

Michael Beer
Ioannis A. Kougioumtzoglou
Edoardo Patelli
Siu-Kui Au
Editors

Encyclopedia of Earthquake Engineering

EXTRAS ONLINE

 **Springer**Reference

Encyclopedia of Earthquake Engineering

Michael Beer
Ioannis A. Kougiumtzoglou
Edoardo Patelli • Siu-Kui Au
Editors

Encyclopedia of Earthquake Engineering

With 2431 Figures and 278 Tables

 **Springer** Reference

Editors

Michael Beer
Institute for Computer Science in
Civil Engineering
Gottfried Wilhelm Leibniz
University Hannover
Hannover, Germany

Ioannis A. Kougiumtzoglou
Department of Civil Engineering &
Engineering Mechanics
Columbia University
New York, NY, USA

Edoardo Patelli
Institute for Risk & Uncertainty
and Centre for Engineering
Sustainability
Liverpool, UK

Siu-Kui Au
Institute for Risk & Uncertainty
and Centre for Engineering
Dynamics
Liverpool, UK

ISBN 978-3-642-35343-7 ISBN 978-3-642-35344-4 (eBook)
ISBN 978-3-642-35345-1 (print and electronic bundle)
DOI 10.1007/978-3-642-35344-4

Library of Congress Control Number: 2015946601

Springer Heidelberg New York Dordrecht London

© Springer-Verlag Berlin Heidelberg 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer-Verlag GmbH Berlin Heidelberg is part of Springer Science+Business Media (www.springer.com)

Preface

The scope of the *Encyclopedia of Earthquake Engineering* covers the interaction between earthquake events and our engineering installations and infrastructures. It is expected to range over buildings, foundations, underground constructions, lifelines and bridges, roads, embankments, and slopes. Although a plethora of references exist in the context of treating/addressing individual earthquake engineering topics, there is no literature dealing with earthquake engineering in a comprehensive, versatile, and unified manner. In this regard, the extreme event of an earthquake has a multifaceted impact on a variety of activities. These include day-to-day operations of public and private services which have greatly increased their exposure to risk.

The Encyclopedia is designed to inform technically inclined readers about the ways in which earthquakes can affect engineering installations and infrastructures and how engineers would go about designing against, mitigating, and remediating these effects. It is also designed to provide cross-disciplinary and cross-domain information to domain experts. Specifically, the proposed work introduces a coupling between traditional topics of earthquake engineering, such as geotechnical/structural engineering, topics of broader interest such as geophysics, and topics of current and emerging value to industrial applications, such as risk management. In this regard, risk management is included in a comprehensive manner, which addresses the dimension and complexity of earthquake hazards. This elucidates the vital connection of the technical contents to the societal context. The main benefit of the Encyclopedia is its breadth of coverage to provide quick information on a substantial level to virtually all groups of readers from academia, industry, and the general population who would like to find out more in the area of earthquake engineering.

Overall, this work is a concerted effort to provide with a holistic perspective on earthquake engineering-related issues of recent currency. Its innovative, modular, and continuously updated form facilitates a potent, self-contained, and readily accessible exposition of multi/interdisciplinary elements in the broad field of earthquake engineering, thus enabling the researcher/practitioner/designer to identify links and potential future research themes in an efficient and timely manner.

Acknowledgments

The Editors greatly appreciate the effort made by the section editors, the contributions of the authors related to work of high standards, as well as the determination of the reviewers to preserve high quality levels. Further, the Editors would like to thank Springer for its guidance and leadership on making this great project a reality.

Michael Beer, Dr.-Ing.
Professor and Head
Institute for Computer Science in Civil Engineering
Gottfried Wilhelm Leibniz University Hannover
Hannover, Germany

Ioannis A. Kougioumtzoglou, Ph.D.
Assistant Professor
Department of Civil Engineering & Engineering Mechanics
Columbia University
New York, NY, USA

Edoardo Patelli, Ph.D.
Lecturer in Uncertainty and Engineering
Institute for Risk & Uncertainty and Centre for Engineering Sustainability
Liverpool, UK

Siu-Kui Au, Ph.D.
Ir Professor of Uncertainty, Reliability and Risk
Institute for Risk & Uncertainty and Centre for Engineering Dynamics
Liverpool, UK

About the Editors

Michael Beer Institute for Computer Science in Civil Engineering, Gottfried Wilhelm Leibniz University Hannover, Hannover, Germany

Michael Beer, Dr.-Ing., obtained his degrees in civil engineering from the Technical University of Dresden, Germany. He worked at Rice University as a fellow of the Alexander von Humboldt-Foundation and held a faculty position at National University of Singapore in the Department of Civil and Environmental Engineering. Dr. Beer was a professor and the founding director of the Institute for Risk and Uncertainty in the University of Liverpool. He is now professor and head of the Institute for Computer Science in Civil Engineering at Gottfried Wilhelm Leibniz University Hannover, Germany, as well as a guest professor in Tongji University, Shanghai, China and a part-time professor in the Institute for Risk and Uncertainty, University of Liverpool.

Ioannis A. Kougiumtzoglou Department of Civil Engineering & Engineering Mechanics, Columbia University, New York, NY, USA

Ioannis Kougiumtzoglou, Ph.D., obtained his M.Sc. and Ph.D. degrees from the civil and environmental engineering department of Rice University, Texas, USA. He also holds a five-year Diploma in Civil Engineering from the National Technical University of Athens (NTUA), Greece, and is a professional/chartered civil engineer in Greece. He is currently an assistant professor at the Department of Civil Engineering and Engineering Mechanics at Columbia University, New York, USA.

Edoardo Patelli Institute for Risk & Uncertainty and Centre for Engineering Sustainability, Liverpool, UK

Edoardo Patelli, Ph.D., obtained his degrees in nuclear engineering and Ph.D. in radiation science and technology from the Politecnico di Milano, Italy. He worked at the Institute for Engineering Mechanics, University of Innsbruck, Austria. He is currently a lecturer in uncertainty and engineering at the Institute for Risk and Uncertainty at the University of Liverpool, UK, and the head of computational technology for the EPSRC Centre for Doctoral Training in Risk and Uncertainty Quantification. Patelli is also an honorary member of the National Tsing Hua University, Taiwan.

Siu-Kui Au Institute for Risk & Uncertainty and Centre for Engineering Dynamics, Liverpool, UK

Siu-Kui Au, Ph.D., is a chartered civil engineer in Hong Kong and obtained his B.Eng. and M.Phil. from the Hong Kong University of Science and Technology and Ph.D. from the California Institute of Technology (USA). He is currently a professor of uncertainty, reliability and risk at the Institute for Risk and Uncertainty. Before joining Liverpool he has held faculty positions at the Nanyang Technological University (Singapore) and the City University of Hong Kong. He was a visiting professor at the Tokyo City University (Japan) and Wuhan University (China).

Section Editors

Section: Artificial and Other Sources and Mechanisms

Hans Thybo Department of Geography and Geology, University of Copenhagen, København K, Denmark

Section: Aseismic Design

Dimitrios G. Lignos Civil Engineering, McGill University, Montreal, QC, Canada

Section: Case Histories

Rafael Riddell Departamento Ingeniería Estructural y Geotécnica, Pontificia Universidad Católica de Chile, Santiago, Chile

Section: Computational and Sensing

Muneo Hori Earthquake Research Institute, University of Tokyo, Bunkyo, Tokyo, Japan

Section: Computational Rock Mechanics

Jui-Pin Wang Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong

Section: Construction Techniques

Polat Gulkan Civil Engineering Department, Cankaya University, Ankara, Turkey

Section: Disaster Recovery and Reconstruction and Loss Modeling

Erica Seville Resilient Organisations, University of Canterbury, Sheffield, New Zealand

John Vargo Department of Accounting and Information Systems, University of Canterbury, Christchurch, New Zealand

Section: Geotechnical Engineering

Kok-Kwang Phoon Department of Civil and Environmental Engineering, National University of Singapore, Singapore, Singapore

Hesham El Naggari Geotechnical Research Centre, Department of Civil and Environmental Engineering, The University of Western Ontario, London, ON, Canada

Dimitrios Zekkos Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, USA

Section: Mathematical Tools

Sondipon Adhikari College of Engineering, Swansea University, Swansea, UK

Subhamoy Bhattacharya Chair in Geomechanics, Faculty of Engineering and Physical Sciences, University of Surrey, Guildford, UK

Section: Paleoseismology

Sarah J. Boulton Centre for Research in Earth Sciences, Fitzroy 115, Plymouth University, Plymouth, Devon, UK

Iain Stewart School of Geography, Earth and Environmental Sciences, Plymouth University, Plymouth, Devon, UK

Section: Random Vibration

Pol D. Spanos Department of Mechanical Engineering and Materials Science, Rice University, Houston, TX, USA

Antonina Pirrotta Dipartimento di Ingegneria Civile, Ambientale, Aerospaziale, dei Materiali, Università di Palermo, Palermo, Italy

Section: Reliability and Robustness

Hector A. Jensen Department of Civil Engineering, Santa Maria University, Casilla, Valparaiso, Chile

Section: Remote Sensing

Salvatore Stramondo Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy

Section: Retrofitting and Strengthening

Agathoklis Giaralis City University London, Room: C174, School of Engineering and Mathematical Sciences, London, UK

Andreas Kappos School of Engineering and Mathematical Sciences, City University London, London, UK

Section: Risk Management: Decision Analysis

Ayhan Irfanoglu School of Civil Engineering, Purdue University, West Lafayette, IN, USA

Section: Risk Mitigation Policies and Approaches

Bijan Khazai Center for Disaster Management and Risk Reduction Technology (CEDIM), Karlsruhe Institute of Technology, Karlsruhe, Germany

Section: Seismic Risk Assessment

Fatemeh Jalayer Department of Structures for Engineering and Architecture, University of Naples “Federico II”, Naples, Italy

Carmine Galasso Department of Civil, Environmental and Geomatic Engineering and Institute for Risk and Disaster Reduction University College London, University College London, London, UK

Section: Sensors and Sensor Systems

Jens Havskov Department of Earth Science, University of Bergen, Bergen, Norway

Gerardo Alguacil Física Teórica y del Cosmos, Universidad de Granada, Beiro, Granada, Spain

Section: Structural Analysis for Earthquake-Resistant Design

Charis J. Gantes Institute of Steel Structures, School of Civil Engineering, National Technical University of Athens, Athens, Greece

Section: Structural Health Monitoring

Eleni Chatzi Institute of Structural Engineering ETH Zürich HIL E, Zürich, CH, Switzerland

Costas Papadimitriou Department of Mechanical Engineering, University of Thessaly, Volos, Greece

Geert Lombaert Department of Civil Engineering, KU Leuven, Leuven, Belgium

Section: Tectonic Sources and Mechanisms

Jan Sileny Institute of Geophysics, Academy of Sciences, Praha, Czech Republic

Section: Volcanic Seismology

Silvio De Angelis Earth, Ocean and Ecological Sciences, School of Environmental Sciences, University of Liverpool, Liverpool, UK

Contributors

S. Abhinav Department of Civil Engineering, Indian Institute of Science, Bangalore, Karnataka, India

Nick Ackerley Nanometrics, Inc, Kanata, Ottawa, ON, Canada

Christoph Adam Department of Engineering Science, Unit of Applied Mechanics, University of Innsbruck, Innsbruck, Austria

Sondipon Adhikari College of Engineering, Swansea University, Swansea, UK

Vadim M. Agafonov Center for Molecular Electronics, Moscow Institute of Physics and Technology, Moscow, Russia

John J. Sanchez Aguilar Departamento de Geociencias, Universidad Nacional de Colombia, Bogotá, Colombia

Nopdanai Ajavakom Department of Mechanical Engineering, Chulalongkorn University, Bangkok, Thailand

H. Serdar Akyüz İstanbul Teknik Üniversitesi, Maden Fakültesi, Jeoloji Müh. Bölümü, Ayazağa, İstanbul, Turkey

David Alexander University College London, London, England

N. A. Alexander Civil Engineering Department, University of Bristol, Bristol, UK

Gerardo Alguacil Instituto Andaluz de Geofísica, University of Granada, Granada, Spain

Clive Allen Faculty of Engineering and Built Environment, The University of Newcastle, Callaghan, NSW, Australia

Hernán G. Alvarado Departamento de Geociencias, Universidad Nacional de Colombia-Sede Bogotá, Bogotá, Colombia

Sotiris Argyroudis Department of Civil Engineering, Aristotle University, Thessaloniki, Greece

Aysegul Askan Civil Engineering Department, Middle East Technical University, Ankara, Turkey

Domenico Asprone Department of Structures for Engineering and Architecture, University of Naples “Federico II”, Naples, Italy

Kuvvet Atakan Department of Earth Science, University of Bergen, Bergen, Norway

Siu-Kui Au Institute for Risk & Uncertainty and Centre for Engineering Dynamics, Liverpool, UK

Luis David Avendaño-Valencia Department of Mechanical & Aeronautical Engineering, Stochastic Mechanical Systems and Automation (SMSA) Laboratory, University of Patras, Patras, Greece

Jack Baker Stanford University, Stanford, CA, USA

Richard J. Bathurst GeoEngineering Centre at Queen’s-RMC Civil Engineering Department, Royal Military College of Canada, Kingston, ON, Canada

Josep Batlló Instituto Dom Luiz (IDL), Faculdade de Ciências da Univ. de Lisboa, Lisbon, Portugal

Sarah Beaven Department of Geological Sciences, University of Canterbury, Christchurch, New Zealand and Natural Hazards Research Group, University of Canterbury, Christchurch, New Zealand

André T. Beck Department of Structural Engineering, São Carlos School of Engineering, University of São Paulo, São Carlos, SP, Brazil

Julia Becker GNS Science, Lower Hutt, New Zealand

Danilo Beli Department of Computational Mechanics, Faculty of Mechanical Engineering, UNICAMP, Campinas, SP, Brazil

Andrew F. Bell School of GeoSciences, University of Edinburgh, Edinburgh, UK

Amadeo Benavent-Climent Department of Structural Mechanics and Industrial Constructions, Polytechnic University of Madrid, Madrid, Spain

Fouad Bendimerad Earthquakes and Megacities Initiative, Quezon City, Philippines

Ninfa L. Bennington Department of Geoscience, University of Wisconsin-Madison, Madison, WI, USA

Djillali Benouar University of Bab Ezzouar, Algiers, Algeria

Mounir Khaled Berrah Ecole Nationale Polytechnique, Algiers, Algeria

Sanjaya Bhatia UNISDR Recovery Platform, Kobe, Japan

Subhamoy Bhattacharya University of Surrey, Guildford, UK

Christian Bignami Istituto Nazionale di Geofisica e Vulcanologia, National Earthquake Center, Rome, Italy

Mustafa Bilal Civil Engineering Department, Middle East Technical University, Ankara, Turkey

Daniel Binder Zentralanstalt für Meteorologie und Geodynamik, Vienna, Austria

Gian Maria Bocchini Institute of Geodynamics, National Observatory of Athens, Athens, Greece

Teddy Boen PT Teddy Boen Konsultan, Jakarta, Indonesia

I. Bondár Research Centre for Astronomy and Earth Sciences of the Hungarian Academy of Sciences, Budapest, Hungary

Franco Bontempi Department of Structural and Geotechnical Engineering, Sapienza University of Rome, Rome, Italy

Moses Kent Borinaga Earthquakes and Megacities Initiative, Quezon City, Philippines

Peter Bormann Formerly GFZ German Research Center for Geosciences, Potsdam, Germany

Jitendra Kumar Bothara NSET, Kathmandu, Nepal and Miyamoto Impact, Christchurch, New Zealand

Sarah J. Boulton Plymouth University, Centre for Research in Earth Sciences, Plymouth University, Devon, UK

Nouredine Bourahla Civil Engineering Department, University Saâd Dahlab, Blida, Algeria

Francesca Bovolo Fondazione Bruno Kessler (FBK), Povo, Trento, Italy

Enrico Brandmayr Department of Mathematics and Geosciences, University of Trieste, Trieste, Italy and The Abdus Salam International Centre for Theoretical Physics, SAND Group, Trieste, Italy

Jochen Braunmiller University of South Florida, Tampa, FL, USA

Ewald Brückl Department of Geodesy and Geoinformation, TU Wien, Vienna, Austria

Marco Breccolotti Department of Civil and Environmental Engineering, University of Perugia, Perugia, Italy

Florent Brenguier Institut des Sciences de la Terre, University of Grenoble, Grenoble, France

Christopher Bronk Ramsey Oxford Radiocarbon Accelerator Unit, University of Oxford, Oxford, UK

Charlotte Brown Christchurch Polytechnic Institute of Technology, Christchurch, New Zealand

David Brunson Kestrel Group Ltd., Wellington, New Zealand

Lorenzo Bruzzone Department of Information Engineering and Computer Science, University of Trento, Povo, Trento, Italy

Christian Bucher Institute of Building Construction and Technology, Vienna University of Technology, Wien, Austria

Pierfrancesco Cacciola School of Environment and Technology, University of Brighton, Brighton, UK

Luigi Callisto Department of Structural and Geotechnical Engineering, Sapienza Università di Roma, Rome, Italy

Luigi Carassale Department of Civil, Chemical and Environmental Engineering, University of Genova, Genoa, Italy

Omar-Dario Cardona Universidad Nacional de Colombia, Manizales, Colombia

Mehmet Çelebi Earthquake Science Center, US Geological Survey, Menlo Park, CA, USA

Yunbyeong Chae Department of Civil and Environmental Engineering, Old Dominion University, Norfolk, VA, USA

Chung-Han Chan Department of Geosciences, National Taiwan University, Taipei, Taiwan, ROC and Earth Observatory of Singapore, Nanyang Technological University, Singapore, Singapore

Ya-Ting Chan Department of Geosciences, National Taiwan University, Taipei, Taiwan

A. A. Chanerley School of Architecture, Computing and Engineering (ACE), University of East London, London, UK and The University of Auckland, Auckland, New Zealand

Alice Yan Chang-Richards The University of Auckland, Auckland, New Zealand

Eleni N. Chatzi Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Institute of Structural Engineering, Zurich, CH, Switzerland

Da-Yi Chen Central Weather Bureau, Taipei, Taiwan

Jianbing Chen School of Civil Engineering & State Key Laboratory for Disaster Reduction in Civil Engineering, Tongji University, Shanghai, China

Chin-Tung Cheng Disaster Prevention Technology Research Center, Sinotech Engineering Consultants, Inc., Taipei, Taiwan, ROC

Tai-Lin Chin Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan

Marco Chini Luxembourg Institute of Science and Technology (LIST), Environmental Research and Innovation Department (ERIN), Belvaux, Luxembourg

Christis Z. Chrysostomou Department of Civil Engineering and Geomatics, Cyprus University of Technology, Limassol, Cyprus

Federico Cluni Department of Civil and Environmental Engineering, University of Perugia, Perugia, Italy

Simona Colombelli Department of Physics, University of Naples Federico II, Federico II – AMRA S.c.ar.l, Napoli, Italy

Joel P. Conte Department of Structural Engineering, University of California at San Diego, CA, USA

Vincenzo Convertito Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Vesuviano, Napoli, Italy

R. Corotis Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, CO, USA

Wayne C. Crawford Institut de Physique du Globe de Paris, Sorbonne Paris Cité, Univ Paris Diderot, UMR 7154 CNRS, Paris, France

Michele Crosetto CTTC Division of Geomatics, Av. C.F. Gauss, Castelldefels, Spain

Laura D'Amico School of Environment and Technology, University of Brighton, Brighton, UK

James Edward Daniell Geophysical Institute, Karlsruhe Institute of Technology, Karlsruhe, Germany

Shideh Dashti Department of Civil, Environmental, and Architectural Engineering, University of Colorado at Boulder, Boulder, CO, USA

Craig A. Davis Waterworks Engineer, Los Angeles Department of Water and Power, Los Angeles, CA, USA

Dina D'Ayala Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK

Silvio De Angelis Earth, Ocean and Ecological Sciences, School of Environmental Sciences, University of Liverpool, Liverpool, UK

José Roberto de França Arruda Department of Computational Mechanics, Faculty of Mechanical Engineering, UNICAMP, Campinas, SP, Brazil

Flavia De Luca Department of Civil Engineering, University of Bristol, Bristol, UK

Raffaele De Risi Department of Structures for Engineering and Architecture, University of Naples Federico II, Naples, Italy

Guido De Roeck Department of Civil Engineering, KU Leuven, Leuven, Belgium

A. Deraemaeker FNRS Research Associate, Building Architecture and Town Planning (BATir), Brussels, Belgium

Mario Di Paola Dipartimento di Ingegneria Civile, Ambientale, Aerospaziale, dei Materiali (DICAM), Università di Palermo, Palermo, Italy

Amod Mani Dixit NSET, Kathmandu, Nepal

Mustapha Djafour Risam Laboratory, Faculty of Technology, University Abou Bakr Belkaïd, Tlemcen, Algeria

Matjaž Dolšek Faculty of Civil and Geodetic Engineering, University of Ljubljana, Ljubljana, Slovenia

Baiping Dong Department of Civil and Environmental Engineering, Lehigh University, Bethlehem, PA, USA

Doug Dreger Department of Earth and Planetary Science, College of Letters and Science, University of California, Berkeley, CA, USA

Hossein Ebrahimian Department of Structures for Engineering and Architecture, University of Naples Federico II, Naples, Italy

Páll Einarsson Institute of Earth Sciences, University of Iceland, Reykjavík, Iceland

Göran Ekström Department of Earth and Environmental Sciences, Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA

Hany El Naggar Department of Civil and Resource Engineering, Dalhousie University, Halifax, NS, Canada

Gaetano Elia School of Civil Engineering and Geosciences, Newcastle University, Newcastle Upon Tyne, UK

E. Elwood Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, CO, USA

Antonio Emolo Department of Physics, University of Naples Federico II, Federico II – AMRA S.c.ar.l, Napoli, Italy

E. R. Engdahl Department of Physics, University of Colorado, Boulder, CO, USA

Murat Altug Erberik Civil Engineering Department, Middle East Technical University, Ankara, Turkey

Niki Evelpidou Faculty of Geology and Geoenvironment/National and Kapodistrian University of Athens, Athens, Greece

Licia Faenza Istituto Nazionale di Geofisica e Vulcanologia, Centro Nazionale Terremoti, Bologna, Italy

Giuseppe Failla Dipartimento di Ingegneria Civile, dell'Energia, dell'Ambiente e dei Materiali, (DICEAM), University of Reggio Calabria, Reggio Calabria, Italy

Michael N. Fardis Department of Civil Engineering, School of Engineering, University of Patras, Patras, Greece

Spilios D. Fassois Department of Mechanical & Aeronautical Engineering, Stochastic Mechanical Systems and Automation (SMSA) Laboratory, University of Patras, Patras, Greece

Gaetano Festa Department of Physics, University of Naples Federico II, Federico II – AMRA S.c.ar.l, Napoli, Italy

Christopher Corey Fischer Wright State University, Dayton, OH, USA

Tomáš Fischer Faculty of Science, Charles University in Prague, Prague, Czech Republic

Steven L. Forman Department of Geology, Baylor University, Waco, TX, USA

G. Fornaro Institute for Electromagnetic Sensing of the Environment, National Research Council, Naples, Italy

G. R. Foulger Department of Geological Sciences, Durham University, Durham, UK

Michalis Fragiadakis School of Civil Engineering, Laboratory for Earthquake Engineering, National Technical University of Athens (N.T.U.A.), Athens, Greece

Tinu Rose Francis Department of Civil and Environmental Engineering, The University of Auckland, Auckland, New Zealand

Guillermo Franco Head of Catastrophe Risk Research – EMEA, Guy Carpenter & Company Ltd., London, UK

Fabio Del Frate Department of Civil Engineering and Computer Science Engineering, University of Tor Vergata, Rome, Italy

Kohei Fujita Department of Architecture and Architectural Engineering, Graduate School of Engineering, Kyoto University, Kyoto, Japan and RIKEN Advanced Institute for Computational Science, Kobe, Japan

N. Ganesh Department of Applied Mechanics, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India

Charis J. Gantes Institute of Steel Structures, School of Civil Engineering, National Technical University of Athens, Zografou Campus, Athens, Greece

Alexander Garcia-Aristizabal Center for the Analysis and Monitoring of Environmental Risk (AMRA), Naples, Italy

Paolo Gasparini Center for the Analysis and Monitoring of Environmental Risk (AMRA), Naples, Italy

Vincenzo Gattulli DICEAA - Dipartimento di Ingegneria Civile, Edile-Architettura, Ambientale, CERFIS - Centro di Ricerca e Formazione in Ingegneria Sismica, University of L'Aquila, L'Aquila, Italy

Lind S. Gee Albuquerque Seismological Laboratory, U.S. Geological Survey, Albuquerque, NM, USA

Eric L. Geist U.S. Geological Survey, Menlo Park, CA, USA

Carmelo Gentile Department ABC, Politecnico di Milano, Milan, Italy

Debraj Ghosh Department of Civil Engineering, Indian Institute of Science, Bangalore, Karnataka, India

Sonia Giovinnazzi Department of Civil and Natural Resources Engineering, University of Canterbury, Christchurch, New Zealand

Konstantinos Gkoumas Department of Structural and Geotechnical Engineering, Sapienza University of Rome, Rome, Italy

Katsuichiro Goda Department of Civil Engineering, University of Bristol, Bristol, UK

Tatiana Goded GNS Science, Lower Hutt, New Zealand

Marco Götz Institute for Structural Analysis, Technische Universität Dresden, Dresden, Germany

Wolfgang Graf Institute for Structural Analysis, Technische Universität Dresden, Dresden, Germany

Ramana V. Grandhi 210 Russ Engineering Center, Wright State University, Dayton, OH, USA

Damian N. Grant Arup Advanced Technology and Research, London, UK

Michael Gray Cast Connex Corporation, Toronto, ON, Canada

Rebekah Green University of Western Washington, Bellingham, WA, USA

Mircea Grigoriu Civil and Environmental Engineering, School of Civil & Environmental Engineering, Cornell University, Ithaca, NY, USA

Luca Guerrieri Department of Geological Survey, ISPRA, Istituto Superiore per la Protezione e la Ricerca Ambientale, Rome, Italy

P. Gülkan Department of Civil Engineering, Çankaya University, Ankara, Turkey

Manu Gupta SEEDS of India, Delhi, India

Sayan Gupta Department of Applied Mechanics, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India

Vittorio Gusella Department of Civil and Environmental Engineering, University of Perugia, Perugia, Italy

Amir M. Halabian Department of Civil Engineering, Isfahan University of Technology, Isfahan, Iran

Achintya Haldar Department of Civil Engineering and Engineering Mechanics, University of Arizona, Tucson, AZ, USA

Matthew M. Haney U.S. Geological Survey, Alaska Volcano Observatory, Anchorage, AK, USA

Youssef M. A. Hashash Department of Civil and Environmental Engineering, 2230c Newmark Civil Engineering Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Jens Havskov Department of Earth Science, University of Bergen, Bergen, Norway

Mark B. Hayman Nanometrics Inc., Ottawa, Canada

Eric M. Hernandez Department of Civil and Environmental Engineering, College of Engineering and Mathematical Sciences, University of Vermont, Burlington, VT, USA

William T. Holmes Rutherford + Chekene Engineers, San Francisco, CA, USA

Andrew Hooper School of Earth and Environment, The University of Leeds, Maths/Earth and Environment Building, Leeds, UK

Josef Horálek Institute of Geophysics, Academy of Sciences of the Czech Republic, Prague, Czech Republic

Muneo Hori Earthquake Research Institute, University of Tokyo, Bunkyo, Tokyo, Japan

Shigeki Horiuchi COE, Home Seismometer Corporation, Shirakawa, Fukushima, Japan

Alicia J. Hotovec-Ellis Department of Earth and Space Sciences, University of Washington, Seattle, WA, USA

Roman Hryciw Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, USA

Nai-Chi Hsiao Central Weather Bureau, Taipei, Taiwan

Pao-Shan Hsieh Disaster Prevention Technology Research Center, Sinotech Engineering Consultants, Inc., Taipei, Taiwan, ROC

Charles Robert Hutt Albuquerque Seismological Laboratory, U.S. Geological Survey, Albuquerque, NM, USA

Luis F. Ibarra Department of Civil and Environmental Engineering, University of Utah, Salt Lake City, USA

Tsuyoshi Ichimura Earthquake Research Institute, University of Tokyo, Bunkyo, Tokyo, Japan

Aristidis Iliopoulos Peikko Greece SA, Marousi, Athens, Greece

Ioanna Ioannou Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK

Antonio Iodice Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione, Università degli Studi di Napoli Federico II, Napoli, Italy

Tatjana Isakovic University of Ljubljana, Ljubljana, Slovenia

Radosław Iwankiewicz Institute of Mechanics and Ocean Engineering, Hamburg University of Technology, Hamburg, Germany

Steinunn S. Jakobsdóttir Faculty of Earth Science, University of Iceland, Reykjavík, Iceland

Li-ju Jang Department of Social Work, National Pingtung University of Science and Technology, Pingtung, Taiwan

Hector A. Jensen Department of Civil Engineering, Santa Maria University, Casilla, Valparaiso, Chile

Randall W. Jibson U.S. Geological Survey, Golden, CO, USA

Sarb Johal Joint Centre for Disaster Research, GNS Science/Massey University, Wellington, New Zealand

Jessica H. Johnson School of Earth Sciences, University of Bristol Wills Memorial Building, Bristol, UK and School of Environmental Sciences, University of East Anglia, Norwich, UK

David Johnston Joint Centre for Disaster Research, GNS Science/Massey University, Wellington, New Zealand and Risk and Society, GNS Science, Lower Hutt, New Zealand

Lucy Johnston University of Canterbury, Christchurch, New Zealand

Bruce R. Julian Department of Geological Sciences, Durham University, Durham, UK

Ioannis Kalpakidis KBR, Houston, TX, USA

George S. Kamaris School of Engineering, University of Warwick, Coventry, UK

Viswanath Kammula Structural Engineer-in-Training, Soscia Engineering Limited, Toronto, Canada

Ram Chandra Kandel Toronto, ON, Canada

Andreas Kappos Department of Civil Engineering, City University London, London, UK

Volkan Karabacak Eskişehir Osmangazi Üniversitesi, Mühendislik-Mimarlık Fakültesi, Jeoloji Müh. Bölümü, Meşelik, Eskişehir, Turkey

Dimitris L. Karabalis Department of Civil Engineering, University of Patras, Patras, Greece

Spyros A. Karamanos Department of Mechanical Engineering, University of Thessaly, Volos, Greece

Theodore L. Karavasilis School of Engineering, University of Warwick, Coventry, UK

Toshihide Kashima Building Research Institute, Tsukuba, Japan

Tsuneo Katayama Professor emeritus of the University of Tokyo, Bunkyo-ku, Tokyo, Japan

Edward Kavazanjian, Jr. School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, USA

Miklós Kázmér Department of Palaeontology, Eötvös University, Budapest, Hungary

Ilan Kelman Institute for Risk and Disaster Reduction and Institute for Global Health, University College London, London, England and Norwegian Institute of International Affairs, Oslo, Norway

Bijan Khazai Karlsruhe Institute of Technology, Center for Disaster Management and Risk Reduction Technology (CEDIM), Karlsruhe, Germany

Christopher R. J. Kilburn Department of Earth Sciences, Aon Benfield UCL Hazard Centre, University College London, London, UK

Anastasia A. Kiratzi Department of Geophysics, Aristotle University of Thessaloniki, Thessaloniki, Greece

Ulrich Klapp AREVA GmbH, Erlangen, Germany

Takaji Kokusho Department of Civil & Environmental Engineering, Chuo University, Tokyo, Japan

Petros Komodromos Department of Civil and Environmental Engineering, University of Cyprus, Nicosia, Cyprus

Indranil Kongar EPICentre, Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK

Ivan Koulakov Trofimuk Institute of Petroleum Geology and Geophysics, SB RAS, Novosibirsk, Russia and Novosibirsk State University, Novosibirsk, Russia

George P. Kouretzis Faculty of Engineering and Built Environment, The University of Newcastle, Callaghan, NSW, Australia

Ioannis Koutromanos Department of Civil and Environmental Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

D. S. Kusanovic Department of Civil Engineering, Santa Maria University, Casilla, Valparaiso, Chile

Simon Laflamme Department of Civil, Construction, and Environmental Engineering, Iowa State University, Ames, IA, USA

Nikos D. Lagaros Institute of Structural Analysis & Antiseismic Research, Department of Structural Engineering, School of Civil Engineering, National Technical University of Athens, Athens, Greece

Andreas P. Lampropoulos School of Environment and Technology, University of Brighton, Brighton, UK

Roberto Leon The Charles Edward Via, Jr. Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA, USA

Jie Li School of Civil Engineering & State Key Laboratory for Disaster Reduction in Civil Engineering, Tongji University, Shanghai, China

Giorgio Antonino Licciardi GIPSA-Lab – INP Grenoble, Grenoble Institute of Technology, Saint Martin d'Hères, France

Po-Shen Lin Disaster Prevention Technology Research Center, Sinotech Engineering Consultants, Inc., Taipei, Taiwan, ROC

Ting Lin Department of Civil, Construction and Environmental Engineering, Marquette University, Milwaukee, WI, USA

Wen Liu Department of Urban Environment Systems, Chiba University, Chiba, Japan

Geert Lombaert Department of Civil Engineering, KU Leuven, Leuven, Belgium

Rafael H. Lopez Civil Engineering Department, Federal University of Santa Catarina, Florianópolis, SC, Brazil

Santiago López Departamento de Geociencias, Universidad Nacional de Colombia-Sede Bogotá, Bogotá, Colombia

Paulo B. Lourenço Department of Civil Engineering, University of Minho, ISE, Guimarães, Portugal

Xiao Lu School of Civil Engineering, Beijing Jiaotong University, Beijing, People's Republic of China

Xinzheng Lu Key Laboratory of Civil Engineering Safety and Durability of China Education Ministry, Department of Civil Engineering, Tsinghua University, Beijing, People's Republic of China

Björn Lund Department of Earth Sciences, Uppsala University, Uppsala, Sweden

Guido Luzzi CTTC Division of Geomatics, Av. C.F. Gauss, Castelldefels, Spain

Lucia Luzzi Istituto Nazionale di Geofisica e Vulcanologia, Milan, Italy

Fai Ma Department of Mechanical Engineering, University of California, Berkeley, Berkeley, CA, USA

John Hugh George Macdonald Civil Engineering Department, University of Bristol, Bristol, UK

Kristof Maes Department of Civil Engineering, KU Leuven, Leuven, Belgium

Andrea Magrin Department of Mathematics and Geosciences, University of Trieste, Trieste, Italy

Sankaran Mahadevan Department of Civil and Environmental Engineering, Vanderbilt University, Nashville, TN, USA

Ian G. Main School of GeoSciences, University of Edinburgh, Edinburgh, UK

Ljubica Mamula-Seadon Centre for Infrastructure Research at the Department of Civil and Environmental Engineering, Faculty of Engineering, The University of Auckland, Auckland, New Zealand

Gaetano Manfredi Department of Structures for Engineering and Architecture, University of Naples “Federico II”, Naples, Italy

Sandeeka Mannakkara Department of Civil and Environmental Engineering, The University of Auckland, Auckland, New Zealand

C. S. Manohar Department of Civil Engineering, Indian Institute of Science, Bangalore, Karnataka, India

Carlo Marin Department of Information Engineering and Computer Science, University of Trento, Povo, Trento, Italy

Rui Marques Engineering Department, Civil Engineering Section, Pontifical Catholic University of Peru, Lima, Peru

Justin D. Marshall Auburn University, Auburn, AL, USA

Mark J. Masia Faculty of Engineering and Built Environment, The University of Newcastle, Callaghan, NSW, Australia

Neven Matasovic Geosyntec Consultants, Huntington Beach, CA, USA

Annibale Luigi Materazzi Department of Civil and Environmental Engineering, University of Perugia, Perugia, Italy

George P. Mavroidis Department of Civil and Environmental Engineering and Earth Sciences, University of Notre Dame, Notre Dame, IN, USA

Garry McDonald Market Economics Ltd, Takapuna, New Zealand

Hamish McLean School of Humanities, Griffith University, Nathan, QLD, Australia

Nuno Mendes ISISE, University of Minho, Guimarães, Portugal

Stefan Mertl Mertl Research GmbH, Vienna, Austria

Alessandro Maria Michetti Dipartimento di Scienza e Alta Tecnologia, Università dell’Insubria, Como, Italy

Danielle Hutchings Mieler Earthquakes and Hazards Resilience Program, Association of Bay Area Governments, Oakland, CA, USA

Leandro F. F. Miguel Civil Engineering Department, Federal University of Santa Catarina, Florianópolis, SC, Brazil

Jon Mitchell Jon Mitchell Emergency Management Ltd., Queenstown, New Zealand

Chara Ch. Mitropoulou Institute of Structural Analysis & Antiseismic Research, Department of Structural Engineering, School of Civil Engineering, National Technical University of Athens, Athens, Greece

Babak Moaveni Department of Civil and Environmental Engineering, Tufts University, Medford, MA, USA

Tracy Monk Families for School Seismic Safety, Vancouver, Canada

Troy A. Morgan Exponent, New York, USA

Lalliana Mualchin Headquarters: Studio legale Avv. Wania Della Vigna, International Seismic Safety Organization (ISSO), Arsita (TE), Italy

Catherine Murray Market Economics Ltd, Takapuna, New Zealand

Giuseppe Muscolino Dipartimento di Ingegneria Civile, Informatica, Edile, Ambientale e Matematica Applicata, Università degli Studi di Messina, Messina, Italy

S. C. Myers Lawrence Livermore National Laboratory, Livermore, CA, USA

Satish Nagarajaiah Department of Civil and Environmental Engineering, Rice University, Houston, TX, USA and Department of Mechanical Engineering, Rice University, Houston, TX, USA

Hesham El Naggar Geotechnical Research Centre, Department of Civil and Environmental Engineering, The University of Western Ontario, London, ON, Canada

Alexander V. Neeshpapa R-sensors LLC, Dolgoprudny, Russia

Meredith Nettles Department of Earth and Environmental Sciences, Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA

Ching Hang Ng US Nuclear Regulatory Commission, Office of Nuclear Reactor Regulation, Washington, DC, USA

Nikolaos Nikitas School of Civil Engineering, University of Leeds, Leeds, UK

Nicola Nisticò Dipartimento di Ingegneria Strutturale e Geotecnica, Università La Sapienza, Rome, Italy

Viviana Novelli Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK

Ilan Noy School of Economics and Finance, Victoria University, Wellington, New Zealand

David D. Oglesby Department of Earth Sciences, University of California, Riverside, Riverside, CA, USA

Kenji Oguni Department of System Design Engineering, Keio University, Yokohama, Japan

Norio Okada Graduate School of Science and Engineering, Kumamoto University, Kumamoto City, Japan

Izuru Okawa Building Research Institute, Tsukuba, Japan

Scott M. Olson Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Rolando P. Orense Department of Civil and Environmental Engineering, University of Auckland, Auckland, New Zealand

Lars Ottemöller Department of Earth Science, University of Bergen, Bergen, Norway

Daria Ottonelli Department of Civil, Chemical and Environmental Engineering, University of Genoa, Genoa, Italy

Francesca Pacor Istituto Nazionale di Geofisica e Vulcanologia, Milan, Italy

Sara Paganoni Ziegert|Roswag|Seiler Architekten Ingenieure, Berlin, Germany

Alessandro Palmeri School of Civil and Building Engineering, Loughborough University, Loughborough, Leicestershire, UK

Bishnu Pandey University of British Columbia, Vancouver, Canada

Chris P. Pantelides Department of Civil and Environmental Engineering, University of Utah, Salt Lake City, UT, USA

Giuliano F. Panza Department of Mathematics and Geosciences, University of Trieste, Trieste, Italy and The Abdus Salam International Centre for Theoretical Physics, SAND Group, Trieste, Italy and Institute of Geophysics, China Earthquake Administration, Beijing, China and International Seismic Safety Organization (ISSO), Arsita, Italy

C. Papadimitriou Department of Mechanical Engineering, University of Thessaly, Volos, Greece

Vissarion Papadopoulos Institute of Structural Analysis and Seismic Research, School of Civil Engineering, National Technical University of Athens (N.T.U.A.), Athens, Greece

Manolis Papadrakakis Institute of Structural Analysis & Antiseismic Research, Department of Structural Engineering, School of Civil Engineering, National Technical University of Athens, Athens, Greece

Thanasis Papageorgiou Department of Structural Engineering, School of Civil Engineering, National Technical University of Athens (N.T.U.A.), Athens, Greece

Ioanna Papayianni Laboratory of Building Materials, Department of Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

Byeongjin Park Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology (KAIST), Yuseong-Gu, Daejeon, Republic of Korea

Douglas Paton School of Psychology, University of Tasmania, Launceston, TAS, Australia

Manuel Pellisetti AREVA GmbH, Erlangen, Germany

George Gr. Penelis Civil Engineering Department, Aristotle University of Thessaloniki, Thessaloniki, Greece

Marla Petal Risk RED (Risk Reduction Education for Disasters), Los Angeles, CA, USA

Francesco Petrini Department of Structural and Geotechnical Engineering, Sapienza University of Rome, Rome, Italy

Matteo Picchiani Department of Civil Engineering and Computer Science Engineering, University of Tor Vergata, Rome, Italy

Paolo Pirazzoli Laboratoire de Géographie Physique, Paris, France

Kyriazis Pitolakis Department of Civil Engineering, Aristotle University, Thessaloniki, Greece

Stephen Platt Cambridge Architectural Research Ltd, Cambridge, UK

Nikos Pneumatikos Department of Civil Engineering, Surveying and Geoinformatics, Technological Educational Institute of Athens, Egaleo-Athens, Greece

Panayiotis C. Polycarpou Department of Civil and Environmental Engineering, University of Cyprus, Nicosia, Cyprus

Silvia Pondrelli Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Italy

Keith Porter Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder and SPA Risk LLC, Denver, CO, USA

Regan Potangaroa Unitec School of Architecture, Auckland, New Zealand

Francesco Potenza DICEAA - Dipartimento di Ingegneria Civile, Edile-Architettura, Ambientale, CERFIS - Centro di Ricerca e Formazione in Ingegneria Sismica, University of L'Aquila, L'Aquila, Italy

Alessandro Proia Dipartimento di Ingegneria Strutturale e Geotecnica, Università La Sapienza, Rome, Italy

Carsten Proppe Institut für Technische Mechanik, Karlsruhe Institute of Technology, Karlsruhe, Germany

Ioannis N. Psycharis School of Civil Engineering, Department of Structural Engineering, National Technical University of Athens (N.T.U.A.), Athens, Greece

Wolfgang Rabbel Institute of Geosciences, University of Kiel, Kiel, Germany

Alin C. Radu Civil and Environmental Engineering, School of Civil & Environmental Engineering, Cornell University, Ithaca, NY, USA

Mircea Radulian National Institute for Earth Physics, Măgurele, Romania

D. Reale Institute for Electromagnetic Sensing of the Environment, National Research Council, Naples, Italy

Klaus Reicherter Institute of Neotectonics and Natural Hazards, Department of Geosciences and Geography, RWTH Aachen University, Aachen, Germany

Robert K. Reitherman Consortium of Universities for Research in Earthquake Engineering, Richmond, CA, USA

Edwin Reynders Department of Civil Engineering, KU Leuven, Leuven, Belgium

Sanaz Rezaeian U.S. Geological Survey, Golden, CO, USA

Daniele Riccio Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione, Università degli Studi di Napoli Federico II, Napoli, Italy

James M. Ricles Bruce G. Johnston Professor of Structural Engineering, Department of Civil and Environmental Engineering, Lehigh University, Bethlehem, PA, USA

Adam T. Ringler Albuquerque Seismological Laboratory, U.S. Geological Survey, Albuquerque, NM, USA

Janise Rodgers GeoHazards International, Menlo Park, USA

Maria Ines Romero-Arduz Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Tiziana Rossetto Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK

Daniel Roten San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA, USA

James Olabode Bamidele Rotimi Auckland University of Technology, Auckland, New Zealand

Theodoros Rousakis Department of Civil Engineering, School of Engineering, Democritus University of Thrace, Xanthi, Greece

Masayuki Saeki Department of Civil Engineering, Tokyo University of Science, Shinjuku, Japan

David Saftner Department of Civil Engineering, University of Minnesota
Duluth, Duluth, MN, USA

Antonella Saisi Department ABC, Politecnico di Milano, Milan, Italy

Simone Salimbeni Istituto Nazionale di Geofisica e Vulcanologia, Sezione
di Bologna, Italy

Zeynep Türkmen Sanduvaç Risk RED, Istanbul, Turkey

Shankar Sankararaman Intelligent Systems Division, SGT Inc., NASA
Ames Research Center, Moffett Field, CA, USA

Evangelos Sapountzakis Department of Structural Engineering, School of
Civil Engineering, National Technical University of Athens (N.T.U.A.),
Athens, Greece

Wendy Saunders GNS Science, Lower Hutt, New Zealand

Richard Sause Joseph T. Stuart Professor of Structural Engineering,
Department of Civil and Environmental Engineering, Lehigh University,
Bethlehem, PA, USA

Mechita C. Schmidt-Aursch Alfred-Wegener-Institut, Helmholtz-Zentrum
für Polar- und Meeresforschung, Bremerhaven, Germany

Johannes Schweitzer NORSAR, Kjeller, Norway

Erica Seville Resilient Organisations, Sheffield, New Zealand

Anastasios G. Sextos Division of Structural Engineering, Department of
Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

Anna S. Shabalina Center for Molecular Electronics, Moscow Institute of
Physics and Technology, Moscow, Russia

Ayman A. Shama Ammann & Whitney, New York, NY, USA

Nikolay Shapiro Institut de Physique du Globe de Paris Laboratoire de
Sismologie, Paris, France

Rajib Shaw University of Kyoto, Kyoto, Japan

P. Benson Shing Department of Structural Engineering, University of
California, San Diego, La Jolla, CA, USA

Nilesh Shome Model Development, Risk Management Solutions, Newark,
CA, USA

Pablo G. Silva Departamento de Geología, Universidad de Salamanca,
Escuela Politécnica Superior de Ávila, Avila, Spain

Priscilla Brandão Silva Department of Computational Mechanics, Faculty
of Mechanical Engineering, UNICAMP, Campinas, SP, Brazil

E. Simoen Department of Civil Engineering, KU Leuven, Leuven, Belgium

Manuel Sintubin Department of Earth and Environmental Sciences,
Geodynamics and Geofluids Research Group, KU Leuven, Leuven, Belgium

Reinoud Sleeman Seismology Division, Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands

Nicola Smith Market Economics Ltd, Takapuna, New Zealand

Patrick Smith Montserrat Volcano Observatory, Flemmings, Montserrat and Seismic Research Centre, University of the West Indies, Trinidad and Tobago, West Indies

Richard Smith Science and Education, Earthquake Commission, Wellington, New Zealand

Andrew W. Smyth Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY, USA

Emily So Department of Architecture, University of Cambridge, Cambridge, UK

Hoon Sohn Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology (KAIST), Yuseong-Gu, Daejeon, Republic of Korea

Minas D. Spiridonakos Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Institute of Structural Engineering, Zurich, CH, Switzerland

Srinivas Sriramula Lloyd's Register Foundation (LRF) Centre for Safety and Reliability Engineering, School of Engineering, University of Aberdeen, Fraser Noble Building, Aberdeen, UK

George Stefanou Institute of Structural Analysis & Antiseismic Research, School of Civil Engineering, National Technical University of Athens, Athens, Greece and Institute of Structural Analysis & Dynamics of Structures, Department of Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

Mark Stirling GNS Science, Lower Hutt, New Zealand

Dieter Stoll Lennartz Electronic GmbH, Tübingen, Germany

Salvatore Stramondo National Earthquake Center, Remote Sensing Lab, Istituto Nazionale di Geofisica e Vulcanologia, Rome, Italy

Kosmas-Athanasios Stylianidis Civil Engineering Department, Aristotle University of Thessaloniki, Thessaloniki, Greece

Xiaodan Sun Southwest Jiaotong University, Chengdu, China

V. S. Sundar Department of Civil Engineering, Indian Institute of Science, Bangalore, Karnataka, India

Ricardo Taborda Department of Civil Engineering, and Center for Earthquake Research and Information, University of Memphis, Memphis, TN, USA

Izuru Takewaki Department of Architecture and Architectural Engineering, Graduate School of Engineering, Kyoto University, Kyoto, Japan

T. P. Tassios Department of Structural Engineering – Reinforced Concrete Laboratory, National Technical University of Athens (N.T.U.A.), Zografou, Athens, Greece

Georgia E. Thermou Department of Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

Geoff Thomas School of Architecture, Victoria University of Wellington, Wellington, New Zealand

Glenn Thompson School of Geosciences, University of South Florida, Tampa, FL, USA

Clifford Thurber Department of Geoscience, University of Wisconsin-Madison, Madison, WI, USA

Lucia Tirca Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC, Canada

G. Tondreau Postdoctoral Researcher, Building Architecture and Town Planning (BATir), Brussels, Belgium

Ikuo Towhata Department of Civil Engineering, University of Tokyo, Bunkyo-ku, Tokyo, Japan

Bruce Townsend Nanometrics, Inc., Kanata, Ontario, Canada

Konstantinos Daniel Tsavdaridis School of Civil Engineering, University of Leeds, Leeds, UK

Yiannis Tsompanakis Computational Dynamics Research Group, School of Environmental Engineering, Technical University of Crete, Chania, Crete, Greece

Angelos S. Tzimas School of Engineering, University of Warwick, Coventry, UK

Marcos Valdebenito Department of Civil Engineering, Santa Maria University, Casilla, Valparaiso, Chile

Suzanne Vallance Lincoln University, Christchurch, New Zealand

Dimitrios Vamvatsikos School of Civil Engineering, National Technical University of Athens (N.T.U.A.), Athens, Greece

John Vargo Resilient Organisations, University of Canterbury, Christchurch, New Zealand

Nick Varley Facultad de Ciencias, Centre of Exchange and Research in Volcanology, Universidad de Colima, Colima, Mexico

Graça Vasconcelos Department of Civil Engineering, ISISE, University of Minho, Guimarães, Portugal

George Vasdravellis Institute for Infrastructure and Environment, Heriot-Watt University, Edinburgh, UK

Marcello Vasta INGEO, Engineering and Geology Department, University of Chieti-Pescara “G. D’Annunzio”, Pescara, Italy

Maria Vathi Department of Mechanical Engineering, University of Thessaly, Volos, Greece

Václav Vavryčuk Institute of Geophysics, Czech Academy of Sciences, Prague, Czech Republic

Ioannis Vayas Traffic Engineering Laboratory, National Technical University of Athens (N.T.U.A.), Athens, Greece

Gerardo M. Verderame Department of Structures for Engineering and Architecture (DiSt), University of Naples Federico II, Naples, Italy

Elizabeth Vintzileou Department of Structural Engineering, Faculty of Civil Engineering, National Technical University of Athens, Athens, Greece

Peter H. Voss Geological Survey of Denmark and Greenland – GEUS, Copenhagen K, Denmark

Christos Vrettos Division of Soil Mechanics and Foundation Engineering, Technical University of Kaiserslautern, Kaiserslautern, Germany

Gregory P. Waite Department of Geological and Mining Engineering and Sciences, Michigan Technological University, Houghton, MI, USA

Jui-Pin Wang Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong

Kai-Shyr Wang Ministry of Science and Technology, Taipei, Taiwan

Friedemann Wenzel Geophysical Institute, Karlsruhe Institute of Technology, Karlsruhe, Germany

Lalith Wijerathne Earthquake Research Institute, University of Tokyo, Tokyo, Japan

Dennis Wilken Institute of Geosciences, University of Kiel, Kiel, Germany

Suzanne Wilkinson Department of Civil and Environmental Engineering, The University of Auckland, Auckland, New Zealand

Thomas Wilson Department of Geological Sciences, University of Canterbury, Christchurch, New Zealand and Natural Hazards Research Group, University of Canterbury, Christchurch, New Zealand

Ben Wisner University College London, London, England and Oberlin College, Oberlin, OH, USA

Guoxi Wu Wutec Geotechnical International, New Westminster, Metro Vancouver, BC, Canada

Yih-Min Wu Department of Geosciences, National Taiwan University, Taipei, Taiwan

Zhen Xu Key Laboratory of Civil Engineering Safety and Durability of China Education Ministry, Department of Civil Engineering, Tsinghua University, Beijing, People's Republic of China

Fumio Yamazaki Department of Urban Environment Systems, Chiba University, Chiba, Japan

Yongchao Yang Department of Civil and Environmental Engineering, Rice University, Houston, TX, USA

Yin-Tung Yen Disaster Prevention Technology Research Center, Sinotech Engineering Consultants, Inc., Taipei, Taiwan, ROC

Ka-Veng Yuen Faculty of Science and Technology, University of Macau, Macau, China

Daniil Yurchenko Institute of Mechanical, Process and Energy Engineering, Heriot-Watt University, Edinburgh, UK

Cengiz Zabcı İstanbul Teknik Üniversitesi, Maden Fakültesi, Jeoloji Müh. Bölümü, Ayazağa, İstanbul, Turkey

Jerome Zayas Earthquakes and Megacities Initiative, Quezon City, Philippines

Djawad Zendagui Risam Laboratory, Faculty of Technology, University Abou Bakr Belkaïd, Tlemcen, Algeria

Christos A. Zeris Department of Structural Engineering, Faculty of Civil Engineering, National Technical University of Athens, Athens, Greece

Enrico Zio European Foundation for New Energy–Electricité de France, Ecole Centrale Paris and Supélec, Paris, France and Energy Department, Politecnico di Milano, Milan, Italy

Aldo Zollo Department of Physics, University of Naples Federico II, Federico II – AMRA S.c.ar.l, Napoli, Italy

Konstantin M. Zuev Department of Computing and Mathematical Sciences, California Institute of Technology, USA

Actively and Semi-actively Controlled Structures Under Seismic Actions: Modeling and Analysis

Nikos Pnevmatikos¹ and Charis J. Gantes²

¹Department of Civil Engineering, Surveying and Geoinformatics, Technological Educational Institute of Athens, Egaleo-Athens, Greece

²Institute of Steel Structures, School of Civil Engineering, National Technical University of Athens, Zografou Campus, Athens, Greece

Synonyms

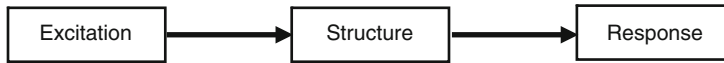
Active control; Control algorithms; Hybrid control; Semi-active control; Structural control

Introduction

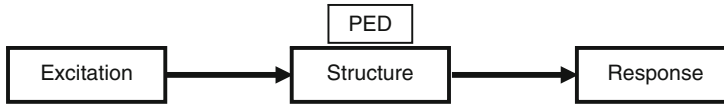
Structural control of seismically excited buildings and other civil structures has attracted considerable attention in recent years. The objective of this entry is to present how seismic design of structures can benefit from the structural control concepts and applications. For this purpose, control theory as applied in other engineering disciplines is adjusted and appropriately modified, where needed, in order to propose integrated control procedures suitable for civil structures subjected to earthquake excitation.

Two approaches can be taken to help buildings withstand seismic excitations. The first involves designing the structure with sufficient strength, stiffness, and inelastic deformation capacity to withstand an earthquake. The choice of material used in construction and the soil beneath the structure are important factors that influence structural vibration and the amount of damage. Because this approach relies on the inherent strength of the structure to dissipate the seismic energy, a certain level of inelastic deformation and associated damage has to be accepted. The second approach relies on using control devices in order to reduce the forces acting on the structure, aiming at reducing all quantities of structural response, that is, floor accelerations, velocities, and displacements. Control systems are categorized according to their energy requirements as passive, active, semi-active, and hybrid. In order to apply a control approach, appropriate control devices are needed. These devices are capable of altering the dynamic characteristics of a structure in real time or applying direct or indirect control forces to the structure, in order to reduce its response. These devices operate as instructed by suitably designed control algorithms.

Several available control algorithms are applied to control structures, many of which were developed by researchers in other fields like electrical or mechanical engineering. However, electrical and mechanical devices are

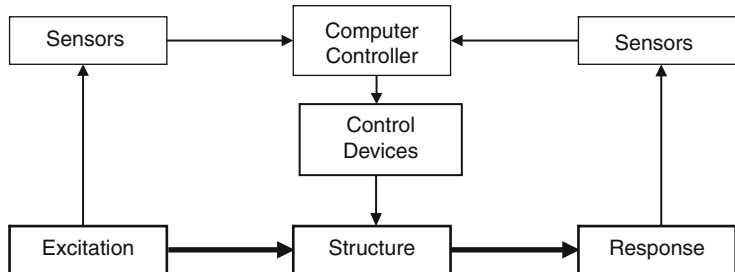


Actively and Semi-actively Controlled Structures Under Seismic Actions: Modeling and Analysis,
Fig. 1 Conventional structure and the response under the seismic excitation

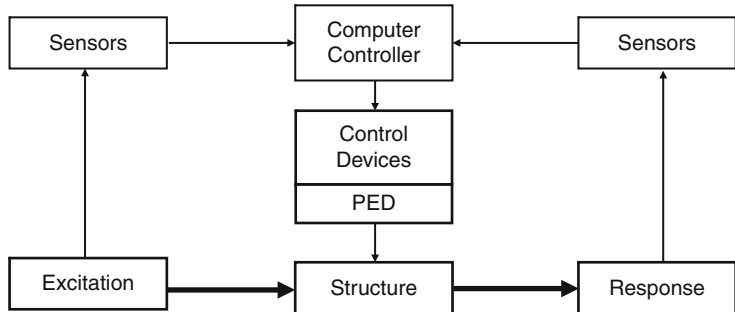


Actively and Semi-actively Controlled Structures Under Seismic Actions: Modeling and Analysis,
Fig. 2 Structure with passive energy dissipation devices (PED)

Actively and Semi-actively Controlled Structures Under Seismic Actions: Modeling and Analysis,
Fig. 3 Structure with active control devices



Actively and Semi-actively Controlled Structures Under Seismic Actions: Modeling and Analysis,
Fig. 4 Structure with semi-active control devices



different from buildings as far as their structural behavior is concerned. The former, in most cases, are mechanisms, while the latter are constructions with a high degree of redundancy. Moreover, in the control of electromechanical devices, the loading is known a priori and is usually harmonic, while for buildings the earthquake loading is unknown and contains a multitude of frequencies. Thus, there is a need of selecting among existing algorithms those that are suitable for the control of buildings and then modifying them accordingly. Along these lines, a common feature of all control strategies that are presented here is that they are applied to structures excited at their base by incoming earthquake waves.

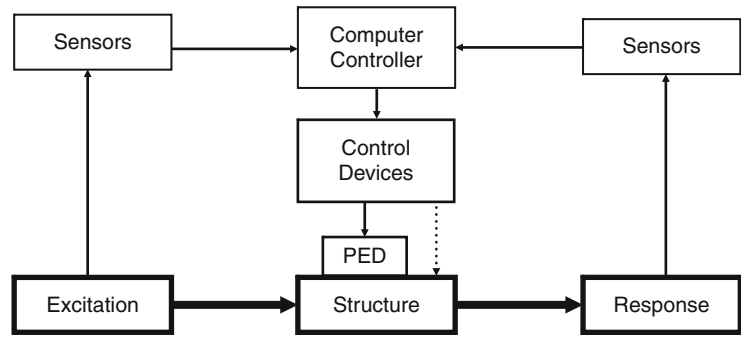
Structural Control (Passive, Active, Semi-active, Hybrid), Corresponding Control Devices, and Practical Applications

Structural control systems fall into four basic categories: passive, active, semi-active, and hybrid control (Soong 1990; Soong and Spencer 2002). These structural control systems are presented in Figs. 1, 2, 3, 4, and 5 below.

Passive Control

Passive control devices are devices that do not require power to operate. Examples of passive devices are base isolation, tuned mass dampers (TMD), tuned liquid dampers (TLD), metallic yield dampers, viscous fluid dampers, and

Actively and Semi-actively Controlled Structures Under Seismic Actions: Modeling and Analysis,
Fig. 5 Structure with hybrid control devices



friction dampers. They dissipate energy using the motion of the structure to produce relative movement within the control device or to alter the dynamic properties of the structure (damping, natural frequencies), so that the earthquake action will be minimized. Since they do not inject energy into the system, they are stable devices. Another advantage of such devices is their low maintenance requirements and the fact that they are unaffected by potential interruptions in power supply. These systems are well understood and well accepted by the engineering community as a means for mitigating the effects of dynamic loadings, such as strong earthquakes and high winds. However, such passive devices have the limitation of not being able to adapt to structural changes and to different earthquake excitations. Active, semi-active, and hybrid control systems aim at addressing these shortcomings.

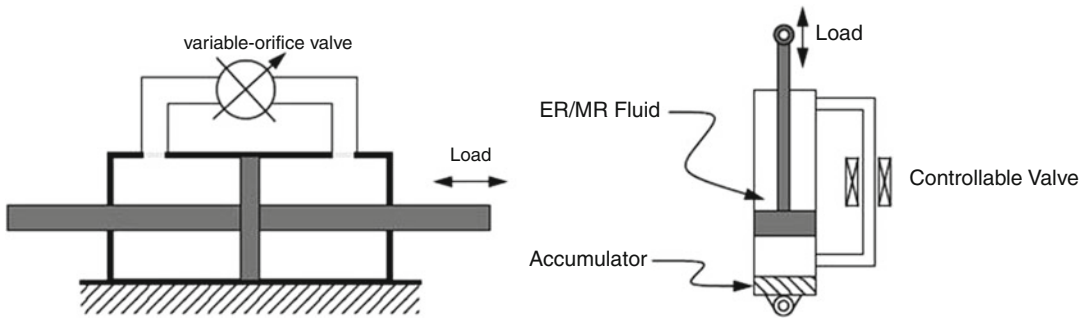
Active Control

Active control strategies have been developed in the 1990s, (Soong 1990; Housner et al. 1997); they operate by using external energy supplied by actuators to impart forces on the structure. The appropriate control action is determined based on measurements of the structural response. Active control devices include the active tendon system (Abdel-Rohman and Leipholz 1983), the active bracing system (Reinhorn et al. 1989), and the active tuned mass damper (Abdel-Rohman and Leipholz 1983).

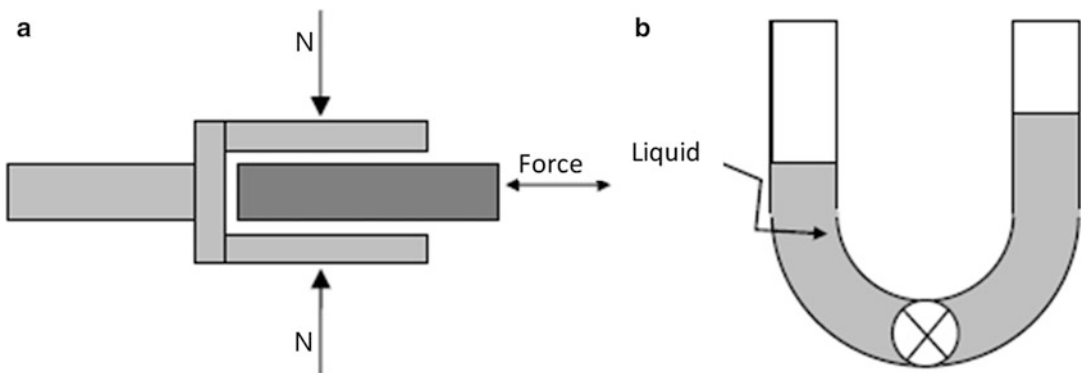
The most famous active control device is the active mass damper (AMD), which uses a mass–spring–damper system combined with

an actuator that moves the mass as needed to increase the amount of damping and the operational frequency range of the device. The first implementation of this control method, and of active control in general, was performed in 1989 in the Kyobashi Seiwa building, Tokyo, Japan, by Kajima Corporation (Kobori et al. 1991). Other applications of such devices include the Applause Tower (Hankyu Chayamachi building) in Osaka, Japan; the Riverside Sumida Central Tower, in Tokyo, Japan; the Nanjing Communication Tower, in Nanjing, China; and the Shin-Jei building in Taipei, Taiwan (Spencer and Nagarajaiah 2003).

Active control devices require considerable amount of external power to operate actuators that supply a control force to the structure. Such power may not always be available during seismic events. Another drawback is that due to their capacity to add energy to the system, they may destabilize it. Cost and maintenance of such systems is also significantly higher than that of passive devices. On the other hand, they are more effective than passive devices because of their ability to adapt to different loading conditions and to control different modes of vibration. Housner et al. (1997) point out the importance of system integration in the design and development of active control systems. Not only is it necessary to consider the individual components of a control system, but the system as a whole must be understood, including the structure, control devices, sensors, and computer control system. Błachowski (2007) uses model-based predictive control to reduce the vibration for guyed mast.



Actively and Semi-actively Controlled Structures Under Seismic Actions: Modeling and Analysis,
Fig. 6 Variable-orifice damper and controllably fluid damper



Actively and Semi-actively Controlled Structures Under Seismic Actions: Modeling and Analysis,
Fig. 7 Friction semi-active device (a) and semi-active tuned liquid column damper (b)

Semi-active Control

Semi-active control devices offer the adaptability of active ones without requiring such high power, since external power is only used to change the device's properties, such as damping or stiffness, and not to generate a control force (Symans et al. 1994). In fact, many semi-active devices can operate on battery power, which is critical during seismic events, when the main power source to the structure may fail.

A semi-active control device cannot inject energy into the controlled system (structure and device) but has properties that can be varied in real time in order to reduce the response of a structural system (Housner et al. 1997). Changes in mechanical properties of the device are based on feedback from measured response and/or ground excitation. Therefore, in contrast to active control devices, semi-active ones do not destabilize the structural system. They offer

stability and reliability, since they function as passive devices in case of power failure (Soong and Spencer 2002). A lot of studies indicate that appropriately designed semi-active systems perform significantly better than passive ones. Moreover, they perform better than active systems for a variety of dynamic loading conditions.

Examples of such devices include variable orifice fluid dampers, controllable friction devices, variable stiffness devices, controllable liquid dampers, and controllable fluid dampers (Figs. 6 and 7). A variable orifice fluid damper uses an electromechanically variable orifice to alter the resistance to flow of a conventional hydraulic fluid (Feng and Shinozuka 1992; Constantinou et al. 1993).

A semi-active controllable fluid device is a combination of dampers with fluids that have the ability to reversibly change their viscosity. The two controllable fluids used in structural

control devices are electrorheological (ER) and magnetorheological (MR) fluids. They consist of dielectric polarizable (electrorheological, ER fluids) or magnetically polarized (magnetorheological, MR fluids) particles suspended in an oil medium. They have the ability to reversibly change from viscous fluids to semi-solids with controllable yield strength in milliseconds, with the application of an electrical or magnetic field, respectively. This property makes them ideal for use in controllable dampers. The advantage of controllable fluid devices is that they contain no moving parts other than the piston, which makes them very reliable and very easy to maintain. Moreover, they require low power to operate.

The discovery of both ER and MR fluids dates back to the late 1940s (Winslow 1947). ER fluid dampers have been developed, modeled, and tested for civil engineering applications (Erhrogott and Marsi 1992, 1993; Makris et al. 1995). Work on MR devices have been done by Spencer et al. (1997), Soong and Spencer (2002), Spencer and Nagarajaiah (2003), Carlson et al. (1995), and Dyke et al. (1996c–f).

Other semi-active devices use the force generated by surface friction to dissipate energy in a structural system.

Other types of semi-active control devices use the dynamic motion of a sloshing fluid or a column of fluid to reduce the response of a structure. These liquid dampers are the evolution of passive tuned sloshing dampers (TSD) and tuned liquid column dampers (TLCD). The TSD uses the liquid in a sloshing tank to add damping to the structural system. Similarly, in a TLCD the moving mass is a column of liquid, which is driven by the vibrations of the structure. These passive systems are not very effective for varying loading conditions. To improve their effectiveness, a semi-active device based on the passive TSD has been proposed, in which the length of the sloshing tank, thus also the properties of the device, and therefore its natural frequency can be changed. Similarly, in semi-active devices based on a TLCD, a variable orifice within the liquid column is used, or the cross section of the sloshing tank is changed.

Semi-active tuned mass dampers are similar to TMDs, but with the capability of varying their level of damping. They are mainly used for wind vibration reduction. Another type of semi-active TMD is the semi-active variable stiffness tuned mass damper (SAIVS-TMD), where the stiffness is also controllable. Their performance is similar to that of AMDs but with less power consumption.

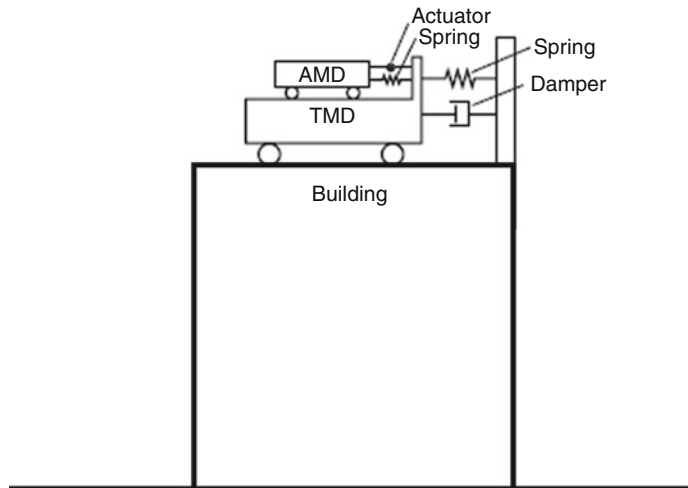
Variable stiffness control devices have the ability to modify the structure's stiffness and therefore its natural frequency to avoid resonant conditions. These systems have been studied by Kobori et al. (1993). They are installed in bracing systems, and with opening or closing a valve, they allow the connection between the brace and beam, thus changing the building stiffness and therefore its frequency, to avoid resonance with the incoming earthquake. Their energy operation is very low, and they are designed so that in the case of power failure, the connection is automatically closed and the structure's stiffness is increased.

The first full-scale application of semi-active control was the installation of variable stiffness devices on both sides of the Kajima Technical Research Institute.

Hybrid Control

Hybrid control refers either to a combination of passive and active systems or, more commonly, to a combination of passive with semi-active systems, aiming at lowering the forces required by active or semi-active systems, respectively. One such device is the hybrid mass damper (HMD), which combines tuned mass dampers with active actuators. The actuator force is only used to increase efficiency and robustness to changes in structural dynamic characteristics. Also in the category of hybrid mass dampers is the active–passive composite tuned mass damper (APTMD) developed by Ohruai et al. (1994) and named DUOX. This device is composed of an active mass damper mounted on a tuned mass damper, Fig. 8. During structural motion, the mass of the AMD is driven in the opposite direction of the TMD, therefore magnifying the motion of the passive device. When the building

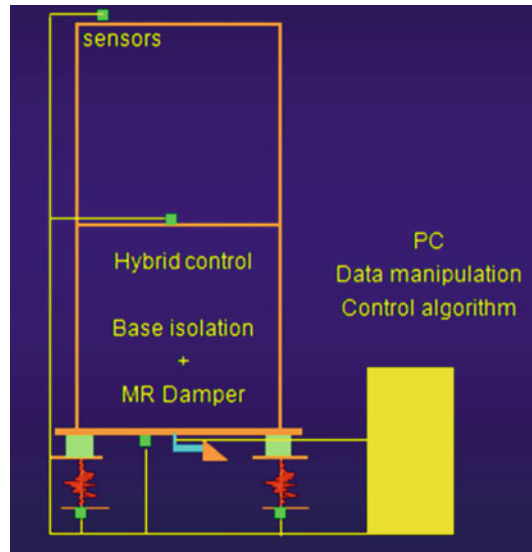
Actively and Semi-actively Controlled Structures Under Seismic Actions: Modeling and Analysis,
Fig. 8 Simplified diagram of DUOX control system



deformation stops, the AMD is used to suppress useless motion of the TMD.

Base isolation systems are passive systems, and they do not have the ability to adapt and change their properties in different external excitation (e.g., near- or far-fault excitation). With the addition of an active or semi-active control device to a base-isolated structure, a higher level of performance can be achieved without a substantial increase in the cost. This thought has led to another type of hybrid control system, referred to as hybrid seismic isolation, consisting of active or semi-active devices introduced in base-isolated structures (see Fig. 9). Although base isolation has the ability to reduce interstory drifts and structural accelerations, it increases base displacement, hence the need for an active or semi-active device. In addition, a semi-active friction-controllable fluid bearing has been employed in parallel with a seismic isolation system (Feng and Shinozuka 1992; Sriram et al. 2003).

Hybrid control strategy, HMD, was first implemented in 1993, in the Ando Nishikicho Building in Tokyo, Japan. During strong winds or moderate earthquakes, when the structure's first mode of vibration can be considered dominant, the control system will simply act as a passive device. However, in the case of a stronger earthquake, where the ground excitation is spread over a wider frequency band and



Actively and Semi-actively Controlled Structures Under Seismic Actions: Modeling and Analysis,
Fig. 9 Hybrid control system

the first mode of vibration may no longer be dominant, the actuator is activated to compensate the response due to higher modes.

A combination of passive and active control systems has been applied to USC University Hospital and includes five of the six buildings of the medical center. Both linear and high damping rubber bearings were chosen for the base isolation, to provide lateral stiffness,

which controls natural vibration period and hysteretic damping. Active control is accomplished by the placement of viscous damping devices at the base of the structure to provide velocity-dependent damping, which controls the overall building displacements.

Structures equipped with hybrid mass dampers are the Kansai International Airport in Osaka, Japan; the Mitsubishi Heavy Industry in Yokohama, Japan; and the RIHGA Royal Hotel in Hiroshima, Japan. An interesting device can be found in the Shinjuku Park Tower consisting of a V-shaped HMD developed by the Ishikawajima-Harima Heavy Industries. This device has an easily adjustable fundamental period.

An application of hybrid control carried out by Lin et al. (2007) included a series of large-scale experimental tests conducted on a mass equipped with a hybrid controlled base isolation system, consisting of a rolling pendulum system (RPS) and a 20-KN magnetorheological (MR) damper. The 12-t mass and its hybrid isolation system were subjected to various intensities of near-fault and far-fault earthquakes on a large shake table. The results showed that a combination of rolling pendulum system and an adjustable MR damper can provide robust vibration control for large civil engineering structures that need protection from a wide range of seismic events.

A benchmark smart base-isolated eight-story building structure has been presented by Narasimhan et al. (2003), similar to existing buildings in Los Angeles, California. The base isolation system includes both linear and nonlinear bearings and control devices. Design and implementation of active semi-active and hybrid systems can also be found in the work of (Chu et al. 2005; Yi et al. 2001).

Modeling

Continuous and Discrete Control

The equation of motion of a controlled structural system with n degrees of freedom u_i , subjected to an earthquake excitation a_g , is given by Eq. 1:

$$\mathbf{M}\ddot{\mathbf{U}}(t) + \mathbf{C}\dot{\mathbf{U}}(t) + \mathbf{K}\mathbf{U}(t) = -\mathbf{M}\mathbf{E}a_g(t) + \mathbf{E}_f\mathbf{F}(t) \quad (1)$$

where $\mathbf{M}$, $\mathbf{C}$, $\mathbf{K}$ denote the mass, damping, and stiffness matrixes of the structure, respectively, $\mathbf{E}$, $\mathbf{E}_f$ are the location matrix for the earthquake and the control forces on the structure, and $\mathbf{F}(t)$ is the control force matrix which is applied to the structure.

In the state space approach, Eq. 1 can be written as follows:

$$\begin{aligned} \dot{\mathbf{X}}(t) &= \mathbf{A}\mathbf{X}(t) + \mathbf{B}_g a_g(t) + \mathbf{B}_f \mathbf{F}(t) \\ \mathbf{Y}(t) &= \mathbf{C}\mathbf{X}(t) + \mathbf{D}\mathbf{F}(t) + \mathbf{v} \end{aligned} \quad (2)$$

The matrixes $\mathbf{X}$, $\mathbf{A}$, $\mathbf{B}_g$, $\mathbf{B}_f$ are given by

$$\begin{aligned} \mathbf{X} &= \begin{bmatrix} \mathbf{U} \\ \dot{\mathbf{U}} \end{bmatrix}_{2n \times 1}, \mathbf{A} = \begin{bmatrix} \mathbf{O} & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix}_{2n \times 2n}, \\ \mathbf{B}_g &= \begin{bmatrix} \mathbf{O} \\ -\mathbf{E} \end{bmatrix}_{2n \times 1}, \mathbf{B}_f = \begin{bmatrix} \mathbf{O} \\ \mathbf{M}^{-1}\mathbf{E}_f \end{bmatrix}_{2n \times 1} \end{aligned} \quad (3)$$

The matrixes $\mathbf{Y}$, $\mathbf{C}$, $\mathbf{D}$, and $\mathbf{v}$ are the output states, the output matrix, the feed forward control force matrix, and the noise matrix, respectively. In the case where the output variables are the same with the states of the system and there is no application of the control forces to the output variables, the matrixes $\mathbf{C}$, $\mathbf{D}$ are the identity and zero matrix, respectively. The noise matrix depends on the sensor that is used to measure the response of the system. The above equation can be solved by any numerical technique for differential equations, like an explicit Runge–Kutta formula, the Dormand–Prince pair, Bogacki–Shampine, and Adams–Bashforth–Moulton PECE solver.

The continuous solution of Eq. 3 is

$$\begin{aligned} \mathbf{x}(t) &= \mathbf{e}^{\mathbf{A}(t-t_0)} \mathbf{x}_0 + \int_{t_0}^t \mathbf{e}^{\mathbf{A}(t-\tau)} (\mathbf{B}_f \mathbf{F}(\tau) + \mathbf{B}_g a_g(\tau) \\ &\quad + \mathbf{B}_p p(\tau)) d\tau \end{aligned} \quad (4)$$

Equation 4 is applied assuming that the displacement or velocity and control force are continuous

functions of time. This does not apply to a real control situation, where the control force is calculated by observed values of displacement, velocity, or acceleration at discrete time intervals. For discrete description, the total time is divided into small intervals $t_0-t_1, t_1-t_2, \dots, t_n-t_f$, with time interval Δt .

An approximate solution of Eq. 3 between two points of time t_j and $t_{j+1} = t_j + \Delta t$ is obtained making the following substitutions in the above continuous solution Eq. 4 and editing the exponential integral:

$$\begin{aligned} t &= t_{j+1} \quad t_0 = t_j \\ \mathbf{X}(t_j) &= \mathbf{X}_j \quad \mathbf{F}(\tau) = \mathbf{F}(t_j) = \mathbf{F}_j \\ p(\tau) &= p(t_j) = p_j \quad a_g(\tau) = a_g(t_j) = a_{g_j} \end{aligned} \quad (5)$$

$$\mathbf{X}_{j+1} = e^{A\Delta t} \mathbf{X}_j + \mathbf{A}^{-1} (e^{A\Delta t} - \mathbf{I}) [\mathbf{B}_F \mathbf{F}_j + \mathbf{B}_g a_{g,j} + \mathbf{B}_p p_j] \quad (6)$$

The above equation provides an estimate of the response at time t_{j+1} based on values of the response at previous time t_j . The first term in the right-hand side represents the free oscillation response (transient state), while the other terms provide the response to load during time Δt (steady state). The above approach can be applied to adaptive systems (term in civil structures: material nonlinearity), wherein the stiffness and damping constants vary with time, while the mass is maintained constant. In that case matrix $\mathbf{A}$ is altered over time but remains constant during the interval Δt . In this case Eq. 6 becomes

$$\mathbf{X}_{j+1} = e^{A_j \Delta t} \mathbf{X}_j + \mathbf{A}_j^{-1} (e^{A_j \Delta t} - \mathbf{I}) [\mathbf{B}_F \mathbf{F}_j + \mathbf{B}_g a_{g,j} + \mathbf{B}_p p_j] \quad (7)$$

where

$$k(t) = k(t_j) = k_j \quad \mathbf{A}(t) = \mathbf{A}(t_j) = \mathbf{A}_j \quad c(t) = c(t_j) = c_j \quad (8)$$

The feedback control force in discrete form is

$$\begin{aligned} \mathbf{F}_j &= -\mathbf{K}_{f,j} \mathbf{X}_j, \quad t_j \leq t \leq t_{j+1} \\ \mathbf{K}_{f,j} &= [k_d(t_j) k_v(t_j)] \end{aligned} \quad (9)$$

Replacing into Eq. 7 yields

$$\begin{aligned} \mathbf{X}_{j+1} &= (e^{A_j \Delta t} - \mathbf{A}_j^{-1} (e^{A_j \Delta t} - \mathbf{I}) \mathbf{B}_F \mathbf{K}_f) \mathbf{X}_j \\ &\quad + \mathbf{A}_j^{-1} (e^{A_j \Delta t} - \mathbf{I}) [\mathbf{B}_g a_{g,j} + \mathbf{B}_p p_j] \end{aligned} \quad (10)$$

The output response is obtained by starting from time t_0 , when the response is known, and calculating subsequent time points. The critical issue here is the determination of time Δt and the distribution of stiffness and feedback parameters k_v and k_d .

Linear and Nonlinear Control

In Eq. 1, the change of the material properties change during loading leads to changes in the stiffness matrix. Then, the differential equations become nonlinear:

$$\mathbf{M}\ddot{\mathbf{U}}(t) + \mathbf{C}\dot{\mathbf{U}}(t) + \mathbf{F}_s(\mathbf{U}(t)) = -\mathbf{M}\mathbf{E}a_g(t) + \mathbf{E}_f \text{sat} \mathbf{F}(t - t_d) \quad (11)$$

In this case the nonlinearity originates from the structure and is described as material nonlinearity.

When the equation of motion is formulated in the deformed configuration to account for the structure's flexibility and associated large displacements, then equation of motion (1) also becomes nonlinear; the nonlinearity also originates from the structure, but now it is described as geometric nonlinearity.

When the control force, $\mathbf{F}$, is a linear function of the response of the structure, then the above equations of motion (1) are linear differential equations, and the control is said to be linear. When the control force is not a linear function, then Eq. 1 are nonlinear differential equations, and the nonlinearity originates from the control force.

Thus, the source of nonlinearity could be either from the structure or from the control force. In Table 1 below, the possible cases are presented.

Practical Considerations

Over the past few decades, various control algorithms and control devices have been developed, modified, and investigated by various groups of researchers. Several well-established algorithms

Actively and Semi-actively Controlled Structures Under Seismic Actions: Modeling and Analysis,
Table 1 Linear and nonlinear cases of analysis

Control force	Structure		
	Linear	Nonlinear	
		Material	Geometric
Linear	Linear structure, linear control	Nonlinear structure, linear control	Nonlinear structure, linear control
Nonlinear	Linear structure, nonlinear control	Nonlinear structure, nonlinear control	Nonlinear structure, nonlinear control

in control engineering have been introduced to control structures. While many of these structural control strategies have been successfully applied, technological problems and challenges relating to time delay, saturation capacity effects, cost, reliance on external power, and mechanical complexity and reliability during the life of the structure have delayed their widespread use, and relatively few actual structures are equipped with control systems.

Another practical issue that influences the effectiveness and the reliability of the proposed control algorithms is the effect of the position of control forces. The selection of locations of control forces influences the location matrix, $\mathbf{B}_f$ of the control force in the differential equation of motion of controlled structure. Thus, this effect can be investigated numerically by parametric variation of the location matrix.

Other practical effects include control–structure interaction, actuator dynamics, and digital control implementation. The reliability of applied semi-active structural control systems and practical applications and verification for active and semi-active vibration control of buildings in Japan have been studied by Ikeda (2009).

Time Delay–Saturation Capacity

Two practical issues that influence the effectiveness and the reliability of the proposed control algorithms are time delay and saturation of the control force. Those parameters come into consideration by solving the differential equation of

motion as a delay differential equation with saturation effects. The expected negative influence of those parameters should be considered in the design process. Thus, there is a need to take them into account in the numerical simulations before the installation of the control system on the real building.

The equation of motion (1) of a controlled structural system considering time delay and saturation becomes

$$\mathbf{M}\ddot{\mathbf{U}}(t) + \mathbf{C}\dot{\mathbf{U}}(t) + \mathbf{K}\mathbf{U}(t) = -\mathbf{M}\mathbf{E}\mathbf{a}_g(t) + \mathbf{E}_f \text{sat}\mathbf{F}(t - t_d) \quad (12)$$

$\text{sat}\mathbf{F}$ is the saturated control force matrix which is applied to the structure with time delay t_d and is given as

$$\text{sat}\mathbf{F}(t - t_d) = \begin{cases} \mathbf{F}(t - t_d) & \mathbf{F}(t - t_d) < \mathbf{F}_{\text{allowable}} \\ \mathbf{F}_{\text{allowable}} & \mathbf{F}(t - t_d) > \mathbf{F}_{\text{allowable}} \end{cases} \quad (13)$$

$\mathbf{F}_{\text{allowable}}$ is the maximum capacity of the control device. In the state space approach, Eq. 12 can be written as follows:

$$\begin{aligned} \dot{\mathbf{X}}(t) &= \mathbf{A}\mathbf{X}(t) + \mathbf{B}_g \mathbf{a}_g(t) + \mathbf{B}_f \text{sat}\mathbf{F}(t_d - t) \\ \mathbf{Y}(t) &= \mathbf{C}\mathbf{X}(t) + \mathbf{D}\mathbf{F}(t_d - t) + \mathbf{v} \end{aligned} \quad (14)$$

This equation can be solved by the technique of delay differential equation, or one can use the following transformation

$$\mathbf{Z}(t) = \mathbf{X}(t) + \int e^{-\mathbf{A}(\eta+t_d)} \mathbf{B}_f \mathbf{F}(t + \eta) d\eta \quad (15)$$

Then:

$$\begin{aligned} \dot{\mathbf{Z}}(t) &= \mathbf{A}\mathbf{Z}(t) + \mathbf{B}_g \mathbf{a}_g(t) + \mathbf{B}(\mathbf{A})\mathbf{F}(t) \\ \mathbf{B}(\mathbf{A}) &= e^{-\mathbf{A}t_d} \mathbf{B}_f \end{aligned} \quad (16)$$

Since the entire control process involves measuring response data, computing control forces through an appropriate algorithm, transmitting data and signals to actuators, and activating the actuators to a specified level of force, time delays

arise and cannot be avoided. The problem of time delay in the active control of structural systems has been investigated from many scientists and engineers. The stability of the structure could be lost due to time delay, and two ways of time-delay compensation can be followed. In the first the gain matrix is redesigned considering the presence of time delay, while in the second low-pass filters are used to filter the velocity measurements from the frequency components of the high-order modes. In the first case, the structure could remain unstable when using control moments as control actions, while in the second a number of vibration modes can be controlled and compensated for time delay, but the higher-order modes remain uncontrolled. Time delay can be compensated with Pade approximations. The allowable time delay is related with natural period and feedback gain. The maximum allowable time delay is decreased with decrease in natural period of the structure, as well as with increase in active damping. Under earthquake excitations, simulation results for the response of multi-degree of freedom structures indicated that the degradation of the control performance due to fixed time delay is significant when time delay is close to a critical value. The time-delay problem is more serious for structures with closely spaced vibration modes.

In optimal control of linear systems, time delay is considered at the very beginning of the control design, and no approximations and estimations are made in the control system. Thus, the system performance and stability can be guaranteed. Instability in the response might occur only if a system with time delay is controlled by an optimal controller that was designed with no consideration of time delay.

For pole assignment algorithm, through varied location of the controlled poles, the control system shows variable performance. However, the locations of the controlled pole pairs should be carefully specified and checked according to the characteristics of the system. Analytical expressions of limiting values of time delay for single degree of freedom systems were derived by Connor (2003); however, such expressions were very difficult to obtain for multi-degree of

freedom systems. Casciati et al. (2006) have taken into consideration the time-delay effect solving numerically delayed differential equations. All of these studies demonstrate how important the issue of time delay is in structural control and how it may result in a degradation of the control performance and may even drive the controlled structure to become unstable. Most studies show that time delays influence negatively the control system; therefore, they should be kept small compared to the fundamental period of vibration of the system and should, if possible, be eliminated.

The second important practical problem is the saturation of the control force. Actuator saturation occurs when the force which is given by the control algorithm is larger than its designed peak capacity. Failure to account for this nonlinear effect can decrease the efficiency of the control system and possibly drive the structure to become unstable. Most control algorithms are linear, assuming that there is no limit in the magnitude of the control force. However, maximum capacity of the control devices is limited. Therefore, designing controllers to account for the bounded nature of the devices is desirable.

The two issues of time delay and saturation of the control device are, in most cases, considered and studied separately. However, in the application of real control systems, these two issues act simultaneously. Pnevmatikos and Gantes (2011) investigated a combined effect of the nonlinear phenomena of bounded capacity of the actuators and time delay of the system, acting simultaneously during the control process, on the systems response. They proposed limits for pair of time delay and saturation capacity that can be used in the design process of controlled structures.

Spillover Effect

With the discretization of a continuous system to a finite degree of freedom system, some information is lost since a real physical system has infinite natural mode shapes and frequencies, while the discrete systems contain some of them. Those frequencies that are not included in the discrete model are called remaining frequencies or

residual modes. If flexible structure is modeled, there is a danger that the control based on a reduced model is destabilized due to the high remaining modes which are not included in the model. Balas (1978) defined and studied the above phenomenon, which is called spillover effect. Meirovich (1990) investigated observation spillover (measurements at some points, without having the total picture of the response) and stated that this can be really dangerous for the controlled structures.

Controllability: Observability

Controllability of system deals with the number of control positions and degrees of freedom of the structure. When the number of control positions is equal to the number of degrees of freedom of the structure, then full control of the system is achieved. In that case, the building performs a rigid body motion, following the imposed ground motion, without relative displacements between the floors. It was also shown that with reduced number of control forces, positioned at appropriate locations, which is a more realistic choice for real buildings, the response can be reduced to a satisfactory level. Observability is associated with measurement positions in the structure to degrees of freedom of the structure. When the measurement positions are equal to the number of degrees of freedom of the structure, then full-state feedback is achieved. Otherwise, feedback with an observer is performed.

Collocated and Not Collocated Control/ Centralized and Decentralized Control

A definition of collocation and centralization is given by Casciati et al. (2006). A control system is collocated when the force generated by an actuator at a point of the structure is measured by a force sensor at the same location, in other words when the actuator and sensor are connected exactly at the same location. Otherwise, control is non-collocated.

A control system is centralized if it is managed by a unique computer that receives the inputs from all sensors and gives the command output to all actuators. The system is decentralized

(noncentralized) if the control system is managed by several computers that take the input from some specific sensors and give the command output to some actuators. It is thus possible that a noncentralized system could be collocated or that a centralized system could be non-collocated.

Analysis

Control Strategies and Algorithms

Several well-established algorithms in control engineering have been introduced to control structures, such as optimal control, LQR or LQG, pole assignment, sliding mode control, H_2 and H_∞ , fuzzy control, and many others. The most suitable algorithms for structural application and the practical considerations that should be taken into account are described by Soong (1990) and Casciati et al. (2006).

Optimal Control, LQR, or LQG

Research in structural control has focused on a variety of control algorithms based on different control design criteria. Some algorithms originate from direct applications of optimal control theory. Some others, however, are specifically proposed for civil engineering structural control applications.

The control force, $\mathbf{F}$, can be applied directly or indirectly to the structure. The way in which the control force is calculated is determined from the control algorithm that is used. If the control force is calculated by linear state feedback:

$$\mathbf{F} = -\mathbf{G}_1 \mathbf{U} - \mathbf{G}_2 \dot{\mathbf{U}} = -[\mathbf{G}_1 \ \mathbf{G}_2] \begin{bmatrix} \mathbf{U} \\ \dot{\mathbf{U}} \end{bmatrix} = -\mathbf{G} \mathbf{X} \quad (17)$$

$\mathbf{G}$ is the gain matrix, which is calculated according to the desired poles of the controlled system. Replacing the force $\mathbf{F}$ into Eq. 1 or 2, the controlled system can be described by

$$\mathbf{M} \ddot{\mathbf{U}}(t) + (\mathbf{E}_f \mathbf{G}_2 + \mathbf{C}) \dot{\mathbf{U}}(t) + (\mathbf{E}_f \mathbf{G}_1 + \mathbf{K}) \mathbf{U}(t) = -\mathbf{M} \mathbf{E} a_g(t) \quad (18)$$

$$\dot{\mathbf{X}} = (\mathbf{A} - \mathbf{B}_f \mathbf{G})\mathbf{X} + \mathbf{B}_g a_g \quad (19)$$

$$\mathbf{X}(t_0) = \mathbf{X}(0) = \mathbf{X}_0$$

From the above equation, it is seen that control of structures can be achieved by changing the stiffness or damping and consequently the dynamic characteristics of the building in a direct or indirect way, depending on the device that is used. The question is how to estimate the control force or the matrix $\mathbf{G}$ in such a way that the desired dynamic characteristics for the controlled building are achieved.

The feedback matrix can be calculated based on optimal control theory like linear quadratic regulator, LQG. Optimal control methods are based on the concept of minimizing a cost criterion. The criterion of cost, represented by J , has a common format:

$$J = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{t_0}^{t_f} \mathbf{e}^T(t) \mathbf{e}(t) dt \quad (20)$$

where

$$\mathbf{e}(t) = \mathbf{X}(t) - \mathbf{X}^*(t) \quad (21)$$

$\mathbf{e}(t)$ is the error between the desired behavior, $\mathbf{X}^*$, and the actual behavior, $\mathbf{X}$, of the system.

The scope of the problem of optimal control is to determine a control force, $\mathbf{F}(t)$, such that it determines the behavior of the control system to minimize some cost criterion while satisfying some physical constraints of the system. The cost criterion is usually formulated so as to express a quantity to have physical significance, e.g., displacement and energy. A specialized form of cost criterion of Eq. 20 is as follows:

$$J = \theta[\mathbf{X}(t), t] \Big|_{t=t_0}^{t=t_f} + \int_{t_0}^{t_f} \phi[\mathbf{X}(t), \mathbf{F}(t)] dt \quad (22)$$

The first term refers to the cost to the ends of the interval or to the boundary condition, while the second term refers to the cost to the entire

space. Depending on the requirements of the problem, functions $\theta(\mathbf{X}(t), t)$, $\phi(\mathbf{X}(t), \mathbf{F}(t))$ are taking specific forms. One of the most common forms of criterion J , which minimizes the energy of the system, is

$$J = \mathbf{X}(t_f)^T \mathbf{S} \mathbf{X}(t_f) + \int_{t_0}^{t_f} [\mathbf{X}(t)^T \mathbf{Q}(t) \mathbf{X}(t) + \mathbf{F}^T(t) \mathbf{R}(t) \mathbf{F}(t)] dt \quad (23)$$

The weighted matrixes $\mathbf{S}$, $\mathbf{Q}(t)$, and $\mathbf{R}(t)$ are selected according to the importance one wants to give to the error vector, $\mathbf{e}(t)$, or to the excitation vector $\mathbf{F}(t)$. The selection of suitable $\mathbf{S}$, $\mathbf{Q}(t)$, and $\mathbf{R}(t)$ for a particular problem is usually a difficult issue that requires experience and engineering insight.

The minimization of the cost criterion can be accomplished using the maximum principle introduced by Pontryagin and the principle of optimality introduced by Bellman. The procedure results in a system of differential equations as follows:

$$\begin{aligned} \dot{\mathbf{P}}(t) + \mathbf{P}(t)\mathbf{A}(t) + \mathbf{A}^T(t)\mathbf{P}(t) - \mathbf{P}(t)\mathbf{B}(t)\mathbf{R}^{-1}(t)\mathbf{B}^T(t) \\ \mathbf{P}(t) = -\mathbf{C}^T(t)\mathbf{Q}(t)\mathbf{C}(t) \\ \mathbf{P}(t_f) = -\mathbf{C}^T(t_f)\mathbf{Q}(t_f)\mathbf{C}(t_f) \end{aligned} \quad (24)$$

$$\begin{aligned} \dot{\boldsymbol{\mu}}(t) + [\mathbf{A}(t) - \mathbf{B}(t)\mathbf{R}^{-1}(t)\mathbf{B}^T(t)\mathbf{P}(t)]^T \\ \boldsymbol{\mu}(t) = -\mathbf{C}^T(t)\mathbf{Q}(t)\mathbf{X}^*(t) \\ \boldsymbol{\mu}(t_f) = -\mathbf{C}^T(t_f)\mathbf{Q}(t_f)\mathbf{X}^*(t_f) \end{aligned} \quad (25)$$

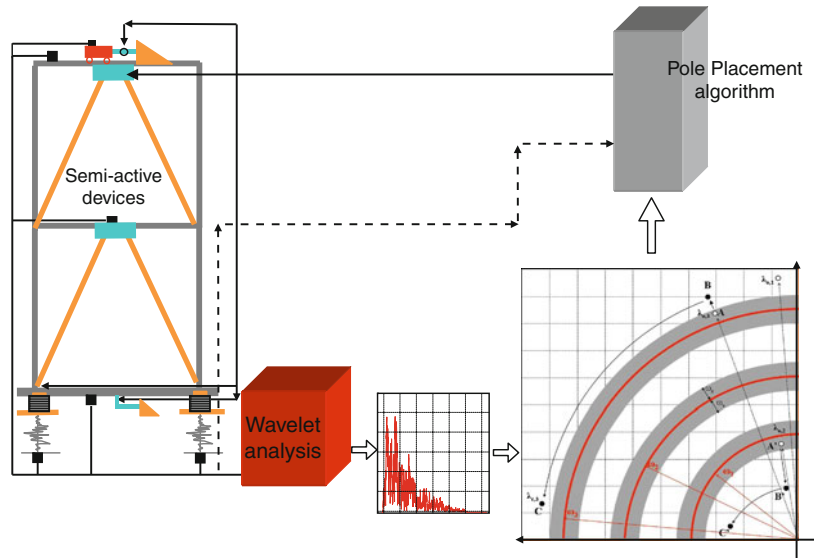
Solving the above differential equations, $\mathbf{P}(t)$ and $\boldsymbol{\mu}(t)$ are computed, and then from Eq. 26, the feedback matrix $\mathbf{K}(t)$ and the control force $\mathbf{F}(t)$ are calculated. Equation 24 is called Riccati equation:

$$\begin{aligned} \mathbf{f}(t) = \mathbf{K}(t)\mathbf{X}(t) + \boldsymbol{\rho}(t) \\ \mathbf{K}(t) = -\ddot{\mathbf{R}}^{-1}(t)\mathbf{B}^T(t)\mathbf{P}(t), \boldsymbol{\rho}(t) = \ddot{\mathbf{R}}^{-1}(t)\mathbf{B}^T(t)\boldsymbol{\mu}(t) \end{aligned} \quad (26)$$

The above control law is under the assumption that all states are available and measurable.

Actively and Semi-actively Controlled Structures Under Seismic Actions: Modeling and Analysis,

Fig. 10 The general flowchart of the pole placement control strategy



A

In civil structures it is unrealistic to expect that the state vector can be fully measured. The case that a few degrees of freedom are measured and used for the calculation of the control force is called output control, in contrast to the full-state control where all degrees of freedom of the system are measured. Suitable control design techniques, like observers with linear quadratic Gaussian (LQG) control, have been developed for output feedback and random disturbances. Many researchers have studied and applied optimal control in civil structures (Abdel-Rohman and Leipholz 1983; Chang and Soong 1980; Yang 1975). An evolutionary control of damaged systems using a rehabilitative, modified LQR algorithm has been proposed by Attard and Dansby (2008).

Pole Placement Algorithm

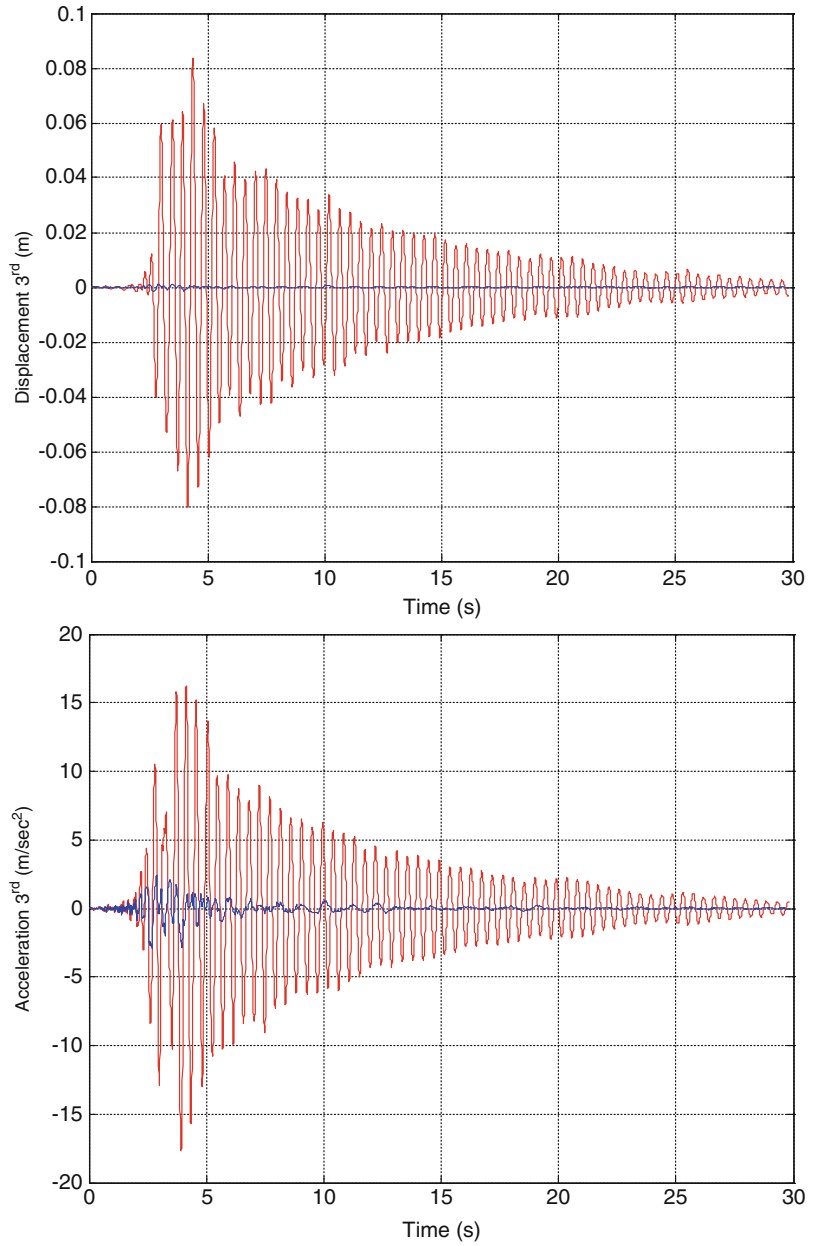
It should be noted that, if the external excitation is ignored or set to zero in the derivation of the Riccati equation, the provided control law is not optimal. In order to include the excitation in the Riccati equation, a priori knowledge of the loading history is required. This is generally not possible for excitations such as earthquakes, wind, or waves, which are common in structural engineering applications.

A control algorithm which addresses this difficulty is pole assignment (pole placement) algorithm. Pole placement algorithms have been studied extensively in the general control literature, while its applications in structural control have been investigated by Martin and Soong (1976), Leonard (1990), Soong (1990), Utku (1998), and Preumont (2002).

A procedure of on-line selection of poles in such a way that first resonance is avoided and secondly sufficient equivalent damping is added, based on the specific characteristics of the incoming dynamic earthquake excitation, have been proposed by Pnevmatikos and Gantes (2010a). This procedure drives the poles to their optimum location and does not need the poles to be predefined and constant during the application of dynamic loading. Numerical simulations show that sufficient reduction of the response, in terms of both displacement and acceleration, can be achieved for all examined earthquakes with reasonable amount of required equivalent control force. The procedure is shown schematically in Fig. 10. The effectiveness to the response of the structure of the control strategy is shown in Fig. 11. It is shown that both the displacement and the acceleration are reduced one order of magnitude.

Actively and Semi-actively Controlled Structures Under Seismic Actions: Modeling and Analysis,

Fig. 11 Displacement and acceleration of the controlled (*blue line*) and uncontrolled (*red line*) system for the third floor of the three-story building subjected to earthquake



H_2/H_∞ Control Algorithm

The objective of the H_2 or H_∞ control algorithms is to design a controller $\mathbf{K}$ that minimizes the H_2 or H_∞ norm of the closed-loop transfer function matrix, $\mathbf{H}$, from the disturbance to the output vector. By definition, the H_2 norm of a stable transfer function matrix is

$$\|\mathbf{H}\|_2 = \sqrt{\text{trace} \left\{ \frac{1}{2\pi} \int_{-\infty}^{\infty} \mathbf{H}(j\omega) \mathbf{H}^*(j\omega) d\omega \right\}} \quad (27)$$

More details regarding the use of control H_2 and LQR methods for civil engineering applications can be found in Zacharenakis et al. (2001).

Sliding Mode Control

Sliding mode control or variable structure strategies were developed specifically for robust control of uncertain nonlinear systems. The fundamental idea of SMC is to design a controller to drive the state trajectory on the sliding surface (or switching surface), whereas the motion on the sliding surface is stable, remains there all the subsequent time, and moves toward the equilibrium position. The first step in SMC is to design the sliding surface on which the response is stable, while the second step is the determination of the control demand which will drive the response trajectory into the sliding surface and force it to stay there all the subsequent time. In most studies the sliding surface, $\mathbf{s}$, is defined as a linear combination of the state vector:

$$\mathbf{s} = \dot{\mathbf{U}} + \lambda \mathbf{U} = [\lambda \quad \mathbf{I}] \begin{bmatrix} \mathbf{U} \\ \dot{\mathbf{U}} \end{bmatrix} = \mathbf{P}\mathbf{X} \quad (28)$$

In the work of Slotin and Li (1991), a more general approach is proposed. The sliding surface can be obtained using optimal control theory. In the work of Pnevmatikos and Gantes (2009), the sliding surface, $\mathbf{s}$, is defined by pole assignment method, where the selection of the poles of the controlled system is based on the frequency content of the incoming earthquake signal.

Following the design of the sliding surface, the control force which will drive the response trajectory into this surface and force it to stay there is calculated using Lyapunov stability theory. To achieve this goal, first, a Lyapunov function is chosen, and then, under the condition that the derivative of the function $\mathbf{V}$ is negative, the control force is obtained. The control forces $\mathbf{F}$ are given by

$$\begin{aligned} \mathbf{F} &= \mathbf{G} - \delta \times \lambda^T, \mathbf{G} = -(\mathbf{P}\mathbf{B}_f)^{-1} \mathbf{P}(\mathbf{A}\mathbf{X} + \mathbf{B}_g \mathbf{a}_g), \\ \lambda &= \mathbf{s}^T \mathbf{P}\mathbf{B}_f \end{aligned} \quad (29)$$

Matrix $\mathbf{G}$ includes the restoring, damping, inertial, and seismic forces. The magnitude of $\mathbf{G}$

is very large for controlling conventional civil engineering structures. Thus, the control force should be restricted to a certain level, and a saturated controller should be considered in the design of SMC. In this case, full compensation of the response cannot be achieved. If the maximum control force is bounded by $\pm \mathbf{f}_{\max}$, the control force is estimated as follows:

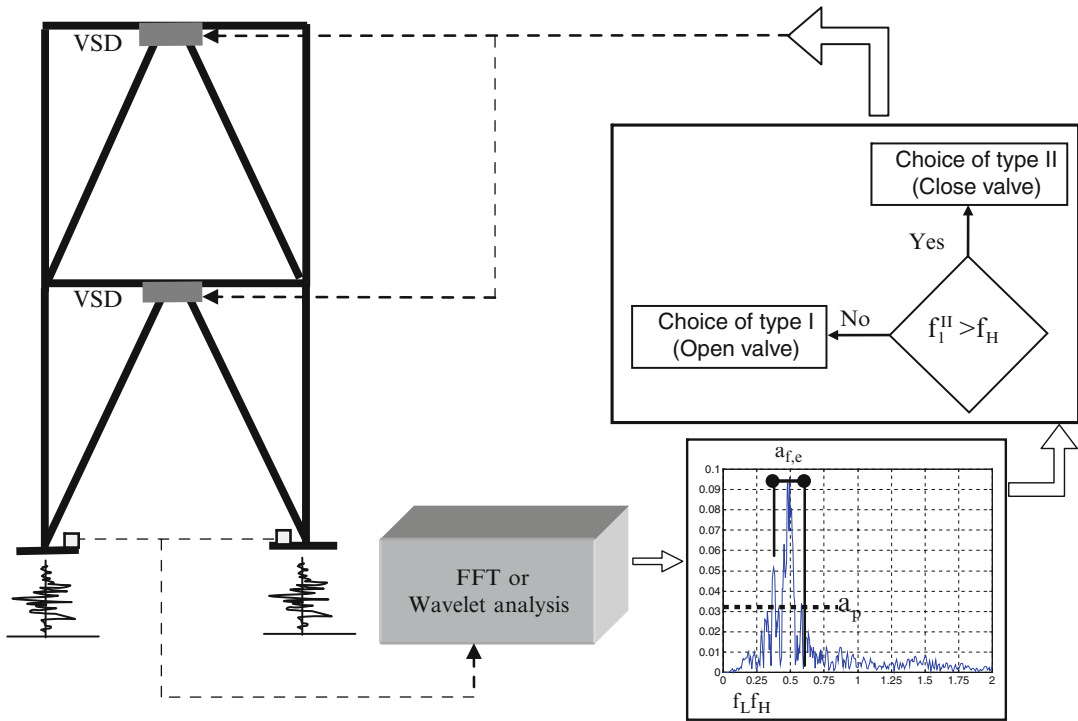
$$\mathbf{F} = \begin{cases} \mathbf{G} - \delta \times \lambda^T, & \text{if } |\mathbf{G} - \delta \times \lambda^T| \leq \mathbf{f}_{\max} \\ \mathbf{f}_{\max} \text{sign}(\mathbf{G} - \delta \times \lambda^T), & \text{otherwise} \end{cases} \quad (30)$$

The force $\mathbf{f}_{\max}$ is specified by the device capacity.

Active Variable Stiffness Control Algorithm

Variable stiffness control has been developed by Kobori and verified experimentally at Kajima Research Institute in Japan, in 1993. The control algorithm is based on the nonresonant state under seismic excitation by altering the stiffness and thus natural frequencies, of a building, based on the nature of the earthquake. Nagarajaiah et al. (1998) developed a semi-active instantaneously variable stiffness (SAIVS) system which varies the structural stiffness continuously and smoothly so as to maintain a nonresonant state. Yang et al. (2000) have developed a resetting semi-active stiffness device (RSASD) and suitable algorithm to regulate RSASD at appropriate time instants.

In Pnevmatikos et al. (2004, 2010a), a control algorithm based on the frequency content of the incoming earthquake is described, suitable for application in active variable stiffness systems installed in buildings designed against seismic actions. The control strategy consists of two stages: (i) the design stage, where the bracing system is designed based on the frequency content of a range of earthquakes, and (ii) the operation phase, where the control algorithm uses the dynamic characteristics of the measured on-line excitation signal, in order to take an appropriate decision regarding the function of the variable stiffness device. Based on the frequency content



Actively and Semi-actively Controlled Structures Under Seismic Actions: Modeling and Analysis, Fig. 12 The integrated control strategy for buildings

equipped with active variable stiffness systems and subjected to earthquake actions

of the incoming signal, the variable stiffness device, by connecting or disconnecting the bracing system, alters the stiffness of the structure and consequently changes its dynamic characteristics in real time, avoiding resonance between the controlled structure and the applied signal. Thus, the structural safety and serviceability against extreme dynamic excitation are enhanced. The general scheme of the proposed algorithm is shown in Fig. 12.

Fuzzy Logic Control

A very useful application of fuzzy logic in civil engineering is fuzzy control. Fuzzy logic, introduced by Zadeh (1965), has to do with the way that the human brain deals with concepts such as uncertainty, vagueness, and imprecision. Boolean logic determines whether an argument definitely belongs or not to some set, while fuzzy logic considers the idea of partial truths, and determines the degree of membership of an argument to a fuzzy set.

Fuzzy control defines fuzzy rules to determine actions to be taken, based on the measured structure's responses. The advantages of fuzzy controller are simple algorithms, suitable for real-time control, no need for information on structural and vibration characteristics, and a robust system in terms of performance and implementation.

Fuzzy control uses experience instead of differential equations to determine desirable control actions. Fuzzy control consists of implementation of rules, in the format of IF...THEN statements, which relate the input variables to the control action. The process begins by first defining membership functions to classify the inputs and outputs using linguistic terms. In structural control applications, the inputs are usually a combination of structural responses: displacement, velocity, and acceleration. The input values are then converted to fuzzy values, using the membership functions. This step is referred to as fuzzification. The next step is

named decision making and consists of using predetermined rules to correlate the fuzzy inputs values to fuzzy outputs. Finally, in the last step, the fuzzy output values are defuzzified, that is, they are converted to values that can be used as control actions.

The rules and the membership functions are important to the effectiveness of the fuzzy controller. Several methods are available for the creation of such rules, including logical reasoning, experiments and simulations, and learning from examples. Fuzzy control rules can be created based on data obtained by an LQG controller. Genetic algorithms can be tuned to the membership functions in order to improve the results. Genetic algorithms can also be used to simultaneously determine the optimal rules and membership functions.

Fuzzy theory can be applied to determine the desired control force to be applied by the actuator. Neural network performance function selection can be used in structural control (Casciati et al. 1993).

Fuzzy logic has been used to vary the mechanical properties of the control device. It has been applied for the reduction of longitudinal bridge displacement due to earthquake or traffic loading with variable dampers or control of structural vibrations with hybrid systems composed of base isolation and semi-active dampers. Fuzzy control is also used to regulate a structure with an MR damper system (Choi Kang-Min et al. 2004, Zhou et al. 2003).

Clip Optimal Control

The clip optimal control has two parts: the first part consists of designing a control law assuming that an ideal device is installed to the building. In the second part, a controller is designed allowing the semi-active damper to develop the force that the control law calculated. In the first part, any control law can be chosen.

Clipped-optimal control strategy has been proposed by Dyke et al. (1996) to control a single MR damper. The control algorithm was extended to control multiple MR devices, and the performance of this algorithm has been experimentally verified.

In the clipped-optimal control algorithm, the control forces f_c is given by the following expression:

$$f_c = L^{-1} \left\{ -K_c(s) L \left\{ \begin{matrix} y_m \\ f_m \end{matrix} \right\} \right\} \quad (31)$$

where $L\{\}$ is the Laplace transform, $K_c(s)$ is a linear optimal controller obtained from H_2 or LQG strategies because of the stochastic nature of earthquake ground motions (any other control law can also be chosen), y_m is the measured structural response vector, and f_m is the measured control force vector.

Because the force generated in the semi-active device is dependent on the local responses of the structural system, the desired optimal control force, f_c , cannot always be produced by the control device. Only the control voltage, v_i , can be directly controlled to increase or decrease the force produced by the device. Thus, a force feedback loop is incorporated to induce the semi-active device to generate approximately the desired optimal control force f_c . To clip the active control law to the semi-active one and to generate approximately the desired optimal control force, the command signal is selected as follows.

When the semi-active device is providing the desired optimal force $f_i = f_c$, the voltage applied to the damper should remain at the present level. If the magnitude of the force produced by the device f_i is smaller than the magnitude of the required target force f_c and the two forces have the same sign, the voltage applied to the current driver is increased to the maximum level so as to increase the force produced by the device to match the desired control force. Otherwise, the commanded voltage is set to zero.

The above procedure that regulates the applied voltage has the following mathematical expression:

$$v_i = V_{\max} H((f_{ci} - f_i)f_i) \quad (32)$$

where $V_{\max}$ is the maximum voltage to the current driver and $H(\)$ is the Heaviside step function

which is a discontinuous function whose value is zero for negative argument and one for positive argument.

In clipped-optimal control algorithm, the command voltage takes two values, either zero or the maximum value. In some situations when the dominant frequencies of the system under control are low, large changes in the forces applied to the structure may result in high local acceleration values. This behavior is dependent on the time lag in the generation of the control voltage. In Yoshida (2003), a modification to the original clipped-optimal control algorithm is proposed to reduce this effect. In the modified version of the control algorithm, the control voltage can be any value between 0 and $V_{\max}$. The control voltage, v_{ci} , is determined using a linear relationship between the applied voltage and the maximum force of MR damper. The modified clipped-optimal control algorithm is given as

$$v_i = V_{ci}H((f_{ci} - f_i)f_i) \quad (33)$$

where

$$V_{ci} = \begin{cases} \mu f_{ci} & \text{for } f_{ci} \leq f_{\max} \\ V_{\max} & \text{for } f_{ci} > f_{\max} \end{cases} \quad (34)$$

Skyhook/Ground Hook Control

On skyhook control, the damper is controlled by two damping values. The choice between high or low state is made using the following law: based on the product of the relative velocity u_{rel} across the damper and the absolute velocity of the system body mass attached to that damper, a damping value is chosen as follows:

$$\begin{aligned} \dot{z}_b \times v_{rel} \geq 0 & \quad c = \text{high state} \\ \dot{z}_b \times v_{rel} \leq 0 & \quad c = \text{low state} \end{aligned} \quad (35)$$

The skyhook control is an on-off strategy. When the relative velocity of the damper is positive, the force of the damper acts to pull on the system body mass, while when the relative velocity of the damper is negative, the force of the damper pushes the body mass.

Since the damping values now are not limited to this state alone and can take any value within this state, a continuous version of skyhook control can be achieved. The choice of damping is now given as follows:

$$\begin{aligned} \dot{z}_b \times v_{rel} \geq 0 & \quad c = a \\ \dot{z}_b \times v_{rel} < 0 & \quad c = \text{low state} \end{aligned} \quad (36)$$

where

$$a = \max\{\text{low state}, \min(g \times \dot{z}_b, \text{high state})\} \quad (37)$$

In the case of ground hook control, the damper is adjusted to its high or low state depending on the product of the relative velocity across the damper and the absolute velocity of lower mass attached to that damper. Contrary to skyhook control, this control law is driving the lower vibration mass to a reduced vibration.

The mathematical expression of the ground hook control is as follows:

$$\begin{aligned} \dot{z}_t \times v_{rel} \leq 0 & \quad c = \text{high state} \\ \dot{z}_t \times v_{rel} > 0 & \quad c = \text{low state} \end{aligned} \quad (38)$$

The continuous ground hook control strategy is similar to the sky hook control.

Summary

Representative control devices and their applications have been presented. Then, mathematical formulations and simulation of controlled structures have been illustrated. Control algorithms suitable for civil structure inspired by the classical control theory were presented. Optimal control, LQR, pole assignment, sliding mode control, H_2 , fuzzy control, clipped-optimal control, and skyhook/ground hook control algorithms were briefly described. It was shown that with the help of modern technology, it is possible to control civil structures in order to protect them from earthquake excitations.

Cross-References

- Buildings and Bridges Equipped with Passive Dampers Under Seismic Actions: Modeling and Analysis
- Early Earthquake Warning (EEW) System: Overview
- Model Class Selection for Prediction Error Estimation
- Parametric Nonstationary Random Vibration Modeling with SHM Applications
- Passive Control Techniques for Retrofitting of Existing Structures
- Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding
- Seismic Resilience
- Tuned Mass Dampers for Passive Control of Structures Under Earthquake Excitations

References

- Abdel-Rohman M, Leipholz HH (1983) Active control of tall buildings. *J Struct Div ASCE* 109:628–645
- Attard TL, Dansby RE (2008) Evolutionary control of damaged systems using a rehabilitative algorithm. In: *Proceedings of tenth pan American congress of applied mechanics PACAM X*. Grand Oasis Resort Cancun, Mexico, pp 166–169
- Balas MJ (1978) Feedback control of flexible structures. *IEEE Trans Autom Control* AC-23 673–679
- Błachowski BD (2007) Model based predictive control for guyed mast vibration. *J Theoret Appl Mech* 45(2):405–423
- Carlson JD, Catanzarite DM, St Clair KA (1995) Commercial magneto-rheological fluid devices. In: *Proceedings of the 5th international conference on ER fluids, MR fluids and associated technology*. University of Sheffield, Sheffield
- Casciati F, Faravelli L, Venini P (1993) A neural-network performance-function selection in active structural control. In: *Proceedings of the international workshop on structural control*. University of Southern California, Los Angeles
- Casciati F, Magonette G, Marazzi F (2006) *Technology of semiactive devices and applications in vibration mitigation*. Wiley, Chichester. ISBN 978-0-470-02289-4
- Chang J, Soong TT (1980) Structural control using active tuned mass dampers. *J Eng Mech Div ASCE* 106(6):1091–1098
- Choi KM, Cho SW, Jung HJ, Lee IW (2004). Semiactive fuzzy control for seismic response reduction using magnetorheological dampers. *Earthq Eng Struct Dyn* 33:723–736
- Choi Kang-Min, Cho Sang-Won, Jung Hyung-Jo, Lee In-Won (2004) Semi-active fuzzy control for seismic response reduction using magnetorheological dampers. *Earthquake Eng Struct Dyn* 33:723–736
- Chu SY, Soong TT, Reinhorn AM (2005) *Active, hybrid, and semi-active structural control a design and implementation handbook*. Wiley, United Kingdom
- Connor JJ (2003) *Introduction to structural motion control*. In: *Mit-prentice hall series on civil, environmental, and system*. Prentice Hall, USA
- Constantinou MC, Symans MD, Tsopelas P, Taylor DP (1993) Fluid viscous dampers in applications for seismic energy dissipation and seismic isolation. In: *NIST workshop 1995. Proceedings of the ATC-17-1 seminar on seismic isolation, passive energy dissipation, and active control*, San Francisco, 2 Mar 1993, pp 581–592
- Constantinou MC, Symans MD, Tsopelas P, Taylor DP (1993) Fluid viscous dampers in application of seismic energy dissipation and seismic isolation. In: *Proceedings of ATC-17-1 seminar of seismic isolation, passive energy dissipation, and active control*, San Francisco, March 2, pp. 581–592
- Dyke SJ, Spencer BF, Sain MK, Carlson JD (1996) Modeling and control of magnetorheological dampers for seismic response reduction. *Smart Mater Struct* 5:565–75
- Ehrgott RC, Masri SF (1992) Modelling the oscillatory dynamic behavior of electrorheological materials in shear. *Smart Mater Struct* 1(4):275–285
- Ehrgott RC, Marsi SF (1993) Structural control applications of an electrorheological device. In: *Proceedings of the international workshop on structural control*, Honolulu, pp 115–129
- Feng MQ, Shinozuka M (1992) Experimental and analytical study of a hybrid isolation system using friction controllable sliding bearings. Report no 92-0009, National center for earthquake engineering Research, Buffalo
- Housner G, Bergman L, Caughey T, Chassiakos A, Claus R, Masri S, Skelton R, Soong T, Spencer B, Yao J (1997) Structural control: past, present, and future. *J Eng Mech* 123(9):897–971
- Ikeda Y (2009) Active and semi-active vibration control of buildings in Japan – practical applications and verification. *Struct Control Health Monit* 16(7–8):703–723
- Kobori T, Koshika N, Yamada K, Ikeda Y (1991) Seismic -response-controlled structure with active mass driver system. Part 1: design. *Earthquake Eng Struct Dyn* 20:133–149
- Kobori T, Takahashi M, Nasu T, Niwa N, Ogasawara K (1993) Seismic response controlled structure with active variable stiffness system. *Earthquake Eng Struct Dyn* 22:925–941
- Leonard M (1990) *Dynamics and control of structures*. Wiley, USA
- Lin C, Chen L, Chen C (2007) RCMAC Hybrid control for MIMO uncertain non linear system using sliding mode technology. *IEEE Trans Neural Netw* 18(3):708–720

- Makris N, Hill D, Burton S, Jordan M (1995) Electrorheological fluid damper for seismic protection of structures. In: Proceedings of the smart structures and materials, San Diego, pp 184–194
- Martin RC, Soong TT (1976) Modal control of multistorey structures. *ASCE J Eng Mech* 102:613–623
- Nagarajaiah S, Mate D (1998). Semi-active control of continuously variable stiffness system. In: Proceedings of the 2nd world conference on structural control, Kyoto, vol 1, pp 397–405
- Narasimhan S, Nagarajaiah S, Johnson E, Gavin H (2003) Smart base isolated building benchmark problem. In: 16th ASCE engineering mechanics conference, 16–18 July 2003. University of Washington, Seattle
- Ohrai S, Kobori T, Sakamoto M, Koshika N, Nishimura I, Sasaki K, Kondo A, Fukushima I (1994) Development of active-passive composite tuned mass damper and an application to the high rise building. In: Proceedings of the first world conference on structural control, Pasadena, California, pp 100–109
- Pnevmatikos N, Gantes C (2009) Sliding mode control for structures based on the frequency content of the earthquake loading. *Smart Struct Syst* 5(3):209–221
- Pnevmatikos N, Gantes C (2010a) Control strategy for mitigating the response of structures subjected to earthquake actions. *Eng Struct* 32:3616–3628
- Pnevmatikos NG, Gantes JC (2010b) Design and control algorithm for structures equipped with active variable stiffness devices. *J Struct Control Health Monit* 17(6):591–613
- Pnevmatikos N, Gantes C (2011) Influence of time delay and saturation capacity to the response of controlled structures under earthquake excitations. *Smart Struct Syst Int J* 8(5):449–470
- Pnevmatikos NG, Kallivokas LF, Gantes CJ (2004) Feed-forward control of active variable stiffness systems for mitigating seismic hazard in structures. *Eng Struct* 26:471–483
- Preumont A (2002) *Vibration control of active structures, an introduction*, 2nd edn. Kluwer, The Netherlands
- Reinhorn AM, Soong TT, Lin RC, Wang YP, Fukao Y, Abe H, Nakai M (1989) 1:4 scale model studies of active tendon systems and active mass dampers for aseismic protection, technical report NCEER-89-0026
- Slotin JJ, Li W (1991) *Applied non linear control*. Prentice Hall, New Jersey, USA
- Soong TT (1990) *Active structural control: theory and practice*. Longman Scientific and Technical/Wiley, London/New York
- Soong TT, Spencer BF Jr (2002) Supplemental energy dissipation: state-of-the-art and state-of-the practice. *Eng Struct* 24:243–259
- Spencer BF Jr, Nagarajaiah S (2003) State of the art of structural control. *J Struct Eng* 129(7):845–856
- Spencer BF, Dyke SJ, Sain MK, Carlson JD (1997) Phenomenological model for magnetorheological dampers. *J Eng Mech* 123(3):230–38
- Sriram N, Satish N, Erik J, Henri G (2003) Smart base isolated building benchmark problem. In: 16th ASCE engineering mechanics conference. University of Washington, Seattle
- Symans MD, Constantinou MC, Taylor DP, Garnujost KD (1994) Semi-active fluid viscous dampers for seismic response control. In: Proceedings of first world conference on structural control, Los Angeles, FA4 pp 3–12
- Utku S (1998) *Theory of adaptive structures: incorporative intelligent into engineering products*. CRC Press, London
- Winslow WM (1947) Methods and means for translating electrical impulses into mechanical forces. US patent 2,417,850
- Yang JN (1975) Application of optimal control theory to civil engineering structures. *J Eng Mech Div ASCE* 819–838
- Yang JN, Kim J-H, Agrawal AK (2000) A resetting semi-active stiffness damper for seismic response control. *J Struct Eng* 126:1427–1433
- Yi F, Dyke SJ, Caicedo JM, Carlson JD (2001) Experimental verification of multi-input seismic control strategies for smart dampers. *J Eng Mech ASCE* 127(11):1152–1164
- Yoshida O (2003) Torsionally coupled response control of earthquake excited asymmetric buildings: development and application of effective control systems using smart dampers. PhD thesis, Washington University, Sever Institute of Technology, Department of Civil Engineering (supervised: Dyke SJ)
- Zacharenakis EC, Arvanitis KG, Soldatos AG, Stavroulakis GE (2001) LQR and H_∞ optimal structural control in aseismic design. In: Proceedings of national mechanics congress, Thessaloniki, pp 240–245
- Zadeh LA (1965) Fuzzy sets. *Info Control* 8:338–353
- Zhou L, Chang C, Wang L (2003) Adaptive fuzzy control for nonlinear building-magnetorheological damper system. *J Struct Eng* 129(7):905–913

Advances in Online Structural Identification

Ka-Veng Yuen

Faculty of Science and Technology, University of Macau, Macau, China

Synonyms

Bayesian inference; Extended Kalman filter; Model class selection; Model updating; Nonparametric identification; Online updating; Outlier, structural health monitoring; System identification

Introduction

Structural health monitoring using dynamic response measurement has received a tremendous attention over the last decades. A number of methods have been developed, including the novelty measure technique (Worden 1997), the GA-based substructural identification methods (Koh and Shankar 2003), and the evolutionary strategy (Franco et al. 2004). On the other hand, the Bayesian inference (Beck and Katafygiotis 1998; Beck 2010; Yuen 2010a) using probability logic provides a rigorous solution to parametric identification and uncertainty quantification for different problems in structural and geotechnical engineering, such as modal identification using nonstationary noisy response measurements (Yuen and Katafygiotis 2005; Yuen et al. 2006a), ambient vibration survey (Yuen and Katafygiotis 2006; Yuen et al. 2006b; Yuen and Kuok 2010a), particulate matters (Hoi et al. 2009), fatigue problem (Papadimitriou et al. 2011), and model class selection (Worden and Hensman 2012). A detailed review of Bayesian methods for structural dynamics and civil engineering can be found in Yuen and Kuok (2011). In particular, the well-known Kalman filter (KF) is one of the most widely applied recursive Bayesian state estimation techniques for trajectory estimation of linear dynamical systems (Kalman and Bucy 1961). Based on the concept of the KF, the extended Kalman filter (EKF) was developed for nonlinear systems (Jazwinski 1970). By introducing an augmented state vector, the EKF can be applied to parametric identification problems, and it has become a standard technique for state tracking, system identification, and control design for dynamical systems.

In the KF or EKF, the covariance matrices of the process noise and measurement noise are required. The conventional way is to tune these noise covariance matrices in a trial-and-error manner. However, ad hoc selection may lead to biased estimation, misleading uncertainty estimation, and even divergence problems (Zhou and Luecke 1994). Insights have been given into the convergence mechanisms, asymptotic

behavior, and estimation performance of the KF and EKF. It was emphasized that discrepancy between the actual and prescribed noise covariance matrices degrades substantially the performance of the filters and induces divergence problems. Existing methodologies for the estimation of these noise covariance matrices can be classified into three categories: covariance matching techniques (Myers and Tapley 1976), correlation techniques (Mehra 1970), and Bayesian techniques (Zhou and Luecke 1994). However, the majority of the literature focused on stationary situation which is not satisfied for general applications.

In this chapter, a computationally efficient Bayesian approach is presented for online estimation of the noise parameters. On the filter performance, it prevents the possible divergence problem and ensures the accuracy of the estimates and the estimated uncertainty. On the adaptability, it takes into explicit consideration the non-stationarity of the excitation, response, and measurement noise. These features of the presented approach enhance the applicability and reliability for the practical usage of the KF and EKF for online structural identification.

Next, the treatment of outliers in dynamic response data for online updating is addressed. In practice, it often occurs that some data points deviate drastically from the model output. The presence of outliers indicates irregularities of the data and/or deficiency of the model or theory. On one hand, outliers may occur due to extraordinarily large measurement error, e.g., human error, sensor noise, sensor failures, unknown environmental disturbances, etc. On the other hand, outliers may also occur due to imperfection of theory, i.e., existence of unmodeled mechanisms.

Since the performance of KF and EKF is severely deteriorated in the presence of outliers, a number of methods have been developed to enhance the robustness to outliers for KF and EKF. The first class of methods uses a non-Gaussian likelihood model for the measurement noise distribution and/or process noise distribution since the Gaussian likelihood model is sensitive to outliers, e.g., the Gaussian sum

approximation (Sorenson and Alspach 1971). However, this class of methods is often computationally very demanding, especially for online updating. Furthermore, the conditional mean of the state vector is not available in some cases. Another class of methods attempts to assign different weights, which are some heuristic function of the data, to different data points (Durovic and Kovacevic 1999). Methods of this class require tuning of the threshold parameters. However, performance is deteriorated with improper choice of the weights due to the difficulty in the choice of the thresholds. Therefore, this class of methods should be utilized with special attention.

In this chapter, an outlier-resistant extended Kalman filter (OR-EKF) is presented for robust online structural parametric identification using dynamic response data, contaminated with outliers in addition to Gaussian noise. In this algorithm, a novel outlier detection algorithm is embedded into the EKF. It is capable for robust online estimation of structural parameters using outlier-contaminated dynamic response data. Instead of definite judgment on the outlieriness of a data point, the OR-EKF provides its outlier probability, which is the extension of the concept in Yuen and Mu (2012). Data points with outlier probability over 0.5 will be regarded as suspicious data points, and they will be discarded for the identification purpose. In contrast to other existing outlier detection criteria that require some subjective threshold (e.g., normalized residual larger than 2.5), the outlier probability threshold of 0.5 is intuitive in the presented approach.

The third issue to be introduced is on the online model class selection. The usual approach in system identification is to find the best/optimal model in a prescribed class of models, e.g., class of shear building models with uncertain inter-story stiffnesses. This problem is commonly referred to as parametric identification. The more general problem of model class selection has not been explored as intensively as parametric identification. It is well known that a more complicated model class often fits the data better than one which has fewer adjustable parameters.

However, an over-fitted model leads to poor predictions because the model parameters depend too much on the detail of the data, and consequently the measurement noise and the modeling error play an important role in the data fitting. Therefore, in order to select a suitable model class for identification purposes, it is necessary to penalize a complicated model albeit the quantification of this penalty is a nontrivial task. This was first recognized by H. Jeffreys who did pioneering work on the application of Bayesian methods (Jeffreys 1961). In the present context, the selected class of models should agree closely with the observed behavior of the system but otherwise be as simple as possible. In recent years, the Bayesian approach to model class selection has been further developed by showing that the evidence for each model class provided by data automatically enforces a quantitative expression of principle of model parsimony or of Ockham's razor (Gull 1988). As a result, no ad hoc penalty is needed as in some of the earlier work on model class selection. Applications in civil engineering include the damage detection (Lam et al. 2006), selection of nonlinear hysteretic models (Muto and Beck 2008), soil compressibility (Yan et al. 2009), ambient effects to structural modal parameters (Yuen and Kuok 2010b), and seismic attenuation relationships (Yuen and Mu 2011). In this chapter, a recently developed recursive algorithm is introduced for Bayesian online model class selection.

In section “[Structural Parametric Identification by Extended Kalman Filter](#),” online structural parametric identification using the EKF will be briefly reviewed. In section “[Online Identification of Noise Parameters](#),” an online identification algorithm for the noise parameters in the EKF is introduced. Then, in section “[Outlier-Resistant Extended Kalman Filter](#),” an online outlier detection algorithm is presented, and it is embedded into the EKF. This algorithm allows for robust structural identification in the presence of possible outliers. In section “[Online Bayesian Model Class Selection](#),” a recursive Bayesian model class selection method is presented for non-parametric identification problems.

Structural Parametric Identification by Extended Kalman Filter

Formulation

Consider a linear dynamical system with N_d degrees of freedom (DOFs):

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{r}(\mathbf{x}, \dot{\mathbf{x}}; \boldsymbol{\varphi}) = \mathbf{T}\mathbf{f}(t) \quad (1)$$

where $\mathbf{M} \in \mathbb{R}^{N_d \times N_d}$ is the mass matrix of the system; $\mathbf{r}(\mathbf{x}, \dot{\mathbf{x}}; \boldsymbol{\varphi})$ is the restoring force that is governed by the unknown structural parameter vector $\boldsymbol{\varphi} \in \mathbb{R}^{N_\varphi}$ to be identified; $\mathbf{f}$ is the zero-mean Gaussian excitation; and $\mathbf{T} \in \mathbb{R}^{N_d \times N_f}$ is the force distribution matrix. The mass matrix $\mathbf{M}$ is assumed known.

Then, the augmented state vector $\mathbf{y}(t) \in \mathbb{R}^{2N_d + N_\varphi}$ is introduced to include the displacement, velocity, and the unknown structural parameters:

$$\mathbf{y}(t) = [\mathbf{x}^T, \dot{\mathbf{x}}^T, \boldsymbol{\varphi}^T]^T \quad (2)$$

where the superscript T denotes the transpose a vector/matrix. As a result, Eq. 1 can be converted to the state-space form:

$$\frac{d}{dt} \mathbf{y}(t) = \begin{bmatrix} -\mathbf{M}^{-1} \mathbf{r}(\mathbf{x}, \dot{\mathbf{x}}; \boldsymbol{\varphi}) \\ \mathbf{0}_{N_\varphi \times 1} \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{N_d \times 1} \\ \mathbf{M}^{-1} \mathbf{T} \\ \mathbf{0}_{N_\varphi \times 1} \end{bmatrix} \mathbf{f}(t) \quad (3)$$

To simplify the notation, the rate vector and input distribution matrix are defined as follows:

$$\mathbf{s} \equiv \begin{bmatrix} -\mathbf{M}^{-1} \mathbf{r}(\mathbf{x}, \dot{\mathbf{x}}; \boldsymbol{\varphi}) \\ \mathbf{0}_{N_\varphi \times 1} \end{bmatrix}; \quad \mathbf{B} \equiv \begin{bmatrix} \mathbf{0}_{N_d \times N_f} \\ \mathbf{M}^{-1} \mathbf{T} \\ \mathbf{0}_{N_\varphi \times N_f} \end{bmatrix} \quad (4)$$

where $\mathbf{0}_{\alpha \times \beta}$ denotes the $\alpha \times \beta$ zero matrix.

The structural response is sampled at N_o DOFs with time step Δt , and the measurement at the k th time step can be represented by the observation vector:

$$\mathbf{z}_k = \mathbf{C}_d \mathbf{y}_k + \mathbf{n}_k \quad (5)$$

Where $\mathbf{y}_k \equiv \mathbf{y}(k\Delta t)$, $\mathbf{C}_d \in \mathbb{R}^{N_o \times (2N_d + N_\varphi)}$ is the observation matrix, and the measurement noise $\mathbf{n}_k$ is assumed to be Gaussian with zero-mean and covariance matrix $\sum \mathbf{n}_k \in \mathbb{R}^{N_o \times N_o}$. Note that the measurement noise $\mathbf{n}$ is assumed to be i.i.d. (independent and identically distributed) and statistically independent to the excitation $\mathbf{f}$.

In order to obtain a discrete state-space equation for computation, the functions in Eq. 4 is linearized for each time step $t \in [k\Delta t, (k+1)\Delta t)$ as follows:

$$\frac{d}{dt} \mathbf{y}(t) = \mathbf{s} \Big|_{\mathbf{y}(t)=\mathbf{y}_k} + \frac{\partial \mathbf{s}}{\partial \mathbf{y}} \Big|_{\mathbf{y}(t)=\mathbf{y}_k} (\mathbf{y}(t) - \mathbf{y}_k) + \mathbf{B}\mathbf{f}(t) \quad (6)$$

And it can be rearranged as follows:

$$\dot{\mathbf{y}}(t) = \mathbf{A}_k \mathbf{y}(t) + \mathbf{B}\mathbf{f}(t) + \mathbf{h}_k, \quad t \in [k\Delta t, (k+1)\Delta t] \quad (7)$$

where the system matrix $\mathbf{A}_k$ is given by

$$\mathbf{A}_k = \frac{\partial \mathbf{s}}{\partial \mathbf{y}} \Big|_{\mathbf{y}(t)=\mathbf{y}_k} = \begin{bmatrix} \mathbf{0}_{N_d \times N_d} & \mathbf{I}_{N_d} & \mathbf{0}_{N_d \times N_\varphi} \\ -\mathbf{M}^{-1} \frac{\partial \mathbf{r}}{\partial \mathbf{x}} & -\mathbf{M}^{-1} \frac{\partial \mathbf{r}}{\partial \dot{\mathbf{x}}} & -\mathbf{M}^{-1} \frac{\partial \mathbf{r}}{\partial \boldsymbol{\varphi}} \\ \mathbf{0}_{N_\varphi \times N_d} & \mathbf{0}_{N_\varphi \times N_d} & -\delta \mathbf{I}_{N_\varphi} \end{bmatrix} \quad (8)$$

and the vector $\mathbf{h}_k$ is given by

$$\mathbf{h}_k = \mathbf{r} \Big|_{\mathbf{y}(t)=\mathbf{y}_k} - \mathbf{A}_k \mathbf{y}_k \quad (9)$$

The variable δ in $\mathbf{A}_k$ in Eq. 8 is chosen as a small positive number to prevent singularity of the matrix $\mathbf{A}_k$; $\mathbf{I}_\alpha$, and $\mathbf{0}_{\alpha \times \beta}$ denote the $\alpha \times \alpha$ identity matrix and the $\alpha \times \beta$ zero matrix, respectively.

State Estimation and Parametric Identification by Kalman Filter

The augmented state vector $\mathbf{y}$ can be updated by the Kalman filter. First, the linearized state-space

equation in Eq. 7 can be discretized to a difference equation:

$$\mathbf{y}_{k+1} = \mathbf{A}_{d,k}\mathbf{y}_k + \mathbf{B}_{d,k}\mathbf{f}_k + \mathbf{h}_{d,k} \quad (10)$$

where $\mathbf{y}_{k+1} = \mathbf{y}((k+1)\Delta t)$, $\mathbf{A}_{d,k} = \exp(\mathbf{A}_k \Delta t)$, $\mathbf{B}_{d,k} = \mathbf{A}_k^{-1}(\mathbf{A}_{d,k} - \mathbf{I}_{2N_d+N_\varphi})\mathbf{B}$, $\mathbf{f}_k = \mathbf{f}(k\Delta t)$, and $\mathbf{h}_{d,k} = \mathbf{A}_k^{-1}(\mathbf{A}_{d,k} - \mathbf{I}_{2N_d+N_\varphi})\mathbf{h}_k$.

With the measurement $D_k \equiv \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k\}$ defined in Eq. 5, the predicted state at the $(k+1)^{\text{th}}$ time step ($\mathbf{y}_{k+1|k} \equiv \mathbb{E}[\mathbf{y}_{k+1} | \mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k]$) can be estimated by taking the conditional expected value of Eq. 10:

$$\mathbf{y}_{k+1|k} = \mathbf{A}_{d,k}\mathbf{y}_{k|k} + \mathbf{h}_{d,k} \quad (11)$$

where $\mathbf{y}_{k|k}$ is the updated state at the k th time step. In addition, the covariance matrix of $\mathbf{y}_{k+1|k}$ is

$$\Sigma_{\mathbf{y},k+1|k} = \mathbf{A}_{d,k}\Sigma_{\mathbf{y},k|k}\mathbf{A}_{d,k}^T + \mathbf{B}_{d,k}\Sigma_{\mathbf{f},k}\mathbf{B}_{d,k}^T \quad (12)$$

where $\Sigma_{\mathbf{f},k}$ is the covariance matrix of the excitation at the k th time step.

With a new data point $\mathbf{z}_{k+1}$, the updated state $\mathbf{y}_{k+1|k+1}$ and its covariance matrix $\Sigma_{\mathbf{y},k+1|k+1}$ can be obtained by using the following Kalman filter recursive formulae (Kalman and Bucy 1961):

$$\mathbf{y}_{k+1|k+1} = \mathbf{y}_{k+1|k} + \mathbf{G}_{k+1}(\mathbf{z}_{k+1} - \mathbf{C}_d\mathbf{y}_{k+1|k}) \quad (13)$$

$$\Sigma_{\mathbf{y},k+1|k+1} = (\mathbf{I}_{2N_d+N_\varphi} - \mathbf{G}_{k+1}\mathbf{C}_d)\Sigma_{\mathbf{y},k+1|k} \quad (14)$$

where $\mathbf{G}_{k+1}$ is called the Kalman filter gain (Kalman and Bucy 1961)

$$\mathbf{G}_{k+1} = \Sigma_{\mathbf{y},k+1|k}\mathbf{C}_d^T(\mathbf{C}_d\Sigma_{\mathbf{y},k+1|k}\mathbf{C}_d^T + \Sigma_{\mathbf{n},k})^{-1} \quad (15)$$

Fading Memory

Fading memory filtering was developed to compensate the modeling error in the identification

process (Sorensen and Sacks 1971). Here, it is employed to track the time-varying structural and noise parameters. The underlying concept of fading memory filtering is to discount the contribution of the past data gradually for adaptive identification of the time-varying parameters. This can be accomplished by assigning an appropriate weighting, which is called the fading factor $\mu(>1)$, to enlarge the estimated covariance matrix at every time step so that the information from the past data will fade out gradually. In particular, the covariance matrix of the filtered augmented state in Eq. 14 is modified as follows:

$$\Sigma_{\mathbf{y},k+1|k+1} = \mu(\mathbf{I}_{2N_d+N_\varphi} - \mathbf{G}_{k+1}\mathbf{C}_d)\Sigma_{\mathbf{y},k+1|k} \quad (16)$$

where μ is the widely adopted exponential fading factor (Sorensen and Sacks 1971)

$$\mu = \exp\left(\frac{\ln\mu_\sigma^2}{N_\mu}\right) \quad (17)$$

where $\mu \geq 1$. The effect of this exponential fading factor is to dilute the information of a data point with a factor of μ_σ^2 , in terms of the contribution to the posterior variance, after every N_μ time steps.

Online Identification of Noise Parameters

In this section, a Bayesian probabilistic approach is presented for online estimation of the noise parameters of the process noise and measurement noise. Section “[Parameterization and Bayesian Formulation](#)” introduces the parameterization of the noise covariance matrices and the Bayesian formulation. Thereafter, the online identification algorithm is presented in section “[Online Estimation of Noise Parameters](#).”

Parameterization and Bayesian Formulation

First, the covariance matrices of the process noise and measurement noise are parameterized

as $\Sigma_{f,k} = \Sigma_{f,k}(\theta_{f,k})$ and $\Sigma_{n,k+1} = \Sigma_{n,k+1}(\theta_{n,k+1})$, respectively. In other words, both matrices are time varying in order to allow online tracking. The uncertain noise parameter vector is given by

$$\theta_{k+1} = \left[\theta_{f,k}^T, \theta_{n,k+1}^T \right]^T \in \mathbb{R}^{N_\theta} \quad (18)$$

This parameterization allows for the tracking problem for general nonstationary situations. In the following, a concurrent procedure is introduced to estimate the noise parameter vector. As a result, not only the optimal estimation can be obtained but also the associated uncertainty can be quantified.

When a new measurement $\mathbf{z}_{k+1}$ is available, the noise parameters can be updated. Using the Bayes' theorem, the posterior probability density function (PDF) of the noise parameter vector given the measurement data set D_{k+1} is given by (Yuen 2010a)

$$p(\theta_{k+1}|D_{k+1}) = \frac{p(\theta_{k+1}|D_k)p(\mathbf{z}_{k+1}|\theta_{k+1}, D_k)}{p(\mathbf{z}_{k+1}|D_k)} \quad (19)$$

where $p(\theta_{k+1}|D_k)$ is the prior PDF of θ_{k+1} ; $p(\mathbf{z}_{k+1}|\theta_{k+1}, D_k)$ is the likelihood function; and $p(\mathbf{z}_{k+1}|D_k)$ is a normalizing constant such that the integral of the posterior PDF over the entire parameter space is unity. The prior PDF is approximated as a Gaussian distribution:

$$p(\theta_{k+1}|D_k) = (2\pi)^{-N_\theta/2} |\Sigma_{\theta,k+1|k}|^{-1/2} \exp \left[-\frac{1}{2} (\theta_{k+1} - \theta_{k+1|k})^T \Sigma_{\theta,k+1|k}^{-1} (\theta_{k+1} - \theta_{k+1|k}) \right] \quad (20)$$

Given the measurement D_k , the one-step ahead predictor for the noise parameters is taken as the updated noise parameters of the previous time step:

$$\theta_{k+1|k} = \theta_{k|k} \quad (21)$$

On the other hand, the associated covariance matrix is assumed to be the following:

$$\Sigma_{\theta,k+1|k} = \mu \Sigma_{\theta,k|k} \quad (22)$$

where the fading factor μ is defined in Eq. 17. Here, it is assumed that the noise covariances are time slowly varying so the ensemble mean can be approximately estimated by the temporal average with certain time delay.

On the other hand, the likelihood function is given by

$$p(\mathbf{z}_{k+1}|\theta_{k+1}, D_k) = (2\pi)^{-N_o/2} |\Sigma_{\mathbf{z},k+1|k}|^{-1/2} \exp \left[-\frac{1}{2} (\mathbf{z}_{k+1} - \mathbf{z}_{k+1|k})^T \Sigma_{\mathbf{z},k+1|k}^{-1} (\mathbf{z}_{k+1} - \mathbf{z}_{k+1|k}) \right] \quad (23)$$

where $\mathbf{z}_{k+1|k}$ and $\Sigma_{\mathbf{z},k+1|k}$ are readily obtained from Eqs. 11 and 12 given the measurement data set D_k :

$$\mathbf{z}_{k+1|k} = \mathbf{C}_d \mathbf{A}_{d,k} \mathbf{y}_{k|k} + \mathbf{C}_d \mathbf{h}_{d,k} \quad (24)$$

$$\Sigma_{\mathbf{z},k+1|k} = \mathbf{C}_d \mathbf{A}_{d,k} \Sigma_{\mathbf{y},k|k} \mathbf{A}_{d,k}^T \mathbf{C}_d^T + \mathbf{C}_d \mathbf{B}_{d,k} \Sigma_{f,k|k} \mathbf{B}_{d,k}^T \mathbf{C}_d^T + \Sigma_{n,k+1|k} \quad (25)$$

By substituting Eq. 20–Eq. 23 to Eq. 19, the posterior PDF becomes

$$p(\theta_{k+1}|D_{k+1}) = \frac{(2\pi)^{-(N_\theta+N_o)/2}}{p(\mathbf{z}_{k+1}|D_k) |\mu \Sigma_{\theta,k|k}|^{1/2}} |\Sigma_{\mathbf{z},k+1|k}|^{-1/2} \exp \left[-\frac{1}{2\mu} (\theta_{k+1} - \theta_{k|k})^T \Sigma_{\theta,k|k}^{-1} (\theta_{k+1} - \theta_{k|k}) - \frac{1}{2} (\mathbf{z}_{k+1} - \mathbf{z}_{k+1|k})^T \Sigma_{\mathbf{z},k+1|k}^{-1} (\mathbf{z}_{k+1} - \mathbf{z}_{k+1|k}) \right] \quad (26)$$

where $\mathbf{z}_{k+1|k}$ and $\Sigma_{\mathbf{z},k+1|k}$ are given by Eqs. 24 and 25, respectively. The objective function $J(\boldsymbol{\theta}_{k+1})$ is defined as the negative logarithm of

$$\begin{aligned} J(\boldsymbol{\theta}_{k+1}) = & \frac{1}{2} \ln |\Sigma_{\mathbf{z},k+1|k}| + \frac{1}{2\mu} (\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_{k|k})^T \Sigma_{\boldsymbol{\theta},k|k}^{-1} (\boldsymbol{\theta}_{k+1} - \boldsymbol{\theta}_{k|k}) \\ & + \frac{1}{2} (\mathbf{z}_{k+1} - \mathbf{z}_{k+1|k})^T \Sigma_{\mathbf{z},k+1|k}^{-1} (\mathbf{z}_{k+1} - \mathbf{z}_{k+1|k}) \end{aligned} \quad (27)$$

The updated noise parameter vector $\boldsymbol{\theta}_{k+1|k+1}$ can be obtained by maximizing the posterior PDF $p(\boldsymbol{\theta}_{k+1}|D_{k+1})$ in Eq. 26, and it is equivalent to minimizing the objective function $J(\boldsymbol{\theta}_{k+1})$ in Eq. 27:

$$\boldsymbol{\theta}_{k+1|k+1} = \arg \min_{\boldsymbol{\theta}_{k+1}} J(\boldsymbol{\theta}_{k+1}) \quad (28)$$

Furthermore, the uncertainty of the updated noise parameter vector $\boldsymbol{\theta}_{k+1|k+1}$ can be represented by its covariance matrix $\Sigma_{\boldsymbol{\theta},k+1|k+1}$ which is equal to the inverse of the Hessian matrix of the objective function calculated at $\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_{k+1|k+1}$:

$$\Sigma_{\boldsymbol{\theta},k+1|k+1} = [\mathbf{H}_J(\boldsymbol{\theta}_{k+1|k+1})]^{-1} \quad (29)$$

where the Hessian matrix is given by $\mathbf{H}_J(\boldsymbol{\theta}_{k+1|k+1}) = [\nabla^2 J(\boldsymbol{\theta}_{k+1})]_{\boldsymbol{\theta}_{k+1}=\boldsymbol{\theta}_{k+1|k+1}}$, and it can be computed using the finite difference method.

Online Estimation of Noise Parameters

In this subsection, a computationally efficient algorithm is presented for online updating of the noise parameter vector $\boldsymbol{\theta}_{k+1|k+1}$ and the associated covariance matrix $\Sigma_{\boldsymbol{\theta},k+1|k+1}$. Starting with an arbitrary initial condition $\boldsymbol{\theta}_{0|0}$ and arbitrarily positive definite matrix $\Sigma_{\boldsymbol{\theta},0|0}$, the noise parameter vector and its associated covariance matrix can be updated based on Eqs. 28 and 29. As aforementioned, the noise parameter vector can be updated by solving the optimization problem in Eq. 28. However, due to the large prior uncertainty and numerical considerations, a training process is necessary for a preliminary solution

the posterior PDF without taking the terms that do not depend on the noise parameters:

before solving directly this optimization problem. In this training process, a heuristic local search method is utilized. Thereafter, a modified Newton's method is used for the operating stage. Details of this estimation scheme are presented as follows.

Training Stage

At the beginning of the identification process, the Gaussian approximation is inaccurate due to the large prior uncertainty, i.e., large difference between the prior mean and the actual values, and large prior variances. Therefore, solving directly the optimization problem in Eq. 28 may result in computational problems. Herein, a training process is introduced to obtain a preliminary solution before the long-term operating stage. During the training process, a heuristic local search method is applied to the optimization problem. This method provides a generic tool for complicated optimization problems. The basic principle is to conduct the optimization within a prescribed parameter set. By exhaustive search from this finite set, the optimal (suboptimal) solution is the one that gives the smallest objective function value. Specifically, the candidate parameter set of the $(k+1)$ th time step is defined as

$$\begin{aligned} \Theta_{k+1} \equiv & \left\{ \boldsymbol{\theta}_{k+1} \in \mathbb{R}^{N_\theta} : \theta_{k+1}^{(l)} = v \theta_{k|k}^{(l)}, l = 1, \dots, \right. \\ & \left. N_\theta, v \in \{1/2, 1, 2\} \right\} \end{aligned} \quad (30)$$

where $\theta_{k+1}^{(l)}$ and $\theta_{k|k}^{(l)}$ are the l th component of the noise parameter vector $\boldsymbol{\theta}_{k+1}$ and the l th

component of the updated noise parameter vector $\theta_{k|k}$, respectively. By considering all the combinations of the N_θ noise parameters, there are 3^{N_θ} candidate solutions in the candidate parameter set Θ_{k+1} . The objective function $J(\theta_{k+1})$ is evaluated for all candidate solutions. Then, the noise parameter vector is updated under the optimization criterion given by

$$\theta_{k+1|k+1} = \arg \min_{\theta_{k+1}} J(\theta_{k+1}) \text{ where } \theta_{k+1} \in \Theta_{k+1} \quad (31)$$

This heuristic local search method is the online version of the half-or-double optimization method in Yuen et al. (2007). If the updated noise parameter vectors of ten consecutive time steps remain unchanged, the training process will be terminated. However, this training process is enforced to be not shorter than one fundamental period of the underlying dynamical system (or its equivalent linear system). This one-period requirement ensures that sufficient information can be gained from the data for the preliminary estimation of the augmented state vector.

Operating Stage

After the training process, the updated noise parameter vector is obtained using a modified Newton's method in the operating stage. Instead of an iterative procedure in the Newton's method, it is necessary to take only one step because the updated parameter vector of the previous time step is very close to the solution. Specifically, the updated noise parameter vector can be obtained as follows:

$$\theta_{k+1|k+1} = \theta_{k|k} - \Sigma_{\theta, k|k} \left(\nabla J(\theta_k) |_{\theta_k = \theta_{k|k}} \right) \quad (32)$$

where the gradient $\nabla J(\theta_k) |_{\theta_k = \theta_{k|k}}$ can be computed using the finite difference method. Since the information carried by one data point is limited, the updated noise parameter vector of the previous time step provides an accurate initial estimation of the current time step. Therefore, Eq. 32 is sufficient without iteration.

Outlier-Resistant Extended Kalman Filter

In this section, the outlier-resistant extended Kalman filter (OR-EKF) is introduced for online outlier detection and robust structural parametric identification using dynamic response data. In this algorithm, an online outlier detection algorithm is embedded into the EKF. Section “[Online Outlier Detection by Outlier Probability](#)” introduces the concept of outlier probability, and it will be utilized for online outlier detection. This outlier detection algorithm is embedded in the EKF for online structural identification. Section “[Procedure of the Outlier-Resistant Extended Kalman Filter](#)” summarizes the procedure of the OR-EKF algorithm.

Online Outlier Detection by Outlier Probability

In this section, the concept of outlier probability and an online outlier detection algorithm are introduced. Instead of definite judgment on the “outlierness” of a data point being an outlier, this algorithm provides the outlier probability, which is the extension of the concept in Yuen and Mu (2012) for linear regression problems, for the measurement in each time step. The outlier probability is a function of the normalized residual, which is defined as the difference between the measured value and the corresponding one-step-ahead predictor, normalized by its standard deviation. Data points with outlier probability over 0.5 are regarded as suspicious data points, and they will be discarded for the identification purpose.

Outlier Probability

Recall that the predicted state $\mathbf{y}_{k+1|k}$ can be calculated using Eq. 11. Therefore, the one-step-ahead predictor of the measurements is readily obtained:

$$\mathbf{z}_{k+1|k} = \mathbf{C}_d \mathbf{y}_{k+1|k} \quad (33)$$

Use $z_{k+1}^{(s)}$ to denote the s th component of the observation vector $\mathbf{z}_{k+1}$ and $z_{k+1|k}^{(s)}$, $s = 1, 2, \dots, N_o$, to denote the s th component of its

one-step-ahead predictor $\mathbf{z}_{k+1|k}$. Hereafter, the superscript (s) is used to denote the sth component of a vector. Then, the normalized residual (prediction error) for $\mathbf{z}_{k+1}^{(s)}$ can be defined (Myers and Tapley 1976):

$$\varepsilon_{k+1}^{(s)} = \left(z_{k+1}^{(s)} - \mathbf{z}_{k+1|k}^{(s)} \right) / \sigma_{\varepsilon, k+1}^{(s)} \quad (34)$$

where $\sigma_{\varepsilon, k+1}^{(s)}$ is the standard deviation for the sth measured channel of the difference between the measurement and the corresponding one-step-ahead predictor. Since the probability model for the measurement noise is Gaussian, the probability of a data point falling outside the interval $(-|\varepsilon_{k+1}^{(s)}|, |\varepsilon_{k+1}^{(s)}|)$ is

$$Q_{k+1} = 2\Phi\left(-\left|\varepsilon_{k+1}^{(s)}\right|\right) \quad (35)$$

where $\Phi(\cdot)$ is the cumulative distribution function (CDF) of the standard Gaussian random variable. Then, a moving time window is introduced to include not more than N_w previous regular data points, and the set $\mathbf{R}_k^{(s)}$ is introduced to include the absolute normalized residuals of the regular data points in this time window:

$$\mathbf{R}_k^{(s)} = \left\{ \left| \varepsilon_l^{(s)} \right| : z_l^{(s)} \text{ is a regular data point, } l = \max(1, k - N_w), \dots, k \right\} \quad (36)$$

where $n(\mathbf{R}_k^{(s)})$ is used to denote the number of elements in the set $\mathbf{R}_k^{(s)}$. Define η_{k+1} as the number of elements, among the set $\mathbf{R}_k^{(s)}$, with absolute normalized residuals larger than $|\varepsilon_{k+1}^{(s)}|$. In other words, η_{k+1} is the number of previous regular data points, within the aforementioned moving time window, that has larger absolute normalized residual than the current one. This is equivalent to counting the numbers of the normalized residuals of the previous regular points falling outside the interval $(-|\varepsilon_{k+1}^{(s)}|, |\varepsilon_{k+1}^{(s)}|)$ in the moving time window.

In order to determine the outlierness of a data point $\mathbf{z}_{k+1}^{(s)}$, the probability that τ points, out of $n(\mathbf{R}_k^{(s)}) + 1$ data points, falling outside the interval $(-|\varepsilon_{k+1}^{(s)}|, |\varepsilon_{k+1}^{(s)}|)$ is considered. Note that τ is a random variable, and it follows the binomial

distribution with probability Q_{k+1} . Then, the outlier probability for $\mathbf{z}_{k+1}^{(s)}$ can be defined as the probability that no more than η_{k+1} samples drawn from the normalized residuals among the $(\mathbf{R}_k^{(s)})$ previous regular data points fall outside the interval $(-|\varepsilon_{k+1}^{(s)}|, |\varepsilon_{k+1}^{(s)}|)$. This probability can be obtained by considering the total probability:

$$P_o\left(\mathbf{z}_{k+1}^{(s)}\right) = \sum_{\tau=0}^{\eta_{k+1}} C\left(n\left(\mathbf{R}_k^{(s)}\right), \tau\right) Q_{k+1}^{\tau} \left(1 - Q_{k+1}\right)^{n\left(\mathbf{R}_k^{(s)}\right) - \tau} \quad (37)$$

where $C\left(n\left(\mathbf{R}_k^{(s)}\right), \tau\right) = \frac{n\left(\mathbf{R}_k^{(s)}\right)!}{\left(n\left(\mathbf{R}_k^{(s)}\right) - \tau\right)! \tau!}$ is the

binomial coefficient. Note that introduction of this moving window is to release the memory and computational burden in considering the entire history. On the other hand, due to possible condition changes of the excitation and the underlying structure, it is suitable to consider only the past data points in a reasonable neighborhood of the current time step. It is suggested to use a value for N_w that corresponds to approximately 100 fundamental periods of the structure.

Efficient Screening Criteria

In order to enhance the computational efficiency, an efficient screening rule is introduced. First, a data point is classified as a regular point when $|\varepsilon_{k+1}^{(s)}| \leq \xi_L$, a conservatively small bound (e.g., $\xi_L = 2$). On the other hand, a data point is classified as an outlier when $|\varepsilon_{k+1}^{(s)}| \geq \xi_U$. This bound can be obtained by solving $[1 - 2\Phi(-\xi_U)]^{N_w+1} = 0.5$. As a result, it is given by

$$\xi_U = -\Phi^{-1}\left[\left(\frac{1}{2} - 2^{-\frac{1}{N_w+1}}\right)\right] \quad (38)$$

where Φ^{-1} is the quantile function which is the inverse function of the CDF of the standard Gaussian random variable. For example, $\xi_U = 3.9787$ for $N_w = 10,000$. It can be easily shown that the outlier probability of the data points with absolute normalized residuals larger than ξ_U is larger than 0.5. Therefore, computation of the outlier probability is necessary only when

the absolute value of the normalized residual of the measurement lies within (ξ_L, ξ_U) . This helps to determine efficiently the outlierlikeness for a large portion of data points without computing the outlier probability.

Procedure of the Outlier-Resistant Extended Kalman Filter

The procedure of the OR-EKF is summarized as follows.

Training Process

The initial value of $\sigma_{e,1}^{(s)}$ can be estimated as follows:

- (i) Implement the EKF for roughly ten fundamental periods of the underlying system.
- (ii) Calculate the residuals $|z_k^{(s)} - z_{k|k-1}^{(s)}|$ for these data points (ignoring the first one period of data points), and sort them in the ascending order. Then, the standard deviation $\sigma_{e,1}^{(s)}$ can be estimated as the value of the 68-percentile point. This is a robust estimator of the standard deviation for possible presence of outliers.

Operating Process

Compute the ξ_U using Eq. 38 and initialize $R_0^{(s)} = \phi$, an empty set for $s = 1, \dots, N_o$.

1. Calculate the one-step-ahead predictor $\mathbf{z}_{k+1|k}$ by Eq. 33, the s th element of the normalized residual $|e_{k+1}^{(s)}|$ by Eq. 34.
2.
 - (a) If $|e_{k+1}^{(s)}| \leq \xi_L$ ($=2$), $z_{k+1}^{(s)}$ will be classified as a regular data point.
 - (b) If $|e_{k+1}^{(s)}| \geq \xi_U$, $z_{k+1}^{(s)}$ will be classified as an outlier.
 - (c) When $\xi_L < |e_{k+1}^{(s)}| < \xi_U$, calculate Q_{k+1} using Eq. 35 and count η_{k+1} from $R_k^{(s)}$. Then, compute the outlier probability $P_o(z_{k+1}^{(s)})$ using Eq. 37. If $P_o(z_{k+1}^{(s)}) < 0.5$, $z_{k+1}^{(s)}$ will be classified as a regular point. Otherwise, it will be concluded as a suspicious data point, and it will be discarded for identification purpose.
3. According to the outlier detection result in step (2), update the sets $R_{k+1}^{(s)}$, $s = 1, \dots, N_o$. If $z_{k+1}^{(s)}$ is an outlier, $R_{k+1}^{(s)} = R_k^{(s)}$.

If $z_{k+1}^{(s)}$ is a regular point, $R_{k+1}^{(s)} = R_k^{(s)} \cup \{|e_{k+1}^{(s)}|\}$ if $n(R_k^{(s)}) < N_w$. Otherwise, $R_{k+1}^{(s)} = R_k^{(s)} \cup \{|e_{k+1}^{(s)}|\} - \{|e_{old}^{(s)}|\}$, where $|e_{old}^{(s)}|$ is the oldest element in the set $R_k^{(s)}$.

4. Update $\sigma_{e,k+2}^{(s)}$ with the elements in $R_{k+1}^{(s)}$. This can be done by using recursive formula.
5. Remove the outliers from $\mathbf{z}_{k+1}$. Note that the observation matrix $\mathbf{C}_d$ and the noise covariance matrix $\Sigma_{n,k+1}$ have to be modified accordingly.
6. Update the state vector and the associated covariance matrix using the measurement obtained from step (5) using Eq. 13 to Eq. 15. If all N_0 elements of the $(k+1)$ th data points are suspicious measurements, $\mathbf{y}_{k+1|k+1} = \mathbf{y}_{k+1|k}$ and $\Sigma_{y,k+1|k+1} = \Sigma_{y,k+1|k}$.
7. Continue for the next time step.

One can refer to Mu and Yuen (2014) for numerical examples.

Online Bayesian Model Class Selection

Bayesian model class selection is utilized for selecting the most plausible model class from a set of N_C dynamic model class candidates $C_1, C_2, \dots, C_{N_C}$ by considering their plausibility $P(C_j|D)$ conditional on the available set of dynamic measurement D (Beck and Yuen 2004; Yuen 2010b):

$$P(C_j|D) = \frac{P(C_j)p(D|C_j)}{p(D)} \quad (39)$$

where the denominator $p(D)$ is the normalizing constant; $P(C_j)$ is the prior plausibility of model class C_j . In general, a noninformative prior can be used, i.e., $P(C_j) = 1/N_C$, where $j = 1, \dots, N_C$, and $p(D|C_j)$ is the evidence given by

$$p(D|C_j) = \int_{\Theta_j} p(D|\boldsymbol{\theta}; C_j) p(\boldsymbol{\theta}|C_j) d\boldsymbol{\theta} \quad (40)$$

where Θ_j denotes the parameter space of model class C_j . However, direct numerical computation of this integral is in general computationally

prohibitive unless the number of uncertain parameters is very small (say 3 or less). To overcome this computational obstacle, an asymptotic expansion was developed for the globally identifiable cases (Beck and Yuen 2004):

$$p(D|C_j) \approx (2\pi)^{N_j/2} p(D|\hat{\boldsymbol{\theta}}; C_j) p(\hat{\boldsymbol{\theta}}|C_j) \left| \mathbf{H}_j(\hat{\boldsymbol{\theta}}) \right|^{-1/2} \quad (41)$$

where N_j is the number of uncertain parameters of model class C_j and $\hat{\boldsymbol{\theta}}$ is the parameter vector that maximizes the posterior PDF, which is proportional to the integrand on the right-hand side of Eq. 40.

In the locally identifiable cases (Katafygiotis and Beck 1998), the product of the likelihood function and the prior PDF can be approximately by weighted Gaussian distributions:

$$p(D|\boldsymbol{\theta}; C_j) p(\boldsymbol{\theta}|C_j) \approx (2\pi)^{N_j/2} \sum_{i=1}^{N_i} p(D|\hat{\boldsymbol{\theta}}_i; C_j) p(\hat{\boldsymbol{\theta}}_i|C_j) \left| \mathbf{H}_j(\hat{\boldsymbol{\theta}}_i) \right|^{-1/2} G\left(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}, \mathbf{H}_j(\hat{\boldsymbol{\theta}}_i)^{-1}\right) \quad (42)$$

where $\hat{\boldsymbol{\theta}}_i$ is the i th local optimal point, $i = 1, \dots, N_i$, and $G\left(\boldsymbol{\theta}; \hat{\boldsymbol{\theta}}_i, \mathbf{H}_j(\hat{\boldsymbol{\theta}}_i)^{-1}\right)$ denotes the multivariate Gaussian distribution for the random vector $\boldsymbol{\theta}$ with mean $\hat{\boldsymbol{\theta}}_i$ and covariance matrix $\mathbf{H}_j(\hat{\boldsymbol{\theta}}_i)^{-1}$. Then, the evidence is given as

$$p(D|C_j) \approx (2\pi)^{N_j/2} \sum_{i=1}^{N_i} p(D|\hat{\boldsymbol{\theta}}_i; C_j) p(\hat{\boldsymbol{\theta}}_i|C_j) \left| \mathbf{H}_j(\hat{\boldsymbol{\theta}}_i) \right|^{-1/2} \quad (43)$$

For general unidentifiable cases, the evidence integral in Eq. 40 can be computed using the transitional Markov chain Monte Carlo (TMCMC) method (Ching and Chen 2007).

Next, an online Bayesian model class selection algorithm is introduced. It was first developed to model the transportation system of particulate matters (Hoi et al. 2011). The plausibility of model class C_j conditional on the measured data up to the $(k+1)$ th time step $D_{k+1} = \{\mathbf{z}_1, \dots, \mathbf{z}_{k+1}\}$ can be rewritten into the following form using the Bayes' theorem:

$$P(C_j|D_{k+1}) = \frac{p(\mathbf{z}_{k+1}|D_k; C_j) P(C_j|D_k)}{p(\mathbf{z}_{k+1}|D_k)} \quad (44)$$

where $p(\mathbf{z}_{k+1}|D_k; C_j)$ denotes the evidence of model class C_j conditional on the data of the

previous k time steps. The denominator $p(\mathbf{z}_{k+1}|D_k)$ is a normalizing constant that does not depend on the model class. Note that the plausibility $P(C_j|D_0)$ is deduced to the prior plausibility of the model class $P(C_j)$. By using the theorem of total probability, the conditional evidence $p(\mathbf{z}_{k+1}|D_k; C_j)$ can be expressed as

$$p(\mathbf{z}_{k+1}|D_k; C_j) = \int_{\Theta_j} p(\mathbf{z}_{k+1}|\boldsymbol{\theta}_k; D_k; C_j) p(\boldsymbol{\theta}_k|D_k; C_j) d\boldsymbol{\theta}_k \quad (45)$$

where $\boldsymbol{\theta}_k$ denotes the parameter vector containing all the uncertain parameters of model class C_j at the k th time step and Θ_j denotes the parameter space for model class C_j . The factor $p(\mathbf{z}_{k+1}|\boldsymbol{\theta}_k; D_k; C_j)$ in the integrand is the conditional likelihood function of model class C_j at the $(k+1)$ th time step. It represents the level of data fitting of the model with a given parameter vector $\boldsymbol{\theta}_k$. The second factor in the integrand is the posterior PDF of the parameter vector $\boldsymbol{\theta}_k$ conditional on the previous data points $\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k$. For globally identifiable cases, an asymptotic expansion of this integral can be obtained in a similar fashion as Eq. 41:

$$p(\mathbf{z}_{k+1}|D_k; C_j) \approx (2\pi)^{N_j/2} p(\mathbf{z}_{k+1}|\boldsymbol{\theta}_k^*; D_k; C_j) p(\boldsymbol{\theta}_k^*|D_k; C_j) \left| \mathbf{H}_j(\boldsymbol{\theta}_k^*) \right|^{-1/2} \quad (46)$$

where N_j denotes the number of uncertain parameters in model class C_j and $\boldsymbol{\theta}_k^* = \boldsymbol{\theta}_{k|k+1}$ is the parameter vector that maximizes the integrand of

Eq. 45 but it will be approximated by $\boldsymbol{\theta}_k^* \approx \boldsymbol{\theta}_{k|k}$. The maximum conditional likelihood of model class C_j evaluated at $\boldsymbol{\theta}_{k|k}$ is given by

$$\begin{aligned} p(\mathbf{z}_{k+1}|\boldsymbol{\theta}_k^*; D_k; C_j) &\approx (2\pi)^{-N_o/2} |\mathbf{C}_d \boldsymbol{\Sigma}_{\mathbf{y}, k+1|k} \mathbf{C}_d^T + \boldsymbol{\Sigma}_{\mathbf{n}, k+1}|^{-1/2} \\ &\times \exp \left[-\frac{1}{2} \left(\mathbf{z}_{k+1} - \mathbf{C}_d \mathbf{y}_{k+1|k} \right)^T (\mathbf{C}_d \boldsymbol{\Sigma}_{\mathbf{y}, k+1|k} \mathbf{C}_d^T + \boldsymbol{\Sigma}_{\mathbf{n}, k+1})^{-1} \left(\mathbf{z}_{k+1} - \mathbf{C}_d \mathbf{y}_{k+1|k} \right) \right] \end{aligned} \quad (47)$$

The maximum posterior probability density of $\boldsymbol{\theta}_k^*$ conditional on D_k for model class C_j is given by

$$p(\boldsymbol{\theta}_k^*|D_k; C_j) \approx (2\pi)^{-N_j/2} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}, k|k}|^{-1/2} \quad (48)$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\theta}, k|k}$ denotes the covariance matrix of the updated parameters. The matrix $\mathbf{H}_j(\boldsymbol{\theta}_k^*)$ denotes the Hessian matrix of the negative natural logarithm of the integrand in Eq. 45 with respect to the parameter vector, evaluated at $\boldsymbol{\theta}_k^*$:

$$\begin{aligned} \mathbf{H}_j(\boldsymbol{\theta}_k^*) &= -[\nabla \ln p(\mathbf{z}_{k+1}|\boldsymbol{\theta}_k; D_k; C_j) \nabla^T] |_{\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k|k}} \\ &+ \sum_{\boldsymbol{\theta}_{k|k}}^{-1} \end{aligned} \quad (49)$$

where ∇ denotes the gradient operator with respect to the parameter vector $\boldsymbol{\theta}_k$. The term $[\nabla \ln p(\mathbf{z}_{k+1}|\boldsymbol{\theta}_k; D_k; C_j) \nabla^T] |_{\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k|k}}$ can be calculated using the finite difference method. Therefore, the conditional evidence in Eq. 46 can be calculated using Eq. 47–49.

the state vector, the structural parameters and their associated uncertainty. This approach allows for the tracking of nonstationary process noise and measurement noise. It estimates the noise parameters recursively for every time step. Then, the outlier-resistant extended Kalman filter (OR-EKF) was presented for robust online outlier detection and structural parametric identification. This method embeds an online outlier detection algorithm into the EKF. It is capable for robust estimation of structural parameters using outlier-contaminated dynamic response data in an online manner. In contrast to other existing robust KF/EKF algorithms, the OR-EKF requires no prior information of the outlier distribution model. Finally, the Bayesian model class selection approach allows for nonparametric online structural identification. This recursive algorithm allows for online model class selection to maintain the optimal balance between the data fitting capability and the robustness to modeling error and measurement noise. Through these recent developments, it is expected that reliable online structural identification using EKF can be achieved.

Summary

This chapter presented several recent advances in online structural identification using the extended Kalman filter (EKF). First, a Bayesian approach was introduced for online identification of the noise parameters. This approach resolves the divergence problem possibly encountered in the conventional EKF due to improper selection of the noise covariance matrices. Furthermore, the presented approach ensures reliable estimation of

Cross-References

- [Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations](#)
- [Model Class Selection for Prediction Error Estimation](#)
- [Nonlinear System Identification: Particle-Based Methods](#)
- [Stochastic Structural Identification from Vibrational and Environmental Data](#)

References

- Beck JL (2010) Bayesian system identification based on probability logic. *Struct Control Health Monit* 17(7):825–847
- Beck JL, Katafygiotis LS (1998) Updating models and their uncertainties. I: Bayesian statistical framework. *J Eng Mech (ASCE)* 124(4):455–461
- Beck JL, Yuen KV (2004) Model selection using response measurements: Bayesian probabilistic approach. *J Eng Mech (ASCE)* 130(2):192–203
- Ching J, Chen YC (2007) Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection and model averaging. *J Eng Mech (ASCE)* 133(7):816–832
- Durovic ZM, Kovacevic BD (1999) Robust estimation with unknown noise statistics. *IEEE Trans Automat Control* 44(6):1292–1296
- Franco G, Betti R, Luş H (2004) Identification of structural systems using an evolutionary strategy. *J Eng Mech (ASCE)* 130(10):1125–1139
- Gull SF (1988) Bayesian inductive inference and maximum entropy. In: Skilling J (ed) *Maximum entropy and Bayesian methods*. Kluwer, Boston, pp 53–74
- Hoi KI, Yuen KV, Mok KM (2009) Prediction of daily averaged PM10 concentrations by statistical time-varying model. *Atmos Environ* 43(16):2579–2581
- Hoi KI, Yuen KV, Mok KM (2011) Iterative probabilistic approach for selection of time-varying model classes. *Procedia Eng* 14:2585–2592. doi:10.1016/j.proeng.2011.07.325
- Jazwinski AH (1970) *Stochastic processes and filtering theory*. Academic, New York
- Jeffreys H (1961) *Theory of probability*, 3rd edn. Clarendon, Oxford
- Kalman RE, Bucy R (1961) New results in linear filtering and prediction theory. *J Basic Eng* 83:95–108
- Katafygiotis LS, Beck JL (1998) Updating models and their uncertainties. II: model identifiability. *J Eng Mech (ASCE)* 124(4):463–467
- Koh CG, Shankar K (2003) Substructural identification method without interface measurement. *J Eng Mech (ASCE)* 129(7):769–776
- Lam HF, Yuen KV, Beck JL (2006) Structural health monitoring via measured Ritz vectors utilizing artificial neural networks. *Comput-Aided Civil Infrastruct Eng* 21(4):232–241
- Mehra R (1970) On the identification of variances and adaptive Kalman filtering. *IEEE Trans Automat Control* 15:175–184
- Mu HQ, Yuen KV (2014) Novel outlier-resistant extended Kalman filter for robust online structural identification. *J Eng Mech (ASCE)*. doi:10.1061/(ASCE)EM.1943-7889.0000810
- Muto M, Beck JL (2008) Bayesian updating and model class selection for hysteretic structural models using stochastic simulation. *J Vib Control* 14(1–2):7–34
- Myers K, Tapley B (1976) Adaptive sequential estimation with unknown noise statistics. *IEEE Trans Automat Control* 21(4):520–523
- Papadimitriou C, Fritzen C-P, Kraemer P, Ntotsios E (2011) Fatigue predictions in entire body of metallic structures from a limited number of vibration sensors using Kalman filtering. *Struct Control Health Monit* 18(5):554–573
- Sorensen SW, Sacks JE (1971) Recursive fading memory filters. *Inform Sci* 3:101–119
- Sorenson HW, Alspach DL (1971) Recursive Bayesian estimation using Gaussian sums. *Automatica* 7(4):465–479
- Worden K (1997) Structural fault detection using a novelty measure. *J Sound Vib* 201(1):85–101
- Worden K, Hensman JJ (2012) Parameter estimation and model selection for a class of hysteretic systems using Bayesian inference. *Mech Systems Signal Process* 32:153–169
- Yan WM, Yuen KV, Yoon GL (2009) Bayesian probabilistic approach for the correlations of compressibility index for marine clays. *J Geotech Geoenviron Eng (ASCE)* 135(12):1932–1940
- Yuen KV (2010a) *Bayesian methods for structural dynamics and civil engineering*. Wiley, New York
- Yuen KV (2010b) Recent developments of Bayesian model class selection and applications in civil engineering. *Struct Safety* 32:338–346
- Yuen KV, Katafygiotis LS (2005) Model updating using response measurements without knowledge of the input spectrum. *Earthq Eng Struct Dyn* 34(2):167–187
- Yuen KV, Katafygiotis LS (2006) Substructure identification and health monitoring using noisy response measurements only. *Comput-Aided Civil Infrastruct Eng* 21(4):280–291
- Yuen KV, Kuok SC (2010a) Ambient interference in long-term monitoring of buildings. *Eng Struct* 32(8):2379–2386
- Yuen KV, Kuok SC (2010b) Modeling of environmental influence in structural health monitoring assessment for reinforced concrete buildings. *Earthq Eng Eng Vib* 9(2):295–306
- Yuen KV, Kuok SC (2011) Bayesian methods for updating dynamic models. *Appl Mech Rev (ASME)* 64(1):010802–1–010802–18
- Yuen KV, Mu HQ (2011) Peak ground acceleration estimation by linear and nonlinear models with reduced order Monte Carlo simulation. *Comput-Aided Civil Infrastruct Eng* 26(1):30–47
- Yuen KV, Mu HQ (2012) A novel probabilistic method for robust parametric identification and outlier detection. *Probab Eng Mech* 30:48–59
- Yuen KV, Beck JL, Katafygiotis LS (2006a) Efficient model updating and monitoring methodology using incomplete modal data without mode matching. *Struct Control Health Monitor* 13(1):91–107
- Yuen KV, Beck JL, Katafygiotis LS (2006b) Unified probabilistic approach for model updating and damage detection. *J Appl Mech (ASME)* 73(4):555–564

- Yuen KV, Hoi KI, Mok KM (2007) Selection of noise parameters for Kalman filter. *Earthq Eng Eng Vib* 6:49–56
- Zhou J, Luecke RH (1994) Estimation of the covariances of the process noise and measurement noise for a linear discrete dynamic system. *Comput Chem Eng* 19(2):187–195

Ambient Vibration Testing of Cultural Heritage Structures

Carmelo Gentile and Antonella Saisi
Department ABC, Politecnico di Milano,
Milan, Italy

Synonyms

Operational modal analysis; Operational modal testing

Introduction

Ambient vibration testing (AVT), long-term dynamic monitoring, and operational modal analysis (OMA, i.e., the identification of modal parameters from ambient vibration data) of Cultural Heritage structures are a rather recent topic, and only a limited number of complete investigations are reported in the literature (Jaishi et al. 2003; Bennati et al. 2005; Ivorra and Pallares 2006; Gentile and Saisi 2007, 2013; Pau and Vestroni 2008; Casarin and Modena 2008; Peña et al. 2010; Ramos et al. 2010; Aras et al. 2011; Oliveira et al. 2012). On the other hand, there is a growing interest on this topic since the preservation of Cultural Heritage is of primary concern in many countries all over the world.

AVT and OMA are especially suitable to historic structures for several reasons: (a) the easy and fully nondestructive way of testing, performed by measuring only the structural response under ambient excitation; (b) the sustainability of testing, which does not interfere with the normal use of the structure and does not induce additional loads rather than those due

to normal conditions (dead loads, wind, micro-tremors); and (c) the multiple-input nature of ambient excitation, ensuring that the response includes the contribution of a certain number of modes. It is indeed true that the response of a historic building to ambient excitation is generally low, but this cannot be considered a prohibitive issue as currently highly sensitive and relatively inexpensive accelerometers are available on the market.

In addition, the popularity of operational modal testing and analysis of civil engineering structures has been favored also by the technological advances (i.e., the availability of data acquisition and storage systems, which are fully computer based) and by the large number of output-only modal identification techniques available in the literature (see, e.g., Magalhães and Cunha 2011).

Furthermore, AVT and OMA seem ideal tools to methodologically complement the investigations currently carried out to assess the structural safety of Cultural Heritage structures. Although general rules, which can be applied to all historic constructions, are very difficult to define, it is generally agreed (see, e.g., Binda et al. 2000) that the first phase of a correct diagnostic approach (i.e., the evaluation of the current health state or performance of the building) involves the collection of all the essential information on the geometry of the building, its evolution from the origin to the present state, the construction technologies, the characteristics of masonry texture, the mechanical characterization of the materials, and the evaluation of their state of preservation.

The above documentary and experimental information – as it is suggested also in current Italian Guidelines for the seismic risk mitigation of Cultural Heritage (DPCM 2011) – provide a 1st-level diagnosis, highlighting the overall state of preservation, the presence of local defects and vulnerabilities, as well as the need of possible repair interventions, especially when local issues are detected; in this case, the intervention design should be addressed by simplified local models (see, e.g., Giuffrè 1993; DPCM 2011).

In principle, the previously collected knowledge of the building should be synthesized in

a finite element (FE) model of the structure. The FE model, in turn, should provide a 2nd-level diagnosis since it could be used for evaluating the structural safety under service loads, predicting the performance under exceptional loads (such as earthquakes), and simulating the effects of structural modifications or repair interventions.

However, FE modeling of historic structures is characterized by well-known issues:

- (a) The correlation between the results of local tests (which indeed provide the mechanical characterization of the materials) and quantitative parameters to build up global structural capacity models is still an open issue (Binda et al. 2000).
- (b) The structural model of a historic structure, even when all the collected information is accurately represented, continues to involve significant uncertainties, e.g., in the material properties (and their distribution) as well as in the boundary conditions. This aspect is especially critical for complex historic buildings evolved in different phases.
- (c) FE models are often used, even in refined nonlinear analyses, without experimental validation, and only occasionally the model validation is roughly performed by using few available local data (such as the stress level evaluated in few points through flat-jack tests).

Within this context, one possible key role of AVT and OMA is to provide effective and accurate validation of the FE model prior to its use in numerical analysis, as demonstrated in different studies on temples and monuments (Jaishi et al. 2003; Pau and Vestroni 2008), churches (Casarin and Modena 2008; Ramos et al. 2010; Gattulli et al. 2013), ancient palaces (Aras et al. 2011), towers, and minarets (Bennati et al. 2005; Ivorra and Pallares 2006; Gentile and Saisi 2007, 2013; Peña et al. 2010; Ramos et al. 2010; Oliveira et al. 2012).

In some cases, AVT can possibly help also to limit the number of on-site and laboratory tests, which are time-consuming and cost-ineffective.

Other possible applications of ambient vibration-based modal analysis in the field of historic structures include periodic or continuous monitoring in order to evaluate the effects of repair interventions or to perform dynamics-based damage assessment (Ramos et al. 2010).

The present entry, after a review of some output-only modal identification techniques, presents the application of ambient vibration-based modal and structural identification to a historic masonry tower. Subsequently, the role of AVT and OMA in the preservation of Cultural Heritage structures is exemplified in a further application (i.e., the long-term dynamic monitoring of a tower) aimed at (1) evaluating the effects of structural modifications, (2) assessing the influence of environmental effects on natural frequencies, (3) identifying the evolution of damage mechanisms.

Modal Identification from Ambient Vibration Data

As previously pointed out, a large number of output-only modal identification techniques are available in the literature, ranging from the simple *peak picking* technique (PP, Bendat and Piersol 1993) to the more advanced *frequency domain decomposition* (FDD, Brincker et al. 2000) and *stochastic subspace identification* (SSI, van Overschee and De Moor 1996). In this section, the PP, FDD, and SSI techniques are briefly described.

Peak Picking and Frequency Domain Decomposition

The PP and FDD techniques work in frequency domain and are based on the evaluation of the spectral matrix $\mathbf{G}_{yy}(f)$ of the recorded responses:

$$\mathbf{G}_{yy}(f) = \mathbf{E} [\mathbf{Y}(f) \mathbf{Y}^H(f)] \quad (1)$$

where the vector $\mathbf{Y}(f)$ collects the responses in the frequency domain, the superscript H denotes the Hermitian transpose operation (i.e., complex conjugate matrix transpose), and $\mathbf{E}$ denotes expected value. The diagonal terms of the matrix $\mathbf{G}_{yy}(f)$ are the (real valued) auto-spectral

densities (ASD), while the other terms are the (complex) cross-spectral densities (CSD). The spectral matrix $\mathbf{G}_{yy}(f)$ is generally computed by using the modified periodogram method (Welch 1967). According to this approach, an average is made over each recorded signal, divided into M frames of $2n$ samples, where windowing and overlapping are applied.

The principle of the two techniques is easiest illustrated by recalling that any response vector $\mathbf{y}(t)$ can be expressed in modal coordinates $\mathbf{q}(t)$ as

$$\mathbf{y}(t) = \boldsymbol{\phi}_1 q_1(t) + \boldsymbol{\phi}_2 q_2(t) + \dots = \boldsymbol{\Phi} \mathbf{q}(t) \quad (2)$$

where $\boldsymbol{\phi}_i$ represents the i -th mode shape vector and $\boldsymbol{\Phi}$ is the mode shape matrix. Hence, the correlation matrix $\mathbf{C}_{yy}(\tau)$ (see, e.g., Bendat and Piersol 1993) of the responses

$$\mathbf{C}_{yy}(\tau) = \mathbb{E}[\mathbf{y}(t + \tau) \mathbf{y}^T(t)] \quad (3)$$

becomes

$$\mathbf{C}_{yy}(\tau) = \boldsymbol{\Phi} \mathbb{E}[\mathbf{q}(t + \tau) \mathbf{q}^T(t)] \boldsymbol{\Phi}^T = \boldsymbol{\Phi} \mathbf{C}_{qq}(\tau) \boldsymbol{\Phi}^T \quad (4)$$

On the other hand, the spectral matrix may be defined as the Fourier transform of the correlation matrix:

$$\mathbf{G}_{yy}(f) = \mathcal{F}[\mathbf{C}_{yy}(\tau)] = \boldsymbol{\Phi} \mathbf{G}_{qq}(f) \boldsymbol{\Phi}^H \quad (5)$$

where $\mathbf{G}_{qq}(f)$ is the spectral matrix of the modal coordinates. It is worth underlining that since the modal coordinates are uncorrelated, the matrix $\mathbf{G}_{qq}(f)$ is diagonal.

The more traditional approach to estimate the modal parameters of a structure is often called *peak picking* method after its key step: the identification of the resonant frequencies as the peaks of ASDs and CSDs. In fact, for a lightly damped structure subjected to a white-noise random excitation, both ASDs and CSDs reach a local maximum at the frequencies corresponding to the system normal modes (Bendat and Piersol 1993); furthermore, if the assumption is introduced of modes having well-separated frequencies, it can be shown (see, e.g., Peeters 2000) that

the spectral matrix can be approximated, in the neighborhood of a resonant frequency f_k , as

$$\mathbf{G}_{yy}(f_k) \approx \alpha_k \boldsymbol{\phi}_k \boldsymbol{\phi}_k^H \quad (6)$$

where α_k depends on the damping ratio, the natural frequency, the modal participation factor, and the excitation spectra. Equation 6 highlights that (a) each row or column of the spectral matrix at a natural frequency f_k can be considered as an estimate of the mode shape $\boldsymbol{\phi}_k$ at that frequency and (b) the square root of the diagonal terms of the spectral matrix at a natural frequency f_k can be considered as an estimate of the mode shape $\boldsymbol{\phi}_k$ at that frequency.

The PP technique leads to reliable results provided that the basic assumptions of low damping and well-separated modes are satisfied; drawbacks of the method are related to the difficulties in identifying closely spaced modes and damping ratios.

Some refinements or variants of the classic PP technique have been proposed in the literature. For example, Felber (1993) suggested obtaining a “global picture” of the eigenfrequencies by evaluating an appropriate function, called averaged normalized power spectral density (ANPSD). More specifically, only the diagonal elements of the spectral matrix $\mathbf{G}_{yy}(f)$ are considered, normalized, and averaged so that a unique function is obtained, summarizing the information on the “modal” peaks contained in all the ASDs.

The FDD technique (Brincker et al. 2000) involves the singular value decomposition (SVD) of the spectral matrix at each frequency and the inspection of the curves representing the singular values, in order to identify the resonant frequencies and to estimate the corresponding mode shape using the information contained in the singular vectors of the SVD.

The SVD of the spectral matrix at each frequency is given by

$$\mathbf{G}_{yy}(f) = \mathbf{U}(f) \boldsymbol{\Sigma}(f) \mathbf{U}^H(f) \quad (7)$$

where the diagonal matrix $\boldsymbol{\Sigma}$ collects the real positive singular values in descending order and $\mathbf{U}$ is a complex matrix containing the singular

vectors as columns. The SVD is used for estimating the rank of $\mathbf{G}_{yy}$ at each frequency with the number of nonzero singular values being equal to the rank; if only one mode is important at a given frequency f_k , as it has to be expected for well-separated modes, the spectral matrix can be approximated by a rank-one matrix:

$$\mathbf{G}_{yy}(f) \approx \mathbf{u}_1(f)\sigma_1(f)\mathbf{u}_1^H(f) \quad (8)$$

By comparing Eq. 5 with Eq. 7, it is evident that (if the mode shapes are orthogonal) Eq. 5 is an SVD of the spectral matrix. Furthermore, the comparison of Eqs. 6 and 8 clearly reveals that the first singular vector $\mathbf{u}_1(f)$ is an estimate of the mode shape. Since the first singular value $\sigma_1(f)$ at each frequency represents the strength of the dominating vibration mode at that frequency, the first singular function can be suitably used as a modal indication function (yielding the resonant frequencies as local maxima). In addition, the successive singular values contain either noise or modes close to a strong dominating one.

The FDD is a rather simple procedure that represents an improvement of the PP because:

1. The SVD is an effective method for separating signal space from noise space, and the evaluation of mode shapes is automatic and significantly easier than in the PP.
2. The FDD technique is able to detect closely spaced modes. In such instances, more than one singular value will reach a maximum in the neighborhood of a given frequency, and every singular vector corresponding to a nonzero singular value is a mode shape estimate. The latter, however, is valid in the strict sense for orthogonal modes.
3. The damping ratios can be identified through the refinement of the FDD technique, namely, the *enhanced frequency domain decomposition* (EFDD, Brincker et al. 2001). The EFDD technique is based on the fact that the first singular value in the neighborhood of a resonant peak is the ASD of a modal coordinate. Hence, moving the partially identified ASD of the modal coordinate back in the time

domain by inverse FFT yields a free decaying time domain function, which represents the autocorrelation function of the modal coordinate. The natural frequency and the related damping ratio are thus simply found by estimating crossing times and employing the logarithmic decrement method.

Stochastic Subspace Identification

The SSI technique lies in the class of time domain methods and is based on the discrete-time stochastic state-space form of the dynamics of a linear time-invariant system under unknown excitation.

The continuous-time state-space equation of motion of a linear time-invariant system can be written as

$$\dot{x}(t) = \mathbf{A}_c x(t) + \mathbf{B}_c f(t) \quad (9)$$

$$\mathbf{A}_c = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C}_1 \end{bmatrix} \quad \mathbf{B}_c = \begin{bmatrix} \mathbf{0} \\ \mathbf{M}^{-1} \end{bmatrix} \quad (10)$$

where

- $x(t) = [\mathbf{u}(t)^T \quad \dot{\mathbf{u}}(t)^T]^T \in \mathfrak{R}^{2N \times 2N}$ is the state vector of the process, containing the displacement $\mathbf{u}(t)$ and the velocity $\dot{\mathbf{u}}(t)$ vectors.
- $\mathbf{A}_c \in \mathfrak{R}^{2N \times 2N}$ is the continuous-time state matrix, which is related to the matrices of mass $\mathbf{M}$, damping $\mathbf{C}_1$, and stiffness $\mathbf{K}$.
- $f(t) \in \mathfrak{R}^N$ is the load vector and $\mathbf{B}_c \in \mathfrak{R}^{2N \times N}$ is the system control influence coefficient matrix.
- The subscript c denotes continuous time.

It should be noticed that the eigenvalues Λ_c and eigenvectors Ψ of the state-space matrix $\mathbf{A}_c$ (solving the eigenvalue problem $\mathbf{A}_c \Psi = \Psi \Lambda_c$) contain the eigenvalues and the eigenvectors of the original second-order system $\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}_1\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = f(t)$.

In dynamic testing, only a subset L of the N responses are measured; hence, the vector of measured outputs $\mathbf{y}(t) \in \mathfrak{R}^L$ can be expressed as

$$\mathbf{y}(t) = \mathbf{C}_a \ddot{\mathbf{u}}(t) + \mathbf{C}_v \dot{\mathbf{u}}(t) + \mathbf{C}_d \mathbf{u}(t) \quad (11)$$

where $\mathbf{C}_a$, $\mathbf{C}_v$, and $\mathbf{C}_d$ are the output location matrices for accelerations, velocity, and displacements, respectively. These matrices contain a lot of zeros and a few unit entries since they are formulated in order to assign the measured degrees of freedom. The vector $\mathbf{y}(t)$ can be written as

$$\mathbf{y}(t) = \mathbf{C}_c \mathbf{x}(t) + \mathbf{D}_c \mathbf{f}(t) \quad (12)$$

$$\begin{aligned} \mathbf{C}_c &= [\mathbf{C}_d - \mathbf{C}_a \mathbf{M}^{-1} \mathbf{K}, \quad \mathbf{C}_v - \mathbf{C}_a \mathbf{M}^{-1} \mathbf{C}_1] \\ \mathbf{D}_c &= \mathbf{C}_a \mathbf{M}^{-1} \end{aligned} \quad (13)$$

By combining the state Eq. 9 and the observation Eq. 12, the classical continuous-time state-space model is found:

$$\begin{cases} \dot{\mathbf{x}}(t) = \mathbf{A}_c \mathbf{x}(t) + \mathbf{B}_c \mathbf{f}(t) \\ \mathbf{y}(t) = \mathbf{C}_c \mathbf{x}(t) + \mathbf{D}_c \mathbf{f}(t) \end{cases} \quad (14)$$

Since real measurements are taken at discrete-time instants and in order to fit model Eq. 14 to the measurements, this model needs to be converted into discrete time. Hence, assuming a constant sampling period Δt and that the input is piecewise constant over the sampling period, the continuous-time equations (Eq. 14) are discretized and solved at all discrete-time instants $t_k = k\Delta t$, obtaining the discrete-time state-space model:

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{A} \mathbf{x}_k + \mathbf{B} \mathbf{f}_k \\ \mathbf{y}_k = \mathbf{C} \mathbf{x}_k + \mathbf{D} \mathbf{f}_k \end{cases} \quad (15)$$

where $\mathbf{x}_k$ is the discrete-time state vector (containing displacements and velocities describing the state of the system at time instant $t_k = k\Delta t$), $\mathbf{f}_k$ and $\mathbf{y}_k$ are the sampled input and output vectors, $\mathbf{A}$ is the discrete state matrix (dependent on the mass, stiffness, and damping properties of the structure), $\mathbf{B}$ is the discrete input matrix, $\mathbf{C}$ is the discrete output matrix (which maps the state vector into the measured output), and $\mathbf{D}$ is the direct transmission matrix. The matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are related to their continuous-time counterparts Eq. 10 and Eq. 13 by the following:

$$\begin{aligned} \mathbf{A} &= e^{\mathbf{A}_c \Delta t} & \mathbf{B} &= (\mathbf{A} - \mathbf{I}) \mathbf{A}_c^{-1} \mathbf{B}_c \\ \mathbf{C} &= \mathbf{C}_c & \mathbf{E} &= \mathbf{E}_c \end{aligned} \quad (16)$$

It is finally observed that the deterministic model (Eq. 15) is not capable of exactly describing real measurement data; consequently, Eq. 15 needs to be modified to account for both the process noise $\mathbf{w}_k \in \mathfrak{R}^{2N}$ due to disturbances and modeling inaccuracies and the measurement noise $\mathbf{v}_k \in \mathfrak{R}^L$ due to sensor inaccuracy:

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{A} \mathbf{x}_k + \mathbf{B} \mathbf{f}_k + \mathbf{w}_k \\ \mathbf{y}_k = \mathbf{C} \mathbf{x}_k + \mathbf{D} \mathbf{f}_k + \mathbf{v}_k \end{cases} \quad (17)$$

Furthermore, both $\mathbf{w}_k$ and $\mathbf{v}_k$ are unmeasurable vectors, assumed to be zero mean, white, and with covariance matrices:

$$\mathbb{E} \left[\begin{pmatrix} \mathbf{w}_p \\ \mathbf{v}_p \end{pmatrix} \begin{pmatrix} \mathbf{w}_p^T & \mathbf{v}_p^T \end{pmatrix} \right] = \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^T & \mathbf{R} \end{bmatrix} \delta_{pq} \quad (18)$$

It is very important to remark that an AVT provides information only on the vibration responses of a structure excited by unmeasured inputs. Consequently, it is impossible to distinguish the input term $\mathbf{f}_k$ from the noise terms $\mathbf{w}_k$ and $\mathbf{v}_k$ in Eq. 17. This results in the following discrete-time stochastic state-space model:

$$\begin{cases} \mathbf{x}_{k+1} = \mathbf{A} \mathbf{x}_k + \mathbf{w}_k \\ \mathbf{y}_k = \mathbf{C} \mathbf{x}_k + \mathbf{v}_k \end{cases} \quad (19)$$

where the input is implicitly modeled by the noise terms and the white noise assumption of these terms turns out to be essential: if the white noise assumption is violated because the input contains some dominant frequency components, these frequency components cannot be distinguished from the eigenfrequencies of the system (of the state matrix $\mathbf{A}$).

The key step of the SSI techniques is the estimation of the state-space matrices $\mathbf{A}$ and $\mathbf{C}$ from the measured output $\mathbf{y}_k$. The estimation of $\mathbf{A}$ and $\mathbf{C}$ can be performed by using different algorithms. The well-known data-driven SSI algorithms are based on linear algebra theorems

(van Overschee and De Moor 1996) demonstrating that the state-space matrices can be calculated from the knowledge of the block Hankel matrix of the measurements, defined as

$$\mathbf{H}_i = \begin{pmatrix} \mathbf{y}_0 & \mathbf{y}_1 & \cdots & \mathbf{y}_{j-1} \\ \mathbf{y}_1 & \mathbf{y}_2 & \cdots & \mathbf{y}_j \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{i-1} & \mathbf{y}_i & \cdots & \mathbf{y}_{i+j-2} \\ \hline \mathbf{y}_i & \mathbf{y}_{i+1} & \cdots & \mathbf{y}_{i+j-1} \\ \mathbf{y}_{i+1} & \mathbf{y}_{i+2} & \cdots & \mathbf{y}_{i+j} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{2i-1} & \mathbf{y}_{2i} & \cdots & \mathbf{y}_{2i+j-2} \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_p \\ \mathbf{Y}_f \end{pmatrix} \quad (20)$$

where $2i$ and j are user-defined quantities, representing the number of output block rows and the number of columns of matrix $\mathbf{H}_i$, respectively.

In Eq. 2, the block Hankel matrix of the measurements is subdivided into two sub-matrices, named as $\mathbf{Y}_p$ and $\mathbf{Y}_f$, which are usually referred to as past and future output block matrices. The orthogonal projection of the row space of $\mathbf{Y}_f$ onto the row space of $\mathbf{Y}_p$ can be directly calculated yielding to matrix $\mathbf{P}_i$, which is known as projection matrix. The observability matrix $\mathbf{O}_i$ of the system – which, in turn, allows to compute $\mathbf{A}$ and $\mathbf{C}$ – can be estimated from the SVD of matrix $\mathbf{W}_1 \mathbf{P}_i \mathbf{W}_2$, where $\mathbf{W}_1$ and $\mathbf{W}_2$ are convenient weight matrices. The available SSI-data algorithms essentially differ for the expressions of the weight matrices (van Overschee and De Moor 1996).

After estimating the model matrices, the modal parameters f_i , ζ_i , and ϕ_i of the structural system are calculated from

$$\mathbf{A} = \mathbf{\Psi} \mathbf{\Lambda} \mathbf{\Psi}^{-1} \quad (21a)$$

$$\mathbf{\Lambda} = \text{diag}(\lambda_i) \quad \lambda_i^c = \frac{\ln \lambda_i}{\Delta t} \quad (21b)$$

$$f_i = \frac{|\lambda_i^c|}{2\pi} \quad \zeta_i = \frac{\text{real}(\lambda_i^c)}{|\lambda_i^c|} \quad (21c)$$

$$\mathbf{\Phi} = [\phi_1, \phi_2, \dots, \phi_i, \dots] = \mathbf{C} \mathbf{\Psi} \quad (21d)$$

It is further noticed that the order $2N$ of the model should equal, in principle, twice the

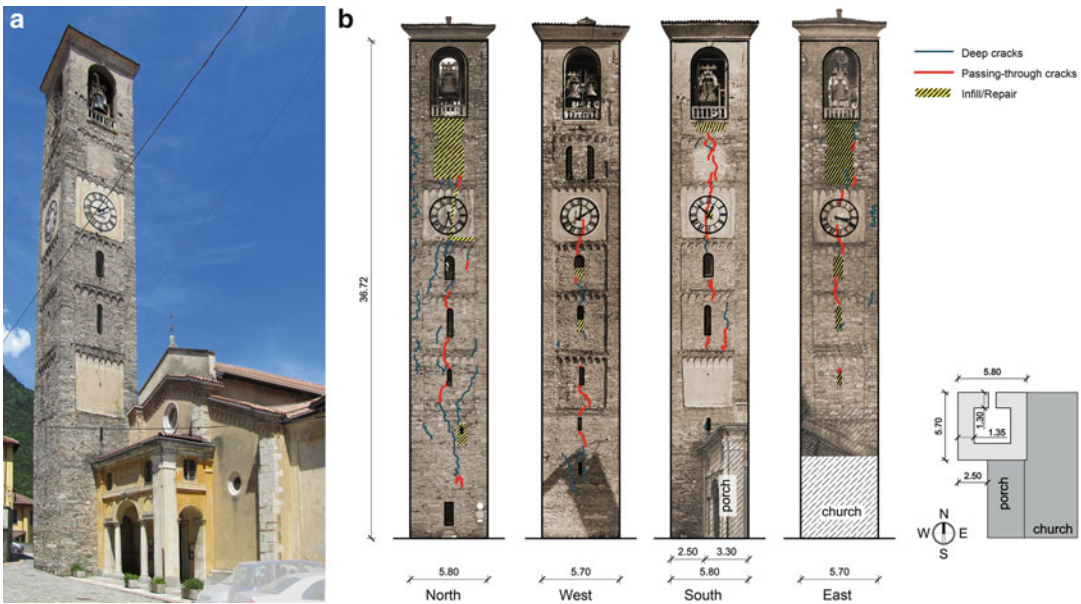
number of modes that are needed to accurately describe the structural response. However, with the purpose of detecting weakly excited modes, it is often necessary to consider more than $2N$ values, which are what would be physically sufficient. On the other hand, over-modeling leads to the appearance of spurious modes associated with the noise content of the measurements. A common practice for identifying the physical modes is based on the creation of the stabilization diagrams, where the modal parameters obtained for increasing model orders are collectively represented. A physical mode is conceivably identified when consistent frequencies, damping, and mode shapes (classified as stable poles) are obtained for models of increasing order.

Ambient Vibration Testing and Structural Identification of a Historic Masonry Tower

Description of the Tower

The first investigated historic building is a bell tower (Fig. 1), about 37.0 m high and built in stonework masonry (Gentile and Saisi 2013). The tower – located in the small town of Arcisate (northern Italy) – has a square cross section, with sides of 5.8 m, and is connected to the church *Chiesa Collegiata* on the East side and partly on the South side. The church, dedicated to St. Vittore, dates back to the fifteenth century and replaced a more ancient church, built in the fourth century and modified in the eleventh century. Probably the tower foundation dates back to the late Roman age, as well.

The first historic document concerning the tower goes back to the sixteenth century and reports St. Carlo Borromeo's request of access modification. Seven orders of floors are present, with five of them being defined by masonry offsets at the corners and by corresponding sequences of small hanging arches marking the floor levels; the last two orders were probably added in the eighteenth century to host the bell trusses. The wall thickness decreases



Ambient Vibration Testing of Cultural Heritage Structures, Fig. 1 (a) View of the bell tower of *Chiesa Collegiata* (Arcisate, Varese); (b) crack patterns on the fronts of the tower (dimensions in m)

progressively along the height, from 135 cm at the ground level up to 65 cm at the top level.

Although extensive visual inspections and few sonic tests generally indicate that the stone masonry is relatively compact and of fairly good execution, the masonry texture appears locally often highly disordered and characterized by the local presence of vertical joints.

The crack pattern (Fig. 1b) has been accurately surveyed also by using an aerial platform. The tower exhibits long vertical cracks on every side, most of them cutting the entire wall thickness and passing through the keystones of the arch window openings. These cracks are mainly distributed between the second/third order of the tower. Many superficial cracks are also diffused, particularly on the North and West fronts, which are not adjacent to the church.

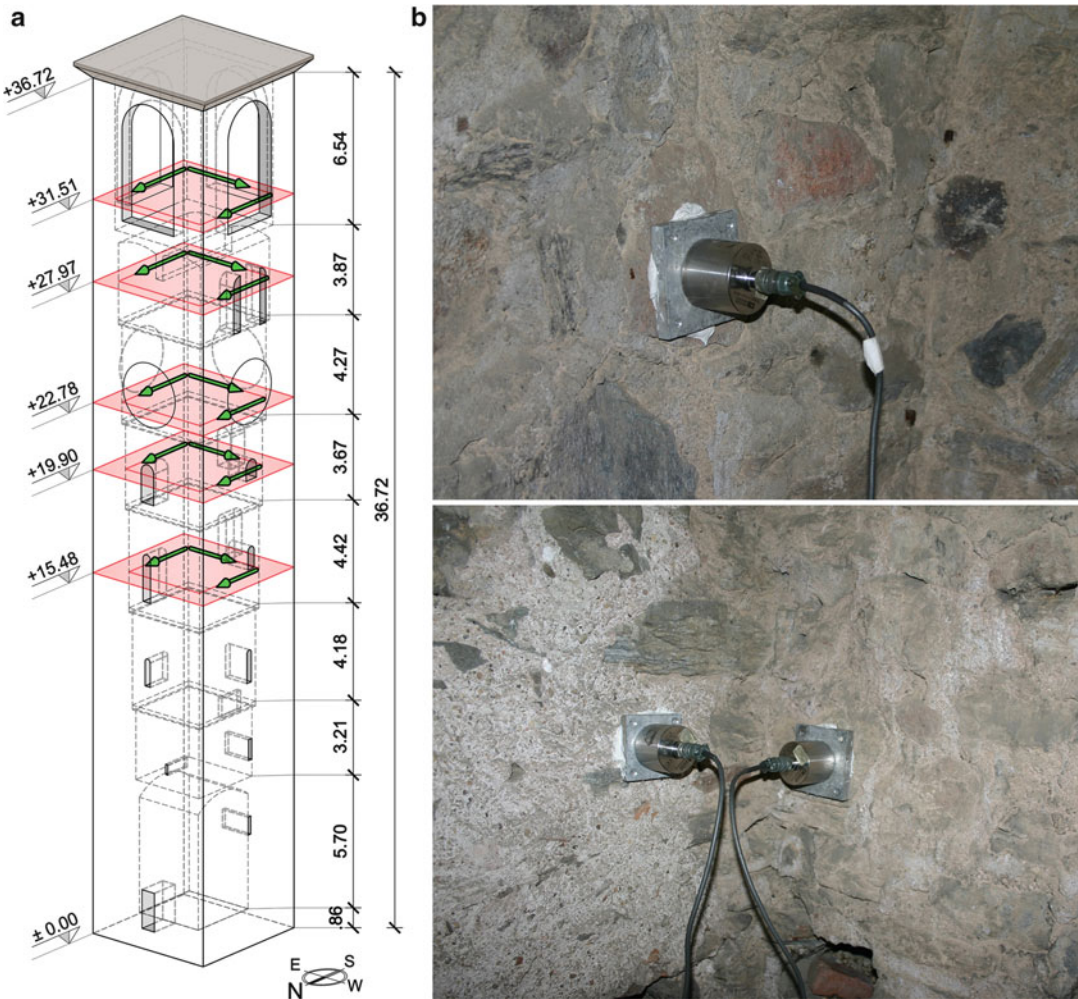
The visual inspection highlighted that the upper part of the tower, beneath the belfry level, could be probably considered the most vulnerable, due to the widespread mortar erosion and the infilled openings (Fig. 1b), mainly on the East and North fronts; in addition, the infillings are often not properly linked to the surrounding load-bearing masonry.

Ambient Vibration Testing Procedures and Modal Identification

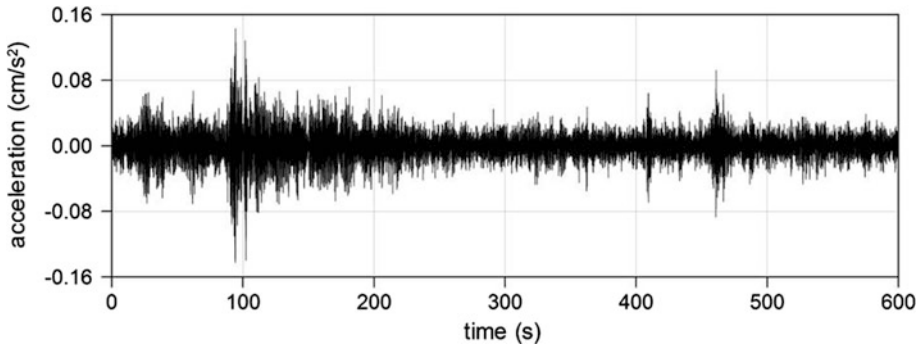
The dynamic tests were carried out in June 2007, using a 16-channel data acquisition system and WR 731A piezoelectric accelerometers (10 V/g sensitivity and 0.5 g peak acceleration). Each accelerometer was connected with a short cable (1 m) to a WR P31 power unit/amplifier, providing the constant current needed to power the accelerometer's internal amplifier, signal amplification, and selective filtering.

The response of the tower was measured in 15 selected points, belonging to 5 different cross sections along the height of the building, according to the sensor layout illustrated in Fig. 2a. Figure 2b shows the mounting of accelerometers in the instrumented cross section at level +22.78 m.

The acceleration time histories induced by ambient excitation were recorded for 3,600 s at a sampling frequency of 200 Hz. A sample of the acceleration time histories recorded during the test in the upper part of the tower is shown in Fig. 3: it should be noticed that very low level of ambient excitation was present during the tests, with the maximum recorded acceleration being always lower than 0.4 cm/s^2 .

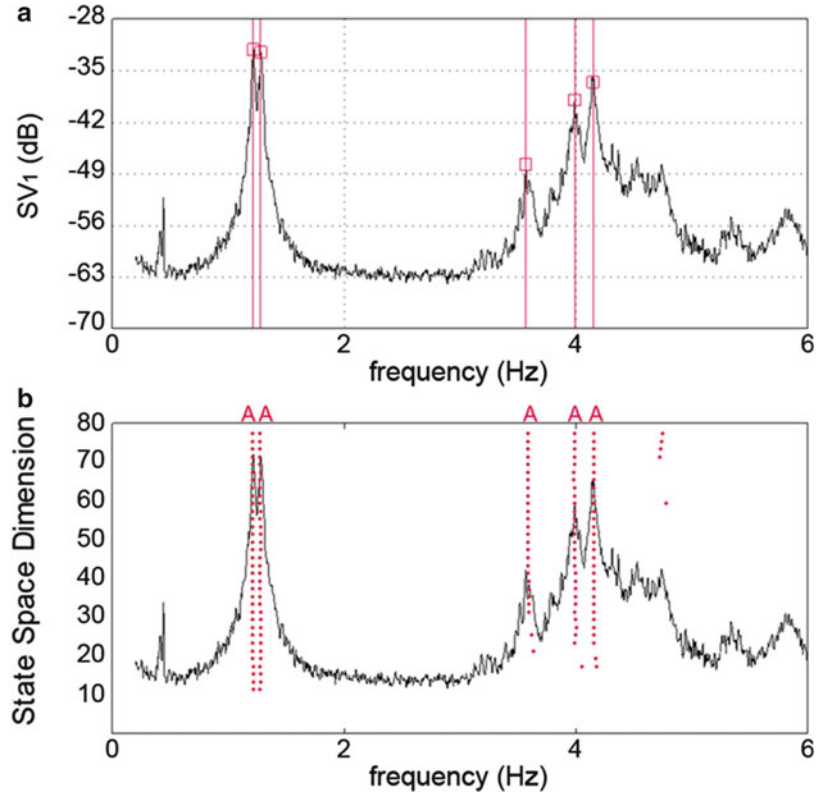


Ambient Vibration Testing of Cultural Heritage Structures, Fig. 2 (a) Sensor layout adopted in the dynamic tests (dimensions in m), (b) mounting of the accelerometers at level +22.78 m



Ambient Vibration Testing of Cultural Heritage Structures, Fig. 3 Typical acceleration time series measured at the upper instrumented level

Ambient Vibration Testing of Cultural Heritage Structures, Fig. 4 (a) First singular value (SV) curve and identification of natural frequencies (*FDD*), (b) stabilization diagram and automatic (*A*) identification of natural frequencies (*SSI*)



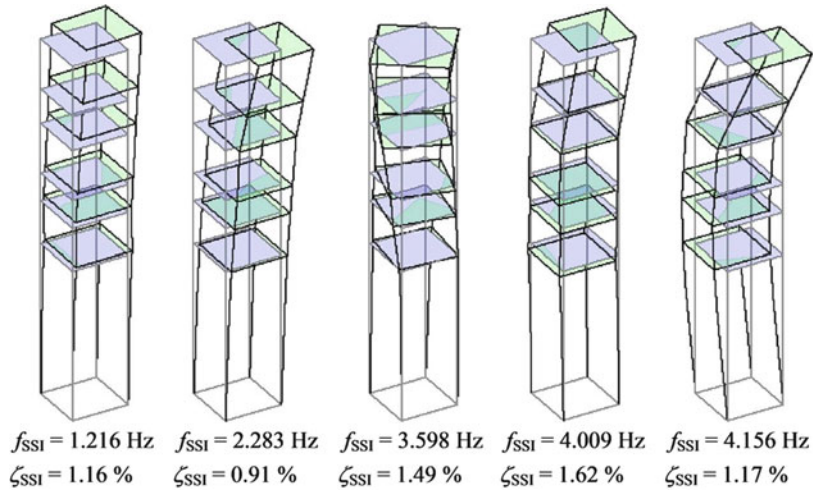
The modal identification was performed using time windows of 3,600 s, in order to comply with the widely agreed recommendation of using an appropriate duration of the acquired time window (ranging between 1,000 and 2,000 times the fundamental period of the structure; see, e.g., Cantieni 2005) to obtain accurate estimates of the modal parameters from OMA techniques. In fact, as already pointed out in section “[Modal Identification from Ambient Vibration Data](#),” OMA methods assume that the excitation input is a zero mean Gaussian white noise, and this assumption is as closely verified as the length of the acquired time window is longer.

The extraction of modal parameters from ambient vibration data was carried out using the FDD and the data-driven SSI techniques available in the commercial software ARTEMIS (SVS 2012). Notwithstanding the very low level of ambient response (Fig. 3) that existed during the tests, the application of both techniques allowed to identify five vibration modes in the frequency range of 0–6 Hz.

The results of OMA in terms of natural frequencies can be summarized through the plots of Fig. 4a, b, showing the first singular value (SV) of the spectral matrix and the stabilization diagrams obtained by applying the FDD and the SSI technique, respectively. The inspection of Fig. 4a highlights that the FDD technique provides a clear indication of the tower modes through well-defined local maxima in the first SV; similarly, Fig. 4b shows that the alignments of the stable poles in the stabilization diagram of the SSI method provide a clear indication of these modes, as well. Furthermore, Fig. 4a, b shows the correspondence of the natural frequency estimates between the two techniques, with the resonant peaks of Fig. 4a being placed practically at the same frequencies of the alignments of stable poles in Fig. 4b.

As it had to be expected, the identified modes can be classified as bending and torsion. Figure 5 shows the identified mode shapes (SSI technique): dominant bending (B) modes were identified at 1.22 (B1), 1.28 (B2), 4.01 (B3), and

Ambient Vibration Testing of Cultural Heritage Structures, Fig. 5 Vibration modes identified from ambient vibration data (SSI)



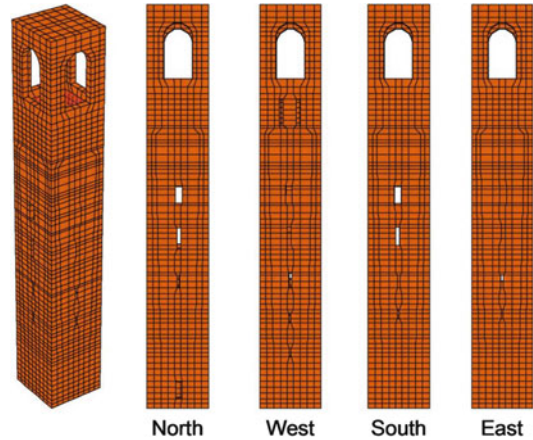
4.16 Hz (B4), while only one torsion mode (T1) was identified at 3.60 Hz. It is observed that the dominant bending modes of the tower involve flexure practically along the diagonals.

FE Modeling and Model Tuning

A 3D structural model (Fig. 6) of the tower was developed (using the FE program Straus7), based on the available geometric survey.

The tower was modeled using 8-node brick elements. A relatively large number of finite elements have been used in the model, so that a regular distribution of masses could be obtained, and all the geometric variations and openings in the load-bearing walls could be reasonably represented. The model consists of 3,475 solid elements with 17,052 active degrees of freedom.

Since the geometry of the tower was surveyed on site and accurately described in the model, the main uncertainties are related to the characteristics of the material and boundary conditions. In order to reduce the number of uncertainties in the model calibration, the following initial assumptions were introduced: (a) homogeneous distribution of the masonry elastic properties; (b) the weight per unit volume of the masonry and Poisson's ratio of the masonry were assumed as 17.0 kN/m³ and 0.15, respectively; and (c) the tower footing was considered as fixed since the soil-structure interaction is hardly

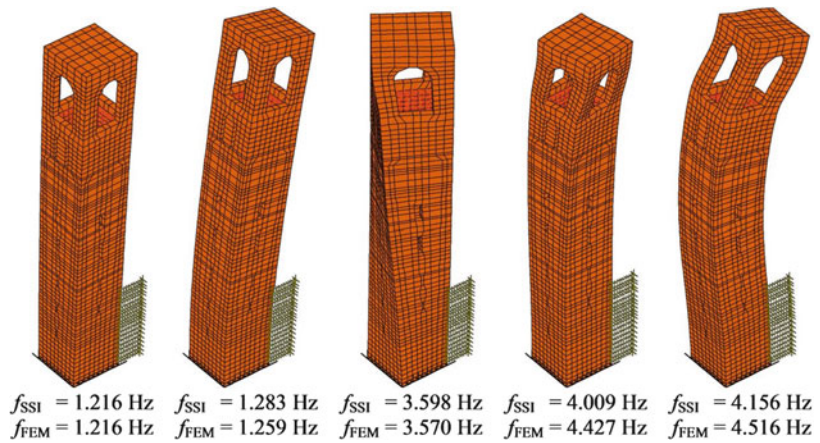


Ambient Vibration Testing of Cultural Heritage Structures, Fig. 6 Finite element model of the tower

involved at the low level of ambient vibrations that existed during the tests.

Under the above assumptions, a sensitivity analysis (see, e.g., Maia and Silva 1997) pointed out that sensitive parameters were (1) the average Young's modulus E of stone masonry, since all natural frequencies are very sensitive to its variation; (2) the ratio $\alpha = G/E$, implying the assumption of orthotropic material and significantly affecting the natural frequency of the torsion mode; (3) the connection between the tower and the neighboring building (represented by linear nodal springs), which highly affects the bending mode shapes.

Ambient Vibration Testing of Cultural Heritage Structures,
Fig. 7 Vibration modes of the initial model



Hence, three steps of manual tuning were carried out to establish a base FE model. In the base model, the following assumptions were adopted:

1. An orthotropic elastic behavior was assumed for the stone masonry, with the average characteristics of the material being $E = 3.00$ GPa and $G_{13} = G_{23} = 0.45$ GPa (corresponding to $\alpha = G_{13}/E = G_{23}/E = 0.15$). It is worth noting that the assumed Young's modulus was in good agreement with the results of the tests performed to characterize the masonry and reported in (Binda et al. 2012): more specifically, the average sonic velocity (P waves) varied between 1,620 and 2,240 m/s², and the values measured by double flat-jack tests at the base of the tower were slightly larger than 3.00 GPa.
2. The effects of the connection between the tower and the church were accounted for through a series of linear (nodal) springs of constant k . After tuning the parameter k in a preselected interval (1×10^4 kN/m $\leq k \leq 10 \times 10^4$ kN/m), the value $k = 4 \times 10^4$ kN/m was assumed since it tends to minimize the average difference between the measured and predicted modal frequencies.

Figure 7 illustrates the dynamic characteristics of the base model and highlights that the mode shapes are fully consistent with the experimental results (Fig. 5). Furthermore, the correspondence with the identified frequencies is fairly good for

the first three modes. The discrepancy is conceivably related to the simplified distribution of the model elastic properties.

As suggested by the mode shapes, the higher bending modes depend on the elastic characteristics of the masonry in the upper part of the tower. In addition, as pointed out in section “[Description of the Tower](#),” the upper region is characterized by a more evident mortar joint erosion and changes of the masonry texture including wide infilled openings on the East and North fronts, beneath the belfry level.

Hence, the distribution of Young's modulus was updated, and the tower was divided in two regions, with the masonry Young's modulus being assumed as constant within each zone. The two regions, denoted as I and II, correspond to the lower five levels of the building (E_I , $h \leq 26.0$ m, including a large number of passing-through cracks) and the upper part (E_{II} , $h > 26.0$ m, including infilled/repared areas and the belfry), respectively. The possible set of updating parameters includes E_I , E_{II} , α , and k . It should be noticed that assuming only one parameter α means that the ratio between the shear moduli in the two regions of the tower equals to the ratio between Young's moduli.

Subsequently, the optimal values of the updating structural parameters were determined. Among the different classes of procedures available in the literature (see, e.g., Friswell and Mottershead 1995), it was decided to evaluate the updating parameters by minimizing the

Ambient Vibration Testing of Cultural Heritage Structures, Table 1 Structural parameters for FE model identification

Structural parameter	Base value	Upper value	Base value	Optimal value
E_I (height ≤ 26 m) (GPa)	2.0	4.0	3.0	2.97
E_{II} (height > 26 m) (GPa)	1.0	3.0	2.0	1.60
α	0.110	0.190	0.150	0.172
k (kN/m)	1.0×10^4	10.0×10^4	4.0×10^4	8.30×10^4

difference between theoretical and experimental natural frequencies, through the simple procedure proposed in (Douglas and Reid 1982). According to this approach, the dependence of the natural frequencies of the model on the unknown structural parameters X_k ($k = I, 2, \dots, N$) is approximated around the current values of X_k , by the following:

$$f_i^*(X_1, X_2, \dots, X_N) = \sum_{k=1}^N [A_{i,k}X_k + B_{i,k}X_k^2] + C_i \quad (22)$$

where f_i^* represents the approximation of the i -th frequency of the FE model. Once the set of approximating functions Eq. 22 has been established, the structural parameters of the model are evaluated by a least-square minimization of the difference between each f_i^* and its experimental counterpart f_i^{EXP} :

$$J = \sum_{i=1}^M w_i \varepsilon_i^2 \quad (23a)$$

$$\varepsilon_i = f_i^{\text{EXP}} - f_i^*(X_1, X_2, \dots, X_N) \quad (23b)$$

where w_i is a weight constant. However, Eq. 22 represents a reasonable approximation in a range, around the “base” value of the structural parameters X_k^B , limited by lower X_k^L and upper values X_k^U ($k = I, 2, \dots, N$); thus, the coefficients A_{ik} , B_{ik} , C_i are dependent on both the base value of the structural parameters and the range in which these parameters can vary. The coefficients A_{ik} , B_{ik} , C_i are readily evaluated from $(2N + I)$ finite element analyses (Douglas and Reid 1982), each with a different choice of the parameters: the first choice of the parameters corresponds to the base

values; then each parameter is varied, one at time, from the base value to upper and lower limit, respectively.

It is further noticed that, in principle, the quadratic approximation Eq. 22 is as better as the base values are closer to the solution; hence, the accuracy and stability of the optimal estimates should be carefully checked either by the complete correlation with the experimental data or by repeating the procedure with new base values.

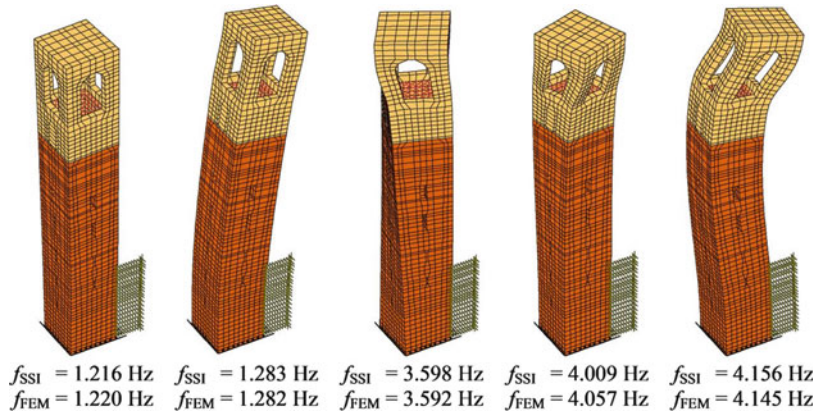
Table 1 summarizes the optimal estimates of the structural parameters, the base values, and the assumed lower and upper limits. By examining the optimal values in Table 1, the following comments can be made: (a) the optimal values of the elastic parameters in the lower region ($E_I = 2.97$ GPa, $G_{13}^I = G_{23}^I = \alpha \times E_I = 0.172 \times 2.97 \cong 0.51$ GPa) are very similar to those of the base model; (b) the optimal estimates of the elastic parameters in the upper region ($E_{II} = 1.60$ GPa, $G_{13}^{II} = G_{23}^{II} = \alpha \times E_{II} = 0.172 \times 1.60 \cong 0.28$ GPa) turned out to be lower than the ones in the lower part, reflecting the observed state of preservation in that region; and (c) the stiffness of the springs becomes larger than in the base model to better fit the ratio between the natural frequencies of the first two modes.

Figure 8 shows the mode shapes of the updated model, corresponding to the experimental ones (Fig. 5), and the correlation with the measured modal behavior. It should be noticed that the updated model represents an excellent approximation of the real structure, with the maximum relative error between predicted and measured modal frequencies being larger than 1 % only for mode B3.

Furthermore, also the correlation between mode shapes – estimated via the MAC (Allemang and Brown 1983) – is very good for

Ambient Vibration Testing of Cultural Heritage Structures,

Fig. 8 Vibration modes of the optimal (updated) model



Ambient Vibration Testing of Cultural Heritage Structures, Fig. 9 Views of the *Gabbia Tower* (Mantua, Italy) from South and East

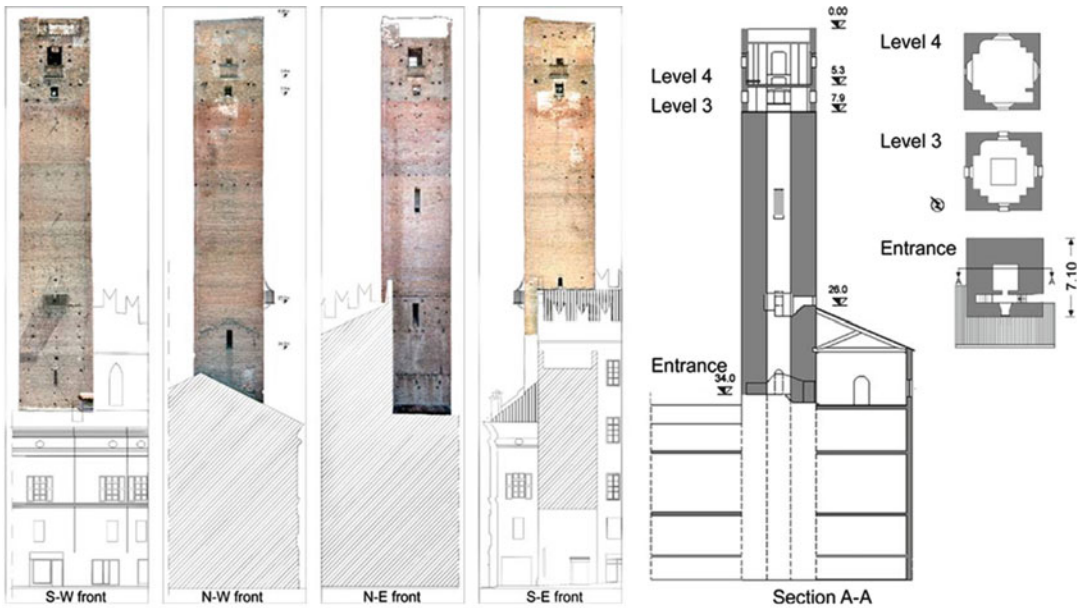
the first two modes (with the MAC being larger than 0.96); for the higher modes, the MAC is in the range 0.711–0.845 so that appreciable average differences are detected and again related to the simplified distribution of the model elastic properties, which were held constant for large regions of the tower.

Ambient Vibration Testing and Dynamic Monitoring of a Historic Masonry Tower

After the Italian earthquakes of May 2012, an extensive research program has been performed to assess the state of preservation of the tallest historic tower in Mantua, Italy. The investigated tower (Fig. 9), about 54.0 m high and dating back to the twelfth century, is known as the *Gabbia Tower*.

Visual inspection of all main-bearing walls (Saisi et al. 2013) clearly indicated that the upper part of the tower is characterized by the presence of several discontinuities due to the historic evolution of the building, local lack of connection, and extensive masonry decay. The poor state of preservation of the same region was confirmed by the observed dynamic characteristics (Saisi et al. 2013), and one local mode involving the upper part of the tower was clearly identified by applying different output-only techniques to the response data collected for more than 24 h on the historic building.

These results clearly highlighted the critical situation of the upper part of the tower, pointing out the need for structural interventions to be carried out. With this motivation, and in order to allow for better indoor inspection of the tower-bearing walls, a metal scaffolding and a light



Ambient Vibration Testing of Cultural Heritage Structures, Fig. 10 Fronts and section of the *Gabbia Tower* (dimensions in m)

wooden roof have been installed inside the tower. Hence, a second dynamic test was performed – aimed at checking the possible effects of scaffolding and wooden roof on the modal characteristics of the structure – and a simple permanent dynamic monitoring system (including three highly sensitive accelerometers and one temperature sensor) was installed in the tower, with structural health monitoring and seismic early warning purposes.

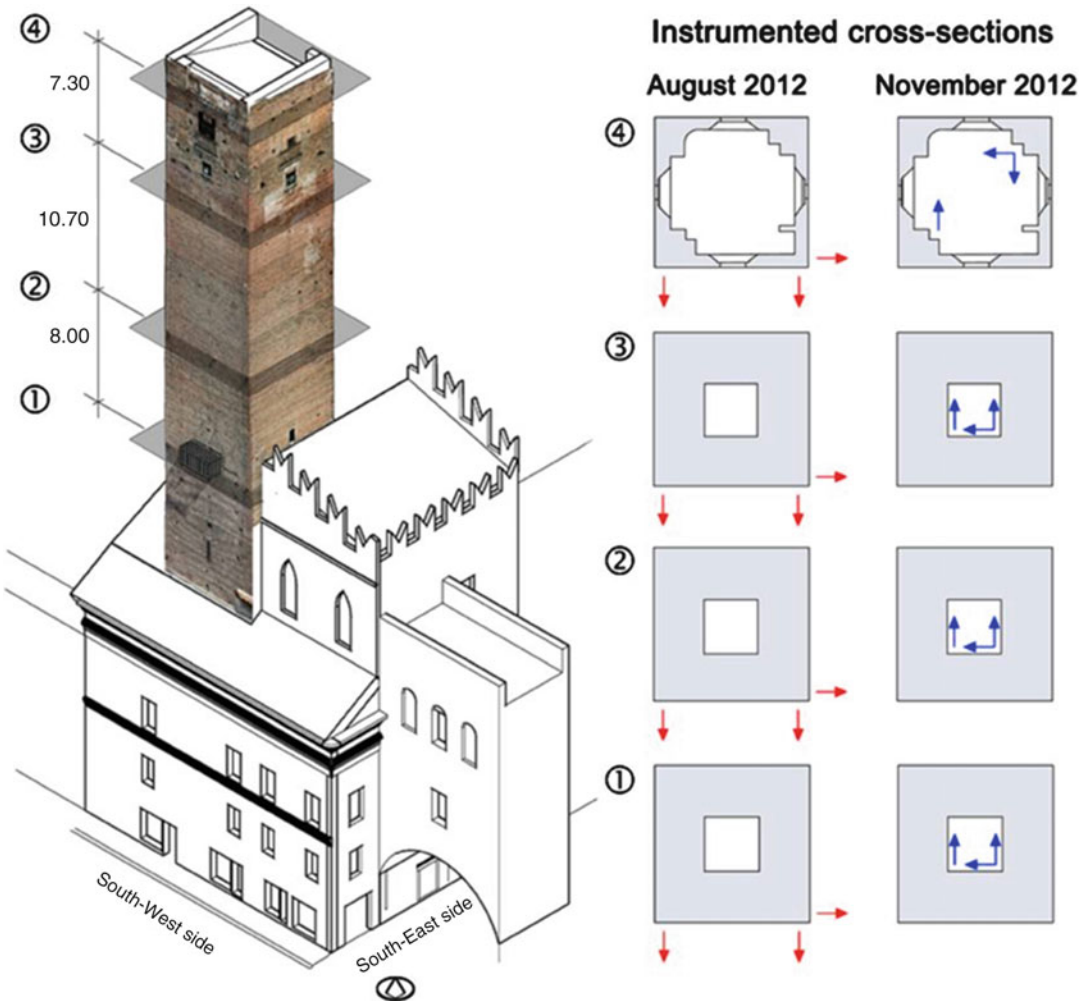
Description of the Tower and On-Site Inspections

The *Gabbia Tower*, about 54.0 m high, is built in solid masonry bricks and has an almost square base; the load-bearing walls are about 2.4 m thick up to the upper levels (Fig. 10) where the thickness of the masonry cross section decreases to about 0.7 m. The top part of the building has a two-level lodge, which hosted in the nineteenth century the observation and telegraph post. A wooden staircase reached the lodge, but it is no more practicable since several years due to the lack of maintenance. The inner access to the tower was reestablished recently (October 2012) through provisional scaffoldings.

The original layout of the surrounding structures is unknown. At present, the tower is part of an important palace, evolved since the thirteenth century, complicating the geometry of the structure and the mutual links.

Few historic documents are available on the past interventions on the tower, but the observation of the masonry texture reveals passing-through discontinuities in the upper region, which are conceivably related to the tower evolution. Traces of past structures are visible on all fronts, and the presence of merlon-shaped discontinuities suggests modifications and further adding at the top of the tower. Moreover, at about 8.0 m from the top, a clear change of the brick surface workmanship (the bricks of the lower part are superficially scratched) could reveal a first addition; in the same region concentrated changes of the masonry texture reveal local repair.

An accurate on-site survey of all fronts of the tower was firstly performed using a mobile platform. The visual inspection, omitting the upper part of the tower (i.e., a portion about 8.0 m high), did not reveal evident structural damage but only superficial decay of the materials (mainly mortar



Ambient Vibration Testing of Cultural Heritage Structures, Fig. 11 Instrumented cross sections and sensors layout during the dynamic tests performed on August and November 2012 (dimensions in m)

joint erosion, due to the natural aging and the lack of maintenance). Subsequent pulse sonic tests, double flat jacks, and laboratory tests on sampled mortars and bricks confirm the soundness and the compactness of the masonry until the height of about 46.0 m. On the contrary, significant damages were observed in the upper 8.0 m of the tower; those damages are related to the abovementioned detachment of the several construction phases and worsened by the natural decay. More specifically, critical areas are the infillings between the merlons, supported only by few courses of thin masonry due to the unusual layout of the scaffolding holes.

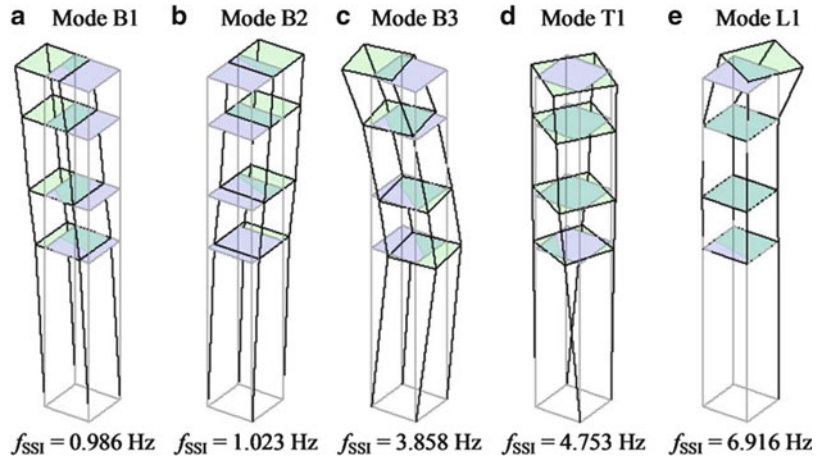
Ambient Vibration Testing Procedures and Modal Identification

Two AVTs were conducted on the tower: between 31/07/2012 and 02/08/2012 and on 27/11/2012. It is worth recalling that the second test was performed after the installation of a metallic scaffolding and a wooden roof inside the tower in order to check the possible effects of those additions on the dynamic characteristics of the structure.

The response of the tower was measured in 12 selected points, belonging to 4 different cross sections along the height of the building, according to the sensor layout shown in Fig. 11.

Ambient Vibration Testing of Cultural Heritage Structures,

Fig. 12 Vibration modes generally identified during the first AVT (SSI, 31/07/2012, 21:00–22:00)



It is worth noting that the same cross sections were instrumented in the two tests, but in the second survey the accelerometers were installed on the inner side of load-bearing walls.

In both tests, the excitation was provided only by wind and micro-tremors. In the first test, acceleration data were acquired for 28 h (between 16:00 and 23:00 of 31/07/2012 and from 9:00 of 01/08/2012 to 6:00 of 02/08/2012), and a second acquisition system was used to measure the temperature in three different points of the tower: on the S-W front both indoor and outdoor temperature were measured, whereas only the outdoor temperature was measured on the S-E front. It is worth mentioning that the changes of outdoor temperature were very significant, ranging between 25 °C and 55 °C, whereas slight variations were measured by the indoor sensor (29–30 °C), due to the high thermal inertia of the load-bearing walls.

The modal identification was performed using 3,600 s long time windows and applying the data-driven SSI algorithm available in the Artemis software (SVS 2012).

Figure 12 shows typical results of the first AVT, in terms of natural frequencies and mode shapes. It should be noticed that (a) two closely spaced modes were identified around 1.0 Hz and these modes (Fig. 12a, b) are dominant bending (B) and involve flexure in the two main planes of the tower, respectively; (b) the third mode (Fig. 12c) involves dominant bending in the

N-E/S-W plane with slight components also in the orthogonal N-W/S-E plane; (c) just one torsion mode (T) was identified (Fig. 12d); and (d) the last identified mode is local (L) and only involves deflections of the upper portion of the tower (Fig. 12e).

The presence of a local vibration mode provides further evidence of the structural effect of the change in the masonry quality and morphology observed on top of the tower during the visual inspection. On the other hand, both visual inspection and operational modal analysis confirm the concerns about the seismic vulnerability of the building and explain the fall of small masonry pieces from the upper part of the tower, reported during the earthquake of May 29, 2012.

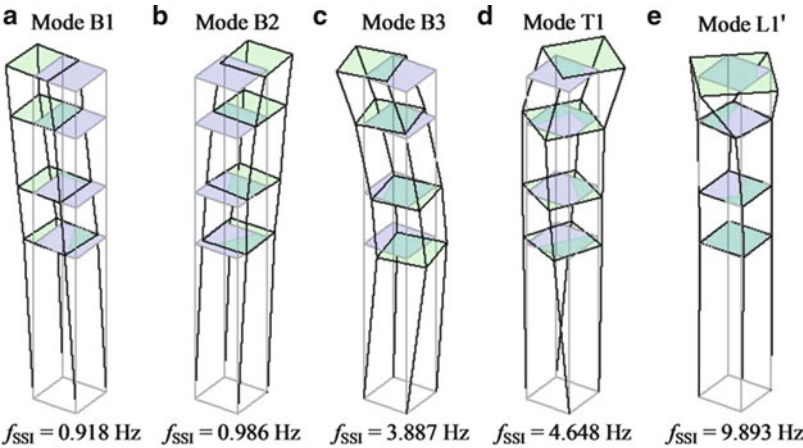
Statistics of the modal frequencies identified between 31/07/2012 and 02/08/2012 are summarized in columns (2)–(5) of Table 2 through the mean value, the standard deviation, and the extreme values of each modal frequency. It should be noticed that the natural frequencies of all modes exhibit slight but clear variation, with the standard deviation ranging between 0.011 Hz (mode B2) and 0.037 Hz (mode L1). The correlation analysis performed to investigate the possible relationships between natural frequencies and temperature (Saisi et al. 2013) clearly indicated that the natural frequencies of the global modes B1–B3 and T1 increase with increased temperature. This behavior, observed also in previous measurements on masonry structures

Ambient Vibration Testing of Cultural Heritage Structures, Table 2 Natural frequencies identified (SSI) in the ambient vibration tests of *Gabbia Tower*

Mode	31/07/2012 – 02/08/2012				27/11/2012
	f_{ave} (Hz)	σ_f (Hz)	f_{min} (Hz)	f_{max} (Hz)	f (Hz)
(1)	(2)	(3)	(4)	(5)	(6)
1 (B1)	0.981	0.018	0.957	1.014	0.918
2 (B2)	1.026	0.011	1.006	1.052	0.986
3 (B3)	3.891	0.025	3.857	3.936	3.887
4 (T1)	4.763	0.022	4.714	4.802	4.648
5 (L1)	6.925	0.037	6.849	6.987	–
6 (L1')	–	–	–	–	9.893

B bending mode, *T* torsion mode, *L* local mode

Ambient Vibration Testing of Cultural Heritage Structures, Fig. 13 Vibration modes identified during the second AVT (SSI, 27/11/2012)



(Ramos et al. 2010; Gentile et al. 2012), can be explained through the closure of superficial cracks, minor masonry discontinuities, or mortar gaps induced by the thermal expansion of materials. Hence, the temporary “compacting” of the materials induces a temporary increase of stiffness and modal frequencies, as well.

As previously stated, the possible effects of scaffolding and wooden roof on the dynamic characteristics of the tower were investigated in the second AVT, performed on November 27, 2012 with the outdoor temperature being almost constant (10–11 °C). The results of this investigation in terms of identified natural frequencies’ main and mode shapes are shown in Fig. 13 and can be summarized as follows:

1. Beyond the difference in terms of natural frequency (that are conceivably related to the temperature effects), the mode shapes of bending

modes B1–B3 did not exhibit significant changes (see Figs. 12a–c and 13a–c). Hence, the metallic scaffolding and the wooden roof practically do not affect those modes.

2. On the contrary, the mode shape T1 (Fig. 13d) now involves both torsion and bending. The identified frequency did not change appreciably with respect to the first dynamic survey, but the mode shape looks significantly different. The torsion component is still dominant in the lower portion of the structure, while the upper part is characterized by dominant bending with significant components along the two main planes of the tower. In other words, after the installation of the wooden roof, mode shape T1 becomes mixture of previous modes T1 (Fig. 12d, lower part of the structure) and L1 (Fig. 12e, upper part of the structure). Furthermore, the previous mode L1 was no more detected.

3. The previous local mode L1 (Fig. 12e) has been “replaced” by another local mode, with higher frequency of 9.89 Hz, involving torsion of the upper part of the tower (Fig. 13e).

As a further comment, it seems that especially the wooden roof, even if very light, affects the dynamic characteristics of the upper part of the building: the roof acts as a mass directly connected to a highly inhomogeneous and weaker portion of the building, so that a possible decrease of the natural frequency of the previous local mode is generated.

Long-Term Dynamic Monitoring and Typical Results

Few weeks after the execution of the second AVT, a simple dynamic monitoring system was installed in the tower. The system is composed by a 4-channel data acquisition system (24-bit resolution, 102 dB dynamic range, and anti-aliasing filters) with three piezoelectric accelerometers (WR model 731A, 10 V/g sensitivity, and ± 0.50 g peak). The response of the tower is measured in three points, belonging to the cross section at the crowning level of the tower. Furthermore, a temperature sensor is installed on the S-W front, measuring the outdoor temperature.

The digitized data are transmitted to an industrial PC on site. A binary file, containing three acceleration time series (sampled at 200 Hz) and the temperature data, is created every hour, stored on the local PC, and transmitted to Politecnico di Milano for subsequent data processing.

The continuous dynamic monitoring system has been active since December 2012. The data files received from the monitoring system are managed by a software developed in LabVIEW and including the following tasks: (a) creation of a database with the original data (in compact format) for later developments; (b) preliminary preprocessing (i.e., de-trending, automatic recognition and extraction of possible seismic events, creation of one dataset per hour); (c) statistical analysis of data, including the evaluation of averaged acceleration amplitudes and temperature trends; (d) low-pass filtering and decimation of the each dataset; and (e) creation of a second

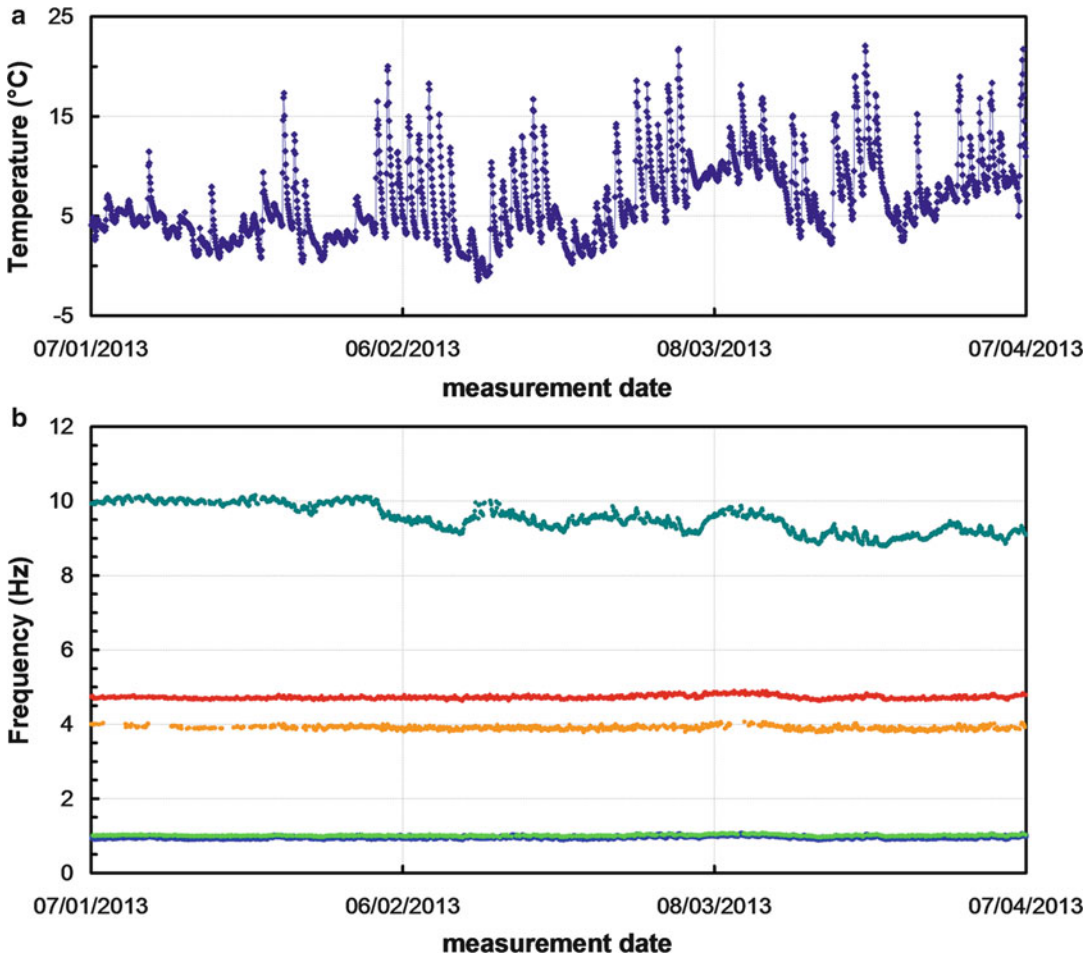
database, with essential data records, to be used in the modal identification phase. The results of modal identification herein presented were obtained applying the SSI technique.

Figure 14a presents the evolution of the outdoor temperature on the S-W front during the period from 07/01/2013 to 06/04/2013 and shows that the temperature changed between -2 °C and 25 °C with significant daily variations in sunny days.

The identification of the modal frequencies from the datasets collected during the same period provided the frequency tracking shown in Fig. 14b. The inspection of Fig. 14b firstly suggests that the slight fluctuation of the natural frequencies of global modes follows the temperature variation. In order to better explore the temperature effect on the modal frequencies, Fig. 15 presents the first four natural frequencies of the tower plotted with respect to temperature, along with linear best fit lines. The plots in Fig. 15 confirm what already observed in the first dynamic survey: the natural frequencies of the global modes tend to increase with increased temperature almost linearly, as a consequence of the temporary increase of the local stiffness due to the thermal expansion of materials.

The time evolution of the natural frequency of the upper mode, i.e., the local mode L1' (Fig. 13e), deserves some concern because the trend of this modal frequency is very different from the others (Fig. 14b). More specifically, the modal frequency exhibits more significant fluctuations and clearly decreases in time, from an initial value of about 10.0 Hz (07/01/2013) to a final value of about 9.0 Hz at the end of the analyzed period (06/04/2013).

A better inspection of Fig. 14b reveals that two clear drops of the modal frequency took place: (a) between 03/02/2013 and 04/02/2013 and (b) between 14/03/2013 and 15/03/2013. These drops divide the analyzed time period in three parts that are also easily identified by plotting the modal frequency versus the measured outdoor temperature, as shown in Fig. 16. The temperature-frequency plot of Fig. 16 highlights that the clouds of temperature-frequency points corresponding to each of the three different periods are characterized by similar slope of the best fit



Ambient Vibration Testing of Cultural Heritage Structures, Fig. 14 Time evolution of (a) the outdoor temperature measured on the S-W front, (b) the natural frequencies identified with the SSI technique

line, whereas the average frequency value significantly decreases. This behavior suggests the quick progress of a damage mechanism, conceivably related to the worsening of the connection between the wooden roof and the masonry walls, and confirms – once more – the poor structural condition and the high seismic vulnerability of the upper part of the tower, highlighting the urgent need for preservation actions to be performed.

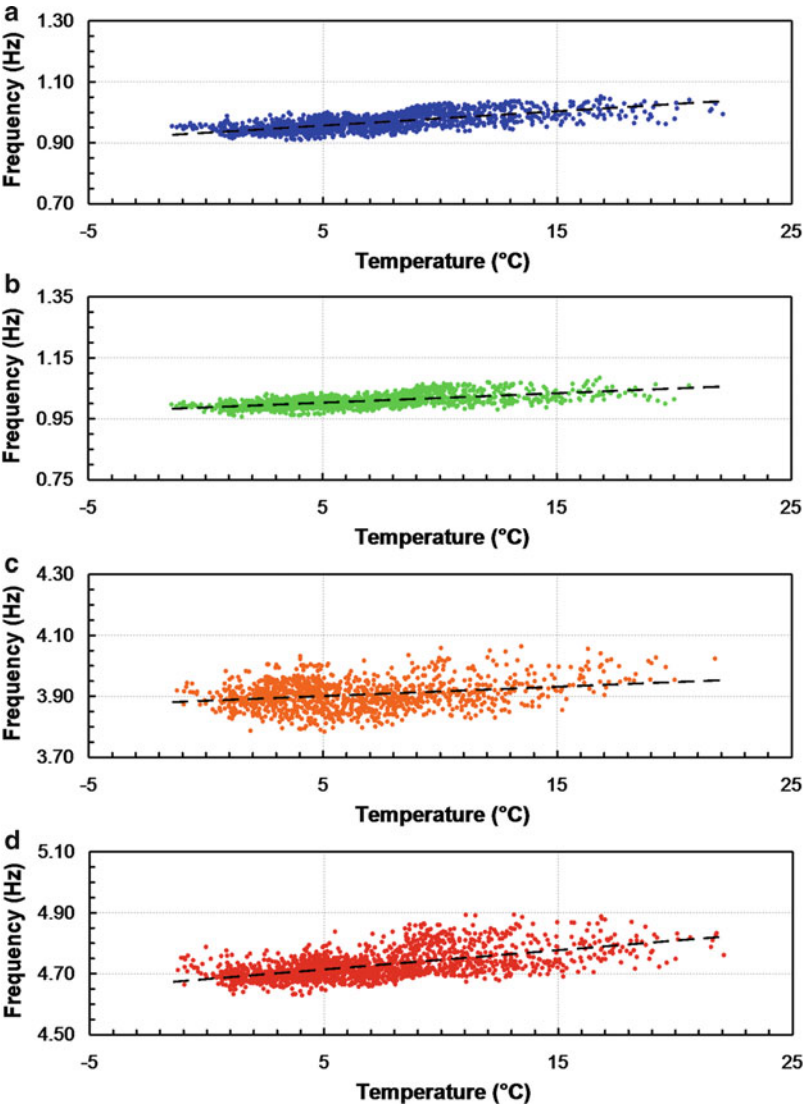
Summary

This entry focuses on the application of ambient vibration testing (AVT) and operational modal analysis (OMA, i.e., the identification of natural

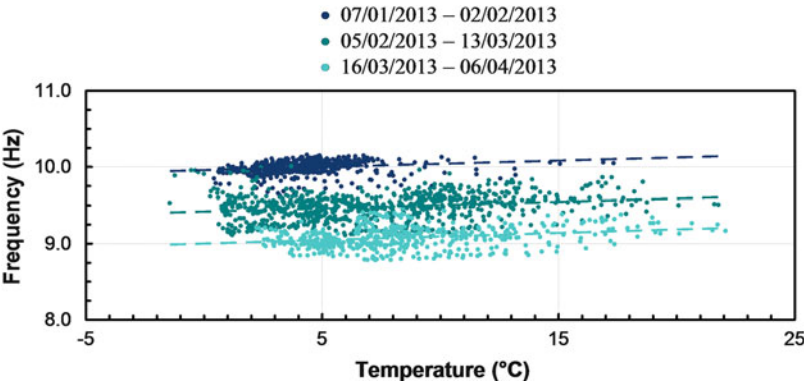
frequencies, mode shapes, and damping ratios from ambient vibration data) to Cultural Heritage structures. Although the topic is rather recent, it is emerging as a subject of great importance in the modern approach to preservation of historic structures.

AVT is a fully nondestructive test, especially suitable to historic structures because it is performed by just measuring the response in operational conditions. In turn, the knowledge of global parameters of the building, such as the modal parameters, provides essential information to validate the numerical models currently used to quantitatively estimate the structural safety of Cultural Heritage structures or to design repair interventions.

Ambient Vibration Testing of Cultural Heritage Structures, Fig. 15 Correlation between the identified modal frequencies and the outdoor temperature (S-W front): (a) mode B1, (b) mode B2, (c) mode B3, (d) mode T1



Ambient Vibration Testing of Cultural Heritage Structures, Fig. 16 Natural frequency of local mode L1' plotted with respect to the outdoor temperature



Furthermore, in some cases, AVT can possibly help also to limit the number of on-site and laboratory tests of materials, which are time-consuming and cost-ineffective. Other possible applications of ambient vibration-based modal analysis in the field of historic structures are for evaluation of the effects of repair interventions or in order to perform dynamics-based damage assessment.

In the first part of this entry, some robust techniques to perform OMA are reviewed, namely, the classical *peak picking* method and the more recent *frequency domain decomposition* and data-driven *stochastic subspace identification*.

Subsequently, the application of these techniques is exemplified with reference to two historic towers. In both cases, notwithstanding the very low level of ambient vibrations that existed during the tests, AVT and OMA have proved to be effective tools for identifying the dynamic characteristics of key vibration modes, provided that appropriate and very sensitive acquisition chain (capable of capturing the “interesting” dynamics embedded in the noise) is used in the tests.

In the first case study (the bell tower of the church *Chiesa Collegiata* in Arcisate, Varese), a rational vibration-based methodology, developed for the calibration of the numerical model, has been discussed as well. The presented methodology – involving systematic manual tuning, sensitivity analysis, and a simple system identification algorithm – provided a linear elastic model of the tower, representing an excellent approximation of the structure in its present condition (i.e., the calibrated model summarizes all the collected documentary and geometric and experimental information and exhibits very good agreement between predicted and measured modal parameters).

In the second case study, typical results from long-term monitoring of the ambient vibration data collected on the *Gabbia Tower* in Mantua, Italy (twelfth century), are summarized. The results clearly demonstrate that the modal identification is capable of highlighting (a) the effects of slight structural modifications; (b) the impact

of temperature on the natural frequencies (i.e., the modal frequencies increase with increased temperature); and (c) the quick progress of a damage mechanism, involving the upper part of the tower, which is clearly identified through the remarkable fluctuations and the significant decrease (about 10 % in 3 months) of the natural frequency corresponding to a local mode.

Cross-References

- [Bayesian Operational Modal Analysis](#)
- [Masonry Structures: Overview](#)
- [Operational Modal Analysis in Civil Engineering: An Overview](#)
- [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)
- [Seismic Vulnerability Assessment: Masonry Structures](#)
- [Stochastic Structural Identification from Vibrational and Environmental Data](#)

References

- Allemang RJ, Brown DL (1983) Correlation coefficient for modal vector analysis. In: Proceedings of the 1st international modal analysis conference (IMAC-I), Orlando, FL, USA
- Aras F, Krstevska L, Altay G, Tashkov L (2011) Experimental and numerical modal analyses of a historical masonry palace. *Constr Build Mater* 25(1):81–91
- Bendat JS, Piersol AG (1993) Engineering applications of correlation and spectral analysis. Wiley Interscience, New York
- Bennati S, Nardini L, Salvatore W (2005) Dynamic behavior of a medieval masonry bell tower. II: measurement and modeling of the tower motion. *J Struct Eng ASCE* 131(11):1656–1664
- Binda L, Saisi A, Tiraboschi C (2000) Investigation procedures for the diagnosis of historic masonries. *Constr Build Mater* 14(4):199–233
- Binda L, Condoleo P, Tiraboschi C, Rigamonti P (2012) On-site investigation and crack monitoring of an ancient bell-tower. In: Proceedings of the 14th international conference on structural faults & repair (SF&R-2012), Edinburgh, Scotland
- Brincker R, Zhang LM, Andersen P (2000) Modal identification from ambient responses using frequency domain decomposition. In: Proceedings of the 18th international modal analysis conference (IMAC-XVIII), San Antonio, TX, USA

- Brincker R, Ventura CE, Andersen P (2001) Damping estimation by frequency domain decomposition. In: Proceedings of the 19th international modal analysis conference (IMAC-XIX), Orlando, FL, USA
- Cantieni R (2005) Experimental methods used in system identification of civil engineering structures. In: Proceedings of the 1st international operational modal analysis conference (IOMAC-2005), Copenhagen, Denmark
- Casarin F, Modena C (2008) Seismic assessment of complex historical buildings: application to Reggio Emilia Cathedral, Italy. *Int J Archit Herit* 2(3):304–327
- Douglas BM, Reid WH (1982) Dynamic tests and system identification of bridges. *J Struct Div ASCE* 108(10):295–312
- DPCM (2011) Valutazione e riduzione del rischio sismico del patrimonio culturale con riferimento alle NTC 14 gennaio 2008. G.U.R.I., Rome
- Felber AJ (1993) Development of a hybrid bridge evaluation system. PhD thesis, University of British Columbia
- Friswell M, Mottershead JE (1995) Finite element model updating in structural dynamics. Kluwer, Boston/London/Dordrecht
- Gattulli V, Antonacci E, Vestroni F (2013) Field observations and failure analysis of the Basilica S. Maria di Collemaggio after the 2009 L'Aquila earthquake. *Eng Fail Anal* 34(X):715–734
- Gentile C, Saisi A (2007) Ambient vibration testing of historic masonry towers for structural identification and damage assessment. *Constr Build Mater* 21(6):1311–1321
- Gentile C, Saisi A (2013) Operational modal testing of historic structures at different levels of excitation. *Constr Build Mater* 48:1273–1285
- Gentile C, Saisi A, Cabboi A (2012) Dynamic monitoring of a masonry tower. In: Proceedings of the 8th international conference on structural analysis of historical constructions (SAHC-2012), Wroclaw, Poland
- Giuffrè A (1993) Sicurezza e conservazione di centri storici in area sismica: il caso Ortigia. Laterza, Bari
- Ivorra S, Pallares FJ (2006) Dynamic investigation on a masonry bell tower. *Eng Struct* 25(5):660–667
- Jaishi B, Ren WX, Zong ZH, Maskey PN (2003) Dynamic and seismic performance of old multi-tiered temples in Nepal. *Eng Struct* 25(14):1827–1839
- Magalhães F, Cunha A (2011) Explaining operational modal analysis with data from an arch bridge. *Mech Syst Signal Process* 25(5):1431–1450
- Maia NMM, Silva JMM (1997) Theoretical and experimental modal analysis. Research Studies Press, Baldock, Hertfordshire, England
- Oliveira CS, Çaktı E, Stengel D, Branco M (2012) Minaret behavior under earthquake loading: the case of historical Istanbul. *Earthq Eng Struct Dyn* 41:19–39
- Pau A, Vestroni F (2008) Vibration analysis and dynamic characterization of the Colosseum. *Struct Control Health Monit* 15(8):1105–1121
- Peeters B (2000) System identification and damage detection in civil engineering structures. PhD thesis, Katholieke Universiteit Leuven
- Peña F, Lourenço PB, Mendes N, Oliveira DV (2010) Numerical models for the seismic assessment of an old masonry tower. *Eng Struct* 32(5):1466–1478
- Ramos LF, Marques L, Lourenço PB, De Roeck G, Campos-Costa A, Roque J (2010) Monitoring historical masonry structures with operational modal analysis: two case studies. *Mech Syst Signal Process* 24(5):1291–1305
- Saisi A, Gentile C, Cantini L (2013) Post-earthquake assessment of a masonry tower by on-site inspection and operational modal testing. In: Proceedings COMPDYN 2013, Kos, Greece
- SVS (2012) ARTeMIS extractor 2012. <http://www.svbs.com>
- van Overschee P, De Moor B (1996) Subspace identification for linear systems: theory, implementation, applications. Kluwer, Boston/London/Dordrecht
- Welch PD (1967) The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short modified periodograms. *IEEE Trans Audio Electro-Acoust* 15(2):70–73

Analysis and Design Issues of Geotechnical Systems: Flexible Walls

Amir M. Halabian

Department of Civil Engineering, Isfahan
University of Technology, Isfahan, Iran

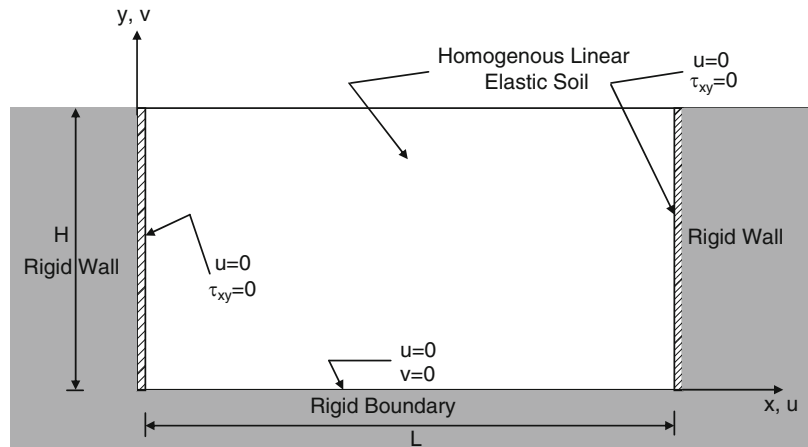
Synonyms

Non-yielding walls; Retaining flexible walls; Soil thrust

Introduction

The effect of ground motion on rigid retaining walls is examined using the method proposed by Mononobe and Okabe, which is based on the Coulomb's theory of static soil pressure. The analysis and design issues for these walls are covered in a separate companion article. M-O method requires that the retaining walls can move freely (slide or rotate) so that active or passive earth pressures develop behind the wall. Nevertheless, there are many cases (such as basement walls) where the free movement of the wall is fully or

Analysis and Design Issues of Geotechnical Systems: Flexible Walls,
Fig. 1 Walls geometry taken in Wood (1973) analysis of pressures on non-yielding walls



A

partially restrained, referred as non-yielding walls; furthermore, in situ retaining wall systems made from sequential excavation process provide stability and minimize movements of the adjacent ground throughout their flexible facing. In the last decades, a great deal of research work both in the analytical and in experimental areas has been performed to evaluate the adequacy of the M-O method or to extend the method for specific applications such as non-yielding walls or flexible retaining structures. Discussion of the existing approaches to evaluate the seismic soil pressure on non-yielding walls and practical design methods based on limit equilibrium concept and numerical algorithms are described below.

Non-Yielding Walls

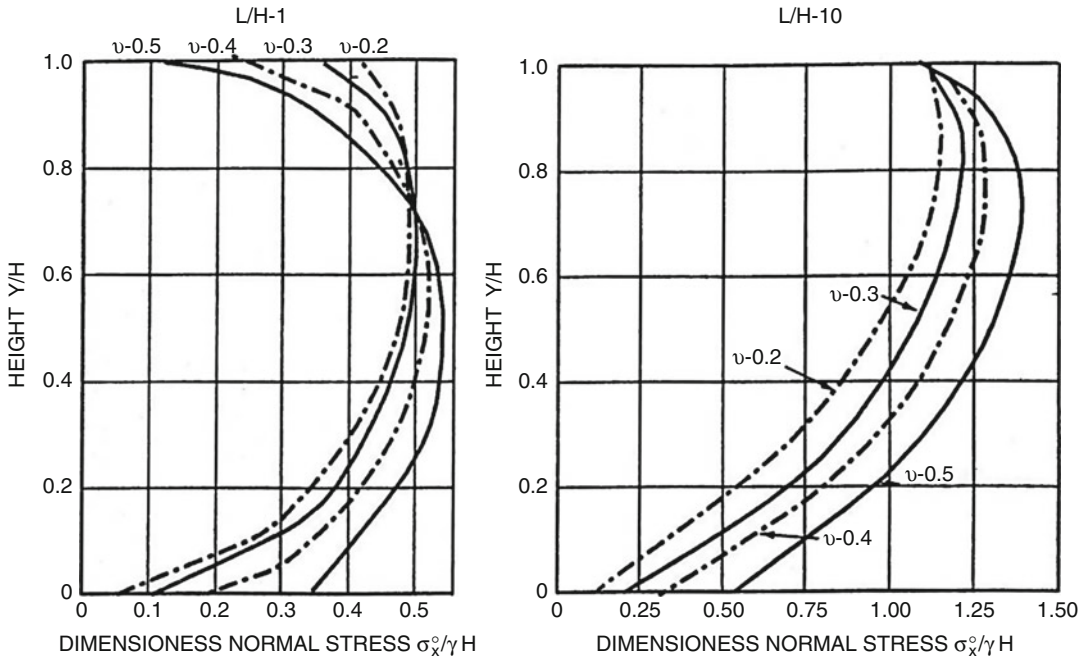
Although there are reports of damage and failure on retaining walls due to earthquakes around the world, the distress has been attributed to some form of soil or foundation failure, such as slope instability or soil liquefaction (Al Atik and Sitar 2010). There have been a few reports of damage to building basement walls as a result of seismic earth pressures in recent earthquakes including damages to basements in two recent earthquakes in Turkey. Gur et al. (2009) reported that basement damage occurred in a half-buried basement of a school building during 1999 Düzce earthquake; the half-buried basement was surrounded

by partial height earth-retaining concrete walls and there were windows between the top of the earth-retaining walls and the beams at the top of the basement.

Wood's Method

M-O method requires that the retaining walls can move freely (slide or rotate) so that active or passive earth pressures develop behind the wall. Nevertheless, there are many cases where the free movement of the wall is fully or partially restrained (such as basement walls, massive gravity walls embedded in rock-like formations, and braced walls). Wood (1973) analyzed the response of a homogenous linear elastic soil trapped between two rigid walls connected to rigid base (Fig. 1). In this figure, u , v , and τ_{xy} are referred to the in-plane deformations and shear stresses, respectively.

If both walls are spaced far apart, the pressures on one wall are not influenced by the presence of the other. For low-frequency input motions with frequency less than half the fundamental frequency of the unrestrained backfill, $V_s/4H$ (V_s is the soil shear wave velocity), the pseudo static conditions are governed (i.e., the dynamic amplification is negligible). For this range of frequencies, wall pressures with plane strain assumption can be obtained from elastic solution for the case of a uniform, constant horizontal acceleration applied throughout the soil. The dynamic earth pressures obtained from this method must be



Analysis and Design Issues of Geotechnical Systems: Flexible Walls, Fig. 2 Normal wall stresses from Wood's Solution (From Ebeling et al. 1992)

added to the static earth pressures to obtain the total earth pressures during an earthquake. Figure 2 presents two examples of the variation in the values for normalized horizontal stresses with normalized elevations above the base of a wall, based on Wood's solution (Wood 1973). An L/H value of 1 (where L = the width of backfill and H = the wall height) where value of 1 corresponds to a narrow backfill placed within rigid containment, and an L/H value of 10 represents a backfill of great width. The horizontal stresses (σ_x^0) at any vertical location Y, along the wall, are normalized by the product of γH in Fig. 2, where γ is the soil unit weight.

With the assumption of smooth rigid walls, the dynamic thrust and dynamic overturning moment about the base of the wall are expressed as:

$$\Delta F_{eq} = \gamma H^2 \frac{a_h}{g} F_p \quad (1)$$

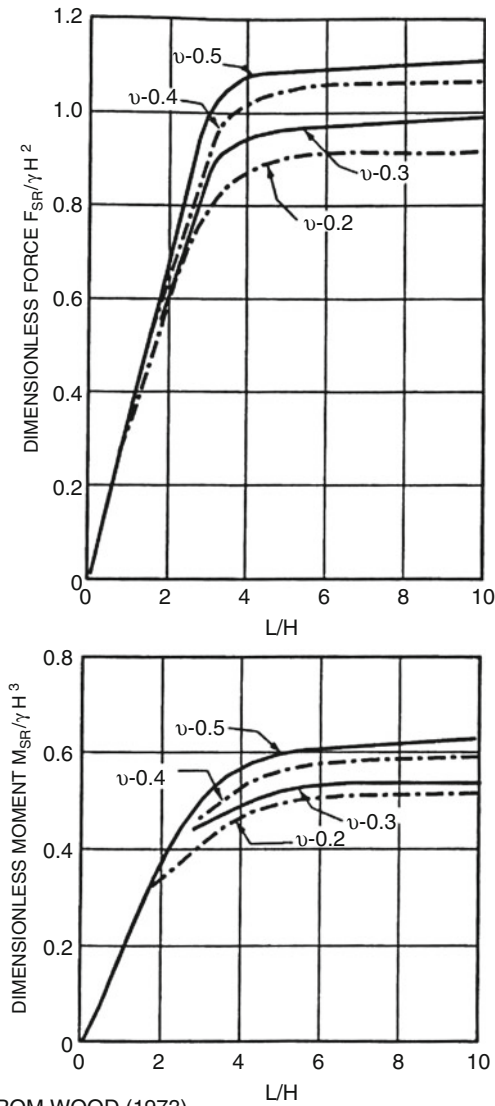
$$\Delta M_{eq} = \gamma H^3 \frac{a_h}{g} F_m \quad (2)$$

where a_h is the amplitude of the harmonic base acceleration, and F_p and F_m are the dimensionless dynamic thrust and moment factors which can be obtained from Fig. 3. The point of application the dynamic thrust will be located at a height

$$h_{eq} = \frac{\Delta M_{eq}}{\Delta F_{eq}} \quad (3)$$

Ostadan-White Approach

The solution by Wood (1973) commonly used for critical facilities are, in fact, based on static "1g" loading of the soil-wall system and does not include the wave propagation and amplification of motion. On the other hand, Wood's solution is mathematically complicated to apply in engineering practice and is limited to harmonic input motions. Employing the finite element technique, in a simplified method proposed by Ostadan and White (1997) and Ostadan (2004) which incorporates the main parameters affecting the seismic soil pressure for buildings, the lateral seismic soil pressure on these structures



Analysis and Design Issues of Geotechnical Systems: Flexible Walls, Fig. 3 Dimensionless thrust and moment factors in Wood's Solution (From Ebeling et al. 1992), ν is the soil Poisson's ratio

can be predicted. The studies showed that seismic soil pressure is affected by the long period part of ground motion and amplified near the resonant frequency of the backfill. The method is focused on the building walls rather than soil-retaining walls and specifically considers the dynamic soil nonlinear properties, relative motion between soil, and the structure and

frequency content of the design motion in its formulation. The method is focused on the building walls rather than soil-retaining walls where the movement of the walls is limited due to the presence of the floor diaphragms representing non-yielding walls. To reach a simplified approach for estimating the lateral seismic pressures on non-yielding walls, Ostadan (2004) performed a series of seismic soil-structure interaction analyses. A typical numerical model of a building basement wall is shown in Fig. 4. The base of the wall is resting on the rock or a firm soil layer.

Having the model shown in Fig. 4 subjected to 1g harmonic acceleration, the frequency contents of the pressure response were evaluated using the pressure transfer functions (TF) amplitude which is the ratio of the amplitude of the seismic soil pressure to the amplitude of the input motion. The pressure transfer functions were evaluated using Poisson's ratio of 1/3 for a wide range of frequencies (Fig. 5a).

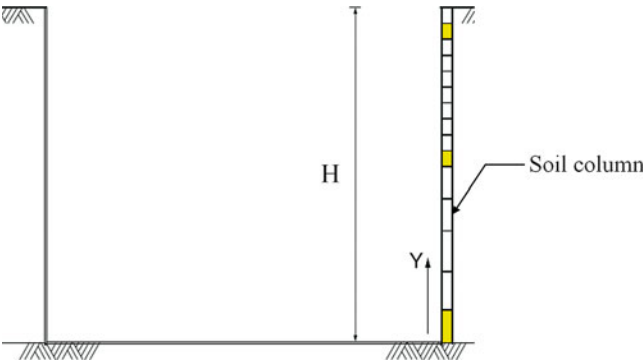
If the transfer functions are expressed in terms of normalized frequencies using the soil column frequency which is

$$f_s = V_s / 4H \quad (4)$$

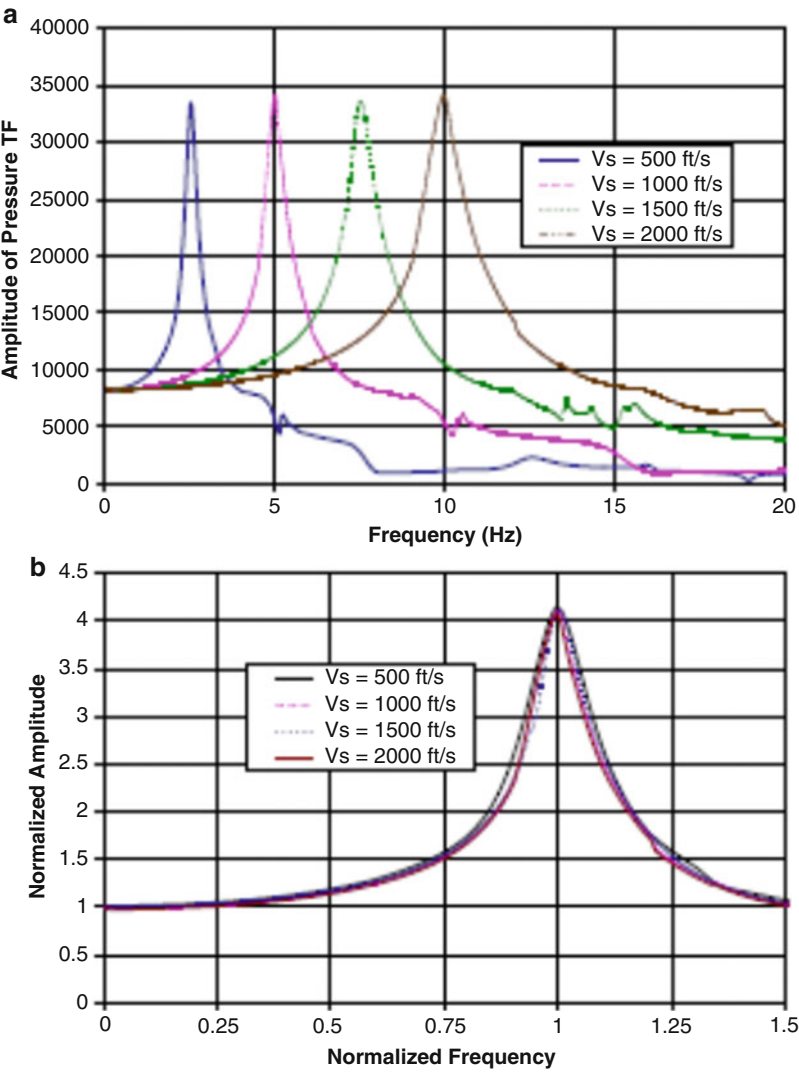
it was shown that the maximum amplifications take place at the frequency corresponding to the soil column frequency (Fig. 5b). V_s is the soil column shear velocity. Comparing the dynamic characteristics of the normalized pressure amplitudes (Fig. 5b), one can conclude that such characteristics are similar to a single degree of freedom (SDOF) system.

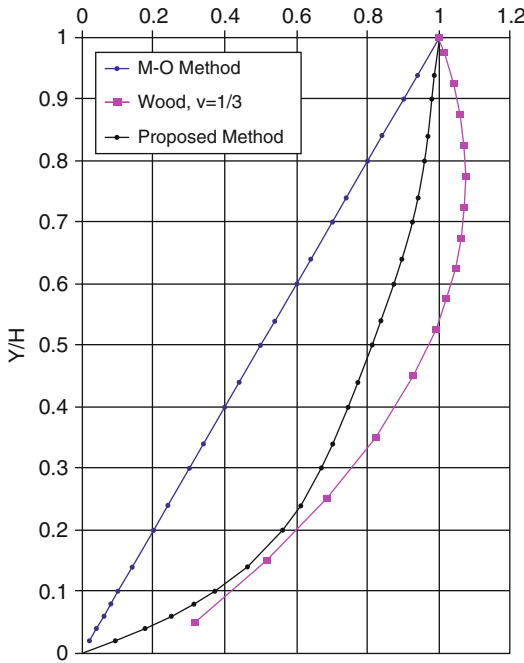
For the parametric studies conducted by Ostadan (2004), the results showed that low-frequency pressure profiles depict the same pressure distribution along the height of the wall. In Fig. 6, the soil pressure distribution obtained from such a nonlinear numerical method is compared with the normalized solutions from Wood (1973) and M-O methods. Using the soil pressure distributions obtained from a parametric study, a polynomial relationship has been developed to fit the normalized pressure curves. This relationship is expressed in terms of the

Analysis and Design Issues of Geotechnical Systems: Flexible Walls,
Fig. 4 A typical numerical model of a building basement wall



Analysis and Design Issues of Geotechnical Systems: Flexible Walls,
Fig. 5 Pressure transfer functions (Ostadan (2004))
(a) A typical pressure transfer functions versus frequency (b) Normalized transfer functions versus normalized frequency





Analysis and Design Issues of Geotechnical Systems: Flexible Walls, Fig. 6 Normalized pressure distributions (Ostadan (2004))

normalized height, $y = Y/H$ (Y coordinate defined in Fig. 4), as:

$$p(y) = -0.0015 + 5.05y + 15.84y^2 + 28.25y^3 + 24.59y^4 + 8.14y^5 \quad (5)$$

The total thrust applied to the wall is corresponding to the area under the normalized pressure profile curve which can be obtained from the integration of the pressure distribution over the height of the wall. The total area is $0.744H$ for a wall with height H . Keeping in mind that the normalized shape of the pressure distribution is similar to the response of SDOF systems, the amplitudes of the seismic pressure can be obtained from the response spectrum analysis of SDOF systems. Accordingly, the total thrust applied to the wall is subsequently obtained from the product of the total mass in the equivalent SDOF system times the acceleration spectral value at the respective frequency of the system. Having the total thrust, the frequency of

the system and the input motion to the SDOF system, the relationship in the form proposed by Veletsos and Younan (1994) can be used to compute the total mass and the damping of the SDOF system. Therefore, the total mass is obtained from

$$m = 0.5\rho H^2\Psi_v \quad (6)$$

in which ρ is the soil density and Ψ_v is a factor to take into account the Poisson's ratio effect. Since the original study performed for Poisson's ratio equals 1/3 (Ostadan (2004)), the pressure distribution should be adjusted for different Poisson's ratios using a factor recommended by Veletsos et al., Ψ_v , defined as

$$\Psi_v = \frac{2}{[(2-v)(1-v)]^{0.5}} \quad (7)$$

Study of the soil pressure transfer functions and the free-field response motions for wall with height of 50 ft showed that spectral values at the soil column frequency and at 30 % damping have the best correlation with the forces computed directly from the SSI analysis (Ostadan (2004)).

The computational steps of the proposed method by Ostadan (2004) to evaluate the lateral seismic soil pressures applied to the basement walls as a type of the non-yielding wall can be summarized as follows:

- Perform seismic free-field soil column analysis and obtain the acceleration response spectrum at the base of the wall using a free-field analysis code such as SHAKE91 (Schnabel et al. 1972) taking into account soil nonlinearities with input motion specified at the ground surface or at the depth of foundation base mat, $S_{a(\text{free})}$.
- Obtain total mass in the representative SDOF using Eq. 6.
- Obtain the total seismic lateral force by multiplying the mass of the representative SDOF by the spectral amplitude of the free-field response at the soil column frequency.

$$F = m S_{a(\text{free})} \quad (8)$$

- The maximum lateral earth pressure at the ground surface is calculated by dividing the total seismic lateral force by the area under normal soil pressure curve (0.744H).
- The lateral soil pressure distribution along the height of the wall is obtained by multiplying Eq. 5 by the maximum lateral earth pressure at the ground surface.

Soil-Reinforced Walls

Limit Equilibrium Methods

The fundamental stability concept of soil-nailed walls is based on reinforcing soil mass with reinforcement elements such as steel rebars so that the soil mass could behave as a unit mass. Therefore, during an earthquake, in addition to static forces, a reinforced soil wall is subjected to a dynamic soil thrust at the back of the reinforced zone and to internal forces within the reinforced zone. Accordingly, the wall should be designed to have global (external) stability such as avoiding sliding or overturning failure of the reinforced zone as well as local (internal) stability such as avoiding pullout failure of the reinforcement. Stability analyses based on limit-force equilibrium methods have been developed to assess the global static stability of soil-nailed walls along with local stability of reinforced soil mass by taking into account different factors influencing the wall performance, i.e., shearing, tension, and pullout resistance of the inclusions. Therefore, the reinforced wall is treated much like a gravity wall. The reinforced zone recommended by several researchers, shown in Fig. 7, is assumed to be acted on by its own weight, W , and the earthquake loading from unreinforced soil mass represented pseudostatically by dynamic soil thrust, F_D , and the inertial force on the reinforced zone, F_i . The normalized width of the reinforced soil mass, the inertial force on the reinforced zone, and the dynamic soil thrust suggested by Seed and Mitchell (1981), Dhouib (1987) and Segrestin and Bastick (1988) are given in Table 1.

Seed and Mitchell (1981) in their recommendation to FHWA suggested that the reinforced soil area should be taken equal to 0.5H independent of the ground horizontal acceleration, a_h . On the other hand, developing a nonlinear finite element code, Dhouib (1987) showed that the incremental dynamic force is proportional to the distribution of static forces, and the geometry of the active failure zone should be expressed in terms of the system's relative stiffness and the ground acceleration.

Segrestin and Bastick (1988) also conducted finite element analyses to understand seismic response of soil-reinforced structures. In their studies, the elastoplastic behavior of the soil was simulated by varying the modulus of elasticity as a function of observed deformations. The results of the study indicated that the distribution of dynamic tensile forces along the strips is fairly uniform and does not give significant change in the position of points of maximum tension. Unlike the FHWA recommendation, Segrestin and Bastick (1988) suggested that the active failure zone is not affected that much by the earthquake loading, and therefore the width of this area should be limited to 0.3 H. In their studies, the elastoplastic behavior of the soil was simulated by varying the modulus of elasticity as a function of observed deformations. The results of their study indicated that the distribution of dynamic tensile forces along the strips is fairly uniform and does not give significant change in the position of points of maximum tension.

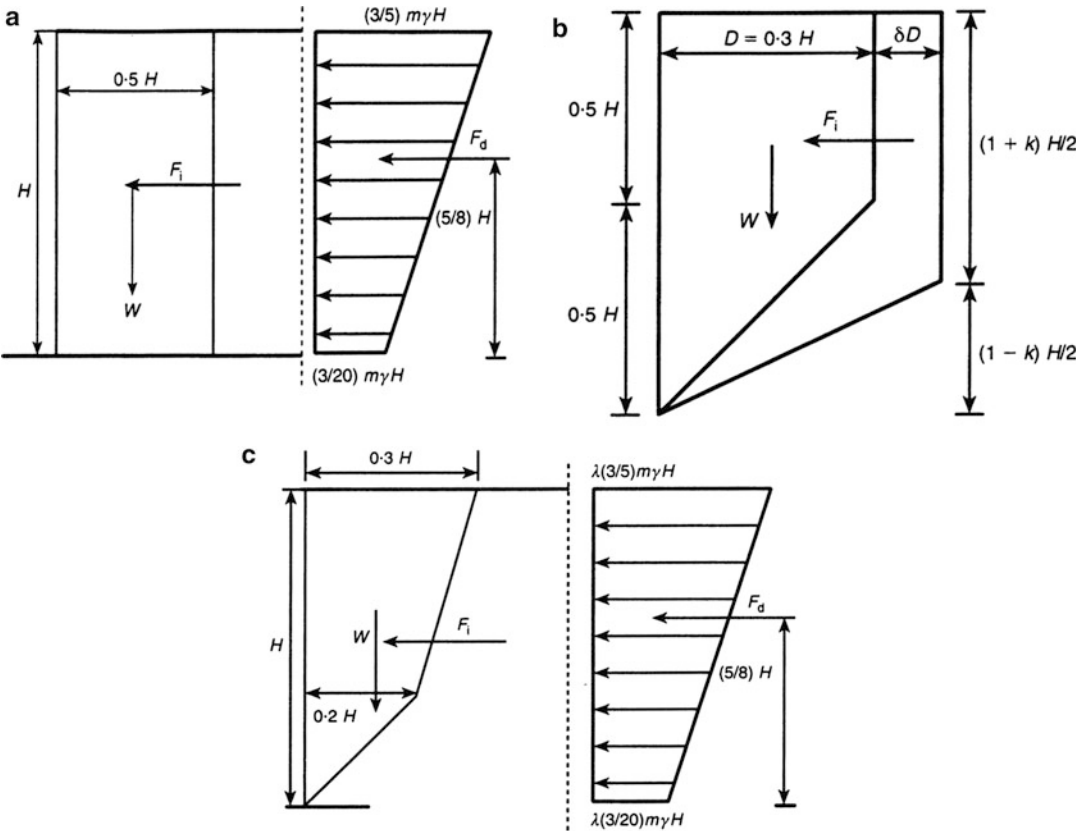
In Table 1, m is the horizontal ground acceleration coefficient calculated at the centroid of the reinforced zone using

$$m = \left(1.45 - \frac{a_h}{g} \right) \frac{a_h}{g} \quad (9)$$

recommended by FHWA.

Pseudostatic Approaches

While limit equilibrium methods fail to give any information regarding developed tensile and shear forces along nails and they cannot evaluate local stability of structure, numerically derived pseudostatic approaches have been developed to investigate behavior of soil-nailed walls under



Analysis and Design Issues of Geotechnical Systems: Flexible Walls, Fig. 7 Dynamic soil thrust acting on soil-reinforced walls’ evaluation using different pseudostatic approaches (the symbols are defined in Table 1)

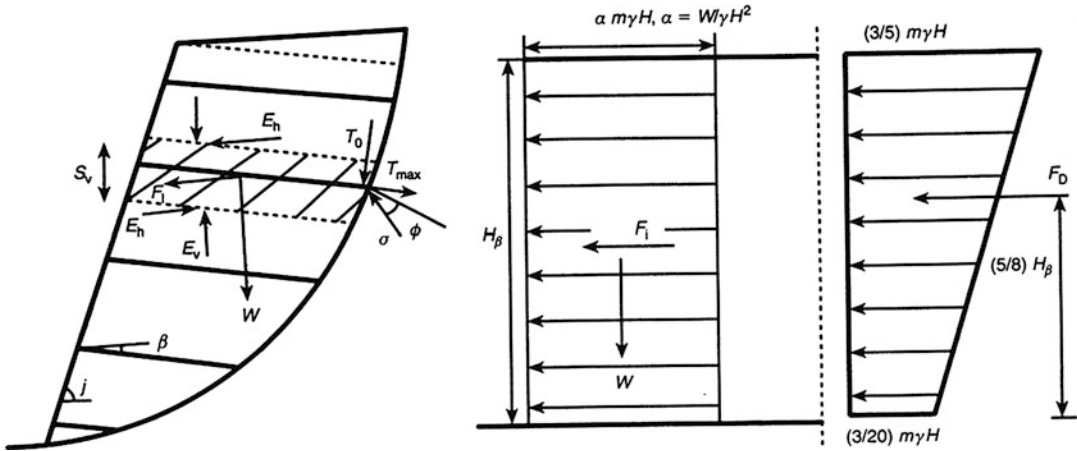
Analysis and Design Issues of Geotechnical Systems: Flexible Walls, Table 1 Inertia and dynamic forces applied to the reinforced soil mass

Design method	S/H	Inertial force F_i	Dynamic force F_D
Seed and Mitchell (1981)	0.50	$0.5 m \gamma H^2$	$(3/8) m \gamma H^2$
Dhouib (1987)	$0.30 + m/2$	$\frac{(0.30+m/2)m\gamma H^2}{4(2K+3)}$	—
Se grestin and Bastick (1988)	0.30	$0.20 m \gamma H^2$	$\lambda(3/8) m \gamma H^2$

$K = 2.5 m, \lambda = 0.6$

dynamic-loading conditions. On the other hand, unlike common gravity retaining structures, soil-nailed walls have also shown remarkably well performance during many ground motions,

which could be attributed to the intrinsic flexibility of soil-nailed wall system and possibly some level of conservatism in current design procedures (Choukeir et al. (1997)). Accordingly, as an extension of static limit equilibrium methods, pseudostatic approaches in which dynamic earth pressure is computed based on conventional Mononobe-Okabe approach were developed. Choukeir et al. (1997) presented a new pseudostatic method to analyze soil-nailed slopes. Their proposed approach is derived as an extension of the kinematical working-stress design method developed by Juran et al. (1990) for static loading of soil-nailed structures. The applicability of this working-stress design method under static loading has been described in details by Juran and Elias (1991) and is incorporated in several design codes (FHWA 2003).



Analysis and Design Issues of Geotechnical Systems: Flexible Walls, Fig. 8 Pseudostatic method suggested by Choukeir et al. (1997), H_β = the projected height of the inclined wall

The main design assumptions of the kinematical analysis are as follows:

- Failure occurs by a quasi-rigid body rotation of the active zone, which is limited by a log-spiral failure surface.
- The locus of the maximum tension and shear forces at failure coincides with the failure surface developed in the soil mass.
- The shearing resistance of the soil, defined by Coulomb's failure criterion, is entirely mobilized along the sliding surface.
- The shearing resistance of stiff inclusions, defined by Tresca's failure criterion, is mobilized in the direction of the sliding surface in the soil.
- The horizontal components of the inter-slice forces, E_h (see Fig. 8), are equal.
- The effect of a slope (or horizontal surcharge), at the upper surface of the nailed soil mass, on the tension forces in the nails is linearly decreasing along the failure surface.

The extension of the kinematical design method for pseudostatic stability analysis involves the following assumptions:

- The failure-reinforced wedge is subjected to horizontal pseudostatic forces specified as the

mass of the wedge multiplied by the pseudostatic horizontal acceleration $a_h = mg$.

- Shear modulus of the soil is constant along the depth.
- Shear strength parameters of the soil are constant during earthquake loading.
- The soil is not saturated, and thus no effect of pore water pressure is present.
- The failure-reinforced wedge is assumed to be subjected by its own weight, W , and the earthquake loading from unreinforced soil mass represented pseudostatically by dynamic soil thrust, F_D , and its own inertial force, F_i (Fig. 8)

$$K_h = n.m \dots, \dots, n = \frac{a_{ps}}{a_m}$$

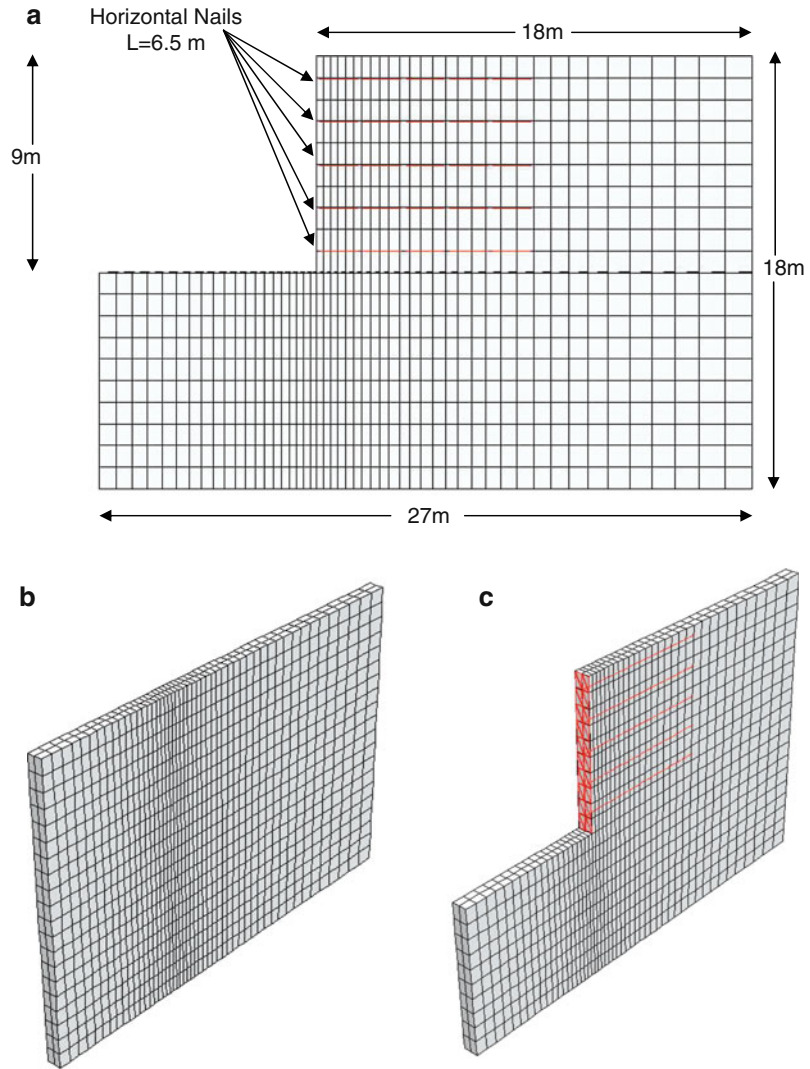
$$= \left[\frac{0.5}{1-r^2} \right]^{\frac{1}{2}} \dots, \dots r = \frac{\omega}{\omega_n} \quad (10)$$

Do check that all symbols in Eq. 10 have been defined somewhere else.

Nonlinear Dynamic Analysis Method

As it was presented, there are several pseudostatic methods by which the magnitude of dynamic forces along reinforcements for a soil-nailed structure under dynamic loading conditions could be determined. To get better

Analysis and Design Issues of Geotechnical Systems: Flexible Walls, Fig. 9 (a) Model geometry (b) Numerical mesh used for analyses (c) Soil-nailed wall after final stage of excavation (Halabian et al. 2010)

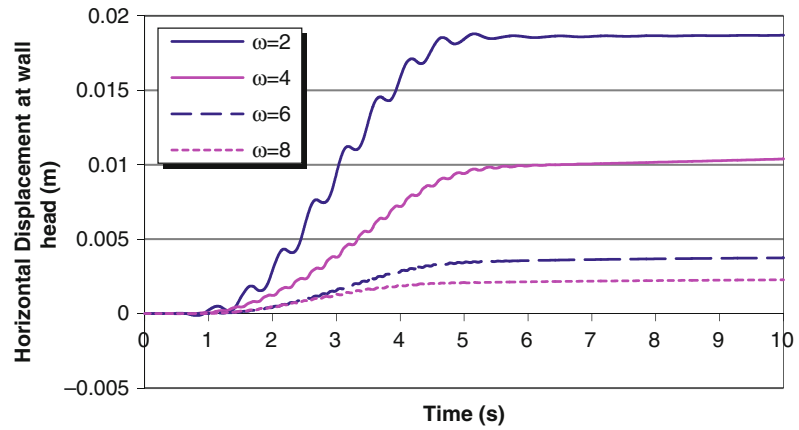


understanding of dynamic response of soil-nailed wall systems under earthquake excitations, a numerical model like what is shown in Fig. 9 for a typical reinforced wall, rather than empirical formulations could be employed to evaluate the wall performance.

To gain better perception of dynamic soil-nailed structures behavior, the numerical model should take into account the wall construction stages which include nail installation and application of shotcrete facing. It should be noted that the model dimensions have to be determined wide enough so that the effects of boundary conditions on the response of structure would be

negligible. Due to repetitive arrangement of nails along the length of the excavation, only a slice of the soil can be modeled (Fig. 9b and c). To have more reliable results of analyses, a somewhat finer mesh should be utilized for those areas near the excavation face. It should be noted that initial state of equilibrium has to be established before any stage proceeds. Afterward as each excavation ends, the equilibrium conditions (according to maximum unbalanced force) will be examined in order to ensure whether or not that model meets the equilibrium conditions to resume the next stage. An appropriate soil constitutive law should be used to simulate the stress

Analysis and Design Issues of Geotechnical Systems: Flexible Walls, Fig. 10 Horizontal displacement response time history at the wall head during the harmonic base excitation (Halabian et al. 2010)



path induced to soil mass during the earthquake loading. After static equilibrium is achieved, the full width of soil subgrade should be subjected to dynamic excitations.

Harmonic Ground Excitation

Despite having a single-frequency content and short duration of the excitation, Halabian et al. (2010) used a variable-amplitude harmonic ground motion to understand the effects of crucial parameters on the performance of soil-nailed structure under actual earthquake acceleration records. The accelerogram follows a trend which has both increasing and decaying peak acceleration parts expressed as:

$$\ddot{U}(t) = \frac{1}{2} \left(1 - \cos \left(\frac{2\pi}{T} t \right) \right) \sin(2\pi f t) \quad (11)$$

in which T and f are the duration of net harmonic excitation and the frequency of excitation.

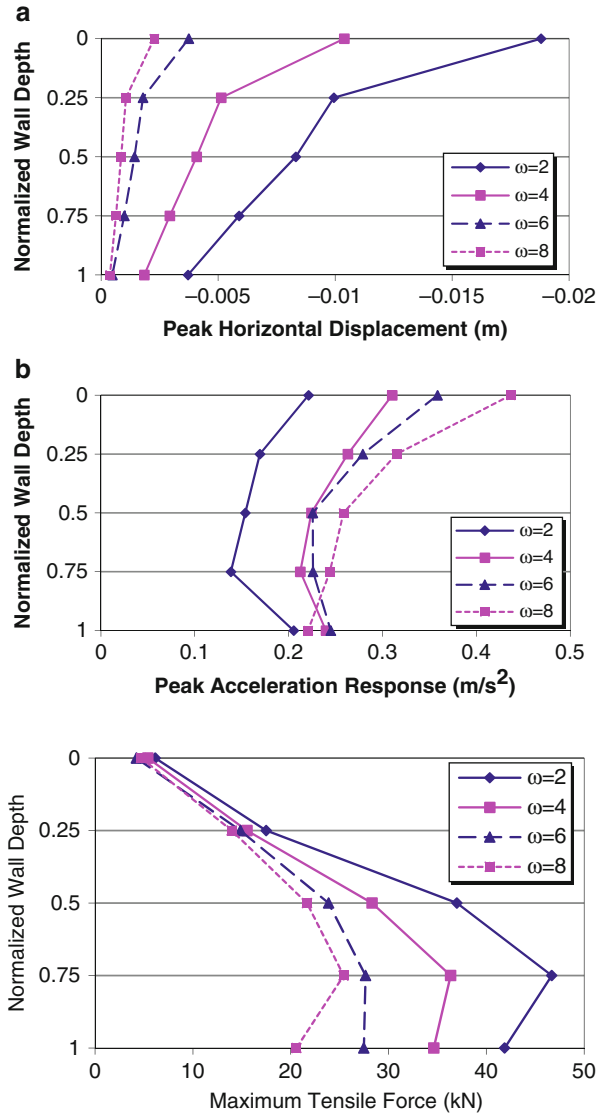
Influence of Harmonic Excitation Frequency The effect of the excitation frequency on the horizontal displacement time histories at the crest of the wall is presented in Fig. 10. The displacement time histories indicate that the horizontal displacement at the crest of the wall was influenced by the excitation frequency and increases monotonically with the time during the harmonic base excitation. However, it can be concluded that as the excitation frequency increases, the horizontal displacement at the

wall crest decreases. Figure 11a indicates that while the input frequency motion increases, the peak outward horizontal displacement of the wall decreases. Conversely, as illustrated by Fig. 11b, the peak horizontal acceleration response along the wall facing is increased by an ascending trend of input motion frequency. From the other results shown in Fig. 12, it is worth noting that increasing input motion frequency made tensile forces to less mobilize along nail bars due to the fact that the critical state becomes more inaccessible since the input motion frequency gets higher than fundamental frequency of the system.

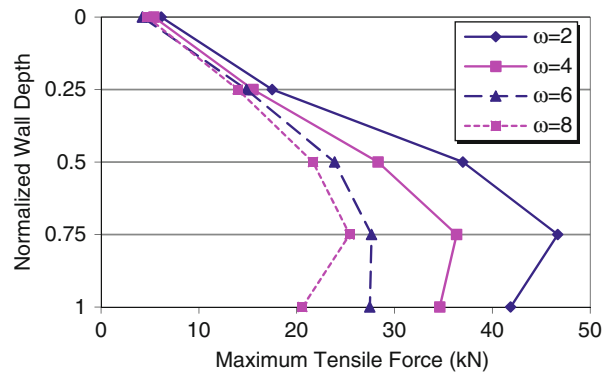
Influence of Angle of Nail Inclination The influence of nail inclination angle on dynamic response of the structure is shown in Fig. 13a. The peak outward horizontal displacement of walls shows decreasing in value by increasing nails inclination angle while it varies in the range of 0–15°. Afterwards, increasing nail inclination angle results in slight increase of wall deformations, especially when it ranges in 25–30°. As a result, it is shown that the increase in nail inclination angle would have destabilizing effect on the performance of structure during harmonic base excitations.

When it comes to the peak horizontal acceleration, as it is demonstrated in Fig. 13b, it can be concluded that increasing nail inclination angle would result in increase of peak horizontal acceleration response of the wall. Figure 14 shows distribution of maximum tensile forces along nail bars.

Analysis and Design Issues of Geotechnical Systems: Flexible Walls, Fig. 11 Normalized peak horizontal deformations distribution along height of wall facing (Halabian et al. 2010) (a) Normalized peak horizontal displacement (b) Normalized peak horizontal displacement



Analysis and Design Issues of Geotechnical Systems: Flexible Walls, Fig. 12 Maximum tensile forces distribution along height of wall facing (Halabian et al. 2010)



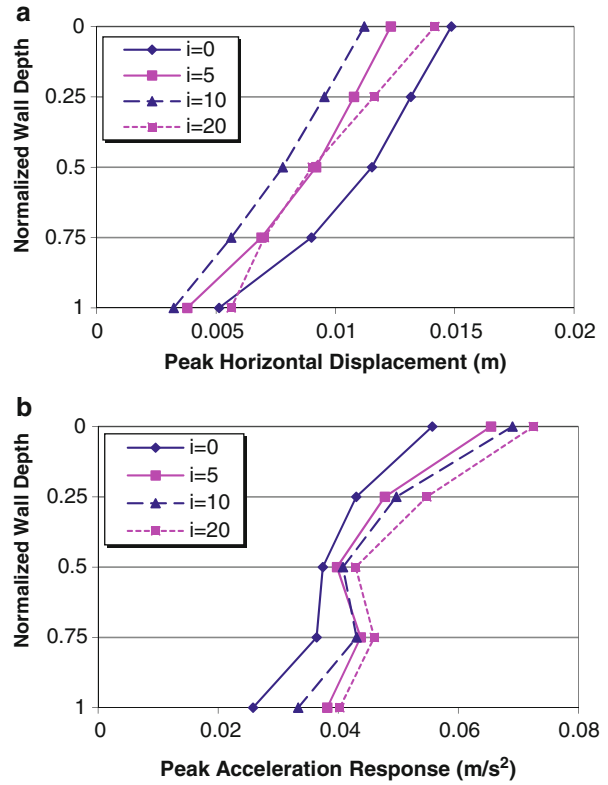
It is of interest to note that tensile forces along nail bars increase for the upper nail bars as their inclination angle increases.

Influence of Nail Length The effect of nail length on the peak outward horizontal displacement of the wall is shown in Fig. 15a. It can be noted that the peak outward horizontal displacement of the wall decreases as the nails length increases; while, the peak horizontal acceleration along the wall facing has been greater for longer nail bars (Fig. 15b). However, the major influence of nail length on increasing the peak

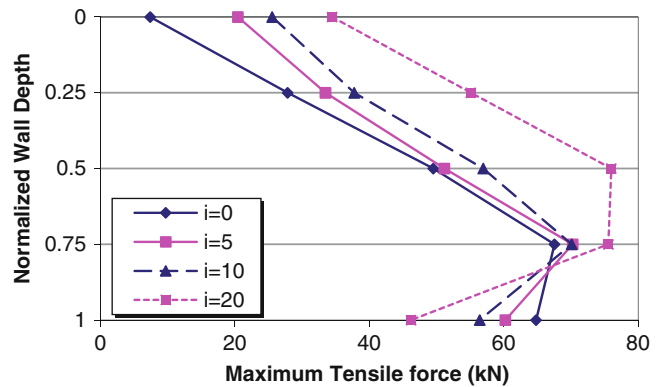
horizontal acceleration response of the wall would be in the range of 7.5–10 m. Therefore, there would be no considerable influence as the nail length increases beyond the value of 12 m. The results of the maximum nails tensile forces (Fig. 16) indicate that there would be no significant differences in the mobilized maximum tensile forces along nail bars (excluding the third nail) as the nails length increases.

Influence of Soil Strength Properties The soil strength properties, soil friction angle, ϕ , and soil cohesion, c , can also influence on the

Analysis and Design Issues of Geotechnical Systems: Flexible Walls,
Fig. 13 Normalized peak horizontal deformations distribution along height of wall facing (Halabian et al. 2010) (a) Normalized peak horizontal displacement (b) Normalized peak horizontal acceleration



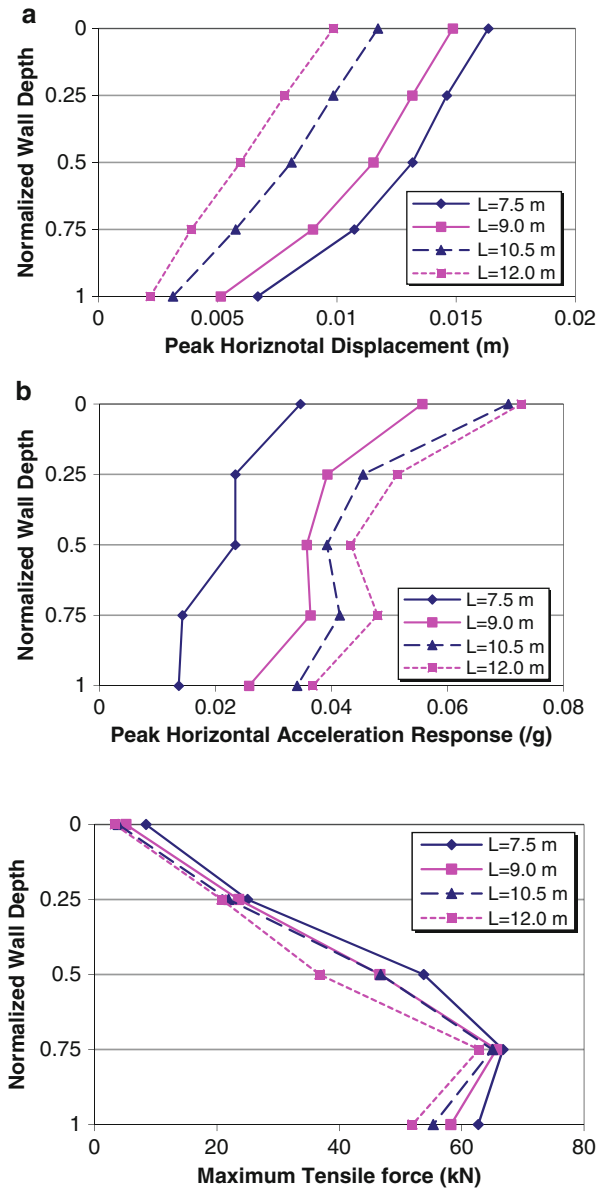
Analysis and Design Issues of Geotechnical Systems: Flexible Walls,
Fig. 14 Maximum tensile forces distribution along height of wall facing (Halabian et al. 2010)



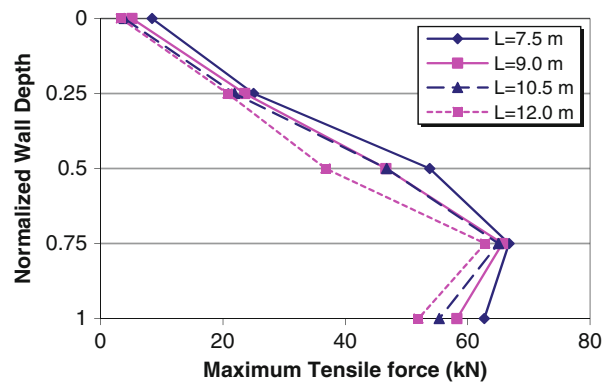
wall performance such as horizontal deflections along the wall facing and the maximum mobilized tensile forces along the nail bars during an earthquake. Global performances of soil-nailed walls during earthquakes have shown that increasing in soil strength properties would improve soil shear strength so that soil-nailed wall would better resist against

dynamic loads due to harmonic base excitations. As illustrated in Fig. 17a and b, peak horizontal displacements along wall facing decrease as the soil cohesion and internal friction angle increase. Figure 17c and d show that increasing soil strength properties would result in reduction of maximum tensile forces along nail bars.

Analysis and Design Issues of Geotechnical Systems: Flexible Walls,
Fig. 15 Normalized peak horizontal deformations distribution along height of wall facing (Halabian et al. 2010) (a) Normalized peak horizontal displacement (b) Normalized peak horizontal acceleration



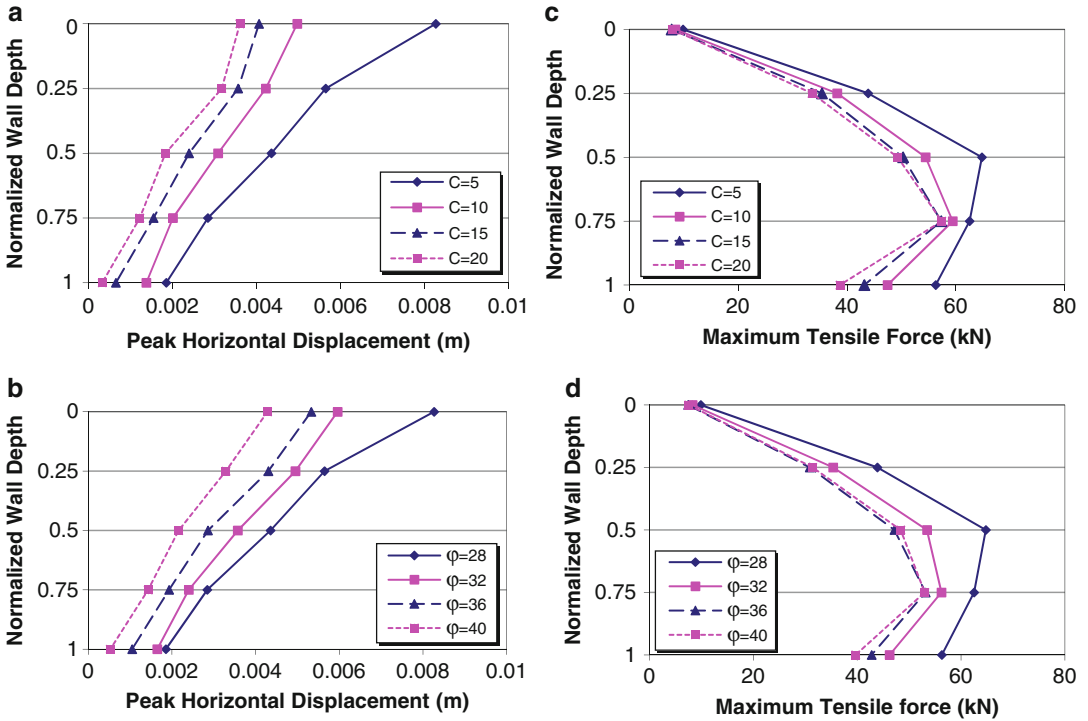
Analysis and Design Issues of Geotechnical Systems: Flexible Walls,
Fig. 16 Maximum tensile forces distribution along height of wall facing (Halabian et al. 2010)



Seismic Base Excitations

Similar to the harmonic base excitations, numerous parameters including nails inclination, nails lengths, soil mechanical properties, and nail spacing can contribute to the performance of soil-nailed walls during seismic excitations. In addition to these parameters, the earthquake motions depending on their frequency content and peak amplitude have also substantial influences on deformations of soil-nailed walls.

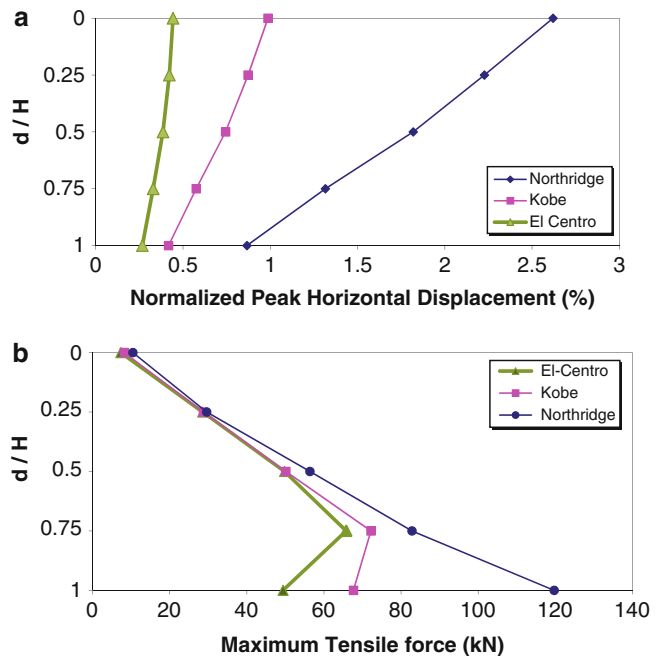
In a study conducted by Halabian et al. (2010) for a soil-nailed wall example, the effects of three different base excitation records (El Centro, Kobe, and Northridge earthquakes) were examined. It is noteworthy that the ground motion with the frequency content close to the soil-nailed system and higher peak amplitude induced considerable lateral displacements along the wall facing compared to the other ground motions (Fig. 18a).



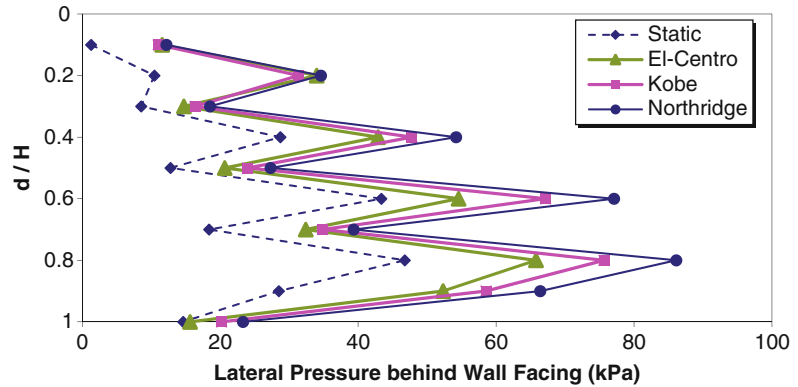
Analysis and Design Issues of Geotechnical Systems: Flexible Walls, Fig. 17 Normalized peak horizontal displacement and maximum tensile forces distribution along height of wall facing (Halabian et al. 2010) (a)

Normalized peak horizontal displacement (b) Normalized peak horizontal displacement (c) Maximum tensile forces distribution along height of wall facing (d) Maximum tensile forces distribution

Analysis and Design Issues of Geotechnical Systems: Flexible Walls, Fig. 18 The peak response distributions along the height of wall facing (Halabian et al. 2010) (a) Normalized peak horizontal displacement distribution (b) Maximum tensile forces distribution



Analysis and Design Issues of Geotechnical Systems: Flexible Walls,
Fig. 19 Lateral earth pressure distribution behind wall facing along height of wall (Halabian et al. 2010)



Besides, as illustrated in Fig. 18b, earthquake motions with high peak amplitude of base acceleration would cause higher nail forces along nails. There is considerable difference between the values of nail forces in the fifth row of nails. Furthermore, the earth lateral pressure distribution behind the wall facing for different earthquake motions is depicted in Fig. 19, and it can be concluded that earth pressure will increase as the base excitation peak amplitude rises. The observed pattern in earth lateral distribution could be interpreted due to stage construction of soil-nailed structures.

Summary

Several methods to evaluate the maximum seismic soil pressures for some cases (such as basement walls) where the free movement of the wall is fully or partially restrained, referred as non-yielding walls, are presented. Furthermore, based on a comprehensive parametric study, the influence of crucial parameters on performance of soil-nailed walls after being subjected to harmonic and seismic base excitations were also given.

Cross-References

- [Bridge Foundations](#)
- [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)
- [Seismic Design of Earth-Retaining Structures](#)

References

- Al Atik L, Sitar N (2010) Seismic earth pressures on cantilevered retaining structures. *J Geotech Geoenviron ASCE Eng* 136:1324–1333
- Choukeir M, Juran I, Hanna S (1997) Seismic design of reinforced-earth and soil-nailed structures. *Ground Improv* 1:223–238
- Dhouib A (1987) Contribution a L'etude du comportement des sols renforces sous sollicitations statiques et dynamiques. These de Docteur-Ingenieur, Universite des Sciences et Techniques de Lille Flandres Artois, France
- Ebeling RM, Morrison EE, Whitman RV, Liam Finn WD (1992) A manual for seismic design of waterfront retaining structures. US Army Corps of Engineers, Technical Report ITL-92-11
- Federal Highway Administration (FHWA), Lazarte CA, Elias V, Espinoza RD, Sabatini PJ (2003) Geotechnical engineering circular no. 7, Soil nailed walls, report no. FHWA0-IF-03-017. Washington, DC
- Gur T, Pay AC, Ramirez JA, Sozen MA, Johnson AM, Irfanoglu A, Bobet A (2009) Performance of school buildings in Turkey during the 1999 Düzce and the 2003 Bingöl Earthquakes. *Earthq Spectra* 25(2):239–256
- Halabian AM, Sheikhbahaei AM, Hashemolhosseini SH (2010) Analysis of nailed walls under seismic excitation using finite difference method. In: 9th US national and 10th Canadian conference on earthquake engineering, Toronto
- Juran I, Baudrand G, Farrag, K, Elias V (1990) Kinematical limit analysis for design of soil nailed structures. *J Geotech Eng* 116(1):54–72
- Juran I, Elias V (1991) Ground anchors and soil nails in retaining structures. *Foundation Engineering Handbook*, 2nd edn, Hasai-Yang Fang, Chapter 26
- Ostadan F (2004) Seismic soil pressure for building walls—an updated approach. In: 11th international conference on soil dynamics and earthquake engineering (11th ICSDEE) and the 3rd international conference on earthquake geotechnical engineering (3rd ICEGE), University of California, Berkeley

- Ostadan F, White WH (1997) Lateral seismic soil pressure-an updated approach. Bechtel Technical Grant Report, Bechtel Corporation, San Francisco
- Schnabel PB, Lysmer J, Seed HB (1972) SHAKE – a computer program for earthquake response analysis of horizontally layered sites. Earthquake Engineering Research Center, University of California, Berkeley, Report no. EERC 72-12, Dec
- Seed HB, Mitchell JK (1981) Earthquake resistant design of reinforced earth walls. International study for the reinforced earth company, Progress report, Berkeley
- Segrestin P, Bastick MJ (1988) Seismic design of reinforced earth retaining walls. In: Proceedings of the international geotechnical symposium on theory and practice of earth reinforcement, Fukuoka Kyushu, pp 577-582
- Veletsos A, Younan AH (1994) Dynamic soil pressure on rigid vertical walls. *Earthq Eng Soil Dyn* 23:275-301
- Wood J (1973) Earthquake-induced soil pressures on structures. Report EERL 73-05, California Institute of Technology, Pasadena

Analysis and Design Issues of Geotechnical Systems: Rigid Walls

Amir M. Halabian
Department of Civil Engineering, Isfahan
University of Technology, Isfahan, Iran

Synonyms

Pseudo-dynamic method; Rigid retaining walls; Yielding walls

Introduction

Earthquakes have caused transient and permanent deformations of retaining structures in past severe earthquakes followed by collapse of walls in some cases. Therefore, knowledge of seismic active earth pressure behind retaining walls is very important in the design of these structures in seismically active regions. This article discusses commonly used pseudo-static approaches such as the Mononobe-Okabe method, which gives the linear distribution of seismic earth pressure on rigid retaining walls in an approximate way. A general pseudo-dynamic method is also

presented to compute the distribution of seismic active earth pressure on a rigid retaining walls supporting cohesionless backfill in more realistic manner by considering time and phase difference within the backfill.

Types of Retaining Walls

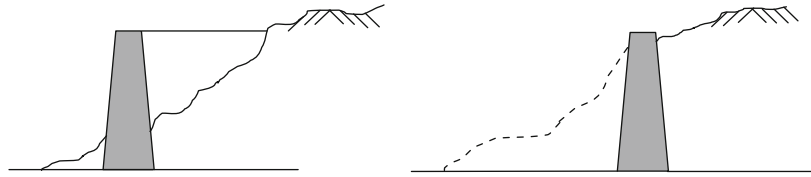
Earth retaining structures are structures that can be used to support backfill along a slope or to support an excavation as illustrated in Fig. 1.

Various types of retaining structures are adopted for different applications. Over the time, the classical gravity rigid retaining walls evolved into reinforced concrete cantilever walls (e.g., sheet piles), with or without buttresses and counter forts (Fig. 2). These were then followed by a variety of crib- and bin-type walls. All these walls are externally stabilized walls or conventional gravity retaining walls.

In order to stabilize the deep excavation projects in metropolitan areas or even slopes where sequential construction in comparison with other common retaining walls is beneficial, the flexible retaining structures could be considered. During the sequential excavation, an in situ flexible retaining wall system is constructed to provide stability and to minimize movements of the adjacent ground. Soldier piles with shotcrete lagging are being used extensively as an excavation support system, particularly in stiff soil conditions and where ground water ingress into the excavated area is not problematic (Tomlinson 1995). Furthermore, since the soldier piles are not contiguous, much fewer soldier piles often need to be driven in comparison with sheet piles, thereby yielding significant savings in time and cost of installation and thus allowing excavation to commence with a minimum of lead time. These systems are usually being constrained against certain types of movements by the presence of external bracing elements such as struts (Figs. 3 and 4). In the case of basement walls or bridge abutments, lateral movements at the top of the retaining structure are restrained by the presence of the structure they support.

Soil reinforcement in various constructions is also used to provide the stability for excavations

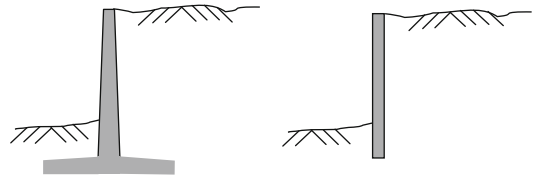
Analysis and Design Issues of Geotechnical Systems: Rigid Walls, Fig. 1 Gravity retaining structures



and is proven to be effective and under practice from the last two decades. Among the reinforced soil retaining structures, geosynthetic reinforced soil retaining walls with modular block facing are functioned well during the past seismic events compared to the other flexible retaining walls. The reinforced layers of soil in the mechanically stabilized earth mass made with this technique allow the modular construction, which was clearly recognized as being advantageous in most practical situations (Fig. 5).

Soil nailing is another in situ reinforcing technique of the soil while it is excavated from the top down. Using soil-nailed walls which consist of passive reinforcement to the existing ground by installing closely spaced steel bars or cables encased in grout has gained a lot of popularity in recent years (Fig. 6). This method is typically used in order to stabilize slopes and excavations where sequential construction is beneficial in comparison with other common retaining walls. The mass of reinforced soil made by an array of soil nails functions to retain the less stable material behind it. In the right soil conditions, soil nailing is a rapid and economical means of constructing excavation support systems and retaining walls. The process of construction of soil-nailed walls commonly involves three important stages: excavation, nail installation, and face stabilization. The fundamental stability concept of soil-nailed walls is based on reinforcing soil mass with reinforcement elements such as steel rebars so that the soil mass could behave as a unit mass.

Due to significant flexibility of soil-nailed walls which is attributed to particular construction procedure of these systems, soil-nailed walls can experience more deflections comparing with other common gravity walls. After the 1989 Loma Prieta, 1995 Kobe, and 2001 Nisqually earthquakes, it was reportedly observed that soil-nailed walls have shown no sign of being



Analysis and Design Issues of Geotechnical Systems: Rigid Walls, Fig. 2 Cantilever retaining structures

distressed or significant permanent deflection, despite having experienced, in some cases, ground accelerations as high as 0.7 g. The observations from post-earthquake investigations imply that soil-nailed walls appear to have an inherent satisfactory seismic response. This has been attributed to the intrinsic flexibility of soil-nailed wall system and possibly some level of conservatism in current design procedures (Choukeir et al. (1997)).

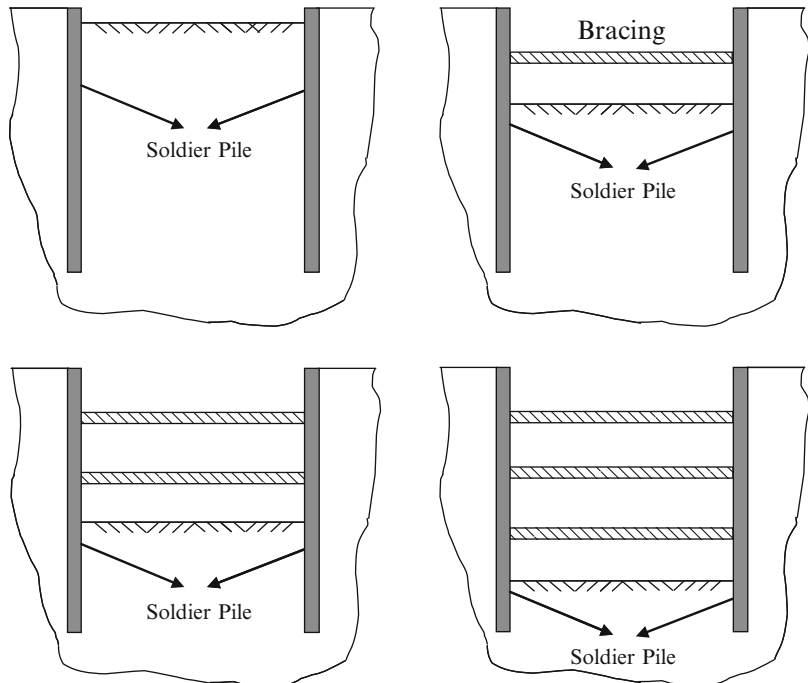
Although soil nailing is now recognized as a viable technique in retaining structures, tiebacks have been used and continue to be employed as a temporarily supporting system for excavations. Permanent tieback walls have been used in cuts, bridge abutments, underpinning of structures, and stabilization of sliding slopes. Essentially, there are two types of tieback walls: the slurry wall that is also called diaphragm wall and the soldier pile wall. The main advantage of using soldier piles is their relatively low cost and ease of installation, compared to other forms of supporting systems such as diaphragm walls and bored piles.

Furthermore, since the soldier piles are not contiguous, much fewer soldier piles often need to be driven in comparison to sheet piles. The soldier pile walls are built by driving piles (usually H-piles) in a line with spacing of the order of 2–3 m. Soldier piles are driven to a depth slightly below the final excavation. Sometimes bored piles are used by drilling a hole, lowering an H-pile in the center and filling the annulus with low strength concrete.

**Analysis and Design
Issues of Geotechnical
Systems: Rigid Walls,
Fig. 3** Braced excavation



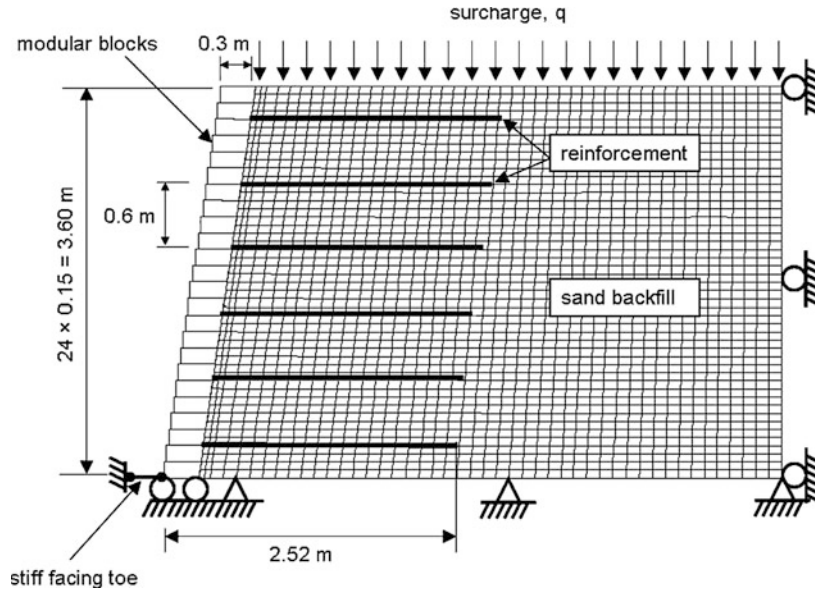
**Analysis and Design
Issues of Geotechnical
Systems: Rigid Walls,
Fig. 4** Braced excavation
steps



As excavation proceeds, wood lagging or sheeting (concrete shotcrete) is placed to retain the soil between the piles. Then, the anchors are installed at regular intervals and grouted in a zone beyond

the failure zone. The anchors are stressed up to a chosen load when the grout has sufficiently cured. The process continues until the final excavation level is reached. Tiebacks eliminate

Analysis and Design Issues of Geotechnical Systems: Rigid Walls,
Fig. 5 Geosynthetic reinforced soil retaining wall with modular block facing

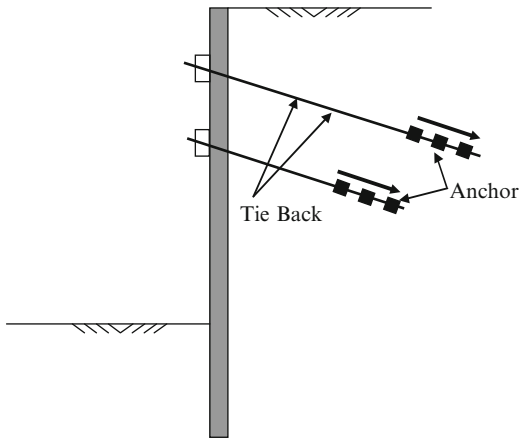


Analysis and Design Issues of Geotechnical Systems: Rigid Walls,
Fig. 6 Soil-nailed walls



obstructions in the excavation inherent in struts. The total structural system acts in tension and receives its support in earth and, therefore, consists of the earth mass, which provides the ultimate support for the system, tension members which transfers the load from the soil-retention system to the earth mass. A stressing unit, which engages the anchors, permits the tieback element to be stressed and allows the load to be maintained in the tieback. A typical tieback wall is shown in Fig. 7.

Depending on soil-induced deformation values, earth retaining structures are broadly categorized into non-yielding, yielding, and self-yielding walls. Non-yielding walls are inherently incapable of and/or constrained against both deformation and displacement in the horizontal direction under static or dynamic loads. Basement walls of buildings, bridge abutments, and free-standing retaining walls that are restrained against horizontal displacement due to physical restraint or structure geometry can be named as



Analysis and Design Issues of Geotechnical Systems: Rigid Walls, Fig. 7 Tieback retaining system

the common examples. In the static loading phases, non-yielding structures can be logically designed assuming the at-rest earth pressure state within the retained soil. However, in the metropolitan areas where one major concern with deep excavations is the potentially large ground deformations in and around the excavation resulting damages to the adjacent buildings and utilities, there are many uncertainties corresponding to the calculation of the lateral earth pressure distribution of braced excavation systems. While active and passive earth pressure theory is applicable in simple cases, multilevel braced excavations tend to experience more complex earth pressures.

Yielding walls define earth retaining systems that can either displace or deform or both in the horizontal direction under design loads. For rigid retaining walls, the deformation mode shapes could be translation, rotation, or both (Fig. 8). These structures are assumed to be capable of developing the active or passive earth pressure states within the retained soil. The third type of earth retaining structures is so-called self-yielding rigid walls whose as a result of thermal changes in their surrounding environment displace horizontally on their own as opposed to displacing (or not) as a reaction from earth loads as in the classic cases of yielding and non-yielding earth retaining walls. Examples include various types of circular water- and wastewater-treatment tanks (Horvath 2005).

Dynamic Earth Pressure on Rigid Retaining Structures

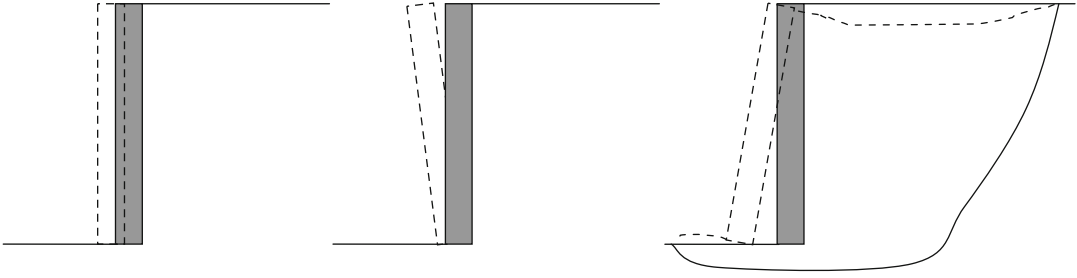
Design of retaining structures to sustain strong earthquakes may further assure the efficient functioning of these structures. Among several aspects to be considered for seismic design of these structures, seismic lateral earth pressure on rigid walls is the important parameter for designing proper reinforcement materials and configuration. Earth pressure theories associated with conventional rigid retaining walls which are based on limit state analysis of yielding walls on the static conditions can be extended to the dynamic states. According to IBC 2006, the natural frequency (in terms of Hz) of a retaining wall with height H can be roughly calculated by

$$f = \frac{20}{H^{0.75}} \quad (1)$$

in which H is the height of the wall in meter. If the natural frequency of wall is considerably larger than the frequency of the input motion applied to the wall, the soil structure system could be assumed to be rigid, and therefore, earth pressure theories, hereafter, are applicable. Methods that are commonly used for design of retaining walls could be classified as force-based methods developed on limit state analysis basis and performance-based approaches.

Force-Based Method

Increasing lateral earth pressures on retaining walls during earthquakes has been one of the major causes of their damage and excessive displacement (Seed and Whitman 1970; Dakoulas and Gazetas 2008). Therefore, correct estimation of the active earth pressure distribution acting on retaining walls during earthquakes is vital for evaluating the safety and designing of the wall. Some simplified approaches have been proposed to calculate the rather complicated dynamic earth pressures on retaining walls during earthquakes (Okabe 1926; Mononobe and Matsuo 1929; Wood 1973;



Analysis and Design Issues of Geotechnical Systems: Rigid Walls, Fig. 8 Deformation mode shapes for yielding rigid retaining walls

Steedman and Zeng 1990; Richards et al. 1999; Choudhury and Singh 2006; Ghosh 2008).

Mononobe-Okabe Approach

Okabe (1926) and Mononobe and Matsuo (1929) extended Coulomb's theory and considered seismic forces by applying earthquake loads as pseudo-static inertial forces to the Coulomb active or passive failure wedge. This pseudo-static method (known as the Mononobe-Okabe method) has been widely used in practical applications (Mylonakis et al. 2007) and is recommended by several building codes and guidelines (e.g., EAU 1996; FHWA 1997; CHBDC 1998; ASCE 4-98 2000; FEMA 369 2000; PIANC 2001; EN 1997 2002; AASHTO 2006; IBC 2006) because of its simplicity in practical applications and reasonable predictions of the actual dynamic pressures acting on walls (Seed and Whitman 1970; Whitman 1990; Ebeling et al. 1992; Veletsos and Younan 1994; Dakoulas and Gazetas 2008).

The dynamic lateral earth pressure on rigid retaining structures proposed in the Mononobe-

Okabe method is obtained from the equilibrium of the active or passive wedges (Fig. 9a and b). In addition to the forces that exist under static conditions of the failure wedge in a dry, cohesionless backfill, the wedge is also subjected to horizontal and vertical pseudo-static forces specified as the mass of the wedge multiplied by pseudo-static accelerations $a_h = k_h g$ and $a_v = k_v g$. The total active and passive thrusts on a rigid wall retaining a dry, cohesionless backfill can be presented in a form similar to what is developed for the static conditions in the Coulomb's theory:

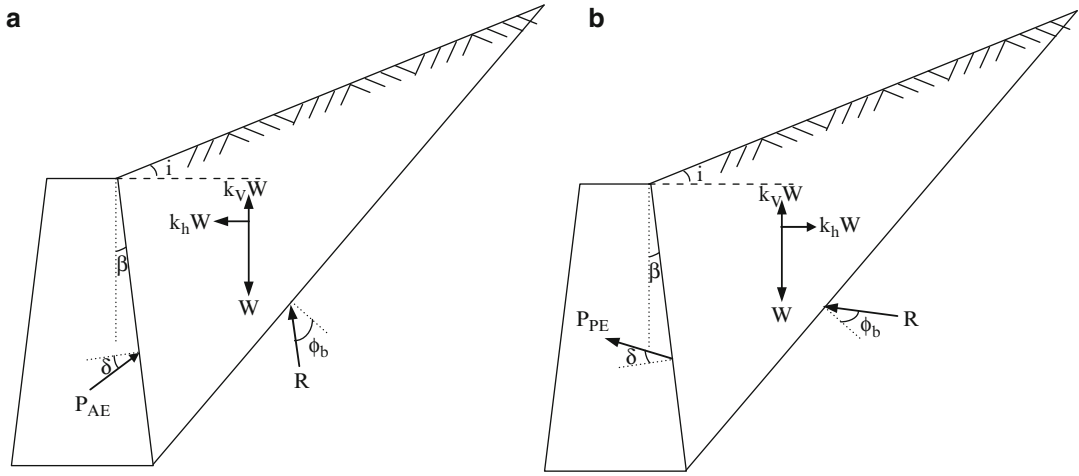
$$P_{AE} = \frac{\gamma_b H^2}{2} (1 - k_v) K_{AE} \quad (2)$$

$$P_{PE} = \frac{\gamma_b H^2}{2} (1 - k_v) K_{PE} \quad (3)$$

in which γ_b is the backfill unit weight, H is the height of the wall, and P_{AE} and P_{PE} are the dynamic active and passive earth pressures. K_{AE} and K_{PE} are the coefficient of dynamic active and passive earth pressures given by:

$$K_{AE} = \frac{\cos^2(\varphi_b - \varphi - \beta)}{\cos \varphi \cos^2 \beta \cos(\delta + \varphi + \beta) \left(1 + \sqrt{\frac{\sin(\varphi_b + \delta) \sin(\varphi_b - \varphi - i)}{\cos(\delta + \beta + \varphi) \cos(i + \beta)}} \right)^2} \quad (4)$$

$$K_{PE} = \frac{\cos^2(\varphi_b - \varphi + \beta)}{\cos \varphi \cos^2 \beta \cos(\delta + \varphi - \beta) \left(1 - \sqrt{\frac{\sin(\varphi_b + \delta) \sin(\varphi_b - \varphi + i)}{\cos(\delta - \beta + \varphi) \cos(i + \beta)}} \right)^2} \quad (5)$$



Analysis and Design Issues of Geotechnical Systems: Rigid Walls, Fig. 9 Mononobe-Okabe approach's failure wedges. (a) Forces acting on active wedge in

Mononobe-Okabe analysis. (b) Forces acting on passive wedge in Mononobe-Okabe analysis

ϕ_b is the internal friction angle of the backfill. i is the slope of the backfill with respect to the horizontal axis. δ is the friction angle between the inner face of the wall and the backfill. β is the angle between the inner face of the wall and the vertical axis. $\varphi = \tan^{-1} \frac{k_h}{1-k_v}$ in which k_h and k_v stand for the horizontal and vertical accelerations of the soil wedge in g unit. The wall is unstable when k_h is higher than $(1-k_v) \tan \varphi$. During the earthquake, k_h and k_v are time varying and chaos so that the seismic earth pressures in both active and passive cases are not constant. The vertical ground excitation has a great influence on dynamic stability of rigid retaining structures during an earthquake. Despite time-varying nature of ground excitations accelerations, the vertical accelerations of earthquakes usually assumed to be a fraction of their horizontal accelerations.

Although the Mononobe-Okabe method suggests that the total dynamic lateral earth pressure on rigid retaining structures should apply at a point $H/3$ above the base of the wall of height, H , experimental results demonstrated that it actually occurs at the higher point under dynamic loading conditions (Fig. 10). The total dynamic lateral earth pressure can be divided into the static active or passive

component, P_A or P_P , and the dynamic contribution, ΔP_{AE} or ΔP_{PE} :

$$P_{AE} = P_A + \Delta P_{AE}, P_{PE} = P_P + \Delta P_{PE} \quad (6)$$

According to the Coulomb theory for linear backfill surfaces with no surcharge loading, the static component acts at a point located $H/3$ above the height of the wall. On the other hand, the dynamic part can be taken to act at a point approximately $0.6H$ above the base of the wall (Seed and Whitman 1970). Therefore, the total lateral earth pressure is applied at a height:

$$h = \frac{P_A (H/3) + \Delta P_{AE} (0.6H)}{P_{AE}} \quad (7)$$

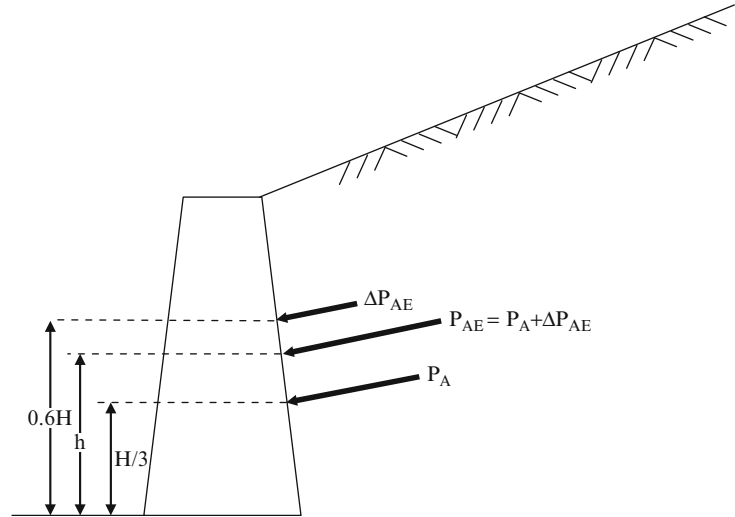
above the base of the wall.

Effects of Water on Dynamic Lateral Pressures

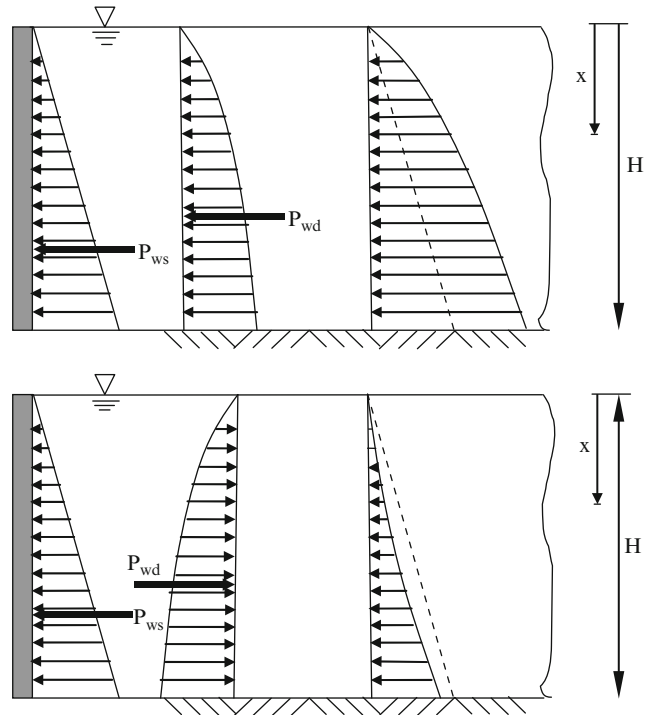
Hydrodynamic Pressure

The retaining walls in practice are usually designed with drainage systems to prevent water from building up. In these cases, the dynamic lateral earth pressure can be

**Analysis and Design
Issues of Geotechnical
Systems: Rigid Walls,
Fig. 10** Acting point of the
total active thrust



**Analysis and Design
Issues of Geotechnical
Systems: Rigid Walls,
Fig. 11** Acting point of the
total active thrust



determined based on soil dry unit weight using the Mononobe-Okabe theory. However, the presence of water in retaining walls in water-front areas and in backfills behind retaining walls can play a significant role in determining seismic loads that act on the wall during and after earthquakes. During any ground

excitations, besides the hydrostatic pressure, p_{ws} , water in front of a retaining wall will exert dynamic pressures, p_{wd} , on the face of the wall (Fig. 11). The hydrostatic pressure is calculated by

$$p_{ws}(x) = \gamma_w x \quad (8)$$

in which γ_w is the water unit weight and x is the distance from the reservoir surface. The total hydrostatic thrust is then determined as

$$P_{ws} = \int_0^H p_{ws}(x) dx = \frac{1}{2} \gamma_w H^2 \quad (9)$$

The application point of the hydrostatic thrust is a point $H/3$ above the base of the wall of height, H .

Hydrodynamic pressure results from the seismic response of the water mass in front of the wall and is usually estimated using Westergaard's theory (Westergaard 1931). Westergaard's theory applies for a vertical rigid wall retaining a very large (theoretically infinite) extent of water basin with no backfill horizontally excited by harmonic motion of its rigid base. The amplitude of hydrodynamic pressure is then determined using Westergaard's theory as

$$p_{wd}(x) = \frac{7}{8} k_h \gamma_w H \sqrt{x/H} \quad (10)$$

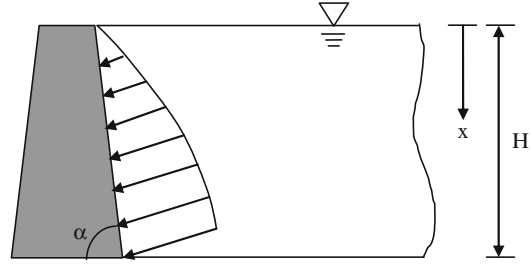
The resultant hydrodynamic force is obtained as

$$P_{wd} = \int_0^H p_{wd}(x) dx = \frac{7}{12} k_h \gamma_w H^2 = (1.17 k_h P_{ws}) \quad (11)$$

According to Westergaard's theory, the hydrodynamic thrust can be taken to act at a point approximately $0.4H$ above the base of the wall. The excess hydrodynamic pressures could be applied toward or outward the wall depending on the base excitation direction. Thus, the direction of the excitation should be taken into account in determining the total hydrodynamic pressure acting on the wall.

Effect of Wall Inclination

The hydrodynamic pressure on the walls with facing inclination (Fig. 12) could be estimated taking the wall inclination using Zangar (1953) and Chwang (1978):



Analysis and Design Issues of Geotechnical Systems: Rigid Walls, Fig. 12 Effect of wall inclination on hydrodynamic pressure

$$p_{wd}(x, \alpha) = C_m(\alpha) k_h \gamma_w H \left[\frac{x}{H} \left(2 - \frac{x}{H} \right) + \sqrt{\frac{x}{H} \left(2 - \frac{x}{H} \right)} \right] \quad (12)$$

or approximately using Westergaard's approach can be estimated by

$$p_{wd}(x, \alpha) = \frac{7}{8} C_m(\alpha) k_h \gamma_w H \sqrt{x/H} \quad (13)$$

in which $C_m(\alpha)$ is the inclination factor given by

$$C_m(\alpha) = 2.0 \frac{\alpha}{\pi} \quad (14)$$

The total hydrodynamic thrust considering the wall inclination acts at a point located $0.4H$ above the base of the wall of height, H with amplitude

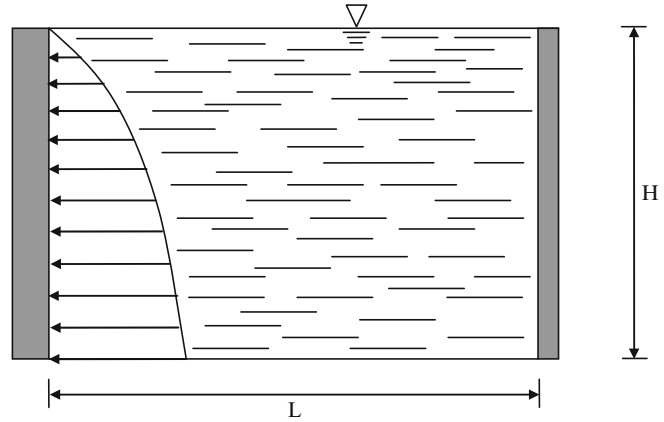
$$P_{wd}(\alpha) = \frac{7}{12} C_m(\alpha) k_h \gamma_w H^2 = (1.17 C_m(\alpha) k_h P_{ws}) \quad (15)$$

Effect of Water Basin Length

Westergaard's theory was developed based on this assumption that the rigid wall is retaining a semi-infinite reservoir of water. For the case a finite water basin (Fig. 13), the hydrodynamic pressure applied to the wall can be evaluated using Werner and Sundquist's (1943) suggestion by intruding the water basin length modification factor, C_n , as

$$p_{wd}(x) = \frac{7}{8} C_n k_h \gamma_w H \sqrt{x/H} \quad (16)$$

**Analysis and Design
Issues of Geotechnical
Systems: Rigid Walls,
Fig. 13** Wall retaining
finite water basin



A

where C_n is defined in terms of ratio of water basin length, L , over the wall height, H , as

$$C_n = \frac{4}{3} \frac{L/H}{1 + L/H} \leq 1.0 \quad (17)$$

$$C_n = 1.0 \quad L/H \geq 2.7$$

The total hydrodynamic thrust considering the water basin length effect acts at a point located $0.4H$ above the base of the wall of height, H with amplitude

$$P_{wd} = \frac{7}{12} C_n k_h \gamma_w H^2 = (1.17 C_n k_h P_{ws}) \quad (18)$$

Dynamic Pressure for Saturated Backfill

Seismic response of rigid walls retaining saturated backfills could be affected by the presence of water changing the inertial forces within the backfill, developing the hydrodynamic pressures with the saturated soil and generating excess pore water pressure resulting in the cyclic deformation of the backfill soil (Matsuo and Ohara 1965). The soil permeability plays an important role in developing hydrodynamic pressures within the backfill soil. For soils with small permeability, the pore water moves within the soil in restrained conditions during the ground shaking; while if the permeability of the backfill is very high, the soil particles move easily through the pore water, and the pore water may remain stationary. Hydrodynamic water

pressures resulting from the seismic response of the water in the backfill soils could be estimated by Westergaard's theory modifying the effect of the soil permeability as

$$P_{wd}(x) = \frac{7}{8} C_e k_h \gamma_w H \sqrt{x/H} \quad (19)$$

in which C_e is the correction factor that expresses the portion the pore water which vibrates freely within the soil:

$$C_e = 0.5 - 0.5 \tan h \left[\log \frac{2\pi n \gamma_w H^2}{7E_w K T} \right] \quad (20)$$

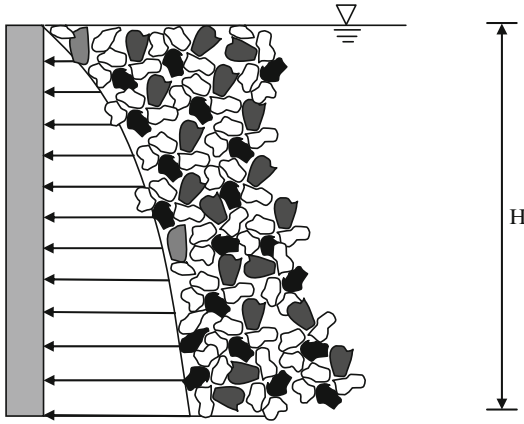
where n is the porosity, H is the water depth, E_w is the bulk modulus of water (2×10^6 kPa), K is the backfill permeability, and T is the predominant period of the excitation (Fig. 14).

The total hydrodynamic thrust considering the water within the backfill acts at a point located $0.4H$ above the base of the wall of height, H with amplitude

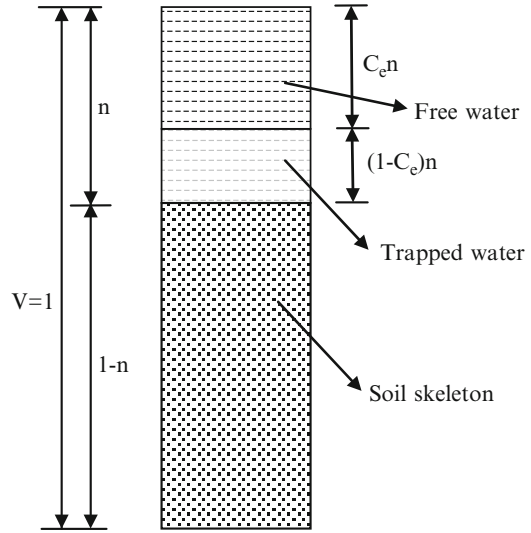
$$P_{wd} = \frac{7}{12} C_e k_h \gamma_w H^2 = (1.17 C_e k_h P_{ws}) \quad (21)$$

Earth Lateral Pressures on Walls Retaining Saturated Backfills

For walls retaining saturated soils with restrained pore water conditions, to take into account the presence of water within the backfill, the Mononobe-Okabe approach is modified using



Analysis and Design Issues of Geotechnical Systems: Rigid Walls, Fig. 14 Wall retaining saturated backfill



Analysis and Design Issues of Geotechnical Systems: Rigid Walls, Fig. 15 Physical analogy proposed by Matsuzawa et al. (1985)

physical analogy proposed by Matsuzawa et al. (1985). The dynamic lateral pressures act on a yielding wall are coming from the soil skeleton, the trapped water which vibrates together with the soil skeleton and free water which vibrates independently from the soil skeleton (Fig. 15). The total active static soil thrust acts at a point located $H/3$ above the height of the wall can be evaluated as

$$P_A = \frac{1}{2} k_A (\gamma_{\text{sat}} - \gamma_w) H^2 \quad (22)$$

Besides the hydrostatic force exerted on the wall, the hydrodynamic pressures considering the wall facing inclination, the water basin length, and the saturated backfill resulting from the seismic response of the water in the backfill soil could be estimated as

$$p_{\text{wd}}(x) = \frac{7}{8} C_n C_m C_e k_h \gamma_w H \sqrt{x/H} \quad (23)$$

and the total hydrodynamic thrust acts at a point located $0.4H$ above the base of the wall (Fig. 16) and is determined by

$$P_{\text{wd}} = \frac{7}{12} C_n C_m C_e k_h \gamma_w H^2 \quad (24)$$

$$H^2 = (1.17 C_n C_m C_e k_h P_{\text{ws}})$$

If the total dynamic earth pressure coefficient, K_{AE} , is written as

$$K_{\text{AE}} = K_A + \Delta K_{\text{AE}}$$

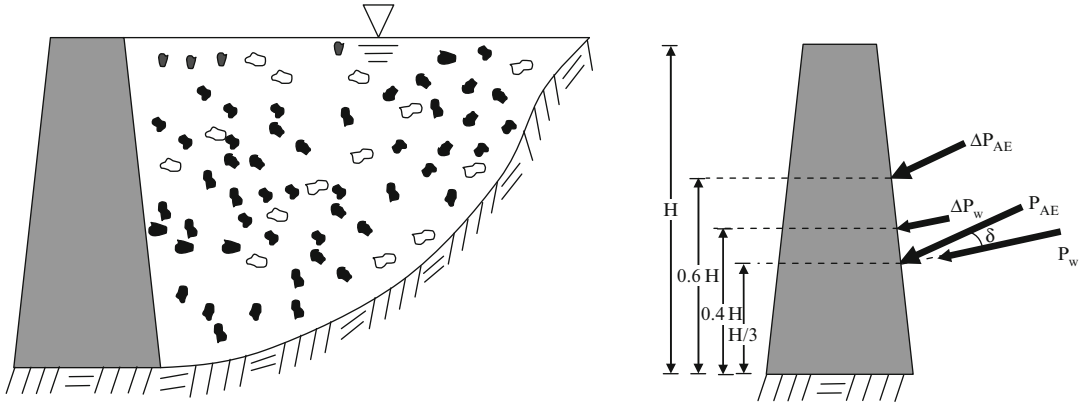
and using an approximation for ΔK_{AE} equal to $(3/4)k_h$ suggested by Seed and Whitman, therefore, the dynamic lateral force component for a rigid wall retaining dry backfill soil would be

$$\Delta P_{\text{AE}} = \frac{1}{2} \gamma H^2 \Delta K_{\text{AE}} = \frac{1}{2} \left(\frac{3}{4} k_h \right) \gamma H^2 \quad (25)$$

For saturated backfill soil, the dynamic component of the earth thrust acting on the wall (Fig. 16), employing $\gamma^* = C_e \gamma_{\text{dry}} + (1 - C_e) \gamma_{\text{sat}}$, should be computed as

$$\Delta P_{\text{AE}} = \frac{1}{2} \left(\frac{3}{4} k_h \right) \gamma^* H^2 \quad (26)$$

P_A computation requires $\gamma_{\text{sat}} - \gamma_w$, while γ^* should be used in determining ΔP_{AE} . However, sometimes it is necessary to compute



Analysis and Design Issues of Geotechnical Systems: Rigid Walls, Fig. 16 Different components of acting thrust on the wall

P_A and ΔP_{AE} with a common unit weight (e.g., EAK 2002). In this case, the buoyant unit weight, $\gamma_{sat} - \gamma_w$, and a modified seismic coefficient as

$$k^* = k \frac{\gamma^*}{\gamma_{sat} - \gamma_w} \quad (27)$$

For impermeable soils such as clayey sand, clayey silts, and clayey gravels, $C_e = 0$, and therefore, should be used $\gamma^* = \gamma_{sat}$ resulting $P_{wd} = 0$ and

$$\Delta P_{AE} = \frac{1}{2} \left(\frac{3}{4} k_h \right) \gamma_{sat} H^2 \quad (28)$$

or

$$\Delta P_{AE} = \frac{3}{8} \left(k_h \frac{\gamma_{sat}}{\gamma_{sat} - \gamma_w} \right) (\gamma_{sat} - \gamma_w) H^2 \quad (29)$$

On the other hand, for permeable soils such as sands, gravels, and cobbles, $C_e = 1$, and therefore, $\gamma^* = \gamma_{dry}$ resulting

$$P_w = \frac{1}{2} \gamma_w H^2, P_{wd} \neq 0$$

$$\Delta P_{AE} = \frac{1}{2} \left(\frac{3}{4} k_h \right) \gamma_{dry} H^2 \quad (30)$$

or

$$\Delta P_{AE} = \frac{3}{8} \left(k_h \frac{\gamma_{dry}}{\gamma_{sat} - \gamma_w} \right) (\gamma_{sat} - \gamma_w) H^2 \quad (31)$$

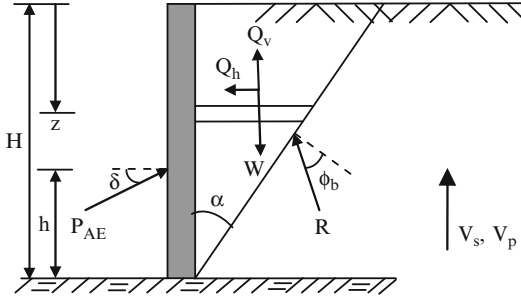
In other approach, by representing the excess pore water pressure in the backfill using the pore pressure ratio, r_u , the active earth pressure acting on wall with cohesionless backfill can be computed using

$$p_{AE} = K_{AE} \gamma_{sub} (1 - r_u) (1 - k_v)$$

$$\phi = \tan^{-1} \left[\frac{\gamma_{sat} k_h}{\gamma_{sub} (1 - r_u) (1 - k_v)} \right] \quad (32)$$

Steedman-Zeng Method

The dynamic nature of earthquake loading in M-O method is considered in a very approximate way without taking any effect of time. To reach the dynamic response characteristics of rigid walls during an earthquake taking into account the phase difference due to finite shear wave propagation within the backfill soil, Steedman and Zeng (1990) proposed a simple pseudo-dynamic approach to determine the seismic response of fixed-base cantilever walls subjected to a harmonic horizontal acceleration considering finite shear wave velocity within backfill material. Choudhury and Nimbalkar (2006) extended the Steedman-Zeng approach to take into account the wall friction angle, soil



Analysis and Design Issues of Geotechnical Systems: Rigid Walls, Fig. 17 Retaining wall considered for computation of pseudo-dynamic active earth pressure

friction angle, shear wave velocity (V_s), primary wave velocity (V_p), and both the horizontal and vertical seismic accelerations ($k_h g$ and $k_v g$) on the seismic active earth pressure behind the wall (Fig. 17).

If the retaining wall is subjected to both the horizontal and vertical seismic accelerations ($k_h g$ and $k_v g$), the accelerations at any depth z and time t , below the top of the wall, can be expressed as

$$\begin{aligned}\ddot{u}(z, t) &= k_h g \sin \left[\omega \left(t - \frac{H - z}{V_s} \right) \right] \\ \ddot{v}(z, t) &= k_v g \sin \left[\omega \left(t - \frac{H - z}{V_p} \right) \right]\end{aligned}\quad (33)$$

The mass of a thin element of wedge at depth z is

$$m(z) = \gamma_b \frac{H - z}{g \tan \alpha} dz \quad (34)$$

The total horizontal and vertical inertial forces acting within the failure zone can be expressed as

$$\begin{aligned}F_h(t) &= \int_0^H m(z) \ddot{u}(z, t) dz \\ &= \frac{\lambda \gamma_b k_h}{4\pi^2 \tan \alpha} [2\pi H \cos \omega \zeta + \lambda (\sin \omega \zeta - \sin \omega t)]\end{aligned}$$

$$\begin{aligned}F_v(t) &= \int_0^H m(z) \ddot{v}(z, t) dz \\ &= \frac{\eta \gamma_b k_v}{4\pi^2 \tan \alpha} [2\pi H \cos \omega \psi + \eta (\sin \omega \psi - \sin \omega t)]\end{aligned}\quad (35)$$

where $\lambda = 2\pi V_s / \omega$ is the wavelength of the vertically propagating shear wave, $\zeta = t - H/V_s$, $\eta = 2\pi V_p / \omega$ is the wavelength of the vertically propagating primary wave, and $\psi = t - H/V_p$. For rigid wedge, the shear wave velocity and the primary wave velocity tend to infinite, and total horizontal and vertical inertial forces will be equivalent to the pseudo-static force assumed in the Mononobe-Okabe method. The total (static + seismic) active thrust, $P_{AE}(t)$, can be obtained by resolving the forces on the wedge and considering the equilibrium of the forces, and hence $P_{AE}(t)$ can be expressed as follows:

$$P_{AE}(t) = \frac{F_h(t) \cos(\alpha - \phi_b) + W \sin(\alpha - \phi_b) - F_v(t) \sin(\alpha - \phi_b)}{\cos(\delta + \phi_b - \alpha)} \quad (36)$$

The seismic active earth pressure coefficient, K_{AE} , is defined as

$$K_{AE} = \frac{2P_{AE}}{\gamma_b H^2} \quad (37)$$

Substituting for F_h and F_v , K_{AE} is derived as

$$\begin{aligned}K_{AE} &= \frac{1}{\tan \alpha} \frac{\sin(\alpha - \phi_b)}{\cos(\delta + \phi_b - \alpha)} \\ &+ \frac{k_h}{2\pi^2 \tan \alpha} \left(\frac{2\pi V_s}{\omega H} \right) \frac{\cos(\alpha - \phi_b)}{\cos(\delta + \phi_b - \alpha)} m_1 \\ &- \frac{k_v}{2\pi^2 \tan \alpha} \left(\frac{2\pi V_p}{\omega H} \right) \frac{\sin(\alpha - \phi_b)}{\cos(\delta + \phi_b - \alpha)} m_2\end{aligned}\quad (38)$$

where

$$m_1 = \left[2\pi \cos 2\pi \left(\frac{t\omega}{2\pi} - \frac{H\omega}{2\pi V_s} \right) + \left(\frac{2\pi V_s}{H\omega} \right) \left(\sin 2\pi \left(\frac{t\omega}{2\pi} - \frac{H\omega}{2\pi V_s} \right) - \sin 2\pi \left(\frac{t\omega}{2\pi} \right) \right) \right]$$

$$m_2 = \left[2\pi \cos 2\pi \left(\frac{t\omega}{2\pi} - \frac{H\omega}{2\pi V_p} \right) + \left(\frac{2\pi V_p}{H\omega} \right) \left(\sin 2\pi \left(\frac{t\omega}{2\pi} - \frac{H\omega}{2\pi V_p} \right) - \sin 2\pi \left(\frac{t\omega}{2\pi} \right) \right) \right]$$

(39)

From the above equation, it can be noted that K_{AE} is function of the dimensionless parameters $\frac{H\omega}{2\pi V_s}$, $\frac{H\omega}{2\pi V_p}$, $\frac{t\omega}{2\pi}$, and the wedge angle, α . The maximum value of K_{AE} is obtained by optimizing K_{AE} with respect to $\frac{t\omega}{2\pi}$ and α . This parameter would be a function of $\frac{H\omega}{2\pi V_s}$ and $\frac{H\omega}{2\pi V_p}$ which is the ratio of time for shear wave and primary wave to travel the full height of the wall to the period of lateral shaking.

The seismic active earth pressure distribution can be obtained by differentiating the total active thrust as

$$P_{AE}(t) = \frac{\partial P_{AE}(t)}{\partial z} = \frac{\gamma z}{\tan \alpha \cos(\delta + \phi_b - \alpha)} \frac{\sin(\alpha - \phi_b)}{\cos(\delta + \phi_b - \alpha)} + \frac{k_h \gamma z}{\tan \alpha \cos(\delta + \phi_b - \alpha)} \sin \left[\omega \left(t - \frac{z}{V_s} \right) \right] - \frac{k_v \gamma z}{\tan \alpha \cos(\delta + \phi_b - \alpha)} \sin \left[\omega \left(t - \frac{z}{V_p} \right) \right]$$

(40)

Design earth pressures determined by the Steedman-Zeng method or in general by Choudhury-Nimbalkar that account for backfill amplification should be considered for design of unusually displacement-sensitive walls such as tall retaining gravity walls.

Performance-Based Design

The observations after severe earthquakes have shown that retaining walls could fail during earthquakes by sliding away from the backfill or due to combined action of sliding and rocking displacements, while the post-event serviceability of such structures is also related to the wall displacements induced by the earthquake. Accordingly, any approaches that predict post-earthquake wall displacements should provide a more indication of retaining walls performances. Therefore, an alternative design approach, so-called performance-based design, could be used on the basis of permanent displacements for the wall when selecting an allowable displacement for design.

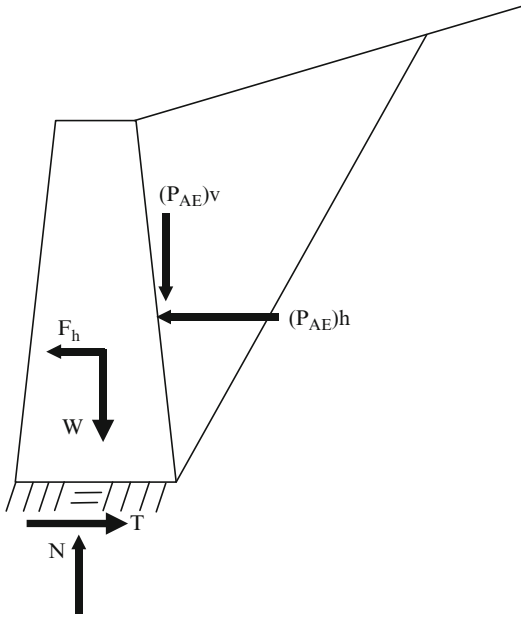
Performance-based design of retaining walls must account for the likely displacements the retaining wall may experience during an earthquake in addition to calculating the usual factors of safety against failure in bearing capacity, sliding, and overturning. There are several procedures available such as Richard and Elms (1979), Whitman and Liao (1985), and Wu and Prakash (2001) to estimate the permanent displacements of rigid walls.

In the Richards-Elms method, only the displacements of the sliding modes are taken into account. Considering the gravity rigid wall shown in Fig. 18, when the active backfill edge is subjected to ground acceleration toward the backfill, the resulting inertial forces will act away from the backfill. Therefore, in the limit state of sliding motion where the level of ground acceleration, so-called yield acceleration, is large enough to cause the wall to slide on its base, the horizontal and vertical equilibrium equations are

$$T = F_h + (P_{AE})_h$$

$$N = W + (P_{AE})_v$$

(41)



Analysis and Design Issues of Geotechnical Systems: Rigid Walls, Fig. 18 Gravity retaining wall considered in Richard-Elms method

Substituting $T = N \tan \phi_b$, $F_h = a_y W/g$, $(P_{AE})_h = P_{AE} \cos(\delta + \beta)$, and $(P_{AE})_v = P_{AE} \sin(\delta + \beta)$, the yield acceleration can be computed by

$$a_y = \left[\tan \phi_b - \frac{P_{AE} \cos(\delta + \beta) - P_{AE} \sin(\delta + \beta)}{W} \right] g \quad (42)$$

In calculating P_{AE} , since the M-O method requires that a_y be known, the solution of the above equation must be obtained iteratively. In the Richard-Elms, based on the sliding block analysis, the permanent block displacement is determined as

$$\delta = 0.087 \frac{V_{\max}^2}{\alpha_{\max}} \left(\frac{a_y}{a_{\max}} \right)^{-4} \quad (43)$$

where $V_{\max}$ is the peak ground velocity and $a_{\max} = k_h g$ is the horizontal peak ground acceleration.

A realistic model for estimating the dynamic displacement must account for the combined action of possible displacement modes including sliding and rocking vibrations and considering (1) soil stiffness in sliding and rocking, (2) geometrical and material damping in sliding and rocking, and (3) nonlinear coupling effects for stiffness and damping. Wu and Prakash (2001) proposed a model for simulating the response of rigid retaining walls subjected to seismic loading. This model (Fig. 19) consisted of a rigid wall resting on the foundation soil and subjected to a horizontal ground motion and analyzed the problem as a case of combined sliding and rocking vibrations including the effect of various important parameters such as soil stiffness in sliding, soil stiffness in rocking, geometrical damping in sliding, geometrical damping in rocking, material damping in sliding, and material damping in rocking.

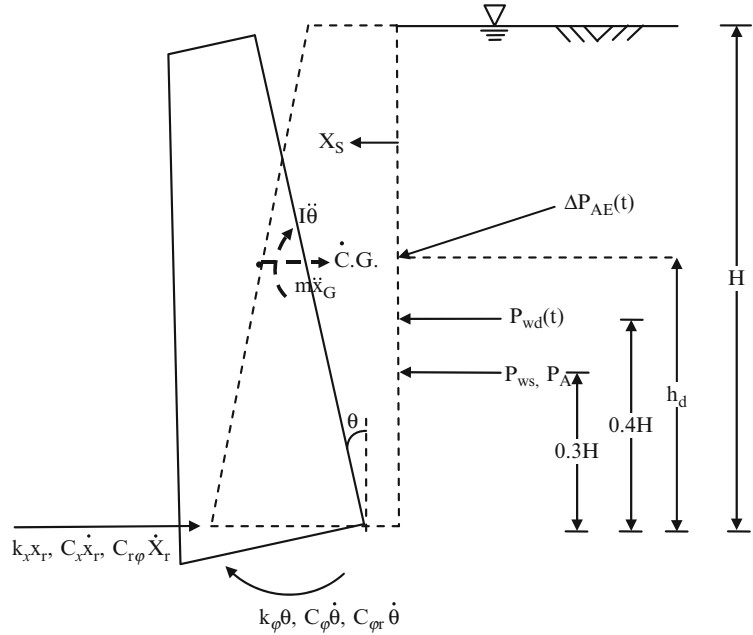
An important step in the performance design procedure is the selection of the permissible displacement. There are some guidelines given available based on experience or judgment (Huang 2005). Eurocode (1994) gives the permissible horizontal displacement equal to $300 a_{\max}$ (mm) where $a_{\max}$ is the maximum horizontal design acceleration, while AASHTO (2006) suggests this parameter to be limited to $250 a_{\max}$ (mm). Wu and Prakash (1996) proposed that the permissible horizontal displacement should be 2 % of the height of the retaining wall, H , while the failure horizontal displacement is equal to $0.1H$. According to the Japanese Railway Technical Research Institute (JRTRI 1999), permissible differential settlement = $0.1\text{--}0.2$ m (damage needing minor retrofit measures).

Severe differential settlement = >0.2 m (damage needing long-term retrofit measures).

It may be noted that Eurocode 8 (1994) and Wu and Prakash (1996) recommend using specified horizontal displacements of the retaining wall for evaluating its seismic performance, while JRTRI suggests the use of vertical differential settlement as the performance criterion which seems reasonable for traffic accessibility and retrofit purposes after the earthquake.

Analysis and Design Issues of Geotechnical Systems: Rigid Walls,

Fig. 19 Free body diagram of forced vibration of a gravity retaining wall retaining saturated backfill soil



Any procedure in the performance-based design of retaining walls should follow these steps:

- Assume a permissible displacement, δ_{per}
- Determine the yield acceleration that corresponds the permissible displacement as

$$a_y = \left(0.087 \frac{V_{max}^2 \alpha_{max}^3}{\delta_{per}} \right)^{1/4} \quad (44)$$

- Having the yield acceleration determined in the previous step, calculate P_{AE} using the M-O method.
- Calculate the wall weight required to limit the wall displacement to the permissible displacement taken in the first step as

$$W = \left[\frac{P_{AE} \cos(\delta + \beta) - P_{AE} \sin(\delta + \beta) \tan \phi_b}{\tan \phi_b - a_y/g} \right] \quad (45)$$

A factor of safety 1.2–1.5 should be applied to the weight of the wall to take into account the probability of exceedance in permissible

displacements of the wall, the importance of the wall, the effects of failure, and the cost of repair.

Cross-References

- [Bridge Foundations](#)
- [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)
- [Seismic Design of Earth-Retaining Structures](#)

References

- AASHTO (2006) Recommended LRFD guidelines for the seismic design of highway bridges. American Association of State Highway and Transportation Officials (AASHTO), National Cooperative Highway Research Program (NCHRP), Transportation Research Board
- ASCE 4–98 (2000) Seismic analysis of safety related nuclear structures and commentary. American Society of Civil Engineers
- CHBDC (1998) Canadian highway bridge design code. Canadian Standards Association, Rexdale, Ontario
- Choudhury D, Nimbalkar SS (2006) Pseudo-dynamic approach of seismic active earth pressure behind retaining wall. *J Geotechn Geol Eng* 24:1103–1113
- Choudhury D, Singh S (2006) New approach for estimation of static and seismic active earth pressure. *Geotech Geol Eng* 24(1):117–127

- Choukeir M, Juran I, Hanna S (1997) Seismic design of reinforced-earth and soil-nailed structures. *Ground Improv* 1:223–238
- Chwang AT (1978) Hydrodynamic pressures on sloping dams during earthquakes. Part 2. Exact theory. *J Fluid Mech* 87:343–348
- Dakoulas P, Gazetas G (2008) Insight into seismic earth and water pressures against caisson quay walls. *Geophys J Roy Astron Soc* 58(2):95–111
- EAK (2002) Empfehlungen des Arbeitsausschusses Kustenschutzwerke. Die Kuste, H. 65
- EAU (1996) Recommendations of the committee for waterfront structures, harbours, and waterways, 7th edn. Ernst & Sohn, Berlin
- Ebeling RM, Morrison EE, Whitman RV, Liam Finn WD (1992) A manual for seismic design of waterfront retaining structures. US Army Corps of Engineers, Technical report ITL-92-11
- EN 1997 (2002) Eurocode 7 geotechnical design – Part 1: General rules. CEN European committee for standardization, Bruxelles
- EUROCODE 8 (EUROPEAN PRE-STANDARD) (1994) Design provisions for earthquake resistance of structures-Part 5: Foundations, retaining structures and geotechnical aspects. The Commission of the European Communities
- FEMA 369 (2000) The 2000 NEHRP recommended provisions for new buildings and other structures. Part 2: Commentary. Federal Emergency Management Agency
- FHWA (1997) Geotechnical engineering circular #3. Design guidance: geotechnical earthquake engineering for highways, design principles. FHWA-SA-97-076, Federal Highway Administration, vol I, U.S. Department of Transportation
- Ghosh P (2008) Seismic active earth pressure behind a non-vertical retaining wall using pseudo-dynamic analysis. *Can Geotech J* 45(1):117–123
- Horvath JS (2005) Integral-abutment bridges: geotechnical problems and solutions using geosynthetics and ground improvement. In: Proceeding IAJB 2005: the 2005 FHWA conference on integral abutment and jointless bridges, Baltimore, pp 281–291
- Huang C (2005) Seismic displacement of soil retaining walls situated on slope. *J Geotech Geoenviron Eng* 131(9):1108–1117
- International Code Council (2006) IBC (International Building Code). International Code Council, Country Club Hills, 664 pp
- Japan Railway Technical Research Institute (JRTRI) (1999) Design guidelines for railway structures-aseismic design. Maruzen Co. Ltd
- Kramer S (1996) Geotechnical earthquake engineering. Prentice-Hall, Upper Saddle River
- Matsuo H, Ohara S (1965) Dynamic pore water pressure acting on quay walls during earthquakes. In: Proceedings of the third world conference on earthquake engineering, New Zealand, vol 1, pp 130–140
- Matsuzawa H, Ishibashi I, Kawamura M (1985) Dynamic soil and water pressures of submerged soils. *J Geotech Eng ASCE* 111(10):1161–1176
- Mononobe N, Matsuo H (1929) On the determination of earth pressures during earthquakes. In: Proceedings of world engineering congress
- Mylonakis G, Kloukinas P, Papantonopoulos C (2007) An alternative to the Mononobe-Okabe equations for seismic earth pressures. *Soil Dyn Earthq Eng* 27:957–969
- Okabe S (1926) General theory of earth pressure. *Journal Japan Society of Civil Engineering*, 12(1)
- PIANC (2001) Seismic design guidelines for port structures. A. A. Balkema, Tokyo (International Navigation Institute)
- Richard R, Elms DG (1979) Seismic behavior of gravity retaining walls. *J Geotech Eng, ASCE* 105:449–464
- Richards R, Huang C, Fishman KL (1999) Seismic earth pressure on retaining structures. *J Geotech Geoenviron Eng, ASCE* 125(9):771–778
- Seed HB, Whitman RV (1970) Design of earth retaining structures for dynamic loads. In: Proceedings, ASCE Specialty Conference on Lateral Stresses in the Ground and Design of Earth Retaining Structures, pp 103–147
- Steedman RS, Zeng X (1990) The influence of phase on the calculation of pseudo-static earth pressure on a retaining wall. *Geotechnique* 40:103–112
- Tomlinson MJ (1995) Foundation design and construction. Wiley, New York
- Veletsos AS, Younan AH (1994) Dynamic soil pressures on rigid retaining walls. *Earthq Eng Struct Dyn* 23:275–301
- Werner PW, Sundquist KJ (1943) On hydrodynamic earthquake effects. *Trans Am Geophys Union* 30(5):636–657
- Westergaard H (1931) Water pressure on dams during earthquakes. In: Transportations of ASCE, Paper no. 1835, pp 418–433
- Whitman RV (1990) Seismic design and behavior of gravity retaining walls. In: Proceedings conference on design and performance of earth retaining structures, ASCE geotechnical special publication, vol 25, pp 817–842
- Whitman RV, Liao S (1985) Seismic design of retaining walls, paper GL-85-1, U.S. Army engineer waterways experiment station, Vicksburg
- Wood J (1973) Earthquake-induced soil pressures on structures. Report EERL 73–05, California Institute of Technology, Pasadena
- Wu Y, Prakash S (1996) On Seismic Displacement of Rigid retaining Walls. In: Prakash S (ed) ASCE Geotechnical special publication: Analysis and Design of Retaining Structures Against Earthquakes, SP editor, pp 21–37
- Wu Y, Prakash S (2001) Seismic displacements of rigid retaining walls. In: State-of-the-art paper number 705, proceedings fourth international conference on recent advances in geotechnical earthquake engineering and soil dynamics, San Diego
- Zangar CN (1953) Hydrodynamic pressures on dams due to horizontal earthquakes. *Proc Soc Exp Stress Anal* 10:93–102

Analytic Fragility and Limit States [P(EDP|IM)]: Nonlinear Dynamic Procedures

Dimitrios Vamvatsikos

School of Civil Engineering, National Technical University of Athens (N.T.U.A.), Athens, Greece

Introduction

The determination of structural response given the seismic loading is of paramount importance in earthquake engineering. Due to the random nature of earthquakes, it is widely accepted that this is best done in probabilistic rather than deterministic terms. Thus, modern frameworks for performance-based earthquake engineering are based on the evaluation of the distribution of structural response, characterized by one or more engineering demand parameters (EDPs, e.g., peak interstory drift ratio or peak floor acceleration), given the level of a (typically scalar) intensity measure (IM, e.g., peak ground acceleration or first-mode spectral acceleration), used to characterize the earthquake loading (Cornell and Krawinkler 2000). This distribution is symbolized by the corresponding probability distribution function (PDF) of EDP given the IM: $P(EDP|IM)$.

Such results are often employed in tandem with one or more distinct limit or damage states that characterize the performance (or state) of the structure, such as immediate occupancy, life safety, or near collapse. Then, the results of structural analysis can be distilled into the so-called building-level fragility functions associated with the violation of each limit state. Formally, a (building-level) fragility function is the cumulative distribution function of the seismic intensity, in terms of the IM, needed to violate the limit state. A limit state is usually tied to specific threshold capacity (EDP_c) values of one or more EDPs that can be either deterministic (i.e., assumed to be perfectly known) or probabilistic, the latter obviously being the more realistic option. When such values are exceeded, the

limit state is deemed to have been violated. In the case where a limit state is defined via a *single* EDP, e.g., exceedance of 3 % peak interstory drift to define life safety (or worse), the fragility function is essentially the same as the probability function of the seismic demand EDP exceeding the capacity EDP_c given the IM, symbolized as $P(EDP_c < EDP|IM)$.

Fragility can be determined in a variety of ways, using, for example, empirical data from post-earthquake surveys or even expert opinion. For the vast majority of cases, though, computer-intensive analytical options are the most suitable choice. In this direction, recent advances in computer technology have allowed the consideration of nonlinear dynamic analysis as a realistic option. In the following, a number of approaches will be discussed on how to assess the distribution of EDP given IM and determine fragility functions using nonlinear dynamic analysis.

Intensity Measure and Ground Motions

The importance of selecting an efficient and sufficient IM (Luco and Cornell 2007) cannot be understated. An efficient IM will generally be well correlated with the EDPs of choice, thus showing low dispersion of demand given the IM, and subsequently allow the determination of EDP demand or IM capacity statistics using a relatively low number of ground motion records. Sufficiency is defined as the independence of the distribution of EDP given the IM from any other seismological parameters that may characterize the ground motion, e.g., duration, magnitude, spectral shape, or the presence of a pulse indicative of near-source forward directivity. A sufficient IM essentially captures all seismological information needed to determine the effect of a ground motion record on the structure being investigated (see also Jalayer et al. 2012). In other words, it frees the analyst from having to be careful regarding the selection (sampling) of ground motion records, as one only needs to match the required level of IM and no other seismological parameter.

A sufficient IM is essentially the premise of scaling, where the acceleration values of a ground motion time history are uniformly multiplied by a constant to reach the intensity needed. Still, it can be safely assumed that no IM for which a seismic hazard curve can be practically estimated (i.e., those for which ground motion prediction equations currently exist) will ever be 100 % sufficient for all sites and structures. Thus, excessive scaling may introduce biased estimates of response (Luco and Bazzurro 2007). Still, given the limitations of the catalogue of recorded ground motion records, scaling is often the only method to be used to reach truly high IM values that can force a modern well-designed structure to experience global collapse.

A standard IM choice is the 5 % damped first-mode (pseudo) spectral acceleration $S_a(T_1)$ (Shome et al. 1998; Shome and Cornell 1999). This is generally adequate for first-mode-dominated structures that do not displace far into the nonlinear region, as is the case of most existing brittle or moderately ductile low-/mid-rise buildings. For structures where higher modes become important or modern buildings that exhibit significant ductility, improved IM alternatives should be sought. One particularly attractive option is using $S_{agm}(T_i)$ geometric mean of spectral acceleration values at several periods T_i (Cordova et al. 2000; Vamvatsikos and Cornell 2005; Bianchini et al. 2009) that can largely alleviate the effect of spectral shape, being able to capture both the period elongation characterizing ductile structures and the effect of higher modes. Another option that can offer similar, if not better, sufficiency is the compound IM proposed by Luco and Cornell (2007) based on the combination of inelastic spectral displacement at the first mode and the elastic spectral displacement at the second mode of the structure. Its only disadvantage is the need for using specialized ground motion prediction equations to run seismic hazard analysis, developed according to Tothong and Cornell (2006).

Whenever an IM is not sufficient enough to remove the influence of other seismological parameters, their distribution at each IM level should be properly accounted for by

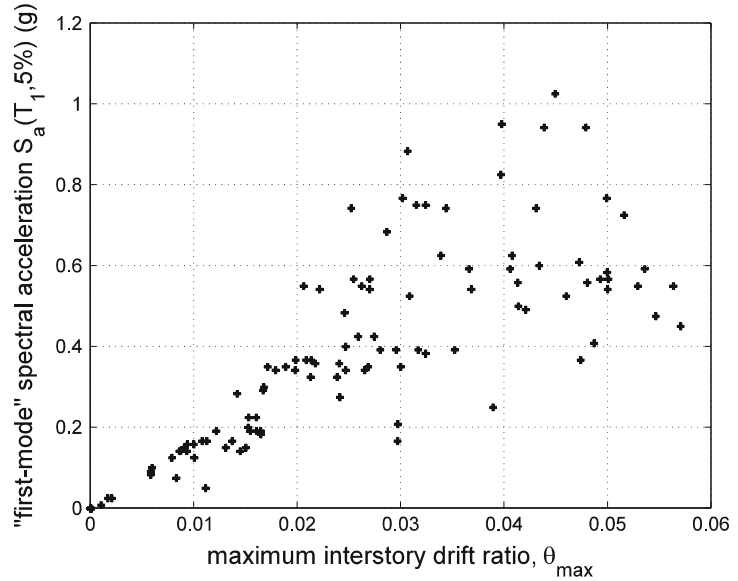
appropriately selecting ground motion records. Thus, for example, if spectral shape, duration, or any other feature that cannot be accounted for by the IM is deemed to be important, the ground motion records used at different intensities should be chosen to reflect the anticipated distribution of those features given the IM. A prime example of a selection method is the conditional spectrum (CS) by Lin et al. (2013a, b), whereby the distribution of spectral shape, characterized by the parameter epsilon (Baker and Cornell 2006), is taken into account. Furthermore, Bradley (2010) has proposed the generalized conditional intensity measure (GCIM) approach to also incorporate other significant characteristics such as the duration and number of significant cycles of the ground motion. For additional information, see also Iervolino and Cornell (2005) and the comprehensive review of record selection provided by Katsanos et al. (2010).

While undoubtedly useful, ground motion selection is not a panacea. It is a method that is site and structure dependent, thus heavily encumbering or even precluding its use whenever a building portfolio or an entire building class (e.g., mid-rise steel frames of the Western USA) is to be assessed to extract fragility or vulnerability functions. Additionally, the boundaries of the recorded ground motions' catalogue put a limit on what can be actually represented by natural records. Thus, it is often the case that a combination of selection and scaling needs to be employed. Alternatively, one could also employ the results of CS or GCIM methods to create artificial ground motions that display the required characteristics. Still, this is an option that is presently available only for highly proficient analysts, as appropriate algorithms have not appeared in the literature so far.

Alternatively, a good rule for a general use is to utilize strong records that when unscaled can still damage the investigated structure, together with a relatively sufficient IM to allow scaling within reason. For most existing low-/mid-rise structures having low-to-moderate ductility that are not subject to near-field motions, this generally means employing $S_a(T_1)$ together with records having naturally high $S_a(T_1)$ values.

Analytic Fragility and Limit States [P(EDP|IM)]: Nonlinear Dynamic Procedures, Fig. 1

The resulting points of a cloud approach for a 9-story steel frame using $S_d(T_1)$ as the IM and the maximum over all stories peak interstory drift $\theta_{\max}$ as the EDP



For modern ductile structures or non-ductile buildings that are not first-mode dominated (e.g., plan-asymmetric or tall structures) or whenever the site of interest is subject to near-field motions, a strong set of records together with an improved IM (see earlier discussion) should be preferred.

Analysis Strategies

Determination of fragility necessitates a wide-range assessment of structural response at multiple levels of intensity. There are several ways to organize the execution of nonlinear dynamic analyses, mainly differing in the manner of postprocessing and in how they can employ ground motion selection and scaling. The main candidates are three (Jalayer and Cornell 2009): cloud analysis, stripe analysis, and incremental dynamic analysis (IDA).

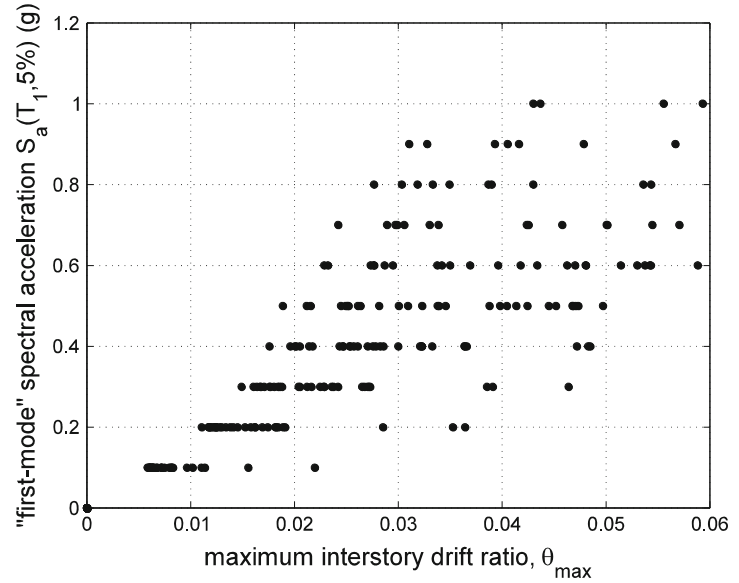
Of the three, cloud analysis is the least restrictive. Its name is derived from the characteristic cloud of results that appears in the IM versus EDP plane, where each point corresponds to one analysis (see Fig. 1). One can employ any combination of scaling and record set selection. At one end, one can employ only scaling, using a fixed

record set that is scaled to several levels of intensity (typically by multiplying all natural accelerograms by the same scale factor). At the other end, scaling can be completely avoided (if possible, considering catalogue limitations) by using a separate set of natural records for each level of IM (within some tolerance of course). Extracting the P(EDP|IM) information can be done in several ways. One approach is to employ parametric regression, taking care to separately fit the non-collapsing IM-EDP points via a linear regression in logarithmic space. “Infinite EDP” points indicative of collapse should be accounted for using logistic regression (see Shome and Cornell 1999; Jalayer and Cornell 2009). Alternatively, a nonparametric approach can be used, employing, e.g., “running” 16/50/84 percentiles to estimate the median and dispersion of the EDP response at each level of the IM.

Stripe analysis is named after the characteristic stripes of points aligned in distinct rows at different IM levels (Fig. 2). Like cloud analysis, it may involve any number of sets of records, and it also needs to employ at least some scaling to make sure that all runs at a given IM level actually perfectly correspond to the IM level requested, without any tolerance. The importance of this detail is that the distribution of EDP given

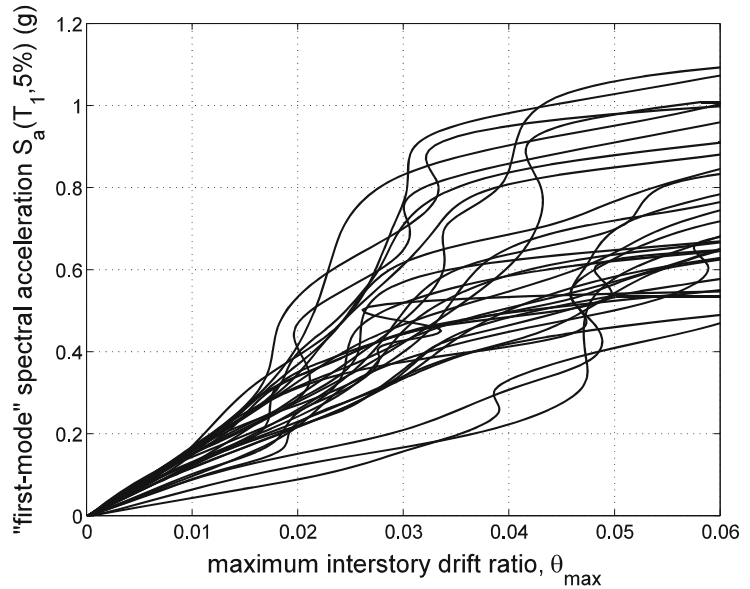
Analytic Fragility and Limit States [P(EDP|IM)]: Nonlinear Dynamic Procedures, Fig. 2

The resulting points of a stripe approach for a 9-story steel frame using $S_d(T_1)$ as the IM and the maximum over all stories peak interstory drift $\theta_{\max}$ as the EDP



Analytic Fragility and Limit States [P(EDP|IM)]: Nonlinear Dynamic Procedures, Fig. 3

The resulting IDA curves for a 9-story steel frame using $S_d(T_1)$ as the IM and the maximum over all stories peak interstory drift $\theta_{\max}$ as the EDP

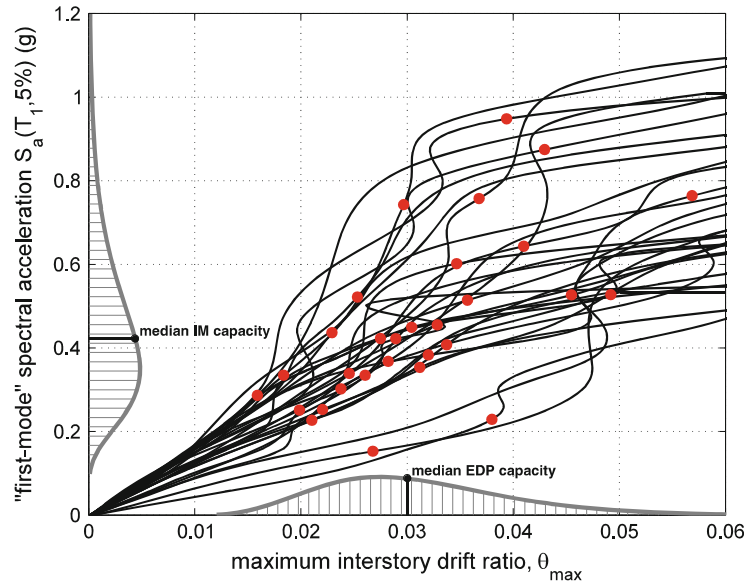


IM is directly represented at each IM level by the empirical distribution of the EDP results extracted from the pertinent analyses. In other words, any statistical quantity of EDP given the IM (mean, 16/50/84 percentile, standard deviation, etc.) can be estimated directly from the corresponding EDP values without any need for parametric or nonparametric regression, thus considerably simplifying postprocessing.

Incremental dynamic analysis (IDA) is based on scaling and scaling only (Vamvatsikos and Cornell 2002). Its focus is on individual records, each of which is scaled to several levels of intensity, typically until collapse is reached. Although closely related to stripe analysis, its main advantage is that the IM levels of the different records do not need to match. Instead, a fixed number of nonlinear dynamic analyses can be utilized for

Analytic Fragility and Limit States $[P(\text{EDP}|\text{IM})]$: Nonlinear Dynamic Procedures,

Fig. 4 Thirty IDA curves, thirty limit-state capacity points, and the corresponding EDP_c , IM_c PDFs for a 9-story steel frame (From Vamvatsikos 2013)



each record to achieve an accurate yet economical representation of response from elasticity to global dynamic instability. Then, interpolation of the results pertaining to each record is employed to generate continuous IDA curves (Fig. 3), as well as stripes, and subsequently derive the needed $\text{EDP}|\text{IM}$ statistics (Vamvatsikos and Cornell 2004).

In practice, any of the above approaches can be used to derive equivalent results. The distinction may become important only in the sense of scaling versus selecting ground motion records that puts IDA at a potential disadvantage: It has to use the same set of ground motions for all intensity levels, needing a sufficient IM to provide unbiased results. On the other hand, the representation of the $\text{EDP}|\text{IM}$ statistics via continuous IDA curves offers a powerful visual tool for understanding structural performance that can often tip the scales in its favor. In the end, the choice lies with the analyst and his/her understanding of the problem at hand.

As a final remark, it is important to remember that the aforementioned methods can only take into account the record-to-record variability. If other sources of uncertainty need to be incorporated, for example, model parameter variability, then more refined (and complex) approaches should be adopted instead (Liel et al. 2009;

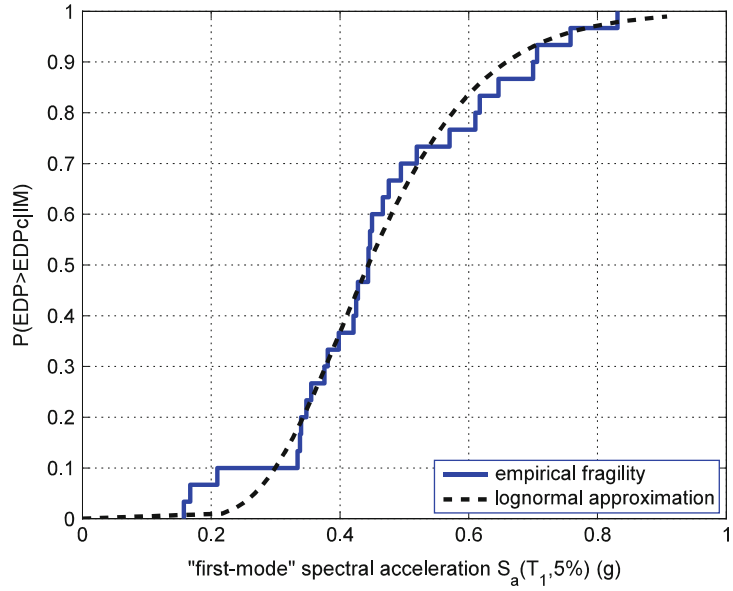
Dolsek 2009; Vamvatsikos and Fragiadakis 2010; Jalayer et al. 2011).

Determination of Fragility

A building-level fragility function is defined as a probability-valued function of the IM that represents the probability of violating a limit state given the IM level. Formally, it may be represented as $P(C < \text{DI}|\text{IM})$, i.e., as the probability of the seismic demand D exceeding the seismic capacity C given the IM. Two alternative yet equivalent formats exist to define the C and D terms, namely, the IM basis and the EDP basis. When a single EDP is used to define the limit state, this dichotomy is best understood by visualizing the first exceedance of the limit state as a single point on each IDA curve, having unique IM and EDP coordinates (Fig. 4).

If we choose to characterize demand and capacity by their IM counterparts, then $P(C < \text{DI}|\text{IM})$ simply becomes the cumulative distribution function (CDF) of the IM-valued capacity, IM_c . Simply plotting the empirical CDF of the IM_c values, as, for example, in Fig. 5, provides us with a ready estimate of the fragility function. By estimating the appropriate statistics (median, IM_{c50} , and standard deviation of the log

Analytic Fragility and Limit States $P(EDP|IM)$: Nonlinear Dynamic Procedures, Fig. 5 The empirical fragility function for exceeding a (deterministic) peak interstory drift of 3 % for the 9-story steel frame versus its lognormal approximation



values, β_{IMc}), one may also represent the results via an analytic lognormal approximation, as shown in Fig. 5, and formally represented by the following expression:

$$P(EDP_c < EDP|IM) = \Phi\left(\frac{\ln IM - \ln IM_{c50}}{\beta_{IMc}}\right). \quad (1)$$

If instead D and C are represented in EDP terms, then they are both potentially random quantities and integration is needed:

$$P(EDP_c < EDP|IM) = \int_0^{+\infty} P(EDP_c < EDP) P(EDP|IM) dIM. \quad (2)$$

The first term of the integrant is essentially the complementary cumulative distribution function (CCDF) of EDP_c , and the second is the probability distribution function (PDF) of EDP given the IM. Despite the apparent complexity of this approach, it is often the preferred method for estimating fragility (although not necessarily for representing it). This is best understood by considering the simpler case where the EDP capacity

is deterministic, i.e., we presume to perfectly know the value of EDP that causes violation of the limit state. Then, the capacity points of Fig. 4 align themselves into a single vertical line that intersects all IDA curves at the limiting (single) EDP_c value, meaning that all the dispersion in the fragility function (or the IM_c values, equivalently) originates from the records themselves. Inversely, if one wants to introduce the (typically considerable) uncertainty in EDP_c , he/she had best go through the EDP route; otherwise, it is not possible to obtain the inflated β_{IMc} dispersion of Eq. 1. Regardless of the route taken, though, the end result can always be represented in terms of the lognormal function parameters IM_{c50} and β_{IMc} of Eq. 1.

Summary

A key ingredient of modern frameworks for performance-based earthquake engineering is the assessment of the distribution of structural demand given the seismic intensity. Expressed as the probability of exceeding specific levels of the structural response indicative of defined limit states, this is also known as fragility. Recent advances in computer technology have allowed

the use of resource-intensive nonlinear dynamic analysis as the tool of choice for analytically estimating such fragility functions. For practically implementing such an assessment, several strategies exist on how to select and/or modify ground motion records and apply them to a structural model at several levels of intensity. Clouds, stripes, and incremental dynamic analysis curves, each characterized by the appearance of the corresponding results in the response versus intensity space, are among the most prominent techniques. Despite their superficial differences, they may all achieve the same fidelity in their results, practically differing only in the ease of postprocessing and interpretation. In all cases, parameterization by the (typically scalar) intensity measure means that considerable attention should be paid to its selection, as well as the selection of the ground motions, to ensure the unbiasedness of the estimates.

Cross-References

- [Analytic Fragility and Limit States \[P\(EDP|IM\)\]: Nonlinear Static Procedures](#)
- [Assessment of Existing Structures Using Response History Analysis](#)
- [Incremental Dynamic Analysis](#)
- [Nonlinear Dynamic Seismic Analysis](#)
- [Seismic Collapse Assessment](#)
- [Seismic Fragility Analysis](#)
- [Seismic Loss Assessment](#)
- [Time History Seismic Analysis](#)

References

- Baker JW, Cornell CA (2006) Spectral shape, epsilon and record selection. *Earthq Eng Struct Dyn* 35(9):1077–1095
- Bianchini M, Diotallevi PP, Baker JW (2009) Predictions of inelastic structural response using an average of spectral accelerations. In: *Proceedings of the 10th international conference on structural safety & reliability (ICOSSAR)*, Osaka
- Bradley BA (2010) A generalized conditional intensity measure approach and holistic ground motion selection. *Earthq Eng Struct Dyn* 39(12):1321–1342
- Cordova PP, Deierlein GG, Mehanny SS, Cornell CA (2000) Development of a two-parameter seismic intensity measure and probabilistic assessment procedure. *Proceedings of the 2nd US-Japan Workshop on Performance-based Earthquake Engineering Methodology for RC Building Structures*, Sapporo, Hokkaido
- Cornell CA, Krawinkler H (2000) Progress and challenges in seismic performance assessment. *PEER Center News*, <http://peer.berkeley.edu/news/2000spring/index.html>. Accessed May 2014
- Dolsek M (2009) Incremental dynamic analysis with consideration of modelling uncertainties. *Earthq Eng Struct Dyn* 38(6):805–825
- Iervolino I, Cornell CA (2005) Record selection for non-linear seismic analysis of structures. *Earthq Spectra* 21(3):685–713
- Jalayer F, Cornell CA (2009) Alternative non-linear demand estimation methods for probability-based seismic assessments. *Earthq Eng Struct Dyn* 38:951–972
- Jalayer F, Elefante L, Iervolino I, Manfredi G (2011) Knowledge-based performance assessment of existing RC buildings. *J Earthq Eng* 15(3):362–389
- Jalayer F, Beck JL, Zareian F (2012) Analyzing the sufficiency of alternative scalar and vector intensity measures of ground shaking based on information theory. *J Eng Mech ASCE* 138(3):307–316
- Katsanos EI, Sextos AG, Manolis GD (2010) Selection of earthquake ground motion records: a state-of-the-art review from a structural engineering perspective. *Soil Dyn Earthq Eng* 30:157–169
- Liel AB, Haselton CB, Deierlein GG, Baker JW (2009) Incorporating modeling uncertainties in the assessment of seismic collapse risk of buildings. *Struct Saf* 31(2):197–211
- Lin T, Haselton CB, Baker JW (2013a) Conditional-spectrum-based ground motion selection. Part I: Hazard consistency for risk-based assessments. *Earthq Eng Struct Dyn* 42(12):1847–1865
- Lin T, Haselton CB, Baker JW (2013b) Conditional-spectrum-based ground motion selection. Part II: Intensity-based assessments and evaluation of alternative target spectra. *Earthq Eng Struct Dyn* 42(12):1867–1884
- Luco N, Bazzurro P (2007) Does amplitude scaling of ground motion records result in biased nonlinear structural drift responses? *Earthq Eng Struct Dyn* 36:1813–1835
- Luco N, Cornell CA (2007) Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions. *Earthq Spectra* 23(2):357–392
- Shome N, Cornell CA (1999) Probabilistic seismic demand analysis of nonlinear structures. RMS program, report no. RMS35, PhD thesis, Stanford University, CA
- Shome N, Cornell CA, Bazzurro P, Carballo JE (1998) Earthquakes, records and nonlinear responses. *Earthq Spectra* 14(3):469–500
- Tothong P, Cornell CA (2006) An empirical ground-motion attenuation relation for inelastic spectral displacement. *Bull Seismol Soc Am* 96(6):2146–2164

- Vamvatsikos D (2013) Derivation of new SAC/FEMA performance evaluation solutions with second-order hazard approximation. *Earthq Eng Struct Dyn* 42:1171–1188
- Vamvatsikos D, Cornell CA (2002) Incremental dynamic analysis. *Earthq Eng Struct Dyn* 31(3):491–514
- Vamvatsikos D, Cornell CA (2004) Applied incremental dynamic analysis. *Earthq Spectra* 20(2):523–553
- Vamvatsikos D, Cornell CA (2005) Developing efficient scalar and vector intensity measures for IDA capacity estimation by incorporating elastic spectral shape information. *Earthquake Engineering and Structural Dynamics* 34(13):1573–1600
- Vamvatsikos D, Fragiadakis M (2010) Incremental dynamic analysis for estimating seismic performance sensitivity and uncertainty. *Earthq Eng Struct Dyn* 39(2):141–163

Analytic Fragility and Limit States [P(EDP|IM)]: Nonlinear Static Procedures

Matjaž Dolšek

Faculty of Civil and Geodetic Engineering,
University of Ljubljana, Ljubljana, Slovenia

Synonyms

Fragility analysis; Fragility curve; Limit state; Nonlinear static procedure; Pushover analysis; SDOF model; Seismic demand; Target displacement; Uncertainty

Introduction

Nonlinear static (pushover-based) procedures were initially developed for the seismic performance assessment of structures (e.g., Saiidi and Sozen 1981; Fajfar and Fischinger 1988; Paret et al. 1996; Chopra and Goel 2002) and later on used also for the fragility analysis (e.g., Fragiadakis and Vamvatsikos 2010; Barbato et al. 2010; Dolšek 2012; Jalayer et al. 2011; Rota et al. 2014; Kosič et al. 2014). For example, Jalayer et al. (2011) proposed the use of Bayesian method for calculating fragility curve of the reinforced concrete frame. Rota et al. (2014)

adopted logic tree approach for evaluating the effects of modelling uncertainties on the fragility curves of masonry buildings. Alternatively, the impact of the modelling uncertainties on the fragility curve can be simulated by utilizing the so-called probabilistic SDOF model (Kosič et al. 2014). The advantage of pushover analyses in comparison with nonlinear response history analysis lies in the former's simplicity, since the results of pushover analysis are very intuitive. This makes pushover-based methods attractive for practical applications. On the other hand, they are based on several assumptions. Thus, the analyst should have good background knowledge in order to be able to adequately interpret results obtained by using pushover-based methods. However, many studies have shown that, in the case of simple bridges and buildings up to around eight stories high, the mean/median response can be sufficiently accurately predicted by means of conventional pushover analysis. In the case of more complicated structures, the analyst should think about taking into account, at least approximately, torsional effects and the so-called higher mode effects (e.g., Reyes and Chopra 2011; Kreslin and Fajfar 2012). Reyes and Chopra (2011) and Kreslin and Fajfar (2012) used linear elastic analysis in estimating higher mode contributions seismic demands. If the target displacement is obtained by nonlinear response history analysis utilizing the single-degree-of-freedom (SDOF) model, then it is possible to consider the different system failure modes obtained from the pushover analyses (Brozovič and Dolšek 2014). In this case, the failure-based SDOF models are used in addition to the modal-based SDOF model associated with the first model.

It should be noted that the main objective of this entry is to introduce the use of conventional pushover-based method for fragility analysis, taking into account the randomness of ground motion and the epistemic (modeling) uncertainty. Thus, not all issues associated with pushover-based methods are precisely addressed.

The entry has been organized into three sub-entries. The theoretical background for the prediction of the parameters of SDOF models is first described. This subentry is subdivided into two

parts. The transformation of the equations of motion to a conventional equation of motion of a SDOF model is presented in the first part, while in the second part, the procedure to determine the parameters of the SDOF model using pushover analysis is described. The first subentry provides an insight into the SDOF model, which represents a key link between the pushover analysis and seismic demand. Quite precise instructions are given on how to determine a SDOF model which can be used to assess the seismic demand by nonlinear dynamic analysis. The first subentry is followed by an overview of the different types of pushover analysis and the limit states. Some variants of pushover analyses are only mentioned, since it is not the aim of this entry to provide an insight into different types of pushover analysis. In the final subentry, pushover-based fragility analysis is explained with emphasis on the step-by-step procedure, which is also demonstrated by means of an example of fragility analysis for a four-story reinforced concrete frame building. The EDP-based and IM-based formulations of the fragility function are introduced. This is followed by explaining three procedures for the estimation of the fragility parameters using pushover-based methods. Finally, the most comprehensive procedure, which takes into account the ground-motion randomness and modeling uncertainty, is demonstrated by explaining each step of the pushover-based fragility analysis.

The Theoretical Background for Predicting Parameters of the SDOF Model

The main objective of this derivation is to introduce the theoretical background for the determination of the parameters of the SDOF model, which is one of the key components of the pushover-based methods. The advantage of the SDOF model is its simplicity. Thus, it can be used to determine the seismic demand based on nonlinear response history analyses. However, it is important to understand the theoretical limitations of such approach.

Transformation of Equations of Motion to a Conventional Equation of a Single-Degree-of-Freedom Model

The seismic response of structures can, in general, be described by the equation of motion

$$\mathbf{M}\ddot{\mathbf{U}} + \mathbf{C}\dot{\mathbf{U}} + \mathbf{K}\mathbf{U} = -\mathbf{M}\ddot{\mathbf{U}}_g \quad (1)$$

where the first, second, and third parts of the equation on the left-hand side represent, respectively, the vectors of the inertial, damping, and resisting forces. The sum of these forces should be in equilibrium with external load, which, in the case of an earthquake, is expressed by the product of the mass matrix $\mathbf{M}$ and the vector of the ground acceleration $\ddot{\mathbf{U}}_g$. Equation 1 is derived for the simulation of the seismic response of a structure relative to the ground. Thus, the displacements $\mathbf{U}$, velocities $\dot{\mathbf{U}}$, and accelerations $\ddot{\mathbf{U}}$ are vectors which are expressed relative to the ground. The kinematic quantities are, in general, simulated at all joints of the finite elements which form the structure. The matrix $\mathbf{C}$ is a damping matrix, which is often modeled in proportion to the mass matrix and/or the stiffness matrix $\mathbf{K}$. It should be noted that all kinematic quantities are functions of time. In the case of the nonlinear response of a structure, the stiffness matrix and, in some cases, also the damping matrix depend on the history of the structure's deformations.

In order to achieve brevity of the derivation, it is assumed that the building structure is symmetric in plan. The problem therefore simplifies to the planar case. It is further assumed that the seismic response of the building can be described just by the horizontal displacements at each story level. The external load of Eq. 1 simplifies as follows:

$$\mathbf{M}\ddot{\mathbf{U}} + \mathbf{C}\dot{\mathbf{U}} + \mathbf{K}\mathbf{U} = -\mathbf{M}\mathbf{1}\ddot{u}_g \quad (2)$$

where $\mathbf{1}$ is a vector with ones, and $\ddot{u}_g$ is the ground acceleration. It should be noted that the vector of the resisting forces,

$$\mathbf{F} = \mathbf{K}\mathbf{U} \quad (3)$$

which are function of time, can now be understood as the external, equivalent, static forces

which produce a vector $\mathbf{U}$ of displacements at any individual floor level, at any instant in time. Despite simplifications, Eq. 2 is a system of n -dependent nonhomogenous second-order differential equations with nonlinear coefficients, where n stands for the number of stories of the building structure. Several methods exist for the solution of this equation at any instant in time, but there are also several reasons to further simplification of the problem.

For example, all nonlinear structural models are just an approximation of the realistic seismic response of a structure. The majority of current structural models are capable of simulating several of the failure modes that can be observed in realistic structures. From this point of view, the question arises as to whether it is justified to treat the whole system as dynamic, if the model used to represent a building structure is already a proxy of reality. Furthermore, although the solution of the equation of motion will provide exact results from the theoretical point of view, the seismic loading is not known in advance. Thus, the selection of inappropriate ground motions could have a greater impact on the results than the selected method of analysis. Finally, solving system of equations of motion is extremely time demanding and, in the case of complex systems, also related to poor convergence of the numerical results obtained. These are just some reasons that it is often not practical to use response history analysis in order to simulate the seismic response of a structure. Therefore, Eq. 2 is often further simplified.

The solution of Eq. 2 would be significantly easier if the motion of the structure could be described with sufficient accuracy by reducing the number of differential equations to just one, e.g., in the case of building structures, by simulating just the top (roof) displacement. From the theoretical point of view, such a simplification can be achieved by assuming a displacement vector $\mathbf{U}$ in the following form:

$$\mathbf{U} = \Psi u_t \quad (4)$$

where Ψ is a displacement vector that is independent of time and is normalized to a top

displacement equal to 1, and u_t is the top displacement, which is a function of time. Based on such an assumption, it becomes clear that a multi-degree-of-freedom (MDOF) model of a structure can be transformed into a single-degree-of-freedom (SDOF) model. This is achieved by substituting all the kinematic quantities in the form of Eq. 4 and pre-multiplying Eq. 2 by the vector Ψ^T :

$$\Psi^T \mathbf{M} \Psi \ddot{u}_t + \Psi^T \mathbf{C} \Psi \dot{u}_t + \Psi^T \mathbf{K} \Psi u_t = -\Psi^T \mathbf{M} \mathbf{1} \ddot{u}_g \quad (5)$$

Equation 5 can be rearranged into the following form:

$$m_\psi \ddot{u}_t + c_\psi \dot{u}_t + k_\psi u_t = -m_* \ddot{u}_g \quad (6)$$

by realizing that the results of the multiplication of the matrices (Eq. 5) are scalars:

$$\begin{aligned} m_\psi &= \Psi^T \mathbf{M} \Psi \\ c_\psi &= \Psi^T \mathbf{C} \Psi \\ k_\psi &= \Psi^T \mathbf{K} \Psi \end{aligned} \quad (7)$$

$$m_* = \Psi^T \mathbf{M} \mathbf{1} \quad (7a)$$

By introducing Eq. 6, it is clear that the system of n -dependent differential equation has been simplified to just one equation of motion, where the unknown quantities are the displacement, velocity, and acceleration at the top of the building. However, it still does not have the form of the conventional equation of motion, since $m_* \neq m_\psi$. In order to make Eq. 6 equivalent to the conventional equation of motion, a so-called transformation factor, which relates the top displacement of the structure to the displacement of the SDOF model, needs to be defined as follows:

$$\Gamma = \frac{m_*}{m_\psi} = \frac{\Psi^T \mathbf{M} \mathbf{1}}{\Psi^T \mathbf{M} \Psi} \quad (8)$$

and the kinematic quantities of the SDOF model as

$$u_* = \frac{u_t}{\Gamma}, \quad \dot{u}_* = \frac{\dot{u}_t}{\Gamma}, \quad \ddot{u}_* = \frac{\ddot{u}_t}{\Gamma} \quad (9)$$

Equation 6 can now be transformed into the conventional equation of motion of the SDOF model by taking into account Eqs. 8 and 9:

$$m_* \ddot{u}_* + c_* \dot{u}_* + k_* u_* = -m_* \ddot{u}_g \quad (10)$$

where

$$c_* = \Gamma c_\psi \quad (11)$$

and

$$k_* = \Gamma k_\psi \quad (11a)$$

Determination of the Parameters of the SDOF Model Using Pushover Analysis

There are no exact instructions as to how to determine the parameters of the SDOF model since the mass m_* , the damping c_* , and the stiffness k_* all depend on the assumed displacement shape. In order to obtain a practical solution to Eq. 10, the analyst has to determine the force-displacement relationship of the SDOF model taking into account certain hysteretic rules, which are often assumed to be similar to those used for the structural components. However, the displacement of the SDOF model can be simply determined according to Eq. 9. Some further explanation is needed in order to be able to better understand the background for the determination of the forces of the SDOF model from the pushover curve. In general, the product of the stiffness k_* and u_* represents a force in the spring of the SDOF model:

$$F_* = k_* u_* \quad (12)$$

which should be expressed in a different form by taking into account Eqs. 11a and 9:

$$F_* = k_\psi u_t \quad (13)$$

Equation 13 can be further transformed by substituting k_ψ using Eq. 7 and with consideration of Eqs. 3 and 4:

$$F_* = \Psi^T \mathbf{F} \quad (14)$$

It should be noted that the force-displacement relationship of the SDOF model is, in general, nonlinear due to the nonlinear response of the structure. Thus, there is no exact answer as to which is the most appropriate shape of the normalized displacement vector Ψ or which are the most appropriate lateral forces $\mathbf{F}$. If it is assumed that the structure oscillates just in one given mode shape, then $\Psi = \Phi$. If it is further assumed that the mode shape Φ does not change significantly with increasing damage to the structure and that the damping does not affect the shape of the buildings' oscillation, then it can be shown, by analogy to the modal analysis, that the shape of the equivalent static forces which causes lateral displacements that are equivalent to the mode shape can be expressed as follows:

$$\mathbf{F} = \mathbf{M} \Phi \quad (15)$$

Substituting Eq. 15 into Eq. 14 and realizing that in such a case $\Psi = \Phi$, it can be shown that

$$F_* = \Phi^T \mathbf{M} \Phi \quad (16)$$

Multiplying and dividing Eq. 16 by $m_* = \Psi^T \mathbf{M} \mathbf{1}$ and taking into account Eq. 8 produces the following result:

$$F_* = \frac{\Phi^T \mathbf{M} \mathbf{1}}{\Gamma} \quad (17)$$

Eq. 18 can be further simplified by realizing that $\mathbf{F}^T = \Phi^T \mathbf{M}$:

$$F_* = \frac{\mathbf{F}^T \mathbf{1}}{\Gamma} = \frac{F_b}{\Gamma} \quad (18)$$

where F_b is the so-called base shear, i.e., the sum of the lateral forces $\mathbf{F}$. The force-displacement relationship of the SDOF model is thus obtained simply by dividing the quantities at the structural level by the so-called transformation factor. However, it should be emphasized that Eq. 18 may not be sufficiently accurate in the case when the pushover curve is based on a higher mode shape. Equation 9 can be used, together with Eqs. 14 or 18, to determine the

force-displacement $F_* - u_*$ relationship of the SDOF on the basis of the pushover curve.

It is often useful to know the period of the SDOF model, which can be obtained by analyzing Eq. 10. It is well known that the radial frequency of an oscillating body is

$$\omega_* = \sqrt{\frac{k_*}{m_*}} = \frac{2\pi}{T_*} \quad (19)$$

whence it follows that

$$T_* = 2\pi \sqrt{\frac{m_*}{k_*}} \quad (20)$$

The stiffness of the SDOF model can be expressed by means of Eq. 12. The period of the SDOF model thus equals

$$T_* = 2\pi \sqrt{\frac{m_* u_*}{F_*}} \quad (21)$$

In order to solve Eq. 10, it is necessary to define the damping model c_* . The so-called Rayleigh damping model is often assumed at the level of the structure. In this case, the damping matrix is expressed as a linear combination of the mass and stiffness matrices:

$$\mathbf{C} = \alpha \mathbf{M} + \beta \mathbf{K} \quad (22)$$

where α and β are constant values. The question arises as to how to relate the damping model at the structural level to the damping model for the SDOF model. Starting from Eq. 11 and taking into account Eq. 7 for c_ψ as well as Eq. 22, it can be shown that

$$\begin{aligned} c_* &= \Gamma c_\psi = \Gamma \Psi^T \mathbf{C} \Psi = \Gamma \Psi^T (\alpha \mathbf{M} + \beta \mathbf{K}) \Psi = \\ &= \Gamma (\alpha \Psi^T \mathbf{M} \Psi + \beta \Psi^T \mathbf{K} \Psi) = \\ &= \alpha \Gamma m_\psi + \beta \Gamma k_\psi \end{aligned} \quad (23)$$

Substituting Eqs. 8 and 11 into Eq. 23, the following equation is obtained:

$$c_* = \alpha m_* + \beta k_* \quad (24)$$

whence it can be shown that the same damping constants α and β of the SDOF model are equal to that assumed for the response history analysis of the entire structure. The damping constants α and β , or only one of these constants, if damping is proportional to the mass or stiffness matrix only, are often determined by assuming a certain ratio of critical damping. Describing the relationship between the damping constants and the ratio of critical damping is, however, outside the scope of this derivation.

Equation 10 can be solved by numerical integration which is implemented in conventional software for response history analysis. However, the solution depends, in general, on the ground motion and assumptions adopted for the determination of the parameters of the SDOF model, i.e., the mass m_* , the damping model, the force-displacement relationship $F_* - u_*$, and the hysteretic rules.

In the simplest and the most commonly used approach, the parameters of the SDOF model are determined by performing pushover analysis taking into account invariant distribution of the lateral forces. In this case, the procedure for the determination of the parameters of the SDOF model can be decomposed into the following steps:

1. Define the distribution of the lateral forces $\mathbf{F}$ for the pushover analysis.
2. Perform the pushover analysis. This results in the relationship between the force vector $\mathbf{F}$ and the displacement vector $\mathbf{D}$ from elastic range to the collapse of the structure. The corresponding base shear F_b and the top displacement u_t form the so-called pushover curve.
3. Assume the shape of the normalized displacement vector Ψ .
4. Idealize the pushover curve $F_b - u_t$ using a bilinear or multi-linear shape or the shape of any other curve which can be used to simulate the $F_* - u_*$ force-displacement relationship of the SDOF model.
5. Compute the $F_* - u_*$ force-displacement relationship of the SDOF model. The force F_* at the characteristic points of the idealized

pushover curve can be obtained, in the case of the SDOF model, according to Eq. 14. In the case when the distribution of displacement vector $\mathbf{D}$ is more or less constant for all levels of damage, then the forces F_* can be determined by Eq. 18.

6. Assume hysteretic behavior of the SDOF model.
7. Compute the mass m_* of the SDOF model according to Eq. 7.
8. Compute the damping constants α and/or β based on the assumed damping model and the ratio of critical damping.

Pushover Analysis and Limit States

Pushover analysis is a kind of nonlinear static analysis which aims at determining the relationship between the lateral forces and the engineering demand parameters (EDPs) for all performance levels of a structure. The results of pushover analysis are very intuitive, which makes it attractive for practical applications. It is clear that the relationship between the forces and the deformations are linearly elastic until damage occurs to the first element. Beyond this level, an increment of lateral forces causes a greater increment in the deformations in comparison with that observed in the case of linear elastic analysis, since the stiffness of the structure is reduced due to the damage which has occurred to some of the elements. When the deformations reach a certain level, the maximum strength of the structure is reached. From here on, the strength of the structure decreases due to the softening of the material (e.g., see Fig. 2) or due to second-order effects, if they are considered in the analysis. In less general case, when the softening of material and second-order effects is not taken into account, the pushover curve becomes a straight line once a plastic mechanism forms.

The strength and deformation capacity, the available ductility, and the system failure mode represent the global results of the pushover analysis, which should always be checked in the case when seismic performance assessment of structure is based on a nonlinear analysis method. The

results of a pushover analysis are usually presented in the form of a pushover curve, which represents a relationship between the base shear and top (roof) displacement of the investigated structure (e.g., see Fig. 2). The pushover curve is, in fact, a characteristic of the structure. However, it depends on the type of pushover analysis used.

The pushover analysis becomes an approximate analysis method if it is used for the prediction of expected seismic performance of a structure (e.g., Saiidi and Sozen 1981; Fajfar and Fischinger 1988; Fajfar 2000). In this case, it is important that the structural model implicitly accounts for the cyclic deterioration. However, the results of pushover analysis are used for the definition of the SDOF model, which represents a link between the seismic action and the seismic demand.

The seismic demand at the level of the SDOF model can be obtained by means of numerous procedures. Once the displacement at the level of the SDOF model has been obtained, it can be transformed to the selected EDP at the structural level. Since several phenomena associated with the response of a structure during an earthquake are neglected in pushover-based methods, it is clear that the results of such simplified procedure are approximate also from the theoretical point of view. Due to the approximate nature of pushover-based procedures, many types of pushover analysis have been developed in recent decades. Some of them are briefly described below.

Types of Pushover Analysis

Conventional pushover analysis (e.g., Saiidi and Sozen 1981; Fajfar and Fischinger 1988) is based on an invariant distribution of the lateral forces. The shape of the lateral forces is therefore independent of deformations, whereas the shape of the deformations is dependent on the level of structural damage. This is the most common type of pushover analysis, which is recommended by some guidelines and building codes (e.g., FEMA 356, Eurocode 8). It is usually prescribed that pushover analysis should be performed for at least two vertical distributions of the lateral loads

in order to assess the range of EDPs that might occur during actual dynamic response. Eurocode 8 suggests the use of the so-called uniform and modal load patterns, which are, respectively, proportional to the story masses and the lateral forces from the elastic analysis. According to FEMA 356, two patterns have to be selected from two groups. The first load pattern should be consistent with the modal pattern, whereas a second pattern can be based on a uniform load pattern or on an adaptive load distribution which accounts for the redistribution of the lateral loads, taking into account the properties of the yielded structure. Due to the simplicity of conventional pushover analysis, even more than two pushover analyses can be easily performed. For example, an iterative pushover-based procedure (IPP) (Celarec and Dolšek 2013) involving model adaptation was recently introduced, which aims at improving the capability of simulations based on simplified nonlinear models. It was shown that the IPP can be used to approximately simulate the shear failure of columns (Celarec and Dolšek 2013; Kosič et al. 2014), although simplified nonlinear models are not capable of direct simulation of such effects.

Adaptive pushover analysis is an alternative to conventional pushover analysis. In force-based adaptive pushover analysis, the load vectors are gradually updated due to the damage occurring to the structure, which affects the mode shapes. One of the first procedures for the adaptive lateral load for pushover analysis was proposed by Bracci et al. (1997). Many different procedures for force-based adaptive pushover analysis followed. Most approaches involve response spectrum analysis which is performed for the different modal characters of the structure which are associated with selected stages of the pushover analysis. However, Aydinoglu (2003) has argued that the load patterns should be based on an inelastic rather than an elastic response spectrum. It was soon realized that force-based adaptive pushover analysis offers only a relatively minor advantage in comparison with conventional pushover analysis. As a result of such an observation, a displacement-based adaptive pushover procedure was proposed (Antoniou and Pinho 2004), which,

however, did not provide significantly improved prediction in comparison to the force-based algorithms that are used for adaptive pushover analysis.

Conventional or adaptive pushover analysis provides a single pushover curve, which can be used in conjunction with the SDOF model to obtain the expected values of EDPs. For some types of buildings and EDPs, such a simulation of the seismic response of a structure provides sufficiently accurate results. However, the analyst has to be careful when interpreting the results of the seismic performance assessment of a structure on the basis of one pushover curve, since it is not possible to assess all types of EDPs on the basis of single pushover analysis. In general, the dynamic response of a structure is far more complex, especially if it suffers significant damage. In general, several system failure modes can be observed due to the randomness of ground motions. Additionally, in the case of taller buildings, several modes can significantly contribute to the results. On the other hand, mode shapes depend on the scale of inflicted damage, which gradually increases during a strong earthquake. Thus, the maximum values of EDPs do not appear at the same instant in time. Furthermore, the impact of mode shapes varies with respect to the EDPs. For example, the higher mode effects can be significant for the prediction of story drifts and for the internal forces in the upper parts of taller buildings. Much effort has been expended in order to try to approximately account for the so-called effect of higher modes. One of the first attempts in this direction has been made by Paret et al. (1996), who performed several pushover analyses using a lateral load pattern based on different elastic mode shapes. A theoretical basis for the modal pushover analysis procedure for estimating the seismic demand for buildings was introduced by Chopra and Goel (2002). The method involves several pushover analysis and corresponding modal-based SDOF models, whereas the overall response is obtained by combining the results of all considered modes. Recently, a general pushover analysis was introduced (Sucuoğlu and Günay 2011). It is based on several pushover analyses, which are performed

on the basis of the lateral forces obtained as a combination of the modal forces. It was also shown that the so-called higher system failure modes can significantly contribute to the response of taller buildings. Brozovič and Dolšek (2014) have shown that system failure modes observed due to the randomness of ground motion can be sufficiently accounted for by the envelope-based pushover analysis procedure, which is based on several pushover analyses and the use of failure-based SDOF models. The failure-based SDOF models utilize the displacement vectors corresponding to the system failure modes that are observed from pushover analyses in the case of second or higher modes. This significantly improves the prediction of the intensity causing a designated limit state, if the structure is sensitive to the effects of higher modes.

Definition of Limit States

As already mentioned above, pushover analysis is the simplest nonlinear analysis method, which has been included in several guidelines and standards for seismic performance assessment and the retrofitting of buildings. The results of pushover analysis can be used, in conjunction with the target displacement for a given seismic action, to check whether the structural performance satisfies the defined performance objectives. For this purpose, building codes provide limit states at the component or structural level. Eurocode 8-3 (CEN 2005) utilizes the following limit states: damage limitation (DL), significant damage (SD), and near collapse (NC). These limit states are explicitly defined at the component level, whereas definitions at the structural level are only descriptive. For example, it is prescribed that the near-collapse limit state at the structural level is attained when a structure is heavily damaged, with some residual lateral strength and stiffness, but is still capable of sustaining vertical loads and aftershocks of moderate intensity. However, moderate permanent drifts are present, and it is not likely, from the economic point of view, that repair of the building could be justified. Note that the NC limit state is often defined at the global level, i.e., when the strength in the post-capping range of pushover curve decreases for 20 %.

Each country can prescribe return periods of the design seismic action associated with these limit states. By default, it is assumed that appropriate levels of protection are achieved when the structural performance for the DL, SD, and NC limit states is checked for seismic actions corresponding to return periods of 225, 475, and 2475 years, respectively. The prestandard for the seismic rehabilitation of buildings FEMA 356 (2000), which was issued a few years before the Eurocode 8-3, is based on similar metrics for checking the target building performance. The basic structural performance levels of a building are immediate occupancy (IO), life safety (LS), and collapse prevention (CP). FEMA 356 states that building performance is a combination of the performance of both the structural and the nonstructural components. The description of performance levels (i.e., limit states), which are discrete damage states selected from among the infinite spectrum of possible damage states that buildings could experience during an earthquake, consists of estimates rather than precise predictions (FEMA 2000).

According to the brief overview of limit states provided in some standards, it is clear that there are many different definitions of them. In general, each limit state is associated with a certain degree of damage. However, there are some restrictions regarding the definition of limit states on the basis of pushover analysis. For example, the pushover analysis is not able to explicitly simulate cycling deterioration. Thus, in the case of pushover analysis, limit states cannot be dependent on the level of cyclic deterioration. Since damage models are often defined by story drifts or the rotations in plastic hinges, such a limitation of pushover analysis, associated with an inability to simulate cycling deterioration, is not critical, provided that the nonlinear model used for pushover analysis implicitly accounts for cyclic deterioration of the structural components. In the case of nonstructural components, it is common to define a limit state by the level of story acceleration, which is not explicitly simulated by pushover analysis. A simplified model for story acceleration is therefore also required.

Although there are no critical restrictions for the definition of limit states in the case when seismic demand is obtained by the pushover-based method, it should be realized that the accuracy of pushover-based methods is limited. The analyst should therefore be careful when checking the limit states, especially in the case when target performance is defined at the component level. It is thus recommended that the limit states should be defined at the global and structural levels, taking into account the global parameters obtained from the pushover curve. Even in this case, it is recommended that the structural damage associated with such defined limit states should be checked, since it is possible that a flexible structure has not suffered significant damage even though the corresponding point on the pushover curve is beyond the post-capping point. Such performance can be observed in the case of flexible frames, where the P- Δ effect has a quite significant impact on structural performance.

Pushover-Based Seismic Fragility Analysis

The result of fragility analysis is the fragility function, which represents the probability that an engineering demand parameter (EDP) will exceed a certain limit-state value edp given an intensity $IM = im$. In the simplest case, it is assumed that the seismic fragility function is based on a lognormal distribution function:

$$P(EDP > edp | IM = im) = 1 - \Phi\left(\frac{\ln edp - \ln \mu_{edp}}{\sigma_{\ln edp}}\right) \quad (25)$$

where $\Phi\left(\frac{\ln edp - \ln \mu_{edp}}{\sigma_{\ln edp}}\right)$ represents the standard normal distribution function, μ_{edp} is the median value of the EDP given the intensity im , and $\sigma_{\ln edp}$ is the standard deviation of the natural logarithms of the EDP given that $IM = im$. The above definition is suitable when the limit state is based on a threshold value of an engineering

demand parameter. In the case of an IM-based definition of a limit state, which is common for the collapse limit state, the fragility function can be defined as follows:

$$P(LS | IM = im) = \Phi\left(\frac{\ln im - \ln \mu_{imLS}}{\sigma_{\ln imLS}}\right) \quad (26)$$

where $P(LS | IM = im)$ is the probability of exceeding the limit state LS if the intensity measure assumes a value equal to im , μ_{imLS} is the median limit-state intensity, and $\sigma_{\ln imLS}$ is the corresponding standard deviation of the natural logarithms. The fragility function is therefore defined by two parameters, the median value μ_{edp} or μ_{imLS} and the corresponding standard deviation $\sigma_{\ln edp}$ or $\sigma_{\ln imLS}$. Many different procedures for the estimation of these parameters based on pushover analysis have been developed. It should also be noted that many different procedures for fragility analysis exist. Some of them can also be used on the basis of seismic performance assessment utilizing pushover analysis (e.g., Jalayer et al. 2011; Rota et al. 2014).

The accuracy of determination of the fragility parameters (μ , $\sigma_{\ln}$) depends on the method for seismic performance assessment of a structure and the level of the simulations, which accounts for the effect of uncertainty.

Deterministic Approach for Assessing the Parameters of the Fragility Function

In the simplest case, the effects of uncertainty can be assumed by the default values of $\sigma_{\ln}$. Consequently, the deterministic seismic performance of a structure can be used. This enables the use of analytical models for the determination of the target displacement. Parametric studies have shown that the $\sigma_{\ln imLS}$ depends on the limit state and the type of seismic intensity measure. If the intensities causing collapse are assessed by the spectral acceleration corresponding to the first vibration period, the $\sigma_{\ln imLS}$ is in the interval between 0.3 and 0.5 (Lazar and Dolšek 2014). The $\sigma_{\ln imLS}$ increases with respect to the period of structure if it is assessed on the basis of peak ground acceleration. It varies between 0.5 and

0.75. The value of $\sigma_{\ln imLS}$ can additionally increase if epistemic uncertainties are considered in addition to the record-to-record randomness.

In Europe, the most commonly used approach for the determination of expected seismic response based on pushover analysis is the N2 method (e.g., Fajfar 2000), which has been implemented in Eurocode 8 (CEN 2004), whereas the so-called coefficient method is used in the United States (e.g., FEMA 356 (2000)). According to Eurocode 8 and FEMA 356, the target displacement (d_t and δ_t), respectively, can be determined as follows:

$$d_t = \Gamma \cdot \frac{\mu(q_u, T^*, T_C)}{q_u} \cdot S_e(T^*) \frac{T^{*2}}{4\pi^2}, \quad (27)$$

$$\delta_t = C_0 \cdot C_1(R, T_S, T_e) C_2 C_3 \cdot S_a(T_e) \frac{T_e^2}{4\pi^2}, \quad (28)$$

where the parameters Γ and C_0 have the same meaning, i.e., they are the so-called transformation or modification factors which relate the spectral displacement of an equivalent SDOF model to the roof displacement of the MDOF model. For the experienced reader, it is also trivial that the last parts of Eqs. 27 and 28 represent the elastic spectral displacement S_{de} , whereas the middle part of Eqs. 27 and 28 corresponds to the so-called inelastic displacement ratio C (e.g., Miranda 2000), which depends on many parameters, such as the ground motion type, the hysteretic behavior, the force-displacement relationship, the masses, damping, and others. For reasons of simplicity, formulas for the determination of the inelastic displacement ratio have been introduced based on the results of parametric studies. For example, in Eurocode 8, the simplified $R - \mu - T$ relationship proposed by Fajfar (2000) is used to define the inelastic displacement ratio, whereas in FEMA 356 (2000), two additional parameters are used to define the inelastic displacement ratio, i.e., the parameter C_2 , which represents the effect of a pinched hysteretic shape, stiffness degradation, and strength deterioration on maximum displacement, and the parameter C_3 , which represents the increased displacement due to dynamic P- Δ effects. Note that the meaning of

all the variables in Eqs. 27 and 28 is described in the corresponding references (CEN 2004; FEMA 2000).

Consideration of Ground-Motion Randomness in Estimation of Parameters of Fragility Function

The first level of simulations involves consideration of the effect of ground-motion randomness. In general, the target displacement according to this approach is based on nonlinear dynamic analysis of the SDOF model. The analyst should therefore select an appropriate set of ground motions and define the SDOF model, which could be used for nonlinear dynamic analysis as previously described in this entry. The results of such analyses are samples of EDP values given the different values of the intensities or the sample of limit-state intensities, which could be used to estimate the parameters of the fragility function. Some alternatives exist which can be used in order to avoid the use of nonlinear dynamic analysis at the level of the SDOF model. Vamvatsikos and Cornell (2006) developed the SPO2IDA software tool, which is capable of recreating the seismic behavior of a single-degree-of-freedom (SDOF) model with a quadri-linear force-displacement relationship. They performed a parametric study by suitably varying the five parameters of the force-displacement relationship, which are negative and nonnegative hardening, the residual plateau, ductility at the beginning of strength degradation, and ductility at collapse. Based on the results of the parametric study, regression analysis was used to define each segment of the approximate 16th, 50th, and 84th fractile IDA curves. Although SPO2IDA is a useful software tool, it is not easy to extend its applicability, since a new regression analysis is required if the seismic response parameters are to be computed for additional ground motion records. More recently, Peruš et al. (2013) introduced a web-based methodology for the prediction of approximate IDA curves, which consists of two independent processes. The result of the first process is a response database of the single-degree-of-freedom model, whereas the second process involves the prediction of approximate

IDA curves from the response database by using n -dimensional linear interpolation. Such an approach makes possible user-friendly prediction of the seismic response parameters with high accuracy. In order to demonstrate the capabilities of the proposed methodology, a web application for the prediction of the approximate 16th, 50th, and 84th fractile response of a reinforced concrete structure was developed (<http://ice4risk.slo-projekt.info/WIDA/>). Note that web-based methodology for the prediction of the approximate IDA curves estimates an error of the predicted IDA curves of a SDOF model.

Consideration of Ground-Motion Randomness and Epistemic Uncertainty in Estimation of Parameters of Fragility Function

In a general case, fragility parameters should be based on simulations which take into account the effects of ground-motion randomness and the other knowledge-based sources of uncertainty, such as physical uncertainties, modeling uncertainties, human errors, uncertainties associated with the knowledge level about the structure, and others. In order to account for knowledge-based uncertainties, the simulations can be performed at the level of a MDOF model (e.g., Fragiadakis and Vamvatsikos 2010; Dolšek 2012) or a SDOF model (Kosič et al. 2014). Many different simulation methods can be used to perform pushover analysis with consideration of knowledge-based uncertainties (e.g., Barbato et al. 2010; Jalayer et al. 2011). It is convenient and straightforward to utilize Monte Carlo simulation with Latin hypercube sampling (e.g., Fragiadakis and Vamvatsikos 2010; Dolšek 2012). The LHS technique uses stratification of the probability distribution function of the random variables X_i and consequently requires significantly fewer simulations in comparison with the ordinary type of Monte Carlo simulation. In general, two steps are needed to determine the sample of random variables, which are directly applied in the structural model. First, each random variable is sampled by inverse method using equidistant points between sample probabilities. If the random variables are correlated, the sample

should be generated in such a way that the correlation structure between the random variables is simulated in the best possible manner. This represents the second step in the process of the determination of the sample of random variables. The problem can be successfully solved by the stochastic optimization method called simulated annealing. More details are available elsewhere (e.g., Dolšek 2012).

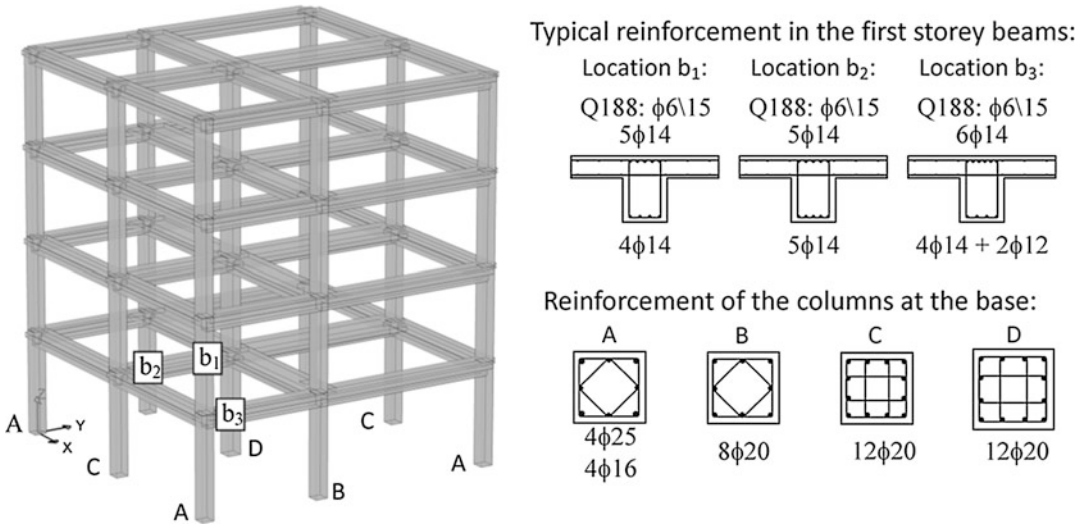
However, all the referenced procedures are based on simulations which are performed at the level of a MDOF model. Recently, an approximate procedure utilizing a deterministic MDOF model and uncertainty analysis at the level of a SDOF model was proposed (Kosič et al. 2014). It can be classified as being somewhere in between the aforementioned procedures for the determination of fragility parameters on the basis of pushover methods, both in terms of computational time and in terms of accuracy. Such an approach is computationally less demanding, since all the simulations are performed at the level of the so-called probabilistic SDOF model. It can therefore be attractive for the fragility analysis of more extensive building stock.

Example: Pushover-Based Seismic Fragility Analysis of the RC Frame Building With Consideration of Ground-Motion Randomness and Modeling Uncertainty

A simple variant of a pushover-based fragility analysis is demonstrated by means of an illustrative example. The procedure involves conventional pushover analyses and the nonlinear dynamic analyses of corresponding SDOF models. The effects of ground-motion and modeling uncertainties are accounted for, respectively, by the set of ground motions and the set of structural models which are generated by Monte Carlo with LHS. The step-by-step procedure for the determination of the fragility curve is as follows:

1. *Preparation of data regarding the structure and the seismic hazard at the location of the structure.* It is important to obtain as accurate as possible data regarding the seismic hazard in order to reduce the effect of uncertainties.

- Information regarding the seismic hazard and the soil type are important for the selection of appropriate ground motions.
2. *Definition of limit states.* In general, limit states can be defined at the component or structural level. As briefly discussed above, there is no critical restriction regarding the definition of limit states if the seismic demand is based on a pushover analysis procedure.
 3. *Identification of epistemic uncertainties.* Uncertainties are often divided into aleatoric and epistemic, although there is no need to make such a distinction. In this example, the aleatoric uncertainties are accounted for by the set of ground motions (step 5), whereas the effects of the epistemic uncertainties are incorporated by the set of structural models. The most important sources of epistemic uncertainty should be modeled by appropriate random variables.
 4. *Determination of the set of structural models utilizing Monte Carlo with LHS.* The LHS technique uses stratification of the probability distribution function of the random variables and consequently requires fewer simulations in comparison with a crude Monte Carlo simulation. However, any appropriate LHS technique can be used in order to generate the sample of the uncertain input parameters. It has been previously shown (e.g., Dolšek 2012) that the effect of epistemic uncertainties can be simulated with sufficient accuracy by a number of structural models, which is at least twice the number of random variables. It is worth emphasizing that the nonlinear structural model should implicitly account for cyclic deterioration.
 5. *Selection of an appropriate set of ground motions.* Many different procedures exist for the selection of an appropriate set of ground motions. Ground motions are usually selected to match the target spectrum, which is, in the simplest case, the elastic acceleration spectrum prescribed by a building code.
 6. *Calculation of pushover curves for all the structural models simulated in step 4.* Different types of pushover analysis can be performed. Conventional pushover analysis, which is based on the invariant lateral load, is the simplest and is the most often applied. The results of the pushover analysis are pushover curves and the seismic demand on the structure given the top displacement. In general, more than one pushover analysis per structural model is required.
 7. *Determination of the SDOF models.* This step requires the transformation of top displacement and base shear to the displacement (Eq. 9) and force (Eq. 14 or Eq. 18) of the SDOF model. The force-displacement relationship of the SDOF model should be multi-linear. Negative post-capping stiffness can also be applied in order to estimate the collapse capacity. It is usually assumed that the hysteretic behavior is the same as that prescribed for the components of the structural models. The mass of the SDOF model should be determined according to Eq. 7a, whereas the damping can be proportional to mass and/or stiffness (Eq. 24).
 8. *Response history analyses for all SDOF models and ground motions.* Since, in the case of SDOF models, the response history analysis is not computationally demanding, it is recommended that the records should be scaled and that the relationship between the ground-motion intensity and the displacement of the SDOF model should be computed for several levels of damage, e.g., from elastic behavior to collapse.
 9. *Assessment of seismic demand at the level of structural model.* The displacement obtained from the response history analysis of the SDOF model should be transformed to the top displacement using Eqs. 8 and 9. Any other engineering demand parameter at the level of structure can be obtained from the results of the pushover analysis on the basis of the known top displacement. However, if more than one pushover analysis is performed per structural model, in order to account for the effect of higher failure modes, then the total demand can be obtained by enveloping the results of all pushover



Analytic Fragility and Limit States [P(EDP|IM)]: Nonlinear Static Procedures, Fig. 1 A view of the investigated four-story building and the reinforcement in selected columns and beams

analyses (e.g., Brozovič and Dolšek 2014) or by combining the results together according to an appropriate combinational rule (Chopra and Goel 2002). The results of this step are samples of the engineering demand parameters given the intensities or samples of limit-state intensities.

10. *Calculation of the fragility parameters and the fragility curve.* Fragility curves can be directly estimated on the basis of the sample values. In the case when the fragility curves are based on an assumed lognormal distribution, the method of moments, the maximum likelihood method, the method of counted percentiles, or any other suitable method can be used to calculate the median engineering demand parameters or the limit-state intensities and the corresponding dispersions.

The application of the above-described step-by-step procedure for the determination of the fragility curves is demonstrated by means of a four-story reinforced concrete frame building (Fig. 1), which was previously analyzed (Dolšek 2012). Just a brief overview of the procedure is presented.

Step 1. The structure was designed for a peak ground acceleration of 0.3 g, soil type B, and

ductility class high (behavior factor = 5). Concrete of quality C25/30 and B500 Tempcore reinforcing steel were used to construct the building. Minimum and maximum mean concrete strengths of 32 MPa and 56 MPa were measured. The modulus of elasticity of the concrete varied from 28.5 to 35.3 GPa. The mean yield strength exceeded the characteristic yield strength by between 10 % and 20 %, depending on the diameter of the reinforcing bars. However, the average value of the yield strength of the reinforcing steel amounted to about 580 MPa.

Step 2. Fragility analysis was performed for the limit states of damage limitation (DL), significant damage (SD), and near collapse (NC), as defined in Eurocode (CEN 2005). It was assumed that the DL limit state is attained at the structural level if the longitudinal reinforcement in all the columns in any story starts to yield. It was further assumed that the SD and NC limit states are reached, at the structural level, when the moment exceeds the maximum moment in the case of the first column or the rotation exceeds the ultimate rotation in the case of first column, respectively. However, additional criteria for the occurrence of the SD and NC limit states were

defined. It was assumed that the top displacement corresponding to the NC limit state should not be less than that associated with 80 % of the maximum base shear measured in the post-capping range of the pushover curve. Similarly, it was assumed that the top displacement corresponding to the SD limit state should be not less than the 75 % of the NC limit-state top displacement.

Step 3. Since full knowledge about the building structure was available, only the following modeling uncertainties were considered in the fragility analysis: mass, strength of the concrete and the reinforcing steel, effective slab width, damping, and the model for determining the initial stiffness and ultimate rotation in the plastic hinges of the beams and columns. A normal or lognormal distribution was assumed for the majority of the nine random variables ($N_{\text{var}} = 9$). The statistical characteristics of the input random variables were taken from literature as reported elsewhere (Dolšek 2012). All the considered input random variables were assumed to be uncorrelated with respect to the story masses, which were assumed to be perfectly correlated and modeled by one random variable. The highest value of the coefficient of variation was adopted for the prediction of the ultimate rotations in the beams (0.6) and columns (0.4).

Step 4. Realization of the random variables was based on the LHS technique. The size of the sample was assumed to be equal to 20, which is slightly larger than twice the number of all the random variables considered in the fragility analysis. The correlation matrix of the sample was almost identical to the target correlation matrix. Note that this is only a necessary condition for the prediction of the seismic response parameters with the required accuracy, whereas the sufficient condition is related to the number of simulations, which is, in the case of the example, relatively low ($N_{\text{sim}} = 20$). However, the results of performed studies have shown that, in the case when N_{sim} is larger than twice the number of random variables (N_{var}), an increase in the number of simulations does not significantly

affect the estimated median seismic response parameters and the corresponding dispersion. The results of this step were 20 structural models based on the sample of random variables. For comparison, an additional “deterministic” structural model was also created by assuming median values of the random variables.

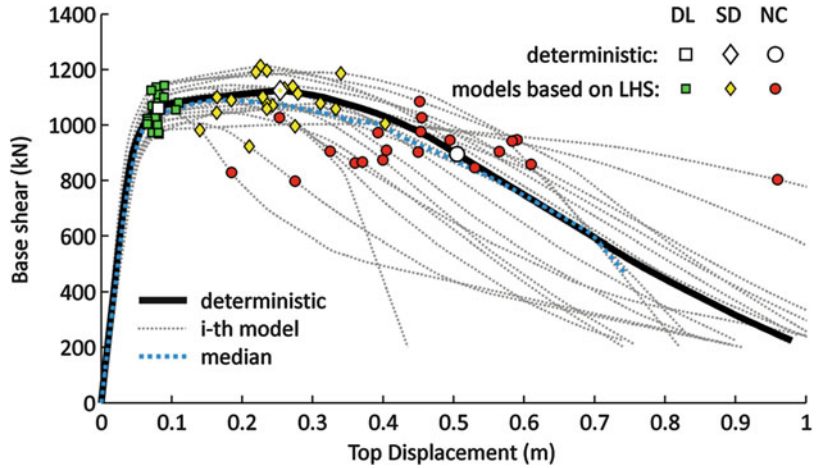
Step 5. A set of 14 ground motions ($N_{gm} = 14$) were selected from European Strong Motion Database to match the Eurocode-based acceleration spectrum. All the selected records were recorded on stiff soil and had peak ground acceleration greater than $0.1g$. The mean acceleration spectrum of the selected ground motions matched reasonably well with the target Eurocode-based spectrum. It should be noted that ground motions selected in this way often overestimate the seismic response with respect to that based on hazard-consistent ground motions.

Step 6. The lateral load pattern for conventional pushover analyses was based on the first vibration mode (Eq. 15). For simplicity, the structure was analyzed in the Y direction only (see Fig. 1). The pushover curves for all structural models from *Step 4* are presented in Fig. 2. An extensive scatter can be observed in the case of the deformation capacity of the realized structural models, whereas the difference in the strength is less. However, no significant differences can be observed between the deterministic model and the so-called median pushover curve. The limit-state top displacement and the corresponding base share are highlighted for the DL, SD, and NC limit states. The high coefficient of variation of the limit-state top displacement, which increases gradually with respect to the severity of the limit state, is partly the consequence of the high coefficient of variation of the input parameters and partly the consequence of the formation of different system failure modes, which were observed from pushover analyses.

Step 7. The pushover curves were idealized by means of a trilinear force-displacement relationship, also taking into account negative post-capping stiffness. Transformation of the

Analytic Fragility and Limit States [P(EDP|IM)]: Nonlinear Static Procedures, Fig. 2

The pushover curves corresponding to the structural models based on the LHS technique, the pushover curve corresponding to the deterministic model, and the median pushover curve. The points on the pushover curve represent the limit-state top displacements



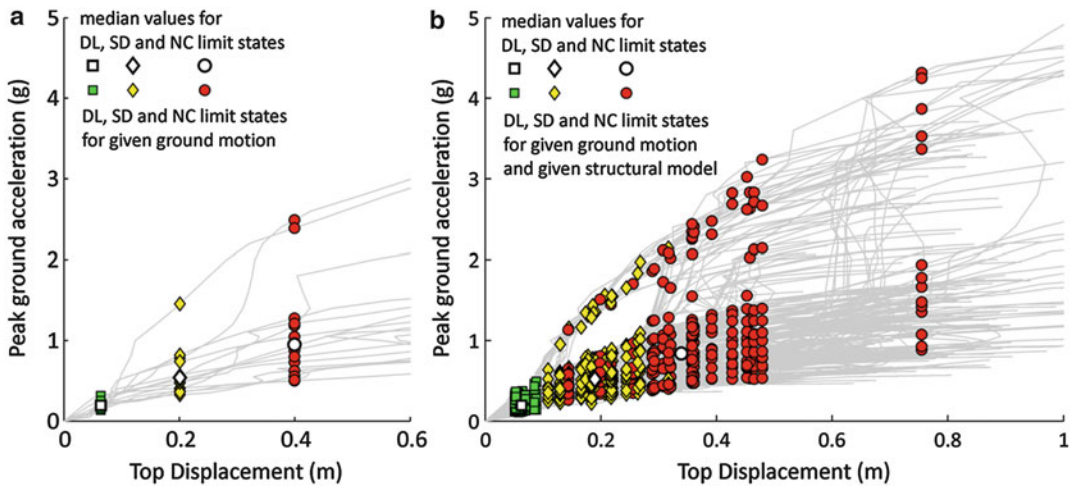
idealized pushover curve to the parameters of an equivalent SDOF model was performed by dividing the base shear and top displacement of the idealized pushover curve by a transformation factor F (Eq. 8). The mass of the SDOF model and the force-displacement relationships of the SDOF model were obtained according to Eq. 7a and Eqs. 9 and 18, respectively. Note that the number of the SDOF models was the same as the number of pushover curves. All the parameters of the SDOF models differed from one another due to the epistemic uncertainties. However, the peak-oriented hysteretic model and damping proportional to the tangent stiffness were assumed to be invariant.

Step 8. The response history analyses were performed for each SDOF model by scaling the ground-motion intensity in order to obtain the complete relationship between the peak ground acceleration and the displacement of the SDOF model. The results are presented in Fig. 3, where the highlighted points indicate the limit-state intensities which were used for the calculation of the fragility parameters.

Steps 9 and 10. The limit-state top displacements highlighted in Fig. 2 were transformed into the displacement of the corresponding SDOF model, which was then used to determine the limit-state intensities. The limit-state displacements and the limit-state intensities

shown in Fig. 3a are presented for the deterministic SDOF model, taking into account just ground-motion randomness. The variability can be observed only for the limit-state intensities, while in the adjacent figure (Fig. 3b), variability can be observed also for the limit-state displacements. The latter variability is the consequence of the effect of epistemic uncertainty. The sample of limit-state intensities presented in Fig. 3b was used to estimate the parameters of the fragility curves. The median values and the corresponding dispersion of natural logarithms were estimated according to maximum likelihood method. The fragility curves were then obtained using Eq. 26. The results are presented in Fig. 4.

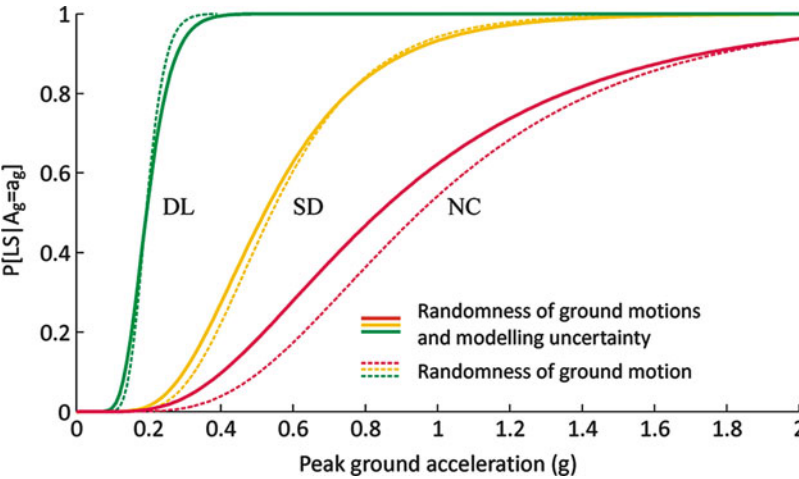
A flatter fragility curve can be observed in the case of the DL limit state if the effects of epistemic uncertainty are considered in the fragility analysis. This is in agreement with expectations and can be modeled by using the mean estimate approach, in which it is assumed that the effects of epistemic uncertainties on the fragility curve can be modeled by simply inflating the dispersion. However, in the case of the NC limit state, it is clear that incorporating the epistemic uncertainties in the determination of the fragility curve causes a shift of the fragility curve to the left, which means that the effects of epistemic



Analytic Fragility and Limit States [P(EDP|IM)]: Nonlinear Static Procedures, Fig. 3 The limit-state peak ground accelerations and displacements associated with the set of structural models and (a) the deterministic

SDOF model and (b) the SDOF models based on the LHS technique. The median limit-state points for the DL, SD, and NC limit states are also presented

Analytic Fragility and Limit States [P(EDP|IM)]: Nonlinear Static Procedures, Fig. 4 The fragility curves for the DL, SD, and NC limit states taking into account only the randomness of ground motions and both, i.e., randomness of the ground motions and modeling uncertainty



uncertainty reduced the median limit-state intensity. This effect was frequently observed, especially in the case of reinforced concrete frames.

Summary

Pushover-based methods are used around the world as a tool for the approximate seismic performance assessment of structures. Due to their

computational efficiency, pushover-based methods are attractive also for fragility analysis, which incorporates the effects of ground motions and other sources of uncertainty. In this entry, emphasis was placed on explaining the theoretical background for predicting seismic demand using a SDOF model, since it is important to understand the basic assumptions of the approximate methods. A variant of the pushover-based fragility analysis is then explained through a

step-by-step procedure and demonstrated by means of an example of a four-story building. It has been shown that pushover-based methods can be successfully used for fragility analysis. However, the analyst should always be aware of the approximate nature of pushover-based methods. Such an awareness will allow him to decide to which type of structures and for what purpose individual pushover-based methods can be applied.

Cross-References

- [Reinforced Concrete Structures in Earthquake-Resistant Construction](#)
- [Seismic Fragility Analysis](#)
- [Seismic Reliability Assessment, Alternative Methods for](#)
- [Uncertainty Theories: Overview](#)

References

- Antoniou S, Pinho R (2004) Development and verification of a displacement-based adaptive pushover procedure. *J Earthq Eng* 8(5):643–661
- Aydinoğlu MN (2003) An incremental response spectrum analysis procedure based on inelastic spectral displacements for multi-mode seismic performance evaluation. *Bull Earthq Eng* 1(1):3–36
- Barbato M, Gu Q, Conte JP (2010) Probabilistic push-over analysis of structural and soil-structure system. *J Struct Eng* 136(11):1330–1341
- Bracci JM, Kunnath SK, Reinhorn AM (1997) Seismic performance and retrofit evaluation of reinforced concrete structures. *J Struct Eng* 123(1):3–10
- Brozović M, Dolšek M (2014) Envelope-based pushover analysis procedure for the approximate seismic response analysis of buildings. *Earthq Eng Struct Dyn* 43(1):77–96
- Celarec D, Dolšek M (2013) Practice-oriented probabilistic seismic performance assessment of infilled frames with consideration of shear failure of columns. *Earthq Eng Struct Dyn* 42(9):1339–1360
- CEN (2004) Eurocode 8: design of structures for earthquake resistance – part 1: general rules, seismic actions and rules for buildings, EN 1998–1. European Committee for Standardisation, Brussels
- CEN (2005) Eurocode 8: design of structures for earthquake resistance – part 3: assessment and retrofitting of buildings, EN 1998–3. European Committee for Standardisation, Brussels
- Chopra AK, Goel RK (2002) A modal pushover analysis procedure for estimating seismic demands for buildings. *Earthq Eng Struct Dyn* 31(3):561–582
- Dolšek M (2012) Simplified method for seismic risk assessment of buildings with consideration of aleatory and epistemic uncertainty. *Struct Infrastruct Eng* 8(10):939–953
- Fajfar P (2000) A nonlinear analysis method for performance-based seismic design. *Earthq Spectra* 16(3):573–592
- Fajfar P, Fischinger M (1988) N2 – a method for non-linear seismic analysis of regular buildings. Proceedings of ninth world conference on earthquake engineering, vol V, Tokyo-Kyoto
- FEMA (2000) Prestandard and commentary for the seismic rehabilitation of buildings, vol 356, FEMA. Federal Emergency Management Agency, Washington, DC
- Fragiadakis M, Vamvatsikos D (2010) Fast performance uncertainty estimation via pushover and approximate IDA. *Earthq Eng Struct Dyn* 39(6):683–703
- Jalayer F, Elefante L, Iervolino I, Manfredi G (2011) Knowledge-based performance assessment of existing RC buildings. *J Earthq Eng* 15(3):362–389
- Kosič M, Fajfar P, Dolšek M (2014) Approximate seismic risk assessment of building structures with explicit consideration of uncertainties. *Earthq Eng Struct Dyn* 43(10):1483–1502
- Kreslin M, Fajfar P (2012) The extended N2 method considering higher mode effects in both plan and elevation. *Bull Earthq Eng* 10(2):695–715
- Lazar N, Dolšek M (2014) Incorporating intensity bounds for assessing the seismic safety of structures: does it matter? *Earthq Eng Struct Dyn* 43(5):717–738
- Miranda E (2000) Inelastic displacement ratios for structures on firm sites. *J Struct Eng* 126(10):1150–1159
- Paret TF, Sasaki KK, Eilbeck DH, Freeman SA (1996) Approximate inelastic procedures to identify failure mechanisms from higher mode effects. In: Proceedings of 11th world conference on earthquake engineering (P.No. 966), Acapulco
- Peruš I, Klinc R, Dolenc M, Dolšek M (2013) A web-based methodology for the prediction of approximate IDA curves. *Earthq Eng Struct Dyn* 42(1):43–60
- Reyes JC, Chopra AK (2011) Three-dimensional modal pushover analysis of buildings subjected to two components of ground motions, including its evaluation for tall buildings. *Earthq Eng Struct Dyn* 40:789–806
- Rota M, Penna A, Magenes G (2014) A framework for the seismic assessment of existing masonry buildings accounting for different sources of uncertainty. *Earthq Eng Struct Dyn* 43(7):1045–1066
- Saiidi M, Sozen MA (1981) Simple nonlinear seismic analysis of R/C structures. *J Struct Div ASCE* 107(5):937–953
- Sucuoğlu H, Günay MS (2011) Generalized force vectors for multi-mode pushover analysis. *Earthq Eng Struct Dyn* 40(1):55–74
- Vamvatsikos D, Cornell CA (2006) Direct estimation of the seismic demand and capacity of oscillators with multi-linear static pushovers through IDA. *Earthq Eng Struct Dyn* 35(9):1097–1117

Ancient Monuments Under Seismic Actions: Modeling and Analysis

Ioannis N. Psycharis

School of Civil Engineering, Department of
Structural Engineering, National Technical
University of Athens (N.T.U.A.), Athens, Greece

Synonyms

Classical monuments

Introduction

Classical monuments are made of structural elements (called drums in case of columns), which lie one on top of the other without mortar. Columns are connected to each other with architraves (epistyles) consisting of stone beams, usually made of marble. A characteristic example is shown in Fig. 1 from the Olympieion of Athens, Greece.

Architrave beams are usually connected to each other with iron clamps and dowels. However, in most cases no structural connections are provided between the drums of the columns. Only in few cases, iron shear connectors (dowels) are provided at the joints, which restrict, up to their yielding, sliding but do not affect rocking. The wooden dowels that were usually placed at the joints among the drum of the columns were aiming at centering the stones during construction and, practically, do not have any effect on their seismic response.

Due to their spinal construction, columns of ancient monuments respond to strong earthquakes with intense rocking and sliding of the drums. Therefore, the dynamic analysis of ancient monuments is a difficult problem to treat, since it does not follow the behavior rules of continuum systems.

Several investigators have examined the seismic response of classical monuments and, in general, of stacks of rigid bodies analytically, numerically, or experimentally, mostly using

two-dimensional models (e.g., Allen et al. 1986; Sinopoli 1989; Psycharis 1990; Winkler et al. 1995; Psycharis et al. 2000; Konstantinidis and Makris 2005; Papaloizou and Komodromos 2009 among others) and lesser three-dimensional ones (e.g., Papantonopoulos et al. 2002; Mouzakis et al. 2002; Psycharis et al. 2003, 2013; Dasiou et al. 2009a, b). These studies have shown that, due to the domination of rocking, the characteristics of the response are similar to the ones of the simplest case, the rocking of a rigid block. They have also shown that these structures, despite their apparent instability to horizontal loads, are, in general, earthquake resistant (Psycharis et al. 2000), which is also proven from the fact that many classical monuments built in seismic prone areas have survived for almost 2,500 years.

In general, the dynamic behavior of ancient monuments is highly nonlinear and complicated. The abovementioned extensive numerical and experimental investigations have shown that such structures do not possess natural modes in the classical sense and the period of free vibrations is amplitude dependent. During a strong earthquake the response alternates between different “modes” of vibration, each one being governed by a different set of equations of motion. As a result, the response is highly nonlinear. An example of this nonlinearity is that a column may collapse under a certain earthquake motion and be stable under the same excitation magnified by a value greater than one.

In addition, the response is very sensitive to even trivial changes of the parameters of the system or the excitation. This was verified during experiments, since “identical” experiments produced significantly different results in some cases. The sensitivity of the dynamic behavior was also evident in the numerical analyses, in which trivial changes in the excitation or the system parameters changed the response significantly. Another effect of the response sensitivity is the significant out-of-plane displacements observed for purely planar excitations: in some experiments, the deformation in the direction normal to the plane of the excitation was of the same order of magnitude with the principal deformation.



Ancient Monuments Under Seismic Actions: Modeling and Analysis, Fig. 1 SE corner of the Olympieion of Athens, Greece

The vulnerability of classical monuments to earthquakes depends on two main parameters: (i) the predominant period of the ground motion and (ii) the size of the structure. The period of the excitation affects significantly the response and the risk of collapse, with low-frequency earthquakes being much more dangerous than high-frequency ones. In the first case, the response is characterized by intensive rocking, while in the latter rocking is usually restricted to small values, but significant sliding of the drums occurs, especially at the upper part of the structure. In this sense, near field ground motions, which contain long-period directivity pulses, might bring these structures to collapse. The size of the structure is another important parameter, with bulkier structures being much more stable than smaller ones with the same aspect ratio of dimensions.

Despite their apparent instability, classical monuments without significant damages are not

vulnerable to usual earthquake motions. However, collapse can occur much easier if imperfections are present, as cutoff of drums, displaced drums, inclined columns due to foundation failure, etc. Such imperfections are common in ancient monuments and may endanger the safety of the structure in future earthquakes.

It is evident that the dynamic analysis of classical monuments for earthquake loads is a very important part of the restoration/preservation process. However, it is not an easy task. Due to the complicated response, such analyses can only be performed using powerful computational codes that can account for rocking, sliding, and even complete separation of the individual stones. Additionally, all such analyses are inevitably based on assumptions concerning the material and joint properties, some of which are difficult to evaluate. Therefore, and taking under consideration the sensitivity of the response, one must have in mind that any dynamic analysis of an ancient monument contains a certain amount of doubt. In any case, the earthquake response of ancient monuments does not follow general rules, and each monument deserves a case-by-case investigation that focuses on the specific monument, recognizing its current condition and local seismic hazard.

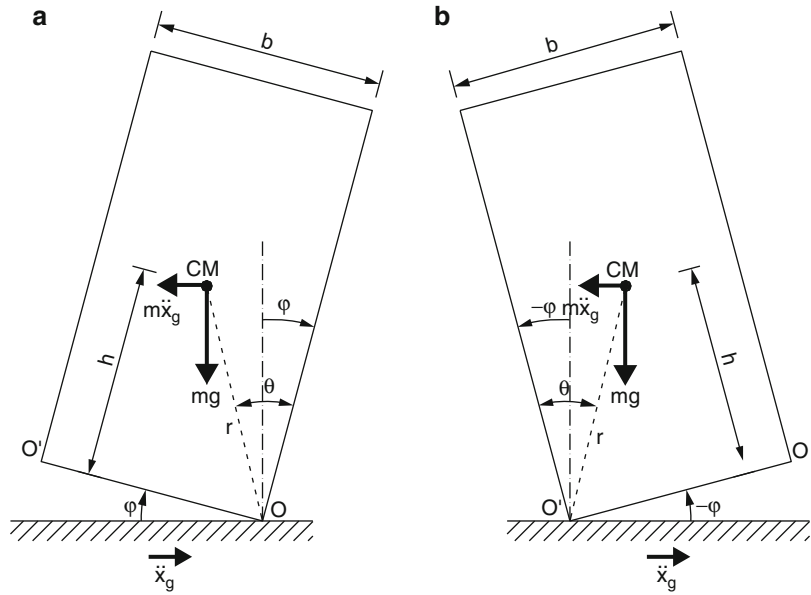
In this article, the basic features of the dynamic response of multi-drum columns and colonnades are presented, and critical points regarding their modeling and analysis for seismic loads are discussed. First, the important features of the simplest case, the rocking block, are presented. Then, the dynamics of multi-drum columns and groups of columns (colonnades) are discussed, while emphasis is given to the sensitivity of the response to several parameters of the numerical model. Finally, guidelines regarding the proper selection of base motion acceleration time histories to be used in the analyses are presented.

Rocking Response of a Rigid Block

Despite its apparent simplicity, the rocking response of a rigid block sitting on a rigid base

Ancient Monuments Under Seismic Actions: Modeling and Analysis,

Fig. 2 Rocking of a rigid block on a rigid base without sliding: (a) positive angle of rotation; (b) negative angle of rotation



is a complicated problem that has attracted the interest of many researchers in the last century. The basic scope of these investigations has been the estimation of the required intensity of the ground motion to overturn slender bodies during strong earthquakes. The first attempts towards this aim were made by Milne (1885), Milne and Omori (1893) and Kirkpatrick (1927). However, it was Housner (1963) who first tackled the problem systematically, deriving the equations of motion that govern the rocking response of a rigid, freestanding block and examining the overturning risk under simple ground pulses.

Equation of Motion

Let us consider the rocking block shown in Fig. 2. The block and the base are considered rigid, and the coefficient of friction is assumed to be large enough to prevent sliding of the block. In Fig. 2, a rectangular block is shown, with dimensions: width of the base = b and height of the center of mass (CM) from the base = h . However, the equations of motion and the basic features of the response that are presented in the following are also valid for any symmetric block in the vertical direction with such dimensions and arbitrary shape.

Assuming that the block has been set into rocking motion, the rotation can either happen

about corner O of the base (Fig. 2a), which corresponds to positive angles of rotation (clockwise), or about corner O' of the base (Fig. 2b), which corresponds to negative angles of rotation (counterclockwise). The distinction between positive and negative angles of rotation is essential, because a different equation governs the motion in each case.

Under a horizontal base excitation $\ddot{x}_g(t)$, the equation of motion can easily be derived by applying Newton's second law for the moments about the pole of rotation (point O for positive rotations, point O' for negative rotations) and taking under consideration that, apart from the weight, a horizontal inertial force (d' Alembert force) acts at CM, equal to $-m\ddot{x}_g(t)$. Then, the equation of motion can be written in the form:

$$I_O \ddot{\varphi} \pm mgr \sin(\theta \mp \varphi) = -m\ddot{x}_g r \cos(\theta \mp \varphi) \quad (1)$$

where φ is the rocking angle, m is the mass of the block, r is the distance of CM from the pole of rotation (point O or O'), θ is the slenderness angle defined by $\theta = \tan^{-1}b/(2h)$, I_O is the mass moment of inertia about O or O', and g is the acceleration of gravity. In this equation, whenever a double sign appears, the upper one

corresponds to $\varphi > 0$ (clockwise rotation) and the lower one to $\varphi < 0$ (counterclockwise rotation).

Introducing the parameter p defined by

$$p^2 = \frac{mgr}{I_O} \quad (2)$$

and using the $\text{sgn}()$ function, Eq. 1 can be written in the following general form:

$$I_O \ddot{\varphi} + p^2 \sin[\theta \text{sgn}(\varphi) - \varphi] = -p^2 \cos[\theta \text{sgn}(\varphi) - \varphi] \quad (3)$$

The parameter p is a characteristic frequency of the system. Actually it is equal to the pendulum frequency of the block when hung about the corner of the base.

As mentioned above, Eqs. 1 and 3 are valid for any rigid block with an axis of symmetry about the vertical axis through CM, with I_O being the mass moment of inertia about the corner of the base. In the special case of a homogeneous rectangular block, I_O is given by the following relation:

$$I_O = \frac{m[b^2 + (2h)^2]}{3} = \frac{4mr^2}{3}, \quad (4)$$

and the characteristic frequency p becomes

$$p = \sqrt{\frac{3g}{4r}} \quad (5)$$

For small angles of rotation ($\varphi \ll 1$), Eq. 1 can be linearized about the equilibrium position assuming that $\cos \varphi \approx 1$ and $\sin \varphi \approx \varphi$ and neglecting the second order term containing the product $\varphi \cdot \ddot{x}_g(t)$. Taking under consideration that $r \sin \theta = b/2$ and $r \cos \theta = h$, the following linearized equation can be derived:

$$I_O \ddot{\varphi} - mgh\varphi = \mp mg \frac{b}{2} - mh\ddot{x}_g \quad (6)$$

For large angles of rotation, when the block is close to the verge of overturning, the linearization should be performed about the point of unstable

equilibrium assuming that $|\theta \mp \varphi| \ll 1$, thus $\cos(\theta \mp \varphi) \approx 1$ and $\sin(\theta \mp \varphi) \approx (\theta \mp \varphi)$. Then, Eq. 1 becomes

$$I_O \ddot{\varphi} - mgr\varphi = \mp mgr\theta - mr\ddot{x}_g \quad (7)$$

It is noted that, as evident from Eqs. 6 and 7, the rocking block corresponds to a system with negative stiffness. Also, for slender blocks with slenderness angle θ less than about 20° , the linearized Eqs. 6 and 7 do not differ significantly, since $\tan \theta \approx \sin \theta \approx \theta$. Analyses of such blocks showed that the linearized equations can predict with acceptable accuracy for engineering purposes the response and the overturning risk of the block.

Impact with the Ground

During rocking, the pole of rotation alternates from point O to O', or vice versa. The alteration of the pole of rotation takes place when the block hits the ground and is accompanied by energy dissipation due to the change in the velocity of the center of mass which is schematically shown in Fig. 3. Let us assume that the angular velocity before impact is $\dot{\varphi}_1$ and after impact is $\dot{\varphi}_2$. Then, one can write

$$\dot{\varphi}_2 = \varepsilon \cdot \dot{\varphi}_1 \quad (8)$$

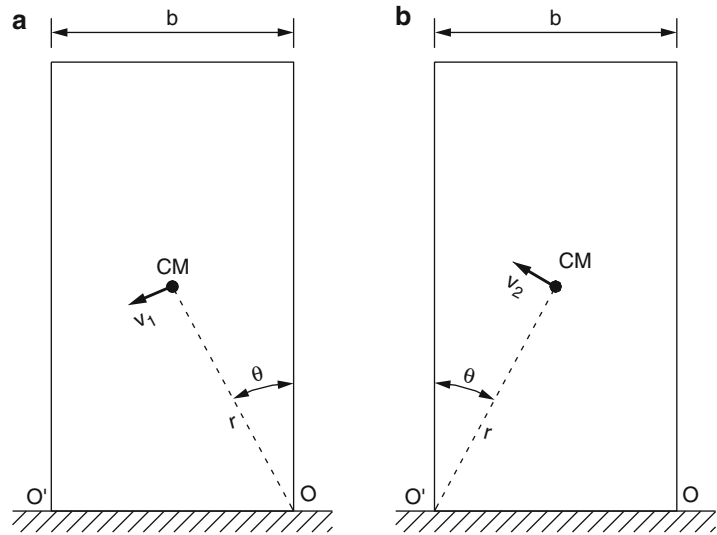
where ε is a coefficient of restitution. If we assume that the impact is fully plastic, i.e., that there is no rebound of the block on the base, the value of ε can be calculated applying the principle of conservation of the angular momentum about point O' before and after impact (Housner 1963), which gives

$$\varepsilon = 1 - \frac{mb^2}{2I_O} \quad (9)$$

Equation 9 implies that the coefficient of restitution ε and the dissipated energy depend solely on the geometry of the block. However, experimental investigation (e.g., Priestley et al. 1978; Aslam et al. 1980) showed that the actual value of ε might be significantly different than the theoretical value of Housner, depending on the

Ancient Monuments Under Seismic Actions: Modeling and Analysis,

Fig. 3 Change in the velocity of the CM during impact from positive to negative angles of rotation: (a) immediately before impact, when the pole of rotation is point O; (b) immediately after impact, when the pole of rotation is point O'



materials of the block and the base. For this reason, in many analyses ε is considered an independent parameter of the problem rather than been calculated from Eq. 9.

Free Vibrations

In case that the block is rocking freely ($\ddot{x}_g(t) = 0$), the linearized equation of motion (7) becomes

$$\ddot{\varphi} - p^2 \varphi = \mp p^2 \theta \quad (10)$$

where p is given by Eq. 2. Note that p is not the eigenfrequency of the system, since, due to the negative stiffness, rocking blocks do not possess eigenfrequencies in the classical sense. The solution of Eq. 10 is

$$\varphi(t) = A \cosh(pt) + B \sinh(pt) \pm \theta \quad (11)$$

where the coefficients A and B are determined from the initial conditions. For example, for an initial tilt of the block, $\varphi_0 < \theta$, the initial conditions are $\varphi(t=0) = \varphi_0$ and $\dot{\varphi}(t=0) = 0$, and Eq. 11 becomes

$$\varphi(t) = \theta - (\theta - \varphi_0) \cosh(pt) \quad (12)$$

Equation 12 describes the motion of the block around point O as it comes back to the

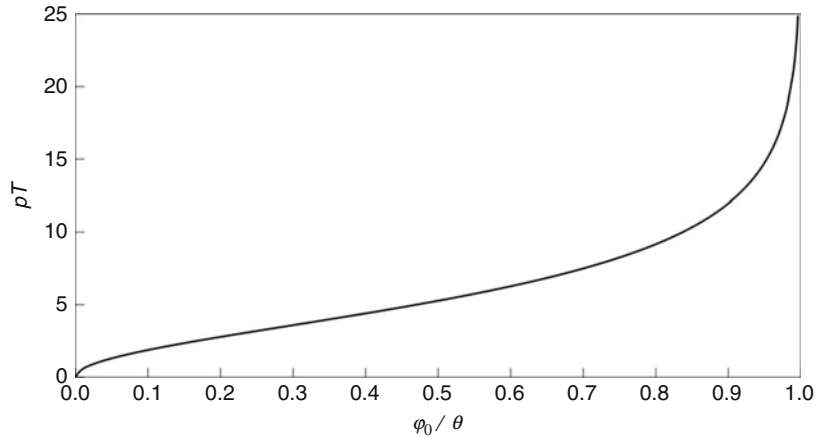
equilibrium position with positive angles of rotation. When φ becomes zero, the block hits the base and the rotation continues around point O' with the motion been described by the equation describing negative rotations. According to Eq. 8, the initial angular velocity of this motion is reduced compared with the velocity of the block when it hits the base. Due to the different equations of motion describing each regime of the response, the overall behavior is nonlinear.

If there was not any dissipation of energy during impact, the time elapsed from the angle of maximum tilt (initially equal to φ_0) up to the point when the block reaches the equilibrium position ($\varphi = 0$) would correspond to one fourth of the period T of free vibrations. Therefore, setting $\varphi = 0$ for $t = T/4$ in Eq. 12, the period of free vibrations can be calculated:

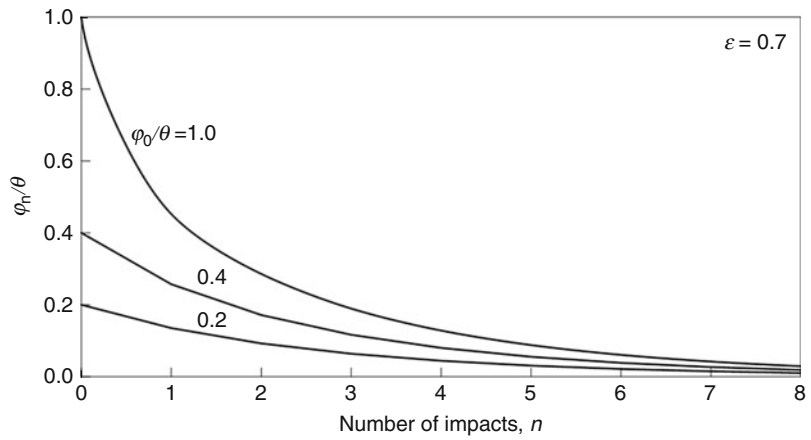
$$T = \frac{4}{p} \cosh^{-1} \left(\frac{1}{1 - \varphi_0/\theta} \right) \quad (13)$$

However, in real structures dissipation of energy does occur at each impact. For this reason, the block will attain a reduced maximum tilt after each impact, and the period of free vibrations will continuously decrease. Using Eq. 8,

Ancient Monuments Under Seismic Actions: Modeling and Analysis, Fig. 4 Dependence of the period of free vibrations on the normalized initial tilt, φ_0/θ



Ancient Monuments Under Seismic Actions: Modeling and Analysis, Fig. 5 Variation of the amplitude of free vibrations with the number of impacts for $\varepsilon = 0.7$



Housner (1963) calculated that the amplitude φ_n of the free vibrations after the n^{th} impact is

$$\frac{\varphi_n}{\theta} = 1 - \sqrt{1 - \varepsilon^n \left[1 - \left(1 - \frac{\varphi_0}{\theta} \right)^2 \right]} \quad (14)$$

The plot of the dimensionless product pT versus the normalized initial tilt angle, φ_0/θ (Eq. 13), is given in Fig. 4, while in Fig. 5 the decrease of the amplitude with the number of impacts is plotted for $\varepsilon = 0.7$ and various values of the normalized initial tilt, φ_0/θ .

Initiation of Rocking Under Earthquake Excitation

Under a seismic ground excitation $\ddot{x}_g$, rocking starts when the base acceleration reaches

a critical value, $(\ddot{x}_g)_{cr}$. At that time, the overturning moment about O or O' due to the d'Alembert inertial force, $M_A = m (\ddot{x}_g)_{cr} h$, becomes equal with the restoring moment due to the weight of the block, $M_E = mgh/2$; therefore

$$(\ddot{x}_g)_{cr} = (\tan \theta)g \quad (15)$$

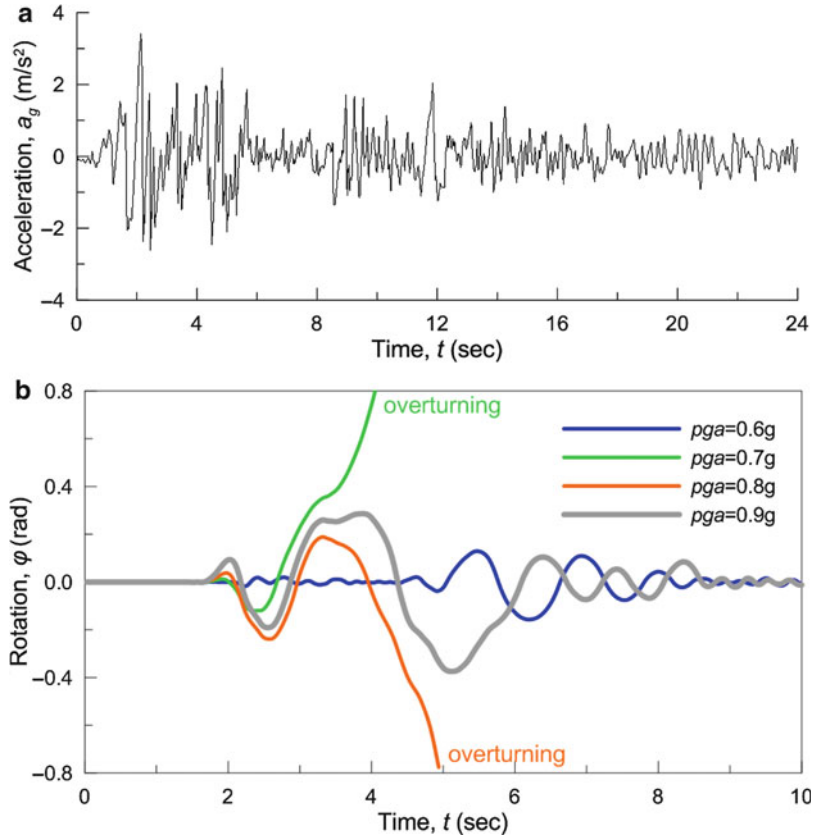
For slender blocks ($\theta < 20^\circ$), the critical base acceleration can be approximated by

$$(\ddot{x}_g)_{cr} = \theta g \quad (16)$$

If the peak acceleration of the base motion, pga , is smaller than $(\ddot{x}_g)_{cr}$, the earthquake is not strong enough to initiate rocking of the block.

Ancient Monuments Under Seismic Actions: Modeling and Analysis,

Fig. 6 (a) El Centro (1940) earthquake. (b) Rocking response of an orthogonal block of dimensions $b = 0.50$ m, $2h = 1.5$ m for the El Centro earthquake amplified to several values of pga (for $\varepsilon = 0.85$)



As evident from Eq. 15, the slenderer the block, the smaller is the required base acceleration to set it into rocking motion.

Nonlinearity and Sensitivity of the Response

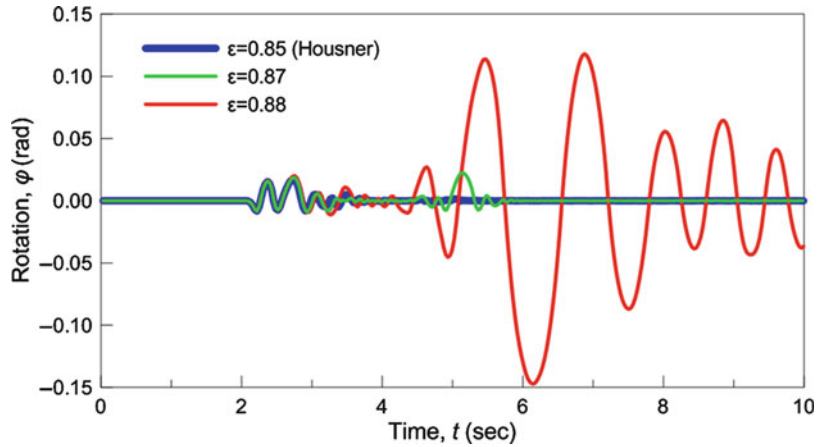
As mentioned above, the rocking response is highly nonlinear. This is illustrated in Fig. 6b, in which the time history of the angle of rotation of an orthogonal block with dimensions $b = 0.50$ m, $2h = 1.50$ m is shown for the El Centro (1940) earthquake (Fig. 6a) amplified to four different values of the peak ground acceleration: $pga = 0.60$ g, 0.70 g, 0.80 g, and 0.90 g. In all cases, the coefficient of restitution was set to $\varepsilon = 0.85$, which corresponds to Housner's theoretical value according to Eq. 9. It is seen that the response of the block is stable for $pga = 0.60$ g (blue line), while the block overturns in the direction of positive rotations for $pga = 0.70$ g (green line). Increasing the base excitation to $pga =$

0.80 g the block overturns in the opposite direction (negative rotations). However, if the base motion is amplified even more to $pga = 0.90$ g, the response is stable again and overturning does not occur (gray line).

Apart from the nonlinearity, another characteristic of the response is its sensitivity to even trivial changes of the parameters. This sensitivity has been proven by the non-repeatability of the same experiment (Yim and Chopra 1984). In Fig. 7, the sensitivity of the response of the abovementioned block to the value of the coefficient of restitution ε is shown. In this plot, the response of the block is shown for the El Centro record amplified to $pga = 0.50$ g and for three values of the coefficient of restitution: $\varepsilon = 0.85$ (Housner's value), $\varepsilon = 0.87$, and $\varepsilon = 0.88$.

It is seen that the response for $\varepsilon = 0.87$ (green line) is very similar with the one for $\varepsilon = 0.85$

Ancient Monuments Under Seismic Actions: Modeling and Analysis, Fig. 7 Rocking response of an orthogonal block of dimensions $b = 0.50$ m, $2h = 1.5$ m for the El Centro earthquake amplified to $pga = 0.50$ g for various values of the coefficient of restitution ε



(blue line), except for an additional small rocking response of the block around $t = 5$ s, which does not occur for $\varepsilon = 0.85$. However, if we slightly increase the coefficient of restitution to $\varepsilon = 0.88$, intense rocking occurs after $t = 4$ s with significantly larger amplitude than the amplitude in the time interval $2.0 < t < 3.5$ s when all the rocking response takes place for $\varepsilon = 0.85$. It is interesting to notice that this intense rocking for $\varepsilon = 0.88$ occurs after the strong motion of the ground excitation (see Fig. 6a).

It must also be noted that although, in general, a decrease in the value of ε leads to smaller rocking amplitude, due to the larger dissipation of energy during impact, it is also possible that a smaller coefficient of restitution produces larger rocking response (Aslam et al. 1980). This counterintuitive phenomenon is attributed to the nonlinearity of the response.

Main Features of the Rocking Response

Except for the abovementioned nonlinearity and the sensitivity of the response, an important feature of the dynamic behavior is its dependence on the dimensionless quantity pT_p with T_p being the predominant period of the ground motion. Actually, for harmonic excitation (which can be extrapolated to pulse-like ground motions) the normalized response can be expressed solely as a function of four dimensionless terms (Zhang and Makris 2001; Dimitrakopoulos and DeJong 2012):

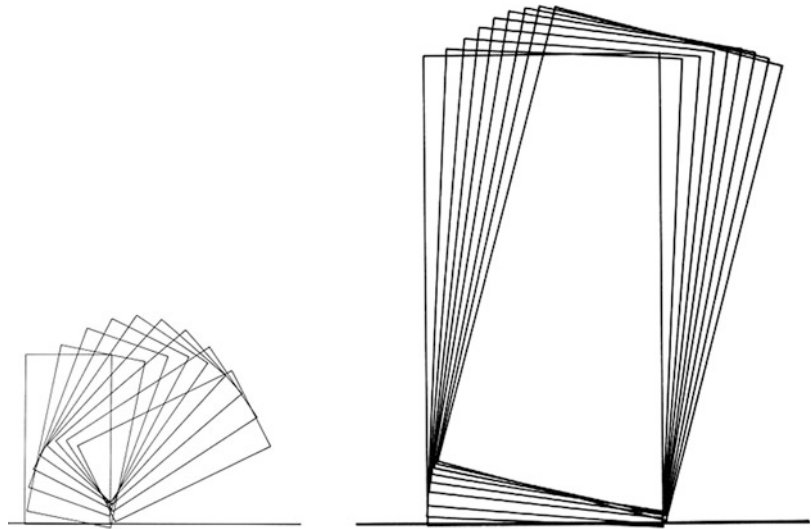
$$\frac{g \cdot \varphi_{\max}}{pga} = f\left(pT_p, \frac{g \tan \theta}{pga}, \tan \theta, \varepsilon\right) \quad (17)$$

Assuming that ε is known, Eq. 17 implies that:

- For a given base excitation (given pga and T_p), the response depends on the slenderness θ and the characteristic frequency p . The latter is a function of the size of the block. The value of p decreases inversely with the size of the block (e.g., see Eq. 5 for an orthogonal block), measured with the distance r . Therefore, for the same slenderness there is an important size effect on the response. Actually, among two blocks with the same slenderness θ but different size, the smaller one will experience more intense rocking than the larger one. This is shown in Fig. 8, in which the response of two blocks with $\tan \theta = 0.5$ but different size ($b = 0.50$ m for the left block and $b = 1.5$ m for the right) is shown for the same impulse base excitation. It is seen that the small block overturns, while the large one does not.
- For a given block (given θ and p), the rocking response and the overturning risk greatly depend on the predominant period of the base excitation. In general (for details see Zhang and Makris 2001; Dimitrakopoulos and DeJong 2012), the required normalized amplitude of the base acceleration, $pga/(g \tan \theta)$, to cause overturning decreases as the

Ancient Monuments Under Seismic Actions: Modeling and Analysis,

Fig. 8 Rocking response of two similar orthogonal blocks of the same slenderness ($\tan \theta = 0.5$) and different size (base width $b = 0.50$ m in the *left* block and $b = 1.50$ m in the *right*) for the same impulse base excitation



period T_p increases. In other words, the block is more vulnerable to long-period earthquakes than to high-frequency ones.

It should be noted that the inequality $\varphi > \theta$ is a necessary but not a sufficient condition for overturning to occur, since it is possible that the rocking angle attains temporarily values larger than θ (i.e., $\varphi_{\max} > \theta$) without overturning. Of course such cases are exceptional, since for $\varphi > \theta$ the weight of the block produces an overturning moment instead of a restoring one; thus, the block will not topple only if at the same time a quite large restoring inertial force develops due to the ground motion, capable to reverse this situation and bring the block back to stable state.

Dynamics of Multi-drum Columns and Colonnades

Difficulties and Uncertainties of the Dynamic Analysis

As mentioned in the Introduction, columns of ancient monuments are made of individual structural elements, called drums, which are put one on top of the other without mortar or any other connecting material (Fig. 9). The wooden dowels that were usually placed at the joints between the

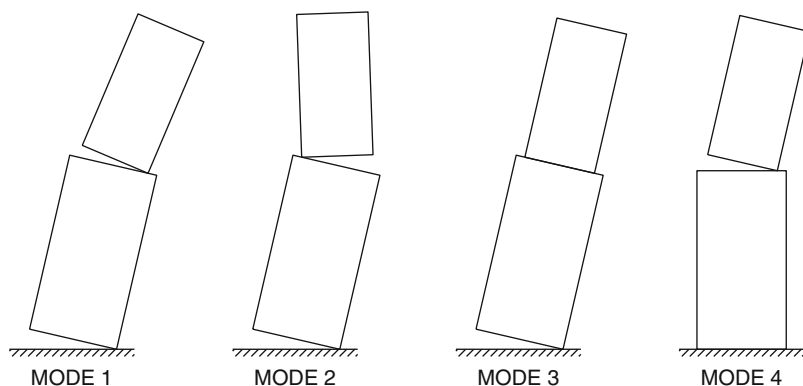


Ancient Monuments Under Seismic Actions: Modeling and Analysis, Fig. 9 Photo of a single standing and a fallen column of the Olympieion of Athens, Greece, showing the multi-drum construction of ancient columns

drums of the columns were aiming at centering the stones during construction and, practically, do not have any effect on the seismic response. In few cases steel connections (dowels) are

Ancient Monuments Under Seismic Actions: Modeling and Analysis,

Fig. 10 The four rocking “modes” of vibration of a two-block assembly (Psycharis 1990)



provided at the joints, which restrict, up to their yielding, sliding but do not affect, in general, rocking.

Due to their spiral construction, ancient columns respond to strong earthquakes with intense rocking and sliding of the drums. In addition, due to the cylindrical shape of the drums, wobbling also occurs during rocking. Usually, rocking dominates the response of multi-drum columns, which, for this reason, is characterized by the strong nonlinearity and the sensitivity discussed in the previous section. In addition, the analysis is extremely difficult and complicated due to the many “modes” of response in which multi-block systems can respond, as it will be discussed in the ensuing.

Furthermore, there is a number of other uncertainties associated with the seismic analysis of ancient monuments: the existing damage, which might be crucial to the stability of the monument to future earthquakes but is difficult to implement in the numerical models; the difficulties in representing accurately the real geometry; and the “vague” properties that must be assigned to the joints in the numerical models. All these issues will be discussed in the following; however, taking into consideration the sensitivity of the rocking response to even small changes in the parameters, one must have in mind that, no matter how accurate an analysis is, there is always an inherent uncertainty in the results.

Analysis of the Dynamic Response

The dynamic response of multi-drum columns, and, in general, of stacks of rigid blocks, is

governed by the motion of the stones, which can rock and slide individually or in groups. In case of cylindrical drums, wobbling also occurs during rocking. The dynamic analysis of such systems is a difficult task that can only be treated numerically, since there are many different “modes” in which the structure can vibrate. For example, for a system of two blocks, there are four “modes” of rocking vibrations, shown in Fig. 10 (Psycharis 1990). In case of multi-drum columns, the corresponding modes increase exponentially with the number of the drums and, for typical columns with many drums, can be tens of thousands.

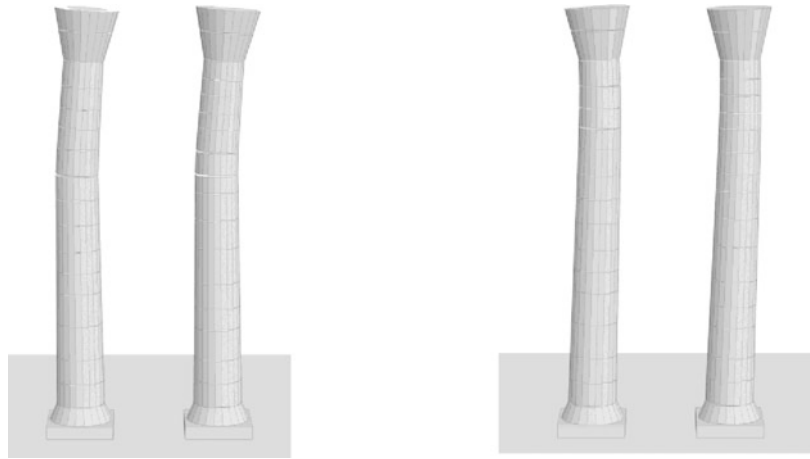
It is noted that the term “modes” of vibration is used here to denote the different patterns of the response. These “modes” of response must not be confused with the “eigenmodes” of continuous systems. Rocking structures do not possess natural modes in the classical sense.

Under an earthquake excitation, the response of the column continuously alternates from one “mode” to another (Fig. 11). Each “mode” is governed by a different set of equations of motion, while criteria must be defined for the transition between “modes.” It is evident, therefore, that sophisticated numerical codes are needed for the dynamic analysis of ancient monuments, able to take under consideration the sliding of the drums, the opening of the joints (rocking), and even complete separation and recontact of the drums.

In general, numerical models for the analysis of masonry structures may be classified into two major conceptual classes: (i) equivalent

Ancient Monuments Under Seismic Actions: Modeling and Analysis,

Fig. 11 Response of two columns of Olympieion of Athens at two different time instances during intense ground shaking. The geometry of the two columns is slightly different (the *left* has 14 drums and the *right* 15) leading to different “modes” of vibration (numerical results obtained with 3DEC software)



A

continuum models, in which the influence of the joints between blocks is introduced by means of special constitutive relations, and (ii) discontinuous models, in which the joints are represented explicitly, leading to an idealization of the structure model as a block assemblage. The finite element method is the preferred numerical tool for models of the first type, while either finite element implementations with joint elements or discrete element methods are capable of handling the second type.

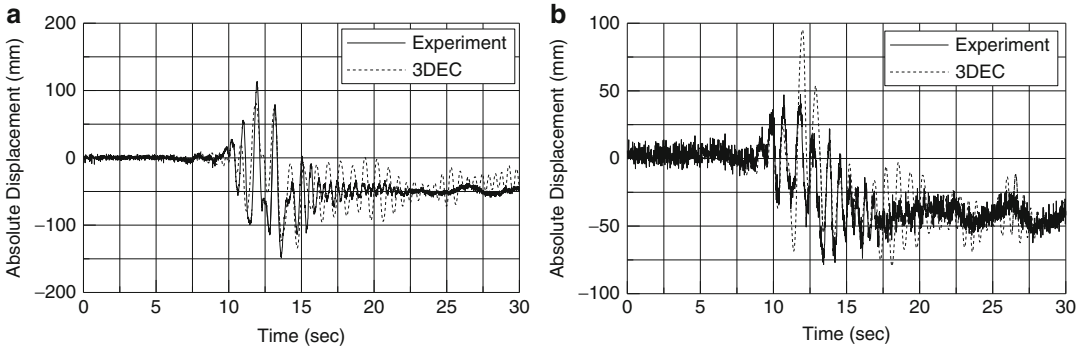
For stone masonry structures, particularly those composed of multi-drum columns and architraves as ancient monuments, block models are an obvious option. Deformation and failure of these structures is mainly governed by the relative movements between blocks. The blocks can be assumed rigid without significant loss of precision.

One method that has been proven to be very efficient in dealing with such systems is the distinct (or discrete) element method (DEM) introduced by Cundall in the 1970s in the context of rock mechanics and later extended to three-dimensional problems (Cundall 1988). The code 3DEC (Itasca 1998), which was used for most of the analyses presented herein, is based on DEM.

This method provides the means to apply the conceptual model of a masonry structure as a system of blocks. The system deformation is concentrated at the joints, where frictional sliding or complete separation may take place.

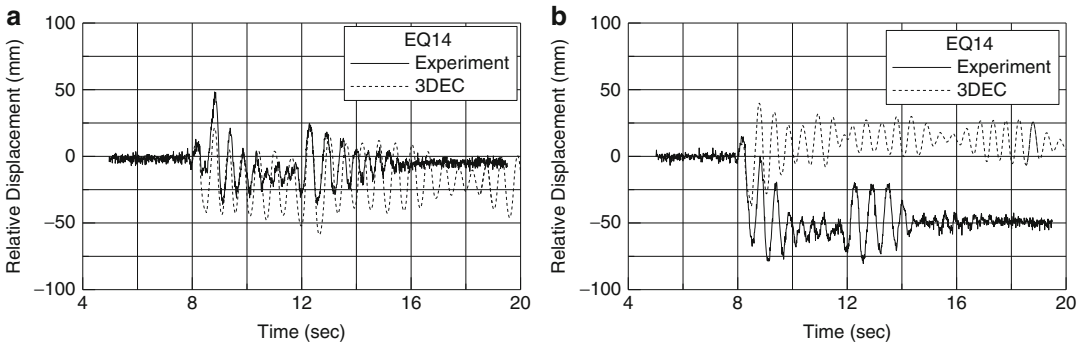
The method calculates the displacements and the forces in the individual blocks and applies compatibility laws to detect new contacts between the bodies. DEM employs an explicit algorithm for the solution of the equations of motion of the blocks, taking into account large displacements and rotations (Itasca 1998; Papantonopoulos et al. 2002).

Comparison of experimental (shaking table) data concerning the seismic response of a marble drummed replica of a column of the Parthenon in 1:3 scale with the numerical results produced by 3DEC (Papantonopoulos et al. 2002) showed satisfactory agreement in the maximum displacements during the seismic motion and in the residual deformation of the column (Fig. 12). Exact agreement could not be obtained, due to the sensitivity of the dynamic response to even trivial changes in the parameters (Mouzakis et al. 2002; Dasiou et al. 2009a). For example, repetition of the same experiment led to different results in many cases. An example is shown in Fig. 13, in which the recorded response during two similar experiments (EQ14 and EQ15) is depicted, showing significantly different residual deformation. The differences between these “similar” experiments were (i) trivial differences in the initial geometry of the column due to slightly different (not visible) relocation of the drums at their initial position and (ii) small difference (less than 2 %) in the shaking table motion. It is interesting to note that the numerical analyses also



Ancient Monuments Under Seismic Actions: Modeling and Analysis, Fig. 12 Comparison of experimental (shaking table) and numerical (3DEC) results concerning the horizontal displacement at the top of the capital of

a 1:3 replica of a column of the Parthenon under seismic excitation: (a) x-direction; (b) y-direction (Papantonopoulos et al. 2002)



Ancient Monuments Under Seismic Actions: Modeling and Analysis, Fig. 13 Horizontal displacement at the top of the capital of a 1:3 replica of a column of the

Parthenon under seismic excitation for the same experiment repeated twice (Papantonopoulos et al. 2002)

showed different response, caused by the abovementioned small difference in the base excitation (the geometry was exactly the same in the numerical models). These comparisons, and additional ones concerning more complicated structural systems comprising of three columns connected with architraves (Dasiou et al. 2009b), showed that DEM, and specifically 3DEC, can reliably predict the seismic response of classical monuments.

It must be emphasized that the application of sophisticated numerical models for the prediction of the seismic response of ancient columns and colonnades requires the knowledge of the value of several parameters which, in general, are not known a priori (this issue is discussed in detail in

the following). In this sense, experimental data of the seismic response of multi-drum columns are extremely valuable since they can be used for the calibration of the numerical models.

Modeling Aspects

Regardless of the numerical method that will be used for the analysis of the seismic response, assigning values to the parameters of the numerical model is not a straightforward procedure. Due to the sensitivity of the response, the parameters of the numerical model and the more or less accurate geometrical representation of the structure can affect the results significantly. For this reason, experimental data are generally needed in order to calibrate these parameters, especially the

ones concerning the properties at the joints. In the following, several critical aspects related to the sensitivity of the response to several parameters are discussed.

Joint Parameters

In the distinct element method, the interaction forces between two blocks are applied at a set of contact points located at the vertex-to-face and the edge-to-edge intersections. These forces depend on the relative displacement between the blocks according to the constitutive model adopted for the joints. In the normal direction, the joint behavior is governed by the normal stiffness coefficient, k_n , which relates the contact stress with the normal contact displacement. No tensile strength is considered, so this spring element is only active in compression. In the shear direction, an elastoplastic stress-displacement law is usually assumed. The elastic range is characterized by the shear stiffness, k_s , while the shear strength is governed by the Coulomb friction coefficient, μ , with no cohesive strength component.

The appropriate values to be assigned to the stiffness coefficients k_n and k_s depend on the material of the drums. Since typical values for various materials are not provided in the literature, the assignment of these values is not easy. If experimental data are available, the coefficients can be calibrated against these data. However, such data are very few and concern specific materials only (usually marble). It is noted that loose contacts at the joints due to deteriorations of the contact surfaces might also affect the results.

One way to overcome this difficulty is to calibrate k_n and k_s so that the natural period of the column for small amplitude oscillations matches the one determined by ambient vibration measurements. It is noted that for low-amplitude vibrations multi-drum columns behave like continuous systems and do possess natural modes. This approach has the advantage that the effect of existing imperfections at the joints can be included in the joint stiffness.

Concerning the coefficient of friction, it varies from a static value μ_s at the initiation of sliding to a lower value, which degrades to the kinetic value

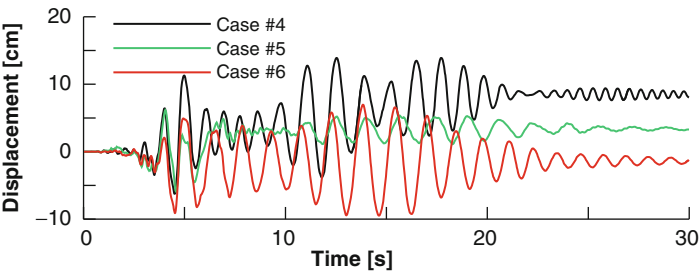
μ_k , where $\mu_k \leq \mu_s$. This change in friction depends not only on the normal stresses on the drum interface during sliding but also on the number of reversals and the amount of cumulative slip. A considerable number of laboratory test results on surfaces of limestone and marble specimens of different roughness can be found in literature. Although the test procedures differ, the evidence is that, with increasing normal stress σ_n , the static friction coefficient increases, while the difference $(\mu_s - \mu_k)$ decreases. This means that, other things being equal, it is easier for drums to slide in the upper parts of a multi-drum column than near its pedestal and also that, once sliding begins, the drop of the static coefficient μ_s to the residual value of μ_k may lead to larger and more rapid displacement in the upper parts than at lower levels of the shaft.

The proper assessment of the joint parameters is essential for the analysis, since they affect the response significantly. An example of the effect of the value of k_n and k_s on the response of a six-drum column with a two-piece capital is illustrated in Fig. 14 for three different combinations of the normal and shear coefficients (Toumbakari and Psycharis 2010). It is evident that the values of k_n and k_s affect both the amplitude of the response and the residual deformation of the column.

Significant might also be the effect of the coefficient of friction. An example is shown in Fig. 15 for the column of Olympieion of Athens (the existing dowels at the joints were not considered in this analysis) under a strong earthquake motion. Three values of the coefficient of friction were examined: $\mu = 0.55$, 0.75 , and 1.05 . It is noted that the typical value of μ for marble is about 0.70 – 0.75 ; thus, the value $\mu = 0.55$ is considered rather low, while the value $\mu = 1.05$ is unrealistically high. It is seen that the value of the coefficient of friction affected mainly the response and the residual dislocation of the drums at the upper part of the column, where the developed accelerations were large enough to initiate sliding.

Damping

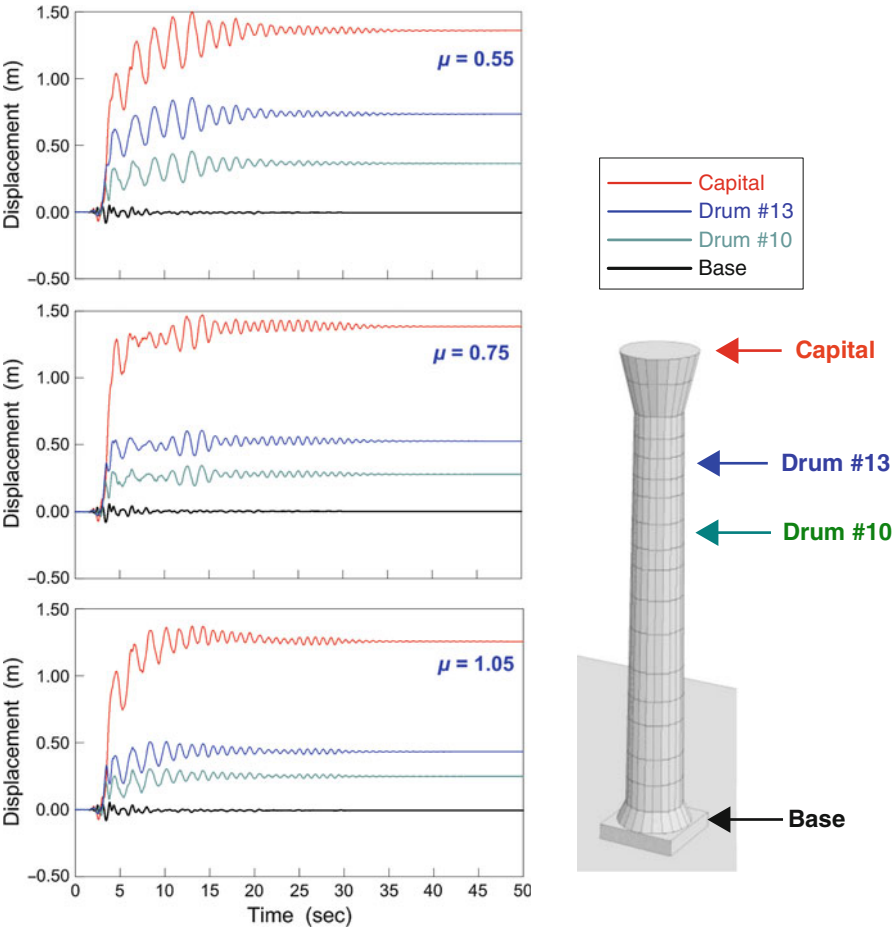
The dissipation of energy due to sliding is automatically taken under consideration with the



Case	k_n [Pa/m]	k_s [Pa/m]
4	5×10^9	1×10^9
5	0.5×10^9	0.1×10^9
6	1×10^9	0.2×10^9

Ancient Monuments Under Seismic Actions: Modeling and Analysis, Fig. 14 Time histories of the displacement at the top of the capital of a freestanding

column for three different sets of the joint stiffness (Toumbakari and Psycharis 2010)

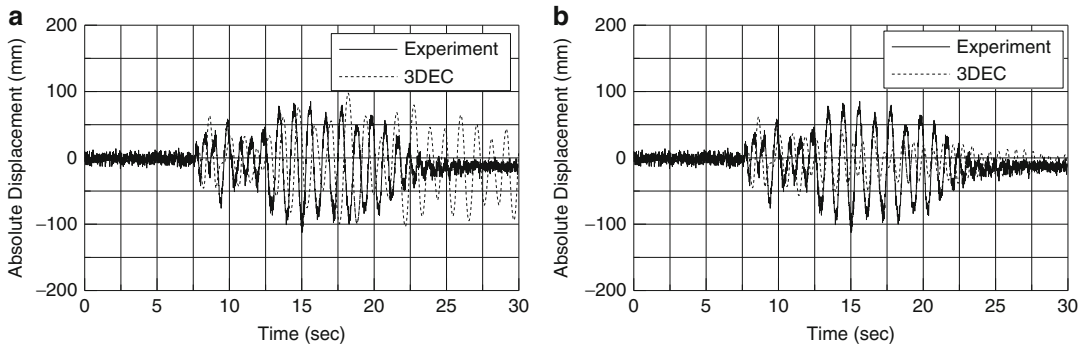


Ancient Monuments Under Seismic Actions: Modeling and Analysis, Fig. 15 Time histories of the displacement at various positions along the height of a freestanding column for three different values of the coefficient of friction

elastoplastic model of the shear stiffness. However, since elastic properties are assumed at the joints, extra damping must be introduced to the numerical model to account for the

dissipation of energy due to the impacts between drums during rocking.

Shaking table experiments on multi-drum columns showed very low attenuation. These results



Ancient Monuments Under Seismic Actions: Modeling and Analysis, Fig. 16 Comparison of numerical results with experimental data for (a) zero damping and (b) 0.5 % stiffness-proportional damping (Papantonopoulos et al. 2002)

lead to the conclusion that, to be conservative, the numerical simulations should be performed with very low or even zero damping, at least during the strong shaking. This is illustrated in Fig. 16, in which the numerical results for zero damping and 0.5 % stiffness-proportional damping are compared with experimental data (Papantonopoulos et al. 2002). It is seen that the introduction of even a small value of damping decreased unrealistically the amplitude during the strong shaking. However, it is necessary to introduce some damping towards the tail of the response in order to attenuate the free vibrations and be able to calculate the residual deformation of the column.

In general, whenever damping is used it is preferable to use the stiffness-proportional component of Rayleigh damping. However, explicit time-stepping algorithms, such as the one in 3DEC, require rather small time steps if stiffness-proportional damping is used. Thus, in order to avoid the increase in the computational effort, only the mass proportional component of Rayleigh damping can be adopted, with a low value though.

Required Accuracy of the Geometric Representation

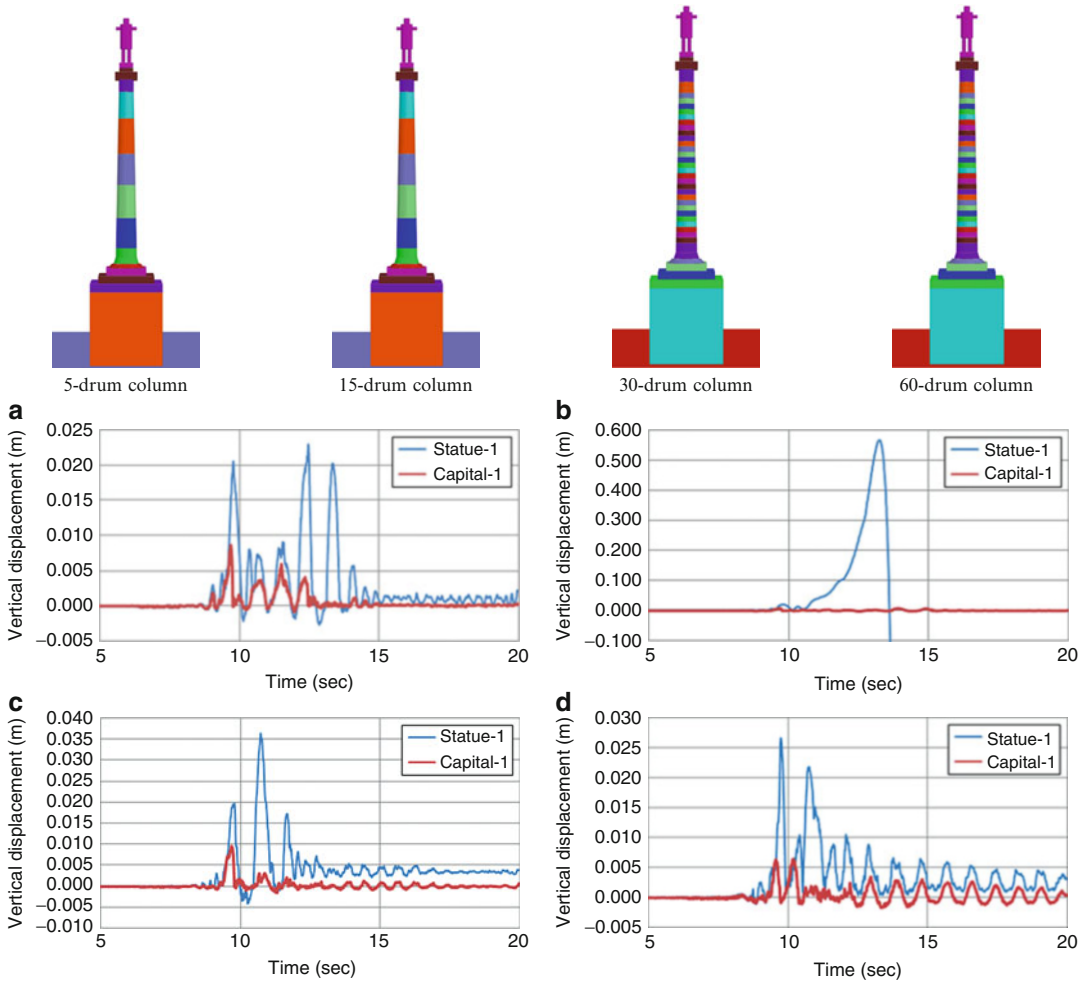
One question that arises in conducting numerical analyses of ancient monuments is how accurately the model must represent the actual geometry. As a general rule, the geometry should be implemented as accurately as possible, due to the sensitivity of the response. However, exact

modeling is practically impossible, due to the complicated geometry of the architectural details of the columns and the random shape of existing damage. In some cases, even the gross geometry (number and dimensions of drums) might not be easy to implement due to lack of information, since, in most monuments, the height of each drum and sometimes even the number of the drums of the columns are not constant (see Fig. 9). This happens because drums were made of marble of superior quality and the height of each drum depended on the available marble pieces, as ancient builders were trying to avoid any unnecessary loss of material.

In addition, numerical models containing large number of blocks might require extremely high computational time; thus, it might be desirable to simplify the model as much as possible.

An example of the effect of the number of drums considered in the analysis on the response of a column with a statue on top is given in Fig. 17 (Ambraseys and Psycharis 2011). In this case, the investigation concerned the risk of toppling of the statue. The column consisted of 15 drums, but analyses were also performed for columns of 5 drums, 30 drums, and 60 drums (see top of Fig. 17). In Fig. 17, the uplift of one corner of the statue is shown. It is seen that the response varied in each case while the statue overturned only in the case of the column with 15 drums.

It is evident therefore that the geometry must be implemented as accurately as possible in order to reach correct conclusions. However, if only gross results are sought, it might be adequate to



Ancient Monuments Under Seismic Actions: Modeling and Analysis, Fig. 17 Time histories of the vertical displacement of corner 1 of the base of the statue (*blue line*) and the capital (*red line*): (a) 5-drum column, (b)

15-drum column, (c) 30-drum column, and (d) 60-drum column. Overturning of the statue occurs only in case (b) (Ambraseys and Psycharis 2011)

apply very simple geometries, which can be analyzed much easier. An example is given in Fig. 18, in which the required amplitude of a harmonic excitation to cause overturning versus the period of the excitation (stability threshold) is shown for (a) the column of the temple of Zeus at Nemea, Greece, and (b) the column of the temple of Apollo at Bassae, Greece (Psycharis et al. 2000). The results obtained for the actual geometry of the multi-block columns are compared with the ones for the equivalent single-block columns, i.e., the monolithic columns with the same overall dimensions with the

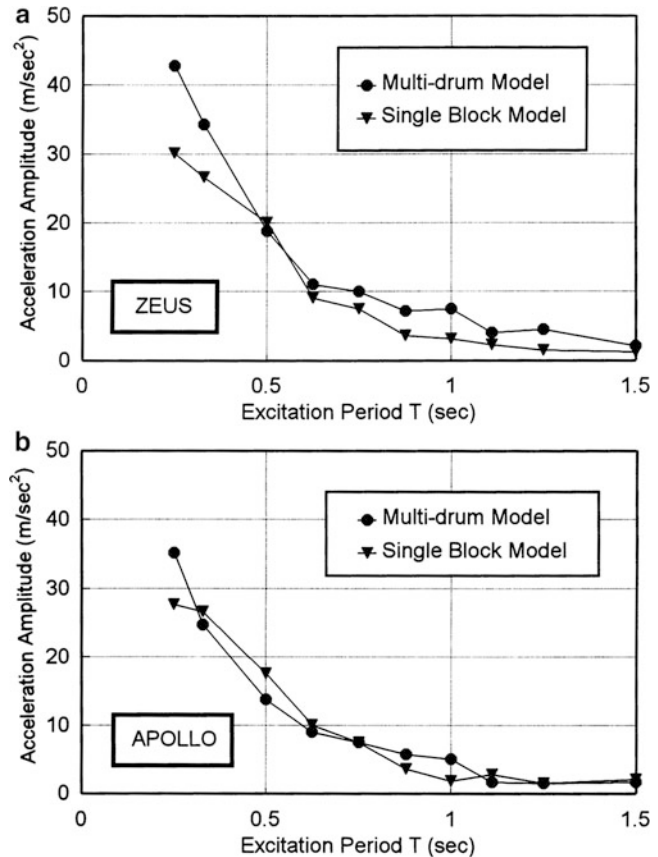
original columns. It is seen that the very simple single-block representation could predict with acceptable accuracy the overturning risk of the columns and thus it could be used as a first-order approximation within a decision making procedure.

Representation of Existing Damage

In their current condition, ruins of ancient structures present many different types of damage. Most common are missing pieces (cutoffs) that reduce the horizontal sections in contact, foundation problems resulting in tilting of the columns,

Ancient Monuments Under Seismic Actions: Modeling and Analysis,

Fig. 18 Comparison of the stability threshold for the overturning of the multi-drum column and the equivalent single-block column: (a) column of the temple of Zeus; (b) column the temple of Apollo (Psycharis et al. 2000)



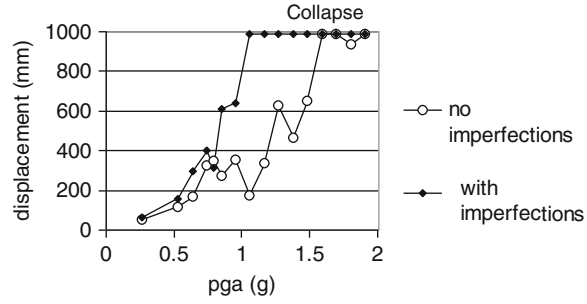
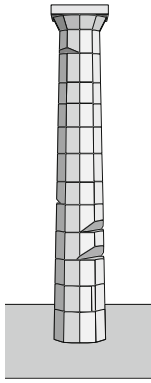
dislocated drums from previous earthquakes, and cracks in the structural elements that in some cases split the block in two. Such imperfections affect significantly the stability of the columns which, thus, are much more vulnerable to earthquake excitations compared with the original intact structures. Therefore, the abovementioned impressive stability of ancient monuments against earthquakes might not be a valid assumption any more, if significant damage is present.

An example of the effect of existing imperfections on the stability of ancient columns is shown in Fig. 19 for the column of the Parthenon in Athens (Psycharis et al. 2003). The maximum permanent displacement of the column is plotted versus the *pga* of the ground motion, and it is seen that the presence of the imperfections shown in the left drawing of Fig. 19 leads to larger displacements and significantly earlier collapse.

It is evident therefore that the damage that can be observed today in the monuments must be implemented in the numerical models. It should be mentioned, however, that this is not an easy task for several reasons: (i) because the existing damage has not always been mapped in the required detail, (ii) because the shape of missing pieces is irregular and very difficult to be modeled accurately, and (iii) because certain types of damage, as cracks in the stone blocks, are either unknown or impossible to take under consideration.

Modeling of Clamps and Dowels

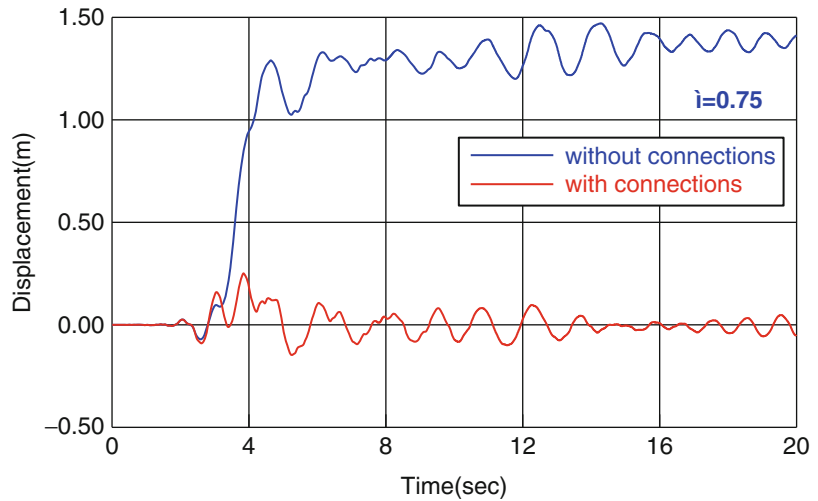
In ancient monuments, iron clamps and dowels exist between the beams of architraves and the stones of walls. In columns, iron dowels were rarely put at the joints, connecting adjacent drums. In such cases, these connectors must be included in the numerical models because they



Ancient Monuments Under Seismic Actions: Modeling and Analysis, Fig. 19 Maximum permanent displacements of a column of the Parthenon under the

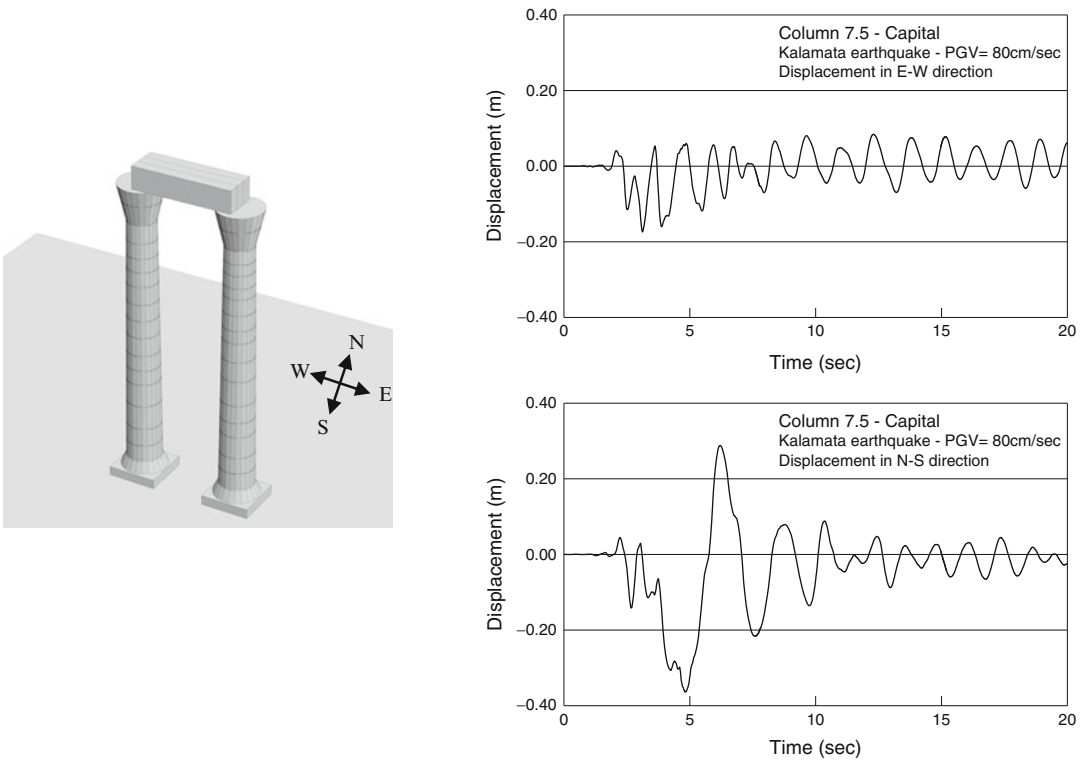
Aigion, Greece (1995), earthquake amplified to several values of *pga* without and with the imperfections shown in the *left* diagram (Psycharis et al. 2003)

Ancient Monuments Under Seismic Actions: Modeling and Analysis, Fig. 20 Displacement at the top of a column of the Olympieion of Athens when the iron dowels at the joints are considered in the analysis (*red line*) and when they are neglected (*blue line*)



affect the response significantly. An example is shown in Fig. 20, in which the displacement of the capital of the column of the Olympieion of Athens is shown: (a) when the two iron dowels of 10 cm² cross section that exist at each joint are considered in the analysis and (b) when they are neglected. The difference in the response of the column, and especially in the amount of permanent deformation, is very large, showing the generally beneficial effect of the dowels on the seismic response. It must be noted, however, that there might be cases in which the prevention of sliding of the upper drums provided by the dowels could be unfavorable to the overall behavior of the column.

In general, dowels should be modeled as nonlinear shear connectors without tensile strength. However, one must have in mind that clamps and dowels were put in inserts carved in the stone blocks and the gap was usually filled with lead. For the dowels placed between the drums of columns, filling the gap with lead was not easy, and, frequently, parts were left without filling. For this reason, it is not usually known whether the dowels are tightly fixed to the drums or small displacements are allowed. This situation introduces an uncertainty in the numerical results, since, if the dowels were loose, the response of the columns would be different.



Ancient Monuments Under Seismic Actions: Modeling and Analysis, Fig. 21 Olympieion of Athens: displacement at the top of the E column in the in plane (E–W

direction – *top* diagram) and the out of plane (N–S direction – *bottom* diagram)

It must also be noted that, for large rocking angles, it is possible that a dowel disengages from the upper drum. In that case, the dowel might not be able to reinsert in the mortise during the reversed motion due to the wobbling and the sliding of the drums, blocking thus the proper sitting of the upper drum. In most cases, however, shear dowels are inserted several centimeters in the drums, and, thus, it is not probable that rocking can cause their disengagement, even for strong ground motions.

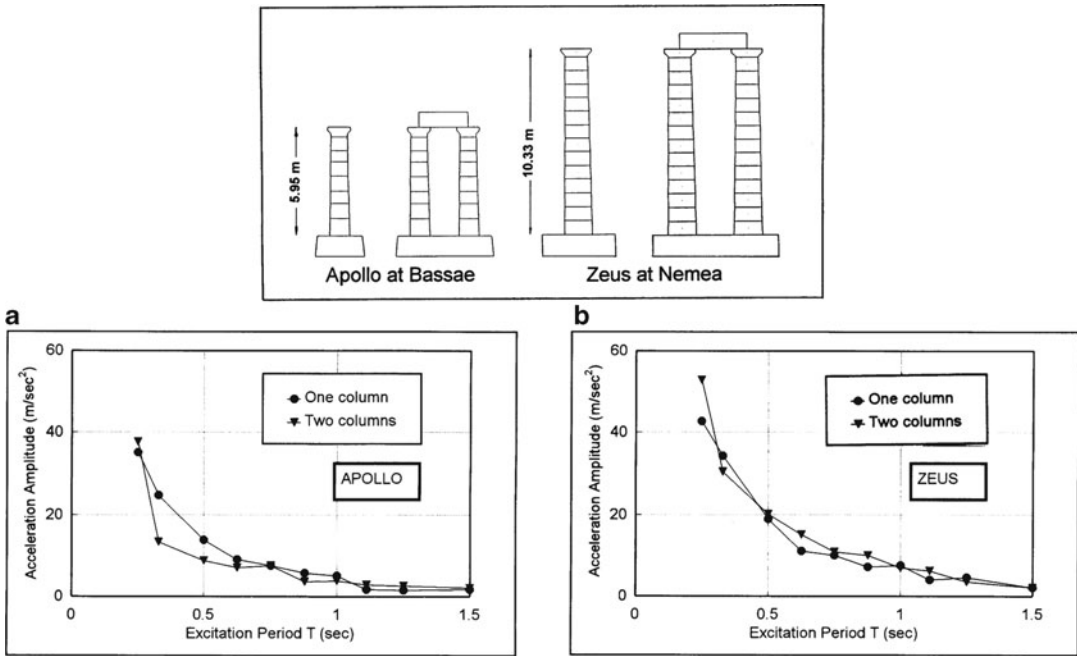
2D Versus 3D Analysis

For freestanding columns, and despite the symmetry of the column about the vertical axis, there is a significant difference in the response between two-dimensional and three-dimensional analysis. Two-dimensional analysis is unable to capture all the aspects of the real response, mainly the rotation of the drums around the vertical axis due to the simultaneous rocking in two normal

directions. For cylindrical blocks, the pole of rotation continuously changes its position, moving along the perimeter of the base of the drum (wobbling). It is interesting to note that it was observed during shaking table experiments for purely plane excitation that even small disturbances in the direction normal to the plane of rocking could cause significant amplification of the response in that direction (Papantonopoulos et al. 2002).

In general, comparisons between 2D and 3D analyses show that the 2D approach underestimates the response, predicting greater stability.

For systems involving many columns in line or in corner (colonnades), 3D analyses are evidently necessary. In such cases, it has been observed that the end column of each row suffers significant out-of-plane deformation which, for strong earthquake motions, might be significantly larger than the in-plane one. An example is shown in Fig. 21 for a set of two columns of the



Ancient Monuments Under Seismic Actions: Modeling and Analysis, Fig. 22 Stability threshold for the collapse of the freestanding column and the set of two

columns under harmonic excitations: (a) columns of the temple of Apollo; (b) columns of the temple of Zeus (Psycharis et al. 2000)

Olympieion of Athens connected with an architrave. Under a three-dimensional earthquake excitation, the displacement at the top of the east column was much larger in the out-of-plane direction (bottom diagram) than in the in-plane one (top diagram).

Size Effect

Similarly to the single rocking block (see section “Main Features of the Rocking Response”), the size of ancient monuments affects their dynamic response and their vulnerability to earthquakes, with larger structures being more stable than smaller ones. This is shown in Fig. 22, in which the minimum required acceleration amplitude of a harmonic excitation of varying period to cause collapse (stability threshold) is shown for two cases: (a) the columns of the temple of Apollo at Bassae, Greece, of height 5.95 m and (b) the columns of the temple of Zeus at Nemea, Greece, of height 10.33 m (Psycharis et al. 2000). Results are given for the freestanding column and the set of two columns. It is seen that, for the same

period of excitation, significantly larger acceleration is needed to overturn the larger columns of Zeus compared with the smaller columns of Apollo.

Another interesting observation is that the stability threshold of each monument was similar for the freestanding column and the set of two columns. This means that restoration of fallen architraves does not necessarily lead to enhanced stability of the monument against future earthquakes. Figure 22 shows that such restoration of the architraves might be favorable or unfavorable depending on the characteristics of the structure: in the case of Apollo, it was generally unfavorable; in the case of Zeus, it was generally favorable.

It should be mentioned that the above observation concerns the in-plane collapse of the columns (2D analyses). However, shaking table tests on sets of columns connected with architraves in line or in corner have shown that the architrave beams are the most vulnerable parts of monuments, as they are the first pieces that fall down

(in the out-of-plane direction). The collapse of the architraves endangers the stability of the whole monument, since it is possible that they hit the columns during their fall.

Selection of Ground Motions

The earthquake response of ancient monuments is dominated by the rocking of the drums of the columns. As mentioned in section “[Main Features of the Rocking Response](#),” rocking is greatly affected by the predominant period of the excitation. It is evident, therefore, that the vulnerability of a monument depends on the frequency content of the ground motion, with long-period earthquakes being much more dangerous than high-frequency ones. This is shown in Fig. 22, in which the stability threshold decreased exponentially as the period of excitation increased in all cases. Previous analyses (Psycharis et al. 2000) have shown that low-frequency earthquakes force the structure to respond with intensive rocking, whereas high-frequency ones produce significant sliding of the drums, especially at the upper part of the structure.

It is evident that the choice of the earthquakes that will be used in the analyses is very important, as the dynamic response of multi-drum columns and colonnades and the danger of collapse are sensitive to the energy and frequency content of the time history of the input ground motion. Apart from the abovementioned strong effect of the predominant period of the ground motion, the time sequence of the various phases in the record might also be significant. In this sense, it is essential to constrain the selection of the base excitations to what one may call suitable surrogate ground acceleration time histories that could replicate as closely as possible the time histories of past and anticipated earthquakes.

Records must be chosen from past earthquakes which are associated specifically with the tectonics and the seismicity of each region or with earthquakes from other regions of similar tectonics. They may not be always available or in sufficient numbers, but they should not be selected

arbitrarily. The identification of active or potentially active faults, including blind faults, the long-term seismic history of the region and reliable magnitudes as well as regionally representative data on strong ground motions are important parameters that must be considered in this choice. Among the important characteristics of ground motion, time histories to be used in 3D numerical analyses are their maximum velocity and associated period as well as the directivity effects which will be present depending on the proximity of the monuments to a potentially active fault.

From the geological point of view, three independent source parameters are required to describe an earthquake: the fault dimension, the static moment M_0 , and either the mean stress or energy released by faulting or strain energy change. For sites close to the causative fault, the ground motions will show pronounced directivity effects in their ground velocity and displacements. These will appear as long-period amplitude pulses linearly related to the seismic moment. Additionally, the effect of the source mechanism on directivity will be that ground accelerations due to strike-slip faulting will be larger than those due to normal faulting, while ground velocities and displacements will be larger for thrust faulting compared with normal earthquakes.

It is evident, therefore, that the choice of which time histories to include and which to exclude in order to constrain ground motions is an important decision. There is a balance to be struck between being not restrictive enough in the time histories used, leading to unreliable results and hence predictions due to errors and uncertainties, and being too restrictive, which leads to a too small set of time histories and hence nonconclusive results.

Summary

In this chapter the main characteristics of the response of ancient monuments to earthquakes and the difficulties and uncertainties encountered in the analysis, mainly in what regards the values assigned to the parameters of the numerical

model and the selection of surrogate earthquake motions, are presented and discussed. Based on the results of previous studies, the main features of the response can be summarized as follows:

- Owing to rocking and sliding, the response is nonlinear. The nonlinear nature of the response is pronounced even for the simplest case of a rocking single block. In addition, multi-drum columns can rock in various “modes,” which alternate during the response increasing thus the complexity of the problem. The word “mode” denotes the pattern of rocking motion rather than a natural mode in the classical sense, since rocking structures do not possess such modes and periods of oscillation.
- The dynamic behavior is sensitive to even trivial changes in the geometry of the structure or the base-motion characteristics. The sensitivity of the response has been verified experimentally, since “identical” experiments produced significantly different results in some cases. The sensitivity of the response is responsible for the significant out-of-plane motion observed during shaking table experiments for purely planar excitations.
- The vulnerability of the structure greatly depends on the predominant period of the ground motion, with earthquakes containing low-frequency pulses being in general much more dangerous than high-frequency ones. The former force the structure to respond with intensive rocking, whereas the latter produce significant sliding of the drums, especially at the upper part of the structure.
- The size of the structure affects significantly the stability, with bulkier structures being much more stable than smaller ones of the same slenderness.
- Classical monuments are not, in general, vulnerable to earthquakes. However, their stability might have been significantly reduced in the damaged condition that they are found today. Types of damage that might increase their vulnerability to earthquakes include cutoff of drums, displaced drums, inclined columns due to foundation failure, cracks in the stones, etc.

Cross-References

- [Ambient Vibration Testing of Cultural Heritage Structures](#)
- [Archeoseismology](#)
- [Damage to Ancient Buildings from Earthquakes](#)
- [Nonlinear Dynamic Seismic Analysis](#)
- [Seismic Actions due to Near-Fault Ground Motion](#)
- [Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring](#)
- [Seismic Strengthening Strategies for Heritage Structures](#)
- [Seismic Vulnerability Assessment: Masonry Structures](#)
- [Selection of Ground Motions for Response History Analysis](#)
- [Sustained Earthquake Preparedness: Functional, Social, and Cultural Issues](#)

References

- Allen RH, Oppenheim IJ, Parker AP, Bielak J (1986) On the dynamic response of rigid body assemblies. *Earthquake Eng Struct Dyn* 14:861–876
- Ambraseys N, Psycharis IN (2011) Earthquake stability of columns and statues. *J Earthquake Eng* 15:685–710
- Aslam M, Godden WG, Scalise DT (1980) Earthquake rocking response of rigid bodies. *J Struct Div ASCE* 106(ST2):377–392
- Cundall PA (1988) Formulation of a three-dimensional distinct element model – part I: a scheme to detect and represent contacts in a system composed of many polyhedral blocks. *Int J Rock Mech Min Sci* 25:107–116
- Dasiou ME, Mouzakis HP, Psycharis IN, Papantonopoulos C, Vayas I (2009a) Experimental investigation of the seismic response of parts of ancient temples. In: Prohitech conference, Rome, 21–24 June 2009
- Dasiou ME, Psycharis IN, Vayas I (2009b) Verification of numerical models used for the analysis of ancient temples. In: Prohitech conference, Rome, 21–24 June 2009
- Dimitrakopoulos EG, DeJong MJ (2012) Revisiting the rocking block: closed form solutions and similarity laws. *Proc R Soc A* 468(2144):2294–2318
- Housner GW (1963) The behaviour of inverted pendulum structures during earthquakes. *Bull Seismol Soc Am* 53:403–417
- Itasca Consulting Group (1998) 3DEC – universal distinct element code, Minneapolis

- Kirkpatrick P (1927) Seismic measurements by the overthrow of columns. *Bull Seismol Soc Am* 17(2):95–109
- Konstantinidis D, Makris N (2005) Seismic response analysis of multidrum classical columns. *Earthquake Eng Struct Dyn* 34:1243–1270
- Milne J (1885) Seismic experiments. *Trans Seismol Soc Jpn* 8:1–82
- Milne J, Omori F (1893) On the overturning and fracturing of brick and columns by horizontally applied motion. *Seismol J Jpn* 17:59–86
- Mouzakis H, Psycharis IN, Papastamatiou DY, Carydis PG, Papantonopoulos C, Zambas C (2002) Experimental investigation of the earthquake response of a model of a marble classical column. *Earthquake Eng Struct Dyn* 31:1681–1698
- Papaloizou L, Komodromos P (2009) Planar investigation of the seismic response of ancient columns and colonnades with epistyles using a custom-made software. *Soil Dyn Earthquake Eng* 29:1437–1454
- Papantonopoulos C, Psycharis IN, Papastamatiou DY, Lemos JV, Mouzakis H (2002) Numerical prediction of the earthquake response of classical columns using the distinct element method. *Earthquake Eng Struct Dyn* 31:1699–1717
- Priestley MJN, Evison RJ, Carr AJ (1978) Seismic response of structures free to rock on their foundation. *Bull N Z Natl Soc Earthquake Eng* 11(3):141–150
- Psycharis IN (1990) Dynamic behaviour of rocking two-block assemblies. *Earthquake Eng Struct Dyn* 19:555–575
- Psycharis IN, Papastamatiou DY, Alexandris AP (2000) Parametric investigation of the stability of classical columns under harmonic and earthquake excitations. *Earthquake Eng Struct Dyn* 29:1093–1109
- Psycharis IN, Lemos JV, Papastamatiou DY, Zambas C, Papantonopoulos C (2003) Numerical study of the seismic behaviour of a part of the Parthenon Pronaos. *Earthquake Eng Struct Dyn* 32:2063–2084
- Psycharis I, Fragiadakis M, Stefanou I (2013) Seismic reliability assessment of classical columns subjected to near-fault ground motions. *Earthquake Eng Struct Dyn* 42:2061–2079
- Sinopoli A (1989) Dynamic analysis of a stone column excited by a sine wave ground motion. *Appl Mech Rev Part 2* 44:246–255
- Toumbakari EE, Psycharis IN (2010) Parametric investigation of the seismic response of a column of the Aphrodite Temple in Amathus, Cyprus. In: 14th European conference on earthquake engineering, Ohrid, 30 Aug 2010–3 Sept 2010
- Winkler T, Meguro K, Yamazaki F (1995) Response of rigid body assemblies to dynamic excitation. *Earthquake Eng Struct Dyn* 24:1389–1408
- Yim CS, Chopra AK (1984) Dynamics of structures on two-spring foundation allowed to uplift. *J Eng Mech ASCE* 110(7):1124–1146
- Zhang J, Makris N (2001) Rocking response of free-standing blocks under cycloidal pulses. *J Eng Mech ASCE* 127:473–483

Archeoseismology

Manuel Sintubin

Department of Earth and Environmental Sciences, Geodynamics and Geofluids Research Group, KU Leuven, Leuven, Belgium

Synonyms

Archeological seismicity; Earthquake archeology; Seismic archeology

Introduction

In 1991 an international conference was held in Athens (Greece), marking the beginning of the modern research field of **archeoseismology**, described as “*the study of ancient earthquakes from the complementary standpoints of their social, cultural, historical and physical effect*” (Stiros and Jones 1996). Besides the term archeoseismology, also the term **seismic archeology** was introduced to emphasize the use of archeological methods in the quest to better understand the effects of earthquakes on historical buildings and archeological remains. Moreover, in analogy with historical seismicity, also the term **archeological seismicity** was suggested.

Archeoseismology can thus be defined as **the interdisciplinary study of ancient earthquakes through evidence in the archeological record, such as destruction layers, structural damage to man-made constructions, cultural piercing features, indications of repairs, abandonment, cultural changes, etc.** Archeoseismology is thus seen as a subdiscipline of paleoseismology with a particular focus on man-made constructions as a potential source of paleoseismological information covering the last few millennia. By doing so, archeoseismology serves objectives proper to seismology and earthquake geology, i.e., parameterizing of earthquakes in an effort to assess the seismic hazard in a region.

Earthquake archeology can be considered as a synonym for archeoseismology. But earthquake archeology can also be seen to serve objectives proper to archeology, i.e., reconstructing human history, in particular attempting to better understand the true impact of earthquakes on human history. The term earthquake archeology can be traced to the Japanese term *jishin kōkogaku*, referring to a research field developed in the mid-1980s in Japan primarily through the initiative of Sangawa Akira, a geomorphologist at the Geological Survey of Japan, focusing on sediment deformation features within archeological contexts (Barnes 2010).

Since the book *Archeoseismology* (Stiros and Jones 1996), a series of special issues of journals has reflected the evolution of the burgeoning discipline over the last two decades towards an ever increasing multidisciplinary discipline (McGuire et al. 2000; Galadini et al. 2006a; Caputo and Pavlides 2008; Sintubin et al. 2010; Silva et al. 2011).

In the current entry, **archeoseismology** is considered as a discipline belonging to the broad research realm of **earthquake sciences**, reflected by a continuum of overlapping and complementary research fields, each focusing a particular source of earthquake data, applying appropriate methods and techniques, and targeting a specific time window. In this respect, archeoseismology bridges the gap between **instrumental** and **historical seismology** on the one side and **paleoseismology** and **earthquake geology** on the other. Archeoseismology focuses on cultural material data spanning the last few millennia. It shares, however, a common goal with the other disciplines, i.e., a better understanding of the earthquake history within a region in an attempt to assess the seismic hazard and mitigate the seismic risk. Most valuable contribution of archeoseismology to seismic hazard assessment is situated in earthquake-prone regions with a long and lasting cultural heritage. Seismic-hazard practitioners are confronted with the problem that the instrumental record is too short (only spanning somewhat over a century) and the historical record too incomplete or even inexistent. By having the potential of determining

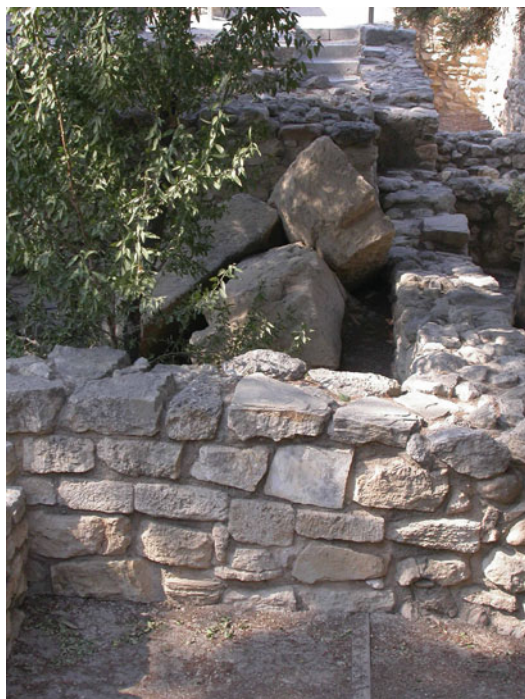
earthquake activity over millennial time spans, archeoseismology can indeed extend the archive of earthquakes beyond written sources, thus becoming a legitimate and complementary source of seismic-hazard information.

After a short historical note, a summary of the different types of archeological evidence for ancient earthquakes, commonly used in archeoseismology, is given. Subsequently, the strengths, challenges, and pitfalls of archeoseismology are discussed. Some new developments in archeoseismology will be introduced. In conclusion, some issues and perspectives in archeoseismology are presented.

A Historical Note

In the first volume of the *Palace of Minos*, published in 1921 (Evans 1921), Sir Arthur Evans did not mention earthquakes as a possible cause for the destructions observed during excavation works at the Bronze Age, Minoan site of Knossos (Crete, Greece). In the second volume of the *Palace of Minos*, published in 1928 (Evans 1928), though, tectonic earthquakes became the primary destructive agent, not only leaving a clear marker horizons in the archeological stratigraphy but also causing cultural change as evidenced by discontinuities in ceramic style and architecture. So, what happened in those 7 years that completely changed Evans' thinking?

On 20 April 1922, during the excavation of the “*House of the Sacrificed Oxen*” and the “*House of the Fallen Blocks*” (Fig. 1) at Knossos, Evans experienced an earthquake, leading him to believe that earthquakes of tectonic nature (so not related to the volcanic activity of the Thera/Santorini volcano) may very well have caused damage to the Minoan buildings during the Bronze Age. But only after experiencing the dramatic earthquake that hit the Eastern Mediterranean on 26 June 1926 and caused severe damage in the region of Heraklion, Sir Arthur Evans got convinced that earthquakes are the primary destructive agent responsible for the main stages of destruction observed at Knossos. He developed a seismic archeological stratigraphy,



Archeoseismology, Fig. 1 The “House of the Fallen Blocks” at Knossos, of which the particular context of the massive blocks inspired Sir Arthur Evans that a major earthquake may have been responsible for this damage (cf. Sintubin 2011) © Sintubin

marked with a number of earthquake-related destruction horizons (cf. Jusseret and Sintubin 2013). This work of Sir Arthur Evans can thus indeed be seen as the earliest attempt to introduce earthquakes into archeological contexts.

Evans’ ideas inspired a number of his colleagues around the Eastern Mediterranean, in particular Claude Schaeffer, excavator of the Bronze Age sites of Ugarit (Syria) and Enkomi (Cyprus). In his book *Stratigraphie Comparée et Chronologie de l’Asie Occidentale*, published in 1948 (Schaeffer 1948), Claude Schaeffer went even one step further by correlating archeological destruction layers attributed to earthquakes between Bronze Age archeological sites throughout the Asia Minor, the Caucasus, and the Middle East, setting the stage for the myths of regional earthquake catastrophes. Incorporating modern concepts of seismic storms, the myth of the Late Bronze Age seismic paroxysm around 1200 BC endured to date (e.g., Nur and Cline 2000).

Ever since, earthquakes became all too easy a “*deus ex machina*” to explain to otherwise inexplicable at archeological sites, eventually even to add drama to a site’s history. While skeptical earthquake scientists portrayed this indiscriminate use of earthquakes as neocatastrophism, advocates see the earthquake hypothesis as the simplest solution, referring to Occam’s razor. Therefore, many earthquake scientists still question the basic principles and practices of archeoseismology.

Ancient Earthquakes

Earthquakes that form the subject of archeoseismology are defined as **ancient earthquakes**, i.e., pre-instrumental earthquakes that can only be identified through indirect evidence in the archeological record (e.g., Sagalassos earthquake; cf. Similox-Tohon et al. 2005). Earthquakes that are documented in the historical record may be included if they left marks in the archeological record.

Earthquakes can basically be subdivided in **instrumental** and **pre- or noninstrumental**; the former are instrumentally recorded by seismometers, while the latter are indirectly recorded. The instrumental record covers a little more than a century since the first modern seismometers were designed in the 1890s.

Of **instrumentally recorded earthquakes**, all physical parameters (e.g., magnitude, seismic source, epicenter, duration, intensity distribution, sequence of aftershocks) can be derived. These earthquakes are the main subject of **seismology**. Recent earthquakes can have a major impact on cultural heritage (e.g., 2003 Bam M_W 6.6 earthquake), historical buildings (e.g., 2009 L’Aquila M_W 6.3 earthquake), and/or archeological remains. The latter earthquakes, affecting archeological remains, are though excluded from the field of archeoseismology.

Pre- or noninstrumental earthquakes are only indirectly recorded, in the historical, archeological, and/or geological record. Earthquake parameters that can be derived from these records concern macroseismological parameters, such as intensity, macroseismic epicenter, date, etc.

Historical earthquakes are pre-instrumental earthquakes of which information can be found in all types of historical – written – records (e.g., reports, epigraphy, epitaphs). All this historical information is compiled into earthquake catalogues. Archeoseismology can complement the knowledge with respect to historical earthquakes, when the evidence of specific, well-documented, historical earthquakes (e.g., 365 AD Crete earthquake) on archeological sites is searched for. Also, paleoseismology can add information to better constrain the macroseismological parameters of historical earthquakes.

Prehistorical earthquakes are earthquakes that have no historical record. They can both be ancient earthquakes or paleo-earthquakes.

Paleo-earthquakes, finally, can be interpreted widely, incorporating all prehistorical earthquakes, and even historical earthquakes. In this respect, it becomes synonymous to pre-instrumental earthquakes. One can opt to narrow down the definition of paleo-earthquakes to earthquakes of which evidence is only found in the geological and/or geomorphological record, being subject of **paleoseismology**. These earthquakes are also called **fossil earthquakes**.

Archeological Evidence for Ancient Earthquakes

Archeoseismology calls upon archeological material, ranging from a single occupation horizon within a Holocene stratigraphical context (e.g., Tuttle and Schweig 1995) to a widespread archeological site with monumental buildings (e.g., Similox-Tohon et al. 2006). Methodological developments in archeoseismology is indeed primarily grafted on archeological work in the Eastern Mediterranean and the Middle East, which depends strongly on identifying structural damage to monumental buildings and other cultural remains at archeological sites (e.g., Stiros 1996).

There are two limiting factors to archeoseismological investigations. The first is related to temporal aspects of the archeological record. On the one hand, the time span of occupancy of a site

largely determines the archeoseismological potential of a site (cf. Sintubin and Stewart 2008). It is obvious that the longer the site's occupancy, the higher chances are that a major earthquake has affected the site and left its marks in the archeological record. On the other hand, the archeological record is not evenly distributed through time, to a large extent dependent on socioeconomic, political, and cultural conditions of an ancient society. Secondly, archeoseismological work is limited to archeological sites, which are commonly rather a rare occurrence. There is though a remarkable spatial bias advantageous to archeoseismology, due to a "*fatal attraction*" (Jackson 2006). On the one hand, there appears to be a close relationship between tectonically active environments – thus prone to earthquakes – and ancient civilizations along the southern boundary of the Eurasian Plate (Force and McFadgen 2010). On the other hand, many settlements are founded in seismic landscapes (Michetti and Hancock 1997), thus in the direct proximity of active earthquake faults, as convincingly illustrated by numerous archeological sites throughout the Mediterranean and the Middle East. **Seismic landscapes** are defined as the cumulative geomorphological and stratigraphical effect of signs left on the environment by its past earthquakes over a geologically recent time interval (Michetti and Hancock 1997).

The archeological record can be used in basically three ways to help confront the seismic-hazard threat. First, where archeological relics are displaced, they can be used to find active faults, show in which direction faults slipped during the earthquake(s), and establish comparative fault-slip rates. Second, archeological evidence can date episodes of faulting and shaking. Third, ancient signs of earthquake-related damage, commonly related to ground shaking or ground instabilities (e.g., liquefaction), can be searched for.

Structural Damage Due to Surface Rupturing or Ground Failure

The most obvious and straightforward archeological evidence of fault activity – and thus of earthquakes – are archeological remains that are



Archeoseismology, Fig. 2 The wall of the *Crusader Fortress of Vadum Iacob* (Ateret, Israel) (cf. Marco et al. 1997) is displaced *left* laterally over more than

2 m. The fortress has been built astride the Dead Sea transform plate boundary between the African and Arabian plates © Sintubin

partly displaced due to coseismic surface rupturing on an active fault. As already mentioned, it is definitively not a coincidence that archeological sites are astride active faults.

As cultural piercing features, these **faulted relics** not only serve to identify active faults, but they are also used to determine the type of faulting (normal, reverse, strike slip), the amount of coseismic slip related to a single earthquake, as well as the cumulative fault slip. All these data eventually allow to derive time-averaged fault-slip rates over time spans of centuries to millennia, very comparable to paleoseismological work. The fault-slip rates obtained from an archeological context can subsequently be compared to long-term slip rates from paleoseismological work, enabling the evaluation of potential slip deficits and thus increased seismic hazards.

The most spectacular cases of such faulted relics can be found astride strike-slip faults, in which case there can be no doubt of the tectonic nature of the displacement. Notorious examples are the Crusader Fortress of Vadum Iacob (Israel) (Fig. 2) (cf. Marco et al. 1997) and the Al Harif Roman aqueduct (Syria) (cf. Sbeinati et al. 2010), both located astride the Dead Sea fault, a left-lateral transform plate boundary between the African and Arabian plates. By the way, as long, linear structures, aqueducts are ideal cultural piercing features because chances are they cross more than one fault. Also in extensional settings, giving rise to the typical seismic landscapes of the Mediterranean (from Turkey to Spain), faulted relics are often encountered in archeological sites (Fig. 3) (e.g., Similox-Tohon et al. 2006). In these cases of dip-slip, normal-fault movements,



Archeoseismology, Fig. 3 The mosaic floor of the Roman *Neon Library* at *Sagalassos* (SW Turkey) has been cut by a normal fault, of which a post-363 AD

activity (*terminus post quem*) can be inferred based on the archeological evidence (cf. Similox-Tohon et al. 2006)
© Sintubin

particular caution should be paid to preclude gravitational mass movement that may or may not be seismically triggered (ground failure). Complementary and independent, usually paleoseismological and/or geophysical, evidence to support a tectonic nature of the normal displacement, observed in the faulted relics, is therefore imperative (e.g., Similox-Tohon et al. 2005).

Besides these on-fault effects, numerous off-fault, coseismic ground-failure features can find their way into the archeological record, such as landslides and rockfalls, subsidence and uplift, and liquefaction (see Rodríguez-Pascua et al. 2011 and references therein). The latter sediment-deformation features form the focus of earthquake archeology Japan style (Barnes 2010).

Archeological Destruction Horizons

An **archeological destruction horizon** (destruction layer/destruction deposit) designates a horizon in the archeological stratigraphy showing evidence of sudden destruction caused by human (e.g., war, vandalism) and/or natural agents (e.g., earthquake, storm, flood). These destruction horizons commonly occur on top of “living surfaces,” evidenced by, e.g., in situ broken vases (Fig. 4), buried valuable objects, and/or skeletons of victims. Other criteria are, e.g., burned material, charcoal, collapsed architectural debris, and crushed, toppled objects.

The use of destruction horizons in archeoseismology goes back to the original work of Sir Arthur Evans in Minoan Knossos (Bronze Age Crete), inspiring several generations of archeologists who, often too easily



Archeoseismology, Fig. 4 In situ broken vessels, evidencing a collapse on a “living surface,” as part of a destruction horizon at the Minoan site of *Sissi* (Crete)

that has been attributed to an earthquake in the thirteenth century BC (cf. Jusseret and Sintubin 2012) © J. Driessen

(“*deus ex machina*”), attributed destruction horizons to catastrophic earthquakes. Identifying the true agent, eventually responsible for the destruction horizon, is rarely clear and unambiguous and remains one of the major challenges of archeoseismology.

Destruction horizons, which can with a high degree of certainty be related to ground failure and/or ground shaking caused by an earthquake, are the most appropriate “proxies” for ancient earthquakes in archeological contexts dominated by rubble architecture and associated stratigraphy, such as the Bronze Age civilizations around the Mediterranean and the Harappan civilization in the Indus Valley. In these contexts, no appeal can indeed be made to structural damage evidence on monumental buildings and constructions.

Besides evidencing ancient earthquakes, material (e.g., charcoal) and artifacts (e.g., ceramics, coins) included in the destruction horizons can be used to date episodes of earthquake damage, by means of, e.g., radiocarbon dating, changes in ceramic styles, and numismatics. A major drawback is the temporal resolution of these dating methods with uncertainties ranging from decennia to centuries. On the one hand, this does not allow to pinpoint a destruction horizon to a specific historical earthquake. On the other hand, imprecise age control leads to discrete multiple earthquakes, among which the aftershock sequence of a major earthquake, being amalgamated to “oversized” earthquake catastrophes. Finally, weak time constraints on destruction horizons hamper any reliable territorial correlation of destruction horizons between



Archeoseismology, Fig. 5 Recycling of building material in repair works of a wall at *Susita* (Golan Heights) (cf. Sintubin 2011), possibly after a destructive earthquake © Sintubin

archeological sites, again giving rise to the danger of amalgamating regionally distinct earthquakes.

Other issues concerning destruction horizons are preservation and disturbance. Little is known about the way ancient societies coped with the aftermath of a major earthquake. Earthquake debris may have been cleared from streets and buildings and disposed at particular dumpsites, so that the earthquake destruction horizon is no longer preserved in the archeological record. Valuable and/or victims may have been recovered from the debris, while material may have been reused in rebuilding, leaving behind a highly disturbed earthquake destruction horizon. It is therefore fair to conclude that the visibility of earthquake destruction in the archeological stratigraphy is rather dependent on social factors than on physical parameters (cf. Jusseret and Sintubin 2012).

Finally, other archeological evidences, such as repairs, recycling of building materials (Fig. 5), complete or partial abandonment, and architectural and/or cultural changes, can further contribute to the identification of ancient earthquakes.

Structural Damage Due to Ground Shaking and Ground Motion

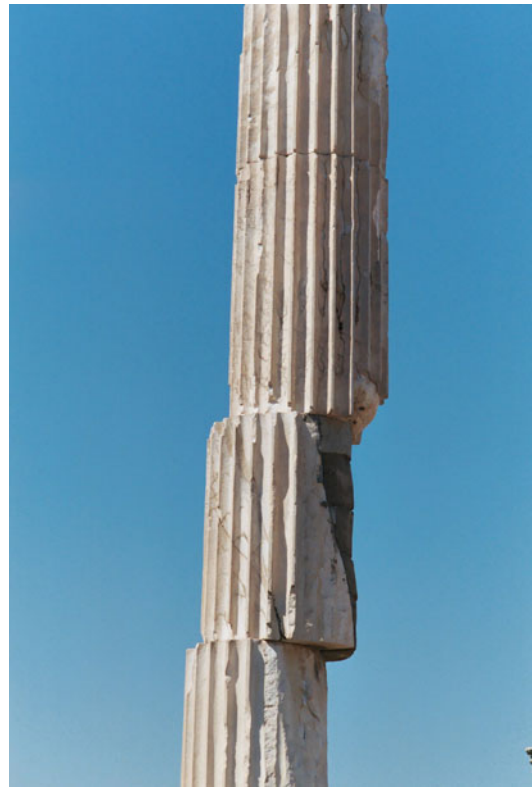
A third type of earthquake evidence in the archeological record are typical **strain structures in the building fabric** that are primarily caused by coseismic ground shaking and ground motion. These earthquake-related damage features are most conspicuous on monumental buildings and constructions, such as temples, fountains, theaters, basilicas, pavements, columns, statues, aqueducts, etc.

A first systematic inventory of possible earthquake-related damage typologies has been

introduced by Stiros (1996). Currently all these potential earthquake indicators are compiled in a comprehensive classification of **Earthquake Archeological Effects** (EAEs) (Rodríguez-Pascua et al. 2011), completely based on the guidelines of the Earthquake Environmental Effects (EEEs) that are used in the framework of the macroseismic Environmental Seismic Intensity scale (ESI2007, Michetti et al. 2007). These guidelines prescribe the difference between “primary (direct) effects” and “secondary (indirect) effects.” In archeological contexts, the latter effects of which evidence can be found in the archeological record and in particular destruction horizons, reflect the way an ancient community copes with the consequences of an earthquake affecting their settlement. Besides the on-fault (e.g., surface rupturing) and off-fault (e.g., liquefaction) geological effects, the former effects are specifically recorded in strain structures in the building fabric.

Different types of strain structures can, on the one hand, be generated by coseismic ground motion: folded and fractured pavement; shock breakouts in flagstone pavement; tilted, rotated, displaced, and bent walls, etc. On the other hand, other types of strain structures in the building fabric can be generated by coseismic ground shaking: penetrative fractures in masonry walls and columns; rotated, displaced, and ejected masonry blocks in walls; rotated and displaced drums in columns (Fig. 6); dropped keystones in arches (Fig. 7); rotated steps in stairways; collapsed stairways; folded curbs; domino-type collapsed walls and columns; directional collapse of columns; collapsed vaults; impact markings on pavements; dipping broken corners and chipping marks; U-shaped gaps in walls, etc.

The obvious difficulty with these building fabric effects is that it remains very challenging to distinguish between damage caused by an earthquake and damage caused by other destructive physical or human agents, such as natural failure of the foundations, vandalism, or warfare. Moreover, standing or partially standing buildings and constructions at archeological sites in earthquake-prone regions most probably experienced ground shaking from numerous – minor to



Archeoseismology, Fig. 6 Displaced drums of a column in the *Temple of Aphrodite* in the Roman city of *Aphrodisias* (SW Turkey), as a potential earthquake archeological effect © Sintubin

moderate to major – earthquakes over the life span of the structure, which makes it nearly impossible to attribute specific building fabric effects to a particular earthquake. Working with these damage typologies, possibly indicative of ground shaking, the investigator has to be very cautious not to “overinterpret” the potential earthquake archeological effects. Uncertainties, inherent to any archeologically based earthquake hypothesis, should moreover be assessed properly (cf. Sintubin and Stewart 2008).

Parameterization of Ancient Earthquakes

Besides the inherent ambiguities and uncertainties of archeological earthquake evidence, primarily resulting from the difficulty to



Archeoseismology, Fig. 7 A dropped keystone in the *Roman Baths at Sagalassos* (SW Turkey), interpreted as a potential earthquake archeological effect (cf. Similox-Tohon et al. 2006) © Sintubin

irrefutably distinguish between damage caused by earthquakes and that caused by other natural agents or human intervention, the main issue archeoseismology is confronted with is the question how earthquake evidence in destruction horizons and/or disturbed buildings can be meaningfully translated into physical earthquake parameters, such as intensity, magnitude, distance to epicenter, date of earthquake, ground acceleration, etc. Ultimately, because the limitations of the archeoseismological record are all too obvious when it comes to claiming its potential role in seismic-hazard studies, it all boils down to the question what the true added value is of archeoseismology.

In Search of a Shared Protocol and Standardized Methodology

Archeoseismology's greatest challenge – and its foremost attraction – remains to date the

integration of principles and practices of a very wide range of sciences, from history, anthropology, archeology and sociology, over geology, geomorphology, geophysics, and seismology to architecture and structural engineering. Arguably the principle difficulty in archeoseismology is the lack of a shared protocol and of a rigorous and transparent standardized methodology (cf. Sintubin and Stewart 2008 and references therein).

Through the years, different, primarily qualitative, archeoseismological schemes have been proposed, consisting of points of interest (Karcz and Kafri 1978; Rapp 1986; Nikonov 1988; Stiros 1996), key research questions (Guidoboni 1996), or flow charts (Galadini et al. 2006b) that ought to be considered during excavation works at an archeological site by collaborative teams of seismologists, geologists, archeologists, etc. (see Sintubin and Stewart 2008 for synthetic

overview of archeoseismological schemes). More recently, the comprehensive classification of earthquake archaeological effects (EAEs) (Rodríguez-Pascua et al. 2011) pursues the integration of archeoseismological evidence in the framework of the macroseismic Environmental Seismic Intensity scale (ESI2007, Michetti et al. 2007). These efforts to develop a shared protocol, however, are commonly designed from within a single scientific discipline.

Most of these schemes have been grafted onto the archeological work in the Mediterranean and the Middle East, strongly relying on strain structures in the building fabric (e.g., Stiros 1996). More quantitative approaches evaluate the probability of the occurrence of a proposed ancient earthquake by using a feasibility matrix for archeoseismological findings (Hinzen 2005) or assess the degree of certainty to which an archeological site has recorded an ancient earthquake by using a logic-tree formalism (Sintubin and Stewart 2008).

In recent years, a clear shift in perspective, from qualitative to more quantitative and multidisciplinary approaches, trying to integrate earthquake evidence from different perspectives (e.g., archeoseismology, geophysics, paleoseismology, geomorphology, geology), has definitively proven a major advancement in the discipline, supporting the reliability of the archeoseismological evidence (e.g., Similox-Tohon et al. 2006).

Because of the wide variety of disciplines involved, from the humanities and the social, natural, and engineering sciences, it seems, though, nearly inevitable that all practitioners who look at earthquake evidence in the archeological record will keep pursuing different objectives. The historian may want to know if an earthquake had any effect on the political, social, or military balance in a region. The engineer may be concerned about mitigating the seismic threat to our cultural and architectural heritage, while the seismologists are attempting to complete the historical catalogue of earthquakes and their physical parameters. Finding a balance between all these interests will also in the future remain archeoseismology's greatest challenge.

Ancient Seismoscopes

For assessing the seismic hazard of a region, an accurate catalogue of earthquakes and their physical parameters is imperative. Seismic-hazard practitioners need exact dates, magnitudes, source areas, etc. of past earthquakes. Taking into account the incompleteness of the archeological record, its limited spatial and temporal resolution, and the uncertainties inherent to archeological earthquake evidence, the skepticism with respect to the applicability of archeoseismology in **seismic-hazard studies** is indeed legitimate (cf. Sintubin 2011).

Some common pitfalls keep adding to the seismologist's skepticism. There is the **preservation problem**. The archeological record is not evenly distributed through time. Ancient history consists of long periods of cultural, social, and political stability and flourishing economies, during which any sign of earthquake is most probably expertly covered up. In contrast, during intervening, short periods of social and political upheaval, and economic crisis, signs of destructive earthquakes may be left extant, primarily because there is no impetus or funds to fully recover from the earthquake disaster. Only then, the earthquake leaves its marks in the archeological record, giving rise to an observational bias that focuses on periods of upheaval, destruction, abandonment, etc.

Another danger exists that "anomalous" earthquake catastrophes, supposedly proven by archeologists, are used uncritically in seismic-hazard assessments as "real" events (Ambraseys et al. 2002). Moreover, confronting the archeoseismological evidence, commonly poorly constrained in time and space, with the evenly incomplete and sometimes poorly constrained historical earthquake catalogues, may carry the risk of an arbitrary correlation, inevitably leading to **circular reasoning** (Rucker and Niemi 2010). When archeologists identify a destruction horizon as caused by an earthquake and consult existing earthquake catalogues to assign a date to the particular earthquake-related destruction horizon, the risk exists that historical seismologists add this particular archeological site to the catalogue as evidence

for that particular historical earthquake. This overreliance on historical catalogues clearly corrupts the usefulness of archeoseismology and should therefore be omitted from archeoseismological practices.

Given all these pitfalls, it becomes apparent that archeoseismological investigations should indeed not start from a seismological perspective but from the archeological earthquake evidence itself, with all its inherent limitations and uncertainties. Rather than simply complementing earthquake catalogues with potentially highly conjectural, ancient earthquakes, archeological sites may have the potential of becoming testing grounds to quantitatively assess site-specific ground effects. In this respect, archeological sites become **ancient seismoscopes** that can be used strategically to examine specific earthquake scenarios in a region (cf. Sintubin 2011). Archeological sites, especially those with a long and lasting history, do have the potential to have recorded the effects of a major earthquake. A quantitative assessment of the ground motions on archeological sites may indeed hold the eventuality to have narrowed down macroseismic parameters associated with the maximum credible earthquake in the region, irrespective of the time of occurrence, the magnitude, the seismic source, etc. With such an approach, archeoseismology enters in the logics of scenario-based, deterministic seismic risk assessment.

New Developments in Archeoseismology

This tendency towards a more standardized and quantitative approach of potential archeological earthquake evidence in archeological sites is fully exemplified in the rapid advances in **quantitative archeoseismology** (cf. Hinzen et al. 2011 and references therein). This recent development in archeoseismology primarily focuses on monumental architecture, the classical field of application of archeoseismology since its beginning (cf. Stiros 1996). Quantitative archeoseismology firstly applies modern techniques, such as 3D laser scanning (e.g., ground-based LIDAR), to obtain a three-dimensional



Archeoseismology, Fig. 8 Perfectly aligned *toppled columns* at *Susita* (Golan Heights), classically interpreted as indicative for the direction of strong ground motion caused by a major earthquake. Such earthquake hypothesis does not pass the test of scenario-based ground motion simulations (cf. Hinzen et al. 2011) © Sintubin

structural model of the archeological damaged building or structure, allowing the construction of a very precise structural damage inventory. Using earthquake engineering models, the dynamic behavior of the ancient structure can be evaluated. Subsequently, scenario-based earthquake ground motion simulations enable testing a realistic earthquake hypothesis to explain the damage observed. This approach has already lead to the conclusion that in a number of cases, alternative, natural, or anthropogenic causes are more probable than the seismic cause that has originally been considered. Moreover, “classical” damage typologies attributed to earthquake-related ground motions, such as the perfectly aligned toppled columns (Fig. 8), could not be validated, even adding to the skepticism towards “classical” archeoseismological practices.

Besides this added value of quantitative archeoseismology with respect to monumental architecture, also the value of rubble architecture and associated destruction horizons get again particular interest in a more quantitative, **integrated territorial approach** that starts from the specific seismotectonic context and the empirical ground-motion relationships of potential earthquake sources and focuses on well-documented, high-visibility archeological contexts, characterized by very rapid ceramic change (~100-year time window) (cf. Jusseret and Sintubin 2012).

Issues and Perspectives in Archeoseismology

Archeoseismology will always be plagued by the ambiguities inherent to the archeological record of ancient earthquakes. In this respect, archeoseismology may very well never be able to deliver reliable and conclusive earthquake evidence that is needed to improve the **assessment of the seismic hazard** in a region. New developments towards more integrated, multidisciplinary approaches as well as quantitative archeoseismology will allow, though, that the potential of archeological sites as **ancient seismoscopes** is fully developed in the future.

The fact that the archeological visibility of earthquakes may be strongly biased towards the relatively short but commonly well-documented (e.g., rapid changes in ceramic styles) periods in a society's history of social, political, and/or economic turmoil opens unique perspectives for archeoseismological research. Preservation of archeological earthquake evidence may indeed rather be related to societal factors than to physical aspects of the ancient earthquakes. Ultimately, the relatively undisturbed, extant archeological record of earthquakes can tell us more about past societies and their attitude to physical disasters (cf. Jusseret and Sintubin 2012). A better appreciation of the complex way by which our ancestors responded to damaging earthquakes might

indeed shed light on the societal factors defining the resilience or vulnerability of past societies. Eventually, by highlighting how our ancestors coped with earthquake disasters, archeoseismology could find new aspirations in establishing local **earthquake cultures** in earthquake-prone regions (cf. Sintubin et al. 2008), possibly providing a substantial contribution to the **mitigation of the earthquake risk**, by improving earthquake risk literacy and awareness.

Summary

Archeoseismology is the interdisciplinary study of pre-instrumental, **ancient earthquakes** through indirect evidence in the archeological record. As a burgeoning discipline within the broad research realm of **earthquake sciences**, archeoseismology bridges the gap between **instrumental** and **historical seismology** on the one side and **paleoseismology** and **earthquake geology** on the other.

The archeological record can be used in basically three ways in the identification and characterization of ancient earthquakes. The most obvious and straightforward archeological evidence of fault activity are archeological remains that are partly displaced due to coseismic surface rupturing on an active fault. As cultural piercing features, these **faulted relics** can be used to determine the type of faulting, the amount of coseismic slip, as well as the cumulative fault slip. Within the archeological stratigraphy, particular horizons can be designated to sudden destruction caused by human and/or natural agents. If the **archeological destruction horizons** can be attributed to an earthquake with a high degree of certainty, material included in the destruction horizons can be used to date episodes of earthquake damage. Destruction horizons are the most useful “proxy” for ancient earthquakes in archeological contexts dominated by rubble architecture and associated stratigraphy. A third type of earthquake evidence in the

archeological record are typical **strain structures in the building fabric** that are primarily caused by coseismic ground shaking and ground motion. These earthquake-related damage features are most conspicuous on monumental buildings and man-made constructions.

Archeoseismology will always be plagued by the ambiguities and uncertainties inherent to the archeological earthquake evidence, primarily resulting from the difficulty to irrefutably distinguish between damage caused by earthquakes and that caused by other natural or human agents. In this respect, archeoseismology may very well never be able to deliver reliable and conclusive earthquake evidence that is needed to improve the assessment of the seismic hazard in a region. Archeological sites have though the potential of becoming testing grounds to quantitatively assess site-specific ground effects. As **ancient seismoscopes**, archeological sites may hold the eventuality to have narrowed down macroseismic parameters associated with the maximum credible earthquake, irrespective of the time of occurrence and the physical parameters of the earthquake. Finally, preservation of archeological earthquake evidence may rather be related to societal factors than to physical parameters. In this respect, the relatively undisturbed archeological record of ancient earthquakes can tell us more about past societies and their attitude to physical disasters.

Cross-References

- [Damage to Ancient Buildings from Earthquakes](#)
- [Damage to Buildings: Modeling](#)
- [Earthquake Location](#)
- [Earthquake Magnitude Estimation](#)
- [Earthquake Recurrence](#)
- [Earthquakes and Their Socio-economic Consequences](#)
- [Intensity Scale ESI 2007 for Assessing Earthquake Intensities](#)
- [Mechanisms of Earthquakes in Aegean](#)

- [Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings](#)
- [Paleoseismology](#)
- [Paleoseismology: Integration with Seismic Hazard](#)
- [Post-Earthquake Diagnosis of Partially Instrumented Building Structures](#)
- [Review and Implications of Inputs for Seismic Hazard Analysis](#)
- [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)
- [Tsunamis as Paleoseismic Indicators](#)

References

- Ambraseys NN, Jackson JA, Melville CP (2002) Historical seismicity and tectonics: the case of the eastern Mediterranean and the middle East. In: Lee WHK, Kanamori H, Jennings PC, Kisslinger C (eds) International handbook of earthquake and engineering seismology, vol 81A, International geophysics series. Academic, Amsterdam, pp 747–763
- Barnes GL (2010) Earthquake archaeology in Japan: an overview. In: Sintubin M, Stewart IS, Niemi TM, Altunel E (eds) Ancient earthquakes, Geological Society of America special paper, 471. Geological Society of America, Boulder, pp 81–96
- Caputo R, Pavlides SB (2008) Earthquake geology: methods and applications. Tectonophysics 453:1–296
- Evans A (1921) The palace of minos. A comparative account of the successive stages of the early cretan civilization as illustrated by the discoveries at knossos. Volume I. The neolithic and early and middle minoan ages. MacMillan and Co., London
- Evans A (1928) The palace of minos. A comparative account of the successive stages of the early cretan civilization as illustrated by the discoveries at knossos. Volume II: Part I. Fresh light son origins and external relations: The restoration in town and palace after seismic catastrophe towards close of M.M. III, and the beginning of the new era. MacMillan and Co., London
- Force ER, McFadgen BG (2010) Tectonic environments of ancient civilizations: opportunities for archaeoseismological and anthropological studies. In: Sintubin M, Stewart IS, Niemi TM, Altunel E (eds) Ancient earthquakes, Geological Society of America special paper, 471. Geological Society of America, Boulder, pp 21–28
- Galadini F, Hinzen K-G, Stiros SC (2006a) Archaeoseismology at the beginning of the 21st century. J Seismol 10:393–537
- Galadini F, Hinzen K-G, Stiros SC (2006b) Archaeoseismology: methodological issues and procedure. J Seismol 10:395–414

- Guidoboni E (1996) Archaeology and historical seismology: the need for collaboration in the Mediterranean area. In: Stiros S, Jones RE (eds) *Archeoseismology*. Institute of Geology & Mineral Exploration and British School of Athens, Athens, pp 7–13
- Hinzen K-G (2005) The use of engineering seismological models to interpret archeoseismological findings in Tolbiacum, Germany: A case study. *Bull Seismol Soc Am* 95:521–539
- Hinzen K-G, Fleischer C, Reamer SK, Schreiber S, Schütte S, Yerli B (2011) Quantitative methods in archeoseismology. *Quat Int* 242:31–41
- Jackson J (2006) Fatal attraction: living with earthquakes, the growth of villages into megacities, and earthquake vulnerability in the modern world. *Philos Trans R Soc* 364:1911–1925
- Jusseret S, Sintubin M (2012) All that rubble leads to trouble: reassessing the seismological value of archaeological destruction layers in Minoan Crete and beyond. *Seismol Res Lett* 83(4):736–742
- Jusseret S, Sintubin M (2013) The origins of an old myth: Sir Arthur Evans, Claude Schaeffer and the seismic destruction of Late Bronze Age Eastern Mediterranean civilizations. *Seismol Res Lett* 84(1):94–100
- Karcz I, Kafr, U (1978) Evaluation of supposed archeoseismic damage in Israel. *J Archaeol Sci* 5:237–253
- Marco S, Agnon A, Ellenblum R, Eidelman A, Basson U, Boas A (1997) 817-year-old walls offset sinistrally 2.1 m by the Dead Sea transform, Israel. *J Geodyn* 24:11–20
- McGuire WG, Griffiths DR, Hancock PL, Stewart IS (eds) (2000) *The archaeology of geological catastrophes*, Special publications, 171. Geological Society, London
- Michetti AM, Hancock PL (1997) Paleoseismology: understanding past earthquakes using quaternary geology. *J Geodyn* 24(1–4):3–10
- Michetti AM, Esposito E, Guerrieri L, Porfido S, Serva L, Tatevossian R, Vittori E, Audemard F, Azuma T, Clague J, Comerci V, Gürpınar A, McCalpin J, Mohammadioun B, Möner NA, Ota Y, Roghazin E (2007) Intensity scale ESI 2007. *Memorie descrittive della Carta Geologica d'Italia* 74:41
- Nikonov AA (1988) On the methodology of archeoseismic research into historical monuments. In: Marinos PG, Koukis GC (eds) *Engineering Geology of Ancient Works*, Balkema, Rotterdam, pp 1315–1320
- Nur A, Cline EH (2000) Poseidon's horses: plate tectonics and earthquake storms in the late bronze age aegean and eastern mediterranean. *J Archaeol Sci* 27:43–63
- Rapp GJ (1986) Assessing archaeological evidence for seismic catastrophes. *Geoarchaeology* 1:365–379
- Rodríguez-Pascua MA, Pérez-López R, Giner-Robles JL, Silva PG, Garduño-Monroy VH, Reicherter K (2011) A comprehensive classification of Earthquake Archaeological Effects (EAE) in archeoseismology: application to ancient remains of Roman and Mesoamerican cultures. *Quat Int* 242:20–30
- Rucker JD, Niemi TN (2010) Historical earthquake catalogues and archaeological data: achieving synthesis without circular reasoning. In: Sintubin M, Stewart IS, Niemi TM, Altunel E (eds) *Ancient earthquakes*, Geological Society of America special paper, 471. Geological Society of America, Boulder, pp 97–106
- Sbeinati MR, Meghraoui M, Suleyman G, Gomez F, Grootes P, Nadeau MJ, Al Najjar H, Al-Ghazzi R (2010) Timing of earthquake ruptures at the Al Harif Roman aqueduct (Dead Sea fault, Syria) from archeoseismology and paleoseismology. In: Sintubin M, Stewart IS, Niemi TM, Altunel E (eds) *Ancient earthquakes*, Geological Society of America special paper, 471. Geological Society of America, Boulder, pp 243–267
- Schaeffer CFA (1948) *Stratigraphie Comparée et Chronologie de l'Asie Occidentale*. Oxford University Press, London
- Silva PG, Sintubin M, Reicherter K (2011) Earthquake archaeology and paleoseismology. *Quat Int* 242(1):1–253
- Similox-Tohon D, Sintubin M, Muchez P, Vanhaverbeke H, Verhaert G, Waelkens M (2005) Identification of a historical morphogenic earthquake through trenching at ancient Sagalassos (SW Turkey). *J Geodyn* 40:279–293
- Similox-Tohon D, Sintubin M, Muchez P, Verhaert G, Vanneste K, Fernandez M, Vandycke S, Vanhaverbeke H, Waelkens M (2006) The identification of an active fault by a multidisciplinary study at the archaeological site of Sagalassos (SW Turkey). *Tectonophysics* 420:371–387. [+Erratum in *Tectonophysics*, 435 (2007), 55–62]
- Sintubin M (2011) Archeoseismology: past, present and future. *Quat Int* 242:4–10
- Sintubin M, Stewart IS (2008) A logical methodology for archeoseismology: a proof of concept at the archaeological site of Sagalassos, southwest Turkey. *Bull Seismol Soc Am* 98(5):2209–2230
- Sintubin M, Stewart IS, Niemi TM, Altunel E (2008) Earthquake archaeology – just a good story? *Seismol Res Lett* 79(6):767–768
- Sintubin M, Stewart IS, Niemi TM, Altunel E (eds) (2010) *Ancient earthquakes*, Geological Society of America special paper, 471. Geological Society of America, Boulder
- Stiros SC (1996) Identification of earthquakes from archaeological data: methodology, criteria and limitations. In: Stiros S, Jones RE (eds) *Archeoseismology*. Institute of Geology & Mineral Exploration and British School of Athens, Athens, pp 129–152
- Stiros SC, Jones RE (1996) *Archeoseismology*. Institute of Geology & Mineral Exploration and British School of Athens, Athens
- Tuttle MP, Schweig ES (1995) Archeological and pedological evidence for large prehistoric earthquakes in the New Madrid seismic zone, central United States. *Geology* 23(3):253–256

Artificial Offshore Seismic Sources

Wolfgang Rabbel and Dennis Wilken
Institute of Geosciences, University of Kiel,
Kiel, Germany

Synonyms

Marine seismic sources; Seismic energy sources

Introduction

Artificial Offshore Seismic Sources are instruments that are used for generating short-duration pressure pulses in the marine water column, usually with constant repetition rate. They provide the acoustic energy for marine seismic surveys. Marine seismic surveys are performed in order to image geological structure (layers and faults) beneath the seafloor and to characterize rocks and rock strength in terms of seismic wave velocity or dynamic elastic constants. Therefore, short pressure pulses are applied to the water column that travel as acoustic waves down to the sea bottom where they are converted into seismic body and interface waves (elastic waves). These seismic waves penetrate deeper into the underground and are partially reflected at geological interfaces and/or refracted back to the surface (Fig. 1, right). This returning part of the wave field is recorded either at the seafloor with seismographs or in the water column with hydroacoustic receivers. In order to allow for profiling both seismic source and receivers are towed behind a vessel carrying out repeated measurements while moving forward. The technique of marine seismic has been developed and improved continuously since the late 1940s. This entry provides an overview of the purpose and principles marine seismic sources. More detailed information can be found, for example, in Jones (1999) and Meunier (2011).

General Considerations

The goal of marine seismic reflection surveys is to image geological structure down to a predefined target depth at highest possible resolution. Therefore, the pressure pulses generated by a marine seismic source are usually designed to have a sharp onset and short duration (some to some 10 ms; exception marine vibrator source) corresponding to a broadest possible frequency spectrum. The short duration, respective large bandwidth, is necessary to enable distinguishing of reflections (echoes) arriving at the receivers subsequently from interfaces at increasing depths. Seismic wave propagation is affected by absorption that leads to a notable continuous decrease of amplitudes starting from the high-frequency end of the spectrum. Therefore, penetration depth increases with decreasing signal frequency. While it propagates, the wave front widens and signal amplitudes decrease accordingly (“geometrical spreading”). This effect needs to be compensated for by increasing the initial strength of the pressure pulse in order to keep reflected signals of deep horizons detectable. Finally, in order to obtain a reliable depth section, the measured profiles have to be sampled spatially with a sufficiently high density realized by repeating the measurements while the vessel is moving. This implies that the marine seismic source needs to be reloadable within a certain time interval depending on the velocity of the ship. Typical repetition intervals are between 0.5 up to some 10 s. The target range relevant for seismic exploration – for example in earthquake engineering – may extend from near-seafloor layering down to several kilometers depth, for example, in order to detect and trace deep reaching fault systems. Correspondingly, different sorts of seismic sources are in use, each accounting for different target depth or infrastructure related situations (Table 1).

Types of Marine Seismic Sources and Their Key Principles

Airguns and GI-guns

Airguns or G-(Generator)-Guns generate an acoustic impulse by releasing a high pressure air

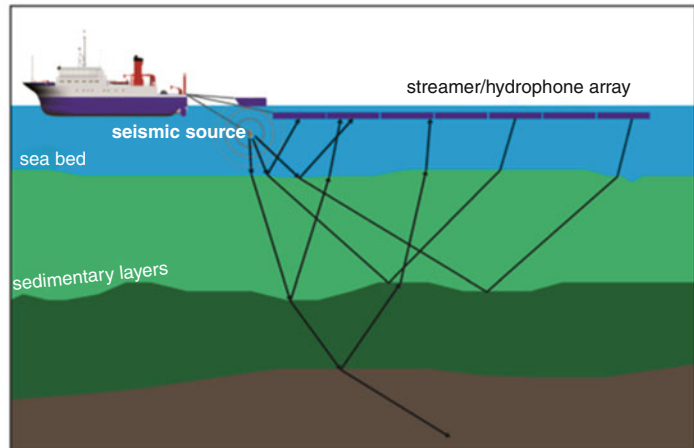
GI-Gun



Sparker



Boomer



A

Artificial Offshore Seismic Sources, Fig. 1 *Right:* Principle of marine seismic data acquisition; pulse-like seismic waves travel through geological layers and are reflected back to the surface at interfaces where density

or propagation velocity change; wave propagation paths indicated by rays (*black lines*). *Left:* examples of seismic sources for near-surface prospecting; from *top to bottom*: Airgun (GI-gun), Sparker, Boomer

bubble underwater. The radiation pattern of the airgun source corresponds to an acoustic dipole because an image-source mirrored at the water surface has to be considered in order to account for the free-surface condition. It causes a so-called ghost reflection following the primary pulse with negative polarity and a time delay depending on source depth. While moving to the sea surface, the generated bubble oscillates around the equilibrium of air- and hydrostatic water pressure with a natural period T . T depends on the energy of the injected air E , the confining hydrostatic pressure P of the water column, and the fluid density d according to the Rayleigh-Willis formula $T \approx 1.14d^{1/2} P^{-5/6} E^{1/3}$. This oscillation generates an undesired lengthening of the seismic signal. In order to overcome the

bubble oscillation, GI-(Generator Injector)-Guns are very common that inject a second air pulse to collapse the initial bubble before it starts to oscillate. G-Guns and GI-Guns can be mounted in arrays to enhance the signal amplitude and, when fired with time offsets, to modulate the shape of the signal.

The generated frequency content mainly depends on the volume of the airgun chamber, the assigned pressure, and the tow depth.

Water Guns

A water gun is a pneumatic seismic source operated by compressed air, too. The air is used to drive a piston pressing water from a chamber inside the gun into the water column. The seismic impulse is generated by the implosion of a cavity

Artificial Offshore Seismic Sources, Table 1 Technical data and properties of typical marine seismic sources. The given value of structural resolution corresponds to a quarter wavelength at 1,600 m/s wave speed

Source type	Chamber volume (l)	Energy per pulse (J) / pressure (dB reference 1 μ Pat 1 m)	Spectral width (Hz)	Main frequency (Hz)	Typical exploration depths	Resolution at main frequency (m)	Power supply unit
Airgun	0.2–32	~10 kJ/170–230 dB	—	10–200	Some 100 m to some km	2–40	Compressor
GI Gun	0.4–4	~10 kJ/170–230 dB	—	50–200	Some 100 m to some km	2–8	Compressor
Water Gun	0.2–2	~30 kJ	—	50–1,000	Some 100 m to some km	0.4–8	Compressor
Sparker	—	300–2,400 J	100–3 k	~1,500	Upto ~200 m	0.25	Capacitor bank
Boomer	—	300–700 J	500–4 k	~3,000	Upto ~100 m	0.15	Capacitor bank
Pinger	—	~100 J	2 k–150 k	—	Upto ~10 m	0.2–0.03	Electrical amplifier
Vibro source	—	Some 10 kW times sweep length (s)	6–100	—	Some 100 m to some km	4–100	Electrical current

created behind the expelled water. It is free of bubble oscillations and contains higher frequencies than typically generated by airguns.

Sparker

The sparker source generates an acoustic impulse by a high-voltage spark discharge between two electrodes in the water. The spark consists of a high-pressure plasma whose expansion and collapse generates the acoustic peak. The high voltage needed to generate such a plasma is stored in a bank of capacitors, which is reloaded for each shot. Sparkers can be used at different energy levels, depending on the capacitor bank. Commonly sparkers are used in arrays to enhance the signal amplitude and to maintain the signal shape which usually changes because the electrodes tend to tip-off during usage. Sparkers can be applied in brackish water and seawater.

Boomer

Boomers are also based on combining electromagnetic and mechanical principles. Similar to the sparker case, electric charges are assembled in a bank of capacitors releasing a high-voltage

pulse to the source. The discharge is led through a flat coil inside the source body. From below an aluminum plate is pressed against the coil by springs. The magnetic field of the coil induces eddy currents inside this plate. These currents produce magnetic fields that repel the plate from the coil (Lenz's law). The spring then moves the plate back, leaving a collapsing cavitation volume that generates the acoustic impulse.

Pinger

Sediment echo sounders (SES) use piezoelectric "pingers" as electroacoustic converters.

Some materials (especially crystalline) tend to show electromechanical properties which means that the material is stretched when voltage is applied. Closely packed piezoelectric crystals are very effective electroacoustic converters and can be used as acoustic sources.

These pingers are applied as single point sources or – even more often – in arrays which enable, when used with time offsets, a modulation of the radiation pattern of the source. Alternative developments are based on the parametric nonlinearity effect to enhance the acoustic efficiency and the stability of the signal.

Marine Vibrator

Marine vibrator sources are basically volumetrically oscillating shells that transmit a chirp signal (“sweep”) to the water column. The chirp is linearly modulated in frequency and lasts typically several seconds. Seismic reflections can only be recognized after the records were cross-correlated with the source signal. This concept was adopted from the onshore VibroseisTM method (e.g., Sheriff and Geldart 1995). The shells are constructed flexible or otherwise expandable; the oscillations are driven by hydraulic, magnetostrictive, or electromagnetic devices placed inside (e.g., Tenghamn 2006; Meunier 2011). Like airguns and water guns, marine vibrators can be combined to form arrays in order to improve energy output and radiation characteristics. They work in a similar frequency range as airguns and release similar seismic energy in total. However, since they provide energy at lower rate over a much longer time interval they are causing less environmental impact than the impulsive source that release energy within some 10 ms. Since the vessel moves significantly while the source is active, Doppler effects occur in the records that need to be corrected in during data processing. For near-surface investigations, Vibroseis-type chirps can be generated through low-energy piezoelectric transducers. However, marine vibrators for deep (km-scale) sounding are still under development.

Summary

Artificial Offshore Seismic Sources are instruments emitting sound signals for marine seismic surveys that are performed for investigating offshore geological structure. Functional principles of marine seismic source depend on the required energy and frequency content of the signals, including piezoelectric and high-pressure air pulsers as low and high energy end members. The structural resolution of the seismic images of geological layering ranges from some 10 cm to some 10 m depending on the depth of the target horizons and the seismic sources applied.

References

- Jones EJW (1999) Marine geophysics. Wiley, Chichester
- Meunier J (2011) Seismic acquisition from yesterday to tomorrow. Society of Exploration Geophysicists, Tulsa
- Sheriff LE, Geldart LP (1995) Exploration seismology, 2nd edn. Cambridge University Press, Cambridge
- Tenghamn J (2006) An electrical marine vibrator with a flextensional shell. *Explor Geophys* 37:286–291

Assessment and Strengthening of Partitions in Buildings

Christis Z. Chrysostomou

Department of Civil Engineering and Geomatics,
Cyprus University of Technology, Limassol,
Cyprus

Synonyms

Infills; Infill walls; Masonry infills; Reinforced concrete infills

Introduction

Partitions in buildings are necessary architectural features that separate spaces in order to facilitate various functions that are required depending on the use of a building. These partitions in modern structures can be lightweight, for example, in case that they are used to subdivide office spaces, or can be heavy masonry partitions that are used mainly in reinforced concrete structures to subdivide the plan area of a building into various rooms. In the latter case, these are mostly built within the frame of the structure filling the gap, and they are therefore called infills or infill walls. Depending on the type of the material that is used to construct them, they can be called masonry infills or reinforced concrete (RC) infills.

Infill walls have attracted the attention of many researchers since the early 1950s, and much work has been undertaken to study their behavior and interaction with the surrounding

frames. In addition, efforts have been made to utilize infill walls as a means of producing economic designs by reducing the sizes of the members of the bounding frames.

A large number of researchers have studied the behavior of infilled frames. It is evident from their studies that infill walls can provide both an economic and practical means for the lateral stability of framed structures and a viable alternative for retrofitting existing structures to resist seismic, wind, and blast loads. Despite this, there is reluctance from the engineering community to use this structural system widely and treat infill walls as structural elements.

Therefore, although the subject of infilled frames has been studied for more than 60 years, there are still no definitive answers either about their behavior and interaction with the bounding frame or about the estimation of their stiffness and strength. Some problems that make it difficult for practicing engineers to use infill walls as structural elements are the inherent nonlinearity, the high degree of variability, and the inherent degradability of infill walls. The fact that infill walls exhibit completely different in-plane and out-of-plane behavior makes the problem even more difficult to tackle. This reveals the difficulties associated with this problem and the need for more research to answer these long-standing questions.

Nevertheless, the presence of infill walls has shown that in most cases they have beneficial effects on the behavior of structures during earthquakes and have contributed in the prevention of collapse of many structures. It is therefore for this reason that efforts have been made to use engineered infills to retrofit existing seismic-deficient buildings, since they present an economically viable solution.

In this entry, the in-plane behavior and in-plane failure modes of infilled frames are presented. Then, macro- and micromodels for infill walls are presented followed by the results of experimental investigations for the replacement of masonry infill walls with reinforced concrete infills, as an economically viable method for retrofitting existing structures. Finally, methods

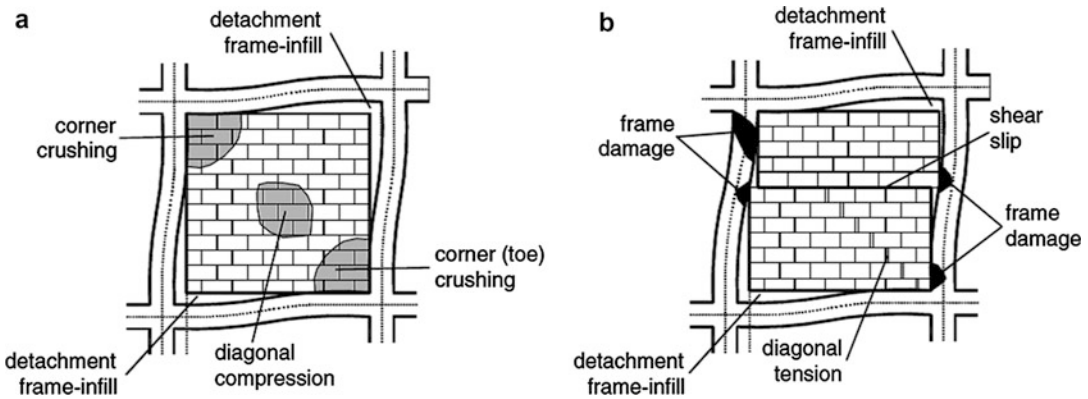
for strengthening existing partition walls are presented.

Behavior of Partitions

The modeling of the behavior of infilled frames under lateral loading (and mainly earthquake-induced loads) is a complex issue because these structures exhibit a highly nonlinear response due to the interaction between the masonry infill panel and the surrounding frame. This results to several modes of failure, each of which has a different failure load, and hence a different ultimate capacity and overall behavior.

At moderate loading levels, the infill of a non-integral infilled frame separates from the surrounding frame and the infill acts as a diagonal strut (Fig. 1). As the racking load is increased, failure occurs eventually in either the frame or the infill. The usual mode of frame failure results from tension in the windward column or from shearing in the column or beams or plastic hinging in columns or beams; however, if the frame strength is sufficient enough to prevent its failure by one of these modes, the increasing racking load eventually produces failure of the infill. In the most common situations, the in-plane lateral load applied at one of the top corners is resisted by a truss formed by the loaded column and the infill along the diagonal connecting the loaded corner and the opposite bottom corner. The state of stress in the infill gives rise to a principal compressive stress along the diagonal and a principal tensile stress in the perpendicular direction. If the infill is made of concrete, successive failures, initially by cracking along the compression diagonal and then by crushing near one of the loaded corners or by crushing alone, will lead to collapse; if the infill is made of brick masonry, an alternative possibility of shearing failure along the mortar planes may arise (Fig. 1).

Based on both experimental and analytical results during the last five decades, different failure modes of masonry-infilled frames were proposed that can be classified into five distinct modes given below:



Assessment and Strengthening of Partitions in Buildings, Fig. 1 Failure modes of non-integral infilled frames: (a) Corner crushing (CC) and diagonal

compression (DK) modes. (b) Sliding shear (SS), frame failure (FF), and diagonal cracking (DK) modes (Asteris et al. 2011 with permission from ASCE)

1. The corner crushing (CC) mode, which represents crushing of the infill in at least one of its loaded corners, as shown in Fig. 1a. This mode is usually associated with infilled frames consisting of a weak masonry infill panel surrounded by a frame with weak joints and strong members (Mehrabi and Shing 1997; El-Dakhakhni 2002; Ghosh and Amde 2002; El-Dakhakhni et al. 2003).
2. The diagonal compression (DC) mode, which represents crushing of the infill within its central region, as shown in Fig. 1a. This mode is associated with a relatively slender infill, where failure results from out-of-plane buckling of the infill.
3. The sliding shear (SS) mode, which represents horizontal sliding failure through bed joints of a masonry infill, as shown in Fig. 1b. This mode is associated with weak mortar joints in the infill and a strong frame.
4. The diagonal cracking (DK) mode, which is seen in the form of a crack across the compressed diagonal of the infill panel and often takes place with simultaneous initiation of the SS mode, as shown in Fig. 1b. This mode is associated with a weak frame or a frame with weak joints and strong members infilled with a rather strong infill (Mehrabi and Shing 1997; El-Dakhakhni 2002).
5. The frame failure (FF) mode, which is seen in the form of plastic hinges developing in the columns or the beam-column connections, as shown in Fig. 1b. This mode is associated with a weak frame or a frame with weak joints and strong members infilled with a rather strong infill.

Ghosh and Amde (2002), based on the finite element method and including interface elements at the frame-infill interface, confirmed the order of occurrence of the above five distinct failure modes. Of the five modes, only the CC and SS modes are of practical importance (Comité Euro-International du Béton CEB 1996), since most infills are not slender (El-Dakhakhni et al. 2003) and therefore the second mode (DC) is not favored. The fourth mode (DK) should not be considered as a failure mode, due to the post-cracking capacity of the infill to carry additional load. The fifth mode (FF) relates to the failure of the frame, and it is particularly important when examining existing structures, which in many cases exhibit frame weakness. It should be noted that these failure modes are only seen/applicable to the case of infill walls without openings on the diagonal of the infill panel.

Kappos and Ellul (2000), based on an analytical study of the seismic performance of masonry-infilled RC-framed structures, found that taking into account the infill in the analysis resulted in

an increase in stiffness as much as 440 %. It is clear that, depending on the spectral characteristics of the design earthquake, the dynamic behavior of the two systems dealt with by the author (bare vs. infilled frame) can be dramatically different. They also presented a very useful global picture of the seismic performance of the studied infill frames by referring to the energy dissipated by each component of the structural system. It is clear that at the serviceability level, over 95 % of the energy dissipation is taking place in the infill walls (subsequent to their cracking), whereas at higher levels, the RC members start making a significant contribution. This is a clear verification of the fact that masonry infill walls act as a first line of defense in a structure subjected to earthquake load, while the RC-frame system is crucial for the performance of the structure to stronger excitations (beyond the design earthquake).

The quantification of the in-plane properties of non-integral infilled frames and the prediction of the failure modes is a rather cumbersome task even when the load is applied monotonically. The problem becomes even more difficult when a dynamic load is applied, since the hysteretic behavior of the materials and their deterioration with the cyclic loading should be taken into consideration. Chrysostomou and Asteris (2012) discuss these issues and present methods for the determination of the in-plane stiffness, strength, and deformation capacity of infills along with the results of a parametric study that compares these methods and checks them against experimental results whenever possible. Based on the above material, recommendations are made for the in-plane material properties, failure modes, strength and stiffness, as well as deformation characteristics, of infilled frames.

Modeling of Partitions

In order for engineers to be able to use infills as an engineering element, a reliable mathematical model needs to be developed that simulates the combined behavior of the infill partition and the bounding frame. As explained in the previous section, the behavior of infilled frames is rather complex and it depends on a large number of

parameters. Therefore, for the mathematical model to accurately simulate the behavior of infilled frames, it should take into consideration all these parameters.

The models for infills can be subdivided into macro- and micromodels. In the former, simple models are developed that are simulating the overall global behavior of infilled frames, while in the latter, surface or solid finite elements are used along with interface ones that simulate both the local and global behavior of both the infills and the bounding frames and their interaction. In the following two sections macro- and micromodels are presented.

Macromodeling

Since the first attempts to model the response of the composite infilled frame structures, experimental and conceptual observations have indicated that a diagonal strut with appropriate geometrical and mechanical characteristics could possibly provide a solution to the problem. Early research on the in-plane behavior of infilled frame structures undertaken at the Building Research Station, Watford (later renamed Building Research Establishment and now simply BRE), in the 1950s served as an early insight into this behavior and confirmed its highly indeterminate nature in terms solely of the normal parameters of design. On the basis of these few tests, a purely empirical interaction formula was later tentatively suggested for use in the design of tall framed buildings.

Diagonal Strut Model

In the early 1960s, Polyakov suggested the possibility of considering the effect of the infilling in each panel as equivalent to diagonal bracing, and this suggestion was later adopted by Holmes, who replaced the infill by an equivalent pin-jointed diagonal strut made of the same material and having the same thickness as the infill panel and a width defined by

$$\frac{w}{d} = \frac{1}{3} \quad (1)$$

where d is the diagonal length of the masonry panel. The “one-third” rule was suggested as

being applicable irrespective of the relative stiffness of the frame and the infill. One year later, Stafford Smith, based on experimental data from a large series of tests using masonry-infilled steel frames, found that the ratio w/d varied from 0.10 to 0.25. On the second half of the 1960s, Stafford Smith and his associates using additional experimental data related the width of the equivalent diagonal strut to the infill-frame contact lengths using an analytical equation, which has been adapted from the equation of the length of contact of a free beam on an elastic foundation subjected to a concentrated load. They proposed the evaluation of the equivalent width λ_h as a function of the relative panel-to-frame-stiffness parameter, in terms of

$$\lambda_h = h \sqrt[4]{\frac{E_w t_w \sin 2\theta}{4EI h_w}} \quad (2)$$

where E_w is the modulus of elasticity of the masonry panel, EI is the flexural rigidity of the columns, t_w is the thickness of the infill panel and equivalent strut, h is the column height between centerlines of beams, h_w is the height of infill panel, and θ is the angle, whose tangent is the infill height-to-length aspect ratio, being equal to

$$\theta = \tan^{-1} \left(\frac{h_w}{L_w} \right) \quad (3)$$

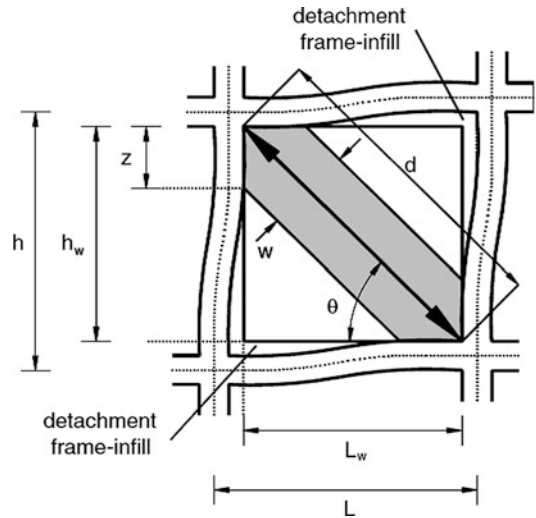
in which L_w is the length of infill panel (all the above parameters are explained in Fig. 2).

The use of this equation to seismic design is recommended for a lateral force level up to 50 % of the ultimate capacity.

Based on experimental and analytical data, Mainstone proposed an empirical equation for the calculation of the equivalent strut width, given by

$$\frac{w}{d} = 0.16 \lambda_h^{-0.3} \quad (4)$$

Mainstone and Weeks and Mainstone, also based on experimental and analytical data, proposed a



Assessment and Strengthening of Partitions in Buildings, Fig. 2 Definitions for the equivalent diagonal strut (Asteris et al. 2011 with permission from ASCE)

slightly modified empirical equation for the calculation of the equivalent strut width:

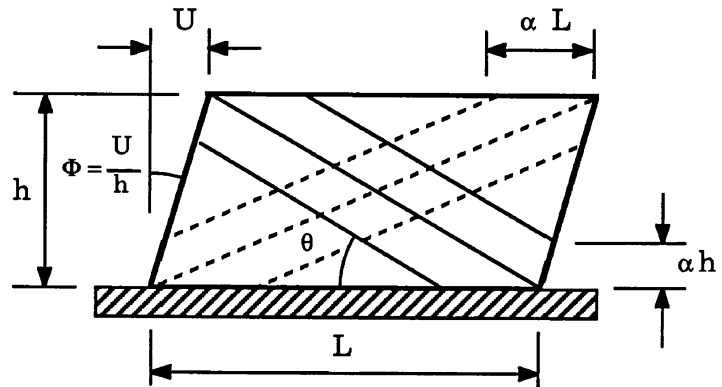
$$\frac{w}{d} = 0.175 \lambda_h^{-0.4} \quad (5)$$

This formula was included in FEMA-274 for the analysis and rehabilitation of buildings as well as in FEMA-306, as it has been proven to be the most popular over the years. Although this equation is accepted by the majority of researchers dealing with the analysis of infilled frames, several variations were presented by various researchers trying to improve its applicability. A discussion and comparison of these proposals as well as references to the work of the various researchers mentioned above is given by Chrysostomou and Asteris (2012), and Asteris et al. (2011).

Multiple-Strut Models

In the last two decades, it became clear that one single-strut element is unable to model the complex behavior of the infilled frames. The bending moments and shear forces in the frame members cannot be adequately represented using a single diagonal strut connecting the two loaded corners.

Assessment and Strengthening of Partitions in Buildings,
Fig. 3 Six-strut model for masonry infill panel in frame structures
 (Chrysostomou 1991)



More complex macromodels were hence proposed, still typically based on a number of diagonal struts.

Thiruvengadam proposed the use of a multiple-strut model to simulate the effect of the infill panel. This particular model consists of a moment-resisting frame with a large number of pin-jointed diagonal and vertical struts. Initially, a perfect frame-infill bond condition is assumed, and the lateral stiffness of the infill, by its shear deformation, is modeled by a set of pin-ended diagonal struts running in both directions. These diagonals represent the shear and axial stiffness of the masonry infill. Similarly, the vertical stiffness contribution is accounted for by providing vertical struts. The objective of the aforementioned study was a realistic evaluation of the natural frequencies and modes of vibration, purposes for which the nonlinear phenomena do not play an important role. This model has been adopted by many researchers to investigate the effect of infill on the behavior of infilled frames and has been also included in FEMA-356, due to the great number of struts, as a method for modeling the special case of infilled frames with openings. Similarly, Hamburger and Chakradeo proposed a multiple-strut configuration that can account for the openings also, but the evaluation of the characteristics of the struts is rather complicated. They showed that for panels of typical configuration, the formation of these struts protects the beam-to-column connections, which have limited capacity in withstanding significant flexural demand, with plastic hinges forming instead within the mid-span region of the beam.

They postulated that this resulted in a system of significant strength, stiffness, and ductility that behaves much like the modern eccentrically braced frame systems. Such behavior could, in part, be responsible for the observed good performance of these buildings in the 1906 San Francisco earthquake (Hamburger and Meyer 2006).

The main advantage of the multiple-strut models, in spite of the increase in complexity, is the ability to represent the actions in the frame more accurately. Symmakezis and Vratsanou (1986) employed five parallel struts in each diagonal direction. It was stressed how different contact lengths play a significant effect on the bending moment distribution in the frame members.

Chrysostomou (1991, Chrysostomou et al. 2002) aimed at obtaining the response of infilled frames under earthquake loading by taking into account both stiffness and strength degradation of infills. Each infill panel was modeled by six compression-only inclined struts (Fig. 3). Three parallel struts were used in each diagonal direction and the off-diagonal ones were positioned at critical locations along the frame members. These locations are specified by a parameter α , which represents a fraction of the length or height of a panel and is associated with the position of the formation of a plastic hinge in a beam or a column. At any point during the analysis of the nonlinear response, only three of the six struts are active, and the struts are switched to the opposite direction whenever their compressive force reduced to zero.

In order to conduct nonlinear analysis, the force-displacement relationships corresponding to the equivalent strut model must be adequately defined. The modeling of hysteretic behavior increases not only the computational complexity but also the uncertainties of the problem.

In Chrysostomou's model, the hysteretic behavior of the six struts is defined by a hysteretic model, which consists of two equations. The first equation defines the strength envelope of a structural element and the second defines its hysteretic behavior. The shape of the envelope and the hysteretic loops is controlled by six parameters, all of which have physical meaning and can be obtained from experimental data. The advantage of this strut configuration over the single diagonal strut is that it allows the modeling of the interaction between the infill and the surrounding frame and it takes into account both strength and stiffness degradation of the infill, which is vital for determining the response of infilled frames subjected to earthquake load.

Saneinejad and Hobbs developed a method based on the equivalent diagonal strut approach for the analysis and design of steel or concrete frames with concrete or masonry infill walls subjected to in-plane forces. The proposed analytical model assumes that the contribution of the infill panel to the response of the infilled frame can be modeled by replacing the panel by a system of two diagonal masonry compression struts. The method takes into account the elastoplastic behavior of infilled frames considering the limited ductility of infill materials. Various governing factors such as the infill aspect ratio, the shear stresses at the infill-frame interface, and relative beam and column strengths are accounted for in this development. This model has been adopted by Madan et al. (1997) for static monotonic loading, as well as quasi-static cyclic loading. This model for masonry infill panels was implemented in IDARC 2D Version 4.0 (Vales et al. 1996), a computer-based analytical tool for the inelastic analysis and damage evaluation of buildings and their components under combined dynamic, static, and quasi-static loading.

El-Dakhkhni (2000, 2002) and El-Dakhkhni et al. (2001) suggested replacing the infill wall by

one diagonal and two off-diagonal struts, on making use of the orthotropic behavior of the masonry wall as well as on some experimental observations and analytical simplifications, in order to simplify the nonlinear modeling of these structures.

Crisafulli investigated the influence of different multiple-strut models on the structural response of infilled reinforced concrete frames, focusing on the stiffness of the structure and the actions induced in the surrounding frame. Numerical results, obtained from the single-, two-, and three-strut models, were compared with those corresponding to a refined finite element. The lateral stiffness of the structure was similar in all the cases considered, with lower values for two- and three-strut models. It must be noted that, for the multiple-strut models, the stiffness may significantly change depending on the separation between struts. The single-strut model underestimates the bending moment because the lateral forces are primarily resisted by a truss mechanism. On the other hand, the two-strut model leads to higher values than those corresponding to the finite element model. A better approximation is obtained from the three-strut model, although some differences arise at the ends of both columns. Although the single-strut model constitutes a sufficient tool for the prediction of the overall response and the triple-strut model is superior in precision, Crisafulli adopted the double-strut model approach, accurate enough and less complicated compared to the other models.

More recently, Crisafulli and Carr (2007) proposed a new macromodel in order to represent, in a rational but simple way, the effect of masonry infill panels. The model is implemented as a four-node panel element which is connected to the frame at the beam-column joints. Internally, the panel element accounts separately for the compression and shear behavior of the masonry panel using two parallel struts and a shear spring in each direction. This configuration allows an adequate consideration of the lateral stiffness of the panel and of the strength of masonry panel, particularly when a shear failure along mortar joints or diagonal tension

failure is expected. Furthermore, the model is easy to apply in the analysis of large infilled frame structures. The main limitation of the model results from its simplicity, since the panel is connected to the beam-column joints of the frame, being thus not able to properly predict the bending moment and shear forces in the surrounding frame. A detailed presentation as well as references to the work on macromodels of the various researchers mentioned above is given by Asteris et al. (2011).

Micromodeling

The ability to model accurately the behavior of a structural system, including the infilled RC frame, depends on experimental work. A review of the available published experimental data describing the effect of infill walls on the overall structural response of RC frames finds significant scatter in the obtained results, due to the large number of uncertainties involved in the various investigations carried out to date. This data can describe qualitatively the effect of infill walls, but it is not yet possible to quantify this effect experimentally. Since the formulation and validation of mathematical (numerical or analytical) models is based on available experimental data, the validity of the yield predictions of current software packages is currently questionable.

Due to the difficulties, limitations, and high costs associated with the testing of infilled RC frames, resort is frequently made to the use of nonlinear finite element analysis (NLFEA). The use of the FE method can provide a more detailed description of the effect of infill walls on the response of infilled frames. At the same time, it allows the investigation to be extended to structural forms more complex than the simple structural elements that are usually studied experimentally (i.e., scaled models of one- or two-story infilled frames with column height to beam length equal to values from 0.75 to 1.25). To qualify for accurately capturing the behavior of a concrete structure, a FEA package must predict the structural response within an error of up to 20 %. Such a package should also employ appropriate constitutive models and adopt nonlinear numerical procedures capable of

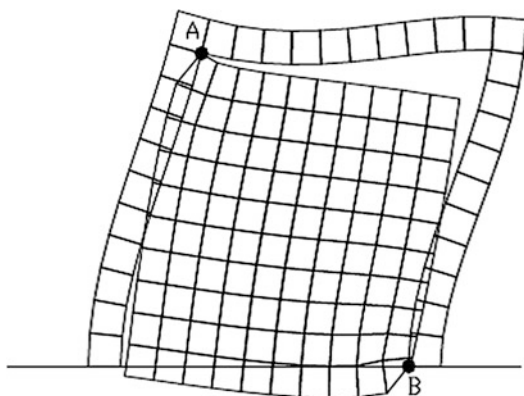
capturing the force redistribution and cracking processes of brittle materials.

To this end, constitutive relationships must be defined for the use of finite element models of infilled frames. The use of 3D solid, instead of linear, elements in constitutive models requires a considerably higher level of model sophistication. Models of concrete behavior are based either on regression analyses of experimental data (empirical models) or on continuum mechanics theories, which should also be verified against experimental data. Many such models have been proposed, but the application of FE packages in practical structural analysis has shown that the majority of constitutive relationships are case dependent, since the solutions obtained are realistic only for specific types of problems. The application of these packages to a different set of problems requires modification, sometimes significant, of the constitutive relationships. The situation is better for the reinforcement. However, complications arise with the introduction of bond-slip laws, which results in large discrepancies in predicted behavior.

Regarding the modeling of the infill panels, the great number of influencing factors, such as dimension and anisotropy of the bricks, joint width and arrangement of bed and head joints, material properties of both brick and mortar, and quality of workmanship, make the simulation of plain brick masonry extremely difficult. The level of complexity of the analytical model depends on whether masonry is considered as one-, two-, or three-phase material, where the three phases are comprised of the units, mortar, and unit-mortar interface.

Another level of complexity is added with the modeling of the infill-frame interface, which determines the boundary conditions of the infill and its interaction with the bounding frame. Springs and interface elements have been used by various researchers to represent the interaction between deformable structures such as these, where interfacial separation and sliding may arise.

Infill walls have also been represented by boundary elements. These are combined with FE, which can represent the bounding frame.



Assessment and Strengthening of Partitions in Buildings, Fig. 4 Finite element mesh for a non-integral infilled frame (Reprinted from Asteris et al. 2013 with permission from Elsevier)

This type of model reduces computational costs considerably, since the infill is represented by a single boundary element. However, the applicability of these models is limited to the elastic behavior of the infill; once cracks are initiated, boundary elements hold a disadvantage as they can be implemented only with extreme difficulty in nonlinear problems.

Mesh sensitivity poses a further challenge (Fig. 4). Very dense meshes are often thought to provide better results, but this is not always true when brittle materials are modeled. The size of the 3D brick FEs used for modeling concrete continua should be equivalent to the size of the cylinders used in testing. These are assumed to constitute a “material unit” for which average material properties are obtained. Hence, the volume of these specimens provides a general guideline for the size of the FE that should be used for the modeling of concrete structures. Furthermore, each Gauss point of a FE must correspond to a volume at least three times the size of the largest aggregate used in the concrete mix, in order to provide a realistic representation of concrete, rather than a description of its constituent materials.

Regarding the crack formation, the smeared-crack approach, rather than the discrete-crack approach, is usually adopted for the modeling of the cracking. With this approach, one may avoid

the complexities linked to remeshing, which is required by the discrete-crack approach.

The discrete element method (DEM) originated from the studies of fractured rock masses and is also used for the modeling of infill walls. Due to its capability of representing explicitly the motion of multiple, intersecting discontinuities, the DEM is particularly suitable for the analysis of discontinuity, such as masonry structures, where significant deformation occurs as relative motion between the blocks. The disadvantage of this element arises in the case of complicated deformation problems (e.g., beam-column component behavior in an infilled frame), where the number of triangular elements into which the area has to be discretized may become very large. The main difference between DEM and other continuum-based methods (e.g., FEM) is that the contact points in DEM are automatically updated and changed based on the “contact overlap” concept, leading to the detection of new contacts and complete detachment of blocks as the calculation process allows.

For the solution of the nonlinear dynamics problem, FEA packages use an iterative scheme, such as the Newton-Raphson method, and an implicit or explicit integration scheme. Some of these procedures have been adapted to take into account the brittle nature of the constitutive model of concrete with satisfactory results.

A large number of procedures/techniques have been used to apply micromodeling to infill frames, while taking into account all of the above parameters. Each of these models has its own merits and limitations, as well as level of complexity, and has been used to study different sets of parameters affecting the behavior of infilled frames.

The knowledge acquired over the years has made possible the development of commercial packages that provide a wide selection of elements, constitutive relationships, and solution schemes that researchers may use as tools for the study of such complex problems as the ones described above. Nevertheless, only recently have such packages been used for the study of infilled frames. A thorough discussion of the above concepts and models is presented by Asteris et al. (2013).

Reinforced Concrete Infills in Reinforced Concrete Frames

The construction of new walls is the most effective and economic method for retrofitting multistory reinforced concrete (RC) buildings, especially those with pilotis (soft ground story). Their structural and economic effectiveness increases when selected bays of an existing RC frame are fully infilled with integral RC walls replacing masonry ones. Such a method is appealing since the intervention is concentrated in only certain bays of the frame and reduces the disturbance to the inhabitants as compared to the case of using concrete jackets on the columns.

Most of the experimental research work performed in the last decades has focused on other frequently used types of retrofitting, in particular on fiber-reinforced polymers (FRP) and concrete jackets. Research on the use of RC infill walls has mainly targeted what is feasible: testing of one- to two-story specimens. However, data is lacking for taller full-scale specimens that reflect real-life applications, due to the practical difficulties associated with the high forces needed for the tests. Regarding code provisions, Eurocode 8 – Part 3 fully covers retrofitting with FRP or concrete jackets, while it does not address the retrofitting of RC frames with the addition of new walls created by infilling selected bays.

Experimental research on reinforced concrete frames converted into walls by infilling with RC has been carried out almost exclusively in Japan and Turkey. The experiments in Japan (Chrysostomou et al. 2013, 2014) were performed on a total of 27 1:3–1:4 scale single-story one-bay RC-infilled frames with RC infill walls with a thickness of 26–60 % (on the average 43 %) of the width of the frame members. The test results were compared in most cases with monolithically cast specimens of the same geometric characteristics (in which the frame and the infill wall were cast at the same time and integrally connected). The connection of the RC infill to the bounding frame was done by means of epoxy-grouted dowels (17 specimens) or through mechanical devices, such as shear keys and dowels without epoxy (6 specimens). In four

other test campaigns, the thickness of a preexisting thin wall was increased by 100–150 % without any direct connection of the new wall with the bounding frame. The failure mode of all specimens was in shear (including sliding at the interface). It is interesting to note that for epoxy-grouted dowels, the force resistance of the infilled frame was on average 87 % of the integral one, while for the mechanical connections it was 80 % on average. For the increased thickness of an existing thin infill wall, the force resistance was on average 92 % of the monolithic specimen, while the displacement at failure was on average 13 % smaller than for the integral specimen. For the epoxy-grouted dowels and for the mechanical connection, the ultimate deformation was on average 55 % and 115 % larger than in the integral specimen, respectively. The results show that although a deformable connection gives a somewhat reduced strength with respect to the monolithic case, the ultimate deformation of the retrofitted structure is considerably increased.

Among the specimens tested in Turkey (Chrysostomou et al. 2013, 2014), some were single-story one-bay 1:2 and 1:3 scale, with RC infill thickness 25 % and 33 % of the width of the frame members. Others were two-story one-bay scaled at 1:3, with infill wall thickness 33 % and 40 % of the width of the members of the bounding frame. The RC infill was in most cases fully connected on the perimeter with dowels, in some cases there was a gap between the infill and the columns, while in some other cases there was no connection other than simple bearing. In one case, the rebars of the infill were welded to those of the members of the frame, instead of using dowels, and monolithic specimens (not exactly similar to the infilled ones) were included for comparison purposes. Finally, there was a two-story three-bay specimen scaled at 1:3, with the middle bay infilled with a wall with 63 % thickness of the width of the frame members. The connection was made with epoxy-grouted dowels and the failure mode was predominantly flexural. In all other cases, the single-story walls failed in shear, while the two-story walls failed in a combination of flexure and shear sliding at the base.

The test specimens used in the experiments described in the previous paragraphs correspond to walls with failure modes dominated by shear, with low aspect ratios not representative of multistory slender walls. In fact, the failure mode of multistory slender walls is controlled by bending and the design is governed by the formation of a plastic hinge at the base. In such a case, shear will not have a detrimental effect on displacement and energy dissipation capacity. In addition, it has been shown numerically that higher modes may increase considerably the shear forces at the upper floors of a wall after the formation of a plastic hinge at the base. This aspect has never been studied experimentally because their height and number of stories has not been large enough to allow higher mode inelastic response. Another common element of past tests is the smaller thickness of the RC infill wall relative to the width of the frame members. As a result, the weak link of the structural system is either the infill wall in diagonal compression or its connection with the surrounding frame.

In order to start bridging the gap of knowledge regarding infilling of existing RC frames with RC walls, the effectiveness of seismic retrofitting of multistory multi-bay RC-frame buildings by converting selected bays into new walls through infilling with RC was studied experimentally at the European Laboratory for Structural Assessment (ELSA) of the Joint Research Centre in Ispra (Italy). Similar tests on scaled 0.75:1 specimens were performed at the Structures Laboratory of the University of Patras. Detailed description of the former results is presented by Chrysostomou et al. (2013, 2014) and for the latter by Fardis et al. (2013).

The test specimen in ELSA was designed based on a four-story prototype building structure consisting of four three-bay frames spaced at 6 m, with RC infilling of the exterior frames only. The specimen was designed at full scale to represent the two exterior frames of the prototype structure, spaced at 6 m and linked by a 0.15 m thick RC slab.

The dimension of the specimen in the direction of testing was 8.5 m (two exterior bays of



Assessment and Strengthening of Partitions in Buildings, Fig. 5 Elevation of the specimen in the lab with the south wall on the right of the picture (Chrysostomou et al. 2014 with kind permission from Springer Science and Business Media)

3.0 m and a central bay of 2.5 m), with an inter-story height of 3.0 m and a total height, excluding the foundation, of approximately 12.0 m. The structure was designed for gravity loads only using the code provisions of the 1970s in Cyprus.

In order to facilitate the study of the effect of as many parameters as possible, the walls in the two frames, which had a thickness of 0.25 m equal to the width of the beams and columns of the bounding frame, were reinforced with different amounts of reinforcement, with the north frame being the stronger of the two. Figure 5 depicts the south and north frames, with the south one being on the right of the picture.

The design approached used by Fardis et al. (2013) was adopted and two parameters were examined: (a) the amount of web reinforcement in the walls and (b) the connection detail between the wall and the bounding frame. Regarding the connection with the bounding



Assessment and Strengthening of Partitions in Buildings, Fig. 6 (a) Dowels and starter bars. (b) Dowels, starter bars, and web reinforcement

(Chrysostomou et al. 2014 with kind permission from Springer Science and Business Media)

frame, two distinct connection details were used. In the first detail, the web bars are connected to the surrounding frame through lap splicing with same diameter starter bars epoxy grouted into the frame members. Short dowels are then used in order to transfer the shear at the interface between the wall and the frame members (see bottom beam of Fig. 6a).

In the second detail, longer dowels were used to act both as dowels and as anchorage of the web panel to the surrounding frame (see left column and top beam of Fig. 6a); to this end, the dowels are considered as lap spliced with the nearest – smaller diameter – web bars. The complete wall reinforcement (including web, starter bars, and dowels) is shown in Fig. 5b.

Since the lapping of the column reinforcement can only take compression forces, a lap-splice failure in tension would be highly detrimental to the whole experiment. Therefore, in order to safeguard against this type of failure and allow the experiment to be performed without any premature failure, it was decided to reinforce the bounding columns of the wall at the first floor with three-sided CFRP (carbon-fiber-reinforced polymer) for a height of 0.60 m from the base of the column (Fig. 7).

The testing campaign consisted of two pseudo-dynamic tests (one at 0.10 g and the



Assessment and Strengthening of Partitions in Buildings, Fig. 7 CFRP reinforcement of the column to safeguard against lap-splice failure (Chrysostomou et al. 2014 with kind permission from Springer Science and Business Media)

other at 0.25 g) and a cyclic test. Some findings regarding the behavior of the structure are:

1. The structure managed to sustain an earthquake of 0.25 g without significant damage.
2. Some column lap splices failed with concrete spalling, but the structure continued to carry load.
3. The three-sided CFRPs protected the wall bounding columns at the first floor and prevented lap-splice failure.

4. The “weak” south frame behaved equally well as the “strong” north frame.
5. The slip displacements at the horizontal interfaces of the ground-floor walls were of the order of 0.8 mm, which is very close to the full engagement of the starter bars but not of the dowels.
6. The slip displacements between the wall and the bounding columns of the ground floor were of the order of 0.4 mm.
7. The behavior of the wall was mainly flexural; yielding took place at both the ground-floor and the first-floor walls.
8. The distribution of strains along the bounding columns of the walls shows that the ones for the ground floor are much larger than those of the first floor, while those of the second and third floors are negligible.
9. The two connection arrangements used behaved satisfactorily.
10. Higher mode effects appeared in the response of the structure.
11. Some vertical cracks appeared at the connection of the beams to both the exterior and the wall columns.
12. A horizontal crack appeared at the ground beam of the walls, which was the main cause for the loss of strength of the south frame.

It was demonstrated that this is a viable method for retrofitting and it can be used to strengthen existing ductility and strength-deficient structures. The recorded global and local behavior of the structure provides data for the development of numerical models, to facilitate the proposal of design guidelines for such a retrofitting method.

Strengthening of Partitions

In general, infill walls and in particular the masonry infill panels exhibit a small plastic region on the stress-strain curve due to significant decrease in stiffness, strength, and energy absorption capacity. In order to enhance the behavior of infill partitions and remove their inherent deficiencies, various researchers

have proposed methods for improving the performance of these elements.

One of the methods is to use shotcreting on the faces of masonry infills which can increase the stiffness and the lateral load capacity of the infilled frame and reduce the lateral drift at the ultimate load. Calvi and Bolognini (2001) have performed full-scale tests in which they have placed a 10 mm plaster on both sides of a masonry infill wall covering either reinforcement ($\varnothing 5$ mm or $\varnothing 6$ mm) or wire meshes and have studied the behavior of the strengthened infilled frame. Based on their observations from the experiments, the introduction of some reinforcement in the mortar layers, with a geometrical percentage lower than 1 %, will almost double the acceleration levels for the occupational and damage limit states.

Another technique is the use of fiber-reinforced polymer (FRP) reinforcement to enhance the overall response of such systems. Several experimental studies have been performed which demonstrate that significant improvement of strength and energy absorption capacity can be achieved if adequate anchoring is provided.

Altin et al. (2008) investigated experimentally the behavior of strengthened masonry-infilled reinforced concrete (RC) frames using diagonal carbon fiber-reinforced polymer (CFRP) strips under cyclic loads. Test results indicated that, CFRP strips significantly increased the lateral strength and stiffness of perforated clay brick-infilled nonductile RC frames. Specimens receiving symmetrical strengthening showed higher lateral strength and stiffness, compared to the ones at which CFRP strips of the same width were applied to one of the interior or exterior surface of the infill wall. In the latter case, similar lateral strength and stiffness was obtained, irrespective of the side of placement of the strips.

Lunn and Rizkalla (2009) used glass fiber-reinforced polymer (GFRP) systems for increasing the out-of-plane resistance of infill masonry walls to loading. They have concluded that GFRP strengthening of infill masonry walls is effective in increasing the out-of-plane load-carrying capacity when proper anchorage of the FRP laminate is provided. They also note that the latter has a significant effect on the failure mode of the assemblage.

Papanicolaou et al. (2007) have used the textile-reinforced mortar (TRM), which is a new structural material, for testing its capability for increasing the load-carrying capacity and deformability of unreinforced masonry walls subjected to cyclic in-plane loading. TRMs comprise fabric meshes made of long woven, knitted, or even unwoven fiber rovings in at least two directions. The density of rovings in each direction can be controlled independently, thus affecting the mechanical characteristics of the textile and the degree of penetration of the mortar matrix through the mesh openings.

Based on the experimental results, the authors stated that, in terms of strength, TRM jackets are at least 65–70 % as effective as FRP jackets with identical fiber configurations. In terms of deformability, of crucial importance in seismic retrofitting of unreinforced masonry walls, TRM jacketing for shear walls is about 15–30 % more effective than FRP. They therefore conclude that TRM jacketing is an extremely promising solution for strengthening and seismic retrofitting of unreinforced masonry walls subjected to in-plane loading.

A more recent development is the use of engineered cementitious composites (ECC) which are a special class of fiber-reinforced cement-based composite materials, typically reinforced with polyvinyl alcohol fibers. Deghani et al. (2013) have tested in diagonal compression a number of specimens with different ECC-strengthening configuration, and they evaluated their in-plane deformation and strength properties, including the post-peak softening behavior. They state that the proposed technique can effectively increase the shear capacity of masonry panels (1.5–2.8 times), improve their deformability, enhance their energy absorption capacity (35 times), and prevent the brittle failure mode.

Summary

The presence of infill walls has shown that in most cases they have beneficial effects on the behavior of structures during earthquakes and

have contributed in the prevention of collapse of many structures. It is therefore for this reason that efforts have been made to use engineered infills to retrofit existing seismic-deficient buildings, since they present an economically viable solution. Several equations have been introduced to define the stiffness of infilled frames and their capacities, which are related to a number of failure modes. These were considered in macro- and micro mathematical models that have been proposed to simulate their behavior and their interaction with the bounding frame. The former are simpler (especially the single-strut models), while the latter require a higher level of modeling sophistication. The construction of new walls is the most effective and economic method for retrofitting multistory reinforced concrete (RC) buildings, especially those with pilotis (soft story). Their structural and economic effectiveness increases when selected bays of an existing RC frame are fully infilled with integral RC walls replacing masonry ones. Recent experimental investigations make possible the quantification of this methodology and the proposal of code equations for the design of such systems. Several methods exist for improving the properties of infilled panels avoiding brittle failure and increasing their load-carrying capacity as well as their energy absorption during an earthquake.

Cross-References

- ▶ [Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings](#)
- ▶ [Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions](#)
- ▶ [Seismic Fragility Analysis](#)
- ▶ [Seismic Strengthening Strategies for Existing \(Code-Deficient\) Ordinary Structures](#)
- ▶ [Seismic Vulnerability Assessment: Reinforced Concrete Structures](#)
- ▶ [Strengthening Techniques: Code-Deficient R/C Buildings](#)
- ▶ [Strengthening Techniques: Code-Deficient Steel Buildings](#)
- ▶ [Time History Seismic Analysis](#)

References

- Altin S, Anil Ö, Kara ME, Kaya M (2008) An experimental study on strengthening of masonry infilled RC frames using diagonal CFRP strips. *Compos Part B Eng* 39(4):680–693
- Asteris PG, Antoniou ST, Sophianopoulos DS, Chrysostomou CZ (2011) Mathematical macromodeling of infilled frames: state of the art. *J Struct Eng ASCE* 137(12):1509–1517
- Asteris PG, Cotsovos DM, Chrysostomou CZ, Mohebbkhah A, Al-Chaar GK (2013) Mathematical micromodeling of infilled frames: state of the art. *Eng Struct* 56:1905–1921
- Calvi GM, Bolognini D (2001) Seismic response of reinforced concrete frames infilled with weakly reinforced masonry panels. *J Earthq Eng* 5(2):153–185
- CEN (2004) Eurocode 2: design of concrete structures—part 1: general rules and rules for buildings (EN-1992-1-1), Comité Européen De Normalisation, Brussels
- Chrysostomou CZ (1991) Effects of degrading infill walls on the nonlinear seismic response of two-dimensional steel frames. PhD thesis, Cornell University, Ithaca
- Chrysostomou CZ, Asteris PG (2012) On the in-plane properties of infilled frames. *Eng Struct* 41:385–402
- Chrysostomou CZ, Gergely P, Abel JF (2002) A six-strut model for nonlinear dynamic analysis of steel infilled frames. *Int J Struct Stab Dyn* 2(3):335–353
- Chrysostomou CZ, Poljanšek M, Kyriakides N, Molina J, Taucer F (2013) Pseudo-dynamic tests on a full-scale 4-storey RC frame seismically retrofitted with RC infilling. *Struct Eng Int J IABSE* 23(2):159–166
- Chrysostomou CZ, Kyriakides CN, Poljanšek M, Taucer F, Molina F (2014) Chapter 17: RC infilling of existing RC structures for seismic retrofitting. *Seismic evaluation and rehabilitation of structures, geotechnical, geological and earthquake engineering*, vol 26. doi:10.1007/978-3-319-00458-7_17
- Comité Euro-International du Béton (CEB) (1996) RC frames under earthquake loading. State of the Art Rep., Tomas Telford Services, London
- Crisafulli FJ, Carr AJ (2007) Proposed macro-model for the analysis of infilled frame structures. *Bull N Z Soc Earthq Eng* 40(2):69–77
- Dehghani A, Fischer G, Alahi FN (2013) Strengthening masonry infill panels using engineered cementitious composites. *Mater Struct*. DOI 10.1617/s11527-013-0176-4
- El-Dakhkhni WW (2000) Non-linear finite element modeling of concrete masonry-infilled steel frame, MSc thesis, Civil and Architectural Engineering Department, Drexel University, Philadelphia
- El-Dakhkhni WW (2002) Experimental and analytical seismic evaluation of concrete masonry-infilled steel frames retrofitted using GFRP laminates. PhD thesis, Drexel University
- El-Dakhkhni, WW, Elgaaly M, Hamid AA (2001) Finite element modeling of concrete masonry-infilled steel frame. In: 9th Canadian masonry symposium, The University of New Brunswick, Canada
- El-Dakhkhni WW, Elgaaly M, Hamid AA (2003) Three-strut model for concrete masonry-infilled frames. *J Struct Eng ASCE* 129(2):177–185
- Fardis MN, Schetakis A, Strepelias E (2013) RC buildings retrofitted by converting frame bays into RC walls. *Bull Earthq Eng* 11:1541–1561
- Ghosh AK, Amde AM (2002) Finite element analysis of infilled frames. *J Struct Eng ASCE* 128(7):881–889
- Hamburger RO, Meyer JD (2006) The performance of steel-frame buildings with infill masonry walls in the 1906 San Francisco earthquake. *Earthq Spectra* 22(S2):S43–S67
- Kappos AJ, Ellul F (2000) Seismic design and performance assessment of masonry infilled R/C frames. In: *Proceedings of the 12th world conference on earthquake engineering*, paper no 989 on CD-ROM. Auckland, New Zealand
- Lunn DS, Rizkalla SH (2009) Strengthening of infill masonry walls with FRP materials. *J Compos Constr* 15(2):206–214
- Madan A, Reinhorn AM, Mander JB, Valles RE (1997) Modeling of masonry infill panels for structural analysis. *J Struct Eng ASCE* 123(10):1295–1302
- Mehrabi AB, Shing PB (1997) Finite element modeling of masonry-infilled RC frames. *J Struct Eng ASCE* 123(5):604–613
- Papanicolaou CG, Triantafyllou TC, Karlos K, Papathanasiou M (2007) Textile-reinforced mortar (TRM) versus FRP as strengthening material of URM walls: in-plane cyclic loading. *Mater Struct* 40(10):1081–1097
- Syrmakizis CA, Vratsanou VY (1986) Influence of infill walls to R.C. frames response. In: *Proceedings of the eighth European conference on earthquake engineering*, vol 3, Lisbon, pp 47–53
- Vales RE, Reinhorn AM, Kunnath SK, Li C, Madan A (1996) IDARC2D version 4.0-a program for the inelastic damage analysis of buildings. Technical report NCEER-96-0010, National centre for earthquake engineering research, State University of New York, Buffalo

Assessment of Existing Structures Using Inelastic Static Analysis

Tatjana Isakovic
University of Ljubljana, Ljubljana, Slovenia

Synonyms

Nonlinear static analysis; Pushover-based analysis

Introduction

Over the last few decades, it has been found that the traditional procedures, which are based on elastic linear analysis, can only approximately estimate the typically nonlinear seismic response. Therefore, the inelastic methods of analysis have been gradually introduced into practice in order to estimate the seismic response more realistically. Contrary to the elastic linear methods which can only implicitly predict the performance, the objective of the inelastic seismic analysis procedures is to directly estimate the magnitude of inelastic deformations.

In general, the most refined and accurate inelastic method is the nonlinear response history analysis (NRHA). Nevertheless, it is only sporadically used in the design practice, since it is, for the time being, still too complex for regular use. It requires substantial experience and knowledge about the modeling of the dynamic response of structures and seismic loading. Specialized software is also needed. The results are typically quite extensive, and their interpretation is often time-consuming and too demanding for an engineer with an average knowledge about seismic engineering.

To keep the inelastic analysis relatively simple and make it more apparent for practicing engineers, different static inelastic methods have been developed. Mostly they are based on the pushover analysis. They are considered more user-friendly and relatively easy to understand. They have a great advantage in specifying the seismic input, i.e., they employ the familiar elastic response spectra instead of selecting and scaling ground motion histories (Kappos et al. 2012).

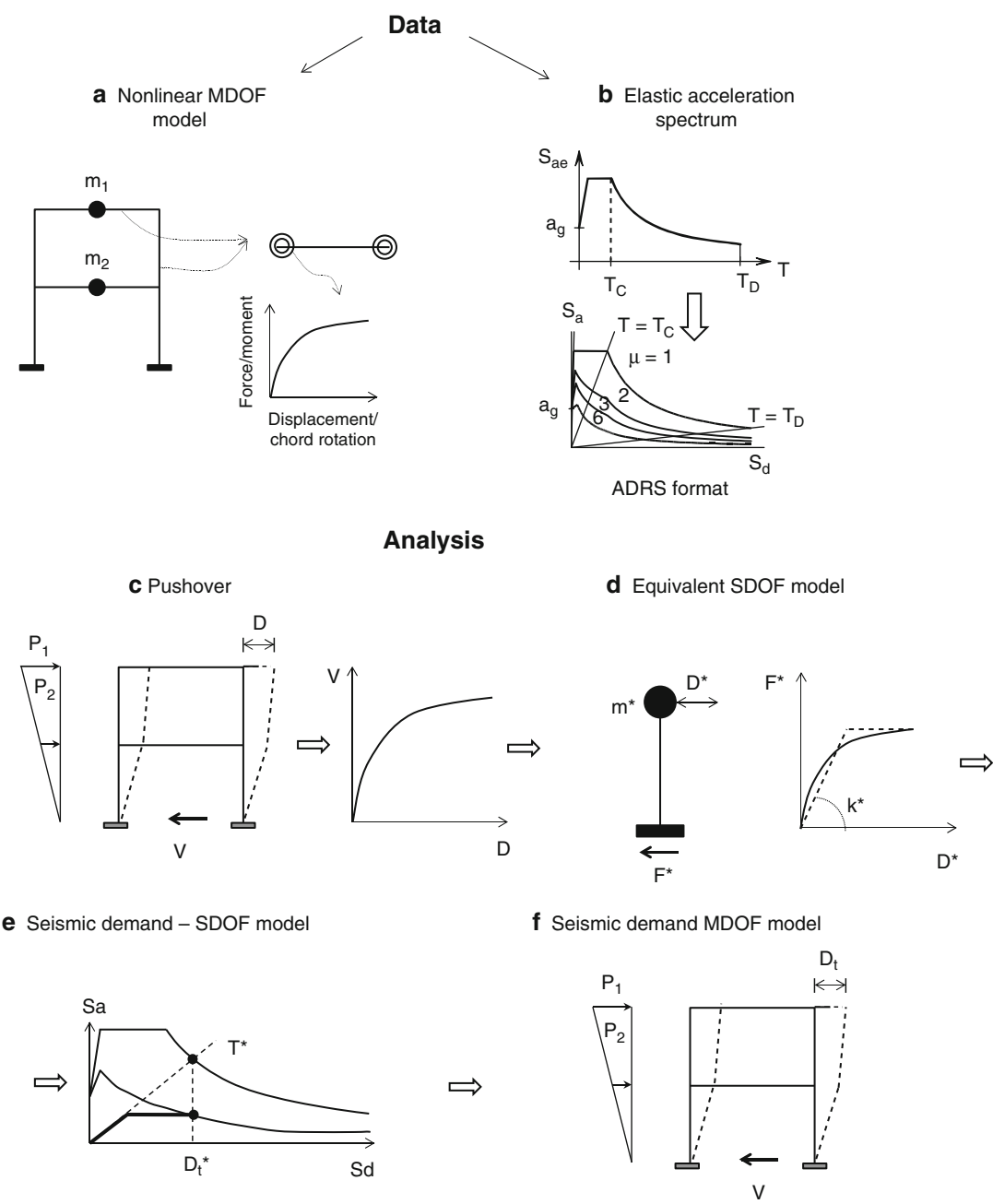
From the historical point of view, pushover analysis has been typically used as a convenient tool to estimate the nonlinear properties and capacity of individual structural components or the whole structure. It typically represents the first step of different static inelastic methods. More specifically, the static nonlinear analysis is performed using the multi-degree-of-freedom (MDOF) model of the analyzed structure. The main purpose of this analysis is to define the properties of the equivalent single-degree-of-freedom (SDOF) model, which is then further

employed to estimate the maximum global displacement demand. This idea was first explored by Saiidi and Sozen (1981). Based on the estimated maximum global displacement demand, other response quantities of interest are then evaluated (see Fig. 1).

In the majority of pushover-based methods, the pushover analysis is performed in a similar manner (see section “[Pushover Analysis: Numerical Models and the Lateral Load Pattern](#)”). However, the techniques that are used to estimate the maximum displacement demand are quite different. The pushover-based methods can be classified with respect to different parameters, e.g., based on (a) the representation of the earthquake input, (b) the type of analysis performed on the SDOF model, and (c) the way how the stiffness of the equivalent SDOF model is defined.

The seismic input is represented by the acceleration response spectrum included in the codes and the displacement spectrum or with a set of accelerograms. The type of analysis is static, dynamic, or response history. The stiffness that is used to calculate the maximum displacement demand is the equivalent pre-yielding stiffness or the equivalent secant stiffness, usually obtained based on the pushover analysis. This variety of solutions can result in a different estimation of the response. In section “[Different Applications of the Inelastic Pushover-Based Analysis, Adopted in the \(Pre\)Standards and Guidelines](#),” those solutions (methods), which are adopted in Eurocode 8/3 (CEN 2005), ATC-40 (ATC 1996), and FEMA-356 (ASCE 2000), the recommended modifications in FEMA-440 (ATC FEMA 2005), and the displacement-based method, included in the NZSEE (2006), are presented. They were selected since they illustrate many of the basic concepts that are currently in use.

These methods are nowadays extensively and successfully used for the analyses of different types of buildings and bridges. Their popularity has increased since they are relatively simple to use but at the same time provide valuable information about the inelastic response, which is not possible to obtain with the elastic methods. They are a useful tool for understanding the general structural behavior corresponding to different



Assessment of Existing Structures Using Inelastic Static Analysis, Fig. 1 Version of the inelastic pushover-based analysis, included in EC8/3

seismic intensity levels. Besides other superior features, they can provide more realistic information about the force demand on brittle elements and the information on the influence of the strength degradation of individual components

on the global behavior of the structure. They can identify the critical regions in the structure. This information is particularly useful when the existing structures have been evaluated and the strengthening techniques have been selected.

However, the pushover-based analysis includes certain assumptions which limit its capabilities. Basic variants of the above code methods can be characterized as single-mode nonadaptive methods, since they assume that the response is controlled by a single predominant mode, which remains unchanged after yielding occurs. This assumption limits their application, particularly when they are used for the assessment of those existing structures which are irregular in plan and/or in elevation. Consequently, the in-plan torsion and higher modes can significantly influence their response. In such cases the extended versions of the single-mode methods are needed. As an alternative, the other types of pushover methods, accounting for the influence of higher modes (multimode methods), can be used. Some issues that influence the applicability and the way of application of the single-mode methods are briefly discussed in section “[Some Issues that Influence the Accuracy of the Inelastic \(Single-Mode\) Pushover-Based Methods.](#)” Some examples of the multimode pushover methods are briefly presented in section “[Some Alternatives to Single-Mode Pushover-Based Methods.](#)”

The assumption that the fundamental mode shape is almost invariant can also limit and complicate the application of nonadaptive methods in existing structures, where the fundamental mode can change considerably depending on the seismic intensity. This can affect the choice of the lateral load pattern in the pushover analysis and make it more complex (see section “[Different Applications of the Inelastic Pushover-Based Analysis, Adopted in the \(Pre\)Standards and Guidelines](#)”). As an alternative, adaptive pushover methods can be used. An example of these methods is briefly presented in section “[Some Alternatives to Single-Mode Pushover-Based Methods.](#)”

The methods presented in sections “[Different Applications of the Inelastic Pushover-Based Analysis, Adopted in the \(Pre\)Standards and Guidelines](#)” and “[Some Alternatives to Single-Mode Pushover-Based Methods](#)” were primarily developed for the analysis of buildings. The response of bridges, particularly in the transverse direction, can be significantly different and more

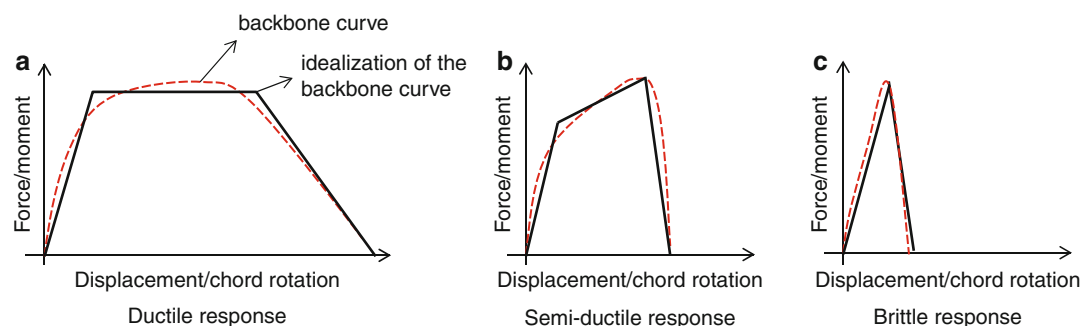
complex. Therefore, some modifications of these procedures are needed. A short overview of the application of the inelastic static methods for the analysis of bridges is presented in section “[Application of the Inelastic Static Methods to the Analysis of Bridges.](#)”

In this contribution, only those inelastic static methods that are included or referred to in the codes and national guidelines are presented. Quite a long list of other inelastic static procedures currently available in the literature can be created. Some of them are general, and some are specialized for certain types of structures. Many of them are presented in different state-of-the-art reports (e.g., FEMA 440 2005; CEB-FIB 2003; Kappos et al. 2012, etc.) or other specialized literature (e.g., Bhatt 2011).

Pushover Analysis: Numerical Models and the Lateral Load Pattern

Pushover analysis is the static nonlinear analysis which is typically performed as the first step of the majority of the inelastic static methods for the seismic assessment of the (existing) structures. The structure is subjected to the lateral load (representing the inertial forces), the intensity of which is gradually increased. The corresponding lateral displacement at a certain location (the reference point or control point) in the structure is recorded. Then the pushover curve, representing the relationship between the base shear of the structure and the registered displacements, is constructed (see Fig. 1c).

The pushover analysis is usually performed employing the MDOF model of the structure which is similar to the linear elastic finite-element models (see Fig. 1a). The most important difference is that the properties of some or all of the components of the model include the post-elastic strength and deformation characteristics in addition to the initial elastic properties. They are usually defined approximating the response observed in the experiments or the response defined by the theoretical analysis of individual components. The envelopes or backbone curves



Assessment of Existing Structures Using Inelastic Static Analysis, Fig. 2 Idealized numerical models

are approximated based on the type of the response. Typical examples are presented in Fig. 2.

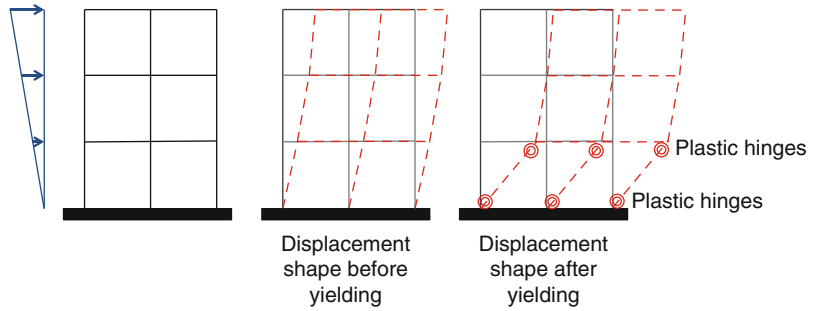
Contrary to the structures designed according to modern codes, where the brittle types of failure are in general avoided, in structures which were designed before the modern principles of seismic engineering were established, different types of brittle or semi-ductile failure of their structural components may be expected. Thus for the reliable estimation of their response, the appropriate numerical models should be employed. Their properties can be defined based on different procedures, defined in the (pre)standards, guidelines, and literature (e.g., CEN 2005, FEMA-356, ATC-40, CEB-FIP 2003). It is, however, worth noting that the available models which describe quite complicated mechanisms that reduce the ductility capacity of the structural components have been less frequently investigated and are often less reliable. The properties of these models are often defined using empirical or semiempirical procedures. The results of these procedures can be considerably different.

For example, the value of the shear strength of RC components can strongly depend on the method used to estimate it. The differences between different methods depend on many parameters, e.g., the amount of the flexural and shear reinforcement, the shear span ratio, the axial force, etc. Moreover, the numerical models that can take into account the complex interaction between shear and flexural response in the nonlinear range are still under investigation and

evaluation (e.g., Mergos and Kappos 2008; Fischinger et al. 2012). Similar observations can be applied to other mechanisms, which reduce the ductility capacity of the structural components and the whole structure. Therefore, it is feasible to explore and compare different available options before the numerical model, and its properties are defined. All uncertainties related to the material properties should also be properly explored.

In the pushover analysis, the MDOF model of the structure is subjected to lateral forces that are intended to simulate the inertial forces expected in the building during an earthquake. They are usually distributed according to a selected (mostly invariant) pattern. It significantly influences the pushover curve and further determines the relative magnitudes of the shear forces, the moments, and deformations within the structure; therefore, it should be carefully selected, depending on the properties of the analyzed structure and the expected response. The distribution of the lateral inertial forces will in general vary continuously during the earthquake response, depending on the gradual formation of the plastic hinges in the structure. To take into account these changes, in many codes and guidelines, it is recommended to perform an analysis, taking into consideration at least two lateral force patterns. Typically the distribution of forces proportional to the fundamental mode of vibration in the elastic range and the uniform distribution are recommended. Then the most adverse results are taken into consideration. Other solutions

Assessment of Existing Structures Using Inelastic Static Analysis, Fig. 3 In some structures (e.g., with “soft” story), the deflection line can considerably change after yielding



recommended in the codes are presented in section “[Different Applications of the Inelastic Pushover-Based Analysis, Adopted in the \(Pre\) Standards and Guidelines](#).”

The approach described in the previous paragraph is in general appropriate for many structures, which are designed according to modern standards. However, in some older existing structures, this solution is not always suitable.

For example, in medium-rise frame buildings, which are designed according to capacity-design principles (where the columns’ flexural strength exceeds the beam flexural strength), the triangular and uniform distribution are typically applied. In existing medium-rise buildings, this is also an appropriate solution, but only if the beam-sway mechanism is the most likely to be developed. In other words, the structure does not include a so-called soft. In opposite cases, the uniform distribution may not be able to frame the range of actions that may occur during the actual dynamic response. Thus a rather complex choice of the lateral force pattern may be needed. The response of such structures in the elastic range will be approximately linear (see Fig. 3). In the nonlinear range, the displacement pattern can considerably change after the formation of the plastic hinges, as is illustrated in Fig. 3. In such cases, some (pre) standards (e.g., FEMA-356, ATC-40) recommend the use of the adaptive load pattern, which can take into account changes of the displacement response shape corresponding to the seismic intensity applied. An alternative solution is the adaptive pushover-based methods (one of which is presented in section “[Some Alternatives to Single-Mode Pushover-Based Methods](#)”). Note, however, that if either the adaptive load

pattern is used or the adaptive methods are applied, the analysis procedure is more complex and more demanding.

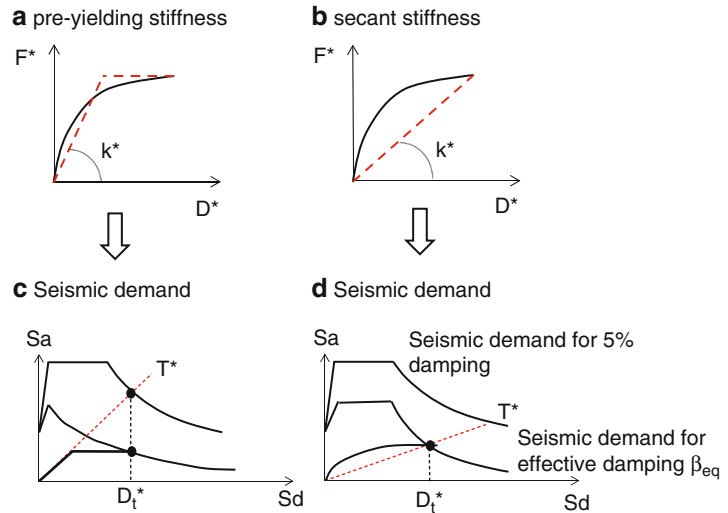
Similar observations to those presented above can be applied, e.g., to the RC dual and wall buildings, with the reinforcement that does not meet the requirements of current codes (e.g., where plastic hinges can form at the elevations above the foundations).

The accuracy of different lateral load patterns in structures where the higher modes have an important influence on the response is discussed in sections “[Some Issues that Influence the Accuracy of the Inelastic \(Single-Mode\) Pushover-Based Methods](#)” and “[Application of the Inelastic Static Methods to the Analysis of Bridges](#)”.

Different Applications of the Inelastic Pushover-Based Analysis, Adopted in the (Pre)Standards and Guidelines

The static pushover analysis has no rigorous theoretical foundation. It is based on the assumption that the response of the structure (MDOF system) can be estimated using the results of the analysis of an equivalent SDOF oscillator (see Fig. 1). This means that it is assumed that the response is governed by one invariant mode of vibration. In general this is incorrect. However, the assumption is approximately fulfilled in many (regular) structures, where the influence of the higher modes is negligible and the deflection shape is almost invariable. Thus the seismic response of these MDOF systems is quite accurately estimated based on the analysis of an equivalent SDOF model. It is not the intention of this article

Assessment of Existing Structures Using Inelastic Static Analysis, Fig. 4 Two different concepts, which are used to define the properties of the equivalent SDOF model



to explain the theoretical background of the formulation of the equivalent SDOF system, since it can be found elsewhere (e.g., Krawinkler and Seneviratna 1998). Here, different applications of this basic concept, which are typically used for the assessment of the existing structures, are overviewed.

In most of the modern (pre)standards and guidelines for the assessment of existing structures, the simplest form of the pushover-based analysis is included. All methods that are included in these documents are based on the same basic concept presented above. However, these methods are not the same. The procedures which are used to define the properties of the equivalent SDOF model and those which are used to define the seismic demand of this model are different. In general two different approaches are used. The properties of the equivalent SDOF system are defined either based on the equivalent pre-yielding stiffness or based on the equivalent secant stiffness (see Fig. 4a, b).

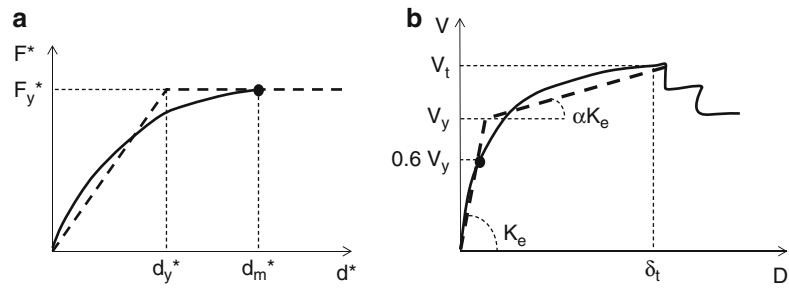
If the first approach is employed, the maximum response of the SDOF oscillator is typically defined based on the 5 % damped acceleration spectra proposed in the codes. The target displacement of the equivalent SDOF system is obtained using the equal displacement rule approximation (see Fig. 4c). Since this approximation is only suitable for the medium- and long-period

structures, the displacements are corrected for short-period structures. This approach is applied in, e.g., Eurocode 8/3 (EC8/3) and implicitly in FEMA-356 (see sections “Eurocode 8/3 (CEN 2005)” and “FEMA-356”). In FEMA-356, the maximum seismic displacements, estimated based on the analysis of SDOF system, are additionally corrected to take into account different issues which are not included in the pushover-based analysis such as strength degradation, P- Δ effect, etc.

If the secant stiffness is used to define the properties of the equivalent SDOF oscillator, typically the overdamped acceleration spectra are used (see Fig. 4d). Actually the capacity spectrum method approach is followed. It is explained later in section “ATC-40” where the method included into the ATC-40 is presented.

In the NZSEE 2006, different inelastic methods are included. The one which is conceptually different from those in other standards is described in this article (see section “FEMA-440”). Contrary to the previously mentioned procedures, the maximum seismic displacement at the reference point is assumed at the beginning of the analysis taking into account the estimated capacity of the structure. It is then used to estimate the properties of an equivalent SDOF model. The dynamic properties of the SDOF model are defined based on the secant stiffness corresponding to the assumed maximum

Assessment of Existing Structures Using Inelastic Static Analysis, Fig. 5 (a) Idealization of the capacity curve in EC8/2, (b) idealization of the pushover curve in FEMA 356



displacement. The seismic demand is estimated using the overdamped displacement spectra, derived from the 5 % damped acceleration spectra.

Since the methods included in different codes are conceptually different, they are presented in more detail in the next subsections. The FEMA-356 and ATC-40 methods were evaluated in the document FEMA-440, where their modifications are proposed. The main observations and the proposed modifications are presented in section “ATC-40.” The matters of accuracy and different issues that can influence the pushover-based analysis of buildings are discussed in section “Some Issues that Influence the Accuracy of the Inelastic (Single-Mode) Pushover-Based Methods.” The estimation of the response of bridges using the inelastic pushover-based analysis is presented in section “Application of the Inelastic Static Methods to the Analysis of Bridges.”

Eurocode 8/3 (CEN 2005)

The method that is included in Eurocode 8/3 (EC8/3) was developed by Fajfar in the 1980s. Its description can be found in (Fajfar 1999). It is included in different parts of the Eurocode 8 standards. Actually the EC8/3 refers to Eurocode 8/1 (CEN 2004) – EC8/1 – and its informative annex B, where the suggested way of application of the method is presented. It is overviewed in Fig. 1.

In the first step, the pushover curve is constructed applying forces proportional to the assumed displacement shape. In EC8/3 it is required to consider at least two force patterns: (a) *modal pattern* – proportional to the lateral forces consistent with the lateral force

distribution determined in an elastic analysis – and (b) *uniform load pattern*.

Based on the pushover analysis, the properties of the equivalent SDOF model are calculated. First, the pushover curve is converted to the capacity curve. Displacements and forces are divided by the transformation coefficient:

$$\Gamma = \frac{\sum m_i \phi_i}{\sum m_i \phi_i^2} \quad (1)$$

m_i is the mass at the location i in the structure (e.g., mass of i th floor of the building), and ϕ_i is a corresponding component of the assumed displacement shape.

Then the *equivalent pre-yielding stiffness* of the equivalent SDOF model k^* is defined as is shown in Fig. 5a. The elastic-perfectly plastic idealization is proposed.

The mass of the SDOF oscillator is determined as

$$m^* = \sum m_i \phi_i \quad (2)$$

The period of vibration of the SDOF model is defined based on the m^* and k^* :

$$T^* = 2\pi \sqrt{\frac{m^*}{k^*}} \quad (3)$$

The target displacement d_{et}^* (inelastic displacement of the equivalent SDOF oscillator) is defined as

$$d_{et}^* = S_e(T^*) \cdot \left(\frac{T^*}{2\pi} \right)^2 \quad (4)$$

$$d_t^* = d_{et}^* \quad (5)$$

for medium- and long-period structures with $T^* \geq T_c$, where T_c is the corner period between the short- and medium-period range. $S_e(T^*)$ is the elastic acceleration in the response spectrum at the period T^* .

The same relationship is used for short-period structures if their response is elastic. If the response is nonlinear, then the target displacement is defined as

$$d_t^* = \frac{d_{et}^*}{q_u} \left(1 + (q_u - 1) \frac{T_c}{T^*} \right) \geq d_{et}^* \quad (6)$$

where q_u is the ratio between the acceleration in the structure with unlimited elastic behavior $S_e(T^*)$ and in the structure with limited strength F_y^*/m^* :

$$q_u = \frac{S_e(T^*)m^*}{F_y^*} \quad (7)$$

In the next step, the seismic displacement of the structure (MDOF model) is defined multiplying the target displacement by the transformation coefficient Γ .

The static nonlinear analysis of the MDOF system is repeated up to the estimated seismic displacement in order to be able to analyze different aspects of the response (e.g., story drifts, shear forces, bending moments). The same loading pattern as in the first pushover analysis is employed.

If the target displacement d_t^* is quite different from the displacement d_m^* used to determine the idealized elastic-perfectly plastic force-displacement relationship, an iterative procedure may be applied.

FEMA-356

In the first step, the pushover analysis is performed. As in the EC8/3, at least two lateral force patterns need to be applied. Three possibilities are defined for the *first load pattern*: (a) forces proportional to the fundamental mode; (b) forces proportional to the values of

coefficient C_{vx} , defined in the standard; and (c) proportional to the story shear distribution calculated by combining the modal responses from a response spectrum analysis of the building. The use of patterns (a) and (b) is limited to structures where more than a 75 % mass participate in the fundamental mode. The pattern (c) can be used in structures where the period of the fundamental mode exceeds 1 s.

For the *second load pattern*, two options are defined: (a) uniform distributions and (b) an adaptive load distribution that changes when the new plastic hinges are formed (see the discussion in section “[Pushover Analysis: Numerical Models and the Lateral Load Pattern](#)”). Different options for the adaptive load distribution are referred to in Fajfar and Fischinger (1988), Eberhard and Sozen (1993), and Bracci et al. (1997).

The pushover curve is idealized as is shown in Fig. 5b. The idealization is bilinear. Note that only the case with a positive post-yield slope is presented in Fig. 5b. In the standard, the case with the negative post-yield slope is also considered. The effective period T_e is defined based on the pre-yielding stiffness, determined in the idealized pushover curve.

The seismic displacement (the term target displacement is used in FEMA-356, but it does not have the same meaning as in the EC8/3) is defined using the so-called coefficient method as

$$\delta_t = C_0 C_1 C_2 C_3 \cdot S_a \frac{T_e^2}{4\pi^2} g \quad (8)$$

where

T_e is the effective fundamental period

S_a response spectrum acceleration at the effective fundamental period

g gravitational acceleration

C_0 is the modification factor to relate the spectral displacement of an equivalent SDOF system to the roof displacement of the building's MDOF system. Actually it has a similar meaning as factor Γ in EC8/3

C_1 is the modification factor to relate the expected maximum inelastic displacements to the displacements calculated for the linear elastic response. The meaning is similar to the

relationship between d_t^* and d_{et}^* in EC8/3. For long- and medium-period structures, C_I is 1, and for short-period structures, it is

$$C_I = \frac{\left(1 + \frac{(R-1) \cdot T_s}{T_e}\right)}{R} \quad (9)$$

R is the ratio of elastic demand to the calculated strength capacity (a similar meaning to q_u in EC8/3).

The C_2 modification factor represents the effect of a pinched hysteretic shape, stiffness degradation, and strength deterioration on the maximum displacement response.

The C_3 modification factor represents increased displacements due to dynamic $P-\Delta$ effects.

ATC-40

While several similarities can be found between the EC8/3 and FEMA-356, the simplified nonlinear procedure in ATC-40 is rather different. It is based on the equivalent linearization. This is a version of the capacity spectrum method, which was first introduced by Freeman et al. (1975). The basic assumption is that the maximum displacement of the nonlinear SDOF system can be estimated from the maximum displacement of a linear elastic SDOF system that has the period and damping ratios that are larger than those of the initial values for the nonlinear system. The elastic SDOF system that is used to estimate the maximum inelastic displacements of the nonlinear system is usually referred to as the equivalent or the substitute system. The period and damping of the equivalent system are referred to as the equivalent period and equivalent damping ratio, respectively.

As in the previous two methods, the pushover analysis is performed first. In general, ATC-40 recommends the distribution of the lateral forces proportional to the fundamental mode pattern. In structures where the response considerably changes after yielding (e.g., in structures with soft stories – see the discussion in section “[Pushover Analysis: Numerical Models and the Lateral Load Pattern](#)”), the adaptive load pattern is required. In high-rise buildings or irregular

buildings, the influence of the higher modes should be properly taken into account. The higher mode effects may be determined by doing higher mode pushover analyses.

The application of the capacity spectrum technique means that both the structural capacity curves and the demand response spectra are plotted in the spectral acceleration versus the spectral displacement domain and compared. Therefore in the next step, the pushover curve is converted to the capacity spectrum curve using the modal shape vectors, participation factors, and modal masses obtained from a modal analysis of the structure. The capacity spectrum curve represents the relationship between accelerations S_a and displacements S_d of the equivalent SDOF oscillator. Then the standard elastic acceleration spectrum (corresponding to 5 % damping) is converted to the ADRS format, where the spectral accelerations are presented as a function of the corresponding spectral displacements (see Fig 4d). In this way, the capacity curve and the seismic demand can be plotted on the same axes and compared.

The capacity spectrum method of equivalent linearization assumes that the equivalent damping of the system is proportional to the area enclosed by the capacity curve. The equivalent period, T_{eq} , is assumed to be the secant period at which the seismic ground motion demands, reduced by the equivalent damping, intersect the capacity curve (FEMA-440). Since the equivalent period and damping are both a function of the displacement, the solution to determine the maximum inelastic displacement (i.e., performance point) is iterative.

The equivalent period T_{eq} and effective viscous damping β_{eq} (which is used to reduce the seismic demand) are defined as

$$T_{eq} = T_0 \sqrt{\frac{\mu}{1 + \alpha\mu - \alpha}} \quad (10)$$

$$\beta_{eq} = 0,05 + \kappa \frac{2 \cdot (\mu - 1) \cdot (1 - \alpha)}{\pi \cdot \mu \cdot (1 + \alpha\mu - \alpha)} \quad (11)$$

where

T_0 is the initial period of vibration of the nonlinear system, α is the post-yield stiffness

ratio, and κ is an adjustment factor to approximately account for the changes in the hysteretic behavior in reinforced concrete structures.

The adjustment factor κ depends on the hysteretic behavior of the system. Three equivalent damping levels are defined. Type A corresponds to structures with reasonably full hysteretic loops, similar to the elastic-perfectly plastic oscillator. Type C corresponds to structures with severely degraded loops, and type B denotes the hysteretic behavior between types A and C. The value of κ and the corresponding equivalent damping is the largest for systems with a hysteretic behavior of type A. Their values decrease for degrading systems B and C. The existing buildings are in general categorized as structures of type C or B, depending on the shaking duration and hysteretic behavior of the structural components.

Based on the comparison of the capacity curve and seismic demand, the target displacement is defined. This displacement is then converted to roof displacement and other aspects of the response are defined.

The ATC-40 method is also well accepted due to the clear and useful visualization of the procedure. Note, however, that the procedure, defined in EC8/3, can also be presented in a similar manner; nevertheless, it is essentially different. The capacity curve as well as the seismic demand, defined in EC8/3, can also be converted to the ADRS format, plotted on the same axes and compared, as is presented in Fajfar (1999) and illustrated in Fig. 4c.

FEMA-440

FEMA-440 is the document where the previously described procedures (FEMA-356 and ATC-40) are evaluated. Improvements of both methods are recommended. Since the document is extended with many important observations, particularly for the existing structures, only some of the conclusions related to the methods included in FEMA-356 and ATC-40 are provided in this section. Some of them are also presented in section “Some Issues that Influence the Accuracy of the Inelastic (Single-Mode) Pushover-Based Methods.”

Some Observations and Proposed Improvements of the Procedure, Included in FEMA-356

It was observed that the characteristic periods which are used to differentiate the response of short- versus medium- and long-period structures were found to be shorter than those observed from nonlinear response history analyses. This can result in an underestimation of the inelastic deformations for the periods between the characteristic period and the periods that are approximately 1.5 times the characteristic period.

The equal displacement rule approximation was found to lead to a relatively good approximation of the maximum inelastic deformations for systems with elastic-perfectly plastic behavior and periods longer than 1 s. This is not always applicable to soft soil sites and near-fault records.

The limiting values of coefficient C_1 , which defines the ratio of elastic and inelastic deformations, can lead to a large underestimation of the displacements in short-period structures. Even if this limitation was not taken into account, the magnification of inelastic displacement demands with a decreasing lateral strength for short-period structures was found to be larger than that suggested by FEMA-356. The corrections of coefficient C_1 were proposed for short-period structures.

It was observed that in many cases the cyclic degradation does not increase the maximum displacements. Thus the use of the related coefficient C_2 is recommended only for structures with significant stiffness and/or strength degradation. Coefficient C_2 is corrected.

It was found that coefficient C_3 , which is intended to represent an increase of displacements due to the dynamic P- Δ effects, cannot adequately take into account the possibility of dynamic instability. It was proposed that this coefficient be eliminated. Instead, it was proposed to use the NRHA for all structures where the strength is below the limit proposed in FEMA-440.

Some Observations and Proposed Improvements in the Procedure Included, in ATC-40

The accuracy of the estimated maximum displacement response in long-period structures depends on the hysteretic behavior type.

In structures with hysteretic behavior type A and periods longer than about 0.7 s, the ATC-40 procedure underestimates the maximum displacements. In structures with a hysteretic behavior type B and periods longer than about 0.6 s, small underestimations or overestimations of the maximum displacements were observed. This depends on the level of lateral strength and on the site class. For structures with hysteretic behavior type C, the ATC-40 procedure leads to an overestimation of the lateral displacements regardless of the period of the structure.

In short-period structures (with periods shorter than noted in the previous paragraph), a significant overestimation of the maximum displacements were observed. The overestimation increases with decreasing strength.

It was also found that the ATC-40 assumption, where it is supposed that the inelastic deformation demands of structures with hysteretic type B will be larger than those in structures with hysteretic type A, does not agree with the results of the NRHA. According to the NRHA, these deformations were approximately the same and in some cases even slightly larger in structures with hysteretic behavior type A. The provisions of ATC-40 do not address the potential dynamic instability that can arise in systems with in-cycle strength degradation and P-D effects.

The suggested improvements of the ATC-40 procedure are presented in chapter 6 of FEMA-440. They are focused on improved estimates of the equivalent period and equivalent damping. It is concluded that generally the optimal effective period is less than the secant period and the optimal effective damping is also less than that specified in ATC-40.

More details about the accuracy, advantages, and drawbacks of both methods can be found in Appendix A of FEMA-440.

NZSEE 2006

The NZSEE 2006 addresses different methods of analysis, referred to as (a) force-based method, (b) displacement-based method, (c) consolidated force-displacement-based method, and (d) method using a nonlinear pushover analysis (the

pushover analysis is performed taking into consideration the same distribution of lateral load as in FEMA-356).

Since the displacement-based method (DBM) is different from those presented in the previous sections, it is described in more detail in the next paragraphs. The basis of the method was developed by Priestley (1995) and further evaluated in, e.g., Priestley et al. (2007). The method places a direct emphasis on establishing the ultimate displacement capacity of the lateral force-resisting elements. Contrary to the methods presented in the previous sections, where the seismic displacements were calculated using acceleration spectra, in the DBM, the displacement spectra are considered.

In the first step, the flexural and shear strengths of the critical sections of the structural components, assuming that no strength degradation occurs due to the cyclic lateral loading in the post-elastic range, are estimated. In the second step, the post-elastic deformation mechanism of the structure and the probable horizontal seismic base shear capacity (V_{prob}) are determined. It is recommended that the post-elastic mechanism is defined using the (a) simple lateral mechanism analysis (presented in the guidelines), (b) inelastic response history analysis, and (c) nonlinear pushover analysis. When the nonlinear pushover analysis is used to estimate the nonlinear response, it is performed in the same manner as in FEMA-356 (the same distribution of the lateral load is taken into account).

In the third step, the plastic rotation capacities of the structural members are determined. It is then eventually corrected if the shear failure occurs before the limits to the flexural plastic rotation capacity are reached. The story inelastic drift capacity is estimated from the plastic rotation capacities.

Considering the critical storey drift and the previously defined post-elastic deformation mechanisms, the overall displacement capacity U_{sc} and the ductility capacity μ are determined at the effective height h_{eff} of the substitute structure (an equivalent SDOF model of the structure). The effective secant stiffness k_{eff} at

the maximum displacement U_{sc} is determined as $k_{eff} = V_{prob}/U_{sc}$. Then the corresponding effective period of vibration T_{eff} is calculated as

$$T_{eff} = 2\pi\sqrt{\frac{M}{k_{eff}}} \quad (12)$$

where M is the effective mass of the substitute structure (an equivalent SDOF model of the structure). Alternatively T_{eff} can be estimated directly using the Rayleigh-Ritz equation.

Based on the evaluated ductility capacity, the equivalent viscous damping ξ_{eff} of the structure is defined. It is recommended to use the method suggested in Pekcan et al. (1999). In general the equivalent viscous damping is determined in a similar manner as in ATC-40, summing the viscous and hysteretic damping.

The structural performance factor S_p is calculated according to NZS 1170.5 (2004) and taking into consideration the detailing used in the structure. It is used to reduce the seismic design actions on a structure. The displacement demand is defined using the overdamped elastic displacement spectrum. It is defined reducing the displacement spectrum $\delta(T)$, derived from the 5 % damped elastic acceleration spectrum. The displacement spectrum $\delta(T)$ is reduced using the correction factor K_ξ :

$$K_\xi = \sqrt{\frac{7}{2 + \xi}} \quad (13)$$

where $\xi = \xi_{eff}$.

Thus the spectral displacement demand of the analyzed structure at height h_{eff} is defined as

$$U_{sd} = S_p(\%NBS)_t \delta(T_{eff}) K_\xi \quad (14)$$

where $(\%NBS)_t$ is the target percentage of the new building standard. The $\%NBS$ is essentially the assessed structural performance of the building compared with the requirements for a new building.

Then the displacement capacity U_{sc} and the demand U_{sd} are compared. If $U_{sc}/U_{sd} \geq 1$, retrofit is unnecessary to achieve $(\%NBS)_t$ and vice versa.

Some Issues that Influence the Accuracy of the Inelastic (Single-Mode) Pushover-Based Methods

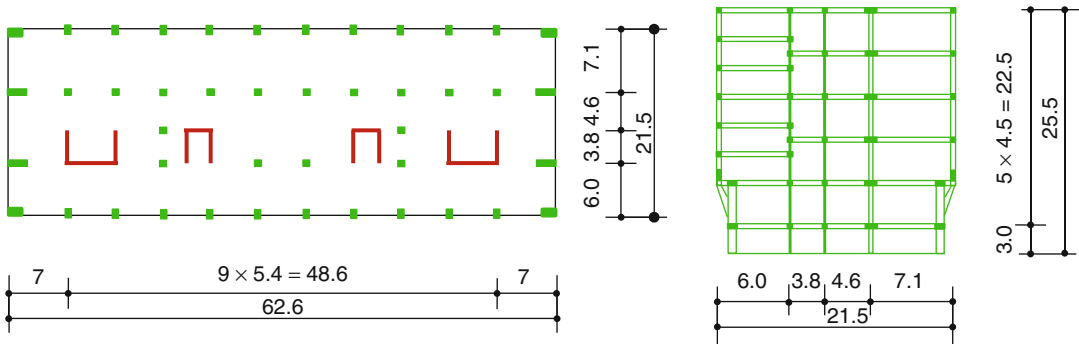
In this section, some issues that can influence the accuracy of the pushover-based methods are presented: the idealization of the capacity curve, the influence of the higher modes, and the influence of the in-plan torsion. In the last two subsections, the basic observations from FEMA-440 about the strength degradation and the soil structure interaction are cited.

Idealization of the Capacity Curve

In some methods presented in section “[Different Applications of the Inelastic Pushover-Based Analysis, Adopted in the \(Pre\)Standards and Guidelines](#),” the actual capacity curve is idealized either by elastic-perfectly plastic or by a bilinear relationship (see Figs. 4 and 5). This approximation is one of the key issues, since it defines the equivalent stiffness of the equivalent SDOF model. This stiffness further influences the period of the SDOF oscillator and the estimated value of the target displacement. Thus the approximation should be carefully performed.

The approximation of the capacity curve is usually performed equating the area bounded (energy) with the actual and idealized curve (see Fig. 5). At this stage, the target displacement should be assumed, because the compared areas depend on the maximum displacement. Thus if the assumed target displacement differs considerably from the value obtained at the end of the analysis of the SDOF system, iterations are strongly recommended. Otherwise the target displacement can be poorly estimated.

In structures that do not exhibit considerable strain hardening in the nonlinear range, the elastic-perfectly plastic idealization is accurate. When the strain hardening is considerable, e.g., in some types of bridges, bilinear idealization is more appropriate (see section “[Application of the Inelastic Static Methods to the Analysis of Bridges](#)” for more details).



Assessment of Existing Structures Using Inelastic Static Analysis, Fig. 6 An example of the structure, irregular in plan and elevation

The Influence of Higher Modes

It was mentioned before that the basic variants of the methods described in section “[Different Applications of the Inelastic Pushover-Based Analysis, Adopted in the \(Pre\)Standards and Guidelines](#)” can be classified as single-mode nonadaptive methods since they assume that the response is controlled by a single predominant mode which remains unchanged after yielding occurs. This assumption limits their application. For example, when they are used for the analysis of existing structures which are irregular in elevation and/or in plan (see an example in Fig. 6), certain extensions are needed in order to take into account the important influence of higher modes and the in-plan torsion. In high-rise buildings, the influence of the higher modes can be important regardless of their irregularity. In such structures, multimode pushover methods (see section “[Some Alternatives to Single-Mode Pushover-Based Methods](#)” for more details) or NRHA can be used as an alternative to single-mode pushover-based methods.

The influence of the higher modes has been the subject of different studies. The one presented in FEMA-440 outlines the most important observations. Some of them are provided in the following paragraphs.

The nonlinear static pushover procedures appeared to be reliable for the design and evaluation of low-rise buildings. In relatively tall frame buildings, where the higher-mode response is significant, interstory drifts, story

shears, and overturning moments can deviate significantly from the NRHA.

In buildings the importance of a higher-mode effect increases with the amount of inelasticity (note that in some bridges quite the opposite trend was observed – see section “[Application of the Inelastic Static Methods to the Analysis of Bridges](#)” for more details). Typical examples are, e.g., RC shear walls, where this phenomenon has been observed years ago. An explanation can be found elsewhere (e.g., Rejec et al. 2012).

Higher-mode contributions become more significant for structures with fundamental periods that fall into the constant-velocity part of the response spectrum. Forces which are developed due to the important influence of higher modes can considerably influence the failure mechanism. Thus the pushover analysis cannot always identify this mechanism.

In FEMA-440, a single first-mode distribution was found sufficient for the estimation of displacement and other quantities that were not significantly affected by higher modes. The adaptive load distribution was sometimes better and sometimes worse than the first-mode distribution. The uniform distribution is recommended in most of the codes as the second choice in order to frame the response quantities. However, in FEMA-440 it was found that it often did not fulfill this role. The uniform distribution was found to be the worst with regard to all the monitored response quantities. Thus it was not recommended as a stand-alone option.

All the codes addressed in this entry recognize that the single-mode pushover-based methods are less efficient when the higher modes are important. Thus they require these methods to be combined with the results of the linear dynamic procedures or the multimode pushover analysis is required. Specific solutions can be found in the particular standard.

The Influence of In-Plan Torsion, Two-Dimensional (2D) Versus Three-Dimensional (3D) Analysis

The in-plan torsion can importantly influence the response of the existing structures which are irregular in plan. In such structures, the 3D analysis is generally needed. The considerable influence of in-plan torsion can be expected in torsionally flexible structures, where the first mode of vibration is torsional. Substantial torsional effects can also be obtained in one direction of structures where the second mode of vibration is torsional.

It was observed that the inelastic static methods can significantly underestimate the displacements on the stiff/strong side of such buildings (see, e.g., Kreslin and Fajfar 2012). Thus the displacements on the stiff/strong side of the structure should be increased.

Most of the (pre)codes recognized this phenomenon. Thus they require an increase in the seismic displacements, defined by inelastic static methods, due to the torsional effects. All of them require a 3D analysis of the structures where the in-plan torsion is important. According to all codes, the accidental eccentricity should be taken into account in all analyses regardless of their torsional flexibility. The specific requirements related to an increase of displacements due to the torsional effects differ from standard to standard.

In general, the 3D pushover-based analysis and torsional effects are topics that require additional investigations. Thus, they are the subjects of numerous researches, reported in the literature. An overview of these researches is recently provided in Bhatt (2011).

The Strength and Stiffness Degradation

The strength degradation, including $P-\Delta$ effects, can lead to an apparent negative post-elastic

stiffness in a force-deformation relationship for a structural model using nonlinear static procedures. The performance implications depend on the type of strength degradation (cyclic or in-cycle strength degradation). For structures that are affected by component strength losses, including $P-\Delta$ effects, occurring in the same cycle as yielding (in-cycle strength degradation), the negative post-elastic slope can lead to the dynamic instability of the structural model (FEMA-440 2005). For this reason, it is suggested that the pushover-based methods are only used if the strength of the structures is above a certain prescribed limit. Otherwise the use of the NRHA is recommended.

The Soil-Structure Interaction

There is a perception among many in the practicing engineering community that short, stiff buildings do not respond to seismic shaking as adversely as might be predicted analytically. There are several reasons why short-period structures may not respond as conventional analysis procedures predict. Among these are (a) radiation and material damping in supporting soils, (b) structures with basements that experience reduced levels of shaking, (c) incoherent input to buildings with relatively large plan dimensions, and (d) inaccuracies in modeling, including dumping of masses, neglecting the foundation's flexibility, and some elements that contribute to the strength (FEMA-440 2006). In FEMA-440 procedures, it is proposed that soil-structure interaction is incorporated into the nonlinear static analyses.

Some Alternatives to Single-Mode Pushover-Based Methods

Multimode procedures are considered as an alternative approach for the analysis of structures where the single-mode methods are less accurate. There are many multimode pushover methods described in the literature. In this section, two of them are briefly presented. The multimode pushover analysis (MPA) (Chopra et al. 2004) is selected as being representative of the

nonadaptive multimode pushover-based methods, and incremental response spectrum analysis (IRSA) (Aydinoğlu 2003) is selected as the representative of the adaptive multimode pushover-based methods. Both methods were addressed in FEMA-440. The MPA was evaluated as an alternative to the single-mode pushover-based methods, and the IRSA was recognized as the potential improvement of the inelastic analysis techniques that can be used to reliably address the MDOF effects.

The MPA Method

The MPA method was developed by Chopra and Goel (e.g., Chopra and Goel 2002). The analysis is performed in a similar manner as that presented in section “Eurocode 8/3 (CEN 2005).” However, the number of the analyses depends on the number of important modes, identified in the initial – elastic – state. The pushover analysis is performed separately for each important mode. The lateral load is proportional to the shape of the vibration mode. Based on each pushover analysis, the contribution of the related mode of vibration to the seismic displacements is defined. These contributions are then combined using the appropriate combination rule (e.g., SRSS). The method supposes that the modes of vibrations are invariant. Thus it can be classified as the nonadaptive method. This limits its applicability. When considerable changes of the mode shapes can be observed in the nonlinear range (an example is described in section “Pushover Analysis: Numerical Models and the Lateral Load Pattern”), the method is in general less reliable.

It was examined in FEMA-440 as an alternative to single-mode pushover-based methods. Taking into account a study of five buildings of very different properties, it was found to be more accurate than the single-mode pushover methods, but not completely reliable. Similar observations were obtained based on other studies, e.g., studies of bridges, presented in Kappos et al. (2012).

The IRSA Method

The IRSA method was developed by Aydinoglu (2003). It is the multimode adaptive pushover

method, which means that it can take into account the influence of the higher modes as well as their changes depending on the seismic intensity. The contributions of the different modes are considered in an incremental pushover analysis.

When the structure enters the nonlinear range, its dynamic properties are changed each time a new plastic hinge is developed. In regular structures, these changes are typically quite small. In irregular structures, the mode shapes as well as their contributions to the overall response can significantly change. IRSA can take into account these changes. More importantly, IRSA is also capable of taking into account the effect of modal coupling to the formation of the plastic hinges.

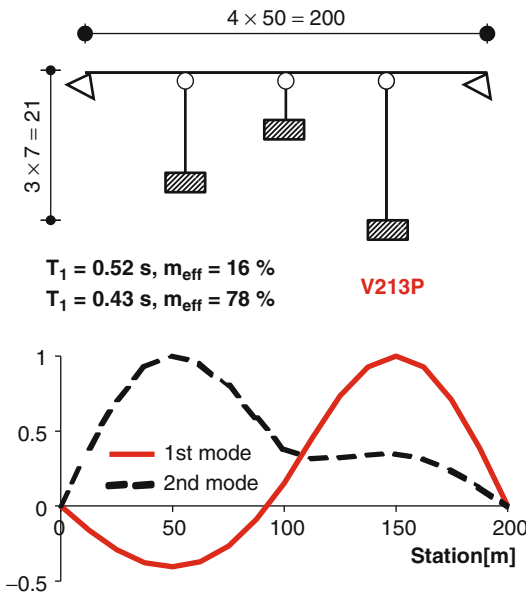
Different studies (e.g., Kappos et al. 2012) have confirmed that the method is quite reliable; however, it also has certain limitations, and thus it is not universal and cannot replace the NRHA in all cases, for example, in certain types of bridges.

Application of the Inelastic Static Methods to the Analysis of Bridges

With regard to their dimensions and structural systems, and in general with regard to their seismic response, bridges are quite different from buildings. Therefore the application of different pushover-based methods, which were originally developed for buildings, is not straightforward, particularly when the bridges are analyzed in the transverse direction. For the analysis of bridges in the transverse direction, the pushover methods in general should be applied in a slightly different manner. The analysis differs mainly regarding: (1) the choice of the reference point where the displacements are registered, (2) the distribution of the lateral load, and (3) the idealization of the capacity curve. These issues are briefly presented in the following subsections. In the last subsection, the applicability of the pushover methods for the analysis of bridges is discussed.

The Reference Point

In buildings, the center of mass of the roof is typically selected as the reference point.



Assessment of Existing Structures Using Inelastic Static Analysis, Fig. 7 An example of a highly irregular bridge

In bridges, this choice is not straightforward. The specialized standards for the design of bridges often recommend the center of mass of the deformed deck as the reference point. An alternative solution could be the top of a certain column. However, in irregular bridges, both of these solutions could be inadequate.

In general, in highly irregular bridges, such as the one presented in Fig. 7, the position of the maximum displacement can considerably vary depending on the seismic intensity. In the bridge, presented in Fig. 7, the mode shapes, their importance, and the ratios are changing depending on the seismic intensity. When the response is in the elastic range, the maximum displacement is observed close to the top of the right pier. When the yielding of the central shortest pier is reached, the station of the maximum displacement is moved toward the center of the bridge (center of mass).

Three quite different pushover curves were obtained when each of the three columns was considered as the reference point (curves P1–P3 in Fig. 8a). Consequently, the dynamic properties of the equivalent SDOF model were also

different. Since the importance of the different modes is changing considerably, depending on the seismic intensity, significantly different displacements of the structure (see curves P1–P3 in Fig. 8c) were estimated based on the pushover curves, presented in Fig. 8a. One can conclude that the pushover curve, corresponding to the column, where the maximum displacements were observed, should be used in the analysis. This is true, so far as this is the station of the maximum displacement of the superstructure, too.

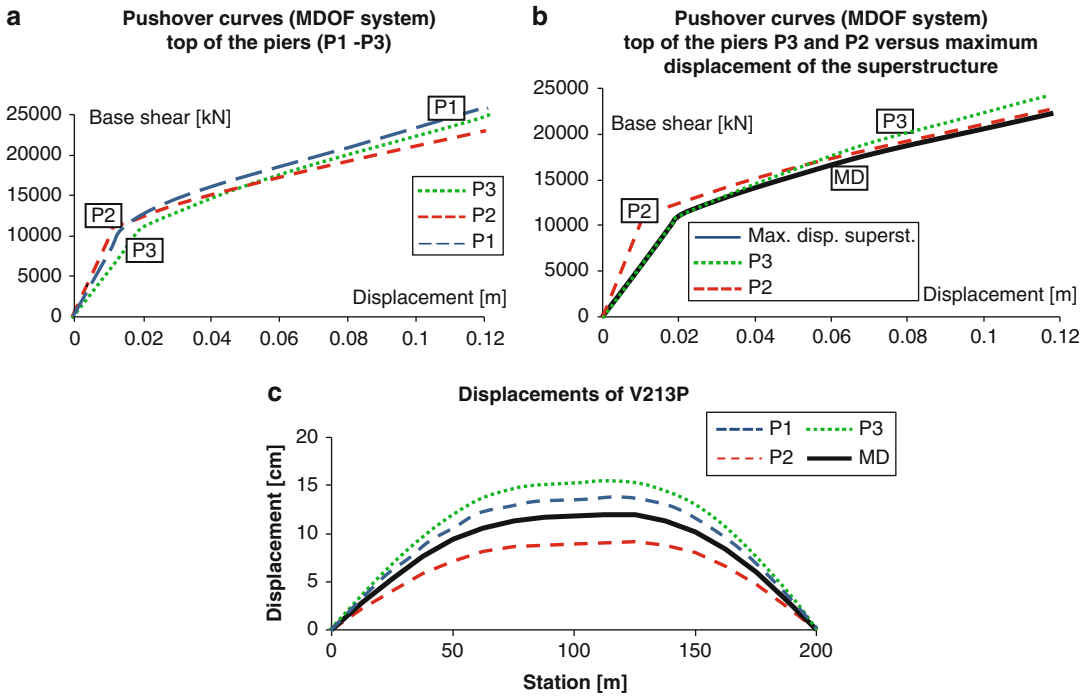
The station of the maximum displacement of the superstructure in a viaduct, presented in Fig. 7, considerably changes depending on the seismic intensity. Thus, the corresponding pushover curve (see curve MD in Fig. 8b) does not coincide with any of the pushover curves constructed based on the displacements monitored at the top of a particular column. The corresponding displacements of the bridge are also significantly different from those calculated using the top of the columns as the reference points (see Fig. 8c).

Since the position of the maximum displacement is variable, it coincides with the center of mass only during stronger seismic intensities. Thus, the same conclusions as those presented in previous paragraphs can be applied for the center of mass, too.

One of the possible solutions is to consider the variable reference point when constructing the pushover curve. This means that the maximum displacement is monitored, wherever it is, in its variable position, which corresponds to a certain load level. In the case presented above, this further means that in the elastic range, the reference point is above the right pier, and after yielding of the central column, it is moved toward the center of mass. The resulting pushover curve is presented by a solid line in Fig. 8b, and the displacements of the superstructure with the bold solid line in Fig. 8c.

The Distribution of the Lateral Load

The lateral load pattern, in general, can be defined following the same basic recommendations as those for buildings: (a) distribution proportional



Assessment of Existing Structures Using Inelastic Static Analysis, Fig. 8 Pushover curves and displacement envelopes, obtained in a highly irregular viaduct, based on different reference points

to the fundamental mode of the bridge in the elastic range and (b) uniform distribution. Note however that the shape of the load pattern, proportional to the fundamental mode, depends on the type of the supports above the abutments as it is presented in Figs. 9 and 10. In regular bridges, pinned at the abutments, the parabolic distribution (see Fig. 9c) is also feasible.

The Idealization of the Capacity Curve

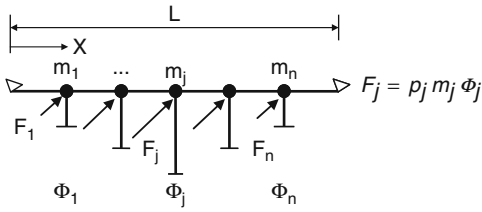
The idealization of the capacity curve can significantly influence the stiffness of the equivalent SDOF model and the estimated value of the maximum seismic displacement. When this stiffness is not adequately estimated, the actual and estimated maximum displacement can be significantly different. In some methods, such as that included in the EC8/3, the capacity curve is approximated using the elastic perfectly plastic idealization. However, in viaducts which are pinned at the abutments, this idealization can be inappropriate, since an underestimated equivalent stiffness of the SDOF system and an

overestimated maximum displacement (see Fig. 11) can be obtained. Namely, in bridges with pinned abutments, the capacity curve can exhibit a considerable strain hardening slope, which should be properly taken into account. This is illustrated in Fig. 11.

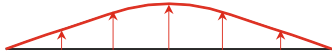
Applicability of Pushover-Based Methods for the Analysis of Bridges

The single-mode methods can accurately predict the response of regular bridges where the influence of higher modes is not important. This is the case where the effective mass of the predominant mode exceeds 80 % of the total mass. When the methods are nonadaptive, they can be accurately used if the mode shapes do not significantly change based on the seismic intensity.

In bridges where the superstructure is considerably stiffer than the supporting elements (piers), the influence of the higher modes is in general negligible. Typical representatives are short- and medium-span bridges (e.g., the length of a bridge is less than 500 m), which are not



a proportional to the 1st mode



$\Phi_j - 1^{\text{st}} \text{ mode}$

b uniform



$\Phi_j = 1$

c parabolic



$$\Phi_j = -\frac{4}{L^2}x^2 + \frac{4}{L}x$$

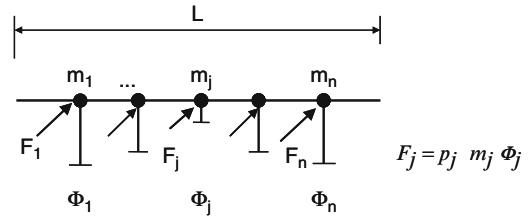
Assessment of Existing Structures Using Inelastic Static Analysis, Fig. 9 Distributions of the lateral load, appropriate for bridges that are pinned at the abutments

supported by very stiff (very short) piers (e.g., in single-column piers, the height of the columns exceeds 10 m).

Considerable changes of the mode shapes can be expected (and were observed) first of all in short bridges, where the displacements of the superstructure above the abutments are not restrained. In such bridges, the predominant mode usually changes considerably, when the damage of the side columns reduces their stiffness to such an extent that the torsional stiffness of the bridge becomes lower than its translational stiffness.

The accuracy of the single-mode pushover-based methods depends on: (a) the ratio of the superstructure stiffness and the stiffness of the bents (the length of the bridge and number and the location of the short columns along the bridge); (b) the relative strength of the columns, compared to the seismic intensity; and (c) the boundary conditions at the abutments (mostly in short bridges).

When the superstructure is stiff compared to the supporting columns (piers), it has a predominant role defining the response of the bridge. In such cases, the response is typically influenced by one predominant mode. However, if the bridge is supported by short stiff columns, they govern the response and cause the



a proportional to the instant 1st mode



$\Phi_j - \text{instant } 1^{\text{st}} \text{ mode}$

b uniform



$\Phi_j = 1$

Assessment of Existing Structures Using Inelastic Static Analysis, Fig. 10 Distributions of the lateral load, appropriate for bridges with roller supports at the abutments

significant influence of the higher modes to the deflection line of the superstructure.

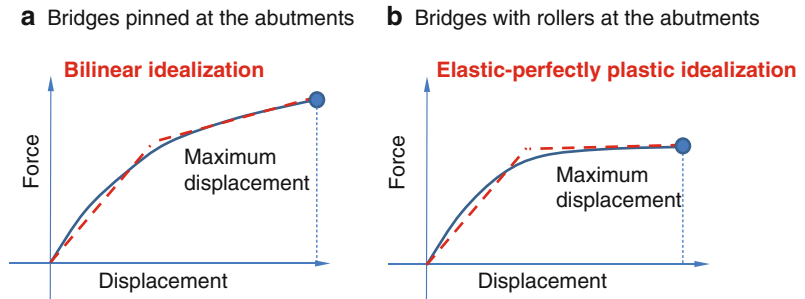
In many bridges, the accuracy of the single-mode methods increases with the seismic intensities (nonlinearity). Columns yield, their stiffness is reduced, and thus the superstructure has a more important role for the overall response. However, this is not the rule. In bridges where the torsional stiffness decreases when the columns yield, the single-mode methods cannot accurately predict the response at high intensity levels, due to the emphasized torsional rotations.

In bridges with roller supports above the abutments, considerable changes of mode shapes can occur, particularly when the side spans are relatively long. When they are supported by short stiff columns in the central part of the superstructure, considerable torsional effects can be obtained.

In long bridges, the influence of the higher modes in the majority of cases does not depend on the stiffness of the columns and their strength. In such bridges, the superstructure of the standard types becomes quite flexible, and consequently the higher modes become important, regardless of the stiffness and the strength of the columns.

When the response of the bridge is considerably influenced by the higher modes or the modes of vibration are changing considerably depending on the seismic intensity, the multimode pushover

Assessment of Existing Structures Using Inelastic Static Analysis, Fig. 11 Idealization of the pushover curve



methods are needed. However, note that they also have certain limitations, which depend on their basic assumptions. For example, even quite accurate methods, such as IRSA (see section “[Some Alternatives to Single-Mode Pushover Based Methods](#)”), can fail to predict the response accurately when the bridge is supported by short stiff central columns. In such cases, the NRHA is needed.

More details and recommendations about the use of the pushover-based methods for the analysis of bridges can be found in Kappos et al. (2012).

Summary

The inelastic static pushover-based methods, which are typically used for the assessment of the existing structures, are presented. Some of the basic concepts that are nowadays used are presented in the example of the four single-mode pushover-based methods included in the codes. Some parameters that influence the accuracy of these methods are presented: selection of the lateral load pattern, idealization of the capacity curve, numerical models, influence of the higher modes, and changes of the shapes of the vibration modes depending on the seismic intensity, in-plan torsion, strength degradation, and soil-structure interaction.

Two multimode pushover-based methods, which are typical examples of nonadaptive and adaptive multimode methods, are briefly overviewed. The application of the inelastic pushover-based methods to the analysis of bridges is discussed.

Cross-References

- ▶ [Assessment and Strengthening of Partitions in Buildings](#)
- ▶ [Assessment of Existing Structures Using Response History Analysis](#)
- ▶ [Earthquake Response Spectra and Design Spectra](#)
- ▶ [Earthquake Return Period and Its Incorporation into Seismic Actions](#)
- ▶ [Equivalent Static Analysis of Structures Subjected to Seismic Actions](#)
- ▶ [European Structural Design Codes: Seismic Actions](#)
- ▶ [Incremental Dynamic Analysis](#)
- ▶ [Modal Analysis](#)
- ▶ [Nonlinear Dynamic Seismic Analysis](#)
- ▶ [Nonlinear Finite Element Analysis](#)
- ▶ [Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings](#)
- ▶ [Response Spectrum Analysis of Structures Subjected to Seismic Actions](#)
- ▶ [Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used](#)
- ▶ [Retrofitting and Strengthening Measures: Liability and Quality Assurance](#)
- ▶ [Retrofitting and Strengthening of Contemporary Structures: Materials Used](#)
- ▶ [Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions](#)
- ▶ [Seismic Analysis of Concrete Bridges: Numerical Modeling](#)
- ▶ [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)
- ▶ [Seismic Analysis of Steel and Composite Bridges: Numerical Modeling](#)

- ▶ **Seismic Analysis of Steel Buildings: Numerical Modeling**
- ▶ **Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling**
- ▶ **Strengthened Structural Members and Structures: Analytical Assessment**
- ▶ **Strengthening Techniques: Bridges**
- ▶ **Strengthening Techniques: Code-Deficient R/C Buildings**
- ▶ **Strengthening Techniques: Code-Deficient Steel Buildings**
- ▶ **Strengthening Techniques: Masonry and Heritage Structures**

References

- ATC 40, ATC (1996) Seismic evaluation and retrofit of concrete buildings, ATC-40 report, funded by the Seismic Safety Commission, published by the Applied Technology Council. Redwood City, California
- Aydinoglu NM (2003) An incremental response spectrum analysis procedure based on inelastic spectral displacements for multi-mode seismic performance evaluation. *Bull Earthq Eng* 1:3–36
- Bhatt CAAF (2011) Seismic assessment of existing buildings using nonlinear static procedures (NSPs) – a new 3D pushover procedure, PhD thesis. Technical University of Lisbon, Instituto Superior Técnico, Lisbon
- Bracci JM, Kunnath SK, Reinhorn AM (1997) Seismic performance and retrofit evaluation of reinforced concrete structures. *J Struct Eng* 123(1):3–10
- CEB FIP (2003) Displacement-based seismic design of reinforced concrete buildings; state-of-the-art-report, fib Bulletin No 25, Lausanne, Switzerland
- CEN (2004) Eurocode 8: design of structures for earthquake resistance – part 1: general rules, seismic actions and rules for buildings. Comité Européen de Normalisation, Brussels
- CEN (2005) Eurocode 8: design of structures for earthquake resistance. Part 3: strengthening and repair of buildings, EN 1998–3. European Committee for Standardization, Brussels
- Chopra AK, Goel RK (2002) A modal pushover analysis procedure for estimating seismic demands for buildings. *Earthq Eng Struct Dyn* 31:561–582
- Chopra AK, Goel RK, Chintanapadtee C (2004) Evaluation of a modified MPA procedure assuming higher modes as elastic to estimate seismic demands. *Earthq Spectra* 20(3):757–778
- Eberhard MO, Sozen MA (1993) Behavior-based method to determine design shear in earthquake-resistant walls. *J Struct Div Am Soc Civil Eng N Y* 119(2):619–640
- Fajfar P (1999) Capacity spectrum method based on inelastic demand spectra. *Earthq Eng Struct Dyn* 28(9):979–993
- Fajfar P, Fischinger M (1988) N2 – a method for non-linear seismic analysis of regular structures. In: *Proceedings of the ninth world conference on earthquake engineering, Tokyo-Kyoto*
- FEMA 356, ASCE (2000) Prestandard and commentary for the seismic rehabilitation of buildings, FEMA 356 report, prepared by the American Society of Civil Engineers for the Federal Emergency Management Agency, Washington, DC
- FEMA 440, ATC, FEMA (2005) FEMA-440, improvement of nonlinear static seismic analysis procedures, applied technology council for department of Homeland Security, Federal Emergency Management Agency, Washington, DC
- Fischinger M, Rejec K, Isakovic T (2012) Modelling Inelastic shear response of RC walls. In: *Proceedings of 15th world conference on earthquake engineering, Sept 2012, Lisbon*
- Freeman A, Nicoletti JP, Tyrell JV (1975) Evaluations of existing buildings for seismic risk – a case study of Puget Sound Naval Shipyard, Bremerton, Washington, DC. In: *Proceedings of the 1st U.S. national conference earthquake engineering, EERI, Berkeley*, pp 113–122
- Kappos AJ, Saiidi MS, Aydinoglu MN, Isakovic T (2012) Seismic design and assessment of bridges: inelastic methods of analysis and case studies. Springer, Dordrecht/Heidelberg/New York/London
- Krawinkler H, Seneviratna GDDPK (1998) Pros and cons of a pushover analysis of seismic performance evaluation. *Eng Struct* 20(4–6):452–464
- Kreslin M, Fajfar P (2012) The extended N2 method considering higher mode effects in both plan and elevation. *Bull Earthq Eng* 10(2):695–715
- Mergos PE, Kappos AJ (2008) A distributed shear and flexural flexibility model with shear-flexure interaction for R/C members subjected to seismic loading. *Earthq Eng Struct Dyn* 37(12):1349–1370. doi:10.1002/eqe.812
- NZS 1170.5 (2004) Structural design actions, part 5: earthquake actions – New Zealand. Standards New Zealand, Wellington
- NZSEE (2006) Assessment and improvement of the structural performance of buildings in earthquake. New Zealand Society for Earthquake Engineering, Wellington
- Pekcan G, Mander JB, Chen SS (1999) Fundamental considerations for the design of non-linear viscous dampers. *Earthq Eng Struct Dyn* 28:1405–1425
- Priestley MJN (1995) Displacement-based seismic assessment of existing reinforced concrete buildings. *Proc Pac Conf Earthq Eng* 2:225–244. Melbourne
- Priestley MJN, Calvi GM, Kowalsky MJ (2007) Displacement-based seismic design of structures. IUSS Press, Pavia
- Rejec K, Isakovic T, Fischinger M (2012) Seismic shear force magnification in RC cantilever structural walls, designed according to Eurocode 8. *Bull Earthq Eng* 10(2):567–586
- Saiidi M, Sozen M (1981) Simple nonlinear seismic analysis of R/C structures. *J Struct Div ASCE* 107(ST5):937–952

Assessment of Existing Structures Using Response History Analysis

Damian N. Grant

Arup Advanced Technology and Research,
London, UK

Synonyms

Nonlinear dynamic analysis; Nonlinear response history analysis; Time history analysis

Introduction

Nonlinear response history analysis (NRHA) is used in both engineering industry and academia to assess the seismic performance of structures and to validate simpler analysis methods used in design or assessment. International building codes for the design of new buildings and assessment guidelines for existing buildings generally allow NRHA to be used, although the level of detail in published guidelines is mixed, and industry practice varies. In the past, NRHA was the domain of only a few specialists, but many structural analysis packages are now available to carry it out, and many engineering consultancies have the capability.

NRHA is the most general and most detailed analytic modeled approach in the earthquake engineer's toolkit, involving the numerical solution of the equation of the motion allowing for both material and geometric nonlinearity. The response of a structure to an earthquake ground motion is evaluated at small increments of time, tracking the development of plasticity in the structure as it deforms into the inelastic range. Different algorithms are available to numerically solve for forces and displacements at the current time step in terms of those at the previous time step and the applied ground acceleration over the time interval. Analysis programs commonly used in earthquake engineering use *implicit* solution algorithms, many of which are unconditionally stable (i.e., any analysis time step will lead to

a stable evaluation of response, although small time steps are still needed for accuracy), but require iteration within each time step. Some packages used more for large-deformation analysis, such as those used in the automotive industry for crash simulation, use *explicit* solution algorithms which do not require iteration, but with the trade-off of requiring a very small time step for numerical stability (depends on application, but often of the order of 10 μ s).

The most appropriate analysis methodology is very application-dependent and can evolve as a project progresses. If it is possible to demonstrate acceptable performance of a building on the basis of simple, conservative analyses, then a very detailed NRHA may require excessive engineering resource and time. On the other hand, for unusual structures or those for which initial findings are unclear, NRHA is often justified. This is especially true for the assessment of existing structures as the potential benefit of demonstrating compliance with target performance requirements (and therefore avoiding retrofit) often outweighs the engineering costs associated with advanced analysis. Existing structures are often deficient with respect to new design requirements – both qualitatively, in terms of code-required detailing, and quantitatively, in terms of seismic demand exceeding seismic resistance. This means that simplified code inputs and methods developed for code-compliant structures (such as allowable ductility levels) may not be appropriate for noncompliant structures, and more detailed analysis may be required to assess their performance.

There are many aspects of NRHA that can affect estimates of structural response, relating to appropriate numerical inputs, modeled methodologies for different materials and structural elements, and treatment of analysis outputs. On the input side, ground acceleration histories are required that are in some way consistent with the seismic hazard at the site (e.g., a code design spectrum). Different authors recommend the use of raw unscaled ground motions recorded in real earthquakes, scaled or otherwise modified real ground motions (including spectrally matched records, which are modified such that their

response spectrum matches a target spectrum), ground motions developed from physics-based simulations of fault rupture and seismic wave passage, and those generated artificially from the modification of white noise or some other mathematical approach. On the methodological side, structures can be modeled using a very detailed 3D finite element model, simplified with 2D shell elements (especially appropriate for structural walls and diaphragms) or simplified even further into 1D “stick” elements, routinely used for moment frame and braced frame structures. Nonlinear material behavior for the latter approach may be taken into account with distributed plasticity (numerical integration of the nonlinear stress–strain behavior over the length of the element) or lumped plasticity (assuming plasticity may be concentrated in a plastic hinge at the end of each member). Finally, on the output side, there are many different response quantities that can be monitored, including interstory drifts, base shear, and plastic deformations, and it is important to be able to link real descriptions of structural performance (from onset of cracking through to full structural failure) to the outputs of the analytic model. Each analyst will treat each of these aspects in different ways, potentially leading to very different estimates of seismic response of a given structure.

This entry outlines the application of NRHA to the assessment of seismic performance of existing structures, with an emphasis on those structures that do not meet prescriptive code requirements. Section “[Performance Objectives and Assessment Methodology](#)” discusses the setting of performance objectives and how these are affected by the assessment methodology adopted. Section “[Modeling Considerations](#)” covers general aspects of nonlinear dynamic analysis, including modeled considerations and interpretation of the outputs. Section “[Published Codes, Standards and Guidelines](#)” is a summary of international published guidelines on the topic, including ASCE/SEI 41-06, Eurocode 8 Part 3, the New Zealand Society for Earthquake Engineering Recommendations, and FEMA P-58. Finally, section “[Summary](#)” is a summary of the entry.

Performance Objectives and Assessment Methodology

Target seismic performance for assessment of existing structures depends on both the needs of building owners or occupants and on regulatory requirements. In many jurisdictions, seismic assessment and retrofit are carried out on a voluntary basis, and in this case, there is significant flexibility in setting performance objectives. This does not remove the responsibility of the engineer to educate the client on the seismic risk represented by different objectives. Many clients come into the assessment process with incomplete performance targets in mind, such as “to ensure life safety” or “functional performance after an earthquake,” without a full appreciation of the uncertainties in seismic hazard definitions or the costs involved in achieving different levels of performance. Indeed, even engineers often misleadingly refer to performance objectives solely in terms of structural performance and do not refer to the fairly arbitrarily defined hazard levels with which these are associated.

Performance objectives will generally be expressed in terms of a target structural performance under one or more seismic hazard levels or a time period of exposure to the hazard. Performance objectives are expressed differently depending on the analysis to be carried out. Three potential analysis types are described in FEMA P-58 (section “[FEMA P-58: Seismic Performance Assessment of Buildings](#)”):

1. *Intensity-based assessment*, in which seismic performance and losses are evaluated for seismic input defined by a given response spectrum
2. *Scenario-based assessment*, in which the seismic input is a specific earthquake scenario, such as an historical event or potential rupture of a particular fault
3. *Time-based assessment*, in which losses are integrated over the occurrence of many potential earthquakes in a fully probabilistic way to allow estimates of, e.g., annualized seismic losses.

For intensity-based assessment (which is the default option in most codes and standards for seismic assessment, including ASCE 41, EC8-3, and the NZSEE Recommendations, reviewed in Section “[Published Codes, Standards and Guidelines](#)”), one or more performance objectives will be checked, each comprising a return period at which the probabilistic seismic hazard is evaluated and the desired structural and nonstructural performance at this hazard level. In ASCE 41, for example, the “basic safety objective” (BSO) is considered to be satisfying a life safety performance level under a 475-year return period ground motion and a collapse prevention performance level under a 2,475-year return period. Nonstructural performance objectives are also given, including a life safety objective to ensure that heavy equipment or façades do not fall and endanger life or impede egress.

Performance objectives both for assessment and (if required) retrofit design for existing structures are often allowed to be less onerous than in the design of new structures. This is justified primarily on the basis of pragmatism: incorporating seismic-resistant detailing into new buildings usually comes at a relatively small premium, whereas the cost of retrofit (including both direct and indirect costs) may be prohibitive. This is especially true of historical structures, where extensive strengthening measures may not be acceptable from a cultural heritage point of view (of course, this cultural heritage is exactly what seismic strengthening is trying to preserve in the long term). Less onerous performance objectives are also sometimes justified on the basis of a lower remaining design life than the nominal value adopted for existing structures (and therefore lower exposure period to seismic hazard). The BSO in ASCE 41, for example, is considered to represent a potentially higher risk than expected for properly designed and constructed new buildings. In jurisdictions where seismic assessment and retrofit are not obligatory, building owners may decide on performance objectives for themselves, based on a cost-benefit analysis of the cost of strengthening and the risk reduction in terms of future repair costs and potential loss of life following an earthquake.

The main goal of the latest generation of performance-based earthquake engineering guidelines (such as FEMA P-58) is to develop tools to convert engineering performance objectives into metrics that are more meaningful to other stakeholders, such as building owners, building tenants, and insurance companies. Performance may be quantified in terms of direct monetary costs (for repair or replacement), human impact (chance of casualties or fatalities), and indirect costs (time to repair, reoccupy, or return to function) – snappily summarized by the 3 Ds: “dollars, deaths, and downtime.” Performance objectives expressed in this format of course place a higher burden on the analyst – particularly for areas outside the United States, where the default fragility functions given in *PACT* for various nonstructural components may not apply. The recently developed REDiTM Rating System (Resilience-based Earthquake Design Initiative; Almufti and Willford 2013) bases its platinum, gold, and silver performance resilience objectives partly on numerical estimates of post-earthquake reoccupancy and functional recovery times and provides a useful framework for discussion of enhanced performance objectives with clients and other stakeholders.

Modeling Considerations

As with any engineering or scientific numerical modeled exercise, the outputs from NRHA are only as reliable as the specific modeled techniques used and the numerical inputs. Although NRHA is perceived as giving the most accurate estimates of building seismic performance, it has the disadvantage that model checking (of both inputs and outputs) may be hampered by complexity and potentially by computational power (e.g., the time taken to envelope results of large structural analysis models over multiple ground motions can discourage the analyst from checking all response quantities thoroughly). Analysts may be overconfident of the results of NRHA and feel that detailed modeled obviates the need to apply engineering principles to the assessment of structures.

The following subsections provide a summary of some modeled decisions and interpretations that face an analyst employing NRHA on an assessment project. Many of these aspects are covered in more detail in other chapters in this entry. The discussion is divided into aspects relating to the seismic input (ground acceleration histories), structural modeled, and treatment of outputs from the analysis.

Seismic Input

From the point of view of the structure, seismic action is felt through acceleration of the ground at the base – in both horizontal and vertical directions – and therefore NRHA requires appropriate ground acceleration histories to apply. In most cases, the seismic hazard is defined in the form of a response spectrum, and therefore compatibility with a target spectrum is one of the most important considerations in definition of ground acceleration histories, but nonstationary characteristics such as duration can also be important. Results can be very sensitive to the ground motions applied, and the ground motions to be used for a given application depend not only on the seismic hazard and expected seismic shaking at the site but also on the objectives of the analysis.

For the assessment of mean (expected) response only, current codes allow as few as three ground motions to be assessed, although if fewer than seven are used, the maximum of each response quantity across all the analyses should be used. These requirements date back to the level of computing power available in the 1980s and 1990s, and nowadays, seven is generally considered to be a sensible minimum, from which the arithmetic mean response can be taken (NEHRP Consultants Joint Venture 2011). For estimates of the statistical distribution of structural response quantities, many more ground motions will be required – NEHRP Consultants Joint Venture (2011) suggests 30 records, although this depends on the variability of structural response and the ground motions affecting the site.

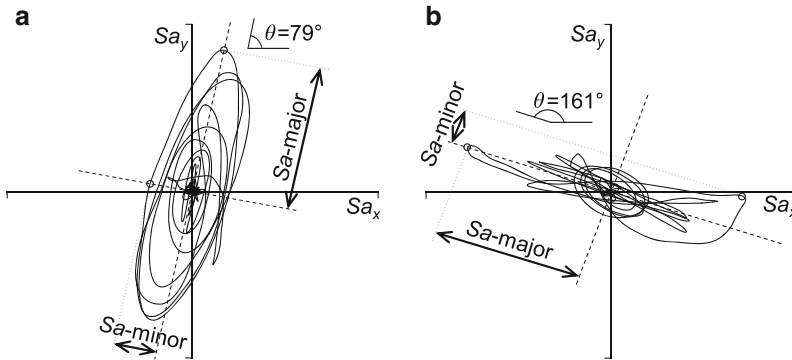
There is a substantial body of literature on the development of ground motions appropriate for the NRHA of structures, and no consensus has yet

been reached among researchers and codes of practice on the most appropriate methods. Categories of ground motions applied in seismic assessment are:

- Recorded by accelerometers in real earthquakes, applied to the structural model unmodified
- Recorded in real earthquakes multiplied by a linear scaling factor to achieve compatibility with the peak ground acceleration or other scalar measure of seismic intensity (e.g., spectral acceleration at the fundamental period of the structure)
- Recorded in real earthquakes modified to achieve spectrum compatibility across a range of structural periods (*spectral matching*)
- Generated from a purely mathematical algorithm, such as modified white noise
- Derived from a physics-based simulation of the seismic wave generation and propagation process.

NEHRP Consultants Joint Venture (2011) provides a discussion of the various considerations for development of ground motions for a project.

Most literature on ground motion development considers only a single component of horizontal input and does not discuss the proper treatment of horizontal variability and the application of multiple components of ground motion to a 3D analysis model. Ground motions recorded in real earthquakes are typically not axisymmetric – i.e., they do not shake each horizontal direction equally strongly. In fact, in some cases, structures with a certain period may be shaken strongly in one direction, whereas structures with a different period may be shaken more strongly in a different direction. Figure 1 shows an example of such a case, where the 1.3-s structural response is strongest at 79° from the x -axis and the 3.0-s structural response is strongest at 161°. Grant (2011) developed the software, *RspMatchBi*, to spectrally match recorded ground motions to retain the axes of strong shaking of the original motion while matching the



Assessment of Existing Structures Using Response History Analysis, Fig. 1 Response orbits (displacement of single degree of freedom oscillators) and definition of

major and minor axes of response for PEER Record 184 from Imperial Valley event; (a) 1.3-s period response and (b) 3.0-s period response (From Grant et al. 2011)

spectra of the record to target spectra. NEHRP Consultants Joint Venture (2011) provides some discussion on the application of multicomponent ground motions to a structural model.

Structural Modeling

The most appropriate type of response history analysis model (or analysis model in general) is very application specific. If a simple and conservative model is adequate to demonstrate acceptable seismic performance, then there is little benefit in introducing extra detail and complexity. In practice, modeled requirements are different for every project, and it is not possible to give a recipe that can be followed for every application. The following are some aspects of structural analysis that are important for the use of NRHA in both the assessment of existing structures and in new design.

Linear Versus Nonlinear

Although the focus of this entry is on nonlinear analysis, the extra effort involved to introduce material and geometric nonlinearity is not always justified – particularly when target performance objectives restrict the response to the elastic or near-elastic range. This also applies at a component level – for example, force-controlled (brittle) elements are usually modeled linearly, given that their nonlinear response is generally unacceptable. Of course, even if heavily nonlinear response is acceptable, if

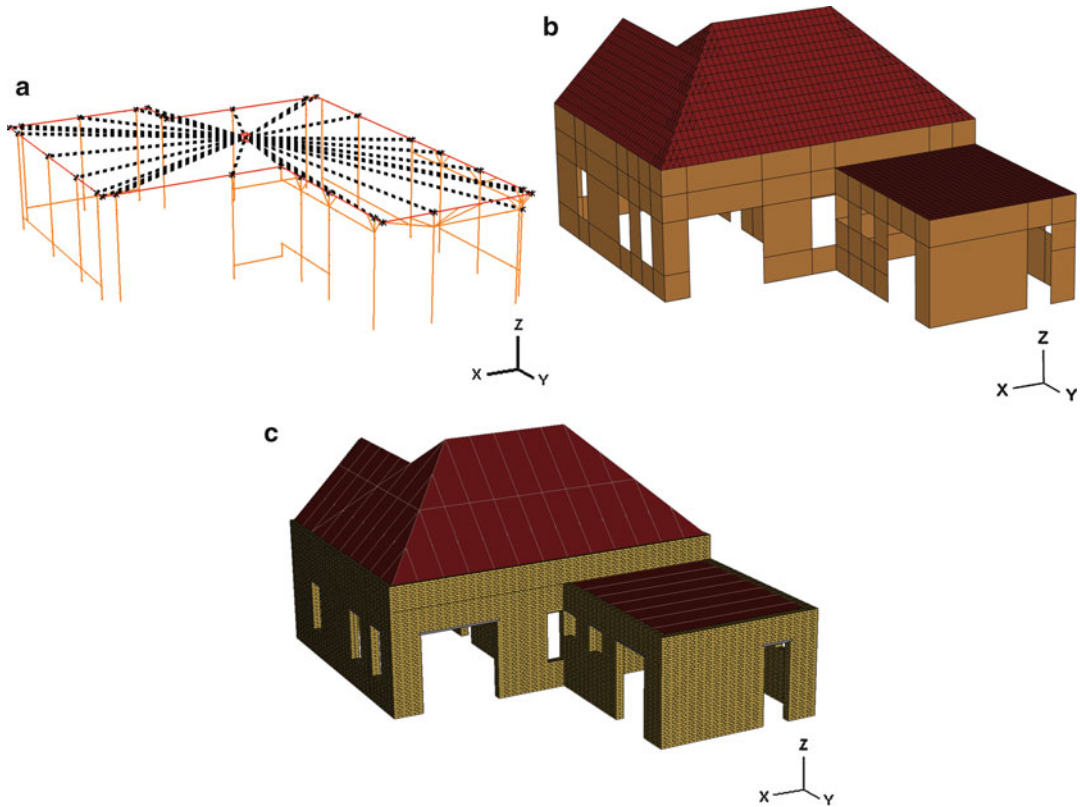
input seismic demand is lower than the elastic response of the structure, nonlinear deformation is not expected, and a linear model would suffice. On the other hand, when a degree of nonlinearity is expected in the structure's seismic response, explicitly modeled this nonlinearity may be necessary, allowing a more accurate estimation of the overall structural response (e.g., peak displacements or storey drifts), but also a better understanding of how seismic demand is plastically redistributed among parallel load paths following first yield.

2D Versus 3D

Regular structures in plan, for which torsional response is expected to be minimal, are often analyzed with separate 2D models for each principal axis of the structure. In engineering practice (in contrast to academic research), projects on which NRHA is considered worthwhile are seldom particularly regular, although there are of course many exceptions. 3D analysis does not introduce significant complexity in pre- and post-processing when using modern analysis packages, with the exception that the ground motion input is more complicated.

Level of Detail

A given structure can be analyzed in different levels of detail: from an equivalent single degree of freedom representation through to a detailed finite element model. Often, several different



Assessment of Existing Structures Using Response History Analysis, Fig. 2 Different levels of analysis detail for modeling an unreinforced masonry house

levels of analysis will be considered for one project: for example, simplified models at an initial stage and for subsequent model checking, a main structural model (e.g., frame members modeled as 2D beam elements) for overall performance assessment, and detailed local models to calibrate aspects of the main model (e.g., modeled force transfer within a RC joint region using 3D solid elements for concrete, 2D beam elements for reinforcing bars, and bond and friction between them defined with a mathematical model).

Three different levels of detail are shown in Fig. 2, for the NRHA of an unreinforced masonry (URM) house. Figure 2a shows an equivalent frame model, in which masonry pier flexural and shear responses are modeled with nonlinear spring or link elements. For this model, the mass is lumped at an equivalent roof height, and a rigid diaphragm connects the mass to the tops of each

of the wall piers. Figure 2b shows masonry walls modeled in 2D shell elements; masonry material models may be of a smeared crack type, in which the development of cracks in the walls is modeled explicitly, or a phenomenological hysteresis model in which backbone and cyclic hysteretic rules, based on experimental behavior, are used. Finally, Fig. 2c shows a model in which individual bricks are modeled with 3D solid elements. In this case, the interaction between bricks is described with a contact surface formulation, which allows for opening and closing of cracks between the bricks and the transfer of compression and traction across the interfaces. If calibrated correctly, this model can explicitly track the development of failure mechanisms in the URM wall up until the point of collapse.

Another example of a multilevel approach, as applied to the new design of an offshore platform, may be found in Gibson et al. (2012).

Boundary Conditions and Soil–Structure Interaction (SSI)

Although structural engineers are primarily interested in the seismic response of the building itself, in many cases, this cannot be accurately predicted without taking into account the interaction between the structure, its foundation, and the soil at the site. Soil behavior is generally nonlinear, at least for the range of strains of interest for earthquake engineering applications, but is often modeled with equivalent linear stiffness and damping properties. SSI literature (e.g., NEHRP Consultants Joint Venture 2012) refers to three main effects:

1. Inertial interaction – the forces and deformations on the structure–foundation–soil interface generated by inertial forces on the structure.
2. Kinematic interaction – the stiffening effect of the building’s foundation on the soil around it and its effect on the soil vibrations.
3. Foundation deformations – foundation elements, such as piles embedded in the soil, are subjected to deformations from the building and soil and must be designed to accommodate these movements.

SSI can be taken into account in NRHA in different levels of detail: from ignoring it (simplest), adding springs at the base of the structure to represent soil and foundation flexibility, through to a direct nonlinear finite element analysis of the soil and its interaction with structural and foundation elements. The recent guide to SSI by the NEHRP Consultants Joint Venture (2012) summarizes the state of practice and requirements of design standards and codes.

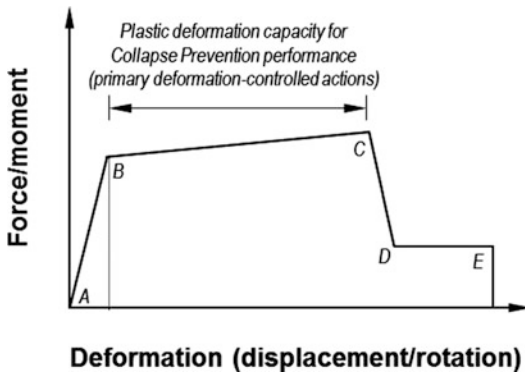
Recognizing that soil properties are typically subject to more uncertainty than structural material properties, design and assessment guidelines sometimes require a bounding approach, wherein separate analyses are carried out with lower and upper bound soil properties, and the analysis results are enveloped. Due to the nature of site response and SSI, however, there is no guarantee that this will envelope the worst-case structural response. Several guidelines suggest a bounding

approach for a serviceability level earthquake ground motion, but the use of the best-estimate soil properties for the evaluation of an ultimate level seismic response.

Material and Section Properties

Material and section properties depend on the type of modeled carried out and the hysteresis models available in the analysis software used. A common requirement of the international guidelines reviewed in section “[Published Codes, Standards and Guidelines](#)” is that member strengths should be taken based on “best-estimate” properties for deformation-controlled (ductile) responses and “lower-bound” properties for force-controlled (brittle) responses. Deformation-controlled actions include structural responses that exhibit a ductile nonlinear behavior and therefore may be permitted to deform into the plastic range without (necessarily) compromising the ability of the structure to support gravity loads and seismic forces. This includes flexure in steel and RC frame elements, axial deformation in steel braces, and rocking in masonry piers. Force-controlled actions include those responses that fail in a brittle manner when their capacity is exceeded and thus should usually be prevented to ensure structural integrity. This includes shear in frame elements and masonry piers.

Stiffnesses should be based on a best estimate, as otherwise the distribution of forces in the structure could be miscalculated. Structural modeled generally requires some degree of rationalization – e.g., plastic hinges are modeled with piecewise linear hysteretic backbones like those shown in Fig. 3 (from ASCE 41) – and the most appropriate simplifications will be application-dependent. In the backbone shown, AB is the nominal elastic range, BC the plastic range (shown with a small positive stiffness due to strain hardening and spread of plasticity), and DE the residual capacity, at which a small proportion of the peak capacity at point C is still carried, prior to final failure at point E. In ASCE 41, branch CD is modeled as very steep to account for sudden loss of strength, which may not be realistic for some component



Assessment of Existing Structures Using Response History Analysis, Fig. 3 Component force–deformation backbone behavior to ASCE 41 (Adapted from ASCE 2007)

responses (ATC 2010). When modeled RC elements with the backbone shown in Fig. 3, the initial branch (AB) is taken to represent the cracked response, with the justification that the uncracked response is only attained on the first cycle (if at all). The stiffness of branch AB is then taken as an approximation of the stiffness up to a nominal first yield point. For elements expected to be pushed far into the inelastic range (beyond point B), the stiffness of AB will usually not have a large effect on the overall structural response, but for elements that are not expected to yield or maybe not even crack, using fully cracked stiffness for AB may lead to an underestimate of forces on these elements.

Assessment standards and guidelines have requirements for material testing (nondestructive or destructive) to establish material properties for NRHA. Generally, at least some destructive testing is required for NRHA to be appropriate, unless material properties are well-documented on design documentation. This follows from the “principle of consistent crudeness” (Elms 1985), according to which excessive detail is not appropriate in one aspect of a design process if significant uncertainties remain in other aspects.

Hysteretic Response of Components

As well as member strengths and backbone response, NRHA requires a model to describe response of members under cyclic loads.

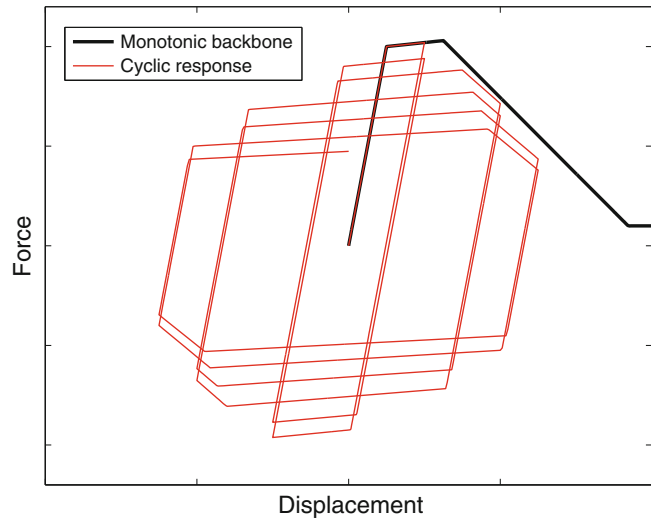
For example, if the backbone response in Fig. 3 is used to represent hysteretic response of a yielding member (such as a RC or steel beam in flexure), then unloading from the backbone may be at a stiffness close to the initial elastic stiffness of the member, and cyclic response may be close to elastoplastic. If, on the other hand, the backbone represents rocking of a structural wall, then unloading may retrace the backbone back through the origin, and the response is effectively nonlinear-elastic. A generic hysteretic response, generated using the Ibarra et al. (2005) hysteresis model, is shown in Fig. 4, under both monotonic and cyclic loading.

Most structural members’ response degrades with nonlinear cyclic loading, meaning that the strength reduces for large deformations (“monotonic degradation”) and that the stiffness and strength reduce over successive cycles out to a similar level of deformation (“cyclic degradation”). Monotonic degradation is illustrated in Fig. 4 by a decreasing monotonic backbone behavior when a certain displacement is exceeded. Cyclic degradation is shown by a reduction in the force carried under successive cycles to the same displacement level.

Degradation is especially important for those members that have not been designed with seismic-resistant detailing, but even well-detailed steel and RC plastic hinges degrade at higher levels of plastic deformation. As with other aspects of modeled, the relevance of modeled this degradation for seismic assessment depends on the type of elements in the structural system and the expected level of nonlinear demand on those elements. The plastic deformation limits tabulated in ASCE 41, for example, are artificially reduced to take into account the fact that most analysts will not include degradation in their analysis models. For more details of modeled aspects of structural degradation, see ATC (2010) and Deierlein et al. (2010). These references are more useful for modeled seismically detailed structural elements; the literature on modeled degradation in seismically deficient elements (poorly detailed RC or steel elements or URM) is not developed to the same level. The analyst should ensure that the structural model

Assessment of Existing Structures Using Response History Analysis,

Fig. 4 Component force–deformation hysteretic behavior, generated using the Ibarra et al. (2005) hysteresis model



adopted is able to accurately describe available experimental data for the level of strain expected in the analysis and the cyclic conditions expected.

Modeling Deficiencies in Structural Design

NRHA is most commonly adopted for the assessment of existing structures when the structural detailing is deficient with respect to modern code requirements, and therefore adequate seismic performance cannot be demonstrated by simpler analysis methods. Typical examples of detailing deficiencies are inadequate development lengths and splice lengths for reinforcing bars, transverse reinforcement in beam–column joints and plastic hinge regions, pre-Northridge steel beam–column connections, presence of unreinforced masonry infills in frame buildings, inadequate shear strength and/or failure to satisfy a capacity design hierarchy in beams and columns, and unreinforced masonry buildings in general. Although it may be desirable to address any of these deficiencies through retrofit when they are present, there are many cases when acceptable behavior can be demonstrated. Generally, detailing deficiencies such as splice lengths and beam–column joint details are addressed by limiting both the strength and plastic deformation capacity of the affected members. For example, when reinforcing bars are anchored with inadequate development length, ASCE 41 requires that

the capacity of the bar in question is reduced by a linear ratio of the provided to required development length and also significantly reduces plastic rotation limits by up to a factor of 10.

Model Checking and Post-Processing

The amount of post-processing required for a given NRHA model is in some ways inversely related to the amount of rationalization and simplification that has gone into the modeled effort. For example, if URM walls are modeled with a discrete element approach (e.g., Fig. 2c) – wherein all bricks and their interaction are explicitly modeled – then crack widths can be evaluated explicitly, and the “no collapse” performance requirement can be checked by observation of the state of the model at the end of the analysis. In frame analyses – with 2D beam elements used for beams and columns – if monotonic and cyclic degradation are properly accounted for, collapse response can also be observed directly. More commonly, adequate performance of deformation-controlled elements is evaluated on the basis of observed peak plastic deformations (e.g., peak plastic hinge rotations in beams and columns, peak plastic axial deformation in braces, peak drifts in RC or URM walls) compared with allowable limits from published guidelines (e.g., those considered in section “Published Codes, Standards and Guidelines”)

or experiments from the literature carried out specifically for the project. Performance of force-controlled elements is evaluated on the basis of observed forces from the analysis compared with capacities evaluated according to guidelines or engineering first principles.

As noted in section “**Seismic Input**”, analyses are typically repeated for multiple earthquake ground motions, and potentially for multiple application orientations of each ground motion, to account for the fact that record-to-record uncertainty is relatively large compared to other sources of uncertainty. Appropriate processing of structural analysis results will depend on the considerations discussed in that section. When the peak response from multiple analyses is considered (say if fewer than seven ground motions are considered as input), this does not introduce significant difficulty – every analysis carried out must demonstrate acceptable response. Consideration of mean response, however, introduces several difficulties. For purely numerical quantities (e.g., plastic hinge rotation), typically codes are interpreted to refer to arithmetic mean, whereas the geometric mean would be more consistent with the common assumption of log-normally distributed structural response. For qualitative responses, such as “collapse” versus “no collapse,” or “uplift at foundation” versus “no uplift,” means cannot be evaluated. In these cases, the median is a more appropriate measure of central tendency of analysis results. Many engineers consider more than just statistics when interpreting analysis results – even if four out of seven analyses demonstrate no collapse, most engineers would be interested in assessing how collapse occurs in the other three and, if economically possible, prevent these modes of collapse from occurring. NEHRP Consultants Joint Venture (2011) provides further discussion of how analysis and performance objectives affect ground motion development and analyses carried out.

Residual deformations can be more important than peak transient deformations for evaluating performance criteria relating to post-earthquake operability or possibility of occupancy. For example, peak deformations during the

earthquake may indicate some cracking that closes up following the earthquake and does not require structural repair. Residual deformations are sensitive to backbone characteristics (such as post-yield stiffness, line BC in Fig. 3), as well as cyclic hysteretic characteristics (such as unloading stiffness and pinching). Furthermore, estimating residual deformations in NRHA can require a significantly longer analysis, allowing time after the earthquake ground motion for the structure to settle. Residual deformations are not explicitly required to be checked in most published guidelines, although approximate limits are given in ASCE 41. FEMA P-58 discussed the estimation of residual interstory drifts in an appendix. On the basis that estimation of residual response is very sensitive to modeled and ground motion characteristics, it suggests using predefined relationships between residual and peak drifts and to not attempt to evaluate them in the main NRHA model.

Nonstructural components are seldom included in a NRHA model, and therefore their assessment is carried out in the post-processing phase. Some so-called nonstructural components, such as URM infill panels in frame buildings, should be included in the structural model, when their inclusion is expected to affect structural response. Nonstructural components are typically classified as either displacement/drift-sensitive (such as partitions or façade panels) or acceleration-sensitive (such as electrical equipment), and the performance of some equipment may be rated on the basis of a limiting peak drift or acceleration level. For more important equipment, response can be evaluated with a response history analysis using floor acceleration histories recorded in the main NRHA model or a response spectrum analysis using secondary response spectra, also from the main model. Finally, statistical performance of nonstructural components is often expressed in the form of a fragility function, which gives the probability of reaching different damage levels for a given level of acceleration or drift. Fragility functions for common types of equipment and other nonstructural components (at least for US practice) are given in FEMA P-58.

Published Codes, Standards, and Guidelines

This section provides a review of available international guidelines on the seismic assessment and retrofit design of existing structures. Regulatory requirements for demonstrating seismic performance vary, but in real projects, it is helpful to demonstrate compliance with documents such as the following to establish credibility – although this should establish a minimum duty of care and should not restrict the engineer from applying state-of-the-art methods and models when they contradict the published guidelines.

ASCE/SEI 41-06: Seismic Rehabilitation of Existing Buildings

ASCE/SEI 41-06 (ASCE 2007) – herein ASCE 41 – is perhaps the most complete published guideline available for assessment and retrofit design for existing buildings, including extensive provisions on NRHA. Acceptable nonlinear seismic performance is evaluated on a component-by-component basis (e.g., plastic rotations in beam and column plastic hinges), and all structural elements must satisfy these checks for compliance. Different limits are quantified for different performance objectives: Immediate Occupancy, Life Safety, and Collapse Prevention. Overall measures of structural performance such as interstory drifts are not explicitly checked, except to evaluate possibility of impact with adjacent buildings and in checking elements that span between stories such as reinforced concrete (RC) shear walls and nonstructural components.

ASCE 41 distinguishes between deformation-controlled and force-controlled actions, as discussed in section “[Material and Section Properties](#)”. Deformation-controlled actions are modeled with their expected (mean) strength and nonlinear response quantities, while force-controlled actions are restricted to their lower-bound (mean minus one standard deviation) strength. The nonlinear response of force-controlled actions is not usually modeled, since exceeding their linear capacity would result in a sudden shedding of their load (although see

note below about the updated version ASCE 41/SEI 41-13). Some deformation-controlled actions – such as flexure in columns of moment frames – may be inhibited for new structures using capacity design procedures, but for existing structures, it would be excessively conservative to condemn a noncompliant structure if adequate seismic performance can be demonstrated.

ASCE 41 tabulates nonlinear deformation acceptance criteria for all types of structural components; in fact, these criteria are routinely applied also for the performance-based seismic design of new structures and are referred to in guidelines for analysis and design of high-rise buildings (e.g., ATC 2010). Allowable plastic deformations are tabulated in terms of relevant characteristics such as performance objective, member slenderness (for steel braces), presence of high shear and axial load and compliance of transverse reinforcement with code requirements (for concrete beams and columns), and other non-compliances with respect to modern codes. A disadvantage of this tabular format is that sometimes small changes in assumptions (e.g., a change from “noncompliant” transverse reinforcement to “compliant” transverse reinforcement) can make a large step change to the allowable deformation, whereas the real effect modeled is likely to be continuous.

Nonlinear behavior of deformation-controlled actions is modeled following the generalized force–displacement or moment–rotation backbone illustrated in Fig. 3. Parameters (deformations and forces) to describe these curves are tabulated for every component in ASCE 41. Note that ASCE 41 does not provide guidance on modeled the cyclic response of components following the backbone in Fig. 3; therefore, modeled of hysteretic phenomena such as stiffness and strength degradation and pinching is not foreseen (see section “[Hysteretic Response of Components](#)”). ATC-72-1 (ATC 2010) notes that the force–displacement backbone represents a “cyclic envelope curve” (i.e., represents cyclically degraded response) rather than the force envelope that one would get from a monotonic test on the component. Furthermore, restricting deformations of primary structural components

to within point C on Fig. 3 ensures that the region in which cyclic degradation dominates is not reached.

The next edition of ASCE 41 – ASCE/SEI 41-13 (ASCE 2014) – has just been published and expands the remit of ASCE 41 to include the multitiered seismic assessment approach previously found in ASCE/SEI 31-03 (ASCE 2003). The nonlinear response history analysis procedures have been expanded somewhat, reflecting significantly increased application of this method and the availability of research efforts such as Deierlein et al. 2010; NEHRP Consultants Joint Venture 2010, since the publication of ASCE/SEI 41-06. Nonlinear behavior of force-controlled actions is now permitted to allow explicit modeled of post-failure redistribution following brittle failure. Finally, and importantly for this entry, the unreinforced masonry (URM) wall provisions have been updated; major changes are that bed-joint sliding is now considered a deformation-controlled action, while the pier rocking mechanism is now limited to lower axial load ratios and for piers at least 6" in thickness.

Eurocode 8 Design of Structures for Earthquake Resistance Part 3: Assessment and Retrofitting of Buildings

Eurocode 8 Part 3 (CEN 2005) – herein EC8-3 – deals with both the seismic assessment and retrofitting of existing buildings. It is interesting to note that other Eurocodes do not cover non-seismic assessment of existing structures, but it was recognized by the code developers that many older structures in Europe did not take into account seismic resistance in their design (whereas presumably gravity and wind have been taken into account – at least by common design practice if not explicit calculation). In common with the other standards considered here, EC8-3 leaves the development and details of a national program for seismic risk mitigation to the member countries, including such critical aspects as what triggers the need for assessment of an existing building, timescales over which assessment and retrofitting must take place (if required), and the performance objectives

(limit states and associated return periods of seismic action) that must be satisfied, either before or after strengthening. Performance objectives considered in the standard are defined as Near Collapse, Significant Damage and Damage Limitation – inconsistently with the performance objectives in EC8-1 for new structures – but the actual objective(s) to be checked is left to the National Annex.

Similarly to ASCE 41, EC8-3 distinguishes between ductile and brittle actions. For nonlinear analyses, capacities are evaluated on the basis of deformation for ductile actions and strength for brittle ones. Column flexure is considered a ductile action, although interestingly “strong column–weak beam” behavior is enforced for steel moment frames only (not reinforced concrete).

Acceptance criteria for nonlinear actions are contained in informative Annexes A, B, and C for RC, steel, and masonry, respectively. No guidance is given for the assessment of timber structures. Generally, for elements classified as primary components of the seismic force resisting system, allowable deformations are explicitly noted as mean minus one standard deviation estimates, and for secondary components, they are the mean values.

For RC beams and columns, allowable total or plastic rotations are expressed in terms of regression relationships on experimental data, including such aspects as section depth, shear and axial force ratios, longitudinal and transverse reinforcement ratio, and material properties. Modifications are also provided for lack of compliance with modern requirements, such as lap splices in plastic hinge zones and the use of smooth reinforcing bars. Alternative expressions are also provided for curvatures rather than rotations. Annex A is also notable for a ductility-dependent shear strength calculation and specific rules for concrete, steel, and fiber-reinforced polymer (FRP) jacketing of seismically deficient elements, including estimates of the strength and deformation capacities provided by these interventions that could be used in NRHA.

Acceptability criteria for steel elements, contained in Annex B, are more comparable to the equivalents in ASCE 41. They are expressed

as multiples of the yield displacement (for braces) or yield rotation (for flexural elements), with multipliers that are in most cases identical to those in ASCE 41, assuming Immediate Occupancy, Life Safety, and Collapse Prevention performance objectives (from ASCE 41) may be mapped directly on to Damage Limitation, Significant Damage, and Near-Collapse limit states (from EC 8-3). Rotation capacities for the three limit states are also provided for retrofitted components, using interventions such as haunched connections and reduced beam section connections.

Masonry acceptance criteria (in Annex C) are quantified in terms of wall drift – in common with ASCE 41 for flexure-controlled walls, but in contrast to ASCE 41 for shear-controlled walls for which strength criteria apply. For flexure-controlled walls, drift limits are given in terms of the slenderness ratio of width to the height of the point of contraflexure, which gives limits similar to those given in ASCE 41 for fix–fix boundary conditions of wall piers but doubles the limits for cantilever walls (mapping the limit states as discussed for steel).

Stress–strain relationships for different materials are contained in each of the other material-specific Eurocodes and may be used to develop force–displacement or moment–rotation relationships for use in nonlinear analysis. However, these typically do not consider the range of strains associated with near-collapse response of structures, and therefore in-cycle degradation of strength may not be considered (see section “[Hysteretic Response of Components](#)”). Furthermore, no guidance is provided for cyclic hysteresis models for a given structural element, except a generic statement that it “should realistically reflect the energy dissipation in the element over the range of displacement amplitudes expected in the seismic design situation.”

New Zealand Society for Earthquake Engineering (NZSEE): Assessment and Improvement of the Structural Performance of Buildings in Earthquakes

The Building Act 2004 required territorial authorities (TAs) in New Zealand to adopt

a policy on “earthquake-prone buildings” (EPBs) within 18 months of the passing of the Act. The corresponding Building Regulations defines an EPB as one that will have its ultimate capacity exceeded in earthquake ground motion one-third as strong as that required for new building design. The New Zealand Society for Earthquake Engineering developed its recommendations (NZSEE 2006) (herein “the NZSEE Recommendations”) to assist TAs in developing their policies and engineers in carrying out assessments and design of seismic rehabilitation.

NRHA is allowed by the NZSEE Recommendations, although they note that NRHA should not be the sole adopted assessment procedure and should be supplemented by the results of a simplified approach. They cite such uncertainties as flexure–shear–axial force interaction, effect of axial force on column stiffness, degrading strength characteristics, and modeled of beam–column joints. Presumably, the restriction on the use of NRHA could be relaxed somewhat if more detailed analysis models are used that reduce these uncertainties. The focus of the NZSEE Recommendations is on linear analysis supplemented by an analysis of expected yielding mechanisms, and therefore the recommendations for NRHA are reasonably limited.

Nonlinear deformation limits for concrete members in the NZSEE Recommendations are evaluated in terms of yield and ultimate curvatures and an assumed nominal plastic hinge length. The yield curvature is given as a function of the yield strain in the reinforcement and the dimension of the section, and the ultimate curvature can be evaluated from a concrete strain limit, taking into account the effect of confinement and the neutral axis depth of the section at ultimate curvature. The plastic rotation capacity of steel sections is tabulated (in the New Zealand standard for design of steel structures) as a function of the section compactness and the axial load on the member. Allowable plastic hinge rotations for steel sections range from 0.5 % to 4 % for seismic load cases (higher values are given for links of eccentrically braced frames).

The NZSEE Recommendations contain useful discussion on the modeled of URM structures; they note the importance of including the distribution of mass in a masonry building and discuss the use of finite element modeled with 2D or 3D elements. However, the guidance is more targeted towards linear analysis; they recommend using a ductility factor of 1.0 and an equivalent viscous damping value of 15 % to account for energy dissipation due to cracking, sliding, and rocking, which do not seem to be appropriate for nonlinear analysis. Nevertheless, deformation limits are given as 1 % for elements failing in a rocking mode and 0.5 % for all other elements. Unique among the documents considered here, the Recommendations also give a rotation limit for spandrels of 0.5 % (based on the clear span).

FEMA P-58: Seismic Performance Assessment of Buildings

The ATC-58 and ATC-58-1 projects developed the so-called next-generation performance-based seismic design guidelines following on from “first-generation” projects such as ASCE 41. The outcome of these projects was FEMA P-58, published in 2012 (ATC 2012). This project takes one of the main original objectives of performance-based design – communicability of seismic performance to non-engineers – to the next level by providing guidance on converting results from analytic models into estimates of casualties, repair costs, and repair time. To achieve this in a meaningful way, the framework explicitly incorporates variability and uncertainty in both structural and nonstructural response. The probabilistic loss calculations are implemented in a computer program, *PACT* (Performance Assessment Calculation Tool).

FEMA P-58 considers the three basic assessment types discussed in section “[Performance Objectives and Assessment Methodology](#).” For each of these assessment types, a similar workflow is followed. Structural performance is assessed with NRHA (or other methods) and is quantified both in terms of collapse fragility (probability distribution of the ground motion level required to cause collapse) and probability

distributions of other demand parameters associated with damage to structural and nonstructural components (e.g., peak floor acceleration and peak storey drift ratio). A building performance model is developed, which takes into account all the nonstructural components in the building and relates them to specific demand parameters (e.g., partitions in a floor are related to the storey drift in that floor). Fragility functions and consequence functions are provided for typical nonstructural components present in US buildings and incorporated in the software, *PACT*. A Monte Carlo analysis is carried out to sample each of these fragility functions to assess the expected damage and losses under the design earthquake ground motion.

FEMA P-58 evaluates economic and human losses as continuous variables, rather than checking for compliance with discrete limit states (such as “Life Safety” and “Collapse Prevention” used in ASCE 41). It is left to the building owner, engineer, or future code developer to define acceptable performance for a given project. Economic losses can be compared with the costs of seismic retrofit or with other insurance or investment costs.

Downtime calculations in FEMA P-58 are limited to estimates of repair time resulting from earthquake damage, which does not account for delays for required inspections, mobilization of engineering and contractors resources, financing, permitting, and utility disruption. It also only focuses on the time to achieve full recovery of the facility, but does not consider intermediate goals such as the time to reoccupy or return to functionality, which are of more direct concern to building owners. The downtime calculation was therefore improved as part of the development of the REDiTM Rating System (Resilience-based Earthquake Design Initiative) guidelines (Almufti and Willford 2013), taking into account input from contractors, cost estimators, bankers, building department officials, and researchers. The REDiTM Rating System uses this extended FEMA P-58 loss assessment approach to demonstrate that the REDiTM resilience objectives are satisfied – primarily for new buildings, but also applicable for existing structures.

Summary

This entry has summarized the assessment of existing structures using nonlinear response history analysis. As noted in the introduction, NRHA is no longer the domain of specialist engineering consultancies, and many commonly used software packages are available to carry it out on projects.

It has been emphasized throughout this entry that the most appropriate modeled techniques and approximations to use are very application-dependent. Published guidance aimed at accurately estimating the collapse risk of well-detailed structures in highly seismic regions is not necessarily appropriate for assessing brittle URM buildings in low to moderate seismic regions. Furthermore, every client will have different requirements, and it is not always appropriate to use the most advanced analysis methods in cases where acceptable performance can be demonstrated by simpler methods or where the level of information available is limited (according to the principle of consistent crudeness). Nevertheless, this entry has attempted to summarize the different considerations that the analyst faces when setting up and interpreting his or her NRHA models for existing structures.

Several other entries in this volume consider different aspects of NRHA, and they should be consulted for more detailed guidance on each of the modeled decisions discussed in this entry.

Cross-References

- [Analytic Fragility and Limit States \[P\(EDPI IM\)\]: Nonlinear Dynamic Procedures](#)
- [Assessment of Existing Structures Using Inelastic Static Analysis](#)
- [Damage to Buildings: Modeling](#)
- [Earthquake Response Spectra and Design Spectra](#)
- [Engineering Characterization of Earthquake Ground Motions](#)
- [Nonlinear Dynamic Seismic Analysis](#)
- [Nonlinear Finite Element Analysis](#)

- [Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings](#)
- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- [Response-Spectrum-Compatible Ground Motion Processes](#)
- [Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions](#)
- [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)
- [Seismic Analysis of Steel Buildings: Numerical Modeling](#)
- [Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling](#)
- [Seismic Collapse Assessment](#)
- [Seismic Response Prediction of Degrading Structures](#)
- [Time History Seismic Analysis](#)

References

- Almufiti I, Willford MR (2013) Resilience-based earthquake design (REDi) rating system, version 1.0. Arup. http://www.arup.com/Publications/REDi_Rating_System.aspx. Accessed Nov 2013
- ASCE (2003) ASCE standard ASCE/SEI 31-03: seismic evaluation of existing buildings. American Society of Civil Engineers, Reston
- ASCE (2007) ASCE standard ASCE/SEI 41-06: seismic rehabilitation of existing buildings. American Society of Civil Engineers, Reston
- ASCE (2014) ASCE standard ASCE/SEI 41-13: seismic evaluation and retrofit of existing buildings. American Society of Civil Engineers, Reston
- ATC (2010) PEER/ATC-72-1: modeling and acceptance criteria for seismic design and analysis of tall buildings. Applied Technology Council, Redwood City
- ATC (2012) FEMA P-58: seismic performance assessment of buildings, vol 1–3. Applied Technology Council, Redwood City
- CEN (2005) Eurocode 8: design of structures for earthquake resistance – part 3: assessment and retrofitting of buildings. European Committee for Standardization (Comité Européen de Normalisation), Brussels
- Deierlein GG, Reinhorn AM, Willford MR (2010) Nonlinear structural analysis for seismic design. NEHRP seismic design technical brief no. 4, NIST GCR 10-917-5. National Institute of Standards and Technology, Gaithersburg
- Elms DG (1985) Principle of consistent crudeness. Workshop on civil engineering application of fuzzy sets. Purdue University, West Lafayette

- Gibson R, Wilcock TR, Grant DN, Gration DA (2012) Performance based seismic design of an offshore platform in the Caspian Sea. *Struct Eng* 90:14–22
- Grant DN (2011) Response spectral matching of two horizontal ground motion components. *Journal of Structural Engineering* 137: 289–297
- Grant DN, Padilla D, Greening PD (2011) Orientation dependence of earthquake ground motion and structural response. In: Dolšek M (ed) *Protection of built environment against earthquakes*. Springer, Dordrecht, 352 pp
- Ibarra LF, Medina RA, Krawinkler H (2005) Hysteretic models that incorporate strength and stiffness deterioration. *Earthq Eng Struct Dyn* 34:1489–1511
- NEHRP Consultants Joint Venture (2010) Applicability of nonlinear multiple-degree-of-freedom modeling for design. NIST GCR 10-917-9. National Institute of Standards and Technology, Gaithersburg
- NEHRP Consultants Joint Venture (2011) Selecting and scaling earthquake ground motions for performing response-history analyses. NIST GCR 11-917-15. National Institute of Standards and Technology, Gaithersburg
- NEHRP Consultants Joint Venture (2012) Soil-structure interaction for building structures. NIST GCR 12-917-21. National Institute of Standards and Technology, Gaithersburg
- NZSEE (2006) Assessment and improvement of the structural performance of buildings in earthquakes. New Zealand Society for Earthquake Engineering, Wellington

B

Base-Isolated Systems, Reliability-Based Characterization of

Hector A. Jensen and D. S. Kusanovic
Department of Civil Engineering, Santa Maria
University, Casilla, Valparaiso, Chile

Synonyms

Advanced simulation techniques; First excursion probability; Isolation systems; Rubber bearings; Stochastic excitation

Introduction

The recent improvements in isolation system products have led to the design and construction of an increasing number of seismically isolated buildings worldwide (Ceccoli et al. 1999; Chopra 2001; Kelly 1986; Zou et al. 2010). Similarly, seismic isolation has been extensively used for seismic retrofitting of existing buildings (De Luca et al. 2001; Mokha et al. 1996). In addition, base isolation concepts are utilized for the protection from shock and vibration of sensitive components of critical facilities such as hospitals, nuclear reactors, and industrial and data center facilities. One of the difficulties in the analysis and design of base-isolated systems has been the explicit consideration of the nonlinear behavior of the isolators. Another challenge has

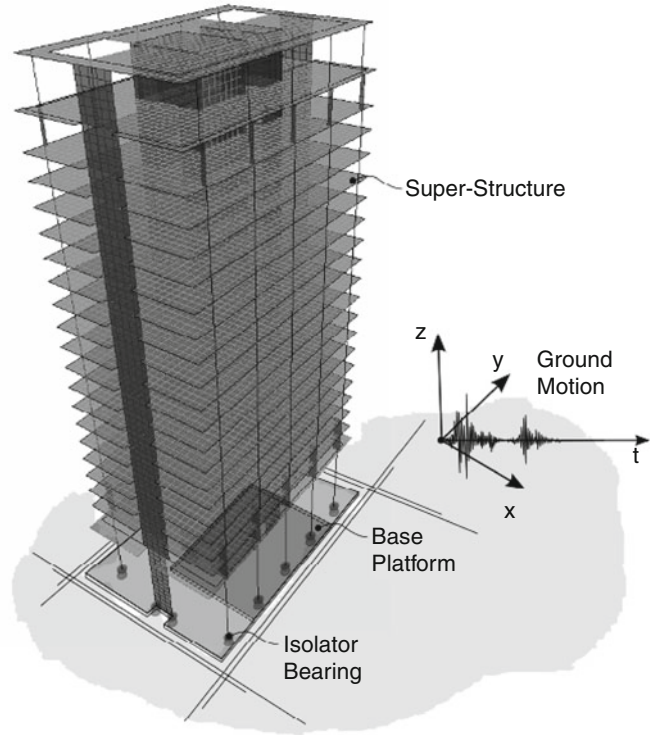
been the efficient prediction of the dynamic response under future ground motions considering their potential variability as well as competing objectives related to the protection of the superstructure and the minimization of base displacement. The objective of this work is to characterize the performance of base-isolated systems from a reliability point of view. In particular, the case of large-scale building models is considered here. Isolation systems composed by rubber bearings are used in the present formulation. The nonlinear behavior of these devices is characterized by a biaxial hysteretic model which is calibrated with experimental data (Yamamoto et al. 2009). First excursion probabilities are used as measures of the system reliability. In this setting, reliability is quantified as the probability that the response quantities of interest (base displacement and superstructure response) will not exceed acceptable performance bounds within a particular reference period (Lutes and Sarkani 2004). Such probabilities are estimated by a stochastic simulation method (Au and Beck 2001). The variability of future excitations is addressed by adopting a probabilistic approach which belongs to the class of point-source models (Boore 2003).

The organization of this entry is as follows: Section “[Structural Model](#)” introduces the structural model for the superstructure and base platform. The characterization of the isolation system is presented in section “[Isolation Model](#).” Section “[Excitation Model](#)” deals with the

Base-Isolated Systems, Reliability-Based

Characterization of,

Fig. 1 Schematic representation of a base-isolated finite element building model



stochastic model for the excitation. The reliability assessment and the structural response of base-isolated buildings are discussed in sections “[Reliability Measures](#)” and “[System Response](#),” respectively. In section “[Application Problem](#),” the proposed characterization is illustrated by means of a based-isolated finite element building model. The entry closes with some final remarks.

Structural Model

Finite element building models with a relatively large number of degrees of freedom are considered for modeling the superstructure, i.e., the structure above the isolation system. For illustration purposes a schematic representation of a base-isolated finite element building model under ground motion is shown in Fig. 1. In general, base-isolated buildings are designed such that the superstructure remains elastic. Hence, the superstructure is modeled as a

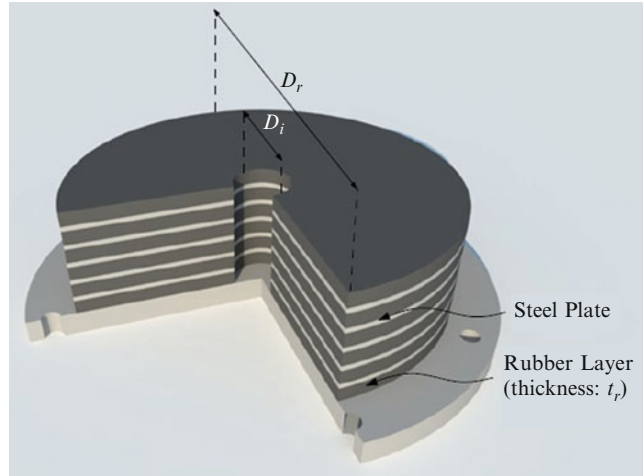
three-dimensional linear elastic system, while the base is assumed to be rigid in plane and it is modeled using 3 degrees of freedom. Let $x_s(t)$ be the n th dimensional vector of relative displacements of the superstructure with respect to the base and M_s , C_s , K_s be the corresponding mass, damping, and stiffness matrices. Also, let $x_b(t)$ be the vector of base displacements with three components and G_s be the matrix of earthquake influence coefficients of dimension $n \times 3$, that is, the matrix that couples the excitation components of the vector $\ddot{x}_g(t)$ to the degrees of freedom of the superstructure. The equation of motion of the superstructure is expressed in the form

$$M_s \ddot{x}_s(t) + C_s \dot{x}_s(t) + K_s x_s(t) = -M_s G_s [\ddot{x}_b(t) + \ddot{x}_g(t)] \quad (1)$$

where $\ddot{x}_b(t)$ is the vector of base accelerations relative to the ground. To complete the formulation of the combined model, the equation of motion for the base platform is written as

Base-Isolated Systems, Reliability-Based

Characterization of,
Fig. 2 Schematic
representation of a rubber
bearing



B

$$\begin{aligned} & (G_s^T M_s G_s + M_b) (\ddot{x}_b(t) + \ddot{x}_g(t)) + G_s^T M_s \ddot{x}_s(t) \\ & + C_b \dot{x}_b(t) + K_b x_b(t) + f_{is}(t) = 0 \end{aligned} \quad (2)$$

where M_b is the mass matrix of the rigid base, C_b is the resultant damping matrix of viscous isolation components, K_b is the resultant stiffness

matrix of linear elastic isolation components, and $f_{is}(t)$ is the vector containing the nonlinear isolation elements forces.

Rewriting the previous equations, the combined equation of motion of the base-isolated structural system can be formulated in the form

$$\begin{aligned} & \begin{bmatrix} M_s & M_s G_s \\ G_s^T M_s & M_b + G_s^T M_s G_s \end{bmatrix} \begin{Bmatrix} \ddot{x}_s(t) \\ \ddot{x}_b(t) \end{Bmatrix} + \begin{bmatrix} C_s & O \\ O & C_b \end{bmatrix} \begin{Bmatrix} \dot{x}_s(t) \\ \dot{x}_b(t) \end{Bmatrix} + \\ & \begin{bmatrix} K_s & O \\ O & K_b \end{bmatrix} \begin{Bmatrix} x_s(t) \\ x_b(t) \end{Bmatrix} = - \begin{bmatrix} M_s G_s \\ M_b + G_s^T M_s G_s \end{bmatrix} \ddot{x}_g(t) - \begin{Bmatrix} 0 \\ f_{is}(t) \end{Bmatrix} \end{aligned} \quad (3)$$

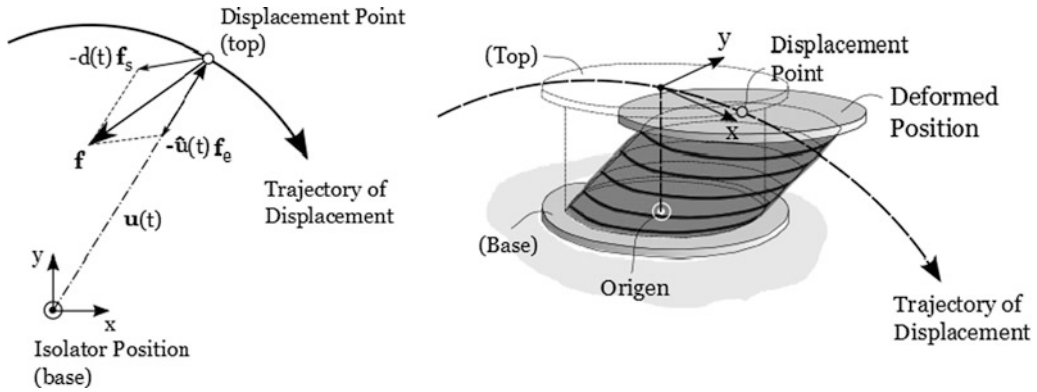
The above formulation can be extended to other cases as well, for example, the consideration of nonlinear models for the superstructure.

Isolation Model

Rubber bearings are considered for modeling the isolation system. Such devices have been used over many years in a number of seismically isolated buildings worldwide (Kelly 1986; Su et al. 1990; Makris and Chang 1998). They require minimal initial cost and maintenance compared to other passive, semi-active, and active energy absorption devices. A rubber bearing consists of layers of rubber and steel, with the

rubber being vulcanized to the steel plates. Rubber bearing systems are able, in principle, to support the superstructure vertically, to provide the horizontal flexibility together with the restoring force, and to supply the required hysteretic damping. Figure 2 shows a schematic representation of a rubber bearing, where D_r represents the external diameter of the isolator, D_i indicates the internal diameter, and $H_r = t_r n_r$ is the total height of rubber in the device, where t_r is the layer thickness and n_r is the number of rubber layers.

An analytical model that simulates measured restoring forces under bidirectional loadings is used in the present formulation. The model is based on a series of experimental tests conducted for real-size rubber bearings (Yamamoto



Base-Isolated Systems, Reliability-Based Characterization of, Fig. 3 Schematic representation of the decomposition of the restoring force

et al. 2009; Minewaki et al. 2009). The loading tests of seven full-scale isolators were carried out using the Caltrans Seismic Response Modification Device Test Facility at the University of California, San Diego. The specimens used in the tests were made with high damping rubber compounds. In particular, horizontal bidirectional loading tests for isolators with a diameter of 0.7 m and 1.3 m were conducted. On the basis of the test results, the model assumes that the restoring force on the rubber bearing is composed of a force directed to the origin of the isolator and another force approximately opposite to the direction of the movement of the isolator. Such decomposition of the restoring force is shown schematically in Fig. 3.

According to the model, the direction of the movement of the isolator $d(t)$ is defined in terms of the components in the x and y direction of the base displacement vector $x_b(t)$ by means of the nonlinear differential equation

$$\dot{d}(t) = \frac{1}{\alpha} \|\dot{u}(t)\| \left[\hat{u}(t) - \|\dot{d}(t)\|^\beta \hat{d}(t) \right], \quad (4)$$

$$u(0) = 0, \quad d(0) = 0$$

where the components of the vector $u(t)$ correspond to the components in the x and y direction

of the base displacement vector $x_b(t)$; $\dot{u}(t)$ is the velocity vector; $\hat{u}(t)$ and $\hat{d}(t)$ are the unit directional vectors of $\dot{u}(t)$ and $\dot{d}(t)$, respectively; and $\|\cdot\|$ indicates the Euclidean norm. The parameters $\alpha = 0.2H_r$ and $\beta = 0.7$ are positive constants that relate to the yield displacement and smoothness of yielding, respectively. Once the vector $d(t)$ has been derived, the restoring force $f(t)$ on the isolator (in the x and y direction) is expressed in terms of the unit directional vector $\hat{u}(t)$ and the vector $d(t)$ as $f(t) = -\hat{u}(t)f_e(t) - d(t)f_s(t)$, where $f_e(t)$ is the nonlinear elastic component and $f_s(t)$ is the elastoplastic component. The previous decomposition of the restoring force to elastic and dissipative components simplifies the restoring force characteristics and results in a simple and accurate model. These two forces have different appearance mechanisms and different influencing factors. For example, the dependency of these forces on factors such as temperature, vertical pressure, and maximum experienced shear strain would differ. Therefore, the previous decomposition has the potential for more general modeling of other dependencies. Based on experimental results, the components of the restoring force on the rubber bearing can be approximated as

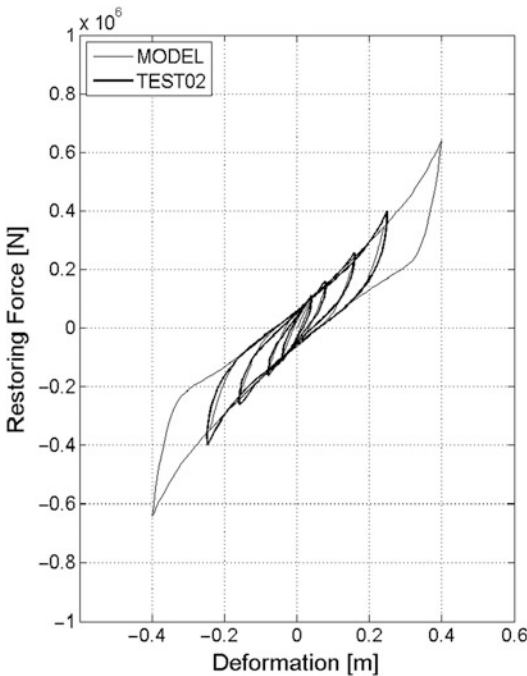
$$f_e(t) = \begin{cases} 0.35\gamma(t)A & \text{if } 0 \leq \gamma(t) \leq 1.8 \\ \left(0.35\gamma(t) + 0.2(\gamma(t) - 1.8)^2\right)A & \text{if } \gamma(t) \geq 1.8 \end{cases} \quad (5)$$

and

$$f_s(t) = \left(0.125 + 0.015\gamma(t) + 0.012\gamma(t)^3\right)A \quad (6)$$

where A is the cross-sectional area of the rubber and $\gamma(t) = \|u(t)\|/H_r$ is the average shear strain. Validation calculations have shown that the analytical model is able to accurately simulate the test results for both bidirectional and unidirectional loading. The reader is referred to Yamamoto et al. (2009) and Minewaki et al. (2009) for a detailed description of the analytical model presented in this section.

Using the previous calibration of the analytical model by means of bidirectional loading tests, the restoring forces and those calculated by the model under unidirectional loading are compared in Fig. 4 for a medium-size rubber bearing. The test results were conducted for a maximum average shear strain of 150 %.



Base-Isolated Systems, Reliability-Based Characterization of, Fig. 4 Comparison of analytical and experimental hysteresis loops for a medium-size rubber bearing. $D_r = 0.8$ m, $H_r = 0.16$ m, $D_i = 0.15$ m

It is observed that the analytical model simulates the test results very well. Additional validation calculations have shown that the analytical model is also able to accurately simulate the test results for bidirectional loadings.

Excitation Model

The uncertain earthquake excitation is modeled as a non-stationary stochastic process. In particular, a point source model characterized by the moment magnitude M and epicentral distance r is considered in this work (Boore 2003; Atkinson and Silva 2000). The time history for a specific event magnitude M and epicentral distance r is obtained by first generating a discrete white noise sequence $z^T = \langle \sqrt{2\pi/\Delta t} z_j, j = 1, \dots, n_T \rangle$, where $z_j, j = 1, \dots, n_T$, are independent, identically distributed standard Gaussian random variables, Δt is the sampling interval, and n_T is the number of time instants equal to the duration of the excitation T divided by the sampling interval. The white noise sequence is then modulated by an envelope function $h(t, M, r)$ at the discrete time instants. Discrete Fourier transform is applied to the modulated white noise sequence. The resulting spectrum is normalized by the square root of the mean square of the amplitude spectrum. Finally, the normalized spectrum is multiplied by a ground motion spectrum $S(f, M, r)$ after which discrete inverse Fourier transform is applied to transform the sequence back to the time domain to yield the desired ground acceleration time history. The envelope function is represented by

$$h(t, M, r) = a_1 \left(\frac{t}{t_n}\right)^{a_2} e^{-a_3 \left(\frac{t}{t_n}\right)} \quad (7)$$

where

$$a_1 = \left(\frac{e}{0.2}\right)^{a_2}, \quad a_2 = \frac{-0.2\ln(0.05)}{1 + 0.2(\ln(0.2) - 1)}, \quad a_3 = \frac{a_2}{0.2} \quad (8)$$

On the other hand, the total spectrum of the motion at a site $S(f, M, r)$ is expressed as the

product of the contribution from the earthquake source $E(f, M)$, path $P(f, r)$, site $G(f)$, and type of motion $I(f)$, i.e., $S(f, M, r) = E(f, M) P(f, r) G(f) I(f)$.

The source component is given by $E(f, M) = C M_0(M) S_a(f, M)$, where C is a constant, $M_0(M) = 10^{1.5M + 10.7}$ is the seismic moment, and the factor S_a is the displacement source spectrum given by

$$S_a(f, M) = \frac{1 - \varepsilon}{1 + \left(\frac{f}{f_a}\right)^2} + \frac{\varepsilon}{1 + \left(\frac{f}{f_b}\right)^2} \quad (9)$$

where the corner frequencies f_a and f_b , and the weighting parameter ε are defined, respectively, as $\log(f_a) = 2.181 - 0.496 M$, $\log(f_b) = 2.41 - 0.408 M$, and $\log(\varepsilon) = 0.605 - 0.255 M$. The constant C is given by $C = UR_\phi VF / 4\pi\rho_s\beta_s^3 R_0$, where U is a unit-dependent factor (10^{-20}), R_ϕ is the radiation pattern, V represents the partition of total shear-wave energy into horizontal components, F is the effect of the free surface amplification, ρ_s and β_s are the density and shear-wave velocity in the vicinity of the source, and R_0 is a reference distance.

The path effect $P(f, r)$ which is another component of the process that affects the spectrum of motion at a particular site is represented by functions that account for geometrical spreading and attenuation $P(f, r) = Z(R(r)) e^{-\pi f R(r)/Q(f)\beta_s}$, where $R(r)$ is the radial distance from the hypocenter to the site given by $R(r) = \sqrt{r^2 + h^2}$, where $\log(h) = 0.15 M - 0.05$. The attenuation quantity $Q(f)$ is taken as $Q(f) = 180f^{0.45}$, and the geometrical spreading function is selected as $Z(R(r)) = 1/R(r)$ if $R(r) < 70.0$ km and $Z(R(r)) = 1/70.0$ otherwise.

The modification of seismic waves by local conditions, site effect $G(f)$, is expressed by the multiplication of a diminution function $D(f)$ and an amplification function $A(f)$. The diminution function accounts for the path-independent loss of high frequency in the ground motions and can be accounted for a simple filter of the form $D(f) = e^{-0.03\pi f}$. The amplification function $A(f)$ is based on empirical curves given for

Base-Isolated Systems, Reliability-Based Characterization of, Table 1 Parameters for the stochastic ground acceleration model

Parameter	Numerical value	Parameter	Numerical value
ρ_s (gm/cc)	2.8	β_s (km/s)	3.5
V	$1/\sqrt{2}$	R_ϕ	0.55
F	2.0	R_0 (km)	1.0

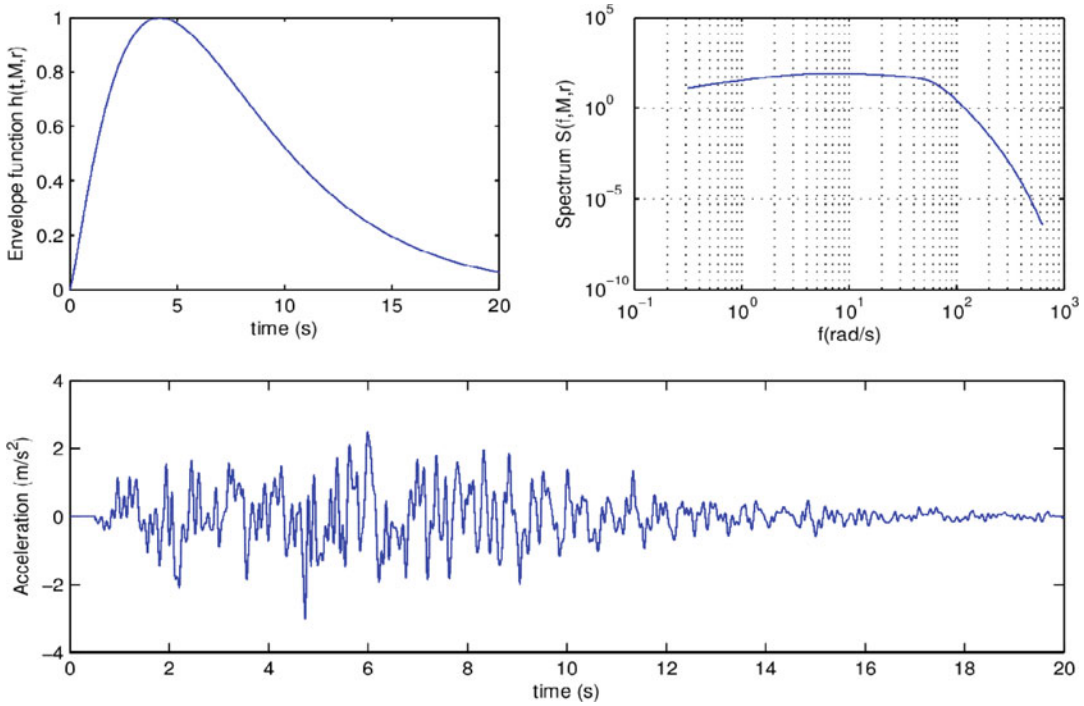
generic rock sites. An average constant value equal to 2.0 is considered.

Finally, the filter that controls the type of ground motion $I(f)$ is chosen as $I(f) = (2\pi f)^2$ for ground acceleration. The particular values of the different parameters of the stochastic ground acceleration model used in this work are given in Table 1. For a detailed discussion of the point-source model, the reader is referred to Anderson and Hough (1984), Boore et al. (1997), Atkinson and Silva (2000), and Boore (2003).

The previous excitation model is complemented by considering the moment magnitude M and epicentral distance r as uncertain. The uncertainty in moment magnitude is modeled by the Gutenberg-Richter relationship truncated on the interval $[6.0, 8.0]$, which leads to the probability density function $p(M) = b e^{-bM} / (e^{-6.0b} - e^{-8.0b})$ where $b = 1.8$ is a seismicity factor (Kramer 2003). For the uncertainty in the epicentral distance r , a lognormal distribution with mean value of 20 km and coefficient of variation equal to 35 % is considered. For illustration purposes Fig. 5 shows a typical envelope function, a ground motion spectrum, and a corresponding sample of ground motion.

Reliability Measures

The performance of base-isolated structural systems is characterized by means of a set of response functions $h_i(t, y, \theta)$, $i \in I_h$, $t \in [0, T]$, where y is the vector of controllable system parameters; θ is the vector of uncertain variables that characterizes the stochastic excitation model, i.e., the white noise sequence z and the



Base-Isolated Systems, Reliability-Based Characterization of, Fig. 5 Envelope function, ground motion spectrum, and a sample ground motion

seismological parameters M and r ; and I_h is a set of indices. The controllable system parameters are taken as the external diameter and the total height of rubber of the isolators, that is, $y^T = \langle D_r, H_r \rangle$. The probability that performance conditions are satisfied within a particular reference period T is used as a reliability measure. In this context a failure event F is defined in terms of a performance function $g(y, \theta)$ as $F(y, \theta) = g(y, \theta) \leq 0$, where

$$g(y, \theta) = 1 - \max_{i \in I_h} \left(\max_{t \in [0, T]} \frac{|h_i(t, y, \theta)|}{h_i^*} \right) \quad (10)$$

and h_i^* , $i \in I_h$ are the corresponding acceptable response levels. In particular, the probability of failure events associated with the base drift and superstructure absolute acceleration are considered in the present formulation. The probability of failure $P_F(y)$ evaluated at some particular value of the isolation system parameters y can be expressed in terms of the probability integral

$$P_F(y) = \int_{g(y, \theta) \leq 0} p(\theta) d(\theta) \quad (11)$$

where $p(\theta)$ is the multidimensional probability density function that characterizes the parameters involved in the excitation model. For systems under stochastic excitation, the probability integral involves in general a large number of uncertain parameters. Thus, the reliability estimation constitutes a high-dimensional problem. This problem is solved by applying an advanced simulation technique called subset simulation (Au and Beck 2001). In the approach, the failure probabilities are expressed as a product of conditional probabilities of some chosen intermediate failure events, the evaluation of which only requires simulation of more frequent events. Therefore, a rare event simulation problem is converted into a sequence of more frequent event simulation problems. Validation calculations have shown that subset simulation can be applied efficiently to first excursion response problems for a wide range of dynamical systems.

Details of this simulation procedure from the theoretical and numerical viewpoint can be found in Au and Beck (2001), Ching et al. (2005), and Zuev et al. (2011).

System Response

The response of base-isolated structural systems is obtained from the solution of the equation of motion that characterizes the combined system, that is, isolation system and superstructure (Eq. 3). The solution of this equation is obtained in an iterative manner due to the nonlinearity of the isolators forces. The solution scheme is implemented as follows:

- (i) At the beginning of the iteration ($l = 0$) within the time interval $[t, t + \Delta t]$, it is assumed that $\mathbf{d}_j^{(l)}(t + \Delta t) = \mathbf{d}_j(t)$ and $\mathbf{u}_j^{(l)}(t + \Delta t) = \mathbf{u}_j(t)$, $j = 1, \dots, n_I$ where $\mathbf{d}_j(t)$ and $\mathbf{u}_j(t)$ represent the direction of the movement and displacement vector of isolator number j , respectively, and n_I is the total number of isolators in the isolation system.
- (ii) The previous information and the characterization of the restoring force given in section “Isolation Model,” allow the characterization of the vector containing all nonlinear isolators forces $\mathbf{f}_{is}^{(l)}(t + \Delta t)$ (in the right-hand side of Eq. 3) is obtained.
- (iii) The solution of Eq. 3 is obtained by any suitable step-by-step integration scheme. Such solution gives the base responses $\mathbf{x}_b^{(l+1)}(t + \Delta t)$ and $\dot{\mathbf{x}}_b^{(l+1)}(t + \Delta t)$. From these quantities the isolator displacement and velocity vectors $\mathbf{u}_j^{(l+1)}(t + \Delta t)$ and $\dot{\mathbf{u}}_j^{(l+1)}(t + \Delta t)$, $j = 1, \dots, n_I$, respectively, are computed.
- (iv) The nonlinear differential Eq. 4 is integrated to obtain a new estimate for $\mathbf{d}_j(t + \Delta t)$, $j = 1, \dots, n_I$, i.e., $\mathbf{d}_j^{(l+1)}(t + \Delta t)$, and the nonlinear forces in the isolation system $\mathbf{f}_{is}(t + \Delta t)$ (in the right-hand side in Eq. 3). The solution of the equation for the evolution of the set of variables $\mathbf{d}_j(t)$, $j = 1, \dots, n_I$ is obtained by

a proper integration scheme such as the Crank-Nicolson method (Burden and Faires 2011).

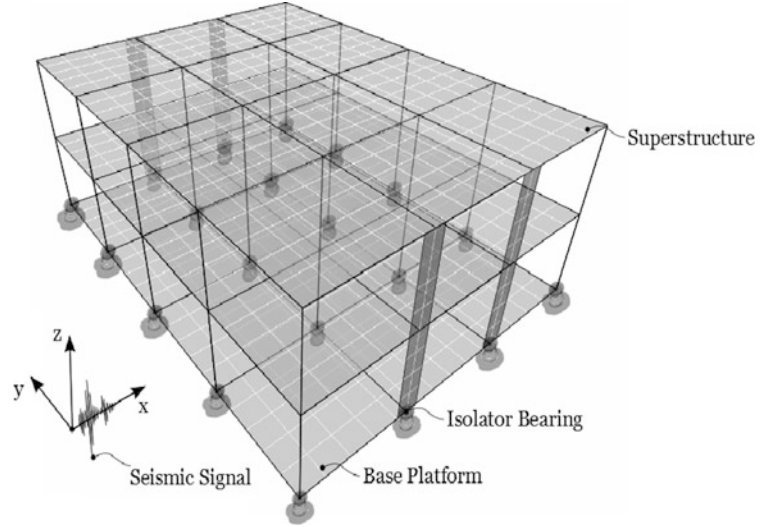
- (v) The iteration starting with solving Eq. 3 needs to be repeated until the variation of the quantities $\mathbf{d}_j(t + \Delta t)$, $j = 1, \dots, n_I$ is sufficiently small between two consecutive iterations (Jensen and Sepulveda 2011).

Application Problem

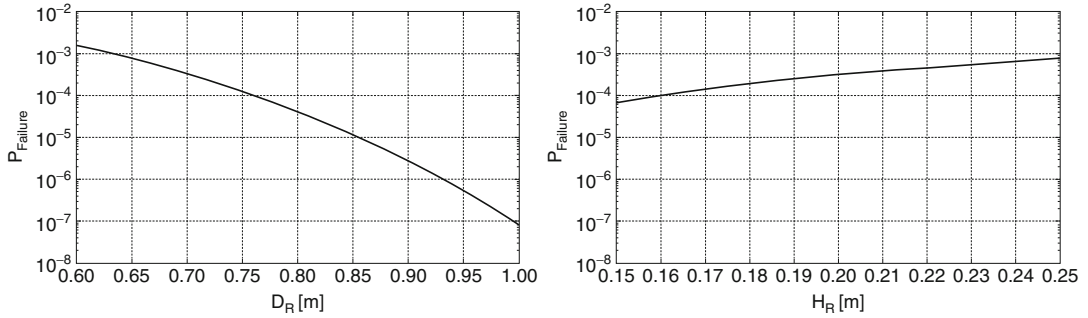
The objective of the application problem is to evaluate the effect of basic design isolator characteristics such as the rubber diameter (D_r) and the height of rubber (H_r) on the reliability of a base-isolated structural system. To this end, a three-dimensional reinforced concrete building model with more than 7000 degrees of freedom is considered for the analysis. The isometric view of the finite element model is shown in Fig. 6. Material properties of the reinforced concrete structure have been assumed as follows: Young's modulus $E = 2.5 \times 10^{10}$ N/m²; and Poisson ratio $\nu = 0.2$. The total mass of the first and second floor is 7.0×10^5 kg and 6.0×10^5 kg, respectively. On the other hand, the total mass of the platform is equal to 6.0×10^5 kg. The height of each floor is 3.5 m, and the floors are modeled with shell elements with a thickness of 0.25 m. Additionally, beam and column elements are used in the model. A 5 % of critical damping is added to the structural system. For anti-seismic design purposes, the structure is equipped with a total number of 20 rubber bearings in its isolation system. The structural system is excited by a ground acceleration in the x direction and modeled as described in section “Excitation Model.” The sampling interval and the duration of the excitation are taken equal to $\Delta t = 0.01$ s and $T = 30$ s, respectively. Thus, the generation of synthetic ground motions involved more than 3000 random variables in this case. The reliability of the base-isolated system is evaluated in terms of the probability of failure events associated with the base drift and superstructure absolute acceleration. The failure event related to the base drift is defined as

Base-Isolated Systems, Reliability-Based

Characterization of,
Fig. 6 Building model
with base isolation system



B



Base-Isolated Systems, Reliability-Based Characterization of, Fig. 7 Probability of the failure event related to the base drift as a function of the isolation system parameters

$$F_{\text{drift}}(y, \theta) = 1 - \max_{t \in [0, T]} \frac{|x_{bx}(t, y, \theta)|}{x_{bx}^*} \quad (12)$$

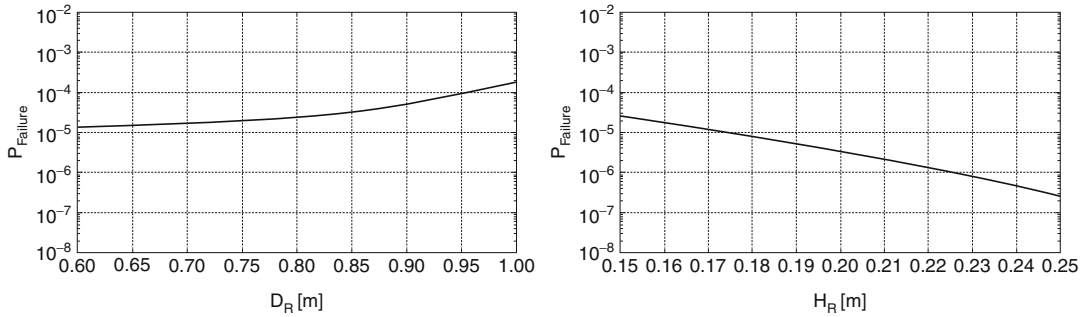
where $x_{bx}(t, y, \theta)$ represents the base displacement in the x direction, x_{bx}^* is the critical threshold level equal to 0.25 m, and $\theta^T = \langle z_1, z_2, \dots, z_{3001}, M, r \rangle$ is the vector of random variables that characterizes the excitation.

On the other hand, the failure event associated with the superstructure absolute acceleration is given by

$$F_{\text{acceleration}}(y, \theta) = 1 - \max_{i \in I} \left(\max_{t \in [0, T]} \frac{|\ddot{x}_{sxi}^{(\text{absolute})}(t, y, \theta)|}{\ddot{x}^*} \right) \quad (13)$$

where $\ddot{x}_{sxi}^{(\text{absolute})}(t, y, \theta), i \in I$ represent the absolute acceleration of the degrees of freedom of the superstructure in the x direction and $\ddot{x}^*$ is the acceptable level of response equal to 3.94 m/s^2 .

Figure 7 shows the probability of failure in terms of the base drift response as a function of the isolation system parameters. In this figure the nominal values of the isolation parameters are set equal to $D_r = 0.75 \text{ m}$, $H_r = 0.168 \text{ m}$, and $D_i = 0.10 \text{ m}$. It is seen that there is a decrease in the probability of failure as the rubber diameter is increased. This is reasonable since the base isolation system becomes stiffer in this case. On the other hand, the probability of failure increases as the height of the rubber



Base-Isolated Systems, Reliability-Based Characterization of, Fig. 8 Probability of the failure event related to the superstructure absolute acceleration as a function of the isolation system parameters

increases. In this case the isolation system becomes more flexible.

The effects of the isolation system parameters on the superstructure absolute acceleration are shown in Fig. 8. In this figure the probability of failure in terms of the absolute acceleration of the superstructure is presented. The probability of failure increases as the rubber diameter increases. In contrast, as the height of the rubber increases, the probability of failure is decreased. Clearly these effects are opposite to the effects of the isolation system parameters on the base response. For example, the introduction of additional stiffness in the isolation system makes the structural response ineffective. In other words, this additional stiffness which controls the base displacement makes the superstructure more vulnerable. This in turn affects the primary objective of the isolation system, namely, the reduction of the superstructure response. The previous results and observations show that the protection of the superstructure and the control of the base displacement are conflicting objectives.

Summary

A reliability-based characterization of base-isolated systems under stochastic excitation has been proposed. The characterization allows to evaluate the effect of important design isolator

characteristics on the reliability of such systems. The results of the example problem indicate that the minimization of the superstructure absolute acceleration and the minimization of the base displacement are conflicting goals. For example, the provision of additional stiffness to the isolation system decreases the superstructure reliability. Contrarily, the flexibility of the isolators has a positive impact on the reliability of the superstructure. However, this flexibility may induce undesirable effects on the base isolation system. Therefore, the performance of the superstructure as well as the response of the base isolation system should be considered simultaneously for a proper characterization of base-isolated systems. It is noted that additional considerations such as the evaluation of the superstructure reliability in terms of interstory drifts can also be taken into account. This could lead to important results and conclusions. Similarly, the explicit consideration of near-field ground motions may have important implications for flexible structures such as base-isolated systems. For these systems near-field ground motions may lead to excessive base deformations and superstructure deformations with important implications for the reliability of the combined structural system (isolation system and superstructure). In summary, the proposed reliability-based characterization of base-isolated systems under stochastic excitation provides a valuable basic tool for the analyst and designer.

Cross References

- [Code-Based Design: Seismic Isolation of Buildings](#)
- [Earthquake Mechanisms and Stress Field](#)
- [Engineering Characterization of Earthquake Ground Motions](#)
- [Integrated Earthquake Simulation](#)
- [Lead-Rubber Bearings with Emphasis on Their Implementation to Structural Design](#)
- [Nonlinear Dynamic Seismic Analysis](#)
- [Nonlinear Finite Element Analysis](#)
- [Probabilistic Seismic Hazard Models](#)
- [Random Process as Earthquake Motions](#)
- [Reliability Estimation and Analysis](#)
- [Robust Control of Building Structures Under Uncertain Conditions](#)
- [Seismic Reliability Assessment, Alternative Methods for](#)
- [Sensitivity of First-Excursion Probabilities for Nonlinear Stochastic Dynamical Systems](#)
- [Stochastic Ground Motion Simulation](#)
- [Structural Reliability Estimation for Seismic Loading](#)
- [Structural Seismic Reliability Analysis](#)
- [Subset Simulation Method for Rare Event Estimation: An Introduction](#)

References

- Anderson JG, Hough SE (1984) A model for the shape of the Fourier amplitude spectrum of acceleration at high frequencies. *Bull Seismol Soc Am* 74(5):1969–1993
- Atkinson GM, Silva W (2000) Stochastic modeling of California ground motions. *Bull Seismol Soc Am* 90(2):255–274
- Au SK, Beck JL (2001) Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilist Eng Mech* 16(4):263–277
- Boore DM (2003) Simulation of ground motion using the stochastic method. *Pure Appl Geophys* 160(3–4):635–676
- Boore DM, Joyner WB, Fumal TE (1997) Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: a summary of recent work. *Seismol Res Lett* 68(1):128–153
- Burden RL, Faires JD (2011) Numerical analysis. Brooks Cole, Pacific Grove, Boston, MA 02210, USA
- Ceccoli C, Mazzotti C, Savoia M (1999) Non-linear seismic analysis of base-isolated RC frame structures. *Earthq Eng Struct Dyn* 28(6):633–653
- Ching J, Au SK, Beck JL (2005) Reliability estimation for dynamical systems subject to stochastic excitation using subset simulation with splitting. *Comput Method Appl Mech Eng* 194(12–16):1557–1579
- Chopra AK (2001) Dynamics of structures: theory and applications to earthquake engineering. Prentice Hall, Englewood Cliffs, Upper Saddle River, NJ 07458, USA
- De Luca A, Mele E, Molina J, Verzeletti G, Pinto AV (2001) Base isolation for retrofitting historic buildings: evaluation of seismic performance through experimental investigation. *Earthq Eng Struct Dyn* 30(8):1125–1145
- Jensen HA, Sepulveda JC (2011) On the effects of vibration control devices in the reliability-based design of stochastic dynamical systems. In: Proceedings of the 8th international conference on structural dynamics (EURODYN 2011), Leuven, 4–6 July 2011
- Kelly JM (1986) Aseismic base isolation: review and bibliography. *Soil Dyn Earthq Eng* 5(4):202–216
- Kramer SL (2003) Geotechnical earthquake engineering. Prentice Hall, Upper Saddle River, Englewood Cliffs, New Jersey, NJ, USA
- Lutes L, Sarkani S (2004) Random vibrations: analysis of structural and mechanical systems. Elsevier Butterworth-Heinemann, Burlington
- Makris N, Chang S (1998) Effects of damping mechanisms on the response of seismically isolated structures. PEER report 1998/06. Pacific Earthquake Engineering Research Center. College of Engineering, University of California, Berkeley
- Minewaki S, Yamamoto M, Higashino M, Hamaguchi H, Kyuke H, Sone T, Yoneda H (2009) Performance tests of full size isolators for super high-rise isolated buildings. *J Struct Eng AIJ* 55(B):469–477
- Mokha AS, Amin N, Constantinou MC, Zayas V (1996) Seismic isolation retrofit of large historic building. *J Struct Eng* 122(3):298–308
- Su L, Ahmadi G, Tadjbakhsh JG (1990) A comparative study of performances of various base isolation systems, part II: sensitivity analysis. *Earthq Eng Struct Dyn* 19:21–33
- Yamamoto M, Minewaki S, Higashino M, Hamaguchi H, Kyuke H, Sone T, Yoneda H (2009) Performance tests of fully size rubber bearings for isolated superhigh-rise buildings. In: International symposium on seismic response controlled buildings for sustainable society, S1-4, Tokyo
- Zou X-K, Wang Q, Li G, Chan C-M (2010) Integrated reliability-based seismic drift design optimization of base-isolated concrete buildings. *J Struct Eng* 136(10):1282–1295
- Zuev K, Beck JL, Au SK, Katafygiotis L (2011) Bayesian post-processor and other enhancements of Subset Simulation for estimating failure probabilities in high dimensions. *Comput Struct* 5:599–607

Bayesian Operational Modal Analysis

Siu-Kui Au

Institute for Risk & Uncertainty and Centre for Engineering Dynamics, Liverpool, UK

Synonyms

Ambient modal identification; Bayes' theorem; Modal identification; Operational modal analysis; Signal-to-noise ratio; System identification; Uncertainty law

Introduction

Operational modal analysis, or ambient modal identification, aims at identifying the modal properties (natural frequency, damping ratio, mode shape, etc.) of an instrumented structure using only the (output) vibration response (acceleration, velocity, etc.). The input excitation to the structure is not measured but is assumed to be broadband random, often referred to as “ambient.” This allows vibration data to be collected when the structure is in its working or “operating” condition without much intervention, therefore implying significant economy over free-vibration (initially excited but no input afterwards) or forced-vibration tests (known input). The broadband random assumption essentially requires that the spectral characteristics (shape) of the measured response reflect the properties of the modes rather than those of the excitation, which is assumed to be constant in the vicinity of the natural frequencies.

In the absence of input loading, the uncertainties associated with the identified modal properties are often significantly higher than those identified from free-vibration tests or forced-vibration tests. The situation is aggravated by the presence of “modeling errors” that can be significant in field tests, i.e., the structure may not obey the modeling assumptions used in identifying the modal properties.

A Bayesian system identification approach provides a fundamental mathematical framework for quantifying the uncertainties and their effects on the identification results. This article describes the basics of Bayesian operational modal analysis, covering issues on formulation, efficient computations, interpretation of results, the quantification of identification uncertainties, and their management.

The methods presented in this article make use of the FFT (Fast Fourier transform) of measured acceleration data on a selected frequency band around the modes of interest. The modes are assumed to be classically damped.

In theory, ambient modal identification can be performed in either the time domain or the frequency domain. In practice, however, a frequency domain approach is preferable because it allows a natural partitioning of information in the data for identifying the modes of interest. It significantly simplifies the identification model because it only needs to model the modes in the selected band. For well-separated modes, one can select the band to cover one mode only, so that it can be identified independently of other modes. In general, the number of modes in the identification model only needs to be equal to the number of closely spaced modes, which rarely exceeds three. In the Bayesian formulation, this does not require any band-pass filtering because it can be done by simply omitting the FFT data of the excluded bands from the likelihood function.

Using only the FFT data within a selected band also significantly reduces modeling error risk because the identification results are invariant to complexities in the excluded bands containing other modes not of interest or even dynamics that are difficult to model. The mechanical response, excitation, and channel noise are assumed to have a flat spectrum within the selected frequency band only, rather than over the whole sampling band from zero to the Nyquist frequency (half of the sampling frequency). The latter is inevitable in time-domain identification approaches.

As a literature note, operational modal analysis was first formulated using a Bayesian approach in the Master's Thesis of Yuen (1999).

This led to the time-domain formulation in Yuen and Katafygiotis (2001a), a spectral density formulation in Yuen and Katafygiotis (2001b) and later an FFT formulation in Yuen and Katafygiotis (2003). As mentioned before, the time-domain formulation is not preferred for its assumption on the prediction error and modal force. The spectral density formulation involves statistical averaging and may not make full use of the information contained in the data. The Bayesian formulation based on FFT is therefore the preferred choice. The above formulations are only computationally feasible for a small number of measured degrees of freedom (dofs). This partly explains why the formulations did not flourish and there were no applications for a number of years after the formulations first appeared. Based on the FFT formulation, fast but equivalent formulations that allow mathematical analysis and feasible implementation in general were developed in Au (2011) (see also Zhang and Au 2013) for well-separated modes, Au (2011, 2012a) for multiple (possibly close) modes, and Au and Zhang (2012a) and Zhang et al. (2015) for well-separated modes but multiple setups. Application to field or experimental data can be found in Au et al. (2012a, b), Au and To (2012), Au and Zhang (2012b), and Fan et al. (2012). See also Yuen and Kuok (2010a, b) for application of the Bayesian spectral density approach. Asymptotic uncertainty laws that provide insights on the achievable limits operational modal analysis have been developed in Au (2014a, b). Recent overviews can be found in Yuen and Kuok (2011) and Au et al. (2013).

Context

The primary objective of ambient modal identification is to identify the natural frequencies, damping ratios, and mode shapes of a constructed structure using (output) vibration response measurement at a number of measured degrees of freedom (dofs) under “ambient conditions.” The latter is a notion relative to the mode of interest. It refers specifically to the situation

where it is justified to assume that the modal (not physical) force has a constant power spectral density in the vicinity of the resonance band dominated by the mode.

Following a Bayesian approach in the frequency domain, use is made of the (scaled) FFT of the measured acceleration response at a number of dofs on the structure, denoted by $\{\hat{\mathbf{x}}_j \in R^n : j = 0, \dots, N-1\}$ and abbreviated as $\{\hat{\mathbf{x}}_j\}$, where n is the number of measured dofs. The (scaled) FFT of $\{\hat{\mathbf{x}}_j\}$ is the complex-valued sequence $\{F_k \in C^n : k = 0, \dots, N-1\}$ where

$$F_k = \sqrt{\frac{2\Delta t}{N}} \sum_{j=0}^{N-1} \hat{\mathbf{x}}_j \exp(-2\pi i j k / N) \quad (1)$$

$$k = 0, \dots, N-1$$

and $i^2 = -1$ and Δt is the sampling interval. For a given k , the FFT F_k corresponds to frequency

$$f_k = k / N \Delta t \quad k = 0, \dots, \text{int}[N/2] \quad (2)$$

where $\text{int}[N/2]$ is the integer part of $N/2$ and is the index corresponding to the Nyquist frequency.

The FFT F_k has been scaled so that $F_k F_k^* \in C^{n \times n}$ gives the one-sided sample spectral density (periodogram) matrix. Summing the i -th diagonal element of $F_k F_k^*$ over $k = 0, \dots, N-1$ and multiplying by the frequency resolution $\Delta f = 1/N \Delta t$ gives twice the mean square value of the i -th measured dof (Parseval equality).

For the purpose of identifying the modes of interest, it is sufficient and preferable to make use of the FFT data only on a selected frequency band around the mode(s). This is because in reality, the data contains a variety of dynamic (colored) activities over its sampling spectrum (up to the Nyquist frequency), most of which are irrelevant to identifying the mode(s) or are difficult to model.

Suppose there are m modes within the selected frequency band and let $\{F_k\}$ denote the collection of FFT data within the band. It is modeled as a sum of contributions from the structural vibration (the target to be measured) and a prediction error

(noise) that accounts for the difference between the theoretical and measured response:

$$F_k = \sum_{i=1}^m \Phi_i \ddot{\eta}_{ik} + \epsilon_k \quad (3)$$

where $\Phi_i \in R^n$ ($i = 1, \dots, m$) is the mode shape of the i -th mode in the selected frequency band; $\ddot{\eta}_{ik} \in C$ is the FFT at frequency f_k of the i -th modal acceleration response; and $\epsilon_k \in C^n$ is the FFT of the prediction error, which may arise due to measure (channel) noise and modeling error (e.g., unaccounted dynamics). Assuming classically damped modes, the time-domain counterpart of $\ddot{\eta}_{ik}$ satisfies the uncoupled equation of motion:

$$\ddot{\eta}_i(t) + 2\zeta_i \omega_i \dot{\eta}_i(t) + \omega_i^2 \eta_i(t) = p_i(t) \quad (4)$$

where $\omega_i = 2\pi f_i$; f_i , ζ_i , and $p_i(t)$ are, respectively, the natural frequency (in Hz), damping ratio, and modal force.

In the above context, the set of modal parameters to be identified from the FFT data $\{F_k\}$, denoted by θ , consists of the parameters:

- (1) Natural frequencies $f_1, \dots, f_m \in R^+$
- (2) Damping ratios $\zeta_1, \dots, \zeta_m \in (0, 1)$
- (3) Mode shapes $\Phi_1, \dots, \Phi_m \in R^n$
- (4) PSD matrix of modal forces $S \in C^{n \times n}$ (assumed constant within the band)
- (5) PSD of prediction error $S_e \in R^+$ (assumed constant within the band)

Each mode shape $\Phi_i = [\Phi_{1i}, \Phi_{2i}, \dots, \Phi_{ni}]^T$ is subjected to a unit norm constraint, i.e.,

$$\|\Phi_i\|^2 = \Phi_i^T \Phi_i = \sum_{j=1}^n \Phi_{ji}^2 = 1 \quad i = 1, \dots, m \quad (5)$$

Formulation from First Principle

In a Bayesian context, the information about the set of modal parameters θ that can be inferred from the FFT data $\{F_k\}$ is encapsulated in the “posterior probability density function” (PDF) of

θ , denoted by $p(\theta|\{F_k\})$. Using Bayes’ theorem, it is given by

$$p(\theta|\{F_k\}) = p(\{F_k\})^{-1} p(\{F_k\}|\theta) p(\theta) \quad (6)$$

The first term on the RHS does not depend on θ . As far as the identification of θ is concerned, it can be ignored. The middle-term $p(\{F_k\}|\theta)$ is called the “likelihood function.” It is derived based on modeling assumptions that relate the response of the theoretical model for a given θ to the FFT data. The middle-term $p(\theta)$ is the “prior distribution” that reflects one’s knowledge about θ in the absence of data. In modal identification problems with sufficient data, it can be taken as a constant because it is slowly varying compared to the likelihood function. As a result, it is sufficient for modal identification problems to use

$$p(\theta|\{F_k\}) \propto p(\{F_k\}|\theta) \quad (7)$$

Deriving the likelihood function requires deriving the joint PDF of $\{F_k\}$ for a given θ . This PDF is generally complicated, but it turns out to admit an asymptotic closed form when the number of data points N is large (Brillinger 1981). In particular, as $N \rightarrow \infty$, $\{F_k\}$ at different k ’s are asymptotically independent. For a given k , the real and imaginary parts of F_k are jointly Gaussian with zero mean. Thus,

$$p(\theta|\{F_k\}) \propto p(\{F_k\}|\theta) = (2\pi)^{-nN_f} \left[\prod_k \det C_k \right]^{-1/2} \exp \left\{ -\frac{1}{2} \sum_k \begin{bmatrix} \text{Re} F_k \\ \text{Im} F_k \end{bmatrix}^T C_k^{-1} \begin{bmatrix} \text{Re} F_k \\ \text{Im} F_k \end{bmatrix} \right\} \quad (8)$$

where the sum is overall frequency ordinates in the selected band, whose number is equal to N_f ; $C_k \in R^{2n \times 2n}$ is the covariance matrix of the augmented vector $[\text{Re} F_k; \text{Im} F_k] \in R^{2n}$; and $\det(\cdot)$ denotes the determinant of the argument matrix. For high sampling rate, it can be shown using random-vibration theory (Lutes and Sarkani 1997) that asymptotically

$$\mathbf{C}_k = \frac{1}{2} \begin{bmatrix} \Phi(\text{Re}\mathbf{H}_k)\Phi^T & -\Phi(\text{Im}\mathbf{H}_k)\Phi^T \\ \Phi(\text{Im}\mathbf{H}_k)\Phi^T & \Phi(\text{Re}\mathbf{H}_k)\Phi^T \end{bmatrix} + \frac{S_e}{2} \mathbf{I}_{2n} \quad (9)$$

where $\Phi = [\Phi_1, \dots, \Phi_m] \in R^{n \times m}$ is the mode shape matrix and $\mathbf{H}_k \in C^{m \times m}$ is the modal transfer matrix whose (i, j) entry is given by

$$\mathbf{H}_k(i, j) = S_{ij} h_{ik} h_{jk}^* \quad (10)$$

$$h_{ik} = [(\beta_{ik}^2 - 1) + (2\zeta_i \beta_{ik})\mathbf{i}]^{-1} \quad (11)$$

$$\beta_{ik} = f_i / f_k \quad (12)$$

and S_{ij} is the (i, j) entry of the PSD matrix of modal force, $\mathbf{S} \in C^{m \times m}$. Note that $\mathbf{H}_k$ and $\mathbf{S}$ are Hermitian (i.e., equal to its conjugate transpose); $\mathbf{C}_k$ is real symmetric.

Posterior Most Probable Value and Covariance Matrix

For convenience in analysis or computation, the posterior PDF is written in terms of the NLLF (negative log-likelihood function):

$$L(\boldsymbol{\theta}) = \frac{1}{2} \sum_k \text{Indet} \mathbf{C}_k + \frac{1}{2} \sum_k \begin{bmatrix} \text{Re} F_k \\ \text{Im} F_k \end{bmatrix}^T \mathbf{C}_k^{-1} \begin{bmatrix} \text{Re} F_k \\ \text{Im} F_k \end{bmatrix} \quad (13)$$

so that

$$p(\boldsymbol{\theta} | \{F_k\}) \propto \exp(-L(\boldsymbol{\theta})) \quad (14)$$

Note that the constant $nN_f \ln(2\pi)$ has been omitted from the NLLF.

With sufficient data often encountered in practice, the posterior PDF has a single peak in the parameter space of $\boldsymbol{\theta}$, which is the most probable value (MPV) of the modal parameters given the data. Equivalently, the NLLF has a unique minimum at the MPV. Approximating the NLLF by a second-order Taylor expansion about the MPV leads to a Gaussian approximation of the posterior PDF. The resulting Gaussian distribution has a mean at the MPV, and its covariance matrix, called the “posterior covariance matrix,” reflects the remaining uncertainties of the modal parameters

in the presence of the data. Mathematically, the posterior covariance matrix is equal to the inverse of the Hessian of the NLLF at the MPV.

The above results can be reasoned as follows. Given the data, let the MPV of the modal parameters be denoted by the vector $\hat{\boldsymbol{\theta}}$. Approximating the NLLF by a second-order Taylor expansion about $\hat{\boldsymbol{\theta}}$,

$$L(\boldsymbol{\theta}) \approx L(\hat{\boldsymbol{\theta}}) + \nabla L(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) + \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \nabla^2 L(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \quad (15)$$

where $\nabla L(\hat{\boldsymbol{\theta}})$ and $\nabla^2 L(\hat{\boldsymbol{\theta}})$ denote, respectively, the gradient (row) vector and Hessian matrix of the NLLF at $\hat{\boldsymbol{\theta}}$. Since the NLLF is minimized at $\hat{\boldsymbol{\theta}}$, $\nabla L(\hat{\boldsymbol{\theta}}) = \mathbf{0}$ (a zero vector) and $\nabla^2 L(\hat{\boldsymbol{\theta}})$ are positive definite matrix. Applying this on Eq. 15 and then substituting into Eq. 14 gives

$$\begin{aligned} p(\boldsymbol{\theta} | \{F_k\}) &\propto \exp \left[-L(\hat{\boldsymbol{\theta}}) - \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \nabla^2 L(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right] \\ &\propto \exp \left[-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \nabla^2 L(\hat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right] \\ &\propto \exp \left[-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \hat{\mathbf{C}}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right] \end{aligned} \quad (16)$$

where

$$\hat{\mathbf{C}} = \nabla^2 L(\hat{\boldsymbol{\theta}})^{-1} \quad (17)$$

is a positive definite matrix. Eq. 16 is just the variable part of a joint Gaussian PDF with zero mean and covariance matrix equal to $\hat{\mathbf{C}}$. The proportionality constant of the equation is $(2\pi)^{-n_0/2} (\det \hat{\mathbf{C}})^{-1/2}$ (n_0 is the number of modal parameters), which can be obtained by integration or by probability reasoning. Thus,

$$\begin{aligned} p(\boldsymbol{\theta} | \{F_k\}) &= (2\pi)^{-n_0/2} (\det \hat{\mathbf{C}})^{-1/2} \\ &\exp \left[-\frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^T \hat{\mathbf{C}}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \right] \end{aligned} \quad (18)$$

Computational Problems

The MPV of modal parameters cannot be obtained analytically because the relationship

between the NLLF and the modal parameters is complicated. Brute-force numerical minimization of the NLLF with respect to the modal parameters is prohibitive or not convergent, primarily because the number of modal parameters is typically large. The latter arises primarily from the number of measured dofs and secondarily from the number of modes. Taking into account the Hermitian nature of the PSD matrix of modal force, the number of modal parameters in $\boldsymbol{\theta}$ is equal to

$$n_{\theta} = (m + 1)^2 + nm \quad (19)$$

For example, a setup with 18 measured dofs and a band with $m = 2$ modes result in $(2 + 1)^2 + (18)(2) = 45$ parameters.

Adding to the difficulty is the ill-conditioned nature of the matrix $\mathbf{C}_k$ in Eq. 9. For good quality data, it is close to being rank deficient since it is then dominated by the first term, which has a rank of at most $2m$ and is often less than $2n$ (the number of modes m is often less than the number of measured dofs n).

Fast Equivalent Formulation

Using linear algebra techniques, it is possible to rewrite the NLLF in a form that facilitates analysis and computations. In particular, the NLLF can be written as a quadratic form of the mode shape, thereby allowing the most probable mode shape to be determined efficiently when the remaining parameters are given. Based on this, the MPV of all modal parameters can be obtained by optimizing different groups of parameters in turn, iterating until convergence.

Using the equivalent forms, analytical expressions for the Hessian of the NLLF can also be derived effectively. This allows the posterior covariance matrix to be evaluated efficiently and accurately, without resorting to finite difference method.

Single Mode

When there is only one mode in the selected band, the set of modal parameters reduces to

$$\boldsymbol{\theta} = \{f, \zeta, S, S_e, \boldsymbol{\Phi}\} \quad (20)$$

where mode index has been dropped for simplicity; $S \in R^+$ is the PSD of modal force; $\boldsymbol{\Phi} \in R^n$ is the mode shape.

It can be shown that the NLLF in Eq. 13 can be rewritten as

$$\begin{aligned} L(\boldsymbol{\theta}) = & -nN_f \ln 2 + (n - 1)N_f \ln S_e \\ & + \sum_k \ln(SD_k + S_e) \\ & + S_e^{-1} (d - \boldsymbol{\Phi}^T \mathbf{A} \boldsymbol{\Phi}) \end{aligned} \quad (21)$$

where

$$\mathbf{A} = \sum_k \left(1 + \frac{S_e}{SD_k} \right)^{-1} (\text{Re} F_k \text{Re} F_k^T + \text{Im} F_k \text{Im} F_k^T) \in R^{n \times n} \quad (22)$$

$$D_k = \left[(\beta_k^2 - 1)^2 + (2\zeta\beta_k)^2 \right]^{-1} \beta_k = f/f_k \quad (23)$$

$$d = \text{Re} F_k^T \text{Re} F_k + \text{Im} F_k^T \text{Im} F_k \quad (24)$$

Note that $\mathbf{A}$ depends on f, ζ, S, S_e ; D_k depends on f, ζ . The significance of Eq. 21 in comparison to Eq. 13 is that it no longer involves the inverse of ill-conditioned matrix. More importantly, the mode shape $\boldsymbol{\Phi}$ only appears in the quadratic form. Given the remaining parameters $\{f, \zeta, S, S_e\}$, minimizing the NLLF with respect to $\boldsymbol{\Phi}$ subjected to the norm constraint $\boldsymbol{\Phi}^T \boldsymbol{\Phi} = 1$ gives the most probable shape as the eigenvector of $\mathbf{A}$ with the largest eigenvalue. This effectively reduces the dimension of the optimization problem for MPV to only four (f, ζ, S, S_e), which can be readily performed numerically (Au 2011).

Multiple Modes

When there is more than one mode in the selected band, the set of modal parameters $\boldsymbol{\theta}$ is given by that stated in section Context, which is repeated here:

- (1) Natural frequencies $f_1, \dots, f_m \in \mathbb{R}^+$
- (2) Damping ratios $\zeta_1, \dots, \zeta_m \in (0, 1)$
- (3) Mode shapes $\Phi_1, \dots, \Phi_m \in \mathbb{R}^n$
- (4) PSD matrix of modal forces $\mathbf{S} \in \mathbb{C}^{n \times n}$
(assumed constant within the band)
- (5) PSD of prediction error $S_e \in \mathbb{R}^+$ (assumed constant within the band)

Each mode shape $\Phi_i = [\Phi_{1i}, \Phi_{2i}, \dots, \Phi_{ni}]^T$ is subjected to a unit norm constraint, i.e.,

$$\|\Phi_i\|^2 = \Phi_i^T \Phi_i = \sum_{j=1}^n \Phi_{ji}^2 = 1 \quad i = 1, \dots, m \quad (25)$$

In this case the most probable mode shape cannot be obtained directly by solving a standard eigenvalue problem. The mathematical structure of the optimization problem is more complex because the mode shapes are not necessarily orthogonal to each other. However, it is possible to reduce the complexity by representing the mode shape via a set of orthonormal basis and noting that the dimension of the subspace spanned by such basis does not exceed the number of modes. The mode shape matrix $\Phi = [\Phi_1, \dots, \Phi_m] \in \mathbb{R}^{n \times m}$ is represented as

$$\Phi = \mathbf{B}' \alpha \quad (26)$$

where

$$\mathbf{B}' = [\mathbf{B}'_1, \dots, \mathbf{B}'_{m'}] \in \mathbb{R}^{n \times m'} \quad (27)$$

contains in its columns a set of orthonormal “mode shape basis” $\{\mathbf{B}'_i \in \mathbb{R}^n : i = 1, \dots, m'\}$ spanning the “mode shape subspace;” $\alpha \in \mathbb{R}^{m' \times m}$ contains in its columns the coordinates of each mode shape with respect to the mode shape basis; $m' \leq \min(n, m)$ is the dimension of the mode shape subspace, which can be determined with the help of a singular-value spectrum. The MPVs of $\mathbf{B}'$ and α need to be determined in the identification process.

Based on Eq. 26, it can be shown that the NLLF can be rewritten as

$$\begin{aligned} L(\theta) = & -nN_f \ln 2 + (n - m')N_f \ln S_e \\ & + S_e^{-1}d + \sum_k \ln |\det \mathbf{E}'_k| \\ & - S_e^{-1} \sum_k F_k^* \mathbf{B}' (\mathbf{I}_{m'} - S_e \mathbf{E}'_k^{-1}) \mathbf{B}'^T F_k \end{aligned} \quad (28)$$

where d is given by Eq. 24 as in the case of single mode, and

$$\mathbf{E}'_k = \alpha \mathbf{H}_k \alpha^T + S_e \mathbf{I}_{m'} \quad (29)$$

is an m' -by- m' Hermitian matrix. The significance of Eq. 28 is that the mode shape basis $\mathbf{B}$ has been segregated out in the last term, which is quadratic in nature. The most probable basis should therefore minimize the quadratic form under orthonormal constraints. This does not lead to a standard eigenvalue problem, but procedures have been developed that allow the most probable basis to be determined efficiently by Newton iteration. This resolves the potentially high-dimensional optimization with respect to the mode shape. Based on this, a strategy has been developed for determining the MPV of different groups of parameters, iterating until convergence (Au 2012a).

Modal Signal-to-Noise Ratio

The term “signal-to-noise ratio” (s/n ratio) is often used to describe the quality of data to indicate the contribution of the signal targeted to be measured compared to the contribution of “noise,” which inevitably arises from sensor, cables transmitting the signal (especially for analog signals), digitizing hardware, etc. The actual quantitative definition differs depending on the particular application, however. In ambient modal identification, the s/n ratio that affects the identification results is related to the spectral contribution of the modal response compared to the spectral contribution of the noise in the vicinity of the resonance frequency band. In the case of well-separated modes, a definition of the s/n ratio that is found to govern fundamentally the

characteristics of the identification results is the PSD of the theoretical acceleration response at resonance ($SD_k = S/4\zeta^2$) divided by the PSD of the prediction error (S_e):

$$\gamma = \frac{S}{4S_e\zeta^2} \quad (30)$$

Governing Factors

The modal s/n ratio in Eq. 30 depends on the following factors:

- (1) The prediction error PSD S_e . This includes contributions from the channel noise as well as modeling errors, e.g., due to unaccounted modes and colored nature of modal force.
- (2) The damping ratio ζ . The smaller the damping, the higher the modal s/n ratio, as a result of higher modal response PSD at resonance.
- (3) The modal force PSD S . This in turn depends on the intensity of environmental excitation and (less trivially) on the measured dofs. Specifically, let ξ denote the “full” mode shape of the structure containing all (possibly an infinite number of) dofs, and it is scaled to have unit norm, i.e., $\xi^T \xi = 1$. Then it can be reasoned that

$$S = S_p \left(\sum_i \xi_i^2 \right) \quad (31)$$

where the sum is over the measured dofs, and

$$S_p = \frac{\xi^T \mathbf{S}_F \xi}{(\xi^T \mathbf{M} \xi)^2} \quad (32)$$

is the modal force PSD consistent with the scaling of the global mode shape ξ ; $\mathbf{M}$ is the mass matrix of the structure. Note that in reality, the full mode shape ξ and the mass matrix $\mathbf{M}$ can hardly be identified and so is S_p . Instead, the measured mode shape Φ and the modal force PSD S consistent with the scaling $\Phi^T \Phi = 1$ is identified. Eq. 31 shows that the modal force PSD S always increases with the number of measured dofs, although the rate depends on the mode shape of the dof incrementally added to the existing set of measured dofs. As a result, the

modal s/n ratio in Eq. 30 always increases with the number of measured dofs.

High s/n Asymptotics

It turns out that the Bayesian modal identification results in terms of the MPV and posterior covariance matrix reduce to intuitive forms when the modal s/n ratio is large (theoretically infinity). This is referred to as “high s/n asymptotics.”

Asymptotically for large modal s/n ratio, the terms in the NLLF can be separated into two terms: one quadratic term depending only on the mode shape (and the prediction error PSD in a multiplicative manner) and the other term depending only on the remaining parameters (natural frequency, damping ratio, modal force PSD, prediction error PSD). The mode shapes are therefore asymptotically uncoupled from the remaining parameters. This has important implications on the behavior of the MPV and the posterior covariance matrix. The former provides a sound basis for setting initial guess and designing fast algorithms for determining the MPV of modal parameters. The latter leads to remarkably simple closed-form formulas for the posterior variance of the modal parameters. These are collectively known as “uncertainty laws,” which govern the achievable limits of operational modal analysis.

The asymptotic behavior of the most probable mode shape is discussed in sections [Single Mode](#) and [Multiple Modes](#) for a single mode and multiple modes in the selected frequency band, respectively. The posterior covariance matrix is discussed in the context of uncertainty laws in section [Uncertainty Laws](#).

Single Mode

For a single mode in the selected band, asymptotically the most probable mode shape Φ is equal to the eigenvector with the largest eigenvalue of the following matrix:

$$\mathbf{A}_0 = \sum_k (\text{Re} F_k \text{Re} F_k^T + \text{Im} F_k \text{Im} F_k^T) \quad (33)$$

This matrix depends only on the data, and so the most probable mode shape can be calculated explicitly without optimization.

The above result can be reasoned as follows. Asymptotically the matrix $\mathbf{A}$ in the NLLF Eq. 21 becomes $\mathbf{A}_0$ in Eq. 33, and so the NLLF becomes

$$\begin{aligned} L(\boldsymbol{\theta}) \sim & -nN_f \ln 2 + (n-1)N_f \ln S_e \\ & + \sum_k \ln(SD_k + S_e) + S_e^{-1}d \\ & - S_e^{-1} \boldsymbol{\Phi}^T \mathbf{A}_0 \boldsymbol{\Phi} \end{aligned} \quad (34)$$

In the above expression, only the last term depends on $\boldsymbol{\Phi}$, which is a quadratic form with a constant coefficient matrix $\mathbf{A}_0$. Minimizing it with respect to $\boldsymbol{\Phi}$ and subjected to the norm constraint $\boldsymbol{\Phi}^T \boldsymbol{\Phi} = 1$ gives the most probable mode shape as the eigenvector of $\mathbf{A}_0$ with the largest eigenvalue.

Multiple Modes

When there is more than one mode in the selected band, asymptotically the most probable mode shape basis is given by the eigenvectors of $\mathbf{A}_0$ in Eq. 33 corresponding to the first m' largest eigenvalues. Note that this does not give directly the most probable mode shapes because the MPV of the coordinates with respect to the basis, i.e., $\boldsymbol{\alpha}$ in Eq. 26, is not known but its determination is still nontrivial.

The above can be reasoned as follows. Asymptotically, the term $(\mathbf{I}_{m'} - S_e \mathbf{E}'_k)$ in the quadratic term of the NLLF in Eq. 28 becomes the identity matrix $\mathbf{I}_{m'}$ and so

$$\begin{aligned} L(\boldsymbol{\theta}) \sim & -nN_f \ln 2 + (n-m')N_f \ln S_e \\ & + S_e^{-1}d + \sum_k \ln|\det \mathbf{E}'_k| \\ & - S_e^{-1} \sum_k F_k^* \mathbf{B}' \mathbf{B}'^T F_k \end{aligned} \quad (35)$$

This expression depends on the mode shape basis $\mathbf{B}'$ only through the sum in the last term, which is quadratic in nature. Substituting $\mathbf{B}' = [\mathbf{B}'_1, \dots, \mathbf{B}'_{m'}]$, the sum can be rewritten as

$$\begin{aligned} \sum_k F_k^* \mathbf{B}' \mathbf{B}'^T F_k &= \sum_{i=1}^{m'} \mathbf{B}'_i{}^T \left(\sum_k F_k F_k^* \right) \mathbf{B}'_i \\ &= \sum_{i=1}^{m'} \mathbf{B}'_i{}^T \mathbf{A}_0 \mathbf{B}'_i \end{aligned} \quad (36)$$

Minimizing the NLLF is equivalent to maximizing the sum on the rightmost. Since it contains no cross terms between $\mathbf{B}'_i$ and $\mathbf{B}'_j$ ($i \neq j$), maximizing it gives the most probable basis as the eigenvectors of $\mathbf{A}_0$ corresponding to the first m' largest eigenvalues.

Quantification of Identification Uncertainty

Mode Shape Uncertainty

For scalar-valued modal parameters such as the natural frequency and damping ratio, the posterior uncertainty can be conveniently described in terms of the posterior standard deviation or in a nondimensional manner the posterior coefficient of variation (c.o.v.), equal to the posterior standard deviation divided by the MPV. The posterior standard deviation is equal to the square root of the posterior variance, which can be obtained from the corresponding diagonal entry of the posterior covariance matrix.

The above quantification cannot be applied to the mode shape because it is vector valued and is subject to the norm constraint. One way to quantify the uncertainty of the mode shape is through the Expected modal assurance criterion (MAC), defined analogously as in the deterministic case. According to the posterior distribution, which is approximated by a Gaussian distribution, given the measured data, the mode shape is a Gaussian vector with mean vector equal to the most probable mode shape and a covariance matrix $\mathbf{C}_\Phi \in R^{n \times n}$. The latter is given by the corresponding $n \times n$ partition of the full covariance matrix (containing all modal parameters). The Expected MAC is defined as the expected value of the cosine of the hyper angle between the random mode shape vector and the most probable mode shape vector. This value ranges between 0 and 1; the higher the value, the smaller the mode shape uncertainty. An exact closed-form expression for the Expected MAC has not been developed, but an asymptotic expression is available, which is adequate for practical purposes. It can be shown that asymptotically

$$\text{Expected MAC} = \left(1 + \sum_{i=2}^n \kappa_i^2\right)^{-1/2} \quad (37)$$

where $\{\kappa_i : i = 1, \dots, n\}$ are the eigenvalues (principal variances) of the posterior covariance matrix of the mode shape, $\mathbf{C}_\Phi$. This formula is asymptotically correct for small $\{\kappa_i\}$ or large n . The sum does not contain κ_1 because it is theoretically equal to zero, resulting from the norm constraint on the mode shape Au 2011.

In well-controlled ambient tests, the Expected MAC is often close to 1. A complementary measure of the mode shape uncertainty is the expected value of the hyper angle between the random mode shape vector and the most probable mode shape vector. It can be shown that under the same asymptotic conditions in Eq. 37, the expected value of the hyper angle is given by

$$\delta_\Phi = \left(\sum_{i=2}^n \kappa_i^2\right)^{1/2} \quad (38)$$

and it is related to the Expected MAC by

$$\text{Expected MAC} = 1 - \frac{1}{2} \delta_\Phi^2 \quad (39)$$

Bayesian Versus Non-Bayesian Measure

Bayesian Measure

In the context of Bayesian system identification, the spread of the posterior PDF is a direct fundamental quantification of the remaining uncertainty associated with the modal parameters for a given assumed identification model and in the presence of the measured data. Since the posterior PDF is typically unimodal and it can be approximated by a joint Gaussian PDF, the uncertainty of the modal parameters can be quantified by the covariance matrix, which is called the “posterior covariance matrix.” The posterior covariance matrix is the inverse of the Hessian matrix of the NLLF, and it can be calculated for a given set of data. Clearly it depends on the particular set of data.

Non-Bayesian Measure

In non-Bayesian identification methods, the identification results are often given in terms of the “best” or “optimal” value calculated from an accepted algorithm. This is analogous to the MPV calculated in a Bayesian identification method. In a non-Bayesian context, one conventional quantification of the uncertainty of the identified parameters is through the ensemble variance (or more generally, the ensemble covariance matrix) of the modal parameters. This can be understood as the variance among the identified (i.e., best) values calculated from a large (theoretically infinite) number of repeated experiments. This is a “frequentist” measure of the uncertainty of the identified parameters. Unlike the posterior variance that is defined for a given set of data, the ensemble variance requires (at least conceptually) the notion of repeated experiments. Because of its ensemble nature, it does not depend on a particular data set. If there were no modeling error, i.e., the data in the repeated experiments indeed results from a process obeying the identification model assumed, the variability of the identified values among repeated experiments reflects the quality of the identification algorithm and the amount of data used. Otherwise, it reflects also the variability due to other factors such as the changes in system properties among the experiments and modeling errors.

Connection

Due to their different definitions, the Bayesian and frequentist measure of the identification uncertainty of modal parameters need not coincide, although intuition suggests that they should be of similar order of magnitude when there is no modeling error. It can be shown that when there is no modeling error then in a weighted sense, the expectation of the posterior covariance matrix (Bayesian) is equal to the ensemble variance of the most probable value among repeated experiments (Au 2012b). In reality where modeling error can exist and conditions can change among the repeated experiments, the Bayesian and frequentist measure can differ significantly.

The Bayesian and frequentist measure of identification uncertainty has different roles. The Bayesian measure reflects fundamentally the remaining uncertainty associated with the parameters for a given identification model and a given set of data. It does not reflect what the identification results would be if the experiment is repeated, because modeling error can exist and conditions of the next experiment can change. It does not describe ensemble variability. On the other hand, the frequentist measure is an aggregate effect of identification uncertainty, modeling error, and system changes over the repeated experiments.

Uncertainty Laws

Under the Bayesian framework, the posterior uncertainty of the modal parameters can be calculated in terms of the posterior covariance for a given set of measured data. This process is computational and implicit in nature. It does not yield much insight about how the identification uncertainty depends on test configurations.

In the case of a single mode in the selected band, it has been possible to derive closed-form analytical expressions for the posterior covariance matrix under asymptotic conditions, namely, small damping and long data duration, which are typically met in applications. The expressions are collectively referred as “uncertainty laws” (Au 2014a, b). They provide insights on the scientific nature of the ambient modal identification problem and its fundamental limits on identification uncertainty. The uncertainty laws can also be used for drafting specifications for ambient vibration tests.

Context

The uncertainty laws have been derived under the following context:

- (1) The damping ratio is small, $\zeta \rightarrow 0$.
- (2) The spectral information for identifying the mode is large, in the sense that $N_f = 2\kappa\zeta N_c \rightarrow \infty$ where $N_c = T_d/T$, T_d is

the data duration, and T is the natural period of the target mode. Note that $N_f \rightarrow \infty$ and $\zeta \rightarrow 0$ imply that $N_c \rightarrow \infty$ as well.

- (3) The selected frequency band is assumed to be $f(1 \pm \kappa\zeta)$, where κ is called the “bandwidth factor.” The bandwidth factor is a trade-off between identification information (the larger, the better) and modeling error risk (the smaller, the better). Typically it ranges between 3 and 10.
- (4) The FFT data $\{F_k\}$ in the selected band is assumed to indeed result from a process obeying the identification model, i.e., no modeling error. Note that prediction error is still present because it is part of the identification model.

The selected frequency bandwidth is assumed to be proportional to the damping ratio because the width of the resonance band around the natural frequency does, e.g., the half power band is $f(1 \pm \zeta)$, corresponding to $\kappa = 1$. An asymptotically small damping ratio implies that the modal s/n ratio $\gamma = S/4S_e\zeta^2$ in Eq. 30 is asymptotically high. It is therefore implied in the uncertainty laws that the modal s/n ratio is high.

Main Results

The asymptotic expressions of the posterior coefficient of variation of the modal parameters are summarized in Table 1. In terms of correlation, except for that between the damping ratio ζ and the modal force PSD S , the correlation between any pair of the modal parameters among $\{f, \zeta, S, S_e, \Phi\}$ is asymptotically small, at most $O(\zeta)$. The correlation between ζ and S is $O(\kappa^{-1/2})$.

Implications

The following implications can be drawn from the uncertainty laws:

- (1) For small damping ratio, the posterior uncertainties of the natural frequency and mode shape are significantly smaller than those of the damping ratio and the modal force PSD. The former are $O(\zeta^{1/2})$ and the latter are $O(\zeta^{-1/2})$.

Bayesian Operational Modal Analysis, Table 1 Summary of uncertainty laws

Parameter	Uncertainty law	Bandwidth factor
Frequency f	Squared posterior c.o.v. $\delta_f^2 = \frac{\zeta}{2\pi N_c B_f(\kappa)}$	$B_f(\kappa) = \frac{2}{\pi} \left(\tan^{-1} \kappa - \frac{\kappa}{\kappa^2 + 1} \right)$
Damping ζ	Squared posterior c.o.v. $\delta_\zeta^2 = \frac{1}{2\pi \zeta N_c B_\zeta(\kappa)}$	$B_\zeta(\kappa) = \frac{2}{\pi} \left[\tan^{-1} \kappa + \frac{\kappa}{\kappa^2 + 1} - \frac{2(\tan^{-1} \kappa)^2}{\kappa} \right]$
Modal force PSD S	Squared posterior c.o.v. $\delta_S^2 = \frac{1}{N_f B_S(\kappa)}$	$B_S(\kappa) = 1 - \frac{2}{\kappa} (\tan^{-1} \kappa)^2 \left(\tan^{-1} \kappa + \frac{\kappa}{\kappa^2 + 1} \right)^{-1}$
Prediction error PSD S_e	Squared posterior c.o.v. $\delta_{S_e}^2 = \frac{1}{(n-1)N_f}$	—
Mode shape Φ	Covariance matrix $\mathbf{C}_\Phi = \frac{v_\zeta^2}{N_c B_\Phi(\kappa)} (\mathbf{I}_n - \Phi \Phi^T)$ Expected hyper angle (squared) $\delta_\Phi^2 = \frac{(n-1)v_\zeta^2}{N_c \tan^{-1} \kappa}$	$B_\Phi(\kappa) = \tan^{-1} \kappa$

- (2) The damping ratio has the greatest uncertainty in terms of posterior c.o.v. and is therefore expected to govern the accuracy requirement in ambient vibration tests.
- (3) The mode shape is the only modal property whose leading order uncertainty is affected by the modal s/n ratio. Its posterior uncertainty is typically small. This implies that in reality the quality of the mode shape is likely to be governed by other sources such as sensor alignment error.
- (4) When the modal s/n ratio is high, reducing the channel noise level (e.g., with better equipment) or increasing the measured dofs has little or no effect on reducing the posterior c.o.v.s of modal parameters. It is because these factors influence only the modal s/n ratio, but the latter has no role in the uncertainty laws.
- (5) Assuming a damping ratio of 1 % and a bandwidth factor of 6, the minimum data duration to achieve a posterior of c.o.v. of 30 % in the damping ratio is 300 natural periods.

Summary

A Bayesian approach for modal identification provides a fundamental means for processing the information contained in the data to make inference on the modal parameters consistent

with probability logic and modeling assumptions. This is especially relevant for ambient modal identification where the input excitation is not measured and where significant uncertainty can exist in the identified modal parameters even in the presence of data. A frequency domain approach is preferred over a time-domain approach because it significantly relaxes the broadband assumption on the modal force and prediction error. By making inference using the FFT on a selected frequency band around the mode of interest, the modal force and prediction error only need to have a flat spectrum within the band. The likelihood function can be formulated from first principles using random-vibration theory and signal processing asymptotics. The resulting form, however, does not allow the posterior statistics of the modal parameters to be computed easily because its dependence on the modal parameters is complicated and there are typically a moderate to large number of parameters. Using linear algebra techniques, fast equivalent formulations have been developed that allow efficient algorithms to be developed. The Bayesian identification results are governed characteristically by the modal signal-to-noise ratio. The posterior covariance matrix also admits remarkably simple closed form when the s/n ratio is high, leading to a set of uncertainty laws that govern the limit of ambient modal identification.

Cross-References

- [Ambient Vibration Testing of Cultural Heritage Structures](#)
- [Building Monitoring Using a Ground-Based Radar](#)
- [Modal Analysis](#)
- [Operational Modal Analysis in Civil Engineering: An Overview](#)
- [Uncertainty Quantification in Structural Health Monitoring](#)

References

- Au SK (2011) Fast Bayesian FFT method for ambient modal identification with separated modes. *J Eng Mech-ASCE* 137(3):214–226
- Au SK (2012a) Fast Bayesian ambient modal identification in the frequency domain, Part I: posterior most probable value. *Mech Syst Signal Process* 26(1):60–75
- Au SK (2012b) Connecting Bayesian and frequentist quantification of parameter uncertainty in system Identification. *Mech Syst Signal Process* 29:328–342
- Au SK (2014a) Uncertainty law in ambient modal identification. Part I: theory. *Mech Syst Signal Process* 48(1–2):15–33
- Au SK (2014b) Uncertainty law in ambient modal identification. Part II: implication and field verification. *Mech Syst Signal Process* 48(1–2):34–48
- Au SK, To P (2012) Full-scale validation of dynamic wind load on a super-tall building under strong wind. *J Struct Eng ASCE* 138(9):1161–1172
- Au SK, Zhang FL (2012a) Fast Bayesian ambient modal identification incorporating multiple setups. *J Eng Mech* 138(7):800–815
- Au SK, Zhang FL (2012b) Ambient modal identification of a primary-secondary structure using fast Bayesian FFT approach. *Mech Syst Signal Process* 28:280–296
- Au SK, Ni YC, Zhang FL, Lam HF (2012a) Full scale dynamic testing of a coupled slab system. *Eng Struct* 37:167–178
- Au SK, Zhang FL, To P (2012b) Field observations on modal properties of two tall buildings under strong wind. *J Wind Eng Ind Aerodyn* 101:12–23
- Au SK, Zhang FL, Ni YC (2013) Bayesian operational modal analysis: theory, computation, practice. *Comput Struct* 126:3–14
- Brillinger DR (1981) *Time series: data analysis and theory*. Holden-Day, San Francisco
- Fan YL, Xiao Y, Kunnath SK (2012) Modal parameter identification of a rigid frame-continuous girders bridge under ambient excitation by fast bayesian FFT method. *Adv Civil Infrastruct Eng* 639–640:985–991
- Lutes DL, Sarkani S (1997) *Stochastic analysis of structural and mechanical vibrations*. Prentice Hall, Upper Saddle River
- Yuen KV (1999) Structural modal updating using ambient dynamic data. Master of philosophy thesis, Department of Civil Engineering, Hong Kong University of Science and Technology, Hong Kong
- Yuen KV, Katafygiotis LS (2001a) Bayesian time-domain approach for modal updating using ambient data. *Probab Eng Mech* 16(3):219–231
- Yuen KV, Katafygiotis LS (2001b) Bayesian spectral density approach for modal updating using ambient data. *Earthq Eng Struct Dyn* 30:1103–1123
- Yuen KV, Katafygiotis LS (2003) Bayesian fast fourier transform approach for modal updating using ambient data. *Adv Struct Eng* 6(2):81–95
- Yuen KV, Kuok SC (2010a) Modeling of environmental influence in structural health assessment for reinforced concrete buildings. *Earthq Eng Eng Vib* 9(2):295–306
- Yuen KV, Kuok SC (2010b) Ambient interference in long-term monitoring of buildings. *Eng Struct* 32(8):2379–2386
- Yuen KV, Kuok SC (2011) Bayesian methods for updating dynamic models. *Appl Mech Rev* 64(1):010802-1–010802-18
- Zhang FL, Au SK (2013) Erratum to “fast Bayesian FFT method for ambient modal identification with separated modes” by Siu-Kui Au. *J Eng Mech, ASCE* 139(4):545–545
- Zhang FL, Au SK, Lam HF (2015) Assessing uncertainty in operational modal analysis incorporating multiple setups using a Bayesian approach. *Struct Control Health Monit* 22(3):395–416

Bayesian Statistics: Applications to Earthquake Engineering

Alin C. Radu and Mircea Grigoriu
Civil and Environmental Engineering, School of
Civil & Environmental Engineering, Cornell
University, Ithaca, NY, USA

Synonyms

Bayesian statistics; Earthquake probabilities; Fragility analysis; Ground motion simulation; Seismic activity matrix; Seismic hazard; Seismological model

Introduction

Bayesian statistics are related to Bayes' rule after Thomas Bayes, a Presbyterian minister, probably born in 1701. Thomas Bayes' fame rests in his paper "An Essay Towards Solving a Problem in the Doctrine of Chances" communicated to the Royal Society in 1763, after his death in 1761, by his friend Richard Price (1723–1791) (Bellhouse 2004). Bayes addresses the following question in his essay: "Given the number of times in which an unknown event has happened and failed: Required the chance that the probability of its happening in a single trial lies somewhere between any two degrees of probability that can be named (Bayes (1763), p. 376)" (Dale 1986; Edwards 1978). The question is answered in its famous Bayes' model table problem (Bellhouse 2004).

Gillies (1987) calls this problem "Bayes' billiard table example." The problem assumes a unit-square table and two balls O and W. The ball W is thrown across the table once, and its rest point defines level p with respect to one of the sides of the table. The ball O is then thrown n times, and it is said that if ball O stops below level p , the event is a success. If we call the number of successes r , then the probability that ball O will stop between levels a and b is given by

$$\frac{\int_a^b x^r (1-x)^{n-r} dx}{\int_0^1 x^r (1-x)^{n-r} dx}.$$

As Bayes' rule becomes popular in many fields of science and engineering, earthquake engineering problems began to be solved in the Bayesian framework. One of the first Bayesian approaches on site seismicity is proposed in Esteva (1969) for the rate of seismic event occurrence and later in Suzuki and Kiremidjian (1991). The Bayesian framework is also used in other seismology-related areas to update ground motion prediction models' parameters to new data in Wang and Takada (2009) and Atkinson and Boore (2011). Earthquake engineering also deals with the probabilistic assessment of the structural response

under seismic loads. The probability of structural response to exceed critical states is known as seismic fragility. Early papers by Goodman (1986) and Mosleh and Apostolakis (1986) deal with failure and fragility analyses in a Bayesian context. Recently, the Bayesian framework has been used in a wide range of applications for calculating fragility curves (Straub and Kiureghian 2008; Koutsourelakis 2010).

Bayes' rule stays at the basis of the Bayesian framework and is a statement of conditional probability. It provides information on the probability of unknown parameters $\Theta = [\theta_1, \dots, \theta_n]$ given some data x ,

$$p(\Theta|x) \propto f(\Theta)l(\Theta|x), \quad (1)$$

where $f(\Theta)$ and $p(\Theta|x)$ are referred to as the *prior* and *posterior* probability density functions of Θ and $l(x|\Theta)$ is called the *likelihood* function, which accounts for the significance of the observed data x on the distribution of Θ (Gelman et al. 2003). The prior probability density function $f(\Theta)$ should contain all possible values of Θ . The probability associated with the values of Θ , i.e., $f(\Theta)$, can be interpreted as the current knowledge about Θ . When all the values of the unknown parameters Θ are assumed to be equally likely, we call the prior *non-informative*. If the prior and the posterior densities have the same functional form for the probability distributions, we refer to the prior as a *conjugate prior*.

Sampling from the posterior distribution of Θ may be challenging. Direct sampling from $p(\Theta|x)$ in the case of non-conjugate prior densities is even more difficult, and sampling techniques have been developed. Markov Chain Monte Carlo (MCMC) is a general algorithm of drawing samples of Θ from a random distribution $p(\Theta|x)$.

Applications of the Bayesian theory in earthquake engineering are shown in the next section. This material is based upon work supported by the National Science Foundation under Grants No. CMMI-0925714 and No. CMMI-1265511. This support is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this material are those

of the authors and do not necessarily reflect the views of the National Science Foundation.

Bayesian Framework in Earthquake Engineering

Seismic Hazard Analysis

Earthquake Probabilities

A seismic activity matrix is defined as a two-dimensional histogram which gives the probability of occurrence p_j of earthquakes characterized by moment magnitude m and source-to-site distance r , denoted by $(m, r)_j$. Each point in this matrix with coordinates $(m, r)_j$ is referred to as a cell $j = 1, \dots, n_c$, where n_c is the total number of cells. Note that $\sum_{j=1}^{n_c} p_j = 1$. Seismic activity matrices can be calculated for each zip code in the United States by using the earthquake probability tool available on the US Geological Survey website (USGS 2009). Figure 1 shows the seismic activity matrix for Los Angeles.

A Bayesian approach is used to update earthquake probabilities in light of new data. For this, we seek a probabilistic parametric model with parameters, the earthquake probabilities. Let Z_j be a random variable which counts the number of new earthquakes recorded in each cell j . The vector $\mathbf{Z} = \{Z_j, j = 1, \dots, n_c\}$ has a

multinomial distribution with parameters $\Theta = \{\theta_j, j = 1, \dots, n_c\}$, a random vector whose coordinates θ_j capture earthquake probabilities in each cell j , such that $\sum_{j=1}^{n_c} \theta_j = 1$. We call $N_Z = \sum_{j=1}^{n_c} Z_j$ the total number of newly recorded earthquakes. The new records are used in the Bayesian framework to update earthquake probabilities p_j .

Numerical example. For illustration, a hypothetical example is considered. It is assumed that at a site only three earthquakes $j = 1, 2, 3$ are likely to occur with known probabilities $p_1 = 0.2, p_2 = 0.3, p_3 = 0.5$. The seismic activity matrix, shown in Fig. 3a, has $n_c = 3$ cells. If an earthquake of type $j = 1$ occurs at that site and it is used for the update of the earthquake probabilities p_1, p_2, p_3 , then $N_Z = 1$ and $\mathbf{z} = [z_1 \ z_2 \ z_3] = [1 \ 0 \ 0]$, which is a sample of $\mathbf{Z}$.

The unknown parameters $\{\theta_1, \dots, \theta_{n_c}\}$ are assumed to have a Dirichlet distribution with parameters $\{\alpha_1, \dots, \alpha_{n_c}\}$, that is,

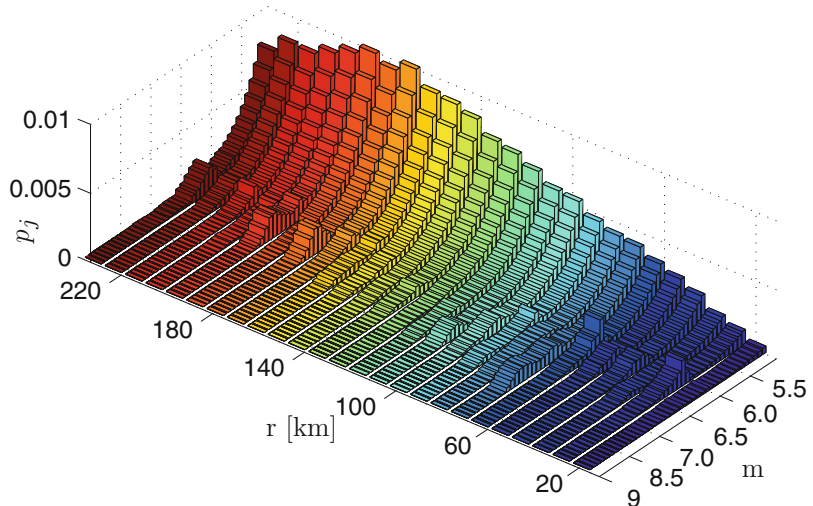
$$f(\theta_1, \dots, \theta_{n_c}) \propto \prod_{j=1}^{n_c} \theta_j^{\alpha_j-1}, \forall \theta_j \geq 0. \quad (2)$$

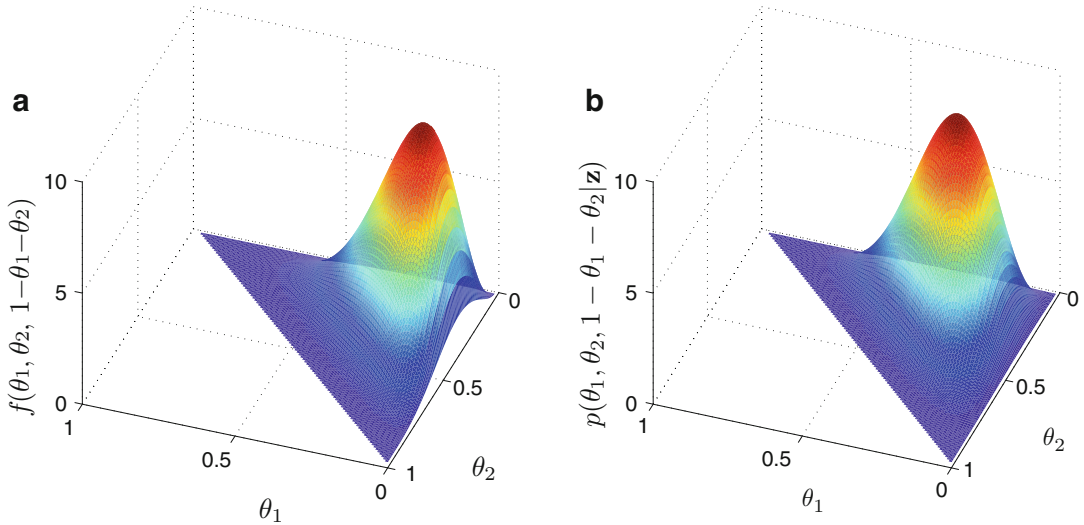
Moreover, we impose that

$$\mathbb{E}[\theta_j] = \frac{\alpha_j}{\alpha_0} = p_j, \quad (3)$$

Bayesian Statistics: Applications to Earthquake Engineering,

Fig. 1 Seismic activity matrix for Los Angeles





Bayesian Statistics: Applications to Earthquake Engineering, Fig. 2 (a) Prior density $f(\theta_1, \theta_2, 1 - \theta_1 - \theta_2)$ and (b) posterior density $p(\theta_1, \theta_2, 1 - \theta_1 - \theta_2 | \mathbf{z})$

in order to reflect the prior knowledge from the seismic activity matrix, where $\alpha_0 = \sum_{j=1}^{n_c} \alpha_j$. For a given value of α_0 , parameters $\alpha_j = p_j \alpha_0$ have deterministic values, and they are known as hyper-parameters. The value α_0 can be interpreted as the total number of recorded earthquakes to have occurred at the site until the new earthquakes in $\mathbf{Z}$ were recorded.

Based on the assumption that $\mathbf{Z}$ has a multinomial distribution, the likelihood function is written as

$$l(\mathbf{z} | \Theta) = \prod_{j=1}^{n_c} \theta_j^{z_j}. \quad (4)$$

The posterior probability density of the unknown parameters Θ is calculated as shown in Eq. 1:

$$p(\Theta | \mathbf{z}) = \prod_{j=1}^{n_c} \theta_j^{\alpha_j + z_j - 1}, \quad (5)$$

which is also a Dirichlet distribution with parameters $\alpha_j + z_j, j = 1, \dots, n_c$. Note that the Dirichlet distribution is the conjugate prior of the multinomial distribution. The seismic activity matrix in light of new data is, then, given by the expectation of Θ :

$$\mathbb{E}[\theta_j] = \frac{\alpha_j p_j + z_j}{\alpha_0 + n_z}. \quad (6)$$

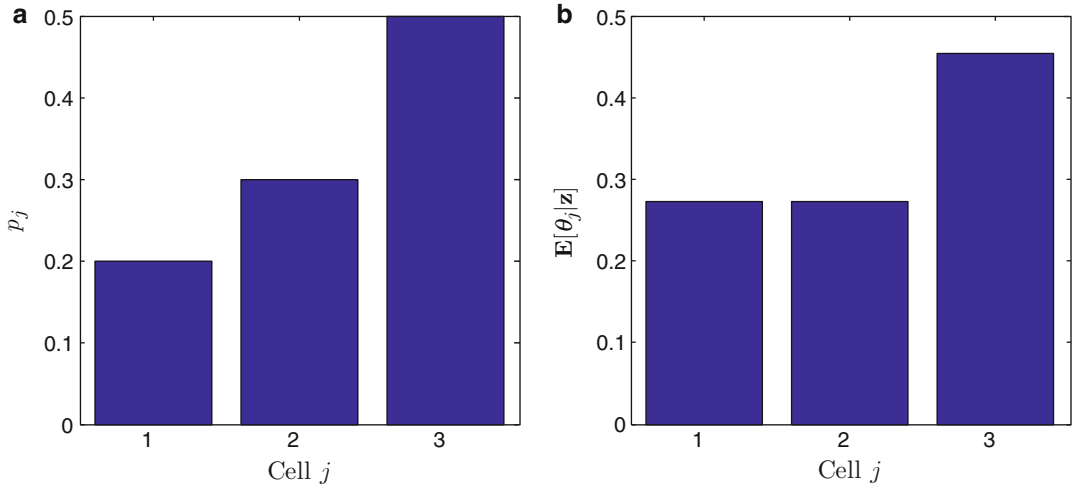
For the hypothetical case presented above, the prior density $f(\theta_1, \theta_2, \theta_3)$ and the posterior density $p(\theta_1, \theta_2, \theta_3 | \mathbf{z})$ are shown in Fig. 2 in the form of $f(\theta_1, \theta_2, 1 - \theta_1 - \theta_2)$ and $p(\theta_1, \theta_2, 1 - \theta_1 - \theta_2 | \mathbf{z})$, respectively, since $\theta_1 + \theta_2 + \theta_3 = 1$. The results are shown for $\alpha_0 = 10$.

The updated seismic activity matrix with the new data $\mathbf{z}$, i.e., $\mathbb{E}[\Theta | \mathbf{z}]$, is shown in Fig. 3b.

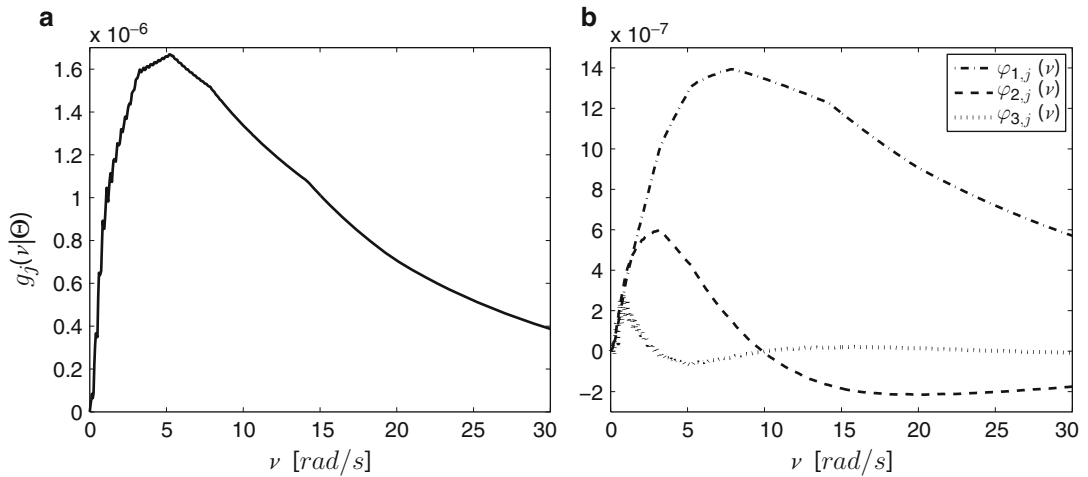
Seismological Model Update

The specific barrier model (Halldorsson and Papageorgiou 2005) is a source model which characterizes the frequency content of the ground motion process in the form of a function $g(v; (m, r)_j)$ of $(m, r)_j$, called power spectral density. The specific barrier model is calibrated to regional data and cannot provide information about ground motions at particular sites. For simplicity, the following notation $g_j(v) := g(v; (m, r)_j)$ is used.

A Bayesian framework is used to update the spectral density model with site-specific records (Radu and Grigoriu 2014). The power spectral density function $g_j(v)$ is only available in



Bayesian Statistics: Applications to Earthquake Engineering, Fig. 3 Two-dimensional seismic activity matrix: (a) original earthquake probabilities $p_j, j = 1, 2, 3$, and (b) updated earthquake probabilities $\mathbb{E}[\theta_j|\mathbf{z}], j = 1, 2, 3$



Bayesian Statistics: Applications to Earthquake Engineering, Fig. 4 (a) Power spectral density $g_j(v)$ and (b) linear components $\varphi_{k,j}, k = 1, 2, 3$, for $(m, r)_j = (8, 200 \text{ km})$

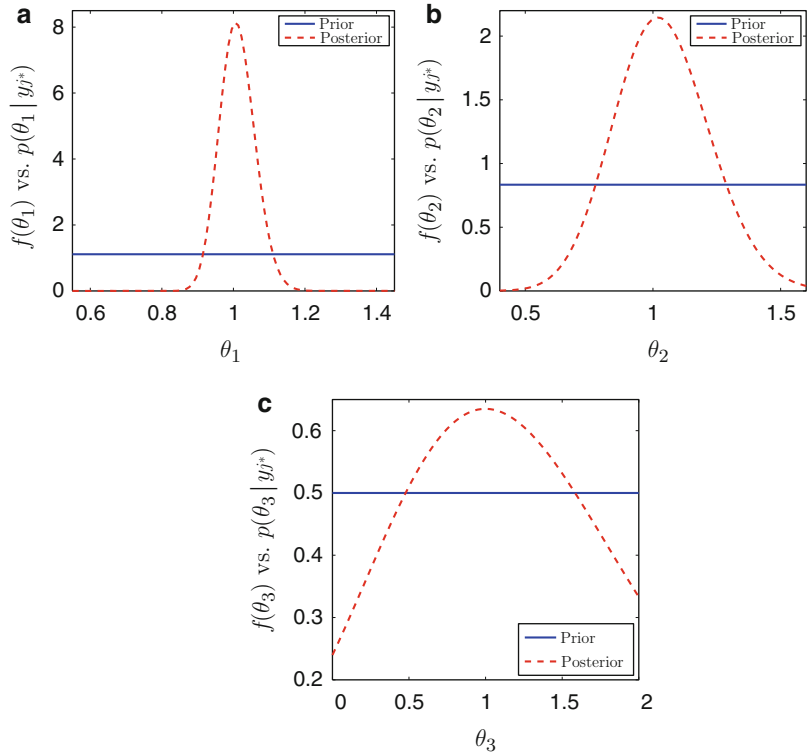
algorithmic form. For the Bayesian approach, a parametric model for $g_j(v)$ is sought, and it is written as

$$g_j(v : \Theta) = \theta_1 \varphi_{1,j} + \theta_2 \varphi_{2,j} + \theta_3 \varphi_{3,j} \quad (7)$$

using the singular value decomposition, where $\{\theta_k = 1, k = 1, 2, 3\}$ are some global parameters and $\varphi_{k,j}, k = 1, 2, 3$, are deterministic

functions which depend on $(m, r)_j$. In the Bayesian context, the vector $\Theta = \{\theta_k, k = 1, 2, 3\}$ is seen as a random vector with mean $\mathbb{E}[\Theta] = [111]$. The range Ψ of Θ is defined such that it ensures the existence of the spectral density, i.e., $g_j(v) \geq 0, \forall_j = 1, \dots, n_c, \forall v \geq 0$, and $\forall \theta \in \Psi$. A uniform prior $f(\Theta)$ on Ψ is assumed for the vector of unknown parameters Θ . Figure 4 shows the spectral density $g_j(v)$ for $(m, r)_j = (8, 200 \text{ km})$ and its components $\varphi_{k,j}, k = 1, 2, 3$.

**Bayesian Statistics:
Applications to
Earthquake Engineering,**
Fig. 5 Marginal prior and
posterior densities for (a)
 θ_1 , (b) θ_2 , and (c) θ_3



Let $\mathbf{x} = \{x_{j,i}, i = 1, \dots, n_j, j = 1, \dots, n_c\}$ be a set of $n = \sum_{j=1}^{n_c} n_j$ ground motion records available at a site, where n_j is the number of earthquakes with $(m, r)_j$ and n_c is the number of cells in the seismic activity matrix in Fig. 1. It is assumed that $x_{j,i}$ are the discrete versions of the records' strong motion parts, i.e., the part of the earthquake where most of its energy is concentrated (Trifunac 1971). Under the assumption that $x_{j,i}$ are samples of a zero-mean, stationary, Gaussian process (Zerva 2009), Chap. 7; Rezaeian and Kiureghian 2010), the logarithmic form of the likelihood function is written as

$$\ln(l(\mathbf{x}|\Theta)) \propto -\frac{n}{2} \ln(|C_j(\Theta)|) - \frac{1}{2} \sum_{i=1}^{n_c} \sum_{j=1}^{n_j} x_{j,i}^T C_j(\Theta)^{-1} x_{j,i}, \quad (8)$$

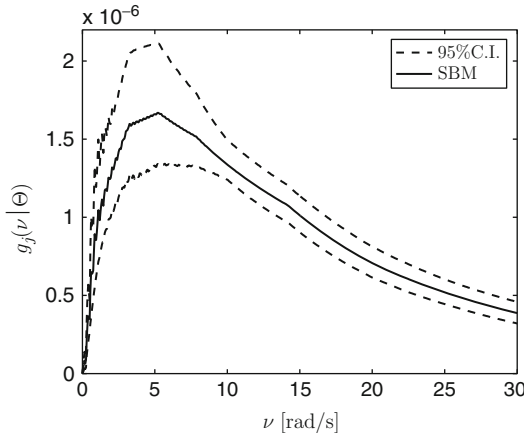
where $\ln$ denotes the natural logarithm and $|C_j(\Theta)|$ is the determinant of the covariance

matrix $C_j(\Theta)$ of the ground motion process in cell j . The components of the covariance matrix can be calculated from the Fourier transform of the spectral density function $g_j(v; \Theta)$ (Grigoriu 2012, Section 3.6).

The posterior probability density function $p(\Theta | \mathbf{x})$ follows directly from Eq. 1.

Numerical example. For numerical results, $n = 5,000$ samples of zero-mean, stationary, Gaussian processes with spectral densities $g_j(v)$, $j = 1, \dots, n_c$, were simulated in cells j according to the multinomial distribution given by the seismic activity matrix of Los Angeles (Fig. 1). The marginal prior and posterior densities of the unknown parameters θ_i , $i = 1, 2, 3$ are shown in Fig. 5.

The resulting posterior 95 % confidence interval for the spectral density $g_j(v; \Theta)$ shown in Fig. 6, i.e., the interval $[g_j^{lw}(v), g_j^{up}(v)]$ such that $\mathbb{P}\{g_j(v; \Theta) \in [g_j^{lw}(v), g_j^{up}(v)]\} = 0.95$, is also calculated based on the density $p(\Theta | \mathbf{x})$.



Bayesian Statistics: Applications to Earthquake Engineering, Fig. 6 Ninety-five percent confidence interval (CI) for the spectral density $g_j(\nu; \Theta)$, $j = 1, 2, 3$

Seismic Fragility Analysis

Seismic fragility of a structural or nonstructural system is the probability that the system's response exceeds a critical value under seismic ground motions of specified intensities, e.g., peak ground acceleration (PGA) or pseudo-spectral acceleration (PSa) (Kafali and Grigoriu 2010). Graphical representations of seismic fragilities are called fragility curves.

Reliable fragility analyses require the use of large numbers of seismic records, which are not usually available. Thus, Monte Carlo simulations of synthetic seismic motions or scaling of actual records to common intensity measures are among the most popular methods. The Bayesian framework allows calculation of seismic fragility curves using just the data available.

It is assumed that n ground motion records $x_i(t)$, $i = 1, \dots, n$, are available at a site. The response displacement $y_i(t)$ of a system to record $x_i(t)$ is characterized by a second-order differential equation:

$$y_i'' = D(y_i, y_i', x_i), \quad (9)$$

where D can be either a linear or a nonlinear operator. For example, for a linear single-degree-of-freedom system with frequency ω_0 and damping ratio ζ_0 ,

$$D(y_i, y_i', x_i) = -(2\omega_0\zeta_0 y_i' + \omega_0^2 y_i + x_i). \quad (10)$$

Two intensity measures for the ground motion $x_i(t)$ are defined:

1. Peak ground acceleration (PGA)

$$\xi_i = \max_{0 \leq t \leq \tau} |x_i(t)| \quad (11)$$

2. Pseudo-spectral acceleration (PSa)

$$\xi_i(\omega; \zeta) = \max_{0 \leq t \leq \tau} |y_i(t; \omega, \zeta)|, \quad (12)$$

where $y_i(t; \omega, \zeta)$ is the solution of Eq. 10.

For illustration purposes, we present the methodology for calculating fragility curves for the single-degree-of-freedom system in Eq. 10. In this context, we define the seismic fragility as

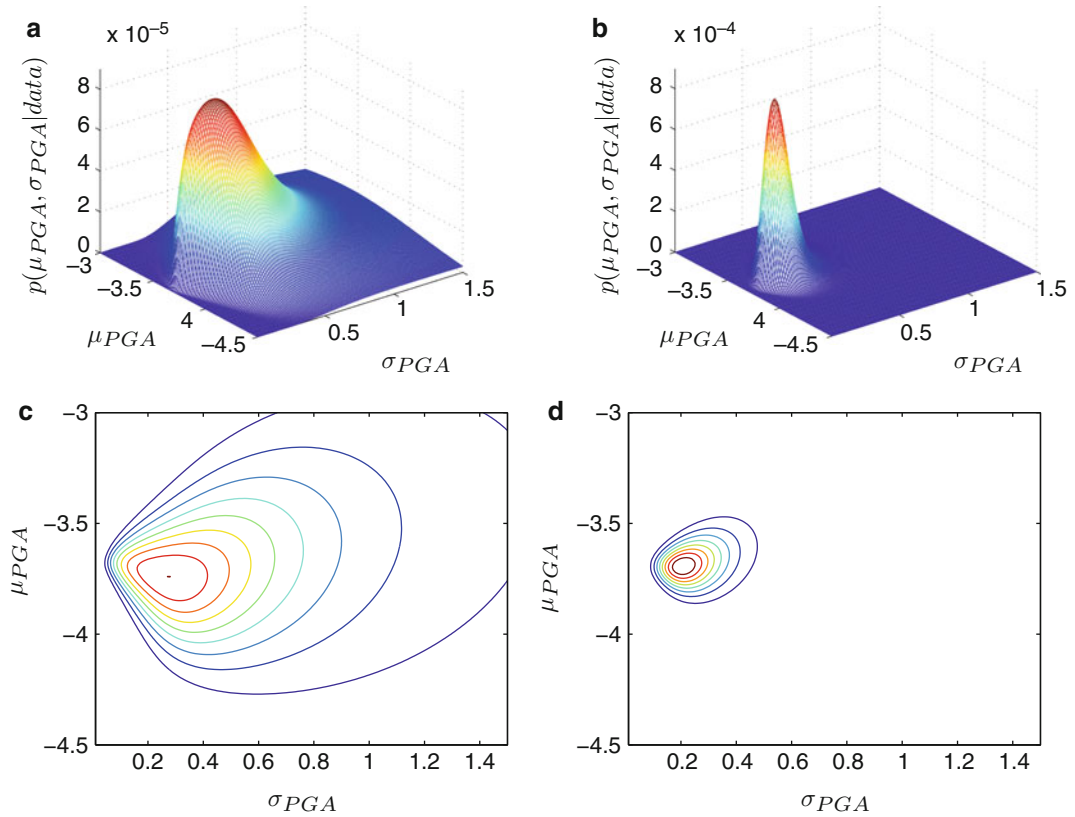
$$P_f(\zeta; y_{cr}) = \mathbb{P} \left(\max_{0 \leq t \leq \tau} |y_i(t; \omega_0, \zeta_0)| > y_{cr} \mid \xi_i = \zeta \right), \quad (13)$$

where y_{cr} is a critical displacement imposed for the system. Fragility curves can be calculated for any other engineering design parameters, e.g., the inter-story drift in the case of multi-degree-of-freedom systems. For simplicity, parameters y_{cr} is dropped from notation $P_f(\zeta; y_{cr})$.

In order to calculate fragility curves in a Bayesian framework, a parametric model for $P_f(\zeta)$ is sought. Fragility curves are increasing functions plotted against positive ground motion intensity measures $\zeta > 0$ and left-bounded at zero.

Therefore, it is common to assume that they are well modeled by log-normal cumulative distribution functions $F(\zeta; \mu, \sigma)$ with mean μ and standard deviation σ . The log-normal cumulative distribution function is defined as

$$F(\zeta; \mu, \sigma) = \int_0^{\zeta} \phi_{LN}(x; \mu, \sigma) dx, \quad (14)$$



Bayesian Statistics: Applications to Earthquake Engineering, Fig. 7 Three-dimensional and corresponding contour plots for the posterior density functions of $(\mu_{PGA}, \sigma_{PGA})$, for (a), (c) $n = 10$ and (b), (d) $n = 50$ records

where

$$s_i = 1\{y_i(t) > y_{cr}\}, \quad (16)$$

$$\phi_{LN}(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left\{-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right\}, x > 0, \quad (15)$$

and $\phi_{LN}(x; \mu, \sigma)$ is the probability density function of the log-normal distribution with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma \geq 0$.

The model for the fragility curve in Eq. 13 is $P_f(\xi; \Theta) = F(\xi; \mu, \sigma)$, where $\Theta = (\mu, \sigma)$ is the vector of unknown parameters to be estimated from the data. Parameters (μ, σ) define uniquely the fragility curve.

A non-informative uniform prior $f(\Theta)$ is assumed for $\Theta = (\mu, \sigma)$ in a custom range within previously defined bounds, i.e., $\mu \in \mathbb{R}, \sigma \geq 0$.

Let $s = \{s_i, i = 1, \dots, n\}$ be a vector with values:

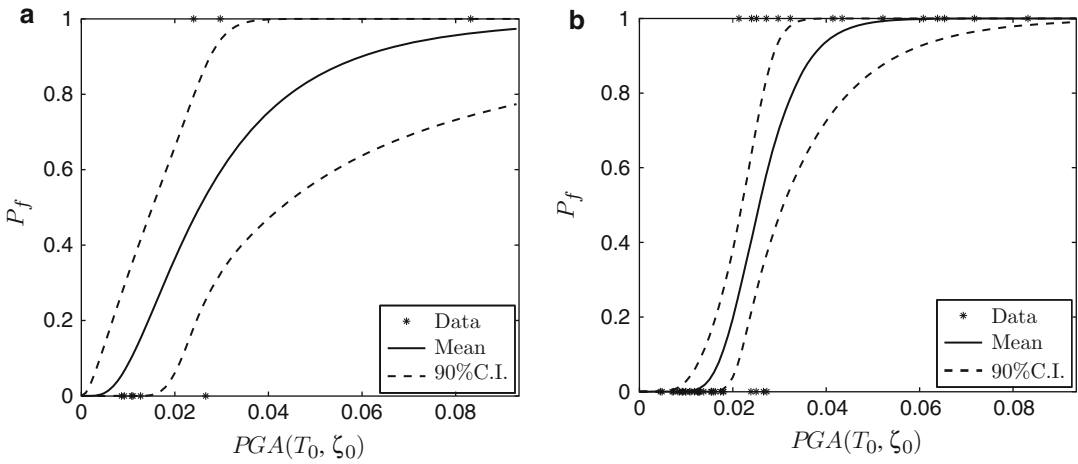
where 1 denotes the indicator function. The vector s is a sample of a binomial random vector with parameters $(n, \{F(\xi_i; \mu, \sigma), i = 1, \dots, n\})$. Then, the likelihood function is defined as

$$l(s|\mu, \sigma) = \prod_{i=1}^n (1 - F(\xi; \mu, \sigma))^{1-s_i} F(\xi; \mu, \sigma)^{s_i}. \quad (17)$$

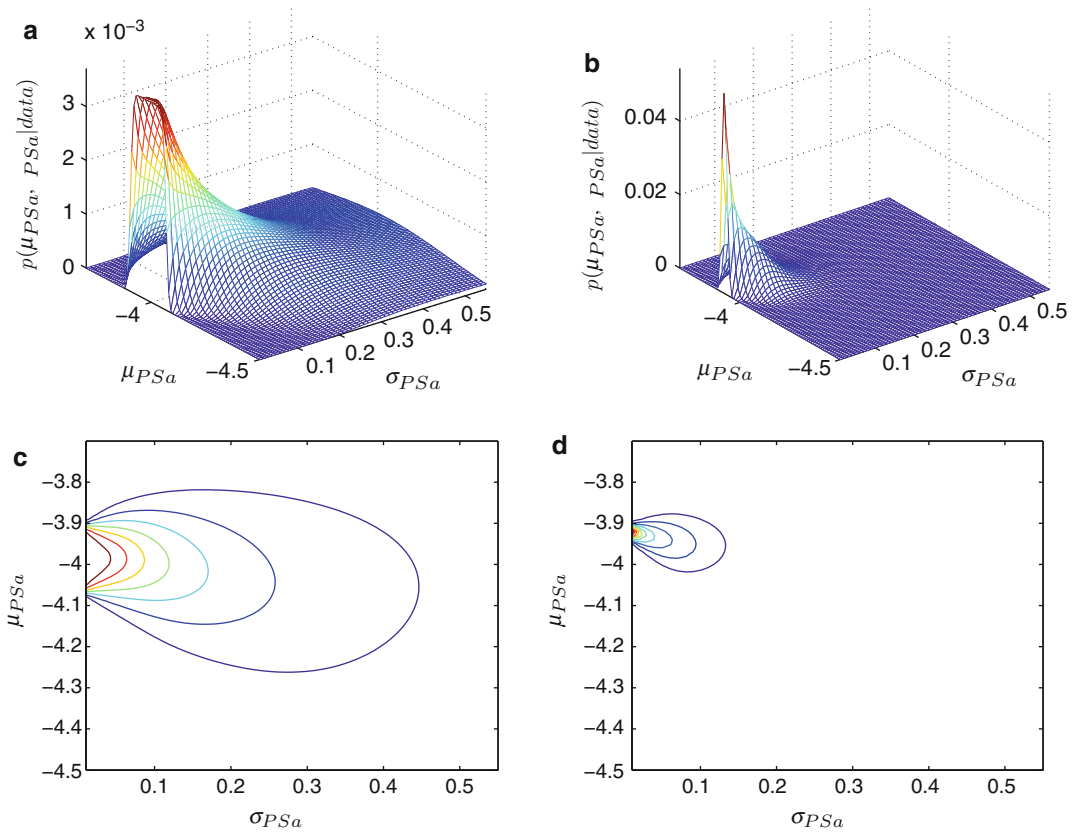
Finally, the posterior density of parameters Θ is

$$p(\mu, \sigma|s) \propto f(\mu, \sigma) l(s|\mu, \sigma). \quad (18)$$

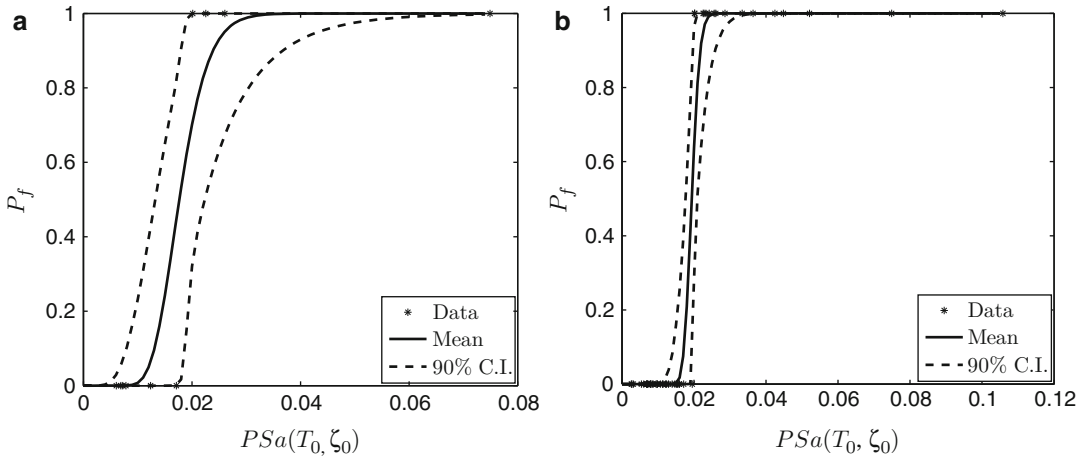
Numerical examples Numerical examples are shown for the linear single-degree-of-freedom system in Eq. 10 with $\omega_0 = \pi \text{ rad/s}$, $\zeta_0 = 2\%$,



Bayesian Statistics: Applications to Earthquake Engineering, Fig. 8 Fragility curves (mean) $P_f(PGA)$ and 95 % confidence intervals (CIs) for (a) $n = 10$ and (b) $n = 50$ records



Bayesian Statistics: Applications to Earthquake Engineering, Fig. 9 Three-dimensional and corresponding contour plots for the posterior density functions of $(\mu_{PSa}, \sigma_{PSa})$, for (a), (c) $n = 10$ and (b), (d) $n = 50$ records



Bayesian Statistics: Applications to Earthquake Engineering, Fig. 10 Fragility curves (mean) $Pf(PSa)$ and 95 % confidence intervals (CIs) for (a) $n = 10$ and (b) $n = 50$ records

and $y_{cr} = 2$, for two sets of $n = 10$ and $n = 50$ ground motion records, respectively. The ground motion samples used in the fragility analysis were simulated as zero-mean, nonstationary Gaussian processes with probability density functions $g_j(v)$ from the specific barrier model in Halldorsson and Papageorgiou (2005). Samples are generated for various (m, r) , according to the seismic activity matrix of Los Angeles. Fragility curves are constructed for both PGA and PSa intensity measures.

Figures 7 and 9 show the posterior distributions of the unknown parameters $\Theta = (\mu_{PGA}, \sigma_{PGA})$ and $\Theta = (\mu_{PSa}, \sigma_{PSa})$, which define the fragility curves for the PGA and PSa intensity measures, respectively. Plots (a) and (c) of each figure show the posterior of Θ obtained for $n = 10$ records and (b) and (d) for $n = 50$ records, respectively. Fragility curves for PGA and their corresponding 90 % confidence intervals calculated based on the posterior densities of Θ shown in Fig. 7a and c are illustrated in Fig. 8a and b, respectively. Fragility curves and their 90 % confidence intervals for the PSa intensity measure are plotted in Fig. 10 for the $n = 10$ and $n = 50$ ground motion sample sets. The posterior densities are less spread on the domain of Θ as the number of samples n increases. Consequently, the more data is used in the analysis, the narrower the confidence intervals for the fragility curves are.

Summary

The Bayesian framework, widely used in earthquake engineering, produces statistical inferences about uncertain parameters of a probabilistic model by combining observable data with available prior information about the parameters. The entry presents applications of Bayes' theorem for several concepts used in earthquake engineering.

The data is usually composed of ground motion records, which are used to develop probabilistic parametric models for fragility curves or to improve and update existent models on earthquake probabilities and frequency content of seismic motions. The Bayesian framework also facilitates the calculations of various statistics by using limited amounts of data.

Cross-References

- [Analytic Fragility and Limit States \[P\(EDP|IM\)\]: Nonlinear Dynamic Procedures](#)
- [Engineering Characterization of Earthquake Ground Motions](#)
- [Nonlinear Dynamic Seismic Analysis](#)
- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- [Physics-Based Ground-Motion Simulation](#)

- ▶ Probabilistic Seismic Hazard Models
- ▶ Seismic Actions Due to Near-Fault Ground Motion
- ▶ Seismic Fragility Analysis
- ▶ Site Response for Seismic Hazard Assessment
- ▶ Stochastic Analysis of Linear Systems
- ▶ Stochastic Analysis of Nonlinear Systems
- ▶ Stochastic Ground Motion Simulation
- ▶ Time History Seismic Analysis

References

- Atkinson G, Boore D (2011) Modification to existing ground-motion prediction equations in light of new data. *Bull Seismol Soc Am* 101(3):1121–1135
- Bayes T (1763) An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions* 53, p. 370–418. [Reprinted (1958) *Studies in the history of probability and statistics IX*. Thomas Bayes' essay towards solving a problem in the doctrine of chances. *Biometrika* 45, p. 296–315]
- Bellhouse D (2004) The reverend Thomas Bayes, FRS: a biography to celebrate the tercentenary of his birth. *Stat Sci* 19(1):3–43
- Dale A (1986) A newly-discovered result of Thomas Bayes. *Arch Hist Exact Sci* 35(2):101–113
- Edwards A (1978) Commentary on the arguments of Thomas Bayes. *Scand J Stat* 5(2):116–118
- Esteva L (1969) Seismicity prediction: a Bayesian approach. In: *Proceedings of the fourth world conference on earthquake engineering*, Santiago de Chile
- Gelman A, Carlin JB, Stern HS, Rubin DB (2003) *Bayesian data analysis*, 2nd edn, *Texts in statistical sciences*. Chapman Hall CRC, Boca Raton
- Gillies D (1987) Was Bayes a Bayesian? *Hist Math* 14:325–346
- Goodman J (1986) Interval estimates of average failure rate and unavailability per demand. *Reliab Eng* 14:107–121
- Grigoriu M (2012) *Stochastic systems. Uncertainty quantification and propagation*. Springer, London
- Halldorsson B, Papageorgiou A (2005) Calibration of the specific barrier model to earthquake to different tectonic regions. *Bull Seismol Soc Am* 95(4):1276–1300
- Kafali C, Grigoriu M (2010) Seismic fragility analysis: application to simple linear and nonlinear systems. *Earthq Eng Struct Dyn* 36(13):1885–1900
- Koutsourelakis P (2010) Assessing structural vulnerability against earthquakes using multi-dimensional fragility surfaces: a Bayesian framework. *Probab Eng Mech* 25:49–60
- Mosleh A, Apostolakis G (1986) The assessment of probability distributions from expert opinions with an application to seismic fragility curves. *Risk Anal* 6(4):447–461
- Radu A, Grigoriu M (2014) A site-specific seismological model for probabilistic seismic-hazard assessment. *Bull Seismol Soc Am* 104(6):3054–3071
- Rezaeian S, Kiureghian A (2010) Simulation of synthetic ground motions for specified earthquake and site characteristics. *Earthq Eng Struct Dyn* 39(10):1155–1180
- Straub D, Kiureghian A (2008) Improved seismic fragility modeling from empirical data. *Struct Saf* 30:320–336
- Suzuki S, Kiremidjian A (1991) A random slip rate model for earthquake occurrences with Bayesian parameters. *Bull Seismol Soc Am* 81(3):781–795
- Trifunac MD (1971) A method for synthesizing realistic strong ground motions. *Bull Seismol Soc Am* 61(6):1739–1753
- USGS (2009) U.S. geological survey: 2009 earthquake probability mapping. Last checked on 9 Sept 2013. <https://geohazards.usgs.gov/eqprob/2009/index.php>
- Wang M, Takada T (2009) A Bayesian framework for prediction of seismic ground motion. *Bull Seismol Soc Am* 99(4):2348–2364
- Zerva A (2009) Spatial variation of seismic ground motions: modeling and engineering applications. In: *Advances in engineering*. CRC Press – Taylor Francis Group, LLC, 6000 Broken Sound Parkway, Boca Raton, FL

Beginner's Guide to Fragility, Vulnerability, and Risk

Keith Porter

Civil, Environmental, and Architectural Engineering, University of Colorado, Boulder and SPA Risk LLC, Denver, CO, USA

Introduction

This entry provides a primer for earthquake-related fragility, vulnerability, and risk. Many of its concepts can be applied to other perils. Section “**Fragility**” discusses fragility – the probability of an undesirable outcome as a function of excitation. Vulnerability (the measure of loss as a function of excitation) is discussed in section “**Vulnerability**.” Section “**Hazard**” presents enough information about seismic hazard for the reader to understand risk, which is discussed in section “**Risk for a Single Asset**.” Section “**Conclusions**” provides brief conclusions. For solved exercises, see Porter (2014).

Fragility

Uncertain Values

Many of the terms used here involve uncertain quantities, often called random variables. “Uncertain” is used here because it applies to quantities that change unpredictably (e.g., whether a tossed coin will land heads or tails side up on the next toss) and to quantities that do not vary but that are not known with certainty. For example, a particular building’s capacity to resist collapse in an earthquake may not vary much over time, but one does not know that capacity before the building collapses, so it is uncertain. In this entry, uncertain variables are denoted by capital letters, e.g., D ; particular values are denoted by lower case, e.g., d ; probability is denoted by $P[\]$; and conditional probability is denoted by $P[A|B]$, that is, probability that statement A is true given that statement B is true.

Meaning and Form of a Fragility Function

The stage is now set to discuss fragility functions. A fragility function expresses the probability that an undesirable event will occur as a function of the value of some (potentially uncertain) environmental excitation. Let X denote the excitation. A fragility function is usually shown on an x - y chart with probability on the y -axis (bounded by 0 and 1) and excitation X on the x -axis (usually bounded below by 0 and above by infinity). (The variable X is used more than once in this entry. It may have different meanings in different places.)

When the undesirable outcome has to do with damage to a whole building, bridge, or other large facilities, excitation is generally measured in terms of ground motion. Common measures of ground motion are addressed later. When the undesirable outcome has to do with damage to a building component, excitation is often measured in terms of one of the following:

Peak floor acceleration (PFA). Maximum zero-period acceleration in any direction at any time during an earthquake at the base of floor-mounted components or at the soffit of

the slab from which a component is suspended.

Peak floor velocity (PFV). Like PFA, except maximum velocity at the base of the floor-mounted components or at the soffit of the slab from which a component is suspended.

Peak transient interstory drift ratio (PTD). This is the maximum value at any time during seismic excitation of the displacement of the floor above relative to the floor below the story on which a component is installed, divided by the height difference of the two stories. The displacements are commonly measured parallel to a horizontal axis of the component, such as along a column line.

Peak residual drift ratio (PRD). Like PTD, except measures the displacement of the floor above relative to the floor below after the cessation of motion.

Some people use the term fragility curve to mean the same thing as fragility function. Some use fragility and vulnerability interchangeably. This work will not do so and will not use the expression “fragility curve” or “vulnerability curve” at all. A function allows for a relationship between loss and one, two, or more inputs, whereas a curve only allows for one input.

The most common form of a seismic fragility function is the lognormal cumulative distribution function (CDF). It is of the form

$$\begin{aligned} F_d(x) &= P[D \geq d | X = x] \quad d \in \{1, 2, \dots, n_D\} \\ &= \Phi\left(\frac{\ln(x/\theta_d)}{\beta_d}\right) \end{aligned} \quad (1)$$

$P[A/B]$ = probability that A is true given that B is true.

D = uncertain damage state of a particular asset such as a building component or a building. It can take on a value in $\{0, 1, \dots, n_D\}$, where $D = 0$ denotes the undamaged state, $D = 1$ denotes the first damage state, etc.

d = a particular value of D , i.e., with no uncertainty.

n_D = number of possible damage states, $n_D \in \{1, 2, \dots\}$.

X = uncertain excitation, e.g., peak zero-period acceleration at the base of the asset in question. Here excitation is called demand parameter (DP), using the terminology of FEMA P-58 (Applied Technology Council 2012). FEMA P-58 builds upon work coordinated by the Pacific Earthquake Engineering Research (PEER) Center and others. PEER researchers use the term engineering demand parameter (EDP) to mean the same thing. Usually $X \in \{\Re \geq 0\}$ but it does not have to be. Note that $X \in \{\Re \geq 0\}$ means that X is a real, nonnegative number.

x = a particular value of X , i.e., with no uncertainty.

$F_d(x)$ = a fragility function for damage state d evaluated at x .

$\Phi(s)$ = standard normal cumulative distribution function (often called the Gaussian) evaluated at s , e.g., `normsdist(s)` in Excel.

$\ln(s)$ = natural logarithm of s .

θ_d = median capacity of the asset to resist damage state d measured in terms of X . Usually $\theta_d \in \{\Re > 0\}$ but it could in principle have a vector value and perhaps negative values. The subscript d appears because a component can sometimes have $n_D > 1$.

β_d = logarithmic standard deviation of the uncertain capacity of the asset to resist damage state d .

For example, see the PACT fragility database at https://www.atcouncil.org/files/FEMAP-58-3_2_ProvidedFragilityData.zip (Applied Technology Council 2012). See the tab PERFORMANCE DATA, the line marked C3011.002c. It employs the lognormal form to propose two fragility functions for Wall Partition, Type: Gypsum + Ceramic Tile, Full Height, Fixed Below, Slip Track Above w/ returns (friction connection). The demand parameter is "Story Drift Ratio," meaning the time-maximum absolute value of the peak transient drift ratio for the story at which partition occurs. For that component, $n_D = 2$, which occur sequentially, meaning that a component must enter damage state

1 before it can enter damage state 2. Damage state 1 is defined as "minor cracked joints and tile." Damage state 2 is defined as "cracked joints and tile." $\theta_1 = 0.0020$, $\beta_1 = 0.70$, $\theta_2 = 0.0050$, and $\beta_2 = 0.40$. The repair for $D = 1$ is described as "carefully remove cracked tile and grout at cracked joints, install new ceramic tile and re-grout joints for 10 % of full 100 foot length of wall. Existing wall board will remain in place." Repair for $D = 2$ is "install ceramic tile and grout all joints for full 100 foot length of wall. Note: gypsum wall board will also be removed and replaced which means the removal of ceramic tile will be part of the gypsum wall board removal."

The Lognormal Distribution

The lognormal is ubiquitous in probabilistic seismic hazard analysis (PSHA) and probabilistic seismic risk analysis (PSRA). To understand it, consider first the normal (*not* the lognormal) distribution. If a quantity X is normally distributed with mean μ and standard deviation σ , it can take on any scalar value in $-\infty < X < \infty$. Its cumulative distribution function (CDF) can be expressed as follows:

$$P[X \leq x] = \Phi\left(\frac{x - \mu}{\sigma}\right) \quad (2)$$

Note that:

$\mu \in \{\Re\}$, meaning that μ is any real scalar value
 $\sigma \in \{\Re > 0\}$, meaning that σ is any positive scalar value.

One can also find the value x associated with a specified nonexceedance probability, p .

$$x = \mu + \sigma \cdot \Phi^{-1}(p) \quad (3)$$

Now consider to the lognormal distribution. If a variable is lognormally distributed, that means its natural logarithm is normally distributed, which means it must take on a positive real value and the probability of it being zero or negative is zero. One can write the CDF several different, equivalent, ways:

$$\begin{aligned}
 P[X \leq x] &= \Phi\left(\frac{\ln x - \ln \theta}{\beta}\right) \\
 &= \Phi\left(\frac{\ln(x/\theta)}{\beta}\right) \\
 &= \Phi\left(\frac{\ln x - \mu_{\ln X}}{\sigma_{\ln X}}\right)
 \end{aligned} \tag{4}$$

where $\mu_{\ln X}$ denotes the mean value of $\ln(X)$, which is the same as the natural logarithm of the median, $\ln(\theta)$. It is sometimes desirable to calculate θ and β in terms of μ and σ . Here are the conversion equations. (The reader can learn more about probability distributions from various sources, such as the free online source NIST/SEMATECH (2013) or Ang and Tang (1975).) Let v denote the coefficient of variation of X . Then

$$v = \frac{\sigma}{\mu} \tag{5}$$

$$\beta = \sqrt{\ln(1 + v^2)} \tag{6}$$

$$\theta = \frac{m}{\sqrt{1 + v^2}} \tag{7}$$

Note well: there is nothing fundamental about the lognormal distribution that makes it ideal or exact for fragility functions, ground motion, and so on. It is commonly used for several reasons:

1. Simplicity. It only requires two parameter values to completely define.
2. Precedent. The lognormal has been used for fragility functions for decades.
3. Information-theory reasons. Given a median and logarithmic standard deviation of some uncertain positively valued quantity like component capacity, the lognormal is the most uncertain distribution, that is, it assumes the least amount of information.
4. Sometimes it fits the data well.

But the lognormal may fit data badly, sometimes worse than other competing parametric and nonparametric forms. Beware oversimplification,

and never confuse a mathematical simplification or model with reality. Ideally one's model approximates reality, but the model is not the thing itself.

Multiple Fragility Functions

Consider situations where a component, asset, or person can experience multiple possible damage states. This entry considers only discrete damage states, meaning that one can number the damage states $D = 1, 2$, etc., but not 1.5. If a component is damaged, one can number damage states in at least one of three ways. First, let $D = 0$ denote the undamaged state. If the component or asset is damaged, let the damage state be denoted by $D \in \{1, 2, \dots, n_D\}$. The three kinds of fragility functions dealt with here are:

1. Sequential damage states. A component must enter damage state 1 before it can enter damage state 2, and it must enter 2 before 3, and so on.
2. Simultaneous damage states. A damaged component can be in more than one damage state at the same time. Order does not matter.
3. Mutually exclusive and collectively exhaustive (MECE) damage states. A damaged component can be in one and only one damage state. Order does not matter.

There are other ways to express fragility, such as with a vector that combines numbers of various types. For example, one might want to talk about a scalar quantity Q of a component that is in a particular sequential damage state D , such as the fraction of a reinforced concrete shear wall that has cracks of at least 3/8 in. width, which would be indicative of the quantity of steel that has to be replaced. But the present discussion is limited to the three kinds noted above.

Sequential Damage States

In sequential damage states, the damage states are ordered (D is therefore an ordinal number, subject to certain mathematical rules of ordinal numbers), so one can talk about lower damage states

and higher ones. The probability of reaching or exceeding a lower damage state is greater than or

equal to the probability of reaching or exceeding a higher damage state.

$$\begin{aligned}
 P[D = d|X = x] &= 1 - P[D \geq 1|X = x] & d = 0 \\
 &= P[D \geq d|X = x] - P[D \geq d + 1|X = x] & 1 \leq d < n_D \\
 &= P[D \geq d|X = x] & d = n_D
 \end{aligned} \tag{8}$$

The first line is the probability that the component is undamaged. The next is the probability that the component is damaged, but not in damage state n_D , called the maximum damage state. The last line is the probability that the component is damaged in the maximum damage state.

Simultaneous Damage States

With simultaneous damage states, one can evaluate the probability that a component is in each damage state independently of the others. Because order does not matter, damage states D are nominal numbers, like the numbers on football jerseys without any order, although $D = 0$ is reserved for the undamaged state.

$$\begin{aligned}
 P[D = d|X = x] &= 1 - P[D \geq 1|X = x] & d = 0 \\
 &= P[D \geq 1|X = x] \cdot P[D = d|D \geq 1] & 1 \leq d \leq n_D
 \end{aligned} \tag{9}$$

where

$P[D \geq 1|X = x]$ = probability that the component is damaged in some way, which can be quantified just like any fragility function, such as with a lognormal CDF:

$$P[D \geq 1|X = x] = \Phi\left(\frac{\ln(x/\theta)}{\beta}\right) \tag{10}$$

where there are only one value of θ and only one value of β – no subscripts as in Eq. 1 – that is, a single median and logarithmic standard deviation of capacity. Note that the fragility function does not have to be a lognormal CDF, but that is common.

$P[D = d|D \geq 1]$ = probability that, if damaged, it is in damage state d . It can be in others as well.

Since under simultaneous damage states a component can be in more than one damage state,

$$\left(\sum_{d=1}^{n_D} P[D = d|D \geq 1]\right) > 1 \tag{11}$$

How can one estimate the probability that a component is in one and only one damage state? Let d_i denote one particular value of D and d_j another particular value of D , but $d_i \neq d_j$, $d_i \neq 0$, and $d_j \neq 0$. Let $P[D = d_i \& D \neq d_j|X = x]$ denote the probability that the component is in damage state d_i and it is not in any other damage state d_j given that $X = x$. It is given by

$$\begin{aligned}
 P[D = d_i \& D \neq d_j|X = x] &= P[D \geq 1, D = d_i|X = x] \cdot \prod_j (1 - P[D = d_j|D \geq 1, X = x]) \\
 &= P[D \geq 1|X = x] \cdot P[D = d_i|D \geq 1, X = x] \cdot \prod_j (1 - P[D = d_j|D \geq 1, X = x]) \\
 &= P[D \geq 1|X = x] \cdot P[D = d_i|D \geq 1] \cdot \prod_j (1 - P[D = d_j|D \geq 1])
 \end{aligned} \tag{12}$$

Consider now the probability that the component is in exactly two damage states. Let D_1 denote the first of two nonzero damage states that the component is in and D_2 the second. Let $P[D_1 = d_i \& D_2 = d_j \& D \neq d_k | X = x]$ denote the probability

that the component is in two damage states $D_1 = d_i$ and $D_2 = d_j$ but not any other damage state d_k ($k \neq i, k \neq j, i \neq j, i \neq 0, j \neq 0$, and $k \neq 0$) given $X = x$. It is given by

$$\begin{aligned} P[D_1 = d_i \& D_2 = d_j \& D = d_k | X = x] \\ = P[D \geq 1 | X = x] \cdot P[D = d_i | D \geq 1] \cdot P[D = d_j | D \geq 1] \cdot \prod_k (1 - P[D = d_k | D \geq 1]) \end{aligned} \quad (13)$$

One could repeat for three damage states i, j , and k by repeating the pattern. It is the product of the probabilities that the component is in each damage state i, j , and k and the probabilities that it is not in each remaining damage state l, m, n , etc.

MECE Damage States

Remember that MECE means that, if the component is damaged (denoted by $D \geq 1$), it is in one and only one nonzero damage state. One can evaluate it by

$$\begin{aligned} P[D = d | X = x] &= 1 - P[D \geq 1 | X = x] & d = 0 \\ &= P[D \geq 1 | X = x] \cdot P[D = d | D \geq 1] & d \in \{1, 2, \dots, N_D\} \end{aligned} \quad (14)$$

$P[D \geq 1 | X = x]$ = probability that the component is damaged in some way, which one evaluates with a single fragility function. If the fragility function is taken as a lognormal CDF, see Eq. 10. Note that the fragility function does not have to be a lognormal CDF, but that is common.

$P[D = d | D \geq 1]$ = probability that, if damaged, it is damaged in damage state d (and not any other value of D). Since under MECE damage states a component can only be in one damage state,

$$\left(\sum_{d=1}^{N_D} P[D = d | D \geq 1] \right) = 1 \quad (15)$$

Creating Fragility Functions

What to Know Before Trying to Derive a Fragility Function

Consider now how fragility functions are made. Much of this section is drawn from Porter et al. (2007). Before trying to derive a fragility

function, the analyst should define failure in unambiguous terms that do not require the exercise of judgment, i.e., where two people observing the same specimen would reach the same conclusion as to whether a specimen has failed or not. Beware damage scales that do not meet this test. Second, define the excitation to which specimens are subjected (maximum base acceleration, peak transient drift ratio, etc.) in similarly unambiguous terms. Third, select specimens without bias with respect to failure or nonfailure. That is, one cannot use data about specimens that were observed *because* they were damaged or because the damage was prominent or interesting in some way. (Failure data gathered by reconnaissance surveys tend to be biased in this way.) Fourth, ensure that specimens were subjected to multiple levels of excitation.

Actual Failure Excitation

In the unusual case where specimens were all tested in a laboratory to failure and the actual excitation at which each specimen failed is known, then one can fit a lognormal fragility

function to the data as follows. This kind of data is referred to here as type A data, A for actual failure excitation. Before using the following math, the analyst must clearly define “failure” and should know both the means of observing specimen excitation and failure, as well as a clear definition of the component or other asset category in question:

n_i = number of specimens, $n_i \geq 2$.
 i = index to specimens, $i \in \{1, 2, \dots, n_i\}$.
 r_i = excitation at which specimen i failed.

$$\theta = \frac{1}{n_i} \sum_{i=1}^{n_i} \ln(r_i) \quad (16)$$

$$\beta = \sqrt{\frac{1}{n_i - 1} \sum_{i=1}^{n_i} (\ln(r_i/\theta))^2} \quad (17)$$

Bounding-Failure Excitation

Suppose one possesses observations where at least one specimen did not fail, at least one specimen did fail, and one knows the peak excitation to which each specimen was subjected, but not the actual excitation at which each specimen failed. These data are referred to here as bounding, or type B, data. Specimens are grouped by the maximum level of excitation to which each specimen was subjected. Assume the fragility function is reasonably like a lognormal cumulative distribution function and find the parameter values θ (median) and β (logarithmic standard deviation) as follows:

m_i = numbers of levels of excitation among the data, referred to here as bins, $m_i \geq 2$.
 i = bin index, $i \in \{1, 2, \dots, m_i\}$.
 r_i = maximum excitation to which specimens in bin i were subjected.
 n_i = number of specimens in bin i , $n_i \in \{1, 2, \dots\}$.
 f_i = number of specimens in bin i that failed, $f_i \in \{0, 1, \dots, n_i\}$.

One proper way to estimate θ and β is by the maximum likelihood method, i.e., by finding the

values of θ and β that have the highest likelihood of producing the observed data. At any level of excitation r_i , there is a probability of any individual specimen failing that is given by the lognormal CDF. Let p_i denote this probability:

$$p_i = \Phi\left(\frac{\ln(r_i/\theta)}{\beta}\right) \quad (18)$$

Assume that the failure of any two different specimens is independent conditioned on excitation. In that case, if one were to estimate the number of failed specimens in bin i , it would be proper to take that number as a random variable with a binomial distribution. Let F_i denote that random variable. The following equation gives the probability that one will observe f_i failures among n_i specimens with the per-occurrence failure probability p_i :

$$P[F_i = f_i] = \frac{n_i!}{f_i!(n_i - f_i)!} \cdot p_i^{f_i} \cdot (1 - p_i)^{n_i - f_i} \quad (19)$$

This is the binomial distribution. One finds the θ and β values that maximize the likelihood of observing all the data $\{n_1, f_1, n_2, f_2, \dots\}$ given excitations $\{r_1, r_2, \dots\}$. That likelihood is given by the product of the probabilities in Eq. 19, multiplied over all the bins. That is, find θ and β that maximize $L(\theta, \beta)$ in:

$$L(\theta, \beta) = \prod_{i=1}^{m_i} P[F_i = f_i] \quad (20)$$

One can explicitly maximize $L(\theta, \beta)$, but it is easier to use Excel or similar MATLAB or other software. Excel's solver is straightforward.

It may be easier to remember a more approximate approach of minimizing the weighted squares of the difference between the observed data and the idealized fragility function. That is, find θ and β that minimize the squared error term $\varepsilon^2(\theta, \beta)$ in:

$$\varepsilon^2(\theta, \beta) = \sum_{i=1}^{m_i} n_i \cdot \left(p_i - \frac{f_i}{n_i}\right)^2 \quad (21)$$

The difference between the fragility functions derived by these two different methods generally appears to be small compared with the scatter about the regression lines, so arguments about what is more proper tend to be academic compared with the choice of which one is easier. There is at least one more reasonable approach, called logistic regression, but again this approach tends to produce roughly the same values of θ and β , with differences that are small compared with data scatter.

Other Data Conditions

There are cases where none of the specimens failed (type C, or capable, data), where failure is derived by structural analysis (type D data), where expert opinion is used (type E), or where Bayesian updating is used to update an existing fragility function with new evidence (type U). For such situations, see Porter et al. (2007).

Dealing with Under-Representative Specimens

If the specimens used to create the fragility function are very few in number or unrepresentative of the broader class whose fragility is desired, or if the excitation to which they were subjected was unlike real-world earthquake shaking, one can reflect added uncertainty associated with unrepresentative conditions by increasing the fragility function's logarithmic standard deviation. The FEMA P-58 guidelines, for example, suggest always increasing the logarithmic standard deviation of a fragility function that is derived from test data or observations, as follows:

$$\beta' = \sqrt{\beta^2 + \beta_u^2} \quad (22)$$

where β' is the new, increased value of the logarithmic standard deviation of the fragility function, β is the value derived using the test or post-earthquake observation data, and β_u is a term to reflect uncertainty that the tests represent real-world conditions of installation and loading or uncertainty that the available data are an adequate sample size to accurately represent the true variability.

FEMA P-58 recommends values of β_u depending on how under-representative are the

data. If any of the following is true, use a minimum value of $\beta_u = 0.25$; otherwise, use $\beta_u = 0.10$.

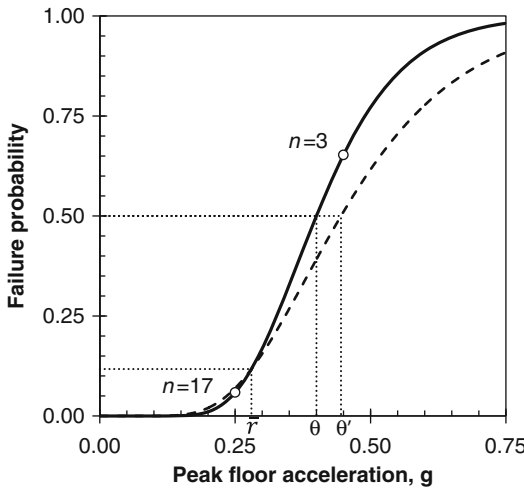
1. Data are available for five or fewer specimens.
2. In an actual building, a component can be installed in a number of different configurations, but all specimens were tested with the same configuration.
3. All specimens were subjected to the same loading protocol.
4. Actual behavior of the component is expected to be dependent on two or more demand parameters (e.g., simultaneous drift in two orthogonal directions), but specimens were loaded using only one demand parameter.

In the case of type B data, increasing β using Eq. 22 introduces a bias in long-term failure probability and can cause the fragility function not to pass through the actual failure data well. If the data generally lie at excitation levels below the derived median capacity θ , which is common, then increasing β without adjusting θ will cause the fragility function to move up (to higher probabilities) relative to the data. To increase β while still ensuring that the derived fragility function passes through the data, one can adjust θ as follows:

$$\bar{r} = \frac{\sum_{i=1}^{N_i} r_i \cdot n_i}{\sum_{i=1}^{N_i} n_i} \quad (23)$$

$$\theta' = \bar{r} \cdot \left(\frac{\bar{r}}{\theta} \right)^{(-\beta'/\beta)} \quad (24)$$

In the case of type A data, use $n_i = 1$ for all i in Eq. 23. For example, imagine that 17 suspended ceilings of identical construction are installed in a large, stiff single-story building, a block away from a 2-story building with three suspended ceilings on the upper floor. A ceiling in the 1-story building and two ceilings in the 2-story building collapse in a particular earthquake. The estimated roof accelerations in the two buildings are 0.25 and 0.45 g, respectively. These type



Beginner's Guide to Fragility, Vulnerability, and Risk, Fig. 1 Increasing beta and adjusting theta to account for under-representative samples

B data are used to derive a fragility function in terms of peak floor acceleration, with $\beta = 0.3$ and $\theta = 0.4$ g. These conditions meet FEMA P-58's criteria 2 and 3, so $\beta_u = 0.25$. Applying the FEMA P-58-recommended value $\beta_u = 0.25$ and evaluating Eqs. 22, 23, and 24 yields

$$\begin{aligned}\beta' &= \sqrt{0.3^2 + 0.25^2} = 0.39 \\ \bar{r} &= \frac{(17 \cdot 0.25 + 3 \cdot 0.45)}{20} = 0.28 \\ \theta' &= 0.28 \cdot \left(\frac{0.28}{0.4}\right)^{(-0.39/0.30)} = 0.45\end{aligned}$$

Figure 1 illustrates the data, the fragility function before increasing β (the solid line) and the fragility function after increasing β . Note how the two curves cross at $r = \bar{r}$ and how $\theta' > \theta$. The solid curve passes through the data because there are only two data points, but it passes over the first data point and below the second, closer to the first because the first represents more data.

Some Useful Sources of Component Fragility Functions

FEMA P-58 produced a large suite of component fragility functions. See https://www.atcouncil.org/files/FEMAP-58-3_2_ProvidedFragilityData.zip.

“Component” means a building component, like an RSMeans assembly such as glass curtain walls. The component fragility functions in FEMA P-58 mostly use as excitation measures the peak transient interstory drift ratio to which a drift-sensitive specimen is subjected or the peak absolute acceleration of the floor or roof to which the specimen is attached, for acceleration-sensitive components. In some cases peak residual drift is used (e.g., doors getting jammed shut). There are other measures of excitation as well. FEMA P-58 failure modes are defined with symptoms of physical damage or nonfunctionality requiring particular, predefined repair measures. They are never vague qualitative states such as “minor damage” that require judgment to interpret. Most of the FEMA P-58 fragility functions are derived from post-earthquake observations or laboratory experiments. Some are based on structural analysis and some are derived from expert opinion. All were peer reviewed. Johnson et al. (1999) also offer a large library of component fragility functions, many based on post-earthquake observations of standard mechanical, electrical, and plumbing equipment in power facilities.

The HAZUS-MH technical manual (NIBS and FEMA 2009) offers a number of whole-building fragility functions, defining for instance probabilistic damage to all the drift-sensitive nonstructural components in the building in four qualitative damage states (slight, moderate, extensive, complete) as a function of a whole-building measure of structural response (spectral acceleration response or spectral displacement response of the equivalent nonlinear SDOF oscillator that represents the whole building).

Vulnerability

Vulnerability Terminology

So far, this entry has discussed damageability in terms of the occurrence of some undesirable event such as a building collapse that either occurs or does not occur. Damageability is also measured in terms of the degree of the

undesirable outcome, called loss here, in terms of repair costs, life-safety impacts, and loss of functionality (dollars, deaths, and downtime) or in terms of environmental degradation, quality of life, historical value, and other measures. When loss is depicted as a function of environmental excitation, the function can be called a vulnerability function. A seismic vulnerability function relates uncertain loss to a measure of seismic excitation, such as spectral acceleration response at some damping ratio and period. A seismic vulnerability function usually applies to a particular asset class.

Vulnerability is not fragility. Vulnerability measures loss; fragility measures probability. Vulnerability functions are referred to many ways: damage functions, loss functions, vulnerability curves, and probably others.

When a vulnerability function measures repair cost, it is commonly normalized by replacement cost new (RCN), a term which here means the cost of a similar new property having the nearest equivalent utility as the property being appraised, as of a specific date (American Society of Appraisers 2013). RCN excludes land value and usually refers to part or all of the fixed components of a building or other assets (structural, architectural, mechanical, electrical, and plumbing components) or to its contents. Repair cost divided by RCN is referred to here as damage factor (DF). Some authors call it damage ratio, fractional loss, or other terms. The expected value of DF conditioned on excitation is commonly called mean damage factor (MDF). Sometimes it is assumed that if DF exceeds some threshold value such as 0.6, the property is not worth repairing and is a total loss, so repair-cost vulnerability functions can jump abruptly from 0.6 (or other threshold value) to 1.0 with increasing excitation. In principle DF can exceed 1.0 because it can cost more to repair a building than to replace it.

When a vulnerability function measures life-safety impacts, it commonly measures the fraction of indoor occupants who become casualties (that is, they are killed or experience a nonfatal injury to some specified degree) as a function of excitation. There are a variety of human-injury

scales, some used by civil engineers and others used by public health professionals. Before using a terms such as minor injury, one should be sure its meaning is entirely clear and meaningful to the intended user of the vulnerability information. Civil engineers sometimes use casualty scales that are ambiguous or not useful to public health professionals.

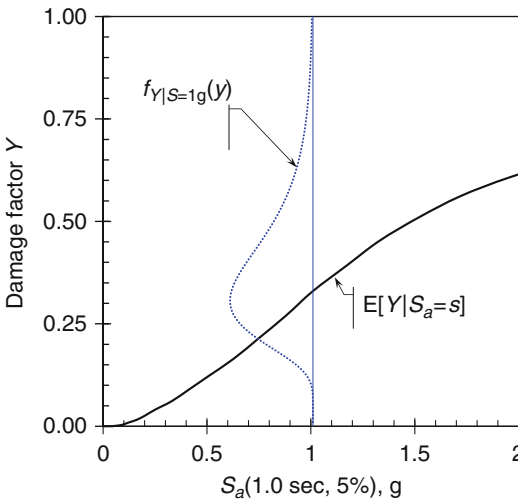
Downtime is commonly measured in terms of days or fractions of a year during which the asset cannot be used for its intended purposes. Sometimes it measures the time from the environmental shock (the earthquake, in the case of a seismic vulnerability function) to the time when all repairs are completed, which includes both the time required to perform the repairs and a previous period during which damage is assessed, repairs are designed, financing is arranged, repairs are bid out, and the repair contractor mobilizes to the site.

Many vulnerability functions are expressed with conditional probability distributions that give a probability that loss will not exceed some specified value given the excitation, for a particular asset class. The distribution is often assigned a parametric form such as lognormal or beta, in which case the parameters of the distribution are all required, some or all of them conditioned on excitation. Figure 2 presents a sample probabilistic vulnerability function.

Brief Summary of Vulnerability Derivation Methods

While this entry does not provide guidance on how to derive vulnerability functions, it is worthwhile to understand the different approaches to doing so and some of the relevant history. There are three distinct approaches – sometimes referred to as empirical, analytical, and expert opinion – and some hybrids that combine aspects of two or more approaches. They are briefly explained here.

An empirical vulnerability function is one derived by regression analysis of observations of pairs of excitation and loss for specimens of an asset class. Professional users of vulnerability functions such as insurance loss estimators tend to prefer empirical vulnerability functions over



Beginner's Guide to Fragility, Vulnerability, and Risk, Fig. 2 Sample vulnerability function that relates damage factor to 1-s spectral acceleration response for some asset class. The *solid line* depicts the expected value of damage factor, and the *dotted curve* depicts the probability density function of damage factor at 1 g of excitation

other approaches, because they are known to be based on real-world observations. One of the best examples of an empirical vulnerability function is the one derived by Wesson et al. (2004), who performed regression analysis of losses to single-family housing from ground motion in the 1994 Northridge earthquake. Among the challenges to deriving empirical vulnerability functions are the scarcity of nearby ground motion observations, the difficulty getting reliable loss information, and relatively few observations at high levels of ground motion.

Analytical methods are useful when empirical data are lacking and one has the time and other resources to create and analyze a model. An analytical approach generally involves defining one or more specimens to represent the class, creating and analyzing a structural model to estimate structural response as a function of ground motion, estimating component-level damage given structural response, and then estimating loss given component damage. The interested reader is referred to FEMA P-58 (ATC 2012) for the current state of the practice for individual buildings and to Porter et al. (2009a, b; 2010,

2014) for building classes. Analytical methods offer the advantage of being able to distinguish the effect of any feature of interest and any asset class the analyst can model, but the disadvantages of cost and often a lack of data to validate the results.

Expert opinion can quickly provide vulnerability functions where empirical data are missing and the analyst lacks the resources for an analytical model. Briefly, one convenes a group of experts familiar with the performance of the asset class of interest and with a structured interview process elicits their judgment of the performance of the class at each of many levels of excitation. ATC-13 (1985) represents one of the earliest and most thorough expert-opinion models of the seismic vulnerability of buildings and other asset classes. Jaiswal et al. (2012) offers a more recent example.

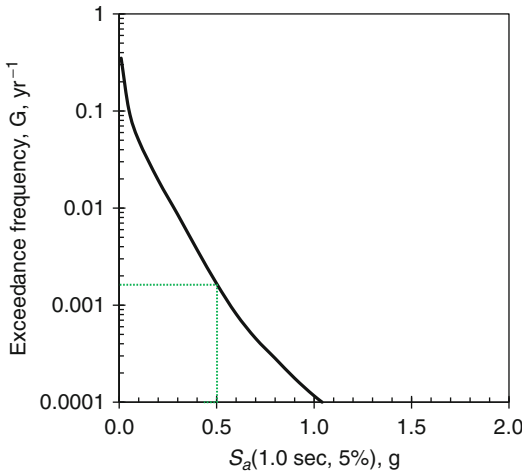
Hazard

Probabilistic Seismic Hazard Analysis

Seismic hazard refers here to an uncertain relationship between the level of some measure of seismic excitation and the frequency or probability of a particular location experiencing at least that level of excitation. It is *not* the measure of excitation, the occurrence of an earthquake, nor the probability or frequency of excitation. These terms are not to be used interchangeably.

Seismic hazard is quantified many ways. One is through a hazard curve, commonly depicted on an x - y chart where the x -axis measures shaking intensity at a site and the y -axis measures either exceedance probability in a specified period of time or exceedance rate in events per unit time. See Fig. 3 for an example Cornell (1968) applied the theorem of total probability to create a hazard curve. What follows here is a summary of current procedures to perform probabilistic seismic hazard analysis (PSHA), but is conceptually identical to Cornell's work.

Engineers sometimes refer to the quantity by which intensity is measured as the intensity measure. Earth scientists call it ground motion. The present work uses the term intensity measure (IM).



Beginner's Guide to Fragility, Vulnerability, and Risk, Fig. 3 Sample hazard curve. The *solid line* shows the mean rate at which various levels of 1 s spectral acceleration response are exceeded. The *dotted line* shows that 0.5 g is exceeded on average about once in 600 years

Some authors distinguish between the intensity measure type (IMT), such as 5 % damped spectral acceleration response at 0.2 s period, and the intensity measure level (IML), a particular value of the IM such as 0.4 g. In any case, IMT must be completely specified. If using damped elastic spectral acceleration response, one states the period, damping ratio, and whether one is referring to the geometric mean of two orthogonal directions or the maximum direction or other directional references.

To estimate seismic hazard, one applies the theorem of total probability to combine the uncertain shaking at the site caused by a particular fault rupture and the occurrence frequency or probability of that rupture. Earth scientists create models called earthquake rupture forecasts that specify the locations and rates at which various fault produce earthquakes of various sizes, e.g., the Uniform California Earthquake Rupture Forecast version 2 (UCERF2, Field et al. 2007). The uncertain shaking given a fault rupture is quantified using a relationship variously called an attenuation relationship or a ground-motion prediction equation, such as the next-generation attenuation (NGA) relationships presented in the

February 2008 issue of *Earthquake Spectra*. Mathematically, a hazard curve is created as follows. Let:

H = uncertain severity of ground motion at the building site, e.g., $S_a(T, 5\%)$.

h = a particular value of H .

$G(h)$ = frequency (events per unit time) with which $H > h$, i.e., the number of events per unit time in which at least once during the event, $H > h$.

n_E = number of earthquake rupture forecast models to consider, $n_E \in \{1, 2, \dots\}$. For example, in UCERF2, there were 480 discrete combinations of fault model, rupture model, magnitude-area relationship, and several other modeling choices that together are represented by a logic tree that begins with fault model, branches to rupture model, then to magnitude-area relationship, etc. For UCERF2, $n_E = 480$.

E = the “correct” earthquake rupture forecast, which is uncertain. $E \in \{1, 2, \dots, n_E\}$.

e = an index to a particular earthquake rupture forecast, $e \in \{1, 2, \dots, n_E\}$.

$P[E = e]$ = Bayesian probability assigned to earthquake rupture forecast e . In UCERF2, this would be the product of the conditional probabilities (weights) of the individual branches in the logic tree that represents UCERF2. Conditional probability means that each branch’s probability can (though only sometimes do) depend on which choices came before it.

n_A = number of attenuation relationships (also called ground motion prediction equations) to be employed, $n_A \in \{1, 2, \dots\}$

A = the “correct” ground-motion-prediction equation, i.e., the one that expresses the true state of nature, which is uncertain.

a = an index to ground-motion prediction equations.

$P[A = a]$ = Bayesian probability assigned to ground-motion-prediction equation a , i.e., the probability that ground-motion-prediction equation a actually reflects the true state of nature and is therefore the correct ground-motion-prediction equation.

$n_F(e)$ = number of fault sections in earthquake rupture forecast e , $n_F(e) \in \{1, 2, \dots\}$.

f = an index to faults section, $f \in \{1, 2, \dots n_F(e)\}$.

m_0 = minimum magnitude to consider.

Δm = an increment of magnitude, say, $\Delta m = 0.1$.

$M_{\max}(f|E = e)$ = maximum magnitude of which fault f is deemed capable under earthquake rupture forecast e , $M_{\max}(f|E = e) \in \{m_0 + \frac{1}{2} \cdot \Delta m, m_0 + \frac{3}{2} \cdot \Delta m, \dots\}$.

$n_o(f, m)$ = number of locations within fault section f that can generate a rupture of magnitude m , $n_o(f, m) \in \{1, 2, \dots\}$.

o = an index to locations on fault section f , in $o \in \{1, 2, \dots n_o(f, m)\}$.

$G(f, m, o|E = e)$ = frequency (events per unit time) with which fault section f ruptures at location o producing earthquake of magnitude $m \pm \frac{1}{2} \cdot \Delta m$ given earthquake rupture forecast e .

V_S = uncertain site soil, potentially measured by NEHRP site soil classification or shearwave velocity in the top 30 m of soil (Vs30) or something else. To find Vs30, see the site data application at www.opensha.org/apps for U.S. locations or Google the USGS's Global Vs30 map server for other locations

v = a particular value of V .

$P[H > h | m, r, v, a]$ = probability that $H > h$ given earthquake magnitude m at distance r on soil with soil type v using ground-motion prediction equation a . Note that, given a known site location, fault segment f , and location along the fault segment o , distance r from the site to the fault is known. Note that ground-motion prediction equations use a variety of distance measures. As of this writing, most ground-motion prediction equations give an equation for the mean of the natural logarithm of H given m , r , and v , for the logarithmic standard deviation of H (i.e., the standard deviation of the natural logarithm of H), and assume a lognormal distribution of H conditioned on m , r , and v . Let $m_{\ln H}$ and $s_{\ln H}$ denote the mean of $\ln(H)$ and the standard deviation of $\ln(H)$, respectively, given m , r , and v , assuming ground-motion-prediction equation a . Under these conditions,

$$P[H > h | m, r, v, a] = 1 - \Phi\left(\frac{\ln(h) - m_{\ln H}}{s_{\ln H}}\right) \quad (25)$$

One can now estimate the hazard curve by applying the theorem of total probability. Suppose one always knew soil conditions V with certainty. Then

$$G(h) = \sum_{e=1}^{n_E} \sum_{a=1}^{n_A} \sum_{f=1}^{n_F(e)} \sum_{m} \sum_{o=1}^{n_o(f, m)} P[H > h | m, r, v, a] \cdot G(m, f, o | E = e) \cdot P[A = a] \cdot P[E = e] \quad (26)$$

where the summation over m means that one considers each $m \in \{m_0 + \frac{1}{2} \cdot \Delta m, m_0 + \frac{3}{2} \cdot \Delta m, \dots M_{\max}(f|E = e)\}$.

Hazard Rate Versus Probability

Seismic hazard is often expressed in terms of exceedance probability, rather than in terms of exceedance rate. The distinction is this: exceedance probability is the probability that shaking of $H > h$ will occur at least once in a given period of time. That means that it is the probability that it $H > h$ will occur exactly once, plus the probability that it will occur exactly twice, etc. For some calculations, the analyst wants rate (number of

events per unit time) not probability (chance that it occurs one or more times in a given time period).

The two can be related using a concept called a Poisson process. From the Wikipedia's article on the Poisson Process: "In probability theory, a Poisson process is a stochastic process which counts the number of events and the time that these events occur in a given time interval. The time between each pair of consecutive events has an exponential distribution with parameter G [the parameter is the occurrence rate per unit time] and each of these inter-arrival times is ... independent of other

inter-arrival times [meaning that the time between the second and third occurrence is independent of the time between the first and second occurrence – knowing one tells you nothing about the other, so a Poisson process is called memoryless]. The process is named after the French mathematician Siméon-Denis Poisson and is a good model of radioactive decay, telephone calls, and requests for a particular document on a web server, among many other phenomena.”

Modeling earthquakes as Poisson arrivals is convenient in part because in a Poisson process, arrival rate and occurrence probability have this relationship:

$$G = \frac{-\ln(1 - P)}{t} \quad (27)$$

where

G = occurrence rate, i.e., events per unit time,
 P = probability that at least one event will occur in time t .

So if a hazard curve were represented as $P(h)$, the probability that $H > h$ at least once in a particular period of time t , one could use Eq. 27 to estimate the occurrence rate $G(h)$, the average number of times that $H \geq h$ per unit time. One can also rearrange Eq. 27 to give probability as a function of rate:

$$P = 1 - e^{-G \cdot t} \quad (28)$$

So with the occurrence rate $G(h)$ of earthquakes causing $H > h$, one can calculate the probability that at least one earthquake with $H > h$ will occur in a given time t .

Measures of Seismic Excitation

These are many common measures of seismic excitation. Some measure ground motion and some measure structural response or excitation to which the components of a building, bridge, or other structural systems are subjected. This section introduces common ones.

Some Commonly Used Measures of Ground Motion

Peak ground acceleration (PGA). This is the maximum value of acceleration of a particular point on the ground at any time during an earthquake. Often PGA is estimated as the geometric mean (the square root of the product) of the maximum values of PGA parallel to each of two orthogonal horizontal axes. PGA is sometimes called zero-period acceleration (ZPA), meaning the spectral acceleration response of a single-degree-of-freedom elastic oscillator with zero, or near-zero, period. Before using a PGA value, be sure you know whether it refers to geometric mean, maximum-direction value, or something else.

Peak ground velocity (PGV). This is like PGA, except maximum velocity of a point on the ground rather than acceleration.

Peak ground displacement (PGD). This is like PGA, except maximum displacement relative to a fixed datum.

Spectral acceleration response ($S_a(T, z)$). This usually refers to damped elastic spectral acceleration response at some specified index period such as $T = 0.3$ s, 1.0 s, 3.0 s, etc., and specified damping ratio such as $z = 5\%$, at a particular point on the ground. To be precise, $S_a(T, z)$ is the absolute value of the maximum acceleration relative to a fixed datum of a damped elastic single-degree-of-freedom harmonic oscillator with period T and damping ratio z when subjected to a particular one-degree-of-freedom ground-motion time history at its base. In practice, it typically refers to the geometric mean of spectral acceleration response parallel to each of two orthogonal horizontal axes. It is so often measured for $z = 5\%$ that damping ratio is often not mentioned. Before using $S_a(T, z)$, be sure you know T , z , and whether it is a geometric mean value or the maximum-direction value or something else.

Spectral displacement response ($S_d(T, z)$). This is like $S_a(T, z)$ but for relative spectral displacement (displacement of the oscillator relative to its base, not relative to a fixed datum) rather than absolute acceleration of the oscillator.

Pseudoacceleration response ($PSA(T, z)$). $PSA(T, z)$ is defined as $S_d(T, z) \cdot \omega^2$, where ω is

angular frequency, $2 \cdot \pi/T$. Some authors prefer to use $PSA(T,z)$ rather than $S_a(T,z)$, but for values of z less than about 20 %, the two measures are virtually identical.

Modified Mercalli Intensity (MMI) and European Macroseismic Scale (EMS). These are macroseismic intensity measures, meaning that they measure seismic excitation over a large area, not at a particular point on the ground. They are measured with an integer scale in Roman numerals from I to XII. They measure whether and how people in a region such as a neighborhood or city felt and reacted to earthquake motion (did they run outside?) and what they observed to happen to the ground, buildings, and contents around them, such as plates rattling and weak masonry being damaged. They are subjective and generally easier for nontechnical audiences to understand than instrumental measures. On the MMI scale, building damage begins around MMI VI and it is rare for an earthquake to produce shaking of $MMI \geq X$. EMS is similar to MMI, but building-damage observations are related to common European building types. A version of EMS defined in 1998 is often referred to as EMS-98. For detail see Table 1, Wood and Neumann (1931) and European Seismic Commission Working Group—Macroseismic Scales (1998).

Japan Meteorological Agency (JMA) seismic intensity scale. This is like MMI, but with a 0–7 scale. It has both a macroseismic sense (observed effects of people and objects) and an instrumental sense (in terms of ranges of PGA). See http://en.wikipedia.org/wiki/Japan_Meteorological_Agency_seismic_intensity_scale for details.

Instrumental intensity measure (IMM). This is a positively valued measure of intensity that can take on fractional values, e.g., 6.4. This is an estimate of MMI using functions of instrumental ground-motion measures such as PGA and PGV.

Conversion Between Instrumental and Macroseismic Intensity

It is often desirable to convert between instrumental ground-motion measures such as PGA or PGV and macroseismic intensity

measures, especially MMI. One reason is that MMI observations can be made by people exposed to shaking or who make post-earthquake observations, whereas instrumental measures require an instrument.

Ground motion to intensity conversion equations (GMICE). These estimate macroseismic intensity as a function of instrumental measures of ground motion. There are several leading GMICEs. When selecting among them, try to match the region, magnitude range, and distance range closest to the conditions where the GMICE will be applied. More data for conditions like the ones in question are generally better than less data, all other things being equal. When considering building response, GMICEs that convert from $S_a(T,z)$ to macroseismic intensity are generally better than those that use PGA or PGV, which do not reflect anything building-specific. Two recent GMICEs for the United States are as follows:

As of this writing, Worden et al.'s (2012) relationships in Eqs. 29 and 30 seem to be the best choice for estimating MMI from ground motion and vice versa for California earthquakes. The reason is they employ a very large dataset of California (ground motion, MMI) observations. The dataset includes 2092 PGA-MMI observations and 2074 PGV-MMI observations from 1207 California earthquakes $M = 3.0\text{--}7.3$, $MMI = 2.0\text{--}8.6$, $R = 4\text{--}500$ km. It includes no observations from continental interior. It includes regressions for $S_a(0.3\text{ s}, 5\%)$, $S_a(1.0\text{ s}, 5\%)$, $S_a(3.0\text{ s}, 5\%)$, PGA, and PGV that operate in both directions, meaning that one can rearrange the relationships to estimate instrumental measures in terms of MMI, as well as MMI in terms of instrumental measures. The reason that the relationships are bidirectional is that Worden et al. (2012) used a total least-squares data modeling technique in which observational errors on both dependent and independent variables are taken into account. Equation 30 includes the option to account for the apparent effects of magnitude M and distance R . The columns for residual standard deviations show a modest reduction in uncertainty when accounting for M and R (Table 2).

Beginner's Guide to Fragility, Vulnerability, and Risk, Table 1 MMI and EMS-98 macroseismic intensity scales (abridged)

MMI	Brief description	EMS-98	Brief description
I. Instrumental	Generally not felt by people unless in favorable conditions	I. Not felt	Not felt by anyone
II. Weak	Felt only by a couple people that are sensitive, especially on the upper floors of buildings. Delicately suspended objects (including chandeliers) may swing slightly	II. Scarcely felt	Vibration is felt only by individual people at rest in houses, especially on upper floors of buildings
III. Slight	Felt quite noticeably by people indoors, especially on the upper floors of buildings. Standing automobiles may rock slightly. Vibration similar to the passing of a truck. Indoor objects may shake	III. Weak	The vibration is weak and is felt indoors by a few people. People at rest feel swaying or light trembling. Noticeable shaking of many objects
IV. Moderate	Felt indoors by many people, outdoors by few. Some awakened. Dishes, windows, and doors disturbed, and walls make cracking sounds. Chandeliers and indoor objects shake noticeably. Like a heavy truck striking building. Standing automobiles rock. Dishes and windows rattle	IV. Largely observed	The earthquake is felt indoors by many people and outdoors by few. A few people are awakened. The level of vibration is possibly frightening. Windows, doors, and dishes rattle. Hanging objects swing. No damage to buildings
V. Rather strong	Felt inside by most or all and outside. Dishes and windows may break. Vibrations like a train passing close. Possible slight damage to buildings. Liquids may spill out of glasses or open containers. None to a few people are frightened and run outdoors	V. Strong	Felt indoors by most, outdoors by many. Many sleeping people awake. A few run outdoors. China and glasses clatter. Top-heavy objects topple. Doors and windows swing
VI. Strong	Felt by everyone; many frightened and run outdoors, walk unsteadily. Windows, dishes, glassware broken; books fall off shelves; some heavy furniture moved or overturned; a few instances of fallen plaster. Damage slight to moderate to poorly designed buildings; all others receive none to slight damage	VI. Slightly damaging	Felt by everyone indoors and by many outdoors. Many people in buildings are frightened and run outdoors. Objects on walls fall. Slight damage to buildings; for example, fine cracks in plaster and small pieces of plaster fall
VII. Very strong	Difficult to stand. Furniture broken. Damage light in buildings of good design and construction; slight to moderate in ordinarily built structures; considerable damage in poorly built or badly designed structures; some chimneys broken or heavily damaged. Noticed by people driving automobiles	VII. Damaging	Most people are frightened and run outdoors. Furniture is shifted and many objects fall from shelves. Many buildings suffer slight to moderate damage. Cracks in walls; partial collapse of chimneys

(continued)

Beginner's Guide to Fragility, Vulnerability, and Risk, Table 1 (continued)

MMI	Brief description	EMS-98	Brief description
VIII. Destructive	Damage slight in structures of good design, considerable in normal buildings with possible partial collapse. Damage great in poorly built structures. Brick buildings moderately to extremely heavily damaged. Possible fall of chimneys, monuments, walls, etc. Heavy furniture moved	VIII. Heavily damaging	Furniture may be overturned. Many to most buildings suffer damage: chimneys fall; large cracks appear in walls and a few buildings may partially collapse. Can be noticed by people driving cars
IX. Violent	General panic. Damage slight to heavy in well-designed structures. Well-designed structures thrown out of plumb. Damage moderate to great in substantial buildings, with a possible partial collapse. Some buildings may be shifted off foundations. Walls can collapse	IX. Destructive	Monuments and columns fall or are twisted. Many ordinary buildings partially collapse and a few collapse completely. Windows shatter
X. Intense	Many well-built structures destroyed, collapsed, or moderately damaged. Most other structures destroyed or off foundation. Large landslides	X. Very destructive	Many buildings collapse. Cracks and landslides can be seen
XI. Extreme	Few if any structures remain standing. Numerous landslides, cracks, and deformation of the ground	XI. Devastating	Most buildings collapse
XII. Catastrophic	Total destruction. Objects thrown into the air. Landscape altered. Routes of rivers can change	XII. Completely devastating	All structures are destroyed. The ground changes

Beginner's Guide to Fragility, Vulnerability, and Risk, Table 2 Parameter values for Worden et al. (2012) GMICE for California

Y	c_1	c_2	c_3	c_4	c_5	c_6	c_7	t_1	t_2	Eq. 29		Eq. 30	
										σ_{MMI}	$\sigma_{\log 10 Y}$	σ_{MMI}	$\sigma_{\log 10 Y}$
PGA	1.78	1.55	-1.60	3.7	-0.91	1.02	-0.17	1.57	4.22	0.73	0.39	0.66	0.35
PGV	3.78	1.47	2.89	3.16	0.90	0.00	-0.18	0.53	4.56	0.65	0.40	0.63	0.38
PSA (0.3 s)	1.26	1.69	-4.15	4.14	-1.05	0.60	0.00	2.21	4.99	0.84	0.46	0.82	0.44
PSA (1.0 s)	2.50	1.51	0.20	2.90	2.27	-0.49	-0.29	1.65	4.98	0.80	0.51	0.75	0.47
PSA (3.0 s)	3.81	1.17	1.99	3.01	1.91	-0.57	-0.21	0.99	4.96	0.95	0.69	0.89	0.64

Beginner's Guide to Fragility, Vulnerability, and Risk, Table 3 Parameter values for Atkinson and Kaka (2007) GMICE for the United States

Y	c ₁	c ₂	c ₃	c ₄	c ₅	c ₆	c ₇	σ_{MMI}	
								Eq. 31	Eq. 32
PGA	4.37	1.32	3.54	3.03	0.47	−0.19	0.26	0.80	0.76
PGV	2.65	1.39	−1.91	4.09	−1.96	0.02	0.98	1.01	0.89
PSA(0.3 s)	2.40	1.36	−1.83	3.56	−0.11	−0.20	0.64	0.88	0.79
PSA(1.0 s)	3.23	1.18	0.57	2.95	1.92	−0.39	0.04	0.84	0.73
PSA(2.0 s)	3.72	1.29	1.99	3.00	2.24	−0.33	−0.31	0.86	0.72

$$\begin{aligned}
 \text{MMI} &= c_1 + c_2 \cdot \log_{10}(Y) & \log_{10}(Y) \leq t_1 & & \text{MMI} &= c_1 + c_2 \log_{10}(Y) & \text{MMI} \leq 5 \\
 &= c_3 + c_4 \cdot \log_{10}(Y) & \log_{10}(Y) > t_1 & & &= c_3 + c_4 \cdot \log_{10}(Y) & \text{MMI} > 5
 \end{aligned} \quad (29)$$

$$\begin{aligned}
 \text{MMI} &= c_1 + c_2 \cdot \log_{10}(Y) + c_5 + c_6 \cdot \log_{10}(R) \\
 &+ c_7 \cdot M & \text{MMI} \leq t_2 \\
 &= c_3 + c_4 \cdot \log_{10}(Y) + c_5 + c_6 \cdot \log_{10}(R) \\
 &+ c_7 \cdot M & \text{MMI} > t_2
 \end{aligned} \quad (30)$$

Units of Y are cm/s or cm/s, 5 % damping. Units of R are km. Columns labeled σ show residual standard deviation and depend on whether the M and R adjustment is used or not. Use $\sigma_{\log_{10}Y}$ with rearranged equations to give $\log_{10}Y$ in terms of MMI.

Atkinson and Kaka's (2007) relationships, shown in Eqs. 31 and 32, employ smaller dataset of California observations than Worden et al. (2012), but they reflect data from central and eastern and US observations. There are 986 observations: 710 from 21 California earthquakes, $M = 3.5\text{--}7.1$, $R = 4\text{--}445$ km, $\text{MMI} = \text{II--IX}$, and 276 central and eastern US observations from 29 earthquakes $M = 1.8\text{--}4.6$, $R = 18\text{--}799$ km. They include regression for $S_a(0.3 \text{ s}, 5 \%)$, $S_a(1.0 \text{ s}, 5 \%)$, $S_a(3.0 \text{ s}, 5 \%)$, PGA, and PGV. Equation 32 accounts for the apparent effects of magnitude M and distance R . As suggested by the difference between the columns for residual standard deviation, the information added by M and R only modestly reduces uncertainty. The Atkinson and Kaka (2007) relationships are not bidirectional, meaning that one cannot rearrange them to estimate ground motion as a function of MMI.

Units of Y are cm/s or cm/s², 5 % damping. Units of R are km. Columns labeled σ show residual standard deviation and depend on whether the M and R adjustment is used or not (Table 3).

Other GMICES of potential interest include the following. Wald et al.'s (1999) relationship draws on 342 (PGA, PGV, MMI) observations from eight California earthquakes. Kaestli and Faeh (2006) offer a PGA-PGV-MMI relationship for Switzerland, Italy, and France. Tselentis and Danciu (2008) offer relationships for MMI as functions of PGA, PGV, Arias intensity, cumulative absolute velocity, magnitude, distance, and soil conditions for Greece. Kaka and Atkinson (2004) offer GMICES relating MMI to PGV and three periods of PSA for eastern North America. Sørensen et al. (2007) offer a GMICE relating EMS-98 to PGA and PGV for Vrancea, Romania.

For relationships that give ground motion as a function of MMI (intensity to ground-motion conversion equations, IGMCE), consider Faenza and Michelini (2010) for Italy, Murphy, and O'Brien (1977) for anywhere in the world and Trifunac and Brady (1975) for the western United States. Unless explicitly stated, GMICE and IGMCE relationships are not

interchangeable – it is inappropriate to simply rearrange terms of a GMICE to produce an IGMCE. The reason is that both GMICE and IGMCE are derived by regression analysis. Given (x,y) data, a least-squares regression of y as a function of x will generally produce a different curve than a least-squares regression of x as a function of y .

Hazard Deaggregation

When evaluating the risk to an asset, it is often desirable to perform nonlinear dynamic structural analyses at one or more intensity measure levels. To do so, one needs a suite of ground-motion time histories scaled to the desired intensity. The ground-motion time histories should be consistent with the seismic environment. That is, they should reflect the earthquake magnitudes m and distances r that would likely cause that level of excitation in that particular place. The reason is that magnitude and distance affect the duration and frequency content of the ground-motion time history, which in turn affects structural response. See McGuire (1995) or Bazzurro and Cornell (1999) for more information.

There is another term (commonly denoted by ε) that also matters. It relates to how the spectral acceleration response at a specified period in a particular ground-motion time history differs from its expected value, given magnitude and distance. Let y denote the natural logarithm of the intensity measure level, e.g., the natural logarithm of the spectral acceleration response at the building's estimated small-amplitude fundamental period of vibration. Let μ and σ denote the expected value and standard deviation of the natural logarithm of the intensity measure level, respectively, calculated from a ground-motion prediction equation. The ε term is a normalized value of y , as follows:

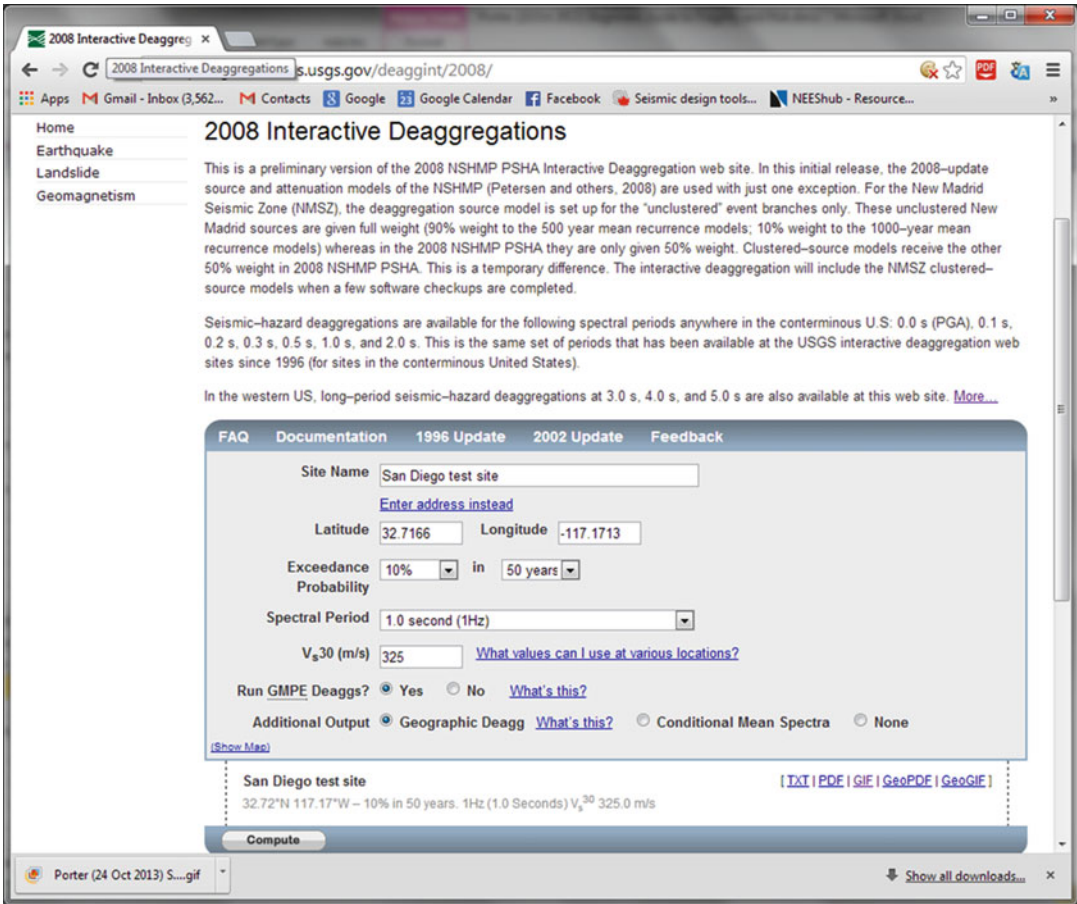
$$\varepsilon = \frac{y - \mu}{\sigma} \quad (33)$$

When calculating the motion y_0 that has a specified exceedance probability p_0 , one labels the ε from a specific source and this particular value of motion y_0 as ε_0 . The equation is the same as

Eq. 33, except with the subscript 0 on y and ε . It is practical to calculate for a given location, intensity measure type, and intensity measure level, the contribution of each fault segment, magnitude, rupture location, and value of ε_0 to the frequency with which the site is expected to experience ground motion of at least the specified intensity measure level. In fact, Eq. 26 shows that the site hazard is summed from such values. (For simplicity that equation omits mention of ε , but the extension is modest.) See Baker and Cornell (2005) for more information.

Rather than leading the reader through the math, suffice it to say that there are online tools to do that hazard deaggregation, and an example is given here. The USGS offers a website that does interactive hazard deaggregation for the United States. As of this writing, the URL includes the year associated with the hazard model, so it will change over time. The most recent tool at this writing is <https://geohazards.usgs.gov/deaggint/2008/>. When that site becomes obsolete, the reader should be able to find the current one by googling “interactive hazard deaggregation USGS.”

Consider an imaginary 12-story building in San Diego, California at 1126 Pacific Hwy, San Diego CA, whose geographic coordinates are 32.7166 N -117.1713 E (North America has negative east longitude). Suppose its small-amplitude fundamental period of vibration is 1.0 s, its V_{s30} is 325 m/s, and its depth to bedrock (defined as having a shearwave velocity of 2,500 m/s) is 1.0 km. One wishes to select several ground-motion time histories with geometric mean $S_a(1.0 \text{ s}, 5 \%)$ equal to that of the motion with 10 % exceedance probability in 50 years. The input data look like Fig. 4. The results look like Fig. 5, which shows that 10%/50-year motion at this site tends to result from earthquakes with Mw 6.6 at 1.8 km distance and a value of $\varepsilon_0 = -1.22$. One can then draw sample ground-motion time histories with approximately these values of magnitude, distance, and ε_0 from a database such as PEER's strong motion database, currently located at http://peer.berkeley.edu/peer_ground_motion_database.



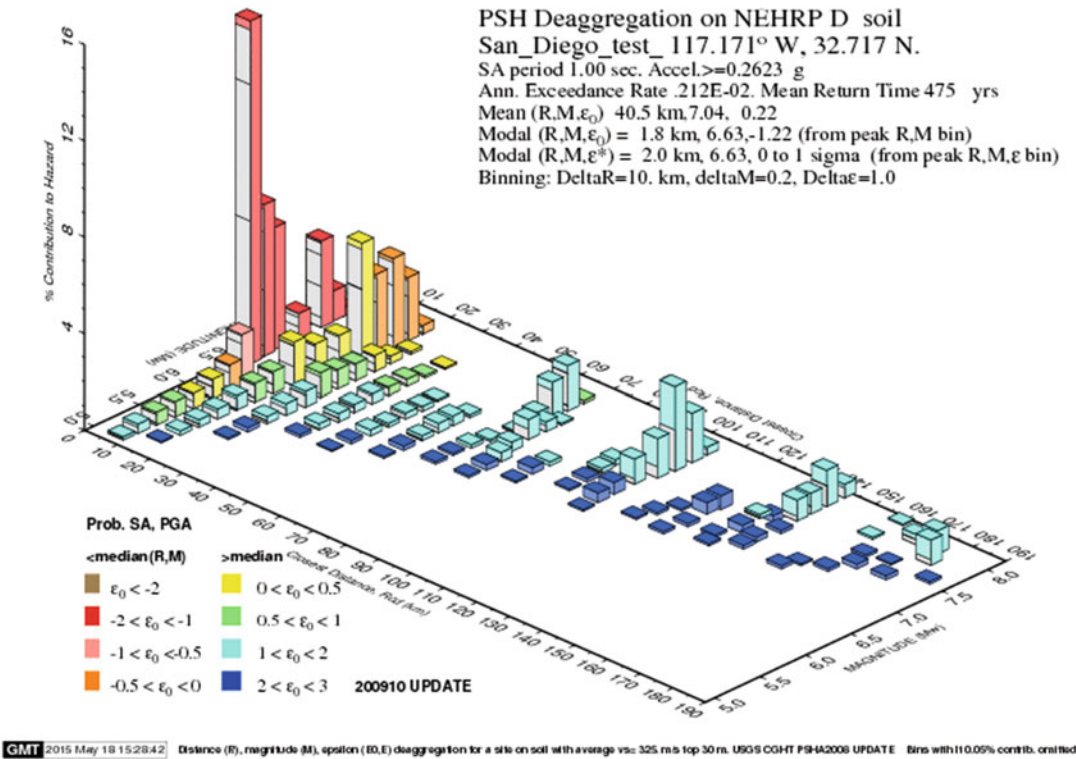
Beginner's Guide to Fragility, Vulnerability, and Risk, Fig. 4 USGS interactive hazard deaggregation website

Convenient Sources of Hazard Data

For California sites, see www.OpenSHA.org/apps for a very powerful hazard curve calculator. See <http://earthquake.usgs.gov/hazards/products/> for several sets of gridded hazard data from the National Seismic Hazard Mapping Program. For example, for the 2008 NSHMP hazard curves for $S_a(1.0 \text{ s}, 5 \%)$, see <http://earthquake.usgs.gov/hazards/products/conterminous/2008/data/> and look for the data file labeled “Hazard curve data 1 Hz (1.0 s).” If you use NSHMP’s hazard curves (as opposed to uniform seismic hazard maps), the data will reflect hazard on site class B, so adjust for other site classes. A good way to do that is by multiplying by the site coefficient F_a (for the 3.33 Hz or 5 Hz curves) or F_v (for the 1 Hz curves) from ASCE 7-10 Tables 11.4-1 and 11.4-2, respectively. You will also find at the

same location gridded uniform seismic hazard data such as the values of $S_a(1.0 \text{ s}, 5 \%)$ with 2 % exceedance probability in 50 years (the so-called MCE_G map). For that map, see the data file labeled “Gridded Hazard Map 1 Hz (1.0 s) 2 % in 50 Years” at <http://earthquake.usgs.gov/hazards/products/conterminous/2008/data/>. For elsewhere in the world, see www.globalquakemodel.org.

The risk analyst typically also needs a convenient and authoritative source of soil information. As of this writing, the USGS offers a global map of estimated average shearwave velocity in the upper 30 m of soil (V_s^{30} in ASCE 7-10 notation). See <http://earthquake.usgs.gov/hazards/apps/vs30/> or Google “Global V_s^{30} Map Server” for this useful resource. Find your latitude and longitude using Google Earth, then look up your



Beginner’s Guide to Fragility, Vulnerability, and Risk, Fig. 5 Sample output of the USGS’ interactive hazard deaggregation website

location at the USGS site by centering a grid around it, adding and subtracting 0.01 degrees of latitude and longitude for a Vs30 map about 2km × 2km grid centered at your location. See Wald and Allen (2007) for its technical basis. In the United States, the site data app at www.OpenSHA.org/apps is more powerful and easier to use. It offers the Wald-Allen Vs30 along with other sources, such as the Wills and Clahan (2006) map of Vs30 in California.

Risk for a Single Asset

Risk

This entry has dealt so far with fragility, vulnerability, and seismic hazard. Risk is analogous to hazard, but as used here it refers to the relationship between probability or frequency of the undesirable outcome and a measure of the degree of that undesirable outcome. If there are only two

possible values of that undesirable outcome – it occurs or it does not occur – one can apply the theorem of total probability, combining fragility and hazard, to estimate the mean frequency with which it occurs or the probability that it will occur in a specified period of time. If the undesirable outcome is measured in terms of loss, then one can apply the theorem of total probability, combining vulnerability and hazard, to estimate the mean annualized loss or the probability that at least a specified degree of loss will occur in a specified period of time. This entry deals with risk only for a single asset.

Expected Failure Rate for a Single Asset

Let $F(s)$ denote a fragility function for a component with a single damage state and let $G(s)$ denote the mean rate of shaking $S \geq s$ (mean number of events per year in which the shaking is at least s at the site of interest). The mean rate of failures (number of times per year that the

component reaches or exceeds the specified damage state) is given by

$$\lambda = \int_{s=0}^{\infty} -F(s) \frac{dG(s)}{ds} ds \quad (34)$$

where $G(s)$ = mean annual frequency of shaking exceeding intensity s . One can also use integration by parts and show that

$$\lambda = \int_{s=0}^{\infty} \frac{dF(s)}{ds} G(s) ds \quad (35)$$

If, for example, $F(s)$ is taken as a cumulate log-normal distribution function, $dF(s)/ds$ is the log-normal probability density function, i.e.,

$$\begin{aligned} F(s) &= \Phi\left(\frac{\ln(s/\theta)}{\beta}\right) \\ \frac{dF(s)}{ds} &= \phi\left(\frac{\ln(s/\theta)}{\beta}\right) \end{aligned} \quad (36)$$

Equation 34 only rarely can be evaluated in closed form. More commonly, $G(s)$ is available only at discrete values of s . If one has $n + 1$ values of s , at which both $F(s)$ and $G(s)$ are available, and these are denoted by s_i , F_i , and G_i ; $i = 0, 1, 2, \dots, n$, respectively, then Eq. 34 can be evaluated numerically by

$$\begin{aligned} \lambda &= \sum_{i=1}^n \left(F_{i-1} G_{i-1} (1 - \exp(m_i \Delta s_i)) \right. \\ &\quad \left. - \frac{\Delta F_i}{\Delta s_i} G_{i-1} \left(\exp(m_i \Delta s_i) \left(\Delta s_i - \frac{1}{m_i} \right) + \frac{1}{m_i} \right) \right) \\ &= \sum_{i=1}^n (F_{i-1} a_i - \Delta F_i b_i) \end{aligned} \quad (37)$$

where

$$\begin{aligned} \Delta s_i &= s_i - s_{i-1} \quad \Delta F_i = F_i - F_{i-1} \\ m_i &= \ln(G_i/G_{i-1})/\Delta s_i \quad \text{for } i = 1, 2, \dots, n \end{aligned}$$

$$a_i = G_{i-1} (1 - \exp(m_i \Delta s_i))$$

$$b_i = \frac{G_{i-1}}{\Delta s_i} \left(\exp(m_i \Delta s_i) \left(\Delta s_i - \frac{1}{m_i} \right) + \frac{1}{m_i} \right)$$

Equation 37 is exact for piecewise linear F and piecewise loglinear G .

Probability of Failure During a Specified Period of Time

If one assumes that hazard and fragility are memoryless and do not vary over time, then failure is called a Poisson process, and the probability that failure will occur at least once in time t is given by

$$P_f = 1 - \exp(-\lambda \cdot t) \quad (38)$$

where λ is the expected value of failure rate, calculated for example using Eq. 37.

Expected Annualized Loss for a Single Asset

Now consider risk in terms of degree of loss to a single asset. There are many risk measures in common use. First, consider the expected annualized loss (*EAL*). It is analogous to mean rate of failures as calculated in Eq. 34. If loss is measured in terms of repair cost, *EAL* is the average quantity that would be spent to repair the building every year. It can be calculated as

$$EAL = V \int_0^{\infty} y(s) \frac{-dG(s)}{ds} ds \quad (39)$$

where V refers to the replacement value of the asset and $y(s)$ is the expected value of loss given shaking s as a fraction of V . Equation 39 can only rarely be evaluated in closed form. More commonly, $y(s)$ and $G(s)$ are available at discrete values of s . If one has $n + 1$ values of s , at which both $y(s)$ and $G(s)$ are available, and these are denoted by s_i , y_i , and G_i ; $i = 0, 1, 2, \dots, n$, respectively, then *EAL* in Eq. 39 can be replaced by

$$\begin{aligned}
 EAL &= V \sum_{i=1}^n \left(y_{i-1} G_{i-1} (1 - \exp(m_i \Delta s_i)) \right. \\
 &\quad \left. - \frac{\Delta y_i}{\Delta s_i} G_{i-1} \left(\exp(m_i \Delta s_i) \left(\Delta s_i - \frac{1}{m_i} \right) + \frac{1}{m_i} \right) \right) \\
 &= V \sum_{i=1}^n (y_{i-1} a_i - \Delta y_i b_i)
 \end{aligned} \tag{40}$$

where

$$\begin{aligned}
 \Delta s_i &= s_i - s_{i-1} \quad \Delta y_i = y_i - y_{i-1} \\
 m_i &= \ln(G_i/G_{i-1})/\Delta s_i \quad \text{for } i = 1, 2, \dots, n
 \end{aligned}$$

$$\begin{aligned}
 a_i &= G_{i-1} (1 - \exp(m_i \Delta s_i)) \\
 b_i &= \frac{G_{i-1}}{\Delta s_i} \left(\exp(m_i \Delta s_i) \left(\Delta s_i - \frac{1}{m_i} \right) + \frac{1}{m_i} \right)
 \end{aligned}$$

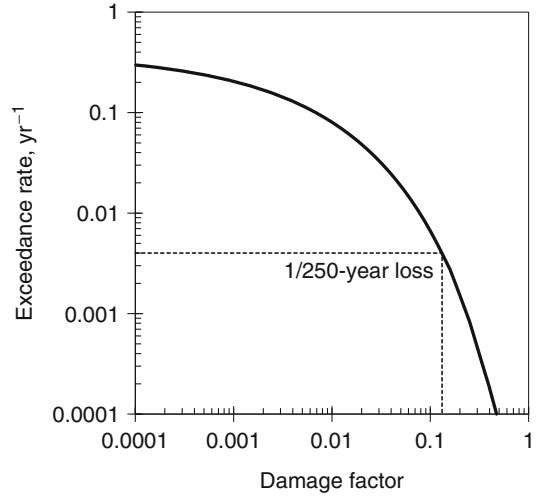
Risk Curve for a Single Asset

It is often desirable to know the probability that loss will exceed a particular value during a given time period t as a function of loss. Here, that function is called a risk curve or a loss-exceedance curve. It is like the hazard curve, except that the x -axis measures loss instead of excitation. Suppose one knows the hazard curve and the uncertain vulnerability function for a single asset. Figure 6 shows an example. The risk curve for a single asset can be calculated as

$$R(x) = \int_{s=0}^{\infty} (1 - P[X \leq x | S = s]) \frac{dG(s)}{ds} ds \tag{41}$$

where

X = uncertain degree of loss to an asset, such as the uncertain damage factor,
 x = a particular value of X ,
 s = a particular value of the excitation, such as the shaking intensity in terms of the 5 % damped spectral acceleration response at some index period of vibration,



Beginner's Guide to Fragility, Vulnerability, and Risk, Fig. 6 Sample risk curve. The *solid line* is the risk curve; the *dotted line* shows that the asset experiences a damage factor of 0.13 with an average exceedance rate of 0.004

$R(x)$ = annual frequency with which loss of degree x is exceeded,

$G(s)$ = the mean annual frequency of shaking exceeding intensity s ,

$P[X \leq x | S = s]$ = cumulative distribution function of X evaluated at x , given shaking s

If X is lognormally distributed at $S = s$, then

$$P[X \leq x | S = s] = \Phi \left(\frac{\ln(x/\theta(s))}{\beta(s)} \right) \tag{42}$$

where

$\theta(s)$ = median vulnerability function, i.e., the value of the damage factor with 50 % exceedance probability when the asset is exposed to excitation s ,

$v(s)$ = coefficient of variation of vulnerability, i.e., the coefficient of variation of the damage factor of the asset exposed to excitation s ,

$\beta(s)$ = logarithmic standard deviation of the vulnerability function, i.e., the standard deviation of the natural logarithm of the damage factor when the asset is exposed to excitation s .

If one has the mean vulnerability function $y(s)$ and coefficient of variation of loss as a function of shaking $v(s)$, use Eqs. 6 and 7 to evaluate $\theta(s)$ and $\beta(s)$.

Suppose the analyst has $y(s)$, $v(s)$, and $G(s)$ at a number n of discrete values of s , denoted here by s_i , where i is an index $i \in \{1, 2, \dots, n\}$. One can numerically integrate Eq. 41 by

$$\begin{aligned} R(x) &= \sum_{i=1}^n \left(p_{i-1}(x) G_{i-1} (1 - \exp(m_i \Delta s_i)) \right. \\ &\quad \left. - \frac{\Delta p_i(x)}{\Delta s_i} G_{i-1} \left(\exp(m_i \Delta s_i) \left(\Delta s_i - \frac{1}{m_i} \right) + \frac{1}{m_i} \right) \right) \\ &= \sum_{i=1}^n (p_{i-1}(x) \cdot a_i - \Delta p_i(x) \cdot b_i) \end{aligned} \quad (43)$$

where

$$\begin{aligned} p_i(x) &= P[X \geq x | S = s_i] \\ &= 1 - \Phi \left(\frac{\ln(x/\theta(s_i))}{\beta(s_i)} \right) \end{aligned} \quad (44)$$

$$\Delta p_i(x) = p_i(x) - p_{i-1}(x) \quad (45)$$

$$\Delta s_i = s_i - s_{i-1} \quad m_i = \ln(G_i/G_{i-1})/\Delta s_i \quad \text{for } i = 1, 2, \dots, n$$

$$\begin{aligned} a_i &= G_{i-1} (1 - \exp(m_i \Delta s_i)) \\ b_i &= \frac{G_{i-1}}{\Delta s_i} \left(\exp(m_i \Delta s_i) \left(\Delta s_i - \frac{1}{m_i} \right) + \frac{1}{m_i} \right) \end{aligned}$$

Equation 43 is exact if $p(x)$ and $\ln G(s)$ vary linearly between values of s_i .

Probable Maximum Loss for a Single Asset

There is no universally accepted definition of probable maximum loss (PML) for purposes of earthquake risk analysis, but it is often understood to mean the loss with 90 % nonexceedance probability given shaking with 10 % exceedance probability in 50 years. For a single asset, PML can be calculated from the seismic vulnerability function by inverting the conditional distribution

of loss at 0.90, when conditioned on shaking with 10 % exceedance probability in 50 years.

For example, assume that loss is lognormally distributed conditioned on shaking s , with median $\theta(s)$ and logarithmic standard deviation $\beta(s)$ as described near Eq. 42, which are related to the mean vulnerability function $y(s)$ and coefficient of variation $v(s)$ as in Eqs. 6 and 7. Under the assumption of Poisson arrivals of earthquakes, shaking with 10 % exceedance probability in 50 years is the shaking with exceedance rate $G(s_{PML}) = 0.00211$ per year, so PML can be estimated as a fraction of value exposed by

$$PML = \theta(s_{PML}) \cdot \exp(1.28 \cdot \beta(s_{PML})) \quad (46)$$

where $s_{PML} = G^{-1}(0.00211 \text{ year}^{-1})$, that is, the hazard curve (events per year) inverted at 0.00211.

Conclusions

This entry has provided a primer on fragility, vulnerability, and risk for the student or professional who is new to the topic. It briefly introduced basic concepts of fragility, vulnerability, hazard, and risk. Fragility is used to relate excitation, such as ground motion, to the probability that some undesirable event will occur, such as the probability that a component will become nonfunctional or that a building will be damaged. In some situations, multiple fragility functions can apply to the same asset; the entry presented three distinct situations and presented several approaches to deriving fragility functions and some useful sources of fragility functions. Vulnerability relates excitation to the degree of loss, such as the uncertain repair cost to a building. Three distinct approaches exist to deriving vulnerability functions: empirical, analytical, and expert opinion, each with its advantages and disadvantages. This entry briefly summarized the approaches, though it did not provide detail.

Hazard relates the exceedance probability or exceedance frequency to various levels of excitation, such as the rate at which earthquakes occur causing or exceeding various levels of spectral

acceleration response. The entry introduced probabilistic seismic hazard analysis, distinguished probability from exceedance rate, summarized some leading measures of excitation and conversion between them, explained hazard deaggregation, and offered some convenient sources of hazard data. Risk relates the degree of loss to its rate of exceedance, or it can express long-term average loss. This entry presented methods for estimating and depicting risk for a single asset.

References

- American Society of Appraisers (2013) Definitions of value relating to MTS assets, Machinery & Technical Specialties Committee of the American Society of Appraisers. Reston VA, <http://www.appraisers.org/Disciplines/Machinery-Technical-Specialties/mts-appraiser-resources/DefinitionsOfValue>. Accessed 18 May 2015
- Ang AHS, Tang WH (1975) Probability concepts in engineering planning and design, 1 – basic principles. Wiley, New York, 409 pp
- Applied Technology Council (ATC) (1985) ATC-13, Earthquake Damage Evaluation Data for California. Redwood City, CA, 492 pp
- Applied Technology Council (ATC) (2012) Seismic performance assessment of buildings volume 1 – methodology FEMA P-58-1. Federal Emergency Management Agency, Washington, DC. <http://goo.gl/QN8AQz>
- Atkinson GM, Kaka SI (2007) Relationships between felt intensity and instrumental ground motion. *Bull Seismol Soc Am* 97:497–510
- Baker JW, Cornell CA (2005) A vector-valued ground motion intensity measure consisting of spectral acceleration and epsilon. *Earthq Eng Struct Dyn* 34(10):1193–1217
- Bazzurro P, Cornell A (1999) Disaggregation of seismic hazard. *Bull Seismol Soc Am* 89(2):501–520
- Cornell CA (1968) Engineering seismic risk analysis. *Bull Seismol Soc Am* 58(5):1583–1606
- European Seismic Commission Working Group—Macroseismic Scales (1998) European macroseismic scale 1998 EMS-98. Luxembourg. http://www.gfz-potsdam.de/pb5/pb53/projekt/ems/eng/index_eng.html. Accessed 17 July 2006
- Faenza L, Michelini A (2010) Regression analysis of MCS intensity and ground motion parameters in Italy and its application in ShakeMap. *Geophys J Int* 180:1138–1152
- Field EH, Dawson TE, Felzer KR, Frankel AD, Gupta V, Jordan TH, Parsons T, Petersen MD, Stein RS, Weldon RJ II, Wills CJ (2007) The Uniform California earthquake rupture forecast, version 2 (UCERF 2). USGS open file report 2007–1437
- Jaiswal KS, Aspinall WP, Perkins D, Wald D, Porter KA (2012) Use of expert judgment elicitation to estimate seismic vulnerability of selected building types. *Proc 15th World Conference on Earthquake Engineering*, September 2012, Lisbon, Portugal
- Johnson GS, Sheppard RE, Quilici MD, Eder SJ, Scawthorn CR (1999) Seismic reliability assessment of critical facilities: A handbook, supporting documentation, and model code provisions, MCEER-99-0008. Multidisciplinary Center for Earthquake Engineering Research, Buffalo, NY, 384 pp
- Kaestli P, Faeh D (2006) Rapid estimation of macroseismic effects and ShakeMaps using macroseismic data. In: *Proceedings of First European conference on earthquake engineering and seismology*, Geneva
- Kaka SI, Atkinson GM (2004) Relationships between instrumental ground-motion parameters and Modified Mercalli intensity in Eastern North America. *Bull Seismol Soc Am* 94(5):1728–1736
- McGuire RK (1995) Probabilistic seismic hazard analysis and design earthquakes: closing the loop. *Bull Seismol Soc Am* 85(5):1275–1284
- Murphy JR, O'Brien LJ (1977) The correlation of peak ground acceleration amplitude with seismic intensity and other physical parameters. *Bull Seismol Soc Am* 67:877–915
- National Institute of Building Sciences and Federal Emergency Management Agency (NIBS and FEMA) (2009) Multi-hazard loss estimation methodology, Earthquake Model, HAZUS[®] MH MR4 Technical Manual. Federal Emergency Management Agency, Washington, DC
- NIST/SEMATECH (2013) e-Handbook of statistical methods. <http://www.itl.nist.gov/div898/handbook/>. Accessed 3 Mar 2015
- Porter KA (2009a) Cracking an open safe: HAZUS vulnerability functions in terms of structure-independent spectral acceleration. *Earthq Spectra* 25(2):361–378
- Porter KA (2009b) Cracking an open safe: more HAZUS vulnerability functions in terms of structure-independent spectral acceleration. *Earthq Spectra* 25(3):607–618
- Porter KA (2010) Cracking an open safe: uncertainty in HAZUS-based seismic vulnerability functions. *Earthq Spectra* 26(3):893–900
- Porter KA (2014) A beginner's guide to fragility vulnerability and risk, with solved exercises. Resource document. <http://spot.colorado.edu/~porterka/Porter-beginners-guide.pdf>. Accessed 28 Apr 2014
- Porter KA, Kennedy RP, Bachman RE (2007) Creating fragility functions for performance-based earthquake engineering. *Earthq Spectra* 23(2):471–489
- Porter K, Farokhnia K, Vamvatsikos D, Cho I (2014) Analytical derivation of seismic vulnerability functions for highrise buildings, global vulnerability

- consortium. Available from www.nexus.globalquakemodel.org/gem-vulnerability/posts/
- Sørensen MB, Stromeyer D, Grünthal G (2007) Deliverable 4.1: generation of area-specific relationships between ground motion parameters (PGA, PGV) at certain sites, magnitude M and distance R to the causative fault and site intensities in terms of EMS-98. Databank of intensity data points and related parameters, seismic early warning for Europe. GFZ Potsdam 19–32, Potsdam, Germany
- Trifunac MD, Brady AG (1975) On the correlation of seismic intensity scales with the peaks of recorded strong ground motion. *Bull Seismol Soc Am* 65:139–162
- Tselentis GA, Danciu L (2008) Empirical relationships between modified Mercalli intensity and engineering ground-motion parameters in Greece. *Bull Seismol Soc Am* 98:1863–1875
- Wald DJ, Allen TI (2007) Topographic slope as a proxy for seismic site conditions and amplification. *Bull Seismol Soc Am* 97:1379–1395
- Wald DJ, Quitoriano V, Heaton TH, Kanamori H (1999) Relationships between peak ground acceleration, peak ground velocity and modified Mercalli intensity in California. *Earthq Spectra* 15(3):557–564
- Wesson RL, Perkins DM, Leyendecker EV, Roth RJ, Petersen MD (2004) Losses to single-family housing from ground motions in the 1994 Northridge, California, earthquake. *Earthq Spectra* 20(3):1021–1045
- Wills CJ, Clahan KB (2006) Developing a map of geologically defined site-conditions categories for California. *Bull Seismol Soc Am* 96(4A):1483–1501
- Wood HO, Neumann F (1931) Modified Mercalli intensity scale of 1931. *Bull Seismol Soc Am* 21:277–283
- Worden CB, Gerstenberger MC, Rhoades DA, Wald DJ (2012) Probabilistic relationships between ground-motion parameters and modified Mercalli intensity in California. *Bull Seismol Soc Am* 102(1):204–221

Behavior Factor and Ductility

Christos A. Zerbis

Department of Structural Engineering, Faculty of Civil Engineering, National Technical University of Athens, Athens, Greece

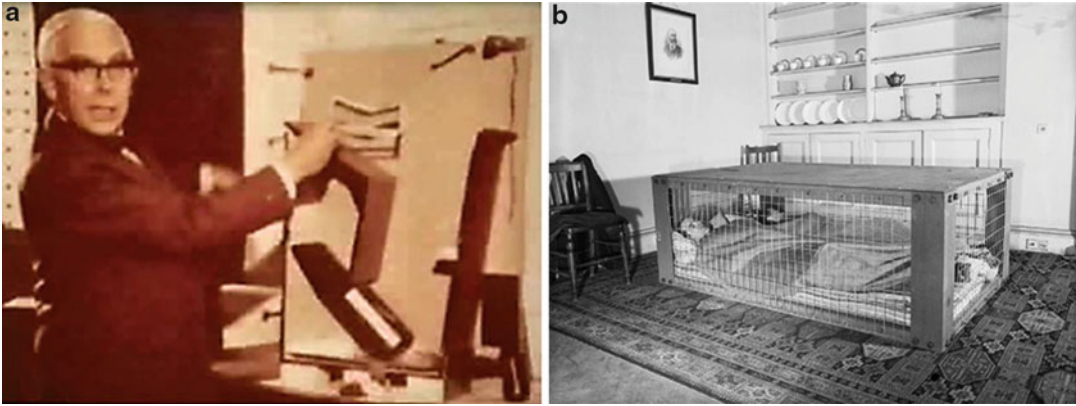
Synonyms

Behaviour factor; Damage; Ductility; Inelastic design; Reinforced concrete; Response spectrum; Seismic design

An Introduction: Evolution of Inelastic Design and Justification for Using a Behavior Factor for Extreme Transient Load Designs

The force-based design philosophy adopted by current design codes prescribes that the design of the structural system and the sizing and reinforcing of its structural elements are developed under prescribed equivalent statically applied forces. Structural design against seismic loads follows an ultimate limit state (ULS) design, since the excitation levels considered have very low probabilities of being exceeded. In this context, limit analysis of the structures is enforced possibly using a load factor and unfactored resistance approach or, in most cases, a partial factor – load and resistance factor – design approach (e.g., EC8 2004). Historically, earlier generations of seismic design codes adopted throughout the world a serviceability level approach, whereby the prescribed seismic loads were applied to the structure unfactored and the corresponding verification at the member level was evaluated using linear elastic analysis – including cracking, for reinforced concrete (RC) structures – and code prescribed allowable stress for the structural materials (concrete, reinforcement, structural steel or timber). These allowable stress limits were a factored percentage of the actual material strength and represented the safety factor built in the design inequality, since the loads were unfactored, thereby justifying linear elastic analysis. Such an allowable (working) stress force-based design has been the traditional design approach up until the 1970s in most seismically affected countries.

Direct inelastic design through limit analysis through the relaxation of the prevailing design assumption of an elastic response, accepting the fact that the structure can survive an extreme transient action by entering into the inelastic range in a controlled damage response, is traced in protective civilian designs against military action during World War II in England. Although plasticity theory, namely, the mathematical formulation of the ideal plastic flow of metal-type



Behavior Factor and Ductility, Fig. 1 (a) Baron J. Baker demonstrating the effectiveness of a model Morrison shelter to survive a collapsed building floor following a bomb hit, with controlled inelastic deformation in order to ensure survival of its occupants (http://en.wikipedia.org/wiki/Baron_Baker, also: <http://www-g.eng.cam.ac.uk/125/1925-1950/baker6.html>). (b) Picture of a Morrison shelter deployed inside a house for civilian protection (http://en.wikipedia.org/wiki/Morrison_shelter#cite_note-21)

materials, was well developed in the early nineteenth century, it was at this time that through the formulation of the plasticity theorems that limit analysis design methods for reduced loading capacity were adopted, for the design of simple steel structures for civilian protection under extreme blast loads (Fig. 1). Specifically for seismic loading and following the observation/evaluation of the response of simple structures following severe earthquakes in California in the 1950s, it was demonstrated by Housner (1956) that a sufficient inelastic deformation supply by the structure through controlled (hierarchical) underdesign of the elements and local detailing resulted in these structures behaving in an acceptable manner during the extreme earthquake event.

Taking therefore into account the transient time-varying nature of the loading and of the response, the type of the loading and inertia effects, and the extreme magnitude of the load intensity, a reduction factor was proposed to be used in order to reduce the design force (therefore resistance) of the structure to levels (often considerably) lower than those required for full elastic response, at the expense of increased inelastic global (structural) and local (element, section) deformations of a controlled magnitude and distribution. This reduction is enforced through

recognizing (and thereby enforcing) the fact that such an intentionally (or accidentally) underdesigned structure with adequate ductility capacity will survive the extreme loading event scenario with tolerable inelastic deformations, to a level that is acceptable for the safety of its occupants and equipment. As an extension of this initial collapse limitation design approach adopted in the 1970s, additional intermediate performance requirements for occupants and equipment have been adopted, leading to a complete performance-based design (PBD) philosophy for the safety and/or operability of the structure, occupants, and equipment.

In order to reduce the design load for elastic response, a factor greater than 1.0 is adopted, which, depending on the code, is denoted as: the behavior factor q (Europe), the response reduction coefficient or response modification factor R (United States and the Americas), the structural performance factor S_p (New Zealand), or the structural characteristic factor D_s (Japan), all these factors denoted as the behavior factor q herein, for simplicity. The magnitude of q , obtained for (design) or by (assessment) the quantification of the limiting structural resistance versus the magnitude of the corresponding inelastic global or local element, section, or material inelastic deformations (axial or shear strains, flexural rotations,

and so on) – expressed as the corresponding ductilities μ – and the interrelation of these two, is the topic of discussion of this entry.

The Behavior Factor: Its Types, Definitions, and Uses in Seismic Design

Behavior Factor for Ductility q_μ of SDOF Equivalent Structural Systems

The behavior factor for ductility (denoted as q_μ) of an inelastic SDOF system is defined, for a particular base excitation $g(t)$, as the ratio of its yield resistances R_y in order to develop two ductility levels, a prescribed ductility level $\mu = \mu_i$ and a ductility $\mu = 1.0$, namely, entirely elastic response (Eq. 1):

$$q_\mu(T_0, \xi, f(R, \mu), g(t), \mu_i) = \frac{R_y(\max|\mu| = 1.0)}{R_y(\max|\mu| = \mu_i)} \quad (1)$$

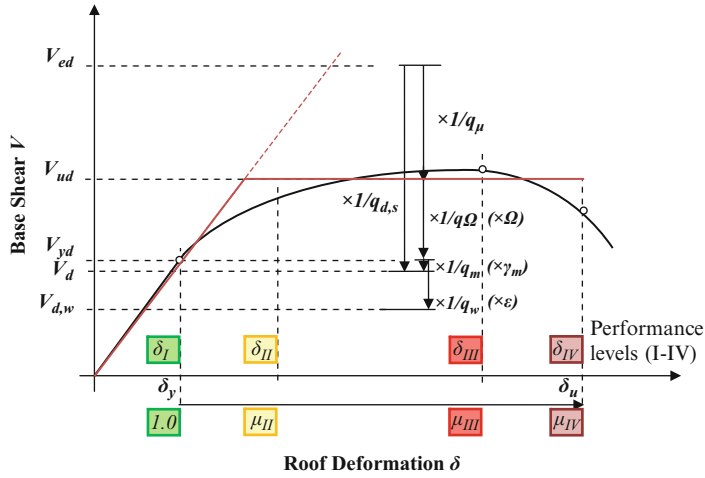
where q_μ , besides $g(t)$, depends on the initial elastic response parameters, namely, elastic period T_0 and critical damping ratio ξ , and on the parameters characterizing the inelastic response, namely, R_y and the cyclic hysteretic response shape $f(R, \mu)$.

Both terms of the fraction above are evaluated from time history integration of the equation of motion under the base excitation $g(t)$. Solution of the equation over a range of periods T_0 yields the ductility spectrum for the given earthquake excitation, also denoted as the q_μ – μ – T spectrum of this excitation. For design applications, such spectra are evaluated for an ensemble of base excitation inputs characterizing the local seismicity of the region and local soil effects; the resulting average plus percent fractile q_μ – μ – T spectrum is obtained for the design and evaluation of structures in the region exhibiting compatible hysteretic response characteristics and initial damping. The definition of q_μ and its dependence on structural and excitation parameters is discussed in section “Evaluation of the Behaviour Factor Due to Ductility q_μ of SDOF Systems” herein.

Behavior Factors for Seismic Design or Evaluation and Redesign of Structural Systems, q_d and q_s

The earliest use of a ductility reduction coefficient can be traced to the seismic design regulations in California, in which a force reduction coefficient was adopted depending on the building properties, namely, seismic resisting system and construction material. The 1959 edition of the Structural Engineers Association of California (SEAOC) Blue Book and its successive editions (e.g., SEAOC 1974) used this horizontal force factor K , in the context of an allowable stress design procedure, for the estimation of the design base shear. ATC (1978) was the first seismic design code that adopted an R factor greater than 1.00 as a divisor of the elastic design response spectrum (EDRS) forces, which depended on the type of structural system but was, as yet, independent of the building period. The early uses of R were based not only on SDOF analysis but, primarily, on consensus values through the earthquake engineering design community, with typical values of R and comparisons among different codes given in ATC (1995). Since then, the reduction coefficient R – initially as R_w , for codes based on working stress design methods such as UBC (1988) – has been adopted in all ultimate level design codes (e.g., NEHRP 1985; FEMA-356 2000; ASCE 2013). In Europe, Eurocode 8, from the initial stages of development (EC8 ENV 1988) up to the final completion (EC8 2004), also adopted the behavior factor q . To date, most modern force-based seismic design codes adopt the behavior factor approach for quantifying the seismic design forces, with different levels of rationalization for the magnitude of q (R) (see, for instance, an overview of R factors adopted in South American seismic design codes in Chavez et al. 2012).

Based on the way the behavior factor is applied in seismic design, one can identify two uses of q , namely: (i) q_d – specified for the design of new buildings in order to control the ductility demand (and thus damage level) for a given site seismicity and performance level. This is specified in the seismic codes as the maximum allowed behavior factor for the design of modern



Behavior Factor and Ductility, Fig. 2 Definition of the design and supplied (available) behavior factors q_d and q_s and the contributing q , for non-base-isolated structures: (i) q for ductility (q_μ), (ii) q for structural overstrength (q_Ω), and (iii) q for material strength reduction (q_m or q_w).

In the context of PBD, $q_{s,d}$ are no longer unique and depend on the local (and therefore global) ductility demands μ_I to μ_{IV} for different performance levels I to IV , respectively

structures. (ii) q_s – supplied (also called available) behavior factor of a given building (new or existing), corresponding to an anticipated elastic demand defined by the building period and the local EDRS, for a given ductility capacity and therefore damage level at specified performance levels (I – IV , Fig. 2) of this structure; for existing structures designed by past code generations, this behavior factor is usually below the currently enforced design q_d due to de facto underdesign coupled with insufficient or complete lack of ductile response of such structures.

Behavior Factor q_d for Seismic Design of New Structures

In the context of force-based seismic design, the design behavior factor q_d is a design force reduction coefficient greater than 1.0 by which the EDRS-specified base shear V_{ed} of the structure's equivalent SDOF system representation (Fig. 2) is divided, in order to establish the seismic design base shear level V_d of this structure (Eq. 2), expressed as the product of the seismic coefficient C_{sd} times the inertial weight W of the structure for the seismic load combination:

$$V_{yd} = \frac{V_{ed}}{q_d} = \frac{C_{sd}W}{q_d} \quad (2)$$

where the EDRS is representative of the local seismicity, earthquake return period, and system equivalent damping. Modern design typically codes specify for RC buildings the 5 % EDRS, giving suitable adjustment multipliers for systems with other damping coefficients such as steel, timber, or prestressed concrete (e.g., the EDRS modification coefficient $\eta = \sqrt{10/(5 + \xi)} \geq 0,55$ in EC8 (2004), where ξ is the percent critical damping).

Considering the typical base shear versus roof deformation response under increasing lateral load in Fig. 2, it can be seen that q_d can be expressed (Uang 1991; Whittaker et al. 1999) as the product of four behavior factors that relate the elastic base shear demand of the EDRS to the design base shear force as follows (Eq. 3):

$$q_d = q_\mu \cdot q_\Omega \cdot q_\xi \cdot \{q_m \text{ or } q_w\} \quad (3)$$

where:

- (i) q_μ is the behavior factor for ductility of the equivalent SDOF system, $q_\mu = V_{ed}/V_{ud}$: this

factor relates the elastic spectral base shear demand V_{ed} to the equivalent SDOF bilinear yield strength V_{ud} , obtained by assuming an elastoplastic equal area approximation of the inelastic base shear–roof displacement response under a given lateral load profile, up to the ultimate deformation δ_u . The value of this factor varies with the base excitation(s), the dynamic and inelastic hysteretic characteristics of the system (including structural damping), and its ductility. It corresponds to the basic design behavior q_s in EC8 (2004): for concrete structures q_s takes values between $1.5 < q_s < 3.0$ and $2.0 < q_s < 4.5$, for ductility classes medium and high, respectively. q_μ is established from SDOF analysis, as described in section “Evaluation of the Behaviour Factor Due to Ductility q_μ of SDOF Systems” of this work.

- (ii) q_Ω is the behavior factor for structural overstrength, $q_\Omega = V_{ud}/V_{yd}$, relating V_{ud} to the base shear V_{yd} at incipient onset of yield within the structural elements, at a roof deformation δ_y . The value of this factor, which in EC8 (2004) is defined as the ratio α_u/α_1 , accounts for the overstrength of the structure and depends on the type of structural system and the degree of redundancy built in the system: for instance, for a typical soft ground story pilotis structure with limited redistribution of inelastic actions in the members of the upper stories, this factor is close to 1.0. For medium-rise RC frame buildings, however, where seismic loads govern the design and members exhibit sufficient local ductility for redistribution, q_Ω is of the order of 1.5–1.8, while for fully infilled RC frames with good-quality masonry construction, q_Ω can easily reach values of 2.50 and higher (Repapis et al. 2006b). For low-rise RC frame buildings, overstrengths of 4.0 have been reported. This factor can be quantified through static or dynamic inelastic analysis

of the structure. Its magnitude and mode of evaluation are discussed in section “Building Behaviour Factor q_s ” herein.

- (iii) q_ξ is the behavior factor contribution due to increased critical damping ξ ; this contributing factor is meaningful for base-isolated or artificially damped systems, since the contribution of the initial mass proportional damping (rather than the hysteretic damping, which is included in q_u) is not as significant in non-base-isolated structures. This factor is therefore not considered further herein.
- (iv) q_m (or q_w) is the behavior factor due to material strength $q_m = V_{yd}/V_d$ or $q_w = V_{yd}/V_{dw}$, relating the base shear at incipient member yield V_{yd} to the design base shear level specified by the code V_d (for ULS design) or $V_{d,w}$ (for working stress design). This factor is practically the ratio of the material strength partial factor in load and resistance factor design, in order to relate the design material stress level (for EC2 2004, the characteristic strength f_{yk}) to the yield strength (f_{yd}) of the material controlling the response, namely, the steel reinforcement in modern structural RC designs; for EC2 (2004), $q_m = f_{yk}/f_{yd} = 1.15$. For buildings in Greece constructed during the 1970s using reinforcement with a yield strength of 400 MPa and designed for an allowable stress in flexural design calculations of 230 MPa with a 20 % increase for the seismic load combination, $q_m = 400/1.2 \cdot 230 = 1.45$ (Repapis et al. 2006a). It has been proposed that the factor q_m may also include strength gain due to strain rate effects, although this increase is not a factor to take into account since it is counterbalanced by strength reduction by cyclic degradation.

This behavior factor is implicitly taken into account in the design of new buildings and is of primary importance since it defines the

transition from allowable design level to peak member resistance in the assessment of existing buildings, in which their members may behave in a brittle manner with very little deformability reserves.

Behavior Factor q_s Supplied by an Existing Structure Under a Given EDRS

In the context of performance-based design (PBD), the available behavior factor q_s supplied by an existing structure can also be established for a given EDRS demand and at specific performance limits of damage. The structure may have been designed using a prescribed design behavior factor q_d (different than q_s) or none at all, as the case is for older existing buildings which have been designed using past force-based design philosophies (e.g., working stress).

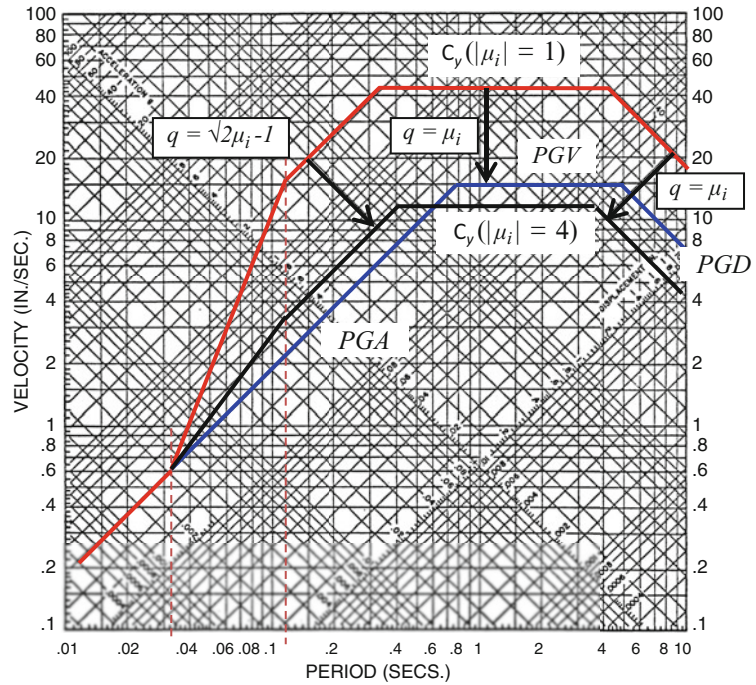
The available behavior factor q_s of the structure *as is* (often reevaluated after strengthening and/or rehabilitation of the structure at hand) is evaluated (i) through dynamic analyses of the structure until the establishment of a collapse base excitation, which is then compared to the design base excitation of the currently enforced EDRS for this type of structural system, or (ii) through static inelastic analyses of the structure, using an often idealized bilinear approximation of the resulting base shear resistance–roof deformation capacity curve under an imposed lateral deformation. Following an initial idealization of the structure as an equivalent SDOF system, with a yield resistance equal to the base shear at the onset of yielding in the building and an effective period equal to the initial secant stiffness from the lateral pushover curve, q_s is the behavior factor which results in a ductility demand equal to the ductility actually supplied by the system equivalent SDOF system. Contributions to q_s in this case are accounted from the structure overstrength or the member overstrength due to material q_m , relating the yield strength (also obtained from in situ evaluation of material capacity) to the design strength.

Evaluation of the Behavior Factor Due to Ductility q_μ of SDOF Systems

The response of nonlinear SDOF systems under seismic excitation has traditionally been formulated in the form of $q_\mu-\mu-T$ (or $R_\mu-\mu-T$) spectra, representing an extension of earlier uses of such spectra, relating the ratio of yield to elastic resistance of the SDOF system (q_μ) to the corresponding displacement ductility μ , under blast-type loads (Biggs 1964). Such inelastic response spectra (also known as shock spectra) have been formulated for different time-varying forcing shapes, such as impulse, triangular, or half sine, and are widely used in military and protective designs for blast. As discussed in the introduction, the use of such design load reductions has been well established in the design of protective structures against blast, where the introduction of a factored elastic design for controlled inelastic deformation was initially introduced.

The mathematics for the evaluation of the dynamic response of SDOF systems under seismic excitation is well established in textbooks of structural dynamics (Clough and Penzien 1975) and is not covered herein. The peak ductility demand μ of a given inelastic system under a certain type of base (or force) excitation is obtained by step-by-step integration using explicit or implicit time integration procedures, depending on the duration of the forcing function. From the way the problem is posed in Eq. 1, given the SDOF characteristics (R_y, T_0, ξ, f, g), only the peak ductility can readily be estimated. Consequently, the numerical estimation of the $q_\mu-\mu-T$ for a prescribed ductility $\mu = \mu_i$ is necessarily an iterative process, involving successive iterative estimates of R_y and solution of the equation of motion to establish $R_y(\max|\mu| = \mu_i)$, for which the SDOF system attains a peak ductility $\mu = \mu_i$, e.g., using a variable bound bisection iterative solution scheme, to within a user-specified tolerance. Subsequent application of the procedure for $\mu = 1$ as well (elastic response) will establish

Behavior Factor and Ductility, Fig. 3 Indirect methods: Variation of q with period for ductility $\mu_i = 4$ from the construction of the EDRS and the IDRS spectrum of the design base shear coefficient C_y , according to the graphic methodology by Newmark and Hall (1973)



$R_y(\max|\mu| = 1)$ and q_μ . It should be noted that in the case of strongly softening systems which are representative of structures with severe second-order effects or strength degradation, such an iterative solution may not converge, implying that the SDOF system may never attain this target ductility.

Indirect Evaluation Methods

For seismic design, the evaluation of the behavior factor through the formulation of an inelastic design response spectrum (IDRS) shape for a given seismicity condition – expressed in the form of a peak ground acceleration (PGA), ground velocity (PGV), and ground displacement (PGD) at the site – has initially been based on indirect methods. Such methods specified amplification factors for the establishment of an EDRS and subsequent spectral reduction factors (i.e., period-dependent behavior factors) for constructing the IDRS. This approach was initially recognized by Veletsos and Newmark (1960), yet, later on, Newmark and Hall (1973) formulated the procedure for the construction of

the IDRS. Such indirect methods essentially provide $q_\mu(\mu, T)$ relations, as a function of target μ_i , for an assumed damping and form of inelastic hysteretic response (initially an elastic perfectly plastic (EPP) oscillator was considered).

Following estimations of SDOF inelastic response assuming EPP behavior and using limited accelerogram data available at the time, Newmark and Hall (1973) observed that (Fig. 3) (i) elastic and inelastic SDOF systems with sufficiently medium-to-long periods (medium-to-low frequencies) tend to displace equally and, consequently, the behavior factor is equal to the required ductility, (ii) elastic and inelastic SDOF systems with very short period (very high frequency) have spectral accelerations (and thus force demands) equal to the PGA and should be designed to remain elastic, and (iii) in between, the energy absorbed by an inelastic system having a yield strength $1/q_\mu$ of the elastic demand and, as a consequence, deforming to a target μ_i is nearly equal to the kinetic energy absorbed by the elastic SDOF system, thereby leading to an equal energy criterion between q_μ and μ_i (Fig. 3).

Behavior Factor and Ductility, Table 1 Period dependent response reduction coefficients q_μ following the Newmark and Hall (1973) procedure

Soil class	I			II			III		
EDRS region	<i>D</i>	<i>V</i>	<i>A</i>	<i>D</i>	<i>V</i>	<i>A</i>	<i>D</i>	<i>V</i>	<i>A</i>
$\mu_i = 1.5$	1.62	1.70	1.44	1.58	1.66	1.50	1.78	1.94	1.47
$\mu_i = 2.0$	2.28	2.21	1.76	2.16	2.19	1.83	2.46	2.70	1.79
$\mu_i = 5.0$	5.63	4.59	2.84	5.04	4.63	2.91	5.82	5.96	2.85

They therefore proposed the indirect construction of an IDRS from an EDRS, for a given target

ductility μ_i , using the following period-dependent response reduction coefficients q_μ (Eq. 4):

$$\begin{aligned}
 0 \leq T < T_1/10 & \quad q_\mu = 1 \\
 T_1/10 \leq T < T_1/4 & \quad q_\mu = \sqrt{2\mu_i - 1} \left(\frac{T_1}{4T} \right)^{2.513 \log \left(\frac{1}{\sqrt{2\mu_i - 1}} \right)} \\
 T_1/4 \leq T < T'_1 & \quad q_\mu = \sqrt{2\mu_i - 1} \\
 T'_1 \leq T < T_1 & \quad q_\mu = \mu_i \frac{T}{T_1} \\
 T_1 \leq T < T_2, \quad T_2 \leq T < 10 \text{ s} & \quad q_\mu = \mu_i
 \end{aligned} \tag{4}$$

where T_1 , T'_1 , and T_2 are the characteristic corner periods of the EDRS, defined from the peak ground response parameters (Newmark and Hall 1973):

$$\begin{aligned}
 T_1 &= 2\pi \frac{\phi_{ev} PGV}{\phi_{ea} PGA}, \quad T_2 = 2\pi \frac{\phi_{ed} PGD}{\phi_{ev} PGV}, \\
 T'_1 &= T_1 \cdot \frac{\sqrt{2\mu_i - 1}}{\mu_i}
 \end{aligned} \tag{5}$$

where ϕ_{ea} , ϕ_{ev} , ϕ_{ed} are the amplification factors defining the EDRS acceleration, velocity, and displacement from the corresponding PGA , PGV , and PGD for the site, defined by local seismicity considerations. Thus, in a log-log format, the Newmark and Hall q_μ - μ - T relation consisted of two constant q_μ regions (Eq. 4) followed by intermediate linear transitions, as shown in Fig. 3. Based on a statistical parameter evaluation, they recommended averaged relations of PGV/PGA (48 in/s/g) and PGD/PGV^2 (6 s/in.) and ϕ_{ea} , ϕ_{ev} , ϕ_{ed} equal to 2.6, 1.9, and 1.4, respectively, for the construction of the EDRS at 5 % damping; different amplification values were given for stiff or soft soil or, in addition to the average values above, for average plus or minus one standard deviation.

Successive improvements of the indirect q_μ - μ - T construction model were proposed by Newmark and colleagues (for a review of the literature, see Miranda and Bertero 1994) (i) using additional ground motions and including the vertical excitation, (ii) for wider sets of critical damping ratios ξ (2–10 %) and target ductilities μ_i (1–10), and (iii) for different hysteretic shapes $f(u, t)$.

Riddell (1995) extended the indirect method of evaluating q_μ for different soil classification conditions, by considering 72 accelerograms which were obtained during the 1985 Chile earthquake, initially classified according to the prevailing soil conditions as records on rock (I), firm (II), and medium stiffness (III) soil sites. Following statistical analysis, soil-dependent corner frequencies and amplification values Ψ_μ were proposed for obtaining the 5 % damping EDRS and the ductility-dependent IDRS, given the site PGA , PGV , and PGD , following the Newmark and Hall (1973) procedure. Based on the amplification factors reported, typical q_μ , evaluated as $q_\mu = \Psi_{\mu=1}/\Psi_\mu$ for $\mu_i = 1.5$, 2.0, and 5.0, are given below for the constant displacement (D), velocity (V), and acceleration (A) regions of the EDRS (Table 1):

Direct Evaluation Methods of q_μ

Since the indirect evaluation of the behavior factor intrinsically included the uncertainty of both the elastic spectral amplification as well as the inelastic reduction, the scatter between target and actual ductilities or strength demands was high. Furthermore, use of a constant q_μ over a period range or over the entire period range, as originally adopted in ATC (1978), was shown to be inadequate to limit the inelastic demands of SDOF systems, particularly for near-field earthquake events with severe velocity pulses (Mahin and Bertero 1981).

Consequently, direct methods of evaluating the behavior factor variation with period and ductility have been proposed, through statistical evaluation of SDOF time history analysis results. According to the direct estimation methodology, the elastic and inelastic response of a set of nonlinear SDOF oscillators over a period range was evaluated under a set of earthquake excitations and given oscillator characteristics, and the ratio of the required resistance of the system to remain elastic over that to develop a given ductility, as defined in Eq. 1, was evaluated in each case. The resulting variation of this ratio (q_μ) over the period range considered was statistically processed in order to establish $q_{\mu-\mu-T}$ functional relations that directly fitted the corresponding spectra for the specific record(s) and oscillator characteristics. Although average prediction functions were often reported, several studies also gave sensitivity analyses or even additional equations to obtain the mean plus standard deviation approximation of the data.

An examination of the indirect method above for the construction of the EDRS and IDRS reveals the physical requirements for the variation of q_μ with period for a given ductility μ_i (Newmark and Hall 1973):

- (i) For very stiff systems, the SDOF system cannot develop any ductility and is therefore elastic; consequently, any functional describing a $q_{\mu-\mu-T}$ variation must satisfy:

$$\lim_{T \rightarrow 0} q_\mu = 1$$

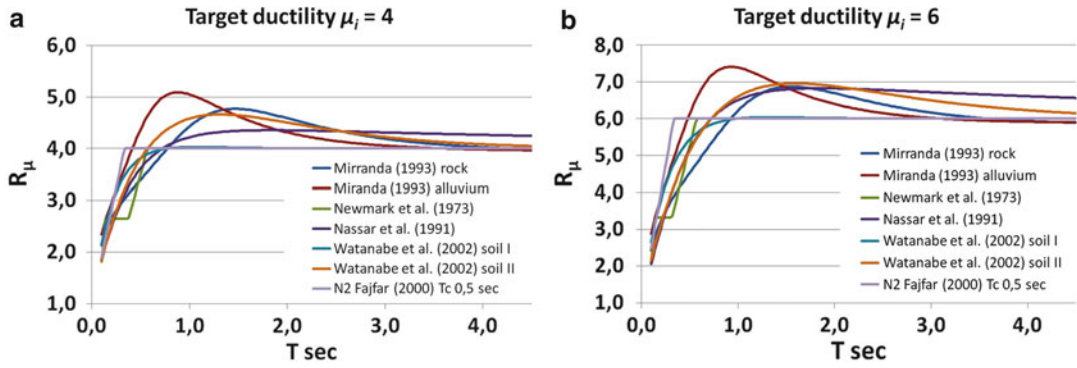
- (ii) For flexible long-period systems, the SDOF deformation of the elastic and the inelastic systems is equal to the imposed ground displacement; consequently q_μ is equal to the SDOF ductility:

$$\lim_{T \rightarrow \infty} q_\mu = \mu$$

- (iii) In between, in the so-called constant velocity region, energies of the elastic and the inelastic systems are comparable. Direct and indirect methods of establishing q_μ yield different requirements in this range.

Therefore, functional relations describing q_μ in terms of period and ductility should satisfy these requirements. Such statistical estimations of q_μ have shown that the latter limit is actually approached asymptotically from above (Nassar and Krawinkler 1991); however, the constraint imposed by the first limit has been relaxed in certain statistical studies in order to obtain better statistical correlations of $q_{\mu-\mu-T}$ in the short period (Riddell 1995). A brief description of several such functional relations and the corresponding parametric studies involved are given herein; a graphical comparison of the different $q_{\mu-\mu-T}$ spectral models for target ductilities $\mu_i = 4$ and 6 is given in Fig. 4, as proposed by these studies.

Influence of the SDOF oscillator inelastic characteristics. One of the earliest proposed $q_{\mu-\mu-T}$ expression that satisfied the functional limitations of the reduction factor spectrum by Newmark and Hall (1973) was the exponential function proposed by Nassar and Krawinkler (1991). In their statistical study, they considered 36 records to establish the attenuation of the ground motion parameters and 15 records from Western US earthquakes on firm-to-medium stiffness soils (plus 10 records from the Whittier Narrows earthquake, for verification) to establish q_μ . Two hysteretic shapes were considered to



Behavior Factor and Ductility, Fig. 4 Comparison of different q_μ - μ - T models for evaluating the behavior factor, for target ductilities of (a) $\mu_i = 4$ and (b) $\mu_i = 6$

model the hysteresis curve of the SDOF oscillator, namely, an EPP and a bilinear stiffness

degrading model. The q_μ - μ - T expression proposed is given in Eq. 6:

$$\begin{aligned}
 q_\mu(T, \mu, \alpha) &= (c \cdot (\mu_i - 1) + 1)^{1/c} \\
 \text{Elastoplastic systems } (\alpha = 0) \quad c(T) &= \frac{T}{T+1} + \frac{0.42}{T} \\
 \text{Elasto-hardening systems } (\alpha = 0.02) \quad c(T) &= \frac{T^{1.01}}{T^{1.01} + 1} + \frac{0.37}{T} \\
 \text{Elasto-hardening systems } (\alpha = 0.10) \quad c(T) &= \frac{T^{0.8}}{T^{0.8} + 1} + \frac{0.29}{T}
 \end{aligned} \tag{6}$$

The authors concluded that, with the exception of the short-period range, the stiffness degrading models with no hardening exhibited a 20 % higher behavior factor compared to the EPP system, with the effect diminishing for the 10 % hardening slope EPP and degrading oscillators. Furthermore, in addition to the quantification of the oscillator bias and based on the attenuation studies and the evaluation of the corresponding q_μ for the subject datasets, they further established that q_μ of SDOF systems for given μ_i is not sensitive to the epicentral distance of the earthquake and the ground motion; this observation, however, is considered in more detail further on.

Lee and Han (1999) developed direct expressions of q_μ for stiff soil and rock sites only, namely, soils exhibiting a shear-wave velocity larger than 750 m/s. In order to evaluate the

influence of the hysteretic response, they performed a statistical analysis of q_μ - μ - T using 40 ground motion records taking into account five different SDOF oscillator hysteretic shapes, namely, (i) EPP oscillator, (ii) bilinear hysteretic oscillator, (iii) degrading strength hysteretic oscillator, (iv) degrading stiffness oscillator (centered upon reversal), and (v) pinching hysteretic oscillator (in all cases initial damping of 5 % was assumed). They concluded that in the absence of the soil bias, additional factors that affect the q_μ demand (given target ductility and period) are the hardening slope and the form and extent of pinching of the oscillator cyclic characteristic, in a statistically independent manner to each other. They subsequently proposed a q_μ - μ - T functional in terms of the hysteretic parameters α_1 , α_2 , α_3 , and α_4 (Eq. 7):

$$\begin{aligned}
q_\mu &= (0.99 \cdot \mu + 0.15) \left(1 - e^{-23.69 \cdot T \cdot \mu^{-0.83}} \right) \cdot C_{\alpha 1} \cdot C_{\alpha 2} \cdot C_{\alpha 3} \cdot C_{\alpha 4}, \text{ with :} \\
C_{\alpha 1} &= 1.0 + (2.07 \cdot \ln(\mu) - 0.28) \cdot \alpha_1 - (10.55 \cdot \ln(\mu) - 5.21) \cdot \alpha_1^2 \\
C_{\alpha 2} &= \frac{1}{(0.2 \cdot \mu + 0.42) \cdot \alpha_2 + 0.005 \cdot \mu + 0.98} \\
C_{\alpha 3} &= \frac{0.85 + (0.03 \cdot \mu + 1.02) \cdot \alpha_3}{1 + (0.03 \cdot \mu + 0.99) \cdot \alpha_3 + 0.001 \cdot \alpha_3^2} \\
C_{\alpha 4} &= \frac{1}{1 + 0.11 \cdot e^{(1.4 \ln(\mu) - 6.6) \cdot \alpha_4}}
\end{aligned} \tag{7}$$

where α_1 measures the hardening slope, α_2 measures the strength degradation upon reversal, α_3 measures the stiffness degradation, and α_4 measures the pinching response.

Influence of the local soil conditions. Miranda (1993) and Miranda and Bertero (1994) considered a wide dataset of ground motion excitations which had been previously classified according to the relevant earthquake parameters (magnitude, distance from the fault) and the local soil conditions prevalent at the site. Statistical equations of q_μ in terms of the local soil conditions were subsequently obtained, proposing the following q_μ - μ - T evaluation function for different soil classifications (Eq. 8):

$$\begin{aligned}
q_\mu &= 1 + \frac{\mu_i - 1}{\Phi(T)} \\
\text{Rock sites} \quad \Phi(T) &= 1 + \frac{1}{10T - \mu_i T} - \frac{1}{2T} e^{-\frac{3}{2}(\ln T - 3/5)^2} \\
\text{Alluvium sites} \quad \Phi(T) &= 1 + \frac{1}{12T - \mu_i T} - \frac{2}{5T} e^{-2 \cdot (\ln T - 1/5)^2} \\
\text{Soft soil sites} \quad \Phi(T, T_s) &= 1 + \frac{T_s}{3T} - \frac{3T_s}{4T} e^{-3 \left(\ln \left(\frac{T}{T_s} \right) - 1/4 \right)^2}
\end{aligned} \tag{8}$$

where, apart from the parameters defined already, the dominant soil period T_s was introduced for the soft soil sites. The analysis showed that there is a distinct trend in the shape of the q_μ - μ curves for (primarily soft) soils that exhibit a predominant period, whereby in the vicinity of the oscillator resonance with the soil dominant period T_s ($T/T_s \sim 1$), the behavior factor q_μ is much larger than the target ductility (e.g., $q_{\mu i=3} = 8.5$ or $q_{\mu i=5} = 18.0$ was obtained); this effect led to

conservative designs based on the q_μ - μ - T relations obtained on firm ground or alluvia proposed by other studies, while for lower T/T_s values, the opposite was observed, with the reduction factors in this case being lower than μ_i (e.g., $q_{\mu i=3} = 2$ or $q_{\mu i=5} = 2.8$ at $T/T_s = 0.5$). Finally, for normalized periods above T_s ($T/T_s > 2$), the behavior factor was close to the target ductility, as for stiff soils and rock, yielding similar or conservative designs compared to the use of rock spectra.

In order to investigate the influence of stiffness degradation on soft soils, Miranda and Ruiz-Garcia (2002) extended this study by expressing q_μ in terms of the structural period normalized by the predominant soil period T_s (namely, q_μ - μ - T/T_s), using a set of 116 ground motions on soft-to-medium soils and two hysteretic systems, a bilinear EPP and a stiffness degrading. It was shown that the influence of stiffness degradation was significant for soft soil excitations, since for period ratios lower than soil resonance, in the range of $T/T_s \sim 2/3$ (also depending on μ_i), the average q_μ of stiffness degrading systems was about 25 % lower than q_μ of EPP systems, leading to unconservative designs based on the q_μ - μ - T relations obtained on firm ground for EPP systems. On the contrary, q_μ of stiffness degrading systems at soil resonant periods and above was higher than that for EPP systems, indicating that the earlier onset of nonlinearity in these structures induced a drop in the required R_y for a given target ductility compared to EPP structures.

Vidić et al. (1994) expressed q_μ versus ductility relations in site-specific form, expressed in

terms of the accelerogram's frequency parameter T_I , obtained at the period of change in slope of the idealized bilinear approximation of the pseudovelocity response spectrum $PS_v(T, \xi)$ of the accelerogram. For their analyses they considered 24 records from California, Chile,

Italy, Montenegro, and the 1985 Mexico City earthquake, analyzing SDOF oscillators with either EPP or stiffness degrading characteristics, yielding the more general q_μ - μ - T exponential expression of Eq. 9 (a tangent stiffness proportional damping equal to 5 % critical was used):

$$\begin{aligned}
 & T < T_0 & q_\mu &= c_1(\mu_i - 1)^{c_R} \cdot \frac{T}{T_0} + 1 \geq 1,0 \\
 & T_0 \leq T & q_\mu &= c_1(\mu_i - 1)^{c_R} + 1 \\
 & \text{where :} & T_0 &= c_2 \cdot \mu^{c_T} T_1 \\
 & \text{and :} & & \\
 & \text{stiffness degrading SDOF :} & c_1 & \quad c_2 \quad c_R \quad c_T \\
 & \text{elastoplastic SDOF :} & 0,75 & \quad 1,0 \quad 0,65 \quad 0,30 \\
 & & 1,10 & \quad 0,95 \quad 0,75 \quad 0,20
 \end{aligned} \tag{9}$$

In the context of PBD, Fajfar (1999) proposed more simplified conservative design envelopes for the estimation of the required target ductility

at a given q_μ (Eq. 10), by adapting the statistical q_μ - μ - T estimations by Vidić et al. (1994) above in the N2 method:

$$\begin{aligned}
 & T < T_0 & q_\mu &= (\mu_i - 1) \cdot \frac{T}{T_0} + 1 \geq 1,0 \\
 & T_0 \leq T & q_\mu &= \mu_i \\
 & \text{where :} & T_0 &= 0,65 \cdot \mu^{0,3} T_c \leq T_c \\
 & \text{or, in a simplified form :} & T_0 &= T_c
 \end{aligned} \tag{10}$$

This expression has been adopted with refinements (soil effect, softening or hardening oscillator characteristics) as the displacement coefficient method (DCM) for PBD (NEHRP 1985; FEMA-356 2000), where q_μ is equal to the coefficient C_I in the target roof estimation equation.

Riddell (1995) examined the influence of local soil conditions on q_μ using 72 records obtained during the Magnitude 7.8 1985 Chile earthquake and aftershocks on rock (I), hard (II), or medium soil (III), as also previously described in the indirect evaluation method for q_μ : following the indirect method analysis and recognizing that the variation in q_μ was high, direct evaluation

expressions for spectral amplification factors $\Psi(T, \mu)$ and therefore q_μ in terms of ductility and period were also proposed, as being more reliable and less prone to statistical scatter than the former indirect method. The function proposed was of the exponential form given in Eq. 11 using five statistical parameters (α_1 - α_5) – although a dominant soil period was not explicitly considered (type IV soft soil records were not available).

$$\begin{aligned}
 q_\mu &= \frac{S_a(T, \mu = 1)}{S_a(T, \mu = \mu_i)}, \text{ where} \\
 \Psi(T, \mu = \mu_i) &= \frac{\alpha_1 + \alpha_2 \cdot T^{\alpha_3}}{1 + \alpha_4 \cdot T^{\alpha_5}}
 \end{aligned}$$

Soil Type	$\mu_i = 2$					$\mu_i = 5$					(11)
	a_1	a_2	a_3	a_4	a_5	a_1	a_2	a_3	a_4	a_5	
I	0,45	5,81	0,66	18,0	2,04	0,60	4,75	0,86	37,3	2,03	
II	0,88	7,77	1,39	24,2	2,79	0,84	1,41	1,44	13,6	2,52	
III	-33,96	35,65	0,005	1,4	2,62	0,86	0,06	2,967	2,303	2,72	

The parameters were obtained from regression analysis after relaxing, for the sake of improved statistical correlation, the requirement that $\Psi(T, \mu)$ should be equal to 1.0 for rigid structures.

Ordaz and Pérez-Rocha (1998) considered the direct evaluation of q_μ as a force reduction parameter from the elastic relative displacement spectrum S_d , since both the inelastic and elastic displacements of the SDOF system are related to the ductility and q_μ , the variation of S_d with period being the controlling factor for the variation of q_μ , and also S_d

includes local soil effects. They therefore proposed a functional for q_μ in terms of the spectral displacement amplification S_d/PGD instead of T (Eq. 12) since, implicitly, it is a function of the period T . For their statistical analysis, they considered 445 records over a wide range of magnitudes, epicentral distance, and local soft soil to rock conditions.

$$q_\mu = 1 + (\mu_i - 1) \cdot \left(\frac{S_d(T)}{PGD} \right)^{\beta(\mu)}, \text{ where} \\ \beta(\mu) = 0.388 \cdot (\mu - 1)^{0,173}$$

Soil Type	$\mu_i = 2$			$\mu_i = 4$			$\mu_i = 6$			(12)
	a	b	c	a	b	c	a	b	c	
I	1,29	2,77	0,0218	1,12	2,18	0,0777	2,35	1,69	0,0080	
II	1,12	2,18	0,0416	,989	1,62	0,2037	1,52	1,05	0,1334	
III	2,35	1,69	0,0418	1,03	1,24	0,2707	1,85	,821	0,1184	

Their expression $q_{\mu-\mu-S_d(T)/PGD}$ yielded similar results as the proposed direct $q_{\mu-\mu-T}$ relations by other investigators (e.g., Nassar and Krawinkler (1991) or Miranda (1993)), for either firm or soft soil, yet, unlike previous studies, is of a more general applicability over the entire range of soil types considered.

Watanabe and Kawashima (2002) investigated in detail the scattering of q_μ from the mean value as well as the influence of initial damping under elastic and EPP inelastic response, using 70 free-field records from a relatively large dataset from Japanese earthquakes: ground motions were classified according to different soil conditions with predominant soil period T_g , following the Japanese seismic design

code classification. They showed that their predictions of the average $\bar{q}_\mu$ approach reasonably well the equal energy and equal displacement approximation (Fig. 3). However, the average minus one standard deviation (σ_{q_μ}) values differ significantly from this simplification and on the unconservative side, since (σ_{q_μ}) increases disproportionately, and therefore, the use of $\bar{q}_\mu$ for design may be unconservative, especially for larger target ductilities. They then proposed an exponential $q_{\mu-\mu-T}$ relation and the corresponding expression for σ_{q_μ} in Eq. 13:

$$q_\mu = 1 + (\mu_i - 1) \cdot \Psi(T), \text{ where} \\ \Psi(T) = 1 + c \cdot \frac{T - a}{e^{b(T-a)}}, c = \frac{1}{(a \cdot e^{ab})}$$

Soil Type	$\mu_i = 2$			$\mu_i = 4$			$\mu_i = 6$		
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
I) $T_g < 0.2\text{s}$	1, 29	2, 77	0, 0218	1, 12	2, 18	0, 0777	2, 35	1, 69	0, 0080
II) $T_g < 0.6\text{s}$	1, 12	2, 18	0, 0416	, 989	1, 62	0, 2037	1, 52	1, 05	0, 1334
III) $T_g > 0.6\text{s}$	2, 35	1, 69	0, 0418	1, 03	1, 24	0, 2707	1, 85	0, 821	0, 1184

B

and $\sigma_{q_\mu} \sim 0.4 \cdot \mu_i - 0.3$ (all soil classes)

Additional soil-dependent $q_{\mu}-\mu-T$ functions were proposed by Genshu and Yongfeng (2007) with a range of application of μ_i between 1.0 and 6.0 and a period range between 0.1 and 6.0 s. A dataset of 370 records was used, divided according to the prevailing soil conditions at the site and classified in accordance with the four soil-type classifications adopted in the Chinese seismic code. Four different hysteretic oscillator characteristics were considered, namely, (i) an EPP model, (ii) a bilinear with hardening (5 %, 10 %, and 20 % hardening stiffness) and pinching

hysteresis, (iii) a shear slip model, and (iv) a bilinear elastic oscillator, in order to establish the influence of energy absorption. Three critical damping ratios were considered, namely, undamped, 3.5 % critical, and 5 % critical. Following a sensitivity study, the resulting $q_{\mu}-\mu-T$ spectra were correlated to the ratio of T/T_g , where T_g is the characteristic ground motion period of each record, defined as the period at which q_μ was maximized. They therefore proposed expressions with T for the estimation of q_μ with 90 % confidence, $q_{\mu,90}$ (Eq. 14):

$$q_{\mu,90} = \begin{cases} 1 + 0.36 \cdot \left[(2.5 \cdot \mu_i - 2)^{0.75} - 1 \right] \cdot \left(\frac{T}{T_g} \right)^{(0,6 - T/T_g)}, & 0 < T \leq 0.6 T_g \\ 1 + \left[(2.5 \cdot \mu_i - 2)^{0.75} - 1 \right] \cdot \left(\frac{T}{T_g} \right)^2, & 0.6 T_g < T \leq T_g \\ (2.5 \mu_i - 2)^{0.75} - \left[(2.5 \mu_i - 2)^{0.75} - 0.5 - 0.6 \mu_i \right] \left(\frac{T_g - 1}{\chi - 1} \right)^{2/\mu_i}, & T_g < T \leq \chi T_g \\ 0.5 + 0.6 \cdot \mu & , \chi \cdot T_g < T \end{cases} \quad (14)$$

Soil Type	A	B	C	D
χ	1.5	2.1	2.5	3.0

where χ is a coefficient depending on the soil class. Their statistical analysis confirmed that expressing q_μ in terms of the ratio T/T_g reduced the statistical scatter, making the governing parameter the target ductility; little influence of the critical damping and the hysteretic shape, apart from the short-period range, was observed on the magnitude of q_μ .

Influence of seismological and ground motion characteristics for near-field excitations. So far, the overall consensus of the SDOF

studies reported has been (and the statistics proved it so) that the earthquake magnitude was not a governing parameter for the quantification of q_μ , thereby the effort for parameter identification concentrated on the target ductility, the oscillator hysteretic shape, the damping, and, primarily, the soil characteristics and its predominant period.

Mavroeidis et al. (2004) investigated further the seismicity bias by considering the pulse duration T_p of near-field ground motions and

its effect on the strength demands for a given μ_i of EPP systems. The pulse duration observed in near-field records was shown on geophysical (source mechanism) terms to be about twice the rise time of the rupture process generating the earthquake, namely, the time it takes for a point on the fault to reach the maximum fault displacement, which, therefore, was well

correlated to the earthquake moment magnitude M_w . Subsequent inelastic SDOF analyses using near-field pulses demonstrated that T_p was an important parameter for near-field events, particularly through the establishment of amplification factors along the lines of a modified Eq. 4 (Newmark and Hall 1973), in terms of T/T_p (Eq. 15):

$$q_\mu = \begin{cases} 1 & \frac{T}{T_p} < \left(\frac{T}{T_p}\right)_a \\ \sqrt{2\mu_i - 1} & \left(\frac{T}{T_p}\right)_b < \frac{T}{T_p} < \frac{\sqrt{2\mu_i - 1}}{\mu_i} \left(\frac{T}{T_p}\right)_c \\ \mu_i & \frac{T}{T_p} > \left(\frac{T}{T_p}\right)_c \end{cases}$$

where:

M_w	$\left(\frac{T}{T_p}\right)_a$	$\left(\frac{T}{T_p}\right)_b$	$\left(\frac{T}{T_p}\right)_c$	ϕ_{ev}	ϕ_{ea}
5.6 – 6.3	0.035	0.35	0.75	2.10	0.95
6.4 – 6.7	0.010	0.2	0.75	2.0	0.75
6.8 – 7.6	0.002	0.055	0.75	1.55	0.43

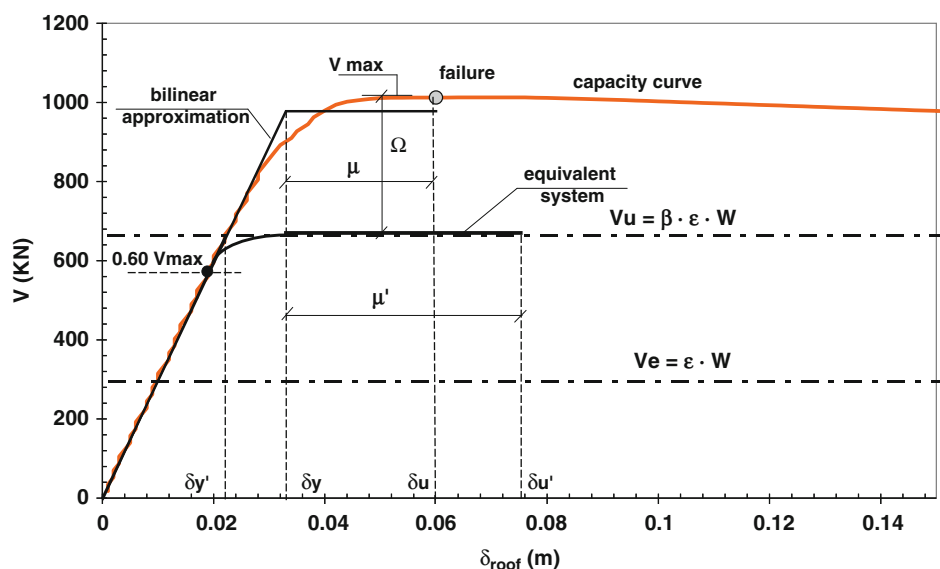
(15)

Gillie et al. (2010) also considered the sensitivity of the q_μ - μ - T relations of EPP systems on the fault directivity effects in near-field earthquake motions, namely, the content of a strong velocity pulse. Their study was based on a dataset of 82 near-field acceleration records with velocity pulse content due to fault directivity effects. They demonstrated that, in the near-field case, unlike the prevailing notions on published q_μ - μ - T functionals from statistical analysis of both near- and far-field events, there was a much stronger dependence of q_μ on the earthquake magnitude and a much smaller dependence on the local soil conditions. Recognizing this dependence, they therefore examined the dependence of q_μ to the predominant period T_p of the velocity pulse content or, equivalently, to T_{RSV} , the period at the peak velocity of the elastic velocity spectrum, which was found to be the influencing parameter for the magnitude of q_μ . Overall, near-field events were proven to demand by the SDOF oscillators

lower q_μ than those with no directivity effects, such as those considered by previous q_μ - μ - T relations proposed so far in the literature. They therefore also proposed q_μ relations for near-field seismic events in terms of M_w (range of 6.0–7.5), since both T_p and T_{RSV} are not known a priori for a given earthquake; however, they are adequately correlated with the earthquake magnitude.

Building Behavior Factor q_s

In addition to the development of IDRS and ductility-dependent behavior factors of SDOF systems, considerable interest has been given to the evaluation of the available behavior factor supplied by entire structural systems (q_s) following different performance criteria, also compared to the behavior factor specified by modern building codes for design (q_d) following allowable stress, ULS, or PBD design procedures. Such studies have been specifically related to typical RC or steel building construction, either modern designs according to a particular seismic design code or existing structures, designed using past regulations. The purpose of such studies aimed for the evaluation, calibration, or reliability analysis (often using stochastic analysis procedures) of the behavior factor(s) specified by the subject



Behavior Factor and Ductility, Fig. 5 Evaluation of the overstrength, the ductility, and the behavior factor of an existing RC structure (From Repapis et al. 2006a)

code or structural system under scrutiny. Although most studies (especially related to code design parameter calibration and reliability analysis) have been aimed toward the evaluation of the overall design behavior factor q_s , studies of specific structural systems (e.g., RC shear wall or steel braced frame structures) have also been concerned with the evaluation of the behavior factor due to overstrength q_Ω .

Generally, the evaluation of q_s of a new or an existing structure is based on either static push-over (SPO) or time history inelastic analyses under an earthquake excitation set. The resulting q is particular to the building type and structural material under scrutiny, the EDRS, and the performance level criteria adopted, as well as the entire set of underlying code-specific design requirements inherent in the seismic design of the structure. Such requirements include, among others, the design load level during earthquake, the relative magnitude of gravity and seismic loads under the seismic load combination, limiting section geometry or reinforcement requirements, the degree of conservatism in the design, the code prescribed detailing rules, and so on. Hence, these behavior factors can only be

used for the corresponding design environment over which they have been calibrated for.

Evaluation of the Behavior Factor Using SPO Analysis

The evaluation of q_s using SPO methods involves the estimation of the capacity curve of the structure under a prescribed lateral load and/or displacement profile (modal, triangular, constant, adaptive) and crucial modeling assumptions (among others) of the strength and stiffness dependence of the member characteristics (also the finite element formulation itself), of the foundation and the joint models, including or not the second-order effects at the global and local element level. Following the definition of a set of local and global failure criteria (the damage indices), the available global ductility of the structural model is evaluated at the minimum roof deformation at incipient satisfaction of any of these indices, over the onset of yield roof deformation. Repapis et al. (2006a) have adopted SPO procedures in order to evaluate the available q_s and also q_Ω of existing RC plane irregular frames, typical of structural designs between the 1960s and 1990s in Greece (Fig. 5). Their results were

compared to dynamic analysis estimates of q_s , as described in the next section.

Evaluation of the Behavior Factor Using Dynamic Analysis

Several analytical estimations of q_s based on dynamic analysis have been published, in order to calibrate or verify the reliability of the behavior factor as a static analysis parameter in force-based seismic design, against local or global damage predictions. Two methods, the direct and the indirect method, have been proposed for evaluating q_s of a building using dynamic analysis (EC8 ENV 1988). Recently, in FEMA P-695 (2009), the procedure is formalized for a uniform reliability of evaluating q_s and q_Ω in PBD.

- (a) **Indirect evaluation of q_s .** A building initially designed for a given EDRS (shape, PGA equal to A_d , critical damping, and a design behavior factor q_d) is analyzed in the time domain under a base excitation whose response spectrum matches the EDRS entirely or locally in the vicinity of the initial period of the structure for the same critical damping. The PGA is linearly increased until incipient violation of any of a set of failure criteria, at the nominal collapse value of A_c . The behavior factor q_s is equal to

$$q_s = q_d \cdot \frac{A_c}{A_d} \quad (16)$$

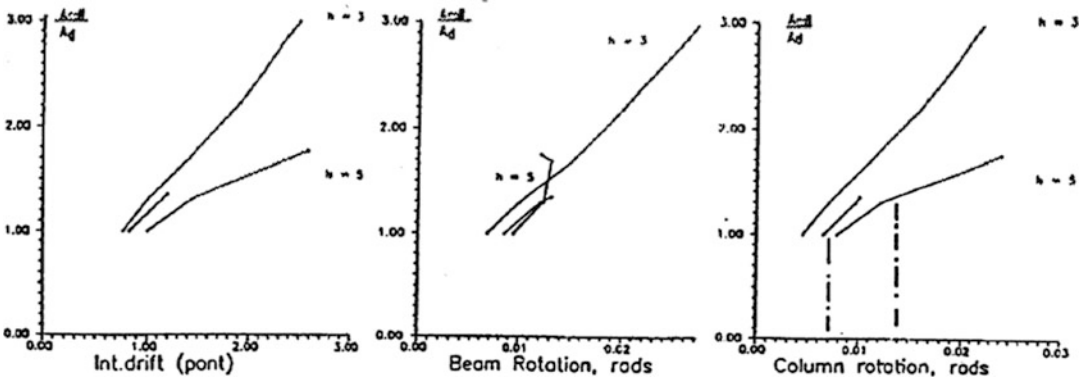
The process is repeated for different earthquake input sets, and a statistical estimate of q_s is established for this type of building, design procedure, and performance criteria. The method suffers from the drawback that the performance indices do not depend linearly on the PGA (Kappos 1991); however, it has been commonly used since q_s is obtained at the expense of only one initial structural design and model formulation. This method falls closely with the application of the incremental dynamic analysis method (IDA) proposed by Vamvatsikos and Cornell (2003) for evaluating existing structural systems.

- (b) **Direct evaluation of q_s .** A set of collapse criteria is adopted together with an EDRS (shape, PGA, and damping) and a set of base excitation records matching the specific EDRS. The structure is designed for the assumed EDRS in a repetitive manner, initially assuming a design q_d as proposed by the code and, subsequently, for increasing (or possibly decreasing) values of q_d^i for each i_{th} design iteration; decreasing values may need to be considered in irregular or non-ductile structures in regions of high seismicity. For each i_{th} design, an inelastic model of the structure is formulated and analyzed over the design earthquake set. This design – analysis procedure – is repeated until any of the preset collapse criteria is incipiently violated: this limiting q_d^i is the behavior factor q_s .

$$q_s = \max(\text{or min}) q_d^i \quad (17)$$

In this case, a direct redesign method is adopted for q_s at the expense of several structural (re)designs. As before, the resulting behavior factor is a function of the design assumptions as well as the performance level evaluation criteria and the design record set.

Kappos (1991) applied the direct method for the evaluation of q_s for two typical RC plane frames, namely, a moment frame and a dual frame-wall system, designed according to EC8 ENV (1988) for medium ductility class using a q_d of 3.50 and 2.10, respectively. Using performance limits of total and interstory drift, member shear capacity, and local curvature ductility μ_{φ} , the resulting available q_s were evaluated to be 4.9 and 3.4 of the bare frame and the dual system, respectively. Different criteria governed for each structure and base excitation: for the bare frame system, the story drift limit governed, while for the dual system, member failure governed the response (the ground story columns). Zeris et al. (1992) applied the direct method to medium-height RC frames with a relatively taller ground story designed according to EC8 ENV (1988) for ductility class II and a q_d of 3.50, observing design code limitations (e.g., member



Behavior Factor and Ductility, Fig. 6 Variation of the structural behavior factor q_s of a tall first-story building with base excitation amplitude (Zeris et al. 1992)

Behavior Factor and Ductility, Table 2 Comparison of building q_s using static and dynamic analysis for three excitations (minimum value, mean, and standard deviation) (Zeris et al. 2005)

Frame building type	q (SPO)	q (IDA)		
		$\bar{q}$	$q_{\min}$	σ_q
Nonconforming ordinary frame, Zone I	2.03	2.18	1.92	0.33
Nonconforming ordinary frame, Zone II	1.77	3.15	2.43	1.25
Nonconforming frame with tall first story	2.34	2.82	2.41	0.70
Nonconforming frame with a recess	1.55	2.78	2.36	0.39
Nonconforming frame with discontinuous column	1.98	4.31	2.87	2.48
Nonconforming frame with discontinuous beam	2.14	2.12	1.81	0.43
Conforming ordinary frame, Zone I	5.44	6.78	5.22	1.84

geometry, reinforcement detailing). Adopting failure limits in interstory drift and μ_φ as a function of confinement, it was shown that for a μ_φ of 5 and 10, the q_s of the regular RC frame was 4.4 and 7.1, while the tall first-story frame was marginally able to satisfy the μ_φ limit of 5.0 at $q_s = q_d = 3.5$, supplying a $q_s = 4.6$ for $\mu_\varphi = 10$ (Fig. 6).

For the case of modern building designs, Kappos (1999) and Borzi and Elnashai (2000) evaluated the ductility and q_s of typical low- to medium-rise RC buildings designed according to the CEB model seismic code or EC8, using both SPO and dynamic analysis SPO. In addition to comparisons of SDOF q_μ predictions of their record datasets with those available in the literature (e.g., by Miranda and Bertero 1994), the problems associated with comparing q_μ of entire structural systems (after excluding q_Ω) with

SDOF analysis were considered in detail. The studies concluded that the design behavior factor in EC8 was conservative for the cases considered, although their results were conflicting (proving the fact that the design and building bias affects the end result), attributing the conservatism of q_s to either the structural overstrength or the ductility supply or both.

In order to establish q_s of typical existing irregular buildings in Greece constructed in the 1960s and 1970s, Zeris et al. (2005) applied IDA analysis methods for evaluating q_s using the indirect method of evaluation; these values were compared with values obtained from SPO analysis by Repapis et al. (2006b), taking into account both interstory drift and local critical region rotational ductility limits comparable to the previous study. The resulting estimates are compared in Table 2, showing that (i) the available behavior

factors are well below currently adopted and (ii) dynamic analysis predictions of q_s for irregular buildings tend to be more conservative than those using dynamic analysis, being closer to the minimum predictions obtained in the latter case.

Chryssanthopoulos et al. (2000) evaluated the reliability of the design behavior factor q_d adopted by the Eurocode for the case of a regular ten-story RC frame designed for ductility class M. Using the indirect method of evaluation of the available q_s , they assumed a statistical variation of the structural material properties (concrete and steel yield strength and ultimate failure strain of the reinforcement), the uncertainty in confinement (and therefore supplied μ_φ in the critical regions), and a set of spectrum compatible base excitation records, in order to evaluate the reliability of ULS design following EC8 recommendations for the subject building. At the same time, they established the variation of the available q_s of the subject frame under the base input set. They concluded that for a seismic hazard model with a return period of 475 years for Southern Europe, the violation of the initial design value (q_d) of 3.75 was to be expected in 0.6 % of the frames exhibiting this type of building and design characteristics and the material property variability assumed.

Summary and Conclusions

It should be noted at the outset that such an intentional underdesign approach (i.e., using a behavior factor) is by no means a socially irresponsible approach from the point of view of the engineers' role to the society. It is well accepted that economic limitations and design feasibility will define the intensity level that is socially acceptable for the design of structures for the extreme earthquake; furthermore, it is well known that the gradual increase of records through more dense earthquake monitoring grids of digital instruments has proven more and more often that the actual seismic load intensity at a site will surprise us, when an extreme

earthquake event happens. Consequently, until sufficient seismic data are available, structures will continue to be constructed in regions in which, historically or otherwise, a fault was not known to exist or the existing faults' ability to induce a given seismic magnitude has been underestimated. Such famous examples, among others, were (i) the unexpected 1983 Coalinga earthquake in a quiet region in California; (ii) the 1971 Sylmar earthquake with its surprising (*at the time*) Pacoima Dam record exceeding 1.0 g; (iii) the numerous strong motion records since then that have well exceeded 1.0 g in Chile, New Zealand, Taiwan, and elsewhere; and (iv) for Greece, the 1999 Athens earthquake that ruptured an unknown extension of a known fault in a low- to medium-seismicity zone with a design effective PGA of 0.16 g, yielding estimated PGAs of 0.5 g, and the 2014 earthquakes in a well-known seismically active region in Kefallinia, exceeding the design effective PGA by as high as 200 %. Such consistent overshoots of our impression of the expected level of shaking in a seismically affected region demonstrate the need for introducing to our designs reasonable reduction factors coupled with excess ductility capacity and structural redundancy, rather than adopting $q_d = 1.0$ -type designs with no built-in ductility.

Extensive statistical studies have demonstrated that the q (or R) factors of SDOF systems are not sensitive to the earthquake magnitude unless near the causative fault, while they show a reasonable similarity (in terms of the estimated q_μ - μ - T relations) between earthquake faults of similar tectonic characteristics. Furthermore, they do not seem to be sensitive to the hysteretic shape of the system, for EPP and hardening systems of conventional new buildings. They are sensitive, however, to softening cyclic response, such as the case of structures with excessively softening or brittle failure mechanisms, such as existing RC frame or frame-wall structures, buildings with strong P - δ effects, or infilled RC frames with weak infills. They also depend strongly on the EDRS (de)amplification in the case of a resonant (or not) structural period near

the predominant period of the local soil conditions, in soft soil sites. These observations make the specification in earlier codes of constant system-dependent q_μ for design unconservative.

Can the behavior factor q_μ be increased to 1.0 (elastic response) at the gain of lack of implementing ductility into the system (i.e., $q_s = q_\Omega$)? Although the incorporation of ductility was initially imperative in seismic designs, this is no longer the case, and the codes are coming at the end of closed circle situation at this point, where one may (at the national level) relax this compulsory requirement. In this context, EC8 (2004), for instance, accounts for the case of a low ductility class (DCL) design, whereby a contemporary seismic design may for all practical purposes ignore all seismic detailing provisions for ductility, provided that a q_d close to 1.0–1.50 is adopted; in fact, the New Greek Code for Concrete Works still in effect in parallel usage with EC8 (2004) during the publication of this work allows for such a non-ductile design, even though the Greek National Annex of EC8 (2004) wisely, in the authors' opinion, precludes the use of DCL buildings in Greece. Given the performance of non-ductile structures in strong earthquakes and the possibly increased cost of introducing ductility detailing into the structure, it is believed that the practice of neglecting to incorporate ductility may lead to unsafe designs; therefore, adequate behavior factors with ductile detailing should be introduced, so that incorporating ductility is financially feasible.

Cross-References

- ▶ [Assessment of Existing Structures Using Inelastic Static Analysis](#)
- ▶ [Equivalent Static Analysis of Structures Subjected to Seismic Actions](#)
- ▶ [Nonlinear Finite Element Analysis](#)
- ▶ [Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings](#)
- ▶ [Seismic Vulnerability Assessment: Reinforced Concrete Structures](#)

References

- ASCE (2013) Minimum design loads for buildings and other structures. Standards SEI/ASCE 7-10. American Society of Civil Engineers, Reston
- ATC (1978) Tentative provisions for the development of seismic regulations for buildings. Report ATC 3-06. Applied Technology Council, Redwood City
- ATC (1995) Structural response modification factors. Report ATC-19. Applied Technology Council, Redwood City
- Biggs J (1964) Introduction to structural dynamics. McGraw-Hill, New York
- Borzi A, Elnashai AS (2000) Refined force reduction factors for seismic design. Eng Struct 22(10):1244–1260
- Chavez J, Khemici O, Khater M, Keshishian P (2012) Building codes and relative seismic vulnerability in Latin American Countries. In: Proceedings of 15th world conference of earthquake engineering, Lisbon
- Chryssanthopoulos MK, Dymiotis C, Kappos AJ (2000) Probabilistic evaluation of behaviour factors in EC8-designed R/C frames. Eng Struct 22(8):1028–1041
- Clough RW, Penzien J (1975) Dynamics of structures. McGraw Hill, New York
- EC2 (2004) Eurocode No. 2, design of concrete structures - part 1-1: general rules and rules for buildings, EN-1992-1-1. European Committee for Standardization, Brussels
- EC8 (2004) Eurocode No. 8, design of structures for earthquake resistance – part 1: general rules, seismic actions and rules for buildings, EN-1998-1. European Committee for Standardization, Brussels
- EC8 ENV (1988) Background documents for Eurocode 8, Part 1. Vol. 2 – Design rules. Commission of the European Communities, Brussels
- Fajfar P (1999). Capacity spectrum method based on inelastic demand spectra. Earthq Eng Struct Dyn 28(9):979–993
- FEMA-356 (2000) Prestandard and commentary for the seismic rehabilitation of buildings. FEMA, Washington, DC
- FEMA-695 (2009) Quantification of building seismic performance factors. FEMA, Washington, DC
- Genshu T, Yongfeng Z (2007) Seismic force modification factors for modified-Clough hysteretic model. Eng Struct 29(11):3053–3070
- Gillie J, Rodriguez-Marek A, McDaniel C (2010) Strength reduction factors for near-fault forward-directivity ground motions. Eng Struct 32(1):273–285
- Housner G (1956) Limit design of structures to resist earthquakes. In: Proceedings, world conference of earthquake engineering. Earthquake Engineering research Institute, Berkeley
- Kappos A (1991) Analytical prediction of the collapse earthquake for R/C buildings: case studies. Earthq Eng Struct Dyn 20(2):177–190

- Kappos A (1999) Evaluation of behaviour factors on the basis of ductility and overstrength studies. *Eng Struct* 21(9):823–835
- Lee LH, Han SW (1999) Determination of ductility factor considering different hysteretic models. *Earthq Eng Struct Dyn* 28(9):957–977
- Mahin SA, Bertero VVB (1981) An evaluation of inelastic seismic design spectra. *J Struct Div ASCE* 107(9):1777–1795
- Mavroeidis GP, Dong G, Papageorgiou AS (2004) Near-fault ground motions, and the response of elastic and inelastic single-degree-of-freedom (SDOF) systems. *Earthq Eng Struct Dyn* 33(9):1023–1049
- Miranda E (1993) Site-dependent strength reduction factors. *J Struct Eng ASCE* 119(12):3503–3519
- Miranda E, Bertero VVB (1994) Evaluation of strength reduction factors for earthquake-resistant design. *Spectra* 10(2):357–379
- Miranda E, Ruiz-Garcia J (2002) Influence of stiffness degradation on strength demands of structures built on soft soil sites. *Eng Struct* 24(10):1271–1281
- Nassar AA, Krawinkler H (1991) Seismic demands for SDOF and MDOF systems, TR 95, The J. Blume Earthquake Engineering Center, Stanford University, Palo Alto
- NEHRP (1985) NEHRP Recommended provisions for the development of seismic regulations for new buildings. Building Seismic Safety Council, Washington, DC
- Newmark NM, Hall WJ (1973) Seismic design criteria for nuclear reactor facilities. Report No. 46, Building Practices for Disaster Mitigation, National Bureau of Standards, US Department of Commerce, pp 209–236
- Ordaz M and Pérez-Rocha LE (1998) Estimation of strength-reduction factors for elastoplastic systems: a new approach. *Earthq Eng Struct Dyn* 27(9):889–901
- Repapis K, Vintzeleou E, Zeri C (2006a) Evaluation of the seismic performance of existing RC buildings: I suggested methodology. *Eur J Earthq Eng* 10(2):265–288
- Repapis K, Zeri C, Vintzeleou E (2006b) Evaluation of the seismic performance of existing RC buildings: II a case study for regular and vertically irregular buildings. *Eur J Earthq Eng* 10(3):429–452
- Riddell R (1995) Inelastic design spectra accounting for soil conditions. *Earthq Eng Struct Dyn* 24(11):1491–1510
- SEAOC (1974) Recommended lateral force requirements and commentary, 4th edn. Struct. Engineers Assoc. of California, San Francisco
- Uang CM (1991) Establishing R (or R_w) and C_d factors for building seismic provisions. *J Struct Eng ASCE* 117(1):19–28
- UBC (1988) Uniform building code. International Conference of Building Officials, Washington, DC
- Vamvatsikos D, Cornell CA (2003) Applied incremental dynamic analysis. *Earthquake Spectra* 20(2):525–533
- Veletsos A, Newmark NM (1960) Effect of inelastic behaviour on the response of simple systems to earthquake motions. In: Proceedings second world conference on earthquake engineering, Japan, pp 895–912
- Vidić T, Fajfar P, Fischinger M (1994) Consistent inelastic design spectra: strength and displacement. *Earthq Eng Struct Dyn* 23(5):507–521
- Watanabe G, Kawashima K (2002) An evaluation of the forced reduction factor in the force-based seismic design. In: Proceedings of 39th joint meeting, panel on wind and seismic effect, UJNR, NIST Special Publication, Gaithersburg
- Whittaker A, Hart G, Rojahn C (1999) Seismic response modification factors. *J Struct Eng* 125(4):438–444
- Zeri C, Tassios TP, Lu Y, Zhang GF (1992) Influence of irregularity on the q factor of RC frames. In: Proceedings, 10th world conference of earthquake engineering, Madrid
- Zeri C, Repapis K, Vintzeleou E (2005) Seismic performance of existing irregular RC buildings. Paper No. 32, Proceedings of the 4th European workshop on the seismic behaviour of irregular and complex structures, Thessaloniki

Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations

Satish Nagarajaiah^{1,2} and Yongchao Yang¹

¹Department of Civil and Environmental Engineering, Rice University, Houston, TX, USA

²Department of Mechanical Engineering, Rice University, Houston, TX, USA

Synonyms

Blind source separation; Modal identification; Sparse representation; Structural health monitoring; Time-frequency analysis

Introduction

Traditional model-based structural modal analysis and damage identification methods are typically parametric and user involved; as such, they are usually associated with demanding computational resources and require quite a lot of prior knowledge of structures. For practical

applications, it would be useful to seek efficient structural identification methods that may be able to extract the salient information directly from the measured structural signals. The recently widely deployed advanced structural health monitoring (SHM) systems in structures with dense sensors also support such an effort: the massive recorded data especially call for efficient data-driven algorithms (Yang and Nagarajaiah 2014c, d) for further structural assessment.

Recently, blind source separation (BSS) has emerged as a new unsupervised machine learning tool (Hyvärinen and Oja 2000) and has been extensively studied in structural dynamics and output-only modal identification (Antoni 2005; Kerschen et al. 2007; Yang and Nagarajaiah 2013a, b, c, 2014a, b; Poncelet et al. 2007; Zhou and Chelidze 2007; McNeill and Zimmerman 2008; Hazra et al. 2010; Hazra and Narasimhan 2010; Sadhu et al. 2011, 2012; Abazarsa et al. 2013; Antoni and Chuahan 2013; Ghahari et al. 2013). Essentially, BSS techniques are able to recover the hidden source signals and their underlying factors using only the observed mixtures; it is thus suitable to perform output-only structural identification when structural input or excitation is usually extremely difficult or expensive to obtain.

This entry presents the authors' recent work on data-driven output-only modal identification and damage detection of structures. It is found that exploiting the sparse essences of modal expansion and damage information can efficiently and effectively address some challenging problems in output-only modal identification and damage detection via BSS. A series of novel algorithms are developed with experimental and real-world structure examples for demonstrations.

Blind Identification of Damage via Sparse Independent Component Analysis (ICA)

The sparse damage features hidden in the structural information can be blindly extracted via a

BSS technique called independent component analysis (ICA), as detailed in the following.

ICA is popularly used to estimate the blind source separation (BSS) model (Hyvärinen and Oja 2000),

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) = \sum_{i=1}^n \mathbf{a}_i s_i(t) \quad (1)$$

using only the observed mixture vector $\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_m(t)]^T$; $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_n(t)]^T$ and $\mathbf{A}$ denote the latent source vector and the unknown constant $m \times n$ linear mixing matrix, respectively, to be simultaneously estimated. $\mathbf{a}_i$ is the i th column of $\mathbf{A}$ and is associated with the corresponding source $s_i(t)$. The assumption of $m = n$ is imposed herein, i.e., the number of mixtures equals that of the sources and $\mathbf{A}$ is square. With only $\mathbf{x}(t)$ known, Eq. 1 may not be mathematically solved by classical methods; additional assumption is thus needed to estimate the BSS model.

The principle of ICA estimation is based on the classical central limit theorem (CLT), which states that a sum of independent random variables tends to distribute toward Gaussian, i.e., a mixture of independent random variables is always more Gaussian than any one of the original variables (except that the mixture only contains one random variable). As seen in Eq. 1, mixtures are expressed as a weighted sum of the sources themselves; they are thus always more or equally Gaussian than the sources. ICA therefore searches for proper demixing matrix $\mathbf{W}$ such that the recovered independent components (ICs) $\mathbf{y}(t) = [y_1(t), y_2(t), \dots, y_n(t)]^T$ obtained by

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \quad (2)$$

are as non-Gaussian as possible and thus approximate $\mathbf{s}(t)$. Each IC $y_i(t)$ is computed by

$$y_i(t) = \mathbf{w}_i \mathbf{x}(t) \quad (3)$$

with $\mathbf{w}_i$ denoting the i th row of $\mathbf{W}$. By seeking those ICs which maximize non-Gaussianity, the

sources (and simultaneously the mixing matrix) can therefore be recovered by ICA.

Non-Gaussianity of a random variable can be measured by some contrast function, e.g., negentropy. The entropy of a discrete random variable $v = \{v^1, v^2, \dots, v^i, \dots\}$ is defined by

$$H(v) = -\sum_i p(v = v^i) \cdot \log p(v = v^i) \quad (4)$$

where $p(\cdot)$ is the probability mass operator. Entropy measures the uncertainty or randomness of a random variable. For example, for a random variable with impulse probability mass function, its entropy is zero, i.e., it is completely determined.

The Gaussian random variable has the largest entropy among all other random variables with equal variance (Hyvärinen and Oja 2000), i.e., it is the most random or uncertain one. On the other hand, a random variable with sparse representation has small entropy as it is less random or easier to be predicted. This conclusion yields the definition of negentropy as a measure of non-Gaussianity given by

$$J(v) = H(v_{\text{gau}}) - H(v) \quad (5)$$

in which v_{gau} is a standardized Gaussian random variable (zero mean and unit variance); it quantitatively evaluates the entropy distance of a (standardized) random variable from a Gaussian variable. Finding the ICs that maximize the negentropy by ICA thus yields random variables with sparse representation (Yang and Nagarajaiah 2014a). This finding turns out very useful for damage identification, as subsequently described.

A simplified approximation to the negentropy is the classical kurtosis, which is defined by

$$\text{kurt}(v) = E[v^4] - 3(E[v^2])^2 \quad (6)$$

where $E[\cdot]$ denotes the expectation operator. The kurtosis of a Gaussian random variable is zero, and that of a non-Gaussian random variable is nonzero. It is easy to estimate and computationally efficient. The FastICA is one of the most

efficient algorithms implementing ICA estimation and is adopted in this study.

Damage may behave as pulse-like information hidden in the structural vibration response signals once they are processed further such as in the wavelet domain. The pulse-like feature containing damage information may be buried in the noisy wavelet-domain signals on a certain scale. Because ICA biases to extract sparse components from the observations, feed the wavelet-domain responses $\mathbf{x}^l(t)$ at the l th scale as mixtures into the BSS model,

$$\mathbf{x}^l(t) = \mathbf{A}\mathbf{s}(t) \quad (7)$$

If there is any pulse-like feature hidden in $\mathbf{x}^l(t)$, then ICA will extract it, which is to be revealed in the recovered sparse component $y_j(t)$ with sharp spikes indicating damage

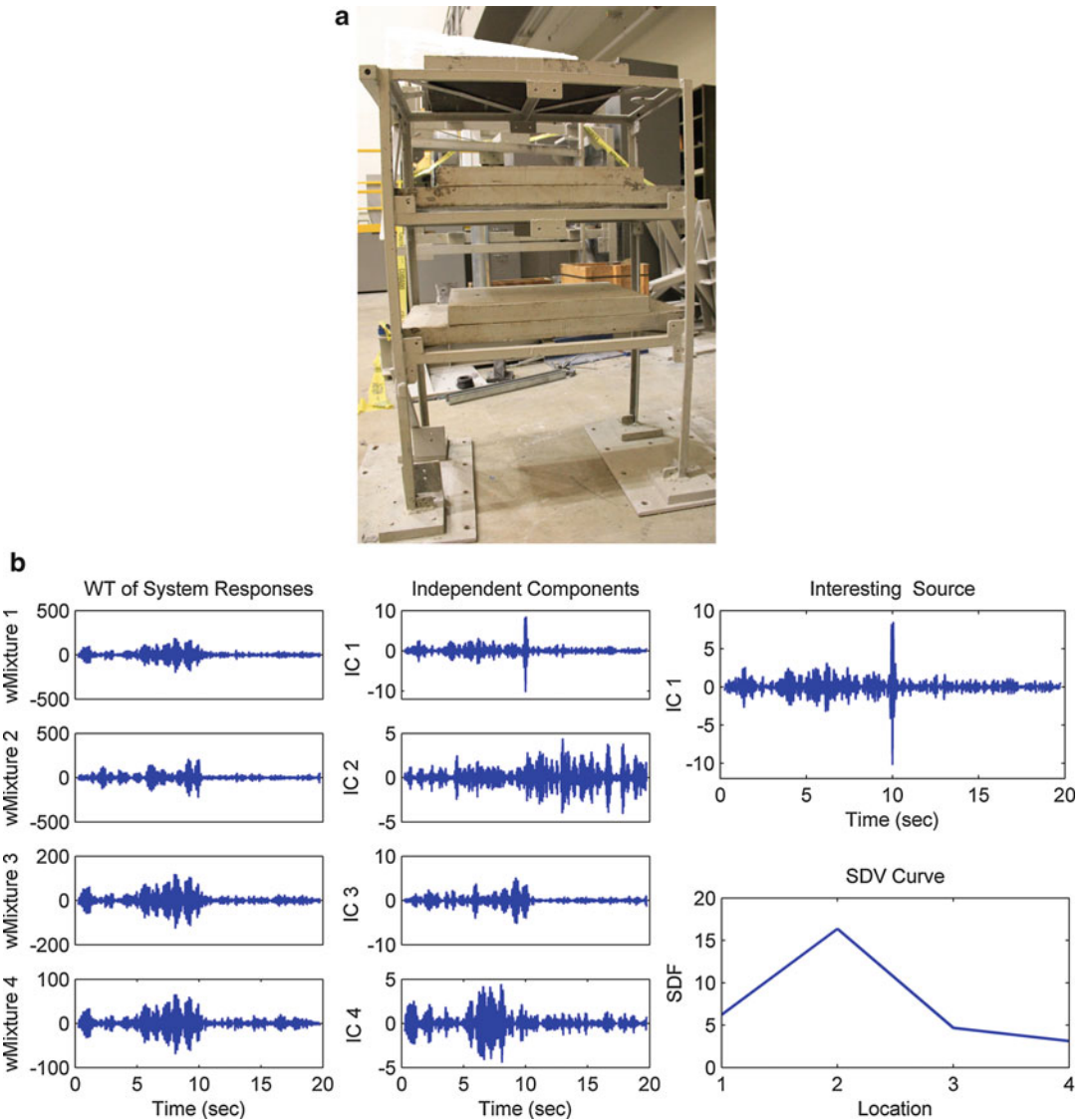
$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}^l(t) \quad (8)$$

Such $y_j(t)$ is proposed as the “interesting” source within the damage identification framework. Note that $\mathbf{x}^l(t)$ inherits the temporal information of the responses; this implies that the recovered “interesting” source $y_j(t)$ retains temporal signatures of the inflicted damage, which is indicated by the time instant location of the sharp spike.

Expanding the WT-BSS model Eq. 7 as

$$\begin{aligned} \mathbf{x}^l(t) &= \sum_{i=1}^n \mathbf{a}_i s_i(t) \\ x_1^l &= \sum_{i=1}^n a_{1i} s_i(t) = a_{11} s_1(t) + a_{12} s_2(t) + \dots + a_{1n} s_n(t) \\ x_j^l &= \sum_{i=1}^n a_{ji} s_i(t) = a_{j1} s_1(t) + a_{j2} s_2(t) + \dots + a_{jn} s_n(t) \\ x_n^l &= \sum_{i=1}^n a_{ni} s_i(t) = a_{n1} s_1(t) + a_{n2} s_2(t) + \dots + a_{nn} s_n(t) \end{aligned} \quad (9)$$

Observe that the mixing coefficient a_{ji} locates the mixture and the source by its indices j and i , respectively, and the columnwise vector $\mathbf{a}_i = [a_{1i}, a_{2i}, \dots, a_{ji}, \dots, a_{ni}]^T$ contains the spatial signature of the corresponding source $s_i(t)$. Herein, $\mathbf{a}_i$ and its element a_{ji} are proposed as the



Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations, Fig. 1 Sparse ICA simultaneous identification of both structural damage instant and damage location: (a) the

experimental structure with four sensors embedded subject to white noise excitation at the base and (b) the WT-ICA identification results

source distribution vector (SDV) and source distribution factor (SDF), respectively; they describe how source $s_i(t)$ is distributed among n mixtures. Specifically, if a_{ji} has the largest (absolute) value among $\mathbf{a}_i = [a_{1i}, a_{2i}, \dots, a_{ji}, \dots, a_{ni}]^T$, then $x_j(t)$ contains most $s_i(t)$ components among all the mixtures (Yang and Nagarajaiah 2014a).

The proposed concepts of SDV and SDF are readily extended to damage localization issues. As the structural response in the vicinity of damage naturally contains most spike-like features, damage can be localized by tracking the spike in the SDV of the recovered impulse-like IC, which is the “interesting” source. Figure 1 shows an experimental structure application of WT-ICA

in simultaneous identification of both damage instant (at the 10th second) and damage location (the second sensor, left column in the first floor) by the spike feature in the recovered “interesting” source and the corresponding spatial signature, respectively; the detail of this experiment is referred to Yang and Nagarajaiah (2014a).

Output-Only Modal Identification by Sparse Time-Frequency ICA

In the above section, the sparse wavelet ICA is introduced for identification of the hidden spike-like features that typically indicate structural damage information. In the following, the sparse properties of modal expansion are further exploited and lead to a new output-only modal identification method STFT-ICA within the BSS framework which can handle even highly damped structures, as detailed in the following.

For a linear time-invariant system with n degrees-of-freedom (DOF), the governing equation of motion is

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{C}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) = \mathbf{f}(t) \quad (10)$$

where $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are the symmetric mass, (diagonalizable) damping, and stiffness matrices, respectively, and $\mathbf{f}(t)$ is the external force.

The system responses $\mathbf{x}(t) = [x_1(t), \dots, x_n(t)]^T$ can be expressed using the modal expansion

$$\mathbf{x}(t) = \Phi \mathbf{q}(t) = \sum_{i=1}^n \varphi_i q_i(t) \quad (11)$$

where $\Phi \in \mathbb{R}^{n \times n}$ denotes the inherent vibration mode matrix (real-valued normal modes) with its i th column (modeshape), and $\varphi_i \in \mathbb{R}^n$ is the modal response (real-valued).

The output-only identification issue pursues to identify the modal parameters by solely relying on the knowledge of $\mathbf{x}(t)$ without any excitation or input information to the system, such like identification of both Φ and $\mathbf{q}(t)$ only from $\mathbf{x}(t)$ in Eq. 11. Such is an ill-posed problem and may

not be solved mathematically. Traditional output-only modal identification methods usually presume a parametric model (e.g., a stochastic state space model) to proceed with the identification; however, they have a serious drawback of the model order determination problem. Other methods may suffer from sensitivity to noise, dependence on expert experience, and heavy computational burden. The BSS technique provides a straightforward and efficient algorithm for output-only modal identification, as detailed in the following.

The BSS problem has a similar pursuit with the output-only modal identification issue. The close similarity is implied between the modal expansion Eq. 11 and the BSS model Eq. 1. If the system responses are fed as mixtures into the BSS model, then the target of output-only identifying Φ and $\mathbf{q}(t)$ in Eq. 11 can be solved by blind recovery of $\mathbf{A}$ and $\mathbf{s}(t)$ using those BSS techniques dependent on the independence assumption since $\mathbf{q}(t)$ typically with incommensurable frequencies are independent on the modal coordinates.

It has been shown, however, that such direct extraction of time-domain modal responses by ICA has failed for higher-damped structures (only within 1 %) (Kerschen et al. 2007; Yang and Nagarajaiah 2013a, b, 2014a, Brewick and Smyth 2015). The primary reason lies in that ICA ignores the temporal information of signals, while the targeted modal responses possess significant time structure – exponentially decaying monotone sinusoids.

To address the aforementioned issues of ICA, it is proposed by Yang and Nagarajaiah (2013a) to transform the time-domain modal expansion Eq. 11 to the time-frequency domain where the target modal responses have sparse representations, using the short-time Fourier transform (STFT) prior to the ICA estimation,

$$\mathbf{X}_{f\tau} = \Phi \mathbf{Q}_{f\tau} = \sum_{i=1}^n \varphi_i Q_{f\tau,i} \quad (12)$$

where f and τ are the frequency and window indices, respectively. ICA then extracts the most

independent (sparse) components that are the time-frequency representation of the modal responses,

$$\tilde{\mathbf{Q}}_{f\tau} = \tilde{\mathbf{W}}\mathbf{X}_{f\tau} \quad (13)$$

The recovered independent sources $\tilde{\mathbf{Q}}_{f\tau} = [\tilde{Q}_{1,f\tau}, \dots, \tilde{Q}_{n,f\tau}]^T$ are supposed to be as sparse as possible and approximate the targeted sparse time-frequency representations of the monotone modal responses. The obtained mixing matrix is therefore the estimated normal mode, i.e., $\tilde{\Phi} = \tilde{\mathbf{A}} = \tilde{\mathbf{W}}^{-1}$. Once the normal modes are estimated, the time-domain modal responses can be recovered using the demixing matrix

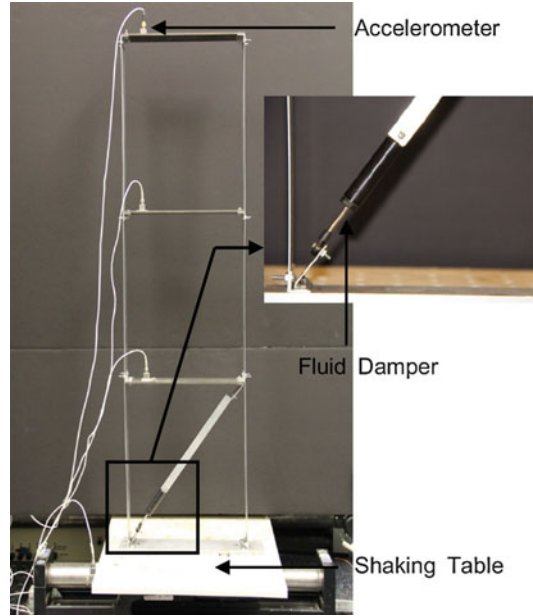
$$\tilde{\mathbf{q}}(t) = \tilde{\mathbf{W}}\mathbf{x}(t) \quad (14)$$

whereby readily estimating the modal frequencies and damping ratios in free vibration are readily estimated by Fourier transform (FT) and Hilbert transform (HT) or logarithm-decrement technique (LT) (Nagarajaiah and Basu 2009; Basu et al. 2008; Nagarajaiah and Li 2004; Dharap et al. 2006; Koh et al. 2005a, b; Li et al. 2007).

Figures 2, 3, 4, and 5 as well as Tables 1 and 2 present the successful application of STFT-ICA in identification of the modal parameters of an experimental structure (Fig. 2) and a real-world structure subject to the Northridge Earthquake 1994 (Nagarajaiah and Dharap 2003; Nagarajaiah and Sun 2000, 2001), both of which are highly damped. The details are referred to Yang and Nagarajaiah (2013a).

Complexity Pursuit on “Independent” Modal Coordinates

Exploiting the properties of the system responses and modal responses can lead to an efficient output-only modal identification method (Yang and Nagarajaiah 2013b) that enjoys even wider success based on a new BSS learning rule complexity pursuit (Stone 2001), which states that the complexity of any mixture signal



Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations, Fig. 2 The experimental highly-damped 3-story steel frame and the zoomed fluid damper

(system response) always lies between that of the simplest source (simplest modal response) and most complicated source (most complex modal response).

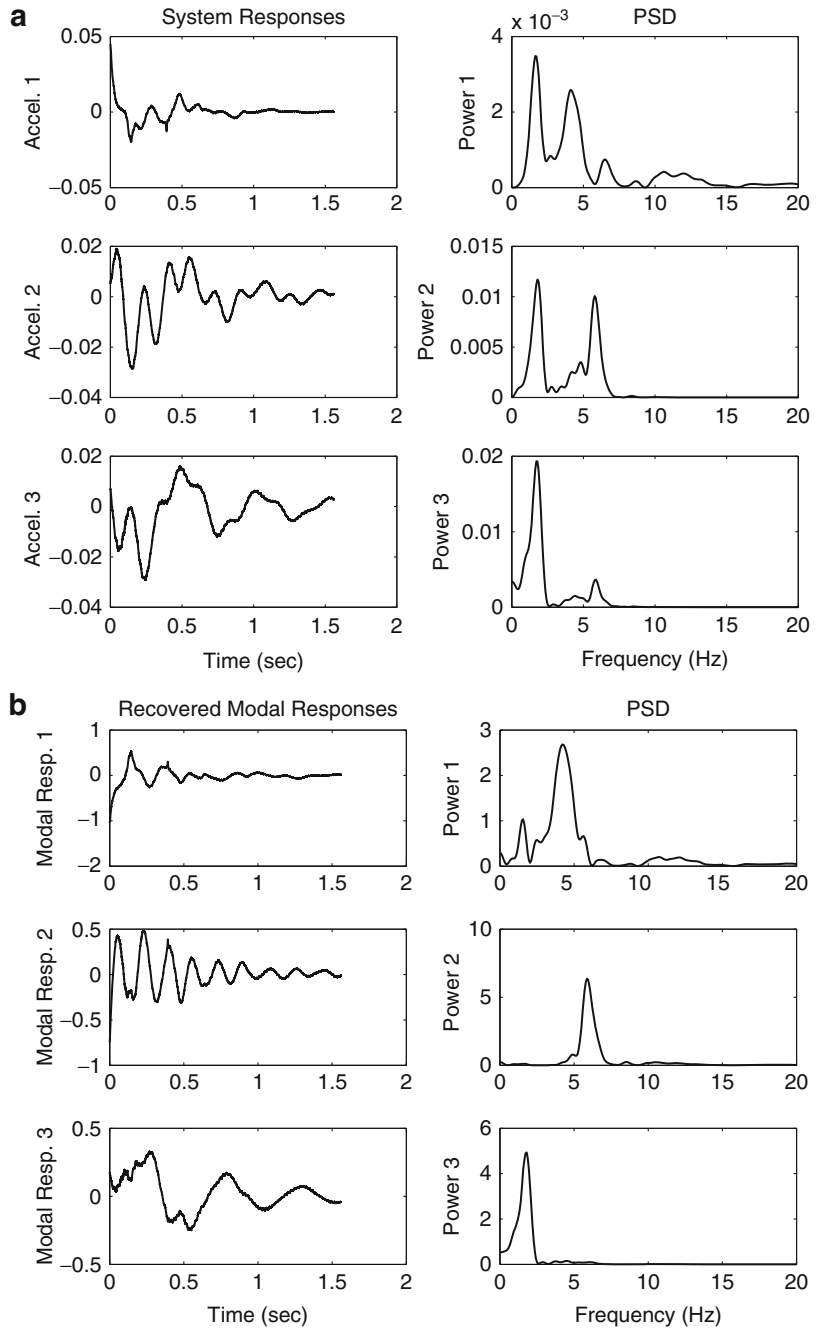
In statistics, the complexity of a signal, say, $y_i = \mathbf{w}_i\mathbf{x}$ (the temporal index is made implicit), is rigorously measured by Kolmogorov complexity. Given that Kolmogorov complexity is not intuitive and difficult to approximate in practice, Stone (2001, 2004) provided a simple yet robust complexity measure of a signal, temporal predictability, which is defined by

$$F(y_i) = \log \frac{V(y_i)}{U(y_i)} = \log \frac{\sum_{t=1}^N (\bar{y}_i(t) - y_i(t))^2}{\sum_{t=1}^N (\hat{y}_i(t) - y_i(t))^2} \quad (15)$$

where the long-term predictor $\bar{y}_i(t)$ and short-term predictor $\hat{y}_i(t)$ are given, respectively, by

Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations,

Fig. 3 (a) System responses (free vibration) of the experimental structure (b) the recovered modal responses by STFT-ICA



$$\begin{aligned}\bar{y}_i(t) &= \lambda_L \bar{y}_i(t-1) + (1 - \lambda_L) y_i(t-1) \quad 0 \leq \lambda_L \leq 1 \\ \hat{y}_i(t) &= \lambda_S \hat{y}_i(t-1) + (1 - \lambda_S) y_i(t-1) \quad 0 \leq \lambda_S \leq 1\end{aligned}\quad (16)$$

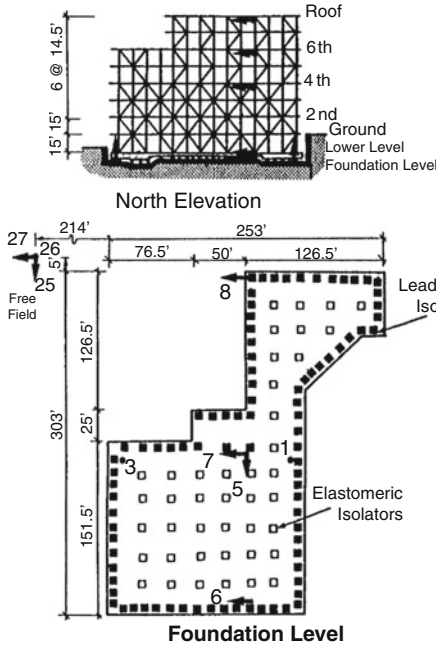
The parameter λ is defined by the half-life parameter h as

$$\lambda = 2^{-1/h} \quad (17)$$

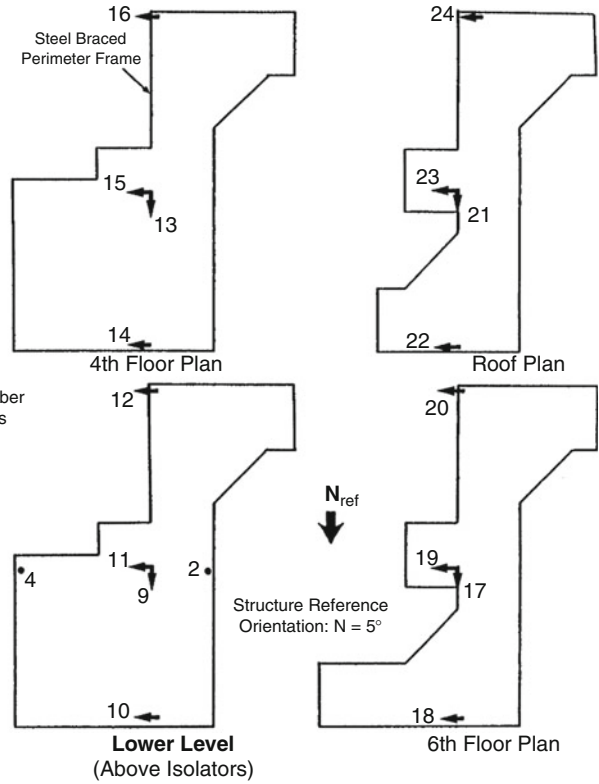
where $h_S = 1$ and h_L is arbitrarily set (say, 900,000) as long as $h_L \gg h_S$ (Stone 2001, 2004).

Incorporate $y_i = \mathbf{w}_i \mathbf{x}$ into Eq. 15,

Los Angeles -7-story University Hospital
(CSMIP Station No. 24605)



SENSOR LOCATIONS



B

Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations,
Fig. 4 Sensor outline of the highly-damped USC hospital building

$$F(y_i) = F(\mathbf{w}_i, \mathbf{x}) = \log \frac{V(\mathbf{w}_i, \mathbf{x})}{(\mathbf{w}_i, \mathbf{x})} = \log \frac{\mathbf{w}_i \bar{\mathbf{R}} \mathbf{w}_i^T}{\mathbf{w}_i \hat{\mathbf{R}} \mathbf{w}_i^T} \quad (18)$$

where $\bar{\mathbf{R}}$ and $\hat{\mathbf{R}}$ are the $n \times n$ long-term and short-term covariance matrix between the mixtures, respectively; their elements are defined as

$$\begin{aligned} \bar{r}_{ij} &= \sum_{t=1}^N (x_i(t) - \bar{x}_i(t)) (x_j(t) - \bar{x}_j(t)) \\ \hat{r}_{ij} &= \sum_{t=1}^N (x_i(t) - \hat{x}_i(t)) (x_j(t) - \hat{x}_j(t)) \end{aligned} \quad (19)$$

Therefore, given a set of mixtures $\mathbf{x}(t)$, the CP learning rule is formulated to search for the demixing vector $\mathbf{w}_i$ which maximizes the temporal predictability contrast function $F(\cdot)$; this can

be solved by the classic gradient ascent technique as described in the following.

Following Eq. 18, the derivative of F with respect to $\mathbf{w}_i$ is

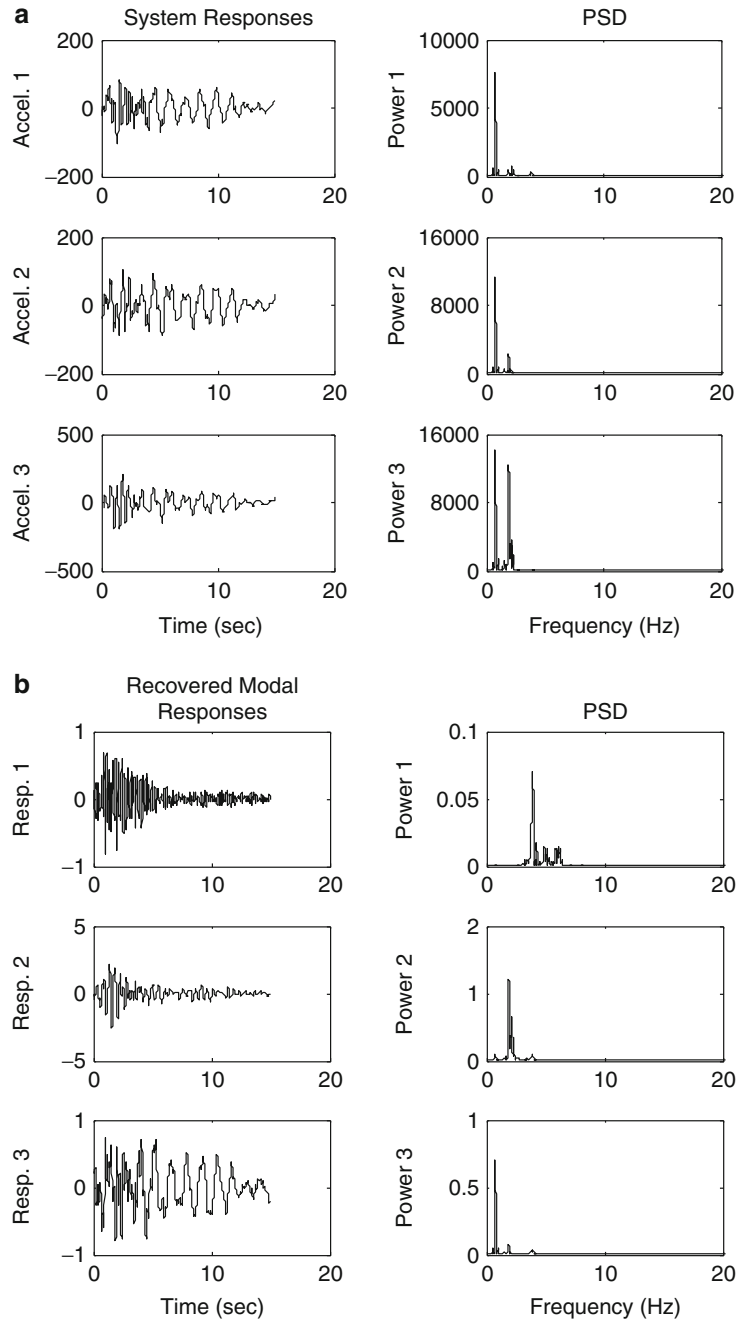
$$\nabla_{\mathbf{w}_i} F = \frac{2\mathbf{w}_i}{V_i} \bar{\mathbf{R}} - \frac{2\mathbf{w}_i}{U_i} \hat{\mathbf{R}} \quad (20)$$

By iteratively updating $\mathbf{w}_i$, a maximum of F is guaranteed to be found; the extracted component $y_i = \mathbf{w}_i \mathbf{x}$ with maximum temporal predictability is the least complex signal and thus approaches the simplest source hidden in the mixtures, according to Stone's theorem in the CP learning rule.

Restricted to Stone's theorem (Xie et al. 2005), however, the sources can only be extracted one by one by maximizing the temporal

Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations,

Fig. 5 (a) System responses (during the Northridge Earthquake 1994) of the USC hospital building (b) the recovered modal responses by STFT-ICA



predictability using the gradient ascent technique. Stone (2001) proposed a more elegant algorithm that can efficiently extract all the hidden sources simultaneously, described as follows.

The gradient of F reaches zero in the solution, where

$$\nabla_{\mathbf{w}_i} F = \frac{2\mathbf{w}_i}{V_i} \bar{\mathbf{R}} - \frac{2\mathbf{w}_i}{U_i} \hat{\mathbf{R}} = 0 \quad (21)$$

Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations,

Table 1 Identified results of the experimental model

Mode	Frequency (Hz)		Damping ratio (%)		MAC
	ERA	STFT-ICA	ERA	STFT-ICA	
1	1.649	1.800	12.9	11.7	0.99
2	4.226	4.126	16.8	16.5	0.69
3	5.999	5.869	5.0	5.1	0.91

Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations,

Table 2 Identified results of the USC building

Mode	Frequency (Hz)		MAC
	Analytical	STFT-ICA	
1	0.746	0.766	0.98
2	1.786	1.907	0.93
3	3.704	3.941	0.30

Rewriting as

$$\mathbf{w}_i \bar{\mathbf{R}} = \frac{V_i}{U_i} \mathbf{w}_i \hat{\mathbf{R}} \quad (22)$$

yields a well-defined generalized eigenproblem (Stone 2001); the solution for $\mathbf{w}_i$ can thus be obtained as the eigenvector of the matrix $\hat{\mathbf{R}}^{-1} \bar{\mathbf{R}}$, with the eigenvalue $\gamma_i = V_i/U_i$. The sources can then be efficiently extracted simultaneously by

$$\mathbf{s}(t) = \mathbf{y}(t) = \mathbf{W} \mathbf{x}(t) \quad (23)$$

where the eigenvector matrix $\mathbf{W}$, with $\mathbf{w}_i$ as its i th row, is the target demixing matrix such that $\mathbf{A} = \mathbf{W}^{-1}$ and $\mathbf{y}(t) = [y_1(t), \dots, y_n(t)]^T$ is the recovered component vector which approaches the source vector $\mathbf{s}(t)$.

The aforementioned CP framework can then be cast into the modal identification whose physical interpretation is based on a new concept of “independent physical system on modal coordinates” (Yang and Nagarajaiah 2013b). The CP method is successful as shown in Figs. 6, 7, 8, and 9 and Tables 3 and 4 in identification of closely spaced couples with high-damped modes and in the presence of nonstationary seismic excitation.

Sparse Component Analysis of Modal Expansion

Existing BSS-based output-only modal identification methods may not be applied to the underdetermined problem where $m < n$ (i.e., only partial observations are available), which also often arises; e.g., for a large-scale structure, compared to its complexity with quite a few active modes, the measurement sensors may be limited. In this situation, the SCA method (Gribonval and Lesage 2006; Yang and Nagarajaiah 2013c) is used to tackle both the determined and underdetermined problem by exploiting the sparsity essence of the modal response. The targeted modal responses, which are viewed as sources in the BSS framework, are monotone, implying that they are active at only one distinct frequency, respectively. Therefore, they are most sparsely and disjointly distributed in the frequency domain and naturally satisfy the source sparsity assumption of SCA. Hence, transform the modal expansion Eq. 11 into the sparse frequency domain to incorporate modal identification to the SCA framework

$$\mathbf{x}(f) = \tilde{\Phi} \mathbf{q}(f) = \sum_{i=1}^n \varphi_i q_i(f) \quad (24)$$

with

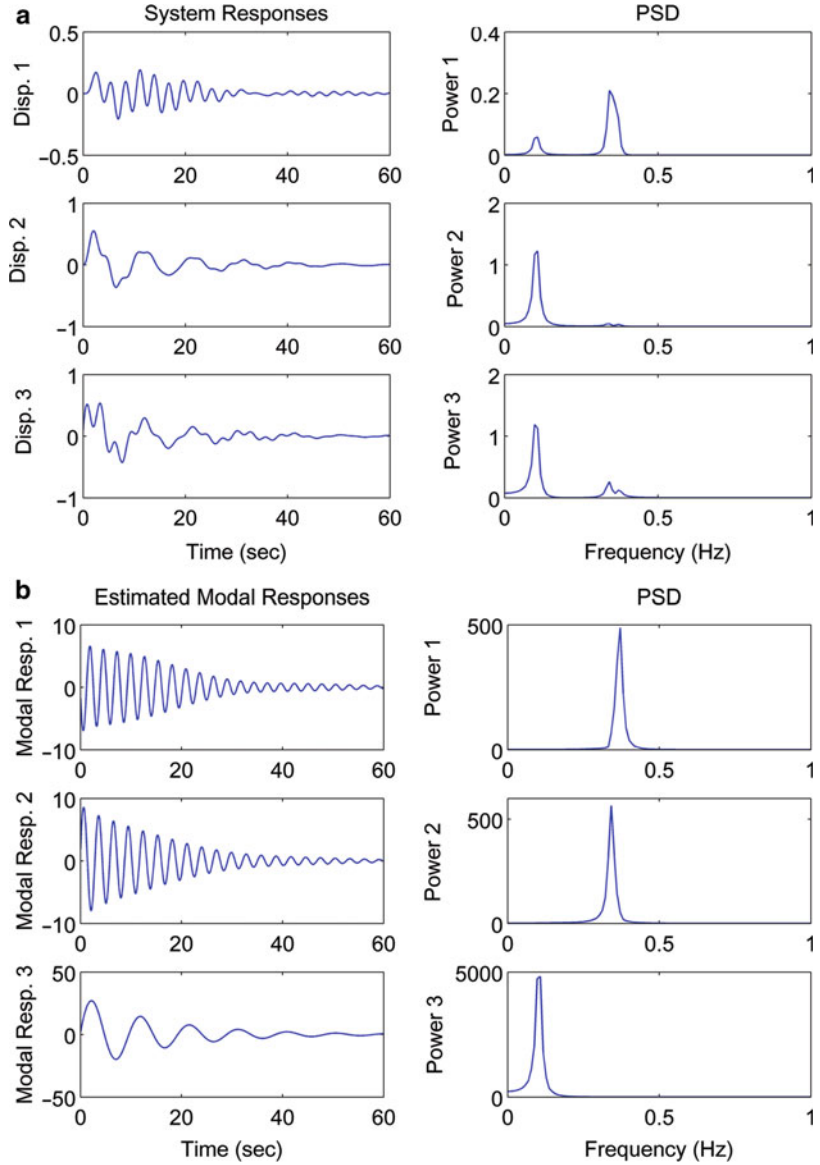
$$\begin{aligned} \mathbf{x}(f) &= \mathcal{F}(\mathbf{x}(t)) = \int_{-\infty}^{\infty} \mathbf{x}(t) e^{-j2\pi f t} dt \\ \mathbf{q}(f) &= \mathcal{F}(\mathbf{q}(t)) = \int_{-\infty}^{\infty} \mathbf{q}(t) e^{-j2\pi f t} dt \end{aligned} \quad (25)$$

where $\mathcal{F}$, f , and j denote the Fourier transform operator, frequency index, and the imaginary operator, respectively. Note that $\mathcal{F}$ is an invertible linear transform; it holds the form of the modal expansion such that Eq. 24 is valid and Φ remains invariant.

To avoid complex elements in Eq. 24, it is more practical to use the cosine transform $\mathcal{F}^c$ (also linear) to yield real-valued $\mathbf{x}(f)$, simply replacing the Fourier basis $e^{-j2\pi f t}$ with the cosine

Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations,

Fig. 6 (a) The system responses of a 3-DOF numerical model and (b) the modal responses recovered by CP in free vibration (closely-spaced coupled with high-damped modes)



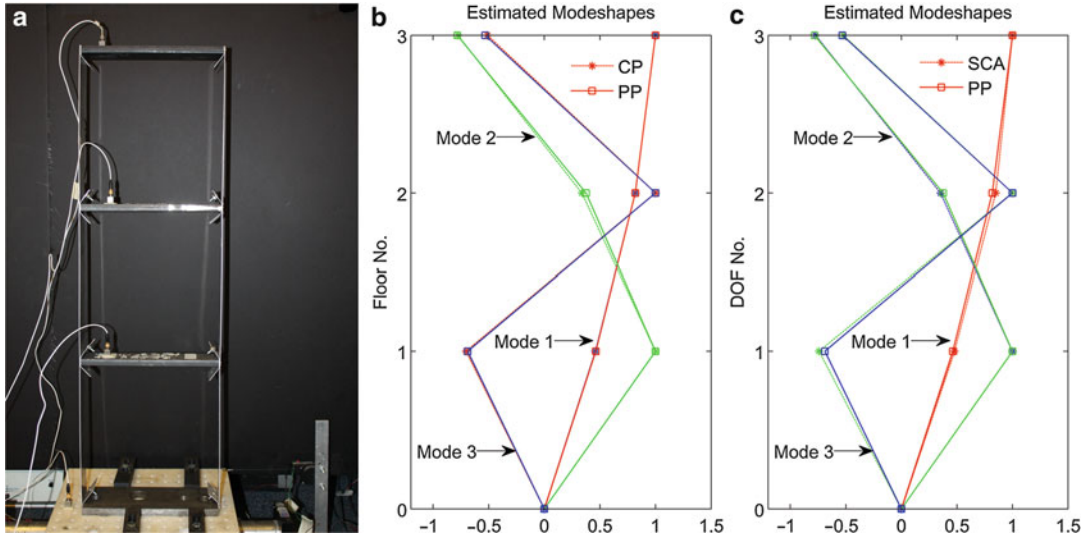
basis $\cos 2\pi ft$ in Eq. 24. The DCT is adopted in the proposed SCA method, then $\mathbf{x}(f)$ and $\mathbf{q}(f)$ are understood as real-valued cosine transform coefficients.

Using the disjoint sparsity property of modal responses with distinct frequencies, at some f_k where only one modal response

$q_j(j = 1, \dots, n)$ is active and $q_i = 0$ for $i \neq j$, Eq. 24 becomes

$$\mathbf{x}(f_k) = \varphi_i q_j(f_k) \quad (26)$$

Therefore, the points of $\mathbf{x}(f)$ will cluster to the direction of the j th modeshape $\varphi_j(j = 1, \dots, n)$



Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations, Fig. 7 (a) The experimental 3-story structure, (b) the modeshapes estimated by CP, and (c) the modeshapes estimated by SCA

such that the estimated vibration mode matrix $\tilde{\Phi}$ can be extracted by the automatic fuzzy C-means clustering algorithm.

For determined cases, $\tilde{\Phi}$ is square, and the time-domain modal responses can be recovered directly by

$$\tilde{\mathbf{q}}(t) = \tilde{\Phi}^{-1} \mathbf{x}(t) \quad (27)$$

from which the frequency and damping ratio can be estimated by straightforward Fourier transform and logarithm decrement, respectively. In underdetermined cases, $\tilde{\Phi}$ is rectangular. The frequency-domain modal sources $\tilde{\mathbf{q}}(f)$ can first be recovered using ℓ_1 -minimization (P_1): at each $f \in \Omega$,

$$\begin{aligned} \tilde{\mathbf{q}}(f) &= \arg \min \|\mathbf{q}(f)\|_{\ell_1} \quad \text{subject to} \quad \tilde{\Phi} \mathbf{q}(f) \\ &= \mathbf{x}(f) \end{aligned} \quad (28)$$

where $\|\mathbf{q}(f)\|_{\ell_1} = \sum_{i=1}^n |q_i(f)|$. The validity of this strategy resides in the ability of the ℓ_1 -minimization to recover the sparsest solution to Eq. 28, which is exactly the desired monotone frequency-domain modal responses since they are the sparsest solution among all feasible solutions to Eq. 28.

Using the inverse cosine transform, the time-domain modal responses can be readily recovered by

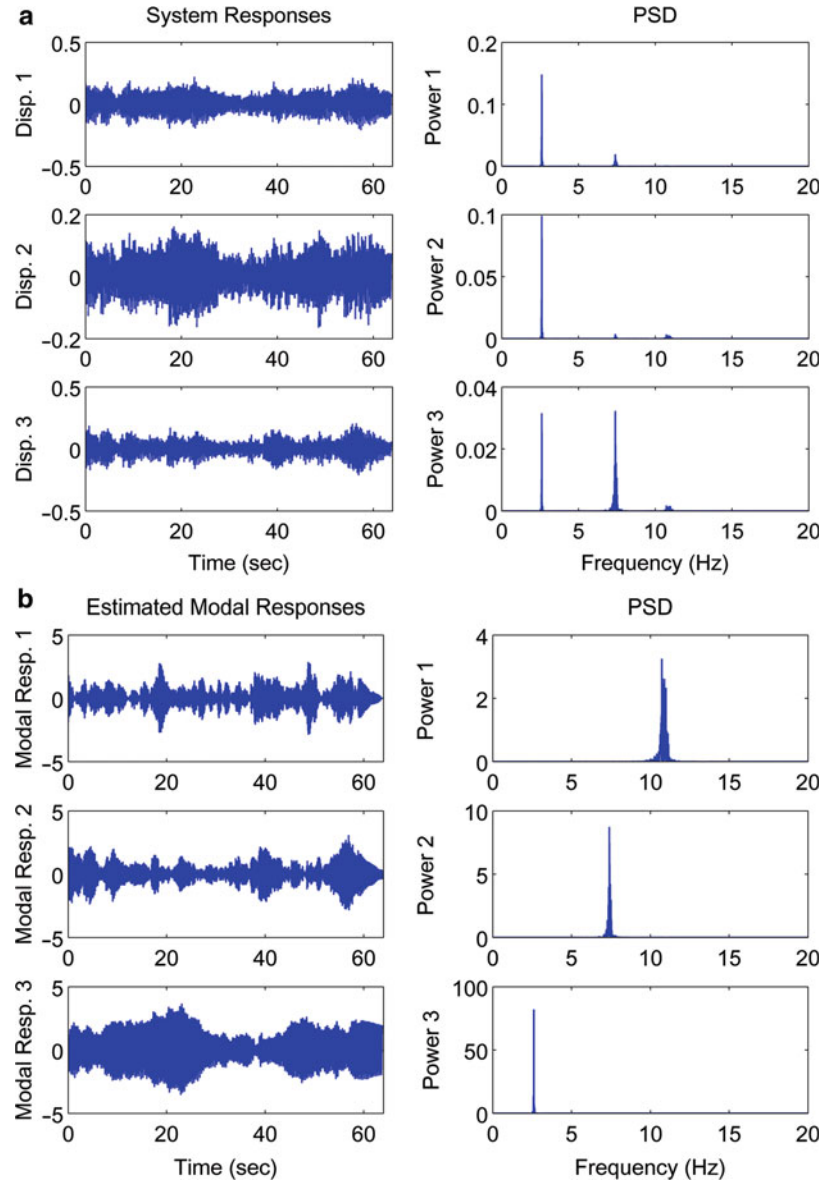
$$\tilde{\mathbf{q}}(t) = \mathcal{F}^{c-1}(\tilde{\mathbf{q}}(f)) \quad (29)$$

thereby estimating the frequency and damping ratio.

For demonstrations, Fig. 10 shows that SCA is able to recover the close modes coupled with high damping using only two sensors of the 3-DOF numerical model.

Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations,

Fig. 8 (a) The measured system responses and (b) the modal responses recovered by CP of the experimental model subject to white noise excitation



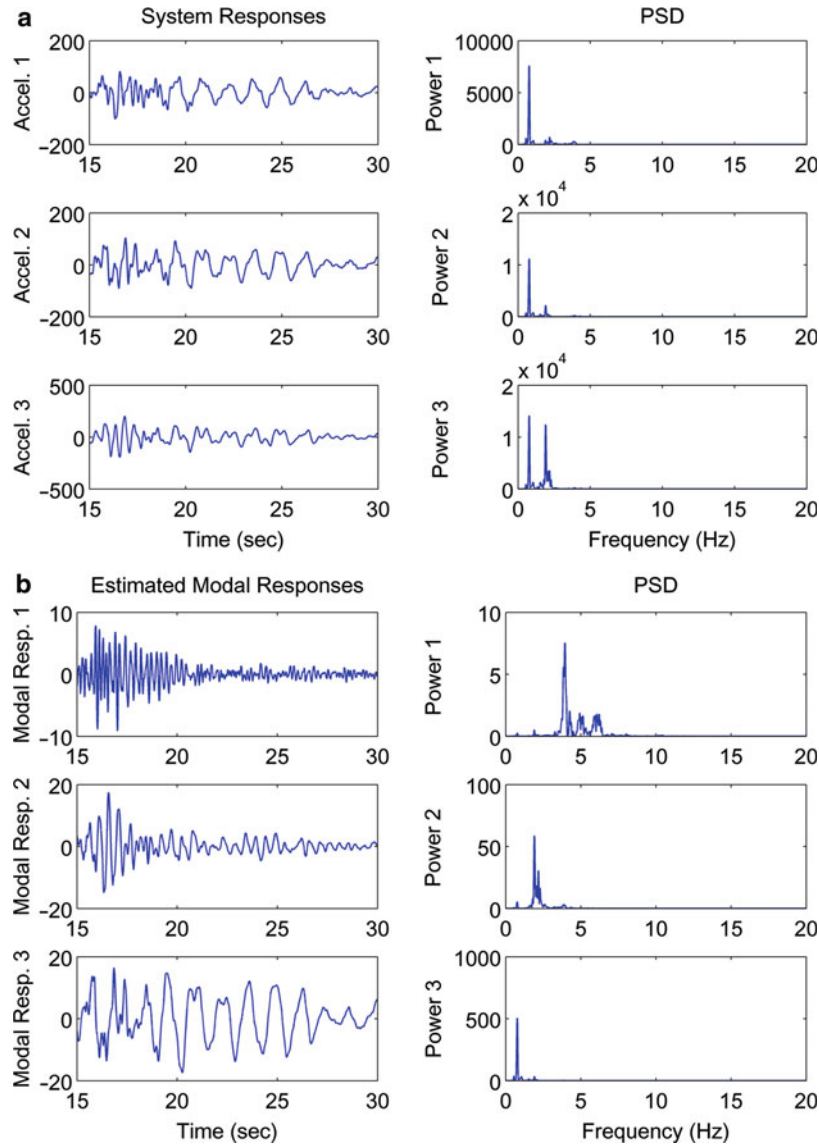
Figures 7c, 11, 12, and 13 and Table 5 show that SCA works very well even with limited sensors in the three-story experimental structure. Details are referred to Yang and Nagarajaiah (2013c).

Summary

This study shows that properly exploiting the sparse essences of modal expansion and damage information could efficiently and

Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations,

Fig. 9 (a) The measured system responses and (b) the modal responses recovered by CP of the USC hospital building in the Northridge earthquake 1994



Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations, Table 3 CP identified results of the experimental model

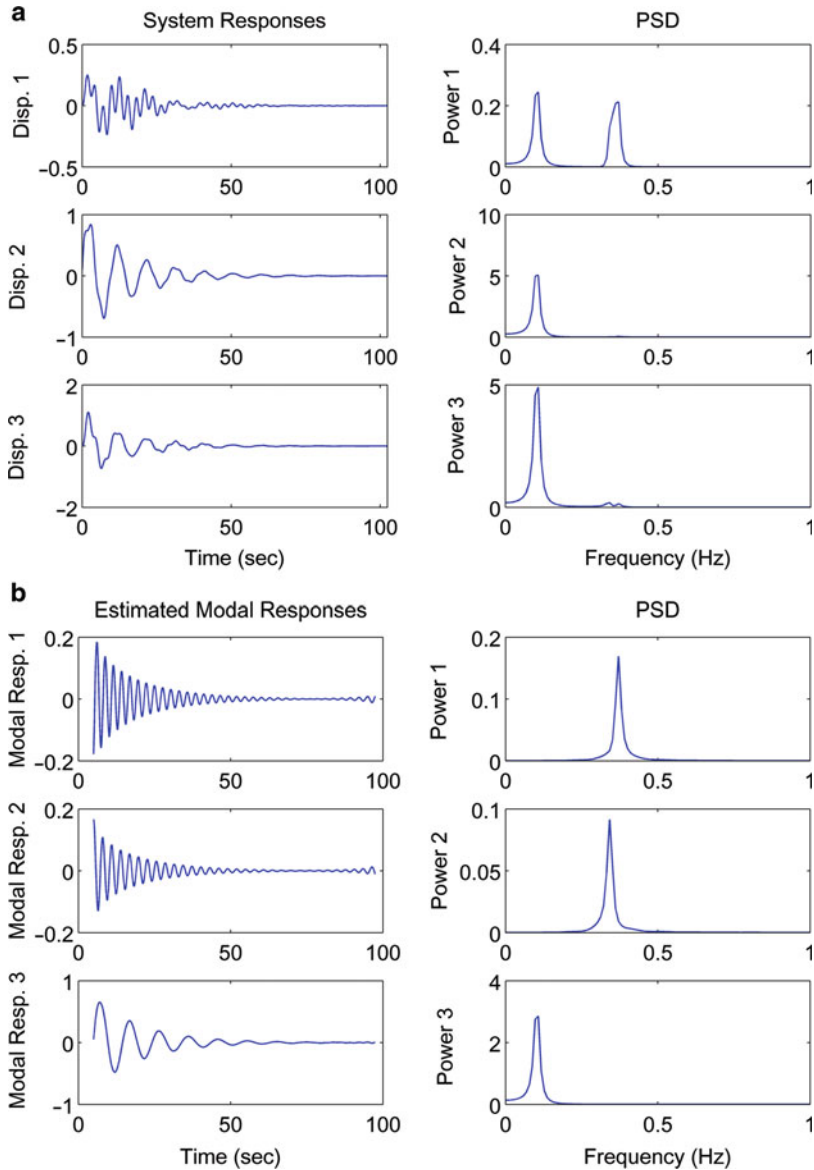
Mode	Frequency (Hz)		MAC
	Peak picking	CP	
1	2.550	2.600	1.0000
2	7.330	7.395	0.9993
3	10.460	10.720	0.9997

Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations, Table 4 CP identified results of the USC building

Mode	Frequency (Hz)		MAC
	Analytical	CP	
1	0.746	0.768	0.9751
2	1.786	1.907	0.9054
3	3.704	3.941	0.7874

Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations,

Fig. 10 The modal responses recovered by SCA using only two sensors (close modes coupled with high damping case)



effectively address some challenging problems in output-only modal identification and damage detection via the unsupervised blind source separation (BSS) method. Sparse ICA is first introduced to simultaneously identify

both damage time instants and damage locations and then further employed to exploit the sparse nature of modal expansion to handle the problem of identification of highly damped structures. What is more, two new

Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations,

Fig. 11 The free-vibration system responses of the 3-story experimental model (Fig. 7a)

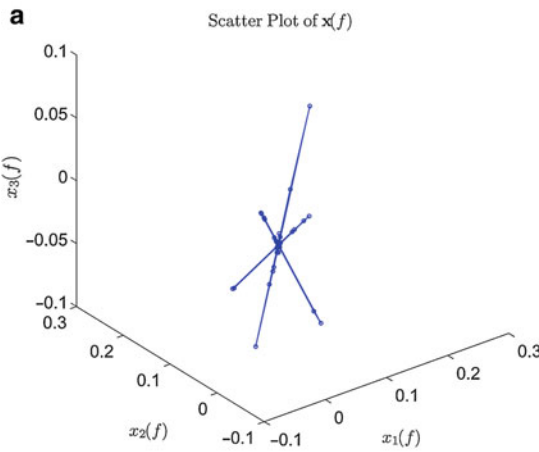
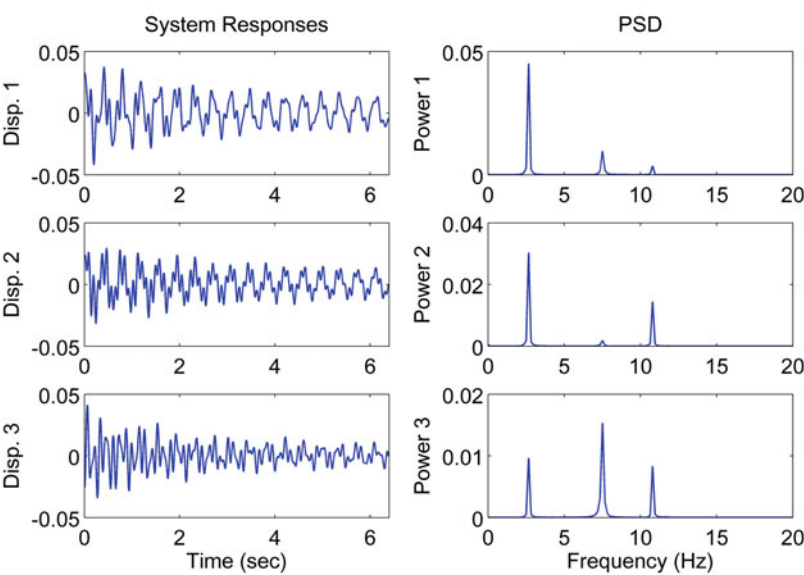
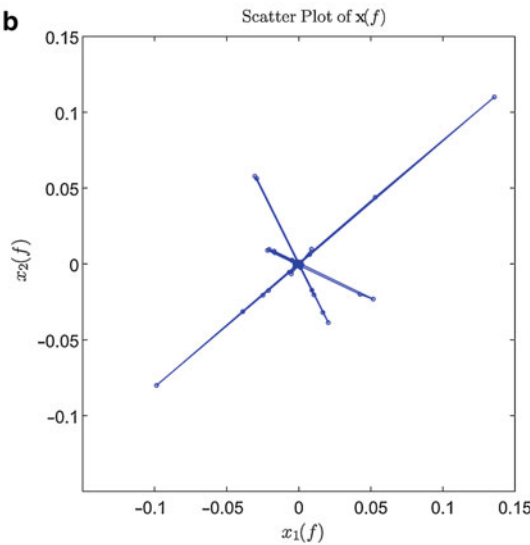


Fig. 12 The scatter plot in frequency domain of the



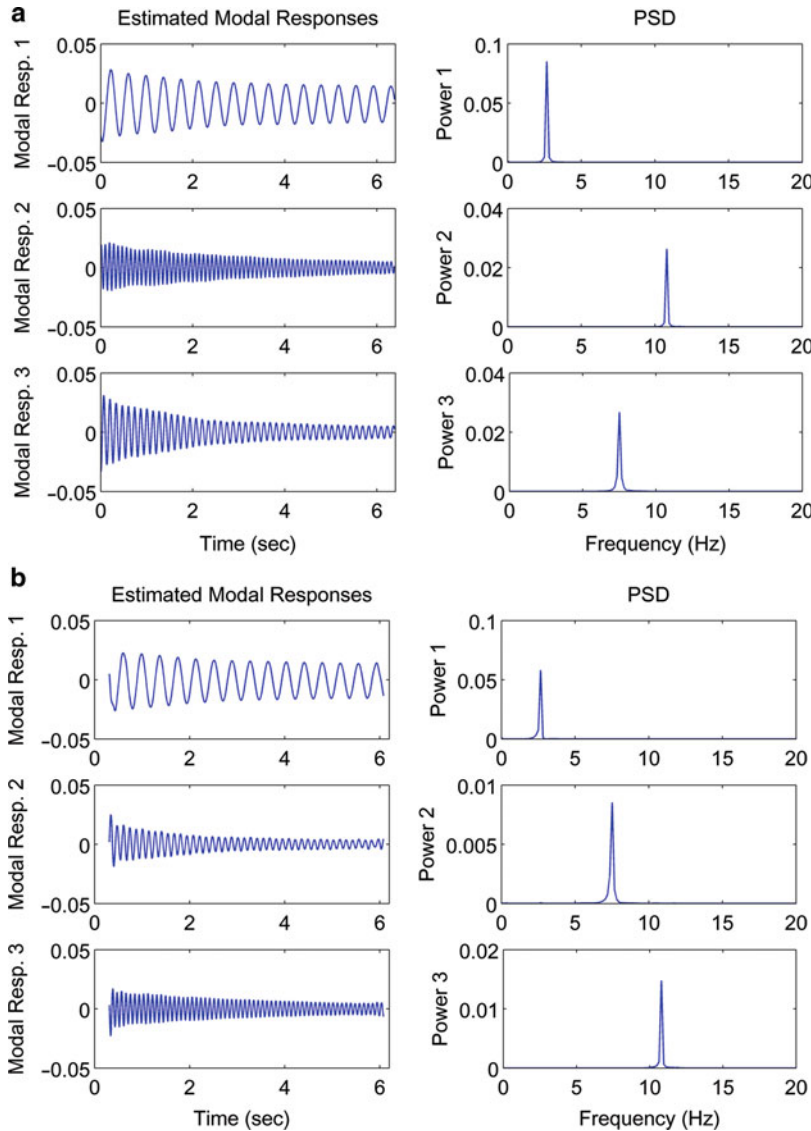
system responses of (a) all three sensors; (b) Sensor 1 and 2 in the 3-story experimental model (shown in Fig. 7a)

output-only modal identification methods are presented: the time-domain CP method and the SCA method which can handle the underdetermined problems with limited

sensors. The interpretations of CP and SCA in output-only modal identification are presented, and the successful implementations are also demonstrated using both

Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations,

Fig. 13 The modal responses recovered by SCA (a) using all the three sensors and (b) using Sensor 1 and 2 of the experimental model



Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations,
Table 5 Identified modal parameters by SCA of the experimental system

Mode	Frequency (Hz)			Damping ratio (%)		MAC	
	PP	Determined	Under-	Determined	Under-	Determined	Under-
1	2.55	2.66	2.66	1.12	1.04	0.9997	0.9995
2	7.33	7.51	7.51	1.08	1.03	0.9997	0.9996
3	10.46	10.80	10.80	0.68	0.67	0.9988	1.0000

experimental and real-world structure exam- efficient computation and little user involve-
ples. The established data-driven output-only ment, which is expected to have potential for
modal identification and damage identification real-time structural identification and health
framework is nonparametric and enjoys monitoring.

Cross-References

- [Advances in Online Structural Identification](#)
- [Ambient Vibration Testing of Cultural Heritage Structures](#)
- [Bayesian Operational Modal Analysis](#)
- [Modal Analysis](#)
- [Stochastic Structural Identification from Vibrational and Environmental Data](#)
- [System and Damage Identification of Civil Structures](#)
- [Vibration-Based Damage Identification: The Z24 Bridge Benchmark](#)

References

- Abazarsa F, Ghahari SF, Nateghi F, Taciroglu E (2013) Response-only modal identification of structures using limited sensors. *Struct Control Health Monit* 20:987–1006
- Antoni J (2005) Blind separation of vibration components: principles and demonstrations. *Mech Syst Signal Process* 19:1166–1180
- Antoni J, Chuah S (2013) A study and extension of second-order blind source separation to operational modal analysis. *J Sound Vib* 332:1079–1106
- Basu B, Nagarajaiah S, Chakraborty A (2008) Online identification of linear time-varying stiffness of structural systems by wavelet analysis. *Struct Health Monit Int J* 7(1):21–36
- Brewick P, Smyth A (2015) Exploration of the impacts of driving frequencies on damping estimates. *J Eng Mech* 141(3): 04014130
- Dharap P, Koh B, Nagarajaiah S (2006) Structural health monitoring using ARMarkov observers. *J Intell Mater Syst Struct* 17(6):467–481
- Ghahari S, Abazarsa F, Ghannad M, Taciroglu E (2013) Response-only modal identification of structures using strong motion data. *Earthq Eng Struct Dyn* 42:1221–1242
- Gribonval R, Lesage S (2006) A survey of sparse component analysis for blind source separation: principles, perspectives, and new challenges. In: *Proceedings of European symposium on artificial neural networks*, Bruges, Apr 2006, pp 323–330
- Hazra B, Narasimhan S (2010) Wavelet-based blind identification of the UCLA Factor building using ambient and earthquake responses. *Smart Mater Struct* 19(2):205005
- Hazra B, Roffel AJ, Narasimhan S, Pandey MD (2010) Modified cross-correlation method for the blind identification of structures. *J Eng Mech* 136(7):889–897
- Hyvärinen A, Oja E (2000) Independent component analysis: algorithms and applications. *Neural Netw* 13:411–430
- Kerschen G, Poncelet F, Golinval J-C (2007) Physical interpretation of independent component analysis in structural dynamics. *Mech Syst Signal Process* 21:1561–1575
- Koh BH, Dharap P, Nagarajaiah S, Phan MQ (2005a) Real-time structural damage monitoring by input error function. *AIAA J* 43(8):1808–1814
- Koh BH, Li Z, Dharap P, Nagarajaiah S, Phan MQ (2005b) Actuator failure detection through interaction matrix formulation. *J Guid Control Dyn AIAA* 28(5):895–901
- Li Z, Koh BH, Nagarajaiah S (2007) Detecting sensor failure via decoupled error function and inverse input-output model. *J Eng Mech ASCE* 133(11):1222–1228
- McNeill SI, Zimmerman DC (2008) A framework for blind identification using joint approximate diagonalization. *Mech Syst Signal Process* 22:1526–1548
- Nagarajaiah S, Basu B (2009) Output only identification and structural damage detection using time frequency and wavelet techniques. *Earthq Eng Eng Vib* 8:583–605
- Nagarajaiah S, Dharap P (2003) Reduced order observer based identification of base isolated buildings. *Earthq Eng Eng Vib* 2(2):237–244
- Nagarajaiah S, Li Z (2004) Time segmented least squares identification of base isolated buildings. *Soil Dyn Earthq Eng* 24(8):577–586
- Nagarajaiah S, Sun X (2000) Response of base-isolated USC hospital building in Northridge earthquake. *J Struct Eng ASCE* 126(10):1177–1186
- Nagarajaiah S, Sun X (2001) Base-isolated FCC building: impact response in Northridge earthquake. *ASCE J Struct Eng* 127:1063–1075
- Poncelet F, Kerschen G, Golinval J-C, Verhelst D (2007) Output-only modal analysis using blind source separation techniques. *Mech Syst Signal Process* 21:2335–2358
- Sadhu A, Hazra B, Narasimhan S, Pandey MD (2011) Decentralized modal identification using sparse blind source separation. *Smart Mater Struct* 20:125009
- Sadhu A, Hazra B, Narasimhan S (2012) Blind identification of earthquake-excited structures. *Smart Mater Struct* 21:045019
- Stone JV (2001) Blind source separation using temporal predictability. *Neural Comput* 13:1559–1574
- Stone JV (2004) Independent component analysis: a tutorial introduction. The MIT Press, Cambridge
- Xie S, He Z, Fu Y (2005) A note on Stone's conjecture of blind source separation. *Neural Comput* 17:321–330
- Yang Y, Nagarajaiah S (2013a) Time-frequency blind source separation using independent component analysis for output-only modal identification of highly-damped structures. *ASCE J Struct Eng* 139(10):1780–1793
- Yang Y, Nagarajaiah S (2013b) Blind identification of modal parameters in time domain based on complexity pursuit. *Earthq Eng Struct Dyn* 42(13):1885–1905

- Yang Y, Nagarajaiah S (2013c) Output-only modal identification with limited sensors using sparse component analysis. *J Sound Vib* 332:4741–4765
- Yang Y, Nagarajaiah S (2014a) Blind identification of damage in time-varying system using independent component analysis with wavelet transform. *Mech Syst Signal Process* 47(1):3–20
- Yang Y, Nagarajaiah S (2014b) Structural damage identification via a combination of blind feature extraction and sparse representation classification. *Mech Syst Signal Process* 45(1):1–23
- Yang Y, Nagarajaiah S (2014c) Blind denoising of structural responses with outliers via principal component pursuit. *Struct Control Health Monit* 21(6):962–978
- Yang Y, Nagarajaiah S (2014d) Data compression of structural seismic responses using principled independent component analysis. *ASCE J Struct Eng* 140(7):04014032
- Zhou W, Chelidze D (2007) Blind source separation based vibration mode identification. *Mech Syst Signal Process* 21(8):3072–3087

Bridge Foundations

Ayman A. Shama¹ and Hesham El Naggar²

¹Ammann & Whitney, New York, NY, USA

²Geotechnical Research Centre, Department of Civil and Environmental Engineering, The University of Western Ontario, London, ON, Canada

Introduction

The poor performance of bridges during the 1971 San Fernando earthquake has urged the engineering community to further investigate methods of analysis and evaluation of seismic loads and their effects on these structures. Later, earthquakes such as Northridge, 1994, and Kobe, 1995, have further increased engineers' awareness of the effects of earthquakes on bridges. The past two decades have witnessed a great evolution in seismic design codes, advance modeling, and analysis. Moreover, the importance of the interaction between the structure and the surrounding soil on the seismic response of bridge structures during earthquakes has been fully realized.

Past performance of foundations was satisfactory in moderate earthquakes and poor in large

earthquakes with magnitudes greater than seven. Examples include the 1964 M9.2 Alaskan Earthquake, the 1991 M7.6 Costa Rica Earthquake, the 1999 M7.8 Izmit-Turkey Earthquake, the 2010 M8.8 Chile Earthquake, the 2011 M9.0 Tohoku-Oki Earthquake in Japan, and the 2012 M7.4 Guatemala Earthquake. In all these earthquakes, spread footings failures were due to loss of bearing capacity and lateral spreading due to soil liquefaction. Figure 1 depicts photo of a collapsed bridge during the Limon Province, Costa Rica Earthquake. The piers have disappeared in the river a result of liquefaction and both bridge spans have collapsed during the earthquake.

The seismic response of bridge foundations depends on: the type, geometry, and embedment of the foundation; the characteristics of the surrounding soil; and the interaction between the foundation and the surrounding soil. Soil–structure interaction is the influence of the soil and structure both on each other simultaneously. Seismic ground motions that are not influenced by the presence of the structure are known as free-field motions. If the structure exists, the deformations of the foundations deviate from the deformations of the free-field motion and the dynamic response of the structure induces deformation of the supporting soil. The process in which the response of the soil influences the motion of the structure and the motion of the structure influences the response of the soil is known as soil–structure interaction (SSI). There are two approaches used to incorporate the effect of SSI in seismic analyses of structures, direct methods and simplified methods. In the direct approach, the entire soil–foundation–structure system is modeled and analyzed simultaneously in one step using finite element or finite difference methods with the seismic waves applied at the model boundaries. This approach requires specific programs that can idealize the nonlinear behavior of the soil and large CPU time for analysis. Although it is conceptually very attractive, it is fairly complex and very seldom been applied for practical bridge problems. Simplified methods include the substructure approach and the Winkler approach. The substructure approach includes subdividing the structure into two

Bridge Foundations,

Fig. 1 A bridge collapse
due to foundation failure
Costa Rica
Earthquake 1993



B

substructures, foundation and superstructure, with a convenient interface between the two. First, the stiffness and damping properties of the foundation system, for each degree of freedom, are first evaluated and then applied to the bridge superstructure in the form of springs and dashpots at the point of interface. The Winkler spring model (Winkler 1867) acts in conjunction with the foundation to eliminate rotational springs. The p-y curve method, which was developed for off-shore structures and used for analysis of deep foundations, originates from the Winkler spring method.

The most common types of bridge foundations are: (i) spread footings for sites of rock and stiff soil; (ii) pile-supported cap foundations for sites with soft soil or for sites with soil layers vulnerable to liquefaction; (iii) drilled shafts for cohesive soils, especially with deep groundwater or when construction of new foundations requires a small footprint wherein, a single drilled shaft under a single column can avoid the large footprint that would be necessary with a group of piles; and (iv) large gravity caissons for complex bridges such as cable supported bridges.

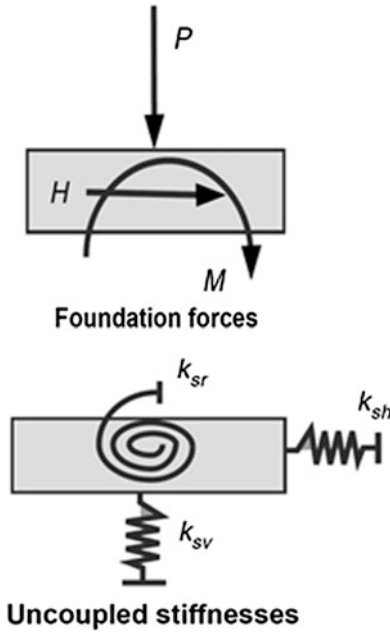
The subsequent sections of this chapter discuss the procedures for modeling bridge foundations for analysis including effects of SSL. Methods for designing these foundations to resist seismically induced forces and displacements are also explained.

Spread Footings

Spread footings are usually suitable for sites of rock and firm soils. The stability of these foundations under seismic loads can be evaluated using a pseudostatic bearing capacity procedure. The applied loads for this analysis can be taken directly from the results of a global dynamic response analysis of the bridge with the soil-foundation-interaction effects represented in the structural model.

Spread Footing Stiffness

A method to be selected for evaluating the foundation stiffness must adequately reflect the shape of the foundation–soil interface; the amount of embedment; the nature of the soil profile; and the mode of vibration and frequencies of excitation. The uncoupled spring approach satisfies all these conditions. It is accomplished through determining the dynamic impedance functions for the foundation. This method is adequate if the seismic foundation loads are not expected to exceed twice the ultimate foundation capacities. As illustrated in Fig. 2, the dynamic impedance model is an uncoupled single node model that represents the foundation element. An upper and lower bound approach to evaluating the foundation stiffness is often used because of the uncertainties in the soil properties and the static loads on the foundations. As a general rule of thumb, a factor



Bridge Foundations, Fig. 2 Analysis model for spread footings

of 4 is taken between the upper and lower bound (ATC-1995).

The dynamic impedance model is based on earlier studies on machine foundation vibrations, in which, it is assumed that the response of rigid foundations excited by harmonic external forces can be characterized by the impedance or dynamic stiffness matrix for the foundation. The impedance matrix depends on the frequency of excitation, the geometry of foundation and the properties of the underlying soil deposit.

The evaluation of the impedance functions for a foundation with an arbitrary shape has been solved mathematically using a mixed boundary-value problem approach or discrete variation problem approach. Currently, there are two commonly used approaches for evaluating the dynamic impedance functions for a shallow foundation. The first is based on the approximate solution for a circular footing rigidly connected to the surface of isotropic homogeneous elastic half space. The second approach is more general and is applicable to a foundation with an arbitrary shape.

In the equivalent circular base approach the impedance function of a foundation is obtained from the elastic solution of a rigid massless circular base resting on the surface of the soil for each degree of freedom independently. The impedance function for each degree of freedom is a frequency dependent complex expression, where its real part represents the elastic stiffness (spring constant) of the soil-foundation system and its imaginary part represents the damping in the soil-foundation system. The impedance function is expressed as:

$$\bar{k}_j = k_j(a_0, \nu) + i\omega c_j(a_0, \nu) \quad (1)$$

where, k_j is the real part represent the dynamic stiffness for a specific degree of freedom j , c_j is the imaginary part represent the damping for a degree of freedom j , ν is Poisson's ratio of the soil medium, a_0 is a dimensionless frequency expressed as:

$$a_0 = \frac{\omega r_j}{V_s} \quad (2)$$

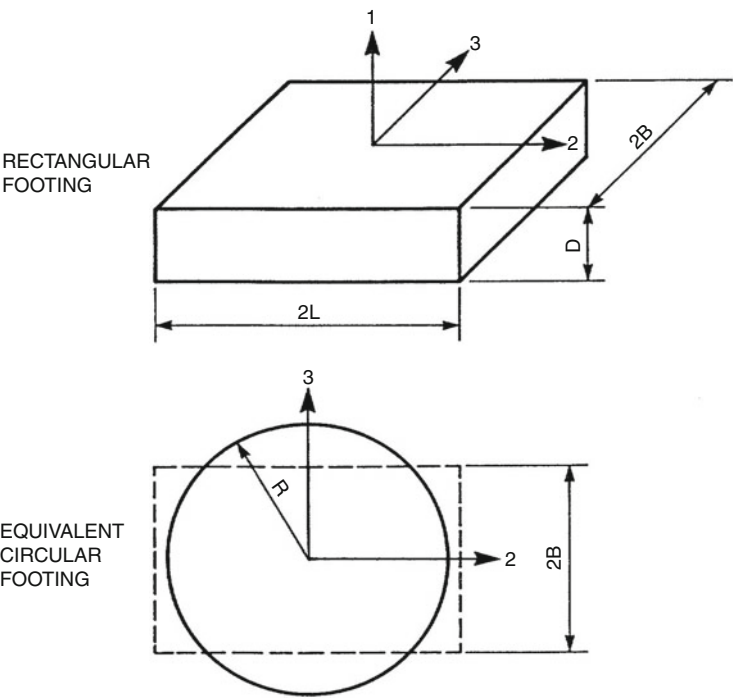
in which, ω is the circular frequency of excitation in rad/s, r_j is equivalent radius for a specific degree of freedom, and V_s is the shear wave velocity of the soil medium. Both the dynamic stiffness k_j and the damping c_j are obtained in terms of the static stiffness K_j as:

$$k_j = \alpha_j K_j \text{ and } c_j = \beta_j \frac{K_j r_j}{V_s} \quad (3)$$

where α_j and β_j are dynamic modifiers are evaluated as function of the dimensionless frequency a_0 and Poisson's ratio ν .

Evaluation of the static stiffness coefficients can be obtained using the equivalent circular footing, wherein it is assumed that the footing is rigid and founded on top of a semi-infinite elastic half space. The rectangular footing is converted to an equivalent circular footing for each degree of freedom. The stiffness coefficients are evaluated for the circular footing and then multiplied

Bridge Foundations,
Fig. 3 Procedure for
evaluating equivalent
radius of a rectangular
footing



EQUIVALENT RADIUS:

TRANSLATIONAL:

$$R_o = \sqrt{\frac{4BL}{\pi}}$$

ROTATIONAL:

$$R_1 = \left[\frac{(2B)(2L)^3}{3\pi} \right]^{1/4} \qquad \text{(x-AXIS ROCKING)}$$
$$R_2 = \left[\frac{(2B)^3(2L)}{3\pi} \right]^{1/4} \qquad \text{(y-AXIS ROCKING)}$$
$$R_3 = \left[\frac{4BL(4B^2 + 4L^2)}{6\pi} \right]^{1/4} \qquad \text{(z-AXIS TORSION)}$$

by shape and embedment factors to obtain the solution for the rectangular base. The solution is carried out in five steps as follows:

Step 1: Determine the equivalent radius for each degree of freedom, which is the radius of a circular footing with the same area as the rectangular footing as shown in Fig. 3.

Step 2: Calculate the stiffness coefficients K_0 for the transformed circular footing (Table 1), where, G and ν in the table are the shear modulus and Poisson’s ratio for the soil-foundation system.

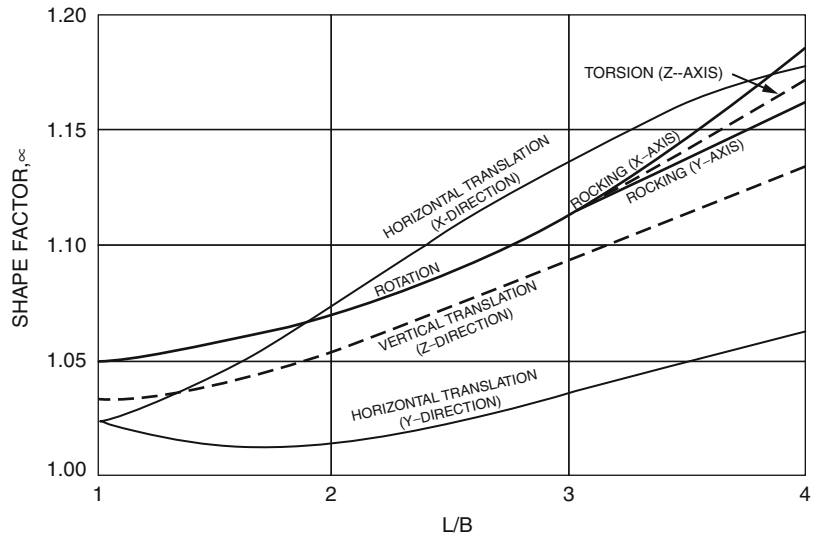
Bridge Foundations, Table 1 Static stiffness coefficients for a circular footing resting on the surface of soil

Displacement degree-of-freedom	k_0
Vertical translation	$\frac{4Gr}{1-\nu}$
Horizontal translation	$\frac{8Gr}{2-\nu}$
Torsional rotation	$\frac{16}{3}Gr^3$
Rocking rotation	$\frac{8Gr^3}{3(1-\nu)}$

Step 3: Multiply each of the stiffness coefficients values obtained in step 2 by the appropriate shape correction factor α from Fig. 4 (Lam and Martin 1986). This figure provides the shape

B

Bridge Foundations,
Fig. 4 Shape factors for
 rectangular footings (Lam
 and Martin 1986)



factors for different aspect ratios L/B for the foundation.

Step 4: Multiply the values obtained from step 3 by the embedment factor β using Fig. 5. D in these figures is the footing thickness.

Step 5: Calculate the dynamic stiffness and the damping coefficients using Eq. 3 and the charts displayed in Figs. 6, 7, and 8 (Veletsos and Verbic 1973),

The static stiffness can be calculated using the second approach for calculating the impedance functions for the soil-foundation system (Gazetas 1991). Using Fig. 9, a two-step calculation process is required. First, the stiffness terms are calculated for a foundation at the surface. Then, an embedment correction factor is calculated for each stiffness term. The stiffness of the embedded foundation is the product of these two terms. According to Gazetas, the height of effective sidewall contact, d , in Fig. 9 should be taken as the average height of the sidewall that is in good contact with the surrounding soil.

Considering the range of frequencies and amplitudes in earthquake ground motions compared to machine foundations, it is reasonable to ignore the frequency dependence of the stiffness as well as the damping parameters.

Based upon the results of the dynamic analysis, the peak dynamic loads and deformations of

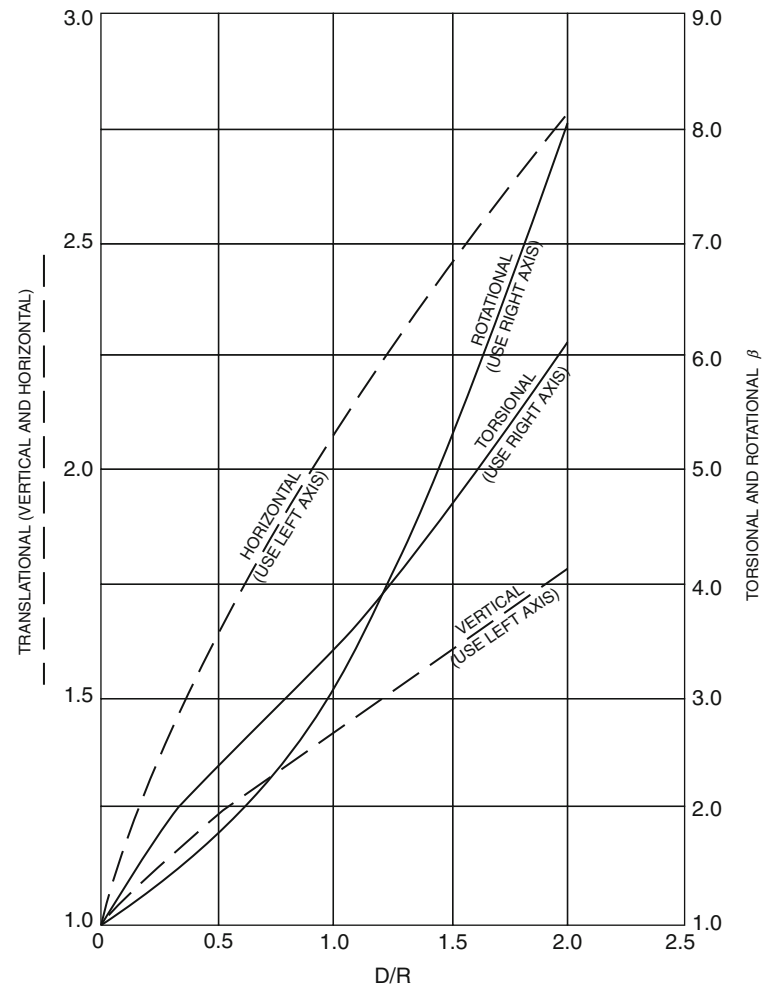
the foundation are determined and compared to acceptable values. The foundation must be evaluated for bearing capacity and sliding resistance due to the seismic loads.

Seismic Design of Spread Footings

Earthquakes will induce moments and horizontal loads in addition to the traditional vertical loads applied to a spread footing. In low seismic hazard areas, spread footings may be proportioned to resist overturning, sliding, flexure, and shear due to forces obtained from the seismic analysis of the bridge. In high seismic regions, the flexural and shear demands are those associated with the over-strength plastic moment capacity of the column or pier attached to the footing. To represent the combined effect of the seismic forces and moments a resultant load that may have to be inclined and applied eccentrically can be applied in lieu of the seismic vertical forces, seismic horizontal forces, and seismic moments. Therefore, a procedure is established to account for load inclination and load eccentricity of footing through which the general bearing capacity equation of shallow footing is adjusted to account for these effects. This procedure is carried out in three steps as follows:

Step 1: Compute the seismic vertical loads, seismic horizontal loads, and seismic moments

Bridge Foundations,
Fig. 5 Embedment factors
for rectangular footings
(Lam and Martin 1986)

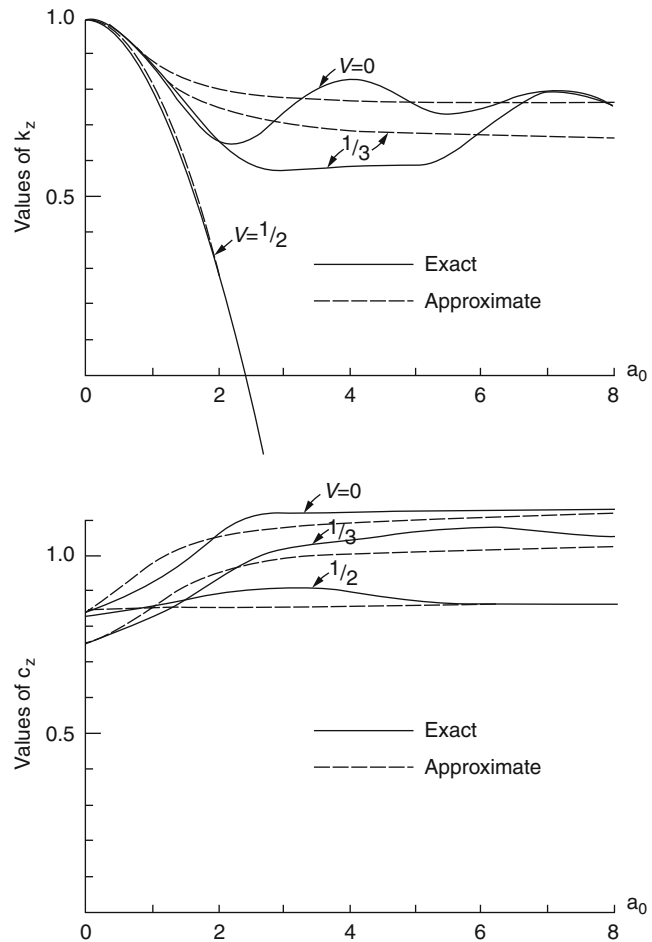


imposed to the footing. These seismic loads and moments can be taken directly from the results of a global dynamic response analysis of the structure with the soil–foundation interaction SFI effects represented in the structural model. For each direction, these forces are then combined into a single resultant force with an inclination angle β with respect to the vertical.

Step 2: Calculate the equivalent dimensions for the footing to account for the load eccentricity, which is caused by the seismic moments applied to the foundation in both directions. The vertical load can be transferred to an eccentric position defined by $e_b = M_b/Q$ and $e_l = M_l/Q$, where Q is the central vertical load due to seismic load plus other service loads;

M_b and M_l = seismic moments about the short and long axes of the footing; and e_b and e_l = eccentricities of the load Q about the centroid of the footing in the direction of the short and long axes respectively.

It is known from basic principles of strength of materials that if the eccentricity in one direction is less than 1/6 of the foundation's length in that direction, the footing is in compression throughout. As eccentricity exceeds this value, a loss of contact occurs. The concept of effective width was introduced by Meyerhof (1953) who proposed that at the ultimate bearing capacity of the foundation, it could be assumed that the contact pressure is identical to that for a centrally loaded foundation but of reduced width. The reduced

Bridge Foundations,**Fig. 6** Dynamic modifiers for vertically excited footing

dimensions as indicated in AASHTO-LRFD (AASHTO 2012) for a footing subjected to two-way eccentricity can be adopted as illustrated in Fig. 10.

Step 3: Adjust the static bearing capacity equation for inclination and eccentricity. Load inclination factors may be calculated as Vesic 1970 and added to the static bearing capacity equation to count for load inclination.

The sliding resistance of a spread footing should be evaluated independently of the bearing capacity, wherein, the unit adhesion and/or frictional resistance of the footing's base to sliding are multiplied by the area of the base to obtain the sliding resistance.

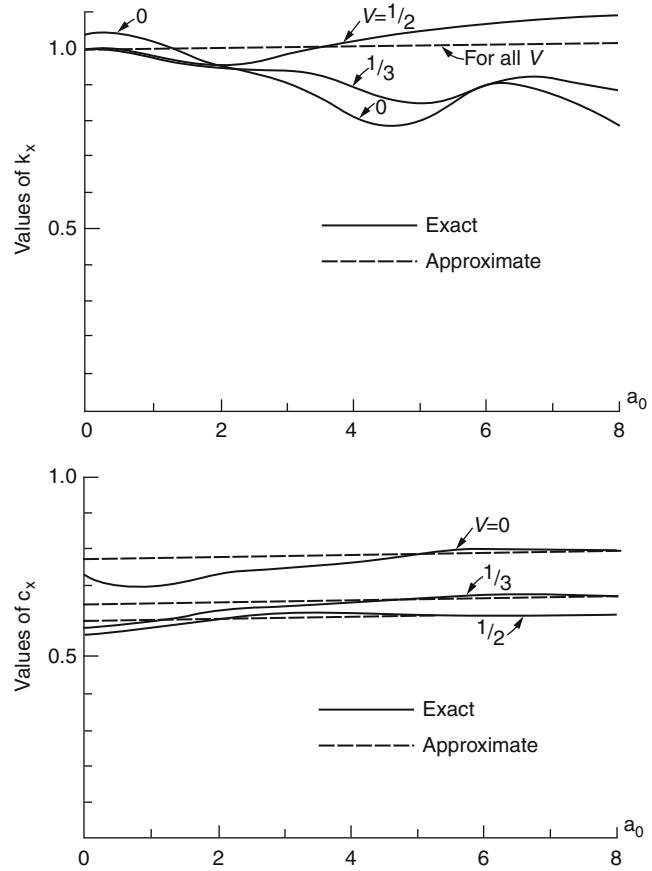
Spread footings are not allowed to be located in soils that are susceptible to liquefaction unless the footing is below the

maximum depth of liquefaction. Soil densifications or other acceptable methods for mitigating the potential for liquefaction may be used. Spread footings can be used in this case only if verification studies confirm that liquefaction potential has been mitigated.

Pile Footings

Pile footings generally consist of pile groups connected to a pile cap with pile diameters usually less than or equal to 24 in. They are preferred if the upper soil layers are weak or susceptible to liquefaction. Essentially, the seismic response of piles requires consideration of six degrees of freedom; that is, three translational components and three rotational components. The lateral soil

Bridge Foundations,
Fig. 7 Dynamic modifiers
for horizontally excited
footing



B

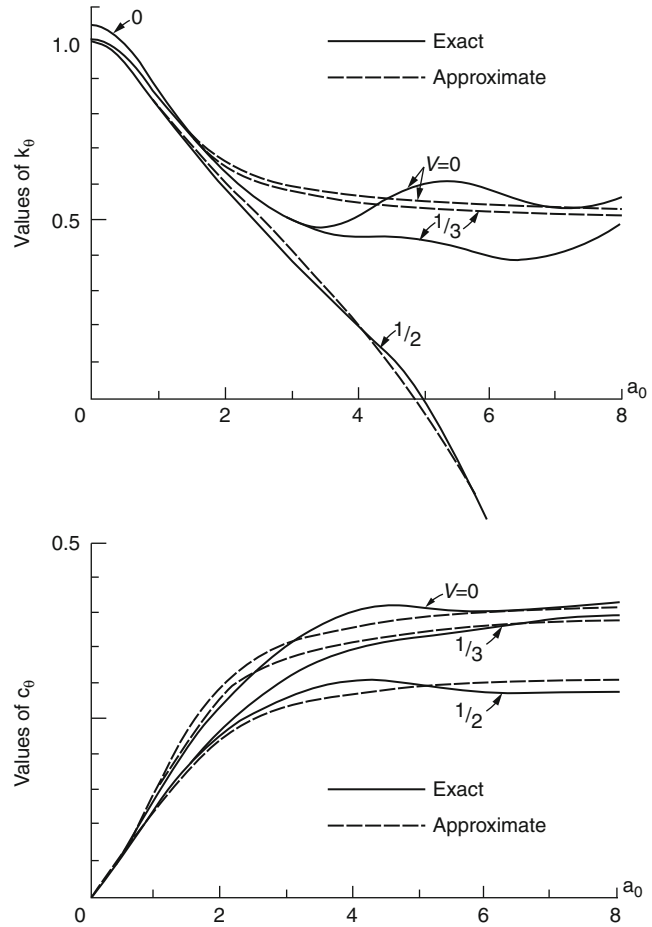
reactions are usually mobilized along the top 5–10 pile diameters. The axial soil resistances, however, develop at greater depths. Hence, the axial and lateral capacities of piles are considered to be uncoupled.

In general, seismic design of pile foundations is evaluated by the substructure method. First, the total structure is divided into two substructure models with a convenient interface between the two at the pile cap level. Next, the stiffness matrix of the pile group, which implies soil effects, is established using the foundation substructure model or simple methods. The stiffness matrix is then implemented in the superstructure model and seismic analysis is conducted using the superstructure model to determine the demands at the interface point. Finally, capacities of individual piles are evaluated and compared to the demands by back substitution in the substructure model. For a proper determination of the seismic

demand, it is imperative to estimate precisely the foundation stiffness to be included in the overall structural model for determination of the demands.

Pile Foundation Stiffness

Soils are inherently nonlinear, starting from incredibly small load levels. Lateral loads on piles are resisted by the surrounding soil. Therefore, piles exhibit nonlinear load deflection characteristics. This behavior is represented assuming the pile as a beam supported on Winkler springs that are characterized by nonlinear p-y curves for lateral loading, or t-z and q-z for vertical loading. These curves characterize the lateral soil resistance per unit length of pile as a function of the displacement. These relationships are generally developed on the basis of semiempirical curves, which reflect the nonlinear resistance of the local soil surrounding the pile at a certain depth.

Bridge Foundations,**Fig. 8** Dynamic modifiers for footing in rocking motion

The two most commonly used p-y models are those proposed by Matlock (1970) for soft clay and by Reese et al. (1974) for sand. The most commonly used t-z and q-z models are those developed by McVay et al. (1998).

There are two methods for modeling the behavior of pile groups for seismic response studies, the soil-pile stiffness as will be explained in the following subsections.

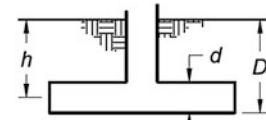
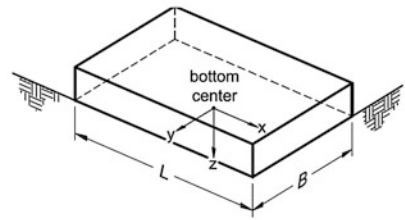
Coupled Pile Foundation Stiffness Matrix

In this method, a quasi-dynamic analysis for the pile group is conducted by applying loading (either as forces or displacements) at the interface node between the superstructure and foundation model using linearized properties for the soils. Linearized properties for a single pile can be achieved by assuming secant foundation stiffness

at 0.5–0.65 of peak deflection (Lam et al. 1998). A stiffness matrix can be obtained by prescribing a unit deformation vector for each degree of the six degrees of freedom, while keeping the other five degrees zero. The resultant force vector corresponding to each unit deformation vector can be used to form the corresponding column vector in the stiffness matrix. The stiffness matrix must be positive definite otherwise, numerical problems may be expected when the stiffness matrix is implemented in the overall structural model. One way to ensure that the stiffness matrix is positive definite is to invert it and check that the diagonal elements in the inverted (compliance) matrix are positive values. Programs such as LPILE (2010), FLPIER (Hoit and McVay 2010), and SWM (Ashour and Norris 2000) may be used to establish the foundation

Stiffness Parameter	Stiffness at Surface, K_i
Vertical Translation, K_z	$\frac{GL}{(1-\nu)} \left[0.73 + 1.54 \left(\frac{B}{L} \right)^{0.75} \right]$
Horizontal Translation, K_y (toward long side)	$\frac{GL}{(2-\nu)} \left[2 + 2.5 \left(\frac{B}{L} \right)^{0.85} \right]$
Horizontal Translation, K_x (toward short side)	$\frac{GL}{(2-\nu)} \left[2 + 2.5 \left(\frac{B}{L} \right)^{0.85} \right] - \frac{GL}{(0.75-\nu)} \left[0.1 \left(1 - \frac{B}{L} \right) \right]$
Rotation, K_{θ_x} (about x axis)	$\frac{G}{(1-\nu)} I_x^{0.75} \left(\frac{L}{B} \right)^{0.25} \left(2.4 + 0.5 \frac{B}{L} \right)$
Rotation, K_{θ_y} (about y axis)	$\frac{G}{(1-\nu)} I_y^{0.75} \left[3 \left(\frac{L}{B} \right)^{0.15} \right]$

Stiffness Parameter	Embedment Factors, e_i
Vertical Translation, e_z	$\left[1 + 0.095 \frac{D}{B} \left(1 + 13 \frac{B}{L} \right) \right] \left[1 + 0.2 \left(\frac{2L + 2B}{LB} d \right)^{0.67} \right]$
Horizontal Translation, e_y (toward long side)	$\left[1 + 0.15 \left(\frac{2D}{B} \right)^{0.5} \right] \left\{ 1 + 0.52 \left[\frac{\left(D - \frac{d}{2} \right) 16(L+B)d}{BL^2} \right]^{0.4} \right\}$
Horizontal Translation, e_x (toward short side)	$\left[1 + 0.15 \left(\frac{2D}{L} \right)^{0.5} \right] \left\{ 1 + 0.52 \left[\frac{\left(D - \frac{d}{2} \right) 16(L+B)d}{LB^2} \right]^{0.4} \right\}$
Rotation, e_{θ_x} (about x axis)	$1 + 2.52 \frac{d}{B} \left(1 + \frac{2d}{B} \left(\frac{d}{D} \right)^{-0.20} \left(\frac{B}{L} \right)^{0.50} \right)$
Rotation, e_{θ_y} (about y axis)	$1 + 0.92 \left(\frac{2d}{L} \right)^{0.60} \left(1.5 + \left(\frac{2d}{L} \right)^{1.9} \left(\frac{d}{D} \right)^{-0.60} \right)$



d = height of effective sidewall contact (may be less than total foundation height)

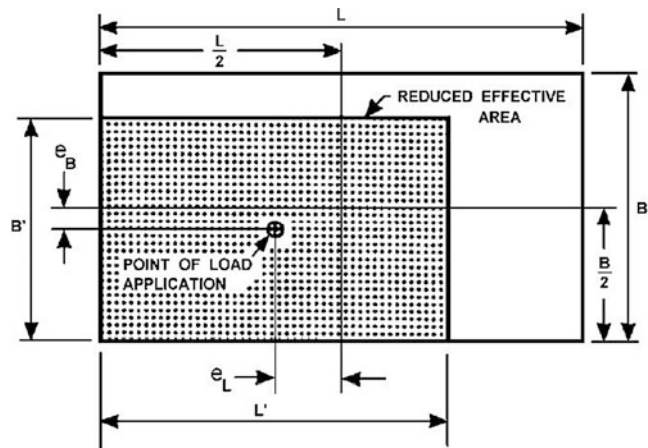
h = depth to centroid of effective sidewall contact

For each degree of freedom, calculate

$$K_{emb} = \beta K_{sur}$$

Bridge Foundations, Fig. 9 Static stiffness coefficients (Gazetas 1991)

Bridge Foundations, Fig. 10 Effective footing dimensions due to eccentric loads (After AASHTO LRFD 2012)



$$B' = B - 2e_B$$

$$L' = L - 2e_L$$

where:

e_B = eccentricity parallel to dimension B (ft)

e_L = eccentricity parallel to dimension L (ft)

stiffness matrix of the pile group through an iterative process.

Simplified Procedure for Pile Group Stiffness Matrix

The simplified method assumes coupling only between shear and overturning moment for the single pile. The general form of a single pile stiffness matrix can take the form:

$$\begin{bmatrix} K_{xx} & 0 & 0 & 0 & -K_{x\theta_y} & 0 \\ 0 & K_{yy} & 0 & K_{y\theta_x} & 0 & 0 \\ 0 & 0 & K_{zz} & 0 & 0 & 0 \\ 0 & K_{\theta_x y} & 0 & K_{\theta_x x} & 0 & 0 \\ -K_{\theta_y x} & 0 & 0 & 0 & K_{\theta_y y} & 0 \\ 0 & 0 & 0 & 0 & 0 & K_{\theta_z} \end{bmatrix} \quad (4)$$

in which, K_{xx} , K_{yy} are the lateral stiffnesses; K_{zz} is the axial stiffness; and $K_{x\theta_y}$ and $K_{y\theta_x}$ are corresponding coupled stiffnesses between shear and overturning moment.

The simplified method involves five basic steps (Lam et al. 1991) as follows:

1. Determine the stiffness coefficient of a single pile under lateral loading.
2. Determine the stiffness coefficient of a single pile under axial loading.
3. Superimpose the stiffness of individual piles to obtain the pile group stiffness.
4. Solve for the stiffness contribution of the pile cap.
5. Superimpose the stiffness of the pile cap to the pile group.

These steps are described herein.

Step 1: Single Pile Under Lateral Load Lam and Martin (1986) came to a realization that lateral load-deflection characteristics representing the overall stiffness of the soil pile system are dominated by the elastic pile stiffness over the nonlinear soil behavior. Moreover, the localized zone of influence is limited to the upper 5–10 pile diameters. Hence, they concluded that linear solutions may be adequate for pile stiffness evaluations. They developed single-layer, pile-head

stiffness design charts for lateral loading as presented in Figs. 11, 12, and 13. These charts are applicable for piles up to 24 in. Two parameters are required to define the soil-pile system: the pile bending stiffness, EI , and the coefficient of variation f of soil reaction modulus E_s with depth. The coefficient f has units of force/unit volume.

Step 2: Single Pile Under Axial Load The recommended LRFD guidelines for the seismic design of highway bridges proposed the following simple equation for the determination of the axial stiffness of the pile:

$$K_v = 1.25 \frac{E_p A}{L} \quad (5)$$

Wherein, E_p is the modulus of elasticity of pile material; A is cross sectional area of pile; and L is length of pile.

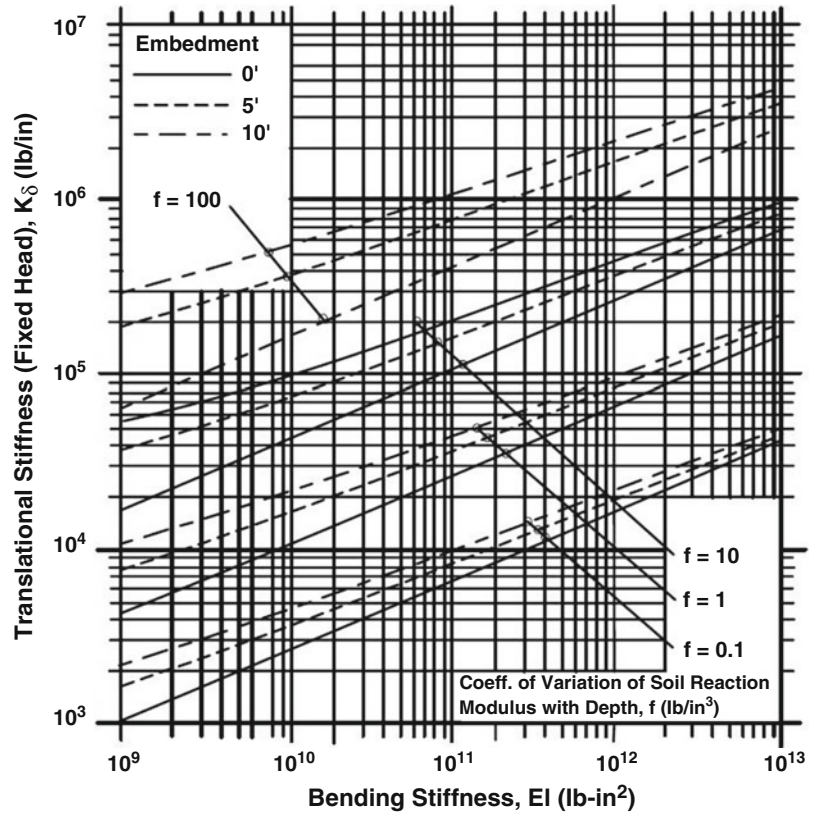
Step 3: Pile Group Stiffness The stiffness of the single pile can be used to establish the pile group stiffness matrix. If the pile group consists of vertical piles, the stiffness summation procedure is relatively straight forward. The stiffness for the translational displacement terms (the two horizontal and the vertical displacements) and the cross-coupling terms can be obtained by multiplying the corresponding stiffness components of an individual pile by the number of piles.

In general, the axial stiffness of the piles will dominate the rotational stiffness of the group. Therefore, the rotational stiffness terms require consideration of this additional stiffness component. Similarly, there is an interaction between the translational stiffness of piles and the torsional stiffness of the group. The following equation (Lam et al. 1991) can be used to develop the rotational terms of a pile group:

$$K_{RG} = NK_{RP} + \sum_{n=1}^N K_{\delta n} S_n^2 \quad (6)$$

In which, K_{RG} and K_{RP} are the rotational stiffness of the pile group and an individual pile respectively; N is the number of piles in the pile group;

Bridge Foundations,
Fig. 11 Pile translational
stiffness



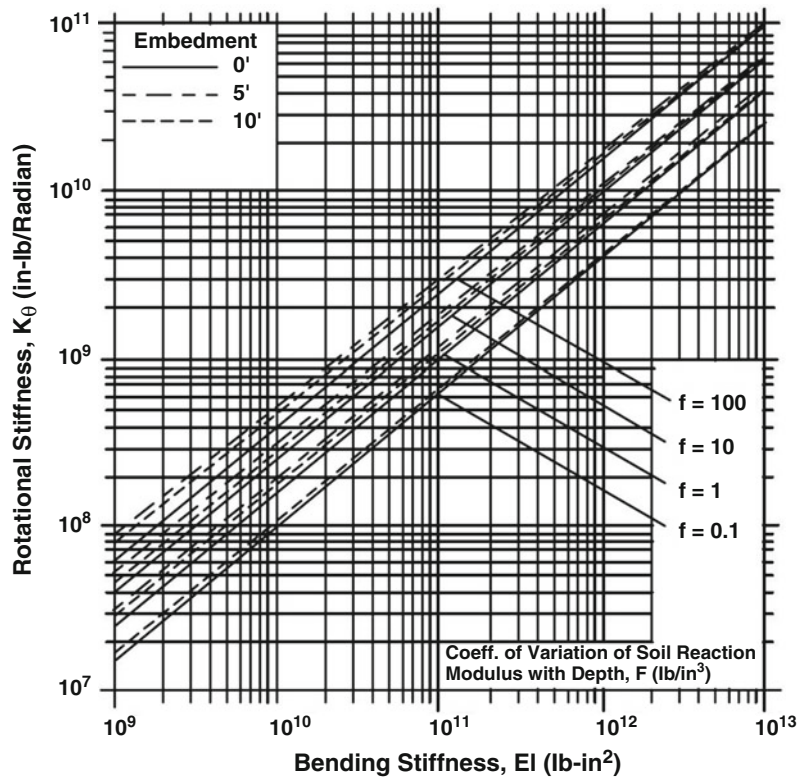
$K_{\delta n}$ is the translational stiffness coefficient of an individual pile (axial for group rocking stiffness and lateral for group torsional stiffness); and S_n is the distance between the n th pile and the axis of rotation.

Step 4: Stiffness Contribution of the Pile Cap In addition to the component of soil resistance acting on piles, the passive pressure soil resistance on the vertical pile cap face can be added to the stiffness and resistance obtained from the pile members. This is contingent on stable level ground conditions. The pile cap stiffness can be estimated as the ultimate soil capacity divided by an estimated displacement to mobilize this capacity. Centrifuge tests (Gadre 1997) showed that the deflection level to reach the ultimate pile cap capacity occurs at about 0.02 times the embedment depth. This equivalent linear secant stiffness can be added to the stiffness of the piles.

Step 5: Superimpose the Stiffness of the Pile Cap to the Pile Group The resultant pile cap stiffness obtained from step 4 can be added to the diagonal lateral translational stiffness coefficients in the pile group stiffness matrix for the total pile group–pile cap stiffness matrix.

The above procedure does not account for group effects which relate to the influence of the adjacent piles in affecting the soil support characteristics. Full-scale tests by a number of investigators demonstrate that the lateral capacity of a pile in a pile group may be less than that of a single pile due to the interaction between closely spaced piles in the group (group efficiency). As the pile spacing reduced, the reduction in lateral capacity becomes more pronounced. In general, pile spacing of less than 3–5 pile diameters are necessary before the effects of pile interaction becomes significant in practical terms. Type and strength of soil, number of piles, and loading level are other factors that may affect the

Bridge Foundations,
Fig. 12 Pile rotational
 stiffness



efficiency and lateral stiffness of the pile. Moreover, in addition to group effect, gapping, and potential cyclic degradation were also subject of many investigations (e.g., Brown et al. 1987; McVay et al. 1995). It has been shown that a concept based on p -multiplier applied on the standard stiffness matrix can work reasonably to account for pile group and cyclic degradation effects.

Seismic Design of Pile Footings

Design of pile footings for seismic effects is often relatively a complex process. The strategy of the seismic design is to determine capacity and deflection of the piles under the action of the seismic lateral loads and ensure the integrity of the pile group against liquefaction.

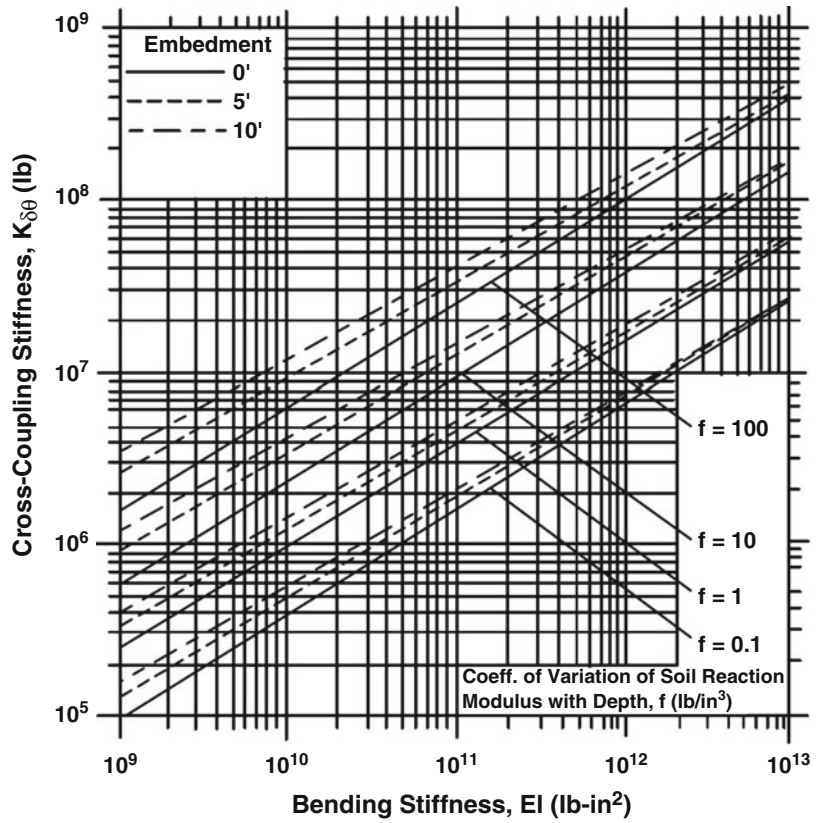
Moment Capacity of Pile Groups

Typically, the foundation system is designed to be capacity protected, which ensures that damage occurs above ground, where it can be accessed and repaired. The seismic moment capacity of the

pile group will govern the design. For a fixed-head pile group, the piles must be designed for the plastic hinging moment capacity of the column/pier connected to the pile cap. The maximum negative moment occurs at the base of the pile cap while the maximum positive moments occur in the pile at a short depth below the base. In this regard, the pile cap has to be designed to behave rigidly. The rigid response of the footing may be assumed if the following equation is satisfied (AASHTO 2011):

$$\frac{L}{D} \leq 2.5 \quad (7)$$

where, L is the cantilever overhang length measured from the face of pier or column to the outside edge of the pile cap and D is the depth of the pile cap. Due to the interaction between the rocking response and the vertical response of a pile group, the axial demands on an individual pile can be estimated as:

Bridge Foundations,**Fig. 13** Pile cross-coupling stiffness

$$F = \frac{P}{N} \pm \frac{M_Y^{\text{col}} C_{X_i}}{I_{pgy}} \pm \frac{M_X^{\text{col}} C_{Y_i}}{I_{pgx}} \quad (8)$$

in which, P is the total axial load due to dead load, footing weight, and seismic load; N is the total number of piles in a group; C_{X_i} and C_{Y_i} are the distances from neutral axis of pile group to i th row of piles measured parallel to the x and y axes respectively; M_X^{col} and M_Y^{col} are the plastic moment capacities of the pier or column connected to the footing about the x and y axes respectively; and I_{pgx} and I_{pgy} are the effective moment of inertia of pile group about the x and y axes evaluated as:

$$I_{pgx} = \sum_{i=1}^{N_x} n_x c_{yi}^2 \quad \text{and} \quad I_{pgy} = \sum_{i=1}^{N_y} n_y c_{xi}^2 \quad (9)$$

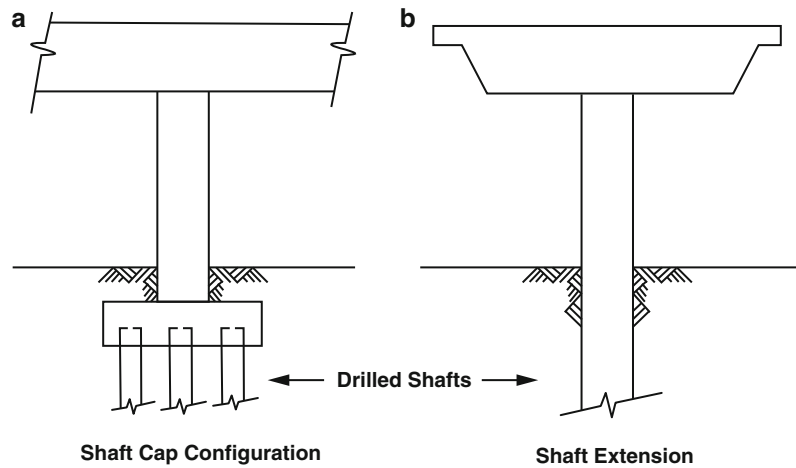
where n_x and n_y are the number of piles in a single row parallel to the x and y axes.

For a foundation protected design strategy, the connection of pile to cap should be designed such that its moment capacity exceeds that of the pile.

Pile-to-Cap Connection

The moment capacity at the connection depends on both the depth of embedment of the pile into the cap and the reinforcement arrangement. Driven precast piles should be constructed with considerable spiral confining steel to ensure good shear strength and tolerance of yield curvatures that may be exerted during the seismic event. Parametric studies (Harris and Petrou 2001) with prestressed piles suggest that the embedment length should be taken as the larger of the pile diameter or 12 in.. Experimental studies (Shama and Mander 2004) have showed that an embedment of one pile diameter of a timber pile into a pile cap was sufficient to develop the full moment capacity of the pile and provide a ductile response. Experimental studies (Rollins and Stenlund 2011) also showed that steel pipe piles

Bridge Foundations,
Fig. 14 Drilled shafts
 configuration



embedded 24 in. into a pile cap can produce a sufficient moment capacity for the connection without the need for a reinforcement detail. These observations agree with the results of cyclic tests on steel HP piles (Shama et al. 2002).

Lateral Capacity of Pile Footings

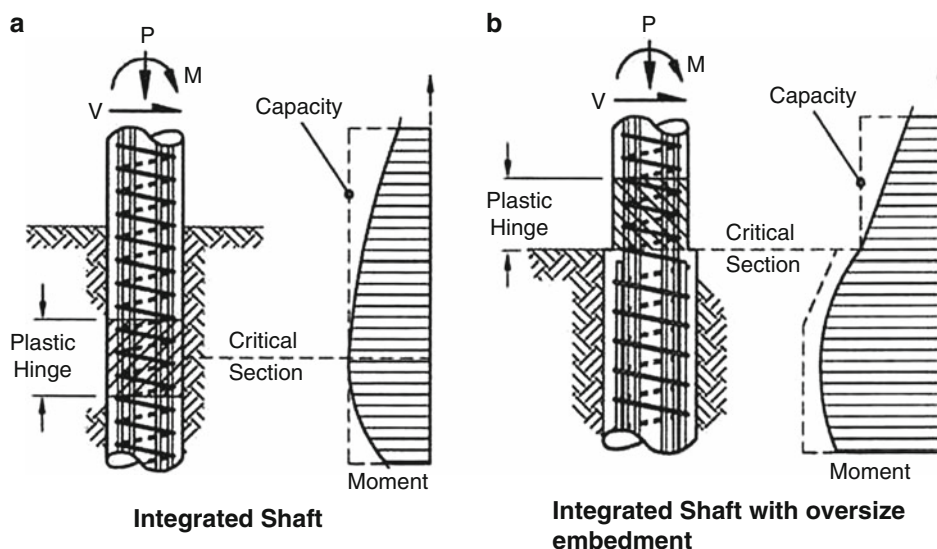
The lateral capacity of piles requires compiling of both geotechnical and structural engineering principles. For a successful design, it is important to get the soil resistance mobilized before structural failure of the pile. The relative importance of the pile cap lateral capacity to the individual piles needs to be considered. Therefore, the lateral capacity evaluation must include the resistance developed by the pile cap in addition to the lateral shear resistance of the piles. Including the lateral capacity of the pile cap is contingent upon efficient compaction of the soils around the pile cap to improve the cap resistance. The lateral capacity of the pile cap must include the passive pressure mobilized at the front face of the cap perpendicular to the seismic load and the interface shear resistance developed along the sides of the cap parallel to the application of the seismic load.

Drilled Shafts

Drilled shafts are large diameter piles that are designed to sustain high axial loads and overturning moments. They are constructed

using drilling (excavating) equipment capable of auguring or coring 30" to 120" diameter excavations into soil and rock. After the excavation is completed, a reinforcing cage is placed in the excavation and the excavation is filled with high slump concrete. Drilled shafts are usually constructed with steel shells.

Drilled shafts are attached to the superstructure in two configurations as illustrated in Fig. 14. In the first configuration a group of drilled shafts, usually 4–6 are connected to a cap that supports the bridge pier or column. In the second configuration, a large diameter shaft is extended as a structural unit to support the bridge superstructure. The second configuration comes in two alternatives as illustrated in Fig. 15 (Priestley et al. 1996). In the first alternative, the section of the column above the ground is similar to that of the shaft. The maximum moment forms at a depth of typically 1.5–2.5 shaft diameter. A disadvantage of this alternative is that the extent of damage to the plastic hinge region will be below ground after an earthquake and needs excavation of the soil material for inspection. With the alternative detail of Fig. 15b, the moment capacity of the shaft below ground is increased above that of the column above ground to ensure that hinging occurs at the base of the column and damage can be expected after an earthquake. The disadvantage of this detail is that it is more expensive with respect to the alternative of Fig. 15a.



Bridge Foundations, Fig. 15 Extended shaft seismic design alternatives (Priestley et al. 1996)

Stiffness of Drilled Shaft Foundations

The most common methods for modeling drilled shaft behavior in a global bridge model for seismic response analysis are categorized to three classes:

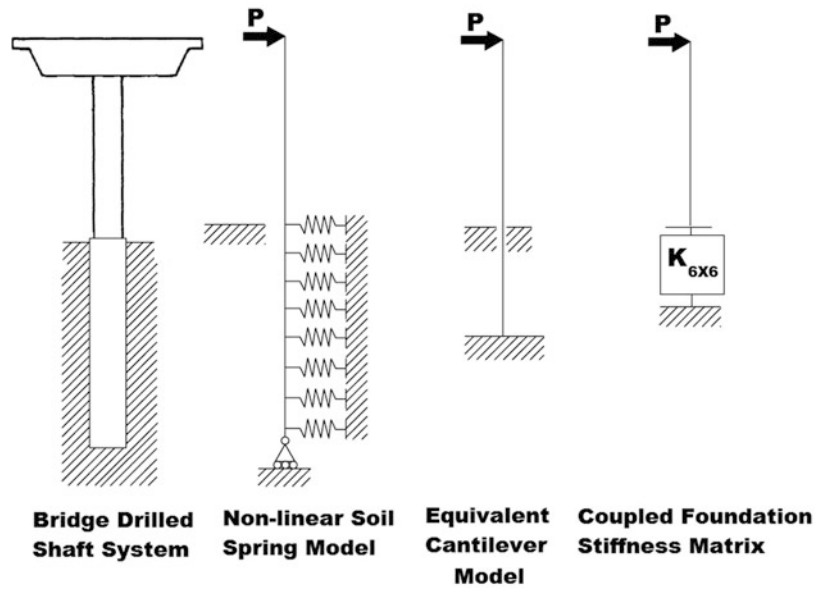
- Beam supported on nonlinear Winkler springs
- Equivalent cantilever model
- Coupled foundation stiffness matrix

Beam Supported on Nonlinear Winkler Model

The beam supported on nonlinear Winkler springs method can be implemented in the structural model without any difficulty as the number of large diameter drilled shafts for a typical bridge is limited. This model, as illustrated in Fig. 16, is based on representing the soil by a series of nonlinear springs. It provides a complete representation of the soil–structure interaction. Furthermore, nonlinear behavior as plastic hinging of the shaft cross-section can be included as well as soil yielding. This approach is usually required for bridges located in highly active seismic regions. Two analytical methods are available to establish the nonlinear springs. These are the p - y method and the strain wedge method. The p - y model, as described in section “[Pile](#)

[Foundation Stiffness](#),” simply describes the force displacement relationship of a soil spring, in which p is the soil reaction per unit length at a certain location along the embedment length of the pile into soil and y is the corresponding deflection of the pile. A major shortage of this method is that the p - y curves are semiempirical relationships and were developed based on a limited number of full-scale lateral load tests on piles of diameters ranging from 12 to 24 in.. Hence, they might underestimate the soil stiffness for larger diameter drilled shafts. Also, the p - y method uses a P -multiplier to account for group effects. Selection of values of this parameter is still controversial as treated by different codes and specifications. The strain wedge method (Ashour and Norris 1998), on the other hand, fills all these gaps. The nonlinear springs, obtained using this method, are based on modeling the three-dimensional interaction between the surrounding soil and the shaft. The method considers the alternative effects of the pile properties (stiffness and cross-section) and soil properties in developing the p - y curve. Furthermore, there is no need to use P -multipliers to the strain wedge method because this method accounts for the group effects through an assessment of the overlap of the passive wedges that develop due to

Bridge Foundations,
Fig. 16 Methods of
representing shaft
foundation stiffness



pushing the shaft in the lateral direction. For a group, the strain wedges for different shafts will interfere together, depending on the spacing, diameter, and location of the shaft within the group leading to strength reduction and softening.

The nonlinear spring modeling technique can include the Masing hysteretic behavior of the soil, which when implemented includes automatically in the model the energy dissipation associated with material damping. The radiation damping component can be included by attaching additional viscous dashpots. Recent studies (Wang et al. 1998) showed that placing the viscous dashpots (representing radiation damping in the far field) in series with the hysteretic component of the soil-structure element (representing the nonlinear soil-pile response in the near field) is technically preferable to a parallel arrangement of the viscous and hysteretic damping components.

Equivalent Cantilever Model

The equivalent cantilever model is the simplest approach to represent the effects of the surrounding soil, wherein the sectional properties of the cantilever are the same as that of the shaft, but its length (depth to fixity) is adjusted to provide the same maximum bending moment as in the actual

soil-pile system. The effective length of the shaft shall be equal to the laterally unsupported length L_C plus an embedded depth to fixity L_F . The depth to fixity below the ground in ft. may be taken as (AASHTO-LRFD 2012):

For clays:

$$1.4 \left[\frac{E_p I_w}{E_s} \right]^{0.25} \quad (10)$$

and for sands:

$$1.8 \left[\frac{E_p I_w}{n_h} \right]^{0.20} \quad (11)$$

Where E_p is the modulus of elasticity of the shaft material; I_w is the weak axis moment of inertia for the shaft; E_s is the soil modulus for clays; S_u is the undrained shear strength of clay; and n_h is the rate of increase of soil modulus with depth for sand.

Coupled Foundation Stiffness Matrix

This approach is usually employed in regions of low to moderate seismicity. As illustrated in Fig. 16, a 6×6 can be determined for a single shaft at the ground line. As indicated in section “[Coupled Pile Foundation Stiffness Matrix](#),”

linearization of p-y curves is required for the development of a 6×6 coupled stiffness matrix.

Seismic Design of Drilled Shafts

The following design requirements shall apply (ATC-32 1996) to bridges on drilled shaft foundations:

- (I) Seismic lateral foundation design forces shall be based on either plastic hinging of the bridge column or pier; linear dynamic analysis; or more advanced nonlinear dynamic response analysis.
- (II) Foundation stiffness must be accounted for in the dynamic response analysis of the overall bridge.
- (III) The capacity of shaft foundations and their individual components shall be based on the safety evaluation earthquake.
- (IV) Earth pressures generated by lateral ground displacements due to liquefaction shall be accounted for at poor soil sites.
- (V) Strong connection details shall be evaluated at poor soil sites.

Large Gravity Caissons

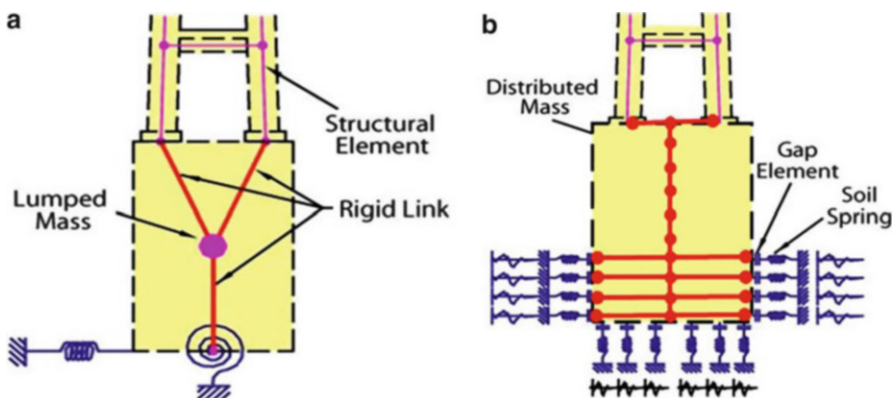
Large caissons are very large concrete boxes that are excavated or sunk to a predetermined depth. They are usually used for the construction of bridge piers or other heavy waterfront structures,

and they often become advantageous where water depths exceed 30–36 ft. Caissons are divided into three major types: (1) open caissons, (2) box caissons (or closed caissons), and (3) pneumatic caissons. A common feature of large caissons produced by the three methods is that they are massive structures that respond to seismic loads in a primarily rocking mode about the base plus some translations.

Seismic Evaluation of Large Caissons

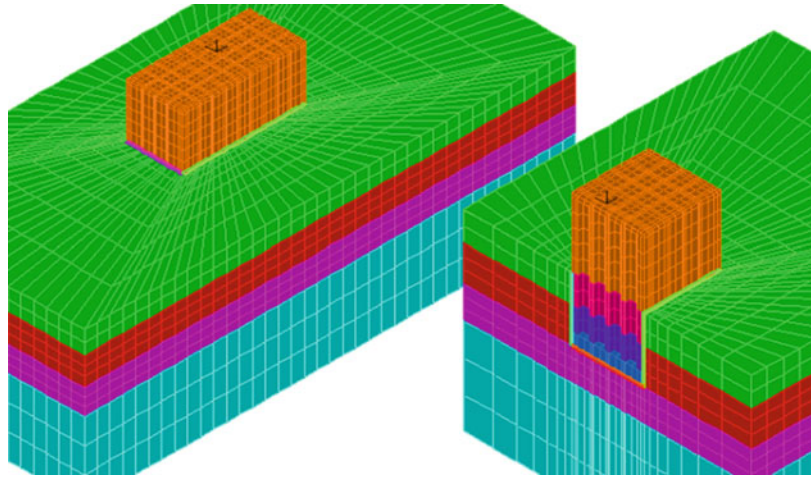
The behavior of caissons under lateral seismic loads is essentially nonlinear. Geometric nonlinearity dominates this behavior due to rocking of the caisson and gapping at the soil-caisson interface. The weight of caissons constitutes the major portion of the entire structure due to its large volume. Hence, the fundamental natural period of vibration of the large caisson is short and very different from the long period response of the bridge. This difference in the two modes of vibrations enables uncoupled seismic analyses of the two structures. However, when evaluating the bridge structural response, the caisson as an element must be included in the total bridge model. Two approaches are usually used for representing large caissons for seismic design.

The uncoupled spring approach represents the caisson as a lumped mass and establishes the interaction between the caisson and soil by the impedance approach. Figure 17a displays a



Bridge Foundations, Fig. 17 Modeling techniques of caisson foundations: (a) Uncoupled spring approach; and (b) Winkler approach

Bridge Foundations,
Fig. 18 Continuum
 models of caissons



schematic of the caisson model using this approach, wherein a rigid link is established between the node located at the center of mass of the caisson where its mass is lumped and another node at the bottom where a set of springs is attached. This method has been presented in section “[Spread Footing Stiffness](#).” This method overlooks the propensity of gapping at the base and at the embedded sides of the caisson during the seismic event and may lead to irrelevant results in the large displacement range of loading. Hence, it is only recommended for regions of low seismicity.

The nonlinear spring Winkler approach (Fig. 17b), as mentioned earlier, treats the soil supporting the foundation as a series of springs. It represents linear and nonlinear soil properties and hence favorable for life safety performance based design. It also accounts for the separation and gapping between the caisson walls and the surrounding soils. The properties of these nonlinear springs are established through static pushover analyses of three dimensional finite element continuum models as. Figure 18 depicts an example of such continuum model, where both the caisson and surrounding soil are modeled by three dimensional brick elements. The detailed 3-D finite element shall include constitutive relationships for the nonlinear behavior of the soil and special interface elements that can capture gapping

between the soil and the caisson at the base and side walls. Nonlinear static pushover analyses shall be carried out to establish the soil–structure interaction behaviors for implementation in the global model. Pushover analyses of the local model shall be performed by applying a point load at the center of gravity of the rigid caisson for each mode of soil resistance. The soil response results from the different pushover analyses of the local model in each direction shall be extracted and distributed to the soil spring elements at various nodal points in accordance with the discretization scheme of the global model. The behavior of the caissons under seismic loads shall be assessed through nonlinear time history analysis of the global model. The performance of the caissons is evaluated by comparing the maximum drifts from the results of the time history analysis to the permissible levels according to the performance based design criteria of the project.

Cross-References

- [Dynamic Soil Properties: In Situ Characterization Using Penetration Tests](#)
- [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)
- [Soil-Structure Interaction](#)

References

- American Association of State Highway and Transportation Officials (2011) AASHTO LRFD seismic guidelines. Washington, DC
- American Association of State Highway and Transportation Officials (2012) AASHTO LRFD bridge design specifications. Washington, DC
- Applied Technology Council (1995) Seismic evaluation and retrofit of concrete buildings volume 1, ATC-40. Applied Technology Council, Redwood City
- Applied Technology Council (1996) Improved seismic design criteria for California bridges: provisional recommendations ATC-32. Applied Technology Council, Redwood City
- Ashour M, Norris G (1998) Lateral loading of a pile in layered soil using the strain wedge model. *J Geotech Eng Div ASCE* 124(4):303–315
- Ashour M, Norris G (2000) Modeling lateral soil-pile response based on soil-pile interaction. *J Geotech Eng Div ASCE* 126(5):420–428
- Brown D, Reese L, O’Niell M (1987) Cyclic lateral loading of a large-scale pile. *J Geotech Eng Div ASCE* 113(11):1326–1343
- Gadre A (1997) Lateral response of pile-cap foundation systems and seat-type bridge abutments in dry sand. PhD dissertation, Rensselaer Polytechnic Institute
- Gazetas G (1991) Formulas and charts for impedances of surface and embedded foundations. *J Geotech Eng Div ASCE* 117(9):1363–1381
- Harris KA, Petrou MF (2001) Behavior of precast prestressed concrete pile to cast-in-place pile cap connections. *PCI Journal*, 46(4)
- Hoit MI, McVay MC (2010) FLPiER user’s manual. University of Florida, Gainesville
- Lam IP, Martin GR (1986) Seismic design of highway bridge foundations. Report no. FHWA/RD-86-102, U.S. Department of Transportation, Federal Highway Administration, McLean, 167 p
- Lam IP, Martin GR, Imbsen R (1991) Modeling bridge foundations for seismic design and retrofitting. *Transp Res Rec* 1290 p.113–126
- Lam IP, Kapuskar M, Chaudhuri D (1998) Modeling of pile footings and drilled shafts for seismic design. Technical report MCEER-98-0018, Multidisciplinary Center for Earthquake Engineering Research. Buffalo
- LPILE (2010) Program LPILE plus, version 6.0. Ensoft, Austin
- Matlock H (1970) Correlations for design of laterally loaded piles in soft clay. In: 2nd offshore technology conference, vol 1. Houston, pp 579–594
- McVay MC, Casper R, Shang T (1995) Lateral response of three-row groups in loose to dense sands at 3D and 5D pile spacing. *Journal of Geotechnical Engineering, ASCE*, 121(5)
- McVay MC, O’Brien M, Townsend FC, Bloomquist DG, Caliendo JA (1998) Numerical analysis of vertically loaded pile groups. In: ASCE foundation engineering congress, Northwestern University, Evanston, pp 675–690
- Meyerhof GG (1953) The bearing capacity of foundations under eccentric loads. In: Proceedings 3rd international conference on soil mechanics and foundation engineering, vol 1. pp 440–445
- Priestley MJN, Seible F, Calvi GM (1996) Seismic design and retrofit of bridges. Wiley, New York
- Reese LC, Cox WR, Koop FD (1974) Analysis of laterally loaded piles in sand. In: Proceedings 6th Offshore Technology Conference, Paper 2080, Houston, pp 473–483
- Rollins K, Stenlund T (2011) Experimental testing of pile-to-cap connections for embedded pipe piles. *J Bridge Eng* 16(2):286–294
- Shama AA, Mander JB, Aref AA (2002) Seismic performance and retrofit of steel pile to concrete cap connections. *ACI Struct J* 99(1):185–192
- Shama AA, Mander JB (2004) Behavior of timber pile-to-cap connection under cyclic lateral loading. *J Strut Eng ASCE* 130(8):1252–1262
- Shama AA, Mander JB, Chen SS (2002) Seismic investigation of steel pile bents: II. Retrofit and vulnerability analysis. *Earthq Spectra EERI* 18(1):143–160
- Veletsos AS, Verbic B (1973) Vibration of viscoelastic foundations. *J Earthq Eng Struct Dyn* 2(1):87–102
- Vesic AS (1973) Analysis of ultimate loads of shallow foundations, chapter 3. In: Fang H-Y (ed) Foundation engineering handbook, 1st edn. Van Nostrand Reinhold, New York, pp 553–593
- Wang S, Kutter BL, Chacko JM, Wilson DW, Boulanger RW, Abghari A (1998) Nonlinear seismic soil-pile-interaction. *Earthq Spectra EERI*, Oakland 14(2) pp. 377–396
- Winkler E (1867) Die lehre von elasticizitat und festigkeit (on elasticity and fixity). Verlag von H. Dominicus Prague, 182 p

Buckling-Restrained Braces and Their Implementation in Structural Design of Steel Buildings

Justin D. Marshall
Auburn University, Auburn, AL, USA

Synonyms

Ductile design of steel frames; Hysteretic energy dissipation; Passive energy dissipation; Steel braced frames

Introduction

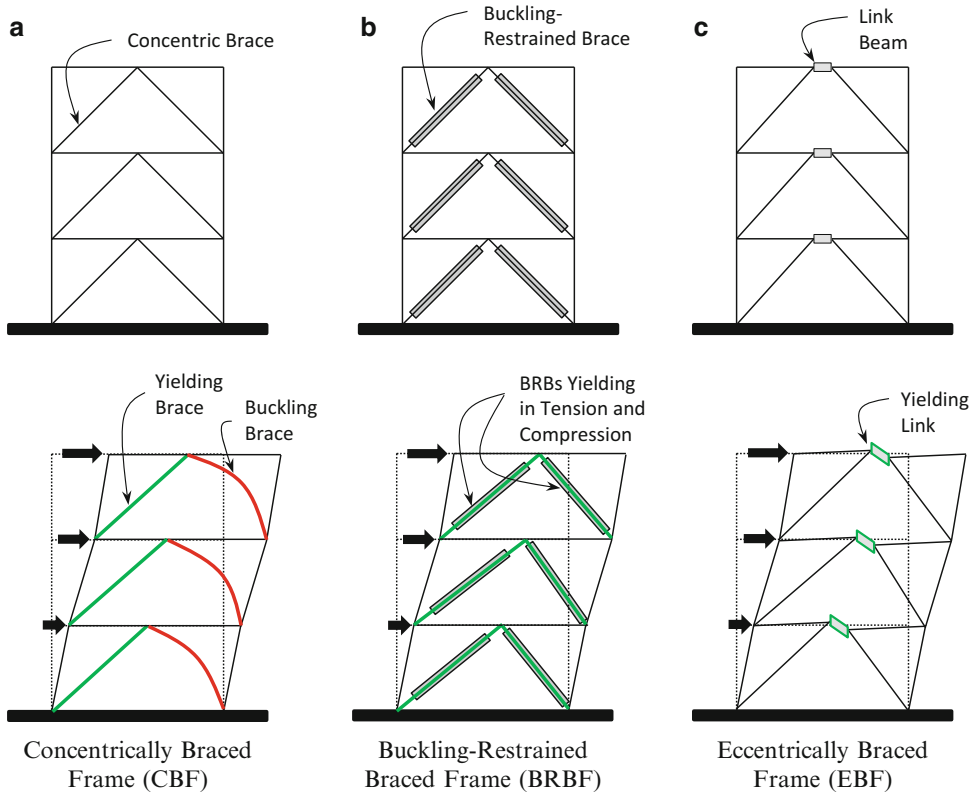
For a structural engineer, selection of the lateral load-resisting system for a building is a critical decision. While it is easy to understand the requirement to select beams and columns to support the gravity loads (i.e., self-weight of the building itself and the live loads related to building occupancy), a lateral system is required for any building to resist horizontal loading. Lateral loads routinely considered in design are wind and earthquake, although blast, flood, and tsunami are also possible. The lateral system can be especially critical when the building is to be constructed in a region where significant seismic activity is expected. In the case of a steel building, a wide variety of lateral resisting systems are available. These tested and approved lateral systems have the capability to maintain structural integrity in the face of significant lateral loads and inelastic behavior. One of the key elements to steel structures resisting seismic loads is a properly designed energy-dissipating fuse element. This critical element yields and subsequently dissipates energy during an earthquake without causing instability or collapse of the structure. In moment-resisting frames, the beams and columns resist both gravity and lateral loads although typically only the perimeter frames resist lateral forces. In this case, the beam elements are the fuse where inelastic behavior is expected. The challenge with these systems is that repairing or replacing beams after an earthquake is extremely difficult and costly. The other common lateral resisting system for steel buildings is a braced frame. The three most common braced frame systems are concentrically braced frames (CBF), eccentrically braced frames (EBF), and buckling-restrained braced frames (BRBF). Representative schematics for all three steel brace systems are shown in Fig. 1. Buckling-restrained braces (BRB) are a somewhat recently developed structural device that has a balanced force-deformation behavior (i.e., nearly equal strength in tension and compression) and are one of the few manufactured, rather than fabricated, products in common use in the structural steel construction industry. This entry will briefly

explore the three common types of braced frames and then discuss in greater depth the mechanics, performance, and design of BRBs and the BRBF system.

Steel Braced Frame Lateral Resisting Systems

The basic premise of braced frame systems is that diagonal braces are placed in certain bays creating a vertical truss system to resist lateral loads. This results in primarily axial forces and axial deformations to control response of the lateral force-resisting system. The most common type of braced frame is the concentrically braced frame (CBF) (Fig. 1a). For CBFs, the centerline of the braces, beams, and columns is concentric at beam-column-brace joints and beam-brace intersections so that axial forces are induced in the braces, beams, and columns and significant moments are not developed in the members. The braces are the fuse elements, and the stiffness and strength of the system are controlled by the geometric and material properties of the braces. There are several drawbacks to this system, the first being the asymmetry of the system due to the braces yielding in tension and buckling in compression. This unbalance results in a significant difference in brace forces as well as degradation in both strength and stiffness during an earthquake. In addition, the potential exists for brace fracture due to multiple cycles of compression buckling and tension yielding. The benefit of this system is that typically the system is stiff resulting in small interstory drifts. CBFs have been extensively studied and tested experimentally such that if proper detailing methods are followed, the structure will be safe and the braces can be replaced following a significant earthquake.

A second type of braced frame is the eccentrically braced frame (EBF) (Fig. 1c). For these systems, the braces are intentionally eccentric relative to the element joints resulting in a combination of axial, shear, and moment in certain structural elements. The fuse element in this case is typically a link beam. The case shown in

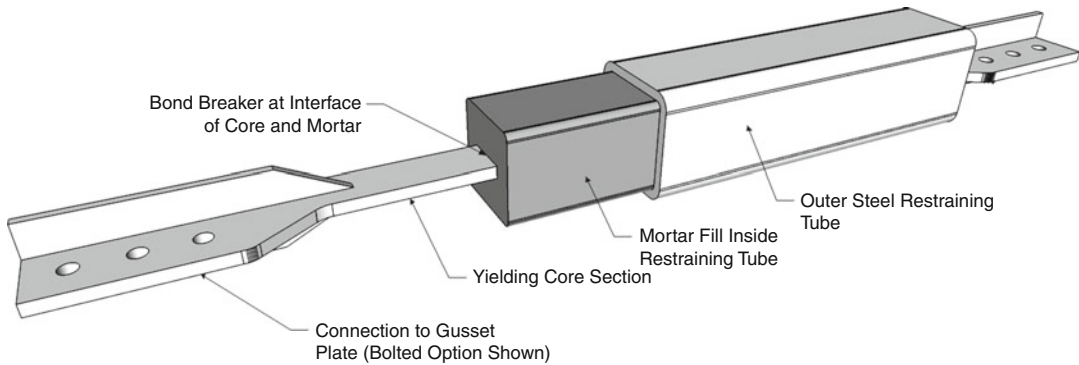


Buckling-Restrained Braces and Their Implementation in Structural Design of Steel Buildings, Fig. 1 Types of steel braced frames

Fig. 1c shows a link in the middle of the beam span although it is also commonly adjacent to the column. The diagonal braces used in EBFs are designed to remain elastic for the maximum forces and moments that can develop in the link beam. An EBF system is more flexible than a CBF and has a symmetric response as the yielding under shear and flexure is symmetric. The primary drawback for EBFs is that replacing the link beam is more difficult and costly than replacing a diagonal brace. There has been recent work in developing and testing replaceable link beams to overcome this problem (Mansour et al. 2011).

The third type of braced frame system is the buckling-restrained braced frame, which is a type of concentrically braced system. The BRB is the primary structural element in the BRBF providing the necessary ductile energy dissipation. The difference between a typical fabricated steel

brace and the BRB is the symmetry of the response. A BRB is specially designed to prevent buckling during the compression cycle. This results in highly ductile behavior in both tension and compression and prevents the force unbalance present in other brace configurations due to the significantly lower buckling strength as compared to the tension yield (see Fig. 1b). Brace fracture is not a problem due to the fact that the large stresses in the core are primarily axial. There is no strength or stiffness degradation for BRBs that are detailed correctly. The stiffness of a BRBF is less than a CBF as the BRB strength is directly related to the tension strength, whereas in a CBF the area generally must be increased to get the required capacity in compression. Buckling-restrained braces are unique in that they are a manufactured device rather than a fabricated steel section. Since their development and introduction into the industry, BRBs have taken a



Buckling-Restrained Braces and Their Implementation in Structural Design of Steel Buildings, Fig. 2 Isometric cutaway of buckling-restrained brace

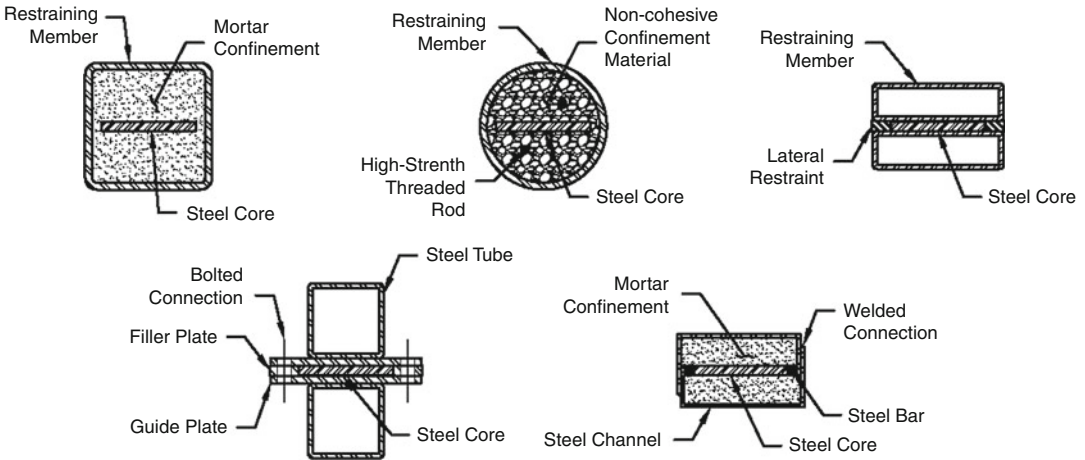
large market share in braced steel frame systems due to their cost competitiveness and their reliable, resilient performance in seismic events.

Buckling-Restrained Brace Description

Because braced frames have very beneficial characteristics, significant work was undertaken to determine ways to overcome the negative aspects of brace buckling. The early beginnings of a brace restrained against buckling were in the 1970s in Japan where steel plates were encased in a precast concrete wall (Wakabayashi et al. 1973). Since then, significant improvements and modifications have been made to the original configuration that has enabled the reliable and ductile BRB commonly used today (Black et al. 2004; Fahnestock et al. 2006; Xie 2005). The key to the mechanism is the axial load going through the steel core while sufficient buckling resistance (flexural rigidity) is provided in another component. This has been accomplished in several ways although the most common is shown in Fig. 2. The basic premise of the BRB is that the core steel element has a cross section designed to yield in tension and compression. In order for this to occur, sufficient flexural rigidity must be provided so the overall buckling load of the composite device is greater than the ultimate strength of the steel core. This is accomplished through the use of a steel tube which is most commonly square or circular. The core steel

runs through the center of the tube which is then filled with a mortar paste to fill the space allowing the steel tube to provide the required flexural resistance to brace the core. It is not desirable for any axial load to go through the mortar or the steel outer tube. This is accomplished by providing a bond breaker between the mortar and the steel core. Several different options have been used for this including a lubricant or a material wrapped around the steel core prior to filling the outer tube. In addition to the BRB shown in Fig. 2, several other systems have been developed by researchers to create a BRB system for lower cost, to simplify manufacturing, or to replace damaged cores more simply after an event. Several of the BRB systems developed through these research efforts are shown in Fig. 3, although this is a sample of the different variations that have been investigated. Figure 4 shows a photograph of a building retrofitted with BRBs.

The geometry and yield strength of the core section play an important role in the performance of BRBs. From a material specification, it is desirable to have a known value or at least a small range where yielding will occur. An A36 material with controlled yield is commonly used for this purpose. The yielding core section can be fabricated to any area specified by the structural engineer. The area of the yielding core is the prime cross-sectional characteristic used to determine strength. There are two primary shapes for the core region, flat plate or cruciform.



Buckling-Restrained Braces and Their Implementation in Structural Design of Steel Buildings, Fig. 3 Various BRB cross sections

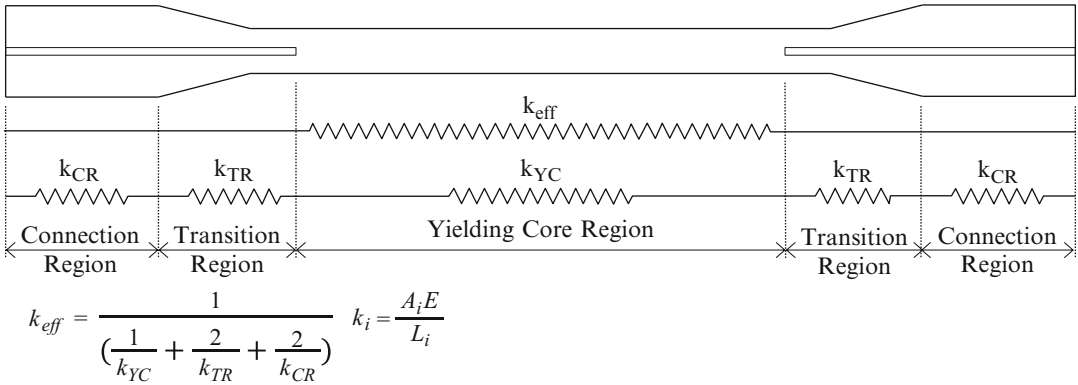
Buckling-Restrained Braces and Their Implementation in Structural Design of Steel Buildings,

Fig. 4 Photograph of building retrofitted with BRBs (Courtesy of Core Brace, LLC)



In addition, some BRBs use multiple cores within the restraining tube to achieve the desired core area. The length of the yielding core is an important part in determination of the stiffness of the brace. Figure 5 shows the breakdown of a typical core with the flat plate yielding core, transition region, and connection region. The flexural stiffness of the transition region and connection region is important as the flexural rigidity of these regions must be sufficient to sustain the

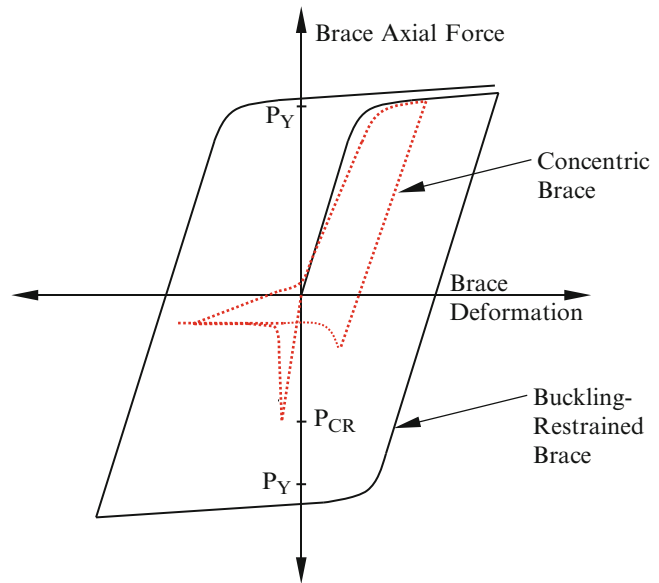
core axial strength. The equation for the equivalent elastic stiffness of the brace is shown in Fig. 4 and is based on a combination of springs in series representing the axial stiffness of each segment. The overall stiffness of the brace is important for accurate analysis of the frame with BRBs and can be modified if the brace stiffness is too low. This modification is typically done by shortening the yielding core length (i.e., increasing the transition length) which is the most flexible segment.



Buckling-Restrained Braces and Their Implementation in Structural Design of Steel Buildings,
Fig. 5 Stiffness calculation and different regions of BRB steel core

**Buckling-Restrained
 Braces and Their
 Implementation in
 Structural Design of
 Steel Buildings,**

Fig. 6 Hysteresis plot for
 buckling-restrained brace
 and concentric brace



However, it is important to recognize that a shorter yielding length results in higher inelastic strains in the core which will result in higher forces in the braces, an important design criterion.

The overall representative hysteretic behavior of a BRB and a concentric brace is shown in Fig. 6. The BRB has full hysteresis loops in both tension and compression including strain hardening. The capacity for significant deformation ductility in both directions is apparent where deformation ductility is defined as the total

deformation divided by the yield deformation. There is a small increase in strength on the compression side, typically about 10 % larger than the brace tension strength. This is due to friction between the core and the mortar. Even with a bond breaker, the core will dilate under compression resulting in greater frictional force and some axial load transfer. By contrast, a concentric brace performs well in tension but experiences buckling in compression which typically results in formation of a flexural plastic hinge near the

center of the brace. Once the brace has buckled one time, it quickly loses strength in compression upon subsequent cycles. The formation of the hinge can result in brace fracture in tension due to the damage occurring during compression buckling. The lower compression capacity and potential brace fracture significantly limit the energy dissipation and deformation ductility capacity of CBFs when compared to BRBs.

As was stated previously, BRBs are a manufactured product. In the United States there are three primary brace producers, Core Brace (<http://www.corebrace.com/>), Star Seismic (<http://www.starseismic.net/>), and Unbonded Brace (<http://www.unbondedbrace.com/>). Each of these companies has tested their braces to meet the requirements put forth in the applicable building codes and reference documents. In addition to the braces, they provide design and engineering guidance associated with their products. While the basic premise of the BRB is the same for each manufacturer, many specific details of the various brace manufacturers differ as they compete based on price and performance in the marketplace.

BRBs, as manufactured products, have a high degree of quality control requirement placed on them in the American Institute of Steel Construction (AISC) *Seismic Provisions for Structural Steel Buildings* Section K3 (AISC 2010), which is a reference standard in the United States (USA) governing the design of steel buildings for seismic loads. The requirements include testing which includes both axial demands on the brace and the consideration of rotations at the connections due to frame deformations. The primary requirements of the testing are to show (1) the required ductility, both for a single cycle and cumulatively; (2) repeatable stable hysteresis with positive post-yield slope; and (3) avoidance of core rupture, brace instability, and connection failure. The required testing for a wide variety of brace sizes has been completed by the three primary producers. The specification however allows for these tests to be run on a project-specific basis if a new or unique brace configuration is specified.

Buckling-Restrained Brace Design

Buckling-restrained braces were first used in the USA in 1999, over a decade after the first usage in Japan in 1987 (UBB 2014). Part of the motivation for this new system was due to the failure of steel moment frames that occurred in the 1994 Northridge, California, and the 1995 Kobe, Japan, earthquakes. As with other new developments in structural engineering practice, it was implemented in building practice in the USA prior to inclusion in building codes or reference standards. The first inclusion of BRB design provisions was in the *NEHRP Recommended Provisions for Seismic Regulations for New Buildings and Other Structures* (FEMA 2003), FEMA 450. The information included in FEMA 450 was developed by a joint committee including AISC and the Structural Engineers Association of California (SEAOC). FEMA 450 is not a building code or reference standard, but inclusion is typically the first step in the process. The buckling-restrained braced frame was included in the 2005 version of ASCE/SEI 7, *Minimum Standard for Design Loads for Buildings and Other Structures* (2005), as an acceptable seismic force-resisting system. In ASCE/SEI 7-05, a BRBF had the following seismic performance factors (SPF): $R = 7$, $\Omega_o = 2$, $C_d = 5.5$. A dual system with BRBs and moment-resisting connections at the beam-column joints without a bracing connection had slightly different SPFs: $R = 8$, $\Omega_o = 2.5$, $C_d = 5$. One change of note that appeared in ASCE 7-10 was that the SPFs for BRBs with and without moment frames were made the same as the dual system based on research that had taken place in the interim. In addition to ASCE 7, BRB provisions were included in AISC Standard 341-05, *Seismic Provisions for Structural Steel Buildings* (2005). This document deals with more detailed requirements for design of not only the BRBs but also the other elements of the lateral force-resisting system. This document also lists the required testing protocol for BRBs. The subsequent versions of the ASCE/SEI 7 and AISC 341 standards continued to include provisions for BRBs with

modifications based on continuing research. An early and still applicable reference for the design of BRBs is by Lopez and Sabelli (2004) which has detailed examples of design provisions. A more recent textbook (Bruneau et al. 2011) also has a chapter on BRBFs.

From the design perspective, sizing of BRBs is a very straightforward process. The size of the core is based on the maximum axial demand according to the limit state of yielding based on the load combinations from the applicable building code. Equation 1 shows the basic premise where P_{ysc} is the nominal axial yield strength, A_{sc} is the area of the steel core, and F_{ysc} is either the minimum specified yield stress or the actual yield stress from a coupon test. If the manufacturer provides coupon tests from the steel used to fabricate the core, this value can be used in design. Otherwise, a range of values should be used to determine the upper and lower limits that could occur.

$$P_{ysc} = F_{ysc}A_{sc} \quad (1)$$

Once determined, the nominal yield strength value is then modified by a resistance factor of 0.9 for LRFD design to determine the design axial strength. One of the benefits of BRBs is that any cross-sectional area can be provided so each floor can have a BRB sized to the demand. Typically the area will be rounded to some degree based on the manufacturer's practice, but this is not generally a significant change. In addition to nominal yield, the adjusted brace strength must also be established as part of the design. The adjusted brace strength has two components, the tension ($\omega R_y P_{ysc}$) and compression adjusted brace strength ($\beta \omega R_y P_{ysc}$), where ω is the strain hardening adjustment factor, β is the compression strength adjustment factor, and R_y is the ratio of expected to specified minimum yield stress. R_y can be set to 1 if coupon tests are used to determine the yield strength. The values for β and ω are taken from coupon tests that are carried out to the strain equivalent to that occurring in the brace at the expected deformation. The expected deformation is defined as the larger of 2 % story drift or two times the design story drift. Figure 7 shows a

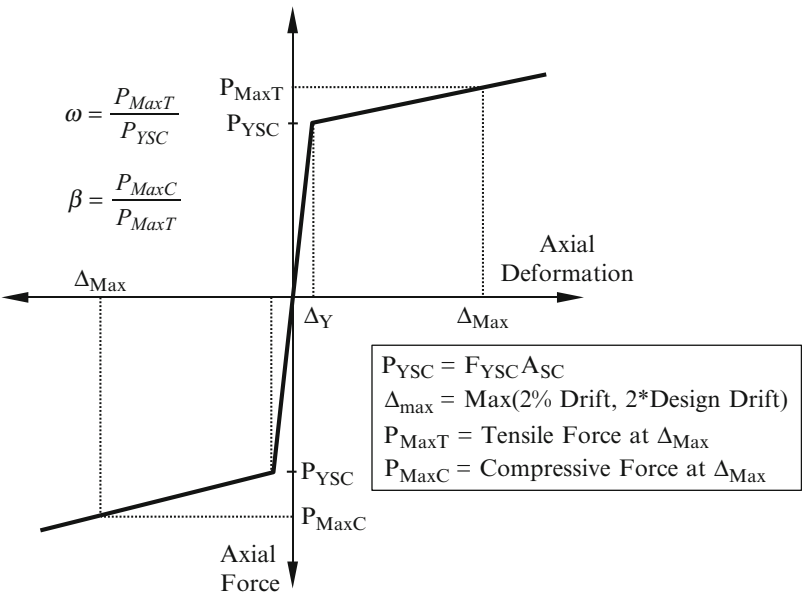
plot of how these values would be determined. This is one of several items that will require input from a brace manufacturer. The adjusted brace strength is used to ensure that the beams, columns, and connections in the frame have the required strength to remain elastic through the expected strength of the BRB. The design of brace connections must be designed to 1.1 times the adjusted brace strength in compression.

The other component of design that is important is accurate modeling of brace stiffness. As discussed previously, the stiffness of the brace represents a combination of the stiffness of the different segments. This versatility provides additional mechanisms to control response by modifying the stiffness of the brace. For example, if the brace is too flexible, the yield length could be shortened. However, this will have an effect on other components as the inelastic strains in the brace for the same deformation will increase which will also increase the adjusted brace strengths. Brace manufacturers will provide assistance to create the appropriate balance of stiffness and required strength.

One of the final design considerations is that of the connection of the BRB to the framing. There are three typical options for connecting BRBs to gusset plates although the specific details of the connections vary based on the different manufacturer's brace geometry. The three options are bolted, welded, and pinned connections. Example details are shown in Fig. 8 although many variations can be found and have been used. As opposed to CBFs where the gusset plate is designed to allow for the out-of-plane rotation due to brace buckling, BRBs can sustain significant moments. These large moments are developed based on the large deformations in the frame inducing rotations in the beam and columns. The pin allows for these rotations to occur in the connection rather than the structural element. This allowance for rotation can also be developed using welded and bolted gusset connections by providing a beam stub off the column with a hinge connection to the remainder of the beam as shown in Fig. 8c. It is also of critical importance to ensure that the connection (gusset plate and end section of BRB) has sufficient

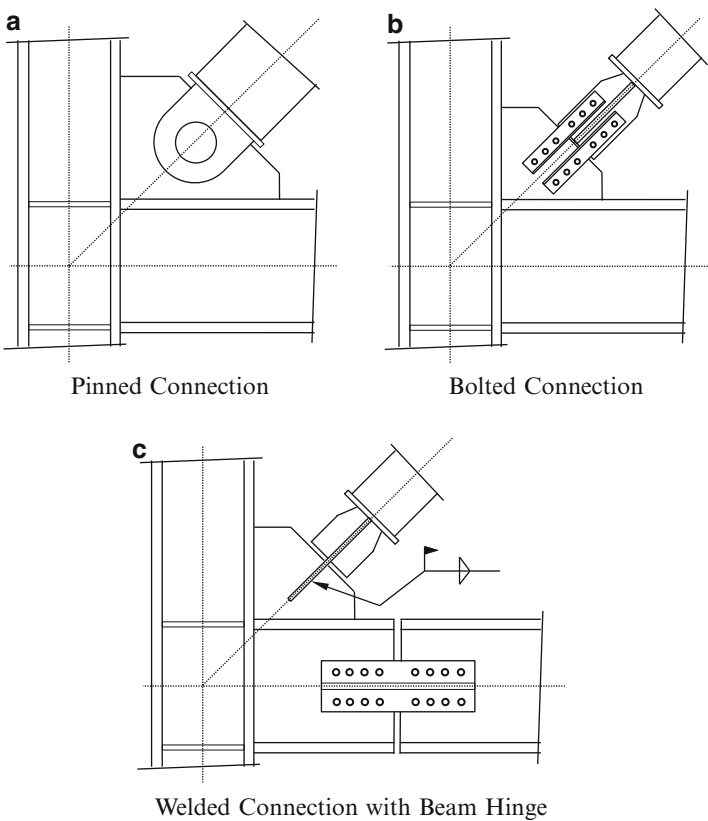
Buckling-Restrained Braces and Their Implementation in Structural Design of Steel Buildings,

Fig. 7 Determination of adjusted brace strength from test data



Buckling-Restrained Braces and Their Implementation in Structural Design of Steel Buildings,

Fig. 8 Typical gusset plate connection details



out-of-plane flexural resistance to resist buckling up to the full strength of the core section. Typical gussets for BRBFs are markedly smaller than those used in CBFs.

In addition to typical frame structures, BRBs can be used anywhere a ductile fuse element is required. Examples include bridges as ductile connections between the superstructure and the substructure, stadiums as part of the lateral resisting systems, and dam intake towers to prevent lateral motion while limiting the forces imparted to the dam. They have also been used more recently in the San Francisco Airport Air Traffic Control Tower which is a self-centering system where the BRBs provide hysteretic energy dissipation (Muthukumar and Sabelli 2013).

It is interesting to note that the design for BRBs in Japan is different from the USA. In Japan, braced frames are designed to higher force levels than moment frames (i.e., the equivalent of the R value penalizes braced frames) due to the potential for buckling. Because of this fact, most low-rise construction in Japan is designed with moment frames. However, for taller buildings, BRBs are used significantly due to the increased stiffness. In order to use BRBs for their stiffness and energy dissipation without the force penalty, they are used as hysteretic energy-dissipating devices as part of a dual system with moment-resisting frames (Xie 2005). The first reported use of a BRB in Europe was in 2006 at the University of Ancona in Italy. Use of BRBs in Europe is much less than in Japan and the USA (Della Corte et al. 2011).

Buckling-Restrained Brace Innovations

Buckling-restrained braces are one of the more impactful developments in earthquake engineering. They are considered today a standard option for both new designs and retrofits of existing buildings in significant seismic hazards. However, as with any innovation, additional improvements are constantly explored to further improve

on the existing concept. One of the biggest weaknesses of BRBs that has been identified in analytical and experimental research is residual displacements following significant earthquakes. The modifications and/or additions to BRBs have primarily aimed to reduce residual deformations in BRBFs. This movement is also centered on a larger goal of providing a higher performance standard than life safety in buildings for major earthquakes. Since BRBs are an effective lateral system, additions and modifications to a high-performance seismic system represent an ideal way to move toward damage reduction in addition to life safety.

The first of these innovations is called the hybrid passive control device (HPCD). The HPCD consists of a BRB in series with a viscoelastic damping device. For small deformations, only the viscoelastic damper is active. After a specified displacement, the viscoelastic device locks out and the BRB becomes active. The device was tested in the laboratory and implemented in analytical models of a nine-story structure. Experimental testing showed the concept worked as designed. The building response to a suite of earthquakes was improved, and residual displacements were reduced when compared to a BRB frame (Marshall and Charney 2010a, b).

One of the ways that has been explored to improve building response and reduce damage is through the use of self-centering systems. The idea of these systems is to provide an element that remains elastic while a fuse element yields and dissipates energy. Analytical and experimental research was completed on a self-centering buckling-restrained brace (SC-BRB) using nickel-titanium-based shape memory alloys. The shape memory alloy's unique properties provide the capacity to self-center in addition to providing additional energy dissipation. The SC-BRB consists of typical BRB with two additional outer concentric tubes and free-floating end plates so the shape memory alloy rods were always loaded in tension. The experimental testing showed significant capacity for energy

dissipation, deformation ductility, and self-centering which would result in a lateral system with the ability to reduce structural response and limit residual damage in the event of a large earthquake (Miller et al. 2012).

An additional innovation associated with BRBs is the use of a multi-core BRB with different steel grades, called a hybrid BRB (HBRB). The idea of using a low yield stress steel core (100 MPa/14.5 ksi) in combination with a typical or higher yield stress steel presents a unique option. Since the modulus of elasticity is the same for the two steels, the low yield stress steel will yield earlier and begin to dissipate energy while the structure still has significant stiffness. The system can be optimized based on the ratio between the area of low yield point steel and the other steel in the BRB. The idea here is to dissipate significant energy before the structure reaches a state where the stiffness has decreased significantly. This concept has been shown to reduce residual displacement and decrease the number of simulated collapses when compared to typical BRBFs (Atlayan and Charney 2012).

Summary

Buckling-restrained braced frames are a resilient lateral system for structural steel buildings. The ability to control both the stiffness and strength, nearly independently of each other, and the large, full hysteresis loops makes a BRBF an effective seismic system that can be designed for many different hazards. BRBs are versatile and can be used in many different types of structures including both new construction and seismic retrofits. This has been demonstrated by both the increase in usage for construction projects and the continued worldwide experimental research, both completed and ongoing. As with any lateral resisting system, there are weaknesses which need to be resolved. Continued improvement through research and development will continue to improve performance of this ductile lateral resistance system.

Cross-References

- [Buildings and Bridges Equipped with Passive Dampers Under Seismic Actions: Modeling and Analysis](#)
- [Cast Fuse Braces: Design and Implementation](#)
- [Earthquake Protection of Essential Facilities](#)
- [Friction Dampers for Seismic Protections of Steel Buildings Subjected to Earthquakes: Emphasis on Structural Design](#)
- [Passive Control Techniques for Retrofitting of Existing Structures](#)
- [Seismic Analysis of Steel Buildings: Numerical Modeling](#)
- [Steel Structures](#)

References

- AISC (2005) ANSI/AISC 341–05 Seismic Provisions for Structural Steel Buildings, American Institute of Steel Construction, Chicago, IL
- AISC (2010) ANSI/AISC 341–10 Seismic Provisions for Structural Steel Buildings, American Institute of Steel Construction, Chicago, IL
- Atlayan O, Charney FA (2012) Hybrid steel frames. In: 15th World conference on earthquake engineering, International Association for Earthquake Engineering, Lisbon, Sept 24–28
- Black CJ, Makris N, Aiken ID (2004) Component testing, seismic evaluation and characterization of buckling-restrained braces. *J Struct Eng* 130(6): 880–894
- Bruneau M, Uang C-M, Sabelli R (2011) Ductile design of steel structures, 2nd edn. McGraw Hill, New York
- Della Corte G, D’Aniello M, Landolfo R, Mazzoniani FM (2011) Review of steel buckling-restrained braces. *Steel Constr* 4(2):85–93
- Fahnestock LA, Ricles JM, Sause R (2006) Experimental study of a large-scale buckling restrained braced frame using the pseudo-dynamic testing method. In: 8th US National conference on earthquake engineering, Earthquake Engineering Research Institute, San Francisco
- Lopez WA, Sabelli R (2004) Seismic design of buckling-restrained braced frames. Structural Steel Educational Council, Moraga
- Mansour N, Christopoulos C, Tremblay R (2011) Experimental validation of replaceable shear links for eccentrically braced steel frames. *J Struct Eng* 137(10):1141–1152
- Marshall JD, Charney FA (2010a) A hybrid passive control device for steel structures I: development and analysis. *J Constr Steel Res* 66(10):1278–1286

- Marshall JD, Charney FA (2010b) A hybrid passive control device for steel structures II: physical testing. *J Constr Steel Res* 66(10):1287–1294
- Miller DJ, Fahnstock LA, Eatherton MR (2012) Development and experimental validation of a nickel-titanium shape memory alloy self-centering buckling-restrained brace. *Eng Struct* 40:288–298
- Muthukumar S, Sabelli R (2013) Nonlinear seismic analysis of a round concrete tower with a post-tensioned self-centering system. *ASCE/SEI Structures Congress 2013*, American Society of Civil Engineers, Pittsburgh, 2–4 May 2013
- UBB (2014) <http://www.unbondedbrace.com/facts.htm>. Accessed Jan 2014
- Wakabayashi M, Nakamura T, Kashibara A, Morizono T, Yokoyama H (1973) Experimental study of elastoplastic properties of precast concrete wall panels with built-in insulating braces. *Summaries of technical papers of annual meeting, Architectural Institute of Japan*, pp 1041–1044 (in Japanese)
- Xie Q (2005) State of the art of buckling-restrained braces in Asia. *J Constr Steel Res* 61(6):727–748

"Build Back Better" Principles for Reconstruction

Sandeeka Mannakkara, Suzanne Wilkinson and Tinu Rose Francis
Department of Civil and Environmental Engineering, The University of Auckland, Auckland, New Zealand

Synonyms

BBB; Reconstruction; Recovery

Introduction

"Build Back Better" signifies an ideal reconstruction and recovery process that delivers resilient, sustainable, and efficient recovery solutions to disaster-affected communities. The motivation behind the Build Back Better concept is to make communities stronger and more resilient following a disaster event. Statistics from the United Nations Environment Programme in 2008 show an increase in the number of natural disasters over time attributing to growing populations,

urban growth in risk-prone areas due to scarcity of land, and global warming. Along with increasing frequency, recent disasters show an increase in magnitude and resulting destruction according to studies by the Red Cross. Both natural and technological/man-made disasters have seen nearly exponential rises in the number of disasters over time.

Despite the increasing number of disaster experiences, post-disaster activities remain inefficient and poorly managed and need to be improved according to Halvorson and Hamilton (2010). Traditionally, post-disaster reconstruction consisted of simply repairing the physical damage that has been induced by a disaster. However, Kennedy et al. (2008) pointed out that rebuilding the built environment and infrastructure exactly as they were prior to a disaster often re-creates the same vulnerabilities that existed earlier. If restored to pre-disaster standards, disaster-affected communities would face the same difficulties if exposed to another disaster event in the future. The reconstruction and recovery period following a disaster poses an opportunity to address and rectify vulnerability issues found in communities.

As a result of witnessing the ongoing impacts of disasters on communities, a concept started to emerge where post-disaster reconstruction was to be taken as an opportunity to not only reconstruct what was damaged and return the community to its pre-disaster state but to also seize the opportunity to improve its physical, social, environmental, and economic conditions to create a new state of normalcy that is more "resilient" (Boano 2009). This concept was termed "Build Back Better," suggesting that successful recovery of communities following disasters needs to amalgamate the rehabilitation and enhancement of the built environment along with the psychological, social, and economic climates in a holistic manner to improve overall community resilience. The phrase "Building Back Better" became popular during the large-scale reconstruction effort following the Indian Ocean Tsunami disaster in 2004 after which it became more officially embraced with the creation of sets of BBB Guidelines to steer recovery and

reconstruction activities toward achieving this goal (Clinton 2006).

This chapter reviews what BBB entails and presents the key elements required to improve post-disaster reconstruction and recovery practices to build back better. First, existing guidelines and reports providing recommendations for BBB are introduced. Key information from the guidelines and reports is used to then identify the key concepts which represent Building Back Better. Finally, each concept and its importance for building back better are reviewed.

The Need for Building Back Better

The South Asia Disaster Report (DNS and PA 2005) states that disasters are produced due to the weaknesses and vulnerabilities of communities, countries, and structures to withstand encountered hazards. Wisner et al. (2004) defines vulnerability as the lack of capacity to anticipate, cope with, resist, and recover from the impact of a hazard. The destruction and loss of human lives from the 2005 Kashmir Earthquake in Pakistan was primarily due to the collapse of inappropriately built structures constructed on earthquake-prone land using substandard building materials and designed with little earthquake resistance. Poorly planned and sometimes illegal developments and their resulting impacts on the environment worsened the damage from the Mumbai Floods in 2005. A similar situation was seen in Sri Lanka after the Indian Ocean Tsunami.

Restoration of the damaged physical, social, economic, and environmental impacts of disasters is a complicated and drawn-out process. Reconstruction and recovery projects often focus on quick restoration of affected communities which can replicate and worsen existing vulnerabilities faced by the community. The Tsunami Evaluation Commission Synthesis Report in 2007 provided examples where escalated pressures and the need for fast rebuilding and recovery processes following a disaster can further increase the vulnerability of a community. Examples include: nonadherence to design and construction policies for buildings and

infrastructure, insufficient focus given to certain aspects of the recovery process such as livelihood development programs and small business support programs, overruling of local government agencies, and neglecting vulnerable groups of people in the community.

Complete recovery requires attention to many different elements. BBB was defined by Khasalamwa (2009) as a way to utilize the reconstruction process to improve a community's physical, social, environmental, and economic conditions to create a more resilient community, where resilience is defined as "the capacity to recover or 'bounce back' after an event" (Twigg 2007). Therefore, what the concept of BBB proposes is a broad holistic approach to post-disaster reconstruction in order to address the wide range of prevalent issues including those mentioned above and ensure that affected communities are regenerated in a resilient manner for the future.

Existing Guidelines for Building Back Better

Clinton's (2006) "Key Propositions for Building Back Better" was the earliest known official document to be published which attempted to provide a comprehensive guideline for implementing BBB practices in post-disaster environments. The report was based on and aimed at the Indian Ocean Tsunami disaster. He introduced ten propositions for building back better.

Clinton's propositions were:

- Proposition 1: Governments, donors and aid agencies must recognize that families and communities drive their own recovery.
- Proposition 2: Recovery must promote fairness and equity.
- Proposition 3: Governments must enhance preparedness for future disasters.
- Proposition 4: Local Governments must be empowered to manage recovery efforts, and donors must devote greater resources to strengthening Government recovery institutions, especially at the local level.

- Proposition 5: Good recovery planning and effective coordination depend on good information.
- Proposition 6: The UN, World Bank, and other multilateral agencies must clarify their roles and relationships, especially in addressing the early stages of a recovery process.
- Proposition 7: The expanding role of NGOs and the Red Cross/Red Crescent Movement carries greater responsibilities for quality in recovery efforts.
- Proposition 8: From the start of recovery operations, Governments and aid agencies must create the conditions for entrepreneurs to flourish.
- Proposition 9: Beneficiaries deserve the kind of agency partnerships that move beyond rivalry and unhealthy competition.
- Proposition 10: Good recovery must leave communities safer by reducing risks and building resilience.
- Monday's "Holistic Recovery Framework" in 2002 which addresses enhancing the quality of life in the community, economic vitality, and the quality of the environment, risk reduction, and participatory decision-making in recovery activities
- Bam's Reconstruction Supreme Supervisory and Policymaking Association's "Bam's Reconstruction Charter" in 2010 which includes policies for reconstruction management; community participation, employing suitable construction technology and materials; preserving cultural and architectural heritage; and ensuring stability of construction
- Victorian Bushfire Reconstruction and Recovery Authority's "Recovery and Reconstruction Framework" in 2011 which focuses on the safety and well-being of the community, needs-based resource allocation, community engagement, equity, and tailored solutions
- Canterbury Earthquake Recovery Authority's "Recovery Strategy" in 2013 which entails leadership and integration to manage recovery activities using a participatory approach, regenerating the economy, restoring and enhancing the community, reconstruction of the built environment, and restoring natural and healthy ecosystems

Several other guidelines directly and indirectly proposing BBB-based recovery and reconstruction operations include:

- United Nations Disaster Relief Organization's "Principles for Settlement and Shelter" in 1982 which addresses stakeholder role allocation, needs-based provision of resources to the community, and risk reduction
- The Government of Sri Lanka's "Post-Tsunami Recovery and Reconstruction Strategy" and "Build Back Better Guiding Principles" in 2005 which include needs-based resource allocation and provision of locally appropriate solutions, community participation and consultation in recovery activities, equity, transparency between stakeholders, risk reduction and consideration of future sustainability, and livelihood support
- Federal Emergency Management Agency's "Rebuilding for a More Sustainable Future: An Operational Framework" in 2005 which mentions role allocation and coordination of stakeholders, community-centered recovery operations, and hazard-based sustainable risk reduction practices

Key Concepts

The concepts proposed to achieve BBB during reconstruction and recovery in the various guidelines in the previous section feature similarities. Aspects such as role allocation of stakeholders, community participation, and risk reduction appeared in most of the guidelines. The key concepts introduced in the guidelines for improving reconstruction and recovery efforts and building back better include: risk reduction, psychosocial recovery, economic recovery, effective implementation, and monitoring and evaluation. The next few subsections describe these key concepts in further detail.

Risk Reduction

Risk reduction identifies all actions taken toward reducing disaster risks in communities to improve the physical resilience in the built environment. Previous post-disaster experiences have emphasized the need to identify prevalent hazards and determine solutions to be undertaken to reduce risks imposed on people. The Red Cross's World Disaster Report in 2010 disclosed that the risks seen in cities are due to a number of reasons such as: growth in informal or illegal settlements, inadequate infrastructure, and building on sites at risk from hazards. The report also stated that many past disasters could have been anticipated and avoided with proper planning, design, and construction methods. The Victorian Bushfires Royal Commission Final Report in 2010 recommended the amendment of the Australian Building Code following the Victorian Bushfires ensuring greater safety standards. The Royal Commission suggested identifying bushfire-prone areas and adopting suitable building and planning controls.

The National Mitigation Strategy produced in Turkey following the Kocaeli and Duzce earthquakes of 1999 also stated the need for site-specific hazard identification before reconstruction as well as retrofitting and updating structural codes and the use of tax incentives to encourage mitigation work (Bakir 2004). The 2008 South Asia Disaster Report by the nongovernmental organizations Duryog Nivaran and Practical Action recommended producing hazard and vulnerability maps and enforcing building codes to avoid development-related disasters in the future. The two primary ways of risk reduction are through improving structural designs and through better land-use planning.

The importance of reviewing and changing building designs and codes to improve the structural integrity of buildings and infrastructure following a disaster is widely understood but is however less frequently attained successfully in practice due to a range of common issues. Poor regulative powers and the lack of strict enforcement can lead to building code changes being disregarded resulting in substandard structures

in the rebuild. When the Indian Ocean Tsunami struck, enforcement of building codes was mainly restricted to urban and suburban areas in Sri Lanka. The rural and coastal areas were the main victims of the disaster, where the lack of strict structural standards resulted in magnified damage (Pathiraja and Tombesi 2009). Extra costs incurred by adopting new technologies and materials to improve structural resilience also discourage compliance of new building codes worldwide (Batteate 2006).

The experiences of post-disaster reconstruction efforts worldwide have provided lessons which can be adopted when implementing structural changes to avoid the above mentioned issues and build back better. BBB theory suggests that hazard-based building regulations should be created using multi-hazard assessments in areas chosen for redevelopment and reconstruction. Consistent regulations and a strong legal framework are necessary to assist the adoption of building codes and regulations and ensure that structural changes improve the built environment (Clinton 2006). As structural improvements are expensive and unaffordable especially in post-disaster settings, long-term funding needs to be made available to cover extra costs for structural improvements and promote adoption. Quality of reconstruction can be maintained by arranging inspections during construction by local governmental authorities. Stakeholders involved in the rebuild such as builders, engineers, and architects should be trained on revised building codes and other specific requirements to avoid inconsistencies and produce good quality results in order to build back better.

A land-use planning strategy was used in the post-disaster recovery efforts following the Indian Ocean Tsunami and the Samoan Tsunami, resulting in the relocation of coastal communities further inland to prevent future impacts of coastal hazards (Kennedy et al. 2008). The mandatory resettlement operations in Sri Lanka and Samoa were problematic due to the lack of consideration given to the lifestyles of the local people which led to the loss of their sea-dependent livelihoods,

dissatisfaction with their new settlements, and illegal return of people to the original coastal lands (Kennedy et al. 2008). A recurring issue with relocation is the focus given to moving communities away from a certain hazard resulting in exposing communities to new unanticipated hazards. Well-intended land-use planning measures can also fail due to the lack of knowledge and awareness of local people who do not conform to new regulations and the lack of experience and knowledge of local governing authorities who do not enforce new regulations (Kennedy et al. 2008).

Therefore, it was recommended by Baradan (2006) that hazard assessments of current land sites and possible new land sites should be carried out, after which risk zone maps are to be created which divide the land into zones based on the level of risk. Appropriate land uses and new planning and building regulations based on the risk zone maps are to be created. The risk zone maps should be legislated and included in council development plans and approval permit procedures to ensure compliance. Examples, such as Taiwan's Mitigation Plans, the Philippines Municipal Maps, and the Christchurch City Plan in New Zealand following the Canterbury earthquakes, display successful application of BBB measures to create safer developments. If resettlement to lower-risk lands is opted for, Mannakkara and Wilkinson (2012b) recommend that a comprehensive resettlement strategy should be created with community consent which takes into account risk levels of new lands, community preferences, and livelihood and lifestyle opportunities offered in the new locations for resettlement to be a success.

DN and PA (2008) encourage educating communities about risks and the importance of risk reduction measures and engaging them in collective risk reduction efforts. The Participatory Flood Risk Communication Support System (PafriCS) developed in Japan to educate locals and other stakeholders including NGOs and local governments on flood risks and risk management strategies is an example of a participatory tool.

Psychosocial Recovery

Supporting psychosocial recovery of affected communities has been identified as essential for building back better (Davidson et al. 2007). Post-disaster recovery often focuses on providing fast solutions in an attempt to reestablish a sense of normality in affected communities as soon as possible (Khasalamwa 2009). The focus on speed results in overlooking the real needs of communities. The community is often not consulted to provide their input on reconstruction and recovery (Boano 2009). The lack of community consultation and participation leads to the provision of recovery solutions that are not suitable. For example, some of the new houses constructed in Sri Lanka by humanitarian agencies during the Indian Ocean Tsunami rebuild featured bathrooms made with half-heighted walls and shared bathrooms for males and females which were culturally unacceptable (Ruwanpura 2009). Locals were unhappy with the reconstruction of homes following the 1999 Marmara Earthquake in Turkey as their local life, culture, and aesthetics were not considered. Khasalamwa (2009) stated that insufficient attention to social, cultural, and ethnic facets of communities during recovery can exacerbate preexisting vulnerabilities. Separation during disasters and resettlement operations disrupt community cohesion and psychological recovery (Florian 2007).

Social issues arising in post-disaster environments are primarily related to social/cultural/religious/ethnic factors, and psychological factors. Reconstruction is a chaotic and stressful time for individuals who are also experiencing trauma. These communities require various forms of assistance as part of building back better. Personalized advice and one-on-one support provided to families in Columbia during the 1999 earthquake recovery were a success. Similar forms of personal assistance were provided during the Victorian Bushfire recovery in Australia as well. James Lee Witt Associates (2005) recommended arranging specialized assistance for vulnerable communities. Providing psychological support and counseling are essential during recovery. The establishment of information centers which

offer easy access to recovery-related information for the community is also recommended. Upholding a sense of community spirit and improving community cohesion through organizing group activities are recommended for social recovery. The Canterbury Earthquake Recovery in Christchurch proposed sports, recreation, arts, and cultural programs to engage the community and provide a sense of normality.

One of the first steps to be taken in post-disaster recovery efforts in order to build back better is to understand the local context of the affected community through needs assessments and surveys in order to provide appropriate assistance to satisfy the community (Khasalamwa 2009). The reconstruction and recovery policies must then be developed based on local requirements to support and preserve the local culture and heritage. Batteate (2006) stated that maintaining community involvement throughout recovery is integral for BBB success. The importance of decentralization to empower disaster-affected communities by enabling them to take responsibility of the recovery effort and become involved in decision-making has been stressed by literature. The establishment of community consultation groups is an effective way to communicate with the community. Community consultation groups consisting of community leaders from preexisting community groups and reputed members of the community to liaise between the wider community and governmental authorities have been successful in Sri Lanka and India. Existing community groups can also be called upon to assist with recovery activities.

Economic Recovery

Supporting economic recovery of the community and supporting livelihood regeneration and entrepreneurship are also an important part of recovery. Disasters cause damage to the economy of communities with the disruption of businesses and income-generating industries leading to issues such as high inflation rates and poverty. The adverse effects of disasters on the economy can also impede the overall recovery of a city. Hurricane Katrina displayed a disaster's long-term impacts on higher education and health

care in New Orleans, which were the foundations of the city's economy, eventually leading to a decline in population numbers as people moved away in search of better opportunities.

Post-disaster recovery efforts to date have shown support for economic recovery with strategies such as: "cash-for-work" programs, provision of business grants, "asset replacement" programs to provide industries with necessary resources, and training programs to up-skill locals and help them find work. In Aceh, Indonesia, tsunami-affected people were trained and employed in reconstruction to provide them with a source of income alongside the opportunity to become involved in their own recovery (Kennedy et al. 2008). In Japan following the 2011 earthquake and tsunami, the government decided to consolidate smaller fishing markets into large fishing centers to enable fishermen to support each other (Okuda et al. 2011). The Christchurch City Council's Central City Plan proposes fast-tracking of building consents for businesses to allow faster repair and construction work. Despite the implementation of such initiatives, post-disaster economic recovery is reportedly slow and below pre-disaster levels. The lack of success in economic recovery initiatives can be attributed to insufficient backing from policies and legislation for employment creation and lack of consideration given to the needs of affected communities.

Clinton (2006) said in his BBB propositions that "a sustainable recovery process depends on reviving and expanding private economic activity and employment and securing diverse livelihood opportunities for affected populations." Thus, the uniqueness of BBB comes from the integrated approach it proposes by giving economic recovery as much importance as reconstruction and aiming to provide solutions to suit local dynamics and preferences.

Monday (2002) stated that one of the first steps needed for effective economic recovery is to obtain accurate information about the local population through data collection and consultation with local governmental authorities, and a comprehensive economic recovery strategy must be created that is tailor-made to suit each

different community based on data obtained. Where applicable, attractive and flexible low-interest loan packages, business grants, and resources should be provided to support the livelihoods of the disaster-affected. Training programs should be held to support people in improving their existing livelihoods or acquire new skills. Mannakkara and Wilkinson (2012a) propose that business support and counseling services should be provided to assist with the economic recovery. Rebuilding of businesses must also be facilitated through special fast-tracked permit procedures. Incentives such as subsidized accommodation must be provided to attract builders from other areas to participate in rebuilding.

Effective Implementation

A successful recovery effort requires effective and efficient recovery solutions as part of building back better. Two ways in which the efficiency and effectiveness of post-disaster recovery can be improved are through better management of stakeholders and through the use of appropriate post-disaster legislation and regulation.

One of the most common issues with post-disaster environments is the difficulty in coordinating between stakeholders to produce a unified outcome. Initially, there is often no organization in charge of the overall recovery effort. The lack of guidance leads different stakeholders to participate disjointedly promoting personal agendas which conflict with the interests of the local community (Batteate 2006). For example, nongovernmental organizations (NGOs) who operated in Sri Lanka following the Indian Ocean Tsunami constructed homes which were unsuitable for locals and were largely abandoned. The pressure for fast results during recovery also prevents well-intentioned stakeholders from considering community needs. Ambiguity about the roles of different stakeholders is another issue. The Victorian Bushfires Royal Commission report, 2010, stated that the roles of personnel involved in the recovery effort were unclear which led to the duplication of some activities. Many stakeholders involved in recovery have no previous experience in post-disaster

environments leading to ad hoc responses (Kennedy 2009). Often post-disaster interventions are governed by the national government without sufficient consultation or power given to local councils (Clinton 2006). Local-level organizations with useful local knowledge lack the capacity to operate to their full extent when impacted by disasters and are therefore excluded from recovery efforts. The lack of proper role allocation, coordination, and involvement of local-level stakeholders is a common issue found in post-disaster reconstruction environments.

A step taken to improve the management of large numbers of stakeholders in major disasters in order to build back better is the creation of a separate body to act as a recovery authority. Examples of recovery authorities created to manage reconstruction include: the Bureau of Rehabilitation and Reconstruction (BRR) in Indonesia following the Indian Ocean Tsunami, Bam's Reconstruction Supreme Supervisory and Policymaking Association (BRSSPA) in Iran following the 2003 Bam Earthquake, the Victorian Bushfire Reconstruction and Recovery Authority (VBRRA) in Australia following the 2009 Victorian Bushfires, and the Canterbury Earthquake Recovery Authority (CERA) following the 2010 and 2011 Canterbury earthquakes in New Zealand. The recovery agencies contributed to the success of recovery to differing extents. Clinton (2006) said that stakeholders must operate with a common set of standards, approaches, and goals in order for recovery to be a success. Twigg (2007) proposes that the recovery authority should be responsible for establishing clear roles and responsibilities for the different stakeholders to divide recovery tasks based on resources and skills and avoid duplication.

Functional partnerships and linkages established between organizations can enhance reconstruction projects. Post-disaster recovery is a unique environment which requires deviation from normal procedures. Information sharing between organizations is one such deviation. The Federal Emergency Management Agency in the United States advocates the sharing of information, contacts, resources, and technical

knowledge between organizations to help recovery activities (FEMA 2000). Knowledge from past disasters should be retained and transferred to the government and other relevant organizations who will be involved in future post-disaster efforts. Twigg (2007) recommends that local government should be included as a key stakeholder in the recovery effort and also given the responsibility to manage local-level activities.

Another obstacle preventing successful BBB-centered recovery is the absence of proper controls to enforce BBB principles. Having BBB knowledge and producing recovery plans in-line with these principles are futile without proper legislation and regulations in place to ensure they are implemented. A common challenge in post-disaster environments is the sudden increased work load, especially in the building industry, along with a drop in the workforce across local organizations which slow down and impede recovery activities. Post-disaster reconstruction requires time-consuming activities such as hazard analysis, land selection, infrastructure development, and rebuilding to be done in a relatively short period of time. It is important to facilitate recovery-related activities by simplifying, fast-tracking, and exempting certain rules and regulations using special legislation.

Post-disaster legislation can be used to ensure compliance with BBB-based activities as well as to facilitate normal operations to improve the efficiency of recovery efforts. The lack of enforcement of hazard-related laws and adequate risk-based building controls contributed to the large-scale devastation caused by the 2004 Indian Ocean Tsunami (DNS and PA 2005). The same was seen in countries like Pakistan, Turkey, Samoa, and Haiti. Enforcing updated risk-based building design standards through the use of compulsory building codes and maintaining construction standards through careful inspections is an important regulatory requirement in reconstruction (James Lee Witt Associates 2005). Lack of awareness and understanding of new legislation can also lead to noncompliance. In the post-tsunami recovery effort in Sri Lanka, external nongovernmental organizations (NGOs) who took part did not comply with local standards

due to unawareness (Boano 2009). The National Post-Tsunami Lessons Learned and Best Practices Workshop held in Sri Lanka in 2005 highlighted the importance of training stakeholders (especially external NGOs) about existing and newly introduced legislation and regulations. The community's support can also be obtained by educating them about legislation and regulations that must be adhered to in reconstruction and recovery.

Post-disaster legislation can also be used to simplify and assist recovery activities to speed up the recovery process. Legislation that is customarily used to impose security and safety controls (such as building consents) can become an obstacle in high-pressure post-disaster environments. Time-consuming procedures, insufficient resources to process permits, and the lack of fast-tracked methods delay reconstruction. Delays in permits were a major reason for the holdup in housing repair and rebuilding following the 2005 Bay of Plenty storm in New Zealand (Middleton 2008). Fast-tracked consenting procedures, collaboration with other local councils, and open access to information between stakeholders can help speed up recovery.

Legislation can be used to remove unnecessary red tape to facilitate recovery activities. Meese III et al. (2005) reported a good example in the recovery following the 1994 Northridge Earthquake, USA, where legislative suspensions and emergency powers greatly reduced highway reconstruction time. The construction work provided employment and opening up the highways soon after the disaster helped boost the economy.

Monitoring and Evaluation

The effectiveness and efficiency of post-disaster reconstruction and recovery activities is crucial to the success of a community's restoration following the impact of a disaster event. Having the knowledge of Build Back Better concepts in designing recovery programs is insufficient without systems in place to overlook and monitor implementation. The creation of a recovery strategy to assist in conducting post-disaster reconstruction and recovery activities is a common response following disaster events. Despite

having recovery strategies and revisions in legislation and regulation to improve recovery activities, the findings by Tas (2010) indicated that compliance was not monitored in the respective recovery efforts in Sri Lanka and Turkey, leading to poorly executed recovery projects. The lack of properly trained professionals who were competent in post-disaster environments and disaster management activities poorly affects the outcome of recovery efforts. The shortage of effective information and knowledge sharing and dissemination are also reasons for unsatisfactory disaster management practices. Findings from the Business Civic Leadership Center in 2012 on "What a Successful Recovery Looks Like" raised concerns that long-term recovery beyond reconstruction often does not take place due to the lack of mechanisms and expertise which prevents affected communities from satisfactorily "building back better" in the long run.

Recommendations to improve post-disaster recovery efforts through monitoring and evaluation have been provided in many sources of literature. The role of monitoring and evaluation is twofold: (1) to monitor and ensure compliance of recovery activities in accordance with the recovery strategy in place and relevant guidelines and regulations (Clinton 2006) and (2) to obtain lessons for the future and improve future disaster management and post-disaster reconstruction and recovery efforts (Monday 2002).

The 2003 Bam earthquake reconstruction provided a good example where rebuilding was monitored by providing construction supervision which assisted in assuring the quality of the rebuild. Clinton (2006) stated that the Tsunami Recovery Impact Assessment and Monitoring System (TRIAMS) was put in place during the Indian Ocean Tsunami recovery for the most affected countries. The recovery strategy in Christchurch, New Zealand, has also been equipped with monitoring mechanisms. Clinton (2006) suggested that long-term recovery should be monitored through continued data collection to ensure that recovery efforts do not leave communities with residual issues.

Monday (2002) pointed out that monitoring can be used to identify problems with post-

disaster interventions and establish lessons learnt. Lessons learnt should be incorporated into revising policy and procedures for future disaster management practices. Bakir (2004) recommends that public education campaigns should be run on lessons learnt, including the community in participatory disaster management. Public seminars have been held and advice notes have been distributed in Australia during the Victorian Bushfires recovery to keep the community informed about revised guidelines and standards. Workshops have been held in the Philippines, Japan, and California involving the community in vulnerability identification which have been successful (Batteate 2006).

Summary

"Build Back Better" is an important concept for post-disaster reconstruction and recovery, signifying the need to use reconstruction as an opportunity to not only recover from the encountered disaster but to improve the resilience of communities to face and withstand future disaster events. BBB represents adopting a holistic approach toward recovery by addressing risk reduction of the built environment, psychosocial recovery of affected people, and rejuvenation of the economy in an effective and efficient manner. Risk reduction can be achieved primarily through the improvement of structural designs in buildings and infrastructure and through better risk-based land-use planning. BBB requires improved building codes and land-use plans to be enforced using a strong legal framework along with financial backing to encourage adoption. Quality assurance of the rebuild is also integral for building back better.

Psychosocial recovery needs to be addressed to assist communities with moving forward with their lives as an important part of overall community recovery. Psychosocial recovery of affected people needs to be assisted through the provision of support services such as personal case management, counseling, and social activities. Inclusion of community members in recovery activities is another way to support psychosocial

recovery and provide recovery solutions that are in-line with community needs as part of building back better.

Economic recovery is essential for the recovery of communities. An informed economic strategy to address and support community-specific issues is the first step toward BBB-based economic recovery. Financial assistance, training, and business rebuilding support need to be provided to assist with economic recovery.

Reconstruction and recovery requires effective management of stakeholders and the use of post-disaster legislation and regulation in order to build back better. The creation of a recovery authority to allocate roles and coordinate and manage stakeholders is recommended. Successful recovery requires local-level partnerships and contribution to provide locally viable recovery solutions. Compliance of BBB-based concepts in recovery needs to be ensured through the use of appropriate post-disaster legislation and regulation to enforce risk reduction and community recovery initiatives. Legislation and regulation can also be used to facilitate post-disaster recovery activities by fast-tracking and exempting normal procedures.

The effective implementation of risk reduction and community recovery initiatives concurrently will result in building back better. Recovery efforts also need to be monitored continuously through short-term and long-term recovery to ensure compliance with BBB concepts and to obtain lessons to improve future disaster management efforts.

Build Back Better: The Way Forward

Understanding and implementing the concept of Building Back Better is integral to improving a community's resilience following a disaster event in order to achieve positive changes for affected communities. The elements required to build back better which are; risk reduction, psychosocial recovery, economic recovery, effective implementation and monitoring and evaluation, have been introduced and discussed in this chapter. This understanding can be used to create

practical guidelines to design future reconstruction and recovery efforts including all these key facets to effectively build back better. Further comprehension of how these strategies for building back better can be more successfully implemented in different environments can be gained by studying different disaster events in the future. It is also suggested that criteria for measuring levels of resilience should be established which can serve as indicators to measure progress and effectiveness of build back better practices.

The long-term sustainability of resilience in communities instilled by using BBB concepts depends on how they are linked with on-going developmental strategies. It is therefore important for the key concepts identified in this chapter to be incorporated into local and national government policies for community planning and development even during non-disaster periods.

Cross-References

- [Building Codes and Standards](#)
- [Building Disaster Resiliency Through Disaster Risk Management Master Planning](#)
- [Community Recovery Following Earthquake Disasters](#)
- [Economic Recovery Following Earthquakes Disasters](#)
- [Land Use Planning Following an Earthquake Disaster](#)
- [Legislation Changes Following Earthquake Disasters](#)

References

- Bakir PG (2004) Proposal of a national mitigation strategy against earthquakes in Turkey. *Nat Hazards* 33(3):405–425
- Baradan B (2006) Analysis of the post-disaster reconstruction process following the Turkish earthquakes, 1999. In: GROUP IR (ed) International conference on post-disaster reconstruction meeting stakeholder interests, University de Montreal, Florence
- Batteate C (2006) Urban disaster risk reduction and regeneration planning: an overview. *Focus J City Reg Plan* Dep 3(1):11–17

- Boano C (2009) Housing anxiety and multiple geographies in post-tsunami Sri Lanka. *Disasters* 33(4):762–785
- Clinton WJ (2006) Lessons learned from tsunami recovery: key propositions for building back better. Office of the UN Secretary-General's Special Envoy for Tsunami Recovery, New York
- Davidson CH, Johnson C, Lizarralde G, Dikmen N, Sliwinski A (2007) Truths and myths about community participation in post-disaster housing projects. *Habitat Int* 31(1):100–115
- DN & PA (2008) Disaster and development in South Asia: connects and disconnects, South Asia disaster report. Duryog Nivaran & Practical Action, Colombo
- DNS & PA (2005) Tackling the tides and tremors: South Asia disaster report 2005. Duryog Nivaran Secretariat & Practical Action – South Asia Programme, Colombo
- FEMA (2000) Rebuilding for a more sustainable future: an operational framework. FEMA report. Federal Emergency Management Agency, Washington, DC
- Florian S (2007) Housing reconstruction and rehabilitation in Aceh and Nias, Indonesia – Rebuilding lives. *Habitat Int* 31(1):150–166
- Halvorson SJ, Hamilton JP (2010) In the aftermath of the Qa'yamat: the Kashmir earthquake disaster in northern Pakistan. *Disasters* 34(1):184–204
- James Lee Witt Associates (2005) Building back better and safer: private sector summit on post-tsunami reconstruction. James Lee Witt Associates, Washington, DC
- Kennedy J (2009) Disaster mitigation lessons from “build back better” following the 26 December 2004 Tsunami. In: Ashmore J, Babister E, Kelman I, Zarins J (eds) *Water and Urban Development Paradigms*, Taylor and Francis Group, London
- Kennedy J, Ashmore J, Babister E, Kelman I (2008) The meaning of ‘build back better’: evidence from post-tsunami Aceh and Sri Lanka. *J Conting Crisis Manag* 16(1):24–36
- Khasalamwa S (2009) Is ‘build back better’ a response to vulnerability? Analysis of the post-tsunami humanitarian interventions in Sri Lanka. *Nor J Geogr* 63(1):73–88
- Mannakkara S, Wilkinson S (2012a) Build back better principles for economic recovery: the Victorian bushfires case study. *J Bus Contin Emerg Plan* 6(2):164–173
- Mannakkara S, Wilkinson S (2012b) Build back better principles for land-use planning. *Urban Des Plan* 166(5):288–295
- Meese E III, Butler SM, Holmes KR (2005) From tragedy to triumph: principled solutions for rebuilding lives and communities, Heritage special report. The Heritage Foundation, Washington, DC
- Middleton D (2008) Habitability of homes after a disaster. In: 4th international i-REC conference on building resilience: achieving effective post-disaster reconstruction. International Council for Research and Innovation in Building and Construction, Christchurch
- Monday JL (2002) Building back better: creating a sustainable community after disaster. *Natural Hazards Informer* [Online], 3. Available <http://www.colorado.edu/hazards/publications/informer/infrmr3/informer3b.htm>. Accessed 27 Feb 2013
- Okuda K, Ohashi M, Hori M (2011) On the studies of the disaster recovery and the business continuity planning for private sector caused by Great East Japan earthquake. *Commun Comput Inf Sci* 219:14–21
- Pathiraja M, Tombesi P (2009) Towards a more “robust” technology? Capacity building in post-tsunami Sri Lanka. *Disaster Prev Manag* 18(1):55–65
- Ruwanpura KN (2009) Putting houses in place: rebuilding communities in post-tsunami Sri Lanka. *Disasters* 33(3):436–456
- Tas M (2010) Study on permanent housing production after 1999 earthquake in Kocaeli (Turkey). *Disaster Prev Manag* 19(1):6
- Twigg J (2007) Characteristics of a disaster-resilient community – a guidance note. DFID Disaster Risk Reduction Interagency Coordination Group, London
- Wisner B, Blaikie P, Cannon T, Davis I (2004) *At risk: natural hazards, people's vulnerability and disasters*. Routledge, New York

Building Codes and Standards

P. Gülkan¹ and Robert K. Reitherman²

¹Department of Civil Engineering, Çankaya University, Ankara, Turkey

²Consortium of Universities for Research in Earthquake Engineering, Richmond, CA, USA

Synonyms

Codes of practice; Guidelines; Regulations

Introduction

Selected key elements of modern seismic codes are reviewed here, along with a brief review of their historical development. What can be considered a single seismic code pertaining to a nation or a subdivision thereof consists of dozens of pages of provisions and by reference includes thousands of pages of standards that relate to loads, testing procedures, manufactured

building products, and materials. As used in this general discussion, “code” includes these other documents. Thus, it is far beyond the feasible scope here to provide an encyclopedic review of current seismic codes. Instead, this contribution is an attempt to delineate key features of most modern seismic codes and trace the development of their underlying concepts as they have evolved through the decades.

Preparing a global review of seismic codes runs the risk of turning out to be country specific and omitting other important references or milestones. This entry will attempt to avoid that. Instead the narrative will describe what a code should look like and what its important elements should contain. A building code places legal requirements on designers and constructors. Codes that fail to keep up with technological change and the advancement of knowledge become irrelevant. Building codes must be revised frequently, often at about 3-year increments. As more earthquakes occur and related building performance is observed, and as more research is conducted, the seismic provisions of a code often are one of the sections that change the most from one edition of a building code to the next.

The main purpose of building codes is to protect public health, safety, and general welfare as they relate to the construction and occupancy of buildings and structures. The building code becomes law of a particular jurisdiction when formally enacted by the appropriate governmental or private authority. The parts of a building code that relate to special requirements that must be fulfilled in order to inject a desired level of protection against the effects of ground shaking are called a seismic code. This broad description applies also to other types of construction, e.g., bridges, water tanks, towers, and port facilities. There are often additional codes or sections of a building code that have more specific requirements that apply to particular occupancies, for example, providing simplified requirements for the typical dwelling and more extensive requirements for the school, theater, or hospital.

Overview of Model Codes

A model building code is a convenient resource that can be adopted by the appropriate jurisdiction as its legal requirement. This makes the cost of maintaining and updating a code more economical and also provides design and construction consistency from one city to the next than would be the case if each developed its own code. That multiplicity of codes in a country was the rule throughout the nineteenth century and in many cases has only gradually trended toward nationally uniform provisions in the twentieth century. Two important interests that have pushed for uniformity are the construction and building materials industries, which can operate more efficiently if they have one set of rules, and the insurance industry, which desires up-to-date code provisions that can be easily evaluated for rate-setting purposes.

The practice of developing, approving, and enforcing building codes varies considerably among nations. Several countries have adopted model codes, including earthquake regulations, on a national basis, such as Japan, New Zealand, and Italy. In such countries, building codes are developed by the government agencies or quasi-governmental standards organizations and then made to apply across the country by the central government. Until 2000, the United States had three major model codes and associated seismic regulations, and even after their integration into the International Building Code, the process of adopting and enforcing the regulations, sometimes with significant variations, is left to the states and local governments. Similarly, in India each municipality and urban development authority has its own building code, which is mandatory for all construction within its jurisdiction. In Europe, the Eurocode is a pan-European building code that has all but superseded the older national building codes. Each country must now develop its own “national country annex” to localize the contents of the Eurocode. The seismic component of the Eurocode is only one small part of that model code. While the consistency of the regulations across European national

boundaries provides a better technical basis for its seismic and other provisions, the more important motive for such a code of European scope is economic. The Euro economic block can compete more effectively against the nations outside it if its design, construction, and building materials industries are guided by consistent provisions.

Prescriptive and Performance-Based or Objective-Based Codes

Building code requirements are usually a combination of prescriptive requirements that spell out exactly how something is to be done, on the one hand, and performance requirements that just outline what the required level of performance is and leave it up to the designer how this is achieved, on the other. An example of the former would be a rule of thumb for spacing of anchor bolts in house construction; an example of the latter would be to have the engineer calculate interstory drift and then design concrete cladding to accommodate that building-specific distortion. In recent years there has been a move among many building codes to move to more performance requirements and less prescriptive requirements. Performance-based code requirements still require tight definitions so that adequate performance can be evaluated by the building regulatory agency. The fire protection field has developed performance-based design approaches for many years, in which test or other data can be used to provide alternate means of fire protection instead of following the prescriptive requirements of a code.

In recent years, several countries, beginning with Australia, have moved to much shorter objective-based building codes. Rather than prescribing specific details, objective-based codes list a series of objectives all buildings must meet while leaving open how these objectives will be met. When applying for a building permit, the designers must demonstrate how they meet each objective. This makes it necessary for approving authorities to employ correspondingly qualified personnel so that a productive synergy can be

created between innovative designs and traditional safety concerns. It also requires a high degree of professionalism, because it gives the architect and engineer more leeway, as compared to more prescriptive requirements, and also requires a higher level of building code enforcement review. Seismic isolation, inclusion of damping devices, response history analyses, and displacement-based design are some of the innovative approaches currently in use in some places where this higher level of design and review capacity is present. Each of these represents challenges in analysis and design that require an intimate knowledge of the underlying mechanics-based mathematical theory and its limitations. As such, they are best performed by professionals with considerable experience because there are alternative approaches that require deep insight on the part of the engineer.

Seismic codes begin with the goal of providing safety, and many stop there in most respects, but some include requirements for protecting the functionality of essential buildings, such as fire stations, hospitals, and emergency communications centers or data processing centers. This is discussed in the separate chapter on Essential Facilities and is only mentioned in passing here. Some of the most stringent regulations of this type were passed in California after the 1971 San Fernando Earthquake, when the Hospital Seismic Safety Act of 1972 was passed. The Veterans Administration adopted its own regulations after that earthquake, regulations that seek to not provide not only safe hospitals but also more functional ones. Some voluntary above-code ("performance-based design") approaches lead an owner to invest in the cost of greater seismic protection to achieve less property damage in earthquakes, but the most common seismic design criteria that go beyond the goal of providing safety are related to protecting essential functions.

Geologic and Geotechnical Topics

Mentioned briefly here are geologic and geotechnical engineering aspects to the seismic

provisions of a building code. Construction regulations that deal effectively with earthquake vulnerabilities must include within their purview geologic hazards such as surface faulting and liquefaction. Geologic and geotechnical engineering provisions to deal with those hazards came later than structural provisions. In 1972, the state of California passed the Alquist-Priolo Fault Zoning Act (the current name, which has changed over the years), and slightly before that the city of Portola Valley, which is bisected by the San Andreas Fault, passed its own legislation. Surface fault zoning is the most clearly delineated of the earthquake ground failure hazards, and thus it was logically the first to be subject to regulations. Liquefaction was not a word that soils engineers (then the term for geotechnical engineers) used prior to the 1964 Alaska and the closely following 1964 Niigata Earthquake, but in the following decades, areas of a jurisdiction that were suspect in terms of poorly compacted granular soils and high water table were frequently zoned as requiring special geotechnical investigations. Geotechnical engineers now have at their disposal field and laboratory investigation techniques and a knowledge base about how soils behave in earthquakes that did not exist in the 1960s. Codes and standards of practice can now require geotechnical evaluations that would have been impossible to perform a few decades ago.

One tool that developed out of non-seismic slope stability concerns that later was seen to have a seismic benefit is the grading ordinance. The city of Los Angeles passed such a law in 1952 controlling how excavation and cut and fill could be performed. Such ordinances have become common, and the geotechnical engineering evaluations required in the design stage now typically include earthquake considerations in regions where a seismic code is enforced as well as non-seismic landslide hazards.

Nonstructural Code Provisions

Coming later than structural regulations were provisions to protect the nonstructural features of a building. The first detailed report on

earthquake nonstructural damage was done for the 1964 Alaska Earthquake (Ayres et al. 1973), with a similar report on the 1971 San Fernando Earthquake (Ayres and Sun 1973). Nonstructural provisions are only briefly mentioned here but are an essential part of the modern seismic building code. Approximately three-quarters of the initial value of a modern building's construction, dwellings excepted, is composed of nonstructural components, such as enclosure systems, elevators, partitions, fire sprinklers, ceilings, and heating-ventilating-air-conditioning equipment. In general it can be said that the leading seismic codes of the world have increasingly included more and more detailed regulations for the calculations of loads that are used to design nonstructural bracing and anchorage, but a lack of coordination of the implementation of these rules still persists. Different design professionals work on their own systems – architects on partitions and ceilings, fire protection engineer on fire sprinkler system, mechanical engineer on the HVAC system, and so on – and yet it is the structural engineer with the most expertise for such work. And then different contractors and trades work on the implementation phase, often with little guidance or oversight. Recent earthquakes have demonstrated that simply having nonstructural damage protection rules on the building code's books is not sufficient to actually protect that myriad of components. A common example is that some codes for many years have contained a clause stating that nonstructural components must be designed for the motions and deformations that the structure imposes upon them, but in actual practice, partitions and other components were constructed without this engineering.

Historic Context

Some accounts of historic building code evolution claim that the Code of Hammurabi (a Babylonian king of some 3,800 years ago who was considered by his subjects to be a god as well) engraved on stone is the first building code ever to have been put into effect. That notion needs to be dispelled because the Code

of Hammurabi is much more of a civil or penal code as it has only a few articles out of several hundreds that even mention buildings. It has harsh penalties for builders of shoddy buildings, but the rest of the provisions are divine orders that deal with the trade of commodities, sale of slaves, marriage unions, inheritance regulations, and the like. King Hammurabi informs a builder that if the house he builds should collapse and kill the person who owns it, then he too stands to lose his life, but there is no explanation for how one might build a house that does not collapse. Thus, it fails the test of belonging to the family of codes because it has no stipulations for how things should be done; only penalties are listed that must have been intended to discourage negligence of good building practices.

Tobriner (1984a) provides a historical survey of the development of building codes, noting that they originated primarily in the form of fire-resistant construction requirements, and then only much later in most countries had seismic regulations added to them. The first systematic national building standard was the London Building Act of 1844. Among the provisions, builders were required to give the district surveyor 2 days' notice before building, and they contained regulations regarding the thickness of walls, height of rooms, the materials used in repairs, and the dividing of existing buildings. The placing and design of chimneys, fireplaces, and drains were to be enforced, and streets had to be built to minimum requirements. The city of Baltimore passed its first building code in 1859. The Great Baltimore Fire occurred in February, 1904. Subsequent changes were made that matched the contemporary fire-resistant regulations of some other large American cities. In Paris, under the reconstruction of much of the city by Baron Haussmann during the Second Empire (1852–70), great blocks of apartments were erected, and the height of buildings was limited by law to six stories. Though height limits were instituted for city planning reasons, they later sometimes became part of seismic codes in determining allowable structural systems for various heights.

Construction Regulations Preceded Engineering Regulations

A historic survey of prescriptive requirements on how to build for better seismic performance shows that these landmarks were always reached after cataclysmic experiences. These early regulations, such as were passed after the 1509 Istanbul Earthquake, 1755 Lisbon Earthquake, or the 1880 Luzon (Manila) Earthquakes, were limited to descriptions of allowable building materials and building heights. They predated the extensive use of reinforced concrete and steel. Unreinforced masonry was an allowable material in these early regulations, and no quantitative method was provided for calculating either demands or capacities. Thus, they are primitive seismic codes in today's light, regulating some features of construction but not employing the quantitative methods civil engineers would later develop. That notwithstanding, these efforts are impressive for making such early attempts to improve the building stock's earthquake resistance.

The major earthquake that shook Istanbul in the summer of 1509 led to the banning of stone masonry construction in the city, because most deaths had occurred when such buildings collapsed. This decision paved the way for the emergence of another form of disaster that plagued the city for the next four centuries: fires consumed not only timber construction but also much of the cultural heritage. Following the 1755 earthquake in Lisbon, which destroyed the city center area known as Baixa, the Marquis of Pombal gathered a group of builders to determine the best manner of earthquake-resistant construction to use for the rebuilding. The type of construction selected became known as the Pombalino wall. In its complete form, it is also referred to as "gaiola" or "cage" construction. Most, if not all, of the buildings reconstructed in the reconfigured planned Baixa area were constructed with Pombalino walls and sometimes (but not always) with complete "gaiola" timber frames (Tobriner 1984b; Langenbach 2003).

The Pombalino system was not based on what today would be known as rational methods of

analysis, applying engineering theory to a given case and calculating loads and resistances. It was an attempt at earthquake-resistant construction, not earthquake engineering per se, because the engineering techniques of the nineteenth and twentieth centuries were far off in the future. It was decreed in royal fashion for use in a city that had experienced one of the most destructive earthquakes in Europe and was based on the construction tradition in the Mediterranean basin and may have borrowed construction lore developed by shipbuilders. It was used on the interior of buildings that consisted of timber frames with vertical and horizontal timbers of approximately 10–12 cm square, with internal braces, forming an “X,” referred to in Italy and Portugal as the “Cross of St. Andrew.” The timbers for the cross were 9 cm by 11 cm in section. The frame was then “nogged” (i.e., filled with brick) in the triangular spaces formed by the crosses with a mixture of stone rubble, broken brick, and square pieces of Roman brick in different patterns in each panel. The interior walls were then covered with plaster, hiding the infill and the timber frame. The exterior facades of the Baixa buildings were reconstructed with load-bearing masonry walls of about 60 cm in thickness, some of which had a timber frame on the inside face (Gülkan and Langenbach 2004).

The 1906 San Francisco Earthquake was something of a nonevent in the history of seismic codes. Although the cause of the earthquake was accurately noted – the San Andreas Fault had been mapped in several areas before the earthquake and its surface rupture was well documented afterward – and although civil engineers trained in universities were in existence, no engineering consensus existed around provisions that would define how to calculate loads and then calculate the lateral resistance of a structure. Unreinforced masonry was not prohibited, and by the time the city took stock of its unreinforced masonry building seismic hazards in the 1980s, over 90 % of the 2,000 buildings that required retrofitting were dangers that were built after the earthquake, not before it. The San Francisco building code even reduced the wind load a few years after the earthquake. In that era, the wind

load was sometimes thought to be a surrogate for seismic load.

A few years later, after the 1910 Cartago Earthquake in Costa Rica, a national code was passed that phased out adobe and trapia (rammed earth) structures, substituting bahareque construction (sawn lumber or bamboo framing with plaster). This was a similar but more regulatory approach to the substitution of wood for masonry that occurred in construction styles after some other earthquakes, such as the 1855 Wairarapa (Wellington, New Zealand) Earthquake.

The First Seismic Codes with Engineering Content

Soon after the 1906 California Earthquake, the 1908 Messina-Reggio Earthquake occurred in southern Italy, but it had a different and more beneficial impact on building codes. Sorrentino (2007, 2011) relates how a committee of engineers developed an equivalent lateral force analysis method that was adopted into the building code of the disaster region. It even had the refinement of applying a different lateral force coefficient (*rapport sismica*) to the ground story (1/12) than the story above (1/8), accounting for greater acceleration in the second or third stories. After the 1915 Avezzano (or L’Aquila) Earthquake, the code was changed to have factors of 1/8 for the ground story and 1/6 for the upper stories. These Italian seismic provisions applied only to the area affected by the earthquakes, a pattern that was common in early codes prior to the availability of national maps depicting the hazard of seismic ground shaking.

At the turn of the nineteenth to twentieth century, Riki Sano was getting his PhD from the University of Tokyo on “Seismic Design Concept for Building Structures,” along with Tachu Naito, his student; Sano was to be instrumental in the development of the Japanese equivalent lateral force method. Their work came to fruition after the 1923 Great Kanto or Tokyo Earthquake of 1923, when the 1924 Urban Building Law Enforcement Regulations were passed. It used a 10 % lateral force factor (*shindo*) applied

uniformly up the height of the building. As an example of how engineering practice in some cases is in advance of the codes, when the 1923 earthquake occurred, three large buildings designed by Tachu Naito had already been built (and performed well in the earthquake, giving the new equivalent lateral force method a boost into the code).

An important threshold was crossed in Japan toward the implementation of scientifically determined national seismic provisions (Otani 2008). The objectives of the Japanese Building Standard Law, proclaimed in May 1950, were “to safeguard the life, health, and properties of people by providing minimum standards concerning the site, structure, equipment, and use of buildings.” The law outlined the basic requirements, and the technical details were specified in the Building Standard Law Enforcement Order (Cabinet Order) and in a series of Notifications by the Minister of Construction.

The seismic design provisions of the Building Standard Law Enforcement Order were significantly revised in 1981, including the following three major changes:

1. Structural calculations are required to examine (a) maximum story drift under design earthquake forces, (b) lateral stiffness distribution along the height, (c) eccentricity of mass and stiffness in plan, and (d) story shear-resisting capacity at the formation of a collapse mechanism.
2. Earthquake resistance is specified (a) in terms of story shear rather than horizontal floor forces, (b) as a function of fundamental period of a building and soil type, and (c) separately for the allowable stress design and the examination of story shear-resisting capacity.
3. Required story shear-resisting capacity is varied for construction materials and with the deformation capacity of hinging members under earthquake forces.

The framework of the law was also significantly revised in 1998, (a) introducing performance-based regulations wherever feasible, (b) allowing private agencies to execute the

building confirmation and construction inspection works, (c) deregulating urban land use, and (d) allowing public survey of design and inspection documents. New technical specifications in the form of the Law Enforcement Order and a series of Notifications of the Minister of Construction were issued in June 2000, including the definition of performance objectives at design limit states and the specifications for verification methods.

California’s contribution to the development of the modern seismic code based on engineering calculations had a small start after the 1925 Santa Barbara Earthquake, and then after the 1933 Long Beach Earthquake, statewide regulations were passed that were derived from the Japanese code. The 1930s was the decade when the lineage of seismic codes in several countries began. Provisional earthquake regulations were enacted in Chile in 1930 after the Talca Earthquake and became more institutionalized after the 1939 Chillán Earthquake. In New Zealand, following the 1931 Hawke’s Bay Earthquake, the 1935 New Zealand Standard Model Building By-Law included earthquake regulations. India adopted its first seismic code in 1935 following destructive earthquakes in 1931 and 1935 in Quetta (now part of Pakistan), and the 1939 Erzincan Earthquake in Turkey led to its first engineering regulations for the earthquake hazard.

Equivalent Static Elastic Lateral Force Method

A seismic load on a building is actually dozens of different significant seismic loads that occur during the 20–60 s during which the ground is strongly shaking. The earliest practical way to “put a number” on that bewildering set of loads, which are not well known in advance of the earthquake, was to represent the worst loading effect as a percentage of the weight of the structure, recognizing that seismic shaking causes inertial loads, and inertia is a product of mass and acceleration. The full descriptive name of the method is the equivalent static elastic seismic lateral force analysis method. It is intended to be

equivalent to or adequately represent the actual earthquake forces; the method computes a single static force to represent the changing dynamic forces; that force is applied to an analytical structural model that remains elastic, even though inelastic behavior obviously occurs; and it is an analysis method for determining only the design loads, not for distributing them through the structural components and connections that provide the resistance to the loads. And finally, the only design loads were lateral or horizontal, whereas strong motion records routinely measure some vertical motions as well. Each of these limitations was to be worked on in the following decades. From these first incarnations, the equivalent lateral force method was refined to include several necessary considerations.

Instead of applying seismic regulations to a region after it had an earthquake, earth scientists compiled maps showing where earthquakes should be expected. This was first done with a small number of large-scale zones for a country, each zone defining the force factor to be applied. Then beginning in the 1950s in Japan, a probabilistic basis to the maps was developed. Kawasumi (1951) developed three maps, depicting the shaking that should be expected to occur in exposure periods of 75, 100, and 200 years. The Applied Technology Council in its ATC-3 document (Applied Technology Council, 1978) included national maps produced by S.T. Algermissen that had a probabilistic basis. In China in 1977 a national seismic design map was based on shaking with 3 % and 10 % probabilities of being surpassed. Formats such as these for tying the probability of occurrence to the severity of shaking have become standard in seismic codes today. Another refinement has been microzonation: zoning small areas as having different ground shaking severities, or ground failure hazard (e.g., liquefaction), based on local soils. It became increasingly known that in general soft soils amplified ground shaking, though the large differential response between soft and hard grounds as measured in low-level shaking was not found to be linear up through high levels of ground motion: a column of soft soil 30 m high simply cannot move back and forth rapidly

enough to track rapid and severe motion of the underlying strata, and soil can behave nonlinearly just as structural materials do. The limiting case of microzonation is the site-specific evaluation, where the construction to be placed on a site is designed for a customized set of earthquake criteria, a more expensive approach usually reserved for major facilities.

On the structural side, knowledge of different structural systems and materials had to evolve for codes to seismically regulate them, and a growing body of structural laboratory research provided that information. The steel and concrete industries in particular were the settings for a large amount of research conducted in the 1950s and later. Research on reinforced concrete in great part conducted at the University of Illinois was the technical basis for an important document that advanced the use of that material in seismic codes (Blume et al. 1961). The other kind of research that provided this knowledge was actual earthquake performance. Earthquake reconnaissance or field surveys of the effects of earthquakes became increasingly common in the 1960s.

The early codes provided loads computed with a slide rule that were low enough to keep the structural model that was on the engineer's drawing board in the elastic range. (Computers did not become widespread in seismic design or other civil engineering practice until the 1970s.) Slowly, inelastic behavior was included in codes by including requirements for ductility. Classifying structural systems and materials was an approximate first approach. A structure thought to have greater ductility was designed for lower elastic-level forces, because there was confidence in how it would perform when pushed into the inelastic range. The ability of ductile connections and members to absorb punishment without failure, and also the softening effect (period lengthening) and its beneficial effect on response, became increasingly recognized in the 1960s, 1970s, and 1980s. Today, the choice of a structural system is heavily influenced by how the code defines its ductility. Seismic structural codes are all aware that ground motions are random and can be determined only approximately

in advance. For this reason they contain provisions to prevent unexpected surprises leading to catastrophic failures. Lessons that have been taught by past earthquakes for how building components behave under ground shaking are embedded in their verbiage. As theory and experiment combine to lead to more refined ways of seismic protection, this knowledge must be incorporated into the newer versions of codes for continuous improvement.

The Response Spectrum Method

Yet another improvement in seismic codes was consideration of the dynamic properties of the structure, its period or periods of vibration, and the amount of damping. The tendency of the structure to respond at its natural frequency was only useful information if the frequency content of the ground shaking was also known: in structural dynamics, it takes two to tango. Because the first strong motion seismograph (accelerograph) was only deployed in 1932, it took years for records to accumulate to provide a statistical picture of earthquake severity in terms of frequency content. Trifunac (2006) traces the history of the response spectrum method of analyzing earthquake ground motions and producing design spectra that engineers could use to proportion required resistance of the structure. Two of the most influential developers of the method were Maurice Biot and George Housner, whose doctoral theses at the California Institute of Technology (1932 and 1941, respectively) were devoted to this topic. The theory became widely applied in engineering practice in the 1970s and 1980s when several conditions were favorable. First, there were more earthquake records. Second, engineers could afford to have the new, more powerful electronic computers on their desks that could do the extensive mathematical work. Housner (1997, p. 33) notes a third factor: "Because of the practicing engineers' reluctance to employ the design spectrum, I think it was essentially the nuclear power business that got the spectrum into widespread use."

Concerning the growing number of earthquake records, the city of Los Angeles played an important role when it passed a revision to its building code in 1965 requiring that three strong motion instruments be installed in buildings six stories or taller (one at the base, one at mid-height, one at the roof). This resulted in the large harvest of ground motion and structural motion records from the 1971 San Fernando Earthquake. The city of Los Angeles also was precocious in its 1943 adoption of a formula in its code that related period of vibration as represented by number of stories to the base shear coefficient calculation in its equivalent lateral force method: the taller the building, the lesser the base shear. In 1957, Los Angeles also had a strong effect on the development of earthquake codes, even though it was not an earthquake code revision: the zoning code was revised to allow buildings taller than 150 ft. Engineers in California thought that taller, more flexible (longer period) buildings could be designed for lower force coefficients, and they also knew that if the same coefficients as were used for low-rise buildings (typically 10 % to 13 %) were applied to every story in a tall building, the large amount of strength, and thus area taken up by shear walls and other structural material, would have very negative architectural and real estate implications.

When it seemed that engineers in the two large urban regions of California, centered on Los Angeles in the south and San Francisco in the north, could be diverging toward significantly different provisions for dealing with the tall building question and the modernization of seismic codes, the statewide structural engineering organization Structural Engineers Association of California (SEAOC) began its influential set of editions of the "Blue Book," the *Recommended Lateral Force Requirements and Commentary* (SEAOC 1959). Its requirements became the seismic provisions of the Uniform Building Code used in various editions and adaptations throughout California and the Western United States until the year 2000 when the UBC was folded into the IBC, as previously described.

The Response History Method

The equivalent static lateral force method leads the engineer to calculate a single quantity, the total shear at the base of the building, the base shear, and then proportion that load up the height of the structure according to procedures that attempt to represent the dynamic response of the building in a simple way. While there is dynamic thinking underlying the method, it still represents the series of motions that occur during the earthquake with a static view or “snapshot” of the overall effect of those motions. The response spectrum method similarly is a way to give the engineer a base shear to use in design. A given overall representation of one record, for example, the peak ground acceleration, might exceed that of another record and yet that other record might be more damaging to a particular structure. A record with greater duration, more pulses above a threshold such as the elastic limit of a portion of the structure, can create more demand than a record of smaller duration, even if the latter has a greater peak value (Johnson 2013). The logical improvement was seen to be to calculate the various forces and deformations that occurred to a structural model split second by split second in response to the changing ground motion.

Three sources of research information that developed from the 1960s to the present day provided the basis for incorporating a fully dynamic method into seismic codes, subjecting the analytical model of the structure to the motions of several recorded earthquakes. One came from completely outside the earthquake engineering field, the development of the modern computer. A second prerequisite was a large library of earthquake records, conveniently supplied by growing coverage of highly seismic areas with strong motion instruments and, unfortunately, by Earth’s frequent earthquakes. The third was improved knowledge of structural behavior, in particular of how structures behaved in the inelastic range. While computers played a role in that research on structural behavior, it has also required the traditional approach of subjecting specimens to simulated seismic loading in the laboratory. Computer simulation is not

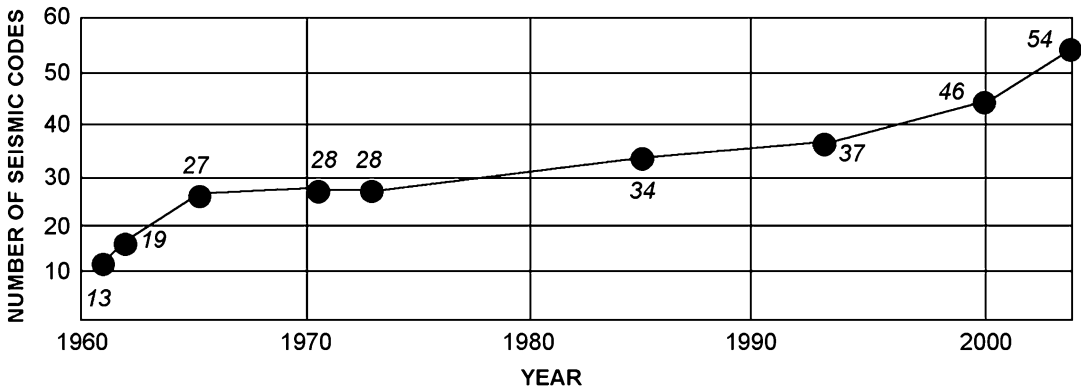
capable of substituting for the simulation of physical testing.

A 1973 benchmark for the development of the response history method is provided by the seismic design manual of the US military (Departments of the Army, the Navy, and the Air Force 1973, pp 3–3): “Since the mechanics of dynamic analysis requires that a separate solution must be obtained for each instant of time during the entire history of interest, computation by computers is necessary. This kind of analysis is generally beyond the kind of effort that can be afforded in the design of almost all but the most critical structures.” Today, the method is more frequently used, though it still is applied to a small minority of all the structures being seismically designed. Chopra (2005, p. 107) provides an assessment of the future of the method:

At the present time, nonlinear RHA [response history analysis] is an onerous task, for several reasons. First, an ensemble of site-specific ground motions compatible with the seismic hazard spectrum for the site must be simulated. Second, despite increasing computing power, inelastic modeling and nonlinear RHA remains computationally demanding, especially for un-symmetric-plan buildings—which require three-dimensional analysis to account for coupling between lateral and torsional motions—subjected to two horizontal components of motion. Third, such analyses must be repeated for several excitations because of the wide variability in demand due to plausible ground motions, and the statistics of response must be considered. Fourth, the structural model must be sophisticated enough to represent a building realistically, especially deterioration in strength at large displacements. Fifth, commercial software is so far not robust enough to predict response with high reliability. Sixth, an independent peer review of the results of nonlinear RHA is required by the FEMA-356 guidelines, adding to the project duration and cost. With additional research and software development, most of the preceding issues should be resolved, and nonlinear RHA may eventually become the dominant method in structural engineering practice.

Growth in Adoption of Seismic Codes

Figure 1 provides a quick way to see how seismic codes have spread from an initially small number of countries. The numbers are drawn from the



Building Codes and Standards, Fig. 1 Growth in the adoption of seismic codes (Reitherman 2012, p. 582)

various editions of *Regulations for Seismic Design: A World List*, published by the International Association for Earthquake Engineering (1960 and later editions).

One could argue that because seismic codes now cover most of the significantly seismic areas of the world, seismic safety has now been achieved on a global scale. This would be overly optimistic for two reasons. First, there are always older, pre-code buildings that did not benefit from a modern seismic code. Secondly, even today in countries with regulations “on the books,” legally binding laws that include the latest earthquake engineering thinking, those regulations are not always thoroughly carried out.

The seismic provisions in a code must be complemented with reliable implementation of the code, its effective enforcement. This remains one of the hardest issues to solve, especially in poorer countries but also in the most developed. Merely making seismic codes more sophisticated as each new edition is promulgated every few years is not a solution, unless comprehensive implementation also occurs. This includes the education and professional training of the engineers and other design professionals involved as well as the thoroughness of the quality control measures of building regulation agencies, such as in the plan review and construction inspection phases. In some instances, the comprehensive protection that the provisions in a building code seem to provide exists mostly on paper. Up until the 1990s, even the most advanced seismic codes

listed only a few specific nonstructural components that needed anchorage and then added a general phrase such as “all other equipment and machinery.” That sounds comprehensive, but in the absence of a well-defined definition of that all-encompassing phrase, it was not particularly meaningful. Another example is the accurate calculation of interstory drift, which in some codes was seemingly covered many decades ago but in practice was not well-implemented in standard practice. The engineer who diligently tried to meet that code requirement had inadequate analytical tools available because of the lack of underlying knowledge of the displacements that would actually occur. Conversely, some engineers incorporated design measures that were only required by the code much later. In the words of George Housner (1986, p. 25), “in some instances, earthquake requirements were adopted in building codes but were not used by architects and engineers. And in other instances earthquake design was done by some engineers before seismic requirements were put in the code.”

Summary

Model building codes are common around the world, although the consistency with which they are applied throughout a jurisdiction can vary greatly. In New Zealand and Japan, there is a high degree of consistency in how a nationally

standardized set of seismic provisions are adopted and enforced at the local level. In India and the United States, there is greater variation. Seismic provisions in model building codes are usually a combination of prescriptive and performance-based procedures. Seismic provisions in building codes evolved from the earlier building codes that were typically motivated by concern over fire resistance and urban conflagrations. The first seismic regulations governed construction characteristics such as wall material or number of stories, but because they preceded civil engineering developments of the 1800s and 1900s, they lacked quantitative methods for calculating loads and capacities. In the early 1900s, the first seismic codes based on engineering methods were developed in Japan and Italy. The equivalent static elastic lateral force method originated in these codes. By the 1930s, seismic regulations based on the equivalent lateral force method were adopted in New Zealand, the United States, Turkey, India, and Chile. Later, as earthquake engineering research accumulated, as strong motion records were collected, and as powerful and inexpensive computers became available, the response spectrum and the response history methods were added to the modern building code's seismic provisions. A fundamental problem in reducing earthquake risks through adoption of up-to-date building code provisions is enforcement and implementation. While the problem is most prominent in poorer countries, it is also an issue in the most technologically advanced areas of the world. It is to be expected that displacement-based procedures will make further inroads into routine structural system analyses. Also, as seismic isolation technology becomes more mature, it will be accepted by more owners on account of observed benefits.

Cross-References

- ▶ [Analytic Fragility and Limit States \[P\(EDPI IM\)\]: Nonlinear Dynamic Procedures](#)
- ▶ [Analytic Fragility and Limit States \[P\(EDPI IM\)\]: Nonlinear Static Procedures](#)

- ▶ [Assessment of Existing Structures Using Response History Analysis](#)
- ▶ [Behavior Factor and Ductility](#)
- ▶ [Code-Based Design: Seismic Isolation of Buildings](#)
- ▶ [European Structural Design Codes: Seismic Actions](#)
- ▶ [Global View of Seismic Code and Building Practice Factors](#)
- ▶ [Learning from Earthquake Disasters](#)
- ▶ [Masonry Modeling](#)
- ▶ [Masonry Structures: Overview](#)
- ▶ [Nonlinear Dynamic Seismic Analysis](#)
- ▶ [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- ▶ [Probabilistic Seismic Hazard Models](#)
- ▶ [Reinforced Concrete Structures in Earthquake-Resistant Construction](#)
- ▶ [Response Spectrum Analysis of Structures Subjected to Seismic Actions](#)
- ▶ [School Seismic Safety and Risk Mitigation](#)
- ▶ [Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling](#)
- ▶ [Seismic Analysis of Steel Buildings: Numerical Modeling](#)
- ▶ [Seismic Risk Assessment, Cascading Effects](#)
- ▶ [Seismic Strengthening Strategies for Existing \(Code-Deficient\) Ordinary Structures](#)
- ▶ [Steel Structures](#)
- ▶ [Timber Structures](#)

References

- Ayres JM, Sun TY (1973) Nonstructural damage. In: The San Fernando, California earthquake of February 9, 1971. National Oceanic and Atmospheric Administration, Washington, DC
- Ayres JM, Sun TY, Brown FR (1973) Nonstructural damage to buildings. In: The great Alaska earthquake of 1964: engineering. National Academy of Sciences, Washington, DC
- Blume JA, Newmark N, Corning LH (1961) Design of multistory reinforced concrete buildings for earthquake motions, Portland Cement Association, Chicago
- Chopra A (2005) Earthquake dynamics of structures: a primer. 2nd ed. Earthquake Engineering Research Institute, Oakland
- Departments of the Army, the Navy, and the Air Force (1973) Seismic design for buildings. Departments of the Army, Navy, and Air Force, Washington, DC

- Gülkan P, Langenbach R (2004) The earthquake resistance of traditional timber and masonry dwellings in Turkey. In: Proceedings, 13th World conference on earthquake engineering, Vancouver, Canada, (Paper no. 2297)
- Housner GW (1986) Historical view of earthquake engineering. In: Proceedings, 8th World conference on earthquake engineering, San Francisco. (The complete set of proceedings for all World Conferences on Earthquake Engineering is freely available at: <http://www.nicee.org/wcee/>)
- Housner GW (1997) George W. Housner, Connections: EERI Oral History Series. Stanley Scott, interviewer. Earthquake Engineering Research Institute, Oakland
- International Association for Earthquake Engineering (IAEE) (1960a, 1963, 1966, 1973, 1976, 1980a, 1980b, 1992, 1996, 2000, 2004). Regulations for seismic design: a world list. International Association for Earthquake Engineering, Tokyo
- Johnson J (2013) Energy methods: what does the future hold? *Structure* 2013:14–15
- Kawasumi H (1951) Measures of earthquake danger and expectancy of maximum intensity throughout Japan as inferred from the seismic activity in historical times. *Bulletin of the Earthquake Research Institute, University of Tokyo*, 21:469–481
- Langenbach R (2003) Crosswalls instead of Shearwalls: a proposed research project for the retrofit of vulnerable reinforced concrete buildings in earthquake areas based on the traditional “hımsı” construction. In: Proceedings of the fifth [Turkish] national conference on earthquake engineering, Istanbul. Available on www.conservationtech.com
- Otani S (2008) The dawn of structural earthquake engineering in Japan. In: Proceedings, 14th world conference on earthquake engineering, Beijing
- Reitherman R (2012) Earthquakes and engineers: an international history. ASCE Press, Reston
- SEAOC (1959) Recommended lateral force requirements. Structural Engineers Association of California, Sacramento
- Sorrentino L (2007) The early entrance of dynamics in earthquake engineering: Arturo Danusso's contribution. *ISCT J Earthq Technol* 44(1):1–24
- Sorrentino L (2011) Engineering solutions after the Messina 1908 earthquake. In: *Engineering history and heritage*. Institution of Civil Engineers, London
- Tobriner S (1984a) The history of building codes to the 1920s. In: Proceedings, SEAOC annual convention. Structural Engineers Association of California, Sacramento
- Tobriner S (1984b) A history of reinforced masonry construction designed to resist earthquakes: 1755–1907. *Earthq Spectra* 1(1):125–149
- Trifunac MD (2006) Biot response spectrum. *Soil Dyn Earthq Eng* 26(2006):491–500

Building Damage from Multi-resolution, Object-Based, Classification Techniques

Marco Chini

Luxembourg Institute of Science and Technology (LIST), Environmental Research and Innovation Department (ERIN), Belvaux, Luxembourg

Synonyms

Earthquake damage mapping; Object-oriented classification; Very high-resolution SAR and optical images

Introduction

During a seismic event a fast and reliable damage map of the urban areas involved is very important to manage civil protection interventions. Satellite remote sensing data can provide valuable pieces of information in this respect, thanks to their capability to have an instantaneous synoptic view of the scene especially if the seismic event is located in remote regions or the main communication systems are damaged. The major limitation that could hamper the operational use of this source of information is the speed of data collection just after a disastrous event. This is a key point for civil protections that need a fast overview of the epicentral area, quick information on to the extension and distribution of damages and the evaluation of infrastructure conditions such as roads, bridges, focal and crucial buildings and so on. A single satellite can provide access time to a specific site in the order of days as a result the necessity to use any kind of available satellite image and making mandatory data integration approaches to increase the chance of collecting information in near real time (Chini 2009).

Since the end of the twentieth century satellite systems and image-analysis techniques have progressed a lot in the aspects where they can contribute significantly to support the

management of major natural disasters, as well as humanitarian crisis situations. Compared to 15 years ago the availability of satellite imagery, amount, timeliness and the capability to cover a certain crisis situation or disaster event with high degree of detail have all improved substantially. There are several factors which have led to this. First of all ground pixel spacing of civil Earth Observation (EO) systems has developed to the meter domain (or less) for optical and radar systems, such as synthetic aperture radar (SAR) (Voigt et al. 2007). Indeed on September 24 1999 the commercial EO satellite IKONOS was launched, the first able to collect openly available high-resolution imageries that were both panchromatic and multispectral with a spatial resolution of 1 and 4 m, respectively. Two years later, on October 18, 2001, another important step forward was made with the launch of one more high-resolution commercial EO satellite, QuickBird, with an increased spatial resolution up to 0.6 m for the panchromatic sensor and 2.4 m for the multispectral one. The first earthquake where this kind of data proved its usefulness in conjunction with automatic image-analysis techniques was the one that struck the city of Bam (Iran) in 2003, provoking a lot of damage to man-made structures and causing the death of many people (Yamazaki et al. 2005; Gusella et al. 2005; Stramondo et al. 2006; Gamba et al. 2007; Chini et al. 2008b; Chini et al. 2009; Bignami et al. 2011). Concerning SAR satellite sensors, the spatial resolution breakthrough arrived in 2007 with the launch of the first two satellites of the two X-band satellite constellations, COSMO-SkyMed and TerraSAR-X, on June 8 and on June 15, 2007, respectively, both having the capability to acquire images in the so-called SpotLight image mode with a spatial resolution of 1 m. TerraSAR-X mission is composed of two satellites, the second one was launched on June 21, 2010, while COSMO-SkyMed has four satellites, the fourth one launched on November 5, 2010. The fact that they are constellations has reduced the site access time down to two and a half days with TerraSAR-X and to 12 h with COSMO-SkyMed. This latter issue is the second

important factor that makes satellite data able to face and provide useful information during the crises phase soon after an earthquake.

Both optical and SAR data have characteristics to be used as a source of information to provide a damage map after an earthquake (Yonezawa and Takeuchi 2001; Matsuoka and Yamazaki 2004; Chini et al. 2008a; Chini et al. 2013), with particular attention to very high-resolution data, which has the potential to monitor wide areas, but at the same time giving information on single structures present in the scene (Stramondo et al. 2008; Turker and Sumer 2008; Brunner et al. 2010; Chini et al. 2011; Dell'Acqua et al. 2011; Ferro et al. 2013). The huge amount of information carried by these kinds of data can hamper the visual inspection, which is time consuming during a crisis phase, fostering the implementation of more automatic approaches.

Methodology

The automatic or semiautomatic classification of remote sensing data is a tool to help understand and monitor a large variety of scenarios, converting them into tangible data, which can be utilized in conjunction with other data sources to provide crucial and accurate information. Even if metric or submetric resolution images provide new opportunities in the damage mapping field, new problems arise at the same time in automatic classification procedures. Indeed, the aim of image classification analysis is to extract information and display the results in a simple and effective way for possible further studies. In metric resolution images a single pixel represents a small area ($0.5 \div 10 \text{ m}^2$) of the acquired scene and is usually smaller than the size of the objects of interest. Hence the information of a single pixel is strongly linked to the information carried by its neighbors, which are part of the same object, pointing out the necessity to perform an object-oriented analysis, although the measureable statistics typically exploited for classification purposes are necessarily pixel

based (Blaschke 2010). This aspect is very interesting, because it makes possible to focus on a single object, for example, a building or also other man-made structure, having the advantage of a valuable amount of pixels to characterize the object itself.

Object-based techniques recognize that important semantic information is not always represented in single pixels but in meaningful image objects and in their contextual relations. For example, it is more likely that a pixel adjacent to a private garden should be classified as a house, road, or private garden rather than as a forest. Usually, object-based classification comprises two steps: the image segmentation and the object classification. Image segmentation subdivides the image into groups of contiguous pixels called objects or segments that correspond to meaningful features or targets in the field (Blaschke 2010). The images are segmented into homogeneous objects based on the spectral information and local patterns or textural information that are included in groups of neighboring pixels. As a result, the object-based classifications can consider a wide range of variables such as mean reflectance, texture, shape and size of objects and can potentially produce more accurate and detailed maps than conventional classification strategies (Mathieu et al. 2007; Blaschke 2010; Chini et al. 2014).

All algorithms exploiting metric resolution data for damage mapping purposes have to deal with object-oriented approaches even if the final target is not the object itself, but maybe just a classification useful for restricting the analysis and the statistics to classes we are interested in. Such an example could be a change detection approach to identify a collapsed building using one pre-earthquake image and one post-earthquake image by a threshold approach. At this resolution many regions will be affected by changes even those not relevant for damage purposes. Thus to fix the threshold used to identify damaged buildings (pixels that have changed their radiometric values because of the earthquake) we should restrict the analysis to objects of interest such as buildings (Chini et al. 2009).

It is worth recalling the optical and SAR image properties that are usually used for damage detection purposes, highlighting characteristics that are relevant to damage mapping analysis such as the physical properties of the measured signal and the spatial resolution that can have an impact on the final products. Basically, the process for forming a SAR image is different from that used for an optical image for two main aspects: the physical properties of the imaged targets that contribute to the measured signal and the signal processing steps used to create the image. Optical sensors measure the radiometric properties of reflected sunlight in spectrally distinct regions of the visual and near-infrared spectrum or integrated in a single panchromatic band. The material properties of objects, the illumination conditions of the scene and the sensor perspective determine the radiometric and geometric appearance of distinct targets in optical imagery. The SAR system is an active sensor and measures the backscatter of a transmitted signal, typically in a narrow microwave frequency band and sampled in the range direction, i.e., the line of sight (LOS) from SAR sensor to target. Backscattering is primarily determined by the geometry and dielectric properties of an object and the transmit/receive configuration of the SAR sensor (Brunner et al. 2010).

Concerning optical images it is easier for the reader to grasp how different spatial resolutions make buildings appear and more in general, cities, where if the resolution is higher, the amount of detail composing land use classes of interest will be greater. Indeed, in high-resolution optical images the radiance of each pixel is not directly related nor can be considered representative of the class it belongs to. However, many studies are able to demonstrate that the advantage of being able to aggregate pixels to an object and then address object characteristics, is that it can discern between different kinds of buildings and different kinds of asphalted areas such as roads or parking lots, turning all of this augmented complexity of understanding into an information luxury (Pacifi et al. 2009). Such an example could be the shadow from a building, which has at this resolution a clear signature in the image,

giving us information about the building height, which could be also an important source of information for damage detection purposes. As a matter of fact, Turker and Sumer (2008) detected damaged buildings from the big earthquake in Turkey in 1999, applying a watershed segmentation on a post-event optical image utilizing the relationship between the buildings and their shadows, labeling 80 % of buildings correctly as either damaged or undamaged. Another example of how object-oriented approaches are useful is presented in Gusella et al. (2005) that quantified the number of collapsed buildings of the Bam earthquake in 2003, starting from the inventory of buildings as objects in QuickBird satellite imagery taken before the event.

Pertaining to SAR data an in depth description of how a building appears is necessary. For this purpose, it is useful to describe how radar returns from different parts of the building are mapped on the SAR image due to the fact that they reach the receiver at different times (Kropatsch and Strobl 1990; Franceschetti et al. 2002; Franceschetti et al. 2003; Guida et al. 2008; Guida et al. 2010). Given an isolated building on ground and proceeding in the SAR image with a constant azimuth from near to far range, we can identify four main singular contributions:

1. A bright stripe corresponding to a mixture between the backscattering from the ground, facade, and roof, which is a typical example of layover in a SAR image. Layover is called “radar mapping,” where different object points having the same time and same range are mapped into one image point, i.e., more than one point on the Earth’s surface are mapped into one image point (many to one mapping). SAR mapping is mainly an integration of reflected signals having the same Doppler frequency (azimuth or along-track measurement) and same distance (range or across-track measurement) (Kropatsch and Strobl 1990).
2. A very bright line corresponding to facade-ground and ground-facade double-bounce scattering, because double-bounce ray paths all have the same length, equal to

the distance between SAR sensor and the facade base (all different scattering points are positioned at a unique distance, corner reflector).

3. The contribution from the roof that could be of two types: (a) flat roof, a unique gray area corresponding to backscattering from the roof, and (b) gamble roof, the roof has a different scattering signature with respect to a flat one. Basically it is different due to the presence of a second bright scattering feature, which is closer to the sensor than to the double-bounce contribution (Brunner et al. 2010). This behavior concerns only the direct backscattering from the part of the roof that is oriented toward the sensor.
4. A dark area corresponding to the SAR building shadow which should be not confused with the shadow concept in an optical image. Shadow in radar imagery is called the region where an object point is not reached by any radar beam. Such object points produce a “zero” signal in the image. Therefore, shadow regions appear as dark areas corrupted by noise (Kropatsch and Strobl 1990)

Of course this is a simplistic schematization of the phenomenon; other parameters also play an important role in the backscattering values of the four building parts such as geometric parameters (height, length, and width of the building and the roughness of the ground) and the electromagnetic parameters (dielectric constants of facade, roof, and ground). The four types of building responses with a distinct backscattering values can be detected as separate objects if the resolution is higher than the dimension of a building itself, while this distinction is not possible if the SAR resolution is lower than the buildings being identified. As a matter of fact an urban area imaged by a SAR sensor with a resolution in the order of tens of meters appears uniformly brighter than other land cover classes since the double-bounce effects are the dominant scattering component in the SAR resolution cell (i.e., the image pixel), that one characterizing the backscattering for such scenario. Whilst in metric resolution SAR sensors such as COSMO-SkyMed and

TerraSAR-X, we still have very bright values in the city, principally double-bounce pixels, but we have gray and dark regions as well, roofs and shadows respectively, that are well identified and with a remarkable spatial extension. On the contrary big uniform surfaces, such as bare soil or asphalted industrial areas, appear the same in both high- and low-resolution SAR since their extension is bigger than both sensor resolutions. Change detection approaches with lower-resolution SAR stem from the concept that the double-bounce effect fails when a building collapses (which is the dominating scattering mechanism in the resolution cell); thus in the comparison between two images, one acquired before and one after the seismic event, a decrease in the backscattering value is expected (Yonezawa and Takeuchi 2001; Matsuoka and Yamazaki 2004; Stramondo et al. 2006). It is a different case with the high-resolution SAR data because we can see a decrease of the backscattering for the pixel where the double-bounce is located, but also an increase in the roof part, since rubble can cause an increase of the backscattering as compared to a smooth roof. Furthermore, where the shadow is present in the pre-event image, some changes are expected in the post-event image, depending on the structures shadowed before the collapse.

As it appears from the aforementioned description of the optical and SAR data, the high resolution has drastically changed the semantic of objects as they were known so far. Indeed the lower resolution caused them to appear differently than the metric or submetric spatial resolution, thus the necessity to introduce new characteristic features and a new vocabulary. For this reason the object-oriented approach seems more promising than per pixel, since with submetric resolution image without having a well-defined semantic for identifying classes of interest we can just identify primitive objects, which are different from objects of interest. Indeed only objects of interest match real-world objects, e.g., the building footprints. Primitive objects are usually the necessary intermediate step before objects of interest can be found by a segmentation and classification process, having

as a smallest image object the pixel. Of course an object can be defined in different ways, depending on our scope and at which scale of detail we want to provide the information. An object can be a building or a neighborhood in a city with a particular type of structure and both of them can give us important information in terms of damage mapping. The following are some examples of applied object-oriented methodologies in real scenario, where very high-resolution optical and SAR data have been used for the seismic events that occurred in Bam (Iran) in 2003 and in the city of L'Aquila (Italy) in 2009.

Case Studies

On December 26, 2003, the southeastern region of Iran was hit by a 6.5 moment magnitude (Mw) earthquake whose epicenter was located very close to the city of Bam. The event had a tremendous impact all over the world, due to a heavy loss of lives, injuries, and destruction; the earthquake caused the death of more than 26,000 residents and injured another 30,000. Additionally, the city of Bam was a UNESCO World Heritage Site and as such it plays an important historical role. The Bam earthquake was the worst seismic event in recent Iranian history. The damage was concentrated in a relatively small area around Bam. Ninety percent of the structures in the city of Bam were left more or less heavily damaged or totally collapsed. In particular, the main historical and traditional adobe buildings collapsed. Contemporary houses demonstrated better resistance to earthquake shake, even though many collapsed due to poor construction practices and the presence of weak stories in the buildings (EERI 2004; Chini et al. 2009). As already highlighted this event was the first one to be imaged by an optical sensor with a submetric spatial resolution (60 cm) such as QuickBird, so for this event we have many papers where the damage mapping was treated making use of these new satellite sensors. Within these, three publications have been selected to show the different scales of products that can be provided and the synergy between the optical and

SAR sensors, but not only and also the impact that the very high resolution can have in the lower-resolution products.

In Chini et al. (2008b) pre- and post-earthquake QuickBird panchromatic images have been used to show the capability of this data to map damage at building scale by means of segmentation approach based on the application of mathematical morphology operators (Pesaresi and Benediktsson 2001; Soille 2003). In this article, a classification procedure was applied to the two panchromatic available data, one taken before (September 30, 2003) and one after (January 3, 2004) the seismic event. The buildings' extraction from one single panchromatic image is not an easy task. Many objects in the imaged scene can be confused because they have very close radiance values: cars with some buildings, shadow with some asphalt roads, soil with some kinds of roofs and so on. Due to the lack of multispectral data with a resolution of 60 cm, a contextual analysis is needed to extract geometric information of objects/classes within the images. Morphological profiles from the original panchromatic images taken before and after the earthquake have been carried out in order to classify all the buildings in the scene. The classification was done so as to avoid false alarms in the change detection process caused by shadow or other temporary objects like cars or recovery tents. For this purpose, the open and close morphological operators have been applied and the derived profiles have been used as inputs to an unsupervised classifier in order to extract the entire map of buildings before and after the earthquake. Hence, after the classification process, each single building has been recognized and then labeled. Finally, the damage estimation can be computed, by comparing the pixel belonging to the same building before and after the seismic event. The percentage of the damaged pixels with respect to the total for each building is an estimation of the damage level (ratio between the number of pixels after and before the earthquake forming the single object). The higher the number of damage pixels, the higher damage level of the buildings. The map deals with a classification criteria where each building of Bam city has

been labeled as belonging to one of three possible damage levels: light or no damage, medium and heavy. Two thresholds have been set in order to identify the three classes. The satellite-based damage map has been compared with a map derived from ground survey, provided by Geological Survey of Iran, which reports the map of collapse ratio. This latter concerns the percentage of completely collapsed buildings with respect to the total number of buildings within a city block. It is worth noting that the information about damage from field survey was given at district scale, while the results from object-oriented approach provided information at building scale. Although the different scales of the maps did not allow a direct comparison, a qualitative accuracy evaluation was done counting the percentage of the three classes from remote sensing data in each regional class obtained from ground survey. It is interesting to highlight that the percentage of heavy damaged buildings (from the object-oriented analysis) increased with the increase of ground survey damage degree.

In Chini et al. (2008b) the semiautomatic classification approach aimed to identify completely or partially collapsed buildings based on the optical satellite images vocabulary; it was basically a change detection approach at building scale, which evaluated how much the radiance value for all pixels of each building changed as an effect of the earthquake. A step forward was attempted in the next article we will describe (Bignami et al. 2011) where an analysis was performed on the sensitivity of textural features at object scale with respect to the damage levels according to the European Macroseismic Scale 1998 (EMS98). Textural features were clearly demonstrated to be a method to overcome the lack of spectral resolution for classification purposes. Indeed texture analysis plays an increasingly important role in digital image processing and interpretation, especially for very high-resolution data (Haralick et al. 1973; Haralick 1979; Pacifici et al. 2009). This is mainly due to supplementary information about image properties it can provide. Such information can be fruitfully exploited for change detection applications as well. In Bignami et al. (2011), the analysis was

focused on the sensitivity of some textural features for change detection purposes; in particular it was focused on earthquake damage mapping. This approach was preferred in order to give information strictly related to each building and to mitigate false alarms weighting per-pixel classification error. Instead of the usual textural features extraction by considering the co-occurrence matrix on a moving window, the proposed method was based on the calculation of textural parameters at object scale. Indeed, the co-occurrence matrix is evaluated by taking into account all the pixels belonging to a single object, i.e., the building, and considering a shift of one pixel in the horizontal and vertical directions. In this way for each building a co-occurrence matrix was obtained. The algorithm makes use of the same dataset as Chini et al. (2008b) and also of some outcomes of the algorithm such as the buildings map, where each building/object was singled out. Stemming from this building map, the textural parameters for each specific object is derived, overcoming the dependency from the window size, which makes the analysis test case dependent. Following the formulation of the textural parameters described in Haralick et al. (1973) and Haralick (1979), for each one of the selected building in the two images, pre- and post-earthquake, four different object textural parameters have been calculated, such as contrast, dissimilarity, homogeneity and entropy. In order to investigate the sensitivity of these parameters to damage degree from EMS98 damage scale, some buildings were selected by the ground truth map described in Hisada and Shibaya (2004), where some buildings, located around eight strong motion stations, were surveyed and classified. The EMS98 classifies five damage levels: G1 (negligible to slight), G2 (moderate), G3 (substantial to heavy), G4 (very heavy) and G5 (destruction). The G1 level can also include the non-damage class, because the difference between G1 and no damaged building cannot be appreciated using this kind of satellite data. Once the four object textural parameters have been derived on both pre-seismic and post-seismic data, the difference between post- and

pre-seismic at object scale has been computed in order to analyze their sensitivity to the damage levels for those building labeled by ground survey. A total of 367 buildings were accounted for in the sensitivity analysis. For each object textural feature the mean value within a damage class from the ground survey was evaluated and the difference between post-seismic and pre-seismic features was calculated. These values have been plotted in respect to the five ground survey damage grade. By observing the plots, some evidence of sensitivity with respect to damage levels are observed. A positive trend in the difference of textural features has been obtained for the contrast, the dissimilarity and the entropy, while a negative trend is visible for the homogeneity. The trends reflect the expected behavior of these textural parameters. In fact, as the damage level of an object increases, the numbers of pixels within the object, which appear scattered, increase. From a textural point of view, the homogeneity of the object should decrease, i.e., the more scattered pixels are, the less homogeneous is the object. On the contrary, the contrast, the dissimilarity, and the entropy increase with the increasing number of scattered pixels. A comparison with textural parameter extracted by using a moving window with a fixed size has also been carried out, highlighting that the trends were not as expected. This work showed that to better understand some discrepancies present between satellite information and ground survey damage grade, an object-by-object analysis is needed. It was stated that the damage level assigned to an object should be validated also from the remote sensing point of view because sometimes the damage level cannot be identified from satellite data. A typical false alarm case is the pancake effect which occurs when severe damage affects a building (G3 and G4 levels), or sometimes it collapses (G5 level), but the roof remains almost undamaged, preventing its identification from space images.

The Bam earthquake was also studied in Chini et al. (2009), where the integration between medium resolution SAR data was analyzed, from 20 m resolution C-band sensor onboard of

Envisat satellite mission and very high-resolution optical data, with the aim of producing damage maps at district scale with three different levels of damage. In this article additional parameters, namely, the InSAR complex coherence and the intensity (or backscattering) correlation, have been computed, since couples of pre- and post-seismic interferometric images were available (two images before and one or two after the event). Both parameters can be derived by combining the pre-seismic pair, the post-seismic, and the coseismic (i.e., one pre-seismic and one post-seismic image) ones. Note that these two features contain slightly different information concerning changes in the scene. The complex coherence is primarily influenced by the phase difference between radar returns, which is a distinctive parameter measured by a coherent sensor like SAR. It is particularly related to the spatial arrangement of the scatterers within the pixel and, thus, to their possible displacements. Conversely, the intensity correlation is more related to changes in the magnitude of the radar return. Unfortunately, for this event, SAR interferometric image pairs have very high values of the perpendicular baseline, around 500 m for both pairs and the resulting high spatial decorrelation prevents their employment from detecting damage levels through the InSAR phase coherence, which has almost the same values both in damaged and undamaged areas when the baseline is too large (Zebker and Villasenor 1992). Concerning the use of the backscattering correlation as an indicator of surface changes, the approach is founded on the idea that the correlation coefficient is high when two images are similar, while it becomes low when changes have occurred in the surface target. Thus, in the latter case, a decrease of correlation of coseismic images is expected to be relative to pre-seismic ones and this phenomenon is expected to increase with the increase of damaged buildings on the ground. The difference between the pre- and coseismic intensity correlations has been averaged within areas of homogeneous damage levels provided from a ground survey map, and it is clearly shown how this difference increases

with the increase of damage level (Yonezawa and Takeuchi 2001; Stramondo et al. 2006). In addition to Envisat SAR data, the information extracted from a high-resolution optical QuickBird image has been used, by extracting the building maps before the earthquake carried out with the adoption of object classification approaches. The building map was used as a mask to restrict the SAR intensity correlation change detection only to built-up areas, in order to reduce possible false alarms. Indeed, in Bam City, there were a lot of vegetated areas, which could affect the value of the correlation. The intensity correlation difference showed an increase of the dynamic range of about 25 % with respect to the non-masked data, using the information on buildings from optical data.

On April 6, 2009, a $M_w = 6.3$ earthquake hit a wide portion of the Abruzzi region in Italy, with the epicenter located 4 km southwest of the city of L'Aquila, whereas the highest damages occurred about 8 km to the southeast. In L'Aquila and in the neighboring villages, the earthquake caused the collapse or irreparable damage to over 15,000 buildings, killing 308 people and causing the relocation of over 65,000 (Stramondo et al. 2011). For this event very high-resolution SAR images, acquired in SpotLight mode (1 m per pixel on ground) from the X-band COSMO-SkyMed satellite mission, were available. The very high resolution images from optical sensors such as QuickBird and IKONOS were available as well. It is interesting to see how information has been exploited in the work of Dell'Acqua et al. (2011), for estimating the damage at districts levels from a post-event SAR image and the fusion with optical images for increasing the accuracy. It is commonly acknowledged that due to speckle effects, single-pixel classification of SAR images leads to unsatisfactory results, and this seems to hold true also when damage assessment is concerned. Satisfying results may be achieved if the damage is assessed at a block level, somehow averaging the unreliable results of pixel-wise comparing pre- and post-event images due to speckle noise. In the Dell'Acqua et al. (2011) article, damage was considered at

pixel scale, in optical data and at block scale, in SAR data, separately. Indeed, the pixel-based approach by using optical data provided useful information on the discrimination between damaged or non-damaged areas. Even though no further classification with respect to damage levels can be retrieved, the information derived from optical imagery can be effective where apparently very high-resolution SAR data is least with the proposed method, and it could be probably the reason why fusion of optical and SAR data in this method is promising. Indeed in this work the SAR textures, which were the features used as damage indicator, were somehow correlated to low damage levels (or affected by other factors than damage), whereas optical data were more suitable to distinguish between damage and non-damage classes. Definitely, the SAR texture was more effective to distinguish between two different levels of damage, in the regions that were identified as damaged by optical data. Some important issues remained still open; such an example could be finding suitable threshold levels for classification in future cases, although significant links between satellite-derived indicators and damage levels have been discovered. It seems that some texture maps computed on SAR data can provide an estimate of damage levels, only if independent information concerning presence or absence of damage in a given block would be available. It is evident from the results that further studies are required and they should be aimed at finding sufficiently independent statistics, extracted from SAR data, which may possibly be linked to presence or absence of damage.

Summary

This entry highlights the characteristics of very high-resolution optical and SAR images that are pertinent for earthquake damage mapping purposes. Automatic or semiautomatic classification algorithms, which make use of this kind of data, have been described as well. Particular attention has been paid to object-based classification approaches, which are more suitable to deal with metric or submetric resolution data.

Relevant case studies, where object-based damage mapping approaches have been considered, are here presented.

Acknowledgment The author would like to thank R. De Meo for reviewing this manuscript.

Cross-References

- [Building Disaster Resiliency Through Disaster Risk Management Master Planning](#)
- [Building Monitoring Using a Ground-Based Radar](#)
- [Damage to Ancient Buildings from Earthquakes](#)
- [Damage to Buildings: Modeling](#)
- [Damage to Infrastructure: Modeling](#)
- [Earthquake Damage Assessment from VHR Data: Case Studies](#)
- [Earthquake Disaster Recovery: Leadership and Governance](#)
- [Estimation of Potential Seismic Damage in Urban Areas](#)
- [Features Extraction from Satellite Data](#)
- [Hyperspectral Data in Urban Areas](#)
- [Remote Sensing in Seismology: An Overview](#)
- [Resilience to Earthquake Disasters](#)
- [SAR Images, Interpretation of](#)
- [Seismic Collapse Assessment](#)
- [Seismic Loss Assessment](#)
- [System and Damage Identification of Civil Structures](#)
- [Urban Change Monitoring: Multi-temporal SAR Images](#)

References

- Bignami C, Chini M, Stramondo S, Pierdicca N, Emery WJ (2011) Objects textural features sensitivity for earthquake damage mapping. In: IEEE GRSS and ISPRS Joint Urban Remote Sensing Event – JURSE 2011 proceedings, pp 333–336
- Blaschke T (2010) Object based image analysis for remote sensing. *ISPRS J Photogramm* 65:2–16
- Brunner D, Lemoine G, Bruzzone L (2010) Earthquake damage assessment of buildings using VHR optical and SAR imagery. *IEEE Trans Geosci Remote* 48:2403–2420

- Chini M (2009) Earthquake damage mapping techniques using SAR and optical remote sensing satellite data. In: Jedlovac G (ed) *Advances in geoscience and remote sensing*. InTech, Vukovar, pp 269–278
- Chini M, Bignami C, Stramondo S, Pierdicca N (2008a) Uplift and subsidence due to the December 26th, 2004, Indonesian earthquake and tsunami detected by SAR data. *Int J Remote Sens* 29:3891–3910
- Chini M, Bignami C, Emery WJ, Pierdicca N, Stramondo S (2008b) QuickBird panchromatic images for mapping damages at building scale caused by the 2003 Bam earthquake. In: 2008 I.E. International Geoscience and Remote Sensing Symposium (IGARSS) – proceedings, vol 2, Issue 1, II1029–II1031
- Chini M, Pierdicca N, Emery WJ (2009) Exploiting SAR and VHR optical images to quantify damage caused by the 2003 Bam earthquake. *IEEE Trans Geosci* 47:145–152
- Chini M, Cinti FR, Stramondo S (2011) Co-seismic surface effects from very high resolution panchromatic images: the case of the 2005 Kashmir (Pakistan) earthquake. *Nat Hazard Earth Syst Sci* 11:931–943
- Chini M, Piscini A, Cinti FR, Amici S, Nappi R, De Martini PM (2013) The 2011 Tohoku-Okii (Japan) tsunami inundation and liquefaction investigated by optical, thermal and SAR data. *IEEE Geosci Remote Sens Lett* 10:347–351
- Chini M, Chiancone A, Stramondo S (2014) Scale object selection (SOS) through a hierarchical segmentation by a multi-spectral per-pixel classification. *Pattern Recognit Lett* 49:214–223
- Dell'Acqua F, Bignami C, Chini M, Lisini G, Polli D, Stramondo S (2011) Earthquake rapid mapping by satellite remote sensing data: L'Aquila April 6th, 2009 event. *IEEE J Selected Top Appl Earth Observations Remote Sens* 4:935–943
- Earthquake Engineering Research Institute – EERI (2004) Preliminary observations on the Bam, Iran, earthquake of December 26, 2003. EERI Special Earthquake Report, [Online]. Available at http://www.eeri.org/lfe/pdf/iran_bam_eeri_preliminary_report.pdf
- Ferro A, Brunner D, Bruzzzone L (2013) Automatic detection and reconstruction of building radar footprints from single VHR SAR images. *IEEE Tran Geosci Remote Sens* 51:935–952
- Franceschetti G, Iodice A, Riccio D (2002) A canonical problem in electromagnetic backscattering from buildings. *IEEE Trans Geosci Remote Sens* 40:1787–1801
- Franceschetti G, Iodice A, Riccio D, Ruello G (2003) SAR raw signal simulation for urban structures. *IEEE Trans Geosci Remote Sens* 41:1986–1995
- Gamba P, Dell'Acqua F, Trianni G (2007) Rapid damage detection in the Bam area using multitemporal SAR and exploiting ancillary data. *IEEE Trans Geosci Remote Sens* 45:1582–1589
- Guida R, Iodice A, Riccio D, Stilla U (2008) Model-based interpretation of high-resolution SAR images of buildings. *IEEE J Selected Top Appl Earth Observations Remote Sens* 1:107–119
- Guida R, Iodice A, Riccio D (2010) Height retrieval of isolated buildings from single high-resolution SAR images. *IEEE Trans Geosci Remote Sens* 48:2967–2979
- Gusella L, Adams BJ, Bitelli G, Huyck CK, Mognol A (2005) Object-oriented image understanding and post-earthquake damage assessment for the 2003 Bam, Iran, earthquake. *Earthq Spectra* 21(S1):225–238
- Haralick RM (1979) Statistical and structural approaches to texture. *Proc IEEE* 67:786–804
- Haralick RM, Shanmuga K, Dinstein IH (1973) Textural features for image classification. *IEEE Trans Syst Man Cybern* 3:610–621
- Hisada Y, Shibaya A (2004) Building damage and seismic intensity in Bam city from the 2003 Iran, Bam, earthquake. *Bull Earthquake Res Inst Univ Tokyo* 79:81–93
- Kropatsch WG, Strobl D (1990) The generation of SAR layover and shadow maps from digital elevation models. *IEEE Trans Geosci Remote Sens* 28:98–107
- Mathieu R, Aryal J, Chong AK (2007) Object-based classification of Ikonos imagery for mapping large-scale vegetation communities in urban areas. *Sensors* 7:2860–2880
- Matsuoka M, Yamazaki F (2004) Use of satellite SAR intensity imagery for detecting building areas damaged due to earthquakes. *Earthq Spectra* 20(975):994
- Pacifici C, Chini M, Emery WJ (2009) A neural network approach using multi-scale textural metrics from very high resolution panchromatic imagery for urban land-use classification. *Remote Sens Environ* 113:1276–1292
- Pesaresi M, Benediktsson JA (2001) A new approach for the morphological segmentation of high-resolution satellite imagery. *Trans Geosci Remote Sens* 39:309–320
- Soille P (2003) *Morphological image analysis – principles and applications*. 2nd edn. Springer, Berlin
- Stramondo S, Bignami C, Chini M, Pierdicca N, Tertuliani A (2006) Satellite radar and optical remote sensing for earthquake damage detection: results from different case studies. *Int J Remote Sens* 27:4433–4447
- Stramondo S, Chini M, Salvi S, Bignami C, Zoffoli S, Boschi E (2008) Ground deformation imagery of the may Sichuan earthquake. *EOS Trans Am Geophys Union* 89:341–342
- Stramondo S, Chini M, Bignami C, Salvi S, Atzori S (2011) X- C- L-Band DInSAR investigation of the April 6th, 2009, Abruzzi earthquake. *IEEE Geosci Remote Sens Lett* 8:49–53
- Turker M, Sumer E (2008) Building-based damage detection due to earthquake using the watershed segmentation of the post-event aerial images. *Int J Remote Sens* 29(11):3073–3089
- Voigt S, Kemper T, Riedlinger T, Kiefl R, Scholte K, Mehl H (2007) Satellite image analysis for disaster

- and crisis-management support. *IEEE Trans Geosci Remote Sens* 45:1520–1528
- Yamazaki F, Yano Y, Matsuoka M (2005) Visual damage interpretation of buildings in Bam city using QuickBird images following the 2003 Bam, Iran, earthquake. *Earthq Spectra* 21(S1):329–336
- Yonezawa C, Takeuchi S (2001) Decorrelation of SAR data by urban damage caused by the 1995 Hiyogoken-Nambu earthquake. *Int J Remote Sens* 22:1585–1600
- Zebker HA, Villasenor J (1992) Decorrelation in interferometric radar echoes. *IEEE Trans Geosci Remote Sens* 30:950–959

Building Disaster Resiliency Through Disaster Risk Management Master Planning

Fouad Bendimerad¹, Jerome Zayas¹,
Bijan Khazai² and Moses Kent Borinaga¹
¹Earthquakes and Megacities Initiative, Quezon City, Philippines
²Karlsruhe Institute of Technology, Center for Disaster Management and Risk Reduction Technology (CEDIM), Karlsruhe, Germany

Introduction

Since impacts of disasters are most felt at the local level, they require proactive action and intervention directly from local authorities and affected communities. However, the extent of impacts varies from one area to another. Highly urbanized areas are typically more vulnerable to natural hazards as a result of environmental and socioeconomic stresses brought about by rapid and often unplanned urbanization. Because of the concentration of population, infrastructure, and resources in urban areas, they stand to lose more in the event of actual disaster. Thus, it is becoming increasingly imperative for city government authorities and local actors in urban areas to develop and enhance their capacity to manage disaster risks in order to protect their development.

Cities have largely and chronically been neglected by national governments and international organizations in terms of dealing with risk

and its contributory factors such as poverty and rapid urbanization (Khazai and Bendimerad 2008). Contrary to popular misconception of local authorities in urban areas having the requisite capacity to address risk on their own, most cities, particularly in the developing world, have actually been ineffective in managing their risk, which remains high and is continuously rising. The relative negligence of disaster risk management as a local responsibility is also exacerbated by the fact that many local authorities are already overburdened by their presently mandated responsibilities.

Disaster risk management (DRM) is defined by the United Nations International Strategy for Disaster Reduction (UN-ISDR) as “*the systematic process of using administrative directives, organizations, and operational skills and capacities to implement strategies, policies and improved coping capacities in order to lessen the adverse impacts of hazards and the possibility of disaster*” (UN-ISDR 2009). Current approaches to managing disaster risk are often characterized by (a) the prevalence of a response or humanitarian perspective rather than a development perspective, (b) the lack of cross-disciplinary and cross-sectoral processes demanded by a more integrated mainstreaming approach that captures the parameters of risk that are relevant to each sector, and (c) the absence of policymaking processes based on consensus building and effective engagement of the concerned stakeholders. These factors are correlated, calling for a holistic approach to urban resilience, one that integrates disaster risk reduction (DRR) with developmental policy and planning, delivery of core services, natural and environmental resource management, and poverty reduction. A methodology that could define needs to be met and why and how to do it is necessary to generate support for allocating resources to disaster risk reduction, especially in view of overwhelming demands from citizens for vital services and to meet day-to-day needs.

The key to manage disaster risks in urban areas without draining the resources of local authorities is to mainstream DRM in the daily functions and services of the city. By integrating

risk reduction parameters in planning processes, such as land use and urban planning, public works, delivery of core services, and emergency response planning, among others, the task of managing disaster risks becomes more attainable. However, mainstreaming DRM in the operations of urban local authorities is easier said than done. Disaster risk management in general is a new field of practice for most city officials and personnel, not to mention that disaster risk reduction implementation itself is complex. It takes time, effort, training, and, most importantly, scientifically designed methods and tools in order to effectively and fully assimilate disaster risk reduction in city functions and ongoing operations. DRM mainstreaming is also closely in line with the commitment of countries to several international conventions and agreements such as the Hyogo Framework for Action for 2005–2015 (HFA) and the UN Millennium Development Goals (MDG), to name a few.

This entry further elaborates the concept of DRM mainstreaming, shows examples of tools that are used for mainstreaming, and provides a case study of an actual application of DRM mainstreaming using the Disaster Risk Management Master Planning (DRMMP) process developed by the Earthquake and Megacities Initiative (EMI).

The Concept of DRM Mainstreaming

Mainstreaming is a core concept of DRM but is seldom understood. Mainstreaming is essential to achieving effectiveness in a DRM structure. It is about defining and distributing roles and responsibilities between national and local authorities and civil society in a way that would result in greater outcome for lower amount of resources. It is about determining the parameters of risk that are most relevant to each sector and function of government and making these parameters an integral part of the governance and management of urban life. An organization that has accomplished mainstreaming has effectively integrated the practice of risk management within its governance, its functions, and operations.

Mainstreaming requires a deliberate and well-founded strategy and consistent applications of policies that strive to pinpoint and enforce roles and responsibilities. Inherent in this process is a policy of decentralization of authority and accountability. The current practice shows an overemphasis on centralization at the national level of all DRM functions. The reasons are twofold:

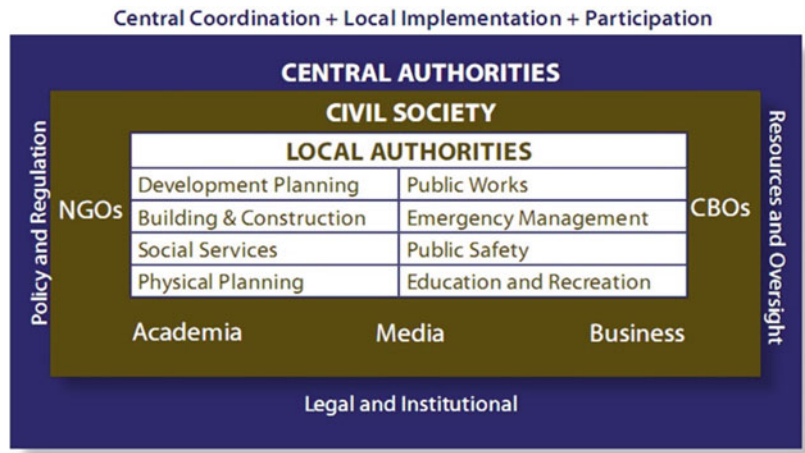
- (a) Lack of decentralization and coordination structures. The DRM practice is an emerging function of government; sublevel government institutional structure and coordination mechanisms have yet to be devised and put in place.
- (b) Lack of capacity of local authorities. The central authorities of the government are reluctant to provide local government with the authority for DRM out of prudence.

Despite these limitations, a distribution of roles and responsibilities between the various levels of government is critical to achieve mainstreaming. One approach to establish a mainstreaming strategy is to follow the scheme shown in Fig. 1 (EMI 2009). The three encompassing boxes provide a descriptive assignment of roles and responsibilities. Fundamentally, the model represented in the figure indicates that, following other more conventional functions of government, DRM implementation has three principles of mainstreaming: central coordination, local implementation, and participation (Bendimerad and Zayas 2015).

- *Local implementation:* Implementation should as much as possible be delegated to local authorities and local actors. Local authorities cannot mainstream DRM without a clear authority and resources.
- *Central coordination:* The role of the central authorities should focus on establishing policy, enacting regulation, putting in place control processes, allocating resources, and exercising oversight.
- *Participation:* Government (central and local) must open the door to the participation of civil

Building Disaster Resiliency Through Disaster Risk Management Master Planning,

Fig. 1 Framework for mainstreaming disaster risk reduction. Functions are indicative (Source: EMI. State of the practice report on urban disaster: open file report. June 2009)



society, which collectively groups all the active agents of society.

The central government delegates aspects of policy, regulatory powers, and other functions and allocates additional resources to local government. It should in turn monitor the performance of local authorities based on sanctioned guidelines and criteria. The boxes do indicate a strict division between the various entities. In fact, flexibility and coordination are keys to success. Within this mainstreaming framework, an optimization of resources and outcome is accomplished when local authorities are strategically establishing internal policies and process where planning, programming, and action for risk reduction is integral to each of their jurisdictional responsibilities either as a service provider or as a regulator. This process is sometimes referred to as “integrated” or “holistic” approach to DRM.

This framework also suggests that risk reduction can be mainstreamed in local governance by harnessing existing legal mechanisms (e.g., local government acts, environmental management acts, DRM acts and related laws and regulatory mechanisms), processes (e.g., licensing, land use, etc.), and service delivery systems (e.g., health, education, transport, utilities, revenue generation) and making use of such resources. For local authorities, this would translate into having processes and practices that inherently incorporate urban risk reduction in these core functions

that they undertake such as construction and building licensing, land use and urban development planning, environmental management, and social welfare as well as in the services that they provide or regulate. To provide necessary authority, additional reforms are often needed, particularly in aligning DRM laws and policies with other acts that regulate local government functions. The point is that wherever responsibility is located, authority and resources should also be provided.

Review of Tools for DRM Mainstreaming

In line with global efforts to reduce disaster risks by focusing on local-level action through mainstreaming, several disaster risk management models have been developed that integrate disaster risk in local development. A review has been conducted of 76 tools, 10 of which were shortlisted for detailed assessment because of their applicability for local application in urban settings (Berse et al. 2009). “Tools” are used interchangeably with models and instruments and in this context refer to a framework that lays out the methodology for undertaking DRM including the mechanism for implementing the elements in the methodology.

Each tool’s relevance to disaster risk reduction and their applicability or feasibility to the local government’s capacity to institutionalize disaster

risk reduction efforts within its functions and operations were carefully considered and weighed in the assessment. Each tool has its own strengths and weaknesses that local authorities in urban areas can learn from in mainstreaming DRM. Evaluation of DRM tools were based on a set of criteria indicators representing the three principles of mainstreaming, i.e., central coordination, local implementation, and participation. The criteria indicators are meant to assess effectiveness and relevance to DRR processes, feasibility requirements, and applicability to the urban setting especially of developing countries. Through this study (which is not exhaustive), a set DRM instruments have been developed—or at least have the potential—for replication in areas outside of their respective origins and were rated for local adoption in the urban context. A summary of the results of this study is provided here.

Effectiveness

Local Implementation

The paradigm shift in the disaster management policies has elicited the participation and commitment of local governments to a more proactive disaster risk management. It is argued that since local stakeholders are the ones greatly affected in the events of disasters, it is critical for local governments to play a major role in the DRM. This is the very reason for giving emphasis on the local implementation of DRM mainstreaming models.

Of the frameworks reviewed, the following have evolved around the concept of DRM mainstreaming: Comprehensive Hazard and Risk Management (CHARM) Programming Approach (SOPAC 2002), Total Disaster Risk Management (TDRM) (ADRC 2005), Central United States Earthquake Consortium's (CUSEC'S) Disaster Resistant Community (DRC) Model (CUSEC 1997), and the Bangladesh DRR Mainstreaming model (Siddiqui 2007). CUSEC's DRC Model was developed by a consortium that includes both the local community and business sectors and not surprisingly gives the most emphasis on the involvement of

these constituencies in the DRM processes. Although designed to be implemented at the national level, Emergency Management Australia's (EMA) Emergency Risk Management Process (Commonwealth of Australia 2004) can also be developed for local government DRM mainstreaming implementation.

Central Coordination

The reviewed DRM mainstreaming models are consistently aligned with the disaster risk reduction strategies formulated by the DRM authorities of their respective countries. The TDRM as in a few other cases was developed by nongovernmental organizations for application in different cities and countries. The Total Disaster Risk Management implementation strongly encourages endorsement from the central DRM authority and adaptation to the DRM institutional framework of the country of application. The Total Disaster Risk Management developed by the Asian Disaster Reduction Center- Office for the Coordination of Humanitarian Affairs (ADRC-OCHA) was endorsed by the government of Thailand and its central DRM authority and adapted for implementation in the country (ADRC 2005).

Participation

The need for stakeholder participation is a well-established approach used in development work. Its main tenet is that stakeholders are collaborators in a development project at every stage of the process. Through this approach, stakeholders influence and share control over the decisions that affect them and their resources. In essence, the stakeholders drive the development process. They are the ones who decide in the planning, design, implementation, monitoring, and evaluation of the project. In this process, they are guided by experts and specialists who also processed the outcome in a scientifically rigorous output. With this, participation does not only ensure stakeholders' ownership but also encourage commitment. In general, the reviewed models emphasize multistakeholder participation at every stage of the DRM planning process. They also recognize the relevance of the participatory process as

enabling communication mechanisms. Models provide more relevance to particular stakeholders: national authorities (e.g., Comprehensive Hazard and Risk Management and Total Disaster Risk Management), local authorities (e.g., Central United States Earthquake Consortium), business sector (e.g., Central United States Earthquake Consortium's Disaster Resistant Community), or community (e.g., BDRRM, Central United States Earthquake Consortium). At the same time, other tools such as the BDRRM provide specific indications of the targeted stakeholders.

Applicability

Relevance

The reviewed DRM mainstreaming models generally adhere to the main goal of integrating risk management in the development and governance processes in order to reduce risks and mitigate the impact of disasters, albeit employing distinct strategies and aiming at varied focus.

The CUSEC model, for example, prioritizes achieving disaster-resistant communities. With this, the model aims to reduce community vulnerability to natural hazards in order to minimize losses and accelerate recovery (CUSEC 1997). The Emergency Management Australia's Emergency Risk Management Process, on the other hand, focuses on reducing risk by making modifications in the causes of risks (Commonwealth of Australia 2004). Similarly, South Pacific Applied Geoscience Commission Disaster Management Unit's Comprehensive Hazard and Risk Management Programming Approach aspires to develop preparedness and reduce vulnerabilities while increasing resilience of the community (SOPAC 2002). The added value of this particular model is that it engages the implementing institution in hazard and risk management in a more holistic manner as it intrinsically links DRR to national development (South Pacific Applied Geoscience Commission SOPAC 2002). Food and Agricultural Organization's DRM Framework (Baas et al. 2008) follows the same line and aims to reduce underlying factors of risk and to prepare for and initiate an immediate response should

disaster hit (Berse et al. 2009). The Asian Disaster Reduction Center- Office for the Coordination of Humanitarian Affairs' Total Disaster Risk Management model, like the Comprehensive Hazard and Risk Management model, is anchored on risk management. However, the TDRM also focuses on "total" approach of balancing the ability to respond to the consequences of disaster with predisaster actions that lessen the need for response (ADRC 2005).

Feasibility/Resource Requirements

Financial and resource requirement are crucial factors for assessing the implementation potential of a particular DRM mainstreaming model. Oftentimes, the availability of resource dictates the scope of DRM mainstreaming implementation and the strategies or tools that will be utilized. Acknowledging this, most of the shortlisted DRM models signified the use of available resources and the development or enhancement of existing resources whether it be financial/material or human resources. While some DRM mainstreaming models identified low-cost tools and strategies for their implementation, others opted for more sophisticated and high-cost tools or mechanisms.

Apart from relying on the engagement of different stakeholders and the local government for human resources, interestingly Central United States Earthquake Consortium uses the strategy of providing incentives to promote steps and measures to reduce the vulnerability of communities to natural hazards (CUSEC 1997). The Food and Agricultural Organization's DRM Mainstreaming model, South Pacific Applied Geoscience Commission-DMU's Comprehensive Hazard and Risk Management, and the Bangladesh DRM Mainstreaming model all give importance on strengthening the capacities of local stakeholders in disaster management and risk reduction and improving advocacy and awareness (SOPAC 2002). They do not specify what would be the required resources for implementation.

Emergency Management Australia's Emergency Risk Management did more than this by identifying a package of special tools for risk

management (i.e., Unique Identifier System, the Risk Register Database, and the Geographic Information Systems) and at the same time by carrying out emergency risk management training for local stakeholders (Commonwealth of Australia 2004).

Disaster Risk Management Master Plan (DRMMP)

The Disaster Risk Management Master Plan was developed by Earthquake and Megacities Initiative (EMI) to support local-level long-term planning and programming of disaster risk reduction activities. It follows the approach set out in Australian/New Zealand Standard 4360-2004, a standard in risk management developed jointly by Australia and New Zealand and recently adopted by the International Standards Organization as ISO31000 for risk management. The DRMMP is based on EMI's mainstreaming framework that optimizes resources by defining and distributing roles and responsibilities across various stakeholders at all levels of government.

The DRMMP provides opportunity for local authorities to systematically and rigorously identify programs, projects, and activities (PPA) aimed at reducing risk caused by natural and man-made hazards and to define processes for implementing these projects and activities. Local authorities are familiar with the master planning process as it is an integral part of their planning activities in particular for land use planning and transport. These plans constitute the basis for budgeting and investments.

Embedding the three principles of mainstreaming (i.e., local implementation, central coordination, and participation), the DRMMP process is designed to align along core planning processes of local governments and is strategically developed for local government DRM implementation. It is anchored on the regulatory framework that defines mandate and authority of government. It is a methodology for integrating risk parameters into various local government functions. EMI's DRMMP

implementation also strongly encourages formal endorsement from the central DRM authority and adaptation to the DRM institutional framework of the country of application. Finally, the DRMMP process has been designed as an approach for organizing the stakeholders in thematic Focus Groups to ensure efficiency and sustainability using a participatory process with targeted stakeholders.

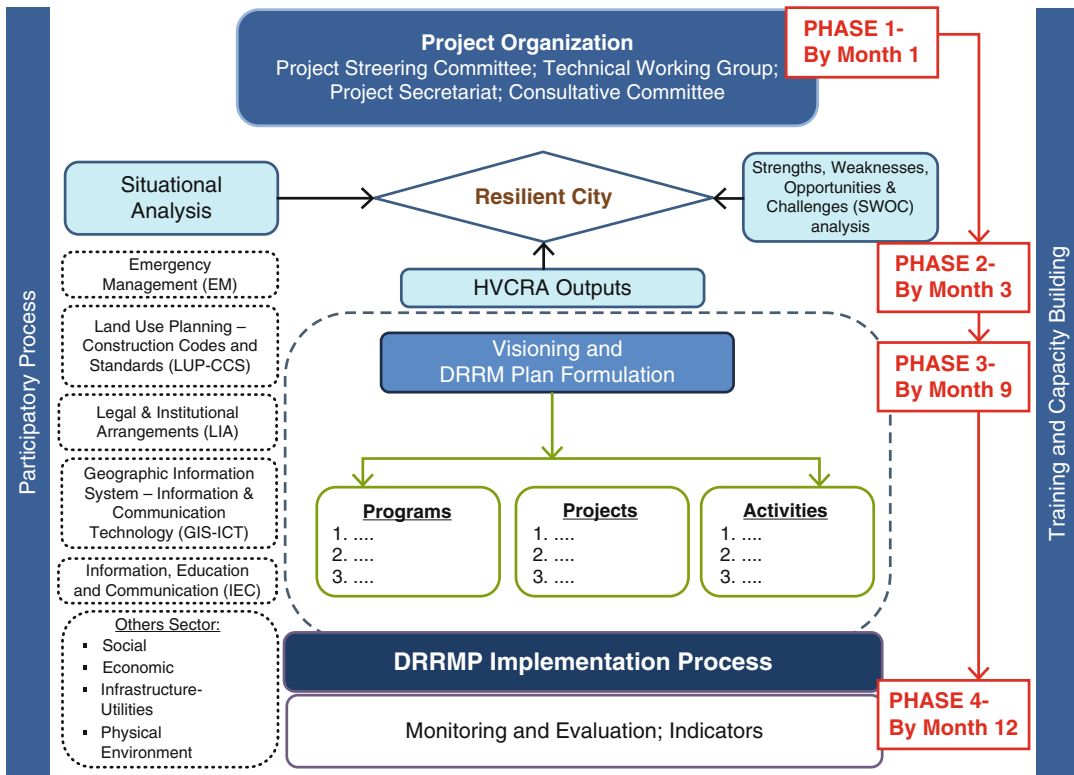
The DRMMP methodology follows similar principles and processes as other more conventional planning processes, e.g., land use planning, namely,

- It is structured and systematic and follows a consistent step-by-step methodology.
- It is science based and data driven.
- It is multisectoral and multidisciplinary.
- It applies to all hazards.
- It is participatory and consensus driven.

By anchoring on mainstreaming, the DRMMP puts in place a structured approach for setting up a DRM system that is based on science, understood by all, and provides the rationale for investments in DRR programs and projects. As of 2014, the DRMMP has already been applied to eight (8) local authorities, namely, Istanbul (Turkey), Kathmandu (Nepal), Metro Manila (Philippines), Amman (Jordan), Mumbai (India), Pasig City (Philippines), Quezon City (Philippines), and Dhaka (Bangladesh).

Process to Implement the DRMMP

The Disaster Risk Management Master Plan provides a *road map* for a proactive DRM approach, which recognizes that disasters are not just “set-backs” or “roadblocks” to development but result from the paths that development is pursuing. Thus, by changing planning processes, and incorporating disaster risk parameters explicitly in the planning of processes and projects, it can be ensured that in future natural hazards will encounter resilient communities that are capable of withstanding their impact, reduce their losses, cope, and recover normal life faster. Ultimately,



Building Disaster Resiliency Through Disaster Risk Management Master Planning, Fig. 2 The DRMMP workflow (Source: Urban Resilience Master Planning: A Guidebook for Practitioners and Policy Makers, March 2015)

impacts of hazards become mere emergencies rather than disasters.

The undertaking of a DRMMP follows a series of analyses along each of four phases, namely,

- Phase 1: Organization and Preparation
- Phase 2: Diagnosis and Analysis
- Phase 3: Plan Development
- Phase 4: Plan Implementation, Monitoring, and Evaluation
- Legal and Institutional Arrangements (LIA)
- Land Use Planning (LUP) and development controls
- DRM Database and Geographic Information System (GIS)
- Hazard, Vulnerability, and Risk Assessments (HVRA)
- Emergency Management (EM)

Figure 2 provides the workflow process for developing the DRMMP, which can be completed in 12–24 months, depending on the elements that need to be covered. First, the elements that need to be included in the master planning process have to be identified. Some components are core to the DRMMP methodology because they provide the essential scientific information on the DRM practice. These include

Further elements/sectors (e.g., housing, transport, sanitation, health, education, social services, and others) can be added depending on the jurisdictional responsibility of the city. This also depends on what is most relevant to local authorities in terms of priority, access to data, and availability of resources and expertise. For example, in the application of the DRMMP in Greater Mumbai Project, there were additional elements that were included such as Water, Wastewater

and Storm Drainage Systems, and Transportation. This is due to the fact that the Municipal Corporation of Greater Mumbai regulates the delivery of these services. It is important to note that not all sectoral elements fall under the authority of the city or of other agencies of the government. Many could be provided by the private sector or utility companies (e.g., water, power). Others would be a mix of public and private entities (e.g., transport). Other elements may also have to be addressed because of the local context and conditions particular to a city. For example, in the Mumbai application, elements such as Slums, Shelter, and Housing, were likewise addressed because of the unique characteristics of the city that has more than 60 % of its population living in the slum areas.

A listing of additional elements included in the DRMMP includes

- Shelter and Housing (including slums and informal settlements) (S&H)
- Health
- Education
- Construction Codes and Standards (CCS)
- Water Systems
- Waste Water and Storm Drainage Systems (WWSD)
- Power Systems (PS)
- Transportation Systems (TS)
- Solid Waste Management (SWM)
- Telecommunications
- Possibly other relevant sectors

There are other elements that support the whole planning process and are crosscutting in nature. These include

- Training and Capacity Building (TCB)
- Information Education and Communication (IEC)

Training and Capacity Building and Information Education and Communication components facilitate the meaningful participation of stakeholders since their aim is to bring the gap in understanding of DRM concepts and principles between the EMI experts and practitioners and

the stakeholders. Throughout the DRMMP process, the stakeholders are made to understand and take ownership of the types of risks that their city faces as they acquire the competency and tools to be able to manage those risks on their own.

Participatory Approach to Develop the DRMMP

The DRMMP allows for the meaningful participation of stakeholders. By design, the DRMMP provides stakeholders with an opportunity for learning the concepts of hazard, vulnerability, and risk and to understand their relevance to planning. It develops consensus among them on the trade-offs and the rationales for investment in urban resilience. With their engagement in the development of the plan, the stakeholders not only provide essential knowledge to the plan but acknowledge their roles and responsibilities in the implementation of the plan, thus taking ownership. The underlying principle to EMI's participatory approach is that urban DRR can only be made meaningful if stakeholders understand their risks. They should be well informed in order to meaningfully contribute. In most cases, there is a gap in how risk is understood and interpreted by various stakeholders with diverse backgrounds and oftentimes competing interests. Once there is shared understanding of risk, stakeholders can contribute to the work at hand in terms of data collection and validating findings and interpretations of risk assessments that are fundamental to any risk reduction program.

EMI's version of participatory approach was first applied in 2008 in the *Makati Urban Redevelopment Planning Project: Mainstreaming Disaster Risk Reduction in Megacities, A Pilot Application in Metro Manila and Kathmandu*. It was then further tested and applied in various cities and has become EMI's cornerstone strategy to stakeholder engagement in urban DRR. Using this approach, the participation by stakeholders is organized along key sectors that allow for generating inputs across a broad range of stakeholders with different agenda and interests but at the same time facilitate decision and action.

Building Disaster Resiliency Through Disaster Risk Management Master Planning, Fig. 3 Project organization structure of the disaster risk reduction in Greater Mumbai project which shows an example of a flat non-hierarchical structure that is composed of stakeholders from different sectors that are working together in a participatory approach



Typically, implementing the DRMMP requires organizing the stakeholders along the following:

- Creation of a **Project Management Team** (PMT) composed of the project managers and technical experts and specialists who make up the main workforce that will develop the DRMMP. It is multidisciplinary and structured as a single unit incorporating EMI’s experts and counterpart from the host local authority.
- Creation of the **Advisory Committee** which is composed of policy-/decision-making officials who are endorsed as representatives to the project by their respective institutions. It provides the platform to receive input and guidance from the stakeholders and share the project issues, findings, and accomplishments that the members can take back to their respective institutions.
- Creation of the **Scientific Consortium** which is composed of local experts in relevant fields

- addressed by the project. The members of the Scientific Consortium are selected on the basis of their credentials. They advise on the validity of the scientific data and approach and help reach scientific consensus.
- Creation of the **Focus Groups** which are composed of local practitioners, researchers, and community leaders. These include midlevel managers and specialists from the various departments and offices of the city. They are organized along each of the sectors that are being analyzed in the DRMMP. Through the Focus Groups, relevant stakeholders get engaged in the development of the project by providing inputs and tackling issues, as well as validating the assumptions, findings, and recommendations relevant to the project.

Figure 3 below shows an example of the project organization structure of the DRMMP that was put in place in Mumbai. It reveals a flat nonhierarchical structure that puts emphasis on

teamwork and consensus building but at the same time assigns responsibility and accountability to the project team. It requires experts, practitioners, policymakers, community leaders, and private sector working together to provide inputs to develop the elements of the DRMMP. In the Disaster Risk Reduction in Greater Mumbai Project, there were more than 100 members of the Advisory Committee representing various organizations of government, civil society, technical agencies, and the private sector. In the Building a Disaster Resilient Quezon City Project, the various Focus Groups were organized under a Technical Working Group (TWG). They are composed of technical specialists from the different departments of the Quezon City Government officially assigned by the Office of the Mayor to work with EMI to develop the various elements of the DRMMP.

Strong Emphasis on Data and Science

The development of the DRMMP requires a strong emphasis on data and science. It includes collecting evidence to understand the DRM context and situation, assessing the inherent risks of the city, developing an information database on disaster risk management, and identifying the gaps and needs. The following sections describe several key activities that are conducted to produce the required assessments to develop the DRMMP.

Development of the DRM Database and City Risk Profile

A significant part of the effort goes into data collection, review, and structuring all data into a single DRM database. Data must be collected from several city departments as well as from relevant government agencies and other service providers. This is the most challenging task in the process as not all departments willingly share data. The data is typically organized at the resolution of the lowest geopolitical boundary of the city. This will enable an effective use and management of the data by the local and sublocal authorities. A list of the data that needs to be

Building Disaster Resiliency Through Disaster Risk Management Master Planning, Table 1 Data typically collected in the development of the DRMMP

Administrative and institutional	Build environment, including:
Demographic and socio economic	Buildings
Physical profiles including:	Critical and essential facilities
Topology; geology; hydrology; soil, climate	Infrastructure: transport, water, drainage, sanitation, power, communication, health
Land use	High loss facilities

collected is indicated in Table 1 below. The data is compiled in a City Risk Profile shared and validated with all the stakeholders. Knowledge-sharing mechanism is defined and endorsed to ensure the data is kept current and is benefiting other planning functions of the city. In addition to facilitating DRM planning, tax mapping provides another motivation for cities to have a centralized database in uniform format.

Analysis of the Legal and Institutional Arrangements for DRM

The DRMMP process is anchored on the legal and institutional arrangements that define the geopolitical boundaries of the concerned planning authority (i.e., the local government), its jurisdictional responsibility, and the potential elements of the plan. An in-depth analysis of the legal and institutional context is essential in providing legitimacy and credibility to the plan and in avoiding potential jurisdictional conflicts. Several outputs are generated such as a network analysis that shows the organizational chart of the DRM undertaking in the city with key agencies and their relationships. A functional mapping is also completed to indicate functions, roles, and responsibilities of each organization in the DRM process. The results are shared and validated with the stakeholders through focus group exercises to ensure that the DRM organization is understood by all concerned agencies and officials. An example of network analysis is shown in Fig. 4.

Hazard, Vulnerability, and Risk Assessment (HVRA)

The DRMMP is informed by the hazard, vulnerability, and risk parameters which provide the scientific basis for the plan and ensures that the plan is responding to the objectives of risk reduction and resilience enhancement. The parameters defining the current practice of the elements of the plan (i.e., sectors) are checked against the risk parameters to define the gaps and the strategies for improving urban resilience. The hazard, vulnerability, and risk parameters are developed using scientific methodologies (e.g., scenario analysis or probabilistic risk assessment) that provide information on the distribution of damage and the social, economic, and material losses to the city considering both physical vulnerability and socioeconomic vulnerabilities. The risk assessment should be done at the highest resolution possible to provide an accurate and reliable understanding of the risks. The results are presented, discussed, and validated by the stakeholders. The greater the understanding of the risks, the more relevance is given to the master plan.

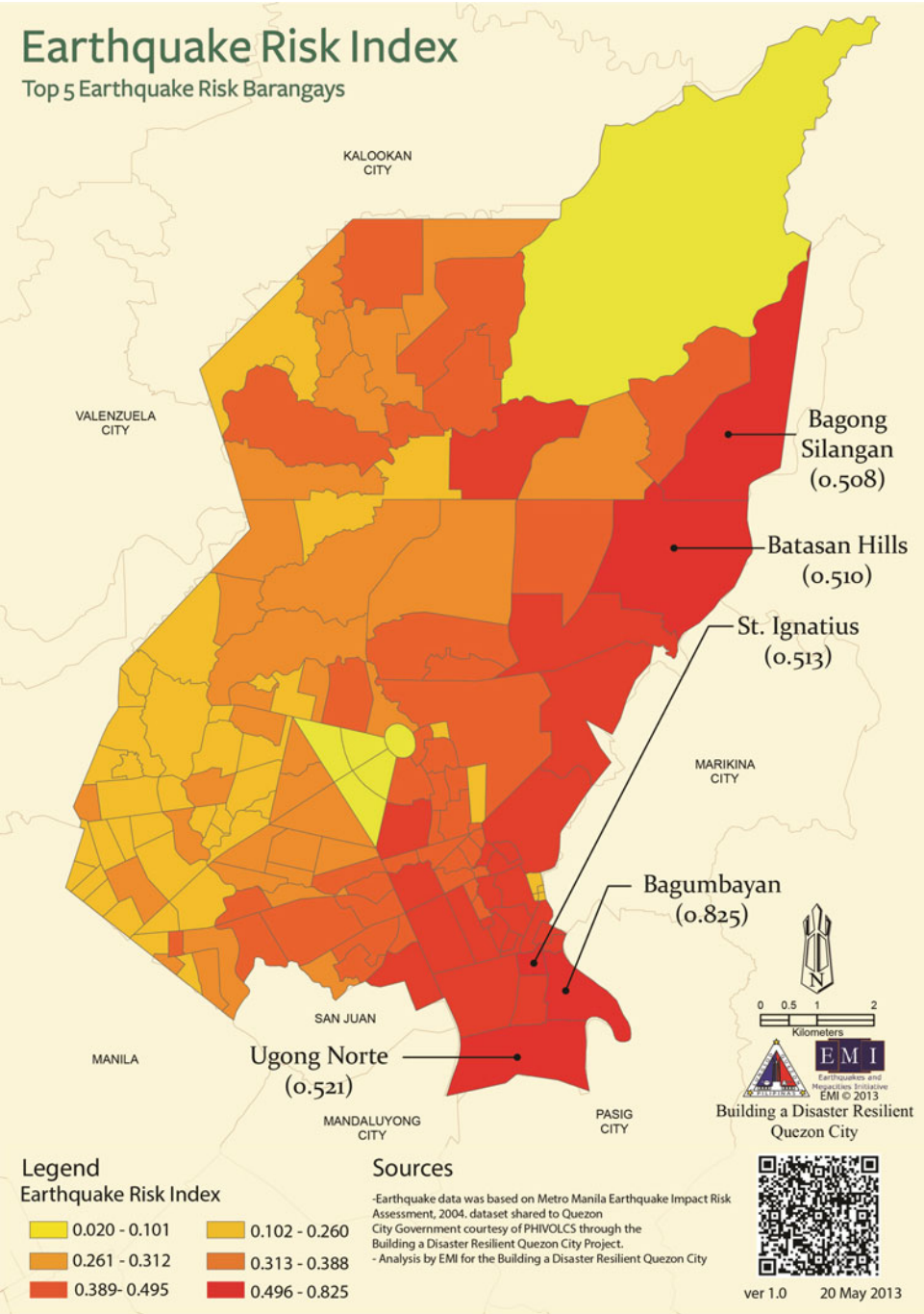
The Hazard, Vulnerability, and Risk Assessment also provides information on “hotspots” which are areas that have the highest risk for the considered hazards. The scientific approach to defining “hotspots” relies on developing dimensional indicators by which parameters of physical risk (e.g., number of heavily damaged buildings in a certain locality) are combined with parameters that represent the capacity of that particular locality to sustain and recover from the impact of an event (Nardo et al. 2005). Through the hotspots, local authorities will know where they should prioritize the DRR projects. Maps are produced to provide illustrations where the potential problems are highest. They indicate priorities for planning in terms of where the investments should be for reducing risk and building resilience that address area-specific issues. An example of hazard mapping and hotspot analysis is provided in Fig. 5.

Situational Analyses of Urban Resilience Elements

Situational analyses on key urban resilience elements are conducted to provide the baseline for understanding the state-of-the-practice urban DRR. With the support of the experts, the stakeholders are guided through workshop-type exercises where they can determine the gaps relative to standards for urban resilience. Essentially, the stakeholders are presented with the risk parameters which are contrasted with their current practices. In small group exercises, they look at how the current systems and practices for particular development sectors are responding to the risk parameters and the requirements for disaster risk management. In the *Building a Disaster Resilient Quezon City Project*, the development sectors that were considered include population, economic activity, social services, emergency management, institutional and land use administration, and physical resources (EMI and QCG 2014). Figure 6 provides an example of framework for analyzing the resilience of lifeline systems in Greater Mumbai. It provides the work that needs to be done in the assessment process starting from the mapping of the exposure (i.e., the asset at risk) into a geographic information system (GIS), then superimposing the hazard information, then understanding the impact both in terms of direct damage and secondary damages, then assessing the impact on the system serviceability taking into consideration measures of resilience of the system including physical resilience and community resilience.

Putting the Elements of the Plan Together

For each of the development sectors, programs, projects, and activities are identified as concrete actions to reduce risks and improve resilience. Stakeholders undertake the consequential analysis together with the experts. The consequential analysis is an interactive exercise which builds on the situational analysis but looking at “what-if” scenarios aimed at developing options to improve



Building Disaster Resiliency Through Disaster Risk Management Master Planning, Fig. 5 Earthquake risk profile for Quezon City, Philippines, showing the hotspot counties. The map and corresponding information

was generated through the conduct of a hazard, vulnerability and risk assessment, which is a vital component of the DRMMP



Building Disaster Resiliency Through Disaster Risk Management Master Planning, Fig. 6 The diagram, as an example of a situational analysis exercise, shows the lifeline system and resiliencies and how they influence community resilience. The elements of resiliency are numbered from ① to ⑦, each with visual representations. The arrows identify prerequisite elements; for example in

order to assess the geospatial systemic impacts ③ from an earthquake, one must first understand the geographic distribution of the system ① being analyzed and the geographic distribution of the hazard ②, in this case earthquake ground motions and permanent ground deformation. Taken from the *Disaster Risk Reduction in Greater Mumbai Project*. 2011

Building Disaster Resiliency Through Disaster Risk Management Master Planning, Table 2 Results of consequential analysis of a magnitude 7.2 earthquake to physical resources of Quezon City

Primary hazards faulting, shaking, liquefaction, landsliding	Primary damage: building/ structural non-structural/ equipment	Primary loss: life/ injury, repair costs, function, communication/ control	Secondary hazard/ damage: liquefaction. landsliding fire, hazmat, flooding. . .	Secondary loss: business/operations interruptions, market share, reputation
Collapse of water sam Extreme flood Loss of lives, physical injuries Damages to main roads, infrastructure	Damage to water reservoir	Trauma (Stampede)	Possible trash slide in dumpsites	Loss of jobs
	Building collapse (Eastwood, Libis area)	Isolation of victims due to no road/ highways access	Loss of property, lives	Loss of income
	Loss of water supply	Operation interruptions in hospitals	Physical injuries	Loss of opportunities
	Loss of lives		Fire	Eastwood, malls
	Trapped inside the building		Houses made from light materials	Loss of livelihood
	Damage to equipment		Biogas plant explosion	Production halts
	Stranded people		Informal settlers	Businesses, offices
	Chaos/stampede (inside buildings)		Food shortage	No classes
	Contamination of waterways			

the current practice. Through the consequential analysis, the stakeholders achieve an understanding of the impacts of the current practices and define strategies and rationales for reducing these impacts. They also identify the “entry points” for mainstreaming DRR in each particular development sector. Table 2 shows an example of the result of a consequential analysis for a magnitude 7.2 to the Physical Resources Sector of Quezon City.

Next is the establishment of a vision for resilience and a mission to reduce risk through a visioning exercise. In the Disaster Risk Reduction and Management Plan (DRRMP) process in Quezon City, the stakeholders took some time to establish consensus around developing the vision for the city. There was a long gridlock in identifying a common descriptor for the natural environment around descriptors such as habitable, sustainable, resilient, “clean and green.” This is followed by the development of resilience strategies through the conduct of Strengths-Weaknesses-Opportunities-Challenges (SWOC) exercise to formulate and prioritize strategies by

taking into account key internal and external elements that may affect the attainment of the resilience vision and mission.

After reaching consensus on the resilience strategies, programs, projects, and activities are then formulated. Apart from listing down and discussing potential investments, stakeholders also prioritize projects and activities. This **prioritization process** is based on a set of criteria that helps stakeholders focus their attention to interventions that will have the greatest impact in terms of reducing their risk while taking into account their inherent limitations such as budget capacity or resource constraints. The final form of the DRMMP is then validated. An example of a DRMMP program and its related projects are shown in Table 3.

Ensuring Implementation of the DRMMP

The time frame for a DRMMP can span anywhere from 6 to 15 years, depending on the planning cycles of cities. For the DRMMP to be fully

Building Disaster Resiliency Through Disaster Risk Management Master Planning, Table 3 Example of DRMMP program, related projects, and priority ranking of each project

Program 2: Mainstreaming DRR in land use planning and land use management	
Mitigation and prevention projects	Priority
Formulating a communications plan to institutionalize HVRA within Quezon City (with focus on hotspot areas)	Immediate
Mainstreaming DRR in the 2010–2030 comprehensive land use plan	Medium-term
Consultation workshops for pilot project identification and feasibility of selected land use management methods for flood and earthquake risk reduction	Short-term
Program for improving Quezon City Government’s capacity and expertise in building and infrastructure construction project development permitting, monitoring and evaluation	Medium-term
Restructuring existing M and E program for stakeholder response(s) to and development impacts of Quezon City Government’s projects on disaster risk reduction and management	Short-term
Pilot projects on preferred land use management methods (e.g., development regulations, zoning earthquake hazard areas, floodway zoning and set-backs, fire zones)	Short-term

B

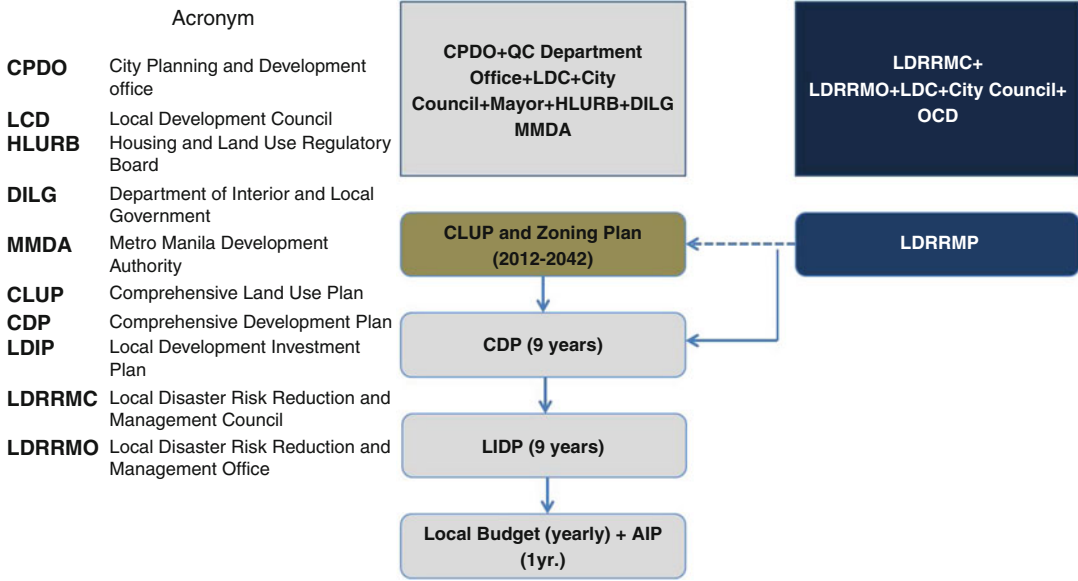
implementable, there has to be a clear understanding of the legal and administrative processes of a city that concerns with approval of mandated plans. This requires knowing who approves, when, and through which step in the approval process should the DRMMP be incorporated so that it becomes integrated in the budgeting cycle. This way, it is ensured that budget for priority projects identified in the DRMMP gets allocated in the overall budget of the city. Figure 7 below provides an example of the budget approval process of Quezon City indicating how the Local Disaster Risk Reduction and Management Plan (LDRRMP) gets integrated in the Annual Investment Plan (AIP). The Annual Investment Plan refers to the indicative yearly expenditure requirements of cities that will be integrated into their annual budget.

Understanding who approves, when, and through which steps in the approval process of cities helps ensure that DRMMP projects and activities are incorporated in the budget and, as such, get funded.

Monitoring, Evaluation, and Communication

To track the progress (or lack of progress) of implementation, monitoring and evaluation tools are developed. One of these tools is to develop a set of Indicators that set initial

benchmarks and can be used to measure progress. The first is the quantitatively derived **Urban Disaster Risk Indicators** (UDRI) which provide a holistic view of disaster risk in a local authority by capturing the direct physical damages and the aggravating social conditions contributing to total risk. The model and methodology referred to here as the Urban Disaster Risk Indicators was originally developed for the Inter-American Development Bank through the IDB-IDEA Indicators Program by the Institute of Environmental Studies (IDEA) of the National University of Colombia, Manizales (NUCM) (Cardona 2003; Carreno et al. 2006). In the context of EMI’s DRMMP methodology and approach, the indicators were used as an innovative risk communication tool to engage stakeholders in understanding their involvement and taking ownership of the risk factors in the city. Besides the initial EMI implementation in Metro Manila (Fernandez et al. 2006), the risk indicators have also been successfully implemented with stakeholders in Istanbul (Khazai et al. 2009), Amman (Bendimerad et al. 2009), Mumbai (EMI and Municipal Corporation of Greater Mumbai 2011a), and Quezon City (Bendimeard et al. 2013). The application of the Urban Disaster Risk Indicators in a city will prove its value if it is “mainstreamed” within the DRM practices and other planning processes that impact DRM in the city. To reach this objective, the indicators have been linked to the quantitative outputs and



Building Disaster Resiliency Through Disaster Risk Management Master Planning, Fig. 7 Example of a budget approval process in Quezon City, Philippines

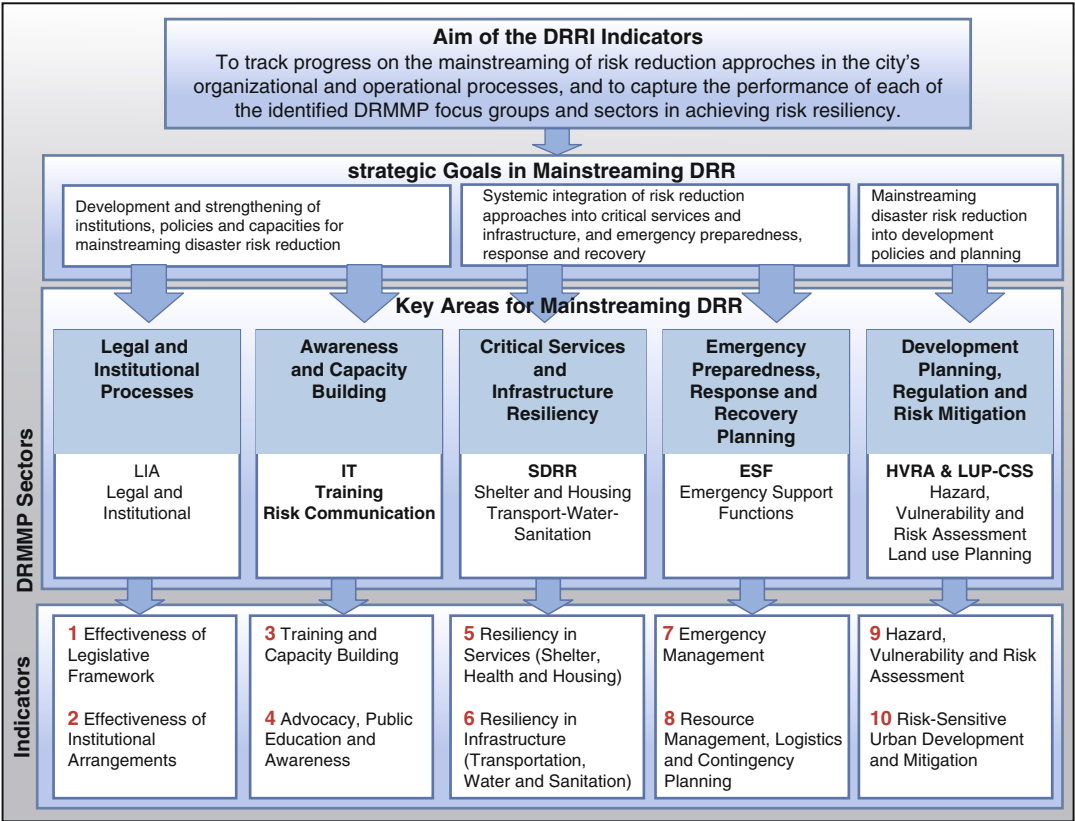
highlighting how the local disaster risk reduction and management plan gets integrated into the annual investment plan

analyses of the DRMMP. Guided by recommendations following implementation of the risk indicators by EMI in different cities, the current methodology encompasses the following three broad areas:

- To develop the system of indicators in close collaboration with the technical staff and officials in a city, to engage them in the development of the data and the understanding of the conceptual framework and methodology, and to ensure ownership over the indicator system and its periodic updating and upgrading
- To work closely with the Focus Group, in evaluating the indicators, such that they are relevant to the megacity scale of analysis, represent the conditions and reality of socio-economic vulnerability in the city, and reflect the outcomes of the various studies undertaken as part of the DRMMP (e.g., earthquake risk assessment, transportation study, housing and shelter, land use planning, construction codes and standards, water supply and water treatment analysis, etc.)
- To implement the system of indicators through a process of engaging a wider group

of stakeholders in the city, such that they are used as an effective risk communication tool that inherently relates to the city-level DRM practices; as a planning tool that aids in correcting, reviewing, and deciding on where to invest resources; and as Disaster Risk Management benchmarks which assist in policy decision-making and monitoring of different risk reduction practices implemented at the local level.

The second group of indicators is the qualitative **Disaster Risk and Resiliency Indicators** (DRRI) that is a set of ten (10) indicators that establishes initial benchmarks to measure to what extent risk reduction approaches have been mainstreamed in the organizational, functional, operational, and development systems and processes of local governments (EMI and MCGM 2011b). The Disaster Risk and Resiliency Indicators are linked to the priorities of the *Hyogo Framework for Action 2015–2015*. The aim of these indicators is to understand how well the local authority is performing in mainstreaming DRR into different sectors based on predefined benchmarks and performance



Building Disaster Resiliency Through Disaster Risk Management Master Planning, Fig. 8 Disaster risk and resiliency indicators (DRRI) (Source: EMI and MCGM (2011b))

targets. The rationale for selecting the 10 DRRRI indicators can be traced in Fig. 8 by following the information from top to bottom of the chart. The main aim of the DRRRI indicators is to track progress on the **mainstreaming** of risk reduction approaches in the city's organizational, functional, operational, and development systems and processes. The mainstreaming goal is further divided into three **strategic goals** shown in the chart. Each of the strategic goals corresponds to one or more **key areas** analyzed in the DRMMP where these goals are to be implemented. Finally, two indicators corresponding to each of the five key areas of mainstreaming are shown. Mainstreaming risk reduction and achieving risk resiliency cuts across all the key areas and all the 10 indicators shown below. Thus, the stakeholders using and scoring the indicators should look at *all 10 indicators* together from the

perspective of their institutions and not only the indicators that relate most closely to their activities.

The scores for the DRRI are derived from a self-assessment along key functional activities/policies undertaken by the DRMMP Focus Groups. A handbook is developed to describe the organization of the indicators, a rationale for their selection, and provide guidelines for their implementation and scoring that is contextualized to the city for which it is implemented by providing "guiding questions" and "evidence for discussion" derived from in-depth interviews with stakeholders along each indicator (EMI and MCGM 2011b). With five Performance Target Levels of attainment: little or no awareness, awareness of needs, engagement and commitment, policy engagement and solution development, and full integration, the DRRI allocates a

ranking for the 10 indicators that fall under the main 5 areas of mainstreaming, resulting in a graphic visualization of the mainstreaming disaster risk reduction at the Municipality at a point in time.

The Indicator System is also used as a risk communication and planning tool to assist in policy development, decision-making, and in monitoring the effectiveness of programs, projects, and activities. In actual practice, however, there is no clear motivation for cities to monitor performance and outcome of DRR investments beyond accounting for the number of projects and activities conducted. This is also because there are no existing tools as yet for monitoring and evaluating performance of DRR programs, projects, and activities that have been developed for cities.

Lessons Learned

Building resilience, especially within the context of urban areas, requires not just raising awareness on risk reduction but actual investments on programs, projects, and activities that reduce risk. For the plan to be fully understood, accepted, and supported, it has to engage a broad range of stakeholders at all stages of the planning process. However, they have to be equipped with the knowledge and tools to be able to participate meaningfully and contribute to resilience building. Stakeholders have to understand what types of risk they face, what are the implications of these risks to their own mandated responsibilities, what options are available to them to reduce their risk, and what possible trade-offs they need to make in order to prioritize limited resources and address inherent constraints.

Full implementation of the DRMMP, however, requires continued technical assistance. Even if funding is available from cities to implement DRR projects, oftentimes, they lack the resources to manage and undertake DRR projects. It is a resource-intensive approach. The DRMMP only provides the roadmap for change in practice from a reactive to a more proactive approach. For the actual change to take place, the

city has to make continued investments in the medium to long term to make sure that it continues on the path toward risk reduction.

Summary

Local authorities have difficulty in dealing with the risks from disasters due to the fact that they are overburdened by their development aspirations and mandated responsibilities. Mainstreaming DRM in the daily functions and services of the city is the key to manage disaster risks in urban areas without exacerbating cost for local authorities. By incorporating disaster risk parameters explicitly in the planning of processes and projects, it can be ensured that in future, natural hazards will encounter resilient cities that are capable of withstanding their impact, reduce their losses, cope, and recover faster. Ultimately, impacts of hazards become mere emergencies rather than disasters.

Mainstreaming necessitates the distribution of roles and responsibilities between the various levels of government and with the society. It gives emphasis to the local authority's implementation role, central government's coordination function (manifested in the control processes, allocation of resources, and exercising oversight on the performance of the local authority), and the participation of the civil society which is crucial in all facets of the process. Thus, it optimizes resources from these stakeholder institutions to reduce the risks.

Following the mainstreaming tenet, the DRMMP provides a road map for a proactive DRM approach to support local-level long-term planning and programming of disaster risk reduction activities. It follows a series of analyses in four phases, namely, (1) Organization and Preparation, (2) Diagnosis and Analysis, (3) Plan Development and Plan Implementation, (4) Monitoring and Evaluation. The DRMMP process is designed to align along core planning processes of local authorities. Furthermore, it provides stakeholders with an opportunity for learning the concepts of hazard, vulnerability, and risk and to understand their relevance to planning.

The DRMMP was developed taking into consideration other related DRM models that focus on local-level action through mainstreaming.

With DRMMP's strong emphasis on data and science, evidences are collected to understand the DRM context and situation, assessing the inherent risks of the city, developing information database on DRM, and identifying the gaps and needs. These tasks are achievable through a participatory approach involving a broad range of stakeholders. Implementing the DRMMP requires organizing these stakeholders into (1) Advisory Committee, (2) Scientific Consortium, and (3) Focus Groups. Guiding these groups is the Project Management Team.

The DRMMP provides opportunity for these stakeholders to understand the parameters of vulnerability and risk in the context of their own responsibilities and functions. With this knowledge, they can systematically and rigorously identify and prioritize programs, projects, and activities to implement in managing and eventually reducing their risk. These projects and activities are institutionalized by integrating these in the budgeting cycle. This way, it is ensured that budget for priority projects identified in the DRMMP gets allocated in the overall budget of the city. Monitoring and evaluation tools are then developed to track the progress (or lack of progress) of implementation.

With their engagement in the development of the plan, the stakeholders not only provide essential knowledge to the plan but acknowledge their roles and responsibilities in the implementation of the plan, thus taking ownership. By anchoring on mainstreaming, the DRMMP puts in place a structured approach for setting up a DRM system that is based on science, understood by all, and provides the rationale for investments in DRR programs and projects.

Cross-References

- ▶ [Community Recovery Following Earthquake Disasters](#)
- ▶ [Earthquake Disaster Recovery: Leadership and Governance](#)

- ▶ [Earthquake Protection of Essential Facilities](#)
- ▶ [Earthquakes and Their Socio-economic Consequences](#)
- ▶ [Economic Recovery Following Earthquake Disasters](#)
- ▶ [Emergency Response Coordination Within Earthquake Disasters](#)
- ▶ [Land Use Planning Following an Earthquake Disaster](#)
- ▶ [Resilience to Earthquake Disasters](#)

References

- Asian Disaster Reduction Center (ADRC) (2005) Total disaster risk management: good practices. ADRC, Kobe
- Baas S, Ramasamy S, De Pryck J, Battista F (2008) Disaster risk management systems analysis. FAO, Rome
- Bendimerad F (2009) State-of-the-Practice Report on Urban Disaster Risk Management. EMI internal report. <http://emi-megacities.org/?emi-publication=state-of-the-practice-report-on-urban-disaster-risk-management>. Quezon City
- Bendimerad F, Zayas J (2015) Urban Resilience Master Planning: A Guidebook for Practitioners and Policy Makers. EMI, Quezon City
- Bendimerad F, Mattingly S, Seva V, Kustu O, Khazai B, Reyes M, Berse K, Al-Soub W, Quran S, Al-Sous G, Khalili H, Dabaseh A (2009) Amman – disaster risk management master plan, support to building national capacities for earthquake risk reduction in Amman. volume 1, Earthquake and Megacities Initiative (EMI), main report. UNDP Jordan – 00051485, 31 Mar 2009, p 130
- Bendimerad F, Khazai B, Zayas J, Daniell JE, Salunat J, Perez B, Lingad L, Dakis K, Bergonia L, Magtaas B, Barba B, Padoa I, Pino L, Lanuya E, Valera A, Villa P, Hisanan T (2013) Building a disaster resilient Quezon City project: hazard, vulnerability and risk assessment report. Earthquake and Megacities Initiative, Quezon City
- Bendimerad F, Zayas J, Padoa I (2015) Legal and institutional arrangements guidebook, Urban resilience guidebook series. EMI, Quezon City
- Berse K, Bendimerad F, Cruz J, Dagli W, Deocariza M, Khazai B, Lopez IG, Pulmano N, Seva V (2009) Desk review of international disaster risk management models and tools intended for local institutional application. GFDRR, p 118, Makati City, Philippines
- Cardona OD (2003) The need for rethinking the concepts of vulnerability and risk from a holistic perspective: a necessary review and criticism for effective risk management. In: Bankoff G, Frerks G, Hilhorst D (eds) Mapping vulnerability: disasters, development and people. Earthscan Publishers, London

- Carreno L, Cardona O, Barbat A (2006) Urban seismic risk evaluation: a holistic approach. *Nat Hazards* 40(1):137–172
- Central United States Earthquake Consortium (CUSEC) (1997) Disaster resistant communities: a community-based approach to hazard mitigation. *CUSEC J* 4(1):2–4
- Commonwealth of Australia (2004) Emergency risk management application guide. Emergency Management Australia, Canberra
- EMI (2009) State-of-the-practice report on urban disaster risk management. Quezon City. Retrieved Mar 2014, from http://emi-megacities.org/home/components/com_booklibrary/ebooks/StateofPractice_UrbanDRM.pdf
- EMI (2014) Legal and Institutional Arrangements Framework Guidebook. EMI, Quezon City
- EMI, Municipal Corporation of Greater Mumbai (MCGM) (2011a) Mumbai DRMMP framework report and implementation work outputs. EMI, Quezon City
- EMI, MCGM (2011b) Final technical report supplement. EMI, Quezon City
- EMI, Quezon City Government (QCG) (2014) Disaster risk reduction and management plan 2014–2020. EMI, Quezon City
- Fernandez J, Mattingly S, Bendimerad F, Cardona O (2006) Application of indicators in urban and megacities disaster risk management: a case study of Metro Manila. EMI topical report TR-07-01. p 30. A report for this application is available from the earthquakes and megacities Initiative website. http://emi-megacities.org/upload/3cd_2007_MISReport_TR0701.pdf. Cited Mar 2014
- Khazai B, Bendimerad F (2011a) Megacity Indicator Systems (MIS) for DRM in Greater Mumbai. In: Bendimerad F, Daclan JM, Dagli W, Zayas J (eds) *Mumbai Disaster Risk Management Master Plan (DRMMP)*. Earthquake and megacities initiative final technical report No CR-1 1-0, 31, MCGM
- Khazai B, Bendimerad F (2011b) Risk and resiliency indicators. Earthquake and megacities report No. TR-1 1-03. MCGM, p 103
- Khazai B, Wenzel F, Kilic O, Basmaci A, Konukcu B, Mentese EY, Sungay B (2009) Megacity indicator systems (MIS) for disaster risk management in Istanbul. International conference on megacities: risk, vulnerability and sustainable development, 7–9 Sept 2009. Helmholtz Centre for Environmental Research – UFZ, Leipzig
- Nardo M, Saisana M, Saltelli A, Tarantola S, Hoffman A, Giovannini E (2005) Handbook on constructing composite indicators. OECD Statistics working paper. OECD, Paris
- Siddiqui KM (2007) National institute of disaster management: 2nd Asian ministerial conference on DRR high level round table 1. Available at the National Institute of Disaster Management in India. http://www.nidm.gov.in/amcd_presentations/HighLevel_Table1/BanglaDesh.pdf. Cited Mar 2014
- South Pacific Applied Geoscience Commission (SOPAC) (2002) Comprehensive Hazard and Risk Management (CHARM): guidelines for pacific Island countries. SOPAC report DM0044, Suva, 32 pp
- UNISDR (2009) 2009 UNISDR terminology on disaster risk reduction. Available at: <http://www.unisdr.org/we/inform/terminology>. Cited Mar 2014

Building Monitoring Using a Ground-Based Radar

Guido Luzi and Michele Crosetto
CTTC Division of Geomatics, Av. C.F. Gauss,
Castelldefels, Spain

Synonyms

Displacements; Measurement; Natural frequencies; Real aperture radar; Structural health monitoring

Introduction

This entry is focused on a terrestrial remote sensing technique that can provide a valuable input to earthquake engineering: the real aperture radar interferometry. In the last decades non-contacting measurement techniques based on optical sensors that can be employed for vibration monitoring have been largely described in the literature (Passia et al. 2008; Nassif et al. 2005; Gueguen et al. 2010). As an example, techniques like the Scanning Laser Doppler Vibrometry (SLDV) appear to be an appropriate non-contacting alternative to accelerometers, offering wide frequency bandwidths (0–200 kHz for the most common models) and extreme sensitivity to velocity and displacement (down to 2 pm). On the other hand, they suffer important limitations, due to constraints in the maximum observation angle and the maximum working range. In addition, they might need external acoustic stimulation to carry out measurements with adequate sensitivity, as described

in Esposito et al. (2004). Finally, they measure velocity, not displacements. Although less frequently, the use of microwave sensors to evaluate the vibration state of structures as buildings, bridges, and wind turbine towers has been investigated (Tarchi et al. 1997; Farrar et al. 1999; Pieraccini et al. 2003; Pieraccini et al. 2006; Bartoli et al. 2008; Pieraccini et al. 2008; Luzi et al. 2012a). Microwave techniques measure directly the displacement and have demonstrated to be an effective tool when the monitored displacements are down to tens of microns. Its capability to provide input data that can be used in a modal testing has been assessed, and the obtained results finally encouraged the production of radar instrumentation aimed at this application (Coppi et al. 2010). A coherent Real-Aperture Radar (RAR) is able to measure and monitor the natural vibrations of buildings, structures, and infrastructures. The main observations provided by the sensor are the displacements between the radar sensor and the observed target. The displacements can be observed at relatively high frequencies, up to 100 or 200 Hz with currently available sensors, thus providing a dense sampling of the displacements over time and space. For this reason, the technique is able to measure and monitor the vibrations of the observed objects. It is possible to derive key information of the observed objects as the main vibration frequencies, the main vibration modes, etc. from the displacement time series. There are two main potential applications of the described technique. The first and more consolidated is structural health monitoring, where RAR can provide input for the characterization of engineering structures, especially of their structural systems. This input can be provided periodically to determine the current state of system health and, in particular, to monitor its aging and degradation. In addition, the RAR can provide a valuable input after extreme events, such as earthquakes or blast loading, for rapid condition screening and assessment of the integrity of the structures at hand. The second potential application is the seismic risk evaluation of populated and industrialized areas that are seismically

active. This can be achieved through the measurement of dynamic parameters of buildings and structures, which can be done using a RAR system, and the dynamic parameters of the underlying soil in order to assess possible resonance phenomena.

This entry is organized in four main sections. The first one explains the RAR working principle and the main characteristics of the RAR measurements. The second to fourth sections describe three examples of vibration measurement: a tall tower, a building, and a bridge.

Working Principle

Radar is a well-known instrument able to detect and range objects. It sends a modulated electromagnetic signal, the transmitted pulse, along the line of sight (LOS), and receives the portion of the signal reflected by the illuminated area. Different objects or parts of the illuminated area will reflect the microwave radiation with different intensities: the basic signal provided by a radar is a profile of the intensity backscattered from the scene as a function of the distance, usually referred as “range profile.” The radar acquires simultaneously the contributions from the different parts of the structure included in its antenna field of view (FOV), and different amplitude peaks correspond to contributions from parts of the observed structure located at different distances. The FOV of the radar is dictated by the pulse width and the antenna characteristics. The elementary sampling volume of a radar measurement is usually called radar bin and different bins are used to isolate different parts of the monitored structure. Such vision of the radar prevents to see details as optical systems do, but at the same time, this averaging allows to enhance the general behavior of the structure. In addition, the radar acquires all the bins simultaneously (with respect to the observed phenomenon) thus allowing to achieve a spatial sampling useful to estimate modal shapes. Reflecting properties of a given structure are strongly related to its geometry and the dielectric characteristics of the medium: for

example, metal objects, corners, and plane surface perpendicularly oriented with respect to the radar LOS display high values of reflectivity. If the reduction of the signal strength during propagation is not dramatic, that is to say in radar nomenclature if a good signal-to-noise ratio (SNR) is preserved, the received radar signal permits to retrieve the distance of the objects. The starting point is to achieve a good acquisition, consisting in selecting an observation geometry providing a valuable range profile, i.e., a signal where the bins of interest show a strong response. The RAR is only sensible to displacements along the LOS. For this reason, another important point is the selection of geometry where the LOS and the displacement are as much parallel as possible.

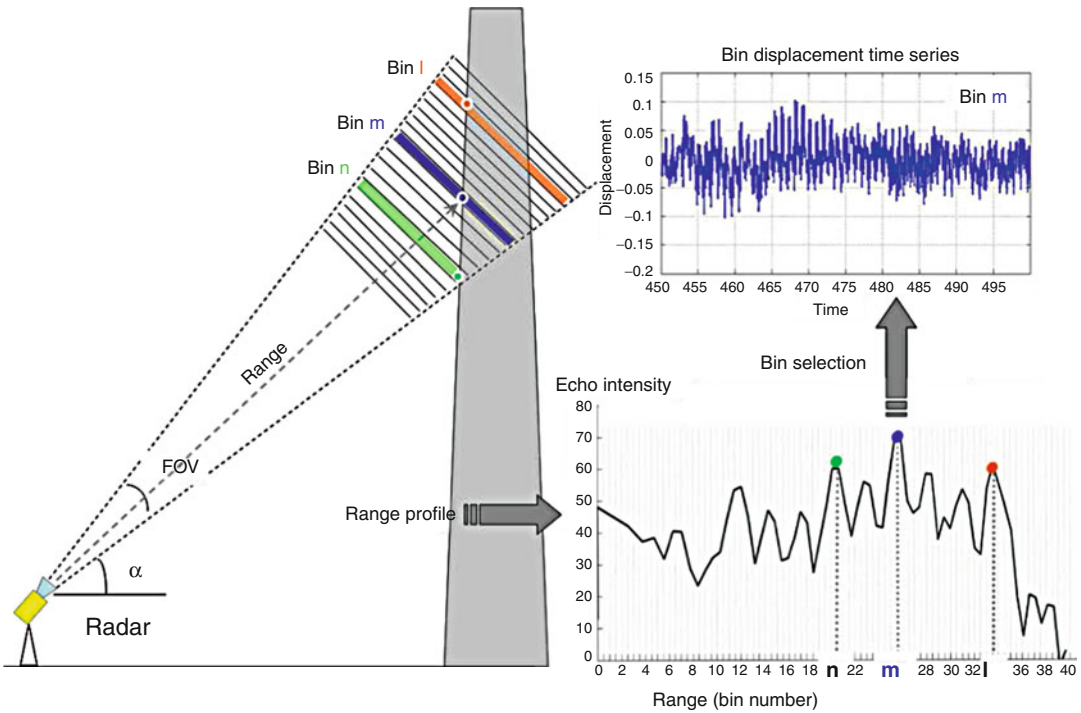
The minimum displacement detectable for conventional radar is dictated by the range resolution, i.e., the capability to distinguish different targets along the LOS, and this is usually not better than some tens of centimeters. Conventional techniques based on amplitude processing are not sufficient to measure the vibration displacements expected from building, which can reach values down to a few microns. On the other hand, interferometric techniques can be exploited to evaluate range variation along the LOS. Interferometric processing can be used only when the radar is coherent, i.e., able to measure not only the amplitude but also the phase of the received echoes. In this way, distance variations of the order of small fractions of the transmitted wavelength, which occurred between two or more acquisitions, can be estimated by comparing the phase of signals acquired at different times. A radar system designed to acquire periodically a set of received echoes, amplitude and phase values, of an entire range profile can provide a phase variation signal of each bin at the corresponding sampling frequency. Now, using the basic formula of differential interferometry, here below drawn as Eq. 1, these phase variations, $\Delta\phi(t)$, can be translated to displacement along the LOS time series, $d_{\text{LOS}}(t)$, for each bin:

$$d_{\text{LOS}}(t) = \frac{\lambda}{4\pi} \cdot \Delta\phi(t) \quad (1)$$

According to Eq. 1, the shorter the wavelength, the higher the sensitivity of the measurement: for a Ku radar, with wavelength $\lambda = 1.76$ cm (operating frequency = 17 GHz), a phase variation of 1° , typically achievable using a state-of-the-art sensor, corresponds to a displacement of ≈ 20 μm .

Equation 1 is a first approximation because it does not take into account some potential sources of error in phase estimation, but it is usually valid in the application here described; for a detailed discussion on this topic, see Luzi et al. (2012b). The standard deviation of the estimated displacements is related to the standard deviation of the measured phase, which in turn depends on the SNR of the radar measurement (Coppi et al. 2010). Vibrations with very small amplitudes, down to ten microns, can be detected when a good range profile with SNR peaks of up to 70 dB is acquired. The SNR must be statistically estimated when secondary undesirable vibrating targets (clutter) are included in the radar bin: its value decreases and the error in phase estimate increases. For this reason, and due to the high variability of the experimental conditions, it is very important to optimize the measuring conditions to acquire a good range profile.

A typical geometry used to monitor buildings is when the radar is positioned close to the base of the building, observing it along a LOS with an angle of inclination α with respect to the horizon. According to the radar equation, the operating distance and the surface characteristics of the backscattering surface strongly affect the radar response. The strength of the received echo must be maximized, case by case, selecting the best observational position. This can be achieved by choosing a distance and an angle of observation able to assure a good range profile and a radar LOS close to the main direction of the observed displacements. Then, the displacement time series can be retrieved from a range profile. The measurement procedure is graphically resumed in Fig. 1 and consists in: (1) collecting an



Building Monitoring Using a Ground-Based Radar,
Fig. 1 Scheme of a standard procedure aimed at retrieving displacement time series from a radar acquisition. The radar collects an amplitude profile as a function of the

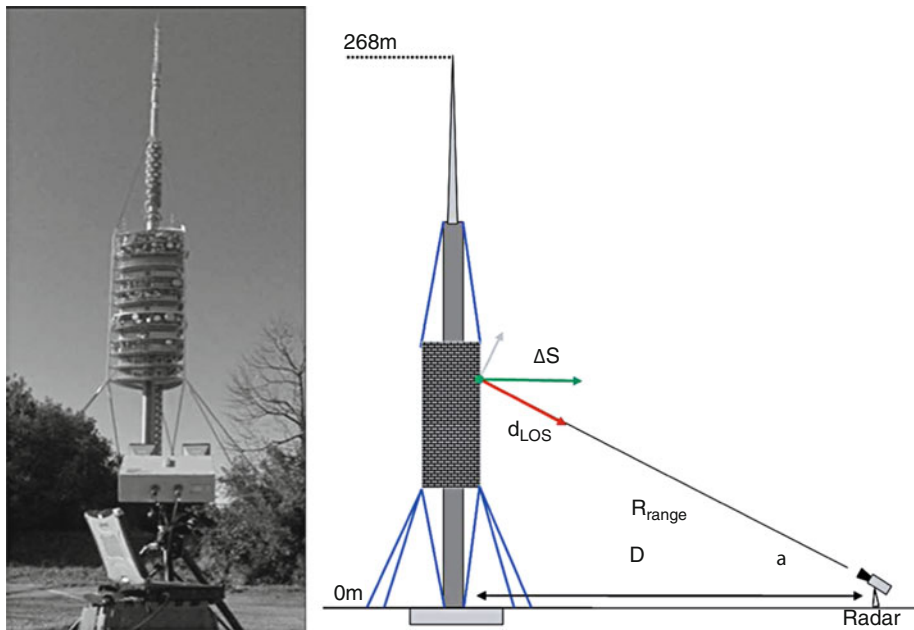
range, sampled at regular spatial steps, the radar bins. A displacement time signal is obtained from the interferometric phase of the echo of the selected bin using Eq. 1

amplitude profile as a function of the range, sampled at regular spatial steps (radar bins); (2) obtaining the interferometric phase of the echo of the selected bin; and, finally, (3) transforming the phase temporal variations into displacement time signal for analysis and processing using Eq. 1, as, for example, to calculate the spectral content of the signal.

The acquired data can be processed through conventional signal processing tools. Due to the statistical characteristics of the sampled phenomena, random and stationary, a typical approach consists in calculating the Power Spectral Density (PSD) using the Welch method (Welch 1967).

Nowadays only a few companies market radars for vibration monitoring; as a reference, the main characteristics of one of the interferometric radar most frequently cited in the literature and available at the CTTC, the IBIS-S,

manufactured and marketed by IDS (Ingegneria dei Sistemi SpA), are summarized. It is important to note that the results here discussed do not basically depend on the radar system used. IBIS-S is designed to monitor large structures as bridges and wind turbine towers. The system consists of a sensor module, a control PC, and a power supply unit. The maximum acquisition rate is currently 200 Hz. This parameter depends on the selected maximum range and decreases as the distance increases. The sensor is installed on a tripod equipped with a rotating head used to adjust the bearing of the sensor towards the investigated structure. The radar equipment is a continuous-wave (CW) step-frequency (SF) transceiver, which transmits continuous waves at discrete frequency values, sampling a bandwidth B at a constant interval. The transceiver transmits an electromagnetic signal at a central frequency of 17.2 GHz with a maximum



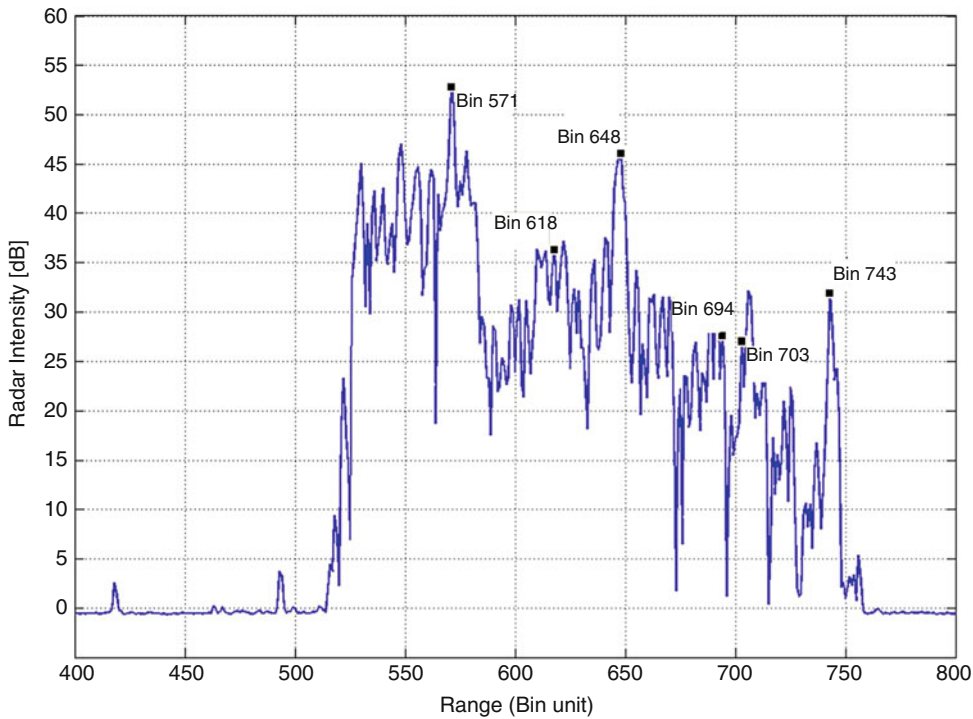
Building Monitoring Using a Ground-Based Radar, Fig. 2 Photography of the tower and the radar sensor (left) and a simple scheme of the measurement geometry

indicating the fraction of the displacement (*bold arrow*), d_{LOS} , detected by the radar (*right*)

bandwidth of 300 MHz, necessary to perform a range resolution of 0.5 m; details on the functioning principle can be found in Gentile (2011). The location and the dimension of the measured radar bins are dictated by the range resolution and the transmitting and receiving antennas. Antennas with different gain and aperture can be exchangeable; the most frequently used are pyramidal horns whose FOV is about 0.18 rad. At 10 m distance from the target, a 3 m² area is approximately illuminated, and this value increases as the observation angle with respect to the horizontal increases. The sensor unit is managed by a control PC, through a standard USB communication, which is provided with system management software used to configure the acquisition parameters, store the measurement data, and show the displacements in real time. The power supply unit provides about 5 h of autonomy. The equipment is quick and easy to install and can be used both day and night, and in almost any weather conditions.

Vibration Measurement of a Tall Tower

The first example here discussed is the monitoring of a tall tower, the Collserola communications tower, designed by the architect Norman Foster, which serves as a telecommunications tower for Barcelona and its neighborhood. The tower stands from an altitude of 445 m a.s.l., reaching a maximum 266 m height above the ground and 286 m above its base, see Fig. 2. It is made up of five main elements: a concrete shaft, a metal mast, block of platforms, pre-tensed metal guys, and fiber guys. The lower guys are steel cables anchored to concrete blocks; the remaining ones are Kevlar fiber guys. This structure represents an ultimate case from the measurement point of view for different reasons. Firstly, the tower is very high; thus, the magnitude of the amplitude of oscillation of its top, in normal wind conditions, is typically of several millimeters. Secondly, this assures a high number of bins available for data



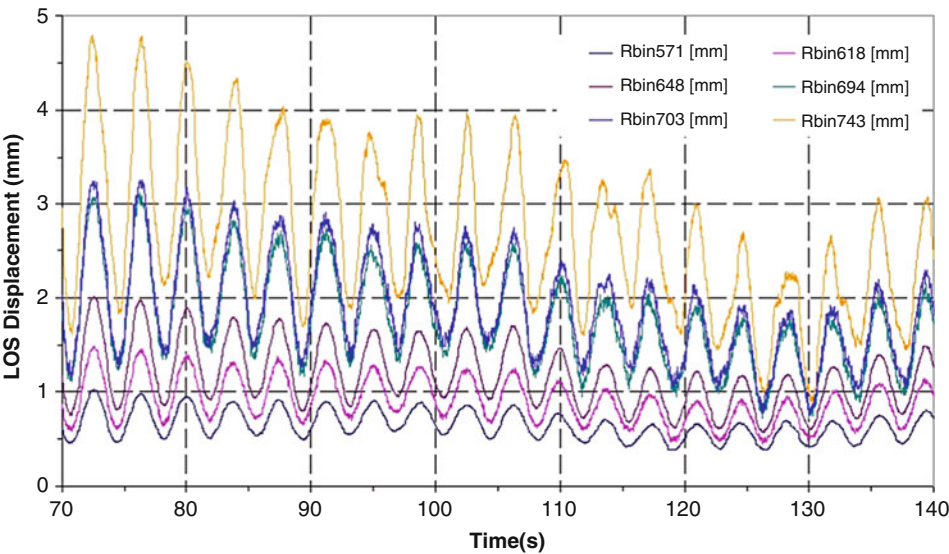
Building Monitoring Using a Ground-Based Radar,
Fig. 3 Range profile, intensity of the radar echoes as
 a function of the range expressed in bins, obtained from

a tower monitoring. A selection of bins, to be processed
 and analyzed, is marked

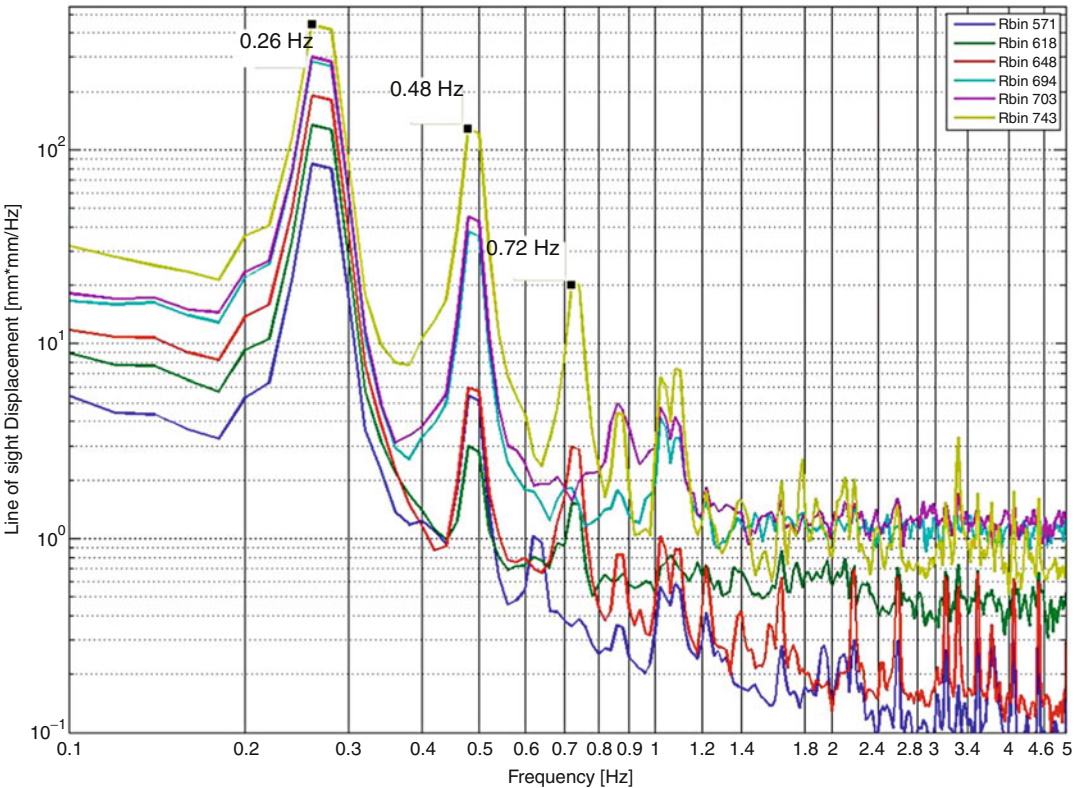
interpretation. Finally, its geometry and composing materials, metal and concrete, offer a strong radar reflectivity and this allows acquiring data from ranges up to hundreds of meters. It is worth noting that an analogous monitoring carried out with conventional point-wise sensors, e.g., accelerometers, would be unadvisable due to its high costs and the strong electromagnetic noise of this environment.

Different contributions from parts located at different heights can be clearly identified and selected for processing and observing a range profile acquired at a distance of 266 m from the tower and with an observation angle of 37° (Fig. 3). A typical approach consists in selecting a set of bins sampling the entire height of the structure and drawing the variation of the measured displacement as a function of the time. Figure 4 represents the starting point of the

analysis of the vibration behavior of the monitored structure. The curves retrieved from this plot indicate that the amplitudes of the displacement increase with the height; this trend fits the main mode of vibration of the tower, similar to the first mode expected from a clamped beam. The period of oscillation is also easily calculable in approximately 3.7 s (frequency 0.27 Hz). A spectral analysis has been carried out through the estimation of the PSD calculated using the Welch method (Welch 1967) to look for further vibration modes and deepen the analysis of the vibrating characteristics of the tower. The PSD calculated with the following configuration is plotted in Fig. 5: windowing duration = 50 s, type = Hamming, overlapping between windows: 66 % for different bins. The results obtained indicate the presence not only of the 0.27 Hz peak but also further peaks at 0.49 and



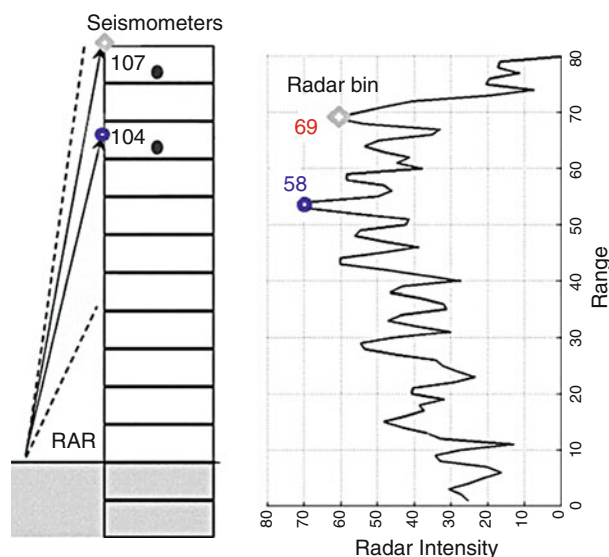
Building Monitoring Using a Ground-Based Radar, Fig. 4 Example of a 70 s sample of the displacement of different bins retrieved from radar measurements as a function of the time



Building Monitoring Using a Ground-Based Radar, Fig. 5 PSD calculated for the bins shown in Fig. 4. Different colors refer to different bins. The three peaks related to tower vibration are marked

Building Monitoring Using a Ground-Based Radar, Fig. 6

Location of the seismometers (#104 and #107) with respect to the building, and radar bins (#53 and #69) selected from the range profile, to perform a radar data validation



0.72 Hz. Some minor peaks, probably caused by the cables and its interaction with the tower, are also present.

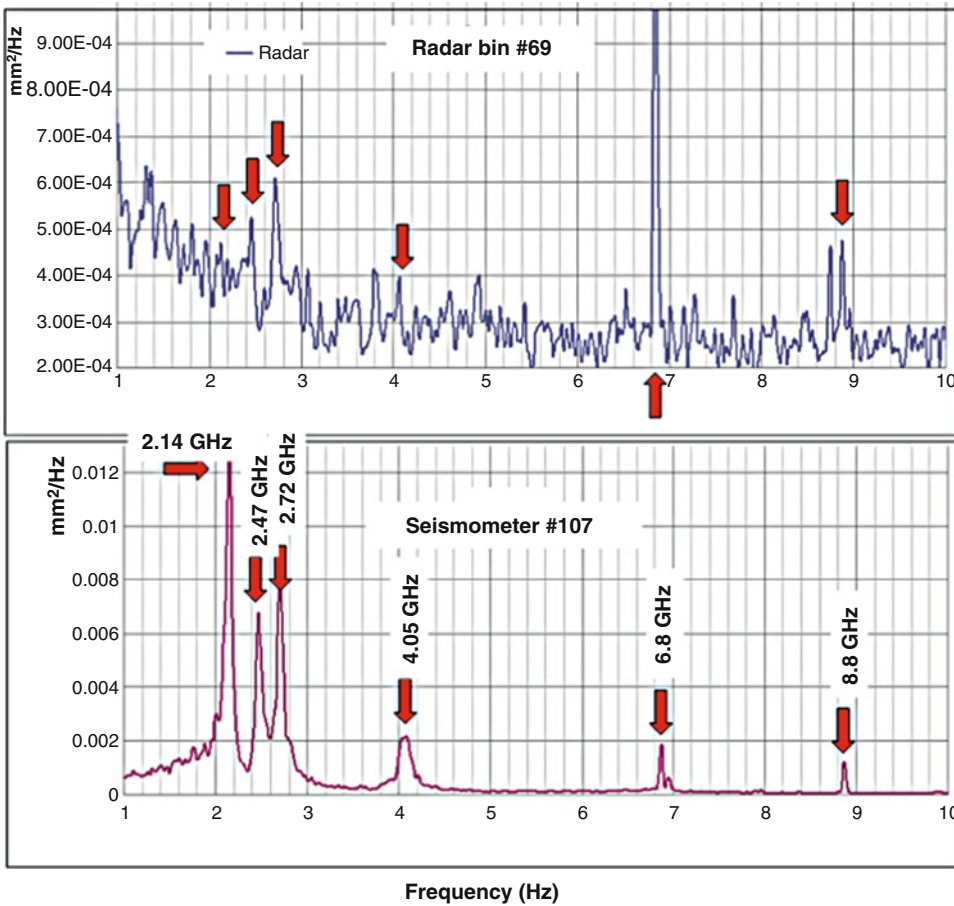
Case studies aimed at estimating vibration characteristics of towers of different shapes and heights, selected among architectural heritages, have been carried out by different authors (Pieraccini et al. 2005; Atzeni et al. 2010; Gentile and Saisi 2011).

Comparison of Radar Observations and Conventional Measurements: The Case of a Building

Radar observations can be integrated with some standard point-wise measurements as a support to model calculation of the modal shapes of the investigated structures. This is the case described in Negulescu et al. (2013), where a building, part of a sportive complex located in Font Romeu (France), was monitored by means of radar and several accelerometers, and the results obtained were compared with a model calculation. The building has a reinforced concrete structure, with double basement, ground floor, 10 storeys, and a total height of 33 m (not including the basement). In this study, several sensors were

placed, in two different configurations, on the roof and at different heights of the structure. The experiment was carried out with the aim of comparing seismometer data with the displacement retrieved from several radar bins. In particular, a direct comparison was carried out between the data obtained using two radar bins and the closest sensors installed in the building, whose position is shown in Fig. 6, where the geometry of the measurement is also depicted. The seismometer acquisition set includes eight sensors and a wireless networking connection was used during the measurements. The normalized PSD calculated for the displacement signal retrieved from the radar measurements and that obtained from microtremor signal are compared in Fig. 7.

The location of the main peaks is almost coincident and the radar signal shows some minor peaks too. The amplitude is a little different; this can be imputed to the fact that the radar only measures the LOS component of the displacement and the two sensors are not exactly positioned in the same point. The agreement between radar observations and seismometer measurements was also confirmed by numerical model calculation and modal shape estimation (Negulescu et al. 2013).



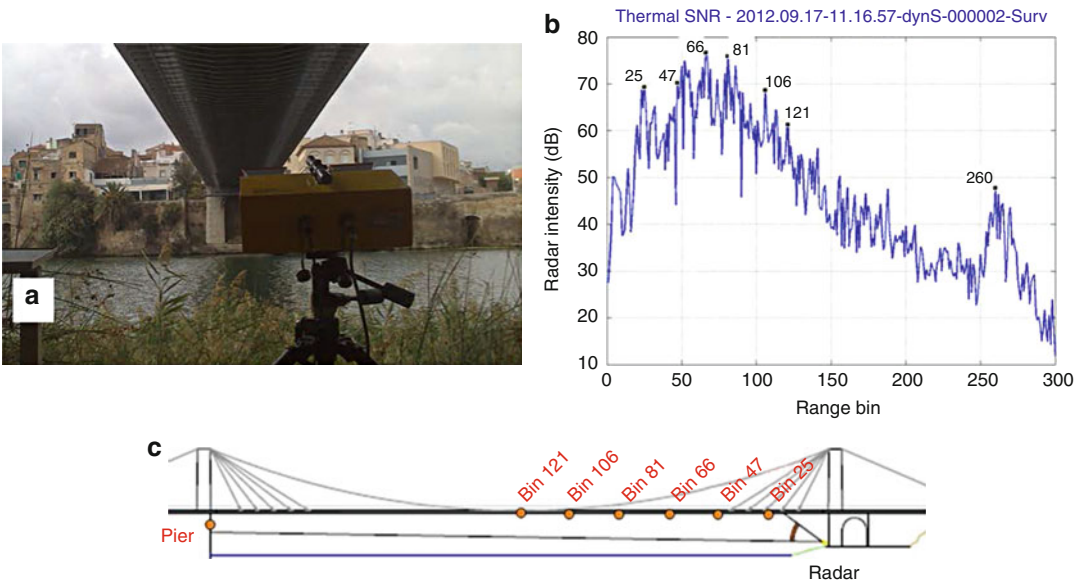
Building Monitoring Using a Ground-Based Radar, Fig. 7 Comparison of the normalized Power Spectral Density (PSD) functions from the records obtained using the radar technique, bin #69 (*top*) and the seismometer #107 (*down*)

Vibration Measurement of a Bridge

Bridge monitoring is an application where the technique has been thoughtfully studied and validated (Gentile 2010). Radar can be used for both static (Pieraccini et al. 2007) and dynamical studies (Stabile et al. 2013). Cable-stayed bridges can especially benefit from radar monitoring, with moderate time and cost recourses with respect to the use of conventional sensors. In this case, the radar monitoring is usually carried out with two different geometries. The first one aims at detecting the main features of the deck vibration, with the radar positioned below the bridge deck acquiring a range profile where the different bins provide a spatial profile of deflection which can occur both due to the wind and the ambient

solicitation as microtremor. The second position is carefully selected to detect the behavior of the stay cables avoiding the direct influence of the deck vibration. In this second case, the main objective is to acquire a range profile where the different stay cables are clearly identified. Figures 8 and 9 show the two geometries used during the monitoring of a cable-stayed bridge in Amposta (Spain), over the Ebro River.

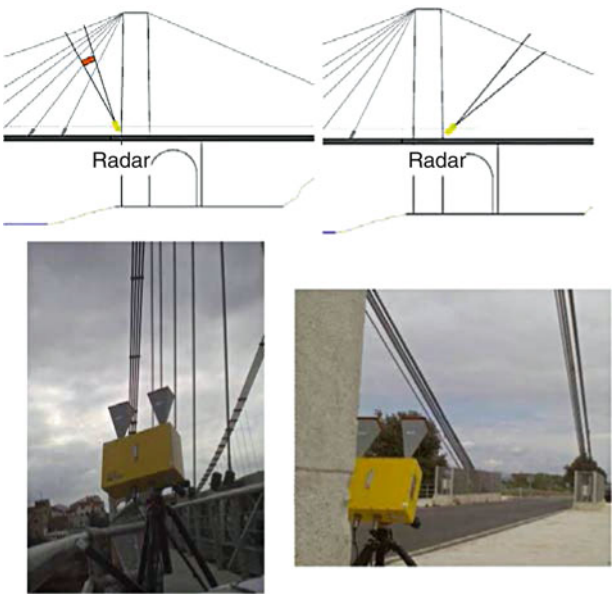
A typical signal acquired during the transit of a vehicle using the geometry depicted in Fig. 8 is shown in Fig. 10. The displacement responses of different bins of the bridge, corresponding to a sampling of half of the bridge deck, are jointly plotted. From Fig. 10 the difference in the deflection for the different bins is noticeable: the central part of the bridge (bin #81) lowers down to 4 mm,



Building Monitoring Using a Ground-Based Radar, Fig. 8 Photography of the radar pointing towards the deck (a); range profile acquired (b); scheme of the

geometry of acquisition indicating the selected bins: bin 25, 47, 66, 81, 106, and 121 belong to the bridge deck, while bin 260 corresponds to the pier (c)

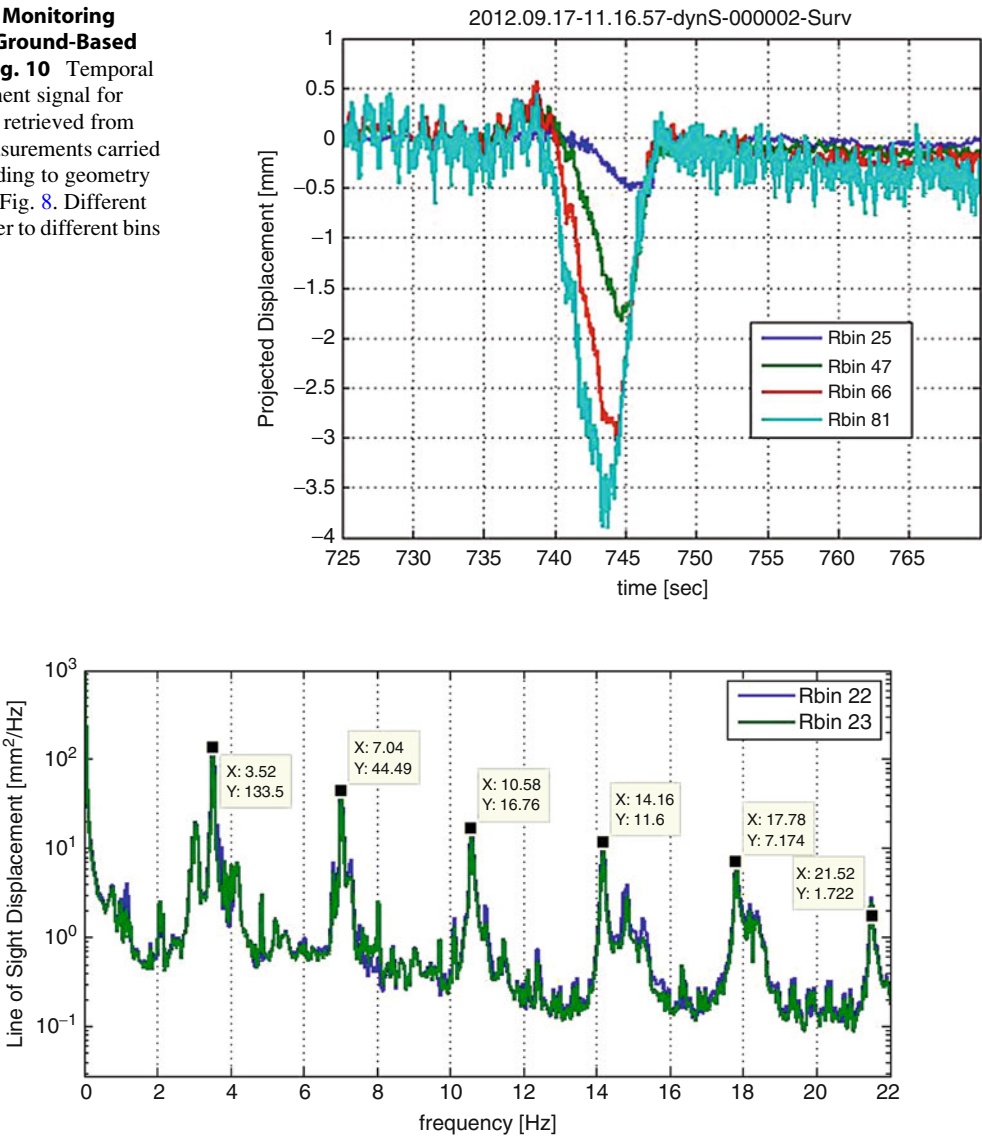
Building Monitoring Using a Ground-Based Radar, Fig. 9 Photos and geometry of the acquisitions with the radar pointing towards the inner stay cables positioned within the two pillars (*left*) and watching the set of guys tied to the soil (*right*)



while the part close to the pier (bin#25) experiences a 500 μm lowering. Using the second geometrical configuration (Fig. 9), the radar observation is focused towards the cables. In this case, the measurement of the natural frequency of the cable can be used to estimate the

tension force of the cable (applying a formula based on the tau string model; see Caetano and Cunha 2011; Gentile and Ubertini 2012; Luzi et al. 2014). The calculation of the PSD of the displacement signal allows detecting the main vibrating frequencies of the cables. Figure 11

Building Monitoring Using a Ground-Based Radar, Fig. 10 Temporal displacement signal for four bins, retrieved from radar measurements carried out according to geometry shown in Fig. 8. Different colors refer to different bins



Building Monitoring Using a Ground-Based Radar, Fig. 11 PSD calculated for two bins corresponding to two stay cables. Different colors refer to different bins

shows the results obtained from the simultaneous measurement of the displacements of two bins corresponding to two cables.

The values of the main peaks displayed in Fig. 11, 3.52, 7.04, 10.58, 14.16, 17.78, and 21.52, show that they are related to each other by a linear relationship. In fact, they are natural multiples of the fundamental frequency, 3.52 Hz, although the higher values gradually deviate. The first frequencies are in agreement with the values

expected from the tau string model, thus allowing the application of the model and the estimation of the tension force of the cables. It is worth noting that obtaining the same information using standard local sensors as accelerometer should be unfeasible. Besides, the knowledge of the spectral behavior of all the stay cables can also be used for evaluating the health state of the structure or comparing these experimental results with model analysis provided by structural engineering analysis.

Summary

In this entry a radar technique applied to the monitoring of vibrations of structures as tower, buildings, and bridges has been introduced. After presenting the working principle and briefly describing a commercial radar system, a measurement procedure underlying the main critical aspects has been summarized. Experimental results from three case studies are then commented. The first is the monitoring of a tall tower which, thanks to the strength of the radar response and the millimeter amplitude of oscillation of its top, allows achieving abundant information. This monitoring also demonstrated that, in the best conditions, the technique is capable of acquiring data at ranges up to hundreds of meters. The second example shows a case where radar data has been validated using a network of standard point-wise measurements (seismometers). The last case study concerns the monitoring of a cable-stayed bridge, an application thoughtfully studied and validated by several authors. In this last case, the technique also allows detecting the vibration characteristics of the cables with moderate time and cost resources.

Cross-References

- [Ambient Vibration Testing of Cultural Heritage Structures](#)
- [Geotechnical Earthquake Engineering: Damage Mechanism Observed](#)
- [Modal Analysis](#)
- [Remote Sensing in Seismology: An Overview](#)

References

- Atzeni C, Bicci A, Dei D, Fratini M, Pieraccini M (2010) Remote survey of the leaning tower of Pisa by interferometric sensing. *IEEE Trans Geosci Remote Sens Lett* 7(1):185–189
- Bartoli G, Facchini L, Pieraccini M, Fratini M, Atzeni C (2008) Experimental utilization of interferometric radar techniques for structural monitoring. *Struct Control Health Monit* 15(3):283–298
- Caetano E, Cunha Á (2011) On the observation and identification of cable-supported structures. In: *Proceedings of the 8th international conference on structural dynamics EUROdyn 2011*. Leuven, Belgium, 4–6 July 2011
- Coppi F, Gentile C, Ricci P (2010) A software tool for processing the displacement time series extracted from raw radar data. In: Tomasini EP (ed) *Proceedings of the 9th international conference on vibration measurements by laser and non-contact techniques and short course*, Ancona, Italy; 22–25 AIP conference proceedings 1253, June 2010
- Eposito E, Copparoni S, Naticchia B (2004) Recent progress in diagnostics of civil structures by laser vibrometry. In: *Proceedings of the 16th World conference on nondestructive testing*, on CD, Montreal, Canada, August 30–September 3, 2004
- Farrar C, Darling TW, Migliorini A, Baker WE (1999) Microwave interferometer for non-contact vibration measurements on large structures. *Mech Syst Signal Process* 13(2):241–253
- Gentile C (2010) Deflection measurement on vibrating stay cables by non-contact microwave interferometer. *NDT & E Int* 43:231–240
- Gentile C (2011) Vibration measurement by radar techniques. In: De Roeck G, Degrande G, Lombaert G, Müller G (eds) *Proceedings of the 8th international conference on structural dynamics, EUROdyn 2011* Leuven, Belgium, 4–6 July 2011 ISBN 978-90-760-1931-4
- Gentile C, Saisi A (2011) Dynamic measurement on historic masonry towers by microwave remote sensing. In: Gentile C, Benedettini F (eds) *Proceeding of the international conference on experimental vibration analysis for civil engineering structures*, vol II, pp 524–530, 3–5 October 2011 Varenna, Italy. ISBN 978-88-96225-39-4
- Gentile C, Ubertini F (2012) Radar-based dynamic testing and system identification of a Guyed Mast. In: *10th international conference on vibration measurements by laser and noncontact techniques – AIVELA 2012* Ancona, Italy 27–29 June 2012, ISBN 978-0-7354-1059-6 ISSN 0094-243X, pp 318–325
- Gueguen P, Jolivet V, Clotaire M, Schweitzer AS (2010) Comparison of velocimeter and coherent lidar measurements for building frequency assessment. *Bull Earthquake Eng* 8:327–338. doi:10.1007/s10518-009-09137-2
- Luzi G, Monserrat O, Crosetto M (2012a) The potential of coherent radar to support the monitoring of the health state of buildings. *Research in Non-destructive Evaluation* 23(3):125–145 <http://dx.doi.org/10.1080/09349847.2012.660241>
- Luzi G, Monserrat O, Crosetto M (2012b) Real aperture radar interferometry as a tool for buildings vibration monitoring: limits and potentials from an experimental study. In: *10th international conference on vibration measurements by laser and noncontact techniques – AIVELA 2012* Ancona, Italy 27–29 June 2012, ISBN 978-0-7354-1059-6 ISSN 0094-243X, pp 309–317

- Luzi G, Crosetto M, Cuevas-González M (2014) A radar-based monitoring of the Collserola tower (Barcelona). *Mech Syst Signal Process* 49:234–248
- Nassif H, Gindy M, Davis J (2005) Comparison of laser Doppler vibrometer with contact sensors for monitoring bridge deflection and vibration. *NDT & Int* 38(3):213–218
- Negulescu C, Luzi G, Crosetto M, Raucoules D, Roullé A, Monfort D, Pujades L, Colas B, Dewez T (2013) Comparison of seismometer and radar measurements for the modal identification of civil engineering structures. *Eng Struct* 51:10–22
- Passia H, Staniek A, Szade A, Motyka Z, Bochenek W, Bartmański C (2008) Comparative studies of principal vibration parameters of a building using an LDV, laser tilt and vibration sensors and piezoelectric accelerometers. In: Tomasini EP (ed) 8th international conference on vibration measurements by laser techniques: advances and applications, Proc. of SPIE vol 7098, 70980J, (2008) 0277-786X/08/\$18 doi: 10.1117/12.803001
- Pieraccini M, Luzi G, Mecatti D, Noferini L, Atzeni C (2003) A microwave radar technique for dynamic testing of large structure. *IEEE Trans Microwave Theory Tech* 51(5):1603–1609
- Pieraccini M, Fratini M, Parrini F, Pinelli G, Atzeni C (2005) Dynamic survey of architectural heritage by high-speed microwave interferometry. *IEEE GRSL* 2(1):28–30
- Pieraccini M, Fratini M, Parrini F, Atzeni C (2006) Dynamic monitoring of bridges using a high-speed coherent radar. *IEEE Trans Geosci Remote Sens* 40(11):3284–3288
- Pieraccini M, Parrini F, Fratini M, Atzeni C, Spinelli P, Micheloni M (2007) Static and dynamic testing of bridges through microwave interferometry. *NDT & E Int* 40(3):208–214. doi:10.1016/j.ndteint.2006.10.007
- Pieraccini M, Fratini M, Parrini F, Atzeni C, Bartoli G (2008) Interferometric radar vs. accelerometer for dynamic monitoring of large structures: an experimental comparison. *NDT & E Int* 41(4):258–264
- Stabile TA, Perrone A, Gallipoli MR, Ditommaso R, Ponzo FC (2013) Dynamic survey of the Musmeci bridge by joint application of ground-based microwave radar interferometry and ambient noise standard spectral ratio techniques. *IEEE GRSL* 10(4):870–874. doi:10.1109/LGRS.2012.2226428
- Tarchi D, Ohlmer E, Sieber A (1997) Monitoring of structural changes by radar interferometry. *Res Nondestruct Eval* 9(4):213–225
- Welch PD (1967) The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. *IEEE Trans Audio Electroacoust* 15(2):70–73

Buildings and Bridges Equipped with Passive Dampers Under Seismic Actions: Modeling and Analysis

Theodore L. Karavasilis, George S. Kamaris and Angelos S. Tzimas
School of Engineering, University of Warwick, Coventry, UK

Synonyms

Bridges; Numerical simulation; Passive dampers; Steel buildings; Structural design

Introduction

The scope of this contribution is the presentation of numerical models and analysis methods for structures equipped with passive dampers. The principal function of passive dampers and their application in buildings and bridges will be discussed. The most common types of passive dampers and the constitutive models used to simulate their hysteretic behavior will be presented. Finally, the analysis methods for seismic design and assessment of structures with passive dampers are covered.

Passive dampers include a variety of materials and devices that can enhance the damping, stiffness, and strength characteristics of structures. Passive dampers are suitable both for retrofitting existing vulnerable structures and for designing new ones and have been developed for a number of years with a rapid increase in applications in the mid-1990s. Passive dampers can significantly reduce plastic deformation demands on the primary lateral load-resisting system in addition to decreasing velocity and acceleration demands on non-structural components (Soong and Spencer 2002; Symans et al. 2008). Different types of passive dampers have been developed for seismic protection of structures such as viscous fluid dampers, viscoelastic solid dampers, elastomeric dampers, friction dampers, and metallic dampers.

Other devices that could be classified as passive dampers include tuned mass and tuned liquid dampers.

Passive dampers are typically used within conventional lateral load resisting systems to dissipate a significant portion of the seismic input energy (Christopoulos and Filiatrault 2006). The degree to which a passive damper is able to achieve this goal depends on the inherent properties of the lateral load resisting system, the properties of the passive damper and its connecting elements, the characteristics of the ground motion, and the limit state being investigated (Symans et al. 2008). Over the past decades, passive dampers have been specified for application to buildings and bridges with a wide variety of structural configurations. The growth in application of passive dampers has been steady to the extent that there are now numerous applications (Soong and Spencer 2002).

The first design/analysis procedure for buildings with passive dampers was published in SEAONC (Whittaker et al. 1993). The intent of these guidelines was to direct engineers in locating dampers in the lateral load resisting system to reduce repair costs and business interruption due to a severe earthquake. In the mid-1990s, the Federal Emergency Management Agency (FEMA) funded the development of guidelines for the seismic rehabilitation of buildings proposing, among others, basic principles for the seismic design with passive dampers. The outcome of this effort was the *NEHRP Guidelines for the Seismic Rehabilitation of Buildings*, FEMA Reports 273 and 274 (ATC 1997a, b). Later on, the 2003 *NEHRP Recommended Provisions* (BSSC 2004) included a chapter on design and analysis methods for buildings with passive dampers. These methods were reformatted and included in the 2005 edition of the ASCE/SEI 7-05 standard *Minimum Design Loads for Buildings and Other Structures* (ASCE 2005). The 2003 *NEHRP Recommended Provisions* (and the current 2010 ASCE/SEI 7 standard (ASCE 2010)) allow a reduced design base shear force for the seismic design of buildings with passive

damping systems where the expected performance is similar or higher than that of buildings with a conventional lateral load resisting system.

Recent analytical and experimental research showed that steel moment-resisting frames (MRFs) with elastomeric dampers can be designed to be lighter and perform better than conventional steel MRFs under the design basis earthquake (DBE; 475-year return period) (Karavasilis et al. 2011a, 2012a). However, it was shown that it is generally not feasible to design steel MRFs with passive dampers at a practical size to eliminate inelastic deformations in main structural members under the DBE (Karavasilis et al. 2011a). To address this issue, a seismic design strategy for steel MRFs that isolates damage in removable steel dampers and uses in parallel viscous fluid dampers to reduce drifts has been proposed by Karavasilis et al. (2012b). A study shows that supplemental viscous damping does not always ensure adequate reduction of residual drifts (Karavasilis and Seo 2011). A recent work evaluates the seismic collapse resistance of steel MRFs with viscous fluid dampers and shows that supplemental viscous damping does not always guarantee a better seismic collapse resistance when the strength of the steel MRF with dampers is lower or equal to 75 % of the strength of a conventional steel MRF (Seo et al. 2014).

In the following section, two major real-world applications of passive dampers in a tall building and a bridge are briefly discussed. The most common types of passive dampers and the models used to simulate their hysteretic behavior are then presented. Finally, the analysis methods that can be utilized to predict the response of structures with passive dampers are discussed.

Application of Passive Dampers in Buildings and Bridges

Two relatively recent applications of passive dampers in a building and a bridge are discussed below.



Buildings and Bridges Equipped with Passive Dampers Under Seismic Actions: Modeling and Analysis, Fig. 1 (a) Torre Mayor Tower under

construction showing partial view of megabraces; (b) view of installed viscous fluid dampers (Symans et al. 2008)

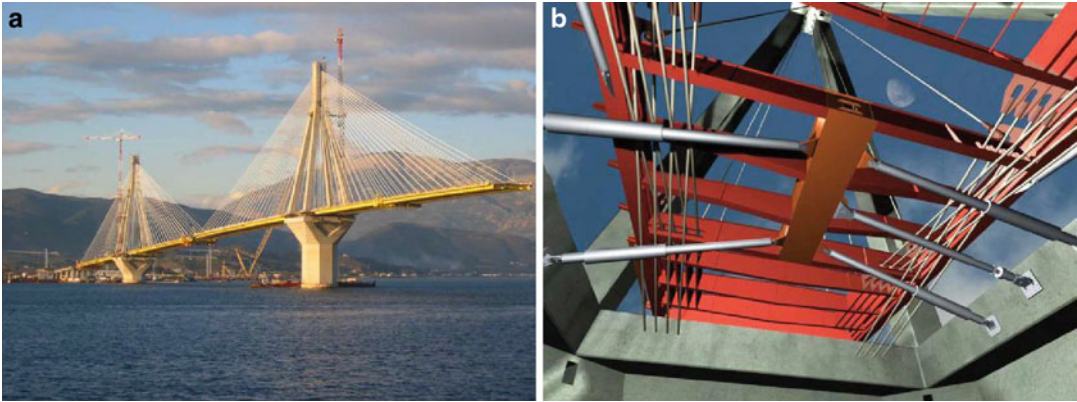
Torre Mayor Tower, Mexico

The Torre Mayor Tower is a 77,000 m², 57-story steel and reinforced concrete office/hotel tower in Mexico City. Its construction was completed in 2003 and its height is 225 m. The lower half of the superstructure is a rectangular tower which consists of steel framing encased in concrete, while the upper half consists of steel framing (Fig. 1a). The seismic design of the structure had the challenging objective of achieving “operational” performance for an expected 8.2 magnitude seismic event (Rahimian and Romero 2003). The seismic design employs 96 large nonlinear viscous fluid dampers with 2,670 and 5,340-kN force output capacities (Fig. 1b). The dampers on the faces of the building are installed in megabraces, which are diagonal braces that span over more than one story, as shown in Fig. 1a. Note that, a first design of the building without dampers resulted in weight of the building that was excessive for the soil capacity. The addition of the dampers resulted in less steel weight and acceptable soil bearing pressure. The structure experienced no damage during a seismic event with 7.8

magnitude and epicenter 500 km from the building site (Rahimian and Romero 2003).

Rion-Antirion Bridge, Greece

The Rion-Antirion Bridge (Fig. 2), that crosses the homonymous strait in Greece, is located in the very active seismic region of the Gulf of Corinth. The deck of this multi-span cable-stayed bridge is continuous and fully suspended from four pylons (total length of 2,252 m). Its approach viaducts comprise 228 m of concrete deck on the Antirion side and 986 m of steel composite deck on the Rion side. Said structures are designed to withstand seismic events generating ground accelerations of up to 0.48 g through the use of viscous fluid dampers and other seismic devices. The Main Bridge seismic protection system consists of fuse restraints and viscous fluid dampers that act in parallel, connecting the deck to the pylons (Fig. 2b). The fuse restraints are designed as a rigid link intended to withstand high wind loads up to a predetermined force. Under the action of the design earthquake, the fuse restraints will fail and leave the dampers free to dissipate the semi



Buildings and Bridges Equipped with Passive Dampers Under Seismic Actions: Modeling and Analysis, Fig. 2 Rion-Antirion Bridge: (a) main bridge

under construction; (b) arrangement of viscous fluid dampers and fuse restraints on the main piers (Infanti et al. 2004)

induced energy. The approach viaducts are seismically isolated utilizing elastomeric isolators and viscous dampers.

Passive Dampers and Constitutive Models

Passive dampers can be classified in three main categories: (1) velocity-activated or rate-dependent, (2) displacement-activated or rate-independent, and (3) motion-activated (Symans et al. 2008; Christopoulos and Filiatrault 2006).

Velocity-activated or rate-dependent dampers have force output that depends on the relative velocity across the damper. Therefore, their efficiency to dissipate energy depends on the frequency of the input ground motion. The behavior of such dampers is commonly described using various models of linear viscoelasticity. Typical dampers falling in this category include viscous fluid dampers and viscoelastic dampers (Symans et al. 2008; Christopoulos and Filiatrault 2006).

Displacement-activated or rate-independent dampers have force output that does not depend on the relative velocity across the damper but on the magnitude of the displacement and possibly the direction of motion. Their behavior is

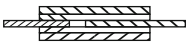
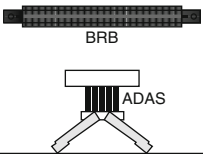
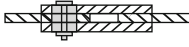
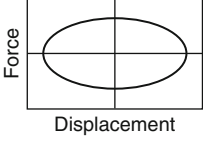
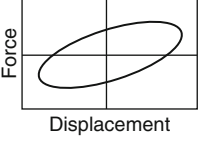
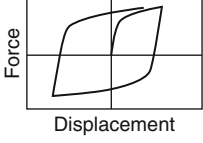
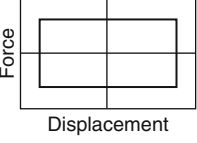
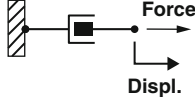
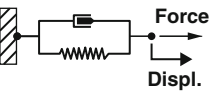
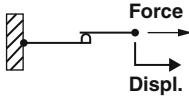
commonly described using various nonlinear hysteretic models. Typical dampers falling in this category include metallic dampers and friction dampers (Symans et al. 2008; Christopoulos and Filiatrault 2006).

Motion-activated passive dampers consist of secondary systems that disturb the flow of energy in the structure. These systems have the same resonance period with the main structure. Tuned-mass dampers (TMDs) are an example of motion-activated dampers. This last category is not discussed further in this contribution, but the reader is referred to documents such as the one by Soong and Spencer (2002) for more details.

A summary of velocity-activated and displacement-activated passive dampers is provided in Fig. 3 (reproduced from Symans et al. (2008)). Figure 3 lists the basic construction, the idealized hysteretic response and the associated physical model, and the major advantages and disadvantages of each passive damper.

Viscous Fluid Dampers

Viscous fluid dampers consist of a hollow cylinder filled with fluid (Fig. 3), the fluid typically being silicone based. As the damper piston rod and piston head are stroked, fluid is forced to flow through orifices either around or through the piston head. The resulting differential in pressure

	Viscous Fluid Damper	Viscoelastic Solid Damper	Metallic Damper	Friction Damper
Basic Construction				
Idealized Hysteretic Behavior				
Idealized physical Model			Idealized Model Not Available	
Advantages	- Activated at low displacements - Minimal restoring force - For linear damper, modeling of damper is simplified - Properties largely frequency and temperature-independent - Proven record of performance in military applications	- Activated at low displacements - Provides restoring force - Linear behavior, therefore simplified modeling of damper	- Stable hysteretic behavior - Long-term reliability - Insensitivity to ambient temperature - Materials and behavior familiar to practicing engineers	- Large energy dissipation per cycle - Insensitivity to ambient temperature
Disadvantages	- Possible fluid seal leakage (reliability concern)	- Limited deformation capacity - Properties are frequency and temperature-dependent - Possible debonding and tearing of VE material (reliability concern)	- Device damaged after earthquake; may require replacement - Nonlinear behavior; may require nonlinear analysis	- Sliding interface conditions may change with time (reliability concern) - Strongly nonlinear behavior; may excite higher modes and require nonlinear analysis - Permanent displacements if no restoring force mechanism provided

Buildings and Bridges Equipped with Passive Dampers Under Seismic Actions: Modeling and Analysis, Fig. 3 Summary of basic construction,

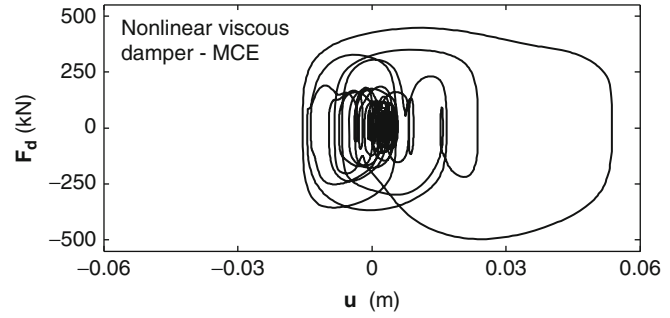
hysteretic behavior, physical models, advantages, and disadvantages of passive dampers for seismic protection applications (Symans et al. 2008)

across the piston head can produce very large forces that resist the relative motion of the damper. The fluid flows at high velocities, resulting in the development of friction between

fluid particles and the piston head, which leads to energy dissipation in the form of heat. Interestingly, although the damper is called a *viscous fluid damper*, the fluid typically has a relatively

Buildings and Bridges Equipped with Passive Dampers Under Seismic Actions: Modeling and Analysis,

Fig. 4 Hysteretic behavior of a nonlinear viscous damper from nonlinear dynamic analysis (Karavasilis et al. 2012b)



low viscosity, e.g., silicone oil with a kinematic viscosity in the order of $0.001 \text{ m}^2/\text{s}$ at 20°C . The term *viscous fluid damper* is associated with the macroscopic behavior of the damper, which is essentially the same as that of an ideal linear or nonlinear viscous dashpot, i.e., the resisting force is directly related to the velocity (Symans et al. 2008).

Experimental testing (Seleemah and Constantinou 1997) has shown that the behavior of viscous fluid dampers corresponds to a nonlinear viscous dashpot and can be modeled by the following nonlinear force-velocity relation:

$$P(t) = C|\dot{u}(t)|^a \text{sgn}[\dot{u}(t)] \quad (1)$$

where $P(t)$ is the force developed by the damper, $u(t)$ is the displacement across the damper, C is damping coefficient, a is an exponent whose value is determined by the piston head orifice design, $\text{sgn}[\cdot]$ is the signum function, and the overdot indicates differentiation with respect to time, t . For earthquake protection applications, the exponent a typically has a value ranging from about 0.3 to 1.0. For a equal to unity, the damper can be described as an ideal linear viscous dashpot (Symans et al. 2008). Figure 4 shows the behavior of a nonlinear viscous damper installed in a prototype steel building subjected to a strong earthquake ground motion from nonlinear dynamic analysis conducted in Karavasilis et al. (2012b).

Viscoelastic Solid Dampers

Viscoelastic solid dampers generally consist of solid elastomeric pads (viscoelastic material)

bonded to steel plates, which are attached to the structure within chevron or diagonal bracing (Fig. 3). As one end of the damper displaces with respect to the other, the viscoelastic material is sheared resulting in the development of heat, which is dissipated to the environment. Viscoelastic dampers provide both a velocity-dependent force, which provides supplemental viscous damping to the system (similar to viscous dampers), and a displacement-dependent elastic restoring force (Symans et al. 2008).

Experimental testing (Bergman and Hanson 1993) has shown that, under certain conditions, a suitable mathematical model for simulating the mechanical behavior of viscoelastic dampers is the Kelvin model of viscoelasticity. This model is described by the following equation:

$$P(t) = Ku(t) + C\dot{u}(t) \quad (2)$$

where K is the storage stiffness of the damper and C is the damping coefficient which is equal to the ratio of the loss stiffness to the frequency of loading. The physical model corresponding to Eq. 2 is a linear spring in parallel with a linear viscous dashpot.

A constitutive model that can be used to describe with more accuracy the behavior of viscoelastic dampers is the generalized Maxwell (GM) model. The GM model consists of a parallel combination of a linear spring, a dashpot and multiple in-series combinations of springs and dashpots as shown in Fig. 5 (Karavasilis et al. 2011b). Figure 5 shows the GM model in terms of shear stress, τ , and shear strain, γ , of the viscoelastic material. Under harmonic loading of

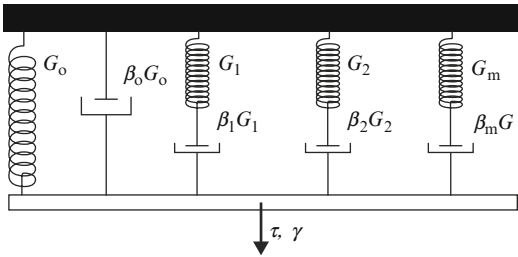
cyclic frequency ω , the GM model provides storage shear modulus equal to

$$G'(\omega) = G_0 + \sum_{m=1}^n \frac{(\omega\beta_m)^2}{1 + (\omega\beta_m)^2} \cdot G_m \quad (3)$$

and loss factor equal to

$$\eta(\omega) = \frac{(\omega\beta_0)G_0 + \sum_{m=1}^n \frac{(\omega\beta_m)}{1 + (\omega\beta_m)^2} \cdot G_m}{G_0 + \sum_{m=1}^n \frac{(\omega\beta_m)^2}{1 + (\omega\beta_m)^2} \cdot G_m} \quad (4)$$

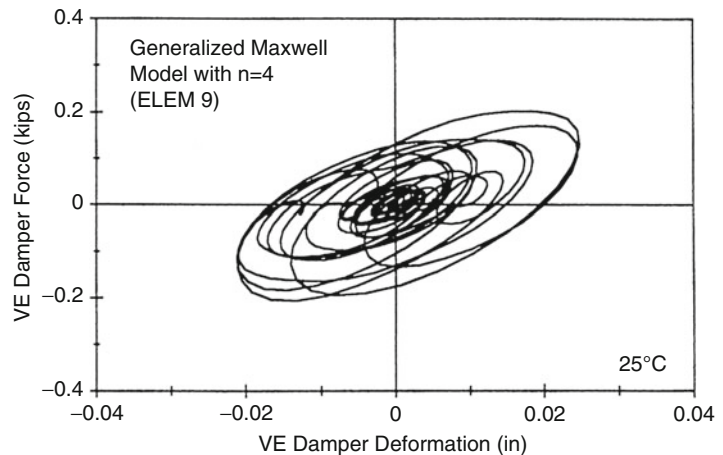
These equations can be used to calibrate the GM model against experimentally obtained values of G' and η for different cyclic frequencies. To use the GM model as a combination of linear springs and dashpots within a structural analysis



Buildings and Bridges Equipped with Passive Dampers Under Seismic Actions: Modeling and Analysis, Fig. 5 Generalized Maxwell model for viscoelastic dampers

Buildings and Bridges Equipped with Passive Dampers Under Seismic Actions: Modeling and Analysis,

Fig. 6 Hysteretic behavior of viscoelastic damper from nonlinear dynamic analysis (Fan 1998)



software, simply multiply G_0 and G_m in Fig. 5 with the ratio A_d/t_d (A_d : horizontal area of the viscoelastic material; t_d : thickness of the viscoelastic material) to transform τ - γ behavior to force ($F = \tau \cdot A_d$) – deformation ($d = \gamma \cdot t_d$) behavior.

Figure 6 shows the hysteretic behavior of a typical viscoelastic damper from nonlinear dynamic analysis conducted in Fan (1998).

Metallic Dampers

Metallic dampers are designed to yield, based on the ability of steel to dissipate energy. Two major types of metallic dampers are the buckling-restrained brace (BRB) and the added damping and stiffness (ADAS) device.

A BRB consists of a steel brace with a cruciform cross section that is surrounded by a stiff steel tube. The region between the tube and brace is filled with a concrete-like material and a special coating is applied to the brace to prevent it from bonding to the concrete. Thus, the brace can slide with respect to the concrete-filled tube. The confinement provided by the concrete-filled tube allows the brace to be subjected to compressive loads without buckling. Under compressive loads, the BRB behavior is essentially identical to its behavior in tension. Since buckling is prevented, significant energy dissipation can occur over a cycle of motion (Symans et al. 2008).

The ADAS damper (Whittaker et al. 1991) consists of a series of steel plates wherein the

bottom of the plates are attached to the top of a chevron bracing arrangement and the top of the plates are attached to the floor level above the bracing. As the floor level above deforms laterally with respect to the chevron bracing, the steel plates are subjected to a shear force. The shear force induces bending moments over the height of the plates, with bending occurring about the weak axis of the plate cross section. The geometrical configuration of the plates is such that bending moments produce a uniform flexural stress distribution over the height of the plates so that inelastic action occurs uniformly over the full height of the plates. For example, in the case that the plates are fixed-pinned, the geometry is triangular; however, if the plates are fixed-fixed, the geometry is an hourglass shape.

Other types of metallic dampers have been developed for beam-column connections, braces and base isolation systems. Early developments include the U-strip hysteretic dampers and the T-ADAS damper. Other examples include the honeycomb damper used as seismic isolation system in bridges, C-shaped and E-shaped hysteretic dampers for bridges, slit-type dampers applied to beam-column connections or brace members, yielding shear panels, cast-iron yielding fuses installed in braces and hourglass shape pins installed in beam-column connections. A complete reference list for the aforementioned metallic dampers is provided in Vasdravellis et al. (2014).

Various mathematical models that describe yielding behavior of metals can represent the hysteretic behavior of metallic dampers such as the standard Bouc-Wen model (Wen 1976). The standard Bouc-Wen model results from the parallel combination of an elastic component and an

elastic-perfectly plastic component. The force output F of the model is

$$F = pku + (1 - p)F_y z \quad (5)$$

where u is the deformation across the model, F_y the yield force, k the elastic stiffness, p the post-yield stiffness ratio, and z a dimensionless hysteretic parameter governed by

$$\dot{z} = \frac{k}{F_y} \dot{u} [1 - |z|^n (\beta \text{sgn}[\dot{u}z] + \gamma)] \quad (6)$$

where β and γ are parameters controlling the shape of the hysteresis, n is a parameter that controls the sharpness of the smooth transition from the elastic to the inelastic region of the hysteresis, $\text{sgn}[\cdot]$ is the signum function, and the overdot denotes derivative with respect to time.

Equation 5 shows that the Bouc-Wen model accounts for kinematic hardening (i.e., post-yield force increases with increasing deformation) due to the post-yield stiffness ratio p . However, the model does not account for the isotropic hardening (i.e., yield force F_y increases due to cyclic inelastic deformation) in the hysteresis of steel energy dissipation devices.

To account for isotropic hardening in the Bouc-Wen model, the mathematical formulation developed by Karavasilis et al. (2012b) is described below:

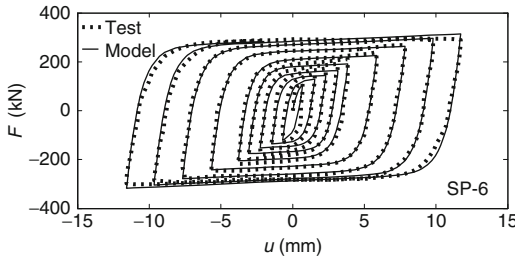
The yield force F_y needs to be updated by considering the history of the imposed cyclic deformation u . Examination of the constitutive Eqs. 5 and 6 reveals that a change in the yield force F_y can be achieved by appropriately including a third shape control parameter Φ in Eq. 6:

$$\dot{z} = \frac{k}{F_y} \dot{u} \left[1 - |z|^n \left(\beta \text{sgn}(\dot{u}z) + \gamma - \frac{\Phi \text{sgn}(\dot{u})(\text{sgn}(z) + \text{sgn}(\dot{u}))}{\text{sgn}(\dot{u})} \right) \right] \quad (7)$$

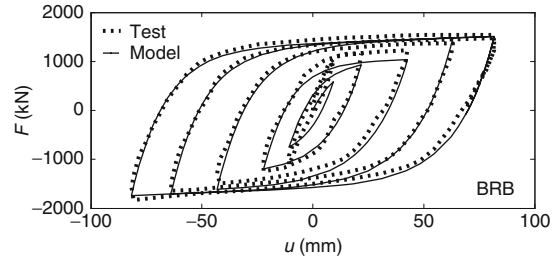
The parameter Φ quantifies isotropic hardening and is calculated using functions that cause Φ to increase exponentially with increasing cumulative plastic deformation $u_{pl,c}$, i.e.,

$$\Phi_p = \Phi_{\max,p} \left[1 - \exp \left(-p_{\Phi,p} \left| \frac{u_{pl,c}}{u_y} \right| \right) \right] \quad (8a)$$

or



Buildings and Bridges Equipped with Passive Dampers Under Seismic Actions: Modeling and Analysis, Fig. 7 Results from the modified Bouc-Wen



model proposed in Karavasili et al. (2012b) and experimental results for a low-yield-strength shear panel and a BRB

$$\Phi_n = \Phi_{\max, n} \left[1 - \exp \left(-p_{\Phi, n} \left| \frac{u_{pl, c}}{u_y} \right| \right) \right] \quad (8b)$$

where $u_y (= F_y/k)$ is the yield deformation, $p_{\Phi, p}$ and $p_{\Phi, n}$ are parameters that control the isotropic hardening rate due to cumulative plastic deformation, and $\Phi_{\max, p}$ and $\Phi_{\max, n}$ are the maximum possible values of Φ for the fully saturated isotropic hardening condition, i.e., for $u_{pl, c} \rightarrow \infty$, $\Phi_p \rightarrow \Phi_{\max, p}$ and $\Phi_n \rightarrow \Phi_{\max, n}$. On the other hand, when $u_{pl, c} = 0.0$, $\Phi_p = 0.0$ and $\Phi_n = 0.0$.

Φ_p and Φ_n are used to independently capture isotropic hardening in different loading directions (positive and negative). Typically, yielding devices exhibit the same isotropic hardening in different loading directions, and therefore, $\Phi_{\max, p} = \Phi_{\max, n}$ and $p_{\Phi, p} = p_{\Phi, n}$. However, the model can simulate different isotropic hardening in different loading directions (e.g., compressive and tensile loading in BRB hysteresis) by using different parameter values in Eqs. 8a and 8b.

To understand the effect of Φ , consider the proposed Bouc-Wen model with $p = 0.0$ (i.e., without kinematic hardening), $n = 1$ and $\beta + \gamma = 1$, under a positive deformation increment, i.e., $\text{sgn}[\dot{u}] = 1$. Assume that in the previous deformation increment, z has reached its positive ultimate value z_u , and therefore, $\text{sgn}[z] = 1$ and $\dot{z} = 0.0$. For this case, Eq. 7 yields $z_u = 1/(1 - 2 \cdot \Phi)$ and Eq. 5 yields $F = F_y \cdot z_u = F_y/(1 - 2 \cdot \Phi)$. When $\Phi = 0.0$ (i.e., without isotropic hardening), F is equal to $F_y/(1 - 0.0) = F_y$. When $\Phi \neq 0.0$ (e.g., $\Phi = 0.1$), F is equal to $F_y/(1 - 2 \cdot 0.1) = F_y/(1 - 2 \cdot 0.1) = 1.25 \cdot F_y$. In that case, Φ

reflects a 25 % increase in the initial yield strength F_y due to isotropic hardening. The term $\text{sgn}[\dot{u}]$ after the parameter Φ in Eq. 7 ensures that the above calculations apply to the case of a negative deformation increment and a negative ultimate value of z .

The state determination procedure requires as an input the previous force and deformation, the previous z value, and the current deformation. The current value of the parameter Φ is then calculated based on the following rules: Eq. 8a is used to update Φ_p when the deformation increment changes from negative to positive within the plastic region of the hysteresis; Eq. 8b is used to update Φ_n when the deformation increment changes from positive to negative within the plastic region of the hysteresis; Φ equals to Φ_p when a positive deformation increment occurs; and Φ equals to Φ_n when a negative deformation increment occurs. With the current Φ value known, Eq. 7 is numerically integrated to obtain the current value of z which is simply used in Eq. 5 to provide the current force output F of the model.

Figure 7 shows how the aforementioned modified Bouc-Wen model captures the behavior of low yield strength shear panels showing significant isotropic hardening in their hysteresis as well as the behavior of BRBs showing different isotropic hardening in tension and compression.

Friction Dampers

Friction dampers dissipate seismic energy by friction that develops between two solid bodies sliding relative to one another. Typical friction

dampers are slotted-bolted (Grigorian et al. 1993), where a series of steel plates are bolted together with a specified clamping force. The clamping force is such that slip occurs at a pre-specified friction force. Another configuration is the Pall friction damper, which consists of cross bracing that connects in the center to a rectangular damper (Pall and Marsh 1982). The damper is bolted to the cross bracing and, under lateral load, the structural frame distorts such that two of the braces are subjected to tension and the other two to compression. This force system causes the rectangular damper to deform into a parallelogram, dissipating energy at the bolted joints via sliding friction (Symans et al. 2008).

Experimental testing (Pall and Marsh 1982) has shown that the idealized Coulomb friction model can model the behavior of friction dampers, i.e., the force output of friction dampers is given by

$$P(t) = \mu N \text{sgn}[\dot{u}(t)] \quad (9)$$

where μ is the coefficient of dynamic friction and N is the normal force at the sliding interface.

Analysis Methods for Buildings and Bridges with Passive Dampers

The analysis methods used to predict the response of buildings and bridges equipped with passive dampers are classified in two categories: (1) linear analysis methods and (2) nonlinear analysis methods.

Linear Analysis Methods

All linear analysis methods for structures equipped with rate-dependent passive dampers involve a simplified calculation of the supplemental (equivalent) damping ratio, ξ_{eq} . For example, given the damping coefficients C_i at each story i of a building and by assuming linear fluid viscous dampers ($a = 1$ in Eq. 1) positioned in a horizontal configuration, ξ_{eq} at the fundamental period of vibration T_1 under elastic conditions can be estimated as (Whittaker et al. 2003):

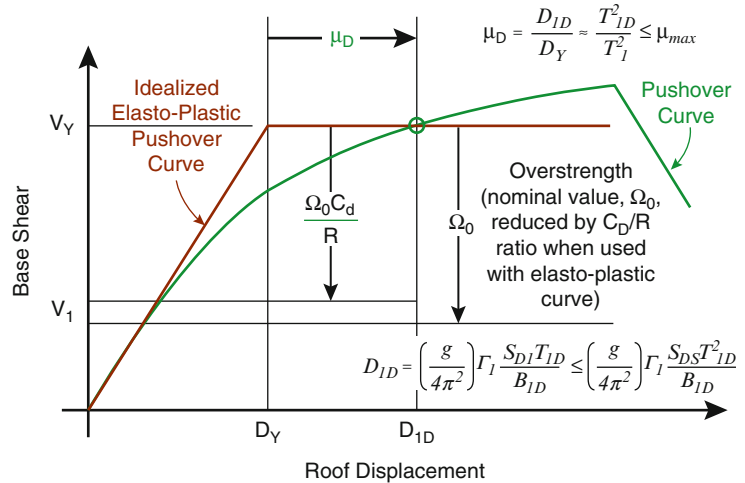
$$\xi_{eq} = \frac{T_1}{4\pi} \times \frac{\sum_i C_i (\phi_i - \phi_{i-1})^2}{\sum_i m_i \times \phi_i^2} \quad (10)$$

where ϕ_i and ϕ_{i-1} are the first modal displacements of stories i and $i - 1$, respectively, and m_i is the mass of story i . Equation 10 assumes that the braces supporting the dampers are stiff enough (essentially rigid) so that story deformation results in damper deformation rather than brace deformation. Alternative equations that take into account the effect of the brace flexibility on ξ_{eq} are provided for various brace-damper configurations in Hwang et al. (2008).

ξ_{eq} is used to calculate the damping reduction factor, which is then utilized to appropriately reduce the ordinates of the elastic response spectrum. This reduced highly damped spectrum is used to estimate the response of the structure with dampers. Lin and Chopra (2003) have evaluated the premise that the peak displacement of an elastic SDOF system with a natural period T_n and equipped with a viscous damper (damping coefficient C) in series with a brace of stiffness K_b , can be estimated by a linear viscous SDOF system with the same period T_n along with a simplified calculation of the added damping ratio (see Eq. 10). This premise was found valid for SDOF systems with $\tau/T_n < 0.02$, where τ is the relaxation time, i.e., $\tau = C/K_b$. As discussed by Lin and Chopra (2003), the relation $\tau/T_n < 0.02$ is satisfied for the practical range of values for the bracing axial stiffness K_b and the damping coefficient C . Therefore, the damper and brace properties necessary to ensure that elastic drift demands do not exceed a prescribed design limit can be determined using the reduced elastic design spectrum associated with the equivalent damping ratio.

In the 2003 *NEHRP Recommended Provisions* (BSSC 2004) for buildings with passive dampers, and for the equivalent lateral force (ELF) analysis (linear static analysis) specifically, the response is defined by two modes: the fundamental mode and the residual mode, which is used to approximate the combined effects of higher modes. For response spectrum analysis, higher modes are

Buildings and Bridges Equipped with Passive Dampers Under Seismic Actions: Modeling and Analysis, Fig. 8 Idealized elastoplastic pushover curve used for linear analysis (Symans et al. 2008)



explicitly evaluated. For both the ELF and the response spectrum analysis procedures, the shape of the fundamental-mode pushover capacity curve is not known and an idealized elastoplastic pushover curve is assumed, as shown in Fig. 8. Note that, in Fig. 8, the parameters Γ_1 and S_{DS} , which are used to compute the design earthquake displacement, D_{ID} , represent the modal participation factor for the fundamental mode and the 5 % damped design spectral response acceleration at short periods, respectively. The idealized pushover curve permits defining the effective global ductility demand due to the design earthquake, μ_D , as the ratio of design roof displacement, D_{ID} , to the yield displacement, D_Y . This ductility factor is used to calculate various design factors and to limit the maximum ductility demand, μ_{max} , in a manner that is consistent with conventional building response limits. Passive dampers should be designed for actual fundamental-mode design earthquake forces corresponding to a base shear value of V_Y , while the elements of the seismic-force-resisting system be designed for a reduced fundamental-mode base shear, V_1 , where the force reduction is based on system overstrength, Ω_0 (Symans et al. 2008).

A simplified linear method of analysis for building frames with elastomeric or viscoelastic dampers has been proposed by Lee et al. (2009) and has been recently extended to the case of fluid viscous dampers by Seo et al. (2014). This method is fully compatible with linear methods

of analysis used for conventional lateral load resisting systems in seismic codes and adopts the use of the strength reduction factor to calculate the required design base shear and the displacement amplification factor to estimate drifts.

Nonlinear Analysis Methods

The 2003 *NEHRP Recommended Provisions* (BSSC 2004) specify procedures for nonlinear static analysis and nonlinear dynamic (response-history) analysis. The nonlinear static analysis procedure is similar to the ELF procedure, in that the pushover capacity curve is used to define the nonlinear behavior of the structure. However, in the nonlinear static analysis procedure, the actual nonlinear force-displacement relation is used, rather than an idealized elastoplastic curve. In addition, since actual pushover strength is known from the nonlinear pushover analysis, the force reduction for design of the seismic-force-resisting system is based on overstrength alone with no additional reduction (Symans et al. 2008).

In general, the nonlinear dynamic analysis procedure is the most robust procedure available for evaluating the behavior of structures with passive dampers. It allows explicit modeling of individual devices, the elements connecting the devices to the structure, and the structure itself. If the connecting elements or the structural framing yields during the response, this behavior must be incorporated into the analytical model.

It is noted that accurate modeling of the flexibility of the floor diaphragm and of the connecting elements (braces) is essential since a loss of effective damping may occur if these elements are overly flexible.

Nonlinear dynamic analysis may be performed using a variety of commercially available software. In addition, there are several *academic* programs available, such as DRAIN-2DX (Prakash et al. 1993) and OpenSees (Mazzoni et al. 2006). Most of these programs can readily be used to model the behavior of linear fluid viscous dampers, viscoelastic dampers, friction dampers, or metallic yielding dampers. However, modeling of some damping devices (e.g., nonlinear viscous dampers and dampers with temperature-dependent or frequency-dependent mechanical properties) can be more challenging or, in some cases, not possible with a given program.

When nonlinear dynamic analysis is used, it is often beneficial to investigate the sensitivity of the structure response to one or more systemic parameters. Examples of parameters to vary include ground motion scaling parameters and parameters of the passive dampers (Symans et al. 2008).

Summary

The scope of this contribution was the presentation of numerical models and analysis methods used for structures equipped with passive dampers. The principal function of passive dampers and their application in buildings and bridges were discussed. The most common types of passive dampers and the constitutive models used to simulate their hysteretic behavior were presented. Finally, the analysis methods used for seismic design and assessment of structures with passive dampers were covered.

Cross-References

- [Assessment of Existing Structures Using Response History Analysis](#)

- [Behavior Factor and Ductility](#)
- [Buckling-Restrained Braces and Their Implementation in Structural Design of Steel Buildings](#)
- [Earthquake Response Spectra and Design Spectra](#)
- [Earthquake Return Period and Its Incorporation into Seismic Actions](#)
- [Equivalent Static Analysis of Structures Subjected to Seismic Actions](#)
- [European Structural Design Codes: Seismic Actions](#)
- [Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis](#)
- [Modal Analysis](#)
- [Plastic Hinge and Plastic Zone Seismic Analysis of Frames](#)
- [Response Spectrum Analysis of Structures Subjected to Seismic Actions](#)
- [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)
- [Seismic Analysis of Steel and Composite Bridges: Numerical Modeling](#)
- [Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling](#)
- [Soil-Structure Interaction](#)
- [Steel Structures](#)
- [Strengthening Techniques: Code-Deficient Steel Buildings](#)
- [Structural Design Codes of Australia and New Zealand: Seismic Actions](#)
- [Time History Seismic Analysis](#)

References

- American Society of Civil Engineers [ASCE] (2005) Minimum design loads for buildings and other structures. ASCE/SEI, Reston, pp 7–05
- American Society of Civil Engineers [ASCE] (2010) Minimum design loads for buildings and other structures. ASCE/SEI 7–10. ASCE, Reston
- Applied Technology Council [ATC] (1997a) NEHRP commentary on the guidelines for the seismic rehabilitation of buildings. 1997 ed, report no FEMA-274, prepared for the Building Seismic Safety Council (BSSC) by the Applied Technology Council (ATC). Federal Emergency Management Agency (FEMA), Washington, DC

- Applied Technology Council [ATC] (1997b) NEHRP guidelines for the seismic rehabilitation of buildings. 1997 ed, report no FEMA-273, prepared for the Building Seismic Safety Council (BSSC) by the Applied Technology Council (ATC). Federal Emergency Management Agency (FEMA), Washington, DC
- Bergman DM, Hanson RD (1993) Viscoelastic mechanical damping devices tested at real earthquake displacements. *Earthq Spectra* 9(3):389–418
- Building Seismic Safety Council [BSSC] (2004) NEHRP recommended provisions for seismic regulations for new buildings and other structures. 2003 ed, report no FEMA-450/1 and FEMA-450/2, prepared by the Building Seismic Safety Council (BSSC) for the Federal Emergency Management Agency (FEMA), Washington, DC
- Christopoulos C, Filiatrault A (2006) Principles of supplemental damping and seismic isolation. IUSS Press, Milan
- Fan C-P (1998) Seismic analysis, behavior, and retrofit of non-ductile reinforced concrete frame buildings with viscoelastic dampers. Phd Lehigh University, Bethlehem
- Grigorian CE, Yang TS, Popov EP (1993) Slotted bolted connection energy dissipators. *Earthq Spectra* 9(3):491–504
- Hwang J-S, Huang Y-N, Yi S-L, Ho S-Y (2008) Design formulations for supplemental viscous dampers to building structures. *J Struct Eng ASCE* 134(1):22–31
- Infanti S, Papanikolas P, Benzoni G, Castellano MG (2004) 13th world conference on earthquake engineering, paper no 2174, Vancouver
- Karavasilis TL, Seo C-Y (2011) Seismic structural and non-structural performance evaluation of highly damped self-centering and conventional systems. *Eng Struct* 33:2248–2258
- Karavasilis TL, Ricles JM, Sause R, Chen C (2011a) Experimental evaluation of the seismic performance of steel MRFs with compressed elastomer dampers using large-scale real-time hybrid simulation. *Eng Struct* 33(6):1859–1869
- Karavasilis TL, Blakeborough T, Williams MS (2011b) Development of nonlinear analytical model and seismic analyses of a steel frame with self-centering devices and viscoelastic dampers. *Comput Struct* 89(11–12):1232–1240
- Karavasilis TL, Sause R, Ricles JM (2012a) Seismic design and evaluation of steel MRFs with compressed elastomer dampers. *Earthq Eng Struct Dyn* 41(3):411–429
- Karavasilis TL, Krawala S, Hale E (2012b) Model for hysteretic behaviour of steel energy dissipation devices and evaluation of a minimal-damage seismic design approach for steel frames. *J Constr Steel Res* 70:358–367
- Lee KS, Ricles J, Sause R (2009) Performance based seismic design of steel MRFs with elastomeric dampers. *J Struct Eng ASCE* 135(5):489–498
- Lin W-H, Chopra AK (2003) Earthquake response of elastic single-degree-of-freedom systems with nonlinear viscoelastic dampers. *J Eng Mech ASCE* 129(6):597–606
- Mazzoni S, McKenna F, Scott MH, Fenves GL (2006) Open system for earthquake engineering simulation: user command language manual. University of California at Berkeley, Pacific Earthquake Engineering Research Center, Berkeley
- Pall AS, Marsh C (1982) Seismic response of friction damped braced frames. *J Struct Div* 108(6):1313–1323
- Prakash V, Powell GH, Campbell S (1993) DRAIN-2DX base program description and users guide. Department of Civil Engineering, University of California at Berkeley, Berkeley
- Rahimian A, Romero EM (2003) Super structure. *Civ Eng NY* 73(6):21–69
- Seleemah A, Constantinou MC (1997) Investigation of seismic response of buildings with linear and nonlinear fluid viscous dampers. Report no NCEER 97-0004. State University of New York at Buffalo, National Center for Earthquake Engineering Research, Buffalo
- Seo C-Y, Karavasilis TL, Ricles JM, Sause R (2014) Seismic performance and probabilistic collapse resistance assessment of steel moment resisting frames with fluid viscous dampers. *Earthq Eng Struct Dyn* 43(14):2135–2154
- Soong TT, Spencer BF Jr (2002) Supplemental energy dissipation: state-of-the-art and state-of-the-practice. *Eng Struct* 24(3):243–259
- Symans MD, Charney FA, Whittaker AS, Constantinou MC, Kircher CA, Johnson MW, McNamara RJ (2008) Energy dissipation systems for seismic applications: current practice and recent developments. *J Struct Eng ASCE* 134(1):3–21
- Vasdravellis G, Karavasilis T, Uy B (2014) Design rules, experimental evaluation, and fracture models for high-strength and stainless-steel hourglass shape energy dissipation devices. *J Struct Eng* 140(11):04014087. doi: 10.1061/(ASCE)ST.1943-541X.0001014
- Wen YK (1976) Method for random vibration of hysteretic systems. *J Eng Mech Div* 102(2):249–263
- Whittaker AS, Bertero VV, Thompson CL, Alonso LJ (1991) Seismic testing of steel-plate energy dissipating devices. *Earthq Spectra* 7(4):563–604
- Whittaker AS, Aiken ID, Bergman D, Clark PW, Cohen J, Kelly JM, Scholl RE (1993) Code requirements for design and implementation of passive energy dissipation systems. In: *Proceedings of ATC-17-1 seminar on seismic isolation, passive energy dissipation, and active control*, vol 2. ATC, Redwood City, pp 497–508
- Whittaker AS, Constantinou MC, Oscar MR, Johnson MW, Chrysostomou CZ (2003) Equivalent lateral force and modal analysis procedures of the 2000 NEHRP provisions for buildings with damping systems. *Earthq Spectra* 19(4):959–980

Cast Fuse Braces: Design and Implementation

Michael Gray

Cast Connex Corporation, Toronto, ON, Canada

Synonyms

Cast steel yielding brace system

Introduction

The yielding brace system (YBS) is a seismic-resistant bracing system for concentrically braced frame building structures designed at the University of Toronto (Gray et al. 2013). The YBS consists of a special yielding connector, or fuse, at the end of a concentric steel brace. When large lateral earthquake forces are imposed on a building designed with the YBS, the connector's specially designed triangular fingers yield in flexure. The yielding of the fingers dissipates seismic energy and limits the forces that the remaining structural elements must resist. This is analogous to the fuse in an electric circuit, which is why the yielding connector is sometimes called a "fuse."

Yielding connectors are comprised of two castings which are connected to the end of a brace member. Each casting includes an elastic arm portion with a series of triangular yielding fingers. The ends of the yielding fingers are

bolted to a fabricated splice plate assembly which is connected to a standard beam-column gusset plate connection. A YBS prototype is illustrated in a full-scale frame test configuration in Fig. 1.

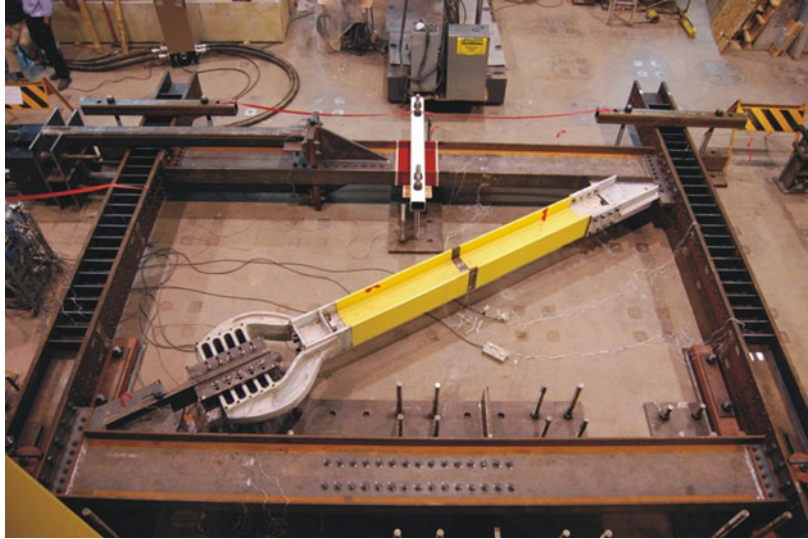
Much like the TADAS system (Tsai et al. 1993), the fingers are triangular to promote yielding along their entire length. The bending moment diagram of each finger is triangular, like a cantilever. The triangular shape of the finger ensures that moment resistance diagram matches the applied moment diagram and the extreme fiber of the finger yields simultaneously at all points along the length of the finger.

The yielding fingers are connected to a fabricated plate assembly via slotted holes. The slotted holes enable the end of the finger to translate in the direction that is normal to the brace axis as the finger is deformed. A standard hole would restrain that movement and cause large axial forces, and strains, to develop in the fingers, which would decrease their displacement capacity.

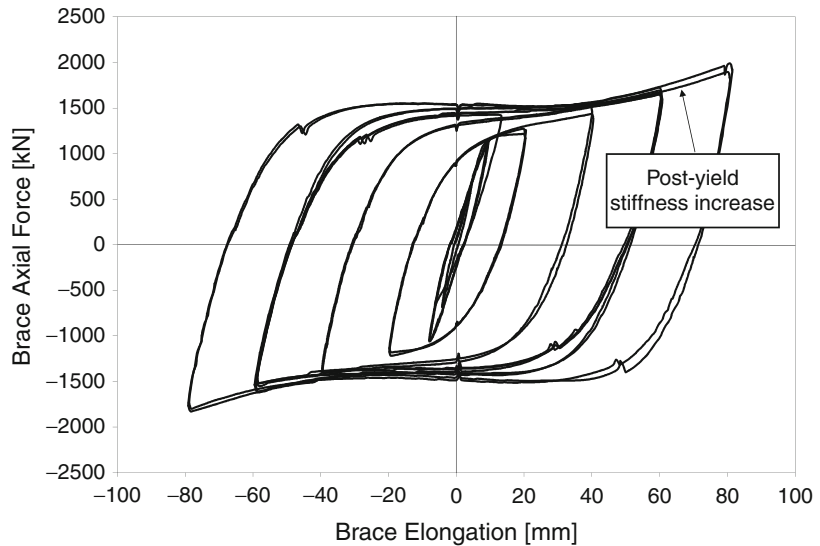
Nonlinear Response

The YBS yielding connector exhibits a full, ductile hysteretic response. A sample hysteresis of the YBS is presented in Fig. 2. The stiffness and displacement capacity of the braces and connectors have been demonstrated to be very similar to other high-ductility braces, like the buckling

Cast Fuse Braces: Design and Implementation,
Fig. 1 YBS prototype in a full-scale frame test



Cast Fuse Braces: Design and Implementation,
Fig. 2 Sample hysteresis of a YBS brace



restrained brace. However, unlike the buckling restrained brace (► [Buckling-Restrained Braces and Their Implementation in Structural Design of Steel Buildings](#)) which has a higher strength in compression than in tension, the YBS has a symmetric response.

Another difference between the two systems is the post-yield response of the YBS. Although the slotted holes, to which the fingers are bolted, reduced the axial forces that develop in the fingers at large displacements, some axial force develops due to the finger's second-order

deformed shape. The small axial force appears in the connector's global response as distinct increase in strength and stiffness at large displacements. The benefit of such a response, when considered in the structural design, is the ability to limit the amount of inelastic demand that can concentrate on a single story of a multistory building. As the lateral deformation of a single story increases, the brace eventually increases in strength, thereby forcing adjacent stories to become engaged and inelastic demand to spread up the height of the building.

Design

To best take advantage of the economies available in the casting process, YBS connectors are available in a series of off-the-shelf configurations for a variety of brace yield loads. The intention is that each available connector size will have an associated maximum brace elongation and brace rotation for which it has successfully completed the testing requirements of Section K3 of AISC 341-10 (AISC 2010).

A designer wishing to utilize the YBS in a building design first selects the appropriate connector for each brace based on strength requirements. As fuse connectors are generally available in force increments of approximately 10 %, any overstrength due to the incremental nature of the YBS is small.

Like all traditional earthquake-resistant steel structures, elements of the lateral force-resisting system which are intended to remain elastic are capacity designed for the expected strength of the yielding element. The adjusted brace strength is the product of three parameters: the nominal connector yield strength, P_y ; material overstrength, R_y ; and the overstrength resulting from cyclic strain hardening and the geometric post-yield stiffening effect, ω .

The nominal connector yield strength is directly computed using the sum of plastic flexural strengths of the connector's yielding fingers, the length of each finger, and minimum allowable yield strength of the cast steel material.

Past experience with manufacture of steel castings (Gray 2012) has indicated that cast steel yield strength can be well controlled through a well-defined heat treatment process. An assumed material overstrength of 1.1 is easily achieved through standard foundry practices and can be verified by tensile coupon testing.

The effect of cyclic strain hardening and post-yield stiffening, ω , can be measured directly through full-scale testing. As the fuse's adjusted strength is determined for an elongation of $2\Delta_{bm}$ (the elongation associated with greater of twice the design level drift or 2 %), the overstrength factor, ω , is the ratio of the force measured during cyclic testing at this deformation to the yield

force determined using the actual material yield strength of the test specimen.

Once the connector sizes are selected and the adjusted brace strength is calculated, the brace members must be designed. Generally, YBS fuse connectors are used with W-section or rectangular/square hollow structural section braces. Brace design is governed by the factored compressive resistance of the brace member, which must exceed the adjusted strength of the yielding connector. The in-plane and out-of-plane effective buckling lengths are different for yielding connector equipped braces. Out of plane, the connector provides continuous stability such that the effective length can be taken as the work point-to-work point length of the brace member. In plane, the yielding finger-to-slotted hole connection provides very little rotational or lateral stiffness, and the brace can be idealized in plane as a flagpole. Thus it is recommended that the in-plane effective buckling length of the brace be determined from the distance from the center of the yielding finger-to-plate assembly connection to the opposite brace end connection with an effective length factor of 2.0. For stability, this requires that the opposite end brace connection provide in-plane flexural stiffness.

After the initial brace design is complete, the lateral stiffness of the building can be evaluated. In order to check drift limits, to conduct a linear response spectrum analysis, or to run a nonlinear dynamic analysis, the total brace stiffness must be calculated. The flexibility of a brace equipped with a yielding connector is determined by summing the flexibilities of the connector and the brace itself. Since the design force for the brace (the adjusted strength of the connector) is determined based on the design story drift, which is a function of the brace stiffness, brace design may be an iterative process. If the initial brace flexibility is higher than expected, the design story drift and adjusted brace strength will also increase, potentially requiring a stronger brace.

With the use of yielding connectors, a structure's lateral stiffness can be increased through the use of heavier brace members, without increasing the brace yield force. Thus, there is not an increase in capacity design forces for

beams, columns, diaphragms, connections, or foundations that is often the case when increasing the stiffness of other traditional lateral force-resisting systems. This may be of particular relevance for structures with strict drift control limits.

Like a buckling restrained braces, bracing connections are to be designed for a force of 1.1 times the adjusted brace strength. Connections can be field bolted or field welded, provided that the connection on the opposite side of the brace from the yielding connector provides in-plane flexural stiffness.

As with other high-ductility lateral force-resisting systems, the beams, columns, diaphragms, and foundations should all be designed with capacity design principles for the adjusted brace strength of the final braced frame design.

Replace-Ability

Replace-ability is a newer concept in earthquake engineering that involves detailing the yielding element of a seismic force-resisting system to be easily removed and replaced after a significant earthquake event. If a seismic lateral force-resisting system is properly designed and detailed, only a small part of the structure is damaged during an earthquake. If those elements are designed to be easily replaced, the post-earthquake repair cost for the building could be significantly less than with traditional seismic force-resisting systems. Replaceable elements have been designed and tested for eccentrically braced frames (Mansour et al. 2011) and moment-resisting frames (Shen et al. 2011). Replaceable eccentrically braced frames have been implemented in steel buildings in Christchurch, since the devastating earthquake of 2011.

The yielding brace system can be designed with bolted joints at both the connector-to-brace and connector-to-gusset plate connections. This enables relatively simple removal and replacement after a major earthquake. Of course the removal of the connectors is only necessary if they have undergone significant cyclic inelastic displacements. The state of a connector can be

easily evaluated post-earthquake, as all of the yielding elements are visible. After cycling at large displacements, but long before low-cycle fatigue failure, cracking becomes visible on the surfaces of the yielding fingers. This would enable the inspection and evaluation of a yielding connector after a major earthquake.

Retrofit

Although originally intended for using in new building structure, yielding connectors can be used as hysteretic dampers for the retrofit of seismically deficient existing structures (Retrofitting and Strengthening: An Overview). The full, symmetric hysteretic response of the yielding connectors provides the added damping that is desirable in many retrofit schemes. In addition, yielding connector equipped braces have a relatively high stiffness, which is beneficial when drift control is an issue, as it often is when retrofitting brittle, non-seismically designed structure.

The ability to modify a brace assembly's stiffness independently from its yield force (discussed in the Design section) is an attribute that is perhaps most valuable in a retrofit design. Excessive drift is a primary source of damage to existing structural elements in non-seismically detailed structures. As such, retrofit schemes often must limit drift during an earthquake to limit resulting damage. This is usually achieved through added stiffness; however, adding stiffness also requires simultaneously adding strength. Increasing the strength of the lateral force-resisting system increases the capacity design force that the existing structural elements must resist, often requiring expensive reinforcement of those elements.

The advantage offered by the YBS is that small connectors, which are stiff yet have large displacement capacities, can be used to provide hysteretic damping but limit the capacity design force the existing elements are required to resist. The brace member can then be selected to optimize the overall brace stiffness to control drifts and limit damage.

The post-yield stiffness of the YBS is valuable when used in deficient buildings with no other source of lateral stiffness. Seismic force-resisting systems with no post-yield stiffness, such as braces with friction dampers, are not suitable without a secondary source of lateral stiffness. On their own, such systems have a negative post-yield stiffness when second-order ($P-\Delta$) effects are considered. Unlike friction systems, the YBS can be used in cases with no other lateral stiffness, such as simple steel frames with unreinforced masonry walls, because of its characteristic post-yield stiffness.

Also, since the brace members are not part of the yielding element, they can have field splices, which enable easier installation in the existing building.

Summary

The yielding brace system (YBS) is a new high-ductility concentric bracing system that dissipates energy through the yielding of fingers within a special cast steel yielding connector. The connector has a full, symmetric hysteretic response that is characterized by a post-yield increase in strength and stiffness at large displacements. The design of systems using the yielding connectors is similar to the design of a buckling restrained brace: the connectors are selected for strength requirements and the remaining structural elements are capacity designed for the adjusted strength of the connector. Yielding connectors can be designed with bolted connections enabling simple post-earthquake replacement. They are suitable for both new buildings and for the retrofit of seismically deficient existing buildings.

Cross-References

- [Buckling-Restrained Braces and Their Implementation in Structural Design of Steel Buildings](#)

References

- American Institute of Steel Construction (AISC) (2010) ANSI/AISC 341-10 seismic provisions for structural steel buildings. AISC, Chicago
- Gray MG (2012) Cast steel yielding brace system for concentrically braced frames. University of Toronto PhD dissertation. University of Toronto, Toronto
- Gray MG, Christopoulos C, Packer JA (2013) Cast steel yielding brace system (YBS) for concentrically braced frames: concept development and experimental validations. *ASCE J Struct Eng* (in press). ASCE, Reston, Vol. 140, pp 04013095-1–04013095-11
- Mansour N, Christopoulos C, Tremblay R (2011) Experimental validation of replaceable shear links for eccentrically braced steel frames. *ASCE J Struct Eng* 137(10):1141–1152
- Shen Y, Christopoulos C, Mansour N, Tremblay R (2011) Seismic design and performance of steel moment-resisting frames with nonlinear replaceable links. *ASCE J Struct Eng* 137(10):1107–1117
- Tsai KC, Chen HW, Hong CP, Su YF (1993) Design of steel triangular plate energy absorbers for seismic-resistant construction. *Earthq Spectra* 9(3):505–528

Classically and Nonclassically Damped Multi-degree of Freedom (MDOF) Structural Systems, Dynamic Response Characterization of

Marcello Vasta

INGEO, Engineering and Geology Department, University of Chieti-Pescara “G. D’Annunzio”, Pescara, Italy

Synonyms

MDOF systems; Nonproportional damping; Reliability analysis; Spectral parameters; Viscous damping

Introduction

Characterization of damping forces in a vibrating structure has long been an active area of research in structural dynamics. The most common approach to model viscous forces is to use the

so-called viscous damping, first introduced by Lord Rayleigh (1877), which relies on the assumption that the damping matrix is a linear combination of the mass and stiffness matrices. Since its introduction, this model has been used extensively and is now usually known as *Rayleigh damping* or *classical damping* (Argyris and Mlejnek 1991; Clough and Penzien 1993). With such a damping model, the modal analysis procedure, originally developed for undamped systems, can be used to analyze damped systems in a very similar manner.

In reality, physical structures or systems are generally comprised of many substructures tied together in various fashions. These substructures can be made up of a variety of materials, i.e., metals, plastics, and woods. Furthermore, these substructures may be connected to one another by rivets, bolts, dampers, springs, friction, etc. Also, the spatial geometry of the structure may be very complicated. All of these factors influence the inherent dynamical properties of the structure. For these structures, mass, damping, and stiffness distribution (matrices) of the system are rather complicated. In general, for real-life structures, the damping matrix for such a system will not always be proportional to the mass and stiffness matrix.

For nonproportional damping, the well-known modal coordinate transformation, which diagonalizes the system mass and stiffness matrices, will not diagonalize the system damping matrix. Therefore, when a system with nonproportional damping exists, the equations of motion are coupled when formulated in n dimension physical space. Fortunately, the equations of motion can be uncoupled when formulated in $2n$ dimension state space, by means of complex analysis.

When dealing with earthquake excitation, the dynamic behavior of structural and mechanical systems subjected to uncertain dynamic excitations can be described, in general, through random processes (Chopra 1995). The probabilistic characterization of these random processes can be extremely complex, when nonstationary and/or non-Gaussian input processes are involved (Lin 1976). In specific applications, an incomplete description of stochastic processes corresponding to dynamic structural response

may suffice, based on the spectral characteristics of the processes under study (Priestley 1987). In particular, spectral characteristics can be employed to evaluate the time-variant central frequency and bandwidth parameters, which characterize in a synthetic way a nonstationary stochastic process. The spectral characteristic has been proved appropriate for describing nonstationary stochastic processes and can be effectively employed in structural reliability applications, such as the computation of the time-variant probability that a random process outcrosses a given limit-state threshold, and is useful for reliability of earthquake-sensible structures (Barbato and Conte 2008).

Classically Damped MDOF Structures

Free Vibrations of MDOF Structures

The free vibration equations of motion of an n -degree-of-freedom elastic structure are

$$\mathbf{M}\ddot{\mathbf{U}}(t) + \mathbf{K}\mathbf{U}(t) = 0 \quad (1)$$

where $\mathbf{M}$ and $\mathbf{K}$ are the $n \times n$ mass and stiffness symmetric and positive definite matrices, while $\mathbf{U}(t)$ is an n vector whose component collects the time-dependent response components of the structure. Equation 1 is a coupled system of n differential equations to be integrated with initial conditions

$$\dot{\mathbf{U}}(0) = \dot{\mathbf{U}}_0, \quad \mathbf{U}(0) = \mathbf{U}_0. \quad (2)$$

If the mass and stiffness matrices are positive definite, the solution of Eq. 1 can be written as

$$\mathbf{U}(t) = \sum_{j=1}^n C_j \exp(i\omega_j t) \phi_j \quad (3)$$

where C_j are n complex coefficients to be evaluated by imposing the initial conditions (Eq. 2), while ϕ_j and ω_j are, respectively, the *normal modes* and the *natural frequencies* of the structure, solution of the associated eigenvalue problem:

$$[\mathbf{K} - \omega_j^2 \mathbf{M}] \phi_j = \mathbf{0} \quad (4)$$

It can be shown that the orthogonality of the normal modes ϕ_j respects the stiffness and mass matrices (see, e.g., Clough and Penzien 1993). By setting with $\phi = [\phi_1 \dots \phi_n]$ the $n \times n$ modal matrix, whose column collects the normal modes of the structure, the following identity holds:

$$\phi^T \mathbf{M} \phi = \mathbf{I}, \quad \phi^T \mathbf{K} \phi = \mathbf{\Omega}^2 \quad (5)$$

In Eq. 5, $\mathbf{\Omega}$ is a diagonal matrix collecting the n natural frequencies of the structure, $\mathbf{\Omega} = [\omega_1 \dots \omega_n]$, while $\mathbf{I}$ is the identity matrix of order n .

Let us define the modal coordinates $\mathbf{q}(t)$ as

$$\mathbf{U}(t) = \phi \mathbf{q}(t) \quad (6)$$

Substituting Eq. 6 in Eq. 1 and left multiplying by ϕ^T , we obtain n uncoupled differential equation of motion in modal coordinates:

$$\ddot{\mathbf{q}}(t) + \mathbf{\Omega}^2 \mathbf{q}(t) = \mathbf{0} \quad (7)$$

which are the n uncoupled equations of motion in modal coordinates:

$$\ddot{q}_j(t) + \omega_j^2 q_j(t) = 0 \quad (8)$$

with associated solution

$$q_j(t) = q_j(0) \cos(\omega_j t) + \frac{\dot{q}_j(0)}{\omega_j} \sin(\omega_j t) \quad (9)$$

Using the orthogonality conditions (Eq. 5), the motion of the structure (Eq. 3) can be explicitly rewritten as follows:

$$\mathbf{U}(t) = \sum_{j=1}^n \left[\phi_j^T \mathbf{M} \mathbf{U}_0 \cos(\omega_j t) + \frac{\phi_j^T \mathbf{K} \dot{\mathbf{U}}_0}{\omega_j} \sin(\omega_j t) \right] \phi_j \quad (10)$$

Equation 10 shows that the free vibration of an n -dof structure can be thought as a linear combination of normal modes ϕ_j , each of them vibrating with the associated circular frequency ω_j .

Furthermore, giving an initial condition in displacement proportional to the j -th modes with zero velocity, the structure oscillates maintaining the shape of the given modes with circular frequency equal to ω_j .

Free Vibrations of Classically Damped MDOF Structures

The free vibration equations of motion of an n degree of freedom elastic structure with viscous damping are

$$\mathbf{M} \ddot{\mathbf{U}}(t) + \mathbf{C} \dot{\mathbf{U}}(t) + \mathbf{K} \mathbf{U}(t) = \mathbf{0} \quad (11)$$

where $\mathbf{C}$ is the $n \times n$ symmetric and positive definite damping matrix. An MDOF linear structure is *classically damped* if the damping matrix is diagonalizable by the modal matrix ϕ :

$$\phi^T \mathbf{C} \phi = [2\xi_1 \omega_1 \dots 2\xi_n \omega_n] = \mathbf{\Sigma} \quad (12)$$

In Eq. 12, ξ_i is the damping ratio of the i -th mode and ω_i is the undamped natural frequency of the i -th mode. The diagonal matrix $\mathbf{\Sigma}$ is called *modal damping matrix*. It is worth noting that for classically damped structural systems, the damping matrix $\mathbf{C}$ is diagonalizable by ϕ , which in turn depends only on $\mathbf{M}$ and $\mathbf{K}$, and is independent of $\mathbf{C}$ itself, regardless of how heavily the system is damped. There are many ways to compute a classical damping matrix from mass and stiffness matrices. A *Rayleigh* damping matrix is proportional to the mass and stiffness matrices:

$$\mathbf{C} = \alpha \mathbf{M} + \beta \mathbf{K} \quad (13)$$

where α and β are related to damping ratios and frequencies by

$$\xi_j = \frac{\alpha}{2\omega_j} + \frac{\beta\omega_j}{2} \quad (14)$$

An extended *Rayleigh* damping matrix, called *Caughey* damping matrix, can be computed from

$$\mathbf{C} = \mathbf{M} \left[\sum_{j=0}^{n-1} a_j (\mathbf{M}^{-1} \mathbf{K})^j \right] \quad (15)$$

where the a_j coefficients are related to damping ratios and frequencies by

$$\xi_k = \frac{1}{2} \sum_{j=0}^{n-1} a_j \omega_k^{2j-1}. \quad (16)$$

As for the free vibrations, the motion of an n -dof classically damped structure can be thought as a linear combination of normal modes ϕ_j , each of them vibrating with the associated circular frequency ω_j and damping ξ_j . Indeed, using the modal coordinates $\mathbf{q}(t)$ as defined in Eq. 6 and Eq. 11 and left multiplying by ϕ^T , we obtain n uncoupled differential equation of motion in modal coordinates:

$$\ddot{\mathbf{q}}(t) + \mathbf{\Sigma} \dot{\mathbf{q}}(t) + \mathbf{\Omega}^2 \mathbf{q}(t) = \mathbf{0} \quad (17)$$

which are the n uncoupled equations of motion in modal coordinates:

$$\ddot{q}_j(t) + 2\xi_j \omega_j \dot{q}_j(t) + \omega_j^2 q_j(t) = 0 \quad (18)$$

with associated solution

$$q_j(t) = e^{-\xi_j \omega_j t} \left(q_j(0) \cos(\bar{\omega}_j t) + \frac{\dot{q}_j(0) + \xi_j \omega_j q_j(0)}{\bar{\omega}_j} \sin(\bar{\omega}_j t) \right) \quad (19)$$

where $\bar{\omega}_j$ is the reduced or damped frequency of the j -th damped natural frequency and is related to the j -th undamped natural frequency and damping ratio by

$$\bar{\omega}_j = \omega_j \sqrt{1 - \xi_j^2}. \quad (20)$$

The response of the entire structure can be explicitly written as follows:

$$\mathbf{U}(t) = \sum_{j=1}^n \phi_j^T \mathbf{M} \left[\mathbf{U}_0 \cos(\bar{\omega}_j t) + \frac{(\dot{\mathbf{U}}_0 + \xi_j \omega_j \mathbf{U}_0)}{\bar{\omega}_j} \sin(\bar{\omega}_j t) \right] e^{-\xi_j \omega_j t} \phi_j \quad (21)$$

Equation 21 shows that the classically damped vibration of an n -dof structure can be thought as a linear combination of normal modes ϕ_j , each of them vibrating with the associated circular frequency ω_j and decaying with damping ξ_j .

Forced Vibrations of Classically Damped MDOF Structures

Let us consider the forced vibration equations of motion of an n degree of freedom elastic structure with classical viscous damping:

$$\mathbf{M} \ddot{\mathbf{U}}(t) + \mathbf{C} \dot{\mathbf{U}}(t) + \mathbf{K} \mathbf{U}(t) = \mathbf{f}(t) \quad (22)$$

where $\mathbf{f}(t)$ is the time-dependent n dimensional force vector. Equation 16 represents a system of n coupled linear differential equations to be solved with the initial conditions (Eq. 2). Using the modal coordinates $\mathbf{q}(t)$ transform, Eq. 6 and Eq. 22, and left multiplying by ϕ^T , because of the orthogonality conditions (Eqs. 5 and 12), the

following n uncoupled forced differential equation of motion in modal coordinates are obtained:

$$\ddot{\mathbf{q}}(t) + \mathbf{\Sigma} \dot{\mathbf{q}}(t) + \mathbf{\Omega}^2 \mathbf{q}(t) = \mathbf{p}(t) \quad (23)$$

having set

$$\mathbf{p}(t) = \phi^T \mathbf{f}(t) \quad (24)$$

to be solved with initial conditions

$$\mathbf{q}(0) = \phi^T \mathbf{U}(0), \quad \dot{\mathbf{q}}(0) = \phi^T \dot{\mathbf{U}}(0). \quad (25)$$

Equation 23 in explicit form becomes

$$\ddot{q}_j(t) + 2\xi_j \omega_j \dot{q}_j(t) + \omega_j^2 q_j(t) = p_j(t). \quad (26)$$

Solution of Eq. 26 can be split in the *homogenous* $q_j^0(t)$ and *particular* $q_j^p(t)$ integrals

$$q_j(t) = q_j^0(t) + q_j^p(t). \quad (27)$$

The homogenous solution depends only by the initial conditions and is given in Eq. 19, while the particular integral depends by the analytical form of the excitation and, for the rest initial conditions, can be simply written as

$$q_j^p(t) = \int_0^t h_j(t - \tau) p_j(\tau) d\tau. \quad (28)$$

In Eq. 28, the integral term is known as the *Duhamel integral*; the kernel of the integral $h_j(\cdot)$ is the response of the j -th oscillator as defined in Eq. 26 to an impulse idealized as Dirac's delta centered in $t = 0$ and is given by

$$h_j(t) = \frac{1}{\omega_j} e^{-\xi_j \omega_j t} \sin(\bar{\omega}_j t). \quad (29)$$

Vibrations of Classically Damped MDOF Structures in the Frequency Domain

Let us apply the Fourier transform operator to the force and response vectors:

$$\begin{aligned} \mathbf{F}(\omega) &= \int_{-\infty}^{+\infty} \mathbf{f}(t) e^{-i\omega t} dt, \\ \mathbf{Q}(\omega) &= \int_{-\infty}^{+\infty} \mathbf{q}(t) e^{-i\omega t} dt. \end{aligned} \quad (30)$$

It can be easily shown that applying the Fourier operator to both sides of Eq. 23, the following input–output relationship is found:

$$\mathbf{Q}(\omega) = \mathbf{H}_M(\omega) \phi^T \mathbf{F}(\omega). \quad (31)$$

In Eq. 31, $\mathbf{H}_M(\omega)$ is the *transfer modal matrix*, diagonal of order n , whose j -th element is given by

$$H_{Mj}(\omega) = \frac{1}{\omega_j^2 - \omega^2 + 2i\xi_j \omega_j \omega}. \quad (32)$$

Once $\mathbf{H}_M(\omega)$ is known, the nodal response transform

$$\mathbf{U}(\omega) = \int_{-\infty}^{+\infty} \mathbf{f}(t) e^{-i\omega t} dt \quad (33)$$

is given by

$$\mathbf{U}(\omega) = \mathbf{H}(\omega) \mathbf{F}(\omega). \quad (34)$$

In Eq. 34, $\mathbf{H}(\omega)$ is the *transfer matrix* of the structure of order n , defined as follows:

$$\mathbf{H}(\omega) = \phi \mathbf{H}_M(\omega) \phi^T = [\mathbf{K} - \omega^2 \mathbf{M} + i\omega \mathbf{C}]^{-1}. \quad (35)$$

Since $\mathbf{H}(\omega)$ is not diagonal, an easy to handle representation is

$$\mathbf{H}(\omega) = \sum_{j=1}^n \phi_j H_{Mj}(\omega) \phi_j^T. \quad (36)$$

Nonclassically Damped Vibrations of MDOF Structures

Nonclassically Damped Free Vibrations of MDOF Structures

When the damping is not classical, $\phi^T \mathbf{C} \phi$ is not a diagonal matrix. Under these circumstances, the natural frequencies, damping ratios, and modal vectors depend on the mass, stiffness, and damping matrices of the structural system. To determine the mode–shape vectors, natural frequencies, and damping ratios from $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$, it is convenient to write the second-order differential equation (Eq. 11) as two sets of first-order differential equations. Defining the velocity $\mathbf{V} = \dot{\mathbf{U}}$, so that $\dot{\mathbf{V}} = \ddot{\mathbf{U}}$, Eq. 11 can be rewritten as a set of two first-order differential equations in matrix form:

$$\frac{d}{dt} \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{M}^{-1} \mathbf{K} & -\mathbf{M}^{-1} \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix} \quad (37)$$

The matrix equation of motion Eq. 37 can be recast into the following first-order matrix equation:

$$\dot{\mathbf{Z}} = \mathbf{GZ} \quad (38)$$

where

$$\mathbf{Z} = \begin{bmatrix} \mathbf{U} \\ \dot{\mathbf{U}} \end{bmatrix}_{2n \times 1} \quad (39)$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{0}_{n \times n} & \mathbf{I}_{n \times n} \\ -\mathbf{M}^{-1}\mathbf{K} & -\mathbf{M}^{-1}\mathbf{C} \end{bmatrix}_{2n \times 2n} \quad (40)$$

The subscripts in Eqs. 39 and 40 indicate the dimensions of the vectors and matrices to which they are attached. Solution of Eq. 38 can be written as

$$\mathbf{Z} = \mathbf{\Psi} e^{\lambda t}. \quad (41)$$

Substituting Eq. 41 in Eq. 38 returns the eigenvalue problem:

$$\mathbf{G}\mathbf{\Psi} = \lambda\mathbf{\Psi} \quad (42)$$

The eigenvalue problem in Eq. 42 admits $2n$ complex eigenvectors $\mathbf{\Psi}_j$ and complex eigenvalues $\lambda_j (j = 1, \dots, 2n)$ occurring in pair complex conjugate. It follows that each complex eigenvector $\mathbf{\Psi}_k$ with complex eigenvalue λ_k is associated with a complex conjugate eigenvector $\mathbf{\Psi}_k^*$ with complex conjugate eigenvalue $\lambda_k^* (k = 1, \dots, n)$.

Once the complex eigenvalues problem in Eq. 42 is solved, the solution of the equation of motion (Eq. 37) is given as follows:

$$\mathbf{U}(t) = \sum_{j=1}^{2n} C_j e^{\lambda_j t} \mathbf{\Psi}_j = 2 \sum_{j=1}^n \Re \{ C_j e^{\lambda_j t} \mathbf{\Psi}_j \}. \quad (43)$$

The $2n$ complex coefficients C_j depend on the initial conditions of motion and are given by

$$C_j = \mathbf{\Psi}_j^T (\lambda_j \mathbf{M} + \mathbf{C}) \mathbf{U}_0 + \mathbf{\Psi}_j^T \mathbf{M} \dot{\mathbf{U}}_0. \quad (44)$$

Vibrations of Nonclassically Damped MDOF Structures in the Frequency Domain

Applying the Fourier operator to both sides of the equation of motion in the nonclassically damped case, the following input–output relationship is found:

$$\mathbf{Q}(\omega) = \mathbf{H}_M(\omega) \phi^T \mathbf{F}(\omega) \quad (45)$$

In Eq. 45, $\mathbf{H}_M(\omega)$ is the $n \times n$ transfer modal matrix, for nonclassically damped systems non-diagonal, defined as follows:

$$\mathbf{H}_M(\omega) = [\mathbf{K} - \omega^2 \mathbf{M} + i\omega \mathbf{C}]^{-1}. \quad (46)$$

If the damping matrix $\mathbf{C}$ is decoupled in the sum of a diagonal matrix $\mathbf{C}_D$ collecting all the elements of the principal diagonal of $\mathbf{C}$ and a non-diagonal matrix $\mathbf{C}_F$ collecting the elements out of the principal diagonal of $\mathbf{C}$

$$\mathbf{C} = \mathbf{C}_D + \mathbf{C}_F \quad (47)$$

then defining the diagonal matrix $\mathbf{H}_D(\omega)$

$$\mathbf{H}_D(\omega) = [\mathbf{K} - \omega^2 \mathbf{M} + i\omega \mathbf{C}_D]^{-1}, \quad (48)$$

the following approximate representation holds:

$$\mathbf{H}_M(\omega) = \mathbf{H}_D(\omega) \left[\mathbf{I} + \sum_{j=1}^n (-i\omega)^j [\mathbf{C}_F \mathbf{H}_D(\omega)]^j \right] \quad (49)$$

Nonclassically Damped Vibrations of MDOF Structures Under Ground Excitation

A state-space formulation of the equations of motion for a linear *MDOF* system is useful to describe the response of both classically and nonclassically damped systems. The general (second-order) equations of motion for an n -degree-of-freedom linear structure under ground excitation are in matrix form:

$$\mathbf{M}\ddot{\mathbf{U}} + \mathbf{C}\dot{\mathbf{U}} + \mathbf{K}\mathbf{U} = \mathbf{P}\mathbf{F}(t) \quad (50)$$

where $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are the $n \times n$ time-invariant mass, damping, and stiffness matrices, respectively; $\mathbf{U}(t)$ is the length- n vector of nodal displacements, $\mathbf{P}$ is the length- n load distribution vector, $\mathbf{F}(t)$ is the scalar function describing the time history of the external loading due to ground motion, and a superposed dot denotes differentiation with respect to time. The matrix equation of motion Eq. 50 can be recast into the following first-order matrix equation:

$$\dot{\mathbf{Z}} = \mathbf{G}\mathbf{Z} + \tilde{\mathbf{P}}\mathbf{F}(t) \quad (51)$$

where $\mathbf{Z}$ and $\mathbf{G}$ are defined in Eqs. 39 and 40, while $\tilde{\mathbf{P}}$ is given by

$$\tilde{\mathbf{P}} = \begin{bmatrix} \mathbf{0}_{n \times 1} \\ \mathbf{M}^{-1} \mathbf{P} \end{bmatrix}. \quad (52)$$

The subscripts in Eq. 52 indicate the dimensions of the vectors to which they are attached. Using the complex modal matrix $\mathbf{T}$ formed from the complex eigenmodes of matrix $\mathbf{G}$, the first-order differential matrix equation (Eq. 51) can be decoupled into the following $2n$ normalized complex modal equations:

$$\dot{S}_i(t) = \lambda_i S_i(t) + F(t), \quad i = 1, 2, \dots, 2n \quad (53)$$

where $\mathbf{S} = [S_1(t) S_2(t) \dots S_{2n}(t)]^T$ is a normalized complex modal response vector and $\lambda_i (i = 1, \dots, 2n)$ are the complex eigenvalues of the system matrix $\mathbf{G}$. The response of the linear MDOF system can be obtained as

$$\mathbf{Z}(t) = \mathbf{T} \mathbf{S}(t) = \tilde{\mathbf{T}} \mathbf{S}(t) \quad (54)$$

in which $\mathbf{\Gamma}$ = diagonal matrix containing the $2n$ modal participation factors Γ_i , defined as the i -th component of vector $\mathbf{T}^{-1} \tilde{\mathbf{P}} = [\Gamma_1 \dots \Gamma_{2n}]^T$, and $\tilde{\mathbf{T}} = \mathbf{T} \mathbf{\Gamma}$ = effective modal participation matrix. Assuming that the system is initially at rest, the solution of Eq. 53 can be expressed by the following Duhamel integral:

$$S_i(t) = \int_0^t e^{\lambda_i(t-\tau)} F(\tau) d\tau, \quad i = 1, 2, \dots, 2n \quad (55)$$

It is worth mentioning that the normalized complex modal responses $S_i(t)$, $i = 1, 2, \dots, 2n$ are complex conjugate by pairs.

Nonclassically Damped Forced Vibrations of MDOF Structures Under Earthquake Excitation

It is well known that the earthquake excitation can be represented as a nonstationary process, and a correct definition has to be addressed through the theory of stochastic process (Chopra 1995).

A nonstationary stochastic process (NSSP) $X(t)$ can be expressed in the general form of a Fourier–Stieltjes integral as (Priestley 1987)

$$X(t) = \int_{-\infty}^{\infty} A_X(\omega, t) e^{j\omega t} dZ(\omega) \quad (56)$$

where t is the time, ω is the frequency parameter, $j = \sqrt{-1}$, $A_X(\omega, t)$ is the complex-valued deterministic time-frequency modulating function, and $dZ(\omega)$ is the zero-mean orthogonal increment process defined so that $E[dZ^*(\omega_1) dZ(\omega_2)] = \Phi(\omega_1) \delta(\omega_1 - \omega_2) d\omega_1 d\omega_2$, where $E[\cdot]$ is the mathematical expectation, $\Phi(\omega)$ is the power spectral density (PSD) function of the embedded stationary process $X_S(t)$, $\delta(\cdot)$ is the Dirac delta function, and the superscript $(\cdot)^*$ denotes the complex-conjugate operator. If the process $X(t)$ is real valued, the following property holds:

$$A_X(-\omega, t) = A_X^*(\omega, t) \quad (57)$$

The process $X(t)$ has the following evolutionary power spectral density (EPSD) function:

$$\Phi_{X,X}(\omega, t) = A_X^*(\omega, t) \Phi(\omega) A_X(\omega, t) \quad (58)$$

In the case of a nonstationary loading process, the loading function $F(t)$ can be expressed as

$$F(t) = \int_{-\infty}^{\infty} A_F(\omega, t) e^{j\omega t} dZ(\omega). \quad (59)$$

It can be shown that the normalized complex modal responses are given by

$$S_i(t) = \int_{-\infty}^{\infty} A_{S_i}(\omega, t) e^{j\omega t} dZ(\omega), \quad i = 1, 2, \dots, 2n \quad (60)$$

where

$$A_{S_i}(\omega, t) = \int_{-\infty}^{\infty} e^{(\lambda_i - j\omega)(t-\tau)} A_F(\omega, \tau) d\tau, \quad i = 1, 2, \dots, 2n. \quad (61)$$

It is also convenient to define the process $Y(t)$ as the modulation (with modulating function

$A_X(\omega, t)$ of the stationary process $Y_S(t)$ defined as the Hilbert transform of the embedded stationary process $X_S(t)$, i.e.,

$$Y(t) = -j \int_{-\infty}^{\infty} \text{sign}(\omega) A_X(\omega, t) e^{j\omega t} dZ(\omega). \quad (62)$$

The cross variance between the response process $X(t)$ and the modulation process $Y(t)$ is used in the definition of the time-variant central frequency, $\omega_c(t)$, and bandwidth parameter, $q(t)$, of the *NSSP* $X(t)$ as

$$\omega_c(t) = \frac{\sigma_{XY}(t)}{\sigma_X^2(t)} \quad (63)$$

$$q(t) = \left(1 - \frac{\sigma_{XY}^2(t)}{\sigma_X^2(t) \sigma_Y^2(t)} \right) \quad (64)$$

The time-variant central frequency and bandwidth parameter are useful in describing the time-variant spectral properties of a real-valued *NSSP* $X(t)$ (Barbato and Conte 2008). The central frequency $\omega_c(t)$ provides the characteristic/predominant frequency of the process at each instant of time. The bandwidth parameter $q(t)$ provides information on the spectral bandwidth of the process at each instant of time. Notice that a *NSSP* can behave as a narrowband and a broadband process at different instants of time. For complex-valued *NSSPs*, the complex-valued central frequency and bandwidth parameter defined in Eqs. 63 and 64 lose the simple physical interpretation available for real-valued *NSSPs*, even though their computation is instrumental to the solution of problems requiring a state-space representation. In addition, the bandwidth parameter $q(t)$ plays an important role in time-variant reliability analysis, since it is an essential ingredient of analytical approximations to the time-variant failure probability for the first-passage reliability problem.

Response Spectral Characterization of Nonclassically Damped MDOF Linear Systems Using Complex Modal Analysis

The state-space formulation of the equations of motion is also advantageous for the computation of the *NGSCs* of response processes of both classically and nonclassically damped linear *MDOF* systems. If only Gaussian input processes are considered, only few spectral characteristics are needed to fully describe the response processes of linear elastic *MDOF* systems, since the response processes are also Gaussian. In particular, if $U_i(t)$ denotes the i -th *DOF* displacement response process of a linear elastic *MDOF* system subjected to Gaussian excitation, the only spectral characteristics required, e.g., for reliability applications, are $\sigma_{U_i}^2(t)$, $\sigma_{\dot{U}_i}^2(t)$, $\sigma_{U_i \dot{U}_i}(t)$ and $\sigma_{U_i \dot{Y}_i}(t)$, where $\dot{Y}_i$ is the first time derivative of the process Y_i defined as

$$\mathcal{Y}(t) = -j \int_{-\infty}^{\infty} \text{sign}(\omega) A_{U_i}(\omega, t) e^{j\omega t} dZ(\omega), \quad i = 1, 2, \dots, n \quad (65)$$

and $A_{U_i}(\omega, t)$ the time-frequency modulating function of the process $U_i(t)$ (Barbato and Conte 2008). The process $\mathcal{Y}(t)$ is the modulation (with the same modulating function $A_{U_i}(\omega, t)$ as process $U_i(t)$) of the Hilbert transform of the stationary process embedded in the process $U_i(t)$. Similarly to the response processes, the following auxiliary state vector process can be defined:

$$\Xi = \begin{bmatrix} \mathcal{Y} \\ \dot{\mathcal{Y}} \end{bmatrix}_{(2n \times 1)} \quad (66)$$

Using complex modal decomposition, the cross-covariance matrices of the response processes and the auxiliary processes can be computed as (Barbato and Vasta 2010)

$$\begin{aligned} E[\mathbf{Z}(t) \mathbf{Z}^T(t)] &= \begin{bmatrix} \mathbf{U}(t) \mathbf{U}^T(t) & \mathbf{U}(t) \dot{\mathbf{U}}^T(t) \\ \dot{\mathbf{U}}(t) \mathbf{U}^T(t) & \dot{\mathbf{U}}(t) \dot{\mathbf{U}}^T(t) \end{bmatrix} \\ &= \tilde{\mathbf{T}}^* E[\mathbf{S}^*(t) \mathbf{S}^T(t)] \tilde{\mathbf{T}}^T \end{aligned} \quad (67)$$

$$\begin{aligned}
 E[\mathbf{Z}(t)\mathbf{\Xi}^T(t)] &= \begin{bmatrix} \mathbf{U}(t)\mathcal{Y}^T(t) & \mathbf{U}(t)\dot{\mathcal{Y}}^T(t) \\ \dot{\mathbf{U}}(t)\mathcal{Y}^T(t) & \dot{\mathbf{U}}(t)\dot{\mathcal{Y}}^T(t) \end{bmatrix} \\
 &= \tilde{\mathbf{T}}^* E[\mathbf{S}^*(t)\mathbf{\Sigma}^T(t)] \tilde{\mathbf{T}}^T
 \end{aligned} \quad (68)$$

where the components of the vector process $\mathbf{\Sigma} = [\Sigma_1(t)\Sigma_2(t) \dots \Sigma_{2n}(t)]^T$ are defined as

$$\begin{aligned}
 \Sigma_i(t) &= -j \int_{-\infty}^{\infty} \text{sign}(\omega) A_{S_i}(\omega, t) e^{j\omega t} dZ(\omega), \\
 i &= 1, 2, \dots, n
 \end{aligned} \quad (69)$$

Eqs. 67 and 68 show that all quantities required for reliability applications can be computed from the following spectral characteristics of complex-valued nonstationary processes ($i, m = 1, 2, \dots, 2n$)

$$\begin{cases} E[S_i^*(t)S_m(t)] = \sigma_{S_i S_m}(t) \\ E[S_i^*(t)\Sigma_m(t)] = \sigma_{S_i \Sigma_m}(t) \end{cases} \quad (70)$$

Response Statistics of MDOF Linear Systems Subjected to Modulated Gaussian-Colored Noise

Time-modulated Gaussian-colored noises constitute an important class of nonstationary dynamic load processes. The expression describing a general nonstationary loading process reduces to

$$F(t) = A_F(t)P(t) \quad (71)$$

where the time-modulating functions $A_F(t)$ are frequency independent, and the process $P(t)$ is a colored noise with PSD function having the following general expression (i.e., rational function):

$$\Phi_P(\omega) = S_0 \sum_{k=1}^N \frac{A_k}{(\omega - \bar{\omega}_k)} \quad (72)$$

In the sequel, it is assumed that the time-modulating functions have the following general expression:

$$A_F(t) = H(t) \sum_{q=1}^M a_q e^{b_q t} \quad (73)$$

Substituting Eq. 73 into Eq. 61 yields ($i = 1, 2, \dots, 2n$)

$$A_{S_i}(\omega, t) = j \sum_{q=1}^M \left\{ a_q e^{b_q t} \left[\frac{e^{-j(\omega - \omega_{1q})t-1}}{\omega - \omega_{1q}} \right] \right\} \quad (74)$$

where $\omega_{1q} = -j\lambda_i$ and $\omega_{1q} = \omega_1 + jb_q$. The spectral characteristics defined in Eq. 70₁ can be computed using Cauchy's residue theorem as ($i, m = 1, 2, \dots, 2n$)

$$\sigma_{S_i S_m}(t) = S_0 \sum_{q,s=1}^M \sum_{k=1}^N a_q a_s A_k \left\{ \left[e^{(b_q + b_s)t} + 1 \right] I_1^{iq, ms, k} - e^{-j\omega_{1q}^* t} I_2^{iq, ms, k} - e^{j\omega_{ms} t} I_3^{iq, ms, k} \right\} \quad (75)$$

in which

$$I_1^{iq, ms, k} = 2\pi j \sum_{r=1}^3 B_r^{iq, ms, k} I(\tilde{\omega}_r) \quad (76)$$

$$I_2^{iq, ms, k} = 2\pi j \sum_{r=1}^3 B_r^{iq, ms, k} e^{j\tilde{\omega}_r t} I(\tilde{\omega}_r) \quad (77)$$

$$I_3^{iq, ms, k} = 2\pi j \sum_{r=1}^3 B_r^{iq, ms, k} e^{-j\tilde{\omega}_r t} I(-\tilde{\omega}_r) \quad (78)$$

where $\tilde{\omega}_1 = \omega_{1q}^*$, $\tilde{\omega}_2 = \omega_{ms}$, $\tilde{\omega}_3 = \bar{\omega}_k$ and

$$\begin{aligned}
 B_r^{iq, ms, k} &= \frac{1}{(\tilde{\omega}_r - \tilde{\omega}_p)(\tilde{\omega}_r - \tilde{\omega}_q)}, \\
 r, p, q &= 1, 2, 3 \quad r \neq p \neq q
 \end{aligned} \quad (79)$$

while

$$I(\omega) = \max \left[\frac{\Im(\omega)}{|\Im(\omega)|}, 0 \right] \quad (80)$$

After extensive algebraic manipulations, the spectral characteristics in Eq. 70₂ are obtained as $(i, m = 1, 2, \dots, 2n)$

$$\sigma_{S_i \Sigma_m}(t) = -jS_0 \sum_{q,s=1}^M \sum_{k=1}^N a_q a_s A_k e^{(b_q+b_s)t} \left\{ \left[e^{j(\omega_{ms}-\omega_{1q}^*)t} + 1 \right] J_1^{iq,ms,k} - e^{-j\omega_{1q}^*t} J_2^{iq,ms,k} - e^{j\omega_{ms}t} J_3^{iq,ms,k} \right\} \quad (81)$$

in which

$$E_1(x) = \int_x^\infty \frac{e^{-u}}{u} du, \quad |\arg(x)| < \pi \quad (85)$$

$$J_1^{iq,ms,k} = -\sum_{r=1}^3 \left\{ B_r^{iq,ms,k} [\log(\tilde{\omega}_r) + \log(-\tilde{\omega}_r)] \right\} \quad (82)$$

A very useful modulating function in earthquake engineering is the one of Shinozuka and Sato, defined as

$$J_2^{iq,ms,k} = 2 \sum_{r=1}^3 B_r^{iq,ms,k} \left\{ e^{\tilde{\omega}_r t} [E_1(j\tilde{\omega}_r t) + j\pi I(\tilde{\omega}_r)] \right\} \quad (83)$$

$$A_F(t) = C [e^{-B_1 t} - e^{-B_2 t}] H(t) \quad (86)$$

where

$$J_3^{iq,ms,k} = 2 \sum_{r=1}^3 B_r^{iq,ms,k} \left\{ e^{-j\tilde{\omega}_r t} [E_1(-j\tilde{\omega}_r t) + j\pi I(-\tilde{\omega}_r)] \right\} \quad (84)$$

$$C = \left[\frac{B_1}{B_2 - B_1} \right] e^{\frac{B_2}{B_2 - B_1} \log\left(\frac{B_2}{B_1}\right)} \quad (87)$$

where $E_1(\cdot)$ denotes the integral exponential function:

and $B_2 > B_1 > 0$. With this position, Eq. 81 reduces to $(i, m = 1, 2, \dots, 2n)$

$$\sigma_{S_i \Sigma_m}(t) = \frac{CS_0}{4\pi} \sum_{q,s=1}^2 \sum_{k=1}^4 (-1)^{k+s+q} e^{(B_q+B_s)t} \left\{ \left[e^{j(\omega_{ms}-\omega_{1q}^*)t} + 1 \right] J_1^{iq,ms,k} - e^{-j\omega_{1q}^*t} J_2^{iq,ms,k} - e^{j\omega_{ms}t} J_3^{iq,ms,k} \right\} \quad (88)$$

where $\omega_{1q} = \omega_1 - jB_q$ ($q = 1, 2$).

Summary

Response analysis of MDOF linear structures showing classical or nonclassical viscous

damping are revised and discussed. Attention is focused on earthquake excitation acting on nonclassically damped linear structures. Spectral characteristics are employed to evaluate the time-variant central frequency and bandwidth parameters, useful for assessing the reliability of earthquake-sensible structures.

Cross-References

- [Structures with Nonviscous Damping, Modeling, and Analysis](#)
- [Tuned Mass Dampers for Passive Control of Structures Under Earthquake Excitations](#)

References

- Argyris J, Mlejnek HP (1991) *Dynamics of structures*. North Holland, Amsterdam
- Barbato M, Conte JP (2008) Spectral characteristics of non-stationary random processes: theory and applications to linear structural models. *Probabilist Eng Mech* 23(4):416–426
- Barbato M, Vasta M (2010) Closed-form solutions for the time-variant spectral characteristics of non-stationary random processes. *Probabilist Eng Mech* 25(1):9–17
- Chopra AK (1995) *Dynamics of structures – theory and applications to earthquake engineering*. Prentice Hall, Upper Saddle River. Amsterdam, 1991
- Clough RW, Penzien J (1993) *Dynamics of structures*. McGraw-Hill, New York
- Lin YK (1976) *Probabilistic theory of structural dynamics*. McGraw-Hill, New York. 1967, Krieger, Huntington
- Priestley MB (1987) *Spectral analysis and time series*, volume 1: univariate series, volume 2: multivariate series, prediction and control. Academic, London. Fifth Printing
- Rayleigh L (1877) *Theory of sound*, 2 vols, 1945th edn. Dover, New York

Code-Based Design: Seismic Isolation of Buildings

Troy A. Morgan
Exponent, New York, USA

Synonyms

Base isolation; Friction pendulum; High-damping rubber; Lead rubber; Passive control

Introduction

One of the most significant developments in structural engineering in the past 20 years has

undoubtedly been the emergence of performance-based design as a means of selecting, proportioning, and building structural systems to resist seismic excitations. This methodology is an ideal framework for design due to its flexibility with respect to the selection of performance objectives, the characterization and simulation of both demand and resistance, and the overarching treatment of uncertainty. A great strength of the methodology is that performance objectives may be defined in terms of structural performance, architectural function, socioeconomic considerations, and environmental sustainability. This framework has the attractive feature of providing a metric of performance that can be implemented by a wide variety of infrastructure stakeholders, including architects, building owners, contractors, insurance providers, capital investment proprietors, and public officials. As structural engineers train their focus on broadly defined solutions to the challenges posed by maintaining and improving civilization, performance-based design will increasingly play a central role. This innovative approach to design demands the use of innovative structural systems to achieve the complex and potentially multi-objective performance goals envisioned by the various stakeholders. Given the uncertainty that is unavoidably present in any earthquake-resistant design framework, innovative systems must not only be capable of predictable response to deterministic input but also be sufficiently robust to respond reliably to a broad range of potential input.

Seismic isolation systems are ideally suited for implementation within a performance-based framework because: (a) robust characterizations of their behavior can be made through experimentation, (b) the variance of observed behavior from expected is often low relative to conventional structural elements, and (c) it can be challenging or even impossible to achieve an enhanced performance objective without the use of seismic isolation. Compared to conventional structural system for seismic resistance, isolation provided a unique and reliable means of simultaneously reducing earthquake damage in both

deformation-sensitive and acceleration-sensitive elements.

Performance-based seismic design of structures is currently undergoing significant development in response to consequences experienced in recent earthquakes. Not only has there been substantial loss of human life as a result of damage caused by major earthquakes, the economic toll resulting from direct losses (repair of infrastructure, replacement of damaged contents) and indirect losses (business disruptions, relocation expenses, supply chain interruption) has also been significant. As a result of the significant socioeconomic turmoil due to the occurrence of earthquakes worldwide, innovative systems such as seismic isolation have emerged as a reliable approach to limiting damage and providing post-earthquake functionality and operability.

Principles of Seismic Isolation

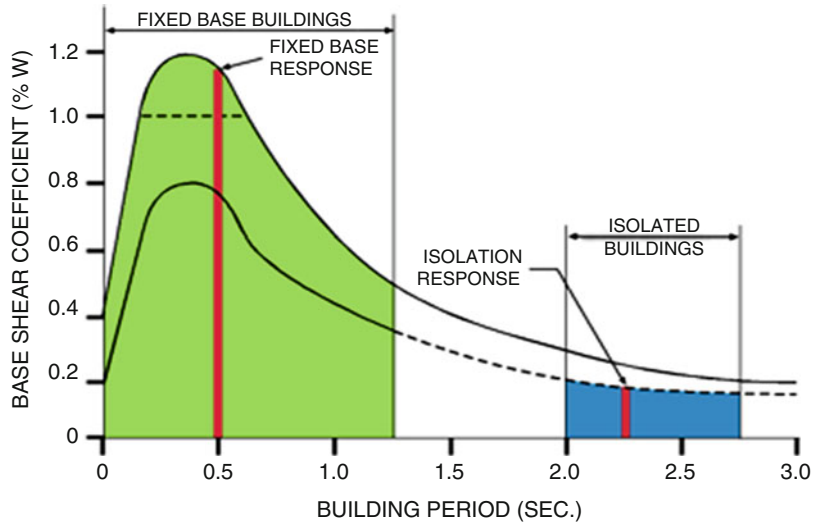
Seismic isolation is rooted in the need to control structural response due to harmonic vibrations. This need stems from (a) the discomfort caused to occupants as a result of oscillatory motion of floor-supported equipment and (b) the potential for damage to sensitive equipment caused by vibrations of the supporting structure. The source of such vibrations has traditionally been rotating machinery, ambient traffic conditions, and walker-induced floor vibrations. A common method of reducing the accelerations due to harmonic excitation is to provide a compliant base (either from steel springs or elastomeric pads) that adjusts the natural frequency of the supported equipment such that it is unable to reach resonance under the operating frequency of the excitation. This idea is behind the concept of *transmissibility*, or the ratio of the response amplitude to the input amplitude. For example, in vibration isolation, one can define the transmissibility as the ratio of the peak total acceleration of the equipment to the peak input acceleration of the support. The transmissibility is generally a function of the forcing frequency, the natural frequency of the supporting hardware,

and the amount and type of damping present. Whereas the amplitude and forcing frequency may be a function of the equipment's operating speed and weight, the natural frequency and damping of the supporting hardware may be adjusted to limit the total acceleration due to harmonic excitation.

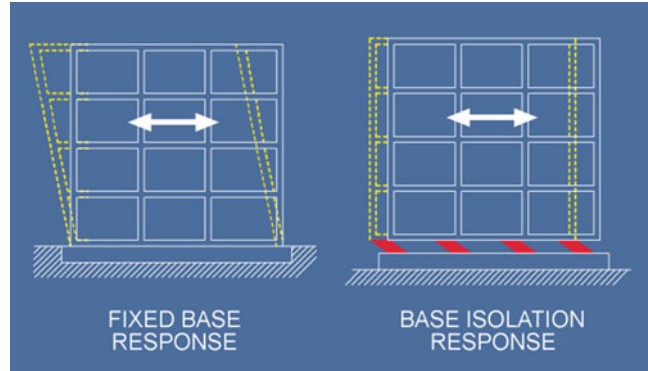
The reduction of transmissibility is also the goal of seismic isolation. Unlike traditional equipment isolation, however, the excitation is due to ground shaking and therefore cannot be characterized by harmonic input. Ground shaking resulting from a seismic event is stochastic in nature, and this excitation may contain a rich array of frequency content. Data collected from numerous historical seismic events has demonstrated that the predominant frequency of seismic excitation is generally above 1 Hz, and hence systems having a lower natural frequency than this will experience a reduction in transmissibility of acceleration from the ground into the structure. The reduction of natural frequency (or elongation of natural period) provides the so-called "decoupling" of the motion of the structure with that of the ground, as shown in Fig. 1. This reduction in natural frequency has the undesirable consequence of increasing deformation demand under the input excitation. These undesirable deformations can be mitigated to some extent through the addition of damping, typically through a combination of velocity-dependent (termed viscous damping) and deformation-dependent (termed hysteretic damping) energy dissipation mechanisms. As a result, a properly designed isolation system will have the appropriate combination of stiffness and damping such that substantial dynamic decoupling is achieved without detrimental deformation demands in the isolation hardware.

Period elongation is not, however, the only necessary component of seismic isolation. Indeed, if adding flexibility were sufficient to reduce seismic response, then an effective design strategy would be to use the most flexible members possible given the consideration of stability under gravity. An equally important aspect of isolation is the change in fundamental mode

Code-Based Design:
Seismic Isolation of
Buildings, Fig. 1 Period
 shift strategy behind
 seismic isolation



Code-Based Design:
Seismic Isolation of
Buildings, Fig. 2 Comparison of
 fundamental mode shape
 for fixed-base and base-
 isolated buildings



shape. The introduction of a layer that is compliant relative to the supported structure introduces a key modification to the free vibration characteristics of the structure. The more similar the fundamental mode shape is to rigid-body behavior, the less mass participation is present in higher modes. Such rigid-body behavior is associated with the relative compliance of the isolation layer. Hence, as the natural period of the isolation system increases relative to the natural period of the supported structure, the participation of higher modes becomes closer to zero and the seismic deformation is concentrated at the isolation layer and not in the superstructure (Fig. 2).

If the isolated structure has total weight W , and the total lateral stiffness of the isolation system is

given by k_b , the isolated natural frequency is defined as

$$\omega_b = \sqrt{\frac{k_b g}{W}} \quad (1)$$

This well-known expression assumes that the mass above the isolation system is rigid, and all lateral deformation takes place in the isolators. This is an approximation and is clearly dependent on the natural frequency of the supported structure, ω_s . However, for reasonable large frequency ratios ω_s/ω_b , this rigid structure assumption is a very good approximation of reality.

Another important parameter for isolated structures is the critical damping ratio of the isolation system, ζ_b . This damping is defined by

the isolation bearings themselves and any supplemental damping devices. The general expression for equivalent damping of an isolation system is given by

$$\zeta_b = \frac{E_d}{4\pi E_s} \quad (2)$$

where E_d is the energy dissipated in one complete cycle of displacement u_0 and E_s is the energy stored at a displacement equal to u_0 . Both parameters of stiffness and damping for specific isolation devices will be further described in the following section.

Preliminary Design of Isolated Structures

The preliminary design of isolated structures generally begins with an assumption of equivalent linear properties for the isolation system and a rigid superstructure. This design approach is sufficient to obtain, for a given level of seismic hazard, design forces and interstory drift demands for the superstructure and peak isolator displacements.

Suppose an isolated structure is being designed at a site characterized by a design-level elastic acceleration response spectrum $A(\omega_n, \zeta)$, where ω_n is the natural frequency and ζ is the viscous damping ratio. The isolation system stiffness k_b is the sum of the individual stiffnesses of each bearing, and the viscous damping ζ_b of the system is computed from Eq. 2. If the total seismic weight supported above the isolation system is W , the isolated natural frequency ω_b is given by Eq. 1. As a result, the design base shear for the isolated structure would simply be computed as

$$V_b = \frac{A(\omega_b, \zeta_b)}{g} W \quad (3)$$

This equation treats the superstructure as a rigid mass single degree-of-freedom system oscillator whose stiffness and damping are entirely defined by the isolation system. This base shear represents the total lateral force delivered to the superstructure given the frequency and damping of the isolation system. This base shear may be

distributed to the superstructure using the first mode shape, and these forces may be used to compute member forces, interstory drifts, etc. While isolated buildings are generally designed to be essentially elastic for design ground motion, the base shear in Eq. 3 may be reduced by an overstrength-based force reduction factor R_Ω . This factor is much less than the force reduction factors R used for ductile fixed-base buildings, since both ductility and overstrength are considered, and conventional buildings are expected to undergo potentially large inelastic deformations in the design earthquake.

The maximum displacement of the isolation system, D_b , is computed from the basic expression for a spring force, $V_b = k_b D_b$. Noting that $k_b = \omega_{n_g}^2 W$ and the relationship between spectral acceleration and displacement is given by $A(\omega_n, \zeta) = \omega_n^2 D(\omega_n, \zeta)$, the maximum isolator displacement is simply

$$D_b = D(\omega_b, \zeta_b)$$

The preliminary design of the individual isolators will be based on the required shear stiffness and damping and the stability of the isolator under the maximum axial load at the maximum displacement D_b . The final design process is more involved and is based on the requirements of the governing building code.

Historical Development of Seismic Isolation

The roots of seismic isolation as a means for protecting structures from earthquakes go back to the latter part of the nineteenth century. While there appeared to be significant interest in creating an innovative seismic protective system, few opportunities were available for practical application. The first documented proposal for a seismic isolation system appears to be in an 1870 US patent filed by Jules Touaillon that describes a double-concave rolling ball bearing (Fenz and Constantinou 2006). Other early proposals include a 1906 US patent by J. Bechtold and a 1909 British patent by J. A. Calantarientis (Buckle

and Mayes (1990), Kelly (1986)). Both patents advocated base isolation through the separation of a building from its foundation, with low-friction material providing the necessary lateral flexibility. Despite the innovative design approaches embodied in these patents, none appear to have seen actual implementation.

The first application of seismic isolation was in 1885 by John Milne, a Professor of Mining Engineering at the University of Tokyo (Naeim and Kelly 1999). Having previously conducted experiments on sliding isolators, Milne designed and constructed a seismic isolation system for a building in Tokyo that incorporated ball bearings and cast iron plates formed in a spherical shape. The first application of seismic isolation using elastomeric bearings was in the Pestalozzi School located in Skopje, Yugoslavia, constructed in 1969 (Kelly 1986). The elastomeric bearings were constructed as blocks of rubber with no intermediate layers of steel reinforcement. This lack of confinement caused substantial vertical deflection under both gravity and lateral loading, resulting in an undesirable combination of horizontal and rocking modes of vibration under earthquake excitation. This early application of rubber in seismic design highlights the development that was underway and identifies the motivation for further research.

Modern seismic isolation originated in the mid-1960s with the New Zealand Department of Scientific and Industrial Research (Skinner et al. 1993). One of the outcomes of this research was the development of the lead rubber bearing, a system that relies on a laminated rubber bearing for horizontal flexibility and a press-fit central lead core to provide energy dissipation. Lead bearings have seen extensive use worldwide, the first application being the William Clayton Building constructed in Wellington, New Zealand, in 1981. Another type of elastomeric bearing is the high-damping rubber bearing, which was first used in the Foothill Communities Law and Justice Center in Rancho Cucamonga, USA, in 1985. The first modern sliding isolation system, the Friction Pendulum System, was developed in 1986 (Zayas et al. 1987) and was used in the retrofit of a wood-framed apartment

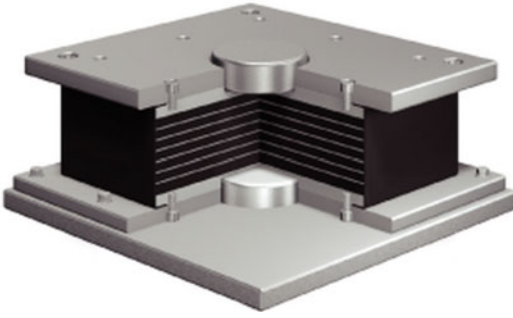
building in San Francisco, USA, that was damaged by the 1986 Loma Prieta earthquake.

Since the initial development of reliable, practical isolation devices in the early 1970s, seismic isolation has seen worldwide adoption. Detailed design provisions have been incorporated in several model building codes, and isolated buildings have reportedly been constructed in over 30 countries with the majority found in Japan, China, Russia, Italy, the USA, France, Armenia, Taiwan, and New Zealand (Martelli et al. 2012). Japan in particular has emerged as the worldwide leader in the application and advancement of seismic isolation technology. This was brought on in large part by the 1995 Hyogoken-Nanbu earthquake near Kobe, an event that highlighted the significant seismic risk to buildings and infrastructure in even the most developed countries (Clark et al. 2000).

Seismic Isolation Devices

A critical ingredient of seismic isolation is the introduction of a specially designed stratum between the structure and the foundation that is both horizontally flexible and vertically stiff. A stratum with these properties is generally achieved through a series of manufactured isolation bearings, whose required performance is specified on a set of contract documents, and acceptance is verified through prototype (quality assurance) and production (quality control) testing. Manufactured bearings primarily fall into two categories: elastomeric based or sliding based. Elastomeric-based bearings take advantage of the flexible properties of rubber to achieve isolation, while sliding-based bearings rely on the inherently low stiffness of a structure resting on its foundation with no connection other than friction at the interface. Devices that are designed and manufactured to effectively achieve the goals of seismic isolation possess three general characteristics:

1. High axial stiffness and strength to resist gravity and other vertical loads (both sustained and transient) without excessive deformation that



Code-Based Design: Seismic Isolation of Buildings,
Fig. 3 Typical laminated rubber bearing with cutaway showing internal construction (courtesy of Takenaka Corp., Japan)

would compromise the serviceability of a structure.

2. Sufficiently low horizontal stiffness such that the fundamental frequency of the isolated structure is substantially lower than the predominant frequency content of the expected ground motions, and adequate separation is provided between the natural frequency of the superstructure and that of the isolated structure.
3. An effective mechanism for energy dissipation to mitigate excessive lateral deformations which might cause instability of the isolation devices and connected structural elements due to combined horizontal and vertical forces.

Categories of devices currently manufactured to exhibit these characteristics are discussed in this section.

Elastomeric Isolation Bearings

Elastomeric isolation bearings are devices that make use of the flexibility of natural or synthetic rubber compounds to achieve the characteristics necessary for an effective isolation system. Such bearings have been widely used in bridge applications since the early part of the twentieth century to accommodate translational and rotational movement resulting from thermal expansion while supporting the weight of the deck (Fig. 3).

Low-Damping Rubber (LDR) Bearings

The most fundamental elastomeric isolation bearing is the low-damping rubber (LDR) bearing. This device consists of alternating layers of rubber and steel, providing the device with both high axial stiffness and low shear stiffness. The effective shear modulus of the rubber compound, G , is generally a function of the imposed shear strain γ . However, for natural rubber, there is a weak dependence of shear modulus on shear strain and may be assumed constant over a typical range of shear strain assumed in the design of isolation bearings subject to earthquakes. To derive the stiffness of an LDR bearing, first consider a single rubber layer having bonded area A_b and thickness t_r . Given an imposed shear stress τ , the total horizontal shear force acting on the layer is given as $V = \tau A_b$. From Hooke's law, the shear stress is $\tau = G\gamma$, so the total shear is $V = GA_b\gamma$. The shear displacement is simply $u = t_r\gamma$, giving shear force in terms of shear displacement

$$V = \frac{GA_b}{t_r} u \quad (4)$$

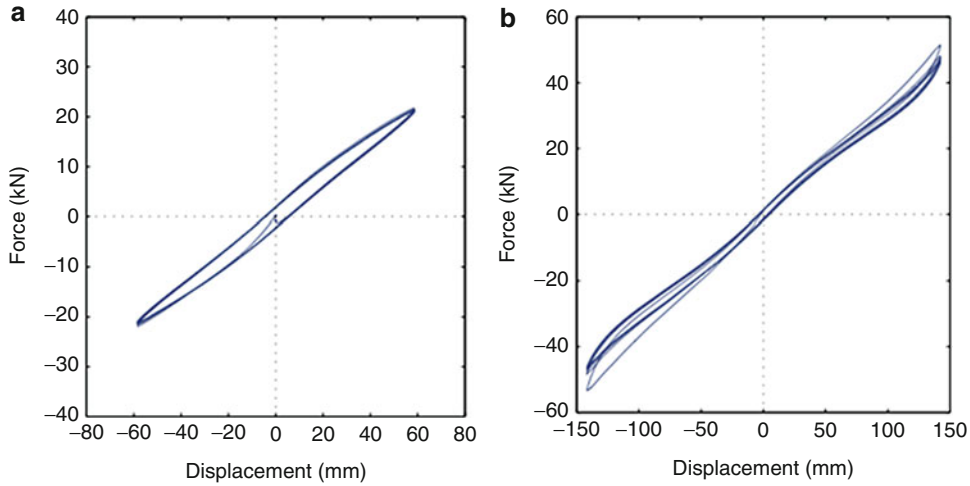
The layer stiffness is therefore GA_b/t_r . Since the total horizontal flexibility of the bearing is the sum of the individual layer flexibilities, considering a bearing with n rubber layers, the horizontal stiffness of an NR bearing is given by

$$k_b = \frac{GA_b}{T_r} \quad (5)$$

where the total rubber thickness is $T_r = nt_r$. Given an isolation system consisting of multiple isolators each having effective stiffness $k_{b,j}$, the natural period of the isolated building would be computed as

$$T_b = 2\pi \sqrt{\frac{W}{g \sum_j k_{b,j}}} \quad (6)$$

where W is the total weight of the supported structure. The vertical stiffness of an LDR bearing is given by



Code-Based Design: Seismic Isolation of Buildings, Fig. 4 Hysteresis for a low-damping rubber bearing (a) $\gamma = 75\%$; (b) $\gamma = 175\%$ (Constantinou et al. 2007)

$$k_v = \frac{E_c A_b}{T_r} \quad (7)$$

where E_c is the compression modulus of the rubber layer. This compression modulus is a function of the *shape factor* S , defined as the ratio of the bonded area to the area free to bulge. For circular bearings with bonded diameter D_b , the shape factor is computed as $S = D_b/4t_r$. For a rectangular bearing having dimensions $a \times b$, the shape factor is computed as

$$S = \frac{ab}{2t_r(a+b)} \quad (8)$$

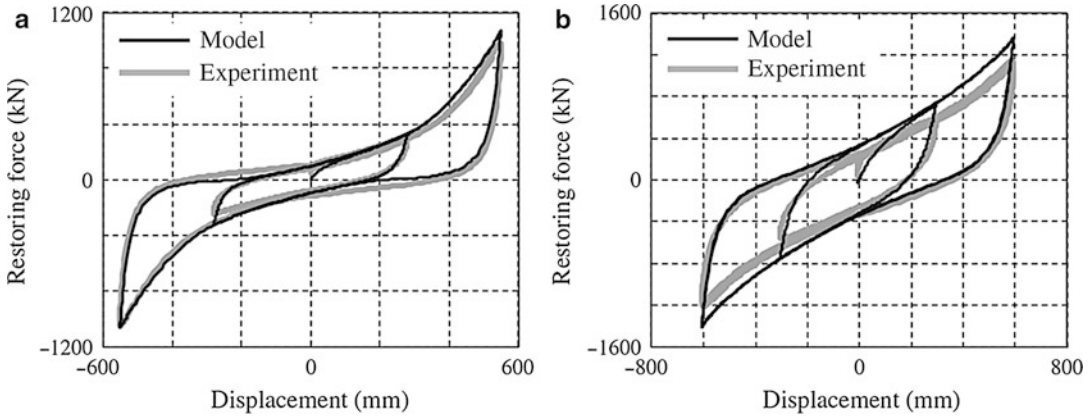
If the shape factor is sufficiently low, for circular bearings, $E_c = 6GS^2$, and for square bearings, $E_c = 6.73GS^2$ (Naeim and Kelly 1999). Typical shape factors for modern elastomeric bearings range from 10 to 30 and result in extremely high compressive stiffness. At these high shape factors, the bulk modulus K of the rubber becomes important. An expression for the compressive modulus incorporating the effect of the bulk modulus is given as

$$E_c = \frac{6GS^2K}{6GS^2 + K} \quad (9)$$

Elastomeric bearings in tension behave much differently than in compression. Whereas laminated rubber bearings can resist large compressive stress with no damage to the rubber matrix, tension stresses cause cavitation and significant loss of axial stiffness. For this reason, the design of isolated buildings is based in part of distributed overturning forces at the isolation layer such that uplift forces are minimized (Fig. 4).

High-Damping Rubber (HDR) Bearings

A class of elastomeric bearings that has seen wide implementation is high-damping rubber (HDR). These bearings are similar in construction to LDR bearings; however, the natural or synthetic rubber is specially compounded with carbon black, oils, resins, and other fillers that increase the inherent damping characteristics of the rubber (Clark and Kelly 1994). HDR bearings are characterized by effective shear modulus G_{eff} and equivalent viscous damping ζ_{eq} that is a function primarily of shear strain but also exhibits dependence on strain rate, strain history, and aging (Morgan et al. 2001). Several highly sophisticated models have been developed that are capable of predicting the behavior of HDR bearings over a wide range of response amplitudes (e.g., Yamamoto et al. (2012), among others). Such models rely on a combination of nonlinear elastic



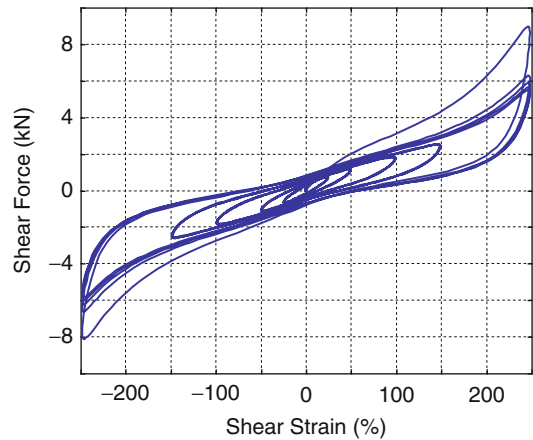
Code-Based Design: Seismic Isolation of Buildings, Fig. 5 Hysteresis of full-scale HDR bearings indicating (a) substantial strain crystallization and (b) low strain crystallization (Yamamoto et al. (2012))

springs, variable yield force plasticity elements, and variable stiffness springs. These models are then calibrated to experimental data.

A unique feature of HDR bearings is a phenomenon known as large strain crystallization, which causes an evident increase in stiffness above 120–150 % shear strain. Hysteresis loops for full-scale HDR bearings with and without significant large strain crystallization are shown in Fig. 5.

Another unique property of HDR bearing is the strain history dependence of both effective stiffness and equivalent viscous damping. Experimental programs (e.g., Morgan et al. 2001) have demonstrated the variability of measured stiffness and damping depending on the nature of the input loading. For example, the effective stiffness measured at 100 % shear strain (a common design parameter) could be markedly different for a test in which the bearing first experienced a shear strain of 250 % compared to a test that only includes strain amplitudes of 100 %. Another related effect is the increased stiffness measured in the first cycle of displacement compared to the stiffness in subsequent cycles. This effect is known as scragging and appears to be a persistent feature of HDR compounds. A hysteresis loop of an HDR bearing showing the effects of scragging is presented in Fig. 6.

Strain rate is also an important consideration in the modeling of HDR bearings. Testing of



Code-Based Design: Seismic Isolation of Buildings, Fig. 6 Hysteresis loop of reduced-scale HDR bearing showing scragging effect

HDR bearings at various frequencies has indicated that the measured stiffness and damping of the HDR compound is a function of both shear strain amplitude and velocity. As a result, any suitable model for an HDR bearing must include both hysteretic and viscous behavior to properly account for the likely energy dissipation characteristics over a broad range of seismic excitations. Further information on simple combined models capable of capturing this behavior is developed in Naeim and Kelly (1999).

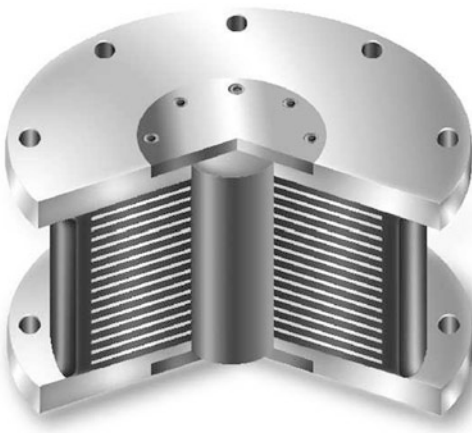
Lead Rubber (LR) Bearings

The lead rubber (LR) bearing consists of a low-damping natural rubber bearing, as described previously, with a press-fit central lead core. A cutaway section of a typical LR bearing is shown in Fig. 7. The bearing has linear behavior identical to the LRB bearing described above, and the stiffness is described by Eq. 5. The lead plug behaves as an elasto-perfectly plastic hysteretic element with a yield shear force $q_d = \tau_y A_{pb}$ where τ_y is the yield stress of lead and A_{pb} is the area of the lead core. Lead is an attractive material for the lead core since it is malleable, has a low yield stress that is stable for high levels of purity, and is readily available. A bilinear force–deformation loop for an LR bearing is shown in Fig. 8.

Most isolation devices are characterized by effective stiffness, k_{eff} , and equivalent viscous damping ζ_{eq} . For LR bearings, the effective stiffness assuming the above bilinear hysteretic model is computed as

$$k_{eff} = \frac{q_d}{u} + k_b \quad (10)$$

The effective period of a system of LR bearings is computed as



Code-Based Design: Seismic Isolation of Buildings, Fig. 7 Typical lead rubber bearing with cutaway showing internal construction (courtesy of Takenaka Corp., Japan)

$$T_b = 2\pi \sqrt{\frac{W}{g \sum_j k_{eff,j}}} \quad (11)$$

where $k_{eff,j}$ is the effective stiffness of the j th isolator. The equivalent viscous damping of a single isolator is given as

$$\zeta_{eq} = \frac{1}{4\pi} \frac{E_d}{E_s} \quad (12)$$

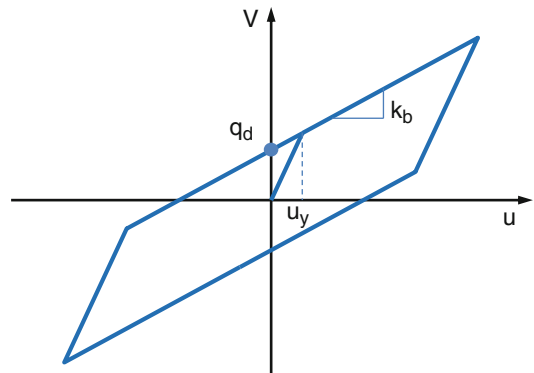
where the energy dissipated per cycle $E_d = 4q_d(u - u_y)$ and the elastic energy stored $E_s = \frac{1}{2}k_{eff}u^2$. Substituting these expressions into Eq. 12 gives an expression for ζ_{eq} that is a function of bearing properties q_d , u_y , k_b , and the displacement u :

$$\zeta_{eq} = \frac{2q_d(u - u_y)}{\pi u(q_d + k_b u)} \quad (13)$$

The equivalent damping of a system of LR bearings is computed as

$$\zeta_{eq} = \frac{\sum_j \zeta_{eq,j} k_{eff,j}}{\sum_j k_{eff,j}} \quad (14)$$

where $\zeta_{eq,j}$ is the effective stiffness of the j th isolator. In other words, the system equivalent



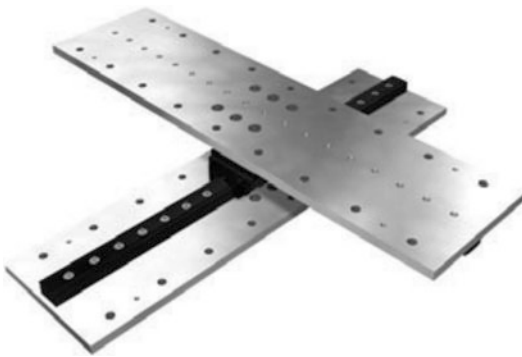
Code-Based Design: Seismic Isolation of Buildings, Fig. 8 Bilinear force–displacement relationship for LR bearings

damping is the weighted arithmetic mean of the individual damping values where the individual stiffnesses are the weighting factors.

Sliding Isolation Bearings

Flat Sliders

In cases where an elastomeric bearing supports low axial loads, it may be either impractical or unnecessarily costly to resist these axial loads using one of the bearings previously described. Instead, it is often preferable to install a flat sliding surface that allows the column to simply move with the isolation level diaphragm. These bearings typically have low friction coefficients and are inexpensive relative to the isolation bearings. One type of flat slider is the cross linear bearing, shown in Fig. 9, which provides the features of a typical flat slider with the addition of tension-resistant behavior. These have been implemented in high-rise buildings in Japan to resist the uplift forces below columns due to overturning effects.



Code-Based Design: Seismic Isolation of Buildings, Fig. 9 Cross linear bearing (courtesy of THK Co., Ltd.)

Single-Pendulum (SP) Bearings

The single-pendulum bearing is the original Friction Pendulum™ System described by Zayas et al. (1987) and represents the first manufactured sliding bearing to make use of the pendulum concept. This bearing consists of an articulated slider resting on a concave spherical surface. The slider is coated with a woven PTFE (polytetrafluoroethylene) composite liner, and the spherical surface is coated with a polished stainless steel overlay. A picture showing an SP bearing and a cross section is shown in Fig. 10, indicating the above-described components.

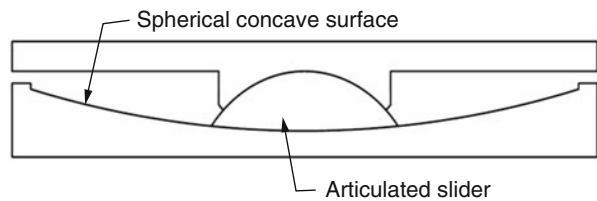
Although there is an abundance of published work on the cyclic behavior of SP bearings (e.g., Zayas et al. 1990 and Mosqueda et al. 2004, among others), it is useful to recapitulate the essential aspects of their behavior. To develop a mathematical model of the force–displacement relationship of an SP bearing, the geometry of the device must be fully understood.

First, we can write the equilibrium equations of the bearing in its displaced condition (as seen above in Fig. 11), summing forces along the horizontal and vertical axes and obtaining

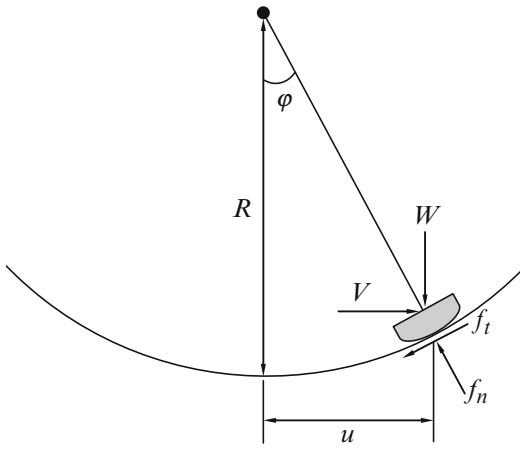
$$\begin{pmatrix} V \\ W \end{pmatrix} = \begin{bmatrix} \cos \varphi & \sin \varphi \\ -\sin \varphi & \cos \varphi \end{bmatrix} \begin{pmatrix} f_t \\ f_n \end{pmatrix} \quad (15)$$

From geometry, it is clear that $\sin \varphi = u/R$ and $\cos \varphi = \sqrt{R^2 - u^2}/R$; hence, Eq. 15 can be written as

$$\begin{pmatrix} V \\ W \end{pmatrix} = \frac{1}{R} \begin{bmatrix} \sqrt{R^2 - u^2} & u \\ -u & \sqrt{R^2 - u^2} \end{bmatrix} \begin{pmatrix} f_t \\ f_n \end{pmatrix} \quad (16)$$



Code-Based Design: Seismic Isolation of Buildings, Fig. 10 Photo (left) and section (right) of a typical SP bearing



Code-Based Design: Seismic Isolation of Buildings,
Fig. 11 Equilibrium of slider in displaced configuration

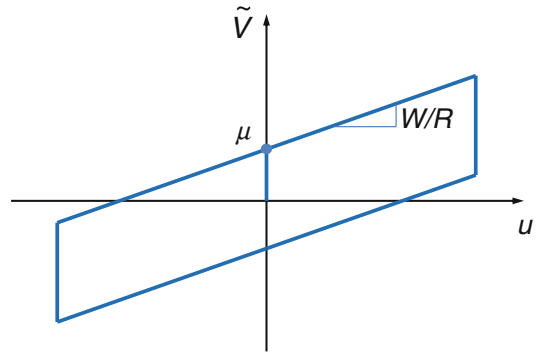
Assuming the tangential traction f_t is related to the normal traction f_n through Coulomb friction by the equation $f_t = \mu f_n$ (where μ is the coefficient of friction) and defining a normalized shear force $\tilde{V} = V/W$, Eq. 16 reduces to

$$\tilde{V} = \frac{\mu\sqrt{R^2 - u^2} + u}{\sqrt{R^2 - u^2} - \mu u} \quad (17)$$

Equation 17 returns a normalized shear force in the bearing given an imposed displacement u . While in some cases it may be important to consider the large displacement formulation of the normalized restoring force, a common simplifying assumption here is useful. For virtually all SP bearings, the effective radius of curvature R is much greater than the maximum expected displacement u . The consequence of this assumption can be seen by expanding $g(u) = \sqrt{R^2 - u^2}$ in a Taylor series about the point $u = 0$. Carrying out this expansion yields the series

$$\sqrt{R^2 - u^2} = R - \frac{1}{2}R\left(\frac{u}{R}\right)^2 - \frac{1}{8}R\left(\frac{u}{R}\right)^4 - \dots \quad (18)$$

From this series, it is clear that for $R \gg u$, $\sqrt{R^2 - u^2} \approx R$. Substituting this approximation



Code-Based Design: Seismic Isolation of Buildings,
Fig. 12 Idealized hysteresis loop of single-concave SP bearing based on Eq. 19

into Eq. 17 and simplifying, the actual restoring force is

$$V = \mu W \text{sgn}(\dot{u}) + \frac{W}{R} u \quad (19)$$

This is the well-known force-deformation relation for a single-concave SP bearing at a particular sliding displacement u and velocity $\dot{u}$. The function $\text{sgn}(\cdot)$ is the *signum* function and returns 1 (or -1) if the argument is positive (or negative). The inclusion of the *signum* function is necessary since the direction of the friction force always opposes that of the sliding velocity. A normalized force-deformation plot of an SP bearing under one complete cycle of loading, based on Eq. 19, is shown in Fig. 12.

An implication of Eq. 19 is that the frequency characteristics of a rigid structure isolated on SP bearings are independent of its mass. This can be seen by writing the equation of motion for an undamped single degree-of-freedom oscillator with an SP bearing as the restoring force (where m is the supported mass; hence, $W = mg$). Neglecting friction, the equation of motion and the resulting natural period is given by

$$m\ddot{u} + \frac{mg}{R}u = 0 \quad \Rightarrow \quad T_b = 2\pi\sqrt{\frac{R}{g}} \quad (20)$$

Most isolation devices are characterized by effective stiffness, k_{eff} , and equivalent viscous

damping ζ_{eq} . For SP bearings it is appropriate to normalize effective stiffness by the supported weight W , or

$$\tilde{k}_{eff} = \frac{\tilde{V}}{u} = \frac{\mu}{u} + \frac{1}{R} \quad (21)$$

Note that $\tilde{k}_{eff}$ has units of $[1/L]$, and the actual effective stiffness is simply $k_{eff} = \tilde{k}_{eff}W$. Similarly, the equivalent viscous damping is given as

$$\zeta_{eq} = \frac{1}{4\pi} \frac{E_D}{E_S} \quad (22)$$

where the energy dissipated per cycle $E_D = 4\mu Wu$ and the elastic energy stored $E_S = \frac{1}{2}k_{eff}u^2$. Substituting these expressions into Eq. 22 gives an expression for ζ_{eq} that is a function of bearing properties R , μ and the displacement u :

$$\zeta_{eq} = \frac{2R\mu}{\pi(R\mu + u)} \quad (23)$$

Note that the equivalent damping is independent of the weight supported by the SP bearing. Another way to express the equivalent damping is by recognizing that the friction shear force at slip is $V_f = \mu W$, and the shear force in the bearing is $V_b = k_{eff}u$. In terms of these physically meaningful forces, the equivalent damping is given as

$$\zeta_{eq} = \frac{2}{\pi} \frac{V_f}{V_b} \quad (24)$$

For an SP system with friction coefficient of 0.06 and a peak base shear coefficient of 0.25, the equivalent damping can easily be computed as

$$\zeta_{eq} = \frac{2}{\pi} \frac{0.06}{0.25} = 0.15 \quad (25)$$

Multi-pendulum Bearings

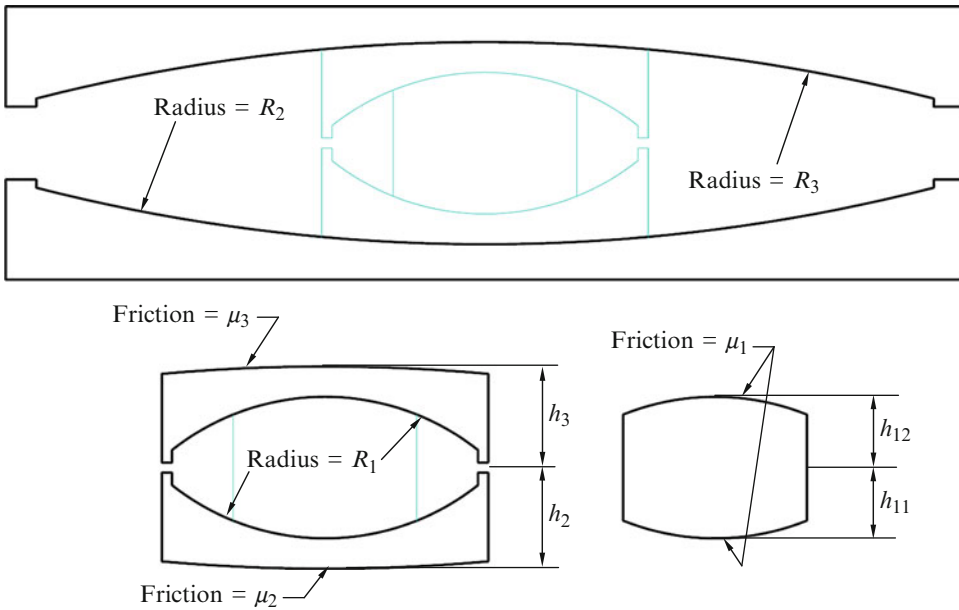
Recent developments in the design and manufacturing of FP bearings have centered around the use of multiple pendulum mechanisms. Whereas the single-concave FP bearing has two key parameters that characterize cyclic

behavior (R and μ), a multistage FP bearing has greater design flexibility because the pendulum lengths and friction coefficients are specific for each independent pendulum mechanism. The triple-pendulum (TP) bearing introduced by EPS, Inc. consists of four concave surfaces and three independent pendulum mechanisms. The outer slider consists of concave surfaces on either side of a cylindrical inner slider with a low-friction interface on both ends. This forms one pendulum mechanism and defines the properties of the isolation system under low levels of excitation. The outer slider also consists of sliding interfaces between top and bottom outer sliders and the major spherical surfaces of the bearing. The bottom sliding surface is in contact with a spherical surface of a particular radius of curvature, forming the second pendulum mechanism. This mechanism defines the primary properties of the isolation system under moderate levels of excitation. The upper sliding surface is in contact with another spherical surface of a particular radius of curvature, forming the third pendulum mechanism. The friction coefficient of this third sliding interface is sufficiently large to prevent sliding until an extreme level of excitation occurs. For this type of bearing, shown in Figs. 13 and 14, the parameters characterizing the cyclic behavior are (R_1, μ_1) for the inner concave sliding interfaces, (R_2, μ_2) for the bottom outer surface, and (R_3, μ_3) for the top outer surface. The slider heights define the kinematic relations for the positions of each sliding surface. The behavior of the triple-pendulum bearing has been described by Morgan (2007) and Fenz and Constantinou (2008a, b).

Given the geometry and properties of the TP bearing, the force–displacement relationship for the element can be derived by taking equilibrium in the deformed configuration. The TP bearing cannot, however, be treated simply as a series of three single-pendulum elements, as this would ignore the relative rotation of the inner slider relative to the outer concave surfaces. A complete derivation of the monotonic force–deformation relation for the TP bearings has been presented by Morgan (2007) and Fenz and Constantinou (2008a, b). This relation is depicted in Fig. 15,

**Code-Based Design:
Seismic Isolation of
Buildings,**

Fig. 13 Triple-pendulum
(TP) bearing showing
internal components
(courtesy of EPS, Inc.)



Code-Based Design: Seismic Isolation of Buildings, Fig. 14 Exploded cross section of triple-pendulum bearing showing geometric and friction parameters

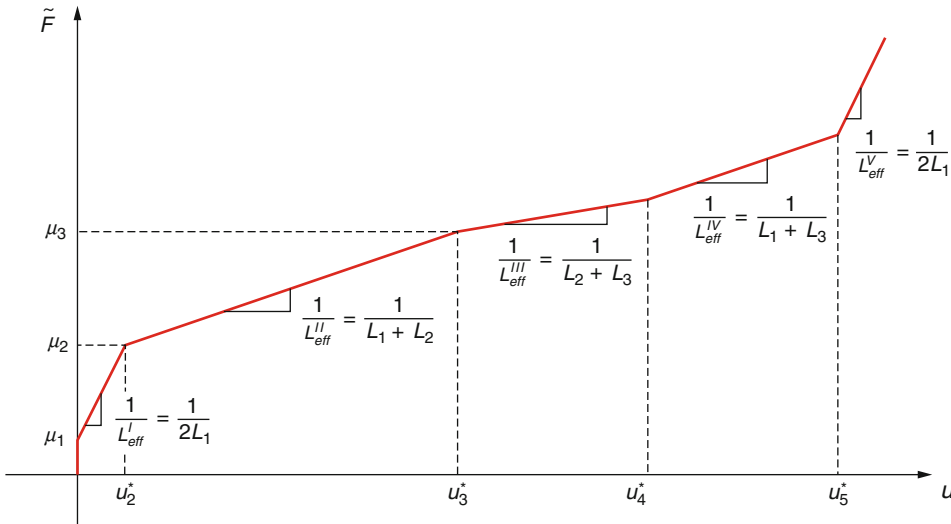
showing each of the five stages of unidirectional sliding. Here the effective pendulum lengths are defined as $L_j = R_j - h_j$ where R_j is the radius of the j th sliding surface and h_j is the height of the j th slider, as shown in Fig. 14.

The main feature of the TP bearing is that stiffness and damping properties evolve with increasing displacement. Moreover, this evolution of stiffness and damping can be controlled by varying the radii and friction coefficients of

each sliding surface. This results in a wide variety of realizable hysteretic relations, and proper optimization of the parameters results in a TP bearing that is “tuned” for a particular seismic hazard and performance objective (Fig. 16).

Energy Dissipation Devices

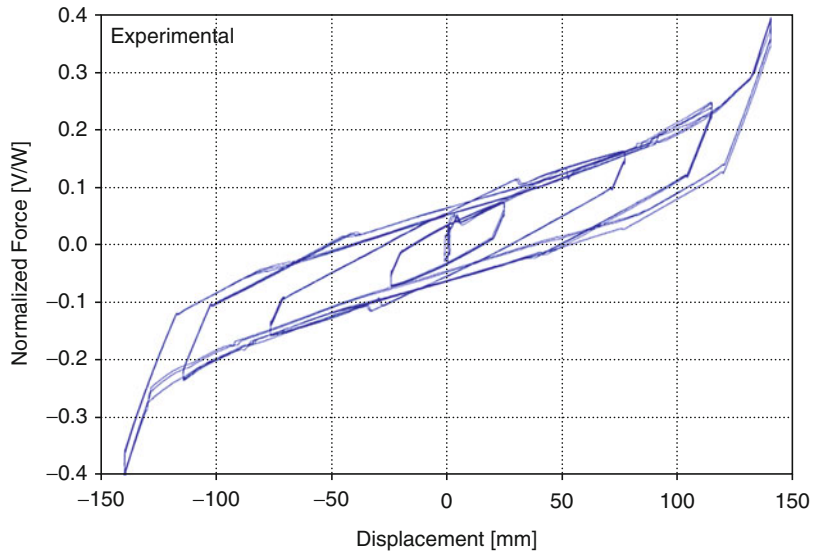
An important consideration in the design of isolated structures is the energy dissipation of the isolation system. While LR, HDR, and FP



Code-Based Design: Seismic Isolation of Buildings, Fig. 15 Monotonic force–deformation relation for triple-pendulum (TP) bearing (Morgan 2007)

**Code-Based Design:
Seismic Isolation of
Buildings,**

Fig. 16 Measured cyclic behavior for reduced-scale triple-pendulum (TP) bearing (Fenz and Constantinou 2008b)



bearings have internal damping mechanisms, the seismicity at a site may be significant and the expected deformation demands may be excessive given the stability requirements of the bearings. In such cases, supplemental energy dissipation devices may be specified to control the expected deformations. These devices can be categorized as either viscous dampers or hysteretic dampers. The details of each are discussed herein.

Fluid Viscous Dampers

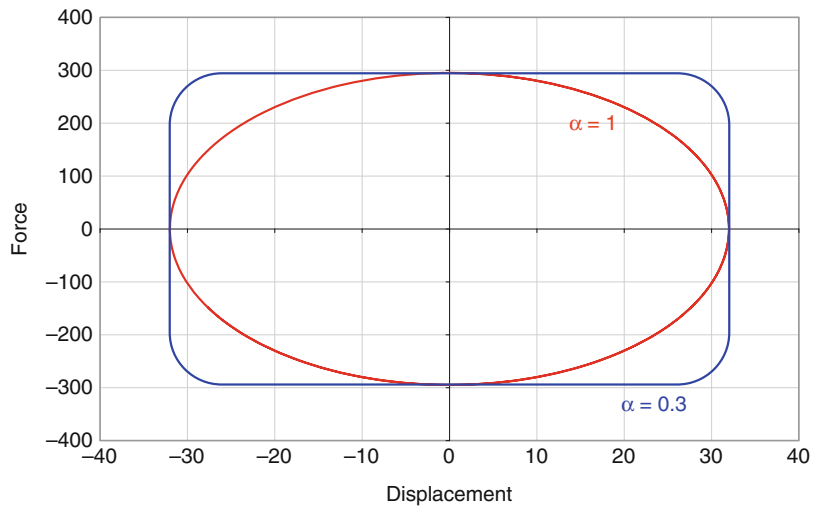
One class of supplementary energy dissipation devices is fluid viscous dampers (FVD). These dampers function as a dashpot, dissipating energy by forcing a viscous fluid through a set of orifices in a piston head, which results in heat generation. These dampers are precision devices, manufactured to achieve a target damping coefficient and velocity dependence.



Code-Based Design: Seismic Isolation of Buildings, Fig. 17 Fluid viscous damper at isolation plane

**Code-Based Design:
Seismic Isolation of
Buildings, Fig. 18**

Force–displacement
relationship for linear
viscous damper ($\alpha = 1$)
and nonlinear viscous
damper ($\alpha = 0.3$)



The force output F_d of a viscous damper is given by

$$F_d = c_d |\dot{u}|^\alpha \text{sgn}(\dot{u}) \quad (26)$$

where c_d is the damping coefficient and α is the velocity exponent. In the case than $\alpha = 1.0$, the damper is classified as a linear viscous damper. Otherwise, the damper is classified as nonlinear. Linear dampers are characterized by an elliptical force–displacement loop for harmonic motion, whereas nonlinear dampers have more rectangular loops with rounded corners (see Fig. 17). In both cases, the force output of the damper at maximum displacement (and hence zero velocity) is equal to zero. A major advantage of viscous

damper is the phase lag in force output of the isolation devices and the damping devices. Since isolation bearings are deformation dependent, the peak force output occurs at peak displacement, and zero force output occurs at zero displacement. This is in contrast to viscous dampers, in which the peak force output occurs at zero displacement (or peak velocity), and zero force output occurs at peak displacement (or zero velocity). For moderate levels of damping, isolator displacements are reduced without an increase in total building base shear (Fig. 18).

The design of isolation systems is based on the natural frequency of the isolation system, ω_b , and the equivalent viscous damping, ζ_b . For some



Code-Based Design: Seismic Isolation of Buildings, Fig. 19 Metallic yielding dampers for use at isolation plane

harmonic displacement with amplitude u_0 and forcing frequency $\bar{\omega}$, the energy dissipation for one cycle, E_d , is given by (Soong and Constantinou 1994)

$$E_d = \lambda(\alpha) C_d u_0^{1+\alpha} \bar{\omega}^\alpha \quad (27)$$

where $\lambda(\alpha)$ is defined in terms of the gamma function and is given by

$$\lambda(\alpha) = 4(2^\alpha) \frac{\Gamma^2(1 + \alpha/2)}{\Gamma(2 + \alpha)}, \quad (28)$$

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$$

It is useful to define a weight-normalized damping coefficient, $\bar{c}_d = c_d/mg$, where m is the total supported mass of the isolation system. From the equation for equivalent viscous damping, the equivalent damping ζ_{eq} for an isolation system with viscous dampers is given by

$$\zeta_{eq} = \frac{\lambda(\alpha)}{2\pi} \bar{c}_d u_0^{\alpha-1} \omega_b^{\alpha-2} g \quad (29)$$

Two special cases of viscous dampers are linear viscous ($\alpha = 1.0$) and pure friction ($\alpha = 0$). For linear viscous dampers, the equivalent damping is given by the simple and well-known expression

$$\zeta_{eq} = \frac{\bar{c}_d g}{2\omega_b} \quad (30)$$

For pure friction dampers, the equivalent damping is given by

$$\zeta_{eq} = \frac{2\bar{c}_d g}{\pi u_0 \omega_b^2} \quad (31)$$

For general nonlinear viscous dampers, a convenient and highly accurate approximation is given by $2\pi/\lambda(\alpha) \approx 0.4\alpha + 1.6$ (Morgan and Mahin 2011). This results in a simplified expression for the equivalent damping

$$\zeta_{eq} \approx \frac{\bar{c}_d u_0^{\alpha-1} \omega_b^{\alpha-2} g}{0.4\alpha + 1.6} \quad (32)$$

Hysteretic Dampers

Devices that dissipate energy through metallic yielding or friction are classified as hysteretic dampers. These dampers are characterized by an elastic stiffness k_e and a yield force f_y and are generally considered to behave elasto-perfectly plastic (much like the lead plug in the LR bearing). Two hysteretic dampers that have been implemented in Japan are shown in Fig. 19. Unlike viscous dampers, hysteretic dampers are displacement dependent, and the force output is in phase with the forces in the isolation system. As a result, the addition of hysteretic dampers, while reducing the peak displacement of the isolators, may introduce substantially higher forces in the superstructure and foundations.

The presence of hysteretic dampers increases both the effective stiffness and equivalent damping of the isolation system. Suppose an isolation system without dampers has effective stiffness k_{eff} and equivalent damping ζ_{eq} . Furthermore, a group of hysteretic dampers are to be added whose aggregate elastic stiffness and yield force are given by k_e and f_y , respectively. For these dampers subjected to a lateral displacement u , the energy dissipated per cycle $E'_d = 4f_y(u - u_y)$, where $u_y = f_y/k_e$. The effective stiffness of the dampers is $k'_{eff} = f_y/u$. From Eq. 2, the added equivalent damping is

$$\zeta'_{eq} = \frac{2}{\pi} \frac{(u - u_y)}{u} \quad (33)$$

Notably, for small yield displacement u_y , the equivalent damping is a constant $\zeta'_{eq} = 2/\pi \approx 0.64$ and is independent of the damper yield force f_y or isolator lateral displacement u .

The total effective stiffness of the combined isolator/damper system is

$$k_{eff,t} = k_{eff} + k'_{eff} \quad (34)$$

From Eq. 14, the total equivalent damping for the combined isolator/damper system is

$$\zeta_{eq,t} = \frac{\zeta_{eq}k_{eff} + \zeta'_{eq}k'_{eff}}{k_{eff} + k'_{eff}} \quad (35)$$

Design Provisions for Isolated Buildings

Throughout the world, there are a number of published design provisions devoted to the design of seismic isolated structures. The principal goal of these provisions is to ensure that isolated structures maintain an acceptable level of safety to the public in the event of an earthquake. The discussion herein focuses on ASCE 7–10, which contains provisions governing design of seismic isolated buildings in the USA.

The commentary to the 2003 National Earthquake Hazards Reduction Program (NEHRP) Provisions (BSSC 2003) states that isolated

structures designed to those provisions should be able to:

1. Resist minor and moderate levels of earthquake ground motion without damage to structural elements, nonstructural components, or building contents
2. Resist major levels of earthquake ground motion without failure of the isolation system, without significant damage to structural elements and nonstructural components, and without major disruption of the function of the facility

This is an implicit two-stage performance objective that seeks to minimize damage to the building for moderate levels of ground shaking while preventing major losses in the event of a design earthquake (DE) event. The commentary goes on to state that “the above performance objectives for isolated structures considerably exceed the performance anticipated for fixed-base structures during moderate and major earthquakes.” It is crucial to distinguish the performance expectations for conventional and isolated building designed to the same ASCE 7–10 provisions, as conventional buildings are only designed to limit the likelihood of collapse in the event of a maximum-considered earthquake (MCE).

Isolation systems which are deemed acceptable by the provisions of ASCE 7–10 have the following general characteristics:

1. Maintain stability when subjected to design displacements
2. Provide increasing lateral resistance with increasing displacement
3. Suffer no strength or stiffness degradation under repeated cyclic loading
4. Have quantifiable force–deflection and damping characteristics

Analysis Methods

The analysis of isolated buildings is generally conducted by one of three possible methods: equivalent lateral force (static) analysis, response spectrum analysis, and response history analysis.

Each analysis method is capable of representing the peak internal forces and deformations in both the superstructure and isolation system given a design level of seismic intensity and may be used to select and proportion both structural elements and isolation devices.

Equivalent Lateral Force Procedure

The simplest analysis procedure is the equivalent lateral force (ELF) procedure, which is analogous to the static design procedures for conventional structures. In this procedure, the 5 % damped design spectral acceleration for a site at a period of 1 s, S_{D1} , represents the design input.

$$D_D = \frac{gS_{D1}T_D}{4\pi^2B_D} \quad (36)$$

where T_D is the DE effective period, g is the gravitational constant, and B_D is a factor to account for the equivalent damping of the isolation system. While B_D is based on tabulated value, this can be approximated by the equation $B_D = 2.4\zeta_{eq}^{0.3}$ where ζ_{eq} is the equivalent damping computed for the isolation system. The DE effective period is computed as

$$T_D = 2\pi\sqrt{\frac{W}{k_{D\min}g}} \quad (37)$$

where W is the total supported weight of the building and $k_{D\min}$ is the lower-bound effective stiffness of the isolation system at a displacement D_D . The base shear of the superstructure is computed as

$$V_s = \frac{k_{D\max}D_D}{R_I} \quad (38)$$

where $k_{D\min}$ is the upper-bound effective stiffness of the isolation system at a displacement D_D and R_I is the response modification factor for the superstructure and is based on 3/8 times the value of R for a conventional structure with the same lateral system (not to exceed $R_I = 2.0$). This R_I factor is substantially smaller than the R factor for conventional structures because of the enhanced

performance objectives and the resulting need to maintain essentially elastic behavior in the structure. This base shear is also subjected to several lower-bound limits to prevent yielding of the building prior to activation of the isolation system and excessive movement under the design wind loads.

Once the base shear V_s has been computed, the lateral forces on the building are distributed vertically using the equation

$$F_x = V_s \frac{w_x h_x}{\sum_{i=1}^n w_i h_i} \quad (39)$$

where F_x is the lateral force at story x , w_i is the story weight at the i^{th} floor, and h_i is the height above the grade to the i^{th} floor.

The isolation system must be designed such that the elements of this system remain stable given an MCE event. Given a 5 % damped MCE spectral acceleration for a site at a period of 1 s of S_{M1} , the displacement of the isolators is computed as

$$D_M = \frac{gS_{M1}T_M}{4\pi^2B_M} \quad (40)$$

where T_M is the MCE effective period and B_M is similar to B_D in Eq. 36, computed at the displacement D_M . The MCE effective period is computed as

$$T_M = 2\pi\sqrt{\frac{W}{k_{M\min}g}} \quad (41)$$

where $k_{D\min}$ is the lower-bound effective stiffness of the isolation system at a displacement D_M . The isolation system must be designed for this displacement D_M plus the additional effects of both actual and accidental torsion.

The ELF procedure is permitted for analysis of isolated buildings provided the following restrictions are met:

1. Building is vertically regular.
2. MCE spectral acceleration $S_{M1} < 0.6$ g.
3. $T_M \leq 3.0$ sec.

4. Soil type A, B, C, or D.
5. Height not greater than four stories or 20 m (65 ft).
6. $T_D/T_s > 3.0$.
7. Positive incremental stiffness of isolation system.
8. $T_D > T_{0.2D}/3$.
9. Isolation system does not limit MCE displacement to less than the total maximum displacement.

Response Spectrum Procedure

Where the ELF procedure is not permitted, it is often permissible to use the response spectrum procedure for design. In this procedure internal forces and deformations are computed using response spectrum analysis, as described by Chopra (2007), among others. In this procedure, the natural frequencies and modes of vibrations of the isolated building are computed in the usual way, and demands are calculated through linear combination of these modes weighted by their spectral ordinates at each natural frequency. The results obtained from response spectrum analysis are not permitted to exceed those computed from the ELF procedure.

This procedure may be used unless the rate-dependence, load history-dependence, or deformation-hardening characteristics of the isolation system necessitate explicit consideration of their nonlinear and/or velocity-dependent force–deflection properties.

Response History Procedure

In any case, an isolated building may be analyzed using the response history procedure. In this procedure, the internal forces and deformations of the structure and isolation system are computed from an ensemble of ground motion acceleration records. Where seven or more records are considered in the analysis, the demands may be based on the ensemble average; otherwise, the demands must be based on the ensemble maximum. No less than three records may be used in the analysis.

The model of the building and isolation system is developed to incorporate any potential nonlinear behavior. Whereas the building is

generally designed to be essential, most isolation devices are characterized by some combination of nonlinear and velocity-sensitive behavior. These behaviors must be included in the response history model.

Limits of Dynamic Results

The demands computed from either the response spectrum or response history procedures may not exceed those from the ELF procedure by more than specified amounts. Figure 20 contains a reproduction of a table from the commentary to ASCE 7–10 that lists the limits imposed on dynamic procedures for a variety of force and deformation demands.

Testing Requirements

ASCE 7–10 requires that the stiffness and damping values for the isolation system used in analysis and design are based on physical tests of the isolation devices. Such testing can be divided into either prototype tests or production tests. Prototype tests are conducted on two identical units or each isolator type in order to establish adequate performance under the entire range of possible demands to which the isolators may be subjected. These tests are prescribed to characterize the behavior of each device under wind loads and both the DE and MCE levels of seismic intensity. These tests must consider maximum and minimum expected axial loads considering both gravity and seismic effects. The prototype test program prescribed by ASCE 7–10 is as follows:

1. Twenty fully reversed cycles of loading at a lateral force corresponding to the wind design force
2. Three fully reversed cycles of loading at each of the following increments of the total design displacement: $0.25D_D$, $0.5D_D$, $1.0D_D$, and $1.0D_M$
3. Three fully reversed cycles of loading at the total maximum displacement, $1.0D_{TM}$
4. $30 S_{D1}/S_{DS}B_D$, but not less than 10, fully reversed cycles of loading at 1.0 times the total design displacement D_{TD}

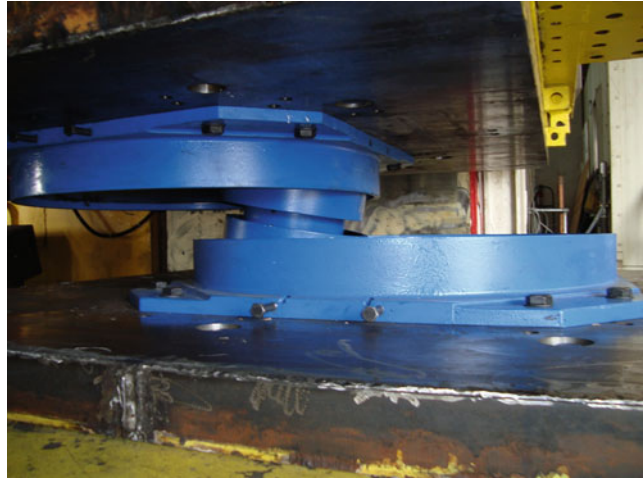
Table C13.2-2 Lower-Bound Limits on Dynamic Procedures Specified in Relation to ELF Procedure Requirements

Design Parameter	ELF Procedure	Dynamic Procedure	
		Response Spectrum	Response History
Design displacement – D_D	$D_D = (g/4\pi^2)(S_{D1}T_D/B_D)$	–	–
Total design displacement – D_T	$D_T \geq 1.1D$	$\geq 0.9D_T$	$\geq 0.9D_T$
Maximum displacement – D_M	$D_M = (g/4\pi^2)(S_{M1}T_M/B_M)$	–	–
Total maximum displacement – D_{TM}	$D_{TM} \geq 1.1D_M$	$\geq 0.8D_{TM}$	$\geq 0.8D_{TM}$
Design shear – V_b (at or below the isolation system)	$V_b = k_{Dmax}D_D$	$\geq 0.9V_b$	$\geq 0.9V_b$
Design shear – V_s ("regular" superstructure)	$V_s = k_{Dmax}D_D/R_I$	$\geq 0.8V_s$	$\geq 0.6V_s$
Design shear – V_s ("irregular" superstructure)	$V_s = k_{Dmax}D_D R_I$	$\geq 1.0V_s$	$\geq 0.8V_s$
Drift (calculated using R_I for C_d)	$0.015h_{sx}$	$0.015h_{sx}$	$0.020h_{sx}$

Code-Based Design: Seismic Isolation of Buildings, Fig. 20 Reproduction of Table C13.2–2 in commentary to ASCE 7–10 (ASCE 2010)

**Code-Based Design:
Seismic Isolation of
Buildings,**

Fig. 21 Prototype test of a triple-pendulum (TP) bearing, conducted by EPS, Inc.



Acceptance criteria for these tests are developed by the engineer of records and are generally focused on meeting a specified range of stiffness and damping values for the DE tests and maintaining stability in the MCE tests. Figure 21 shows a triple-pendulum (TP) bearing during prototype testing.

Production testing, while not explicitly prescribed by ASCE 7–10, is specified on most

projects. These tests are meant to ensure that fabricated isolators meet the established range of stiffness and damping properties prior to installation. A common production test program requires three fully reversed cycles at displacement amplitude D_D , considering the average gravity loads acting on the isolator type. These tests are often conducted on each isolator prior to installation but could be

limited to some sample subset at the discretion of the engineer and owner.

Design Review Requirements

ASCE 7–05 mandates a design review of the isolation system and related testing program by an independent panel composed of qualified individuals with experience in earthquake engineering principles and analysis and design of seismic isolation systems. This process typically includes a review of the following: the seismic hazard at the site and any ground motions that may be developed, the preliminary design of the isolation system, the observation of the prototype and production testing programs, and the final analysis and design of the entire structural system. The design review is often an interactive process where comments and suggestions are provided to the design team, who can then incorporate these in early schematic phases. This continuous interaction largely prevents significant divergence in the design approach adopted by the design team and the preferred approach advocated by the review panel. A healthy design review process is generally one that promotes a combination of sound engineering judgment and technically accurate analysis methods. As seismic isolated buildings are expected to achieve enhanced performance in a design earthquake, the review panel's principal role is to give the owner their independent assessment of the expected performance of the building and whether it meets the building code's intended performance objectives.

References

- American Society of Civil Engineers (ASCE) (2010) Minimum design loads for buildings and other structures. ASCE/SEI 7–10, ASCE, Reston
- BSSC (2003) FEMA 440: NEHRP recommended provisions for seismic regulations for new buildings and other structures. Building Seismic Safety Council, Washington, DC
- Buckle IG, Mayes RL (1990) Seismic isolation: history, application and performance – a world view. *Earthquake Spectra* 6(2):161–201
- Chopra AK (2007) Dynamics of structures: theory and application to earthquake engineering, 3rd edn. Pearson Prentice Hall, Englewood Cliffs
- Clark PW, Kelly JM (1994) Mechanical properties of BTR/Andre high-damping rubber seismic isolation bearings. EERC lab report 94–701, Earthquake Engineering Research Center, University of California, Berkeley
- Clark PW, Aiken ID, Nakashima M, Miyazaki M, Midorikawa M (2000) The 1995 Kobe (Hyogo-ken Nanbu) earthquake as a trigger for implementing new seismic design technologies in Japan. In: Lessons learned over time, vol 3, Learning from earthquake series. Earthquake Engineering Research Institute, Oakland, pp 79–109
- Constantinou MC, Whittaker AS, Kalpakidis Y, Fenz DM, Warn GP (2007) Performance of seismic isolation hardware under service and seismic loading. MCEER-07-0012, Multidisciplinary Center for Earthquake Engineering Research, State University of New York, Buffalo
- Fenz DM, Constantinou MC (2006) Behavior of double concave friction pendulum bearing. *Earthq Eng Struct Dyn* 35(11):1403–1424
- Fenz DM, Constantinou MC (2008a) Mechanical behavior of multi-spherical sliding bearings. MCEER-08-0007, Multidisciplinary Center for Earthquake Engineering Research, State University of New York, Buffalo
- Fenz DM, Constantinou MC (2008b) Spherical sliding isolation bearings with adaptive behavior: experimental verification. *Earthq Eng Struct Dyn* 37(2):185–205
- Kelly JM (1986) Aseismic base isolation: review and bibliography. *Soil Dyn Earthq Eng* 5(4):202–216
- Kelly JM (1993) Earthquake-resistant design with rubber. Springer, London
- Martelli A, Forni M, Clemente P (2012) Recent worldwide application of seismic isolation and energy dissipation and conditions for their correct use. In: Proceedings of 15th world conference on earthquake engineering, Lisbon
- Morgan TA (2007) The use of innovative base isolation systems to achieve complex seismic performance objectives. PhD dissertation, Department of Civil and Environmental Engineering, University of California, Berkeley
- Morgan TA, Mahin SA (2011) The use of innovative base isolation systems to achieve complex seismic performance objectives. PEER 2011/06, Pacific Earthquake Engineering Research Center, University of California, Berkeley
- Morgan TA, Whittaker AS, Thompson A (2001) Cyclic behavior of high-damping rubber bearings. In: Proceedings of 5th world congress on joints, bearings and seismic systems for concrete structures. American Concrete Institute, Rome
- Mosqueda G, Whittaker AS, Fenves GL (2004) Characterization and modeling of friction pendulum bearings subjected to multiple components of excitation. *J Struct Eng ASCE* 130(3):433–442

- Naeim F, Kelly JM (1999) Design of seismic isolated structures: from theory to practice. John Wiley, New York
- Skinner RI, Robinson WH, McVerry GH (1993) An introduction to seismic isolation. Wiley, Chichester
- Soong TT, Constantinou MC (1994) Passive and active structural vibration control in civil engineering. Springer-Verlag, Wien
- Yamamoto M, Minewaki S, Yoneda H, Higashino M (2012) Nonlinear behavior of high-damping rubber bearings under horizontal bidirectional loading: full-scale tests and analytical modeling. *Earthq Eng Struct Dyn* 41(13):1845–1860
- Zayas VA, Low SS, Mahin SA (1987) The FPS earthquake resisting system: experimental report. UCB/EERC-87/01, Earthquake Engineering Research Center, University of California, Berkeley
- Zayas VA, Low SS, Mahin SA (1990) A simple pendulum technique for achieving seismic isolation. *Earthquake Spectra* 6:317–333

Community Recovery Following Earthquake Disasters

Douglas Paton¹, Sarb Johal² and David Johnston^{2,3}

¹School of Psychology, University of Tasmania, Launceston, TAS, Australia

²Joint Centre for Disaster Research, GNS Science/Massey University, Wellington, New Zealand

³Risk and Society, GNS Science, Lower Hutt, New Zealand

Synonyms

Community leadership; Community processes; Community recovery; Resilience

Introduction

The effectiveness and efficiency of community recovery following a disaster is shaped not just by nature and scale of the physical impacts of the event but also on how the social environment supports the complex and protracted processes of recovery (Olshansky et al. 2006; Paton et al. 2014). Recent research following a number of disasters has highlighted the complex nature of community recovery and the

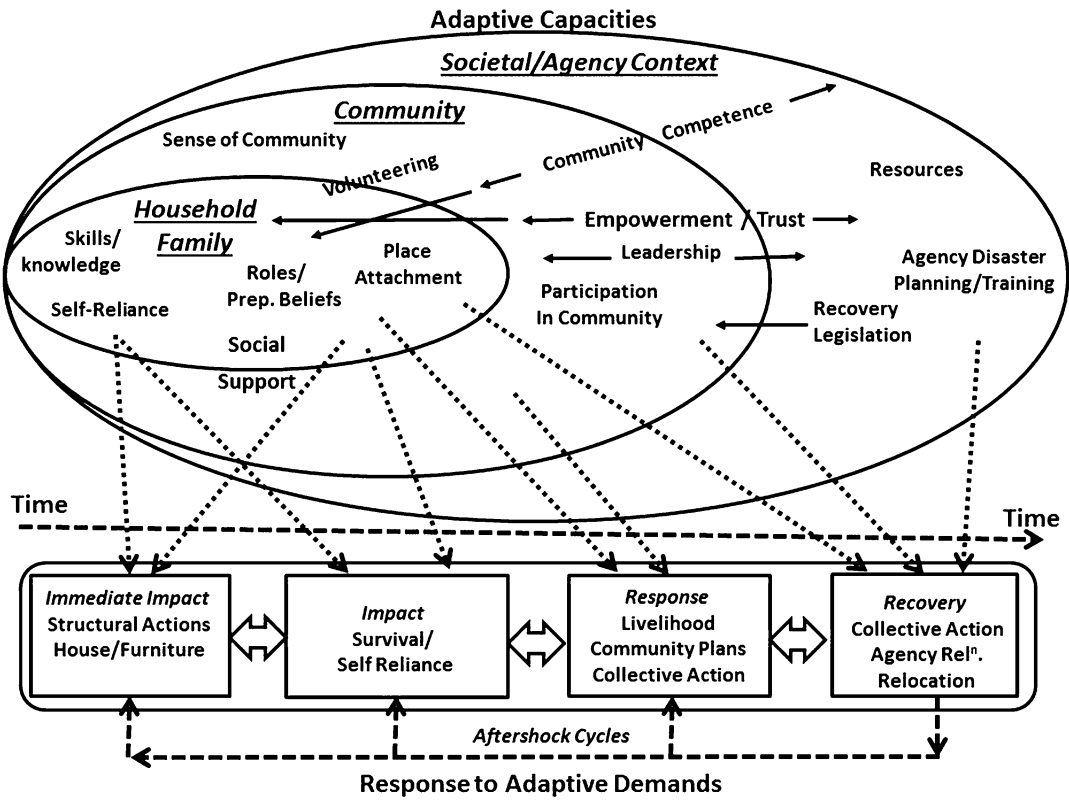
importance of not only strong local government capacity but also a cohesive system of public, private, and volunteer groups integrated into the community (Olshansky et al. 2012). Research consistently shows that community recovery from disasters can be greatly enhanced by ensuring that the preexisting social environment supports the recovery process.

Resilience in a Recovery Context

In a recovery context, a resilient community is one that can cope with, adapt to, and recover from the loss and disruption encountered through the experience of a disaster. How effectively this is done is a function of how well people, communities, and societies can work together and use their resources to deal with the problems encountered. This entry discusses recovery in relation to community and societal recovery. In the entry we draw largely on recent experiences following the 2011 Christchurch, New Zealand, earthquake from the perspective of people's accounts of what they had to contend with and how they responded to the disruption and challenges encountered during the recovery. We also draw on experiences gained from research in other cultures to illustrate the similarities and differences that culture introduces into how earthquake recovery can be understood and managed. The first issue addressed concerns what people had to contend with and how these demands changed over the course of recovery.

The Adaptive Demands of Earthquake Recovery

The Christchurch recovery experience illustrates several issues that need to be considered when conceptualizing community recovery following earthquakes. The first is that recovery is a process rather than any specific activity and one that extends over time as people deal with demands and challenges that differ depending on the stage of response and recovery they are at. This is depicted in Fig. 1 as the time arrow



Community Recovery Following Earthquake Disasters, Fig. 1 A summary of how adaptive capacities at person, family, household, and societal levels interact to influence earthquake recovery (Adapted from Paton et al. 2014)

illustrates how the relative salience of recovery issues and demands changes with time (Paton et al. 2014). For example, during the immediate impact phase (e.g., 3–14 days depending on nature and intensity of impact), the destruction of infrastructure and loss of essential services means that the demands people must cope with and adapt to relate people’s ability (“self-reliance”) to meeting survival needs on their own or with their neighbors. Over time, as essential services are progressively restored but while normal societal and economic activities remain affected, the challenges and issues people have to contend with change. As people move into the response phase, they find themselves working with neighbors and community members to deal with local issues (e.g., effecting basic repairs, looking after more vulnerable members of their community, etc.) and dealing with issues relating

to employment and livelihood. As the event moves into the recovery phase, people can increasingly find themselves dealing with demands from dealing with, for example, government agencies and insurance companies. The relative salience of these “adaptive demands” (i.e., what people have to adapt to and cope with) changes over time (Fig. 1). A second issue highlighted by the Christchurch experience was the fact that people may have transition through response and recovery cycles several times as they deal with the consequences of multiple aftershocks. This is illustrated in the aftershock feedback loop in Fig. 1. A third issue arises from the fact that differences were evident in how well people and communities dealt with these adaptive demands (Olshansky et al. 2006; Paton et al. 2014). This introduces a need to identify the sources of such variability in “adaptive capacity.”

Mitigation and Structural Preparedness

While recovery takes place in the aftermath of the hazard event, how well people, communities, and societies recover is a function of what they have to mitigate their risk and how well they have prepared to deal with those aspects of risk that cannot be readily anticipated (e.g., the degree of localized losses). Mitigation activities such as land use planning and building codes and the steps residents have taken to secure the structural integrity of their homes (Fig. 1) have an important bearing on the speed of earthquake recovery. With regard to community recovery, the importance of these pre-event activities reflects the fact that having a home that survives the initial impact reduces the risk of injury and death to its inhabitants. It also increases the likelihood of people having a habitable building in which to base their recovery activities and increases people's availability to offer assistance to neighbors and others and to contribute to economic recovery. However, people differ in the degree to which they take steps to safeguard their homes. This provides insights into the predictors of preparedness that can be defined as adaptive capacities (Paton and Johnston 2006). For example, the level of people's earthquake knowledge and family beliefs about the importance of structural preparedness offers insights in differences in community recovery capacity (Fig. 1). Thus, prevailing levels of structural and survival readiness provide the foundation for effective recovery when people face the emergent issues that arise.

Personal and Community Adaptive Capacity

The effectiveness of initial earthquake recovery is also influenced by people's capacity for self-reliance and their ability to deal with the immediate challenges of coping with earthquake consequences. The adaptive capacities that help explain differences in people's ability to do so are influenced by factors such as their earthquake knowledge, self-reliance, and family beliefs in preparedness (Fig. 1). Emergent recovery issues

include being able to develop self-help groups with neighbors and others to deal with local problems (e.g., removing rubble, setting up community meeting places, etc.). Subsequently, reestablishing community groups or forming new ones (e.g., organizing local efforts to repair homes, identifying and meeting local needs) affects the quality of recovery (Paton et al. 2014). Differences in how well these emergent demands are dealt with are influenced by family preparedness beliefs and factors such as levels of community competence and community leadership (Fig. 1). For example, confidence in self-help capability and the ability to engage with others under novel circumstances to confront issues is important (Paton et al. 2014). Developing a group identity, by increasing people's commitment to the recovery of the community and its members, which makes an additional and important contribution to recovery (Winstanely et al. 2011), and willingness to volunteer all contribute to explaining differences in levels of community earthquake recovery.

Key characteristics of community members (Mamula-Seadon et al. 2012; Norris et al. 2008; Paton and Johnston 2006) include willingness to be proactive despite the novelty in their circumstances, taking steps to include as many residents as possible, and empowering, encouraging, and supporting community initiatives. A need for new, community-specific ways of communication that address specific and diverse community needs and that are available when people need it was also identified as playing a significant role in expediting community recovery (Winstanely et al. 2011). A common characteristic of community groups is that the key roles within them are filled by volunteers.

Volunteering

Volunteering includes things ranging from assisting welfare services to organizing social events to working directly with agencies. Volunteering has secondary effects in that it helps people feel a sense of control in otherwise uncontrollable circumstances. This can be very

important in earthquake recovery as aftershock sequences can result in people cycling through response and recovery several times over what can be a prolonged period of recovery. Volunteering also increases opportunities for people to be actively involved in community life (cf. community participation) in meaningful ways (e.g., doing something of value for others) and helping rebuild the social fabric of the community. Volunteering not only provides community members with access to social support but also provides opportunities for people to work on issues of shared significance and develop problem-solving and planning skills. These activities result in the development and maintenance of constructs such as community participation and collective efficacy that have been recognized as playing key roles in community resilience and the development of social capital (e.g., Norris et al. 2008; Paton 2008). The quality of people's recovery is also influenced by how well community members and recovery agencies interact.

Community-Agency Adaptive Capacity

Given the atypical and threatening nature of the demands encountered, recovery is influenced by how well citizens and groups can represent their needs to agencies and can secure the resources each community needs to take responsibility for their unique recovery (Collins et al. 2011). These needs range from dealing with loss of utilities and dealing with residents with special needs to dealing with government agencies over the course of a prolonged period of recovery.

A pivotal predictor of recovery is proactive community leadership. Emergent community-based leaders help bring people together, facilitate a coherent community response, and link the community with external agencies and specialists to secure the resources and help required to meet local needs (Paton et al. 2014; Paton and Jang 2014). Thus, leadership is a crucial community resource and one that plays a critical role in empowering community recovery (Norris et al. 2008; Paton 2008). This is illustrated by the two-way arrow that links community and

society through community leadership. Community recovery may call for multiple leaders (e.g., managing community recovery versus managing specialist activities such as organizing repair work). Because those filling leadership roles will be volunteers, training and support for them is important.

The leadership, community processes, and volunteer activities discussed above represent resources that contribute to creating empowered communities. However, in complex disaster recovery environments, communities also need information and resources that cannot be met from within the community but only from the wider societal context: from NGOs, government agencies, and businesses. The quality of the relationships between communities and NGOs, government, and businesses that evolve in the recovery environment will thus influence the degree to which communities are empowered to enact their own recovery initiatives. This introduces a need to consider how relationships with NGOs, government, and businesses influence recovery.

The important role that relationships with government and businesses played in meeting community needs was endorsed in people's accounts of what helped and hindered community recovery in Christchurch (Mamula-Seadon et al. 2012; Paton et al. 2014). It is, however, important to appreciate that the relationship between community recovery and business and economic recovery is one characterized by interdependence (e.g., community members can be employees and consumers) and not just as one defined in terms of agencies and businesses being unidirectional providers of resources to communities (Paton and McClure 2013). Notwithstanding, the latter is important.

Communities can influence their relationship with government agencies and businesses by being proactive. This is illustrated by the two-way arrows in Fig. 1 that depict how community competency and leadership help link people and communities to societal agencies. In Christchurch, recovery was facilitated by community leaders taking active steps to represent community needs to government departments,

NGOs, and businesses rather than waiting for them to come to the community. Facilitation was increased also by agencies working through community leaders. However, when it comes to the ability and effectiveness of a proactive community-led approach to dealing with government agencies, the extent to which this relationship marginalizes or facilitates community recovery is affected by things such as the governance mechanisms setup as part of the formal recovery response (Bach 2012; Collins et al. 2011). Bureaucracies often do not adapt well to this circumstance, and new governmental organizations typically emerge to provide more resources. It is important to appreciate that the quality of earthquake recovery is a reciprocal process. That is, it will only arise when communities and government agencies, NGOs, and businesses play complementary roles in facilitating comprehensive recovery. Thus, while societal-level agencies play key roles in passing appropriate legislation and providing resources, these activities must be developed in ways that empower communities. This is depicted in Fig. 1 by the two-way arrow linked to empowerment and trust.

Earthquake Recovery and Community Development

Recovery is also about facilitating development. The concept of “Building Back Better” applies to the social and the physical environment. The physical issues requiring attention include land use planning, building codes, and architecture. Social rebuilding requires the inclusion of other strategies. For example, in Christchurch, a Psychosocial Recovery Advisory Group (PRAG) was developed to provide expert, timely, quality evidence-based advice on psychosocial issues (Mooney et al. 2011). The PRAG adopted a strengths-based approach for conceptualizing recovery and facilitating longer-term social development. The strengths-based approach mobilizes cultural, spiritual, and social resources with an emphasis on development through community participation and engagement. It is thus

an example of a strategy that contributes to creating empowered communities and providing a platform for social development.

Development will be further facilitated by developing a comprehensive Canterbury Recovery Strategy. Long-term recovery is facilitated by planning that covers, for example, built heritage, education, Maori, and community development in ways that build social and cultural capital.

The Christchurch experience also highlights caution in making recovery decisions when a community is under stress. There are challenges in getting communities to participate in complex decision-making in times of stress immediately after a disaster event, as well as delaying such decisions during the recovery period. Ensuring that communities are participating in decision-making processes prior to an event can help alleviate some of these problems. The importance of the latter is reinforced by the inclusion of a cross-cultural perspective on earthquake recovery.

Because earthquakes occur in many different countries, developing a comprehensive understanding of earthquake recovery makes it pertinent to consider the similarities and differences that the associated social and cultural diversity introduces into the recovery process. While this is an area requiring more research, it is possible to advance this understanding by comparing countries that differ in their respective cultural characteristics. For example, in countries with a relatively individualist cultural orientation (e.g., New Zealand), people act consistently across situations in accordance with a self-concept that is relatively independent of social situation and in which achieving personal goals is a prominent objective. In contrast, in collectivist countries like Taiwan, actions are underpinned by culturally embedded beliefs that are reflected in shared purpose and activity involving alignment with social norms, achieving collective goals, and engaging in activities related to future goals that emphasize social relations (Paton and Jang 2014). Comparing these countries can provide a foundation for development of a comprehensive understanding of earthquake recovery.

Based on some preliminary research comparing New Zealand (populations affected by the 2011 Christchurch earthquake) and Taiwan (those affected by the 1999 Chi Chi earthquake) (Paton and Jang 2014), people in Christchurch and Taiwan (Tung Shia) endorsed the fact that earthquake recovery was influenced by survival, household planning, neighborhood self-help, psychological, and community-agency factors. Irrespective of culture or location, people face the same challenges. However, both similarities and differences were evident in how people responded and recovered.

The work in Taiwan involved an ethnic minority group, the Hakka that account for some 20 % of the Taiwanese population. They were the group most significantly affected by the Chi Chi earthquake. Similarities between the social processes that influenced earthquake recovery were evident in both Taiwan and New Zealand. Both populations endorsed the need for self-reliance, active community involvement, leadership, and communities being proactive in managing their own recovery activities. The fact that government agencies could help or hinder community recovery was also identified as an important determinant of community recovery following a significant seismic event. However, some cultural differences in how this was achieved were evident when the two countries were compared.

In Taiwan, the nature of the social recovery was strongly influenced by the Hakka spirit, the essence of Hakka culture, which comprises characteristics such as frugality, diligence, self-reliance, shared responsibility, and persistence (Paton and Jang 2014). Residents highlighted the key role that Hakka spirit played in their recovery following the Chi Chi earthquake. For example, it was identified as generating a sustained sense of purpose, self-reliance, and perseverance (which creates a kind of culture-specific self-efficacy) throughout the several years of their recovery and facilitated people's efforts to rebuild both their homes and their communities.

Recovery was also facilitated by the influence of Hakka beliefs regarding the importance of accepting disasters as part of life experience,

accepting that the earthquake and its consequences had occurred, and the importance of moving on in ways that emphasized adopting a problem-focused coping approach to their adaptive recovery. Spiritual beliefs were also implicated. Spiritual beliefs helped people cope with challenges, be more optimistic, to focus on the positive aspects of their experience (e.g., how everyone pulled together to promote the recovery of the whole community), accept their losses, and adopt a problem-focused approach to their problem solving. Spirituality influenced residents' acceptance of earthquake consequences, enhanced their sense of community, and strengthened their social support networks.

The comparative New Zealand-Taiwan analyses illustrate how everyday or mainstream community and cultural competencies and characteristics (e.g., the community participation and collective efficacy variables reflect the outcomes of people's accumulated experience of community life) influence on people's ability to cope with, adapt to, and recover from earthquakes. However, the discussion also highlighted a need to accommodate cultural differences in the development and delivery of earthquake recovery interventions.

Summary

In conclusion, the quality of people's earthquake recovery is a function of their ability to deal with survival and planning issues (e.g., having water and being able to develop household and neighborhood plans to deal with the immediate impact of an earthquake), being able to confront psychological and livelihood/economic disruption, and being able to work with government agencies and NGOs. How well people and communities can deal with these demands and recover effectively is a function of the degree to which several personal and community competencies are present. These include self-efficacy, place attachment, active community participation, collective efficacy, empowerment, trust, community leadership, inclusivity, and sense of community.

Cross-References

- ▶ “Build Back Better” Principles for Reconstruction
- ▶ Building Disaster Resiliency Through Disaster Risk Management Master Planning
- ▶ Community Recovery Following Earthquake Disasters
- ▶ Earthquake Disaster Recovery: Leadership and Governance
- ▶ Economic Recovery Following Earthquakes Disasters
- ▶ Interim Housing Provision Following Earthquake Disaster
- ▶ Reconstruction Following Earthquake Disasters
- ▶ Resilience to Earthquake Disasters

References

- Bach RL (2012) Mobilising for resilience: from government to governance. *Tephra* 23:36–39
- Collins S, Glavovic B, Johal S, Johnston D (2011) Community engagement post-disaster: case studies of the 2006 Matata debris flow and 2010 Darfield Earthquake, New Zealand. *N Z J Psychol* 40:17–25
- Mamula-Seadon L, Selway K, Paton D (2012) Exploring resilience: learning from Christchurch communities. *Tephra* 23:5–7
- Mooney MF, Paton D, de Terte I, Johal S, Karanci AN, Gardner D, Collins S, Glavovic B, Huggins TJ, Johnston L, Chambers R, Johnston DM (2011) Psychosocial recovery from disasters: a framework informed by evidence. *N Z J Psychol* 40:26–39
- Norris FH, Stevens SP, Pfefferbaum B, Wyche KF, Pfefferbaum RL (2008) Community resilience as a metaphor, theory, set of capacities, and strategies for disaster readiness. *Am J Community Psychol* 41:127–150
- Olshansky RB, Johnson LA, Topping KC (2006) Rebuilding communities following disaster: lessons from Kobe and Los Angeles. *Built Environ* 32(4):354–374
- Olshansky RB, Hopkins LD, Johnson LA (2012) Disaster and recovery: processes compressed in time. *Nat Hazards Rev* 13:173–178
- Paton D (2008) Risk communication and natural hazard mitigation: how trust influences its effectiveness. *Int J Glob Environ Issues* 8:2–16
- Paton D, Jang L-J (2014) Earthquake readiness and recovery: an Asia-Pacific perspective. In: D’Amico S (ed) *Earthquakes and their impact on societies*. Springer, London
- Paton D, Johnston D (2006) Disaster resilience: an integrated approach. Charles C. Thomas, Springfield
- Paton D, McClure J (2013) *Preparing for disaster: building household and community capacity*. Charles C. Thomas, Springfield
- Paton D, Johnston D, Mamula-Seadon L, Kenney CM (2014) Recovery and development: perspectives from New Zealand and Australia. In: Kapucu N, Liou KT (eds) *Disaster & development: examining global issues and cases*. Springer, New York
- Winstanley A, Cronin K, Daly M (2011) Supporting communication around the Canterbury earthquakes and other risks. GNS science miscellaneous report 2011/37, 39 p

Complex Fractional Moments and Their Use in Earthquake Engineering

Mario Di Paola

Dipartimento di Ingegneria Civile, Ambientale, Aerospaziale, dei Materiali (DICAM), Università di Palermo, Palermo, Italy

Synonyms

Complex fractional moments; Earthquake ground motion; Fractional calculus; Gaussian zero-mean excitation; Mellin transform; Riesz fractional integrals; Stochastic analysis

Introduction

Fractional calculus is referred to derivatives and integrals of order $\beta \in \mathbb{R}^+$ or more generally to complex order $\gamma = \rho + i\eta$, $\rho \in \mathbb{R}^+$, $\eta \in \mathbb{R}^+$. There are many different definitions of fractional operators such as Riemann-Liouville, Riesz, Marchaud, Caputo, etc. (see, e.g., Podlubny 1999; Samko et al. 1993). The various definitions differ with each another by intervals of integration or are simply adaptations of the Riemann-Liouville ones. In any case all the fractional operators share some common points:

- All the fractional operators are convolution integrals with power law kernel.
- For all fractional operators, the rules of classical derivatives and integrals (*linearity, semi-group, Leibniz rule*, etc.) hold.

- In the Fourier and in Laplace domains, they exactly behave like the derivatives and integrals of integer order.

For all these reasons, it will be more appropriate to call such operators as *generalized derivatives* and *integrals*, but for historical reasons all the people continue to call them as *fractional* ones.

In the last half century, fractional calculus becomes more and more popular also among the communities of physicists and engineers. This is due to its versatility to describe dynamic properties of the systems in many fields such as electrochemistry, thermal engineering, electricity, stochastic dynamics, and mechanics and especially to describe viscoelasticity.

In this entry some relevant examples on the use of fractional calculus to earthquake ground motion modeled as a stationary normal colored noise are presented. Applications of fractional calculus for the description of mono- and multivariate earthquake accelerations and exact filter equation are obtained in an integral form involving Riesz fractional operator in zero. The latter are related to complex spectral moments by Mellin transform operator. Other relevant application in probability may be found in Cottone et al. (2010) and Di Paola and Pinnola (2012).

Some Basic Elements of Fractional Calculus

Hereinafter some few preliminary definitions of fractional calculus are introduced for clarity sake's as well as for introducing appropriate symbologies. Let us start with the definition of the Cauchy formula for multiple integrals

$$\begin{aligned} (I_{a^+}^n f)(t) &= \int_a^t dt_1 \int_a^{t_1} \dots \int_a^{t_{n-1}} f(\tau) d\tau \\ &= \frac{1}{(n-1)!} \int_a^t (t-\tau)^{n-1} f(\tau) d\tau; \quad n \in \mathbb{N} \end{aligned} \quad (1)$$

Equation 1 shows that the n -th primitive of $f(t)$ may be obtained as a convolution integral

with power law kernel. The fractional integral is the generalization of Eq. 1 to an arbitrary order β in the form

$$(I_{a^+}^\beta f)(t) = \frac{1}{\Gamma(\beta)} \int_a^t (t-\tau)^{\beta-1} f(\tau) d\tau; \quad \beta > 0 \quad (2)$$

where $\Gamma(\cdot)$ is the Euler-Gamma function that interpolates the factorials ($\Gamma(n-1) = n!$). $(I_{a^+}^\beta f)(t)$ is referred as *Riemann-Liouville (RL) fractional integral*. The suffix a^+ stands for the initial time at which the operator applies and the apex $+$ means that t in Eq. 2 is greater than a . For this reason $(I_{a^+}^\beta f)(t)$ is called *left RL fractional integral*. The value a may be also $-\infty$ and in this case in literature is written $(I_{-\infty^+}^\beta f)(t)$ or simply $(I_{+}^\beta f)(t)$. The *right RL fractional integral* still exists and the integration is performed in $t \div a$, $a > t$. The *right RL* is usually written as $(I_{a^-}^\beta f)(t)$, and if $a \rightarrow \infty$ then is written as $(I_{-\infty^-}^\beta f)(t)$ or simply $(I_{-}^\beta f)(t)$.

As the derivative of order n may be viewed as the derivative of order $m+n$ of the m -th primitive, the Riemann-Liouville fractional derivative is defined as

$$\begin{aligned} (D_{a^+}^\beta f)(t) &= \frac{1}{\Gamma(n-\beta)} \frac{d^n}{dt^n} \int_a^t (t-\tau)^{n-\beta-1} f(\tau) d\tau; \\ \beta > 0 \end{aligned} \quad (3)$$

where $n = [\beta] + 1 \in \mathbb{N}$ and $[\beta]$ means *integer part* of β . The definition of left and right RL fractional derivative is like for the RL fractional integrals, then we denote $(D_{a^+}^\beta f)(t)$ if $t < a$, or if $a \rightarrow \infty$, we denote $(D_{-\infty^+}^\beta f)(t)$ or simply $(D_{+}^\beta f)(t)$ and similarly for the right RL derivative. Other notations are used by different authors, and here the notation of Samko et al. (1993) has been adopted. RL fractional integrals and derivatives remain meaningful also for complex order $\gamma = \rho + i\eta$.

In the following the fractional operators are always defined for complex order. All the rules of classical derivatives and integrals of integer order

still apply also for fractional order operators including semigroup properties, integration by parts, etc. Moreover, in Fourier and Laplace domains, RL fractional operators exactly behave like the classical derivatives and integrals. Such an example, by denoting the Fourier transform operator of $f(t)$ as $F_{\mathcal{F}}(\omega)$ defined as

$$\mathcal{F}\{f(t); \omega\} = \int_{-\infty}^{\infty} \exp(i\omega t) f(t) dt = F_{\mathcal{F}}(\omega) \quad (4)$$

we have

$$\begin{aligned} \mathcal{F}\left\{\frac{d^n f(t)}{dt^n}; \omega\right\} &= (-i\omega)^n F_{\mathcal{F}}(\omega); \\ \mathcal{F}\left\{(D_{0+}^{\gamma} f)(t); \omega\right\} &= (-i\omega)^{\gamma} F_{\mathcal{F}}(\omega) \end{aligned} \quad (5a)$$

$$\begin{aligned} \mathcal{F}\{I^n f(t); \omega\} &= (-i\omega)^{-n} F_{\mathcal{F}}(\omega); \\ \mathcal{F}\left\{(I_{0+}^{\gamma} f)(t); \omega\right\} &= (-i\omega)^{-\gamma} F_{\mathcal{F}}(\omega) \end{aligned} \quad (5b)$$

Similar behavior may be readily found in Laplace domain (see, e.g., Podlubny 1999).

In order to derive the filter equations, we need of other two relevant definitions: the *Riesz fractional operator* and the *Mellin transform*.

The RL fractional operators are related to the evolution of $f(t)$ into an assigned interval $(a \div t)$. In some cases, like in the probability domain or in power spectral density domain, the functions are defined in the whole range $-\infty \div \infty$. In these cases it is important to extend the RL fractional operators to unbounded domain. The extensions of the RL operator in unbounded domain are the *Riesz fractional integrals and derivatives*. These operators are a combination of left and right RL operators. Riesz fractional integrals and derivatives are in the following denoted as $(I^{\gamma} f)(t)$ and $(D^{\gamma} f)(t)$, respectively, and are defined as

$$\begin{aligned} (I^{\gamma} f)(t) &= \frac{1}{2v_c(\gamma)} \int_{-\infty}^{\infty} |t - \tau|^{\gamma-1} f(\tau) d\tau \\ &= \frac{\Gamma(\gamma)}{2v_c(\gamma)} [(I_{+}^{\gamma} f)(t) + (I_{-}^{\gamma} f)(t)] \end{aligned} \quad (6)$$

$$\begin{aligned} (D^{\gamma} f)(t) &= \frac{1}{2v_c(1-\gamma)} \int_{-\infty}^{\infty} |\tau|^{-\gamma-1} (f(t-\tau) - f(t)) d\tau \\ &= \frac{\Gamma(\gamma)}{2v_c(\gamma)} [(D_{+}^{\gamma} f)(t) + (D_{-}^{\gamma} f)(t)] \end{aligned} \quad (7)$$

where $v_c(\gamma) = \Gamma(\gamma) \cos(\gamma\pi/2)$, $\gamma = \rho + i\eta$, $\rho > 0$, $\rho \neq 1, 3, \dots$.

Fourier transform of the Riesz operator for $0 < \rho < 1$ is given as (see Samko et al. 1993, p. 217)

$$\mathcal{F}\{(I^{\gamma} f)(t); \omega\} = |\omega|^{-\gamma} F_{\mathcal{F}}(\omega) \quad (8a)$$

$$\mathcal{F}\{(D^{\gamma} f)(t); \omega\} = |\omega|^{\gamma} F_{\mathcal{F}}(\omega) \quad (8b)$$

From Eqs. 8a and 8b it may be stated that, for Fourier transformable functions, $(I^{-\gamma} f)(t) = (D^{\gamma} f)(t)$. This property is valid only for $0 < \rho < 1$ and Fourier transformable functions.

Lastly the Mellin transform operator is defined as

$$\begin{aligned} \mathcal{M}\{f(t); \gamma\} &= \int_0^{\infty} f(t) t^{\gamma-1} dt = F_{\mu}(\gamma); \\ \gamma &= \rho + i\eta \end{aligned} \quad (9)$$

Inverse Mellin transform returns $f(t)$ in the form

$$\begin{aligned} f(t) &= \mathcal{M}^{-1}\{F_{\mu}(\gamma); t\} \\ &= \frac{1}{2\pi i} \int_{\rho-i\infty}^{\rho+i\infty} F_{\mu}(\gamma) t^{-\gamma} d\gamma; \quad t > 0 \end{aligned} \quad (10)$$

From Eq. 9 it may be observed that integration is performed along the imaginary axis while ρ remains fixed. This observation has important consequences on how it will be better explained later on. Moreover, it may be observed that $F_{\mu}(\gamma)$ is a complex function having even real part and odd imaginary part with respect to η , that is, $F_{\mu}(\rho + i\eta) = F_{\mu}^*(\rho - i\eta)$, where the star means complex conjugate.

Condition of existence of Mellin transform and its inverse is $-\rho < \rho < -q$, where p and

q refer to the asymptotic behavior of $f(t)$ at $t \rightarrow 0$ and $t \rightarrow \infty$, respectively:

$$\lim_{t \rightarrow 0} x(t) = \mathcal{O}(t^p); \quad \lim_{t \rightarrow \infty} x(t) = \mathcal{O}(t^q) \quad (11)$$

where $\mathcal{O}(\cdot)$ means *order* of the term in parenthesis. Such an example if $f(t) = 1/(1+t)$, $\lim_{t \rightarrow 0} f(t) = 1$, then $p = 0$, $\lim_{t \rightarrow 0} f(t) = t^{-1}$ then $q = -1$. It follows that $f(t) = 1/(1+t)$ is Mellin transformable as the real part of γ is selected in the range $0 < \rho < 1$. The strip $-p \div -q$ is called *fundamental strip* (FS) of the Mellin transform. In discretized form Eq. 10 may be also rewritten as

$$\begin{aligned} f(t) &\simeq \frac{\Delta\eta}{2\pi} \sum_{k=-m}^m F_{\mu}(\gamma_k) t^{-\gamma_k} \\ &= \frac{1}{2b} t^{-\rho} \sum_{k=-m}^m F_{\mu}(\gamma_k) \left[\cos\left(k \frac{\pi}{b} \xi\right) - i \sin\left(k \frac{\pi}{b} \xi\right) \right] \end{aligned} \quad (12)$$

where $\gamma_k = \rho + ik\Delta\eta$, $b = \pi/\Delta\eta$, and $\xi = \ln t$. The number m of the summation in Eq. 12 gives a cutoff value along the imaginary axis and is selected in such a way that terms of order $m+n$ do not produce sensible variations on $f(t)$. It is to be stressed that Eq. 12 is independent of the value of ρ selected, provided it belongs to the FS of the Mellin transform. This is due to the fact that the Mellin transform $F_{\mu}(\gamma)$ is holomorph into the FS.

From Eq. 12 some considerations may be withdrawn:

- Inverse Mellin transform returns $f(t)t^{\rho}$ in the domain ξ and looks as a Fourier series, and $F_{\mu}(\gamma_k)/2b$ are the (complex) coefficients of the series.
- In the range $e^{-b} \div e^b$, the trigonometric functions $\exp(-ik \frac{\pi}{b} \ln t)$ are orthogonal, and as $m \rightarrow \infty$ they form a basis in such an interval. It follows that the orthogonality condition in ξ domain is in $-b \div b$.
- The factor $t^{-\rho}$ in Eq. 12 ensures us that $f(t)$ tends to zero as $t \rightarrow \infty$, while in zero Eq. 12, unless $f(0) = 0$, gives an infinite value due to the factor $t^{-\rho}$. Apart from this pathological behavior in $t = 0$, all the function $f(t)$ is

overlapped by summation (12) including the trend at $t = \infty$.

For more informations on fractional calculus, integration of fractional differential equation, readers are referred to the books of Miller and Ross (1993), Samko et al. (1993), Gorenflo and Mainardi (1997), and Podlubny (1999).

Power Spectral Density, Correlations, and Spectral Moments

The simplest model of earthquake acceleration $\ddot{Z}_g(t)$ is that it is represented as a stochastic stationary zero mean colored normal process. In this case the power spectral density (PSD) function, or its inverse Fourier transform, namely, the autocorrelation function, fully describes the process $\ddot{Z}_g(t)$ in probabilistic setting. In the following, for simplicity of notation, $\ddot{Z}_g(t)$ will be indicated as $Y(t)$. More realistic models may be found in literature giving a variety of nonstationary models. Among them the so-called *uniformly* modulated (Shinozuka and Sato 1967) have been proposed. These processes are constructed as the product of a stationary process, with assigned PSD, and a deterministic modulating function. Modulation in frequency and time has been proposed by Solomos and Spanos (1984), Shinozuka and Deodatis (1988), and Conte and Peng (1997).

For simplicity's sake, in this entry, we confine ourselves to the stationary model. In order to define the stationary seismic motion, two strategies are commonly assumed:

- The seismic ground acceleration is thought as a filtered normal white noise.
- The PSD is modeled in such a way that the corresponding response spectrum of accelerograms generated by the PSD matches the target one.

In the former case, the equation of filter, usually a system of differential equation of integer order, is already known, but the selection of the various parameters to model the effective

behavior of the ground deposit requires a level of knowledge that is seldom available.

On the other hand, the PSD consistent with the target response spectrum gives a shape of PSD (Falsone and Neri 1999; Cacciola et al. 2004) in frequency domain, whose parameters of the filter equation may not be easily obtained by using classical differential equations. In this regard discrete autoregressive (AR) or moving-average (MA) or their combination ARMA models have been proposed in literature (Spanos 1983, 1986) to generate sample functions with assigned PSD or correlation function.

In section “Filter Equation,” the issue of the definition of the filter equations for any shape of PSD will be addressed.

Fractional Spectral Moments

In the analysis of stochastic processes, an important role is played by the so-called spectral moments (SM), introduced by Vanmarcke (1972), important for the definition of some characteristics of stochastic processes and in reliability analysis (Michaelov et al. (1999)). These quantities are defined as the moments of the one-sided PSD with respect to the frequency origin. Let $S_Y(\omega)$ be the PSD of the earthquake acceleration with $S_Y(\omega) = S_Y(-\omega)$. Let $G_Y(\omega)$ be the *one-sided* PSD defined

$$G_Y(\omega) = 2U(\omega)S_Y(\omega) \quad (13)$$

being $U(\omega)$ the unit step function.

The SM are the moments of $G_Y(\omega)$ with respect to the frequency origin that is

$$\begin{aligned} \lambda_Y^j &= \int_{-\infty}^{\infty} 2U(\omega)\omega^j S_Y(\omega) d\omega \\ &= \int_0^{\infty} G_Y(\omega)\omega^j d\omega \end{aligned} \quad (14)$$

As a particular case, λ_Y^0 and λ_Y^2 represent the variance of the process $Y(t)$ and of $\dot{Y}(t)$, respectively. More informations on spectral moments may be found in Di Paola and Petrucci (1990), Muscolino (1991), and Spanos and Miller (1994). The spectral moments λ_Y^j are not able to describe

neither the PSD nor the correlation of the process $Y(t)$ itself because they may be divergent quantities starting from an assigned order. As an example for the classical Tajimi-Kanai filter, the spectral moments diverge for $j \geq 4$.

An extension of the definition given in Eq. 14 is obtained in the form $\lambda_Y^{\gamma-1}(\gamma \in \mathbb{C})$ in the following denoted as $\Lambda_Y(\gamma - 1)$ defined as

$$\Lambda_Y(\gamma - 1) = \int_0^{\infty} G_Y(\omega)\omega^{\gamma-1} d\omega \quad \gamma = \rho + i\eta \quad (15)$$

that is a complex function representing the Mellin transform of the one-sided PSD $G_Y(\omega)$. The function $\Lambda_Y(\gamma - 1)$ will be called *fractional spectral moment* (FSM) function. On the other hand, the inverse Mellin transform returns $G_Y(\omega)$ and $S_Y(\omega)$ in the form

$$\begin{aligned} G_Y(\omega) &= \frac{1}{2\pi i} \int_{\rho-i\infty}^{\rho+i\infty} \Lambda_Y(\gamma - 1)\omega^{-\gamma} d\gamma \\ &\simeq \frac{1}{2b} \sum_{k=-m}^m \Lambda_Y(\gamma_k - 1)\omega^{-\gamma_k}; \quad \omega \\ &> 0 \end{aligned} \quad (16a)$$

$$S_Y(\omega) \simeq \frac{1}{4b} \sum_{k=-m}^m \Lambda_Y(\gamma_k - 1)|\omega|^{-\gamma_k} \quad (16b)$$

where $\gamma_k = \rho + ik + \Delta\eta$, $b = \pi/\Delta\eta$, and $m\Delta\eta = \bar{\eta}$ is a cutoff value along the imaginary axis appropriately selected.

Equations 16a and 16b remain valid provided ρ belongs to the FS of the Mellin transform of $G_Y(\omega)$. Such an example for the Tajimi-Kanai (1960) PSD, since $\lim_{\omega \rightarrow 0} G_Y(\omega) = 4\pi S_0(\mathcal{O}(\omega^0))$, being S_0 the (finite) PSD of the white noise, and $\lim_{\omega \rightarrow \infty} G_Y(\omega) = \omega^{-2}(\mathcal{O}(\omega^{-2}))$, then the FS for the Tajimi-Kanai filter is $0 < \rho < 2$. Since $\Lambda_Y(\gamma_k - 1)$ never diverge provided ρ belongs to the FS of $G_Y(\omega)$, the whole $G_Y(\omega)$ is restored with the expansion in Eq. 16 with exception of the value in zero. It is to be stressed that truncation of the series up to the first m terms does not produce the divergence at $\omega \rightarrow \infty$.

Now we define the correlation function of the ground acceleration as $R_Y(\tau) = E[Y(t)Y(t+\tau)]$, where $E[\cdot]$ denotes ensemble average. By making the Fourier transform of the Riesz integral, according to Eq. 8a, we obtain

$$\mathcal{F}\{(I^{1-\gamma}R_Y)(t), \omega\} = |\omega|^{\gamma-1}S_Y(\omega); \quad 0 < \rho < 1 \quad (17)$$

It follows that $(I^{1-\gamma}R_Y)(t)$ is the inverse Fourier transform of $|\omega|^{\gamma-1}S_Y(\omega)$, that is,

$$(I^{1-\gamma}R_Y)(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega t} |\omega|^{\gamma-1} S_Y(\omega) d\omega \quad (18)$$

Since $S_Y(\omega)$ is symmetric, by putting $t = 0$ in Eq. 18, we get

$$\begin{aligned} (I^{1-\gamma}R_Y)(0) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |\omega|^{\gamma-1} S_Y(\omega) d\omega \\ &= \frac{1}{2\pi} \int_0^{\infty} |\omega|^{\gamma-1} G_Y(\omega) d\omega = \frac{1}{2\pi} \Lambda_Y(\gamma-1) \end{aligned} \quad (19)$$

As a conclusion $2\pi(I^{1-\gamma}R_Y)(0)$ is the Mellin transform of the one-sided PSD, namely, the FSM of the one-sided PSD $G_Y(\omega)$. From Eq. 19 it may be recognized that the FSM may be evaluated as the moment of the one-sided PSD or by evaluating the Riesz fractional integrals in zero of the correlation function. It is to be remarked that if $\Lambda(\gamma-1)$ is evaluated for a certain value ρ , the FSM for any other value of $\rho - \alpha$ may be easily evaluated provided $\rho - \alpha$ belongs to the FS of Mellin transform (see “Appendix”).

By inserting Eq. 19 in 16, the one-sided PSD may also be represented as follows

$$G_Y(\omega) \simeq \frac{\pi}{b} \sum_{k=-m}^m (D^{\gamma_k-1}R_Y)(0) \omega^{-\gamma_k}; \quad (20a)$$

$$\gamma_k = \rho + ik\Delta\eta; \quad \omega > 0$$

$$S_Y(\omega) \simeq \frac{\pi}{2b} \sum_{k=-m}^m (D^{\gamma_k-1}R_Y)(0) |\omega|^{-\gamma_k} \quad (20b)$$

The representation of $G_Y(\omega)$ in the form of Eq. 20a reveals that $G_Y(\omega)$ is expressed in a

generalized Taylor series. The difference with the classical Taylor series is that in the latter series the coefficients are the derivatives of the given function in zero and in Eqs. 20a and 20b are the fractional derivative in zero. Since in the FS the various FSMs never diverge and $|\omega|^{-\gamma_k}$ remains finite (provided ρ belongs to the FS), the whole PSD is restituted by Eq. 16b or 20b.

Fractional Spectral Moments of the Correlation function

It is to be emphasized that by using integer order derivatives of the correlation function, we get

$$\mathcal{F}\left\{\frac{d^j R_Y(\tau)}{d\tau^j}; \omega\right\} = (-i\omega)^j S_Y(\omega) \quad (21)$$

and its inverse Fourier transform returns $d^j R_Y(\tau)/d\tau^j$ in the form

$$\left[\frac{d^j R_Y(\tau)}{d\tau^j}\right] = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\omega\tau} (-i\omega)^j S_Y(\omega) d\omega \quad (22)$$

By evaluating $d^j R_Y(\tau)/d\tau^j$ in $\tau = 0$, we get $[d^j R_Y(\tau)/d\tau^j]_{\tau=0} = (-i)^j \lambda_Y^j / \pi (j = 0, 2, \dots)$ and is zero for $j = 1, 3, 5, \dots$. It follows that the classical Taylor series of the correlation is given as

$$R_Y(\tau) = \frac{1}{\pi} \sum_{j=0,2}^{\infty} \frac{(-i)^j}{j!} \lambda_Y^j \tau^j \quad (23)$$

We have already observed that the classical spectral moments may be divergent quantities, and then the classical Taylor series (23) becomes meaningless. On the other hand, also if the spectral moments do not diverge, truncation of the series (23) always produces the undesired phenomenon, that is, $\lim_{\tau \rightarrow \infty} R_Y(\tau) = \infty$.

Starting from the correlation function, we may define the one-sided correlation

$$C_Y(\tau) = 2U(\tau)R_Y(\tau) \quad (24)$$

and then the Mellin transform of $C_Y(\tau)$ is defined as

$$\mathcal{M}\{C_Y(\tau); \bar{\gamma}\} = \int_0^\infty \tau^{\bar{\gamma}-1} C_Y(\tau) d\tau = \Theta_Y(\bar{\gamma} - 1) \quad (25)$$

that are the spectral moments of the correlation function. Inverse Mellin transform of Eq. 25 returns $C_Y(\tau)$ in the form

$$C_Y(\tau) = \frac{1}{2\pi i} \int_{\bar{\rho}-i\infty}^{\bar{\rho}+i\infty} \Theta_Y(\bar{\gamma} - 1) \tau^{-\bar{\gamma}} d\bar{\gamma} \\ \cong \frac{\Delta\eta}{2\pi} \sum_{k=-m}^m \Theta_Y(\bar{\gamma}_k - 1) \tau^{-\bar{\gamma}_k}; \quad \tau > 0 \quad (26)$$

and as usual $\bar{\gamma}_k = \bar{\rho} + ik\Delta\eta$. Because of the symmetry of $R_Y(\tau)$, the latter may be found in the form

$$R_Y(\tau) \cong \frac{\Delta\eta}{4\pi} \sum_{k=-m}^m \Theta_Y(\bar{\gamma}_k - 1) |\tau|^{-\bar{\gamma}_k}; \quad \tau > 0 \quad (27)$$

Fourier transform of Eq. 27 restitutes $S_Y(\omega)$ in the form

$$S_Y(\omega) = \frac{\Delta\eta}{2\pi} \sum_{k=-m}^m \Theta_Y(\bar{\gamma}_k - 1) \Gamma(1 - \bar{\gamma}_k) \sin\left(\frac{\bar{\gamma}_k \pi}{2}\right) |\omega|^{\bar{\gamma}_k-1} \quad (28)$$

By letting $\bar{\gamma}_k - 1 = -\gamma_k$, Eq. 28 may be rewritten as

$$S_Y(\omega) = \frac{1}{2b} \sum_{k=-m}^m \Theta_Y(-\gamma_k) v_c(\gamma_k) |\omega|^{-\gamma_k} \quad (29)$$

Comparison with the expression given in Eq. 16b reveals that

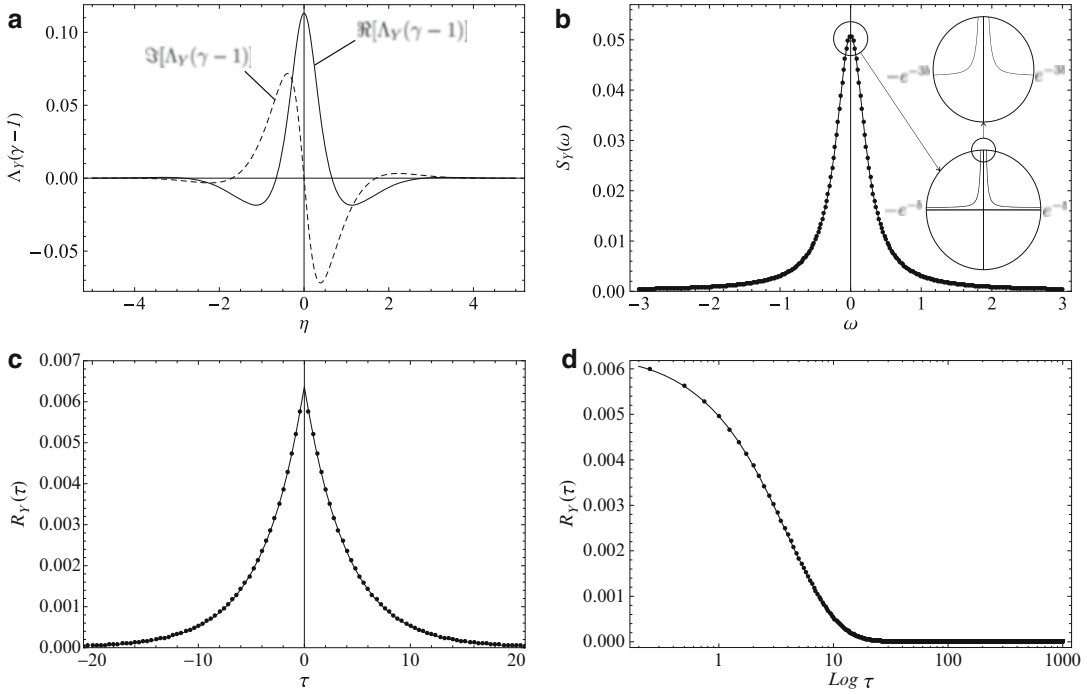
$$\Lambda_Y(\gamma_k - 1) = 2\Theta_Y(-\gamma_k) v_c(\gamma_k) \quad (30)$$

From this equation it may be recognized that the FSM may be also evaluated directly from the correlation function simply by evaluating the Mellin transform of the correlation. If the

Mellin transform is evaluated for a certain value of $\bar{\gamma}_k$, then we may evaluate directly the Mellin transform of the one-sided correlation function in $-\gamma_k$ (which is for a value of ρ translated into the fundamental strip by using the technique described in appendix).

As a concluding remark of this section, we may assert that:

- The FSMs are able to reconstitute both power spectral density (see Eq. 16) and the correlation function (see Eq. 27) in the whole domain with exclusion of $\omega = 0$ or $\tau = 0$, respectively. This pathological behavior, due to the discretization in η axis, may be overcome by assuming the value in zero as the value in $\omega = \exp(-b)$.
- The FSMs never diverge provided ρ lies in the fundamental strip of the Mellin transform, then the cutoff value along the imaginary axis may be selected also for a finite value of m , and the truncation of the complex series does not produce divergences as $\omega \rightarrow \infty$ or $\tau \rightarrow \infty$.
- Since $\Lambda_Y(\gamma_k - 1) = \Lambda_Y^*(\gamma_{-k} - 1)$ and $\omega^{-\gamma_k} = (\omega^{-\gamma_{-k}})^*$, then the summation may run from zero up to m with a very simple manipulation.
- For any Mellin transformable function $\lim_{n \rightarrow \pm\infty} \Lambda_Y(\gamma - 1) \rightarrow 0$, this happens since the $\Lambda_Y(\gamma - 1)$ is holomorph into the fundamental strip. This means that the truncation of the summations in Eq. 16 or in Eq. 27 may be properly performed.
- A very important property that allows us to solve fractional differential equations, Fokker-Planck equation, and many other problem is that once the Mellin transform is evaluated for a fixed value of ρ , say ρ_1 , then it may be easily evaluated for any other value of ρ , say ρ_2 , provided ρ_1 and ρ_2 belong to the FS (see “Appendix”).
- Since $\Lambda_Y(\gamma - 1)$ is able to reconstruct both $R_Y(\tau)$ and $S_Y(\omega)$, then it may be stated that $\Lambda_Y(\gamma - 1)$ is another representation that fully describes the process $Y(t)$. It follows that we may now assert that, for a stationary normal process, we have three equivalent



Complex Fractional Moments and Their Use in Earthquake Engineering, Fig. 1 FSM, PSD, and correlation: (a) FSM along the imaginary line $\rho = 0.5$; (b)

PSD (Eq. 31) continuous line, approximated (Eq. 16) dotted line; (c) correlation function (Eq. 32) continuous line, approximated (Eq. 27) dotted line; (d) Fig. 1c in log plot

representations that fully describe the process in statistic sense that are the PSD, the correlation, and the FSM. The first and the second are related to each other by the Fourier Transform operator, while the third representation is related to both the previous ones by the Mellin transform.

Examples

Such an example let $Y(t)$ be a normal zero mean process with a PSD given as

$$S_Y(\omega) = \frac{\alpha \sigma^2}{\pi(\alpha^2 + \omega^2)} \quad (31)$$

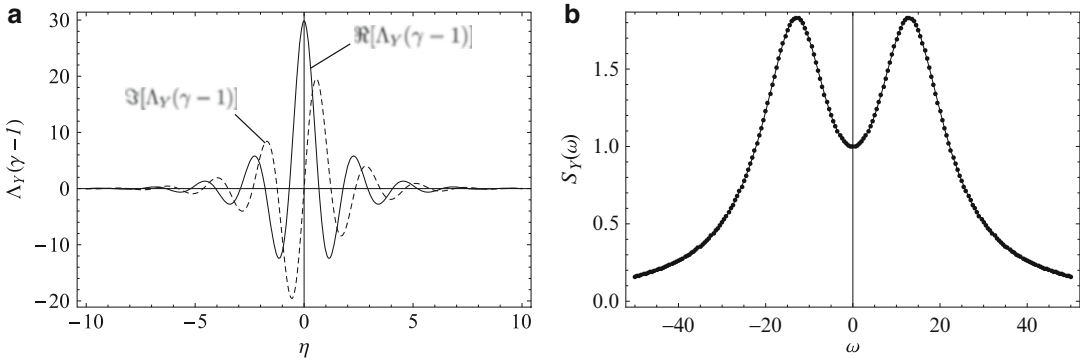
The FS for this PSD is $0 < \rho < 2$. The corresponding correlation function is

$$R_Y(\tau) = \frac{1}{2\pi} \sigma^2 \exp(-\alpha|\tau|) \quad (32)$$

The FSM function is

$$\begin{aligned} \Lambda_Y(\gamma - 1) &= 2 \int_0^\infty \frac{\alpha \sigma^2}{\pi(\alpha^2 + \omega^2)} \omega^{\gamma-1} d\omega \\ &= \alpha^{\gamma-1} \sigma^2 \csc(\pi\gamma/2) \end{aligned} \quad (33)$$

Equation 33 may also be obtained from Eq. 32 by using Eq. 30. We evaluate $\Lambda_Y(\gamma - 1)$ in the discrete sequence $\gamma_k = \rho + ik\Delta\eta$. The parameters selected are $\alpha = 0.25$, $\sigma = 0.2$, $\rho = 0.5$, $m = 20$, and $\Delta\eta = 0.5$. The plot of $\Lambda_Y(\gamma - 1)$ along the imaginary axis is reported in Fig. 1a; it may be observed that $\Lambda_Y(\gamma - 1) \rightarrow 0$ at $\eta \rightarrow \pm \infty$. In Fig. 1b, c the PSD and the correlation function given in Eqs. 31 and 32, respectively, are contrasted with those evaluated by Eqs. 16b and 27. In Figure 1d the log plot of the correlation function is contrasted with the results of Eq. 27, showing that the trend at ro is guaranteed. In Fig. 1b the plot of the PSD in the range $-e^{-b} \div e^{-b}$ is plotted, that for the present case ($\Delta\eta = 0.5$) leads to $-1.87 \cdot 10^{-3} \div 1.87 \cdot 10^{-3}$.



Complex Fractional Moments and Their Use in Earthquake Engineering, Fig. 2 FSM and PSD of the Tajimi-Kanai ground acceleration: (a) FSM versus η

($\rho = 0.5$); (b) PSD evaluated by Eq. 34 continuous line, PSD evaluated with Eq. 16 dotted line

From this figure it appears that into the interval $-e^{-\pi/\Delta\eta} \div e^{-\pi/\Delta\eta}$ Eq. 16b shows a divergence phenomenon, but outside the two curves totally overlaps each another. However, the aforementioned problem may be easily overcome by assuming in $\omega = 0$ (or in $\tau = 0$) the value in $e^{-\pi/\Delta\eta}$ so avoiding the pathological behavior in zero. Moreover, the divergence phenomenon in $-e^{-b} \div e^{-b}$ is self-similar in the sense that in the range $-e^{-3b} \div e^{-3b}$, $-e^{-5b} \div e^{-5b} \dots$ the shape of the curve remains the same and then into $-e^{-b} \div e^{-b}$ we have a fractal structure. However, for technical purposes this fact is not influent and is produced by the fact that $\Delta\eta$ is finite. This phenomenon disappears if $\Delta\eta \rightarrow 0$ and $m \rightarrow \infty$ that corresponds to the integral form of the inverse Mellin transform.

As another example the Tajimi-Kanai (1960) PSD given as

$$S_Y(\omega) = 2\pi S_0 \frac{\omega_g^4 + 4\zeta_g^2 \omega_g^2 \omega^2}{(\omega_g^2 - \omega^2)^2 + 4\zeta_g^2 + \omega_g^2 \omega^2} \quad (34)$$

is considered. In Eq. 34 S_0 is the PSD of the white noise at bedrock, and ω_g and ζ_g are the relevant parameters of the ground deposit. In Fig. 2 the PSD given in Eq. 20 is plotted for $S_0 = 1/2\pi$, $\zeta_g = 0.6$, and $\omega_g = 5\pi$. In Fig. 2a the FSM function is plotted along the imaginary axis for $\rho = 0.5$. In Fig. 1b the PSD given in Eq. 20 is

plotted (continuous line) and contrasted with the PSD given in Eq. 16 (dotted line) for $\Delta\eta = 0.5$ and $m = 20$.

From the previous examples, it is evident that the level of accuracy reached by the use of FSM is quite good including the trend at infinity. It is to be emphasized that smaller values of $\Delta\eta$ do not produce sensible variations on both correlation and PSD evaluated by FSM. By contrast the cut-off value $m\Delta\eta = \bar{\eta}$ on the summation in Eq. 16 or in Eq. 27 may produce variations and then has to be appropriately selected. Finally the number of FSM that is necessary to describe the Tajimi-Kanai PSD is very limited; with these complex quantities, all the PSD is restored (with the exception of the value in zero).

Other examples may be found in Cottone and Di Paola (2010) and Di Paola and Pinnola (2012).

Cross-Correlations and Cross-PSD

As the effect of spatial variability and propagation of seismic ground motion on the response of multiply supported structure is considered, the cross-correlation and cross-PSD of the ground acceleration have to be taken into consideration.

Let $Y(x_1, t)$ and $Y(x_2, t)$ the acceleration of the ground at any pairs of different locations x_1 and x_2 . Once x_1 and x_2 are fixed, we have two distinct processes that will be simply indicated as $Y_1(t)$ and $Y_2(t)$. Because the two processes come out

from a propagation of the seismic waves in the medium, they are correlated with each another. Then in order to define the two processes in probabilistic setting, due to the hypothesis of Gaussianity and stationarity, we need of the correlation function matrix

$$\mathbf{R}_Y(\tau) = E[\mathbf{Y}(t)\mathbf{Y}^T(t+\tau)]$$

$$= \begin{bmatrix} R_{Y_1}(\tau) & R_{Y_1Y_2}(\tau) \\ R_{Y_1Y_2}(-\tau) & R_{Y_2}(\tau) \end{bmatrix} \quad (35)$$

where $\mathbf{Y}^T(t) = [Y_1(t)Y_2(t)]$, $R_{Y_j}(\tau)$ is the correlation of Y_j ($j = 1, 2$), and $R_{Y_1Y_2}$ is the cross-correlation between $Y_1(t)$ and $Y_2(t)$. Since $R_{Y_jY_j}(\tau) \in \mathbb{R}$ and $R_{Y_jY_j}(\tau) = R_{Y_jY_j}(-\tau)$, then the Fourier transform of the correlation matrix, namely, the PSD matrix, is given as

$$\mathbf{S}_Y(\omega) = \begin{bmatrix} S_{Y_1}(\omega) & S_{Y_1Y_2}(\omega) \\ S_{Y_1Y_2}^*(\omega) & S_{Y_2}(\omega) \end{bmatrix} \quad (36)$$

where $S_{Y_1Y_2}(\omega)$ is the cross-PSD of the two processes $Y_1(t)$ and $Y_2(t)$. The PSD matrix (36) is Hermitian and $S_{Y_1Y_2}(\omega)$ is a complex function whose real part is even in ω and the imaginary part is odd. Under the far-field assumption, usually it is assumed that $S_{Y_1Y_2}(\omega) = S_{Y_2Y_1}(\omega) = S_Y(\omega)$ and $S_{Y_1Y_2}(\omega)$ is usually defined as (Harichandran and Vanmarcke 1986; Zerva 1991)

$$S_{Y_1Y_2}(\omega) = S_Y(\omega) |\rho^{(1,2)}(\omega, |\xi|)| \exp(i\omega\xi/v) \quad (37)$$

where $\rho^{(1,2)}(\omega, \xi)$ is the so-called coherence function depending of the radian frequency ω and of the distance $\xi = x_2 - x_1$ between the different points as well. The coherence function is even $\rho^{(1,2)}(\omega, |\xi|) = \rho^{(1,2)}(-\omega, |\xi|)$ and is usually assumed to be exponentially decaying with $|\xi|$. The various parameters are selected in order to fit experimental data. Moreover the factor $\exp(i\omega\xi/v)$, where v is the velocity of the traveling waves in the medium, takes into account for the retardation time of the waves from two different points. In any case, whatever the coherence function is selected, the

representation of the cross-PSD with the FSM requests the definition of the real and imaginary part of Eq. 37:

$$\Re[S_{Y_1Y_2}(\omega)] = S_Y(\omega) |\rho^{(1,2)}(\omega, |\xi|)| \cos \frac{\omega\xi}{v} \quad (38a)$$

$$\Im[S_{Y_1Y_2}(\omega)] = S_Y(\omega) |\rho^{(1,2)}(\omega, |\xi|)| \sin \frac{\omega\xi}{v} \quad (38b)$$

where $\Re[\cdot]$ and $\Im[\cdot]$ are the real and imaginary part of the term in parentheses. The Mellin transform borns for functions that are defined in the range $0 \div \infty$. For $S_{Y_j}(\omega)$ Eq. 16b restores the PSD in terms of FSM. For the cross-PSD, we may consider the real and the imaginary part separately. By taking into account Eq. 16b for the real part given in Eq. 38a, because it is an even function like $S_Y(\omega)$, we may write

$$\Re[S_{Y_1Y_2}(\omega)] = \frac{1}{4b} \sum_{k=-m}^m A_{Y_1Y_2}^{\Re}(\gamma_k - 1) |\omega|^{-\gamma_k} \quad (39)$$

Regarding the imaginary part since $\Im[S_{Y_1Y_2}(\omega)] = -\Im[S_{Y_1Y_2}(-\omega)]$, once the imaginary part of $S_{Y_1Y_2}(\omega)$ is defined in the positive frequency range, it may be immediately evaluated also in the negative one, that is,

$$\Im[S_{Y_1Y_2}(\omega)] = \frac{1}{4b} \text{sgn}(\omega) \sum_{k=-m}^m A_{Y_1Y_2}^{\Im}(\gamma_k - 1) \text{sgn}(\omega) |\omega|^{-\gamma_k} \quad (40)$$

where $\text{sgn}(\cdot)$ is the signum function.

In Eqs. 39 and 40, the FSM $A_{Y_1Y_2}^{\Re}(\gamma - 1)$, $A_{Y_1Y_2}^{\Im}(\gamma - 1)$ are given as

$$A_{Y_1Y_2}^{\Re}(\gamma - 1) = 2 \int_0^{\infty} \Re[S_{Y_1Y_2}(\omega)] \omega^{\gamma-1} d\omega \quad (41)$$

$$A_{Y_1Y_2}^{\Im}(\gamma - 1) = 2 \int_0^{\infty} \Im[S_{Y_1Y_2}(\omega)] \omega^{\gamma-1} d\omega \quad (42)$$

It follows that the cross-PSD is given as

$$S_{Y_1 Y_2}(\omega) = \frac{1}{4b} \sum_{k=-m}^m \left[A_{Y_1 Y_2}^{\Re}(\gamma_k - 1) + i \operatorname{sgn}(\omega) A_{Y_1 Y_2}^{\Im}(\gamma_k - 1) \right] |\omega|^{-\gamma_k} \quad (43)$$

Inverse Fourier transform gives the cross-correlation function in the form

$$R_{Y_1 Y_2}(\tau) = \frac{1}{4\pi b} \sum_{k=-m}^m A_{Y_1 Y_2}^{\Re}(\gamma_k - 1) v_c(1 - \gamma_k) + \operatorname{sgn}(\tau) A_{Y_1 Y_2}^{\Im}(\gamma_k - 1) v_s(1 - \gamma_k) |\tau|^{\gamma_k - 1} \quad (44)$$

where $v_s(1 - \gamma_k) = \Gamma(1 - \gamma_k) \sin[(1 - \gamma_k)\pi/2]$. As in the case of autocorrelation also for the cross-correlation, the FSM may be directly evaluated by starting from $R_{Y_1 Y_2}(\tau)$ instead of $S_{Y_1 Y_2}(\omega)$. In order to do this, we may decompose $R_{Y_1 Y_2}(\tau)$ in an even and an odd function

$$R_{Y_1 Y_2}(\tau) = R_{Y_1 Y_2}^{ev}(\tau) + R_{Y_1 Y_2}^{od}(\tau) \quad (45)$$

where

$$R_{Y_1 Y_2}^{ev}(\tau) = \frac{1}{2} (R_{Y_1 Y_2}(\tau) + R_{Y_1 Y_2}(-\tau)) \quad (46a)$$

$$R_{Y_1 Y_2}^{od}(\tau) = \frac{1}{2} (R_{Y_1 Y_2}(\tau) - R_{Y_1 Y_2}(-\tau)) \quad (46b)$$

Now we may define

$$\begin{aligned} C_{Y_1 Y_2}^{ev}(\tau) &= 2U(\tau) R_{Y_1 Y_2}^{ev}(\tau); \\ C_{Y_1 Y_2}^{od}(\tau) &= 2U(\tau) R_{Y_1 Y_2}^{od}(\tau) \end{aligned} \quad (47)$$

whose Mellin transform gives

$$\begin{aligned} \mathcal{M} \left\{ C_{Y_1 Y_2}^{od}(\tau), \bar{\gamma} \right\} &= \int_0^{\infty} C_{Y_1 Y_2}^{od}(\tau) \tau^{\bar{\gamma}-1} d\tau \\ &= \Theta_{Y_1 Y_2}^{od}(\bar{\gamma} - 1) \end{aligned} \quad (48)$$

then discrete inverse Mellin transform gives

$$C_{Y_1 Y_2}^{ev}(\tau) = \frac{1}{4\pi b} \sum_{k=-m}^m \Theta_{Y_1 Y_2}^{ev}(\bar{\gamma}_k - 1) |\tau|^{-\bar{\gamma}_k}; \quad (49a)$$

$$C_{Y_1 Y_2}^{od}(\tau) = \frac{\operatorname{sgn}(\tau)}{4\pi b} \sum_{k=-m}^m \Theta_{Y_1 Y_2}^{od}(\bar{\gamma}_k - 1) |\tau|^{-\bar{\gamma}_k} \quad (49b)$$

The cross-spectral moments of the correlation function may be rewritten in terms of cross-FSM $\Lambda_{Y_1 Y_2}^{ev}(\gamma_k - 1)$ and $\Lambda_{Y_1 Y_2}^{od}(\gamma_k - 1)$, respectively, in the form

$$\begin{aligned} \Lambda_{Y_1 Y_2}^{ev}(\gamma_k - 1) &= 2\Theta_{Y_1 Y_2}^{ev}(-\gamma_k) v_c(\gamma_k); \\ \Lambda_{Y_1 Y_2}^{od}(\gamma_k - 1) &= 2\Theta_{Y_1 Y_2}^{od}(-\gamma_k) v_s(\gamma_k) \end{aligned} \quad (50)$$

Filter Equation

In order to find the filter equation by using FSM, fractional calculus, and related concepts, we suppose that the ground acceleration $Y(t)$ is a stationary normal stochastic process characterized by $S_Y(\omega)$. We now suppose that $Y(t)$ is the output of a linear differential equation

$$\mathcal{N}[Y(t)] = W(t) \quad (51)$$

or by using its inverse, we get

$$Y(t) = \mathcal{M}[W(t)] \quad (52)$$

where $\mathcal{N}[\cdot]$ and $\mathcal{M}[\cdot]$ are a priori unknown linear operator and $W(t)$ is a normal white noise process $E[W(t)W(t + \tau)] = 2\pi S_0 \delta(\tau)$. The operator $\mathcal{N}[\cdot]$ or $\mathcal{M}[\cdot]$ has to be selected in such a way that the PSD of the output of Eq. 51 coalesces with the target PSD $S_Y(\omega)$, that is, given as

$$S_Y(\omega) = |H(\omega)|^2 S_0 \quad (53)$$

where $H(\omega)$ is the (unknown) transfer function of the linear operator $\mathcal{N}[\cdot]$. Without loss of generality, we assume a unitary S_0 (in L^2/T^3). Unfortunately from Eq. 53 the only information on the

stochastic process $Y(t)$ is the PSD, and from this unique information, the (complex) function $H(\omega)$ may not be reconstructed. However because we are interested to represent the stationary process $Y(t)$, then we may assume that

$$H(\omega) = \sqrt{S_Y(\omega)} \quad (54)$$

With this choice the corresponding impulse response function of $\mathcal{N}[\cdot]$ is symmetric in time domain, and then it violates the causality condition. Even though the causality condition is violated, in the frame of stationary processes, this does not produce any problem since the output of Eq. 51 remains a normal strictly stationary process with PSD $S_Y(\omega)$ that is the target one.

As the impulse response function is assumed as in Eq. 54, we may evaluate the FSM of $\sqrt{S_Y(\omega)} = H(\omega)$

$$\Lambda_H(\gamma - 1) = \int_0^\infty 2\sqrt{S_Y(\omega)}\omega^{\gamma-1}d\omega \quad (55)$$

and the inverse Mellin transform gives

$$\sqrt{S_Y(\omega)} = H(\omega) = \frac{1}{4b} \sum_{k=-m}^m \Lambda_H(\gamma_k - 1) |\omega|^{-\gamma_k} \quad (56)$$

By denoting the impulse response function corresponding to the transfer function $H(\omega) = \sqrt{S_Y(\omega)}$ as $h(t)$, an inverse Fourier transform of Eq. 56 gives

$$h(t) = \frac{1}{4\pi b} \sum_{k=-m}^m v_c(1 - \gamma_k) \Lambda_H(1 - \gamma_k) |t|^{\gamma_k-1} \quad (57)$$

This equation has been obtained observing that $h(t)$ is the inverse Fourier transform of $H(\omega)$, and then it may be treated as the case of correlation function of a PSD defined as $\sqrt{S_Y(\omega)}$ (see Eq. 34).

Once the impulse response is obtained, then the filter equation is readily found as

$$Y(t) = \frac{1}{4\pi b} \sum_{k=-m}^m v_c(1 - \gamma_k) \Lambda_H(1 - \gamma_k) \int_{-\infty}^{\infty} |t - \tau|^{1-\gamma_k} W(\tau) d\tau = \frac{1}{4b} \sum_{k=-m}^m \Lambda_H(1 - \gamma_k) (I^{\gamma_k} W)(t) \quad (58)$$

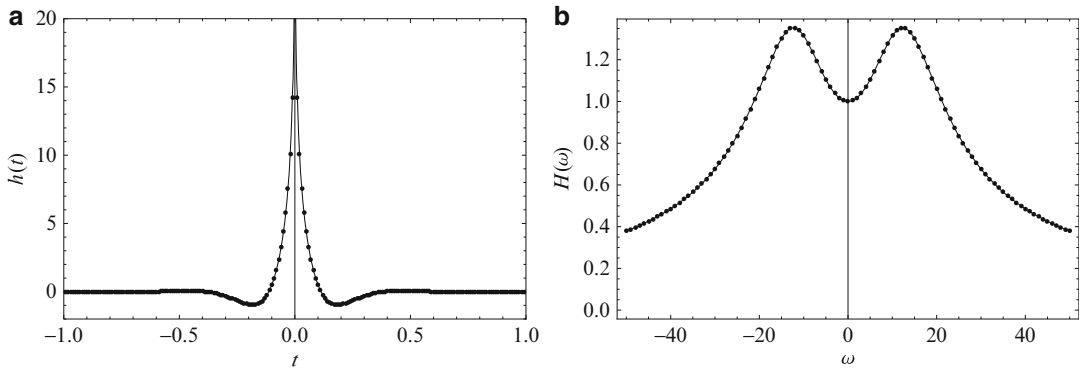
that is the discrete filter equation. An exact representation is given in the form

$$Y(t) = \frac{1}{4\pi i} \int_{\rho-i\infty}^{\rho+i\infty} \Lambda_H(1 - \gamma) (I^{\gamma} W)(t) d\gamma \quad (59)$$

Some comments are necessary to highlight the peculiar aspect of the filter equation in Eq. 58 or 59:

- In order to find the process $Y(t)$, we need of the quantities $\Lambda_H(1 - \gamma_k)$ that play the role of amplification factor of the Riesz fractional derivatives of the white noise $W(t)$. Even though $W(t)$ is nowhere differentiable, the fractional derivative exists since the kernel of the Riesz integral smooths the singularities (see West et al. 2003); see also Chechkin and Gonchar (2001), Grigoriu (2007), and Ortigueira and Batista (2008)).
- Equation 58 looks like a moving-average model, as in fact the actual value of $Y(t)$ depends of the entire past and future history of $W(t)$ as well.
- The value of ρ selected for the definition of the filter equation influences both $\Lambda_H(1 - \gamma_k)$ and $(I^{\gamma_k} W)(t)$, but the result is not affected by the value of the ρ selected provided it belongs to the FS of the Mellin transform.

In Fig. 3a the impulse response function evaluated by Eq. 57 with the PSD of the earthquake acceleration modeled as the Tajimi-Kanai filter given in Eq. 34 with $S_0 = 1/2\pi$, $\omega_g = 5\pi$, $\zeta_g = 0.6$ is plotted. Results are compared with the inverse Fourier transform of $\sqrt{S_Y(\omega)}$. In Fig. 3b the transfer function, obtained as $\sqrt{S_Y(\omega)}$, with $\sqrt{S_Y(\omega)}$, the Tajimi-Kanai PSD is compared with the results obtained by Eq. 56.



Complex Fractional Moments and Their Use in Earthquake Engineering, Fig. 3 $H(\omega)$ and $h(t)$ of the Tajimi-Kanai ground acceleration: (a) $h(t)$ evaluated as inverse Fourier transform of Eq. 54 (continuous line), $h(t)$

evaluated with Eq. 57 (dotted line); (b) $H(\omega)$ evaluated with Eq. 54 (continuous line), $H(\omega)$ evaluated with Eq. 56 (dotted line)

For more details on the filter equations and generation of sample function, see Cottone et al. (2010).

Summary

Fractional calculus opens new perspectives for solving problems that with the classical calculus may not be solved. The initial effort to understand the rules is fully rewarded from the elegance and the unbelievable accuracy obtained.

In this entry only two problems have been presented: representation of the power spectral density (PSD) and correlation by means of complex fractional moments and filter equations. Regarding the first problem, it has been shown that both PSD and correlation may be represented by the so-called fractional spectral moments (FSMs). The latter are the extension of the classical spectral moments introduced by Vanmarcke (1972) to complex order. The appealing in using these FSM is that they are able to reconstruct both PSD and correlation. As a result, it may be stated that FSM function is another equivalent representation of PSD and correlation. Moreover it has been shown that with a limited number of informations (few FSM), the whole PSD and correlation may be restored, including the trend at infinity. Extension to multivariate seismic process has been also provided in both frequency

and time domain. By using FSM also the filter equations have been found. It is shown that both impulse response function and its Fourier transform, namely, the transfer function of the filter, may be found by evaluating the FSM of the squared root of the target PSD.

By using the same concepts, a very large number of other problems may be solved. Such an example the probability density function of a random variable may be obtained with the same technique here used for representing cross-correlations in terms of FSMs. It follows that Fokker-Planck equation, Kolmogorov-Feller equation, Einstein-Smoluchowski equation, and path integral solution (Cottone et al. 2008) may be solved in terms of FSM. Moreover, wavelet transform and classical or fractional differential equations may be easily solved by using fractional calculus and Mellin transform in complex domain.

Acknowledgments The author is indebted to the Ph.D. student Gioacchino Alotta for the help in preparing the numerical applications.

Appendix

In this Appendix it is shown that once the Mellin transform of a given function is evaluated for a certain value ρ_1 (into the fundamental strip), the Mellin transform of the function may be

evaluated for any other value ρ_2 (into the FS) as a linear combination of the fractional spectral moments of the function evaluated in ρ_1 . This fact is very important in order to solve a lot of problems like solution of fractional differential equation, Fokker-Planck equation, and Einstein-Smoluchowski equation and also to evaluate FSM for different value of ρ ; see Eq. 26. Results of this appendix have been found by the author (Di Paola 2014; Di Matteo et al. 2014) for solving Fokker-Planck equation and Kolmogorov-Feller equation.

We denote as $f_{\mathcal{M}}(\gamma^{(1)} - 1)$ the Mellin transform of the function $f(t)$ which evaluates for $\rho = \rho_1$, that is,

$$f_{\mathcal{M}}(\gamma^{(1)} - 1) = \mathcal{M}\{f(t); \gamma\} = \int_0^{\infty} f(t) t^{\gamma^{(1)}-1} dt; \quad (60)$$

$$\gamma^{(1)} = \rho_1 + i\eta$$

Inverse Mellin transform operator returns $f(t)$ in the form

$$f(t) \simeq \frac{\Delta\eta}{2\pi} \sum_{k=-m}^m f_{\mathcal{M}}(\gamma_k^{(1)} - 1) t^{-\gamma_k^{(1)}} \quad (61)$$

Since Eqs. 60 and 61 remain valid also for $\rho = \rho_2$, we may write

$$\begin{aligned} & \frac{\Delta\eta}{2\pi} \sum_{k=-m}^m f_{\mathcal{M}}(\gamma_k^{(1)} - 1) t^{-\gamma_k^{(1)}} \\ & \simeq \frac{\Delta\eta}{2\pi} \sum_{k=-m}^m f_{\mathcal{M}}(\gamma_k^{(2)} - 1) t^{-\gamma_k^{(2)}} \end{aligned} \quad (62)$$

where $\gamma^{(2)} = \rho_2 + i\eta$. By multiplying both sides of Eq. 62 for $t^{-1/2}$ and remembering that $\gamma^{(j)} = \rho^{(j)} + i\eta$ ($j = 1, 2$), we obtain

$$\begin{aligned} & t^{-1/2} \sum_{k=-m}^m f_{\mathcal{M}}(\gamma_k^{(1)} - 1) e^{-ik\frac{\pi}{b}\text{Int}} \simeq t^{-(\Delta\rho+1/2)} \\ & \sum_{k=-m}^m f_{\mathcal{M}}(\gamma_k^{(2)} - 1) e^{-ik\frac{\pi}{b}\text{Int}}; \quad x > 0 \end{aligned} \quad (63)$$

where $\Delta\rho = \rho_2 - \rho_1$. It is to be remarked that Eq. 63 strictly holds for $t > 0$ because in zero singularities appear. Now we want to evaluate coefficients $f_{\mathcal{M}}(\gamma_k^{(2)} - 1)$ from the knowledge of $f_{\mathcal{M}}(\gamma_k^{(1)} - 1)$. Because Eq. 61 is an approximation, we require that Eq. 63 holds true in a weak sense in the interval $t_1 > 0, t_2 \gg 0$:

$$\begin{aligned} & \int_{t_1}^{t_2} \frac{1}{t} \left\{ \left[\sum_{s=-m}^m f_{\mathcal{M}}(\gamma_s^{(1)} - 1) e^{-is\frac{\pi}{b}\text{Int}} - t^{-(\Delta\rho)} \sum_{k=-m}^m f_{\mathcal{M}}(\gamma_k^{(2)} - 1) e^{-ik\frac{\pi}{b}\text{Int}} \right] \times [C.C.] \right\} dt \\ & = \min(f_{\mathcal{M}}(\gamma_k^{(2)} - 1)) \end{aligned} \quad (64)$$

where $C.C.$ stands for complex conjugate. It is convenient to make a change of variables

$$\xi = \text{Int}, d\xi = \frac{dt}{t}; \xi_j = \text{Int}_j, j = 1, 2 \quad (65)$$

Moreover, a good choice of t_1 and t_2 in Eq. 64 is $t_1 = e^{-b} = e^{-\frac{\pi}{\Delta\eta}}$, $t_2 = e^b = e^{\frac{\pi}{\Delta\eta}}$. With this

choice two goals are achieved. First the interval $e^{-b} \div e^b$ is very wide for the usual values of $\Delta\eta$ ($\Delta\eta = 0.25 \div 0.5$). As in fact in the worst case ($\Delta\eta = 0.5$) $e^{-b} = 1.87 \cdot 10^{-3}$, $e^b = 5.35 \cdot 10^2$; second in the ξ domain the extremes of the integral (64) become $-b \div b$ and in this interval $\exp(-i\frac{\pi}{b}s\xi)$ are orthogonal.

Then Eq. 64 in ξ domain is rewritten as

$$\int_{-b}^b \left\{ \left[\sum_{s=-m}^m f_{\mathcal{M}}(\gamma_s^{(1)} - 1) e^{-is\frac{\pi}{b}\xi} - e^{-\Delta\rho\xi} \sum_{k=-m}^m f_{\mathcal{M}}(\gamma_k^{(2)} - 1) e^{-ik\frac{\pi}{b}\xi} \right] \times [C.C.] \right\} d\xi$$

$$= \min(f_{\mathcal{M}}(\gamma_s^{(1)} - 1)) \quad (66)$$

Due to the orthogonality conditions in $-b \div b$ after minimization, we obtain

$$2bf_{\mathcal{M}}(\gamma_s^{(1)} - 1) = \sum_{k=-m}^m f_{\mathcal{M}}(\gamma_k^{(2)} - 1) a_{ks}(\Delta\rho) \quad (67)$$

where

$$a_{ks}(\Delta\rho) = \int_{-b}^b e^{-\Delta\rho\xi} e^{-i(k-s)\frac{\pi}{b}\xi} d\xi$$

$$= \frac{2b \sin[\pi(k-s) - ib\Delta\rho]}{\pi(k-s) - ib\Delta\rho} \quad (68)$$

Inspection of Eqs. 67 and 68 reveals that the Mellin transform $f_{\mathcal{M}}(\gamma_s^{(1)} - 1)$ is related to the Mellin transform $f_{\mathcal{M}}(\gamma_k^{(2)} - 1)$ as a linear combination of $f_{\mathcal{M}}(\gamma_k^{(1)} - 1)$ with the coefficients in Eq. 68.

References

- Cacciola P, Colajanni P, Muscolino G (2004) Combination of modal response consistent with seismic input representation. *J Struct Eng ASCE* 130(1):47–55
- Chechkin AV, Gonchar VY (2001) Fractional Brownian motion approximation based in fractional integration of white noise. *Chaos Solitons Fractals* 12:391–398
- Conte JP, Peng BF (1997) Fully nonstationary analytical earthquake ground motion model. *J Eng Mech ASCE* 123:15–24
- Cottone G, Di Paola M (2010) New representation of power spectral density function and correlation function by means of fractional spectral moments. *Probab Eng Mech* 25:348–353
- Cottone G, Di Paola M, Pirrotta A (2008) Path integral solution by fractional calculus. *J Phys Conf Ser* 96(1):012007
- Cottone G, Di Paola M, Santoro R (2010) A novel exact representation of stationary colored Gaussian processes (fractional differential approach). *J Phys A Math Theor* 43:1–16
- Di Matteo A, Di Paola M, Pirrotta A (2014) Probabilistic characterization of nonlinear systems under Poisson white noise via complex fractional moments. *Nonlinear Dyn.* doi:10.1007/s11071-014-1333-1
- Di Paola M (2014) Fokker Planck equation solved in terms of complex fractional moments. *Probab Eng Mech* 38:70–76
- Di Paola M, Petrucci G (1990) Spectral moments and pre-envelope covariances of non separable processes. *ASME J Appl Mech* 112:218224
- Di Paola M, Pinnola FP (2012) Riesz fractional integrals and complex fractional moments for the probabilistic characterization of random variables. *Probab Eng Mech* 29:149–156
- Falsone G, Neri F (1999) Stochastic modeling of earthquake excitation following the EC8, power spectrum and filtering equations. *Eur Earthq Eng* 13:45–54
- Gorenflo R, Mainardi F (1997) Fractional calculus integrals and differential equations of fractional order. In: Carpinteri A, Mainardi F (eds) *CISM courses and lectures*, vol 378. Springer, New York
- Grigoriu M (2007) Linear systems with fractional Brownian motion and Gaussian noise. *Probab Eng Mech* 22:276–284
- Harichandran RS, Vanmarcke EH (1986) Stochastic variation of earthquake ground motion in space and time. *J Eng Mech ASCE* 112:154–174
- Michaelov G, Sarkani S, Lutes LD (1999) Spectral characteristics of non-stationary random processes-response of a simple oscillator. *Struct Saf* 21(3):245–267
- Miller KS, Ross B (1993) An introduction to the fractional calculus and fractional differential equations. Wiley, New York
- Muscolino G (1991) Nonstationary pre-envelope covariances of nonclassically damped systems. *J Sound Vib* 149:107–123
- Ortigueira MD, Batista AG (2008) On the relation between the fractional Brownian motion and the fractional derivatives. *Phys Lett A* 372:158–198
- Podlubny I (1999) Fractional differential equations, vol 198, *Mathematics in science and engineering*. Academic, London

- Samko SG, Kilbas AA, Marichev OI (1993) Fractional calculus integrals and derivatives. Gordon and Breach Science Publishers, Amsterdam
- Shinozuka M, Deodatis G (1988) Stochastic process models for earthquake ground motion. *Probab Eng Mech* 3(3):114–123
- Shinozuka M, Sato Y (1967) Simulation of nonstationary random process. *J Eng Mech ASCE* 93:11–40
- Solomos GP, Spanos PD (1984) Oscillator response to non-stationary excitation. *Am Soc Mech Eng Appl Mech Div AMD* 65:159–170
- Spanos PD (1983) ARMA algorithms for ocean wave modeling. *Trans ASME J Energy Resour Technol* 105:300–309
- Spanos PD (1986) Filter approach to wave kinematics approximations. *Appl Ocean Res* 8:2–7
- Spanos PD, Miller SM (1994) Hilbert transform generalization of a classical random vibration integral. *J Appl Mech* 61:575–581
- Tajimi H (1960) A standard method of determining the maximum response of a building structure during an earthquake. In: *Proceedings of 2nd world conference on Earthquake Engineering*, vol 2, Tokyo
- Vanmarcke EH (1972) Properties of spectral moments with applications to random vibration. *J Eng Mech ASCE* 98:425–446
- West JW, Bologna M, Grigolini P (2003) *Physics of fractal operators*. Springer, New York
- Zerva A (1991) Effect of spatial variability and propagation of seismic ground motions on the response of multiply supported structures. *Probab Eng Mech* 6:212–221

Conditional Spectra

Ting Lin¹ and Jack Baker²

¹Department of Civil, Construction and Environmental Engineering, Marquette University, Milwaukee, WI, USA

²Stanford University, Stanford, CA, USA

Synonyms

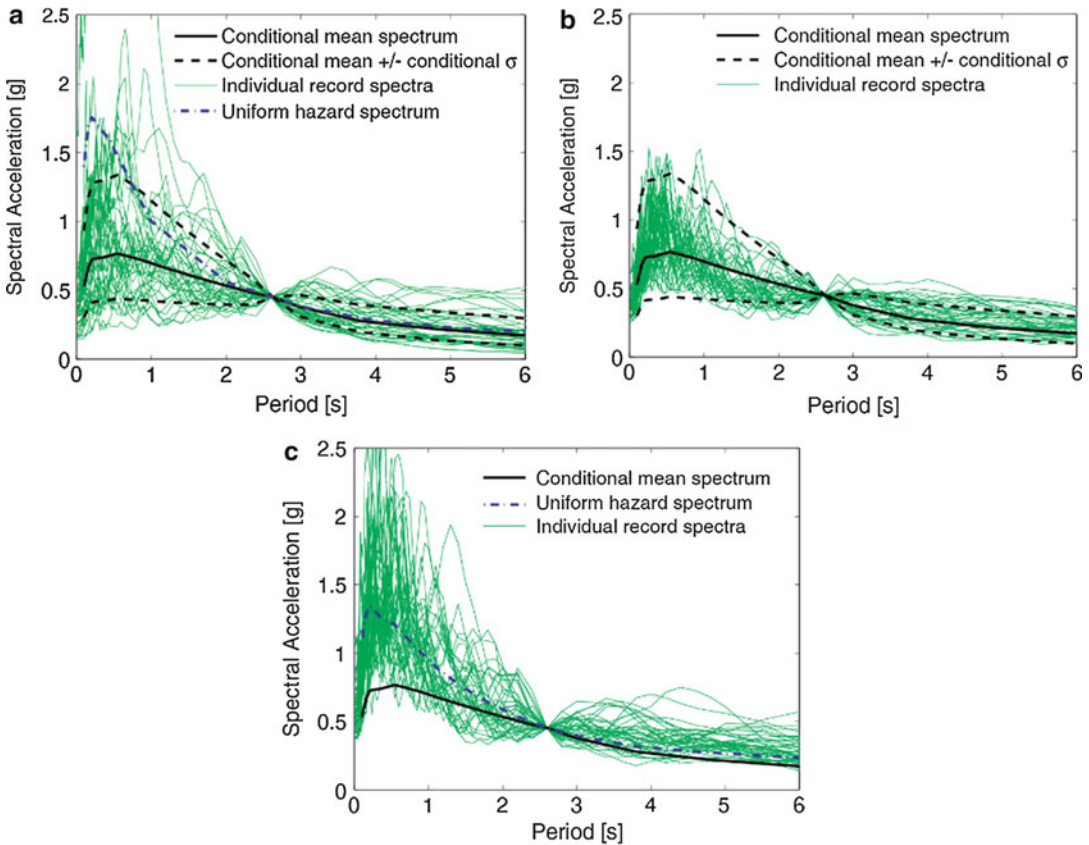
Conditional mean spectrum; Hazard-consistent ground motion selection; Response spectrum; Scenario spectrum; Spectral shape

Introduction

The conditional spectrum (CS) is a response spectrum that specifies the probability distribution of spectral accelerations, S_a , over a range of

periods of vibration, T_i , conditioned on spectral acceleration at a conditioning period, T^* , of interest. The conditional spectrum utilizes the correlations between spectral accelerations at different periods (e.g., Baker and Jayaram 2008) to compute the expected (logarithmic mean) response spectrum (Baker and Cornell 2006; Baker 2011) and additionally account for the variability (variance) of the response spectra (Jayaram et al. 2011; Lin et al. 2013a). Assuming the distribution of logarithmic spectral accelerations is multivariate normal (Jayaram and Baker 2008), then the first two moments (i.e., the means, standard deviations, and correlations) fully describe the conditional spectrum. The CS provides the link between seismic hazard and structural response, the first two elements of performance-based earthquake engineering (PBEE). To maintain consistency with probabilistic seismic hazard analysis (PSHA), the computation of the CS is refined by incorporating multiple causal earthquakes and ground motion prediction models (GMPMs, also known as ground motion prediction equations, attenuation relationships) (Lin et al. 2013a). The CS can be used as a target response spectrum for site- and structure-specific ground motion selection (e.g., Lin et al. 2013b, c).

Baker and Cornell (2006) and Baker (2011) described the “conditional spectrum” as the “conditional mean spectrum” (CMS) with an emphasis on the mean spectrum and its associated ground motion selection where the best ground motion candidates have response spectra as close to the CMS as possible. Bradley (2010) extended the concept of the CMS to ground motion parameters other than response spectra and referred the resulting distribution as a “generalized conditional intensity measure” (GCIM). Bradley (2012) also developed an algorithm to select ground motions based on the GCIM. Abrahamson and Al Atik (2010) coined the term “conditional spectrum”, but represented the CS distribution directly by realizations of the CS rather than an analytical distribution. Wang (2011) developed a ground motion selection algorithm that captures response spectrum characteristics and variability of scenario earthquakes. Gulercce and Abrahamson (2011) extended the CS concept to vertical ground



Conditional Spectra, Fig. 1 Response spectra of ground motions selected and scaled with (a) conditional spectrum, (b) conditional mean spectrum, and (c) uniform

hazard spectrum as target spectra for $S_a(2.6\text{ s})$ associated with 2 % in 50 years probability of exceedance for an illustrative structure in Palo Alto, California, USA

motions. Jayaram et al. (2011) developed a computationally efficient algorithm to select ground motions that match the conditional mean and variance of the response spectrum.

Lin et al. (2013a) for the first time used the term “conditional spectrum” which was used to represent an analytical distribution of response spectra in the manner described here. The CS builds upon the CMS (which focuses on the mean) and includes a measure of the variance in addition to the mean. Figure 1a depicts the analytical distribution of the conditional spectrum with conditional mean spectrum as the thick solid line as well as conditional mean $\pm$ conditional standard deviation as the thick dashed lines. Lin et al. (2013b, c) utilized the conditional spectrum to select ground motions for nonlinear dynamic analysis (also known as response history

analysis) in risk-based and intensity-based assessments respectively. Intensity-based assessment estimates structural response given ground motions with a specified intensity level. Risk-based assessment (also known as PEER integral, drift hazard, time-based assessment, and probabilistic seismic demand analysis) estimates the mean annual rate of exceeding a specified structural response amplitude and can be obtained by integrating intensity-based assessment results at various intensity levels with the associated seismic hazard curve. As illustrated in Fig. 1a, the thin lines represent response spectra from individual ground motion records selected to match the mean and variance of the CS (thick solid and dashed lines) for a specified intensity level, i.e., spectral acceleration at the period of 2.6 s (“pinched” conditioning period) associated with

2 % in 50 years probability of exceedance in Palo Alto, California, USA.

The consideration of the variability in the CS is important to capture the full distribution of the target S_a and its subsequent structural response analysis based on a target response spectrum. Lin et al. (2013c) compared the conditional spectrum (Fig. 1a) to such alternative response spectra as the conditional mean spectrum (Fig. 1b) and the uniform hazard spectrum (UHS) (Fig. 1c); the issue of the choice of the conditioning period was also discussed. The CMS (thick solid line) and UHS (thick dash-dot line) are superimposed on the CS as reference spectra in Fig. 1a. Figure 1b depicts the CMS as the target response spectrum, whereas the thin lines represent individual record response spectra selected to match the mean. Figure 1c illustrates the UHS as the target response spectrum and individual record spectra selected to match the UHS. Unlike the UHS (Fig. 1c), the CS (Fig. 1a) has a lower mean at periods away from the conditioning period. Unlike the CMS (Fig. 1b), the CS additionally considers the spectral uncertainty about the mean.

Background

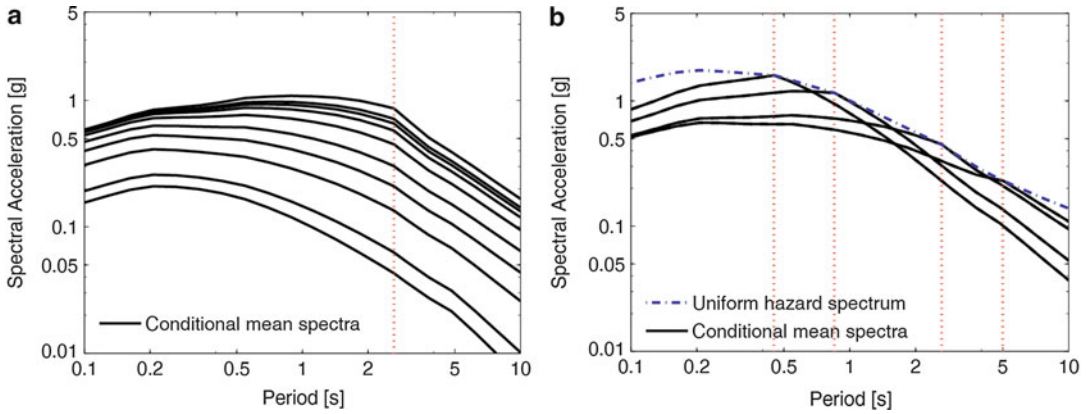
Performance-based earthquake engineering quantifies the seismic hazard, predicts the structural response, estimates the damage to building elements (structural, non-structural, and content), in order to assess resulting losses in terms of dollars, downtime, and deaths (e.g., Cornell and Krawinkler 2000). One key input to seismic design and analysis of structures is earthquake ground motion. Ground motion selection determines ground motion input for a structure at a specific site for nonlinear dynamic analysis. Ground motion selection provides a significant basis for conclusions regarding structural safety, since ground motion uncertainty contributes considerably to uncertainty in structural analysis output.

The source of ground motion inputs can come from (1) simulations, either physics-based (physics-based ground motion simulation),

stochastic (stochastic ground motion simulation), or hybrid; (2) spectral matching; (3) selection and scaling of recorded ground motions from ground motion databases (selection of ground motions for response history analysis). Physics-based simulations attempt to depict the physical phenomenon of earthquake fault rupture and wave propagation using analytical models while stochastic simulations rely heavily on empirical calibration with an attempt to generate ground motions using fewer random variables. Hybrid simulations, on the other hand, combine the two by utilizing physics-based simulations for low-frequency components and stochastic simulations for high-frequency components. A semi-artificial ground motion modification technique is spectral matching that changes the frequency content of the response spectra to be compatible with the target. Another modification technique is selection and scaling of ground motions (e.g., Katsanos et al. 2010) that involves selecting recorded ground motions from ground motion databases and applying amplitude scaling to match the target.

To ensure the proper quantification of ground motion uncertainty, it is important to link ground motion selection to probabilistic seismic hazard analysis (e.g., Cornell 1968). PSHA (probabilistic seismic hazard models) is commonly used in structural dynamic analysis and geotechnical earthquake engineering to identify the ground motion hazard for which structural and geotechnical systems are analyzed and designed. PSHA accounts for the aleatory uncertainties (which are inherently random) of causal earthquakes with different magnitudes and distances with predictions of resulting ground motion intensity in order to compute seismic hazard at a site. PSHA also incorporates epistemic uncertainties (which are due to limited knowledge) in ground motion predictions, by considering multiple ground motion prediction models using a logic tree that also includes seismic source models.

Ground motion selection is often associated with a target response spectrum. For instance, the target response spectrum in current building codes (e.g., ASCE 2010) is based on the uniform



Conditional Spectra, Fig. 2 (a) Conditional mean spectra at $T^* = 2.6$ s at various intensity levels (from 50 % in 30 years to 1 % in 200 years) and (b) conditional mean

spectra at various conditioning periods (compared to uniform hazard spectrum) at the 2 % in 50-year intensity level

hazard spectrum, which assumes equal probabilities of exceedance of spectral accelerations at all periods. However, no single ground motion is likely to produce a response spectrum as high as that of the UHS over a wide range of periods (e.g., Naeim and Lew 1995). The conditional mean spectrum utilizes the correlations of S_a across periods to more realistically compute the expected S_a values at all periods given a target S_a value at a period of interest (e.g., Baker and Cornell 2006). The conditional spectrum quantifies the conditional distribution (mean and variance) of S_a at all periods given S_a at a period of interest (e.g., Lin et al. 2013a). The spectral shapes of conditional mean spectra (mean of conditional spectra) change with amplitudes (Fig. 2a). Uniform hazard spectrum is the envelope of conditional mean spectra at various periods (Fig. 2b).

The target response spectrum used when selecting ground motions has an impact on the distribution of structural responses from resulting nonlinear dynamic analysis (e.g., Jayaram et al. 2011; Lin et al. 2013c). Matching target response variance (in the CS compared to the CMS) increases the dispersion of structural responses, thereby affecting the distribution of structural responses and consequently the damage state and loss estimation computations in performance-based earthquake engineering. The increase in the

dispersion leads to higher and lower extremes of structural responses and the associated damage states and losses. The increased dispersion can also lead to a larger probability of structural collapse. Incorporating the variability in the CS has significant implications for applications where the dispersion of the responses is an important consideration.

Basic Computation

The basic computation of the conditional spectrum requires (1) input earthquake parameters such as magnitude and distance (similar to those required by a GMPM), (2) specification of a conditioning period, (3) a ground motion prediction model to estimate mean and standard deviation of logarithmic spectral acceleration, and (4) a correlation model of spectral accelerations across periods. The computation procedure is outlined as follows:

- (a) Obtain input earthquake parameters: Obtain the target spectral acceleration at the conditioning period T^* , $S_a(T^*)$, from PSHA, and its associated mean causal earthquake magnitude (M), distance (R), and other parameters (if available) such as rupture mechanism and site condition (θ), from deaggregation (e.g., McGuire 1995).

- (b) Compute mean and standard deviation of $\ln Sa(T_i)$: Use a GMPM (e.g., Abrahamson et al. 2008) to compute the logarithmic mean and standard deviation of Sa at all periods T_i , denoted as $\mu_{\ln Sa}(M, R, \theta, T_i)$ and $\sigma_{\ln Sa}(M, \theta, T_i)$.
- (c) Compute epsilon, $\epsilon(T_i)$: For any $Sa(T_i)$, compute $\epsilon(T_i)$, the number of standard deviations by which $\ln Sa(T_i)$ differs from the mean spectral ordinate predicted by a given GMPM, $\mu_{\ln Sa}(M, R, \theta, T_i)$, at T_i

$$\epsilon(T_i) = \frac{\ln Sa(T_i) - \mu_{\ln Sa}(M, R, \theta, T_i)}{\sigma_{\ln Sa}(M, \theta, T_i)} \quad (1)$$

The target $\epsilon(T^*)$ can also be computed using this equation. Spectral shape is characterized by ϵ .

- (d) Compute conditional mean $\mu_{\ln Sa(T_i)|\ln Sa(T^*)}$: Using a correlation model (e.g., Baker and Jayaram 2008) that specifies the correlation coefficient between pairs of epsilon values at two periods, $\rho(\epsilon(T_i), \epsilon(T^*))$ (hereinafter referred to as $\rho(T_i, T^*)$), the conditional mean logarithmic spectral acceleration at other periods T_i , $\mu_{\ln Sa(T_i)|\ln Sa(T^*)}$ is computed

$$\begin{aligned} \mu_{\ln Sa(T_i)|\ln Sa(T^*)} &= \mu_{\ln Sa}(M, R, \theta, T_i) \\ &+ \rho(T_i, T^*) \sigma_{\ln Sa}(M, \theta, T_i) \end{aligned} \quad (2)$$

The spectrum defined by $\mu_{\ln Sa(T_i)|\ln Sa(T^*)}$ is termed the “conditional mean spectrum”, as it specifies the mean values of $\ln Sa(T_i)$ conditional on the value of $\ln Sa(T^*)$. Assuming Sa is lognormally distributed, the exponentials of $\mu_{\ln Sa}$ are equivalent to the median values of Sa .

- (e) Compute conditional standard deviation $\sigma_{\ln Sa(T_i)|\ln Sa(T^*)}$: The standard deviation of logarithmic spectral acceleration at period T_i conditioned on the value of $\ln Sa(T^*)$ is

$$\sigma_{\ln Sa(T_i)|\ln Sa(T^*)} = \sigma_{\ln Sa}(M, \theta, T_i) \sqrt{1 - \rho^2(T_i, T^*)} \quad (3)$$

The conditional standard deviation $\sigma_{\ln Sa(T_i)|\ln Sa(T^*)}$, when combined with the

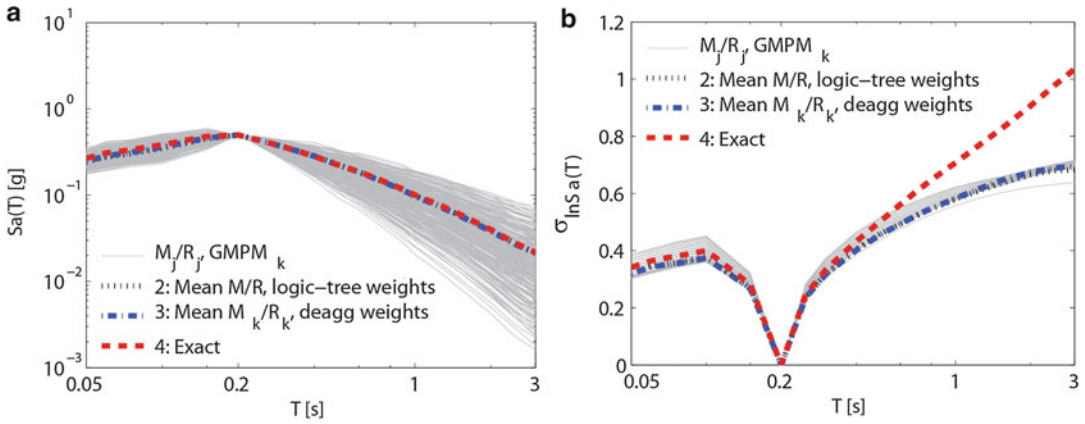
conditional mean value $\mu_{\ln Sa(T_i)|\ln Sa(T^*)}$, specifies a complete distribution of logarithmic spectral acceleration values at all periods (where the distribution at a given period is Gaussian, as justified by Jayaram and Baker 2008). The resulting spectrum distribution is termed the “conditional spectrum”, to be distinguished from the CMS that does not consider the variability.

Deaggregation of Ground Motion Prediction Models

Probabilistic seismic hazard deaggregation of GMPMs links the computation of a target spectrum (for ground motion selection) to the total hazard prediction (from PSHA). Computation of a target spectrum requires deaggregation (via Bayes’ rule) to identify the causal ground motion parameters, along with the predictions from multiple GMPMs. Hazard-consistent ground motion selection incorporates the aleatory uncertainties from earthquake scenarios and the epistemic uncertainties from multiple GMPMs, through deaggregation of GMPMs (Lin and Baker 2011).

This GMPM deaggregation is consistent with the probabilistic treatment of the magnitude and distance random variables in traditional PSHA. The deaggregation of GMPMs provides additional insights into which GMPM contributes most to prediction of Sa values of interest. To match the contribution of each GMPM to its associated ground motion parameters, separate deaggregation of earthquake parameters for each GMPM is also performed. First, seismic hazard is estimated using PSHA that incorporates multiple GMPMs. Next, the relative contributions of events and GMPMs to the hazard prediction are identified using the refined deaggregation procedures.

The deaggregation of GMPMs is performed as an intermediate step to develop a refined mean and, more importantly, a refined standard deviation of the CS. This can also be used to develop, in general, any new target response spectrum that is consistent with PSHA. The side product of this computation is the deaggregation (posterior) weights of GMPMs, instead of the



Conditional Spectra, Fig. 3 (a) Conditional means and (b) conditional standard deviations of the CS computed using Methods 2–4 for $Sa(0.2\text{ s})$ with 10 % probability of exceedance in 50 years at Bissell

logic-tree (prior) weights. This methodology for deaggregation of prediction models can also be immediately applicable to other procedures which require multiple prediction models in an earlier stage of total prediction and a later stage of new target computation.

Refined Computational Methods

Approximate and exact implementations of CS computations are outlined and used for example calculations at Stanford, Bissell, and Seattle (Lin et al. 2013a). Figure 3 illustrates CS mean and standard deviation computational methods using multiple GMPMs at Bissell. Exact CS mean and standard deviation calculations (Method 4) can incorporate multiple GMPMs and causal earthquake M/R combinations. Varying levels of approximations are also considered, that replace multiple M/R combinations with the mean M/R from deaggregation (Methods 1–3), and either consider only a single GMPM (Method 1) or perform an approximate weighting of several GMPMs (Methods 2–3). Methods 1–2 utilize the overall mean M/R , whereas Method 2 considers multiple GMPMs with logic-tree weights. The deaggregation of GMPMs provides GMPM-specific mean M/R and GMPM deaggregation weights, both of which are incorporated in Method 3. The most accurate computation,

Method 4, accounts for the contribution that each GMPM and each earthquake M/R makes to the seismic hazard (Fig. 3). Subscript k denotes that the calculations are based on a specific GMPM indexed by k , whereas subscript j denotes that the parameters such as M/R are associated with a specific earthquake indexed by j . For each GMPM k and causal earthquake combination M_j/R_j , mean and standard deviation of the CS can be obtained (using Eqs. 12 and 13, respectively) that are then combined with corresponding deaggregation weights, $p_{j,k}^d$ (Eqs. 14 and 15). The weight for model k is denoted as p_k^l where the superscript l refers to logic-tree; similarly, the superscript d in p_k^d and $p_{j,k}^d$ refers to deaggregation.

Method 1: Approximate CS using mean M/R and a single GMPM (basic computation)

$$\begin{aligned} \mu_{\ln Sa, k(T_i) | \ln Sa(T^*)} &\approx \mu_{\ln Sa, k}(\bar{M}, \bar{R}, \bar{\theta}, T_i) \\ &+ \rho(T_i, T^*) \sigma_{\ln Sa, k}(\bar{M}, \bar{\theta}, T_i) \bar{\epsilon}(T^*) \end{aligned} \quad (4)$$

$$\sigma_{\ln Sa, k(T_i) | \ln Sa(T^*)} \approx \sigma_{\ln Sa, k}(\bar{M}, \bar{\theta}, T_i) \sqrt{1 - \rho^2(T_i, T^*)} \quad (5)$$

Method 2: Approximate CS using mean M/R and GMPMs with logic-tree weights

$$\mu_{\ln Sa(T_i) | \ln Sa(T^*)} \approx \sum_k p_k^l \mu_{\ln Sa, k(T_i) | \ln Sa(T^*)} \quad (6)$$

$$\sigma_{\ln Sa(T_i)|\ln Sa(T^*)} \approx \sqrt{\sum_k p_k^l \left(\sigma_{\ln Sa, k(T_i)|\ln Sa(T^*)}^2 + \left(\mu_{\ln Sa, k(T_i)|\ln Sa(T^*)} - \mu_{\ln Sa(T_i)|\ln Sa(T^*)} \right)^2 \right)} \quad (7)$$

Method 3: Approximate CS using GMPM-specific mean M/R and GMPMs with deaggregation weights

$$\sigma_{\ln Sa, k(T_i)|\ln Sa(T^*)} \approx \sigma_{\ln Sa, k}(\bar{M}_k, \bar{\theta}_k, T_i) \sqrt{1 - \rho^2(T_i, T^*)} \quad (9)$$

$$\begin{aligned} \mu_{\ln Sa, k(T_i)|\ln Sa(T^*)} &\approx \mu_{\ln Sa, k}(\bar{M}_k, \bar{R}_k, \bar{\theta}_k, T_i) \\ &+ \rho(T_i, T^*) \sigma_{\ln Sa, k}(\bar{M}_k, \bar{\theta}_k, T_i) \bar{\epsilon}_k(T^*) \end{aligned} \quad (8) \quad \mu_{\ln Sa(T_i)|\ln Sa(T^*)} \approx \sum_k p_k^d \mu_{\ln Sa, k(T_i)|\ln Sa(T^*)} \quad (10)$$

$$\sigma_{\ln Sa(T_i)|\ln Sa(T^*)} \approx \sqrt{\sum_k p_k^d \left(\sigma_{\ln Sa, k(T_i)|\ln Sa(T^*)}^2 + \left(\mu_{\ln Sa, k(T_i)|\ln Sa(T^*)} - \mu_{\ln Sa(T_i)|\ln Sa(T^*)} \right)^2 \right)} \quad (11)$$

Method 4: “Exact” CS using multiple causal earthquake M/R and GMPMs with deaggregation weights

$$\sigma_{\ln Sa, j, k(T_i)|\ln Sa(T^*)} = \sigma_{\ln Sa, k}(M_j, \theta_j, T_i) \sqrt{1 - \rho^2(T_i, T^*)} \quad (13)$$

$$\begin{aligned} \mu_{\ln Sa, j, k(T_i)|\ln Sa(T^*)} &= \mu_{\ln Sa, k}(M_j, R_j, \theta_j, T_i) \\ &+ \rho(T_i^*, T_i) \epsilon_j(T^*) \sigma_{\ln Sa, k}(M_j, \theta_j, T_i) \end{aligned} \quad (12) \quad \mu_{\ln Sa(T_i)|\ln Sa(T^*)} = \sum_k \sum_j p_{j, k}^d \mu_{\ln Sa, j, k(T_i)|\ln Sa(T^*)} \quad (14)$$

$$\sigma_{\ln Sa(T_i)|\ln Sa(T^*)} = \sqrt{\sum_k \sum_j p_{j, k}^d \left(\sigma_{\ln Sa, j, k(T_i)|\ln Sa(T^*)}^2 + \left(\mu_{\ln Sa, j, k(T_i)|\ln Sa(T^*)} - \mu_{\ln Sa(T_i)|\ln Sa(T^*)} \right)^2 \right)} \quad (15)$$

Implementations

The refined CS computational methods increase with accuracy and complexity. Several factors contribute to the accuracy of approximations. First, the importance of considering multiple causal earthquakes depends upon how many differing causal earthquake M/R values contribute significantly to the hazard. In addition, variation in other parameters besides M and R , such as

depth to the top of rupture for different source types, can affect the accuracy of the approximation since they contribute to the Sa prediction. Second, the similarity of the GMPMs affects approximations in their treatment. For cases where the ground motion predictions for a given M/R vary significantly between models, approximate treatment of GMPMs is less effective. Third, as the GMPM deaggregation weights differ more from the GMPM logic-tree weights,

approximate treatment of those weights works less effectively. The exact conditional standard deviation is always higher than approximate results (Fig. 3b) because of the additional contribution from the variance in mean logarithmic spectral accelerations (Fig. 3a) due to variation in causal earthquakes and GMPMs. The impact of CS approximations on engineering decisions depends upon the difference between approximate and exact spectra: as the difference increases, more refined target spectra will have increasing benefits for ground motion selection and structural response assessment (Lin et al. 2013a).

The deaggregation of ground motion prediction models and the exact conditional mean spectrum (incorporating multiple causal earthquakes and multiple GMPMs) have been implemented in the United States Geological Survey hazard mapping tools (<http://geohazards.usgs.gov/deaggint/2008/>). Carlton and Abrahamson (2014) proposed an alternative Method 2.5 for a simplified implementation of the CS, which gives similar results to Method 3 in Lin et al. (2013a). Deaggregation weights, instead of logic-tree weights, of GMPMs were recommended when multiple GMPMs were used to compute the CS (e.g., Lin et al. 2013a; Carlton and Abrahamson 2014). Gulercce and Abrahamson (2011) extended the CS concept to vertical ground motions. Bradley (2010) developed a generalized conditional intensity measure approach to include intensity measures other than S_a .

Ground Motion Selection Algorithms

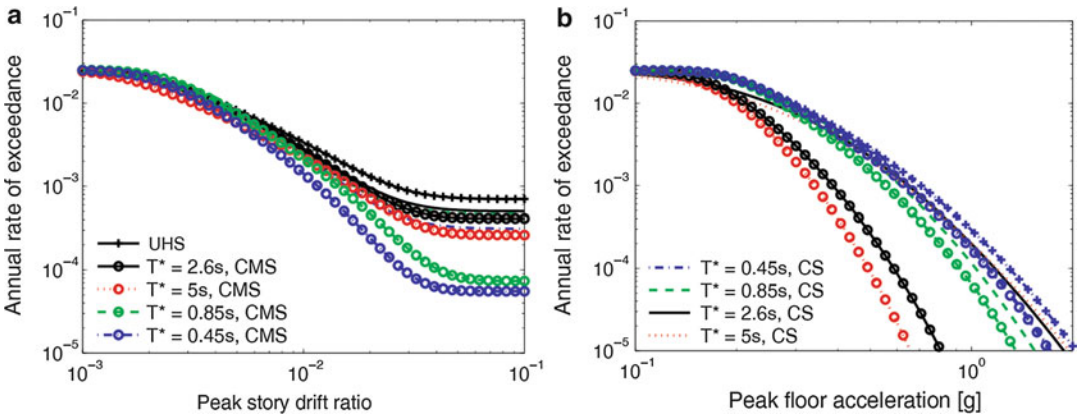
The conditional spectrum can be used as the target spectrum to select “representative” ground motions from a ground motion database for performance-based earthquake engineering. A computationally efficient algorithm (Jayaram et al. 2011) suitable for such CS-based ground motion selection (http://www.stanford.edu/~bakerjw/gm_selection.html) computes a target spectrum, and then selects and scales ground motions from the PEER NGA database to match the target spectrum mean and variance. This

selection algorithm (1) probabilistically generates multiple response spectra from a target distribution, and then (2) selects recorded ground motions whose response spectra individually match the simulated response spectra. (3) A greedy optimization technique further improves the match between the target and the sample means and variances. This algorithm has been adopted by the global earthquake model (<http://www.globalquakemodel.org/>) for ground motion selection. Alternative algorithms that can be used for CS-based ground motion selection include those by, e.g., Wang (2011) and Bradley (2012).

If the uniform hazard spectrum or the conditional mean spectrum, instead of the conditional spectrum, is used as the target spectrum to select (and optionally scale) ground motions, the PEER ground motion database (http://peer.berkeley.edu/products/strong_ground_motion_db.html) is an online tool appropriate for such applications. In addition to CS-based ground motion selection (as shown in Fig. 1a), Jayaram et al. (2011) algorithm can also be modified to accommodate CMS- and UHS-based ground motion selection, as illustrated in Fig. 1b, c respectively.

Hazard Consistency in Risk-Based Assessments

Risk-based assessment estimates the mean annual rate of exceeding a specified structural response amplitude. The conditional spectrum computation requires specification of a conditioning period. The sensitivity of risk-based assessment results to the choice of conditioning period using CS-based ground motion selection is investigated in Lin et al. (2013b). This study focuses on risk-based assessments, with a specific emphasis on the rates of exceeding various levels of peak story drift ratio (i.e., drift hazard calculations) in the structure. Some additional engineering demand parameters (EDPs) are also considered, such as the peak floor acceleration over the full building heights, a single-story story drift ratio, and a single-story floor acceleration. The primary structure considered is a 20-story reinforced concrete frame structure



Conditional Spectra, Fig. 4 Risk-based assessments of (a) peak story drift ratio and (b) peak floor acceleration of a 20-story reinforced concrete frame structure obtained

from ground motions selected to match the CS, the CMS, and the UHS

assumed to be located in Palo Alto, California, using a structural model (Haselton and Deierlein 2007) with strength and stiffness deterioration that is believed to reasonably capture the responses up to the point of collapse due to dynamic instability.

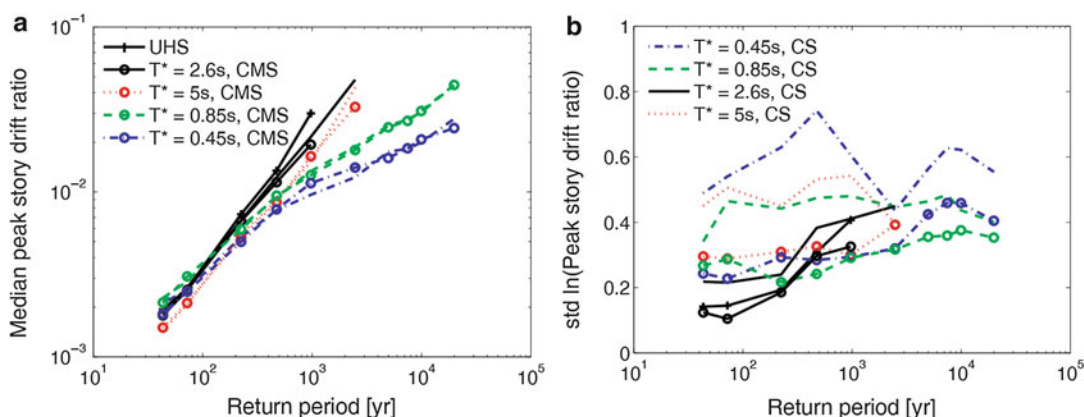
The risk-based assessments are performed several times, using ground motions selected and scaled to match conditional spectra, where the conditioning period used for these calculations is varied from 0.45 to 5.0 s (i.e., the building's third-mode structural period up to approximately twice the first-mode period). For each case, the risk-based assessment results are found to be similar (Fig. 4). The similarity of the results stems from the fact that the careful record selection ensures that the distributions of response spectra at all periods are nominally comparable, so the distribution of resulting structural responses should also be comparable (to the extent that response spectra describe the relationship between the ground motions and structural responses).

From these results, it is observed that if the analysis goal is to perform a risk-based assessment obtaining the mean annual frequency of exceeding an EDP, then one should be able to obtain an accurate result using any conditioning period, provided that the ground motions are selected carefully to ensure proper representation of spectral values and other ground motion

parameters of interest. Here “proper representation” refers to consistency with the site ground motion hazard curves at all relevant periods, and this is achieved by using the CS approach to determine target response spectra for the selected ground motions. The reproducibility of the risk-based assessment results, for varying conditioning periods, then results from the fact that the ground motion intensity measure used to link the ground motion hazard and the structural response is not an inherent physical part of the seismic reliability problem considered; it is only a useful link to decouple the hazard and structural analysis. If this link is maintained carefully then one should obtain a consistent prediction of the risk-based assessment in every case.

Intensity-Based Assessments and Evaluation of Alternative Target Spectra

Intensity-based assessment estimates structural response given ground motions with a specified intensity level. A study on the sensitivity of intensity-based assessment results to the choice of conditioning period when the CS is used as a target for ground motion selection and scaling is conducted by Lin et al. (2013c). This study also investigates the sensitivity of both risk-based and intensity-based assessments to the choice of target spectrum, including evaluation of the uniform



Conditional Spectra, Fig. 5 Intensity-based assessments of peak story drift ratio (median and logarithmic standard deviation in (a) and (b) respectively) of a

20-story reinforced concrete frame structure obtained from ground motions selected to match the CS, the CMS, and the UHS

hazard spectrum and the conditional mean spectrum.

The study shows the different effects of conditioning periods on intensity-based and risk-based assessments. In contrast to risk-based assessments, which are relatively insensitive to the choice of conditioning period in the CS, the choice of conditioning period for the CS can substantially impact structural response estimates for an intensity-based assessment (NIST 2011). It is therefore critical to specify the structural analysis objective clearly.

The study also demonstrates the importance of target spectrum. For intensity-based assessments, use of the CMS, instead of the CS, does not significantly affect the median response estimates (Fig. 5a) but does decrease both the dispersion of the response (Fig. 5b) and the probability of collapse, while use of the UHS typically results in higher median response (Fig. 5a). For risk-based assessments, use of the CMS, instead of the CS, results in underestimation of structural response hazard due to the omission of spectral variability (Fig. 4), while use of the UHS results in overestimation in the structural response hazard (Fig. 4).

An important issue regarding conditioning period arises when an intensity-based assessment is being used and the purpose is to compute the mean or median response associated with an $Sa(T^*)$ having a specified probability of

exceedance (e.g., for a building-code-type check). In this extremely common case, the response prediction will always change depending upon the choice of conditioning period. This comes from the fact that the choice of conditioning period is an inherent part of the problem statement, and so in this case changing the conditioning period changes the question that is being asked. For example, computing the median drift response for a building subjected to a 2 % in 50 year exceedance $Sa(1 \text{ s})$ is not the same as computing the median drift response for a building subjected to a 2 % in 50 year exceedance $Sa(2 \text{ s})$; these are two different questions. Resolution of this issue likely lies in identifying a conditioning period and performance check that, when passed, confirms satisfactory reliability of the structural system.

Summary

The conditional spectrum (CS) is a target response spectrum (with mean and variance) for ground motion selection to perform nonlinear dynamic analysis. The computation of the CS requires (1) input earthquake parameters of magnitude, distance, relevant seismological and site characteristics (e.g., fault type and soil type); (2) specification of a conditioning period; (3) ground motion prediction model(s) (GMPMs)

to estimate mean and standard deviation of logarithmic spectral accelerations; and (4) correlation model(s) to relate spectral accelerations between different periods of vibration. Refined computation of the CS incorporates multiple causal earthquakes and multiple ground motion prediction models. The level of complexity in computation increases from Methods 1 (single earthquake and single ground motion prediction model) to 4 (multiple earthquakes and multiple ground motion prediction models with corresponding deaggregation weights). Factors that affect the accuracy of the approximations include (1) the input causal earthquake parameters, (2) the GMPMs used, and (3) the GMPM deaggregation weights. Hence, exact calculation methods may be needed for locations with hazard contributions from multiple earthquake sources especially with multiple source types, and/or for sites with larger variation in predictions from various GMPMs (Lin et al. 2013a). An intermediate method, Method 2.5, was proposed by Carlton and Abrahamson (2014) as a simpler practical alternative to Method 3 in Lin et al. (2013a). Probabilistic seismic hazard deaggregation of GMPMs (Lin and Baker 2011), which computes the relative contribution of GMPMs to a given intensity level, facilitates the refined CS computation. Deaggregation weights, instead of logic-tree weights, of GMPMs were recommended when multiple GMPMs were used to compute the CS (e.g., Lin et al. 2013a; Carlton and Abrahamson 2014). Notable extensions of the CS concept include the CS for vertical motions (Gulerce and Abrahamson 2011) and the generalized conditional intensity measure (Bradley 2010).

The computation and use of the CS advances hazard-consistent ground motion selection for performance-based earthquake engineering. The automation via USGS hazard mapping interface and ground motion selection web-based algorithms enhances practical applications of the CS. Challenges in the CS remain in (1) the choice of the conditioning period, (2) the appropriate question to ask in design-orientated intensity-based assessment, (3) a larger number of distinct ground motions that are used at various intensity levels, and (4) issues related to scaling ground

motions. Progress has been made towards the effect of conditioning period in intensity- and risk-based assessments (Lin et al. 2013b, c; Carlton and Abrahamson 2014), the performance goal and design check via design spectrum calibration (Loth and Baker 2013), and a new method termed “adaptive incremental dynamic analysis” (AIDA) (Lin and Baker 2013) that allows overlap of ground motions across intensity levels. Research is ongoing to provide insights on the current engineering practice of ground motion scaling and the seismological counterpart in direct ground motion simulation.

Cross-References

- [Physics-Based Ground-Motion Simulation](#)
- [Probabilistic Seismic Hazard Models](#)
- [Selection of Ground Motions for Response History Analysis](#)
- [Stochastic Ground Motion Simulation](#)

References

- Abrahamson NA, Al Atik L (2010) Scenario spectra for design ground motions and risk calculation. In: 9th US National and 10th Canadian conference on earthquake engineering, Toronto, 12 p
- Abrahamson NA, Bozorgnia Y, Boore D, Atkinson G, Campbell K, Silva W, Chiou B, Idriss IM, Youngs R (2008) Comparisons of the NGA ground-motion relations. *Earthq Spectra* 24(1):45–66
- ASCE (2010) Minimum design loads for buildings and other structures. ASCE 7–10. American Society of Civil Engineers/Structural Engineering Institute, Reston
- Baker JW (2011) Conditional mean spectrum: tool for ground-motion selection. *J Struct Eng* 137(3):322–331
- Baker JW, Cornell CA (2006) Spectral shape, epsilon and record selection. *Earthq Eng Struct Dyn* 35(9):1077–1095
- Baker JW, Jayaram N (2008) Correlation of spectral acceleration values from NGA ground motion models. *Earthq Spectra* 24(1):299–317
- Bradley BA (2010) A generalized conditional intensity measure approach and holistic ground-motion selection. *Earthq Eng Struct Dyn* 39(12):1321–1342
- Bradley BA (2012) A ground motion selection algorithm based on the generalized conditional intensity measure approach. *Soil Dyn Earthq Eng* 40:48–61

- Carlton B, Abrahamson N (2014) Issues and approaches for implementing conditional mean spectra in practice. *Bull Seismol Soc Am* 104(1):503–512
- Cornell CA (1968) Engineering seismic risk analysis. *Bull Seismol Soc Am* 58(5):1583–1606
- Cornell CA, Krawinkler H (2000) Progress and challenges in seismic performance assessment. *PEER Center News* 3(2):1–3
- Gulerce Z, Abrahamson NA (2011) Site-specific design spectra for vertical ground motion. *Earthq Spectra* 27(4):1023–1047
- Haselton CB, Deierlein GG (2007) Assessing seismic collapse safety of modern reinforced concrete frame buildings. Technical report, 2007/08. Pacific Earthquake Engineering Research Center, University of California, Berkeley
- Jayaram N, Baker JW (2008) Statistical tests of the joint distribution of spectral acceleration values. *Bull Seismol Soc Am* 98(5):2213–2243
- Jayaram N, Lin T, Baker JW (2011) A computationally efficient ground-motion selection algorithm for matching a target response spectrum mean and variance. *Earthq Spectra* 27(3):797–815
- Katsanos EI, Sextos AG, Manolis GD (2010) Selection of earthquake ground motion records: a state-of-the-art review from a structural engineering perspective. *Soil Dyn Earthq Eng* 30(4):157–169
- Lin T, Baker JW (2011) Probabilistic seismic hazard deaggregation of ground motion prediction models. In: 5th international conference on earthquake geotechnical engineering, Santiago, 12 p
- Lin T, Baker JW (2013) Introducing adaptive incremental dynamic analysis: a new tool for linking ground motion selection and structural response assessment. In: 11th international conference on structural safety and reliability. CRC Press, New York, 8 p
- Lin T, Harmsen SC, Baker JW, Luco N (2013a) Conditional Spectrum computation incorporating multiple causal earthquakes and ground-motion prediction models. *Bull Seismol Soc Am* 103(2A):1103–1116
- Lin T, Haselton CB, Baker JW (2013b) Conditional spectrum-based ground motion selection. Part I: hazard consistency for risk-based assessments. *Earthq Eng Struct Dyn* 42(12):1847–1865
- Lin T, Haselton CB, Baker JW (2013c) Conditional spectrum-based ground motion selection. Part II: intensity-based assessments and evaluation of alternative target spectra. *Earthq Eng Struct Dyn* 42(12):1867–1884
- Loth C, Baker JW (2013) Reliability-based calibration of design seismic response spectra and structural acceptance criteria. In: 11th international conference on structural safety and reliability. CRC Press, New York, 8 p
- McGuire RK (1995) Probabilistic seismic hazard analysis and design earthquakes: closing the loop. *Bull Seismol Soc Am* 85(5):1275–1284
- Naeim F, Lew M (1995) On the use of design spectrum compatible time histories. *Earthq Spectra* 11(1):111–127
- NIST (2011) Selecting and scaling earthquake ground motions for performing response-history analyses. NIST GCR 11-917-15 (ATC 82). Prepared by the NEHRP Consultants Joint Venture for the National Institute of Standards and Technology, Gaithersburg, 256 p
- Wang G (2011) A ground motion selection and modification method capturing response spectrum characteristics and variability of scenario earthquakes. *Soil Dyn Earthq Eng* 31(4):611–625

D

Damage Detection in Built-Up Areas Using SAR Images

Lorenzo Bruzzone¹, Carlo Marin¹ and
Francesca Bovolo²

¹Department of Information Engineering and
Computer Science, University of Trento, Povo,
Trento, Italy

²Fondazione Bruno Kessler (FBK), Povo, Trento,
Italy

Synonyms

Collapsed building detection; Crisis manage-
ment; Synthetic aperture radar data

Part of this entry has been taken from: Marin, C., Bovolo, F., Bruzzone, L.: *Building Change Detection in Multitemporal Very High Resolution SAR Images*, has been accepted for publication to IEEE Transaction on Geoscience and Remote Sensing.

It is worth noting that the incidence angle can be different from the incidence angle of SM images.

Equations 4 and 5 hold assuming flat-roof buildings with rectangular layout and SAR images in slant-range geometry.

Vice-versa a new buildings (ω_{c2}) is identified if the order of appearance of the pair increased/decreased is inverted with respect to the range direction as described in section “Behavior of Changed Buildings in Multitemporal HR and VHR SAR Images”

Introduction

Earthquakes have struck cities around the globe causing injury, loss of life, and damage to human settlement since the ancient times. In the last years large earthquakes have been reported globally affecting wide geographical areas and a relevant number of people. A remarkable example was the 2011 Tohoku earthquake with moment magnitude scale (MMS) of 9.0. The earthquake caused 15,884 deaths, 6,147 injured, and 2,636 people missing across twenty prefectures, as well as 129,316 buildings totally collapsed, 263,845 buildings half collapsed, and 725,760 buildings partially damaged (World Bank 2012). The earthquake triggered tsunami waves that caused the nuclear accidents at three reactors in the Fukushima Daiichi Nuclear Power Plant complex and the associated evacuation zones affecting hundreds of thousands of residents (World Bank 2012). Moreover, the earthquake and tsunami also caused extensive and severe structural damage in northeastern Japan, including heavy damage to roads and railways, as well as fires in many areas, and a dam collapse. The World Bank has estimated an economic damage of USD 210 billion, making it the costliest natural disaster in the world history (World Bank 2012). Another relevant example is the 2009 L’Aquila earthquake (Italy) with 6.3 MMS. The earthquake caused damages to a relevant number of medieval buildings in the city of L’Aquila: several buildings collapsed, about 300 people died, and about

1,500 people were injured making this event the deadliest earthquake to hit Italy since 1980. The list of earthquakes and their damages that occurred globally can be continued. Indeed, according to the United States Geological Survey (available online: <http://earthquake.usgs.gov/>) records, the frequency of such catastrophic events is high: 332 earthquakes with an MMS of 7 or higher have been reported in the last 20 years.

In order to understand and possibly mitigate the impact of such events on the citizenry, several governmental initiatives and research activities have been carried out in the past (e.g., the international charter on space and major disasters (available online: <http://www.disasterscharter.org/>), UN-OSAT (available online: <http://www.unitar.org/unosat/>), or the geohazard supersites and natural laboratories (available online: <http://supersites.earthobservations.org/>) to tackle each of the phases of such events. In detail a catastrophic event can be partitioned in three characteristic phases: (1) before the event in which a system of risk assessment for early warning is essential, (2) the moment the event occurs in which a disaster-alerting system is the main priority, and (3) after the event in which an emergency response with impact assessment has to be initiated.

Focusing on the third phase, among all the possible damages caused by natural disasters, the building damages contribute mostly to the casualties (Dell'Acqua and Gamba 2012). Nonetheless, the identification of struck built-up areas may be particularly complex because of both the possible large extension of the affected region and the difficulties to access remote villages. In this scenario Earth observation (EO) data acquired by remote-sensing satellites represent a useful tool to support decisions of civil protection authorities after the event. Indeed, satellites equipped with optical or synthetic aperture radar (SAR) sensors can monitor large regions in a quick, synoptic, and uncensored manner. In particular, SAR systems present the advantage over optical sensors to be independent from sun illumination and to be relatively insensitive to atmospheric weather conditions. This ensures the data

availability during the crisis event even though the site affected by the crisis event is covered by clouds or if the satellite is passing during night.

Since 2008 a new generation of SAR missions (i.e., TerraSAR-X, TanDEM-X, and COSMO-SkyMed) has regularly acquired data over Earth. These systems can perform acquisitions using two high-resolution modes called StripMap (SM) and SpotLight (SL). SM is characterized by a wide coverage (around 1,500 Km²) with a high resolution (3–5 m), whereas SL is characterized by a moderate coverage (around 100 Km²) with a very high resolution (1 m). This means that the two SAR modes are complementary to each other and can be used for performing different duties during the emergency response phase. SM mode can be used for identifying the blocks of the city more hit by the earthquake, whereas the SL images can be used to identify detailed information about collapsed buildings. It is worth noting that, these SAR missions are daily populating a huge archive of multitemporal images. This archive can be employed to detect changes in the state of an object or phenomenon by jointly observing the data acquired at different times over the same geographical area. This process, which is known in the literature as change detection, has got an increase of interest from the scientific community in the last decade, and a relevant number of novel techniques were developed to achieve this goal. Nonetheless, SAR images with a resolution ranging from 5 m (referred in this work as high-resolution images) of the SM images up to 1 m (referred in this work as very high-resolution images) of the SL images are more heterogeneous than old-generation sensor images, which have resolution generally ranging from 10 to 30 m (Soergel et al. 2006; Brenner and Roessing 2008). In high-resolution (HR) and very high-resolution (VHR) images, objects that are considered homogeneous from a semantic point of view (e.g., buildings) show a signature that is inhomogeneous at high spatial resolution because of the scattering contributions from sub-objects (e.g., facade and roof of a building). Furthermore, on the one hand the side-looking illumination

required by SAR systems leads to phenomena such as layover, shadow, and multipath signals (Dong et al. 1997; Franceschetti et al. 2002), which are very marked in urban areas. On the other hand, the appearance of a ground object depends on radar system parameters (i.e., wavelength, polarization, pulse length, incidence angle, look direction, etc.), surface feature properties (e.g., dielectric constant), and environmental variables (e.g., groundwater content) (Xia and Henderson 1997). On top of these aspects, SAR images are corrupted by speckle noise.

All these factors, within the context of the multitemporal analysis, make the problem of the detection of changes complex. Due to the high resolution, a large set of possible changes with different semantic meanings and scales are detectable in multitemporal HR and VHR SAR images. In general, each change may be associated with the cause of the change itself (Bruzzone and Bovolo 2013) and may present an extension that varies from the single pixel to a relevant portion of the entire scene. For instance, it is possible to distinguish among changes due to the anthropogenic activity, the phenological evolution of the vegetation, and the natural disaster effects. Depending on the application, some of these may be of interest to the end users, whereas others may not (Bovolo et al. 2013). In addition, external factors, such as different content of water on the ground due to different weather conditions, may affect the local back-scattering behaviors at the two considered dates also in absence of any change. Therefore, the same object may show different values of back-scattering even though it is not affected by a relevant change. This invalidates the assumption that two SAR images acquired on the same geographical area at different times are similar to each other except for the presence of changes occurred on the ground, which is often considered for medium-resolution SAR images. Thus, the use of standard pixel-based change detection techniques (Bazi et al. 2005) is not possible as they would be affected by a large amount of false alarms. Given the complexity of the problem, also standard context-based techniques based on a local neighborhood

analysis (Inglada and Mercier 2007) would fail to solve the problem since they are not able to properly take into account the high geometrical detail of HR and VHR data. To achieve change detection in HR and VHR SAR images, the analysis between the multitemporal images should be performed at a higher conceptual level that models the source of change from the prospective of interactions with the incidence electromagnetic wave.

Numerous studies have been recently presented in the literature that deal with the problem of recognition of changes in urban areas using VHR SAR images (Matsuoka and Yamazaki 2004; Chini et al. 2009; Ehrlich et al. 2009; Brunner et al. 2010; Balz and Liao 2010; Dekker 2011; Dell'Acqua et al. 2011; Dong et al. 2011; Jin et al. 2011; Dong and Shan 2013; Poulain et al. 2011; Tao et al. 2012; Brett and Guida 2013). In detail, Matsuoka and Yamazaki (2004), Chini et al. (2009), Ehrlich et al. (2009), Brunner et al. (2010), Balz and Liao (2010), Dekker (2011), Dell'Acqua et al. (2011), Dong et al. (2011), Jin et al. (2011), and Dong and Shan (2013) are focused on the detection of earthquake damages, Poulain et al. (2011) are focused on the building databases' updating, and Tao et al. (2012) address the detection of changes due to the urban evolution. The strategies that are exploited in these works include post-event supervised analysis (Ehrlich et al. 2009; Balz and Liao 2010; Dong et al. 2011; Jin et al. 2011; Dong and Shan 2013), joint use of optical and SAR data (Chini et al. 2009; Brunner et al. 2010), joint use of LiDAR and SAR data (Tao et al. 2012), unsupervised detection of damages at the level of aggregated blocks (Matsuoka and Yamazaki 2004; Dekker 2011), or GIS polygons (Dell'Acqua et al. 2011). Despite the great interest, the only work that addresses the problem of building change detection using multitemporal VHR SAR data in an unsupervised way is presented in (Brett and Guida 2013). In detail, the authors present an approach to the detection of damaged structures in urban areas from VHR SAR images, which is based on the multitemporal detection of the double-bounce

line generated by the multiple backscattering between the wall of the building and the ground. If a double-bounce line appears (disappears) between two acquisitions, a new (destroyed) building is recognized. Double bounce is one of the salient primitives characterizing a building in SAR images. It is known from the literature that at least other two primitives characterize the building footprint: the layover and the shadow. Therefore, the partial modeling of the source of change used in (Brett and Guida 2013) may generate a relevant number of missed alarms. Furthermore, the double-bounce line of a building alone is not a reliable feature for the identification of buildings because in several cases it may be not visible (Ferro et al. 2011).

In this work we present an approach that exploits the characteristics of the two complementary acquisition modes (i.e., StripMap and SpotLight) to: (i) quickly and automatically identify the areas severely affected by a catastrophic event (i.e., hot spots), such as an earthquake, by analyzing images characterized by a large coverage and a medium to high geometrical resolution and (ii) analyze images characterized by very high geometrical resolution acquired over hot spots in order to detect collapsed buildings. The effectiveness of the proposed approach is demonstrated in experimental results obtained on COSMO-SkyMed multitemporal images acquired in 2009 on the city of L'Aquila (Italy) before and after the earthquake that hit the region.

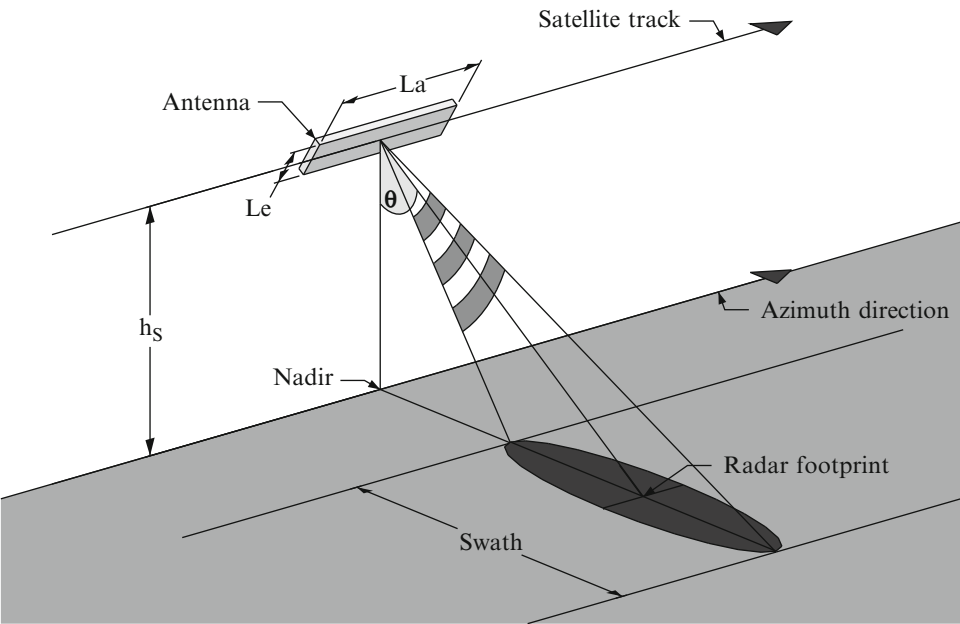
This entry is organized into five sections. Section “[Backscatter of Buildings in Synthetic Aperture Radar](#)” presents and reviews the fundamentals of the backscattering mechanism of buildings in monotemporal images. Moreover, it proposes an analysis of this mechanism for bitemporal SAR images, which is based on the concepts expressed for the monotemporal case. The proposed approach to building change detection is described in section “[Proposed Approach to Damage Detection](#).” Section “[Dataset Description and Experimental Results](#)” presents the dataset and shows the experimental results. Section “[Summary](#)” draws the conclusions of the work.

Backscatter of Buildings in Synthetic Aperture Radar

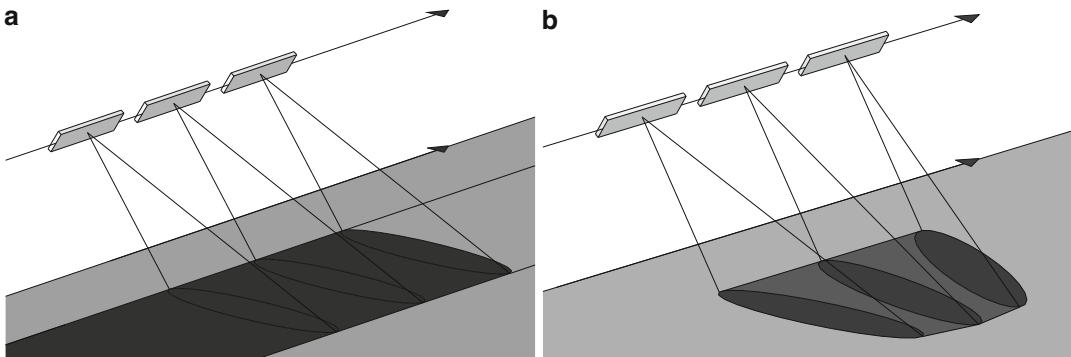
SAR is an active radar system operating in the microwave region of the electromagnetic spectrum generally between P band (i.e., 0.1–0.39 GHz) and Ka band (i.e., 26–40 GHz). The inherent characteristics of SAR measurements render this sensor sensitive to dielectric and roughness properties of the investigated natural media. The active operating mode makes this kind of sensor independent of solar illumination allowing day and night acquisitions. Moreover, the effect of clouds, fog, rain, smokes, etc., is mostly avoided due to the operation of the sensor in the microwave portion of the electromagnetic spectrum. SAR systems illuminate the scene using a side-looking geometry (see Fig. 1). The antenna of the radar system is mounted on a flying platform. Its horizontal and vertical axes are parallel and orthogonal to the azimuth direction, respectively. The angle between nadir and the radar beam direction is called incidence angle and it is usually denoted by θ . Such a system illuminates the Earth surface with microwave pulses and receives the electromagnetic signal backscattered from the illuminated scene. By means of signal processing techniques that exploit the Doppler shifts of the received electromagnetic echoes, SAR systems are able to synthesize a two-dimensional high-spatial resolution image from all the received signals.

The side-looking geometry of SAR together with non-flat terrain causes geometric distortions, such as foreshortening, layover, and shadow effects, which are visible as relatively bright and dark regions in SAR imagery, respectively. Furthermore, SAR data are affected by a characteristic noise called speckle, which causes the granular appearance of SAR imagery. This is the result of constructive and destructive interferences between the complex returns from the scatterers in a resolution cell.

Several sensor operation modes for acquiring SAR data were developed in the past. The most common modes which are implemented in spaceborne SAR missions are StripMap (SM) and SpotLight (SL). In the SM (see Fig. 2a) the



Damage Detection in Built-Up Areas Using SAR Images, Fig. 1 Side-looking geometry of SAR systems (Courtesy of Dominik Brunner)



Damage Detection in Built-Up Areas Using SAR Images, Fig. 2 SAR operation modes. (a) StripMap mode. (b) SpotLight mode (Courtesy of Dominik Brunner)

radar antenna points along a fixed direction with respect to the flight path. As the platform moves, the resulting image covers a strip of the ground surface. The extension of the imaged area in azimuth direction is limited only by the power and free memory available onboard of the satellite, whereas the resolution in range direction is constrained by half of the actual antenna length. Generally speaking for the new SAR generation, the standard coverage is of about 1,500 Km² with

a resolution that varies between 1.5 and 5 m in ground range and about 3 m in azimuth depending on the platform and acquisition parameters. In the SL mode (see Fig. 2b) the azimuth resolution is improved by increasing the length of the synthetic antenna. This is done by steering the antenna beam so that to illuminate the same area on the ground during the acquisition. Nonetheless, the improvement in the azimuth resolution is at the expense of the spatial

coverage. The standard coverage is in fact of about 100 Km^2 (instead of $1,500 \text{ Km}^2$) with a resolution that varies between 1 and 3.5 m (instead of 1.5 to 5 m) in ground range and 1 m (instead of 3 m) in azimuth depending on the platform.

In the rest of the section, we analyze the backscattering mechanism with respect to the variation of the acquisition geometry when isolated buildings with rectangular layout are sensed by HR or VHR SAR (section “[Building Backscattering Mechanism in Single-Detected HR and VHR SAR Images](#)”). This analysis is used to interpret multitemporal HR and VHR SAR images in order to study and model the effects associated with new/demolished building (section “[Behavior of Changed Buildings in Multitemporal HR and VHR SAR Images](#)”). This study gives the base for the development of the proposed approach.

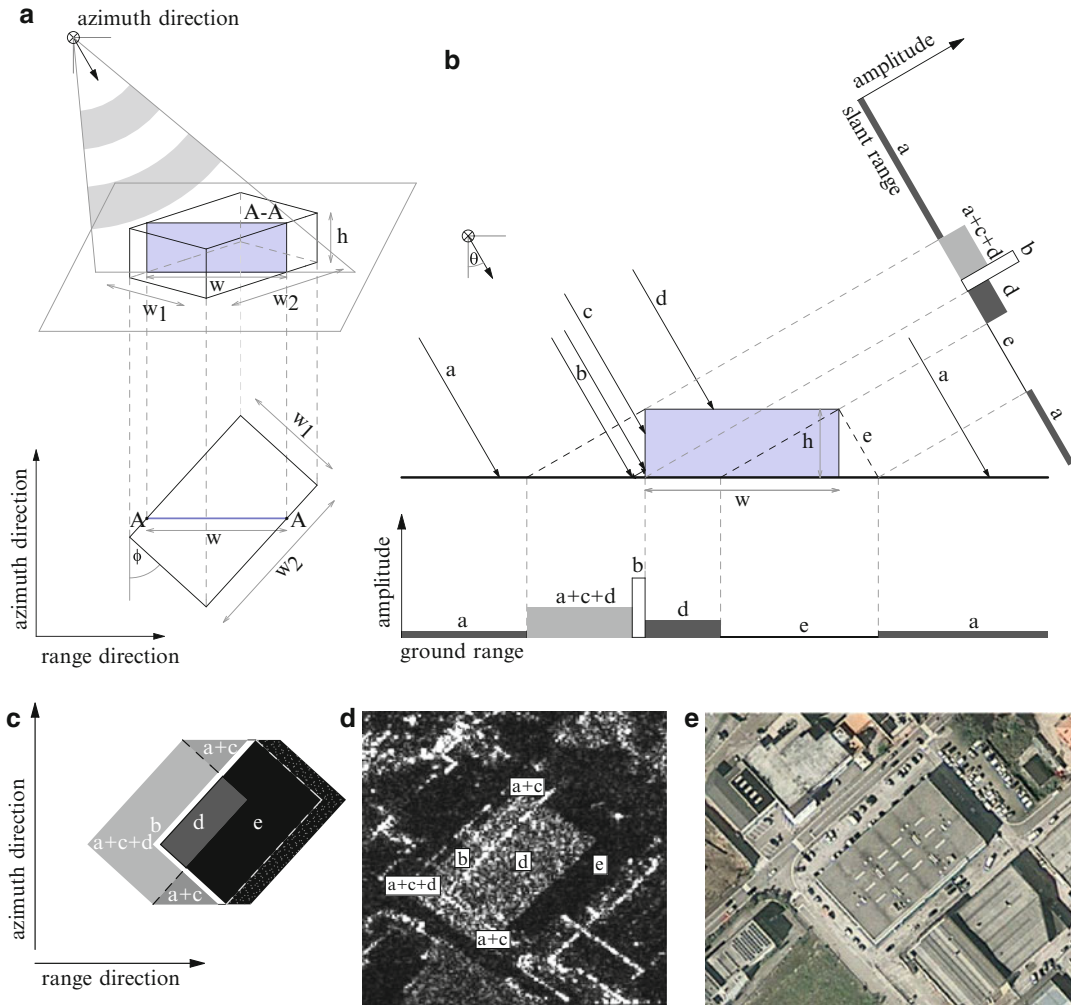
Building Backscattering Mechanism in Single-Detected HR and VHR SAR Images

In the literature several works have been presented that analyzed the scattering mechanism for different building models in order to derive their appearance in SAR images (Bennett and Blacknell 2003; Thiele et al. 2007; Brunner et al. 2010). These analyses make use of the geometrical optics (ray theory) approximation in order to model the electromagnetic scattering. This has the power to be intuitive and geometrically accurate even though the electromagnetic interactions and the transmission configurations of the antenna are not taken into account. In geometrical optics the wave propagation is described by rays, which are modeled as lines perpendicular to the wavefronts, that may be reflected, absorbed, or split at the interface between two different media. By knowing the geometry of acquisition of the SAR system and by exploiting trigonometric functions, it is possible to simulate the appearance of a building in HR and VHR SAR images.

Let us consider an isolated flat-roof building with dimensions $w_1 \times w_2 \times h$ illuminated by a SAR sensor that is moving along the azimuth direction and is illuminating the building from the left (Fig. 3a). Let us consider a building

model that does not take into account the building features such as windows, eaves, ridges, railings, and so on. This is equivalent to analyze the image at the scale level comparable with the building, and it allows us to derive the appearance of a building without losing generality. The acquisition geometry of a SAR system (Fig. 3a) is characterized by two parameters: the incidence angle θ (i.e., the angle defined by the incident radar beam and the normal to the intercepting surface) that is generally included between 20° and 55° and the aspect angle $0^\circ \leq \theta < 90^\circ$ (i.e., the angle between the azimuth direction and the orientation of an object in the horizontal plane). For the sake of argument, let us first assume that the SAR sensor illuminates the section A-A (light-blue area in Fig. 3a) of the building from a fixed position in the azimuth. Figure 3b shows in a qualitative way the amplitude of the backscattering projected in ground range and slant range of A-A. Following (Brunner et al. 2010) the main contributions that can be identified are the return a from the ground, the double-bounce effect b caused by the dihedral reflector formed by the building wall and the ground, the backscattering c from the front wall, the returns d from the building roof, and the shadow area e (see Fig. 3b). As one can observe the contributions from the ground, the wall and the roof are summed up into an area in front of the building ($a + c + d$), which appears brighter than each of these contributions taken singularly because of their superposition. This phenomenon, called layover, occurs when SAR sensors are imaging a surface with a slope steeper than the incidence angle θ , such as the wall of a building. It is worth noting that the length of the layover, the roof contribution, and the shadow depends on the width (w) and the height (h) of the section A-A and the incidence angle θ .

In order to derive the appearance of the whole building in a VHR SAR image, it is necessary to perform the aforementioned analysis for all the sections parallel to A-A that form the building. Nevertheless, as the width of each parallel section changes, the length of each mentioned contribution changes accordingly. Three cases can be observed on the basis of the relationship between



Damage Detection in Built-Up Areas Using SAR Images, Fig. 3 Example of backscattering mechanism for a flat-roof building. (a) Simplified model of a flat-roof building oriented with a given aspect angle θ . The SAR sensor is moving on the left of the building. Theoretical scattering model for a given incidence angle θ in the case the sensor is illuminating. (b) Only the section A-A of the building from a fixed position in the azimuth (the

backscattering is reported in both slant and ground range). (c) The whole extension of the building (the backscattering is in slant-range geometry). Different gray levels represent different amplitudes. Example of (d) a real flat-roof building in Trento (Italy) acquired at 1 m resolution by TerraSAR-X in slant range (SAR illumination is from the left courtesy of DLR). (e) The same building in an optical image (GeoEye – Google)

the width w of the section A-A and the limit value w_b defined as (Bennett and Blacknell 2003; Soergel et al. 2003):

$$w_b = \frac{h}{\tan \theta} \quad (1)$$

where h and θ are the height and the incidence angle of the building, respectively. Figure 3a

shows the case in which $w > w_b$. As the width of the section w decreases, the layover and shadow result larger, whereas the backscattering from the roof will be absorbed in the layover. This behavior is valid up to the limit condition $w = w_b$ for which the whole roof contribution is sensed in the layover area before the double bounce. In the case $w < w_b$ two contributions to layover areas can be distinguished: one due to the

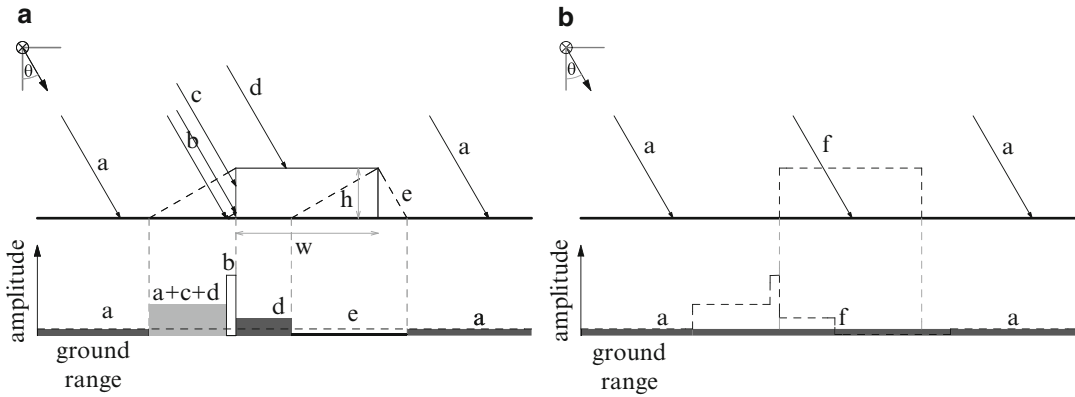
ground, the front wall, and the roof ($a + c + d$) and one due to the backscattering from the ground and the front wall of the building ($a + c$). Figure 3c shows the building radar footprint generated from the building of Fig. 3a. It is a convex polygon made up of (i) a bright L-shaped region due to the layover, (ii) a dark L-shaped region due to the shadow, (iii) a bright L-shaped line due to the double bounce, (iv) a rectangular region due to the direct return from the roof, and (v) two bright triangular regions due to the layover of only ground and wall ($a + c$). It is worth noting that since in the acquisition phase the radar is moving, the shadow casted by a ground object is moving as well resulting in a blurring of the border of the shadow (Zhang et al. 2013; Groen et al. 2009) (see the dotted area in Fig. 3c). By comparing Fig. 3c, d, it is possible to observe that the geometrical optics approximation can effectively describe the real behavior of scattering in VHR SAR images.

The same building may appear differently in SAR images according to the value of the aspect angle ϕ . When ϕ is approaching the limit values $\phi = 0^\circ$ and $\phi = 90^\circ$, the layover, the double bounce, and the shadow change by approaching a rectangular shape. A special attention has to be given to the double-bounce line with respect to the variation of ϕ (Franceschetti et al. 2002; Dong et al. 1997). In Ferro et al. (2011) an empirical study on the relationship between the strength of the double bounce and the aspect angle highlighted that the double-bounce contribution drops off significantly if the aspect angle increases from $\phi = 0^\circ$ up to 10° , whereas it decays moderately for higher angles. θ also affects the appearance of a building in VHR SAR images. By reverting Eq. 1, it is possible to derive a limit value θ_b that is equivalent to w_b . Nonetheless, in urban areas the choice of θ can be critical for two reasons: (i) the value of backscattering is related to θ , i.e., small value of θ generates higher value of backscattering and vice versa; and (ii) a large value of θ generates long shadow areas and small layover areas and vice versa. Thus, the backscatter has a dependence on the incidence angle, and there is potential for choosing optimum configurations for different applications.

The same analysis described in this section for flat-roof building can be conducted for other building models, e.g., gable-roof buildings (Brunner et al. 2010). The outcomes of such a study can be summarized as follows: the footprint of any type of isolated buildings with a rectangular base is given by a specific convex pattern made up of a bright area (due to layover and double-bounce effects) followed by a dark area (due to the shadow effect). These features may have different thickness, shape, and internal variability of backscattering on the basis of the considered building structure and material. Nevertheless, they will systematically arise when SAR systems are sensing an isolated building with an adequate resolution.

Behavior of Changed Buildings in Multitemporal HR and VHR SAR Images

The aim of this section is to analyze the behavior of the radar backscattering in multitemporal images when a building changes by taking into account the single data analysis carried out in the previous section. This analysis introduces the basic concept on which the proposed approach is based. In order to properly illustrate the problem, let us focus on a building that fully disappears between two acquisitions. Let us assume to sample this situation acquiring one image when (i) the building is standing and (ii) the building is totally dismissed. Let us assume to use a VHR SAR system configured with the same geometrical parameters (i.e., same incidence angle θ and same azimuth path) for the two acquisitions. As in section “Building Backscattering Mechanism in Single Detected HR and VHR SAR Images,” let us start the analysis by comparing the backscattering considering the illumination source fixed at a given point along the azimuth direction corresponding to section A-A (Fig. 1a) in the case the building is present. The backscattering profile obtained for the standing building is reported in Fig. 4a. As one can notice, the backscattering behavior is the same as the one obtained in the previous section and reported in Fig. 3b. Whereas for the case of totally dismissed building, by assuming bare ground as depicted in Fig. 4b, the value of backscattering is approximately



Damage Detection in Built-Up Areas Using SAR Images, Fig. 4 Building scattering mechanisms for a fixed illumination source corresponding to: (a) a section of a building and (b) bare ground. The comparison of Figures (a) and (b) describes the multitemporal scattering

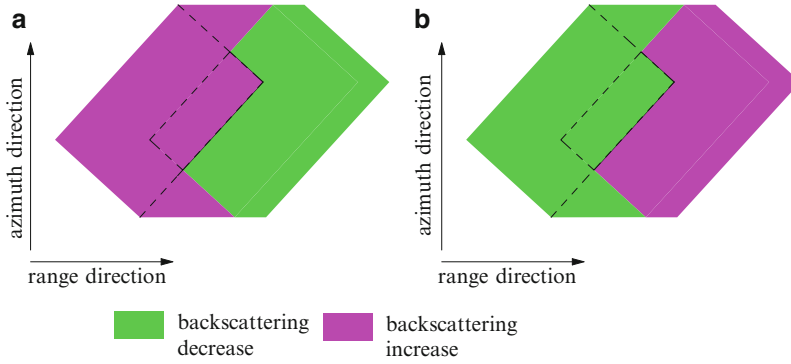
mechanism in the cases (a) a new building is built up (the envelope of backscattering of bare soil is reported for comparison in *dotted line*) and (b) a building is dismissed (the envelope of backscattering of the building in (a) is reported in *dotted line*)

constant. By comparing Fig. 4a, b, one can observe that the region with a high value of backscattering due to the layover $a + c + d$, the double bounce b , and the roof contribution d decreases its value when the building disappears, whereas the region hidden by the shadow e becomes visible to the line of sight of the radar and therefore increases its value (the dashed line in Fig. 4a represents the envelope of backscattering of bare ground). In other words, in the case of a new building we are likely to observe a structured pattern made up of two regions having increase and decrease of the backscattering oriented from near-to-far range. Vice versa if we consider the case in which a building disappears between two acquisitions, we expect that a structured pattern made up of two regions having decrease and increase in the backscattering values arises in near-to-far range (the dashed line in Fig. 4b represents the envelope of the backscattering amplitude of a building).

This specific multitemporal behavior obtained when the illumination source is fixed at a given position in azimuth can be used to retrieve the appearance of new/demolished buildings in multitemporal VHR SAR images by repeating the same analysis for all the sections parallel to A-A that form the scene. The obtained new/destroyed building radar footprint is made up of a pattern included in a convex polygon.

The pattern includes two regions that can be classified as: (i) area of increase of the value of backscattering and (ii) area of decrease in the value of backscattering. The order of appearance of these two regions along the near-to-far-range direction defines if the pattern is due to new buildings (see Fig. 5a), i.e., the increase area (in this work depicted conventionally in magenta) is closer to the sensor than the area of decrease (in this work depicted conventionally in green) or demolished building (see Fig. 5b), i.e., the decrease area is closer to the sensor than the increase area.

By taking into account the analysis done in section “Building Backscattering Mechanism in Single Detected HR and VHR SAR Images” on the radar signature of a general isolated building, it is possible to derive the appearance of the radar footprint of any type of newly built-up or destroyed buildings in VHR SAR multitemporal images. In detail, the change building radar footprint can be identified checking: (i) the presence of both the regions of increase and decrease in backscattering, (ii) the proportion between the areas (in pixels) of the regions of increase and decrease in backscattering (the proportion depends on the incidence angle of the acquisition), (iii) the equality between the lengths of the regions of increase and decrease in backscattering in the azimuth direction, and (iv) the alignment of the regions of increase and decrease in



Damage Detection in Built-Up Areas Using SAR Images, Fig. 5 Building scattering mechanisms in multitemporal VHR SAR images (illumination source on the *left*). Idealized map highlighting decrease and increase

in the value of backscattering (unchanged pixels are in white) obtained in the case of (a) new and (b) fully demolished building

backscattering with respect to the range direction, i.e., the barycenters of the two regions lay on the line with the range (as the regions in the model are not regularized, the barycenters were calculated by means of the geometric decomposition method). The order of appearance of the two regions of increase and decrease in backscattering value determines if the changed building radar footprint is due to a new or a destroyed building.

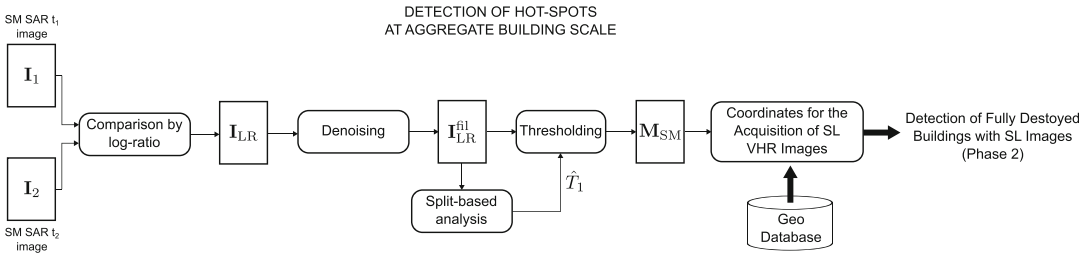
It is worth noting that the regions of change may have different thickness, shape, and variability of backscattering change on the basis of the considered building structure and acquisition geometry. Despite this variability, the pattern will systematically arise when SAR systems are sensing an isolated building with an adequate resolution (Bovolo et al. 2012). Hence, a method for the detection of the changed building based on this concept results to be robust to both the noise and the uncertainty (i.e., the impossibility of modeling the reality precisely) of multitemporal VHR SAR data.

Proposed Approach to Damage Detection

Let us assume to analyze two HR amplitude SAR images I_1 and I_2 acquired in StripMap mode before (pre-event) and after the event

(post-event), respectively, which have the same incidence angle and centered over the epicenter of the event. The goal is to detect the areas severely damaged by the event (i.e., hot spots) in order to quickly organize the rescue operations. The most critical issue for the identification of the damages is related to the presence of many possible kinds of changes on the ground. A single site may include many kinds of change that show significantly different characteristics in terms of size, shape, and semantic meaning. Note that not all the changes are associated to the built-up site, and thus, they are not relevant for the rescue operation. The analysis of SM image allows the detection of large portions of changed areas. Here we refer them as hot spots. The obtained hot spots are then exploited to drive the selection of a pre-crisis VHR SAR $X_1^1, \dots, X_1^Q$ and the acquisition of post-crisis VHR SAR images $X_2^1, \dots, X_2^Q$ taken with the same incidence angle. Note that the acquisition of the post-crisis images $X_2^1, \dots, X_2^Q$ should be scheduled on the basis of the detected hot spots. These images are used to perform change detection at high spatial resolution with the goal to detect collapsed buildings with a high level of detail.

Hence, the proposed approach is made up of two main phases: (i) a coarse detection of the areas of change (hot spots) and (ii) a fine detection of the fully collapsed buildings inside each hot spot. Figures 6 and 7 show the block schemes



From $\mathbf{I}_{LR}^{\text{fil}}$ it is possible to isolate the areas affected by a significant backscattering change and that present an extension equal or larger than the resolution of the selected wavelet level $N + 1$. Therefore, in order to detect the hot spots, the selected level has to be comparable to the scale of aggregated buildings' or city blocks' size. As the analysis of SAR images acquired in SM mode involves the processing of large multitemporal images, it is possible that the population of changed pixels is in sharp minority when considering $\mathbf{I}_{LR}^{\text{fil}}$. This may affect the accuracy of the adopted automatic threshold selection technique. In order to address this issue, we use the thresholding method presented in (Bovolo and Bruzzone 2007). This method divides the considered image into sub-images of a given size (splits) and performs the threshold selection considering only the splits that present the highest probability of containing changed splits. This selection allows the definition of a subset of splits in which the class of change shows a higher prior probability than in the whole image. For our purpose, $\mathbf{I}_{LR}$ is split into a set of S sub-images of a user-defined size. The choice of the split size $S_R^{\text{SM}} \times S_A^{\text{SM}}$ is driven by the average extension of the expected changes. Once the size S_R^{SM} and S_A^{SM} are defined, the splits are identified and selected according to their probability to contain a significant amount of changed pixels. The selection is done according to the value of the variance σ_s^2 , $s = 1, \dots, S$, computed on each split. This is a reasonable index for predicting the presence of changes in the log-ratio image since the residual multiplicative noise of the ratio image becomes additive due to the log operator. The desired set $\mathcal{P}_s$ of splits with the highest probability to contain changes is defined by selecting the splits that satisfy the following inequality:

$$\sigma_s^2 \geq \bar{\sigma}^2 + B\sigma_{\sigma^2}, \quad s = 1, \dots, S \quad (3)$$

where $\bar{\sigma}^2$ denotes the average variance of splits, σ_{σ^2} is the standard deviation of the variance of splits, and $B > 0$ is a constant. High values of B result in the selection of only those splits with high variance (i.e., the ones with the highest probability to contain changes); vice versa, low

values of B results in the selection of more splits. The split-based selection allows defining a subset of pixels $\mathbf{I}_s = \{\mathbf{I}_{LR} | \mathbf{I}_{LR} \in \mathcal{P}_s\}$ in which the classes of change show a higher prior probability than in the whole image. Consequently, the statistical estimation of the parameters related to the three probability density functions associated with no change, increase, and decrease of backscattering between the two dates (i.e., ξ_u , ξ^+ , ξ^-) can be correctly derived and used to separate the three classes. Here the thresholding method described in (Bazi et al. 2007) is adopted and the Bayes decision rule for minimum error is applied to separate the three classes. To this end a statistical model for class distributions is required together with an approach for class statistical parameter estimation. The Gaussian model and the well-known expectation-maximization (EM) algorithm (Dempster et al. 1977) are employed to derive $\mathbf{M}_{\text{SM}}$.

It is worth noting that $\mathbf{M}_{\text{SM}}$ does not report any information about the source of the changes. Indeed the changes may be associated to the crisis event as well to other source of change such as phenological changes. By exploiting prior information (e.g., cadastral map) and focusing on changes related to built-up areas, the hot spots possibly related to the crisis event that are in urban areas can be isolated. Nevertheless, among these changes some may be more critical than others. The severity of the change can be inferred considering the prior information available to civil protection or exploiting the intensity and texture information of the SAR images as proposed in (Dell'Acqua and Gamba 2012; Dekker 2011). Once the most relevant hot spots are detected, an algorithm for solving the "packing problem" of acquiring the minimum number of SL acquisitions $\mathbf{X}_2^1, \dots, \mathbf{X}_2^Q$ by covering the highest number of critical hot spots can be applied. Under the assumption the pre-event $\mathbf{X}_1^1, \dots, \mathbf{X}_1^Q$ images have been already acquired, when a crisis occurs the acquisition of $\mathbf{X}_1^1, \dots, \mathbf{X}_1^Q$ can be effectively planned (if not already done) by the outcome of this phase.

Detection of Fully Collapsed Buildings

The aim of the second phase is to detect collapsed buildings considering higher resolution data

acquired in SpotLight mode over the hot spots detected in the previous phase. This is the most challenging part of this work and it requires advanced techniques for change detection. For the sake of notation, and without any loss of validity in the following, we assume that $\varrho = 1$ and thus that critical hot spots can be fully covered by means of a single couple of images $\mathbf{X}_1$ and $\mathbf{X}_2$ acquired with the same incidence angle on the same geographical area at different times t_1 and t_2 , respectively. Let $\Omega = \{\omega_u, \Omega_c\}$ be the set of classes of changes to be identified: ω_u represents the class of unchanged pixels, whereas $\Omega_c = \{\omega_{c1}, \omega_{c2}, \dots, \omega_{cK}\}$ is a meta-class that gathers all the K possible classes (kinds) of change that may arise on the ground. One of the most critical issues dealing with this kind of scenario is related to the presence of many kinds of change on the ground. Nevertheless, in this work we are interested to investigate an urban area with the goal to detect collapsed buildings which usually represent the most critical kind of change in an earthquake emergency. In detail, we consider $K = 4$ classes of change: (i) fully destroyed buildings (ω_{c1}), (ii) new buildings (ω_{c2}), (iii) changes that have a size comparable to a building but do not present the typical pattern of full new/demolished buildings (ω_{c3}), and (iv) all the other changes that do not show a size comparable to a building and are therefore not related to changed buildings (ω_{c4}). In order to achieve this classification, we introduce an approach made up of two stages: (i) identification of the areas affected by changes in the backscattering at the scale of buildings and (ii) exploitation of the backscattering models presented in section “Behavior of Changed Buildings in Multitemporal HR and VHR SAR Images” in order to detect the classes $\Omega_c = \{\omega_{c1}, \omega_{c2}, \dots, \omega_{c4}\}$. Each stage of the proposed method for SL images analysis is explained in detail in the following. The block scheme is reported in Fig. 7.

As described in section “Behavior of Changed Buildings in Multitemporal HR and VHR SAR Images,” under the hypothesis of using a SAR sensor with a resolution comparable to the building size, the signature of isolated new/demolished buildings in multitemporal data

is given by a pattern made up of increase and decrease of backscattering regions. The first stage of the proposed approach is thus devoted to extract the areas that present both a significant increase or decrease in the backscattering value in SL (i.e., ξ_u, ξ^+, ξ^-). Nonetheless, at a resolution of a meter or less, the small objects that form the buildings such as windowsills or rain drains are visible. This results in an inhomogeneous signature of the building, which may affect also the appearance of new/destroyed buildings in multitemporal images rendering the regions of increase and decrease far from being homogeneous. It is worth noting that working at the full resolution of SM images (i.e., between 3 and 5 m) is also problematic due to the massive presence of noise. Therefore, we propose to work at an optimum scale level for representing buildings and not subparts of them, derived from the VHR images. This can be used to identify the regions of increase and decrease of backscattering, which characterize the buildings, with the advantages of (i) reducing the impact on the detection of small changes and thus reducing the false alarm rate and (ii) making it possible a mitigation of the speckle effect on the detection.

In order to work at the scale of a building, we propose to build a multiscale representation of the multitemporal information made up of N scale level. The $(N - 1)$ th resolution level, which represents the optimum scale level, is selected according to the minimum size of the building in the investigated scene. The multiscale representation can be obtained by applying several methods, e.g., Laplacian/Gaussian pyramid decomposition (Burt and Adelson 1983) or wavelet transform (Mallat 1989). Here the two-dimensional discrete stationary wavelet transform (2D-SWT) is exploited, which has the advantage of avoiding the down-sampling after each convolution step (Nason and Silverman 1995). In detail, we used the wavelet-based decomposition and reconstruction of the log-ratio image approach presented in (Bovolenta and Bruzzone 2005) for change detection in medium-resolution SAR images. The log-ratio image $\mathbf{X}_{LR}$, which is derived as $\mathbf{X}_{LR} = \log \mathbf{X}_2 / \mathbf{X}_1$, efficiently points out the differences in

backscattering values. From $\mathbf{X}_{LR}$ a set of multilevel images $\mathcal{X}_{MS} = \{\mathbf{X}_{LR}^0, \dots, \mathbf{X}_{LR}^n, \dots, \mathbf{X}_{LR}^{N-1}\}$ are computed, where the superscript n , $n = 0, \dots, N - 1$ indicates the resolution level. This is done in two steps: a decomposition phase in which $\mathbf{X}_{LR}$ is filtered through a cascade of N filters and a reconstruction phase in which only the information of interest is used to reconstruct the original image at the n th resolution level. According to (Bovolo and Bruzzone 2005), the decomposition is based on 2D-SWT, which applies to the considered image-appropriate level-dependent high- and low-pass filters at each resolution level n . Filter impulse response depends on the selected wavelet family. For each approximation sub-band $\mathbf{X}_{LR}^{LL_{n+1}}$, the inverse stationary wavelet transform (2D-ISWT) is applied $n + 1$ times in order to reconstruct in the image space the set $\mathcal{X}_{MS} = \{\mathbf{X}_{LR}^0, \dots, \mathbf{X}_{LR}^n, \dots, \mathbf{X}_{LR}^{N-1}\}$. For n ranging from 0 to $N - 1$, the images are characterized by resolution that is degraded

approximately by a factor of 2^n . Therefore, as expected by decreasing the resolution, small changes tend to disappear and only changes of a given size are fully preserved. More details on this strategy are given in (Bovolo and Bruzzone 2005).

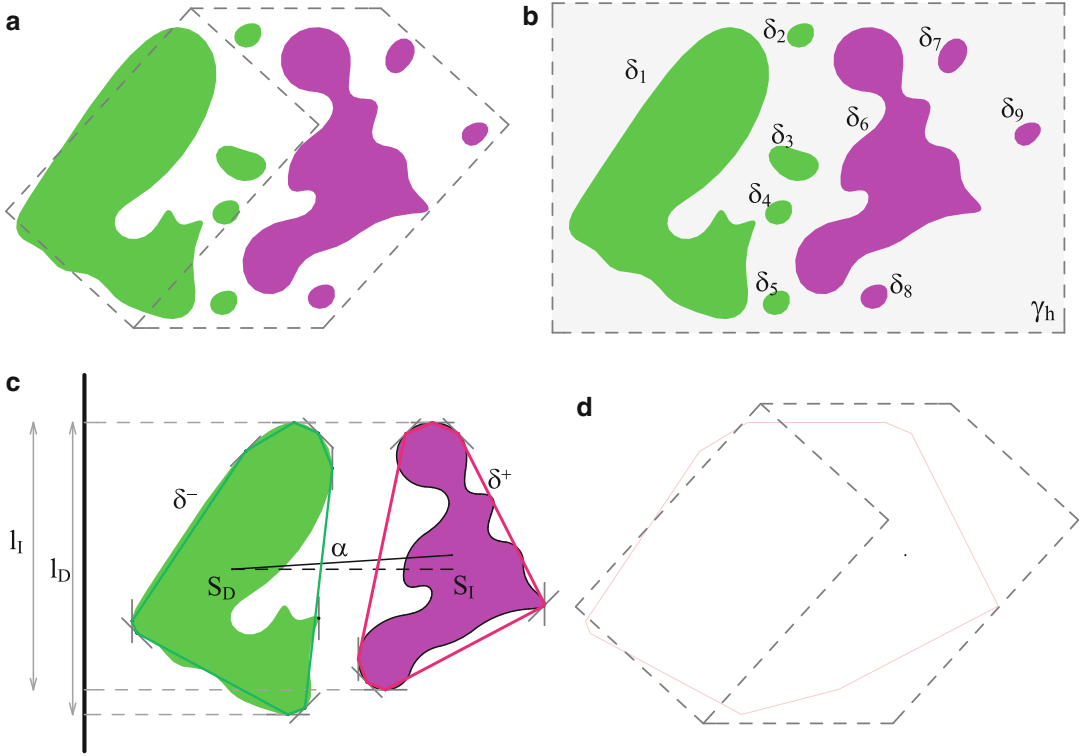
Once the optimum scale level (opt) has been selected, a change detection (CD) map $\mathbf{M}_{SL}^{opt}$ is derived from $\mathbf{X}_{LR}^{opt}$ according to the split-based thresholding procedure described in the previous section. Since we are considering changes related to buildings, we can determine S_R^{SL} and S_A^{SL} (i.e., the size of split for building analysis at the SL resolution) taking into account the average size of the radar footprint of the changed buildings in the considered radar image. This information can be inferred by considering the actual average size of the buildings, i.e., w_1 , w_2 , and h (see Fig. 3a) in the considered area, and converting it into SAR geometry by applying the following equations (Bennett and Blacknell 2003):

$$S_R^{SL} = \begin{cases} \frac{h}{\cos \theta} + w_1 \sin \theta, & \text{if } \theta \geq \tan^{-1}\left(\frac{h}{w_1}\right) \\ \left(w_1 + \frac{h}{\tan \theta}\right) \sin \theta + h \cos \theta & \text{Otherwise.} \end{cases} \quad (4)$$

$$S_A^{SL} = w_2 \quad (5)$$

Figure 8a reports an example of a change detection map $\mathbf{M}_{SL}^{opt}$ obtained with the proposed method. As one can notice, in this example the change detection map $\mathbf{M}_{SL}^{opt}$ verifies the assumptions defined for the model presented in section “Building Backscattering Mechanism in Single Detected HR and VHR SAR Images,” which are (i) the detected regions of increase or decrease in backscattering at the building scale are homogeneous and (ii) the change detection map presents a reduced number of changes smaller than the size of buildings, i.e., changes that are not related to changed buildings (ω_{c4}) are filtered out. Hence, from $\mathbf{M}_{SL}^{opt}$ it is possible to identify the building radar footprints of new or destroyed buildings by locating the pattern of increase (ξ^+) and decrease (ξ^-) of backscattering (building candidates) and

by evaluating the matching between the proprieties of the building candidates with respect to the proprieties of new and destroyed building models described in section “Behavior of Changed Buildings in Multitemporal HR and VHR SAR Images.” In detail, this task is based on two steps: first, the areas of change candidate to be associate to one of the classes ω_{c1} (demolished building), ω_{c3} (new buildings), or ω_{c3} (changes that have a size comparable to a building but do not present the typical pattern of full new/demolished buildings) are detected among all the backscattering changes highlighted in $\mathbf{M}_{SL}^{opt}$. The set of changed building candidates is denoted by $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_H\}$. Then the matching between the expected backscattering behavior of changed buildings (presented in section “Behavior of Changed Buildings in Multitemporal HR and VHR SAR Images”) and



Damage Detection in Built-Up Areas Using SAR Images, Fig. 8 Conceptual example of detection of a destroyed building (illumination source on the left). (a) M_{SL}^{opt} . The pixels associated to the classes of backscattering decrease (ξ^-) and increase (ξ^+) are reported in green and magenta, respectively. The idealized dismissed building radar footprint is drawn in dashed line. (b) In gray the bounding box of changed building candidate γ_h containing

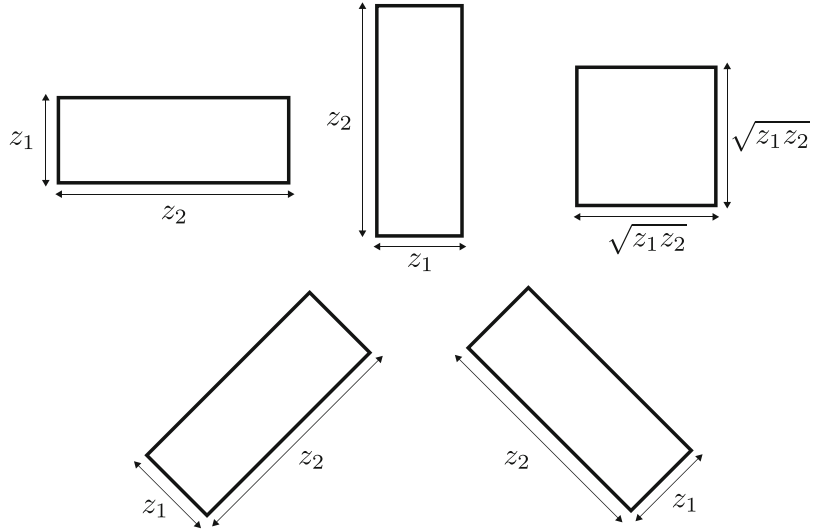
the regions of change $\delta_1, \delta_2, \dots, \delta_8$ that form the candidate. (c) Illustration of the parameters S_I, S_D, l_I, l_D , and α , which are evaluated by the fuzzy rules described in Table 1 in order to check the matching between the candidate and the expected pattern. (d) Identified demolished building footprint (red continuous line) compared with the actual footprint (gray dashed line)

the characteristics of the changed building candidates is evaluated. This is done by considering the physical proprieties and the relation of the regions of change inside each candidate, i.e., $\Delta = \{\delta_1, \delta_2, \dots, \delta_K\}$. In order to properly model the uncertainty inherent in the process (that can be due, for example, to cluttered objects placed in front of buildings), fuzzy theory is used (Zadeh 1965). In Fig. 8 an illustrative example that depicts the detection process of the proposed approach is reported for the case of a demolished building. This example will be used in the paper to better illustrate the proposed approach. Figure 8b–d illustrates the process of the proposed building change detection stage for the case of a demolished building. In the following each step is described in detail.

The first step of the detection of individual collapsed building phase aims at detecting the set of changed building candidates $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_H\}$. Ideally each changed building in M^{opt} is made up of regions of both increase and decrease in backscattering, whose total extension is comparable with the expected size of buildings. One simple and effective approach to identify the changed building candidates $\gamma_1, \dots, \gamma_H$ is to use a sliding window algorithm for each pixel in the image. Pixels within the moving window are taken and used to compute the detector output. In this work, the window is moved in M^{opt} from the left to the right by one pixel, and the number of changed pixels (labeled both as increase ξ^+ or decrease ξ^-) is counted. The process is repeated over five moving windows: four rectangular

Damage Detection in Built-Up Areas Using SAR Images,

Fig. 9 Windows used to derive the changed building candidates



moving windows showing different directions and a square window having the same area of the rectangular ones. This choice allows us to properly capture the most of the building orientations. The detector output is given by the window that results in the maximum value. The five windows are depicted in Fig. 9. The size of the windows $z_1 \times z_2$ is chosen according to the minimum size of the buildings in the considered scene in order to minimize missed alarms. Hence, let $\mathbf{M}_{\mathbf{W}_{i,j}^\beta}^{\text{opt}} = \left\{ \mathbf{M}^{\text{opt}} | \mathbf{M}^{\text{opt}} \subset \mathbf{W}_{i,j}^\beta \right\}$ be the set of pixels of $\mathbf{M}_{\text{SL}}^{\text{opt}}$ included in the window $\mathbf{W}_{i,j}^\beta$ centered at the pixel (i, j) $i = 1, \dots, I; j = 1, \dots, J$, with $\beta = 1, \dots, 5$ indicating the different windows. Let $\mathbf{C}$ be the image with size $I \times J$ that reports for each pixel $C_{i,j}$ an index of the size of the changes. $C_{i,j}$ is computed as the maximum on the five windows as follows:

$$C_{i,j} = \arg \max_{\beta \in 1, 2, \dots, 5} \left\{ \left| \mathbf{M}_{\mathbf{W}_{i,j}^\beta}^{\text{opt}} \in \zeta^+ \vee \mathbf{M}_{\mathbf{W}_{i,j}^\beta}^{\text{opt}} \in \zeta^- \right| \right\} \quad \forall \quad i = 1, \dots, I; \quad j = 1, \dots, J \quad (6)$$

where $|\cdot|$ indicates the cardinality of a set. The obtained index image $\mathbf{C}$ exhibits relatively high values when the sliding window contains a large amount of changes oriented in the same direction of the moving window. $\mathbf{C}$ exhibits relatively low

values when the sliding window contains a small amount of changes or the windows do not match the orientation of the change. Therefore, it is possible to detect the changed areas that can be associated in terms of size to the expected radar footprint of changed buildings by selecting the areas of $\mathbf{C}$ exceeding a certain threshold τ_C , thus obtaining the image $\overline{\mathbf{C}}$ as follows:

$$\overline{\mathbf{C}} \in \begin{cases} 1, & \text{if } \mathbf{C} \geq T_C \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

T_C is selected according to the expected minimum size of buildings $z_1 \times z_2$ in the investigated scene. The map $\overline{\mathbf{C}}$ represents the areas containing changes with size comparable or bigger than the minimum building size and therefore they belong to one of the classes $\omega_{c1}, \omega_{c2}, \omega_{c3}$. As one can notice, the pixels belonging to ω_{c4} are implicitly identified in this operation. From the map $\overline{\mathbf{C}}$ the set of changed building candidates $\Gamma = \{\gamma_1, \gamma_2, \dots, \gamma_H\}$ is extracted by calculating the connected components considering an 8-connected neighborhood. This is done by means of a flood-fill algorithm (Heckbert 1994). Each candidate γ_h ($h = 1, \dots, H$) contains a subset of regions $\Delta = \{\delta_1, \delta_2, \dots, \delta_K\}$ labeled as ξ_u, ξ^+ or ξ^- . The information associated to the kind of change is derived considering $\mathbf{M}_{\text{SL}}^{\text{opt}}$ in the areas delimited by the region γ_h . Figure 8b shows an example of

candidate γ_h that contains $\delta_k (k = 1, \dots, 8)$ regions for the case of a demolished building. In general the number of regions, k , may vary from 1 to K depending on the size and proximity of the changes. For the sake of clarity, in the next step of the individual collapsed building detection procedure, we assume that only two regions belonging to increase (i.e., $\delta^+ = \delta_k \in \xi^+$) or decrease (i.e., $\delta^- = \delta_k \in \xi^-$) occur inside a candidate. If more than two regions are present, all the combinations among the regions of increase and decrease are automatically analyzed by the proposed approach and the most reliable(s) selected.

In order to properly classify each collapsed building candidate according to the classes ω_{c1} , ω_{c2} , and ω_{c3} , the physical characteristics and the spatial arrangement of the pattern formed by the regions $\delta_k (h = 1, \dots, K)$ have to match with the four characteristics of the models of destroyed building discussed in section “[Behavior of Changed Buildings in Multitemporal HR and VHR SAR Images](#).” The matching is tested exploiting four fuzzy rules called here: (i) completeness, (ii) proportionality of areas, (iii) equivalence of lengths, and (iv) alignment. These rules aim at associating a grade of membership to the changed building candidate for each of specific performed test. On the basis of the aggregate membership, the final classification decision is taken. In the following each rule is presented in detail. It is worth noting that the theoretical model of a new building is just alike to the model of a destroyed building but the orientation of the pattern is reversed (see section “[Behavior of Changed Buildings in Multitemporal HR and VHR SAR Images](#)”). Therefore, the proposed method can detect collapsed buildings as well as new constructed buildings. Nonetheless, since the focus of this work is on the detection of collapsed building due to earthquakes, we concentrate in the following on this case.

Completeness

A pair of regions of increase and decrease is strictly needed to identify a destroyed building (ω_{c1}). The first rule evaluates the simultaneous

presence of δ^+ and δ^- inside the candidate γ_h . This is done by means of a crisp membership function μ_p that takes value 0 or 1 ($\mu_p(p) = \{0, 1\}$) depending on whether regions of increase and decrease are simultaneously present $\mu_p(p) = 1$ or not $\mu_p(p) = 0$.

Proportionality of Areas

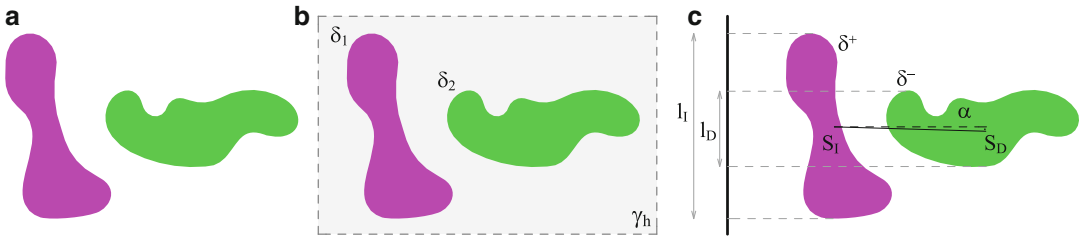
This rule aims at verifying that the area δ^+ is not prevailing with respect to the area of δ^- and vice versa. The mathematical modeling of this rule depends on the parameters of acquisition and the size of the building as discussed in section “[Building Backscattering Mechanism in Single Detected HR and VHR SAR Images](#).” It can be effectively represented by a sigmoid membership function that evaluates the attribute $r_s = \min\{s_I/s_D, s_D/s_I\}$, where s_I and s_D indicate the areas of δ^+ , i.e., $S_I = |\delta^+|$, and δ^- , i.e., $S_D = |\delta^-|$, respectively. The sigmoid function is defined as follows:

$$\mu_{r_s}(r_s) = \frac{1}{1 + e^{-a_1(r_s - b_1)}}. \quad (8)$$

The constant $a_1 > 0$ tunes the slope of function and the constant $b_1 > 0$ locates the center of the function. Equation 8 returns values in $[0, 1]$. Figure 11 shows an example of sigmoid function. As one can notice this is in accordance with what it is expected: the smaller r_s is, the smaller the membership grade of the candidate.

Equivalence of Lengths

This rule aims at checking that the lengths of the regions δ^+ and δ^- in the azimuth direction are equivalent. This is done by identifying the extrema of the regions δ^+ and δ^- . To test the reliability of a candidate with respect to this criterion, the attribute $r_l = \min\{l_I/l_D, l_D/l_I\}$ (where l_I and l_D are the lengths in azimuth direction of δ^+ and δ^- , respectively) is used in the sigmoid membership function defined in Eq. 8 with appropriate parameters $a_2 > 0$ and $b_2 > 0$. Since the farther r_l is from 1 the smaller the membership grade of the candidate to be associated to a new/destroyed building, the slope of the sigmoid is steep and the center moved toward 1. Figure 8c illustrates

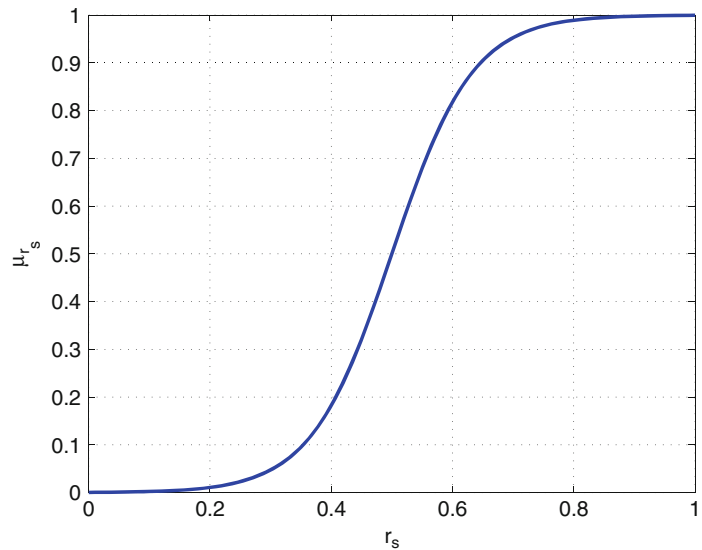


Damage Detection in Built-Up Areas Using SAR Images, Fig. 10 Example of a candidate γ_h that does not respect the equivalence of lengths. (a) M_{ST}^{opt} . The pixels associated to backscattering decrease (ζ^-) and increase (ζ^+) classes are reported in *green* and *magenta*,

respectively. (b) In *gray* the changed building candidate γ_h containing the regions δ_1 and δ_2 , which form the candidate. (c) Analysis of psychological proprieties and relations of the regions δ^+ and δ^-

Damage Detection in Built-Up Areas Using SAR Images,

Fig. 11 Example of sigmoid function $\mu_{r_s}(r_s)$ defined according to Eq. 8. The slope of the sigmoid is tuned by the parameter a_1 . b_1 defines instead the center of the sigmoid, i.e., $\mu_{r_s}(b_1) = 0.5$. In the represented case $a_1 = 15$ and $b_1 = 0.5$



an example of correct alignment between δ^+ and δ^- , whereas Fig. 10c depicts a case in which this rule is not satisfied.

Alignment

This rule is devoted to test the alignment of the entire pattern increase/decrease (which is made up of regions δ^+ and δ^-) with respect to the range direction. The alignment of the centroids of the regions of increase and decrease is considered, which is expected to lay on the same line. In detail, the absolute value of the angle α defined as the angle included by the line that connects the centroids of the regions δ^+ and δ^- and the SAR range direction is used in the equation of the

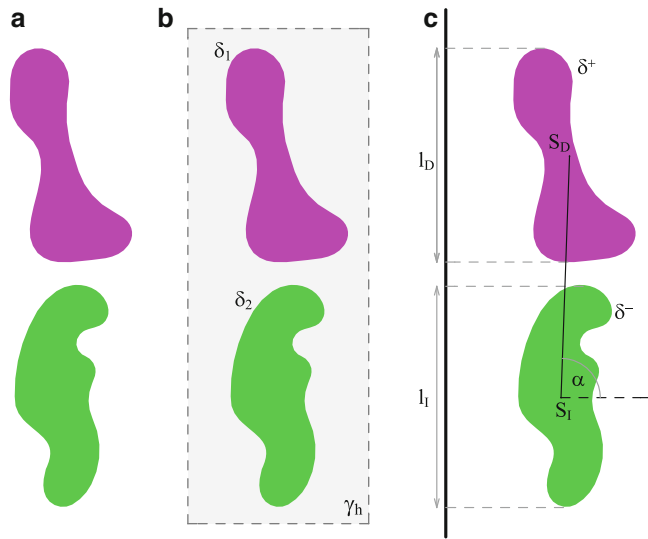
sigmoid Eq. 8. α ranges from $[-\pi/2, \pi/2)$. As expected, by imposing $a_3 < 0$ the larger is $|\alpha|$ the smaller is the membership grade of the candidate to be associated to a new/destroyed building. Figure 8c illustrates an example of correct orientation of the pattern increase/decrease, whereas Fig. 12c depicts a case in which this rule is not satisfied.

A summary reporting the description of each rule, the parameters, and the membership functions used to represent the rules is reported in Table 1.

The final decision on the label of the candidate is done according to the aggregated membership grade Y , which is calculated by combining the

Damage Detection in Built-Up Areas Using SAR Images,

Fig. 12 Example of a candidate γ_h that does not respect the fourth rule. (a) $\mathbf{M}_{SL}^{opt}$. The pixels associated to backscattering decrease (ζ^-) and increase (ζ^+) classes are reported in green and magenta, respectively. (b) In gray the changed building candidate γ_h containing the regions δ_1 and δ_2 , which form the candidate. (c) Analysis of psychical proprieties and relations of the regions δ^+ and δ^-



Damage Detection in Built-Up Areas Using SAR Images, Table 1 Fuzzy rules used to classify γ_h as new building (ω_{c1}), demolished buildings (ω_{c2}), or change comparable to the size of a building but not related to fully new/demolished buildings (ω_{c3})

Rule	Parameter	Membership function
1. Presence of both regions δ^+ and δ^- (completeness)	p	$\mu_p(p) = \{0, 1\}$
2. Proportionality of areas of δ^+ and δ^- (proportionality of the areas)	$r_s = \min\left(\frac{s_l}{s_D}, \frac{s_D}{s_l}\right)$	$\mu_{r_s}(r_s) = \frac{1}{1+e^{-a_1(r_s-b_1)}}$
3. Equivalence of the lengths of δ^+ and δ^- along the azimuth direction (equivalence of the lengths)	$r_l = \min\left(\frac{l_l}{l_D}, \frac{l_D}{l_l}\right)$	$\mu_{r_l}(r_l) = \frac{1}{1+e^{-a_2(r_l-b_2)}}$
4. Alignment of the barycenters of the regions δ^+ and δ^- (alignment)	$ \alpha $	$\mu_{ \alpha }(\alpha) = \frac{1}{1+e^{-a_3(\alpha -b_3)}}$

four membership functions. Among the fuzzy aggregation methods presented in the literature in this work, the Larsen product implication (Larsen 1980) is used, i.e.,

$$\mathbf{M} = \mu_p \mu_{r_s} \mu_{r_l} \mu_{|\alpha|} \quad (9)$$

In this way, one low value among the individual membership grades yields the final result to be low regardless of the other values. The candidates that present a value of $\mathbf{M}$ greater than a user-defined threshold T_M are selected as fully demolished (ω_{c1}) if δ^- is appearing before than δ^+ in range direction. The others are classified as ω_{c3} (i.e., they are not fully demolished or new buildings). It is worth noting that T_M will be a small value since the algebraic product of value smaller than 1 results in value smaller than the

factors in Eq. 9. Nonetheless, the value of $\mathbf{M}$ obtained for each of these classes is representing a grade of reliability of the detection that can be used by the final users.

For each changed building candidate recognized as demolished building, the 8-extrema points (i.e., from top left to bottom right) of both the regions δ^+ and δ^- are calculated and used to find the smallest convex polygon that contains the two regions, i.e., the convex hull (see Fig. 8d). The computation of the convex hull is done using the Quickhull algorithm (Barber et al. 1996). This allows a generation of an envelope of a dismissed building radar footprint as much as similar to the ideal expected one. Once the building footprint candidate is classified a demolished building, reverting the equations 4 and 5, it is possible to obtain the actual size of the

building affected by the change, which is expressed in terms of $w_1[m] \times w_1[m] \times h[m]$ (see Fig. 3a).

Selection of the Parameters

In this section the role of the parameters used in the proposed approach is discussed in detail. The parameters can be divided into three groups: the first two groups are related to the building size and the third to the fuzzy rules. The first group includes s_r and S_A , which represent the size of the split used for the automatic thresholding procedure of the log-ratio image at the optimal building scale. As discussed in the paper, the split size is selected according to the average size of the building in the considered area. It is worth noting that the split-based thresholding approach has demonstrated to be tolerant to relative large variations of the selected split size with respect to the expected average size of buildings (see section “Dataset Description and Experimental Results”). The second group of parameters includes $N - 1$, z_1 , z_2 , and T_C that are related to the minimum size of the buildings in the investigated site. In general, the minimum size of buildings is driven by the typology of urban area (industrial or residential) and the zone of the world under analysis (e.g., typically buildings in US metropolis are larger than buildings in European towns). Nonetheless, by selecting large values of $N - 1$, z_1 , z_2 , and T_C , building candidates with size smaller than the chosen values can be excluded from the analysis. The third group of parameters is associated to the membership functions of the fuzzy rules. In detail the center of the sigmoids (regulated by b_1, b_2, b_3) represents the boundary situation that has half membership with respect to the theoretical model described in section “Behavior of Changed Buildings in Multitemporal HR and VHR SAR Images,” while the slope (regulated by a_1, a_2, a_3) represents the flexibility of the rule to cope with clutter and noise. In the considered case the centers are fixed from the model, whereas the slope can be chosen considering real situations, e.g., urban density.

On the basis of this analysis and of our experimental results, we can conclude that the tuning

of the parameters to give as input to the proposed approach requires very basic prior information and is not critical.

Dataset Description and Experimental Results

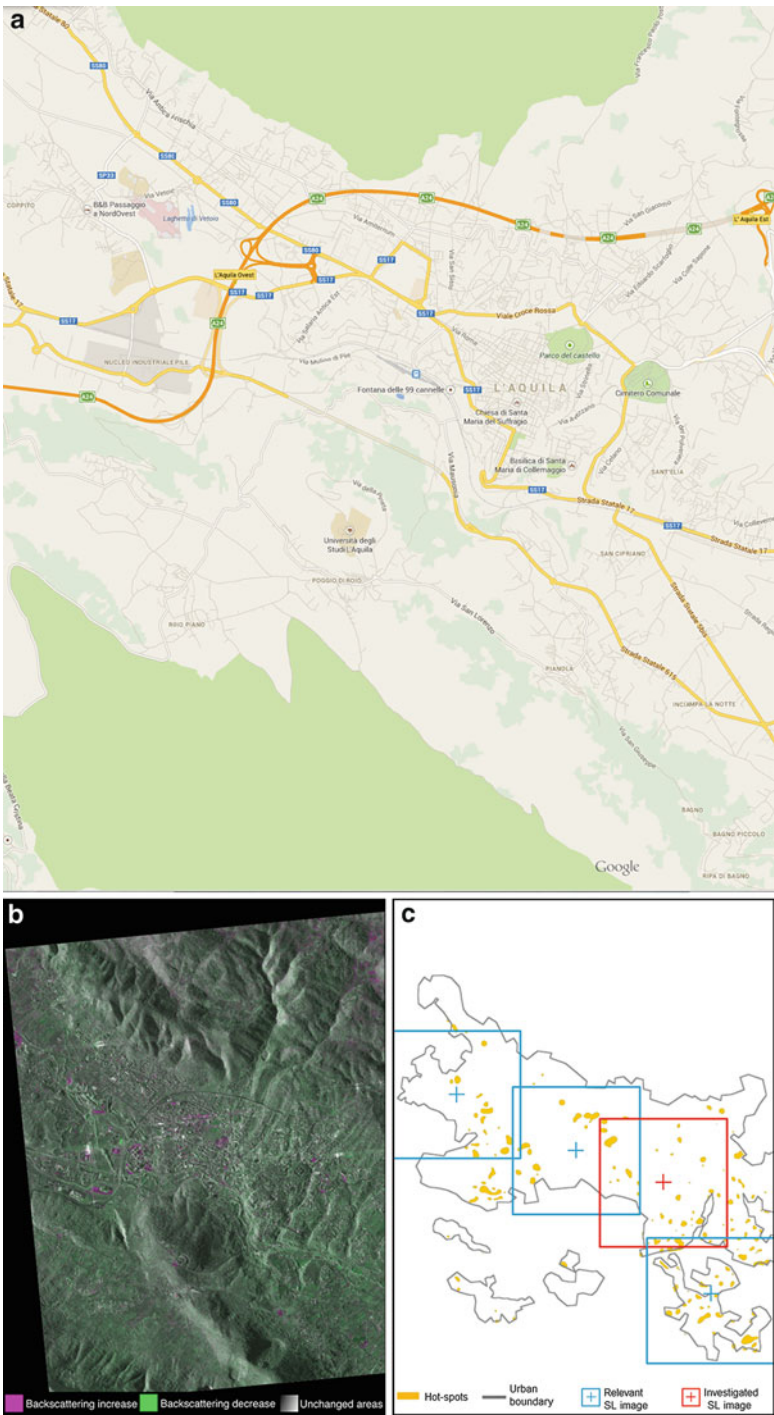
On April 6, 2009, an earthquake of 6.3 moment magnitude scale (MMS) struck central Italy with its epicenter near L’Aquila (42°21’N 13°24’E). The earthquake caused damages to a relevant number of buildings in the city of L’Aquila: several buildings collapsed, and about 1,500 people were injured, and about 300 people died, making this event the deadliest earthquake to hit Italy since the 1980 Irpinia earthquake.

At the time of the earthquake, no SM images were available on the area of the event. Nonetheless, two SL (1-look, 0.5 m $\times$ 0.5 m pixel spacing, X-band) amplitude images (X_1 and X_2) acquired in HH-polarization in ascending orbit with 57–58° incidence angle and processed according to the standard processing level 1C (geo-coded ellipsoid corrected) acquired on April 5, 2009 and April 21, 2009, respectively. This is due to the fact that the COSMO-SkyMed constellation was operational for few years at the time and that only two out of the four satellites were fully operational. Therefore, the COSMO-SkyMed archive was just started to be populated with images. In order to assess the effectiveness of the proposed method, the SL images were used to generate a couple of images with a degraded resolution (pixel spacing of 2.5 m) that can be assimilated to SM images (I_1 and I_2). This was done by using a nearest neighbor interpolation without multilooking operations and thus obtaining in images with size of 5,000 $\times$ 5,000. In the following, we refer to these images as SM images. Moreover, in order to be consistent with the real relationship between the size of SM and SL images, in the following we considered SL images with size five times smaller than the obtained SM images, i.e., 5,000 $\times$ 5,000.

Figure 13a shows the map of the area covered by the two SM images. Figure 13b shows a false color composition of the two SM images

Damage Detection in Built-Up Areas Using SAR Images,

Fig. 13 L'Aquila: (a) map of the area covered by the SM COSMO-SkyMed images (Google). (b) RGB multitemporal composition of SM COSMO-SkyMed images (*red channel*, 21/04/2009, *green channel*, 04/05/2009, *blue channel*, 21/04/2009). (c) Hot spots map M_{SM} . In *red* and *blue* the most damaged regions. The size of these regions corresponds to the coverage guarantee by SL images



(red channel, 21/04/2009, green channel, 04/05/2009, blue channel, 21/04/2009) in which pixels with an increase in the value of backscattering appear in magenta tone, pixels with a decrease in the value of backscattering appear in green tone, and unchanged pixel in gray scale. It is worth noting that a relevant number of changes are present between the two acquisitions (see Fig. 13b). This increases the complexity of the detection of collapsed buildings.

In order to apply the proposed method, the two SM images were calibrated and coregistered (with sub-pixel accuracy), and the log-ratio image $\mathbf{I}_{LR}$ calculated. $\mathbf{I}_{LR}$ was filtered by applying recursively a 2D-SWT transformation and reconstructing considering only the low-pass information at the 4th level. From $\mathbf{I}_{LR}^{fil}$ the hot spots of changes were derived by applying the split-based thresholding. The size of splits $S_R^{SM} \times S_A^{SM} = 64 \times 64$ pixels (corresponding to $160 \times 160 \text{ m}^2$ on the ground) was selected according to average size of expected changes in the test site. The obtained map is reported in Fig. 13c. From $\mathbf{M}_{SM}$ and the cadastral map $Q = 4$, areas with extension comparable to SL images were selected to be further analyzed in order to detect the presence of collapsed buildings. Among these areas we selected a representative case (red box in Fig. 13c) to be used in the following in order to effectively illustrate how the proposed method for the detection of collapsed buildings works. Indeed, the considered area presents a relevant number of collapsed buildings, differently from the other areas that present changes due to the construction of refugee camps.

As described in the methodological part, the proposed approach for the identification of collapsed buildings firstly calculates the log-ratio image $\mathbf{X}_{LR}$. This was computed from the two calibrated and coregistered CSK SL images $\mathbf{X}_1$ and $\mathbf{X}_2$. In detail, the coregistration was performed with sub-pixel accuracy, and the radar brightness calibration was applied in order to render the two images comparable. From $\mathbf{X}_{LR}$ the set $\mathbf{X}_{MS} = \{\mathbf{X}_{LR}^0, \dots, \mathbf{X}_{LR}^n, \dots, \mathbf{X}_{LR}^{N-1}\}$ was computed by applying sequentially the 2D-SWT and the 2D-ISWT with an 8-length *Daubechies* filter. The impulse response of low-pass

decomposition *Daubechies* filter of order 4 is given by the following coefficient set:

$$\{-0.0105974, 0.0328830, 0.0308414, -0.187035, -0.0279838, 0.630881, 0.714847, 0.230378\}$$

The finite impulse response of the high-pass filter for the decomposition step can be computed by satisfying the properties of the quadrature mirror filters. The level $(N - 1)\text{th} = 3\text{rd}$, which gives a resolution of approximately 8 m, was selected as the optimum one. This allows to observe radar footprints of buildings as small as 8 by 8 m, which is compatible with the minimum size of building footprints in the considered scene. $\mathbf{X}_{LR}^3$ was thresholded according to the unsupervised split-based procedure described in section “[Detection of Damages \(Hot-Spots\)](#).” The split size was calculated starting from the average building size, which was estimated to be $w_1 \times w_2 \times h = 25 \times 20 \times 15 \text{ m}^3$. By substituting these values in Eq. 4 S_{SR} was obtained to be equal to 49.50 m in slant-range geometry. Since the CSK images are ground projected, S_{SR} was projected in ground range considering the reference incidence angle, i.e., 58° obtaining in this way $S_{GR} = 58.36 \text{ m}$. From the calculation and taking into account a pixel spacing in a range of 0.5 m, S_R results to be equal to 120 pixels. By applying Eq. 5 and considering a pixel spacing in azimuth of 0.5 m, S_A was estimated to be equal to 40 pixels. It is worth noting that the split-based thresholding technique has a relative high tolerance to the selection of the values of S_R and S_A . This was proven by a sensitivity analysis in which S_A was ranged from 20 to 60 pixels and s_r from 80 to 170 pixels. Despite the large range (about $\pm 1/3$ of the optimal selected size), the error on the estimation of the thresholds is small. Moreover, in all the cases the generated maps did not present any critical behavior and resulted in a very similar final detection of changed buildings. The obtained CD map $\mathbf{M}^{\text{opt}}$ highlighting the three classes ξ_u, ξ^+, ξ^- is reported in Fig. 14c. The detector described by Eq. 6 was then applied to $\mathbf{M}^{\text{opt}}$ in order to obtain the image C. With regard to the considered scene, it is expected that the minimum size of building footprint is 40×20 pixels. Therefore, the set of

5 windows reported in Fig. 9 were set to size $z_1 \times z_2 = 40 \times 20$ pixels in order to derive $\mathbf{C}$ (see Fig. 14d). A set of changed areas $\Gamma = \{\gamma_1, \dots, \gamma_{38}\}$ compatible with building footprint changes were derived by thresholding $\mathbf{C}$ with $T_C = 160$ (which corresponds to 20 % of maximum value of $\mathbf{C}$) and by applying a flood-fill algorithm. It is worth noting that both the window size and the threshold value were selected according to minimum size of buildings in the

considered area in order to limit the missed alarms. Each of the candidate areas was analyzed in order to detect the changed buildings. To this end the four fuzzy rules were automatically evaluated for each candidate γ_h , ($h = 1, \dots, 38$). Finally, for each building candidate that presents an aggregated membership T_M greater than 0.125, the building radar footprint was approximated by the convex hull containing the candidate. It is worth noting that T_M was selected in

D



Damage Detection in Built-Up Areas Using SAR Images, Fig. 14 (continued)



Damage Detection in Built-Up Areas Using SAR Images, Fig. 14 L'Aquila (Italy) dataset: (a) orthophoto acquired few days after the earthquake on April 2009 (Geoportale Abruzzo) and (b) RGB multitemporal composition of SpotLight COSMO-SkyMed images (R, 21/04/2009, G, 04/05/2009, B, 21/04/2009). The images are geo-projected according to the reference ellipsoid.

Actual range and the azimuth directions are reported in white arrows. (c) Change detection map M_S^{pt} obtained at the level $N - 1 = 3$. (d) Map showing the index of the size of changes C . (e) Changed building map overlapped to the RGB multitemporal composition of CSK data. (f) Cadastral map of the area (destroyed buildings are reported in red)

order to minimize the missed alarms. This is done by taking into account the definition of the membership functions and the aggregation strategy (i.e., algebraic product). In the present experiment $\mu_{rs} = \mu_{rl} = \mu_a = 0.5$ were selected as limit case, which results in an aggregate membership function $M = 0.125$. Figure 14e shows the final result. In the specific case, all the reconstructed building radar footprints have an aggregate membership function greater than 0.615. Table 2a reports the settings of the parameters used for the dataset of L'Aquila.

By comparing the changed building map with the false color composition of the multitemporal images (Fig. 14e), one can notice that the proposed approach can effectively detect the backscattering changes associated to disappeared buildings as described in section “Behavior of Changed Buildings in Multitemporal HR and VHR SAR Images,” which patterns present a decrease and increase in backscattering oriented near-to-far range (see Fig. 14e). From a quantitative point of view, Table 3 reports the number of

Damage Detection in Built-Up Areas Using SAR Images, Table 2 Parameters used in the experiments carried out on L'Aquila dataset

Parameter	Value
$N - 1$	3
S_R	120 pixels
S_A	40 pixels
z_1	40 pixels
z_2	20 pixels
T_C	160 pixels
a_1, b_1	10, 0.3
a_2, b_2	10, 0.5
a_3, b_3	$-10, \frac{\pi}{3}$
T_M	> 0.125

Damage Detection in Built-Up Areas Using SAR Images, Table 3 Performance of the proposed approach for L'Aquila dataset in terms of number of detected destroyed buildings, missed alarms, and false alarms

Total number of buildings	Detected destroyed (ω_{c1})	Missed destroyed	False destroyed
200	6	0	1

Damage Detection in Built-Up Areas Using SAR Images,

Fig. 15 (a–d) Zoom of the 6 buildings destroyed and correctly detected by the proposed method (Bing Maps, Microsoft Corporation). See Fig. 14 (f) for the correspondence of the buildings on the map. As one can notice the building was totally demolished after the earthquake



building radar footprints of demolished building detected by the proposed approach. Even though the dataset presents a relevant number of changes in the backscattering value, none of the actual demolished buildings (see Fig. 15a–d) were misdetected, and one of the building was wrongly

identified as totally destroyed (see Fig. 15e) even though it was half collapsed on itself. Figure 14f shows the positions of the destroyed buildings in the cadastral map obtained by transforming the geometry of SAR images to the geo-referenced cadastral map. The analysis confirms the high

accuracy of the proposed technique. It is worth noting that even though the density of buildings in the considered site is high, the proposed approach can effectively discriminate between demolished and standing buildings. This points out the high robustness of the proposed technique to the possible deviations from the adopted ideal model of changes. A similar analysis has been performed for changes that are not associated with buildings, i.e., ω_{c3} (yellow areas in Fig. 14d). The orthophotos point out that these changes are not associated to fully destroyed buildings but to the post-earthquake activities, e.g., emergency housing units or to partially damaged buildings. This confirms the effectiveness of the proposed method also in the detection of changes different from demolished buildings.

Summary

In this entry an approach for the rapid and accurate detection of damages in built-up areas after a catastrophic event using SAR data has been presented. SAR systems onboard of new generation SAR missions are able to acquire images with a geometrical resolution that varies from 5 m up to 1 m using two different operational modes called StripMap and SpotLight. By increasing the resolution the maximum ground coverage guaranteed decrease accordingly. In this entry we have proposed to jointly exploit these two acquisition modes in order to: (i) quickly identify the areas severely affected by the crisis event (i.e., hot spots) by analyzing large areas using data characterized by a high geometrical resolution (3–5 m) and (ii) accurately detect collapsed buildings inside each hot spot by analyzing small areas using data with very high geometrical resolution (1 m).

In the first phase of the proposed approach, the hot spots, which are areas with both a significant increase or decrease in the backscattering value, are detected from the SM images. Since the goal during this phase is to quickly identify the areas severely affected by the crisis event and not perform detection at a fine resolution, a comparison between the SM images with a log-ratio operator

and a de-noising operation based on the wavelet transformation was applied in order to obtain the hot spots. By exploiting the available prior information about the scene, i.e., cadastral maps, it was possible to define the regions to be further analyzed in order to detect the collapsed buildings.

The analysis at the level of single building is conducted in the second phase of the proposed approach by taking advantage of the theoretical behavior of collapsed building in HR and VHR SAR images. Hence, from this modeling it is known that destroyed buildings are identified as a pattern made up of an area of increase and decrease of backscattering with specific physical proprieties and a specific alignment. In order to extract the changes associated with increase and decrease in backscattering in the SL images, the proposed approach makes use of a multiscale representation of the multitemporal information. This allows a detection of changes at the optimal building scale. This means that changes smaller than the building scale are rejected, while changes related to building size are made homogeneous. This information is used to identify the candidates to be destroyed buildings. The building candidates are analyzed in order to properly detect the collapsed building by means of four fuzzy rules. The fuzzy rules are formulated by taking into account the ideal backscattering mechanisms that arise when a building appears or disappears between two acquisitions. The aggregated membership resulting from the application of the fuzzy rules makes it possible to identify the class of each building candidate (i.e., collapsed building, new building, or general change with size comparable to the building size but not related to new or destroyed building). Moreover, the membership value gives an indication of the reliability of the detected class. The proposed approach requires the tuning of some parameters that depend on the approximated size of the buildings in the investigated area. This is a prior information usually available and easy to include in the processing. It is worth remarking that after this tuning the method is completely automatic.

The proposed approach was tested using SAR data related to the earthquake of L'Aquila (Italy) in 2009 acquired by COSMO-SkyMed constellation.

The proposed method demonstrated to be effective in detecting all the demolished buildings with a high accuracy. In detail, only one building was misclassified over 200 present in the two datasets with no missed detection. This was possible because the high robustness of the proposed method to deviations from the ideal model considered for changed buildings. This robustness is due to both the use of adequate fuzzy rules and the intrinsic reliability obtained when working on the comparison of two images rather than on the modeling of building in each image.

As future development of this work, we plan to test the proposed approach considering SAR dataset related to other earthquakes. Moreover, we plan to engineer the proposed approach to be compliant with the cogent requirements of real emergency scenarios.

Cross-References

- [Building Damage from Multi-resolution, Object-Based, Classification Techniques](#)
- [Earthquake Damage Assessment from VHR Data: Case Studies](#)
- [Estimation of Potential Seismic Damage in Urban Areas](#)
- [InSAR and A-InSAR: Theory](#)
- [SAR Images, Interpretation of](#)
- [SAR Tomography for 3D Reconstruction and Monitoring](#)
- [Urban Change Monitoring: Multi-temporal SAR Images](#)

References

- Balz T, Liao M (2010) Building-damage detection using post-seismic high-resolution SAR satellite data. *Int J Remote Sens* 31(13):3369–3391
- Barber CB, Dobkin DP, Huhdanpaa H (1996) The quickhull algorithm for convex hulls. *ACM Trans Math Softw* 22(4):469–483
- Bazi Y, Bruzzone L, Melgani F (2005) An unsupervised approach based on the generalized Gaussian model to automatic change detection in multitemporal SAR images. *IEEE Trans Geosci Remote Sens* 43(4):874–887
- Bazi Y, Bruzzone L, Melgani F (2007) Image thresholding based on the EM algorithm and the generalized Gaussian distribution. *Pattern Recogn* 40(2):619–634
- Bennett A, Blacknell D (2003) Infrastructure analysis from high resolution SAR and InSAR imagery. In: *Remote sensing and data fusion over urban areas*. 2nd GRSS/ISPRS joint workshop on, May 2003, pp 230–235
- Bovolo F, Bruzzone L (2005) A detail-preserving scale-driven approach to change detection in multitemporal SAR images. *IEEE Trans Geosci Remote Sens* 43(12):2963–2972
- Bovolo F, Bruzzone L (2007) A split-based approach to unsupervised change detection in large-size multitemporal images: application to tsunami-damage assessment. *IEEE Trans Geosci Remote Sens* 45(6):1658–1670
- Bovolo F, Marin C, Bruzzone L (2012) A novel approach to building change detection in very high resolution SAR images. In: *Proceedings of SPIE*, vol 8537, Bellingham, WA, pp 85 370Y–85 370Y–12
- Bovolo F, Marin C, Bruzzone L (2013) A hierarchical approach to change detection in very high resolution SAR images for surveillance applications. *IEEE Trans Geosci Remote Sens* 51(4):2042–2054
- Brenner A, Roessing L (2008) Radar imaging of urban areas by means of very high-resolution sar and interferometric sar. *IEEE Trans Geosci Remote Sens* 46(10):2971–2982
- Brett P, Guida R (2013) Earthquake damage detection in urban areas using curvilinear features. *IEEE Trans Geosci Remote Sens* 51(9):4877–4884
- Brunner D, Lemoine G, Bruzzone L (2010) Earthquake damage assessment of buildings using VHR optical and SAR imagery. *IEEE Trans Geosci Remote Sens* 48(5):2403–2420
- Bruzzone L, Bovolo F (2013) A novel framework for the design of change-detection systems for very-high-resolution remote sensing images. *Proc IEEE* 101(3):609–630
- Burt P, Adelson E (1983) The laplacian pyramid as a compact image code. *Commun IEEE Trans* 31(4):532–540
- Chini M, Pierdicca N, Emery W (2009) Exploiting SAR and VHR optical images to quantify damage caused by the 2003 Bam earthquake. *IEEE Trans Geosci Remote Sens* 47(1):145–152
- Dekker R (2011) High-resolution radar damage assessment after the earthquake in Haiti on 12 January 2010. *IEEE J Sel Top Appl Earth Obs Remote Sens* 4(4):960–970
- Dell’Acqua F, Gamba P (2012) Remote sensing and earthquake damage assessment: experiences, limits, and perspectives. *Proc IEEE* 100(10):2876–2890
- Dell’Acqua F, Bignami C, Chini M, Lisini G, Polli D, Stramondo S (2011) Earthquake damages rapid mapping by satellite remote sensing data: L’Aquila April 6th, 2009 event. *IEEE J Sel Topics Appl Earth Obs Remote Sens* 4(4):935–943

- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B* 39(1):1–38
- Dong L, Shan J (2013) A comprehensive review of earthquake-induced building damage detection with remote sensing techniques. *ISPRS J Photogramm Remote Sens* 84:85–99
- Dong Y, Forster B, Ticehurst C (1997) Radar backscatter analysis for urban environments. *Int J Remote Sens* 18(6):1351–1364
- Dong Y, Li Q, Dou A, Wang X (2011) Extracting damages caused by the 2008 Ms 8.0 Wenchuan Earthquake from SAR remote sensing data. *J Asian Earth Sci* 40(4):907–914, the 2008 Wenchuan Earthquake, China and Active Tectonics of Asia
- Ehrlich D, Guo H, Molch K, Ma J, Pesaresi M (2009) Identifying damage caused by the 2008 Wenchuan earthquake from VHR remote sensing data. *Int J Digit Earth* 2(4):309–326
- Ferro A, Brunner D, Bruzzone L, Lemoine G (2011) On the relationship between double bounce and the orientation of buildings in VHR SAR images. *IEEE Geosci Remote Sens Lett* 8(4):612–616
- Franceschetti G, Iodice A, Riccio D (2002) A canonical problem in electromagnetic backscattering from buildings. *IEEE Trans Geosci Remote Sens* 40(8):1787–1801
- Groen J, Hansen R, Callow H, Sabel J, Sabo T (2009) Shadow enhancement in synthetic aperture sonar using fixed focusing. *Ocean Eng IEEE J* 34(3):269–284
- Heckbert P (1994) Graphics gems IV, ser. Graphics gems series. AP professional
- Inglada J, Mercier G (2007) A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis. *Geosci Remote Sens IEEE Trans* 45(5):1432–1445
- Jin D, Wang X, Dou A, Dong Y (2011) Post-earthquake building damage assessment in Yushu using airborne SAR imagery. *Earthq Sci* 24(5):463–473
- Larsen PM (1980) Industrial applications of fuzzy logic control. *Int J Man–Mach Stud* 12(1):3–10
- Mallat S (1989) A theory for multiresolution signal decomposition: the wavelet representation. *Pattern Anal Mach Intell IEEE Trans* 11(7):674–693
- Marin C, Bruzzone L, Bovolo F (2014) Rapid and accurate damage detection in built-up areas combining stripmap and spotlight SAR images. In: Geoscience and remote sensing symposium (IGARSS), 2014 Québec City, Canada
- Matsuoka M, Yamazaki F (2004) Use of satellite SAR intensity imagery for detecting building areas damaged due to earthquakes. *Earthq Spectra* 20: 975–994
- Nason G, Silverman B (1995) The stationary wavelet transform and some statistical applications. In: Antoniadis A, Oppenheim G (eds) *Wavelets and statistics*, ser. Lecture notes in statistics, vol 103. Springer, New York, pp 281–299
- Poulain V, Inglada J, Spigai M, Tourneret J-Y, Marthon P (2011) High-resolution optical and SAR image fusion for building database updating. *IEEE Trans Geosci Remote Sens* 49(8):2900–2910
- Rignot E, van Zyl J (1993) Change detection techniques for ERS-1 SAR data. *IEEE Trans Geosci Remote Sens* 31(4):896–906
- Soergel U, Thönnessen U, Stilla U (2003) Iterative building reconstruction from multi-aspect insar data. In: Maas HG, Vosselmann G, Wayne Streilein J (eds) *3-D reconstruction from airborne laserscanner and InSAR data*. International archives of Photogrammetry and Remote Sensing, vol 34. Institute of Photogrammetry and Remote Sensing, Dresden, Germany, pp 186–192
- Soergel U, Thönnessen U, Brenner A, Stilla U (2006) High-resolution sar data: new opportunities and challenges for the analysis of urban areas. *Radar Sonar Navig IEE Proc* 153(3):294–300
- Tao J, Auer S, Reinartz P (2012) Detecting changes between a dsm and a high resolution sar image with the support of simulation based separation of urban scenes. In: *Synthetic aperture radar, 2012. EUSAR. 9th European conference on, 2012, Nuernberg, Germany*, pp 95–98
- Thiele A, Cadario E, Schulz K, Thönnessen U, Soergel U (2007) Building recognition from multi-aspect high-resolution InSAR data in urban areas. *IEEE Trans Geosci Remote Sens* 45(11):3583–3593
- World Bank (2012) The great east Japan earthquake learning from megadisasters: knowledge notes, executive summary. In: *EAP DRM knowledge notes*
- Xia Z-G, Henderson F (1997) Understanding the relationships between radar response patterns and the bio- and geophysical parameters of urban areas. *IEEE Trans Geosci Remote Sens* 35(1):93–101
- Zadeh L (1965) Fuzzy sets. *Inf Control* 8(3):338–353
- Zhang Y, Chen H, Ding C, Wang H (2013) A novel approach for shadow enhancement in high-resolution SAR images using the height-variant phase compensation algorithm. *IEEE Geosci Remote Sens Lett* 10(1):189–193

Damage to Ancient Buildings from Earthquakes

Miklós Kázmér

Department of Palaeontology, Eötvös University, Budapest, Hungary

Synonyms

Archeoseismology

Introduction

Ever since man-made structures have been erected, earthquakes have left their marks on these constructions. The study of earthquake damage can contribute to multidisciplinary efforts to assess parameters of ancient earthquakes from archeological evidence (archeoseismology). Initially these studies were aimed at enriching earthquake catalogs; however, recent interest has been targeted to provide quantitative earthquake parameters and describe site effects (Galadini et al. 2006; Sintubin 2011).

Yet, there is widespread skepticism about whether structural damage to man-made constructions, displaced structures, and indications of repair can be used as earthquake indicators. Damage to ancient buildings is often hard to identify, even for experienced archeologists and engineers (Hinzen 2011). Here we offer a brief overview of earthquake-induced damages observed in ancient structures.

The fabric of the building (masonry, brick, rubble infill between retaining walls, built with or without mortar) has to be thoroughly understood before the analysis and interpretation of archeoseismological data can be undertaken (Rodríguez-Pascua et al. 2011). Each building material behaves differently under seismic loading; for example, dressed stone walls tend to be rather rigid, while rubble walls have limited elastic properties.

Following the seminal works of Karcz and Kafri (1978) and Stiros (1996), a multitude of studies have listed and illustrated supposed seismic-induced damage of ancient buildings, mostly from the Mediterranean. Korjenkov and Mazor (2003) and Marco (2008) discussed proven or supposed seismic origin of a rich variety of damage features, arranging them in groups of sliding and shifting blocks, fallen columns, chipped block corners, and fractured and deformed walls and floors. Rodríguez-Pascua and his team (2011) established a comprehensive classification of earthquake archeological effects, of which building damages form the most significant part. However, in-depth analysis, including physical or numerical modeling, is only available for just a

few types of structural damage (shifted blocks: Vasconcelos et al. 2006; dropped keystone: Kamai and Hatzor 2008; toppled columns: Hinzen 2009; Yagoda-Biran and Hatzor 2010; for a review, see Hinzen et al. 2011). Complex systems for classifying damage are available (e.g., the genetic approach of Rodríguez-Pascua et al. 2011), but here a simple, descriptive system is used based on the shape of the damage and on whether it involves a single block, multiple blocks, a single wall, adjacent walls, or a whole building.

Typology of Earthquake-Induced Damages in Ancient Buildings

Archeoseismology is far from being a science with established methods. However, a few considerations taken into account will help the researcher to recognize, analyze, and interpret earthquake damage in proper context. (1) The older a building, the higher is the chance it has been damaged by seismic activity. (2) Constructions made of meticulously dressed stone provide better records than coarsely worked stone. (3) Buildings neglected since an earthquake offer better record than repaired ones; however, the repair itself can be a source of information on previous structural damage. (4) The history of successive restoration and architectural modification must be understood before seismic analysis is undertaken. In general, ancient structures show no difference in deformation styles than those observed on masonry buildings due to modern earthquakes (Galadini et al. 2006).

Furthermore, finding only one piece of evidence cannot be considered conclusive (Stiros 1996; Mazor and Korjenkov 2001). The identification of different types of deformations, or the repetition of a certain deformation on numerous edifices of a settlement, can be considered as consistent with the occurrence of an earthquake, once other possible natural causes have been excluded (Galadini et al. 2006).

Damage Affecting a Single or Multiple Blocks

The vertical component of a seismic wave moves masonry blocks (ashlars or columns) rapidly up



Damage to Ancient Buildings from Earthquakes,

Fig. 1 Typology of earthquake-induced damages in ancient buildings. (a) Cracks/fractures within a block (second century AD Palmyra theater, Syria). Photo Kázmér #4281. (b) Chipped edges of column and capital in the sixth-century Euphrasius cathedral, (Poreč, Croatia), damaged by the 1440 earthquake #0080. (c) Through-cutting fractures (repaired) in the eight- to ninth-century Brahma temple of the Prambanan complex, Yogyakarta, Indonesia, caused by the 2006 earthquake #6488. (d) Horizontal shift of large ashlars; vertical joints opened

up between them. Shiva temple, Prambanan complex, Yogyakarta, Indonesia. Tape measure extended 20 cm #6498. (e) Dropped ashlars in a Roman arch (Damascus, Syria) #2017. (f) Broken lintels in twelfth-century al-Marqab citadel, coastal Syria #4663. (g) Clockwise-rotated blocks in twelfth-century al-Marqab citadel, coastal Syria #5168. (h) Displaced drums of masonry columns of the fifth-century BC Hephaisteion temple, Athens, Greece #1132. (i) Extruded portion of house wall. Bosra, Syria #4183

and down. The hammering effect of the upper block on the lower one yields cracks in either or both blocks (Fig. 1a). Penetrating cracks allow a chip, corner, or edge to be separated from the block (Fig. 1b). The direction of the crack is influenced by the lithology of the rock and shape of the block and by direction of hammering. Chipping is not to be confused with rounding

or peeling due to weathering as seismically broken chips have sharp edges.

Fractures cutting through two or more blocks are often oriented close to vertical, this being the weakest plane to resist bending forces (Fig. 1c). Fractures through individual blocks are more or less connected to each other, resulting from sudden seismic shaking.

Damage Affecting a Single Wall

Where blocks are arranged as walls, two categories of failure occur: *in-plane* failures, where loads act in the plane of the wall, and *out-of-plane* failures, where loads act at an angle to the wall's face.

The most common in-plane failures are *gaps* between shifted blocks (Fig. 1d), produced by lasting vibrations acting parallel to the wall. Cracks can develop across single or multiple blocks, even across whole walls. Brick walls display *diagonal cracks* near doors and windows. Masonry walls – dressed stones being significantly larger than bricks – generally do not display this feature.

Dropped keystones in arches are widely held to be the most reliable evidence of earthquake damage (Kamai and Hatzor 2008) (Fig. 1e). These are formed during horizontal shaking, if there is no significant vertical load (i.e., when higher parts of the building have already collapsed). In multistory buildings, arches in upper floors get damaged, while lower floors do not exhibit damaged arches. Significant vertical loading tends to hold masonry blocks together through high friction at interfaces.

Broken lintels (Fig. 1f) and thresholds occur widely. These features are not firm evidence for seismic shaking because differential settlement of walls due to inadequate foundation can also produce similar structures.

Common out-of-plane failures are *rotated blocks* (Fig. 1g), whose angle or rotation (clockwise or counter clockwise) reflects the direction of strong motion and the friction between adjacent blocks.

Displaced drums in rows of masonry columns (Fig. 1h) are products of either in-plane or out-of-plane seismic loads. Both shifting and rotation of drums occur. Historically, the observation of aligned, fallen monolithic columns has been used to infer directions of earthquake epicenters. However, this interpretation is erroneous, since column alignment in fallen structures can also be produced by heavy storms acting on deteriorated buildings (Ambraseys 2006).

Extruded blocks (Fig. 1i) indicate loads at a high angle to a wall. Displacement occurs along a more-or-less irregular pattern of masonry blocks

(mostly without through going fractures), reflecting the failure of the wall core. Masonry maintains coherence during displacement, indicating that not the mortar between blocks but mortar between wall and core maintains wall integrity.

The *arcuate collapse* of walls (Fig. 2a) is sign of loads acting at high angle to a wall that is confined (fixed) at both ends. Wall terminations (usually supported by cross walls) maintain their full elevation, while the free-standing middle part loses coherence and collapses during heavy vibration.

A *warped wall* is produced by out-of-plane loads acting at high angle to the wall (Fig. 2b). Warping can be to one or both sides. Thick walls made of Roman concrete can behave this way, while thin walls collapse in arcuate form (Fig. 2a). Generally, there is no change of wall geometry at foundation level.

A *V-shaped extrusion* near the top of a cylindrical building (Fig. 2c) indicates strong motion in the direction of extrusion.

Fallen walls lying on the ground are more-or-less preserving original coherence (Fig. 2d). A seismic load acting at a high angle to a wall can cause simultaneous collapse along the full length of the structure.

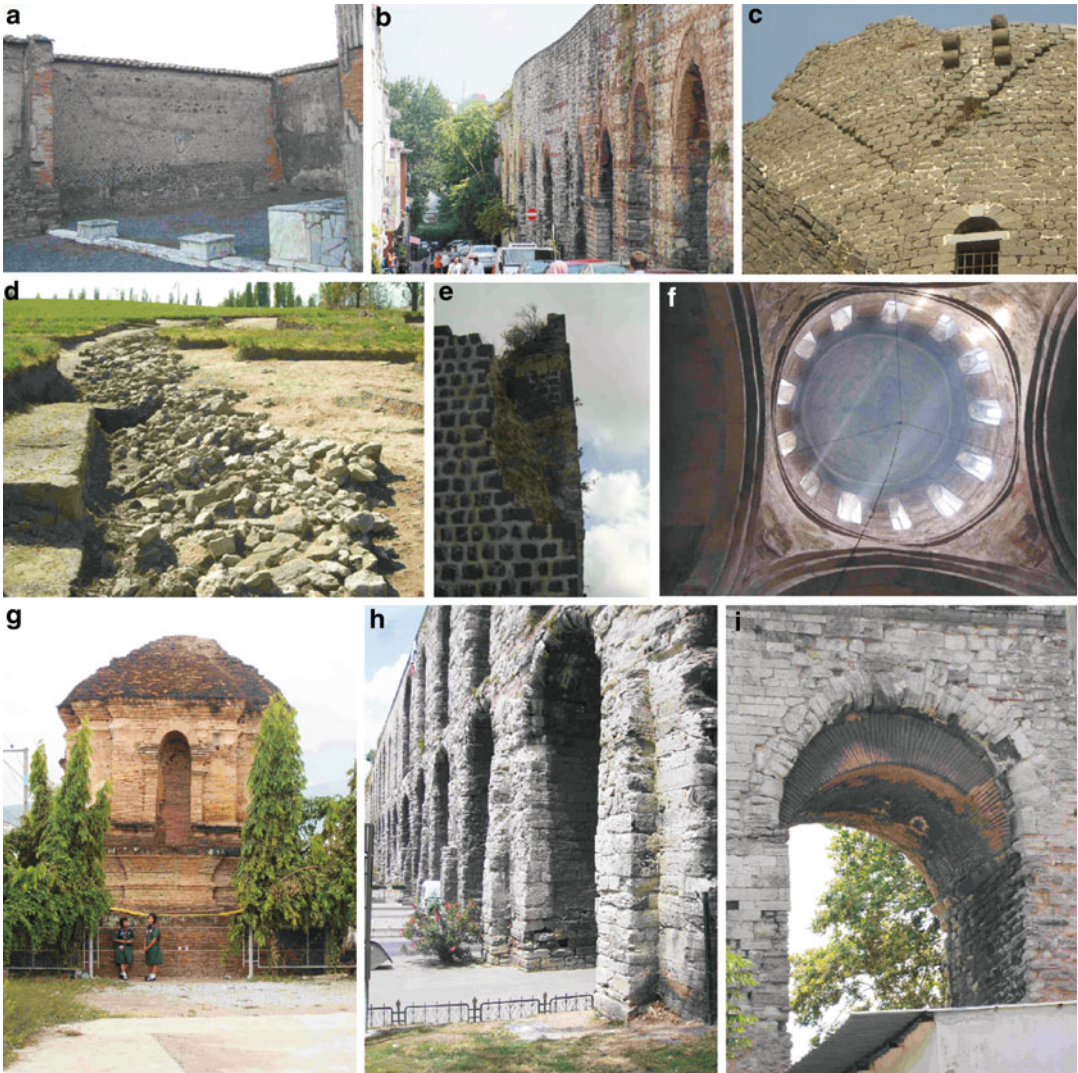
Damage Affecting Adjacent Walls

Triangular collapses in the corners of walls (Fig. 2e) are due to differential loading of perpendicular walls. Shaking in different directions at the same time shears off the masonry joints, allowing collapse of the weakened corner.

Simple *deformed geometrical structures deformed* (rectangular to parallelogram, circular to elliptical, etc.) (Fig. 2f). These features are hard to recognize, since deformations are subtle: parallels and right angles deviate a few degrees only (higher deformation would yield total collapse).

Whole Buildings

Tilted buildings are signs of uneven loading of subsoil (Fig. 2g). When seismic vibrations act on water-saturated soil, it loses coherence and behaves like a fluid. Building foundations – losing support during liquefaction events – suffer differential



Damage to Ancient Buildings from Earthquakes, Fig. 2 (a) Arcuate out-of-plane collapse in the market, Pompeii, Italy, repaired before the 79 AD eruption of Vesuvius, which buried the city. Photo: Kázmér #9266. (b) Warped northern side of the Valens aqueduct (Istanbul, Turkey). Completed in 368 AD, several subsequent earthquakes left their traces on the massive construction #0340. (c) V-shaped damage on the SW wall of the twelfth-century donjon of al-Marqab citadel, coastal Syria. The damage was inflicted by the 1202 earthquake (Kázmér and Major 2010) #4589. (d) Fallen masonry fence in the Roman city of Carnuntum (Deutsch-Altenburg, Austria) #5496. (e) Triangular missing parts

in corners of walls (al-Marqab citadel, coastal Syria) #4671. (f) Deformed circular dome of fourth- to eleventh-century Samtavro cathedral, Mtskheta, Georgia #1633. (g) Tilted Buddhist stupa (approximately fifteenth century) in Chiang Mai, Thailand. While a single tilted building does not indicate seismic origin, this stupa is one of 21 sites displaying tilting in the city (Kázmér et al. 2011) #3411. (h) Buttresses support the pillars of the Valens aqueduct in Istanbul, Turkey, completed in 368 AD #0335. (i) Valens aqueduct in Istanbul, completed in 368 AD. A brick arch was built to support the weakened stone arch after one of the frequent earthquakes in the region (Istanbul, Turkey) #0335

settlement, tilting, and collapse. Traces of total failure are mostly removed, but minor tilting – not affecting the use of the building – is often left unrepaired.

Very rarely a building – if built across an active fault – is sheared by fault movement. The displaced walls indicate the sense and cumulative offset of the fault since construction time. Walls are either displaced along a single fault (i.e., Vadum Iacob crusader tower, Israel: Ellenblum et al. 1998) or along a splay of faults, where the building maintains integrity, but the floor plan gets deformed (i.e., St. Simeon monastery, Syria: Karakhanian et al. 2008).

Repairs

Ruins of completely collapsed buildings are usually removed after an earthquake. Moderate damages are repaired on the spot, and additions, new walls, *buttresses* (Fig. 2h), *reinforced arches* (Fig. 2i), and other *supporting structures* can be identified as evidence of past earthquakes. Repairs are often built of different, often inferior material (Fig. 2i: brick arc supporting a stone arch), poorly fitted to masonry patterns, obstructing windows – these are features of repair following structural damage.

Study of damaged ancient buildings is a major source of evidence for past earthquakes: various seismic parameters (e.g., date, intensity, strong motion direction) can be assessed from a careful, critical study. However, before anyone jumps head-fast into archeoseismology, here is a cautionary note: attributing seismic origin to observed damage needs careful consideration and exclusion of all other potential causative agents (Ambraseys 2006).

Summary

Damage to ancient buildings is a major source of information about the parameters of past earthquakes (date, intensity, strong motion direction, just to name a few). Features to aid in the identification of seismic damage are briefly described

and illustrated and arranged in a simple, descriptive system. Deformations affecting single and multiple blocks of dressed masonry, single walls, adjacent walls, and whole buildings record evidence of seismic damage. Their recognition, careful analysis, and distinction from other types of damage (by aging, warfare, poor construction, etc.) need thorough and systematic study but provide much needed data on past earthquakes.

Cross-References

- ▶ [Ancient Monuments Under Seismic Actions: Modeling and Analysis](#)
- ▶ [Archeoseismology](#)
- ▶ [Damage to Buildings: Modeling](#)
- ▶ [Geotechnical Earthquake Engineering: Damage Mechanism Observed](#)
- ▶ [Masonry Structures: Overview](#)
- ▶ [Paleoseismology](#)
- ▶ [Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used](#)
- ▶ [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)
- ▶ [Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring](#)
- ▶ [Seismic Vulnerability Assessment: Masonry Structures](#)

References

- Ambraseys NN (2006) Earthquakes and archaeology. *J Archaeol Sci* 33:1008–1016
- Ellenblum R, Marco S, Agnon A, Rockwell T, Boas A (1998) Crusader castle torn apart by earthquake at dawn, 20 May 1202. *Geology* 26:303–306
- Galahini F, Hinzen K-G, Stiros S (2006) Archaeoseismology: methodological issues and procedure. *J Seismol* 10:395–414
- Hinzen K-G (2009) Simulation of toppling columns in archaeoseismology. *Bull Seismol Soc Am* 99(5):2855–2875
- Hinzen K-G (2011) Archeoseismology. In: Gupta HK (ed) *Encyclopedia of solid earth geophysics*. Springer, Berlin, pp 11–15
- Hinzen K-G, Fleischer C, Reamer SK, Schreiber S, Schütte S, Yerli B (2011) Quantitative methods in archaeoseismology. *Quat Int* 242:31–41

- Kamai R, Hatzor YH (2008) Numerical analysis of block stone displacements in ancient masonry structures: a new method to estimate historic ground motions. *Int J Numer Anal Methods Geomech* 32:1321–1340
- Karakhanian AS, Trifonov VG, Ivanova TP, Avagyan A, Rukieh M, Minini H, Dodonov AE, Bachmanov DM (2008) Seismic deformation in the St. Simeon monasteries (Qal'at Sim'an), northwestern Syria. *Tectonophysics* 453:122–147
- Karcz I, Kafri U (1978) Evaluation of supposed archaeoseismic damage in Israel. *J Archaeol Sci* 5:237–253
- Kázmér M, Major B (2010) Distinguishing damages of two earthquakes – archaeoseismology of a Crusader castle (Al-Marqab citadel, Syria). In: Stewart I, Sintubin M, Niemi T, Altunel E (eds) *Ancient earthquakes*, Geological Society of America Special Paper, 471. Geological Society of America, Boulder, pp 186–199
- Kázmér M, Sanittham K, Charusiri P, Pailoplee S (2011) Archaeoseismology of the AD 1545 earthquake in Chiang Mai, Thailand. In: Grützner C, Pérez-López R, Fernández Steeger T, Papanikolaou I, Reicherter K, Silva PG, Vött A (eds) *Earthquake geology and archaeology: science, society and critical facilities*. 2nd INQUA-IGCP 567 international workshop on active tectonics, earthquake geology, archaeology and engineering, Corinth, 19–24 Sept 2011. *Proceedings vol 2*, pp 102–105
- Korjenkov AM, Mazor E (2003) Archaeoseismology in Mamshit (Southern Israel): cracking a millennium-old code of earthquakes preserved in ancient ruins. *Archaeol Anz* 2003(2):51–82
- Marco S (2008) Recognition of earthquake-related damage in archaeological sites: examples from the Dead Sea fault zone. *Tectonophysics* 453:148–156
- Mazor E, Korjenkov A (2001) Applied archaeoseismology: decoding earthquake parameters recorded in archaeological ruins. In: Krasnov B, Mazor E (eds) *The Makhteshim country: laboratory of nature*. Pensoft, Sofia/Moscow, pp 123–153
- Rodríguez-Pascua MA, Pérez-López R, Giner-Robles JL, Garduño-Monroy VH (2011) A comprehensive classification of Earthquake Archaeological Effects (EAE) in archaeoseismology: application to ancient remains of Roman and Mesoamerican cultures. *Quat Int* 242:20–30
- Sintubin M (2011) Archaeoseismology: past, present and future. *Quat Int* 242:4–10
- Stiros SC (1996) Identification of earthquakes from archaeological data: methodology, criteria and limitations. In: Stiros S, Jones RE (eds) *Archaeoseismology*. Institute of Geology and Mineral Exploration & The British School at Athens, Athens, pp 129–152
- Vasconcelos G, Lourenço PB, Mouzakis H, Karapitta L (2006) Experimental investigations on dry stone masonry walls. In: 1st international conference on restoration of heritage masonry structures, Cairo, 24–27 Apr 2006, pp P31-1–P31-10
- Yagoda-Biran G, Hatzor YH (2010) Constraining paleo-PGA values by numerical analysis of overturned columns. *Earthq Eng Struct Dyn* 39:463–472

Damage to Buildings: Modeling

Sonia Giovinazzi¹, Indranil Kongar²,
Gian Maria Bocchini³ and Daria Ottonelli⁴

¹Department of Civil and Natural Resources Engineering, University of Canterbury, Christchurch, New Zealand

²EPICentre, Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK

³Institute of Geodynamics, National Observatory of Athens, Athens, Greece

⁴Department of Civil, Chemical and Environmental Engineering, University of Genoa, Genoa, Italy

Synonyms

Building damage; Building taxonomy; Physical impact; Seismic fragility; Seismic vulnerability

Introduction

While the seismic performance of a single building can be assessed by a detailed structural analysis, the assessment of the earthquake-induced damage to building stocks, i.e., hundreds or thousands of buildings within an urban or regional area, relies on the use of probabilistic fragility functions. Fragility functions represent the typical behavior of a group of buildings characterized by similar physical characteristics (i.e., structural, geometric, constructive, technological, etc.) in response to an earthquake. This entry describes how these functions are derived and how they are used to estimate damage scenarios for building stocks.

The modeling of earthquake-induced damage to buildings can be either probabilistic or deterministic or, since seismic risk assessment is a multi-step process, a combination of both. The benefit of probabilistic analysis is that they can account for the many uncertainties associated with seismic risk assessments of building stocks.

These include uncertainties in the quantification of seismic hazard, the identification of different building types and their characteristics within the building stock, the relationship between seismic hazard and physical damage, and the relationship between the physical damage and socioeconomic consequences.

In the case of seismic hazard definitions, there are uncertainties over the time, location, depth, and magnitude of future earthquakes. One approach to account for these is probabilistic seismic hazard assessment (PSHA) (Bommer 2002). In PSHA, maps of ground motion intensity are produced for a study area in which the values of the intensity measure at each location represent the ground motion that is predicted to occur on average once, within a specified return period, based on all possible earthquake sources. The earthquake consequences (e.g., damages, losses, or casualties) are computed for these ground motions. A feature of PSHA maps is that the evaluated ground motions are not necessarily predicted to occur simultaneously due to the same earthquake and therefore are not representative of a real earthquake. An alternative approach, which allows for more realistic representation of individual event, is to use a deterministic seismic hazard assessment (DSHA), i.e., to use scenario earthquakes. A single scenario earthquake does not account for the uncertainties in seismic hazard. However, the distribution of uncertainties can be represented by using a large-scale simulation technique such as Monte Carlo simulation or Latin hypercube sampling to undertake analysis for multiple scenario earthquakes, each with attributes determined probabilistically from the historic earthquake catalogue for the study area. In simulations such as this, where multiple potential outcomes are computed, statistical measures are needed to make inferences from the results.

There are also uncertainties in the relationship between seismic hazard and damage – two buildings of the same typology will not necessarily experience the same damage if exposed to the same ground motion due to inestimable factors such as constructive details, the age, and the maintenance history of each specific building.

A well-established tool used to relate building damage to seismic hazard probabilistically is the fragility function, commonly referred to as the fragility curve when presented graphically (e.g., Rossetto and Elnashai 2003). For a given level of ground motion intensity, fragility functions determine the probability that a building or a group of buildings will be in, or exceed, the i th damage state, d_i . The probabilistic relationship is commonly assumed to be a cumulative lognormal distribution. In this case, the fragility function can be expressed conditionally in the form of Eq. 1.

$$P[d \geq d_i | IM] = \Phi \left[\frac{1}{\beta_{di}} \ln \left(\frac{IM}{IM_{di}} \right) \right] \quad (1)$$

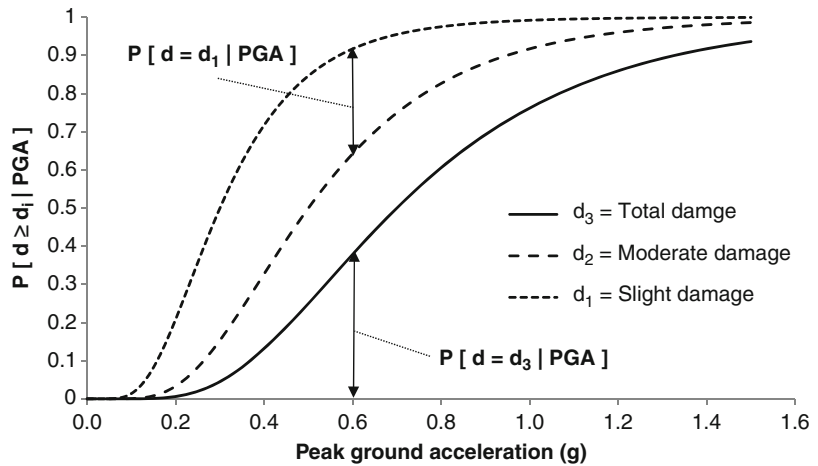
In Eq. 1, P is the exceedance probability for the i th damage state, IM is the ground motion intensity (e.g., peak ground acceleration), Φ is the standard normal cumulative distribution function, β_{di} is the standard deviation of the natural logarithm (dispersion) of ground motion intensity for the i th damage state, and IM_{di} is the median value of ground motion intensity at which the subject reaches the i th damage state. If different damage states are defined for the building or building class, then distinct fragility curves exist for each damage state with variation of the dispersion and median values of ground motion intensity. Therefore, the risk to a building or building class being assessed for multiple damage states will be represented by a set of fragility curves. Fragility curves define exceedance probabilities, but it is also desirable to know the probabilities that a building or building class will be in specific damage states. For a specified value of ground motion, these can be calculated from the intervals between the fragility curves, so the probability a structure is in the i th damage state is the difference between the exceedance probabilities for the i th damage state and the $i + 1$ th damage state (Eq. 2):

$$p_i = P[d \geq d_{i+1} | IM] - P[d \geq d_i | IM] \quad (2)$$

Figure 1 shows a set of fabricated fragility curves for a structure whose damage is defined

Damage to Buildings:

Modeling, Fig. 1 A fabricated set of fragility curves for a fictitious structural typology



by three damage states against peak ground acceleration. Although this form of fragility function has been widely adopted in academic literature and by risk assessment platforms, probabilistic functions do not necessarily have to assume a cumulative lognormal shape derived, as specified in section “[Assigning Damage to Buildings Based on Ground Motions](#).”

Six steps are necessary toward the modeling of building damage (Fig. 2). The first step is to classify the building stock under analysis, so that buildings that are expected to have a similar seismic performance are considered uniformly. To do so, it is necessary to select an existing building taxonomy or define an ad hoc building taxonomy, classify the building within the taxonomy, and enumerate the buildings in each one of the identified building classes. The second step is to define the appropriate scales for classifying damage to building structural and nonstructural components. The third step is to determine the appropriate earthquake hazard parameters against which damage should be evaluated. The fourth step is to generate the hazard scenarios (ground motions and land deformations) for which the building damage needs to be assessed. The fifth step is to identify the appropriate hazard-damage relationship and assign damage to buildings. The sixth step is to estimate the direct losses and functional impacts related to the earthquake-induced physical damage to buildings.

Classifying Buildings

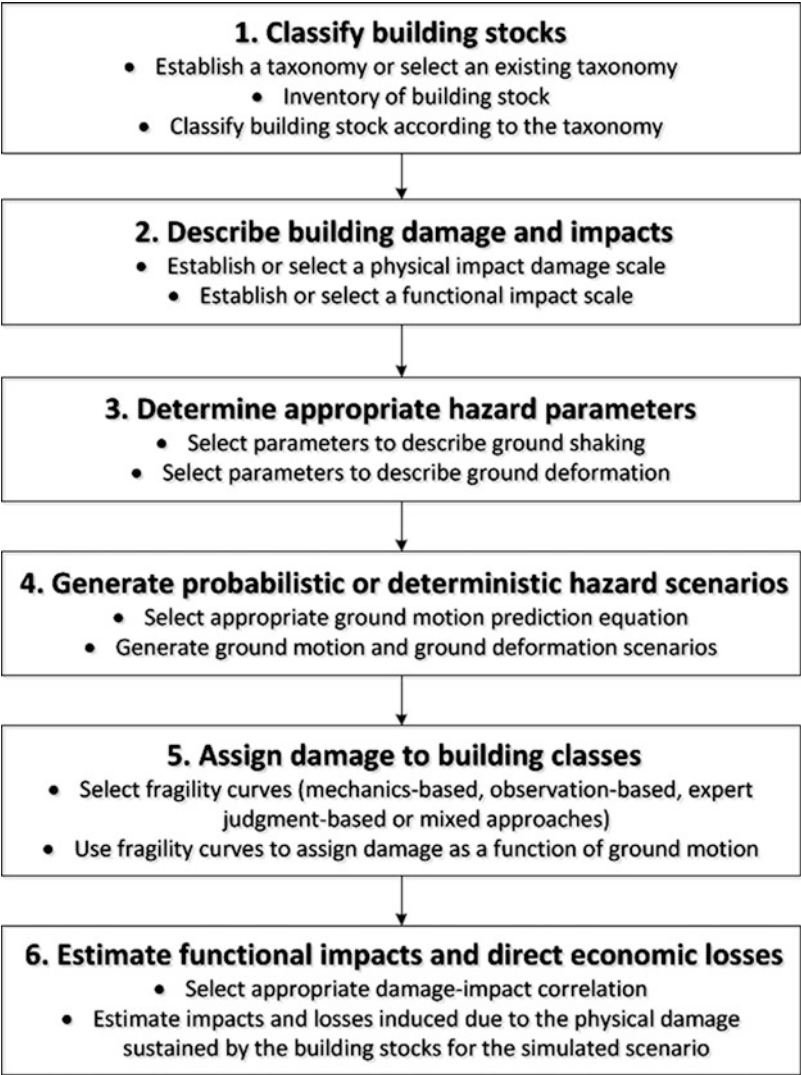
The first step to predict expected damage to buildings, in the event of an earthquake, is to develop an accurate classification of the building stock. The purpose of classifying building stock is to group together buildings that can be expected to behave similarly, sustaining a comparable level of damage, when subjected to comparable levels of earthquake-induced ground shaking and deformation. To this aim, relevant data and information on the building stock should be collected. The following subsections provide a non-exhaustive overview of the steps that are needed to classify building stocks for seismic risk analysis purposes.

Building Taxonomies

Building classification framework for seismic risk analysis purposes, referred here as building taxonomies, groups buildings by their construction attributes, geometry, and age/maintenance conditions, among many other factors that can influence their seismic behavior. Furthermore, building taxonomies group buildings by their occupational and social function (i.e., residential, commercial, industrial). In fact, after the earthquake-induced physical damage to building, the number of casualties, the impact on content and machineries, the direct economic losses, and the business interruption aspects are strictly correlated with the building occupational and social function.

Damage to Buildings: Modeling,

Fig. 2 Suggested steps for assessing the earthquake-induced damage, and related functional impacts and direct losses, to buildings



D

This section will provide a literature review on different building taxonomies that have been proposed worldwide to classify building stocks according to their expected seismic performance and occupational function.

Early building taxonomies grouped building typologies according to their building material only. Table 1 reports, for example, the three building classes identified and defined by the Medvedev–Sponheuer–Karnik (MSK-64) macroseismic scale.

The need for a deeper diversification of building behaviors, when subjected to an earthquake event led to the definition of more

elaborated classification systems, such as the one proposed within HAZUS (NIBS 2003) for the United States and the European Macroseismic Scale EMS-98 (Grünthal 1998) for Europe, where, further to the material, consideration is given to primary parameters influencing the building damage, namely, (1) the building structural system, (2) the seismic design criteria (design code level), (3) the building height (low rise, mid rise, high rise), and (4) the nonstructural elements. Table 2 reports and compares HAZUS and EMS-98 taxonomies for unreinforced masonry and reinforced concrete buildings.

Damage to Buildings: Modeling, Table 1 Classes of buildings according to the MSK-64 macroseismic scale

A	Buildings with dry stone or clay, adobe, or mud walls
B	Buildings with brick walls, mortar blocks, masonry and mortar, stone block, timber frame
C	Buildings with metal structure or reinforced concrete

Recently, international projects and initiatives worldwide have developed and defined several further taxonomy systems aiming to represent the variety of building types around the world and the continued upgrading/modification of construction techniques. Pitilakis et al. (2013) proposed a modular and flexible taxonomy for classifying European masonry and reinforced concrete buildings, where buildings are classified according to some main categories, including the force-resisting mechanism (FRM), material of FRM, plan (regular or irregular), elevation (regular or irregular), cladding, detailing (ductile, non-ductile, with tie-rod beam, without tie-rod beam), floor system, roof system, height level, code level, and material. Subcategories (e.g., FRM, material of FRM, cladding characteristics, floor system material, roof system material, number of stories) were introduced for some of the aforementioned categories, to provide a more refined classification, when more accurate data are available for the building stock under analysis.

The Prompt Assessment of Global Earthquakes for Response (PAGER) project (Porter 2005) developed a comprehensive building taxonomy with an international coverage including structural types common in non-developed countries (i.e., adobe, clay, poor quality wood constructions). The Earthquake Engineering Research Institute (EERI) developed a World Housing Encyclopedia (WHE) (EERI 2000) collecting information related to housing construction practices in seismically active areas worldwide.

The Global Earthquake Model (GEM) developed a global building taxonomy aiming to represent all the building types around the globe (Brezev et al. 2013; Gallagher et al. 2013) by

Damage to Buildings: Modeling, Table 2 HAZUS (NIBS 2003) and EMS-98 (Grünthal 1998) classifications for unreinforced masonry and reinforced concrete buildings

	EMS 98	HAZUS
Unreinforced masonry	Rubble stone	Unreinforced masonry bearing walls (URM)
	Adobe (earth bricks)	
	Simple stone	
	Massive stone	
	U Masonry (old bricks)	
Reinforced concrete	U Masonry – RC floors	Concrete moment frame
	Reinforced concrete frame	
	Shear walls	
		Concrete frame with U Masonry infill walls

building on a comprehensive review of existing taxonomies (Charleson 2011) and on the findings from WHE. The effectiveness of existing taxonomy to represent building typologies worldwide was checked against different criteria (Porter 2005) aiming to define a simple, collapsible, and detailed taxonomy, international in scope, user-friendly, and able to distinguish differences in seismic performance, but extensible to future needs (e.g., to other natural hazards). Following this comprehensive and critical review, the Syner-G taxonomy (Pitilakis et al. 2013) was selected as the benchmark to define the GEM building taxonomy. The GEM taxonomy is organized as a series of expandable tables, describing various attributes at increasing levels of detail (up to 5 depending on the specific attribute). Fabricated examples of building taxonomy and building occupancy taxonomy are presented, respectively, in Tables 3 and 4.

Further to taxonomies for building structures, specific building taxonomies have been developed for nonstructural components and content (e.g., NIBS 2003; NIST 1999; Porter 2005). HAZUS (NIBS 2003) provided a classification for architectural components (e.g., veneer and

Damage to Buildings: Modeling, Table 4 Example of building occupancy taxonomy (Adapted from GEM taxonomy, Brezev et al. 2013)

Building occupancy class									
Unknown occupancy	Residential	Mixed use	Industrial	Agriculture	Assembly	Government	Education	Other occupancy type	Commercial and public
Level 1 attributes									
									Offices, professional/technical services ^a
									Hospital/medical clinic
									Entertainment
									Public building
									Covered parking garage
									Bus station
									Railway station
									Airport
									Recreation and leisure
									Higher-level attributes

^a An example of higher-level attributes for the “Commercial and Public” occupancy class

finishes, cantilever elements and parapets, etc.), mechanical and electrical components (e.g., piping systems, elevators, etc.), and contents (e.g., file cabinets, bookcases etc.). NISTIR 6389 (NIST 1999) provided taxonomy for shell, interiors, services, equipment, furniture, and special constructions. Porter (2005) proposed an expendable and flexible taxonomy for nonstructural elements based on the review, combination, and improvement of the existing ones, including NISTIR 6389 (NIST 1999).

Building Inventories

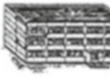
A building inventory for seismic risk analysis purposes can be defined as an enumeration and characterization of the building stock under analysis according to the assumed taxonomy. Preparation of a building stock inventory is a critical part of a damage and loss estimation study and can be both time and resource consuming. A compromise needs to be sought between the time and cost required for higher-quality data on the building stock and the uncertainties arising from low-quality data. In order to obtain an inventory affordable both from the cost and the time point of view, it would be worth to collect, as a first step, all the already available data sources, while being conscious that these data are often overlapping and incomplete and that additional work might be required to reconcile multiple sources and missing data. Possible available sources, in order to built-up regional building inventory, are databases belonging to state, regional, and local authorities and private sector sources as well as inventory information arising from previous loss or available within seismic risk assessment tools or platforms (e.g., HAZUS NIBS 2003). Census data, for example, can provide aggregated data on the size, the number, and the age of residential buildings; conversely specific databases prepared, for example, for insurance purposes or for civil defense planning purposes can provide information on commercial and/or strategic buildings.

Further to available databases, there are different methods for collecting, processing, and archiving building inventory datasets. Currently, the most common and widely used approaches

(Tenerelli et al. 2013) include (1) ground screening procedures (e.g., field reconnaissance, sidewalk survey, street-front direct visual observation, interpretation of street-front photographs, etc.), (2) satellite-remote sensing approaches, or (3) a combination of ground screening procedures and satellite-remote sensing approaches. Ground screening procedures collect building attributes through direct observations, usually via (a) paper forms, ad-hoc formatted to quickly record relevant building characteristics and attributes (e.g., FEMA 154 1988), and (b) handheld equipment and software (e.g., tablet, phone) that enables user-defined electronic inventory data. For both the approaches, a critical aspect is the definition of the survey form that has to be coherent with the adopted building taxonomy (section “[Building Taxonomies](#)”) and with the adopted approach to assign damage to buildings as a function of the ground motion (section “[Assigning Damage to Buildings Based on Ground Motions](#)”). Furthermore, proper training of staff those performing the ground screening survey is essential since identifying structural types and other building attributes from the street can be challenging. Satellite and airborne remote sensing techniques may provide information on building stocks, e.g., building footprint and height information, in a timely and cost effective manner. However, these techniques must be closely calibrated to the particular region under study, and the analysis must be supported by expert judgment, coupled with some field reconnaissance.

Building Classification

The final aim of the building classification process is to identify and enumerate the buildings composing the building stock under analysis in each of the building types and subtypes considered by the assumed classification system. Although a lot of information sources on the building stock might be available at regional level, and further data can be collected (section “[Building Inventories](#)”), they might be insufficient for a direct classification of the building in terms of the construction types and occupancy classes identified by an assumed taxonomy (section “[Building Taxonomies](#)”).

	Slight (D_1)	Moderate (D_2)	Heavy (D_3)	Very heavy (D_4)	Collapse (D_5)
Masonry					
Reinforce Concrete					

Damage to Buildings: Modeling, Fig. 3 Damage grades for masonry and reinforced concrete buildings according to the EMS-98 macroseismic scale (Adapted from Grünthal 1998)

Therefore, inferences have to be established between larger groups of buildings (referred to as categories), recognized on the basis of more general information (such as land use patterns or building age) and building types or occupational classes. Inferences may take the form of “for masonry buildings built before 1919, 40 % belong to rubble stone typologies and 60 % are old brick masonry buildings.” The assumed inferences should be verified on the basis of the “expert opinion” of local engineers and building officials in order to verify that they reflect the construction features of the region. A ground screening procedures can be used to develop or check the inference rules, used to characterize the building stock into categories, as well to assess the accuracy of the available information.

Describing Building Damage

Earthquake-induced damage to building structures and nonstructural elements is usually described in a scale of multiple increasing damage states, also referred to as damage grades, from no damage to total collapse. The damage scale adopted in the HAZUS (NIBS 2003) platform, for example, accounts for four discrete damage states further to “no damage,” namely, minor damage, moderate damage, extensive damage, and collapse. In the European Macroseismic Scale EMS-98 (Grünthal 1998), the damage is represented in a discrete form through five-damage grades D_i ($i = 1 \div 5$), further to “no damage,” D_0 . Figure 3 represents the EMS-98 (Grünthal 1998) five-damage grades for masonry and reinforced building types. For each

damage grade, the EMS-98 scale provides a qualitative description of the expected physical impacts to structural and nonstructural elements.

Waiiles and Horner (1933) propose a five-damage state scale, where for some of the damage states, further to qualitatively describing the expected impacts to structural and nonstructural elements, an estimate of the damage extension is provided (Table 5). ATC-13 (1985) proposes a six-damage grade scale, plus a “none damage” (Table 5).

In order to estimate the functional impacts and consequences deriving from the earthquake-induced physical damage to buildings, functional impact scales are needed and a correspondence between physical damage grades/states and functional levels needs to be established. Vision 2000 (SEOAC 1995) proposes a five-level performance objective scale, where the following functional level/objectives for buildings are identified: fully operational, operational, life safety, and near collapse. Rossetto and Elnashai (2003) propose damage scales specific for reinforced concrete buildings (referred to as homogenized reinforced concrete, HCR) and a correspondence between physical damage states and performance objectives, namely, serviceability, damage control, and collapse prevention. The Applied Technology Council Report (ATC-21 1988) defines a tagging system, corresponding to a three-level scale, to be used for post-earthquake safety assessment aiming to identify buildings, which are safe to occupy (green tag) being undamaged or with light damage, temporarily unusable building (yellow tag) being affected by significant damage and requiring repairs before reoccupation, and unusable,

Damage to Buildings: Modeling, Table 5 Examples of damage and functional scales: comparison of damage grades and functional levels (Adapted from Rossetto and Elnashai 2003)

HRC	HAZUS	Vision2000	EMS-98	ATC-13	ATC-21	Wailes and Horner ^a
None	No damage	Fully operational	No damage (D ₀)	None	Green tag	A
Slight	Slight		Slight (D ₁)	Slight		
Light		Operational		Light		B
Moderate	Moderate	Life safe	Moderate (D ₂)	Moderate	Yellow tag	C
Extensive	Extensive	Near collapse	Heavy (D ₃)	Heavy		
Partial collapse			Very heavy (D ₄)	Major	Red tag	D
Collapse	Collapse	Collapse	Collapse (D _{G5})	Destroyed		E

^aA – Undamaged or minor cracking, no significant structural damage, minor veneer damage

B – Parapet failure or separation of veneer, major wall cracking, and interior damage

C – Failure of portion of exterior walls, major damage to less than 50 % of walls

D – Major damage to more than 50 % of walls

E – Non-repairable damage, demolition probably appropriate

dangerous buildings (red tag) being affected by heavy structural damage up to partial or total collapse, deemed in some cases for demolition. Table 5 presents few examples of damage scales and functional scales and compares their damage grades and functional levels.

Hill and Rossetto (2008) critically reviewed and compared building damage scales, identifying their potentialities and shortcomings for their use in seismic risk analysis and for loss assessment applications.

Selecting Appropriate Earthquake Hazard Parameters

The selection of earthquake hazard parameters, commonly referred to as intensity measure (IM), is a crucial step when assessing seismic risk for buildings. Performance, integrity, or damage of structures is, in fact, influenced by the intensity, frequency content, and duration of the ground motion (Weatherill et al. 2011a). The choice of a suitable IM should possibly account for the characteristics of the building stock analyzed (section “Classifying Buildings”) and the method used to predict the building performance, when subjected to seismic solicitations (section “Assigning Damage to Buildings Based on Ground Motions”). Physical–mechanical

parameters or qualitative parameters can be selected for the ground motion description.

Physical–mechanical parameters for the ground motion description aim to characterize the amplitude, the frequency content, and the duration of strong ground motion; some of them describe only one of these characteristics, while others may reflect several of them. In engineering application the most commonly used measure of acceleration for a particular ground motion is the peak horizontal acceleration, PHA, the maximum horizontal acceleration at the ground surface is also known as the peak horizontal ground acceleration. For most earthquakes, the horizontal accelerations is greater than the vertical acceleration, and thus the peak horizontal ground acceleration also turns out to be the peak ground acceleration (PGA). The PGA for a given component of motion is the largest (absolute) value of horizontal acceleration obtained from accelerogram of that component. By taking the vector sum of two orthogonal components, the maximum resultant PGA can be obtained. Horizontal accelerations have commonly been used to describe ground motions because of their natural relationship to inertial forces; however they do not provide any information about the dynamic behavior of a structure. This can be obtained by the use of response spectra. Response spectrum describes the maximum response of

a single-degree-of-freedom (SDOF) system to a particular input motion as a function of the natural frequency (or natural period) and damping ratio of the SDOF. The response may be expressed in terms of acceleration, velocity, and displacement. The maximum values of each of these parameters depend only on the natural frequency and damping ratio of the SDOF system (for a particular input motion). The maximum values of acceleration, velocity, and displacement are referred to, respectively, as spectral acceleration, S_a ; spectral velocity, S_v ; and spectral displacement, S_d . Note that for a SDOF system of zero natural periods the S_a would be equal to the PGA.

When adopting mechanical-based approaches for assessing damage to buildings (section “[Assigning Damage to Buildings Based on Ground Motions](#)”), S_a at the natural period of the structure and/or S_d should be the preferred IMs. Where a capacity spectrum method is used, the full spectrum is essential along with the definition of damping reduction factors (Weatherill et al. 2011a).

Differently from physical–mechanical parameters, the macroseismic intensity (I) is a qualitative description of the effects of the earthquake at a particular location, as evidenced by observed damage on the natural and built environment and by the human and animal reactions at that location. Thanks to its qualitative nature, macroseismic intensity is the oldest measure for earthquake impacts; I scales have been extensively used when adopting empirical-based fragility functions (section “[Assigning Damage to Buildings Based on Ground Motions](#)”). Different macroseismic intensity scales exist all over the world. Foulser-Piggott and Spence (2013) recently attempted the development of a uniform and worldwide recognized International Macroseismic Scale.

Weatherill et al. (2011a) provides a detailed and critical review of the recommended and more common and suitable intensity measures to use in seismic risk analysis for different types of buildings, including masonry, reinforced concrete, steel frame, and timber building types (Table 6).

Damage to Buildings: Modeling, Table 6 Most commonly used and recommended intensity measures (IMs) for seismic risk assessment of building stocks (Modified from Weatherill et al. 2011a)

Building type	Common IMs	Recommended IMs
Masonry	S_a , S_d , PGA, IEMS-98	S_a , S_d
Reinforced concrete	S_a , S_d , PGA, PGV, IEMS-98	S_a , S_d
Steel	S_a , S_d	S_a , S_d
Timber/wood	S_a , S_d , PGA, PGV	S_a , S_d

Generating Ground Motions

Two commonly recognized approaches exist for generating ground motions for seismic risk analysis purposes, namely, the deterministic seismic hazard assessment (DSHA) and the probabilistic seismic hazard assessment (PSHA). The DSHA considers each seismogenetic source separately and assume the occurrence of an earthquake of specified size at a specified location. PSHA combines the contributions from all the relevant seismogenetic sources pertinent to the region or urban area under analysis and allow for characterizing the rate at which earthquakes and particular levels of ground motions occur.

The maps of ground shaking generated by PSHA or DSHA approaches represent future possible “ground motion fields.” Ground motion fields, resulting from PSHA, provide for each location possible levels of ground shaking (logarithmic mean and residual) with the associated probability of being exceeded in a given time period, giving the rate at which particular levels of ground motion might occur (e.g., 10 % probability of being exceeded in 50 years, which correspond to a return period of 475 years). Results from a PSHA can be, as well, represented in terms of hazard curves, giving, at each location, the probabilities of exceedance within a given time period for a continuous range of ground shaking intensities. Ground motion fields obtained via DSHA represent at each site the level of ground shaking (logarithmic mean and residual) resulting from a single, deterministic

earthquake (e.g., the maximum historical event from a pertinent seismogenetic source or the maximum earthquake compatible with the known tectonic framework) at a specific location (e.g., the shortest distance from the source to the studied region, when the most unfavorable situation is considered).

It is well recognized that there is not a mutually exclusive dichotomy between DSHA and PSHA (Bommer 2002) and different approaches combining DSHA and PSHA have been proposed in the literature. Bazzurro and Cornell (1999) propose to select the event that is most likely to contribute to the hazard for a given intensity measure at a site, from the disaggregation of PSHA outputs. Bazzurro and Luco (2005) proposed the use of Monte Carlo simulations to reproduce a series of single earthquake scenarios within PSHA.

For PSHA, DSHA, and any combination of them, it is necessary, as a first step, to identify the potential seismogenetic sources pertinent to the region under analysis and characterize them in terms of locations, geometry, seismic activity, and possibly released energy. As a second step, it is necessary to represent the propagation of the ground motion by the use of a suitable ground motion prediction equation (GMPE). GMPEs, also referred to as “attenuation” relationships, are equations derived from observed strong motion records and provide a means for predicting the level of ground shaking (in terms of the desired intensity measure, e.g., PGA, PGV) and its associated uncertainty at any given site or location, based on the earthquake magnitude, source-to-site distance, local soil conditions, fault mechanism, etc. Due to the empirical nature of GMPEs, the ground motion at each site/location is estimated by the median, or logarithmic mean, of the ground motion plus two random variables, referred to as residuals. The residuals represent the aleatory variability, namely, the difference between the predicted and observed ground motions in the regression analysis, occurring intra-event, i.e., within a single event, and inter-event, i.e., between separate events (Douglas 2011). For this reason, when GMPEs are applied to a large area,

adjacent predicted ground motions do not correlate with each other and the resulting ground motion field is random. This is fine when dealing with spatially clustered analysis (i.e., analysis at a single site). Conversely, when the analysis is targeting spatially distributed building stocks and the simultaneous generation of ground motions at multiple sites is required, the ground motion field should be spatially correlated to more realistically represent the earthquake scenario. Crowley et al. (2008) found that the spatial correlation of ground motion field can significantly influence the outputs of a seismic loss analysis for distributed portfolios, when simulating a single earthquake scenario. A number of spatial correlation models are available in the literature. Weatherill et al. (2011b) present a comprehensive discussion and description of different methods to produce spatially correlated ground motion fields and to account for site amplifications.

Assigning Damage to Buildings Based on Ground Motions

The damage that might affect a building stock, when subjected to a seismic event, can be predicted via different approaches, including (a) actual damage observation, (b) expert judgment, (c) simplified mechanical and analytical models, or a combination of the three aforementioned approaches. Methods based on actual damage observations are derived statistically by processing the damage data from different earthquakes and by summarizing the outcomes in terms of (a) damage probability matrices (DPM) (e.g., Whitman et al. 1973, Table 7), (b) vulnerability curves (e.g., Giovinazzi 2005, Fig. 4a, c), or fragility curves (e.g., Rossetto and Elnashai 2003). A DPM is a matrix that summarizes the seismic damage that occurred in a set of buildings, represented as the statistical distribution for each of the damage grades D_i considered and for the different macroseismic intensities I . In other words, DPM expresses, in a discrete form, the conditional probability $p[D_i/T, I]$ to obtain a damage grade D_i due to an earthquake of

Damage to Buildings: Modeling, Table 7 Example of DPM, representing the probability of damage in percent by macroseismic earthquake intensity and damage states (Adapted from ATC-13 1987)

Damage state		Earthquake intensity				
		VI	VII	VIII	XI	X
1	None	95.0	49.0	30.0	14	3.0
2	Slight	3.0	38.0	40.0	30	10
3	Light	1.5	8.0	16.0	24	30
4	Moderate	0.4	2.0	8.0	15	26
5	Heavy	0.1	1.5	3.0	10	18
6	Major	–	1.0	2.0	4.0	10
7	Destroyed	–	0.5	1.0	2.0	3.0

macroseismic intensity I and for a defined building type T (Table 7). Examples of vulnerability and fragility curves obtained from the processing of damage observations are provided in section “Assigning Damage to Buildings Based on Macroseismic Intensities.”

Due to the inherent difficulty to retrieve comprehensive damage data for all the building types considered within an assumed taxonomy and for a complete range of damaging earthquake intensities, “hybrid” methodologies have been defined, where the available observed damage data have been completed with the use of different approaches, including numerical analysis, neural network system, fuzzy set theory, expert judgment, etc.

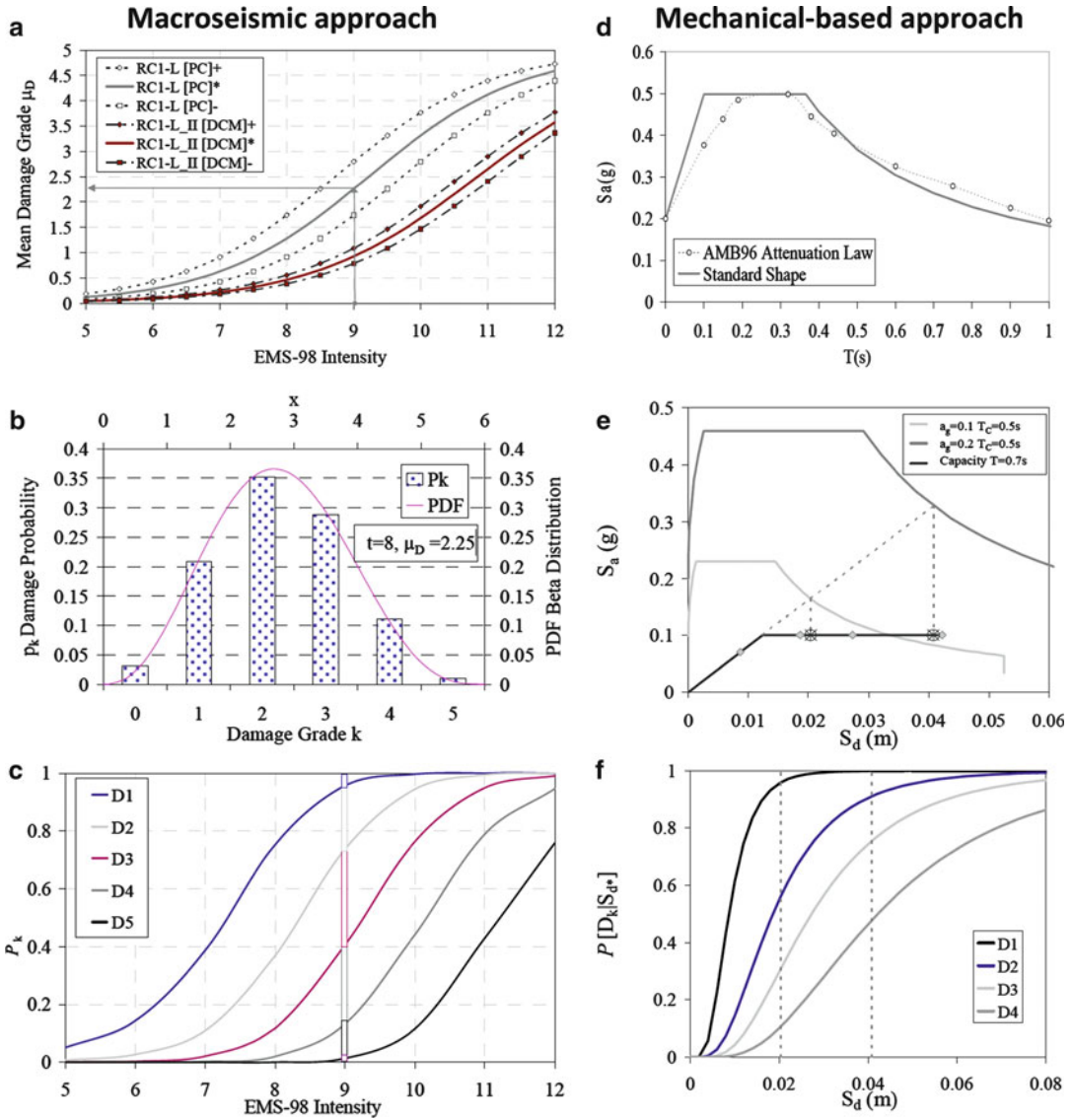
Expert-based methods imply the human judgment to completely replace the processing of observed data, leading to expert-defined *DPM* (e.g., ATC-13 1987) or score assignment procedures (e.g., ATC-21 1988; FEMA 154 1988).

Mechanical-based models for territorial scale analysis on classes of buildings can be defined on the basis of either traditional force-based procedures (e.g., capacity spectrum method implemented in HAZUS, in NIBS (2003), or in RISK-UE (2004)) or according to displacement-based designed approaches (e.g., Calvi et al. 2005). According to force-based procedures, the building performance is identified, within acceleration-displacement response spectra (ADRS) by the intersection point between the capacity curve of an equivalent nonlinear SDOF

system and the earthquake demand curve, adequately reduced to account for the inelastic behavior and energy dissipation capacity of the system (Fig. 4e). On the other hand, according to displacement-based approaches, the periods associated to the boundary of different limit states can be evaluated by the intersection between capacity curves, represented in terms of period displacement relationship, and the displacement spectrum demand curves, scaled by equivalent viscous damping factors. Other proposals for mechanical-based methods are based on the evaluation of collapse multipliers associated to alternative collapse mechanisms (e.g., Cosenza et al. 2005) or on the derivation of fragility curves directly from the results of extensive numerical analyses (e.g., Elnashai and Jeong 2005). The examples of an observed-based approach and a mechanical-based approach are summarized in the following subsections (sections “Assigning Damage to Buildings Based on Macroseismic Intensities” and “Assigning Damage to Buildings Based on Physical–Mechanical Ground Motion Parameter,” respectively). Despite being different for derivation and conception, a quantitative comparison and reciprocal calibration has been possible for these two specific methods, thanks to their closed-form formulations (Lagomarsino and Giovinazzi 2006). As a useful result, refinements of the methods based on numerical/experimental analysis were directly implemented into an equivalent macroseismic approach. Dually, the reliability of assumed force-based capacity curves was cross-validated on the basis of real observed damage data (Giovinazzi et al. 2006).

Assigning Damage to Buildings Based on Macroseismic Intensities

Combining classical probability theory and fuzzy set theory, a macroseismic approach was derived (Giovinazzi and Lagomarsino 2004; Giovinazzi 2005), based on the EMS-98 macroseismic scale definitions (Grünthal 1998) that summarize the damage observations following several European earthquakes. Numerical and complete *DPM* have been evaluated, in the range of damaging



Damage to Buildings: Modeling, Fig. 4 Macroseismic- and mechanic-based approaches: basic steps. Macroseismic: (a) medium, upper, and lower vulnerability curves for RC building types; (b, c) damage

probabilities p_k and fragility curves for RC types for $I_{EMS-98} = 9$ ($\mu_D = 2.25$). Mechanic: (d) elastic response spectrum; (e) evaluation of the performance point through capacity spectrum method; (f) lognormal fragility curves

EMS-98 intensities, I_{EMS-98} , and damage grades D_i ($i = 1 \div 5$), for the EMS-98 building classes and types (Table 2). Fuzzy set theory was effective to consider, within the proposed macroseismic approach, the epistemic uncertainties affecting the damage assessment and the building classification procedures (section “Classifying Buildings”). Upper and lower bounds of

the expected damage, as provided by the DPM_{EMS-98} , have been in terms of a mean damage grade, μ_D , defined as the mean value of the DPM_{EMS-98} damage distributions:

$$\mu_D = \sum_{i=1}^5 p_i i \quad 0 \leq \mu_D \leq 5 \quad (3)$$

where p_i is the probability of having damage of grade i , in the set of buildings.

The relationship between the mean damage grade, μ_D , and the intensity, $I_{\text{EMS-98}}$, has been represented in the form of $I_{\text{EMS-98}}-\mu_D$ damage curves (Fig. 4b) as

$$\mu_D = 2.5 \left[1 + \tanh \left(\frac{I + 6.25 * V - 13.1}{Q} \right) \right] \quad (4)$$

where Q is a ductility-based index and V is a vulnerability index, i.e., $V = V^* + \Delta V_m + \Delta V_r + \Delta V_s$, where V^* , the function of the building typology; ΔV_m , the behavior modification factor; ΔV_r , the regional vulnerability factor; and ΔV_s , the soil amplification factor. ΔV_s has been evaluated for different building types, class of height, and soil classes (Giovinazzi 2005), while the values of the V^* , ΔV_m , ΔV_r factors have been calibrated on the basis of observed damage data and expert judgment. A beta probability density function (Fig. 4b) has been assumed to represent the damage distribution around the mean damage grade μ_D . Fragility curves (Fig. 4c) defining the probability of reaching or exceeding each damage grade $P[d_i/I, (V, Q)]$ can be directly derived. Different scatter can be associated to the beta damage distributions depending on the level of the cognitive uncertainties and the reliability of the data used for the building classification (section “Classifying Buildings”).

Assigning Damage to Buildings Based on Physical-Mechanical Ground Motion Parameter

The mechanical method proposed in the framework of RISK-UE project (2004) is essentially a capacity spectrum-based method, similar to the one adopted by HAZUS (NIBS 2003) with few modifications including (1) the definition of capacity curves for non-designed European masonry typologies, accounting for the prevailing collapse modes, geometrical features, and mechanical and dynamic characteristics; (2) the definition of capacity curves for seismically designed buildings according to the Eurocode 8 (CEN 2003) and to older European design codes; and (3) the representation of the cognitive uncertainties. In order to facilitate the operative implementation, the mechanical method was defined with a closed-form solution (Lagomarsino and Giovinazzi 2006). Simplified bilinear elastic-perfectly plastic capacity curves were defined, given the yielding acceleration a_y , the fundamental period T , and the structural ductility capacity μ . Constant ductility inelastic response spectra were derived from a 5 % damped elastic response spectrum $S_{ae}(T)$ by means of a ductility-based reduction factor, $R\mu$. The displacement corresponding to the performance point S_d can thus be directly evaluated (Fig. 4e), without any further iteration as

$$S_{d*} = \begin{cases} \left[1 + \left(\frac{S_{ae}(T)}{a_y} - 1 \right) \frac{T_C}{T} \right] d_y & \text{if } T < T_C \quad \text{and } \frac{S_{ae}(T)}{a_y} > 1 \\ \frac{S_{ae}(T)}{a_y} d_y & \text{if } T_C \leq T < T_D \quad \text{and } \frac{S_{ae}(T)}{a_y} \leq 1 \\ \frac{S_{ae}(T_D) T_D^2}{4\pi^2} d_y & \text{if } T \geq T_D \end{cases} \quad (5)$$

where T_C and T_D define the onset of constant spectrum velocity and displacement range within the elastic response spectrum $S_{ae}(T)$ evaluated from deterministic or probabilistic hazard analyses (Fig. 4d).

Estimating Losses and Consequences Due to Building Damage

Earthquake-induced structural and nonstructural damage to buildings is the root cause for several

Damage to Buildings: Modeling, Table 8 Example of weight factors to assess consequences from earthquake-induced physical damage to buildings, referred to EMS-98 damage scale

Damage grade	Structural (S), nonstructural (NS)	w_k	MRD ^a	MRD ^b	MRD ^c	UFU ^d	S ^e
Slight (D1)	S = no, NS = slight	w_1	0.02	0.01	0.01	0	0
Moderate (D2)	S = slight, NS = moderate	w_2	0.1	0.1	0.1	0	0
Heavy (D3)	S = moderate, NS = heavy	w_3	0.5	0.35	0.35	0.40	0
Very heavy (D4)	S = heavy – NS = very heavy	w_4	1	0.75	0.75	1	0
Collapse (D5)	S = very heavy	w_5	1	1	1	1	0.3

^aMean damage ratio, derived from HAZUS (NIBS 2003)

^bMean damage ratio, derived and adapted from ATC-13 (1987)

^{c-e}Mean damage ratio, unfit for use buildings, severely injured from Bramerini et al. (1995)

losses and impacts. Consequences to buildings and their contents, direct economic losses (e.g., cost of repair/reconstruction of building structural or nonstructural elements or replacement of machinery and content), and consequences to inhabitants and building occupants can be estimated once the physical damage sustained by the buildings has been assessed. Different empirical correlations, translating structural and nonstructural damage into percentage of losses, are proposed by different authors and in the framework of different seismic risk analysis methods. These correlations apparently different can be, actually, represented according to a common formulation (Giovinazzi 2005).

Having the physical damage described in terms of a discrete damage scale (see Table 6 for some examples), the probability of having consequences p_c induced by an earthquake can be obtained as

$$p_c = \sum_i w_i p_i \quad (6)$$

where p_i is the probability of having a damage grade d_i and w_i is a weight factor, specifically defined for a specific consequence c and a damage grade i .

Table 8 summarizes weight factors for the evaluation of different consequences, including (a) the ratio between cost of repairing and cost of replacement defined as mean damage ratio, MDR ; (b) the number of buildings unfit for use, UFU ; and (c) the number of severely injured people, S .

The MRD weight factors in Table 8 are intended for the evaluation of economic losses due to the physical damage of buildings. Similar correlations can be derived for building contents and for business interruption.

Summary

Earthquake risk is a public safety issue that requires appropriate risk management measures and means to protect citizens, properties, infrastructures, and the built cultural heritage. The aim of a seismic risk analysis is the estimation and the hypothetical, quantitative description of the consequences of seismic events upon a geographical area (a city, a region, a state, or a nation) in a certain period of time. The effects to be predicted are the physical damage to buildings and other facilities, the number and type of casualties, the potential economic losses due to the direct cost of damage along with the indirect economic impacts (loss of the productive capacity and business interruption), the loss of function for lifelines and critical facilities, as well as social, organizational, and institutional impacts. The results provided by a seismic risk analysis could be regarded as helpful guidelines with respect to all the phases of risk management: during normal periods, during crisis periods, as well as in the recovery and post-emergency periods.

This entry has summarized the basic steps needed for modeling earthquake-induced damage to buildings at territorial or urban scale as

part of seismic risk analysis. It has described the importance and purpose of inventorying and classifying buildings. It has explained the importance of using an appropriate damage scale for classifying physical damage to building structural and nonstructural components or for describing the expected performance objectives for the buildings. The recommended parameters for describing seismic ground motion, against which damage should be evaluated, have been introduced. The use of either probabilistic seismic hazard assessment or deterministic scenario earthquakes or a combination of the two, to evaluate ground motions for building damage assessment, has been discussed. A concise review of hazard-damage relationships has been provided along with the presentation of the basic steps for the two of them, a macroseismic- and a mechanic-based approach, for which a closed-form formulation assures an easy computational implementation within seismic risk analysis tools and platforms. For both the methods, the classification of the building stock can be performed with data of different origin and quality (e.g., from data properly surveyed to statistical poor existent data), and a different uncertainty can be associated with the resulting damage evaluation depending on the quantity and on the consistency of dataset used. Finally, a simple approach for a first estimation of the expected direct economic losses and consequences to people, based on the estimation of the physical damage to buildings, has been presented.

Cross-References

- [Building Damage from Multi-resolution, Object-Based, Classification Techniques](#)
- [Building Disaster Resiliency Through Disaster Risk Management Master Planning](#)
- [Earthquake Mitigation Strategies Through Insurance](#)
- [Earthquake Response Spectra and Design Spectra](#)

- [Earthquakes and Their Socio-economic Consequences](#)
- [Economic Impact of Seismic Events: Modeling](#)
- [Empirical Fragility](#)
- [Estimation of Potential Seismic Damage in Urban Areas](#)
- [Engineering Characterization of Earthquake Ground Motions](#)
- [Insurance and Reinsurance Models for Earthquake](#)
- [Learning from Earthquake Disasters](#)
- [Probabilistic Seismic Hazard Models](#)
- [Seismic Fragility Analysis](#)
- [Seismic Actions Due to Near-Fault Ground Motion](#)
- [Seismic Loss Assessment](#)
- [Seismic Risk Assessment, Cascading Effects](#)
- [Seismic Vulnerability Assessment: Masonry Structures](#)
- [Seismic Vulnerability Assessment: Reinforced Concrete Structures](#)
- [Site Response for Seismic Hazard Assessment](#)

References

- ATC-13 (1985) Applied technology council report. Earthquake damage evaluation data for California. Redwood City
- ATC-21 (1988) Applied technology council report. Rapid visual screening of buildings for potential seismic hazards: a handbook. Palo Alto
- Bazzurro P, Cornell CA (1999) Disaggregation of seismic hazard. *Bull Seismol Soc Am* 89:501–520
- Bazzurro P, Luco N (2005) Accounting for uncertainty and correlation in earthquake loss estimation. In: Augusti G, Schueller GI, Ciampoli M (eds) *Proceedings 9th international conference structural safety and reliability (ICOSSAR)*, 20th–23rd June 2005, Rome, Italy. Millpress, Rotterdam, pp 2687–2694
- Bommer JJ (2002) Deterministic versus probabilistic seismic hazard assessment: an exaggerated and obstructive dichotomy. *J Earthq Eng* 6:43–73
- Bramerini F, Di Pasquale G, Orsini A, Pugliese A, Romeo R, Sabetta F (1995) Assessing seismic risk for the Italian territory. Servizio Sismico Nazionale technical report, SSN/RT/95/01, Roma (in Italian)
- Brezev S, Scawathorn C, Charleson AW, Allen L, Greene M, Jaiswal K, Silva V (2013) GEM building taxonomy version 2.0. GEM technical report 2013–02, p 180

- Calvi GM, Magenes G, Pinho R (2005) Displacement based methods to predict earthquake damage at variable geographical territorial scale. In: Proceedings of 250th anniversary of the 1775 Lisbon earthquake, Lisbon
- CEN (2003) Comité Européen de Normalisation, prEN 1998-1-Eurocode 8: design of structures for earthquake resistance. Part 1: general rules, seismic actions and rules for buildings. Pr-EN 1998-1. Draft N°6, Doc CEN/TC250/SC8/N335, Jan 2003, Brussels
- Charleson A (2011) Review of existing structural taxonomies. <http://www.nexus.globalquakemodel.org/gem-building-taxonomy/posts/review-of-existing-structural-taxonomies>. Accessed Feb 2014
- Cosenza E, Manfredi G, Polese M, Verderame GM (2005) A multi-level approach to the capacity assessment of existing RC buildings. *J Earthq Eng* 9(1):1–22
- Crowley H, Bommer JJ, Stafford PJ (2008) Recent developments in the treatment of ground-motion variability in earthquake loss models. *J Earthq Eng* 12(S2):71–80
- Douglas J (2011) PEER report 2011/102: ground motion prediction equations 1964–2010. Pacific Earthquake Engineering Centre, Berkeley
- EERI (2000). World housing encyclopaedia. www.world-housing.net. Accessed Feb 2014
- Elnashai AS, Jeong SH (2005) Rapid assessment of structural system in earthquake regions. In: Proceedings of 250th anniversary of the 1775 Lisbon earthquake, Lisbon (CD-Rom)
- FEMA 154 (1988) Rapid visual screening of buildings for potential seismic hazards: supporting documentation. FEMA, Washington, DC
- Foulser-Piggott R, Spence R (2013) Extending EMS-98 for more convenient application outside Europe I: review of field experience using EMS-98: development of the international macroseismic Scale. In: Proceeding of Vienna congress on recent advances in earthquake engineering and structural dynamics 2013 (VEESD 2013)
- Gallagher H, Farmer B, Mendoza C, Lee C, Dickson H, Greene M (2013) GEM building taxonomy v2.0: evaluation and testing report, GEM building taxonomy global component. GEM foundation
- Giovinazzi S (2005) The vulnerability assessment and the damage scenario in seismic risk analysis. PhD dissertation. University of Florence (I) and Technical University of Braunschweig (D)
- Giovinazzi S, Lagomarsino S (2004) A macroseismic model for the vulnerability assessment of buildings. In: 13th world conference on earthquake engineering, Vancouver
- Giovinazzi S, Pampanin S, Lagomarsino S (2006) Alternative retrofit strategies for pre' 70 buildings: vulnerability models and a damage scenarios. In: First European conference on earthquake engineering and seismology, Geneva
- Grünthal G (1998) European macroseismic scale 1998. European Seismological Commission, Luxembourg
- Hill M, Rossetto T (2008) Comparison of building damage scales and damage descriptions for use in earthquake loss modelling in Europe. *Bull Earthq Eng* 6(2): 335–365
- Lagomarsino S, Giovinazzi S (2006) Macroseismic and mechanical models for the vulnerability assessment of current buildings. *Bull Earthq Eng (Spec Iss Risk-Ue Project)* 4(4)
- NIBS (2003) HAZUS MR4 technical manual. National institute of Building Sciences, Washington, DC
- NIST (1999) National institute of standards and technology, UNIFORMAT II elemental classification for building specifications, cost estimating, and cost analysis, NISTIR 6389, Washington, DC, p 93. <http://www.bfrl.nist.gov/oae/publications/nistirs/6389.pdf>
- Pitilakis K, Argyroudis S, Kakderi K, Gehl P, Desramaut N, Hancilar U, Taucer F, Kaynia AM, Fardis M, Tsionis G, Crowley H, Colombi M, Broglio S, Monteiro R, Franchin P, Lupoi A (2013) D.8.8 guidelines for typology definition of European physical assets. Syner-G project: systemic seismic vulnerability and risk analysis for buildings, lifeline networks and infrastructures safety gain. <http://www.vce.at/SYNER-G/pdf/deliverables>. Accessed 20 Jan 2014
- Porter KA (2005) A taxonomy of building components for performance-based earthquake engineering, Pacific Earthquake Engineering Research Center, PEER 2005/03, Berkeley, p 58
- RISK-UE (2004) An advanced approach to earthquake risk scenarios with applications to different European towns. Research project, European commission, DG XII2001-2004, CEC contract number: EVK4-CT-2000-00014
- Rossetto T, Elnashai A (2003) Derivation of vulnerability functions for European-type RC structures based on observational data. *Eng Struct* 25:1241–1263
- SEOAC (1995) Vision 2000, performance based seismic engineering of buildings. Structural Engineers Association of California, Sacramento
- Tenerelli P, Hancilar U, Taucer F, Ehrlich D, Polli D, Dell'Acqua F, Lisini G, Crowley H, Colombi M, Argyroudis S, Kakderi K, Raucoules D, Gehl P, Wenzel H, Schafer D (2013) D.8.9 development of inventory datasets through remote sensing and direct observation data for earthquake loss estimation, SYNER-G reference report 3 (D8.9)
- Wales CD, Horner AC (1933) Earthquake damage analysed by long beach building officials, Engineering Xeu's record, 25 May 1933
- Weatherill G, Crowley H, Pinho R (2011a) Efficient intensity measures for components within a number of infrastructures, deliverable D2.12 SYNER-G project: systemic seismic vulnerability and risk analysis for buildings, lifeline networks and infrastructures safety gain. University of Pavia (part of SYNER-G Consortium). <http://www.vce.at/SYNER-G/pdf/deliverables>. Accessed 20 Jan 2014

- Weatherill G, Crowley H, Pinho R, Franchin P, Cavalieri F, Iervolino I, Esposito S (2011b) A review and preliminary application of methodologies for the generation of earthquake scenarios for spatially distributed systems, D2.13. SYNER-G project: systemic seismic vulnerability and risk analysis for buildings, lifeline networks and infrastructures safety gain. <http://www.vce.at/SYNER-G/pdf/deliverables>. Accessed 20 Feb 2014
- Whitman RV, Reed JW, Hong ST (1973) Earthquake damage probability matrices. In: Proceedings of the 5th European conference on earthquake engineering, Rome, p 2531

Damage to Infrastructure: Modeling

Indranil Kongar¹ and Sonia Giovinazzi²

¹EPICentre, Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK

²Department of Civil and Natural Resources Engineering, University of Canterbury, Christchurch, New Zealand

Synonyms

Economic impact; Functional impact; Infrastructure systems; Lifelines; Physical impact; Seismic fragility; Seismic vulnerability

Introduction

Infrastructure systems, also commonly known as *lifelines*, can be defined as the “systems or networks, which provide for the circulation of people, goods, services and information upon which health, safety, comfort and economic activity depend” (Platt 1991). Infrastructure systems include utility networks such as energy, water, telecommunications, and roads or discrete critical facilities such as hospitals, ports, and airports. They are crucial to the routine functioning of society and play an important role in emergency relief, reconstruction, and recovery after an earthquake.

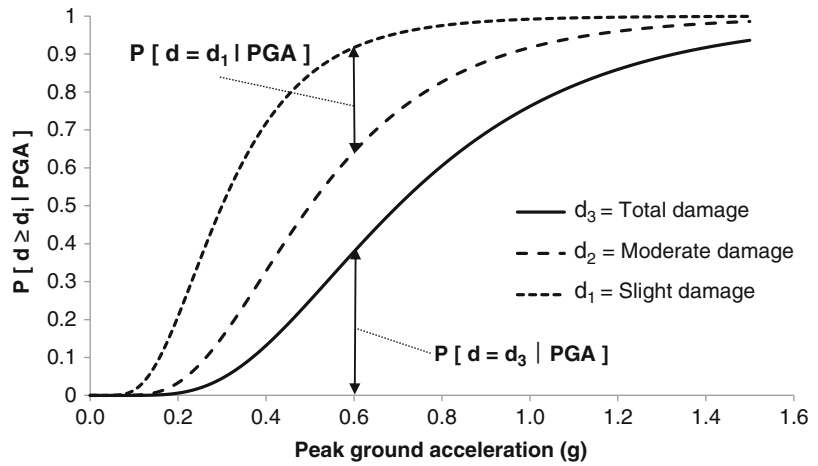
Infrastructure systems are made up of different types of components, including “point”

components, which include facilities used for generation, storage, treatment, or redistribution of a utility, and linear components that connect these facilities acting as conduits for the distribution and transmission of the service. Earthquake-induced physical damage to infrastructure generates direct losses due to the cost of repairs to these components. Furthermore operational failure resulting from physical damage diminishes the capacity of the infrastructure system to provide services to its consumers, resulting furthermore in indirect losses due to business interruption and subsequent long-term economic impacts.

The modeling of earthquake-induced damage can be either probabilistic or deterministic or, since seismic risk assessment is a multistep process, a combination of both. The benefit of probabilistic analysis is that they can account for the many uncertainties associated with seismic risk assessments of infrastructure. These include uncertainties in the quantification of seismic hazard, the relationship between seismic hazard and component damage, and the relationship between component damage and component functionality.

In the case of seismic hazard, there are uncertainties over the time, location, depth, and magnitude of future earthquakes. One approach to account for this is probabilistic seismic hazard assessment (PSHA) (Bommer 2002). PSHA maps of ground motion intensity are produced for a study area in which the values of the intensity measure at each location represent the ground motion that is predicted to occur on average once within a specified return period based on all possible earthquake sources. The earthquake consequences (e.g., damages, losses, or casualties) are computed for these ground motions. A feature of PSHA maps is that the evaluated ground motions are not necessarily predicted to occur simultaneously due to the same earthquake and therefore are not representative of a real earthquake. An alternative approach which allows for more realistic representation of individual events is to use scenario earthquakes. A single scenario earthquake does not account for the uncertainties in seismic

Damage to Infrastructure: Modeling, Fig. 1 A fabricated set of fragility curves for a fictitious structural typology



hazard. However, the distribution of uncertainties can be represented by using a large-scale simulation technique such as Monte Carlo simulation or Latin hypercube sampling to undertake analysis for multiple scenario earthquakes, each with attributes determined probabilistically from the historic earthquake catalog for the study area. In simulations such as this where multiple potential outcomes are computed, statistical measures are needed to make inferences from the results. The simulation approach to loss estimation is common in catastrophe modeling in the insurance sector (Bazzurro and Luco 2005).

There are also uncertainties in the relationship between seismic hazard and damage – two structures of the same typology will not necessarily experience the same damage if exposed to the same ground motion due to inestimable factors such as the maintenance history of a structure. A well-established tool used to assess this risk is the fragility function (or fragility curve when presented graphically) and relates damage to hazard probabilistically (Rossetto and Elnashai 2003). For a given level of ground motion intensity, fragility functions determine the probability that a structure or component will be in, or exceed, the i th damage state, d_i . The probabilistic relationship is commonly assumed to be a cumulative lognormal distribution. In this case the fragility function can be expressed conditionally in the form of Eq. 1:

$$P[d \geq d_i | IM] = \Phi \left[\frac{1}{\beta_{d_i}} \ln \left(\frac{IM}{IM_{d_i}} \right) \right] \quad (1)$$

In Eq. 1, P is the exceedance probability for the i th damage state, IM is the ground motion intensity (e.g., peak ground acceleration), Φ is the standard normal cumulative distribution function, β_{d_i} is the standard deviation of the natural logarithm (dispersion) of ground motion intensity for the i th damage state, and IM_{d_i} is the median value of ground motion intensity at which the subject reaches the i th damage state. If different damage states are defined for a component, then distinct fragility curves exist for each damage state with variation of the dispersion and median values of ground motion intensity. Therefore, the risk to a component being assessed for multiple damage states will be represented by a set of fragility curves. Fragility curves define exceedance probabilities, but it is also desirable to know the probabilities that a structure will be in specific damage states. For a specified value of ground motion, these can be calculated from the intervals between the fragility curves, so the probability that a structure is in the i th damage state is the difference between the exceedance probabilities for the i th damage state and the $i + 1$ th damage state. Figure 1 shows a set of fabricated fragility curves for a structure whose damage is defined by three damage states against peak ground acceleration. Although this form of fragility function has been widely adopted in academic literature and

by risk assessment platforms, probabilistic functions do not necessarily have to assume a cumulative lognormal shape. Moreover, fragility functions can also be deterministic, exemplified by the repair rate functions used to evaluate damage to pipelines (Kakderi and Argyroudis 2014).

Seven steps are necessary toward the modeling of infrastructure damage. The first step is to produce taxonomy of infrastructures and their constituent components so that elements that are expected to behave similarly are considered uniformly. The second step is to define the appropriate scales for classifying damage of each component. The third step is to determine the appropriate earthquake hazard parameters against which damage should be evaluated. The fourth step is to generate the appropriate ground motions necessary to assess damage. The fifth step is to identify the appropriate hazard-damage relationship and assign damage to each component based on the ground motion. The sixth step is to assess the performance of the whole infrastructure system based on the damages that have been assigned to each component within it. The final step is to model the restoration times for infrastructure components, which is necessary to evaluate indirect losses from business interruption. This process is summarized as a flow chart in Fig. 2, which shows the steps involved in the risk assessment of a single infrastructure system due to a single earthquake.

Classifying Infrastructure

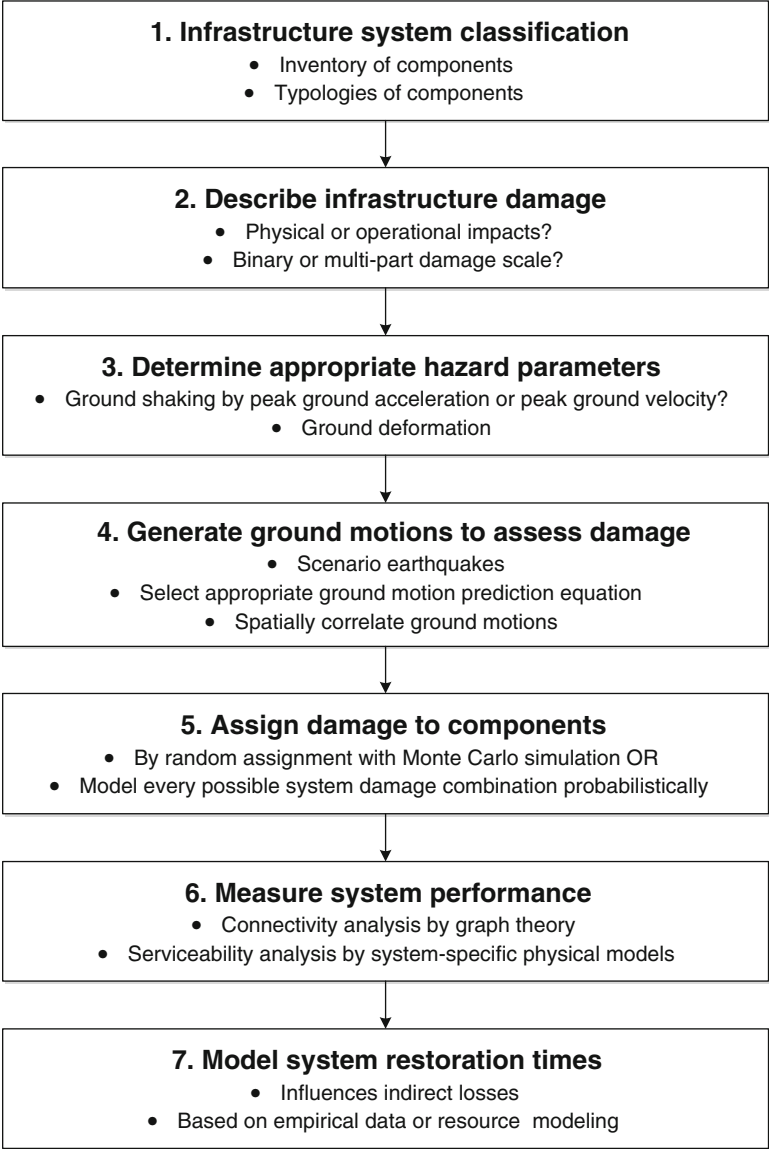
The purpose of classifying infrastructure into taxonomy is to group together elements that can be expected to behave similarly following an earthquake. While buildings are discrete structures which can be grouped together based on certain structural characteristics (e.g., material, geometry, year of construction), infrastructure systems are made up of different components whose differences are based on their function as well as their constructive characteristics. As such classification of infrastructure systems is in two parts: firstly, identification of the different

components that exist in the system and, secondly, identification of the typologies of that component.

For example, an electric power network may comprise three components: generation plants, substations, and conduits. Each of these serves a different function in the system. Generation plants are responsible for production of electric power, conduits transport the electricity from one location to another, and substations allow for voltage change and/or redistribution. Then for each of these components there are different typologies, relating to structural or operational attributes. Electrical conduits, for example, can vary based on their position (overhead lines or underground cables), their insulation material, or their size. Substations may vary by load capacity or voltage.

For infrastructure systems it is not just the damage to the individual components that is important but also how the whole system consequently performs, e.g., is a water supply system able to supply water in sufficient quantity at adequate pressure to all consumers? While damage informs direct losses, system performance is needed to inform indirect losses as well as emergency management decision-making. To ensure that system performance can be properly evaluated, classification of a single infrastructure system should therefore include all constituent components whose damage condition may affect service provision. It is essential also to understand the connections between these components and their spatial distribution relative to urban populations, since some components will be more critical than others in terms of their impact on system performance. Another feature of infrastructure systems that does not apply to building risk assessments is the interdependency between systems, for example, the reliance of the water supply system on electricity in order to power pumping stations. It is good practice to identify and document these connections in order to fully understand the potential risk to an infrastructure. Although it is possible to model all infrastructure systems as a single composite system of systems, this is computationally intensive. There is a growing field of research investigating

Damage to Infrastructure: Modeling,
Fig. 2 Procedure for assessing the risk to a single infrastructure system due to a single earthquake



modeling methods that can account for such interdependencies quantitatively, but at present little of this has been adopted in risk management practice.

Table 1 shows a fabricated example of infrastructure taxonomy, including some important infrastructure systems, their key constituent components, and component attributes. The component attributes are the descriptors that could be used to group components into distinct typologies. Components can sometimes be pieces of equipment housed inside a building. In this case

the attribute “seismic design level” refers to whether the building is seismically designed or whether the component is secured. Sometimes components are made up of a systemic arrangement of smaller subcomponents. In this case “seismic design level” refers to whether or not these subcomponents are secured to protect them from ground shaking.

An inventory is an enumeration of the components and facilities in each of the typologies considered by the assumed taxonomy. Preparation of an inventory for infrastructure can be difficult

Damage to Infrastructure: Modeling, Table 1 An example infrastructure system taxonomy

Infrastructure system	Components	Component attributes
Electric power	Generation plants	Capacity, seismic design level
	Substations	Voltage, seismic design level
	Cables	Material, size
Potable water	Wells	Seismic design level
	Water treatment plants	Capacity, seismic design level
	Pumping stations	Capacity, seismic design level
	Storage tanks	Elevation, material, geometry, quantity of contents, seismic design level
	Pipelines	Material, joint type, age, diameter
Waste water	Lift stations	Capacity, seismic design level
	Treatment plants	Capacity, seismic design level
	Pipelines	Material, joint type, age, diameter
Natural gas	Pipelines	Material, joint type, age, diameter
	Compressor stations	Capacity, seismic design level
Fuel	Refineries	Capacity, seismic design level
	Pumping stations	Capacity, seismic design level
	Storage tanks	Elevation, material, geometry, quantity of contents, seismic design level
	Pipelines	Material, joint type, age, diameter
Telecommunications	Central offices	Seismic design level
	Cables	Material, size
Highways	Roadways	Importance level
	Bridges	Structural system, material, age, geometry, seismic design level
	Tunnels	Construction method, geometry, local geology
	Embankments	Height, soil type

since attributes cannot always be identified by simple visual inspection, and in many cases the components are not visible (e.g., buried pipelines). The inventory should therefore be prepared by collecting data from available sources. In the case of infrastructure, however, usually the only source is the system operator. To protect commercial interests and due to concerns over security, this type of information is not usually available in the public domain to a high level of detail, and so depending on the granularity of the modeling exercise, it may only be feasible to work in partnership with infrastructure owners.

Some infrastructure components may need to be assessed in more detail individually because of their criticality or damage consequences. An example of this would be a nuclear power station within an electric power network, since the consequences of damage to nuclear power stations have the potential to be extremely serious over

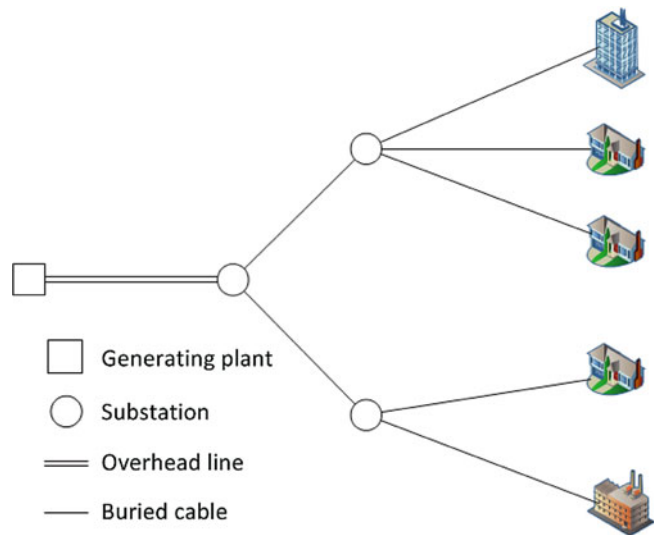
a wide geographical area. If such analysis of an infrastructure component is required, then the arrangement of its constituent subcomponents may also need to be documented in the taxonomy.

Describing Infrastructure Damage

The damage scales used to model infrastructure damage vary depending on the type of component being assessed. All networked infrastructures can be represented as an arrangement of “point” components, also known as *nodes* and linear components, also known as *edges*. The nodes are the facilities within the infrastructure system and the edges are the components that connect the facilities and connect consumers (or end users) to the service. Figure 3 illustrates a small fictitious electric power transmission system represented diagrammatically as a network, where the

Damage to Infrastructure:

Modeling, Fig. 3 A schematic diagram of a fictitious electric power transmission network. The network would continue to divide into further branches as power is distributed to customers and voltage is stepped down



D

generation plant and substations are nodes and the overhead lines and buried cables are edges. In a potable water system, the nodes could be wells, treatment plants, or pumping stations and the edges would be pipelines. This classification of nodes and edges determines the way in which damage is described. Damage to edges is usually described per unit length, whereas damage to nodes is described for the whole component.

Damage to infrastructure nodes can be described either in terms of physical damage or operational failure. Where components are housed inside buildings, it may be necessary to separately assess the operational state of the equipment and physical damage to the building. For infrastructure components housed in buildings, it is possible for the structure to be significantly damaged yet the component is fully operational as none of the equipment is damaged. Conversely it is possible for the structure to be sound yet the component does not function because equipment inside has been damaged. The purpose of infrastructure is to provide a service to consumers, and so while assessing structural damage is important for calculating direct losses, it is also important to assess the operability of the component after an earthquake, noting that the two do not necessarily correlate.

Damage to buildings is usually described in a scale of multiple increasing damage states from

no damage to total collapse. The damage scale adopted in the HAZUS (NIBS 2003) platform, for example, accounts for no damage, minor damage, moderate damage, extensive damage, and collapse. For buildings it is desirable for damage scales to correspond to damage limit states including serviceability, damage control, and collapse prevention (Rossetto and Elnashai 2003). Damage scales for equipment should relate to its operational state or functionality since this is most critical for evaluating indirect losses. The difficulty in identifying distinct physical damage states for equipment means that damage to equipment is often assessed only in terms of its operational state. The post-earthquake functionality can be expressed simply in binary terms, i.e., whether the equipment is operational or has failed, which is equivalent to a single damage state, or a set of fragility curves may be used to represent varying degrees of functionality.

While damage to infrastructure nodes is usually classified qualitatively, damage to infrastructure edges can be assessed quantitatively. The only type of infrastructure edge that has been studied extensively in terms of seismic risk is pipelines, which exist in water supply, waste water, natural gas, and fuel delivery networks. Fragility curves are not used to assess risk to pipelines, and instead the common parameter for pipeline damage is the repair rate. The repair

rate is a deterministic calculation of the mean number of damages that the pipeline is expected to experience per unit of length, usually per kilometer. For the damages, no distinction is made between breakages or leakages. The relationship between repair rate and earthquake hazard commonly follows a power law relationship although more complex functions do exist (Kakderi and Argyroudis 2014). Most repair rate functions have been derived empirically, and the typical form of a power law repair rate function is shown in Eq. 2, where RR is the repair rate, IM is the earthquake hazard parameter, and a and b are coefficients determined using some regression technique (ALA 2001):

$$RR = a \times IM^b \quad (2)$$

To account for different material properties or soil conditions, the repair rate function may include additional multiplying factors which vary according to attribute, or a set of functions may be proposed for different conditions.

Selecting Appropriate Earthquake Hazard Parameters

Ground shaking hazard is commonly represented by either peak ground acceleration (PGA) or peak ground velocity (PGV). It is usual for above-ground components such as substations to be assessed against PGA, while buried components such as pipelines are assessed against PGV. Various studies (e.g., Isoyama et al. 2000; O'Rourke et al. 2001) have shown that pipeline damage correlates better with PGV. This may be because PGV relates to ground strains, which are more relevant to buried components, whereas PGA relates to inertia forces, which are more relevant to above-ground structures (Pineda-Parros and Najafi 2010).

Infrastructure risk can also be assessed for earthquake-induced ground deformation hazard such as liquefaction or lateral spreading, although this is less common since the prediction of permanent ground deformation is more complex. There are very few predictive relationships for

permanent ground deformation (Tromans 2004), and those that do exist often require detailed geotechnical data (Bird et al. 2006). While this may be feasible for single-point analysis such as for buildings, it is not always practical to obtain this data across a spatially distributed infrastructure system. Furthermore, even if ground deformation can be predicted, there are very few relationships that can then be used to relate this to infrastructure damage (Bird and Bommer 2004).

Generating Ground Motions

For modeling building damage, ground motion intensities are usually evaluated by probabilistic seismic hazard assessment (PSHA) to account for the uncertainty in the attributes of potential earthquakes over the assessment period. However, this method is not suitable for spatially distributed infrastructure systems. Since the ground motion intensity map produced by PSHA is a composite statistical representation of multiple possible earthquakes, the ground motions do not replicate those that might be produced by a single real earthquake. However, a true earthquake representation is important for infrastructure systems. In order to evaluate the performance of a system after an earthquake, it is necessary to know the damage state and functionality at each component simultaneously and hence also the ground motion at each site simultaneously. Because PSHA ground motions are usually evaluated for long return periods, they are conservative compared to the ground motions that would be generated by most possible earthquakes in the region. This increases the likelihood that the model will predict component failures, leading to larger reductions in system performance and a conservative estimation of indirect losses accrued.

Infrastructure systems require a method that can evaluate seismic hazard simultaneously over a large area and should therefore be modeled using scenario earthquakes, with Monte Carlo simulation or a similar technique used to account for the distribution of potential earthquake

attributes. In the Monte Carlo simulation, multiple earthquake events are simulated, the attributes of which are assigned randomly but probabilistically based on data from historic earthquakes in the region. The result is a hybrid between deterministic and probabilistic hazard analysis (Bommer 2002). Damage, operational performance, and losses are assessed for each event, generating a distribution of impacts. Statistical measures can be used to make inferences from the distribution, ranging from very simple measures such as median or percentile values to more complex statistical techniques. An important criterion for this type of analysis is that the number of events in the simulation is sufficiently large to represent the likely distribution of attributes.

For scenario earthquakes, ground motions can be evaluated using ground motion prediction equations (GMPEs). However, since GMPEs are derived empirically with resulting uncertainties, GMPEs contain a deterministic part, which calculates the median value of ground motion and added to this are one or two random variables representing the variability that occurs within a single event (intra-event) and between separate events (inter-event) (Douglas 2011). Therefore, when GMPEs are applied to a large area, adjacent ground motions do not correlate with each other and the resulting ground motion map is random. This is not a problem for analysis at a single site, but for analysis of spatially distributed systems, the requirement for simultaneous ground motions at multiple sites means that the ground motions must be spatially correlated to more realistically represent an earthquake scenario. It is possible to generate a spatially correlated ground motion map by ignoring the random variables, hence using only median ground motions, or assuming a constant value for the random variables. This approach is straightforward, but the distribution of uncertainties is then not represented. More complex geostatistical methods can be used to empirically determine the spatial correlation of ground motions in a region based on historic data and apply them to simulated earthquake scenarios. These are not discussed here but one method for

producing spatially correlated ground motion fields is included in the European seismic risk platform, SYNER-G, and is described by Weatherill et al. (2014).

Assigning Damage to Infrastructure Components Based on Ground Motions

D

The method for assigning damage to infrastructure components for a scenario earthquake varies depending on whether the relationship between hazard and component damage is defined by a probabilistic fragility function or by a repair rate function. For a component whose damage is defined by a fragility function, described below are two approaches that can be applied to ensure the assignment of damages to each component reflects the distribution of damages defined by the fragility function.

The taxonomy produced in the first step of modeling process has already defined and grouped the different typologies of all components by attributes. The first method for damage assignment involves subdividing the typologies further by grouping together the occurrences of each typology that are exposed to the same ground motion. If there are a large number of components, it may be possible to only group occurrences exposed to exactly the same level of ground motion, but if they are few in number, then it may be necessary to define ground motion intervals and group together the occurrences that fall within the same interval. With the components grouped by attributes and ground motion, it is possible to assign damage states to all components that reflect the probabilistic distribution of the relevant fragility functions at the corresponding ground motion level. So, for example, if the fragility function defines that 20 % of a particular component typology experiencing a PGA of 0.5 g should fail, then 20 % of the components of that typology being exposed to 0.5 g should be assigned to the “failure” damage state. This is repeated for each damage state and then for groups of the same typology exposed to different ground motions. The whole process can then be repeated for other

typologies, resulting in a numerical count of the different component typologies in each damage state.

To measure system performance in infrastructure systems, however, it is important to know not just how many components of a particular typology are damaged but also specifically which ones, since different occurrences of a component will have different systemic influence either because of varying service area sizes or different levels of local network redundancy. An example might be an electric power network with two generation plants, one which serves 75 % of a city and another which serves the other 25 %. This method for damage assignment might tell us that one of these has failed but not which one, with the two possibilities producing very different outcomes in terms of system performance impacts. It is possible to assign the damage states to specific components on a random basis, but in that case all assignment permutations should be modeled for each scenario earthquake to ensure that the distribution of potential impacts is represented. The system performance and loss calculations should then also be computed for each of these permutations.

An alternative method for damage assignment from fragility functions involves modeling each component individually rather than in groups. The first step in this procedure is to assign a random number to each component by sampling from a uniformly distributed variable, u , where $0 \leq u \leq 1$. The random number is then compared to the damage state exceedance probabilities defined by the fragility function at the corresponding ground motion level for that component. The condition described in Eq. 3 can then be used to assign the damage state. In Eq. 3, u_j is the sampled random variable for component j and P is the exceedance probability of i th damage state, d_i , for component j at the corresponding ground motion intensity, IM . So the component is assigned to a specific damage state if the sampled random number is less than or equal to the exceedance probability for that damage state but greater than the exceedance probability for the next most damaging state.

$$P[d \geq d_{i+1}|IM] < u_j \leq P[d \geq d_i|IM] \quad (3)$$

This procedure is repeated for every component individually, but this represents just one possible system damage response. To reflect the full distribution of possible responses, a technique such as Monte Carlo simulation or Latin hypercube sampling should be used to repeat the damage assignment for all components and generate multiple system damage responses. The system performance and loss calculations should be calculated for each system damage response.

For pipelines, whose damage is defined by a repair rate function, occurrences of the same typology exposed to the same ground motion are grouped together. The appropriate function for that typology is then used to calculate the average repair rate for that group. It is commonly assumed that the location of repairs follows a Poisson distribution (e.g., Hwang et al. 1998; Adachi and Ellingwood 2008). The Poisson distribution can be used to determine the probability that a specified number of repairs will occur on a particular section of pipe based on its length and repair rate. The expression for the Poisson distribution is shown in Eq. 4, where n is the number of repairs, RR is the repair rate, and l is the pipe length:

$$P[N = n] = e^{-RR \times l} \frac{(RR \times l)^n}{n!} \quad (4)$$

Adachi and Ellingwood (2008) proposed a simple method to determine pipeline failure. For a section of pipeline, Eq. 4 can be used to determine the probability of at least one repair. Then a random number is sampled from a uniformly distributed random variable between 0 and 1. If the random number is below this probability, then that section of pipeline is assumed to have failed. This is repeated for all sections of pipeline within a group and then for all groups to simulate a pipeline system response. To reflect the probability distribution, however, multiple pipeline system responses should be generated using a Monte Carlo simulation technique.

A more detailed analysis can be undertaken to determine how many repairs occur on each section of pipe, which is useful for evaluating flow reductions. In a Poisson process the probability distribution of the interval between events can be described exponentially. Therefore by repeated sampling from an exponentially distributed random variable with a rate parameter equal to the repair rate, it is possible to simulate the locations of repairs along a section of pipe (Hwang et al. 1998). The sampling continues until the sum of all the length intervals between repairs exceeds the length of pipe being assessed. Alternatively the number of repairs on each pipe section can also be evaluated directly by using Poisson distribution random number generation with the repair rate and pipe length as parameters. The repairs can be further classified as either a break or a leak by sampling from a uniformly distributed random variable between 0 and 1 and comparing this value to the corresponding probabilities for breaks and leaks, which are deduced empirically from historic data. Since the simulation of the number of repairs on each pipe section is just one possible network response, the simulation should be repeated multiple times.

Measuring System Performance Based on System Damage Response

There are two paradigms for measuring system performance: connectivity analysis and serviceability analysis. Connectivity analysis determines whether two points in the system remain connected and so can be used to determine whether a demand node (customer) remains connected to a source node. Serviceability analysis determines not just whether a customer is connected to a service but also what the quality of that service is.

To determine the connectivity between two points only requires knowledge of the damage or failure state of all the components between those points. If there is a path from a source node to a customer node in which all components are functional, then a connection exists. For large networks, however, it may not be feasible to

manually identify connections between all source-demand pairs, and computational techniques may be required. A simple metric of connectivity for infrastructure is to count the number of customers who are disconnected from an infrastructure system after an earthquake. However, since infrastructure systems can be modeled as networks of nodes and edges, it is possible to apply a field of mathematics known as graph theory to compute more complex metrics of connectivity that capture phenomena such as redundancy and path length (Duenas-Osorio et al. 2007), although these are not discussed further here.

For serviceability analysis, each infrastructure system will have its own unique metrics and methodologies for evaluating system performance in terms of service flow levels or quality. In general physical flow evaluations can be used to quantify the post-earthquake capacity of an infrastructure system. For the electric power system, for example, it is possible to determine flow at each node by solving a set of power flow equations simultaneously to find an equilibrium state across the network (Ang et al. 1996), while hydraulic analysis can be used for flow calculations in the potable and waste water systems. The methods for each individual infrastructure system are not described here but they are based on established physical models which can be found by consulting relevant texts.

A simple serviceability metric that can be applied to all infrastructure systems is to compare capacity or supply to customer demand as a simple ratio. Another type of metric is to determine whether an infrastructure system is able to provide service at some threshold condition, e.g., potable water at sufficient pressure and quality or electric power at sufficient voltage. Table 2 lists some examples of metrics that may be used to measure serviceability for different infrastructure systems, although this list is not exhaustive.

Irrespective of the system performance metric selected, it should be evaluated for all system damage responses that exist for each scenario earthquake. The final calculations for loss estimation should include appropriate statistical methods to make use of the full distribution of impacts that this generates.

Damage to Infrastructure: Modeling, Table 2 List of serviceability system performance metrics for infrastructure systems

Infrastructure system	Serviceability metrics
Electric power	Power supply/demand ratio, voltage reduction
Potable water	Water supply/demand ratio, hydraulic pressure reduction
Waste water	Capacity/demand ratio, probability of overflow
Natural gas	Gas supply/demand ratio
Fuel	Fuel supply/demand ratio
Telecommunications	Probability of call not being connected
Highways	Travel time delay, increase in travel distances

Restoration Functions

The purpose of restoration functions is to evaluate the time it might take for damaged infrastructure components to be repaired. This is important for determining indirect losses, which are affected by the outage time of infrastructure systems. A restoration function will provide information on the average repair time for a component based on its damage state or the average percentage of repairs that will take place within a specified time period (Cagnan et al. 2006). Such functions may be derived empirically or be based on assumptions for available resources and work rate. If derived empirically there will be some uncertainty associated with the function, and this should be accounted for in its application. Figure 4 shows a fabricated example of a component restoration function. By evaluating the repair times for components, it is possible to reevaluate system performance metrics at time steps after the earthquake based on improved system conditions. This in turn means that business interruption impacts and indirect losses can be better estimated.

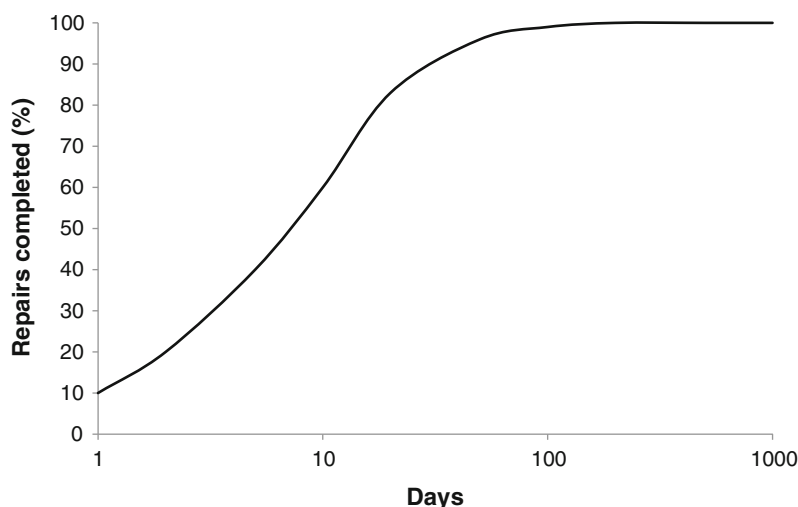
Summary

This entry has summarized the importance of modeling earthquake-induced damage to infrastructure systems and the steps involved

including the uncertainties in the process. It has described the purpose of infrastructure system classification, the types of infrastructure components that exist, and the procedure for creating an inventory of infrastructures. It has explained the different uses of probabilistic fragility curves and deterministic repair rate functions to evaluate damage for infrastructure nodes and edges, respectively, as well as the differences between descriptions of physical damage and operational failure. Damage to infrastructure can be compared to both peak ground acceleration and peak ground velocity, with the former most relevant for above-ground components and the latter most appropriate for buried components. Although many infrastructure components are susceptible to damage caused by ground deformation, there are no established methods for predictive modeling of this risk. Although it is common to use probabilistic seismic hazard assessment to evaluate ground motions for building damage assessment, for infrastructure systems, because of their systemic and spatially distributed nature, it is more valuable to use realistic scenario earthquakes reinforced by simulation techniques to account for uncertainties in earthquake attributes if this is required. The systemic nature of infrastructure means that it is important to know which, and not just how many, components are damaged. Damage can be assigned to specific components by either grouping components together by typology or by considering each component individually, but in both cases uniform random assignment is used and so a simulation technique is required to account for uncertainties. For pipelines it is often assumed that the distribution of damage follows a Poisson distribution. Similar methods involving random assignment and simulation techniques are used to locate damages making use of exponentially distributed rather than uniformly distributed variables. Because of the uncertainties in damage assignment, there should be multiple damage state scenarios for each earthquake. System performance can be measured either in terms of connectivity between source nodes and consumers or as a measure of the level of service being provided. Connectivity analysis is more straightforward

Damage to Infrastructure:

Modeling, Fig. 4 An example of a restoration function to evaluate repair time of infrastructure components



D

since it only requires knowledge of the network topology and does not require physical models to be solved. However level of service analysis can generate a more informative view of the impacts of damage to the system. Finally restoration functions can be used to estimate the length of time it will take to repair the components. Since this method for modeling infrastructure damage uses simulation techniques for both earthquake hazard and damage state assignment, it will not produce a single outcome but rather a distribution of potential outcomes. Statistical measures and techniques therefore need to be applied to make inferences from this distribution.

- [Probabilistic Seismic Hazard Models](#)
- [Resiliency of Water, Wastewater, and Inundation Protection Systems](#)
- [Seismic Actions Due to Near-Fault Ground Motion](#)
- [Seismic Design of Pipelines](#)
- [Seismic Fragility Analysis](#)
- [Seismic Loss Assessment](#)
- [Seismic Vulnerability Assessment: Lifelines](#)
- [Seismic Vulnerability Assessment: Reinforced Concrete Structures](#)
- [Site Response for Seismic Hazard Assessment](#)

Cross-References

- [Earthquake Risk Mitigation of Lifelines and Critical Facilities](#)
- [Economic Impact of Seismic Events: Modeling](#)
- [Empirical Fragility](#)
- [Estimation of Potential Seismic Damage in Urban Areas](#)
- [Insurance and Reinsurance Models for Earthquake](#)
- [Liquefaction: Performance of Building Foundation Systems](#)
- [Post-Earthquake Diagnosis of Partially Instrumented Building Structures](#)

References

- Adachi T, Ellingwood BR (2008) Serviceability of earthquake-damaged water systems: effects of electrical power availability and power backup systems on system vulnerability. *Reliab Eng Syst Saf* 93:78–88
- ALA (2001) Seismic fragility formulations for water systems: part 1 – guidelines. American Lifelines Alliance, Federal Emergency Management Agency, Washington D.C.
- Ang AH, Pires JA, Villeverde R (1996) A model for the seismic reliability assessment of electric power transmission systems. *Reliab Eng Syst Saf* 51:7–22
- Bazzurro P, Luco N (2005) Accounting for uncertainty and correlation in earthquake loss estimation. In: Augusti G, Schueller GI, Ciampoli M (eds) *Proceedings 9th international conference structural safety and reliability (ICOSSAR)*, 20–23 June 2005, Rome. Millpress, Rotterdam, pp 2687–2694

- Bird JF, Bommer JJ (2004) Earthquake losses due to ground failure. *Eng Geol* 75(2):147–179
- Bird JF, Bommer JJ, Crowley H, Pinho R (2006) Modeling liquefaction-induced building damage in earthquake loss estimation. *Soil Dyn Earthq Eng* 26(1):15–30
- Bommer JJ (2002) Deterministic versus probabilistic seismic hazard assessment: an exaggerated and obstructive dichotomy. *J Earthq Eng* 6:43–73
- Cagnan Z, Davidson R, Guikema SD (2006) Post-earthquake restoration planning for Los Angeles electric power. *Earthq Spectra* 22(3):589–608
- Douglas J (2011) PEER report 2011/102: ground motion prediction equations 1964–2010. Pacific Earthquake Engineering Center, Berkeley
- Duenas-Osorio L, Craig JI, Goodno BJ (2007) Seismic response of critical interdependent networks. *Earthq Eng Struct Dyn* 36:285–306
- Hwang HM, Lin H, Shinozuka M (1998) Seismic performance assessment of water delivery systems. *J Infrastruct Syst* 4:118–125
- Isoyama R, Ishida E, Yune K, Shirozu T (2000) Seismic damage estimation procedure for water supply pipelines. In: Proceedings 12th world conference on earthquake engineering, 30th Jan–4th Feb 2000, Auckland, New Zealand. New Zealand Society for Earthquake Engineering, Upper Hutt
- Kakderi K, Argyroudis S (2014) Fragility functions of water and waste water systems. In: Pitilakis K, Crowley H, Kaynia A eds. SYNER-G: Typology definition and fragility functions for physical elements at seismic risk, Geotechnical, geological and earthquake engineering, vol 27. Springer, Dordrecht, Ch. 8, pp. 221–258
- NIBS (2003) HAZUS MR4 technical manual. National Institute of Building Sciences, Washington, DC
- O'Rourke TD, Stewart HE, Jeon S-S (2001) Geotechnical aspects of lifelines engineering. *Geotech Eng* 149(1):13–26
- Pineda-Parros O, Najafi J (2010) Seismic damage estimation for buried pipelines: challenges after three decades of progress. *J Pipeline Syst Eng Pract* 1:19–24
- Platt RH (1991) Lifelines: an emergency management priority for the United States in the 1990s. *Disasters* 15(2):172–176
- Rossetto T, Elnashai A (2003) Derivation of vulnerability functions for European-type RC structures based on observational data. *Eng Struct* 25:1241–1263
- Tromans I (2004) Behaviour of buried water supply pipelines in earthquake zones. PhD thesis, Imperial College, London
- Weatherill G, Esposito S, Iervolino I, Franchin P, Cavalieri F (2014) Framework for seismic hazard analysis of spatially distributed systems. In: Pitilakis K, Franchin P, Khazai B, Wenzel H eds. SYNER-G: Systemic seismic vulnerability and risk assessment of complex urban, utility, lifeline systems and critical facilities, Geotechnical, geological and earthquake engineering, vol 31. Springer, Dordrecht, Ch. 3, pp. 57–88

Design of Cable-Supported Bridges: Control Strategies

Konstantinos Gkoumas, Francesco Petrini and Franco Bontempi

Department of Structural and Geotechnical Engineering, Sapienza University of Rome, Rome, Italy

Synonyms

Cable-supported bridges; Cable-stayed bridge control; Performance-based design; Structural control; Suspension-bridge control

Introduction

The introduction of cable-supported bridges in modern-day bridge engineering, in the form of cable-stayed and suspension bridges, allowed to overcome spans unimagined decades ago. This was helped by advances in construction, materials, equipment, and numerical methods. Nowadays, the implementation of cable-supported structural schemes allows to span lengths in the range from 200 m to 2,000 m (and beyond), thus covering approximately 90 % of the present span range (Gimsing and Georgakis 2012).

Cable-supported bridges are basically divided into two types: cable-stayed and suspension bridges. Although both have as common factor the cable support, their geometry and structural configuration are radically different.

Cable-stayed bridges, in which the bridge deck is supported by straight cables directly by the pylons, emerged in their modern form initially in Europe after the Second World War. Since then, their span has been increased significantly. As of 2014 the bridge to Russky Island (Russia) is the world's longest cable-stayed bridge, with a 1,104 m long central span, followed closely by the Sutong Bridge (China) which spans 1,088 m, some four times the span of the 1958 Theodor Heuss Bridge over the Rhine, an epitome of the modern cable-stayed scheme.

Even though evolution in cable materials towards strength and weight is nowadays impressive, cable-stayed bridges are limited by the stiffness of their superstructure. Simply put, the weight of the bridge superstructure would create a sag in the cable, leading to nonlinear behavior. For this reason, for longer spans, the suspension scheme, in which the deck is supported by suspension cables on vertical hanger ropes, is preferred. The suspension bridge with the longest central span in the world is the Akashi Kaikyo Bridge in Japan, completed in 1998, with a central span of 1,991 m. The feasibility of longer spans has been demonstrated, with different design paradigms at a conceptual or preliminary stage.

Nowadays, cable-suspended bridges are designed in different schemes (e.g., self-anchored, multi-span), in different materials (essentially steel and concrete), and with different deck types (e.g., truss, plate). Several variations to the structural form have been successfully built as a variation of either forms (self-anchored suspension bridges, side-spar cable-stayed bridges, etc.) or as a combination (cable-stayed-suspension hybrid bridges).

In general, cable-supported bridges, due to the fact that they are supported by cables which constitute elastic supports, behave very well under earthquake excitation. Nevertheless, structural control for seismic loads is equally a concern for seismically active regions (with several cases in the USA, Japan, China, Turkey, and Greece). The issue is the structural behavior to long-period and long-duration ground motion. The principal seismic concern is the pounding between the deck and the abutments. This also accounts for the differential motion at the support system (towers), influenced also by the spatially varying soil conditions.

Performance Requirements and Design Issues

Both cable-stayed and suspension bridges have unique characteristics compared to ordinary structures and are characterized by a flexible

design, which mainly depends on the specific structural configuration, and by a low damping of the stay cables or hanger ropes. The flexible design may lead to unwanted motion or vibrations that could compromise the bridge safety and serviceability performance and even lead to immediate structural failure and also have an influence on the fatigue performance. Excitations due to wind, earthquake, or anthropic loads (railway, traffic, pedestrian loads, etc.) are the principal causes of unwanted vibrations.

Unlike conventional bridges, cable-suspended bridge design is not covered by bridge codes, standards, and specifications. Nowadays, performance-based design (PBD) is mainly employed for the seismic design of cable-suspended bridges (Bontempi 2006).

The dynamic wind action is the principal design concern (especially for longer spans), since it affects different parts of the bridge (deck, pylons, main cables, and hanger ropes of a suspension bridges; stay cables of a cable-stayed bridge). Thus, the principal concern is the alleviation of wind-induced vibration. The commonly observed aeroelastic phenomena (wake galloping, torsional and coupled flutter, and vortex-induced vibrations) have a strong influence on the bridge behavior and are usually suppressed or controlled by appropriate design measures and control strategies. Furthermore, bridge substructures and components have been shown to act as bluff bodies, resulting in self-excited vibration of the bridge.

Vibrations due to anthropic (pedestrian, railway, and traffic) loads can be of issue on a case-by-case basis, depending on the bridge type and use. Other additional effects can be an issue (e.g., the rain-induced vibration that mostly affects cables). Design solutions and countermeasures for these vibration effects have been applied, after meticulous studying during the years, starting from notable disasters and failures (e.g., the wind-induced collapse of the Tacoma Narrows Bridge in 1940 caused by aerodynamic flutter or the synchronous lateral excitation of the pedestrian London Millennium Footbridge in 2000). In the latter, frame-mounted viscous dampers and tuned mass

vibration absorbers were retrofitted after the initial undesirable performance of the bridge.

Regarding earthquake excitation, due to its (typically) very long fundamental periods, seismic behavior of cable-supported bridges is in general satisfactory and easily reparable with extraordinary repairs (replacement of fuse restrainers, etc.). Seismic loads have in general a direct influence on the foundations and towers, while the deck is indirectly influenced.

For seismic design, performance-based earthquake engineering (PBEE) design entitles bridge owners and stakeholders to decide on the behavior of the suspension bridge under different design solutions in terms of seismic vulnerability assessment. Usually, two seismic hazard levels are identified, with performance criteria related to safety or serviceability: the design earthquake and the maximum considered earthquake. These hazard levels are characterized by different probabilities of exceedance (or return periods) for the earthquake. All other hazards (e.g., high wind) are fulfilled also on the basis of appropriate limit states (e.g., ULS, Ultimate Limit States; SLS, Serviceability Limit States), while additional limit states (e.g., durability, fatigue, impact, etc.) are appropriately defined on the basis of the tender design document and the bridge site.

For long-span bridges, one of the principal design issues is the asynchronous earthquake excitation. While a normal structure frequently has a rather limited imprint, a long-span bridge is supported by towers placed sometimes hundreds of meters away. Thus, the seismic excitation from pylon to pylon can vary in time and intensity, and a uniform ground motion assumption is not realistic anymore.

Although the phenomenon is complex, the spatial variation of an earthquake mainly is divided for simplification in three effects and conditions: the geometric incoherence effect, the wave passage effect, and the local soil conditions. The asynchronous earthquake motion in bridges has shown to have an effect on the deck displacement during earthquake, which in cable-supported bridges is in part mitigated by the cables.

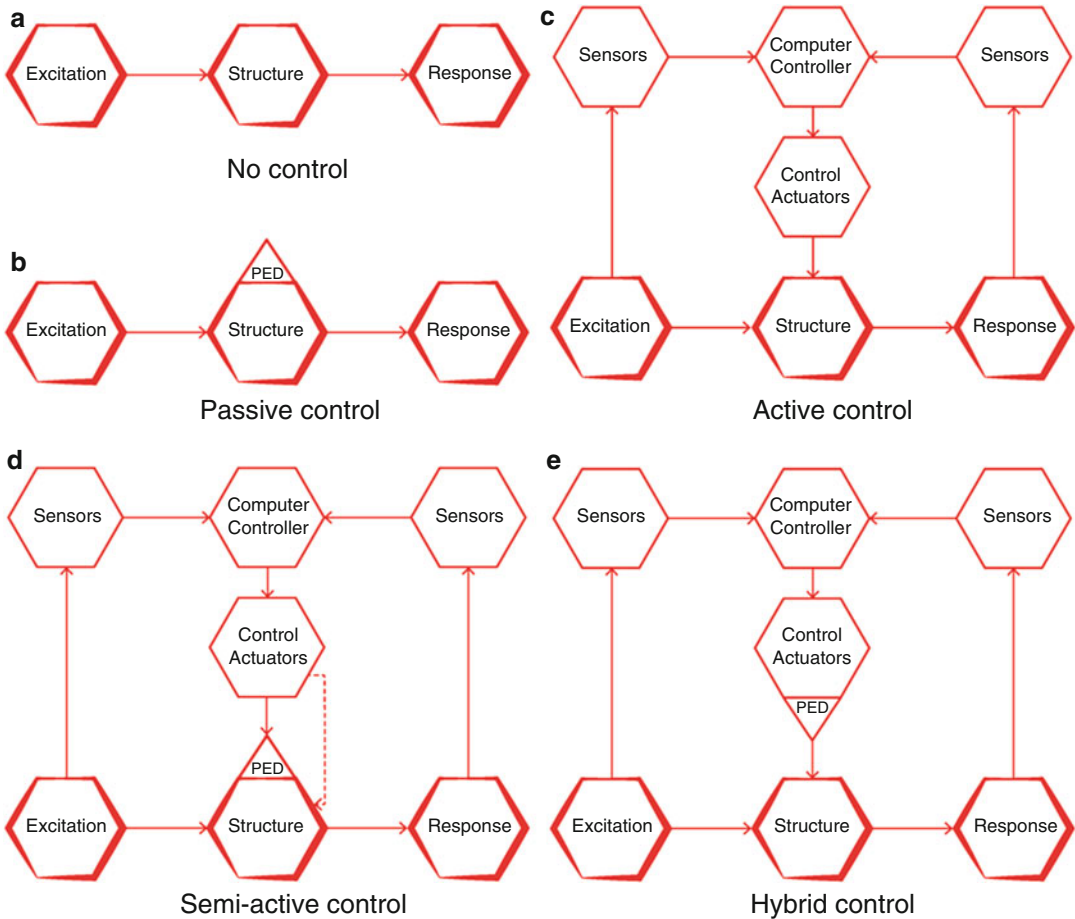
There are reports of ordinary bridge failures that may have been caused or aided by differential support motion during earthquakes (Burdette et al. 2008). However, it is difficult to fully attribute these failures to asynchronous earthquake motion, even though the phenomenon contributes to large longitudinal displacements between pylons. Nevertheless, different numerical analyses have shown that in long-span suspended bridges, the spatially varying ground motion may result in bigger structural response than the synchronous ground motion.

Structural Control

The mitigation of extensive vibration or motion has been a concern in the civil engineering field since the early conception of complex structural systems (high-rise buildings, long-span bridges), characterized by nonlinear behavior. The issue is to control how input energy (from strong wind, earthquake, etc.) is absorbed by a structure, by means of different methods and techniques, as an alternative to conventional design methods based on ductile response. Most commonly implemented methods include installing isolators or passive energy dissipation devices to dissipate vibration energy and reduce dynamic responses. In addition to these, with the advent of advanced computational methods (hardware and software), it is now possible to alter the structural configuration for mitigating the induced energy. In long-span bridge engineering, recent bridges mostly in Japan (e.g., the 1993 Rainbow Bridge in Tokyo) deploy such more advanced, “active” control strategies, especially for the transitory phase of construction.

Structural Control Strategies

In the civil engineering field, the term “structural control” with explicit reference to “active control” was introduced in 1972 by Yao (1972). Since then, the increase in the conception and design of complex structural systems and the increased demand for structural performance lead to numerous studies, design cases, and applications of structural control for such structures.



Design of Cable-Supported Bridges: Control Strategies, Fig. 1 Different control strategies (Adapted from Spencer and Soong 1999)

These were also motivated by devastating disasters through the years (earthquakes, tsunamis, typhoons, etc.) that put in evidence the issues of structural safety and reliability.

Structural control systems can be grouped into four broad categories (Housner et al. 1997):

- Passive control
- Active control
- Semi-active control
- Hybrid control

Figure 1 (adapted from Spencer and Soong 1999) provides an overview of the different control strategies, together with the no-control case (Fig. 1a).

Passive control is achieved by employing a passive energy device (PED), e.g., a Tuned Mass Damper or mechanical energy dissipation tool, that responds to the motion of the structure to dissipate energy or by isolating the structure from the ground and horizontal loads (base isolation systems). Passive control is employed exclusively by the motion of the structure, and no external force or energy is applied to effect the control (Fig. 1b). It is nowadays a broadly accepted vibration control strategy, with many consolidated cases worldwide, while in the recent years, innovative approaches have been applied in long-span bridge engineering. Typical applications of passive control in cable-supported bridges are in the form of dampers, placed principally

on the deck, pylons, and the hangers, and base isolation schemes on the foundations.

Active control, on the other hand, is a response-triggered control method. In other words, an external force is applied to the structure every time a displacement or vibration is measured by means of a control actuator. The latter takes input from sensors that measure the external excitation (open-loop control) and, in some cases, the structural response (closed-loop control) – (Fig. 1c). Although in the last three decades numerous methods have been proposed and many researches have been carried out, practical applications in cable-supported bridges are rather limited due to cost and reliability issues. Most of the applications regard pylon control during erection, as in the case of the Akashi Kaikyo Bridge. In fact, a change in the pylon height leads to a change in its natural frequency, something that poses as a challenge to deal with using a small number of mass dampers.

In between, two other methods have been widely studied (and applied in a few cases worldwide) that try to combine advantages and exclude shortcomings of the abovementioned control strategies.

In **semi-active control**, the control actuator does not directly apply force to the structure (as in the case of active control), but instead it is used to control the properties of a passive energy (Fig. 1d). Within the class of semi-active control, the use of controllable electrorheological (ER) and magnetorheological (MR) fluids in dampers has found ground in the recent years, with applications mostly on the stay cables of cable-stayed bridges.

Finally, **hybrid control** consists in the combined application of both passive and active methods. This means that a part of the control derives from a passive control device, something that alleviates the required effort in terms of forces and energy for the active device (Fig. 1e). Hybrid control is applied as an alternative to active control during the erection of towers of suspension bridges. Spencer and Nagarajaiah (2003) report several cases of hybrid control in bridges during tower erection.

Compared to pure active control, semi-active and hybrid control strategies have the advantage

that the passive control part can function also during a power failure, thus providing a minimum degree of protection.

Considering the above, there are significant advantages and disadvantages for each control method. Also, optimal fields of application can be significantly different: in general, consolidated passive methods are preferred for their cost-effective solution and reliability. However, in the last years, alternative active, semi-active, or hybrid methods are implemented in case a passive method results inadequate (two cases worth mentioning are the implementation of active methods as a support to passive ones during the tower erection and the use of semi-active tendon control). Finally, different control methods and devices are preferred for different parts of the structure.

Structural Control Issues

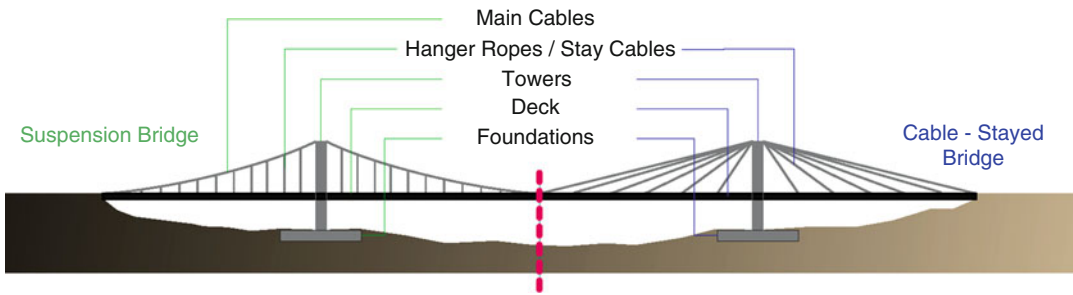
Taking into account what was said above, the following sections overview control devices and strategies with respect to different parts of cable-supported bridges. On the basis of consolidated cases, Fig. 2 synthesizes the control options.

It should be noted that many of the issues can be dealt with an improved design or structural modifications. This is the case, for example, for flutter- and vortex-induced vibration suppressions on bridge deck, obtained by means of section modifications or aerodynamic countermeasures (e.g., deflectors, fairings, and flaps). Similarly, an appropriate designed cable cross section and surface can minimize cable vibration.

Foundation Isolation

Foundations transmit lateral loads to the entire bridge through the connection to the other parts (principally deck and cables). The typical control strategy in long-span bridge design is base isolation. Base isolation, that is, to decouple the bridge superstructure from its foundations, is now considered a mature technology. A modern interpretation of this strategy has been applied in the recent years in several long-span suspension and cable-stayed bridges.

Two notable cases are the 2004 Rio-Antirrio multi-span cable-stayed bridge in Greece and the



Design of Cable-Supported Bridges: Control Strategies, Fig. 2 Control strategies for different substructures of cable-supported bridges

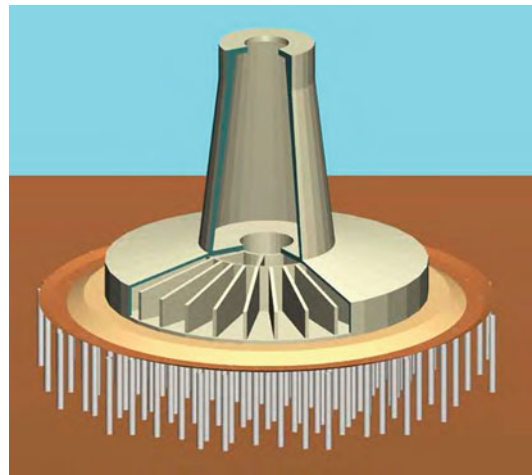
currently under construction Izmit Bay Suspension Bridge in Turkey. Although both bridges are entirely different for conception and structural scheme, a similar base isolation system is applied to the tower foundations of both of them.

The Rio-Antirrio cable-stayed bridge spans approximately 2.9 km, including the approach viaducts, with the three central spans being 560 m long and the adjacent spans 286 m, all supported by four towers. The bridge is designed to withstand earthquakes with a P.G.A. of 0.48 g and tectonic movements up to 2 m between consecutive pylons (Jacques 2011). It is positioned in a highly seismic area, less than 10 km from the Psathopyrgos fault and the Rion-Patras transfer fault.

The Izmit Bay Suspension bridge has a 1,550 m long main span and two 566 m side spans. It is supported by towers placed in 40 m water depth, with a top elevation of 252 m above the sea level.

Both bridges were designed for severe environmental conditions, in deep water depth and for strong winds, and are positioned in highly seismic areas, with nearby active faults. Their base isolation system at the foundation level constitutes an emblematic example of the base isolation concept in modern long-span bridge designs.

The tower foundations (Fig. 3 shows the scheme of the Rio-Antirrio Bridge tower foundations) are placed on a gravel bed on soil reinforced with tubular steel pile inclusions not connected to the foundations. The gravel layer behaves as a plastic hinge (with inelastic deformation and energy dissipation taking place),

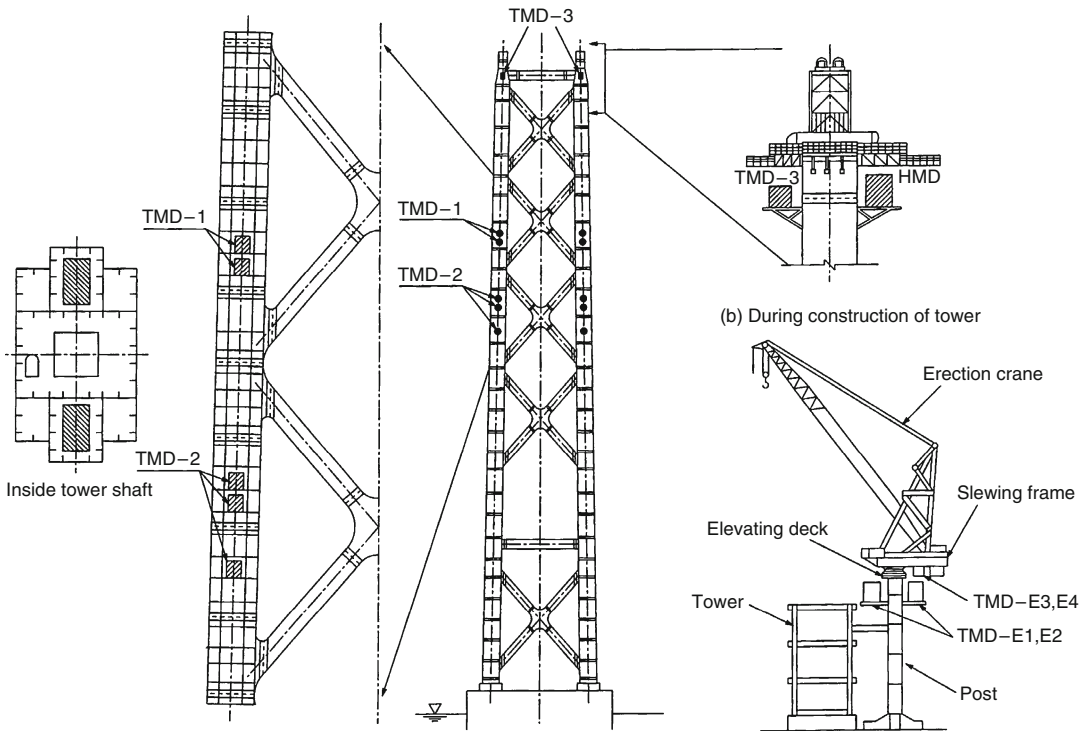


Design of Cable-Supported Bridges: Control Strategies, Fig. 3 Overview of the Rio-Antirrio Bridge foundation (From Jacques 2011)

while the ground reinforcement provides an overstrength which prevents the development of rotational failure modes of the foundation. In short, this solution allows for rocking, gapping, and sliding of the foundation. Should the seismic force exceed the design one, the failure mode would be pure sliding and not rotational, easily accommodated by the bridge.

Pylon Control

Pylons are connected to the bridge foundations and thus are affected by ground motion. However, due to their high vibration modes, they are prone mostly to wind-induced vibrations. Consolidated practices implemented are based on passive control, using different systems.



Design of Cable-Supported Bridges: Control Strategies, Fig. 4 TMDs on the Akashi Kaikyo towers (Courtesy of “Honshu-Shikoku Bridge Expressway Co., Ltd.”)

TLDs (Tuned Liquid Damper systems) dissipate energy using a secondary mass in the form of a body of liquid introduced into the structural system and tuned in such a way to act as a dynamic vibration absorber. TLDs are particularly effective for the suppression of lateral vibrations (e.g., due to wind or pedestrian loads) and were used for vibration control of the Toda Park Bridge in Japan. Similar devices to TLDs are TSDs (Tuned Sloshing Dampers) and TLCDs (Tuned Liquid Column Dampers).

Another important issue is the control of the pylons during erection and also during the cable spinning, since their frequency varies significantly in that phase. The use of passive dampers (mostly Tuned Liquid Dampers) proves to be ineffective since they have to be constantly tuned to the new pylon configuration. In the past 20 years, this problem has been dealt with the implementation of active and hybrid control strategies. This is a particularly consolidated practice in Japan, with several cases reported

(Fujino and Siringoringo 2013). The first application of active mass dampers (AMD) to a bridge pylon was on the 1991 Rainbow suspension bridge in Tokyo, Japan. The AMDs were placed on top of the outer edge of the 126 m high pylons to help reduce large-amplitude vortex-induced vibration. During erection, the Akashi Kaikyo Bridge pylons were equipped with two sets of Hybrid Mass Dampers (HMD), a combination of a combination of passive TMD and active control actuator. Figure 4 (Courtesy of “Honshu-Shikoku Bridge Expressway Co., Ltd.”) outlines the position of the vibration control devices on the bridge tower.

Deck Control

As in the case of the pylons, the deck of a long-span bridge is exposed to turbulent wind and thus to different wind-induced phenomena. Depending on the bridge configuration, the bridge deck is subjected from minor to average solicitation from earthquakes. The former is

particularly truth for bridges with a deck totally suspended (without deck-pylon connections). Vibrations and displacements due to anthropic loads (railway, highway, and pedestrian) have to be controlled as well.

An important issue is the pounding effect that usually takes place during strong earthquakes between the main span and the approaching span of long suspended bridges or the main span and the pylon of self-anchored suspended bridges and may cause considerable damage of the colliding bridge deck segments, in correspondence to the position of the expansion joints. This is important considering that the vibration periods of the flexible and usually unrestricted main span are different from the vibration period of the pylon or the approaching spans.

For long spans the usual design practice nowadays for seismic isolation is to have a fully or semi-floating deck (where the deck is suspended from stay cables or hanger ropes and there is little or no contact at the pylons through lateral control devices and bearings).

Control devices are placed usually between the deck and the pylons and the main span and the approaching span, and consolidated practice indicates the implementation of passive devices. Again, design decisions drive the adoption of control devices. In fact, a compromise has to be made for what regards the longitudinal movement and the possible deck-pylon connection, considering both impact forces (earthquake, wind, vehicle braking) and slow developed stresses due to the temperature gradient. The compromise lies in choosing between a fully floating system, prone to an elevated movement, and a fixed deck-pylon that results to a more rigid system. For very long spans either case is impossible, and in many cases, an intermediate solution is adopted, with the deck-pylon being connected with damping systems that provide limitations to the free movement of the deck.

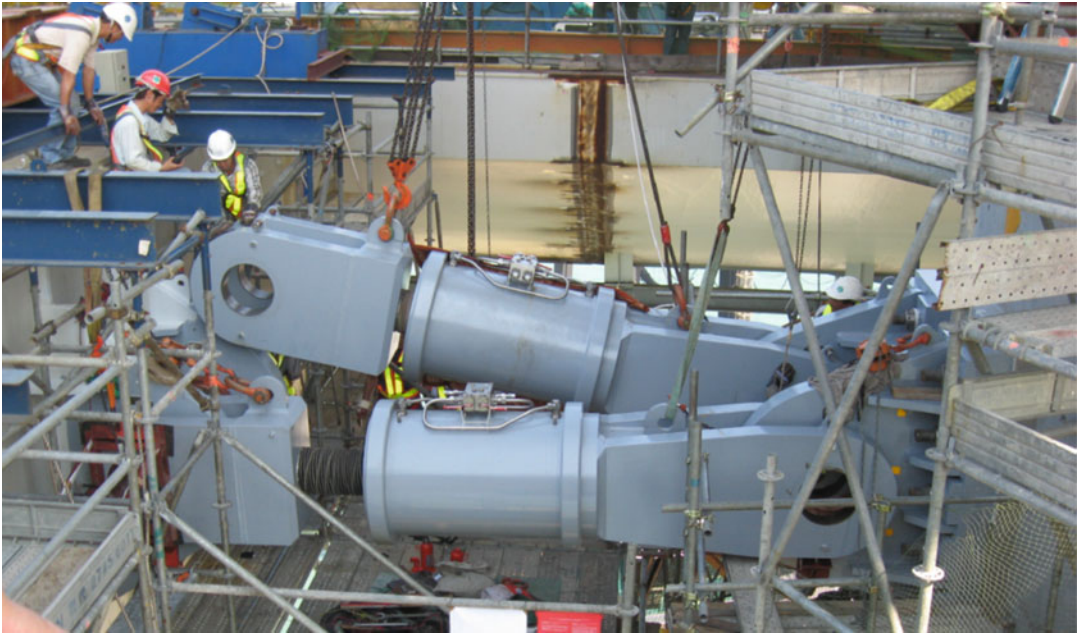
Shock Transmission Units and (usually fluid) dampers are implemented to limit longitudinal deck movements, of the otherwise undamped deck, and, in the case of suspension bridges, to avoid pounding between the stiffening trusses and the towers.

Shock Transmission Units (STUs), otherwise known as Lock-Up Devices (LUDs), are devices that allow the slow movement of the deck when the bridge structure slowly expands and contracts (under thermal loads or creep) but lock-up (under seismic or wind). In this way, the structural system responds differently to slowly applied loads (daily temperature variation and creep) and fast loads (impact, earthquake, wind gusts). Thus, large seismic or extreme event forces are restricted from the bridge towers.

STUs have been used broadly in major cable-supported bridges. Patel (2013) reports several applications both in suspension bridges (the Storebaelt in Denmark) and cable-stayed bridges (the Sidney Lanier in the USA, the Second Severn Bridge in the UK, the Seohae Grand Bridge in South Korea, and the Stonecutters Bridge in Hong Kong). In many cases, the concern was not only seismic but wind loads as well (Fig. 5).

Fluid viscous dampers (FVDs) are passive control devices that dissipate energy by applying a resisting force over a finite displacement. The damper's output force acts in a direction opposite to that of the input motion. Such dampers consist usually of a hollow cylinder filled with a fluid (typically silicon based). As the damper piston rod and piston head are stroked, fluid is forced to flow through orifices either around or through the piston head, something that produces a force in the opposite direction. The friction between fluid particles and the piston head leads to energy dissipation in the form of heat. Fluid viscous dampers can be tuned to have either linear or nonlinear behavior, depending on the specific application. A strongly nonlinear behavior means that the opposing force is very low for low applied forces but can be high as the applied loads increase.

FVDs have been implemented in many long-span suspended bridges in the last years, especially in China. An example is the Xihoumen Bridge with a main span of 1,650 m (currently the second longest suspension bridge in the world) that relies on four stroke dampers that have a range of movement of $\pm 1,100$ mm installed at the end of the bridge deck.



Design of Cable-Supported Bridges: Control Strategies, Fig. 5 STU units under installation in the Stonecutters Bridge (From Patel 2013)

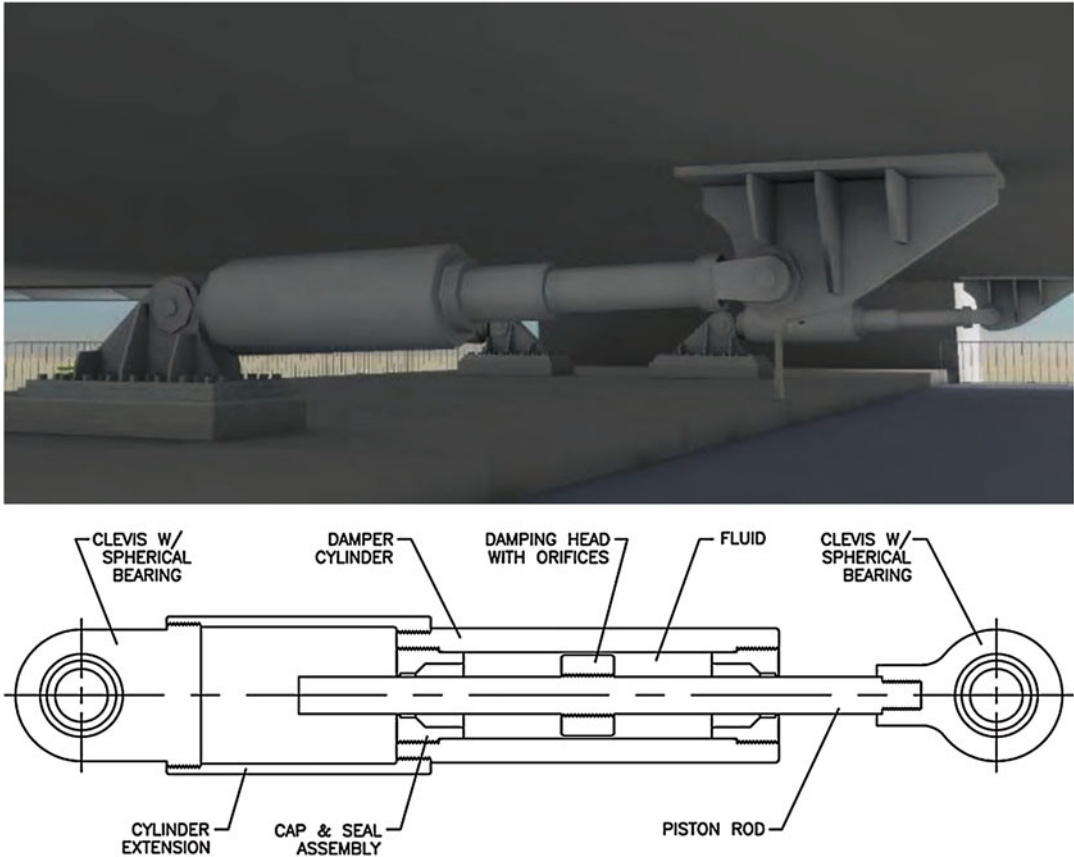
Other innovative ideas include the case of the Sutong Bridge, the world's second longest cable-stayed bridge with a span of 1,088 m and a totally suspended deck; eight isolators are used to reduce the axial motion of the suspended deck (Fig. 6). Each isolator consists in a combined spring-damper element (consisting in a fluid damper, an elastomer spring, and an internally placed gapping mechanism). For the initial 85 % of the available displacement from the neutral position, the isolator has only damping response (that accommodates also the daily thermal expansion and contraction of the suspended deck) and for the rest 15 %, combined spring and damping response. This, since the long period of the deck, provides intrinsic vibration control in the longitudinal direction, while additional displacement has to be limited in the longitudinal direction for other events (e.g., vehicle braking forces, typhoon loads). The control device has a maximum displacement of 85 cm of which about 10 cm is the final combined spring and damping.

Lateral movement is also an issue, especially for floating deck configuration. For example, in the Rio-Antirrio cable-stayed bridge, the

continuous deck is free to accommodate the movements from seismic excitation in the longitudinal direction, since it is fully suspended. The movements in the transverse direction are controlled by nonlinear fluid viscous dampers and fuse restraints installed parallel to the dampers (Fig. 7). In this way, the deck is linked rigidly to the substructure when subjected to lateral loads not exceeding the design capacity. In major seismic events, the fuse restrainers are designed to fail and leave the dampers free to dissipate the earthquake-induced energy acting upon the structure. After their failure the deck is free to swing coupled to the dampers.

Other passive control devices (solid viscoelastic dampers, metallic dampers, friction dampers, etc.) have a different hysteric behavior with several advantages and disadvantages also for what concerns their reliability.

In the case of the Tianxingzhou Yangtze River Bridge in China (a bridge with a main span of 504 m, the longest combined road and rail cable-stayed span in the world), a combination of magnetorheological (MR) and hydraulic dampers is used in correspondence to the piers, to limit



Design of Cable-Supported Bridges: Control Strategies, Fig. 6 Sutong Bridge pylon-deck viscous damper system (above) and device schematic (below) (From You et al. 2008)

deck vibrations. MR dampers consist of a hydraulic cylinder containing a solution that, as their name suggests, can reversibly change, when exposed to a magnetic field or electric current, from a free-flowing liquid to a semisolid with controllable yield strength. These devices allow to actively control impact forces (due to train braking and earthquakes), where the hydraulic damper provides a velocity-related recovery (damping) force that relieves slow movements (e.g., due to daily temperature change).

Cable Control

Stay cables and hanger ropes are characterized by very low internal damping, relatively small mass, large flexibility, and considerable length. They are susceptible to vibrations due to direct loading (caused by wind, rain, and other extreme events

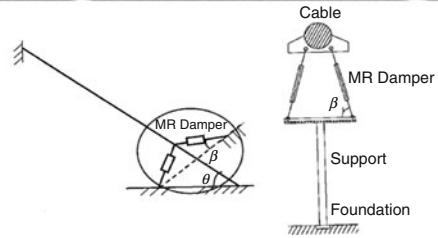
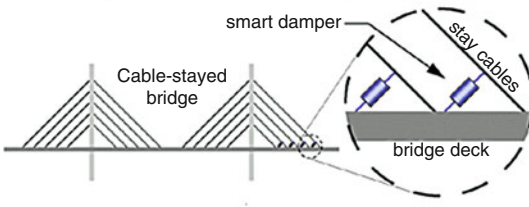
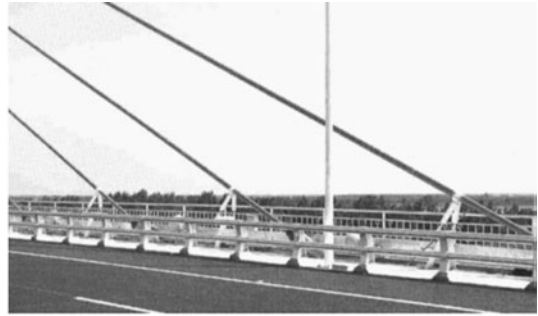
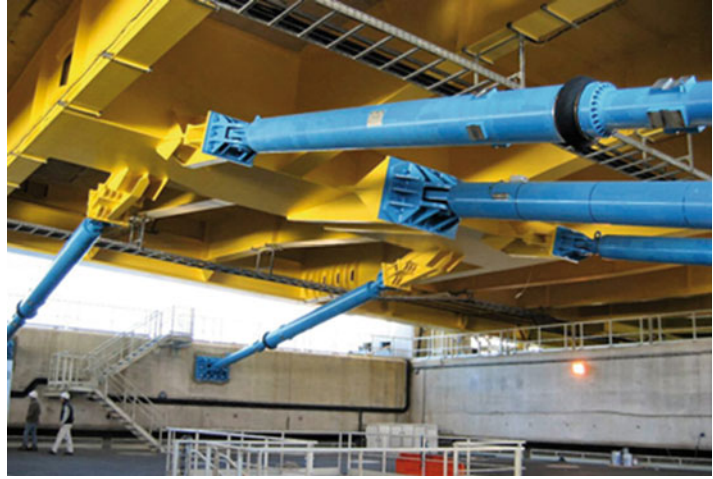
such as snowstorms) or, indirectly, due to the vibrations of the supported deck. To suppress these vibrations, commonly passive devices are used (hydraulic, oil, rubber, or viscous dampers). Usually these devices are attached to the cables near the deck anchorage, although other configurations are possible.

More recently, adaptive dampers have been implemented on many occasions. Their principal advantages compared to passive dampers lie in the easy tuning (by adjusting the current or voltage) both before and after the installation and for different modes of vibration.

Duan et al. (2006) report the first successful application of MR dampers for the retrofitting of the Dongting Lake Bridge in China. A total of 312 dampers were installed on 156 cables of the bridge (two for each cable), at a height of 1.8 m

Design of Cable-Supported Bridges: Control Strategies,

Fig. 7 Vibration control on the deck of the Rio-Antirrio Bridge (courtesy of Gefyra SA)



Design of Cable-Supported Bridges: Control Strategies, Fig. 8 Cable-stayed vibration control with MR dampers at the Dongting Lake Bridge (left, from Duan

et al. 2006) and the Shandong Binzhou Yellow River Bridge (right, from Li et al. 2007)

from the bridge deck, and in an inclined T-shape support configuration (Fig. 7). A similar approach has been used in the retrofitting of the Shandong Binzhou Yellow River Highway Bridge (a three-tower cable-stayed bridge in China, with two main spans of 300 m). The bridge suffered unaccepted vibrations after its opening in 2004 and was retrofited with MR dampers on 20 of its longest stay cables (Li et al. 2007; Fig. 8).

Other recent applications of MR dampers for cable control include the Jiashao Bridge in China, the world's longest and widest multi-span cable-stayed bridge, and the Sutong Bridge.

Overview of Control Methods

The above are reassumed in Table 1. Many issues can be minimized by adopting an appropriate structural shape. For example, deck

Design of Cable-Supported Bridges: Control Strategies, Table 1 Overview of principal control issues and methods for cable-suspended bridges

Substructure	Principal issue(s)	Control type	Control device
Foundations	Earthquake	Passive (seismic isolation)	Foundation design
Towers	Vortex-induced vibrations	Passive	TMD (Tuned Mass Dampers)
	Earthquake	Active	TLD (Tuned Liquid Dampers)
		Semi-active	TSD (Tuned Sloshing Dampers)
		Hybrid	TLCD (Tuned Liquid Column Dampers) HMD (Hybrid Mass Dampers)
Stay cables and hanger ropes	Wind	Passive	Viscous dampers
	Rain	Active	MR (magnetorheological) dampers
	Ice		
Deck	Vortex-induced vibrations	Passive	FVD (fluid viscous dampers)
	Earthquake	Semi-active	STU (Shock Transmission Units)
	Pedestrian		TMD
	Vehicle		TLD
	Train		

vortex-induced vibration can be suppressed by changing the section or by using aerodynamic appendages or flaps.

Seismic Response of Cable-Suspended Bridges

In the last 30 years, many new bridges were constructed, and many old bridges were retrofitted with instrumentation systems. Measurements from these systems during major seismic events are allowed to assess the seismic performance of the structures, the performance of the control systems, and the post-earthquake behavior of the bridge. A number of significant case studies based on SHM (Structural Health Monitoring) data or field investigations are reported below.

Akashi Kaikyo Bridge

The Akashi Kaikyo Bridge, presently the suspension bridge with the longest span in the world, was hit by the Hanshin Kobe earthquake while under construction. The epicenter of the 6.8 magnitude earthquake was located just about 2–3 km

east of the bridge. The construction of the foundations, both towers, and main cables had been completed before the earthquake occurred, but neither hangers nor girders had been attached to the main cables, while the stiffening girders were not yet in place. The bridge towers were equipped with temporary semi-active mass dampers, tuned for wind excitation.

There was no damage to bridge structures, but displacements of the four foundations were caused by crustal movements. There was also an 80 cm elongation at the center span and 30 cm at one of the side spans. There was also a change of angle horizontally and vertically at the towers and at the anchorages. The construction reprised accounting the above facts in the updated design (Kitagawa 2004).

Higashi-Kobe Bridge

The Higashi-Kobe Bridge is a cable-stayed bridge with a center span of 485 m. With the aim of augmenting the fundamental period of bridge, the bridge deck is supported by towers and piers in such a way that the girder is movable in the longitudinal direction independently from the towers; thus, the displacements are mainly

restricted by the cables. The bridge is equipped with two vane-type oil dampers at the end piers, and the cables are arranged in a harp pattern, in order to prevent an excessive longitudinal displacement. The vane-type dampers help in preventing excessive displacement since its effect increases with the velocity of the girder movement. During the 1995 Hanshin Kobe earthquake, the bridge suffered significant yet repairable damage (Naganuma et al. 2000). The upper set of bolts of the transverse wind brace on the end pier broke, and the upper shoe was taken off, something that put excessive force on the vane-type dampers that were not designed to carry transversal loads. It should be observed that the design acceleration was significantly lower than the one monitored during the event.

Yokohama Bay Bridge

The Yokohama Bay Bridge with a total length of 860 m and a central span of 460 m is a densely instrumented cable-stayed double deck highway bridge in Japan. The bridge opened in 1989 and was already retrofitted in 2005 for additional seismic safety. Specifically, to accommodate longitudinal girder displacement in case of large earthquake, LBCs (link-bearing connections) were implemented: 10 m long end-links at the two end piers and 2 m long tower links at the two towers. The bridge is located 180 km from the closest fault area of the 2011 9.0 magnitude Great East Japan (Tohoku) Earthquake, notably the largest earthquake in Japan's modern history. The bridge experienced large transverse displacements during the earthquake (55 and 62 cm in the top of the tower and the mid-span, respectively) and very small longitudinal displacements (15 cm). No damage was detected. The excessive transversal and longitudinal motion leads to scratch marks on the wind shoes and a few crushed nuts and bolts on one of the tower links (Siringoringo et al. 2013).

Rio-Antirrio Bridge

The Rio-Antirrio Bridge is a multi-pylon cable-stayed bridge designed to withstand seismic

events by means of different mechanisms between the pylon and the soil and the deck and the pylons and abutments (where a dissipation system with hydraulic dampers is installed). The seismic design was tested during the 2008 6.5 magnitude earthquake about 35 km from the bridge location, which had no permanent effects to the bridge. The bridge suffered very mild damage, principally in the lateral restrainers (sacrificial elements) that yielded (Papanikolas et al. 2010). As a remedial, the deck was realigned and six fuses were replaced.

Ji-Lu Bridge

The Ji-Lu Bridge is a single pylon cable-stayed bridge with prestressed concrete girders that suffered important damage in September 1999 during the 7.7 magnitude Chi-Chi, Taiwan earthquake (Kosa and Tasaki 2003). The 120 + 120 m bridge, supported by a single concrete pylon, was under construction at the time. The damage was widespread, extending over the main girder, the main pylon, and the two piers, due to the seismic force that acted in both the axial and the perpendicular directions of the bridge. Pounding occurred between the main span and the approaching spans and between the superstructure and the towers. Also, one cable was put out from the anchorage section of the pylon, and many sockets were also pulled out from the anchorage section.

Shipshaw Bridge

The Shipshaw Bridge is an asymmetrical single pylon cable-stayed bridge in Quebec, Canada, with a deck 183 m long supported by a double-leg steel tower. At a later inspection after the 1988 Saguenay 6.0 magnitude earthquake (Filiatrault et al. 1993), it was found that one of the four interior anchorage plates in one abutment failed. Although the damage cannot be linked directly to the earthquake, later analysis point towards this, together with the fact that the anchorage plates were already subjected to high stress concentrations. The bridge was immediately closed, and all four plates in correspondence to that abutment were replaced (Table 2).

Design of Cable-Supported Bridges: Control Strategies, Table 2 Notable earthquake events and bridge behavior

Bridge	Type	Length (m) (total/longest span)	Seismic event	Magnitude (Mw)	Damage	References
Akashi Kaikyo Bridge ^a	Suspension	3,911/1,991	1995 Hanshin Kobe earthquake	6.8	No damage	Kitagawa 2004
Higashi-Kobe Bridge	Cable stayed	885/485	1995 Hanshin Kobe earthquake	6.8	Vane-type dampers	Naganuma et al. 2000
Yokohama Bay Bridge	Cable stayed	860/460	2011 Great East Japan Earthquake	9.0	No damage	Siringorino et al. 2013
Rio-Antirrio Bridge	Cable stayed (multi-pylon)	2,252/560	2008 Achaia-Ilia earthquake	6.5	Fuse replacement	Papanikolas et al. 2010
Ji-Lu Bridge	Cable stayed (single pylon)	240/120	1999 Chi-Chi, Taiwan earthquake	7.7	Pounding, cable detachment,	Kosa and Tasaki 2003
Shipshaw Bridge	Cable stayed (single pylon)	183/13,725	1988 Saguenay earthquake	6.0	Anchorage plate failure	Filiatrault et al. 1993

^aDuring construction

Future Trends

Future trends of cable-suspended bridge vibration control will run alongside other developments for smart bridges and SHM. SHM was in fact pioneered in long-span cable-suspended bridges in Japan, China, and Korea parallel to active or semi-active control methods. A bigger integration of control methods with SHM systems and instrumentation is expected parallel to current developments of bridge SHM in major areas around the world (Korea, Koh et al. 2009; Hong Kong, Ni and Wong 2012; Japan, Fujino and Siringorino 2008; and China, Sun et al. 2009).

Another aspect is the integration of Energy Harvesting (EH) methods with vibration suppression and control devices, for supplying power to autonomous sensors for SHM applications or for self-powered damper systems. Applications are expected using electromagnetic dampers (Shen and Zhu 2014) and piezoelectric or other materials once issues related to fatigue and real-life stresses are resolved.

The implementation of active, semi-active, and hybrid control strategies is expected to increase, even though passive vibration

suppression methods could benefit from innovative research (e.g., using passive nonlinear energy sinks – Wierschem et al. 2011).

Summary

Vibration control is a major topic for cable-supported bridges. Even though this structural configuration is less sensitive to earthquake-induced vibrations, issues related mostly to pounding have to be accounted for in the design and retrofitting stage. In the last years there are consolidated cases of active or semi-active control in major bridges around the world, with significant performance gains compared to classic passive methods. This trend is expected to increase with the increase of reliability of such methods.

References

- Bontempi F (2006) Basis of design and expected performances for the Messina Strait Bridge. In: International conference on bridge engineering – challenges in the 21st century, Hong Kong, 1–3 Nov.

- Burdette NJ, Elnashai AS, Lupoi A, Sextos AG (2008) The effect of asynchronous earthquake motion on complex bridges. Part 1: methodology and input motion. *J Br Eng ASCE* 13(2):158–165
- Duan YF, Ni YQ, Ko JM (2006) Cable vibration control using magnetorheological dampers. *J Intell Mater Syst Struct* 17(4):321–325
- Filiatrault A, Tinawi R, Massicotte B (1993) Damage to cable-stayed bridge during 1988 Saguenay earthquake. I. Pseudostatic analysis. *J struct Eng* 119(5):1432–1449
- Fujino Y, Siringoringo D (2008) Structural health monitoring of bridges in Japan: an overview of the current trend. In: Fourth international conference on FRP composites in civil engineering (CICE2008), Zurich, 22–24 July.
- Fujino Y, Siringoringo D (2013) Vibration mechanisms and controls of long-span bridges: a review. *Struct Eng Int* 3:248–268
- Gimsing NJ, Georgakis CT (2012) Introduction, in *Cable Supported Bridges: Concept and Design*, Third Edition, John Wiley & Sons, Ltd, Chichester, UK. doi: 10.1002/9781119978237.ch
- Housner G, Bergman L, Caughey T, Chassiakos A, Claus R, Masri S, Skelton R, Soong T, Spencer B, Yao J (1997) Structural control: past, present, and future. *J Eng Mech* 123(9):897–971
- Jacques C (2011) The Rion-Antirion bridge – when a dream becomes reality. *Front Archit Civil Eng China* 5(4):415–426
- Kitagawa M (2004) Technology of the Akashi Kaikyo bridge. *Struct Control Health Monit* 11:75–90
- Koh H-M, Lee H-S, Kim S, Choo JF (2009) Monitoring of bridges in Korea. In: Christian B, Fou-Kuo C, Yozo F (eds), *Encyclopedia of structural health monitoring*. Wiley. <http://onlinelibrary.wiley.com/doi/10.1002/9780470061626.shm163/abstract>
- Kosa K, Tasaki K (2003) Detailed investigation of PC Cable-stayed Bridge damaged in the 1999 Taiwan earthquake. In: *Proceedings of the 19th US – Japan Bridge Engineering Workshop*, Tsukuba Science City, Japan, 27–29 Oct.
- Li H, Liu M, Li J, Guan X, Ou J (2007) Vibration control of stay cables of the Shandong Binzhou Yellow River Highway Bridge using magnetorheological fluid dampers. *J Br Eng ASCE* 12(4):401–409
- Naganuma T, Kitazawa M, Adachi Y, Noguchi J (2000) Seismic design and behavior of the Higashi-Kobe Bridge and restoration after the 1995 Kobe Earthquake. 12 WCEE, paper no 2793, Auckland
- Ni YQ, Wong KY (2012) Integrating bridge structural health monitoring and condition-based maintenance management. Lecture 06: workshop on civil structural health monitoring (CSHM-4) “SHM systems supporting extension of the structures’ service life”, Berlin, 6–8 Nov
- Papanikolas P, Stathopoulos-Vlamiis A, Panagis A, Pecker A, Infanti S (2010) The behavior of Rion-Antirion Bridge during the Earthquake of “ACHAIA-ILIA” on June 8, 2008. In: Third international fib congress incorporating the PCI annual convention and bridge conference, Washington, DC, 29 May–2 June
- Patel DJ (2013) Shock transmission units in construction. ICE, London
- Shen WA, Zhu S (2014) Harvesting energy via electromagnetic damper: application to bridge stay cables. *J Intell Mater Syst Struct*. doi:10.1177/1045389X13519003, Published online before print 21 Jan 2014
- Siringoringo D, Fujino Y, Namikawa K (2013) Seismic response analyses of the Yokohama Bay cable-stayed bridge in the 2011 Great East Japan Earthquake. *J Br Eng ASCE*. doi:10.1061/(ASCE)BE.1943-5592.0000508, A4014006
- Spencer B, Nagarajaiah S (2003) State of the art of structural control. *J Struct Eng ASCE* 129(7):845–856
- Spencer BF, Soong TT (1999) New application and development of active, semi-active and Hybrid control techniques for seismic and non seismic vibration in the USA. In: International post-SmiRT conference seminar on seismic isolation, passive energy dissipation and active control of vibration of structures, Cheju, 23–25 Aug
- Sun L, Sun Z, Dan D, Zhang Q, Huang H (2009) Researches and implementations of structural health monitoring systems for long span bridges in China. *Doboku Gakkai Ronbunshuu A* 65(1):15–29
- Wierschem NE, Spencer B, Bergman LA, Vakakis AF (2011) Numerical study of nonlinear energy sinks for seismic response reduction. In: ANCRISST 2011 – The sixth international workshop on advanced smart structures and technology, Dalian
- Yao JTP (1972) Concept of structural control. *J Struct Div* 98(7):1567–1574
- You Q, He P, Dong X, Zhang X, Wu S (2008) Sutong bridge – the longest cable-stayed bridge in the world. *Struct Eng Int* 18(4):390–395

Design of Mechanically Stabilized Walls (MSWs)

Richard J. Bathurst
GeoEngineering Centre at Queen’s-RMC Civil Engineering Department, Royal Military College of Canada, Kingston, ON, Canada

Synonyms

Dynamic finite difference method; Dynamic finite element method; Earthquakes; Geosynthetics; Metallic reinforcement; Pseudostatic; Reinforced soil; Retaining walls; Seismic design; Sliding block

Introduction

Mechanically stabilized walls (MSWs) are a class of earth retaining walls that use horizontal layers of metallic or geosynthetic reinforcement placed within the wall backfill to create a composite mass of soil, reinforcement layers, and facing. This mass (or block) acts as a gravity structure to resist the destabilizing earth forces that act at the back of the reinforced soil zone. Details of some MSW types are illustrated in Fig. 1. The function of the gravity structure (reinforced zone including the facing) can be imagined equivalent to conventional gravity retaining wall structures including reinforced concrete T-sections. The advantage of MSWs over conventional gravity wall structures, based on North American experience, is that they can be constructed for as little as 50 % of the cost of traditional solutions.

History

The modern history of MSW structures can be traced back to the 1960s with the advent of steel strip walls first constructed in France. In the 1970s, modern polymeric sheet materials (geotextiles) were first employed as the soil reinforcement material in wall structures located in Europe and North America. Today, MSWs are ubiquitous for the earth retaining wall function. Their use and the methodologies to design these structures are well established and accepted. Conventional static design methods for structures in non-seismic areas have been extended to accommodate the additional loads that develop as a result of earthquake.

Reinforcement

The soil reinforcement materials that are used in MSWs can be broadly classified into metallic and geosynthetic categories. Metallic reinforcement includes steel strips, steel bar mats and ladders, welded wire, and steel anchor plates attached to the facing using steel rods. Almost all steel reinforcement products are galvanized for corrosion

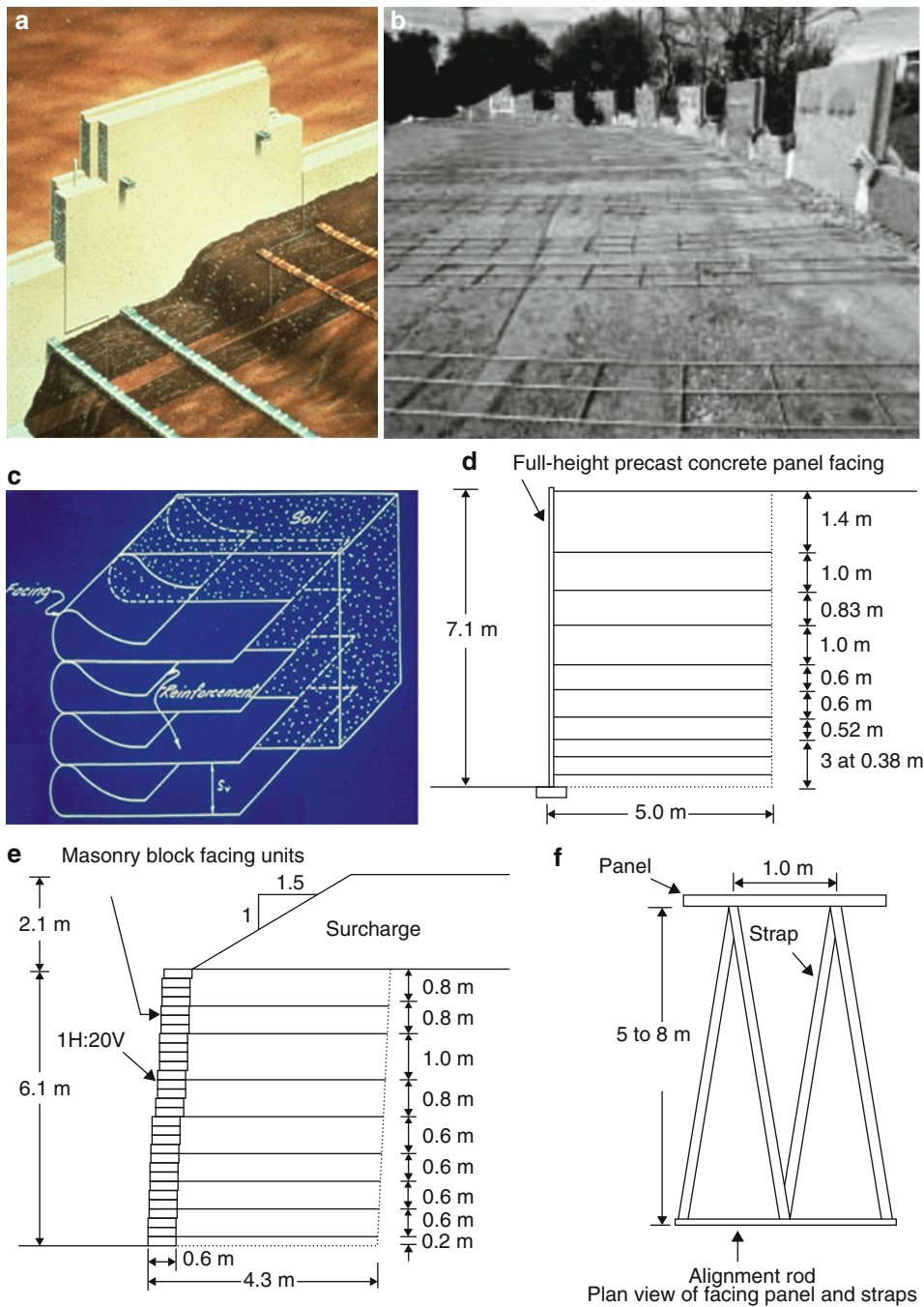
protection. Geosynthetic reinforcement materials are high-performance polymeric materials whose primary constituent is typically polypropylene (PP), high-density polyethylene (HDPE), or polyester (PET). Geosynthetic reinforcement products are classified as geotextiles, geogrids, and strips. Woven geotextiles are sheetlike materials that have the appearance of a textile. Some geogrids are manufactured from punched and drawn polymeric sheets to create an open integral grid-like structure with square, rectangular, or triangular openings. Other geogrid products are manufactured from bundled polyester filaments that are knitted or woven together in longitudinal and transverse directions to form the geogrid structure. Geogrid strips are typically belts of aligned polyester filaments. The polyester core is protected by a polymeric sheath. Most often, geogrids are used in geosynthetic MSW structures because the geogrid apertures can better engage the backfill soil to transfer load from the soil to the reinforcement, particularly if the soil is a cohesionless sand or gravel.

Backfill Soil

Cohesionless soils such as sands and gravel are the preferred material for MSWs compared to fine-grained soils because they are easier to place and compact, are better draining, and have higher shear strength and stiffness. In fact, most government design guidelines in North America restrict the soil in the reinforced soil zone to these “select” materials. Limitations on soil type also apply for steel-reinforced MSWs in order to minimize corrosion of the steel.

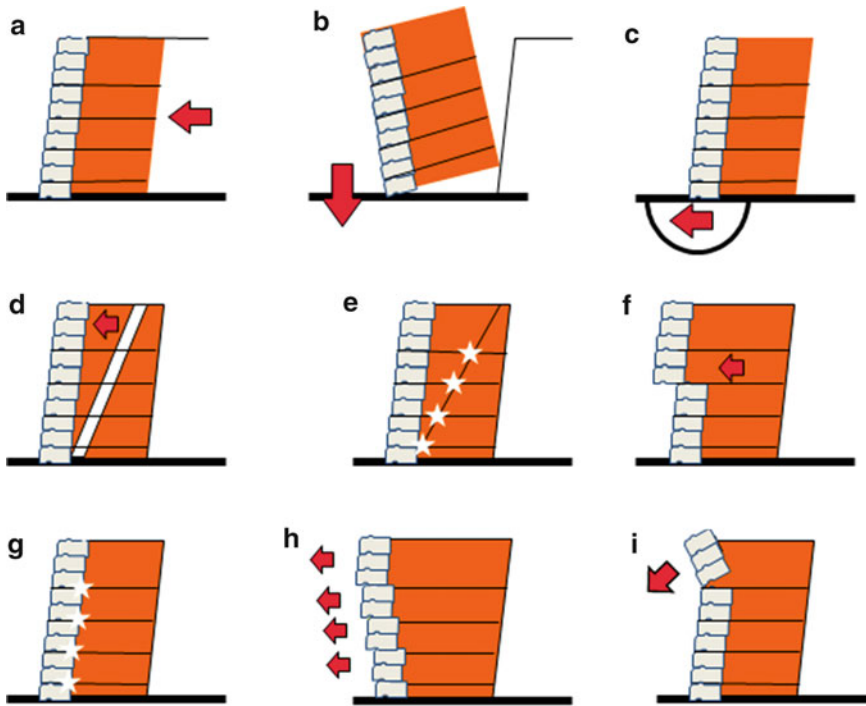
Facing

MSWs must include a facing that acts as a construction aid to assist with placement and compaction of the soil at the wall face and ensures that the wall face will be stable at a vertical or near-vertical orientation over its design life. The simplest facing can be constructed by simply extending the primary reinforcement layers in a



Design of Mechanically Stabilized Walls (MSWs),
Fig. 1 Example MSW structures and components (a) Steel strip reinforced soil wall with incremental concrete panel. (b) Steel ladder reinforcement with incremental

concrete panel. (c) Geotextile wrapped face wall. (d) Full-height concrete panel wall. (e) Modular block wall. (f) Polyester strap reinforcement



Design of Mechanically Stabilized Walls (MSWs), Fig. 2 Failure modes for MSWs with modular block facing (a) Base sliding. (b) Overturning. (c) Bearing

capacity. (d) Pullout. (e) Tensile over-stress. (f) Internal sliding. (g) Connection failure. (h) Column shear failure. (i) Toppling

wrapped configuration at the wall face and tucking the free end of the wrap back into the reinforced soil mass. A temporary moving formwork is required to support each wrap during construction. These facings are flexible and cost-effective and best used for temporary structures when long-term durability of the facing is not a concern. More robust facings can be formed using welded wire cages, masonry or wet-cast concrete blocks, incremental concrete panels, or full-height concrete panels. Historically, steel soil reinforcement materials have been used in conjunction with incremental concrete panels that come in square or hexagonal shapes.

Design Concepts

Static Load Environments

MSWs are first designed to have adequate margins of safety against collapse and excessive deformation under static loading conditions.

Potential failure mechanisms are illustrated in Fig. 2 for the case of geosynthetic and metallic reinforced soil walls constructed with a modular block facing. External failure modes are related to the stability of the composite reinforced soil zone comprised of the backfill soil, reinforcement layers, and facing. The composite acts as a gravity block to resist sliding at the base of the block and overturning at the toe in response to active earth forces from the retained soil located immediately behind the block (reinforced soil zone). In addition, the block must not generate a bearing capacity failure in the foundation soil or excessive settlement. For walls with simple geometry, sliding resistance typically controls the length of the block in the direction perpendicular to the running length of the wall face and consequently the length of the reinforcement. Regardless of the computed minimum reinforcement length, design codes typically restrict the minimum length of the reinforcement to be not less than 60–70 % of the height of the wall.

Additional stability calculations for global stability must always be carried out, as is the case for any retaining wall structure, to ensure that the structure is not part of a larger failure mechanism that extends beyond the reinforced soil zone and into the foundation. These calculations can be carried out using conventional slope stability programs.

In order for the reinforced zone to maintain its coherence, it is necessary to design the reinforcement layers against internal stability modes of failure. These include tensile overstressing of the reinforcement leading to excessive deformation or even rupture and pullout of the reinforcement layers from the anchorage zone which extends from the location of the potential internal failure wedge to the free end of the reinforcement layers. In some cases it may be possible for a failure mechanism to propagate at an angle from the heel of the wall facing or at a higher elevation at the back of the facing column through the reinforced soil zone and into the retained soil zone. These failure mechanisms are called composite failures and are best investigated using circular or noncircular slope analysis methods implemented in slope stability computer programs that have been modified to include the tensile capacity of any intersected reinforcement layer. For reinforced soil walls with incremental panel or full-height panels, the internal sliding failure through the concrete facing is unlikely (Fig. 2f).

The design tensile strength of the reinforcement will vary depending on the reinforcement type. For steel reinforcement the unfactored tensile strength of the reinforcement is easily computed as the product of yield strength of the tensile member and the cross-sectional area. Most steel reinforcement products are galvanized for cathodic protection, and a sacrificial thickness of steel is included in the calculation of tensile load capacity to account for possible corrosion. This thickness is computed based on the soil type, porewater chemistry and resistivity, and design life. For geosynthetic MSWs, the reference tensile strength of the reinforcement from standard laboratory testing must be reduced to account for possible installation

damage, chemical degradation, and creep. Modern geosynthetic reinforcement products have been specially formulated to minimize degradation. Polyester geogrids and strips are manufactured with a coating for additional stability and for protection during installation and exposure to UV light.

The facing must also be stable over the life of the structure, and the connections between the facing and the reinforcement layers must have sufficient tensile capacity. Adequate connection capacity is most important for walls with a hard facing (e.g., concrete blocks and panels). For modular block walls, stability against toppling of the units at the top of the facing column and interface shear must also be ensured. The failure modes for column shear and toppling in Fig. 2h, i are unique to modular block walls and thus do not apply to incremental concrete panel, full-height concrete panel, and wrapped face walls.

Post-construction wall deformations for well-constructed MSWs should be less than 1.5 % of the wall height. Deformations for the same walls but using much stiffer steel reinforcement will be less. This difference may be a factor when deciding which category of reinforcement to use in a particular project.

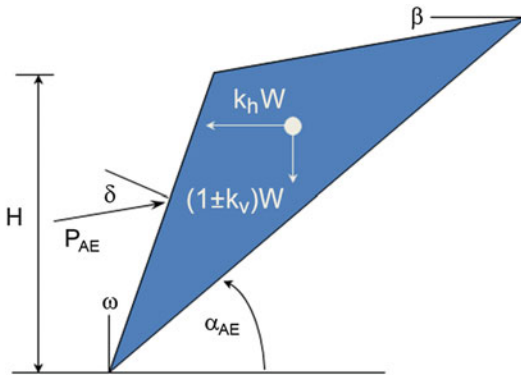
Seismic Design

The potential modes of failure identified for static design of MSWs are equally applicable to the same structures subjected to additional loads due to earthquake.

Current methods for the earthquake design of MSWs fall into three categories: (1) pseudo-static methods, (2) sliding block displacement methods, and (3) dynamic finite element or finite difference methods.

Pseudo-Static Methods

The calculations for earthquake loads used in internal and external stability design of MSWs have historically been based on extensions of the classical (static) Coulomb wedge approach for active earth forces acting against retaining walls. The advantage of the method is that in the limit of no earthquake-induced ground accelerations, the equations for active earth force



Design of Mechanically Stabilized Walls (MSWs), Fig. 3 Active wedge geometry and interpretation of total active earth force P_{AE} due to wedge self-weight W and peak ground acceleration coefficients k_h and k_v

(or active earth pressure) are the same as for the static case. The disadvantages of the approach are that it is limited to simple wall geometry and ground conditions, and analytical solutions are limited to peak ground accelerations often not more than 0.3 g. Where analytical solutions are possible, the computed total earth forces are often very high resulting in excessively conservative design with respect to number of reinforcement layers, strength of the reinforcement, and length. Nevertheless, the general Coulomb approach is familiar to geotechnical engineers, and this makes it an attractive method.

The total active earth force (P_{AE}) acting against the back of the reinforced soil zone for external stability wall calculations and the back of the wall facing for internal stability calculations can be equated to the wedge with weight (W) shown in Fig. 3. For the simple geometry shown in this figure, the total active earth force (P_{AE}) acting against an inclined surface can be computed analytically as:

$$P_{AE} = \frac{1}{2}(1 \pm k_v)K_{AE}\gamma H^2 \quad (1)$$

Here, γ = unit weight of the soil and H = height of the wall. The total earth pressure coefficient (K_{AE}) is calculated using the Mononobe-Okabe equation which is an extension of the Coulomb equation:

$$K_{AE} = \frac{\cos^2(\phi + \omega - \theta) / \cos \theta \cos^2 \omega \cos(\delta - \omega + \theta)}{\left[1 + \sqrt{\frac{\sin(\phi + \delta) \sin(\phi - \beta - \theta)}{\cos(\delta - \omega + \theta) \cos(\omega + \beta)}} \right]^2} \quad (2)$$

Here, ϕ = peak soil friction angle, ω = wall/slope face inclination taken as positive in a clockwise direction from the vertical, δ = interface friction angle at the back of the wall (or back of the reinforced soil zone), β = backslope angle taken from horizontal, and θ = seismic inertia angle given by:

$$\theta = \tan^{-1} \left(\frac{k_h}{1 \pm k_v} \right) \quad (3)$$

Quantities k_h and k_v are peak horizontal and vertical seismic coefficients, respectively, expressed as fractions of the gravitational constant, g . The peak ground acceleration components are selected based on a site- or region-specific design accelerogram at bedrock, the distance from the earthquake, and the ground conditions above bedrock. The values of k_h and k_v are assumed to be constant at the back and through the entire reinforced soil mass.

There are different strategies to separate the static component of load on the back of the reinforced soil zone and the static load component on the reinforcement layers from the corresponding additional dynamic load components due to earthquake. For example, the additional horizontal component of dynamic load (ΔP_{dyn}) against the wall face or back of the reinforced soil zone can be expressed as the difference between the total earth force (P_{AE}) and static earth force component (P_A) as follows:

$$\Delta P_{dyn} = P_{AE} - P_A \quad (4)$$

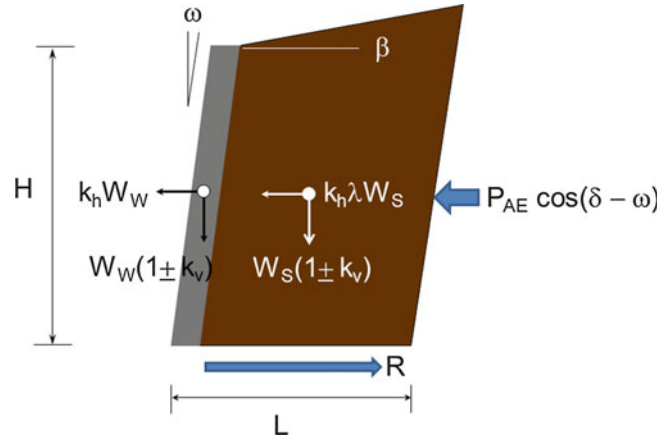
Equivalently,

$$\Delta K_{dyn} = (1 \pm k_v) K_{AE} - K_A \quad (5)$$

Here, K_A = static active earth pressure coefficient and ΔK_{dyn} = incremental dynamic active earth pressure coefficient. In stability calculations the

Design of Mechanically Stabilized Walls (MSWs),

Fig. 4 External stability geometry and forces



forces or coefficients of earth pressure shown above must be multiplied by $\cos(\delta - \omega)$ when factor of safety expressions are for loads and resistances acting in the horizontal direction. As distance from the earthquake epicenter increases, the most important ground acceleration is the horizontal component, and the vertical component is typically ignored (i.e., $k_v = 0$). Furthermore, the likelihood of peak horizontal and vertical acceleration components of an earthquake arriving simultaneously is small. In practice, the choice of k_h is typically based on local experience or prescribed by design codes that are applicable to the site area. In current US practice, the location of the dynamic earth force component is assumed to act at the same elevation as the static component (AASHTO 2012). If the consequences of failure are high, then the same code recommends locating the resultant at a higher elevation.

The calculation of factor of safety against external base sliding is carried out by substituting P_{AE} for P_A in a conventional block sliding analysis and adding horizontal inertial forces due to the mass of the reinforced zone (sliding block) as illustrated in Fig. 4. Quantities W_s and W_w are the weight of the reinforced soil and the facing column, respectively. Typically, the horizontal inertial force due to W_s is reduced by setting $\lambda = 0.5$. The contribution of the facing column to horizontal inertial loads is usually applied to facings constructed with modular blocks and is ignored for thinner incremental or full-height concrete

panel wall systems. Quantity R in Fig. 4 is the sliding resistance at the base of the reinforced soil zone.

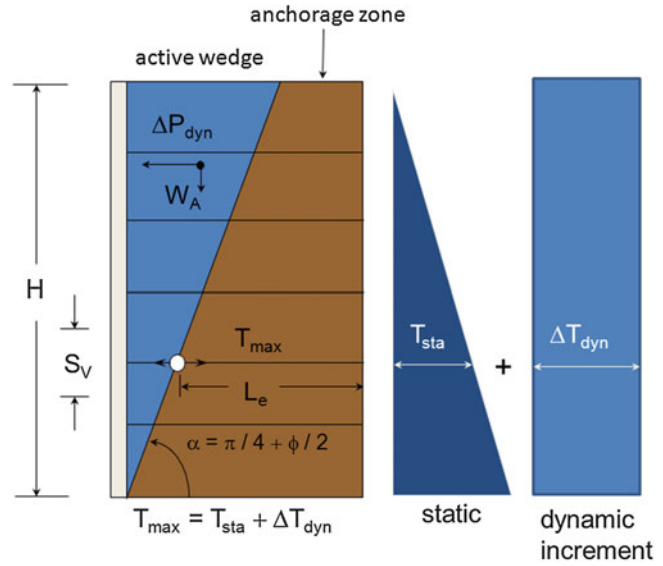
During an earthquake, additional dynamic transient tensile loads will be carried by the reinforcement. In the simplest approaches to internal stability seismic design, the dynamic load (T_{dyn}) due to earthquake is most often added to the static load (T_{sta}) calculated for the same structure as illustrated in Fig. 5. For example, the tensile reinforcement load T_{max} in a layer can be calculated as:

$$T_{max} = T_{sta} + T_{dyn} = K_r \sigma_v S_v + \Delta K_{dyn} \bar{\sigma}_v S_v \quad (6)$$

Tensile load is expressed with units of force per unit running length of wall. Here S_v is the reinforcement spacing and σ_v is the vertical stress acting at the elevation of the reinforcement due to soil self-weight and any uniformly distributed surcharge. The quantity $\bar{\sigma}_v$ is the average vertical stress over the height of the wall. For modular block wall systems, the horizontal inertial force due to the facing units over the layer spacing height S_v should be included in the calculation of T_{max} . The example here is simplified by conservatively ignoring wall-soil friction, wall inclination, and weight of the facing. Parameter K_r is the static earth pressure coefficient. For geosynthetic reinforcement, $K_r = K_A$. For steel reinforcement cases, $K_r \geq 1.2 K_A$ at the top of the wall (with actual values depending on steel reinforcement type) and decreases linearly with depth

Design of Mechanically Stabilized Walls (MSWs),

Fig. 5 Calculation of maximum internal reinforcement loads



D

to $1.2K_A$ at a depth of 6 m. The distribution of dynamic horizontal pressure in the above example is assumed to be constant over the height of the wall. However, assumptions regarding the distribution of this dynamic pressure component can vary based on reinforcement type (geosynthetic or steel) and between different design codes. As noted in the previous section on static design principles, the tensile load T_{max} in the reinforcement must not exceed a critical value to prevent overstressing that can lead to rupture. For static design of geosynthetic MSWs, this value is largely determined by the constant load creep behavior of these polymeric materials. However, under transient dynamic loading, the residual strength of the reinforcement is assumed to be available. This strength is computed by ignoring the creep-induced strength loss that is assumed for the design life of a wall based on static loading only. In other words, the residual strength value is calculated as the original available strength of the reinforcement immediately after the end of construction.

In order for the internal active wedge to be stable, the reinforcement layers must have adequate anchorage (pullout) capacity. The pullout capacity (T_{pull}) can be computed as:

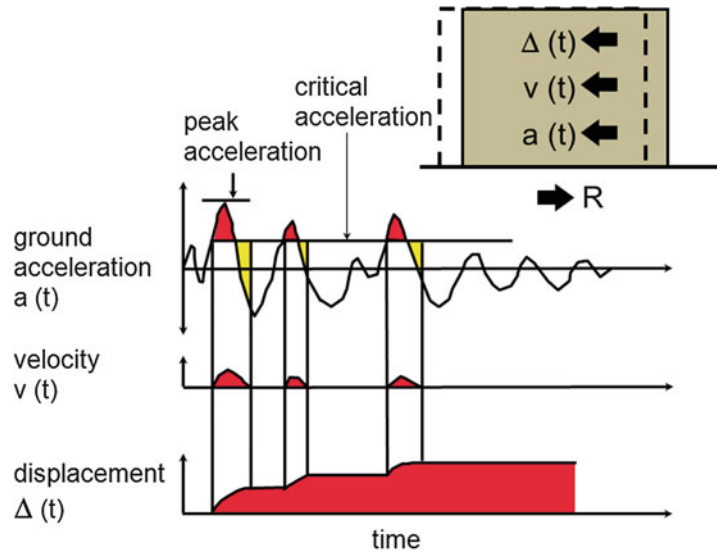
$$T_{pull} = 2Cf^*\sigma_v L_e \quad (7)$$

Here, C is the coverage ratio of the reinforcement in the plane of the reinforcement and is equal to one for continuous sheet reinforcement (e.g., geogrid, woven geotextile). Parameter f^* is the interface shear coefficient and is a function of the reinforcement type and soil type, and for the case of steel strips, it will also be a function of depth. Parameter σ_v is vertical stress acting at the elevation of the reinforcement and L_e is the anchorage length. The anchorage length can be computed knowing the geometry of the internal active wedge. In Fig. 5 this wedge is shown as triangular shape which is the typical assumption for MSWs constructed with relatively extensible geosynthetic reinforcement materials. For steel reinforcement, the active wedge is assumed to be bilinear in shape with a constant distance from the back of the wall equal to $0.3H$ to a depth of $H/2$ below the crest of the wall and then tracing a straight line to the heel of the wall face thereafter. For MSWs constructed with select backfills, there is no evidence to suggest that the value of f^* should be less for the case of earthquake loading than for the static loading condition.

Rigorous analytical or numerical solution of the active Coulomb wedge problem for the case of addition seismic inertial forces shows that the orientation of the internal failure wedge (α_{AE})

Design of Mechanically Stabilized Walls (MSWs),

Fig. 6 Double integration method to calculate displacements in sliding block analysis



illustrated in Fig. 3 will decrease with increasing magnitude of k_h . This can lead to excessive reinforcement lengths. Conventional practice for geosynthetic reinforced walls is to limit this angle to the static loading value which is reasonably approximated using the Rankine solution $\alpha = \pi/4 + \phi/2$. Similarly, the bilinear wedge shape for steel-reinforced MSWs is assumed to be applicable to both static and dynamic loading cases. Pullout may be a concern at the top of the wall where vertical stress (σ_v) is low. The solution is to locally increase the length of these layers (L_c) to meet an acceptable factor of safety against this mode of failure.

Sliding Block Displacement Methods

A shortcoming of pseudo-static design and analysis methods is that they cannot be used to predict serviceability performance of a wall with respect to displacement. Sliding block displacement methods have been used for many years to estimate the displacement of earth structures during earthquake. This general approach is easily adaptable to failure limit states for MSWs where the failure surface is horizontal (see Fig. 2a, f). The first step is to find the critical acceleration, k_c , which is the value of k_h that will give a factor of safety of unity in the corresponding static factor of safety expression. The critical acceleration is then

applied to the horizontal ground acceleration record at the structure, and double integration is performed to calculate the cumulative displacement, as illustrated in Fig. 6. Specifically, the cumulative displacement is calculated by integrating those portions of the accelerogram that are above the critical acceleration and those portions that are below until the relative velocity between the sliding mass and the sliding base reduces to zero.

Dynamic Finite Element or Finite Difference Methods

MSWs are complex mechanical systems and this complexity is compounded when the structures are subjected to transient dynamic loading due to earthquake. Pseudo-static and displacement methods may be sufficient for simple structures, for stable ground conditions and MSWs in low seismic risk areas, and/or for preliminary design. Otherwise, more sophisticated analyses may be warranted using advanced dynamic finite element model or finite difference computer programs. Today commercially available computer programs offer the user a suite of constitutive models for the component materials and in some cases allow the user to implement their own constitutive models. In order to accurately model the soil in a MSW, it may be necessary to use nonlinear cyclic stress-strain models. The equivalent-linear

method (ELM) is one example. The ELM requires strain-dependent shear stiffness degradation curves and damping curves to capture the reduction of shear modulus of the soil and the increase in soil damping that occurs with increasing shear strain. These parameters can be found in published databases or from project-specific resonant column testing for small shear strain amplitudes and cyclic triaxial tests for larger strain amplitudes. Another important consideration for the correct numerical modeling of MSWs with polymeric reinforcement materials is the strain-rate-dependent stiffness of these materials. Often forgotten is that the stiffness of polymeric PP and HDPE reinforcement products increases significantly with rate of loading. Hence, stiffness values deduced from conventional laboratory tensile tests at low strain rates will underestimate the stiffness of the reinforcement during dynamic loading. Consequently, numerical-predicted reinforcement loads will be nonconservative for design (i.e., underestimated).

Additional Considerations

Frequency Response

The seismic response of MSWs of typical heights (say $H < 10$ m) can be dominated by the natural frequency of the structure. A recommended check of the response of any soil structure is the proximity of the natural (resonant) frequency of the system (f_ω) to the predominant frequency of the input acceleration record (f_1). An approximate upper-bound estimate of the natural frequency of a MSW structure can be computed from linear-elastic theory as:

$$f_\omega = V_s/4H = 1/4H \sqrt{\frac{G}{\rho}} \quad (8)$$

where V_s = shear-wave velocity of the soil back-fill, G = shear modulus, and ρ = density of the soil. Modifications to this formula are available in the literature to adjust this one-dimensional solution to the case of two-dimensional rectangular volumes to match the reinforced soil zone shape.

Factors of Safety and Load and Resistance Factor Design (LRFD)

The concepts introduced in this entry for seismic design of MSWs have been related to margins of safety using classical notions of factor of safety. In this approach, factor of safety is simply the ratio of resistance capacity to load. In North American practice, the factors of safety against the modes of failure identified in Fig. 2 are reduced to values that are typically 75 % of static values or as low as 1.1. The exception is the factor of safety for overstressing (rupture) for geosynthetic reinforcement products which is taken as unity for seismic design.

Modern US design codes for MSWs now adopt a load and resistance factor design (LRFD) approach (AASHTO 2012; FHWA 2009). Load factors are applied to each load term in a limit state function, and a single resistance factor is applied to the resistance term in the same function. The factors have been computed using rigorous reliability theory to satisfy an acceptable minimum target probability of failure or to give design outcomes with equivalent factors of safety to those using the conventional factor of safety approach. Load and resistance factors are different for seismic and static design consistent with the expectation that severe earthquakes are relatively rare events over the design life of a typical MSW. In Europe, a partial factor of safety approach is used in which partial factors are applied to material properties on the resistance side of each limit state function. At present, partial factors have not been updated for seismic design of MSW systems in the British code (BSI 2010). Eurocode guidelines are limited to a summary of the general approaches to the design and analysis of geotechnical soil structures.

Performance of MSWs During Earthquake

In recent years, surveys of MSW structures have been carried out in the Americas and in Japan after earthquakes. These surveys and project-specific case studies of reinforced soil walls after earthquake have demonstrated generally good behavior, and, where direct comparisons are available, they have behaved better than conventional gravity structures and at least as well as

traditional reinforced concrete structures (Koseki et al. 2006; Yen et al. 2011). Geosynthetic MSWs constructed with full-height panel facings for railway track embankments and bridge abutments (including bullet train lines) were virtually undamaged and exhibited good seismic resistance during the recent 11 March 2011 Great East Japan Earthquake compared with conventional unreinforced embankments (JGS 2011).

The reason for this good performance is that MSWs are generally more flexible than conventional walls. They are more forgiving of the uneven ground displacements that may occur due to surface faulting during an earthquake event. Importantly, the ductility of these systems has prevented collapse of structures during earthquakes so that adjacent structures, property, or persons were not put at risk even though supporting infrastructure (e.g., railway lines) was no longer in service. Where there have been detectable performance problems, these are typically related to inadequate details at the wall corners, the joints between full-height panels, and the wall top due to inadequate coping details or toppling of masonry concrete blocks.

In addition to field surveys, there is now a large body of results from centrifuge and 1-g gravity tests on model walls subjected to simulated earthquake loading using laboratory shaking tables. Overturning about the toe of a MSW structure is rare and can be safely omitted for the external stability design of MSW structures. Similarly, bearing capacity failure is unlikely and this mode of failure is not expected for walls built on competent foundations. Field surveys and laboratory shaking table tests have demonstrated that MSWs designed to satisfy current factors of safety for static design are typically stable at peak ground accelerations up to 0.4 g if the structures are seated on competent level ground below and in front of the walls and the walls support horizontal backslopes (Mikola and Sitar 2013). Consequently, for these scenarios additional earthquake-induced loads need not be considered in design. The explanation for this recommendation is that the static design of these structures is very conservative. Hence, these walls have sufficient excess load capacity

so that static designs do not need to be augmented for peak ground accelerations less than 0.4 g. Details of the conditions that meet this waiver can be found in the American Association of State Highway and Transportation Officials design guidelines (AASHTO 2012). For wall cases exceeding the 0.4-g criterion, more sophisticated analyses executed by experienced and knowledgeable design engineers will be required. Often this will include displacement methods and numerical modeling using commercial computer software programs that use dynamic finite element or finite difference methods.

Summary

Mechanically stabilized walls (MSWs) are a well-accepted technology to perform the earth retaining wall function in civil engineering earthwork applications. The design of these structures for earthquakes is most often based on conventional extensions of static wall design that are well known and accepted by geotechnical engineers. The performance of MSWs during earthquake loading has been excellent in most cases due to their relative flexibility compared to conventional gravity wall structures and their reserve load capacity due to conservatism in the static design of these structures.

Cross-References

- [Analysis and Design Issues of Geotechnical Systems: Rigid Walls](#)
- [Engineering Characterization of Earthquake Ground Motions](#)
- [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)
- [Seismic Design of Earth-Retaining Structures](#)

References

- AASHTO (2012) AASHTO LRFD bridge design specifications, 6th edn. American Association of State Highway and Transportation Officials, Washington, DC

- BSI (2010) Code of practice for strengthened/reinforced soil and other fills (BS8006). British Standards Institution, Milton Keynes
- FHWA, Berg RR, Christopher BR, Samtani NC (2009) Design of mechanically stabilized earth walls and reinforced soil slopes. vol 1. FHWA GEC 011-VolI. National Highway Institute – Federal Highway Administration, U.S. Department of Transportation, Washington, DC
- JGS (2011) Lessons and recommendations from the 2011 Great East Japan earthquake. In: 2011 committee for Geo-hazards during earthquakes and mitigation measures. The Japanese Geotechnical Society, Tokyo, 91 p
- Koseki J, Bathurst RJ, Güler E, Kuwano J, Maugeri M (2006) Seismic stability of reinforced soil walls, Invited keynote paper. In: Proceeding of the 8th international conference on geosynthetics, Yokohama, Japan, 18–22 Sept 2006, 30 p
- Mikola RG, Sitar N (2013) Seismic earth pressures on retaining structures in cohesionless soils. Report No, UCB GR 13-01. Department of Civil and Environmental Engineering, University of California, Berkeley
- Yen W-HP, Chen G, Buckle I, Allen T, Alzamora D, Ger J, Arias JG (2011) Postearthquake reconnaissance report on transportation infrastructure impact of the February 27, 2010, Offshore Maule Earthquake in Chile. Publication No. FHWA-HRT-11-030. Office of Infrastructure Research and Development, Federal Highway Administration, McLean, VA

Design, Testing, and Evaluation of the Web Plastifying Damper for the Aseismic Protection of Buildings

Amadeo Benavent-Climent
Department of Structural Mechanics and
Industrial Constructions, Polytechnic University
of Madrid, Madrid, Spain

Synonyms

Earthquake resistant structure; Energy dissipation; Hysteretic damper; Metallic damper; Passive control

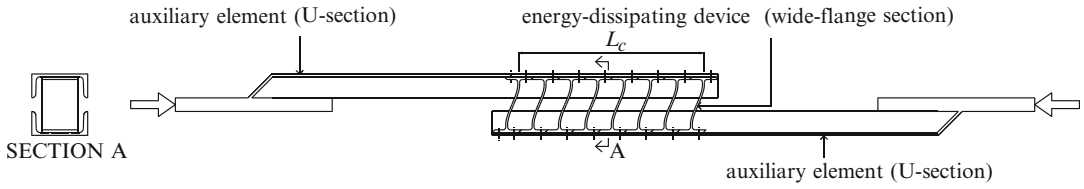
Introduction

Over the last three decades, structural control has achieved significant progress. Structural control includes three categories: passive, active, and

semi-active. Among them, passive control systems are less expensive and have the advantage of not requiring an external source of power. The passive control systems consist basically on installing energy-dissipating devices (dampers) in the main structure, in order to increase its energy dissipation capacity and to prevent or limit the damage in the main structure (plastic deformations) in case of severe earthquakes. The use of seismic energy-dissipative devices for passive control is increasing exponentially in recent years for both new and existing buildings (Martelli 2006). The challenge now is to produce low-cost and effective dampers feasible for massive use in developing countries.

Several mechanisms have been used for passive energy dissipation: yielding of metals, phase transformation of metals, friction sliding, fluid orificing, and deformation of viscoelastic solids or liquids. The former is one of the most popular and there are a number of ways to induce the yielding of the metal. In the ADAS damper (Berman and Goel 1987), yielding is attained by subjecting metallic plates to out-of-plane bending. In the honeycomb damper (Kobori et al. 1992) or the slit damper (Benavent-Climent et al. 1998) a metallic plate with openings is subjected to in-plane shear deformations. Another popular damper is the so-called Buckling Restrained Brace (Chan and Albermani 2008), which consists on yielding a metal core subjected to axial deformations, preventing buckling under compressive loads.

A new type of hysteretic damper, called Web Plastifying Damper (WPD), was recently proposed by Benavent-Climent et al. (2011) that is based on subjecting the web of standard wide-flange or I shape sections to out-of-plane bending, to dissipate energy in seismic applications. Yielding the web under out-of-plane bending is a potential source of energy dissipation because: (i) the transition from the web-ends to the flanges has a smooth curve shape (preventing stress concentrations), and (ii) the union between flange and web is weld-free, which avoids uncertainties and imperfections associated with welding. This energy-dissipating device requires a very simple manufacturing process (cutting a wide-flange or I-section in short segments and drilling openings



Design, Testing, and Evaluation of the Web Plastifying Damper for the Aseismic Protection of Buildings,
Fig. 1 Details of a typical WPD

in the flanges for the bolts). Further, leftovers of wide-flange or I-sections can be employed. For these reasons, the new damping device is quite economic.

Hereafter, the new type of seismic damper is presented and the guidelines and formulae for its design are provided. The energy dissipative characteristics of the web of wide-flange sections are investigated experimentally, and a simple hysteretic model for representing the load–displacement curve under arbitrary deformations is presented.

Design of the Web Plastifying Damper

The WPD is installed in a frame structure as a conventional brace or diagonal bar. It is constructed by assembling short length segments of wide-flange or I-shaped sections and two steel bars that remain elastic and function as auxiliary elements. If the brace damper is subjected to forced deformations in the axial direction, the web of the wide-flange or I-shape section undergoes out-of-plane flexural deformations. Figure 1 shows a prototype of the WPD, where standard U-shaped steel sections have been used as auxiliary elements.

The relative displacement δ_y between flanges at yielding, the yielding load Q_y , and maximum load Q_B of a single segment of wide-flange or I-shaped section of length l , clear height of the web h , and web thickness t are given by:

$$\delta_y = \frac{f_y h^2}{2tE} \quad (1)$$

$$Q_y = \frac{lt^2 f_y}{2h} \quad (2)$$

$$Q_B = \frac{lt^2 f_B}{2h}. \quad (3)$$

In the above equations, f_y and f_B are the yield and maximum tensile stress of the steel, and E the modulus of elasticity. The axial yield strength Q_{Ty} and apparent maximum strength Q_B of a damper constituted of n wide-flange of I-shaped segments are given by:

$$Q_{Ty} = n \frac{lt^2 f_y}{2h} \quad (4)$$

$$Q_B = n \frac{lt^2 f_B}{2h} \quad (5)$$

By summing the δ_y and the axial elastic deformation of the steel bars that function as auxiliary elements, the total axial displacement of the brace damper at yielding, δ_{Ty} , is obtained. The later contribution can be approximated by $f_{aux}L/E$, where L is the total length of the brace damper and f_{aux} is the normal stress in the cross-section of the steel bar caused by the axial force Q_{Ty} , that is:

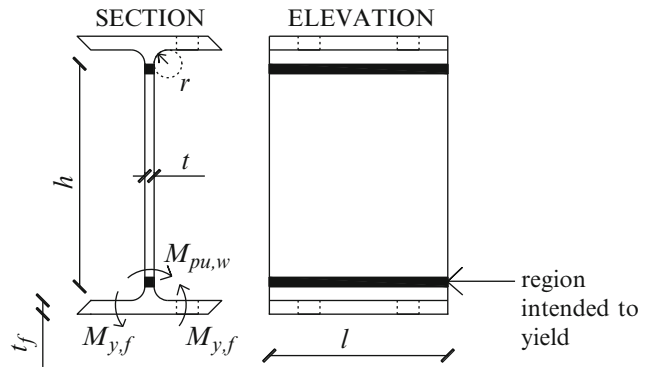
$$\delta_{Ty} = \frac{f_y h^2}{2tE} + \frac{f_{aux} L}{E} = \frac{1}{E} \left[0.5 f_y \left(\frac{h^2}{t} \right) + f_{aux} L \right] \quad (6)$$

From Q_{Ty} and δ_{Ty} , the axial stiffness K_T of the brace damper is given by:

$$K_T = \frac{Q_{Ty}}{\delta_{Ty}} = \frac{nl^3 f_y E}{f_y h^3 + 2thL f_{aux}} \quad (7)$$

In order to design a brace damper of given length L with target strength and yield

Design, Testing, and Evaluation of the Web Plastifying Damper for the Aseismic Protection of Buildings, Fig. 2 Test specimen



displacement, Q_{Ty} and δ_{Ty} , respectively, the following paths must be followed:

- Choose the type of steel and the stress level in the auxiliary elements f_{aux}
- Using Eq. 6 determine the required h^2/t
- Choose a commercial wide-flange or I-shape section that fits the required h^2/t
- Determine the value of the product nl to attain the target strength Q_{Ty} using Eq. 4
- Determine an appropriate combination of n and l , compatible with architectural constraints of the buildings
- Ensure that the following design requirements are fulfilled:

$$\frac{h}{t} > \frac{3\sqrt{3}}{4} \quad (8)$$

$$\left(\frac{t_f}{t}\right)^2 > \frac{3f_B}{4f_y} \quad (9)$$

Above requirements ensure that the WPD fails under out-of-plane flexural yielding of the web of the wide-flange or I-shape section, near the round-shaped transition zone from the web to the flanges (see Fig. 2).

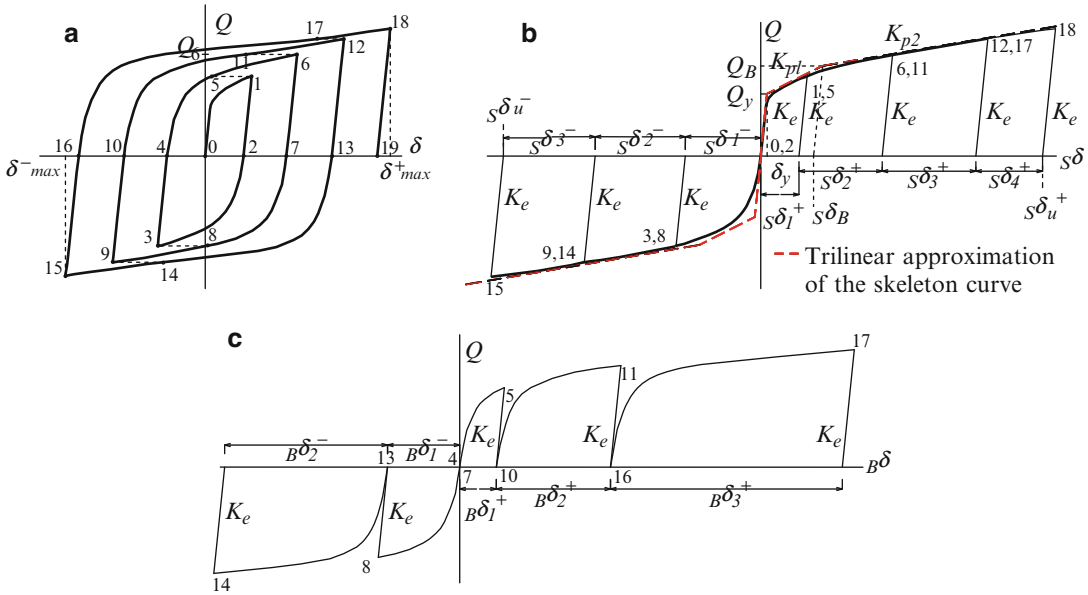
Testing the Web Plastifying Damper

Six component tests were conducted to investigate the energy dissipation capacity and to characterize the hysteretic behavior of the WPD. Each test consisted in subjecting 90 mm length specimens cut from a single segment of a structural

wide-flange section of $140 \times 73 \text{ mm}^2$ (total depth $\times$ flange width) – referred to as the IPE-140 section hereafter – to different histories of imposed out of plane bending, until failure. Figure 2 shows the test specimen.

The mechanical characteristics of each test specimen, in terms of length l , the clear height of the web h , and the web thickness t are: $l = 90 \text{ mm}$, $h = 112 \text{ mm}$, and $t = 5.4 \text{ mm}$. The specimens were made of mild steel S275, with $f_y = 340 \text{ N/mm}^2$, $f_B = 441 \text{ N/mm}^2$ and $E = 2.1 \times 10^5 \text{ N/mm}^2$. Applying above equations, the lateral displacement δ_y , the load Q_y at yielding, and the apparent maximum lateral load Q_B of a single IPE-140 section damping device are $\delta_y = 1.88 \text{ mm}$, $Q_y = 3.98 \text{ kN}$ and $Q_B = 5.17 \text{ kN}$.

The only variable considered in the tests was the loading history. Patterns of cyclic displacements were applied to the specimens until they failed. Failure was assumed to occur at the onset of strength degradation under increasing forced displacements. The cyclic rate applied by the actuator was 1/80Hz. In the first, second, and third tests cycles of incremental amplitude were applied to the specimens, with the increment of amplitude in each consecutive cycle normalized by the yielding displacement δ_y being $\phi = 2$, $\phi = 5$ and $\phi = 10$, respectively. The specimens were subjected to cycles of constant amplitude in the fourth and fifth tests; the amplitude normalized by the yielding displacement δ_y was $\Delta = 20$ and $\Delta = 30$, respectively. The loading protocol ATC-24 developed by the Applied Technology Council (Applied Technology Council 1992) was applied to the sixth specimen. This protocol



Design, Testing, and Evaluation of the Web Plastifying Damper for the Aseismic Protection of Buildings, Fig. 3 Decomposition of a Q - δ curve: (a) Q - δ curve; (b) skeleton part; (c) Bauschinger part

consists of several sets of three cycles; the amplitude of the cycles is constant within each set but it is increased every consecutive set of cycles, following the sequence $0.5\delta_y, 0.75\delta_y, 1.0\delta_y, 2\delta_y, 3\delta_y, 4\delta_y$ and so on.

Hysteretic Behavior

Figure 3a shows a typical load–displacement Q - δ hysteretic curve obtained from the test. It can be seen that the WPD has stable hysteretic behavior and high energy dissipation capacity.

Ultimate Energy Dissipation Capacity

Following the procedure formerly proposed by Kato et al. (1973) a Q - δ hysteretic curve such that is shown in Fig. 3a can be decomposed into: (i) the segments controlled by the Bauschinger effect, so-called “Bauschinger part”; (ii) the segments controlled by the strain-hardening phenomena, so-called “skeleton part”; (iii) and the unloading paths, so-called “unloading segments”. It has been shown (Kato et al. 1973) that the curve obtained by connecting subsequently the segments of the skeleton part coincides with the relation obtained from a monotonic test.

The scheme proposed by Kato et al. (1973) was applied for decomposing the Q - δ hysteretic curves obtained during the cyclic tests, as illustrated in Fig. 3a, b, and c. In these figures, the segments 0–1, 5–6, 11–12, 17–18 in the positive, and 2–3, 8–9, 14–15 in the negative domain are the paths that exceed the load level attained by the preceding cycle in the same domain of loading. Connecting these paths sequentially, the two curves called “skeleton part” (Fig. 3b) are obtained. These curves can be approximated by a three-segment curve that is shown with dash and dot lines in Fig. 3b. The trilinear approximation is defined by the yielding load, Q_y , the yielding displacement, δ_y , the first and second plastic stiffnesses K_{p1} and K_{p2} ($K_{p1} \geq K_{p2}$), and the load Q_B corresponding to the transition point from K_{p1} to K_{p2} . The unloading paths are the segments 1–2, 6–7, 12–13, 18–19, 3–4, 9–10, and 15–16. The slope of the unloading paths coincides with the initial elastic stiffness $K_e (=Q_y/\delta_y)$. In Fig. 3b, $s\delta_u^+$ and $s\delta_u^-$ are the plastic deformation accumulated in each skeleton curve when the WPD device fails. On the other hand, $s\delta_B$ is the plastic deformation accumulated

Design, Testing, and Evaluation of the Web Plastifying Damper for the Aseismic Protection of Buildings,
Table 1 Ultimate energy deformation capacity of the damping devices

Load pattern	Q_y kN	δ_y mm	$ep\eta$	$s\eta$	$B\eta$	η	$B\eta/s\eta$
$\phi = 2$	3.98	1.88	44.32	67.19	532.79	599.98	7.93
$\phi = 5$	3.98	1.88	60.93	97.62	355.60	453.53	3.64
$\phi = 10$	3.98	1.88	75.12	124.54	269.45	393.99	2.16
$\Delta = 20$	3.98	1.88	53.27	79.25	555.08	634.32	7.00
$\Delta = 30$	3.98	1.88	67.86	107.63	338.05	445.68	3.14
ATC-40	3.98	1.88	20.77	29.31	923.11	952.41	31.50

D

in the approximate trilinear skeleton curve at $Q = Q_B$. The area enveloped in Fig. 3b for each domain of loading by the skeleton curve is called hereafter sW_u^+ and sW_u^- . The segments 4–5, 10–11, and 16–17 in the positive domain and 7–8 and 13–14 in the negative domain of loading begin at $Q = 0$ and terminate at the maximum load level previously attained in

preceding cycles in the same loading domain; they constitute the so-called “Bauschinger part” (Fig. 3c). The sum of the area enveloped by the Bauschinger part for each domain of loading will be referred to as BW_u^+ and BW_u^- .

In addition, sW_u^+ , sW_u^- , BW_u^+ , BW_u^- , $s\delta_u^+$ and $s\delta_u^-$ can be expressed in a nondimensional by,

$$s\bar{\eta}^+ = \frac{sW_u^+}{Q_y\delta_y}; \quad s\bar{\eta}^- = \frac{sW_u^-}{Q_y\delta_y}; \quad B\bar{\eta}^+ = \frac{BW_u^+}{Q_y\delta_y}; \quad B\bar{\eta}^- = \frac{BW_u^-}{Q_y\delta_y}; \quad ep\bar{\eta}^+ = \frac{s\delta_u^+}{\delta_y}; \quad ep\bar{\eta}^- = \frac{s\delta_u^-}{\delta_y}, \quad (10)$$

Next, the following ratios are defined:

$$\begin{aligned} s\eta &= |s\bar{\eta}^+| + |s\bar{\eta}^-|; \quad B\eta \\ &= |B\bar{\eta}^+| + |B\bar{\eta}^-|; \quad ep\eta \\ &= |ep\bar{\eta}^+| + |ep\bar{\eta}^-|; \quad \eta = s\eta + B\eta \end{aligned} \quad (11)$$

$ep\eta$ denotes the apparent ultimate cumulative plastic deformation ratio on the skeleton part; $s\eta$ is the ultimate cumulative plastic deformation ratio on the skeleton part; $B\eta$ is the ultimate cumulative plastic deformation ratio on the Bauschinger part; η is the total ultimate cumulative plastic deformation ratio. η represents in nondimensional form the total energy dissipated by the damper until failure, i.e. $W_u (= sW_u^+ + sW_u^- + BW_u^+ + BW_u^-)$.

Table 1 shows the values of the parameters defined above obtained from each test. It can be seen in the table that the total energy dissipated by the WPD as expressed by η , and the distribution of this energy between the skeleton and the

Bauschinger parts (i.e., $B\eta/s\eta$) depend on the loading pattern. If the WPD is subjected to cycles of small constant amplitude (i.e., the loading pattern with $\Delta = 20$) or to cycles whose amplitude grows in small increments (i.e., the loading patterns ATC-40 or with amplitude $\phi = 2$), η and $B\eta/s\eta$ are larger than when applying cycles of large constant amplitude (i.e., the loading pattern with $\Delta = 30$) or than cycles whose amplitude enlarges in large increments (i.e., the loading pattern with $\phi = 10$).

It can be seen that the relation between the values $ep\eta$ and $B\eta$ shown in Table 1 is almost linear and can be formally expressed by:

$$B\eta = a_{ep}\eta + b. \quad (12)$$

The value of parameters a and b for the WPDs tested are $a = -12$ and $b = 1140$. It has been shown in previous research that while the relation between $B\eta$ and $ep\eta$ is linear irrespective of the steel type, the values of a and b depend on the

material properties and must be determined for each type of steel. a and b can be obtained by testing only two specimens under different patterns of cyclic loading.

As for the skeleton part, the approximated trilinear curve shown in Fig. 3b can be normalized by dividing the abscissa by δ_y given by Eq. 1, and the ordinate by Q_y given by Eq. 2. The resulting values of the normalized plastic stiffness $k_{p1} = K_{p1}/(Q_y/\delta_y)$, and $k_{p2} = K_{p2}/(Q_y/\delta_y)$ for the WPD tested, are $k_{p1} = 1/5$ and $k_{p2} = 1/30$. Now, a new parameter τ_B is defined to characterize the normalized trilinear skeleton curve as

$\tau_B = Q_B/Q_y$. Using these parameters, the energy dissipated on the skeleton part expressed in nondimensional form by $s\eta$ can be related to $ep\eta$ assuming that the skeleton curves in the positive and negative domains of loading are equal and that $|_{ep\eta}| = |_{ep\eta}| = |_{ep\eta}| = 0.5ep\eta$:

$$\text{for } ep\eta \leq \frac{2(\tau_B - 1)(1 - k_{p1})}{k_{p1}} : \quad (13)$$

$$s\eta = 0.25ep\eta^2 \frac{k_{p1}}{1 - k_{p1}} + ep\eta$$

$$\text{for } ep\eta > \frac{2(\tau_B - 1)(1 - k_{p1})}{k_{p1}} : \quad (14)$$

$$s\eta = \frac{(\tau_B^2 - 1)(1 - k_{p1})}{k_{p1}} + \left(\frac{ep\eta}{2} - \frac{(\tau_B - 1)(1 - k_{p1})}{k_{p1}} \right) \left[2\tau_B + \frac{k_{p2}}{(1 - k_{p2})} \left(\frac{ep\eta}{2} - \frac{(\tau_B - 1)(1 - k_{p1})}{k_{p1}} \right) \right]$$

By adding Eqs. 12, 13, and 14, the ultimate energy dissipation capacity in terms of the nondimensional ratio η is obtained as follows:

$$\text{for } ep\eta \leq \frac{2(\tau_B - 1)(1 - k_{p1})}{k_{p1}} : \quad (15)$$

$$\eta = 0.25ep\eta^2 \frac{k_{p1}}{1 - k_{p1}} + ep\eta(1 + a) + b$$

$$\text{for } ep\eta > \frac{2(\tau_B - 1)(1 - k_{p1})}{k_{p1}} : \quad \eta = \frac{(\tau_B^2 - 1)(1 - k_{p1})}{k_{p1}} + \left(\frac{ep\eta}{2} - \frac{(\tau_B - 1)(1 - k_{p1})}{k_{p1}} \right) \left[2\tau_B + \frac{k_{p2}}{(1 - k_{p2})} \left(\frac{ep\eta}{2} - \frac{(\tau_B - 1)(1 - k_{p1})}{k_{p1}} \right) \right] + aep\eta + b \quad (16)$$

Numerical Model for Characterizing the Hysteretic Behavior of the WPD

The load–displacement relationship of the WPD under an arbitrary cyclic loading pattern can be predicted with a simple hysteretic model, and a procedure for predicting its failure. In this hysteretic model, the shape of the skeleton part and the shape of each segment of the Bauschinger are determined as follows.

For characterizing the shape of the skeleton part, the trilinear idealization proposed above is adopted. The shape of the Bauschinger parts is studied next. Focusing on Fig. 3a, consider for example that under a given loading history, the WPD has covered the path 0-1-2-3-4-5-6-7-8-9-10, and an approximation for the Bauschinger segment 10–11 located in the positive loading domain is sought. The ordinate at the origin of the segment (i.e., point 10) is $Q = 0$, and the

ordinate at the end (i.e., point 11) is known because it coincides with the maximum force Q_m attained in the skeleton part in previous cycles of deformation in the same domain of loading (i.e., the ordinate at point 6 which is referred to as $Q_m = Q_6$ in Fig. 3a). The Bauschinger segment 10–11 can be further approximated with a bilinear curve determined by three parameters: (i) the stiffness of the first line; (ii) the displacement ${}_B\delta$ of the segment (i.e., ${}_B\delta_2^+$ in Fig. 3c for the example segment 10–11); and (iii) the ordinate of the intersection point of the two lines, which will be expressed as a fraction α of the maximum force attained in the skeleton part in previous cycles of deformation in the same domain of loading αQ_m (i.e., αQ_6 for the example segment 10–11). For the WPD, and based on the results of the test, these parameters can be set as follows: (i) the first stiffness is taken equal to the initial elastic stiffness K_e ; (ii) ${}_B\delta$ is taken as a fraction β of the displacement accumulated in the skeleton part $\sum {}_S\delta$ up to the start of the Bauschinger segment under consideration (i.e., for the example segment 10–11, ${}_B\delta_2^+ = \beta\{|_S\delta_1^+| + |_S\delta_1^-| + |_S\delta_2^+| + |_S\delta_2^-|\}$); and (iii) the value of α is determined so that the area (i.e., the dissipated energy) under the Bauschinger segment obtained from the test and the area under the approximated bilinear curve are equal. Based on the test results, for the WPD the following values can be adopted: $\alpha = 0.66$ and $\beta = 0.61$.

Once the shape of the skeleton part and the Bauschinger part are determined by the parameters Q_y , Q_B , δ_y , k_{p1} , k_{p2} , α and β , the hysteretic curve of the WPD under an arbitrary history of displacements can be constructed as follows. The first time the damper is loaded in a given domain (positive or negative), the proposed model considers that the Q - δ relationship follows the skeleton curve. Referring to Fig. 3a, consider for example that, first – starting from point 0 – the damper is deformed an amount ${}_S\delta_1^+$ in the positive domain; it will follow the skeleton curve moving from point 0 to 1. At point 1 the deformation accumulated on the skeleton part is $\Sigma_S\delta = |_S\delta_1^+|$ and the corresponding force attained

is Q_I . If the damper is unloaded at point 1, the Q - δ curve will follow the unloading line of stiffness K_e ($=Q_y/\delta_y$) until point 2. If the WPD is now loaded in the opposite domain of loading, the Q - δ curve follows the skeleton curve, from point 2 to 3. $\Sigma_S\delta = |_S\delta_1^+| + |_S\delta_1^-|$ is the deformation accumulated on the skeleton part at point 3. If the damper is unloaded at point 3 it will again follow a line of stiffness K_e until point 4. Reloading the damper in the positive domain, the Q - δ curve will follow a Bauschinger path whose amplitude ${}_B\delta$ is ${}_B\delta_1^+ = \beta\{|_S\delta_1^+| + |_S\delta_1^-|\}$. Within this Bauschinger path, initially the Q - δ curve follows a line of stiffness K_e until the point of ordinate αQ_I ; from here, the Q - δ curve follows a line oriented to a point determined by ${}_B\delta_1^+$ and Q_I . Following this rule, the complete hysteretic curve under arbitrary loading can be constructed until failure.

For the prediction of the failure of the WPD, deformation accumulated on the skeleton curve in each step i of the loading process, $\Sigma_S\delta_i$, and the total energy W_i dissipated by the WPD up to this step, must be monitored. The corresponding ratios ${}_{ep}\eta_i$ and η_i at step i are: ${}_{ep}\eta_i = \Sigma_S\delta_i/\delta_y$ and $\eta_i = W_i/(Q_y\delta_y)$. Failure is assumed to occur when this point reaches the “failure line” defined by Eqs. 15 and 16 (Benavent-Climent 2007).

Summary

A new type of seismic damper based on plastifying the web of short length segments of wide-flange or I-shaped steel sections under out-of-plane bending is presented. The damper can be installed in a frame like a conventional brace. The damper device was tested to characterize its hysteretic behavior and ultimate energy dissipation capacity. The results of the tests indicate that the seismic damper has stable restoring force characteristics and high energy dissipation capacity. Further, the load–displacement curve of the damper under arbitrarily applied cyclic load can be easily predicted with a simple numerical model. Also, its ultimate energy dissipation

capacity and the proximity to failure can be anticipated with a procedure that takes into account the influence of the loading path.

References

- Applied Technology Council (1992) Guidelines for cyclic seismic testing of components of steel structures ATC-24. Applied Technology Council, Redwood City
- Benavent Climent A, Oh S, Akiyama H (1998) Ultimate energy absorption capacity of slit-type steel plates subjected to shear deformations. *J Struct Constr Eng* 503:139–147
- Benavent-Climent A (2007) An energy-based damage model for seismic response of steel structures. *Earthq Eng Struct Dyn* 36:1049–1064
- Benavent-Climent A, Morillas L, Vico JM (2011) A study on using wide-flange section web under out-of-plane flexure for passive energy dissipation. *Earthq Eng Struct Dyn* 40:473–490. doi:10.1002/eqe.1031
- Bergman DM, Goel SC (1987) Evaluation of cyclic testing of steel plate devices for added damping and stiffness. Report UMCE87-10, University of Michigan, Ann Arbor
- Chan RWK, Albermani F (2008) Experimental study of steel slit damper for passive energy dissipation. *Eng Struct* 30:1058–1066
- Kato B, Akiyama H, Yamanouchi H (1973) Predictable properties of structural steels subjected to incremental cyclic loading. IABSE Symposium on resistance and ultimate deformability of structures acted on by well defined loads, Lisbon
- Kobori T, Miura Y, Fukusawa E, Yamada T, Arita T, Takenake Y, et al (1992) Development and application of hysteretic steel dampers. In: *Proceedings of 11th world conference on earthquake engineering*, Madrid, pp 2341–2346
- Martelli A (2006) Modern seismic protection systems for civil and industrial structures. An advanced approach to earthquake risk scenarios, with applications to different European towns. At: http://www.samco.org/network/download_area/paper_martelli.pdf

Downhole Seismometers

Mark B. Hayman
Nanometrics Inc., Ottawa, Canada

Synonyms

Borehole seismometers; Direct burial installation; Posthole seismometers

Introduction

This entry reviews the topic of downhole seismometers. The entry is split into three sections. The first section reviews the different types of downhole seismometers. The second section provides an overview of seismic noise sources and their impact on downhole installations. The last section discusses the downhole installation methods.

There are three main motivations for deploying seismometers in a downhole environment. The first is to provide quieter seismic noise environment to improve signal-to-noise ratios. The second motivation is to lower the cost of deployment of a seismic station. The third motivation is to reduce the environmental impact and land use costs by minimizing the physical footprint of a seismic station.

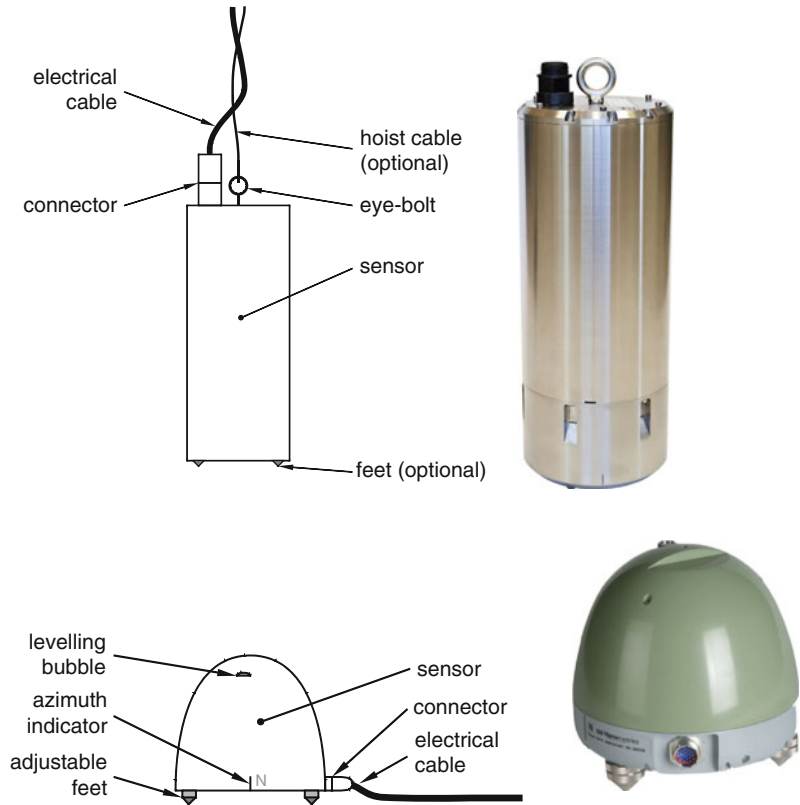
Downhole Seismometer Types

Seismometers can be categorized by the installation environment into three broad classes of instruments, surface (or vault) seismometers, ► [ocean bottom seismometers](#) (OBS), and downhole seismometers. Downhole seismometers are seismometers that are designed for installation below the surface, ocean bottom seismometers are designed for deployment underwater to depths of 7,000 m or more, and vault seismometers are designed for installation in seismic vaults at or near the surface. The construction of vaults is described in some detail in the “► [Principles of Broadband Seismometry](#)” entry.

Although these three classes of seismometers are markedly different in appearance, the internal sensing transducers and associated electronics are typically similar in design and performance. The difference is in the packaging and topology: in vault seismometers, the three axes are normally side by side, in downhole seismometers the axes are often stacked vertically, and in OBS seismometers either topology can be utilized. Typically, downhole seismometers are packaged in tall cylindrical stainless steel

Downhole Seismometers,

Fig. 1 Posthole and vault seismometers



pressure tubes and are watertight to a depth of 50–1,000 m. Vault seismometers are normally not designed for continuous submersion in water. The diameter of a downhole seismometer is minimized to enable the seismometer to fit down narrow boreholes. The diameter of a vault seismometer is not as critical as a downhole seismometer, and vault seismometers are typically shorter and wider for stability (see Fig. 1).

Vault seismometers are usually leveled by hand using a reference bubble level and adjusting three leveling feet. Downhole seismometers are normally not accessible for manual leveling and have automatic leveling mechanisms where leveling is required.

Downhole seismometers can be further broken down into two different instrument types, namely, borehole seismometers and posthole seismometers.

Borehole seismometers are downhole seismometers that are designed for deployment in steel-cased deep boreholes. These seismometers

typically include a holelock assembly as part of the instrument. A holelock is used to mechanically support and connect the seismometer to the borehole casing. It includes an actuator or a passive spring that pushes the entire seismometer against the side of the borehole with a substantial force. This force holds the seismometer rigidly in place and prevents the seismometer sliding down the hole.

A posthole seismometer is a downhole seismometer that does not include a holelock and is designed for direct burial or installation in shallower, uncased holes, known as postholes. A posthole seismometer is designed to be installed in sand at the bottom of a posthole. The sensor will typically also have three fixed feet, which can be useful in side-by-side testing on a pier, or placement at the bottom of a cased hole, as long as that surface is reasonably hard and level. These seismometers are designed for holes up to 50 m deep and may not have the underwater pressure rating of a borehole

seismometer. Furthermore, a posthole seismometer's drivers and power supplies may not support the cable lengths required for a borehole seismometer.

A downhole seismometer can be a single-component seismometer, meaning that the seismometer only measures vertical motion or that it can be a three-component seismometer which measures vertical, north–south, and east–west motion. Three-component seismometers are also known as triaxial seismometers.

An important parameter in selecting a downhole seismometer is the self-noise level of the instrument. The self-noise of a seismometer defines the limit of the instrument's ability to resolve ground motion. The self-noise of a seismometer is often graphed with respect to the minimum earth noise as defined by the new low-noise model (NLNM) (Peterson 1993). Although a self-noise graph provides a comprehensive view of seismometer performance over frequency, the long-period performance of an instrument is typically the limiting parameter of the self-noise. This is because all seismometers have self-noise lines that steadily curve upward at long periods.

Seismometers can be further categorized by the lower corner frequency (–3 dB point) of the amplitude frequency response of the instrument. Geophones have lower corner frequencies from 1 to 40 Hz. Short-period seismometers have lower corner frequencies from 1 to 4 Hz, and broadband seismometers have lower corner frequencies from 0.027 to 1 Hz. Broadband seismometers typically have lower noise floors over wider bandwidths than short-period seismometers, and short-period seismometers are typically quieter than geophones. Accelerometers are approaching the performance levels of some geophones and short-period seismometers and can be considered for downhole applications too.

Nevertheless, in most downhole applications today, broadband seismometers are typically used since they can measure all three components of motion, are quieter instruments, and can be realized in packages that fit downhole. Broadband seismometers are more challenging to install downhole than short-period seismometers

due to lower tilt tolerances. This entry will focus on broadband downhole seismometers and their installation. However, many of the installation techniques and challenges presented in this entry apply to short-period seismometers, accelerometers, and geophones in varying degrees.

Leveling

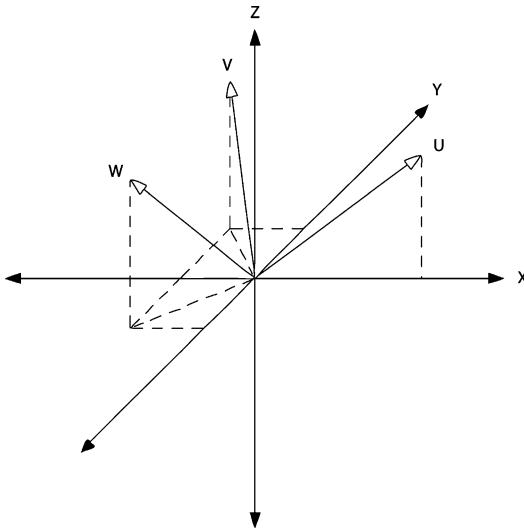
In order for a seismometer to measure true vertical and north–south, east–west horizontal motion, the verticality error due to the as-drilled borehole must be compensated for along with the north–south orientation. The methods of north–south orientation will be discussed in a later section (see section “[Water Egress](#)”), with this section focusing on leveling. Since a downhole seismometer cannot be leveled manually because it is down a borehole or posthole, downhole seismometers have automatic leveling mechanisms that can adjust the sensing elements or axes to measure true vertical and horizontal motion.

The leveling range of a borehole seismometer is typically 5° to cover the range of verticality error due to the drilling. Some downhole seismometers have wider leveling ranges, but this is only needed in special situations.

Leveling mechanisms in seismometers are challenging to design due to the space constraints inside the seismometer, the need to have high fidelity coupling to ground motion, and the need to maintain a design that is not subject to spontaneous micro-mechanical movements. Different instrument designers have taken different approaches to designing leveling mechanisms for seismometers.

There are four main methods for leveling sensing axes within a seismometer, namely, gimbal leveling, horizontal axis leveling, mainspring adjustment, and electronic centering. Each of these methods has benefits and drawbacks that a user should be aware of in selecting a downhole seismometer. Some of the drawbacks relate to the axis topology within the seismometer. More details will be given below.

Within a triaxial seismometer, there are three sensing components or axes which measure motion in three orthogonal directions. These



Downhole Seismometers, Fig. 2 Axis orientations for XYZ and UVW topologies

axes can be configured in a familiar XYZ topology wherein X measures east–west motion, Y measures north–south motion, and Z measures vertical motion. The axes can also be configured in a Galperin or symmetric triaxial topology. In this topology the three axes are tilted at 54.7° from horizontal plane and 120° apart around the vertical axis. These axis directions are known as UVW, wherein U is 52.5° elevation, north 90° west azimuth; V is 54.7° elevation, north 30° east; and W is 54.7° elevation, north 120° east azimuth (Graizer 2009). The symmetric triaxial topology is an orthogonal arrangement where coordinate system rests on its corner (see Fig. 2). The transformation between the UVW coordinate system and the XYZ coordinate system is shown below:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \frac{1}{\sqrt{6}} \cdot \begin{bmatrix} 2 & 0 & \sqrt{2} \\ -1 & \sqrt{3} & \sqrt{2} \\ -1 & -\sqrt{3} & \sqrt{2} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{\sqrt{6}} \cdot \begin{bmatrix} 2 & -1 & -1 \\ 0 & \sqrt{3} & -\sqrt{3} \\ \sqrt{2} & \sqrt{2} & \sqrt{2} \end{bmatrix} \cdot \begin{bmatrix} u \\ v \\ w \end{bmatrix}$$

Mathematically, leveling is the rotation of the seismometer axes about the X-axis and Y-axis

only. Rotation about the Z-axis is not handled by the leveling system. Z-axis rotations of a seismometer align the seismometer to true north. This will be discussed separately. A basic rotation of angle θ about the X-axis can be written in 3×3 matrix form:

$$R_x(\theta) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix} \quad (1)$$

Similarly, a single rotation of angle α about the Y-axis is

$$R_y(\alpha) = \begin{bmatrix} \cos \alpha & 0 & \sin \alpha \\ 0 & 1 & 0 \\ -\sin \alpha & 0 & \cos \alpha \end{bmatrix} \quad (2)$$

The leveling process is the combination of the two rotations and can be written as

$$R_t = R_x(\theta) \cdot R_y(\alpha) \quad (3)$$

$$R_t = \begin{bmatrix} \cos \alpha & 0 & \sin \alpha \\ \sin \theta \sin \alpha & \cos \theta & -\sin \theta \cos \alpha \\ \cos \theta \sin \alpha & \sin \theta & \cos \theta \cos \alpha \end{bmatrix} \quad (4)$$

If the starting orientation of the seismometer axes is the vector, $X'Y'Z'$, and the leveled position is XYZ, then it can be written:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = R_t \cdot \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad (5)$$

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \cos \alpha & 0 & \sin \alpha \\ \sin \theta \sin \alpha & \cos \theta & -\sin \theta \cos \alpha \\ \cos \theta \sin \alpha & \sin \theta & \cos \theta \cos \alpha \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \quad (6)$$

Expanding these equations results in:

$$x = x' \cos \alpha + z' \sin \alpha \quad (7)$$

$$y = x' \sin \theta \sin \alpha + y' \cos \theta - z' \sin \theta \cos \alpha \quad (8)$$

$$z = x' \cos \theta \sin \alpha + y' \sin \theta + z' \cos \theta \cos \alpha \quad (9)$$

The Y-component is a more complicated expression (Eq. 8), but that is the outcome of the order of the rotations. Changing the order of the rotations, namely,

$$R_t = R_y(\alpha) \cdot R_x(\theta) \quad (10)$$

results in the following expression for the Y-component:

$$y = y' \cos \theta - z' \sin \theta \quad (11)$$

Notice this equation has the same form as the X-component in Eq. 7, so for conceptualizing, this form can be used for both the X-component and Y-component.

Equations 7, 8, and 9 are important because some leveling methods do not actually level the seismometer axes. This results in the seismometer not outputting a true XYZ signal, but rather an XYZ as defined by Eqs. 7, 8, and 9 where an axis output is mixed with signals from the other two axes. These equations will be referred to as each leveling method is discussed in detail.

Gimbal Leveling

With gimbal leveling, there is a leveling platform on which all three seismometer axes are mounted. The seismometer axes are orthogonal to one another in both a symmetric triaxial and a horizontal-vertical topology, and the seismometer axes are fixed to the platform. The platform is leveled around the X and Y axes to eliminate tilt in the X and Y directions. Typically, motorized adjusters are used for leveling in the two directions. There are some passive leveling designs that rely on gravity to shift the gimbal to level using a mass under the gimbal. The seismometer axes can be stacked vertically or placed side by side on the leveling platform. Once the gimbal platform is leveled, the seismometer axes would need to be mass centered to eliminate any residual leveling error and to correct for temperature and spring relaxation. This is a more complex design as compared to the other methods, but it is more accurate in that the seismometer is outputting true vertical and horizontal signals. Referring to

Eqs. 7, 8, and 9, the angles θ and α are zero after leveling, and the three equations simplify to:

$$x = x'$$

$$y = y'$$

$$z = z'$$

Horizontal Axis Leveling

The horizontal axis leveling method is a variation of gimbal leveling where the vertical axis is fixed and not leveled, and the two horizontal axes are leveled in one direction only. This method can only be used on XYZ topology seismometers. It cannot be used in symmetric triaxial seismometers because it would affect the mixing ratios when converting from UVW to XYZ. Referring to Eqs. 7, 8, and 9 again, the vertical axis is fixed in position, so Eq. 9 describes the Z-component signal. The east-west axis or X-axis is leveled about the Y-axis, so Eq. 7 with α equal to zero holds. Lastly, the north-south axis or Y-axis is leveled about the X-axis, so Eq. 8 with θ equal to zero describes the Y-axis signal:

$$x = x' \cos \alpha + z' \sin \alpha, \alpha = 0$$

$$x = x'$$

$$y = x' \sin \theta \sin \alpha + y' \cos \theta - z' \cos \theta \cos \alpha, \theta = 0$$

$$y = y'$$

$$z = x' \cos \theta \sin \alpha + y' \sin \theta + z' \cos \theta \cos \alpha$$

With this leveling method, the X- and Y-components measure true X and Y motions, respectively, but the Z-component does not measure true Z motion. This is not as bad as it seems as the angles are small. Substituting a typical maximum tilt range of 5° in two directions results in:

$$\alpha = 5^\circ, \theta = 5^\circ$$

$$z = x' \cos (5) + y' \sin (5) + \cos (5)$$

$$z = 0.087x' + 0.087y' + 0.992z'$$

The vertical sensitivity is reduced by 0.76 % which is not much at all given that it is hard to calibrate a seismometer to better than 1 %. The Z signal also has X- and Y-components mixed in, but at a level ten times smaller than the Z-component. Expressed a different way, there is a loss of orthogonality between the axes. Herein lie the downsides of this method, the loss of orthogonality between axes resulting in the mixing of the Z-axis by X- and Y-components. If the tilt angles were known, then the vertical axis could be corrected after the measurement, but typically the tilt angles are not recorded so it cannot be corrected.

The main problem with this method is that the horizontal and vertical ground motions do not have the same magnitudes at long periods. Long-period horizontal signals are affected by tilt noise, which result in the apparent horizontal ground motion being much larger than the vertical. It is common for horizontal components to have amplitudes 20–50 dB higher at long periods than vertical components. With these levels, the horizontal components mixed into the vertical become significant and can swamp out the vertical signal at long periods. The user needs to be careful about how much tilt noise there is on the horizontals and the degree of tilt of the sensor. This complicates the deployment and places a burden of understanding on the user.

The primary benefit of this method is that it is simpler and less expensive to realize in a seismometer than gimbal leveling. This is because axes in downhole seismometers are typically stacked vertically due to space constraints on the diameter of the borehole. It is challenging to level three axes stacked together, so each axis is leveled individually. If each axis was leveled individually in the X and Y directions, there would need to be six leveling actuators in all: three axes times two directions. The horizontal axis leveling method only requires two leveling actuators, one on the X-axis and one on the Y-axis. There is a saving of four leveling actuators, hence the cost benefit.

Mainspring Adjustment

The mainspring adjustment method can be used in symmetric triaxial seismometers. This method does not actually involve leveling any part of the seismometer. Instead it compensates for the seismometer being off-level by adjusting the tension in the mainspring of the seismometer. All inertial sensors see an apparent horizontal acceleration when they are tilted. For small angles the apparent horizontal acceleration H corresponding to a tilt θ in radians is simply

$$H = g_0 \theta$$

where $g_0 = 9.8 \text{ m/s}^2$ is the acceleration due to gravity.

The equilibrium position is maintained by adjusting the tension in the mainsprings of each of the axes to exactly compensate for this apparent acceleration. The tension in the mainspring is adjusted by means of an actuator connected to the mainspring. Many seismometers already have a mainspring actuator to compensate for variations in the elastic modulus with temperature. Expanding the functionality of this actuator to include leveling does not require any more components so there is no impact on cost. However, there are trade-offs in the design which limit the usefulness of this approach.

There is a limit to how far a mainspring can be adjusted due to the geometry and size of a seismometer axis. In addition, with active broadband seismometers, there is a trade-off between mainspring adjustment leveling range and the noise floor of the seismometer due to noise in the seismometer feedback electronics. A low-noise floor seismometer using mainspring adjustment cannot have the required leveling range. A broadband seismometer with a high noise floor would be able to have the required leveling range, so this method is limited to high-noise floor seismometers.

For this method of leveling, Eqs. 7, 8, and 9 apply:

$$x = x' \cos \alpha + z' \sin \alpha \quad (7)$$

$$y = x' \sin \theta \sin \alpha + y' \cos \theta - z' \cos \theta \cos \alpha \quad (8)$$

$$z = x' \cos \theta \sin \alpha + y' \sin \theta + z' \cos \theta \cos \alpha \quad (9)$$

Downhole Seismometers, Table 1 Comparison of leveling methods

Leveling method	Maintains orthogonality	True leveling	Cost	Compatibility
Gimbal leveling	Yes	Yes	High	All
Horizontal axis leveling	No	Partial	Medium	XYZ only
Mainspring adjustment	Yes	No	Low	UVW only
Electronic centering	Yes	No	Low	Active only

The seismometer axes are fixed in orthogonal positions within the seismometer, so there is no loss of orthogonality, but the XYZ components do not measure true vertical and horizontal signals.

This method of leveling is less costly to implement, but it is limited to higher-noise seismometers.

Electronic Centering

Electronic centering can only be implemented in force balance seismometers. With this method, the electromagnetic coil circuit known as a forcing coil is used to hold the inertial mass in the equilibrium position and counterbalance change in the gravity vector. Any change in the apparent horizontal acceleration due to gravity is compensated by the forcing coil and electronic feedback. As an axis is tilted, the current in the coil changes which changes the force on the inertial mass. For a horizontal axis, the apparent acceleration H produced is related to the feedback current I , motor constant H , and boom mass M by

$$H = \frac{G}{M} I$$

This method preserves the orthogonality of the axes as the axes are in fixed positions and again, Eqs. 7, 8, and 9 apply, so the XYZ components are not true verticals and horizontals. This is the same as the mainspring adjustment method.

This method has a few disadvantages. The seismometer consumes more power since it is the forcing coil current that is counterbalancing the gravity vector now. In addition, inertial mass must be small to keep the current low and still achieve a wide tilt range, which results in higher levels of Brownian noise (see “► [Principles of Broadband Seismometry](#)”).

There are a number of advantages of this method. There are no leveling actuators or mainspring adjustment actuators, making the design more reliable and lower in cost. Depending on the lower corner frequency of the seismometer (see “► [Principles of Broadband Seismometry](#)”), the seismometer can have a wide leveling range of up to 10° .

Downhole Seismometer Selection

There are a number of criteria to consider in selecting a downhole seismometer. These include self-noise floor and axis topology, the cost, the leveling method, the leveling range, the type of seismometer, the physical dimensions of the seismometer, and the reliability and consistency of the seismometer.

The first item to consider in the selection process is the self-noise floor of the seismometer. The choice should depend on the noise environment that the sensor will be deployed in and on nature of the application. Ideally, site noise should be the limiting factor in the overall station noise, not seismometer self-noise.

However, the performance needs to be balanced with the cost of the instrument. Instruments with lower self-noise cost more than those with higher levels of self-noise. It is up to the user to trade off performance with cost for their application.

For the axis topology, the symmetric triaxial topology is particularly advantageous in the downhole environment due to the ease of differentiating between installation- and sensor-related noise (see “► [Symmetric Triaxial Seismometers](#)”). The four leveling methods each have cost and performance trade-offs which are summarized in Table 1.

For high-performance broadband seismometers, the most suitable leveling method is the

Downhole Seismometers, Table 2 A sample of downhole broadband seismometers available on the market in 2013

Seismometer name	Manufacturer	Lower corner frequency	Axis topology	Leveling method	Leveling range
Trillium 120 Borehole	Nanometrics	120 s	Symmetric triaxial	Gimbal	$\pm 5^\circ$
Trillium Posthole	Nanometrics	120 s	Symmetric triaxial	Gimbal	$\pm 5^\circ$ standard $\pm 10^\circ$ optional
Trillium Compact Posthole	Nanometrics	120 s	Symmetric triaxial	Electronic centering	$\pm 2.5^\circ$ $\pm 10^\circ$
KS-54000	Geotech	360 s	XYZ	Unknown	$\pm 10^\circ$
KS-2000BH	Geotech	120 s (standard)	XYZ	Unknown	$\pm 15^\circ$
CMG-3TB	Guralp	120 s	XYZ	Horizontal axis leveling	$\pm 3^\circ$ standard $\pm 10^\circ$ optional
CMG-40T-B	Guralp	60 s (standard)	XYZ	Electronic centering	$\pm 8^\circ$
CMG-Flute	Guralp	60 s (standard)	XYZ	Mainspring adjustment	$\pm 12^\circ$
STS-5A	Streckeisen	120 s	Symmetric triaxial	Gimbal	$\pm 5^\circ$

gimbal leveling as it does not compromise the quality of the data while maintaining orthogonality. The next most suitable method is horizontal axis leveling. Although the data will be neither as accurate nor as orthogonal, there are cost benefits to such an approach. Mainspring adjustment and electronic centering are best reserved for higher-noise, lower-cost seismometers.

For the leveling range, the user needs to ensure that the seismometer has sufficient leveling range to cover their deployment conditions. Insisting on a larger leveling range than needed unnecessarily limits the choice of seismometer. Most downhole applications do not require large leveling ranges.

The type of seismometer, namely, whether it is a borehole or posthole type, will be determined by the type of installation (see section “[Types of Installations](#)”). Borehole seismometers tend to cost more than posthole seismometers but can be deployed in deeper holes.

The user needs to ensure that the seismometer will fit into the posthole or borehole that has been drilled. Deeper holes tend to have smaller diameters to reduce the cost per meter of drilling; hence a narrower diameter instrument may be required.

Lastly, a user needs to consider the reliability and consistency of a seismometer. Some seismometers are more reliable than other

seismometers due to factors in the design and the quality of the manufacturing. The consistency of the seismometer refers to the spread in the self-noise floor over time and from instrument to instrument. This can best be assessed by huddle testing at least three instruments over a period of time (Sleeman et al. 2006). There are a number of published papers on self-noise floor of seismometers (Ringler and Hutt 2010).

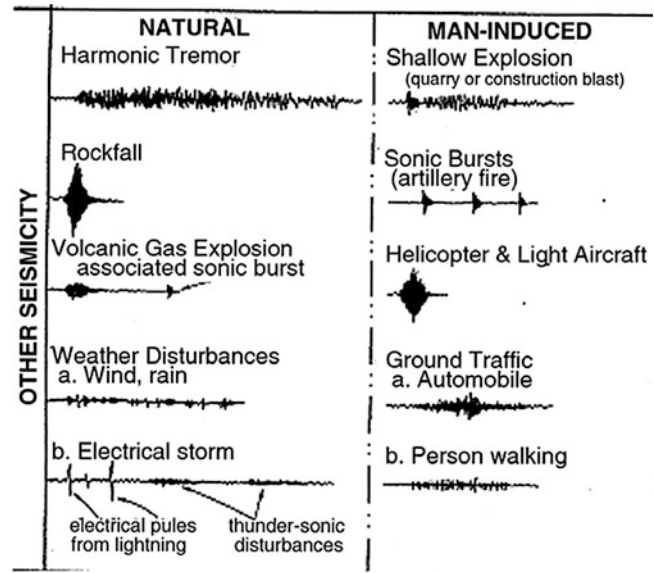
Table 2 documents some commercially available downhole broadband seismometers to be considered in a selection process.

See company websites for further information.

Downhole Seismic Noise Environment

One of the goals of a seismic station installation is to minimize seismic noise and maximize the level of a seismic signal. In other words, the goal is to maximize the signal-to-noise ratio (SNR) that can be achieved at a seismic station. A higher SNR ratio will enable a station to detect smaller-magnitude seismic events. This goal applies equally to surface vaults and downhole seismic installations, but the means to accomplish this goal are different for vaults and downhole installations due to the characteristics of the seismic noise, the seismic signal, and the local geology.

Downhole Seismometers,
Fig. 3 USGS graphic of seismic noise types



Seismic noise can be broken down into different types or sources of noise. These include wind-induced seismic noise, cultural seismic noise (noise generated by human activities), microseismic noise, moving-water-induced seismic noise, and tilt-induced noise (Fig. 3). Each of these noise sources will be briefly described.

Microseismic Noise

Microseismic noise is caused by wave actions on large bodies of water such oceans, seas, and lakes. As waves enter the continental margins and crash onto shorelines, energy is transferred from the water into the earth, generating Rayleigh waves. Microseismic noise ranges from 1 Hz to 30 s, with a peak at 4–8 s known as the microseismic peak. Microseismic noise is a global phenomenon, but has higher amplitudes closer to coastlines and shorelines.

Tilt Noise

Inertial sensors, such as seismometers, measure accelerations due to the motion of a seismic mass. However, inertial sensors are also sensitive to static accelerations such as gravity. Normally, accelerations due to gravity are counterbalanced by the mainspring in the suspension of a seismometer, but when a seismometer is tilted (rotated around a horizontal axis), gravity

appears as an apparent horizontal acceleration. The resulting accelerations due to tilt are indistinguishable from accelerations due to seismic motion (Rodgers 1968; Graizer 2005). This effect is relatively minor on the vertical axis of a seismometer as long as this axis is substantially vertical. On the horizontal axes, the effect is significant and is described as

$$H = g_0 \cdot \sin \theta$$

Since the angles are small, the approximation $\sin \theta = \theta$ can be used:

$$H = g_0 \cdot \theta$$

where

- $g_0 = 9.8 \text{ m/s}^2$ is the acceleration due to gravity,
- H is the horizontal acceleration in m/s^2 , and
- θ is the tilt in radians.

Ground tilts can be caused by human activities, wind, or changes in local barometric pressure, temperature, or soil moisture content. The magnitude of ground tilts is influenced by the stiffness of the ground. Hard-rock sites will be much less susceptible to tilt as compared to a site

on clay. Tilt noise causes horizontal components to be typically 20–50 dB noisier than the vertical components at long periods (Bormann 2002).

Tilt noise tends to decrease with depth as many sources of tilt noise are surface effects, namely, wind, air pressure, soil moisture, and human activity. Tilt noise can also be caused by air currents circulating around a seismometer. This effect can be significant in borehole installations if strong vertical convection air currents are allowed to flow. Precautions must be taken to prevent the air moving. This problem can be mitigated with a sand installation, which eliminates air spaces around a seismometer, or with insulation methods, such as foam plugs, to limit the airflow below and above a seismometer.

Wind-Induced Noise

Wind is another major source of seismic noise. Wind transfers energy into the ground as it decelerates from colliding with the terrain, water bodies, buildings, vegetation, and other objects coupled to the ground. The chaotic nature of wind often excites resonances in objects it strikes. Trees and buildings sway in the wind, waves are generated on water, and the ground is compressed. Although the wind blows predominantly in a horizontal direction, interactions with the surface translate the motion into the vertical component. For example, as trees sway back and forth in the wind, alternating pressure is placed on opposite sides of the root bundle imparting vertical motion. Wind-induced seismic noise is short period in nature, meaning that is above 1 Hz. Wind noise is dependent on the wind velocity. Higher velocities generate more noise. For reference, a wind speed below 3 m/s (11 km/h) does not raise the background seismic noise at the surface for grassland sites on sedimentary rock in rolling hills (Withers et al. 1996). Wind induces predominantly surface waves (Love and Rayleigh waves) along with a small component of body waves onto the ground.

Moving-Water-Induced Noise

Seismic noise is generated in watercourses from small streams to large rivers. As water flows downhill, turbulence is created which imparts

energy into the ground. Regions with steeper gradients, such as waterfalls and rapids, generate more seismic noise. Moving-water seismic noise is continuous but varies in amplitude depending on water levels. Moving water induces surface waves similar to wind noise.

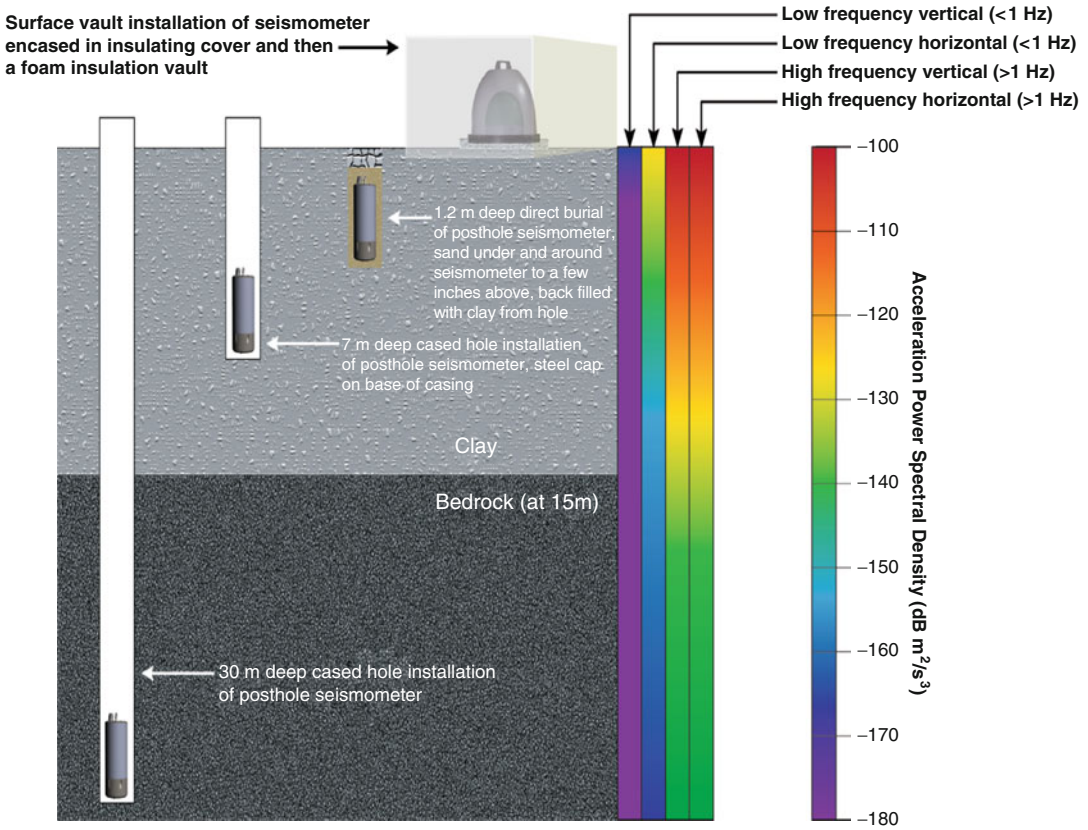
Cultural Noise

Seismic noise generated by human activities is referred to as cultural noise. The principal sources of cultural noise are related to transportation, construction, military, and mineral extraction activities. However, any human activity can generate cultural noise, even a person walking down a path. Cities have higher levels of cultural noise than rural areas. Cultural noise can be stationary or nonstationary, periodic or random, and generated by sources distributed or localized. Cultural noise does tend to be diurnal, meaning it varies over the day with quietest times in the middle of the night when there is the least human activity. Like wind-induced noise, cultural seismic noise is predominantly composed of high-frequency surface waves but also includes a body wave component.

Seismic Noise Mitigation with Depth

The common factor in wind-induced, cultural, moving-water, and microseismic noise is that the noise is predominately generating surface waves. At short periods, the amplitude of surface waves decays exponentially with depth due to amplitude drop from the surface and due to the upper layers of the earth being highly attenuating or “lossy.” It is well known that high-frequency seismic noise decreases with depth making borehole sites typically quieter than surface vault sites from 1 to 100 Hz (Carter et al. 1991). At long periods, the seismic noise also drops with depth, particularly on the horizontal channels due to a reduction in the tilt noise (Sorrells 1971).

While seismic noise attenuates with depth, seismic signal amplitudes are dependent on a number of factors relating to the geology of the site. As a seismic signal propagates toward the surface, it typically travels through progressively lower Q layers and loses energy. The layers at the surface have the lowest Qs and are responsible for



Downhole Seismometers, Fig. 4 Vault installation and posthole installations at various depths as measured in Ottawa, On, Canada (Nanometrics Inc 2013)

a significant portion of the loss in energy (Aster and Shearer 1991). However, as the velocity of the signal slows, the amplitude increases, and there is also surface amplification effect. In summary, it is difficult to estimate the amplitude of the signal with depth.

Nevertheless, the effect of seismic noise attenuation with depth results in improved signal-to-noise ratios especially when the final layers of earth are highly lossy (Young et al. 1996). Ideally, a downhole seismometer should be placed in hard rock below a highly lossy overburden.

The question that remains is, how deep should the downhole seismometer be placed? The answer is as deep as economically and practically possible as signal-to-noise ratios typically improve with depth in the first few hundred meters. However, the question can be turned around to: how shallow can a downhole

seismometer be placed to result in an improvement in the signal-to-noise ratio? The answer is within 3–10 m of the surface depending on the geology due to the attenuation of temperature, tilt noise, and surface noise.

Consider an extreme example of 15 m clay above quartzite bedrock in an urban environment shown in Fig. 4. This was measured in Ottawa, Ontario, Canada, using Trillium120 Posthole seismometers (Nanometrics Inc 2013). Clay has a very low Q , and the quartzite has a much higher Q . Unsurprisingly, the surface vault has the highest noise at low and high frequencies. The 1.2 m direct burial posthole seismometer installation shows improved low-frequency vertical and horizontal noise and improved high-frequency noise. At 7 m, the posthole seismometer installation shows further improvements in low-frequency horizontal noise and improved high-frequency noise. At 30 m from

the surface in bedrock, the horizontal low-frequency noise has dropped 30 dB, the low-frequency vertical noise has improved 10 dB, and the high-frequency noise on the vertical and horizontal channels has dropped 40 dB as compared to the surface vault. This example shows that improvements in noise levels can be achieved in relatively shallow downhole installations. Deeper installations could possibly yield further improvements, but in this example an improvement was achieved in the first 15 m. It should be noted that these results are specific for this site. In general, noise levels will be site specific.

In summary, even shallow posthole installations can provide SNR improvements over surface vaults.

Downhole Seismometer Installations

Types of Installations

Downhole seismometers can be deployed into boreholes or shallow holes.

Boreholes are steel-cased holes that are cemented into competent rock. Typical seismic boreholes are 50–150 m deep, but there are seismic boreholes 2,000 m deep and beyond (Young et al. 1994). Although a cemented steel casing extends to the surface, a borehole poorly propagates surface noise down the borehole. There are two reasons for this. First, the coupling of the casing to the upper softer layers of material is not as good as the coupling to stiffer materials lower down. Second, the borehole is relatively compliant horizontally over the length of the borehole.

These holes can be drilled using water-well or oil-well drilling equipment. The steel casings are cemented into the borehole to stabilize the casing, provide good coupling to the ground, and maintain the integrity of underground aquifers.

Shallow holes are typically holes 0.5–5 m deep that are dug by shovel or lightweight drilling equipment such as augers, small drill rigs, screw-pile equipment, and the like (Fig. 5). These holes are typically larger in diameter as compared to boreholes to enable a wider range of seismometers to be deployed and to improve access to the seismometer. At permanent sites, the hole may be

cased in steel or pvc and cemented in. Shallow holes drilled into hard rock do not need to be cased. Seismometers are typically deployed into shallow holes using a sand installation method.

Shallow holes that are used for temporary deployments are typically not cased. This minimizes the cost of the installation. A seismometer deployed in an uncased hole is backfilled with sand or the excavated materials from the hole. This type of deployment is also known as a direct burial installation. A direct burial installation is quicker and easier to install than an installation using a vault seismometer. It is often challenging leveling a vault seismometer in a field deployment since it requires a hard surface, access to the leveling feet, and bubble level. With a posthole seismometer, the procedure is to dig a small hole, place the seismometer in the hole, backfill with earth or sand, and start the autoleveling in the seismometer. There is no further work required to insulate or weatherproof the installation.

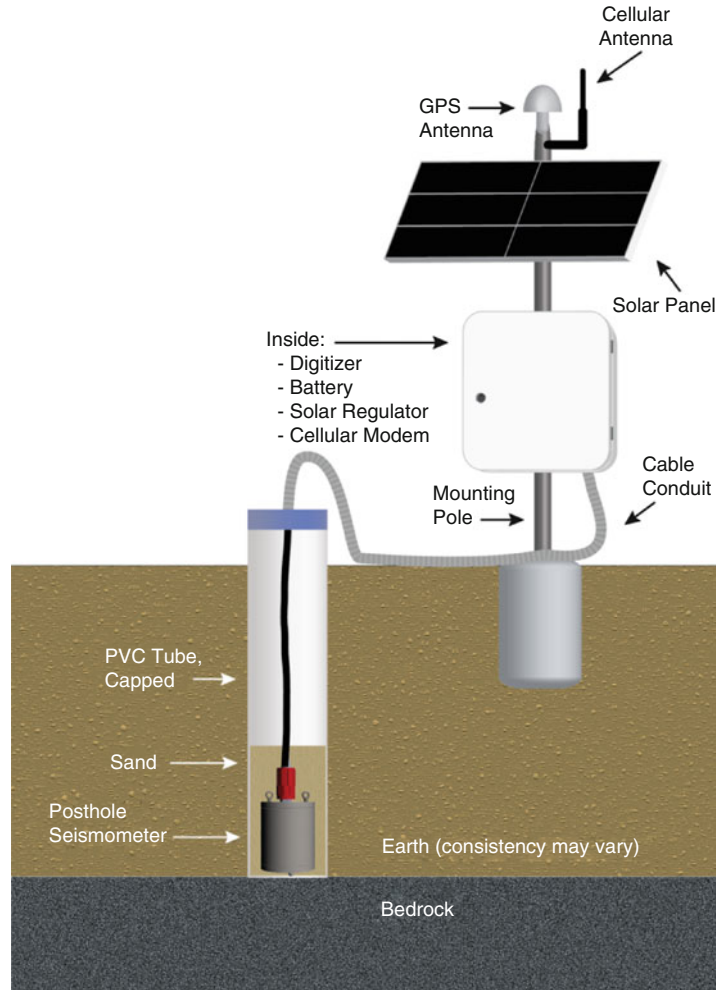
Seismometer Emplacement

Once a seismometer is lowered into a borehole, it must be secured rigidly in the hole to ensure good seismic coupling to the borehole casing and the ground. There are two methods for accomplishing this. The first is with a mechanical borehole lock and the second is with a sand installation.

A borehole lock is a device attached to the seismometer which clamps the seismometer to the side of the borehole. Typically, a borehole lock has three fixed pins on one side of the holelock which create a three-point contact with the borehole and one moving pin on the opposite side which clamps the holelock to the side of the borehole. The moving pin pushes the holelock three-point contact pins against the borehole with a force sufficient to hold the seismometer in place. The seismometer is then supported by the four pins. This is a kinematic coupling. The seismometer must not be touching the bottom of the borehole; otherwise the coupling will be over-constrained which can result in undesired motions. With this method of locking, the seismometer takes on the vertical angle of the borehole at that point. The moving pin on the holelock

Downhole Seismometers,

Fig. 5 Components of a typical posthole station with the sensor 1–6 m below the surface



can be actuated with an electric motor or a mechanical system controlled from the surface.

The sand installation method involves lowering the seismometer to the bottom of a borehole or shallow hole and pouring dry sand into the hole sufficient to cover the seismometer in sand. The sand flows around the entire seismometer filling the entire gap between the seismometer and the hole. This method of installation has a few benefits and drawbacks. A sand installation is simpler and less costly than a holelock installation. It eliminates convection air currents around the seismometer and effectively couples the seismometer to the casing. However, sand is a loose

material that is prone to spontaneous movements that can be detected by a seismometer during a settling period. Another issue with sand installations is that after a period of time, moisture and dissolved minerals in the sand can solidify the sand making it challenging to remove the seismometer.

Economics of Installations

One of the primary considerations in deploying a downhole seismic station is the cost of the installation, especially in comparison to vault installations. In the past, downhole installations have been substantially more costly than vault

installations because downhole installations were predominately 100–200 m-deep borehole installations with high-cost, top-performance borehole seismometers. A typical borehole was drilled using oil-well drilling equipment which costs approximately US\$100,000 per borehole. A borehole station of this kind would have had a total installed cost of US\$150,000–300,000 depending on the location and ancillary equipment included. This limited borehole installations to those who had budgets large enough to cover cost of a borehole network and had a requirement for excellent signal-to-noise ratios.

Recently, newer instruments and installation methods have dramatically lowered the cost of a downhole installation making downhole stations cost competitive with vault installations.

One of the main innovations is switching away from using oil-well drilling equipment and materials. Oil-well drilling equipment is designed and sized to drill deep wells (up to many kilometers) in complex geologies. This equipment typically includes steerable drill bits to enable horizontal drilling. For a typical vertical seismic borehole of 50–150 m deep, this equipment is expensive to use and in most geologies unnecessary.

Water-well drilling rigs are more suitable and substantially cheaper to operate as compared to oil-well drilling equipment. Water-well drilling equipment is designed for drilling only vertical boreholes 10–300 m deep reducing rig size, complexity, and cost. Water-well drilling services are typically available in most locations reducing deployment costs and are competitively priced due to the demand for water wells. A typical 60 m water well can be drilled in North America for \$US 5,000–25,000.

A seismic borehole includes costs for casing, cementing, and pressure testing, in addition to the drilling costs. Casing and cementing costs will vary with depth, casing type, and diameter. Local drilling contractors who are often very familiar with the geology in the region can readily provide up-to-date prices and guidance on drilling.

For posthole installations, cased boreholes, screw piles, and uncased shallow holes can be

used. Uncased shallow holes are the least expensive and can be dug with hand or powered augers. Screw-pile installations have costs between boreholes and shallow holes. Screw piles are installed with skid-steers and have the added benefit of being removable at the end of a deployment. Local construction contractors can provide current prices for comparison.

Security of Installations

Boreholes and shallow holes are in general more secure than vault installations for the seismometers. The overall footprint of a downhole installation is much smaller and less visible than a typical permanent vault installation and can be camouflaged more easily. If desired, the entire borehole wellhead can be buried. Buried underground conduits can be routed from the borehole or shallow hole to the communications and power hub. The curious would assume that a visible wellhead is that of a water well which is uninteresting. Often winches and tripods (or masts) are required to extract seismometers from the hole providing a level of physical security that is hard to overwhelm even for the most determined thief.

Nevertheless, a downhole seismometer installation only protects half the equipment at the station. The communications and power equipment is still highly visible and attractive to a thief. Solar panels and backup batteries are popular and useful everywhere in the world and are routinely stolen from less-secure sites (see Fig. 6). A solution to station security must cover all the equipment at a site (Fig. 7).

Typically, there is a trade-off to be made between a rural site which has low cultural seismic noise but is less secure due to its isolation and a secure urban site, such as police or fire station, which has higher cultural seismic noise. Ideally, an operator wants a low cultural noise site that is also secure. With downhole installations, a secure urban site enables the reduction of seismic noise by providing vertical separation from seismic noise which originates at the surface. The degree of attenuation of seismic noise with depth will depend

Downhole

Seismometers,

Fig. 6 Rural posthole installation with cellular communications in Alberta, Canada



Downhole

Seismometers,

Fig. 7 Rural vault installation with satellite communications in Fiji



on the local geology and nature of the cultural noise. An urban borehole or posthole may not be as quiet as a good rural site, but it is secure and can be quickly serviced if there is a problem.

A network operator will have to trade security and ease of access for uptime and data quality based on the local geology, cultural seismic noise, economic conditions, cultural norms regarding property, and cost.

Station Footprint

Downhole installations typically have a smaller footprint as compared to vault installations. The land area required for a borehole or shallow hole

can be as little as the area of the hole. Typically, a borehole or shallow-hole installation will require a $0.25\text{--}1\text{ m}^2$ of space (Fig. 9). Seismic vaults typically use much more space, particularly if the vault is a walk-in type of vault.

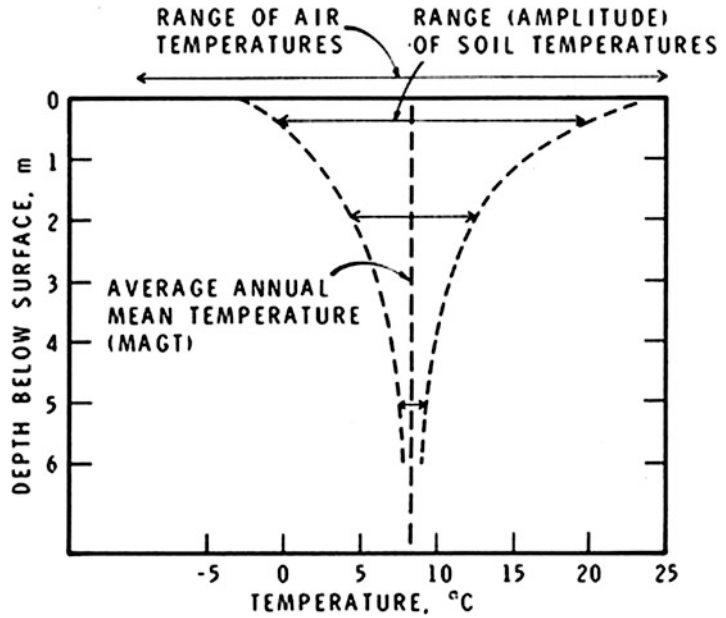
The smaller footprint of a downhole site offers more flexibility in station placement, especially in an urban environment where available land is limited and costly and local building codes restrictive.

Site Environmental Conditions

The site environmental conditions, namely, weather, water, and wildlife, have a large impact

Downhole Seismometers,

Fig. 8 Ground temperature variations with depth for Ottawa reproduced from Williams and Gold (1976)



on the performance of seismic stations. While great care must be taken with the design of vault installations to mitigate the effects of temperature, wind, air pressure, water egress and flooding, and wildlife, downhole installations are much less susceptible to these effects and require less design effort.

Temperature

Seismometers are sensitive to temperature changes, and ideally, a seismometer should be exposed to as little temperature variation as is economic. At the surface, the temperature can range from -20°C in the winter to $+35^{\circ}\text{C}$ in the summer in a northern continental climate, with daily fluctuations between 10°C and 20°C . Other climates can be more or less extreme. A vault installation uses thermal insulation and structures with high thermal mass to reduce the temperature variation of the seismometer. This requires design work and cost.

A downhole installation relies on the heat capacity and thermal conductivity properties of the ground to reduce the temperature fluctuations to a minimum. The depth where the temperature fluctuation is 1/100 of the surface temperature variation for a given time period is known as the

penetration depth and depends on the composition of the soils and the moisture content of the soils. For daily temperature variations, the penetration depth is less than 1.2 m for all soil types. The yearly penetration depths vary from 5 to 20 m depending on the soil conditions. For most downhole seismometer installations, an annual temperature variation of $\pm 3^{\circ}$ is acceptable and can be achieved in as little as 3–6 m below-ground. See Fig. 8; ground temperature variations with depth for Ottawa, reproduced from Williams and Gold (1976).

Water Egress

Seismic vaults are usually located at or just below the surface on undisturbed soils and are subject to water egress and flooding from time to time. While vaults are often designed to be waterproof, there is usually some weakness in the implementation that enables water to enter the vault over time or during flood situations. Some vaults are equipped with pumps, but over time, pumps fail. This would not be an issue except that vault seismometers and ancillary equipment are not designed for continuous submersion. Flooding causes the temporary shutdown of many vault seismic stations.



Downhole Seismometers, Fig. 9 Rural borehole seismometer deployment in Alberta, Canada

Downhole seismometers are designed for continuous submersion and are unaffected by periodic flooding. Generally, boreholes and postholes are designed to be dry, but it is not strictly necessary for the equipment. Shallow hole can be dry or wet, or even below the water table. Water may degrade the seismic performance due to movement of the water, but it will not cause the total loss of data from a station.

Wildlife

Animals from time to time cause problems at seismic stations. This can range from large animals walking in the vicinity of the station or scratching themselves against the station down to mice building nests in the insulation surrounding a seismometer or chewing cables and every critter and its activities in between. Being at the surface, vault installations are simply more accessible and susceptible to these invasions. Downhole Installations are away from the surface and difficult to access

even for the smallest animals. Surface noise from animals is attenuated with depth to some degree.

Deployment Considerations

Borehole Verticality

Boreholes are typically not vertical holes but have small tilt angles of up to 10° from vertical. This is due to the accuracy of the driller's equipment and technique and the nature of the material being drilled. Drilling practices that affect the verticality of a borehole include the weight on the drill bit, the speed of drilling, the degree of hole cleaning, drill bit selection, whether a steerable drill bit is used, and the use of verticality monitoring equipment while drilling. Experienced drillers understand the trade-offs in maintaining a vertical borehole. In maintaining a vertical borehole, a driller will want to understand the geology of the area prior to drilling. It is harder for a driller to maintain verticality with complex borehole profiles. For example, it is more challenging to maintain verticality when drilling soft sediments such as clay or sand or in sediments with materials with differing hardness such as glacial till or limestone with cavities. With glacial till, the drill bit can be deflected by rocks and boulders in the sand, and with limestone, the drill bit can drift in the cavities. Other challenging borehole profiles include highly fractured rock and sedimentary rocks with steep dip angles.

In most situations, drillers can maintain the tilt to less than 3° off vertical. Typically, drilling specifications stipulate that the borehole shall be within 3° off vertical or less. USGS suggests a drilling specification of 2° or less (Albuquerque Seismological Laboratory 2001). Even with a vertical error of 3° , the vertical offset or drift at the bottom of the 100 m borehole is just over 5 m. The drift at the bottom of a borehole may be a concern in some installations that have underground features that must be avoided.

Another issue related to borehole verticality is the tilt of the seismometer in the borehole. It is not possible to maintain the verticality of a

seismometer as it is lowered; a seismometer will always be resting against the side of the borehole as it goes down and will tend to assume the angle of the borehole when it reaches the bottom. For example, if the seismometer is lowered to 50 m in the borehole, and the borehole at 50 m is 2° off vertical, the seismometer will tend to be 2° off vertical when it comes to rest. See section “[Levelling](#)” for how to compensate for the vertical error.

Aligning to North

Seismometers are typically aligned to north. Vault seismometers are aligned to north manually using a compass. A compass bearing is taken, a north–south line is scribed on the vault floor, and the seismometer is manually rotated to align with the north–south line. There are many variations of this method, but overall the process is straightforward.

Aligning downhole seismometers to north is more challenging. The user has only limited access to the instrument downhole, and a compass cannot be used downhole due to the magnetic shielding by the steel borehole casing. Nevertheless, methods have been developed to align downhole seismometers.

The downhole alignment process can be broken down into two steps. The first step is to determine the seismometer orientation relative to north. This can be accomplished using a number of different methods including downhole gyroscopic alignment, surface seismometer referencing, rifle scope alignment, and north-finder alignment. See below:

- With gyroscopic alignment, a gyroscope is activated and aligned to north at the surface. It is then lowered down the borehole and engages with a north-aligned feature on the seismometer or holelock. A reading is taken from the gyroscope to determine the rotational offset from north.
- Surface seismometer alignment involves collecting a seismic data set from the downhole seismometer and a surface seismometer located at the wellhead. The downhole data set is

correlated with the surface data set to determine the offset from north.

- Rifle scope alignment involves mounting a rifle scope vertically pointing down the borehole. The rifle scope cross hair is focused on the north–south marking on the seismometer downhole. A reading of the rifle scope rotation relative to north is taken.
- North-finder alignment uses a north-finder instrument which measures earth’s rotation vector to determine north. The north finder can be mounted on the seismometer or lowered down the borehole to engage the seismometer.

The second step is to rotate the seismometer to be north aligned. This can be done digitally, mechanically with a “bishop’s hat” mechanism, or manually with an alignment rod.

- Digital alignment involves mathematically rotating the digital data set to align to north. This can be done with no loss of precision. The user enters a rotation parameter into the seismometer or digitizer configuration, and the instrument subsequently rotates all time-series data.
- Bishop’s hat mechanical alignment is a multistep process. A holelock with a bishop’s hat mechanism is lowered into the borehole and clamped in place. A north-measuring device is lowered onto the bishop’s hat mechanism to determine the orientation of the holelock. The north-measuring device is removed, and the seismometer bishop’s hat mate is rotated and locked at the surface. The seismometer is lowered down onto the bishop’s hat and is now aligned to north. A bishop’s hat mechanism is a device that forces the mating assembly (the seismometer) to a particular orientation. It is called a bishop’s hat because it is shaped like a bishop’s hat.
- Lastly, a seismometer can be aligned with an alignment rod. The alignment rod is attached to the seismometer and the seismometer is lowered into the hole. From the surface, the user rotates the alignment rod and the

seismometer together. Once the seismometer is aligned, the alignment rod is removed from the hole. An alignment rod is limited to use in postholes and shallow holes.

Summary

In the past, downhole seismic stations were limited to deep borehole installations which were so expensive that they only justified the deployment of very costly borehole seismometers. These stations demonstrated the reduction of seismic noise and the improvement of signal-to-noise ratios with depth. However, the cost of ownership of these stations limited their widespread adoption.

In the last few years, posthole and direct burial seismometers have been introduced, a better understanding of seismic noise and its dependence with depth has been developed, and more economic installation methods, such as those using water-well drilling equipment, have been established. These instruments and methods have made downhole seismometer installations cost competitive with vault installations, while at the same time improving on the signal-to-noise ratios. In some cases, downhole installations can cost much less than vault installations of similar performance.

Shallow-hole direct burial deployments are an improvement over temporary surface deployments in many aspects. These include improved signal-to-noise ratios, better site security, simpler installation and servicing, smaller footprint, and improved station reliability.

With new borehole and shallow-hole deployment methods feasible, network operators can now balance more finely the cost of a downhole station with the quality of the seismic data enabling operators to collect better data for less money.

Cross-References

- [Principles of Broadband Seismometry](#)
- [Symmetric Triaxial Seismometers](#)

References

- Albuquerque Seismological Laboratory (2001) Specifications for seismic instrumentation boreholes. Retrieved 1 Dec 2013, from U.S. Geological Survey: http://earthquake.usgs.gov/regional/asl/pubs/files/tech_boreholes.pdf
- Aster R, Shearer P (1991) High-frequency borehole seismograms recorded in the San Jacinto Fault zone, Southern California Part 2. Attenuation and site effects. *Bull Seismol Soc Am* 81(4):1081–1100
- Bormann P (2002) Seismic signals and noise, Chap 4. In: Bormann P, Bormann P (eds) *New manual of seismological observatory practice*, vol 1. GeoForschungsZentrum, Potsdam
- Carter J, Barstow N, Pomeroy P, Chael E, Leahy P (1991) High-frequency seismic noise as a function of depth. *Bull Seismol Soc Am* 81(4):1101–1114
- Graizer V (2005) Effect of tilt on strong motion data processing. *Soil Dyn Earthq Eng* 25(3):197–204
- Graizer V (2009) The response to complex ground motions of seismometers with Galperin sensor configuration. *Bull Seismol Soc Am* 99(2B):1366–1377
- Nanometrics Inc (2013) Comparison study between vault seismometers and a new posthole seismometer. Retrieved from www.nanometrics.ca: http://www.nanometrics.ca/ckfinder/userfiles/files/VaultvsPostholePaper.pdf
- Peterson J (1993) Observations and modeling of seismic background noise. Open-file report 93-322. U. S. Geological Survey, Albuquerque
- Ringler AT, Hutt CR (2010) Self-noise models of seismic instruments. *Seismol Res Lett* 81(6):972–983
- Rodgers P (1968) The response of the horizontal pendulum seismometer to Rayleigh and Love waves, tilt, and free oscillations of the Earth. *Bull Seismol Soc Am* 58(5):1385–1406
- Sleeman R, van Wettum A, Trampert J (2006) Three-channel correlation analysis: a new technique to measure instrumental noise of digitizers and seismic sensors. *Bull Seismol Soc Am* 96(1):258–271
- Sorrells GG (1971) A preliminary investigation into the relationship between long-period seismic noise and local fluctuations in the atmospheric pressure field. *Geophys J Roy Astron Soc* 26:71–82
- Williams G, Gold L (1976) Ground temperatures. *Can Build Dig*, CBD-180
- Withers M, Aster R, Young C, Chael E (1996) High-frequency analysis of seismic background noise as a function of wind speed and shallow depth. *Bull Seismol Soc Am* 86(5):1507–1515
- Young C, Chael E, Zagar D, Carter J (1994) Variations in noise and signal levels in a pair of deep boreholes near Amarillo, Texas. *Bull Seismol Soc Am* 84(5):1593–1607
- Young C, Chael E, Withers M, Aster R (1996) A comparison of the high-frequency (>1 Hz) surface and subsurface noise environment at three sites in the United States. *Bull Seismol Soc Am* 86(5):1516–1528

Drift Issues of Tall Buildings in 2011 Tohoku Earthquake

Mehmet Çelebi¹, Izuru Okawa² and Toshihide Kashima²

¹Earthquake Science Center, US Geological Survey, Menlo Park, CA, USA

²Building Research Institute, Tsukuba, Japan

Synonyms

Drift; Earthquake records; Long-distance effects; Long-period structure; Tohoku earthquake

Introduction

One of the most significant effects of the M9.0 Tohoku, Japan, earthquake of March 11, 2011, is the now well-known long-duration (>10 min) shaking of buildings in Japan – particularly those in Tokyo (~350–375 km from the epicenter) and in places as far as Osaka (~770 km from the epicenter). Although none collapsed, the strong shaking caused many tall buildings not to be functional for days and weeks.

The purpose of this entry is to discuss the drift ratio issues related to tall buildings in Japan affected by long-period ground motions at far distances from the epicenter of the Tohoku earthquake of March 11, 2011 ($M = 9.0$). We review the inferred long-period long-distance effects of the ground motions from recorded response data from four tall buildings in Osaka and Tokyo and the spectral properties of ground motions recorded from stations nearby these buildings.

Therefore, three other relevant topics are also discussed in this entry:

- (a) Past observations of long-distance long-period effects and implications for the United States
- (b) Long-period effects as demonstrated by response spectra of ground motions recorded during the same event
- (c) Maximum allowable drift ratios in Japan and other seismic areas of the world

The performances of buildings discussed in this entry are inferred from computations of average drift ratios because the accelerometer deployed in the tall buildings is sparse. It is important to note that drift ratios are the best measure to infer damage occurrence or likelihood during design/analysis processes and also during analyses of recorded response data if they exist (Çelebi et al. 2004; Çelebi 2008). The effect of long-period ground motions to buildings and other long-period structures is demonstrated by the four cases presented herein.

All of the four examples indicate that considerably higher drift ratios for relatively low (~3 % g) to moderate (10–20 % g) input ground motions were affecting the functionality of tall buildings at large distances from the epicenter of the Tohoku event. Published documents to date on performances of tall buildings during the Tohoku event in general do not describe actual observed or unobserved possible hidden damages. While no collapses were reported, there is no certainty that hidden damages in some of the buildings do not exist.

Past Observations of Long-Distance Long-Period Effects and Implications for the United States

One of the earliest observations in the United States was during the $M = 7.3$ Kern County earthquake of July 7, 1952 that shook many taller buildings in Los Angeles and vicinity, about 100–150 km away from the epicenter (http://earthquake.usgs.gov/earthquakes/states/events/1952_07_21.php, last accessed July 15, 2011; Hodgson 1964). The March 28, 1970, $M = 7.1$ Gediz earthquake in inland western Turkey damaged several buildings at a car-manufacturing factory in Bursa, 135 km northwest from the epicenter (Tezcan and Ipek 1973). One of the most dramatic examples of long-distance effects of earthquakes is from the September 19, 1985, Michoacan, Mexico, $M 8.0$ earthquake during which, at approximately 400 km from the coastal epicenter, Mexico City suffered more destruction and fatalities than the epicentral area due to

amplification and resonance (mostly around 2 s) of the lakebed areas of Mexico City (Anderson et al. 1986; Çelebi et al. 1987). To the best knowledge of the authors, however, there are no publicly available records of the responses of tall structures from these past earthquakes. Therefore, records obtained from numerous instrumented tall buildings during the Great East Japan earthquake of March 11, 2011 offer a rare opportunity to study and understand how structures characterized by predominantly long-period responses behave during medium to large events originating at long distances. Such effects have consequences for large metropolitan areas in Japan but also in other parts of the world, including the United States (e.g., Los Angeles area from Southern California earthquakes, Chicago from NMSZ, and the Seattle (WA) area from large Cascadia subduction zone earthquakes). For example, the recent $M = 5.8$ Virginia earthquake of August 23, 2011, was felt in 21 states of the Eastern and Central United States that include large cities such as New York and Chicago (<http://earthquake.usgs.gov/earthquakes/eqinthenews/2011/se082311a/#summary>, last accessed September 11, 2014). During this event, occupants of a 28-story building at MIT Campus in Cambridge, MA, experienced the shaking (Toksoz, 2011, “personal communication”).

Discussion on Long-Period and Long-Distance Effects During Tohoku Earthquake and Response Spectra Issues

Based on a study by Okawa and others (2010), 4 months before the Tohoku event, the Japanese Government Ministry of Land, Infrastructure, Transport and Tourism [MLIT] released a regulation to upgrade the design of high-rise buildings under long-period ground motions (2010). With this release, MLIT solicited public and professional opinions. However, immediately after the end of this period, the Tohoku earthquake occurred. Following the Tohoku event, MLIT started to revise the document with collected opinions and new data from the earthquake.

In the MLIT (2010a) document, a total of 9 areas in Japan were specified (4 in Tokyo, 3 in Nagoya, and 2 in Osaka). For each one of these areas, simulated acceleration and velocity time-histories were specified. Takewaki and others (2011) report that velocity response spectra of all the simulated ground motions have significant large peaks within 2–8 s range. It is important to state that Takewaki and others (2011) report that in Shinjuku area of Tokyo, where tall buildings are discussed later in this entry, the KNET Shinjuku station (TKY007) motions exhibit predominant peaks between 3 and 8 s range.

In this entry, we present acceleration and velocity response spectra of ground motions recorded during the Tohoku earthquake from two stations. As an example, for Osaka Bay, spectra of motions from KIKNET station OSKH02 are provided in Fig. 1 and display the long periods both at surface and downhole. This station is closest to the example building in Osaka Bay discussed also later in the entry. In Fig. 2, we present spectra of motions from KNET stations TKY007 (representative of motions in the Shinjuku area of Tokyo, as discussed above). These spectra clearly display the long periods capable of affecting responses of the tall buildings. TKY007 was also examined by Takewaki and others (2012).

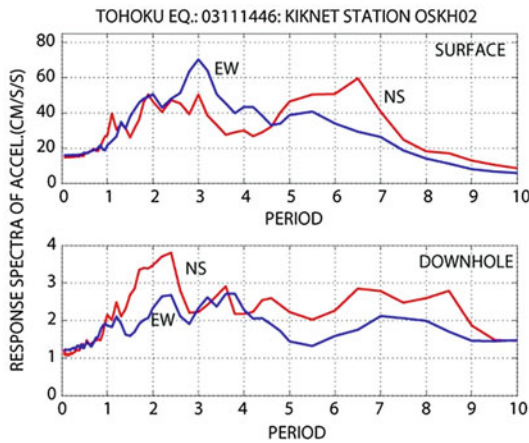
Drift Ratios of Different Countries

Before presenting drift issues relative to the four buildings, it is important to review practiced drift ratio limits of a few seismically active countries and also that of Eurocode 8 (2008). Comparison of the drift ratios of five countries and Eurocode 8 is provided in Table 1. The table exhibits drastic variability in practice of drift ratio limits from 0.2 % to 2.5 %.

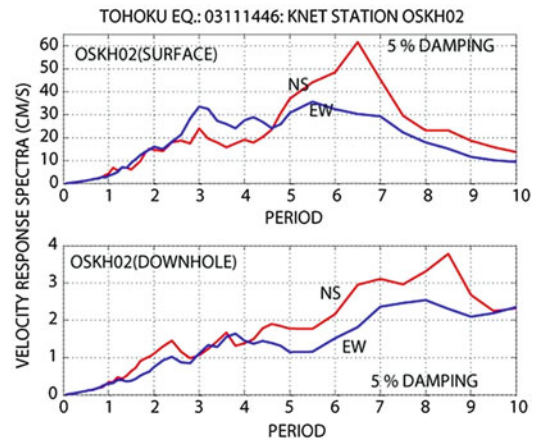
Buildings Studied

Building A at Osaka Bay ~ 770 km from the Epicenter of the Earthquake

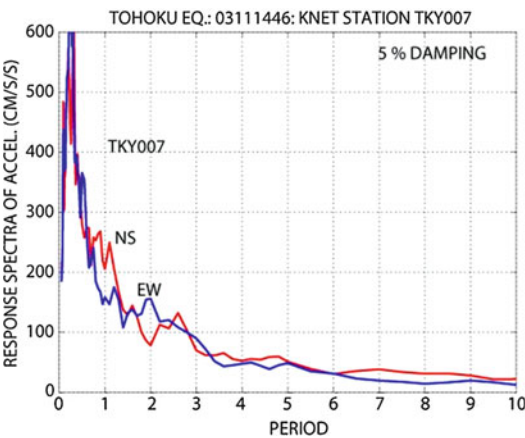
Detailed discussion of the building, data, and analyses is provided in Çelebi et al. (2012).



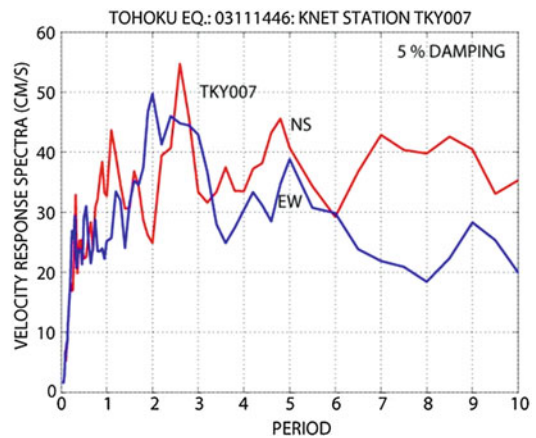
Drift Issues of Tall Buildings in 2011 Tohoku Earthquake, Fig. 1 (Left) NS and EW acceleration response spectra and (right) velocity response spectra of ground motions recorded at surface and downhole during Tohoku



event at KIKNET OSKH02 station in Osaka Bay. Velocity response spectra exhibit the large amplitudes between 5 and 7 s



Drift Issues of Tall Buildings in 2011 Tohoku Earthquake, Fig. 2 (Left) NS and EW acceleration response spectra and (right) velocity response spectra of ground motions recorded during Tohoku event at KNET



TKY007 station in Shinjuku area of Tokyo where tall buildings are concentrated. Velocity response spectra exhibit the large amplitudes particularly for periods >2 s

Figure 3 shows vertical sections and locations of triaxial accelerometers. Figure 4 shows plan views, orientation of the building, and location of accelerometers at the 52nd floor.

The building was subjected to an input motion at ground level with peak accelerations of $\sim 3\%$ g during the Tohoku earthquake. Furthermore, the building was founded at a site with a site frequency (period) of 0.13–0.17 Hz (5.88–7.69 s) as seen in Fig. 5 (left) (Çelebi et al. 2012).

The parameters used in computing the site transfer functions are the (depth vs. Vs) Profiles A, B, and C shown in Fig. 3 (left). Profile A is an approximation based on the geotechnical data for free-field KIK-NET station OSKH02 that is near (~ 2.5 km) the building. In this profile, the upper and softer layers have been ignored. By way of comparison with the transfer functions computed for Profiles B and C, which underlie the building, it is concluded that the upper layers

Drift Issues of Tall Buildings in 2011 Tohoku Earthquake, Table 1 Upper limits of drift ratios (Japan, Chile, the United States, Turkey, New Zealand, and Eurocode 8)

Code of	Upper limit drift ratio (%)	Comment	Reference
Chile	0.2	Results in elastic design	NCh433.Of96 (1996)
Japan	1.0	Max for buildings taller than 60 m. For collapse prevention (level 2) motions	Building Center of Japan 2001a, b
The United States	2.0	No collapse state	ASCE7-10 (2007)
Turkey	2.0	Can be increased by 50 % in case of some steel frames	DBYYHY (2007): Section 2.10.1.3, Eqn: 2.19 in Turkish Code (2007)
New Zealand	2.5		NZS1170.5(2004)
Eurocode 8	1.0 (max)	For buildings having nonstructural elements fixed in a way so as not to interfere with structural deformations or without nonstructural elements	Section 4.4.3.2, Eqn: 4.33 Eurocode 8 (2008)

do not significantly alter the computed fundamental frequency of the site of this building. Q values used in calculating the transfer functions range between 25 and 60 for shear wave velocities between 200 and 600 m/s – having been approximately interpolated to vary linearly within these bounds. As seen in Fig. 5 (left), the site fundamental frequency is computed to be in the range of 0.13–0.17 Hz due to the dominant characteristics of layers three and four (typically of the area of the site of the building and KIKNET OSKH02 strong-motion station as described in Fig. 5 [left]).

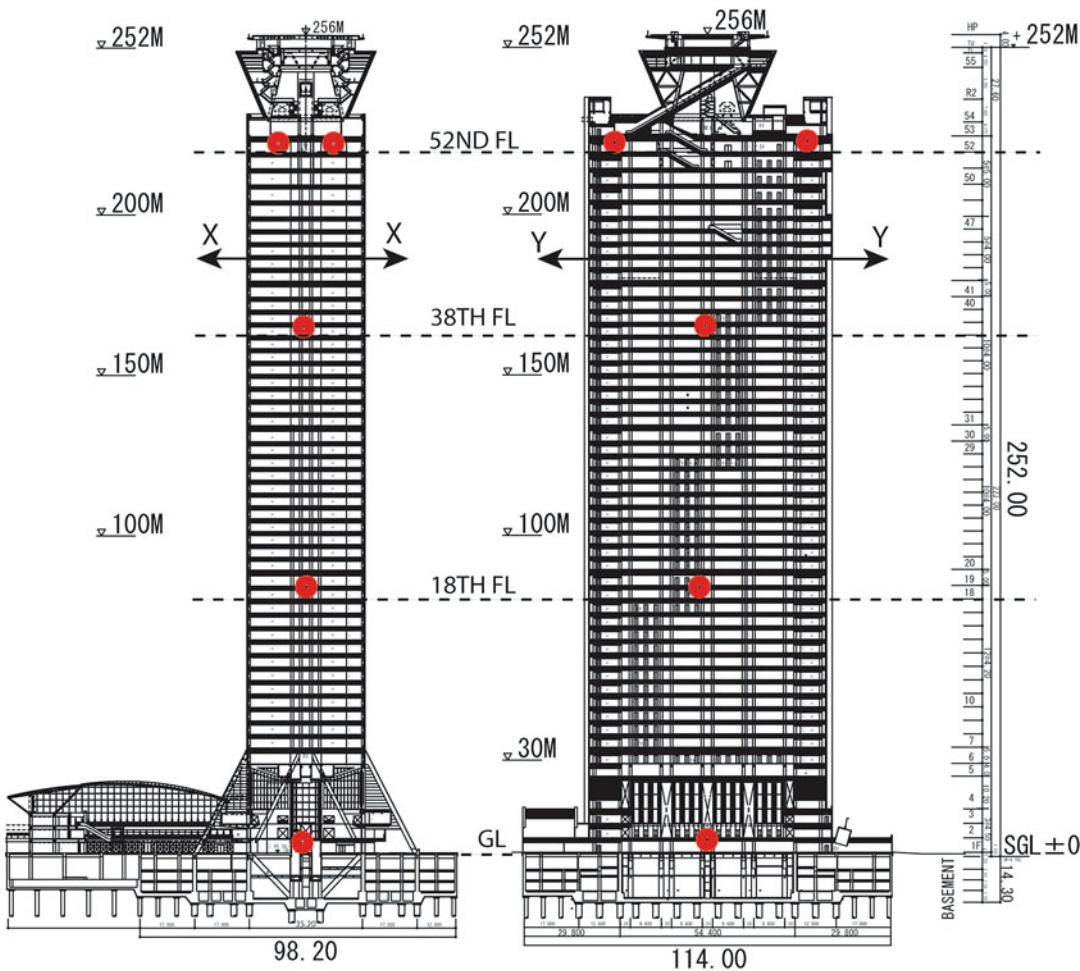
The building exhibited fundamental frequencies (periods) of 0.152Hz (6.58 s) in the X-direction and 0.145 Hz (6.90) in the Y-direction as seen in Fig. 5 (right). Such close frequencies are clear confirmation of resonance that led to prolonged responses (~1,000s recorded) and long-duration high-amplitude shaking as exhibited in Fig. 4 which shows average drift ratios.

The average drift ratios (Fig. 6) computed from relative displacements between many floors indicate that maximum average drift ratios experienced during the mainshock were between 0.5 % and 1.0 % for the X-direction and 0.2–0.4 % for the Y-direction. These average drift ratios are less than the maximum 1 % limit usually used in Japan for collapse protection level motions (level 2 used for buildings 60 m

or taller (The Building Center of Japan 2001a, b)). However, average drift ratios are much larger than expected for an input motion with a small peak acceleration in the order of only 3 % g. In the United States, the comparative maximum drift ratio for tall buildings for Risk Category 1 or 2 is 2 % (Table 12.12, ASCE7-10 2007).

Detailed studies of the Osaka basin are presented in Yamada and Horike (2007), Sekiguchi et al. (2007), and Iwaki and Iwata (2008), but the results of these studies likely do not accurately characterize the local site transfer function. Instead, we computed the site transfer functions using software developed by C. Mueller, 1997 (“personal communication”), which is based on Haskell’s shear wave propagation method (Haskell 1953, 1960). In this method, the transfer function is computed using linear propagation of vertically incident SH waves and has, as input, data related to the layered media (number of layers, depth of each layer, corresponding Vs, damping, and density), desired depth of computation of transfer function, computation frequency (df), half space substratum shear wave velocity, and density. Damping (ξ) in the software is introduced via the quality factor (Q), a term used by geophysicists that is related to damping by $\xi = 1/(2Q)$.

In summary, during the Tohoku event, the building was subjected to ground level input



Drift Issues of Tall Buildings in 2011 Tohoku Earthquake, Fig. 3 Vertical sections of the building showing major dimensions and locations of triaxial accelerometers

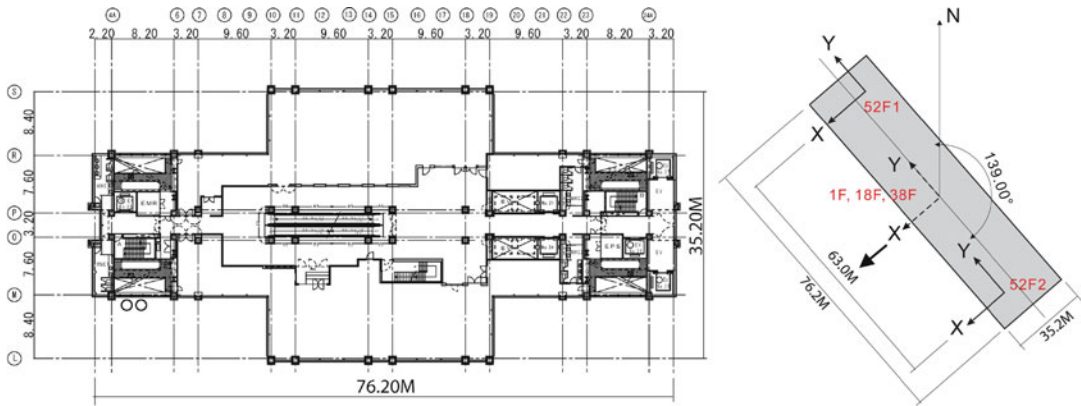
on the 52nd, 38th, 18th, and ground level (First floor). *X* and *Y* denote principal axes of the building (see Fig. 4) (From Çelebi et al. 2012)

acceleration of 3 % of *g*. But the deformations were large enough (to realize sizeable average drift ratios) that the building lost its functionality for several weeks. Elevator cables were entangled. This case, along with others, deserves further studies as to how tall buildings would respond if the ground input accelerations were 20–30 % of *g* with similar low-frequency content that could be caused by a closer earthquake (e.g., ~100 km or less) [e.g., similar to 1923 Kanto and 1994 Kobe earthquakes]. It is important to note that same elevator cable tangling problems also occurred in several tall buildings in the Shinjuku area of Tokyo, Japan (Hisada et al. 2012a).

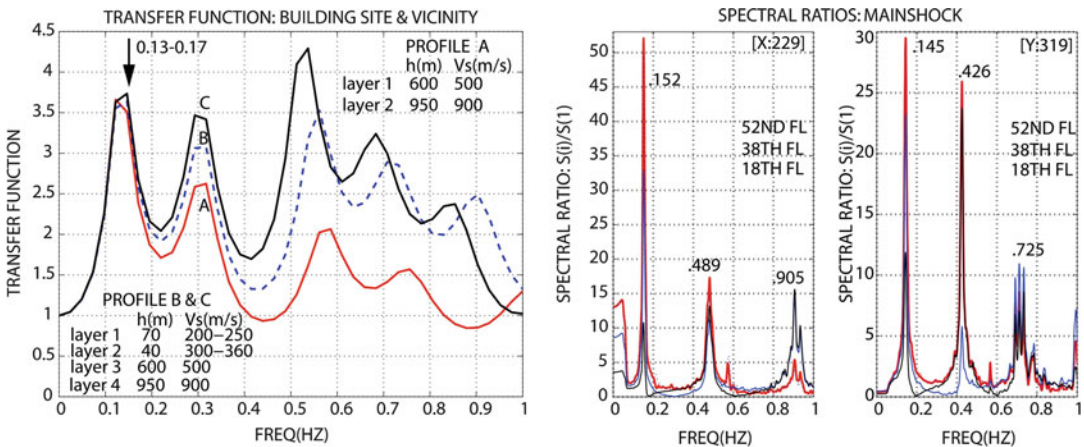
Furthermore, the actual drift ratios computed from relative displacements divided by story heights between some of the pairs of two consecutive floors are certainly to be larger than the average drift ratios computed from differential displacements between several floors, due to sparse deployments of instruments, as in this case.

Building B: 55-Story Shinjuku Center Building in Shinjuku, Tokyo

According to EERI Special Earthquake Report (EERI Newsletter 2012), the 54-story Shinjuku Center Building was constructed in 1979.



Drift Issues of Tall Buildings in 2011 Tohoku Earthquake, Fig. 4 (Left) Typical plan view (the figure shows 52nd floor) and (right) orientations (From Çelebi et al. 2012)



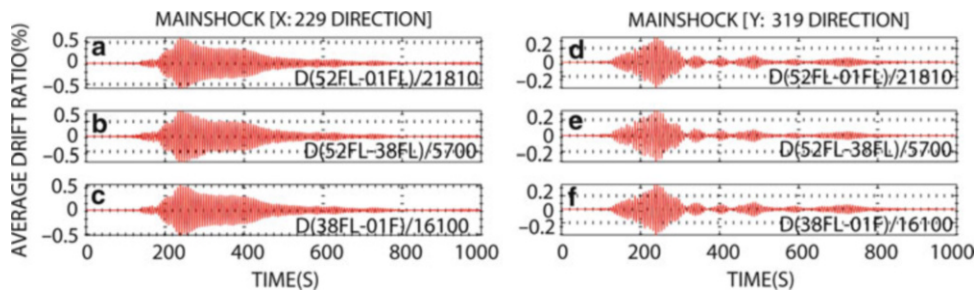
Drift Issues of Tall Buildings in 2011 Tohoku Earthquake, Fig. 5 (Left) Transfer functions computed for Profile A (near the OSKH02 strong-motion site; www.kik.bosai.go.jp/, last accessed 09/16/2011) and Profiles B and C below the building. The depth of the softer upper two layers (to about 1,500 m depth) below the building does not significantly change the position of the

peaks in the transfer function, particularly for the fundamental mode of the site. (Right) Spectral ratios of amplitude spectra at 52nd floor, 38th floor, and 18th floor with respect to that at first floor. Note that the third mode in X-direction is identified from the ratios. Different colors (red, black, and blue) are used only to distinguish lines corresponding to different floors in descending order

The report states: “The structure’s height is 223 m, and the first natural period of the structure is 5.2 and 6.2 s in two perpendicular directions. In 2009, the building was retrofitted from the 15th to 39th floor with 288 oil dampers that were configured to exhibit a form of deformation dependency, in addition to velocity dependency. In the March quake, the dampers were calculated to have reduced the maximum accelerations by 30 % and roof displacements by 22 %.”

Figure 7 shows a picture of the building and the oil dampers installed in 2009.

Figure 8 shows recorded acceleration time history at first floor and computed displacements at the roof during the Tohoku earthquake at ~375 km from the epicenter. Without the actual data being available, as shown in Fig. 8, the maximum roof displacement is in the transverse (Y) direction and reported as 54.2 cm (Figure adopted from Moehle, 2012, “personal



Drift Issues of Tall Buildings in 2011 Tohoku Earthquake, Fig. 6 Average drift ratios computed from displacements between 52nd, 38th, and 1st floors. In each

frame, the numbers in denominators are distances (in cm) between the designated floors

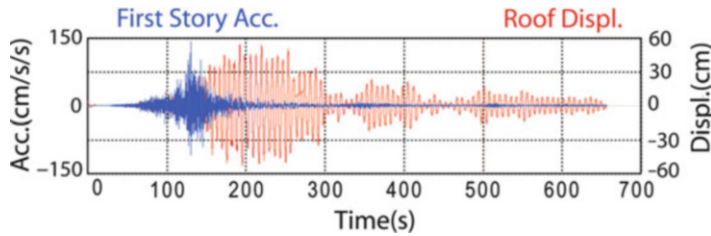


Drift Issues of Tall Buildings in 2011 Tohoku Earthquake, Fig. 7 (Left) Picture of Shinjuku Center Building and (right) dampers installed in 2009 (From EERI Newsletter 2012)

communication” and revised Sinozaki (2012), “personal communication”). Hosozawa et al. (2012) as well as Takewaki et al. (2011) also confirm these displacements. Assuming that this is the same as the maximum relative displacement between ground floor and the roof, with a height of 216 m (Sinozaki, personal communication), the average drift ratio can be assumed to be $\sim 54/21600 \sim .25\%$ (see Table 2). This is the average. Hence, the maximum drift ratio is $> .25\%$ even with claimed reductions in

maximum accelerations and displacements thanks to the dampers. Without actual data, it is not possible to compute possibly how much larger than 0.25% is the drift ratio between any two consecutive floors.

Again, similar to the discussion related to Building A, the actual drift ratios computed from relative displacements divided by story heights between some of the pairs of two consecutive floors are certainly to be larger than the average drift ratio computed using the maximum



Drift Issues of Tall Buildings in 2011 Tohoku Earthquake, Fig. 8 Recorded first-story accelerations (max $\sim 14\%$ of g) and computed roof displacements

(max ~ 54 cm) during Tohoku earthquake (Figure adopted from J. Moehle and Y. Sinozaki)

roof displacement divided by the height of the building, due to unavailable detailed data, as done in this case.

Buildings C and D: Two Tall Buildings at Kogakuin University Campus, Shinjuku, Tokyo

Without mentioning drift ratios, detailed studies of two buildings on the Kogakuin University Shinjuku, Tokyo, campus are provided by Hisada et al. (2012a, b). Figure 9 shows the vertical sections of the two subject buildings (Fig. 10).

Drift Issues of Tall Buildings in 2011 Tohoku Earthquake, Table 2 Maximum observed responses (From Hosozawa et al. 2012 and Sinozaki, 2012, “personal communication”)

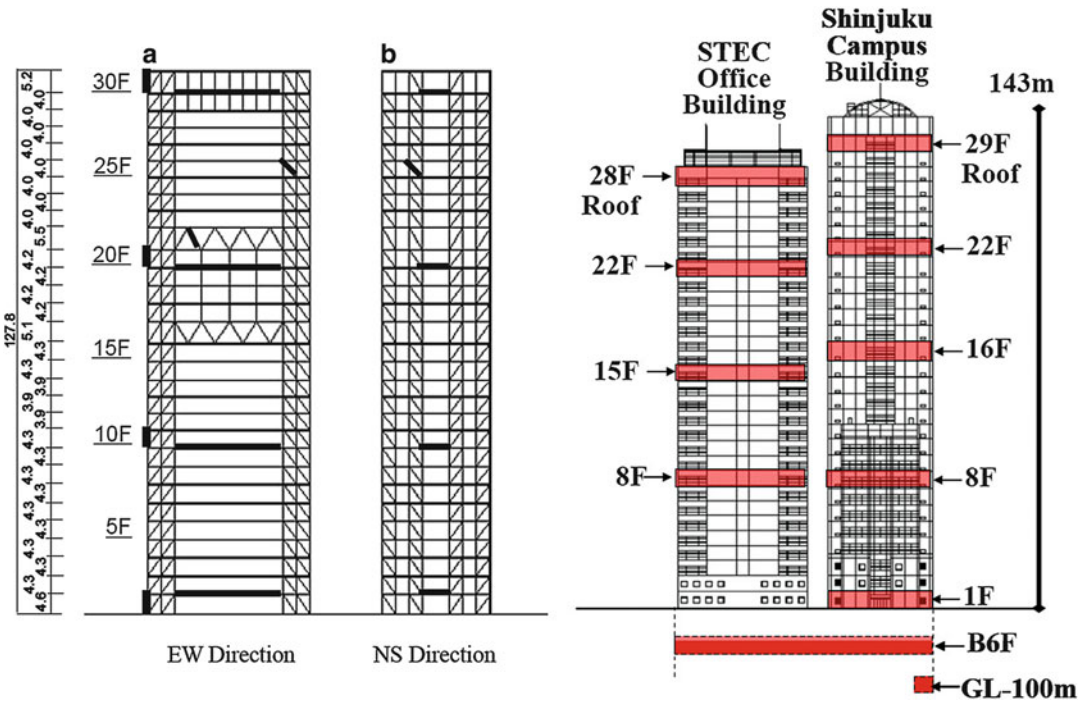
	Max. Accel. (cm/s/s)		Max displ. (cm)	
	Long. (X)	Trans. (Y)	Long. (X)	Trans. (Y)
Roof	236.0	161.3	49.4	54.2
28th floor	112.7	171.3	26.3	33.3
1st floor	94.3	142.1		

Discussion and Conclusions

The four examples presented herein demonstrate that tall buildings at far distances from the epicenter of M9.0 Tohoku earthquake of March 11, 2011 were affected by long-period ground motions. Tall buildings in Tokyo (~ 350 – 375 km from epicenter) and Osaka (~ 770 km from the epicenter) shook for long durations. In addition to these examples, other examples are discussed in Takewaki et al. (2011).

Long-period dominance of ground motions at far distances from the epicenter of the earthquake is confirmed by velocity response spectra of ground motions. These spectra confirm also the recommended design spectra issued 4 months before the Tohoku earthquake by MLIT that accounts for the effects of long periods on tall buildings in 9 regions of Japan that include the buildings discussed herein. Building A is in one of these regions in Osaka and Buildings B, C, and D are all in another region in Shinjuku area of Tokyo.

For small ground-level input ground motions as in the cases presented herein, these four tall buildings deformed significantly to experience sizeable drift ratios as summarized in Table 3. The drift ratios discussed herein are computed from recorded response data from accelerometers that are sparsely deployed in both buildings, hence, the reason for referring them as average drift ratios. Therefore, it is deduced that, for some pairs of consecutive floors, the actual drift ratios are larger than the average. Collection of such data is essential (a) to assess the effect of long-period ground motions on long-period structures caused by sources at large distances and (b) to consider these effects and discuss whether the design processes should consider reducing drift limits to more realistic percentages and lower them to significantly less than 1 % as in Japan for collapse protection level motions (level 2 used for buildings 60 m or taller (The Building Center of Japan 2001a, b)). This ratio is much lower than the 2 % in the United States which is the comparative maximum drift ratio for tall



Drift Issues of Tall Buildings in 2011 Tohoku Earthquake, Fig. 9 Vertical sections showing instrumented floors of 30 storeys Shinjuku Campus Building and 28

storeys STEC Office Building in Shinjuku area of Tokyo, Japan (Figure adopted from Hisada et al. (2012a and b))

buildings for Risk Category 1 or 2 (Table 12.12, ASCE7-10 2007). Even better, drift ratios closer to the 0.2 % imposed by the Chilean Code for Earthquake Resistant Design of Buildings (NCh433.Of96 1996) should be considered. And (c) finally, further applications of unique response modification features are feasible to reduce the drift ratios.

Finally, the following points are made:

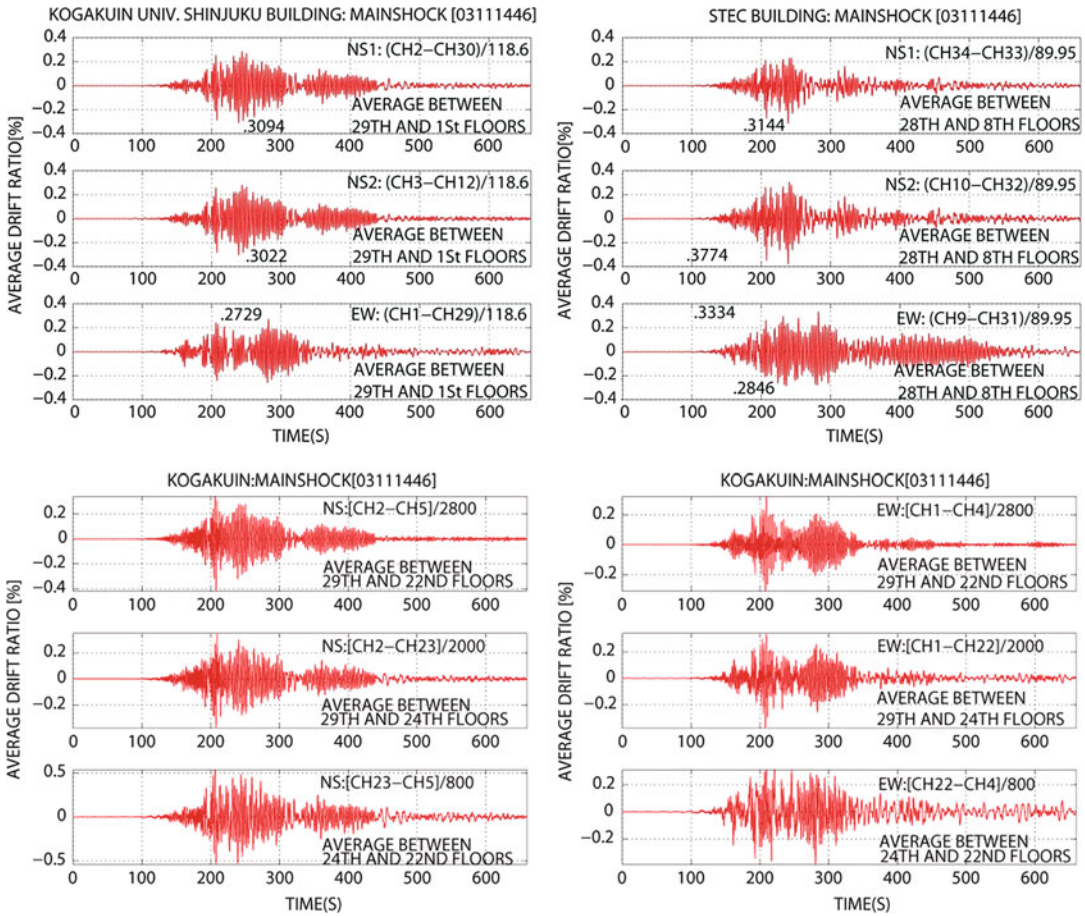
1. Behavior and performances of these particular tall buildings far away from the strong shaking source of the M9.0 Tohoku earthquake of 2011 and large-magnitude aftershocks should serve as a reminder that, in the United States as well as in many other countries, risk to such built environments from distant sources must always be considered.
2. The potential risk from closer large-magnitude earthquakes that could subject the buildings to larger peak input motions should be assessed in light of the substantial drift

ratios under the low peak input motions experienced during and following the Tohoku earthquake of 2011.

3. Prolonged cyclic behavior of the buildings, even at moderate to small amplitudes with relatively acceptable elastic drift ratios, can be a cause for low-cycle fatigue. This issue is significant, and unless structural dynamics characteristic of the buildings is modified, low-cycle fatigue can be an important concern for the health of the building in the future.

Summary

One of the most significant effects of the M9.0 Tohoku, Japan, earthquake of March 11, 2011, is the now well-known long-duration (>10 min) shaking of buildings in Japan – particularly those in Tokyo (~350–375 km from the epicenter) and in places as far as Osaka (~770 km from the epicenter). Although none collapsed, the



Drift Issues of Tall Buildings in 2011 Tohoku Earthquake, Fig. 10 Combinations of average drift ratios computed from recorded data of 30 storeys Shinjuku Campus Building and 28 storeys STEC Office Building

strong shaking caused many tall buildings not to be functional for days and weeks.

The purpose of this entry is to discuss the behavior and performance of four tall buildings, considered to be representative of most tall buildings in Tokyo and other locations and from which shaking data were retrieved. Of particular interest is a building in Osaka that almost reached an average of 0.5 % drift ratio – this, with a ground level input motion of ~ 3 % g is significant. What might have happened during an event with input level motions with similar low-frequency content and in 10–20 % g range is a legitimate question that must be pondered. The particular building had serious site effects and was in resonance.

in Shinjuku area of Tokyo, Japan. Highest drift ratio ($\sim 0.5\%$) is computed from displacements in the NS direction between 24th and 22nd floors (bottom left frame)

The other three examples are from Tokyo. For example, based on records obtained from a 54-story building retrofitted with 288 oil dampers on 24 floors, computations show that average drift ratio may have reached ~ 0.3 % and maximum drift ratio likely was $> .3$ %. The maximum allowed by Japanese practice for buildings taller than 60 m and for collapse protection (level 2) motions is 1 % (The Building Center of Japan, 2001a, b).

Performances of tall buildings in many seismically active regions of the world (e.g., Chile, Turkey) or those tall buildings affected by long-distance long-period effects by sources at a distance (e.g., Abu Dhabi, Dubai) are of

Drift Issues of Tall Buildings in 2011 Tohoku Earthquake, Table 3 Epicentral distances, peak accelerations and displacements, and computed average drift ratios of the buildings supplemented by peak accelerations of ground motions recorded at stations in proximity to the buildings

Bldg	Epicentral distance (km)	Peak accel. at ground level (gals)	Peak accel. at top floor (gals)	No of floors	Rel. displ (cm)	H Height (m)	(Ave) DR (%)
A	767	30	130	54	~130	218	~.5
B ^a	~375	142	236	54	54.2	216	~.2
C ^b	~375	92	340	31	33.4	119	~.5
D ^b	~375	92	302	29	34.9	90	~.4
Peak accelerations (gals) at Free-Field Stations in the vicinity of the buildings							
Free-Field Station	NS	EW					
KIKNET:OSKH02 \surface	13	15					
OSKH02\downhole	1	1					
KNET: TKY007	198	165					

H height of building used to computed average drift ratios, *DR* drift ratio

^aFrom Sinozaki, 2012, (“personal communication”)

^bFrom Hisada and others (2012a, b)

interest to the earthquake engineering community. Chile imposes 0.2 % drift limit that results in elastic design. The United States and Turkey impose 2 % drift limit. Such wide variations of drift limits in design practices deserve discussion in light of functionality and performance of tall buildings during the 2011 Tohoku event.

Cross-References

- [Building Monitoring Using a Ground-Based Radar](#)
- [System and Damage Identification of Civil Structures](#)

References

- Anderson JP, Bodin P, Brune JN, Prince J, Singh SK, Quass R, Onate M (1986) Strong ground motion from the Michoacan, Mexico, earthquake. *Science* 233:1043–1049
- ASCE/SEI Standard 7–10 (2007) Minimum design loads for buildings and other structures. American Society of Civil Engineers, 1801 Alexander Bell Drive, Reston, Virginia 20191, USA
- Çelebi M (2008) Real-time monitoring of drift for occupancy resumption. In: *Proceedings of the 14WCEE (CD-ROM)*, Beijing, 13–17 Oct 2008
- Çelebi M, Prince J, Dietel C, Onate M, Chavez G (1987) The culprit in Mexico City – amplifications of motions. *Earthq Spectra*, J EERI 3:2. 315–328
- Çelebi M, Sanli A, Sinclair M, Gallant S, Radulescu D (2004) Real-time seismic monitoring needs of a building owner and the solution – a cooperative effort, *Journal of EERI. Earthq Spectra* 20(2):333–346
- Çelebi M, Okawa I, Kashima T, Koyama S, Iiba M (2014) Response of a tall building far from the epicenter of the March 11, 2011 M = 9.0 Great East Japan earthquake and its aftershocks. *J Design Tall Build Spec Struct (a John Wiley Journal)* 23:427–441
- DBYYHY (2007) Deprem bolgelerinde yapılacak binalar hakkında yönetmelik (Code for buildings to be built in earthquake prone regions of Turkey)
- EERI Newsletter (2012) EERI special earthquake report: Learning from earthquakes Performance of Engineered Structures in the Mw 9.0 Tohoku, Earthquake of 11 Mar 2011 in EERI Newsletter (January 2012)
- Eurocode 8 (2008) Design of structures for earthquake resistance: Part1: general rules, seismic actions and rules for buildings, European standard, prEN1998-1
- Haskell NA (1953) The dispersion of surface waves on multi-layered media. *Bull Seismol Soc Am* 43 (1):17–34
- Haskell NA (1960) Crustal reflection of plane SH waves. *J Geophys Res* 65(12):4147–4150
- Hisada Y, Yamashita T, Murakami M, Kubo T, Shindo J, Aizawa K, Arata T (2012a) Seismic response and damage of high-rise buildings in Tokyo, Japan, during the Great East Japan earthquake. In: *Proceedings of the International symposium on engineering lessons learned from the 2011 Great East Japan Earthquake*, Tokyo, 1–4 March 2012, pp 1110–1119
- Hisada Y, Yamashita T, Murakami M, Kubo T, Arata T, Shindo J, Aizawa K (2012b) Seismic response and

- damage of high-rise buildings in Tokyo, Japan, during the 2011 Tohoku earthquake. In: Proceedings of the 15WCEE, Lisbon, Sep 2012
- Hodgson JJ (1964) Earthquakes and earth structure. Prentice-Hall, Englewood Cliffs
- Hosozawa O, Kimura Y, Aono H (2012) Seismic retrofit of high-rise building with deformation-dependent oil dampers. In: Proceedings of the 2nd international conference on archi-cultural translations through the silk road, Mukogawa Women's University, Nishinomiya, 14–16 July 2012
- Iwaki A, Iwata T (2008) Validation of 3-D basin structure models for long-period ground motion simulation in the Osaka basin, Western Japan. *J Seismol* 12: 197–215
- MLIT (2010a) Code draft for the retrofit of existing high-rise buildings and design guidelines for new high-rise buildings (in Japanese) (http://www.mlit.go.jp/report/press/house05_hh_000218.html)
- MLIT (2010b) Request for public comments, on earthquake countermeasures for high-rise buildings
- NCh433.Of96 (1996) Earthquake resistant design of buildings. Instituto de Normalization (INN), Chile
- NZS 1170.5 (2004) Structural design actions: Part 5: earthquake actions – New Zealand
- Okawa I (coordinator) (2010) Study on the safety countermeasures for super-high-rise buildings etc. against the long-period earthquake ground motions. *BRI Res Data* 127:1–453 (in Japanese)
- Sekiguchi H, Yoshimi M, Horikawa H, Yoshida K, Kunimatsu S, Satake K (2007) Prediction of ground motion in the Osaka sedimentary basin associated with the hypothetical Nankai earthquake. *J Seismol* 12:185–195
- Takewaki I, Murakami S, Fujita K, Yoshitomi S, Tsuji M (2011) The 2011 off the Pacific coast of Tohoku earthquake and response of high-rise buildings under long-period ground motions. *Soil Dyn Earthq Eng* 31:1511–1528
- Tezcan SS, Ipek M (1973) Long distance effects of the 29 March 1970 Gediz, Turkey earthquake. *Earthq Eng Struct Dyn* 1:203–215
- The Building Center of Japan (2001a) Time history response analysis building performance evaluation manual, Technical appraisal Department, Structural safety Section Report No: BR KO-02-01 (adopted 1 June 2000, amended 25 Apr 2001)
- The Building Center of Japan (2001b) Manual for time history response analysis of building performance evaluation manual, Technical appraisal Department, Structural safety Section Report No: BR KO-02-01 (adopted 1 June 2000, amended 25 Apr 2001)
- Turkish Code (2007) Specification for structures to be built in disaster areas. TEC-07. Ankara, Turkey: Turkish Ministry of Public Works and Settlement
- Yamada K, Horike M (2007) Influence of Q-values below 1 Hz from borehole and surface data in the Osaka basin by three-component waveform fitting. *Bull Seism Soc Am* 97(4):1267–1278

Dynamic Soil Properties: In Situ Characterization Using Penetration Tests

David Saftner¹ and Roman Hryciw²

¹Department of Civil Engineering, University of Minnesota Duluth, Duluth, MN, USA

²Department of Civil and Environmental Engineering, University of Michigan, Ann Arbor, MI, USA

Synonyms

Cone penetration test; Dilatometer test; Shear wave velocity; Small strain shear modulus; Standard penetration test

Introduction

For geotechnical earthquake engineering analysis, soil properties are determined through laboratory and in situ testing. Generally, geotechnical investigations include both components since they provide complementary information. The advantages of in situ testing include testing the soil in its native state, without the disturbance inherent in soil sampling, as well as being more efficient in terms of evaluating a greater volume of soil. While these advantages apply to many soil properties, the effect of disturbance is particularly relevant to dynamic (small strain) soil properties since they are most affected by it. Therefore, correlations have been developed to determine dynamic soil properties from in situ penetration tests including the standard penetration test (SPT), the cone penetration test (CPT), and the Marchetti (Flat Plate) dilatometer test (DMT).

This chapter provides a brief overview of in situ geotechnical penetration tests. Methods of estimating or determining dynamic soil properties from the SPT, CPT, and DMT are explained. A historical overview of correlations, as well as methods for directly measuring the shear wave velocity (V_s), is presented.

The content discussed in this chapter relates to several other entries in the *Encyclopedia of Earthquake Engineering*. While this chapter discusses methods of obtaining dynamic soil properties from penetration tests, closely related entry include ► [Dynamic Soil Properties: In Situ Characterization Using Penetration Tests](#). Because dynamic soil properties are used in seismic analysis and design, the entries covering geotechnical seismic design and analysis may also be of interest. These entries include ► [Seismic Analysis of Masonry Buildings: Numerical Modeling](#), ► [Seismic Design of Dams](#), ► [Seismic Design of Earth-Retaining Structures](#), and ► [Seismic Design of Tunnels](#).

Overview of Penetration Tests

This section describes the SPT, CPT, and DMT equipment and test procedures. Also discussed are the readings taken during each test and the engineering parameters deduced from them.

Standard Penetration Test

The SPT was developed in the 1920s and is the most common in situ penetration test. ASTM D1586 details the test's equipment and procedure. Due to its maturity and popularity, there is vast experience in interpreting SPT results and an extensive database of correlations between the SPT and soil properties. An advantage of the SPT by comparison to other penetration tests is that the SPT is conducted in boreholes where, in addition to the disturbed soil samples obtained by the SPT split-spoon sampler, "undisturbed" soil samples may also be collected. While no sampling method provides a truly undisturbed specimen, some methods and devices such as piston samplers with thin-walled tubes are less disruptive than others. In terms of relating SPT results to dynamic soil properties, this advantage means that in situ SPT results can be compared to laboratory test results from samples collected from the same borehole. There is, however, a caveat: it is virtually impossible to collect undisturbed samples of cohesionless soil without expensive ground freezing and coring. As such, laboratory tests only bracket the soil

properties, while in situ tests reflect the actual combined effects of both material (particles size and shape distribution, mineralogy, etc.) and state (confining stresses, void ratio, etc.) properties.

A drilling rig, shown in Fig. 1, is used to advance boreholes. Depending on soil and groundwater conditions, hollow stem augers (Fig. 2), casing, or bentonite slurry (drilling mud) may be used to maintain borehole stability. The SPT can also be conducted using light-weight portable equipment, such as shown in Fig. 3. Stable, near-surface soils can be tested in shallow boreholes advanced by hand augers or post-hole diggers.

When a test depth is reached, a split-spoon sampler (Fig. 4) is connected to the end of a string of drill rods and lowered into the hole. The split-spoon sampler is cylindrical with a 1.5 in. inner diameter and 2.0 in. outer diameter. Its length is between 1.5 and 2.5 ft. As shown in Fig. 4, the sampler is made of two symmetrical halves, splitting the cylinder lengthwise. This allows for visual inspection and removal of the soil after sampling. Most split-spoon samplers can accommodate a liner, creating a uniform inner diameter. When the liner is not used, the inner diameter is slightly larger, reducing the friction between the incoming soil and sampler. The top of the rod assembly is connected to a drive-weight assembly consisting of a 140 lb hammer, anvil, and a lift-drop system. After lowering the split-spoon sampler to the bottom of the borehole, the drill rod string is attached to the drive-weight assembly.

In the early years of the SPT, equipment and procedural differences made it impossible to compare test results. However, the procedures described in ASTM D1583 and recommendations summarized by Seed et al. (1985) have largely standardized the test, hence its name. The test requires the 140 lb hammer to be raised 30 in. and dropped, driving the sampler into the soil. This process is repeated until the sampler has penetrated 18 in. into the soil at the bottom of the borehole. The number of blows required to advance the sampler each of three 6 in. increments are recorded. After lifting the split-spoon sampler from the borehole and removing the soil

**Dynamic Soil Properties:
In Situ Characterization
Using Penetration Tests,**

Fig. 1 Typical drill rig used to perform the standard penetration test



sample, the borehole is advanced to the next test depth. The SPT is typically performed at 3–5 ft intervals although continuous sampling (1.5 ft test intervals) may also be used.

The SPT blow count (N) is computed by summing the blows from the final two 6 in. increments. The readings from the first 6 in. increment are discarded because a number of factors unrelated to soil properties and state conditions influence the results. These include disturbance of the bottom of the borehole and soil cuttings that have fallen to the bottom of the borehole.

Early SPT hammers transferred approximately 60 % of the theoretical impact energy to the split-spoon sampler. Therefore, many correlations between SPT- N and soil properties are based on this 60 % efficiency. Other factors affecting N include borehole diameter, type of sampler, and rod length. The blow count standardized to 60 % efficiency and corrected for the other factors is designated as N_{60} . Skempton (1986) used Eq. 1 to convert N to N_{60} :

$$N_{60} = \frac{E_m C_B C_S C_R N}{0.60} \quad (1)$$

where:

E_m = hammer efficiency (0.43–0.85; ~0.6 in US).

C_B = correction for borehole diameter (1.0 for 2.5"–4.5" diameter).

C_S = correction for sampler type (1.0 for standard sampler).

C_R = correction for rod length (1.0 for greater than 30' total length).

N = field measured SPT blow count.

Additional information on compensating for hammer efficiency and the correction factors are found in virtually all modern foundation engineering textbooks.

Cone Penetration Test

The CPT was developed in Europe in the 1930s. While many subsequent modifications have been made, the current testing method can be traced to



Dynamic Soil Properties: In Situ Characterization Using Penetration Tests, Fig. 2 Advancing a hollow stem auger through which an SPT will be conducted

Holland. Thus, the test is also referred to as the Dutch cone test. By comparison to the SPT, which provides data only at discrete elevations, the CPT provides continuous soil data with depth. Unlike the SPT, it also exhibits excellent test repeatability.

The CPT device is a probe with a conical tip as shown in Fig. 5. The cone typically has either a 10 cm^2 or 15 cm^2 projected area and a 60° apex angle. A 13.3 cm long friction sleeve is located behind the tip. The disassembled cone, pictured in Fig. 6, shows the tip and friction sleeve separately. The white ring located between the tip and friction sleeve in Fig. 5 is a porous stone which must be saturated prior to the test. A pressure transducer mounted within the probe measures pore water pressure during cone advance and may also be used to monitor excess pore water pressure dissipation when the cone is not advancing in order to estimate soil hydraulic conductivity.

A CPT rig's hydraulic actuator pushes the probe into the ground at a steady speed not exceeding 2 cm/s . The rig's dead weight provides the reaction. As shown in Fig. 7, rigs range in size, and smaller rigs may require anchoring into the ground to provide the needed reaction. The cone is pushed into the ground using continuous hollow rods having the same diameter as a 10 cm^2 cone. To minimize friction along the side of the rods, a larger diameter "friction reducer" may be added between the CPT probe and push rods. Because all readings are taken at the probe and because the friction reducer is located above the probe, it does not affect the test results. If a 15 cm^2 cone is used, the cone itself creates an enlarged hole thus eliminating the need for a friction reducer.

An electronic cable threaded through the hollow rods connects the probe to a data acquisition system at the surface. Readings are taken at small regular depth intervals, typically 2 cm or 5 cm . Three measurements are recorded at each depth: the tip resistance (also called cone bearing), q_c , the side friction (also called local friction), f_{cs} , and the pore water pressure, u . When the test is completed, the probe is withdrawn from the ground, and the hole is filled with grout or bentonite slurry.

Flat Plate Dilatometer Test

The DMT is the newest of the three penetration tests. It was developed in the late 1970s (Marchetti 1980). It is commonly used alongside the CPT since it uses the same rods and pushing system.

The dilatometer blade is shown in Fig. 8. It is 320 mm tall, 95 mm wide, and 14 mm thick. A 60 mm diameter flexible metal diaphragm is located at the center on one side of the blade. The blade is bored to provide access for an electrical line and gas pressure to an area behind the diaphragm. An electrical-pneumatic cable is strung from the blade through the rods and attached to a control unit, shown in Fig. 9.

Unlike the CPT probe which is advanced continuously and paused only for addition of the next rod at 1 m increments, the DMT blade is stopped at predetermined intervals, typically 20 cm , for testing. During advance, the diaphragm is pressed



Dynamic Soil Properties: In Situ Characterization Using Penetration Tests, Fig. 3 Typical light-weight SPT equipment, including a portable tripod SPT system with safety hammer and gas powered auger

Dynamic Soil Properties: In Situ Characterization Using Penetration Tests, Fig. 4 Opened split-spoon sampler revealing clayey soil being tested with a pocket penetrometer



against the blade by lateral earth pressure, an electric circuit is completed, and an audible tone sounds from the dilatometer control unit. When the blade reaches a test depth, its advance is halted, and pressure is added until the diaphragm is unseated from the blade. This breaks a circuit, and the buzzer in the control unit ceases to sound. The gas pressure at which this occurs is recorded

as the “A” reading. When the center of the diaphragm expands to 1.1 mm from the blade, another circuit is completed, and the buzzer begins to sound again. The corresponding pressure is recorded as the “B” reading. Lastly, the diaphragm is slowly deflated until contact with the blade is restored and the buzzer sounds again. This pressure, taken as the “C” reading,

corresponds to pore water pressure in free draining materials as water rapidly fills the gap between the deflating membrane and coarse grained soil. After recording the “C” reading, all pressure is released, and the blade is advanced to the next test depth.

Following data collection, the DMT readings are corrected for the stiffness of the steel diaphragm and gauge offset at atmospheric pressure. With these corrections, the gage readings A, B, and C become corrected pressures p_0 , p_1 , and p_2 ,

respectively. The corrected pressures are used to compute three DMT indices: the lateral stress index, K_D ; the dilatometer modulus, E_D ; and the material index, I_D , as defined by Eqs. 2, 3, and 4:

$$K_D = \frac{p_0 - u_0}{\sigma'_v} \quad (2)$$

$$E_D = 34.7(p_1 - p_0) \quad (3)$$

$$I_D = \frac{p_1 - p_0}{p_0 - u_0} \quad (4)$$

where:

u_0 = hydrostatic water pressure at the test depth.

σ'_v = effective vertical stress at the test depth.

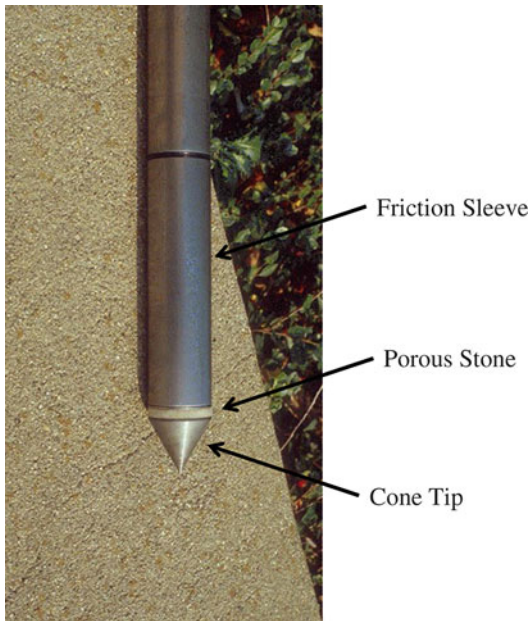
The DMT has a smaller database of results with which to build correlations. However, the test provides a wealth of unique data, affording interesting possibilities for estimating dynamic soil properties through empirical correlations.

Dynamic Soil Properties from SPT

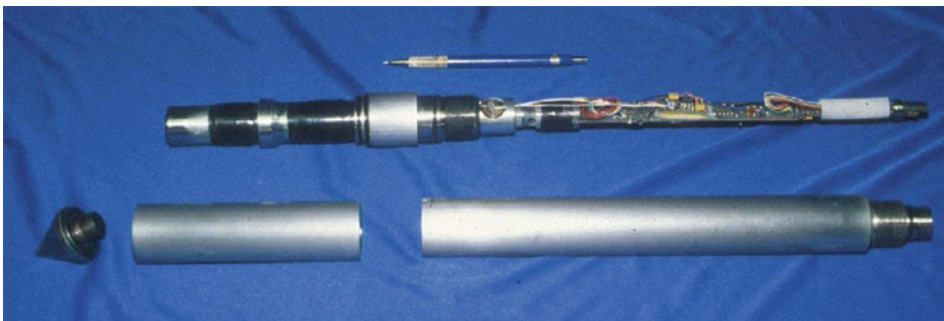
The majority of methods for determining dynamic soil properties from the SPT correlate N to V_s . One can then compute the small strain shear modulus, G_0 by

$$G_0 = \rho V_s^2 \quad (5)$$

where ρ = mass density.



Dynamic Soil Properties: In Situ Characterization Using Penetration Tests, Fig. 5 Electronic cone penetrometer



Dynamic Soil Properties: In Situ Characterization Using Penetration Tests, Fig. 6 Disassembled cone penetrometer



Dynamic Soil Properties: In Situ Characterization Using Penetration Tests, Fig. 7 Types of CPT rigs: (a) trailer mounted, (b) small light-weight CPT rig, (c) large commercial rig, (d) tracked CPT rig

While G_0 is commonly used in dynamic soil analysis, other small strain soil moduli related to G_0 may be of interest. These relationships are given by Eqs. 6 through 8 for the Young's modulus, E ; bulk modulus, B ; and constrained modulus, M , respectively:

$$E = 2G(1 + \nu) \quad (6)$$

$$B = \frac{2G(1 + \nu)}{3(1 - 2\nu)} \quad (7)$$

$$M = \frac{2G(1 - \nu)}{1 - 2\nu} \quad (8)$$

where ν = Poisson's ratio.

Due to their dependence on strain, it is important to remember that all of the soil moduli determined from V_s , which is measured at strains of $\sim 10^{-4}\%$ or

smaller, are small strain moduli. The SPT is a large strain test, and, therefore, correlations between N and V_s should be used with caution. Correlations developed in one geologic setting, and particular soil type should not be expected to accurately predict V_s in other situations.

Many such correlations have been developed, and some are presented in Table 1. The following review summarizes the conditions under which these relationships were developed and is intended to aid the practitioner in choosing an appropriate relationship if no other means of determining V_s exist. As shown in Figs. 10, 11, and 12, for sand, clay, and all soils, respectively, the different V_s versus N correlations do not show very good agreement. The authors therefore recommend using the correlations that were developed under conditions most similar to those of the users' site.



Dynamic Soil Properties: In Situ Characterization Using Penetration Tests, Fig. 8 Flat plate (Marchetti) dilatometer blade

Kanai et al. (1966) used the predominant period of microtremors measured during the Matsushiro earthquake swarm in Nagano, Japan, to estimate V_s of sand deposits. They compared these values to 52 SPTs conducted in the area to develop their correlation. Ohba and Toriuma (1970) developed their relationship from SPT results in alluvial soil deposits in the Osaka area. Fujiwara (1972) reported data collected after the 1968 Tokachi-oki earthquake and base his relationship on that investigation. Ohsaki and Iwasaki (1973) developed their relationship based on approximately 220 sets of data relating N to G_0 from a database collected by the Committee for Evaluation of Structural Safety of High-Rise Buildings, a part of the Japanese Ministry of Construction. Imai (1977) published results of 244 adjacent SPT and down-hole tests across Japan. For the relationship by Ohta and Goto (1978), 300 sets of data from tests conducted in alluvial plains in Japan were sorted by depth, soil type, and age of deposit.

The Japanese Road Association (1980) offered equations for sand and clay, but did not explain the process of developing the relationships. Imai and Tonouchi (1982) used data from 250 sites and 386 SPT boreholes across Japan.

Dynamic Soil Properties: In Situ Characterization Using Penetration Tests, Fig. 9 DMT control unit



Dynamic Soil Properties: In Situ Characterization Using Penetration Tests, Table 1 Summary of correlations between SPT data and shear wave velocity

Citation	Shear wave velocity, V_s (m/s)		
	Sand	Clay	All soil
Kanai et al. 1966	$V_s = 19N^{0.6}$		
Imai and Yoshimura 1970			$V_s = 76N^{0.33}$
Ohba and Toriuma 1970			$V_s = 84N^{0.31}$
Shibata 1970	$V_s = 32N^{0.5}$		
Fujiwara 1972			$V_s = 92.1N^{0.337}$
Ohta et al. 1972	$V_s = 87N^{0.36}$		
Ohsaki and Iwasaki 1973	$V_s = 59N^{0.47}$		$V_s = 82N^{0.39}$
Imai et al. 1975			$V_s = 90N^{0.341}$
Imai 1977			$V_s = 91N^{0.337}$
Ohta and Goto 1978	$V_s = 88N^{0.34}$		$V_s = 85.35N^{0.348}$
JRA 1980	$V_s = 80N^{0.33}$	$V_s = 100N^{0.33}$	
Imai and Tonouchi 1982			$V_s = 97N^{0.314}$
Seed et al. 1983	$V_s = 56N^{0.5}$		
Sykora and Stokoe 1983	$V_s = 100N^{0.29}$		
Fumal and Tinsley 1985	$V_s = 152 + 5.1N^{0.27}$		
Jinan 1987			$V_s = 116.1(N + 0.3185)^{0.202}$
Lee 1990	$V_s = 57N^{0.49}$	$V_s = 114N^{0.31}$	
Athanasopoulos 1995			$V_s = 108N^{0.36}$
Sisman 1995			$V_s = 32.8N^{0.51}$
Iyisan 1996			$V_s = 51.5N^{0.516}$
Kayabali 1996	$V_s = 175 + 3.75N$		
Jafari et al. 1997			$V_s = 22N^{0.85}$
Pitilakis et al. 1999	$V_s = 145(N_{60})^{0.178}$	$V_s = 132(N_{60})^{0.271}$	
Wride et al. 2000	$V_{s1} = X(N_1)_{60}^{0.25}$ Where $74.8 \leq X \leq 102.1$		
Jafari et al. 2002		$V_s = 27N^{0.73}$	
Andrus et al. 2004			$V_{s1cs} = 87.8N_{1,60cs}^{0.25}$
Hasancebi and Ulusay 2007	$V_s = 90.82N^{0.319}$	$V_s = 97.89N^{0.268}$	$V_s = 90N^{0.308}$
Hasancebi and Ulusay 2007	$V_s = 131N_{60}^{0.205}$	$V_s = 107.63N_{60}^{0.237}$	$V_s = 104.79N_{60}^{0.26}$
Hanumantharao and Ramana 2008			$V_s = 82.6N^{0.43}$
Anbazhagan and Sitharam 2008			$V_s = 78N_{60}^{0.4}$
Uma Maheshwari et al. 2010			$V_s = 95.641N^{0.3013}$
Anbazhagan et al. 2013			$V_s = 68.96N^{0.51}$
Sun et al. 2013			$V_s = 65.64N^{0.407}$

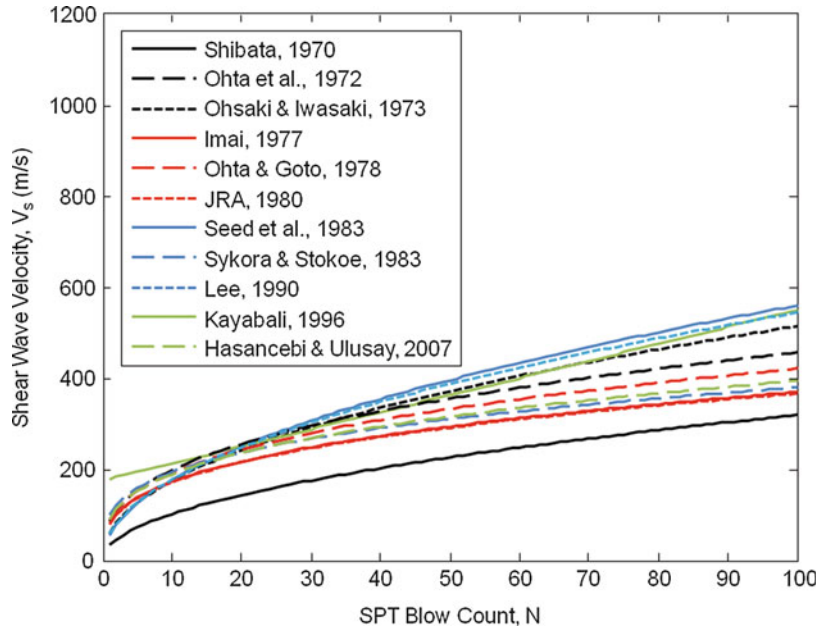
Seed et al. (1983) adapted Ohaski and Iwasaki's (1973) relationship based on data collected following liquefaction events. Sykora and Stokoe (1983) measured V_s using the cross-hole method and used 229 data pairs to develop their correlation. Fumal and Tinsley (1985) based their

correlation on V_s down-hole measurements at 84 sites across the Los Angeles basin. Jinan (1987) proposed a correlation based on approximately 100 measurements in Shanghai.

Lee (1990) conducted 15 SPTs and used 88 data pairs to construct his relationship for

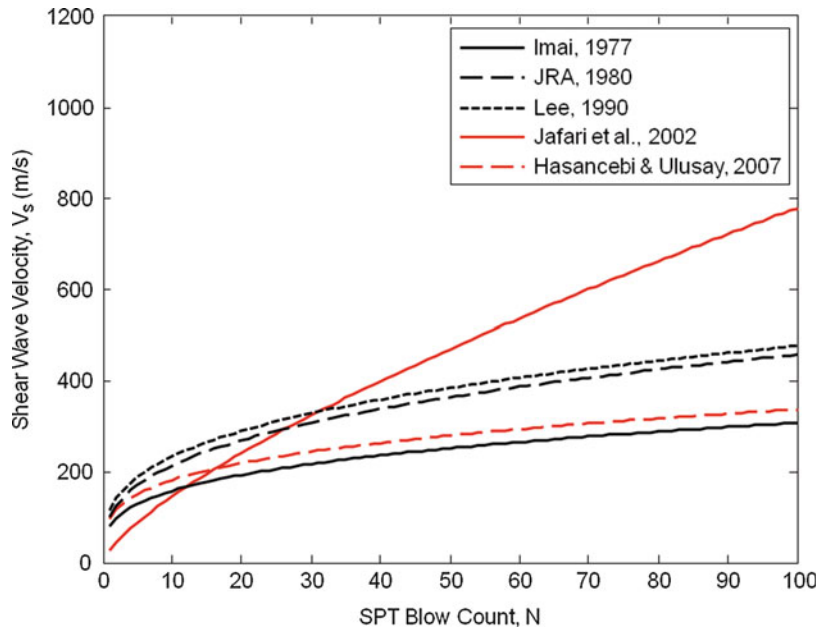
**Dynamic Soil Properties:
In Situ Characterization
Using Penetration Tests,**

Fig. 10 Shear wave velocity versus SPT- N correlations for sands



**Dynamic Soil Properties:
In Situ Characterization
Using Penetration Tests,**

Fig. 11 Shear wave velocity versus SPT- N for clays



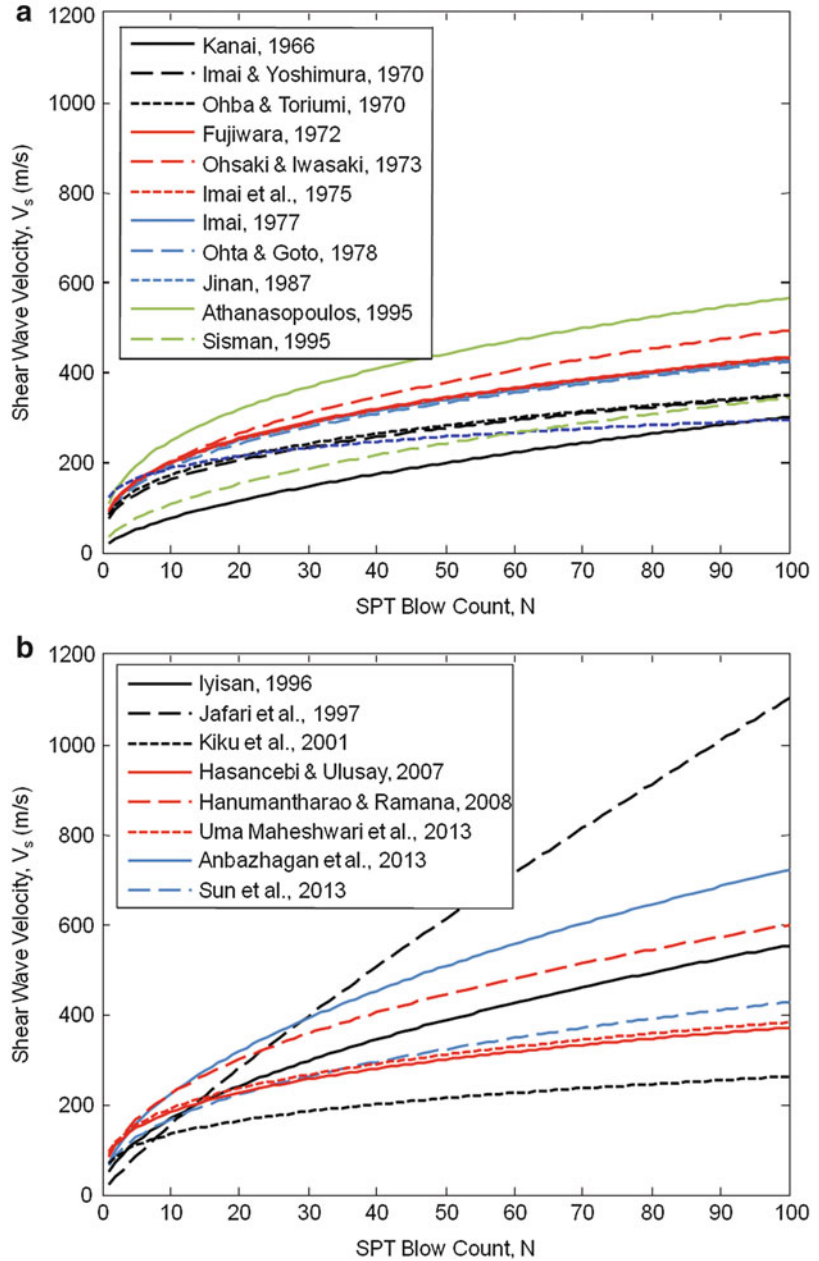
N - V_s in Taipei. Athanasopoulos (1995) developed his correlation based on 221 data pairs from Greek soils. Kayabali (1996) proposed a relationship based on 30 alluvial silty sand deposits from an unspecified location. Pitilakis et al.'s (1999) correlation was developed using 176 pairs of N and V_s data points from the

EURO-SEISTEST database gathered from northern Greece near Thessaloniki.

Wride et al. (2000) proposed a correlation based on results from experiments conducted as part of the Canadian Liquefaction Experiment (CANLEX) at six sand sites in Western Canada. The range of data shown in their equation

**Dynamic Soil Properties:
In Situ Characterization
Using Penetration Tests,**

Fig. 12 Shear wave velocity versus SPT- N correlations for all soils: (a) correlations developed prior to 1996, (b) more recent correlations



represents the minimum and maximum site-specific values. Jafari et al. (2002) used 57 spectral analyses of surface waves (SASW) tests, 9 down-hole tests, and 26 SPT boreholes to create a correlation for south Tehran. The database used in Andrus et al.'s (2004) correlation came from published work performed in 21 soil layers from eight sites in California, one site in South

Carolina, six sites in Canada, and eight sites in Japan. Hasancebi and Ulusay (2007) correlated 97 pairs of N and V_s measurements from Yenisehir, Turkey. Using V_s measurements from multichannel analysis of surface waves (MASW) testing points adjacent to 38 SPT borehole locations, 162 data pairs were used by Anbazhagan and Sitharam (2008) to develop an N - V_s

relationship for the Bangalore Region. An N - V_s relationship for Delhi was proposed by Hanumantharao and Ramana (2008) based on a subset of SASW-SPT data compiled by the city government from public and private entities.

Uma Maheshwari et al. (2010) used MASW to determine V_s and compared the results to 200 SPTs in order to create their correlation for the Chennai region. Anbazhagan et al. (2013) correlated results from 11 SPTs to MASW results in Lucknow City, India. A correlation for Korean soils was developed by Sun et al. (2013) using cross-hole, down-hole, and up-hole tests from 26 sites. Some references listed in Table 1 were published in languages other than English (Imai and Yoshimura 1970; Shibata 1970; Ohta et al. 1972; Imai et al. 1975; Sisman 1995; Iyisan 1996; Jafari et al. 1997) and are, therefore, not explained in detail.

Other Factors Influencing the SPT Correlations

As shown in Table 1, the majority of the empirical SPT- V_s correlations are in the form of a power fit:

$$V_s = XN^Y \quad (9)$$

This is logical since both V_s and N_{60} depend on confining stress. However, recent research has shown that σ'_v may need to be added explicitly to the empirical N - V_s relationship (Schnaid et al. 2004; Brandenburg et al. 2010). Schnaid et al. (2004) incorporated σ'_v into the following relationships between G_0 and N_{60} :

$$\begin{aligned} G_0 &= 1200(N_{60}\sigma'_{v0}p_a^2)^{1/3} \\ \text{upper bound : cemented sands} \\ G_0 &= 450(N_{60}\sigma'_{v0}p_a^2)^{1/3} \\ \text{lower bound : cemented sands and upper bound :} \\ &\text{uncemented sands} \\ G_0 &= 200(N_{60}\sigma'_{v0}p_a^2)^{1/3} \\ \text{lower bound : uncemented sands} \end{aligned} \quad (10)$$

where p_a = atmospheric pressure.

Additionally, geologic age has been shown to influence the N - V_s relationship (Ohsaki and Iwasaki 1973; Lee 1991). It is often difficult to

accurately determine the geologic age of soil deposits. Therefore, while this factor influences the correlation, it is difficult to apply in practice. When determining which relationship to use in practice, the authors recommend matching the geologic age of the deposit of interest to the geologic age of deposit used to develop the correlation.

In summary, while the SPT is a large strain test, correlations have been developed for over four decades relating corrected and uncorrected blow counts to small strain soil properties, such as V_s and G_0 . Local experience has demonstrated the utility of some of the relationships. However, caution should be used in applying these correlations outside of the conditions for which they were developed.

Dynamic Soil Properties from CPT

The relationships between V_s , G_0 , E_0 , and M_0 were presented in the SPT section of this chapter. The research summarized here compares CPT results to various soil moduli, but other properties of interest can be determined from the relationships above. In the section describing the CPT equipment, procedures, and results, q_c , f_s , and u were introduced. Some relationships described in this section use dimensionless, normalized cone parameters. The normalized cone tip resistance, Q_t ; normalized friction ratio, F_N ; and normalized pore pressure, B_q , were proposed by Robertson (1990) and are given by

$$Q_t = \frac{q_c - \sigma_{v0}}{\sigma'_{v0}} \quad (11)$$

$$F_N = \frac{f_s}{q_t - \sigma_{v0}} * 100\% \quad (12)$$

$$B_q = \frac{\Delta u}{q_t - \sigma_{v0}} \quad (13)$$

where:

σ_{v0} = total stress.

Δu = the difference between measured pore pressure and hydrostatic pore pressure.

Additionally, some of the CPT-based correlations depend on the soil behavior type index, I_c , (Robertson and Wride 1998):

$$I_c = \left[(3.47 - \log Q_t)^2 + (\log F_N + 1.22)^2 \right]^{0.5} \quad (14)$$

Finally, Robertson et al. (1992) introduced an overburden stress-corrected shear wave velocity, V_{s1} :

$$V_{s1} = V_s * \left(\frac{p_a}{\sigma'_{v0}} \right)^{0.25} \quad (15)$$

Just as with the SPT, the following correlations between CPT parameters and small strain dynamic properties are best utilized under the same soil conditions under which they were developed. Relationships between CPT and dynamic soil properties are presented in Table 2.

Baldi et al. (1989) proposed a V_s - G_0 relationship developed with Po River Sand and is meant for uncemented, silica sand. Rix and Stokoe (1992) updated the Baldi et al. (1989) relationship between G_0 and V_s by including laboratory cone penetration tests in calibration chambers:

$$\frac{G_0}{q_c} = 1634 \left(\frac{q_c}{\sigma'_{v0}{}^{0.5}} \right)^{0.75} \quad (16)$$

Hegazy and Mayne (1995) used the general form of equations developed by Mayne and Rix (1993) and replaced initial void ratio, e_0 , and σ'_v and with f_s . This correlation was based on V_s from cross-hole, down-hole, SASW, and the seismic cone penetration tests (SCPT) from 30 sites in the United States, Canada, Norway, and Italy. Mayne and Rix (1995) compiled data from cross-hole, down-hole, SASW, and SCPTs conducted at 31 clay sites in Europe and North America and proposed a V_s - q_c relationship.

Simonini and Cola (2000) reported SCPT results from the Venice Lagoon. After investigating several relationships, the best fit relationship included the B_q pore pressure parameter:

$$G_0 = 21.5 q_c^{0.79} (1 + B_q)^{4.59} \quad (17)$$

Wride et al.'s (2000) relationship was based on tests conducted as part of CANLEX at six sites in Western Canada with loose sand deposits. The denominator in the equation is listed as a range of values because Wride et al. (2000) presented a different value for each of the six sites. The database used in Andrus et al.'s (2004) correlation came from published work from eight sites in California, seven South Carolina, six from Canada, and four from Japan. Schnaid et al. (2004), similar to their N - G_0 relationship, presented a q_c - G_0 relationship based on data from sites in South America, Europe, and the United States:

$$\begin{aligned} G_0 &= 800 (q_c \sigma'_{v0} p_a)^{1/3} \\ &\text{upper bound : cemented sands} \\ G_0 &= 280 (q_c \sigma'_{v0} p_a)^{1/3} \\ &\text{lower bound : cemented sands and upper bound} \\ &\text{: uncemented sands} \\ G_0 &= 110 (q_c \sigma'_{v0} p_a)^{1/3} \\ &\text{lower bound : uncemented sands} \end{aligned} \quad (18)$$

Hegazy and Mayne (2006) added 12 new sites to databases assembled by Mayne and Rix (1993) and Hegazy and Mayne (1995) in order to develop their correlation. Andrus et al. (2007) used 229 data pairs from sites in California, South Carolina, and Japan.

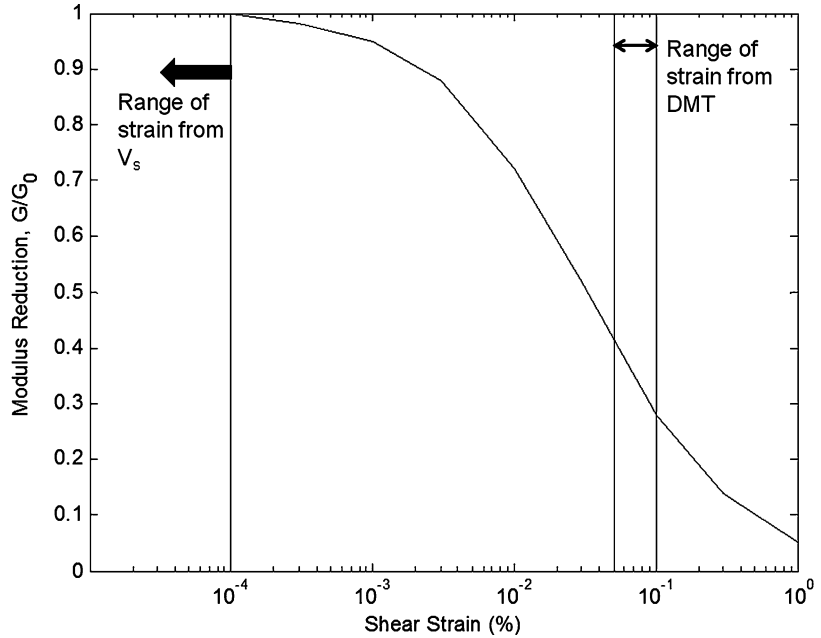
Robertson (2009) utilized a database of more than 100 SCPTs and over 1,000 data pairs in his correlation. Stuedlein's (2010) CPT- V_s correlation is based on 12 SCPTs, yielding 254 data pairs, in the Puyallup River delta near the port of Tacoma, Washington. SCPTs from 17 sites in five areas along the western coast of South Korea were used to develop Sun et al.'s (2013) correlation.

McGann et al. (2014) compared predicted V_s from Andrus et al. (2007), Hegazy and Mayne (2006), and Robertson (2009) to V_s measured using the SCPT following the 2010–2011 Canterbury earthquakes. They found that the relationships over-predicted V_s and recommend future research to develop a Canterbury-specific CPT- V_s relationship. McGann et al. (2014) note that

Dynamic Soil Properties: In Situ Characterization Using Penetration Tests, Table 2 Summary of correlations between CPT data and shear wave velocity

Citation	Shear wave velocity, V_s (m/s)		
	Sand	Clay	All soil
Baldi et al. 1989	$V_s = 277(q_c)^{0.13}(\sigma'_{v0})^{0.27}$ (stress units in MPa)		
Hegazy and Mayne 1995	$V_s = 12.02q_c^{0.319}f_s^{-0.047}$	$V_s = 3.18q_c^{0.549}f_s^{0.025}$	$V_s = (10.1\log q_c - 11.4)^{1.67}\left(\frac{f_s}{q_t} * 100\right)^{0.3}$
Mayne and Rix 1995		$V_s = 1.75(q_c)^{0.627}$	
Wride et al. 2000	$V_{s1} = (95.6 - 110.8)q_{cl}^{0.25}$		
Andrus et al. 2004			$V_{S1cs} = 62.6q_{clN,cs}^{0.231}$
Hegazy and Mayne 2006			$V_s = 0.0831q_{clN}\left(\frac{\sigma'_{v0}}{p_a}\right)^{0.25}e^{1.786f_c}$
Robertson 2009			$V_s = [10^{0.55f_c + 1.68}(q_t - \sigma_v)p_a]^{0.5}$
Stuedlein 2010			$V_s = 0.033Q_t\left(\frac{p_a}{\sigma'_{v0}}\right)^{0.25}e^{2.11f_c}$
Sun et al. 2013	$V_s = [32.159q_t^{0.053}f_s^{0.131}(\sigma'_{v0})^{0.158}(1 + B_q)^{0.169}]$	$V_s = [14.716q_t^{0.168}f_s^{-0.018}(\sigma'_{v0})^{0.246}(1 + B_q)^{-0.126}]$	$V_s = [24.215q_t^{0.091}f_s^{0.037}(\sigma'_{v0})^{0.232}(1 + B_q)^{-0.181}]$

**Dynamic Soil Properties:
In Situ Characterization
Using Penetration Tests,
Fig. 13** Degradation of
shear modulus with shear
strain



the relationships particularly over-predict V_s at shallow depths due to a scarcity of shallow depth data in the examined correlations.

Other Factors Influencing the CPT Correlations

While most SPT- V_s correlations were based solely on N , the CPT- V_s correlations sometimes included tip resistance, side friction, pore pressure, and effective stress. However, there are yet other factors which influence the relationship between CPT results and dynamic soil properties. Projects in which these factors were investigated are summarized below.

The age of the deposit plays a significant role on the relationship between CPT results and V_s (Andrus et al. 2007). Indeed, small strain properties, such as V_s , have been shown to increase with time after deposition more rapidly than penetration resistance does (Andrus et al. 2009). Therefore, when applying correlations between CPT and dynamic soil properties, it is important to match the deposit's geologic age to the geologic age of the sites used to develop the correlation.

Cementation is another factor that affects V_s . Similar to the effects of geologic age, V_s is more

sensitive to cementation than penetration resistances are. Schnaid et al.'s (2004) CPT- G_0 correlation (Eq. 18) shows that the upper bound prediction of G_0 in uncemented soils is equal to the lower bound prediction in cemented soils.

It is well known that a soil's V_s is influenced by the average number of particle contacts per grain as these transmit the seismic waves. Since the number of contacts is impossible to determine, void ratio, e_0 , is used as its surrogate. Therefore Mayne and Rix (1993) proposed this relationship for predicting V_s based on CPT results:

$$G_0 = 99.5 p_a^{0.305} q_c^{0.695} e_0^{1.130} \quad (19)$$

The difficulty in applying this relationship in practice is that in situ e_0 is also difficult to predict. Rix and Stokoe (1992), Hegazy and Mayne (1995), and Mayne and Rix (1995) also discuss the importance of e_0 in CPT- V_s correlations.

As shown in Fig. 13, soil modulus degrades with strain. Fahey and Carter (1993) proposed this simple approach to estimating degradation of G :

Dynamic Soil Properties: In Situ Characterization Using Penetration Tests,

Fig. 14 Striking the pad of the CPT rig to generate a shear wave in an SCPT



$$\frac{G}{G_0} = 1 - f \left(\frac{q}{q_{ult}} \right)^g \quad (20)$$

where:

f and g = empirical constants depending on soil type and stress history.

q = applied load causing modulus reduction.

q_{ult} = failure load.

Mayne (2007) suggested that f equals 1 and g is approximately 0.3 ± 0.1 . Mayne (2010) demonstrated the utility of Eq. 20 by using data from laboratory tests and finding good agreement to results from case histories given proper selection of f and g . Combining V_s data from in situ tests with laboratory tests results provides an estimate of G/G_0 . However, there is currently no method for obtaining this relationship entirely from CPT data.

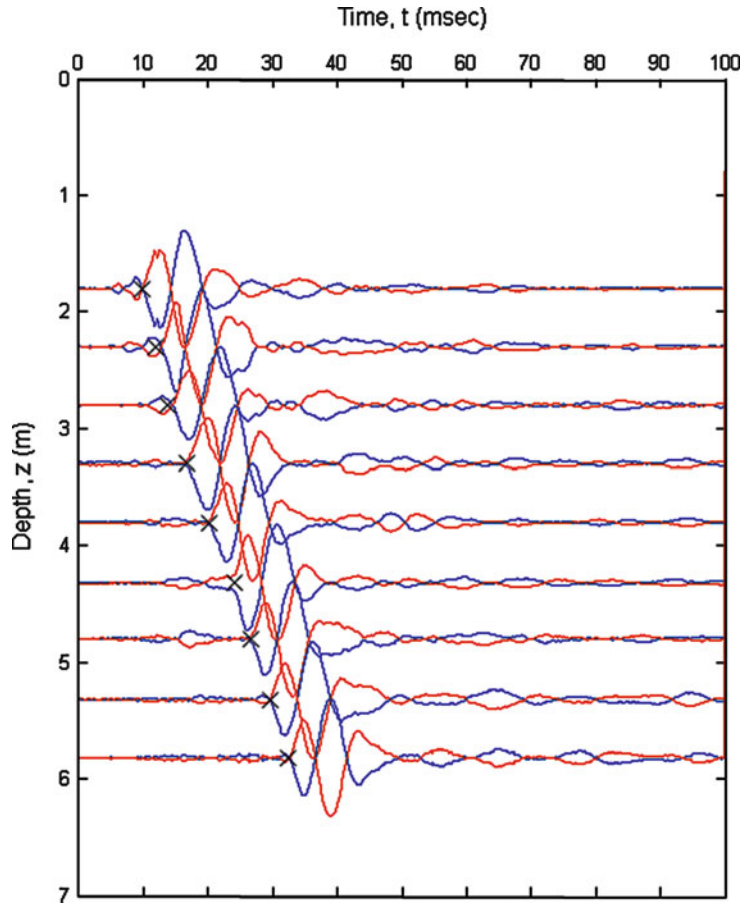
Seismic Cone Penetration Test

The seismic cone penetration test (SCPT) allows for direct measurement of dynamic soil properties rather than through empirical

correlations through the basic CPT readings. An accelerometer placed in the CPT probe is used to record the arrival times of shear waves generated at the ground surface as shown in Fig. 14. A wired sledge hammer completes an electrical circuit when it strikes a support pad of the CPT rig. This triggers the beginning of the data acquisition system's recording and sends the seismic signal into the ground. The arrival of the shear wave at the accelerometer is marked by the onset of a large pulse in the signal trace. The probe is then advanced to the next test depth, and the process is repeated. Common practice is to conduct seismic tests at CPT rod changes (i.e., 1.0 m intervals). However, the authors find that use of 1.0 m intervals does not define soil stratigraphy very well and therefore should not be used for detailed site characterization or liquefaction hazard assessment. However, a 1.0 m interval does provide a reasonable estimate of average dynamic properties and is therefore very reliable for earthquake site response analysis.

An example of SCPT data is presented in Fig. 15. The first step in interpreting the data is

**Dynamic Soil Properties:
In Situ Characterization
Using Penetration Tests,**
Fig. 15 Wave recordings
versus accelerometer depth
in a seismic CPT



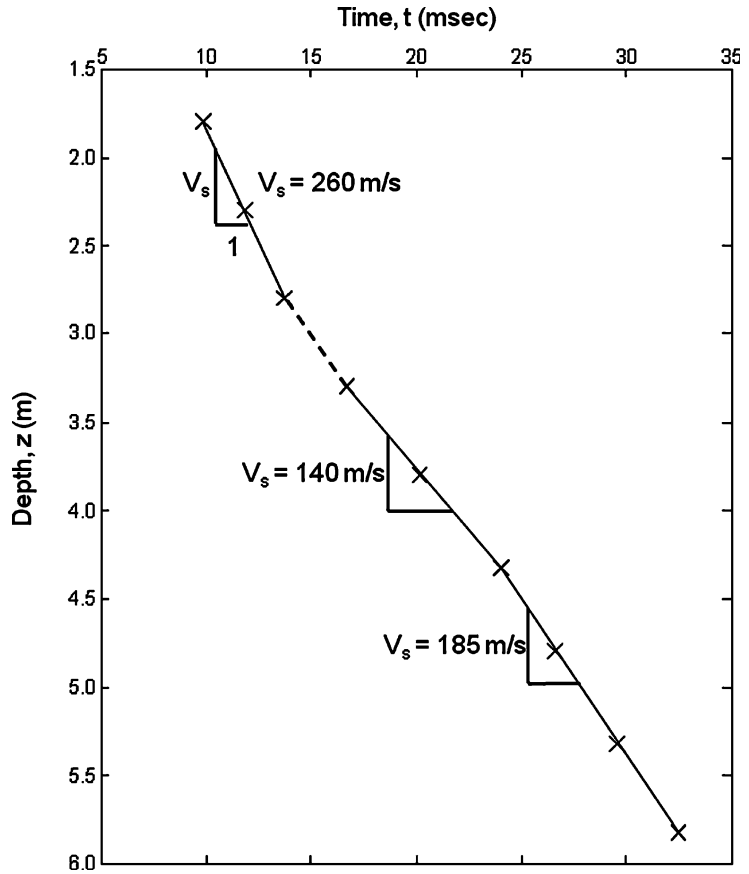
to determine the arrival time of the shear wave. Campenella and Stewart (1992) and Brignoli et al. (1996) explain methods of picking shear wave arrival times. The authors use the first intense energy excursion, as described by Brignoli et al. (1996). The determined shear wave arrival times are plotted versus accelerometer depth as shown in Fig. 16. When more than two data points line up, there is reasonable certainty that the elevation increment can be represented by constant V_s . However, this does not necessarily mean that the soil over this increment is the same; it merely shows that the shear wave velocity is constant. By contrast, a break in the V_s profile generally means a change in soil conditions. Figure 16 shows some uncertainty in the layer interface between a depth of 2.8 and

3.3 m. It could also mean that the soil between 2.8 and 3.3 m is different from the soils above and below. This illustrates the need for smaller test intervals if better stratigraphic resolution is desired.

Determining dynamic soil properties directly from measured shear wave velocities such as by the SCPT is obviously more accurate than estimates of V_s based on correlations with CPT resistances. Additionally, research has shown excellent agreement between V_s measured using the SCPT with other methods of measuring V_s , such as down-hole tests, cross-hole tests, SASW, and laboratory measurements (Robertson et al. 1992; Lunne et al. 1997; Schneider et al. 1999, 2001; Mayne 2000; Saftner et al. 2011; etc.).

**Dynamic Soil Properties:
In Situ Characterization
Using Penetration Tests,**

Fig. 16 Shear wave arrivals versus depth and identification of depth intervals with constant shear wave velocity



D

Dynamic Soil Properties from DMT

The first correlations between DMT results and dynamic soil properties used E_D . Baldi et al. (1986) conducted resonant column tests on Ticino sand and compared the results to DMT results from a calibration chamber to develop Eq. 21.

$$\frac{G_0}{E_D} = 2.72 \pm 0.59 \quad (21)$$

Bellotti et al. (1986) used cross-hole testing in Po River sand to determine the correlation presented in Eq. 22.

$$\frac{G_0}{E_D} = 2.2 \pm 0.7 \quad (22)$$

Jamiolkowski et al. (1988) conducted experiments using Ticino sand and found that relative

density, D_r , played a significant role in relating E_D to G_0 , suggesting that previous relationships were too simple.

Hryciw (1990) presented a method of predicting G_0 by the DMT based on results from nine sites in Europe and North America. Hryciw's (1990) relationship included the confining stresses and soil unit weights determined from DMT indices:

$$G_0 = 530 \left(\frac{\sigma'_v}{p_a} \right)^{-0.25} \frac{\frac{\gamma_D}{\gamma_w} - 1}{2.7 - \frac{\gamma_D}{\gamma_w}} K_0^{0.25} (\sigma'_v p_a)^{0.5} \quad (23)$$

where:

γ_D = unit weight as determined by DMT indices.
 γ_w = unit weight of water.

Monaco et al. (2009) collected data from 34 sites in Europe and the United States. The following equations provided estimates of G_0 :

$$\begin{aligned} \text{Clay} \quad \frac{G_0}{M_{DMT}} &= 26.177K_D^{-1.0066} \\ \text{Silt} \quad \frac{G_0}{M_{DMT}} &= 15.686K_D^{-0.921} \\ \text{Sand} \quad \frac{G_0}{M_{DMT}} &= 4.5613K_D^{-0.7967} \end{aligned} \quad (24)$$

where M_{DMT} = constrained dilatometer modulus.

Marchetti (1980) suggested that M_{DMT} can be obtained through the DMT modulus, E_D by

$$M_{DMT} = R_M E_D \quad (25)$$

In turn, R_M is a function of the DMT material index, I_D , and horizontal stress index, K_D :

$$\begin{aligned} \text{if } I_D \leq 0.6 \quad R_M &= 0.14 + 2.36 \log K_D \\ \text{if } I_D \geq 3 \quad R_M &= 0.5 + 2 \log K_D \\ \text{if } 0.3 < I_D < 0.6 \quad R_M &= R_{M0} + (2.5 - R_{M0}) \log K_D \\ &\quad \text{with } R_{M0} = 0.14 + 0.15(I_D - 0.6) \\ \text{if } K_D > 10 \quad R_M &= 0.32 + 2.18 \log K_D \\ \text{if } R_M < 0.85 \quad R_M &= 0.85 \end{aligned} \quad (26)$$

Numerous case histories (i.e., Lacasse 1986; Schmertmann 1986; Schnaid et al. 2000; etc.) have demonstrated that the DMT gives a reliable estimate of M .

Seismic Dilatometer Test

The seismic dilatometer test (SDMT) is similar to the SCPT. The descriptions of equipment, test procedure, and data reduction presented for the SCPT are essentially the same for the SDMT. The advantage of performing SDMT is that one can obtain dynamic soil properties, in this case, V_s , directly, without the uncertainty of correlations. Monaco et al. (2009) emphasized that the availability of SDMT makes direct measurement of V_s the best method for determining dynamic properties from the DMT.

Whereas the SPT and CPT are large strain tests (~10 %), the DMT is a medium strain test (0.05–0.1 %) as shown in Fig. 13. Therefore, there is also an opportunity to develop a method

of determining G/G_0 degradation with strain amplitude (γ) by the DMT. Marchetti et al. (2008) suggest a path for future research, using two points, G (or G_{DMT}) as determined from M_{DMT} and G_0 as determined from V_s to construct the G/G_0 curve. Suggestions for methods of accomplishing this include using I_D to determine soil type and comparing G_{DMT} to G - γ curves recommended for the given soil type, as well as by comparing settlement predictions based on M_{DMT} to predictions from G - γ curves with variable decay rates to determine the best match. Of the three in situ penetration tests discussed in this chapter, the DMT appears to have the most potential for developing G/G_0 - γ relationships from in situ penetration tests.

Summary

Numerous empirical correlations exist between geotechnical in situ penetration test results and dynamic soil properties. These relationships have been developed in various locations around the world and are best applied to sites geologically similar to those at which the correlations were developed. In situ penetration tests that measure V_s directly, such as the SCPT and SDMT, allow determination of small strain dynamic soil properties more accurately than empirical relationships based on the basic SPT-N, CPT- q_c , or DMT indices.

Cross-References

- [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)
- [Seismic Design of Dams](#)
- [Seismic Design of Earth-Retaining Structures](#)
- [Seismic Design of Tunnels](#)

References

- Anbazhagan P, Sitharam TG (2008) Mapping of average shear wave velocity for Bangalore Region: a case history. *J Environ Eng Geophys* 13(2):69–84

- Anbazhagan P, Kumar A, Sitharam TG (2013) Seismic site classification and correlation between standard penetration test N value and shear wave velocity for Lucknow City in Indo-Gangetic Basin. *Pure Appl Geophys* 170:299–318
- Andrus RD, Piratheepan P, Ellis B, Zhang J, Juang H (2004) Comparing liquefaction evaluation methods using penetration- V_s relationships. *Soil Dyn Earthq Eng* 24:713–721
- Andrus RD, Mohanan NP, Piratheepan P, Ellis BS, Holzer TL (2007) Predicting shear wave velocity from cone penetration resistance. In: *Proceedings of the 4th international conference on earthquake geotechnical engineering, The ssaloniki, Paper no 1454*
- Andrus RD, Hayati H, Mohanan NP (2009) Correcting liquefaction resistance for aged soils using measured to estimated velocity ratio. *J Geotech Geoenviron Eng* 135(6):735–744
- Athanasopoulos GA (1995) Empirical correlation V_s -N for soils of Greece: a comparative study of reliability. In: *Proceedings of the 7th international conference on soil dynamics & earthquake engineering, Crete, pp 19–36*
- Baldi G et al (1986) Flat dilatometer tests in calibration chambers. In: *Proceedings of the use of in situ tests in geotechnical engineering, Blacksburg, pp 47–73*
- Baldi G, Bellotti R, Ghionna V, Jamiolkowski M, Lo Presit DCF (1989) Modulus of sands from CPTs and DMTs. In: *Proceedings of the XII ICSMFE, vol 1, Rio de Janeiro, pp 165–170*
- Bellotti R, Ghionna V, Jamiolkowski M, Lancellotta R, Manfredini G (1986) Deformation characteristics of cohesionless soils from in situ tests. In: *In situ '86 specialty conference in situ. ASCE, Virginia Tech, Blacksburg*
- Brandenberg SJ, Bellana N, Shantz T (2010) Shear wave velocity as function of standard penetration test resistance and vertical effective stress at California bridge sites. *Soil Dyn Earthq Eng* 30:1026–1035
- Brignoli EG, Gotti M, Stokoe KH (1996) Measurement of shear waves in laboratory specimens by means of piezo-electric transducers. *Geotech Test J* 19(4):384–397
- Campenella RG, Stewart WP (1992) Seismic cone analysis using digital signal processing for dynamic site characterization. *Can Geotech J* 29(3):477–486
- Campenella RG et al (1985) Recent developments in in situ testing of soils. In: *Proceedings of the XI international conference on soil mechanics and foundation engineering, San Francisco, pp 849–854*
- Chein LK, Lin MC, Oh YN (2000) Shear wave velocity and SPT-N values of in situ reclaimed soil in west Taiwan. *Geotech Eng* 31:63–77
- Fahey M, Carter JP (1993) A finite element study of the pressuremeter test in sand using a nonlinear elastic plastic model. *Can Geotech J* 30(2):348–362
- Fujiwara T (1972) Estimation of ground movement in actual destructive earthquakes. In: *Proceedings of the 4th European symposium on earthquake engineering, London, pp 125–132*
- Fumal TE, Tinsley JC (1985) Mapping shear wave velocities of near surface geological materials. Predicting areal limits of earthquake induced landsliding. In: *Evaluation of earthquake hazards in the Los Angeles region, USGS aper 1360, pp 127–150*
- Hanumantharao C, Ramana GV (2008) Dynamic soil properties for microzonation of Delhi, India. *J Earth Syst Sci* 117(S2):719–730
- Hasancebi N, Ulusay R (2007) Empirical correlation between shear wave velocity and penetration resistance for ground shaking assessment. *Bull Eng Geol Environ* 66:203–213
- Hegazy YA, Mayne PW (1995) Statistical correlations between V_s and cone penetration data for different soil types. In: *Proceedings, CPT '95, vol 2. Swedish Geotechnical Society, Linköping, pp 173–178*
- Hegazy YA, Mayne PW (2006) A global statistical correlation between shear wave velocity and cone penetration data. In: *Proceedings of the GeoShanghai, site and geomaterial characterization (GSP 149), ASCE, Reston, pp 243–248*
- Hryciw RD (1990) Small strain shear modulus of soil by dilatometer. *J Geotech Eng* 116(11):1700–1716
- Imai T (1977) P and S wave velocities of the ground in Japan. In: *Proceedings of the 9th international conference on soil mechanics & foundation engineering, vol 2, Toyko, pp 127–132*
- Imai T, Tonouchi K (1982) Correlation of N value with S wave velocity and shear modulus. In: *Proceedings of the 2nd European symposium on penetration testing, Amsterdam, pp 57–72*
- Imai T, Yoshimura Y (1970) Elastic wave velocity and soil properties in soft soil. *Tsuchito-Kiso* 18(1):17–22. As cited in Hasancebi N, Ulusay R (2007) Empirical correlation between shear wave velocity and penetration resistance for ground shaking assessment. *Bull Eng Geol Environ* 66:203–213
- Imai T, Fumoto H, Yokota K (1975) The relation of mechanical properties of soil to P and S wave velocities in Japan. In: *Proceedings of the 4th Japan earthquake engineering symposium, Toyko, pp 89–96. As cited in Uma Maheshwari R, Boominathan A, Dodagoudar GR (2010) Use of surface waves in statistical correlations of shear wave velocity and penetration resistance of Chennai soils. Geotech Geol Eng* 28:119–137
- Iyisan R (1996) Correlation between shear wave velocity and in situ penetration test results. *Tech J Chamb Civ Eng Turk* 7:1187–1199. As cited in Anbazhagan P, Kumar A, Sitharam TG (2013) Seismic site classification and correlation between standard penetration test N value and shear wave velocity for Lucknow City in Indo-Gangetic Basin. *Pure Appl Geophys* 170:299–318
- Jafari MK, Asghari A, Rahmani I (1997) Empirical correlation between shear wave velocity and SPT-N values for south of Tehran soils. In: *Proceedings of the 4th international conference on civil engineering, Tehran.*

- As cited in Uma Maheshwari R, Boominathan A, Dodagoudar GR (2010) Use of surface waves in statistical correlations of shear wave velocity and penetration resistance of Chennai soils. *Geotech Geol Eng* 28:119–137
- Jafari MK, Shafiee A, Ramzkhah A (2002) Dynamic properties of the fine grained soils in south of Tehran. *J Seismol Earthq Eng* 4(1):25–35
- Jamiolkowski M, Ghionna V, Lancellotta R, Paqualini E (1988) New correlations of penetration tests for design practice. In: Proceedings of the international symposium of penetration testing, vol 1, Orlando, pp 263–296
- Japanese Road Association (1980) Part V: earthquake resistant design. Specification for highway bridges, Nihon Doro Kyokai, Tokyo
- Jinan Z (1987) Correlation between seismic wave velocity and the number of blows of SPT and depth. Selected papers from the Chinese *J Geotech Eng*, ASCE, pp 92–100
- Kanai K, Tanaka Y, Morishita T, Osada K (1966) Observation of microtremors. XI: Matsushiro earthquake swarm area. *Bull Earthq Res Inst* 44:1297–1333
- Kayabali K (1996) Soil liquefaction evaluation using shear wave velocity. *Eng Geol* 44(1):121–127
- Lacasse S (1986) In situ site investigation techniques and interpretation for offshore practice, report 40019-28 Sept. Norwegian Geotechnical Institute, Oslo
- Lee SHH (1990) Regression models of shear wave velocities. *J Chin Inst Eng* 13:519–532
- Lunne T, Robertson PK, Powell JJM (1997) Cone penetration testing in geotechnical practice. Blackie Academic & Professional, London, 312 pp
- Marchetti S (1980) In situ tests by flat dilatometer. *J Geotech Eng* 106(3):299–321
- Marchetti S, Monaco P, Totani G, Marchetti D (2008) In situ tests by seismic dilatometer. In: From research to practice in geotechnical engineering. Geotechnical special publication 170, ASCE, Reston
- Mayne PW, Rix GJ (1995) Correlations between shear wave velocity and cone tip resistance in natural clays. *Soils Found* 35(2):193–194
- Mayne PW (2000) Enhanced geotechnical site characterization by seismic piezocone penetration tests. In: Proceedings of the 4th international geotechnical conference, Cairo, pp 95–120
- Mayne PW (2007) NCHRP synthesis 368: cone penetration testing. Transportation Research Board, Washington, DC, 118 pp
- Mayne PW (2010) Soil parameter evaluation from hybrid in situ penetration geophysics testing. In: *GeoFlorida 2010*. Geotechnical special publication 199, ASCE, Reston, pp 1049–1058
- Mayne PW, Rix GJ (1993) $G_{\max}$ - q_c relationships for clays. *Geotech Test J* 16(1):54–60
- McGann CR, Bradley BA, Cubrinovski M, Taylor ML (2014) Comparison of existing CPT-Vs correlations with Canterbury-specific seismic CPT data. In: 2014 New Zealand Society for Earthquake Engineering conference, Quincy, Auckland, paper O13
- Monaco P, Marchetti S, Totani G, Marchetti D (2009) Interrelationship between small strain modulus G_0 and operative modulus. In: International conference on performance based design in earthquake geotechnical engineering, Tokyo
- Ohba S, Toriuma I (1970) Dynamic response characteristics of Osaka Plain. In: Proceedings of the annual meeting AIJ, Tokyo
- Ohsaki Y, Iwasaki Y (1973) Dynamic shear moduli and Poisson's ratio of soil deposits. *Soils Found* 13:61–73
- Ohta Y, Goto N (1978) Empirical shear wave velocity equations in terms of characteristic soil indexes. *Earthq Eng Struct Dyn* 6(2):167–187
- Ohta T, Goto N, Kagami H, Shiono K (1972) Shear wave velocity measurement during a standard penetration test. *Earthq Eng Struct Dyn* 6:43–50. As cited in Hanumantharao C, Ramana GV (2008) Dynamic soil properties for microzonation of Delhi, India. *J Earth Syst Sci* 117(S2):719–730
- Pitilakis K, Raptakis D, Lontzetidis K, Tika-Vassilikou T, Jongmans D (1999) Geotechnical and geophysical description of euro-seistest using field and laboratory tests and moderate strong ground motions. *J Earthq Eng* 3(3):381–409
- Rix GJ, Stokoe KH (1992) Correlation of initial tangent modulus and cone resistance. In: International symposium on calibration chamber testing, Potsdam, pp 351–362
- Robertson PK (1990) Soil classification using the cone penetration test. *Can Geotech J* 27(1):151–158
- Robertson PK (2009) Interpretation of cone penetration tests – a unified approach. *Can Geotech J* 46(11):1337–1355
- Robertson PK, Wride CE (1998) Cyclic liquefaction and its evaluation based on the SPT and CPT, NCEER-97-0022. In: Proceedings of the NCEER workshop on evaluation of liquefaction resistance of soils, Salt Lake City, pp 41–87
- Robertson PK, Woeller DJ, Addo KO (1992) Standard penetration test energy using a system based on the personal computer. *Can Geotech J* 29:551–557
- Saftner DA, Jung Y, Hryciw RH, Green RA (2011) Comparison of predicted cyclic resistance ratios from CPT, DMT, and shear wave velocity measurements in Griffin, Indiana. In: *Geo-Frontiers 2011: advances in geotechnical engineering*, Geotechnical special publication 211. American Society of Civil Engineers, Reston, pp 2298–2307
- Schmertmann JH (1986) Dilatometer to compute foundation settlement. In: Proceedings of in situ '86, a specialty conference on use of In Situ Tests in Geotechnical Engineering, Geotechnical Special Publication no 6. ASCE, Reston, pp 303–321
- Schnaid F, Kratz de Oliveira LA, Gehling WYY (2000) Use of pressuremeter tests in the assessment of soil collapse. *Soils Rocks* 3(1):12, Brazil
- Schnaid F, Lehane BM, Fahey M (2004) In situ test characterization of unusual geomaterials. In: Proceedings of the 2nd international conference on site characterization, vol 1, Porto, pp 49–74

- Schneider JA, Hoyos L, Mayne PW, Macari EJ, Rix GJ (1999) Field and laboratory measurement of dynamic shear modulus of piedmont residual soils. In: Behavioral characteristics of residual soils. Geotechnical special publication 92, ASCE, Reston, pp 12–25
- Schneider JA, Mayne PW, Rix GJ (2001) Geotechnical site characterization in the greater Memphis area using cone penetration tests. *Eng Geol* 62:169–184
- Seed HB, Idriss IM (1981) Evaluation of liquefaction potential sand deposits based on observation of performance in previous earthquakes. In situ testing to evaluate liquefaction susceptibility, pp 81–544
- Seed HB, Idriss IM, Arango I (1983) Evaluation of liquefaction potential using field performance data. *J Geotech Eng* 109:458–482
- Seed HB, Tokumatsu K, Harder LF, Chung R (1985) Influence of SPT procedures in soil liquefaction resistance evaluations. *J Geotech Eng* 111(12):1425–1445
- Shibata T (1970) Analysis of liquefaction of saturated sand during cyclic loading, disaster prevention. *Res Inst Bull* 13:563–570. As cited in Uma Maheshwari R, Boominathan A, Dodagoudar GR (2010) Use of surface waves in statistical correlations of shear wave velocity and penetration resistance of Chennai soils. *Geotech Geol Eng* 28:119–137
- Simonini P, Cola S (2000) Use of piezocone to predict maximum stiffness of Venetian soils. *J Geotech Geoenviron Eng* 126(4):378–381
- Sisman H (1995) An investigation on relationships between shear wave velocity and SPT and pressuremeter test results. MS Thesis, Ankara University, Geophysical Engineering Department. As cited in Hasancebi N, Ulusay R (2007) Empirical correlation between shear wave velocity and penetration resistance for ground shaking assessment. *Bull Eng Geol Environ* 66:203–213
- Skempton AW (1986) Standard penetration test procedures and effects in sands of overburden pressure, relative density, particle size, aging and over consolidation. *Geotechnique* 36(3):425–447
- Studelein AW (2010) Shear wave velocity correlation for Puyallup River Alluvium. *J Geotech Geoenviron Eng ASCE* 136(9):1298–1304
- Sun CG, Cho CS, Minkyung S, Shin JS (2013) Correlations between shear wave velocity and in situ penetration test results for Korean Soil Deposits. *Pure Appl Geophys* 170:271–281
- Sykora DW, Stokoe KH (1983) Correlations of in situ measurements in sands with shear wave velocity. Geotechnical Engineering Report GR83-33, University of Texas at Austin, Austin. As cited in Sykora DW (1987) Examination of existing shear wave velocity and shear modulus correlations in soils. Miscellaneous Paper GL-87-26, US Army Corps of Engineers, WES, Vicksburg
- Uma Maheshwari R, Boominathan A, Dodagoudar GR (2010) Use of surface waves in statistical correlations of shear wave velocity and penetration resistance of Chennai soils. *Geotech Geol Eng* 28:119–137
- Wride CE, Roertson PK, Biggar KW, Campanella RG, Hofmann BA, Hughes JMO, Kupper A, Woeller DJ (2000) Interpretation of in situ test results from the Canlex sites. *Can Geotech J* 37(3):505–529

Early Earthquake Warning (EEW) System: Overview

Tsuneo Katayama¹ and Shigeki Horiuchi²

¹Professor emeritus of the University of Tokyo, Bunkyo-ku, Tokyo, Japan

²COE, Home Seismometer Corporation, Shirakawa, Fukushima, Japan

Synonyms

Earthquake alarm system (ElarmS); Earthquake alert system; Earthquake early alarm system; Nakamura's method; Real-time earthquake warning; Urgent earthquake detection and alarm system (UrEDAS)

Introduction

Earthquake early warning (EEW or Real-time earthquake warning) is a system to provide warning of potentially significant shaking in neighboring regions of a strong earthquake within a few seconds upon detection of a severe earthquake. It differs from earthquake prediction which is defined as the specification of the time, location, and magnitude of a future earthquake. EEW is aimed at mitigating earthquake-related damage by allowing countermeasures such as promptly slowing down running trains, controlling

elevators, and enabling people to quickly protect themselves in various environments. Operational EEW systems are currently running in Mexico, Taiwan, and Japan. For well over one decade the Japan Meteorological Agency (JMA) had been developing early warning system to generate immediate alerts after earthquakes, and on October 1, 2007, JMA officially launched EEW service to the residents in Japan through a number of media outlets such as TV and radio. JMA's EEW provides advance announcement of the estimated seismic intensities and expected arrival time of principal motion. The system presently uses about 1,000 seismometers. EEW essentially makes automatic calculation of the (1) focus and (2) magnitude of an earthquake, and estimates, before the arrival of strong motions, (3) the distribution of seismic intensity in the regions where strong tremors are expected. Quickness and robustness are the two key factors. Estimation starts when a few seconds' data from a small number of recording stations are obtained, and the results are updated with time as the number of data increases. The JMA's EEW is divided into two grades: "forecast" and "warning." In the case of the March 11, 2011, off the Pacific coast of Tohoku Earthquake (M_w 9.0), a total of 15 forecasts were issued, and the fourth forecast at 8.6 s after the first trigger became a warning to the general public. When EEW was launched in 2007 in Japan, JMA notified the public the limitations associated with the present EEW,

including (1) Timing, (2) False alarms, (3) Magnitude estimation, and (4) Seismic intensity estimation (JMA 2007). These are all inherent to EEW and difficult to easily solve by the present technology.

There are several ways to define the size of an earthquake. The earthquake magnitude defined by most of the common methods saturates for a large event greater than 8, whereas M_w , or a moment magnitude, is known that such saturation is not observed. In the case of the March 11, 2011, Tsunami earthquake of Japan, JMA released a magnitude as 7.9 soon after the occurrence and updated it to 8.4 during the same day. But the final value released 2 days after the earthquake was M_w 9.0. In this article, M or M_{JMA} , and M_w will be used at different places for different purposes.

Definition

Earthquake early warning (EEW or Real-time earthquake warning) is a system to provide information about potentially significant shaking in neighboring regions of a strong earthquake within a few seconds upon detection of a severe earthquake. EEW differs from earthquake prediction which is defined as the specification of the time, location, and magnitude of a future earthquake. EEW is aimed at mitigating earthquake-related damage by allowing countermeasures such as promptly slowing down running trains, controlling elevators, and enabling people to quickly protect themselves in various environments. When an earthquake occurs, tremors extend out from the seismic focus in a wave-like motion. There are two main types of seismic waves: P -waves with travelling speed of about 7 km/s, and S -waves with travelling speed of about 4 km/s. P -waves are the first to travel outward from the focus and are followed by S -waves, which cause stronger tremors. Most earthquake-induced damage results from S -waves. EEW quickly announces to the public that an earthquake “has” occurred and to inform them of the estimated seismic intensity several (or more) seconds before the arrival of strong tremors.

History

Since 1965, the Japanese National Railway (JNR) has developed and operated an onsite warning system named the Urgent Earthquake Detection and Alarm System (UrEDAS) to control running trains at the time of the occurrence of a large event (Nakamura 1988). Thirty two stations with an average distance of 20 km are used to detect large earthquakes. An alert is issued when the horizontal ground acceleration exceeds 40 cm/s². Although the estimate is rough because the hypocenter is determined by using only the data from one station, the system has nonetheless proven itself effective in estimating the expected ground shaking in the area near each station.

The first prototype of an EEW system in the United States was proposed (Bakun et al. 1994) and developed to rapidly detect the aftershocks of the 1989 Loma Prieta earthquake in California by sending an alert to the crews working in the damaged area. The system worked for 6 months, during which time 19 events with magnitude greater than 3.5 occurred, 12 alerts issued with 2 missed triggers and 1 false alarm.

Operational EEW systems are currently running in Mexico, Taiwan, and Japan. An earthquake alert system in Mexico has been in operation since 1989 that uses 12 stations along the 300 km long Pacific coastline to determine earthquake parameters for large events occurring along the Middle American trench. The main target of the system is Mexico City situated about 300 km south-west of the system (Espinosa-Aranda et al. 1995). Mexican researchers later showed the effectiveness of the warning system. In Taiwan, the Taiwan Central Weather Bureau (CWB) developed an early warning system based on a seismic network consisting of 79 strong motion stations installed across Taiwan and covering an area of 100 ~ 300 km² (Wu and Teng 2002). Since 1995 the network has been used to report event information (location, size, strong motion map) within 1 min after an earthquake occurrence (Teng et al. 1997). To reduce the report time, the concept of a virtual sub-network was

introduced (Wu and Teng 2002). As soon as an event is triggered by at least seven stations, the signals coming from the stations less distant than 60 km from the estimated epicenter are used to characterize the event. The system successfully characterized all the 54 events occurred during a test period of 7 months (December 2000 ~ June 2001), with an average reporting time of 22 s.

For well over one decade the Japan Meteorological Agency (JMA) had been developing early warning system to generate immediate alerts after earthquakes, and on October 1, 2007, JMA officially launched EEW service to the residents in Japan through a number of media outlets such as TV and radio. For the first estimation, JMA introduced an empirical method called B- Δ method (Odaka et al. 2004), which determines epicentral distance from the time changes of *P*-wave amplitudes, measures the direction of hypocenter from the *P*-wave orbital motions and locates hypocenter and magnitude from single station data, though, estimation error is large. During the test period before the official launching (February 2004 ~ July 2006), JMA sent out 855 warnings, only 26 of which were found totally false (Tsukada 2006). JMA's EEW provides advance announcement of the estimated seismic intensities and expected arrival time of principal motion. The system presently uses about 1,000 seismometers, of which about 200 are maintained by JMA and about 800 by the National Research Institute for Earth Science and Disaster Prevention (NIED).

Seismologists in California are currently planning real-time testing of EEW across the state using the ElarmS (Earthquake Alarms Systems) methodology (Allen 2007). The approach uses a network of seismic instruments to detect the first-arriving energy at the surface, the *P*-waves, and translate the information contained in these low amplitude waves into a prediction of the following peak ground shaking.

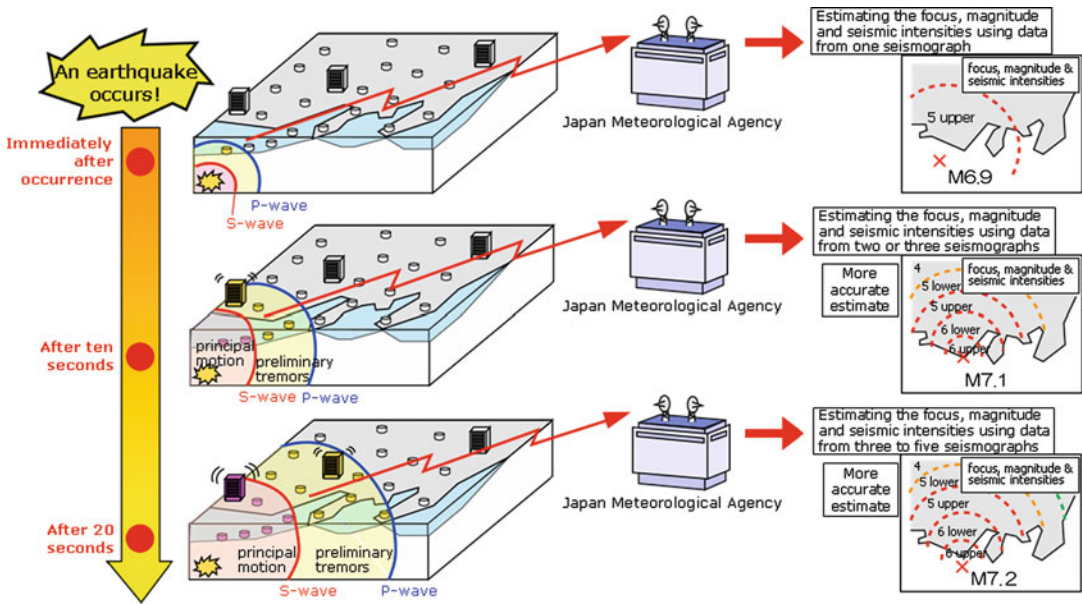
Development and testing of EEWS is being carried out in several other seismically active regions, for example, Bucharest, Cairo, Istanbul and Naples (Zollo et al. 2009).

Basic Principle of EEW

EEW essentially makes automatic calculation of the (1) focus and (2) magnitude of an earthquake, and estimates, before the arrival of strong motions, (3) the distribution of seismic intensity in the regions where strong tremors are expected. Although there are many standard methodologies available for such purposes, they generally use as many records as possible and are not suited for EEW applications. Quickness and robustness are the two key factors. Without quickness a system cannot satisfy the purpose of mitigating damage, and robustness is critical, because the system should be able to determine the earthquake location with a small number of data often with false triggers, picks from other events, and misidentified picks from energetic, secondary phases.

Figure 1 was taken from a public release material by JMA (2007). Computation starts when a few seconds' data from a small number of recording stations and the results are updated with time as the number of data increases. Previous work on earthquake location for early warning includes several approaches to gain constraints on the location at an earlier time and with fewer observations than for standard earthquake location. A single station approach (or onsite approach) was first proposed by Nakamura (1988) using the correlation between the event magnitude and the characteristic period of *P*-phase defined as the ratio between the energy of the signal and its first derivative. The method is often called the Nakamura's method.

In the standard approaches to EEW, the initial portion of the *P*-wave signal is used to rapidly characterize the earthquake magnitude and to estimate the expected ground shaking at target sites before the arrival of the most damaging waves. Based on the analysis of strong motion records, several empirical scaling relationships between EEW parameters and earthquake size have been derived (Allen and Kanamori 2003) and are implemented or being tested. The original Nakamura method was modified to describe the correlation between the predominant period and the event magnitude for Southern California events. In the ElarmS methodology (Allen 2007),



Early Earthquake Warning (EEW) System: Overview, Fig. 1 Flow of JMA's EEW (JMA 2007)

when the first station triggers, the event is temporarily located beneath that station; after a second station trigger the location moves to a point between the two stations, based on the timing of the arrivals; with three or more triggered arrivals, the event location and origin time is estimated using trilateration and a grid search algorithm.

A group of Japanese seismologists (Horiuchi et al. 2005) successfully developed an earthquake early alarm system that determines earthquake parameters of large events within a few seconds from the arrival of *P*-waves at the closest stations by using the real-time data from the Hi-net and Kanto-Tokai seismic networks that employs about 750 high dynamic range seismic stations covering the whole of the Japan Islands. These networks are maintained by the National Institute for Earth Science and Disaster Prevention (NIED). The alarm system employs a novel method to locate hypocenter using not only the *P*-wave arrival times but also the fact that *P*-waves have not yet arrived at other stations in the network.

The final output of EEW is the shaking intensity at locations where customers live and/or work. In the standard procedures, intensity is

determined from estimated values of hypocenter parameters, magnitude, and empirical attenuation equations. There, however, are difficulties in the quick determination of the magnitude for a large earthquake. Generally speaking, $M8$ -class earthquakes have faults with size of about 100 km. It takes about 30 s until all fault areas are ruptured. Numerical simulation indicates that the movement of faulting continues, for nearly the same duration of time as its fault rupture time, even after the stoppage of the rupture propagation. This means that it takes almost 1 min to observe a complete waveform for an $M8$ -class earthquake making quick determination of the magnitude for large events difficult.

To overcome this difficulty, a parameter has been proposed that quickly and accurately estimates shaking intensity (Yamamoto et al. 2008). The shaking intensity is defined from the amplitude of the filtered acceleration. Usually, displacement defines magnitude from which intensity is estimated. The new principle is based on the assumption that the intensity could be better estimated from the observed acceleration data than from the displacement data. The parameter indicating the earthquake size is named Shaking Intensity Magnitude M_I which is

intended to represent the amplitude of short period seismic waves. Station correction factors are incorporated, because short period waves are greatly amplified in the sedimentary layer just beneath stations. It has been shown that M_I grows much faster to its final value compared with the moment magnitude by conventional methods. Test result for the 2003 Off Miyagi earthquake (M 7.0) indicated that M_I grows to the 95 % of the final size by only 3 s, whereas it required 18 s for M_{JMA} to grow to the final value. Similar results were obtained for several different earthquakes. In the case of 2003 Off Tokachi earthquake of M 8.0, M_I reached about 95 % of its final value at 10 s, and M_{JMA} at about 31 s.

An Example: March 11 Tsunami Earthquake of 2011

The present JMA intensity scale has 10° from intensity 1 that people can barely detect to Intensity 7 the upper limit. Intensity 5 and 6 are divided into two degrees, namely Lower-5, Upper-5, Lower-6, and Upper-6. The JMA's EEW is divided into two grades: "forecast" and "warning." The "forecast" is issued to advanced users when events are estimated to be M 3.5 or larger, or when the expected seismic intensity is 3 or stronger. In the "forecast," the regions are specified where seismic intensity 4 or greater is expected. When estimated intensity becomes Lower-5 or greater at any observation station of the seismic intensity networks covering all over the country, "warning" is issued to the general public in regions where intensity 4 or greater is expected. The "warning" (i.e., forecasts and warnings) is updated as available data increases.

In the case of the March 11, 2011, off the Pacific coast of Tohoku Earthquake (M_w 9.0), which will be hereafter abbreviated as the March 11 Tsunami earthquake caused strong ground motion in the north-eastern part of the Honshu Island. JMA's EEW was first triggered when station OURI detected the initial P -wave at 14:46:40.2, Japan Standard Time, March 11, 2011 (Hoshiba et al. 2011). JMA issued a total of 15 EEW forecasts, but the first forecast issued at 5.4 s after the first trigger considerably underestimated the magnitude as 4.3. The fourth

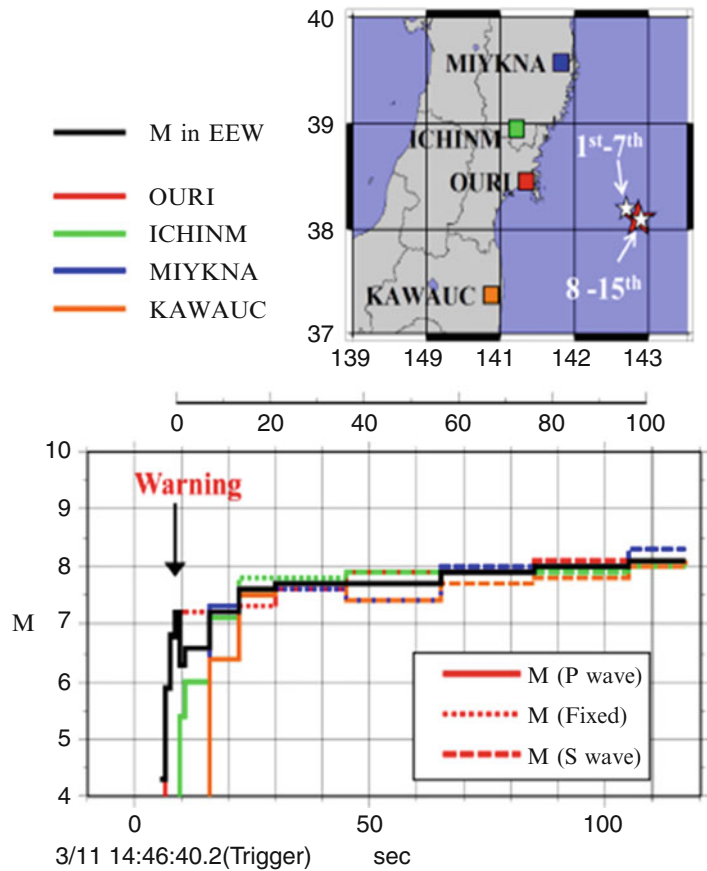
forecast issued at 8.6 s after the first trigger estimated magnitude to be 7.2 and the expected intensity increased to Lower-5 for central Miyagi prefecture. The fourth forecast, therefore, became a warning to the general public in the Tohoku district, and the warning was automatically broadcast nationwide through TV, radios and cellular phone mails. The EEW arrived the coastal areas where people live at least 15 s earlier than the strong phase of ground motion (intensity Lower-5 or greater on the JMA scale).

Figure 2 shows how the hypocenter and magnitude were updated (Hoshiba et al. 2011). Magnitudes were estimated from maximum displacement amplitude at four stations. Although the JMA's standard procedure uses records from five stations, in the case of March 11 Tsunami earthquake, the instrument at one station was found not properly working and the record from that station was discarded. Color lines represent the different stations, and the black line the median value which was used for EEW. The horizontal axis shows the elapsed time from the first trigger, and offset axis at top the elapsed time from the public warning, i.e., the fourth forecast in this particular case. The solid line indicates the magnitude determined by the P -wave, the broken line that by the S -wave, and the dotted line shows the period during which the magnitude is kept unchanged around the S -wave arrival. White stars show epicenters determined by EEW, one for the first to the seventh forecast (5.4–11.0 s after the first trigger) and another for the next eight (15.9–116.8 s). Focal depth was estimated to be 10 km for all 15 announcements. Red star indicates the epicenter location from the unified JMA catalog (focal depth is 24 km). The resolution of the JMA's EEW is 0.1° for latitude and longitude, and 10 km for focal depth. The fifteenth forecast at 116.8 s after the first trigger estimated the magnitude as 8.1 which is almost the upper limit of JMA displacement magnitude because of amplitude-magnitude saturation.

Figure 3 shows regions where forecasts and public warnings were issued (Hoshiba et al. 2011), and the distribution of estimated seismic intensities at three time points, i.e., at 14:46:48.8 when the "warning" was issued, at

**Early Earthquake
Warning (EEW) System:
Overview,**

Fig. 2 Sequence of
epicenter and magnitude
determinations of the
March 11 Tsunami
earthquake by JMA’s EEW
(Hoshiba et al. 2011)



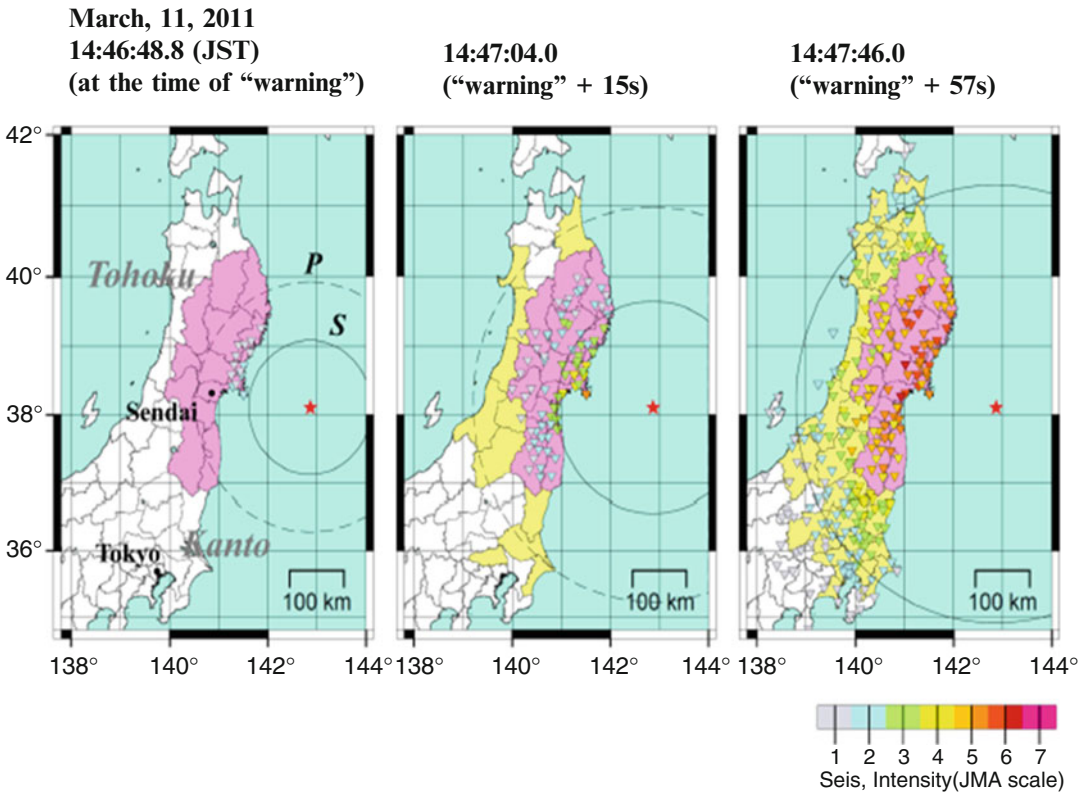
14:47:04.0 when intensity Lower-5 or greater first observed, and at 14:47:46.0 when Tokyo region was first specified as the “forecast” area. Pink areas indicate the region where the “warning” was issued, and the yellow areas are those specified as the target of the “forecast.” The distribution of the observed intensity is shown in Fig. 4. For the Tokyo region, JMA’s EEW estimated intensity 4, and warning was not issued. This was an underestimation, and actual observations reached Upper-5, which corresponds to the level of the warning issuance. After the mainshock, the system did not properly operated for several hours because of high background noise from the coda waves and active aftershocks, and because of power failure and wiring disconnections. For several days, when earthquakes occurred sometimes simultaneously over the wide source region, the system became

confused, and did not always determine the location and magnitude correctly.

In the 19 days from the main shock to March 29, 2011, JMA appropriately issued 15 warnings for the 22 events for which seismic intensity Lower-5 or greater was actually observed. On the other hand, during the same period of time, 45 forecasts were issued, but actual observed intensities did not exceed 2 at any observation station in 11 of the 45 events.

Problems and Prospects

When EEW was launched in 2007 in Japan, JMA notified the public the limitations associated with the present EEW, including (1) Timing, (2) False alarms, (3) Magnitude estimation, and (4) Seismic intensity estimation (JMA 2007). These still are difficult problems to solve in the present technical state.



Early Earthquake Warning (EEW) System: Overview, Fig. 3 Regions of EEW “warning” & “forecast,” and distribution of estimated seismic intensities (Hoshiba et al. 2011)

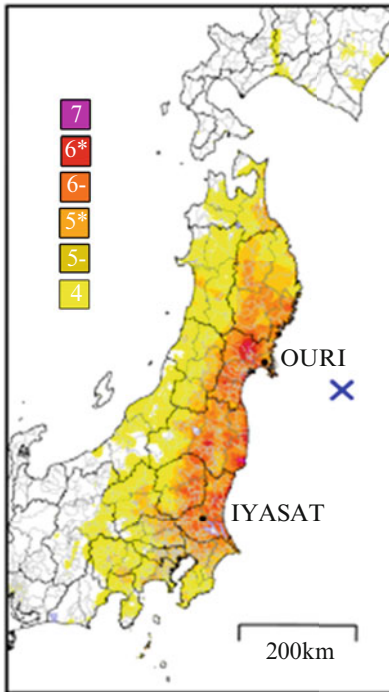
The window of time from the announcement of an EEW until the arrival of the main tremors is very short, i.e., a matter of seconds (or between several seconds and a few tens of seconds). In areas that are close to the focus of the earthquake, the warning may not be transmitted before strong tremors hit.

Because of limitations imposed by station spacing and the determination and transmission of earthquake source parameters in real-time, JMA system cannot issue an EEW to areas within about 30 km of the earthquake’s hypocenter. It is estimated that about 10 times the number of stations would be needed to issue an effective EEW near focal areas.

JMA’s EEW adopts the single station estimate for the first warning (see Fig. 1), and false EEW’s may be transmitted as a result of noise from accidents, lightning, or device failure. If the

number of stations increases, i.e., if the spacing between stations becomes much smaller than that of today, EEW need not rely upon the single station estimation from the very early stage of warning.

A conventional receiving/alarm unit of an EEW system is equipped with a CPU and memory and is on-line via the Internet. By adding an inexpensive MEMS seismometer and A/D converter, this unit is transformed into a real-time seismic observatory, which may be named a home seismometer (Horiuchi et al. 2008). If the home seismometer is incorporated in the standard receiving unit of EEW, the number of observing stations will dramatically increase. In order to properly cancel the influence of background noise caused by human activity, specially developed software is put in the system. When an event occurs, its predominant



Early Earthquake Warning (EEW) System: Overview, Fig. 4 Distribution of observed intensities during 11, 2011, earthquake (Hoshiba et al. 2011)

frequency, the number of zero crossings, and the time-duration of the triggered event are computed.

There are limits to the accuracy of estimating magnitude, especially for large earthquakes. It is difficult to separate earthquakes and provide accurate warnings when multiple earthquakes occur almost simultaneously or in close proximity to each other. These limitations were actually observed in the case of the March 11 Tsunami earthquake. There are limits to the accuracy of estimating seismic intensity by statistical attenuation formula, as well as limits to the prediction of the effect of the subsurface sedimentary layers beneath stations. Although JMA is continuing to improve the accuracy and timing of its EEW, the above-listed limitations are inherent to EEW and it may take substantial time to make the warning truly practical and accurate.

It is well known that all kinds of magnitude except M_w have the problem of magnitude saturation in magnitude range above about 8.0,

because they are defined by the maximum amplitude of displacement in frequency range higher than several seconds which becomes almost constant against the increase of earthquake size beyond about $M8.0$. Since magnitude by EEW system is defined by the displacement amplitude of P -waves in a few initial seconds, EEW systems also have the problem of magnitude saturation. As shown in Fig. 2, owing to the problem of magnitude saturation, estimated magnitude of the March 11 Tsunami earthquake was 8.1, while M_w is 9.0. While Kanamori proposed that a large earthquake is associated with fault rupture of the order of several tens of seconds, and that fault motion may be estimated from the initial P -wave phase, several seismologists opposed such methodologies are difficult in principle (Kanamori 2005; Rydelek and Horiuchi 2006; Rydelek et al. 2007).

Broadband velocity seismometers are currently well distributed around the globe and might represent an alternative to accelerometric sensors. In other words, a broadband seismometer is an accelerometer equipped with servo-type velocimeter. Velocity is obtained within the instrument itself by single integration of acceleration waveform. And one more integration operation is only needed to retrieve displacement waveform. A limitation to the use of broadband velocity seismometers, however, is that these instruments will saturate at relatively short distances from the source during the occurrence of large earthquakes.

Whether the final magnitude of an earthquake can be predicted while the rupture process is underway remains a controversial issue. For the clarification of this kind of problems, GPS methods may provide an evolutionary measurement of ground motion quantity which is directly related to the permanent ground deformation resulting from coseismic displacement after the ground motion has subsided. Since GPS stations are able to register directly the ground displacement without any risk of saturation without any need of complicated corrections, geodetic displacement time series represent an important complementary contribution to the high-frequency information provided by seismic data (Allen and Ziv 2011; Ohta et al. 2012).

Summary

The ultimate goal of the seismological and earthquake engineering studies is to protect human lives from earthquakes. Earthquake early warning systems (EEW) provide warning to the public that a potentially damaging earthquake “has” occurred before the arrival of strong motions, whereas earthquake prediction is defined as the specification of the time, location, and magnitude of a “future” earthquake. Operational EEW systems are currently running in Mexico, Taiwan, and Japan. When EEW was launched in 2007 in Japan, JMA notified the public the limitations of the present EEW including (1) Timing, (2) False alarms, (3) Magnitude estimation, and (4) Seismic intensity estimation (JMA 2007). These are all inherent to EEW and difficult to easily solve by the present technology.

Cross-References

- ▶ [Earthquake Early Warning System in Taiwan](#)
- ▶ [Earthquake Magnitude Estimation](#)
- ▶ [Principles of Broadband Seismometry](#)
- ▶ [Seismic Event Detection](#)
- ▶ [Source Characterization for Earthquake Early Warning](#)

References

- Allen RM (2007) The ElarmS earthquake early warning methodology and its application across California. In: Gasparini P, Manfredi G, Zschau J (eds) *Earthquake early warning systems*. Springer, Berlin, pp 21–44
- Allen RM, Kanamori H (2003) The potential for earthquake early warning in Southern California. *Science* 300:786–789
- Allen RM, Ziv A (2011) Application of real-time GPS to earthquake early warning. *Geophys Res Lett* 38, L16310, doi:10.1029/2011GL047947
- Bakun W, Fischer HF, Jensen E, Van Schaack J (1994) Early warning system for aftershocks. *Bull Seismol Soc Am* 4(2):359–365
- Espinosa-Aranda JM, Jimenez A, Ibarrola G, Alcantar F, Aguilar A, Inostroza M, Maldonado S (1995) Mexico City seismic alert system. *Seismol Res Lett* 66:42–53
- Horiuchi S, Negishi H, Abe K, Kamimura A, Fujinawa Y (2005) An automatic processing system for broadcasting earthquake alarms. *Bull Seismol Soc Am* 95(2):708–718
- Horiuchi S, Horiuchi Y, Yamamoto S, Nakamura H, Wu C, Rydelek PA, Kachi M (2008) Home seismometer for earthquake early warning. *Geophys Res Lett* 36, L00B04, doi:10.1029/2008GL036572
- Hoshiba M, Iwakiri K, Hayashimoto N, Shimoyama T, Hirano K, Yamada K, Ishigaki Y, Kikuta H (2011) Outline of the 2011 off the Pacific coast of Tohoku Earthquake (Mw 9.0) – Earthquake early warning and observed seismic intensity. *Earth Planets Space* 63:547–551
- Japan Meteorological Agency (2007) Earthquake early warning. Tokyo, 30 Aug 2007
- Kanamori H (2005) Real-time seismology and earthquake damage mitigation. *Ann Rev Earth Planet Sci* 33:195–214
- Nakamura Y (1988) On the urgent earthquake detection and alarm system (UrEDAS). In: *Proceedings of the 9th world conference on earthquake engineering*, vol VII. Tokyo, pp 673–678
- Odaka T, Ashiya A, Tsukada S, Sato S, Ihtake K, Nozaka D (2004) A new method of quickly estimating epicentral distance and magnitude from a single seismic record. *Bull Seismol Soc Am* 93:526–532
- Ohta Y, Kobayashi T, Tsushima H, Miura S, Hino R, Takane T, Fujimoto H, Iinuma T, Tachibana K, Demachi T, Sato T, Ohzono M, Umino N (2012) Quasi real-time fault model estimation for near-field tsunami forecasting based on RTK-GPS analysis: application to the 2100 Tohoku-Oki earthquake (Mw = 9.0). *J Geophys Res* 117, B02311, doi:10.1029/2011JB008750
- Rydelek P, Horiuchi S (2006) Is earthquake rupture deterministic? *Nature* 442:E5–E6
- Rydelek P, Wu C, Horiuchi S (2007) Comment on “Earthquake magnitude estimation from peak amplitudes of very early seismic signals on strong motion records” by Aldo Zollo, Maria Lancieri, and Stefan Nielsen. *Geophys Res Lett* 34, L20302, doi:10.1029/2007GL029387
- Teng TL, Wu Y-M, Shin TC, Tsai YB, Lee WHK (1997) One minute after: strong-motion map, effective epicenter, and effective magnitude. *Bull Seismol Soc Am* 87(5):1209–1219
- Tsukada S (2006) Earthquake early warning system in Japan. In: *Proceedings of 6th joint meeting UJNR panel on earthquake research*, Tokushima
- Wu Y-M, Teng T (2002) A virtual subnetwork approach to earthquake early warning. *Bull Seismol Soc Am* 2(5):2008–2018
- Yamamoto S, Rydelek P, Horiuchi S, Wu C, Nakamura H (2008) On the estimation of seismic intensity in earthquake early warning systems. *Geophys Res Lett* 35, L07302, doi:10.1029/2007GL033034
- Zollo A, Iannaccone G, Convertito V, Elia L, Iervolino I, Lancieri M, Lomax A, Martino C, Satriano C, Weber E, Gasparini P (2009) Earthquake early warning system in southern Italy. *Encyclopedia of complexity and systems science*. Springer, New York, pp 2395–2421

Earthquake Damage Assessment from VHR Data: Case Studies

Christian Bignami

Istituto Nazionale di Geofisica e Vulcanologia,
National Earthquake Center, Rome, Italy

Synonyms

Destruction; Loss; Quake; Seism

Introduction

The assessment of damage caused by earthquakes is of primary importance for emergency planning and human rescue, especially in remote areas where infrastructures are weak and communication systems easily break down after natural disasters. In the last decade, the remote sensing instruments, either spaceborne or airborne sensors (i.e., aerial digital cameras), have demonstrated their capabilities to provide valuable tools for damage mapping purposes, especially thanks to the availability of very high-resolution (VHR) images, characterized by pixel size around 1 m or less which are able to capture highly detailed information of the Earth surface.

A wide variety of methodologies can be used to exploit VHR data for building damage assessment, based on visual interpretation approaches, if dealing with optical imagery, and on completely or semiautomated methods as well, especially when Synthetic Aperture Radar is also considered.

The following section focuses on some relevant case studies where VHR images were adopted to map damage caused by destructive earthquakes. The dataset, the adopted methods, and the related achievements are presented for each seismic event.

The January 26, 2001, Gujarat, India, Earthquake

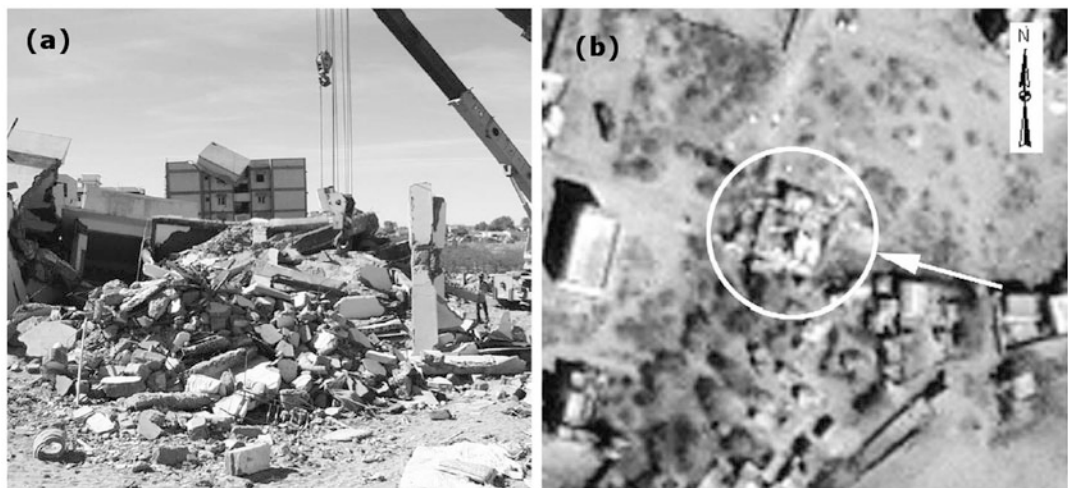
The Gujarat earthquake can be considered the first case study where damage mapping has been

attempted by means of VHR optical images. The region, northwest India, was hit by a very strong seism of 7.7 magnitude, causing destructions in more than 20 districts and killing more than 12,000 people.

This event was studied by using three images acquired by the IKONOS-2 satellite (Saito et al. 2004). One post-event image was taken on the city of Bhuj on February 2, a few days after the earthquake, and no pre-seismic data were available. A second set composed of two images, one before and one after the earthquake, acquired on November 8, 2000, and January 27, 2001, respectively, cover the Ahmedabad city. The used images are 1 m resolution in the panchromatic channel.

A visual analysis of the Bhuj satellite image was performed, aided by the comparison with ground pictures of the damaged buildings. This experiment points out that not all the occurred damage can be easily recognized in a single satellite image. In general, totally collapsed isolated buildings (especially tall buildings) and debris are clearly visible on IKONOS-2 imagery, but other types of damage, such as heavily tilted, partially collapsed buildings or low-story structures, require additional information, e.g., ground photographs, pre-earthquake satellite data, or building heights and footprints (Figs. 1 and 2). Another limitation about the use of optical imagery is, of course, the presence of shadows that prevent the detection of any damage.

The visual interpretation allowed the production of a damage levels map, too. The satellite image was divided in square blocks of one hectare, and the percentage of the collapsed or partially collapsed buildings (which correspond to the grades 4 and 5 of the European Macroseismic Scale 1998, EMS98) within each block is calculated, with respect to the total number of buildings. A damage level classification is then assigned taking into account four levels of damaged buildings percentage: 0–25 % level 1, 25–50 % level 2, 50–75 % level 3, and 75–100 % level 4. A total of 126 blocks were assessed and the final map was obtained in about 4 h by one operator. This satellite extracted map



Earthquake Damage Assessment from VHR Data: Case Studies, Fig. 1 Collapsed single building: ground photo (a) and satellite image (b); the ruins of the building

are visible in the satellite image; *white arrow* in the satellite image shows the look angle of the photo (From Saito et al. 2004)

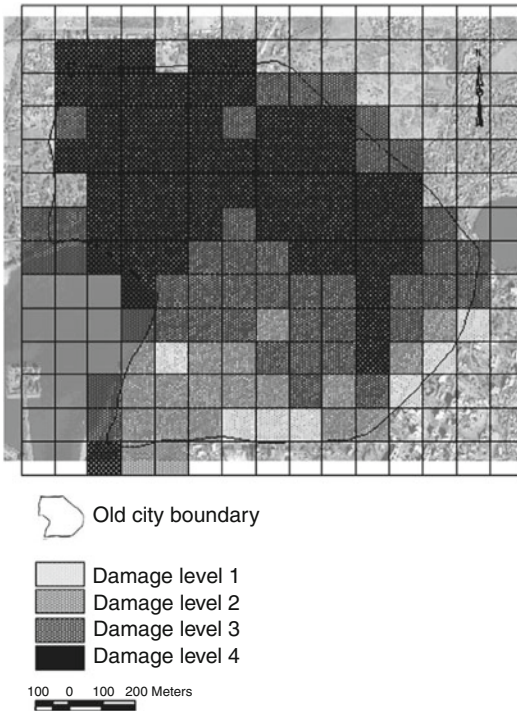


Earthquake Damage Assessment from VHR Data: Case Studies, Fig. 2 A building with collapsed ground and first floors: ground photo (a) and the same building

circled in white in the satellite image (b); the low-story collapse is not visible in the IKONOS-2 acquisition (From Saito et al. 2004)

was compared with a map derived by a ground survey. Such a comparison highlights that the map obtained from remote sensing data underestimates, on average, the level of damage by one level, and the standard deviation of the difference is 1.58. Despite this discrepancy the overall pattern of the two maps is very similar (Fig. 3).

The higher values of differences (2 and 3 levels) can be ascribed mainly to the resampling of the ground survey map, from its original format (i.e., aggregated at street block scale) to the square block type of satellite map. Indeed, in this step a reclassification of damage level was computed, according to the levels defined in the IKONOS-2 derived assessment.



Earthquake Damage Assessment from VHR Data: Case Studies, Fig. 3 Damage classification levels map obtained from IKONOS-2 post-seismic image (From Saito et al. 2004)

A similar analysis were performed to assess the damage occurred on the west Ahmedabad. As already mentioned, pre- and post-earthquake IKONOS-2 images were available for this case study. The visual interpretation was done considering a grid of 300×300 m. In Ahmedabad less than 60 buildings were totally or partially collapsed; therefore, the inspection was focused on the detection of individual damaged buildings among the intact ones. Twelve locations were marked as damaged buildings, and ten of them were confirmed collapsed by a filed survey.

These first outcomes of VHR images exploitation demonstrated that such type of images can provide useful information to identify the most damaged areas after a destructive earthquake. Even if the level of accuracy is limited by various factors, the information from satellite data, if timely delivered to emergency management stakeholders, can be of great value.

The December 26, 2003, Bam, Iran, Earthquake

On December 26, 2003, at 5:56 local time, a strong earthquake of magnitude 6.6 devastated the city of Bam, in southeast Iran. More than 26,000 people died, 30,000 were injured, and about 85 % of the buildings and infrastructures were destroyed (Hisada et al. 2004).

The city of Bam was observed by the VHR optical sensor of QuickBird satellite before the seismic event on September 30, 2003, and after the earthquake on January 3, 2004. QuickBird can take images at a pixel resolution of 0.6 m in the panchromatic channel and at 2.4 m in the multispectral bands.

To estimate the damage occurred in Bam, the QuickBird data were exploited with different methodologies in a wide number of research papers. Bam is the first case where automated procedures were attempted to extract damage index from VHR optical data.

The visual comparison of the two QuickBird images was adopted to map damage buildings of Bam, following the same methodology and damage classification proposed for the Gujarat earthquake (Saito et al. 2004). A damage classification was also performed at single building scale, by assigning the EMS98 damage grades to each of the detected damaged edifice (Yamazaki et al. 2005). A total of 12,063 buildings were classified (see Fig. 4) after 20 h working for images interpretation.

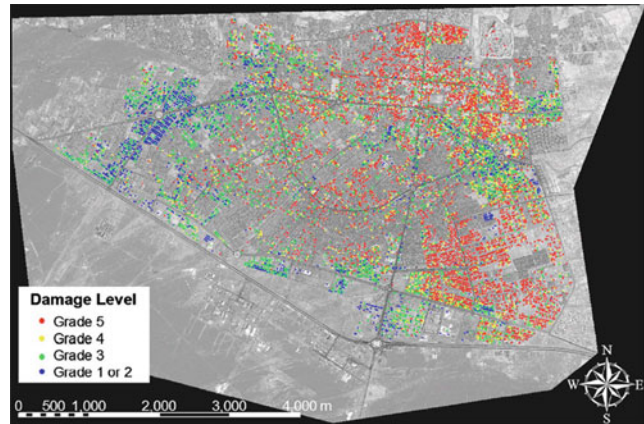
Besides this extremely time-consuming approach, some automatic methods were tested.

A different approach for damage mapping based on the use of edge information and textural features was also tested (Vu et al. 2005). The edge information is obtained by applying the Prewitt filter (Takagi and Shimoda 1991). In particular, edge intensity (E_i), its variance (E_v), and the ratio of the predominant direction (E_d) of the edge intensity were derived. In addition, the textural parameters *angular second moment* (T_a) and *entropy* (T_e) are derived from edge intensity by calculating the co-occurrence matrix (Baraldi, and Parmiggiani 1995). Pixels within the range of predetermined threshold

Earthquake Damage Assessment from VHR

Data: Case Studies,

Fig. 4 Damage mapping results from pre- and post-seismic data visual comparison. The numbers of identified damaged buildings are 1,597 (grades 1 or 2: blue points), 3,815 (grade 3: green), 1,700 (grade 4: yellow), and 4,951 (grade 5: red) (From Yamazaki et al. 2005)



values of the four parameters E_v , E_d , T_a , and T_c correspond to damaged buildings.

The method was applied to IKONOS-2 and QuickBird post-event data. The vegetated areas were masked out by thresholding the Normalized Difference Vegetation Index (NDVI). The results were compared to the damage map obtained by visual inspection, and it was found that the distribution of detected damage from IKONOS-2 was consistent with these results, while less damage information was detected from the QuickBird image.

In Huyck et al. 2005, a similar method based on textural analysis is performed. On both pre- and post-seismic images from QuickBird, the edge detection is performed, and from the resulting images, the *dissimilarity* textual parameter is calculated. At this stage, the change detection of *dissimilarity* is done. The damage in the post-event data gives a high dissimilarity of edges, while intact buildings yield a low dissimilarity of edges. The image textural difference is then classified, on the base of standard deviation values, and finally, the damage map is extracted by considering the mean value of the class within 200×200 pixel block. The textural analysis was also applied in a multisensor scenario, where the pre-seismic image is the QuickBird acquisition and the post-event one is the IKONOS-2 data, begin the latter resampled at the same resolution of QuickBird image, i.e., from 1 to 0.6 m.

The two damage maps were compared with damage observations obtained from an independent source compiled by US-AID where damage

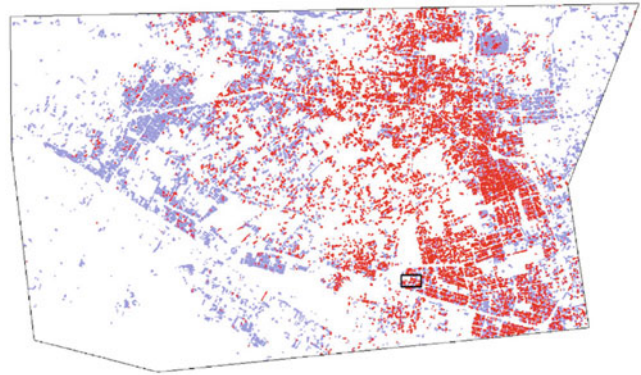
is categorized by the percentage of collapsed structures: 80–100 %, 50–80 %, 20–50 %, and <20 %. The observed the accuracy rate is high, demonstrating the potential of this damage detection algorithm.

Generally, the use of a pixel by pixel approach with VHR optical data has some limitations because of the presence of some restrictive factors such as different viewing angles of the pre- and post-images (and the related geometric distortions), different sun illumination, and shadows due to the different seasons of the satellite acquisitions. These are the reason why the visual interpretation is considered the most reliable approach, but such problems can be reduced by adopting specific methods. One of the proposed in literature is the object-based technique. This technique analyzes each building (the object in the satellite scene) instead of considering individual pixels. In Gusella et al. 2005, an object-oriented technique was adopted for Bam case study. The object selection is done on the pre-earthquake image, where intact buildings are identified by using, first, a segmentation procedure with a bottom-up region merging strategy and then a supervised classification method by using *eCognition* software. The aim is to separate building objects from nonbuilding objects (e.g., vegetation, cars, and shadows). Then the classified buildings in the pre-event image are superimposed on the post-event one, and to distinguish between collapsed and non-collapsed structures, the post-earthquake buildings are classified, with a supervised approach, for a second time using training set of

Earthquake Damage Assessment from VHR

Data: Case Studies,

Fig. 5 Building damage map for the city of Bam, based on object-oriented analysis of QuickBird imagery captured before and after the event. Collapsed structures are shown in *red* and non-collapsed in *blue* (From Gusella et al. 2005)



Earthquake Damage Assessment from VHR Data: Case Studies, Fig. 6 Building damage map: (a) detail of panchromatic image taken on September 30, 2003; (b) detail of panchromatic image taken on January 3, 2004;

(c) damaged pixels, in *red*, superimposed to the difference between pre- and post-seismic images (From Chini et al. 2009)

damaged and non-damaged objects from the building class. Note that these automated methods require a preliminary co-registration of the two satellite images, to match the geometry of one image on the other (Richards 2006).

The resulting damage map is shown in Fig. 5. The accuracy assessment, performed on a random set of 136 buildings, reaches an overall value of 70.5 %.

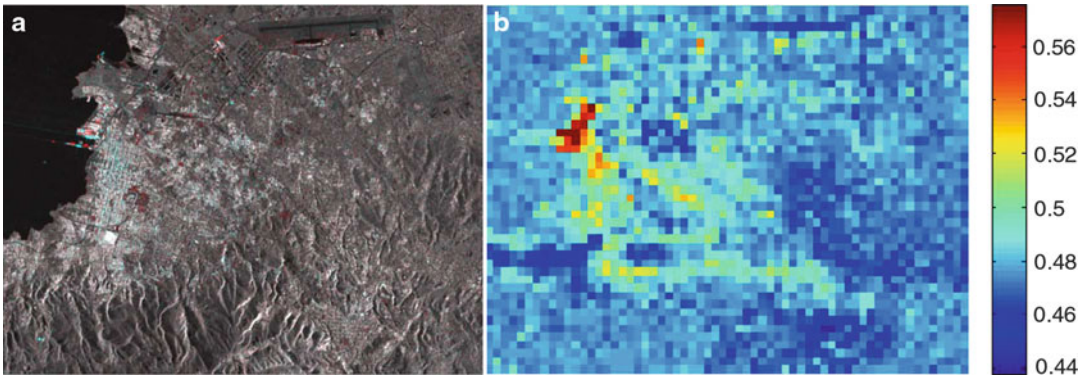
A similar way is proposed in the work of Hutchinson and Chen 2005. Here the object selection is performed through a hierarchical segmentation procedure.

The objects selection can also be obtained by following different methods. An alternative solution is the exploitations of the morphological operators (Soille 2003). It is accomplished by applying the closing and the opening filters, with different size, to extract additional features from panchromatic band of VHR pre-seismic

image (Chini et al. 2009). This results in a feature vector which is used as input dataset of an unsupervised classifier. The output of the classifier is therefore the mask of the buildings (i.e., the objects) of Bam city. At this stage, the pixel by pixel normalized difference (ND), i.e., a simple change detection, is computed only on the mask-selected objects, and by setting a threshold on ND value, the damaged buildings are detected. The result of this methodology is presented in Fig. 6. Such approach allowed improving roughly of 15 % the accuracy with respect to the change detection computed over the entire scene.

The January 12, 2010, Port-au-Prince, Haiti, Earthquake

On January 12, 2010, a devastating earthquake of magnitude 7.2 hit the city of Port-au-Prince and



Earthquake Damage Assessment from VHR Data: Case Studies, Fig. 7 (a) Cosmo-SkyMed color composite of Port-au-Prince (R = January 13, G = B = 26 April);

(b) average normalized difference map of Cosmo-SkyMed pair (From Dekker 2011)

E

the surrounding villages in Haiti. In this very destructive event, more than 300,000 people were killed and over one million were displaced. Over 310,000 houses were destroyed or damaged, and medical facilities, educational buildings, and basic infrastructure such as roads, bridges, ports, and airports were severely damaged (Duda and Jones 2011).

As for the case of Bam and Gujarat earthquakes, the VHR optical images were exploited adopting some of the methodologies (or similar) already presented (e.g., Corbane et al. 2011; Tiede et al. 2011; Voigt et al. 2011). For such reason, this section will focus on the methods and results carried out by using VHR images captured from SAR sensors.

During the emergency phase the area of Port-au-Prince was captured by the German SAR sensor of TerraSAR-X mission (Pitz and Miller 2010) and by the Italian constellation COSMO-SkyMed (Covello et al. 2010). Both satellites can acquire VHR images, at X band, reaching resolution of about 1 m on ground, allowing urban area analysis, not only at city block scale but also at single building scale.

SAR data overcome the typical limitations of optical images that cannot provide information in case of cloud coverage, and only daylight satellite passes are suitable for image collection. Actually, SAR instruments, being active sensors, can operate in all weather conditions and day or night orbital

passages, allowing better performances in terms of time latency between satellite tasking and image availability that reflects a prompt response for damage map delivery. It is worth noticing that usually the visual interpretation of VHR SAR images is a very hard task to be done; therefore, most of the approaches rely on automated methods, mainly based on change detection techniques.

For the Haiti case, pre-seismic and post-seismic images from both TerraSAR-X and COSMO-SkyMed satellites were available. The TerraSAR-X images are dated October 13, 2009, and January 20, 2010, while the COSMO-SkyMed pair was acquired on April 26, 2009, and January 15, 2010. The two datasets are on descending orbit, in strip-map mode at 3 m resolution (Dekker 2011).

One of the proposed method relies on a change detection approach which is based on the normalized difference (ND) between the pre- and post-earthquake images. The damage estimation was computed at grid level, by considering the average value of the ND within a grid cell or area. A cell size of 200×200 m was adopted. An alternative damage index proposed for the Haiti case study was estimating the de-correlation coefficient between pre-seismic and post-seismic images (Yonezawa and Takeuchi 2001; Stramondo et al. 2006). It can be obtained by correlating the intensity pixels within a certain grid cells, which was again of 200×200 m for this analysis. In Fig. 7 an example of the damage

index derived from the ND average evaluation obtained from COSMO-SkyMed data and the original SAR image are shown.

The resulting maps were compared with a damage assessment carried out in the framework of the Post Disaster Needs Assessment and Recovery Framework (PDNA) at request of the Government of Haiti (<http://oneresponse.info/Disasters/Haiti/Early%20Recovery/>). The PDNA map was derived from VHR satellite and airborne optical images, on a visual interpretation base. Such map is at building level and on the classification scheme of the EMS98 scale. In order to compare the results, the PDNA map was also averaged with the same grid cells of the SAR-derived maps. The average PDNA damage map shows the same damage distribution pattern as in the average damage maps calculated from TerraSAR-X and Cosmo-SkyMed, with some discrepancies in the eastern area. In addition to this visual comparison, the regression analysis between PDNA map and SAR maps was performed for both ND and grid-cell de-correlation maps. The regression analysis highlighted that, for both satellites, the ND outperforms the de-correlation estimation, but the two indices are not well correlated with the PDNA damage map. Probably not all the damage reflects in a strong enough change of radar backscattering. It can be explained by the different scattering mechanisms and radar orientation (Dekker 2011).

Besides the grid-based approach, a detailed SAR damage assessment was also carried out. This experiment was done on TerraSAR-X data, adopting the logarithmic ratio between the pre- and post-event images as a damage index. The logarithmic ratio is also filtered to remove the well-known speckle effect, and two thresholds for the increase and decrease of radar backscatter are set (6.0 dB and -6.0 dB, respectively). Comparing the resulting map with the PDNA damage assessment, a similar trend in the center of Port-au-Prince was observed, but TerraSAR-X damage map seems weakly correlated at building level, especially in the shanty area of the city. Reason for this latter phenomenon is that shanty areas are less regular and more

compact, and it result in less open double-bounce scattering mechanism.

Summary

This entry describes the main achievements about the very high-resolution (VHR) remote sensing data in the fieldwork of earthquake damage estimation, from their first application. Both optical and SAR imagery exploitation is reported, with the description of the main steps that were adopted to obtain damage detection and assessment maps. Some relevant case studies that can be considered the basis of the VHR remote sensing methodologies are here presented.

Cross-References

- [Building Damage from Multi-resolution, Object-Based, Classification Techniques](#)
- [Building Disaster Resiliency Through Disaster Risk Management Master Planning](#)
- [Building Monitoring Using a Ground-Based Radar](#)
- [Damage to Ancient Buildings from Earthquakes](#)
- [Damage to Buildings: Modeling](#)
- [Damage to Infrastructure: Modeling](#)
- [Earthquake Disaster Recovery: Leadership and Governance](#)
- [Earthquake Location](#)
- [Earthquake Magnitude Estimation](#)
- [Estimation of Potential Seismic Damage in Urban Areas](#)
- [Features Extraction from Satellite Data](#)
- [Hyperspectral Data in Urban Areas](#)
- [Remote Sensing in Seismology: An Overview](#)
- [Resilience to Earthquake Disasters](#)
- [SAR Images, Interpretation of](#)
- [Seismic Collapse Assessment](#)
- [Seismic Loss Assessment](#)
- [System and Damage Identification of Civil Structures](#)
- [Urban Change Monitoring: Multi-temporal SAR Images](#)

References

- Baraldi A, Parmiggiani F (1995) An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Trans Geosci Remote Sens* 33(2):293–304. doi:10.1109/36.377929
- Chini M, Pierdicca N, Emery WJ (2009) Exploiting SAR and VHR optical Images to quantify damage caused by the 2003 Bam earthquake. *IEEE Trans Geosci Remote Sens* 47(1):145–152. doi:10.1109/TGRS.2008.2002695
- Corbane C, Carrion D, Lemoine G, Broglia M (2011) Comparison of damage assessment maps derived from very high spatial resolution satellite and aerial imagery produced for the Haiti 2010 earthquake. *Earthq Spectra* 27(S1):S199–S218. doi:10.1193/1.3630223
- Covello F, Battazza F, Coletta A, Lopinto E, Fiorentino C, Pietranera L, Valentini G, Zoffoli S (2010) COSMO-SkyMed an existing opportunity for observing the Earth. *J Geodyn* 49(3–4):171–180. doi:10.1016/j.jog.2010.01.001
- Dekker RJ (2011) High-resolution radar damage assessment after the earthquake in Haiti on 12 January 2010. *IEEE J Select Top Appl Earth Observ Remote Sens* 4(4):960–970. doi:10.1109/JSTARS.2011.2164516
- Duda KA, Jones BK (2011) USGS remote sensing coordination for the 2010 Haiti earthquake. *Photogramm Eng Remote Sens* 77(9):899–907
- Gusella L, Adams BJ, Bitelli G, Huyck CK, Mognol A (2005) Object-oriented image understanding and post-earthquake damage assessment for the 2003 Bam, Iran, earthquake. *Earthq Spectra* 21(S1):225–238. doi:10.1193/1.2098629
- Hisada Y, Shibaya A, Ghayamghamian MR (2004) Building damage and seismic intensity in Bam City from the 2003 Iran, Bam, earthquake. *Bull Earthq Res Inst Univ Tokyo* 79:81–93
- Hutchinson TC, Chen ZQ (2005) Optimized estimated ground truth for object-based urban damage estimation using satellite images from the 2003 Bam, Iran, earthquake. *Earthq Spectra* 21(S1):239–254. doi:10.1193/1.2098937
- Huyck CK, Adams BJ, Cho S, Chung H-C, Eguchi RT (2005) Towards rapid citywide damage mapping using neighborhood edge dissimilarities in very high-resolution optical satellite imagery – application to the 2003 Bam, Iran, earthquake. *Earthq Spectra* 21(S1):255–266. doi:10.1193/1.2101907
- Pitz W, Miller D (2010) The TerraSAR-X satellite. *IEEE Trans Geosci Remote Sens* 48(2):615–622. doi:10.1109/TGRS.2009.2037432
- Richards JA (2006) Remote sensing digital image analysis: an introduction, 4th edn. Springer, Berlin
- Saito K, Spence RJS, Going C, Markus M (2004) Using high-resolution satellite images for post-earthquake building damage assessment: a study following the 26 January 2001 Gujarat earthquake. *Earthq Spectra* 20(1):145–169. doi:10.1193/1.1650865
- Soille P (2003) Morphological image analysis: principles and applications, 2nd edn. Springer, Berlin/New York
- Stramondo S, Bignami C, Chini M, Pierdicca N, Tertulliani A (2006) Satellite radar and optical remote sensing for earthquake damage detection: results from different case studies. *Int J Remote Sens* 27(20):4433–4447. doi:10.1080/01431160600675895
- Takagi M, Shimoda H (1991) Handbook of image analysis. University of Tokyo Press, Tokyo (in Japanese)
- Tiede D, Lang S, Fuereder P et al (2011) Automated damage indication for rapid geospatial reporting. *Photogramm Eng Remote Sens* 77(9):933–942
- Voigt S, Schneiderhan T, Twele A, Gähler M, Stein E, Mehlert H (2011) Rapid damage assessment and situation mapping: learning from the 2010 Haiti earthquake. *Photogramm Eng Remote Sens* 77(9):923–931
- Vu TT, Matsuoka M, Yamazaki F (2005) Detection and animation of damage using very high-resolution satellite data following the 2003 Bam, Iran, earthquake. *Earthq Spectra* 21(S1):319–327. doi:10.1193/1.2101127
- Yamazaki F, Yano Y, Matsuoka M (2005) Visual damage interpretation of buildings in Bam City using QuickBird images following the 2003 Bam, Iran, earthquake. *Earthq Spectra* 21(S1):329–336. doi:10.1193/1.2101807
- Yonezawa C, Takeuchi S (2001) Decorrelation of SAR data by urban damages caused by the 1995 Hyogoken-Nanbu earthquake. *Int J Remote Sens* 22(8):1585–1600. doi:10.1080/01431160118187

Earthquake Disaster Recovery: Leadership and Governance

Ljubica Mamula-Seadon
Centre for Infrastructure Research at the
Department of Civil and Environmental
Engineering, Faculty of Engineering,
The University of Auckland, Auckland,
New Zealand

Synonyms

Adaptive management; Collaborative planning;
Integrated disaster management; Public-private
engagement; Social capital and leadership;
Transformative governance

Introduction

Disaster management practice around the world evolved gradually, mainly in response to major

events such as earthquakes, volcanic eruptions, floods, or hurricanes. In the last few decades, an ever-increasing vulnerability and complexity of risks brought to prominence integrated and comprehensive, all-hazards approaches, supported by frameworks and structures based upon a shared system of governance and policy making. These approaches presume democratic environment with decentralized and deliberative planning and decision-making processes that integrate risk management into broader sustainable development strategies. The system requires extensive coordination and cooperation among all levels of government, as well as with the many other stakeholders, such as private sector, NGOs, and community groups involved in disaster management. The supporting legislation, policies, guidelines, and plans are commonly based upon sustainable development principles; they also introduce the concept of shared responsibility and empowerment of local communities, reflecting the recognition that it is essential to develop society's overall "adaptive capacity" or resilience (Tierney 2012).

The all-hazards, comprehensive approach to risk and disaster management means that the practice has evolved from its historical sphere of natural hazard events and now incorporates an array of risks, including major environmental and technological, as well as terrorist, events. Furthermore, the approach comprises risk prevention or risk reduction, preparedness or readiness, and response to and recovery from events, internationally recognized as PPRR of risk prevention, preparedness, response, and recovery, or 4Rs of risk reduction, readiness, response, and recovery. These elements are not sequential or mutually exclusive, but each element contains components of the other three and may, at least in part, be simultaneously operational. Each element should be fully integrated through broader planning programs and management processes for disaster management (MCDEM 2005; FEMA 2011).

The subsequent sections of this article focus on one of the elements of the integrated, comprehensive disaster management – disaster recovery framework, with the accent on governance and

leadership as the key elements of successful recovery. The sections provide a discussion on recovery and earthquake recovery, its principles, and its relation to other elements of disaster management (see section "[Recovery: Definition and Principles](#)"); an identification of specific characteristics of governance and leadership in post-earthquake recovery (see section "[Recovery: Governance](#)"); a discussion on governance transformation and leadership in recovery (see section "[Governance Transformation and Leadership in Recovery](#)"); considerations of strengths and weaknesses of different approaches to recovery governance, as well as strategies for successful recovery governance and leadership (see section "[Strengths and Weaknesses of Different Approaches to Earthquake Recovery Governance and Leadership](#)"); and a discussion on recovery governance and social capital and its implications for governance and leadership for resilience and future trends in disaster recovery (see section "[Recovery Governance, Leadership, Social Capital \(or Social Capacities\), and Resilience](#)").

The Role of Governance and Leadership in Disaster Recovery

Recovery: Definition and Principles

Within a broader comprehensive, integrated framework for disaster management, recovery is generally defined as a coordinated process that involves supporting affected communities in reconstruction of the built environment and the restoration of emotional, social, cultural, economic, and natural environment well-being following an emergency. It is recognized that recovery is a complex social and developmental process that involves much more than simply the replacement of what has been destroyed and the rehabilitation of those affected. Effective recovery presents an opportunity to reduce future risks, "build back better," and foster a more resilient community (MCDEM 2005; FEMA 2011; Australian Government 2011; UK Government 2013). Current scholarship recognizes that the actual recovery processes vary with each disaster and that the so-called stages of recovery overlap

with one another, with different groups and sub-regions within disaster-affected areas recovering at different times and in different ways, rendering recovery nonlinear and often iterative (Tierney and Oliver-Smith 2012). Furthermore, recovery is affected not only by local circumstances but by macro-societal and even global factors and, importantly, by their interconnection (*ibid.*). Recovery is not only a set of activities that occur after a disaster but a continuum that involves both pre-disaster and post-disaster planning, as well as risk-reduction and resilience-building activities that link recovery to other aspects of a comprehensive risk management framework (Smith 2011).

It is now widely recognized that recovery is best achieved when the affected communities are actively involved and able to exercise a high degree of self-determination or even take the leadership in developing recovery priorities and activities (MCDEM 2005; FEMA 2011; Australian Government 2011).

Earthquake recovery epitomizes all the above characteristics of recovery. In particular, due to the sudden and devastating nature of earthquakes, especially when they impact urban environments, earthquake recovery presents communities and governments with an enormous and complex long-term process that often challenges preexisting concepts and frameworks.

Historically, scholars postulated that the amount of physical damage from a disaster best correlates with recovery speed, arguing that areas which suffered a higher degree of damage to the built environment and higher injuries and casualties will recover more slowly, requiring more time for reconstruction, with housing in shorter supply (Aldrich 2011). This is particularly relevant for disasters caused by earthquakes. However, there are theories that the scale of damage and loss does not necessarily correlate with recovery rates, and factors such as population density, pace of rebuilding, and socioeconomic status play equally important role (Aldrich 2011). Tierney and Oliver-Smith (2012) define recovery as a socially configured process that is inextricably linked to the recovery of not only the built environment and infrastructure but also

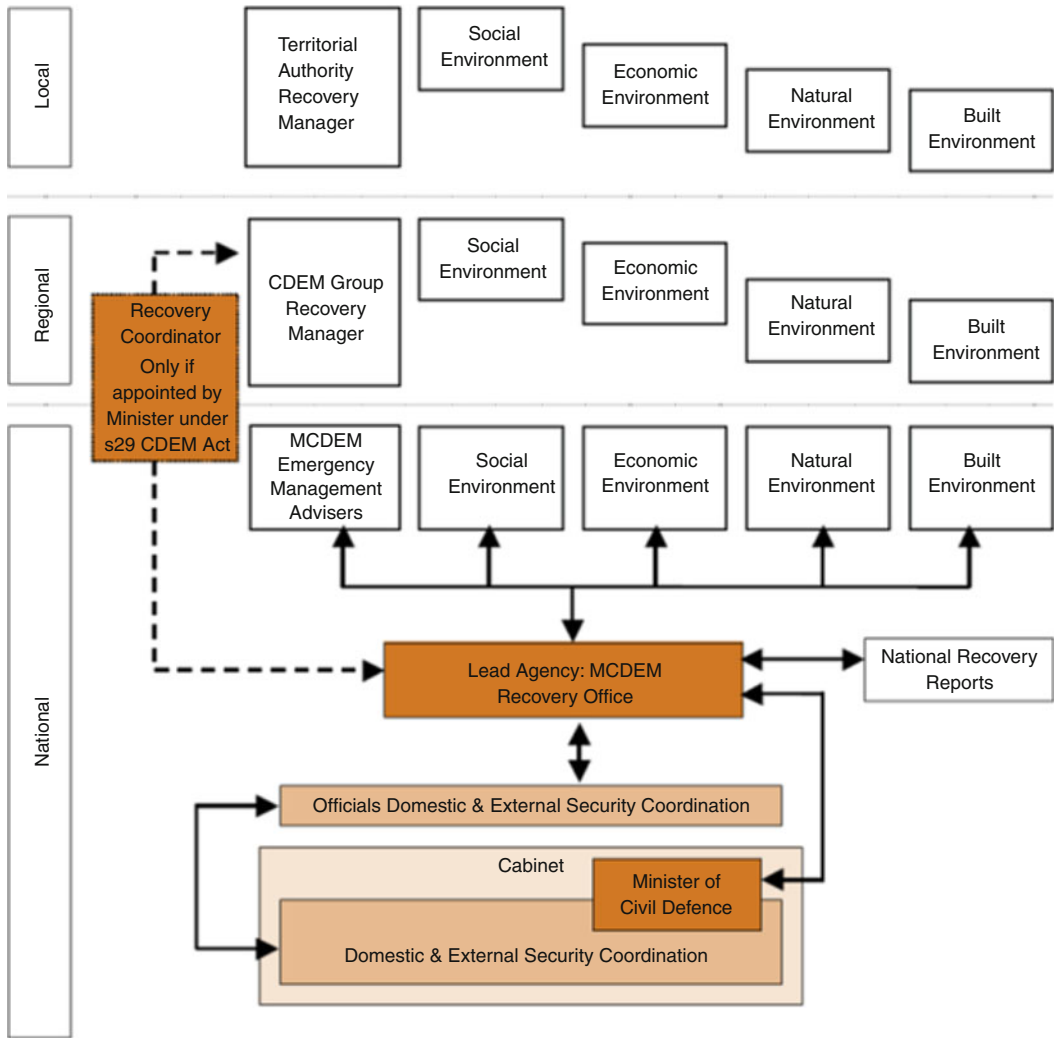
ecosystems, organizations and institutions, economic activity, and culture and, hence, cannot be solely or closely equated with post-disaster reconstruction of the physical environment. It is, rather, a holistic, intricate process requiring complex, integrated, and adaptive governance structures to succeed.

Recovery: Governance

Disaster governance is often defined as a system that enables collaborative activities that bring together multiple organizations (government and nongovernment) to solve problems that extend beyond the purview of any single organization (Tierney 2012). More specifically, Tierney (2012) defines disaster governance as interrelated sets of norms (such as laws and regulations, frameworks, standards), organizational and institutional actors, and practices – across the 4Rs of risk reduction, readiness, response, and recovery – designed to reduce the impacts associated with disasters.

Commonly, governance structures include public sector organizations (federal or national, state or regional, and local governments), nongovernmental organizations (community development corporations, homeowners' associations, special service districts, regional planning organizations, professional associations, and colleges and universities), nonprofit relief organizations (nonprofits, community-based organizations, and foundations), private sector organizations (businesses and corporations, financial and lending institutions, insurance, and media), international relief organizations and nations, and emergent groups and individuals, who together provide three types of resources: financing, policies, and technical assistance (Smith and Birkland 2012). The degree of coordination and integration, underlying values and processes, timing, capacity, and capability of those actors are some of the key factors that determine the effectiveness of recovery processes (Smith and Birkland 2012).

Many countries have developed national frameworks, governance models, and administrative guidelines for recovery management and governance (MCDEM 2005; FEMA 2011;



Earthquake Disaster Recovery: Leadership and Governance, Fig. 1 Multilevel recovery management structure from the New Zealand CDEM group recovery guidelines (MCDEM 2005)

Australian Government 2011; UK Government 2013). In general, all these models reflect the holistic, comprehensive, and integrated risk management approaches based on the principles for sustainable development. They are characterized by decentralized, “bottom-up” structures, with local governments having primary responsibility for supplying disaster-related resources and regional, subnational, national, and international agencies providing support as requested. A similar approach is advocated by the United Nations as well (Fraser-Molekati 2012). For example, New Zealand was one of the first

countries to formalize the approach in its national recovery framework, as illustrated in Fig. 1.

The recovery framework diagram in Fig. 1 depicts the national arrangements in New Zealand, showing the national government as supporting recovery activities lead by the local government in partnership with an array of stakeholders, including private and nongovernment organizations. Depending on the scale of event, the lead agency changes from local to national. The holistic nature of recovery is achieved through integration across different “environments”: social, economic, natural, and built.

Before generalizing recovery arrangements, however, it is important to note that recovery is strongly influenced by global and societal factors that include the economic strength and GDP, the overall ability of societies to address recovery challenges, and the nature and extent of involvement in recovery on the part of international agencies and nongovernment organizations and the outcomes of that involvement, among other factors. If societal systems are weak, this will reflect on recovery arrangements and recovery efforts, and the affected nations will probably be highly dependent on international aid (Tierney and Oliver-Smith 2012).

Researchers found out that, following major disasters, governments are required to undertake complex rebuilding and societal activities in a compressed period of time. This often leads to changes in governance (Olshansky et al. 2012). The same authors summarize recent research findings that identified the effects of time compression on governments and governing in post-disaster recovery as: insufficient capacity to solve problems, provide resources, and take actions; excess demands for information and stakeholder deliberation to make timely decisions; greater needs for organizational integration and coordination; and immediate demands for large amounts of money with existing distribution systems challenged to meet the needs. In the time-compressed recovery environment, bureaucracies often struggle with these and other demands, and as a result, new organizations, both governmental and nongovernmental, emerge to provide more capacity, information, money, and other resources. Sometimes governments create new institutions to manage recovery, as in Louisiana (2005 Hurricane Katrina), Queensland (2010 floods), and Tohoku (2011 earthquake and tsunami). Other times, existing institutions and procedures are modified or adapted, as in Los Angeles (1994 earthquake), Kobe (1995 earthquake), and Chile (2010 earthquake and tsunami) (Johnson and Olshansky 2013). It is important to note that, historically, in times of need, high expectations are placed on governments. Commonly, governments have fronted recovery in all major disasters worldwide, demonstrating different degrees of leadership and effectiveness.

Governance Transformation and Leadership in Recovery

In recent years Australia and New Zealand have experienced large natural disasters. The State of Victoria had bushfires burning for over a month in early 2009, culminating in “Black Saturday” on February 7 and devastating rural communities. For 3 months through the summer of 2010–2011, heavy rains brought extensive flooding to Queensland, compounded by damage from Cyclone Yasi in February 2011. An earthquake sequence devastated central Christchurch in New Zealand, with thousands of quakes occurring over a period of 2 years, between 2010 and 2012. On September 4, 2010, an earthquake with a magnitude of 7.1 occurred only 40 km from the Christchurch City center. The earthquake did not directly cause life loss, but damage to infrastructure and buildings was substantial. It was followed by thousands of aftershocks, some strong enough to cause further damage to the city. On February 22, 2011, a magnitude 6.3 earthquake struck only 6 km from the Christchurch CBD and less than 5 km below ground. This earthquake caused far greater damage than the September earthquake, 185 people died, and the government declared a state of national emergency for the first time in New Zealand history. Christchurch City and the surrounding districts had heavy damage to homes, businesses, infrastructure and education, medical, and community facilities. Further aftershocks caused additional damage, ground failures, and building and infrastructure damage. In all, more than 4,400 aftershocks (of magnitudes greater than three) have been registered from September 2010 to September 2012 (GeoNet 2013).

Both Australia and New Zealand are developed economies with similar political and legal institutions and both had developed recovery frameworks, including governance arrangements, prior to the latest emergency events. However, following these emergencies, the governments in both countries felt compelled to introduce a number of changes to preexisting recovery arrangements, including changes to legislation and governance structures.

Establishment of the special recovery authority in Victoria – Victorian Bushfire Reconstruction and Recovery Authority (VBRRA) – followed Black Saturday, within the preexisting legislative framework (Smart 2012). In the case of the Queensland floods and the Canterbury earthquakes, however, it was perceived that the magnitude and duration of events exhausted the capacity of preexisting institutions to cope, and recovery authorities with greater powers were deemed necessary. In Queensland the Queensland Reconstruction Authority (QldRA) was established (Smart 2012). The QldRA built on the experience by VBRRA, taking onboard recommendations to align with the highest level of government, acquire statutory powers to expedite decision-making and have clearly defined terms of reference (Smith and Birkland 2012). Both Queensland and New Zealand circumvented existing regulations, New Zealand effectively emulating the Queensland experience into its arrangements for the Canterbury Earthquake Recovery Authority (CERA). Both agencies were given similar powers; both could undertake works, enter and acquire land, decide the priorities for reconstruction, and override local government plans and policies; and both had legislated sunset clauses (Smart 2012). The key differences were in their status – CERA was a government department reporting directly to the newly appointed Recovery Minister, while the QldRA was created as a statutory body with a governing board. Effectively, the New Zealand arrangement allowed the government to exert a more direct influence over the recovery process (Smart 2012).

Similar to the governance changes in Australasia, the Great East Japan earthquake of March 11, 2011, triggered a number of major initiatives to transform the current system of disaster mitigation governance in Japan. New government organizations were established: the Reconstruction Agency to deal with recovery and reconstruction (in February 2012) and the Nuclear Regulatory Authority to deal with nuclear safety (September 2012). The role of the Reconstruction Agency is to integrate across different sectors of the national government, as well as to

coordinate between the central and local government (Suzuki and Kaneko 2013).

Following the devastation caused by Hurricane Sandy in October 2012, on January 29, 2013, President Obama signed into law the Sandy Recovery Improvement Act 2013. The legislation introduces significant changes to the way the Federal Emergency Management Agency (FEMA) delivers disaster assistance and empowers FEMA to develop processes and procedures for newly established authorities and their implementation. Prior to the introduction of the act, in December 2012, President Obama issued an executive order that established the Hurricane Sandy Rebuilding Task Force, atypically chaired by the Secretary of Housing and Urban Development, thus allocating the lead role to the Department of Housing and Urban Development, the first for this agency (White House 2012).

It is important not to underestimate the influence Hurricane Katrina had on disaster management in the USA and internationally. The perceived poor recovery following Hurricane Katrina was investigated by the House Select Committee who identified failures at all levels of government to support otherwise exemplary efforts by nongovernment organizations, groups, and individuals. The most important failure for the government was lack of leadership, expressed as lack of imagination, lack of initiative, and lack of adaptive capacity (Waughn and Streib 2006).

The examples from the USA, Australia, New Zealand, and Japan all reflect an acute awareness of the political and fiscal risks and the recognition on the part of respective governments that there is an expectation on governments to demonstrate leadership and deliver on recovery needs. The changes, or transformation of governance that occurred following major disasters, testify to governments taking a leadership role. The newly introduced arrangements were designed to build and sustain institutional capacity for recovery by engaging across a broad sector of government and nongovernment organizations, to allow for formation and growth of partnerships through enabling processes, and to facilitate fueling of the recovery process with money and

information through coordination between different levels and stakeholders.

A closer look at the transformation of governance and arrangements for recovery from Canterbury, New Zealand, earthquakes and the role of the Canterbury Earthquake Recovery Authority (CERA) provides further insights into both requirements for and potential issues with recovery governance and leadership.

Strengths and Weaknesses of Different Approaches to Earthquake Recovery Governance and Leadership

Scholarship has identified that supportive governance is necessary to ensure coping capacities in societies. Governance determines the ways in which a variety of actors at all levels engage and coordinate their efforts in recovery. For engagement and coordination to lead to effective recovery, governance has to ensure broad participation, transparency, accountability, efficiency, and responsiveness (Chen et al. 2013).

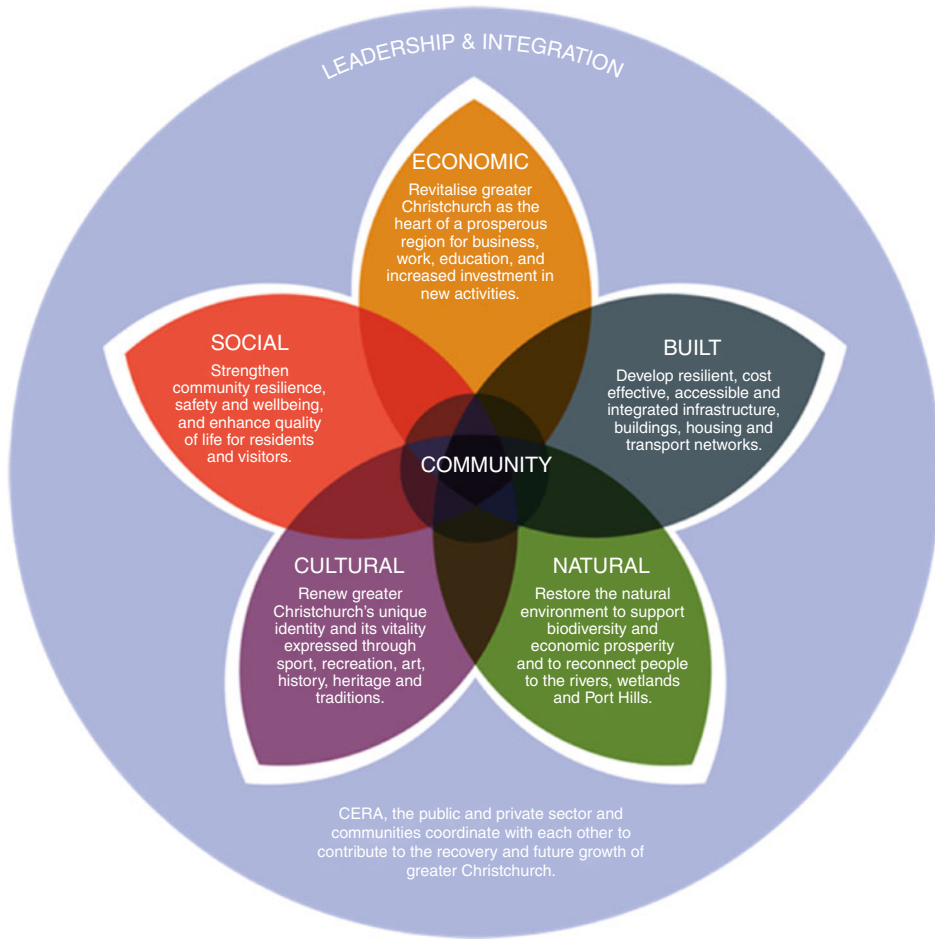
Two days after the September 4, 2010, earthquake in Canterbury, Prime Minister John Key announced the appointment to the new Cabinet position of the Minister for Canterbury Earthquake Recovery (CER Minister) and a newly appointed ad hoc Cabinet Committee on Canterbury Earthquake Recovery. Following this, on September 14, 2010, the Canterbury Earthquake Response and Recovery Bill was passed by the Parliament, with urgency and unanimously, and became an Act (CERR Act 2010). The Bill was exempt from the normal parliamentary examination procedures that precede the introduction of new legislation. Its purpose was to facilitate better coordination between impacted local authorities and the central government. The Act also established the Canterbury Earthquake Recovery Commission, comprised of the mayors of the three impacted local authorities and four government appointees (NZ Parliament 2010). Following the devastating earthquake on February 22, 2011, the ad hoc Cabinet Committee on Canterbury Earthquake Recovery began considering new national governance arrangements to manage the recovery effort. The rationale for the proposed change was given as experience from

historical and, especially, more recent events, such as the Victoria bushfires in 2009, Queensland floods in 2011, and Hurricane Katrina in 2005. The Committee recommended that a new, central government public service department be created to ensure leadership and coordination and to balance the central government's contribution to funding while carefully managing its fiscal situation. It was also recommended that normal parliamentary procedures and the normal timeline for establishing a new national department be waived given the emergency situation, allowing for the Canterbury Earthquake Recovery Authority (CERA) to be provisionally established the following day (Johnson and Mamula-Seadon 2014). At the same time, the CERR Act 2010 was repealed, including the dissolution of the Canterbury Earthquake Recovery Commission (CERC), and a new legislation was introduced, effectively upholding the authorities granted under the CERR Act, extending the expiration date of those authorities where appropriate and conferring on the Minister for CER and/or CERA a series of powers that would enable the timely and coordinated recovery of greater Christchurch. The Parliament agreed to an abbreviated legislative process and passed the Canterbury Earthquake Recovery (CER) Act 2011 with a near unanimous vote on April 14, 2011 (Johnson and Mamula-Seadon 2014).

One of the first tasks for the newly established CERA was to develop the Canterbury Recovery Strategy (CERA 2012a) that sets up the vision for the region's recovery and identifies key goals for the recovery. The strategy was developed through a consultative process with other government agencies, local government, and local communities. The key goals of the recovery process are community centered and expressed through the five components of recovery, pertaining to the social, cultural, natural, built, and economic environments, as shown in Fig. 2.

To effect delivery of the complex recovery tasks, the New Zealand government has proposed the following governance structure for CERA (Fig. 3).

As can be seen from Fig. 3, CERA, as a central government agency, and CERA's Chief Executive



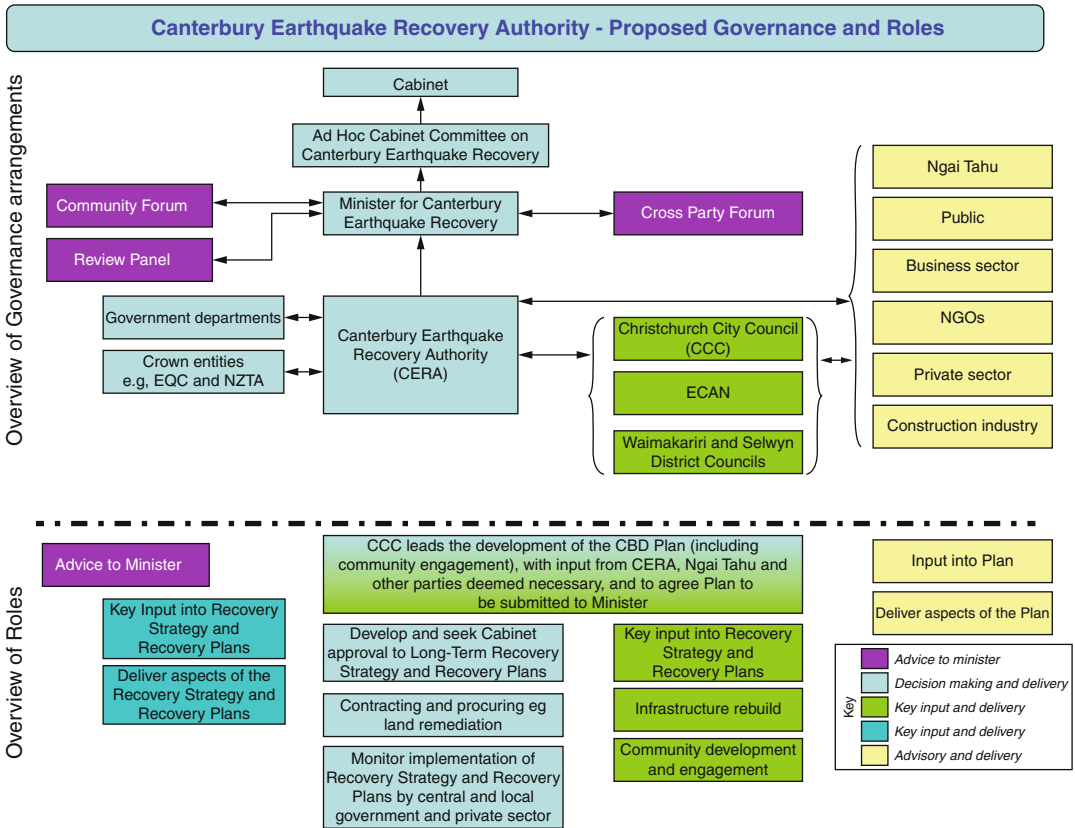
Earthquake Disaster Recovery: Leadership and Governance, Fig. 2 Key goals for the six components of recovery following Canterbury earthquakes in 2011 (CERA 2012a)

both report directly to the Minister for Canterbury Earthquake Recovery. A four-person Canterbury Earthquake Recovery Review Panel, appointed by the Minister, was tasked with reviewing and recommending Orders of Council, a mechanism allowing the government to make decisions requiring legal force by bypassing the standard legislative procedure. There is also a cross party parliamentary forum (CPPF), comprised of members of the Parliament from the Greater Christchurch area, established to provide the Minister with information and advice about the operation of the CER Act, and a community forum (CF) of at least 20 members to provide information and advice to the Minister that he must consider in carrying out his duties (CERA 2012b).

To a degree, the proposed recovery governance structure in Canterbury reflects recent societal and political changes, such as the emergence of public-private partnerships, outsourcing, and contracting, involving not only governmental institutions but also private sector and civil society entities, as discussed in research literature (Tierney 2012).

It also reflects the intent to deliver on findings of the World Bank that identified good recovery governance demonstrated during Queensland recovery effort as:

- Introduction of specific agencies to deal with recovery
- Commitment to community engagement



Earthquake Disaster Recovery: Leadership and Governance, Fig. 3 The proposed governance structure and roles for CERA (2012b)

- Government agencies working with the local government in recovery planning
- Incorporation of risk mitigation strategies into recovery planning
- Provision of advice to individuals
- Learning and improving preexisting arrangements (World Bank 2011)

When analyzing recovery governance arrangements pre-earthquake and post-earthquake, Johnson and Mamula-Seadon (2014) identify that the governance structures for managing recovery following the Canterbury earthquakes have transformed significantly at both the national and local levels, in comparison to preexisting arrangements. At the national level, recovery governance has transformed from a national service delivery coordination and support role for locally affected areas to one of increasing national-level

decision-making authority and operational responsibility for recovery activities. The findings demonstrate how post-disaster “time compression” can spotlight the strengths and weaknesses of recovery management processes and governance frameworks or cause distortions. The authors conclude that the effectiveness of vertical coordination of multiple levels of government, capacity building at the local and regional level, and public consultation and deliberation of key decisions are some critical areas of concern (Johnson and Mamula-Seadon 2014). The research revealed that over the course of the first 2 years following the February 22, 2011, earthquake, the transformation of governance structures for managing recovery has mainly focused on strengthening national-level centralization. The driving force behind this transformation was the need to make timely and effective decisions, add capacity and capabilities for the

recovery, and ensure accountability for the considerable public expenditure, among others (Johnson and Mamula-Seadon 2014). The research indicates that there are signs that the centralization may have helped to strengthen coordination among national agencies and expedite policy and decision-making; but the effectiveness of coordination among multiple levels of government, capacity building at the local and regional levels, and public engagement and deliberation of key decisions are some areas where the transformations may not have been as effective. At the local level, recovery governance transformed from a locally led and collaborative approach to a supportive role with, in some cases, little to no involvement in recovery decision-making or implementation. The authors conclude that long-term outcomes of the CERA model and the Canterbury recovery governance transformations require better understanding. They also provide some cautionary notes for future disaster recovery policy making and the development of effective and sustainable integrated recovery governance models for large-scale disaster recovery (Johnson and Mamula-Seadon 2014).

Reflecting on governance and governance transformation that often occurs in recovery, it is important to bear in mind that the last decade or so has seen a dramatic change in the way public service is delivered. Kapucu (2012) observes that the most important reform has occurred not only in the range of agencies involved but in the tools and forms of service delivery. Government agencies are not anymore the sole providers of services that traditionally were considered “public.” In the current environment, often nonprofit and for-profit agencies as well as ordinary citizens have become the stakeholders taking on the roles and responsibilities of service provision at all stages of the process (Kapucu 2012). The term governance depicts the complex web of interactions and interdependencies among a multiplicity of actors that deliver specific services to citizens (Kapucu 2012). The term is also synonymous with collaborative governance and collaborative public management, formed to deliver on complex problems and describe network relationships across different sectors and different levels of the government, all brought together for the purpose of delivering

public good. The mode of operation presupposes consensus-oriented and deliberative processes, with shared decision-making directed toward shaping public policy (Kapucu 2012).

Thus, one of the themes when considering governance structures, and particularly so in the recovery context, is the necessity to understand and address the state-civil society relationships.

Recovery Governance, Leadership, Social Capital (or Social Capacities), and Resilience

Different forms of governance including authority, power, influence, degree of decentralization, and accountability affect disaster recovery. It is important to understand, for instance, how policy initiatives with local communities can be influenced by the way local and national governments work together and how larger trends in national policies affect the capacities of communities to engage with their governmental agencies. Defining the forms of governance that take the energy and commitment individuals and communities show during an emergency and sustain it through appropriately aligned government efforts, especially in response and recovery situations, is one of the key challenges of successful recovery (MNRPG 2013). Furthermore, if successful recovery requires informed, coordinated, and sustained leadership throughout all levels of the government and community and phases of the recovery process (FEMA 2011), it is paramount to design governance structures that foster meaningful engagement between the state and civil society (MNRPG 2013).

Understanding social capital and the power of social groups to organize, generate, and deploy individuals within a community to take decisive action when facing hardships is one of the critical points in recovery governance. Leadership at all levels, including spontaneous, emergent leadership typical of disasters, requires governments to be able to anticipate and adapt. In nearly every national experience, policy leaders have encountered situations in which community members have spontaneously organized and begun to respond and recover to a disaster well before government agencies and programs have mobilized (MNRPG 2013). It is important to understand how various forms of governance align with these

mobilizations that occur largely outside the control and direction of particular agencies. To the extent that community leaders are the potential allies and partners for government authorities, governments need to learn to identify and work with them – and even help create effective leaders (MNRPG 2013).

State-civil society interactions are complex and take many forms – some nations have strong civil society sectors, while in other nations civil society institutions are nascent, weak, or unable to function openly (Tierney and Oliver-Smith 2012). Recent scholarship clearly links the effectiveness of the recovery process to levels of trust and social capital, including social networks or ties, the degree of political prowess and organization, and the degree to which social networks are embedded into recovery processes (Aldrich 2011). There are many examples of successful state-civil society partnership in earthquake recovery, in Haiti, Taiwan, and China, to mention a few (Tierney and Oliver-Smith 2012; Chen et al. 2013).

When, following the earthquake on February 22, 2011, the citizens of Christchurch were asked to identify what makes them resilient, they singled out the ability to take action and engage with others in determining their future as one of the key ingredients for resilience and successful recovery. They felt that the emergent, spontaneous community leaders were the catalyst for the process. When asked to summarize what transformed community members into community leaders and made them so readily accepted by other members of the community, the participants responded that these leaders were found to have local knowledge and connections and have the ability to listen to the community and had a good grasp of community dynamics and needs; they acted on their own initiative and had little need for formal structure in order to be effective; those leaders were described as able to function with little resource, to be flexible, and able to adapt quickly. They were skilled, open minded, selfless, and focused on the well-being of the community. Whereas those leaders were recognized (and subsequently awarded) by the community, the governance mechanisms that had flexibility and adaptability needed to actively

engage community initiatives in the early stages of recovery were seen as lacking (Mamula-Seadon et al. 2012).

Following on the discussion of social capacities (capital), it can be argued that the strength of local relationships is expressed as the degree of prevalent, strong, and sustained-over-time involvement of local groups and individuals in the recovery decision-making process and is a measure of horizontal integration within the governance framework (Smith and Birkland 2012). Participants may include local government officials, nongovernment groups, citizen groups, business owners, local financial institutions, the media, and individual community members. The empowering process usually involves deliberative planning and policy-making strategies at the community level. Recovery governance with poor horizontal integration lacks stakeholder involvement (Smith and Birkland 2012). Vertical integration is defined by strong linkages to state and national organizations, including national foundations, corporations, national lending institutions, and insurance companies (Smith and Birkland 2012). Integrating social capital, ensuring meaningful deliberation and engagement with affected communities, achieving effective vertical coordination with multiple levels of government and other stakeholders, and capacity building at the local and regional level are among key factors for effective recovery governance and leadership.

International policy makers identified that, to a large extent, the challenge of building disaster resilience is a matter of democratic governance. It involves partnering with communities, building mutual support within communities and across jurisdictional boundaries, and sustaining involvement. In general, the work of building resilience, both pre- and post-disaster, demands cooperation among citizens and between subnational and national levels of the government and integration of both the public and private sectors and thus closely resembles the collaborative nature of community recovery. Furthermore, research and policy making have been advanced by views of resilience as a construct formed through the interdependencies that evolve from established

societal patterns and not as a replica of institutional, group, or program arrangements (MNRPG 2013). Governance for recovery requires leadership capable of transforming historical hierarchical government models into adaptable, integrated, networked, multi-stakeholder environments with deliberative and collaborative models of state-civil society – in other words effective governance for recovery is synonymous with effective governance for resilience.

Summary

The shift in disaster management practice toward more integrated and comprehensive, all-hazards approaches based on principles of sustainable development has seen an introduction of frameworks and processes with a shared system of governance and policy making. Common and overlapping responsibilities are apportioned among layers of government and other nongovernment entities in democratic governance systems with decentralized and deliberative planning and decision-making processes. National frameworks, governance models, and administrative guidelines on recovery management reflect the holistic, comprehensive, integrated risk management approach based on the principles for sustainable development and decentralized, “bottom-up” structures, with local governments having primary responsibility for supplying disaster-related resources and regional, subnational, and national agencies providing support as requested. Recovery is considered a continuum that involves both pre-disaster and post-disaster activities. Recovery preplanning is as crucial as risk-reduction and resilience-building activities that link recovery to other aspects of the comprehensive risk management framework. To a varying degree recovery governance structures also reflect recent societal and political changes, such as the emergence of public-private partnerships, outsourcing, and contracting, involving not only governmental institutions but also private sector and civil society entities.

Disaster governance is often defined as a system that enables collaborative activities that

bring together multiple organizations, including the public sector, nongovernmental organizations, nonprofit relief organizations, the private sector, and the community. The degree of coordination and integration, underlying values and processes, timing, capacity, and capability of those actors are some of the key factors that determine effectiveness of recovery processes.

The immediate post-disaster environment creates enormous pressures on governments and societies that require complex rebuilding and societal activities to happen in a compressed period of time. Governments are acutely aware of political and fiscal risks and the expectation to demonstrate leadership and deliver on recovery needs. They rally to provide and unify communities around a shared vision for recovery. Faced with insufficient capacity to solve problems, make timely and effective decisions, add capacity and capabilities for the recovery, ensure accountability for the considerable public expenditure, deliver on excess demands for information and stakeholder deliberation, greater needs for organizational integration and coordination, and immediate demands for large amounts of money with existing distribution systems challenged to meet the needs, governments often struggle. As a result, this may lead to significant transformation in governance. The new arrangements are designed to build and sustain institutional capacity for recovery by engaging across a broad sector of government and nongovernment organizations, to allow for formation and growth of partnerships through enabling processes, and to facilitate fueling of the recovery process with money and information through coordination between different levels and stakeholders. Supportive governance embracing of stakeholder engagement is necessary to ensure coping capacities in societies. For engagement and coordination to lead to effective recovery, governance has to ensure broad participation, transparency, accountability, efficiency, and responsiveness.

Understanding social capital and the power of social groups to organize, generate, and deploy individuals within a community to take decisive action when facing hardships is one of the critical points in the future of recovery governance

and leadership. Integrating social capital, ensuring meaningful deliberation and engagement with affected communities, achieving effective vertical coordination with multiple levels of government and other stakeholders, and capacity building at the local and regional level are among the key factors for effective recovery governance and leadership. An adaptable, integrated, networked, multi-stakeholder environment with deliberative and collaborative models of state-civil society governance for recovery is synonymous with effective governance for resilience.

Cross-References

- ▶ “Build Back Better” Principles for Reconstruction
- ▶ Community Recovery Following Earthquake Disasters
- ▶ Learning from Earthquake Disasters
- ▶ Legislation Changes Following Earthquake Disasters

References

- Aldrich DP (2011) The power of people: social capital's role in recovery from the 1995 Kobe earthquake. *Nat Hazards* 56:595–611, Springer Science
- Australian Government Attorney-General's Department (2011) Community recovery handbook 2. The Australian Emergency Management Institute, part of the Commonwealth Attorney-General's Department, Canberra
- CERA, Canterbury Earthquake Recovery Authority (2012a) Recovery strategy for greater Christchurch: Mahere Haumanutanga o Waitaha. CERA. <http://cera.govt.nz/recovery-strategy/overview/read-the-recovery-strategy>
- CERA, Canterbury Earthquake Recovery Authority (2012b) Proposed governance arrangements (cabinet paper). <http://cera.govt.nz/sites/cera.govt.nz/files/common/cabinet-paper-1-proposed-governance-arrangements-annex-2-march-2011.pdf>
- Chen J, Chen THY, Vertinsky I, Yumagulova L, Chansoo P (2013) Public-private partnership for the development of disaster resilient communities. *J Conting Disaster Manag* 21(3):130–143, Wiley
- FEMA, Federal Emergency Management Agency (2011) National disaster recovery framework: strengthening disaster recovery for the nation. Federal Emergency Management Agency. <http://www.fema.gov/pdf/recoveryframework/ndrf.pdf>
- Fraser-Molekati GJ (2012) Democratic governance at times of crisis: rebuilding our communities and building on our citizens. *Int Rev Adm Sci* 78(2):191–208, Sage
- GeoNet (2013) Aftershocks – our news – GeoNet. GeoNet: EQC and GNS Science. <http://info.geonet.org.nz/display/home/Aftershocks>
- UK Government, Cabinet Office (2013) Emergency response and recovery guideline. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/253488/Emergency_Response_and_Recovery_5th_edition_October_2013.pdf
- Johnson L, Mamula-Seadon L (2014) Transforming governance: organizing for and managing disaster recovery following the September 4, 2010 and February 22, 2011 Canterbury earthquakes. *Earthq Spectra* 30(1):577–605, special edition Canterbury Earthquakes
- Johnson LA, Olshansky TB (2013) The road to recovery: governing post-disaster reconstruction. Land Lines, Lincoln Institute of Land Policy, Cambridge, MA, pp 14–21
- Kapucu N (2012) Disaster and emergency management systems in urban areas. *Cities* 29:S41–S49, Elsevier
- Mamula-Seadon L, Paton D, Selway K (2012) Exploring resilience: learning from Christchurch communities, in Community resilience: the Christchurch experience. MCDEM Wellington, Tephra. [http://www.civildefence.govt.nz/memwebsite.nsf/Files/Tephra-vol23/\\$file/Tephra2012web.pdf](http://www.civildefence.govt.nz/memwebsite.nsf/Files/Tephra-vol23/$file/Tephra2012web.pdf)
- MCDEM, Ministry of Civil Defence & Emergency Management (2005) Focus on recovery: a holistic framework for recovery in New Zealand. Director's Guidelines, Ministry of Civil Defence & Emergency Management, Wellington. [http://www.civildefence.govt.nz/memwebsite.nsf/Files/FocusOnRecovery/\\$file/FocusOnRecovery.pdf](http://www.civildefence.govt.nz/memwebsite.nsf/Files/FocusOnRecovery/$file/FocusOnRecovery.pdf)
- MNPRG The Multi-national Resilience Policy Group (2013) The New Zealand discussion, notes from the meeting. Wellington/Christchurch
- New Zealand Parliament (2010) Canterbury Earthquake Response and Recovery Act 2010. Available at <http://www.legislation.govt.nz/act/public/2010/0114/latest/DLM3233004.html>
- Olshansky RB, Hopkins LD, Johnson LA (2012) Disaster and recovery: processes compressed in time. *Nat Hazards Rev* 13(3):173–178
- Smart J (2012) The role of post-disaster institutions in recovery and resilience: a comparative study of three recent disasters – Victorian Bushfires (2009), Queensland Floods (2010–2011), and Canterbury Earthquakes (2010–2012). Working paper 12/02. Institute for Governance and Policy Studies, Victoria University Wellington
- Smith G (2011) Planning for post-disaster recovery: a review of the United States disaster assistance framework. Public Entity Risk Institute, Fairfax

- Smith G, Birkland T (2012) Building a theory of recovery: institutional dimensions. *Int J Mass Emerg Disasters* 30(2):147–170
- Suzuki I, Kaneko Y (2013) Japan's disaster governance how was the 3.11 crisis managed? Springer, New York/Heidelberg/Dordrecht/London
- Tierney K (2012) Disaster governance: social, political, and economic dimensions. *Annu Rev Environ Resour* 37:341–363
- Tierney K, Oliver-Smith A (2012) Social dimensions of disaster recovery. *Int J Mass Emerg Disasters* 30(2):123–146
- Waughn WL, Streib G (2006) Collaboration and leadership for effective emergency management. *Public Adm Rev* 66(Spec Issue):131–140
- White House (2012) Executive order establishing Hurricane Sandy Rebuilding Task Force. <http://www.whitehouse.gov/the-press-office/2012/12/07/executive-order-establishing-hurricane-sandy-rebuilding-task-force>
- World Bank (2011) Queensland recovery and reconstruction in the aftermath of the 2010/2011 flood events and cyclone Yasi, report. The International Bank for Reconstruction and Development, Washington, DC, USA and Queensland Reconstruction Authority, Australia

Earthquake Early Warning System in Taiwan

Yih-Min Wu¹, Nai-Chi Hsiao², Tai-Lin Chin³,
Da-Yi Chen², Ya-Ting Chan¹ and
Kai-Shyr Wang⁴

¹Department of Geosciences, National Taiwan University, Taipei, Taiwan

²Central Weather Bureau, Taipei, Taiwan

³Department of Computer Science and Information Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan

⁴Ministry of Science and Technology, Taipei, Taiwan

Introduction

Taiwan is located on the plate boundary where Philippine Sea Plate collides with Eurasian Plate with a very high plate convergence rate of 8.0 cm/year (Yu et al. 1999; Fig. 1). Nearly 18,000 seismic events occur around Taiwan and the vicinity every year (Wu et al. 2008).

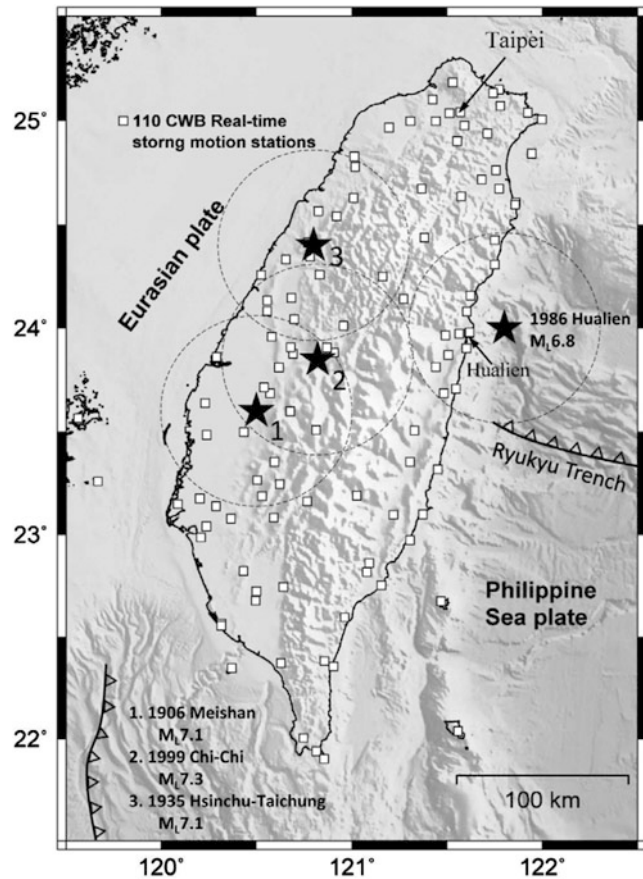
Some of the destructive earthquakes caused large casualties, such as the 1906, $M_L = 7.1$ Meishan earthquake (1,258 death); the 1935, $M_L = 7.1$ Hsinchu-Taichung earthquake (3,276 death); and the 1999, $M_L = 7.3$ Chi-Chi earthquake (2,455 death). Earthquake is one of the most serious disasters in Taiwan. Therefore, it is important to seek scientific resolutions to alleviate earthquake hazards.

The earthquake early warning (EEW) system (Wenzel and Zschau 2013; “► [Early Earthquake Warning \(EEW\) System: Overview](#)”) is considered as a useful tool for real-time seismic hazard mitigation. An EEW system provides a few seconds to tens of seconds of advanced warning time for the damage shakings. It allows some mitigation reactions to be taken in a short time. For instance, high-speed trains can slow down to resist strong ground shakings, gas pipelines can be turned off automatically, elevators can be moved to nearby floor and open the doors, surgeries in hospitals can be suspended in time, etc. The early warning time may be too short for people to evacuate from buildings. However, through careful planning, training, and drills, an educated general public can still take necessary response in time to avoid loss of lives from large earthquakes. Two types of EEW systems are operated in the world. One is a front-detection EEW system in which seismometers installed in the earthquake source area give early warnings for more distant urban areas. The other is an on-site EEW system which determines the earthquake parameters from the initial portion of the P waves and predicts the more severe ground shakings of the following S wave trains (Kanamori 2005; Wu and Kanamori 2005a).

Motivation for Taiwan to develop the EEW is the painful experience of the 15 November 1986 Hualien, Taiwan, earthquake of M_L 6.8 (or M_W 7.8). Although the epicenter of that earthquake was located near Hualien, the most severe damage occurred in the Taipei metropolitan area. The distance between Hualien and Taipei is about 120 km. The shear waves traveling over this distance should take more than 30 s. Thus, if a seismic monitoring system can reliably determine the

Earthquake Early Warning System in Taiwan,

Fig. 1 Distribution of the real-time strong-motion stations (*squares*) of the Central Weather Bureau (CWB) and sample virtual subnetwork (VSN) with 60-km radius from the earthquake early warning (EEW) system. *Stars* show epicenters of the four major damage earthquakes



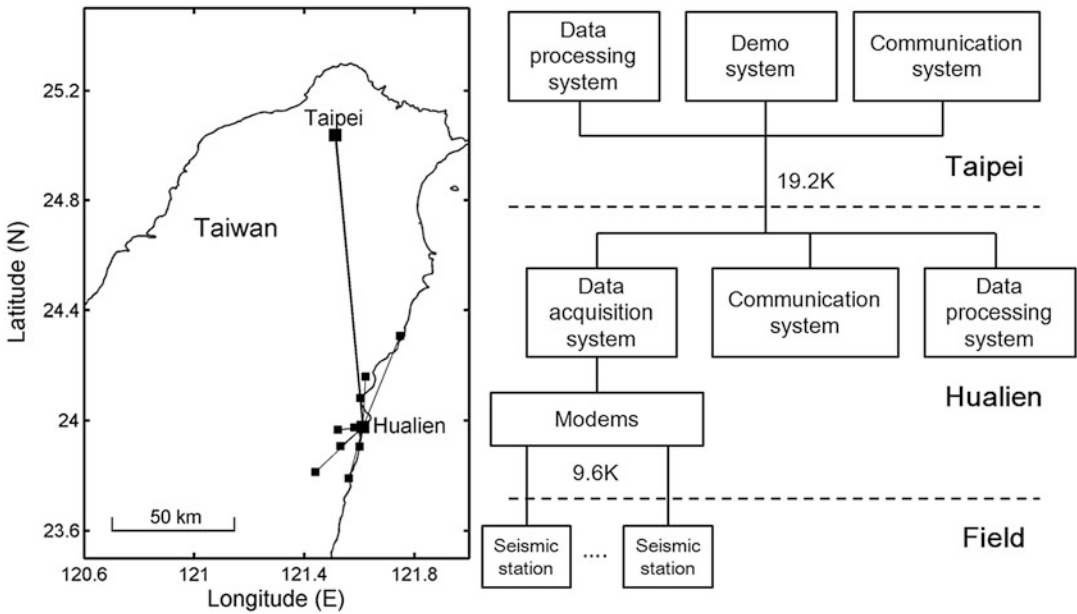
E

location and magnitude within 20 s for a large earthquake that could threaten the metropolitan area, then more than 10 s will be available for emergency response. Consequently, the Central Weather Bureau (CWB), Taiwan, started to develop an experimental system to explore the technology of EEW for the highly seismic area of Hualien in 1993 (Wu et al. 1999). After 20 years efforts, the EEW system is becoming one of the important tools for seismic hazard mitigation in Taiwan, and Taiwan is one of the leading countries in developing EEW system. Progress of the EEW systems in Taiwan will be reported in this entry.

The Hualien Experimental EEW System

The Hualien experimental EEW system is a front-detection type prototype, which was implemented in Hualien area to explore the use

of modern technologies for earthquake early warning purposes by the CWB in 1993. This system consisted of 10 three-component force balanced accelerometer stations (distributed in an aperture of about 20 km) which were telemetered digitally via 9600-baud telephone lines to the CWB Hualien Station. At the Hualien Station, the incoming digital signals are processed in real time and the results are telemetered to the CWB Headquarters in Taipei (Fig. 2). A computer in Hualien handles the communication with Taipei via a dedicated telephone line (19.2 k baud). A parallel computer system in Taipei is used for displaying, processing, and analyzing the received data. It also allows operators in Taipei to access, manage, and control the system in Hualien. A commercial company implemented both the hardware and software under the CWB specifications and directions. The goal of this system was to detect, locate,



Earthquake Early Warning System in Taiwan, Fig. 2 A prototype earthquake early warning system in Hualien. This system consists of 10 three-component accelerometers telemetered digitally via 9600-baud

telephone lines to the CWB Hualien Station. At the Hualien Station, the incoming digital signals are processed in real time and the results are telemetered to the CWB Headquarters in Taipei

and transmit the source information to Taipei within 10 s after the occurrence of a strong earthquake in Hualien. This information could then be disseminated in Taipei prior to the arrival of the strong ground shaking (Lee et al. 1996).

During the test period from May 1994 to October 1994, 32 earthquakes with magnitude ranging from 3.1 to 6.5 were detected. The performance of this system has achieved the goal of rapid detection. It could provide information of earthquake around the Hualien region at about 10 s after the earthquake occurrence. It may offer more than 10 s early warning time to Taipei metropolitan area. However, this dense array cannot provide good station coverage. It results in poor earthquake localization and leads to large uncertainties in magnitude determination (Wu et al. 1999). Base on this experience and the real-time strong-motion signals used for felt earthquake monitoring since 1995, the CWB started to develop EEW system using a telemetered seismic network of digital accelerographs distributed over the entire island (Teng et al. 1997; Wu et al. 1997, 1999).

Front-Detection EEW Systems of the CWB

The Taiwan CWB has utilized its real-time strong-motion network since 1995 (Teng et al. 1997; Wu et al. 1997, 1998) as a basis for the development of rapid earthquake information reporting and EEW capabilities. This real-time strong-motion system consists of 110 telemetered stations (Fig. 1) distributed throughout Taiwan, an area of 100×300 km. Each station has three-component force-balanced accelerometers with signals digitized at 50 samples per sec per channel at 16-bit resolution and telemetered to the CWB Headquarters in Taipei. The full recording dynamic range is ± 2 g, and the sensitivity is sufficient to record events with $M > 4.0$ at distance of 100 km or more.

In order to reduce the earthquake reporting time, a virtual subnetwork (VSN; Wu and Teng 2002) approach and an M_{L10} (Wu et al. 1998) method based on the front-detection early warning approach were utilized in the first stage. The experiment in Hualien, Taiwan, has

demonstrated that earthquake reporting time can be significantly reduced by using a smaller network (Wu et al. 1999). This leads to the design and configuration of a VSN approach. The VSN, automatically configured by the monitoring system, is event-dependent and its configuration varies with time. In hypocenter and magnitude determinations, only stations close to the epicenter (less than 60 km) contribute to the crucial information. Within the framework of the real-time strong-motion network, only signals from a subset of stations are chosen to process and form a VSN network surrounding an event. As soon as the system is triggered by an event, the system automatically extracts signals from a subset of stations and configures a VSN with a 60-km radius centered on that event. Figure 1 also gives a number of possible VSN configurations; each normally consists of about a dozen stations. The extracted data stream for this event forms the basic VSN input data for the subsequent EEW work. By working with the VSN, the reporting time could be substantially reduced, which makes it possible to cover the entire Taiwan region with effective earthquake early warning.

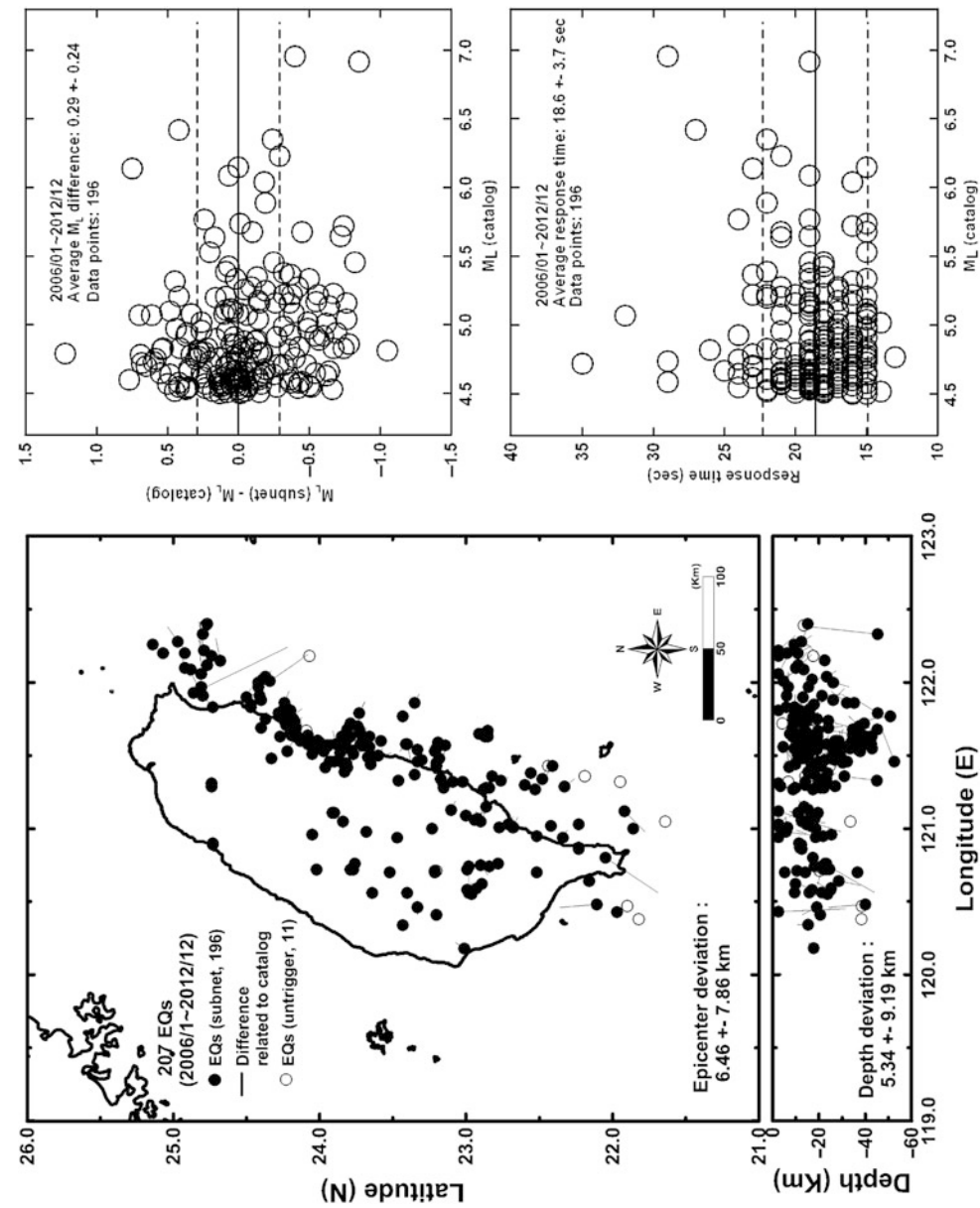
Two major requirements for an EEW system are near real-time estimation of the earthquake location and magnitude. The first requirement on rapid location can be achieved readily in a 10-s time window immediately following the first arrival of P waves. On the other hand, the second requirement of rapid determination of the earthquake magnitude would be more difficult because the shear-wave trains may not be recorded completely within this time window. Therefore, a method for quick magnitude determination for large events needs to be developed. The M_{L10} (Wu et al. 1998) is one of the approaches to determine M_L using first 10 s of the signals. Signals collected by the stations within a 60-km radius from an event are grouped and extracted from the VSN input, which will then be processed in parallel through the VSN software in a dedicated computer. A series of experiments have been conducted to determine the optimum recording time for a 60-km radius network. Results show that 10 s is about the optimum timing. As soon as the 10-s waveforms are

presented at the VSN system, they will be immediately processed to give simulated Wood-Anderson seismograms for magnitude determinations as M_{L10} . The earthquake magnitude M_L cannot be determined in the same time frame due to the incomplete recording of shear waves at some stations. However, the magnitude based on the first 10 s of the signals, i.e., M_{L10} , can be found to correlate with M_L as follows:

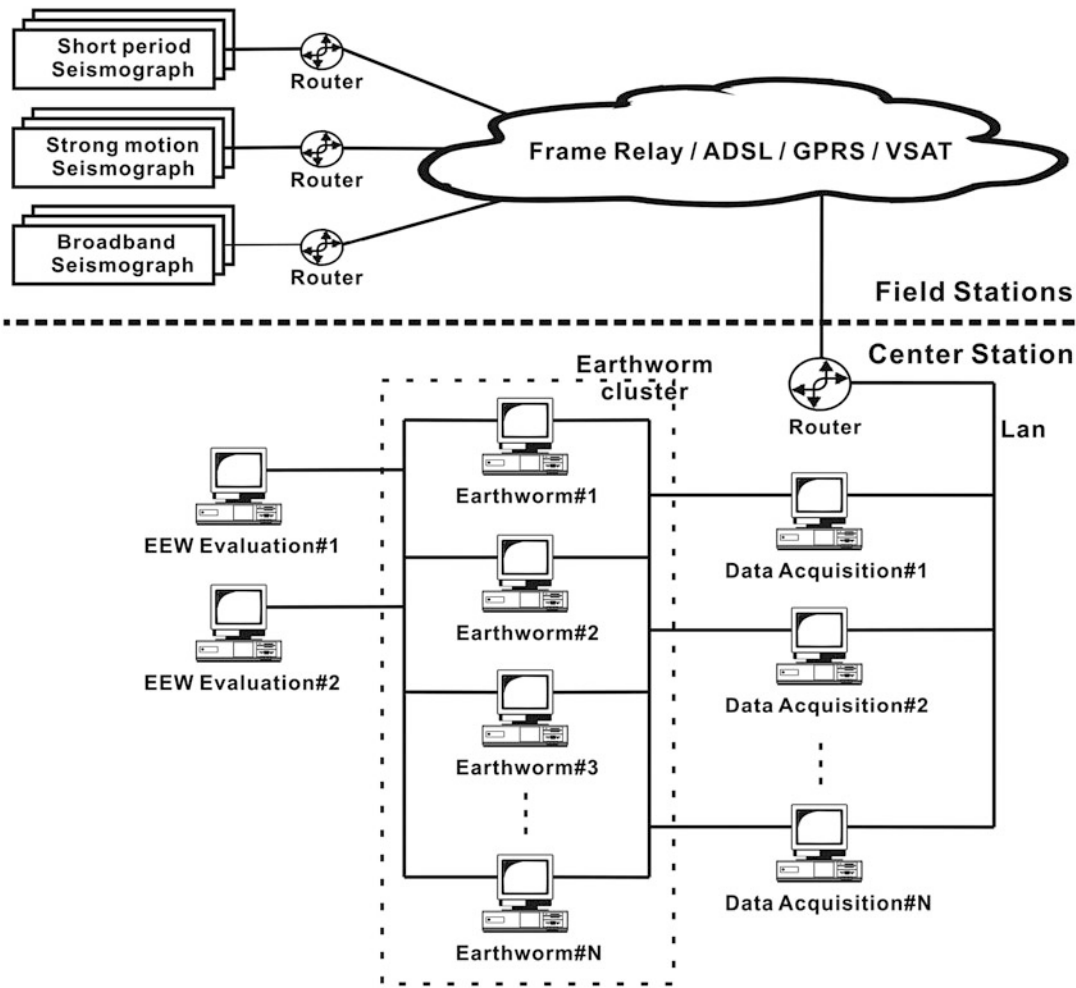
$$M_L = 1.28 \times M_{L10} - 0.85 \pm 0.13 \quad (1)$$

By applying Eq. 1 to magnitude determination, the VSN system can determine hypocenter and magnitude with tolerable uncertainty in about 20 s after the occurrence of an earthquake and early warning is thus possible in Taiwan.

During the period from 2006 to 2012, a total of 207 earthquakes with magnitude M_L ranging from 4.5 to 6.9 occurred on land or within 50 km from the coastlines of the Taiwan Island (Fig. 3). A total of 196 earthquakes were detected, processed, and reported in real time. Detecting successful rate is about 95 %. Only 5 % offshore and magnitude smaller than 5.0 events could not be detected automatically. If the off-line manual measurements give correct values, the VSN results give average errors of 6.5 ± 7.9 km and 5.3 ± 9.2 km in epicenter and focal depth, respectively. The error of magnitude is 0.3 ± 0.2 units. The reporting time is 18.6 ± 3.7 s (Fig. 3). Large uncertainties in the estimated magnitude using the initial portion of the seismograms are inevitable. Similar results are also presented in many other studies (e.g., Nakamura 1988; Grecksch and Kumpel 1997; Allen and Kanamori 2003; Kanamori 2005; Wu and Kanamori 2005a, 2008; Wu et al. 2007). For larger offshore earthquakes, the VSN method may underestimate the magnitudes due to the limited intervals of the waveform signals used. To avoid this problem, the magnitude M_T obtained from the average period τ_c (Kanamori 2005) and the dominant period $\tau_p^{\max}$ (Allen and Kanamori 2003) of the initial P waves may offer a satisfactory solution (Hsiao et al. 2009). However, the VSN system has achieved a short earthquake reporting time about 20 s. This can offer



Earthquake Early Warning System in Taiwan, Fig. 3 Comparison of the real-time automatic event locations and magnitudes of the EEW system (subnet) and the manual off-line event locations of the CWB catalog. *Right-bottom* figure shows the EEW reporting times versus magnitudes



Earthquake Early Warning System in Taiwan, Fig. 4 Schematic diagram of the new prototype system for earthquake early warning in the CWB

earthquake early warning for metropolitan areas located 70 km from the epicenter and beyond.

In order to provide warning for regions within 70 km from the epicenter, a new prototype EEW system using a P-wave method is developed using a real-time seismographic network currently in operation in the CWB of Taiwan. This system is based on the Earthworm environment which is capable to carry out integrated analysis of real-time broadband, strong-motion, and short-period signals. The Earthworm system distributed by the United States Geological Survey (USGS) is developed as a multifunctional seismic data processing system. Figure 4 shows the configuration of

prototype EEW system. Real-time seismic signals are packaged and transmitted to the headquarters via various IP-based networks, such as Frame Relay, ADSL, GPRS, or satellite telemetry. Different telemetered networks can be arranged as a secure environment for seismic data transmission. A cluster of computers running the Earthworm system is installed at the central station in Taipei. Instead of M_{L10} , the “ P_d magnitude,” M_{Pd} (Wu and Zhao 2006; Hsiao et al. 2011), is adopted as magnitude indicator in the new system in order to shorten reporting time. P_d is defined as the peak amplitude of the initial P-wave displacement (Wu and Kanamori 2005b). Incoming signals are

processed in real time. When a large earthquake occurs, P-wave arrival times and P_d will be estimated for location and magnitude determinations for EEW purpose. In a test of 54 felt earthquakes, this system can report earthquake information in 18.8 ± 4.1 s after the earthquake occurrence with an average difference in epicenter locations of 6.3 ± 5.7 km and an average difference in depths of 7.9 ± 6.6 km from catalogs. The magnitudes approach a 1:1 relationship to the reported magnitudes with a standard deviation of 0.5 (Hsiao et al. 2011).

VSN system is the current system operated by CWB. When the new system performs as well as the VSN system, it will become one of the systems operating together with the VSN. Currently, the new system is capable of offering earthquake warning information in less than 10 s after the occurrence of events inland. Therefore, the new system will become the major EEW system of the CWB in the future.

A Hybrid of Both On-Site and Front-Detection EEW System Using Low Cost Accelerometers

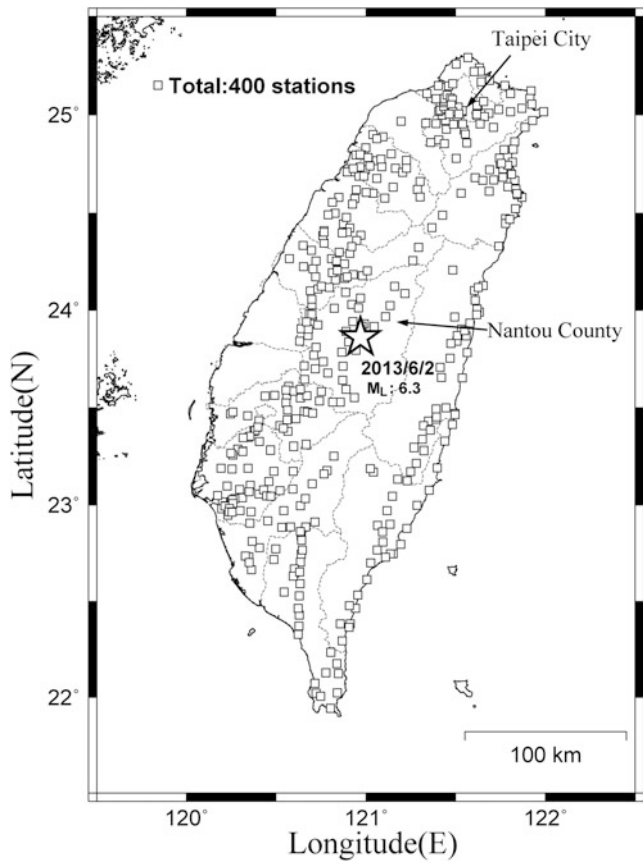
The EEW system of the CWB can provide earthquake information within 20 s following an earthquake occurrence. Although a 20-s reporting time is already very short, the reporting time can still be reduced if the density of seismic stations in the network can be increased. However, to build such a high-density seismic network by traditional force-balance seismometers, the cost is extremely high. In contrast, Micro Electro Mechanical Systems (MEMS) accelerometers introduced in seismic applications (Holland 2003) are ideal and cost-saving devices for recording strong ground motions. The EEW research group at National Taiwan University (NTU) worked with a technology corporation to develop a P-wave alert device, named “Palert,” that uses MEMS accelerometers for on-site and front-detection EEW purposes. The cost of a Palert device is less than 1/10 of the cost for a traditional accelerometer. The Palert device can record real-time, three-component signals, and

the signal resolution is 16 bits with a -2 g to $+2$ g range. The sampling rate is 100 samples per second. Once an earthquake is detected by the trigger algorithm stored in a Palert, a corresponding earthquake watch or warning alarm is sent by the device. The major trigger algorithms include procedures for continuous monitoring peak ground acceleration (PGA) and the peak amplitude of the filtered (high pass at corner 0.075 Hz) vertical displacement of the P wave (e.g., P_d). Palert also has networking capabilities for streaming real-time acceleration signals to the data collection servers and synchronizing clocks by the Network Time Protocol (NTP). Therefore, it is possible to connect Palert devices to build a front-detection EEW system.

Since 2010, NTU began to build up the Palert real-time strong-motion network. A total of 400 stations were deployed and configured in 2013. Figure 5 provides the distribution of the Palert stations. Most of the stations are located in elementary schools where power and Internet connections are available. Figure 6 provides the configuration of the Palert network. At the field site, real-time signals are processed by the Palert devices. Three-component acceleration signals are processed on-site for detecting P-wave arrival and continuously double integrated to displacement signals for calculating P_d . The real-time signals are sent to the central stations every second simultaneously via the Internet. When an earthquake occurs, the Palert automatically detects P-wave arrival, and once P_d or PGA are larger than 0.35 cm or 80 gal, respectively, the Palert device begins its alert with a warning sound on site. Tasks of the central stations include data clustering and processing. For real-time seismic data processing, real-time signals are processed and stored in the Earthworm system. In the Earthworm system, signals are again processed in order to detect the arrival of the P wave (Hsiao et al. 2011; Chen et al. 2012). P-wave information detected by the Earthworm platform is also sent to a shared memory for later fusion process. During the final stage, once eight Palert stations are triggered, an event is declared. The TcPd.c program fetches the event parameters stored within the shared memory and computes

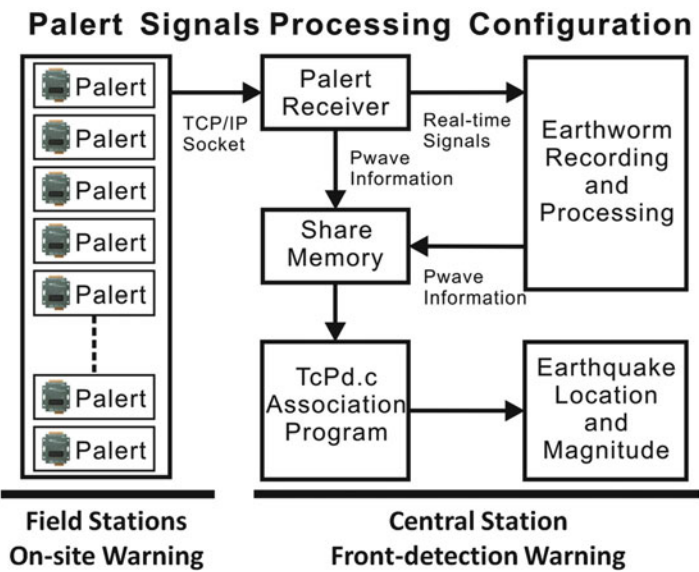
Earthquake Early Warning System in Taiwan,

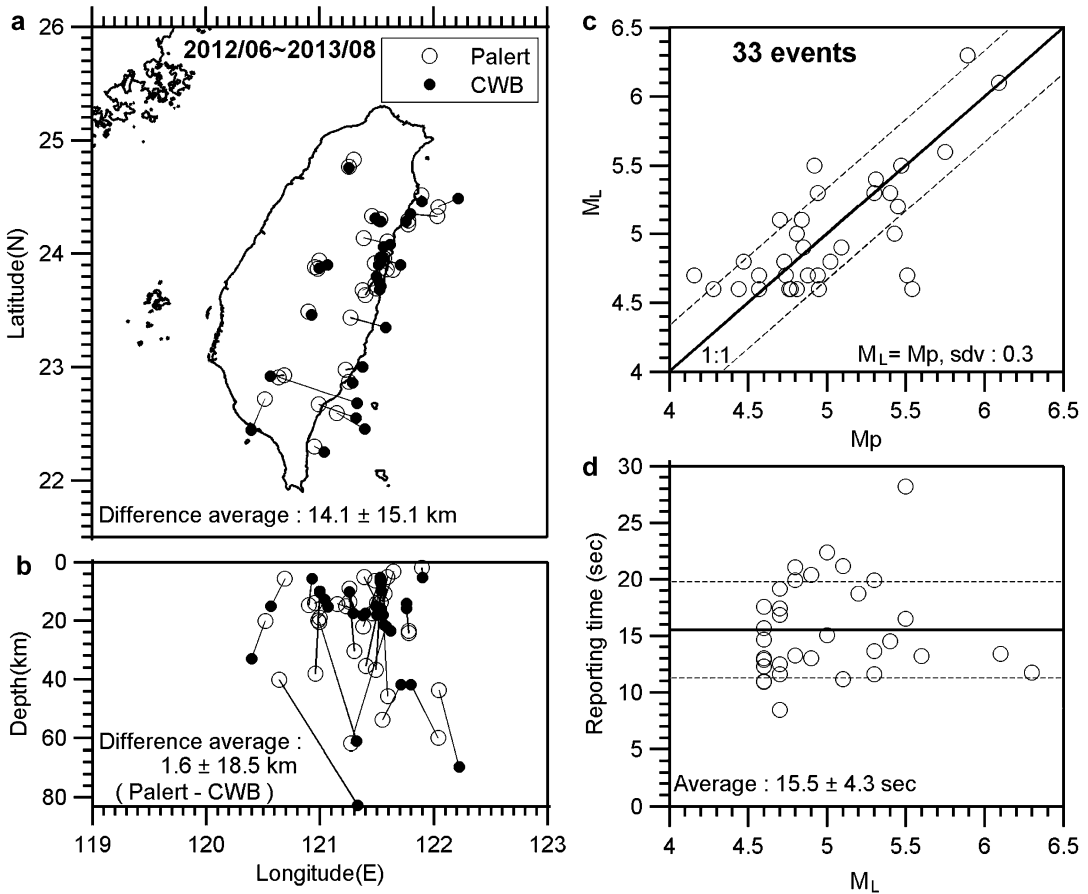
Fig. 5 Distribution of the Palert stations. Star shows the epicenter of the 2 June 2013 Nantou M_L 6.3 earthquake



Earthquake Early Warning System in Taiwan,

Fig. 6 Configuration of the Palert network





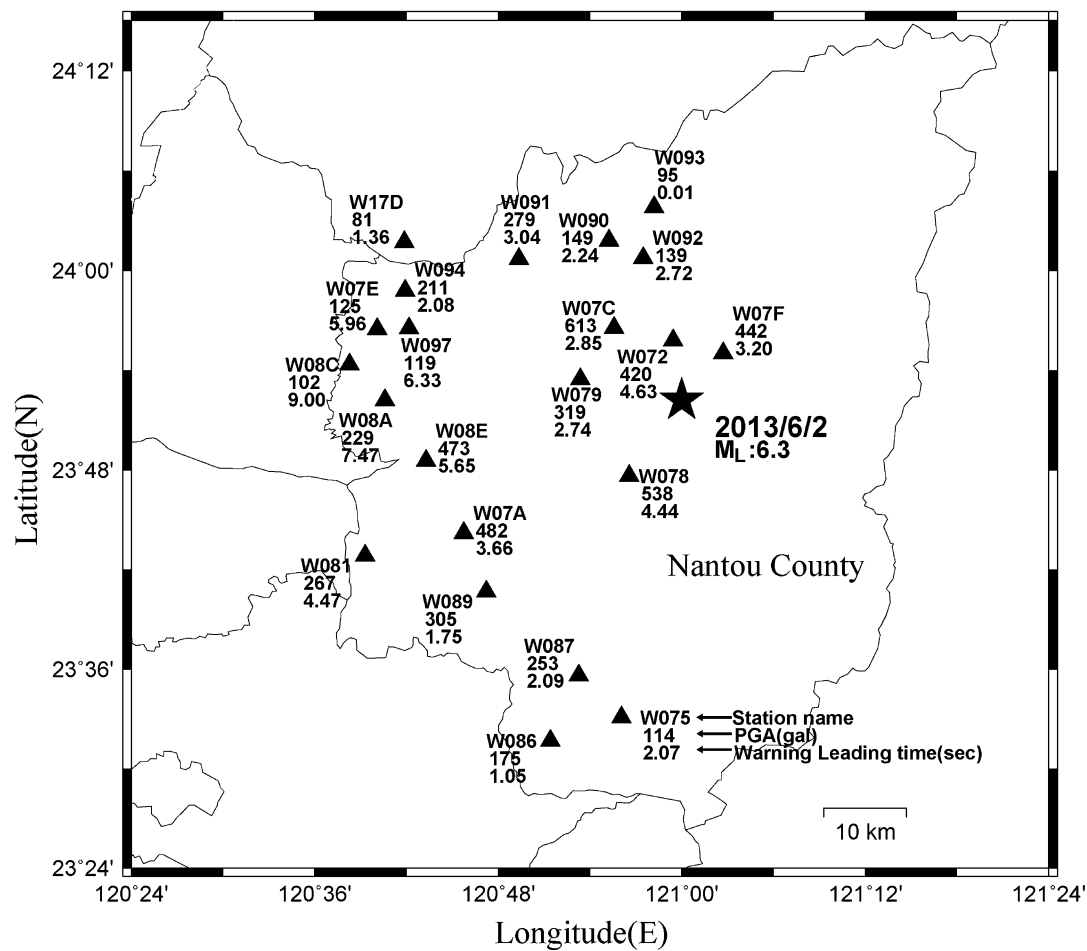
Earthquake Early Warning System in Taiwan, Fig. 7 A comparison of earthquake information, as follows: (a) epicenters, (b) focal depths, and (c) the

magnitudes given by the real-time *Palert* network and the CWB published earthquake catalogs. (d) The reporting times of the *Palert* network

the earthquake early warning information. The hypocentral location is determined using a traditional earthquake location algorithm with a half-space linear increasing velocity model (Wu and Lin 2013). The magnitude (M_p) determination is based on the relationship of peak displacement and velocity amplitudes of initial three P waves and hypocentral distance (Wu and Zhao 2006; Wu et al. 2013).

During the period from June 2012 to August 2013, a total of 33 earthquakes with M_L magnitudes ranging from 4.6 to 6.3 have been detected. Front-detection EEW performance of the *Palert* system is summarized in Figure 7. Results of the *Palert* network were compared with those reported by the CWB. Hypocenters provided by

the *Palert* network versus those from the CWB catalogue are plotted in Fig. 7. The hypocenters determined by the two systems are generally in agreement. Only a few offshore events have larger differences. The average difference in epicenter location is 14.1 km with a standard deviation of 15.1 km. The average difference in focal depth is 1.6 km with a standard deviation of 18.5 km. The magnitudes determined by the *Palert* network and the corresponding M_L values from the CWB earthquake catalog are listed. Figure 7 also provides a plot of M_p values against those of the CWB earthquake catalog. In general, magnitude uncertainty is in the order of 0.3, which is an acceptable value for EEW application. More than 1 year operation, the *Palert*



Earthquake Early Warning System in Taiwan, Fig. 8 The early warning leading times for *Palert* stations close to the epicentral region of the 2 June 2013 Nantou

M_L 6.3 earthquake. The leading time is defined as the interval between the time of on-site sound alert and the time of PGA

network achieves an early warning reporting time, the time between earthquake occurrence and the time of the system providing earthquake information, from 9 to 28 s, with an average of 15.5 s and a standard deviation of 4.3 s (Fig. 7). The results indicate that for the majority of events the *Palert* EEWs can issue an early warning report within approximately 20 s after an earthquake occurrence. It shows that it is feasible to use *Palert* devices to build a practical front-detection EEW system. The system could offer EEW messages for the regions at distance more than 60 km from the earthquake epicenter.

Front-detection EEW system could not offer early warning messages to the epicentral regions.

On-site approach may possibly offer early warning for regions close to the epicenter. *Palert* has an on-site alert function that triggers a warning sound once P_d exceeds 0.35 cm or PGA is larger than 80 gal (Wu et al. 2011). The major advantage of the on-site P_d approach is that it can provide a valuable early warning for epicentral regions. Figure 8 shows the resulting early warning leading times near the epicentral regions for the 2 June 2013 Nantou M_L 6.3 earthquake. So far, this earthquake is the largest event that has been detected by the *Palert* network. The early warning leading time is defined as the interval between the time when the filtered vertical displacement exceeds 0.35 cm or acceleration large than

80 gal and the time of the PGA. Only one station (W093) has a very short leading time. However, the majority of stations which have PGA larger than 80 gal have a 1–9 s leading time. It is worth noting that those stations which are very close to the epicenter (W07F, W072, W07C, W078, W08E, and W07A) and recorded of PGA larger than 400 gal had leading times from 3 to 6 s. Such times are long enough for automatic systems to make an emergency safety response.

Summary

In this paper, the progress on EEW in Taiwan is introduced in particular for the systems run by CWB and NTU. Both systems are able to reduce the EEW reporting within 20 s. This represents a significant step toward the capability of providing realistic EEW in Taiwan. The CWB is the agency responsible agency for earthquake reporting and NTU is a research and education institute. So, the front-detection EEW message is officially announced by the CWB. However, the technique developed by NTU will be transferred to the CWB for routine operation. Currently, front-detection EEW systems in Taiwan still cannot offer EEW messages for regions within 50 km from an earthquake epicenter. However, an on-site method can provide 3–6 s leading time before the largest shaking based on the results from the NTU EEW system. Thus, it is envisioned that a hybrid approach such as the NTU EEW system will be one of the future systems for real operation.

Cross-References

- [Early Earthquake Warning \(EEW\) System: Overview](#)
- [Seismic Accelerometers](#)

References

Allen R, Kanamori H (2003) The potential for earthquake early warning in South California. *Science* 300: 786–789

- Chen DY, Lin TL, Wu YM, Hsiao NC (2012) Testing a P-wave earthquake early warning system by simulating the 1999 Chi-Chi, Taiwan, Mw 7.6 earthquake. *Seismol Res Lett* 83:103–108. doi:10.1785/gssrl.83.1.103
- Grecksch G, Kumpel HJ (1997) Statistical analysis of strong-motion accelerogram and its application to earthquake early-warning systems. *Geophys J Int* 129:113–123
- Holland A (2003) Earthquake data recorded by the MEMS accelerometer: field testing in Idaho. *Seismol Res Lett* 74:20–26. doi:10.1785/gssrl.74.1.20
- Hsiao NC, Wu YM, Shin TC, Zhao L, Teng TL (2009) Development of earthquake early warning system in Taiwan. *Geophys Res Lett* 36:L00B02. doi:10.1029/2008GL036596
- Hsiao NC, Wu YM, Zhao L, Chen DY, Huang WT, Kuo KH, Shin TC, Leu PL (2011) A new prototype system for earthquake early warning in Taiwan. *Soil Dyn Earthq Eng* 31:201–208. doi:10.1016/j.solidyn.2010.01.008
- Kanamori H (2005) Real-time seismology and earthquake damage mitigation. *Annu Rev Earth Planet Sci* 33:195–214. doi:10.1146/annurev.earth.33.092203.122626
- Lee WHK, Shin TC, Teng TL (1996) Design and implementation of earthquake early warning system in Taiwan. Paper no. 2133. In: 11th world conference of earthquake engineering. Acapulco
- Nakamura Y (1988) On the urgent earthquake detection and alarm system (UrEDAS). In: Proceedings of the 9th world conference on earthquake engineering, Tokyo-Kyoto
- Teng TL, Wu YM, Shin TC, Tsai YB, Lee WHK (1997) One minute after: strong-motion map, effective epicenter, and effective magnitude. *Bull Seismol Soc Am* 87:1209–1219
- Wenzel F, Zschau J (2013) Early warning for geological disasters – scientific methods and current practice. Springer, Berlin/Heidelberg/New York. ISBN 978-3-642-12232-3
- Wu YM, Chen CC, Shin TC, Tsai YB, Lee WHK, Teng TL (1997) Taiwan rapid earthquake information release system. *Seismol Res Lett* 68:931–943
- Wu YM, Shin TC, Tsai YB (1998) Quick and reliable determination of magnitude for seismic early warning. *Bull Seismol Soc Am* 88:1254–1259
- Wu YM, Chung JK, Shin TC, Hsiao NC, Tsai YB, Lee WHK, Teng TL (1999) Development of an integrated seismic early warning system in Taiwan – case for the Hualien area earthquakes. *Terr Atmos Ocean Sci* 10:719–736
- Wu YM, Teng TL (2002) A virtual sub-network approach to earthquake early warning. *Bull Seismol Soc Am* 92:2008–2018
- Wu YM, Kanamori H (2005a) Experiment on an onsite early warning method for the Taiwan early warning system. *Bull Seismol Soc Am* 95:347–353
- Wu YM, Kanamori H (2005b) Rapid assessment of damaging potential of earthquakes in Taiwan from

- the beginning of P Waves. *Bull Seismol Soc Am* 95:1181–1185
- Wu YM, Zhao L (2006) Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning. *Geophys Res Lett* 33, L16312
- Wu YM, Kanamori H, Allen R, Hauksson E (2007) Determination of earthquake early warning parameters, Tc and Pd, for southern California. *Geophys J Int* 170:711–717
- Wu YM, Kanamori H (2008) Development of an earthquake early warning system using real-time strong motion signals. *Sensors* 8:1–9
- Wu YM, Chang CH, Zhao L, Teng TL, Nakamura M (2008) A comprehensive relocation of earthquakes in Taiwan from 1991 to 2005. *Bull Seismol Soc Am* 98:1471–1481. doi:10.1785/0120070166
- Wu YM, Lin TL, Chao WA, Huang HH, Hsiao NC, Chang CH (2011) Faster short-distance earthquake early warning using continued monitoring of filtered vertical displacement – a case study for the 2010 Jiasian earthquake, Taiwan. *Bull Seismol Soc Am* 101:701–709. doi:10.1785/0120100153
- Wu YM, Lin TL (2013) A test of earthquake early warning system using low cost accelerometer in Hualien, Taiwan. In: Wenzel F, Zschau J (eds) *Early warning for geological disasters – scientific methods and current practice*. Springer Berlin Heidelberg New York, pp 253–261. doi:10.1007/978-3-642-12233-0_13
- Wu YM, Chen DY, Lin TL, Hsieh CY, Chin TL, Chang WY, Li WS, Ker SH (2013) A high density seismic network for earthquake early warning in Taiwan based on low cost sensors. *Seismol Res Lett*
- Yu SB, Kuo LC, Punongbayan RS, Ramos EG (1999) GPS observation of crustal deformation in the Taiwan-Luzon region. *Geophys Res Lett* 26:923–926

Earthquake Location

I. Bondár¹, S. C. Myers² and E. R. Engdahl³

¹Research Centre for Astronomy and Earth Sciences of the Hungarian Academy of Sciences, Budapest, Hungary

²Lawrence Livermore National Laboratory, Livermore, CA, USA

³Department of Physics, University of Colorado, Boulder, CO, USA

Synonyms

3D velocity models; Earthquake; Hypocenter; Location methods

Introduction

Accurate estimation of an earthquake location is an essential starting point for quantitative seismological analyses, such as seismic hazard analyses and seismotectonics. Although the fundamentals of earthquake location were established nearly a century ago, improvements in robustness to data errors and nonoptimal network configuration have enabled semiautomation of the event location process. Further, algorithmic advancements and improved models of seismic wave speed in the Earth's interior continue to improve event location accuracy.

In the context of this entry, earthquake epicenter will be taken to mean the surface projection of the estimated latitude and longitude where fault rupture initiates. Earthquake location means epicenter and depth, and hypocenter means location and origin time. After a short introduction to the history of the methods for the determination of earthquake locations, a description of modern methods for single-event location is given. The final sections discuss event locations with 3D velocity models and multiple-event location methods.

A Short History of Catalogs/Bulletins

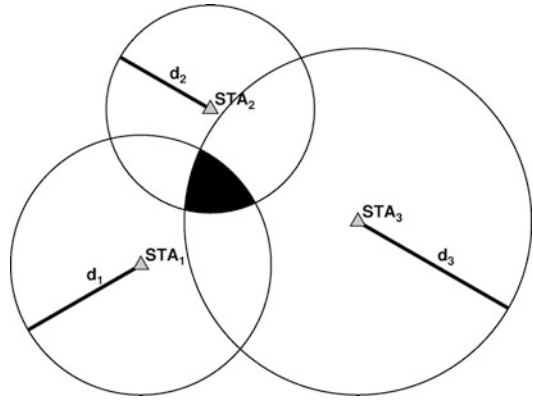
Excellent primary and secondary sources of catalog information about earthquake hypocenters and magnitudes can be found in the archives of the International Seismological Centre (ISC) and the US Geological Survey's National Earthquake Information Center (NEIC). These archives contain earthquake source parameters reported by a number of catalogs, bulletins, and studies of individual earthquakes. The archives are available in computer-readable format that are either directly determined by the ISC and NEIC or derived from published papers and institutional contributions. The most valuable catalog included in these databases for the historical period (before 1964) is the one derived from Gutenberg and Richter's book *Seismicity of the Earth* (1954), which provides hypocenters and magnitudes for most of the larger earthquakes occurring between 1904 and 1952. Other catalogs of large earthquakes are used mainly as sources of magnitude information.

Bulletins differ from catalogs in that they contain reported arrival times for seismic phases that are associated with each event. Seismic phases include first-arriving P-waves, S-waves, and additional seismic phases that are used to estimate a hypocenter. Prior to 1964 the primary sources of historical phase data are the bulletins of the International Seismological Summary (ISS) and the Bureau Central International de Seismologie (BCIS). Unfortunately, until only recently, these bulletins were mainly preserved in printed form and were not in a computer-ready, digital format. For the modern period (1964–present), instrumental phase data for moderate-to-large earthquakes worldwide are readily available in digital format from both the ISC and the NEIC.

In 1997, a project to relocate instrumentally recorded earthquakes during the period 1900–1963 was initiated. In this project the printed ISS bulletins were converted into digital form by scanning the ISS bulletin pages and applying an optical character recognition procedure to produce a digital arrival-time database of events. The initial processing of larger events in this database, combined with recent earthquakes in the ISC bulletin, resulted in a preliminary catalog for the twentieth century, the so-called Centennial Catalog (Engdahl and Villaseñor 2002). Recently, this catalog was substantially revised and expanded for the period 1900–2009 under the Global Earthquake Model (GEM) project (Storchak et al. 2013). The ISC-GEM catalog provides substantially improved earthquake hypocenters and magnitudes, along with their formal uncertainties.

Early Location Methods

John Milne, one of the founding fathers of instrumental seismology, devised some of the first quantitatively based methods for earthquake location, such as the methods of circles and hyperbolas (Milne 1886). The method of circles is similar in concept to the S minus P time location method. Assuming that a velocity model exists to specify the average P and S velocities, the difference between S and



Earthquake Location, Fig. 1 Illustration of the method of circles. If circles are drawn around the recording stations (triangles) with a radius defined by Eqs. 1 or 2, then the intersection of the circles determines the area containing the epicenter

P arrival time provides an estimate of distance from the station to the epicenter d :

$$d = \frac{v_P v_S (\tau_S - \tau_P)}{v_P - v_S} \quad (1)$$

where v is the velocity and τ is the arrival time for the phases indicated with the index. The expression further simplifies with the assumption that the ratio of P and S velocities is typically $\sqrt{3}$. Then Eq. 1 becomes

$$d = \frac{v_P (\tau_S - \tau_P)}{\sqrt{3} - 1} \quad (2)$$

Thus, for typical values of average P velocities, 8–10 times the S – P arrival-time difference gives a reasonable estimate of the epicentral distance from the station. Figure 1 shows that if there are several stations, the intersection of the circles drawn around the stations with the corresponding radius of the epicentral distance will define the most likely position of the epicenter. Note that because the velocity model is never perfect and the arrival times suffer from measurement errors, the circles do not intersect in a point. The area of the intersection is an indication of epicenter uncertainty. One of the major limitations of the method is that it gives little constraint on the source depth. Furthermore, measurement error

for the later-arriving S phase can be large, resulting in a correspondingly large error in the estimate of epicentral distance.

These early graphical methods evolved into a method of earthquake location by triangulation, drawing arcs from station locations with a compass on a large globe using a tape measure calibrated for the travel time. For many years the US Coast and Geodetic Survey (USC&GS) used this method to produce the Preliminary Determination of Epicenters (PDE) bulletin.

The Basic Location Problem

The location estimate is determined by finding the location and origin time (hypocenter) that minimizes the difference between observations and predictions of phase arrival times as measured at a network of seismographic stations. Methods to solve the location problem typically assume that both the phase associations (i.e., all associated phases belong to the event) and the phase identifications (or at least whether a phase is P or S type) are correct. The arrival-time prediction for each observation is the event origin time plus a calculation of travel time between the event and the observing station for the designated phase. Historically, travel time was read from an empirical travel-time table. Modern methods integrate time along ray paths for each phase within a model of wave speed for the Earth's interior. To ease the potentially heavy computational burden of computing ray paths and travel times in a heterogeneous Earth model, most location algorithms use radially symmetric velocity models, for which velocity is only a function of depth. For local velocity models, this translates to a horizontally layered structure.

Computationally efficient methods have been developed to compute travel times in radially symmetric Earth models (Buland and Chapman 1983). To further increase computational speed, precomputed travel times for various phases are often collected in travel-time tables for a set of depth and distance values. These travel-time tables are then used as lookup tables and interpolated to get travel-time predictions for a specific

phase, depth, and distance triplet. The predicted travel times may be adjusted by various corrections, to account for the ellipticity of the Earth and topographical effects, as well as path corrections to account for three-dimensional velocity heterogeneities. Thus, the time residual for a phase at the i th station is defined as

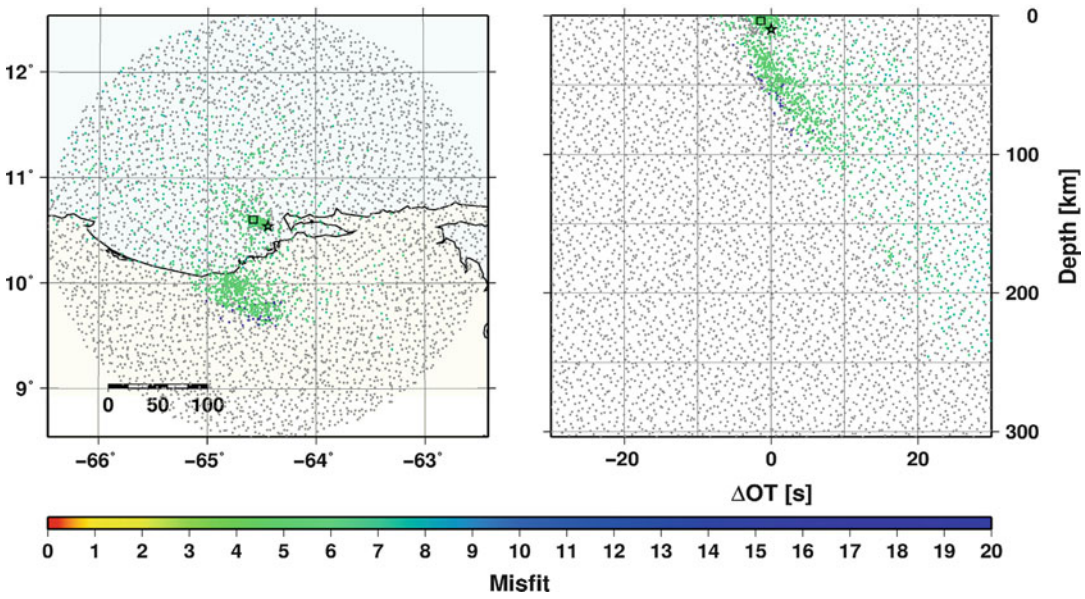
$$d_i = t_i^{obs} - t_i^{pred} = t_i^{obs} - (T^o + t_i^{model} + t_i^{corr}) \quad (3)$$

where d is the time residual; t^{obs} and t^{pred} are the observed and predicted arrival times, respectively; T^o is the origin time; t^{model} is the travel-time table value; and t^{corr} is the sum of applicable travel-time corrections. The predicted travel time for a phase arrival is a function of the station and source coordinates as well as the velocity model. For local earthquake location problems, where the epicentral distances do not exceed a few hundred kilometers, a horizontally layered, or “flat” Earth, model can be used without imparting significant computational error. Representing the Earth as a series of horizontal layers with constant or slowly varying velocities dramatically simplifies the calculation of travel-time predictions as they can be expressed in closed form in a Cartesian coordinate system. In either case, travel time is a nonlinear function of event location, requiring the use of nonlinear inversion methods.

Geiger's Method

One of the earliest numerical methods for earthquake location by the ISS consisted of least-squares estimations using a desk calculator. With the advent of computer processing, the method of Geiger (1910) was used to develop programs for routine earthquake location. Among the first of these programs were those developed by Bolt (1960) for the ISS, by Engdahl and Gunst (1966) for the USC&GS, and by Edouard Arnold for the ISC.

Geiger's insight was that if the initial source coordinates (i.e., the user's initial guess for the starting position for the hypocenter, $[x_0, y_0, z_0, T^o]_0^T$) are sufficiently close to the true



Earthquake Location, Fig. 2 Neighborhood algorithm search results for the $M_w = 6.7$, January 17, 1929, 11:45:44 earthquake in Venezuela. The misfit surface exhibits multiple minima (also note the trade-off between

origin time and depth). The *square* represents the best solution from the neighborhood search; the star shows the final location

hypocenter, the residuals can be expanded in a Taylor series with the higher-order terms considered to be negligible.

$$d_i = \frac{\partial t_i}{\partial x} \Delta x + \frac{\partial t_i}{\partial y} \Delta y + \frac{\partial t_i}{\partial z} \Delta z + \Delta T^o \quad (4)$$

where d is the travel-time residual, z is event depth, y and x are local event coordinates in the direction of latitude and longitude, T^o is event origin time, and t is phase travel time. This yields a linear system of N equations (N = number of arrival-time observations) with $M \leq 4$ model parameters written in matrix form as

$$\mathbf{G}\mathbf{m} = \mathbf{d} \quad (5)$$

where $\mathbf{G}$ is the $(N \times M)$ design matrix containing the partial derivatives for N data with respect to M model parameters, $\mathbf{m}$ is the $(M \times 1)$ model adjustment vector, $[\Delta x, \Delta y, \Delta z, \Delta T^o]^T$, and $\mathbf{d}$ is the $(N \times 1)$ vector of time residuals. Geiger proposed to solve the system of equations with an iterative least-squares method by minimizing the root mean square of residuals. After each

iteration k , the model parameters are adjusted by the current estimate of the adjustment vector, $\mathbf{m}_{k+1} = \mathbf{m}_k + \mathbf{m}_{\text{est}}$, where $\mathbf{m}_{\text{est}}$ is the least-squares solution of Eq. 5. Textbook linear algebra informs us that the solution to Eq. 5 is

$$\mathbf{m}_{\text{est}} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{d} \quad (6)$$

Practically all linearized location algorithms build on Geiger's method. Note that Geiger's linearization does not make the location problem linear; it assumes that the solution to the nonlinear problem can be found by iteratively solving a linear approximation to the problem, which enables application of standard numerical methods.

The assumption that the initial source coordinates are sufficiently close to the true hypocenter makes linearized location algorithms very sensitive to the initial guess for the starting hypocenter. Figure 2 shows the misfit surface (the root mean square of time residuals) obtained with the neighborhood algorithm (Sambridge and Kennett 2001), a nonlinear grid search algorithm (a more

detailed description is given later in the section Nonlinear Methods) for the $M_w = 6.7$, January 17, 1929, event in Venezuela. The station coverage for this event was quite poor as most of the recording stations were in Europe and North America. The misfit surface exhibits multiple minima, and a linearized location algorithm could easily be trapped in any of them, depending on the starting hypocenter. In this case the neighborhood algorithm provided a good starting location (square) for the linearized location algorithm that further refined the initial guess (star). The right panel clearly demonstrates the trade-off between depth and origin time. It should be noted that the problem of multiple local minima in the misfit surface is a particularly serious problem for local event locations in areas with strong 3D velocity heterogeneities.

Velocity Models/Travel-Time Tables

Since knowledge of Earth structure is derived primarily from earthquake data, the earliest Earth models were at best rudimentary, often inaccurate and incomplete. Travel-time predictions determined using early Earth models were valid only for shallow-depth earthquakes. For many years the standard travel-time tables used by the ISC and the NEIC were the Jeffreys and Bullen tables published in 1940. Although the limitations of these tables had been known for some time, until only the 1990s no other tables could provide such a complete representation of travel times for P, S, and other later-arriving phases.

In 1987 the International Association of Seismology and Physics of the Earth's Interior (IASPEI) initiated a major international effort to construct new global travel-time tables for earthquake location and phase identification. A model resulting from this effort was *iasp91* (Kennett and Engdahl 1991). Subsequently, Kennett et al. (1995) constructed an improved model for the average P and S radial velocity profile of the Earth (*ak135*). The *ak135* model has proven very suitable for predicting the arrival times of a wide variety of seismic phases for use in event location and phase identification procedures, and the *ak135* model is presently used by both the NEIC and ISC.

For the flat Earth approximation, widely used in local earthquake location algorithms, many of the above corrections are unnecessary. The travel-time predictions can also, in many cases, be expressed in closed forms. For instance, for a two-layer velocity model with a source at the surface, the travel times for the direct, reflected, and refracted waves are given by

$$\begin{aligned} t_{\text{direct}} &= \frac{x}{v_1}; & t_{\text{reflected}} &= \frac{\sqrt{x^2 + 4h^2}}{v_1}; \\ t_{\text{refracted}} &= t_0 + \frac{x}{v_2} \end{aligned} \quad (7)$$

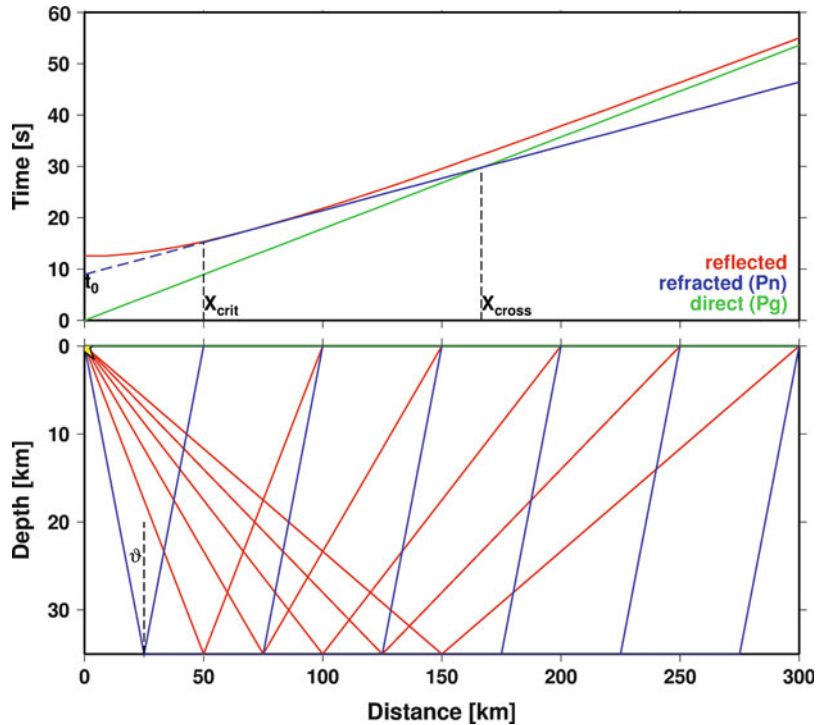
where x is distance, h is the depth of the layer boundary, v_1 and v_2 are the velocities above and below the layer boundary, $\vartheta = \arcsin(v_1/v_2)$ is the critical angle, and $t_0 = 2h \cos(\vartheta)/v_1$. Figure 3 shows the travel-time curves and various ray paths. Figure 3 also illustrates one of the major difficulties of the local earthquake location problem – despite the fact that the travel times are easily computed, one can generate a plethora of crustal phases, all arriving within a few seconds from each other. This can make phase identification difficult, especially around the crossover distances. It should be noted that ray tracing to obtain travel-time predictions could be still necessary if the velocity model becomes complicated, or the source is not in the uppermost layer of the model. Furthermore, the crust is the most heterogeneous part of Earth's interior, so 1D velocity models tend to do a poor job of adequately describing the local velocity structure. The result is increased nonlinearity, and it is not uncommon to have misfit surfaces with multiple local minima.

Depth Resolution

Until only recently the ISC and NEIC relied almost entirely on the times of first-arriving P phases to determine the locations of earthquakes that are observed globally, which for P rays bottoming in the lower mantle (teleseismic) the *ak135* travel times are particularly accurate. These agencies now use all later-arriving phases for earthquake location. In the absence of near-station phase data, direct teleseismic phases alone provide little depth

Earthquake Location,

Fig. 3 Ray paths and travel-time curves for a surface source in a two-layer velocity model. The refracted wave takes over the direct wave at X_{cross} , the crossover distance. The travel-time curves of the direct and refracted waves are *straight lines* with a slope of the slowness (inverse velocity) of the first and the second layer, respectively



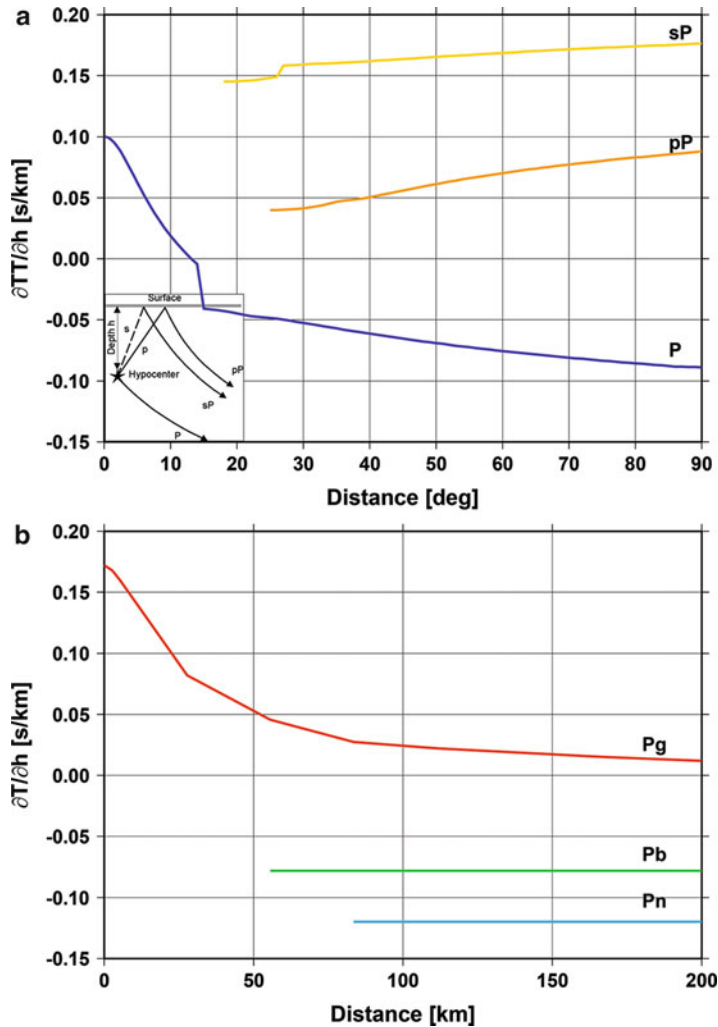
resolution, as the derivative of travel time with respect to depth at teleseismic distances is only a small, slowly varying function. P-waves and S-waves that initially travel upward and reflect off of the Earth's surface before traveling to the recording station provide important constraints on event depth because their travel-time derivatives with respect to depth are opposite in sign to those of direct teleseismic phases, as illustrated in Fig. 4a. The surface-reflected phases are commonly referred to as “depth” phases. The trade-off between event depth and origin time is also avoided by the inclusion of depth phases. The arrival times of depth phases can either be used directly as an independent observation or the depth can be estimated based on differential pP-P times. Alas, due to interference with the direct phase, as well as potentially strong reverberations in the Earth's crust, a distinct depth phase is only reliably observed for events below approximately 50 km.

For local problems depth phases are little help as there is not enough separation between the direct P arrival and the depth phase, which

makes picking the depth phases extremely difficult, and therefore they are hardly ever used. However, if there are stations where both the direct (Pg) and refracted (Pb or Pn, traveling along the Conrad or the Moho discontinuities in the crust, respectively) phases are picked, they could provide sufficient resolution to constrain the depth as their travel-time derivatives with respect to depth have opposite signs. Figure 4b shows that at close distances the travel-time derivative of the direct phase changes rapidly, thus providing depth resolution. As a rule of thumb, stations closer than twice the event depth are needed to reliably constrain event depth.

Error Budget

The error budget in seismic location problems is traditionally described as the combination of “measurement” and “model” errors. The effect of errors due to ignoring the higher-order terms in the Taylor expansion in Geiger's method, except for some highly degenerate network geometries, is typically second order compared to the model and measurement errors.



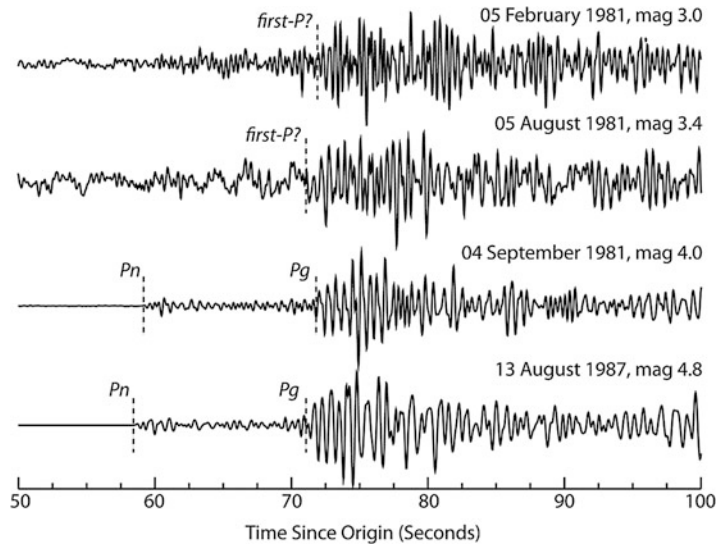
Earthquake Location, Fig. 4 (a) Travel-time derivatives with respect to depth as a function of epicentral distance for a source depth of 600 km for direct teleseismic P and the teleseismic depth phases pP and sP. The direct teleseismic P phase cannot resolve the depth in itself as the depth derivative is a slowly varying function, but when depth phase picks are available, together they can resolve the depth. The inset shows that the depth phases pP and sP first reflected from the Earth surface and then travel along a similar ray path as the

direct teleseismic P phase. (b) Travel-time derivatives with respect to depth at local distances for a source depth of 15 km for the direct Pg and the refracted Pb and Pn from the Conrad and Moho discontinuities. At close distances the depth derivatives of the travel times of the direct phase are rapidly changing which provides depth resolution. Further away the direct and refracted phases together provide depth resolution owing to the opposite signs of their derivatives

Measurement Errors

Measurement errors describe the errors in picking the onset times of seismic phase arrivals and are typically modeled as normally distributed, zero-mean random variables. However, the distribution of travel-time residuals is skewed and suffers from heavy tails, which indicates that the

assumption of normally distributed errors is violated. Measurement errors also suffer from systematic errors as onset times of seismic waves traveling along the same ray paths are systematically picked late with decreasing signal-to-noise ratio. Furthermore, large earthquakes tend to produce more complex and emergent



Earthquake Location, Fig. 5 Waveforms of nuclear explosions carried out at the Nevada test site recorded at Elko, Nevada, about 400 km distance. The explosions were detonated within 15 km of one another. The waveforms are band-pass filtered between 1 and 3 Hz and shown with increasing event magnitude. While for the

two larger explosions the first-arriving Pn can be easily picked, for the two smaller explosions the first-arriving Pn is completely masked by noise, and the more energetic, later-arriving Pg could be erroneously picked as the first arrival

waveforms, due to their longer source time and larger volume of energy release. In such situations often a more energetic later arrival (e.g., *Pg* or *pP*) is picked as the first-arriving phase, instead of the emergent first arrival (e.g., *Pn* or *P*) whose onset may be obscured by noise. Figure 5 shows the waveforms for several nuclear explosions that are within 15 km of each other from the Nevada test site recorded at Elko, Nevada, about 400 km distance from the source. Note the similarity of the waveforms. While for larger magnitude events the first-arriving Pn phase is clearly seen, for smaller events the first-arriving Pn is obscured by noise, increasing the chance of picking the more energetic, but later-arriving Pg as the first arrival. Such phase identification errors have plagued event locations since the beginnings of instrumental seismology. Some location methods remove phases with ambiguous phase labels. Alternatively, one phase label can be randomly chosen from a list of candidates in hopes that averaging errors will mitigate event location bias (Engdahl et al. 1998).

Model Errors

Model errors are caused by travel-time prediction errors due to unknown velocity heterogeneities in the Earth and may introduce location bias. After phase identification errors are accounted for, model errors represent the most significant contribution to the error budget, especially at local and regional distances. Representing model errors as zero mean, Gaussian processes commonly lead to erroneous formal uncertainty estimates, as the distribution of model errors is not zero mean and is often multimodal. Because of their systematic nature (a velocity model will always produce the same travel-time prediction error along the same ray path), model errors can only be reduced by improved travel-time predictions that account for the 3D velocity structure of the Earth.

Another major contribution to the error budget originates from the assumption that observational errors are independent. Unfortunately most location algorithms, either linearized or not, make this very assumption. Velocity heterogeneities inadequately modeled by the underlying velocity

model will generate correlated travel-time prediction errors along similar ray paths. Thus, the assumption of independent errors is violated when rays sample the same velocity anomaly. Ignoring the correlation structure of the travel-time predictions results in both location bias and increasingly underestimated location uncertainty estimates.

Linearized Inversion Methods

All linearized location algorithms originate from Geiger's method (1910). The differences are typically in the details, such as how the inverse of the $\mathbf{G}$ matrix is obtained, what kind of weighting scheme is applied, and how the formal uncertainties are calculated. Expressed in a probabilistic framework and assuming independent, normally distributed data, linearized location algorithms maximize the likelihood function

$$\mathcal{L}(\mathbf{m}) = \exp \left\{ -\frac{1}{2} (\mathbf{d} - \mathbf{Gm})^T \mathbf{C}_d^{-1} (\mathbf{d} - \mathbf{Gm}) \right\} \quad (8)$$

where $\mathbf{C}_d$ is the data covariance matrix describing the uncertainties in the data (picking and model errors). Maximizing $\mathcal{L}(\mathbf{m})$ is equivalent to solving the equation

$$\mathbf{G}_w \mathbf{m} = \mathbf{W} \mathbf{G} \mathbf{m} = \mathbf{W} \mathbf{d} = \mathbf{d}_w \quad (9)$$

where $\mathbf{W} = \mathbf{C}_d^{-1/2}$ is the diagonal ($N \times N$) weight matrix. Some early location algorithms even make the quite simplistic assumption that all data variances are the same and collapse the data covariance matrix into a single scalar value.

Popular location algorithms solve Eq. 9 in various ways, such as standard least squares or calculating the general inverse of the $\mathbf{G}$ matrix. More recent algorithms tend to obtain the general inverse by singular value decomposition. Seismic arrays offer reliable estimates of back-azimuth (station-to-event azimuth) and apparent velocity (or more precisely, its inverse, the horizontal slowness vector) as a plane wave sweeps through the array. Several algorithms make use of the

slowness and back-azimuth measurements from arrays (Bratt and Bache 1988; Lienert and Havskov 1995; Schweitzer 2001). While early locators allowed only for first-arriving P and S phases, recent algorithms use most of the phases allowed for by the *ak135* model (e.g., Bratt and Bache 1988; Engdahl et al. 1998; Bondár and Storchak 2011). Several popular algorithms focus on the local earthquake location problem and use travel-time predictions from a local velocity model (Lee and Lahr 1972; Lienert and Havskov 1995; Schweitzer 2001). Almost all linearized location algorithms rely on one-dimensional velocity models or travel-time tables to obtain phase arrival-time predictions. To account for large-scale three-dimensional effects, the EHB algorithm (Engdahl et al. 1998) applies patch corrections, Bratt and Bache (1988) allows for local and regional source-station-phase-specific corrections as well as slowness-azimuth corrections, and the ISC locator (Bondár and Storchak 2011) is capable taking travel-time predictions directly from a 3D velocity model (Myers et al. 2010).

Uniform Reduction

One of the most successful location algorithms is Jeffreys' uniform reduction method (Jeffreys 1932). It was in use with minor modifications (Bolt 1960; Buland 1986) up until the introduction of a new location algorithm at the ISC (Bondár and Storchak 2011). Uniform reduction also relies on Geiger's method but solves the linear system of equations in an iteratively reweighted least-squares fashion. This means that each arrival time will get a weight depending on its residual. The weighting function is defined as

$$w(d) = \left[1 + \mu \exp \left\{ \frac{(d - m)^2}{2\sigma^2} \right\} \right]^{-1} \quad (10)$$

where μ , m , and σ are parameters estimated from the data (Jeffreys 1932). The weights are recalculated after each iteration. Hence, outliers get progressively smaller weights and their contribution to the final hypocenter will be

gradually reduced. However, Buland (1986) pointed out that the introduction of a nonlinear weighting function further complicates an already nonlinear problem. Indeed, while the uniform reduction performs robustly when only first-arriving phases are used, the fact that it treats all phases as equal makes using less reliably picked later phases in the location quite difficult.

Formal Uncertainty Estimates

One of the major advantages of Geiger's method is that a closed-form solution exists for formal uncertainties, defined by the a posteriori model covariance matrix. Let the general inverse of $\mathbf{G}_w$ obtained by singular value decomposition be $\mathbf{G}_w^{-1} = \mathbf{V}_w \mathbf{\Lambda}_w^{-1} \mathbf{U}_w^T$, where $\mathbf{\Lambda}_w$ is the ($N \times N$) matrix of singular values (eigenvalues), $\mathbf{U}_w$ is an ($M \times N$) orthonormal matrix, and $\mathbf{V}_w$ is an ($N \times N$) orthonormal matrix whose columns are the corresponding eigenvectors of $\mathbf{\Lambda}_w$. The model covariance matrix, $\mathbf{C}_m$, is written as

$$\mathbf{C}_m = \mathbf{G}_w^{-1} \mathbf{C}_d \mathbf{G}_w^{-1^T} = \mathbf{V}_w \mathbf{\Lambda}_w^{-2} \mathbf{V}_w^T \quad (11)$$

The model covariance matrix defines a four-dimensional error ellipsoid, whose projections provide the two-dimensional epicenter error ellipse and the one-dimensional estimates of depth and origin time uncertainties. The formal uncertainties represent the two-dimensional confidence interval for the epicenter and the one-dimensional confidence intervals for depth and origin time. These uncertainties are typically scaled to a specified percentile level (90 % or 95 %). In other words, the error ellipse scaled to the 90 % confidence level encompasses the region that contains the true location with 90 % probability. The surface of an error ellipsoid scaled to the p th percentile confidence level is defined as the hypocenters ($\mathbf{m}$) satisfying

$$(\mathbf{m} - \mathbf{m}_h)^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_h) = \kappa_p^2 \quad (12)$$

where $\mathbf{m}_h$ is the point solution for the hypocenter and κ_p^2 defined by Jordan and Sverdrup (1981) as

$$\kappa_p^2 = M \hat{s}^2 F_p[M, K + N - M] \quad (13)$$

where F_p is an F statistic with M and $K + N - M$ degrees of freedom, M is the number of model parameters, N is the number of observations, and $\hat{s}^2$ is the a posteriori estimate of the variance scale factor given by $\hat{s}_e^2 = \frac{K \hat{s}_K^2 + \hat{s}_N^2}{K + N - M}$, where $\hat{s}_N^2$ is the a posteriori sample variance of residuals and $\hat{s}_K^2$ is the a priori estimate of $\hat{s}^2$ from previous experiments. Hence, the choice of K offers a balance between a variance scaling factor obtained from purely a posteriori residuals when $K = 0$ and from purely a priori estimates of data variances when K approaches infinity.

The formal uncertainties provide a statistical estimate of precision, but give no information about accuracy. Obviously, the precision of the measurement increases with the number of data, resulting in a smaller error ellipsoid. However, systematic measurement and model errors may introduce bias, which cannot be accounted for by the error ellipsoid. The accuracy of the measurement is the deviation from its true value and can only be measured when ground truth information is available.

Correlated Errors

All the equations above rely on the assumption that the errors are independent. When correlated travel-time predictions are present, the data covariance matrix $\mathbf{C}_d$ is no longer diagonal. The correlation structure implies that linear combinations of residuals may exist. In this case $\mathbf{W}$ is no longer defined as a diagonal weight matrix, but a projection matrix that orthogonalizes the data set and projects redundant observations into the null space (Bondár and McLaughlin 2009). Consequently, instead of the total number of observations, only the number of independent data contributes to the calculation of the formal uncertainties. This effectively decouples the size of the error ellipsoid from the number of correlated observations, thus preventing the error ellipsoid shrinking beyond control when increasing number of redundant observations are added. Heavily unbalanced, dense networks are prone to have enough similar ray paths to introduce location bias. Accounting for correlated travel-time

predictions not only produces more reliable formal uncertainty estimates, but also reduces location bias.

Nonlinear Methods

Poor network configuration and/or the use of a model with discontinuous travel-time derivatives can result in multiple local minima in the arrival-time misfit function, which is problematic for algorithms based on Geiger's method. Numerous strategies have been developed to mitigate the tendency of Geiger's method to get "stuck" in a local minimum. Most commonly, a number of widely varying starting locations are iterated to convergence, and the solution with the lowest misfit is selected. However, there is no guarantee that the global minimum is found and the formal uncertainty at the chosen hypocenter is not an accurate representation of the non-Gaussian misfit function.

A number of nonlinear inversion procedures have been developed with the goal of reliably finding the global minimum in the misfit function (i.e., the best location) and better representing location uncertainty. The vast majority of nonlinear methods use a strategy of explicitly sampling the misfit function on a discrete grid of points (grid search). If the sample points provide a high-fidelity representation of the misfit function, then it is straightforward to find the hypocenter, which is the global minimum in the misfit function.

Using grid search, it is only slightly more difficult to determine probability contours that define hypocenter uncertainty. Sambridge and Kennett (1986) provide a succinct description of how misfit may be mapped to probability. Assuming Gaussian data errors, the likelihood at hypocenter grid points, $\mathcal{L}(\mathbf{m})$, that satisfy the inequality

$$\frac{[\mathcal{L}(\mathbf{m}) - \mathcal{L}(\hat{\mathbf{m}})]}{\mathcal{L}(\hat{\mathbf{m}})} \leq \frac{\chi_4^2(p)}{(N - M)} \quad (14)$$

are within the p probability region, where p ranges between 0 and 1. The number of

hypocenter parameters ($M \leq 4$) determines the order of the chi-squared function (χ^2), and $\hat{\mathbf{m}}$ is the global minimum of the misfit function. The chi-squared function is divided by the degrees of freedom, which is the number of observations (N) minus the number of free parameters (M). The shape of the probability region determined in this way is nonparametric, which can provide an accurate assessment of location uncertainty. However, the nonparametric form can be difficult to summarize into a compact form that is convenient for distribution in a bulletin format.

Obviously, a grid search on hypocenter parameters is the most straightforward approach to map the misfit function. However, a global grid search can have computational limitations, even in the case of single-event location. For example, a grid with 1 km spacing on Earth's surface extending to a depth of 35 km (*ak135* crustal thickness) and sampling 0.1 s for 1 min would result in a grid of over 10^{13} points. Even if travel times were computed in a microsecond, a location based on 100 arrival times would require years to compute on a single processor. Clearly, strategies to reduce the grid dimensions are needed in order to make a grid search on all four hypocenter parameters practical.

Sambridge and Kennett (1986) limit the search domain by determining hypocenter bounds based on the station locations at which P-waves arrive first, S-wave minus P-wave arrival time, and plausible ranges in wave speed. In this case the domain of the grid search is determined by the data set, so computational demand can vary greatly. However, if the search domain were limited to an epicenter domain of 100 km by 100 km by 35 km by 60 s on a km spatial grid and a 0.1 s time grid, then computational time can be a few minutes on a single processor.

Rodi (2006) describes an adaptive grid search method that further reduces computational time by starting with a course grid and progressively densifying the grid in the neighborhood of points where the misfit function is lowest. Adaptive gridding has the benefit of enabling an arbitrarily fine grid in the neighborhood of the event, which

removes limitations on hypocenter accuracy that is inherent for a fixed grid.

Sambridge and Kennett (2001) use a natural neighbor approach to further refine the sampling strategy. The natural neighbor method starts with a random sample in hypocenter space. More random samples are then generated in the neighborhood of hypocenter samples with the lowest misfit values. This process is iterated until a global minimum is found. Similar to the grid search method, the misfit function can be contoured to assess location uncertainty. As will be discussed further, the misfit may also be used to guide numerical integration of multidimensional probability of the hypocenter, which provides a high-resolution estimate of the hypocenter volume.

It should be noted that features in the misfit function, including the global minimum, can be missed if the spacing of the sample points is too large. For example, if a grid with 10 km spacing is used, then a feature in the misfit function with spatial dimension of a few km might not be sampled. This is particularly problematic for adaptive grid approaches, for which computational gains are derived from a coarse starting grid. The safest practice is to initialize sampling with spacing smaller than the expected dimension of location uncertainty. That way the edges of the depression in the misfit function will be reliably sampled.

Estimating the probability density function of the hypocenter via stochastic sampling can more accurately characterize the probability volume and can, in some instances, improve computational efficiency. Lomax et al. (2000) introduce the Metropolis-Gibbs method to sample the a posteriori location probability density function. The Metropolis-Gibbs method begins with a starting hypocenter (state), which is commonly selected to be the location of the station at which the first P-wave arrived. A new hypocenter is proposed based on a random perturbation of the current state. If the probability of predicted data (weighted data fit) is higher at the proposed hypocenter, then the proposed hypocenter becomes the new "state." If the probability is lower at the proposed hypocenter, then it is accepted at a rate proportional to the ratio of the probabilities at the

proposed hypocenter and the current state. If the procedure of proposing a new hypocenter adheres to the critical balance condition, which requires reciprocity of probability in the proposal processes, then the series of accepted states comprise a Markov chain, for which each state is a random sample of the multidimensional hypocenter probability function. Therefore, high sample density equates to high probability. Many Markov chains with widely varying starting locations can be used to ensure robustness (Myers et al. 2007). The process of proposing, accepting, and rejecting new hypocenters typically finds the neighborhood of the global minimum in a few hundred iterations. The process is iterated thousands of more times to adequately characterize hypocenter probability, but the number of forward calculations is still orders of magnitude fewer than in a full grid search. Sample density may be contoured to characterize the nonparametric probability volume over hypocenter space. Alternatively, summary statistics (e.g., probability ellipses) may be computed by assuming that the samples are drawn from an analytical distribution (e.g., Gaussian).

Event Location with 3D Velocity Models

As discussed in the section on model errors, travel-time prediction errors are particularly problematic because they are repeatable, not random. This property of model error can result in a similar vector shift (bias) in hypocenter parameters and unrepresentative location uncertainty estimates (Bondár and McLaughlin 2009). Accounting for data covariance can mitigate the effects of model error, but a more accurate location and smaller uncertainty bounds can be achieved by improving travel-time prediction accuracy with improved models of seismic velocity. Practical application of 3D models depends greatly on the distance domain of the network: teleseismic, regional, or local.

Teleseismic

Teleseismic travel times can be accurately computed by integrating slowness along a ray that is

computed using a 1D model. In other words, the derivative of teleseismic travel time with respect to observed variations in lower mantle velocity is approximately linear. As a result, ray templates could be used to compute travel times on the fly. However, the most common approach is to precompute differences between travel times for a 3D and a 1D model at a fixed event depth. The travel time for the 3D model may then be approximated by using the 3D-1D model difference as a correction to the computationally efficient travel-time lookup for the 1D model. For example, a zero-depth correction surface may be applied to locate events throughout the crust. Travel-time corrections are attractive at teleseismic distance because deviations from 1D travel times vary slowly and interpolation of corrections is accurate even when points of precomputation are on the order of 100–200 km apart. Studies of improvement in epicenter accuracy generally find that average error can be reduced from approximately 15 km to approximately 8 km (e.g., Yang et al. 2004).

Regional

Due to strong velocity anomalies, minimum-time rays turning in the upper mantle can differ significantly for 3D and 1D models. Ray shooting techniques have been used to compute regional travel times. Shooting methods iterate on the direction and inclination of a ray leaving the event until the ray hits the station. However, a geometrical ray may not exist for some paths, resulting in “shadow zones” where travel time is not defined, but diffracted waves may be observed. Shadow zones cause significant difficulty for location algorithms because they are discontinuities in the domain of the travel-time calculation. Shadow zones can be avoided by using ray bending methods (e.g., Zhao and Lei 2004; Simmons et al. 2012), where ray endpoints are fixed and the ray is iteratively bent to find the least-time path. Bending methods have the advantage of providing an approximation to the travel time of diffracted waves, thus providing a travel time between any two points. Although regional ray paths and travel times can be

computed in about one-half second on current computers, real-time location may require precomputation. Precomputation approaches at regional distance are the same as those described for teleseismic phases, but the grid spacing must be smaller in order to capture strong gradients and first-order discontinuities in the travel-time corrections. Myers et al. (2010) achieve milli-second regional travel-time computation for Pn and Sn phases by parameterizing upper mantle velocity as a linear gradient, which enables flexible, on-the-fly implementation in a real-time location procedure. However, this approach is not applicable for rays interacting with velocity discontinuities at approximately 410 and 660 km depth.

Because crust and upper mantle velocity anomalies can be large compared to anomalies in the lower mantle, regional travel-time errors for 1D models can be many times larger than those at teleseismic distances. Epicenter error can be reduced from approximately 20 km on average for a 1D model to approximately 6–8 km for a 3D model (e.g., Ritzwoller et al. 2003; Yang et al. 2004; Flanagan et al. 2007; Myers et al. 2010).

Local

Velocity anomalies are largest in the crust, and important velocity discontinuities can be unique to local study areas. Unless local velocity models are highly smoothed, numerical solutions to the equation of a propagating wave front are the best approach for local travel-time calculations (e.g., Hole and Zelt 1995; Rawlinson and Sambridge 2004). Eikonal solutions are computationally expensive and require precomputation from each station of the network to a dense 3D grid of predefined points. Relying on travel-time reciprocity, a travel time from any hypothesized event location to each station of the network can be computed using 3D interpolation of the travel-time grid. The derivative of travel time with respect to event location is appreciably nonlinear at local distances, making nonlinear locators the best option (e.g., Lomax et al. 2000).

Multiple Event Location Methods

Simultaneous analysis of arrival times that are associated with many events provides information that can help to constrain unknowns beyond those of the event locations. Unknowns such as corrections to travel-time predictions, arrival-time measurement precision, and phase names may be inferred if the data set contains sufficient information. Because the unknown parameters include travel-time prediction corrections, uncertainties for each datum, and hypocenters, it is clear that the number of unknowns in the most general formulation of the multiple-event inversion will always outnumber the data. Therefore, the most general formulation will always be underdetermined. The choice of parameters to be included in the inversion and the approaches to constrain the inversion distinguish the various multiple-event applications.

Douglas (1967) introduced multiple-event location with the joint epicenter determination (JED) formulation. If the data set contains one phase type, the multiple-event formulation is

$$d_{ij} = \Delta T_i^o + \frac{\partial t_{ij}}{\partial z_i} \Delta z_i + \frac{\partial t_{ij}}{\partial y_i} \Delta y_i + \frac{\partial t_{ij}}{\partial x_i} \Delta x_i + \Delta s_j \quad (15)$$

where i specifies the event, j specifies the station, d is the travel-time residual, z is event depth, y and x are local event coordinates in the direction of latitude and longitude, T^o is event origin time, t is phase travel time, and S is a station-specific travel-time correction also called station correction. The linear system of equations is iterated to convergence. There is a need to impose constraints on the solution to Eq. 15 to remove ambiguity in the determination of the station terms. S. Douglas (1967) chose to force the sum of station corrections equal to zero, which remedies a trade-off between a shift in origin time of all events and the average station correction. Zero-mean station corrections make the implicit assumption that base-model travel-time prediction errors are zero mean. Still, absolute locations

can be poorly constrained if events are tightly clustered and event-station distances are large. Douglas (1967) suggested determining station corrections based on the residuals of one event whose hypocenter parameters are assumed known. All other events are then located relative to the fixed event. This technique becomes known as the master event location method and is frequently used to constrain absolute locations in an event cluster.

A number of methods improve the robustness and efficiency of solving the multiple-event system of equations. Dewey (1972) expanded on JED to include other phases and data weighting based on estimated pick uncertainty. Dewey's formulation is the widely used Joint Hypocenter Determination (JHD) method. Jordan and Sverdrup (1981) and Pavlis and Booker (1983) employed matrix projection operators to optimally compute unbiased estimates of station corrections. Jordan and Sverdrup (1981) solved the system of equations including event locations and station corrections in one iteration (hypocentroidal decomposition, HDC). Pavlis and Booker (1983) located each event independently, determined unbiased station corrections based on residuals, and repeated the process until convergence (progressive multiple-event location, PMEL). HDC is better suited for use with regional and teleseismic networks for which partial derivatives with respect to event location are relatively linear. PMEL has the advantage being more robust to highly nonlinear derivatives, which is generally true at local distances.

Recognizing the extreme precision of relative arrival times afforded by waveform correlation, Got et al. (1994) directly utilized differential arrival times to obtain highly precise relative event locations. The method minimizes the difference between observed and predicted differential arrivals. Thus, the approach is popularly referred to as double difference. Waldhauser and Ellsworth (2000) extended the double difference method to larger regions by introducing an inter-event length over which the implicit station/phase corrections decorrelate (HypoDD program).

Computational limitations prevent application of HypoDD to an arbitrarily large data set.

Myers et al. (2007, 2009) introduced a Bayesian nonlinear inversion framework to multiple-event analysis (Bayesloc). The nonlinear Markov chain Monte Carlo method enables Bayesloc to simultaneously assess the joint posterior distribution spanning event locations, travel-time corrections, phase names, and arrival-time measurement precision. Myers et al. (2011) demonstrated that Bayesloc can be applied to data sets of tens of thousands of events and millions of arrivals because the solution does not involve direct inversion of a matrix and computation demands grow linearly with the number of arrivals.

Multiple location methods when used properly will ordinarily produce precise relative earthquake locations. However, these locations will almost always have an unresolved bias, depending on the scale of lateral heterogeneity in the underlying Earth model. Lateral heterogeneity, especially in subduction zones, will also set limitations on the maximum inter-event distance when using multiple-event location methods. If the inter-event distances are too large, then travel-time predictions from events to stations will not be precisely correlated, potentially degrading the precision of relative locations.

Summary

The history of earthquake location methods is intertwined with the evolution of computers. Even though Geiger formulated the problem just 4 years after the great San Francisco earthquake, it was not implemented in routine operations until the first mainframe computers became available. Similarly, direct search methods became computationally viable only in the late twentieth century. Before the 1980s location algorithms typically were able to deal with a few dozen first-arriving P phases; nowadays, owing to the much higher station density, it is not uncommon that an event is reported with several

thousands of associated phase arrivals. Even just a decade ago, it would not have been possible to perform near-real-time waveform correlation and double difference location in a dense local network (Waldhauser 2009) or simultaneously relocate thousands of globally distributed events with a multiple location algorithm (Simmons et al. 2012). Hence, it can be expected that earthquake location methodologies will continue to evolve with ever-increasing computational power.

Cross-References

- [Earthquake Magnitude Estimation](#)
- [Seismic Event Detection](#)
- [Seismic Network and Data Quality](#)
- [Seismometer Arrays](#)

References

- Bolt BA (1960) The revision of earthquake epicentres, focal depths and origin times using a high-speed computer. *Geophys J Roy Astron Soc* 3:433–440
- Bondár I, McLaughlin K (2009) Seismic location bias and uncertainty in the presence of correlated and non-Gaussian travel-time errors. *Bull Seism Soc Am* 99:172–193
- Bondár I, Storchak D (2011) Improved location procedures at the International Seismological Centre. *Geophys J Int* 186:1220–1244. doi:10.1111/j.1365-246X.2011.05107.x
- Bratt SR, Bache TC (1988) Locating events with a sparse network of regional arrays. *Bull Seism Soc Am* 78:780–798
- Buland R (1986) Uniform reduction error analysis. *Bull Seism Soc Am* 76:217–230
- Buland R, Chapman C (1983) The computation of seismic travel times. *Bull Seism Soc Am* 73:1271–1302
- Dewey JW (1972) Seismicity and tectonics of western Venezuela. *Bull Seism Soc Am* 62:1711–1751
- Douglas A (1967) Joint epicentre determination. *Nature* 215:47–48
- Engdahl ER, Gunst RH (1966) Use of a high speed computer for the preliminary determination of earthquake hypocenters. *Bull Seism Soc Am* 56:325–336
- Engdahl ER, Villaseñor A (2002) Global seismicity: 1900–1999. In: Lee WHK, Kanamori H, Jennings PC, Kisslinger C (eds) *International handbook of earthquake and engineering seismology, part A, chapter 41*. Academic Press, Amsterdam, pp 665–690

- Engdahl ER, van der Hilst R, Buland R (1998) Global teleseismic earthquake relocation with improved travel times and procedures for depth determination. *Bull Seism Soc Am* 88:722–743
- Flanagan MP, Myers SC, Koper KD (2007) Regional travel-time uncertainty and seismic location improvement using a three-dimensional a priori velocity model. *Bull Seism Soc Am* 97:804–825
- Geiger L (1910) Herdbestimmung bei Erdbeben aus den Ankunftszeiten. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse* 4:331–349
- Got J-L, Fréchet J, Klein FW (1994) Deep fault plane geometry inferred from multiplet relative relocation beneath the south flank of Kilauea. *J Geophys Res* 99:15375–15386
- Gutenberg B, Richter CF (1954) *Seismicity of the earth and associated phenomena*. Princeton University Press, Princeton
- Hole JA, Zelt BC (1995) 3-D finite-difference reflection travel times. *Geophys J Int* 121:427–434
- Jeffreys H (1932) An alternative to the rejection of observation. *Mon Not R Astr Soc Geophys Suppl* 2:78–87
- Jeffreys H, Bullen KE (1940) *Seismological tables*. British Association of the Advancement of Science, Gray-Milne Trust, London
- Jordan TH, Sverdrup KA (1981) Teleseismic location techniques and their application to earthquake clusters in the South-central Pacific. *Bull Seism Soc Am* 71:1105–1130
- Kennett BLN, Engdahl ER (1991) Travel times for global earthquake location and phase identification. *Geophys J Int* 105:429–465
- Kennett BLN, Engdahl ER, Buland R (1995) Constraints on seismic velocities in the Earth from traveltimes. *Geophys J Int* 122:108–124
- Lee WHK, Lahr JC (1972) HYPO71: a computer program for determining hypocenter, magnitude and first motion pattern of local earthquakes. U.S. Geological Survey Open File Report, Menlo Park, pp 72–224
- Lienert BR, Havskov J (1995) A computer program for locating earthquakes both locally and globally. *Seism Res Let* 66:26–36
- Lomax A, Virieux J, Volant P, Berge C (2000) Probabilistic earthquake location in 3D and layered models: introduction of a Metropolis–Gibbs method and comparison with linear locations. In: Thurber CH, Rabinowitz N (eds) *Advances in seismic event location*. Kluwer, Hingham, pp 101–134
- Milne J (1886) *Earthquakes and other earth movements*. Appleton, New York
- Myers SC, Johannesson G, Hanley W (2007) A Bayesian hierarchical method for multiple-event seismic location. *Geophys J Int* 171:1049–1063. doi:10.1111/j.1365-246X.2007.03555.x
- Myers SC, Johannesson G, Hanley W (2009) Incorporation of probabilistic seismic phase labels into a Bayesian multiple-event seismic locator. *Geophys J Int* 171:193–204. doi:10.1111/j.1365-246X.2008.04070.x
- Myers SC, Begnaud ML, Ballard S, Pasyanos ME, Phillips WS, Ramirez AL, Antolik MS, Hutcheson KD, Dwyer JJ, Rowe CA, Wagner GS (2010) A crust and upper-mantle model of Eurasia and North Africa for Pn travel-time calculation. *Bull Seism Soc Am* 100:640–656
- Myers SC, Johannesson G, Simmons NA (2011) Global-scale P-wave tomography optimized for prediction of teleseismic and regional travel times for Middle East events: 1. Data set development. *J Geophys Res* 116, B04304. doi:10.1029/2010JB007967
- Pavlis GL, Booker JR (1983) Progressive multiple event location (PMEL). *Bull Seism Soc Am* 73:1753–1777
- Rawlinson N, Sambridge M (2004) Wave front evolution in strongly heterogeneous layered media. *Geophys J Int* 156:631–647
- Ritzwoller MH, Shapiro NM, Levshin EA, Bergman EA, Engdahl ER (2003) Ability of a global three-dimensional model to locate regional events. *J Geophys Res* 108(B7):2353. doi:10.1029/2002JB002167
- Rodi WL (2006) Grid-search event location with non-Gaussian error models. *Phys Earth Planet Int* 158:55–66
- Sambridge MS, Kennett BLN (1986) A novel method of hypocenter location. *Geophys J Roy Astr Soc* 87:679–697
- Sambridge MS, Kennett BLN (2001) Seismic event location: nonlinear inversion using a neighbourhood algorithm. *Pure Appl Geophys* 158:241–257
- Schweitzer J (2001) HYPOSAT – an enhanced routine to locate seismic events. *Pure Appl Geophys* 158:277–289
- Simmons NA, Myers SC, Johannesson G, Matzel E (2012). LLNL-G3Dv3: global P wave tomography model for improved regional and teleseismic travel time prediction. *J Geophys Res* 117(B10). doi:10.1029/2012JB009525
- Storchak DA, Di Giacomo D, Bondár I, Engdahl ER, Harris J, Lee WHK, Villaseñor A, Bormann P (2013) Public release of the ISC-GEM global instrumental earthquake catalogue (1900–2009). *Seism Res Let* 84:810–815
- Waldhauser F (2009) Near-real-time double-difference event location using long-term seismic archives, with application to Northern California. *Bull Seism Soc Am* 99:2736–2848
- Waldhauser F, Ellsworth WL (2000) A double-difference earthquake location algorithm: method and application to the Northern Hayward Fault, CA. *Bull Seism Soc Am* 90:1353–1368
- Yang X, Bondár I, Bhattachryya J, Ritzwoller M, Shapiro N, Antolik M, Ekström G, Israelsson H, McLaughlin K (2004) Validation of regional and teleseismic travel-time models by relocating GT events. *Bull Seism Soc Am* 94:897–919
- Zhao D, Lei J (2004) Seismic ray path variations in a 3D global velocity model. *Phys Earth Planet Int* 141: 153–166

Earthquake Magnitude Estimation

Peter Bormann

Formerly GFZ German Research Center for
Geosciences, Potsdam, Germany

Synonyms

Body waves; Calibration functions; Earthquake size; Earthquake strength; Magnitude scale; Richter scale; Seismic energy; Seismic moment; Standardization of measurements; Surface waves; Wave attenuation

Introduction

The magnitude of an earthquake (see ► [Earthquake Mechanisms and Tectonics](#)), or of other types of seismic events such as explosions, rock bursts, cave collapses, mining, or other by human activities induced events, is a number that characterizes the relative size of the event or of the amount of elastic seismic wave energy released by it. Therefore, the earthquake magnitude is, together with ► [Earthquake Location](#), the most important and most frequently determined seismic source parameter. Usually, magnitude values are based on the instrumental measurement of the maximum ground motion recorded by seismographs at Seismic Stations, ► [Seismic Networks](#), or ► [Seismometer Arrays](#), sometimes for a particular wave type and in a specific frequency range. However, these measured values have to be corrected for the decay of amplitudes with source distance and depth due to geometric spreading and attenuation during the propagation of seismic waves so as to become a representative value for the size of the seismic source or of the radiated seismic energy and thus of the related strength of seismic shaking. These correction or calibration values are usually given as formulas or in tabulated form. However,

sometimes magnitude values are also attributed to seismic events based on field observations of earthquake effects such as length and displacement of related surface ruptures or macroseismic effects such as earthquake damages in the near source area (see [Earthquake Intensity and Intensity Scales](#) by Musson and Cčić 2011; Grünthal 2011), using correlation relationships with instrumentally measured magnitudes. So derived macroseismic magnitudes M_{ms} are particularly important for the analysis and statistical treatment of historical earthquakes (see 3.2.6.6 in Bormann et al. 2013) but in this essay only instrumentally determined magnitudes will be considered.

The first paper on earthquake magnitude was published by Charles F. Richter (1935). Therefore, often the term *Richter scale* is used as a synonym. Yet, this may be incorrect, especially for very large earthquakes and magnitudes measured at so-called teleseismic distances of more than 1,000–2,000 km away from the source for which the original Richter scale is not applicable. Richter's aim was (quote):

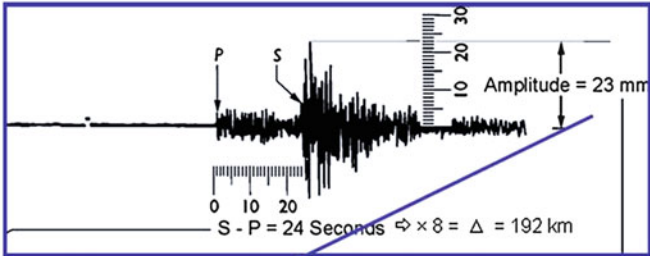
In the course of historical or statistical study of earthquakes in any given region it is frequently desirable to have a scale for rating these shocks in terms of their original energy, independently of the effects which may be produced at any particular point of observation.

The term *magnitude scale* was recommended to Richter by H.O. Wood in distinction to the term *intensity scale* which classifies the severity of earthquakes mainly on the basis of felt shaking or damage manifestations at different localities.

Richter compared earthquake records in Southern California made at different distances and azimuths from the earthquake source with a standard type of short-period horizontal component Wood-Anderson (WA) torsion *seismometer*. All these instruments had an identical frequency response (see red curve in Fig. 3 and ► [Seismic Instrument Response, Correction for](#)) and a static magnification, according to the manufacturer, of 2800. Many stations of the Southern California seismic network had already in the early 1930s been equipped with such seismometers. Figure 1 shows on top a typical WA-record of a near

According to this original definition of the local magnitude scale $M_L = M_L$ by Richter (1935), only the *maximum trace amplitude* B in mm as recorded in standard records of a Wood-Anderson seismometer is measured, i.e.: $M_L = \log B(WA) - \sigma(\Delta)$

⇒ no period T is measured, and **NO CONVERSION** to “ground motion” amplitude A by correcting for the frequency-dependent seismograph magnification is made.



$$M_L = \log B(WA) - \sigma(\Delta)$$

$$M_L = 1.36 + 3.46 = 4.82$$

Calibration function $\sigma_L(\Delta) = -\log A_0$ for local magnitudes M_L according to Richter (1958). A_0 are the amplitudes in mm recorded by a Wood-Anderson Standard Torsion Seismometer from an earthquake of $M_L = 0$. Δ - epicentral distance in km.

Δ (km)	$\sigma_L(\Delta)$	Δ (km)	$\sigma_L(\Delta)$	Δ (km)	$\sigma_L(\Delta)$	Δ (km)	$\sigma_L(\Delta)$
0	1.4	90	3.0	260	3.8	440	4.6
10	1.5	100	3.0	280	3.9	460	4.6
20	1.7	120	3.1	300	4.0	480	4.7
30	2.1	140	3.2	320	4.1	500	4.7
40	2.4	160	3.3	340	4.2	520	4.8
50	2.6	180	3.4	360	4.3	540	4.8
60	2.8	200	3.5	380	4.4	560	4.9
70	2.8	220	3.65	400	4.5	580	4.9
80	2.9	240	3.7	420	4.5	600	4.9

Earthquake Magnitude Estimation, Fig. 1 Summary of the classical procedure of measuring M_L . Note that the letter B was given here to record trace amplitudes,

highlighting that they were not converted to ground motion displacement amplitudes A (Modified copy of Fig. 3.28 in Bormann et al. 2013; © IASPEI)

earthquake. By plotting, for many events, the logarithm of the recorded maximum trace amplitudes A_{max} over the *epicentral distance* Δ (see ► [Earthquake Location](#)), Richter derived empirically average distance-dependent calibration values $\sigma(\Delta)$. They were needed in order to compensate for distances up to 600 km for the decay of the reference amplitude A_0 with distance from the seismic source. It holds $\sigma(\Delta) = -\log A_0$. Tabulated values for $\sigma(\Delta)$ are given in Fig. 1. Adding them to the logarithm of the maximum trace amplitude $B(WA)$ of a seismic event recorded by a WA seismograph (see Fig. 1) makes the event amplitudes measured at different distances comparable and yields the event magnitude. Moreover, the corrected amplitudes for different events are then indicative for the differences in event size. Accordingly, the local magnitude M_L has been defined as

The magnitude of a shock is ... the logarithm of the calculated maximum trace amplitude, expressed in microns, with which the standard short-period torsion seismometer ... would register that shock at an epicentral distance of 100 km. (Richter 1935)

This implies that an amplitude of 1 mm on the WA seismogram corresponds to a magnitude 3 earthquake and this provides the anchoring of the scale.

Figure 1 illustrates the determination of the local magnitude M_L according to this procedure. Since then several other magnitude scales have been developed, pioneered with three publications by Beno Gutenberg (1945a, b, c). He was interested in the use of the magnitude concept for classifying the size of earthquakes worldwide, based on *body-wave* and *surface-wave* measurements also at teleseismic distances. These magnitudes were later termed m_B and M_S

by Gutenberg and Richter (1956) who also published in this paper calibration values for three types of body waves: (1) in propagation direction oscillating pressure waves P, (2) PP waves which are P waves that have been reflected back on the Earth surface at half way distance between the source and the seismic station and (3) S waves, which are slower travelling shear waves of larger amplitudes (see Fig. 1) that oscillate perpendicular to the direction of wave propagation. Yet, already Gutenberg and Richter realized that their original dream of a unique size scaling of earthquakes was not achievable with magnitude scales based on different types of seismic waves, measured at different source distances and frequencies. Although they mutually scaled M_L , m_B , and M_S to agree at values around 6.5 these magnitudes differ elsewhere, up to about one magnitude unit (m.u.) (Fig. 15). This problem became even more obvious when seismographs operating at much wider, or narrower, ranges of frequency with much higher amplification of ground motion than available at Gutenberg's and Richter's time, made it possible to measure magnitudes up to 9.5 and down to less than zero. Magnitudes measured at periods much less than the duration of earthquake rupture or of another seismic waves generating process were then found to systematically underestimate the overall earthquake size and energy release and thus also its magnitude. This effect is often referred to as "magnitude saturation." The controlling parameters are outlined in section [Earthquake Rupture, Radiated Seismic Spectrum, Seismic Moment, Energy and Related Magnitudes](#).

It often occurs that different agencies publish magnitudes with the same nomenclature although they had been measured with different filter procedures within different measurement time windows and corrected with different calibration functions. Thus it could happen that the first available values of the most widely measured short-period teleseismic m_b magnitude (see section [Body-Wave Magnitudes](#)) for the great devastating Sumatra-Andaman tsunami generating earthquake of December 26, 2004, ranged between 5.7 and 7.5 (Bormann et al. 2007, 2009).

In view of magnitudes being logarithmic numbers this corresponds to a difference in estimated short-period seismic energy release of this event by a factor of almost 2,000 times. Moreover, even the largest of these m_b values, measured around 2 s period for a rupture with an estimated duration of 500–600 s, saturated already at 7.5. This is almost two magnitude units less than the much later available largest *moment magnitude* value (see section [Earthquake Rupture, Radiated Seismic Spectrum, Seismic Moment, Energy and Related Magnitudes](#)) of $M_w = 9.3$ and explains the initial underestimation of the size and related hazard of this earthquake. These differences become less for smaller earthquakes but remain nevertheless often still unacceptably large (Granville et al. 2005; Bormann et al. 2007). They are a possibly source of bias or large data scatter when such data, if they enter catalogs, are used in estimates of magnitude-dependent [Earthquake Recurrence](#) frequency, or of Deterministic Seismic Hazard Analysis or [Probabilistic Seismic Hazard Analysis](#) in terms of [Seismic Hazard: Ground Shaking](#).

The confusion that is sometimes associated with apparently well-established common practise of magnitude determination has highlighted the urgent need for:

- (a) Internationally agreed standards of measurement procedures for the most widely used types of magnitudes and
- (b) Wider published and easily accessible critical reviews on the physical basis of the diverse approaches to magnitude measurements, their potentials, biases, and mutual relationship.

The former issue has been taken up by an international Working Group on Magnitude Measurement of the Commission on Seismological Observation and Interpretation (CoSOI) of the International Association of Seismology and Physics of the Earth's Interior (IASPEI). It resulted in the adoption of IASPEI (2013) resolution No. 1 (Fig. 2).

The second issue has been elaborated in detail in Chap. 3 of the second, now electronic edition



IASPEI Resolutions

2013 Scientific Assembly, Gothenburg, Sweden

Resolution 1: Magnitude Standards

RECOGNISING the importance of the magnitude standards proposed by the Working group on Magnitude measurements of the Commission on Seismological Observation and Interpretation and published in the New Manual of Seismological Observatory Practice (second edition),

IASPEI

RECOMMENDS the station and network operators and data centres to adopt these standards in day to day operations, and

ENCOURAGES the developers of waveform processing programmes to incorporate these standards within their software packages.

Earthquake Magnitude Estimation, Fig. 2 IASPEI resolution 2013 on magnitude measurement standards (© IASPEI)

of the IASPEI New Manual of Seismological Observatory Practice (NMSOP-2; <http://nmsop.gfz-potsdam.de>; Bormann et al. 2013) and related complementary NMSOP publications. Therefore, the focus will be in the following on the principles, procedures, and advantages of the recommended new measurement standards for the most important types of classical magnitudes. This will be complemented by an introduction into the more physically defined and nearly non-saturating seismic moment and energy broadband magnitudes, M_w , and M_e , as well as into most recently developed faster proxy estimates for these magnitudes that are applicable in (near) real-time earthquake and tsunami early warning schemes. But beforehand the essential features of earthquake rupture and the radiated seismic wave spectrum as well as their controlling parameters are explained in order to understand the often striking differences between different magnitude scales. For any further details, also on other types of magnitude, readers should consult NMSOP-2.

Earthquake Rupture, Radiated Seismic Spectrum, Seismic Moment, Energy, and Related Magnitudes

An earthquake is a shaking of the Earth that is either *tectonic* or volcanic in origin or caused by collapse of cavities in the Earth. A tectonic earthquake is caused by *fault slip*. Faults are fractures or fracture zones in the Earth along which the two sides have been displaced relative to one another parallel to the fracture. The friction between the rock masses adjacent to the fault usually prevents slip or creep along the fault even under tectonic stress load. Thus stress can accumulate until the breaking strength of the fault material is reached and fault rupture occurs. During the fault slip part of the accumulated tectonic stress is released. The difference between the stress in the source volume before and after the fault slip is called *stress drop* $\Delta\sigma$. How much stress can be accumulated and released also depends on the density of the rock material along the fault and its resistance against shearing, which is usually expressed by

its *shear modulus* μ , also termed *rigidity*. The frictional conditions, μ and $\Delta\sigma$ also control the velocity or speed of the fault rupture and thus the *rupture duration*, which varies between the smallest and the largest earthquakes between about 1/100th of a millisecond to several minutes for rupture lengths between centimeters and more than 1,000 km with related magnitudes between about -2 to 9.5 . An approximate relationship between non-saturated magnitude M and average rupture duration τ_{rup} has been given by Bormann et al. (2009):

$$\log \tau_{\text{rup}} \approx 0.6M - 2.6. \quad (1)$$

Thus, earthquake ruptures of magnitude $M = 4, 5, 6, 7, 8$, and 9 would roughly last on average 0.4 s, 1.6 s, 6 s, 25 s, 100 s and 400 s, i.e., τ_{rup} increases by about by a factor of four per magnitude unit. However, depending on the rigidity of the material in the source volume and the related stress conditions the rupture velocity v_{rup} and thus τ_{rup} of individual earthquakes may deviate from this average by a factor of about 2 – 3 shorter or longer (see Fig. 5).

As shown in section [Magnitude Scales and the New IASPEI Measurement Standards](#) the size estimates of different magnitude scales may differ for the same earthquake. These differences (or none) usually depend on the magnitude itself. They can only be understood when considering – besides possible differences in the time windows within which magnitude is measured on the seismic record – how the measurement parameters amplitude and period, used by these scales, differ in relation to the radiated seismic source spectrum.

Source spectra observed in the far field, i.e., several wavelengths away from the rupture, depend on the rupture size, $\Delta\sigma$, the related v_{rup} and the ratio between released seismic wave energy E_s and seismic moment M_0 . The scalar seismic moment M_0 of a shear rupture is proportional to the constant displacement plateau amplitude u_0 of the radiated source spectrum and defined as

$$M_0 = \mu \bar{D} A \quad (2)$$

where $\bar{D}$ = average slip over the fault and A = rupture area. All these values, with the exception of the rigidity μ , can be estimated via model assumptions from measured spectral displacement plateau amplitudes and the corner frequency f_c of the source spectrum (Baumbach and Bormann 2011). Therefore, it has been suitable to scale the displacement spectra in Figs. 3 and 4 to seismic moment and the related velocity spectra to the moment rate, i.e., the time derivative of M_0 .

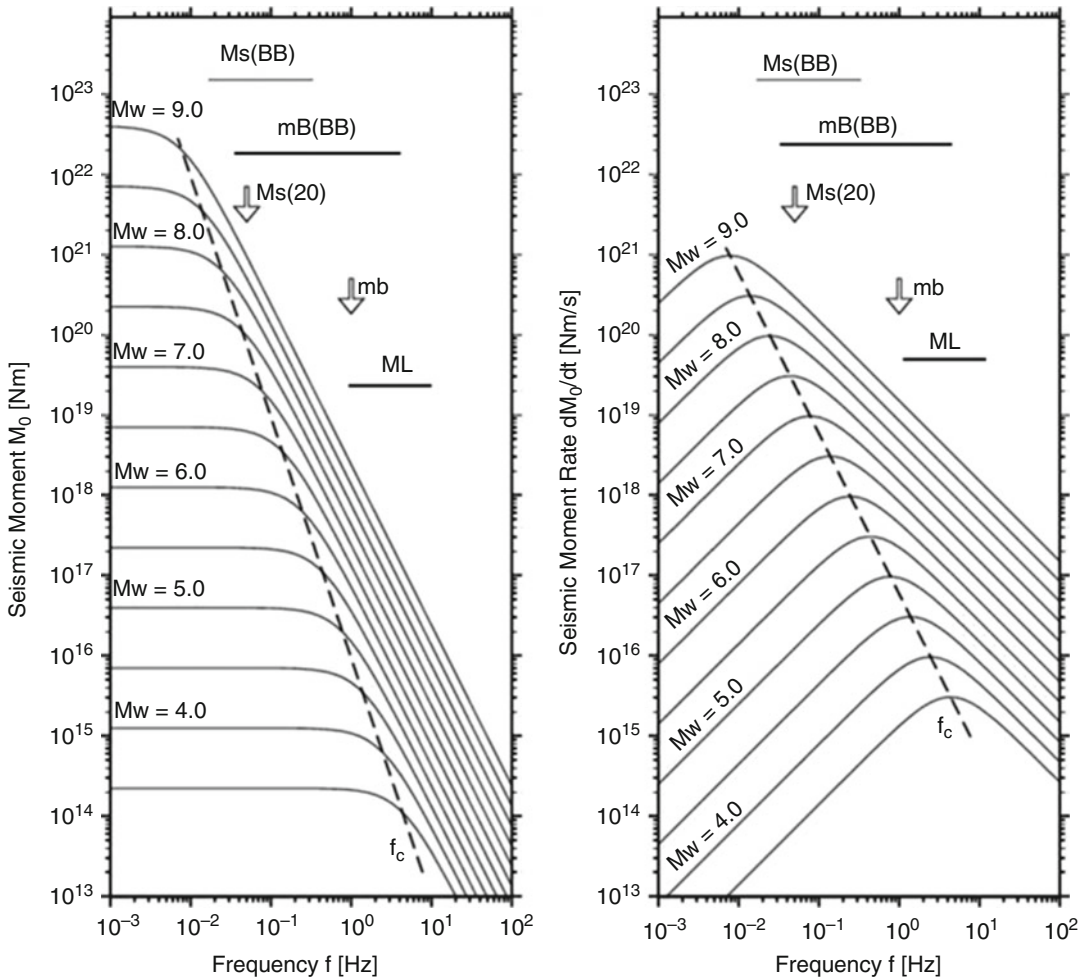
Fig. 3 shows this for an average shear rupture model of equal stress drop for events of different size and Fig. 4 for earthquakes of equal size in terms of seismic moment but different in stress drop. For model details see Bormann et al. (2009), but the fundamentals can be understood with reference to these figures without knowing these details.

M_0 increases linearly with the product $\bar{D} A$, also termed *earthquake potency*, which is a static measure of the overall earthquake “size,” But the potency does not account for differences in rupture velocity or stress drop and thus related differences in the radiated frequency content and amount of seismic energy for a given M_0 . None the less, this lead Kanamori (1977) to propose, via relationship (2) and by assuming constant average values of μ and $\Delta\sigma$ and thus of the ratio E_s/M_0 , a non-saturating “ w = work” or moment magnitude scale, termed M_w . It is most consistently determined by the recommended IASPEI (2013) standard way of writing the original Hanks and Kanamori (1979) formula, and reads

$$M_w = (\log M_0 - 9.1)/1.5. \quad (3)$$

M_0 and thus M_w will not saturate with growing rupture size provided that they are measured via the plateau amplitude u_0 of the radiated displacement source spectrum. This requires, however, to measure u_0 at frequencies much smaller than the corner frequency f_c , i.e., at wavelengths larger than the length of the source rupture.

M_0 calculations are nowadays routine at major seismological centers, either by measuring the displacement amplitude u_0 on the long-period asymptote (plateau) of the source spectrum,

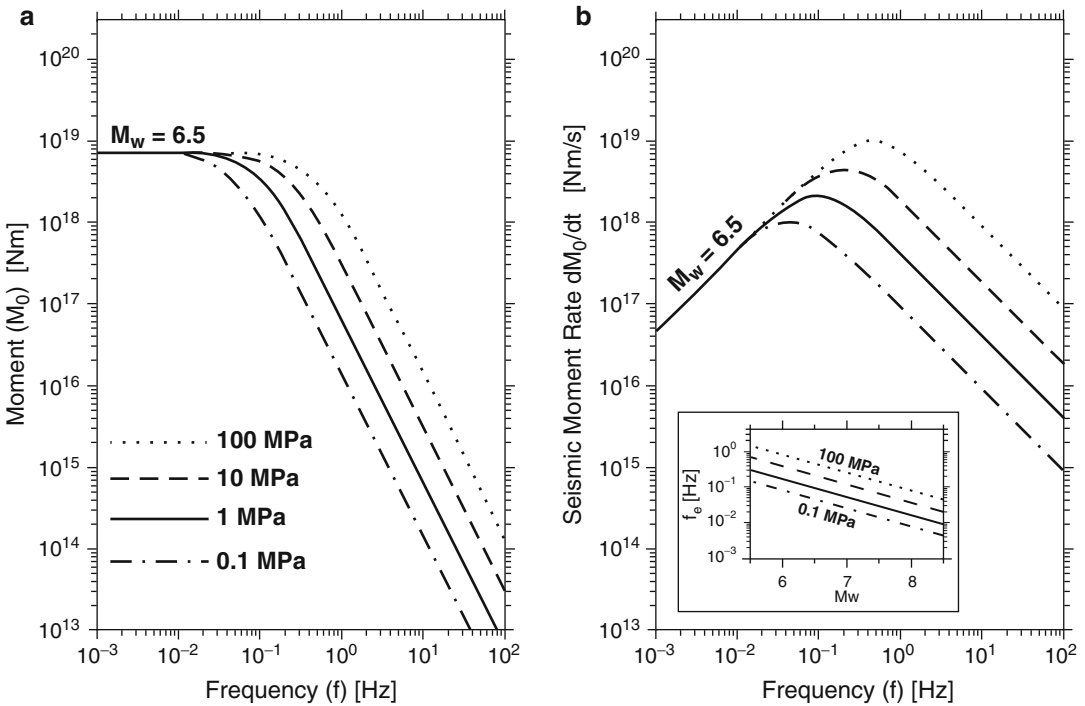


Earthquake Magnitude Estimation, Fig. 3 Theoretical source spectra of radiated displacement (*left*) and velocity amplitudes (*right*) scaled to seismic moment M_0 , respectively seismic moment rate dM_0/dt in relation to the period ranges within which the IASPEI (2013) standard magnitudes are measured (see section [Magnitude Scales and the](#)

[New IASPEI Measurement Standards](#)). The frequencies at which the plateau amplitudes, respectively M_0 have dropped to about 70 % of their plateau values are termed corner frequencies f_c (Modified Fig. 1 from Bormann et al. 2009, BSSA © Seismological Society of America)

which is proportional to M_0 after correcting for the source radiation and wave propagation effects (as done in Fig. 3), or, as a spin-off of seismic moment tensor calculations (see ► [Moment Tensors: Decomposition and Visualization](#)), by fitting long-period synthetic seismograms for different source models best to long-period filtered real seismic records, using wavelengths that are larger than the rupture length. But differences in procedures are the reason that not all yield fully compatible and non-saturating M_0 and M_w

values. For example, for the great 2004 Sumatra-Andaman M_w 9.3 earthquake the reported values for M_w differed up to about one magnitude unit (m.u.). The most important procedure, elaborated by Dziewonski et al. (1981), forms, with some later modifications, the computational basis for the authoritative Global Centroid Moment Tensor Catalog. Values for all major earthquakes with $M_w > 5.5$ are regularly published by the GCMT project (<http://www.globalcmt.org/CMTsearch.html>) and available



Earthquake Magnitude Estimation, Fig. 4 Source spectra for the same model as in Fig. 10 but for earthquakes of the same size yet different in stress drop

(in MPa) (Copy of Fig. 2 from Bormann and Di Giacomo 2011 © Springer; with kind permission from Springer Science + Business Media)

within about 2 h after the event. Less frequently and incomplete the GCMT catalog now also reports M_w values down to about 4.5. Regional centers, such as Swiss Earthquake Service, provide – with larger time delay – reliable moment tensor and M_w solutions also down to significantly smaller values, e.g., when dense quality networks close to the source are available, down to $M_w \approx 3.0$ (Braunmiller et al. 2002).

A new method for rapid non-saturating moment tensor and M_w estimates is based on very long-period W-phase observations between the P- and S-wave arrivals (Kanamori and Rivera 2008 and Sect. 3.2.8.6 in Bormann et al. 2013). First real-time implementations are operational at the U.S. Geological Survey’s NEIC (Hayes et al. 2009) as well as at the GFZ Potsdam and the Indonesian Tsunami Early Warning System (InaTEWS) since 2011.

The question arises why the radiated displacement amplitude spectra in Figs. 3 and 4 are flat at

frequency much lower and drop strongly with frequencies higher than f_c . As in optics, lower frequencies with wavelengths larger than the object, in our case the causing rupture, cannot resolve any smaller detail of the rupture. They “see” the earthquake only as an undifferentiated “point source,” integrating over the whole rupture length. Accordingly, there is no difference in the respective displacement amplitudes at frequencies $f \ll f_c$. In contrast, higher frequencies with wavelengths smaller than the rupture length are radiated by smaller rupture patches. Their generated displacement amplitudes are lower the smaller the size of these patches. This explains the rapid decay of amplitudes towards higher frequencies. The length of rupture can be estimated from f_c via some model assumptions and corrections made about the seismic source and the wave propagation effects.

The most common source spectra model is the ω^{-2} model, in which the displacement

amplitudes decay at frequencies $f = \omega/2\pi \gg fc$ with the second order, as in Figs. 3 and 4. However, there are other models, where the amplitudes drop steeper, with third or even higher order. This is also observed, besides the ω^{-2} decay, in empirical data. If the linear dimensions of the rupture increase with increasing size of the earthquake, fc moves towards lower and lower frequencies. Average values of fc for magnitude 4 earthquakes are on average around 2.5 Hz and for a magnitude 8 earthquakes around 0.02 Hz. However, as seen from Fig. 4, for a given scalar seismic moment and thus static measure of earthquake “size”, fc depends also on $\Delta\sigma$ in the source volume. Depending on μ in the source volume and the ambient stress conditions, $\Delta\sigma$ may vary by 3 to 4 orders of magnitude. Since $f_c \sim \Delta\sigma^{1/3}$ (Brune 1970) this may result in differences of fc for equal M_0 by a factor of 10 or even more.

These possible modifications notwithstanding, Figs. 3 and 4 explain why any magnitude scale which measures amplitudes at frequencies much higher than fc will underestimate the earthquake size in terms of seismic moment. The same applies with respect to the underestimation of earthquake “strength” in terms of released seismic energy, with the latter being proportional to the squared seismic moment rate, when the largest ground motion velocity $V_{\max}$, or $(A/T)_{\max}$, is measured at frequencies much higher or lower than f_c (see the right hand panels of Figs. 3 and 4). On average, spectral underestimation with respect to M_w begins to become significant for short-period (1 Hz) m_b and M_L already at values >5.5 – 6 , in contrast to the two surface-wave magnitudes measured at periods around 20 s, termed $M_S(20)$, respectively <60 s, termed $M_S(\text{BB})$, and the P-wave broadband magnitude $mB(\text{BB})$ that is measured at $T < 30$ s, for which spectral saturation begins only at $M_w > 7.5$ – 8 . However, $M_S(20)$, measuring $\log(A/T)$ around 20 s, and not $A_{\max}$ or $V_{\max}$, is expected to be smaller than M_w also for events with $M_w < 6$ – 6.5 . This is in agreement with empirical data in Figs. 7, 15, and 17 as well as Eq. 17.

Similarly to formula (3), the recommended standard formula for calculating the energy

magnitude M_e proposed by Choy and Boatwright (1995) is now written as

$$M_e = (\log Es - 4.4)/1.5. \quad (4)$$

with Es in Joules. Modern velocity-broadband records allow nowadays to estimate radiated seismic wave energy Es by integrating squared velocity amplitudes over a wide range of periods around the peak values of the velocity spectrum (Fig. 3, right hand panel), and correcting them for wave propagation effects. For more details of these not yet standardized procedures see Choy and Boatwright (2012) and Di Giacomo and Bormann (2011).

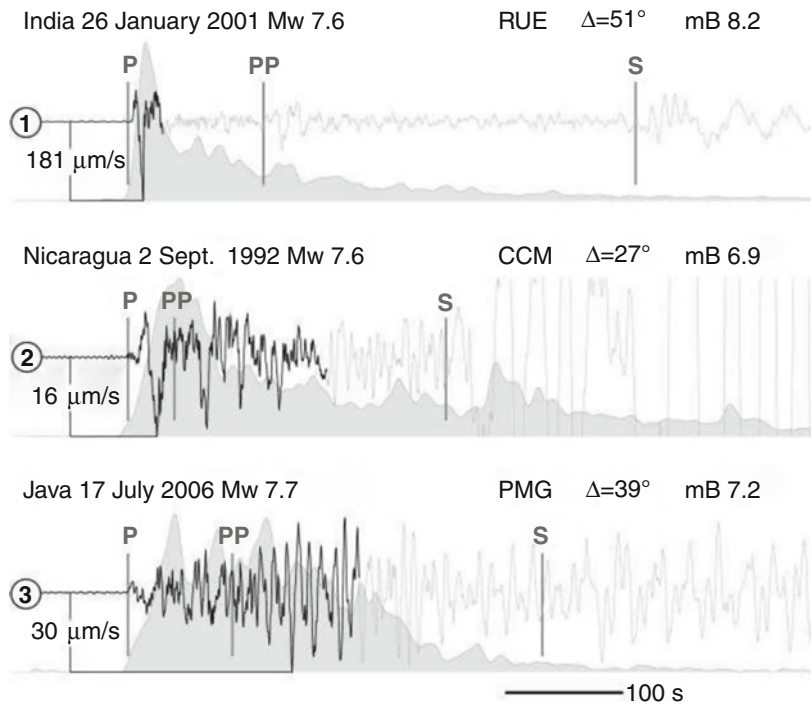
In contrast to measuring M_0 , for which a sufficient condition is that the spectral plateau amplitude u_0 is measured at $f \ll fc$, a reliable estimation of Es requires to integrate the seismic moment rate for about one decade in frequency towards both sides of fc . This requires to take much more higher frequencies into account than in M_0 determination. Since higher frequencies are much more affected by heterogeneities of the real source process and Earth medium and also more difficult to treat in the model calculations, Es and thus M_e estimates are generally less accurate than M_0 and M_w estimates. But in contrast to M_w , M_e is a dynamic measure of earthquake size, which should better be called earthquakes strength. It is controlled by differences in stress drop and the related Es/M_0 ratio, which are associated with differences in rupture velocity and duration (see Fig. 5). Accordingly, the M_e formula can also be written (Bormann and Di Giacomo 2011)

$$\begin{aligned} M_e &= M_w + [\log(\Delta\sigma/2\mu) + 4.7]/1.5 \\ &= M_w + [\log(Es/M_0) + 4.7]/1.5. \end{aligned} \quad (5)$$

Because of the variability of the term in brackets M_e and M_w may differ by several tenths m.u., in the extreme by even more than one m.u. Therefore, M_w and M_e are very important complementary magnitudes reflecting the difference in static and dynamic properties of the earthquake source. M_w alone, being independent on the corner frequency and thus the relative amount

Earthquake Magnitude Estimation,

Fig. 5 Velocity broadband records of three earthquakes with similar M_w but exceptionally different rupture durations and large differences in $m_B(\text{BB})$ values. The gray shaded area marks, as in Fig. 14, the envelope of high-frequency radiation for estimating the rupture duration (Copy of Fig. 3 in Bormann and Saul 2008, SRL © Seismological Society of America)



of high-frequency energy, is not the most suitable magnitude for assessing the actual seismic hazard associated with an earthquake, be it of its damage potential due to strong shaking or of its potential to generate a tsunami. Here even the relationship between the classical long-period M_S and m_b , respectively M_L , may already be more telling than M_w alone. For example, M_L is measured at frequencies of particular interest to earthquake engineers, because they are in the preferred vulnerability range of man-made structures.

Recent publications have shown that considering the difference between M_w and M_e is particularly suitable for a more realistic near real-time assessment of potential effects (e.g., Choy and Kirby 2004; Di Giacomo et al. 2010a; Bormann and Di Giacomo 2011). Generally holds, suitable seismotectonic setting, site conditions, and exposure of vulnerable objects provided, that $M_e \ll M_w$ may be indicative for high tsunamigenic potential and related risk, and vice versa, $M_e \gg M_w$ for high shaking damage potential. But the differential magnitude $M = M_w - M_e$ is also of interest for basic research, e.g., the classification of seismotectonic regions with

respect to their fault maturity and thus capability (Choy and Boatwright 2012). For more details on the relationship between M_w and M_e see also Choy and Boatwright (2012), Bormann and Di Giacomo (2011) and Bormann et al. (2013).

Similarly important it is to consider the difference between M_w and $m_B(\text{BB})$ because the latter correlates very closely with M_e (Fig. 6). Figure 5 shows that for a given M_w $m_B(\text{BB})$ is inverse proportional to the rupture duration and thus proportional with rupture velocity and the related stress drop, as does the ratio E_S/M_0 and thus M_e .

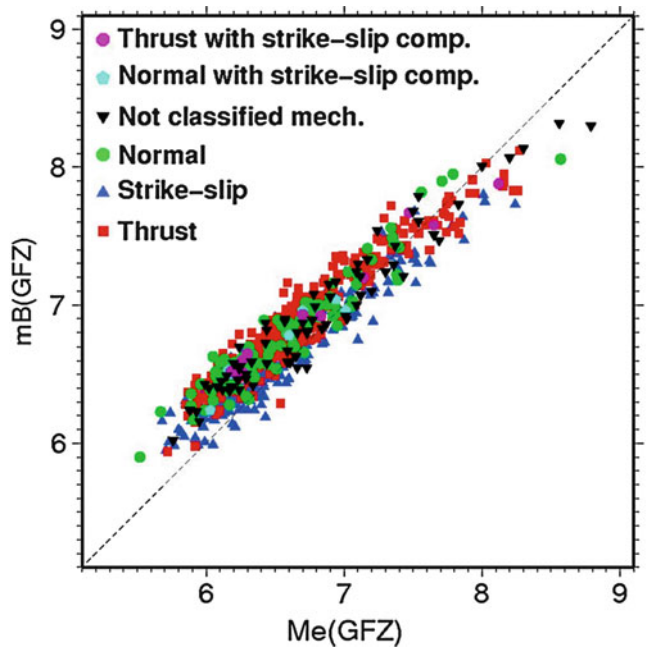
The orthogonal relationship and standard deviation between M_e and $m_B(\text{BB})$ is:

$$M_e = 1.27 \, m_B(\text{BB}) - 1.98 \pm 0.11. \quad (6)$$

From the classical magnitudes, M_S relates best to M_w , also due to the fact that the M_w relationship has been derived via the log E_S - M equation (20) in section Relationships Between Magnitude Scales and Released Seismic Energy E_S . However, Bormann et al. (2009) showed that the new IASPEI broadband standard $M_S(\text{BB})$ correlates on average better with M_w than $M_S(20)$, reducing

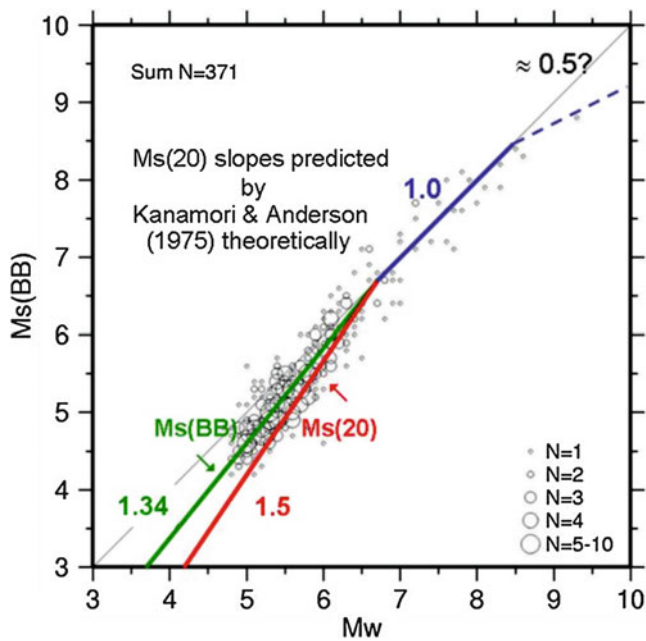
Earthquake Magnitude Estimation,

Fig. 6 Relationship between automatically determined values of $m_B(BB)$ and M_e for earthquakes with different rupture mechanism according to procedures developed at the GFZ German Research Centre for Geosciences by Bormann and Saul (2008) and Di Giacomo et al. (2010a and b) (Courtesy of D. Di Giacomo and J. Saul (2010))



Earthquake Magnitude Estimation, Fig. 7

Data and linear orthogonal regressions between the new IASPEI standard $M_S(BB)$ and $M_w(GCMT)$ as well as the theoretical regression lines for $M_S(20)$ in red, blue, and blue hatched according to Kanamori and Anderson (1975) (Copy of Fig. 3.38 from Bormann et al. 2013 © IASPEI)



to half the systematic underestimation of M_w by $M_S(20)$ for magnitudes less than about 6.7 (Fig. 7). The $M_S(20)$ bias had already been predicted theoretically by Kanamori and Anderson (1975) and empirically confirmed by Ekström and Dziewonski (1988).

Magnitude Scales and the New IASPEI Measurement Standards

Local Magnitude M_L

The definition of M_L and the traditional procedure of its determination have already been

outlined in the introduction. For crustal earthquakes in regions with attenuative properties similar to those of Southern California the formula for the new IASPEI standard M_L (= M_L when written in the IASPEI Seismic Format ISF for magnitude data), has the following form:

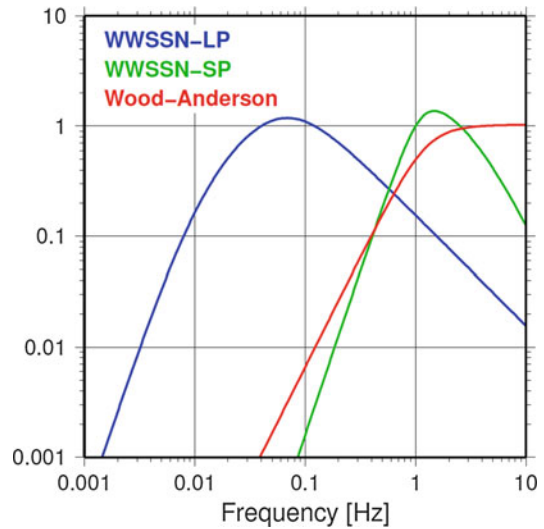
$$M_L = \log_{10} A_{\max} + 1.11 \log_{10} R + 0.00189 * R - 2.09 \quad (7)$$

where $A_{\max}$ = maximum trace amplitude in nm that is measured on the output from a horizontal-component instrument that is filtered so that the response of the seismograph/filter system replicates that of a WA standard seismograph (see ► [Seismic Instrument Response, Correction for](#)) but with a static magnification of 1 (as in Fig. 8), and R = hypocentral distance (see ► [Earthquake Location](#)) in km, typically less than 1,000 km. The constant term in (1), -2.09 , is based on an experimentally determined average static magnification of the WA seismograph of 2080 (Uhrhammer and Collins 1990) and aims at making reported M_L amplitude data unaffected by its uncertainty (Bormann and Dewey 2012).

The calibration function (7) is an expansion of the Hutton and Boore (1987) formula for Southern California (see Fig. 9). For crustal earthquakes in regions with different attenuative properties and for measuring magnitudes with vertical-component seismographs, the standard equation is of the form:

$$M_L = \log_{10}(A_{\max}) + C(R) + D \quad (8)$$

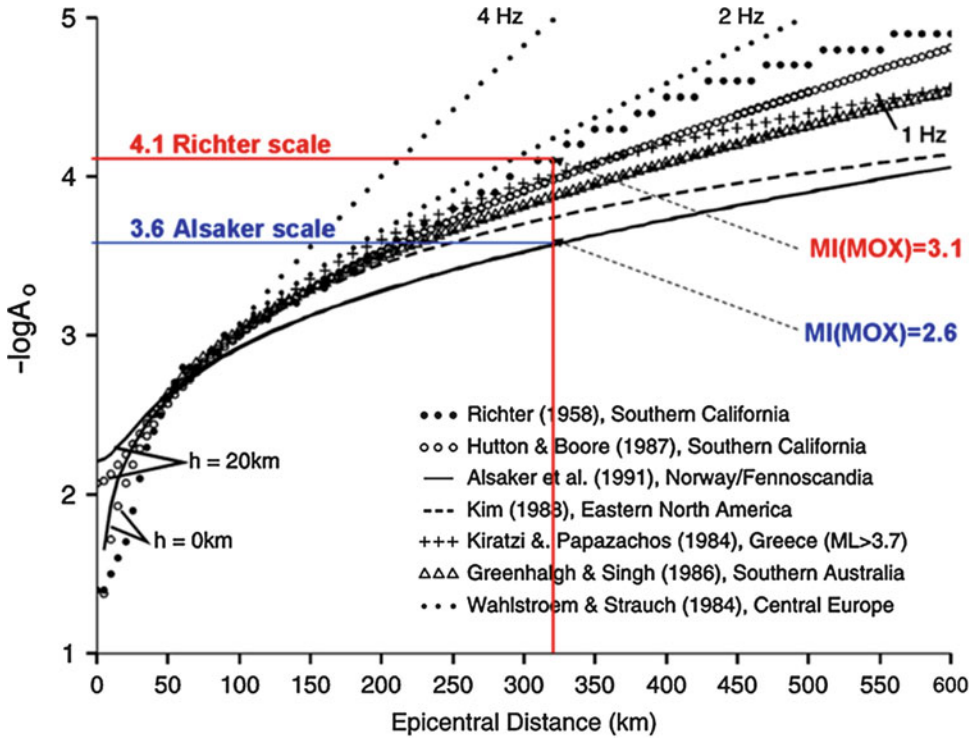
where $A_{\max}$ and R are defined as in Eq. 7, except that A may be measured from a vertical-component instrument, and where $C(R)$ and D have been calibrated to adjust for the different regional attenuation and for any systematic differences between amplitudes measured on vertical instead of horizontal components. Often a station correction term S is still added in those magnitude calibration relationships in order to correct for abnormal (non-average) geologic and/or topographic condition on the recording site.



Earthquake Magnitude Estimation, Fig. 8 Displacement amplitude-frequency responses of the classical seismographs: Wood-Anderson (WA), WWSSN-SP and WWSSN-LP (short-period and long-period standard seismographs of the former US World-Wide Standard Seismograph Network), normalized to a static magnification of 1 at 0.04 Hz for the WWSSN-LP, at 1 Hz for WWSSN-SP and at 4 Hz for the WA seismograph (Copy of Fig. 1 in Bormann and Dewey 2012 © IASPEI)

Figure 9 illustrates how much the M_L calibration functions may differ for regions of different geological age, structure, heat flow and thus attenuation conditions. This highlights the need for scaling other regional M_L formulas to formula (7) so as to assure globally compatible M_L values. Note in this figure the calibration curves published by Wahlström and Strauch (1984). They include besides the curve for calibrating amplitude measurements at 1 Hz also curves for measurements at 2 and 4 Hz. They illustrate the strong influence of frequency-dependent attenuation which has not been considered in Richter's original and all other later developed M_L formulas although $A_{\max}$ may also occur at frequencies in this range. By not accounting for frequency dependence increases the scatter of derived M_L values.

Besides M_L scales that measure $A_{\max}$, typically of relatively weak motion, also M_L^{SM} scales have been developed that measure peak ground accelerations on *strong-motion records*. Also the duration d of the event record (see Fig. 1) can be used for estimating magnitude. But any such



Earthquake Magnitude Estimation, Fig. 9 Calibration functions $\sigma(\Delta) = -\log A_0$ for M_L determination in different regions. Some of them extend well beyond the 600 km distance plotted. But even at much shorter distance, such as the 320 km chosen here for demonstration, the

difference between M_L according to the Richter scale and other regional scales may already amount to 0.5 m.u. for equal measurement amplitudes $A_{\max}$ (Copy of Fig. 3.3 in Bormann et al. 2013 © IASPEI)

alternative local M_d scale has to be developed individually for different regions or seismic networks because of regionally variable attenuation, scattering, signal-to-noise, and other conditions. Moreover, for assuring the compatibility of their values with M_L , these scales have to be calibrated with amplitude-based M_L .

Teleseismic Magnitude Scales

Surface-Wave Magnitudes

IASPEI (2013) recommends to determine for shallow earthquakes with source depth $h < 50\text{--}60$ km two surface-wave magnitude standards, namely

- (a) A purely teleseismic magnitude in the distance range $20^\circ \leq \Delta \leq 160^\circ$ ($1^\circ = 111.12$ km) termed $M_S(20)$, or, in the ISF format

$$M_{S_20} = \log_{10}(A/T) + 1.66 \log_{10} \Delta + 0.3 \quad (9)$$

where A = vertical-component ground displacement in nm measured from the maximum surface-wave trace-amplitude having a period between 18 s and 22 s on a waveform filtered so as to replicate the record of the WWSSN-LP seismograph (Fig. 8), and

- (b) A broadband (BB) surface-wave magnitude that is measured in the wider regional to teleseismic distance range of $2^\circ \leq \Delta \leq 160^\circ$, termed $M_S(BB)$, or in the ISF format:

$$M_{S_BB} = \log_{10}(V_{\max}/2\pi) + 1.66 \log_{10} \Delta + 0.3 \quad (10)$$

where $V_{\max}$ = ground velocity in nm/s associated with the maximum trace-amplitude in the surface-wave train as recorded on a vertical-component seismogram that is proportional to velocity, and where the period T is in the range between $3 \text{ s} < T < 60 \text{ s}$.

Equations 9 and 10 are based on the M_S calibration relationship proposed by Vaněk et al. (1962) which is IASPEI recommended standard since 1967. But in the new standards it is applied to vertical motion measurements only and in Eq. 10 the $\log_{10}(V_{\max}/2\pi)$ term replaces the $\log_{10}(A/T)_{\max}$ term of the original. Equation 9 is since the 1970s exclusive M_S standard at the US National Earthquake Information Center (NEIC). However, most surface-wave amplitude maxima occur at periods outside of the period range accepted by $M_S(20)$ (Fig. 10; Vaněk et al. 1962; Bormann et al. 2009). Moreover, using the standard M_S calibration function exclusively for amplitude readings around 20 s introduces in some ranges systematic distance and magnitude-dependent biases of up to 0.5 m.u. Therefore, the determination of $M_S(\text{BB})$, most easily measured on nowadays widely available velocity BB records, is preferable.

Besides these standards there exist still other surface-wave magnitudes. Bonner et al. (2006) proposed a largely theory-based version, applying multi-bandpass filtering and frequency-dependent attenuation corrections. Yet, according to Bormann et al. (2009), their sophisticated $M_S(V_{\max})$ procedure yields practically the same results as the much simpler $M_S(\text{BB})$ which can be performed easily by any station analyst (Bormann et al. 2009). Another one uses very long-period surface-wave amplitudes with periods between 60 and 400 s, which penetrate into the earth mantle. This “mantle magnitude” M_m (Okal and Talandier 1989) is not prone by saturation and thus suitable also for tsunami early warning (Weinstein and Okal 2005). The Comprehensive Test-Ban Treaty Organization (CTBTO) applies another 20 s surface-wave calibration formula by Rezapour and Pearce (1998) which avoids the systematic distance-dependent biases of $M_S(20)$.

Body-Wave Magnitudes

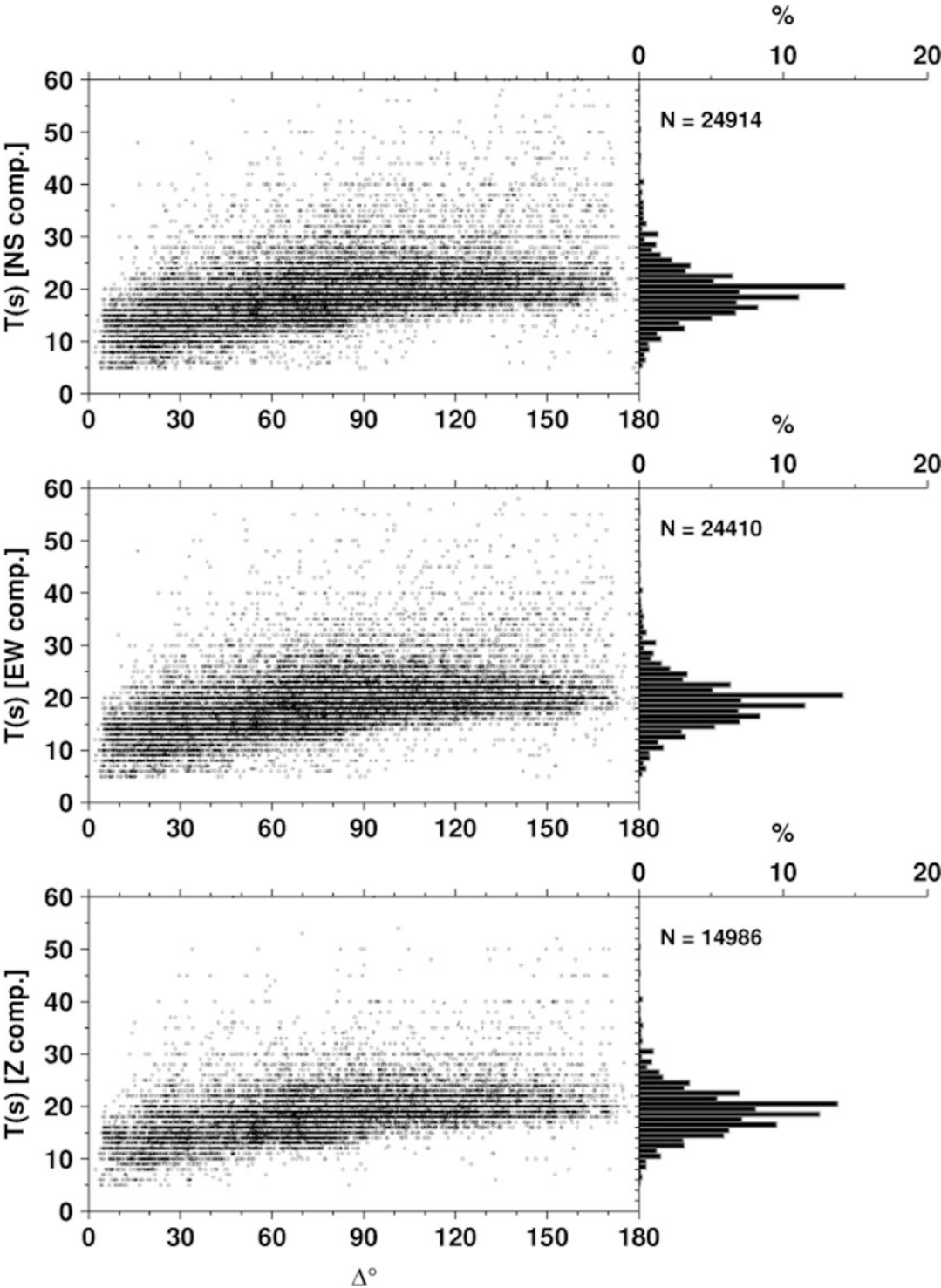
Gutenberg (1945a, b) proposed teleseismic magnitude scales for body waves of type P, PP, and S. This allowed him to assess also the seismic energy released by deep earthquakes. Calibration functions $Q(\Delta, h)$ for amplitude readings on both

vertical and horizontal component records of P and PP waves, respectively on horizontal component S-wave records, were published by Gutenberg and Richter (1956) and the respective magnitudes were termed m_B . Figure 11 shows the Q -values for vertical component P waves as a function of Δ and h . Regrettably, it is the only one still in wide routine use, although S-wave amplitude, period and magnitude data would be of great complementary value for improved studies of the physical properties of Earth matter and seismic source mechanisms. And PP waves would allow to measure body-wave magnitude still with good signal-to-noise ratio in the “core-shadow” range beyond 100° distance where due to the influence of the liquid outer Earth core direct P and S waves are no longer recorded or only recorded as weak long-period diffracted waves.

When elaborating these scales, Gutenberg realized that body-wave magnitudes measured at different sites are compatible and stable only when measuring the ratio $(A/T)_{\max}$ in a wide period range instead of just measuring $A_{\max}$ or measuring $A_{\max}$ in a narrow frequency range. Measuring $(A/T)_{\max}$ corresponds to measuring $V_{\max}/2\pi$ in the IASPEI standard formula (10) for $M_S(\text{BB})$. Accordingly, his general body-wave magnitude formula reads, with A measured in μm :

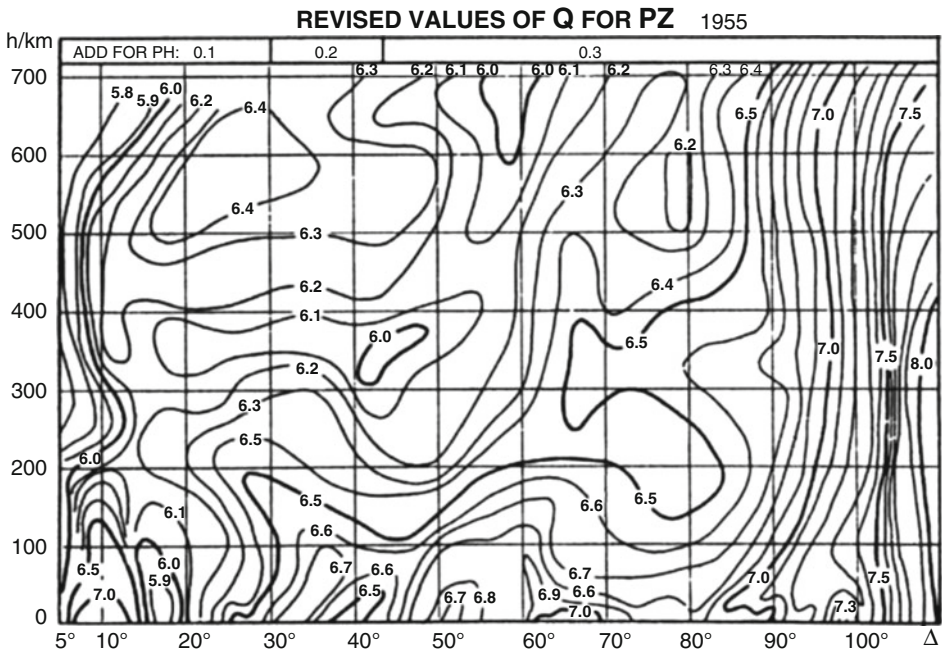
$$m_B = \log_{10}(A/T)_{\max} + Q(\Delta, h). \quad (11)$$

Gutenberg derived the Q -functions by analyzing mostly medium- to long-period body-wave amplitudes in the range $2 \text{ s} < T < 30 \text{ s}$. Therefore he did not account for frequency-dependent attenuation, which becomes significant for P-waves with $T < 4 \text{ s}$ (see Sect. 2.5.4.2 in Bormann et al. 2013). Further, Gutenberg measured always $(A/T)_{\max}$ within the whole P-wave train and not only within a fixed limited time window after the P-wave onset. Yet, this was not considered when introducing, in conjunction with the deployment of the WWSSN in the 1960s, another short-period P-wave magnitude scale m_b . However, as compared to m_B , m_b had the clear advantage of enabling teleseismic



Earthquake Magnitude Estimation, Fig. 10 Three-component distance-period plots for $(A/T)_{\max}$ surface wave readings retrieved at the ISC from printed station/network bulletins for earthquakes that occurred before

1971 and therefore were not dominated by the more narrow-band WWSSN records (Copy of Fig. 3.35 in Bormann et al. 2013 © IASPEI)



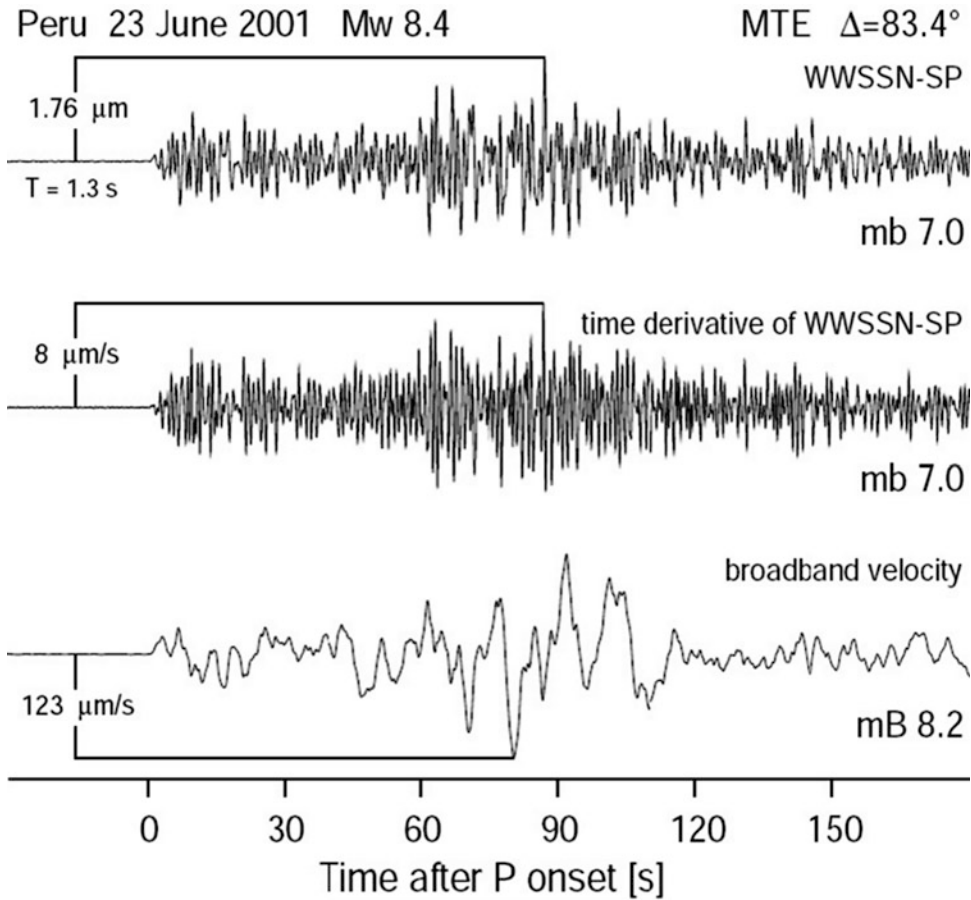
Earthquake Magnitude Estimation, Fig. 11 Calibration values $Q(\Delta, h)$ for vertical (Z) component P-wave amplitudes as a function of epicentral distance Δ and source depth h . They are used for the calculation of the

two body-wave magnitudes m_b and m_B (Figure redrawn from Gutenberg and Richter (1956). For standard tabulated QPZ values see Table 2 in Bormann and Dewey (2012))

magnitudes measured down to about magnitude 4, and thus to classify the many more small earthquakes now being recordable with the much more sensitive instruments now available than those available at Gutenberg’s time.

But m_b still uses the same formula and $Q(\Delta, h)$ values for medium-period P waves as in (5) and measures $A_{\max}$ only on narrow-band WWSSN-SP (see Fig. 3) or similarly filtered short-period records at periods <3 s (mostly around 1 s). This practice became standard in all western countries, which discontinued to measure m_B , although m_b heavily underestimates the body-wave magnitude for large earthquakes (see Introduction). This is mainly due to the earlier-cited phenomenon of spectral saturation (see section Relationships Between Magnitude Scales and Released Seismic Energy E_s). This effect of underestimating earthquake size has been further aggravated by the often too short time window for measuring $A_{\max}$ after the P-wave onset. In early years of the WWSSN, this window had been fixed

to the first few cycles within about 6 s after the onset of P. It is still used by several agencies such as the International Data Center (IDC) of the CTBTO or the China Earthquake Network Center (CENC), despite the 1978 IASPEI recommendation that for small and medium size earthquakes $A_{\max}$ should be measured within 20 s after the first onset and that this time window be extended up to 60 s for very large earthquakes. Even a simple relationship between rupture duration and magnitude, such as Eq. 1, makes it obvious that m_b measurements within the very first few seconds necessarily tend on average to underestimate the size of events with $M > 6$, because it becomes increasingly unlikely that they are based on the true $A_{\max}$ radiated by the considered rupture. But also the extension of the measurement time window up to 60 s will not always be sufficient for really great earthquakes that rupture for several minutes. The new standards for body-wave magnitudes, measuring always the true $A_{\max}$ within the whole P-wave train, avoid this measurement



Earthquake Magnitude Estimation, Fig. 12 Difference between standard $m_B(\text{BB})$ and m_b . Plotted are the vertical-component P-wave records of an M_w 8.4 Peru earthquake at station MTE, Portugal ($\Delta = 83.4^\circ$). On the top trace, m_b is determined by measuring $A_{\max}$ and the related period T from the WWSSN short-period recording that has been simulated from the original BB velocity trace on the bottom. On the second trace the time derivative of the

top trace is depicted in order to make the measurement units directly compatible with the bottom trace. On the bottom trace $V_{\max}$ is measured, yielding $m_B(\text{BB}) = 8.2$, in contrast to standard $m_b = 7.0$. If, however, $A_{\max}$ is measured on the top trace within the first 5 s then one would get $m_b = 6.4$ only (Copy of Fig. 1 in Bormann and Saul 2008, SRL © Seismological Society of America)

time-window component of saturation. For the great Mw 9.3 Sumatra-Andaman earthquake of December 26, 2004, the smallest reported m_b was 5.7, the new standard m_b yielded 7.5. But what also standard m_b cannot avoid is the spectral component of saturation which is so obvious when comparing the lower two traces and related velocity amplitudes of $8 \mu\text{m/s}$ and $123 \mu\text{m/s}$, respectively, in Fig. 12. This spectral component of saturation is strongly reduced by $m_B(\text{BB})$ that is measured for large earthquakes at significantly larger periods (Figs. 3 and 12).

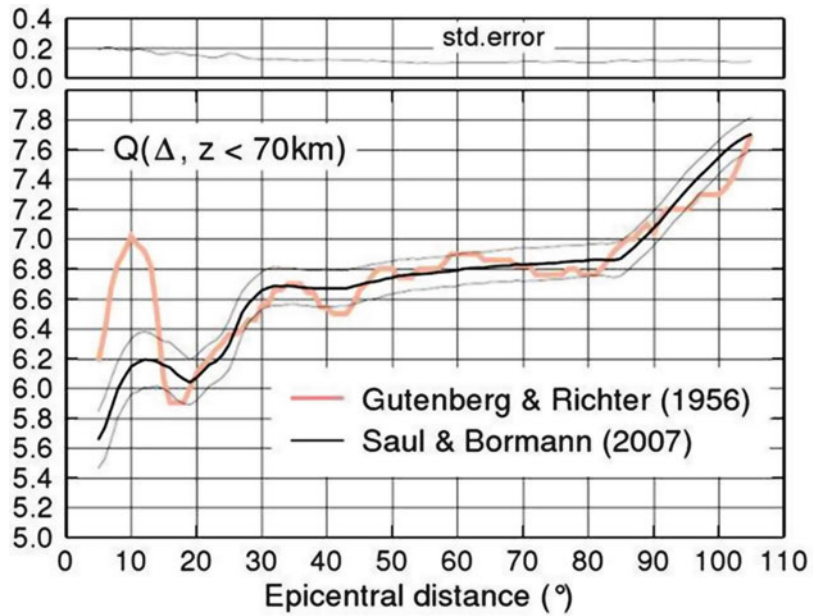
Taking the results of several related studies into account, IASPEI (2005; latest version 2013) recommends two complementary body-wave magnitude standards, namely, a short-period version m_b , or mb in the ISF format:

$$mb = \log_{10}(A/T) + Q(\Delta, h) - 3.0 \quad (12)$$

where A = P-wave ground amplitude in nm, calculated from the maximum trace-amplitude in the entire P-phase train (time spanned by P, the depth phases pP, sP, and possibly also the core

Earthquake Magnitude Estimation,

Fig. 13 Comparison between the original Gutenberg and Richter (1956) QPZ values for vertical component P waves from earthquakes at depth <70 km with preliminary data based on broadband velocity P-wave amplitude measurements by Saul and Bormann (2007). The Saul and Bormann curve is associated with curves representing $\pm$ the standard error, which is also plotted separately at the top of the figure (Copy of Fig. 16 in Bormann and Dewey 2012 © IASPEI)



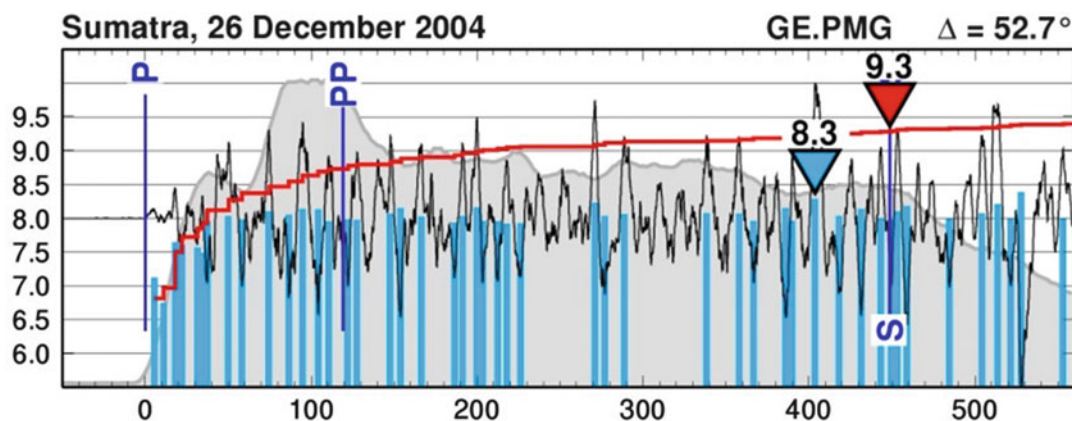
reflection PcP and their codas, and ending preferably before PP); T = period of A in seconds with $T < 3$ s, Δ = epicentral distance in the range $20^\circ \leq \Delta \leq 100^\circ$ and h = focal depth between $0 \text{ km} < h \leq 700 \text{ km}$. Both T and A are measured on output from a vertical-component instrument/filter that simulates a WWSSN-SP seismograph (Fig. 8). Although the early saturation of mb has been reduced by this new definition up to about one magnitude unit, the short-period body-wave magnitude still needs to be complemented – at least for $mb > 5.5$ – by a broadband $m_B(\text{BB})$, or mB_BB in ISF format:

$$mB_BB = \log_{10}(V_{\max}/2\pi) + Q(\Delta, h) - 3.0. \quad (13)$$

with $V_{\max}$ = ground velocity in nm/s associated with the maximum trace-amplitude in the entire P-phase train (as for standard m_b), recorded on a vertical-component seismogram that is proportional to velocity at least in the period range $0.2 \text{ s} < T < 30 \text{ s}$; $Q(\Delta, h)$. The ranges for Δ and h are the same as for standard m_b . As seen from Fig. 11, Gutenberg and Richter (1956) had given calibration values down to $\Delta = 5^\circ$. Since $m_B(\text{BB})$ has proven to be a very suitable earliest available proxy estimator for both seismic moment and

energy magnitude, M_w and M_e (see section [Earthquake Rupture, Radiated Seismic Spectrum, Seismic Moment, Energy and Related Magnitudes](#); Bormann and Saul 2008; 2009a; Bormann et al. 2013), it would be very beneficial to measure $m_B(\text{BB})$ already from distances 5° onwards, e.g., in tsunami early warning schemes. However, the original Gutenberg and Richter QPZ values are not reliable enough in the regional range. A first version of an improved regional calibration relationship has been developed for $m_B(\text{BB})$ (Fig. 13) and is used already operationally in the German-Indonesian Tsunami Early Warning System (GITEWS).

Besides these two body-wave standards there exist several other broadband P-wave magnitudes by various authors which aim at fast estimates of a non-saturating moment magnitude M_w (see Sects. 3.2.7 and 3.2.8 in Bormann et al. 2013). The most successful ones account in different ways also for the rupture duration. One example is given for the great Mw 9.3 Sumatra-Andaman earthquake of 2004 in Fig. 14, analyzed by the fully automated cumulative body-wave magnitude mBc procedure (Bormann and Saul 2009b). Its principle was proposed by Bormann and Khalturin (1975). The justification for the new scale was that Gutenberg and Richter (1956) assumed for



Earthquake Magnitude Estimation, Fig. 14 Plot that illustrates the measurement of the cumulative body-wave magnitude m_{Bc} for the great Sumatra-Andaman earthquake with a rupture duration of about 540 s. The strong P-wave radiation continues well into the time after the first theoretical PP onset. The rupture duration D is estimated from the envelope of high-frequency radiation on a 1–4 Hz filtered record channel (gray shaded area). D is assumed to correspond approximately to the time window from the first P-wave onset until the envelop amplitude

drop to 40 % of its maximum values, i.e., some 16 % of the released high-frequency seismic energy. Blue columns give the m_B values related to subsequent new energy arrivals radiated from sub-rupture events; the red curve shows the development of m_{Bc} with time by summing up significant sub-event amplitudes until saturation is achieved towards the end of the rupture (Cut-out of Fig. 1 from Bormann and Saul 2009b, SRL © granted by Seismological Society of America)

the m_B determination and the relationship derived between m_B and E_S that the P-wave train has on average a simple symmetric waveform radiated by a “point source” (see section [Earthquake Rupture, Radiated Seismic Spectrum, Seismic Moment, Energy and Related Magnitudes](#)). They explicitly stated that they were not sure whether this still holds for very great earthquakes for which they had not enough data. But great earthquakes rupture in a more or less complicated sequence of sub-events. The velocity amplitudes for m_B should therefore be measured individually for major sub-events and summed up toward the end of the rupture process for getting a cumulative m_B estimate, termed m_{Bc} .

Earthquake Early Warning Magnitudes

In recent years great attention has been paid to the development of [Earthquake Early Warning Systems](#) (EEWS). They allow – besides very quick event location – near real-time magnitude estimates from the very first few seconds of acceleration, velocity or displacement records at stations close to the earthquake source. Based on such data, rapid public alarms may be issued

and/or automatically triggered risk mitigation actions taken between the detection of a strong earthquake at the close-up stations and the arrival of strong ground motion at locations slightly further away. EEWS aim at minimizing the area of so-called “blind zones” which are left without advanced warning before the arrival of the S waves which have usually the largest strong-motion amplitudes. This requires very dense and robust seismic sensor networks within a few tens of km from potentially strong earthquake sources. Such dense networks are at present already available in several regions, e.g., in California, Japan, Taiwan, Turkey (Istanbul), and Italy (see [Source Characterization for Earthquake Early Warning](#)).

The principles of rapid magnitude estimates are rather different from those mentioned above and largely based on much debated concepts such as the hypothesis of the deterministic nature of earthquake rupturing. For example, Olson and Allen (2005) claim to be able to estimate, from the maximum period of the first arriving P waves within the first 4 s, the size of earthquakes up to magnitude 8 with an average absolute deviation

of 0.54 m.u. using many low-pass filtered velocity records within 100 km epicentral distance. However, Rydelek and Horiuchi (2006) could not confirm that such a dominant frequency scaling with magnitude exists. Also extensive experiments run at the Taiwan EEWS, aimed at reducing the reporting time from 25 s down to about 8 s by using only the records of the first 8 stations within less than 21 km epicentral distance (Wu and Kanamori 2005) came to sobering results. Estimating the magnitude via the magnitude dependence of the average period τ_c of the P-wave, as well as the peak displacement amplitudes (if > 0.1 cm) within the first 3 s of the low-pass filtered displacement signal after the P-wave onset the authors concluded:

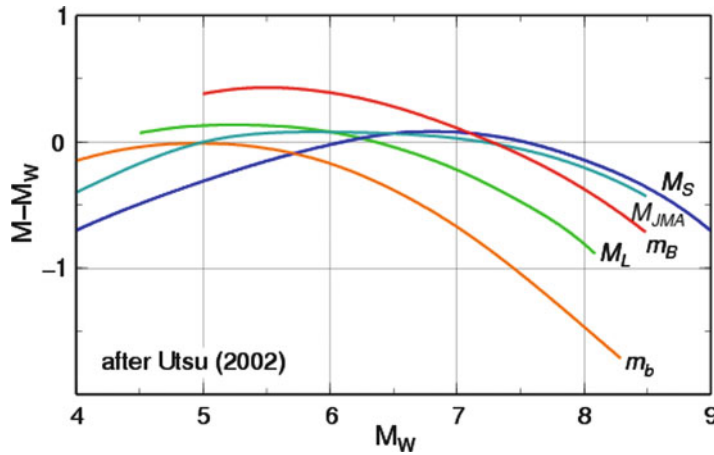
*“... For EWS applications, if $\tau_c < 1$ sec, the event has already ended or is not likely to grow beyond $M > 6$. If $\tau_c > 1$ sec, it is likely to grow, but how large it will eventually become, cannot be determined. In this sense, the method provides a **threshold warning**.” And: “... If $\tau_c > 1$ sec, the event is potentially damaging. If it is larger than 2 sec, the event is almost certainly damaging. Combining this information with other data, such as the initial velocity and displacement amplitudes, would allow the damage potential of the event to be assessed more accurately”.*

More can indeed not be said within the first 3–4 s of a record, especially not for strong and great earthquakes. This fully agrees with relationship (8) between average rupture duration τ_{rup} and M . For earthquakes with $M \leq 6$ and $\tau_{rup} \leq 6$ s it can indeed be expected that, within a measurement time window of 3–4 s, the maximum amplitude and related period of the P-wave train can be measured and are representative for the earthquake magnitude but not for magnitudes much larger than 6 and rupture durations of several 10 to several 100 s. Then, inevitably, as for the current near real-time magnitude estimates in tsunami early warning systems (TEWS), much larger measurement time windows up to about one minute (e.g., Lomax and Michelini 2012) or even longer for mega-earthquakes may be required. This general conclusion is fully confirmed by results from Southern California where the magnitude has been estimated by measuring the P-wave amplitudes

within the first three seconds of the record. These Pd magnitudes agreed with catalog magnitudes with a standard deviation of ± 0.18 m.u. for events with magnitudes < 6.5 but saturated for larger ones and could not be used to estimate magnitudes ≥ 6.7 (Wu and Zhao 2006). For a summary of related studies and information on several representative EEWS, their magnitude concepts and performance see Sect. 3.2.6.5 in Bormann et al. (2013).

Relationships Between Magnitude Scales and Released Seismic Energy E_s

The local and teleseismic magnitudes discussed above are not identical (see Fig. 15). They only agree at some specific magnitude values and differ at others. Moreover, all classical magnitude scales show for large or very great earthquakes more or less pronounced underestimation of earthquake size as compared to a non-saturating purely displacement amplitude-based magnitude such as the moment magnitude M_w . But underestimation of the moment related earthquake size may also occur if magnitudes that are based on the measurement of A/T , such as m_b and $M_S(20)$, are measured outside of the period range of maximum seismic energy release. This is obvious from the nonlinear relationships of these classical magnitudes with M_w (Fig. 15). The reasons have been explained in section [Earthquake Rupture, Radiated Seismic Spectrum, Seismic Moment, Energy and Related Magnitudes](#). This notwithstanding, most classical relationships between different types of magnitudes have been published only as simple linear standard regression relationships that maybe rather biased by noisy data because they assume that only one of the compared magnitudes is afflicted with measurement errors. Only recently has it become common to calculate general orthogonal linear and even nonlinear regression relationships. For a summary on this issue see Sect. 3.2.9 in Bormann et al. (2013). Only the most important classical and a few recently derived relationships of interest are presented here.



Earthquake Magnitude Estimation, Fig. 15 Average relationships of different common types of magnitude scales with respect to the non-saturating moment magnitude M_w . M_{JMA} is determined by the Japanese

Meteorological Agency (Modified Fig. 1 of Utsu (2002). Reproduction of Fig. 4 in Bormann and Saul 2009a © Springer; with kind permission from Springer Science + Business Media)

According to Gutenberg and Richter (1956)

$$m_B = 0.63 M_S + 2.5. \quad (14)$$

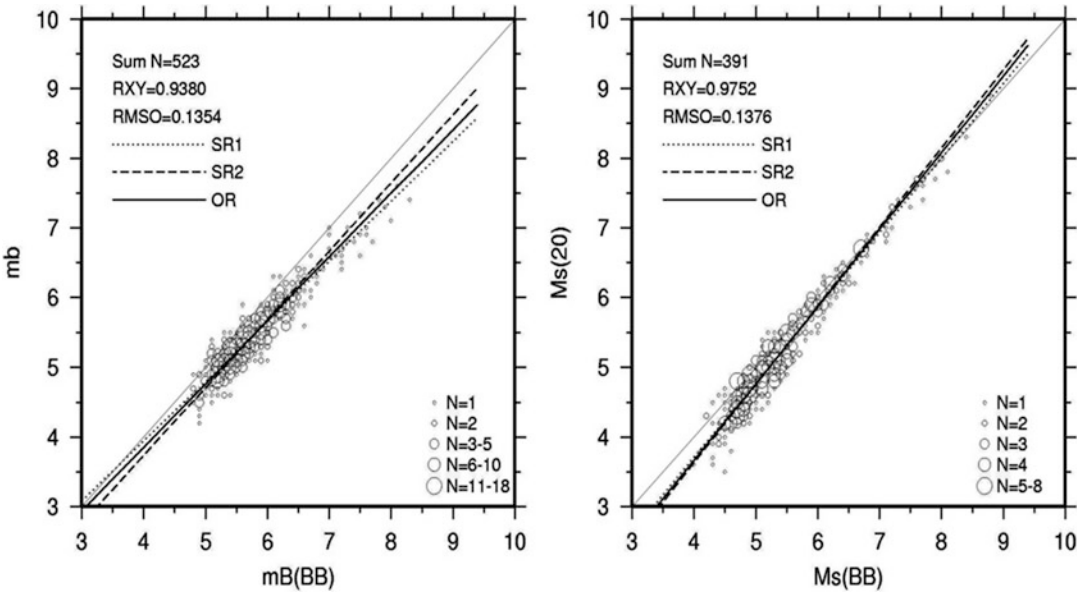
This relationship has been very well reproduced between the new standard magnitudes $m_B(BB)$ and $M_S(BB)$ (Bormann et al. 2009) in an even wider magnitude range than that investigated by Gutenberg and Richter (1956). But Eq. 14 differs strongly from the relationship between short-period m_b and $M_S(20)$ (Gordon 1971):

$$m_b = 0.47 M_S(20) + 2.79. \quad (15)$$

The relationship between the two new IASPEI standard magnitudes for body waves and surface-waves, respectively, are depicted in Fig. 16 and their formulas are given in the figure caption.

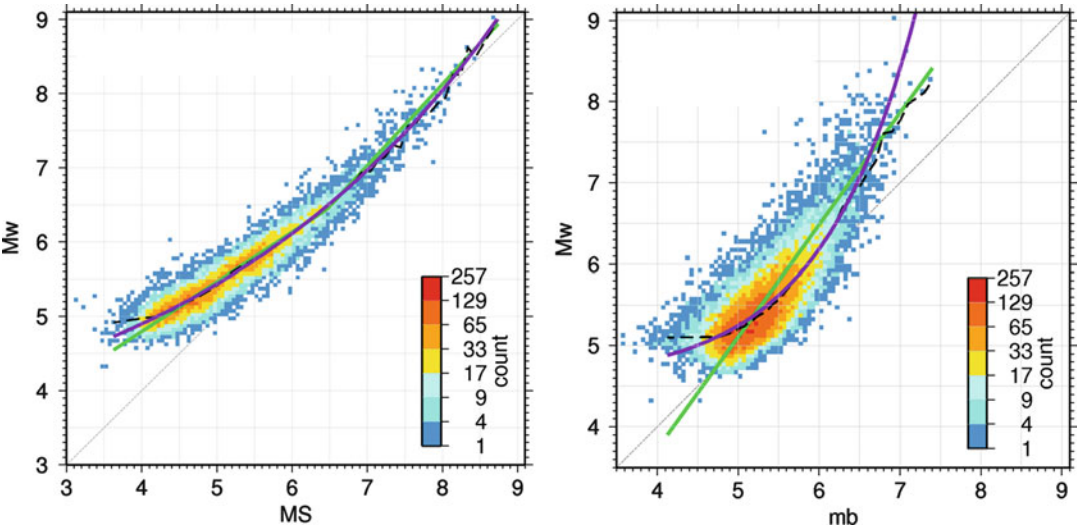
Regression relationships between magnitudes are frequently used to convert data of one type of magnitude into another one. For example, Gutenberg used – in an effort to unify different magnitude estimates per event into a single number – relationship (9) for converting M_S data into m_B which he considered superior to M_S , better physically defined, applicable also to deep earthquakes and less affected by uncertainties in estimates of source depth. Nowadays it

has become common to convert classical magnitudes via empirical conversion relationships into M_w which is often considered as the physically best defined and with smallest measurement errors of about $\leq \pm 0.1$ m.u. afflicted magnitude. Such a unification of magnitude estimates is desirable especially for *seismic hazard assessment* (see ► [Probabilistic Seismic Hazard Models](#)). The latter strongly depends on the parameters **a** and **b** derived from a Gutenberg and Richter (1954) relationship $\log N = \mathbf{a} + \mathbf{b}M$ between magnitude M and the number N of earthquakes that occur on average in a given period of time, usually normalized as of per year (see ► [Earthquake Recurrence](#)). However, reliable instrumentally measured M_w data as in the Global Centroid Moment Tensor catalog (<http://www.globalcmt.org>) are available and complete only since 1975 and for $M_w \geq 5.5$. This time span is too short for reliable long-term seismic hazard assessment. To include earlier events into the $\log N - M_w$ analysis one has to try the conversion of classical M_S , or of the most frequently determined m_b , into M_w . The rather good correlation between M_S and M_w for larger earthquakes is no wonder because the M_w formula (3) has been scaled by Kanamori (1977) and Hanks and Kanamori (1979) via the $\log E_S - M_S$ relationship (20) to M_S . The most recent, both linear and



Earthquake Magnitude Estimation, Fig. 16 Relationships between the two new IASPEI standard body-wave magnitudes (*left panel*) and surface-wave magnitudes $M_s(20)$ and $M_s(BB)$ (*right panel*). The orthogonal

regression relations are $m_b = 0.91 m_B(BB) + 0.20$ and $M_s(20) = 1.10 M_s(BB) - 0.75$ (Cut-out of Fig. 11 in Bormann et al. 2009, BSSA © Seismological Society of America)



Earthquake Magnitude Estimation, Fig. 17 Training data sets for deriving the ISC-GEM $M_s - M_w$ (*left*) and $m_b - M_w$ (*right*) general orthogonal linear (*green*) and exponential (*blue*) conversion relationships. The dashed

black polygon line represents the median values for separate magnitude bins (Figure modified and compiled according to Figs. 14 and 17 of Di Giacomo et al. 2014 © PEPI)

nonlinear regression relationships of M_s and m_b with M_w have been calculated at the International Seismological Centre (ISC) in the framework of *The Global Earthquake Model (GEM)* project for

deriving a Global Instrumental Earthquake Catalogue (1900–2009) (Di Giacomo et al. 2014). Figure 17 shows these respective data together with the orthogonal linear and standard

exponential regression lines. The related equations for $M_w - M_S$ are given in (16–18),

One realizes, when compared with $M_S - M_w$, the much larger data scatter and thus much larger uncertainty of estimating M_w via m_b . Therefore, the latter conversion should be avoided whenever M_S data are available. But even such proxy M_w estimates via M_S are afflicted with at least twice the error of direct high quality instrumental M_w estimates. For most practical applications, however, it may be still acceptable to use for M_S conversion into M_w instead of the more accurate exponential relationship

$$M_w = e^{(-0.222+0.233 \times M_S)} + 2.863 \quad (16)$$

the two orthogonal linear relationships

$$M_w = 0.67M_S + 2.13 \quad \text{for } M_S \leq 6.47 \quad (17)$$

and

$$M_w = 1.10M_S - 0.67 \quad \text{for } M_S > 6.47 \quad (18)$$

However, because of the incompleteness of GCMT data for $M_w < 5.5$ these relationships should not be applied for converting $M_S < 5$ into M_w . With respect to deviations between standard and orthogonal linear and exponential conversion relationships, especially for m_b and M_S into M_w and their extension into lower magnitude ranges on the basis of regional M_w catalogs, see Lolli et al. (2013) and Sect. 3.2.9.1 in Bormann et al. (2013) with the references given therein.

Gutenberg and Richter (1956) published a semiempirical relationship between the classical m_B and released seismic energy Es

$$\log Es = 2.4m_B - 1.2 \quad (\text{with } Es \text{ in units of Joule}), \quad (19)$$

because only m_B measured the maximum ground motion velocity, proportional to $(A/T)_{\max}$, in a one decade wide period range. However, since in western countries m_B was no longer measured after the 1960s, a proposal by Richter (1958) to insert (14) into (19) was followed and this gave

$$\log Es = 1.5 M_S + 4.8 \quad (20)$$

with Es in units of Joule and M_S meaning an $M_S(20)$. However, 20 s displacement amplitudes are not very representative for the radiated seismic energy, especially not for smaller earthquakes that radiate dominantly higher-frequency energy.

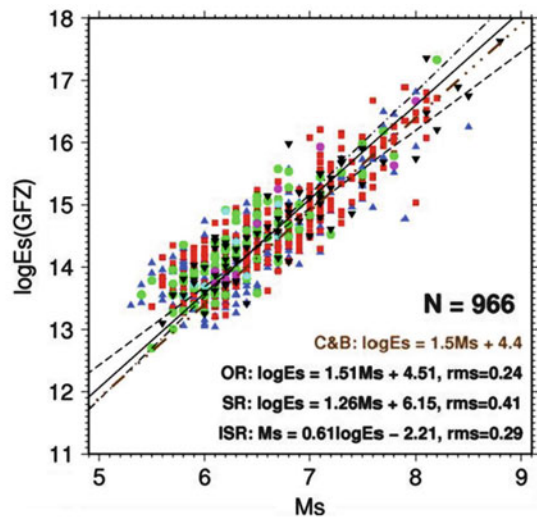
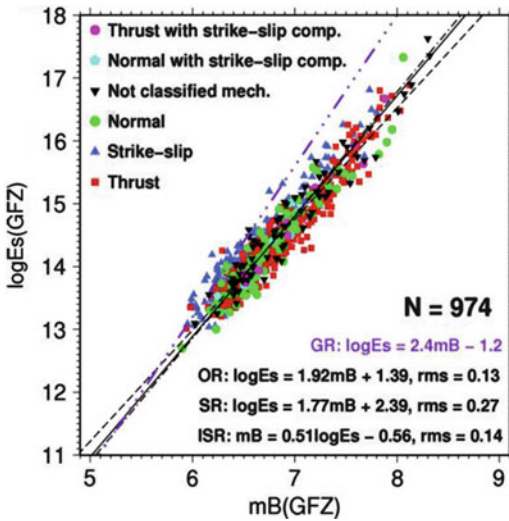
Nevertheless, Eq. 20 has become since the 1960s the most widely used rough estimator for seismic energy release via magnitude and it was instrumental for deriving the non-saturating moment magnitude scale M_w (see section [Earthquake Rupture, Radiated Seismic Spectrum, Seismic Moment, Energy and Related Magnitudes](#)). Moreover, it has been modified by Choy and Boatwright (1995) when correlating direct Es measurements with NEIC $M_S(20)$ data and assuming the same slope as in (20) [see relationship (22)].

Another relationship has been derived by Kanamori et al. (1993) between M_L and Es for Southern California, valid in the range $1.5 < M_L < 6.0$:

$$\log Es = 1.96 M_L + 2.05. \quad (21)$$

Formula (21) is very close to a theoretically expected general relationship $\log Es \propto 2.0 M$ because Es should be proportional to squared ground motion velocity amplitudes. This is obviously well approximated by high-frequency $A_{\max}$ but not ideally by more long-period P- and surface waves [see formulas (19) and (20)]. Differences in the spreading terms and attenuation of body- and surface waves may also influence the slope of empirically determined regression relationships. Anyway, all these relationships allow only rough estimates of the released seismic wave energy, probably within an order of magnitude.

The uncertainty of Es estimates may nowadays be reduced, down to a factor of about 3–5, by integrating squared velocity broadband records with good signal-to-noise ratio in the teleseismic range (see section [Earthquake Rupture, Radiated Seismic Spectrum, Seismic Moment, Energy and Related Magnitudes](#);



Earthquake Magnitude Estimation, Fig. 18 Orthogonal (OR) and standard regressions (SR and its inverse ISR) of $\log E_s$ data for different earthquake source mechanisms, determined by the procedure of Di Giacomo et al. (2008 and 2010b) plotted over m_B (BB) data (left-hand panel),

respectively M_s (20) data of the NEIC (right-hand panel). GR Gutenberg and Richter (1956) relationship, C&B Choy and Boatwright (1995) relationship (Compiled from Figs. 3.85 and 3.86 in Bormann et al. 2013 © IASPEI)

Choy and Boatwright 2012; Di Giacomo and Bormann 2011). Regressing such modern energy measurements with different magnitudes it could be shown that indeed standard m_B (BB) correlates, as expected by Gutenberg, much better with $\log E_s$ than M_s (20). This allows rather reliable energy estimates via m_B (BB). But the classical relationship (19) could not be reproduced well with modern E_s data. On the other hand, although with double data scatter than with m_B (BB), the Choy and Boatwright (1995) relationship

$$\log E_s = 1.5M_s(20) + 4.4 \quad (22)$$

could be well reproduced by orthogonal regression (Fig. 18).

Summary

Magnitudes are one of the most important earthquake parameter data. When properly determined and applied they may allow quick estimates of the tectonic effect of actual earthquakes in terms of rupture area and slip and/or of the released seismic energy in different frequency ranges and thus

of the potential strength of earthquake shaking to be expected. Moreover, in the long run, they also permit to quantify the seismic activity and its variability in space and time, which is a prerequisite for seismic hazard assessment. This necessitates, however, the compatibility and long-term stability of magnitude estimates of the various types on a global scale. The essay shows that just this is often not the case. Procedural differences in magnitude determination, even of supposedly the same type, may occasionally reach even more than one magnitude unit, especially for the most seldom but therefore most important and devastating great earthquakes. In view of the logarithmic scaling of magnitude values this may correspond to wrong estimates of released seismic energy by a factor of 1,000 times or even more. This highlights the urgency of international standards in magnitude measurements. For the most important classical magnitudes such standards have recently been agreed upon by IASPEI and recommended for global operational application. Their principles and underlying physical considerations, advantages and limitations are outlined in the essay, also in relation to other types of magnitudes such as modern M_w and M_e

as well as of their near real-time proxy estimates in the earthquake and tsunami early warning context. Also given are some of the most important correlation relationships between different types of magnitudes but with cautioning remarks on their often uncritical application in efforts to homogenize earthquake catalogs in terms of magnitude. In fact, there is no single type of magnitude that can sufficiently well describe both the geometric size and kinematic-dynamic effects of earthquakes. What is needed is a deeper understanding and conscious use of the beneficial complementarity of different magnitude scales. For comprehensive information on this complex issue reference is made to the online IASPEI New Manual of Seismological Observatory Practice (NMSOP-2; <http://nmsop.gfz-potsdam.de>).

Cross-References

- [Earthquake Location](#)
- [Earthquake Mechanism and Seafloor Deformation for Tsunami Generation](#)
- [Earthquake Mechanisms and Stress Field](#)
- [Earthquake Mechanisms and Tectonics](#)
- [Earthquake Recurrence](#)
- [Moment Tensors: Decomposition and Visualization](#)
- [Probabilistic Seismic Hazard Models](#)
- [Seismic Actions Due to Near-Fault Ground Motion](#)
- [Seismic Instrument Response, Correction for](#)
- [Seismic Network and Data quality](#)
- [Sensors, Calibration of](#)
- [Site Response: Comparison Between Theory and Observation](#)

References

- Baumbach M, Bormann P (2011) EX 3.4: Determination of source parameters from seismic spectra. In: Bormann P (ed) (2012) <http://nmsop.gfz-potsdam.de>
- Bonner JL, Russel DR, Harkrider DG, Reiter DT, Herrmann RB (2006) Development of a time-domain, variable-period surface-wave magnitude measurement procedure for application at regional and teleseismic distances, part II: application and Ms-mb performance. *Bull Seismol Soc Am* 96(2):678–696
- Bormann P, Khalturin VI (1975) Relations between different kinds of magnitude determinations and their regional variations. In: Proceedings XIVth general assembly of the European seismological commission, Trieste, pp 16–22 September 1974. Published by the National komitee Geod. Geophys, AdW der DDR, Berlin, pp 27–39
- Bormann P, Saul J (2008) The new IASPEI standard broadband magnitude m_B . *Seismol Res Lett* 79: 698–705
- Bormann P, Saul J (2009a) Earthquake magnitude. In: Meyers A (ed) *Encyclopedia of complexity and systems science*, vol 3. Springer, Heidelberg/New York, pp 2473–2496
- Bormann P, Saul J (2009b) A fast, non-saturating magnitude estimator for great earthquakes. *Seismol Res Lett* 80:808–816
- Bormann P, Di Giacomo D (2011) The moment magnitude M_w and the energy magnitude M_e : common roots and differences. *J Seismol* 15:411–427
- Bormann P, Dewey JW (2012) IS 3.3: The new IASPEI standards for determining magnitudes from digital data and their relation to classical magnitudes. In: Bormann P (ed) (2012) <http://nmsop.gfz-potsdam.de>
- Bormann P, Liu R, Ren X, Gutdeutsch R, Kaiser D, Castellaro S (2007) Chinese national network magnitudes, their relation to NEIC magnitudes, and recommendations for new IASPEI magnitude standards. *Bull Seismol Soc Am* 97:114–127
- Bormann P, Liu R, Xu Z, Ren K, Zhang L, Wendt S (2009) First application of the new IASPEI teleseismic magnitude standards to data of the China National Seismographic Network. *Bull Seismol Soc Am* 99(3): 1868–1891
- Bormann P, Engdahl ER, Kind R (2012) Chapter 2: Seismic wave propagation and Earth models. In: Bormann P (ed) (2012) <http://nmsop.gfz-potsdam.de>
- Bormann P, Di Giacomo D, Wendt S (2013) Chapter 3: Seismic sources and source parameters In: Bormann P (ed) (2012) <http://nmsop.gfz-potsdam.de>
- Braunmiller J, Kradolfer U, Baer M, Giardini D (2002) Regional moment tensor determinations in the European-Mediterranean area – initial results. *Tectonophysics* 356(1–3):5–22
- Brune JN (1970) Tectonic stress and the spectra of shear waves from earthquakes. *J Geophys Res* 75:4997–5009
- Choy GL, Boatwright J (1995) Global patterns of radiated seismic energy and apparent stress. *J Geophys Res* 100:18205–18228
- Choy GL, Kirby S (2004) Apparent stress, fault maturity and seismic hazard for normal-fault earthquakes at subduction zones. *Geophys J Int* 159:991–1012
- Choy GL, Boatwright J (2012) IS 3.6: Radiated seismic energy and energy magnitude. In: Bormann P (ed) (2012) <http://nmsop.gfz-potsdam.de>

- Di Giacomo D, Bormann P (2011) Earthquake energy. In: Gupta H (ed) *Encyclopedia of solid earth geophysics*. Springer, Berlin, pp 233–236
- Di Giacomo D, Grosser H, Parolai S, Bormann P, Wang R (2008) Rapid determination of M_e for strong to great shallow earthquakes. *Geophys Res Lett* 35, L10308. doi:10.1029/2008GL033505
- Di Giacomo D, Parolai S, Bormann P, Grosser H, Saul J, Wang R, Zschau J (2010a) Suitability of rapid energy magnitude estimations for emergency response purposes. *Geophys J Int* 180:361–374
- Di Giacomo D, Parolai S, Bormann P, Grosser H, Saul J, Wang R, Zschau J (2010b) Erratum to “Suitability of rapid energy magnitude estimations for emergency response purposes.”. *Geophys J Int* 181:1725–1726
- Di Giacomo D, Bondár I, Storchak DA, Engdahl ER, Bormann P, Harris J (2014) ISC-GEM: global instrumental earthquake catalogue (1900–2009), III. Re-computed M_S and m_b , proxy M_W , final magnitude composition and completeness assessment. *Phys Earth Planet Int* 239:33–47
- Dziewonski AM, Chou TA, Woodhouse JH (1981) Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *J Geophys Res* 86(2):2825–2852
- Ekström G, Dziewonski AM (1988) Evidence of bias in estimations of earthquake size. *Nature* 332:319–323
- Gordon DW (1971) Surface-wave versus body-wave magnitude. *Earthquake Notes* 42(3–4):20–28
- Granville JP, Richards G, Kim W-Y, Sykes LR (2005) Understanding the differences between three teleseismic m_b scales. *Bull Seism Soc Am* 95(5):1809–1824
- Grünthal G (2011) Earthquakes, Intensity. In: Gupta H - (ed) *Encyclopedia of solid earth geophysics*. Springer, Berlin, pp 207–2018
- Gutenberg B (1945a) Amplitude of surface waves and magnitude of shallow earthquakes. *Bull Seismol Soc Am* 35:3–12
- Gutenberg B (1945b) Amplitudes of P, PP, and S and magnitude of shallow earthquakes. *Bull Seismol Soc Am* 35:57–69
- Gutenberg B (1945c) Magnitude determination of deep-focus earthquakes. *Bull Seismol Soc Am* 35:117–130
- Gutenberg B, Richter CF (1954) *Seismicity of the Earth and associated phenomena*, 2nd edn. Princeton University Press, Princeton
- Gutenberg B, Richter CF (1956) Magnitude and energy of earthquakes. *Annali di Geofisica* 9:1–15
- Hanks C, Kanamori H (1979) A moment magnitude scale. *J Geophys Res* 84:2348–2350
- Hayes GP, Rivera L, Kanamori H (2009) Source inversion of the W-phase: real-time implementation and extension to low magnitudes. *Seismol Res Lett* 80(5):817–822
- IASPEI (2013) Summary of magnitude working group recommendations on standard procedures for determining earthquake magnitudes from digital data. http://www.iaspei.org/commissions/CSOI/Summary_WG_recommendations_20130327.pdf
- Kanamori H (1977) The energy release in great earthquakes. *J Geophys Res* 82:2981–2987
- Kanamori H, Anderson DL (1975) Theoretical basis of some empirical relations in seismology. *Bull Seismol Soc Am* 65(5):1073–1095
- Kanamori H, Mori J, Hauksson E, Heaton TH, Hutton LK, Jones LM (1993) Determination of earthquake energy release and M_L using TERRASCOPE. *Bull Seismol Soc Am* 83(2):330–346
- Kanamori H, Rivera L (2008) Inversion of W phase: speeding up seismic tsunami warning. *Geophys J Int* 175:222–238
- Lolli A, Gasperini P, Vannucci G (2013). Empirical conversion between teleseismic magnitudes (m_b and M_s) and moment magnitude (M_w) at the global, Euro-Mediterranean and Italian scale. *Geophysical Journal International* (submitted), 199(2):805–828
- Lomax A, Michelini A (2012) Tsunami early warning within 5 minutes. *Pure and Applied Geophysics*, 169 pp
- Musson RMW, Cécic I (2011) Chapter 11: Intensity and intensity scales. In: Bormann P (ed) (2012) <http://nmsop.gfz-potsdam.de>
- Okal EA, Talandier J (1989) M_m : a variable-period mantle magnitude. *J Geophys Res* 94:4169–4193
- Olson EL, Allen R (2005) The deterministic nature of earthquake rupture. *Nature* 438:212–215
- Rezapour M, Pearce RG (1998) Bias in surface-wave magnitude M_S due to inadequate distance corrections. *Bull Seismol Soc Am* 88(1):43–61
- Richter CF (1935) An instrumental earthquake magnitude scale. *Bull Seismol Soc Am* 25:1–32
- Richter CF (1958) *Elementary seismology*. W.H. Freeman, San Francisco/London
- Rydelek P, Horiuchi S (2006) Is earthquake rupture deterministic? *Nature* 444:E5–E6
- Saul J, Bormann P (2007) Rapid estimation of earthquake size using the broadband P-wave magnitude m_B . *Eos, Transactions, American Geophysical Union*, AGU 2007 fall meeting, 88(52), Suppl., Abstract S53A-1035
- Uhrhammer RA, Collins ER (1990) Synthesis of Wood-Anderson seismograms from broadband digital records. *Bull Seism Soc Am* 80:702–716
- Utsu T (2002) Relationships between magnitude scales. In: Lee WHK, Kanamori H, Jennings PC, Kisslinger C (eds) *International handbook of earthquake and engineering seismology*, Part A. Academic, Amsterdam, pp 733–746
- Vaněk J, Zátopek A, Kárník V, Kondorskaya NV, Riznichenko YV, Savarenski EF, Solov'ev SL, Shebalin NV (1962) Standardization of magnitude scales. *Izvestiya Academy of Sciences of the USSR, Geophys Series*, pp 108–111 (English translation)
- Wahlström R, Strauch W (1984) A regional magnitude scale for Central Europe based on crustal wave attenuation. *Seismological Department, University of Uppsala, Report No. 3–84*, 16 pp

- Weinstein SA, Okal EA (2005) The mantle wave magnitude M_m and the slowness parameter Θ : Five years of real-time use in the context of tsunami warning. *Bull Seismol Soc Am* 95:779–799
- Wu YM, Kanamori H (2005) Experiment on an onsite early warning method for the Taiwan early warning system. *Bull Seismol Soc Am* 95(1):347–353
- Wu YM, Zhao L (2006) Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning. *Geophys Res Lett* 33. doi:10.1029/2006GL026871

Earthquake Mechanism and Seafloor Deformation for Tsunami Generation

Eric L. Geist¹ and David D. Oglesby²

¹U.S. Geological Survey, Menlo Park, CA, USA

²Department of Earth Sciences, University of California, Riverside, Riverside, CA, USA

Synonyms

Seismogenic tsunami; Tsunami forcing; Tsunami initial conditions; Tsunamigenic

Introduction

Tsunamis are generated in the ocean by rapidly displacing the entire water column over a significant area. The potential energy resulting from this disturbance is balanced with the kinetic energy of the waves during propagation. Only a handful of submarine geologic phenomena can generate tsunamis: large-magnitude earthquakes, large landslides, and volcanic processes. Asteroid and subaerial landslide impacts can generate tsunami waves from above the water. Earthquakes are by far the most common generator of tsunamis. Generally, earthquakes greater than magnitude (M) 6.5–7 can generate tsunamis if they occur beneath an ocean and if they result in predominantly vertical displacement. One of the greatest uncertainties in both deterministic and probabilistic hazard assessments of tsunamis is computing seafloor deformation for earthquakes

of a given magnitude. This entry reviews past methodologies and current developments of seismogenic tsunami generation models.

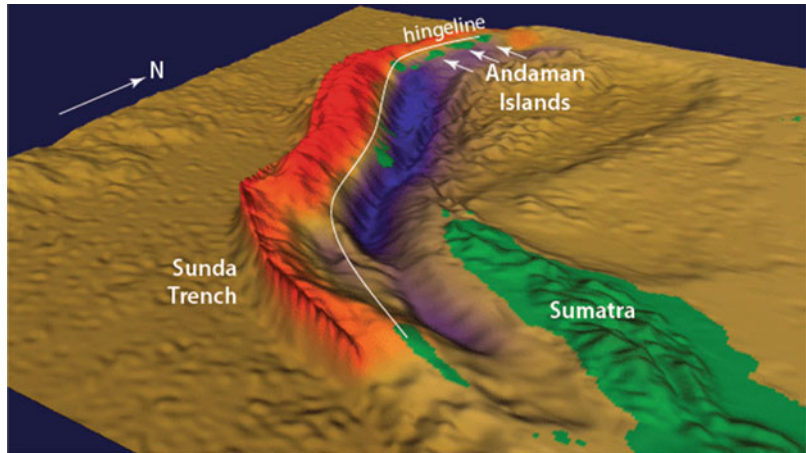
Subduction zones are the tectonic plate boundaries where most seismogenic tsunamis are sourced. Three factors combine to explain this observation: (1) most subduction zones occur beneath oceans, (2) most of the world's largest earthquakes occur along subduction zones, (3) the dominant mechanism of subduction zone earthquakes is thrusting, resulting in significant vertical displacement of the seafloor. Normal-faulting earthquakes in the outer rise and outer trench slope of subduction zones and thrust faulting in the back arc can also generate tsunamis. Other tectonic environments where seismogenic tsunamis can occur are oceanic convergence boundaries as distinguished from subduction zones by Bird (2003) and local, submarine dip-slip fault zones in intraplate settings.

To simulate the tsunami generation process, elastic deformation models are needed to determine the magnitude and pattern of seafloor displacement from earthquake source parameters. To calculate the initial tsunami amplitude distribution on the ocean surface, a $1/\cosh(kh)$ low-pass filter is applied to the seafloor displacement field, where k is the wave number and h is the water depth (Kajiura 1963). This filter primarily has an effect for small earthquakes and surface-rupturing earthquakes that generate a scarp or short-wavelength displacement. Otherwise, the initial wave field of a tsunami is almost identical to the vertical seafloor displacement field. Horizontal displacements also contribute to tsunami generation in regions of steep bathymetry (Tanioka and Satake 1996). Although the rupture process of large subduction zone earthquakes is extremely complex, simple mechanical models have traditionally been used to simulate seafloor displacement and tsunami generation up until recently. These models assume that the earthquake occurs instantaneously relative to the phase speed of the tsunami, often referred to as static or coseismic models.

A significant advancement in simulating tsunami generation is the use of dynamic rupture models. These models simulate the rupture process

Earthquake Mechanism and Seafloor Deformation for Tsunami Generation,

Fig. 1 Perspective view (NW) of vertical displacement field associated with the 2004 M_w 9.2 Sumatra-Andaman earthquake, showing uplift (red) near the trench and subsidence (blue) toward the Sumatran shore



E

along the fault itself, as well as dynamic deformation of the surrounding rocks, including dynamic stress transfer from seismic waves as well as the final static seafloor displacement. The time-varying displacement of the seafloor is extracted from these models and is used to force the tsunami in the source region. The time-varying aspect of the seafloor motion is not likely to be a large factor in tsunami generation, but as will be shown below, the final earthquake size, slip distribution, and rupture path do depend critically on the spatiotemporal details of the rupture process.

First, the most common class of static mechanical models based on dislocation theory is discussed. Static crack models are then described that have several advantages over traditional dislocation models. General methods for incorporating heterogeneous slip, particularly from stochastic slip models, are next discussed. Finally, recent advances in using dynamic rupture models for tsunami generation are reviewed. Because details among the different tsunami generation models are smoothed out in the far field, the focus in this entry is on near-field effects on the tsunami wave field.

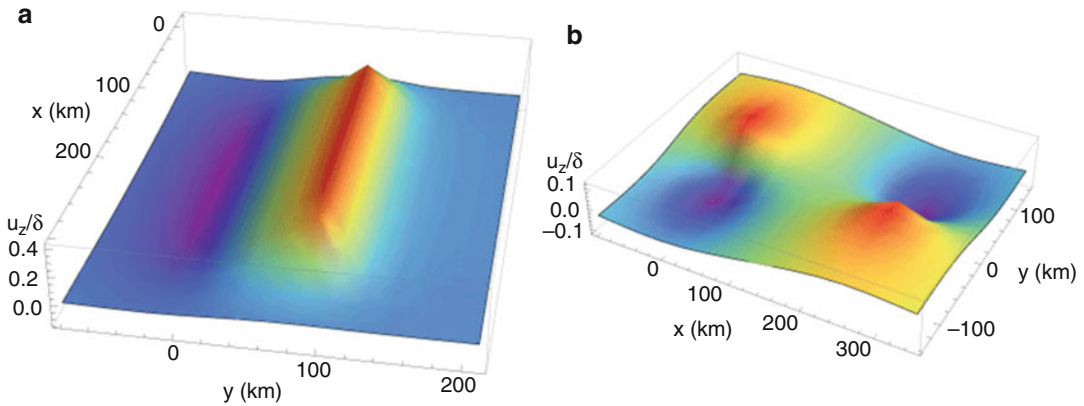
Dislocation Models of Tsunami Generation

Figure 1 shows an example of the coseismic vertical displacement field for a thrust fault: the Sunda thrust during the 2004 M_w 9.2

Sumatra-Andaman earthquake (Geist et al. 2007). In the near field along subduction margins, the field is characterized by negative amplitudes toward shore (i.e., toward the island of Sumatra in Fig. 1) and positive amplitudes toward the open ocean. The zero displacement line is often termed the hinge line. In most cases, the hinge line is offshore, such that coastal subsidence occurs with large-magnitude subduction zone earthquakes, and the first motion of the tsunami broadside from the source is negative (trough phase). Conversely, at locations in the open ocean, the first motion is positive (peak phase). In some cases, the hinge line is onshore, such that there is a coastal uplift and no initial withdrawal of the ocean (e.g., along the Andaman Islands in Fig. 1). In other cases where the rupture area is near the trench, there is no subsidence or uplift of the coastal region (Geist 1999).

Dislocation models were first used to model tsunami generation in the 1970s (e.g., Abe 1973; Fukao and Furumoto 1975; Kanamori 1972). It is a curious historical circumstance that crack models were already well established to address various geophysical problems by this time (e.g., Eshelby 1957), but were never widely used to model tsunami generation. The main advantage of using uniform-slip dislocation models, termed Volterra dislocations (Bilby and Eshelby 1969), is in reference to the mechanical definition of seismic moment (m_0)

$$m_0 = \bar{\mu} \bar{d}_s \Sigma,$$



Earthquake Mechanism and Seafloor Deformation for Tsunami Generation, Fig. 2 Comparison of vertical displacement field from (a) thrust fault (dip=18°) and (b) strike-slip fault (dip=90°). In both cases, depth to

up-dip edge of dislocation is 19 km. Maximum normalized vertical displacements (u_z/δ) are 0.43 and 0.12 for (a) and (b), respectively

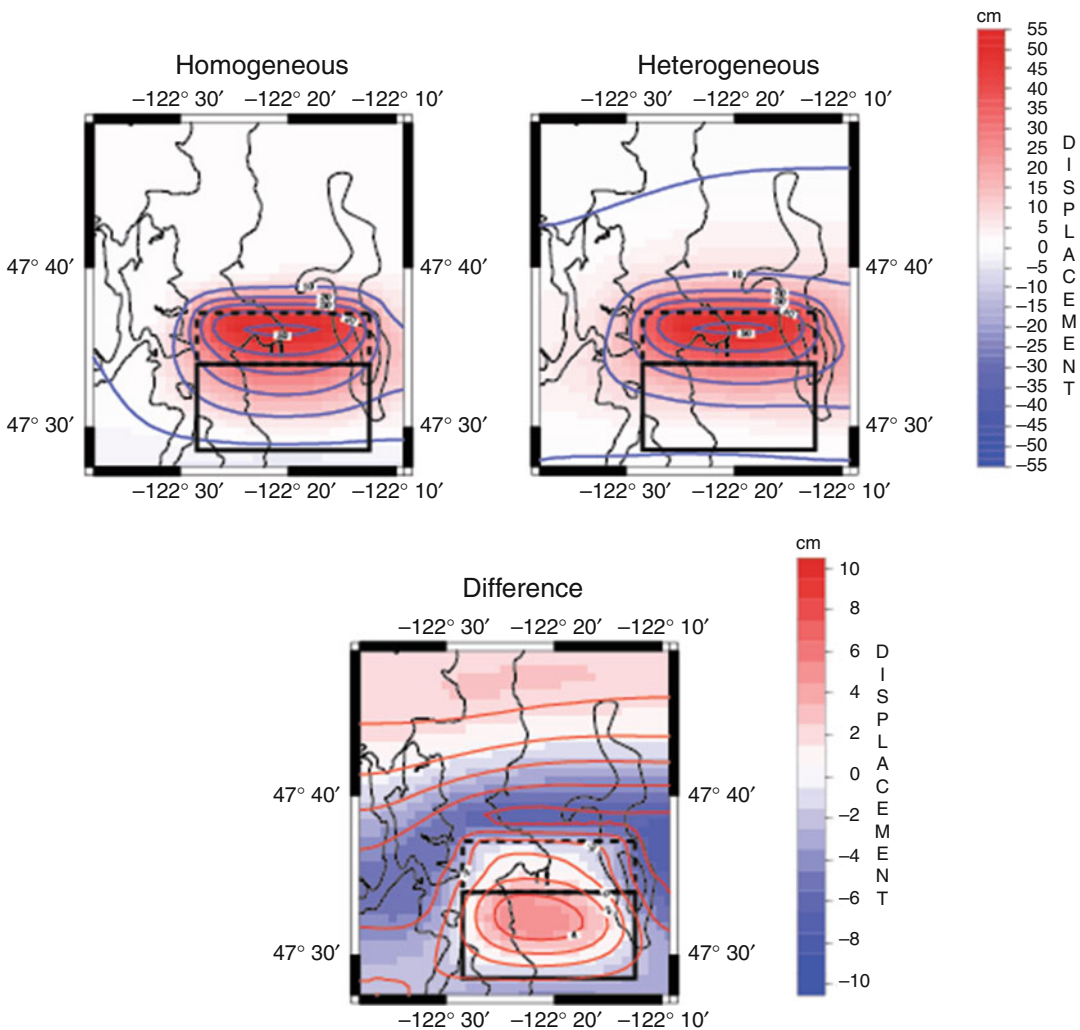
where $\bar{\mu}$ is the average shear modulus of rocks surrounding the rupture and $\bar{\delta}_s$ is the average static slip over the area Σ of the fault that ruptured during the earthquake. However, Madariaga (1979) shows that m_0 can also be related to the static stress drop for a crack. The area Σ of a rupture can be estimated from the aftershock distribution or scaled from earthquake magnitude (e.g., Blaser et al. 2010; Geller 1976). Although mean slip is indicated in the definition of m_0 , this is often taken as the value for uniform slip in Volterra dislocation models (i.e., $\delta_{\text{uniform}} = m_0/(\bar{\mu} \Sigma)$).

Analytical expressions have been developed to calculate the coseismic displacement field $\mathbf{u}$ from dislocations with slip vector $\mathbf{s} \equiv \mathbf{u}_i^+ - \mathbf{u}_i^-$, where the plus and minus sign indicate opposite sides of the dislocation. Okada (1985) reviews the expressions derived through 1983 and derives compact expressions applicable to all fault types and non-Poisson solids. Most importantly for tsunami generation, both finite-fault and point-source representations are given for dip-slip faults, the latter useful for computing displacement from heterogeneous slip as described below. The finite-fault expressions can be compared to those developed earlier by Savage and Hastie (1966) for dip-slip faults. Most of the analytic expressions used for tsunami generation are applied to a homogeneous elastic half space, although expressions have also been

developed for a layered media (e.g., Ma and Kuszniir 1994) and for spherical geometry (e.g., Pollitz 1996).

The effect that individual earthquake source parameters have on seafloor displacement varies considerably. The amount of uniform slip assigned to a Volterra dislocation has the largest effect on the magnitude of seafloor displacement, with an approximately linear relationship between slip and maximum vertical displacement (Geist 1999). The dip of the fault primarily affects the polarity of the tsunami wave. With increasing dip of a thrust fault, the amplitude of the negative polarity phase decreases over the hanging wall and increases over the footwall. Geist (1999) gives a complete description of the effect that these and other earthquake source parameters have on seafloor displacement and the associated near-field tsunami.

The effectiveness of strike-slip faults in generating tsunamis relative to dip-slip faults has recently been debated. Although the slip vector is parallel to the surface, there is vertical displacement associated with strike-slip faults, occurring in a quadrupole pattern (see also Fig. 2b). Using a dislocation model, the vertical displacement pattern for dip-slip and strike-slip faults is compared in Fig. 2 for the same amount of slip and depth to the up-dip edge of the dislocation. For this case, the maximum vertical



Earthquake Mechanism and Seafloor Deformation for Tsunami Generation, Fig. 3 Effect of structural heterogeneity of elastic constants on vertical displacement field in the Puget Sound region. Solid rectangle: down-dip

rupture of Seattle fault through high shear modulus rocks. Dashed rectangle: up-dip rupture through low shear modulus sedimentary rocks of the Seattle sedimentary basin

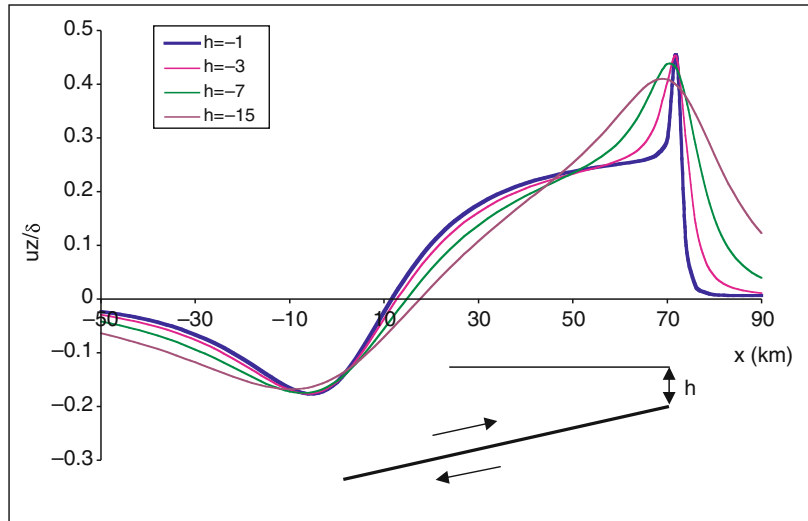
displacement associated with the strike-slip dislocation is approximately a quarter of that for the thrust dislocation. A similar comparison is made using normal-mode theory by Ward (1980) and Okal (1988), which is useful to gauge the far-field response to the different mechanisms.

Rather than using an elastic half-space or layered-earth model, realistic geologic structures and heterogeneous physical properties can be used to calculate seafloor displacement, most often by finite element methods (Geist and Yoshioka 1996;

Masterlark 2003; Zhao et al. 2004). Static elastic deformation models developed for the Nankai subduction zone were adapted to calculate tsunami generation along the Cascadia subduction zone (Geist and Yoshioka 1996). The development of thrust faulting in the presence of subducted seamounts, also at the Nankai subduction zone, is examined using a finite element model by Baba et al. (2001). The effect of sedimentary basins on static elastic displacements is shown for the Seattle basin in Fig. 3 (Geist and Yoshioka 2004). The effect of rupturing into lower-than-average shear

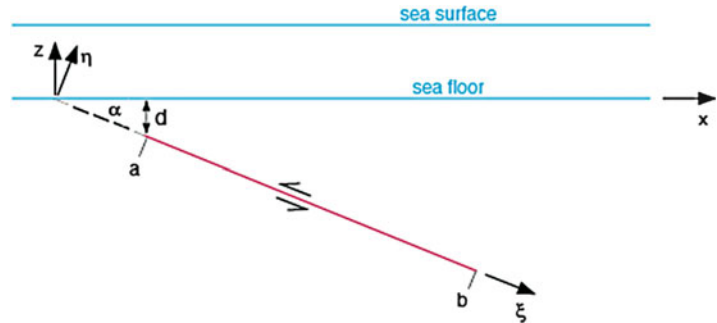
Earthquake Mechanism and Seafloor Deformation for Tsunami Generation,

Fig. 4 Effect of decreasing depth to up-dip edge of rupture (h in kilometers) on vertical displacement (u_z) profile (normalized with respect to slip, δ) for a shallow thrust Volterra dislocation



Earthquake Mechanism and Seafloor Deformation for Tsunami Generation,

Fig. 5 Coordinate system and parameters for crack model



modulus is to increase vertical displacement at the surface (blue region, bottom of Fig. 3), consistent with the results of Ma and Kusznir (1994) and Savage (1998). Finite element models are also critical for calculating dynamic displacements as will be described in the Dynamic Rupture Models section.

Finally, it is important to recognize a prominent artifact associated with dislocation models of shallow imbedded thrust faults, as shown in Fig. 4. Although the expressions for surface displacement developed by Okada (1985) do not include the dislocation edge singularity for faults that rupture to the surface, the short-wavelength effect of the singularity is evident for cases where the up-dip edge is near the surface, relative to the down-dip width of the rupture. In a later study, Okada (1992) discusses

these and other internal singularities further. Application of the $1/\cosh(kh)$ low-pass filter will attenuate this artifact but will lead to a general underestimation of tsunami amplitude. This artifact can be avoided by using crack models for tsunami generation as described in the next section.

Crack Models of Tsunami Generation

Unlike dislocation models for tsunami generation in which slip is prescribed, the slip distribution associated with crack models is calculated and physically consistent with the applied tectonic stresses and frictional conditions on the fault. For dip-slip faults of geometry shown in Fig. 5, the slip gradient in the dip direction $\delta'(\xi)$ is given

by the following singular integral equation (Dmowska and Kostrov 1973):

$$\frac{1}{\pi} \int_a^b \frac{\delta'(s)}{s - \xi} ds + \frac{1}{\pi} \int_a^b K(\xi, s) \delta'(s) ds = \frac{2(1 - \nu)}{\mu} f(\xi),$$

where λ and μ are Lamé's constants, ν is Poisson's ratio, and $f(\xi)$ represents the initial shear stress distribution. In terms of principal stresses, $f(\xi)$ is given by

$$f(\xi) \equiv \sigma_{\xi\eta}^0 = \frac{1}{2}(\sigma_{xx}^0 - \sigma_{zz}^0) \sin 2\alpha + \sigma_{xz}^0 \cos 2\alpha$$

with the zero superscript indicating initial stress.

The effect of decreased normal stress on a shallow imbedded crack makes it more likely to rupture to the surface above a critical depth as shown by Rudnicki and Wu (1995). The Coulomb friction condition is introduced such that slip occurs when $|\sigma_{\xi\eta}| = k\sigma_{\eta\eta}$, where k is the apparent coefficient of friction, including effects of pore-pressure differences in the fault zone. The right-hand side of integral equation above is modified such that

$$f_k(\xi) = f(\xi) - k\sigma_{\eta\eta}^0$$

For shallow imbedded faults, any changes in shear stress at depth will result in normal stress changes up dip, such that the slip distribution becomes skewed up dip until a critical depth when slip is unstable and ruptures through to the surface (Dmowska and Kostrov 1973; Rudnicki and Wu 1995). However, rupture propagation to the surface may be inhibited by inelastic deformation (Roering et al. 1997) and effects of fluids in the fault zone (Chambon and Rudnicki 2001). The added effect of dynamic stress transfer from seismic waves interacting with the fault zone is shown using dynamic rupture models (Oglesby and Archuleta 2000; Oglesby et al. 2000) as detailed below.

Although the singularity in the above equation is integrable, Freund and Barnett (1976) developed a slip distribution associated with the smooth closure condition crack model (Bilby and Eshelby 1969) that has been used for

subduction zone earthquakes and tsunami generation (e.g., Geist and Dmowska 1999; Wang and He 2008). The slip gradient function prescribed by Freund and Barnett (1976) is given by

$$\delta' = \Delta \begin{cases} \frac{12}{q^3} (q\gamma - \gamma^2), & \gamma < q \\ \frac{12}{(1-q)^3} (\gamma^2 - \gamma(1+q) + q), & \gamma > q \end{cases},$$

where Δ is a scaling coefficient, $\gamma \equiv (\xi - a)/(b - a)$, and q ($0 < q < 1$) controls the skewness of the slip distribution. This function satisfies the smooth closure condition $\delta'(a) = \delta'(b) = 0$.

For the specific one-dimensional version of the fault geometry shown in Fig. 5, the displacement profile associated with cracks can be calculated analytically. As indicated by Freund and Barnett (1976), the displacement is calculated by integration based on the slip gradient

$$u_z = \int_a^b U_z(x, s) \delta'_\xi(s) ds,$$

where

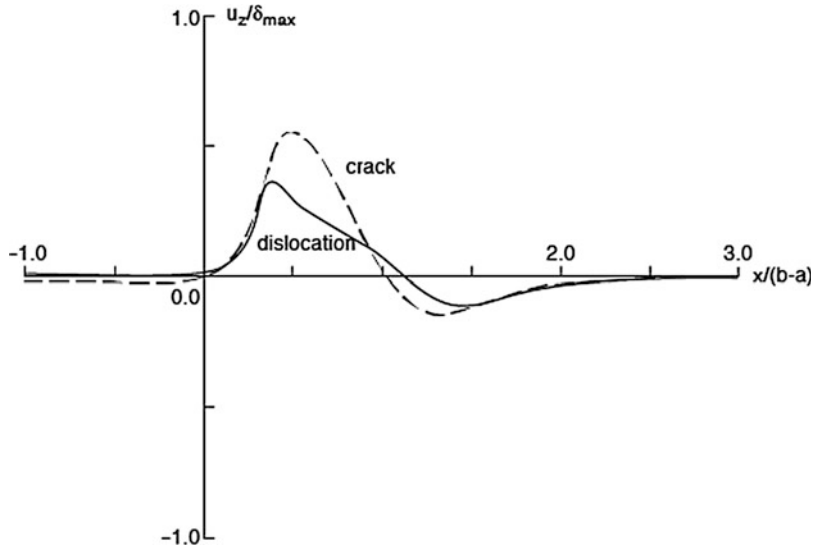
$$U_z(x, s) = \frac{xs \sin^2 \alpha}{\pi(x^2 + s^2 - 2xs \cos \alpha)} + \frac{\sin \alpha}{\pi} \tan^{-1} \left(\frac{x - s \cos \alpha}{s \sin \alpha} \right).$$

Calculation of more general surface displacement fields in two dimensions is discussed in the next section.

Using these formulas, the slip gradient can be related to the aspects of the displacement profile. For example, Fig. 6 compares the vertical displacement profile from a shallow imbedded crack to an equivalent dislocation. Crack models avoid the strong fault-tip singularities associated with dislocation models and the associated effects on the seafloor displacement field. The strain singularity at the edge of the dislocation manifests as a short-wavelength spike in the displacement as shown in Fig. 4. Conversely, the concentration of slip in the middle of a crack

Earthquake Mechanism and Seafloor Deformation for Tsunami Generation,

Fig. 6 Comparison of displacement profiles between imbedded crack (dashed line) and dislocation (solid line) models with same average slip (uniform slip for the dislocation model). See Fig. 5 for fault geometry



results in more seafloor displacements over the center of the rupture when compared to the displacement field from the dislocation.

One can also compare a surface-rupturing crack with an equivalent very shallow imbedded crack to show the effect of surface rupture (Fig. 7). The dip-directed slip profile radically changes from a skewed half ellipse to a quarter ellipse with twice the maximum slip when the surface is ruptured (Fig. 7a). The stress drop for the surface-rupturing case is lower under the tectonic stress conditions given above, and the vertical displacement is similar for the two cases. The reason for the latter is that the up-dip slip gradient is higher for the imbedded case than for the surface-rupturing case. The surface-rupturing case results in a fault scarp at the seafloor that is filtered through the water column as mentioned in the Introduction. As the depth to the top of the imbedded crack increases, the peak vertical displacement decreases.

Heterogeneous Static Slip

The surface displacement field associated with cracks and other heterogeneous slip models can generally be computed using the static elastic Green's functions. Consider three components of displacement at a point on the seafloor given

by the location vector $\mathbf{r}(x, y)$ in response to slip on a fault at location $\mathbf{r}_0(\xi, y)$ (Fig. 8). The three components of seafloor displacement in two dimensions can then be given by (Rybicki 1986)

$$u_m(\mathbf{r}) = \iint_{\Sigma} D_i(\mathbf{r}_0) v_j(\mathbf{r}_0) U_m^{ij}(\mathbf{r}, \mathbf{r}_0) d\Sigma \quad m = 1, 2, 3,$$

where $v(\mathbf{r}_0)$ is the surface normal of a fault plane with arbitrary geometry and Σ is the area of the fault plane. For an isotropic medium, the six Green's functions are given by

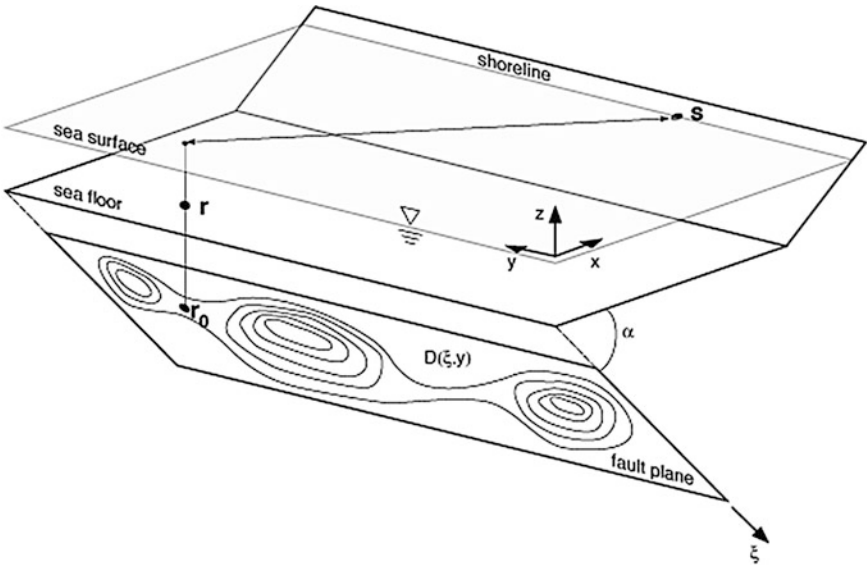
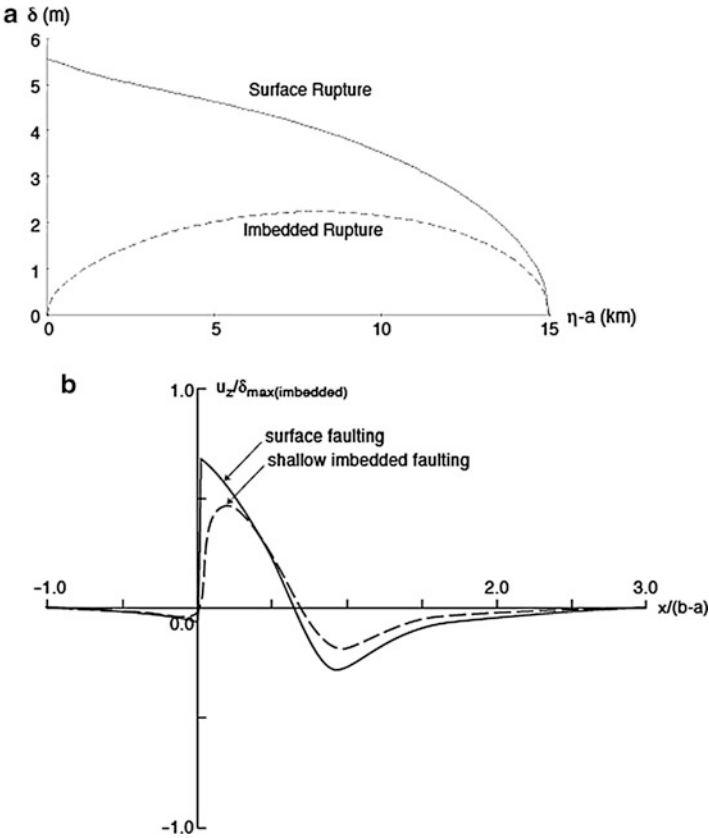
$$U_m^{ij} = U_m^{ji} = \lambda \delta_{ij} G_m^{n,n} + \mu [G_m^{i,j} + G_m^{j,i}],$$

where λ is Lamé's first constant and δ_{ij} is the Kronecker delta. A review of the static elastic Green's functions G_m^i under different assumptions can be found in Segall (2010).

The displacement field can also be approximated by small dislocation fault elements. Because the elastic deformation is linear at large scales, the displacement field from individual small elements can be superimposed to construct the full displacement field from a rupture. For example, Okada (1985) provides point-source analytic expressions for an element of surface area $\Delta\Sigma$. As shown in Fig. 9, this technique provides a good approximation to the analytic

Earthquake Mechanism and Seafloor Deformation for Tsunami Generation,

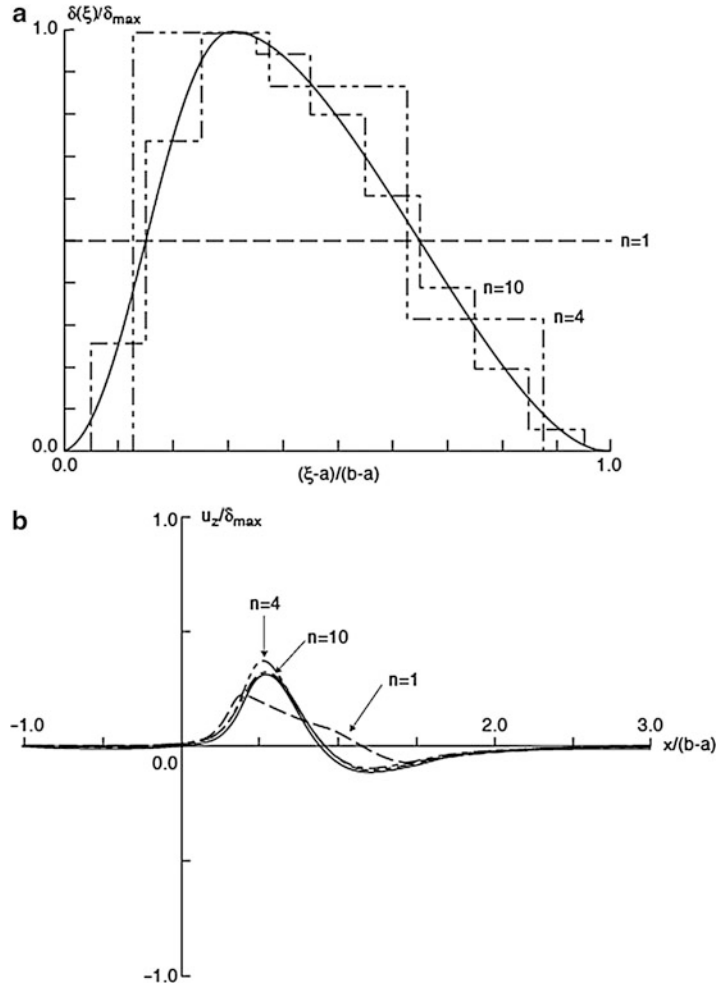
Fig. 7 Comparison of slip profiles (a) and displacement profiles (b) between a surface-rupturing crack (solid line) and a very shallow imbedded crack (dashed line). Static stress drop for the two models is 4.8 and 6.3 MPa, respectively. See Fig. 5 for fault geometry



Earthquake Mechanism and Seafloor Deformation for Tsunami Generation, Fig. 8 Coordinate system and parameters for heterogeneous slip model

Earthquake Mechanism and Seafloor Deformation for Tsunami Generation,

Fig. 9 (a) Smooth closure slip profile approximated by 1 (Volterra dislocation), 4, and 10 elements. (b) Associated vertical displacement profile. See Fig. 5 for coordinate system ($\alpha=20^\circ$)

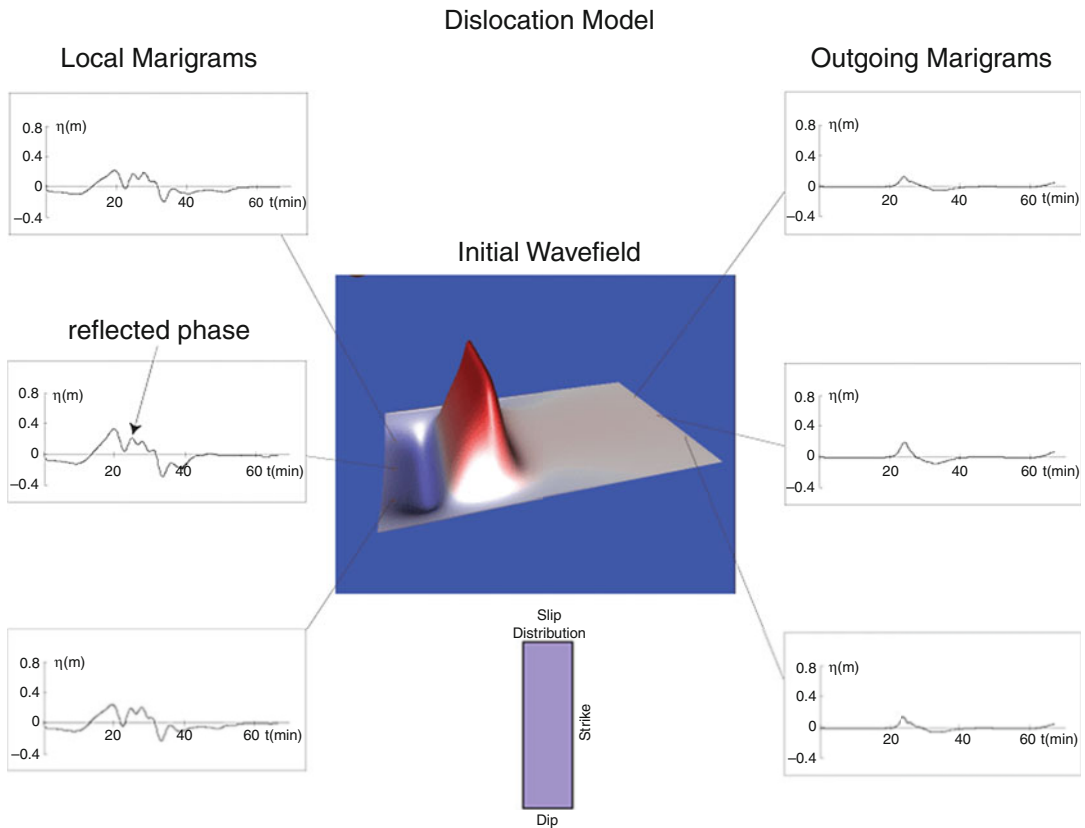


displacement profile calculated for a crack if small enough elements are used. Linear superposition of subfault elements also provides the basis for finite-fault inversion using tsunami waveforms (Satake and Kanamori 1991).

The difference in the local tsunami wave field between crack and dislocation models is shown by example in Figs. 10 and 11. In both cases, a 270 km long and 70 km wide, pure thrust rupture is used, with a dip of 20° and average slip of 1 m. For the crack model, the slip distribution is that derived from Freund and Barnett (1976) under smooth closure conditions ($q=0.3$). Details of the modeling procedure are given in Geist and Dmowska (1999). Marigrams (time series plot of wave amplitude) are displayed for both the

local tsunami approaching shore (left) and the outgoing tsunami (right). For all marigrams, the maximum amplitude is higher for the crack model, even though the scalar seismic moment is identical. The dominant period of the tsunami associated with the crack model is also shorter, leading to a higher wave steepness and larger run-up compared to the tsunami from the dislocation model.

Earthquake rupture is often sufficiently complex that it cannot be adequately described by either uniform slip (in the case of the dislocation model) or uniform static stress drop (in the case of the crack model). For this reason, the slip distribution along a fault is difficult to predict without modeling the rupture process itself



Earthquake Mechanism and Seafloor Deformation for Tsunami Generation, Fig. 10 Tsunami wave field using a dislocation model. Uniform slip within a rectangular rupture zone shown at the bottom middle. Initial

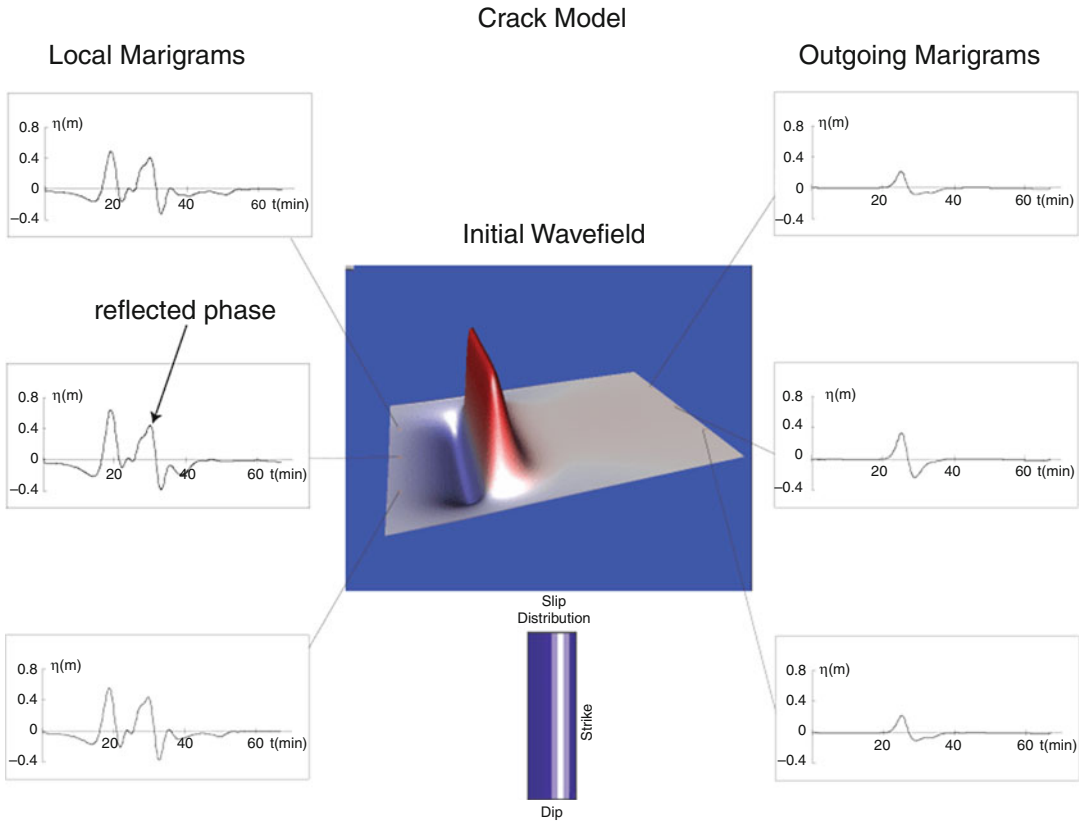
wave field, assuming instantaneous rupture, shown in the center. Marigrams on the left side are time series plots of local tsunami approaching shore. Marigrams on the right side represent propagation toward the ocean basin

(next Section). Several stochastic slip models have been developed to simulate static slip in a probabilistic sense (Andrews 1980; Lavallée et al. 2006).

Application of Dynamic Rupture Models to Tsunami Generation

Dynamic models of the earthquake source push the physical assumptions of dislocation models back a step; rather than assuming an instantaneous distribution of slip on the fault, dynamic spontaneous earthquake rupture models start with a distribution of stress and frictional parameters on the fault, along with material properties throughout the medium. A key assumption in

such models is the constitutive law, which determines how the shear stress on the fault depends on quantities such as the normal stress across the fault, the instantaneous slip or slip rate on the fault, as well as the integrated history of the rupture. Such models use the laws of motion and elastodynamics to calculate the time-dependent propagation of rupture and the evolution of slip at each point on the fault, as well as the time-dependent deformation of the earth's surface. In this way, these models are more similar to the crack models mentioned above, but with the added feature that the rupture path, slip distribution, and final size of the earthquake are calculated as a result of the interaction of dynamic stress waves with the frictional and geometrical characteristics of the fault.



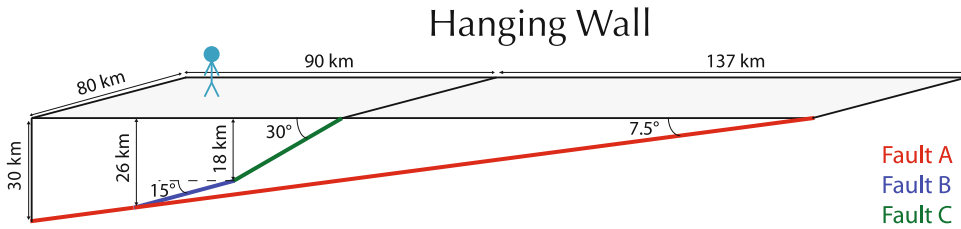
Earthquake Mechanism and Seafloor Deformation for Tsunami Generation, Fig. 11 Tsunami wave field using a crack model. Heterogeneous slip within a rectangular rupture zone shown at bottom middle (lighter colors represent higher slip). Initial wave field, assuming

instantaneous rupture, shown in the center. Marigrams on the left side are time series plots of local tsunami approaching shore. Marigrams on the right side represent propagation toward the ocean basin

While dislocation models may assume slip distributions that are not consistent with any physically plausible stress distribution, such consistency is at the heart of dynamic models. Dynamic models have been used extensively in earthquake modeling for both basic physics and ground motion purposes (e.g., Andrews 1976; Das 1981; Day 1982; Harris et al. 1991; Oglesby et al. 1998). A common feature of such models is that if the frictional stress on the fault drops rapidly enough as a function of slip, slip rate, or time, unstable motion is a result. All other factors being equal, high stress drops lead to fast rupture propagation and rapid fault slip, along with high final slip. Zones in which the frictional stress remains constant or increases over the course of the earthquake (such as in the very deep and shallow

portions of a fault) tend to cause rupture and slip to slow down and die out (Okubo 1989).

Spontaneous dynamic rupture models of megathrust earthquakes can shed considerable light on both the physics of the earthquake process and the generation of tsunamis. For example, in such earthquakes, reflected waves from the free surface can reduce the normal stress on the fault during slip, facilitating rupture (Brune 1996; Oglesby et al. 1998). Kozdon and Dunham (2013) show that in the case of the 2011 Tohoku Earthquake, this effect can explain the rupture and high slip all the way to the trench, even in the near-surface zone in which the frictional coefficient is expected to increase with slip rate. Somewhat similarly, Noda and Lapusta (2013) show that ruptures can tunnel through stably



Earthquake Mechanism and Seafloor Deformation for Tsunami Generation, Fig. 12 Fault geometry of Wendt et al. (2009). The splay fault (Faults B and C)

branches off the plate boundary thrust (Fault A) close to shore and at a significantly steeper dip than the plate boundary thrust

E

sliding regions of a fault (i.e., zones with low seismic coupling) if pore-fluid pressurization takes place. The possibility of earthquake rupture and slip propagating into areas that appear to be stably sliding can have important implications for both the distribution of slip in an earthquake and its overall size, with obvious consequences for tsunami generation. Recent dynamic models by Ma and Hirakawa (2013) have argued that failure of material in the hanging-wall wedge might lead to slow rupture propagation and slow slip near the trench, leading to low-radiation, high-tsunami “tsunami earthquakes.”

Fault dynamics can also play an important role in determining rupture path and partitioning of slip between different fault segments in a geometrically complex fault system. An issue of great importance for tsunami generation is that of splay faults – whether a splay fault is activated coseismically. In a standard dislocation or crack model, the decision of whether a splay fault will slip is made a priori and is unconnected with the slip process elsewhere on the fault system. However, Wendt et al. (2009) show that the ability of rupture to propagate to a splay fault is a result of the dynamic rupture process and can have a first-order effect on tsunami generation and near-field run-up. Their fault geometry, motivated by the Nankai Trough, is shown in Fig. 12.

The researchers used the dynamic finite element method (FEM) (Oglesby 1999; Whirley and Engelmann 1993) to perform spontaneous dynamic rupture models of potential earthquakes on this fault system. They found that if stress on the faults is largely homogeneous, rupture proceeds only on the plate boundary thrust. However, if a barrier is present on the plate boundary thrust, rupture

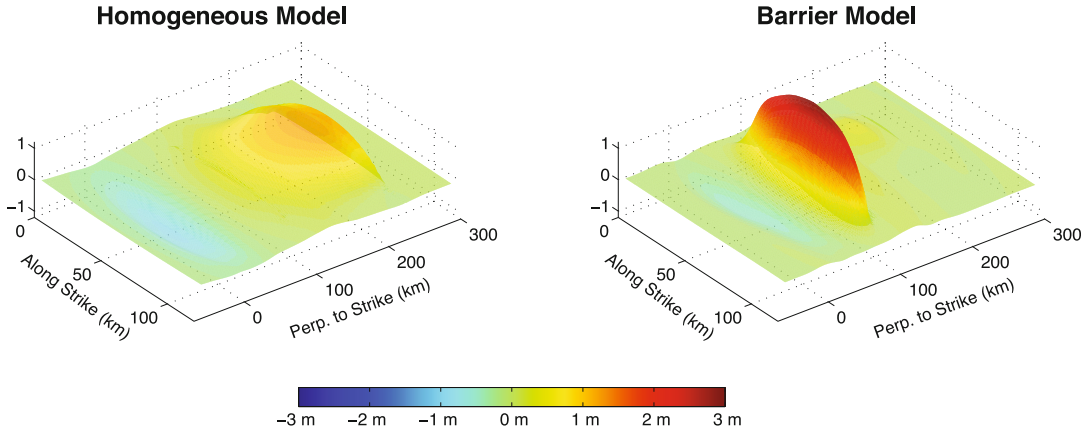
transfers to the steeper splay fault. A comparison between the vertical surface displacement fields in these two cases is shown in Fig. 13.

Because of the more vertically directed hanging-wall movement, the splay faulting case produces a significantly larger vertical seafloor displacement than the homogeneous case. For this reason, the local tsunami in the barrier case produces a significantly larger wave in the near-source region (in this case, along the shore of Japan) than the homogeneous case, as seen in Fig. 14, and it also reaches the near shore much sooner.

It should be noted that in this simple model, the moment of the barrier model (6.8×10^{20} Nm) is less than half that of the homogeneous case (1.8×10^{21} Nm), so the difference in tsunami height is purely due to the rupture path and slip distribution, not to the overall size of the earthquake. Another advantage of the dynamic modeling method in this case is that the partitioning of slip between the plate boundary thrust and the splay fault is a calculated result of the problem rather than an assumption; the method correctly accounts for the fact that slip on the splay suppresses slip on the plate boundary thrust, and vice versa.

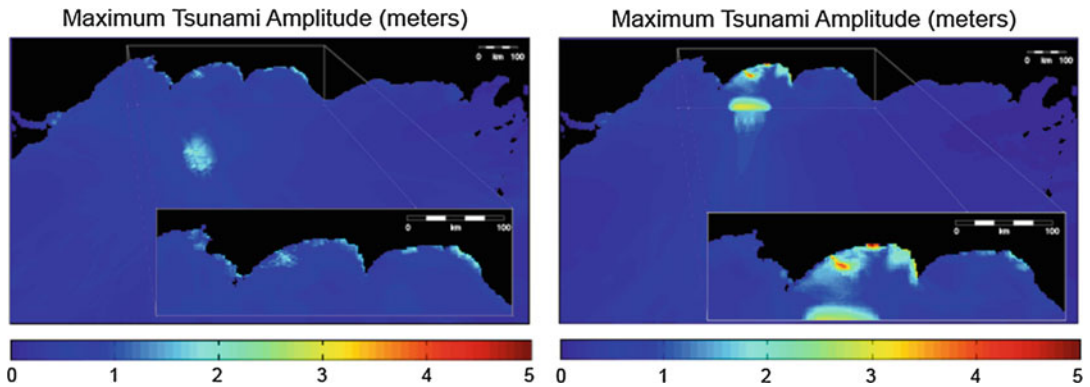
The examples given above raise a tantalizing possibility: if enough is known about the geology and fault geometry of a subduction zone, dynamic modeling has the promise to help predict (or at least constrain) the size and severity of potential earthquakes and tsunamis in the region. Someday it may be possible to constrain the likelihood of rupture to the trench or whether rupture is likely to propagate to a splay fault rather than (or in addition to) the main plate boundary thrust

Vertical Seafloor Displacement



Earthquake Mechanism and Seafloor Deformation for Tsunami Generation, Fig. 13 Comparison of vertical seafloor displacement for homogeneous-stress case (*left*) and barrier case (*right*). Note that the barrier case

produces much higher vertical displacement than the homogeneous case and that this displacement is closer to shore



Earthquake Mechanism and Seafloor Deformation for Tsunami Generation, Fig. 14 Comparison of wave amplitude (m) for the homogeneous case (*left*) and

the barrier case (*right*). The rupture of the splay produces a higher initial tsunami as well as higher wave amplitude as the tsunami approaches the shore (*insets*)

fault. In this manner, dynamic models may help to compensate for lack of historical data, especially in the case of extremely large ($M\ 9+$) events that happen only once or twice a millennium. Dynamic models are not a panacea, however Like dislocation and crack models, they rest on assumptions about physical properties (such as stress and friction) that are quite poorly constrained in many cases. However, they do have the advantage of producing physically self-consistent slip and rupture patterns that take into

account important interactions and processes on and near faults. As such, they can serve an important complement to existing standard methods of tsunami source estimation.

Summary

Tsunamis are generated from earthquakes by vertical displacement of the seafloor during earthquake rupture. Four representations of the

earthquake rupture process have been described, in increasing levels of consistency with earthquake physics, for use as tsunami generation models. First, simple uniform-slip dislocation modeling, scaled to the seismic moment of an earthquake, has historically been the most common representation used for tsunami modeling. Second, crack models avoid artifacts that dislocation models have associated with shallow, imbedded thrust faults. Crack models also directly relate slip distribution to tectonic stress conditions and static stress drop of an earthquake. Third, heterogeneous static slip models account for the observed complexity of distribution slip (and stress drop) throughout a rupture zone, primarily through the use of small dislocation elements. Finally, dynamic rupture models provide the most complete and physically consistent representation of the tsunami generation process. Each of these models has been used in both deterministic and probabilistic hazard assessments, although advanced tsunami generation models provide more accuracy and consistency with the known processes of earthquake physics.

Cross-References

- [Earthquake Mechanisms and Stress Field](#)
- [Earthquake Mechanisms and Tectonics](#)
- [Earthquake Mechanism Description and Inversion](#)
- [Probabilistic Seismic Hazard Models](#)
- [Site Response for Seismic Hazard Assessment](#)
- [Tsunamis as Paleoseismic Indicators](#)

References

- Abe K (1973) Tsunami and mechanism of great earthquakes. *Phys Earth Planet In* 7:143–153
- Andrews DJ (1976) Rupture velocity of plane strain shear cracks. *J Geophys Res* 81:5679–5687
- Andrews DJ (1980) A stochastic fault model I. Static case. *J Geophys Res* 85:3867–3877
- Baba T, Hori T, Hirano S, Cummins PR, Park J-O, Kameyama M, Kaneda Y (2001) Deformation of a seamount subducting beneath an accretionary prism: constraints from numerical simulation. *Geophys Res Lett* 28:1827–1830
- Bilby BA, Eshelby JD (1969) Dislocations and the theory of fracture. In: Liebowitz H (ed) *Fracture*. Academic, New York, pp 99–182
- Bird P (2003) An updated digital model of plate boundaries. *Geochem Geophys Geosyst* 4. doi:10.1029/2001/GC000252
- Blaser L, Krüger F, Ohmberger M, Scherbaum F (2010) Scaling relations of earthquake source parameter estimates with special focus on subduction environment. *Bull Seismol Soc Am* 100:2914–2926
- Brune JN (1996) Particle motions in a physical model of shallow angle thrust faulting. *Proc Indian Acad Sci* 105:L197–L206
- Chambon G, Rudnicki JW (2001) Effects of normal stress variations on frictional stability of a fluid-infiltrated fault. *J Geophys Res* 106:11353–11372
- Das S (1981) Three-dimensional spontaneous rupture propagation and implications for the earthquake source mechanism. *Geophys J Roy Astron Soc* 67:375–393
- Day SM (1982) Three-dimensional simulation of spontaneous rupture: the effect of nonuniform prestress. *Bull Seismol Soc Am* 72:1881–1902
- Dmowska R, Kostrov BV (1973) A shearing crack in a semi-space under plane strain conditions. *Arch Mech* 25:421–440
- Eshelby JD (1957) The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proc R Soc Lond A Math Phys Sci* 241:376–396
- Freund LB, Barnett DM (1976) A two-dimensional analysis of surface deformation due to dip-slip faulting. *Bull Seismol Soc Am* 66:667–675
- Fukao Y, Furumoto M (1975) Mechanism of large earthquakes along the eastern margin of the Japan Sea. *Tectonophysics* 25:247–266
- Geist EL (1999) Local tsunamis and earthquake source parameters. *Adv Geophys* 39:117–209
- Geist EL, Bilek SL (2001) Effect of depth-dependent shear modulus on tsunami generation along subduction zones. *Geophys Res Lett* 28(7):1315–1318
- Geist EL, Dmowska R (1999) Local tsunamis and distributed slip at the source. *Pure Appl Geophys* 154:485–512
- Geist EL, Yoshioka S (1996) Source parameters controlling the generation and propagation of potential local tsunamis along the Cascadia margin. *Nat Hazards* 13:151–177
- Geist EL, Yoshioka S (2004) Effect of structural heterogeneity and slip distribution on coseismic vertical displacement from rupture on the Seattle fault. Open-File Report 2004–1010 Rep 2004–1010, U.S. Geological Survey
- Geist EL, Titov VV, Arcas D, Pollitz FF, Bilek SL (2007) Implications of the December 26, 2004 Sumatra-Andaman earthquake on tsunami forecast and assessment models for great subduction zone earthquakes. *Bull Seismol Soc Am* 97:S249–S270
- Geller RJ (1976) Scaling relations for earthquake source parameters and magnitudes. *Bull Seismol Soc Am* 66:1501–1523

- Harris RA, Archuleta RJ, Day SM (1991) Fault steps and the dynamic rupture process – 2-D numerical simulations of a spontaneously propagating shear fracture. *Geophys Res Lett* 18(5):893–896
- Kajiura K (1963) The leading wave of a tsunami. *Bull Earthq Res Inst* 41:535–571
- Kanamori H (1972) Mechanism of tsunami earthquakes. *Phys Earth Planet In* 6:346–359
- Kozdon JE, Dunham EM (2013) Rupture to the trench: dynamic rupture simulations of the 11 march 2011 Tohoku Earthquake. *Bull Seismol Soc Am* 103(2B): 1275–1289
- Lavallée D, Liu P, Archuleta RJ (2006) Stochastic model of heterogeneity in earthquake slip spatial distributions. *Geophys J Int* 165:622–640
- Ma S, Hirakawa ET (2013) Dynamic wedge failure reveals anomalous energy radiation of shallow subduction earthquakes. *Earth Planet Sci Lett* 375:113–122, 110.1016/j.epsl.2013.1005.1016
- Ma XQ, Kusznir NJ (1994) Effects of rigidity layering, gravity and stress relaxation on 2-D subsurface fault displacement fields. *Geophys J Int* 118:201–220
- Madariaga R (1979) On the relation between seismic moment and stress drop in the presence of stress and strength heterogeneity. *J Geophys Res* 84:2243–2250
- Masterlark T (2003) Finite element model predictions of static deformation from dislocation sources in a subduction zone: sensitivities to homogeneous, isotropic, Poisson-solid, and half-space assumptions. *J Geophys Res* 103. doi:10.1029/2002JB002296
- Noda H, Lapusta N (2013) Stable creeping fault segments can become destructive as a result of dynamic weakening. *Nature*. doi:10.1038/nature11703
- Oglesby DD (1999) Earthquake dynamics on dip-slip faults. PhD thesis, University of California, Santa Barbara, 223 pp
- Oglesby DD, Archuleta RJ (2000) Dynamics of dip-slip faulting: explorations in two dimensions. *J Geophys Res* 105(13):643–613653
- Oglesby DD, Archuleta RJ, Nielsen SB (1998) Earthquakes on dipping faults: the effects of broken symmetry. *Science* 280:1055–1059
- Oglesby DD, Archuleta RJ, Nielsen SB (2000) The three-dimensional dynamics of dipping faults. *Bull Seismol Soc Am* 90:616–628
- Okada Y (1985) Surface deformation due to shear and tensile faults in a half-space. *Bull Seismol Soc Am* 75:1135–1154
- Okada Y (1992) Internal deformation due to shear and tensile faults in half-space. *Bull Seismol Soc Am* 82:1018–1040
- Okal EA (1988) Seismic parameters controlling far-field tsunami amplitudes: a review. *Nat Hazards* 1:67–96
- Okubo PG (1989) Dynamic rupture modeling with laboratory-derived constitutive relations. *J Geophys Res* 94:12321–12335
- Pollitz FF (1996) Coseismic deformation from earthquake faulting on a layered spherical earth. *Geophys J Int* 125:1–14
- Roering JJ, Cooke ML, Pollard DD (1997) Why blind thrust faults do not propagate to the Earth's surface: numerical modeling of coseismic deformation associated with thrust-related anticlines. *J Geophys Res* 102:11901–911912
- Rudnicki JW, Wu M (1995) Mechanics of dip-slip faulting in an elastic half-space. *J Geophys Res* 100: 22173–22186
- Rybicki R (1986) Dislocations and their geophysical application. In: Teisseyre R (ed) *Continuum theories in solid earth physics*. Elsevier, Amsterdam, pp 18–186
- Satake K, Kanamori H (1991) Use of tsunami waveforms for earthquake source study. *Nat Hazards* 4:193–208
- Savage JC (1998) Displacement field for an edge dislocation in a layered half-space. *J Geophys Res* 103:2439–2446
- Savage JC, Hastie LM (1966) Surface deformation associated with dip-slip faulting. *J Geophys Res* 20:4897–4904
- Segall P (2010) *Earthquake and volcano deformation*. Princeton University Press, Princeton
- Tanioka Y, Satake K (1996) Tsunami generation by horizontal displacement of ocean bottom. *Geophys Res Lett* 23:861–865
- Wang K, He J (2008) Effects of frictional behavior and geometry of subduction fault on coseismic seafloor deformation. *Bull Seismol Soc Am* 98:571–579
- Ward SN (1980) Relationships of tsunami generation and an earthquake source. *J Phys Earth* 28:441–474
- Wendt J, Oglesby DD, Geist EL (2009) Tsunamis and splay fault dynamics. *Geophys Res Lett* 36, L15303, 15310.11029/12009GL038295
- Whirley RG, Engelmann BE (1993) DYN3D: a nonlinear, explicit, three-dimensional finite element code for solid and structural mechanics – user manual. University of California, Lawrence Livermore National Laboratory, California
- Zhao S, Müller RD, Takahashi Y, Kaneda Y (2004) 3-D finite-element modeling of deformation and stress associated with faulting: effect of inhomogeneous crustal structures. *Geophys J Int* 157:629–644

Earthquake Mechanism Description and Inversion

Doug Dreger

Department of Earth and Planetary Science,
College of Letters and Science, University of
California, Berkeley, CA, USA

Synonyms

Radiation pattern; Source mechanism; Moment tensor; Source inversion

Introduction

Seismic waves can be excited by a variety of sources and processes such as impacts, explosions, volcanic mass ejection, underground cavity collapses, fluid migration, atmosphere/ocean coupling, and of course earthquakes. The nature of seismic excitation and subsequent radiation can be isotropic, the same in all directions as in the case of an explosion, or anisotropic, having distinct directional dependence as in the case for earthquakes. It is commonly found that inversions of seismic waveform data for source parameters for earthquakes, explosions, and induced seismicity deviate from either of these fundamental types and are composed of mixtures that may be real, or due to noise in the data, from sparse station coverage, and imperfect knowledge regarding the Earth model used to describe the propagation of seismic energy from source to station. Thus, recovered source parameters are dependent upon the approximations and assumptions made about the wave propagation side of the problem, and care is needed in interpretation, and thorough error and uncertainty analysis is needed, particularly in studies in which deviations from idealized sources are of interest.

There are many approaches for estimating seismic source parameters ranging from solving nonlinear inversions or grid searches for fault parameters such as strike, rake, and dip to linear inversions for a general seismic moment tensor. Gilbert (1970) first proposed the seismic moment tensor for describing the strength of fundamental forces responsible for seismic excitation from earthquakes, and subsequently there have been a multitude of publications describing various approximations, methods, and applications. The seismic moment tensor can be used to describe the radiation from earthquakes, explosions, and more complicated processes that might involve compound fault rupture and combinations of shear tensile failure that occur in mining, geothermal, volcanic, and inducing environments during operations for energy resource recovery. Julian et al. (1998) and Miller et al. (1998) provide a review of seismic moment tensor theory and applications to non-tectonic seismicity.

Considering earthquake sources, methods have been developed for spatially distributed point sources to address the complexity of large earthquakes (e.g., Kikuchi and Kanamori 1991), spatial point sources with time-dependent moment release allowing for both asynchronous (each moment tensor term has an independent time history) and synchronous moment rate (e.g., Stump and Johnson 1977), and the most common application which involves both a spatial and temporal point source, of which there are numerous examples in the literature. There are methods that utilize complete three-component waveforms, the amplitude spectra of surface waves, those that isolate specific seismic phases, and methods that utilize only specific wave amplitudes or ratios of those amplitudes. There are body wave methods, surface wave methods, and normal mode methods. Moment tensor inversions have been used to learn about the largest megathrust earthquakes and are now commonly applied to study microseismicity, with magnitudes less than zero, in a variety of environments. Methods are applied at teleseismic ($R > 3,000$ km), regional, and local distances. There have been applications related to mining and drilling operations that have considered distances much less than 1 km. Seismic moment tensor methods have been developed to operate autonomously, automatically, and in near real time with either segmented waveform data or on continuous streaming data sets. Derived moment tensor solutions are presently being used to inform on strong shaking hazard shortly after the occurrence of an earthquake, where automated solutions are available minutes following earthquake occurrence in regional distance monitoring situations and within 15–30 min in global/teleseismic monitoring situations. Methods have been implemented in this fashion by international government organizations tasked with seismic monitoring as well as by academia and the private sector. There are many applications and subtleties in approach described in an extensive literature. Jost and Herrmann (1989) provide an excellent theoretical overview and tutorial on the seismic moment tensor method. In this chapter an overview of seismic mechanisms and the

seismic moment tensor method is presented for the case of spatial and temporal point sources.

Construction of Single-Force, Single-Couple, and Double-Couple Radiation Patterns

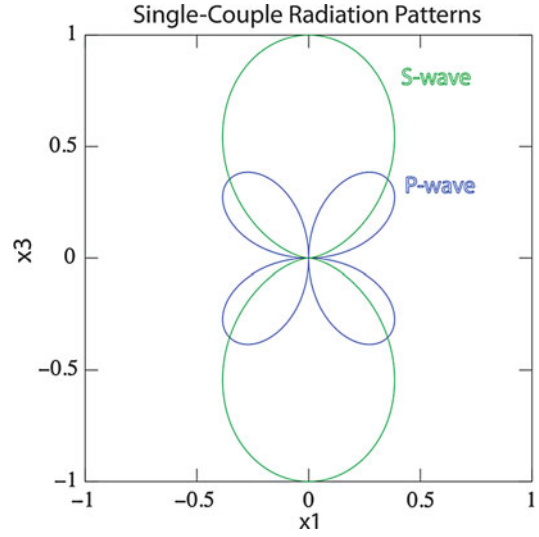
To study the earthquake mechanism, a Green's function solution of the vector wave (Eq. 1) that assumes a homogeneous isotropic elastic whole space is begun with:

$$(\lambda + 2\mu)\nabla(\nabla \cdot \vec{u}) - \mu\nabla \times \nabla \times \vec{u} + \vec{f} = \rho \ddot{\vec{u}} \quad (1)$$

u is the vector displacement for a unit force, f , in an isotropic elastic medium defined by the Lamé coefficients, λ and μ , and the density, ρ . The solution to the forced vector wave equation may be written as follows (Aki and Richards 1980):

$$\begin{aligned} u_n(r, t) = & \frac{3\gamma_n\gamma_p - \delta_{np}}{4\pi\rho r^3} \int_{r/\alpha}^{r/\beta} \tau F(t - \tau) d\tau \\ & + \frac{\gamma_n\gamma_p}{4\pi\rho\alpha^2 r} F(t - r/\alpha) \\ & - \frac{\gamma_n\gamma_p - \delta_{np}}{4\pi\rho\beta^2 r} F(t - r/\beta) \end{aligned} \quad (2)$$

This equation relates the n th component of displacement to the near-field (first term on right), the far-field P-wave, and the far-field S-wave radiation for a unit force in the p -direction. α and β are the P- and S-wave velocities, δ is the Kronecker delta function, and γ are direction cosines relating the direction of ground motion, n , to the direction of the unit force, p . The far-field terms are proportional to the applied force, and the near-field term is proportional to its integral. Single-force solutions may be used to study the seismic excitation due to an impact, mass ejection (Kanamori et al. 1984), or landslides (e.g., Kawakatsu 1989). Other types of seismic excitation are described by combinations of dipole forces, which are obtained by taking spatial derivatives of the single-force solutions, as described below.



Earthquake Mechanism Description and Inversion, Fig. 1 P- and S-wave radiation patterns for a single couple

The distance between the observation point and the location of the unit force is given by

$$r = \left[(x_1 - \zeta_1)^2 + (x_2 - \zeta_2)^2 + (x_3 - \zeta_3)^2 \right]^{1/2}. \quad (3)$$

The solution for the double-couple earthquake mechanism is obtained by taking spatial derivatives of Eq. 2 with respect to the source coordinate, where $\partial r / \partial \zeta_j = -\gamma_j$. Partial derivatives with respect to $1/r$, $1/r^3$ and the direction cosines yield additional near-field terms, whereas the partial derivative of the force time history yields

$$\frac{\partial F(t - r/\alpha)}{\partial \zeta_j} = -\frac{\gamma_j \dot{F}(t - r/\alpha)}{\alpha}, \quad (4)$$

and the far-field term for the P-wave may be written as

$$u_n^{SC}(r, t) = \frac{\gamma_n\gamma_p\gamma_j \dot{M}(t - r/\alpha)}{4\pi\rho\alpha^3 r}, \quad (5)$$

where $\dot{M}$ is the moment rate (Nm/s).

Figure 1 compares the single-couple radiation patterns for P- and S-waves. While the P-wave displays a four-lobed radiation pattern consistent

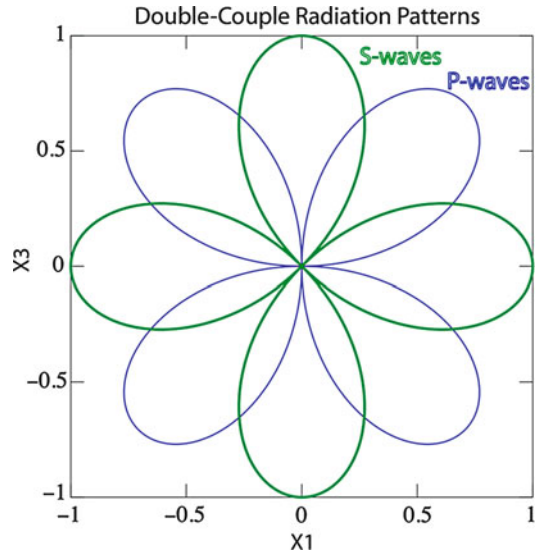
with observation, the S-wave has only a two-lobed pattern. A double-couple solution is found simply by constructing single-couple solutions for two orthogonal directions and summing. For both the P- and S-wave cases, if a single force in the $p = 1$ direction is considered and a derivative taken in $j = 3$ and summed up with a single force for $p = 3$ and a derivative in $j = 1$, the following far-field, double-couple solutions are found:

$$u_n^{\text{DC-Pwave}}(r, t) = \frac{\gamma_n \gamma_1 \gamma_3 \dot{M}(t - r/\alpha)}{4\pi\rho\alpha^3 r} \hat{x}_n + \frac{\gamma_n \gamma_3 \gamma_1 \dot{M}(t - r/\alpha)}{4\pi\rho\alpha^3 r} \hat{x}_n$$

$$u_n^{\text{DC-Swave}}(r, t) = \frac{(\delta_{n1} - \gamma_n \gamma_1) \gamma_3 \dot{M}(t - r/\beta)}{4\pi\rho\beta^3 r} \hat{x}_n + \frac{(\delta_{n1} - \gamma_n \gamma_3) \gamma_1 \dot{M}(t - r/\beta)}{4\pi\rho\beta^3 r} \hat{x}_n \quad (6)$$

Figure 2 compares the P- and S-wave radiation patterns for a double couple in the x_1x_3 plane using Eq. 6. Both the P- and S-waves display two-lobed radiation patterns consistent with observation. As shown in Aki and Richards (1980), the spatial derivatives may be carried out on all terms leading to a solution to Eq. 1 comprised of near-field, intermediate, and far-field terms ($1/r^4$, $1/r^2$, $1/r$), all of which can be represented in terms of direction cosines, strike, rake, and dip parameters of a double-couple mechanism, or more generally as a seismic moment tensor (Aki and Richards 1980).

It is common to represent the mechanism of an earthquake in terms of the P-wave radiation pattern using a focal sphere representation. An example of such a representation is given in Fig. 3 in which the lower hemisphere is presented. There are four quadrants as found in Eq. 6 and Fig. 2. The shaded quadrants indicate a P-wave first arrival with a vertical component polarity that is up, whereas the unshaded quadrants indicate a first motion that is down. In this



Earthquake Mechanism Description and Inversion, Fig. 2 P- and S-wave radiation patterns for a double couple

hemispherical representation, the horizontal angle represents the azimuth from north, and the vertical angle represents the direction that a P-wave ray leaves the source, where an angle of zero indicates a ray leaving straight down and an angle of 90° indicates a wave leaving horizontally. This representation is useful for conveying earthquake mechanism information in map view for analysis of the tectonics of a region.

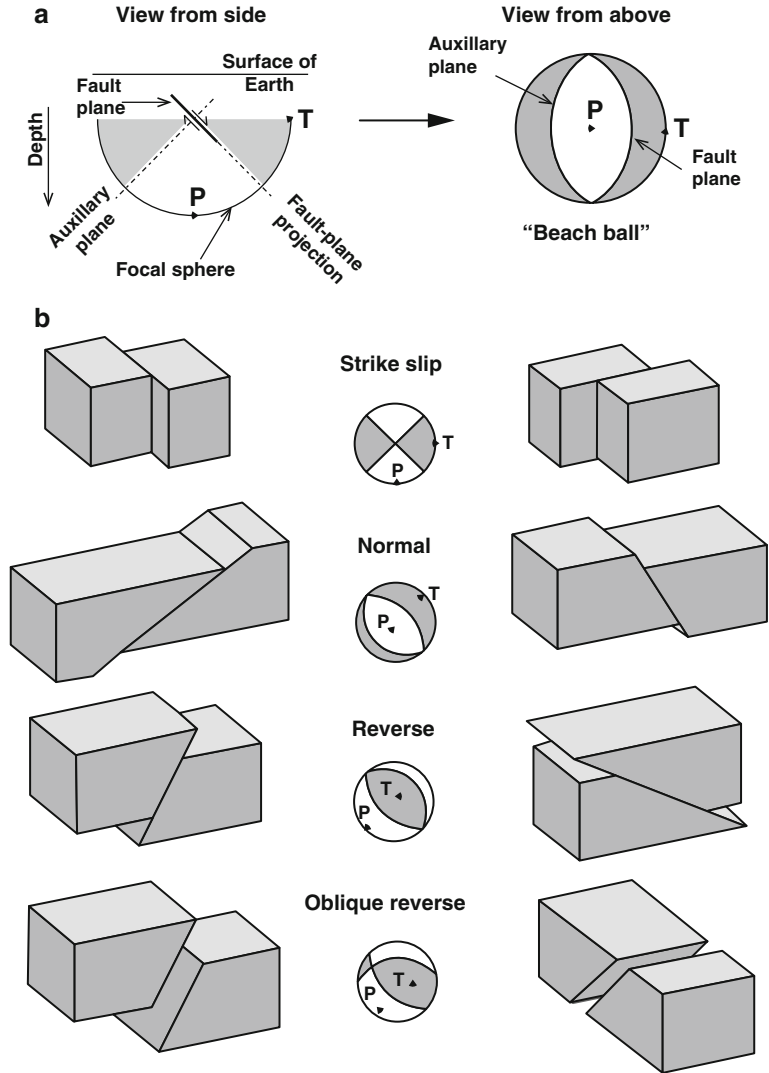
The solutions given by Eq. 6 are for a homogeneous and isotropic elastic whole space. In the next section it is shown that more general source solutions can be characterized by the seismic moment tensor for more complex elastic structure using appropriate numerical methods to compute Green's functions.

Generalized Seismic Moment Tensor

The impulse response to Eq. 1 or more generally for 1-D, 2-D, and 3-D earth structures is referred to as the Green's function solution to the elastodynamic equation of motion. Green's functions may be found for volume distributed forces, for distributed shear dislocation on an interior

Earthquake Mechanism Description and Inversion, Fig. 3

Lower hemisphere projection of the double-couple P-wave radiation pattern and relationships with block-faulting models. Shaded quadrants represent vertical component P-waves with positive polarity. The P (compression) and T (tension) axes are shown. This figure was prepared by the USGS



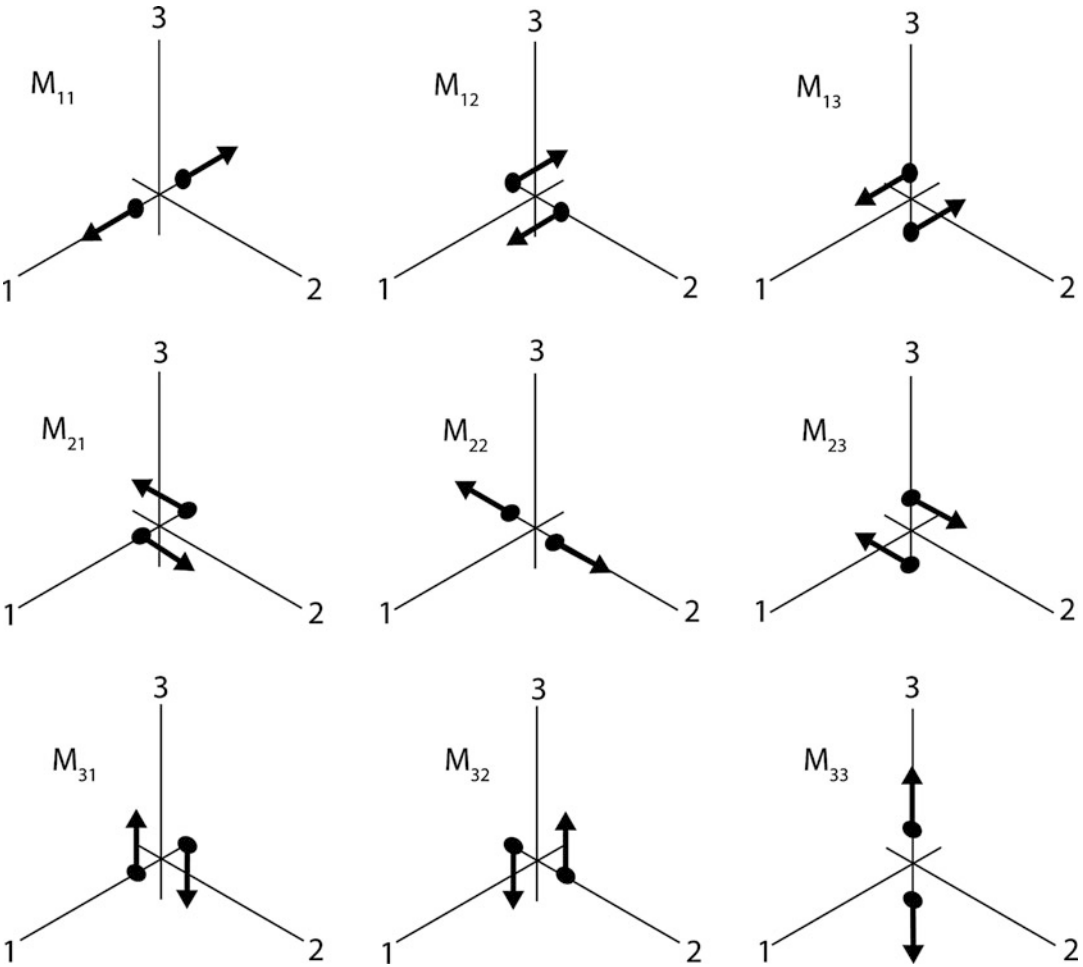
surface, or for traction acting on an exterior surface. The general representation theorem for a shear dislocation over an interior surface (Σ) within an elastic body is written as

$$u_n(\bar{x}, t) = \int d\tau \iint [u_i(\tau, \bar{\zeta})] \hat{v}_j C_{ijkl} \cdot G_{nk,l}(t - \tau, \bar{x}; 0, \bar{\zeta}) d\Sigma$$

$$u_n(\bar{x}, t) = \int d\tau \iint M_{ij}(\tau, \bar{\zeta}) \cdot G_{ni,j}(t - \tau, \bar{x}; 0, \bar{\zeta}) d\Sigma \quad (7)$$

where $[u_i(\tau, \bar{\zeta})]$ is the displacement discontinuity across a surface oriented with the unit vector normal, $\hat{v}_j$. C_{ijkl} is the elastic tensor, and $G_{nk,l}$ is

Green's tensor of single-couple solutions where n is the ground motion component, k is the direction of a unit force, and l is the direction



Earthquake Mechanism Description and Inversion,
Fig. 4 Schematic showing the system of single couples and vector dipoles comprising the seismic moment tensor.

Note that the values of M_{ij} , representing the strength of these source terms, are both real and symmetric

of the spatial derivative of the single force. See Aki and Richards (1980) for a thorough discussion of the general representation theorem of seismic sources. Equivalently, the representation theorem can be written in terms of the moment tensor (M_{ij}) distributed spatially and temporally over Σ (Eq. 7). Figure 4 illustrates the system of single couples and linear vector dipoles represented by both $G_{ni,j}$ and M_{ij} . Given a number of observations, the representation theorem provides a system of equations that if given Green's functions for 1-D, 2-D, or

even 3-D earth models, the spatiotemporal distribution of fault slip, $[u]$, may be obtained through inversion (e.g., Hartzel and Heaton 1983). There are numerous linear, linearized, and nonlinear approaches to solve this system of equations each considering different approximations on how to kinematically represent the spatiotemporal integration of the seismic source.

If a source is considered to be a point in space, in which the wavelengths of observed seismic waves are considerably larger than the

source dimension, then the spatial integral is unnecessary, and

$$\begin{aligned} u_n(\vec{x}, t) &= \int M_{ij}(\tau) G_{ni,j}(t - \tau, \vec{x}) d\tau & (a) \\ u_n(\vec{x}, t) &= \int M_{ij} \cdot S(\tau) G_{ni,j}(t - \tau, \vec{x}) d\tau & (b), \\ u_n(\vec{x}, t) &= M_{ij} G_{ni,j}(t, \vec{x}) & (c) \end{aligned} \quad (8)$$

which accounts for the general case in which each term of the moment tensor can have a different source time history (Eq. 8a), as well as the simplified case in which each element is assumed to have the same, or synchronous, time history (Eq. 8b).

The most commonly applied solution is a synchronous source in which the time history is considered to be a delta function yielding the

linear system of equations employed in seismic moment tensor analysis (Eq. 8c).

The strength of the individual single-couple and vector-dipole Green's function solutions ($G_{ni,j}$) is given by M_{ij} which is a second-order tensor that is both real and symmetric ($M_{12} = M_{21}$) enabling the decomposition into three orthogonal eigenvectors:

$$M_{ij} = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{21} & M_{22} & M_{23} \\ M_{31} & M_{32} & M_{33} \end{pmatrix} \quad (9)$$

Equation 8c is a linear system of equations that is overdetermined with three-component seismic waveform data, even with low-pass filters applied. Writing this in terms of a system of linear equations, the following is obtained:

$$GM = u$$

$$\begin{pmatrix} G_{11,1} & G_{11,2} & G_{11,3} & G_{12,2} & G_{12,3} & G_{13,3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ G_{N1,1} & G_{N1,2} & G_{N1,3} & G_{N2,2} & G_{N2,3} & G_{N3,3} \end{pmatrix} \begin{pmatrix} M_{11} \\ M_{12} \\ M_{13} \\ M_{22} \\ M_{23} \\ M_{33} \end{pmatrix} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix} \quad (10)$$

Here, N refers to the total of number of stations by components by samples of the observed waveforms and computed Green's functions. This system of equations can be solved by finding the generalized inverse where

$$M = (G^T G)^{-1} G^T u, \quad (11)$$

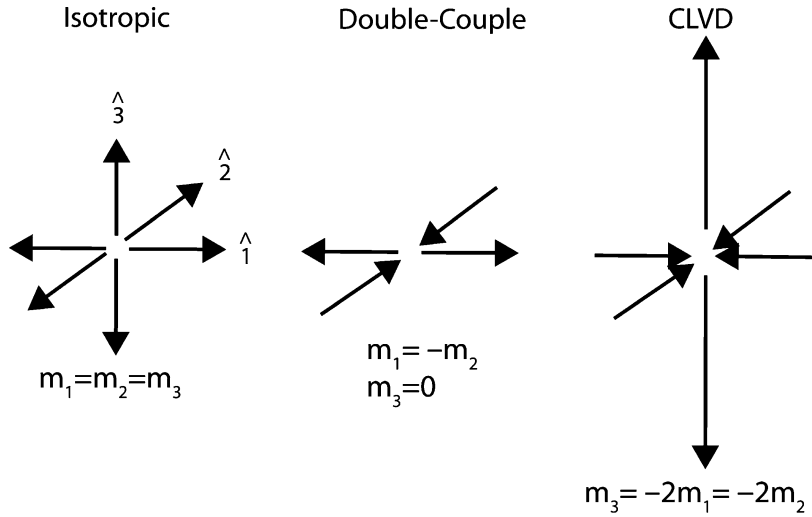
in which G is an $N \times 6$ or $N \times 5$ matrix where there are 5 or 6 independent moment tensor source parameters representing the strength of fundamental couples. All 6 parameters must be solved for if volumetric sources are to be considered. If a deviatoric solution is desired, then there are only 5 independent elements of the moment tensor and the constraint $M_{33} = -(M_{11} + M_{22})$ is applied.

Decomposition

While the solution of Eq. 11 is unique, the decomposition and interpretation of the moment tensor is not. The seismic moment tensor with 6 independent elements can be decomposed into its principle axes by performing an eigenvector decomposition. Each eigenvalue then represents the strength of the principle eigenvectors or source vector dipoles. Thus, a moment tensor can be decomposed into three mutually orthogonal vector dipoles. These vector dipoles can be combined to produce a double couple in which two of the eigenvalues have equal magnitude but opposite sign and where the third has zero strength. In the case of three nonzero eigenvalues, there are numerous ways that the vector dipoles can be combined to produce combinations of double couples, systems of volume compensated

Earthquake Mechanism Description and Inversion,

Fig. 5 Relationships between eigenvalues for the three fundamental moment tensor types



E

vector dipoles (CLVD; Knopoff and Randall 1970), and combinations of DCs and CLVDs that sum to yield the same moment tensor. Figure 5 illustrates combinations of vector dipoles leading to spherically symmetric isotropic radiation, a double couple, and a CLVD. Julian et al. (1998) give a detailed description of the different decomposition possibilities and the radiation characteristics of each type of source, and Jost and Herrmann (1989) give a thorough discussion of the moment tensor theory, seismic waveform inverse procedure, and decomposition schemes for several common cases.

Here, the isotropic, double-couple, and CLVD decomposition are given. For the general 6 parameter moment tensor (M), the isotropic moment tensor ($M_0^I I$) representing volume change is subtracted leaving the deviatoric moment tensor (M^D), where

$$M^D = M - M_0^I I \quad (12)$$

and I is the diagonal matrix. The isotropic scalar moment is defined as

$$M_0^I = (M_{11} + M_{22} + M_{33})/3. \quad (13)$$

M^D is then solved for its three orthogonal eigenvectors (a_i) and eigenvalues (m_i), which are then ranked.

$$|m_3| \geq |m_2| \geq |m_1| \quad (14)$$

There are several methods for determining the scalar seismic moment either directly from the moment tensor or from its eigenvalues (Aki and Richards 1980). Traditionally the method of Dziewonski and Woodhouse (1983) has been used where

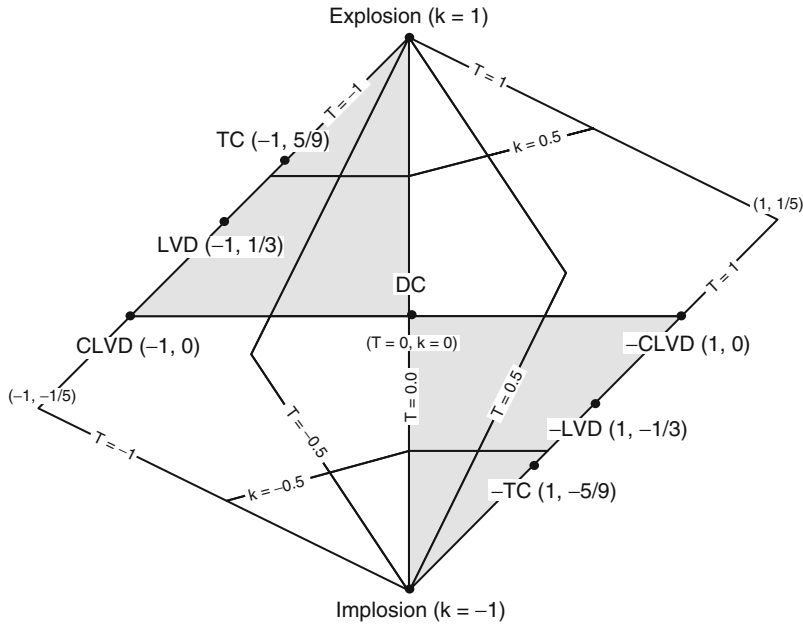
$$M_0^{DW} = \frac{|m_3| + |m_2|}{2}. \quad (15)$$

Alternatively Bowers and Hudson (1999) define the total scalar moment (M_0^T) as the sum of the isotropic moment and the largest deviatoric eigenvalue, which provides an estimate equivalent to M_0^{DW} for double couples, but provides better estimates for sources that deviate from a simple double couple for tectonic earthquakes such as the CLVD and tensile crack.

$$M_0^T = M_0^I + |m_3| \quad (16)$$

As Julian et al. (1998) point out, the decomposition of the seismic moment tensor is necessary for interpretation but is also non-unique. Today the most common decomposition approach is one in which the deviatoric moment tensor is decomposed into a double couple and a CLVD, each with the same compression (P) and tension (T) axes, following Knopoff and Randall (1970).

To decompose the moment tensor in this way, after first removing the isotropic component



Earthquake Mechanism Description and Inversion, Fig. 6 Equal area source-type plot from Bowers and Hudson (1999). Specific (T, k) values are plotted in the source-type space with the corresponding fundamental source types, where the DC point represents all double couples, CLVD has a major vector dipole in tension and

–CLVD in compression, TC and –TC are tensile and closing cracks, LVD are linear vector dipoles, and the *top* and *bottom* of the plot show the position of explosion and implosions. All other points can be interpreted as mixtures of these fundamental source types

(Eq. 12), ranking the deviatoric eigenvalues (Eq. 14), the ratio of the maximum to minimum deviatoric eigenvalue is found where

$$F = \left| \frac{m_1}{m_3} \right| \quad (17)$$

F is used to scale the scalar moment of the double-couple and CLVD components, and the complete moment tensor (M) is then represented as

$$M = M_0^I \cdot I + (1 - 2F) \cdot m_3 \cdot (a_3 a_3 - a_2 a_2) + F \cdot m_3 \cdot (2a_3 a_3 - a_2 a_2 - a_1 a_1), \quad (18)$$

$a_3 a_3$, etc. are dyadics of the eigenvectors, matrices describing the orientations of the principal vector dipoles.

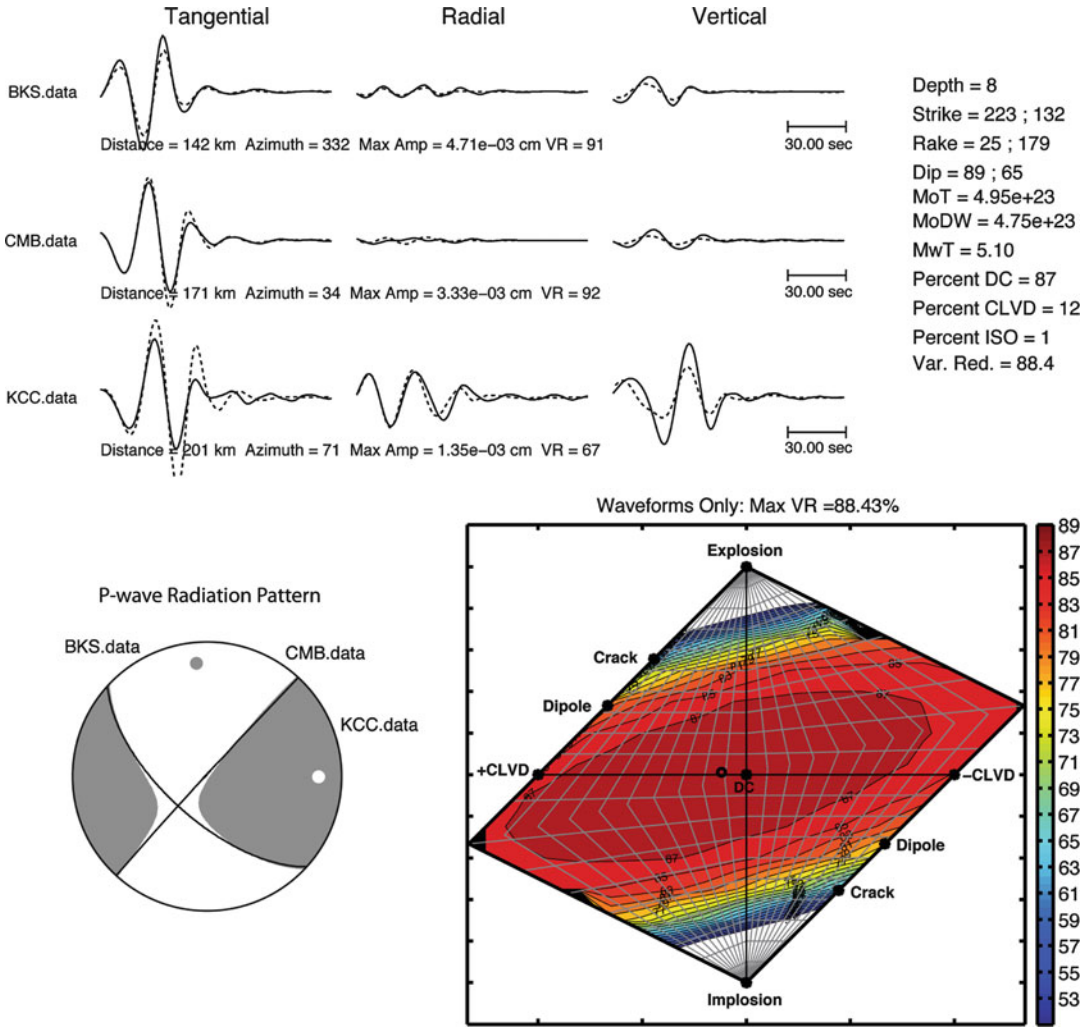
The resulting decomposition can also be characterized in terms of the percentage of moment in

each of the fundamental components. Based on the total scalar moment (Eq. 16), the following defines the percentages of the isotropic, double couple, and CVLD for the decomposition given by Eq. 18:

$$\begin{aligned} \text{Piso} &= \frac{|M_0^I|}{|M_0^T|} \cdot 100\% \\ \text{Pdc} &= (100 - \text{Piso}) \cdot (1 - 2F)\% \\ \text{Pclvd} &= (100 - \text{Piso}) \cdot 2F\% \end{aligned} \quad (19)$$

Representation Methods

Because the choice of the particular decomposition method is non-unique (e.g., Jost and Herrmann 1989; Julian et al. 1998), there is a growing interest in using graphical representations of the moment tensor eigenvalues (Hudson et al. 1989; Riedesel and Jordan 1989; Tape and Tape 2012). For example, the source-type diagram of Hudson et al. (1989) examines ratios of the three eigenvalues of the moment tensor decomposition to



Earthquake Mechanism Description and Inversion, Fig. 7 Comparison of long-period (20–50 s) three-component, complete waveforms for a tectonic earthquake on the Hayward Fault, California. The solution is

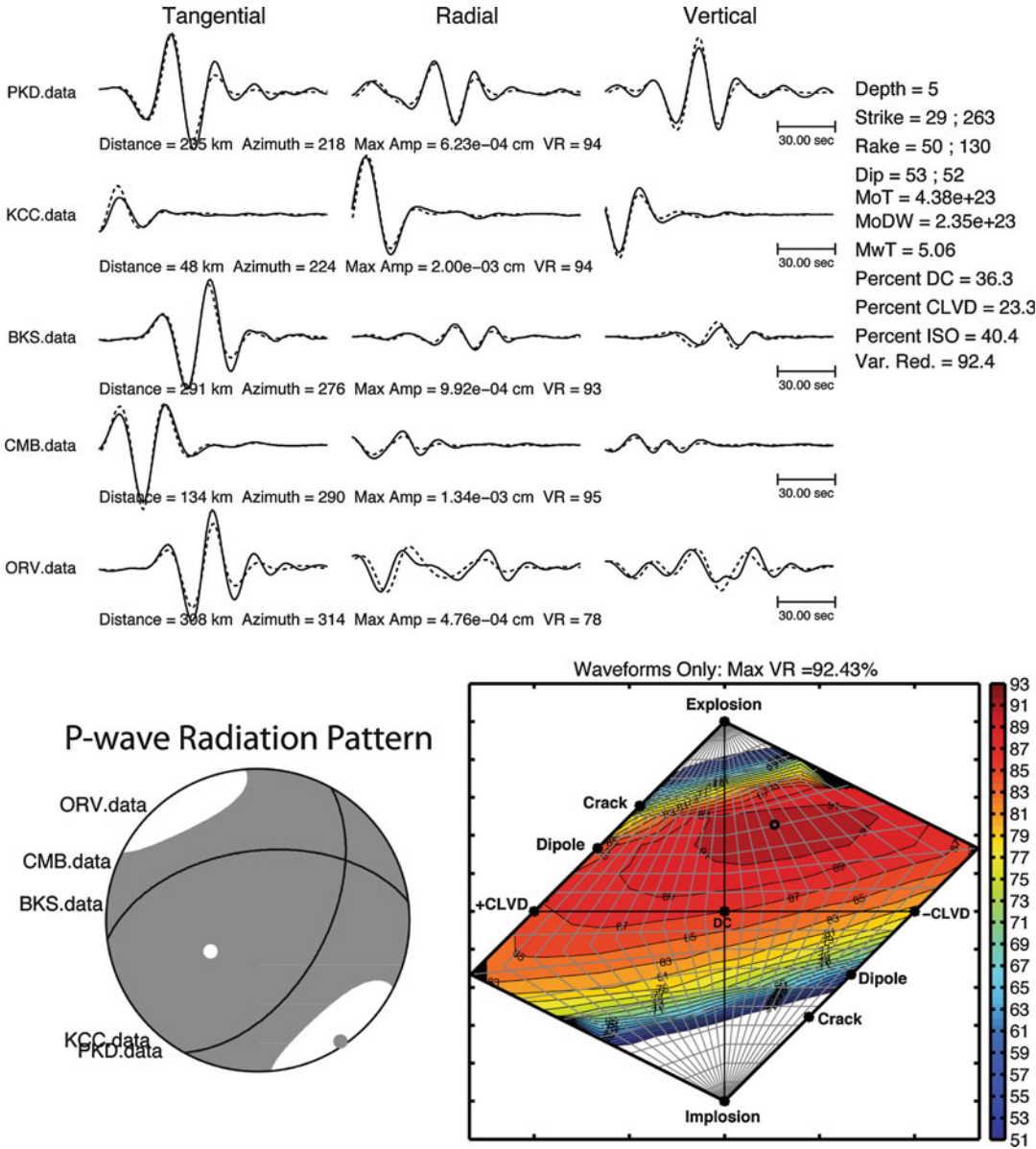
predominantly double couple. The NSS shows a wide area of acceptable solutions due to the relatively small number of stations used. The distribution of best solutions is centered on a double couple

find a scaling for a volumetric component (κ) and a deviation from a pure double-couple deviatoric solution (T) (Eq. 20). The values are plotted in a Cartesian coordinate system that is deformed to provide an equal area representation of κ and T . The advantage of this approach and the others is that a solution may be analyzed in terms of its proximity to idealized, fundamental solution types on the diagram, as well as providing a representation enabling the analysis of both

epistemic and aleatoric uncertainties in the solutions (e.g., Ford et al. 2010):

$$T = \frac{2m_1}{|m_3|}, \quad k = \frac{M'_0}{M'_0 + |m_3|} \quad (20)$$

Figure 6 illustrates the source-type projection with the position of fundamental source types identified (DC, CLVD, TC, LVD, and explosion).



Earthquake Mechanism Description and Inversion, Fig. 8 Comparison of long-period (20–50 s) three-component, complete waveforms for an earthquake in Long Valley Caldera, California. The deviatoric solution has a large CLVD, and the full moment tensor (shown) has

a large volume increase component. The NSS shows the best solutions are substantially shifted from the deviatoric line (+CLVD-DC-CLVD). The improvement in fit afforded by the volume increase component is statistically significant based on the F-test

The DC point represents all possible double-couple solutions. Positive and negative CLVD sources are for the cases with a major vector dipole in tension (example given in Fig. 5) and compression, respectively. Similarly, the linear-vector-dipole (LVD) and tension crack

(TC) solutions plot in distinct locations on the graph. All other points can be considered to be mixtures of these fundamental types. Shear tensile cracks (e.g., Šílený, J., and J. Kozák (1988) can be thought to lie along the line connecting the TC and -TC source types that passes through the

origin (DC). Julian et al. (1998) demonstrate that other locations can be interpreted as shear tensile cracks in which the two elements of the crack are not coplanar and have different geometries.

Example of a Seismic Moment Tensor Inversion for Double-Couple and Non-Double-Couple Cases

This section illustrates the inversion of three-component waveform data for the seismic moment tensor. A full moment tensor is solved for an earthquake located on the Hayward Fault in California. The broadband velocity data were instrument corrected, integrated to displacement, and filtered with an acausal, bandpass, Butterworth filter with corners set at 0.02 and 0.05 Hz. Green's functions were computed for an appropriate layered velocity structure using a frequency-wavenumber integration method and were bandpass filtered with the same filter applied to the data. Figure 7 shows the fit to the data, the P-wave radiation pattern, and pertinent source information. This earthquake is a tectonic earthquake and the solution is predominantly double couple. The small CLVD and volume increase components are due to noise and epistemic uncertainty.

The network sensitivity solution (NSS; Ford et al. 2010) is also presented. An NSS represents a maximum fit surface in the source-type space. It is computed by finding the best fit solution for each T and κ point. The fit is measured by a variance reduction. The NSS for this event shows a relatively wide distribution that is centered at a double couple but extends substantially into regions with either non-double-couple deviatoric solutions or solutions that have volume change components. This is due to the fact that only three stations were used and demonstrates uncertainty in the solution in terms of the more difficult to recover non-double-couple components. In this case non-double-couple and volumetric sources are not statistically significant.

In Fig. 8 a moment tensor solution for an earthquake from Long Valley Caldera, California, during a period of heightened seismic,

deformation, and CO_2 gas flux is presented (Dreger et al. 2000; Minson and Dreger 2008). The data processing is the same as in the previous example. The solution indicates a large volume increase component that has been found to be statistically significant using an F-test. The NSS shows the best fit region is shifted substantially from the deviatoric line.

References

- Aki K, Richards PG (1980) Quantitative seismology, vol 1. W. H Freeman and Company, San Francisco
- Bowers D, Hudson JA (1999) Defining the scalar moment of a seismic source with a general moment tensors. *Bull Seismol Soc Am* 89:1390–1394
- Dreger DS, Tkalčić H, Johnston M (2000) Dilational processes accompanying earthquakes in the Long Valley Caldera, Science 288:122–125
- Dziewonski AM, Woodhouse JH (1983) An experiment in systematic study of global seismicity: centroid-moment solutions for 201 moderate and large earthquakes. *J Geophys Res* 88:3247–3271
- Ford SR, Dreger DS, Walter WR (2010) Network sensitivity solutions for regional moment-tensor inversions. *Bull Seismol Soc Am* 100:1962–1970
- Gilbert F (1970) Excitation of the normal models of the Earth by earthquake sources. *Geophys J Roy Astron Soc* 22:223–226
- Hartzell SH, Heaton TH (1983) Inversion of strong ground motion and teleseismic waveform data for the fault rupture history of the 1979 Imperial Valley, California, earthquake. *Bull Seismol Soc Am* 73(4):1161–1171
- Hudson JA, Pearce RG, Rogers RM (1989) Source type plot for inversion of the moment tensors. *J Geophys Res* 94:765–774
- Jost ML, Herrmann RB (1989) A student's guide to and review of moment tensors. *Seismol Res Lett* 60:37–57
- Julian BR, Miller AD, Foulger GR (1998) Non-double-couple earthquakes 1. Theory. *Rev Geophys* 36:525–550
- Kanamori H, Given JW, Lay T (1984) Analysis of seismic body waves excited by the Mount St. Helens eruption of May 18, 1980. *J Geophys Res* 89(B3):1856–1866
- Kawakatsu H (1989) Centroid single force inversion of seismic waves generated by landslides. *J Geophys Res* 94(B9):12363–12374
- Kikuchi M, Kanamori H (1991) Inversion of complex body waves – III. *Bull Seismol Soc Am* 81:2335–2350
- Knopoff L, Randall MJ (1970) The compensated linear-vector dipole: a possible mechanism for deep earthquakes. *J Geophys Res* 75:4957
- Miller AD, Foulger GR, Julina BR (1998) Non-double-couple earthquakes 2. Observation. *Rev Geophys* 36:551–568

- Minson S, Dreger D (2008) Stable inversions for complete moment tensors. *Geophys J Int* 174:585–592
- Riedesel MA, Jordan TH (1989) Display and assessment of seismic moment tensors. *Bull Seismol Soc Am* 79:85–100
- Šílený J, Kozák J (1988) Shallow seismic events with combined source mechanism. *Tectonophysics* 152:283–296
- Stump BW, Johnson LR (1977) The determination of source parameters by the linear inversion of seismograms. *Bull Seismol Soc Am* 67(6):1489–1502
- Tape W, Tape C (2012) A geometric comparison of source-type plots for moment tensors. *Geophys J Int* 190:499–510

Earthquake Mechanisms and Stress Field

Václav Vavryčuk

Institute of Geophysics, Czech Academy of Sciences, Prague, Czech Republic

Synonyms

Dynamics and mechanics of faulting; Earthquake source observations; Focal mechanism; Rock failure; Shear faulting; Tectonic stress; Tensile faulting

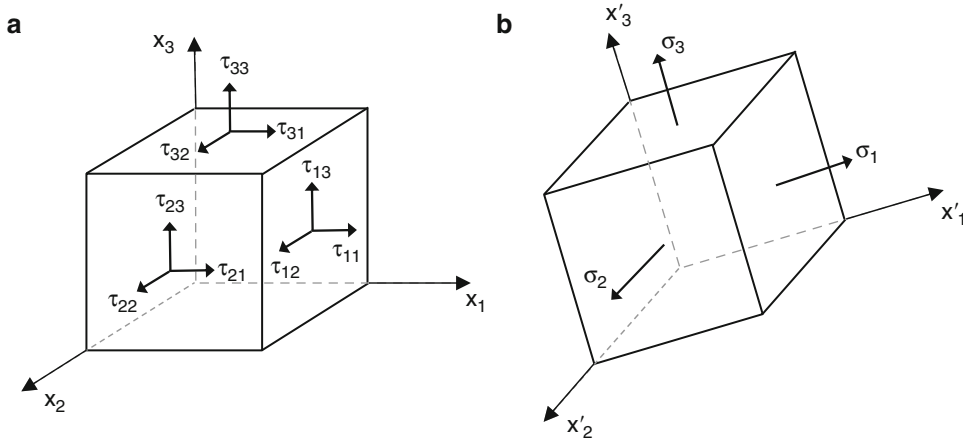
Introduction

Earthquakes are processes associated with the sudden rupture of rocks along cracks, fractures or faults exposed to stress field in the Earth's crust and lithosphere. If stress reaches a critical value exceeding strength of faults or fractures in rock, accumulated energy of elastic deformation is partially spent for anelastic deformations in the focal zone and partially released and radiated in the form of seismic waves. Stress in the Earth's crust causing earthquakes can be of tectonic or non-tectonic origin (Ruff 2002; Zoback 2007; Zang and Stephansson 2010). The main source of non-tectonic stress within the Earth is gravitational loading. This stress is vertical with largest lateral variations near the Earth's surface being more homogeneous at depth. On the other hand,

tectonic stress is mostly horizontal and originates in forces driving the plate motions (Heidbach et al. 2008). These forces typically cause “ridge push” processes where lithospheric plates are pushed to move away from spreading ridge or “slab pull” processes where two plates are in collision and one of them is subducting into the asthenosphere (Fowler 1990). The global plate tectonics successfully explained concentration of large earthquakes on margins of the plates characterized by the presence of low strength zones exposed to large stresses. On regional and local scales, stress in the Earth's crust is also affected by topography and its compensation at depth, sediment loading or presence of any heterogeneity caused by variations of density, rigidity and rock rheology, by fluid flow in rocks or by presence of faults, cracks and micro-cracks in rocks.

Properties of the stress field and of the associated fracture processes in the Earth's crust are closely related (Scholz 2002). The most common type of fracturing is shear faulting but under special stress conditions also tensile faulting can be observed (Julian et al. 1998; Vavryčuk 2011b; see entry “► [Non-Double-Couple Earthquakes](#)”). Type of faulting depends on the stress field but also on the orientation of activated fractures or faults with respect to the stress (Vavryčuk 2011a). In addition, the slip vector for shear faulting is close to or coincides with the direction of the maximum shear stress acting on the fault (Wallace 1951; Bott 1959). Therefore, type of faulting, orientation of activated faults and direction of slip along activated faults serve as an important source of information about the stress field and its spatial and lateral variations within the Earth's crust.

In this entry, the basic concept of stress is introduced and its relation to earthquake mechanisms is explained (for description of focal mechanisms, see entry “► [Earthquake Mechanism Description and Inversion](#)”). Mohr's circle diagram and simple failure criteria are described and used for defining the fault instability, principal faults and principal focal mechanisms. Methods of determining stress from observed earthquake mechanisms are reported and their robustness is



Earthquake Mechanisms and Stress Field, Fig. 1 Definition of the stress tensor in original Cartesian coordinate system (a) and in the rotated system of principal stress directions (b)

exemplified on numerical tests. Finally, several applications of stress inversions from earthquake mechanisms are listed.

of forces in a body which causes its deformation but not rotation, the stress tensor must be symmetric

$$\tau_{ij} = \tau_{ji}, \quad (5)$$

Mathematical Description of Stress

Definition of Stress

Stress describes forces acting on a unit surface in a body (see Fig. 1a). Since the acting force and the normal of the unit surface are vectors, the stress is a tensor described by nine components

$$\boldsymbol{\tau} = \begin{bmatrix} \tau_{11} & \tau_{12} & \tau_{13} \\ \tau_{21} & \tau_{22} & \tau_{23} \\ \tau_{31} & \tau_{32} & \tau_{33} \end{bmatrix}. \quad (1)$$

The force acting on surface S with normal $\mathbf{n}$ is called traction $\mathbf{T}$, and it is expressed as

$$T_i = \tau_{ij}n_j \quad (2)$$

with its normal and shear components σ_n and τ

$$\sigma_n = T_i n_i = \tau_{ij}n_i n_j, \quad (3)$$

$$\begin{aligned} \tau N_i &= T_i - \sigma_n n_i = \tau_{ij}n_j - \tau_{jk}n_j n_k n_i \\ &= \tau_{kj}n_j(\delta_{ik} - n_i n_k), \end{aligned} \quad (4)$$

where $\mathbf{N}$ is the direction of shear component τ and lies in surface S . Since stress is defined as that part

being described by six independent components only.

Mohr's Circle Diagram

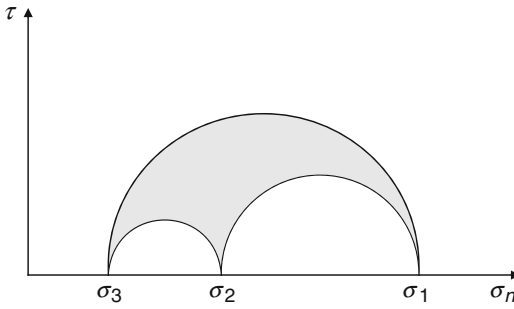
The values of the stress tensor components depend on the system of coordinates, in which the components are measured. The coordinate system can always be rotated in the way that the stress tensor diagonalizes:

$$\boldsymbol{\tau} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}, \quad (6)$$

where σ_1 , σ_2 and σ_3 are called the maximum, intermediate and minimum principal stresses (compression is positive):

$$\sigma_1 \geq \sigma_2 \geq \sigma_3, \quad (7)$$

and the vectors defining this special coordinate system are called the principal stress directions or principal stress axes (see Fig. 1b). Mathematically, the principal stresses and their directions are found by calculating the eigenvalues and eigenvectors of the stress tensor.



Earthquake Mechanisms and Stress Field, Fig. 2 Mohr's circle diagram. Quantities σ_n and τ are the normal and shear stresses along a fault, σ_1 , σ_2 and σ_3 are the principal stresses. All permissible values of σ_n and τ acting on a fault must lie in the shaded area of the diagram

The normal and shear components σ_n and τ of traction **T** (also called the normal and shear stresses) read in the system of principal stress directions

$$\sigma_n = \sigma_1 n_1^2 + \sigma_2 n_2^2 + \sigma_3 n_3^2, \quad (8)$$

$$\tau^2 = \sigma_1^2 n_1^2 + \sigma_2^2 n_2^2 + \sigma_3^2 n_3^2 - \sigma_n^2. \quad (9)$$

If principal stresses σ_1 , σ_2 and σ_3 are fixed, then normal and shear stresses σ_n and τ are just functions of normal **n** of a fault and can be plotted in the Mohr's circle diagram (see Fig. 2). All permissible values of σ_n and τ must lie in the shaded area of the diagram (Jaeger et al. 2007; Mavko et al. 2009). If the sign of shear stress τ is important in the stress analysis, the Mohr's diagram is not plotted using semi-circles but full circles, see below.

When stress conditions in real rocks in the Earth's crust are studied, the Mohr's diagram is modified. Rocks are typically porous and contain pressurized fluids which influence overall stress in the rock (Scholz 2002). For example, fluids present inside a fault reduce the normal stress along the fault. If the fluid pressure is sufficiently high, fluids can even cause opening of the fault (see Fig. 3a). For this reason, effective normal stress σ is introduced as a difference between normal stress σ_n and pore-fluid pressure p

$$\sigma = \sigma_n - p, \quad (10)$$

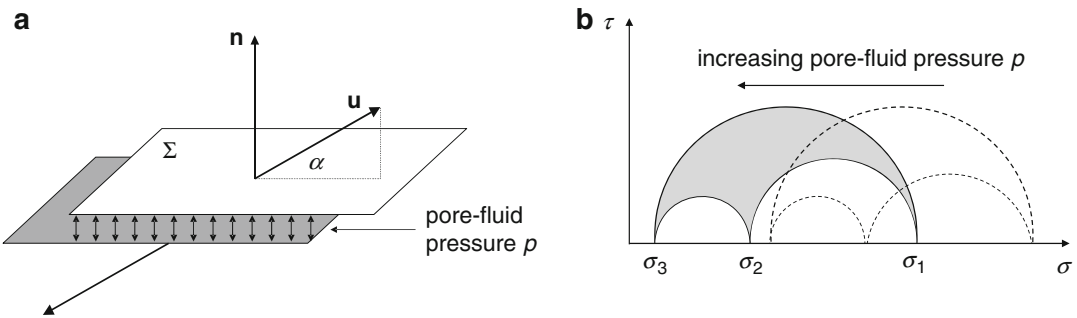
and the Mohr's circle diagram is plotted in the σ - τ plane. Variation of pore-fluid pressure in the rock is projected into shifting of the Mohr's diagram along the σ -axis. The pore-fluid pressure acts against the compressive normal stress and Mohr's circles are moved to the left with increasing pore-fluid pressure (see Fig. 3b).

Tectonic Stress and Faulting Regimes

Principal stress directions in the Earth's crust are frequently close to vertical and horizontal directions. This lead Anderson (1951) to developing a simple scheme connecting the basic stress regimes in the Earth's crust with type of faulting on a pre-existing fault in the crust (see Fig. 4). Anderson (1951) distinguishes three possible combinations of magnitudes of principal stresses: the vertical stress is maximum, intermediate or minimum with respect to the horizontal stresses. If the vertical stress is maximum, the hanging wall is moving downwards with respect to the foot wall and the normal faulting is observed along a deeply steeping fault. If the vertical stress is minimum, the crust is in horizontal compression and the hanging wall is moving upwards with respect to the foot wall and reverse faulting is observed along a shallow dipping fault. Finally, if the vertical stress is intermediate, the foot and hanging walls are moving horizontally and the strike slip faulting is observed along a nearly vertical fault. Obviously, the Anderson's classification is simple and does not cover all observations but still it proved to be valid for many seismically active regions and helpful for rough assessment of stress regime (Simpson 1997; Hardebeck and Michael 2006; see entry "► Earthquake Mechanisms and Tectonics").

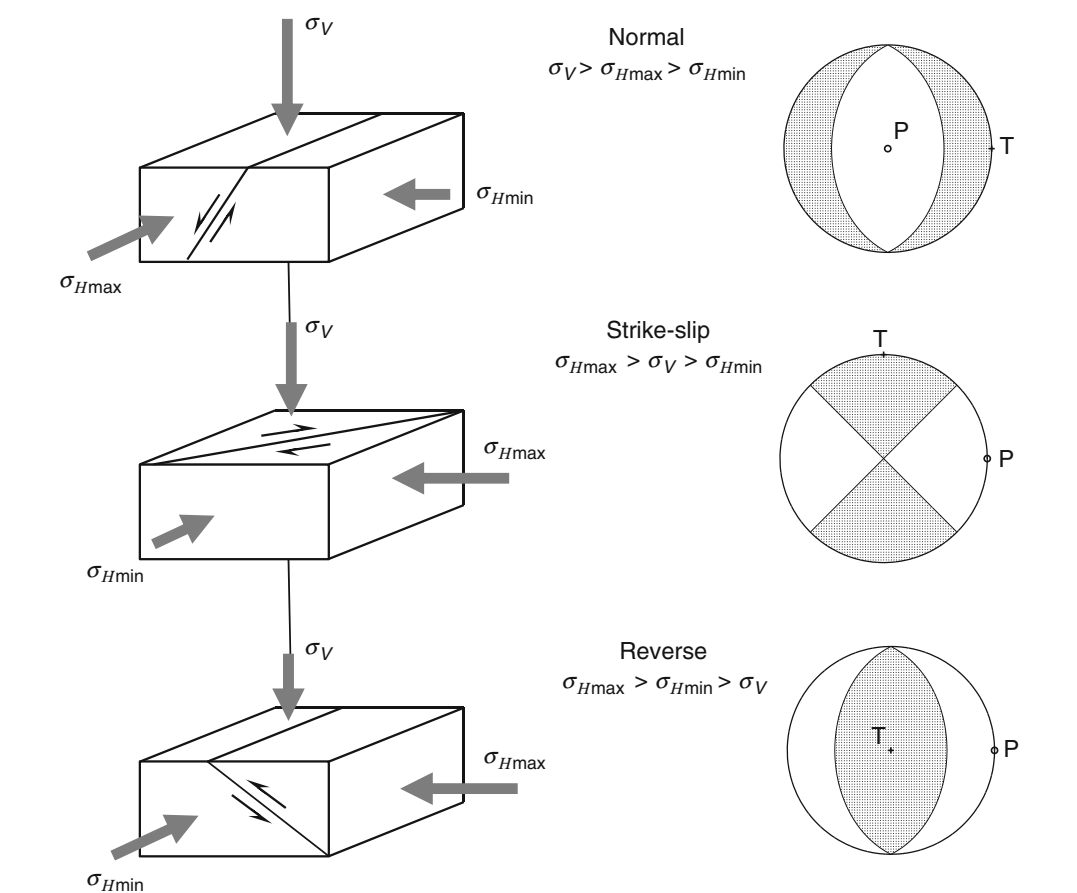
Rock Failure and Earthquakes

If a rock is critically stressed in the Earth's crust, the rock is fractured and this fracturing is associated with an earthquake. In principle, an



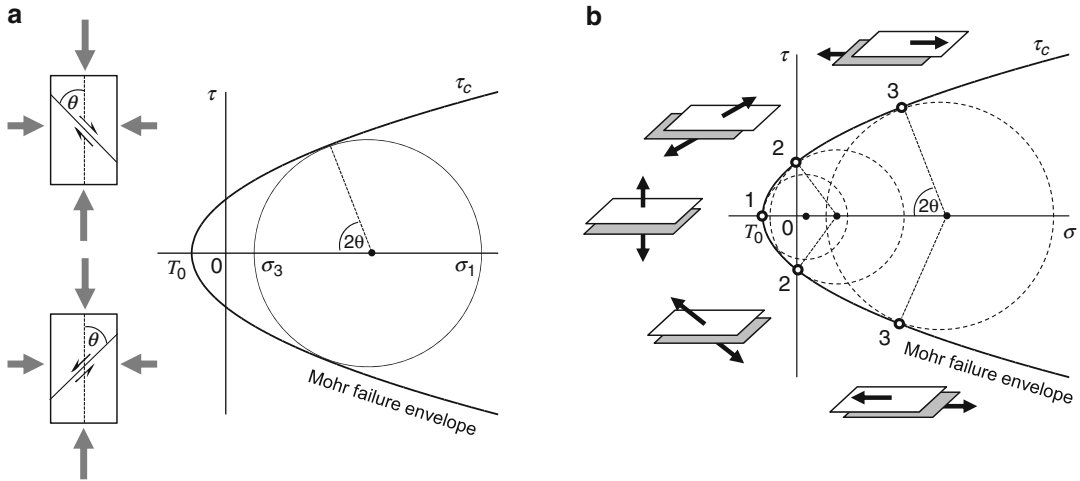
Earthquake Mechanisms and Stress Field, Fig. 3 Role of pore-fluid pressure in the rock. (a) Opening of the fault caused by fluids. (b) Shift of Mohr's circles due to increase of pore-fluid pressure. Σ is the fault, u is

the slip vector and α is the deviation between the slip vector and the fault (the so-called slope angle, see Vavryčuk 2011b)



Earthquake Mechanisms and Stress Field, Fig. 4 Anderson's classification scheme of stress in the Earth's crust (left) and corresponding faulting regimes

(right). The focal mechanisms with the P and T axes are shown in the lower-hemisphere equal-area projection



Earthquake Mechanisms and Stress Field, Fig. 5 Griffith failure criterion. (a) Scheme of a cylindrical specimen with a fracture created by loading with stresses $\sigma_1 > \sigma_2 = \sigma_3$ (left) and the corresponding Mohr's circle diagram (right). (b) Position of different

faulting regimes on the Mohr failure envelope $\tau_c(\sigma)$. Point 1 – pure tensile faulting, points 2 – transition between tensile and pure shear faulting, points 3 – pure shear faulting. Angle θ defines the deviation of the fracture from the σ_1 direction, T_0 is the uniaxial tensile strength

earthquake can occur on a newly developed fracture or on a pre-existing fault in the Earth's crust which is re-activated. The condition under which fracturing or faulting occurs is described by the so-called failure criteria. The simplest and most known criteria are the Griffith failure criterion derived from energy conditions imposed on propagating cracks in a rock, and the Mohr-Coulomb failure criterion (also called the Coulomb failure criterion) based on the concept of friction between two sliding blocks simulating the case of faulting on a pre-existing fault. Both failure criteria predict critical values of shear stress as a function of normal stress which lead to failure

$$\tau_c = f(\sigma). \quad (11)$$

They can be plotted as the so called Mohr failure envelopes in the σ - τ diagram together with Mohr's circles and employed for analysis of stability of fractures or faults under given stress conditions. If the outer Mohr's circle touches the failure envelope, there is one fracture or fault which is unstable and can fail, its orientation being defined by inclination θ of the fault from the maximum stress direction (see Fig. 5a).

Griffith Failure Criterion

The Griffith theory of fracture predicts the failure envelope in the following form (Jaeger et al. 2007, Eq. 10.139):

$$\tau_c^2 = 4T_0(T_0 + \sigma), \quad (12)$$

where T_0 is the uniaxial tensile strength. The criterion is described by a parabola with the most curved part close to tensile normal stresses. The point at the parabola with the lowest value of normal stress (point 1 in Fig. 5b) corresponds to fracture lying along the maximum principal stress direction (maximum compression) σ_1 and opening during the fracture process. This physically describes the case of a pure tensile crack created, for example, by hydrofracturing due to fluid injection into the rock mass (Zoback 2007). The intersection of the parabola with the τ -axis (points 2 in Fig. 5b) defines the transition between tensile and shear modes of faulting and corresponds to fracture orientation which can be found using the following simple formula (Fischer and Guest 2011):

$$\tan 2\theta_{\text{tensile}} = \left. \frac{\partial \tau_c}{\partial \sigma} \right|_{\sigma=0} = \left. \frac{\sqrt{T_0}}{\sqrt{T_0 + \sigma}} \right|_{\sigma=0} = 1, \quad (13)$$

and consequently $\theta_{\text{tensile}} = 22.5^\circ$. Hence, tensile faulting (called also shear-tensile faulting) associated with fault opening can occur only for fractures which deviate from the direction of maximum compression with angle lower than θ_{tensile} (Mohr failure envelope between points 1 and 2 in Fig. 5b). For angles equal to or higher than θ_{tensile} only shear faulting mode can be observed (e.g., points 3 in Fig. 5b). As seen from Fig. 5b, the parabola is less curved for angles higher than θ_{tensile} and the Griffith failure criterion can be well approximated by the linear Mohr-Coulomb failure criterion, see below.

Mohr-Coulomb Failure Criterion

According to the Mohr-Coulomb failure criterion (Scholz 2002; Zoback 2007), shear stress on an activated fault must exceed critical value τ_c , which is calculated from cohesion C , fault friction μ and effective normal stress σ :

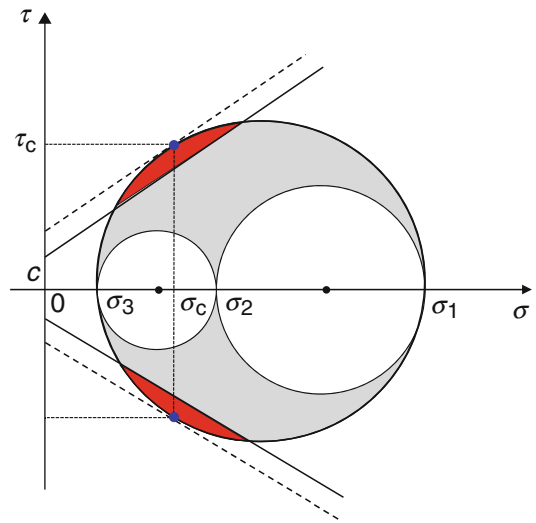
$$\tau_c = C + \mu\sigma, \quad (14)$$

or equivalently

$$\tau_c = C + \mu(\sigma_n - p), \quad (15)$$

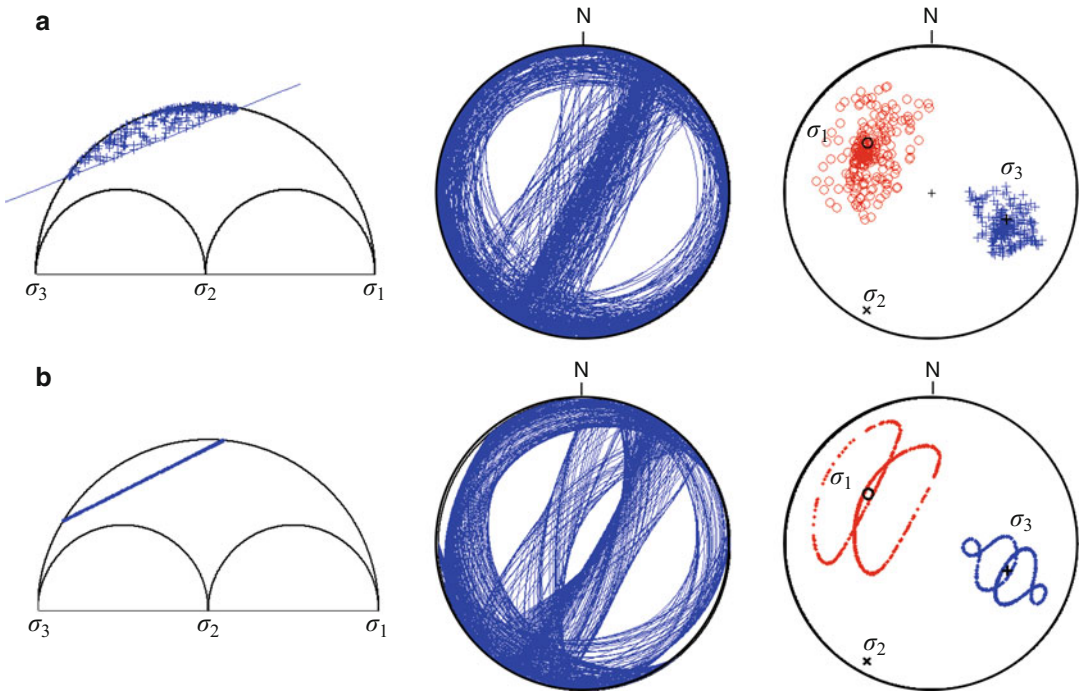
where σ_n is the normal stress and p is the pore pressure. Friction μ for fractures was measured on rock samples in laboratory and ranges mostly between 0.6 and 0.8 (Byerlee 1978). The values of friction of faults in the Earth's crust are similar (Vavryčuk 2011a) but lower values like 0.2–0.4 have also been reported for some large-scale faults like the San Andreas fault (Scholz 2002).

If the Mohr-Coulomb failure criterion is satisfied (red area in Fig. 6), the fault becomes unstable and an earthquake occurs along this fault. The higher the shear stress difference, $\Delta\tau = \tau - \tau_c$, the higher the instability of the fault and the higher the susceptibility of the fault to be activated. A fault most susceptible to failure is called the “principal” fault (Vavryčuk 2011a) being defined by the point in which the Mohr-Coulomb failure criterion touches the Mohr's circle diagram (blue point in Fig. 6)



Earthquake Mechanisms and Stress Field, Fig. 6 Mohr-Coulomb failure criterion. The red area shows all possible orientations of fault planes which satisfy the Mohr-Coulomb failure criterion. The blue dot with shear and normal stresses τ_c and σ_c denotes the principal fault plane which is optimally oriented with respect to stress, and C denotes the cohesion

A variety in possible orientations of unstable fault planes is demonstrated in Fig. 7. The left-hand plot of Fig. 7a shows the Mohr's diagram, the failure criterion and the positions of randomly distributed unstable fault planes satisfying the failure criterion. The middle and right-hand plots of Fig. 7a show the nodal lines and the P (pressure) and T (tension) axes for the corresponding focal mechanisms. The nodal lines and P/T axes inform about the predominant type of faulting and about the scatter in orientations of the unstable fault planes. Predominant faulting and its scatter are also projected into a scatter of the P/T axes, which form clusters of a specific shape and size (Fig. 7a, right-hand plot). The form of the clusters is best demonstrated using the projection of the Mohr-Coulomb failure criterion (Fig. 7b, blue line in the left-hand plot) onto the focal sphere (Fig. 7b, right-hand plot). The failure criterion curve splits into two closed curves for the P axes as well as for the T axes on the focal sphere corresponding to the failure conditions in the upper and lower half-planes in the Mohr's circle diagram. The two P/T failure



Earthquake Mechanisms and Stress Field, Fig. 7 Focal mechanisms associated with unstable fault planes. (a) Randomly distributed fault planes inside the unstable area of the Mohr's diagram (left), corresponding nodal lines (in the middle) and the P/T axes (right). The P (pressure) axes are marked by the red circles, the T - (tension) axes by the blue crosses. (b) Fault planes corresponding to the Mohr failure envelope in the

Mohr's diagram (left), corresponding nodal lines (in the middle) and the P/T axes (right). The P/T axes form the P (red)/T (blue) failure curves displaying the two-wing pattern. The σ_1 , σ_2 and σ_3 stress axes have directions (azimuth/plunge): $308^\circ/44^\circ$, $209^\circ/9^\circ$ and $110^\circ/44^\circ$. The shape ratio R is 0.5. The azimuth angle is measured clockwise from north (denoted as N), and the plunge angle from the horizontal plane

curves may intersect each other or be separated. The area inside the failure curves define positions of the P/T axes of focal mechanisms conceivable under the given stress regime. The pattern when the P/T axes form two distinct sub-clusters is called as the “two-wing” or the “butterfly-wing” pattern (Vavryčuk 2011a). The butterfly wings are well separated provided that friction is high (0.5 or higher). If friction is low, the wings come closer or they overlap.

Fault Instability

Since differently oriented faults have a different susceptibility to be activated and thus being differently unstable in the given stress field, a quantity which measures this instability can be introduced. For example, the fault instability I of all fault orientations can be defined in

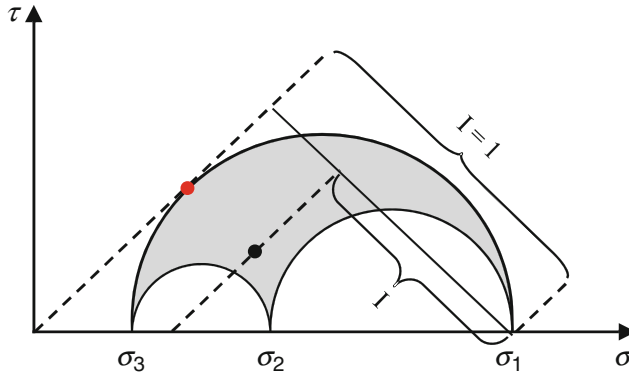
the range from 0 to 1 (see Fig. 8) using the following formula (Vavryčuk et al. 2013):

$$I = \frac{\tau - \mu(\sigma - \sigma_1)}{\tau_c - \mu(\sigma_c - \sigma_1)}, \quad (16)$$

where τ_c and σ_c are the shear and effective normal stresses along the principal fault (blue dot in Fig. 6, red dot in Fig. 8), and τ and σ are the shear and effective normal stresses along the analysed fault (black dot in Fig. 8).

Since Eq. 16 is independent of absolute stress values, the fault instability I can be evaluated just from friction μ , shape ratio R ,

$$R = \frac{\sigma_1 - \sigma_2}{\sigma_1 - \sigma_3}, \quad (17)$$



Earthquake Mechanisms and Stress Field, Fig. 8 Definition of the fault instability in the Mohr's diagram. The *red dot* marks the principal fault characterized by instability $I = 1$. The *black dot* marks an arbitrarily

oriented fault with instability I . Quantities τ and σ are the shear and the effective normal stresses, respectively; σ_1 , σ_2 and σ_3 are the effective principal stresses

and from directional cosines $\mathbf{n}$ defining the inclination of the fault plane from the principal stress axes (i.e., $\mathbf{n}$ being expressed in the coordinate system of the principal stress directions). If the reduced stress tensor is scaled as follows:

$$\sigma_1 = 1, \sigma_2 = 1 - 2R, \sigma_3 = -1, \quad (18)$$

where positive values mean compression, stresses τ_c and σ_c along the principal fault read

$$\tau_c = \frac{1}{\sqrt{1 + \mu^2}}, \sigma_c = -\frac{\mu}{\sqrt{1 + \mu^2}}, \quad (19)$$

and consequently

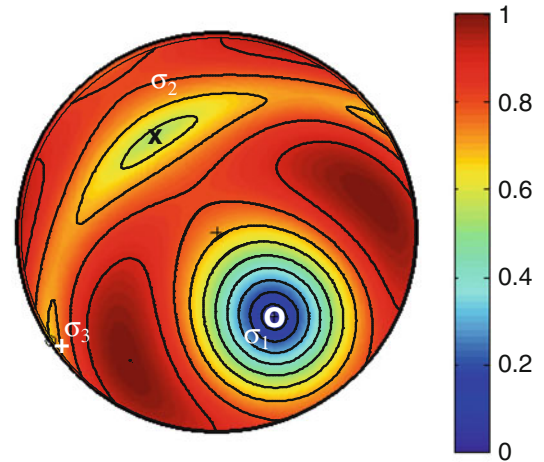
$$I = \frac{\tau - \mu(\sigma - 1)}{\mu + \sqrt{1 + \mu^2}}, \quad (20)$$

where

$$\sigma = n_1^2 + (1 - 2R)n_2^2 - n_3^2,$$

$$\tau = \sqrt{n_1^2 + (1 - 2R)^2 n_2^2 + n_3^2 - (n_1^2 + (1 - 2R)n_2^2 - n_3^2)^2}. \quad (21)$$

The fault instability can be calculated using Eqs. 20 and 21 for all orientations of fault planes and projected into the focal sphere (see Fig. 9). Such figure is, in particular, helpful when analysing



Earthquake Mechanisms and Stress Field, Fig. 9 The fault instability is shown for all possible fault normals; it is colour-scaled and ranges between 0 (the most stable plane) and 1 (the most unstable plane). The lower-hemisphere equal-area projection is used. Directions of the σ_1 , σ_2 and σ_3 axes are (azimuth/plunge): $146^\circ/48^\circ$, $327^\circ/42^\circ$ and $237^\circ/1^\circ$. The shape ratio R is 0.80 and the fault friction μ is 0.5 (Modified after Vavryčuk et al. (2013))

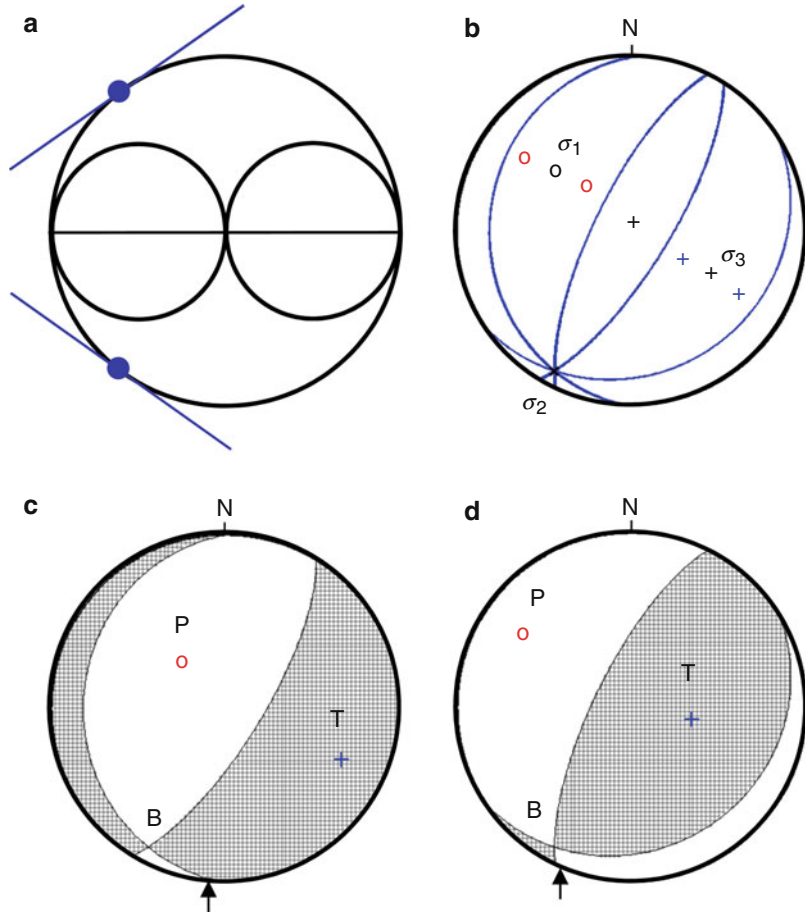
orientations of unstable faults under stress field with inclined principal stress directions.

Principal Focal Mechanisms

As mentioned above, the higher the instability, the higher the probability of the fault being activated. Obviously, a prominent fault plane is the

Earthquake Mechanisms and Stress Field,

Fig. 10 Principal focal mechanisms. (a) Full Mohr's circle diagram, (b) principal nodal lines and P/T axes, and (c, d) principal focal mechanisms. The blue dot in (a) marks the principal fault plane, the arrows in (c) and (d) denote the nodal lines corresponding to the principal fault plane. Point B in (c) and (d) denotes the neutral axis, N denotes north (After Vavryčuk (2011a))



principal fault (Vavryčuk 2011a). This plane is optimally oriented in the stress field and has the highest instability $I = 1$ (see Figs. 6 and 8). Each stress allows for the existence of two distinct principal faults and two distinct principal focal mechanisms (Fig. 10). The reason for the existence of two principal mechanisms is the same as for observing the two-wing pattern of the failure curves and two clusters of focal mechanisms – one principal fault lying in the upper half-plane and the other principal fault in the lower half-plane of the Mohr's circle diagram. The P/T axes of the principal focal mechanisms lie within the two P/T butterfly wings. If the area of unstable faults (see Fig. 6) is decreased, for example, by decreasing pore pressure, the wings become smaller being shrunk to the P/T axes of the principal focal mechanisms in the limit.

The orientation of the principal fault planes and principal focal mechanisms depends on stress and fault friction. The B (neutral) axes of both principal focal mechanisms coincide with the σ_2 axis (Fig. 10b). The P/T axes lie in the σ_1 – σ_3 plane: the σ_1 axis is in the centre between the two P axes, and the σ_3 axis is in the centre between the two T axes. The principal fault planes are those nodal planes whose deviation from the σ_1 axis is less than 45° . Deviation θ between the two principal fault planes is expressed by a formula similar to that for the 2-D stress:

$$\mu^{-1} = \tan 2\theta. \quad (22)$$

Since the relation between the principal focal mechanisms and the orientation of stress is

simple, the inversion for stress orientation from principal focal mechanisms is straightforward. On the other hand, the shape ratio cannot be retrieved from the principal focal mechanisms. The shape ratio constrains the mutual relation between the size and shape of the P and T butterfly wings. Therefore, a rather large set of focal mechanisms is needed to map the wings accurately. Methods for inversion of stress including the shape ratio from a set of focal mechanisms are described below.

Inversion for Stress From Earthquake Mechanisms

Several methods have been proposed for the determination of stress from a set of focal mechanisms of earthquakes (Maury et al. 2013). These methods usually assume that (1) tectonic stress is uniform (homogeneous) in the region, (2) earthquakes occur on pre-existing faults with varying orientations, (3) the slip vector points in the direction of shear stress on the fault (the so-called Wallace-Bott hypothesis; see, Wallace 1951; Bott 1959), and (4) the earthquakes do not interact with each other and do not disturb the background tectonic stress. Obviously, these conditions might not be always satisfied. Firstly, in case that stress is not uniform, the area should be subdivided into smaller areas in which the condition of uniform tectonic stress is reasonable to assume. Secondly, if a high variety of focal mechanisms is not observed, the inversion yields less accurate results. The smaller variety of the observed focal mechanisms, the larger uncertainty in the stress orientations retrieved. Thirdly, the Wallace-Bott assumption of the slip vector parallel to the stress on the fault is valid only in isotropic media. In anisotropic media, both vectors need not be parallel and the problem becomes more involved, particularly, if anisotropy in the focal zone is not known. Finally, stress changes due to the occurrence of small earthquakes are usually negligible but large earthquakes can significantly affect the

background stress field. In this case, inverting for stress should be performed from earthquakes clustered not only in space but also in time and periods between and after a large earthquake should be distinguished.

If the above-mentioned assumptions are reasonably satisfied, the stress inversion methods are capable to determine four parameters of the stress tensor: three angles defining the directions of the principal stress directions, σ_1 , σ_2 and σ_3 , and shape ratio R . The methods are unable to recover the remaining two parameters of the stress tensor. Therefore, the stress tensor is usually searched with the normalized maximum compressive stress

$$\sigma_1 = +1, \quad (23)$$

and with the zero trace

$$Tr(\boldsymbol{\tau}) = \sigma_1 + \sigma_2 + \sigma_3 = 0. \quad (24)$$

This implies that no information about pore pressure, which is isotropic, can be gained from the focal mechanisms.

Next, the individual approaches to stress inversion are described and their pros and cons are mentioned.

Michael Method

A simplest approach to stress inversion is the method of Michael (1984). This method employs Eq. 4 expressed in the following form:

$$\tau_{kj}n_j(\delta_{ik} - n_in_k) = \tau N_i. \quad (25)$$

In order to be able to evaluate the right-hand side of the equation, Michael (1984) applies the Wallace-Bott assumption that direction $\mathbf{N}$ of shear stress component of traction $\mathbf{T}$ on a fault is identical with the slip direction $\mathbf{s}$, and he further assumes that shear stress τ on activated faults has the same value for all studied earthquakes. Since the method cannot determine absolute stress values, τ in Eq. 25 is assumed to be 1 and

Eq. 23 is not applied. Subsequently, Eq. 25 is expressed in the matrix form

$$\mathbf{A}\mathbf{t} = \mathbf{s}, \quad (26)$$

where $\mathbf{t}$ is the vector of stress components

$$\mathbf{t} = [\tau_{11} \quad \tau_{12} \quad \tau_{13} \quad \tau_{22} \quad \tau_{23}]^T, \quad (27)$$

$\mathbf{A}$ is the 3×5 matrix calculated from fault normal $\mathbf{n}$,

$$\begin{bmatrix} n_1(n_2^2 + 2n_3^2) & n_2(1 - 2n_1^2) & n_3(1 - 2n_1^2) & n_1(-n_2^2 + n_3^2) & -2n_1n_2n_3 \\ n_2(-n_1^2 + n_3^2) & n_1(1 - 2n_2^2) & -2n_1n_2n_3 & n_2(n_1^2 + 2n_3^2) & n_3(1 - 2n_2^2) \\ n_3(-2n_1^2 - n_2^2) & -2n_1n_2n_3 & n_1(1 - 2n_3^2) & n_3(-n_1^2 - 2n_2^2) & n_2(1 - 2n_3^2) \end{bmatrix}, \quad (28)$$

and $\mathbf{s}$ is the unit direction of the slip vector. Extending Eq. 28 for focal mechanisms of K earthquakes with known fault normals $\mathbf{n}$ and slip directions $\mathbf{s}$, a system of $3 \times K$ linear equations for five unknown components of stress tensor is obtained. The system is solved using the generalized linear inversion in the L2-norm (Lay and Wallace 1995, their Section 6.4)

$$\mathbf{t} = \mathbf{A}^{-g}\mathbf{s}. \quad (29)$$

The basic drawback of this method is the necessity to know orientations of the faults. Usually, when determining the focal mechanisms, orientations of the two nodal planes are calculated: one nodal plane corresponding to the fault and the other nodal plane (called the auxiliary plane) defining the slip direction. The inherent ambiguity of focal mechanisms does not allow distinguishing easily which of the nodal planes is the fault. If the Michael's method is used with incorrect orientations of the fault planes, the accuracy of the retrieved stress tensor is decreased. On the other hand, the method is quite fast and it can be run repeatedly. Therefore, the confidence regions of the solution are determined using the standard bootstrap method (Michael 1987). If the orientation of fault planes in the focal mechanisms is unknown, each nodal plane has a 50 % probability of being chosen during the bootstrap resampling.

Recently, Vavryčuk (2014) modified the Michael's method and removed the difficulty with the unknown fault orientations. He proposed inverting jointly for stress and for the

fault orientations by applying the fault instability constraint. According to this constraint, the fault is identified with that nodal plane which has a higher value of instability calculated using Eq. 20. The stress is calculated in iterations and overall friction μ on faults is also determined. Numerical tests show that the iterative stress inversion is fast and accurate and performs better than the standard Michael's inversion.

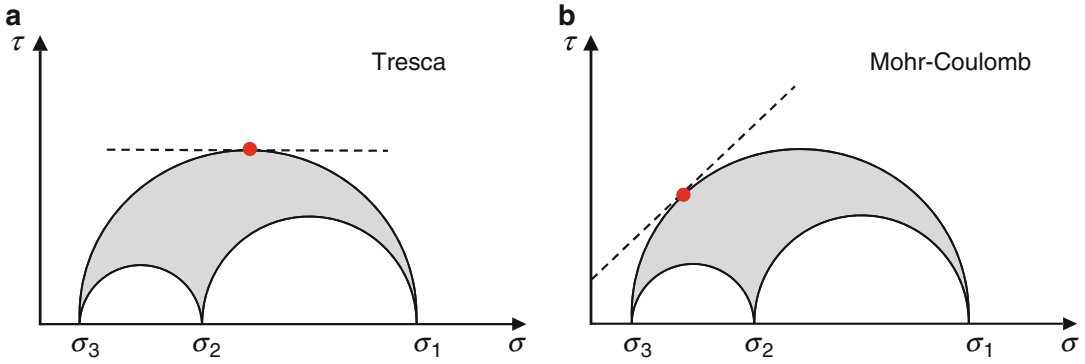
Angelier Method

The difficulty with determining the fault plane in the focal mechanisms is also overcome in the method developed by Angelier (2002). This method is based on maximizing the slip shear stress component (SSSC) along the fault

$$\tau_s = T_i s_i = \tau_{ij} n_i s_j. \quad (30)$$

Since the SSSC value is symmetric with respect to vectors $\mathbf{n}$ and $\mathbf{s}$, it is invariant of the choice of the fault plane from the two nodal planes. The condition of the maximum shear stress Eq. 30 is not fully correct and physically means that faults should obey the so-called Tresca failure criterion where faults are assumed to have zero friction. Only faults with zero friction can achieve maximum shear stress and satisfy Eq. 30; for faults with friction, the SSSC value is always reduced (see Fig. 11).

For K focal mechanisms, the total SSSC value is maximized as follows:



Earthquake Mechanisms and Stress Field, Fig. 11 Tresca and Mohr-Coulomb failure criteria. The Tresca failure criterion is unphysical and corresponds to a fault with zero friction. The red dot marks the principal fault plane

$$\tau_s^{\text{total}} = \frac{\tau_{ij}}{\tau_{\text{max}}} \sum_{k=1}^K n_i^{(k)} s_j^{(k)}, \quad (31)$$

where

$$\tau_{\text{max}} = \frac{\sigma_1 - \sigma_3}{2}. \quad (32)$$

and the optimum stress tensor is calculated analytically (see Angelier 2002). An alternative approach defines the total SSSC value in the form:

$$\tau_s^{\text{total}} = \tau_{ij} \sum_{k=1}^K n_i^{(k)} s_j^{(k)}, \quad (33)$$

with the normalization condition

$$\sigma_1^2 + \sigma_2^2 + \sigma_3^2 = 1. \quad (34)$$

and with the zero trace condition Eq. 24. In this case, the inversion for the optimum stress tensor can be performed using a grid search through the principal stress directions and shape ratio R . The fit function is maximized in the L1-norm.

Although the condition of maximizing the SSSC value looks as a rather rough approximation, its use in practice is advocated by Angelier (2002) for the following reasons. Firstly, the activated fault planes are located mostly close to the top of the Mohr's circle diagram. For example, friction of

0.6 reduces the maximum shear stress by 13 % only. Secondly, numerical tests show that if focal mechanisms inverted display a sufficient variety, maximizing the SSSC value leads to a small bias in the retrieved stress directions. Thirdly, maximizing the SSSC value is quite robust with respect to the errors present in focal mechanisms.

Gephart and Forsyth Method

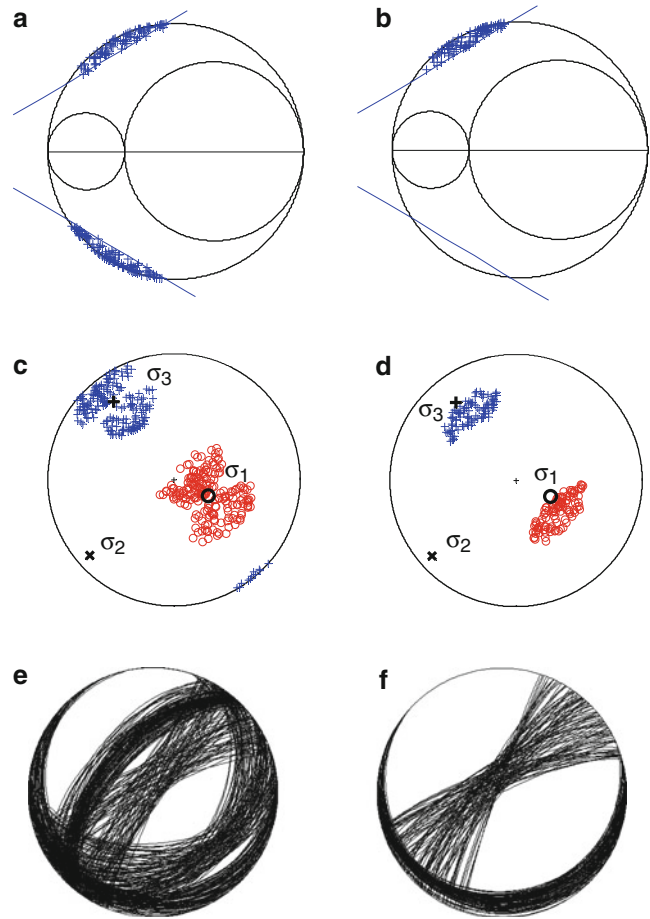
Another way how to solve the problem with ambiguous fault plane orientations in stress inversions is proposed by Gephart and Forsyth (1984). The method is based on minimizing the deviation between direction $\mathbf{N}$ of shear stress and slip direction $\mathbf{s}$ along the fault. This deviation is calculated for both options of the fault orientation and the lower value enters into the total sum over all K focal mechanisms minimized during the inversion:

$$S = \sum_{k=1}^K \text{acos}(\mathbf{N}^{(k)} \cdot \mathbf{s}^{(k)}). \quad (35)$$

The method was further modified and improved by Lund and Slunga (1999) who applied a more physical criterion for identifying the true fault planes. The criterion is based on the evaluation of instability of the fault using Eq. 20 and the fault plane is identified with that nodal plane which is more unstable and thus more susceptible to failure under the given stress field. The misfit function is minimized in the L1-norm by using

Earthquake Mechanisms and Stress Field,

Fig. 12 Example of data used in numerical tests of stress inversions. The plots show 200 noise-free focal mechanisms selected to satisfy the Mohr-Coulomb failure criterion. *Left/right* plots – dataset with a full/reduced variety of focal mechanisms. (a, b) Mohr's circle diagrams, (c, d) P/T axes and (e, f) corresponding nodal lines. The P axes are marked by the red circles and T axes by the blue crosses in (c) and (d). The σ_1 , σ_2 and σ_3 stress axes are (azimuth/plunge): $115^\circ/65^\circ$, $228^\circ/10^\circ$ and $322^\circ/23^\circ$, respectively. Shape ratio R is 0.70, cohesion C is 0.85, pore pressure p is zero and friction μ is 0.60. The minimum instability of faults is 0.82



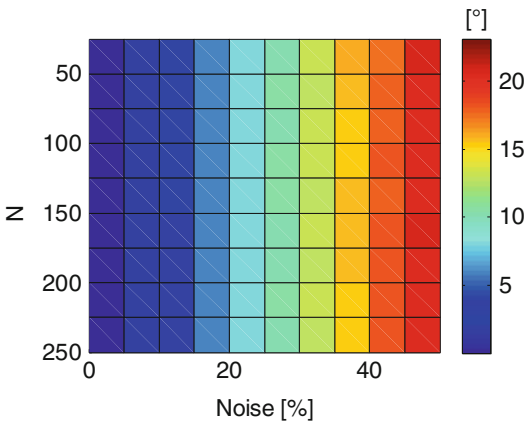
the robust grid search inversion scheme. It is also possible to exclude a predefined number of outliers from the misfit function the result of the inversion to be less sensitive to large errors in input data. Since the misfit function is usually smooth and simple, the time-consuming grid search can be substituted by any non-linear inversion method. For example, the gradient or simplex methods can be applied with a success. This applies also to the Angelier method defined by Eqs. 33 and 34.

Accuracy of Stress Inversion Algorithms: Numerical Tests

The robustness and accuracy of the stress inversion methods can be efficiently tested using numerical modelling. The key parameters of

any stress inversion are the number of focal mechanisms inverted, their accuracy and variability. The inversions yield orientations of the principal stress axes and shape ratio R .

Numerical tests performed with sets of 25–250 focal mechanisms are presented. The stress tensor is fixed for all datasets. The focal mechanisms are selected to satisfy the Mohr-Coulomb failure criterion (see Fig. 12a, b) and subsequently they are used for the calculation of moment tensors. The moment tensors were contaminated by uniform noise ranging from 0 to $\pm 50\%$ of the norm of the moment tensor (calculated as the maximum of absolute values of the moment tensor eigenvalues). The noisy moment tensors were decomposed back into strikes, dips and rakes of noisy focal mechanisms inverted for stress. The deviation between the true and noisy fault normals and slips attained values from 0° to 25°



Earthquake Mechanisms and Stress Field, Fig. 13 Mean deviation of noisy fault normals as a function of noise superimposed on the moment tensors. The deviation is colour-coded and evaluated in degrees. The noisy slip directions display the same deviations

(see Fig. 13). Since the inverted focal mechanisms were calculated from moment tensors, knowledge of the orientation of the fault planes was lost in the data set. The inversion was run repeatedly using 50 realizations of random noise and for two types of sets of focal mechanisms. The first set consisted of focal mechanisms which projected into both half-planes of the Mohr's circle diagram and thus covered both butterfly wings in the P/T plot (see Fig. 12, left plots). The second set consisted of focal mechanisms projected just into upper half-plane of the Mohr's circle diagram and covered just one butterfly wing in the P/T plot (see Fig. 12, right plots). The inverted principal stress directions and shape ratios for both data sets and all realizations of random noise were compared with the true values and the errors were evaluated (see Figs. 14 and 15).

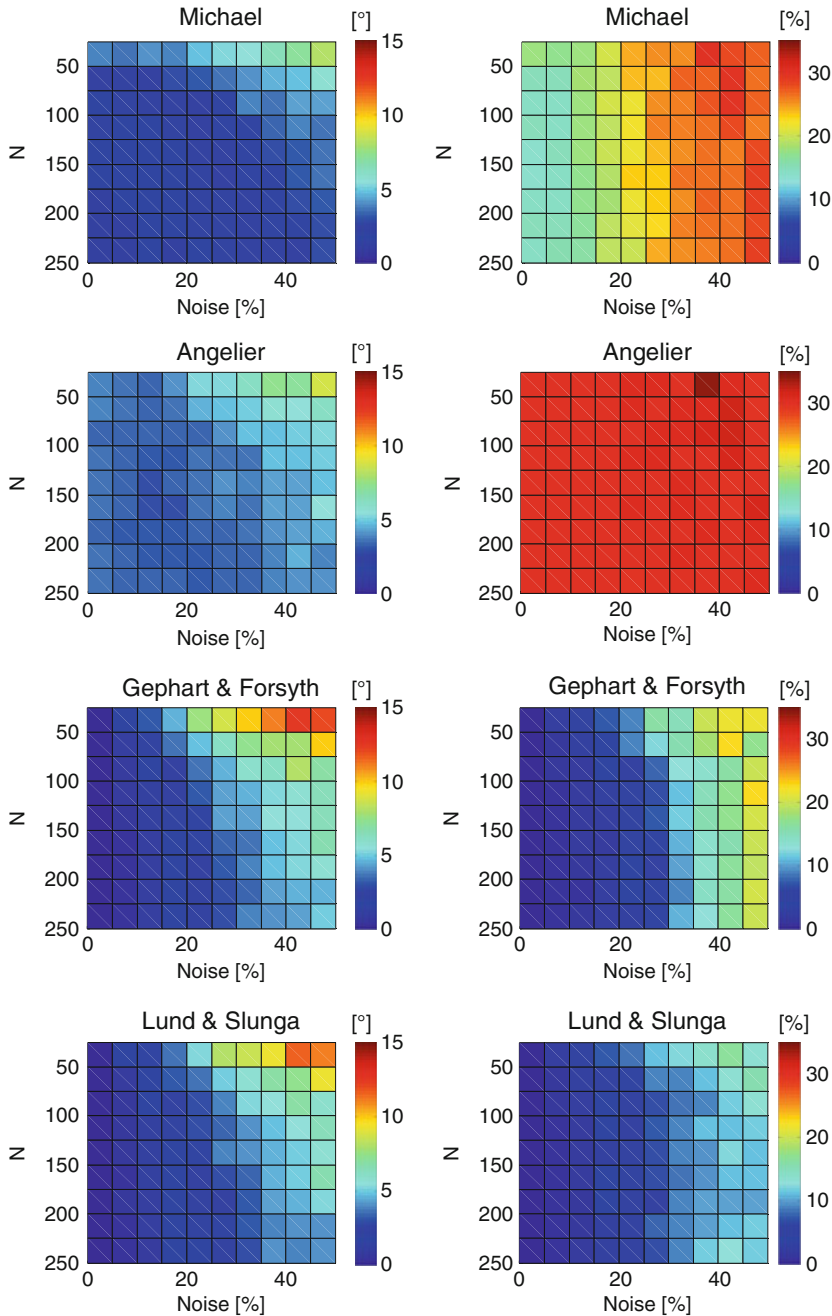
The tests indicate that the accuracy of the stress inversions significantly varies and strongly depends on noise in the data and on the number of focal mechanisms inverted. All inversions yield satisfactory results for the principal stress directions with an average error less than 15° for 25 noisy focal mechanisms inverted (see Fig. 14, left plots). If a reduced variety of focal mechanisms was used (see Fig. 15, left plots), the method of Angelier (2002) produced

systematically worse results than the other methods. The errors occurred even in the case of 250 noise-free focal mechanisms. Slightly biased results for noise-free data were produced also by the method of Michael (1984), because this method is sensitive to the ambiguity in the orientation of fault planes.

In addition, the tests reveal that the shape ratio is a more critical parameter than the orientation of the principal stress directions (see Figs. 14 and 15, right plots). The method of Angelier (2002) completely failed even when the high number of noise-free focal mechanisms of the full variety were inverted (see Fig. 14) yielding the shape ratio of about 0.5 instead of 0.7. Also the method of Michael (1984) works significantly worse than the method of Gephart and Forsyth (1984) and its modification proposed by Lund and Slunga (1999). The latter two methods produce results of similar accuracy, the method of Lund and Slunga (1999) working slightly better and being less sensitive to the low number of focal mechanisms, high noise in the data and the limited variability of focal mechanisms (see Fig. 15, right plots).

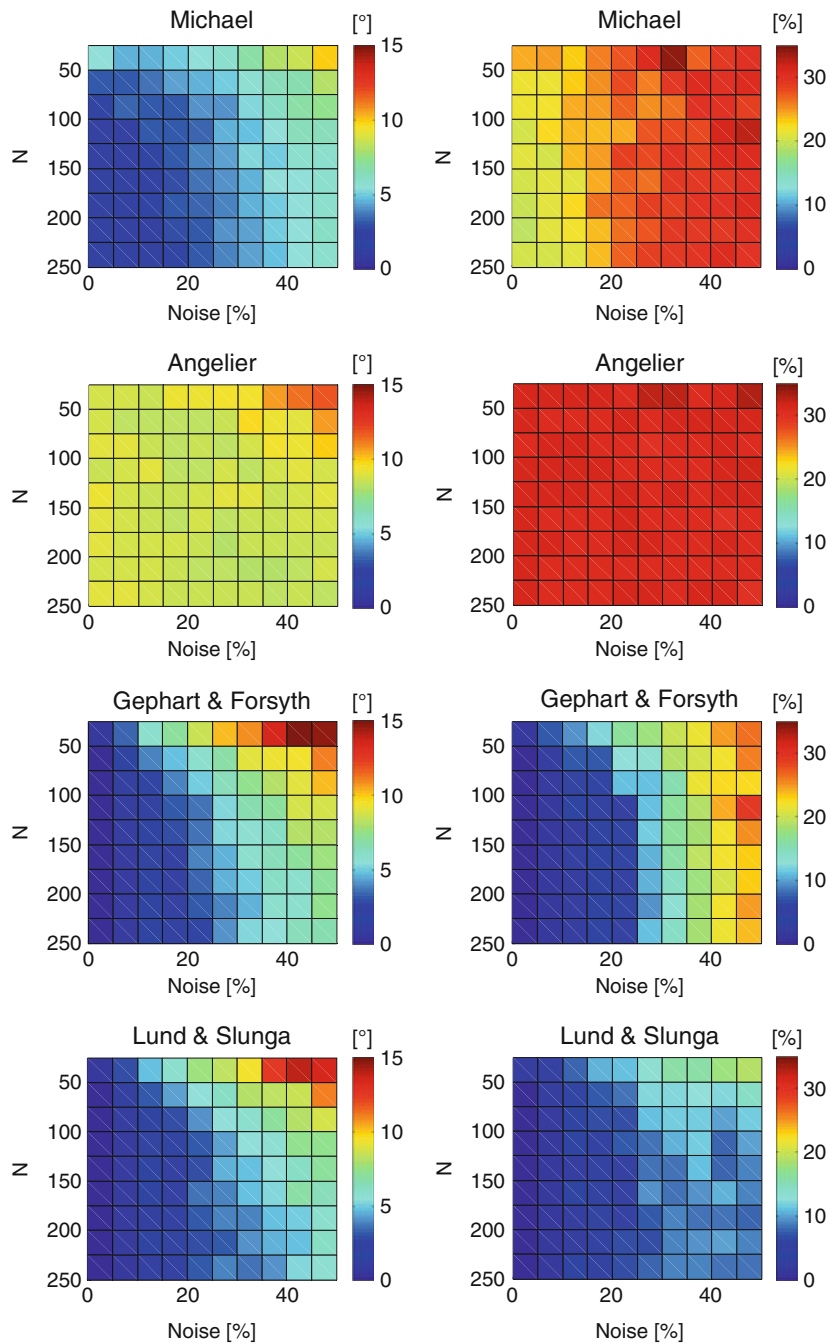
Inversion for Temporal or Spatial Variation of Stress

With an increasing number and accuracy of focal mechanisms of earthquakes occurring in seismically active regions, it becomes possible to study spatial variations of stress or its temporal evolution. The simplest approach how to study the spatial variation of stress is to subdivide the study area into subareas or cells of a similar stress regime and invert for stress of individual cells. Similarly, in case of the time evolution of stress, the total time period of observations can be split into several phases of the seismic activity and stress can be inverted for individual time windows. In order to increase the number of focal mechanisms inverted, the individual cells or time windows can partly overlap. This produces smooth results highlighting their large-scale lateral variations or long-period trends being very analogous to the moving average method.



Earthquake Mechanisms and Stress Field, Fig. 14 Mean error of principal stress directions (*left*) and of the shape ratio (*right*) as a function of the number of inverted focal mechanisms and noise in moment tensors. The data set with the full variety of focal mechanisms (two-wings data) is inverted. The errors are shown for

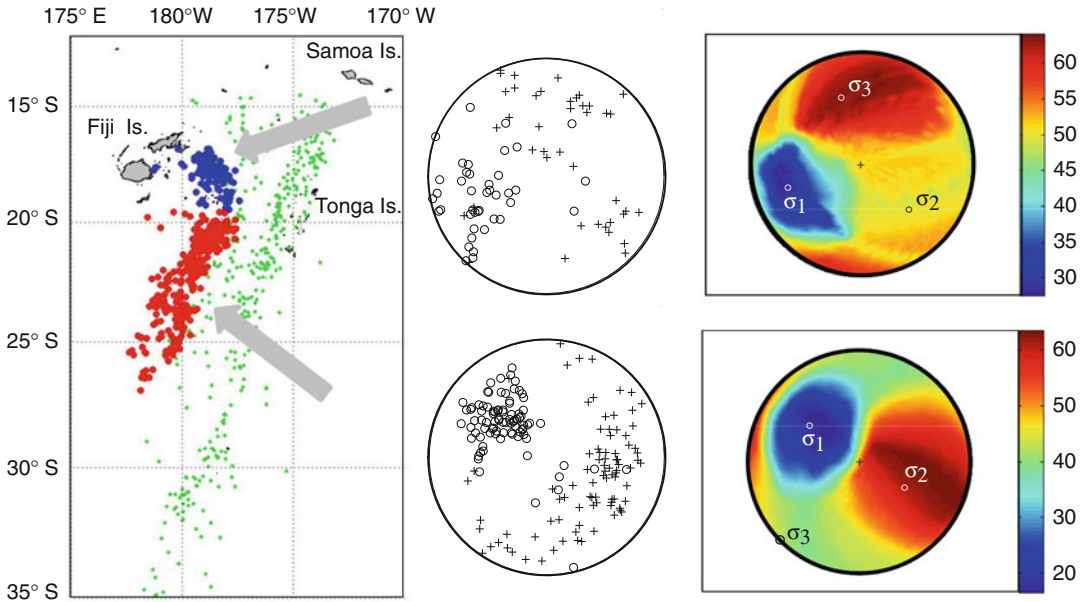
the method of Michael (1984), Angelier (2002), Gephart and Forsyth (1984) and Lund and Slunga (1999). The method of Angelier (2002) employs Eqs. 33 and 34. The errors are calculated from 50 random realizations of noise. The errors in the stress directions are in degrees, the errors in the shape ratio are in percent



Earthquake Mechanisms and Stress Field, Fig. 15 The same as for Fig. 14, but for the data set with the reduced variety of focal mechanisms (one-wing data)

The mentioned stress inversions can be exemplified on several applications. The simplest approach, when the area under study is subdivided just into two subareas of different

stress regimes, was presented, for example, by Vavryčuk (2006) for deep earthquakes in the Tonga subducting slab (see Fig. 16). The slab is considerably deformed at depths between



Earthquake Mechanisms and Stress Field, Fig. 16 Inversion for stress in the Tonga subduction zone. *Left plot* – the map view with epicentres of earthquakes. *Green dots* mark earthquake foci at depths between 100 and 500 km, *blue/red dots* mark deep-focus earthquakes at depths greater than 500 km in the northern/southern cluster. The *grey arrows* show the horizontal projections of the maximum compression in the northern/southern cluster. *Middle plots* – P axes (*circles*) and T axes

(*plus signs*) of focal mechanisms distributed over the focal sphere for earthquakes from the northern (*upper plot*) and southern (*lower plot*) slab segments. *Right plots* – the misfit function for the principal stress axis σ_1 , defined as the average deviation (in degrees) between the predicted shear stress directions and the observed slips at the faults. The Gephart and Forsyth inversion method (1984) was applied. The optimum directions of the principal stresses are marked by *circles* (Modified after Vavříček (2006))

500 and 670 km and can be subdivided into two segments of different orientations characterized by different principal stress directions. The stress inversion shows that the maximum compressive stress is directed along the down-dip motion of the slab in both segments. A more detailed mapping of lateral variations of stress obtained from focal mechanisms was published by Hardebeck (2006) for Southern California who employed the stress inversion code SATSI developed by Hardebeck and Michael (2006). This code inverts simultaneously for stress in all subareas of the region under study and minimizes the difference in stress between adjacent subareas to suppress over fitting of noisy data (see Fig. 17). A modification of this code, called MSATSI and working in the Matlab environment, was applied by Ickrath et al. (2014) to analyse spatial and temporal variations of stress prior and after the 1999 Mw 7.4 Izmit earthquake in Turkey. This analysis

revealed systematic rotations of stress related to the occurrence of the Izmit earthquake. The rotation of the principal stress directions have been observed also after the 1999 Mw 7.6 Chi-Chi earthquake in Taiwan (Wu et al. 2010) and for the 2011 Mw 9 Tohoku-oki earthquake in Japan (Yang et al. 2013). Spatial and temporal variations of stress from focal mechanisms have been studied also for New Zealand (Townend et al. 2012) with focus on the 2010 Mw 6.2 Christchurch earthquake, and for other seismically active regions including geothermal fields in which seismicity is induced by fluid injections (Martínez-Garzón et al. 2013).

Summary

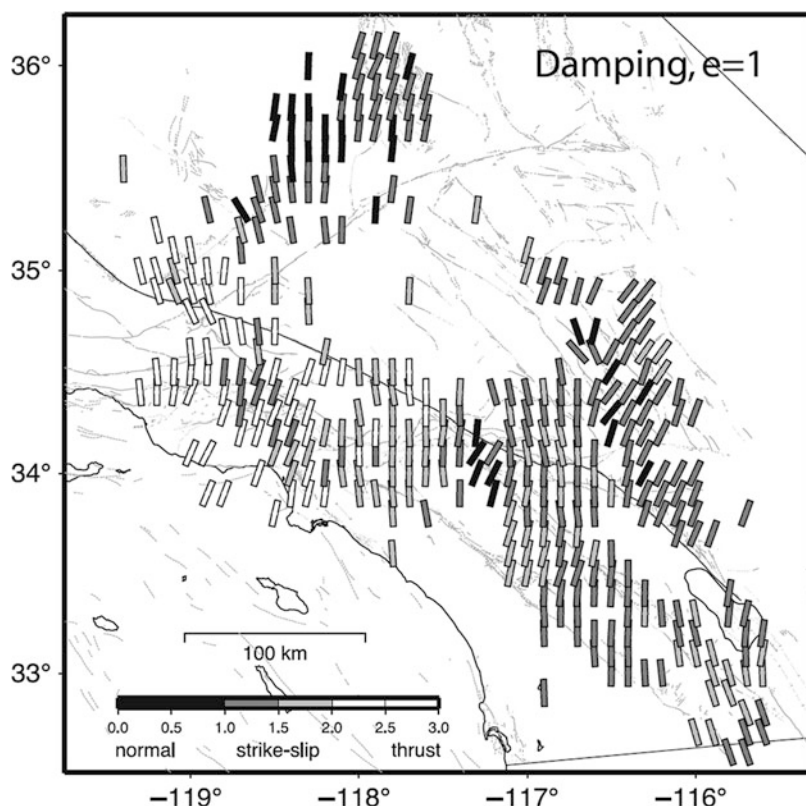
The stress field and fracture processes in the Earth's crust are closely related phenomena.

Earthquake Mechanisms and Stress Field,

Fig. 17 Stress inversion for southern California.

A data set of 6957 focal mechanisms is inverted for the stress tensor at points on a 0.1° grid. The bar orientation shows the direction of the maximum horizontal stress axis, and the shading of the bar indicates the stress regime, following Simpson (1997).

Light grey lines show mapped fault traces, and the black line shows the San Andreas Fault. The damped inversion method of Hardebeck and Michael (2006) is used with damping parameter $e = 1$ (After Hardebeck and Michael (2006), their figure 5c)



Fracturing is controlled by shear stress on the fault, pore-fluid pressure, cohesion and fault friction. In addition, the orientation of the fault with respect to the principal stress directions governs susceptibility of the fault to be ruptured. An essential role in the stress analysis play principal faults – the faults which are optimally oriented for shear faulting and thus being the most unstable under the given stress conditions. The type of faulting, the orientation of the activated faults and the direction of the slip along the faults can be inverted for stress and its spatial and temporal variations within the Earth's crust. The stress inversions work best if the fault orientations are known. If the fault plane cannot be uniquely identified from the focal mechanisms, the inversions are less accurate. The inversions are capable to retrieve four stress parameters: the principal stress directions and the shape ratio. The accuracy of the results depends on the applied inversion method, on the number of inverted focal mechanisms, their errors and variety. The most critical parameter is the shape ratio, which

can be successfully recovered only by the Gephart and Forsyth (1984), Lund and Slunga (1999) and Vavryčuk (2014) methods provided that a high number of accurate focal mechanisms is available.

Cross-References

- [Earthquake Mechanism Description and Inversion](#)
- [Earthquake Mechanisms and Tectonics](#)
- [Non-Double-Couple Earthquakes](#)

References

- Anderson EM (1951) The dynamics of faulting. Oliver & Boyd, Edinburgh
- Angelier J (2002) Inversion of earthquake focal mechanisms to obtain the seismotectonic stress IV – a new method free of choice among nodal lines. *Geophys J Int* 150:588–609
- Bott MHP (1959) The mechanics of oblique slip faulting. *Geol Mag* 96:109–117

- Byerlee J (1978) Fiction of rocks. *Pure Appl Geophys* 116:615–626
- Fischer T, Guest A (2011) Shear and tensile earthquakes caused by fluid injection. *Geophys Res Lett* 38, L05307. doi:10.1029/2010GL045447
- Fowler CMR (1990) *The solid earth. An introduction to global geophysics.* Cambridge University Press, Cambridge
- Gephart JW, Forsyth DW (1984) An improved method for determining the regional stress tensor using earthquake focal mechanism data: application to the San Fernando earthquake sequence. *J Geophys Res* 89:9305–9320
- Hardebeck JL (2006) Homogeneity of small-scale earthquakes faulting, stress, and fault strength. *Bull Seismol Soc Am* 96:1675–1688
- Hardebeck JL, Michael AJ (2006) Damped regional-scale stress inversions: methodology and examples for southern California and the Coalinga aftershock sequence. *J Geophys Res* 111, B11310
- Heidbach O, Tingay M, Barth A, Reinecker J, Kurfeß D, Müller B (2008) The world stress map database release 2008. doi:10.1594/GFZ.WSM.Rel2008
- Ickrath M, Bohnhoff M, Bulut F, Dresen G (2014) Stress rotation and recovery in conjunction with the 1999 Izmit Mw 7.4 earthquake. *Geophys J Int* 196:951–956
- Jaeger JC, Cook NGW, Zimmerman RW (2007) *Fundamentals of rock mechanics*, 4th edn. Blackwell, Oxford
- Julian BR, Miller AD, Foulger GR (1998) Non-double-couple earthquakes. 1. Theory. *Rev Geophys* 36:525–549
- Lay T, Wallace TC (1995) *Modern global seismology.* Academic, New York
- Lund B, Slunga R (1999) Stress tensor inversion using detailed microearthquake information and stability constraints: application to Ölfus in southwest Iceland. *J Geophys Res* 104:14947–14964
- Martínez-Garzón P, Bohnhoff M, Kwiatek G, Dresen G (2013) Stress tensor changes related to fluid injection at The Geysers geothermal field, California. *Geophys Res Lett* 40:2596–2601
- Maury J, Cornet FH, Dorbath L (2013) A review of methods for determining stress fields from earthquake focal mechanisms: application to the Sierentz 1980 seismic crisis (Upper Rhine graben). *Bull Soc Geol Fr* 184(4–5):319–334
- Mavko G, Mukerji T, Dvorkin J (2009) *The rock physics handbook. Tools for seismic analysis in porous media.* 2nd edition. Cambridge University Press, Cambridge
- Michael AJ (1984) Determination of stress from slip data: faults and folds. *J Geophys Res* 89:11517–11526
- Michael AJ (1987) Use of focal mechanisms to determine stress: a control study. *J Geophys Res* 92(B1):357–368
- Ruff LJ (2002) State of stress within the earth. In: Lee WHK, Kanamori H, Jennings PC, Kisslinger C (eds) *International handbook of earthquake and engineering seismology. Part A.* Academic, New York
- Scholz CH (2002) *The mechanics of earthquakes and faulting.* Cambridge University Press, Cambridge
- Simpson RW (1997) Quantifying Anderson's fault types. *J Geophys Res* 102:17909–17919
- Townend J, Sherburn S, Arnold R, Boese C, Woods L (2012) Three-dimensional variations in present-day tectonic stress along the Australia–Pacific plate boundary in New Zealand. *Earth Planet Sci Lett* 353–354:47–59
- Vavryčuk V (2006) Spatially dependent seismic anisotropy in the Tonga subduction zone: a possible contributor to the complexity of deep earthquakes. *Phys Earth Planet Inter* 155:63–72. doi:10.1016/j.pepi.2005.10.005
- Vavryčuk V (2011a) Principal earthquakes: theory and observations from the 2008 West Bohemia swarm. *Earth Planet Sci Lett* 305:290–296. doi:10.1016/j.epsl.2011.03.002
- Vavryčuk V (2011b) Tensile earthquakes: theory, modeling, and inversion. *J Geophys Res* 116, B12320. doi:10.1029/2011JB008770
- Vavryčuk V (2014) Iterative joint inversion for stress and fault orientations from focal mechanisms. *Geophys J Int* 199(1):69–77. doi:10.1093/gji/ggu224
- Vavryčuk V, Bouchaala F, Fischer T (2013) High-resolution fault image from accurate locations and focal mechanisms of the 2008 swarm earthquakes in West Bohemia, Czech Republic. *Tectonophysics* 590:189–195. doi:10.1016/j.tecto.2013.01.025
- Wallace RE (1951) Geometry of shearing stress and relation to faulting. *J Geol* 59:118–130. doi:10.1086/625831
- Wu Y-M, Hsu Y-J, Chang C-H, Teng LS, Nakamura M (2010) Temporal and spatial variation of stress field in Taiwan from 1991 to 2007: insights from comprehensive first motion focal mechanism catalog. *Earth Planet Sci Lett* 298:306–3016
- Yang Y-R, Johnson KM, Chuang RY (2013) Inversion for deviatoric absolute crustal stress using focal mechanisms and coseismic stress changes: The 2001 M9 Tohoku-oki, Japan, earthquake. *J Geophys Res* 118:5516–5529
- Zang A, Stephansson O (2010) *Stress field of the earth's crust.* Springer, New York
- Zoback MD (2007) *Reservoir geomechanics.* Cambridge University Press, Cambridge

Earthquake Mechanisms and Tectonics

Jochen Braunmiller

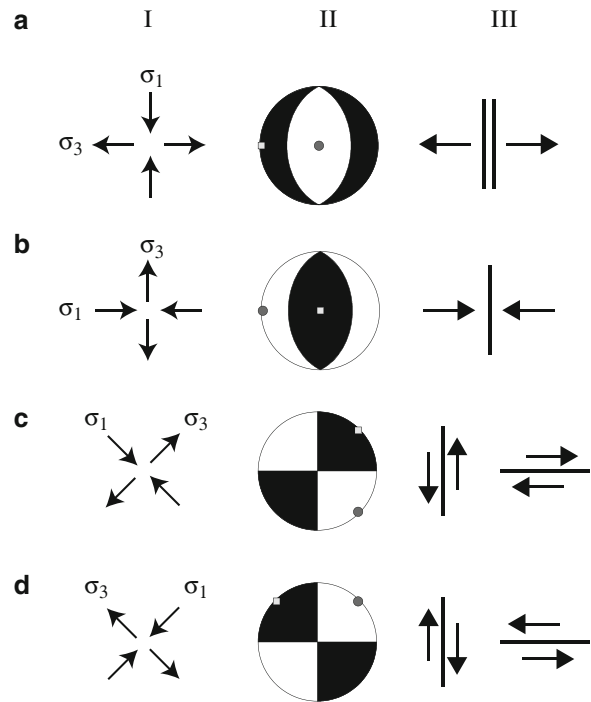
University of South Florida, Tampa, FL, USA

Synonyms

Earthquake source mechanisms; Fault plane solutions; Focal mechanisms; Plate tectonics; Seismic moment tensors; Seismicity and tectonics

Earthquake Mechanisms and Tectonics,

Fig. 1 Examples of basic earthquake focal mechanisms. Rows: (a) normal faulting, (b) thrust faulting, (c and d) strike-slip faulting. Column I: principal (compression and tension) axes orientation, cross-sectional view (rows a and b) and map view (rows c and d). Column II: lower hemisphere focal mechanisms with compressional quadrants shaded, square is T-axis and circle is P-axis. Column III: arrows show relative motion direction relative to fault (heavy line in center, double line for spreading)



Introduction

Seismology has played a key role in the development of the plate tectonic theory, which describes and explains the motions of lithosphere plates. The global distribution of earthquakes outlines the plate boundaries, the occurrence of deep earthquakes indicates regions where plates are recycled back into Earth's mantle, and earthquake fault plane solutions provide information about the relative sense of motion at plate boundaries.

Three different types of plate boundaries exist: convergent, divergent, and transform. Relative motion of two plates toward each other occurs at convergent, away from each other at divergent, and parallel to the boundary at transform boundaries (Fig. 1). Convergent plate boundaries are subduction zones where oceanic lithosphere of one plate dives beneath another plate. The overriding plate can be either oceanic lithosphere, such as in the Philippine Sea, or continental lithosphere, such as in South America. Continental collision zones are another form of convergent boundary. Continental lithosphere material, in

particular continental crust, is not dense enough to sink into the asthenosphere and where two continents collide, such as in the Himalayas where the Indian plate is actively colliding with the Eurasian plate, compressional mountain belts are formed. Divergent plate boundaries in the ocean are mid-ocean ridge spreading centers where new oceanic lithosphere is generated. The Mid-Atlantic Ridge and the East Pacific Rise are examples. Mature divergent plate boundaries are always associated with seafloor spreading; divergent plate boundaries in continents, such as the East African Rift System, indicate rift initiation, young or failed rifting. At transform plate boundaries, new material is neither created nor destructed, rather two plates move sideways (laterally) past each other. The motion sense along a transform fault can be either right lateral (dextral) or left lateral (sinistral). The best-known transform plate boundaries occur on continents, e.g., the San Andreas Fault in California where the North American and Pacific plates slide past each other in a right-lateral sense. Transform faults are also abundant in oceanic lithosphere where they often connect two spreading segments.

Most earthquakes occur at the boundary between two tectonic plates. These are interplate earthquakes. Fewer earthquakes occur inside plates, and these are intraplate earthquakes.

The depth of earthquakes within the lithosphere is usually restricted to the shallow part of the lithosphere that deforms in a brittle fashion. In the continental crust, the depth cutoff is often associated with the 350 °C isotherm which is reached at mid-crustal depth; the actual cutoff depth (between about 10 and 25 km) depends on the temperature gradient. At temperatures above 350 °C, the bulk material of continental crust deforms in a ductile, plastic fashion. In the oceanic lithosphere, due to different mineral composition, ductile deformation sets in above about 600 °C, which, in oceanic lithosphere, is usually reached in the uppermost mantle (the depth depends on lithosphere cooling and can be estimated using a simply half-space cooling model). Some examples exist where earthquakes occur in the lower crust or uppermost mantle of continental lithosphere, such as in parts of the East African Rift System and of the Tibetan Plateau. Some researchers suggest that these events are located in the uppermost mantle, which, at about the same depth and temperature, is stronger than crustal material, while others argue that under certain circumstances earthquakes can occur throughout the continental crust; this is a topic of active research. Deep earthquakes within the oceanic lithosphere of subducting plates are another exception; they are discussed below.

The source mechanisms (also called fault plane solutions and focal mechanisms or, using a more general description, seismic moment tensors) of tectonic earthquakes are generally due to shear faulting on a fault plane. In the moment tensor description, this is a pure double-couple source, i.e., no volume change or other non-shear components (such as a compensated linear vector dipole) are present at the source. The earthquake mechanisms that correspond to extension, compression, and lateral motion are normal, thrust (reverse), and strike-slip fault plane solutions; basic examples are shown in Fig. 1. Normal faults strike perpendicular to the extension direction, and fault dips commonly found from the analysis

of earthquake data are in the 30°–60° range. Thrust (reverse) faults strike perpendicular to the convergence directions and often have a shallow-dip angle (<30°); this is typical for megathrust subduction zone earthquakes. Transform faults are oriented parallel to the relative motion direction and have right- or left-lateral strike-slip mechanisms with near-vertical fault planes. Earthquake slip vectors describe the relative motion of two fault surfaces relative to each other. They have been used extensively in the past to estimate relative plate motion directions; however, we know now that spreading directions and slip on subduction zone thrust earthquakes are not necessarily parallel to plate motion directions.

Earthquake fault plane solutions provide an overall, average description of the seismic source effectively treating it as a point source. Effects of slip on a finite fault or complex ruptures involving several faults or fault segments are observed seismically and in the field. However, these topics are not addressed in this entry.

Plate boundaries in oceanic lithosphere are often narrower and simpler than in continental lithosphere. They will be discussed first with Fig. 2 showing a simple conceptual model.

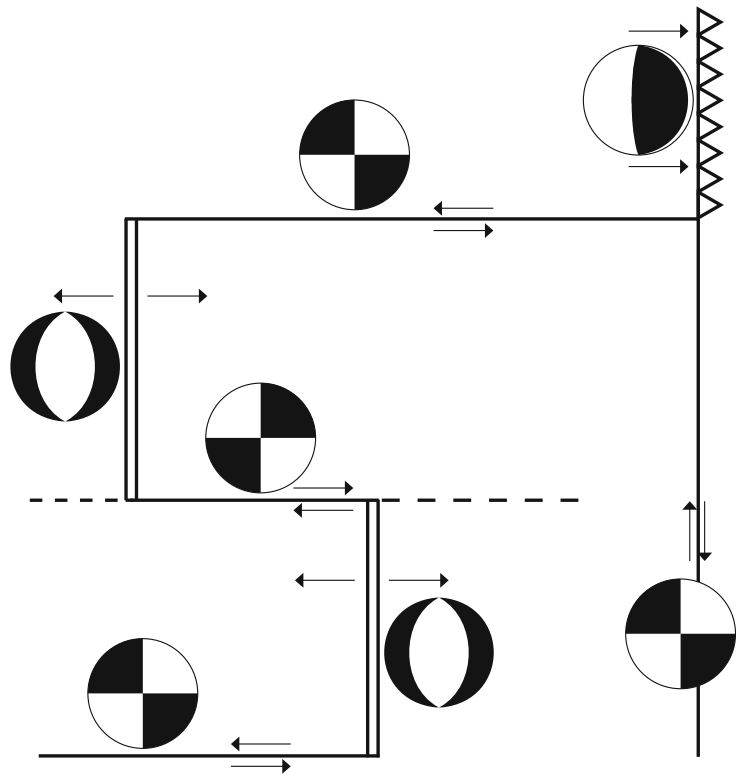
Mid-Ocean Ridges and Oceanic Transform Faults

New oceanic lithosphere at mid-ocean ridge spreading centers is due to extrusive and intrusive volcanism. The extensional nature of spreading is reflected in the observed normal-faulting earthquake mechanisms. Figure 3 shows the global distribution of spreading ridges and the normal-faulting mechanisms of larger (moment magnitude M_w about 5 and larger) earthquakes along them. The mechanisms have nodal planes parallel to the ridge axes.

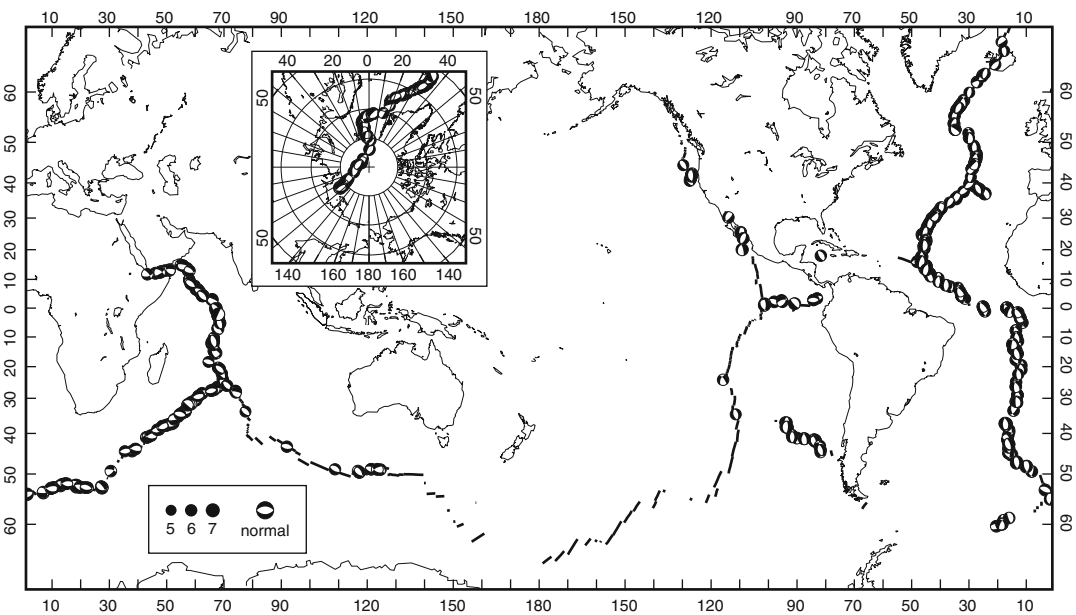
By design, Fig. 3 shows the distribution of $M \geq 5$ earthquakes along the global ridge system. Most of these events occur along ridges with slow (≤ 55 mm/year) spreading rates. Few normal-faulting earthquakes on ridges exceed $M_w = 6$ and the largest event, which occurred

Earthquake Mechanisms and Tectonics,

Fig. 2 Earthquake mechanisms in oceanic environments. Earthquakes occur mainly on active faults (*double line*: spreading center; *single line*: transform fault; *barbed line*: subduction zone, barbs point in direction of subducting plate). *Dashed line* is an inactive fracture zone. *Arrows* show relative motion sense



E



Earthquake Mechanisms and Tectonics, Fig. 3 Global distribution of spreading ridges (*heavy lines*) and oceanic normal-faulting focal mechanisms (beach balls) for 1977–1998 from the Global CMT

project. Note lack of earthquakes along the fast-spreading ridges (e.g., East Pacific Rise) compared to slow-spreading ridges (e.g., Mid-Atlantic) (From Bird et al. 2002) from AGU publications

Earthquake Mechanisms and Tectonics, Table 1 Some examples of significant earthquakes for different tectonic environments with source mechanism type and size (moment magnitude). Type: NF normal faulting, TF thrust faulting, and SS strike-slip faulting. Events are largest (or among largest) of a given type since about 1900. Name gives source region. Ref refers to literature (for further details on a given earthquake). Unless noted otherwise, all earthquake sizes (M_w) in this entry are from the GCMT or USGS sources

Environment	Date	Type	Size (M_w)	Name	References
Mid-ocean ridge	21 Mar 1988	NF	6.4 ^{gcmt}	Gakkel Ridge	
Ridge-to-ridge oceanic transform	14 Mar 1994	SS	7.1 ^{gcmt}	Romanche Transform	1
Megathrust subduction	22 May 1960	TF	9.5 ^{usgs}	Chile	2
Outer rise (subduction) – intraplate	29 Sep 2009	NF	8.1 ^{usgs}	Samoa Islands	3
Overriding plate (subduction) – intraplate	25 Apr 1992	TF	7.2 ^{gcmt}	Petrolia, California	4
Intermediate deep (subduction) – intraplate	9 Dec 2007	TF	7.9 ^{gcmt}	Fiji-Tonga	
Deep (subduction) – intraplate	24 May 2013	NF	8.4 ^{gcmt}	Sea of Okhotsk	5
Continental – collision	15 Jan 1934	TF?	8.1 ^{usgs}	Bihar-Nepal	6
Continental – extension	28 Mar 1970	NF	7.1 ⁷	Gediz, Turkey	7
Continental – strike-slip	14 Nov 2001	SS	7.8 ^{gcmt}	Kunlunshan, Tibet	8
Continental – intraplate	26 Jan 2001	TF	7.6 ^{gcmt}	Gujarat, India	9
Oceanic – intraplate	11 Apr 2012	SS	8.6 ^{usgs}	Indian Ocean	10

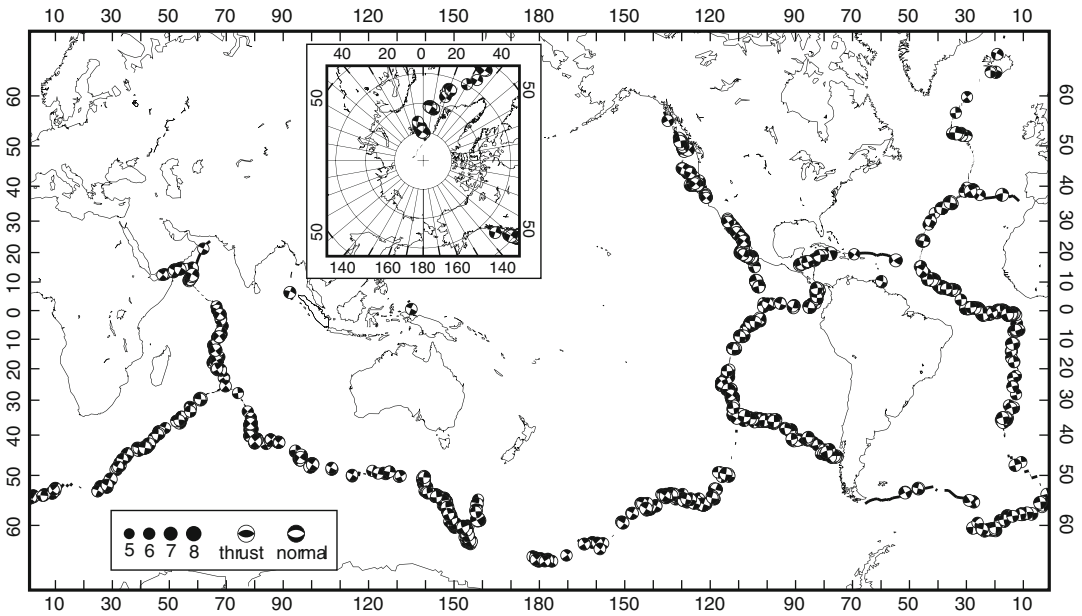
gcmt size from Global Centroid Moment Tensor (GCMT 2014) project, *usgs* size from the U.S. Geological Survey (USGS 2014), 1 Abercrombie and Ekström (2003), 2 Kanamori and Cipar (1974), 3 Lay et al. (2010), 4 Oppenheimer et al. (1993), 5 Ye et al. (2013), 6 Bilham et al. (2001), 7 Braunmiller and Nabelek (1996), 8 Antolik et al. (2004), 9 Antolik and Dreger (2003), 10 Meng et al. (2012)

along the ultraslow (≤ 20 mm/year) Gakkel Ridge separating the North American and Eurasian plates in the Arctic region, reached only $M_w = 6.4$ (Table 1). Maximum earthquake size and the frequency of large earthquake occurrence are related to spreading rate. Note the continuous band of earthquake mechanisms along the slow-spreading Mid-Atlantic and Central and Southwest Indian ridges and the lack of events along the fast-spreading (here set as 80 mm/year and faster) ridges south of Australia and the East Pacific Rise. Along the East Pacific Rise, normal-faulting earthquakes rarely exceed $M = 5$. Differences between slow- and fast-spreading ridges are due to the thermal structure of the lithosphere.

Microearthquake activity along the entire ridge system – particularly along fast-spreading ridges – is probably vigorous based on observations from still relatively few, temporary seismic observations from the ocean floor. Even after considering smaller earthquakes, the seismic moment released by all earthquakes along the

ridges accounts for only a small percentage of the relative plate motions indicating the importance of aseismic slip.

The mid-ocean ridge segments are connected to each other by oceanic transform faults. Ridge-transform intersections are often nearly perpendicular. Motion along transforms is lateral and parallel to relative plate motion directions. Transforms offset ridge segments, and the offset direction (i.e., whether the next segment is to the right or left) determines motion sense and thus whether a fault is right lateral or left lateral (Fig. 2). The offset-parallel nodal plane is usually the fault plane, and both nodal planes are usually near vertical. Parts of a transform fault not between two ridge segments become seismically inactive and are called fracture zone. The Mendocino fracture zone, extending halfway across the northeast Pacific Ocean, is a famous example; its active transform fault part is only about 200 km long and connects a spreading ridge with the Mendocino triple junction in northern California.



Earthquake Mechanisms and Tectonics, Fig. 4 Global distribution of oceanic transform faults (heavy lines) and oceanic strike-slip focal mechanisms (beach balls) for 1977–1998 from the Global CMT

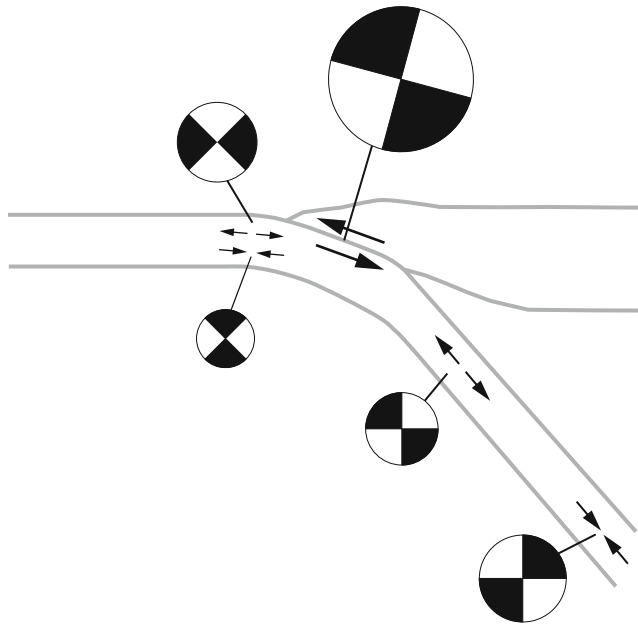
project. Note earthquakes occur along almost all transforms independent of plate motion rate (From Bird et al. 2002) from AGU publications

The global distribution of oceanic transform faults and oceanic transform earthquakes ($M_w \geq 5$) shows these strike-slip earthquakes occur along all oceanic transform faults independent of motion rate (Fig. 4). However, the size of the largest earthquake along an oceanic transform fault decreases with increase in plate motion rate. This is due to the thermal structure of the lithosphere. The thin seismogenic part of the lithosphere (above 600 °C) limits the maximum fault area and thus the maximum size to about $M_w = 7-7.5$. Some of the largest ridge-to-ridge strike-slip earthquakes have been observed along the Romanche transform in the equatorial Mid-Atlantic (Table 1). Global studies indicate that earthquakes account for only part of the predicted plate motions along most oceanic transform faults, requiring, as along spreading ridges, significant aseismic slip. How the seismic slip deficit is accommodated, whether as fully seismic patches imbedded in creeping fault patches or as faults that alternate between seismic and aseismic slip, is a matter of current research.

Subduction Zones

The largest earthquakes on Earth are shallow-depth megathrust earthquakes at plate interfaces of subduction zones. The complex mechanics of subduction, though, generates strong earthquakes also away from the plate interface: (i) As the down-going plate approaches a subduction trench, the plate has to bend from a horizontal to an inclined direction generating bending stresses. (ii) Earthquakes occur within the overriding plate near plate interfaces. (iii) Within the cold subducting slab, earthquakes can occur at greater depth than elsewhere. Cases (i) to (iii) are intraplate earthquakes that occur within the overriding or the subducting plate (cf. interplate or plate boundary earthquakes). Figure 5 illustrates earthquake types observed in subduction zones, and Table 1 lists an example for each type. Note that not all types are observed at each subduction zone.

Plate interface earthquakes have thrust mechanisms indicating underthrusting of the



Earthquake Mechanisms and Tectonics, Fig. 5 Schematic overview of some earthquake types observed in subduction zones. Not all event types occur at each subduction zone. Focal mechanisms are in cross-sectional view and are roughly scaled according to maximum observed size. *Arrows* show relative motion sense

for interplate megathrust earthquake and state of stress for intraplate earthquakes (*diverging arrows*, extension; *converging arrows*, compression). Intermediate-deep event is an example for in-plane extension; some plates show in-plane compression throughout the slab (then intermediate and deep mechanisms are the same)

subducting plate beneath the overriding plate along a shallow-dipping fault. They occur landward from the trench (or, in general, beneath the leading edge of the overriding plate). The mechanisms have nodal planes parallel to the strike of the trench (and coast), and slip vectors, which indicate the relative motion sense on the fault, are often subparallel to the relative plate motion directions.

The largest earthquake measured with seismometers is the 1960 Chile ($M_w = 9.5$) subduction zone earthquake. The rupture dimensions corresponding to an earthquake of this size are about 1,000 km rupture length, 200 km down-dip rupture width, and 10–30 m average slip on the entire area that ruptured. Additional infamous examples include the 1964 Alaska ($M_w = 9.2$), the 2004 Sumatra-Andaman ($M_w = 9.1$), and the 2011 Tohoku, Japan ($M_w = 9.0$), earthquakes. Such gigantic earthquakes can be very devastating, and they are often accompanied by large, deadly tsunamis that travel across ocean basins

to cause destruction far away from the earthquake rupture area itself.

As an oceanic plate starts to descend beneath another plate, it has to bend to accommodate the change in plate motion from horizontal to an inclined direction. The bending stresses are largest seaward of the trench at the outer rise and the outer trench slope. The stresses due to bending and slab-pull forces cause earthquakes. Plate bending in the upper part of a subducting lithosphere plate is extensional resulting in normal faulting; such earthquakes are observed at many subduction zones. Bending stresses in the lower part are compressional resulting in thrust faulting; such earthquakes occur less frequently. The depth of transition from extension to compression within a plate is an indicator for the stress state within oceanic lithosphere. T-axes of normal faulting and P-axes of thrust earthquakes are usually pointing in the direction of plate motion with nodal planes subparallel to the strike of the trench system.

The largest normal-faulting earthquakes observed globally are the plate-bending earthquakes seaward of trenches, and they often generate tsunamis. Examples include the 1933 Sanriku, Japan ($M_w = 8.3$), the 1977 Sumbawa, Indonesia ($M_w = 8.3$), and the 2009 Samoa Island ($M_w = 8.1$) events. In several instances modeling of seismologic data has shown that outer rise earthquakes can occur as “doublets” with a shallow normal-faulting sub-event and an almost coincident deeper thrust sub-event reflecting stress changes transmitted through the plate. The December 2012 Honshu, Japan, earthquake seaward of the great 2011 Tohoku earthquake is an example.

Seismicity landward of a trench within the overriding plate is common. The overall tectonic environment is compressional, but within these regions of complex deformation, a variety of earthquake source mechanisms may be observed. Most earthquakes are relatively small because the deforming edge of a plate usually cannot sustain large strains. A potential example for a larger earthquake within the overriding plate is the April 1992 Petrolia earthquake where the main $M_w = 7.2$ thrust earthquake ruptured within the North American plate and above the subducting Gorda plate.

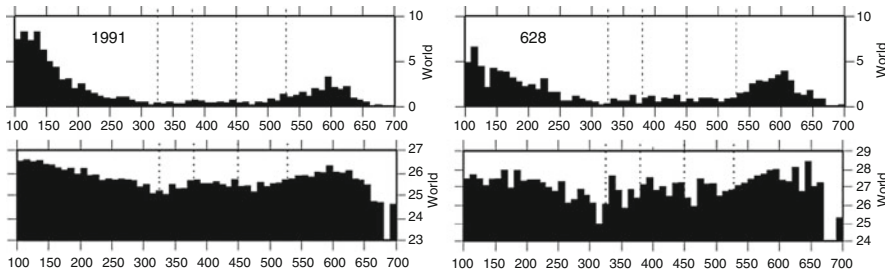
Earthquakes in the subducting slab have been recorded down to about 700 km depth as the cold lithospheric slab penetrates into the mantle and is slowly heated. Temperatures within the slab are significantly lower than in the surrounding mantle, which allows earthquakes to occur within the colder core of the slab at greater depth than anywhere else on Earth.

The mechanisms for generating earthquakes at depths of more than 100 km are not fully understood and are the topic of current research. The observed earthquake source parameters are usually consistent with pure shear dislocations (or fault slip as for shallow earthquakes). However, the normal stresses due to the high confining pressures at depth should actually prohibit shear faulting. Various mechanisms have been presented trying to explain the existence of intermediate (70–300 km depth) and deep (300–700 km depth) earthquakes including

sudden phase changes, breakdown of hydrous minerals, lubrication of fault surfaces by melting of rock, or a combination of these mechanisms.

The earthquake fault plane solutions for deep earthquakes usually show either down-dip extension (T-axis parallel to slab-dip direction) or down-dip compression (P-axis parallel to slab-dip direction). It is thought that deep (300–700 km depth) earthquakes are under compression because the slab is penetrating into stronger material beneath the asthenosphere, which leads to increased subduction resistance. The state of stress for intermediate-deep events (70–300 km depth), however, varies for different subduction zones and can be either under compression or tension. Down-dip compression at intermediate depth could occur if compressive stresses are transmitted from deeper parts of the slab as it encounters strong subduction resistance. Down-dip tension is presumably the result of a slab extending under its own weight as it sinks into less dense surrounding mantle when its leading edge is not experiencing strong subduction resistance. Examples for the former (compressive stresses throughout down-going slab) might be the Tonga and North Honshu subduction zones, while the Kuriles, Chile, and Sunda subduction zones might represent slabs where down-dip tension at intermediate depth and compressive stress at depths of more than 300 km are observed.

The global depth distribution of deep earthquakes (Fig. 6) shows a strong decrease of activity to about 300 km depth, a broad, weak maximum at about 400 km depth, and a significant increase in seismic activity below about 530 km (down to the maximum observed depth of about 700 km). Variations to this general pattern between subduction zones are likely due to differences in slab thermal structure, which depends on relative plate motion rate, age of the subducting plate, and whether the slab is continuous or not. Intermediate-deep earthquakes (70–300 km depth) reach about $M_w = 8$ (Table 1). The largest deep earthquakes occur close to the base of the seismically active part of the slab below 600 km depth and can exceed $M_w = 8$. Examples include the 1994 Bolivia ($M_w = 8.2$) and the 2013 Sea of Okhotsk ($M_w = 8.4$) earthquakes (Table 1).



Earthquake Mechanisms and Tectonics, Fig. 6 Depth versus frequency of earthquake occurrence (*top*) and cumulative seismic moment release (*bottom*) for earthquakes below 100 km depth binned in 10 km steps. See reference for data source. *Left* earthquakes with $M_w < 5.75$ and *right* larger events. Numbers in the *top* are total number of earthquakes shown; vertical axis is

percent of total. Note rapid decay of activity from 150 km to 300 km depth, a small local maxima of events near 400 km, and strong increase below about 530 km depth. Scale at the *bottom* is in units of $\log(M_0)$ dyne cm (subtract 7 for units of [Nm]) (From Estabrook 2004) from AGU publications

Earthquakes in Continental Plate Boundary Settings

Plate boundaries in continental environments are generally more complex and much broader than their oceanic counterparts. This is likely due to differences in crustal thickness, composition, density, and mechanical properties of continental relative to oceanic lithosphere. For example, the low density of continental crust prevents it from subducting and causes mountain building when two continental lithosphere plates collide such as the India-Eurasia collision, which is forming the Himalayan Mountains.

Continental plate boundary zones include convergent, divergent, and transform boundaries. The earthquake focal mechanisms within each zone follow their oceanic counterparts with thrust faulting in convergent, normal faulting in divergent, and strike-slip faulting in transform settings.

The Alpine-Himalayan belt is an example of a large-scale zone of (primarily) continental collision that stretches from the Gibraltar Arc in the west to Southeast Asia in the east. Along this belt, the Eurasian plate collides with the African (Nubian), Arabian, and Indian plates. The seismic deformation is released primarily by thrust-faulting earthquakes with slip vectors that are usually subparallel to the plate motion direction. Large, destructive thrust-faulting earthquakes, for example, are observed in the Caucasus

Mountains (such as the 1988 Spitak, Armenia, earthquake) and within a wide fold-and-thrust belt of the Zagros Mountains, Iran. Underthrusting of the Indian plate beneath Eurasia along the Himalayan front has generated some great (magnitude $M \geq 8$) earthquakes such as the 1934 Bihar-Nepal, India, earthquake ($M_w = 8.1$, Table 1), and possibly even larger earthquakes may have occurred in historic times.

Continental lithosphere is less dense than upper mantle and does not subduct resulting in the absence of deep earthquakes in continental collision zones. However, intermediate-deep earthquakes in the Hindu Kush region and the Vrancea zone of Romania (such as the destructive 1977 Bucharest earthquake) as well as infrequent deep earthquakes beneath southern Spain (such as the 2010 $M_w = 6.4$ earthquake) occur within the Alpine-Himalayan belt. These earthquakes are the result of past subduction of oceanic lithosphere, and their mechanisms are as expected in subduction zones (i.e., P- or T-axis subparallel to presumed slab-dip direction where known).

A consequence of the absence of a subducting slab in continent-continent collision, in addition to mountain building, is the lateral motion or “escape” of large crustal blocks. This is observed, for example, in the India-Eurasia collision as eastward escape of the over-thickened Tibetan Plateau. Deformation is accommodated along left-lateral strike-slip faults (e.g., Altyn Tagh and Kunlun faults) at the northern boundary of

the plateau and as east–west-oriented extension along north–south trending grabens internal to the plateau. Examples for large earthquakes include the 2001 $M_w = 7.8$ strike-slip event on the Kunlun Fault in northern Tibet (Table 1) and the 2008 $M_w = 7.1$ event in northern Tibet, which is one of the largest continental normal-faulting earthquakes in recent years. As crustal material is extruded toward east, it encounters resistance from the stable South China Block resulting in compression at the eastern edge of the Tibetan Plateau that is released in earthquakes. An example is the destructive 2008 $M_w = 7.9$ Sichuan, China, earthquake that occurred on the northeast trending Longmenshan thrust fault.

A similar situation is encountered at the Arabia-Eurasia collision zone where the Anatolian Block, bounded by the left-lateral East Anatolian Fault at the southeast and the right-lateral North Anatolian Fault at the north, is escaping westward as the Arabian plate bulldozes into Eurasia generating the Caucasus Mountains. The destructive 1999 $M_w = 7.4$ Izmit, Turkey, earthquake is an example of activity on the North Anatolian Fault. The Anatolian Block has room to escape westward because the Aegean Sea is a weak region of active back-arc extension behind the southward retreating Hellenic subduction zone. Extension in the Aegean Sea and in westernmost Anatolia is subparallel to the convergence direction, which for the most part is roughly north–south oriented. The 1970 $M_w = 7.1$ Gediz, Turkey, earthquake is a normal-faulting example (Table 1). The broad region of extension in the Basin and Range province of the western United States perhaps represents similar tectonics landwards (east) of the North American plate boundary.

The East African Rift System is an example for a large-scale zone of continental extension where two lithospheric plates slowly spread apart (at about 5 mm/year). The earthquake distribution is broader than at mid-ocean ridges, and the rift actually splits into an eastern and western branch when it encounters the relatively strong and deep-rooted Tanzanian Craton. Earthquakes have normal-faulting mechanisms with strikes parallel to the direction of the East African Rift

as expected. Unusually deep normal-faulting earthquakes, either in the lowermost crust or in the uppermost mantle, has been observed in some parts where the rift is proximal to the cold craton; their origin and exact location relative to the crust-mantle boundary are research topics. Infrequent but rather strong normal-faulting earthquakes seem to occur in the least developed and most slowly spreading southern part of the rift such as the 2006 $M_w = 7.0$ Mozambique earthquake. The rift could eventually develop into an oceanic spreading center and produce new oceanic crust. The geologic record, however, shows that not all continental rifts evolve into oceanic spreading centers and may simply die such as the failed Midcontinental Rift System in the south-central part of the North American plate.

The Pacific-North America plate boundary in the western United States is an example of a continental transform plate boundary. The right-lateral San Andreas Fault takes up most relative plate motion, while the remainder is distributed over faults subparallel to the San Andreas Fault within a wide plate boundary zone. The boundary is wider than observed in oceanic settings. The largest historic earthquakes along the San Andreas Fault are the 1906 $M_w = 7.8$ San Francisco and the 1857 $M_w = 7.9$ Fort Tejon earthquakes rupturing the northern and the central-southern segment of the fault, respectively.

Intraplate Earthquakes

Most earthquakes occur at (narrow or broad) plate boundaries. However, earthquakes are observed to occur within plates far from any boundary. These earthquakes show the limits of plate rigidity and indicate plate internal deformation. Intraplate stresses causing the earthquakes are thought to be transmitted from the plate boundaries or to originate from addition (mountains) or removal (melting glaciers) of internal loads accompanied by the viscoelastic response of the lithosphere. It is assumed that intraplate earthquakes mainly reactivate preexisting faults or occur in zones of preexisting crustal weakness.

Away from plate boundaries, loading (stressing) rates are low and intraplate earthquakes consequently should occur less frequently than plate boundaries events – their recurrence intervals should be longer.

Seismicity in the New Madrid, Missouri, area seems to utilize zones of weakness that were originally generated by the Precambrian failed Midcontinent Rift System. Three $M_w \geq 7.5$ earthquakes occurred in the area in 1811–1812. The source mechanisms are not directly known but probably involved strike-slip and thrust motions on several faults.

Source mechanisms of intraplate earthquakes can be normal, thrust, or strike-slip depending on the stresses inside the lithosphere plate. As such, intraplate source mechanisms are indicators of the stress field, and their analysis is one of the main tools available to estimate stresses in the crust. In many regions, such as North America, internal stresses are compressional, and most earthquakes have thrust and strike-slip source mechanisms. The 2001 $M_w = 7.6$ Gujarat, India, earthquake (Table 1) is an example of intraplate thrust faulting that is probably a consequence of collisional plate boundary forces that are exerted several hundred kilometers away from the epicenter. A counterexample is the European Cenozoic Rift System in the foreland of the Alps that formed in response to the Alpine orogeny; most earthquakes north of the Alps have either normal or strike-slip mechanisms. Intraplate regions pose a challenge to earthquake preparedness because large, potentially destructive earthquakes occur very infrequently (on the order of thousand years or more) but could possibly occur almost anywhere within a plate.

Intraplate earthquakes also occur in oceanic plates. In many cases these earthquakes are located along the inactive fracture zones of transform fault systems or relatively close to the oceanic ridge-transform system. The former are possibly caused by stresses due to differential cooling on either side of a fracture zone, and mechanisms are either strike-slip or vertical dip-slip (one nodal plane is vertical with up or down motion along it). The latter likely represent local stress field variations due to formation of

new oceanic crust, cooling of oceanic lithosphere, and geometrical complexities in plate motions; observed focal solutions include normal, thrust, and strike-slip mechanisms. These earthquakes are usually not very large.

A different type of oceanic intraplate earthquakes is possibly associated with initial formation of new oceanic plate boundaries. The largest strike-slip earthquake ever recorded ($M_w = 8.6$, Table 1) occurred in April 2012 in the Indian Ocean a few hundred kilometers off the west coast of northern Sumatra, Indonesia. The epicenter is located in a broad zone of deformation that marks the diffuse India-Australia plate boundary. Plate boundary faults are not (yet) developed and the stresses due to the slow relative plate motion causes ruptures along zones of preexisting weakness such as fracture zones. The 2012 sequence (including a $M_w = 8.2$ aftershock) shows a complex rupture pattern involving ruptures on several faults. The 1998 $M_w = 8.1$ earthquake near the Balleny Islands, which occurred a few hundred kilometers from the Australia-Antarctic plate boundary, also showed complex faulting behavior and may represent a similar case of reactivation or rupturing of new faults as the 2012 Indian Ocean earthquakes.

Summary

Focal mechanisms of tectonic earthquakes generally follow the deformation style that is expected from the plate tectonic settings they occur in. Thrust mechanisms are observed in subduction zones, collision zones, and other regions of compressional deformation. Normal-faulting earthquakes occur along mid-ocean ridges and in zones of continental extension. Strike-slip earthquakes are observed along transform faults in the ocean and on land where plates or crustal blocks slide laterally past each other. Most earthquakes occur at plate boundaries, but some do occur inside plates (intraplate earthquakes) away from boundaries. Plate boundaries in continental environments are usually broader than their oceanic counterparts. Deep earthquakes are restricted to regions where ocean

lithosphere is recycled back into the mantle at subduction zones. The largest earthquakes are shallow-dipping subduction zone thrust earthquakes.

This entry only provides a general overview of tectonic earthquakes and their source mechanisms. Analysis of seismological data of individual earthquakes and earthquake sequences has in many cases shown complex rupture behavior involving several active faults, which may have different orientations or even a different sense of motion (such as strike-slip and normal faulting or strike-slip and thrust faulting on nearby faults). In addition, slip on faults is not necessarily equally distributed on a fault patch that ruptures during an earthquake; instead slip may show large spatial variations. The reader is referred to the seismological literature for more information on these topics.

Cross-References

- [Earthquake Mechanism and Seafloor Deformation for Tsunami Generation](#)
- [Earthquake Mechanism Description and Inversion](#)
- [Earthquake Mechanisms and Stress Field](#)
- [Long-Period Moment-Tensor Inversion: The Global CMT Project](#)
- [Mechanisms of Earthquakes in Aegean](#)
- [Mechanisms of Earthquakes in Iceland](#)
- [Mechanisms of Earthquakes in Vrancea](#)
- [Moment Tensors: Decomposition and Visualization](#)
- [Non-Double-Couple Earthquakes](#)

References

- Abercrombie RE, Ekström G (2003) A reassessment of the rupture characteristics of oceanic transform earthquakes. *J Geophys Res* 108:2225. doi:10.1029/2001JB000814
- Antolik M, Dreger DS (2003) Rupture process of the 26 January 2001 Mw 7.6 Bhuj, India earthquake from teleseismic broadband data. *Bull Seismol Soc Am* 93:1235–1248. doi:10.1785/0120020141
- Antolik M, Abercrombie RE, Ekstrom G (2004) The 14 November 2001 Kokoxili (Kunlunshan) Tibet

- earthquake: rupture transfer through a large extensional step-over. *Bull Seismol Soc Am* 94:1173–1194
- Bilham R, Gaur VK, Molnar P (2001) Himalayan seismic hazard. *Science* 293:1442–1444. doi:10.1126/science.1062584
- Bird P, Kagan YY, Jackson DD (2002) Plate tectonics and earthquake potential of spreading ridges and oceanic transform faults. In: Stein S, Freymueller JT (eds) *Plate boundary zones*. American Geophysical Union, Washington, DC. doi:10.1029/GD30p0203
- Braunmiller J, Nabelek J (1996) Geometry of continental normal faults: seismological constraints. *J Geophys Res* 101:3045–3052
- Estabrook CH (2004) Seismic constraints on mechanisms of deep earthquake rupture. *J Geophys Res* 109, B02306. doi:10.1029/2003JB002449
- GCMT (2014) Global CMT web page, <http://www.globalcmt.org>. Accessed 20 March 2014
- Kanamori H, Cipar JJ (1974) Focal process of the great Chilean earthquake May 22, 1960. *Phys Earth Planet Inter* 9:128–136
- Lay T, Ammon CJ, Kanamori H, Rivera L, Koper KD, Hutko AR (2010) The 2009 Samoa-Tonga great earthquake triggered doublet. *Nature* 466:964–968. doi:10.1038/nature09214
- Meng L, Ampuero J-P, Stock J, Duputel Z, Luo Y, Tsai VC (2012) Earthquake in a maze: compressional rupture branching during the 2012 Mw 8.6 Sumatra earthquake. *Science* 337:724–726. doi:10.1126/science.1224030
- Oppenheimer D et al (1993) The Cape Mendocino, California earthquake of April 1992: subduction at the triple junction. *Science* 261:433–438
- USGS (2014) Magnitude 8 and greater earthquakes since 1900, http://earthquakes.usgs.gov/earthquakes/eqarchives/year/mag8/magnitude8_1900_date.php. Accessed 20 March 2014
- Ye L, Lay T, Kanamori H, Koper KD (2013) Energy release of the 2013 Mw 8.3 sea of Okhotsk earthquake and deep slab stress heterogeneity. *Science* 341:1380–1384. doi:10.1126/science.1242032

Earthquake Mitigation Strategies Through Insurance

Guillermo Franco

Head of Catastrophe Risk Research – EMEA,
Guy Carpenter & Company Ltd., London, UK

Synonyms

Natural catastrophe risk management; Risk financing; Seismic risk mitigation

Introduction

Insurance can be a powerful *ex ante* strategy in an earthquake risk mitigation framework. Its primordial objective is to provide monetary compensation for damaged assets or lost income but, in addition, it can help achieve other important goals for society such as the establishment of safer building practices, the dissemination of risk information, and the promotion of financial responsibility. Recent scenarios in which insurance has proved most successful suggest strong synergistic effects with other risk mitigation strategies, such as construction regulation and socioeconomic policy.

The study of catastrophe insurance has grown in the last few decades to a burgeoning field bringing together contributions from the private, public, nonprofit, and academic sectors and spanning many varied disciplines. This paper aims at providing a general overview of earthquake (re) insurance strategies without being exhaustive. As earthquake perils are rarely considered in isolation, a review of earthquake insurance necessarily overlaps with broader catastrophe insurance literature.

The usage of earthquake and catastrophe insurance throughout the world is still limited. Guy Carpenter (2014) estimates that 70 % of global economic losses from natural catastrophes were uninsured between 1980 and 2013. Moreover, as the risk exposures grow at a faster pace than insurance, the gap keeps widening.

The insurance and reinsurance industry addresses this challenge via a variety of solutions. Private homeowners may have access to earthquake insurance products that can be purchased individually or through collective, often compulsory, mechanisms such as insurance pools or catastrophe funds. Insurance companies aggregate risks and have access to international reinsurance but also to new instruments in the form of catastrophe bonds or derivatives. These solutions, at the frontier between the traditional insurance market and the financial markets, have also made it possible for sovereign governments to access reinsurance through more varied capital providers.

The first section of this paper establishes the general workings of the modern (re)insurance sector, with a brief look at its history and highlighting the experience in Chile after the 2010 Maule earthquake, as an example of a coordinated financial and political strategy.

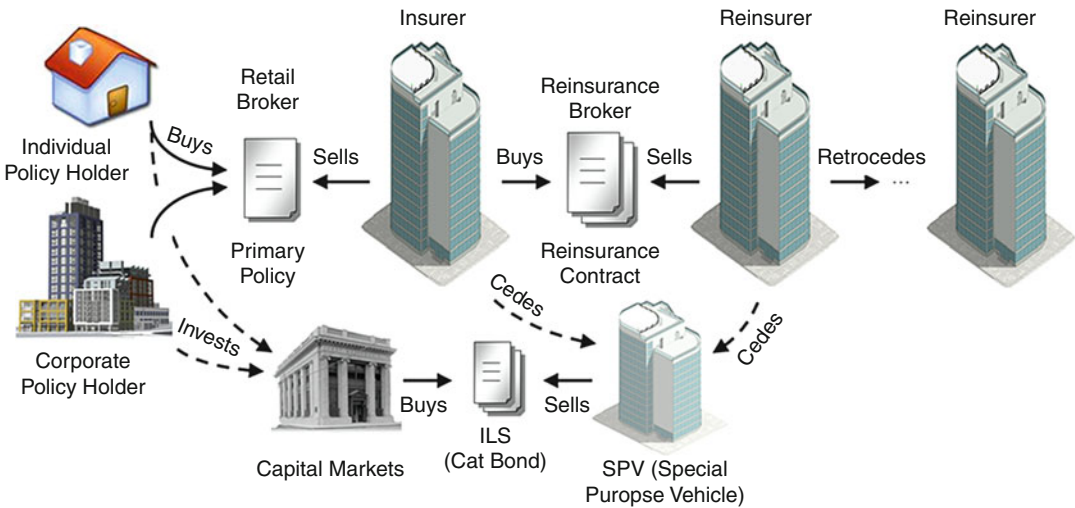
The second section focuses on catastrophe risk models, the technology that makes global risk transfer more seamless and which since the 1980s has had an enormous impact on the industry. The discussion addresses earthquake risk models in particular. These models have played an important role in advances in financial tools that have brought an affluence of capital into earthquake reinsurance. Alternative risk transfer, the convergence market, CAT bonds, parametric solutions, industry indices, and modeled loss triggers, and their importance in earthquake risk transactions are addressed in the third section of the paper.

The aggregation of individual risks and their international transfer is typically facilitated by national private insurance markets. In cases where domestic insurers are unable to accomplish these tasks, state or governmental schemes can be an alternative or a complementary approach. Risk pools, catastrophe funds, or sovereign risk transfer are strategies that a country might adopt with the aim of aggregating and sometimes transfer individual risks. These systems are treated in the fourth and fifth sections of this work where the varied experiences of Turkey, New Zealand, Mexico, and the Caribbean are briefly described.

Broader access to insurance, like broader access to safe building practices, health, or other societal infrastructure, remains an important challenge for developing countries. Micro-insurance systems and community-based parametric schemes are the topic of the final section of the paper.

The Insurance and Reinsurance System

Insurance has a long history. Forms of risk sharing have been documented since early human civilization in Mesopotamia, but the origins of what is known today as the modern insurance and



Earthquake Mitigation Strategies Through Insurance, Fig. 1 A simplified schematic view of the (re)insurance market

reinsurance industry can be traced back to the seventeenth and eighteenth centuries (Swiss Re 2013; Wirtz 2013). During that period, large catastrophic events such as the Great Fire of London of 1666 or the Lisbon earthquake of 1755 cemented the importance of insurance principles considered fundamental today, such as the need for global diversification of risk and the necessity of understanding the underlying physical mechanisms. Probability theory, pioneered by Blaise Pascal, Jacob Bernoulli, and Pierre de Fermat, marked the beginning of risk quantification using mathematical approaches and establishing the foundation of modern actuarial practices.

As the concentration of valuable assets has increased in locations susceptible to natural disasters, the risk exposure has also grown. The 1906 San Francisco earthquake and the ensuing widespread fire constituted the largest catastrophe loss the (re)insurance sector had experienced at the time. Since then, catastrophes have grown more and more costly reaching the estimated US\$210 billion loss caused by the 2011 Tohoku earthquake in Japan (Munich 2013). Moreover, a highly connected international economy has uncovered new vulnerabilities derived from supply chains that can be disrupted by disasters once considered distant (Nanto et al. 2011).

Since catastrophe events are, however, relatively rare, there is a pervasive lack of reliable experience data on their likelihood and consequences. Therefore, assessing and pricing catastrophe risk constitute challenges that would persist until the appearance of numerical simulation models during the second half of the twentieth century. Thanks to this innovation, the quantification of probabilities and effects of physically complex catastrophes slowly starts becoming a standard practice.

Today, the (re)insurance industry is a sophisticated amalgam of specialties and entities connected through a variety of global links. To understand how this risk management engine works, it is useful to consider a simplified view of the transactional relationships that exist in the (re)insurance market (Fig. 1).

Insurance is an arrangement by which a party (the insurer) agrees to compensate another party (the insured) for a set of specific future losses in exchange for a periodic payment (premium). Earthquake insurance policies may cover damages to the built environment, injury, casualty, and business interruption. Both private individuals and companies purchase insurance products to protect their assets and in so doing cede their risk to an insurance company, which effectively

acts as a risk aggregator and diversifies the risk across the population.

An insurance company may in turn cede part of its risk to one or several reinsurance companies, who trade globally and therefore diversify risks further. To share some of its acquired risks, a reinsurer may also “retrocede” parts of their portfolios to other reinsurance companies.

Insurance or reinsurance brokers are intermediaries whose objective is to optimize access to affordable coverage. A retail broker assists individual policyholders in obtaining the most appropriate policy to fulfill their needs from the most adequate and affordable insurance provider. Similarly, a reinsurance broker assists insurance companies to access reinsurance worldwide and find the most appropriate coverage at the best prices. They also work with reinsurers who wish to obtain retrocessional coverage. In addition, brokers may provide a series of services such as claims processing, analytics, and strategy consulting among others.

Today’s reinsurance market has a very strong link with the broader financial and capital markets, through what is known as alternative risk transfer or ART. The most common ART technique consists of securitizing the risk in such a way that it can be sold in the form of a bond to financial investors, who in return for providing capital, obtain an interest on their principal. Sovereign governments can also interact with the capital markets to obtain reinsurance funds through catastrophe bonds. This niche “convergence market” brings capital from financial institutions into reinsurance operations. Note that, as shown in Fig. 1, individuals and companies not only participate actively in the subscription of insurance but, through hedge funds that purchase catastrophe bonds, also contribute to the provision of capital for catastrophe reinsurance.

Today, the reinsurance sector includes large, diversified companies that actively trade in a variety of risks worldwide. Table 1 shows the 15 largest reinsurers according to the net reinsurance premium written in 2012 (Standard and Poor’s 2014).

An example of how insurance systems provide earthquake risk mitigation was recently seen in

Earthquake Mitigation Strategies Through Insurance, Table 1 Largest reinsurers by 2012 net reinsurance written premium (US\$ m)

Reinsurer	Premium
Munich Reinsurance Co. (Germany)	35,797.2
Swiss Reinsurance Co. (Switzerland)	25,344.0
Hannover Rueckversicherung AG (Germany)	16,346.0
Berkshire Hathaway Reinsurance Group/General Re Corp. (US)	16,145.0
Lloyd’s of London (UK)	11,372.7
SCOR (France)	11,285.6
Reinsurance Group of America, Inc. (US)	7,906.6
Partner Re (Bermuda)	4,573.0
Everest Re (Bermuda)	4,081.1
MS&AD Holdings (Japan)	3,446.3
Korean Re (South Korea)	3,390.2
Transatlantic Holdings, Inc. (US)	2,840.7
Nipponkoa Sampo Japan Holdings (Japan)	2,717.4
Tokio Marine Group (Japan)	2,626.5
General Insurance Corp. of India (India)	2,533.6

Source: Standard & Poor’s (2014)

Chile in 2010, as the country faced an economic loss estimated at US\$30 billion arising from the February 27, M_w 8.8 Maule earthquake off the coast of Concepción (Siembieda et al. 2012). Chile enjoys a strong presence of insurance companies that were able to assume close to one third of the total economic loss, about US\$8.4 billion in insured damages (Marsh 2014). These insurance companies were in turn well connected to the global reinsurance markets that provided them coverage to an astounding 95 % (Aon Benfield 2011). This means that most of the insured loss was swiftly transferred to the international reinsurance markets, where it had little impact in the finances of large capital providers. With this robust system in place, Chile’s private market effectively transferred about 30 % of the total economic loss caused by the Maule event outside its borders without any direct government intervention.

This success story worked hand in hand with Chile’s strong enforcement of good construction practices over the decades coupled with compulsory subscription of earthquake insurance for all properties purchased with a mortgage. Without a strong legacy of proactive construction

regulation, insurers would not be easily drawn to this market. And without the compulsory system, the growth of insurance penetration would be severely hampered. Indeed, many owners who had purchased their property outright with cash, a well-established investment strategy in Chile, had no insurance protection at the time of the event.

Regulation and social policies stimulated the growth of the insurance market, and the combination of these “hard” and “soft” resilience strategies was identified as highly effective during the early recovery period (Franco and Siembieda 2010). Much of the success of the insurance system in the Maule event was ultimately attributed to the fast claims settlement process. Ninety percent of the claims were reported by affected households within the first 30 days after the event, and more than 75 % had been settled within less than a year (Marsh 2014).

Quantifying Earthquake Risk and the Rise of CAT Models

The transfer of earthquake risk from Chile to the rest of the world is possible at a global level in great part due to the existence of tools that help to quantify it. These tools are known today as catastrophe or CAT risk models.

The insurance and reinsurance industry has traditionally used historical loss records to price fire, life, theft, and other risks on which there is plenty of experience data, and catastrophes were originally considered in conjunction with more common hazards. Ignoring their specific potential loss contribution resulted in unforeseen insolvencies for the insurance industry in events like hurricane Andrew in 1992. In order to better assess catastrophe risks, companies needed experience data that wasn't available.

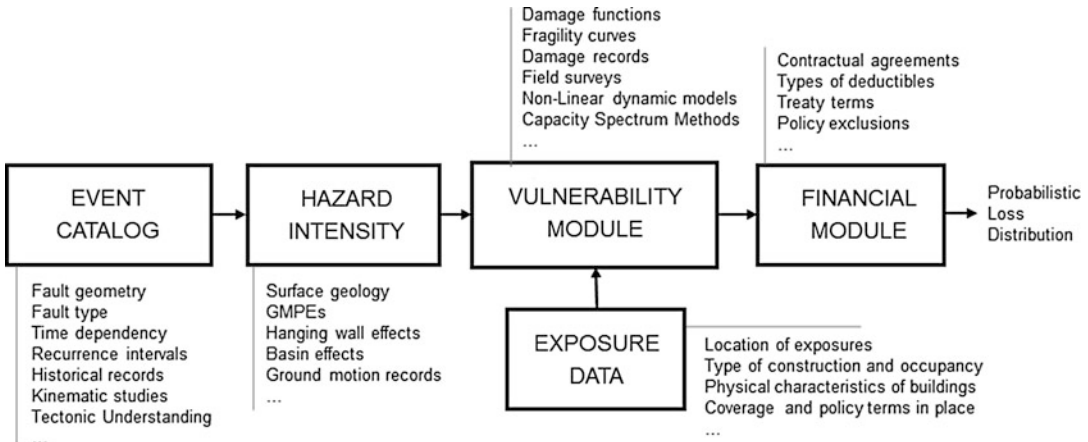
During the late 1980s and 1990s, new risk assessment platforms were being developed and deployed in the market. Grossi and Kunreuther (2005) present a comprehensive view of the history and evolution of numerical CAT models that simulate the physical processes involved in catastrophe events.

The paper by Clark (1986) brought the concept of Monte Carlo simulation applied to the assessment of hurricane losses to the actuarial community and paved the way for the foundation of the first catastrophe modeling company, Applied Insurance Research, in 1987. Clark relied on the visionary numerical approaches developed by Friedman (1975) and others. Early earthquake simulation models were simultaneously being developed in California supported by new probabilistic methods (Cornell 1968; Esteva 1970; McGuire 2008).

Earthquake models have become extraordinarily complex, and teams that develop them have grown to include a wide array of specialists such as seismologists, geologists, engineers, statisticians, and financial and software experts. Most commercial earthquake models fundamentally consist of five important modules as shown in Fig. 2. The first generates an event catalog, a set of simulated events compatible with the available knowledge on the seismicity of a specific territory. This knowledge may include fault location data, recurrence intervals, fault geometry, kinematic analyses, etc.

The second module uses ground motion prediction equations, soil data, and fault-related effects, among other parameters, to generate fields of hazard intensities associated with each of the earthquakes contained in the event catalog. These ground motions are in turn fed into a process that calculates possible damage distributions across the affected area for different types of structures. The crux of this vulnerability module consists of a set of mathematical relationships in the form of damage functions or fragility curves derived from structural engineering techniques such as the capacity spectrum method or nonlinear dynamic analysis. Exposure data from the user serves to define the types of properties for which the analysis needs to be executed. Lastly, the financial module applies the terms contained in (re)insurance policies in order to obtain a monetary estimate of insured loss based on the computed physical damage.

Only three model development companies, AIR Worldwide, EQECAT, and Risk Management Solutions (RMS), had a serious foothold in



Earthquake Mitigation Strategies Through Insurance, Fig. 2 Basic anatomy and elements of an earthquake risk model

the industry during the early years of CAT modeling. Later, other groups started operations in the market occasionally specializing in regional or peril-specific models such as Empresa de Riesgos Naturales (ERN) based in Mexico and with a very strong local relevance in Latin America or Risk Frontiers based out of Macquarie University in Australia and specializing in Asia-Pacific perils. Initiatives like the Oasis Loss Modeling Framework aim at lowering the barrier to entry for academics, consultants, and other market participants. The philosophy of this effort is to create a standardized platform that liberates potential entrants from the common pains of software development and from the relatively specialized financial computations involved in reinsurance analysis. In this way, scientists, engineers, and risk experts may be able to deploy their knowledge more seamlessly onto channels easily accessible by the industry.

During this technological revolution that started in the late 1980s, catastrophe risk (re) insurance has changed enormously. Rare would be the case today that an insurance or reinsurance company managed their catastrophe risks without any usage of CAT models. In the absence of actual loss experience data, these numerical tools have been successful at leveraging scientific information to produce a large body of simulations from which the market can now objectively assess and price risks – albeit with significant

limitations and uncertainties. As models become more ubiquitous, the quantification of uncertainty has gained in importance. Several studies have addressed this topic in the context of hurricane (Miller 1999; Major 1999a, b) and earthquake models (Spence et al. 2003). These tools are also still of limited use to address the occurrence of very large or very rare events. When the scientific knowledge is insufficient, all models tend to fail, as was evident in Japan when the M_W 9.0 Great Tohoku earthquake struck in an area where only up to M_W 8.4 events were thought possible.

The numerical estimates from CAT models make it possible to calculate the frequency of loss, which is expressed in the industry as an exceedance probability curve or “EP curve,” a relationship between the value at risk and its probability. This information is critical to price an insurance policy or a reinsurance contract. The pure premium of an insurance policy is defined as its associated average annual loss, and solvency requirements are often expressed as probable maximum losses at specific return periods, typically at 100 years or 250 years.

By using numerical models to estimate the necessary metrics, the industry has managed to overcome the immediate problem of lack of historical experience. However, the battle is not over. Data on physical and economic losses that would help produce better models continues to be scarce, often at disparate resolutions, and

of varied quality. Therefore, the industry continues to look at diverse scientific or professional initiatives such as the Global Earthquake Model or the Learning from Earthquakes initiative spearheaded by the Earthquake Engineering Research Institute (EERI) as extremely valuable contributions.

Alternative Risk Transfer Solutions

CAT models have improved the way in which the market manages and prices its catastrophic risks, and moreover, they have had a great impact on the industry by opening new avenues to transfer risks through the capital markets with more transparency.

The financial markets have increasingly shown interest in catastrophe reinsurance operations as a source of diversification for investment portfolios. Through a series of new instruments, investors have traded in catastrophe risks thus acting as an alternative source of reinsurance capital. These alternative risk transfer (ART) solutions are today a common way to interact with the “convergence market,” the space where finance meets insurance (World Economic Forum 2008).

CAT bonds have probably been the most successful instrument used to “package” catastrophe risks into a securitization. This solution consists of creating a special purpose vehicle, which acts as a reinsurer to a single entity and which holds the bond principal invested in safe, low-risk, short-term securities. Cummins (2008) presents a comprehensive view of the history and types of securitizations and derivatives that have been used in the market.

Capital providers have had an interest in isolating catastrophe risk to avoid other components of risk that are already present in usual financial portfolios. Fully collateralized CAT bonds, for instance, have attracted interest since they eliminate the risk of default of the parties involved.

Similarly, moral hazard present in traditional indemnity-based (re)insurance processes is also undesirable for the capital markets. This problem arises when a party can potentially misrepresent

the actual loss with the aim of obtaining a financial advantage. When the covered party experiences a damaging event that causes loss, it issues a claim to the (re)insurer. This, in turn, assesses the damage through a claim-adjusting process and establishes the corresponding payment, and the (re)insured is compensated accordingly and depending on the policy conditions of the agreement. In addition to being very expensive and cumbersome, claim-adjusting processes provide opportunities for the claimant to exaggerate the loss or for the insurer to downplay it. Capital providers who do not have the mechanisms to assess the quality of the data provided during this process have a vested interest in avoiding it.

With the aim of reducing moral hazard, new techniques for risk transfer became popular during the 1990s (Bisping 2013). Indexed, modeled loss, or parametric solutions remove the need to assess particular losses to a portfolio by tying the payment of recoveries to an independent proxy.

PROPERTY CLAIM SERVICES (PCS), a company operating in the USA, or PERILS, a similar entity operating in Europe, collect loss data from the major insurance companies after a catastrophe and aggregate them into an overall loss estimate. Risk transfer solutions based on these market estimates tie particular recoveries to the overall loss reported by the insurance market. By removing the direct connection between recoveries and particular portfolio losses, moral hazard is greatly diminished. If the event created a large loss for the insurance industry as a whole, it seems intuitive that recoveries for a particular insurer should be somewhat proportional to the overall loss. Naturally, this type of index works best for large companies whose portfolios mimic the distribution of the entire industry. For companies that operate within an industry niche, these types of solutions need to be tailored somehow since they may involve too much “basis risk,” the potential difference between actual losses and recoveries.

Modeled loss solutions base their payments on the loss produced not by the event itself but on the loss estimated by a numerical CAT model. The model is run after an event assuming certain parameters that match measurements from the

event. For instance, an earthquake model that accounts for epicentral location, magnitude, depth, and rupture length of the event may be run setting these parameters equal to estimates provided by the United States Geological Survey (USGS). Further constraining of the model may be accomplished by generating a series of stochastic footprints and identifying those that match more closely the observations obtained from a network of seismometers, if these data are available.

Although modeled loss solutions bypass the claims process, ambiguities may still creep up into the post-event loss calculation (PELC), a series of activities undertaken by a consultant or calculation agent in order to run the model and determine trigger conditions and index values. These steps often involve models that contain significant intellectual property from their developers and whose technical details may not be fully disclosed. Therefore, the intimate workings of modeled loss indices are usually not well understood by parties external to the transaction, and outside investors may feel that they don't have enough information about the likely range of outcomes.

Parametric (re)insurance bases payments neither on the direct observed damages nor on modeled estimates of loss. Rather, these solutions use measurable parameters of the physical event to determine the payouts. Within this class, two types of solutions have been used in the market successfully: *first-generation* CAT-in-a-box triggers and *second-generation* parametric indices.

CAT-in-a-box triggers associate discrete payments to a set of physical event parameters such as magnitude, epicenter location, and focal depth. They are discontinuous in nature and typically involve significant basis risk. Attempts to reduce the basis risk associated with these transactions have consisted of optimizing the trigger thresholds (Franco 2010) or discretizing the parameter domain (Franco 2013).

Transactions that rely on geographically distributed parameters of the event, such as ground motions, have also appeared in the market in the form of *second-generation* parametric indices. These indices are computed as a weighted sum of some mathematical expression, often

polynomials, evaluated at a set of locations using a ground motion metric as the main variable. Typical metrics are peak ground accelerations, spectral acceleration at a particular natural period, or some other descriptors of site-specific shaking like *Shindo* intensity in Japan.

Second-generation parametric indices are generally considered “better” than first-generation solutions because the correlation between losses and local shaking intensity is assumed to be stronger than between losses and the main event parameters of magnitude and epicentral location (Goda 2013). However, in a real scenario, the uncertainties affecting ground motions used in second-generation parametric indices are typically much greater than those affecting the main parameters of the event. Quality comparisons between triggers often remain difficult to pursue since both solutions are dependent on a plethora of design parameters such as resolution used, number of stations, etc.

The continued appearance of both, first- and second-generation solutions, suggests that there is space in the market for a variety of products that may coexist and complement one another. Table 2 collects all the parametric earthquake transactions of first and second generation (trigger type) since 1997 (compiled from www.artemis.bm). Cases in which there is high uncertainty in the exposure distribution could favor first-generation approaches, whereas cases in which reliable networks of seismometers are present and the exposures are well known could tend to support second-generation approaches. Edmonds (2013) provides a market perspective on some of the transactions used in the recent past, and in particular the Muteki Ltd. CAT Bond that was issued in 2008 to protect the large mutual Zenkyoren in Japan, whose trigger mechanism was activated during the M_w 9.0 March 11th 2011 Tohoku event and which provided US\$300 million in recoveries. In some occasions, the lack of a reliable network of seismometers has been circumvented by simulating the response of a set of “virtual” stations (identified as “V” in the trigger type of Table 2) through ground motion models such as the USGS’ ShakeMap or PAGER systems (Wald et al. 2008).

Earthquake Mitigation Strategies Through Insurance, Table 2 List of parametric earthquake transactions (www.artemis.bm). *Note:* Details on the transactions are approximately inferred from information provided through public sources, and they may not always be accurate or complete

Issuer – cedant/sponsor	EQ region	Size in US\$m	Trigger type	Year
Parametric Re – Tokio Marine & Nichido	JPN	100	1st	1997
Mitsui Marine & Fire Swap – Mitsui M&F	JPN	30	1st	1998
Concentric – Oriental Land Co.	JPN	100	1st	1999
Prime Cap. CalQuake/EuWind – Munich Re	US	135	1st	2001
Atlas Reinsurance II – SCOR	JPN/US	150	1st	2001
Fujiyama – Nissay Dowa Gen. Ins. Co.	JPN	70	1st	2002
PIIONEER 2002 – Swiss Re	JPN/US	255	1st/2nd	2002
Studio Re – Vivendi Universal SA	US	175	2nd	2002
Phoenix Quake – Zenkyoren	JPN	193	2nd	2003
Phoenix Quake Wind – Zenkyoren	JPN	85	2nd	2003
Phoenix Quake Wind II – Zenkyoren	JPN	193	2nd	2003
Arbor I – Swiss Re	JPN/US	95	2nd	2003
Arbor II – Swiss Re	JPN/US	27	2nd	2003
Sequoia Capital – Swiss Re	US	23	2nd	2003
Sakura Capital – Swiss Re	JPN	15	2nd	2003
Arbor I (2) – Swiss Re	JPN/US	60	2nd	2003
Arbor I (3) – Swiss Re	JPN/US	9	2nd	2003
Sequoia Capital (2) – Swiss Re	US	12	2nd	2004
Arbor I (4) – Swiss Re	JPN/US	21	2nd	2004
Arbor I (5) – Swiss Re	JPN/US	18	2nd	2004
GI Capital – Swiss Re	JPN	125	2nd	2004
Sequoia Capital (3) – Swiss Re	US	11	2nd	2004
Arbor I (6) – Swiss Re	JPN/US	32	2nd	2004
Arbor I (7) – Swiss Re	JPN/US	15	2nd	2004
Arbor I (8) – Swiss Re	JPN/US	20	2nd	2005
Cascadia – FM Global	US	300	1st	2005
Arbor I (9) – Swiss Re	JPN/US	25	2nd	2005
Arbor I (10) – Swiss Re	JPN/US	18	2nd	2005
Australis – Swiss Re	AUS	100	2nd V	2006
CAT-Mex – FONDEN	MEX	160	1st	2006
Successor – Swiss Re	US	950	2nd	2006
Cascadia II – FM Global	US	300	1st	2006
Successor I – Swiss Re	US	29	2nd	2006
Redwood Capital IX – Swiss Re	US	300	2nd	2006
Australis (2) – Swiss Re	AUS	50	2nd V	2007
Successor II – Swiss Re	US	100	2nd	2007
MedQuake – Swiss Re	EU-Med	100	2nd V	2007
Fremantle – Brit Insurance	JPN	200	2nd	2007
Fusion 2007 – Swiss Re	MEX	140	1st	2007
Midori – East Japan Railway Co.	JPN	260	1st	2007
Successor II (3) – Swiss Re	US	100	2nd	2007
Redwood Capital X – Swiss Re	US	499	2nd	2007
Muteki – Zenkyoren	JPN	300	2nd	2008
Vega Capital (2008–1) – Swiss Re	US	150	2nd	2008
Topiary Capital – Platinum Underwriting	JPN	200	2nd	2008

(continued)

Earthquake Mitigation Strategies Through Insurance, Table 2 (continued)

Issuer – cedant/sponsor	EQ region	Size in US\$m	Trigger type	Year
Successor II (4) – Swiss Re	US	60	2nd	2009
MultiCat Mexico 2009 – FONDEN	MEX	290	1st	2009
Successor X – Swiss Re	US	150	2nd	2009
Atlas VI Capital (2009–1) – SCOR	JPN	100	2nd	2009
Atlas VI Capital (2010–1) – SCOR	JPN	100	2nd	2010
Vega Capital (2010–I) – Swiss Re	JPN/US	107	2nd	2010
Montana Re (2010–1) – Flagstone Re	JPN/US	210	2nd	2010
Successor X (2011–1) – Swiss Re	AUS/US	170	2nd	2010
Successor X (2011–2) – Swiss Re	US	305	2nd	2011
Kibou (2012–1) – Hannover Re/Zenkyoren	JPN	300	2nd	2012
MultiCat Mexico (2012–1) – Swiss Re	MEX	315	1st	2012
Bosphorus 1 Re – TCIP	TURKEY	400	2nd	2013

Earthquake Insurance Pools

It is convenient for the effective usage of risk transfer tools that risks are properly aggregated, or “pooled,” in the first place. Individual property risks are grouped into larger portfolios through the standard operations of insurance companies, but penetration of voluntary catastrophe insurance is often low. To boost the take-up rate and aggregation of individual risks, other instruments can be used, most prominently the creation of risk pools.

The main purpose of a pool is to create a compensation fund that works on the principle of solidarity. If an entire community is exposed to a set of perils, all members in the community contribute a premium to the pool, which later can be used to pay for damages incurred. Pools differ by their premium collection mechanism, by the split of financial responsibilities between participating actors, by their relationship with broader reinsurance networks, and by the level of involvement of public versus private entities (Consortio de Compensación de Seguros 2008). On this last item, it should be noted that incentives and subsidies provided through public entities for the management of risks may have strong beneficial effects on society, but they may also lead to difficulties in assessing the actual costs and value associated with such measures.

Switzerland’s insurance sector formed the first formal catastrophe insurance pool in 1939,

known as the Swiss Natural Perils Pool. Although this pool does not yet cover damages from earthquakes across all Swiss cantons, it covers a variety of other perils such as windstorm, hail, avalanche, and snow pressure (Schweizerisches Versicherungsverband or SVV 2014). The Swiss pool is managed by the association of insurers acting in coordination with the national and regional governments.

The Consorcio de Compensación de Seguros, another early example of a pool, was created in 1941 after the Spanish Civil War (1936–1939) to compensate losses derived from rioting during the war years. In 1945, the Consorcio expanded its mission and turned into a more general compensation pool that includes damages from catastrophic perils. It paid, for instance, a total of €489 million to compensate losses from the Lorca M_w 5.1 2011 earthquake in southern Spain (CCS 2014). The Spanish pool is a public entity that operates independently from the insurance industry but that receives the appropriate contribution from policies underwritten in the market. While it does not seek reinsurance in the international market (thus providing no international risk transfer), other pools actively look for this financial support.

New Zealand’s Earthquake Commission (EQC), for instance, expects to obtain about NZ\$4.5 billion in reinsurance recoveries for the Christchurch earthquake sequences that started in February of 2011 (EQC 2013). In June 2013, the

Earthquake Mitigation Strategies Through Insurance, Table 3 Existing national catastrophe pools (World Forum of Catastrophe Programs 2014)

Country or region	Year of inception	Pool name	Summary of characteristics
California (USA)	1996	California Earthquake Authority (CEA)	Provides direct insurance for earthquake residential exposures, participation of insurers and reinsurers is voluntary
France	1982	Catastrophes Naturelles (CatNat)	State-owned system, compulsory participation, premiums collected through commission on standard fire policies
Iceland	1975	Vidlagatríggjng Islands	State-owned system, compulsory participation, premiums collected through commission on standard fire policies
Japan	1966	Japan Earthquake Reinsurance Co. Ltd. (JER)	Privately owned by association of reinsurers, compulsory excess of loss cover, residential exposures, state backed
Mexico	1996–1999	Fondo de Desastres Naturales (FONDEN)	State-owned system, publicly financed, for ex ante and ex post activities
New Zealand	1993	Earthquake Commission (EQC)	State-owned system, compulsory participation, premiums collected through commission on standard fire policies
Norway	1979	Norsk Naturskadepool/ Statens Naturskadefond	State-owned system with participation of all nonlife insurers covering natural disaster damages, premium established by pool, collected by insurers
Romania	2009	Natural Disaster Insurance Pool (PAID)	Jointly owned company by insurers, compulsory dwelling cover for earthquake, flood, and landslide
Spain	1941–1945	Consorcio de Compensación de Seguros (CCS)	State-owned system, compulsory participation, premiums collected through commission on standard fire policies
Switzerland	1939	Swiss Natural Perils Pool	System promoted by the association of Swiss insurers, premiums collected in conjunction with standard policies
Taiwan	2001	Taiwan Residential Earthquake Insurance Fund (TREIF)	State-owned system, compulsory participation, premiums collected through commission on standard fire policies
Turkey	2000	Turkish Catastrophe Insurance Pool (TCIP)	State-owned system, compulsory participation, TCIP policies distributed by accredited insurers

EQC had disbursed a total of NZ\$5.7 billion in claims payments and had collected a total of NZ\$1.8 billion in reinsurance. The pool still contained NZ\$2.9 billion then although the expectation was that it would be completely depleted by the time all claims have been settled. The fund would have been further impaired to respond to the Christchurch events without the support of reinsurance.

In Asia, two prominent programs in Japan and in Taiwan address the pooling of risks through two distinct strategies, the first through the participation of private companies in a state-backed reinsurance provider and the second using a

similar mechanism to EQC in New Zealand, through a compulsory fee on standard fire policies. Many of the national pools currently in operation (summarized in Table 3) show distinct degrees of involvement of different governments (Nguyen 2013). The design criteria of each pool depend on history, on the existing insurance mechanisms, on government capacity, and on the objectives of each particular effort.

While the main purpose of the pool mechanism is to provide monetary compensation, it can have a strong socioeconomic effect as well. The Turkish Catastrophe Insurance Pool (TCIP), established in 2000 after the Kocaeli and Düzce

earthquakes of 1999, has been a conduit to accomplish many objectives such as international risk transfer, reduction of liability of the government, prevention of further taxation, promotion of good construction practices, and an increase of social cohesion. Similarly to the EQC, the pool has also financed education and research projects in order to foster the understanding and communication of risks.

TCIP has typically accessed the traditional international markets for reinsurance. In 2013, it decided to complement this coverage through an ART tool called Bosphorus 1 Re Ltd. This parametric earthquake transaction is based on a ground motion index that relies on a network of seismometers around Istanbul managed by the Kandilli Observatory and Earthquake Research Institute. The CAT bond was oversubscribed when it hit the market, showing the great interest of global financial entities to participate in the reinsurance of TCIP as a diversifying risk (Artemis 2013). It also provided TCIP with coverage at advantageous prices.

The example of Turkey's securitization appeal is often compared to the success of a series of CAT bonds issued by FONDEN in Mexico. The FONDEN system is fully backed by the Mexican government, without any direct participation of the insurance industry. Originally established as an agency to provide funds after catastrophic events, it has grown to incorporate risk financing among its responsibilities. Mexico has been tremendously innovative in governmental risk transfer and is an often-cited success story of sovereign risk transfer, an alternative solution to traditional reinsurance for governments.

Sovereign Earthquake Risk Transfer

The cases of Chile, New Zealand, or Turkey illustrate how countries can effectively transfer catastrophe risk to the global markets via their domestic insurance systems. A country may seek financial protection from the global markets in a more direct fashion in what is referred to as "sovereign" risk transfer. The development of insurance solutions in the sovereign space has

been strongly guided by the Hyogo Framework for Action 2005–2015, which identified the need to promote the development of financial risk-sharing mechanisms, particularly insurance and reinsurance against disasters (Cummins and Mahul 2009). The impact that the Global Facility for Disaster Reduction and Recovery (GFDRR) and the World Bank have had on advances in this field is patent in the solutions implemented during the last decade.

In 2006, the Government of Mexico (through FONDEN) sponsored CAT-Mex Ltd., a parametric catastrophe bond designed to finance rescue and rebuilding costs after an earthquake. The bond successfully transferred US\$160 million of Mexico's earthquake risk to the global financial markets on a parametric basis over a period of 3 years. Several geographic zones were defined and associated with threshold conditions on the magnitude and location of the earthquake that need to be met in order for the coverage to be triggered.

The transaction was deemed successful and the experience was repeated after the bond matured in 2009. Mexico then issued the MultiCat Mexico 2009 bond expanding its coverage to hurricane risk for a total of US\$290 million until 2012, at which point the transaction was reissued for US\$315 million under the name MultiCat Mexico Ltd. (Series 2012-1) maturing in 2015.

The 2009 bond was oversubscribed, with demand for the Mexican bonds outpacing supply. The bond originally required \$250 million of coverage but was upsized to \$290 million to satisfy market demand, which in turn translated into more attractive pricing for Mexico.

Over the last decade, Mexico has established itself as the first sovereign country that has effectively transferred some of its catastrophe risk outside of its borders via these novel instruments. The Mexican experience has been pioneering and has attracted much attention and interest. It has also been the subject of several studies that aimed to describe the features of the first transaction (Cardenas et al. 2007), optimize the efficiency of these tools in the context of other reinsurance strategies (Härdle and López Cabrera 2010), or

give a more comprehensive perspective on the construction, implementation, and lessons learned from these solutions (Michel-Kerjan et al. 2011). In addition, trade journals such as the Bermuda-based specialist website www.Artemis.bm have followed the activity related to the Mexican bond issuances and have made available plenty of interesting information as to how markets have reacted to these transactions over the years.

The financial markets, by welcoming Mexican catastrophe risk, have proved that the issuance of these bonds is a plausible strategy for a country to transfer its risks. However, the case of Mexico involves one single country with a strong legal and financial system in place and with a long history of insurance experience. Other emerging countries with fewer systems in place require other strategies to successfully leverage risk transfer solutions.

This was the motivation for the creation of the Caribbean Catastrophe Risk Insurance Facility (CCRIF) in 2007. This institution was formed by 16 Caribbean countries, under guidance from the World Bank, to pool and transfer regional hurricane and earthquake risks to the global financial markets (World Bank 2008). The trigger mechanisms were also parametric, conceptually similar to those used in Mexico. However, the compensations were, in general, not designed to finance long-term operations but to provide a cash injection to address what the CCRIF defined as the “liquidity gap,” the period of time directly after the emergency response and before other assistance funds can be accessed for long-term recovery. Countries like Haiti with nonexistent or feeble financial systems in place would have found it hard to access insurance on a stand-alone basis like Mexico did. However, CCRIF has been able to access the reinsurance community and has effectively responded to several events during its existence, benefitting the governments of Anguilla, St. Vincent and the Grenadines, Barbados, Dominica, St. Lucia, Turks and Caicos Islands, and Haiti.

The case of Haiti also brought to light that in order for insurance mechanisms to be effective, some basic governance and finance infrastructure

needs to be in place. After the M_w 7.0 earthquake of January 2010, the country was devastated (DesRoches et al. 2011). The government fell in complete disarray, and chaos ensued. When CCRIF communicated that it was ready to make a payment of about US\$7.7 million (about 20 times the yearly premium paid by Haiti) as per the pool arrangement, it took some time, in the absence of an operational government and banking community, to find mechanisms to deploy the funds.

Insurance-pooling schemes have provided developing nations with a foothold in the global reinsurance networks. As CCRIF has established itself as a successful operation, similar initiatives have been set in motion in Africa (African Risk Capacity 2012) and in the South Pacific (Pacific Disaster Risk Financing and Insurance Program 2011). Recently, cyclone Ian triggered the payment of US\$1.27 million to the government of Tonga, which participates in this newly created system.

Remarks on Micro-Insurance

Low-income populations in developing nations frequently use community-based financing or insurance schemes, often with great creativity (Rutherford 1999). However, due to the covariant nature of catastrophes, community-based financing techniques are not well suited to respond to events that may affect the entire participating community at once. This has spawned an exploration into whether formal insurance tools can overcome this challenge for developing nations (Linnerooth-Bayer and Mechler 2009).

In general, the few experiences that exist and the analyses carried out suggest that insurance can indeed be a good complement to other risk mitigation initiatives. However, when analyzing the benefits of insurance in a developmental setting, it should not always be considered evident that cash by itself is helpful for long-term recovery or for building future resilience. Monetary support often needs to be accompanied by other initiatives in order to have a truly positive impact on society.

In a recent reconstruction survey in Sri Lanka, eight and a half years after the island nation was hit by the devastating Indian Ocean tsunami of December 2004, Franco et al. (2013) visited several communities along the coast. Differences in recovery outcomes were observed between sites that had been managed by third-party donors versus settlements where owners had been empowered to rebuild through a gradual injection of cash. Through this process, owners were provided a set of cash packages that allowed them to rebuild on their original plot in subsequent phases, assuming that the land remained fit for settlement. At each phase, an inspector would survey the construction quality and would give approval for the next stage, at which point the owner would obtain a transfer of cash through prearranged “tsunami bank accounts.” The outcomes of settlements built by donors on land provided by the government varied in quality, but the feedback on the owner-driven construction scheme was almost universally positive, which suggests that in the presence of other supporting mechanisms, access to cash can truly play a powerful role in a successful recovery.

Micro-insurance initiatives have started addressing part of this problem by tying insurance products to other broader services that integrate the population in formal financial networks. Mechler et al. (2006) discuss several property insurance “bundles” connected to other social and financial services in Nepal, Bangladesh, India, Pakistan, and Malawi. Insurance can have greater societal impact if its purchase links communities with educational initiatives, construction oversight, or risk mitigation, as echoed in the objectives of Turkey’s TCIP program.

Since micro-insurance solutions made their appearance in the early 2000s, the experience so far is very limited. In the meantime, significant challenges have been identified in the marketing and deployment of insurance solutions within populations that are remote or where great educational and cultural barriers exist. This has been partly addressed by the usage of parametric solutions, which as described above, suffer from basis risk but also help bypass claim processes that

would be especially costly in locations with difficult access and lack of skilled adjusters.

In this context, India has made important advances by implementing indexed schemes for crop insurance, successful in great part thanks to the governmental policy that requires the insurance sector to service the low-income populations (Mechler et al. 2006; Giné et al. 2010). Other countries in Africa, Asia, and Latin America are making use of similar schemes (Ibarra 2009).

Earthquake-specific insurance programs have been studied for implementation in China, where these types of solutions have the potential to fit well. China has a sophisticated insurance market, well connected to the reinsurance industry (Wang et al. 2009) but with many needs in low-income communities in earthquake-prone areas. In particular, parametric solutions have been studied for deployment using modern CAT modeling techniques (Del Re et al. 2008).

Summary

Insurance successes, large or small, have been witnessed in events like the Maule earthquake in Chile or the Christchurch earthquakes in New Zealand. Many experiences remain highly imperfect, but in the meantime, countries like Turkey and Mexico have made great strides in advancing their earthquake risk management strategy. Embracing ex ante risk management tools, many countries are slowly assuming more and more responsibility over their future.

Earthquake (re)insurance systems offer exciting opportunities for development along all the dimensions presented in this paper. At the core of national economies, insurance industries need to be established and strengthened in order to formally manage existing risks. In countries that lack a strong socioeconomic foundation, multi-national risk pools or sovereign risk transfer remain open options to slowly reshape or complement ex post response strategies with proactive risk management. Technological advances made in the last few decades have helped communities assess their exposures to a much larger

degree than they were once able to, but refinements of these computational tools are continuously being made to improve the quantification risk. In parallel, new promising paths in alternative risk transfer continue to be charted in the financial markets, and new products and distribution channels continue to be explored at the community level through micro-insurance initiatives.

Experiences suggest that availability of capital after catastrophe events at an individual and governmental level can support a better and faster recovery and can ultimately increase a society's resilience. Systems that facilitate the rapid deployment of funds can provide almost instant liquidity at a time of need. Earthquake insurance can therefore be a strong component of any seismic risk mitigation framework with potentially synergistic effects with other coexisting measures such as risk education, construction regulation, and socioeconomic policies.

The (re)insurance sector and society as a whole are still learning important lessons from earthquake events today, and much work remains to be done. Earthquake and tsunami perils are often perceived as "distant" in time and space, and this often hinders the implementation of risk management policies and measures. However, the combination of documented growing economic losses, small overall global earthquake insurance penetration rates, and a large supply of capital means that an increasing portion of the financial burdens imposed by earthquakes worldwide could be assumed by global financial networks rather than by often fragile local economies. This represents a challenging task but also an exciting prospect.

Cross-References

- ▶ "Build Back Better" Principles for Reconstruction
- ▶ Building Disaster Resiliency Through Disaster Risk Management Master Planning
- ▶ Community Recovery Following Earthquake Disasters
- ▶ Damage to Buildings: Modeling
- ▶ Damage to Infrastructure: Modeling

- ▶ Earthquake Disaster Recovery: Leadership and Governance
- ▶ Economic Impact of Seismic Events: Modeling
- ▶ Economic Recovery Following Earthquakes Disasters
- ▶ Insurance and Reinsurance Models for Earthquake
- ▶ Learning from Earthquake Disasters
- ▶ Probabilistic Seismic Hazard Models
- ▶ Seismic Risk Assessment, Cascading Effects
- ▶ Site Response for Seismic Hazard Assessment

References

- African Risk Capacity (2012) African risk capacity – sovereign disaster risk solutions, a specialized agency of the African Union. Accessible on-line at http://www.africanriskcapacity.org/documents/350251/371107/ARC_Overview_Brief_EN.pdf
- Aon Benfield (2011) Chile: one year on. Chicago. Accessible on-line at www.aon.com/attachments/reinsurance/201102_chile_one_year_on_report.pdf
- Artemis (2013) Bosphorus 1 Re CAT bond parametric trigger details emerge. Accessible on-line at <http://www.artemis.bm/blog/2013/04/04/bosphorus-1-re-cat-bond-parametric-trigger-details-emerge/>
- Bisping M (2013) Risk transformation through capital markets. In: Lane M (ed) Alternative (re)insurance strategies, 2nd edn. Risk Books, Chicago
- Clark K (1986) A Formal Approach to Catastrophe Risk Assessment and Management, in the Proceedings of the Casualty Actuarial Society, Vol. 73, Part 2, No. 140
- Consorcio de Compensación de Seguros or CCS (2008) Natural catastrophes insurance cover – a diversity of systems. Accessible on-line: http://www.wfcatprogrammes.com/c/document_library/get_file?folderId=13442&name=DLFE-553.pdf
- Consorcio de Compensación de Seguros or CCS (2014) Estadística Riesgos Extraordinarios 1971–2013. Accessible on-line at http://www.consorseguros.es/web/c/document_library/get_file?uuid=548d4f59-b6c5-40dd-b06b-98dbcefd790f&groupId=10124
- Cardenas V, Hochrainer S, Mechler R, Pflug G, and Linnerooth-Bayer J (2007) Sovereign financial disaster risk management: The case of Mexico. *Environmental Hazards* 7, pp. 40–53. Elsevier
- Cornell CA (1968) Engineering seismic risk analysis. *Bull Seismol Soc Am* 58(6):1583–1606
- Cummins JD (2008) CAT bonds and other risk-linked securities: State of the market and recent developments. *Risk Manag Insur Rev* 11(1):23–47
- Cummins JD, Mahul O (2009) Catastrophe Risk Financing in Developing Countries –Principles for Public

- Intervention. The International Bank for Reconstruction and Development/The World Bank, Washington DC
- Del Re D, Zhang L, Zhao Z, McEwan P (2008) Testing of a probabilistic model of seismic risk in China to enable micro-insurance viability and promote mitigation. In: Proceedings, 14th world conference on earthquake engineering, Beijing, China
- DesRoches R, Comerio M, Eberhard M, Mooney W, Rix GJ (2011) Overview of the 2010 Haiti earthquake. *Earthq Spectra* 27(S1):S1–S21
- Edmonds C (2013) Learning it the hard way: an analysis of cat bond investing pitfalls. In: Lane M (ed) *Alternative (re)insurance strategies*, 2nd edn. Risk Books, Chicago
- EQC Earthquake Commission (2013) Annual report 2012–2013. Accessible on-line at http://www.eqc.govt.nz/sites/public_files/documents/EQC011-Annual-Report-2013-A-WEB.pdf
- Esteva L (1970) Regionalización sísmica de México para fines de ingeniería. Instituto de Ingeniería, Universidad Nacional Autónoma de México, México
- Franco G (2010) Minimization of trigger error in cat-in-a-box parametric earthquake catastrophe bonds with an application to Costa Rica. *Earthq Spectra* 26(4):983–998
- Franco G (2013) Construction of customized payment tables for cat-in-a-box earthquake triggers as a basis risk reduction device. In: Proceedings, 11th international conference on structural safety & reliability, New York
- Franco G, Siembieda W (2010) Chile's 2010 M8.8 earthquake and tsunami: initial observations on resilience. *J Disaster Res* 5(5):577–590
- Franco G, Sheth A, Meyer M (2013) Observations on the recovery and reconstruction in Sri Lanka following the December 26, 2004 Tsunami. An Earthquake Engineering Research Institute (EERI) field report. Accessible on-line at <https://www.eeri.org/wp-content/uploads/WB-LFE-BBB-Sri-Lanka.pdf>
- Friedman DG (1975) Computer simulation in natural hazard assessment. University of Colorado: Program on Technology, Environment and Man. Monograph #NSF-RA-E-1 S-002
- Giné X, Menand L, Townsend R, Vickery J (2010) Microinsurance: a case study of the Indian rainfall index insurance market. The World Bank. Policy research working paper 5459. Accessible on-line at <https://openknowledge.worldbank.org/bitstream/handle/10986/4002/WPS5459.txt?sequence=2>
- Goda K (2013) Basis risk of earthquake catastrophe bond trigger using scenario-based versus station intensity-based approaches: a case study for Southwestern British Columbia. *Earthq Spectra* 29(3):757–775
- Grossi P, Kunreuther H (2005) Catastrophe modeling: a new approach to managing risk. Huebner international series on risk, insurance and economic security
- Guy Carpenter (2014) Chart: gap between economic and insured losses. GC Capital Ideas. Accessible on-line at <http://www.gccapitalideas.com/2014/01/20/chart-gap-between-economic-and-insured-losses/>
- Härdle WK, López Cabrera B (2010) *Calibrating CAT bonds for Mexican Earthquakes*. *J Risk Insur* 77(3):625–650. DOI: 10.1111/j.1539-6975.2010.01355.x
- Ibarra H (2009) Parametric insurance: general market trends and perspectives for the African insurance sector. Available at <http://www.africa-re.com/WEATHER INSURANCE.pdf>
- Linnerooth-Bayer J, Mechler R (2009) Insurance against losses from natural disasters in developing countries. DESA working paper No. 85. Accessible on-line at http://www.un.org/esa/desa/papers/2009/wp85_2009.pdf
- Major JA (1999a) Uncertainty in catastrophe models part I: what is it and where does it come from? Financing risk & reinsurance. International Risk Management Institute
- Major JA (1999b) Uncertainty in catastrophe models part II: how bad is it? Financing risk & reinsurance. International Risk Management Institute
- Marsh (2014) Comparing claims from catastrophic earthquakes. Accessible on-line at <http://uk.marsh.com/NewsInsights/MarshRiskManagementResearch/ID/36121/Comparing-Claims-from-Catastrophic-Earthquakes.aspx>
- McGuire RK (2008) Probabilistic seismic hazard analysis: early history. *Earthq Eng Struct Dyn* 37(3):329–338
- Mechler R, Linnerooth-Bayer J, Peppiatt D (2006) *Disaster Insurance for the Poor? A Review of Microinsurance for Natural Disaster Risks in Developing Countries*. ProVention Consortium & IIASA, Geneva, Switzerland/Laxenburg, Austria
- Michel-Kerjan E et al (2011) Catastrophe financing for governments: learning from the 2009–2012 MultiCat Program in Mexico. OECD working papers on finance, insurance and private pensions, No. 9, OECD Publishing. doi:10.1787/5kgecj7wkvhb-en
- Miller D (1999) Uncertainty in hurricane risk modeling and implications for securitization. Casualty Actuarial Society Discussion Paper Program
- Munich Re (2013) Significant natural catastrophes 1980 – 2012. Available at http://www.munichre.com/site/corporate/get/documents_E1241756001/mr/assetpool.shared/Documents/0_Corporate%20Website/2_Reinsurance/Business/Non-Life/Georisks/NatCatService/_Significant%20Natural%20Catastrophes/2012/natcatservice_significant_ins_en.pdf. Accessed Mar 2014
- Nanto DK, Cooper WH, Donnelly JM (2011) Japan's 2011 earthquake and tsunami: economic effects and implications for the United State. Congressional Research Service 7–5700, www.crs.gov. R41702
- Nguyen T (2013) Insurability of catastrophe risks and government participation in insurance solutions. Background Paper prepared for the Global Assessment Report on Disaster Risk Reduction 2013. Accessible on-line at <http://www.preventionweb.net/english/hyogo/gar/2013/en/bgdocs/Nguyen.%202012.pdf>

- Pacific Disaster Risk Financing and Insurance Program (2011) Overview document. Accessible on-line at http://siteresources.worldbank.org/INTLACREGTOP/URBDEV/Resources/840343-1319570618921/PDRFIS_pager.pdf
- Rutherford S (1999) The poor and their money: an essay about financial services for poor people. Institute for Development Policy and Management. University of Manchester. Accessible on-line at <http://www.jointokyo.org/mfdl/readings/PoorMoney.pdf>
- Siembieda W, Johnson L, Franco G (2012) Rebuild fast but rebuild better: Chile's initial recovery following the 27 February 2010 earthquake and tsunami. *Earthq Spectra* 28(S1):S621–S641
- Spence R, Bommer J, Del Re D, Bird J, Aydinoglu N, Tabuchi S (2003) Comparing loss estimation with observed damage: a study of the 1999 Kocaeli earthquake in Turkey. *Bull Earthq Eng* 1(1):83–113
- Standard & Poor's (2014) Global reinsurance highlights 2013. Standard & Poor's ratings services. www.standardandpoors.com
- SVV (2014) The Swiss Natural Perils Pool. Accessible on-line at <http://www.svv.ch/en/consumer-info/non-life-insurance/swiss-natural-perils-pool>
- Swiss Re (2013) A history of insurance. Authored for Swiss Re by Niels Vigo Haueter. Edited by Swiss Re Corporate History. Accessible on-line <http://150.swissre.com/about/history/>
- Wald DJ, Earle PS, Allen TI, Jaiswal K, Porter K, Hearn M (2008) Development of the U.S. Geological Survey's PAGER system (Prompt Assessment of Global Earthquakes for Response). In: Proceedings, 14th world conference on earthquake engineering, Beijing
- Wang Z, Lin T, Walker G (2009) Earthquake risk and earthquake catastrophe insurance for the People's Republic of China. Asian Development Bank. Munich Personal RePEc Archive MPRA paper #21248. Accessible on-line at <http://mpra.ub.uni-muenchen.de/21248/>
- Wirtz A (2013) Natural disasters and the insurance industry. In: Guha-Sapir D, Santos I, Borde A (eds) *The economic impacts of natural disasters*
- World Bank (2008) *The Caribbean Catastrophe Risk Insurance Facility: providing immediate funding after natural disasters*, vol 2, issue 1, Operational innovations in Latin America and the Caribbean. The World Bank, Washington, DC, pp 1–23
- World Economic Forum (2008) Convergence of insurance and capital markets. World Economic Forum. Accessible on-line at http://www3.weforum.org/docs/WEF_InsuranceCapitalMarketsConvergence_Report_2008.pdf
- World Forum of Catastrophe Programs (2014) Natural catastrophe cover systems – comparative table. Accessible on-line at http://www.wfcprogrammes.com/generalpresentation?p_p_id=101_INSTANCE_MNZ3&p_p_action=0&p_p_state=normal&p_p_mode=view&p_p_col_id=column-2&p_p_col_count=1&_101_INSTANCE_MNZ3_struts_action=/

Earthquake Protection of Essential Facilities

William T. Holmes¹ and Robert K. Reitherman²

¹Rutherford + Chekene Engineers,
San Francisco, CA, USA

²Consortium of Universities for Research in
Earthquake Engineering, Richmond, CA, USA

Introduction

Essential facilities, such as hospitals, emergency communications centers, and fire stations, are intended to provide an essential service immediately following an earthquake, and thus, a key criterion in their design that distinguishes these facilities from ordinary occupancy construction is the protection of their functionality. The hospital should continue to provide medical care; the dispatching of emergency vehicles should continue without an interruption caused by damage to a communications facility; and fire stations should be able to respond to the need for fire fighting, search and rescue, and hazardous materials releases. There are also nongovernmental facilities whose services or products are so valuable to a company that functionality is protected with on-site structural and nonstructural measures discussed here. Multifacility organization-wide or corporate-wide planning has also been employed so that redundant facilities are sited so that they are not all in the same seismic region.

Building codes as well as the design standards for many components of utility and transportation systems typically set safety as a minimum objective. In some cases, for example, an art museum or historic building, protection of the property itself is another criterion that is also included. As a practical matter, achieving the objective of post-earthquake functionality typically requires meeting and exceeding those two other kinds of objectives, safety and property protection, making it difficult to achieve. If controlling damage to a very low level and also providing backup



Earthquake Protection of Essential Facilities, Fig. 1 One of two completely collapsed buildings at the Veterans Administration Hospital, 1971 San Fernando

earthquake (Source: US Army Corps of Engineers, NISEE-PEER library, University of California, Berkeley)

on-site utility services were easy to achieve without a significant cost premium, most facilities would already be designed for this performance level, whereas in fact, comprehensive protection of functionality of a facility is relatively rare.

Historic Examples

There have been many examples of earthquake disruption of essential facility operations, dating back decades and continuing to the present. One of the earliest particularly instructive earthquakes was the 1971 San Fernando earthquake in the Los Angeles area. Seven major medical facilities were badly damaged: Foothill Medical Center, Pacoima Lutheran Medical Center, Pacoima Memorial Lutheran Hospital, Holy Cross Hospital, Indian Hills Medical Center, the Sylmar Veterans Administration (VA) Medical Center, and Olive View Medical Center (Jennings 1971; Lew et al. 1971; Johnston and Strand 1973). As a result, 1,553 hospital beds were out of service (Oakeshott 1975), and instead of being able to provide the post-earthquake supply of medical care for injuries in the area, damaged hospitals

became additional demand on the medical and rescue resources of other facilities. The Veterans Administration complex of buildings in Sylmar was the scene of two complete structural collapses. See Fig. 1. At the Olive View Medical Center nearby, the 750-bed main hospital building suffered such severe soft story damage that it was evacuated and torn down, only 2 months after it had been built (Fig. 2). In addition, the two-story Olive View Psychiatric Building at that complex experienced a complete ground-story collapse (Fig. 3); several unreinforced masonry buildings on the grounds collapsed; damage to the fire station located on the grounds required lengthy extrication of the equipment; reinforced concrete canopies collapsed onto ambulances (Fig. 4); damage to the backup power system (Fig. 5) and to the entire array of nonstructural components – elevators, windows, partitions, ceilings, piping, and medical gasses – occurred (Ayres and Sun 1973); and the equipment and piping in the on-site power plant building was so damaged that its functions were not available (Fig. 6).

Following the San Fernando earthquake, the Veterans Administration adopted seismic



Earthquake Protection of Essential Facilities, Fig. 2 Destruction of a non-ductile reinforced column, main building, Olive View Medical Center, 1971 San Fernando Earthquake (Source: Karl V. Steinbrugge Collection, NISEE-PEER library, University of California, Berkeley)

standards for new facilities that have set the basic standard for protecting essential facilities, including these provisions: structures with more strength and stiffness than normal buildings to minimize damage, protection of nonstructural components to minimize internal disruption, and particular attention to protection of critically important equipment and utilities, including on-site backup capability such as emergency generators and water storage. The Veterans Administration, having a system of healthcare facilities across the United States, also started a program of evaluation of existing facilities and developed new design criteria for new ones (Veterans Administration 1972; Bolt et al. 1975). The state of California enacted the Hospital Seismic Safety Act in 1972 that covered non-federal-government hospitals with similar

requirements (Holmes 2002), and that act clearly stated the goal of keeping hospitals functioning after earthquakes:

It is the intent of the Legislature that hospitals, which house patients who have less than the capacity of normal healthy persons to protect themselves, and which must be reasonably capable of providing services to the public after a disaster, shall be designed and constructed to resist, insofar as practical forces generated by earthquakes, gravity, and wind.

In both approaches, merely setting code criteria was only the beginning, because implementation of the criteria requires a detailed level of quality assurance well beyond that employed for the design and construction of an ordinary occupancy facility. Because at a given point in history the majority of essential buildings that are in existence will almost always greatly outnumber the number of new buildings that will be built in the next few years, to improve the earthquake protection of a region's inventory of essential facilities requires dealing the difficult problems of retrofits to existing construction. In 1994, following the 1994 Northridge earthquake, California took that step in passing SB (Senate Bill) 1953. Although this has resulted in many hundreds of millions of dollars of cost among California's hospitals to upgrade existing hospitals or replace them with new ones, in the process, not only their safety but their ability to function after earthquakes has been greatly improved.

In another earthquake of the early 1970s, the 1972 Nicaragua earthquake that caused great damage in the capital city of Managua, the scope of damage and disruption to essential facilities was extensive (see Fig. 7):

The damage in Managua provides many examples of the need to construct buildings housing essential facilities in a manner such that they will be functional when desperately needed following an earthquake. Most critical facilities in Managua were destroyed. Examples are [figure citations deleted here]: the fire station which collapsed upon the fire equipment, the Red Cross building, which collapsed upon their ambulances and emergency supplies, the INSS Hospital, which suffered severe structural and nonstructural damages making it not just useless, but definitely hazardous to its occupants. The General Hospital also suffered very severe damage. ... (Wright and Kramer 1973, p. 19)



Earthquake Protection of Essential Facilities, Fig. 3 Collapse of ground story of psychiatric building, Olive View Medical Center, 1971 San Fernando

earthquake (Source: Karl V. Steinbrugge Collection, NISEE-PEER library, University of California, Berkeley)

Earthquake Protection of Essential Facilities, Fig. 4 Destruction of ambulances by collapsed concrete canopy, Olive View Medical Center, 1971 San Fernando earthquake (Source: Karl V. Steinbrugge Collection, NISEE-PEER library, University of California, Berkeley)



Many more examples could be cited – see FEMA (2007) – but in this brief review, these two earthquakes are adequate to illustrate the range of disruptive effects, and the 1971 and

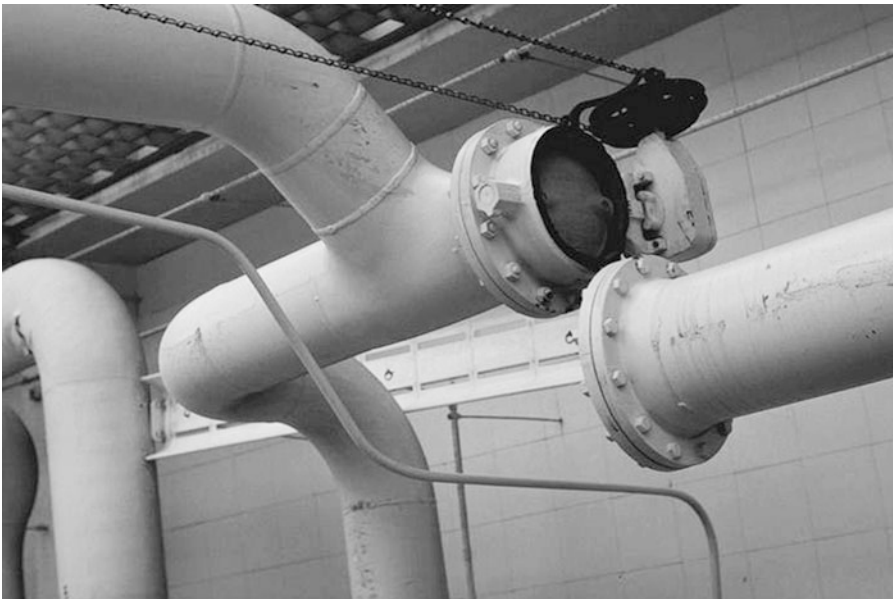
1972 earthquakes also approximately date the beginning of the era when focused attention has been given to protecting essential facility functions. Also during this post-1971 or

Earthquake Protection of Essential Facilities,

Fig. 5 Damage to unrestrained batteries, necessary to start the emergency power generator, Olive View Medical Center, 1971 San Fernando earthquake (Source: Karl V. Steinbrugge Collection, NISEE-PEER library, University of California, Berkeley)



E



Earthquake Protection of Essential Facilities,
Fig. 6 Broken piping, Olive View Medical Center, 1971 San Fernando earthquake (Source: Karl

V. Steinbrugge Collection, NISEE-PEER library, University of California, Berkeley)

post-1972 era, an awareness of the need for functional protection of other kinds of essential facilities evolved. In California, the Essential Services Buildings Seismic Safety Act was passed in 1986.

Structural Protection

Better performance of the structure of an essential building has often been sought in the provisions of seismic codes by increasing the

Earthquake Protection of Essential Facilities,

Fig. 7 Collapsed fire station in Managua, 1972
Nicaragua earthquake
(Source: Karl V. Steinbrugge Collection, NISEE-PEER library, University of California, Berkeley)



design forces. The Uniform Building Code, which was the common code of the Western United States until the inauguration of the International Building Code in 2000, an I or Importance Factor of 1.5 was introduced in the 1976 edition having the effect of increasing the design forces by 50 %. Because inelastic levels of response are still anticipated even for these stronger buildings if severe ground shaking occurs, only increasing the elastic-based design forces is currently thought to be helpful but not sufficient in attaining the performance level of remaining operational. Nuclear power plants are typically designed with the objective of providing so much strength that their structural, as well as key nonstructural, components remain elastic, but that specialized topic is covered in another article and that level of protection is generally infeasible – too expensive and too restricting on the configuration of ordinary buildings. Some inelastic behavior in the structure does not necessarily mean that the operation of the facility has been greatly curtailed, if the design has confined the inelasticity, i.e., damage, to portions of the structure that can absorb that earthquake punishment while still allowing the structure to remain occupiable and operational. A key criterion, and a difficult one to meet, is preventing interstory

drift from reaching levels that result in permanent distortion of the overall structure, a damage state that triggers evacuation.

Although it carries a cost premium, seismic isolation has been increasingly employed since it was introduced in modern form in the 1980s, the first isolated building in the world being the William Clayton Building in Wellington, New Zealand, completed in 1982. Typically, reducing the response of the building to a third of what it would otherwise be, isolation can provide high levels of functional protection, assuming the structure and nonstructural systems are still designed to appropriate levels for that degree of motion. As noted below, because nonstructural components are typically more fragile than the structure, a level of earthquake motion in a building “only” one-third as great as that of the ground’s unreduced level of motion can still be damaging unless appropriate design measures are implemented.

Nonstructural Protection

While it is necessary for the structure of the essential facility to remain safe and serviceable if the facility’s functionality is to be maintained,

it is not sufficient. There are many nonstructural items that must be protected from damage as well, such as emergency power generators, ceilings, fire sprinkler pipes, and elevators, to name but a few. The bare structure of a building, or the supporting framework or tower of a utility component, by itself provides no essential service. The degree of difficulty in protecting essential functions thus is greatly increased beyond the already formidable performance level of confining any structural damage to a negligible level, especially because most nonstructural components are more damageable, that is, they have higher fragilities than the structure itself. Unsecured reagents can fall off the shelves in a laboratory, which can cause a hazardous material release or fire, or both, at very low levels of shaking, to cite one example.

The first comprehensive study of the effects of an earthquake on nonstructural components was for the 1964 Alaska earthquake (Ayres et al. 1973), and that study can still be consulted for a valid overview of the range of damage. Because of the difficulty of providing comprehensive nonstructural protection, unfortunately, many subsequent earthquakes have provided the same basic sorts of examples of damage. Protecting nonstructural components, a topic discussed in a separate article, requires a major effort to go beyond the objective of safety for the occupants and prevention of serious damage to the structure. In a sense, structural engineers deal with one coherent entity in the design of the structure, an integrated structure shown entirely on the structural design drawings and documents, whereas the nonstructural items number in the hundreds or thousands and they are not all under the purview of one design professional. The architect, electrical and mechanical engineers, and specialty designers, such as for fire protection systems or elevators, each separately design nonstructural components. The design professional with the most expertise with regard to bracing and anchorage of nonstructural components, the structural engineer, is only employed when regulations or special attentiveness on the part of the client set that requirement. (Bracing, for example, refers to diagonal struts bracing

a rack containing equipment; anchoring refers to anchor bolts connecting the rack to the floor and or wall.) Adding to the problem is the fact that many kinds of equipment are introduced by the occupant after the building is built and outside the purview of the building code or a design professional, such as the furnishings, file cabinets, and computer equipment in an office setting or the chemicals and related apparatus in a laboratory. Making the problem even more difficult is the fact that even if the design professionals collectively cover all of the necessary nonstructural protective measures, there are then multiple specialty construction trades involved who build those features into the facility. This entry cannot delve deeply into the topic of nonstructural protection, other than to cite in passing a basic evaluation and design manual on the subject, *Reducing the Risks of Nonstructural Earthquake Damage* (FEMA 2011), and a study of the difficulties encountered in developing a coordinated approach to implementing nonstructural protection (Masek and Ridge 2009). The importance as well as the difficulty of ensuring adequate nonstructural performance in an essential facility can be briefly illustrated here using the example of the backup power system.

Picturing a modern telecommunications or medical facility, we immediately note that if the electrical power from the local utility is disrupted by the earthquake – a very plausible earthquake planning assumption – the facility cannot function, unless its own backup power system operates reliably. There are at least six components to an emergency electrical power system; the disabling of any one of which will cause the entire backup power system to fail. Batteries are needed to start the motor-generator set, and these heavy batteries, which thus experience large inertial forces in the earthquake, are only strongly braced if earthquakes are specifically taken into account. The massive motor-generator set itself, often weighing as much as a truck in the case of systems that must supply large amounts of power such as for a hospital, is typically mounted on vibration isolation springs. The springs are designed to filter out the very high frequency vibration of the motor, which would otherwise

propagate through the building and bother occupants, but when the floor is subjected to the lower frequency motions with large displacements of the earthquake, spring-mounted equipment can violently lunge off its supports. Seismically designed products that solve both the motor vibration and seismic response problem are the typical solution today, and existing mounts have been retrofitted with restrainers (“snubbers”). If the batteries remain stable and the motor generator is adequately restrained, the system still needs to have a functional exhaust flue – breakage of which has resulted in filling the generator room with fumes that choke off the engine. In some cases, a generator has been reliably run in tests or for brief power outages, but when run continuously for a day or more, it overheats. The heat from the motor must be dissipated, in the case of large units with water-fed radiator systems. The electrical panel that distributes the power through the facility is another critical component, and on-site fuel, such as oil for a diesel generator, sufficient for several days or longer is necessary as well. In places and eras where providing protection of essential functions is a new approach that is being phased in, singling out the emergency power system has been a strategic place to start (Applied Technology Council 1990). Another way to look at the seismic protection of backup power systems is as a means to evaluate the effectiveness of seismic protection programs. If discussions and planning meetings have gone on for more than a year about how to deal with all the various potential kinds of damage and outages in an essential facility, but if nothing has been done to evaluate and if needed retrofit its backup power system, the program’s effectiveness could be questioned.

Mentioned again here, although it is the topic of a separate article, is the nuclear power plant. The typical seismic design basis is aimed at keeping both structural and nonstructural components elastic, a design approach prohibitively expensive for most essential facilities. Expensive shake table testing has qualified or certified key types of equipment for that industry, and although testing standards vary, some of the sophisticated types of analysis, testing, and

design in that industry have been gradually adapted for use with less critical facilities. In some cases today, key kinds of equipment have been subjected to shake table qualification testing, an approach only relatively recently introduced into some codes and standards, such as the 2005 edition of ASCE 7, the standard referred to by building codes in the United States for determining design loads.

Backups for Essential Systems

Essential buildings are normally dependent on one or more essential systems to provide adequate functionality after an earthquake. Almost all such buildings require electrical power to be delivered to its communications, computer, and other equipment. Hospitals are a special case and in addition require water, medical gasses, and in some cases, waste disposal to continue care for their patients and incoming injured.

Community or regional power systems are susceptible to damage and interruption from earthquake shaking. On-site backup power generation systems, including the generator, starting batteries, fuel and day tanks, cooling systems, and exhaust systems, are typically provided at facilities such as hospitals and fire stations for this reason. However, beyond the anchorage and bracing of this equipment for seismic purposes, discussed above, the amount of on-site fuel must be considered. The California Hospital Program requires enough fuel to run essential systems for 36 hours. The Pan American Health Organization (PAHO), considering hospitals in less-developed countries, recommends fuel for five full days. Almost all hospitals have emergency generators to cover their essential systems for normal system interruptions, which are frequent in some regions, but these outages are typically only hours long, and sufficient fuel for multiple days is often not provided or locally available. In addition, when generators are installed to run for hours, they will often overheat if run continuously for a full day or more. The adequacy of the cooling system therefore must also be considered for generators for essential buildings.

Uninterrupted Power Supply (UPS) systems are often required for computer and/or communications systems and for certain medical equipment in hospitals. These systems are heavy and must be carefully braced and anchored to maintain function during and after earthquake shaking.

Water is another utility that is often necessary for the continued function of an essential facility and also one that is susceptible to interruption of the off-site supply. Besides drinking water, water may also be required at the site of essential facilities for cooling or operation of equipment and for fire sprinklers. For hospitals, PAHO recommends that a hospital site have an emergency supply of 300 l per bed per day, for 3 days. Pumps and on-site filtration, if required, may also require power and supplies that must be provided in the post-earthquake setting.

Most essential facilities are dependent on effective communication after an earthquake. Hard line and cell telephone may not be functional, and backup systems should be provided. Equipment racks used for communication equipment must be able to withstand the shaking. Radio frequencies should be checked for compatibility with outside entities that must be contacted. Shortwave radio and satellite telephones are possible backup systems. Large or multi-site facilities may also need to maintain internal communication, and the viability of the existing systems should be considered.

Access to off-site equipment and supplies may also be cut off after an earthquake. Vehicles that may be needed in the post-earthquake environment must be protected and available. Adequate quantities of supplies needed for post-earthquake function should be stored on-site or reliably available. For example, supplies of medical gasses (oxygen, nitrogen, compressed air) are needed for many functions in a hospital, and medical stores for 3–5 days should be available. Multi-facility systems, for example, a collection of hospitals run by the same organization or agency or several electronics plants or data processing centers of a company, can include in their planning how the services of one or more facilities isolated by a transportation or other off-site outage, or

disabled by on-site damage, can be temporarily supplied by redundant facilities located elsewhere.

Emergency Response

The severity of ground motion from an earthquake is typically unevenly distributed over a region, based on distance from the rupturing fault, direction and sense of fault rupture, local soils and deeper geology, and topography. Vulnerabilities of facilities are also highly variable, if only because the facilities in a region were built at different times to varying seismic criteria. This leads to the likelihood that some essential facilities will be significantly damaged and nonfunctional, while others will remain intact. Mutual aid planning and emergency response agreements are an effective technique for coping with the outages that do occur and capitalizing on those facilities that remain online.

In parallel with the engineering approach of preventing damage is the emergency response approach of quickly controlling damage and providing alternate means of providing service. The key difference between an exercise of a hospital disaster plan for an earthquake scenario as compared to one based on a presumed plane crash or other incident is that the hospital that is expected to suddenly provide an enhanced level of emergency service to the community will simultaneously find itself in a disruptive situation. The hospital that cannot protect itself from disruptive damage in the earthquake is of little use and perhaps a source of demand for emergency services rather than a supplier. Thus, to be realistic, seismic emergency response procedures cannot merely be copied from procedures for other kinds of disasters that do not simultaneously affect the region and the individual facility.

Three examples of communication outages and their effect on emergency response are cited to illustrate the issues involved. In the 1971 San Fernando earthquake, it took 1 h and 22 min before the key response agency, the fire department, knew that two buildings had collapsed at the Veterans Administration Hospital, a hospital

that was part of the urbanized Los Angeles area and was not in an isolated area (Reitherman 1986). Telephone service was out, radio communication was not functional, and although the in-person messenger sent to the nearest fire station to report the need for rescue operations at the hospital delivered that message, it was then confused with the problems at the nearby Olive View Hospital. In retrospect, the single biggest emergency response need in that earthquake was at the VA hospital, where 49 of the 58 fatalities in the earthquake occurred and a massive rescue operation was needed.

Eighteen years later, in the 1989 Loma Prieta earthquake in northern California, the most badly damaged hospital was Watsonville Community Hospital, which had to evacuate its buildings and provide services for in-patients and arriving emergency patients in the parking lot. Outside aide was not requested for several hours, because initial news reports that the hospital heard made it seem that the San Francisco metropolitan area 100 km away was totally devastated and hence aid resources would be focused there. In fact, there were many outside resources that eventually made their way to the hospital when its needs were accurately known.

In the 1994 Northridge earthquake in the Los Angeles area, water lines at the Sepulveda Veterans Administration Medical Center ruptured, which with efficient emergency response should result in prompt water shutoffs of affected zones to minimize water damage and disruption. Instead, patients and personnel, including facilities engineers, were evacuated and did not reenter the building until structural engineers arrived several hours later and declared the building safe for emergency operations such as turning off the water. Effective emergency response planning should include checklists for rapid utilities actions, such as water and gas shutoffs if certain observable criteria are met, and also for rapid safety inspections by engineers under prearranged plans.

New facilities intended for critical post-earthquake functions can be designed in accordance with recommendations discussed here and international standards, but older existing facilities are much more difficult to prepare for adequate

post-earthquake service. Structural retrofit is often expensive, and replacement may take years to plan and build. But honest evaluation of expected structural and nonstructural performance will lead to necessary and often affordable improvements, and more importantly, development of plans to offset functional difficulties.

Summary

Earthquakes have caused disruption of the services provided by essential facilities such as hospitals, emergency communications centers, and fire stations. For the functions of these facilities to be available in the post-earthquake setting, the structure must perform safely and key nonstructural components must be protected. In addition to these engineering approaches, emergency response planning is necessary to identify off-site lifeline systems that could be disrupted by the earthquake, because a facility may need backup systems and supplies to contend with outages to the region's water, power, and transportation systems.

References

- ATC (1990) General acute care hospital earthquake survivability inventory for California, funded by the Office of Statewide Health Planning and Development. Applied Technology Council, Redwood City
- Ayres JM, Sun TY (1973) Nonstructural damage. The San Fernando, California Earthquake of February 9, 1971. National Oceanic and Atmospheric Administration, Washington, DC
- Ayres JM, Sun TY, Brown FR (1973) Nonstructural damage to buildings. The great Alaska earthquake of 1964: Engineering. National Academy of Sciences, Washington, DC
- Bolt B, Johnston RG, Lefter J, Sozen MA (1975) The study of earthquake questions related to Veterans Administration hospital facilities. *Bull Seismol Soc Am* 65(4):937–949
- FEMA (2007) Design guide for improving hospital safety in earthquakes, floods, and high winds, FEMA 577. Federal Emergency Management Agency, Washington, DC
- FEMA (2011) Reducing the risks of nonstructural earthquake damage – a practical guide, FEMA E-74. Federal Emergency Management Agency, Washington, DC

- Holmes WT (2002) Background and history of the seismic hospital program in California. In: Proceedings of the seventh national earthquake engineering conference, Earthquake Engineering Research Institute, Oakland
- Jennings PC (ed) (1971) Engineering features of the San Fernando Earthquake of February 9, 1971, California Institute of Technology, Earthquake Engineering Research Laboratory, Pasadena, EERL 71-02
- Johnston RG, Strand DR (1973) Olive view hospital. In: Murphy L (ed) San Fernando, California, earthquake of February 9, 1971, vol I-A. National Oceanic and Atmospheric Administration, Washington, DC
- Lew HS, Leyendecker EV, Dikkers RD (1971) Engineering aspects of the 1971 San Fernando Earthquake. National Bureau of Standards, Building Science series 40. U.S. Government Printing Office, Washington, DC
- Masek J, Ridge R (2009) Identification of methods to achieve successful implementation of nonstructural and equipment seismic restraints. Earthquake Engineering Research Institute, Oakland
- Oakeshott G (ed) (1975) San Fernando Earthquake, February 9, 1973. California Division of Mines and Geology, Sacramento
- Reitherman R (1986) How to prepare a hospital for an earthquake. *Disaster Manag* 4:119–131
- Veterans Administration (1972) Post earthquake emergency utility services and access facilities, CD-54. Veteran's Administration, Washington, DC
- Wright R, Kramer S (1973) Building performance in the 1972 Managua Earthquake. National Bureau of Standards, Washington, DC

shaking at a specified site will exceed some intensity-measure level of engineering interest in a given time period t . The results of a PSHA have a broad spectrum of end users, ranging from decision-makers who have to provide new codes for designing safe buildings to engineers who must predict potential damage to critical structures. Among all the variables involved in the PSHA computation, this section will focus on the description of the probability models used to build earthquake rupture forecast (ERF) from earthquake rate model in the area under study. ERF allows a modeling of the probability of occurrence of all possible damaging earthquakes in a given region during a specified time span and represents the basis for PSHA. In practice, an incorrect or incomplete representation of the seismicity in space, time, and size may bias the analysis leading to underestimation or overconservatism in the final hazard assessment and consequential direct and nontrivial implications for the compilation of regulations for seismic design.

The Physics of the Problem: From Deterministic Toward Probabilistic Perspective

The elastic rebound theory proposed by Reid (1910) after the 1906 San Francisco earthquake suggests that large tectonic earthquakes occur when the stress exceeds a critical value of the rock strength, producing fractures and relieving crustal strain. In a given area, stress builds up due to tectonic processes ultimately connected to the plate movements. Thus, assuming a constant rate in the stress buildup, the critical value and the consequent earthquake nucleation will be reached at more or less regular intervals. This periodic buildup and release of stress on a given fault represents what is known as *seismic cycle* whose duration is given by the ratio of event-strain release to tectonic strain rate. More precisely, a seismic cycle controls the long-term timing of large earthquakes and comprises three different stages, namely, (i) the interseismic stage during which there is an accumulation of strain

Earthquake Recurrence

Vincenzo Convertito¹ and Licia Faenza²

¹Istituto Nazionale di Geofisica e Vulcanologia, Osservatorio Vesuviano, Napoli, Italy

²Istituto Nazionale di Geofisica e Vulcanologia, Centro Nazionale Terremoti, Bologna, Italy

Synonyms

Hazard function; Recurrence time; Seismic hazard; Time-dependent recurrence models; Time-independent recurrence models

Introduction

Probabilistic seismic hazard analysis (PSHA) aims at estimating the probability that ground

energy; (ii) coseismic stage, which corresponds to the stage of principal rupture; and (iii) post-seismic stage during which a new equilibrium is reached. However, comparison of seismic slip distribution with potential slip inferred from geodetic motion measurements demonstrates that not all the stored energy is released through seismic slip but there is a deficit which is attributed to fault creep and aseismic deformation (Bakun et al. 1986).

In this respect, earthquake recurrence refers to the time between subsequent events rupturing on a given segment fault, hence to the period of the loading cycle. Thus, in theory one would think that it is possible to predict the occurrence of future events if one knows the previous history. Indeed, the scheme described so far provides a deterministic representation of the occurrence of a single earthquake on a given structure and may be too simple to depict the real nature of the earthquake recurrence. In this respect, Nishenko and Buland (1987) described three possible factors for the observed variation in the recurrence period. A first factor is a possible change in source size from event to event which would require a different time for restoring the initial stress value, a second factor is a nonlinear strain accumulation in which the strain field initially recovers rapidly soon after the event and then progressively lowers down until the next event, and a third factor is given by the faults' interaction in which strain energy released on one fault segment affects the strain on nearby faults. In addition, the uncertainties of estimations of the earthquake parameters, and the relatively shortness of the seismic records available, make the definition of the seismic cycle even more complicated, leading to a chaotic behavior of the sequences. In this perspective, models of earthquake occurrence are developed in the framework of the probability theory, on which also the PSHA is established.

From a practical point of view, when performing a PSHA, once the seismic zones which represent the potential threat for the site of interest have been identified and modeled (as points, lines, areas, volumes, or single faults) based on the available data, the recurrence of the

earthquakes, in particular of their magnitude should be defined. Two recurrence relationships are generally used: the Gutenberg-Richter (GR) relationship (Gutenberg and Richter 1944) and the characteristic earthquake model (Schwartz and Coppersmith 1984). The GR model assumes that in a given seismic zone, events of all magnitudes (from a minimum to a maximum magnitude value) may occur and are expressed through the relationship:

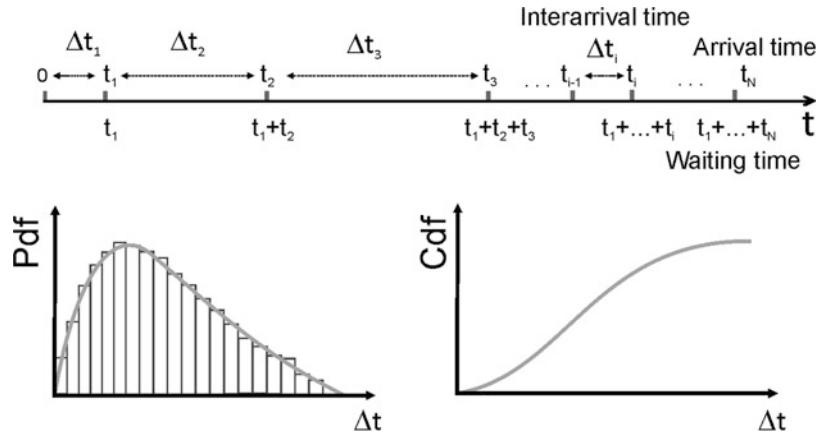
$$\log N(m) = a - bm \quad (1)$$

where $N(m)$ is the cumulative number of events having magnitude larger than m occurring in a given time. The parameter a depends on the number of events in the selected time and region. The slope parameter b , also known as “ b -value,” provides an estimate of the ratio between the number of small and large events and is generally about one. It can be shown that, when the GR is assumed as recurrence relationship, the probability density function of magnitude is the exponential function $f(m) = e^{(\alpha - \beta m)}$ with $\alpha = a \ln 10$ and $\beta = b \ln 10$. As this function covers an infinite range of magnitude, from $-\infty$ to $+\infty$, for practical applications, a lower and upper bounded version is used, which requires the identification of both a minimum and a maximum magnitude value. The definition of the minimum magnitude depends on the catalog completeness, while the definition of the maximum magnitude is still an open question, and its quantification may play an important role in PSHA for the mitigation against very large and rare events.

The characteristic earthquake model is instead based on the assumption that individual faults and fault segments tend to generate similar size or characteristic earthquakes having a relatively narrow range of magnitudes near the maximum with frequency of occurrence larger than that predicted by simply extrapolating the GR (Schwartz and Coppersmith 1984). The resulting probability density function of magnitude to be used in the PSHA has a more complex formulation with respect to the exponential probability density function derived from the GR (Young and Coppersmith 1985).

Earthquake Recurrence,

Fig. 1 Definition of the variables generally used to analyze time distribution of earthquake occurrence (*upper panel*). Lower panels depict a schematic representation of a possible distribution of observed inter-event times together with the estimated probability density function (*pdf*) and the corresponding cumulative function (*cdf*) (see text for the definition of pdf and cdf)



One of the assumptions of the PSHAs, derived from the Reid (1910) theory, is that the seismicity is stationary; as a consequence, the recurrence relationships derived for the current available seismic catalog are valid also in the future. However, as mentioned above, this assumption may not be valid on the geological time scale, thus requiring models that account for nonstationary behavior. As a fully deterministic approach is not yet feasible, the solution is to refer to the earthquakes as the outcome of stochastic processes. In particular, the earthquake occurrence is regarded as *renewal process* (Cornell and Winterstain 1988) in five dimensions: time, magnitude, and location. It is obvious that, given this assumption, the physical processes governing rupture initiation, propagation, and stopping are neglected. On the other hand, real nature of the earthquake process is too complex to be modeled by using deterministic approaches.

In a renewal process, the times between subsequent events are considered independent and identically distributed random variables. This implies that the expected time till the next event does not depend on any of the details of the last event but its time of occurrence. Once the earthquake occurs, the renewal process resets from the beginning.

Although a complete statistical analysis of earthquakes should consider the distribution of all the five dimensions (i.e., time, magnitude, and location), the physical variable largely used to make statistical inferences is the inter-event time, $\Delta t_i = t_i - t_{i-1}$, i.e., the time between two

consecutive shocks. Statistical distribution of Δt_i is analyzed to determine if the observed seismicity can be modeled by using a specific function (Fig. 1).

For simplicity, the inter-event time, Δt , will be indicated as t throughout the text. Inspection of the distribution of the observed t may help to establish some features of the seismicity, such as the presence of any regularity in time and possibly the influence of different tectonic/physics factors that regulate the spatial occurrence of earthquakes. For example, a relatively large percentage of small values of inter-event times may suggest that, in the area under study, earthquakes tend to cluster in time and thus cannot be modeled as statistically independent events.

Formally, the inter-event time distribution is defined unambiguously by its probability density function $f(t)$ (pdf). By definition, the pdf has to be positive in the range of existence of the random variable and its integral under the whole space must be one. The integral between $t = 0$ and $t = T$ defines the cumulative density function $F(t)$ (cdf), while integral between $t = T$ and $t = +\infty$ defines the survivor function $S(t)$, which represents the probability that the system will survive beyond time T , i.e., at least time T will elapse between consecutive events. The cumulative and the survivor functions are defined, respectively, as

$$F(T) = P\{t \leq T\} = \int_0^T f(u) du \quad (2)$$

$$S(T) = P\{t > T\} = \int_T^{+\infty} f(u)du \quad (3)$$

where $F(t)$ is a monotone not-decreasing function and $\lim_{t \rightarrow 0} F(t) = 0$ and $\lim_{t \rightarrow \infty} F(t) = 1$. Recalling the properties of the pdf, it is easy to find that $S(t)$ is the complementary function of $F(t)$. An alternative characterization of the distribution of the random variable t , widely used in seismic hazard calculation, is the hazard function $h(t)$, which is defined as the instantaneous rate of occurrence of the events:

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \quad (4)$$

where the numerator expresses the conditional probability that the event will occur in the interval $(t; t + \delta t)$ given that it has not occurred before, and the denominator is the – infinitesimal – width of the interval. The hazard function is not a density or a probability function. However, we can think of $h(t)\delta t$ as the conditional probability of failure in an infinitesimally small time period between t and $t + \delta t$, given that the considered system has survived up till time t . The concept of the hazard function is borrowed from the analysis of failure of engineering systems. By analogy, earthquakes may be regarded as mechanical failures of rocks that resist to failure through friction and cohesion. There is a biunivocal relationship between probability density function, the cumulative function, the survivor function, and the hazard function; nevertheless, the hazard function is often preferred because a simple analysis of its trend can provide useful insights on the physics of the process under study. Moreover, it has two additional properties: it is additive for two independent subprocesses, and if no event occurs between t and $t + u$ and $v \geq u \geq 0$, then $h_t(v) = h_{t+u}(v-u)$ (Zhuang et al. 2012).

An increasing trend would stand for an increase in the probability of occurrence with the time elapsed since the last event. This case would support, in general, the seismic gap hypothesis, as well as any general model based on the “characteristic earthquake” with almost constant recurrence times. On the other hand, a

monotonic decreasing trend of the hazard function would stand for a “clustered” sequence, where the probability of occurrence decreases with the time elapsed since the last large event. A constant hazard function characterizes the Poisson process; in such case the conditional probability of having an earthquake in a fixed time interval is “time independent,” i.e., it is independent from the time elapsed since the last event.

As mentioned in the introduction section, earthquake rupture forecast aims at modeling the probability of occurrence of all possible damaging earthquakes in a given region during a specified time span. The functions defined so far provide the tools to reach the goal. Indeed, the requested probabilities are the conditional probabilities to have one or more earthquakes in the next τ years of interest, given a time t since the last event; this can be

$$P(T \leq t \leq T + \tau | t > T) = \frac{S(T) - S(T + \tau)}{S(T)} \quad (5)$$

where $S(T)$ is the survivor function. In practical applications the time scales of interest are generally shorter than the mean recurrence intervals of the major faults present in the area of interest. Thus, several values of τ are selected based on the aim for which the analysis is conceived. For example, the reference seismic hazard map of Italy has τ equals 50 years based on the assumption that this time corresponds to the average lifetime of a civil structure.

From the point of view of modeling the temporal behavior of earthquakes, the recurrence models can be divided in two main classes. The first class contains the “memoryless” model, i.e., the Poisson model, whereas the second class includes the non-Poissonian models, i.e., time-dependent models. While the Poisson model is extensively used in PSHAs, its adequacy is limited since the mechanism of earthquake interactions at regional scale has highlighted the tendency to a spatiotemporal clustering of earthquake occurrence, even for moderate earthquakes. Moreover, due to the “memoryless”

property, it is not adequate to model the occurrence of earthquakes on an individual fault where non-Poisson models perform better. In the following section, the most common statistical models for earthquake occurrence are described. Because of the intrinsic complexity of the problem of earthquake forecasting, in PSHA usually several models are used at the same time, the contributions of which are then weighted using a logic tree scheme. On one hand, this approach represents the scarce knowledge about earthquake occurrence process, and it is highlighted by the fact that, in some cases, antithetical models are simultaneously used. On the other hand, it is a way to contemplate the complexity of the earthquake occurrence, since different probability models incorporate different physical attributes of the earthquake process that a single statistical distribution could not incorporate. Generally speaking, in any model there may be two types of uncertainty: *aleatory uncertainty* and *epistemic uncertainty*. The *aleatory* uncertainty is associated to the intrinsic stochastic nature of the process resulting in an unavoidable impossibility of predicting deterministically its evolution. The *epistemic* uncertainty represents the limited knowledge of the system. To the extent that a process is knowable, in principle its epistemic uncertainty is reducible, while *aleatory* uncertainty is not. The inclusion of different models for earthquake forecasting, and the adoption of a logic tree scheme, is a way to quantify the epistemic uncertainty of the process.

Time-Independent Models

The Poisson Model

Because of its simplicity and small number of parameters to be estimated, the Poisson model has been long used to model seismicity, which, on different space and time scales, shows a high degree of randomness. Moreover, it is still one of the most used models in standard PSHA applications. The Poisson model provides a simple scheme for estimating the probability of occurrence of events governed by a Poisson process.

A counting process is said to be a Poisson process with rate λ , $\lambda > 0$, if it is characterized by:

- (i) *Stationarity*. The probability of an event in a short interval of time t to $t + h$ is approximately λh , for any t .

$$P\{N(h)=1\} = \lambda h + o(h) \quad (6)$$

- (ii) *Nonmultiplicity*. The probability of two or more events in a short interval of time is negligible compared to λh .

$$P\{N(h) \geq 2\} = o(h) \quad (7)$$

- (iii) *Independence*. The number of events in any interval of time is independent of the number in any other (nonoverlapping) interval of time. The process is thus said to be “memoryless.”

- (iv) The number of events in any interval of length t is Poisson distributed with mean λt . That is, for all $s, t \geq 0$,

$$P\{N(t+s) - N(t) = k\} = \frac{(\lambda s)^k e^{-\lambda s}}{k!} \quad (8)$$

for $k = 0, 1, 2, \dots; \lambda > 0$

where λ is the rate of occurrence of events, i.e., the number of events in unit of time.

Starting from (iv), it can be shown that, for a Poisson process, the inter-event times τ (or Δt) are exponentially distributed with cdf $F(\tau)$ and pdf $f(\tau)$ given by

$$F(\tau) = 1 - e^{-\lambda \tau} \quad \tau \geq 0, \lambda > 0 \quad (9)$$

$$f(\tau) = \lambda e^{-\lambda \tau} \quad \tau \geq 0, \lambda > 0 \quad (10)$$

Using Eqs. 9 and 10, it is possible to calculate the hazard function $h(\tau)$, which for the specific model is given by

$$h(\tau) = \frac{f(\tau)}{1 - F(\tau)} = \lambda \quad (11)$$

This result implies that the probability of occurrence of earthquakes in a future small increment

of time remains constant regardless of the elapsed time since its occurrence or the size of the last event (in case of spatial Poisson model).

The shape of the pdf and the hazard function for three different values of λ are shown in Fig. 2, panel a. As previously noted, the hazard function is a constant that depends only on the rate of occurrence of the events in the area under study or on the fault, i.e., the number of target earthquakes in the unit of time. Areas or faults with higher rate of occurrence have thus higher probability to experience the next earthquake, regardless the time occurred since the last one.

A direct generalization of the Poisson process is the *nonhomogeneous Poisson process*. As the regular Poisson process, it is based on the assumption of independent increments and at most one occurrence at any instant of time. Different from the homogeneous Poisson process, rate parameter λ depends on time, i.e., $\lambda(t)$. For a nonhomogeneous Poisson process, the probability of having n events in a given time interval (t_1, t_2) is given by

$$P[N(t_2) - N(t_1) = n] = \frac{\left[\int_{t_1}^{t_2} \lambda(u) du \right]^n}{n!} e^{-\left[\int_{t_1}^{t_2} \lambda(u) du \right]} \quad n = 0, 1, 2, \dots, \quad (12)$$

where $\lambda(t)$ is the instantaneous rate of occurrence and $\int_{t_1}^{t_2} \lambda(u) du$ is the expectation value over the considered time interval.

The inter-event time random variable for a set of events modeled as a nonhomogeneous Poisson process has the following cumulative distribution:

$$F(\tau) = 1 - e^{-\left[\int_0^\tau \lambda(u + t_s) du \right]} \quad \tau \geq 0 \quad (13)$$

where t_s is the time since the last event. Once the appropriate $\lambda(t)$ has been identified by analyzing

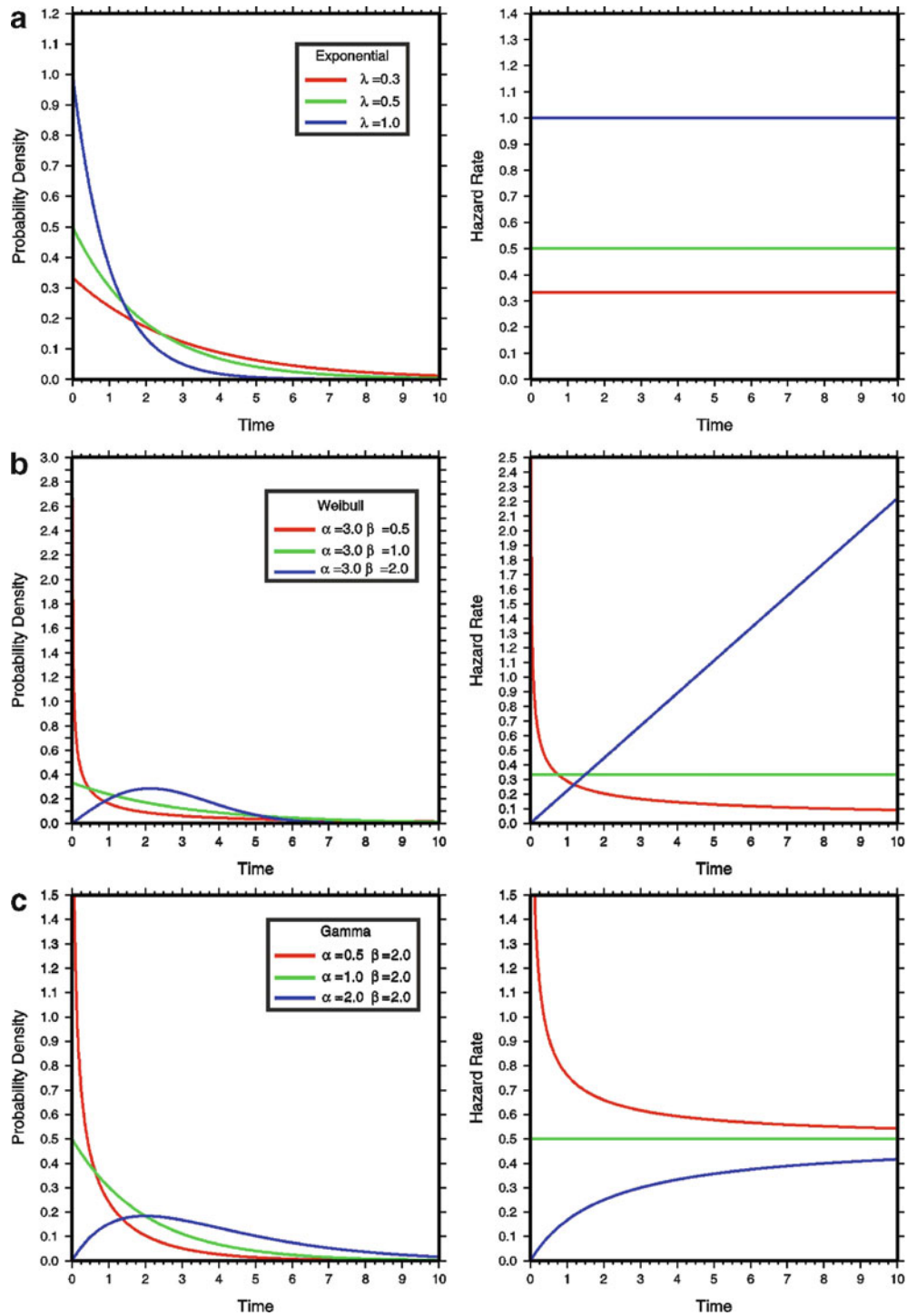
the observed data, the hazard function can be obtained noting that

$$h(\tau) = \frac{f(\tau)}{1 - F(\tau)} = \frac{d}{d\tau} \{-\ln[1 - F(\tau)]\} \quad \tau \geq 0 \quad (14)$$

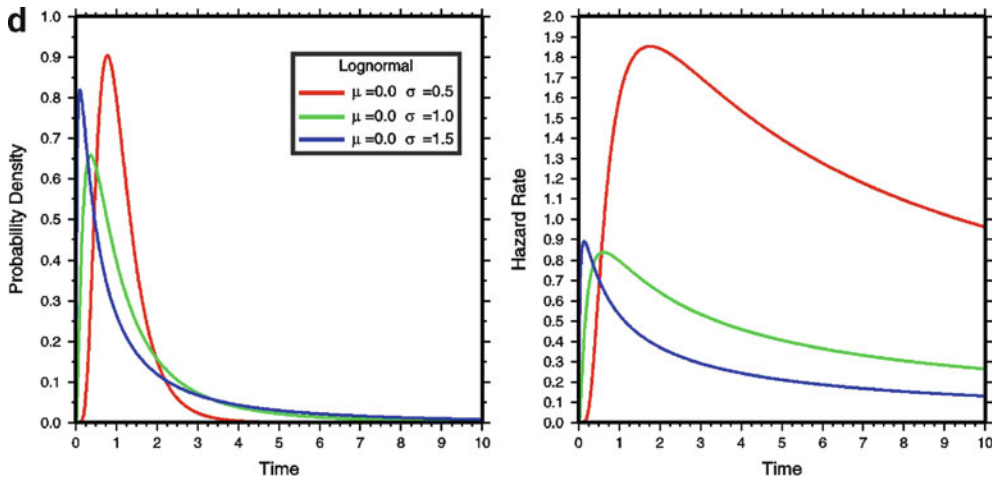
As an example, this model has been applied by Reasenber (1985) to model aftershock sequences and formulate a declustering procedure for seismic catalogs. Concerning the $\lambda(t)$ function, Hong and Guo (1995) by analyzing the seismicity near Taiwan have demonstrated that a U-shaped function is more suited to model the occurrence of clustering earthquakes with respect to a constant rate or a monotonic increasing or decreasing function. The U-shaped function prescribes a constant hazard during the waiting period between cluster of events and higher values at times immediately after the last event or before the next sequence of events.

Time-Dependent Models

Unlike the time-independent models, in time-dependent models, the probability of occurrence of the next event (either in time or space) is modified by the features of the preceding events. In particular, a large class of statistical models assumes that the probability of future events may increase with elapsed time since the last earthquake and is in general applied to define the distribution of earthquakes on single faults. Moreover, for a fault that is in the early stage of the seismic cycle, these models prescribe a lower probability of occurrence compared to the Poisson model, whereas for faults in the late stage of the seismic cycle, they predict a higher probability of next event with respect to the Poisson model. Other models imply that earthquake occurrence probability is higher immediately after a previous event, and it decreases with the time elapsed; this behavior represents the clustered renewal models, and it is more suitable to represent the seismicity of extended areas, which enclose more than a single fault. In both cases, after the occurrence of each event at a



Earthquake Recurrence, Fig. 2 (continued)



Earthquake Recurrence, Fig. 2 Probability density functions and hazard rate functions for the Poisson model (panel a), the Weibull model (panel b), the gamma model (panel c), and the lognormal model

(panel d). In each panel, line colors refer to different combinations of the parameters that characterize the specific distribution

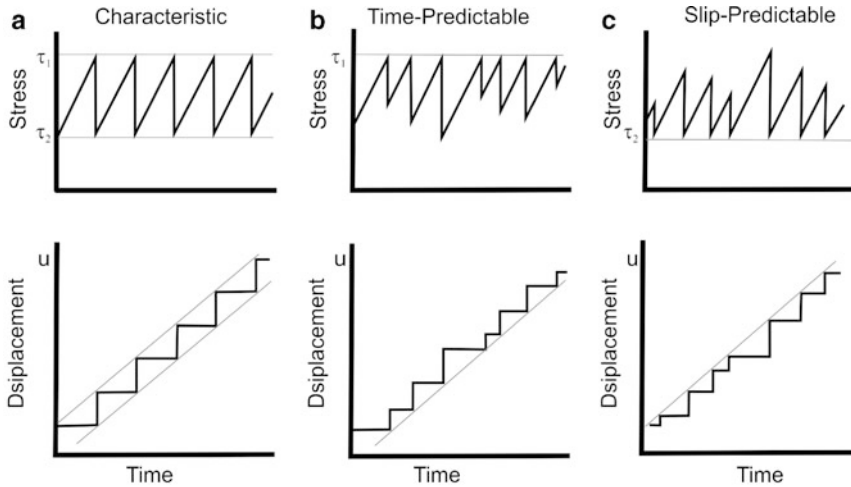
given time t , the system is brought back to the conditions it presented at the time of the last renewal (i.e., the system is perfectly renewed).

Early examples of time-dependent models applied to single fault as an alternative to the standard Poisson model are dating back to the 1970s (Bufe et al. 1977). Analyzing data relative to large thrust-fault earthquakes in Japan, Shimazaki and Nakata (1980) observed that the larger the coseismic slip displacement measured in an earthquake, the longer is the following quiet period. Based on simple assumptions about the initial stress τ_1 and the final stress τ_2 of faulting, the authors suggested two recurrence models known as *time-predictable* model and *slip-predictable* model. In the following, the main features of both models are described. Considering for simplicity that cyclic strain, which is accumulated in time with a constant rate due to tectonic processes is relieved on the same fault, then the stress drop $\Delta\tau$ is assumed to be proportional to the coseismic slip $\overline{\Delta u}$:

$$\Delta\tau \cong \mu \frac{\overline{\Delta u}}{L} \quad (15)$$

where μ is the shear modulus and L is a characteristic fault dimension. Now, if τ_1 and τ_2 are time

independent, then the sequence of earthquakes is a perfect periodic process, and the model is a *characteristic* recurrence model (Fig. 3a). On the other hand, when τ_1 or τ_2 or both are not constant, the process will be no longer periodic. In the assumption that fault strength is constant and that the fault will always rupture when the stress reaches the maximum stress capacity of the fault τ_1 (i.e., τ_1 is constant), if τ_2 is not constant, slip on the fault can vary with each event. Of course, greater slip releases more energy and the recurrence time (time between slip events) is longer because those stresses need to build up once again. This model is said to be *time predictable*, because knowing the amount of slip during the past earthquake, it allows prediction of the time of the next earthquake, but not its magnitude (Fig. 3b). Differently, suppose that τ_1 is not constant, that is, the fault does not rupture at the same stress level each time. If τ_2 is constant, each event reduces the stress on the fault to the same stress level. After an earthquake, stress on the fault will increase at a constant rate from τ_2 . The potential fault slip at any time is proportional to the shear stress on the fault. Thus, if the time of the last rupture is known, the shear stress on the fault and the potential displacement can be determined at any



Earthquake Recurrence, Fig. 3 Earthquake occurrence models. For each model, the upper panels show the stress pattern during different seismic cycles, while the lower panels show the corresponding structure of fault slip. (a) Perfectly periodic model of constant stress increase to a certain level (τ_1) and constant stress drop back to a certain

level during the earthquake (τ_2). (b) Time-predictable model illustrating stress buildup to a certain level and nonuniform stress drops. (c) Slip-predictable model showing nonuniform stress buildups and stress drops to a certain level

particular time of interest. This model is said to be *slip predictable*, in that, it cannot be used to predict when rupture will occur, but it can be used to predict the magnitude of the earthquake that would occur at any given time (Fig. 3c). The application of these models to earthquake forecast has encountered criticism in the seismological community. As an example, Mulargia and Gasperini (1995) have tested the two models on the Italian seismic catalog and found no significant correlation between the coseismic slip of an event and that previous or subsequent inter-event time.

In the following, the most common applied statistical models to model earthquake occurrence are described.

Weibull Model

One of the first applications of the Weibull distribution in geosciences is the one proposed by Hagiwara (1974). Based on the assumption that the Earth crust is strained with a constant speed, he modeled crustal-rupture occurrence time of the South Kanto District in Japan by using the Weibull model. The same distribution was also used by Rikitake (1976) to model the recurrence

times of great earthquakes at the Pacific subduction margins. The formula for the probability density function of the general Weibull distribution is

$$f(t) = \frac{\beta}{\alpha} \left(\frac{t - \mu}{\alpha} \right)^{\beta-1} e^{-\left(\frac{t - \mu}{\alpha} \right)^\beta} \text{ for } t \geq \mu; \beta, \alpha > 0 \quad (16)$$

where β is the shape parameter, μ is the location parameter, and α is the scale parameter. When $\mu = 0$ and $\alpha = 1$, the distribution is called the standard Weibull distribution. The case where $\mu = 0$ is called the 2-parameter Weibull distribution which is generally used for seismological applications. For the 2-parameter distribution, the corresponding cdf and the hazard function are defined as follows, respectively:

$$F(t) = 1 - e^{-\left(\frac{t}{\alpha} \right)^\beta} \quad (17)$$

$$h(t) = \frac{\beta}{\alpha} \left(\frac{t}{\alpha} \right)^{\beta-1} \quad (18)$$

Equations 16, 17, and 18 show that the shape parameter β has a marked effect on the shape of

the functions. However, some interesting insights can be drawn from the analysis of the hazard function (Eq. 18) as shown in Fig. 2b, in which the pdf and the hazard functions are displayed for three different values of β (i.e., 0.5, 1.0, and 2.0) and an α value of 3.0. Indeed, if the inter-event time distribution has a Weibull best-fit model, then the value of β can be used to depict some features of the analyzed seismicity. If, for example, β is equal to 1, the pdf reduces to an exponential function with constant hazard function and rate equal to $1/\alpha$; the process is thus a random process. If $\beta < 1$, then the system exhibits early-type failures, with a decreasing hazard function, with time elapsing since the most recent event, while for $\beta > 1$ the system exhibits wear-out type failures, with an increasing hazard function. The analog for the earthquake generation process for $\beta < 1$ is a clustering property of the earthquake occurrence, while for $\beta > 1$ the pdf can be approximated by a Gaussian or normal distributions and the events, based on the value of α , are quasiperiodic in time.

Gamma Model

Modeling of nonindependent events through a gamma distribution was first presented by Udias and Rice (1975) in their analysis of microearthquake activity near San Andreas Fault, California. Looking at the distribution of the inter-event times, they observed a maximum amount of low values of inter-time intervals and a higher number of events with respect to that predicted by a Poisson model. The deviation from a simple Poisson model was ascribed to a clustering property of the seismicity. The authors thus proposed to fit the observed distribution of the inter-event times by using the gamma distribution. The gamma distribution is a two-parameter function and for $t \geq 0$ the formulation of the pdf is as follows:

$$f(t) = \frac{t^{\alpha-1} e^{-t/\beta}}{\beta^\alpha \Gamma(\alpha)} \quad \text{for } \alpha > 0, \beta > 0 \quad (19)$$

In the above equation, β is a scale parameter reflecting the units of measure, while α controls the shape of the distribution. The gamma

distribution allows to combine a power law, which is usually used to model short inter-event times, that is, clustered events, and long inter-event times through the exponential function which is typical of Poisson seismicity. The corresponding cdf is

$$F(t) = \frac{\gamma(\alpha, t/\beta)}{\Gamma(\alpha)} \quad (20)$$

where

$$\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du \quad \gamma(\alpha, t) = \int_0^t u^{\alpha-1} e^{-u} du \quad (21)$$

while the hazard function is given by

$$h(t) = \frac{t^{\alpha-1} e^{-t/\beta}}{\beta^\alpha (\Gamma(\alpha) - \gamma(\alpha, t))} \quad (22)$$

It can be shown (e.g., Denker and Woyczynski 1998) that the parameters α and β are related to the mean and the variance of the empirical distribution of the inter-event times. In particular, $\alpha \cdot \beta = \mu$ and $\alpha \cdot \beta^2 = \sigma^2$ where μ and σ^2 are the mean and the variance of the observed distribution, respectively. Figure 2c shows the pdf and the hazard function for the gamma distribution for different values of α (i.e., 0.5, 1.0, and 2.0) and a β value of 2.0. From the earthquake occurrence point of view, the α -parameter is the regularity parameter, while beta-parameter is the time scaling parameter. When α is less than 1, the process is clustered; when $\alpha = 1$, the process is Poisson, neither clustered nor regular (the distribution is the exponential distribution, with constant hazard function equals $1/\beta$); and when $\alpha > 1$, the process tends to regular. For large values of α (let us say larger than 100), the inter-event time is almost fixed.

Lognormal Model

Nishenko and Buland (1987) found that normalized earthquake recurrence intervals, for large earthquakes in the circum-Pacific region, along

specific faults or plate boundary segments, can be described by a lognormal distribution. Combining both historic and geologic recurrence data, the authors found that the lognormal distribution fits the observed data significantly better than either the Gaussian or the Weibull distributions.

Since the study by Nishenko and Buland (1987), it has become common practice to use the lognormal distribution in several applications in seismology. One of the most relevant was in computing earthquake probabilities in California (Working Group on California Earthquake Probabilities 2007 and previous works) and for time-dependent seismic hazard analysis (e.g., Cramer et al. 2000).

The pdf of the distribution is given by the equation:

$$f(t) = \frac{1}{\sqrt{2\pi} \sigma_{\ln T} t} e^{-\frac{(\ln t - \mu)^2}{2\sigma_{\ln T}^2}} \quad (23)$$

where $\sigma_{\ln T}$ is the standard error deviation of the logarithm of the recurrence time. The corresponding cdf and the hazard function $h(t)$ have the following expression:

$$F(t) = \frac{1}{2} + \frac{1}{2} \operatorname{erf} \left(\frac{\ln t - \mu}{\sqrt{2\sigma_{\ln T}^2}} \right) \quad (24)$$

$$h(t) = 2 \left[\sqrt{2\pi} \sigma_{\ln T} t \left(1 - \operatorname{erf} \left(\frac{\ln t - \mu}{\sqrt{2\sigma_{\ln T}^2}} \right) \right) \right]^{-1} e^{-\frac{(\ln t - \mu)^2}{2\sigma_{\ln T}^2}} \quad (25)$$

One of the main features of the lognormal distribution is that it is a heavy-tailed distribution, that is, there is a larger probability of getting next earthquakes with respect to the exponential function for the same time window. This feature makes the distribution suitable for fitting inter-arrival times of those events whose return period is larger than the average return period. This is visually explained in Fig. 2d, where the pdfs and the hazard functions for $\mu = 0$ and three different values of $\sigma_{\ln T}$ (e.g., 0.5, 1.0, and 1.5) are shown. The pdfs are peaked, and the

hazard functions have zero value $t = 0$, i.e., the probability of having an earthquake is zero at the beginning of the cycle, then they reach their maximum, and finally decrease asymptotically to zero.

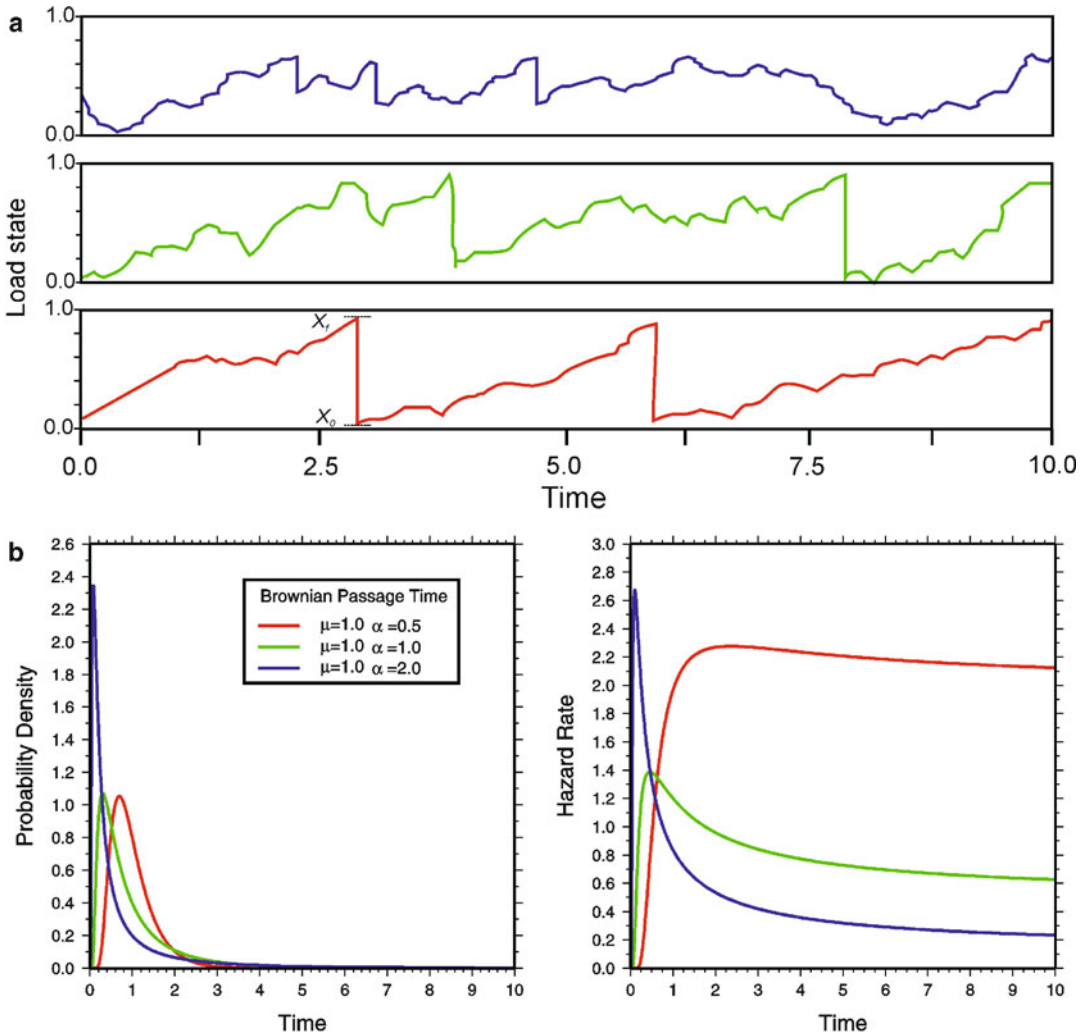
Brownian Passage-Time Model

Brownian passage-time model (BPT model hereinafter) has been developed by Matthews and coworkers in the early 2000s in the framework of earthquake forecasting. The aim was to make a connection between the rebound theory, with the main characteristics of the stress and strain accumulation process, and the observable distribution of the real inter-event data and build up a physically motivated renewal model for earthquake recurrence (Ellsworth et al. 1999; Matthews et al. 2002). In BPT model, the constant tectonic loading is superimposed by a Brownian perturbation, which makes the stress loading a Brownian relaxation oscillator (BRO). The Brownian perturbation could be interpreted as a physical representation of the interaction of external perturbation of the cycling stress load of a fault. In the BRO model, a new earthquake will occur when a state variable (or a set of them) reaches a threshold (X_p); it relaxes the load state to the characteristic ground level (X_0) and begins a new cycle (panel a in Fig. 4).

In the BPT model, the loading of a fault is the composition of two elements: a constant-rate loading component, λt , and a random component, $\epsilon(t) = \sigma W(t)$, as a Brownian motion (where $W(t)$ is a standard Brownian motion and σ is the non-negative scale parameter). Standard Brownian motion is simply the integration of stationary increments where the distribution of the increments is Gaussian (which might be motivated by central-limit arguments if the perturbations are considered as the sum of many small, independent contributions), with zero mean and constant variance:

$$X(t) = \lambda t + \sigma W(t) \quad (26)$$

The BRO model belongs to the family of stochastic renewal processes identified by four



Earthquake Recurrence, Fig. 4 Panel (a) shows three examples of loading state variable trend for the three different aperiodicity values indicated in the inset of panel (b). The levels indicated possible values of the

ground state X_0 and the failure state X_f , respectively. Panel (b) shows the pdf and the hazard rate as a function of three different aperiodicity values

parameters: the mean loading (drift) rate (λ), the perturbation rate (σ^2), the ground state (X_0), and the failure state (X_f). The recurrence properties of the BRO (repose times) are described by a Brownian passage-time distribution which is characterized by two parameters: the mean time or period between events (μ) and the aperiodicity of the mean time (α) which is equivalent to the coefficient of variation ($\alpha = \sigma/\mu$). The probability density function for the BPT model is given by

$$f(t; \mu, \alpha) = \left(\frac{\mu}{2\pi\alpha^2 t^3} \right)^{1/2} e^{-\frac{(t-\mu)^2}{2\mu\alpha^2 t}} \quad (27)$$

The cumulative distribution is

$$F(t) \triangleq P\{T \leq t\} = \Phi[u_1(t)] + e^{2/\alpha^2} \Phi[-u_2(t)] \quad (28)$$

where $\Phi(t)$ is the cumulative Gaussian probability function, u_1 and u_2 are defined as follows:

$$u_1 = \alpha^{-1} [t^{1/2} \mu^{-1/2} - t^{-1/2} \mu^{1/2}] \quad (29)$$

$$u_2 = \alpha^{-1} [t^{1/2} \mu^{-1/2} + t^{-1/2} \mu^{1/2}] \quad (30)$$

and the first moment is $E(T) = \mu$ with $\text{var}(T) = (\mu\alpha)^2$. The mean inter-event gauges the typical frequency at which events occur, and in BPT parameterization it is the scale parameter (i.e., changing μ , the distribution is rescaled in time but its shape is not altered). The aperiodicity is the shape parameter of the BPT family; it is a dimensionless measure of irregularity in the event sequence, and it modifies the shape of the distribution, from a perfectly periodic sequence with $\alpha = 0$ to a random process with an increasing α . As an example panel a in Fig. 4 shows the shape of the pdf and the hazard function for three different values of the aperiodicity parameter α , when μ is set to one.

Compared to the other statistical distributions proposed so far, the $h(t)$ of a BPT distribution is always zero at $t = 0$ (see panel b in Fig. 4), and this implies that after an earthquake the cycle is returned to its starting state and the probability of having a new earthquake is zero; it increases to achieve its unique maximum value at some finite time to the right of the density's mode, and from there it decreases asymptotically at level $h_\infty = (2\mu\alpha^2)^{-1}$ (panel b in Fig. 4). This last property is very important, since it implies that a long time after the previous event, the probability of having a new one does not depend any more from the time since the last event, but it is a constant that depends on the propriety of the fault (i.e., it depends on the mean of the distribution and its aperiodicity $(2\mu\alpha^2)^{-1}$).

It is important to notice that the five models proposed so far imply different mechanisms for earthquake occurrence. All but Poisson model are applied to mark the occurrence of large earthquakes on faults, while the Poisson model can also be used to model the seismicity into homogeneous seismotectonic zonation, where each single zone includes several faults. Trigger models described in the next section are applied on extended areas, where the interaction among faults plays an important role and the clustering of the seismicity is the main feature in

earthquake occurrence, at least in the short time after the most recent event.

To compare the statistical and physical proprieties of the models, their behavior soon after the last event (i.e., at the beginning of the cycle) and after a long time since the last event is analyzed. At $t = 0$, lognormal distribution, BTP, gamma, and Weibull distributions with shape parameters larger than 1 have zero probability of occurrence of next earthquakes. This is conceptually in agreement with the interpretation of the elastic rebound theory and implies that after a large earthquake, the fault is unloaded at its ground level. The lognormal distribution and the BPT remain with lower values of the pdf in the first part of the main recurrence interval, while gamma and Weibull have steep increases, and the probability of new events increases rapidly with time. For the gamma and Weibull distributions with shape parameters less than 1, at $t = 0$ the probability of having a new event is high, indicating a clustering behavior of the seismicity, and for shape parameters equal to 1, they behave as a time-independent Poisson model. All functions are unimodal, with the mode located on the left of the mean of the distributions, implying that most of the events occur at times less than the average time of occurrence.

After long time since the most recent event, the behavior of the distributions can be divided into three classes, except for the case of time-independent distribution. In one case the hazard function reaches a constant value, in the second it goes to zero, and only for the Weibull distribution with shape parameter larger than 1 it grows theoretically to infinity. The important point that should be emphasized in this encyclopedia is the implication that these characteristics have in the definition of the probability of occurrence of earthquakes. The hazard function reaches zero as time goes to infinity for the lognormal distribution and the Weibull distribution with shape parameter less than 1. According to some researches, this characteristic disqualifies these distributions as reliable models, especially if the mechanism of earthquake occurrence is contextualized within the process of steadily increasing load on fault. Some other researchers point out that the stress can likely be dissipated by

alternative mechanisms (seismic or aseismic) and not through the occurrence of the event, and so a decreased failure rate may be appropriate. On the contrary, models with hazard function reaching a constant value as time goes to infinity, as for the BPT and the gamma distribution, imply that after the time corresponding to the mean recurrent interval is passed, the distributions become “time independent,” and the time since the most recent event does not influence any more the occurrence of the next event. The only factors able to control the earthquake occurrence are in the model parameters, as, for example, the mean and the aperiodicity of the BPT distribution.

Trigger Models

One of the most important characteristics of the earthquake time series, which must be considered in the models of recurrence, is the tendency of the events to clustering. A possible definition of cluster is a set of earthquakes originating from a relatively small seismogenic volume and separated in time by very small intervals. Thus, the model should contemplate a finite probability of more than one event to occur in a given time increment. This feature is not taken into account by the Poisson process. A possible solution, which allows for events grouping, is the use of a compound Poisson process. This process permits that groups of independent and identically distributed events occur together at the instants of a simple Poisson process. However, the compound Poisson process may be not satisfactory because groups of events tend to spread or to diffuse over space and time. As an alternative class of models, those named trigger models have been proposed by Vere-Jones and Davies (1966). Basically, the models assume that events initiating groups of shocks – triggering events – occur by following a Poisson process with a constant rate, while the number of events in each group decays by following a prescribed function.

Generalized Poisson Model

In the framework of the trigger models, Shlien and Toksöz (1970) proposed to model the

seismicity contained in the United States Coast and Geodetic Survey (USCGS) catalog from 1961 to 1968, by using the generalized Poisson model. Assuming that cluster occurrence is governed by a Poisson process and that in each cluster the number of subsequent earthquakes follows a negative power law distribution, Shlien and Toksöz (1970) provided a recursive relation to compute the distribution of the number of days with n earthquakes based on two main parameters named k' and E . The probability distribution has the following formulation:

$$p_q(n+1, t) = \frac{kt}{n+1} \sum_{j=0}^n (n-j+1) q(n-j+1) p_q(j, t) \quad (31)$$

where $q(n)$ is the selected negative power law distribution also known as the Pareto distribution, given by

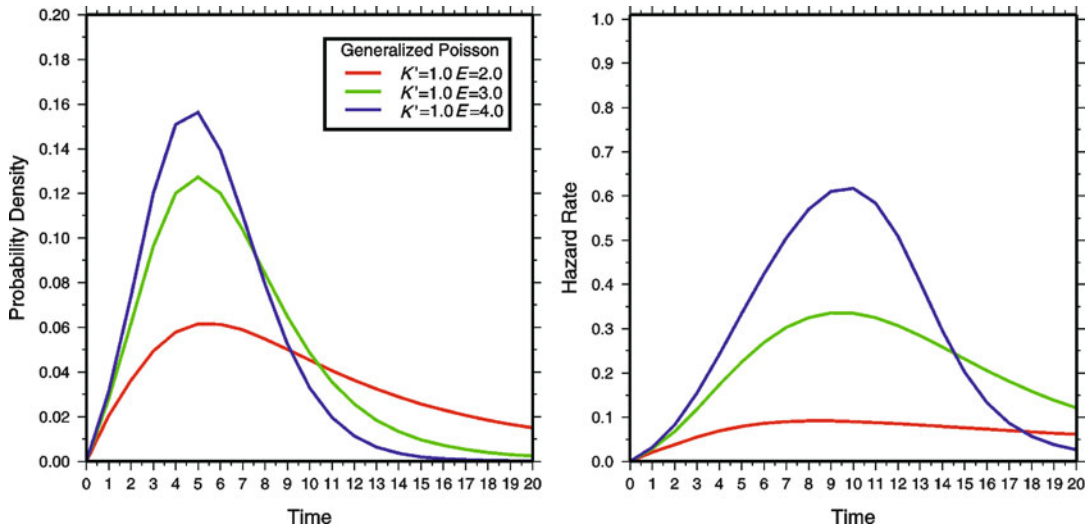
$$q(n) = \frac{1}{\zeta(E)} n^{-E} \quad (32)$$

where the normalization factor $\zeta(E)$ is the Riemann Z function.

The k' parameter controls the mean rate of cluster occurrence. The parameter E is a measure of the departure from the simple Poisson process. For large value of E (let us say larger than 5), the generalized Poisson model tends to a simple Poisson distribution. Figure 5 shows the shape of the pdf for $k' = 1$ and three different values of E (i.e., 2.0, 3.0, and 4.0). In the same figure, the right panel shows the hazard function obtained by numerical integration of the pdf for the same values of the two parameters. Among the several applications, the generalized Poisson model has been also used to model earthquake sequences and compute seismic hazard in volcanic contexts (De Natale and Zollo 1986; Convertito and Zollo 2011).

The ETAS Model

The spatiotemporal clustering of small-to-moderate size seismicity is a concept worldwide accepted in seismology. Among the few



Earthquake Recurrence, Fig. 5 Probability density function and hazard rate function for the generalized Poisson model. The selected values of the parameters k' and E are reported in the inset of the right panel

spatiotemporal models proposed in literature, the epidemic type aftershock sequence (ETAS) model is by far the most studied and applied, and, in some cases, it has been defined as the best way to describe short-term seismicity (Console et al. 2007; Zhuang et al. 2011, 2011) since its formulation in 1988 by Ogata (1988). ETAS model is grounded on empirical statistics and represents a stochastic implementation of the physically based model of Dieterich (1994), based on the interaction among faults. It assumes that there is no difference in triggering seismicity among foreshocks, mainshocks, and aftershocks, and event triggers its own offspring independently according to some probability rules. The hazard rate of the ETAS model is defined as

$$\lambda(t) = \mu + K_0 \sum_{t_i < t} \frac{e^{\alpha(m-m_0)}}{(t-t_i+c)^p} \quad (33)$$

where μ is the background rate (unit of time) of the seismicity and m_0 is the magnitude of completeness of the seismic catalog. The parameter α is a measure of the efficiency of a shock in generating an aftershock activity and it is relative to its magnitude; K_0 represents the productivity of an event of threshold magnitude m_0 , whereas c and p are parameters

of the Omori-Utsu law, which describes the temporal behavior the aftershock sequence. The 5 parameters of the ETAS model are generally estimated by using a maximum likelihood procedure. In literature, different parameterizations of the ETAS model are available, especially when also spatial distribution of the events is taken into account. For the purpose of this encyclopedia, the main features common to the ETAS models developed by different researchers are summarized in five points (Zhuang et al. 2011):

- The rate of background events is assumed to be a function of space and magnitude; the background events are generated following a specific pdf function (e.g., Poisson, BPT, etc.), and so they are considered as “immigrants” in the branching process.
- Each event produces direct offspring events independently of the others, and the productivity depends on its magnitude.
- The pdf of the time until the appearance of a child is a function only of the elapsed time since its parents.
- The pdf of the location and magnitude of children are a function of the location and magnitudes of the parents.

- Magnitude, of both offspring and background events, is a random variable, with a certain distribution, the most common of which is the Gutenberg-Richter relation.

ETAS model is widely adopted in short-term hazard studies, for its ability in the forecasting of aftershock sequence; also, it is adopted as declustering method for separating the background seismicity to the triggered one (Zhuang et al. 2002, 2004). Moreover, it has been found that the inclusion of a more realistic behavior of the seismicity in the temporal parameterization of earthquake occurrence in PSHA, with the well-known structure of the sequence, provides an increment of almost 10 % in the hazard, in low seismicity area, where the identification of active faults or background seismicity is trivial (Faenza et al. 2007).

Estimation of Parameters

It is worth noting that the models described in this entry may fit the available data set collected in the region under study equally well. This is especially true in case of large earthquakes, where the amount of data available may be not sufficient to test any hypotheses due to the limited time extent of the available seismic catalog. Before performing the parameter estimation, the completeness of the catalog has to be determined. This could be trivial in case of historical earthquakes; an erroneous evaluation of the completeness can affect the value of the model parameters and their uncertainties. It must be reminded that large events are rare, especially in area of moderate seismicity, and therefore a larger part of the seismic catalog are derived from pre-instrumental time.

In practical applications, the maximum likelihood estimate (MLE) is by far the most used technique to evaluate the parameter's models. The MLE is based on the use of the likelihood function L for the estimation of parameters. The L function represents the joint probability density of the analyzed variable, for example, the waiting times or the inter-arrival times. Given N data points and a model whose parameters are

identified by a vector θ , the best model parameters $\hat{\theta}$ are those maximizing the L function:

$$\hat{\theta} = \max_{\theta} L(N; \theta) \quad (34)$$

Generally speaking, the MLE method maximizes the agreement of the selected model with the observed data.

When several competing models are used to fit the same data set, the best model can be discriminated by using the Akaike information criterion (AIC), which uses the following statistic to evaluate the fitting performance of each model:

$$AIC = -2 \max_{\theta} \log L(\theta) + 2N_p \quad (35)$$

where N_p is the number of parameters involved in each investigated model. The model with the lowest AIC value is identified as the optimal model. AIC balances between the goodness of fit of the model and its complexity, and it provides penalties to models with a lot of numbers of parameters, to discourage overfitting. AIC does not provide a test of the model in absolute sense, but provides only a tool for the comparison of different models; if all models have poor fitting, AIC will not give any warning of it.

The residual analysis allows the evaluation of the degree at which the adopted model fits the data. The analysis uses a transformation to map general event times, in times having the same properties as homogeneous Poisson time with unit rate parameter (Zhuang et al. 2012). Thus, comparing the observed seismic rate with that obtained in the transformed time domain allows to evaluate the quality of the fit. The comparison allows also to recognize periods of decreased or increased seismic activity (Ogata 1988; Zhuang et al. 2012).

Summary

Seismic hazard in earthquake-prone regions can be expressed in several forms. The exceedance of a given ground-motion threshold value or the probability of having a given number of future

earthquakes that can affect a specific structure is of primary interest in seismic engineering. As a consequence, being able to model present and past seismicity that occurred in the area under study may be essential to make reliable forecast.

In this framework, the present work discusses the basic assumptions and the origin of the main statistical models used in the literature to fit the observed seismicity. Among the five dimensions which concur to basically characterize the earthquakes (i.e., time, magnitude, and location in 3D space), the physical observable of interest for statistical modeling is the earthquake recurrence, which is the time between subsequent events. In this respect, earthquakes are considered as specific stochastic processes named *renewal process* (Cornell and Winterstain 1988) in which after the occurrence of an event, the system process restarts from the beginning (i.e., is completely renewed).

The models used to describe the recurrence of the earthquakes can be divided in two main classes: the “memoryless” model, i.e., Poisson model, and the time-dependent models. While Poisson model is extensively used in PSHAs, its adequacy is limited since the mechanism of earthquake interactions at regional scale has highlighted the tendency to a spatiotemporal clustering of earthquake occurrence, even for moderate earthquakes. Moreover, due to the “memoryless” property, it is not adequate to model the occurrence of earthquakes on single faults in which case non-Poisson models perform better.

As a consequence, since the work of Bufe et al. (1977), a suite of models able to account for earthquake interaction and clustering has been introduced. We have presented those most used in practical applications such as Weibull, gamma, and lognormal and described their main features and differences in terms of their probability density functions and hazard functions. In addition we described a more physical model, that is, the Brownian passage time, which is currently used to perform time-dependent seismic hazard analysis in several areas, tectonic and volcanic areas such as California and Italy.

Finally we have discussed the generalized Poisson model and the ETAS model which, in addition to model single large-to-moderate events, allow to account for sequences of events clustering in time and space.

As previously mentioned, the selection of “the best model” among several competitors is not straightforward. In general, dealing with large earthquakes implies working with small numbers of data available for the fitting; this is especially true in case of recurrence on single fault or in area with moderate seismicity. In this perspective, in recent years, the seismological community has developed rigorous and truly perspective forecast experiments, as, for example, the CSEP experiment – Collaboratory for the Study of Earthquake Predictability, for different regions worldwide.

Cross-References

- [Earthquake Mechanisms and Tectonics](#)
- [Probabilistic Seismic Hazard Models](#)
- [Site Response for Seismic Hazard Assessment](#)

References

- Bakun WH, King GCP, Cockerham RS (1986) Seismic slip, aseismic slip, and the mechanics of repeating earthquakes on the Calaveras fault, California. In: Das S, Boatwright J, Scholz CH (eds) Earthquake source mechanics, vol 37, Geophysical monograph. American Geophysical Union, Washington, DC
- Bufe CG, Harsh PW, Burford RO (1977) Steady-state seismic slip-A precise recurrence model. *Geophys Res Lett* 4:91–94. doi:10.1029/GL004i002p00091
- Console R, Murru M, Catalli F, Falcone G (2007) Real time forecasts through an earthquake clustering model constrained by the rate-and-state constitutive law: comparison with a purely stochastic ETAS model. *Seismol Res Lett* 78:49–56. 3, 15
- Convertito V, Zollo A (2011) Assessment of pre-crisis and syn-crisis seismic hazard at Campi Flegrei and Mt. Vesuvius volcanoes, Campania, southern Italy. *Bull Volcanol* 73:767–783. doi:10.1007/s00445-011-045-2
- Cornell CA, Winterstain SR (1988) Temporal and magnitude dependence in earthquake recurrence models. *Bull Seismol Soc Am* 78:1522–1537

- Cramer CH, Petersen MD, Cao T, Toppozada TR, Reichle MS (2000) A time-dependent probabilistic seismic-hazard model for California. *Bull Seismol Soc Am* 90:1–21
- De Natale G, Zollo A (1986) Statistical analysis and clustering features of the Phlegraean fields earthquake sequence (May 1983–May 1984). *Bull Seismol Soc Am* 76(3):801–814
- Dieterich JH (1994) A constitutive law for rate of earthquake production and its application to earthquake clustering. *J Geophys Res* 99:2601–2618
- Denker M, Woyczynski WA (1998) *Introductory statistics and random phenomena*. Birkhäuser, Boston/Basel/Berlin
- Ellsworth WL, Matthews MV, Nadeau RM, Nishenko SP, Reasenberg PA, Simpson RW (1999) A physically based earthquake recurrence model for estimation of long-term earthquake probabilities. U.S. Geological Survey open-file rept. 99–522
- Faenza L, Hainzl S, Scherbaum F, Beauval C (2007) Statistical analysis of time-dependent earthquake occurrence and its impact on hazard in the low seismicity region Lower Rhine Embayment. *Geophys J Int* 171(2):797–806
- Gutenberg B, Richter CF (1944) Frequency of earthquakes in California. *Bull Seismol Soc Am* 34:185–188
- Hagiwara Y (1974) Probability of earthquake occurrence as obtained from a Weibull distribution analysis of crustal strain. *Tectonophysics* 23:313–318
- Hong L-L, Guo S-W (1995) Nonstationary Poisson model for earthquake occurrences. *Bull Seismol Soc Am* 85:814–824
- Matthews MV, Ellsworth WL, Reasenberg PA (2002) A Brownian model for recurrent earthquakes. *Bull Seismol Soc Am* 92:2233–2250
- Mulargia F, Gasperini P (1995) Evaluation of the applicability of the time- and slip-predictable earthquake recurrence models to Italian seismicity. *Geophys J Int* 120(2):453–573
- Nishenko SP, Buland R (1987) A generic recurrence interval distribution for earthquake forecasting. *Bull Seismol Soc Am* 77:1382–1399
- Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *J Am Stat Assoc* 83:9–27
- Reasenberg P (1985) Second-order moment of central California seismicity, 1969–82. *J Geophys Res* 90:5479–5495
- Reid HF (1910) On mass-movements in tectonic earthquakes. In: *The California earthquake of April 18, 1906: report of the state earthquake investigation commission*. Carnegie Institution of Washington, Washington, DC
- Rikitake T (1976) Recurrence of great earthquakes at subduction zones. *Tectonophysics* 35:335–362
- Schwartz DP, Coppersmith J (1984) Fault behaviour and characteristic earthquakes: examples from Wasatch and San Andreas faults. *J Geophys Res* 89:5681–5698
- Shimazaki K, Nakata T (1980) Time-predictable recurrence model for large earthquakes. *Geophys Res Lett* 7:279–282
- Shlien S, Toksöz MN (1970) A clustering model for earthquake occurrences. *Bull Seism Soc Am* 60:1765–1787
- Udias A, Rice J (1975) Statistical analysis of microearthquake activity near San Andreas Geophysical Observatory, Hollister. *Calif Bull Seism Soc Am* 65:809–827
- Vere-Jones D, Davies RB (1966) A statistical survey of earthquakes in the main seismic region of New Zealand, 2, time series analysis. *N Z J Geol Geophys* 9:251–284
- Working Group on California Earthquake Probabilities, 2007, Uniform California earthquake rupture forecast, USGS open-file report 2007-1437
- Youngs RR, Coppersmith K (1985) Implications of fault slip rates and earthquake recurrence models to probabilistic seismic hazard estimates. *Bull Seism Soc Am* 58:939–964
- Zhuang J, Ogata Y, Vere-Jones D (2002) Stochastic declustering of space-time earthquake occurrences. *J Am Stat Assoc* 97:369–380
- Zhuang J, Ogata Y, Vere-Jones D (2004) Analyzing earthquake clustering features by using stochastic reconstruction. *J Geophys Res* 109(B5):B05301. doi:10.1029/2003JB002879
- Zhuang J, Werner MJ, Hainzl S, Harte D, Zhou S (2011) Basic models of seismicity: spatiotemporal models, community online resource for statistical seismicity analysis, doi:10.5078/corssa-07487583. Available at <http://www.corssa.org>
- Zhuang J, Harte D, Werner MJ, Hainzl S, Zhou S (2012) Basic models of seismicity: temporal models community online resource for statistical seismicity analysis. doi:10.5078/corssa-79905851. Available at <http://www.corssa.org>

Earthquake Recurrence Law and the Weibull Distribution

Jui-Pin Wang

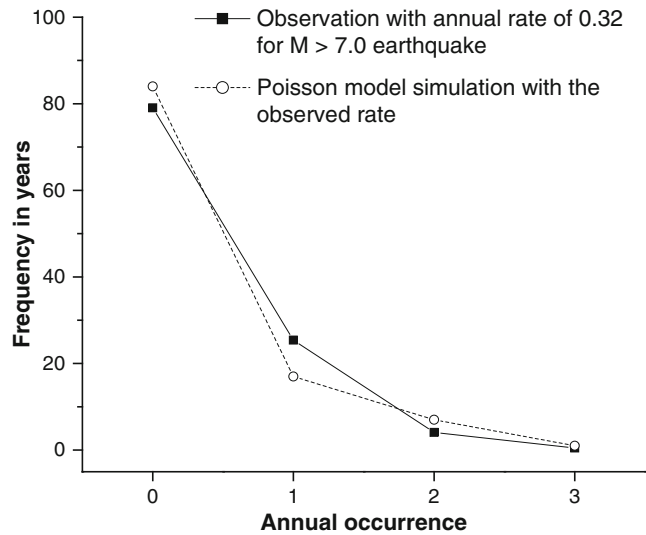
Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Kowloon, Hong Kong

Synonyms

Earthquake; Weibull distribution

Earthquake Recurrence Law and the Weibull Distribution, Fig. 1

The observed frequency and the Poisson frequency of $M > 7.0$ earthquakes around Taiwan since 1978; this statistical study offered a tangible support to the use of the Poisson distribution in modeling the earthquake's randomness in time (Wang et al. 2014)



Introduction

Earthquakes are unpredictable given the difficulty of measuring the stresses and strains on a fault plane (Geller et al. 1997). The situation reflects the reality that recent catastrophic events went unpredicted, such as the 2008 Wenchuan earthquake in China and the 2011 Japan earthquake. Under the circumstances, a few alternatives for earthquake hazard mitigation are developed and practiced, including earthquake early warning, seismic hazard analysis, and earthquake statistics analysis (e.g., Hsiao et al. 2011; Chen et al. 2013; Wang et al. 2011, 2012a, b; Wu and Kanamori 2008).

With statistical models, earthquake probability can be best estimated accounting for the inevitable earthquake uncertainty in nature. One of the famous examples is the use of the Poisson model to estimate the earthquake probability with time. However, besides the support from mainly engineering judgments (earthquakes are rare), more tangible support was not available until a recent statistical study analyzed the seismicity around Taiwan since 1900 (Wang et al. 2014). That study carried out a series of statistical goodness-of-fit tests and indicated that the observed earthquake distribution showed a good agreement with the Poisson distribution. Figure 1 shows an instance from that study providing the

comparison between observation and simulation based on the earthquake data around Taiwan since 1900.

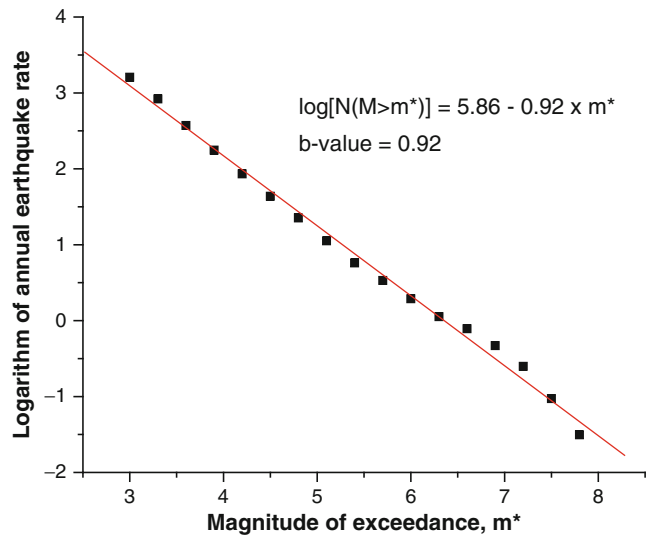
In addition to the modeling of the earthquake temporal uncertainty, some suggestions were proposed to model the probabilistic distribution of earthquake size, including the use of a specific “M-A” algorithm proposed by McGuire and Arabasz (1990), governed by the b -value of the Gutenberg-Richter (G-R) recurrence law:

$$\Pr(m_1 \leq M < m_2 | m_0, m_{\max}) = \frac{10^{-bm_1} - 10^{-bm_2}}{10^{-bm_0} - 10^{-bm_{\max}}} \quad (1)$$

where m_0 and $m_{\max}$ are the magnitude threshold and the maximum magnitude, respectively. With the algorithm, for instance, Fig. 2 shows the b -value of 0.92 from the G-R relationship given the seismicity in Taiwan since 1900. Based on the b -value, Fig. 3 shows the modeling of the earthquake-size probability distribution around Taiwan with the M-A algorithm expressed in Eq. 1. It shows that this approach overall offers a satisfactory fit to the earthquake observations, although the model could not capture the probability at lower magnitudes. One of the reasons causing this problem could be the “de-log” process involved in the calculation. That is, as shown in Fig. 2, the b -value calibrated is on the basis of

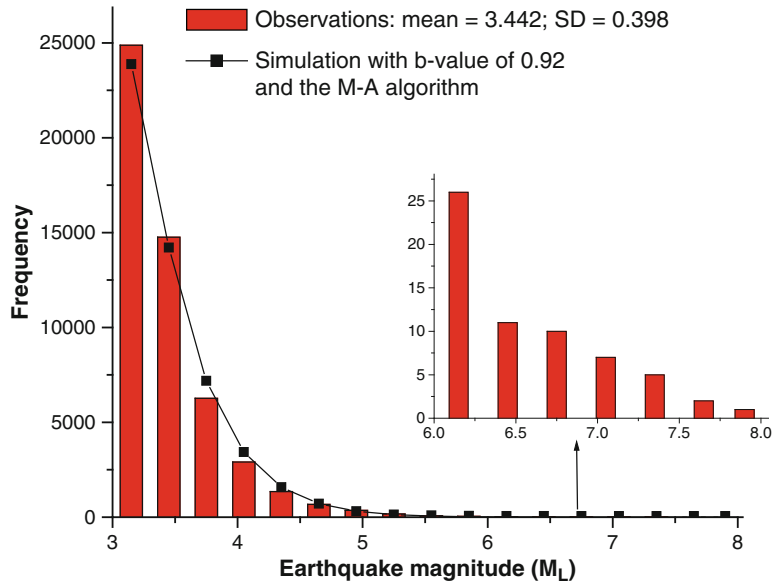
Earthquake Recurrence Law and the Weibull Distribution, Fig. 2

The G-R recurrence law for the seismicity around Taiwan since 1900; the b -value, the slope $\times -1$ in such a dimension, was found around 0.92



Earthquake Recurrence Law and the Weibull Distribution, Fig. 3

Observed earthquake-size distribution given $M_L > 3.0$ events in Taiwan since 1900; given the b -value of 0.92, the plot also shows the simulated earthquake-size distribution with a specific M-A algorithm expressed with Eq. 1



a logarithmic scale, rather than in the arithmetic scale used to present the earthquake-size probability or frequency.

Therefore, in this technical note another option for the earthquake analysis is introduced. Different from the existing approach previously mentioned, the alternative is an application of the Weibull distribution to earthquake statistical studies. The details of the methodology and a case are given in the following.

The Probability Density Function of Earthquake Size Around Taiwan

As shown in Fig. 3, the probability density function for earthquake size was extracted from a catalog listing the earthquakes around Taiwan since 1900 (Wang et al. 2011, 2012a). For example, above all M_L 3.0 (local magnitude) earthquakes, the probability density for earthquakes larger than 7.0 is around 0.01 % among the total

of 55,000 events. In addition, the data also show that the earthquake-size probability distribution is nonlinear and asymmetrical, with the frequency or density decreasing rapidly as magnitudes increase. Therefore, an asymmetrical probability model such as the Weibull distribution should have more potential to model such an earthquake statistic, than symmetrical distributions like the normal distribution.

The Probability Function of the Weibull Distribution

The probability density function (PDF) of the Weibull distribution is expressed as follows (Devore 2008):

$$f(x; \alpha, \beta) = \begin{cases} \frac{\alpha}{\beta} x^{\alpha-1} e - (x/\beta)^\alpha & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (2)$$

where α and β are the model parameters, which can be calibrated with the mean (i.e., μ_X) and SD (i.e., σ_X) of X . The relations between μ_X , σ_X , α , and β are as follows:

$$\mu_X = \beta \Gamma\left(1 + \frac{1}{\alpha}\right) \quad (3)$$

$$\sigma_X^2 = \beta^2 \left(\Gamma\left(1 + \frac{2}{\alpha}\right) - \left[\Gamma\left(1 + \frac{1}{\alpha}\right) \right]^2 \right) \quad (4)$$

where $\Gamma()$ denotes the gamma function:

$$\Gamma(a) = \int_0^{\infty} x^{a-1} e^{-x} dx \quad (5)$$

Therefore, given the mean and SD from observations or samples, α and β of the Weibull distribution can be calibrated with Eqs. 3 and 4. Afterward, the corresponding Weibull distribution can be developed for modeling earthquake-size distributions. However, although this is a two-equation-two-unknown problem, the analytical solution of α and β could be fairly challenging to be developed for such a mathematical

problem with the involvement of the gamma function. As a result, an in-house optimization procedure was developed for the task, and it was introduced in the following section for the reader's information.

Model Parameter Calibrations

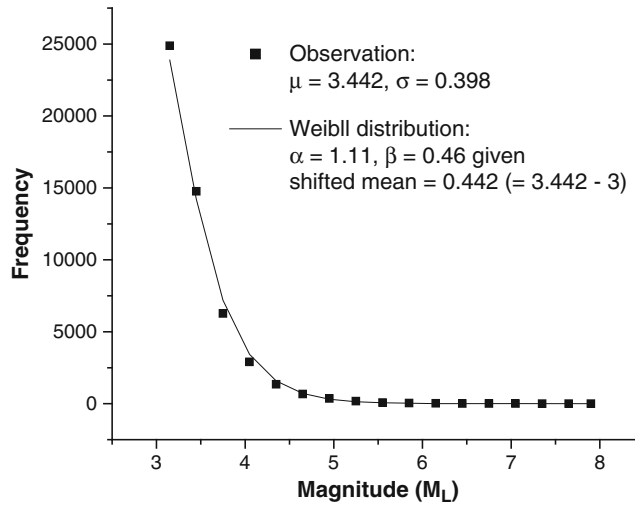
Since the solution could be difficult to be developed analytically, a numerical procedure to calculate α and β was developed and the details were given in the following. Using the first given information (i.e., μ_X), from Eq. 3 the relationship between α and β can be rewritten as follows:

$$\beta = \frac{\Gamma\left(1 + \frac{1}{\alpha}\right)}{\mu_X} \quad (6)$$

With this relationship, the calibration can be followed by a trial-and-error process by varying the value of α (so β changes as well) and substituting both into Eq. 4 to calculate the variance of X . As a result, the absolute difference, $D = f(\alpha)$, between the calculated variance and given variance (i.e., the second given information) in each trial can be calculated.

With the preliminary analysis showing the change in D given different α values, it was found that this variable would converge with the increase of α from a nominal value near zero (α cannot be zero; see Eq. 4) to the solution of α . Therefore, the searching process will be stopped when $D_{i+1} = f(\alpha_i + \Delta\alpha)$ is greater than $D_i = f(\alpha_i)$. For a better resolution in results, a smaller increment of α can be used, say, 0.005.

Apparently, this numerical solution is a computer-based procedure, and the tedious iteration should be complete with computers in a fraction of time. Therefore, this algorithm proposed was programmed into two user-defined functions for Excel. Note that the two functions are nearly identical except for returning α or β at the end. For example, the syntax of the function to calibrate α was created as follows: WEIBULL _ AFA(mn , sd), and obviously



Earthquake Recurrence Law and the Weibull Distribution, Fig. 4 The comparison between the observed earthquake-size distribution and the Weibull distribution with $\alpha = 1.11$ and $\beta = 0.46$ calibrated with the observed mean and standard deviation in the randomness of

earthquake size; this statistical analysis indicated that the Weibull distribution should be suitable to the modeling of earthquake-size distributions based on the empirical finding with earthquake data from Taiwan

“*mn*” and “*sd*” are the function input, standing for the mean and standard deviation of the random variable, respectively. The complete VBA (Visual Basic for Application) scripts of the function are given in the Appendix for the reader’s information.

The in-house function was subject to verification before it was used in the following statistical analysis. The first attempt is to calibrate α and β given $\mu = 10$ and $\sigma = 1$. With the in-house tool, α and β were computed at 12.16 and 10.43, respectively. Substituting them back into Eqs. 3 and 4, μ and σ were found equal to 10 and 0.999, which matched the given values almost perfectly. The second attempt is to test the function’s reliability under a large-variability condition with the standard deviation increased from 1 to 50. Given $\mu = 10$ and $\sigma = 50$, the corresponding α and β are equal to 0.31 and 0.25, respectively. Again substituting both into Eqs. 3 and 4, μ and σ were computed at 10 and 50.45, also close enough to the given conditions. As a result, after two verifications the in-house algorithm and user-defined function were proved reliable and functional.

Case Study with Earthquake Data Around Taiwan

As shown in Fig. 3, given magnitude threshold = 3.0, the mean earthquake size and standard deviation from about 55,000 earthquakes in Taiwan are equal to 3.442 and 0.398, respectively. With Eqs. 3 and 4, α and β of the Weibull distribution are 1.11 and 0.46, respectively. It must be noted that such a calculation was performed on the basis of shifted data, because the Weibull distribution is about the modeling of a random variable ranging from zero to infinity. As a result, the α and β values calibrated are corresponding to the shifted data with mean = 0.442 and SD = 0.398. (The shifted data are the original data subtracting the magnitude threshold of 3.0.) Note that because the variability of data is not affected by data shifting, the SD used for calibrating α and β values is the same as the original SD.

With the model parameters calibrated from observations, Fig. 4 shows both the observed earthquake-size probability density and the

modeling with the Weibull distribution. This analysis indicated that the earthquake-size distribution in Taiwan shows a good agreement with the Weibull distribution.

Discussions

Applications and Verifications with Seismicity in Different Regions

The statistical study of this note offered tangible evidence about the earthquake size's randomness potentially following the Weibull distribution, with the earthquake data around Taiwan since 1978. Although it is believed that this statistical finding should be generic and could be applied to seismicity of different region, more earthquake statistics studies are necessary to provide more support or disagreement on this empirical finding: the earthquake-size distribution follows the Weibull distribution.

Curve-Fitting Parameters

Like the G-R relationship, the new earthquake model on the basis of the Weibull distribution is an empirical relation, without physical indications to why the earthquake's size should be a random variable following such a mathematical function. In other words, the model parameters (i.e., α and β) hold no physical explanations or meanings to earthquake mechanisms, although this statistical model provide a very satisfactory fit to the earthquake data around Taiwan. After all, it was still a statistical study or curve fitting, which was not an analytical study derived with basic laws of physics.

Summary

This entry presents a statistical study regarding earthquake-size distributions in Taiwan since 1978. From earthquake observations, the probability density of earthquake size was found highly asymmetrical, with the majority of seismicity being small earthquakes. This asymmetrical observation motivated the use of the Weibull distribution in an attempt to model the data and

to see if it could provide a reasonable fit to the randomness of earthquake size observed in Taiwan.

With such a statistical study, it was found that the Weibull distribution could fit the earthquake data in Taiwan reasonably well. For instance, given a magnitude threshold of 3.0, mean magnitude of 3.442, and standard deviation of 0.398, the corresponding parameters α and β of the Weibull distribution are 1.11 and 0.46, respectively. With the parameters calibrated with the seismicity, the observation and the model are in a satisfactory agreement in earthquake-size distribution.

Although this statistical study suggests a possible relationship between earthquake size and the Weibull distribution, this empirical finding offers no explanation to earthquake mechanisms and to why the earthquake's size distribution should follow the specific mathematical function. In other words, the Weibull distribution could be useful and reliable for modeling the earthquake-size probability distribution from a statistical point of view, but the model itself offers no explanation to the underlying earthquake mechanics.

Appendix: VBA Scripts of the In-House Weibull Functions Calibrating Model Parameters

As mentioned, the two in-house functions are basically the same except for returning α or β . Therefore, the script for calculating α is only given in the following.

```
Public Function weibull_afa(mn, sd)
' *****
'This function to calculate afa of Weibull
'mn = mean value of X
'sd = standard deviation of X
'inc = increment of afa
'k0 = initial value of afa
'k and c = interim variables
' *****
inc = 0.005
k0 = 0.01
```

```

k = 0
Do
k = k + 1
afa1 = inc * k + k0
beta1 = mn / gamma_function(1 + 1 / afa1)
c1 = beta1 ^ 2 * (gamma_function(1 + 2 /
afa1) - (gamma_function(1 + 1 / afa1)) ^ 2)
c1 = Abs(c1 - sd ^ 2)
afa2 = inc * (k + 1) + k0
beta2 = mn / gamma_function(1 + 1 / afa2)
c2 = beta2 ^ 2 * (gamma_function(1 + 2 /
afa2) - (gamma_function(1 + 1 / afa2)) ^ 2)
c2 = Abs(c2 - sd ^ 2)
If c2 > c1 Then
Exit Do
End If
Loop
weibull_afa = afa1
End Function

```

References

- Chen CH, Wang JP, Wu YM, Chan CH, Chang CH (2013) A study of earthquake inter-occurrence times distribution models in Taiwan. *Nat Hazards*. doi:10.1007/s11069-012-0496-7
- Devore JL (2008) Probability and statistics for engineering and the sciences, 8th ed. Brooks/Cole CENGAGE Learning, Independence, KY
- Geller RJ, Jackson DD, Kagan YY, Mulargia F (1997) Earthquakes cannot be predicted. *Science* 275:1616
- Hsiao NC, Wu YM, Zhao L, Chen DY, Huang WT, Kuo KH, Shin TC, Leu PL (2011) A new prototype system for earthquake early warning in Taiwan. *Soil Dyn Earthq Eng* 31:201–208
- McGuire RK, Arabasz WJ (1990) An introduction to probabilistic seismic hazard analysis. In: Ward SH (ed) *Geotechnical and environmental geophysics*, vol 1. Society of Exploration Geophysics, Tulsa, pp 333–353
- Wang JP, Chan CH, Wu YM (2011) The distribution of annual maximum earthquake magnitude around Taiwan and its application in the estimation of catastrophic earthquake recurrence probability. *Nat Hazards* 59:553–570
- Wang JP, Chang SC, Wu YM, Xu Y (2012a) PGA distributions and seismic hazard evaluations in three cities in Taiwan. *Nat Hazards* 64:1373–1390
- Wang JP, Huang D, Yang Z (2012b) The deterministic seismic hazard map for Taiwan developed using an in-house Excel-based program. *Comput Geosci* 48:111–116
- Wang JP, Huang D, Chang SC, Wu YM (2014) New evidence and perspective to the Poisson process and earthquake temporal distribution from 55,000 events around Taiwan since 1900. *Nat Hazards Rev ASCE* 15:37–48
- Wu YM, Kanamori H (2008) Exploring the feasibility of on-site earthquake early warning using close-in records of the 2007 Noto Hanto earthquake. *Earth Planets Space* 60:155–160

Earthquake Response Spectra and Design Spectra

Dimitris L. Karabalis

Department of Civil Engineering, University of Patras, Patras, Greece

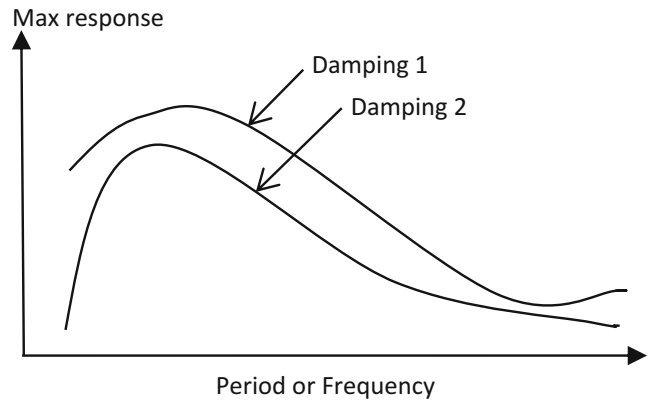
Synonyms

Characterization of ground motions; Earthquake ground motions; Structural response

Introduction

The design of structures capable to perform satisfactorily when subjected to earthquake ground motions requires a characterization of such motions far more detailed than the information provided by popular seismological data, e.g., the Modified Mercalli Intensity. Key role to the seismic response of a structural system plays, among others, the frequency content of the induced ground motion. All ground motions contain a broad range of frequencies and each structure is more or less sensitive to some of them. Of all the rigorous mathematical or engineering tools available to describe the frequency dependence of a seismic signal, none is simpler or more valuable, for design purposes, than the response spectrum which is a plot of the maximum response of all possible single-degree-of-freedom (SDOF) systems to a given dynamic excitation. The term “maximum response” may imply any quantity of interest, e.g., displacement, acceleration, base shear force or moment, etc., while “dynamic

Earthquake Response Spectra and Design Spectra, Fig. 1 A typical response spectrum plot



excitation” usually suggests some earthquake ground motion. Since an SDOF system is fully specified by its period (or frequency) and damping, the response spectrum is usually plotted as a function of period T (or frequency ω) for a set of damping values, as is depicted in Fig. 1.

As it is well known, the maximum response of a structure is usually dominated by its fundamental mode of vibration of, say, period T . Entering this value as the coordinate of the spectrum plot in Fig. 1, a proportional amplitude of the response of that mode, to a given earthquake ground motion, is obtained. Thus, the response spectrum can be seen as a measure of the destructive potential of a given seismic excitation. Of course, the above argument for the fundamental mode can also be extended to any of the “important” modes of vibration of a structure. Furthermore, the spectra of various recorded earthquake ground motions have been appropriately combined into design spectra, by which the ground excitation is specified in all contemporary seismic design codes. For these and other reasons, which will become obvious later, the response spectrum is a convenient and effective engineering tool which has been used universally in structural analysis and design. Apparently, Biot (1933) was the first to introduce the concept of response spectrum and Housner (1959) was the one who applied it to produce earthquake design spectra. Detailed discussions on these matters can be found in, among many others, Gupta (1990), Clough and Penzien (1993), and Chopra (1995).

Response Spectra of Earthquake Ground Motions

An SDOF elastic system of mass m , stiffness k , and damping c is shown in Fig. 2. It can be viewed as a simple single-story structure or an approximation of the first mode of a multistory building. If the SDOF system is subjected to a base motion $u_g(t)$, the equation of motion of the mass m can be written as

$$m[\ddot{u}(t) + \ddot{u}_g(t)] + c\dot{u}(t) + ku(t) = 0 \quad (1)$$

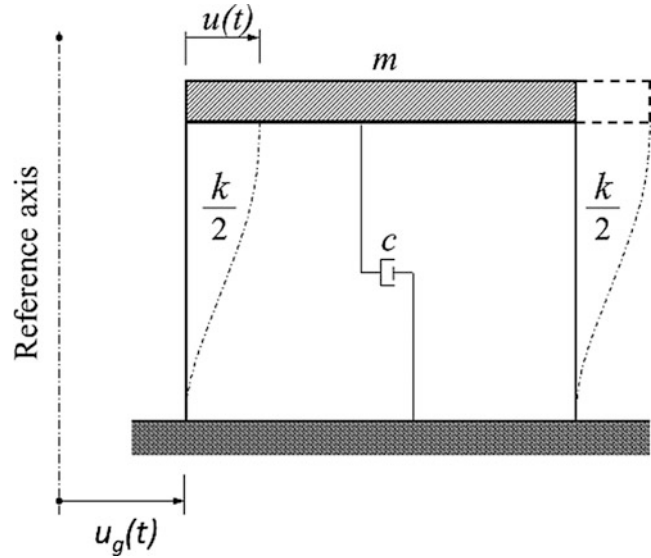
where a dot indicates differentiation with respect to time, $\ddot{u}_g(t)$ is the acceleration of the ground motion, and $u(t)$ is the relative displacement between the mass and the ground. After dividing by the mass m and rearranging terms, Eq. 1 becomes

$$\ddot{u}(t) + 2\xi\omega\dot{u}(t) + \omega^2u(t) = -\ddot{u}_g(t) \quad (2)$$

where $\omega^2 = k/m$ is the circular frequency of the SDOF and $\xi = c/2m\omega$ is its damping ratio (fraction of critical damping). Assuming the motion starts at time $t = 0$ while the structure is at rest, i.e., $u(0) = \dot{u}(0) = 0$, the solution of Eq. 2 can be expressed by the well-known Duhamel integral as

$$u(t) = \frac{-1}{\omega} \left\{ \frac{1}{\sqrt{1-\xi^2}} \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \sin[\omega_d(t-\tau)] d\tau \right\} \quad (3)$$

Earthquake Response Spectra and Design Spectra, Fig. 2 Single-degree-of-freedom system under earthquake excitation



where $\omega_d = \omega\sqrt{1-\xi^2}$ is the damped circular frequency of the SDOF system. Differentiation of Eq. 3 and simple manipulations with the trigonometric identities yield expressions for the velocity and the accelerations in a similar form, i.e.,

$$\dot{u}(t) = \frac{1}{\sqrt{1-\xi^2}} \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \sin[\omega_d(t-\tau) - \alpha] d\tau \quad (4)$$

$$\ddot{u}(t) + \ddot{u}_g(t) = \omega \left\{ \frac{1}{\sqrt{1-\xi^2}} \int_0^t \ddot{u}_g(\tau) e^{-\xi\omega(t-\tau)} \sin[\omega_d(t-\tau) - \beta] d\tau \right\} \quad (5)$$

where $\alpha = \tan^{-1} \left[\frac{\xi}{\sqrt{1-\xi^2}} \right]$ and $\beta = \tan^{-1} \left[\frac{2\xi\sqrt{1-\xi^2}}{1-2\xi^2} \right]$.

A close examination of Eqs. 3, 4, and 5 reveals that all three of them are nearly identical except for the multiplying factor in front of the right-hand side and the phase shift in the argument of the sine term in the integrand. Apparently the only difference in the values of the integrals at any time t is due to the phase shift in the sine

function. While for the velocity Eq. 4, the phase shift α is large, i.e., 90° for $\xi = 0$ and 84.3° for $\xi = 10\%$, the phase shift β in the acceleration Eq. 5 remains small, i.e., 0° for $\xi = 0$ and 11.5° for $\xi = 10\%$. For engineering purposes, the time variation of the response parameters is of no general interest. The crucial information is conveyed by their extreme values as they are related to maximum forces, deformations, etc. In addition, analyses of many strong-motion accelerograms have shown that the maximum values of the integrals in Eqs. 3 and 5 are almost the same, indicating the small influence of the phase shift on the maximum value of the integral. Similar comments can be made even for the large phase shift of the velocity Eq. 4, except for SDOF systems with large period T (small ω). In any case, in view of the random form of the ground acceleration $\ddot{u}_g(t)$, only a numerical evaluation of the integral of Eq. 3 will yield practical results. Alternatively, a solution of Eq. 2 can be obtained directly in a numerical form by a number of well-known methods, e.g., finite difference, Newmark, Runge-Kutta, etc.

It is evident from Eqs. 2 and 3 that the response $u(t)$ to a given ground motion $u_g(t)$ depends only upon the circular frequency ω and the damping ratio ξ and not upon the particular values of the stiffness k and the mass m , i.e., any

two SDOF systems with the same ratio k/m and damping ratio ξ would respond in an identical manner. Therefore, the absolute maximum response of the system can be symbolically written as

$$\max|u(t, \omega, \xi)| = SD(\omega, \xi) \quad (6)$$

where SD is a single-valued function of ω and ξ . A plot of SD versus ω or the period $T = 2\pi/\omega$ or the natural frequency $f = 1/T = \omega/2\pi$, for a fixed value of ξ , is called the *relative displacement* (or simply *the displacement*) *response spectrum* of the given ground motion $u_g(t)$. In view of Eqs. 4 and 5, similar rigorous definitions, like the one of Eq. 6, can be given for the *relative velocity* (or simply *velocity*) *response spectrum* SV and the *absolute acceleration response spectrum* SA of the ground motion $u_g(t)$. However, the form of Eqs. 3, 4, and 5 and the preceding comments about the small-to-negligible influence of the corresponding shift phases suggest the following approximate, but quite accurate for practical purposes, expressions:

$$SV = \omega SD \quad (7)$$

$$SA = \omega^2 SD \quad (8)$$

Due to their approximate nature, the SV and SA spectra defined in Eqs. 7 and 8 are frequently referred to as pseudovelocity and pseudoacceleration response spectra, respectively, and, in order to underline the difference with the rigorously defined SV and SD , the symbols PSV and PSA are also used. It should be noted that the prefix “pseudo” (from the Greek $\psi\epsilon\acute{\upsilon}\delta\omicron$, false) should not be misleading, as in this case it is used to indicate an approximation and not an incorrect concept. In this entry the symbols SV and SA will be used as defined by Eqs. 7 and 8. Figure 3 shows the Lefkada 2003 earthquake record (strong motion, T component) and the corresponding SD , SV , and SA response spectra for two values of the damping ratio, $\xi = 5\%$ and $\xi = 10\%$. Such plots illustrate the maximum response of all SDOF oscillators of practical interest to a given ground motion and, thus,

provide an almost complete description of the effects of the specific ground motion on structural response.

The motivation for the development of the response spectrum concept is rooted on the need to compute in a straightforward manner design related quantities such as maximum base shear forces, moments, etc. For example, the maximum base shear $V_{\max}$ of the SDOF system of Fig. 2 due to a specified ground motion $u_g(t)$ is

$$V_{\max} = ku_{\max} = k SD \quad (9)$$

By direct substitution from Eq. 8 and using the definition of the circular frequency ω , Eq. 9 can be also written in another useful form as

$$V_{\max} = m SA \quad (10)$$

Apparently the maximum shear force can be computed directly from the SA or SD spectra. However, it should be noted that the quantity $m SA$ is the maximum elastic shear force, while theoretically the maximum inertia force is $m \max|\ddot{u}(t) + \ddot{u}_g(t)|$. As explained earlier, the spectrum acceleration SA and the maximum absolute acceleration $\max|\ddot{u}(t) + \ddot{u}_g(t)|$ are nearly equal for almost all practical cases, except for long periods and large damping ratios. The small difference between the elastic shear force and the maximum inertia force is the damping force $c \max|\dot{u}(t)|$.

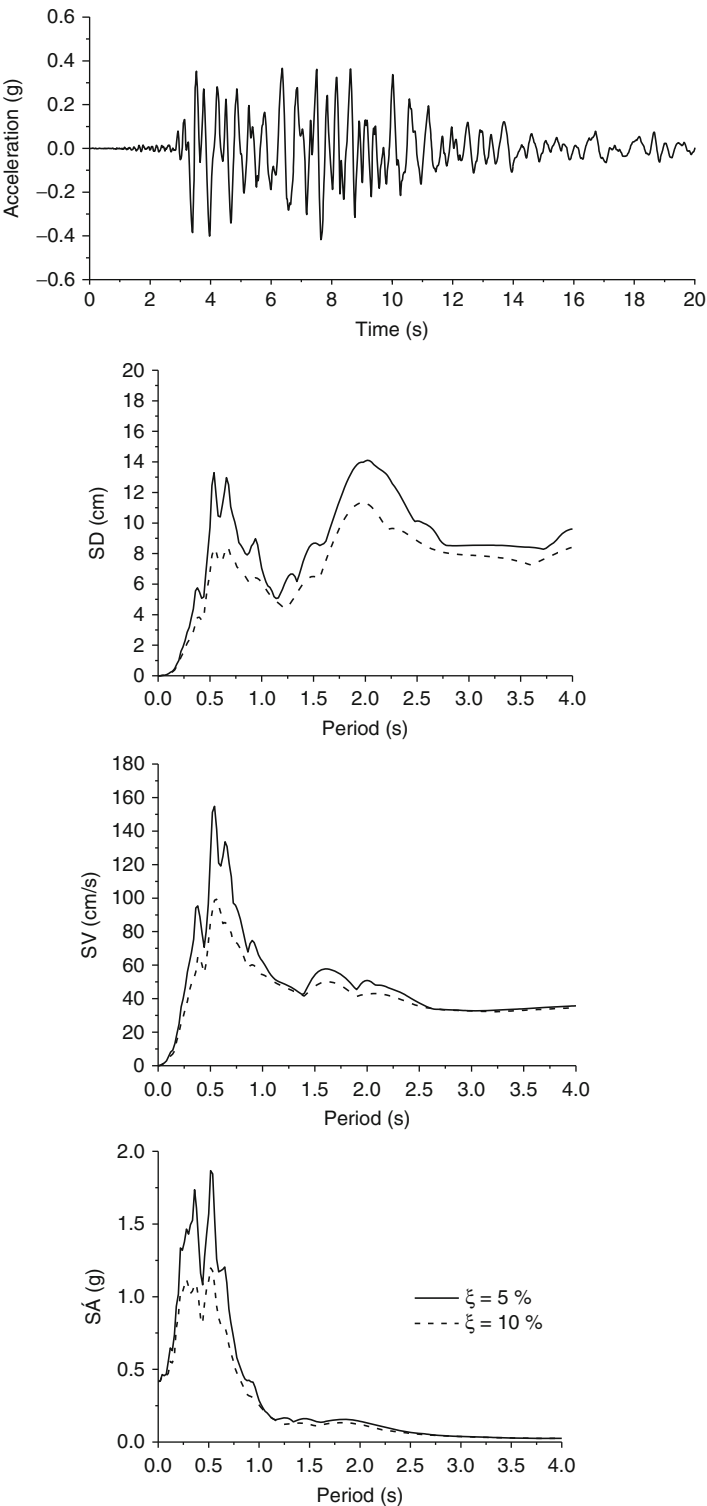
The relative velocity response spectrum, usually of secondary importance, is related to the maximum strain energy of the SDOF system, i.e., SV is equal to the velocity that gives the SDOF system kinetic energy equal to the maximum strain energy due to the earthquake ground motion. This follows from the definition of strain energy and substitution from Eq. 7, i.e.,

$$E_s = \frac{1}{2}k SD^2 = \frac{1}{2}m SV^2 \quad (11)$$

For a wide range of periods, the values of SV are quite close to those of the maximum velocity. However, a considerable difference exists between the response spectrum SV and the

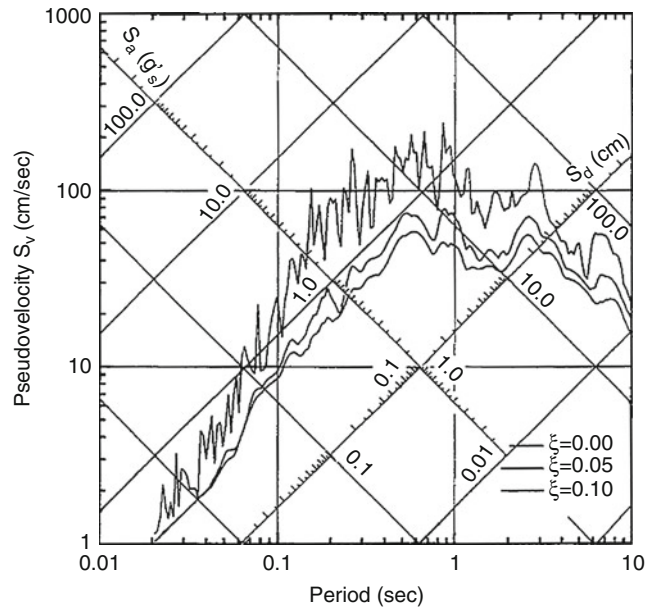
Earthquake Response Spectra and Design Spectra,

Fig. 3 Acceleration record of the Lefkada (Greece) 2003 earthquake (14/08/2003 5:18:48 (GMT), OTE building, Lefkada, T horizontal component, $PGA = 0.42\text{ g}$), and the corresponding response spectra



Earthquake Response Spectra and Design Spectra

Spectra, Fig. 4 Tripartite response spectra plot of the 1940 El Centro (comp. NS) record (After Anagnostopoulos (1997))



maximum relative velocity $\max|\dot{u}(t)|$ for long-period structures. This can be seen on the basis of physical grounds, since for very flexible structures (large T), the elastic forces are small, the mass will exhibit very small displacements, and, thus, the relative displacements between the mass and the ground will be almost identical to the ground displacements, i.e., for $T \rightarrow \infty$, $SD \rightarrow \max|u_g(t)|$, and $\max|\dot{u}(t)| \rightarrow \max|\dot{u}_g(t)|$, but, because of the definition of SV in Eq. 7, $SV = \omega SD = \frac{2\pi}{T} SD \rightarrow 0$, since SD reaches a finite value.

From the definitions of the three response spectra provided by Eqs. 6, 7, and 8 follows the exact mathematical relationship

$$SD = \omega SV = \omega^2 SA \quad (12)$$

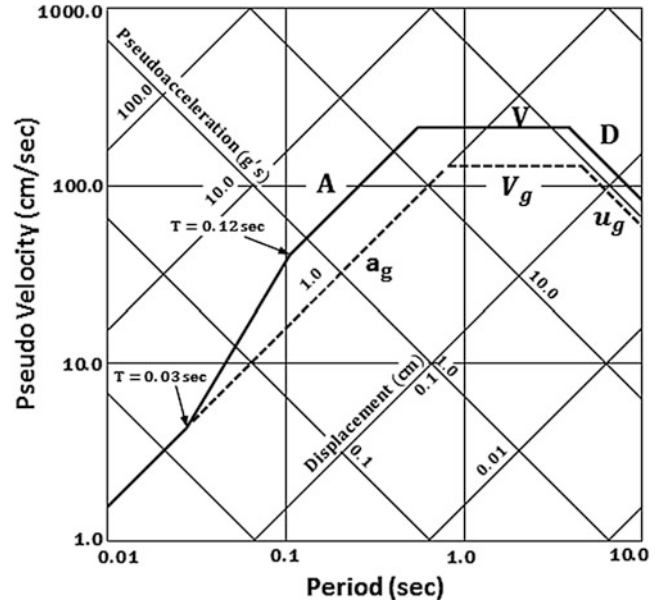
which implies that each of the three response spectra provides the same information as the others about a specific ground motion. Equation 12 also permits plotting, e.g., Anagnostopoulos (1997), all three response spectra by a single curve on a three-way logarithmic graph, as shown in Fig. 4 for the El Centro record and three damping values $\xi = 0, 0.05$, and 0.10 . Such plots become useful because they reveal some general trends, common to all ground

motions. A detailed discussion on these matters can be found in several sources, e.g., Chopra (1995). As explained in the following section, these common trends permit the approximation of average response spectra and the subsequent development of design spectra with high probability of accounting for the general characteristics of future seismic events.

Elastic Design Spectra

The development of earthquake design spectra is the most important application of the response spectra concept. As defined by Anagnostopoulos (1997), a design spectrum is a specification of seismic actions for design purposes, aimed at accounting for the general characteristics of expected ground motions at the site and at minimizing the dependence of earthquake resistant design on the peculiarities of specific ground motions (that will never be repeated in the future). At a first glance, an obvious characteristic of the spectral plots produced from various ground motions is their jagged shape with successive sharp peaks and valleys within small intervals of periods. In addition, each ground motion yields a unique spectral plot with

Earthquake Response Spectra and Design Spectra, Fig. 5 Newmark and Hall (1982) elastic design spectrum ($a_g = 1.0\text{ g}$)



substantial differences to the respective spectral plots of other ground motions, e.g., valleys at one of them may appear as peaks at another, while their magnitudes and periods of occurrence are usually completely unrelated. Therefore, a structural design based on any single ground motion or, equivalently, its response spectrum would assume a substantial risk.

However, spectral plots corresponding to ground motions recorded under similar conditions, e.g., similar soil profiles, comparable earthquake magnitudes, nearly equal epicentral distances, etc., exhibit, in spite of their obvious differences, some characteristic common trends. In addition, it is well known that spectral plots, in general, normalized by their respective peak ground acceleration (PGA) tend to similar values for very short or very long periods. These observations allow the development of design spectra representative of broad classes of soil, epicentral distances, etc. Thus, design ground motions are specified in the form of smooth design spectra disregarding the usual peaks and valleys produced by individual seismic records. The first design spectrum was produced by Housner (1959) on the basis of the four strongest ground motions recorded in the USA up to that time. Later, as the number of recorded ground motions

increased, a number of similar design spectra became available representing a wider range of conditions. Modern design spectral shapes are constructed as smoothed averages of response spectra of many strong ground motions, grouped according to few broad categories of soil conditions and normalized with respect to their probabilistically estimated PGA, e.g., Newmark et al. (1973), Seed et al. (1976), Mohraz (1976), and Newmark and Hall (1982).

The set of rules for the construction of design spectra proposed by Newmark and Hall (1982), the NH spectra, have played a central role in the development of modern seismic design codes. It is based on the specification of a ground motion envelope, at any given site, defined in terms of three parameters, i.e., peak ground acceleration a_g , peak ground velocity v_g , and peak ground displacement u_g . On a three-way logarithmic plot, the ground motion envelope becomes a trapezoid, as shown by the dashed line in Fig. 5.

The values of all three ground parameters can be computed by a seismic hazard analysis, but usually only a_g is computed this way, while v_g and u_g are estimated on the basis of a_g . For example, Table 1, proposed by Newmark et al. (1973), is frequently used to estimate the peak values of ground velocity and displacement.

The actual NH spectrum is an amplification of the ground motion envelope. Its ordinates A, V, and D, as shown in Fig. 5, are obtained by multiplying those of the ground motion envelope by the corresponding amplification factors proposed by Newmark and Hall (1982). These amplification factors are reproduced in Table 2 for two levels of non-exceedance probability (50 % and 84.1 %) and eight values of critical damping.

For periods up to 0.03 s, the design spectra coincide with the ground motion envelope. From this point on the design spectrum slopes until it meets the A line at a period 0.12 s. It is well known that the NH spectrum is not conservative enough for structures with long periods, e.g., periods above 1.0 s, due to several factors, such as ambiguities regarding the combination of modal responses and insufficient depiction of faraway earthquake motions. To remedy this problem, Newmark and Hall (1982) proposed a modification of the NH spectrum for the period range 1.0–6.0 s. It should be noted that although the amplification factors are constant, the shape of the ground motion envelope is not fixed, since the values of the parameters involved depend on local soil conditions or can be determined using detailed seismic hazard analyses taking into

consideration other factors, such as epicentral distance, earthquake magnitude, etc.

Seed et al. (1976) studied the effect of soil conditions based on a number of strong-motion records for various soil profiles, i.e., soft to medium clays and sands, deep cohesionless soils, and stiff soils and rock. The final recommendation of these studies, in the form of a set of smooth, normalized design spectra curves, for 5 % damping, is shown in Fig. 6.

Smooth design spectra, such as those of Newmark et al. (1973), Newmark and Hall (1982), Seed et al. (1976), and Mohraz (1976), provide the basis for most of the design codes in use today, e.g., ATC-3 (1978), UBC (1994), and EC-8 (2004). In particular, the European Standard EC-8 defines the elastic acceleration design spectrum S_e by the following set of explicit formulae:

$$\begin{aligned} 0 \leq T \leq T_B & \quad S_e = a_g S \left[1 + \frac{T}{T_B} (2.5\eta - 1) \right] \\ T_B \leq T \leq T_C & \quad S_e = 2.5a_g S \eta \\ T_C \leq T \leq T_D & \quad S_e = 2.5a_g S \eta \left(\frac{T_C}{T} \right) \\ T_D \leq T \leq 4\text{s} & \quad S_e = 2.5a_g S \eta \left(\frac{T_C T_D}{T^2} \right) \end{aligned} \quad (13)$$

where a_g is the design ground acceleration in g's, η is the damping correction factor specified in reference to the percentage of the critical damping ratio ξ as $\eta = \sqrt{10/(5 + \xi)} \geq 0.55$, S is the soil parameter, and T_B , T_C , and T_D are

Earthquake Response Spectra and Design Spectra,

Table 1 Relationships between ground motion parameters

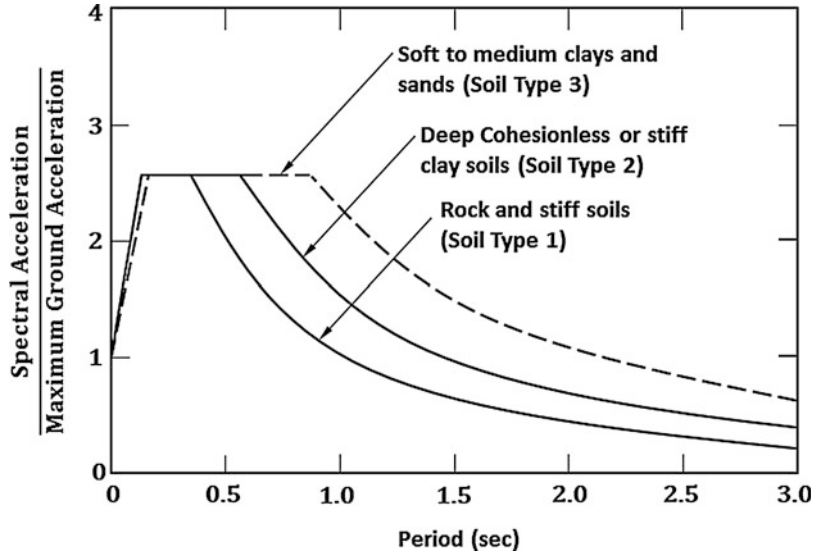
	v_g/a_g (cm/s/g)	$a_g u_g/v_g^2$
Rock	71.0	5.9
Alluvium	122.0	6.0

Earthquake Response Spectra and Design Spectra, Table 2 Newmark and Hall (1982) spectrum amplification factors

Critical damping ratio (%)	One sigma (84.1 %)			Median (50 %)		
	A	V	D	A	V	D
0.5	5.10	3.84	3.04	3.68	2.59	2.01
1	4.38	3.38	2.73	3.21	2.31	1.82
2	3.66	2.92	2.42	2.74	2.03	1.63
3	3.24	2.64	2.24	2.46	1.86	1.52
5	2.71	2.30	2.01	2.12	1.65	1.39
7	2.36	2.08	1.85	1.89	1.51	1.29
10	1.99	1.84	1.69	1.64	1.37	1.20
20	1.26	1.37	1.38	1.17	1.08	1.01

Earthquake Response Spectra and Design Spectra,

Fig. 6 Normalized spectral shapes for a number of soil conditions (After UBS (1994))



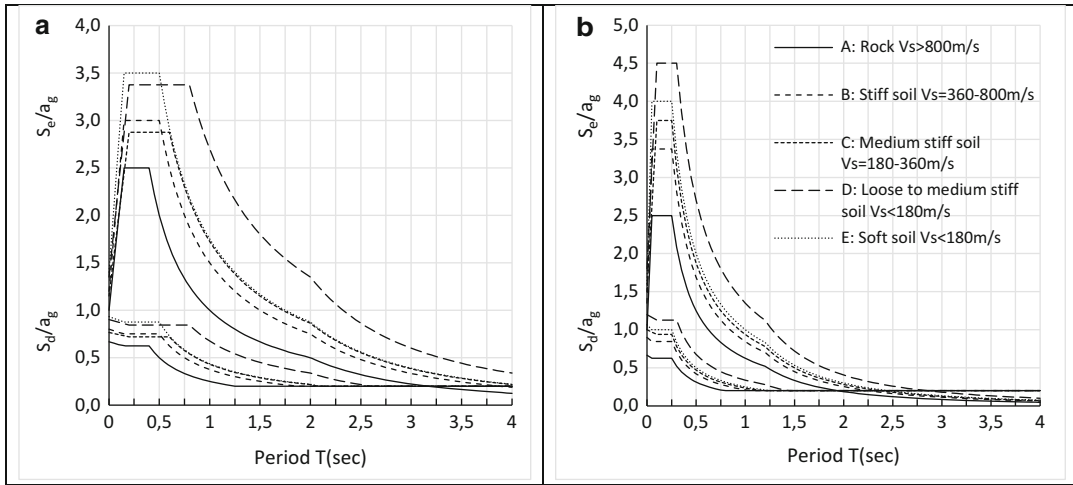
Earthquake Response Spectra and Design Spectra, Table 3 EC-8: parameters involved in elastic response spectra

Soil type	Type 1				Type 2			
	S	$T_B(s)$	$T_C(s)$	$T_D(s)$	S	$T_B(s)$	$T_C(s)$	$T_D(s)$
A	1.0	0.15	0.4	2.0	1.0	0.15	0.4	2.0
B	1.2	0.15	0.5	2.0	1.2	0.15	0.5	2.0
C	1.15	0.20	0.6	2.0	1.15	0.20	0.6	2.0
D	1.35	0.20	0.8	2.0	1.35	0.20	0.8	2.0
E	1.4	0.15	0.5	2.0	1.4	0.15	0.5	2.0

corner points. The recommended values of the parameters S , T_B , T_C , and T_D , for five soil types classified as A, B, C, D, and E are given in Table 3 for the Type 1 and Type 2 spectra. Type 2 spectrum is recommended for sites where the earthquakes that contribute most to the seismic hazard defined for the site for the purpose of probabilistic hazard assessment have a surface-wave magnitude, M_s , not greater than 5.5. The five soil types vary from rock soil profile A with a shear wave velocity $v_s > 800$ m/s to soft soil profile E with $v_s < 180$ m/s. A plot of the EC-8 elastic response spectra S_e , for $\eta = 1$, as shown in Fig. 7.

The design spectra discussed above do not take into account the influence of key seismological factors, such as the epicentral distance or the magnitude of the earthquake, on the frequency content of the motion. This is due to the fact that their shape is fixed, while the actual ordinates of the spectra are obtained by a multiplication with a

scaling factor (the design ground acceleration, obtained by site seismic hazard analysis). As a result, design spectra defined through these procedures fail to provide a uniform non-exceedance probability for all spectral regions. As an example, McGuire (1978) has pointed out that such design spectra undervalue spectral velocities, except for high frequencies, in regions affected by distant earthquakes. A partial cure to this problem has been proposed in ATC-3 where a second variable related to the peak ground velocity is introduced, in addition to the peak ground acceleration. This procedure permits the usage of a variable spectral shape but it requires the availability of two seismic hazard maps, one for the peak ground acceleration and another for the peak ground velocity. However, uniform hazard spectra, providing uniform probability of non-exceedance, are possible nowadays via regression analyses on response spectra of large



Earthquake Response Spectra and Design Spectra, Fig. 7 EC-8 elastic response spectra and design spectra: (a) Type 1 and (b) Type 2 ($\eta = 1$, $q = 4$)

numbers of strong ground motion records available on almost every region of the world, e.g., McGuire (1977) and Joyner and Boore (1982).

Inelastic Response and Design Spectra

In a manner similar to the previously described elastic response spectrum, one can define the inelastic response spectrum as a plot of the maximum response of all nonlinear single-degree-of-freedom (SDOF) systems to a given dynamic excitation. Of course, in this case the type of the nonlinearity of the oscillators of interest should be specified a priori. Inelastic response spectra, based on some nonlinear force–displacement relationship of an SDOF oscillator, are computed similarly to the elastic response spectra. Thus, in Eq. 1 the term $ku(t)$, the elastic resistance force, is substituted by the displacement-depended force $F[u(t)]$ and the equation of motion (2) is now written as

$$\ddot{u}(t) + 2\zeta\omega\dot{u}(t) + \frac{F[u(t)]}{m} = -\ddot{u}_g(t) \quad (14)$$

The simplest and most common case refers to an oscillator whose nonlinear resistance is assumed to be elastoplastic with an initial stiffness k and yield displacement u_y corresponding to a

maximum force F_y , as shown in Fig. 8. Provided the yield quantities F_y and u_y are known properties of the SDOF under investigation, the search for the quantity of interest $u_{\max}$, the maximum displacement, is facilitated by the introduction of the nondimensional ductility ratio μ defined by $u_{\max} = \mu u_y$. Thus, the ductility ratio μ , i.e., the inelastic deformation demand imposed by a ground motion $\ddot{u}_g(t)$, becomes the primary quantity of interest even for design purposes, replacing the maximum force or stress used in elastic design.

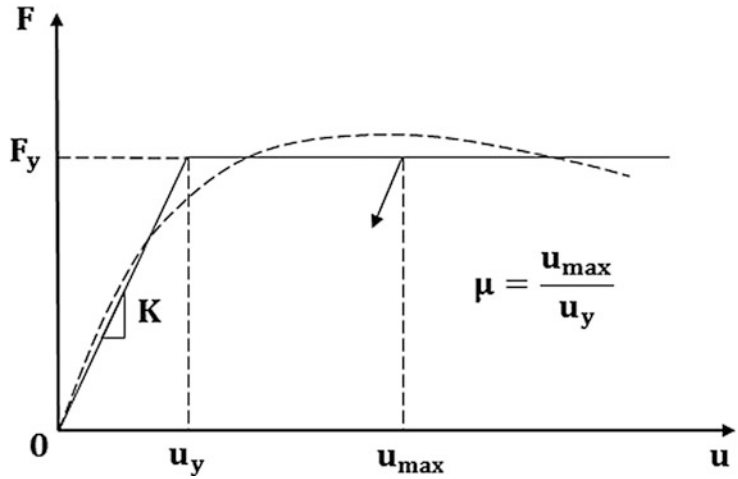
The definitions of the inelastic displacement ($SD = u_{\max}$), (pseudo)velocity (SV) and (pseudo) acceleration (SA) response spectra, and their inter-relationships, follow those of the elastic case. The SA is obtained from the formula providing the maximum elastic force, i.e., $F_y = mSA$, and, as in the elastic case, is a close approximation of the maximum absolute acceleration spectrum. In view of the relationships

$$F_y = ku_y = m\omega^2 u_y, \quad (15)$$

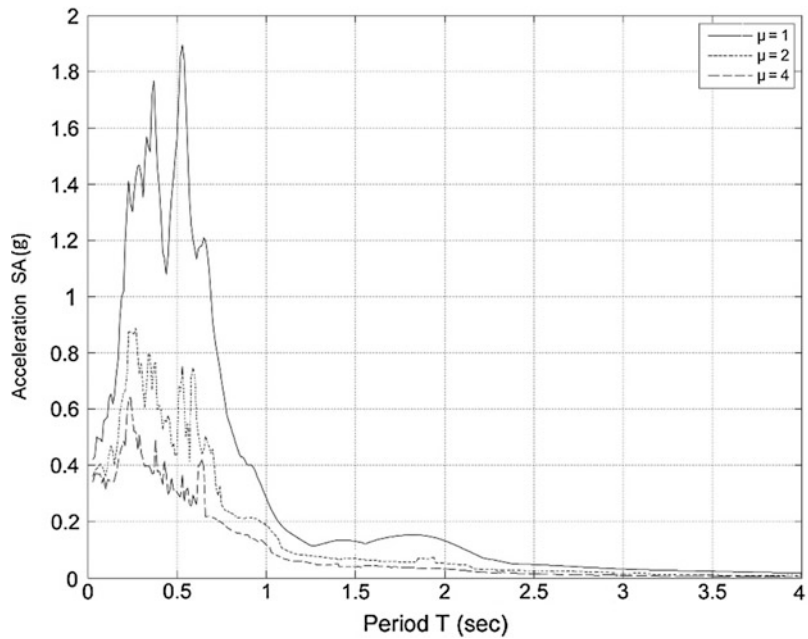
the above definition of SA can be written in alternative forms as

$$SA = \omega^2 u_y = \omega^2 \frac{u_{\max}}{\mu} = \omega^2 \frac{SD}{\mu} \quad (16)$$

Earthquake Response Spectra and Design Spectra, Fig. 8 Idealized elastoplastic system



Earthquake Response Spectra and Design Spectra, Fig. 9 Inelastic response spectra (SA) of the Lefkada (Greece) 2003 earthquake (14/08/2003 5:18:48 (GMT), OTE building, Lefkada, T horizontal component, $PGA = 0.42$ g)



The (pseudo)velocity spectrum can be defined in exactly the same manner as for the elastic case, i.e., $SV = SA/\omega$, in which case Eq. 16 can be expanded to include all three response spectra as

$$SA = \omega SV = \omega^2 \frac{SD}{\mu} \quad (17)$$

Apparently, Eq. 17 expresses the same relation between the three response spectra as its elastic

counterpart Eq. 12, except for the entrance of the ductility ratio μ between SD and the other two response spectra SA and SV.

For specified values of viscous damping ξ and ductility ratio μ , the computation of inelastic response spectra, for any value of ω or T , is a trial-and-error process which aims at calculating the yield force F_y which corresponds to the specified ductility ratio μ . In Fig. 9 the inelastic

response spectra (SA) of the Lefkada 2003 strong-motion record is shown for several values of the ductility factor μ . Of course, the case $\mu = 1$ coincides, as it was expected, with the linear response spectra shown in Fig. 3.

Regarding the development of inelastic design spectra, Newmark and Hall (1982) proposed a set of simple rules to obtain them from the corresponding elastic design spectra. They are based on observations related to the tendencies of the inelastic response spectra, as compared to the corresponding elastic ones, when they are plotted in a tripartite logarithmic graph. Thus, for the various characteristic regions of the elastic design spectra, proposed by Newmark and Hall (1982), the inelastic design spectra are obtained from the corresponding elastic ones by a simple multiplication with a factor which is a function of the ductility ratio μ . For more details the reader is directed to the original publication by Newmark and Hall (1982) and a detailed discussion by Chopra (1995). Following the Newmark and Hall proposals, inelastic design spectra have been developed into the most popular way of estimating the inelastic behavior of structures via standard elastic modal analyses. Modern codes have, in general, adopted this approach by simplifying the Newmark and Hall rules for inelastic design spectra and, at the same time, introducing the so-called capacity design provisions for the purpose of allowing control over the yield patterns of a structure and avoiding excessive local ductility demands, e.g., Anagnostopoulos et al. (1978). Thus, the response modification factor or behavior factor is used to account for the inelastic building behavior. Almost all of the contemporary design codes, e.g., ATC-3 (1978), NEHRP (1994), UBC (1994), and EC-8 (2004), employ a constant response modification factor, dependent upon the building material and type of structure, by which the elastic design spectrum is divided in order to obtain the corresponding inelastic one. In particular, the EC-8 inelastic acceleration design spectra S_d , used in conjunction with standard modal analysis, are specified (see for

comparison the corresponding elastic spectra of Eq. 13) as

$$\begin{aligned}
 0 \leq T \leq T_B & \quad S_d = a_g S \left[\frac{2}{3} + \frac{T}{T_B} \left(\frac{2.5}{q} - \frac{2}{3} \right) \right] \\
 T_B \leq T \leq T_C & \quad S_d = 2.5 a_g \frac{S}{q} \\
 T_C \leq T \leq T_D & \quad S_d = 2.5 a_g \frac{S}{q} \left(\frac{T_C}{T} \right) \\
 T_D \leq T & \quad S_d = 2.5 a_g \frac{S}{q} \left(\frac{T_C T_D}{T^2} \right) \\
 \text{For all periods} & \quad S_d \geq 0.2 a_g
 \end{aligned} \tag{18}$$

The symbols used in Eq. 18 are the same as those used in the corresponding elastic spectra of Eq. 13, while the values of S , T_B , T_C , and T_D remain the same to those listed in Table 3. The behavior factor q , depending on the type of the structure and the building material used, varies from 1.5 for unreinforced masonry structures to 6.0 for the most ductile steel structures. The viscous damping factor η of Eq. 13 is removed since the behavior factor q accounts for the influence of the viscous damping as well. A plot of the EC-8 design spectra S_d , for $q = 4$, is shown in Fig. 7.

Design spectra based on specific values of the ductility factor μ are introduced in the New Zealand seismic code (1992) with certain restrictions that rule out the possibility the ductility demand to exceed the provided capacity.

Summary

The response spectrum concept is defined, and its key role in the characterization of seismic motions, for engineering purposes, is described. The development and the importance of seismic design spectra, both elastic and inelastic, are analyzed and on this basis the main features of several contemporary design codes are outlined. A number of numerical examples, related to response and design spectra, are shown and discussed.

Cross-References

- ▶ [Building Codes and Standards](#)
- ▶ [Conditional Spectra](#)
- ▶ [Engineering Characterization of Earthquake Ground Motions](#)
- ▶ [European Structural Design Codes: Seismic Actions](#)
- ▶ [Probabilistic Seismic Hazard Models](#)
- ▶ [Response-Spectrum-Compatible Ground Motion Processes](#)
- ▶ [Seismic Actions Due to Near-Fault Ground Motion](#)
- ▶ [Selection of Ground Motions for Response History Analysis](#)
- ▶ [Site Response for Seismic Hazard Assessment](#)
- ▶ [Spatial Variability of Ground Motion: Seismic Analysis](#)
- ▶ [Structural Design Codes of Australia and New Zealand: Seismic Actions](#)

References

- Anagnostopoulos SA (1997) Buildings. In: Beskos DE, Anagnostopoulos SA (eds) *Computer analysis and design of earthquake resistant structures – a handbook*. Computational Mechanics Publications, Southampton, pp 369–440
- Anagnostopoulos SA, Haviland RW, Biggs IM (1978) Use of inelastic spectra in aseismic design. *J Struct Div ASCE* 104(ST1):95–109
- Applied Technology Council (1978) Tentative provisions for the development of seismic regulations for buildings, ATC3-06. National Bureau of Standards, Washington, DC
- Biot MA (1933) Theory of elastic systems under transient loading with an application to earthquake proof buildings. *Proc Natl Acad Sci* 19:262–268
- Chopra A (1995) *Dynamics of structures – theory and applications to earthquake engineering*. Prentice Hall, Upper Saddle River
- Clough RW, Penzien J (1993) *Dynamics of structures*, 2nd edn. McGraw-Hill, New York
- European Committee for Standardisation (2004) Eurocode-8: design of structures for earthquake resistance, European Standard EN 1998–1: 2004. European Committee for Standardisation, Brussels
- Gupta AK (1990) *Response spectrum method in seismic analysis and design of structures*. Blackwell Scientific Publications, Boston
- Housner GW (1959) Behavior of structures during earthquakes. *J Eng Mech Div ASCE* 85:109–129
- Joyner WB, Boore DM (1982) Prediction of earthquake response spectra. U.S. Geological Survey, Open-File Report 82–977, p 16
- McGuire RK (1977) Seismic design spectra and mapping procedures using analysis based directly on oscillator response. *Earthq Eng Struct Dyn* 5:211–234
- McGuire RK (1978) Seismic ground motion parameter relations. *J Geotech Eng Div ASCE* 104(GT4):481–490
- Mohraz B (1976) A study of earthquake response spectra for different geological conditions. *Bull Seismol Soc Am* 66:915–935
- NEHRP (1994) Recommended provisions for seismic regulations for new buildings. Building Seismic Safety Council, Washington, DC
- New Zealand Standards (1992) New Zealand Standard 4203: 1992 – code of practice for general structural design and design loadings for buildings, vol 1, Code of practice. Standards New Zealand, Wellington
- Newmark NM, Blume JA, Kapur KK (1973) Seismic design spectra for nuclear power plants. *J Power Div ASCE* 99(PO2):287–303
- Newmark NM, Hall WJ (1982) *Earthquake spectra and design*. Earthquake Engineering Research Institute, Pasadena
- Seed HB, Ugas G, Lysmer J (1976) Site-dependent spectra for earthquake-resistant design. *Bull Seismol Soc Am* 66:221–243
- UBC – Uniform Building Code (1994) International Council of Building Officials, Whittier

Earthquake Return Period and Its Incorporation into Seismic Actions

Yiannis Tsompanakis

Computational Dynamics Research Group,
School of Environmental Engineering, Technical
University of Crete, Chania, Crete, Greece

Synonyms

Earthquake return period; Hazard curves; Performance-based design; Probability of exceedance; Seismic actions; Seismic hazard analysis; Seismic norms

Introduction

Seismic-resistant design of structures aims primarily at preventing loss of life and global collapse.

The most common approach used worldwide is the design of structures with a large ductility capacity, thus, to be able to withstand inelastic deformation during strong ground shaking with a small collapse probability. However, the anticipated damage to structural and nonstructural components and contents that is caused by large deformations can have serious impact on post-earthquake functionality and often require, if not demolition, certainly costly and disruptive repairs. On the other hand, the design of stiffer and stronger, non-ductile structures could minimize deformations and therefore damages. However, issues such as increased material quantities, structural accelerations, and damage to acceleration-sensitive nonstructural components and building contents have to be carefully assessed in such design approaches. An alternative approach toward the limitation of structural damages under strong ground shaking is the modification of structural characteristics, in a way that the capacity is increased and/or demand is reduced. Modern technologies that can be implemented for these purposes are damping devices, base isolation schemes, or a combination of both.

Earthquake frequency occurrence or return period is directly incorporated into seismic actions and contemporary seismic norms, since in general, seismic-resistant structures must have (i) the capacity to resist frequent (with return period 50–100 years), minor earthquakes without damage, (ii) the capacity to resist with limited structural and nonstructural damage infrequent (with return period approximately 500 years) moderate earthquakes, and (iii) the capacity to resist without collapse and life safety endangerment very strong and rare earthquakes (with return period approximately 2,500 years). For more critical infrastructures, even higher seismic demand levels are imposed which is interpreted in terms of a very large return period, e.g., 10,000 years for maximum design earthquake (MDE) motions for major dams in New Zealand.

The majority of seismic norms belong to the category of prescriptive design codes (or limit state design procedures), as their aim is to ensure that the structural design satisfies a number of

checks in order to be considered safe and reliable against collapse. A typical limit state-based design can be viewed as a one-limit state, ultimate strength, or a two-limit state approach, serviceability and ultimate strength. Most of the prescriptive design codes are based on the assumption that structures are capable of dissipating energy through inelastic deformations. It is recommended to design ductile structures provided that ductility remains within the prescribed limits. Seismic loads can significantly be reduced (via the so-called behavior factor) if the structure is designed to be dissipative.

Furthermore, serviceability is usually checked to ensure that the structure will not deflect or vibrate excessively during its service life. Structural assessment is typically performed considering ultimate strength limit state under the design earthquake (typically with return period of 475 years and probability of exceedance 10 % in 50 years) and damage limitation under an occasional earthquake for the serviceability limit state (typically with return period of 72 years and probability of exceedance 50 % in 50 years). These two hazard levels are the most generally used worldwide. In certain norms (e.g., US guidelines (FEMA-750 2009), it is also specified that structures should be capable to withstand (i.e., to avoid global collapse) the so-called maximum considered earthquake (typically, having return period equal to 2,475 years and probability of exceedance 2 % in 50 years).

The most common approach for the definition of the seismic input is the use of a design code response spectrum which is created for a certain return period of earthquake hazard. This is a general approach, which is easy to implement for conventional buildings. Moreover, if higher precision is required, the use of spectra derived from suitable earthquake records is more appropriate. Note that significant dispersion on the structural response due to the use of different records even scaled to the same earthquake intensity can be observed, since ground motions with equal peak ground acceleration (PGA) but different frequency content can have very different demands upon a structure.

Response spectrum method (RSM) has become the basic computational tool in earthquake-resistant design of structures, mainly because of its simplicity as it basically uses two parameters: spectral period and damping ratio. The principal weakness of the RSM is that it does not include the duration and other aspects related to frequency content of a strong motion, thus, especially for higher seismic intensities, and it is incapable to fully cover all aspects of dynamic nonlinear response. To overcome this deficiency, dynamic analysis methods (linear and preferably nonlinear) can be used utilizing historic, simulated, or synthetic acceleration time histories. Toward this goal, recent guidelines (e.g., FEMA-445 2006) suggest the use of more accurate dynamic analysis procedures for structural design and assessment. Input motions should be representative of regional seismicity and must be compatible with each earthquake hazard level; thus, various methodologies have been proposed to select and scale accelerograms (e.g., Katsanos et al. 2010).

The modern conceptual approach for structural design and assessment is the so-called performance-based design (PBD) approach, which is based on the principle that a structure should meet certain performance objectives for multiple seismic hazard levels. Regarding seismic hazard, it ranges from frequent (i.e., having a small return period) earthquakes of low intensity to more catastrophic and rare events that have large period of occurrence. This framework has been introduced in order to increase structural safety against natural hazards. According to PBD, the structures should be able to resist earthquakes in a quantifiable manner and to present levels of possible damage in a prescribed preferable manner (e.g., avoiding soft-storey collapse). In any case, the adoption of a proper stochastic model (usually the Poisson distribution) for the occurrence of earthquake events has to be used within an elaborate seismic hazard analysis. In addition, the construction of a seismic hazard curve is needed to associate a particular ground motion parameter (e.g., peak ground acceleration, spectral acceleration, etc.) with an annual frequency of exceedance or

the probability of exceedance in a given time period (typically for a 50-year period).

Seismic Hazard Analysis

Seismicity at a site is typically defined as a quantitative estimation of ground shaking, which may be represented most commonly by peak ground acceleration (PGA) (as well as velocity (PGV) and displacement (PGD)), or response spectrum ordinates (mostly accelerations). It has to be stressed that in this work, the emphasis on seismic hazard in terms of return period is given on ground shaking. Nonetheless, similar analysis related to kinematic distress and various soil failure types (fault rupture (Angell et al. 2003), liquefaction (Kramer and Mayfield 2007), landslides (Saygili 2008), etc.) due to earthquakes needs to be considered in certain occasions, especially for lifelines that are frequently exposed to such perils. The severity of these hazards (i.e., probability of exceedance) used in performance evaluation should be compatible with that used in the specification of ground shaking hazard levels (FEMA-350 2000).

In any case, the main aim in seismic hazard analysis is the estimation of the most crucial parameters related to earthquake hazard that is expected to occur at a given site during a predefined time interval, the so-called return or recurrence period. Seismic hazard analysis requires the knowledge of certain important data for the wider neighborhood of the examined site, such as geologic evidence of earthquake sources, fault activity, historical seismicity data, etc. Past earthquake data provide valuable information for predicting the magnitude and occurrence frequency of future earthquakes and, thus, are most valuable in seismic hazard analysis. Recently a substantial effort has been devoted to the development of a global earthquake model (GEM) and open-source software for seismic hazard and risk assessment (for more details, visit <http://www.globalquakemodel.org/openquake/about/engine/>).

At this point, it is necessary to clarify the difference between the two commonly used terms “hazard” and “risk.” These terms are

often used interchangeably, which is misleading. For instance, one of the main hazards associated with earthquakes is ground shaking irrespective of its consequences. On the other hand, seismic risk is associated with structural damage and collapse and, possibly, loss of life. Hence, in general, seismic hazard analysis describes the potential for dangerous, earthquake-related natural phenomena, mainly for ground shaking but also for fault rupture, liquefaction, etc. Utilizing this information, seismic risk analysis assesses if the probability of occurrence of losses (human, social, financial) associated with such seismic hazards will exceed a target value in a specified period of time.

Seismic hazard analysis can be performed using two approaches, namely, deterministic seismic hazard analysis (DSHA) and probabilistic seismic hazard analysis (PSHA). Both use similar information to determine the “design earthquake,” while other hazard levels can also be established. Basic elements of deterministic analysis are also included in the probabilistic approach. The main difference is that the probabilistic approach takes into account the uncertainties and the likelihood of an actual earthquake exceeding the design ground motion. Examples of the outcome of the two approaches are as follows (FEMA-451B 2003):

- Deterministic seismic hazard analysis: “The earthquake hazard for the site is a peak ground acceleration of 0.24 g resulting from an earthquake of magnitude 6.0 from a specific fault at a distance of 50 km from the examined site.”
- Probabilistic seismic hazard analysis: “The earthquake hazard for the examined site is a peak ground acceleration (PGA) of 0.50 g with a 2 % probability of being exceeded in a 50-year period.”

Comparisons among PSHA and DSHA have been made by many researchers in the field (e.g., Krinitzsky 1995; Bommer et al. 2000). Overall, DSHA approach provides a clear and simple method for computing seismic hazard, and its assumptions can be easily discerned. Using DSHA, it is relatively easy to derive

comprehensible hazard scenarios. Nonetheless, it does not have any mechanism to account for uncertainties, while findings based on DSHA can easily be upset by the occurrence of new earthquakes. On the other hand, PSHA is capable of integrating a wide range of data and uncertainties into a flexible framework. Unfortunately, its highly integrated framework can obscure those elements which influence more the results, and its highly quantitative nature can lead to false impressions of high accuracy which is not always true.

In general, none of the two methods is superior to the other for all cases, although PSHA is more frequently used. DSHA is still used for some significant structures, such as nuclear power plants, large dams and bridges, hazardous waste containment facilities, etc. Occasionally, the best solution is to use DSHA and PSHA in a complimentary manner, i.e., DSHA may be used as a “cap” for probabilistic analyses to define the “maximum expected earthquake” and PSHA for the more frequent operational basis earthquake. Nevertheless, the most important is to realize that both methods are simply tools to assist in earthquake mitigation. In any case, the results of a seismic hazard analysis, performed either deterministically or probabilistically, are needed by engineers to perform seismic analysis and design. In the sequence, the basic ingredients of the two methodologies are briefly described, with emphasis on earthquake return period and its incorporation into both approaches.

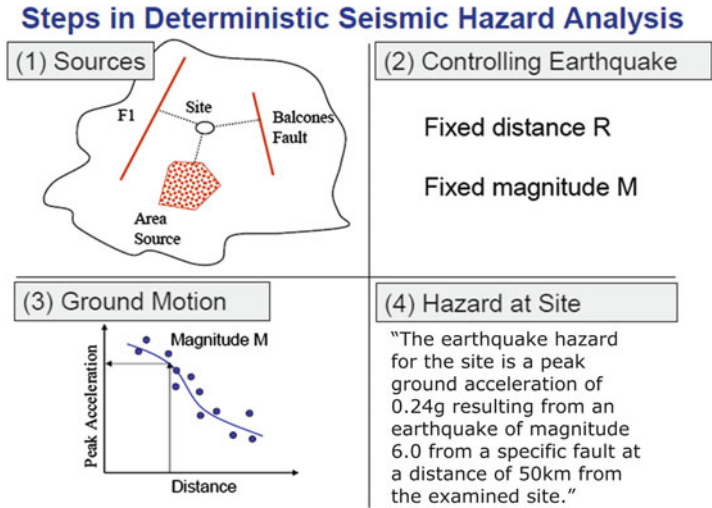
Deterministic Seismic Hazard Analysis

DSHA is based on the magnitude and distance, which yields the largest ground motion amplitude at the site of interest. Under this perspective, the four primary steps in DSHA (see Fig. 1) are as follows (Kramer 1996):

1. Identification and characterization of all sources.
2. Selection of “controlling earthquake”.
3. Selection of source-site distance parameter.
4. Definition of hazard using controlling earthquake.

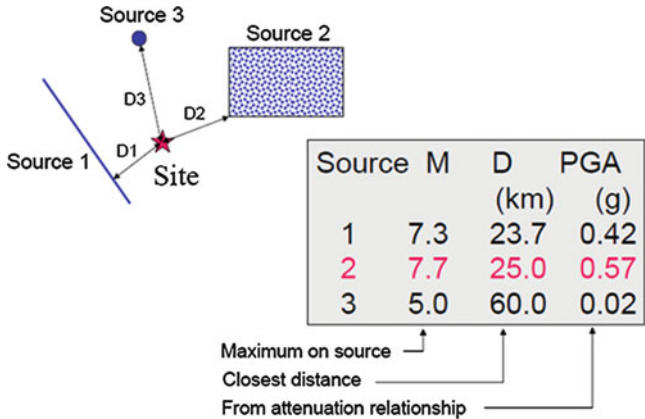
Earthquake Return Period and Its Incorporation into Seismic Actions,

Fig. 1 Basic steps in DSHA (Modified from FEMA-451B (2003))



Earthquake Return Period and Its Incorporation into Seismic Actions,

Fig. 2 Example of DSHA (Adopted from FEMA-451B (2003))



The first step is to identify all possible sources of ground motion, while some may be rather easy to identify (e.g., a known active fault), others may be more difficult to describe. Subsequently, the controlling earthquake needs to be defined, and this involves engineering judgment. For instance, if one wants to design for the largest earthquake that could ever occur at the site (using perhaps an estimate of seismic moment) or the largest motion that has occurred within a time period (e.g., the past 500 years). Note that as aforementioned, there is no inclusion of aspects related to probability of occurrence related to regional seismicity.

Selection of source-site distance parameter is also important for the DSHA process. In DSHA, it is common to use the closest distance from

a source to the examined site. As the known earthquakes will have occurred at a distance that is not likely to be the same as the distance to the site, certain correction needs to be made to take into account geometric spreading of seismic waves and seismic energy absorption. This is done utilizing attenuation models that have been established for each region, as it is very important to use attenuation relationships that characterize local geotectonics.

Figure 2 presents a very simple example of DSHA, in which three sources were considered, while in each case the closest distance to the site was used. Note the different magnitudes for the different sources, which are consistent with the selected “maximum earthquake”. The peak ground acceleration (PGA) at the site was

obtained from an appropriate attenuation relationship. The motion with the greatest resulting PGA is chosen as the controlling earthquake. Note that this is not sufficient to establish seismic risk, since the effect of the motion on the examined structure is not addressed. It is worth mentioning that under certain circumstances, the more distant earthquake, though having a lower effective PGA, could produce waves at a frequency that could be more damaging for a specific structure.

Identification and characterization of all sources can be performed in DSHA utilizing slip rate approach, in which if average displacement over the slip surface, D , relieves stress/strain buildup by movement of the lithospheric plates over a certain time period, t . Therefore, $D = S \cdot t$, where S is the slip rate at the examined fault. Subsequently, the seismic moment is given by

$$M_o = \mu \cdot A \cdot S \cdot t = \mu \cdot A \cdot D \quad (1)$$

where μ is the rigidity or shear modulus of the bedrock, usually taken equal to $\approx 3 \times 10^{11}$ dyne/cm², and ζ is the rupture area on the fault plane where the slip occurs during the earthquake. In addition, the “moment rate” can be defined as $M_o/t = \mu \cdot A \cdot S$. The seismic moment is transformed to earthquake magnitude or energy according to simple logarithmic expressions.

The resulting hazard statement of DSHA is basically a “scenario” earthquake for seismic design (i.e., design earthquake). The rather conservative way that DSHA is commonly used produces a worst-case scenario. DSHA provides no indication of how likely design earthquake is to occur during lifetime of a structure. Design earthquakes may occur every 200 years in some places and every 10,000 years in others. Furthermore, DSHA require subjective opinions of the seismologist on input parameters, while variability of effects is not rationally accounted for.

Probabilistic Seismic Hazard Analysis

Probabilistic seismic hazard analysis (PSHA) was introduced by Cornell (1968). In contrast to DSHA, it assumes many scenarios, in which it

considers all magnitudes, all distances, and all effects. Furthermore, PSHA considers uncertainty in location, size, rate of earthquake occurrence, predictive relationship, etc., and combines them to compute probabilities of different levels of ground shaking. In particular, PSHA integrates site’s total seismicity for a given time period and provides strong-motion parameters within predefined confidence levels, in the form of hazard curves, uniform response spectra, etc., for various return periods. In general, as shown in Fig. 3, PSHA consists of four primary steps (Kramer 1996):

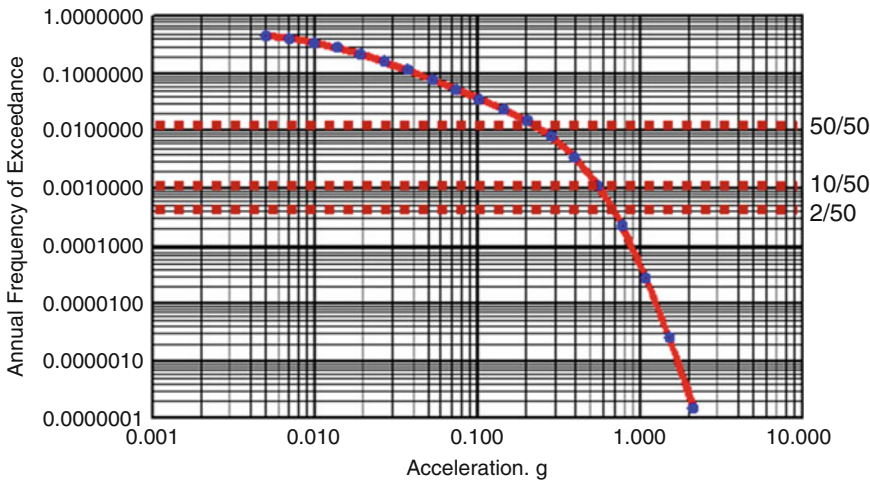
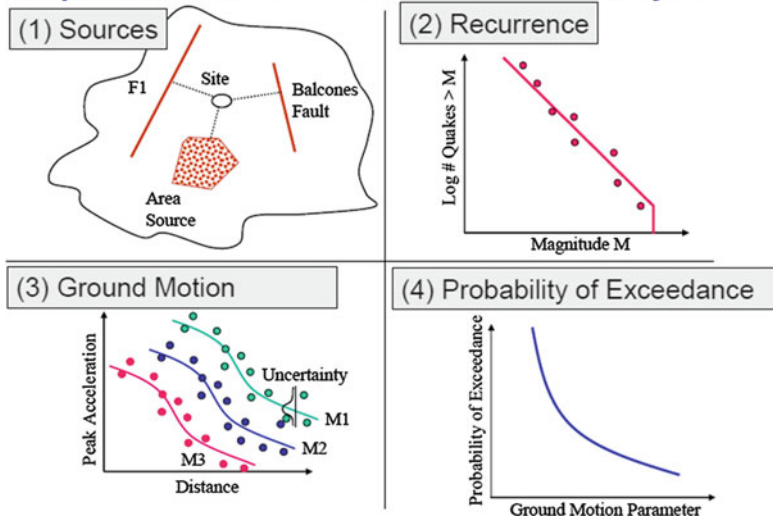
1. Identification and characterization of all source zones (areas or faults) based on historical seismicity, tectonics, and geology.
2. Characterization of seismicity of each source, i.e., to determine an upper magnitude limit and develop a magnitude-recurrence relationship for each source zone.
3. Determination of a predictive relationship for seismic parameter motions (such as PGA) from each source.
4. Perform probabilistic calculations to compute the probability distribution of ground motion at the site and select the design value at the appropriate probability level.

Ground motion parameters that can be derived from a PSHA include: PGA and spectral acceleration (S_a) hazard curves, uniform response spectra (URS), de-aggregation of hazards by magnitude and distance, and parameters for assisting in time-history selection for dynamic analysis. A seismic hazard curve is constructed by plotting the rate of exceedance of the seismic parameter for different seismic intensity levels. In order to assess that a certain engineering demand parameter (e.g., drift) will be exceeded in a building located at a certain site, it is essential to evaluate the accuracy in terms of the mean annual frequencies (MAF) of exceedance given in a form of hazard curve, as in Fig. 4 in terms of PGA values (or alternatively using spectral acceleration values). Similar curves can be constructed for a kinematic type of seismic hazard. For example, utilizing the so-called probabilistic fault

Earthquake Return Period and Its Incorporation into Seismic Actions,

Fig. 3 Basic steps in PSHA (Adopted from FEMA-451B (2003))

Steps in Probabilistic Seismic Hazard Analysis



Earthquake Return Period and Its Incorporation into Seismic Actions, Fig. 4 Typical hazard curve of PGA values in terms of annual frequency rate

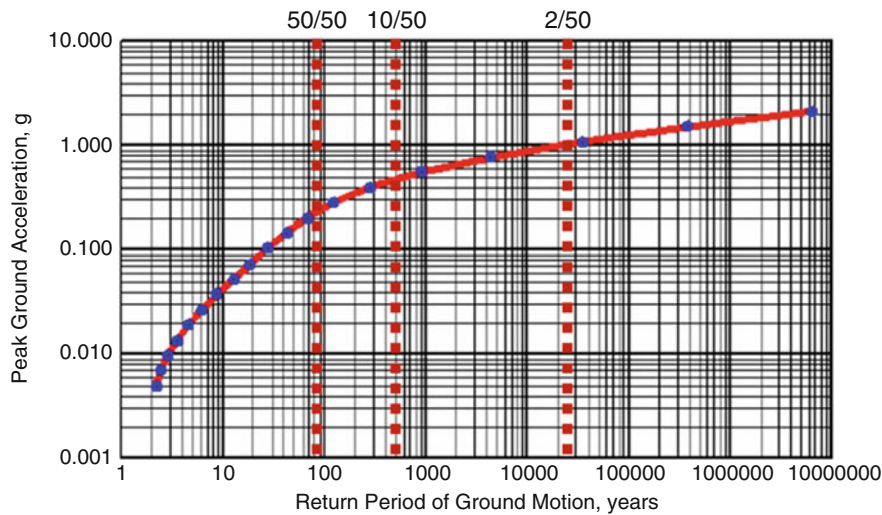
displacement hazard analysis (PFDHA), one can derive fault displacement hazard curves that represent annual frequencies (and corresponding return periods) of certain fault slippage level (Angell et al. 2003).

Frequently, fractile curves (e.g., 15th, 50th, 85th percentiles) of hazard curves are also plotted together with the mean curve to illustrate data scattering. Note that differences between mean and median curves can be observed. Equivalently, seismic parameters can be interpreted as

a function of return periods, as shown in Fig. 5. Tools for producing hazard curves as those shown in Figs. 4 and 5 as well as response spectra, etc., can be found at US Geological Society (USGS) site: <http://earthquake.usgs.gov/hazards/>.

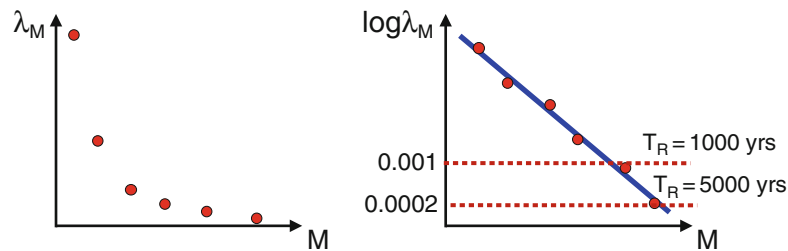
Earthquake Occurrence Frequency

According to PSHA, a given source can produce different earthquakes: low-magnitude earthquakes occur often and cause fewer damages,



Earthquake Return Period and Its Incorporation into Seismic Actions, Fig. 5 Typical hazard curve of PGA in terms of return periods

Earthquake Return Period and Its Incorporation into Seismic Actions, Fig. 6 Schematic representation of magnitude and mean annual recurrence in exponential and log-linear forms



while large magnitude and more severe events occur rarely. The latter have longer return periods and lower probability of exceedance, yet they are more catastrophic; thus, they are referred as low probability – high consequence events. In a probabilistic analysis, the recurrence relationship (e.g., frequency of earthquakes above a certain magnitude) needs to be included as it is the major uncertainty of the process. All major uncertainties are included in PSHA, and the result is a seismic hazard curve that relates the design motion parameter to a probability of exceedance. Hence, for the design of a bridge for a ground motion with a 2 % probability of being exceeded in a 50-year period (which is transformed to annual probability), the ground motion parameter (e.g., PGA) is obtained from the seismic hazard curve (such as the one in Fig. 4).

Typically, the seismicity is interpreted using a specified relationship between magnitude and likelihood of occurrence and is called “recurrence relationship,” which indicates the average rate at which an earthquake of a particular magnitude, M , will be exceeded as shown in Fig. 6. Such relationships are determined from regression analysis of historical earthquake data. The most popular is Gutenberg-Richter model which is a rather simple mathematical formulation that is reliable for intermediate scale events but may be insufficient for small or for large earthquakes; thus, more advanced models have been proposed (Kramer 1996).

According to the Gutenberg-Richter recurrence, law earthquake magnitudes are exponentially distributed and the mean rate of occurrence (i.e., the number of earthquake events/year equal

to or greater than magnitude M), λ_M , is defined in log-linear or exponential form (see plots in Fig. 6) as follows:

$$\log \lambda_M = a - b \cdot M \Rightarrow \lambda_M = 10^{(a-b \cdot M)} = e^{(\alpha - \beta \cdot M)} \quad (2)$$

and following the above formulation, the event return period is directly given by

$$T_R = \frac{1}{\lambda_M} = \frac{10^{b \cdot M}}{10^a} \quad (3)$$

where a and b (while $\alpha = 2.303 \cdot a$ and $\beta = 2.303 \cdot b$) are regional parameters determined from tectonic and geological data, rate of seismicity, etc.; typically, b is close to unity ($b = 1.0 \pm 0.5$) and represents the relative likelihood of large and small events.

Temporal Earthquake Uncertainty

The temporal uncertainty of earthquakes is included in PSHA by considering the distribution of their occurrence in time, which can be modeled by random models, such as Bernoulli, Poisson, and Markov processes. In this manner, it is possible to calculate the probability of the number of occurrences over a certain time interval. The occurrence of an earthquake for a certain time interval is usually modeled as a Poisson process that describes the number of occurrences of an event (not only earthquakes but floods and other natural disasters as well) during a specified time interval and provides also the return period of such event. The main assumptions of Poisson model are as follows (Kramer 1996):

1. The number of occurrences over a given time interval are independent of the number of earthquakes that occur in any other time interval (i.e., the system has no memory from past seismic history, while there are other time-dependent models that keep track of occurred events).
2. The probability of occurrence in a very short time interval is proportional to the length of the interval.

3. The probability of more than one occurrence in a very short time period is negligible.

According to the popular Poisson model, the temporal distribution of earthquake recurrence for a certain intensity level (magnitude, PGA, etc.) is given by

$$P[N = n] = \frac{\mu^n e^{-\mu}}{n!} \text{ or } P[N = n] = \frac{(\lambda \cdot t)^n e^{-\lambda \cdot t}}{n!} \quad (4)$$

where n is the number of occurrences and $\mu = \lambda \cdot t$ is the average number of occurrences in the time interval of interest, t , and λ is the annual rate, which is directly related to the return period of such event as shown in Eq. 3. Hence, the probability that an earthquake having a specific ground motion parameter, such as magnitude, higher than a certain level, will occur at least once in t years is given by

$$\begin{aligned} P[N \geq 1] &= P[N = 1] + \dots + P[N = \infty] \\ &= 1 - P[N = 0] = 1 - e^{-\lambda \cdot t} \\ &= 1 - e^{-t/T_R} \end{aligned} \quad (5)$$

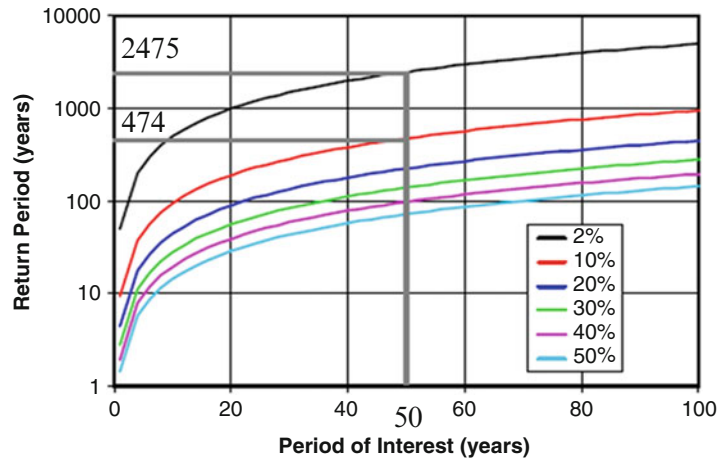
The above expression can be easily combined with a recurrence law (as the one in Eq. 2) to predict the probability of at least one exceedance of a specific magnitude M , λ_M , in a period of t years as follows:

$$P[N \geq 1] = 1 - e^{-\lambda_M \cdot t} \quad (6)$$

An example of using Eq. 5 is to calculate the probability that an event that occurs on average every 1,000 year will occur at least once in a $t = 100$ -year period. Thus, $\lambda = 1/1,000 = 0.001 \Rightarrow P = 1 - \exp[-(0.001)(100)] = 0.0952 = 9.52\%$, while the probability that one such event will occur at least once in a 1,000 year period is $P = 1 - \exp[-(0.001)(1,000)] = 0.632 = 63.2\%$. Obviously, this probability value (63.2 %) is the same for all hazard level return periods when the examined time interval is equal to the return period: $t = T_R$.

Earthquake Return Period and Its Incorporation into Seismic Actions,

Fig. 7 Example of relationship between return period, time interval of interest, and probability of exceedance PSHA (Adopted from FEMA-451B (2003))



Apart from the probability of exceedance, the level of ground shaking that has a certain probability of exceedance in a given period can be determined via Eq. 5 and a hazard curve as it will be shown in the next section. Furthermore, solving Eq. 5 for annual rate (events per year), λ , gives

$$P[N \geq 1] = 1 - e^{-\lambda t} \Rightarrow \lambda = -\frac{\ln(1 - P)}{t} \quad (7)$$

and the corresponding return period, T_R , is given by

$$T_R = \frac{1}{\lambda} = -\frac{t}{\ln(1 - P)} \left(\approx \frac{t}{P} \right) \quad (8)$$

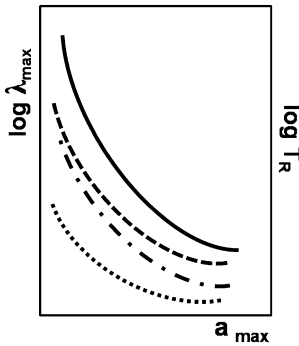
For example, the annual rate of exceedance for an event with a 10 % probability of exceedance in 50 year, or in short 10/50 hazard level, is given by $\lambda = -(\ln(1-0.1)/50) = 0.0021/\text{year}$, while the corresponding return period is $T_R = 1/\lambda = 475$ years. Similarly, for 2 % in 50 year, $\lambda = 0.000404/\text{year}$ and $T_R = 2,475$ years.

As indicated in the last part of Eq. 8, for low probabilities (or long return periods), the return period is approximately equal to t/P . For example, for a period of 100 years and return period of 2,500 years, the 100-year probability is roughly equal to $P = t/T_R = 100/2,500 = 4$ % approximately equal to 3.96 % which is the exact outcome of Eq. 8. Nonetheless, this approximation is

not valid for higher probabilities. For instance, the actual return period of 50 % probability of exceedance in 50 years is 72 years, not 100 years as suggested by the approximate formula. Note also that since λ depends on P and t values, there are many pairs of P and t that result to the same return period.

Figure 7 depicts the relationship between return period, time interval of interest, and the related probability of exceedance. For instance, a 2 % probability for an event of being exceeded in 50 years has a return period of 2,475 years, while a 10 % probability of exceedance in 50 years has approximately a 475-year return period. Such curves are very useful for assessing multiple hazard levels in performance-based design approaches. For structural design, a 50-year base line is generally used, since this is estimated as the service life of a typical building. For critical infrastructures, such as dams, a longer return period is more appropriate as the required functioning life is much longer and failure consequences more severe.

Probability of exceedance must be associated to a time interval of interest and to be connected to a certain return period, which is simply described by its value. For example, an earthquake with a 2 % probability of being exceeded in 100 years has a return period of $T_R = -100/\ln(1-0.02) = 4,950$ years. Note that since the precision of knowledge of this natural phenomenon is not perfect; thus, the return periods can be



Earthquake Return Period and Its Incorporation into Seismic Actions, Fig. 8 Example of seismic hazard curves derived from PSHA

rounded, e.g., the computation of a return period of 474 years for the 10 %/50 year hazard level can be rounded to 475 years and even more for higher hazard levels (i.e., $T_R = 2,475$ to be set 2,500 for 2/50 level). In addition, utilizing Poisson model, one can compare different probabilities for different time intervals. For instance, utilizing Eq. 7 it is straightforward to compare a 2 %/50-year event with a 10 %/250-year event. For 2 %/50 years, $T_R = 50/(-\ln(1-0.02)) = 2,475$ years, while for 10 %/250 years, $T_R = 250/(-\ln(1-0.10)) = 2,372$ years. Hence, these events (expressed with different probabilities) are not exactly equal from a statistical point of view, yet they can be considered equivalent from an engineering perspective.

Seismic Hazard Curves

As shown in Fig. 8, a seismic hazard curve represents the mean annual rate of exceedance of a particular ground motion parameter (peak or spectral accelerations, etc.). In other words, the vertical axis provides the (desired) level of probability, while the horizontal axis depicts a ground motion parameter, such as peak ground acceleration. A seismic hazard curve is the ultimate result of a PSHA after combining all uncertainties and performing necessary probability computations. Contribution of independent sources can be interpreted in different seismic hazard curves (to illustrate which sources are more important) together with the total seismic

hazard curve, as presented with the continuous line which is the outcome of all dashed lines in Fig. 8.

The potential usages of seismic hazard curves are illustrated in Fig. 9. In the left plot (Fig. 9a), the aim is to calculate the probability of exceeding $a_{\max} = 0.35$ g in a 50-year period. As it is shown in the graph, $\lambda = 0.001$; thus, $P = 1 - e^{-\lambda t} = 1 - \exp[-(0.001)(50)] = 0.049 = 4.9$ %. Similarly, in a 500-year period, the probability of exceedance is $P = 1 - \exp[-(0.001)(500)] = 0.393 = 39.3$ %. In a reverse manner, as depicted in the right plot (Fig. 9b), to calculate what peak acceleration has a 10 % probability of being exceeded in a 50 year period, i.e., 10 % in 50 year, then $\lambda = 0.0021$, or $T_R = 475$ years and as it can be derived from the seismic hazard curve, the $a_{\max}$ value corresponding to $\lambda = 0.0021$ is equal to 0.24 g.

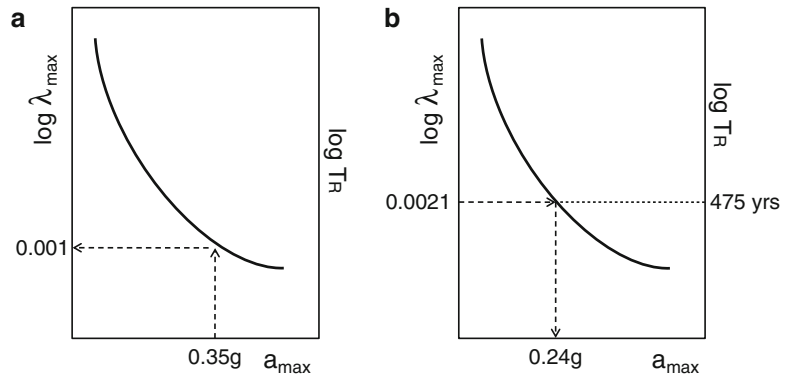
Uniform Response Spectra

Uniform response spectra (UHS) represent spectral ordinates (most frequently accelerations) and are widely used in seismic norms in simplified forms, as the one shown with the continuous line in Fig. 10. Developed from PSHA, UHS is a cumulative graph (constructed from many response spectra derived from records as the one with the dashed line shown in Fig. 10 which is the outcome of a pulse-like accelerogram) that provides, for each spectral period, the spectral amplitude that has a specified probability of exceedance or return period. Therefore, the probability of exceeding a UHS is constant (or uniform) as a function of return period. In other words, a UHS represents a composite of all earthquakes that contribute to site-specific hazard curve; thus, response spectra have the same statistical characteristics as the hazard curve used for their generation. Note that a UHS may not resemble the response spectrum from any specific earthquake magnitude and distance of the site. In addition, it may be too conservative for dynamic analysis and may not be appropriate for artificial ground motion generation.

If a series of seismic hazard curves is developed for a range of different parameters (e.g., PGA or spectral acceleration for 0.2 s or 1 s)

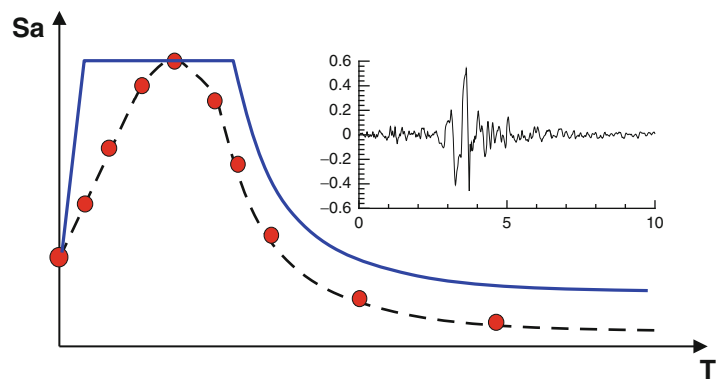
Earthquake Return Period and Its Incorporation into Seismic Actions,

Fig. 9 Examples of using seismic hazard curves for calculating: (a) exceedance probabilities, and (b) ground motion parameters



Earthquake Return Period and Its Incorporation into Seismic Actions,

Fig. 10 Example of typical response spectrum pattern in seismic norms (continuous line) and a response spectrum derived from a pulse-like accelerogram (dashed line)

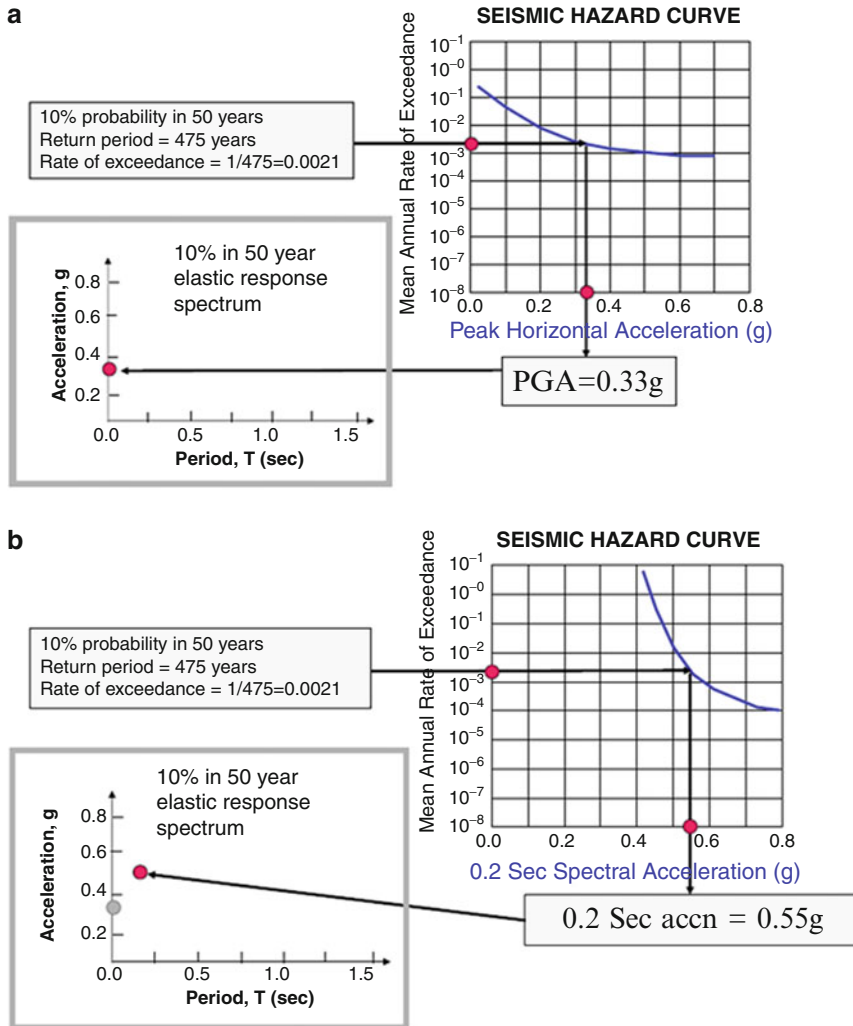


and values are extracted for a certain confidence level (e.g., a 10 % in 50 year probability of exceedance), then a response spectrum may be created by plotting the parameter magnitudes in terms of spectral period, T . The plot at the lower left in Fig. 11a shows the first point on such a spectrum, in this case: the PGA with a 10 % probability of being exceeded. Similarly, in Fig. 11b, the point representing the 0.2 s (5 % damped) spectral ordinate with the same (10 %) probability of being exceeded in 50 years is plotted. In this manner, the whole response spectrum is constructed as shown in Fig. 10.

If the process is repeated for every return period of interest, then design spectra for different hazard levels can be constructed, as shown in Fig. 12. If the aforementioned procedure is continued for many earthquake records, the result is a complete UHS that can be used in a simplified form in seismic norms (see Fig. 10). Such multiple spectra can be used in response spectrum analysis of a structure within a performance-based design

framework. Similar plots can be produced for spectral displacements or velocities.

Note that the information derived from PSHA may be used to assist in the development (selection and scaling) of suitable ground motion time histories (Katsanos et al. 2010). For this purpose, the ground motion amplitudes is scaled in a way that the spectrum of the scaled ground motion closely matches a target UHS. Nonetheless, this can lead to inaccurate results, since the UHS is a composite of hundreds if not thousands of earthquakes, any two of which are very unlikely to occur simultaneously. Typically, the short period part of the UHS is representative of moderate nearby earthquakes, while the long-period part reflects the hazard from larger, more distant events. Figure 13 illustrates more clearly this point. Although it is possible to generate a realistic earthquake ground motion that matches small or large earthquakes, a generated earthquake that matches the whole UHS would be unrealistic and most probably too demanding for the structure.



Earthquake Return Period and Its Incorporation into Seismic Actions, Fig. 11 Example of using seismic hazard curves for generating a design response spectrum

to determine: (a) PGA and (b) S_a for 0.2 s (Adopted from FEMA-451B (2003))

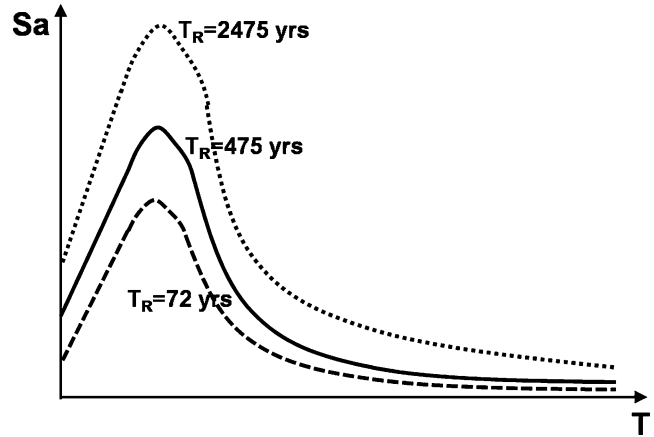
Proper hazard maps can be used to construct uniform response spectra that can be used for seismic design of structures. For instance, in the United States (US), seismic provisions are based on probabilistic maps developed by USGS after performing elaborate probabilistic seismic hazard analyses of the entire USA (Petersen et al. 2008). Using these maps, which provide spectral values for different spectral periods (e.g., for 0.2 s in the S_a 2/50 map of Fig. 14), it is possible to develop damped response spectra (for a specific structural damping ratio, typically

5 %) for a variety of exceedance probabilities (e.g., 2/50, 10/50, etc.) as illustrated in Fig. 14.

In many countries worldwide, there are hazard maps for different PGA or spectral ordinates and for various mean return periods (e.g., details for the USA are given at the theme issue of Earthquake Spectra (2000)). Such European hazard maps have been recently developed (both for PGA and spectral accelerations for various return periods: 101, 475, 975, 2,475 years) in the framework of SHARE project (www.share-eu.org). Figure 15 depicts the UHS for 10 % probability

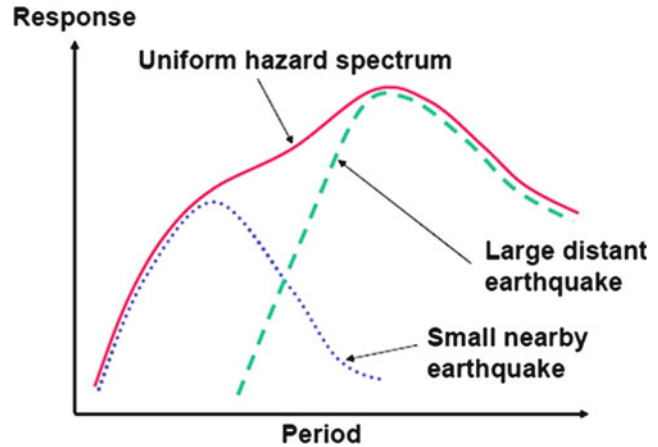
Earthquake Return Period and Its Incorporation into Seismic Actions,

Fig. 12 Example of design spectra for different hazard levels



Earthquake Return Period and Its Incorporation into Seismic Actions,

Fig. 13 Example of a URS for various types of earthquakes (Adopted from FEMA-451B (2003))

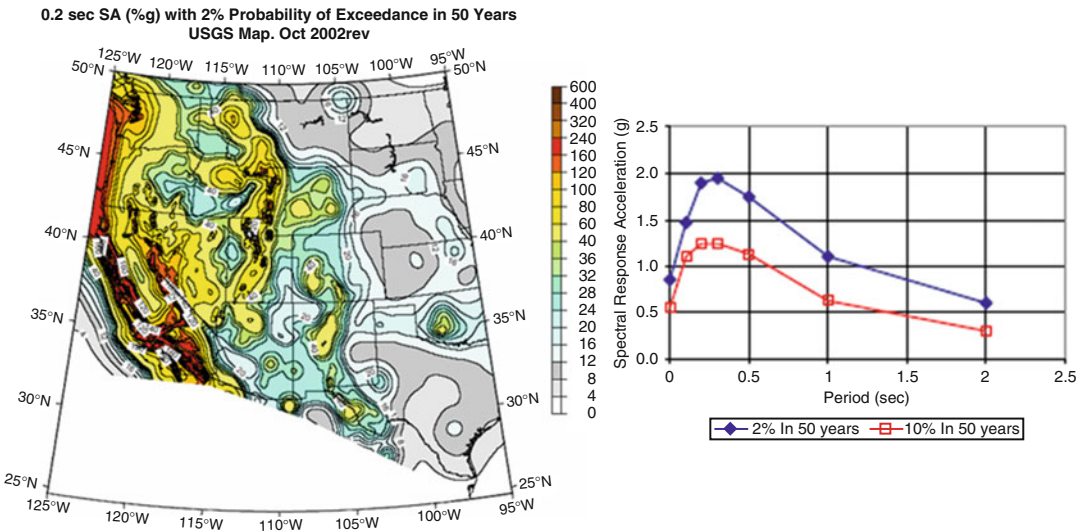


of exceedance in 50 years (right column plots) for Thessaloniki and Istanbul derived from their regional hazard curves (left column plots).

The most frequently used return periods are 475 years (corresponding to 10 % probability of exceedance in 50 years; abbreviated as 10 % in 50 years or 10/50) and 2,475 years (corresponding to 2 % probability of exceedance in 50 years, or 2 % in 50 years, or simply 2/50). The latest US hazard maps can be found at the USGS website <http://earthquake.usgs.gov/hazards/>. These maps are updated every few years. The USGS hazard maps have quantitatively revealed important differences between the ground motion characteristics in different regions of the USA. In particular, they illustrate that the difference between the ground motions for 10 %

in 50 years and 2 % in 50 years in Western US (WUS) is in general less than the difference between these two hazard levels in Central and Eastern USA (CEUS).

In the US seismic design, provisions (IBC 2000; FEMA-368 2001; FEMA-450 2003; etc.) provide, as a minimum, a margin of about 1.5 times the design earthquake ground motion (denoted as the light blue curves in spectra in Fig. 16b, c). More specifically, to prevent collapse for a relatively rare but catastrophic event (associated with a 2 % in 50 year probability) to occur and taking into account the approximate minimum seismic margin of 1.5 against collapse, the IBC provisions generally define design ground motion as $1/1.5 (=2/3)$ times the 2 % in 50-year ground motion. In other words, if



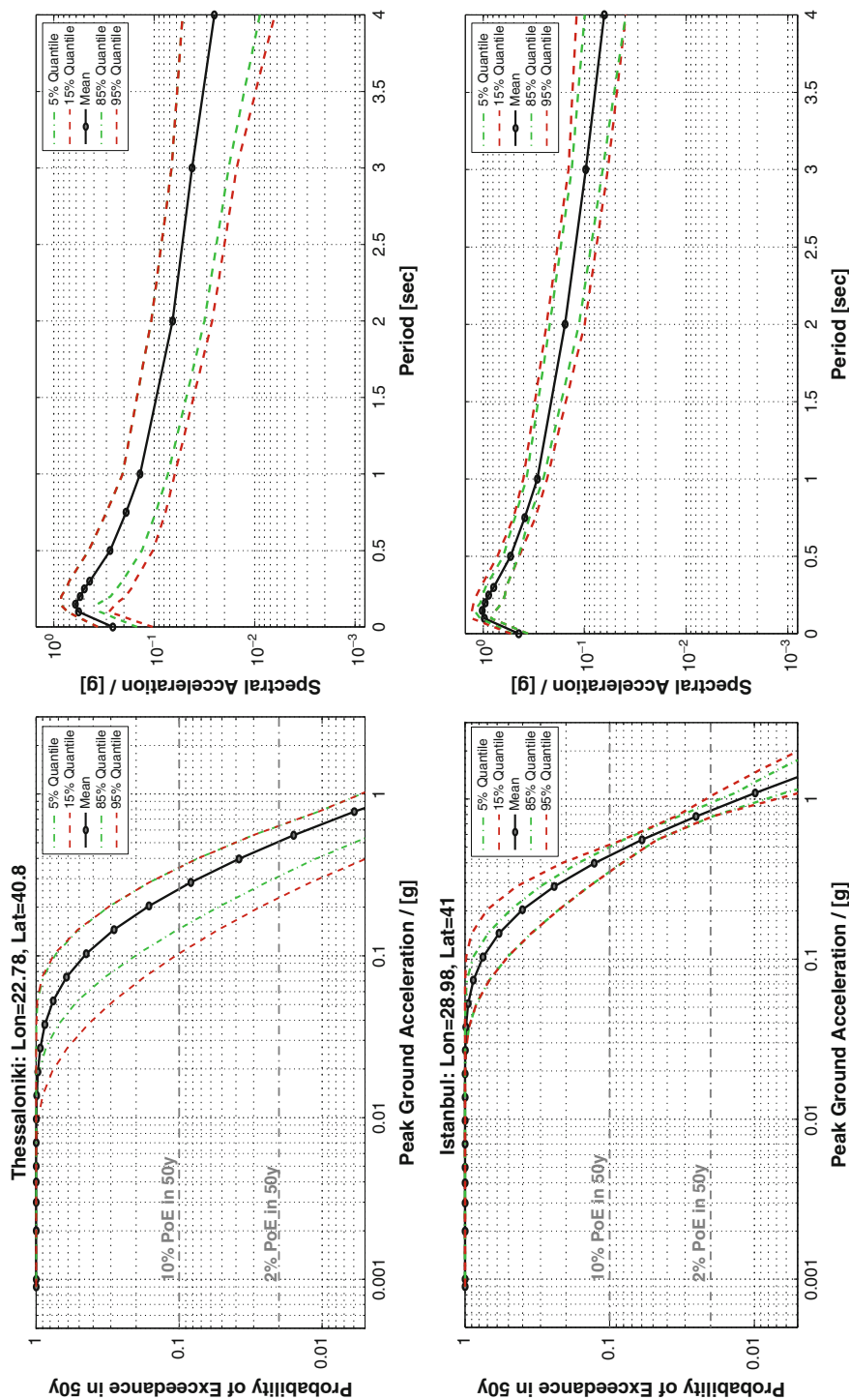
Earthquake Return Period and Its Incorporation into Seismic Actions, Fig. 14 Example of uniform response spectrum generation using probabilistic hazard maps (Adopted from FEMA-451B (2003))

a structure is subjected to a ground motion 1.5 times the standard design level of 10/50, the structure should have a low likelihood of collapse. Certainly, this margin depends on the type of structure, detailing requirements, etc. Nevertheless, this 1.5 factor can be conservative or not depending on the region, e.g., there are important exceptions to this rule, especially near active faults in coastal California.

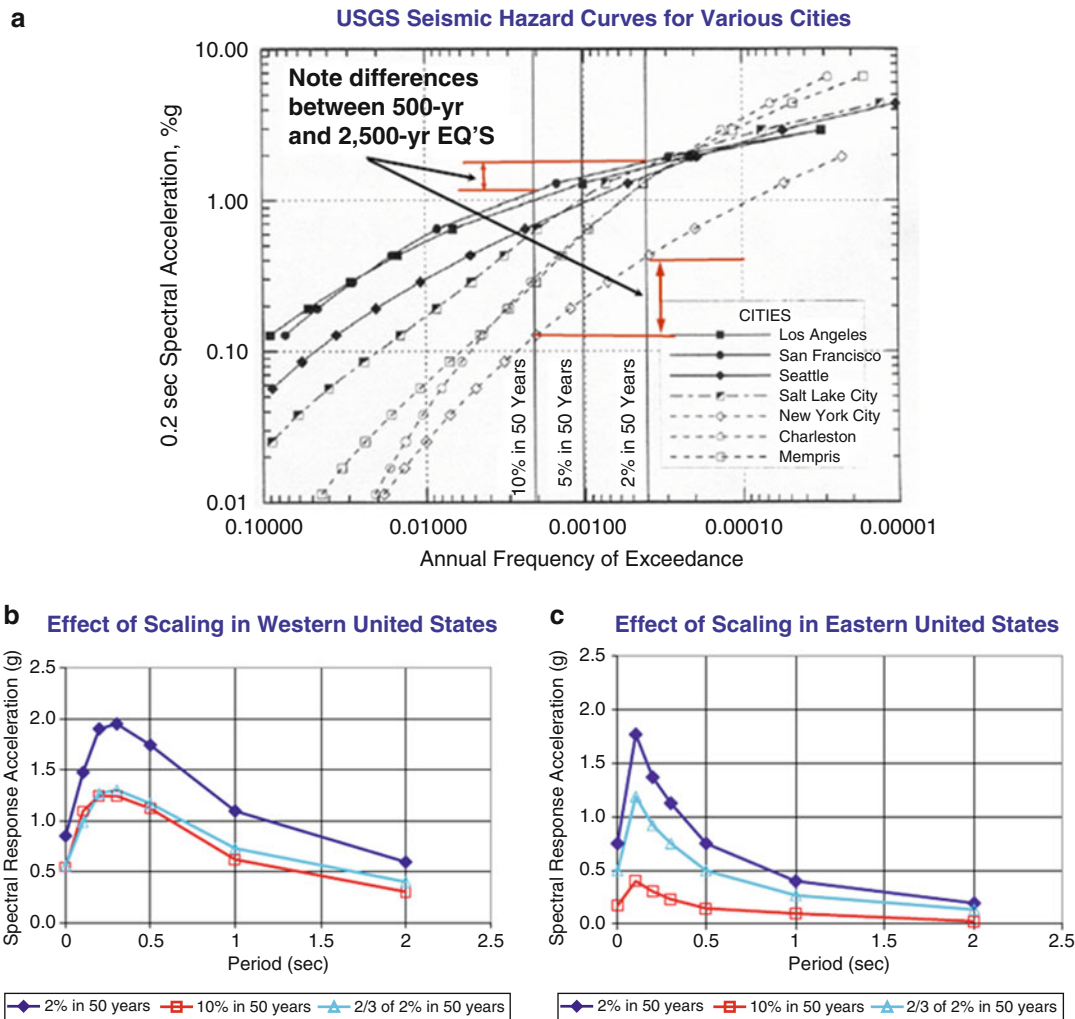
The plot in Fig. 16a presents several seismic hazard curves for major cities in the USA. The main difference to note is the slope of the curves for WUS compared to CEUS. In particular, for WUS, there is a much smaller increase in seismic demand when shifting to the right side of the plot (from a 10/50 to a 2/50 probability) compared to CEUS. Obviously, frequent occasional and rare events are more probable to happen more often in seismic-prone regions compared to low seismicity regions. Nonetheless, it can be easily noticed that for two regions of low and high seismicity, very rare earthquakes (2/50 and less) may have a similar occurrence rate or even to have a higher probability (but very small in absolute numbers) for the low seismicity regions. Due to the shape of the hazard curves, it can be seen that there is a substantial difference in CEUS between the 500- and 2,500-year hazard levels. Thus, many

buildings in CEUS designed for the 500-year event are vulnerable to collapse in a large, very rare earthquake (a 2,500-year period event).

The plot in Fig. 16b illustrates the UHS for San Francisco, in which the 2/50 and 10/50 spectral curves are shown. Note that, in general, two third times the uniform hazard spectrum for 2 % in 50 years is generally comparable to the design spectrum for 10 % in 50 years (corresponding light blue curve coincides with red curve). On the other hand, the plot in Fig. 16c depicts the UHS for Charleston, South Carolina in EUS, where a severe event (with a magnitude of 7.3) occurred in 1886. The differences are significant, e.g., two third times the spectrum for 2 % in 50 years is much higher than that for 10 % in 50 years. Actually, the 10 % in 50-year spectrum provides roughly one fourth of the corresponding values of 2 % in 50 year spectrum. Note also that the shape of the spectrum is different, having its peak at a shorter spectral period. Conclusively, by comparing hazard curves and UHS in Fig. 16, it can be easily concluded that the two third (2,500-year. EQ) $\approx$ 500-year motions in WUS as suggested by US norms, while in contrast, two third (2,500-year. EQ) $\approx$ 1,600-year motions in EUS. Therefore, in EUS the application of IBC design philosophy leads, to a certain extent, to the



Earthquake Return Period and Its Incorporation into Seismic Actions, Fig. 15 UHS for 10 % probability of exceedance in 50 years (right column) and PGA hazard curves (left column) for Thessaloniki and Istanbul (Adopted from www.share-eu.org)



Earthquake Return Period and Its Incorporation into Seismic Actions, Fig. 16 Comparison of hazard curves and UHS in US (Adopted from FEMA-451B (2003))

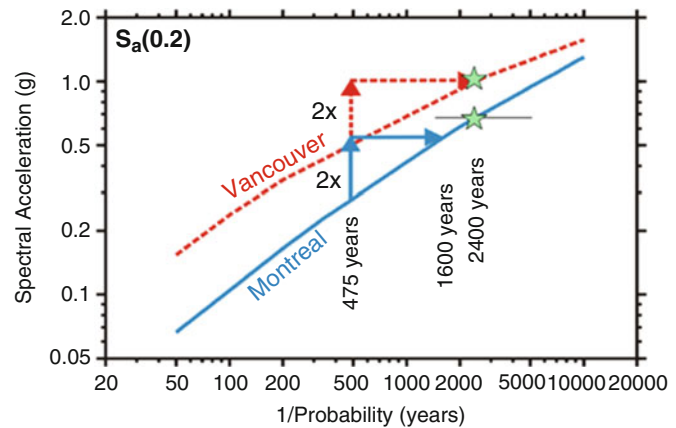
possibility to overcome a more rare and catastrophic earthquake than the design action of 10/50.

Similar observations can be drawn in other seismic-prone countries. For instance, Fig. 17 illustrates the impact of a change in hazard level in Eastern and Western Canada. More specifically, it depicts $S_a(0.2\text{ s})$ hazard curves for Vancouver and Montreal, showing how increasing the 10 % in 50 years hazard by a factor of two ($2\times$) produces different changes in seismic demand for the two cities. In particular, for Vancouver, this would lead to ground motions with

a 1/2,400 per annum exceedance probability, while for Montreal the ground motions would have a 1/1,600 annual exceedance probability. Clearly, the same level of safety would not be achieved. Even if different constants were used for east and west, the geographical variation across Canada would prevent achieving a constant level of safety for medium to severe hazard levels. Nonetheless, as in the USA, hazard curves indicate that the more severe (2 % in 50 years) hazard level leads to a more uniform seismic demand for the whole country (Ventura 2006).

Earthquake Return Period and Its Incorporation into Seismic Actions,

Fig. 17 Expected ground motion increases with increasing return period (i.e., decreasing probability) in Canada (Adopted from Ventura (2006))



Logic Tree Methods in PSHA

Not all types of uncertainties involved in PSHA can be described by probability distributions, e.g., regarding attenuation relationship, seismicity distribution, etc. In addition, experts may disagree on model parameters, such as fault segmentation, maximum magnitude, etc. Aleatory variability is included directly in the PSHA calculations by means of mathematical integration. Epistemic uncertainty, on the other hand, is included in the PSHA by explicitly including alternative hypotheses and models. Modeling epistemic uncertainty due to imperfect knowledge regarding correct model parameters can be performed through sensitivity analysis or logic tree approaches. The consideration of various seismotectonic scenarios via a logic tree includes multiple alternative hypotheses in a single model, taking into account uncertainties, by explicitly including alternative interpretations, models, parameters, etc., that are weighted according to their probability of being correct.

In general, a “logic tree” consists a classical statistical tool for dealing with this type of assessment problems, in which based of the overall knowledge available, the engineer sets up a number of alternative models and associates to each choice a weight factor representative of his subjective (based on judgment or experience) belief on the validity of the choice. In this manner:

- Each model provides an outcome of the assessment, characterized by a probability

which is the product of the probabilities assigned to the branches of the tree.

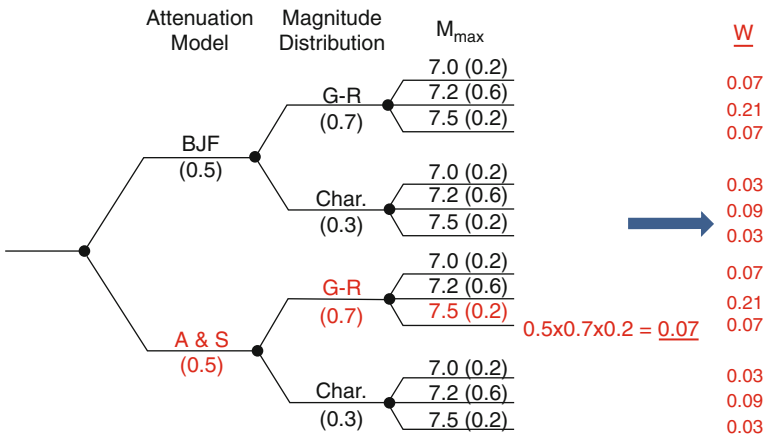
- The ensemble of the outcomes provides an approximate discrete distribution of the system’s characteristics, from which statistics such as mean, dispersion, and confidence interval can be obtained.

The US seismic hazard logic tree models can be found in the relevant USGS report (Petersen et al. 2008). An example of a simple logic tree is shown in Fig. 18 (Kramer 1996) having 12 branches in total, as the outcome of $2 \times 2 \times 3$, number of different attenuation, magnitude distributions and maximum value, respectively. Similar logic trees can be created for structural assessment as well as for a kinematic type of seismic hazard. For example, logic tree assessment of fault parameters can be utilized for various fault rupture scenarios of a certain region in order to define more reliably potential exposure of lifelines crossing it (Angell et al. 2003).

Obviously, computational effort increases when increasing number of nodes and branches. Each alternative hypothesis, indicated by a branch of the logic tree, is given a subjective weight corresponding to its assessed likelihood of being correct. Each branch of the logic tree represents alternative hypotheses for the seismic source, earthquake recurrence frequency, ground motion attenuation, and occurrence probability at a site and their weight factors (shown in parentheses). In each “node” of the logic tree, the sum

Earthquake Return Period and Its Incorporation into Seismic Actions,

Fig. 18 Example of a logic tree model (Adopted from Kramer (1996))



of the weight factors is equal to unity, while the outcome of each branch is the product of the weight factors, as shown in the right column of Fig. 18. Hence, final outcome can be obtained as a weighted average of the values obtained at the terminal branches of the logic tree.

Incorporation of Earthquake Return Period in Seismic Norms: The Case of Eurocode 8

In this section, a brief overview will be given regarding the incorporation of earthquake return period into seismic actions as described in the provisions of European seismic design code Eurocode 8 or EC8 (EN 1998), which is composed of six parts:

- EN 1998-1: General rules, seismic actions and rules for buildings
- EN 1998-2: Bridges
- EN 1998-3: Assessment and retrofitting of buildings
- EN 1998-4: Silos, tanks, and pipelines
- EN 1998-5: Foundations, retaining structures, and geotechnical aspects
- EN 1998-6: Towers, masts, and chimneys

This standard provides alternative procedures, values, and recommendations with notes indicating where national choices may be made. Therefore, in each country the National Standard

implementing EC8 should have a National Annex containing all nationally determined parameters and other supplementary guidelines to be used for the design of buildings and civil engineering works. Additionally, as many European countries have very low seismicity, EC8 includes also the provision that in such cases (defined in every National Annex) it is not obligatory to implement seismic design rules.

In general, as prescribed in Eurocode 8 – Part 1 (EN 1998-1) – a new structure in seismic-prone regions is checked against two limit states: the ultimate no-collapse limit state (corresponding to life safety requirement) and the damage limitation limit state (corresponding to damage limitation requirement). Regarding the first limit state, in order to prevent local failures, the design seismic action in terms of reference probability of exceedance in 50 years is $P_{NCR} = 10 \%$ (or for a reference return period of the seismic action equal to $T_{NCR} = 475$ years). These are the recommended values for $P_{NCR} (=10 \%)$ and $T_{NCR} (=475 \text{ years})$; however, each National Annex can ascribe other values. The value of the probability of exceedance, P_R , in T_L years of a specific level of the seismic action is related to the mean return period, T_R , of this level of the seismic action in accordance with Eq. 8 as follows: $T_R = -T_L / \ln(1 - P_R)$. Thus, for a given T_L , the seismic action may equivalently be specified either via its mean return period, T_R , or its probability of exceedance, P_R in T_L years as previously discussed in section “Seismic Hazard Analysis.”

Earthquake Return Period and Its Incorporation into Seismic Actions, Table 1 Importance classes and factors for buildings and reduction factor for damage limitation limit state according to EC8 (EN 1998-1)

Importance class	Building type	Importance factor (γ_I)	Reduction factor (v)
I	Buildings of minor importance for public safety, e.g., agricultural buildings, small warehouses, sheds, barns, etc.	0.8	0.5
II	Ordinary buildings, not belonging in the other categories	1.0	0.5
III	Buildings whose seismic resistance is of importance in view of the consequences associated with a collapse, e.g., schools, assembly halls, cultural institutions, etc.	1.2	0.4
IV	Buildings whose integrity during earthquakes is of vital importance for civil protection, e.g., hospitals, fire stations, power plants, etc.	1.4	0.4

E

Furthermore, the value of importance factor, γ_I (which for buildings can take a maximum value of 1.4), influences the reference seismic action. In particular, the design value of seismic action is equal to $A_{Ed} = \gamma_I \cdot A_{Ek}$, in which A_{Ek} denotes the characteristic value of the seismic action for the reference return period (i.e., peak ground acceleration, a_{gR}). The importance factor $\gamma_I = 1.0$ is associated with a seismic event having the reference return period. A higher importance factor ($\gamma_I > 1.0$) corresponds to a higher value of earthquake return period and vice versa. Through this modification enhanced performance for critical facilities can be achieved compared to conventional structures.

Each importance factor γ_I is assigned to an importance class of structures (see Table 1). Buildings are classified in four importance classes, depending on the consequences of collapse for human life, on their importance for public safety and civil protection in the immediate post-earthquake period, and on the social and economic consequences of collapse. Wherever feasible, this factor should be derived so as to correspond to a higher or a lower value of the return period of the seismic event (with regard to the reference return period) as appropriate for the design of the specific category of structures. Target reliabilities for the no-collapse (and for the damage limitation) requirement are established by the National Authorities for different types of buildings or infrastructure on the basis of the consequences of failure, i.e., by classifying them into proper importance classes. The different levels of reliability are obtained as

mentioned by multiplying the reference seismic action or, when using linear analysis, the corresponding action effects by the matching importance factor.

It is also noted in EC8 that at most sites, the annual rate of exceedance, $H(a_{gR})$, of the reference peak ground acceleration a_{gR} can vary, as a function of a_{gR} , as follows:

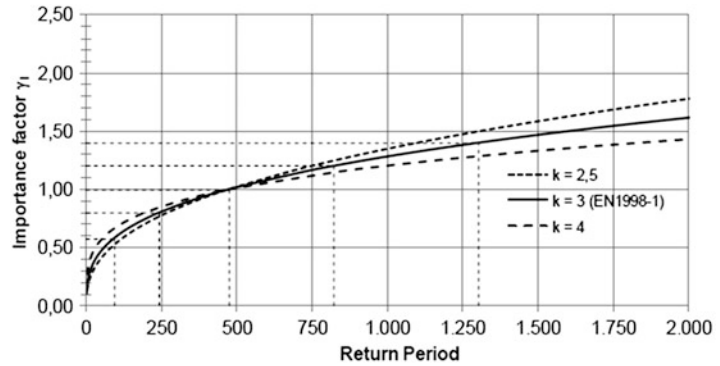
$$H(a_{gR}) \approx k_0 \cdot a_{gR}^{-k} \quad (9)$$

in which the value of the exponent k depends on regional seismicity but being generally of the order of 3, as shown in Fig. 19. In addition, to determine the parameter k_0 at least one a_{gR} (or spectral acceleration) value, corresponding to a specific return period, should be known. This value can be obtained from the seismic hazard map for this region for a given $H(a_{gR})$, e.g., 10 % in 50 years or, alternatively, for a certain return period (e.g., $T_{NCR} = 475$ years). Subsequently, using these $H(a_{gR})$ and a_{gR} values (or analogous spectral acceleration data), k_0 can be easily obtained utilizing Eq. 9. Such expressions have been developed in US within SAC/FEMA project (FEMA-350 2000), while recently more efficient second-order hazard approximations have been proposed to capture more closely the curvature of the hazard function (Vamvatsikos 2012).

Furthermore, if the seismic action is defined in terms of the reference peak ground acceleration, a_{gR} , the value of the importance factor γ_I multiplying the reference seismic action to achieve the same probability of exceedance in T_L years as in

Earthquake Return Period and Its Incorporation into Seismic Actions,

Fig. 19 Influence of importance factor on the return period (Adopted from Carvalho (2011))



the T_{LR} years for which the reference seismic action is defined via the following expression:

$$\gamma_I \approx \left(\frac{T_{LR}}{T_L} \right)^{-1/k} \quad (10)$$

Similarly, the value of the importance factor γ_I that needs to multiply the reference seismic action to achieve a value of the probability of exceeding the seismic action, P_L , in T_L years other than the reference probability of exceedance P_{LR} , over the same T_L years, may be estimated as follows:

$$\gamma_I \approx \left(\frac{P_L}{P_{LR}} \right)^{-1/k} \quad (11)$$

For the damage limitation limit state, the values for the probability of exceedance in 10 years and the return period are $P_{DLR} = 10\%$ and $T_{DLR} = 95$ years, respectively. In the absence of more precise information, the seismic action for the verification of the damage limitation requirement is derived from $A_{Ed} (= \gamma_I \cdot A_{Ek})$ using a reduction factor, v , equal to 0.5 or 0.4 to account for the lower return period for damage limitation performance level. As shown in Table 1, this factor depends on the importance class. It is worth noticing that by applying v , the influence of γ_I is reduced compared to the no-collapse limit state, since the differences among the four classes are not so substantial (i.e., multiplying v and γ_I the values of seismic actions are 0.4, 0.5, 0.48, 0.56 for the four classes, respectively). In addition, the seismic action is

slightly less for importance class III compared to class II. The values of v in each country are provided in its National Annex, while different values of v (that can be associated to a certain time interval similarly to γ_I by an exponential expression, analogous to Eq. 10) may be defined for the various seismic zones of each country, depending on the seismic hazard conditions and on the preferred protection level.

The reduction factor, v , is applied on the design seismic action in accordance with limitation of inter-storey drifts as follows (unless otherwise specified):

- For buildings having nonstructural elements of brittle materials attached to the structure, $d_r \cdot v \leq 0.005 h$;
- For buildings having ductile nonstructural elements, $d_r \cdot v \leq 0.0075 h$;
- For buildings having nonstructural elements fixed in a way so as not to interfere with structural deformations or without nonstructural elements, $d_r \cdot v \leq 0.010 h$;

where d_r is the design inter-storey drift and h denotes storey height.

Additionally to the two basic limit states, EC8 implies that a third limit state can be used to check the safety margin against a global structural collapse (Carvalho 2011). In this case, specific measures should be taken to reduce the uncertainty and ensure a robust behavior of the structure, even under seismic actions more severe than the design seismic action of 10/50. This

performance level is implicitly equivalent to the satisfaction of a third performance requirement: prevention of global collapse under a very rare event (having a much higher return period than $T_{\text{NCR}} = 475$ years), equivalently to the 2/50 event of US norms. This performance level is denoted as near collapse (NC) limit state in EC8 Part 3 (EN 1998-3), which is very close to the actual collapse of the structure as it will be discussed in the sequence.

Regarding the other parts of EC8 and their provisions for seismic return periods, EC8 Part 2 for bridges also follows general rules for reference seismic action return period (there are differences, e.g., in importance classes as well as in damage limitation limit state (Fardis et al. 2012), etc.) and also includes informative Annex A with specific details on the reference seismic action and the corresponding return period as well as on the design seismic action during the construction of bridges. EC8 Part 3 (EN 1998-3) for the assessment and retrofitting of buildings provides specific guidelines for the incorporation of return period into seismic actions, which will be presented in a subsequent section. EC8 Part 4 (for silos, tanks, and pipelines), EC8 Part 5 (for foundations, retaining structures, and geotechnical aspects), and EC8 Part 6 (for towers, masts, and chimneys) do not differentiate with respect to return periods and seismic actions with the basic principles and general rules of EC8 Part 1.

Seismic Zones and Representation of the Seismic Action

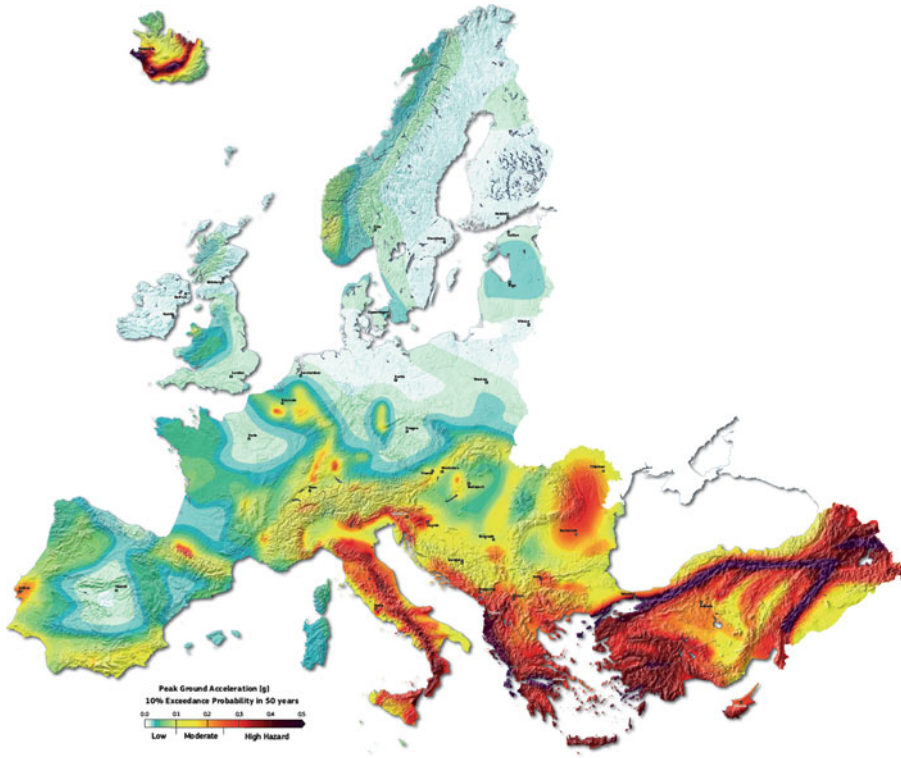
According to EC8, national territories shall be subdivided by the National Authorities into seismic zones, depending on the local seismicity, while the hazard within each zone is assumed to be constant. For most of the applications of EC8, the hazard is described in terms of a single parameter, i.e., the value of the reference peak ground acceleration on type A ground (i.e., rock), a_{gR} . Additional parameters required for specific types of structures are given in the relevant parts of EC8.

The reference peak ground acceleration on type A ground, a_{gR} , may be derived from zonation maps contained in the National Annex of

each country. The reference peak ground acceleration, chosen by the National Authorities for each seismic zone, corresponds to the reference return period T_{NCR} of the seismic action for the no-collapse requirement (or equivalently the reference probability of exceedance in 50 years, P_{NCR}) chosen by the National Authorities. As aforementioned, the values of γ_I may be different for the various seismic zones of the country, depending on the seismic hazard conditions and on public safety considerations. In general, an importance factor γ_I equal to 1,0 is assigned to the reference return period. For other return period values, the design ground acceleration on type A ground is equal to $a_g = \gamma_I \cdot a_{gR}$.

For instance, the National Annex of Greece uses the default values (10 % in 50 years event, etc.) and divides the country into three zones, in which the seismic action, A_{Ek} , in terms of a_{gR} is set equal to: 0.16 g, 0.24 g, 0.36 g, for zones Z1, Z2, and Z3, respectively. An objective of the future updating of EC8 (Carvalho 2011) is to create European zonation maps with spectral values for different hazard levels, similar to provisions in the USA. As aforementioned, such European hazard maps have been recently developed for various return periods, within SHARE project (www.share-eu.org), as the one for 10/50 (475 years return period) mean PGA values shown in Fig. 20.

In EC8 the earthquake motion at a site is represented by an elastic response spectrum in terms of spectral acceleration values (Annex A of EC8 Part 1 additionally represents seismic action in the form of an elastic displacement response spectrum). The shape of the elastic response spectrum for each country may be found in its National Annex, and it is the same for the two levels of seismic performance, i.e., for the no-collapse requirement as well as for the damage limitation requirement. It is also noted that in selecting the appropriate shape of the spectrum, consideration should be given to the magnitude of earthquakes that contribute most to the seismic hazard as defined in PSHA, rather than on conservative upper limits (e.g., the maximum credible earthquake). Regarding magnitude, two types of spectra are recommended,



Earthquake Return Period and Its Incorporation into Seismic Actions, Fig. 20 PGA mean values in Europe for 475 years return period (Adopted from Giardini et al. (2013))

type 1 for magnitude bigger than 5.5 and type 2 for smaller earthquakes. These two types have the same pattern but different spectral amplifications factors that depend also on the stromatography (of the top 30 m surface layer) of the examined region. EC8 provides also recommendations for the vertical acceleration response spectra.

When the earthquakes affecting a site are generated by widely differing sources, the possibility of using more than one spectrum, each having different pattern, should be considered to enable the design seismic action to be adequately represented. In such circumstances, different values of a_g will be required for each type of spectrum and earthquake. Furthermore, for important structures ($\gamma_I > 1.0$), topographic amplification effects should be taken into account, as described in Informative Annex A of EC8 Part 5 (EN 1998-5). Finally, it is also stated that time histories (at least three

accelerograms) representative of the earthquake motion may be used in certain cases, taking into account spatial variability whenever needed.

Return Periods in EC8 Part 2

As aforementioned EC8 Part 2 contains Informative Annex A that defines probabilities related to the reference seismic action as well as guidance for the selection of design seismic action during the construction phase of a bridge. In particular, the reference seismic action can be defined by selecting an acceptably low probability, p , that it can be exceeded within the design life, t_L , of the bridge. Therefore, the return period of the event, T_R , is given by

$$T_R = 1 / \left(1 - (1 - p)^{1/t_L} \right) \quad (12)$$

The reference seismic action (corresponding to $\gamma_I = 1.0$) usually reflects a seismic event with

a reference return period, T_{NCR} , of 475 years. Such an event has a probability of exceedance between 0.10 and 0.19 for a design life ranging between 50 and 100 years, respectively. This level of design action is applicable to the majority of the bridges of average importance.

Regarding the design seismic action for the construction phase, assuming that t_c is the duration of the construction phase of a bridge and p is the acceptable probability of exceedance of the design seismic event during this phase, the return period T_{Rc} is given by Eq. 12, using t_c instead of t_L . For the relatively small values usually associated with t_c ($t_c \leq 5$ years), expression Eq. 12 may be approximated by the following simpler relationship: $T_{Rc} = t_c/p$, while it is recommended that the value of p does not exceed 0.05.

The value of the design ground acceleration, a_{gc} , corresponding to a return period, T_{Rc} , depends on the seismicity of the region. In many cases the following relationship offers an acceptable approximation:

$$\frac{a_{gc}}{a_{gR}} = \left(\frac{T_{Rc}}{T_{NCR}} \right)^k \quad (13)$$

where a_{gR} is the reference peak ground acceleration corresponding to the reference return period T_{NCR} . The value of the exponent k depends on the seismicity of the region, and values in the range of 0.30–0.40 can be used. Finally, it is stated that the robustness of all partial bridge structures should be ensured during the construction phases independently of the design seismic actions.

Limit States and Return Periods in EC8 Part 3

As prescribed in Eurocode 8 Part 3 (EN 1998-3), a retrofitted existing structure is checked against three limit states (LS), namely, near collapse (NC), significant damage (SD), and damage limitation (DL). Note that the definition of the LS of near collapse given in Part 3 of EC 8 is closer to the actual collapse of the building than the one given in Part 1 of EC 8 and corresponds to the maximum exploitation of the deformation capacity of the structural elements. Actually, the LS associated with the “no-collapse” requirement in

Part 1 is roughly equivalent to the one that is defined as LS of significant damage in Part 3.

The National Authorities decide whether all three or less (and which) limit states shall be checked, as defined in National Annex. The appropriate levels of protection are achieved by selecting, for each of the aforementioned LS, a proper return period for the seismic action. The return periods ascribed to the various LS to be checked in a country may be found in its National Annex. The protection is normally considered appropriate for ordinary buildings by selecting the following values for the corresponding return periods:

- LS of near collapse (NC): 2,475 years, corresponding to a probability of exceedance of 2 % in 50 years
- LS of significant damage (SD): 475 years, corresponding to a probability of exceedance of 10 % in 50 years
- LS of damage limitation (DL): 225 years, corresponding to a probability of exceedance of 20 % in 50 years

For instance, the Greek code (KANEPE 2012) for structural interventions on reinforced concrete buildings, after harmonization with Eurocode 8 Part 3, defines two hazard levels. The first has probability of exceedance of 50 % (maximum tolerable) in 50 years and corresponds to an average return period of 70 years, while the second has a probability of exceedance of 10 % in 50 years that corresponds to a 475 years return period. For the 10/50 level, the design seismic action of EC8 Part 1 is taken into account, while for the 50/50 level, 60 % of the design seismic action of EC8 Part 1 is taken into account. It is also stated that a Public Authority shall define the cases where a 50 % probability of exceedance within the reference 50-year period will not be allowed.

Seismic Actions in Performance-Based Design

Severe damages caused by recent earthquakes in seismic-prone areas made the engineering

Earthquake Return Period and Its Incorporation into Seismic Actions, Table 2 Various performance and hazard levels in a typical PBD framework

		Design Performance Level		
		Operational performance level	Life safety performance level	Collapse prevention performance level
Earthquake Hazard Level	50% / 50years (mean return period 72 years)	Performance Objectives for Standard Occupancy Buildings		
	10% / 50years (mean return period 475 years)			
	2% / 50years (mean return period 2475 years)	Performance Objectives for Emergency Response Facilities		

community to question the effectiveness of contemporary seismic design codes. In general, in strong events the economic losses were vast, while in many cases the losses in terms of human lives were also excessive, especially in developing countries. Given that the primary goal of contemporary seismic design is the protection of human life, it is evident that additional performance targets should be considered. Under this perspective, performance-based design (PBD) has been quite recently introduced in earthquake engineering and guidelines (SEAOC 1995; FEMA-273 1997; FEMA-356 2000; FEMA-445 2006). PBD is a holistic framework that includes: local site conditions examination; development of conceptual, preliminary, and final design stages; and construction and maintenance of the building over its life to ensure reliable and predictable seismic performance under varying demand conditions. Consequently, the structure is designed to withstand certain earthquake hazard levels with a predefined behavior for each case.

According to PBD procedures, the structures should be able to resist earthquakes in a quantifiable manner and to present specific levels of possible damages. They are multilevel design approaches, where various levels of

hazard and structural performance are considered. In PBD more accurate analysis procedures are usually employed compared to conventional seismic design to represent more accurately the nonlinear structural response for higher seismic intensities. PBD has the following distinctive characteristics with respect to the prescriptive design codes: (i) it allows the owner and the engineer to choose both the appropriate level of seismic hazard and the corresponding performance level of the structure, and (ii) the structure is designed to meet a series of combinations of hazard levels with corresponding performance levels. As shown in PBD illustrative Table 2, the horizontal axis represents step by step to the right structural performance degradation, while in the vertical axis the size of earthquake increases as we move downward. Although several hazard and performance levels have been proposed, the 3 × 3 PBD framework as the one shown in Table 2 is considered as more effective (e.g., Elnashai and Di Sarno (2008)).

The first step in a performance-based seismic design procedure is the definition of the performance objectives that will be used. One performance objective is defined as the combination of a certain performance level for a specific hazard

level. Typical structural design performance levels are the following:

- (i) *Operational* performance level: the structure retains its strength and stiffness as the overall damage level is very light and none permanent drift is observed. Only minor cracking in facades, partitions, and ceilings are encountered, while all components that are important to normal operation are functional. A green placard will be placed in the building by post-damage inspectors.
- (ii) *Life safety* performance level: the overall damage level is characterized as moderate, i.e., some structural elements and components may be damaged and permanent drifts may be encountered. Nevertheless, structural elements in all stories maintain a considerable residual strength and stiffness and gravity load-bearing elements continue to function. Injuries may occur, however, the overall risk for casualties due to structural damage is low. The damaged structure shouldn't be at imminent collapse risk, but certainly temporary supports and structural repairs will be necessary prior to re-occupancy (i.e., it should be possible to repair the structure although this may not be economical). A yellow placard will be placed in the building by post-damage inspectors.
- (iii) *Collapse prevention* performance level: the structure is heavily damaged and has large permanent drifts. Structural elements have low residual lateral strength and stiffness, although vertical elements are still capable of sustaining gravity loads. Most nonstructural components have collapsed. The structure is near collapse, and most probably it will not be techno-economically practical and safe to repair and reuse it. A red placard will be placed in the building by post-damage inspectors.

The first level aims to maintain structural stiffness, the second is focused on strength, and the last one aims to exploit ductility of the system.

Under this perspective, the structural response is altered from initial elastic to limited and highly nonlinear for the last two performance levels, respectively.

An initial *fully operational* performance level can also be used as the first performance state in a PBD framework, if no damage is allowed to structural and nonstructural components and full functionality of the structure is maintained without any repairs. An additional performance level, the so-called damage limitation, can also be used in PBD as an intermediate state between *operational* and *life safety* performance levels. In this state, the structure may be only slightly damaged having negligible permanent drifts. Structural elements retain their strength and stiffness properties, do not have any significant yielding, and do not need any repair measures. Nonstructural components, such as partitions and infills, may have several cracks; however, such damages could be economically repaired.

The second step in a performance-based seismic design procedure is to determine the earthquake hazard levels. Earthquake-related hazards include mainly ground shaking but also – in certain cases as aforementioned in section “[Seismic Hazard Analysis](#)” – various kinematic distress types (fault rupture, liquefaction, lateral spreading, and landslides). For instance, soil liquefaction hazard in a PBD perspective has been presented by (Kramer and Mayfield 2007) to compare likelihood of liquefaction occurrence for different seismicity level (associated with earthquake return period).

Typically, ground shaking is the earthquake hazard that provisions of the seismic design norms directly address, and it is also the focus of this work. As described in section “[Seismic Hazard Analysis](#),” ground shaking (as well as any other type of) hazard is usually characterized by a hazard curve, which indicates the probability that a given value of a ground motion parameter, for example, peak ground acceleration, will be exceeded over a certain period of time. Hence, the earthquake design levels are expressed in terms of a mean recurrence interval or a probability of exceedance for a specified

Earthquake Return Period and Its Incorporation into Seismic Actions, Table 3 PGA hazard levels (50/50 and 10/50) in Greece utilizing (Papazachos et al. 1993) approach

Seismic zones	Parameters		Occasional (recurrence interval 72 years or 50 % in 50 years)	Rare (recurrence interval 475 years or 10 % in 50 years)
	B ₁	B ₂	PGA (g)	PGA (g)
Zone I	3.28	0.61	0.08	0.14
Zone II	3.64	0.64	0.13	0.215
Zone III	4.01	0.61	0.17	0.29
Zone IV	4.64	0.55	0.29	0.46

number of years that covers the design life of the structure, e.g., typically 50 years for buildings. The most commonly used ground shaking hazard levels are the following:

- (iv) *Occasional* hazard level: with probability of exceedance 50 % in 50 years with recurrence interval equal to 72 years and annual probability of exceedance approximately equal to 1.4 %. This level is the so-called serviceability earthquake (SE), which is approximately 0.5 times the design earthquake (DE) (ATC-40 1996).
- (v) *Rare* hazard level: with probability of exceedance 10 % in 50 years with recurrence interval equal to 475 years and annual probability of exceedance approximately equal to 0.2 %. This level is the so-called design earthquake (DE) or design basis earthquake (DBE), and (as aforementioned in section “[Seismic Hazard Analysis](#)”) for certain norms, it is approximately the two third of the more severe 2/50 level.
- (vi) *Maximum considered event* hazard level: with probability of exceedance 2 % in 50 years with recurrence interval equal to 2,475 years and annual probability of exceedance approximately equal to 0.04 %. This level is the so-called maximum considered earthquake (MCE).

According to the adopted PBD formulation (e.g., SEAOC (1999)), other hazard levels can also be used. For instance, a *frequent event* hazard level having a very low return period (i.e., 43 years or probability of exceedance 50 % in 30 years, or 2.3 % annual probability of

exceedance) or a *very rare event* hazard level having an intermediate return period (i.e., 970 years, or probability of exceedance 10 % in 100 years, or 0.1 % annual probability of exceedance) between the two higher ones. In addition, it is worth noticing the distinction among maximum earthquake (ME) level and maximum considered earthquake (MCE) (ATC-40 1996). The ME is deterministically defined as the maximum level of ground shaking that may ever be expected at the construction site based on regional seismotectonic data. If ME is expressed in probabilistic terms, as the aforementioned hazard levels, then it has a return period of about 1,000 years; thus, it is equivalent to *very rare event* hazard level.

The above hazard levels can be associated to major seismic demand parameters, such as the peak ground acceleration (PGA). For instance, in the study by Papazachos et al. (1993), Greece was divided into four zones, and based on this work, Table 3 depicts the PGA levels for two return periods, T_R , utilizing the following log-linear formula:

$$SI = B_1 + B_2 \cdot \log T_R \quad (14)$$

in which SI denotes a characteristic measure of the seismic intensity (macroseismic intensity or the logarithm of strong ground motion parameters, such as PGA or PGV), and parameters B_1 and B_2 vary for each SI.

Such relationships can be derived based on a Cornell-type attenuation formula of the form (Papazachos et al. 1993)

$$SI = C_1 + C_2 \cdot M - C_3 \cdot \ln(D + C_4) \quad (15)$$

in which M is the earthquake magnitude, D is the epicentral or hypocentral distance, and C_1 , C_2 , C_3 , and C_4 are regional constants. Based on such formulations, updated seismic hazard maps (10/50 and 10/100) for Greece have been constructed (Papaioannou et al. 2008). Alternative relationships for PGA and T_R for design earthquakes in Greece have been recently proposed, utilizing a uniform reliability index approach (Makarios 2013).

In addition, spectral values (accelerations or displacements) taken from a design spectrum obtained for a specific return period (e.g., for reference $T_{NCR} = 475$ years and $S_{10/50}$ spectral values) can be used to obtain approximate spectral values for other return periods utilizing an expression similar to Eq. 13 of EC8 Part 2 for bridges, based on SEAOC (1999) as follows (Kanellopoulos 2007):

$$\frac{S_x}{S_{10/50}} = \left(\frac{T_x}{T_{NCR} = 475} \right)^n \quad (16)$$

where n depends on regional seismotectonic characteristics and depends on the hazard level. For example, for California and lower hazard levels, $n = 0.44$ and for severe events $n = 0.29$; thus, if $Sa_{10/50} = Sa(T = 0) = PGA_{10/50} = 0.30$ g, then for $T_x = 72$ years utilizing Eq. 16 it can be easily derived that, for $n = 0.44$, $Sa_{50/50} = Sa(T = 0) = PGA_{50/50} = 0.13$ g.

Return period can also be used for the assessment of structural capacity and vulnerability of a structure in order to design proper mitigation measures. In other words, it can be used to quantify the safety levels and structural deficiencies in a more straightforward manner and to assess the improvement of structural performance after certain retrofitting interventions. For this purpose, Raffaele and Fiore (2013) presented a simple methodology (based on the reverse application of the well-known N2 method) that allows the assessment of the return period when the elastic period and the capacity of the examined structure for a given limit state are available. Subsequently, this so-called “capacitive” return period can be used to determine the seismic capacity and the reference life for any limit state considered.

Structural Performance Assessment in PBD

The main objective of multistep PBD procedures is to define a relation between the seismic intensity level and the corresponding maximum response quantity of the structural system. The intensity level and the structural response are described via suitable intensity measures (IM) and engineering demand parameters (EDP), respectively. Since EDPs actually represent the level of structural damage, they are usually utilized in the form of various damage indices (DI). Very frequently in seismic assessment of buildings, the maximum inter-storey drift and maximum floor acceleration are chosen as they can be directly associated with structural performance levels.

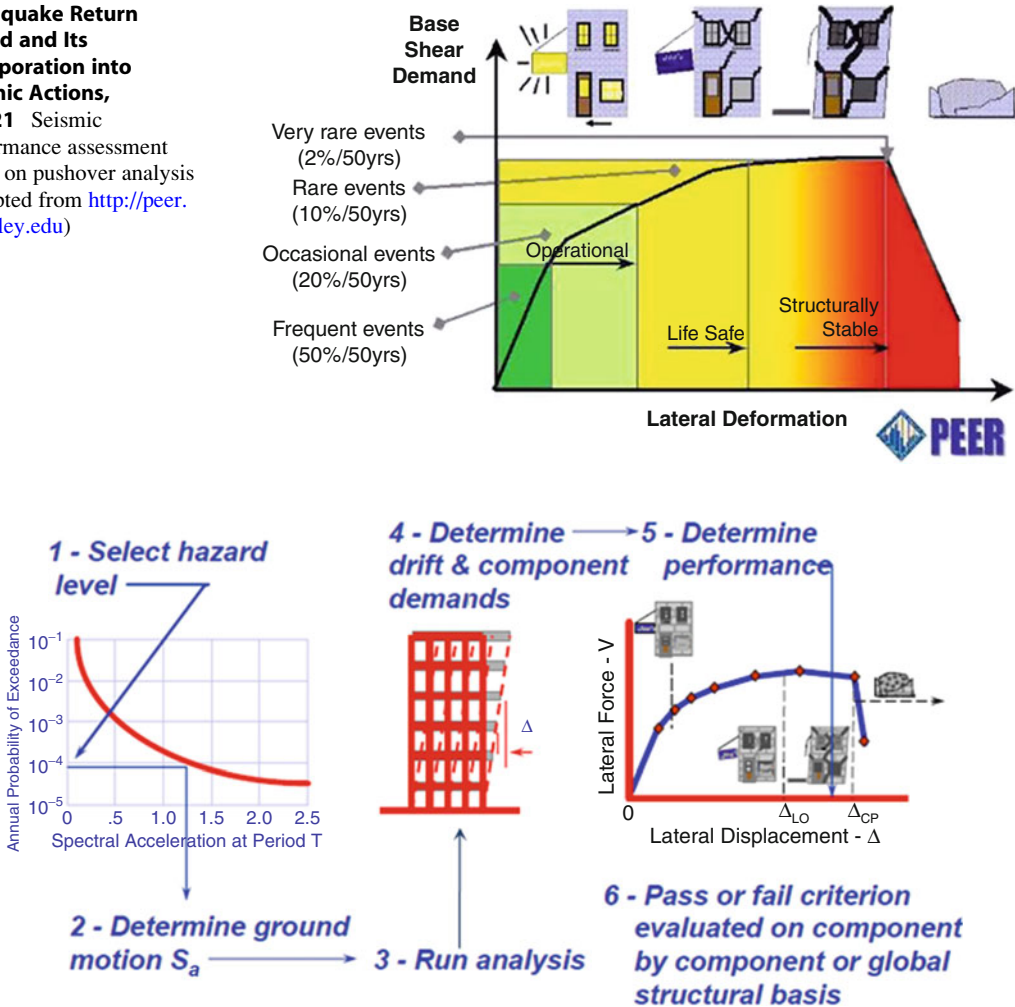
Elnashai and Di Sarno (2008) have presented a very well-established direct association of the structural performance levels with earthquake hazard level (in terms of magnitude and return period), structural characteristics as well as engineering and socioeconomic limit state characterizations (i.e., drifts, damage levels, etc.). In other studies, earthquake return periods have been directly used in presenting and comparing (in terms of EDPs) the performance of different structural configurations, e.g., the height-wise distribution of inter-storey drifts and peak floor accelerations for fixed base vs. isolated buildings (Chimamphant and Kasai 2013).

Despite the continuous advances in computational methods as well as in software and hardware, nonlinear time-history analyses are not widely used in engineering practice, and seismic norms suggest to employ them in special cases. In contrast, a nonlinear static procedure, the so-called “pushover” method, is very frequently used for seismic performance assessment of structures. A schematic representation of a typical pushover curve and its utilization in PBD framework based on the hazard level is shown in Fig. 21. The major steps of the whole evaluation process for each hazard level are depicted in Fig. 22.

It is worth mentioning that in recent years PBD has also been integrated into optimization approaches, in some of which (e.g., Fragiadakis and Papadrakakis (2008)) the aforementioned

Earthquake Return Period and Its Incorporation into Seismic Actions,

Fig. 21 Seismic performance assessment based on pushover analysis (Adopted from <http://peer.berkeley.edu>)



Earthquake Return Period and Its Incorporation into Seismic Actions, Fig. 22 Basic steps in performance evaluation (Adopted from FEMA-451B (2003))

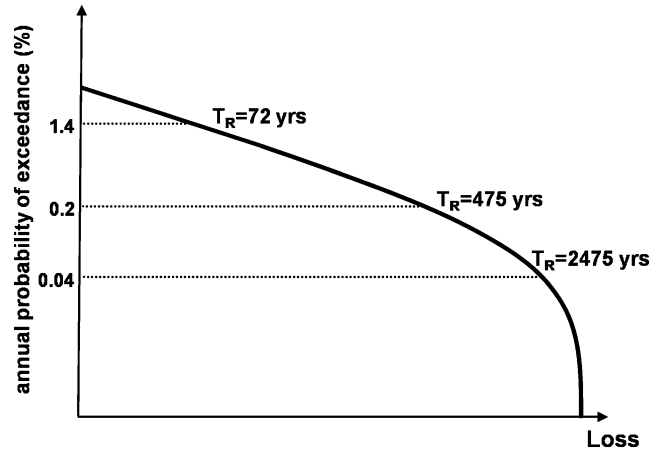
recurrence periods of PBD hazard levels have been used as probabilistic constraints in the optimization process. Life cycle cost analysis (LCCA) is also considered nowadays in conjunction with PBD and optimization to achieve best possible solutions. LCC represents the total cost, including that related to the expected damages caused by earthquake events that may occur during the design life of the structure. It has been proven that the objective to satisfy all checks imposed by contemporary norms aiming to minimize initial construction cost is not always efficient in terms of overall long-term

cost and safety. In a holistic probabilistic LCCA and PBD framework, all aspects are considered, and the outcome is given in terms of fragility curves and loss hazard curves. As shown in Fig. 23, the latter are similar to seismic hazard curves (e.g., Fig. 4), as the losses (or repair cost), expressed in terms of monetary units, and are also associated to earthquake return periods or probabilities of exceedance.

To accomplish this goal, it is necessary to obtain curves illustrating the so-called mean annual frequencies (MAF) of exceedance for

Earthquake Return Period and Its Incorporation into Seismic Actions,

Fig. 23 Characteristic repair loss curve for various hazard levels

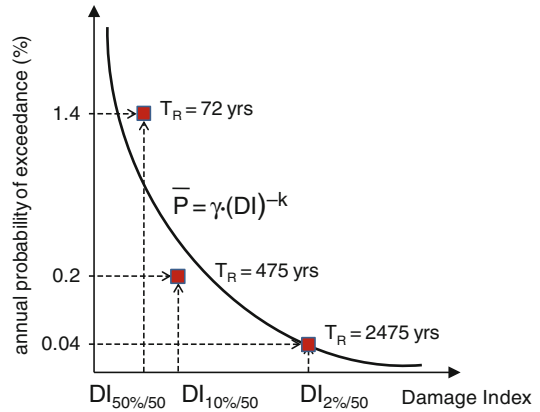


different performance levels. For this purpose, as aforementioned in seismic hazard analysis section, based on Poisson's stochastic model to describe earthquake occurrence, the probabilities for damage indices (DI) (e.g., drifts, maximum accelerations, etc.) are calculated as follows:

$$P(DI > DI_i) = -\frac{1}{\lambda \cdot t} \cdot \ln(1 - P_t(DI > DI_i)) \quad (17)$$

in which $P_t(DI > DI_i)$ is the exceedance probability over a period $[0, t]$, λ is the annual earthquake occurrence rate, and t is the functioning period (typically 50 years) of a new structure or the remaining life of an existing one. In order to calculate $P(DI > DI_i)$, it is necessary to evaluate the t -year exceedance probability $P_t(DI > DI_i)$. For $t = 1$ year, the annual exceedance probability $\bar{P}_i(DI > DI_i)$ is obtained for the three most commonly used PBD earthquake hazard levels (as aforementioned, having exceedance probabilities: 2 %, 10 %, and 50 % in 50 years, or in terms of annual exceedance probabilities approximately equal to 0.04 %, 0.2 %, and 1.4 %, respectively), and the maximum values of the chosen DIs are calculated.

Consequently, the maximum DIs calculated at each hazard level have an annual exceedance probability equal to that of the hazard level. In this manner, one common curve can be constructed for all hazard levels using pairs of maximum DIs and their respective annual exceedance



Earthquake Return Period and Its Incorporation into Seismic Actions, Fig. 24 Best fitting of annual probabilities and damage indices for various hazard levels

probabilities, based on an exponential formula of the form (Lagaros and Mitropoulou 2013):

$$\bar{P}_i(DI > DI_i) = \gamma \cdot (DI_i)^{-k} \quad (18)$$

where k represents the hazard decay and γ is a scale factor. Mathematically, k is the slope of the hazard curve and indicates the way that ground motion gets more intense as the probability of exceedance decreases. Parameters γ and k are obtained by a best fitted curve (as the one shown in Fig. 24), which is the outcome of the available $\bar{P}_i$ and DI_i pairs for each of the examined DIs. Subsequently, for each i th limit state, the annual exceedance probability of DI_i , $\bar{P}_i(DI > DI_i)$ can be easily derived from this curve.

Closure

According to contemporary seismic norms and engineering practice worldwide, seismic design is performed for the functioning period of a structure, depending on its importance and use, for different limit states, very frequently defined in terms of the return periods of seismic actions. The most commonly used seismic intensity measures include: earthquake magnitude and energy as well as number of earthquakes above a certain magnitude per year(s) and the probability of exceedance, annually or in a specified time interval, or the return period of such event. Based on these data, when designing a new structure or when assessing an existing one, seismic hazard and seismic risk can be evaluated. As aforementioned, seismic hazard defines the expected occurrence of future seismic events, while seismic risk describes the expected consequences. There are two methods for the assessment of seismic hazard: deterministic (DSHA), magnitude “ x ” earthquake due to “ y ” fault, and probabilistic (PSHA), “ x ” % probability of exceedance in “ y ” years of the design ground shaking level.

Since hazard analysis should incorporate uncertainties related to seismic ground motions parameter and their modeling, PSHA is the most popular choice. The uncertainties considered in PSHA are the maximum magnitude, earthquake recurrence rate, seismic source characteristics, attenuation relationships, effects of local site conditions, etc. PSHA is used for the calculation of the mean hazard curve (MHC) utilizing the Poisson model to predict the probability of earthquakes based on the average rate of earthquakes of a given size that occur in a region. Such curves provide the annual probability or recurrence period that a seismic ground motion exceeds a particular level at the specific region.

Structural capacity and demand depend on the excitation characteristics and structural properties. Typically, engineers consider the performance of a structural system for a given period of time. The demand, such as structure’s global drift, is clearly a quantity that fluctuates in time and is highly uncertain depending on the seismic

excitation during a certain time period. The capacity characterizes the system and also depends on the imposed seismic demand. It is a common practice to use the maximum response or seismic distress over a given time period (annual or per 50 years) as the demand variable. For example, a 10 % probability of exceedance in 50 years is used worldwide as the threshold for selecting the design earthquake parameters.

Seismic intensity is typically expressed by means of spectral acceleration or peak ground acceleration. According to the contemporary PBD perspective, depending on the seismic intensity or ground motion return period, different (and often not negligible) levels of structural damage and, consequently, repairing costs shall be expected and accepted as a result of the nonlinear structural response. In this context, the specification of each performance objective is given for each pair of hazard level (x % in 50 years) and an acceptable performance level (i.e., maximum acceptable damage if that event occurs) depending on the importance of the structure.

Summary

Seismic hazard (related to ground shaking, fault movement, etc.), is a natural threat that can produce adverse effects to buildings and infrastructure. On the other hand, seismic risk defines the probability of the consequences of an earthquake, such as structural damage, will be equal or exceed a target value for a given period of time. In other words, risk is defined as the product of the hazard and the exposure of structures to the hazard, while the latter is directly related to structural vulnerability. Very frequently magnitude and other seismic intensity parameters are associated with earthquake return period which depends on the seismotectonic conditions of the examined region. Ground motion intensity at a site depends on proximity of sources and other earthquake source effects including travel path effects and local site conditions. Site conditions (stromatography, subsurface bedrock morphology (basin effects), surface topography) can

have a significant effect on ground shaking and structural response. Relative distance of the site to the fault is also important and is usually incorporated in the form of ground motion attenuation relationships that consider source effects, travel path effects, and local site effects (at least basic ground condition, e.g., rock or soil) and also take into account various uncertainties. There are two types of uncertainties related to seismic hazard: (i) epistemic uncertainties due to incomplete knowledge about natural phenomena, which affects our ability to model them, and (ii) aleatory uncertainties that include any inherent randomness which arises from physical variability of the natural process. There are two basic methods for performing seismic hazard assessment: deterministic seismic hazard assessment (DSHA) and probabilistic seismic hazard assessment (PSHA). The latter is the most commonly used in contemporary norms and engineering practice, in which seismic hazard is expressed in terms of mean recurrence of events, the so-called return period, or the probability of exceeding a certain seismic intensity level over a given time period. The return period (e.g., typically 475 years for design basis earthquake) defines the average time interval to have events (not only earthquakes but floods, hurricanes, etc.) of a certain or greater severity. Alternatively, the probability of exceedance (e.g., typically 10 % in 50 years for design basis earthquake) is a statistical representation of the likelihood that earthquake effects of a certain level will be experienced at the site of interest within the prescribed number of years. With proper statistical models these two parameters are directly related and are equivalently used. In this work, a brief overview of these methodologies will be given in order to illustrate the incorporation of earthquake return period into seismic actions on structures due to ground shaking. More specifically, the focus will be given on contemporary seismic norms, such as the European seismic design code (Eurocode 8), as well as performance-based design (PBD) framework, in which the probabilities of exceeding performance criteria is directly related to probabilities of various seismic events occurrence in terms of their return period.

Cross-References

- [Earthquake Recurrence](#)
- [Earthquake Response Spectra and Design Spectra](#)
- [EEE Catalogue: A Global Database of Earthquake Environmental Effects](#)
- [Engineering Characterization of Earthquake Ground Motions](#)
- [European Structural Design Codes: Seismic Actions](#)
- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- [Physics-Based Ground-Motion Simulation](#)
- [Probabilistic Seismic Hazard Models](#)
- [Random Process as Earthquake Motions](#)
- [Response Spectrum Analysis of Structures Subjected to Seismic Actions](#)
- [Review and Implications of Inputs for Seismic Hazard Analysis](#)
- [Site Response for Seismic Hazard Assessment](#)
- [Spatial Variability of Ground Motion: Seismic Analysis](#)
- [Stochastic Ground Motion Simulation](#)
- [Structural Design Codes of Australia and New Zealand: Seismic Actions](#)

References

- Angell MM, Hanson K, Swan FH, Youngs R, Abramson H (2003) Probabilistic fault displacement hazard assessment for flowlines and export pipelines, Mad Dog and Atlantis Field Developments, Deepwater Gulf of Mexico. Offshore Technology Conference (OTC-2003), Houston, 5–8 May 2003, paper OTC 15402
- ATC-40 (1996) Seismic evaluation and retrofit of concrete buildings. Prepared by the Applied Technology Council (Report No. ATC-40), Redwood City, CA, for the California Seismic Safety Commission (Report No. SSC 96-01), Sacramento, CA
- Bommer JJ, Scott SG, Sarma SK (2000) Hazard-consistent earthquake scenarios. *Soil Dyn Earthq Eng* 19:219–231
- Carvalho EC (2011) Overview of Eurocode8, Eurocode 8. Seismic design of buildings training workshop, 10–11 Feb 2011, Lisbon
- Chimamphant S, Kasai K (2013) Seismic performance investigation and comparison of base isolated and fixed base structures. In: 13th world conference on

- seismic isolation, energy dissipation and active vibration control of structures, Sendai, 24–27 Sept 2013
- Cornell CA (1968) Engineering seismic risk analysis. *Bull Am Seismol Soc* 58(5):1583–1606
- Earthquake Spectra (2000) Seismic design provisions and guidelines. *Earthq Spectra Theme Issue* 16(1): 1–310
- Elnashai A, Di Sarno L (2008) *Fundamentals of earthquake engineering*. Wiley, New York
- EN 1998 (2004) Eurocode 8 (EC8): design provisions of structures for earthquake resistance. European Committee for Standardization (CEN), Brussels
- Fardis MN, Koliass B, Panagiotakos T, Katsaras C, Psychogios T (2012) Guide for bridge design with emphasis on seismic aspects. Report No SEE 2012–01, Report Series in Structural and Earthquake Engineering, Department of Civil Engineering, University of Patras, Patras
- FEMA-273 (1997) NEHRP guidelines for the seismic rehabilitation of buildings. Federal Emergency Management Agency, Washington, DC
- FEMA-350 (2000) Recommended seismic design criteria for new steel moment-frame buildings. SAC Joint Venture, Federal Emergency Management Agency, Washington, DC
- FEMA-356 (2000) Prestandard and commentary for the seismic rehabilitation of buildings. Federal Emergency Management Agency, Washington, DC
- FEMA-368 (2001) NEHRP recommended provisions for seismic regulations for new buildings and other structures, part 1, provisions. Federal Emergency Management Agency, Washington, DC
- FEMA-445 (2006) Next-generation performance-based seismic design guidelines, program plan for new and existing buildings. Federal Emergency Management Agency, Washington, DC
- FEMA-450 (2003) NEHRP Recommended provisions for seismic regulations for new buildings and other structures. Federal Emergency Management Agency, Washington, DC
- FEMA-451B (2003) NEHRP recommended provisions: instructional and Training materials. Building Seismic Safety Council (BSSC), Washington, DC
- FEMA-750 (2009) NEHRP Recommended seismic provisions for new buildings and other structures. Federal Emergency Management Agency, Washington, DC
- Fragiadakis M, Papadrakakis M (2008) Performance-based optimum seismic design of reinforced concrete structures. *Earthq Eng Struct Dyn* 37(6): 825–844
- Giardini D, Woessner J, Danciu L, Crowley H, Cotton F, Grünthal G, Pinho R, Valensise G, SHARE consortium (2013) SHARE European seismic hazard map for peak ground acceleration, 10% exceedance probabilities in 50 years. SHARE European seismic harmonization project. Available at www.share-eu.org, Aug 2013
- IBC (2000) International building code. International Code Council, Falls Church
- Kanellopoulos A (2007) Seismic design and retrofitting of reinforced concrete buildings, Cubus Hellas, Athens (in Greek)
- KANEPE (2012) Code of structural interventions on reinforced concrete buildings, English version after harmonization with Eurocode 8 Part 3, Earthquake Planning and Protection Organization, Athens. Available at <http://ecpfe.oasp.gr/>
- Katsanos EI, Sextos AG, Manolis GD (2010) Selection of earthquake ground motion records: a state-of-the-art review from a structural engineering perspective. *Soil Dyn Earthq Eng* 30:157–169
- Kramer SL (1996) *Geotechnical earthquake engineering*. Prentice Hall, Upper Saddle River
- Kramer S, Mayfield RT (2007) Return period of soil liquefaction. *ASCE J Geotech Geoenviron Eng* 133(7):802–813
- Krinitzsky EL (1995) Deterministic versus probabilistic seismic hazard analysis for critical structures. *Eng Geol* 40:1–7
- Lagaros ND, Mitropoulou CC (2013) The effect of uncertainties in seismic loss estimation of steel and reinforced concrete composite buildings. *Struct Infrastruct Eng* 9(6):546–566
- Makarios TK (2013) Peak ground acceleration functions of mean return period for known reliability index of Hellenic design earthquakes. *J Earthq Eng* 17(1):98–109
- Papaoiannou C, Voulgaris N, Karakaisis G, Koutrakis S, Latousakis J, Makropoulos K, Papazachos B, Sokos E, Stavrakakis G, Tselentis G (2008) The utilization of new seismological data in the compilation of the new seismic hazard map of Greece. In: Third Greek conference on earthquake engineering & engineering seismology, Athens, 5–7 Nov 2013, paper 2025 (in Greek)
- Papazachos BC, Papaoiannou CA, Theodulidis NP (1993) Regionalization of seismic hazard in Greece based on seismic sources. *Nat Hazards* 8(1):1–18
- Petersen MD, Frankel AD, Harmsen SC, Mueller CS, Haller KM, Wheeler RL, Wesson RL, Zeng Y, Boyd OS, Perkins DM, Luco N, Field EH, Wills CJ, Rukstales KS (2008) Documentation for the 2008 update of the United States national seismic hazard maps. US Geological Survey, Open-file report 2008–1128, p 61
- Raffaele D, Fiore A (2013) A simplified algorithm for evaluating the seismic return period of structural capacity. In: 4th conference on computational methods in structural dynamics and earthquake engineering, Kos, 12–14 June 2013
- Saygili G (2008) A probabilistic approach for evaluating earthquake-induced landslides. PhD Thesis, The University of Texas at Austin, TX
- SEAOC (1995) Performance based seismic engineering of buildings, Vision 2000 Committee, Structural Engineers Association of California, Sacramento, CA
- SEAOC (1999) Recommended lateral force requirements and commentary, Structural Engineers Association of California, Sacramento, CA

- Vamvatsikos D (2012) Accurate application and second-order improvement of the SAC/FEMA probabilistic formats for seismic performance assessment. *ASCE J Struct Eng*. doi:10.1061/(ASCE)ST.1943-541X.0000774
- Ventura CE (2006) Fundamental concepts in seismic engineering. Short course for CSCE Calgary 2006 annual conference, Canada

Earthquake Risk Mitigation of Lifelines and Critical Facilities

Kyriazis Pitilakis and Sotiris Argyroudis
Department of Civil Engineering, Aristotle
University, Thessaloniki, Greece

Synonyms

Infrastructures; Readiness; Response; Restoration; Retrofitting; Risk reduction; Seismic risk assessment; Transportation networks; Utilities

Introduction

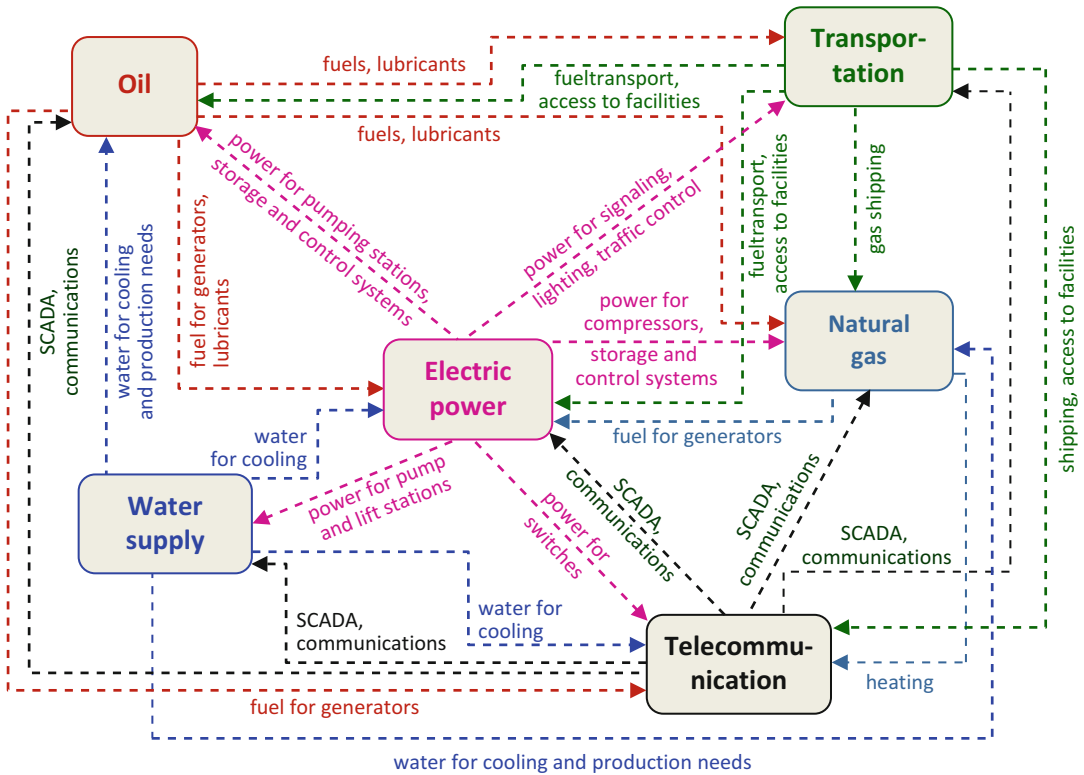
Lifelines and critical facilities refer to the complex system and network assets of connected components usually interacting with other components and systems, which are performing vital functions that are essential to sustain the life and the growth of a community, such as producing, transporting, and distributing goods or services. Their global value for the society and economy is permanently increased in our modern, technologically advanced, highly demanding, and fragile world. In case of strong earthquake motions, their physical damages and the consequent disruption of the services they provide may contribute seriously to the global economic losses. At the same time, their restoration cost may be very high, reaching in several cases to 10 %, 15 %, or even more in some extreme cases (e.g., the water system in Christchurch, New Zealand) of the initial construction cost of the whole system they belong. Therefore, in terms of resilience, it is important to recognize and quantify the seismic

risk and global losses associated to damages of lifelines and critical facilities and establish efficient seismic risk mitigation strategies. Mitigation actions include, among others, enhancement of emergency preparedness, strengthening of existing structures, upgrade of seismic design guidelines, and standards and improvement of the recovery planning.

The issue of resilience of lifelines and critical facilities is of utmost priority for the preparedness and mitigation of seismic risk; a good evaluation of the resilience of the lifelines for different seismic scenarios and the implementation of pro-seismic actions to improve it are of major importance. In other words, lifeline companies, public administration, and community need to be “prepared” and less “vulnerable,” in order to achieve a high “resilience.” In the last years, as the idea of the necessity of building disaster resilient communities gains acceptance, new methods have been proposed to quantify resilience beyond estimating losses. Because of the vastness of the definition, resilience necessarily has to take into account its entire complex and multiple dimensions, which includes technical, organizational, social, and economic facets (Cimellaro et al. 2010).

Lifelines may be distinguished in two major categories: (i) utility systems including potable water, wastewater, natural gas, oil, electric power, and communication and (ii) transportation systems comprising roadways, railways, airport, and port facilities. Sometimes the terms infrastructures or critical facilities are used instead, at least for some of them. Lifeline systems present three distinctive features: (a) spatial variability and exposure to different geological and geotechnical conditions and hazards (e.g., fault crossing, landslides, liquefaction); (b) wide variety of component typologies and materials used; and (c) specific functionality requirements, which makes them highly hierarchical networks. For practical reasons, they are generally distinguished to links (e.g., pipelines, roads) and nodes (e.g., tanks, power substations, bridges).

An important issue in understanding and defining the risk mitigation of lifelines and critical facilities is the extensive interdependencies



Earthquake Risk Mitigation of Lifelines and Critical Facilities, Fig. 1 Example of infrastructure interdependencies (Modified, Rinaldi et al. 2001)

within utility networks and infrastructures as well as infrastructure business, which are schematically shown in Fig. 1. Infrastructure outages due to seismic damages can cascade through the economy and considerably increase the overall losses. Moreover, interactions between systems can severely impact the emergency response, especially when the disruptions concern critical facilities, such as hospitals, command centers, etc. Therefore, the key interactions should be recognized and considered in the risk mitigation planning.

Several devastating earthquakes occurred during the last century, causing extensive losses in terms of fatalities, injuries, damages, and disruptions. The experiences from past earthquakes show that the short- and long-term economic impact due to damages of lifelines can be significant. In a recent study by CATDAT (Daniell and Vervaeck 2013), the disaggregation of direct economic losses of 61 selected major earthquakes

from 1900 to 2012 showed that around 30 % of the direct losses came from infrastructure losses in transport, communications, pipelines, energy supply systems, etc. Knowledge, experience, and lessons learned from past earthquakes have significantly contributed to understanding the structural and functional behavior of lifelines and critical facilities and to develop more appropriate methods and tools toward effective risk mitigation.

The lessons learned from past earthquakes also concern the readiness of lifeline companies and authorities to restore damages as well as to understand the interactions between different systems. In 1989 M_w 6.9 Loma Prieta earthquake, the damages occurred in the water network greatly affected the firefighting capacity in the area of San Francisco, Oakland, and Berkeley. In 1994 M_w 6.7 Northridge earthquake, all of the 1.5 million power customers and about 50,000 water customers lost service, many for a week or

more (Schiff 1995). Earthquake-related damage to electric power systems can cause considerable direct repair and restoration costs, business interruption, inconvenience, and permanent loss of data, food, and perishable goods. Similarly, earthquake damage to water supply systems can have serious economic and health consequences. After the 1995 M_w 6.8 Hyogoken-Nanbu earthquake, it was observed that the restoration period of water system in Kobe was much longer compared to other major earthquakes due to traffic congestion, street blockades, damaged buildings, and water that flowed into gas pipelines (Kameda 2000). In general, the functional restoration for the damaged networks in Kobe needed five to ten times more time than those experienced in other major earthquakes. The final recovery period in Kobe was 82 days for water supply, 85 days for natural gas supply, 7 days for power supply, 20 months for highway bridges, 6 months for major lines of railways, and about 2 years for the main port facilities.

Main Actions for Risk Mitigation of Lifelines and Critical Facilities

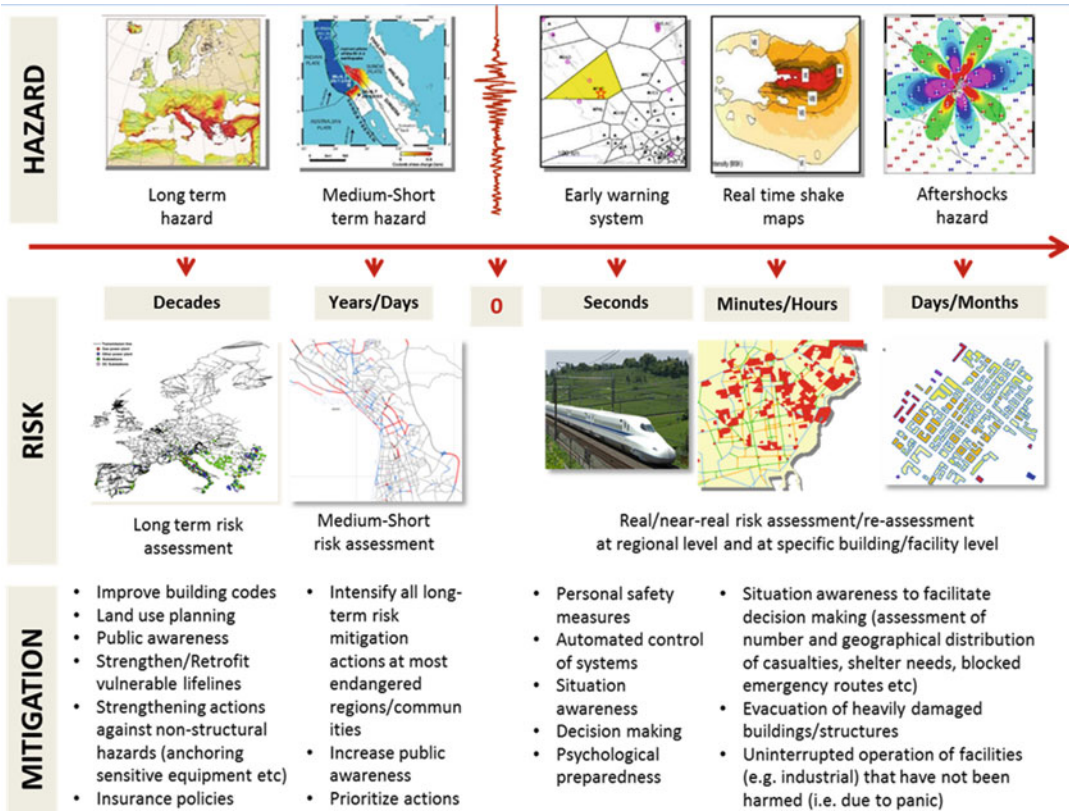
The general mitigation framework of lifelines and critical facilities includes the following main actions: risk assessment (1), risk reduction (2), readiness for emergency response (3), and readiness for recovery (4). Each of these actions is shortly described in the next paragraphs.

1. **Risk assessment** The estimation of seismic risk is the fundamental step in risk management and decision-making for risk mitigation. Risk assessment refers to the probable future losses (physical, performance, economic, social) not only due to structural damages but also considering the business interruption or interdependencies of lifelines. Lifeline companies and civil protection authorities are asked to determine how to best allocate the limited resources for risk reduction measures (actions 2, 3, and 4) based on the results of seismic risk analyses for their facilities and the

selection of “acceptable” risk. Still the decision by itself is very challenging, because seismic risk assessment includes various uncertainties related to seismic hazard, vulnerability, performance, and loss estimation procedures. Moreover the selection of acceptable risk needs appropriate balance between cost, risk, and benefits, which is different for each network and society.

Seismic hazard and therefore seismic risk are considered as variable in space and time. Risk assessment in different time scales is illustrated in Fig. 2 with reference to the occurrence of a disastrous earthquake. Examples of mitigation actions appropriate for each time period are also given. Long-term assessment is based on the seismic hazard model providing, for example, PGA values for outcrop rock conditions with 475 mean return period for Euro-Mediterranean region (Giardini et al. 2013). Such maps can be used for the risk assessment of common structures; for essential facilities (e.g., communication towers, hospitals, major bridges), the hazard assessment and risk require longer return periods (e.g., up to 5000 years). Short to medium term hazard and risk assessment may also be computed often using defined scenario type earthquakes.

Recent technological advances in seismic instrumentation and telecommunication permit the implementation of early warning systems (Fig. 2c) and real-time hazard and risk assessments (Fig. 2d), which are related to the emergency response actions in the co-seismic phase (Gasparini et al. 2007; Allen et al. 2009; Zschau et al. 2009; Allen 2011; Erdik and Fahjan 2006). They involve the generation of alarm signals directly from online instrumental data and the generation at a first stage of real-time ground estimation maps and in the second stage the vulnerability and risk assessment of structures provided that fragility curves are available. The alarm signals alert for the coming earthquake in a time window of *few seconds* and enable safe shutdown of critical facilities (e.g., nuclear plants, gas network) or train and traffic stopping systems.



Earthquake Risk Mitigation of Lifelines and Critical Facilities, Fig. 2 Risk assessment and examples of mitigation actions in different time scales with reference to the occurrence of a disastrous earthquake (EQ)

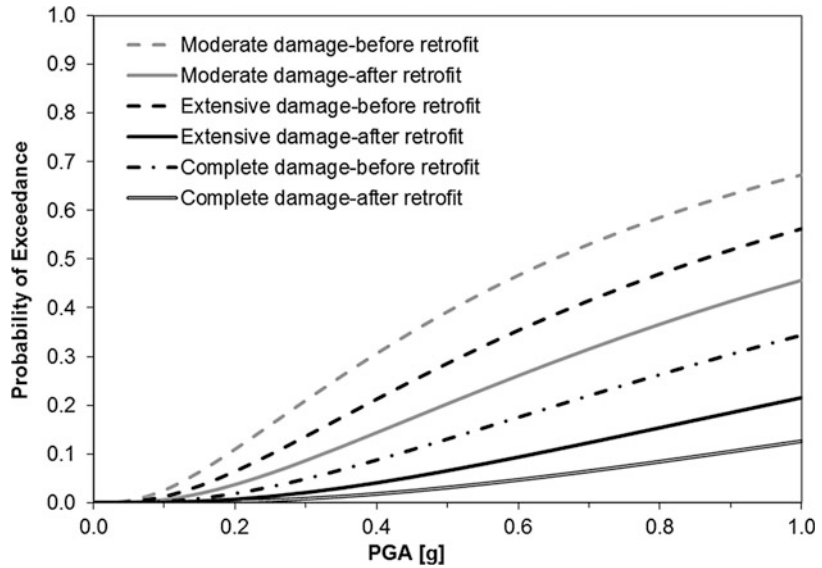
The assessment of ground shaking intensity maps and/or physical damage and casualties takes part immediately after earthquake (*few minutes/hours*) and assists the lifeline owners and stakeholders for the emergency actions (rapid response system). In this way the impact of a disastrous earthquake can be reduced due to timely and effective operations (e.g., allocation of resources and prioritization of interventions). Such dense strong ground motion monitoring systems capable of computing damage and casualties in near real time have been implemented in regional or city scale in California, Yokohama, Tokyo, Istanbul, Taiwan, Bucharest, and Naples (Erdik and Fahjan 2006). In other cases, early warning systems are implemented for the monitoring of individual structures and critical facilities (see REAKT project, <http://www.reaktproject.eu/>).

Finally, a critical issue in the risk management and reduction of impact after a seismic event is to assess the functionality of mainshock damaged infrastructures (e.g., major bridges, hospitals) considering the aftershock hazard (Fig. 2e). This is a challenging issue due to limited time availability (*days/months*), the lack of adequate models for the fragility assessment of damaged structures, and the inherent uncertainties in aftershock hazard assessment.

2. **Risk reduction** is related to the identification of the components or parts of the networks and infrastructures that present particular vulnerability and can result to high loss of network performance due to seismic damages. The definition of priorities for retrofitting should be based on vulnerability and risk assessment studies. Specific risk reduction measures should be taken before the earthquake in

Earthquake Risk Mitigation of Lifelines and Critical Facilities,

Fig. 3 Fragility curves for a five spans bridge, with an expansion joint in the center span, supported by four columns, before and after retrofit (Kim and Shinozuka 2004)



order to improve the seismic performance of infrastructure assets. These include the following:

- **Strengthening of vulnerable infrastructure assets** to improve their seismic performance. Several utility systems and networks are now outdated and often without proper seismic design and protection, as, for example, parts of water and wastewater systems in several cities were built more than a century ago. State-of-the-art technical guidelines for seismic assessment, retrofit, and rehabilitation have been published around the world, with emphasis to buildings. Retrofitting techniques are generally case dependent and the description of their typologies is beyond the present essay. Few typical examples are given below:

Geo-structures: Stabilization of embankments, retaining walls, and steep natural slopes; ground improvement, especially in case of liquefaction susceptibility.

Pipelines and wastewater networks: Replacement of vulnerable segments, upgrade of valves, local ground improvement, and increase of dead load of the pipes in particular in liquefaction-prone areas.

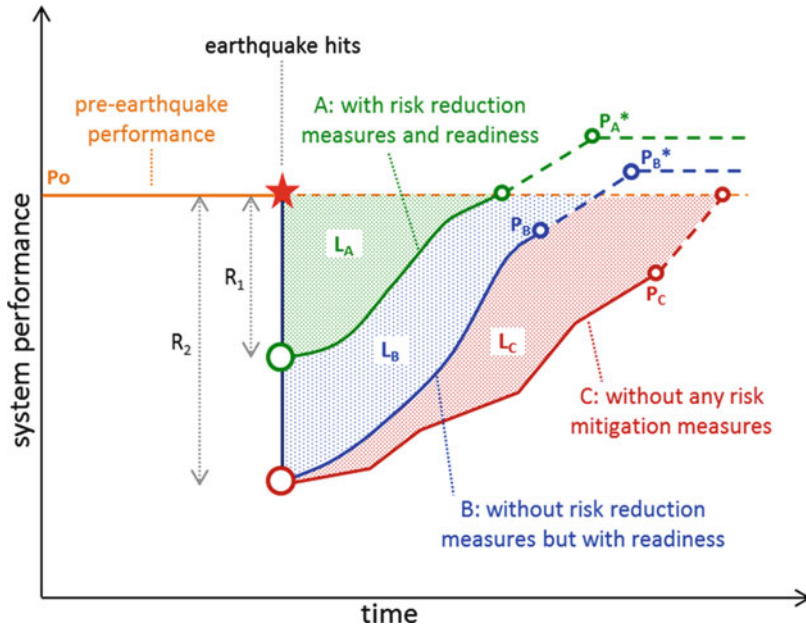
Bridges: Retrofitting for continuity to multiple simple spans through restrainers

and/or shear blocks; replacement of old bearings with ductile, multi-rotational, and multidirectional bearings; retrofitting of concrete columns with insufficient confinement through steel jacketing to provide passive confinement. Several studies have been conducted showing the effect of retrofitting. An example is shown in Fig. 3, where the effect of retrofit by means of steel jacketing of columns is illustrated in the fragility curves of typical bridges in southern California. In the USA, the Federal Highway Administration (FHWA) has published a manual including most current state of practice in assessing the vulnerability of highway structures to the effects of earthquakes and implementing retrofit measures to improve performance (FHWA 2004, 2006).

- **Undertake countermeasures against geotechnical hazards** such as liquefaction (later spreading, settlement) and landslides/rockfalls or **other induced phenomena** such as the tsunami waves (e.g., protecting walls).
- **Improve the design and construction practices** for new lifeline components. The design of critical infrastructures should be based on advanced and detailed seismic

hazard analyses, considering the topographical and geotechnical conditions of the site at local or regional scale. Soil-foundation-structure interaction effects are of high importance and should be considered in the design of critical infrastructures (e.g., major bridges). The uses of special materials and techniques that improve the seismic performance of structures (e.g., ductile materials for pipelines, soil-rubber mixtures for embankments) and the use of base isolation systems are other practices for seismic risk reduction of several lifeline components.

- **Relocate components** if this is cost beneficial or more feasible instead of taking large-scale strengthening countermeasures, for example, to stabilize landslide zones. The case of active seismic fault crossing is another typical example where sometimes relocation of road or railway lines should be examined considering all other parameters involved.
 - **Installing restraints** to restrict the failure (toppling) of sensitive equipment and nonstructural components, such as computer equipment, communication devices, batteries, health-care equipment, diesel tanks, or equipment installed on roofs and ceilings.
 - **Improve the redundancy** of specific components where this is cost-effective. For example, redundant diesel generators and diesel tanks in hospitals, main water pumping stations, and strategic buildings.
3. **Readiness and resourcefulness for emergency response**, which is related to the repairs and measures for the safety of the public and the functionality of the critical facilities immediately after the disaster. Particular actions are:
- **Response planning**, including prioritization for restoration interventions, availability of resources (inspection and repair crews) and means (equipment, materials, and spare parts), organization, and coordination of the emergency response. The prioritization for repairs in this phase is related to the emergency needs, such as accessibility and supply of energy to hospitals and critical facilities. Other specific actions that can improve the readiness and rapidity include the arrangements to have helicopters committed for use in post-earthquake inspections of extended networks (e.g., railroads), the regular training of staff, and check and maintenance of material and equipment. The use of advanced tools like GIS and remote sensing techniques and continuous monitoring of the functionality of the networks are nowadays indispensable.
 - The need for **coordinated actions** between the different lifeline owners and stakeholders is also emphasized due to the interactions between different lifeline elements. For example, the restoration of utilities such as gas, water, wastewater pipes, and communication or electric power cables that exist underneath urban roads requires a coordinated action. A representative case of interactions among different lifeline systems during the restoration period is from 1995 Kobe earthquake. The restoration activities of the water and gas networks in the Kobe area have been delayed due to traffic congestion, street blockades, damaged buildings, and water that flowed into gas pipelines. It is clear that an efficient rescue policy and an optimum recovery strategy require the evaluation and the efficient governance of the interactions between different lifeline systems and authorities.
 - Measures to mitigate losses due to **after-shocks**, such as short-term supporting countermeasures of heavily damaged structures.
 - Installation of **early warning systems**, to alert for the coming earthquake and cut off the operation of high-risk facilities such as the gas network or to slow down high-speed trains. Such systems are currently implemented in Bay Area Rapid Transit system (www.bart.gov) and Japan Railways (Nakamura and Saita 2007).



Earthquake Risk Mitigation of Lifelines and Critical Facilities, Fig. 4 Effect of mitigation measures and readiness (A, B, C) in the recovery process after the occurrence of a disastrous earthquake and representation of a system with high (R_1) and low (R_2) resilience. Ideally, in case A, the performance of the network can reach the pre-disaster

performance level (P_O) or higher (P_A^*). In case B, the system's performance can reach lower (P_B), P_O , or higher (P_B^*) level, while in case C the performance will most probably reach P_O or lower (P_C) level. The losses are represented by L_A , $L_A + L_B$ and $L_A + L_B + L_C$ for mitigation measures A, B, and C, respectively

4. **Readiness and rapidity for recovery** is related to the restoration period and until the functionality of all lifelines and facilities returns to the pre-disaster level. It can be described in terms of time needed to reorganize services, activities, and functions. An efficient planning is required for the prioritization and organization of restoration works. The post-earthquake restoration approaches are briefly described in a next section. The implementation of pre-earthquake risk reduction measures together with the readiness for recovery considerably affects the evolution of the recovery process (Fig. 4).

The resilience of the system (e.g., water supply, electric power) is defined by its ability to withstand the disastrous event without suffering devastating losses and to recover shortly. The recovery time is the period necessary to restore the functionality of the system to a desired level that can operate the

same, close to, or better than the pre-disaster one. Several uncertainties are associated to the recovery time such as the structural restoration time and the business interruption time. It typically depends on the earthquake intensities and on the location of the system with its given resources such as capital, materials, and labor as well as public policies, following the major seismic event (Cimellaro et al. 2010).

Figure 4 illustrates the response and recovery time of a system with lower and higher resilience in relation with the risk reduction and readiness measures. Three cases are described: with risk reduction and readiness measures (A), without risk reduction measures, but with readiness (B), and without any measures (C). Obviously, in reality there are several variations of risk mitigation measures, resilience levels, and recovery time. Ideally, in case A, the performance of the network can reach the pre-disaster

performance level (P_O) or higher (P_A^*). In case B, the system's performance can reach lower (P_B), P_O , or higher (P_B^*) level, while in case C the performance will most probably reach P_O or lower (P_C) level. Losses are represented by L_A , $L_A + L_B$, and $L_A + L_B + L_C$ for mitigation measures A, B, and C, respectively. Both losses and recovery time are increasing from A to C.

Prioritization for Mitigation Actions

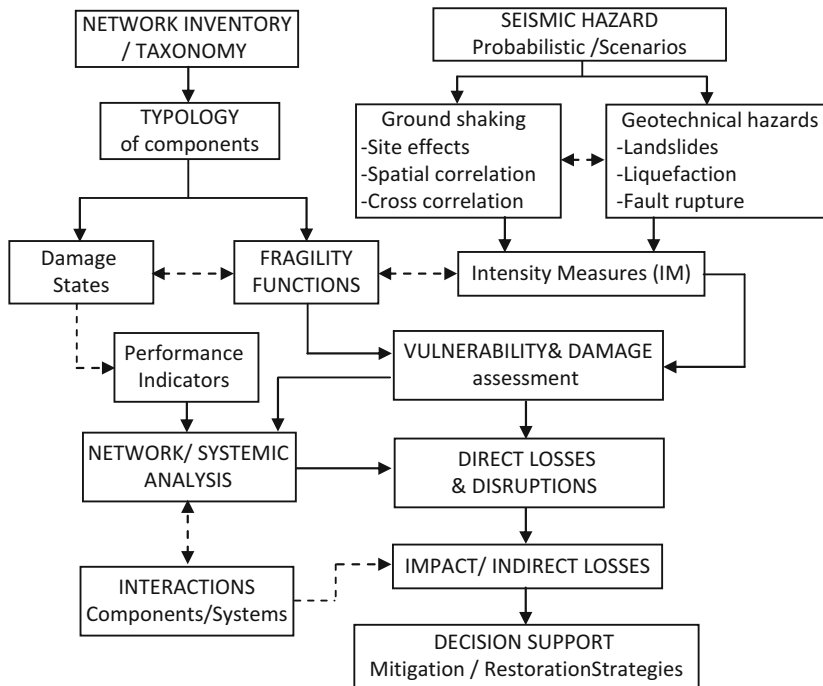
Pre-seismic (retrofitting and strengthening actions), co-seismic, and post-seismic strategies and policies must be based on advanced approaches that quantify the seismic risk of lifelines and critical facilities considering seismic hazard, vulnerability, and exposure of the networks. The final prioritization criteria combine engineering techniques, economic analysis tools and decision-making, or political aspects. The available engineering techniques for seismic risk assessment and prioritization for mitigation actions are classified in two main groups:

(A) *Screening approaches based on multi-criteria value models.* A rating system is used to assign a score in each attribute of the factors that describe the overall risk of the network component. The scaling values are defined based on expert judgment and the experience of past earthquakes, while weighting factors are commonly used in order to account for the relative contribution of each attribute to the overall risk of the component. This type of models has been mainly developed for buildings and bridges, and the main factors considered are (i) the level of seismicity, usually described by the peak ground acceleration (PGA) value as derived by national provisions or based on site-specific studies required for important structures and infrastructures; (ii) the vulnerability of each element at risk, based on the typology and other geometrical and structural characteristics of the element at risk and soil

characteristics, such as local site conditions and the potential for liquefaction or landslides; and (iii) the importance of the structure, which is evaluated based on its function within the network, the impact to the economy, the interdependencies with other lifelines, the public safety, or its importance for the emergency response. Such models have been proposed for buildings, bridges, as, for example, in the USA by the Federal Highway Administration (FHWA 1995) and in Japan by Kawashima and Unjoh (1990) and for utility networks (Wang et al. 1995; Muhlbauer 1996).

(B) *Advanced risk assessment studies* aiming to the estimation of probable direct and indirect losses (physical, performance, economic, social) due to future earthquakes considering also interactions among different systems. A general layout for the seismic risk assessment of lifelines is outlined in Fig. 4 and is further described in Pitilakis and Argyroudis (2014). Inventory and taxonomy/typology definitions of all elements at risk and systems are the necessary initial steps to describe the elements and networks exposed to seismic risk. The estimation of site-specific seismic ground motion and other geotechnical hazards and the selection estimation of appropriate seismic intensity measures (IM) is the basis of the seismic hazard analysis. Vulnerability is the expected response of an element or a network to a given seismic intensity. It is commonly assessed based on fragility models, which estimate the probability of exceeding certain damage states for given seismic intensity. Fragility functions are adapted with respect to the typological characteristics of each element at risk. An extensive review of available fragility functions and state-of-the-art methods for vulnerability assessment of buildings and lifeline components can be found in Pitilakis et al. (2014a).

The expected damage and loss of the entire network are estimated through appropriate systemic analysis, which takes into account the

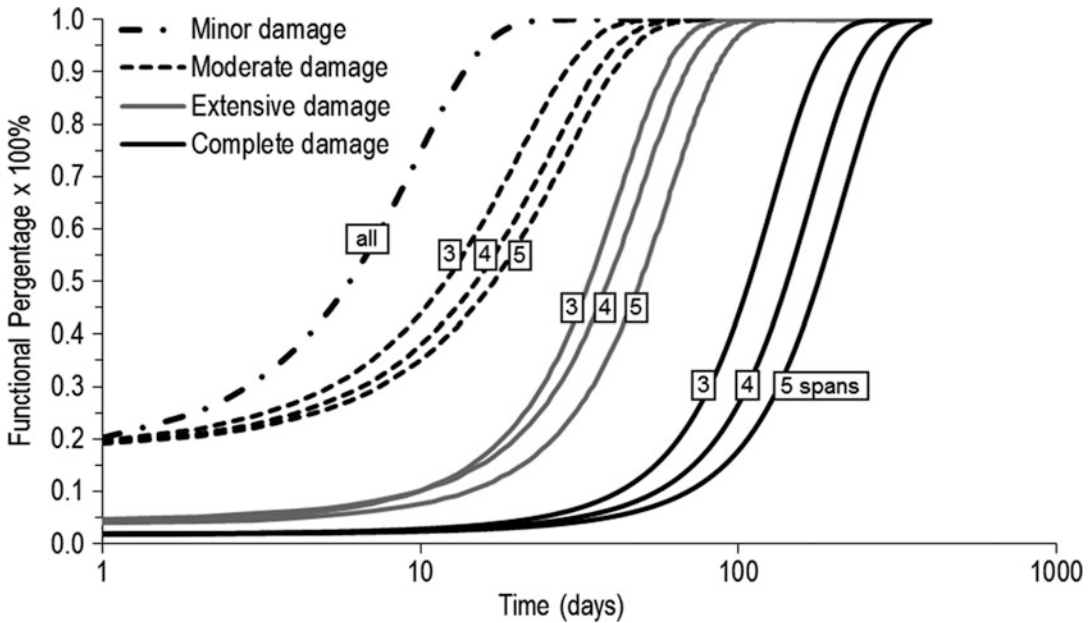


Earthquake Risk Mitigation of Lifelines and Critical Facilities, Fig. 5 General layout for the seismic risk assessment of lifelines

physical damage and the relations and interactions between the different components and systems (Pitilakis et al. 2014b). The interdependencies among different systems may considerably increase the overall impact. The results, which are commonly provided in terms of performance indicators that describe the network functional losses or the expected economic or socioeconomic losses (Khazai et al. 2014), provide the means for an efficient mitigation or recovery planning. In particular, priorities for retrofitting or adding redundancy of the elements with greater risk can be defined. The level of correlation of damaged components with the network loss performance along with additional indicators and criteria for the characterization of the importance of each component can be used in order to classify the mitigation priorities. Obviously, several sources of uncertainties are inherent in the various definitions, models, tools, and methods for estimation of losses, which make the risk assessment of lifelines a complex and challenging topic (Fig. 5).

Post-earthquake Restoration Modeling of Lifeline Systems

Technically the restoration process is part of the recovery process and commonly is expressed by restoration curves, which define the recovered functionality of a damaged network component (e.g., bridge, water pumping station, or harbor crane) as a function of elapsed time since the earthquake occurrence, which normally is the main event. However in case of very strong earthquakes, the aftershock activity is expected quite high for few or several months, a fact that is seriously affecting most actions in the post-earthquake recovery period. For any system, restoration depends on the characteristics and the damage level of the individual components as well as the network performance and the interactions with other lifelines. In general, the estimation of the required time for recovery is a complicated task and varies significantly between different countries or cities, as it depends on numerous factors.

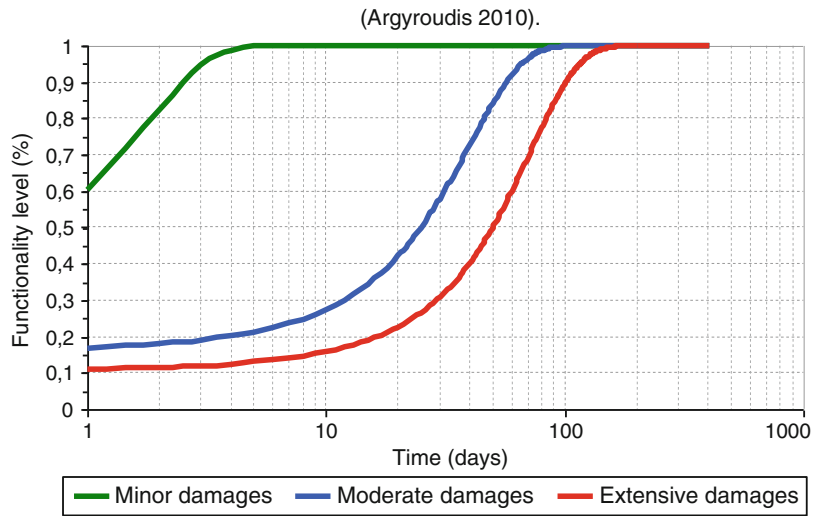


Earthquake Risk Mitigation of Lifelines and Critical Facilities, Fig. 6 Restoration curves for reinforced concrete bridges as a function of number of spans and damage state (Argyroudis 2010)

Restoration process is based on the available construction techniques, local experience and expertise, the regulations and regional politics, the available manpower, material and equipment together with the knowledge of the system, the economic resources, the interactions of individual components and networks, the accessibility of the damaged components, and all these in relation with the intensity, typology, and spatial variability of the damages. Hence it is obvious that restoration curves for lifelines may differ considerably from one country or city to another. Modeling of post-earthquake lifeline restoration can be achieved using experts' opinion, empirical data, theoretical models, or a combination of these. In ATC-25 (ATC 1992), restoration curves are constructed using ATC-13 (ATC 1985) data sources that are based on regression analysis of expert opinion data. These restoration curves were adapted also in HAZUS methodology (FEMA 2003). Examples of empirical approaches include the functions developed for the electric power system in Memphis based on the 1994 Northridge earthquake and ATC-25 data (Shinozuka et al. 1998) or for the water system in

same area, based on data from the 1971 San Fernando earthquake (Chang et al. 1996). Figures 6 and 7 show examples of restoration curves for highway bridges and port facilities based on experts' opinion. In some approaches, resource constraints are accounted for by specifying the number of repairs that can be made in any time period as a function of the personnel available, assuming that all necessary material and expertise are available. In this way, the progress of the restoration across time and space is depicted, while it is possible to investigate the effect of prioritization of repairs. This approach has been employed by Isumi and Shibuya (1985), Ballantyne (1990), Chang et al. (1999), and HAZUS (EMA 2003) for water supply system. The personnel and rate of repairs are assumed to be constant or not with time. However, the actual restoration time of each component may be very different depending on the parameters influencing the restoration process. Moreover, the recovery of the whole system is an even more complicated issue as it depends on the restoration of all individual components. Therefore, the aforementioned models are deterministic and

Earthquake Risk Mitigation of Lifelines and Critical Facilities, Fig. 7 Restoration curves for cargo handling equipment (Kakderi 2011)



generally of moderate reliability, as the various uncertainties related with the restoration time are not accounted.

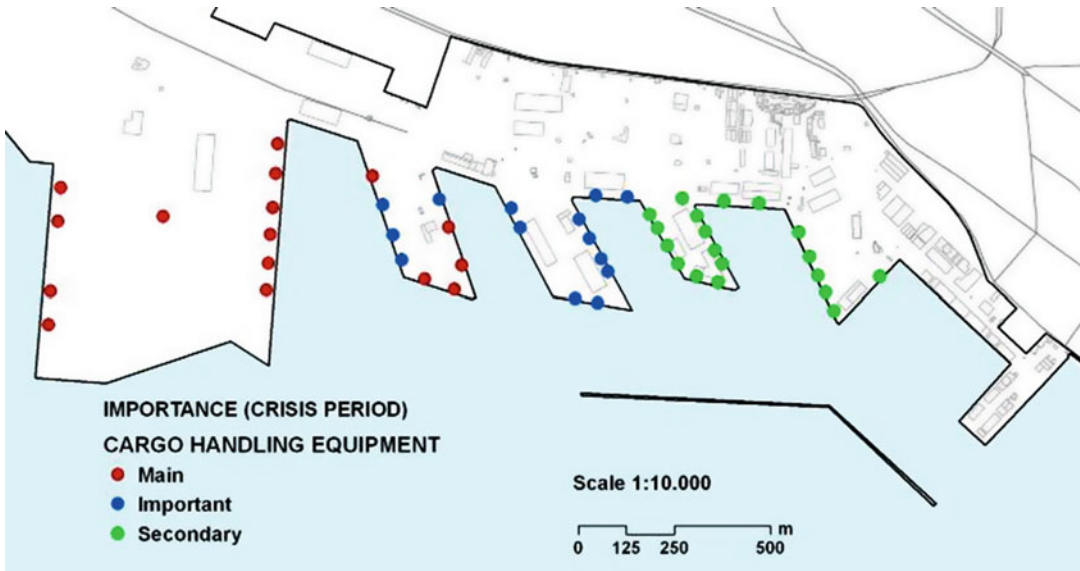
Dynamic simulation approaches that can be either deterministic or stochastic have been employed recently to represent utility company's decision variables (e.g., repair prioritization plans, mutual aid agreements, repair resources) and to produce spatially disaggregated restoration curves, so as one can end up with restoration curves for each region of the service area, rather than just one for the entire lifeline system (Kozin and Zhou 1991; Zhang 1992). Statistical variability and uncertainties associated with key parameters, such as post-earthquake damage inspection durations, repair durations for each damaged component, time needed to replace components that cannot be repaired, and amount of available resources, are taken into account by defining these parameters as random variables with specified probability distributions (Nojima and Kameda 1992; Cagnan and Davidson 2007). However, these approaches require several simplifying assumptions and numerous input data, which are not easily and always available.

As a conclusion, the construction of restoration curves for any component and each lifeline system must be defined by local actors, using basically expert elicitation, considering local conditions and constraints. The reduction of the time delays is actually a political decision of great

importance, where local authorities, government, and lifeline actors have a major role. It is also important to notice that any pre-earthquake actions (upgrading, retrofitting of the most vulnerable elements and sectors) have a very beneficial impact on the definition of the specific restoration curves.

Example of Prioritization in the Restoration Process

An example of seismic risk assessment and prioritization for restoration of damages is presented for the port of Thessaloniki in Greece. In this application, the port facilities are classified into main, important, and secondary through appropriate ranking of the value of the exposed elements, based on various factors that describe the role of each element in the system. In that way, the "global value" of each element at risk depends not only on its direct specific value or content (physical and human) but also upon its indirect/immaterial value that is represented by the usefulness and relative role in the whole urban system, at a specific time. Three periods are identified in respect to the occurrence of an earthquake event: normal, crisis, and recovery. "Global value" evaluation in different periods could be a powerful tool for the prioritization of pre-earthquake actions and quantification of the



Earthquake Risk Mitigation of Lifelines and Critical Facilities, Fig. 8 Classification of the importance of cargo handling equipment of Thessaloniki's port during the crisis period

overall importance of different complex and coupled lifeline systems. Several criteria for this are used, such as operational attributes, land use, population influenced, human losses, economic and social weight under normal, crisis and recovery circumstances, identity/radiance, environmental impact, and others. Appropriate qualitative or quantitative indicators can then be defined for each period, while relevant measuring units are used for their evaluation and the identification of “main,” “important,” and “secondary” elements and system’s weak points (RISK-UE project: Mouroux and Le Brun 2006; Pitilakis et al. 2010).

The cargo handling equipment of Thessaloniki’s port are classified based on indicators that describe the capacity, location, type of cargo handled, and redundancy (Pitilakis et al. 2010), and representative maps illustrating the classification of elements at risk are constructed (Fig. 8).

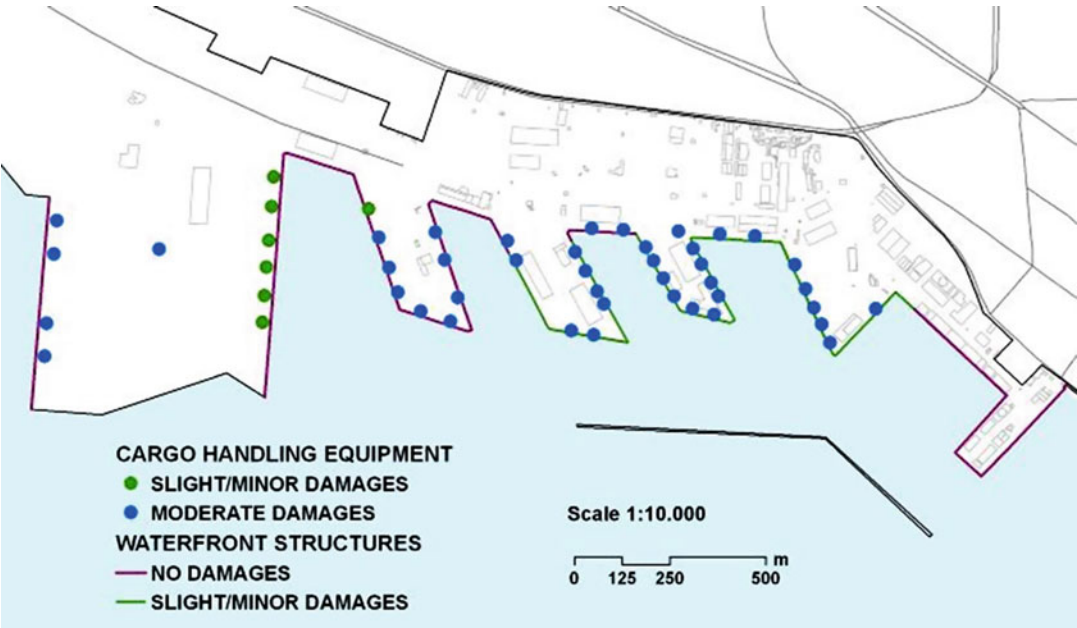
Combining “global value” evaluation and vulnerability assessment and using, if necessary, an “expert opinion,” it is possible to estimate priorities and to account for the economic and social losses, for a specific utility system and a given seismic scenario. Recovery activities

Earthquake Risk Mitigation of Lifelines and Critical Facilities, Table 1 Risk analysis matrix showing cargo handling equipment seismic retrofit priorities

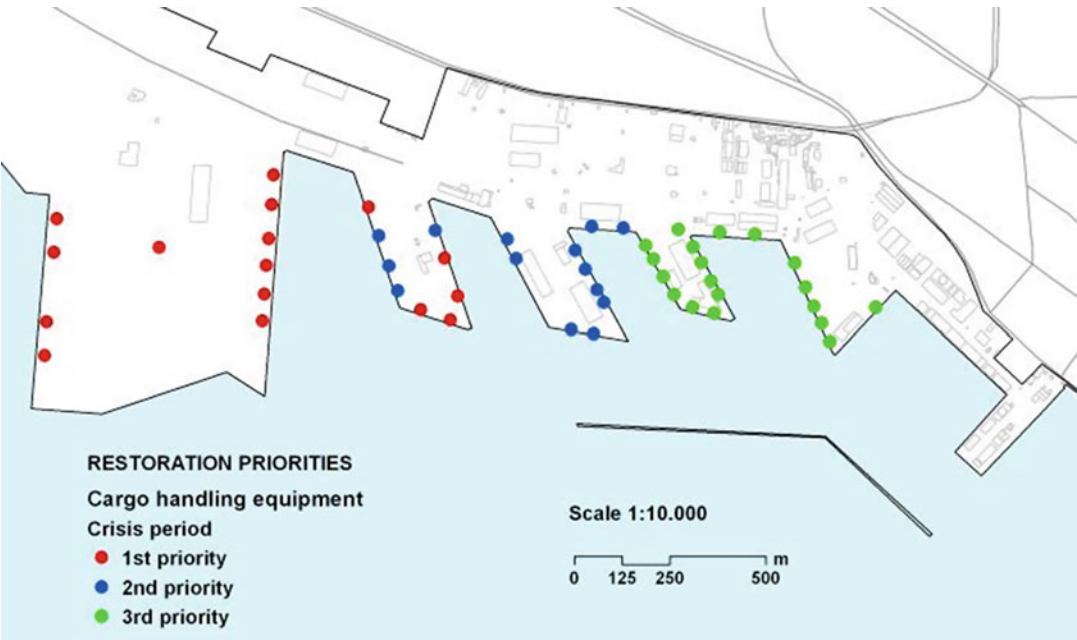
Urban risk/ seismic hazard	Priorities		
	Main	Important	Secondary
Complete/ extensive damages	1st priority	1st priority	2nd priority
Moderate damages	1st priority	2nd priority	3rd priority
Slight/minor damages	1st priority	2nd priority	4th priority

could also follow these priorities aiming at efficient seismic risk management procedures. Table 1 summarizes the application of the proposed approach for the cargo handling equipment of Thessaloniki’s port. Figure 10 illustrates the restoration priorities for a damage scenario with mean return period of 475 years (Fig. 9) (Pitilakis et al. 2010).

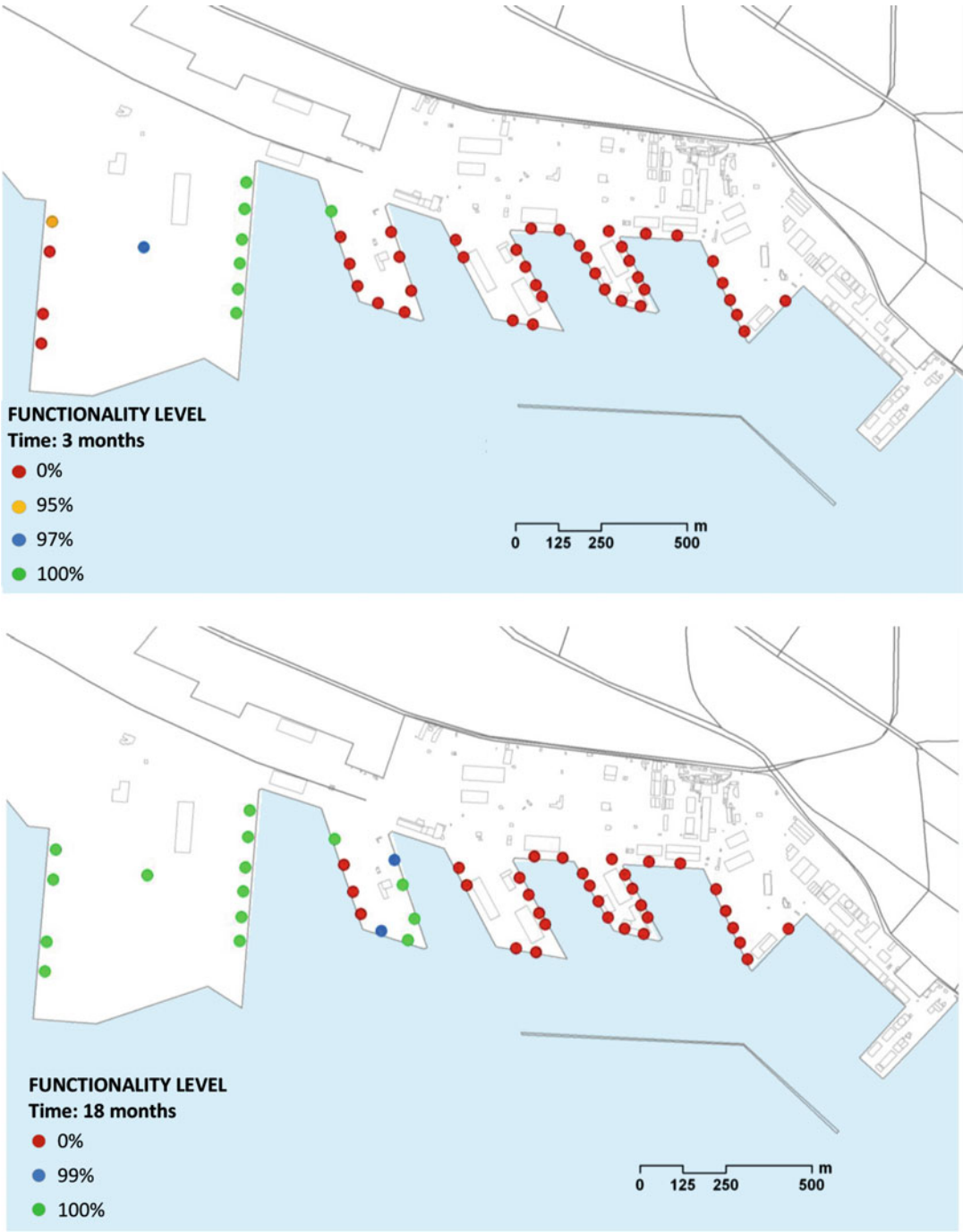
Restoration curves have been defined in collaboration with the port authority (Fig. 7), based on available manpower, capabilities, local experience, and expertise. Figure 11 illustrates the functionality level of cargo handling equipment at 3 and 18 months after the seismic event,



Earthquake Risk Mitigation of Lifelines and Critical Facilities, Fig. 9 Distribution of damages to waterfront structures and cargo handling equipment of Thessaloniki's port ($T_m = 475$ years)



Earthquake Risk Mitigation of Lifelines and Critical Facilities, Fig. 10 Restoration priorities of cargo handling equipment of Thessaloniki's port during the crisis period



Earthquake Risk Mitigation of Lifelines and Critical Facilities, Fig. 11 Functionality percentage of cargo handling equipment of Thessaloniki's port in 3 and 18 months' time after the seismic event

Earthquake Risk Mitigation of Lifelines and Critical Facilities, Table 2 Infrastructure upgrades in San Francisco area (Zoback 2014)

Infrastructure provider	Number clients served	Scope of upgrades	Total cost/source of funding
Pacific Gas and Electric	15M throughout North CA	System upgrade of underground gas, electrical components, substations, and administration building	\$2.5B Ratepayers
Bay Area Rapid Transit	400,000 daily ridership	Retrofit core system-aerial structures, stations, trans-bay tube (completion system by 2018, tube by 2023)	\$1.3B Bonds and taxpayers/ \$3 M from FEMA
East Bay Municipal Utility District	1.3M in East Bay	Entire system upgrade: pipelines, fault crossings, dams, administration building, pumping, and treatment plants (completed 1999)	\$0.189B Ratepayers
California Department of Transportation	38.3M statewide	Structurally upgraded and seismically retrofit over 2000 bridges and overpasses, new E span Bay Bridge	\$13.08B CA taxpayers
San Francisco Public Utilities Commission	2.6M residential, commercial, and industrial	Upgrade of >100-year-old Hetch Hetchy water system (pipelines), fault crossings treatment facilities, and reservoirs (completion by 2016)	\$4.6B Bond measure
Total investment			\$21.6B

assuming that the restoration process starts immediately after the earthquake and that there are two available working groups for the restoration process.

Lifeline Seismic Retrofitting Programs

Some examples of risk reduction measures through retrofitting programs taken by lifeline stakeholders over recent years to reduce the impact of earthquakes are given in the following.

Lifelines Council in San Francisco

The city and county of San Francisco convene a Lifelines Council with a scope focused on coordinated risk reduction strategies and post-disaster reconstruction and recovery. The Lifelines Council seeks to develop and improve collaboration in the city and across the region; understand intersystem dependencies to enhance planning, restoration, and reconstruction; share information about recovery plans, projects, and priorities; and establish coordination processes for lifeline restoration and recovery following a major disaster event (<http://sfgsa.org/index.aspx?page=4964>).

In the 25 years since Loma Prieta, San Francisco Bay area infrastructure providers have

dramatically increased their disaster resilience through substantial system upgrades, an investment of \$21 billion funded almost exclusively by ratepayers and taxpayers – almost \$1 billion/year (see Table 2). A five-step strategy has been employed by the main infrastructure providers to increase their earthquake resilience (Zoback 2014):

1. Assess vulnerabilities of their systems to expected earthquakes.
2. Set performance goals for the systems after the earthquake.
3. Communicate risks/benefits and secure funding.
4. Develop creative and innovative solutions for these complex problems. These include building redundancy into the system at the start and taking advantage of, and in some cases funding, research to improve solutions.
5. Continue to reassess system performance as upgrades proceed. Develop real-time damage assessment capability using near real-time maps of earthquake shaking intensity provided by the USGS overlain by their system fragility functions.

Bridges in California

After the 1971 Sylmar earthquake that struck the Los Angeles area and damaged several bridges,

the California Department of Transportation (Caltrans) initiated a bridge seismic safety retrofit program focused on bridge expansion joints. The mid-span hinge and support joint retrofit program was completed in 1989 at a cost of \$55 million. With the significant loss of life and closure of major routes during the 1989 Loma Prieta earthquake, the Caltrans established a new seismic retrofit program for the more than 12,000 bridges in the California State Highway system. This program is focused on identifying and retrofitting existing bridges statewide, bringing them up to the latest seismic safety retrofit standards established to prevent collapse during future earthquakes. In particular, after the Loma Prieta earthquake, the Department initially identified 1,039 state highway bridges (mostly single-column bridges) in need of being retrofitted to meet seismic safety standards (Phase 1 Program). After the 1994 Northridge earthquake, the Caltrans identified another 1,155 state-owned bridges (mostly multicolumn bridges) in need of being retrofitted (Phase 2 Program). The state's bridge earthquake strengthening program, which has been almost completed, involves more than 2,200 structures and costs more than \$12 billion. A separate program was established for toll bridges which are the largest and most complicated bridges in the State. An estimated \$2.6 billion is allocated for this program, which also includes the replacement of the east span of the San Francisco-Oakland Bay Bridge (<http://www.dot.ca.gov/hq/paffairs/about/retrofit.htm>).

Bay Area Rapid Transit (BART) System

BART is one of the Bay Area's most vital transportation links, carrying about 360,000 travelers every day. BART's success in maintaining continuous service directly after the 1989 Loma Prieta earthquake reconfirmed the system's importance as transportation "lifeline." However, because of the likelihood that BART will be subject to a major future earthquake, an Earthquake Safety Program has been recently initiated. The program will upgrade vulnerable portions of the original system to ensure safety for the public and BART employees. Portions of the original system with the highest traffic will be upgraded

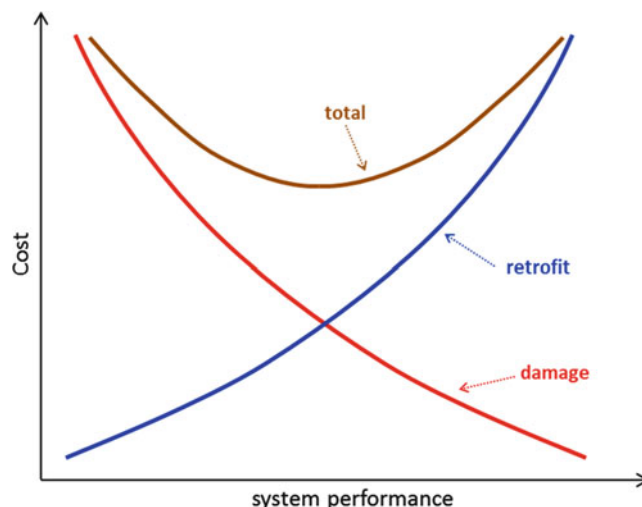
not only for life safety but also to ensure that they can return to operation shortly after a major earthquake. The upgrades will be accomplished by using the latest seismic standards to improve the structural integrity of BART facilities. The completion of all earthquake upgrades is anticipated by 2018. The total project is \$1.2 billion (in 2004 dollars). Preliminary upgrade concepts include: Vibro replacement to compact soil backfill, increasing seismic joint capacity and sealing around joints, and new concrete shear walls in the Oakland Ventilation Structure. Potential upgrading concepts for elevated structures include enlarged or reinforced concrete foundations, "jackets" around concrete columns, additional shear keys, and additional foundation piles where poor soil conditions exist (<http://www.bart.gov/about/projects/eqs>).

Lifelines in New Zealand

Lifeline vulnerability to natural hazards including earthquakes had been given prominence in a comprehensive project undertaken by the Christchurch Engineering Lifelines Group in the first half of the 1990s. Utility networks were described, broken down to component level, with each one assessed for vulnerability to the various natural hazards. Assessments considered "importance" to the network as well as "vulnerability" and "damage impact" estimated in three time frames: immediately after the event, period following, and time for return to normality. Different lifeline systems were covered, namely, energy, telecommunications, transport, and water/sewerage. The project also included emergency buildings like broadcasting, police and fire stations, and ambulance bases. Several mitigation measures have been undertaken in the following years. For example, in the water system restraint actions taken for spare parts and equipment, seismic strengthening of reservoirs and pumping stations, valve shut off improvements; in the electric power network, insecure buildings were removed and the control room was relocated, while major transformers and other essential equipment as well as bridge approaches were strengthened to protect cables. In transportation networks, improvements were made at bridges including

Earthquake Risk Mitigation of Lifelines and Critical Facilities,

Fig. 12 Superposition of cost curves as a function of the system performance for seismic retrofit cost, expected damage cost, and total cost



E

strengthening connections between superstructures and substructures, increasing column strength and ductility, strengthening or renewing retaining and approach structures, and strengthening lateral/longitudinal restraint mechanisms. Planning was improved in relation to key roads, bridges, and the important Lyttelton Tunnel. Improvements were made at the airport and port. Additional equipment, such as generators and portable pumps, were also purchased.

The substantial program of seismic mitigation undertaken by Christchurch lifeline utilities has served Christchurch in reducing losses and facilitating emergency responses and recovery during the 2010 (M7.1) and 2011 (M6.3) earthquakes. The damage would have been greater and the response slower if the emergency planning and other preparatory work had not been undertaken.

For example, the electricity distribution seismic strengthening program, commenced in 1996 and progressed systematically each year, costs \$6 million and is estimated to have saved \$60–\$65 million in direct asset replacement costs and repairs. Total repair costs from the 2011 earthquake amounted to only \$150,000. Seventeen bridges in Canterbury, including many in or near Christchurch, were identified as at risk and had been retrofitted at a cost of \$2 million. These measures prevented critical damage, as only one bridge was closed due to approach settlement and

reopened within a day. Similar cases have been noted for other networks and components in the area. It is a representative example of worthwhile investment, as the costs of seismic risk mitigation in Christchurch would have been repaid many times over if these measures were not have been taken place (Fenwick 2012).

Europe

In European countries with high seismicity (i.e., Italy, Greece, Turkey), risk assessment programs have been undertaken in the last decade for public buildings (i.e., hospitals, schools) and upgrades have been implemented in some cases. For other infrastructures (e.g., bridges), risk assessment studies and screening approaches have been partially conducted; however, retrofitting is rather limited. A short overview of the activity of bridge upgrading currently ongoing in Italy is given by Pinto and Franchin (2010).

Economic Efficiency of Seismic Retrofitting Measures

The economic efficiency of seismic retrofitting measures is related to benefit-cost analysis. Figure 12 shows an idealization of economic efficiency based on the concept of performance levels of the system under consideration. The cost

of retrofitting measures is increasing when higher performance levels are desired. On the other hand, the cost of the expected seismic damages, when the retrofitting measures are applied, is decreasing. The total cost (initial construction cost, without considering depreciation and retrofitting cost) is lower in the intersection of the two curves. In order to determine the economic efficiency, both the repair costs after expected earthquakes of different intensities and return periods and the costs of retrofitting measures to achieve certain performance levels have to be considered, usually as a function of the initial construction costs.

Benefit-cost analysis is a well-established tool to assess the viability of seismic risk reduction projects and policies. In particular, the appraisal of earthquake mitigation projects is crucial in making community, policy- and decision-makers, and insurance and construction industry to understand and quantify the effects of earthquake mitigation. The analysis for the economic feasibility of seismic retrofitting considers the achievable loss reduction, investment return period, and retrofit cost. Overall, the organization of mitigation policies for lifelines is a political and economic decision and surely not a decision that is based on emotion or public feeling, which, however, should be seriously considered. The selection of the acceptable risk considering appropriate seismic scenario or scenarios is always a major decision to make. It should take into account and combine different parameters like the detail of the seismic hazard, the will of central and local actors, the available funds, and the financial and technical capability of the community and the country. Epistemic and physical uncertainties are involved in every step of any earthquake risk reduction policy and should be taken into account.

The increase in safety and reliability of lifelines and the reduction of direct and indirect losses against future earthquakes are the main benefits of efficient retrofitting programs. However, they are not the only ones. Retrofitting strategies followed by investment policies are also a serious and efficient tool for short- and long-term growth of economic development, increasing

employment and properties' value, with benefits to the insurance and construction industry.

Summary

Lifelines and critical facilities are essential public infrastructures and networks that support citizen's life, economy, and industry. Experience from past earthquakes reveals that their components are quite vulnerable and the induced damages can have high impact in the emergency response as well as serious economic and health consequences. Losses can be even higher because of the extensive interdependencies between utility networks and infrastructures. Therefore, high priority should be given to secure safety of lifeline facilities against earthquakes and reduce the overall impact.

The decision for mitigation of lifelines against seismic hazard is a quite complex issue as different institutions from public and private sector are involved: governmental and local authorities, lifeline companies, industries, and insurances. Moreover, the decision by itself is very difficult because it should be made based on the results of the seismic risk assessment of lifelines which includes various uncertainties owing to seismic hazard vulnerability, performance, and loss estimation procedures. The risk assessment is the fundamental step for the prioritization of mitigation measures, which in general include the following actions: risk reduction, readiness for emergency response, and readiness for recovery. Risk reduction is related to the strengthening and upgrade of vulnerable assets, improvement of design and construction practices, relocation of components, improvement of redundancy, and security of sensitive nonstructural equipment. Readiness and resourcefulness for emergency response is related to the repairs and measures for the safety of the public and the functionality of the critical facilities immediately after the disaster. Typical measures include installation of early warning systems, advanced response planning, coordinated actions between different entities/stakeholders, and measures to reduce losses due to aftershocks. Readiness and rapidity

for recovery is related to the restoration period and until the functionality of all lifelines and facilities returns to the pre-disaster level. The implementation of pre-earthquake risk reduction measures together with the readiness for recovery considerably affects the evolution of the recovery process.

Pre-seismic (retrofitting and strengthening actions), co-seismic, and post-seismic strategies and policies must be based on advanced approaches that quantify the seismic risk of lifelines and critical facilities considering seismic hazard, vulnerability, and exposure of the networks. The final prioritization criteria should combine engineering techniques, economic analysis, and decision-making tools.

Cross-References

- ▶ [Community Recovery Following Earthquake Disasters](#)
- ▶ [Damage to Infrastructure: Modeling](#)
- ▶ [Early Earthquake Warning \(EEW\) System: Overview](#)
- ▶ [Earthquake Disaster Recovery: Leadership and Governance](#)
- ▶ [Earthquake Early Warning System in Taiwan](#)
- ▶ [Earthquake Mitigation Strategies Through Insurance](#)
- ▶ [Economic Recovery Following Earthquakes Disasters](#)
- ▶ [Emergency Response Coordination Within Earthquake Disasters](#)
- ▶ [Estimation of Potential Seismic Damage in Urban Areas](#)
- ▶ [Learning from Earthquake Disasters](#)
- ▶ [Liquefaction: Countermeasures to Mitigate Risk](#)
- ▶ [Probabilistic Seismic Hazard Models](#)
- ▶ [Reconstruction Following Earthquake Disasters](#)
- ▶ [Resilience to Earthquake Disasters](#)
- ▶ [Resourcing Issues Following Earthquake Disaster](#)
- ▶ [Seismic Fragility Analysis](#)
- ▶ [Seismic Loss Assessment](#)
- ▶ [Seismic Resilience](#)

- ▶ [Seismic Risk Assessment, Cascading Effects](#)
- ▶ [Site Response for Seismic Hazard Assessment](#)
- ▶ [Strengthening Techniques: Bridges](#)
- ▶ [Uncertainty Theories: Overview](#)

References

- Allen RM (2011) Earthquakes, early and strong motion warning. In: Gupta H (ed) Encyclopedia of solid earth geophysics. Springer, Dordrecht, pp 226–233
- Allen RM, Gasparini P, Kamigaichi O, Böse M (2009) The status of earthquake early warning around the world: an introductory overview. *Seism Res Lett* 80:682–693. doi:10.1785/gssrl.80.5.682
- Applied Technology Council [ATC] (1985) Earthquake damage evaluation data for California. Report no: ATC-13, Redwood City
- Applied Technology Council [ATC] (1992) A model methodology for assessment of seismic vulnerability and impact distribution of water supply systems. Report no: ATC-25-1, Redwood City
- Argyroudis S (2010) Contribution to the seismic vulnerability and risk assessment of transportation networks in urban environment. PhD dissertation, Aristotle University of Thessaloniki, Greece
- Ballantyne DB (1990) Earthquake loss estimation modeling of the Seattle water system. Report to the U.S. Geological Survey, Federal Way
- Cagnan Z, Davidson R (2007) Discrete event simulation of the post-earthquake restoration process for electric power systems. *Int J Risk Assess Manage* 7(8):1138–1156
- Chang SE, Seligson HA, Eguchi RT (1996) Estimation of the economic impact of multiple lifeline disruption: memphis light, gas and water division case study. Technical report no. NCEER-96-0011, Buffalo
- Chang SE, Shinozuka M, Svekla W (1999) Modeling Post – disaster urban lifeline restoration. In: Elliott WM, McDonough P (eds) Optimizing post earthquake lifeline system reliability: proceedings of the 5th U.S. conference on lifeline earthquake engineering. ASCE Technical Council on Lifeline Earthquake Engineering Monograph no. 16, pp 602–611
- Cimellaro GP, Reinhorn AM, Bruneau M (2010) Framework for analytical quantification of disaster resilience. *Eng Struct* 32:3639–3649
- Daniell J, Vervaeck A (2013) The CATDAT damaging earthquakes database - 2012 - the year in review, CEDIM Research Report 2013–01, Karlsruhe, Germany
- Erdik M, Fahjan Y (2006) Early warning and rapid damage assessment. In: Assessing and managing earthquake risk. Geotechnical, geological and earthquake engineering, vol 2. Springer, Dordrecht, pp 323–338
- Federal Emergency Management Agency [FEMA] (2003) Multi hazard loss estimation methodology: earthquake model – HAZUS-MH MR3 technical manual. FEMA, Washington, DC

- Fenwick T (2012) The value of lifeline seismic risk mitigation in Christchurch. New Zealand Earthquake Commission under its 'Science-to-Practice' Programme, ISBN 978-0-473-21631-3
- FHWA (1995) Seismic retrofitting manual for highway bridges: report FHWA-RD-94-052, Federal Highway Administration, Washington, DC
- FHWA (2004) Seismic retrofitting manual for highway structures: part 2 – retaining structures, slopes, tunnels, culverts and roadways, FHWA-HRT-05-067
- FHWA (2006) Seismic retrofitting manual for highway structures: part 1 – bridges, FHWA-HRT-06-032
- Gasparini P, Manfredi G, Zschau J (eds) (2007) Earthquake early warning systems. Springer, Berlin, 362 pp
- Giardini D et al (2013) Seismic Hazard Harmonization in Europe (SHARE). Online Data Resource, doi:10.12686/SED-00000001-SHARE
- Isumi M, Shibuya T (1985) Simulation of post earthquake restoration for lifeline systems. *Int J Mass Emerg Disaster* 3(1):87–105
- Kakderi K (2011) Seismic vulnerability assessment and seismic risk management of complex transportation networks. Application to port facilities. PhD Dissertation, Aristotle University of Thessaloniki, Greece
- Kameda H (2000) Engineering management of lifeline systems under earthquake risk. In: *Proceeding of the 12th world conference on earthquake engineering*, New Zealand Society for Earthquake Engineering, Upper Hutt
- Kawashima K, Unjoh S (1990) An inspection method of seismic vulnerability of existing highway bridges, *Structural Engineering and Earthquake Engineering*, JSCE, 7(1):143–150
- Khazai B, Daniell JE, Düzgün Ş, Kunz-Plap T, Wenzel F (2014) Framework for systemic socio-economic vulnerability and loss assessment. In: Ptilakis K, Franchin P, Khazai B, Wenzel H (eds) *SYNER-G: systemic seismic vulnerability and risk assessment of complex urban, utility, lifeline systems and critical facilities. Methodology and applications*. Springer, Heidelberg. ISBN 978-94-017-8834-2
- Kim SH, Shinozuka M (2004) Development of fragility curves of bridges retrofitted by column jacketing. *Probabilistic Engineering Mechanics* 19:105–112
- Kozin F, Zhou H (1991) System study of urban response and reconstruction due to earthquake. *J Eng Mech ASCE* 116:1959–1972
- Mouroux P, Le Brun B (2006) Risk-UE project: an advanced approach to earthquake risk scenarios with application to different European towns. In: Oliveira CS, Roca A, Goula X (eds) *Assessing and managing earthquake risk*. Springer, Dordrecht, pp 479–508. doi:10.1007/978-1-4020-3608-8_23
- Muhlbauer WK (1996) Pipeline risk management manual. A tested and proven system to prevent loss and assess risk. Gulf Publishing, Houston
- Nakamura Y, Saita J (2007) UreDAS, the earthquake warning system: today and tomorrow. In: Manfredi G, Zschau J, Gasparini P (eds) *Earthquake early warning systems*. Springer, Berlin/Heidelberg, pp 249–281
- Nojima N, Kameda H (1992) Optimal strategy by use of tree structure for post-earthquake restoration of lifeline network systems. In: *Proceeding of the 10th world conference on earthquake engineering*, Balkema, Rotterdam, pp 5541–5546
- Pinto PE, Franchin P (2010) Issues in the upgrade of Italian highway structures, *Journal of Earthquake Engineering* 14(8):1221–1252
- Ptilakis K, Argyroudis S (2014) Seismic vulnerability assessment: lifelines. In: Jalayer F, Galasso C (eds) *Seismic risk assessment, encyclopedia of earthquake engineering*. Springer-Verlag Berlin Heidelberg
- Ptilakis K, Anastasiadis A, Kakderi K, Alexoudi M, Argyroudis S (2010) The role of soil and site conditions in the vulnerability and risk assessment of lifelines and infrastructures. The case of Thessaloniki (Greece). In: *5th International conference on recent advances in Geotechnical Earthquake engineering and Soil Dynamics and Symposium in honor of prof. I.M. Idriss*, 24–29 May, San Diego, Paper no SPL13
- Ptilakis K, Franchin P, Khazai B, Wenzel H (eds) (2014a) *SYNER-G: systemic seismic vulnerability and risk assessment of complex urban, utility, lifeline systems and critical facilities. Methodology and applications*, Geotechnical, geological and earthquake engineering. Springer, Heidelberg. ISBN 978-94-017-8834-2
- Ptilakis K, Crowley E, Kaynia A (eds) (2014b) *SYNER-G: typology definition and fragility functions for physical elements at seismic risk*, vol 27, Geotechnical, geological and earthquake engineering. Springer, Heidelberg. ISBN 978-94-007-7872-6
- Rinaldi SM, Peerenboom JP, Kelly TK (2001) Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Contr Syst Mag* 21(6):11–25
- Schiff AJ (ed) (1995) Northridge earthquake: lifeline performance and post-earthquake response. *ASCE Technical Council on Lifeline Earthquake Engineering Monograph No. 8*, New York, p 339
- Shinozuka M, Rose A, Eguchi RT (1998) Engineering and socioeconomic impacts of earthquakes: an analysis of electricity lifeline disruptions in the New Madrid area. *Monograph no. 2*, MCEER, Buffalo
- Wang L, Li H, Ishibashi I (1995) Prioritization for rehabilitation of buried lifelines. In: O'Rourke MJ (ed), *proceedings of the 4th US conference on lifeline earthquake engineering*, ASCE Technical Council on Lifeline Earthquake Engineering Monograph No. 6, pp 588–595
- Zhang RH (1992) Lifeline interaction and post earthquake urban system reconstruction. In: *Proceeding of 10th World conference on earthquake engineering*, Balkema, Rotterdam, pp 5475–5480
- Zoback ML (2014) Written communication. Stanford University, Stanford, CA
- Zschau J, Gasparini P et al (2009) The SAFER-Project and seismic early warning in Europe, AGU, Fall meeting 2009, abstract #S21C-02

Earthquake Swarms

Josef Horálek¹, Tomáš Fischer², Páll Einarsson³ and Steinunn S. Jakobsdóttir⁴

¹Institute of Geophysics, Academy of Sciences of the Czech Republic, Prague, Czech Republic

²Faculty of Science, Charles University in Prague, Prague, Czech Republic

³Institute of Earth Sciences, University of Iceland, Reykjavík, Iceland

⁴Faculty of Earth Science, University of Iceland, Reykjavík, Iceland

Synonyms

b-value; Clustering of events; Driving forces; Earthquake location; Earthquake swarms; Focal migration; Mainshock-aftershock sequence; Source mechanism, Fluids; Triggering mechanisms; Volcanic activity

Introduction

Earthquake swarms are a specific type of seismicity. They are generally defined as sequences of seismic events closely clustered in space and time, without a single outstanding earthquake (Mogi 1963). Globally, most seismic energy is released in ordinary earthquakes, represented by a single strong event succeeded by a series of smaller aftershocks (in some cases also preceded by foreshocks) with magnitudes of one or more magnitude units lower than that of the main event. These are so-called mainshock-aftershock sequences. In contrast, an earthquake swarm consists of numerous, mostly shallow, earthquakes, which are missing a single large event. They produce few dominant earthquakes of similar magnitudes so that smaller events are not associated with any identifiable mainshock. Earthquake swarms distinctively cluster in time and space and last from several hours to several months, with some exceptional cases lasting more than a year. Swarms originate in the crust from local stresses generated by volcanic or tectonic activity. In

many cases, there is an overlapping signature of both volcanic and tectonic causes (Scholz 2002). The term “earthquake swarm” (“Schwarmbeben” in German) was very likely first used in the 1870s to describe this peculiar seismicity that occurred repeatedly in W-Bohemia/Vogtland (border of the Czech Republic, Saxony, and Bavaria).

Occurrence of Earthquake Swarms in the World

Earthquake swarms occur worldwide at boundaries of the lithospheric plates (interplate), within the plates (intraplate), as well as in subduction zones, and they are very often related to volcanic areas, areas with high activity of crustal fluids (particularly geothermal fields), and ocean ridges. Additionally, earthquake swarms are also observed in relation to human-induced stress and pressure perturbations like fluid-injection into deep boreholes for purposes of the geothermal energy exploitation. Quite frequent earthquake swarms occur in the Yellowstone volcanic field, which is one of the most seismically active areas of the western U.S. Further examples of “volcanic” swarms include the Hawaiian Islands, Alaska, Japan, New Zealand, and Canary Islands.

The strongest events of earthquake swarms are usually of local magnitudes $M_L < 5$, with few reaching $M_L \approx 6.0$. An extraordinarily strong earthquake swarm with $M_{w_{\max}} = 6.4$ and many $M_w > 5.0$ earthquakes accompanied the volcanic eruption on Miyakejima in Japan, in 2000 (Minson et al. 2007). The earthquake swarm with magnitudes reaching $M_w = 6.2$ has been reported in southern Chile in 2007. The 1965–1967 Matsushiro earthquake swarm in southwest Honshu (central Japan) was one of the most energetic swarms in the world (Cappa et al. 2009). A series of more than 700,000 events (about 60,000 earthquakes were felt) with the largest earthquake of magnitude $M_w = 5.4$ produced ground surface deformations and cracking of the topsoil and enhanced spring outflows, with changes in chemical composition and carbon dioxide degassing.

The most intense interplate earthquake swarms in Europe occur in Iceland, which lies directly on

the mid-Atlantic ridge, which separates Iceland over the two major lithosphere plates, the Eurasian plate to the east and the North American plate to the west. An exceptionally intense earthquake swarm has accompanied the Bárðarbunga volcanic eruption. The seismic crisis started in August 2014 and has been going on for 8 months. The largest earthquake reached magnitude $M_w = 5.6$ and more than 70 events had magnitudes $M_w \geq 5.0$. Further examples of European earthquake swarm areas somehow connected with the plate or microplate boundaries are western Alps, a few areas in Apennines, and also a few areas in Greece. A swarm-like activity preceded a severe $M_w = 6.3$ earthquake in L'Aquila (Abruzzo region, central Italy) in 2009.

Intraplate earthquake swarms occur mainly in Quaternary volcanism areas, which are often characterized by geodynamic unrest and other phenomena like diffuse degassing, geothermal anomalies, and chemical or dissolution anomalies.

A typical European intraplate swarm area is W-Bohemia/Vogtland. Other intraplate earthquake swarm areas affected by Quaternary volcanism are, for example, the French Massif Central, Colorado, and Long Valley in California. Vidale and Shearer (2006) examined earthquake bursts in southern California between 1984 and 2002 and identified 18 sequences as swarm like. An example of intraplate earthquake swarm areas where Quaternary volcanism is missing is the Vosges in France. One of the most intensive nonvolcanic earthquake swarm area in Japan is located in the Wakayama district in southwest Japan. The Wakayama swarm activity was initiated more than 40 years ago and has continued to the present day; the occurrence rate of earthquakes with magnitudes greater than 3.0 is >10 events per year. Earthquake swarms of magmatic origin occur frequently at mid-oceanic spreading centers, which is discussed in more detail in section “[Earthquake Swarms Induced by Magmatic and Volcanic Activity](#).”

Primary Earthquake Swarm Models

Earthquake swarms are of diverse nature; they are usually considered to be of two types: volcanic

and tectonic. Accordingly, several models have been proposed to explain the conditions and mechanisms leading to the formation of earthquake swarms. Mogi (1963) proposed their origination as consequence of stress concentration in unusually heterogeneous rock and considered magma intrusions to be the likely cause of stress accumulation. Hill (1977) proposed a model for the generation of earthquake swarms in volcanic environments, in which they occur on planes connecting the edges of parallel and offset inflating dikes. The magma intrusion causes dike inflation, which increases the stress difference $\sigma_1 - \sigma_3$ (σ_1 is maximum and σ_3 minimum local stress) and drives slip on the shear planes. Another model based on mechanical coupling between the rupture process and fluid migration was proposed by Yamashita (1999). Here, numerous pores are created due to the rupture and this allows fluids to migrate from the high-pressure compartments into the ruptured zone. Swarm earthquakes are generated if the rate of the pore creation is high, and consequently, the pore volume rapidly increases. Since the pores are not filled with fluids during their creation, fluids flow into the fracture zone causing a pore pressure decrease with time and ceasing the rupture growth. But the consecutive rise of the pore pressure reactivates the rupture process and the swarm activity goes on. In such a case, the seismic b -value (basic parameter of the frequency–magnitude distribution; for more details, see section “[Statistical Characteristics of Swarm Earthquakes](#)”) is unusually large, which corresponds to the high b -values observed in volcanic earthquake swarms. In case the rate of the pore creation is small or there are no new pores at all, the model provides the foreshock-mainshock sequences and small b -values. A prerequisite for Yamashita's model are continuously open faults, which serve for the fluid supply.

Local Networks: A Tool for Understanding Earthquake Swarms

All the models mentioned above imply that earthquake swarms are phenomena which occur in diverse tectonic environments. Their underlying

causes are not easy to disclose. For the same reasons, earthquake swarms are defined rather empirically than deterministically and their triggering mechanisms and driving forces are also unclear. A promising way to get a deeper insight into their driving mechanisms is a detailed analysis of swarm seismicity including the smallest possible events, over a long time period. A prerequisite for this is high-quality data from long-term observations by means of a dense network of local seismic stations providing a sufficiently low completeness magnitude ($M_L < 0$ as an optimum). Among others, similar networks are operating in the W-Bohemia earthquake swarm region (WEBNET consisting of 22 stations), in Iceland (SIL-network of about 70 stations), and in Yellowstone National Park (Yellowstone seismic network consisting of 26 stations).

Statistical Characteristics of Swarm Earthquakes

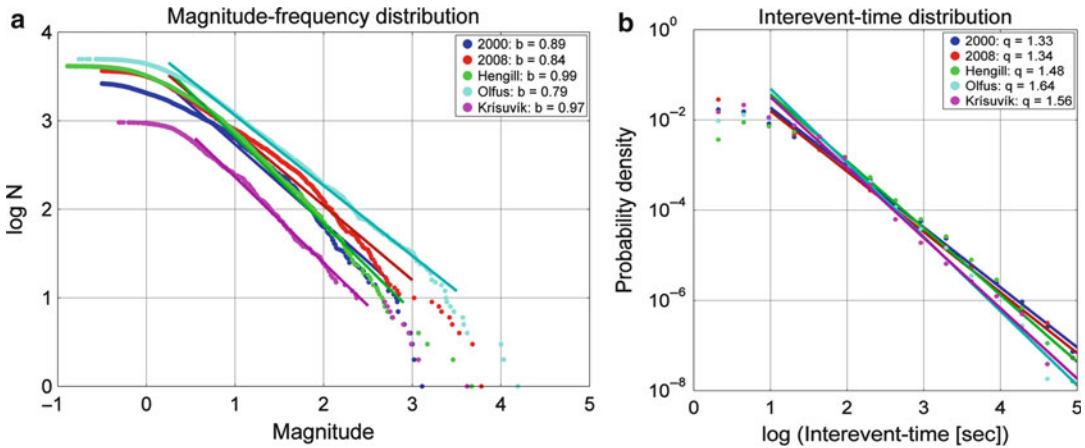
The Gutenberg–Richter law, $\log N = a - bM$, where N is the number of events having a magnitude $\geq M$ and a and b (b -value) are constants, expresses the relation between earthquake size and the frequency of their occurrence in a particular region and time period, hence it is the most important scaling relationship in the earthquake statistics. The constant b signifies the rate of smaller to larger events. In general b -values for ordinary mainshock–aftershock sequences as well as for seismicity in particular region and given time period typically range from 0.8 to 1.0. The b -value of 1.0 implies that there will be a tenfold increase in the number of events for every unit drop in magnitude. It means that distribution of earthquakes with respect to magnitudes is scale invariant, appears to be self-similar and obeys a power law. Computation of the magnitude–frequency distribution is easy to do but its interpretation may be tricky. Reliable evaluation of b -value requires a complete data set over at least three orders of magnitudes; an incompleteness of the analysed data may result in false b -value.

It is generally assumed and often claimed in literature that b -values of earthquake swarms

typically exceed 1.0 and are often as high as 2.5 (e.g., Lay and Wallace 1995), which implies prevalence of small events against larger ones in individual swarm activities. However, relevant studies show that b -values of both tectonic and volcanic swarm activities vary frequently between common values of 0.7–1.5. The b -values ≈ 0.85 –1.0 are characteristic for intraplate earthquake swarms in W-Bohemia/Vogtland, similar b -values were found for several earthquake swarms on the Reykjanes Peninsula and in the Hengill volcanic complex in SW-Iceland (Fig. 1). Also some volcanic earthquake swarms show b -value < 1.0 , e.g., the earthquake swarm prior to the 1995 eruption of Mt Ruapehu in New Zealand (Hurst and McGinty 1999) having b -values ≈ 0.75 . Note that the b -value close to 1.0 indicate that the earthquake swarm concerned is a mixture of mainshock–aftershock sequences that are largely controlled by a local tectonic stress.

The b -values > 1.0 are also observed quite often, particularly for volcanic swarms. For example, b -values of the Yellowstone earthquake swarms range from 0.6 to 1.5 (Farrel et al. 2009), b -value ≈ 1.2 are typical for swarms in Long Valley Caldera and the adjacent Sierra Nevada (Hill et al. 2003). The $M_{L\max} = 4.6$ earthquake-swarm activity associated with an eruption of El Hierro (Canary Islands) showed b -value ≈ 1.2 , however, the early pre-eruptive activity was characterized by a b -value of 2.25 (Ibáñez et al. 2012). Those b -values > 1.0 occur in nonvolcanic areas such as the Ubaye swarm (French Alps) in 2003–2004, $b = 1.2$ (Jenatton et al. 2007) or the Sampeyre swarm (Western Alps, Italy) in 2010, $b = 1.4$ (Godano et al. 2013). Extremely high b -values of 1.5–2.0 have been indicated for earthquake swarms in circum-Pacific subduction zones (Holtkamp and Brudzinski 2011) and an even higher value of 2.3 was determined by Einarsson (1986) for divergent segments of the Mid-Atlantic Ridge (see also section “Earthquake Swarms Induced by Magmatic and Volcanic Activity”).

For temporal evolution of seismicity at plate boundaries, it holds that the aftershock rate decays with the time from the mainshock,



Earthquake Swarms, Fig. 1 Statistical characteristics of interplate earthquake swarms from three different areas in SW-Iceland: Hengill volcanic complex in 1997, $M_L \leq 4.5$, Ölfus area in 1998, $M_L \leq 5.1$, and Krísuvík area on the Reykjanes Peninsula in 2003, $M_L \leq 4.2$, and of intraplate W-Bohemian earthquake swarms in 2000, $M_L \leq 3.3$ and in 2008, $M_L \leq 3.8$, which were located on the same

fault segment (Fischer et al. 2014). (a) The magnitude–frequency distribution of all these swarms shows b -value between 0.8 and 1.0, which suggests an occurrence of the mainshock–aftershock sequences in these swarms. (b) The interevent-time probability distribution signifies a similarity of the event-rate (rapidity) of the individual swarms

according to the power law $N(t) \propto (t + c)^{-p}$, where t is the time after the mainshock, c is constant, and exponent $p \approx 1$ (modified Omori law). Yet no adequate law was proposed for earthquake swarms. The reason is simply the absence of a single mainshock; alternatively the swarm can be considered as a number of overlapping aftershock sequences in time and space (Fischer et al. 2014). To describe this behavior the epidemic type aftershock sequence model (ETAS) was proposed by Ogata (1988) which considers each earthquake being capable of generating aftershocks and its aftershock productivity increases with its magnitude.

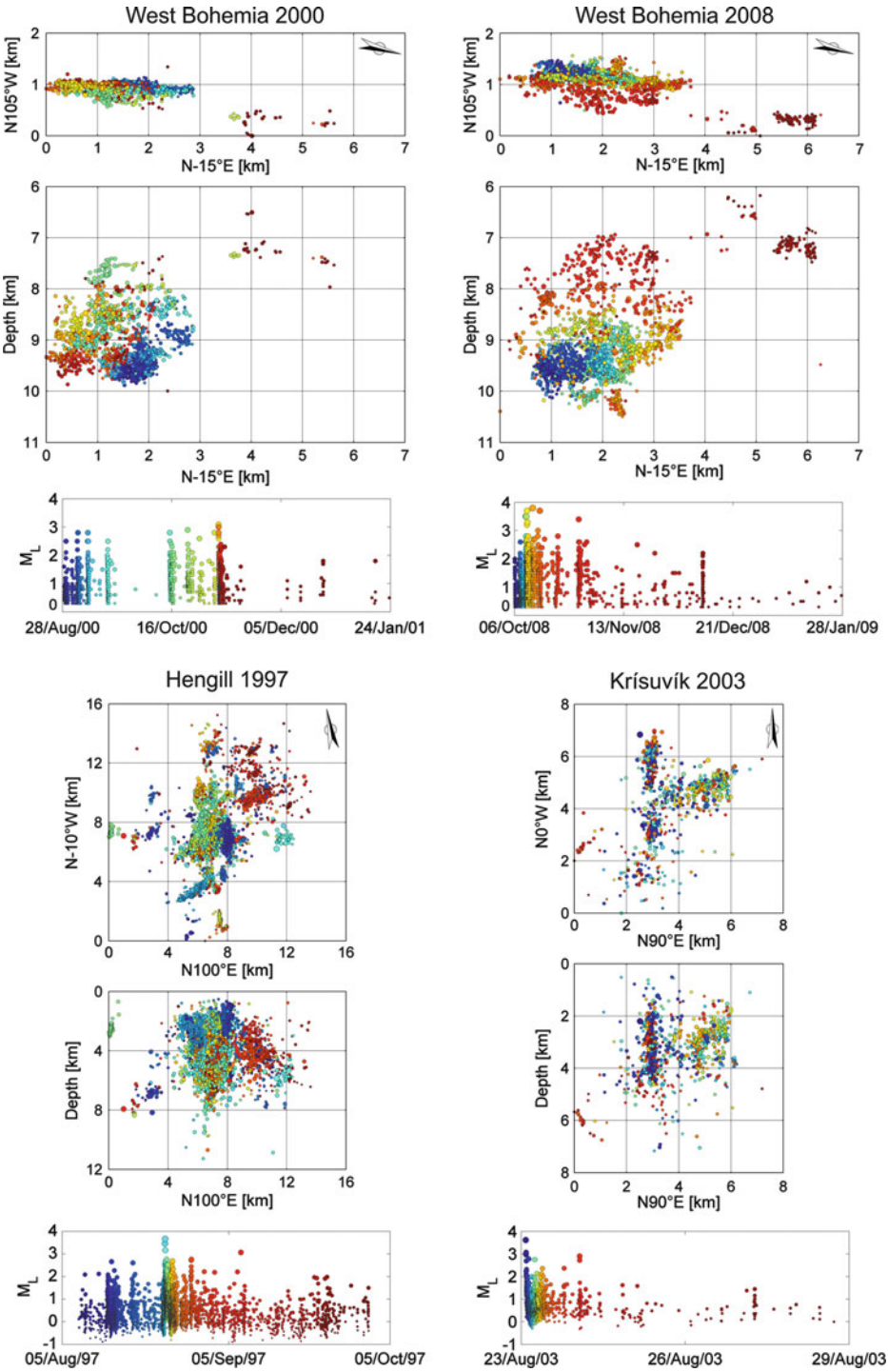
Space and Time Distribution of the Foci

Earthquake swarms frequently originate in the upper part of the Earth's crust. Depths of swarm earthquakes all over the world range approximately between 2 and 20 km, while deeper earthquake swarms exist but rather infrequently. The majority of swarm events are located around 10 km and shallower. For example, earthquake swarms in the Yellowstone volcanic field occur at depths between 2 and 12 km (Farrel et al. 2009),

similar depths are reported for earthquake swarms in continental rifts of the Rio Grande and Kenya (Ibs-von Seht et al. 2008). The hypocenters of the 2000 volcanic Miyakejima swarm in Japan were located in a large range of depths between 5 and 20 km (the strongest $M_w = 6.4$ earthquake occurred at depth of 10 km), while foci of the nonvolcanic Matushiro earthquake swarm of 1965–1967 in central Japan took place from near surface down to about 12 km.

Focal depths of the intraplate earthquake swarms in W-Bohemia/Vogtland vary from 5 to 20 km but those in the main focal zone Nový Kostel, where more than 35,000 $M_L < 4.0$ events were recorded during the last 25 years, range between 6 and 13 km. Interplate earthquake swarms in SW-Iceland – on the Reykjanes Peninsula, ($M_L \leq 4.2$), in the Hengill volcanic complex, ($M_L \leq 4.5$), and Ölfus area ($M_L \leq 5.1$) are a bit shallower compared to those in W-Bohemia/Vogtland (Fig. 2).

Extremely shallow earthquake swarms have been observed in the caldera of Hakone volcano in central Japan, the hypocenters of the 2009 swarm ($M_L \leq 3.2$) were located at depths between 0.5 and 2.8 km (Yukutake et al. 2011). Even shallower was the Tricastin earthquake swarm,



Earthquake Swarms, Fig. 2 Space and time distribution of earthquakes in the W-Bohemian earthquake swarms of 2000 a 2008 and the SW-Icelandic swarms in the Hengill volcanic complex in 1997 and in the K  suv  k

area (Reykjanes Peninsula) in 2003. The spatial distribution of foci for each earthquake swarm is represented by two plots: (*Top*) the map view, and (*Middle*) depth view along the strike. (*Bottom*) The time development of swarm

the Rhône Valley in France, ($M_L < 1.7$), in 2002–2003, that took place at depths from 200 m to 1 km (Thouvenot et al. 2009). A striking phenomenon is meteorologically triggered earthquake swarms at Mt. Hochstaufen in SE-Germany ($M_L \leq 2.4$), with depths from 100 m down to 2.5 km (Kraft et al. 2006).

Earthquake swarms in the lower crust, where ductile deformation is expected, are relatively rare but occasionally observed beneath active volcanoes. An example is the SW-margin of Long Valley Caldera in California (Shelly and Hill 2011), where both shallow brittle-failure swarms at depth down to ≈ 7 km and deep brittle-failures at depths between ≈ 18 and 30 km occurred. Another example is the 2007–2008 swarm beneath Mt. Upptyppingar (a part of the Iceland Northern Volcanic Zone) that was of the magmatic intrusion origin; the focal depths range between 10 and 20 km, see Fig. 3. (It is worthy of notice that this swarm activity was evidenced by a significant deformation, which was observed at permanent GPS stations in the region).

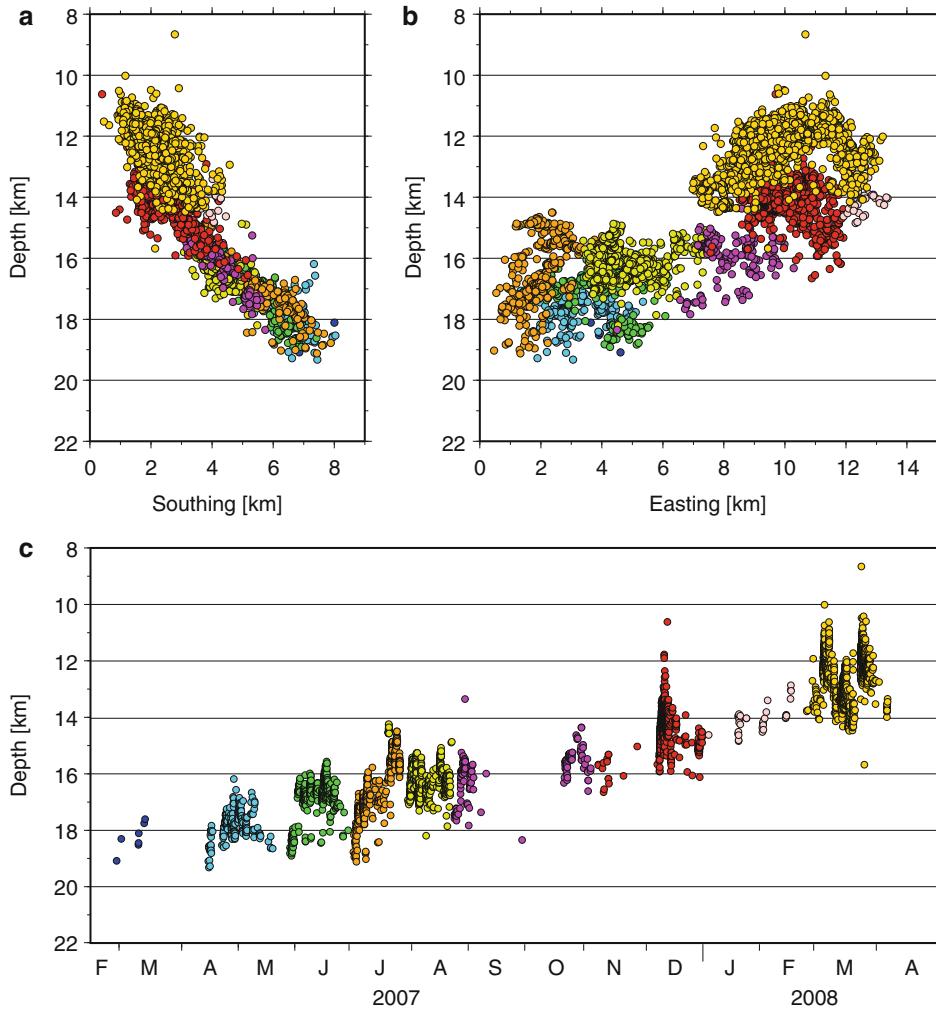
Earthquake swarms, regardless of the tectonic environment, are characterized by a dense clustering of the foci in time and space and in most cases by pronounced migration of the foci. Individual swarms usually initiate abruptly, event rate accelerates and magnitudes of events increase. The event rate may differ from case to case considerably, but quite often exceeds 100 $M_L > 0$ events per hour when swarm activity is culminating. Larger earthquake swarms are commonly comprised of several swarm phases, which are quite often separated by even longer periods of inactivity, during which the swarm activity shifts between different segments of the fault. Large events tend to occur in later swarm phases rather than in the beginning of the swarm.

Spatial clustering and focal migration of swarms are related to each other by their origin. In the beginning of a particular swarm and/or individual swarm phases, the foci cluster tightly in the vicinity of the activity nucleation point and gradually migrate outwards. The extent of the focal migration depends largely on switching the activity among several portions of the fault. Individual swarms show quite different migration patterns, even if they occur in the same place. Two W-Bohemian swarms of 2000 and 2008, which took place on the same fault segment (they represent “swarm twins”), are an informative example of an unusually different earthquake swarm development (Fischer et al. 2014). The former, with maximum magnitude $M_L = 3.3$, was of a distinctly episodic character: most of the events occurred during eight swarm phases with abrupt onsets lasting 2–4 days, which were followed by longer quiescence periods. The swarm activity culminated in the last phase during which all the $M_L > 3.0$ earthquakes occurred. The latter, with maximum magnitude $M_L = 3.8$ showed a much faster seismic-moment release, which resulted in the overlapping of the subsequent swarm phases and thus the 2008 swarm displayed only two distinct phases. Most of the seismic moment, which was more than three times larger than that in 2000, was released during the first 8 days. Both the 2000 and 2008 swarms initiated at the bottom of their focal clusters but they showed different patterns of focal migration (Fig. 2).

The focal-migration rate in earthquake swarms depends on the driving force. In general it differs for individual swarm areas and can be highly variable for a particular swarm. Extremely high migration rates between 0.5 and 1.2 m/s (approximately a common walking speed) are

Earthquake Swarms, Fig. 2 (continued) activity is depicted by the magnitude–time plot. Color coding from dark blue to dark red corresponds to the course of seismicity in time. The north direction is marked by arrow. Note that both W-Bohemia swarms, which happened on the same fault segment, initiated on the bottom of the focal cluster but their migration pattern is different. Further note that the 2003 K  isuv  k swarm began by the strongest

earthquake whereas in the remaining swarms the strongest events occurred during some later swarm phase. The events were localized by applying the double-difference method (HypoDD program package) to data of WEBNET for the West Bohemia swarms and of SIL for Icelandic swarms (section “Local Networks: A Tool for Understanding Earthquake Swarms”)



E

Earthquake Swarms, Fig. 3 A deep earthquake swarm beneath Mt. Upptyppingar, $M_L \leq 2.2$, in 2007–2008 that accompanied magmatic intrusion into a dike in the lower crust of Iceland’s Northern Volcanic Rift Zone (for more detailed information of earthquake swarms induced by magmatic activity see section “[Earthquake Swarms Induced by Magmatic and Volcanic Activity](#)”). The swarm exhibited extremely high b -value of 2.1. The spatial distribution of foci is shown in two depth sections: (a) view along the strike and (b) view across the strike. In (a)

and (b) the coordinate system is rotated 9° counter clockwise, so the easting is along the strike of the focal belt. It is obvious that the swarm is of planar character (the dip of the fault plane is $\approx 40^\circ$). The temporal development of the swarm activity and focal depth changes are depicted in plot (c). The time-axis is scaled in months. It is apparent that during an 11-month period the swarm activity migrated gradually upwards. The events are located by the double-difference method applied to the SIL data

observed in earthquake swarms that are closely connected with intrusion of dikes, such as in the rift zone of Krafla in Iceland or in the Afar rift in Ethiopia (Wright et al. 2012). Massin et al. (2013) reports a fairly high migration rate of 1 km per day for the Yellowstone Lake swarm in 2008–2009. The focal-migration rate of intensive

intraplate swarms amounts to several hundreds of meters per day. The fact that earthquake swarms mostly occur in volcanic or postvolcanic areas, which exhibit high fluid activity in crust, leads to a general assumption that swarming with the seismicity migration are governed by three possible processes: (a) magma migration within the

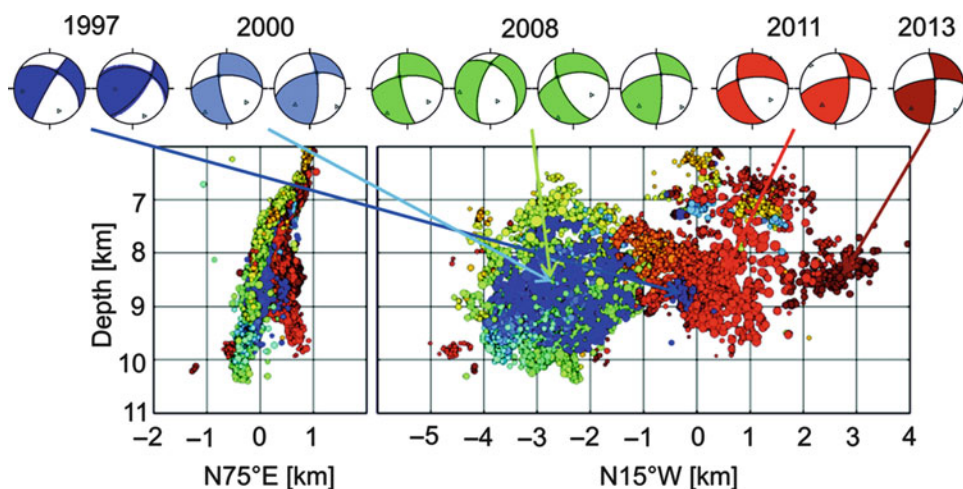
Earth's crust, (b) dike intrusion propagation (for more details see section “[Earthquake Swarms Induced by Magmatic and Volcanic Activity](#)”), or by (c) pressurized hydrothermal fluids diffusing within permeable fracture zone along the fault. Accordingly, Vidale and Shearer (2006) suggest that the fluid pressure fluctuations or aseismic slip as plausible driving mechanisms of the swarm-like sequences in southern California (section “[Occurrence of Earthquake Swarms in the World](#)”).

Source Mechanisms

Knowledge of source mechanisms is of crucial importance for understanding of rupturing of earthquake swarms and revealing the forces driving the swarm activity. The fact that earthquake swarms mostly occur in volcanic areas, ocean ridges, or in areas with high activity of crustal fluids begs the question of possible differences between forces which bring about and govern rupturing of swarm-like and ordinary earthquakes. The source mechanism is an equivalent of the body forces acting in the point source in an elastic medium, hence it provides the first information about rupturing process on the fault (Dreger, this issue). Its unconstrained mathematical representation is the seismic moment tensor which is in its completeness 9 dipole forces but only 6 of them are independent due to symmetry. The moment tensor is usually decomposed into three force systems (components): double-couple (DC) corresponding to shear slip on a fault, the isotropic (ISO) representing volumetric changes due to explosion or implosion, and compensated linear-vector dipole (CLVD), which has no physical interpretation (for more details see Vavryčuk 1, this issue). For example, the CLVD model may approximately represent an opening of cracks due to the magma injection. But in the CLVD model it is implicitly assumed that material flows rapidly into the opening crack from the nearby area so that no net volume change is involved. However, this flow may not occur very rapidly in viscous magma and on the short timescale (Minson et al. 2007).

Ordinary tectonic earthquakes are mostly shear slips along a fault, i.e., the slip direction is parallel to the fault plane, which is represented by the double-couple (DC) force equivalent, while the non-double-couple components (ISO and CLVD) are insignificant. However, earthquakes where crustal fluids act in the rupturing process need not be pure shear slips along a fault. If the fluid pressure is high enough, it can facilitate the slip by reducing the normal stress on the fault. The pressurized fluids may also cause fracturing by tensile cracks. Rupturing during which the shear slip is coupled with the simultaneous opening of fault results in a slip inclined to the fault plane by the slope angle α (so-called tensile earthquake, for more details see Vavryčuk 2011). Such earthquake is represented by a more complex force equivalent, which is manifested by significant non-DC components (both ISO and CLVD) of the source mechanism (for more information refer to Julian et al., this issue).

The DC component includes information about magnitude of the double-couple forces and focal mechanism (DC constraint of source mechanism). Focal mechanism is represented by three DC angles – strike, dip, and rake – which define geometry of the fault plane and the slip angle along the fault plane. It depends on local stress, i.e., the orientation of the principal stress axes σ_1 (maximum stress), σ_2 (intermediate stress), and σ_3 (minimum stress) relative to the fault plane. There are three basic faulting types: strike-slip (σ_1 and σ_3 are horizontal, σ_2 is vertical), normal (σ_2 and σ_3 are horizontal, σ_1 is vertical) and thrust (σ_1 and σ_2 are horizontal, σ_3 is vertical). Other focal mechanisms, the oblique normal and oblique thrust, are a combination of strike slip with normal and thrust type (for more details see Vavryčuk 2, this issue). In general, all types of focal mechanisms occur in earthquake swarms depending on the tectonic and local-stress conditions in a particular area. Nevertheless, when analyzing earthquake swarms in the world, it appears that oblique normal strike-slip faulting prevails. A characteristic feature of earthquake swarm is fairly significant spatial variation of the focal mechanisms in a particular swarm area as well as during individual swarms; very often each focal



Earthquake Swarms, Fig. 4 Distribution of the earthquake swarms in the main focal area of W-Bohemia/Vogtland and typical source mechanisms in the individual swarms. The section along the strike is depicted on the *right*, the cross-section is on the *left*. It is obvious that the focal area consists of a few shorter faults. All but one source mechanisms are pure shears; the DC angles fit the geometry of the corresponding fault segments. The

mechanism, which is depicted as the second from the left, prevailed in the second phase of the 1997 swarm. It contains significant non-DC components (both ISO and CLVD), the pattern indicates tensile earthquakes. The events are located by the same way as in Fig. 2. The source mechanisms are retrieved in the full moment-tensor description by inversion of amplitudes of the direct P- and S-waves from stations of the WEBNET network

cluster is characterized by its own focal mechanism family (see Fig. 4). It points out a rather complex system of differently oriented shorter faults with stress heterogeneities at small spatial scale, which is very likely one of the prerequisites for the strain-energy release in the form of earthquake swarms.

Existence of tensile swarm earthquakes is still a question at issue. The percentage of the non-DC part in the moment tensor of tensile events is fairly sensitive to the deviation of the slip vector, for example, a deviation of $\alpha = 5^\circ$ out of the fault plane generates up to 20 % of the non-DC components (Fig. 2 of Vavryčuk 2001). However, even large non-DC components can be an artifact of the inconsistency of data, which are inverted for moment tensor. Typically, insufficient distribution of sensors around the focus gives rise to apparently large fictitious non-DC components, particularly of CLVD. Further causes can be inaccurate or noisy data used for the moment-tensor inversion, inaccurate localization of the hypocenter, or a poor seismic-velocity model of the area under study.

The non-DC earthquakes have been identified at a number of hydrothermal and magmatic centers, e.g., in Long Valley Caldera (Foulger et al. 2004) or Hengill volcanic complex in SW-Iceland (Julian et al. 1997), as well as in the intraplate area of W-Bohemia/Vogtland (Horálek et al. 2002). However, there are only few papers dealing with a detailed analysis of non-DC earthquakes in individual earthquake swarm areas or in a particular swarm. So an existence of tensile swarm earthquakes and their possible connection to magma intrusion or to fluid transport is still unclear. This can be demonstrated by the following examples: (1) The Long Valley Caldera, which is still volcanically active and hosts an extensive hydrothermal system. The source mechanisms of swarm events from there were investigated by several teams of authors, however, the results appear ambiguous. For example, Stroujkova and Malin (2002) analyzed the mid-1997 swarm and arrived to the conclusion that majority of the resulting mechanisms appear to be of the DC nature, while Foulger et al. (2004) reported the occurrence of significantly non-DC

source mechanisms, inconsistent with simple shear faulting, for selected events of this swarm. Afterwards, Templeton and Dreger (2006) showed that fluid-influenced earthquakes are quite rare in the Long Valley volcanic region and that majority of local earthquakes are best characterized by a simple DC source model. (2) W-Bohemia/Vogtland, a European intraplate area with extinct Quaternary volcanism manifests present geodynamic unrest by persistent swarm seismicity and degassing of big amount of CO₂ (of about 500 m³ per hour) of the upper-mantle origin. The moment tensors of earthquake swarms in the 1997, 2000 and 2008 events signify fairly different amounts of non-DC components of respective source mechanisms. It appears that swarm earthquakes, which occurred on the optimally oriented faults relative to the local tectonic stress, are pure shears. It is the case of both 2000 and 2008 swarms in which the faulting was governed merely by local tectonic stress (Horálek and Šílený 2013). A question is the 1997 swarm in which two types of source mechanisms occurred (Horálek et al. 2002): pure shears in the first swarm phase but significantly non-DC mechanisms in the second swarm phase (thrust-mechanism type colored dark blue in Fig. 4), which may indicate the occurrence of tensile earthquakes on a less favorably oriented fault segment. However, one cannot rule out that the non-DC components are due to the small number of stations which were used in the moment-tensor inversion of the 1997-events. (3) Minson et al. (2007) tried to find non-double-couple mechanisms of selected earthquakes in a strong Miyakejima volcanic swarm in Japan during the year 2000 and concluded that the source mechanisms of the most analyzed earthquakes appear to be related to inflation of the dike yet their source mechanism solutions are ambiguous.

Possible Triggering Mechanisms and Driving Forces

As far as the mainshock-aftershock sequences at plate boundaries are concerned, there is a general understanding that the main shock is triggered by

plate motion loading and aftershocks are a result of stress redistribution due to the main shock. Yet there is no common knowledge how it works in the case of earthquake swarms. In the case of earthquake swarms it appears useful to distinguish the triggering mechanisms that lead to the initiation of the activity and driving forces that keep the activity running.

Compared to tectonic earthquakes at the plate boundaries, the intraplate earthquake swarm areas miss the plate motion that could periodically bring the fault to failure. Recent studies have tried to understand the possible triggering mechanisms by probing/imaging anomalous crustal structures within seismic swarms linked to fluids (e.g., Lin and Shearer 2009; Kato et al. 2010). It is assumed that a fluid overpressure builds up and acts as a triggering mechanism of earthquake swarms. The subsequent swarm activity is then probably driven both by the fluid flow along the fault zone and stress transfer from previous earthquakes. This idea is based on characteristic features of earthquake swarms that help us to disclose their background mechanisms.

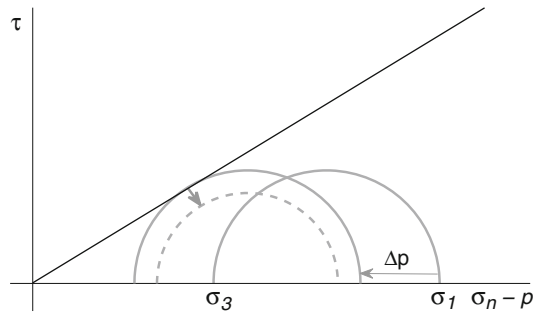
One of the most striking features of earthquake swarms is their focal migration, which is often attributed to the fluid flow along a quasipermeable fault zone. The most frequent model of this driving force is based on diffusive propagation of the pore-pressure front from a fluid source within a poroelastic medium defined in a simplified way by the formula $r = \sqrt{4\pi Dt}$ (Shapiro et al. 1997). Here, r and t are the distance and time interval of the foci from the first event, which corresponds to the source of fluids, and D is the hydraulic diffusivity. A different fluid-related model employs the preferential upwards migration of the hypocenters, which is attributed to the buoyancy of fluids that results in hydraulic fracturing of the rock. This allows for estimation of overpressure responsible for the swarm triggering which Hainzl et al. (2012) obtained an overpressure of 20 MPa for the W-Bohemian swarm in 2008. Daniel et al. (2011) analyzed the effective stress necessary to drive the 2003 swarm in Ubaye, France, and obtained a maximum excess pore pressure of 8 MPa.

Clustering of earthquakes in space and time (see Fig. 2) cannot be explained by fluid flows that appear too slow to be responsible for fast earthquake interactions over the first seconds and distances of hundreds of meters. That is why a similar mechanism that drives aftershock series, the static and dynamic Coulomb stress change due to previous earthquakes, is considered. The point is that the stress released within the rupture area of an earthquake brings about the stress increase in some areas outside the rupture. This results in loading of the adjacent ruptures and their subsequent rupturing. A similar mechanism was considered for example by Hainzl (2004) who used numerical simulation based on fluid diffusion and stress transfer for modeling the activity of the 2000 swarm in W-Bohemia. He found that while the hypocenter spreading is explainable by fluid diffusion, coseismic and postseismic stress transfer is necessary to explain the temporal clustering. Fast stress interactions between consecutive earthquakes, which prevailed in the direction of seismic slip are also documented by Fischer and Horálek (2005).

The repeated loading of the fault plane by successive pressure increase is illustrated in Fig. 5. Here the effective stress decreases due to fluid pressure increase until the shear stress exceeds the failure envelope, which results in shearing and decrease of the differential stress represented by the diameter of the Mohr circle. Further increase of pore pressure leads to repeated failure.

Earthquake Swarms Induced by Magmatic and Volcanic Activity

A general relationship between swarm activity and volcanic regions has been recognized for a long time, but yet, the causal relationship has not been clear, i.e., whether magmatic injection is a direct driving force of the swarm or whether the strain release in an earthquake swarm is governed by the inhomogeneous crustal structure and stress field (possibly influenced by pressurized hydrothermal fluids) in a given region. Recent case histories from Iceland revealed that earthquake



Earthquake Swarms, Fig. 5 Mohr–Coulomb diagram illustrating the reactivation of earthquake slip due to increase of the pore pressure Δp . The failure envelope, which is indicated by the inclined line, shows the failure shear stress $\tau_f = \mu(\sigma_n - p)$, where μ is the friction coefficient. The increasing pore pressure results in decrease of the effective stress ($\sigma_n - p$), this way reaching the failure strength. Shear stress is released by the seismic slip, which is expressed by decreasing the stress difference corresponding to the diameter $\sigma_1 - \sigma_3$ of the Mohr circle

swarms, which are directly a consequence of magmatic intrusion into the crust, are accompanied by more extensive strain (both in magnitude and spatial extent) than earthquakes that are caused by fracturing in a crustal stress field (Pedersen et al. 2007). The ratio of total seismic moment of the swarm earthquakes to the volumetric moment of the episode, measured geodetically, is a sensitive parameter that distinguishes between the swarm induced tectonically and that caused by a magmatic intrusion and volume change. A distinction can be made between at least three types of intrusive activity leading to swarm-like earthquake activities:

1. Inflation of a crustal magma chamber. A preexisting molten batch of magma will inflate if injected by fresh magma. The increasing pressure sets up a stress field in the surrounding crust that will eventually lead to fracturing and earthquakes, mainly in the volume above the chamber. The rate of the seismic moment release is governed by the rate of flow into the chamber. This behavior has been documented for the inflation periods of the Krafla volcano in 1975–1988; the Grímsvötn volcano in 1983, 1998, 2004, and 2011; and the Hrómundartindur volcanic

system in 1994–1998, in Iceland (e.g., Pederesen et al. 2007, and references therein). These periods may last for days, to months, to even a few years. The largest swarm earthquakes may be between magnitude $M = 4$ and 5.

2. Intrusion of dikes or sills into the crust. If increasing magma pressure fractures the surrounding crust a dike propagates away from the magma source. The dike tip changes the local stress and thus induces fracturing and earthquakes. The rate of propagation depends on the regional stress, the width of the dike, and viscosity of the magma. Large variation in the waveforms is one of the characteristics of earthquake swarms of this origin, especially if the fracture front approaches the surface. Propagating swarms of this type have been documented in the Afar rift in Ethiopia (Wright et al. 2012), in the rift zone of Krafla, and Bárðarbunga volcanic system in Iceland, where an intense seismic crisis has accompanied the large lava eruption that began in August 2014 (Sigmundsson et al. 2015). In all these cases the dikes are vertical and propagate horizontally away from the magma source. Sill propagation accompanied by earthquakes preceded the eruptions of Eyjafjallajökull in 2010.
3. Extrusion of lava domes at volcanoes, and its subsurface equivalent, the intrusion of cryptodomes, is frequently accompanied by earthquakes. Domes may have steady or episodic growth, with emplacement times ranging from a few hours to many decades. The accompanying earthquakes therefore show a swarm-like behavior on various time scales. They typically have the characteristics of B-type volcanic earthquakes (Chouet 1996), which have indistinct P-waves, weak or no S-waves, and a low-frequency coda.

Several cases are known of earthquake swarm activity accompanying caldera collapses. The latest case is that of the slow collapse of the Bárðarbunga caldera in 2014–2015 following dike propagation and a large lava eruption. More than 70 $M \geq 5.0$ earthquakes occurred on

the caldera fault in response to more than 60 m of caldera subsidence.

Ever since the discovery of mid-oceanic plate boundaries, it has been recognized that there is a fundamental difference between the seismic activity of transform boundaries and that of divergent boundaries, i.e., spreading centers. The transform-zone seismicity is characterized by mainshock-aftershock earthquake sequences having b -values of 0.8–1.0 and strike-slip faulting mechanisms, with maximum magnitudes $M \approx 6.5$ –7.0. The spreading-center earthquakes tend to occur in swarms, fault plane solutions show normal faulting, b -values are high, as high as 2.3, and the largest earthquakes rarely exceed magnitude $M = 6.0$ (e.g., Einarsson 1986).

The association here with an area of volcanic activity is obvious, but even with all of these mentioned studies the causal relationship is still debated. Experience from divergent boundaries on land, in Iceland and East Africa, appears to indicate that both explanations apply. Dike intrusions associated with rifting do produce earthquake swarms (Wright et al. 2012), but swarms that do not seem to be associated with magmatic activity are also common.

Conclusions

Earthquake swarms represent a specific type of seismicity without a mainshock, usually with a gradual increase of seismic activity, clustering in time and space, and migration of their hypocenters. In contrast to mainshock-aftershock sequences, earthquake swarms occur in various tectonic environments including both purely volcanic areas and areas with no active volcanism. Their recognition is still in progress and thanks to more detailed observations it turns out that they are more common than thought before. Their diverse behavior is expressed in multiple models suggesting their origin. Besides the models based on fluid injection and migration along faults, recent hypotheses suggest aseismic loading by creeping faults. Some recent observations show that earthquake swarms can share the same faults

with other types of seismicity. In terms of the wide range of their behavior it appears possible that volcanic swarms and mainshock-aftershock sequences represent end members of a continuous variety of seismic activity that is governed by a combination of two basic types of external loading: fluid injections on one end and tectonic loading on the other end.

Summary

Earthquake swarms are sequences of numerous events closely clustered in space and time and do not have a single dominant mainshock. A few of the largest events in a swarm reach similar magnitudes and usually occur throughout the course of the earthquake sequence. These attributes differentiate earthquake swarms from ordinary mainshock-aftershock sequences. Earthquake swarms occur worldwide, in diverse geological units. The swarms typically accompany volcanic activity at margins of the tectonic plate but also occur in intracontinental areas where strain from tectonic-plate movement is small, but large heterogeneities of stress and fault strength may exist. It implies that earthquake swarms generally originate from volcanic and tectonic episodes. The swarm earthquakes are usually of magnitudes $M < 5$, with some exceptional events reaching or exceeding $M \approx 6.0$. The b -value of the frequency–magnitude distribution $\log N = a - bM$ of both tectonic and volcanic swarms varies mostly between 0.7 and 1.5 but many earthquake swarms show $b \approx 1$ which is a typical b -value for mainshock-aftershock sequences. Depths of swarm earthquakes range approximately between 2 and 20 km with the majority of swarm events located around 10 km and shallower. Deeper earthquake swarms also exist but rarely occur. Earthquake swarms exhibit usually pronounced migration of the foci, particularly the swarms that are closely connected with intrusion of dikes. The source mechanisms, which are an equivalent of the body forces acting in the point-source approximation, signify that the swarm earthquakes are

mostly shear-slips along a fault just as ordinary tectonic earthquakes. In special cases of magma- or fluid-influenced earthquakes, more complex tensile rupturing might occur. However, the existence of tensile swarm earthquakes and their possible connection to movements of magma or fluids is still an open question. Also, triggering mechanisms and driving forces of earthquake swarms remain unclear. These mechanisms are mainly related to activity of pressurized hydrothermal fluids, magma injection, or dike intrusion, and also to aseismic loading by creeping faults. Accordingly, numerous models trying to explain the seismic-energy release in the form of earthquake swarms have been proposed.

Acknowledgments We are grateful to both anonymous reviewers for their valuable suggestions, which helped us to improve the paper substantially. Most of the work was accomplished within the Grant Project P210–12–2336 of the Grant Agency of the Czech Republic “Earthquake swarms and their triggering mechanisms and driving forces in diverse tectonic environments (Bohemian Massif, Mid-Atlantic ridge, French Alps).”

Cross-References

- [Earthquake Mechanisms and Stress Field](#)
- [Earthquake Mechanism Description and Inversion](#)
- [Moment Tensors: Decomposition and Visualization](#)
- [Non-Double-Couple Earthquakes](#)

References

- Cappa F, Rutqvist J, Yamamoto K (2009) Modeling crustal deformation and rupture processes related to upwelling of deep CO₂-rich fluids during the 1965–1967 Matsuhiro earthquake swarm in Japan. *J Geophys Res* 114, B10304. doi:10.1029/2009JB006398
- Chouet B (1996) Long-period volcano seismicity: its sources and use in eruption forecasting. *Nature* 380:309–316
- Daniel G, Prono E, Renard F, Thouvenot F, Hainzl S, Marsan D, Helmstetter A, Traversa P, Got JL, Jenatton L, Guiguet R (2011) Changes in effective

- stress during the 2003–2004 Ubaye seismic swarm, France. *J Geophys Res* 116, B01309. doi:10.1029/2010JB007551
- Einarsson P (1986) Seismicity along the eastern margin of the North American Plate. In: Vogt PR, Tucholke BE (eds) *The Western North Atlantic region*, vol M, *The Geology of North America*. Geological Society of America, Boulder, pp 99–116
- Farrel J, Husen S, Smith RB (2009) Earthquake swarm and *b*-value characterization of the Yellowstone volcano-tectonic system. *J Volcanol Geotherm Res* 188(1–3):260–276
- Fischer T, Horálek J (2005) Slip-generated patterns of swarm microearthquakes from West Bohemia/Vogtland (central Europe): evidence of their triggering mechanism? *J Geophys Res* 110:B05S21. doi:10.1029/2004JB003363
- Fischer T, Horálek J, Hrubcová P, Vavryčuk V, Bräuer K, Kämpf H (2014) Intra-continental earthquake swarms in West-Bohemia and Vogtland: a review. *Tectonophysics* 611:1–27. doi:10.1016/j.tecto.2013.11.001
- Foulger GR, Julian BR, Hill DP, Pitt AM, Malin P, Shalev E (2004) Non-double-couple microearthquakes at Long Valley caldera, California, provide evidence for hydraulic fracturing. *J Volcanol Geotherm Res* 132:45–71
- Godano M, Larroque C, Bertrand E, Courboux F, Deschamps A, Salichon J, Blaud-Guerry C, Fourteau L, Charléty J, Deshayes P (2013) The October–November 2010 earthquake swarm near Sampeyre (Piedmont region, Italy): a complex multicluster sequence. *Tectonophysics* 608:97–111. doi:10.1016/j.tecto.2013.10.010
- Hainzl S (2004) Seismicity pattern of earthquake swarms due to fluid intrusion and stress triggering. *Geophys J Int* 159:1090–1096. doi:10.1111/j.1365-246X.2004.02463.x
- Hainzl S, Fischer T, Dahm T (2012) Seismicity-based estimation of the driving fluid pressure in the case of swarm activity in Western Bohemia. *Geophys J Int* 191:271–281. doi:10.1111/j.1365-246X.2012.05610.x
- Hill DP (1977) A model for earthquake swarms. *J Geophys Res* 82(8):1347–1352
- Hill DP, Langbein JO, Prejan S (2003) Relations between seismicity and deformation during unrest in Long Valley Caldera, California, from 1995 through 1999. *J Volcanol Geotherm Res* 127(3–4):175–193
- Holtkamp SG, Brudzinski MR (2011) Earthquake swarms in circum-Pacific subduction zones. *Earth Planet Sci Lett* 305:215–225
- Horálek J, Šílený J (2013) Source mechanisms of the 2000 earthquake swarm in the West Bohemia/Vogtland region (Central Europe). *Geophys J Int* 194:979–999. doi:10.1093/gji/ggt295
- Horálek J, Šílený J, Fischer T (2002) Moment tensors of the January 1997 earthquake swarm in West Bohemia (Czech Republic): double-couple vs. non-double-couple events. *Tectonophysics* 356:65–85
- Hurst AW, McGinty PJ (1999) Earthquake swarms to the west of Mt Ruapehu preceding its 1995 eruption. *J Volcanol Geotherm Res* 90:19–28
- Ibáñez JM, De Angelis S, Díaz-Moreno A, Hernández P, Alguacil G, Posada A, Pérez N (2012) Insights into the 2011–2012 submarine eruption off the coast of El Hierro (Canary Islands, Spain) from statistical analyses of earthquake activity. *Geophys J Int* 191:659–670. doi:10.1111/j.1365-6X.2012.05629.x
- Ibs-von Seht M, Plenefisch T, Klinge K (2008) Earthquake swarms in continental rifts – a comparison of selected cases in America, Africa and Europe. *Tectonophysics* 452:66–77
- Jenatton L, Guiguet R, Thouvenot F, Daix N (2007) The 16,000-event 2003–2004 earthquake swarm in Ubaye (French Alps). *J Geophys Res* 112, B11304. doi:10.1029/2006JB004878
- Julian BR, Miller AD, Foulger GR (1997) Non-double-couple earthquake mechanisms at the Hengill-Grensdalur volcanic complex, Southwest Iceland. *Geophys Res Lett* 24:743–746
- Kato A, Sakai S, Iidaka T, Iwasaki T, Hirata N (2010) Non-volcanic seismic swarms triggered by circulating fluids and pressure fluctuations above a solidified diorite intrusion. *Geophys Res Lett* 37, L15302. doi:10.1029/2010GL043887
- Kraft T, Wassermann J, Schmedes E, Igel H (2006) Meteorological triggering of earthquake swarms at Mt. Hochstaufen, SE-Germany. *Tectonophysics* 424:245–258
- Lay T, Wallace TC (1995) *Modern global seismology*. Academic Press, Inc. San Diego
- Lin G, Shearer PM (2009) Evidence for waterfilled cracks in earthquake source regions. *Geophys Res Lett* 36, L17315. doi:10.1029/2009GL039098
- Massin F, Farrell J, Smith RB (2013) Repeating earthquakes in the Yellowstone volcanic field: implications for rupture dynamics, ground deformation, and migration in earthquake swarms. *J Volcanol Geotherm Res* 257:159–173
- Minson SE, Dreger DS, Bürgmann R, Kanamori H, Larson KM (2007) Seismically and geodetically determined nondouble-couple source mechanisms from the 2000 Miyakejima volcanic earthquake swarm. *J Geophys Res* 112, B10308. doi:10.1029/2006JB004847
- Mogi K (1963) Some discussion on aftershocks, foreshocks and earthquake swarms – the fracture of semi-infinite body caused by an inner stress origin and its relation to the earthquake phenomena. *Bull Earthquake Res Inst* 41:615–658
- Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *J Am Stat Assoc* 83:9–27
- Pedersen R, Sigmundsson F, Einarsson P (2007) Controlling factors on earthquake swarms associated with magmatic intrusions; constraints from Iceland. *J Volcanol Geotherm Res* 162:73–80

- Scholz CH (2002) The mechanics of earthquakes and faulting. Cambridge University Press, Cambridge
- Shapiro SA, Huenges E, Borm G (1997) Estimating the crust permeability from fluidinjection-induced seismic emission at the KTB site. *Geophys J Int* 131:F15–F18
- Shelly DR, Hill DP (2011) Migrating swarms of brittle-failure earthquakes in the lower crust beneath Mammoth Mountain, California. *Geophys Res Lett* 38, L20307. doi:10.1029/2011GL049336
- Sigmundsson F, Hooper A, Hreinsdóttir S, Vogfjörð KS, Ófeigsson BG, Heimissson ER, Dumont S, Parks M, Spaans K, Guðmundsson GB, Drouin V, Árnadóttir T, Jónsdóttir K, Guðmundsson MT, Högnadóttir T, Fridriksdóttir HM, Hensch M, Einarsson P, Magnússon E, Samsonov S, Brandsdóttir B, White RS, Ágústadóttir T, Greenfield T, Green RG, Hjartardóttir ÁR, Pedersen R, Bennett RA, Geirsson H, La Femina PC, Björnsson H, Pálsson F, Sturkell E, Been CJ, Möllhoff M, Braidon AK, Eibl EPS (2015) Segmented lateral dyke growth in a rifting event at Bárðarbunga volcanic system, Iceland. *Nature* 517:191–195. doi:10.1038/nature14111
- Stroujkova A, Malin PE (2002) Moment-tensor statistics of the 1997 Long Valley microearthquake swarm. *J Seismol* 6:69–85
- Templeton DC, Dreger DS (2006) Non-double-couple earthquakes in the Long Valley volcanic region. *Bull Seimol Soc Am* 96:69–79
- Thouvenot F, Jenatton L, Gratier J-P (2009) 200-m-deep earthquake swarm in Tricastin (lower Rhone Valley, France) accounts for noisy seismicity over past centuries. *Terra Nova* 21(3):203–210
- Vavryčuk V (2001) Inversion for parameters of tensile earthquakes. *J Geophys Res* 106(B8):16339–16355. doi:10.1029/2001JB000372
- Vavryčuk V (2011) Tensile earthquakes: theory, modeling, and inversion. *J Geophys Res* 116(B12), B12320. doi:10.1029/2011JB008770
- Vidale JE, Shearer PM (2006) A survey of 71 earthquake bursts across southern California: exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers. *J Geophys Res* 111, B05312. doi:10.1029/2005JB004034
- Wright TJ, Sigmundsson F, Ayele A, Belachew M, Brandsdóttir B, Calais E, Ebinger C, Einarsson P, Hamling I, Keir D, Lewi E, Pagli C, Pedersen R (2012) Geophysical constraints on the dynamics of spreading centres from rifting episodes on land. *Nat Geosci*. doi:10.1038/NGEO1428
- Yamashita T (1999) Pore creation due to fault slip in a fluid-permeated fault zone and its effect on seismicity: generation mechanism of earthquake swarm. *Pure Appl Geophys* 155:625–647
- Yukutake Y, Ito H, Honda R, Harada M, Tanada T, Yoshida A (2011) Fluid-induced swarm earthquake sequence revealed by precisely determined hypocenters and focal mechanisms in the 2009 activity at Hakone volcano, Japan. *J Geophys Res* 116, B04308. doi:10.1029/2010jb008036

Earthquakes and Their Socio-economic Consequences

Emily So¹ and Stephen Platt²

¹Department of Architecture, University of Cambridge, Cambridge, UK

²Cambridge Architectural Research Ltd, Cambridge, UK

Synonyms

Casualties; Consequences; Earthquake risk; Economic; Recovery; Social

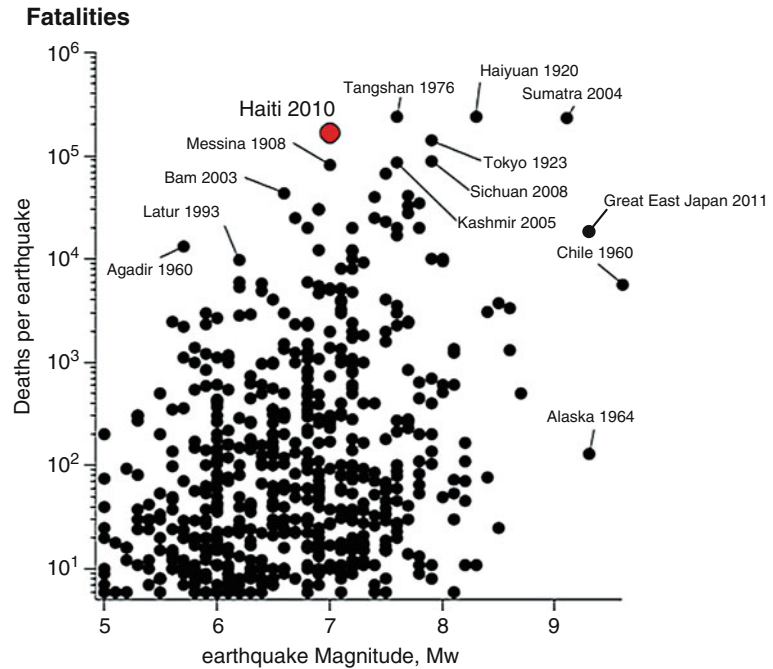
Introduction

One does not have to go far back in time to be reminded of the great force of Mother Nature and the havoc an earthquake can cause in terms of direct damage to the natural and built environment (Bam Earthquake in 2003, Kashmir Earthquake in 2005, Wenchuan Earthquake in 2008, Haiti Earthquake in 2010, Great East Japan Earthquake and Tsunami in 2011, Kobe Earthquake in 1995). Over the past decade, earthquakes have become costlier, both in terms of social and economic costs. Figure 1 shows a graph of earthquake fatalities. Even allowing for population growth during that period, of around 2 % per annum globally, it would be difficult say it is reducing. The current decade is the worst in the last 50 years.

In this chapter we shall first examine what are the socio-economic consequences of earthquake and then discuss briefly how socio-economic consequences are recorded and the difficulties in collecting and highlight the discrepancies in these data. This will be followed by a review of recent efforts to standardize the definitions of social and economic parameters. The causes and the contributing factors to differences in socio-economic consequences are also explored in section “[Characteristics of Socio-economic Consequences](#).” An assessment of the socio-economic

Earthquakes and Their Socio-economic Consequences,

Fig. 1 Earthquake fatalities as a function of earthquake magnitude. Post-2,000 events are named (Adapted from Bilham and Hough 2006)

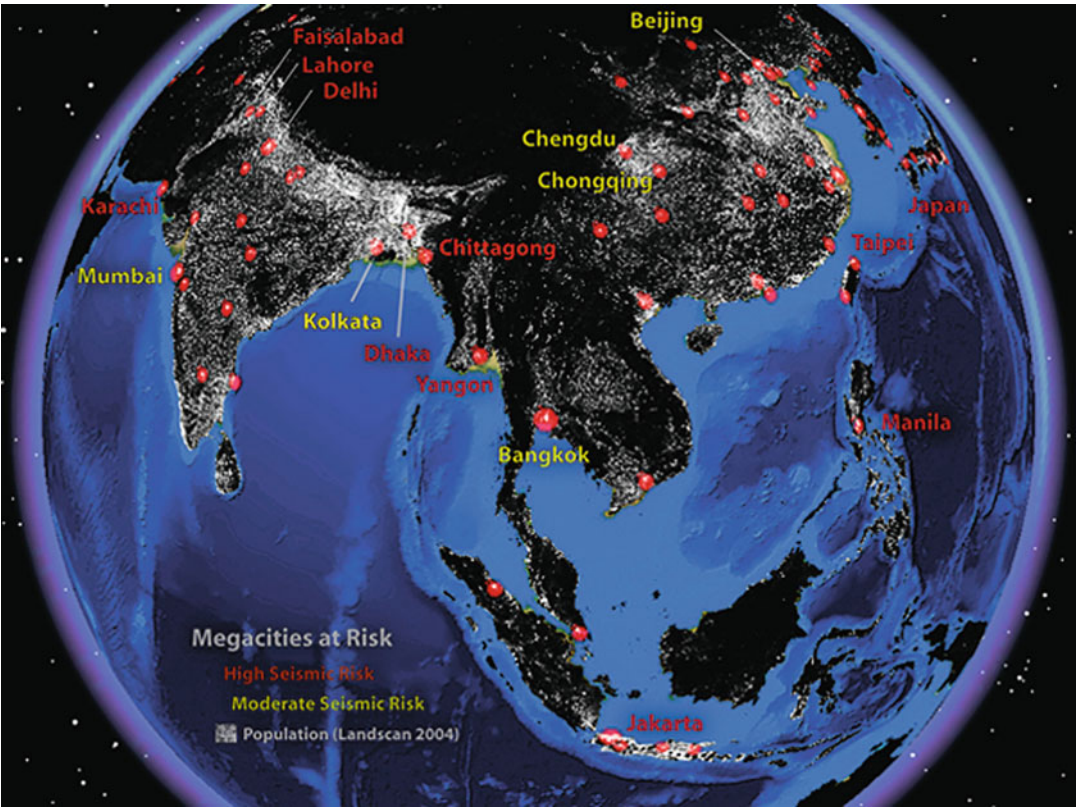


consequences of two case studies, the L'Aquila earthquake of 2009 in Italy and the more recent Great East Japan earthquake and tsunami of 2011, will be examined in detail in section “Two Events in the Twenty-First Century.” The main objective of this section is to not only compare and contrast these events but also to draw important conclusions on the factors affecting recovery and therefore long-term socio-economic impacts on societies affected by earthquakes. The intention is to draw this chapter to a close by looking to the future and examining what as a community of seismologists, engineers, social scientist, public health practitioners, economists, and politicians can be done to eliminate the prospect of a million-death earthquake (Musson 2012)?

Since 2000, more than 1,000,000 people have died from earthquakes and 2010 has seen a single event killing more than 200,000 people (Haiti 2010). Bilham suggested in the conclusions of his Mallet Milne lecture in 2009 that the global death toll from moderate earthquakes near human settlements is likely to average 8,000–10,000 a year in the next several decades and catastrophic events may bring this up to 50,000 deaths per

year. A direct “hit” of a $MW \geq 7$ earthquake in Tehran will cause more than a million deaths (Bilham 2009). As shown in Fig. 2, the number of megacities in this figure depicts cities with over a million people are aplenty and the seismic risk of some of these cities are high. For many cases traversed by active faults, this is due to what Jackson (2006) calls it the *fatal attraction*. For many centuries, an uneasy accommodation was reached between human needs and the earthquake-controlled landscape, sometimes brilliantly exploited by local hydrological engineering, as in Iran. In the past occasional moderate earthquakes would occur, killing a proportion of relatively small population. However, many once-small rural communities have now grown into towns, cities, or megacities while retaining their vulnerability through poor building standards.

As witnessed in recent years, earthquakes that occur in these places now kill many more than they did in the past and there are other cities like capitals as well, like Kathmandu in Nepal which is highly at risk. Extreme catastrophes have been rare only because the exposure of modern megacities to earthquake hazards has been relatively



Earthquakes and Their Socio-economic Consequences, Fig. 2 Population map of Asia showing the megacities at risk (Credit: GEM)

short (approx. 50 years); an increase in the number of such catastrophes now seems to be inevitable., e.g., Tehran, as shown in Fig. 3.

The most used parameter to measure the social impact or severity of an earthquake is fatalities, but in terms of social consequences, it is the survivors that need the humanitarian assistance in the short term and committed recovery and reconstruction plans to rebuild shattered lives in the long term. After the initial relief phase, the “fate” of those affected is the responsibility of those in governance, e.g., the establishment of the Canterbury Recovery Authority (CERA) after the Canterbury Sequence in 2010–2011. Possible indicators to help measure and monitor long-term social and economic recovery are highlighted in the section “[Recording Socio-economic Consequences.](#)”

Table 1 shows how earthquake severity is usually reported in terms of total losses, insured

loss, and fatalities. Apart from the two earthquakes in Japan (Kobe 1995; Tohoku 2011), there is an inverse relationship between property losses and casualties – high losses and low fatalities in developed counties and low insured losses and high fatalities in undeveloped or developing countries. The Great East Japan earthquake and tsunami as shown is the single most costly event since 1980, reporting overall losses of over US \$210 billion. This earthquake was exceptional in many ways, not least because of the high fatality rate in a highly developed exceptionally well prepared country. It is important to note here that with a better understanding of parameters like vulnerability and exposure, losses and casualties are now only loosely associated to physical variables such as magnitude.

More assessments on the economic and insured losses can be found in Chapter 29 of this volume.



Earthquakes and Their Socio-economic Consequences, Fig. 3 The city of Tehran with the active North Tehran Fault as a backdrop (Credit GEM)

Earthquakes are usually reported by insured losses and casualties, though the effects of an earthquake can be generally classified into the following categories:

1. Physical impacts
2. Social impacts (the impact on people)
3. Economic impacts (the impact on the wealth of an area)
4. Environmental impacts (the impact on the landscape)

These can be immediate or long-term consequences and most are entirely interlinked. For example, damage to a transportation system due to ground shaking in an earthquake can take years to repair and therefore contribute significantly to the business and economic losses in the area affected. In a study on transportation disruptions from the Northridge earthquake in 1995, it was reported that 43 % of the businesses surveyed attributed their losses at least partially to

transportation disruption (Boarnet 1995). In the future, disruption of electronic networks may also be crucial, as there is a huge dependency of businesses on ICT.

After the February 2011 earthquake in Christchurch, New Zealand, the majority of businesses in the center business district (CBD) had to relocate. In the week after the quake, people at the Canterbury Earthquake Recovery Authority (CERA) described how everyone was on the phone trying to find a place and negotiate a lease. In some cases this involved multiple locations. Inland Revenue staff, for example, are scattered in 300 different offices. Some firms have moved to residential or semi-industrial areas and others have relocated to other cities. Tourism declined and tertiary education was under pressure. A City Council survey suggests that over 60 % of businesses that moved out wanted to return to the city and the CBD was already under pressure from a number of peripheral shopping malls, though 51,000 people were

Earthquakes and Their Socio-economic Consequences, Table 1 Table of the most costly earthquakes since 1980 (MunichRe)

			Overall losses	Insured losses	Fatalities
Date	Event	Affected area	US\$ m, original values		
11.3.2011	Earthquake, tsunami	Japan: Honshu, Aomori, Tohoku; Miyagi, Sendai; Fukushima, Mito; Ibaraki; Tochigi, Utsunomiya	210,000	40,000	15,840
17.1.1994	Earthquake	USA: CA, Northridge, Los Angeles, San Fernando Valley, Ventura, Orange	44,000	15,300	61
22.2.2011	Earthquake	New Zealand: South Island, Canterbury, Christchurch, Lyttelton	16,000	13,000	185
27.2.2010	Earthquake, tsunami	Chile: Bió Bió, Concepción, Talcahuano, Coronel, Dichato, Chillán; Del Maule, Talca, Curicó	30,000	8,000	520
4.9.2010	Earthquake	New Zealand: Canterbury, Christchurch, Avonside, Omihi, Timaru, Kaiapoi, Lyttelton	6,500	5,000	
17.1.1995	Earthquake	Japan: Hyogo, Kobe, Osaka, Kyoto	100,000	3,000	6,430
29.5./3.6.2012	Earthquake	Emilia-Romagna, San Felice sul Panaro, Cavezzo, Rovereto di Novi, Carpi, Concordia, Bologna	16,000	1,600	18
26.12.2004	Earthquake, tsunami	Sri Lanka, Indonesia, Thailand, India, Bangladesh, Myanmar, Maldives, Malaysia	11,200	1,000	220,000
17.10.1989	Earthquake	USA: CA, Loma Prieta, Santa Cruz, San Francisco, Oakland Berkeley, Silicon Valley	10,000	960	68
13.6.2011	Earthquake	New Zealand: Canterbury, Christchurch, Lyttelton	2,000	800	1

© 2013 Münchener Rückversicherungs-Gesellschaft, Geo Risks Research, NatCatSERVICE. As at: March 2013

still working in the center. This real-life scenario is very reminiscent of Kates and Pijawka (1977) model of recovery from earthquakes.

The main actors helping businesses after the earthquakes of 2010–2011 were the Chambers of Commerce, Canterbury Development Corporation, Christchurch City Council and other local authorities, central government, and CERA. The City Council relaxed regulations allowing businesses to relocate to residential areas in the west and to new retail hubs. This however only aggravated the problem of attracting business back to the center. In Christchurch it is estimated that NZ\$20 billion will be invested in the city and 15,000 jobs created in construction (Fig. 4).

The overall vision for the city was to give investors confidence. However, some of the risks and problems the city still faces include a flight of capital in which property owners are

choosing not to reinvest in the city center or Christchurch; the scale of the demolition and the time taken to rebuild and owners of tall buildings that require demolition are holding the city to ransom for months and months. The last cordons to the CBD were lifted in June 30, 2013, 859 days after the earthquake. Whether Christchurch will recover financially from the earthquakes remains to be seen.

In stark contrast to pure financial costs are social and psychological consequences that are hard to quantify but the effects of which will be felt for a long time after the disaster. The inundation of the tsunami in schools in the East Coast of Japan in 2011 and the collapse of school buildings after the Wenchuan earthquake in China in 2008 killed an entire generation of school children (Fig. 5).

Effects are often ordered as primary and secondary impacts. Primary effects occur as a direct

Earthquakes and Their Socio-economic Consequences,

Fig. 4 Christchurch city center closed to the public (Taken by the authors, February 2012, a year after the second more devastating event)



Earthquakes and Their Socio-economic Consequences,

Fig. 5 Unclaimed children's satchels are left in a former gymnasium at Arahama School in Japan after the Great East Japan earthquake and tsunami in 2011 (Authors' own)



result of the ground shaking, e.g., buildings collapsing. Secondary effects occur as a result of the primary effects, e.g., tsunamis or fires due to ruptured gas pipelines. The table below summarizes the range of possible consequences from earthquakes (Table 2).

With urbanization, the world's population is increasing in hazardous areas and people are

living more densely. Rapid development has meant that even with the strongest of political and economic will, the pace of regeneration of building stock is outrunning the return periods of earthquakes (Bilham 2009). Perhaps the only way to engage with earthquake-prone communities in mitigation is through social and fiscal means, rather than purely physical interventions.

Earthquakes and Their Socio-economic Consequences, Table 2 Summary of possible consequences of earthquakes

	Physical impacts	Social impacts	Economic impacts	Environmental impacts
Short-term (immediate) impacts	Buildings, transportation, and infrastructure systems damaged as well as critical facilities	People may be killed or injured. Homes may be destroyed. Transport and communication links may be disrupted. Water pipes may burst and water supplies may be contaminated. Families may be separated	Shops and business may be destroyed. Looting may take place. The damage to transport and communication links can make trade difficult	The built landscape may be destroyed. Fires can spread due to gas pipe explosions. Fires can damage areas of woodland. Landslides may occur. Tsunamis may cause flooding in coastal areas
Long-term impacts	Relocation of villages and towns	Lack of healthcare provisions for the affected. People may have to be rehoused, sometimes in refugee camps. Psychological distress and chronic injuries. Family stress and disruption	Job losses and financial strain on the community for reconstruction may be significant. Investment in the area may be focused only on repairing the damage caused by the earthquake. Long-term fiscal impacts and business failures	Important natural and cultural landmarks may be lost

Characteristics of Socio-economic Consequences

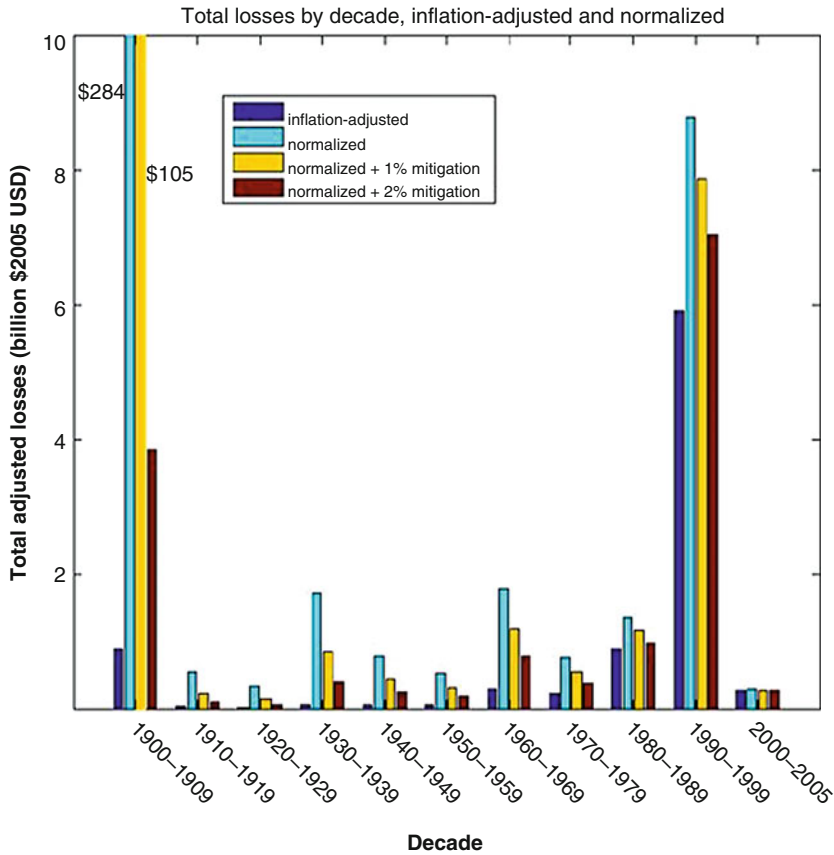
The direct, indirect, and long-term socio-economic consequences of earthquakes need to be carefully considered when devising disaster mitigation plans and strategies for urban development. The figures of losses from earthquakes shown in Table 1 raise an interesting question of how these parameters are calculated and recorded, for example, what do overall losses mean? How are other socio-economic measurements made? Decision makers charged with recovery need to understand the important socio-economic consequences of earthquakes.

What is the definition of a displaced person? In Alexander (1985), he discussed the difficulty of defining the dead and injured. What are the cutoff points of reporting? An important agenda for the Integrated Research on Disaster Risk (IRDR) organization, whose goal is “to develop trans-disciplinary, multi-sectorial alliances for in-depth practical disaster risk reduction research studies, and the implementation of effective evidence based disaster risk policies and practice,”

would be to improve the definition and accuracy of social and economic parameters so that options for investment in mitigation and disaster preparedness might be compared and contrasted

One of the organization’s priorities is to focus on the collation and standardization of loss data through a data working group. The ambition is to enable decision makers to make informed decisions on actions to reduce the impacts of natural disasters. The existing records and a review of global efforts, including that of the Global Earthquake Model (GEM), to standardize the collection of socio-economic consequences through the Earthquake Consequences Database (GEMECD) in collaboration with CRED are described in the section below.

Recording Socio-economic Consequences
Records of socio-economic consequences can be difficult to obtain and hard to compare. Individual case studies carried out by engineers and social scientists give details of the exact consequences of the event at the time of reporting, but these are seldom comparable with other events and therefore cannot be relied on for deriving specific



Earthquakes and Their Socio-economic Consequences, Fig. 6 Graph shows normalized damages in the USA 1900–2005 from Vranes and Pielke (2009) where the 1906 was found to be the most costly event in US history

mitigation strategies for other areas. The need for accurate and comparable numbers cannot be underestimated as they help a community of seismologists and engineers communicate risk to economists and politicians. What are the true costs of earthquakes? Socio-economic figures are meaningless unless placed in context and normalized in terms of population size and demography. For example, what was the absolute number of people killed, as a percentage of local population? What is the cost of the earthquake in today's terms? For example, a study by Vranes and Pielke (2009) on twentieth-century US earthquakes found that if normalized to 2005 dollars by adjusting for inflation, increased in wealth, and changes in population, the 1906 San Francisco earthquake would most likely to have been the most costly US earthquake, at

US\$328 billion, though some would still argue about the definition of "cost." Clear standardized definitions would help to reduce this ambiguity and inconsistency. This question could be answered partially by distinguishing between direct and indirect costs and short- and long-term costs. Some studies report the number of households affected, while others report the number of people. The reporting of social data is also dependent on the experience and training of interviewers and surveyors. Often samples are poorly constructed, interviewing is inconsistent, and questionnaire forms are badly designed (Fig. 6).

Knowing when essential services and livelihood are restored would capture some key aspects of recovery. But another issue with existing socio-economic data is that they lack

date stamping, for example, it is insufficient to know the number of homes without electricity or water supply or the number of homeless or unemployed; one also needs to know the time period of loss.

The indicator “number of people in homeless,” for example, is also time dependent. The term “homeless” can apply to any of the following:

- People taking shelter in an organized camp or in an organized shelter building (e.g., school, sports hall, etc.) – these are facilities providing short-term shelter, having communal kitchens and bathrooms.
- Number of people taking or given the option of alternative accommodation (with or without public subsidy), e.g., rent another home, rent a room in a hotel or hostel, and be hosted by relatives or friends or volunteers (usually after the first week following the event).
- Number of people provided temporary housing units (usually more than 1 month after the event) – these are usually prefabricated homes with minimal facilities intended for occupation no longer than 1–2 years.

In the simplest case, there should be two categories for the homelessness data: “number of homeless in the immediate aftermath” and “number of homeless for a significant duration of time.” This is a crucial differentiation. If not directly available these two categories of homelessness can be derived from damage statistics. Long-term homelessness would be related to the estimated population in destroyed houses or collapsed and red-tagged houses; homelessness in the immediate aftermath would be related to the estimated population in houses with serious structural damage that were deemed unsafe for immediate occupancy after post-earthquake damage/safety assessments, e.g., the number of yellow-tagged dwellings – depending on the damage reporting style in each event, country, and region. Long-term homelessness (usually more than 1 month) would imply people who have lost their homes due to ground shaking damage or due to secondary and induced hazards,

such as liquefaction, landslide, tsunami, or fire following, or because of abandonment/relocation of their neighborhood, village, or town (due to occurrence or high risk of occurrence of landslide, mudslide, rockfalls, liquefaction, fire following, tsunami, risk of dam collapse, nuclear disaster, other hazardous substance release, etc.), even if their homes were not destroyed.

An additional (and important) indicator, related to recovery, is the speed of reconstruction of the dwellings that were destroyed. Recovery can be expressed as a percentage of destroyed houses that have been reconstructed by a given time after the event, e.g., 1 or 2 years after the event (high-consequence events take several years until full recovery is achieved even in the most developed regions of the world).

An example of the recovery of the destroyed dwellings during the first 2 years after the September 7, 1999, Athens earthquake is given below:

- On the first anniversary of the earthquake, the Greek Ministry of Environment and Public Works released the following information: more than 50 % of the red-tagged housing units (i.e., housing units that are not habitable or usable in its current state) are being reconstructed (2,503 out of 4,220).
- On the second anniversary of the earthquake the Greek Ministry of Environment and Public Works released the following information: out of a total of 4,682 red-tagged buildings, 70 % had received reconstruction funding, and 400 (8.5 %) were part of a government-funded regeneration plan in the worst affected and poorest municipality of Ano Liosia and in Nea Philadelphia municipality (expected to be completed 42 months after the occurrence of the earthquake).

Records of socio-economic consequences help track progress of recovery through time that can be used as an accounting measure for donors and provide an insight into the possible issues contributing to delays in recovery of the economy (lack of housing, infrastructure, educational facilities, etc.).

Towards a Standardized Process of Recording and Reporting Socio-economic Consequences

GEMECD is the Global Earthquake Model (GEM) Earthquake Consequences Database. GEMECD is GIS relational database and is a repository of the most relevant and validated data on consequences of the significant events of the future and for the last 40 years around the world. Under the auspices of GEMECD and IRDR, the Centre for Research in Epidemiology and Disasters (CRED) has devised, together with MunichRe and CAR Ltd., a set of 45 key socio-economic indicators for earthquakes. The earthquake impact indicators are divided by type into social, social-recovery, and economic indicators. Generally indicators are mostly applied to entire affected area where impact has occurred, but in certain situations, the indicator may only refer to a reduced area because data is unavailable. Indicator specification, therefore, is based on an optimal situation. Moreover, indicators can be measured for the affected area or for the entire affected country. Which of these is used for a defined area is governed by the source of information.

Attempting to measure all 45 socio-economic indicators in any given post-disaster context is the desirable objective, but data availability constraints mean this is not always feasible. Quite often multiple sources of data are required to measure all indicators (e.g., reinsurance companies such as MunichRe or SwissRe, international organization situation reports such as produced from the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) or the International Federation of Red Cross and Red Crescent Societies (IFRC), Post-Disaster Needs Assessment (PDNA) case studies, such as the ones conducted by the World Bank (WB)). Consequently, this may lead to slight variations exist in estimates and totals.

The socio-economic consequences data for the 18 events in the GEMECD is the most comprehensive set of verified information to be stored in one single repository. The database sets a global standard for future earthquake consequences data collection with a standardized set of 45 indicators, which now also forms part of the IRDR loss data guidelines document (So et al. 2012).

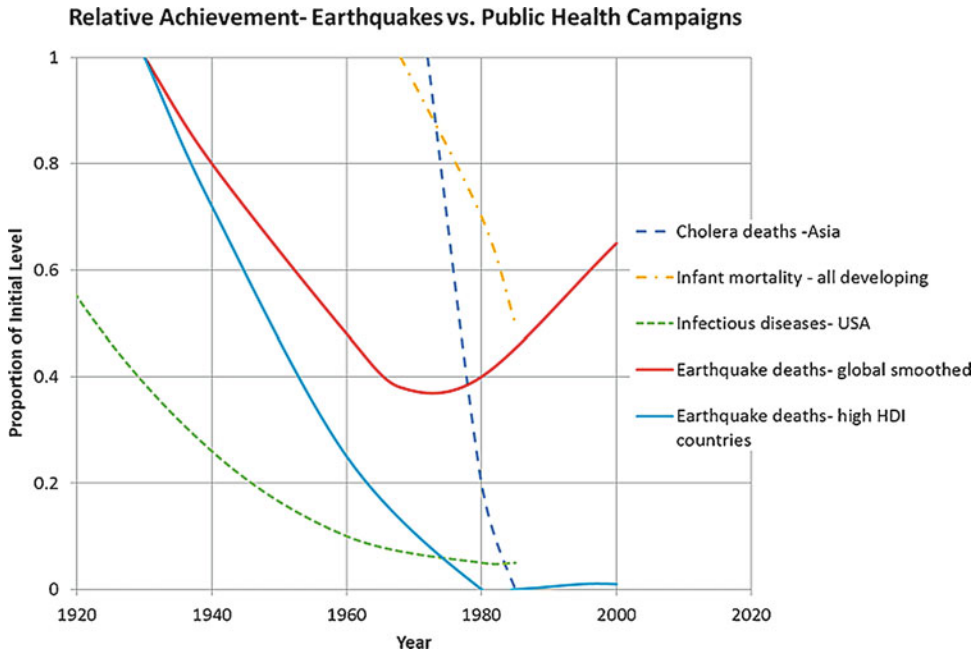
Causes of Socio-economic Consequences

Underlying all of the 45 indicators is the social status and structure of an affected community. The severity of social and economic consequences on a society and its ability to recover depends very much on the sociodemographic characteristics, perception of and experiences with earthquakes, and its overall capacity to respond.

In simplistic terms one can categorize socio-economic effects according to state of development of different countries – high-, medium-, and low-income countries. The chart below is taken from a study carried out by Spence (2007), comparing the relative progress made in the reduction of preventable deaths with public health campaigns and earthquake deaths in the same period. As shown, for high Human Development Index (HDI) countries, the line shows a relatively stable reduction of earthquake deaths from its initial levels, but globally, the red line shown suggests the community is not doing enough to prevent deaths in undeveloped and developing countries. The Human Development Index (HDI) is a composite statistic of life expectancy, education, and income indices used to rank countries into four tiers of human development (Fig. 7).

If one selects just some of the most significant earthquakes of the last two decades and reviews the number of fatalities and the direct economic costs (GEMECD 2014), one can immediately see the differences between low-HDI countries and high-HDI countries, in terms of, crudely put, cost per fatality. Obviously there are many gross assumptions made here, and there are many other factors such as governance, culture, and other social factors that come into play as well as the size of the event. The numbers are not directly comparable; direct economic losses are at the time of the event and not inflation adjusted and the HDI is taken in 2012, but a vague relationship does emerge. It costs more per fatality in high-HDI countries (Table 3).

The disproportionate effects of an earthquake on the poor, for example, are well researched and not at all surprising, as the socially vulnerable group reside in older, poorly built housing more prone to damage during an earthquake and the



Earthquakes and Their Socio-economic Consequences, Fig. 7 Graph comparing the progress made in reduction of deaths in earthquakes and other public health issues (up to 2000)

likely capital required to rebuild and recover is not readily available after the event. Broadly speaking one perhaps can venture to summarize the likely socio-economic consequences of earthquakes for different countries in the following way (Table 4).

Though perhaps the overall level of wealth of a country is not as fundamentally diagnostic of its earthquake vulnerability as is its differential between rich and poor.

The recent earthquake and tsunami in East Japan has also highlighted the need to anticipate the risk to the future population in earthquake-prone regions. About half of victims of this earthquake were elderly people of 65 years old or more, unable to respond to warnings and evacuate in time. Based on the statistics presented by Statistical Information Institute for Consulting and Analysis, depopulation has extended over the whole region, excluding large metropolitan areas such as Sendai. Forty percent of all municipalities will experience a population decrease of 20 % or more, accentuating the pre-disaster condition. In coastal areas of the Pacific Ocean, population is anticipated to decrease even further

since local economies were destroyed and young people will leave as a result (Masateru 2011).

The preexisting socio-economic trends in a region are likely to be amplified by a major disaster. Places that were shrinking before the disaster may be pushed into severe decline. Industries that were struggling to survive may fold and young people may move elsewhere in search of work. Conversely places that were booming before the disaster, or where the economic conditions were propitious, may experience a spurt in growth as opportunities open up and money flows into the area, e.g., Kobe, Japan. Much of the Tohoku region, for example, was in economic and demographic decline before 2011 and, despite the huge investment in the region, the earthquake has exacerbated the situation.

The risk of vulnerability and the impacts of disaster are disproportionately borne by those who are already socioeconomically and physically disadvantaged and who have fewer resources to enable them to “bounce back” to some measure of normality. These include foreign nationals, the very young and very old, those

Earthquakes and Their Socio-economic Consequences, Table 3 List of 10 significant earthquakes in the last 20 years and their socio-economic consequences (high HDI, 0.758; medium HDI, 0.64; low HDI, 0.466)

Earthquake	Fatalities	Direct economic losses (million US\$) ^a	“Cost”/ fatality (million US\$)	HDI	Lessons learned (from reconnaissance)
Northridge, Mw 6.7 (1994)	61	44,000	721	0.937	Severe damage to hitherto deemed safe steel-framed buildings
Kobe, Mw 6.9 (1995)	5,502 (fires)	96,000	17.4	0.912	Revised building code and nationwide retrofit Sch. of traditional timber residences
Kocaeli, Mw 7.4 (1999)	17,118	11,400	0.67	0.722	Building control called into question; new RC buildings performed worse than older traditional ones
Athens, Mw 5.9 (1999)	143	3,395	24.1	0.86	45 % of fatalities were in industrial buildings
Bhuj, Mw 7.7 (2001)	13,800	4,405	0.32	0.554	Site effects were significant on extremely weak masonry
Bam, Mw 6.6 (2003)	26,271	500	0.019	0.742	Extremely vulnerable dry mud brick buildings. 26 % of Bam died
Wenchuan, Mw 7.6 (2008)	>73,000	141,000	1.93	0.699	Bad siting of towns and schools killed many
Haiti, Mw 7.0 (2010)	>230,000	4,245	0.018	0.416	Extremely vulnerable substandard buildings
Christchurch, Mw 6.1 (2011)	185	16,000	86.5	0.919	Collapses of two mid-rise RC buildings accounted for 70 % of the deaths. People were killed on the streets by out-of-plane failure of masonry debris
Great East Japan, Mw 9.0 (2011)	>18,100 (tsunami)	211,753	11.69	0.912	Deaths from earthquake were ~150; tsunami heights were significantly above design allowances. Revisit nuclear facilities design

^aAll the data has been taken from the GEM Earthquake Consequences Database (GEMECD)

Earthquakes and Their Socio-economic Consequences, Table 4 Table relating HDI to socio-economic consequences of earthquakes

Consequences	High-HDI countries	Low-HDI countries
Casualties	Low	High
Property damage	Low	High
Cultural change	High	Low
Economic disruption	High	Low
Environmental damage	Moderate	High

living in poverty, ethnic minorities, the physically and mentally disabled, and women – especially those who are poor, pregnant, lactating, or elderly.

Observations on Gender

For some types of natural disaster, in particular earthquakes and tsunamis, there is evidence of a higher fatality rate in women than in men, and it seems that women may be more vulnerable than men to major unexpected hazards such as earthquakes, fires, tsunamis, or terrorist attacks. This vulnerability may occur because women have a slower reaction time than men or due to poor decision making. Alternatively, it may result from cultural factors, such as the way that women are subservient to or dependent upon men in so-called “traditional” societies.

Although there is a considerable body of evidence that building quality is the most significant factor in the number of fatalities in sudden-

impact disasters such as earthquakes, this fact alone does not account for gender differences in casualty rates. Because the gender division of labor keeps women in or close to homesteads, Byrne and Baden (1995) argued that it is generally they who suffer disproportionately from the collapse of poorly constructed dwellings.

A clear relationship has been found between the number of buildings that collapse in an earthquake and the number of fatalities caused. Mortality is therefore related to the chance of being in a poorly constructed building at the time of the disaster. Timing is clearly important. For example, there may be a higher death toll of both men and women if a disaster happens at night when people are asleep and a higher death toll among women if it happens during the day when men are away from the home. If death rates are higher among women, it could be because:

- Women are more likely than men to be caught in poorly built buildings, for example, by being more likely to be at home.
- Women are more likely to be caring for young children or elderly relatives who slow their evacuation.
- Women are likely to be older, since women live longer.
- In some societies, women tend to be less well educated and make poorer decisions in a crisis.
- Women less likely to evacuate rapidly, because they are more inhibited about leaving the home.
- Women are more likely to be caring for dependent children or elderly relatives.
- Again in some societies women are generally of lower economic and social status and have less access to resources.

It has been also suggested that women have a greater propensity to sacrifice themselves, for example, by rescuing young children or elderly relatives, rather than escaping on their own, but this has never been proven by field evidence. Alternatively, there may be gender differences in attitudes to risk-taking behavior that influence mortality.

Hatch and Dohrenwend (2007) also pointed to the lack of research evidence on the distribution of the events according to gender, age, racial and ethnic background, and socio-economic status (SES). They suggest that such information would help identify groups at greatest risk so that they could be investigated further. Although, theoretically, the impacts of a disaster are not randomly distributed across socio-economic classes, empirical evidence of this claim is scarce. In a population-based cohort study of the 1999 Taiwan earthquake, the multivariate results suggest that health status and socio-economic status (SES) are likely to be two key predictors of earthquake mortality. Indeed, lower SES, major disease status, and moderate physical disability status increased the risk of earthquake death 1.5– 2.5-fold (Chou et al. 2004). It is necessary to understand how socio-economic status and health interact with gender in such cases. In one of the greatest modern disasters, the Indian Ocean tsunami in 2004, more women than men died.

Women's and men's decisions in times of crisis have been shown to differ along risk-taking lines, nature of coping strategies, adaptability, and advice-taking and information-seeking behaviors (Dankelman et al. 2008). Hazeleger (2013) examined gendered disaster recovery issues relevant to Australian policy and also found differences in disaster risk taking as well as economic differences and inequalities that affected survival and recovery between men and women. This calls into question the assertion that high levels of economic development iron out the risk levels between men and women.

Two Events in the Twenty-First Century

In this section the socio-economic consequences of two real events – the L'Aquila earthquake of 2009 in Italy and the Great East Japan earthquake and tsunami of 2011 – are examined. These two earthquakes represent the most socially controversial and most costly events in the past decade. The intention here is not to generalize about how certain decisions pre- and post-event would help a

society recover more effectively. The purpose of this section is to describe the implications of earthquakes on two very different communities and in doing so draw the readers' attention to the complex long-term social and economic consequences of earthquakes. The consequences of earthquakes go far beyond the immediate aftermath of the event and can adversely or positively impact on future generations. Comparing the management of these two disasters and assessing the key overriding factors on recovery. What are the true costs of earthquakes and at whose expense? In reviewing the evidence on the recovery of the, four central themes were found namely:

1. Government plans for reconstruction
2. Socio-economic characteristics of the region
3. Citizen involvement in decision making
4. Politics and the media

The following sections are therefore organized accordingly.

Reflections on 2009 L'Aquila in Italy

The Mw 6.3 L'Aquila earthquake caused 308 deaths and nearly 1,200 injuries of which 202 were serious. This was the deadliest earthquake in Italy since the November 23, 1980, earthquake in the region of Campania and the most destructive earthquake in L'Aquila since the series of three earthquakes that devastated the wider region in January and February 1703 that killed at least 10,000 people. In 2009 203 people died in L'Aquila, a city of 73,000 inhabitants, and many more thousands are students of the L'Aquila University, not living in the city, while in the village of Onna, 38 out of around 350 inhabitants were killed. The hospital at L'Aquila, where many of the victims were brought, suffered damage in the 4.8 aftershock which occurred one hour after the main earthquake and had to be partly evacuated. Damage to L'Aquila's and wider regions' historic buildings and churches was also quite severe.

In total, almost 4,000 buildings collapsed or were damaged beyond repair, while another 70,000 buildings were damaged but repairable, and 65,000 were made homeless. In a report to

the EU, the Italian government estimated the cost of the earthquake at just over €10 billion while insured losses reached \$US250 million. Four months after the event, the Italian Civil Protection reported 49,000 people were still homeless.

Government Plans for Reconstruction

Unlike the management of past earthquakes in Italy, the central government under Berlusconi promised that the affected population would not be placed in transitional housing which usually came in the form of mobile trailers or prefabricated housing. The homeless were housed in camps around the affected area and in various hotels along the coast. The process of reconstruction was entirely led by the Italian Civil Protection Authority (DPC), and the aim was to reconstruct and rehouse the affected population within a year in 19 new towns. This was led entirely by the central government at a cost of €1.6 billion. The name of this project was the C. A.S.E. (Complessi Antisismici Sostenibili ed Ecocompatibili (Calvi and Spaziant 2009). According to the Italian government, nearly €8 billion has been spent on L'Aquila and the 56 surrounding communities that were affected in the quake (BBC 2013). This housing provision has sparked controversy though. The fact that all of these new towns are far away from the main city and are not accessible by public transport has been a source of major discontent (Alexander 2012), though the rehousing of the affected population has taken place very swiftly (Fig. 8).

Socio-economic Characteristics of the Region

L'Aquila is the largest city in the region and, with a GDP of €27,703.41 million, is the wealthiest region in Central Italy, hosting a university. The surrounding villages like Onna and Paganica are predominantly home to small businesses or second homes. The city's work force is mainly concentrated in public services such as defense, manufacturing, service industries, and the building industry. The death toll might have been much worse, if many more day visitors and students had been in the city, inside and outside vulnerable buildings.



Earthquakes and Their Socio-economic Consequences, Fig. 8 A C.A.S.E. town on the outskirts of L'Aquila

Abruzzo has traditionally relied on agriculture, but there has been some industrial growth especially in engineering, food, transportation, and telecommunications sectors in recent years; though by comparison with its northern neighbors, the Abruzzo region is not a major industrial center and the industry is on the coast, two hours' drive from L'Aquila. The fact that there were relatively little financial losses in terms of business interruption may be a key factor in the slow pace of overall recovery and reconstruction. Contrasting L'Aquila's situation is the recovery of the wealthy Emilia-Romagna region, one of the most industrialized areas of the country. Two major earthquakes Mw5.8–5.9 caused 27 deaths and more than €12 billion euros of damage in May 2012, seriously affecting small- and medium-sized companies. The earthquake disrupted more than 10,000 companies in the affected areas. There was an urgency to be back in business and private companies were the main drivers of the recovery of the region. The symbolic "solidarity cheese" fund brought in around €1 million. People all over Italy were encouraged to purchase the damaged wheels of Parmigiano-Reggiano and Grana Padano cheese, produced in and around the city of Parma, to help with the region's economic recovery.

Unlike other Northern Italian cities, the charming historic center of L'Aquila does not attract many tourists, probably due to its geographical location. Despite the initial heroic remediation measures, the historic center still lies derelict with scaffolds supporting the structures, but little else is being reconstructed or considered. All attention was focused on building new settlements, and perhaps L'Aquila's reconstruction was never the major focus of attention from the start since the town center does not have the social clout to demand the attention of the central government. In a recent article by the BBC (2013), it was reported that some 22,000 of the 70,000 people of the city who were made homeless are still unable to return to their homes and are living in new blocks of flats on the outskirts. Some 70 % of the reconstruction needed on the outskirts has been completed, but the thirteenth-century town center remains crippled, guarded by the military who only allow those with authorization to enter.

What will become of L'Aquila's historic center? 4 years on and the center is still littered with unusable buildings; commercial activity is much reduced and public services deteriorated; it is hard to imagine how its fortunes can be reversed (Fig. 9).

Earthquakes and Their Socio-economic Consequences,

Fig. 9 L'Aquila city center 2012 – abandoned and cordoned off, the medieval buildings propped by scaffolding (BBC 2013)



Citizen Involvement in Decision Making

There has been a lot of criticism of the authorities handling of the reconstruction process after the L'Aquila earthquake. A recent report suggested that there had been no public participation in the reconstruction planning of the new towns around L'Aquila. A top-down approach was used under the pretext of a more rapid reconstruction strategy and recovery of the region. However "Once people have been housed in accommodation that will last for years, they can be forgotten, or at least left to better their circumstances without further state intervention" (Özdem and Jacoby 2006). The decision to create the surrounding new settlements seems to lack long-term vision and will inevitably change the residents' daily lives. As recounted in an interview for the BBC:

I miss my former life, the neighbours I grew up with, the woman who lived upstairs who used to treat me like a daughter . . . I want the L'Aquila we all know to return and be even better, I want people to come back. (L'Aquila_) is an oasis among dozens of empty, damaged buildings," says Mariella Riccobono, who now lives 10 min outside the town. Wiping a tear from her cheek, she says: "It's worth fighting for your home. If you give up, you're abandoning the city. We've continued paying a mortgage all this time and it's my livelihood, it's everything I've created.

Politics and the Media

The media and politics were never far from L'Aquila immediately following the earthquake

in April 2009, with the Prime Minister's 23 visits to the region and his decision to bring the 35th G8 summit to L'Aquila in July 2009, 3 months after the earthquake hit the region. The need to secure votes in provincial elections by Prime Minister Berlusconi's Government only weeks after the disaster was perhaps the incentive behind a centralized top-down approach to the rehousing projects in L'Aquila, crucially ignoring local public opinion (Fig. 10).

However, once the cameras and world leaders left and votes for the provincial elections were secured, the town center was left alone. In a recent article in the BBC (2013), Mayor Massimo Cialente accused the government of forgetting about L'Aquila. He says the delay has been due to paralyzing bureaucracy, a lack of organization between the local and national governments, and a lack of funds. But reconstruction is more easily said than done in a city where many of the buildings are centuries old, where homes are interwoven with historically important buildings, and where the narrowness of the streets means cranes and other construction vehicles can only be brought in a few at a time.

Reflections on the Great East Japan Earthquake and Tsunami of 2011

The earthquake and tsunami devastated Tohoku and other regions. In total by the 18th month after the event (September 2012), around 1,230,000 buildings were recorded as having

Earthquakes and Their Socio-economic Consequences,

Fig. 10 World leaders at the G8 visit L'Aquila July 2009 (Photo courtesy of Reuters)



E

Earthquakes and Their Socio-economic Consequences,

Fig. 11 Devastation caused by the Great East Japan earthquake and tsunami (Copyright EPA)

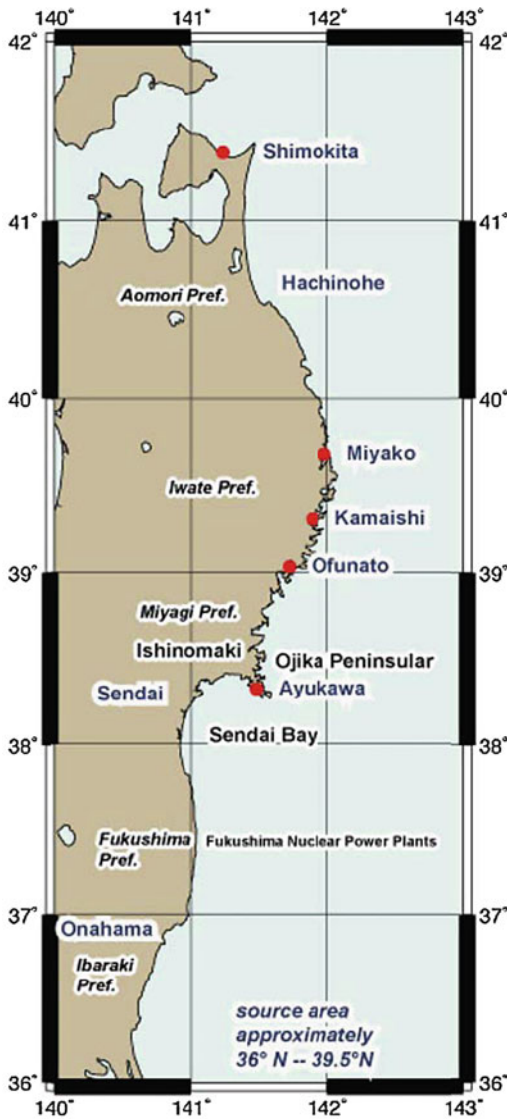


been damaged or destroyed due to tsunami and ground shaking (which is equivalent to 2.4 % of all of Japan's existing buildings) and the tally continues. Life loss is reported as 20,721 and includes 2,303 earthquake-related deaths that occurred due to various causes after the day of the earthquake.

In terms of cause of death (21/4/2011):

- 92.5 % drowned.
- 4.4 % crushed/died of injuries.
- 1.1 % burnt.
- 2.0 % unknown (Fig. 11).

According to MunichRe, the overall losses of this event are estimated to be over US\$210 billion. The monetary cost associated with the failure of the Fukushima Daiichi nuclear power plant as a result of the 14 m tsunami is still not quantified to this day, though Tokyo Electric Power Company (TEPCO) announced a ¥1.02 trillion cost of natural disaster. Immediately after the disaster, approximately 85,000 people were evacuated from 20 km evacuation zone and 185,000 people were screened for contamination. Less serious problems occurred in the Onagawa and Fukushima Daini nuclear power plants.



Earthquakes and Their Socio-economic Consequences, Fig. 12 Locations visited by the UK Earthquake Engineering Field Investigations Team mission (EEFIT 2011)

The strongest tsunami consequences occurred (north to south) from the town of Fudai (in northern part of Iwate Prefecture) down to Iwaki (in the southern part of Fukushima Prefecture). In total it was estimated that around 600 km² of land was flooded, where around 600,000 people lived. Around 18,500 people are dead or missing as a result of the tsunami, with

the worst effects occurring in Miyako, Yamada, Ofunato, Otsuchi, Kamaishi, Rikuzentakata, Minami Sanriku, Kesennuma, Onagawa, Ishinomaki, Higashi Matsushima, Tagajo, Sendai, Natori, Watari, Iwanuma, Yamamoto, Soma, Minami Soma, Namie, Shintchi, and Iwaki (Fig. 12).

In terms of critical facilities, networks, and infrastructures, the power, gas, and water supplies were disrupted in many areas, mostly in the Tohoku region. Roads, railways, ports, airports, and other infrastructure were also damaged. Damage to roads was recorded in all eastern prefectures between Iwate and Kanagawa (in total 4,188 damaged roads of which 2,343 were in Chiba Prefecture). Almost all roads have been reopened quite rapidly, including the Tohoku Expressway, most of the Joban Expressway, as well as national highways. Damage to 116 bridges was recorded, of which 97 were in the Kanto region.

There were also significant secondary hazards other than the tsunami. Landslides were recorded in 208 locations and caused the loss of 19 lives (including 12 killed by the Hanokidaira landslide in Shirakawa city of Fukushima Prefecture). Extensive liquefaction occurred on the northern shores of Tokyo Bay particularly in Urayasu town (Chiba Prefecture). Fires occurred in the Cosmo Oil's refinery in Ichihara town (Chiba Prefecture) as well as in Kesennuma and Natori (Miyagi Prefecture) during and after the tsunami.

It must be noted that the damage, loss, and human casualty data given here originate from Japan's Fire and Disaster Management Agency (FDMA) report issued on the first-year anniversary of the event and are not the final figures, as efforts continue to assess all the consequences of this tremendous earthquake.

The information and reflections in the subsequent sections on the factors affecting the recovery of the Great East Japan earthquake and tsunami have been put together by the authors' visit to the affected area as part of the UK's Earthquake Engineering Field Investigation Team (EEFIT) reconnaissance mission in June 2013.

Government Plans for Reconstruction

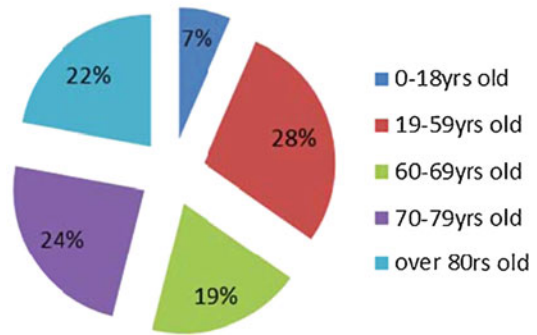
One of the key issues found when reviewing the plans for reconstruction is how much room for adaptation there is in the application of the central government's template for recovery (EEFIT 2014). Japan is a compliant society, and there may be more flexibility than bureaucrats or residents realize. But no one has any inkling about the cost-benefit of the huge investment. This is a national response to disaster, but there are differing opinions in the local area in terms of priorities and decision making.

The Reconstruction Agency was established by the Prime Minister's Office. It advises government on all basic strategies based on quick lessons. But hard solutions give a false sense of increased security and early warning is an issue. The reaction of government has been self-critical. Instead of defending the system, they have been frank about what failed and it is in the character of Japan to always review and look back and learn lessons (Bhatia et al. 2013). For example, ideas about evacuation are changing. Japan national broadcasting has changed the way it announces the early warning. Rather than giving precise information that is open to error and misinterpretation, it will from now on give much simpler direct warning to evacuate immediately.

The focus, at the time of the EEFIT mission in June 2013, has been on relocating housing and safety measures, whereas the imperative is to strengthen the local economy and address economic and demographic decline. Measures that would strengthen existing local businesses and city center shops, attract new industry, and encourage young people to the city might also have been considered. One thing the central government might have considered is founding a college of higher education, either a new university or a branch of a university in Sendai, preferably one that focused on technology and had practical links with industry and enterprise.

There is a proposal from Kobe University to revitalize small business, but people in the affected area do not have the resources or money to take action or to exploit new technology. New people would be most welcome. Do people take the initiative and accept

Ages of those who lost their lives in the Great East Japan Earthquake and Tsunami, 2011



Earthquakes and Their Socio-economic Consequences, Fig. 13 A breakdown of ages of the dead in the GEJ earthquake and tsunami of March 2011, of those 5,971 were male and 7,036 female

responsibility or do they expect people to come and help solve the problem? People with initiative would be a good thing. People have been here a long time and cannot see how to fix the problem.

Socio-economic Characteristics of the Region (Pre- and Post-disaster)

A breakdown of the ages of those who died in the disaster revealed a disproportionate percentage of the elderly were affected; 65 % of those that lost their lives were over 60 years old as shown in this graph (Fig. 13).

Population emigration due to the disaster is largely occurring among young people. The International Recovery Platform pointed out the issues being faced by the affected areas following the Great East Japan earthquake are compounded by the problem of shrinkage confronting most rural towns in Japan. In addition to issues of safety and relocating housing, population decline, ageing, and economic shrinkage pose special planning challenges (International Recovery Platform 2012).

It is hoped that tackling these issues by reordering land use, improved transportation links, and urban center regeneration projects will have a positive impact on the prospects of these places as well as make them more resilient to a future disaster.

Not all places that were affected by the tsunami are the same, however. The area around Sendai in Miyagi Prefecture is a flat plain and has a strong economy, good transport links, and a growing population. Further north in Iwate Prefecture, there are steep slopes and fiords, a declining population, and a weak economy. In each there are differences of scale with a few larger cities and towns and many smaller settlements and villages. This suggests that different places face different issues of recovery. These differences in socio-economic prospects, demography, topography, and scale suggest that approaches to both safety issues and economic development assistance might be fine-tuned to meet local circumstances.

Citizen Involvement in Decision Making

In the areas affected by the 2011 tsunami, consultations between governments and communities were the rule, and community representatives were invited to serve alongside experts on recovery planning committees from the earliest stages. The most common ways of collecting residents' opinions were surveys and workshops. The central government and local governments outside the disaster-affected area helped affected municipalities plan their recovery by conducting research, seconding staff, and hiring professionals to provide technical support. University faculty members, architects, engineers, lawyers, and members of NGOs participated in the municipal planning process (World Bank 2012).

Along the Rias coast the response of the majority is that the government has already decided so they can't do anything. Some even admire the massive colossal infrastructure. But the younger generation, in their forties, is opposed to large embankments and tall sea-walls, but they are not the decision makers. In Japanese community associations, it is elderly men who make the decisions. In Ohyakaigan near Kesennuma the community association meets twice a month and tries to involve children as well as older people. The plan is to collectively relocate the 120 households and to have the land cleared by 2015–2016. The

Japanese Institute of Architects (AIJ) is considering using this as a model of participation. Unfortunately people cannot wait and they are now down to 100 households and the community may fall apart because of the delay. The group decided they would not oppose the proposed embankment but suggested it be moved back. Initially the city was not happy but changed their minds after receiving the petition. The proposed municipal plan is now for a much lower embankment further back, but this needs cooperation between the Ministry of Forestry, Japan railways, the National Highways Agency, and the prefecture.

In Kesennuma the citizens' committee oppose the planned harbor embankment, and they are in talks with the municipality and prefectural government. The majority of residents are against the proposal and it is not settled yet. Planning arrangements cannot proceed while there is a dispute, but city officials are making land-use plans assuming the embankment will go ahead. Because there has been so much opposition, a new deadline has been set for October 2013. In other places plans are proceeding more rapidly.

In Kamaishi the three community workers for the prefecture explained that local authorities have to accept what citizens want. Ideally they would simulate different heights of embankment since communities in some places have opposed the plans. If people want something outside, the government's recommended solution planners have to be careful that safety measures are in place and that the community has collectively relocated.

Partly because of citizen opposition, reconstruction of sea embankments, which suffered extensive damage, has been considerably delayed. Local governments in devastated areas cannot decide on the details of restoration plans, as discussions continue on whether to prohibit people from returning to coastal areas. Reconstruction work has started on only 31 % of destroyed embankments. According to the Fisheries Agency, which has jurisdiction over sea embankments, the design of embankments will depend on whether people will live nearby (Daily Yomiuri 2013).

Earthquakes and Their Socio-economic Consequences,

Fig. 14 Photograph of the moment the tsunami inundated the Natori plains by the Sendai Airport on the March 11, 2011 (Credit AP)



E

Local governments were tasked with recovery by the National Government who asked them to develop local plans based on consultation. The problem is they lack the technical capacity, especially in effective methods of involving citizens in strategic decision making. Voluntarily urban planners and architects from all over Japan surged to provide missing capacity. Local government has lots of problems with consultation, which is time-consuming. And it is not easy to convince communities to relocate. Local governments want to consolidate communities to make it more efficient and economical to deliver services, but many of these places were in decline before the tsunami. The fundamental problem is that the authorities do not really know what size population they are reconstructing for. They have to provide facilities to each community, so the cost is enormously considerable (Bhatia et al 2013).

Politics and the Media

There was inevitably a great deal of media attention surrounding the Great East Japan earthquake, especially over the Fukushima nuclear disaster. In addition to the coverage of the aftermath of the earthquake and tsunami, there was intense attention on how the leaders of the then fragile government would respond to the

disaster. The central government was keen to act fast and one of the first projects completed was the 12 m high seawall along the Natori plains by the Sendai Airport. This structure is twice as high as its maximum height than the previous wall. Figure 14 shows the inundation of the tsunami into the Natori plains by the airport.

Although the need to protect the coast from future risks is clear and the government would argue that it is doing what is right in saving towns that have been here a long time, the public pressure was there to spend as quickly as possible and there was money in the economy to support this. But there was a time limit for people (central and local) to decide on what to do, as the government had set stringent deadlines for spending.

The media has mainly focused their attention on the Fukushima Daiichi nuclear disaster. Following the earthquake, a 15 m tsunami disabled the power supply and cooling of three reactors, causing a nuclear accident. All three cores melted in the first 3 days and the accident was rated seven on the INES scale, due to high radioactive releases over following 3 days. Eventually four of the six reactors had failed. The days and months after the failure, the handling of the disaster by the government and TEPCO, the Tokyo Electric Power Company, was under heavy

scrutiny. To this day there is mistrust of the statements made by TEPCO on the safety of the surrounding areas, and some residents in and around the region have instead chosen to self-evacuate or purchased their own Geiger meters to track the radiation levels.

The disaster changed the national debate over energy policy almost overnight. Japan, a country heavily reliant on nuclear power, due to its own shortage of fossil fuels had by June 2012 shut down all but two of the country's 54 nuclear reactors. This discussion continues and it remains to be seen what the true long-term policies and health effects are from the nuclear accident.

For this devastating event, the socio-economic consequences have had local, national, and global effects and will continue for decades to come, raising the question: which is the biggest disaster – the earthquake (concentrated and powerful), the tsunami (vast area affected), or the radiation release (contamination for centuries)?

Looking to the Future

Earthquakes and their socio-economic consequences will continue to dominate headlines as highlighted in the chapter. The need to consider the severity and long-term impacts of socio-economic factors in earthquake risk reduction and disaster management is therefore crucial. Echoing the recommendations drawn from investigating socio-economic impacts of Iranian earthquakes by Hosseini et al. (2013) includes:

1. Promoting public participation
2. Organizing community-level disaster management groups and systems
3. Developing a natural hazards insurance system
4. Establishment of a catastrophe insurance pool
5. Socio-economic impact assessment for future earthquake scenarios

As one anticipates for a future with the possibility of a “million-death earthquake,” below are some additional recommendations to those listed

above, if initiatives scientists and engineers should focus their efforts on for earthquake mitigation and preparedness.

Understanding Human Behaviors in Emergencies

Risk perception is not a precise science and even with the best of intentions, the way one acts when experiencing an earthquake. In an interview with a survivor of the Great East Japan earthquake and tsunami, he recounted how he knew it was the largest earthquake he had personally ever experienced and that he knew that a large tsunami would follow. However, against every rational and educated decision he could have made to run to higher ground, he chose deliberately to rush to the coast, to where his family was. He survived, but many others including his family did not. This is just one of many similar stories (EEFIT 2014). Understanding the needs and reasoning behind actions after a disaster would help us devise appropriate strategies for mitigation. What physical and social provisions would have changed this man's decision?

Incorporating Socio-economic Impacts in Earthquake Loss Models

To better understand and address earthquake risk in its entirety, one needs to go beyond seismic hazard, exposure, and the vulnerability of the built environment. The socio-economic characteristics of populations exposed to earthquake threats need to be integrated within loss estimation models to an integrated and holistic estimate of risk in an area. A realistic socio-economic impact assessment can be used to examine how an earthquake will change the lives of the current and the future residents of a community. However, it is not easy to quantify the changes in social and economic conditions that will arise from future earthquakes. At present these are done as modelled with reconstruction costs per square meter. The main indicators include changes in demographics, characteristics of the community, potential effects of an earthquake on services and the housing market, changes in employment and income levels, and alterations to the quality of life.

As part of the GEM, there has been an emphasis for the first time to review how one could assess the socio-economic impact of earthquakes in loss estimation. The ambition of GEM is to deliver methods, metrics, and open source software that can be used worldwide to explore the compounded nature of earthquake events where environmental processes blend with factors such as urban growth, marginalization, and poverty to bring about and magnify earthquake impacts. The integration of physical risk with (a) the social characteristics of populations and (b) the susceptibility of people to the adverse impacts of earthquakes leads to an encompassing perspective on global risk assessment which:

- Considers loss and damage as part of a dynamic system, where interactions between natural systems and societal factors redistribute risk before an event and redistribute loss after an event
- Mainstreams socio-economic vulnerability and resilience in policy discussions on earthquake loss and damage
- Evaluates loss and damage, taking social factors into account at different time and space scales
- Uses risk assessments in benchmarking exercises to monitor trends in earthquake risk over time
- Recognizes that both causes and solutions for earthquake loss are found in human-(built) environmental interactions (GEM 2014).

Summary

Recovering from a disaster involves all kinds of development and planning activities – land-use plans, building norms, and transport plans. What is unique in post-disaster situations is that all these activities happen in a compressed period of time. Communities must rebuild as quickly as possible to maintain existing social networks and get the economy back on its feet. But they must also be deliberate in trying to maximize the opportunities disasters provide for improvement.

Disasters in the past are often seen as catalysts for change in the built environment. The urgency of rehousing a large number of people affected by the event mean that funds are diverted and attention is paid to the area in need. Time is compressed and here is an urgency to “get back to normal” – to rebuild livelihoods, clear up the debris, repair the damage, and, among the more farsighted, to “build back better.” A challenge in post-disaster recovery and key to reducing the long-term social and economic consequences of earthquakes is balancing the need for speed and deliberation.

Earthquake losses will continue to rise if the community of practitioners working in the field do not come together and try to mitigate these losses together. Earthquake engineers and scientists have the knowledge to dramatically reduce the number of deaths from earthquakes. Preventable deaths are attributed to damage and collapses of buildings and buildings should not fall down. It is clear that earthquakes can have a significant impact on the social and economic fabric of society, both in the short and the long term. The ability of a society to recover depends very much on the precondition of the affected area but even more so on the immediate decisions and subsequent policies made by the governing authorities. Some would argue that losses mainly rise from political factors, not from a lack of knowledge or inability to apply it. Mitigation is not a priority in the short term, and therefore governments and communities are not necessarily interested in devoting adequate resources to mitigation. Several authors would further argue that corruption is the main contributing factor to losses from earthquakes (Escaleras et al. 2007; Ambraseys and Bilham 2011), and unless we tackle this sensitive issue, there is little hope of mitigating catastrophic losses. This is not disputed here. However, one should and must remain positive to strive for a better understanding of the direct and indirect causes of losses in earthquakes and devise appropriate mitigation strategies for the future, tailoring for local sociodemographics and focusing on the either physical or social interventions, or both. Inaction is not an option.

Cross-References

- [Seismic Loss Assessment](#)
- [Insurance and Reinsurance Models for Earthquake](#)

References

- Alexander DE (1985) Death and injury in earthquakes. *Disasters* 9(1):57–60
- Alexander DE (2012) A tale of three cities and three earthquake disasters. *Tafer J* 50:1–10
- Ambraseys N, Bilham R (2011) Corruption kills. *Nature* 469:153–155
- Bhatia et al (2013) Personal interview, June 2013
- Bilham R (2009) The seismic future of cities. *Bull. Earthq. Eng.* 7(4):839–887
- Bilham R, Hough S (2006) Future earthquakes on the Indian continent: inevitable hazard, preventable risk. *South Asia Mag* 12:1–9
- Boarnet MG (1995) Economic impacts of the Northridge Earthquake's transportation damage: results from a survey of firms. Final Report on Federal Highway Administration Contract number NRS-3041(125)-1
- Byrne B, Baden S (1995) Gender emergencies and humanitarian assistance. BRIDGE Report No 35, IDS, Sussex
- Calvi GM, Spaziente V (2009) Reconstruction between temporary and definitive: the CASE project. *Progett Sism* 3:221–250
- Chou Y-J, Huang N, Lee C-H, Tsai S-LC, Chang H-S (2004) Who is at risk of death in an earthquake. *Am J Epidemiol* 160:688–695
- Dankelman I, Alam K, Ahmed WB, Gueye YD, Fatema N, Mensah-Kutin R (2008) Gender, climate change and human security lessons from Bangladesh, Ghana and Senegal, report prepared by The Women's Environment and Development Organization (WEDO) with ABANTU for Development in Ghana, Action Aid Bangladesh and ENDA in Senegal
- Earthquake Engineering Field Investigation Team (2011) Post earthquake field investigation of the Mw9.0 Tōhoku Earthquake of 11th March 2011. A field report by EEFIT
- Earthquake Engineering Field Investigation Team (2014) Recovery two years after the 2011 Tōhoku Earthquake and Tsunami. A return mission report by EEFIT
- Escaleras M, Anbarci E, Register CA (2007) Public sector corruption and natural disasters: a potentially deadly interaction. *Public Choice* 132(1–2):209–230
- Hatch SL, Dohrenwend BP (2007) Distribution of traumatic and other stressful life events by race/ethnicity, gender, SES and age: a review of the research. *Am J Community Psychol* 40:313–332
- Hazeleger T (2013) Gender and disaster recovery: strategic issues and action in Australia. *Aust J Emerg Manage* 28(2):40–46
- Hosseini KA, Hosseini S, Pooyan Z (2013) An investigation into the socio-economic aspects of two major earthquakes in Iran. *Disasters* 37:516–535
- International Recovery Platform (2012) Transitional shelter: knowledge notes, cluster 4: recovery planning note 4–3
- Jackson J (2006) Fatal attraction: living with earthquakes, the growth of villages into megacities, and earthquake vulnerability in the developing world. *Philos Trans Roy Soc A* 364:1911–1925
- Kates RW, Pijawka D (1977) From rubble to monument: the pace of reconstruction. In: Haas JE, Kates RW, Bowden MJ (eds) *Disaster and reconstruction*. M.I.T. Press, Cambridge, pp 1–23
- Masateru H (2011) Demographic trends should not be overlooked: aging and depopulation in Tohoku. The 2011 East Japan Earthquake Bulletin of the Tohoku Geographical Association, 9 Apr 2011
- Musson R (2012) The million death quake: the science of predicting earth's deadliest natural disaster. Palgrave Macmillan, London
- Özderem A, Jacoby T (2006) *Disaster management and civil society: earthquake relief in Japan, Turkey, and India*. I.B Tauris, London
- So E et al (2012) An introduction to the global earthquake consequences database (GEMECED) 15th world conference in earthquake engineering, Lisbon, 24–28 Sept 2012
- Vranes K, Pielke RA Jr (2009) Normalized earthquake damage and fatalities in the United States: 1900–2005. *Nat Hazards Rev* 10(3):84–101

Online

- BBC (2013) L'Aquila after the earthquake: why flags do not fly. 25 June 2013. <http://www.bbc.co.uk/news/world-europe-22986058>. Accessed 15 Jan 2014
- Chew L, Ramdas KN (2005) Caught in the storm: the impact of natural disasters on women. <http://www.globalfundforwomen.org/storage/images/stories/downloads/disaster-report.pdf>. Accessed 4 Feb 2014
- GEM (2014) Global earthquake model. <http://www.globalquakemodel.org/>
- GEMECED (2014) <http://www.globalquakemodel.org/what/physical-integrated-risk/consequences-database/>
- Daily Yomiuri editorials (2013) 2 years on, reality of recovery differs from official data. 11 Mar 2013. <http://www.accessmylibrary.com/archive/434970-yomiuri-shimbundaily-yomiuri/march-2013-page4.html>. Accessed 7 Sept 2013
- World Bank (2012) Reconstruction policy and planning, Knowledge note 4–2, Cluster 4: Recovery Planning [Online] Sept 2012. Available from: http://wbi.worldbank.org/wbi/Data/wbi/wbicms/files/drupal-acquia/wbi/drm_kn4-2.pdf. Accessed 4 Sept 2013

Economic Impact of Seismic Events: Modeling

Garry McDonald, Nicola Smith and
Catherine Murray
Market Economics Ltd, Takapuna, New Zealand

Synonyms

Agent-based modeling; Computable general equilibrium; Economic impact assessment; Economy; Employment; Gross domestic product; Input–output analysis; Integrated assessment; Loss modeling; Macroeconomic impacts; Market valuation; Microeconomic impacts; Nonmarket valuation; Social accounting matrices; System dynamics

Introduction

People live in earthquake-prone areas. Tectonically active landscapes may have produced greater biodiversity, more food and water, through disturbed drainage patterns and wetland creation. Nevertheless over the last 200 years, global population growth has increasingly placed pressure on human settlements and resulted in the proliferation of urban areas worldwide, many in seismically active zones. The opportunities provided by living close to a seismic area are tempered by the risks of loss and destruction during a seismic event. Attitudes toward seismic risk are informed not only by understanding of the geophysical characteristics underpinning seismic events but also by socioeconomic conditions, such as historical attachment to the area, property ownership, and community connectedness. Although risks of damages to property may be handled on an individual basis through insurance markets, risks of seismic events on a societal level are seldom channeled through formal markets, creating a strong mandate for policy analysis and intervention.

Economic modeling of seismic events is based on a pragmatic anthropocentric view of managing risk and uncertainty. Economic modeling can be undertaken *ex post* (after an event) or *ex ante* (prior to an event). *Ex post* economic modeling is often used to grasp the intensity and extent of disasters, to appraise costs of recovery and reconstruction, and to help build a body of information to assist in future management of seismic risk. *Ex ante* economic modeling is hypothetical and is often used to support planning and decision making, particularly pertaining to land use and infrastructure provision, with an emphasis on mitigating effects of, and creating response strategies for, seismic events. Any economic assessment of a seismic event is only a partial representation of the real world, not a comprehensive replication of the reality.

Types of Impacts Captured by Economic Analyses of Seismic Events

The underlying rationale for applying economic analysis to understand the consequences of seismic events is concern for the well-being of those persons affected or potentially affected. For the most part, economic analysis does not aim to measure changes in well-being directly, but rather changes in the level of activity and assets held within the formal *market*-based economy. Commonly used market-based indicators of well-being include business output or turnover, value added or gross domestic product, capital assets including property values, household income, employment, and consumption of goods and services.

Seismic events may also result in a wide range of *nonmarket* social and environmental impacts. These nonmarket impacts may be both negative and positive. On the one hand, negative impacts may result from loss of life and physical health, psychological problems and stress, destruction of family and social networks, inconvenience, and degradation of ecological systems. On the other hand, seismic mitigation measures put in place

Economic Impact of Seismic Events: Modeling, Table 1 Typology of economic impacts associated with seismic events

		Examples of direct impacts	Examples of higher-order impacts
Market-based analyses	Stocks	Damage to buildings, infrastructure, inventories	Reduced purchases of public assets as finances diverted toward disaster recovery
	Flows	Loss of production by businesses with damaged capital	Loss of production by suppliers and customers of businesses unable to operate Reconstruction diverts resources away from more efficient production activities
Nonmarket analyses	Stocks	Loss of life	Loss of social capital via disruption of social networks and relocation of people
		Loss of physical and mental health	Loss of trust in authorities
		Damage to natural capital	
	Flows	Interruption of ecosystem services Loss of leisure as a result of inconveniences experienced during the event and its recovery	Stress associated with lost incomes and jobs

may lead to positive impacts on well-being by creating a greater sense of security and safety for local residents. These social and environmental impacts may be associated with changes in economic activity, e.g., physical or psychological injuries sustained as a result of a seismic event constitute a loss of human capital which may, in turn, result in a loss of economic production. However, methods based solely on analysis of market outcomes are unlikely to capture the full extent of the effects. In such cases nonmarket valuation techniques may be utilized for impact assessment.

Table 1 provides a typology for the various types of economic impacts associated with seismic events. As with most existing typologies (e.g., Rose (2004) and Hallegatte and Przulski (2010)), Table 1 distinguishes not only between *stock* and *flow* impacts but also between *direct* and *higher-order* impacts. The distinction between stocks and flows is well recognized within economics. A *stock* is a quantity of a system variable that exists at a single point in time, while a *flow* is a movement of materials or information to or from a stock over time. Stock losses associated with seismic events include the physical destruction of houses and other buildings, infrastructure, machinery, inventories, and other physical assets, as well as loss of human capital from deaths, injuries, and relocation.

Flow losses include reduction in the rate of output by businesses, reduction in the quantities of goods and services consumed by households over a given period, or the hours of lost leisure as a result of various inconveniences following a disaster.

Direct economic impacts include damage to buildings, infrastructure, and inventories, loss of production by businesses with damaged capital, loss of life, loss of health, and so on. Such impacts are a direct physical consequence of a seismic event. *Higher-order* impacts of earthquakes are those not provoked by an earthquake itself but by its consequences. Sometimes higher-order impacts are referred to as “secondary” or “indirect” effects; higher order is used here to avoid any confusion with “indirect effects” as used in input–output (IO) analysis (Rose 2004; Okuyama 2007). Regardless of the particular term used, the separation between direct and higher-order impacts is often blurred. Particularly contentious is how business interruptions are classified. On the one hand, business interruption caused directly by damage to capital used in production, e.g., damage to the Sendai fishing fleet during the 2011 Tōhoku earthquake and tsunami, may be considered a higher-order impact. On the other hand, this could be referred to as a direct impact, with the term higher-order reserved for business interruption *not* directly caused by damaged capital. These

would include, for example, a number of car factories in Korea and the United States that experienced shutdowns following the 2011 Tōhoku event because of the disruption in supply of key components. Hallegatte and Przulski (2010) propose that higher-order impacts extend to cover those costs or benefits that span a longer period of time and a larger spatial scale or occur in a different economic sector than the disaster itself. Accordingly, business interruptions are direct losses if caused by damage to buildings, plants, machinery, or critical infrastructure or disruption to employees' availability for work, while losses in production that spread to other firms via backward and/or forward linkages are higher-order impacts.

A further point of distinction in the categorization of earthquake impacts, which is not considered in Table 1, is the separation of costs and benefits that arise from the occurrence, or potential occurrence, of an event itself and those that arise from mitigation measures, e.g., earthquake strengthening of buildings, training of urban search and rescue personnel, and so on.

Market-Based Methods

Up until the 1990s, the economic impacts of earthquakes and other disasters had received relatively little attention within mainstream economics (Okuyama 2007). Subsequently, a plethora of studies extended and modified conventional market-based economic models. IO analysis, for example, has been extensively applied in the assessment of higher-order impacts. Other conventional market-based economic methods used in assessment of seismic events include econometrics, social accounting matrix (SAM), and computable general equilibrium (CGE). With largely the exception of econometrics, the remaining methods have principally focused on determining higher-order impacts. Economic impact assessment undertaken using conventional methods must thus rely on either empirical data, engineering models, or estimation to calculate direct impacts. In recent years there has been growing interest in modeling the

economic impacts associated with seismic events using simulation models that do not originate solely from within the economics discipline. These have included inter alia the use of system dynamics (SD), agent-based models (ABMs), and integrated assessment methodologies.

Engineering-Based Models for Valuing Capital Losses

Studies concerned only with the estimation of flow impacts often require, as a direct modeling input, the assessment of the value of earthquake damages to buildings, infrastructure, stocks, and other physical capital. The value of lost capital is required to determine the likely changes in economic production and the expenditure patterns during recovery and reconstruction. A common approach used to estimate the direct costs from seismic events is the use of susceptibility or damage functions. Such functions describe the relationship between one or more seismic parameters (e.g., ground motion) with damage for a certain type or use of object at risk, which can then be directly correlated to economic losses.

Econometrics

Econometrics focuses on providing an empirical basis to economic relationships, through the use of statistical analysis, for testing economic theory, forecasting, decision making, and ex post evaluation. Although perhaps best known for its use in generating short-term forecasts of the performance of national economies, econometric modeling is widely utilized in both microeconomics and macroeconomics. Econometric modeling is also undertaken in combination with other forms of modeling, particularly for the purposes of parameter estimation, e.g., for CGE, SD, and ABMs.

In the analysis of seismic events, several econometric studies have attempted to understand the determinants of the direct costs (e.g., mortality, capital losses). According to Cavallo and Noy (2009), these studies typically employ a model framework that assumes direct costs are related to some measure(s) of the physical magnitude of the disaster (e.g., ground movement) and variables that capture the "vulnerability" of the region to the

disaster (i.e., susceptibility to the impact of natural hazards). Possible determinants that may increase a country's vulnerability include the level of economic development, population, and various political and institutional conditions.

A number of authors have also investigated the higher-order impacts of earthquake disasters. Many of these studies look at changes in conventional macroeconomic indicators, such as GDP per capita, although there has also been some research at the sector level and to changes in poverty as measured by indicators such as the Human Development Index. Other studies on natural disasters, although not necessarily earthquake specific, have also focused on impacts at the village or household level, including effects on consumption, nutrition, and capital. Macroeconomic studies may be generally separated into those that are interested in short-run (up to several years) and those that consider long-run (at least 5 years, sometimes decades) impacts. Examples of short-run analyses include those authored by Hochrainer (2009) and Jaiswal and Wald (2013), while long-run studies include those produced by Skidmore and Toya (2002) and Crespo Cuaresma et al. (2008). The focus of much of the debate in long-run analyses is on the extent to which natural disasters may contribute toward faster economic growth via the replacement of damaged capital with newer and more productive technologies.

Given the highly variable nature of earthquake events, obtaining sufficient data (time series and cross sectional) upon which to build such models poses a significant challenge. By its very nature, econometric modeling also depends upon the assumption that the underlying causal mechanisms do not change in form, strength, or stochastic properties from the estimation period to the forecasting period or from one cross-sectional sample to another. Econometric models are, however, statistically rigorous and may provide stochastic and forecasting capabilities.

Input–Output Analysis

IO analysis is attributed to Wassily Leontief, the 1973 Nobel Laureate for Economics, who in 1936 published the first IO table for the US

economy (Leontief 1941). Arguably, IO models are the most widely applied tool for assessing higher-order impacts of seismic events. A significant advantage of IO analysis is its ability to capture economic interdependencies within a regional or national economy. In its most basic form, an IO model consists of a system of linear equations, each one describing the distribution of economic goods among industries and final consumers. An IO model is typically used to evaluate the way in which changes in production by one economic industry may, through supply chain linkages, ripple through an economic system. Although mainly concerned with flow impacts, IO models have some potential to capture stock impacts when constructed as a dynamic model.

There are several well-known limitations associated with IO analysis, including linearity, absence of quantity–price interactions, rigid structure with respect to input and import substitutions, and neglect of resource (e.g., capital, labor, and land) constraints. In particular, IO models utilize “technical coefficients,” which describe, for any given industry, the quantity of each type of good or service required per unit of production. Although technical coefficients are relatively stable under normal economic conditions, it may be unrealistic to assume that these will remain constant following a disaster, when an economy is likely to be subject to supply shortages and significant adjustments. A variety of ad hoc extensions and refinements have been put forward to help address IO weaknesses. For example, Cochrane (1997) allows for a more flexible treatment of imports to deal with shortages in local production. Similarly, Hallegatte (2008) adapts the IO framework to incorporate production bottlenecks and changes in production capacity, prices, and labor demand.

Recently, the inoperability IO model (e.g., Haimes et al. 2005) has received significant attention in the context of disaster analysis. In mathematical terms, the evaluation of higher-order impacts within the inoperability IO model is similar to other IO models, except that results may be stipulated in normalized terms, with the point of reference being the as-planned level of production for each economic industry. Inoperability IO

models are commonly used to explain the dynamics that underlie business operation losses. For example, Barker and Santos (2010) explore the role of inventory in muting disruption following an event.

Social Accounting Matrices

SAMs are closely related to IO models – both are typically constructed from a country's System of National Accounts. While the traditional IO model focuses on flows among economic actors involved in production (i.e., industries), a SAM additionally describes the way in which income generated by factors of production (e.g., capital, labor) is allocated among different economic actors (e.g., businesses, households). Secondary distributions of income among actors through, for example, tax payments and government transfers, along with allocation of income to savings, are also included, thus providing a complete picture of the circular flow of income within an economy. SAM-based models thus have the potential to evaluate a greater range of higher-order impacts than an IO model. SAM-based models are however subject to the same limitations as IO models, including the use of rigid technical coefficients.

Computable General Equilibrium

Starting with Leif Johansen's pioneering (1960) multi-sectoral study of economic growth, and aided by rapid advances in computer technology, CGE models have now become a standard part of the toolbox of economists concerned with policy-oriented research. In general terms, a CGE model seeks to simultaneously determine commodity and factor prices within an economic market and the quantities of commodities produced and consumed, subject to budget (e.g., household, government) and resource (e.g., capital, labor, land) constraints. Like IO models, CGE models are typically used to assess flow impacts, for example, changes in economic output, gross domestic product, or household consumption of commodities. Dynamic CGE models also have the potential to track changes in stocks (e.g., capital, labor) over a recovery period. While CGE

models have been identified as promising tools for earthquake and other disaster analysis for some time, the number of actual examples in the literature is still relatively few.

When compared to IO and SAM-based models, CGE models exhibit a number of advantages. Firstly, the modeling framework provides for flexibility in the specification of functions describing the production of each type of commodity, including allowance for substitution between inputs in response to relative changes in price. For example, if a particular good becomes scarce following an event due to damage to local production facilities, a CGE model will allow for industries requiring that good for production to increase purchases of the imported form of that good. Secondly, the framework allows for consumers to alter consumption patterns in response to changes in relative prices of commodities and/or income levels. Despite this generic flexibility, practitioners typically concentrate on specifying production functions with substitution only between factor inputs (i.e., labor and capital), between imported goods and domestic goods, and possibly between some other key inputs (e.g., substitution between energy commodities is often included). Practitioners typically rely on IO tables to specify purchases of intermediate goods.

Whereas IO models are mainly suited to the analysis of demand-side supply chain and some income-related effects, CGE models are also suited to the analysis of supply-side effects (i.e., impacts on customers) and price effects. For example, if half of the capital used in producing a particular type of commodity is damaged during an earthquake, among the impacts considered by a CGE model will be the loss of income for the firms directly impacted by lost production capacity, loss of income for suppliers of directly impacted firms, gains in income for remaining firms supplying the commodity as a result of higher market prices, and gains in income for firms supplying commodities that experience increasing market demand as a result of input substitution. Due to the wide ambit of economic impacts considered, with impacts often occurring in opposite directions (i.e., counterbalancing),

CGE models typically generate smaller estimates of overall impact than IO or SAM-based models.

In terms of limitations, CGE models are constrained to the evaluation of economies at equilibrium. A typical CGE analysis is undertaken by “shocking” an economic system (e.g., adjusting capital stocks to reflect damage and reconstruction) and, in turn, evaluating the change in the system necessary to produce a new equilibrium. While the equilibrium nature of CGE models is not a problem with long-run analysis, it is however a significant shortcoming for short-run analysis following a seismic event, when economies are likely to undergo significant adjustment and exhibit out-of-equilibrium dynamics. Rose and Guha (2004) therefore distinguish different response phases following an event and identify some of the variations in model parameters suitable for evaluation at each phase.

The extensive data requirements for a CGE model present a significant limitation to the more widespread application in earthquake economic impact assessment. In addition to all of the data requirements of a typical IO model, a CGE model requires the input of numerous behavioral functions and substitution parameters. Parameter estimation is also often based on data from historic behavior under business-as-usual conditions, and such behavior may be ill suited to describing behavior following a seismic event. Although some advancements have been made toward refinement of model parameters to better suit disaster analysis (Rose and Liao 2005), the examples of such work are still relatively few.

System Dynamics

SD is a dynamic modeling approach for understanding the behavior of complex systems. Jay Forrester, of the Massachusetts Institute of Technology Sloan School of Management, developed SD in the mid- to late 1950s. Examples of complex systems include the environment, economy, and society. Such systems are characterized by the presence feedback loops and time delays and typically exhibit nonlinear behavior. Feedback loops are classified as either reinforcing (i.e.,

exacerbate dynamics away from an equilibrium) or balancing (i.e., regulate dynamics toward a new equilibrium). At the heart of the SD approach is the modeling of feedback loops through time using causal loop and stocks and flows diagramming.

SD has been used to model the economic impacts associated with disruptions to critical infrastructure. Dauelsberg and Outkin (2005), for example, calculate economic impacts as non-equilibrium events with critical infrastructure interdependencies captured through event propagation. Similarly, Portante et al. (2011) describes *EpFast*, a simulation and impact analysis tool developed at Argonne National Laboratory. *EpFast* is a suite of simulation models which focus on evaluating the vulnerability of energy infrastructure to events. They describe an ex ante application of the model to a high-intensity New Madrid seismic event of the US Eastern Interconnection. Kachali (2013) has further applied SD to study the 2010 and 2011 Canterbury (New Zealand) earthquakes. This work coalesce hazard, recovery, and organization disciplinary fields with primary data collected via surveys, case studies, and interviews of impacted organizations, to depict how systemic interactions within and between economic industries affect recovery post-disaster.

Agent-Based Modeling

Closely aligned with understanding the economic consequences of seismic events as complex systems is the use of ABMs. Thomas Schelling is often attributed with the creation of ABMs in the early 1970s (Schelling 1971). ABMs simulate the actions of autonomous agents with a view to assessing their consequences on the system as a whole. Unlike the conventional economic impact assessment approaches, which typically seek to describe movement from, or to, an economic equilibrium, ABMs capture the dynamic (equilibrium or otherwise) interaction of agents typically across space and through time. It is the emergence of patterns and trends in behavior, either as collectives of agents or for the system as a whole, which is of most interest to agent-based modelers.

The generative ability of ABMs is well suited to simulating the behavior of agents at the time during, immediately after, and in the recovery and reconstruction phases of a seismic event when economies are likely to exhibit out-of-equilibrium dynamics. Miles and Chang (2007), for example, developed an ABM which focuses not only on estimating losses but also on the anticipated recovery trajectories following an event. Specifically, the draw on empirical information from previous disasters, rather than regional economic theory, focused on spatial units or aggregate sectors, to model emergent trajectories of an economy. Their models are calibrated using the 1995 Kobe (Japan) and 1994 Northridge (California, USA) earthquakes. As another example, the US Department of Homeland Security applies a large-scale micro-simulation model, termed N-ABLE™, to capture complex supply chain, critical infrastructure interdependencies, and market dynamics associated with seismic events. N-ABLE™ is used primarily to gain insight into network vulnerabilities, economic security, and continuity.

Integrated Assessment Methods

An emerging area of investigation in the economic consequences of natural hazard events, including seismic events, is the use of integrated assessment methodologies. Such methods typically incorporate critical feedbacks between existing disciplinary models (e.g., scientific, engineering, planning, economics, and so on), thus, broadening the boundaries of investigation. They also tend to be spatially explicit in nature, applying geographical information systems to facilitate data entry and reporting. Feedbacks exist between land use, transport, and economic activity which are salient to understanding how communities recover from seismic events (Cho et al. 2001; Ham et al. 2005). Implicit in integration is the view that local communities, cities, and regions are complex systems requiring a holistic approach to their study. Blanco et al. (2009, p. 206) note a key recovery phase priority “consists of developing

methods to better understand how recovery is influenced by the underlying dynamics of change in the natural, built and human environments, notably including influences of planning (e.g., land use, built form) on disaster risk.”

Nonmarket Methods

Nonmarket valuation techniques are essentially concerned with measuring the extent to which changes brought about by an event either subtract from or add to values experienced by individuals. Nonmarket impacts of seismic events, such as changes to safety and security, environmental systems, human health, and cultural heritage, are often difficult to identify and measure. Nevertheless, nonmarket impacts are an important component of the range of effects resulting from seismic events. Noting that the classifications of nonmarket valuation methods vary between authors, three broad categories of nonmarket valuation techniques are identified:

- *Stated Preference Methods.* These methods utilize surveys or experiments to elicit what would hypothetically be paid for values which are not traded on markets. The most well-known stated preference methods are contingent valuation (CV) and choice modeling (CM). CV uses surveys to elicit the monetary value people are willing to pay to avoid a decrement of some type of nonmarket good or service or the value they are willing to accept in exchange for its deterioration. CM is similar to contingent valuation, except that willingness to pay is elicited from experiments in which respondents are asked to rank preferences for different attributes of nonmarket goods or services. Although CV is the more commonly used method, CM has gained popularity over recent years. In the context of seismic events, CV and CM have mainly been used to help ascertain socially optimum policies for risk reduction. Asgary et al. (2007), for example, apply CV to evaluate the benefits of hypothetical earthquake early warning systems, suggesting that their results

could be used by policymakers and businesses to determine the optimal level of investment in such systems.

- *Revealed Preference Methods.* These methods estimate the value of a nonmarket good or service based on observed behavior toward some closely connected marketed good or service. Among these the travel cost method is frequently used to infer the value of recreation, including experience of cultural heritage sites, based on the costs of travel to these sites. Although seldom employed in seismic-related studies to date, the method has the potential to evaluate nonmarket losses and justify the replacement of public goods and services. Loss in beach recreation value at tourist destinations in Thailand following the 2004 Asian tsunami is, for example, one possible application for the method. Hedonic pricing is concerned with inferring the value of individual attributes of a market commodity. In the context earthquakes, it has been used to ascertain the value people implicitly pay for housing with reduced seismic risk, or the loss in value of housing following changing perceptions of risk, based on information from real estate markets.
- *Cost-Based Valuation Methods.* These methods assume that some set of observed market costs can be used as a proxy for the value of an asset. The cost of illness approach has been utilized to determine expenditures on health care and lost wages from illness as a proxy measure of the physical injuries and psychological trauma of seismic events. Replacement cost or restoration cost methods may also be relevant to the valuation of nonmarket assets following an earthquake. Previous valuations of natural capital have included determining the costs of building infrastructure to perform similar functions. For example, many studies have examined the value of wetlands in improving water quality in terms of replacement costs.

Nonmarket valuation methods are subject to considerable limitations. Only stated preference methods are capable of capturing nonuse (e.g.,

option, bequest, and existence) values. However, these studies can be very difficult to implement and must be carefully designed to avoid biases. Additionally, the “free-rider” problem can create incentives for respondents to misrepresent real preferences. Travel cost and hedonic pricing methods are known to be sensitive to model specification, including choice of functional forms. Cost-based valuation methods are not considered to have a sound basis in welfare theory but nevertheless may be of use when there is limited time and resources available and only approximate estimates are necessary.

Despite the relatively widespread recognition of the opportunities for use of nonmarket techniques in the analysis of natural disasters, the number of actual seismic-related applications is still relatively few. Consequently, the use of “value transfer,” i.e., values estimated in one study applied to a later study, is common among practitioners of nonmarket valuation. Unfortunately, there are few guidelines available upon which to evaluate the appropriateness of value transfer methods.

Research Challenges

The economic impact typology outlined in Table 1, and the preceding description of assessment methodologies, provides a general picture of the approaches currently being employed in assessment of the economic impacts of seismic events. Nevertheless, a number of key challenges remain to be addressed.

Defining Model Scope and Reference Mode

Ultimately, the selection of impacts for assessment, and by corollary the choice of economic method used in evaluation, depends on the purpose of the assessment. As stated above, economic impact modeling is primarily undertaken with a view to maintain or enhance well-being; however, some impacts will be much more relevant to a particular audience than others. Insurance agencies, for example, will be mainly concerned with losses that are potentially insured. Local authorities and governments,

however, are charged with considering wider impacts on employment, poverty and inequity, fiscal balances, and so on. Nevertheless, some impacts that arise following a seismic event may be outside the scope of evaluation.

A further point to note in relation to the setting of model scope is that impact is a relative concept, identified by comparing an actual or a hypothetical trajectory of a disaster, with some kind of counterfactual baseline trajectory or reference model, i.e., that which would have occurred in the absence of the disaster. Either trajectory may include or exclude mitigation measures or response strategies, depending on the purpose of the analysis. Although many economic assessments concentrate on comparing situations before and after an event, this approach is only likely to be valid when considering short-run impacts. Economic systems are complex and constantly evolving over time; thus the longer the length of time for the analysis, the more difficult it is to define an appropriate reference mode, with many different alternatives possible. In a survey of businesses one and a half years after the 1994 Northridge (California, USA) earthquake, for example, one quarter of businesses had failed to recover to the level of operation prior to the event (Tierney 1997). Some of these businesses may, however, have experienced a state of decline even without the event. Similarly, following an event it has also been proposed that new or more productive economic sectors may emerge, assisted by the investment in reconstruction and opportunities to replace capital with newer and more productive capital (e.g., Skidmore and Toya 2002).

Determining Economic Impact Assessment Boundaries

It is very important in undertaken economic impact assessment of seismic events that the boundaries of the study are clearly defined and communicated with intended audiences. Generally speaking, boundary challenges fall into three categories: spatial, temporal, and distributional.

First, the spatial extent of any assessment requires careful consideration if impacts are to be appropriately accounted for locally,

regionally, and nationally. Until the mid-1990s most economic impact studies of seismic events focused mainly on the location directly impacted. Since the early 2000s, with growth in the development of multiregional IO tables and, more recently, spatial CGE models, spatial impacts have increasingly been incorporated in economic impact assessment. A number of studies (e.g., Okuyama 2004, Tatano and Tsuchiya 2008) note that there are often significant economic transfers between regions brought about through the relocation of people, business activity, and recovery operations. While an impacted area may experience significant economic losses in the period immediately following a seismic event, locations a little farther out may experience the opposite. Moreover, there are dynamic higher-order economic feedbacks which exist between regions, for example, competition for labor or capital between an earthquake-damaged area in recovery and an adjoining area experiencing growth. Inappropriate boundary selection can thus easily lead to the unintentional netting out of impacts.

Second, the economic consequences of seismic events are often felt most in the period immediately following an event but nevertheless continue into the recovery period and longer. Seismic events are also often felt as a series rather than as a single event, e.g., the 2010 and 2011 Canterbury (New Zealand) earthquakes. These characteristics change not only economic flows but also stocks, resulting in lags that lead to oscillation in key economic variables through time. The conventional market-based econometric, IO, SAM, and most CGE models require significant adjustments if such time-dependent processes are to be included. Attempts to include temporal dynamics into economic assessment have been made by Okuyama et al. (2004) in the specification of the sequential interindustry model, by Barker and Santos (2010) in the formulation of dynamic inoperability models, and by Rose and Liao's (2005) CGE model. It is worth noting that dynamic analysis is also a key strength of SD and ABM approaches. The temporal analysis of disaster impacts is, as Rose (2004) and Okuyama (2007) point out, still ad

hoc in nature – lacking an underlying theory to support their application.

Finally, a common criticism made of many economic impact assessments is that only aggregate or sector impacts are reported. There is however growing interest in understanding how economic impacts, both direct and higher order, are distributed across businesses and communities. Understanding economic consequences by socioeconomic and demographic variables including household, family, ethnic, and income types is potentially important to identifying socioeconomic vulnerability and, in turn, through mitigation and adaptation, building resilient businesses and communities. It is worth noting that SAMs, CGE, SD, and ABM models, if appropriately configured to include disaggregated household types (or similar), are ideally suited to addressing this challenge. Nevertheless, there is a current dearth of assessments that explicitly address distributional issues.

Building Confidence in Economic Impact Assessment

The validation of an economic impact assessment can only be judged against its intended purpose. Unfortunately, like all models of complex systems, there is no mathematical or absolute proof that an economic impact assessment may fulfill its purpose. Furthermore, we can conclude that any economic impact assessment is inherently imperfect. Like all models, any assessment represents a simplification of reality. It is not so much whether an assessment is valid or not but whether it is useful. Methodological pluralism is required when assessing the economic impacts of seismic events as there is “no one size fits all” assessment method. Nevertheless, confidence may be built in applying methods through the use of sensitivity and uncertainty analysis. Sensitivity analysis is essential for learning how general patterns of behavior exhibited by impact models change in response to key input parameters, narrowing down key drivers of change in a simulation. While sensitivity analysis may be used to identify the key exogenous drivers of change, uncertainty analysis adds value by

investigating model parameters through numerous experiments based on appropriate sampling methods, i.e., Monte Carlo.

Summary

In the context of responding to seismic risk and events, the behaviors of formal economic markets, alone, are unlikely to produce outcomes that are overall in the best interests of communities. Since the 1990s, economic modeling of seismic events, and their associated responses, has risen as an important tool in assisting policymakers in managing risk and adopting appropriate mitigation and response strategies. The application of economic methods to date can be grouped based on whether the assessment is *ex post* or *ex ante* an event and based on market or nonmarket analysis. The range of impacts that may be considered in assessments are vast, from analysis of only direct damages to physical assets to analysis of higher-order upstream and downstream business disruptions, competition for scarce materials during rebuild, and even valuation of destruction of natural capital and psychological trauma. Although conventional IO and econometric models dominated much of the early literature, during the last decade there has been a strong interest in extended IO, SAM, and CGE models. Modeling techniques applied in disciplines outside of economics, particularly SD, ABM, and integrative modeling approaches, are also now being used to assess the economic consequences of seismic events. Although the potential for use of nonmarket valuation techniques is also recognized, to date their uptake has been limited. It is important to note that “no one size fits all” method exists and that all methods are inherently imperfect. In this regard, there are many research challenges still to be addressed.

Cross-References

- [Community Recovery Following Earthquake Disasters](#)
- [Damage to Infrastructure: Modeling](#)

- ▶ **Economic Recovery Following Earthquakes Disasters**
- ▶ **Resilience to Earthquake Disasters**
- ▶ **Resourcing Issues Following Earthquake Disaster**

References

- Asgary A, Levy JK, Mehregan N (2007) Estimating willingness to pay for a hypothetical earthquake early warning systems. *Environ Hazards* 7(4):312–320
- Barker K, Santos JR (2010) Measuring the efficacy of inventory with a dynamic input–output model. *Int J Prod Econ* 126:130–143
- Blanco H, Alberti M, Olshansky R, Chang S, Wheeler SM, Randolph J, Watson V (2009) Shaken, shrinking, hot, impoverished and informal: emerging research agendas in planning. *Progr Plan* 72:195–250
- Cavallo E, Noy I (2009) The economics of natural disasters: a survey: IDB working paper series, no IDB-WP-124
- Cho S, Gordon P, Moore JE II, Richardson HW (2001) Integrating transportation network and regional economic models to estimate the costs of a large urban earthquake. *J Reg Sci* 41(1):39–65
- Cochrane HC (1997) Forecasting the economic impact of a midwest earthquake. In: Jones BG (ed) *Economic consequences of earthquakes: preparing for the unexpected*. National Center for Earthquake Engineering Research, Buffalo, pp 223–247
- Crespo Cuaresma J, Hlouskova J, Obersteiner M (2008) Natural disasters as creative destruction? Evidence from developing countries. *Econ Inq* 46(2):214–226
- Dauelsberg L, Outkin A (2005) Modeling economic impacts to critical infrastructures in a systems dynamics framework. Paper presented at the 23rd international conference of the system dynamics society, Boston
- Haimes YY, Horowitz BM, Lambert JH, Santos JR, Lian C, Crowther KG (2005) Inoperability input–output model for interdependent infrastructure sectors. I: theory and methodology. *J Infrastruct Syst* 11:67–79
- Hallegatte S (2008) An adaptive regional input–output model and its application to the assessment of economic cost of Katrina. *Risk Anal* 28:779–799
- Hallegatte S, Przyluski V (2010) The economics of natural disasters: concepts and methods. Policy research working paper 5507. The World Bank, Sustainable Development Network, Office of the Chief Economist, Washington, DC
- Ham H, Kim JT, Boyce D (2005) Assessment of economic impacts from unexpected events with an interregional commodity flow and multimodal transportation network model. *Transp Res Part A Policy Pract* 39(1): 849–860
- Hochrainer S (2009) Assessing the macroeconomic impacts of natural disasters – are there any? vol 4968, World Bank policy research working paper. The World Bank, Washington, DC
- Jaiswal K, Wald DJ (2013) Estimating economic losses from earthquakes using an empirical approach. *Earthq Spectra* 29(1):309–324
- Kachali H (2013) Key elements of sectoral and resilience after the Canterbury earthquakes: a system dynamics approach. PhD, University of Canterbury, Christchurch
- Leontief W (1941) The structure of the American economy, 1919–1929. Harvard University Press, Cambridge, MA
- Miles S, Chang SE (2007) Simulation model of urban disaster recovery and resilience: implementation for the 1994 Northridge earthquake: MCEER-07-0014
- Okuyama Y (2004) Modeling spatial economic impacts of an earthquake: input–output approaches. *Disaster Prev Manag* 13(4):297–306
- Okuyama Y (2007) Economic modeling for disaster impact analysis: past, present, and future. *Econ Syst Res* 19(2):115–124
- Okuyama Y, Hewings GJD, Sonis M (2004) Measuring economic impacts of disasters: interregional input–output analysis using sequential interindustry model. In: Okuyama Y, Chang SE (eds) *Modeling spatial and economic impacts of disasters*. Springer, Berlin
- Portante EC, Craig BA, Talaber Malone LE, Kavicky JA, Folga SF, Cedres S (2011) A model for simulating impacts of seismic events of large power systems. Paper presented at the EAIA'11 proceedings of the 2011 emerging M&S applications in industry and academia symposium, San Diego
- Rose A (2004) Economic principles, issues, and research priorities in hazard loss estimation. In: Okuyama Y, Chang S (eds) *Modeling spatial and economic impacts of disasters*. Springer, Berlin, pp 14–36
- Rose A, Guha G-S (2004) Computable general equilibrium modeling of electric utility lifeline losses from earthquakes. In: Okuyama Y, Chang S (eds) *Modeling spatial and economic impacts of disasters*. Springer, Berlin, pp 119–138
- Rose A, Liao S-Y (2005) Modeling regional economic resilience to disasters: a computable general equilibrium analysis of water service disruptions. *J Reg Sci* 45:75–112
- Schelling T (1971) Dynamic models of segregation. *J Math Sociol* 1(143–186)
- Skidmore M, Toya H (2002) Do natural disasters promote long-run growth? *Econ Inq* 40(4):664–687
- Tatano H, Tsuchiya S (2008) A framework for economic loss estimation due to seismic transportation network disruption: a spatial computable general equilibrium approach. *Nat Hazards* 44:253–265
- Tierney KJ (1997) Business impacts of the Northridge earthquake. *J Conting Crisis Manag* 5(2):87–97

Economic Recovery Following Earthquakes Disasters

Erica Seville¹, John Vargo² and Ilan Noy³

¹Resilient Organisations, Sheffield, New Zealand

²Resilient Organisations, University of Canterbury, Christchurch, New Zealand

³School of Economics and Finance, Victoria University, Wellington, New Zealand

Synonyms

Business disruption; Business interruption; Consumer; Economy; Employment; Fiscal impacts; Macroeconomic impact; Microeconomic impact

Introduction

Earthquakes can affect economies and economic activity in a number of ways. Directly, an earthquake can damage assets and infrastructure critical for the production and delivery of goods and services. Earthquakes can also directly affect human capital, impacting the availability and productivity of the workforce. These however are not the only ways in which earthquakes impact an economy; indirect (or secondary) effects of the disaster can magnify the disruption caused by direct damage. Here, we describe the ways in which earthquake disasters can impact an economy, starting with macroeconomic impacts (how the local economy as a whole is affected) and then focusing in on microeconomic impacts (by looking at how different actors within an economy can be affected). We conclude by looking at ways in which economic recovery can be supported following a major disaster.

Macro-impact of Disaster and Recovery

The economic research on the process of recovery in the aftermath of natural disasters is mostly of recent vintage. The typical view of a recovery process is that the initial impact entails

a downturn in economic activity, as much productive capacity and infrastructure are destroyed and supply networks are severed. However, this initial period of difficulty is assumed to be followed by a period of enhanced activity, largely as a result of additional spending on reconstruction of infrastructure and housing.

The increased spending on reconstruction creates additional economic activity as these resources are cycled through personal consumption and investment decisions. The exact magnitude of this post-disaster boom in economic activity depends both on the quantity of the additional spending, the ways in which this spending is financed, and on the magnitude of the multiplier effect (the multiplier is the beneficial feedback loop from current spending that generates income for others, which is then cycled back into the economy as future spending). Thus, a post-disaster wage subsidy, for example, provides additional income to wage workers, who then spend at least some of this additional income on consumption goods (food, hospitality services, durable goods, etc.). The recipients of this additional spending in turn increase their own spending as their incomes increase, and the money keeps being recycled through the local economy.

Once the reconstruction phase is complete, the economy returns to its “normal” state, typically assumed to be the one that would have prevailed had the disaster itself not happened. This view can be traced at least back to Adam Smith’s description, in the *Wealth of Nations*, of the process of recovery in Swiss cantons that experience periodic flooding.

This view was largely unexamined in the economic literature until the last decade. The empirical evidence that has now begun to surface (Cavallo and Noy 2011) includes several empirical regularities that are described in the sections below. Many questions about the precise description of a typical post-disaster recovery, as well as the idiosyncratic and contingent differences between different recoveries, however, remain open.

Some of the regularities are the following: In the immediate aftermath of large disasters, low- and middle-income countries typically experience some slowdown in economic activity, but

the magnitude of the slowdown is only partly a function of the nature of the disaster (e.g., hydrometeorological vs. geophysical), its spatial extent, its magnitude, and its timing (e.g., with relation to the crop cycles). The slowdown is also, maybe mostly, a function of the economic vulnerabilities of the affected region (e.g., the degree and nature of diversification in the economic sectors that dominate the affected region).

In high-income economies, however, even large disasters seem to have very little macroeconomic aggregate effects. Thus, even a megadisaster, such as the triple earthquake-tsunami-nuclear event in Japan in 2011, has had only very temporary impact on the aggregate Japanese economy. In the long run, it appears that disasters have little measureable impact on the national economy (Cavallo et al. 2013).

Impacts on prices are not well documented. The scant evidence that is available appears to suggest that goods and services that are directly required for the reconstruction efforts seem to experience price increases as a consequence of increased demand and limited supply. Most typical is an increase in the cost of construction and construction workers' wages. These increases do not seem, however, to translate into more general price increases unless the event itself generated specific bottlenecks and supply constraints, e.g., the destruction of oil refineries in the Gulf of Mexico during hurricane Katrina of 2005.

Fiscal Impacts of Disasters

The fiscal patterns are clearer, and maybe more obvious. Disasters generally entail an increase in spending and a decrease in tax revenue, as some of the new reconstruction spending is paid for by the public sector, and the supply bottlenecks and destruction of productive capital reduce economic activity, and hence taxable income. As already noted, the decline in aggregate economic activity is temporary, and thus the observed decline in tax revenue also does not endure unless tax rates are different across economic activities or the disaster itself leads to significant shifts in tax policy.

Public spending on reconstruction, in contrast, can endure as the duration of the reconstruction efforts can last for more than a decade. Thus, the observed increase in the budget deficit, and the consequent increase in government borrowing, is largely due to increased spending rather than decreased revenue.

This picture, however, does not describe accurately the ways in which very poor countries have to conduct their fiscal policy in the aftermath of a catastrophic disaster. Low-income countries have less access to credit markets; and their high cost of borrowing and their narrow tax base prohibit them from spending adequately on reconstruction. Yet, even in these cases when the government is not able to spend on an aggressive recovery effort, this failure does not necessarily imply reduced public sector deficits as spending does not increase. In these cases, the inability to spend on recovery implies that incomes do not recover, and the tax revenue is decreased. This is frequently compounded by the fact that in poor countries, much tax revenue comes from taxes on imports, and imports typically decline precipitously when incomes are significantly decreased. Thus, in both high- and low-income countries, earthquakes will entail increasing government budget deficits. It is only in rare cases, when, for example, much of the damage was privately insured or reinsured internationally that the government finances may not suffer much (e.g., Doyle and Noy *in press*).

Consumption and Trade

Since government spending and public and private investment will tend to increase in the aftermath of large disasters, consumption and/or the trade balance has to decrease (as the sum of these four components is equivalent to national income in the national income accounts). Both of these declines are well documented.

Research that has examined household spending patterns in post-disaster environments generally concludes that one can observe decreases in spending that are more concentrated in poorer households and that the decrease

in spending frequently is concentrated in luxury goods (less surprising) and health and education (clearly more worrying as this decrease may entail long-term adverse consequences). This pattern is not unique to low-income countries but is also documented, for example, for households in the Japanese city of Kobe after the 1995 earthquake (Sawada and Shimizutani 2008).

For trade, it is not unusual to observe declines in imports, and especially in imported luxury goods, but these are typically offset by increases in importing of building materials, fuels, and other goods associated with the reconstruction effort. After an initial decline in exports, as the productive capacity of the affected region is damaged, the overall and longer-term impact of a typical earthquake shock on exports is largely contingent on the types of exports the country specializes in and the nature of the earthquake damage to production facilities, transportation infrastructure, etc. In some cases, one can potentially observe a counterintuitive increase in exports, especially if the earthquake disaster led to a depreciation of the currency, making exported goods more affordable for foreign buyers.

While predicting the typical impact of an earthquake on exports of goods is fraught with uncertainty, exports of services are almost always harmed by a disaster. Especially vulnerable is the tourism sector, as the tourist trade is mostly dependent on image branding and perception rather than on the actual capacity to supply services. Thus, tourism declined significantly in places where earthquake damages occurred in tourist areas (Sri Lanka and Southern Thailand in 2004 or Japan, Chile, and New Zealand in 2011).

Regional/Local Economic Impacts of Disasters

While the aggregate impact of disasters at the national level may be muted, as productive capacity and labor are able to relocate away from the disaster-affected areas, the impact of the event locally may be felt quite differently.

The limited evidence seems to suggest potentially very long-term adverse impacts at the local level that are quite distinct from the relatively benign impacts at the national level, even for fairly short horizons. At the very extreme, some regions never recover at all from a large event. Such an extreme example is the island of Kefalonia, in which a devastating earthquake in 1953 led to mass exodus from the island, and the population on the island was permanently reduced by more than 20 %. While such an extreme case is unusual, earthquakes and other geophysical events can lead to a redistribution of population and economic activity that may persist for long periods of time.

The exact nature of the long-term impact of an earthquake on the regional/local economy depends on the structure of the economy of the region. For example, the agricultural sector is relatively resilient to earthquakes, and even if the earthquake damages some of the processing and transportation infrastructure, this can be quickly rebuilt. Other sectors may be more vulnerable, especially if these were already facing difficulty before the event. In the city of Kobe, for example, the heavy industry of the region was already facing increasing competition from elsewhere in Asia (in particular from China) before the 1995 earthquake. The earthquake itself led to a speeding up of a process that would have potentially occurred even without the earthquake. Much of the heavy industry was not rebuilt and reopened in Kobe in the years after the quake, in spite of massive government assistance for the region.

Paying for Reconstruction

The ways in which post-disaster reconstruction is financed is an important aspect of the recovery process. In general, public infrastructure investment during the reconstruction effort can be financed by: government borrowing (debt financing), government printing of money (the so-called seigniorage tax: the profits the government received from printing money and then using it), international aid, and insurance

Economic Recovery Following Earthquakes Disasters, Table 1 Earthquake losses insured by country for some recent earthquakes (Source: Swiss Re sigma catastrophe database)

Date	Country	Losses	% GDP	Insured losses %
April 2009	Italy	\$4B	0.2	14
January 2010	Haiti	\$8B	121	1
February 2010	Chile	\$30B	18	27
April 2010	Mexico	\$0.95	0.1	21
September 2010	New Zealand	\$6B	5.3	81
February 2011	New Zealand	\$15B	10	80
March 2011	Japan	\$300B	5.4	17
October 2011	Turkey	\$0.75B	0.1	4

(explicit insurance, insurance-like financial instruments such as CAT bonds, or more informal insurance arrangements). Some of these channels may be more efficient and lead to more equitable distribution of resources than others. It is also quite difficult to assess the need before an event and the implementations after an event of these different funding strategies, as government spending on recovery efforts is typically only a small part of overall government spending, and the coincidental benefits and drawbacks of this funding are difficult to attribute.

In recent decades, governments rely less on seigniorage as a means of paying for a post-earthquake reconstruction and appear to favor borrowing. International aid is typically not that large, even after very catastrophic events, and in many cases much of it is not channeled through the government (e.g., Katz 2013).

Investment by the private sector, and in particular business investment, can be financed by insurance receipts but also by previously accumulated savings, by remittances received from abroad, or through borrowing on international or domestic capital markets. The ability of businesses to recover from the initial destruction associated with the earthquake is, in addition to other things, partly a function of its ability to invest in reconstruction and may also be a function of the type of financing available. Yet, our knowledge of any general conclusions about this process is quite limited – see an exception on post-tsunami Sri Lanka in De Mel et al. (2012).

Households' ability to finance their own recovery is also dependent on the availability of financing. It appears that households that have access to borrowing, or to resources from insurance policies or remittances, are able to recover more quickly and more fully than households whose access to credit is constrained (e.g., Sawada and Shimizutani 2008).

It is worth noting that the degree to which households are insured against earthquake risks differs dramatically across countries. In California, for example, earthquake insurance is bought separately and is very expensive, so that most households are not insured. In contrast, in New Zealand affordable earthquake coverage is part of a standard residential housing insurance policy and is partly handled by a public insurance entity backed by the government (the New Zealand Earthquake Commission). Earthquake risk in New Zealand is thus almost fully insured at the residential level. In most middle- and low-income countries, earthquake insurance policies are either unavailable or prohibitively expensive. See Table 1 showing a comparison for recent earthquake events by country of information collated by Swiss Re. Swiss Re reports financial losses, both insured and uninsured, attributable to the event, including physical property damage, and business interruption as a result of property damage caused by the event. Indirect losses are not included in estimates of total economic losses. Sources for their information comes from newspapers, direct insurance and reinsurance periodicals, specialist publications, and reports from insurers and reinsurers.

Changes in Consumer Behavior

Large natural disasters, including earthquakes, can result in significant shifts in population and changes to consumer behavior, all of which affect the potential for economic recovery.

Large losses of housing can lead to permanent losses of population. Past experience has shown that even when less than 10 % of housing units are lost, the time to repair or rebuild uninhabitable housing units can still be very long (SPUR 2012). In the absence of adequate interim housing, some outmigration is likely, and the loss of population will impact the local economy and overall recovery. Permanent loss of population, and the accompanying social and economic impacts, can occur in communities where housing losses are as low as 5–10 % of the total housing stock (SPUR 2012).

Loss of community facilities (such as access to health-care and education facilities) and disruption to critical infrastructure (such as power, water, sewage, and transportation services) can also trigger certain parts of the population to either temporarily or permanently leave the area. For example, following the Canterbury earthquakes in New Zealand, many pregnant women and people with young children chose to leave the area, many to stay with family and friends in other parts of the country, to get away from the ongoing aftershocks, and to wait for services to be restored.

Population loss however is not an inevitable consequence of disasters, and the incoming workforce to undertake the reconstruction effort can lead to an overall increase in population. The demographic mix of the community however can evolve over this process as the characteristics of the incoming population differ from those who leave.

These rapid changes in both the number and mix of people living in a community have implications for organizations operating within the region, as inevitably the customer base and their needs change. Localized areas of damage or disruption to transportation networks can see marked shifts in where people shop. In any disaster, there are always winners and losers, and

sometimes this can be dictated by simply which part of town an organization is operating from. In a post-disaster environment, organizations need to revisit their business plans to identify what strategies will work in their new environment. For example, organizations may need to shift from in-store to online retailing or from retail to whole-sale operations; they may also need to change their mix of product lines to meet changed customer needs. Adapting to the changed circumstances and finding new ways to thrive in the post-disaster environment is vital for business recovery.

While some sectors are more vulnerable than others to natural disasters, the type of industries that are most affected by a natural disaster depends on the nature of the specific event and the industrial structure of the affected region (Venn 2012). Some particular examples noted by Brown et al. (2013) relating to sector-specific earthquake vulnerabilities include:

- Manufacturing businesses often have specialist machinery and premises making it difficult to find suitable alternative premises in the event of property damage. If specialist machinery is damaged, there may be a long lead time for replacement equipment.
- Tertiary education and tourism sector organizations are reliant on customers from outside the affected areas and can be affected by a perception that the area is no longer a fun or safe place to visit.
- Retail and hospitality can suffer from both drops in visitor numbers and changes in discretionary spending by locals.
- The vulnerability of the agricultural sector is very seasonally dependent, with an earthquake having potentially minimal impact during some times of the year and more severe impacts at others.

In addition, Stevenson et al. (2011) found that businesses located in central business districts or densely populated areas face a greater chance of disruption due to neighborhood effects, with greater potential for access problems due to damage because of greater building densities and/or higher concentrations of historic buildings.

Workforce Issues

Common disaster effects on labor markets include: job and worker displacement, loss of income, and additional participation or entry barriers into the workforce. In addition, labor markets can be fundamentally changed due to restructuring of the economy as a consequence of the disaster itself, or during the reconstruction.

Current understanding of the impact of disasters on labor markets largely relies on indicators of labor market outcomes, such as job losses, labor-force participation, and unemployment. It can be difficult however to distinguish disaster-related effects on the labor market from those of other shocks and developments which occur simultaneously.

In general, natural disasters tend to create large-scale shifts in the demand for skills. This can lead to a paradox of skill shortages coexisting with relatively high unemployment and low rates of labor-force participation. In some cases, there is a marked decline in employment in certain industries while demand for other services, particularly construction, increases. Apart from the sectoral vulnerabilities, women and young people are often most negatively affected. Female workers tend to be badly affected as they are less likely to be able to take advantage of new job opportunities during the reconstruction phase. Time needed for construction training means young people, particularly school leavers, have less chance to take up immediate employment in construction.

The way natural disasters can disrupt labor supply is by displacing people from their normal lives and jobs. Damage to property and infrastructure often force people to evacuate to areas outside the disaster zone. According to Venn (2012), mass evacuations can lead to severe labor market disruptions, making it difficult for evacuees to retain their pre-disaster jobs and putting a strain on local labor markets in the areas to which people have been evacuated.

Businesses are the core medium through which a workforce participates in the labor market. Businesses may be subjected to temporary or permanent closure as they deal with physical damage and loss of customers. Some businesses

will be able to reopen quickly, while others may take much longer or may never be able to reopen.

Family and health issues can have negative impacts on a worker's performance. In a survey of over 300 businesses in the Canterbury region of New Zealand following the September 2010 earthquake, more than a quarter of respondent organizations reported that managing staff well-being in the weeks following the earthquake is the biggest challenge affecting their organizations (Kachali et al. 2012). Heightened emotional strains can be related to other factors such as disruption to childcare arrangements, disruption to public transportation, the slow process of receiving insurance payouts and repair information, and a lack of suitable accommodation while houses are being repaired or rebuilt. While businesses are trying to reopen, relocate, and return to normal, staff may be struggling with personal recovery demands.

Post-disaster labor market indicators show that many of the people who are temporarily displaced from jobs soon find new jobs on their own, either in the same area or elsewhere, but help is needed to manage their transitions back to the labor market. It is a complex undertaking for policy makers, firstly, to identify these groups and individuals and, secondly, to provide the support that is needed to address the barriers to their labor-force participation.

Research looking at workforce policy post-disaster suggests the importance of a multifaceted approach to post-disaster labor market response by combining short-term job creation with longer-term skills development and combining short-term job protection with longer-term economic promotion strategies (Chang-Richard et al. 2013).

Common building blocks of "best practice" identified within APEC economies for managing post-disaster labor markets are shown in the Table 2 below.

Economic Stimulus Effects of Reconstruction

In a post-disaster environment, there is strong pressure to act quickly to get back to normal.

Economic Recovery Following Earthquakes Disasters, Table 2 Principles for natural disasters' sound workforce strategies (Chang-Richard et al. 2013)

Principles	Guidelines
1. Target and tailor measures with flexibility	Employment strategies (e.g., public works, employer subsidies) should be targeted for the most troubled sectors and/or needed groups and individuals; if fiscal capacity is constrained, resources should be directed to sector groups that have the most potential to generate employment
	Social protection measures should be as flexible as possible to meet the needs of beneficiaries, reducing application "red tape," providing easy access, extending the coverage and length of support
2. Institutional and organizational innovation	An overarching recovery strategy with employment being one of key components provides a framework for an efficient, responsive labor market response
	A truly dedicated response team in respective policy domains (regional development organizations, rebuilding advisory services, skills specialist officers) will drive efficiency and higher levels of engagement from the industry and community
	One-stop design is critical to the success and accessibility of employment services
3. A diversity of multiagency collaboration	A diversity of collaboration among multiple agencies is essential to ensure coverage of assistance and combination of capacity (financial, technical, and "know-how") for effective service delivery
	Expertise and "know-how" knowledge of nongovernment organizations, particularly the International Labour Organization (ILO), should be drawn on, promoting a collaborative approach to supporting employment and livelihood assistance
4. A focus on sustainable employment outcome	A commitment to sustainable employment outcomes requires social protection measures to be coupled with training and skills development programs, job creation activities to be combined with productivity and skills enhancement, and disaster labor market response to be integrated in regional workforce development strategies
	Workplace-based employment retention and livelihood support play an important role in overcoming workers' participation barriers
	Self-efficacy and employability concepts should be built into the design of self-employment assistance and training programs, respectively, with a focus on linking support measures with direct employment outcomes
5. Capacity-based preparedness and response	Predesigned labor market response policies which can be quickly activated ensure timeliness and efficiency of employment assistance
	Existing networks and relationships, together with a strong leadership, allow resources and capabilities to be quickly assembled and/or redirected in response to the changing needs over the recovery time
	Fiscal, technical, staffing and technological capacity is essential to ensure efficiency within service delivery
6. Local input and place-based solutions	Economic and business recovery should increase the engagement of local organizations and people with local and industry knowledge
	Employment assistance provided to displaced workers and job seekers, particularly those disadvantaged (youth, women, people with disability, long-term unemployed), can be based on their individual needs and opportunities identified by themselves
7. Robust labor market intelligence	An evidence-based approach, building on what has proved to be good practice in international experience is instrumental in designing a similar measure. Business surveys will allow for more responsive and targeted assistance
	Labor market programs should be based on robust labor-force information which will contribute to a sustainable and resilient employment services system

The need to replace lost homes, infrastructure, and commercial properties generates a surge in demand for reconstruction that can act as a short-term stimulus to the local economy. This stimulus benefits not only the construction sector but also other sectors that directly support the construction sector (such as lawyers and accountants, office-supplies providers, and accommodation providers) as well as those sectors that indirectly benefit from the wages and salaries paid to those working on the rebuild (such as the hospitality sector, retailers, and arts and recreation providers).

While demand surge in the construction market can provide a buffer to an economy affected by disaster, it can also create imbalance and dysfunction. The key characteristic that distinguishes the post-disaster condition from normal times is time compression which requires an unusual pace of capital expenditures and reconstruction (Olshansky et al. 2012). In many circumstances, this leads to shortages in capital (machinery, trucks, diggers, etc.), materials (concrete, steel, timber, etc.), and skills (engineers, builders, plumbers, painters, etc.) (Chang-Richard et al. 2012). Such resourcing bottlenecks can lead to price increases and inflows of large amounts of money, and new players in the construction market can also lead to increases in profiteering, corruption, and fraud.

Business Recovery

Business recovery following an earthquake is a crucial aspect for community recovery. Businesses provide employment, goods and services, as well as social support for their employees and for the community at large by way of gathering places. Key aspects impacting on business recovery include levels of earthquake damage to their own premises and surrounding buildings, whether they have to relocate, disruption to their supply chain, changes in consumer behavior, and the impact of infrastructure disruption (Tierney 1997).

In the immediate aftermath of a damaging earthquake, a business owner may need to make the decision to relocate their business due to

damage or safety cordons that have been put in place. This is a major decision for an organization as it may contribute to employee decisions to stay or leave as well as impacting on customer flows through patterns of foot traffic and agglomeration effects (where similar businesses are colocated thus increasing the customer traffic for all of them). The decision to relocate is a complex decision and is complicated further by such factors as: does the business own, rent, or lease the building; level of damage to the building; damage to the neighboring buildings and the presence of a safety cordon (e.g., Chang et al. 2014) preventing access to the building or to nearby buildings; and availability of infrastructure services.

If the business owns the building, it may have more control over building access, whereas the landlord or the landlord's insurer may prevent a business in rented premises from gaining access. This also applies to speed of repairs, how insurance payouts are used to repair or upgrade the building, and the disruption this has on access to the building. These decisions are typically in the hands of the owner (with some constraints from emergency services and the insurer), while a renter or lessee may find that they have limited influence over such decisions.

Level of damage to the building clearly has implications for the relocation decision. A badly damaged building is a safety and reputational risk to the business and will typically force a closure or relocation. Less obvious however is the situation where the business' building is undamaged but where nearby buildings are badly damaged, closed, and cordoned. In this situation, there is a neighborhood effect in which customer numbers drop significantly as they see the badly damaged neighborhood and assume that all businesses in the area are impacted equally (Chang and Falit-Baiamonte 2002). This situation is particularly difficult for the undamaged business as business interruption insurance is typically dependent on physical damage to trigger a payout. Since they are not physically damaged, there may be no payout, but their cash flow is definitely impacted due to the neighborhood effect.

Availability of infrastructure services such as power, water, and wastewater is a crucial feature for business recovery and the decision to relocate. A building may be undamaged and the business owner desiring to reopen as quickly as possible, but the lack of services may make this impossible. In addition, if infrastructure services are functioning, but at a lower than normal level (e.g., electricity “brownouts” or poor quality water), this may have an impact on ability to conduct business at a competitive quality and undermine the rate of customer return and loyalty to the business.

In the aftermath of an earthquake, it is quite possible that a business and its neighborhood are largely undamaged with full services working. However, there may be significant impacts on their supply chain, both upstream (disrupted suppliers) and downstream (customer impacts and behavior changes). A supermarket may be fine following an earthquake with only minor impacts on their part of town, but their primary supplier’s warehouse may have collapsed and deliveries to the supermarket heavily disrupted, so customers go to other supermarkets and get in the habit of doing so. Even after the supplier can once again deliver reliably (or new suppliers are brought on-stream by the supermarket), customers may be slow to return due to maintained perception and new habits.

Business insurance is a two-edged sword. A swift settlement at expected levels of payout can be a lifesaver to a business, allowing rapid repair or replacement of the building, inventory, and equipment, while also providing much needed cash flow through a business interruption policy. This is a likely scenario if the earthquake damage is not widespread and the business owner has a thorough understanding of their insurance policy and a good relationship with their insurer. Unfortunately these provisos are often not met. Damaging earthquakes tend to be regional in nature resulting in widespread disruption. As a result, there are a large number of claims that the insurers are attempting to deal with at the same time. Insurers become overwhelmed with the consequent workload with the result that settlements are slow in coming, require much

attention from the business owner, and endanger the business further as the owner may focus too much effort on the insurance difficulties and too little on keeping their business running effectively (Brown et al. 2013).

Government Policy Impacts on Economic Recovery

The impact of government policy on economic recovery can be significant. Policies such as national or regional earthquake insurance schemes, direct government loans and grants, employment subsidies, and the establishment of government recovery bodies are all examples of government interventions that can be hugely beneficial or undermine long-term recovery if misguided.

National or regional earthquake insurance schemes can provide much needed capital to an earthquake-hit region following a major event. Examples include the New Zealand EQC (Earthquake Commission), the Japanese JER (Japanese Earthquake Reinsurance) scheme, and the California CEA (California Earthquake Authority). These schemes were all formed through government legislation but vary considerably on how they function. Characteristics that differentiate them include: whether policies cover both residential and commercial buildings, whether policies are mandatory for building owners, the extent to which the government acts as insurer of last resort (it covers further losses once the scheme’s resources are exhausted), and who is responsible for administering the scheme (public or private entities).

An alternative to earthquake insurance schemes is direct government loans and grants following an earthquake. In many countries, earthquake insurance is not available as part of a building owner’s policy and may be unavailable in any form. In countries such as Italy or China, low levels of earthquake insurance are the norm (Asian Development Bank 2008). Where this is the case, a government may still intervene by reconstruction of homes and infrastructure and, much less often, commercial buildings. This

intervention may come in the form of direct government intervention or in the form of grants and loans to the building owners. Such intervention is a form of national self-insurance. This provides the capital to rebuild regionally, while sharing the economic burden with the rest of the country. This risk sharing is of course a characteristic of any insurance scheme, whether explicit or implicit.

Following an earthquake, governments may intervene to support business directly. This may include business recovery loans (such as from the US Small Business Administration) or through employment subsidy schemes (such as the Earthquake Support Subsidy used following the Christchurch earthquakes in New Zealand). The benefit to such government intervention may be the long-term retention of businesses in a region (Hatton et al. 2012). This will typically have a number of positive flow-on effects including higher employment, reduced risk of population dislocation, and generally an increase in the speed of economic recovery. The potential drawbacks to such direct government intervention comes when the intervention is longer term (thus creating a business-government dependency), where investment is inefficient (keeping businesses alive that were on the verge of failing prior to the earthquake, and the government grant/loan simply keeps them alive until the aid runs out), or where the intervention has insufficient accountability resulting in budget blowouts and increased potential for corruption.

In very large events, the national/central government may form a Recovery Authority such as the Canterbury Earthquake Recovery Authority (CERA) in New Zealand following the 2010–2011 earthquakes or the Badan Rehabilitasi dan Rekonstruksi (BRR) in Indonesia following the 2004 earthquake-tsunami in Aceh. Such an intervention provides a central authority to coordinate services, aid, and policy during the difficult recovery period. This will more typically be the case where the event overwhelms local government capabilities or where central government is providing significant direct funding to support the recovery (e.g., in rebuilding housing and critical infrastructure), getting

local government back on its feet, and supporting local social needs. In a few cases, even the central government may be overwhelmed by the reconstruction needs; that was the case in Haiti after the 2010 Port-au-Prince earthquake when the Interim Haiti Recovery Commission led by former US President Bill Clinton was established jointly by the United Nations and the Haitian government.

Metrics of Economic Recovery

How does a city, a region, or a country know they have recovered? What are the ways of measuring the progress from devastation to normality? The complex and multifaceted process of economic recovery has winners and losers, and it has waves of progress and ebbs of disappointment. Metrics that help us to see and measure the progress in population change, consumer spending and economic growth, extent of business relocations, and levels of reconstruction all contribute to the tapestry that is economic recovery.

A key determinant of economic recovery is the state of the population. If there was significant population dislocation and loss, to what extent has the population returned and resettled? If population did not leave, will the population be retained through the difficult recovery period? As noted earlier in this entry, an important feature of population recovery is the employment opportunities available. Other key measures are availability and quality of utilities and amenities (power, water, communications, shopping centers, schools, and higher education) and direct measures of population gain and loss (immigration, emigration, growth of nearby towns, number of working visas issued, etc.). Each of these can be measured and contribute to the complex picture of economic recovery (Table 3).

The condition of a city or region's critical infrastructure is a primary element in economic recovery. Earthquakes are damaging to road, power, water, and telecommunication networks. The recovery of these services can be measured in terms of total level of damage (in physical and monetary terms) and the progress of

Economic Recovery Following Earthquakes Disasters, Table 3 Metrics of economic recovery

Description	Measure
Business birth and death rate	Percentage of total number of businesses
Direct population statistics: immigration, emigration, working visas issued	Number of people and percentage of base population
Levels of employment and unemployment	Total numbers employed and unemployed, workforce participation rates, and percentage of total population of working age
Critical infrastructure reconstruction and recovery	Total km to be repaired, total monetary value of losses, km of repairs now complete, percentage of reconstruction complete in km and dollars. Percentage of rebuild complete
Residential and community building reconstruction	Number of damaged houses, number of houses repaired, demolished, or rebuilt. Percentage of total work completed
Commercial building reconstruction	Number of damaged buildings, number of buildings repaired, demolished, or rebuilt. Percentage of total work completed. Similar but using square meters of floor space rather than number of buildings
Business growth and stability	Total sales in retail and wholesale, sector-specific measures of growth, profitability, and employment
Levels of business relocation	Percentage of businesses in temporary accommodation
Overall measures of economic recovery	Regional GDP trends, total tax take, retail sales, and employment (as noted above)

reconstruction (e.g., number of km of water pipes repaired and percentage of total work completed). Similarly the reconstruction of damaged residential and community buildings (e.g., schools, churches, and other religious buildings) and commercial buildings is important for the economic recovery as well as the flow-on effects to community recovery.

An important measure of economic recovery is the state of business growth, stability, and

profitability. Are levels of temporary relocation still high? Are agglomerations (groups of like businesses such as a fashion district) recovering and regathering? More general economic measures such as levels of exports, total retail sales, and other measures of economic activity round out the broader picture of business recovery. Measures that capture the state of the various economic sectors will also highlight the winners and losers in the aftermath of earthquakes and provide a picture of sectoral recovery that can inform various business and government interventions. For example, tourism and education are export sectors that are often hard hit following earthquake events. Interventions to support the recovery of such sectors (e.g., collaborative international marketing programs) can break the cycle of bad news and accelerate the recovery of such hard-hit sectors.

The growth or shrinkage of an economy can be measured in a multifaceted way to gain a rich picture of general economic recovery. This general economic recovery is dependent on all of the earlier characteristics of population, state of amenities, business stability, and employment and levels of reconstruction. The state of the economy can of course be measured directly as well. Growth of regional GDP, levels of employment, tax revenues, and total retail spending are all useful measures of the state of the overall economy, and the trends on these measures provide a moving picture of the recovery.

Summary

The impact of a large earthquake on a regional economy can be dramatic and complex. It may have devastating short-term regional effects while having limited impact on the national accounts of the country. There are inevitable winners and losers both during the initial response to the disaster and during the long and complex recovery period. The assumed trajectory of recovery: a drop in economic activity in the immediate aftermath of the event followed by a boom during the reconstruction phase may be a reasonable description in relatively richer

countries. However, the picture for local businesses may be far more complex. They must deal with the economic vulnerabilities that existed prior to the quake compounded by the impacts of the disaster, including the need to relocate, loss of key staff, changes in consumer behavior, disruption to their supply chain, and overwhelmed insurers.

In the midst of the challenges faced locally will be opportunities for those able to tap into the changes caused by the disrupted economy including location of consumers, changed demographics, insurance payouts, geographic dislocations, and other changes. Then there is the inevitable reconstruction boom, regardless of whether it is fuelled by insurance payouts, personal savings, or direct government intervention. There will be both disruptions and opportunities. Tracking this unfolding and complex picture is an important feature of economic resilience as policy makers and business owners review the metrics that can warn of troubled economic sectors needing help as well as highlight opportunities for growth.

Cross-References

- ▶ [Community Recovery Following Earthquake Disasters](#)
- ▶ [Earthquake Disaster Recovery: Leadership and Governance](#)
- ▶ [Economic Impact of Seismic Events: Modeling](#)
- ▶ [Reconstruction Following Earthquake Disasters](#)
- ▶ [Resilience to Earthquake Disasters](#)
- ▶ [Resourcing Issues Following Earthquake Disaster](#)

References

- Asian Development Bank (2008) Earthquake insurance: lessons from international experience and key issues for developing earthquake insurance in the PRC, observations and suggestions, 24 June 2008. http://reliefweb.int/sites/reliefweb.int/files/resources/F_R.pdf. Accessed 5 Dec 2013
- Brown C, Seville E, Vargo J (2013) The role of insurance in organisational recovery following the 2010 and 2011 Canterbury earthquakes. Resilient Organisations Research Report 2013/04. <http://www.resorgs.org.nz/Publications/research-reports.html>. Retrieved 12 May 2013
- Cavallo E, Noy I (2011) The economics of natural disasters – a survey. *Int Rev Environ Resour Econ* 5(1):63–102
- Cavallo E, Galiani S, Noy I, Pantano J (2013) Catastrophic natural disasters and economic growth. *Rev Econ Stat* 95(5):1549–1561
- Chang SE, Falit-Baiamonte A (2002) Disaster vulnerability of businesses in the 2001 Nisqually earthquake. *Glob Environ Change Part B Environ Hazard* 4(2–3):59–71
- Chang SE, Taylor JE, Elwood KJ, Seville E, Brunsdon D, Gartner M (2014) Urban disaster recovery in Christchurch: the central business district cordon and other critical decisions. *Earthquake Spectra*. doi:10.1193/022413EQS050M
- Chang-Richard AY, Wilkinson S, Seville E, Brunsdon D (2012) Resourcing issues in past disaster recoveries: some perspectives. Resilient Organisations Research Report 2012/07. <http://www.resorgs.org.nz/Publications/research-reports.html>. Retrieved 12 May 2013
- Chang-Richard AY, Seville E, Wilkinson S, Walker B (2013) Building natural disaster response capacity: sound workforce strategies for recovery and reconstruction in APEC economies. APEC Human Resources Development Working Group, Asia Pacific Economic Cooperation. http://publications.apec.org/publication-detail.php?pub_id=1534. Retrieved 15 August 2014
- De Mel S, McKenzie D, Woodruff C (2012) Enterprise recovery following natural disasters. *Econ J* 122(559): 64–91
- Doyle L, Noy L (in press) The short-run nationwide macroeconomic effects of the Canterbury earthquakes. New Zealand Economic Papers, Taylor and Francis
- Hatton T, Seville E, Vargo J (2012) Improving the resilience of SMEs: policy and practice in New Zealand. Resilient Organisations Research Report 2012/12. <http://www.resorgs.org.nz/Publications/research-reports.html>. Retrieved 12 May 2013
- Kachali H, Stevenson JR, Whitman Z, Seville E, Vargo J, Wilson T (2012) Organisational resilience and recovery for Canterbury organisations after the 4 September 2010 earthquake. *Australas J Disaster Trauma Stud* 2012(1):11–19
- Katz J (2013) The big truck that went by. Palgrave Macmillan, New York
- Olshansky RB, Hopkins LD, Johnson LA (2012) Disaster and recovery: processes compressed in time. *Natl Hazard Rev (ASCE)* 13(3):173–178
- Sawada Y, Shimizutani S (2008) How do people cope with natural disasters? Evidence from the great Hanshin-Awaji (Kobe) earthquake in 1995. *J Money Credit Bank* 40(2–3):463–488

- SPUR (2012) Safe enough to stay. Spur Report 01/2012. www.spur.org. Retrieved 12 May 2013
- Stevenson JR, Kachali H, Whitman Z, Seville E, Vargo J (2011) Preliminary observations of the impacts the 22 February Christchurch earthquake on organisations and the economy: a report from the field (22 February – 22 March 2011). *N Z Soc Earthq Eng Bull* 44(2):65–76
- Tierney KJ (1997) Business impacts of the Northridge earthquake. *J Crisis Conting Manag* 5(2):87–97
- Venn D (2012) Helping displaced workers back into jobs after a natural disaster: recent experiences in OECD countries, OECD social, employment and migration working papers, vol 142 OECD Publishing, Paris

EEE Catalogue: A Global Database of Earthquake Environmental Effects

Luca Guerrieri¹, Alessandro Maria Michetti² and Pablo G. Silva³

¹Department of Geological Survey, ISPRA, Istituto Superiore per la Protezione e la Ricerca Ambientale, Rome, Italy

²Dipartimento di Scienza e Alta Tecnologia, Università dell'Insubria, Como, Italy

³Departamento de Geología, Universidad de Salamanca, Escuela Politécnica Superior de Ávila, Avila, Spain

Synonyms

Earthquake ground effects; Environmental seismic intensity; Macroseismic scale; Paleoseismology; Seismic hazard

Introduction

The recent destructive earthquakes that occurred in Japan (2011; e.g., Lay et al. 2013) and New Zealand (2010–2011 earthquake sequence; e.g., Quigley et al. 2012) have clearly pointed out that traditional seismic hazard assessment based only on vibratory ground motion data needs to be integrated with information about the local vulnerability of the territory to earthquake occurrence. Nowadays, a huge amount of information about the characteristics of Earthquake Environmental

Effects (EEEs; i.e., any phenomena generated by a seismic event in the natural environment; Michetti et al. 2004, 2007; Guerrieri et al. 2007) is available for a very large number of earthquakes that occurred not only in the instrumental period and in historical time but also in the prehistorical period (paleoearthquakes, e.g., Mc Calpin 2009; Reicherter et al. 2009). However, available information located in different sources (scientific papers, historical documents, professional reports) can often be difficult to access.

The EEE Catalogue is a free accessible database designed at global level to collect information about the characteristics, size, and spatial distribution of Earthquake Environmental Effects in a standard way from modern, historical, and paleoearthquakes. This database aims to properly retrieve the available information about EEEs at a global level and archive it into a unique database, in order to facilitate its use for seismic hazard purposes. The database implementation has been endorsed by the INQUA TERPRO Project #0811 (“A global catalogue and mapping of earthquake environmental effects”; 2008–2011), through a working group coordinated by ISPRA – Geological Survey of Italy. The first official release of the EEE Catalogue was done under the framework of the XVIII INQUA Congress, held in Bern in July 2011.

The information included in the EEE Catalogue relates to Earthquake Environmental Effects triggered by individual seismic events, classified according to the EEE types considered in the Environmental Seismic Intensity scale ESI-07. For each event, information is collected at three levels of increasing detail: earthquake, locality, and site. Also available imagery documentation (photographs, videos, sketch maps, stratigraphic logs) can be stored in the database. Moreover, within the EEE Catalogue, the epicentral and local intensity values based on the ESI-07 scale are recorded. This allows an objective comparison of earthquake intensity for events that occurred in different areas and/or in different periods.

The quality of the database in terms of completeness, reliability, and location resolution is obviously time dependent and, therefore, is

generally less reliable with increasingly more ancient events. Nevertheless, even where the information is less accurate (historical earthquakes, archeoseismological evidence), the documented effects are typically the most relevant (i.e., most diagnostic for intensity assessment). Similarly, the information from paleoseismic investigations, although poorly representative of the entire scenario, still includes significant data (i.e., local coseismic fault displacements), which are very helpful in estimating a minimum size for the earthquake. At present about one hundred earthquakes are included in the EEE Catalogue.

The implementation of the EEE Catalogue is always in progress through a Web Implementation Interface developed for remote data collection at the following URL: <http://www.eeecatalog.sinanet.apat.it/login>

Earthquake records can be edited only by authorized compilers. Data compilation must be done at three different levels of generalization (earthquake, locality, and site). Moreover, it is currently possible to upload (i) the pattern of coseismic ruptures as linear elements (previously edited as .kml file), at earthquake level; (ii) photographs, sketches, or videos documented a single effect, at site level.

In the next section the structure of the database including mandatory fields at earthquake, locality, and site level is described.

Data can be explored on a public interface (Fig. 1) based on Google Earth; the web page address is available at <http://www.eeecatalog.sinanet.apat.it/terremoti/index.php>

From this URL, it is possible to view (Fig. 2) some general information about the earthquake visible on a Google Balloon, by clicking on the epicenter (yellow star, frame A) including the pattern of coseismic ruptures (primary or secondary), surveyed or interpreted, by clicking on the gray triangle corresponding to rupture zone and activating the box rupture zone (frame B). The geographic distribution of localities (green drops) and sites (classified in 7 EEE categories) can be viewed by clicking on the gray triangle corresponding to locality level and activating the box locality (frame C). By zooming in detail on

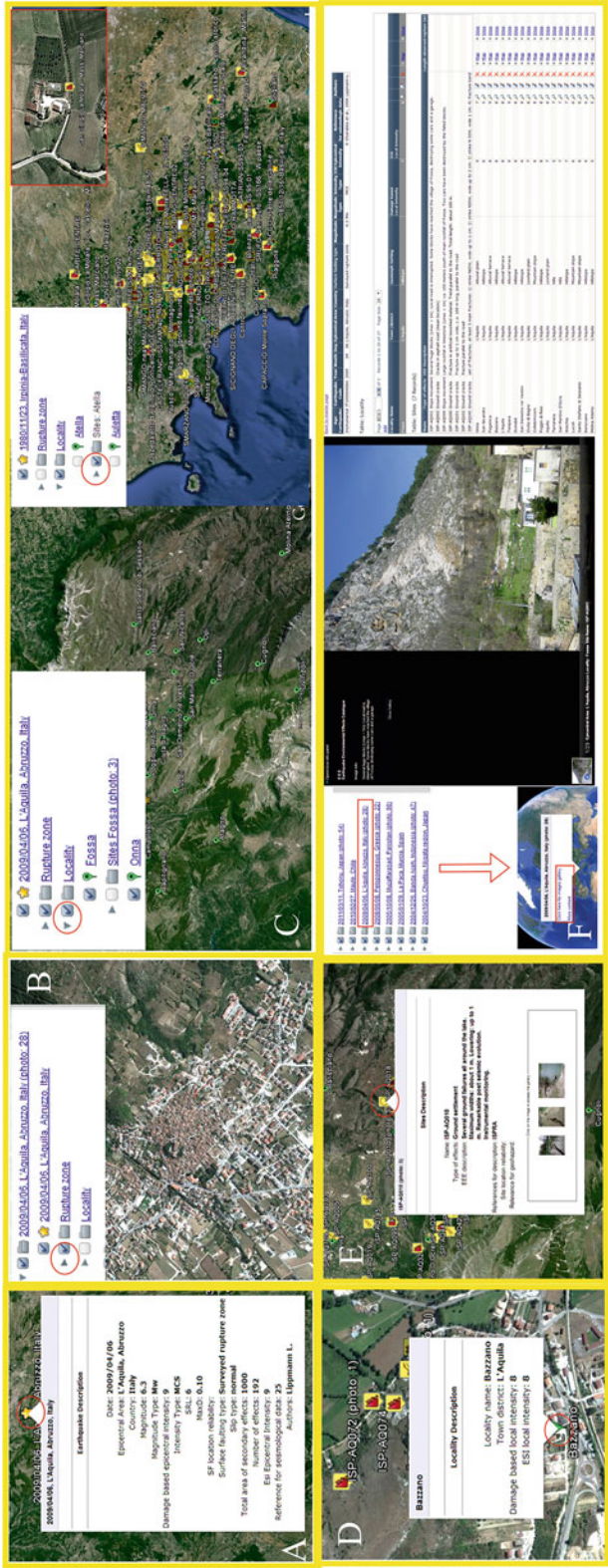
each locality and site, it is possible to better locate each individual element (red box in detail in C). Some general information about the locality (e.g., local intensity) and site (e.g., EEE classification, description, reference, and photograph) are visible on a Google Balloon, by clicking on the locality (frame D) and site (frame E). The entire database and photogallery (including maps, video, and other .pdf documents) related to a specific earthquake can be consulted by clicking on the earthquake string in the table of contents (frame F).

The EEE Catalogue is also available on the IAEA-ISSC (International Atomic Energy Agency- International Seismic Safety Centre) portal (<https://issc.iaea.org/home.php>) for use and dissemination within the worldwide nuclear engineering community. To this end, a new database infrastructure has been developed by the adaptation and customization of the EEE Catalogue to the specific IAEA-ISSC standard requirements (e.g., content, reliability, magnitude thresholds, geographic priorities, IPR (intellectual property rights), security policies).

For a quick field survey of EEE data consistent with the EEE Catalogue standards, an application has been developed named “Earthquake Geo Survey” that can be used by IOS and android mobile devices (e.g., smartphones and tablets). Via this application, the user is able to document on-fault and off-fault effects, collect GPS waypoints, and indirectly geo-tag the collected photos. The types of effects that can be reported and described in detail are surface faulting (on-fault effects) and off-fault effects including slope failures, liquefaction, tsunamis, ground cracks, hydrological anomalies, and other effects like trees shaking. For example, regarding the earthquake-induced liquefaction phenomena, the user can select a subtype of liquefaction from a list including the three most characteristic ones (ejection of sandy material, subsidence, and lateral spreading), describe in detail the failure within an extra field, report the maximum diameter of sand boils, select either water or sand ejection, and finally locate the site by activating the GPS and take a photo of the liquefaction manifestation. In addition, the



EEE Catalogue: A Global Database of Earthquake Environmental Effects, Fig. 1 The public interface of the EEE Catalogue, developed on Google Earth. <http://www.eeeecatalogue.sinanet.apat.it/terremoti/index.php>



EEE Catalogue: A Global Database of Earthquake Environmental Effects, Fig. 2 How to explore data into the EEE Catalogue (see text for details)

macroseismic intensity can also be evaluated using the ESI-07 scale and the relevant chart that is provided.

Structure of the EEE Catalogue

The EEE Catalogue is structured at three levels of progressive detail: earthquake, locality and site. Following this structure, the database has been developed using three main tables linked with a one-to-many relation: each record in the earthquake table is associated with one-to-many records in the locality table and each record in locality is associated to one-to-many records in its site table.

A site corresponds to the place where the single EEE of a certain type was observed, for example, a landslide or a ground crack. This is the level at which EEE descriptions (verbal or measured) have to be compiled. As these effects are strongly dependent not only on the amount of shaking but also on many other outer factors, it is possible to assign only an interval of probable intensity values to the effect observed at the site.

Locality includes one-to-many sites and presents a level of generalization, to which intensity can be assigned. It is assumed that at the spatial frame of a locality, random effects associated with the site are removed. Locality can refer to any place: either inhabited or natural. As a natural locality can be a river valley, a mountain slope, a large hill, etc., it is difficult to define exactly the size of a locality: in fact, it has to be small enough to keep separated areas with significantly different site intensities, but large enough to include several sites and consequently to be representative for intensity assessment.

Therefore, the locality has to be defined by expert judgment.

The main table earthquake is intended to present general information on the seismic event, including surface faulting parameters and total area of secondary EEEs. The locality table reports all the information about the characteristics of the locality where one or many coseismic

effects according to traditional macroseismic scales have occurred, i.e., location (coordinates, altitude) and local expression of the earthquake (local macroseismic intensity). In the site table all the characteristics of the site (location, geomorphological environment, etc.) as well as the type of effect (local surface faulting, slope movements, ground cracks, ground settlements, hydrological anomalies, etc.) should be reported.

The list of fields associated with the earthquake table is shown in Fig. 3 (above). Modern and historical earthquakes include the following mandatory fields:

- (a) General data: only year (with estimated age uncertainties $\pm$ and dating method), country, epicentral area, epicentral latitude, and epicentral longitude. Coordinates of the epicenter will be on the fault trace estimated from existing fault-trench analyses. In case the paleoearthquake will be only and/or mainly featured by secondary EEEs, the coordinates of “epicenter” will be coincident with the coordinates of the barycenter of the paleoseismic sites where this event was documented.
- (b) Surface faulting data only if $M > 6.5$: SRL, MaxD, surface faulting type, slip type, and rupture zone (if recorded).
- (c) EEE total distribution: total area of secondary effects (if recorded).
- (d) Seismic parameters: ESI epicentral intensity and estimated moment magnitude (M_w).

At locality level (Fig. 3, below) mandatory fields considered in the EEE Catalogue also are different for modern and historical earthquakes and paleoearthquakes. For modern and historical events, the following mandatory fields are considered: locality name, locality latitude, locality longitude, and ESI local intensity. For paleoearthquakes, the locality level will include off-fault localities recording secondary EEEs.

At site level all the possibilities considered are included in the classification of EEEs indicated in the ESI-07 scale, listed in Fig. 4. For modern and

EEE Catalogue: A Global Database of Earthquake Environmental Effects,

Fig. 3 Above: the earthquake table provides general information about the seismic event and a summary of primary (pattern of surface faulting) and secondary effects (total area). Below: the locality table allows to compare local intensities from traditional scales (based on damages to buildings) and from ESI scale (based on EEEs)

Id Earthquake Code	39
Age Earthquake	Instrumental
Earthquake Code	IT20090406m
Year	2009
Date uncertainty	
Month	04
Day	06
Hour	01
Min	32
Sec	
Epicentral Area	L'Aquila, Abruzzo
Country	Italy
Epicentral Latitude	42.33
Epicentral Longitude	13.33
SRL	6
Max D	0.10
Surface faulting type	☐ Hypothesized based on site observations ☒ Surveyed rupture zone
Slip Type	normal
Rupture zone kmf	2009 L'Aquila rupture zone(1).kmf * Keep * Remove * Replace Scegli file Nessun file selezionato
Total Area Of Secondary Effects (km2)	1000
Number Of Effects	192
Magnitude	6.3
Magnitude Type	Mw
Damage Based Epicentral Intensity	9
Intensity Type	MCS
ESI Epicentral Intensity	9
Reference for seismologic data	Cestani, 1986 Chatzipetros et alii, 2008 Cherubini et al., 1981 Charabba et al., 2009
Authors	1 September 1

Table: Locality

Page 1 of 2 Records 1 to 20 of 27 Page Size 20

Locality Name	Town /district	Geomorph Setting	Damage based Local intensity	ESI Local Intensity						
Fossa	L'Aquila	Hillslope	7	8					* Map	* Sites
Onna	L'Aquila	Alluvial plain	9	7					* Map	* Sites
San Nicandro	L'Aquila	Hillslope	6	5					* Map	* Sites
Paganica	L'Aquila	Alluvial terrace	8	9					* Map	* Sites
Bazzano	L'Aquila	Hillslope	8	8					* Map	* Sites
L'Aquila	L'Aquila	Alluvial terrace	8	8					* Map	* Sites
Tempera	L'Aquila	Alluvial terrace	9	9					* Map	* Sites
Ovindoli	L'Aquila	Hillslope	6	6					* Map	* Sites
San Demetrio ne' Vestini	L'Aquila	Hillslope	6	8					* Map	* Sites
Civita di Bagno	L'Aquila	Lowland plain	7	6					* Map	* Sites
Collebrincioni	L'Aquila	Mountain slope	6	8					* Map	* Sites
Poggio di Roio	L'Aquila	Hillslope	8	8					* Map	* Sites
Aquilio	L'Aquila	Lowland plain	7	6					* Map	* Sites
Terranera	L'Aquila	Hills	5	7					* Map	* Sites
San Martino D'Ocre	L'Aquila	Hills	7	7					* Map	* Sites
Arischia	L'Aquila	Hillslope	7	7					* Map	* Sites
Lucoli	L'Aquila	Mountain slope	6	6					* Map	* Sites
Santo Stefano di Sessanio	L'Aquila	Mountain slope	6	6					* Map	* Sites
Barisciano	L'Aquila	Hillslope	6	6					* Map	* Sites
Molina Aterno	L'Aquila	Hillslope	5	5					* Map	* Sites

>EEE Catalogue Web Implementation<>

View Table: Sites

[Back to list](#) [Add](#) [Edit](#) [Delete](#) [Images](#)

Sites	Surface faulting	Slope movements	Ground settlements	Ground cracks	Hydrological anomalies	Anomalous waves	Other effects
Name	ISP-AQ010						
Type of effects	Ground settlement						
EEE Description	Several ground failures all around the lake. Maximum widths: about 1 m. Lowering: up to 1 m. Remarkable post seismic evolution. Instrumental monitoring.						
References for description	ISPRA						
Site Latitude	42.2909						
Site Longitude	13.5765						
Site Altitude							
Site Geomorphologic Setting	Hillslope						
Picture							

EEE Catalogue: A Global Database of Earthquake Environmental Effects, Fig. 4 The site table provides a description of the single effect as available in the original source and a first classification according to a standard

terminology (surface faulting, slope movements, ground settlements, ground cracks, hydrological anomalies, anomalous waves, other effects)

historical earthquakes, the following general mandatory fields are considered: site name, type of effects, site latitude, site longitude, EEE description, and reference for description. Then, according to the type of effect, the following fields are highlighted:

- (a) Hydrological anomalies: surface water effects and/or groundwater effects
- (b) Anomalous waves and tsunamis: max wave height, anomalous wave width, and length of affected coast
- (c) Ground cracks: strike, dip, max opening, and length
- (d) Liquefactions: max diameter and max lowering
- (e) Slope movements: type, total volume, and max size of blocks (only for rockfalls)

Where local surface faulting has been observed, strike, dip, type of movement, and local offset (vertical and/or horizontal) are also recorded. For paleoearthquakes, at site level only general fields and fields related to local surface faulting (expected to be found by paleoseismic investigation) must be filled.

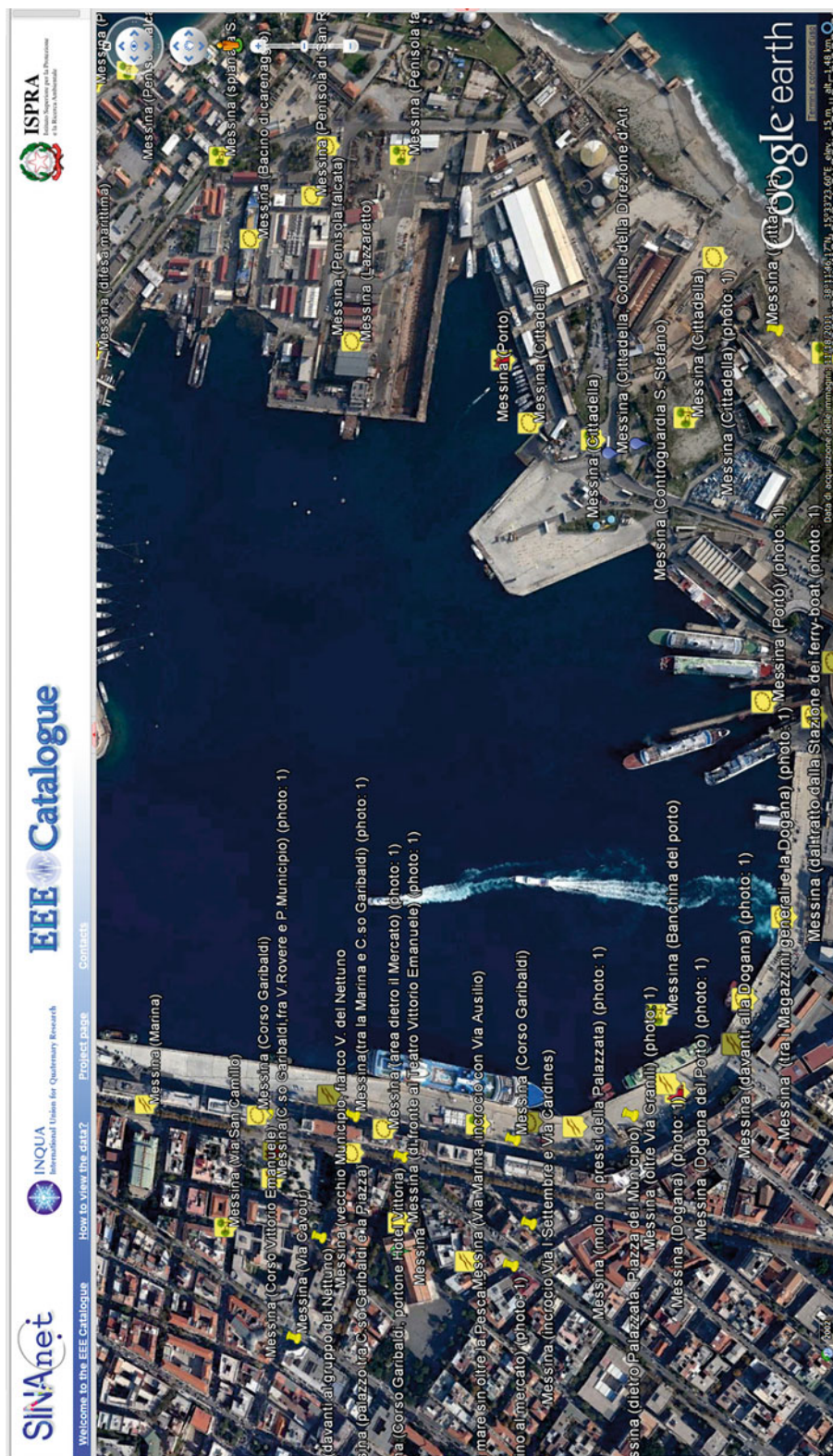
Case Studies

The major added value of the EEE Catalogue for seismic risk assessment is the possibility to explore the scenarios of environmental effects

induced by past earthquakes and therefore to identify the areas where the human settlements and infrastructures are more exposed to this hazard. To this end, a high accuracy for the location of EEEs becomes crucial. Typically, EEEs from recent earthquakes are mapped with good accuracy immediately after the event. Nevertheless, even for some historical earthquakes, it is possible to retrieve with very good detail this information (cfr. Serva et al. 2007).

It is the case of the December 28, 1908 (M 7.1), Messina-Reggio Calabria earthquake and consequent tsunamis (Fig. 5), where the EEE Catalogue allows to locate the earthquake/tsunami effects over the present urban texture with a spatial resolution of a few meters, pointing out the areas characterized by the highest risk.

Comerci et al. (2008) have provided a careful review of all available contemporary documents, i.e., technical and photographic reports, newspapers, and other archive material related to the seismic and tsunami event, allowing to recognize and classify a large number (365) of independent environmental effects. These have been categorized as slope movement, ground settlement, ground vertical movement and coastal retreat, liquefaction, ground crack, hydrological anomaly, gas emission, light, and rumble. A large part of the achieved information has allowed to evaluate the epicentral intensity, besides local intensities at 67 sites, by applying the Environmental Seismic Intensity scale ESI 2007.



EEEE Catalogue: A Global Database of Earthquake Environmental Effects, Fig. 5 EEEs induced by the December 28, 1908, Messina Straits, Italy, earthquake in the area of the Messina harbor. If information from contemporary sources is very precise, it is possible to use the EEE Catalogue also for local seismic microzonation

This study has led to the most comprehensive picture of the distribution and characteristics of coseismic environmental effects of that earthquake, a crucial knowledge to better estimate the future impact of a similar earthquake on a region that has seen a broad, and often poorly concerned, urban, and infrastructural development since then.

Furthermore, the EEE Catalogue can reveal possible trends in the spatial distribution of primary and secondary effects. The spatial distribution of EEEs induced by the October 8, 2005, Muzaffarabad, Pakistan, earthquake shows that the location and amount of surface faulting are consistent with the spatial distribution of coseismic slope movements, in terms of both areal density and size (Fig. 6).

Ali et al. (2009) provide a detailed field survey of EEEs in the epicentral area of the Muzaffarabad event. The displacement and surface expression of the causative fault and accompanying secondary environmental effects were observed at a number of places along a capable thrust fault structure (Kashmir Thrust). A complex, clearly segmented, at least 112-km-long surface rupture was mapped with maximum values of vertical displacement on the order of 4–7 m mainly along the central segment of the rupture (52 km between Muzaffarabad and Balakot) associated with maximum slip at depth and a major portion of the energy release. The adjacent segments are instead characterized by scattered minor surface ruptures with a few centimeters of displacement, accompanied by extensive surface cracking, landslides, and severe damage, concentrated in a narrow belt along the fault trace.

A maximum intensity of XI on the Modified Mercalli Intensity (MMI) scale and on the Environmental Seismic Intensity scale (ESI 2007) was recorded in the epicentral area between Muzaffarabad and Balakot. Extremely severe damage and very important secondary environmental effects in the hanging wall adjacent to the trace of the causative fault plane are mainly due to near-fault strong motion and rupture directivity effects.

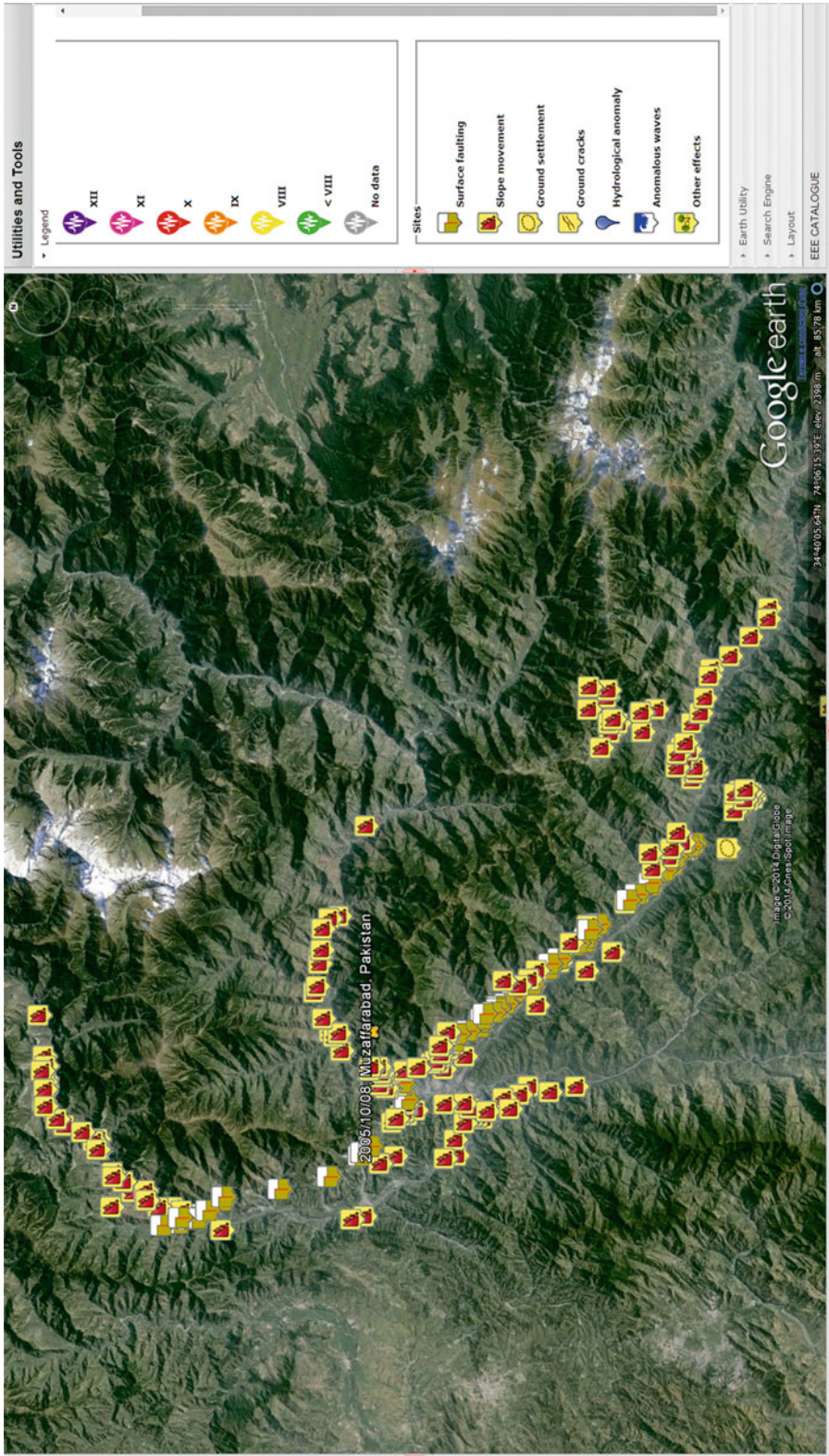
A similar result is shown by the spatial distribution of EEEs induced by the 1811–1812 New Madrid, Missouri, earthquakes, mapped in Fig. 7. Indeed, the most relevant primary and secondary effects are located along the Mississippi Valley near New Madrid, consistently with the surface projection of the causative faults, and unquestionably provide diagnostic elements for assessing an epicentral intensity equal to XI (Guerrieri et al. 2011).

In the winter of 1811 and 1812, the New Madrid Seismic Zone generated an earthquake sequence that included three very large to great earthquakes. The three largest earthquakes in this sequence are estimated to be of moment magnitude M 7.8–8.1 and were felt as far away as Hartford, Connecticut, Charleston, South Carolina, and New Orleans, Louisiana. In the northern part of the Mississippi embayment, these earthquakes caused widespread liquefaction (Fuller 1912; Saucier 1977; Obermeier 1989) and severe ground failure (Tuttle and Barstow 1996). Liquefaction was reported about 200 km northeast of the New Madrid Seismic Zone in White County, Illinois; 240 km to the north-northwest near St. Louis, Missouri; and 250 km to the south near the mouth of the Arkansas River. Although the exact magnitudes of the earthquakes are debatable, there is no doubt that a repeat of an 1811–1812 New Madrid sequence today would cause a great deal of damage in the central United States (Tuttle 1999).

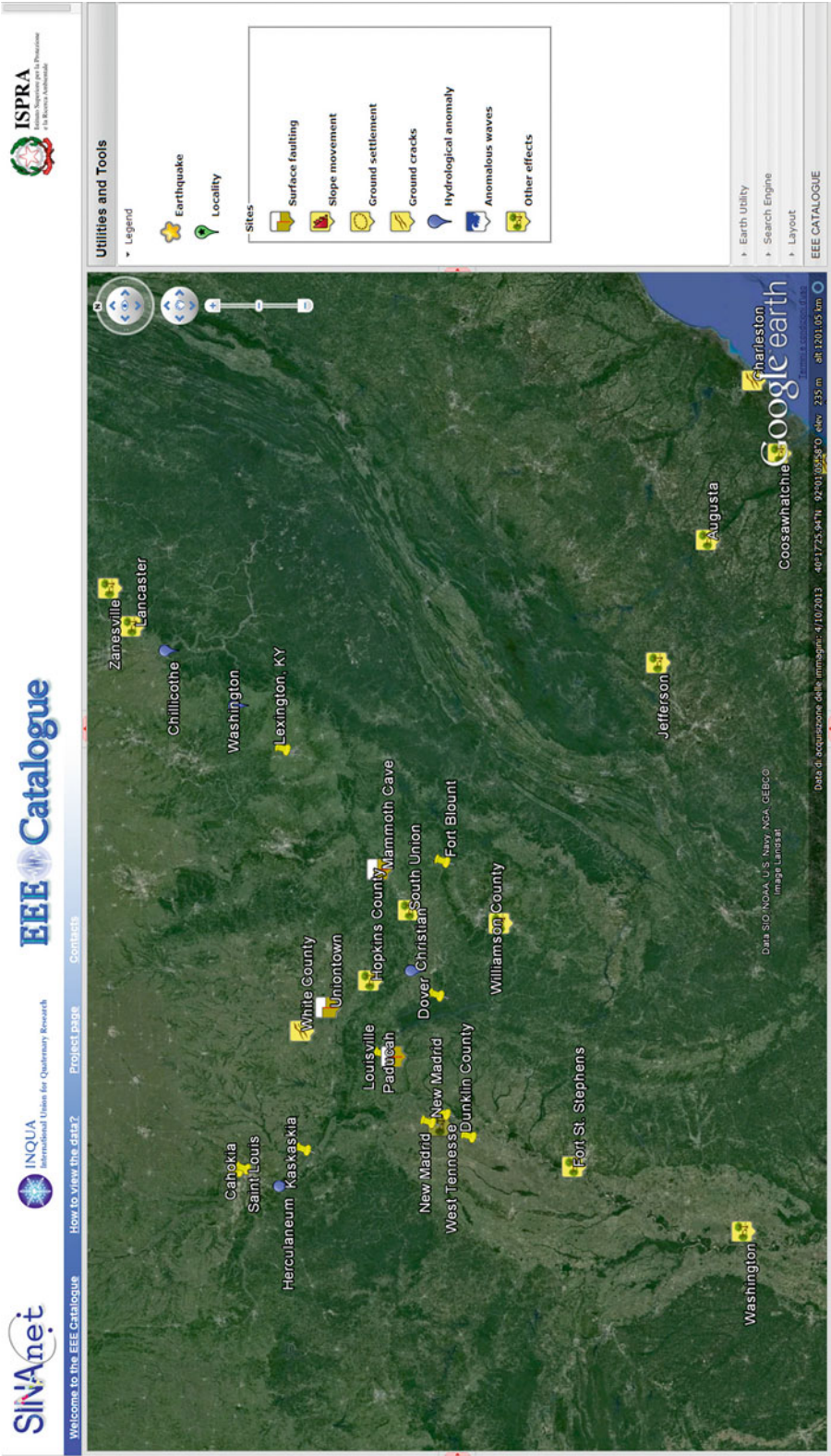
Future Developments

At June 2014, 101 earthquake records have been collected into the EEE Catalogue: among them, 44 are modern (after 1900), 39 are historical, and 18 are paleoearthquakes. Other 31 earthquake records are currently under implementation. Most of the compiled earthquake records have occurred in Europe (62), then in South America (19), in Asia (16), and in North America (4).

Nevertheless, the compilation of the EEE Catalogue is still ongoing with the aim to collect any available information resulting from



EEE Catalogue: A Global Database of Earthquake Environmental Effects, Fig. 6 Surface faulting and slope movements induced by the October 8, 2005, Muzaffarabad, Pakistan, earthquake



EEE Catalogue: A Global Database of Earthquake Environmental Effects, Fig. 7 EEE induced by the December 16, 1811, New Madrid, Missouri, USA, earthquake. Primary and secondary effects indicative of intensity XI in the ESI 2007 scale are located in the epicentral area along the Mississippi Valley

specific field surveys after recent earthquakes, as well as from historical and paleoseismic studies. For the next future, the goal will be to improve the completeness of the EEE Catalogue with the aim to properly use collected data for an improved assessment of seismic hazard based on a credible scenario of EEEs.

Summary

The EEE Catalogue is a free accessible database (<http://www.eeecatalog.sinanet.apat.it/terremoti/index.php>) designed at global level to collect information about the characteristics, size, and spatial distribution of Earthquake Environmental Effects in a standard way from modern, historical, and paleoearthquakes.

This database has been promoted with the aim to properly retrieve the available information about EEEs at a global level and archive it into a unique database, in order to facilitate their use for seismic hazard purposes. The information included in the EEE Catalogue pertains to Earthquake Environmental Effects triggered by individual seismic events, classified according to the EEE types considered in the Environmental Seismic Intensity scale ESI-07. For each event the information is collected at three levels of increasing detail (earthquake, locality, and site). Also, available imagery documentation (photographs, videos, sketch maps, stratigraphic logs) can be stored in the database. Moreover, within the EEE Catalogue the epicentral and local intensity values based on the ESI-07 scale are recorded. This allows an objective comparison of earthquake intensity for events which occurred in different areas and/or in different periods.

The implementation of the EEE catalogue is always in progress through a Web Implementation Interface developed for remote data collection at the following permanent URL: <http://www.eeecatalog.sinanet.apat.it/login>

The major added value of the EEE Catalogue for seismic risk assessment is the possibility to explore the scenarios of environmental effects induced by past earthquakes and therefore to

identify the areas where the human settlements and infrastructures are more exposed to this source of potential hazard.

Cross-References

- Archeoseismology
- Earthquake Location
- European Structural Design Codes: Seismic Actions
- Intensity Scale ESI 2007 for Assessing Earthquake Intensities
- Paleoseismic Trenching
- Paleoseismology
- Paleoseismology and Landslides
- Paleoseismology: Integration with Seismic Hazard
- Probabilistic Seismic Hazard Models
- Seismic Actions Due to Near-Fault Ground Motion
- Seismic Risk Assessment, Cascading Effects
- Time History Seismic Analysis
- Tsunamis as Paleoseismic Indicators

References

- Ali Z, Qaisar M, Mahmood T, Shah MA, Iqbal T, Serva L, Michetti AM, Burton PW (2009) The Muzaffarabad, Pakistan, earthquake of 8 October 2005: surface faulting, environmental effects and macroseismic intensity. *Geol Soc Lond Spec Publ* 316:155–172
- Comerci V, Blumetti AM, Brustia E, Di Manna P, Fiorenza D, Guerrieri L, Lucarini M, Serva L, Vittori E (2008) The effects on the environment of the 1908 Southern Calabria – Messina earthquake (Southern Italy). In: 1908–2008 Scienza e Società a cento anni dal Grande Terremoto, Reggio Calabria, 10–12 dicembre 2008. *Miscellanea INGV*, n. 03
- Fuller ML (1912) The New Madrid earthquake. *US Geol Surv Bull* 494:119
- Guerrieri L, Tatevossian R, Vittori E, Comerci V, Esposito E, Michetti AM, Porfido S, Serva L (2007) Earthquake environmental effects (EEE) and intensity assessment: the INQUA scale project. *Boll Soc Geol It (Ital J Geosci)* 126(2):375–386
- Guerrieri L, Blumetti AM, Brustia E, Esposito E, Lucarini M, Michetti AM, Porfido S, Serva L, Vittori E, The INQUA TERPRO Project #0811 Working Group (2011) Earthquake environmental effects, intensity and seismic hazard assessment: the EEE catalogue (INQUA Project #0418). In: *Proceedings of the*

2nd INQUA – IGCP 567 workshop on active tectonics, earthquake geology, archaeology and engineering. Corinto (Greece) 18–25 Sept 2011, pp 62–65. ISBN: 978-960-466-093-3

Lay T, Fujii Y, Geist E, Koketsu K, Rubinstein J, Sagiya T, Simons M (2013) Introduction to the special issue on the 2011 Tohoku earthquake and Tsunami. *Bull Seismol Soc Am* 103:1165–1170

Mc Calpin JP (2009) *Paleoseismology*, 2nd edn. Elsevier, Amsterdam, 613 pp

Michetti AM, Esposito E, Gürpinar A, Mohammadioun B, Mohammadioun J, Porfido S, Roghazin E, Serva L, Tatevossian R, Vittori E, Audemard F, Commerci V, Marco S, McCalpin J, Mörrer NA (2004) The INQUA scale. An innovative approach for assessing earthquake intensities based on seismically-induced ground effects in natural environment. In: Vittori E, Comerci V (eds) *Memorie Descrittive della Carta Geologica d'Italia*, vol LXVI-116. APAT, Roma

Michetti AM, Esposito E, Guerrieri L, Porfido S, Serva L, Tatevossian R, Vittori E, Audemard F, Azuma T, Clague J, Commerci V, Gürpinar A, Mc Calpin J, Mohammadioun B, Mörrer NA, Ota Y, Roghazin E (2007) Intensity scale ESI 2007. In: Guerrieri L, Vittori E (eds) *Memorie Descrittive della Carta Geologica d'Italia*, vol 74. Servizio Geologico d'Italia, Dipartimento Difesa del Suolo, APAT, Rome

Obermeier SF (1989) The New Madrid earthquakes: an engineering-geologic interpretation of relict liquefaction features. *US Geol Surv, Professional Paper* 1336-R, 114 pp

Quigley M, Van Dissen R, Litchfield N, Villamor P, Duffy B, Barrell D, Noble D (2012) Surface rupture during the 2010 Mw 7.1 Darfield (Canterbury) earthquake: implications for fault rupture dynamics and seismic-hazard analysis. *Geology* 40(1):55–58

Reicherter K, Michetti AM, Silva PG (eds) (2009) *Palaeoseismology: historical and prehistorical records of earthquake ground effects for seismic hazard assessment*, Special Publications, 316. Geological Society of London, London, 60 pp

Saucier RT (1977) Effects of the New Madrid earthquake series in the Mississippi Alluvial Valley. *US Army Corps of Engineers Waterways Experiment Station, miscellaneous paper* S-77-5

Serva L, Esposito E, Guerrieri L, Porfido S, Vittori E, Commerci V (2007) Environmental effects from some historical earthquakes in Southern Apennines (Italy) and macroseismic intensity assessment. Contribution to INQUA EEE scale project. *Quat Int* 173–174:30–44

Tuttle M (1999) The use of liquefaction features in paleoseismology: lessons learned in the New Madrid seismic zone, central United States. *J Seismol* 5(3):361–380

Tuttle M, Barstow N (1996) Liquefaction-related ground failure: a case study in the New Madrid seismic zone, central United States. *Bull Seismol Soc Am* 86(3): 636–645

Electrochemical Seismometers of Linear and Angular Motion

Vadim M. Agafonov¹, Alexander V. Neeshpapa² and Anna S. Shabalina¹

¹Center for Molecular Electronics, Moscow Institute of Physics and Technology, Moscow, Russia

²R-sensors LLC, Dolgoprudny, Russia

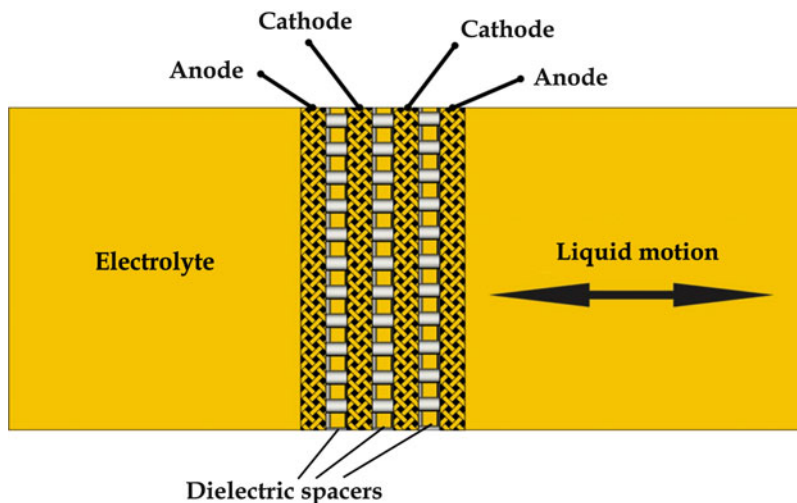
Synonyms

MET seismic sensors; Molecular electronic angular and linear seismometers; Molecular electronic transfer seismic sensors

Introduction

Traditionally seismic sensors use a mechanical system with moving solid-state inertial mass in combination with either velocity or displacement transducers. The latter convert mechanical motion of the inertial mass into the electrical signal. An alternative approach based on employment of liquid inertial mass and four-electrode electrochemical transducer with potential difference applied between the electrodes as a converter of the liquid motion into electrical signal was developed in recent years (Huang et al. 2013; Agafonov et al. 2014). Such devices are known as MET (molecular electronic transfer) or electrochemical sensors. Basically, the operation of the MET sensors uses the sensitivity of an interelectrode current in the electrochemical cell to the movement of a liquid electrolyte relative to the electrodes fixed on the sensor housing. Along with the solid-state electronics (charge transfer by electrons and holes in a solid conductor or semiconductor) and vacuum electronics (charge transfer by free electrons in an ionized gas or vacuum), the MET sensors could be considered as the third class of fundamental electronic devices, characterized by charge transfer via ions in solution – hence the older name of the

Electrochemical Seismometers of Linear and Angular Motion, Fig. 1 Schematic of the basic MET sensing element



E

sensing technology “Solion.” Solion technology was first implemented into a practical device in the 1950s by a US Navy-sponsored project. The early applications of Solion devices were to detect low-frequency acoustic waves, either in the form of an infrasonic microphone or limited-band seismometer (Hurd and Lane 1957; Wittenborn 1958; Collins et al. 1964). The first theoretical model describing the physical principles of operation of such devices was developed by C. W. Larcam in 1965 (Larcam 1965) who studied a simple one-dimensional model of four electrodes (two anodes and two cathodes) placed along a channel filled with an electrolyte moving under the action of external inertial forces.

Significant work on Solion motion detectors was continued in Russia, where the term “molecular electronic transfer” was firstly introduced to describe such a device (Lidorenko et al. 1984). Inspired by the exceptionally high rate of mechanical signal conversion to electric current in MET involving mass and charge transport, pioneer studies of MET (Kozlov et al. 1991; Kozlov and Safonov 2003; Panferov and Kharlamov 2001; Reznikova et al. 2001; Agafonov and Krishtop 2004; Volgin et al. 2003) provided an alternative paradigm in the development of seismic sensors. The advantages of MET motion sensors include small size, lack of fragile moving parts, high shock

tolerance, wide operating frequency range, high sensitivity, and low noise, as well as low dependence or complete independence of the response on installation angle. These features are widely used in practice especially in field and ocean-bottom applications, giving important benefits in using a relatively non-expensive, easy to install, compact and at the same time sensitive and low-noise device.

The main part of the current paper includes the following divisions: operating principles, description of the method, how the transfer function is built, and examples of technical parameters for a commercial MET, including velocimeters, seismic accelerometers, and unique highly sensitive compact rotational seismometers.

Molecular Electronic Transducer: Principle of Operation

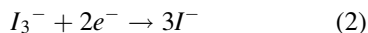
Figure 1 illustrates the basic concept of a MET sensing element, consisting of two transforming electrochemical cells (Lidorenko et al. 1984). Four electrodes configured as anode–cathode–cathode–anode (ACCA), separated by dielectric spacers, which protect adjacent electrodes from contact, span the channels filled with an electrolyte containing iodide ions. Holes through the electrodes and dielectric spacers allow fluid to flow through the channels.

The sensing mechanism of the MET is based on reversible chemical reactions transferring charge between anode and cathode via the electrolyte ions in the solution (Sun and Agafonov 2010). A solution with a high concentration of a neutral electrolyte (taking no participation in the electrode reactions) and a small fraction of the active component that is responsible for the charge transport over the fluid–electrode metal interface is usually used.

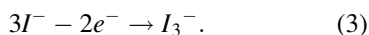
Most of the practical devices use iodine–iodide electrochemical systems with platinum electrodes. The electrolyte consists of a high-concentration aqueous solution of potassium iodide KI (lower temperature range boundary at $-15\text{ }^{\circ}\text{C}$) or lithium iodide LiI (lower temperature range boundary at $-55\text{ }^{\circ}\text{C}$) as neutral electrolyte and a small quantity of molecular iodine I_2 . If there is an excess supply of iodide, iodine enters a freely soluble complex compound (triiodide) according to the following scheme:



When the current flows through the MET, the following electrochemical reactions proceed at the electrodes: the reduction of iodine at the cathode



and the oxidation of iodine at the anode



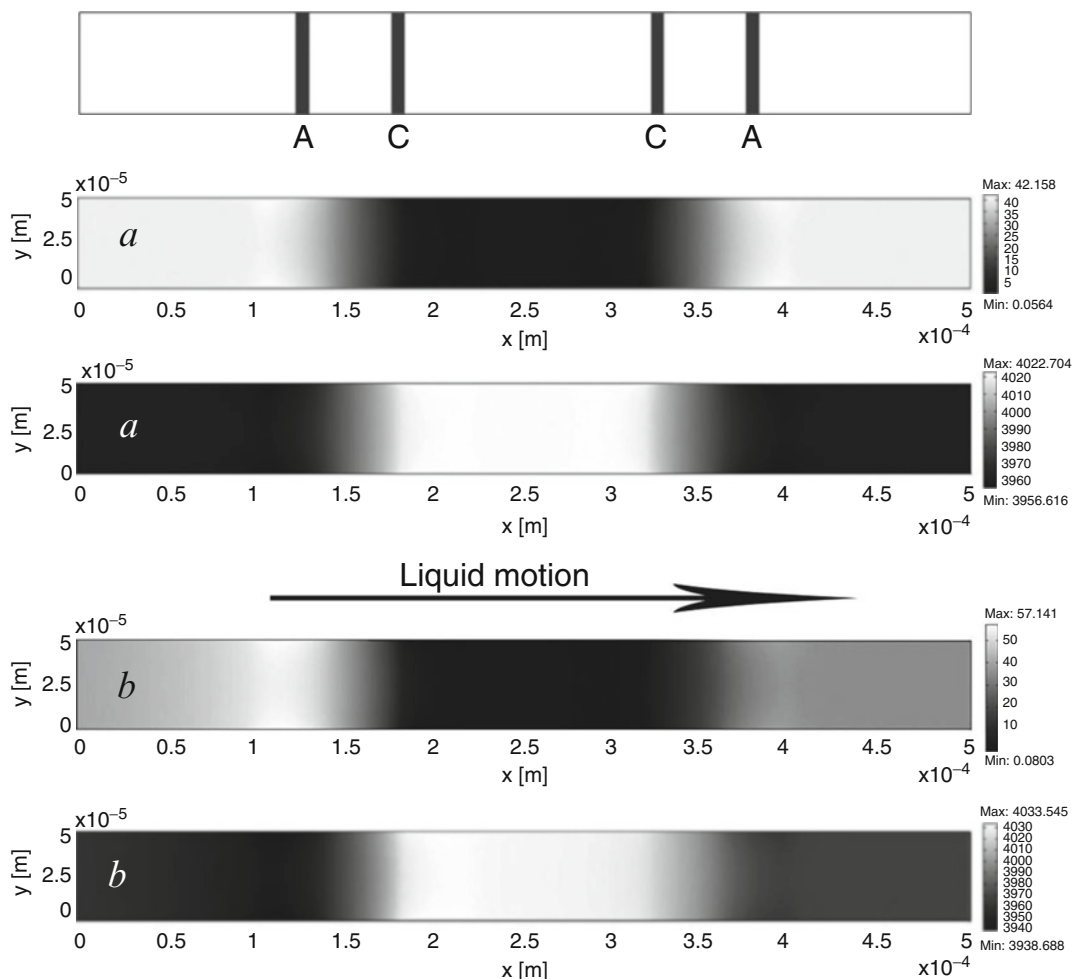
K and Li ions do not participate in the reactions.

The sensing element operation is based on the fact that rates of the electrochemical reaction at the electrodes are significantly higher than the reactant supply rate. The electrode reactions in this case lead to the emergence of a reactant concentration gradient. Figure 2a shows the distribution of I_3^- and I^- in the stationary electrolyte. Because of the electrode reactions, triiodide concentration maximizes at the anodes and minimizes at the cathodes, while iodide concentration maximizes at the cathodes and minimizes at the anodes. Since the ion concentrations shown in

Fig. 2a are symmetric about the central plane of the cell at $x = 250\text{ }\mu\text{m}$, left and right parts of the sensing element are equivalent, and there is no electric current difference between the two pairs of electrodes.

The ion distributions in the moving fluid are shown in Fig. 2b. From Fig. 2b, the maximum of the triiodide concentration is found at the upstream anode, while the minimum is at the downstream cathode. For iodide, as shown in Fig. 2b, the maximum concentration occurs at the upstream cathode, while the minimum concentration is found at the upstream anode. The maximums and minimums found in Fig. 2b are due to the fact that the triiodide generated at the upstream anode is consumed at the downstream cathode. By comparing Fig. 2a, b, it is found that both the maximum and the minimum concentrations of triiodide are greater when the fluid flows; for iodide the maximum concentration is greater when the fluid flows, while the minimum is greater in the stationary electrolyte. In turn, changes in the ion distribution result in the changes of the electrode current. Specifically, increase of the triiodide concentration in the vicinity of a cathode means acceleration of the electrochemical reaction according to Eq. 2. Inversely, decrease of the triiodide concentration results in the deceleration of the same reaction. So there is a difference between electrical currents for the two electrode pairs in the moving electrolyte. Usually the difference between cathode currents is used as an output signal of the transducer.

As the interelectrode potential difference is increased, the rate of electrochemical reactions increases gradually until such a condition occurs that all the triiodide ions reaching the electrode enter the electrochemical reaction instantly. This condition corresponds to the saturation current regime, and further increase in the potential difference does not alter the current. Under this condition the conversion coefficient of the sensing cell achieves its maximum value. That is why the saturation regime is typically used as a working point for the transducer. The current is limited by the volume rate of supply of the active component to the electrodes. In turn, the supply



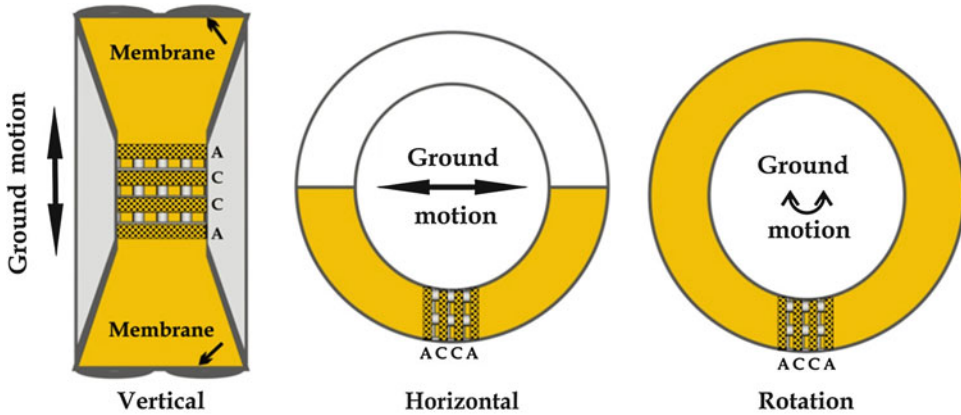
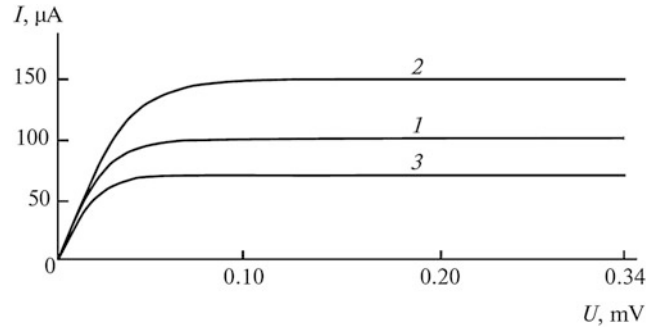
Electrochemical Seismometers of Linear and Angular Motion, Fig. 2 Schematic diagram of a four-electrode conduit (the electrodes are denoted by the shaded portions) and ion distributions in (a) stationary

and (b) moving electrolyte. Both in (a) and (b), I_3^- concentration is shown on the upper graph, I^- on the lower one

rate is determined by the ion diffusion in the stationary electrolyte. The flow of the solution introduces an additional convective ion transport that changes the saturation current at each pair of electrodes (see the volt–ampere characteristics (VACs) of the electrode system in the stationary and flowing electrolyte shown in Fig. 3). The convective flow is directed from cathode to anode for the first pair of electrodes and in opposite direction for the second one. That's why changes of the concentration are not symmetrical, and cathode current difference (output signal) is generated.

In the general case, the process is characterized by a high conversion factor. This is reflected in a strong electric response that exceeds the noise of accompanying electronic circuitry significantly, even when the input mechanical influence level is low. A high signal-to-noise ratio is ultimately achieved throughout the entire measurement channel. Numerous experimental and theoretical studies revealed a considerable dependence of the output parameters of the device on the conversion system geometry and the broad possibilities for optimizing the conversion system response depending on the problems being solved.

Electrochemical Seismometers of Linear and Angular Motion, Fig. 3 VACs of the transducer electrode system in a stationary electrolyte (1) and in an electrolyte flowing from the anode to the cathode (2) and cathode to anode (3)



Electrochemical Seismometers of Linear and Angular Motion, Fig. 4 Schematic of different type of MET motion sensor

Motion sensors based on MET technology include linear and angular accelerometers, rate sensors, gyroscopes, and seismometers. MET sensors can be configured for horizontal, vertical, and rotational sensing, as shown in Fig. 4. Despite of the difference in construction, all of them can be described in common rules-based approach. A vertical-axis sensor is considered in the following as an example.

Transfer Function of the MET Seismometers

A MET motion sensor can be considered as a superposition of two systems: first, input motion is converted into fluid motion of the electrolyte by a mechanical system. Next, the electrolyte's velocity is measured by the electrochemical system, resulting in an output current of the sensor.

Therefore, the frequency-dependent transfer function of the entire device can be written as:

$$H(\omega) = H_{\text{mech}}(\omega)H_{\text{ec}}(\omega), \quad (4)$$

where $H_{\text{mech}}(\omega)$ describes the mechanical response of the fluidic system and $H_{\text{ec}}(\omega)$ is the electrochemical transfer function describing the conversion of the fluid flow into an output current.

Considering the design of the vertical sensor (Fig. 4a) in addition to the sensing element, the highly flexible rubber membranes are assembled at the two ends of the channel to seal the electrolyte, providing a restoring force for the mechanical system. In the presence of an external ground acceleration, an inertial driving force is applied on the liquid electrolyte. The electrolyte moves along the channel once the driving force exceeds the hydrodynamic drag force. Let $x(t)$ be the

displacement of the liquid inertial mass m relative to the ground. There are two real forces acting on the liquid mass:

Restoring force from the flexible membrane, $-kV(t)$, is negative since it opposes the flow motion, where $V(t)$ is the volume of fluid passing through the channel and k is the flexibility coefficient of the membrane and depends only on the characteristics of the membrane.

Hydrodynamic damping force, $-\frac{R_h S_{ch} dV(t)}{dt}$, is linearly proportional to the volumetric flow rate of the liquid associated with the hydrodynamic resistance R_h and is negative since it also opposes the flow motion; S_{ch} is the cross-section area of the channel. In this consideration we suppose the cross-section area is constant.

The acceleration of the mass relative to an inertial reference frame will be the sum of the acceleration $\frac{d^2 x(t)}{dt^2}$ with respect to the ground and the ground acceleration $a(t)$. The center of mass acceleration is assumed equal to average liquid acceleration in the channel. Since the sum of forces must be equal to the mass times the acceleration, the equation that governs the motion of the liquid can therefore be expressed as

$$-R_h S_{ch} \frac{dV(t)}{dt} - kV(t) = m \frac{d^2 x(t)}{dt^2} + ma(t). \quad (5)$$

Given the relations:

$$V(t) = S_{ch} x(t), \quad m = \rho L S_{ch}, \quad (6)$$

where m is the mass of the electrolyte, ρ is the density of the electrolyte, and L represents the channel length, Eq. 5 is transformed to

$$\frac{d^2 V(t)}{dt^2} + \frac{R_h S_{ch}}{\rho L} \frac{dV(t)}{dt} + \frac{k}{\rho L} V(t) = -S_{ch} a(t). \quad (7)$$

By changing the variable of $Q(t) = \frac{dV(t)}{dt}$, where $Q(t)$ is the volumetric flow rate, and transforming Eq. 7 to frequency domain, the magnitude of the transfer function of mechanical system in MET transducer in the frequency domain can be obtained as follows:

$$|H_{mech}(\omega)| = \left| \frac{Q(\omega)}{a(\omega)} \right| = \frac{\rho L}{\sqrt{\left(\frac{\rho L}{S_{ch}} \right)^2 \frac{(\omega^2 - \omega_0^2)}{\omega^2} + R_h^2}}, \quad (8)$$

where $\omega_0 = \sqrt{\frac{k}{\rho L}}$ is the mechanical resonance frequency of the device.

An approximation of the electrochemical transfer function in frequency domain is

$$|H_{ec}(\omega)| = \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_D} \right)^2}}, \quad (9)$$

where $\omega_D = \frac{D}{d^2}$ is the diffusion frequency and d is the interelectrode distance (Kozlov et al. 1991).

The overall frequency-dependent transfer function of a MET seismometer is a superposition of the transfer functions of the mechanical and electrochemical system, which can be written as:

$$|H(\omega)| = \frac{\rho L}{\sqrt{\left(\frac{\rho L}{S_{ch}} \right)^2 \frac{(\omega^2 - \omega_0^2)}{\omega^2} + R_h^2}} \cdot \frac{1}{\sqrt{1 + \left(\frac{\omega}{\omega_D} \right)^2}}. \quad (10)$$

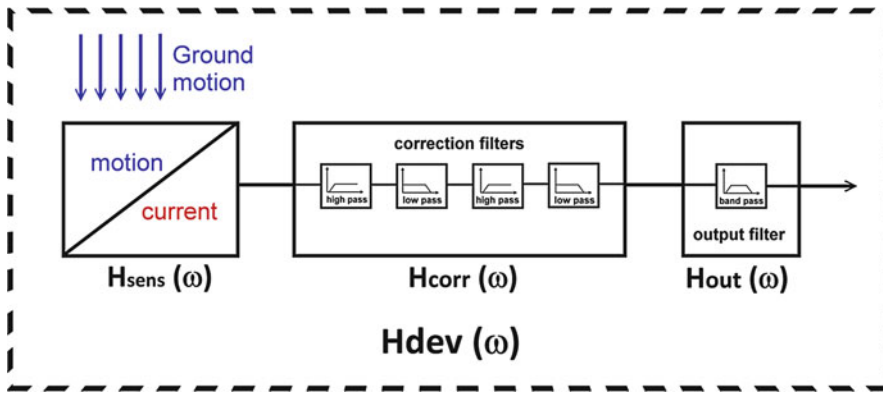
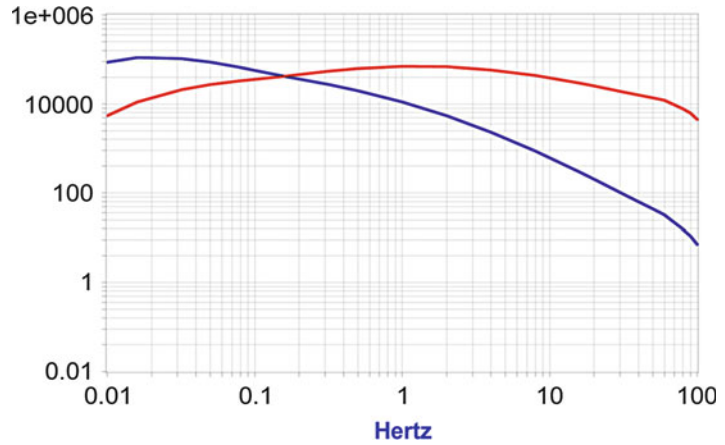
Figure 5 illustrates the typical response of a MET sensor to the applied acceleration (blue) and velocity (red). Both plots have marked maximums in sensitivity and do not have any flat region. To produce the flat response in the operating frequency range, a procedure of frequency equalization is required.

Two main approaches to equalize the frequency response of a MET sensor have been developed so far:

1. Equalization with a combination of low-pass and high-pass filters (filter equalization)
2. The use of force feedback (feedback equalization)

The filter equalization was used in earlier MET seismometers. The actual parameters of a specific sensing cell are difficult to predict and

Electrochemical Seismometers of Linear and Angular Motion, Fig. 5 Responses of a typical MET transducer to applied acceleration (*blue*) and velocity (*red*). Units on vertical axis are $\mu\text{A}/\text{m}/\text{s}^2$ and $\mu\text{A}/\text{m}/\text{s}$ for the *blue* and *red* curves, correspondingly



Electrochemical Seismometers of Linear and Angular Motion, Fig. 6 Functional diagram of a seismometer with the filter equalization

display a noticeable spread, so the task of achieving the desired accuracy can only be performed by means of the measuring the cell's transfer function directly. Shake tables of various types are necessary equipment to implement this procedure.

The approach is based on the directly measured response of a particular sensing cell in the frequency range of interest and adjustment of the corrective filters to get the desired response in the specified operating frequency band.

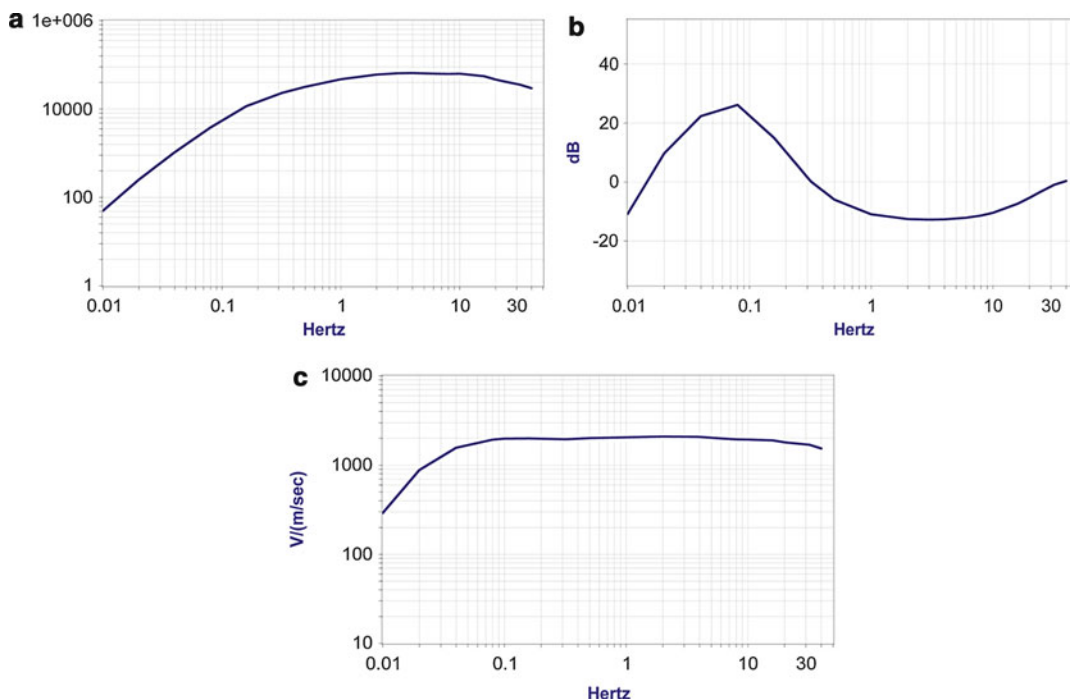
Figure 6 presents the functional diagram for the seismometer where the transfer function was built using the filter equalization procedure. Mathematically, the transfer function of a seismometer $H_{\text{dev}}(\omega)$ is presented as a product of the transfer function of the transducer by the transfer functions of the successive electronic filters:

$$H_{\text{dev}}(\omega) = H_{\text{sens}}(\omega) \cdot H_{\text{corr}}(\omega) \cdot H_{\text{out}}(\omega). \quad (11)$$

The adjustment procedure is the following: First the seismometer transfer function is determined by measuring the seismometer response with some initial parameters of the correcting circuit. After that, the transducer transfer function is calculated by dividing the seismometer transfer function by a product of the electronic filter transfer functions:

$$H_{\text{sens}}(\omega) = \frac{H_{\text{dev}}(\omega)}{H_{\text{corr, initial}}(\omega) \cdot H_{\text{out}}(\omega)}. \quad (12)$$

After that the parameters of the correcting filters are remade using adjustable electronic parts. The mathematical modeling is used to find the required modifications. The result of the



Electrochemical Seismometers of Linear and Angular Motion, Fig. 7 (a) Transfer function $H_{\text{sens}}(\omega)$ of a horizontal MET transducer, (b) product of the transfer

functions $H_{\text{corr}}(\omega) \cdot H_{\text{out}}(\omega)$ of equalization filters, (c) resulting transfer function $H_{\text{dev}}(\omega)$ of a horizontal MET seismometer

correction is checked on the shake table. If necessary, the adjustment procedure is repeated successively until the measured response function $H_{\text{dev}}(\omega)$ meets the standard curve within a tolerance predetermined as acceptable. The iterative adjustment process eliminates the influence of natural spread of electronic component values. The equalization procedure and the final result of the adjustment are illustrated on Fig. 7a–d, where real data are used as an example. The amplitude vs. frequency characteristics are shown, where Fig. 7a is the measured transfer function to applied velocity of the transducer, Fig. 7b is the transfer function for the electronic filters, Fig. 7c is the final amplitude vs. frequency curve, and Fig. 7d is the final phase vs. frequency curve for the seismometer.

The accuracy of the filter equalization procedure is limited by the tolerance of the electronic components and the accuracy of the shake table used to determine the transfer function. The usual variation in the transfer function for a broadband seismometer with filter equalization stays

within ± 1 dB for sensitivity and $\pm 30^\circ$ in the passband. What's more, the filter equalization is unable to correct any changes in the transducer's transfer function that may occur in the long term after the factory adjustment. Now this type of equalization is used only in a few models of electrochemical seismometers.

More precise devices use the feedback design. The commonly used technique for a practical feedback implementation is a moving coil system. The flexible membrane (see Fig. 4a) is equipped with a coil placed in a magnetic field of a firmly installed strong magnet. Applying a current into the coil gives rise to a force acting on the moving liquid inside the sensor. Usually, negative feedback is implemented, so that the force produced by the coil acts oppositely to the inertial force. It is important to note that for this approach, due to the high original conversion coefficient of the transducer, the decrease in the conversion factor due to the negative feedback, practically, does not affect the instrument resolution.

The block diagram of the seismometer equipped with a feedback mechanism is presented in Fig. 8, where $H_{\text{corr}}(\omega)$ represents the transfer function of filters and amplifiers in direct circuit, $H_{\text{fb}}(\omega)$ is the same of the feedback circuit, and $H_{\text{out}}(\omega)$ is the transfer function of the output filter.

The most important feature of a feedback is that, provided the system remains stable and the amount of feedback is sufficient, the output function $H_{\text{dev}}(\omega)$ depends mostly on the feedback transfer function $H_{\text{fb}}(\omega) \cdot H_{\text{coil}}(\omega)$ rather than on the transducer transfer function $H_{\text{sens}}(\omega) \cdot H_{\text{corr}}(\omega)$.

Indeed

$$H_{\text{dev}}(\omega) = \frac{H_{\text{sens}}(\omega) \cdot H_{\text{corr}}(\omega)}{1 + H_{\text{sens}}(\omega) \cdot H_{\text{corr}}(\omega) \cdot H_{\text{fb}}(\omega) \cdot H_{\text{coil}}(\omega)} H_{\text{out}}(\omega). \quad (13)$$

In case the amplification inside the closed-loop system is much higher than 1 in absolute value $|H_{\text{sens}}(\omega) \cdot H_{\text{corr}}(\omega) \cdot H_{\text{fb}}(\omega) \cdot H_{\text{coil}}(\omega)| \gg 1$ and Eq. 13 is approximated by the following formula:

$$H_{\text{dev}}(\omega) = \frac{1}{H_{\text{fb}}(\omega) \cdot H_{\text{coil}}(\omega)} H_{\text{out}}(\omega). \quad (14)$$

The feedback circuit results in a counterforce (i.e., a force compensating effect of the acting acceleration) proportional to the current flow. So $H_{\text{coil}}(\omega)$ can be expressed as a function that converts voltage to acceleration.

The practical implementation of the electrochemical seismometer equipped with feedback is:

Task 1: Providing sufficient amplification to have a “deep” feedback

Actually, this is the simplest task among others listed below. First, the transducer itself has a high original conversion factor. Additionally, the electronic industry produces a variety of low-noise operational amplifiers which can be used for additional amplification of a signal.

Task 2: Producing an output proportional to either acceleration or velocity, depending on the instrument type

In case the electronics in the feedback has frequency-independent response, the transfer function of the instrument would be proportional to acceleration. The output filter transfer function $H_{\text{out}}(\omega)$ is used to shape the low- and high-cutoff frequencies.

A velocimeter is more common for broad-band seismometers. In this case, an additional circuit, an integrator as a part of the output filter ($H_{\text{out}}(\omega) \sim \frac{1}{\omega}$ in the passband), or a differentiator at the feedback amplifier ($H_{\text{fb}}(\omega) \sim \omega$ in the passband) should be added (both shown in dashed lines in Fig. 8).

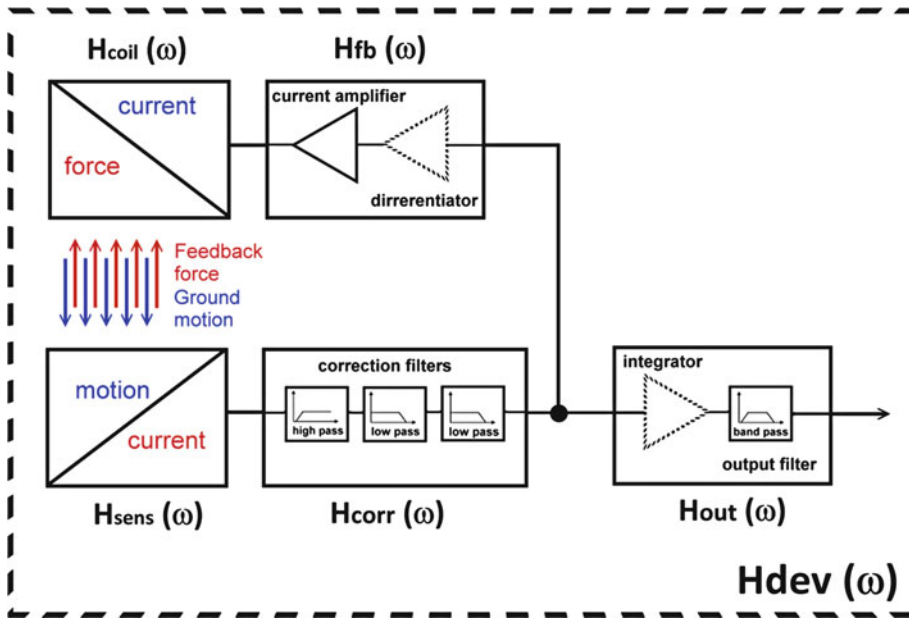
Whenever an integrator or differentiator is used, the transfer function of the instrument would give an output voltage proportional to the ground velocity.

The introduction of an integrator into circuits at lower frequencies is limited by electronic parts leakage and op-amps offset. Also, the substantial variation of capacitors' impedance over the passband in frequency-dependent circuits may impair the noise performance of a device. Therefore special schematic techniques and careful amplifier selection should be made in order to maintain the optimal noise performance of the integrator or differentiator stage.

Another aspect of introducing a differentiator into the feedback circuit is that it brings a leading $\pi/2$ phase shift in the loop amplification. Typically, this phase shift improves feedback stability, and velocimeters with a differentiator in the feedback have better performance at higher frequencies in comparison with devices equipped with an output integrator.

Task 3: Ensuring the stability of the closed-loop system

Maintaining the feedback loop stable is a comparatively sophisticated task.



Electrochemical Seismometers of Linear and Angular Motion, Fig. 8 Block diagram of a closed-loop electrochemical seismometer

Being negative in the middle of the frequency band, the feedback may even change its sign and become positive at some frequency outside the high- or low-frequency band limits due to the phase response variation as a function of frequency. To avoid self-oscillation, it is important to keep the amplification inside the closed-loop system at such a frequency less than 1 in absolute value by all means. This task not only requires considerations of stability when the system is being adjusted, but it's also essential to maintain the closed-loop stable over all temperature ranges and under transitions processes after the device is just powered on or overloaded by a strong input signal.

Changes of an electrochemical seismometer frequency response with variation of the feedback depth are illustrated in Fig. 9. The device response at a low feedback depth is close to the curve of an uncorrected sensor with no visible flat region at all. The deeper the feedback, the wider is the flat region in the device response, but at the same time the lower is the overall sensitivity of the device.

Very strong feedback depth (dark red curve on the Fig. 9) could result in the feedback loop becoming unstable during the transition processes or under variations in temperature.

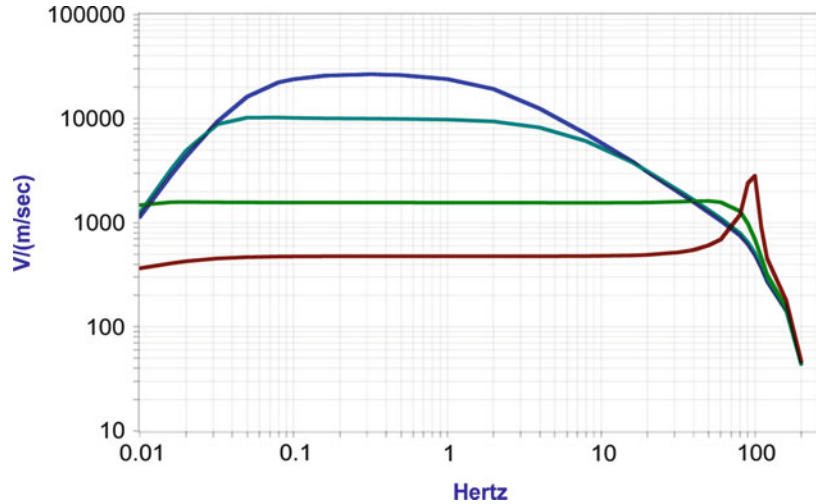
Most electrochemical seismometers have maximum feedback depth in the range from 10 to 50. The feedback for the electrochemical accelerometers could be as big as few hundred.

Task 4: Sufficiency in feedback dynamic range

For the sake of keeping the dynamic range at a maximum, the feedback amplifier should provide sufficient current into the coil. The main limitation comes from the fact that modern systems are continuously using a lower and lower voltage supply, while coil size and weight of the moving body cannot be correspondingly reduced. Moreover, a heavier moving body of a sensor gives less thermal self-noise. A low-noise seismometer usually needs a larger current for the feedback coil.

The procedure of adjusting a sensor with feedback equalization consists of two stages. During the first step, a proper frequency

Electrochemical Seismometers of Linear and Angular Motion, Fig. 9 Device response vs. feedback depth (depths 1, 5, 40, 120)



response is made paying no attention to actual device sensitivity. In the second step, the required device sensitivity is adjusted.

The open-loop characteristics can be obtained by means of an external digital-to-analog converter (DAC) and analog-to-digital converter (ADC) with a harmonic response scan. The frequency response of the open loop outside the operational range is essential to ensure the system stability. After the initial open-loop response is obtained, the filter correction is made to adjust the feedback amplification, to make a flat response in the passband and predetermined behavior at low- and high-cutoff frequencies. The process is iterative in the same way as it was for the non-feedback devices.

The main difference compared to the non-feedback case is that the adjusting procedure gives a much more accurate response of a seismometer within the passband than for the non-feedback sensor. Generally the high- and low-cutoff frequencies are extended by 2–3 octaves. In parallel with the feedback adjustment, a fine-tuning of the output filter is done. In case the closed-loop response is actually flat over the passband, the device characteristics coincide with the output filter characteristics. This decreases the complexity of further device response presentation in terms of poles and zeroes.

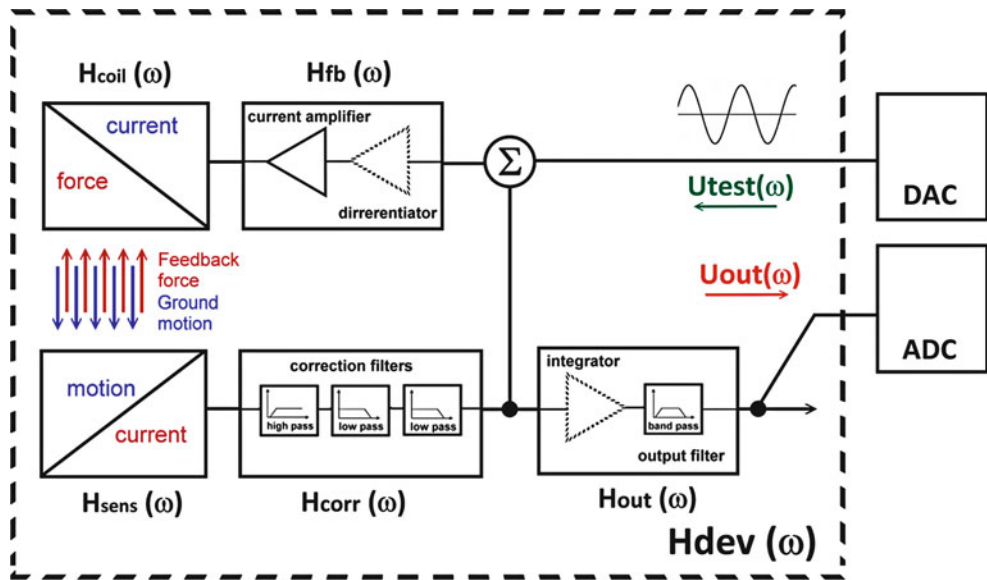
The last step in manufacturing the device is the adjustment of its sensitivity. The operation can be carried out by adjusting the overall gain of the output filter's $H_{out}(\omega)$.

After this final adjustment, the resulting frequency response characteristic of a closed-loop device is measured and verified. The main point of this method is in simulating ground motion by adding of an external signal into the closed-loop circuit. This signal addition works in the same way as a real input signal. An external sine wave generator can be used to provide such a signal at various frequencies and thereby get a full calibration function.

The block diagram of the closed-loop calibration procedure is illustrated in Fig. 10 where the sine wave test signal from a generator is added to the closed-loop signal in the summation unit Σ . The signal disturbance introduced into the closed-loop and the output response are measured by an external 2 channel analog-to-digital converter.

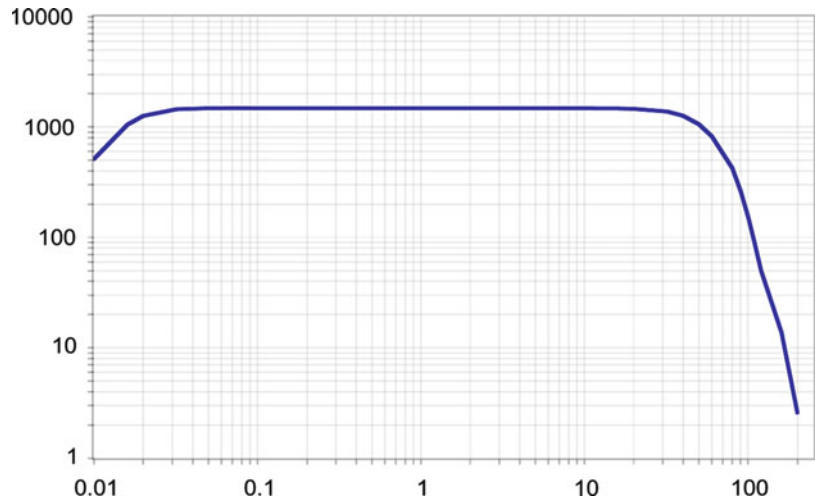
The examples of the resulting calibration curves are shown in Figs. 11 and 12.

The closed-loop calibration procedure makes the subsequent maintenance and validation of a device possible with a common DAC–ADC and a reference sensor with verified sensitivity and noise performance. The test is carried out only by means of the



Electrochemical Seismometers of Linear and Angular Motion, Fig. 10 Closed-loop self-calibration

Electrochemical Seismometers of Linear and Angular Motion, Fig. 11 Seismometer response obtained by the self-calibration test. The actual CME-6011 sensor was used to generate an example



standard device connector and does not require any intervention into the sensor or disassembling the device. The presence of a shake table becomes unnecessary. The device sensitivity can be determined using either factory-determined calibration constants or by analyzing the simultaneous recording made with a sensor with known sensitivity. The noise performance of a device can only be obtained by analyzing the simultaneous record made by at least

two sensors placed close to each other. One of the approaches is to calculate non-correlating components in the signal of two sensors of the same type with identical parameters (sensitivity, phase response, passband). Another approach is to make simultaneous recording with a sensor with far better noise performance. In the latter case the noise can be calculated by subtracting the low-noise sensor's signal from the tested sensor's signal.



Electrochemical Seismometers of Linear and Angular Motion, Fig. 12 MET geophones and accelerometers

Overview of Commercial Electrochemical Seismometers

Broadband Seismometers

The MET seismometers available cover a wide area of applications from permanent seismic stations to seismic field surveys, borehole measurements, and autonomous ocean-bottom systems (OBS). All devices are characterized by robustness, reliability, light weight, and easy installation procedure. None of them require any type of mass lock in transportation or mass centering during installation. They are operable under tilts up to $\pm 10^\circ$ and more, and some models are able to operate under any tilt angles. This feature is extremely useful for OBS operation. The lightest and the most power-efficient models of three component seismometers like CME-4211OBS have weight of 2.2 kg and power consumption

of 90 mW during operation. The dynamic range for the best models is in the range 135–150 dB, while self-noise of the devices goes down below the lowest observed vertical seismic noise levels estimated by the New Low Noise Model (NLNM) in the 15 s–1 Hz range. The majority of models are equipped with closed-loop force feedback equalization. However, devices with filter equalization are lighter and use less power, which is essential for self-contained systems.

Seismometers for permanent installation have better noise performance due to heavier moving parts and are therefore more complex in design. These devices have the lowest cutoff frequency and due to that are less suitable for portable applications. However, for vault installation such parameters as power consumption and weight have less significance. Like other MET devices, they don't require any special arrangements during installation and operation. Models designed for self-contained systems, like field geological surveys with frequent location changes or for oceanic exploration, feature energy effectiveness, light weight, and compact size. Low-frequency cutoff as high as 0.1 Hz is more common for these instruments (Tables 1 and 2).

MET-Based High-Sensitive Geophones and Seismic Accelerometers

The main advantages of MET-based seismic geophones and accelerometers with negative feedback lie in their extremely high coefficient of conversion of mechanical motion into an electrical signal, the possibility of a significant (compared to a traditional geophone) expansion of the frequency range both in the low- and high-frequency directions, and the simplicity and reliability of the design (Agafonov et al. 2010; Agafonov et al. 2014). Until recently, the instruments of this kind were used almost exclusively for measurements in the low-frequency domain. However, the developments of recent years (Krishtop et al. 2012) have made it possible to produce the devices with an upper operating frequency that exceeds the values typical for most geophones.

The experience learned from the development of such sensors indicates that the optimal

Electrochemical Seismometers of Linear and Angular Motion, Table 1 The comparison chart for MET sensors for permanent seismic stations (sensors are sorted in alphabetical order)

Sensor type, producer, country	BB603 PMD Scientific, Inc., USA http://www.pmdsci.com/pdfs/BB603.pdf	CME-6211, R-sensors LLC, Russia http://www.r-sensors.ru/1_products/Descriptions/CME-6211.pdf	EP-300, Eentec, USA http://www.eentec.com/EP-300_data.htm
Operating principle	Electrochemical transducer	Electrochemical transducer	Electrochemical transducer
Equalization type	Force-balancing feedback	Force-balancing feedback	Force-balancing feedback
Output signals	Triaxial	Triaxial	Triaxial
Sensitivity	2,000 V/(m/s)	2,000 V/(m/s)	2,000 V/m/s
Frequency bandwidth	Standard: 0.0167–50 Hz	Standard: 0.0167–40 Hz	Standard: 0.0167–50 Hz
	Optional: 0.0083–50 Hz	Optional: 0.0083–40 Hz	Optional: 0.0083–50 Hz
Maximum output swing	± 10 V	± 20 V	± 10 V
	(± 20 V _{p-p} differential)	(± 40 V _{p-p} differential)	(± 20 V _{p-p} differential)
Self-noise	Below USGS NLNM in 0.05–5 Hz range	Below NLNM in 0.1–7 Hz band	Below NLNM in 0.9–12 Hz band
Temperature range	Standard –12 °C to +55 °C	Standard –12 °C to +55 °C	Standard –12 °C to +55 °C
		Extended –40 °C to +55 °C	
Nominal supply voltage, consumption	Standard: 12Vdc, 28 mA	Standard: 12Vdc, 55 mA	Standard: 12Vdc, 30 mA
	Low power: 12Vdc, 12 mA	Low power: 12Vdc, 27 mA	Low power: 12Vdc, 12 mA
Mass lock, centering	None required	None required	None required
Maximum installation tilt	$\pm 10^\circ$	$\pm 15^\circ$	$\pm 10^\circ$
Case diameter/height	200 mm/220 mm	272 mm/240 mm	200 mm/220 mm
Weight	11 kg	12 kg	11 kg

operating characteristics are attained when MET-based sensors are equipped with an electrodynamic negative feedback mechanism. With a feedback, the instrument scale factor of 300 V/m/s is achieved without using any additional amplification, which is equal to or greater than that achieved by other known low-frequency geophones. At the same time, for a frequency range starting at 1 Hz, the MET geophones are much smaller than typical electrodynamic geophones.

The external view of sensors is shown in Fig. 12. MTSS-1001 is a single-axis, ultracompact, high-gain, 1-Hz seismic sensor. MTSS-2003 contains three orthogonal oriented MTSS-1001 seismic sensors. It is probably the

world's smallest tri-axis 1-Hz seismometer. MTSS-2003 is designed primarily for measuring small low-frequency signals. It combines high gain, small size, and low power consumption.

The 3-component high-performance accelerometer MTSS-1033A and its 1-component version MTSS-1031A are designed for strong motion seismic measurements, industrial vibration monitoring, vibration isolation analysis, etc. MET accelerometers combine very small size and high sensitivity over wide frequency band. The wide dynamic range, low distortion, high temperature stability, and competitive price make it ideal and cost-effective solution for various applications such as earthquake, strong

Electrochemical Seismometers of Linear and Angular Motion, Table 2 The comparison chart for seismic MET sensors for field work (sensors are sorted in alphabetical order)

	BB313	CME-4211	CME-6011	MP403
Sensor type, manufacturer	PMD Scientific, USA http://www.pmdsci.com/pdfs/BB313.pdf	R-sensors, Russia http://www.r-sensors.ru/1_products/Leaflets/4211.pdf	R-sensors, Russia http://www.r-sensors.ru/1_products/Descriptions/CME-6011.pdf	PMD Scientific, USA http://www.pmdsci.com/pdfs/MP403.pdf
Operating principle	Electrochemical motion transducer	Electrochemical motion transducer	Electrochemical motion transducer	Electrochemical motion transducer
Equalization type	Force-balancing feedback	Vertical: force-balancing feedback Horizontal: filter equalization	Force-balancing feedback	Force-balancing feedback
Output signals	Triaxial	Triaxial	Triaxial	Triaxial
Sensitivity	Standard: 2,000 V/(m/s)	Standard: 2,000 V/(m/s)	Standard: 2,000 V/(m/s)	Standard: 2,000 V/(m/s)
Frequency bandwidth	Standard: 0.033–50 Hz	Standard: 0.033–50 Hz Optional: 0.1–100 Hz	Standard: 0.033–40 Hz Optional: 0.1–50 Hz	Standard: 0.1–50 Hz
Maximum output swing	± 10 V (± 20 Vp–p diff.)	± 15 V (± 30 Vp–p diff.)	± 20 V (± 40 Vp–p diff.)	± 10 V (± 20 Vp–p diff.)
Temperature range	Standard -12 °C to $+55$ °C	Standard -12 °C to $+55$ °C	Standard -12 °C to $+55$ °C	Standard -12 °C to $+55$ °C
		Extended -40 °C to $+55$ °C	Extended -40 °C to $+55$ °C	
Nominal supply voltage, consumption	Standard: 12Vdc, 28 mA	Standard: 12Vdc, 27 mA	Standard: 12Vdc, 26 mA	Standard: 12Vdc, 28 mA
	Low power: 12Vdc, 12 mA	Low power: 12Vdc, 8 mA	Low power: 12Vdc, 10 mA	Low power: 12Vdc, 12 mA
Mass lock, centering	None required	None required	None required	None required
Maximum installation tilt	$\pm 12^\circ$	$\pm 15^\circ$	$\pm 15^\circ$	$\pm 12^\circ$
Case diameter/height	155 mm/185 mm	180 mm/140 mm	204 mm/210 mm	200 mm/220 mm
Weight	5.5 kg	4.3 kg	6.6 kg	9 kg

motion measurements, or structural monitoring. At the same time, unlike mechanical force-balanced accelerometers, the MET accelerometers cannot measure the DC component of the seismic signal. Although, the seismic signal is oscillating, the DC signal has a significant practical importance. The DC signal makes it possible to calibrate the sensor by rotating or tilting it relative to the gravity vector and to determine the sensor's axes orientation with respect to the vertical direction. In addition, the difference in

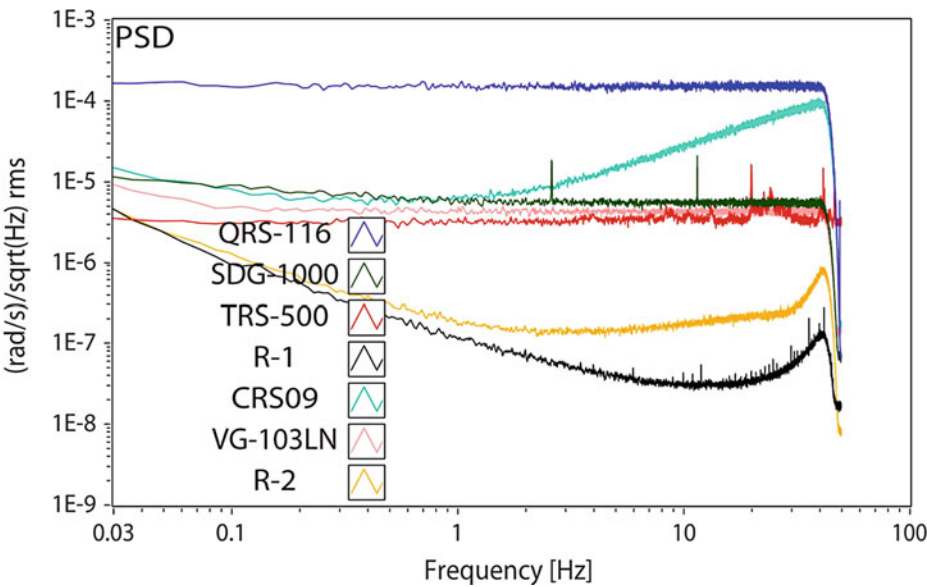
low-frequency phase behavior makes it difficult to compare the responses of a MET accelerometer and a mechanical DC accelerometer (Table 3).

Electrochemical Angular Motion Seismometers

Interest in the pure rotational seismometer is driven by the suggestion that adding a new type of measurements would allow to better characterize the response of the structures to seismic

Electrochemical Seismometers of Linear and Angular Motion, Table 3 The technical performance parameters for MET compact low-frequency geophones and accelerometers

Performance	MTSS-1001	MTSS-1041A	MTSS-1031A
Frequency range	1–300 Hz	0.1–120 Hz	0.1–120 Hz
Sensor noise	150 nm/s, rms over frequency range	110 ng/√Hz	150 ng/√Hz
Conversion factor	250 V/(m/s)	6 V/g	2.4 V/g
Maximum negative feedback depth	20	100	150
Maximum applied acceleration	–	±2 g	±4 g
Dynamic range	>110 dB	>110 dB	>120 dB
Non-linearity	<1 %	<1 %	<1 %
Orientation	Any direction	Any direction	Any direction
Weight	180 g	180 g	180 g
Temperature	“–55 ~ +65 °C”	“–55 ~ +65 °C”	“–55 ~ +65 °C”
Power	<80 mW	<80 mW	<80 mW



Electrochemical Seismometers of Linear and Angular Motion, Fig. 13 Comparison of the self-noise for some electrochemical angular seismometers (models R1 and R2) with other angular motion sensors (Lin

et al. 2013). The models QRS-116, SDG-1000, and CRS-09 are angular motion sensors based on MEMS technology. The TRS-500 and VG-103LN are fiber-optic gyros

input, to measure more accurately the seismic field spatial distribution, to be able separate between modes of seismic waves based on their polarization, and to determine better the site effect (Lee et al. 2011; Bernauer 2012a; Kapustian et al. 2013). Detailed information can be found at web page www.rotational-seismology.org, especially devoted to different

aspects of the seismic rotational measurements for a variety of applications. A special issue “Advances in rotational seismology: instrumentation, theory, observations, and engineering” was issued in Journal of seismology, volume 16, # 4 in October 2012.

The mechanical configuration of the rotational seismometer, based on the electrochemical

Electrochemical Seismometers of Linear and Angular Motion, Table 4 Technical parameters of commercially available rotational seismometers

Model	METR-03/01, LLC R-sensors, http://www.r-sensors.ru/1_products/Seismic_sensors_METR-03_CME-106C.pdf , R1 eentec, http://www.eentec.com/R-1_data_new.htm ;	R2, Eentec, (Bernauer et al. 1012)	METR-13/11, LLC R-sensors, http://www.r-sensors.ru/1_products/Seismic_sensors_METR-03_CME-106C.pdf ,
Output	Angular rate	Angular rate	Angular rate
Sensor type	Electrochemical	Electrochemical	Electrochemical
Noise at 1 Hz	10^{-7} rad/s/ $\sqrt{\text{Hz}}$ (for R1, see Fig. 13)	$2 \cdot 10^{-7}$ rad/s/ $\sqrt{\text{Hz}}$ (see Fig. 13)	
Full-scale range	0.1 rad/s	0.3 rad/s	0.1 rad/s
Bandwidth	0.033–50 Hz	0.033–50 Hz	1–150 Hz
Size	$120 \times 120 \times 102$	$120 \times 120 \times 102$	$120 \times 120 \times 102$

sensing technology, is presented in Fig. 4 above. The sensor consists of the toroidal channel, filled with an electrolyte, and the sensitive cell placed across the channel, converting liquid motion into the electrical response. An expansion volume is used to compensate for the temperature expansions of the liquid. Electrochemical sensors are probably the only rotational sensors which are commercially available at a reasonable cost with a resolution appropriate to seismic applications. Figure 13 shows the spectral noise density spectrum of rotational electrochemical sensors compared with other commercial devices measuring rotation. The parameters of some of the commercially available electrochemical rotational seismometers are shown in Table 4. The independent test data can also be found in Bernauer et al. (2012b).

Summary

The electrochemical seismometers offer an alternative to regular electromechanical devices. They are compact, robust, and easy to install and provide good quality data. The operation principles of the electrochemical seismometers are based on using the charge transfer variations due to electrolyte motion in the four-electrode

electrochemical cell. Among the commercial devices developed on the base of this technology are broadband seismometers, compact high-sensitive MET geophones, accelerometers, and unique angular seismometers.

Cross-References

- Downhole Seismometers
- Ocean-Bottom Seismometer
- Principles of Broadband Seismometry
- Seismic Accelerometers
- Sensors, Calibration of

References

Agafonov VM, Krishtop VG (2004) Diffusion sensor of mechanical signals: frequency response at high frequencies. *Russ J Electrochem* 40:537–541

Agafonov VM, Egorov EV, Zaitsev DL (2010) Molecular electronic linear accelerometers. Preliminary test results. *Gyroscopy Navigation* 1(4):246–251

Agafonov VM, Egorov IV, Shabalina AS (2014) Operating principles and technical characteristics of a small-sized molecular–electronic seismic sensor with negative feedback. *Seismic Instrum* 50(1):1–8

Bernauer M, Fichtner A, Igel H (2012a) Measurements of translation, rotation and strain: new approaches to seismic processing and inversion. *J Seismol* 16(4):669–681

- Bernauer F, Wassermann J, Igel H (2012b) Rotational sensors—a comparison of different sensor types. *J Seismol* 16:595–602. doi:10.1007/s10950-012-9286-7
- Collins JL, Richie WC, English GE (1964) Solion infrasound microphone. *J Acoust Soc Am* 36:1283–1287
- Huang H, Agafonov V, Yu H (2013) Molecular electric transducers as motion sensors: a review. *Sensors* 13:4581–4597. doi:10.3390/s130404581
- Hurd RM, Lane RN (1957) Principles of very low power electrochemical control devices. *J Electrochem Soc* 104:727–730
- Kapustian N, Antonovskaya G, Agafonov V, Neumoin K, Safonov M (2013) Seismic monitoring of linear and rotational oscillations of the multistory buildings in Moscow. *Geotechnical Geological Earthquake Eng* 24:353–363
- Kozlov VA, Safonov MV (2003) Self-noise of molecular electronic transducers. *Tech Phys* 48:1579–1582
- Kozlov VA, Korshak AN, Petkin NV (1991) Theory of diffusion-type transducers for ultralow electrolyte flow-rates. *Sov Electrochem* 27:16–21
- Krishtop VG, Agafonov VM, Bugaev AS (2012) Technological principles of motion parameter transducers based on mass and charge transport in electrochemical microsystems. *Russ J Electrochem* 48(7):746–755
- Larcam CW (1965) Theoretical analysis of the solion polarized cathode acoustic linear transducer. *J Acoust Soc Am* 37:664–678
- Lee WHK, Evans JR, Huang B-S, Hutt CR, Lin C-J, Liu C-C, Nigbor RL (2011) Measuring rotational ground motions in seismological practice. Information sheet. IS 5.3. Feb 2011; doi:10.2312/GFZ.NMSOP-2_IS_5.3
- Lidorenko NS, Ilin BI, Zaidenman IA, Sobol VV, Shchigorev IG (1984) An introduction to molecular electronics. Enegoatomizdat, Moscow
- Lin C-J, Liu G (2013) Calibration and applications of a rotational sensor www.iris.edu/hq/sits_13_docs/mon/Lin.pptx
- Panferov AP, Kharlamov AV (2001) Theoretical and experimental study of an electrochemical converter of a pulsing electrolyte flow. *Russ J Electrochem* 37:394–398
- Reznikova LA, Morgunova EE, Bograchev DA, Grigin AP, Davydov AD (2001) Limiting current in iodine–iodide system on vertical electrode under conditions of natural convection. *Russ J Electrochem* 37:382–387
- Sun Zh, Agafonov VM (2010) 3D numerical simulation of the pressure-driven flow in a four-electrode rectangular micro-electrochemical accelerometer. *Sensors Actuators B Chem* 146:231–238. www.elsevier.com/locate/snb
- Volgin VM, Volgina OV, Bograchev DA, Davydov AD (2003) Simulation of ion transfer under conditions of natural convection by the finite difference method. *J Electroanal Chem* 546:15–22
- Wittenborn AF (1958) Analysis of a logarithmic solion pressure detector. *J Acoust Soc Am* 30:683–683

Emergency Response Coordination Within Earthquake Disasters

Jon Mitchell¹ and David Brunson²

¹Jon Mitchell Emergency Management Ltd., Queenstown, New Zealand

²Kestrel Group Ltd., Wellington, New Zealand

Synonyms

All hazards; Command and control; Community resilience; Comprehensive emergency management; Crisis management; Disaster management; Gold silver bronze; ICS; Incident command system; Infrastructure resilience; Life utilities coordinator; Politics; Risk management

Introduction

Responses to significant earthquakes are challenging and complex. This entry explores the key aspects of earthquake response that are particularly pertinent to the context of earthquake engineering.

Managing the risks of hazards, such as earthquakes, is not merely responding to them when they occur or even preparing effectively for those responses. Contemporary emergency management applies a much broader brush than the 1950s era, response-focused “civil defense” approach that held sway in most jurisdictions until the 1990s or early 2000s. Contemporary comprehensive emergency management has increasingly encompassed: developing adequate appreciations of the risks that communities face; working to enhance resilience, reduce or mitigate exposure to those risks, and build readiness or

Jon Mitchell was NZ Coordinated Incident Management National Steering Committee from 2007 to 2014 and Massey University, Joint Centre for Disaster Research, 2014 to present

David Brunson was Chair, NZ Lifelines Committee 1999 to present and NZ USAR Engineering Leader from 2004 to 2012

preparedness to respond and recover; respond in a coordinated, adaptive, and effective manner including all relevant organizations and effected communities; and initiate regenerative recovery parallel with response so as to further enhance resilience and make the most of the opportunities that disaster provides to communities.

The 1980s and 1990s saw the development of the “all hazards” approach to emergency management, intended to broaden the application of preparedness for hazards beyond a handful of locally relevant natural hazards and the then diminishing threat of nuclear conflagration, with the fall of the Berlin Wall and the end of the cold war. The all hazards approach tended to suggest that the arrangements, capabilities, and resources put in place by government agencies at all levels for the major hazards they faced could then be applied any and all hazards that might then arise.

The world has experienced increasing exposure to natural hazards and their consequences since the 1960s, as populations have boomed in climatic and geological risk-prone areas in less developed countries and as communities and industries in developed economies became increasingly dependent on technology, particularly electricity and telecommunications. Increasing exposure to risks has resulted in continual improvement and refinement of approaches to managing risks internationally, culminating in the United Nations making the 1990s the “International Decade for Natural Disaster Reduction” and the initiation of programs such as the United Nations Office for International Disaster Risk Reduction (UNISDR).

UNISDR projects, including the “Hyogo Framework for Action 2005–2015 Building the Resilience of Nations and Communities to Disasters,” have seen emergency management, and disaster risk management in general take a more comprehensive turn. Nations and communities, in their own ways, are increasingly attempting to understand the natural and human-induced hazards that they face; to inform risk reduction legislation, regulations, and programs; to build community and organizations’ resilience and

response capabilities; and, of more relevance to this entry, to develop and deliver effective response and recovery capabilities and outcomes.

This entry commences with an overview of more formal leadership and coordination of responses. It then goes on to cover the topics of monitoring and analysis of ongoing seismic activity during response; acquiring, analyzing, managing, and sharing information during disaster response; capturing and making use of physical and social impact data; urban search and rescue (USAR); medical response; emergency welfare; and emergency food, water, and sanitation services. This entry concludes with discussion of post-quake building management and critical infrastructure or lifeline utility coordination: essential elements of response and recovery.

Leadership and Coordination

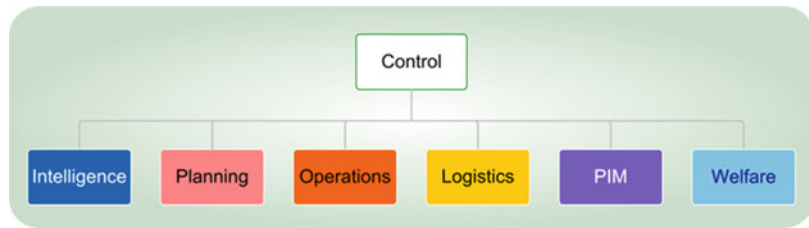
When significant earthquakes strike, the challenges for those in positions of leadership or who are responsible for response coordination are many and varied. Not the least due to the reality that affected communities, ethnic and social networks, businesses, service clubs, and faith organizations, emergency services in the area at the time will be mounting their own responses, while official responses will generally still be regrouping and reestablishing communication with their field resources and external support. The very nature of damaging quakes makes them some of the most difficult hazards to manage responses to, as earthquakes tend to undermine the ability of response organizations to deliver their services more than other hazards tend to.

Leadership and coordination, nonetheless, are crucial to ensuring that the needs of affected communities and the services they depend upon are met in as timely, effective and equitable manner as possible. Key to effective emergency leadership and coordination is legislation that adequately provides for the situations that are likely to arise from major hazards.

The legislative requirement to identify the major hazards and risks that respective national,

Emergency Response Coordination Within Earthquake Disasters,

Fig. 1 Basic response coordination structure – NZ CIMS manual (Ministry of CDEM 2014)



state, regional, and local communities face is an essential foundation to emergency management – as are legal responsibilities to plan at national and subnational levels to manage hazards and their risks holistically.

Coordination of responses to the immediate and longer-term consequences of earthquakes is considerably more important than command and control-style leadership, as no single individual or organization can have adequate knowledge of impacts and response to be able to comprehensively lead or direct responses. More implicit coordination within and between organizations and communities is more influential than individual or organizational leadership and direction. The concept of disaster response management that is increasingly seen as most effective in understanding and delivering responses that actually work is “network-centric operations” (Von Lubitz et al. 2008) or, in short, net-centric operations.

The impacts of earthquakes tend to not only cause buildings and other vertical and horizontal infrastructure to fail or be compromised but also to cause injuries, distress, entrapments, and fatalities. They also disrupt response capabilities, despite how well planned and resourced those capabilities are, due to overload and damage to telecommunications systems, the initial confusion in the community and official responders, and direct impacts on responders and their equipment and facilities – as well as their families, their homes, and the services they depend upon.

Effective coordination depends on a range of factors. An essential prerequisite is a shared understanding of and commitment to response coordination structure, roles, terminology, and processes. Each nation tends to have adopted or developed an incident coordination, command, or management system adapted to suit the national

context and, if necessary, that of its neighboring or closely related nations. All incident management systems provide a common structure, roles, processes, etc. for overall response coordination at various levels of response. Each response organization applies the standard system to their own operating procedures to greater or lesser degrees, depending on the requirements of the overall system and any statutory or regulatory mandate attached to it.

Incident management systems tend to provide an organizational structure and hierarchy with a response director or controller in the senior position, albeit often under the direct influence of senior appointed or elected officials. Reporting to the “controller,” “director,” “commander,” or similar, are managers of response management functions, including public information management, operations, intelligence, planning, and logistics (Fig. 1).

Note: “PIM” is public information management in the New Zealand context and “public affairs” or similar in other contexts.

Response Coordination Principles

Levels of response coordination tend to include:

- **Site** – management closest to the response field activities themselves. In an earthquake context, this will tend to be a large building, block or buildings, or a small geographic area encompassing areas of structural damage and geological damage, including liquefaction and rockfall, landslide, etc. or social impact or social support.
- **Local** – coordination and direction of response activities and resources in support of site-level response, often at suburb or

municipal level. Management of local critical infrastructure will usually be represented in incident management organizations at this level of coordination.

- **Regional** – support, coordination, and direction of local response activities and resources, often at the political level above municipalities, such as county, region, or prefecture.
- **State and or national** – support and coordination of resources and activities in support of lower levels of response. Statewide or national critical infrastructure providers, in some cases clusters like utilities and service providers, link into the requisite level of official response coordination.

Incident Coordination

Incident or site-level coordination tends to occur close to the area of impact and focus of response, from what is usually referred to as an “incident command post,” “incident control point,” or similar. Incident management at this level is usually populated by low-level managers of the agencies contributing to the response and may include supervisors of field components of critical infrastructure providers. Despite this use of the term “incident,” the same term is often applied to all levels of emergency and disaster.

An incident control point can be provided by one person or an incident management team of very few to tens of individuals from the range of agencies responding directly to the incident and those supporting the response organizations. Higher-level coordination facilities, which tend to be preestablished and equipped, are ideally post-earthquake operational in areas where that hazard is prevalent and may be staffed by a few individuals and agencies or by hundreds of personnel from a wide range of government agencies, nongovernment organizations, and the private sector.

Key tools of incident management systems include:

- Collective development of “action plans” to provide a basis for explicit response

coordination and collaborative delivery of tasks assigned within actions plans at each level of response

- “Planning cycles” providing a systematic process for the development, delivery, and review of action plans and resultant actions
- “Situation reports” compiling relevant contextual, impact, and response information within the respective level of response coordination to inform decision-making
- “Status reports” and “incident reports” that inform situation reports and the maintenance of ongoing shared situational awareness
- “Resource requests” from one level of response coordination to that above it for resources that are not able to be acquired at sites or within the respective level of coordination
- Situation briefings, information sharing, informal communication, cooperation, and shared situational awareness support “implicit coordination”
- Intelligence collection, analysis and dissemination cycles, and action plan and task plan development are components of “explicit coordination”

Legislation

All nations have response management legislation of some sort. Emergency management legislation tends to provide the requirement or mandate to plan for emergencies and appoint appropriately trained and experienced individuals to response direction or control roles for which a range of emergency powers are provided for in the legislation. These powers may be retained at the national or state level, but are usually reflected in similar powers within subsidiary levels of government. Emergency or disaster management legislation usually provides for a senior elected official to declare a “state of emergency,” “disaster declaration,” or similar for the respective level and/or lower levels of government – releasing emergency powers and requiring response organizations, critical infrastructure providers, and the media to act in a coordinated manner.

Phases of Emergency Response

Initial objectives of emergency response are around providing support to injured, trapped, isolated, and evacuated individuals and groups. Much of the initial active field response will be focused on search and rescue in areas where buildings and facilities have suffered damage during the initial or subsequent shaking, structural failure, liquefaction, inundation, fire, or other consequential hazard. Search and rescue may continue for weeks in some cases following the initial quake, if there is likelihood of survivors being located and extricated.

While search and rescue activities are underway, those who are either sheltering in place or have evacuated, either under their own auspices, assisted by other members of the community, or by way of a managed or compulsory evacuation, will require further support. This is discussed further under “Emergency Welfare and Emergency Food and Water Supplies” later in this entry.

Planning for and establishment of much longer-term social/community, infrastructure, economic, and environmental recovery management programs is required to be activated parallel with initial and ongoing response activities. Recovery efforts that wait for the official response phase to be formally completed will not be able to maintain or make the most of the focus and momentum established during response (Ministry of CDEM 2004). Recovery management, as with effective response management, can only be efficiently achieved when sufficient planning, legal mandate, resourcing, and professional development have been put in place and maintained prior to any significant emergency occurring.

Public Information Management

A key component of managed responses is public information management, the process of engaging and communicating with those affected by or who have an interest in the disaster and its aftermath, through the media, directly, and, increasingly, through social media.

Whether response organizations engage with, manage, monitor, or support the media and, increasingly, social media, both traditional and social media will have a significant interest in the emergency itself, consequences, pre-organized and emergent community-based, and official responses. This will be all the more so when there has been appreciable damage, injuries, or deaths.

Ensuring that the affected community, those supporting those affected, the wider area, the nation, and, in significant enough emergencies, other parts of the world are aware of what has occurred, what those affected should do, what others should do, and what official organizations are doing fall under the heading of “public information management,” “public affairs,” or a similar title within emergency responses.

Public information management in major emergencies, just like overall responses themselves, is most effectively managed in a multiagency context, in a Joint Information Centre, under the coordination of the leading government agency. This allows for a consistent and well-directed messaging to be delivered by a composite team of communications officials and the wider range of organizations involved in the response.

Traditional print media, particularly newspapers and electronic media, including social media, are relatively vulnerable to the impacts of earthquake, either directly when local production and distribution or transmission infrastructure are compromised or when electricity or access to or within the affected communities is interrupted. Purpose-designed flyers produced by the response organizations and delivered directly to affected homes, communities, evacuation centers, and receiving communities may be essential during response, particularly where media that communities usually depend on may be disrupted by the event and its consequences.

Ensuring that appropriate people are adequately trained and available to act as spokespeople for various organizations involved in response is essential. Some spokespersons will be the response leaders within the organizations involved themselves, such as CEOs [or city managers] of affected and responding local

authorities and infrastructure providers, as well as response controllers and senior emergency service managers actually managing the response. Politicians have an important role to play in response, with mayors usually taking spokesperson role at a local level and responsible cabinet members doing so at state or national level, and prime ministers, governors, or presidents doing so in more challenging emergencies that warrant higher-level political engagement.

One of the risks of attempting to assimilate community, agency, media, and social media information on behalf of the entire response and to produce the outputs to affected communities in a timely manner in particular is that the processes of doing will be slowed by the need to develop outputs in a collaborative manner. Bureaucratic development of outputs combined with compromised means of communication can result in critical information not reaching key audiences in a timely enough manner – something that needs to be kept in mind and mitigated prior to and during responses.

Engaging with affected communities is a crucial aspect of public information management, particularly where community responses are not well integrated into the official response and recovery activities. Including spontaneous and organized community response and preexisting and post-event community groups as partners to public information management is increasingly seen as critical to effective response and recovery efforts.

Significant events will not only receive widespread coverage outside the affected area but may result in hundreds and possibly thousands of reporters and production crews converging on the affected area and coordination centers and sites of particular interest. Dealing constructively with this number of additional personnel and the activities, some potentially risky or at least intrusive or distracting, has to be taken into account in planning for and delivery of response and recovery public information management.

Event Hazard (Seismicity) Monitoring

While the aftershock sequence cannot be prescribed or predicted following a major

earthquake, it is important that there be as good an understanding of the characteristics of the likely aftershock sequence as possible. The communication of this information in a timely way requires the specific integration of national seismological representatives within operational structures. Information about the possible magnitude, intensity, and location of aftershocks is of particular interest to USAR teams, those involved with rapid building assessments, and for public information purposes.

Through their training, the engineers involved both with USAR (Urban Search and Rescue) teams and the rapid building assessment operations have an expectation of major aftershocks of typically up to a magnitude less than the main shock to occur in the days, weeks, or months following, but possibly up to the same magnitude as the main shock.

For USAR teams, as the likelihood of rescuing live people from collapsed structures diminishes with time following the main shock, the “risk vs. reward” equation changes, and therefore a sense of the likelihood of significant aftershocks is a useful input.

Engineers and others undertaking rapid building assessments have a longer time frame to consider as they assess the suitability of damaged buildings for continued occupancy. The placards they post on buildings should take into account the additional damage that could occur from aftershocks in the days, weeks, and months following the earthquake and prior to detailed damage evaluations and repairs being undertaken.

Assessment of the short-, medium-, and long-term seismicity effects after a major earthquake is important as the response transitions to recovery. This may, for example, lead to an increase in the seismic design coefficients used for repairs and rebuilds due to the heightened seismicity, as happened following the Christchurch earthquake in February 2011.

The nature of the expected aftershocks can broadly be characterized from the fault activity which caused the earthquake itself. For example, earthquakes that occur on established and active faults close to plate boundaries typically generate few significant aftershocks, and these will

generally have the same directionality. In contrast, earthquakes that occur away from plate boundaries (intraplate earthquakes) on “new” faults that have not previously shown evidence of surface expression can generate aftershocks which migrate, with much less predictability about the number of significant aftershocks and their intensity and directionality.

The post-event seismicity is typically expressed in probabilistic terms, which can take the form of “the probability of an aftershock of magnitude range (e.g., M6.0–6.4) occurring in a given region within a specified period of time.” This form of estimation often uses different seismicity models, depending whether the time frame is short term (e.g., months or a year) or longer term (e.g., 20 or 50 years).

Assessment of secondary risks should also be undertaken and must be considered within the post-quake context. These include those that are consequential in addition to the primary seismic hazard, such as tsunami and landslide/rock-fall, liquefaction, natural and technology-related environmental contamination, as well as consequentially unstable slopes and land susceptible to subsidence and increased flood risk.

Emergency Information Management

Managing impact and response information is a particular challenge during damaging earthquakes, particularly if inadequate thought and preparation has gone into this function prior to the event or key components of information collection, analysis, sharing, management, storage, or dissemination are compromised by the event itself. The benefits of a well-developed incident management system and supporting information management systems are vital to effective quake response activities.

Contemporary information management relies on networks and equipment which when, unless designed, constructed, supported, and used appropriately can be particularly vulnerable to direct or indirect earthquake interruption.

Telecommunication providers are an essential element in the resilience-building work of such

critical infrastructure groups prior to and their efforts to restore services during emergencies and after emergencies.

Organizations with roles in response, including critical infrastructure providers, are better able to manage and share information during emergencies when they have invested in post-quake accessible and resilient data backup services, ruggedized or robust and seismically secured information management hardware, and alternate means of communications, such as voice and data over satellite or radio. Even resilient data storage, management and communication organizations with response roles also have to have paper-based systems to fall back to, at least temporarily, in the event of their first line of alternate arrangements being compromised or overwhelmed.

Effective response management depends on emergency-related information management applications and operational procedures that are compatible with partner organizations, fit for purpose, well trained, and exercised.

Cloud-based data storage and email systems are becoming increasingly popular in not only emergency response-focused information management but in business continuity and disaster recovery processes. Software and equipment that are intended for emergency purposes and personnel required to use or manage them are considerably better able to meet response and recovery needs if the response roles are integrated as much as possible into business-as-usual practices.

Impact Overview: Rapid Impact Assessment

One of the first tasks of affected communities and organizations, and the emergency response organizations supporting them, after a hazard such as an earthquake causes damage is to carry out a rapid impact assessment. These usually occur within 8–48 h on the incident occurring to:

- Assist in determining what ongoing or new risks may be present
- Planning initial response actions

- Acquiring, activating, and tasking necessary initial resources
- Provide the basis for more detailed needs and structural assessments

Rapid impact assessment is much less about building-by-building, household-by-household, and facility-by-facility assessment generally initiated several days into the response, but about quickly gaining a sufficiently detailed and reliable intelligence picture of the known and likely direct and indirect impacts of the hazard. The process of planning, tasking, collecting, analyzing, and disseminating rapid impact assessment data should follow the intelligence cycle process used in all aspects of emergency response to inform initial and later action planning processes.

Analysis of earthquake risk, including estimates of likely impact and, particularly in relation to critical infrastructure and vulnerabilities of particular and interrelated infrastructure, can greatly assist rapid impact assessment. Knowledge of quake likelihoods and consequences can change relatively rapidly as aspects of earthquake science and knowledge improve.

It is essential for those preplanning or actually managing rapid impact assessment to use quake risk studies to inform their work where any such studies are available. An example of such a study carried out recently is that for the Wellington region, New Zealand (Cousins 2013), which provides estimates of between \$NZ1 billion and \$NZ16.5 billion in damage in varying scenarios generated by the seven major fault lines known in the city and in the nation's capital, with a population of approximately 460,000 and 195,000 buildings where they are built on or adjacent to. The damage would include up to almost 7,000 building failures caused by shaking, geological failure, or tsunami, up to approximately 2,000 fatalities and 9,300 serious injuries from earthquake, and 3,500 fatalities and 7,000 injuries from tsunami. Pre-event impact estimations such as these are of considerable assistance in determining impact assessment, response, and recovery resources that are likely to be necessary when real disasters strike.

In the first few hours and days after a significant earthquake, telecommunications are likely to be compromised, and personnel available to carry out physical and social impact assessments will be in short supply. Therefore, for impact assessment to be conducted efficiently and effectively, it may be necessary for the coordinating emergency management agencies and emergency services to plan and prioritize the deployment of resources in pre- or response-determined sectors.

Impact data collection has traditionally been carried out in a paper-based environment, but electronic data collection, often with digital photographic records and GPS location data, is becoming increasingly affordably feasible – even in the immediate wake of earthquakes and similar hazards. Handheld devices, including smartphones, tablets, iPads, etc., and purpose-designed or business-as-usual applications are revolutionizing the collection, management, and application of data in virtually all emergency, incident, and disaster management contexts.

Although standard telecommunications services will often be compromised or at least overloaded in the first few days after an earthquake, data can be readily transferred to recipient databases by way of direct data download at the impact assessment base of operations or respective emergency operations center.

Rapid Physical Impact Assessment

In the emergency response coordination context, post-earthquake, rapid impact assessment tends to focus first on identifying sites and areas when search and rescue may need to occur – usually multistory buildings that have failed or been seriously compromised and buildings, roads, bridges, ports, airports, and other geological features that may be required for evacuation and emergency response access and egress.

Identification of actual and potential environmental contamination, from analysis of preexisting records of contamination risks, contamination identified or reported in the immediate response or in subsequent managed reconnaissance tasks.

Early assessment also includes determining whether buildings, other infrastructure, or geological features potentially damaged in early shocks may fail later, either of their own accord or due to later seismic activity.

Rapid impact assessment informs immediate planning and management of tactical functions such as search and rescue, evacuation orders and support, public information messaging, etc., as well as more planning for more strategic activities and resourcing.

Rapid physical and social impact assessment tasking, data collection, analysis, and dissemination are usually managed by the emergency management arm of the local or regional/county government involved, in conjunction with local emergency services, particularly fire and police resources working in or supporting immediate response in the affected areas.

Rapid Social Impact Assessment

The process of rapid social impact assessment is activated alongside rapid physical impact assessment – informally by communities and more formally by government and nongovernment relief, social service, and emergency management organizations. Social impact assessment and emergency social/welfare services are usually coordinated by the emergency management arm of the respective local or regional government, supported by state, national, and international entities.

Even in post-quake international disaster relief situations, the responsibility for authorizing the cross-border entry of search and rescue, impact assessment, and relief personnel and their deployment and activities within the affected areas lies within the government of the recipient country and, where delegated authority exists, within the government of the local jurisdiction involved.

Social impact assessment includes gathering data on known needs and making estimates of likely needs based on known data and knowledge of needs generated in similar events in comparable communities elsewhere.

Emergency social needs include the basic hierarchy of needs: safe air, water, and food, a

safe place to sleep, and adequate sanitation. Immediate needs then extend to social reconnection, particularly for family members separated as a consequence of the emergency and identification and support of the vulnerable, whether due to pre-emergency circumstances or as a result of the emergency.

Most social impact data is collected by personnel assigned that responsibility in teams tasked to carry out survey of individuals, families, or groups sheltering in place, seeking support at relief centers, or evacuees in transit or at reception centers or group lodging. However, rapid social need assessments can also be augmented by data collected by those carrying out rapid physical impact assessment, particularly if those gathering data on physical impacts have been prepared to observe and gather social impact information.

Residential impact assessment can often be most effectively carried out in conjunction with physical impact assessment, through the establishment of composite teams of building assessors and social services workers. In areas where there are larger buildings, impact assessors and engineers are likely to be deployed on behalf of building owners and insurance companies relatively early in the response. Opportunities to coordinate and share at least outline impact data between public and private assessors should be pursued. Private assessors should at least work under the safety and access requirements imposed by emergency management agencies.

Access to damaged areas is likely to be controlled by emergency services or emergency management agencies early in the response. If that is the case, impact assessors will need to be required to carry pre-authorized means of identification or identification and access documentation or printed cards developed for the particular response. Assessors will also be required to use personal protective equipment and regular reporting procedures as prescribed by the coordination entity or at standard occupational health and safety practice for the respective jurisdiction.

Impact and needs information is of little use if it is not made available to those delivering

immediate support services or planning for larger-scale longer-term search and rescue, safety, relief, and eventual recovery programs. It is essential that all impact assessment data captured during response are captured and collated at a local level and then made available to response and relief organizations in a timely manner. Some users of impact assessment data, such as rescue teams and risk managers, will need detailed reports and data feeds, while others, such as social service providers and governance bodies, will require more aggregated, area-wide data.

Impact assessment is most effective when it has been planned and resourced in sufficient detail prior to an emergency and has ideally been tested through exercising both data collection and analysis in particular.

Urban Search and Rescue

Urban search and rescue (USAR) involves the location, rescue, and initial medical stabilization of victims trapped in confined spaces following a structural collapse. USAR is an integrated multiagency response beyond the capability of normal rescue arrangements in most jurisdictions. In addition to rescue technicians, USAR teams include medical and canine search and rescue engineering capabilities.

The canine search component of a USAR team focuses on carrying out rapid searches over and within collapsed or seriously damaged structures. Being trained to operate remotely from their handlers, the search dogs are able to go into areas that may be considered too dangerous for rescue technicians. By alerting when they encounter trapped live people, they provide a focal point for search operations and the other equipment that they have, thereby assisting with prioritization and saving significant time in a search and rescue operation that covers a large area or numbers of buildings.

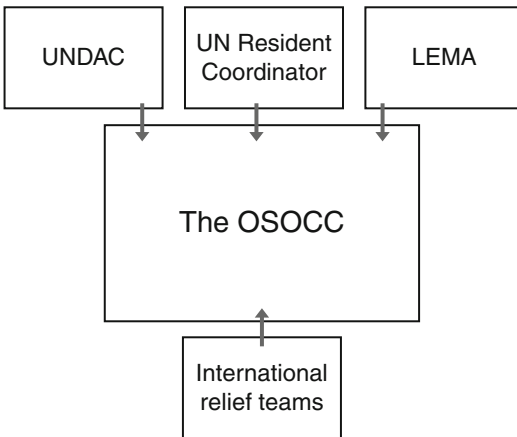
The engineering component of a USAR team assists with the identification, monitoring, and reduction of hazards (within site) and provides detailed/specific structural engineering advice where required. This includes advice on the least dangerous points of access to areas of

structural collapse where the most complex rescues may be required, having due regard to both the current stability of the structure and potential impacts from following aftershocks. Ideally USAR engineering teams should contain a geotechnical engineer, as the threats of further movement posed by unstable hillsides (landslip and rockfall) add considerable complexity to earthquake-affected areas.

The principal source of reference in relation to USAR practice and arrangements is the UN-based International Search and Rescue Advisory Group (INSARAG) www.insarag.org. The focus of INSARAG is the development of standardized guidelines and procedures and sharing of best practices among national and international USAR teams and defining standards for minimum requirements of international USAR teams. The INSARAG Guidelines provide frameworks for the response operation for the country affected by a sudden-onset disaster causing large-scale structural collapse, as well as for international USAR teams responding to the affected country. They also outline the role of the UN assisting affected countries in on-site coordination.

Irrespective of the scale of the USAR response to a significant earthquake event, the Local Emergency Management Authority (LEMA) is the ultimate responsible authority for the overall command, coordination, and management of the response operation. The LEMA can refer to local, regional, or national authorities or combinations thereof, which are collectively responsible for the disaster response operation.

In the rescue response following a major earthquake, the operations focal point is the On-Site Operations Coordination Centre (OSOCC). The OSOCC is established close to the LEMA and as close to the disaster site as is safely possible. It provides a platform for the coordination of international responders and LEMA. The OSOCC is established by the UNDAC team or by the first arriving international USAR team, who will then hand over the OSOCC to the UNDAC team when they arrive. An outline representation of the relationships between these aspects of response is provided as Fig. 2.



Emergency Response Coordination Within Earthquake Disasters, Fig. 2 On-Site Operations Coordination Centre (OSOCC) Relationships (Source: C3Fire.org)

The main purpose of the OSOCC is to assist the LEMA with the coordination of international and national USAR teams as well as establishing intercluster coordination mechanisms (e.g., health, water/sanitation, and shelter).

The United Nations Disaster Assessment and Coordination (UNDAC) Team is dispatched by the UN to sudden-onset emergencies when requested to do so by the affected Government or the UN Resident Coordinator in the affected country. UNDAC Team members are trained emergency managers from countries, international organizations, and UN OCHA. The UNDAC Team assists the LEMA with the coordination of international response including USAR, assessments of priority needs, and information management by establishing an OSOCC.

The activities and processes within each of the USAR phases of *Preparedness*, *Mobilisation*, *Operations*, *Demobilisation* and *Post-mission* are elaborated upon within the INSARAG Guidelines.

In addition to national and international USAR teams, many jurisdictions have local or regional response teams that are trained in a range of operational capabilities, including surface search and rescue. Their role is typically to assist the emergency services and the LEMA and to provide liaison with and support for USAR

teams. The input from response teams following a major earthquake varies with time following the event and ranges from assisting with initial surface rescues in the immediate aftermath prior to the arrival of national USAR teams to detailed searches of less damaged buildings to ensure no trapped persons are present through to providing safety advice and support for engineers undertaking rapid building assessments.

Medical Response

Most injuries as a result of earthquake are caused by falling furniture or internal or external fittings and architectural features. These can range from minor cuts and sprains to more serious lacerations, fractures, crush injuries, and, in some instances, burns. A smaller proportion of injuries are caused by individual falls caused directly by shaking. The most serious injuries and fatalities tend to be caused by failure of parts of structures that casualties were either in or adjacent to at the time that the quake or significant aftershocks occurred.

Mass injuries and fatalities, and infrastructure and environmental damage for that matter, tend to also occur in the more rare instances when tsunami accompany or follow a serious earthquake, as the tsunami surge reaches shore either near to or distant from the epicenter of the quake. As demonstrated in the recent earthquakes and tsunami in Japan and in 2004 in the Indian Ocean, earthquakes and tsunami can present substantial initial and consequential fatalities, injuries, infrastructure, and ongoing environmental contamination that may continue to have effects for a very long time.

Earthquakes bring multiple challenges to those tasked with providing medical support to those injured in the quake and its aftermath. The most challenging situations are those where casualties are entrapped within damaged building or within debris from damaged buildings or failed built or geological structures. Access to and treatment, or extraction and then treatment, of casualties may be achieved in the early stages of response by others on the scene at the time or

spontaneous formal and informal responders converging to sites of structural collapse.

Significant medical support to entrapped, ambulatory, and evacuated casualties will be required following earthquake and may extend as far as limb amputation to enable timely extraction and specialist support to sufferers of crush syndrome or occasionally prolonged dehydration following protracted entrapment.

Earthquakes will often have serious implications for health services and supporting infrastructure in the vicinity, as utilities services and access for ambulances are compromised; community-based health services, general practices, and hospitals, particularly multi-floor facilities are damaged; the homes, places of learning or care, and work places of health workers and their families are also damaged or isolated; and health workers or family or friends become casualties themselves. Even in earthquake-resilient health facilities, damage will be experienced, not the least to supporting infrastructure and often older neighboring facilities that are less resilient.

Following significant earthquakes and similarly damaging emergencies, the scale of injured who self-report or are transported to existing and possibly damaged on-site or makeshift triage and treatment centers setup on-site will be beyond normal mass-casualty planning and experience. Innovative approaches will be required to meet needs generated. Preplanning, training, and exercising for earthquake-scale medical responses will substantially enhance site response, field triage, medical center and hospital resilience and response, and local, regional, and higher-level health service emergency coordination.

Engineering support is critical to health-care providers, initially in terms of damage assessment and operational advice, in the immediate aftermath of earthquakes, as decisions are made on the least risk-prone facilities in which to conduct emergency medical care and whether evacuations are required from possibly damaged facilities.

Hospital emergency responses, ideally planned for and resourced well prior to them

being required, will include not only triaging and treating the injured but also dealing with damage to their own facilities and impacts of staff capabilities, coordination of service delivery and patient care with partner health and social service providers, decanting of less-urgent patients from the service, and transfer of higher acuity patients to unaffected or less affected facilities elsewhere.

Retirement homes and associated hospital and support services are a growing part of health services in many communities, due to aging populations and decreasing family size. The concentration of vulnerable people in elderly care facilities or receiving support in the community becomes a priority for local health services, both in the area affected and, potentially, recipient communities in the event of a significant self- or managed evacuations occurring.

Emergency Welfare

Direct and indirect impacts on humans are the main focus of response and recovery activities following earthquake. Emergency welfare or social services are the formal and informal processes put in place to assist those affected to meet their basic needs following emergencies. These needs are created when people no longer have access to the essentials of life, including accommodation or shelter; personal safety and security; food, water, medication, sanitation, and hygiene; care for dependents and vulnerable people; clothing; means of income; and psychosocial support.

The majority of emergency welfare services tend to be provided informally by the impacted communities themselves, along with their preexisting social networks; extended families; neighborhoods; faith, ethnic, and community service groups; business associations; and spontaneous volunteer groups formed in the wake of the emergency. These efforts are augmented and, where possible, supported by more formal service delivery arrangements.

Initial emergency welfare response tends to focus on making emergency shelter, food,

water, and relatively hygienic toileting services available to affected resident, transient, or visitor communities. These services are provided in community emergency centers, evacuations centers, or other facilities established to centralize services and physical resources in relatively safe and secure locations or dispersed within the affected communities if residents are sheltering in place but with access to normal services compromised by the effects of the earthquake.

Public facilities, such as community centers, sports facilities, halls, or schools, will often be used in formal and informal welfare support – assuming these facilities remain relatively intact. People who have evacuated from their homes, workplaces, or visitor accommodation will tend to congregate open spaces such as parks and playing fields, particularly where built facilities are unavailable or insufficient to meet local needs or community members are reluctant to enter buildings following earthquake. Temporary tent communities will often be established by people who are reluctant or unable to remain in their homes but who want to remain in their communities.

Emergency Food and Water Supplies

The images of food products broken and strewn across supermarket floors, collapsed shelving in warehouses, damaged roads and bridges, burst water pipes, and lines of vehicles at petrol stations are common scenes in news coverage after earthquakes. The realities of this for the affected communities run much deeper.

Earthquakes cause damage and contamination to water reticulation systems and to food supplies in people's homes, retail and wholesale food facilities, as well as food processing plants. Transport routes, including road, rail, water-based, and airport infrastructure into the affected areas, will be damaged to varying degrees, reducing supplies for at least days and often weeks or months.

Regional, state, or national food supply wholesalers are better able to reestablish supply chains more readily than emergency management

agencies sometimes give them credit, with alternative food supply logistics arrangements seldom being required to be established as part of official responses. It is essential, however, that food suppliers are included in emergency preplanning, relationship-building, and exercises to ensure that their needs and capabilities are realistically taken into account of and included in responses.

Distribution of food within affected communities tends to be more difficult to reestablish as retail premises and transport routes within the quake zone may well be out of commission for at least several days. Community engagement, preparedness, or resilience-building programs that encourage community members, groups, and businesses to have food and water supplies for several days will enable them to support themselves more readily, while formal and spontaneous arrangements are put in place.

Post-earthquake Building Management

In broad terms, a post-earthquake building management program is comprised of three phases:

- Overall damage survey
- Rapid building assessments
- Detailed damage evaluations

The *overall damage survey* is typically carried out as part of the immediate response by emergency services personnel and local authority staff, as referred to under *Rapid Physical Impact Assessment* above.

Rapid building assessments are generally undertaken by a number of small teams comprising volunteering engineers and other building professionals working with local authority building officials.

The objectives of a rapid building assessment process are to:

- Confirm where damage to buildings is concentrated to assist response and recovery decision-making
- Indicate whether physical action is to be taken to enable, restrict, or prevent access to individual buildings

- Commence the process of systematically gathering data on damaged buildings that will facilitate the planning and monitoring of longer-term recovery actions and reoccupation

The outputs from a rapid building assessment process are posted placards on the buildings and completed forms summarizing the damage observed and placards posted that are returned for entry into the database of the respective local authority or other responsible agency.

The focus of the rapid building safety evaluation process is on *immediate public safety*, not the provision of an engineering assessment service to building owners. It is an initial triaging process, akin to initial medical assessments at the scene of a large emergency or in a hospital emergency department. Buildings that have sustained visible and significant structural damage are marked via a placarding operation as not being suitable for reoccupancy. Other buildings that do not show signs of visible damage or movement may be suitable for occupancy, but they require further attention subsequently from building owners and their engineers when resources become available.

Rapid building assessments are just that – undertaken quickly in the face of potentially large numbers of buildings to be appraised within an affected area. Rapid assessments of only the exterior are expected to take of the order of 10–20 min, while rapid assessments involving exterior and interior inspections can take anywhere from 1 to 4 h, depending on the size of the building. These assessments are almost always undertaken without recourse to structural drawings.

The coordination of a rapid building assessment operation requires an operational structure that links between the local jurisdictions' building control unit and the emergency management leadership arrangements. As any such operation requires considerable follow-up activity by the affected council(s), it is important that the operation is integrated as much as possible with building control officials, particularly the leadership of the operation and management of data. Hence, a

considerable amount of planning and preparation is required prior to an event.

The following key roles are suggested for inclusion within the rapid building assessment operational arrangements and leadership structure:

- Building assessment manager
- Technical and process adviser
- Central business district/commercial technical lead
- Residential technical lead
- Engineering resource coordinators
- Information management coordinator
- Regulatory process support (transition management)

The scale of the event (and hence rapid building assessment operation) will dictate whether all of these roles are required to be filled.

Detailed damage evaluations are generally carried out by structural and geotechnical engineers specifically engaged by building owners. They involve accessing all available information, detailed interior inspections, and performing specific calculations where required. They can take anywhere from one day to a week or more, depending on the size of the building and the type of damage. Detailed damage evaluations for critical facilities, particularly hospitals, should however be undertaken in the immediate aftermath of an earthquake, and this requires appropriate priority response agreements to be established with engineers familiar with the facilities prior to an earthquake.

Lifeline Utility Coordination

The utility services sector is invariably complex in a day-to-day context, much less in a post-earthquake situation. Within the key sectors including water, energy, telecommunications, and transport, there are typically a range of providers.

The coordination of the information and actions across the utility sector following a major earthquake is accordingly complex and requires specific planning. The specific

coordination arrangements will vary depending on the legislative structures and operating arrangements in place. The key common feature is for such arrangements to enable a strong linkage between the utility sector (collectively) and the emergency management sector. This linkage needs to be active in nonemergency times, given the emphasis that should be given to risk reduction and preparedness.

The objectives of post-earthquake utility coordination are to:

- Coordinate the actions of lifeline utilities and emergency/disaster management organizations to provide a safe and effective response
- Support the restoration of lifeline services as quickly as possible

Utility coordination functions include:

- Managing information flows
- Monitoring impacts
- Communicating lifeline utility status information within the respective coordination center (s)
- Facilitating any support required by utilities for emergency/disaster management or vice versa

All countries have emergency management legislation and related regulations and plans which, to varying degrees of detail and compulsion, define key infrastructure providers as “lifeline utilities,” “critical Infrastructure,” or similar and require organizations falling under this definition to be able to continue to operate during and after an emergency. This has been spelled out more clearly in the New Zealand context than some other, where a wide range of life utilities are required to have planned and resourced their organizations, facilities, and services to be able to continue to operate to the fullest extent possible during and after emergencies.

Lifeline utility coordination is a function defined usually carried out in response coordination centers under the respective incident management system’s *operations* function. Lifeline

utility coordination is most likely to be activated at both regional and national levels, rather than local levels.

The Lifeline Utility Coordinator (LUC) tends to be a non-statutory regional and national emergency management position responsible for coordinating lifeline utilities on behalf of the response controller or director during emergency response and the recovery manager throughout recovery activities.

The roles are assigned before an emergency occurs and activated if required during an emergency.

Summary

In summary, responses to significant earthquakes are challenging and complex. This entry has explored the key aspects of earthquake response that are particularly pertinent to the context of earthquake engineering. Although responses to virtually all emergencies and disaster start and finish with the affected communities, formal leadership and coordination of responses are vitally important for effective response to an earthquake event. When significant earthquakes strike, the challenges for those in positions of leadership or who are responsible for response coordination are many and varied. This is due in large part to the reality that affected communities will be mounting their own responses while official responses are still regrouping and reestablishing communications. The very nature of damaging quakes makes them some of the most difficult hazards to manage responses to, as the sudden and widespread impacts significantly affect the ability of response organizations to deliver their services more than other hazards tend to.

Cross-References

- [Resilience to Earthquake Disasters](#)
- [Resiliency of Water, Wastewater, and Inundation Protection Systems](#)

- ▶ [Resourcing Issues Following Earthquake Disaster](#)
- ▶ [Social Media Benefits and Risks in Earthquake Events](#)

References

- C3Fire.org. Bridging cultural barriers to collaborative decision making in on site operations coordination centers. <http://www.c3fire.org/c3fire/examples/examples-bcb.en.shtml>. Accessed 16 Apr 2015
- Cousins WJ (2013) Earthquake damage and casualties due to large earthquakes impacting the Wellington Region. GNS science report 2013/41 July 2013
- Ministry of Civil Defence and Emergency Management (2004) Recovery management directors guideline DGL 04/05. Wellington. <http://www.civildefence.govt.nz/assets/Uploads/publications/dgl-04-05-recovery-management.pdf>
- Ministry of Civil Defence and Emergency Management (2014) Coordinated incident management system. Lifeline Utilities and Civil Defence Emergency Management Groups. Director's Guideline DGL16/14. Wellington. New Zealand. <http://www.civildefence.govt.nz/resources/new-zealand-coordinated-incident-management-system-cims-2nd-edition/>
- Von Lubitz DKJE, Beakley JE, Patricelli F (2008) 'All hazards approach' to disaster management: the role of information and knowledge management, Boyd's OODA Loop, and network-centricity. *Disasters* 32(4):561–585

Further Reading

- Applied Technology Council (1989) ATC-20 procedures for post-earthquake safety evaluation of buildings. Applied Technology Council (and related documents and updates)
- Ardagh M (2013) The initial health-system response to the Christchurch Earthquake. Research into Health Impacts of Seismic Events (RHISE) symposium, Nov 2013
- Ministry of Civil Defence and Emergency Management (2013) Rapid impact assessment. Information for the CDEM sector. IS 14/13. Wellington. [http://www.civildefence.govt.nz/memwebsite.nsf/Files/Rapid-Impact-Assessment-December-2013/\\$file/Rapid-impact-assessment-IS-Dec-2103.pdf](http://www.civildefence.govt.nz/memwebsite.nsf/Files/Rapid-Impact-Assessment-December-2013/$file/Rapid-impact-assessment-IS-Dec-2103.pdf)
- New Zealand Society for Earthquake Engineering (2011) Building safety evaluation following the Canterbury earthquakes report for the Canterbury earthquakes royal commission of inquiry
- Von Lubitz DKJE (2008) Disaster management: the structure, function, and significance of network-centric operations. *J Homel Secur Emerg Manag* 5(1): ISSN (Online) 1547–7355, doi:10.2202/1547-7355.1411, August 2008 International Search and Rescue Advisory Group, 2012 guidelines and methodology. <http://www.insarag.org/en/methodology/guidelines>

Empirical Fragility

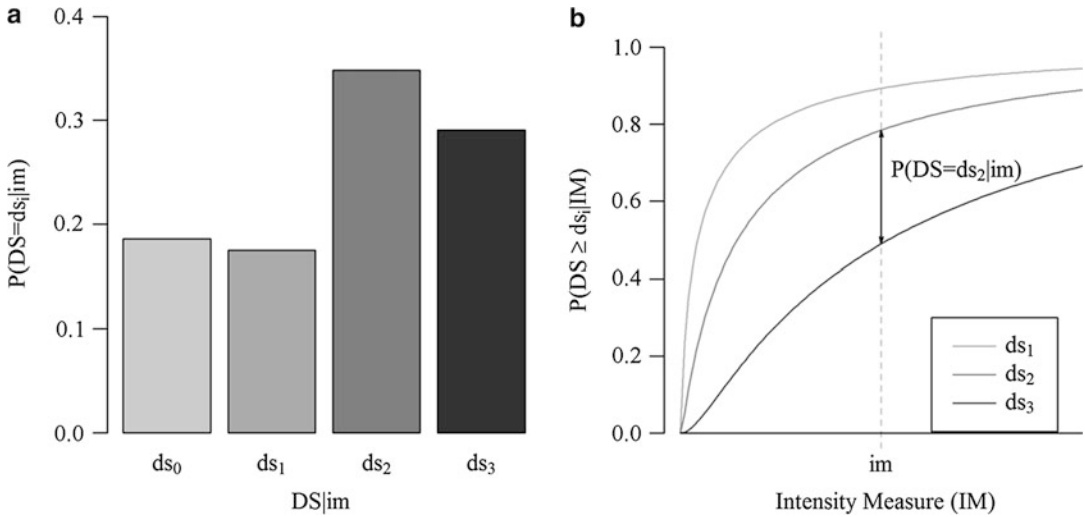
Ioanna Ioannou and Tiziana Rossetto
Department of Civil, Environmental and
Geomatic Engineering, University College
London, London, UK

Synonyms

Damage probability matrices; Empirical fragility assessment; Empirical seismic fragility; Fragility curves

Introduction

The seismic fragility of assets, i.e., buildings and lifelines, is defined as the likelihood of damage to the assets from future earthquake events. Fragility is an important component of seismic risk assessment and is key to the determination of the seismic vulnerability of the exposed assets (i.e., likelihood of economic loss, life loss, or downtime in future earthquake events). Relationships for defining fragility are typically classified into the categories of empirical, analytical, judgment-based, or hybrid according to the type of data used in their determination. Empirical fragility relationships are based on the statistical analysis of post-earthquake observations of the seismic damage sustained by exposed assets. A number of sources for post-earthquake asset damage or loss data are available (e.g., surveys commissioned by authorities in order to assess structural safety or evaluate cost of repair for insurance, tax reductions, or government aid distribution purposes) and the reliability of empirical relationships is heavily dependent on the quality and size of the adopted observational databases. Analytical and judgment-based fragility relationships are based on the interpretation of the results of numerical analyses and expert opinion of the behavior of assets in earthquakes, respectively. Hybrid fragility relationships combine two or more of these data sources in their definition.



Empirical Fragility, Fig. 1 Illustration of (a) a column of a DPM for intensity im , (b) fragility curves corresponding to $n = 3$ damage states for the same building class

Empirical fragility is typically expressed either in terms of damage probability matrices (DPMs) that express the probability of a class of assets with similar attributes suffering given damage states under specified ground motion or a continuous relationship between a measure of damage (damage state) and ground motion or ground failure intensity (see Fig. 1) for a specified asset class. The continuous functions are overwhelmingly expressed in terms of fragility curves. These curves are obtained by fitting a statistical model to the post-earthquake damage observations. The two expressions of empirical fragility are related as depicted in Fig. 1b, where it can be noted that the distances between two successive fragility curves for a given intensity measure level relate to the corresponding values in the damage probability matrix.

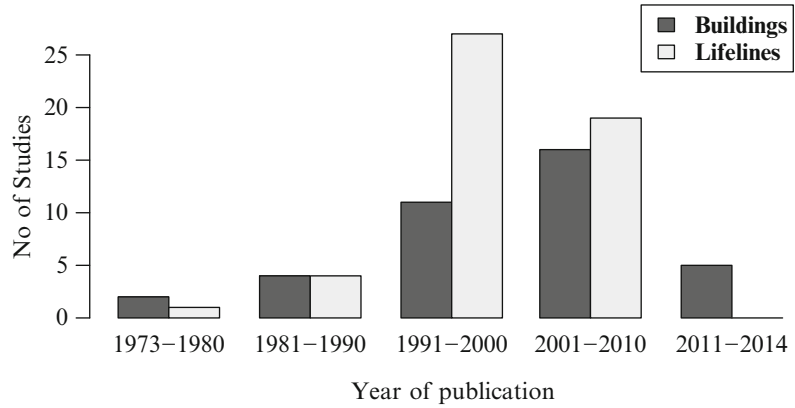
Empirical fragility assessment was firstly introduced by Whitman et al. (1973), who constructed DPMs for tall buildings by relating an 8-state discrete damage scale to loss data collected from 1,240 questionnaires in the aftermath of the 1971 San Fernando earthquake, USA. Since then, approximately 89 studies are seen in the published literature (i.e., academic journals, conference proceedings, reports, and book chapters) to have focused on the development of empirical fragility functions for the global building and

lifeline inventory. Many of these studies include fragility functions for multiple classes of assets.

Figure 2 shows a steady increase in the number of studies on empirical fragility of buildings over the past 40 years. Recently, the Global Earthquake Project (GEM) (Rossetto et al. 2013) compiled a database of empirical fragility functions for the global building inventory. GEM's database consists of 175 empirical fragility functions constructed for the global building inventory. Figure 3 shows that almost half of existing functions have been constructed for the European building stock. Most empirical functions are based on post-earthquake observations collected from a single earthquake event (e.g., Braga et al. 1982). Few exceptions exist of functions based on databases from multiple global (e.g., Spence et al. 1991; Rossetto and Elnashai 2003) or national (e.g., Rota et al. 2008) events. Recently, GEM (Rossetto et al. 2014) built on the state-of-the-art research and proposed new guidelines for the construction of empirical fragility curves from single or multiple earthquake event databases. Within these guidelines procedures are proposed for the assessment of the quality of the post-earthquake observation databases and for fragility relationships to be derived using statistical approaches of increasing complexity according to the characteristics of the adopted observational databases.

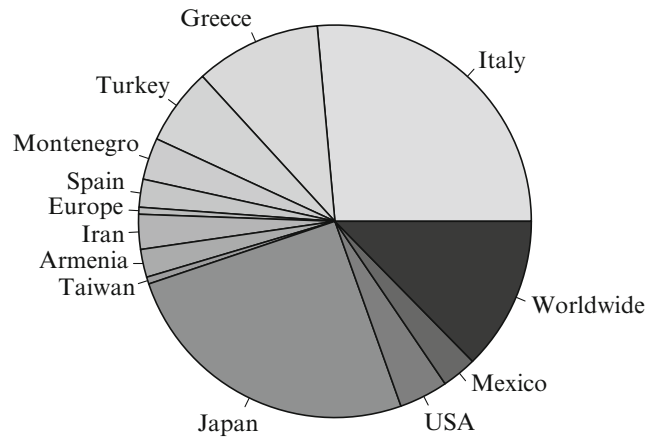
Empirical Fragility,

Fig. 2 Studies focused on empirical fragility assessment of buildings and lifeline networks published in the last four decades (Data sourced from the compendium of empirical fragility functions of buildings in GEM (Rossetto et al. 2013) and Syner-G (Pitilakis et al. 2014))



Empirical Fragility,

Fig. 3 Country of origin of existing fragility functions for buildings



Empirical fragility research for lifelines peaked in the 1990s. In the context of the empirical fragility literature, fragility curves and DPMs for lifelines have mainly been derived for use in transportation (e.g., railway, roadway, or harbors) and utility (e.g., gas and oil, potable water and wastewater, or electric power) networks. Existing studies identify the main components of each network (see Fig. 4) and assess their empirical fragility from post-earthquake data using similar procedures to those used in the empirical fragility assessment of buildings. The fragility of each component is then used in order to estimate the overall fragility of the network (see ► [Seismic Vulnerability Assessment: Lifelines](#)). Recently, Syner-G (2014) compiled a global database of fragility functions for the components of the global lifeline inventory presented in Fig. 4.

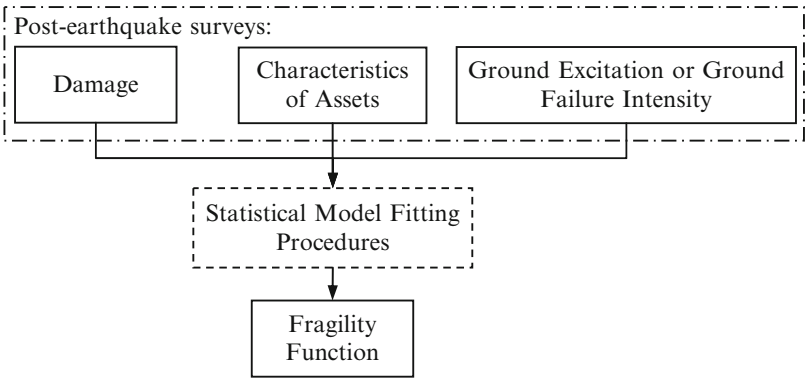
Empirical Fragility Assessment Framework

The general framework for constructing fragility functions for a class of assets is presented in Fig. 5. According to this framework, the construction of empirical fragility functions requires the collection of three types of post-earthquake data: data on the level of seismic damage suffered by the assets, the characteristics of these assets, and the level of ground motion or ground failure intensity the assets were exposed to. A statistical model is commonly fitted to the data to generate the mean fragility function and its confidence bounds. The following sections provide a brief overview of each component of the empirical fragility framework.

Elements of Transportation Networks: • Railway Tracks (1) • Road Pavement (1) • Bridges (5) • Tunnels (2) • Embankments (1) • Quay walls (1) • Cargo handling and storage components (1) • Fuel facilities (1) • Waterfront structures (1)	Elements of Utility Networks: • Water Source (1) • Potable-Water Treatment Plant (2) • Waste-Water Treatment Plant (2) • Pumping Plant (2) • Canals (1) • Pipes (20) • Storage Tanks (4) • Electric power substations (3) • Electric micro-components (7) • Electric power grid (1)
---	--

Empirical Fragility, Fig. 4 Elements of transportation and utility networks for which there are empirical fragility functions and the number of the studies (in brackets) that provide these functions (This figure is based on information from Syner-G (2014))

Empirical Fragility, Fig. 5 Empirical fragility assessment framework



Post-Earthquake Surveys

The reliability of an empirical fragility function is strongly dependent on the quantity and quality of post-earthquake damage observations used for its construction. Ideally post-earthquake damage assessment data should be available for all the assets within the desired asset class that were exposed to the earthquake event. If observations on all the assets are not available, at least a statistically significant representative sample of the exposed assets should be used. In practice, databases of post-earthquake damage observations rarely include the total number of exposed assets and may be very small in size or subject to bias. In general, the quality and quantity of these databases depend on why they were carried out, how, and by whom. The four main types of surveys that have to date provided post-earthquake damage observations for empirical fragility studies are described

in Table 1. Such surveys are typically commissioned by authorities or conducted by reconnaissance organizations and are briefly discussed here. Authorities often commission a post-earthquake “rapid” survey, which aims to assess the safety of assets. These surveys commonly adopt damage scales consisting of three limit states (such as in ATC-20 2005 in the USA). However, some may only adopt two broad damage states (i.e., damaged and undamaged, or safe and unsafe), grouping together a wide range of seismic damage levels. Such surveys aim to assess all exposed assets in an affected location and commonly do not capture many details of the characteristics of the assets (i.e., may only record the construction material of a building but not its height and age). Due to the rapid assessment often from the street by small groups of perhaps inexperienced engineers, these databases are also

Empirical Fragility, Table 1 Main survey types and their main properties

Survey method	Typical sample sizes	Typical building classes	Typical no. of damage states	Reliability of observations	Typical issues
Rapid surveys	Large	All buildings	2–3	Low	Safety not damage evaluations. Misclassification errors
Detailed “engineering” surveys	Large to small	Detailed Classes	4–6	High	Possibility of under-coverage errors
Surveys by reconnaissance teams	Very small	Detailed Classes	4–6	High	Possibility under-coverage errors or very small samples
Remotely sensed	Very large	All buildings	2–3	Low	Only collapsed or very heavy damaged states may be reliable. Misclassification errors

commonly found to suffer from two main errors, namely, misclassification and nonresponse errors. The former type of errors is introduced when the safety level of a building is different from the actual damage suffered by the assets. For example, a building sustained substantial damage and should be classified as “unsafe” but the survey form indicates it is “safe.” This error can potentially have a significant impact on fragility functions (see Ioannou and Rossetto 2013). Nonresponse errors are instead introduced when the forms used for collecting the survey data are incomplete. For example, the location or address of surveyed assets is missing from the form. The mechanism of the missing data determines how this error is treated. For example, due to the speed of “rapid;” surveys, some boxes in the survey forms are left blank completely at random. In this case, the forms with incomplete data can be removed from the database. This leads to, sometimes substantially, smaller sample sizes. By contrast, the missingness mechanism could depend on the missing observations. For example, the safety assessment on the form is left blank, if the building is classified as safe. In this case, the knowledge of this mechanism is essential in order to effectively reduce the bias introduced by this nonresponse error.

“Rapid” surveys can be followed by detailed “engineering” surveys. These surveys collect detailed data concerning the characteristics of assets and damage levels. Although the

“engineering” surveys cover large number of assets, they tend to concentrate on the damaged assets, ignoring the undamaged ones. This practice introduces an error, termed under-coverage, in the database. This error introduces a bias in the database, which can be reduced by the use of data from alternative sources when the source of under-coverage error is known. For example, if it is known that only damaged buildings have been surveyed, it may be reasonable to estimate the number of undamaged buildings from census data.

Reconnaissance surveys are conducted by small teams of highly experienced engineers. These surveys are therefore of high quality but often suffer from small sample sizes, given the limited time the reconnaissance teams spend in the field and the small areas surveyed.

Finally, surveys based on remote sensing techniques have been recently introduced for the empirical fragility assessment of buildings. Of note is the recent establishment of the Virtual Disaster Viewer (Bevington et al. 2010), supported by several international institutions that provide images and a forum for experts to interpret damage from satellite images. The subsequent GEO-CAN initiative (Ghosh et al. 2011) also brought together experts from all over the world to interpret damage in the 2010 Haiti and 2011 Christchurch, New Zealand earthquakes. These surveys have the ability to gather data for large sample sizes of the assets of interest. However, as they are based on observations of

changes to building roofs and plans from comparison of pre- and post-earthquake aerial images, they are unable to distinguish between building classes and are only able to assess the collapse (and to lesser extent heavy damage) state in buildings. Consequently, the inaccuracies involved in the assessment of the damage lead to substantial misclassification errors for the non-collapse damage states (Booth et al. 2011).

Damage Assessment

The seismic damage suffered by assets is typically expressed in terms of a damage scale with n number of discrete qualitative states. The damage states can be defined in terms of structural and nonstructural damage levels. They vary from no damage to collapse and are often considered to be mutually exclusive and collectively exhaustive. The scales, adopted in the literature, vary in the number of damage states and the level of detail in the description of damage for each damage state according to the needs of the survey.

Damage scales with $n = 4\text{--}6$ states are mostly adopted for the classification of damage observed in buildings and lifeline components. Some of the most commonly used damage scales for post-earthquake damage surveys for buildings are the European Macroseismic Scale (Grünthal 1998) and its predecessor the Medvedev-Sponheuer-Karnik scale (Medvedev 1977). For lifelines, HAZUS proposes qualitative damage scales for most components, which are observed to be similar to those for buildings. Syner-G (2014), instead, recently proposed damage scales expressed in terms of operational states, which describe whether or not the component functions. Despite the presence of well-established damage scales, many studies adopt national damage assessment scales (e.g., AeDES in Italy) or bespoke scales. This wealth of damage scales is problematic when databases or fragility functions need to be combined or compared. This problem has been recently addressed by Rossetto (2014) who, building on the work of Hill and Rossetto (2008), has drawn equivalence tables between the various damage scales found in the literature.

In the construction of empirical fragility functions, ideally a damage scale should contain 3 to

6 damage states. Fragility functions are constructed for each damage state, and thus there needs to be a significant number of observations for each damage state in order to ensure the reliability of the corresponding empirical function. In practice, a balance needs to be reached between the number of damage states and the number of observations. In other words, for the same sample size, the fewer the damage states, the more data in each, but the wider the damage states, the less informative the fragility functions in the context of loss estimation. Furthermore, due to the rarity of large earthquake events near urban areas, higher damage states (e.g., collapse) tend to contain few observations, and hence fragility functions constructed for these states may be less reliable than the functions corresponding to less severe damage states.

A notable exception to the above discrete definition concerns the pipes, whose levels of damage are measured in terms of the number of repairs per mile.

Characteristics of Assets

Assets affected by similar intensity measure levels often perform very differently. These differences can be, at least partially, attributed to their unique structural characteristics. Empirical fragility assessment requires the availability of statistically significant sample sizes in order to construct reliable fragility functions. Thus, classes of assets are often determined based on the characteristics that significantly influence their seismic performance. In theory, the more detailed the class, the more homogenous the group of assets and the smaller the variation in seismic response. However, in practice narrowly defined classes often result in small sample sizes. A careful balance is, therefore, needed between level of detail in the class definition and sample size.

For example, the structural characteristics of buildings, which influence their seismic performance, are well documented (e.g., FEMA - 547 2007) and include (but are not limited to):

- Construction material
- Lateral load resisting system (e.g., moment-resisting frame, shear wall, etc.)

- Layout in plan and elevation
- Presence of irregularities (e.g., soft storey, short column, etc.)
- Seismic design code used (i.e., capacity design, level of structural detailing)
- Height of structure
- Weight of structure
- Redundancy of structure
- Type and amount of strengthening intervention

Existing studies overwhelmingly classify buildings according to their structural characteristics. The crudest classes are based on the construction material. The most detailed classes include the construction material, height, structural system, and design code. Most empirical functions, however, are based on classes that use only two structural characteristics, namely, the construction material as well as (in descending order of frequency) age, structural system, or height. Despite the existence of various building classification systems (e.g., adopted by HAZUS (FEMA 1999) or PAGER (Porter et al. 2008)), these are not commonly adhered to by existing studies; rather, bespoke building classes are defined.

Similar rules apply for the classification of the components of the lifeline networks. For instance, bridges are mostly classified according to their construction material and their structural system.

Intensity of Earthquake-Induced Hazards

Assets located in seismic areas are exposed to two main classes of earthquake-induced hazards, namely, ground shaking and ground failure (i.e., liquefaction, landslides and surface fault rupture, tsunami, and fire). Empirical fragility of buildings is typically expressed in terms of ground shaking, which is found to be the main cause of damage to buildings (Bird and Bommer 2004). By contrast, both hazards are found to cause damage to the lifelines, which are distributed over extensive areas.

With regard to ground shaking, three main groups of intensity measures have been adopted in past empirical fragility studies and a brief

introduction to these is presented in what follows (a comprehensive review of these measures can be found in Rossetto et al 2013). In the early empirical fragility studies for buildings, popular measures of ground motion intensity were macroseismic intensities, such as the Modified Mercalli Intensity scale or the Medvedev-Sponheuer-Karnik scale. Macroseismic intensities are discrete scales and a value of macroseismic intensity is assigned to an area based on the observed effects of the earthquake on the area (e.g., the level of observed damage to buildings and experiences of people affected by the event). However, the criteria defining the earthquake effects are often worded ambiguously, and their interpretation is subjective leading to potentially large discrepancies in their evaluation by different groups. A continuous alternative to the discrete macroseismic measures is the parameterless scale of seismic intensity (PSI), proposed by Spence et al. (1991).

A different approach for the selection of ground motion intensity measures has been promoted by the current seismic risk assessment framework, where it is important to decouple the uncertainties associated with the seismic demand from those associated with the structural fragility. Recent empirical fragility studies therefore adopt ground motion parameters that can be evaluated from ground motion recordings, e.g., peak or spectral values of ground motion. However, due to the scarcity of strong ground motion recording stations and the need to quantify the ground motion parameters at each asset site, in practice ground motion parameter values are obtained from ground motion prediction equations. Due to the large number of ground motion prediction equations available and due to its traditional use in the definition of design loads for structures and seismic hazard maps, peak ground acceleration (PGA) is the main parameter used to represent ground motion intensity in empirical fragility studies of buildings and elements of lifeline networks. However, it is well known that PGA does not correlate well with observed damage (especially for ductile structures or severe damage states). In engineering terms, PGA is related

to the dynamic loads imposed on very stiff structures and, therefore, may not be adequate for the characterization of seismic demand in medium or high rise building populations. Peak ground velocity (PGV) and peak ground displacement (PGD) are more representative of the seismic demand on structures of intermediate and long natural period of vibration than PGA, respectively. Both PGV and PGD are generally calculated through the integration of accelerograms and may be sensitive to both the record noise content and the filtering process.

The peak ordinates of a ground motion record, however, are unable to capture features of the earthquake record, such as frequency content, duration, or number of cycles, which affect the response of a structure and its consequent damage and loss. Hence, in an attempt to better capture the influence of accelerogram frequency content, more recent empirical fragility studies have favored the use of response spectrum-based parameters as measures of ground motion intensity. These parameters include the spectral acceleration for the fundamental period and less often the spectral displacement.

Ground failure (particularly liquefaction, lateral spreading, and fault rupture) is expressed in terms of permanent ground deformation, which can be measured laterally, vertically (i.e., settlement), or volumetrically. Vertical settlement can be measured in terms of total settlement or differential settlement, which has greater implications for damage to structures and lifelines. There are currently no standardized parameters for describing permanent ground deformation, though Bird et al. (2006) summarize some more common parameters, which vary in terms of input requirements and methodological approach. There have been few empirical damage observation studies related to ground failure incidents and therefore fewer studies on fragility estimation to these hazards (e.g., Tokimatsu et al. 1994).

Statistical Model Fitting Procedures

Having collected post-earthquake data, the empirical fragility assessment concludes with

Empirical Fragility, Table 2 Example of a damage probability matrix for a theoretical construction class

Bins of intensity (e.g., PGA in g)	0.02–0.04		0.04–0.06	
	Counts	DPM (%)	Counts	DPM (%)
No damage	100	50	60	30
Slight	80	40	90	45
Moderate	18	9	45	23
Collapse	2	1	5	3
<i>Total</i>	<i>200</i>	<i>100</i>	<i>200</i>	<i>100</i>

the construction of fragility functions for classes of assets. Empirical fragility functions are essentially stochastic relationships, which relate the damage to an asset class with the corresponding ground motion or ground failure intensity. These relationships can be discrete (i.e., DPMs) or continuous (e.g., fragility curves).

Damage Probability Matrices

In its simplest form, empirical fragility is expressed in terms of a discrete function. Its construction requires the aggregation of the damage data for assets of a given class into bins of similar ground motion or ground failure intensity levels. The counts of buildings suffering given levels of damage for an examined bin, normalized by the total number of buildings in this bin, provide the values of the damage probability matrix (see Table 2).

Continuous Fragility Functions

Fragility functions are typically constructed by fitting a statistical model to the available post-earthquake data. Practically, a statistical model consists of two components: a random and a systematic component. The random component expresses the probability distribution of the response variable (e.g., the counts of buildings reaching or exceeding a damage state) given the explanatory variable (i.e., the ground motion or ground failure intensity). The systematic component expresses the mean response as a function of the explanatory variable, in other words the mean empirical fragility function.

There are two main types of continuous fragility functions, described below, depending on whether the examined assets can be considered point-like or line-like. The difference between the two types is that point-like assets occupy a relatively small area where the intensity is assumed constant for each asset. Point-like assets include standard buildings and all components of lifeline networks presented in Fig. 4, with the exception of pipes which are considered line-like assets.

Fragility Functions for Point-Like Assets

Fragility functions for point-like assets are expressed in terms of fragility curves. In the literature, fragility curves are typically constructed by fitting parametric statistical models to the available data. A parametric model represents a family of probability distributions of the response conditioned on the explanatory variable defined by a small number of parameters. The most commonly used approach, which is described below, is to fit a generalized linear model to the data (e.g., Shinozuka et al. 2000). These models are based on strong assumptions, which, when they are satisfied, lead to reliable predictions. Nonparametric models have also been adopted for the construction of empirical fragility curves, but they suffer from the danger of over-fitting, especially if the data are aggregated in only small number of bins of ground motion or ground failure intensities. These models include the generalized additive models (see Rossetto et al. 2014) and the Gaussian Kernel Smoothers (see Noh et al. 2013).

The generalized linear model is based on the underlying assumption that the seismic response of each asset is a Bernoulli trial, i.e., the response is equal to 1 if the sustained damage is $DS \geq ds_i$ and 0 otherwise. Given that the damage data typically are aggregated in bins of similar ground motion or ground failure intensity, the response is considered to follow a binomial distribution for intensity level, im :

$$N_i | im \sim \binom{n}{n_i} \mu_i^{n_i} [1 - \mu_i]^{n-n_i} \quad (1)$$

where N_i are the counts of assets that have been damaged to a level equal or exceeding a predetermined damage state, $DS \geq ds_i$; n is the total number of assets in a bin associated with intensity level im . This distribution is essentially the fragility curve and is characterized by the mean response for given intensity, μ_i , as:

$$g(\mu_i) = \begin{cases} \theta_{1i} \ln(im) + \theta_{0i} & (.1) \\ \theta_{1i} im + \theta_{0i} & (.2) \end{cases} \quad (2)$$

where $g(\cdot)$ is termed the link function and relates the mean response with the intensity or the natural logarithm of the intensity. The link function can be expressed in three forms:

$$g(\mu_i) = \begin{cases} \Phi^{-1}(\mu_i) & \text{probit} & (.1) \\ \log\left(\frac{\mu_i}{1 - \mu_i}\right) & \text{logit} & (.2) \\ \log[\log(1 - \mu_i)] & \text{complementary loglog} & (.3) \end{cases} \quad (3)$$

The first two functions presented in Eq. 3 are symmetrical functions, and the last one is asymmetrical.

The parameters of the generalized linear models can be estimated by two procedures. In the first procedure, the parameters are estimated by maximizing the likelihood function through an iterative least squares algorithm, as:

$$\begin{aligned} \theta_i^{opt} &= \text{maxarg}[L(\theta_i)] \\ &= \text{maxarg} \left[\prod_{j=1}^{m_{bins}} \binom{n_j}{n_{ij}} \mu_{ij}^{n_{ij}} [1 - \mu_{ij}]^{n_j - n_{ij}} \right] \end{aligned} \quad (4)$$

In the second approach, a Bayesian regression analysis (e.g., Straub and Der Kiureghian 2008; Ioannou and Rossetto 2013) is adopted in order to take into account prior information regarding the model's parameters, θ , especially when the available number of observations is small. Prior information is obtained from existing fragility functions or independent post-earthquake data of similar groups of assets. In addition, this

analysis allows for the increase in the model's complexity by explicitly accounting for the measurement error in the intensity measure levels (e.g., Straub and Der Kiureghian 2008) or the misclassification of the damage (e.g., Ioannou and Rossetto 2013). The aim of the Bayesian analysis is the estimation of the posterior distribution of the model parameters, $f(\theta|y, x)$, from Bayes' theorem:

$$f(\theta|y, x) = \frac{L(\theta; y, x)f(\theta)}{f(y)} = \frac{L(\theta; y, x)f(\theta)}{\int L(\theta; y, x)f(\theta)d\theta} \quad (5)$$

where $f(\theta)$ is the prior distribution of the regression parameters accounting for existing prior knowledge.

Fragility Functions for Line-Like Assets The continuous fragility functions, constructed for groups of line-like assets, measure the damage in terms of the number of repairs per unit length (termed repair rate) of these assets for given levels of intensity. The repair rate (RR) is related to the intensity through the use of a linear (i.e., the parameters of the model are linearly combined) or nonlinear (i.e., the parameters of the model are nonlinearly combined) statistical model. These models can be expressed in different forms:

$$\begin{aligned} &RR|im \sim Normal(\mu(im), \sigma), \text{ where } \mu(im) = a im + b \\ &\text{or} \\ &\log_{10}(RR|im) \sim Normal(\mu(im), \sigma), \text{ where } \mu(im) = a im + b \\ &\text{or} \\ &\ln(RR|im) \sim Normal(\mu(im), \sigma), \text{ where } \mu(im) = \ln(a) + b \ln(im) \end{aligned} \quad (6)$$

where σ is the standard deviation, which is constant for each intensity level; $\mu(im)$ is the mean repair rate or the lognormal repair rate or the natural logarithm of this rate, which is expressed as a function of the intensity. The model is fit to the available data using a least squares method.

The number of breaks or leaks for the total length (L) of the line-like asset is considered to follow a Poisson distribution. Therefore, if we assume that the examined asset fails if there is at least one break, then the probability of failure for the total length of the asset is obtained, as:

$$\begin{aligned} &P(\text{line-like asset segment fails}) \\ &= 1 - P(\text{No failure}) = 1 - e^{-RR \cdot L} \end{aligned} \quad (7)$$

Summary

The empirical seismic fragility of an asset (i.e., buildings or components of lifeline networks) expresses the potential for its damage in future earthquake events. Empirical fragility is assessed by the statistical analysis of post-earthquake

observations on damage to assets, their characteristics, and the ground motion or ground failure they are subjected to. This typically includes the fitting of a parametric statistical model to the observational data. The reliability of empirical fragility functions depends on two main conditions: firstly, on the quality of the post-earthquake observational data and whether the database includes statistically representative samples of the exposed assets and, secondly, on whether the assumptions on which the selected statistical model is based upon are satisfied. To date, studies struggle to satisfy both of these conditions.

Cross-References

- [Analytic Fragility and Limit States \[P\(EDPI IM\)\]: Nonlinear Dynamic Procedures](#)
- [Analytic Fragility and Limit States \[P\(EDPI IM\)\]: Nonlinear Static Procedures](#)
- [Seismic Risk Assessment, Cascading Effects](#)
- [Seismic Vulnerability Assessment: Lifelines](#)

References

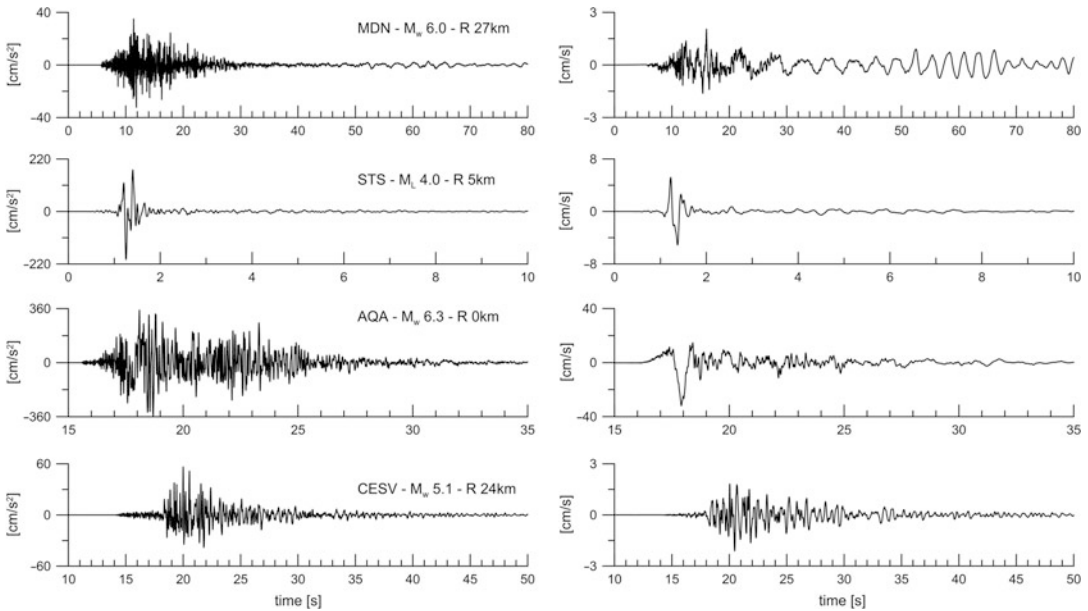
- ATC-20 (2005) Field manual: post-earthquake safety evaluation of buildings. Report. Applied, Technology Council, Oakland
- Bevington J, Adams B, Eguchi R (2010) GEO-CAN debuts to map Haiti damage. *Imaging Notes* 25:2 (online)
- Bird JF, Bommer JJ (2004) Earthquake losses due to ground failure. *Engineering Geology* 75:147–179, ISSN:0013-7952
- Bird J, Bommer J, Crowley H, Pinho R (2006) Modelling liquefaction-induced building damage in earthquake loss estimation. *Soil Dyn Earthq Eng* 26(1):15–30
- Booth E, Saito K, Spence R, Madabhushi G, Eguchi RT (2011) Validating assessments of seismic damage made from remote sensing. *Earthq Spectra* 27(S1): S157–S177
- Braga F, Dolce M, Liberatore D (1982) Southern Italy November 23, 1980 earthquake: a statistical study of damaged buildings and an ensuing review of the M.S. K.-76 scale. Report, Rome
- FEMA - 547 (2007) Techniques for the seismic rehabilitation of existing buildings. Report
- Ghosh S, Huyck CK, Greene M, Gill SP, Bevington J, Svekla W, DesRoches R, Eguchi RT (2011) Crowdsourcing for rapid damage assessment: the global earth observation catastrophe assessment network (GEO-CAN). *Earthq Spectra* 27(S1): S179–S198
- Grünthal G (1998) European macroseismic scale 1998. Report, vol 15. Cahiers du Centre Européen de Géodynamique et de Séismologie, Luxembourg
- HAZUS (1999) Earthquake loss estimation methodology, Technical manual SR2. Report, Federal Emergency Management Agency through agreements with National Institute of Buildings Science, Washington, DC
- Hill M, Rossetto T (2008) Comparison of building damage scales and damage descriptions for use in earthquake loss modelling in Europe. *Bull Earthq Eng* 6(2):335–365
- Ioannou I, Rossetto T (2013) Sensitivity of empirical fragility assessment of buildings to the misclassification error of damage. In: Proceedings of 11th international conference on structural safety & reliability, New York
- Medvedev SV (1977) Seismic intensity scale MSK-76. Report. Institute of Geophysics, Polish Academic Society, Warsaw
- Noh H-Y, Kiremidjian AS, Lallémand D (2013) Development of empirical fragility functions using Gaussian Kernel smoothing methods. In: Proceedings of 11th international conference on structural safety and reliability, New York
- Pitilakis K, Crowley H, Kaynia AM (2014) SYNER-G: typology definition and fragility functions for physical elements at seismic risk. Springer
- Porter K, Jaiswal K, Wald D, Earle PS, Hearn MG (2008) WHE-PAGER project: a new initiative in estimating global building inventory and its seismic vulnerability. In: Proceedings of 14th world conference on earthquake engineering, Beijing
- Rossetto T (2014) The conversion of damage and loss in GEM-VEM indirect vulnerability curve derivation. Report. GEM Foundation, Pavia
- Rossetto T, Elnashai A (2003) Derivation of vulnerability functions for European-type RC structures based on observational data. *Eng Struct* 25(10):1241–1263
- Rossetto T, Ioannou I, Grant DN (2013) Existing empirical fragility and vulnerability relationships: compendium and guide for selection. Report. GEM Foundation, Pavia
- Rossetto T, Ioannou I, Grant DN (2014) Guidelines for empirical vulnerability assessment. Report, GEM Foundation, Pavia, Italy
- Rota M, Penna A, Strobbia CL (2008) Processing Italian damage data to derive typological fragility curves. *Soil Dyn Earthq Eng* 28(10–11):933–947
- Shinozuka M, Feng MQ, Lee J, Naganuma T (2000) Statistical analysis of fragility curves. *J Eng Mech* 126(12):1459–1467
- Spence RJS, Coburn AW, Sakai S, Pomonis A (1991) A parameterless scale of seismic intensity for use in seismic risk analysis and vulnerability assessment. In: Proceedings of earthquake, blast and impact: measurement and effects of vibration. Elsevier, Manchester
- Straub D, Kiureghian DA (2008) Improved seismic fragility modeling from empirical data. *Struct Saf* 30(4):320–336
- Tokimatsu K, Kojima H, Kuwayama S, Abe A, Midorikawa S (1994) Liquefaction-induced damage to buildings in 1990 Luzon earthquake. *J Geotech Eng* 120(2):290–307
- Whitman RV, Reed UW, Hong ST (1973) Earthquake damage probability matrices. In: Proceedings of 5th world conference on earthquake engineering, Rome

Engineering Characterization of Earthquake Ground Motions

Francesca Pacor and Lucia Luzi
Istituto Nazionale di Geofisica e Vulcanologia,
Milan, Italy

Synonyms

Acceleration time series; Processing procedure;
Strong ground motion parameters; Strong-motion
database



Engineering Characterization of Earthquake Ground Motions, Fig. 1 Example of acceleration (*left*) and velocity waveforms (*right*): from *top to bottom* M_w 6.0 2012-05-29 07:00:03 UTC, Italy, recorded at MDN (deep alluvium); M_L 4.0 2003-01-2620-01-16 UTC, Italy,

recorded at STS(rock); M_w 6.3 2009-04-06 01:32:40 UTC, Italy, recorded at AQA (stiff soil); M_w 5.1 1998-04-0307:26:36 UTC, Italy, recorded at CESV (alluvium). (Records are obtained from the Italian strong motion database (<http://itaca.mi.ingv.it>). R is the epicentral distance)

Introduction

A knowledge of the seismic ground motion is essential to understanding the earthquake behavior of structures. The basic data of earthquake engineering are the records of ground acceleration during moderate-to-strong earthquakes (Housner 1970). Acceleration time series (also known as accelerograms and strong-motion records) are the acceleration of the ground sampled many dozens of times per second in three mutually orthogonal directions. For decades, the reference ground motion for earthquake engineers was the renowned El Centro accelerogram, recorded in California in 1940. It was only after the Loma Prieta (1989), Northridge (1994), and Kobe (1995) earthquakes that hundreds of strong-motion records were made available worldwide through databases accessible from the Internet (Pacor et al. 2011). The huge increase of accelerograms highlights the significant variability of ground motion, depending on several factors related to size and focal mechanism of the

event, distance and position of the recording site with respect to the source, properties of the propagation medium, and geological condition underlying the recording station.

In Fig. 1 examples of accelerometric and velocimetric time series recorded close to the source at sites located on rock and stiff soil (STS and AQA stations) and intermediate distances on soft sites (MDN and CESV) are plotted. Each time series is characterized by peculiar features, such as (i) presence of surface waves (MDN); (ii) impulsive behavior (STS); (iii) low-frequency velocity pulse (AQA); and (iv) absence of low-frequency waves (CESV).

The ability to characterize strong ground motion and capture the main features that can potentially cause damage to engineering and geotechnical structures is one of the main inputs in development methods to mitigate seismic risk.

In this chapter, several topics related to accelerometric data are illustrated, including (i) issues related to instrumental characteristics; (ii) evaluation of data quality; (iii) processing

procedures to compute acceleration, velocity, and displacement time series for use for seismic hazard studies and engineering applications; and (iv) definition of the main intensity measures computed from the accelerometric time series.

Analog and Digital Recording Instruments

Accelerometric signals are recorded by two types of instruments (Trifunac and Todorovska 2001):

- Analog accelerographs are optical-mechanical devices that produce traces of the ground acceleration on film or paper. These were in use since the early 1930s to the end of the twentieth century, although the 2004 Parkfield earthquake was recorded principally on such instruments. They have several disadvantages: (i) they operate on standby and recording is triggered by a specific acceleration threshold, so they do not preserve the pre- and post-event time series; (ii) the natural frequency of transducers is generally limited to about 25 Hz or even less, limiting the usable frequency band; and (iii) if numerical analyses are planned, in order to be able to use the recording, it is necessary to process the film and then digitize the traces.
- Digital accelerographs are in general use since the 1980s and record ground motion in continuous mode in digital media (e.g., cassettes or solid-state memory), and, even when they operate on standby, they preserve the pre- and post-event portions of the time series. The precision of the recorded ground motion depends on the instrument settings, such as the digitizer's dynamic range, sampling rate and sensor full scale. The analog-to-digital conversion (ADC) process can produce offsets in the acceleration baseline if the ground motion is slowly varying compared with the quantization level of the digitization (Boore 2003).

Figure 2 shows an example of the same event recorded by an analog and a digital instrument

that were colocated. In this particular example, the analog instrument was triggered by the S-waves.

Strong-Motion Data Processing

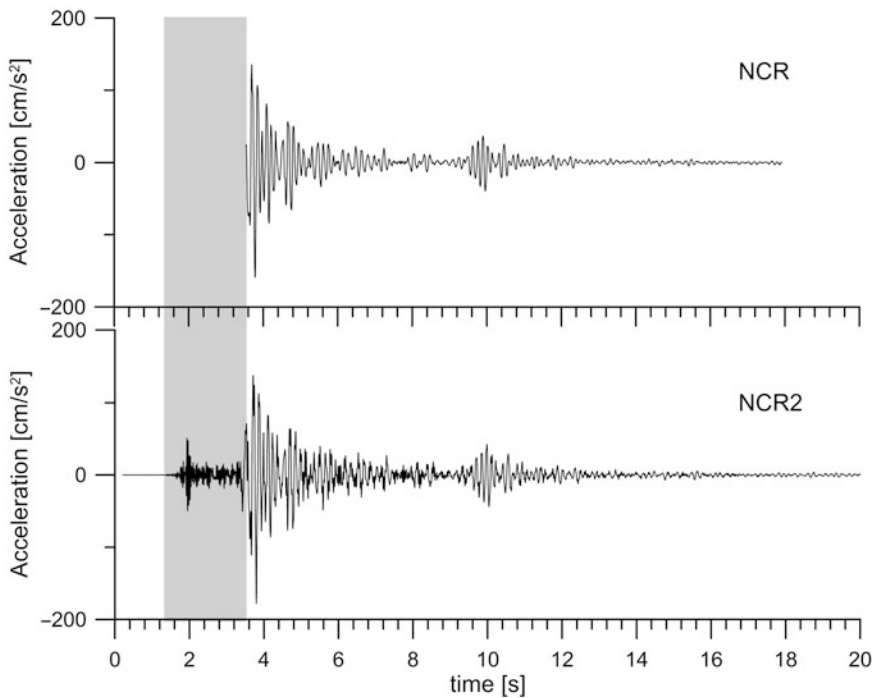
It is important for users of strong-motion data to understand that digitized accelerograms are not exact reproductions of the ground motion, as they contain noise of various origins at both high and low frequencies. The purpose of the signal processing is the estimation and the removal of the noise present in the record, in order to evaluate reliable ground motion for standard and specific applications. Generally it is not possible to identify the “ideal” processing for an individual record, as assumptions always need to be made and a certain degree of subjectivity is involved. Boore and Bommer (2005) provide an extensive overview of techniques for processing strong-motion data and the consequences of applying different approaches, from the perspective of engineering applications. Special cases, such as the processing of near-source strong-motion recordings are discussed by Graves (2004).

Data Quality and Data Processing

The first step before data processing is a quality check of the records. The user can decide, depending on the specific application, if the record should be disregarded or not before further processing.

Douglas (2003) reports the features that identify strong-motion recordings as poor quality (see Table I of his paper), based on an analysis of the records of the European Strong-Motion Database (Ambraseys et al. 2004). Based on these results, the most relevant features include (i) insufficient digitizer resolution; (ii) S-wave trigger, which can be a concern for recordings from analog instruments (Fig. 2); (iii) insufficient sampling rate; (iv) early termination during coda; (v) spikes (Fig. 3); and (vi) presence of multiple baselines.

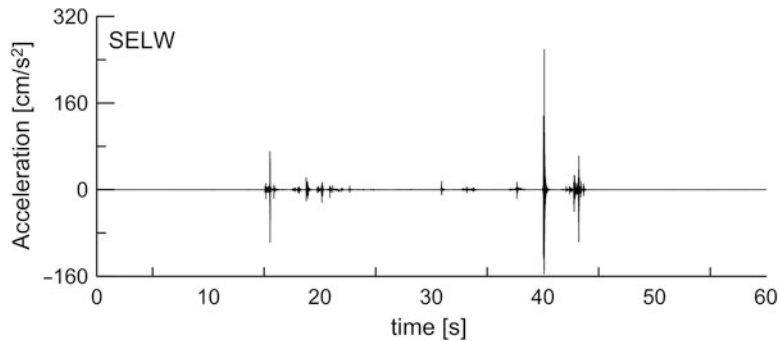
Some of these characteristics, so-called non-standard errors, do not correspond to the usual



Engineering Characterization of Earthquake Ground Motions, Fig. 2 Aftershock of the Umbria-Marche seismic sequence, central Italy (event 1997-10-0705:09:56

UTC); acceleration time series at the analog station NCR (*top*); acceleration time series at the digital station NCR2 (*bottom*) (Source <http://itaca.mi.ingv.it>)

Engineering Characterization of Earthquake Ground Motions, Fig. 3 Example of spurious spikes (raw record of event 1998-01-1301:49:16 UTC at SELW) (Source <http://itaca.mi.ingv.it>)



sources of noise and they can often be visually identified and removed prior to further processing.

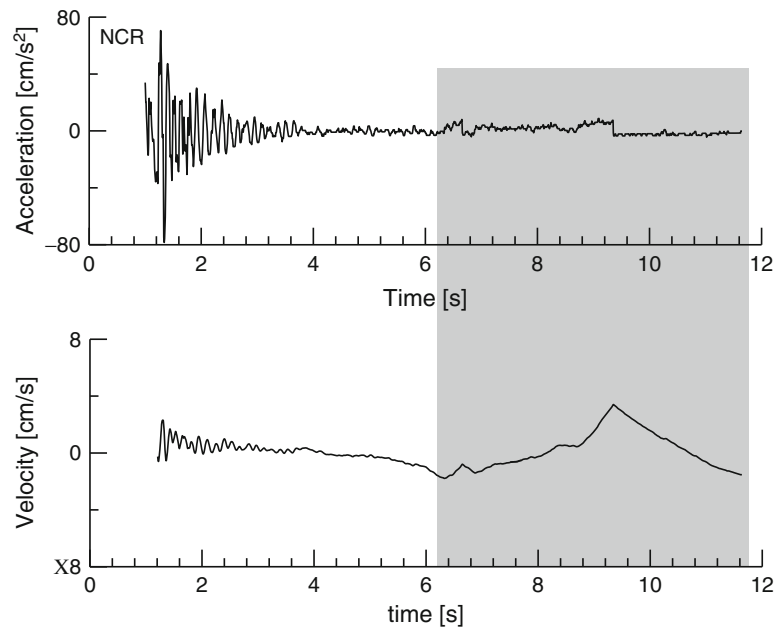
As an example, spurious “spikes” in the digitized record can be replaced with the mean of the accelerations of the data points on both sides of the spike. Another problem encountered with both analog and digital accelerograms are distortions and shifts in the baseline (Fig. 4), which result in unphysical velocities and displacements (Boore 2003). Boore and Bommer (2005) detail

an approach to compensate for these problems consisting in the use of baseline adjustments (Graizer 1979; Iwan et al. 1985; Boore 2001; Boore et al. 2002; Paolucci et al. 2008). In particular, the use of baseline-fitting techniques can sometimes recover permanent ground displacements from accelerograms.

Filtering

The process of filtering generally involves the removal of most of the signal at frequencies

Engineering Characterization of Earthquake Ground Motions, Fig. 4 Example of baseline shift and its effect on the velocity time series (event 1979-05-21 12:36:40 UTC recorded at NCR) (Source <http://itaca.mi.ingv.it>)



where the Fourier amplitude spectrum (FAS) shows a low signal-to-noise ratio. Therefore, the critical issue for end users is to appreciate the record limitations after certain frequencies have been removed.

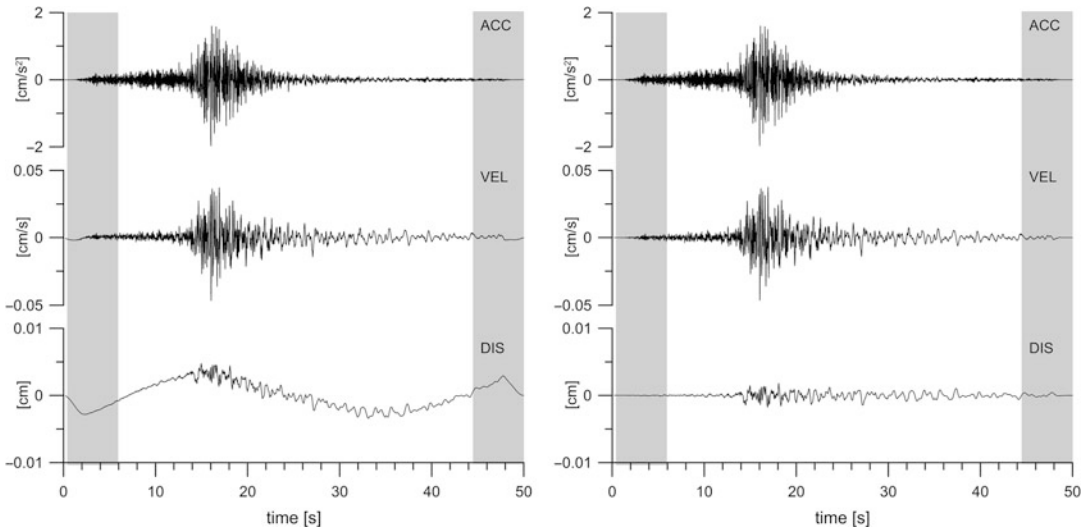
Filters can be applied in the frequency or time domains; exactly the same results should be obtained in both cases. While the choice of filter type is not crucial, the way in which the filter is applied to the accelerograms is very important. The fundamental choice is between causal and acausal filtering, where the latter does not produce any phase distortion in the signal, whereas causal filters do result in phase shifts in the record. The influence of causal and acausal filters on both elastic and inelastic response spectra has been investigated by Boore and Akkar (2003), who found that both elastic and inelastic response spectra computed from causally filtered accelerations can be sensitive to the choice of filter corner periods even for oscillator periods much shorter than the filter's corner periods. Finally, Bazzurro et al. (2005) performed a statistical study on the influence of various signal-processing techniques used for low-cut filtering ground motion records on both elastic and inelastic response spectra.

Low Frequency Noise Removal

The impact of low-frequency noise (<1 Hz) can have strong influence on strong-motion intensity measures such as ground velocities, displacements, and response spectral ordinates, and therefore most of the studies have investigated the effects of low-cut (high-pass) filters.

The low-frequency cutoff, the lowest frequency for which the data are reliable in terms of signal-to-noise ratio, can be determined in several ways: (i) comparison of the FAS of the record with that of a model of the noise obtained from the pre-event memory for digital records, for example; (ii) judgment of the frequency where the record FAS deviates from the tendency to decay in proportion to the reciprocal of the frequency squared, based on a seismological model of the radiated energy (omega-square model, Brune 1970); or (iii) visual inspection of the velocity and displacement time series obtained by double integration of the filtered acceleration.

Figure 5 shows the effect of low-cut filtering. A digital record is filtered using two different low-cut frequencies. The lowest frequency generates a low signal-to-noise ratio and leads to unrealistic velocity and displacement time series (gray boxes highlight the differences).



Engineering Characterization of Earthquake Ground Motions, Fig. 5 Example of low-cut filtering (event 2010-01-1213:35:45UTC recorded at AQUA). *Left:*

low-cut equal to 0.01 Hz; *right:* low-cut = 0.2 Hz (ACC, acceleration; VEL, velocity; DIS, displacement)

Acausal low-cut filters are generally preferred (Boore 2005) and, in order to achieve the zero phase shift, they need to start prior to the beginning of the record and finish after the end of the accelerogram, which can be accomplished by adding points of zero amplitude, known as pads, to the two ends of the record. The length of the pads depends on the filter frequency and its order (Boore 2005).

After filtering, the processed accelerograms are integrated once to obtain velocities and again to obtain displacement time series. They can also be used to obtain other intensity measures (see below).

High-Cut Filters

Ground motion at periods of less than roughly 0.05 s is generally only relevant to particular engineering problems, such as the design and analysis of nonstructural elements, equipment, and pipework (e.g., in nuclear power plants).

Before applying high-cut (low-pass) filters one should consider that the sampling rate determines the upper usable frequency, termed the Nyquist frequency, equal to $(1/2\Delta t)$ where Δt is the sampling interval. High-cut filter applied at

frequencies greater than the Nyquist will have no effect on the record.

Douglas and Boore (2011) give an exhaustive overview of the approaches and effects of high-cut filtering. While low-cut filters have large influence on ground motion intensity measures, high-cut filtering is sometimes even not necessary and accelerometric response spectra can often be used up to frequencies much larger than the applied high-cut corner used to remove the noise. As in the case of low-cut filters, acausal high-cut filters are generally preferred.

Applicability of Processed Data

The user should be aware that different corrected versions of the same record may be distributed in different strong-motion databases. Foti and Paolucci (2012) studied the influence of correction procedures by performing numerical simulations (seismic behavior of a diaphragm wall). The results show large differences, especially in terms of displacements, which is a parameter of particular relevance when dealing with performance based design.

Sometimes database providers distribute processed accelerograms that are incompatible

with the provided velocity and displacement time series because they apply additional processing to reduce noise that is still present in the record. In other cases, time series are not compatible because the pads added for the application of the filter are removed. The consequence of their removal is to weaken the effect of the filter and this can result in offsets and trends in the base-lines of the velocity and displacements obtained by integration.

In order to avoid the latter effects the correct initial conditions in the pad-stripped time series have to be used when computing displacements, velocities, and linear oscillator response. Another way of ensuring compatibility is to post-process the pad-stripped acceleration time series.

Boore et al. (2012) examined the post-processing adopted for two databases, ITACA (Italian Accelerometric Archive, Paolucci et al. 2011; Pacor et al. 2011) and PEER NGA (Pacific Earthquake Engineering Research Center–Next Generation Attenuation), and demonstrated that any biases and distortions are small for the vast majority of the records of the two databases, so data can be used with confidence.

Strong-Motion Intensity Measures

For many engineering and scientific applications, it is useful to describe the seismic ground motion in terms of a few simple parameters instead of the entire waveform. These parameters are evaluated both in the time and frequency domains, each of them capturing some features of the recorded accelerograms.

Time-Domain Parameters

The most common parameter is the *peak ground acceleration*, PGA, which denotes the maximum ground acceleration (absolute value) observed on the accelerometric time series, usually evaluated for each component of the ground motion. It can be evaluated directly either from raw or corrected signals, since PGAs change slightly with respect to the processing procedure.

Similarly, the *peak ground velocity*, PGV, and *peak ground displacement*, PGD, are defined.

Differently from PGA, PGV and PGD can be read only after processing the accelerometric record, and their values are strictly related to the band-pass filtering, especially in case of PGD, which strongly depends on the low-cut corner frequency.

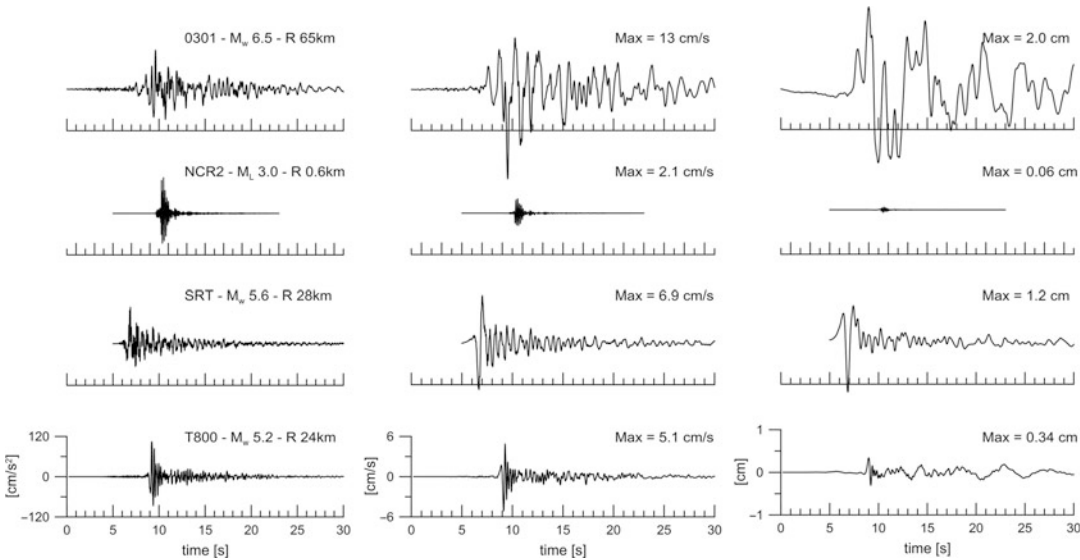
Although the PGA is widely used in earthquake engineering, as it is directly related to the inertia force, it poorly characterizes the ground motion, as demonstrated in Fig. 6. Records with the same PGA can have completely different features in the other ground motion parameters, depending on the magnitude of the event, the source-to-site distance and the geological condition of the site where the recording station is located.

A preliminary estimation of the frequency content of the records may be inferred by the PGA/PGV ratio (Zhu et al. 1988; Sawada et al. 1992; Kwon and Elnashai 2006). Low PGA/PGV ratios generally indicate earthquakes with low predominant frequencies, broader response spectra, and longer duration.

The *duration* of the accelerogram is a parameter used to identify the portion of record in which ground motion amplitude can potentially cause damage to engineering and geotechnical structures.

Several definitions are proposed (Bommer and Martinez-Pereira 1999), as many different meanings can be attached to “duration.” The most commonly used are as follows:

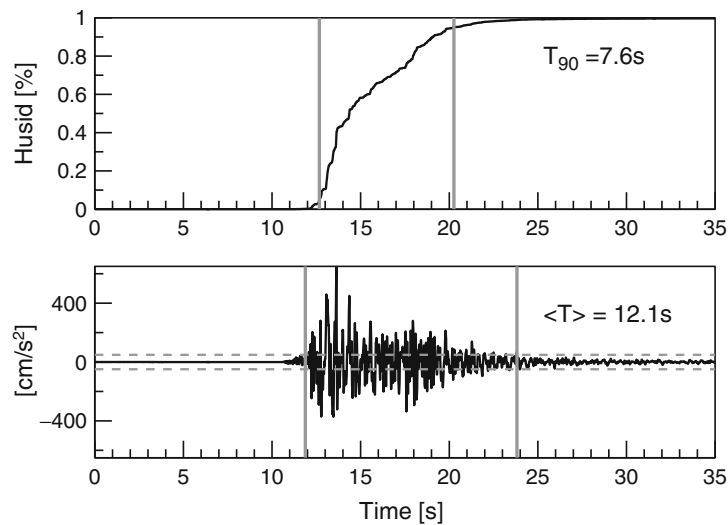
- (a) *Bracketed duration*, introduced by Bolt (1973): a threshold is fixed, typically 0.05 g or 0.1 g, above which it is deemed that the motion has relevance for engineering purposes; the duration is the time interval between the first and the last exceedance of this value.
- (b) *Significant duration*, introduced by Trifunac and Brady (1975) on the basis of the study by Husid (1967), is defined as the time interval over which the integral of the square of the ground acceleration (Husid plot) is within a given range of its total value. Usually this range is between 5 % and 95 % or between 5 % and 75 %. An example of the two



Engineering Characterization of Earthquake Ground Motions, Fig. 6 Left: examples of accelerometric waveforms having PGAs in the range 105–112 cm/s², event 2012-05-29 11:00:25 UTC, Italy, recorded at T0800 (deep alluvium); event 1990-12-13 00:24:26 UTC, Italy, recorded at SRT (rock); event 1997-11-12 15:24:23 UTC, Italy, recorded at NCR2 (shallow alluvium); event 2002-

02-03 07:11:29 UTC, Turkey, recorded at 0301 (alluvium). Middle and right: velocity and displacement waveforms derived from the accelerometric signals in the left panel. For each waveforms, the peak values (Max) is reported (Source <http://itaca.mi.ingv.it> and Turkish National Strong Motion Program http://kyhdata.deprem.gov.tr/2K/kyhdata_v4.php)

Engineering Characterization of Earthquake Ground Motions, Fig. 7 Accelerometric waveform (bottom) and corresponding Husid plot (top) at station Aqv (event 2009-04-06 01:32:32 UTC, Italy). The values of bracketed duration (bottom) and significant duration (top) are also reported



definitions of duration is presented in Fig. 7 for the L'Aquila, Italy, earthquake (M_w 6.3) recorded at Aqv. As marked in this figure, the time interval between 5 % and 95 % of the Husid plot represents the significant duration of the record.

The *Arias intensity* is directly related to the definition of the significant duration. It is an integral parameter adopted to measure the severity of ground motion and it has been found to be a useful measure for thresholds of shaking that trigger landslides (Jibson et al. 1998).

Arias intensity is defined as

$$I_A = \frac{\pi}{2g} \int_0^{T_d} a^2(t) dt \quad (1)$$

where $a(t)$ is the acceleration at time t and g the gravity acceleration, the Arias intensity is the maximum value of this function and T_d is the time corresponding to the end of the accelerogram. Arias intensity has the dimension of a velocity.

The root mean square acceleration is a parameter used in the past to measure the radiated energy. It is defined as the square root of the mean of the squared acceleration computed for a specific time interval of the record (i.e., $\Delta T = T_2 - T_1$):

$$a_{rms} = \left\{ \frac{1}{T_2 - T_1} \int_{T_1}^{T_2} a^2(t) dt \right\}^{1/2} \quad (2)$$

This parameter strongly depends on the choice of the duration and, for this reason, is of limited use inside the engineering community.

Frequency Domain Parameters

Fourier Spectra

A measure of the frequency content of the accelerogram can be obtained through the Fourier transform.

The Fourier transform of an accelerogram $a(t)$ with total duration T_d is defined as follows:

$$F(f) = \int_0^{T_d} a(t) e^{-i2\pi ft} dt \quad (3)$$

Since $F(f)$ is a complex function, it can be expressed as $F(f) = A(f) - iB(f)$

where

$$\begin{aligned} A(f) &= \int_0^{T_d} a(t) \cos(2\pi ft) dt; \quad B(f) \\ &= \int_0^{T_d} a(t) \sin(2\pi ft) dt \end{aligned} \quad (4)$$

Of particular significance to engineering seismologists is the Fourier amplitude spectrum, $FAS(f)$, defined by

$$FAS(f) = \sqrt{A^2 + B^2} \quad (5)$$

The FAS is widely used to investigate the ground motion amplitude and energy content at different frequencies.

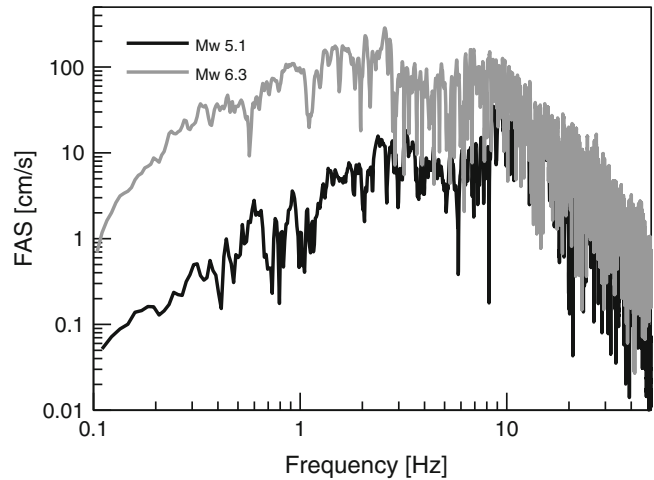
Acceleration FAS are often analyzed to decompose the observed ground motion into the spectral functions describing the source of the event, the propagation from the source to the site, and the site effect (Castro et al. 1990; Bindi et al. 2009). Acceleration FAS are also used to estimate the empirical site response, through the Standard Spectral Ratio (SSR; Tucker and King 1984), based on the ratio between the FAS between the horizontal (or vertical) components of the selected site and a reference site or the Horizontal to Vertical Spectral Ratio (HVSr; Lermo and Chávez-García 1993) based on the ratio of the FAS between the horizontal and vertical component of a single station, that provides reliable estimates of the fundamental frequency when the site effects are mainly controlled by resonance phenomena.

Records of large earthquakes and/or from soft sites usually feature larger motions at low frequencies than records of small events and/or from rock sites. Figure 8 displays the Fourier spectrum amplitude of two acceleration signals recorded at AQV for two events of magnitude M 5.1 and M 6.3, located at the same distance from the station.

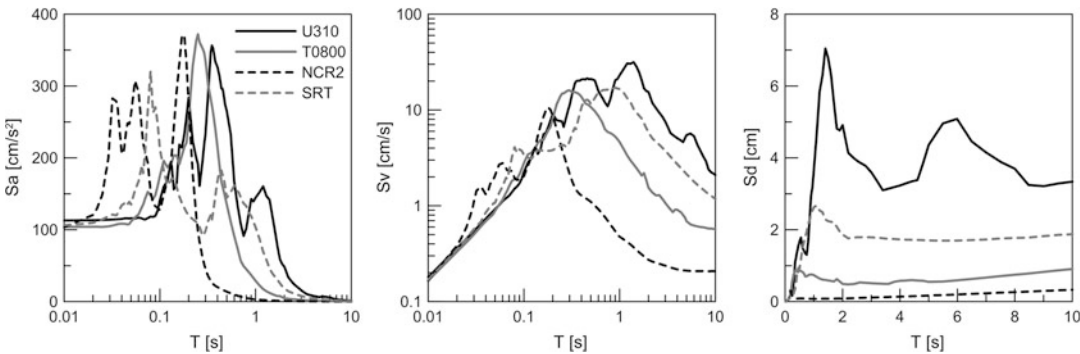
Response Spectra

For engineering purposes, the most important parameter to represent seismic motion is the response spectrum, in terms of relative displacement, relative velocity or absolute acceleration. The response spectrum is defined as the maximum response amplitude of a single-degree-of-freedom (SDOF) system, subject to the acceleration time series, as a function of the system period T_n . The response spectrum can be calculated for

Engineering Characterization of Earthquake Ground Motions, Fig. 8 Fourier amplitude spectra (FAS) for the accelerometric waveforms recorded at AQV station for event 2009-04-07, M 5.1 (Italy) at epicentral distance $R = 5.8$ km and event 2009-04-06, M 6.3 (Italy) at epicentral distance $R = 5.1$ km



E



Engineering Characterization of Earthquake Ground Motions, Fig. 9 Acceleration response spectra (SA) (*left*), velocity response spectra (SV) (*center*) and displacement response spectra (SD) (*right*) for the

accelerometric time series of Fig. 6. Different axis scales are adopted to highlight the different spectral shapes associated to the signals

different damping ratios, ξ (usually a value equal to 5 % of the critical damping is assumed, which is roughly applicable to most structures).

The maximum amplitude of the response is obtained by integrating the equation of motion of the SDOF system:

$$\ddot{x}(t) = -\omega_n^2 y(t) - 2\xi\omega_n^2 \dot{y}(t) \tag{6}$$

where

$y(t)$ and $\dot{y}(t)$ are the relative displacement and velocity of the oscillator with respect to the ground, $\ddot{x}(t)$ is the absolute acceleration of the oscillator, given by $\ddot{y}(t) + a(t)$; $\omega_n = \frac{2\pi}{T_n}$ is the

natural circular frequency of the oscillator. When acceleration, velocity, or displacement is considered, the following definitions are introduced:

Displacement spectrum (relative)	$SD(T_n, \xi) = \max_t y(t) $
Velocity spectrum (relative)	$SV(T_n, \xi) = \max_t \dot{y}(t) $
Acceleration spectrum (absolute)	$SA(T_n, \xi) = \max_t \ddot{x}(t) $

The acceleration response spectra are very common among earthquake engineers since the maximum absolute value of $\ddot{x}$, when multiplied by m , represents the maximum force imposed on the structure.

Engineering Characterization of Earthquake Ground Motions, Table 1 Strong-motion parameters inferred from accelerometric records plotted in Fig. 6

	PGA [cm/s ²]	PGV [cm/s]	PGD [cm]	AI [cm/s]	HI [cm]	T90 [s]
U301	113	13.0	2.60	17.0	43.7	16.6
NCR2	103	2.1	0.06	2.5	2.1	1.5
T0800	104	5.1	0.35	5.3	11.1	11.6
SRT	106	7.0	1.20	5.8	23.1	12.7

Figure 9 shows the acceleration, SA, velocity, SV, and displacement, SD, response spectra (5 % damping) for the acceleration time series shown in Fig. 6.

At period $T = 0$ s, the spectral displacement and velocity are zero and the spectral acceleration equals the PGA; this is the reason for which the PGA is largely adopted in seismic codes to anchor design spectra. On the other hand, when T approaches infinity, SA approaches zero and SD approaches PGD.

The pseudo-acceleration and pseudo-velocity spectra are also widely used in practice. They are defined as a function of the displacement spectrum as follows:

pseudo-acceleration spectrum

$$: PSA(T_n, \xi) = \left(\frac{2\pi}{T_n}\right)^2 SD(T_n, \xi) \quad (7)$$

pseudo-velocity spectrum:

$$PSV(T_n, \xi) = \left(\frac{2\pi}{T_n}\right) SD(T_n, \xi) \quad (8)$$

The response spectra (SA or PSA) of the processed ground accelerations are usually published alongside the accelerograms by various recording agencies.

The pseudo-spectra well approximate the absolute response spectra, in case of small damping values and intermediate and high frequencies, while they exactly match the absolute response spectra in case of an undamped oscillator. Pseudo-spectra and spectra differ at very low frequencies, since SV tends to PGV and PSV tends to zero (Hudson 1979).

A ground motion parameter related to PSV is the Housner intensity (HI) or spectral intensity

(SI) defined as the area under the pseudo-velocity response spectrum (PSV) over the period range 0.1–2.0 s (Housner 1952):

$$SI(\xi) = \int_{0.1}^{2.5} PSV(T, \xi) dT \quad (9)$$

This parameter measures the severity of seismic motion in terms of potential damage, since the majority of structures have a fundamental period of vibration in the range 0.1–2.5 s. Dimensionally, the Housner intensity is a displacement.

Table 1 lists the main strong-motion parameters relative to the records plotted in Fig. 6.

Public Sources of Strong-Motion Records

Various governmental and local agencies, as well as many universities and research institutes, operate strong-motion networks. Depending on the agency, accelerometric data are distributed in various ways (real time or triggered by an event, raw or processed, with basic metadata or revised metadata). The main sources of strong-motion data (worldwide, European, national) for engineering application are discussed in detail in this paragraph. Other sources are listed in Table 2.

Worldwide Databases

The major source of US strong-motion data is the California Strong Motion Instrumentation Program (CSMIP). The Centre for Engineering Strong Motion Data (CESMD <http://www.strongmotioncenter.org/>) distributes processed and unprocessed strong-motion data from US and major international earthquakes.

Engineering Characterization of Earthquake Ground Motions, Table 2 List of the main public sources for archiving and distributing strong-motion data

Database name	Database code	URL
Centre for Engineering Strong Motion Data	CESMD	http://www.strongmotioncenter.org/
COSMOS Virtual Data Center	COSMOS	http://www.cosmos-eq.org/VDC/index.html
Pacific Earthquake Engineering Research Center database	PEER	http://peer.berkeley.edu/peer_ground_motion_database
European Strong-Motion Database	ESMD	http://www.isesd.hi.is/ESD_Local/frameset.htm
European Integrated Data Archive	EIDA	http://eida.gfz-potsdam.de/webdc3/
Italian Accelerometric Archive	ITACA	http://itaca.mi.ingv.it/
Real-time INGV strong- motion data	ISMD	http://ismd.mi.ingv.it/
Italian Strong Motion Network – RAN Download	MOT1	http://www.mot1.it/randownload/EN/index.php
Strong-motion database of Turkey		http://kyhdata.deprem.gov.tr/2K/kyhdata_v4.php
Hellenic Accelerogram Database	HEAD	http://www.itsak.gr/en/head
Kyoshin network and Kiban Kyoshin network	K-NET and Kik-NET	http://www.kyoshin.bosai.go.jp/
Swiss Seismological Service	SED	http://www.seismo.ethz.ch/index
French accelerometric network	RAP	http://www-rap.obs.ujf-grenoble.fr/
The Guerrero Accelerograph Network (Mexico)		http://crack.seismo.unr.edu/guerrero/description.html
GeoNet (New Zealand)		http://info.geonet.org.nz/display/appdata/Applications+and+Data
Euroseistest database (Greece)		http://euroseisdb.civil.auth.gr/
Iran Strong Motion Network	ISNM	http://www.bhrc.ac.ir/Portal/Default.aspx?tabid=635
USGS National Strong- Motion Project	NSMP	http://nsmp.wr.usgs.gov/

The COSMOS Virtual Data Center (<http://www.cosmos-eq.org/VDC/index.html>) is a web portal to access strong ground motion records. It distributes records from the United States and 14 contributing countries for use by the engineering and scientific communities. It contains flexible search methods, including map-based, parameter-entry, and earthquake- and station-based searches. Display and download options allow users to view data in multiple contexts, extract and download data parameters, and download data files in convenient formats.

The Pacific Earthquake Engineering Research Center (PEER) database (http://peer.berkeley.edu/peer_ground_motion_database) has been created for engineering applications (e.g., development of ground motion prediction equations in the framework of the Next Generation Attenuation, NGA, program). The latest version of the PEER Ground Motion Database contains 3,182 three-component recordings of

104 shallow crustal earthquakes worldwide compiled from over 1,000 stations, extensive metadata (including different distance measures, various site characterizations and earthquake source data) with corrections to the information contained in the original databases. The web interface provides tools for searching, selecting, and downloading ground motion data.

European/Pan-European Databases

The European Strong-Motion Database (http://www.isesd.hi.is/ESD_Local/frameset.htm) contains more than 3,000 uniformly processed and formatted waveforms. Each waveform is associated to earthquake and station parameters. The user can search the database and databank interactively and download selected strong-motion records and associated parameters.

WebDc (<http://eida.gfz-potsdam.de/webdc3/>) is a portal to the European Integrated Data Archive (EIDA) infrastructure, which is a

distributed data center that aims to securely archive seismic waveform data gathered by European research infrastructures. Through the EIDA portal, the user can request real-time data streams from permanent networks that are mostly acquired over the internet and are immediately accessible as continuous data streams by SeedLink in real time. Raw seismological data (waveform) and station lists (Dataless SEED, Inventory XML) can be obtained from different data centers spread across Europe.

National Databases

1. *Italy*. The Italian Accelerometric Archive (ITACA, <http://itaca.mi.ingv.it>) contains about 7,500 three-component processed waveforms generated by about 1,200 earthquakes with magnitudes greater than 3, in the time span 1972–2013. Corrected and uncorrected time series and response spectra are available from the download pages, where the parameters of interest can be set and specific events, stations, waveforms, and their metadata can be retrieved. Quasi real-time Italian strong-motion data recorded by the two national accelerometric networks (RSN, managed by the National Institute of Geophysics and Volcanology, INGV and RAN, operated by Department of Civil Protection, DPC) can be accessed at ISMD (<http://ismd.mi.ingv.it/>) and MOT1 (<http://www.mot1.it/randownload/EN/index.php>).
2. *Turkey*. The strong-motion database of Turkey (http://kyhdata.deprem.gov.tr/2K/kyhdata_v4.php) contains about 10,000 (processed and unprocessed) waveforms recorded, from 1976 to the present, by the network operated by the Prime Ministry for Disaster and Emergency Management. Corrected and uncorrected time series and response spectra are available from the download pages, where events, stations, and waveforms can be searched with respect to basic parameters.
3. *Greece*. The Greek strong-motion database from 1973 to 1999 is contained in the Hellenic Accelerogram Database, HEAD (<http://www.itsak.gr/en/head>). The dataset includes 677 three-component records associated to 319 earthquakes, obtained from 117 strong-motion stations. For the major earthquakes after 2000, an additional web page has been created (<http://www.itsak.gr/en/db/data/sm/after2000>). Waveforms can be searched specifying event date and time and station code. Corrected and uncorrected time series and response spectra are available from the download pages.
4. *Japan*. Japanese records are available at <http://www.kyoshin.bosai.go.jp/>. K-NET (Kyoshin network) is a strong-motion seismograph network that consists of more than 1,000 observation stations distributed roughly every 20 km. K-NET is operated by the National Research Institute for Earth Science and Disaster Prevention (NIED) since 1996. At each K-NET station, a seismograph is installed on the ground surface. KiK-net (Kiban Kyoshin network) is a strong-motion seismograph network that consists of pairs of seismographs installed in a borehole and on the ground surface deployed at roughly 700 locations. The strong-motion data recorded by K-NET and KiK-net are transmitted in real-time to the NIED data management center and can be downloaded by users. The soil conditions at K-NET stations and the geological and geophysical data derived from drilling boreholes at KiK-net stations are available.

Codes to Select Spectrum-Compatible Strong-Motion Data

The strong-motion databases are also employed to select records that match a target spectrum and that can represent a response spectrum for a scenario event computed from a ground motion prediction equation or which may be derived from probabilistic seismic hazard analysis and presented as a uniform hazard spectrum (UHS) or which may be developed from building code provisions. The selected records, conveniently scaled to fit the target spectrum, are then used as input into numerical simulations for structural analysis and response site study. Several methods

Engineering Characterization of Earthquake Ground Motions, Table 3 List of some computer codes to select spectrum-compatible records

Name	Reference	Main characteristics
SigmaSpectra	Kottke and Rathje (2008)	It selects suites of earthquake ground motions from a library of ground motions such that the median of the suite matches a target response spectrum at all defined periods, then scales the suite such that the standard deviation agrees with the target standard deviation
	http://nees.org/resources/sigmaspectrav	The library of ground motions and the spectrum should be supplied by users
REXEL/REXEL-DISP	Iervolino et al. (2009); Smerzini and Paolucci (2013); Smerzini et al. (2013)	REXEL allows to define the design spectra according to the Eurocode 8, the new Italian Building Code (NTC08) or user defined. The strong-motion datasets contained in REXEL come from ESMD, ITACA and SIMBAD (Selected Input Motions for Displacement-Based Assessment and Design, Smerzini and Paolucci 2013)
	http://wpagina.unina.it/iuniervo/	REXEL-DISP allows to select suites of natural accelerograms compatible with <i>displacement spectra</i> of NTC08, EC8, or user defined. The records library is SIMBAD
CMS – Conditional Mean Spectrum	Baker (2011); Jayaram et al. (2011)	CSM allows one to compute the expected response spectrum associated with a target spectral acceleration (S_a) value at a single period; the records selection is based on PEER database; records compatible with predicted response spectra from NGA model (also as a conditional spectrum) can be directly selected from the web-based tool in the PEER database
ASCONA	Corigliano et al. (2012)	It allows to select records which are compatible both with acceleration and displacement response spectra The strong-motion dataset is extracted from NGA dataset, ITACA, ESMD, and K-net databases, and it only includes waveforms recorded on rock site condition
RSPMatch09	Al Atik and Abrahamson (2010)	It performs time-domain spectral matching by adding adjustment wavelets to an initial acceleration time series to generate a modified time series whose response spectrum is compatible with a specified target response spectrum
	http://nees.org/resources/rpsmatch09	The library of ground motions and the spectrum should be supplied by users

and computer codes are proposed in literature to this aim, some of them together their characteristics are reported in Table 3; other information can be found in the section devoted to this topic.

Summary

Accelerometric time series recorded during moderate and strong seismic events are fundamental for earthquake engineers and engineering

seismologists since they are immediately usable to estimate damage. Furthermore, from accelerometric waveforms, several strong-motion parameters can be inferred, such as peak values, spectral ordinates, measures of the frequency content, and the shaking duration, which are fundamental elements for seismic design and evaluation of the seismic response of structural and geotechnical systems.

This chapter presents engineering characteristics of strong ground motion. In section “[Analog](#)

and Digital Recording Instruments,” issues related to the instrumental characteristics of the analog and digital accelerometers are described.

In section “[Strong-Motion Data Processing](#),” an overview on the processing of strong-motion data is presented focusing on the evaluation of data quality and on the procedures that are commonly adopted to remove high- and low-frequency noise in the signals. In section “[Strong-Motion Intensity Measures](#),” the main strong-motion parameters in the time and frequency domains and their characteristics and applications are illustrated. These parameters include peak values (acceleration, velocity, and displacement), integral parameters (Arias and Housner intensities) and Fourier and elastic response spectra. Finally, in section “[Public Sources of Strong-Motion Records](#),” the main worldwide, European and national public strong-motion databases are listed, together with a short description of their content in terms of strong-motion data and associated information.

Cross-References

- [Conditional Spectra](#)
- [Earthquake Location](#)
- [Earthquake Magnitude Estimation](#)
- [Earthquake Response Spectra and Design Spectra](#)
- [Engineering Characterization of Earthquake Ground Motions](#)
- [European Structural Design Codes: Seismic Actions](#)
- [Probabilistic Seismic Hazard Models](#)
- [Recording Seismic Signals](#)
- [Response-Spectrum-Compatible Ground Motion Processes](#)
- [Response Spectrum Analysis of Structures Subjected to Seismic Actions](#)
- [Review and Implications of Inputs for Seismic Hazard Analysis](#)
- [Seismic Actions due to Near-Fault Ground Motion](#)
- [Seismic Instrument Response, Correction for](#)
- [Seismic Network and Data Quality](#)
- [Site Response for Seismic Hazard Assessment](#)
- [Seismic Noise](#)
- [Time History Seismic Analysis](#)

References

- Al Atik L, Abrahamson NA (2010) Nonlinear site response effects on the standard deviations of predicted ground motions. *Bull Seismol Soc Am* 100(3):1,288–1,292
- Ambraseys NN, Smit P, Douglas J, Margaris B, Sigbjörnsson R, Olafsson S, Suhadolc P, Costa G (2004) Internet site for European strong-motion data. *Boll Geofis Teorica Appl* 45(3):113–129
- Baker JW (2011) Conditional mean spectrum: tool for ground motion selection. *J Struct Eng* 137(3): 322–331
- Bazzurro P, Sjöberg B, Luco N, Silva W, Darragh R (2005) Effects of strong motion processing procedures on time histories, elastic and inelastic spectra. In: *Proceedings of the workshop on strong-motion record processing*, Richmond, 26–27 May 2004, pp 1–39
- Bindi D, Pacor F, Luzi L, Massa M, Ameri G (2009) The Mw 6.3, 2009 L’Aquila earthquake: source, path and site effects from spectral analysis of strong motion data. *Geophys J Int* 179:1573–1579. doi:10.1111/j.1365-246X.2009.04392.x
- Bolt BA (1973) Duration of strong ground motion. In: *Proceedings, 5th world conference on earthquake engineering*, Rome, pp 1304–1313
- Bommer JJ, Martinez-Pereira A (1999) The effective duration of earthquake strong motion. *J Earthq Eng* 3:127–172
- Boore DM (2001) Effect of baseline corrections on displacements and response spectra for several recordings of the 1999 Chi-Chi, Taiwan, earthquake. *Bull Seismol Soc Am* 91(5):1199–1211
- Boore DM (2003) Analog-to-Digital conversion as a source of drifts in displacements derived from digital recordings of ground acceleration. *Bull Seismol Soc Am* 93(5):2017–2024
- Boore DM (2005) On pads and filters: processing strong-motion data. *Bull Seismol Soc Am* 95(2): 745–750
- Boore DM, Akkar S (2003) Effect of causal and acausal filters on elastic and inelastic response spectra. *Earthq Eng Struct Dyn* 32(11):1729–1748. doi:10.1002/eqe.299
- Boore DM, Bommer JJ (2005) Processing of strong-motion accelerograms: needs, options and consequences. *Soil Dyn Earthq Eng* 25:93–115. doi:10.1016/j.soildyn.2004.10.007
- Boore DM, Stephens C, Joyner WB (2002) Comments on baseline correction of digital strong-motion data: examples from the 1999 Hector Mine, California, earthquake. *Bull Seismol Soc Am* 92(4):1543–1560

- Boore DM, Azarisi A, Akkar S (2012) Using pad-stripped acausally filtered strong-motion data. *Bull Seismol Soc Am* 102(2):751–760. doi:10.1785/0120110222
- Brune JN (1970) Tectonic stress and the spectra of seismic shear waves from earthquakes. *J Geophys Res* 75(26):4997–5009
- Castro RR, Anderson JG, Singh SK (1990) Site response attenuation and source spectra of S waves along the Guerrero, Mexico, subduction zone. *Bull Seismol Soc Am* 80:1481–1503
- Corigliano M, Lai CG, Rota M, Strobbia CL (2012) ASCONA: automated selection of compatible natural accelerograms. *Earthq Spectra* 28(3):965–987
- Douglas J (2003) What is a poor quality strong-motion record? *Bull Earthq Eng* 1:141–156
- Douglas J, Boore DM (2011) High-frequency filtering of strong-motion records. *Bull Earth Eng* 9(2):395–409. doi:10.1007/s10518-010-9208-4
- Foti S, Paolucci R (2012) Influence of strong motion processing on numerical simulations of soil-structure interaction. In: Second international conference on performance-based design in earthquake geotechnical engineering, Taormina, 28–30 May 2012
- Graizer VM (1979) Determination of the true ground displacement by using strong motion records. *Izvestiya Acad Sci USSR, Phys Solid Earth* 15(12):875–885
- Graves RW (2004) Processing issues for near source strong motion recordings. In: COSMOS workshop on archiving and web dissemination of geotechnical data – 2004, pp 1–9. http://www.cosmos-eq.org/events/wkshop_records_processing/index.html
- Housner GW (1952) Spectrum intensities of strong-motion earthquakes. In: Duke CM, Feigen M (eds) *Proceedings, symposium on earthquake and blast effects on structures*, Los Angeles, pp 21–36
- Housner GW (1970) Design spectrum. In: Wiegel RL (ed) *Earthquake engineering*. Prentice-Hall, Englewood Cliffs
- Hudson DE (1979) Reading and interpreting strong motion accelerograms. *Earthquake Engineering Research Institute, Berkeley (Monograph)*
- Husid R (1967) Gravity effects on the earthquake response of yielding structures. *Earthquake Engineering Research Laboratory, California Institute of Technology, Pasadena*
- Iervolino I, Galasso C, Cosenza E (2009) REXEL: computer aided record selection for code-based seismic structural analysis. *Bull Earthq Eng* 8:339–362. doi:10.1007/s10518-009-9146-1
- Iwan WD, Moser MA, Peng C-Y (1985) Some observations on strong-motion earthquake measurements using a digital accelerograph. *Bull Seismol Soc Am* 75(5):1225–1246
- Jayaram N, Lin T, Baker JW (2011) A computationally efficient ground-motion selection algorithm for matching a target response spectrum mean and variance. *Earthq Spectra* 27(3):797–815
- Jibson RW, Harp EL, Michael JA (1998) A method for producing digital probabilistic seismic landslide hazard maps: an example from the Los Angeles, California, area. *US Geological Survey O-F report* 98–113
- Kottke A, Rathje EM (2008) *Earthq Spectra* 24:911. doi:10.1193/1.2985772
- Kwon O, Elnashai AS (2006) The effect of material and ground motion uncertainty on the seismic vulnerability curves of a RC structure. *Eng Struct* 28:289–303
- Lermo J, Chávez-García FJ (1993) Site effect evaluation using spectral ratios with only one station. *Bull Seismol Soc Am* 83:1574–1594
- Pacor F, Paolucci R, Ameri G, Massa M, Puglia R (2011) Italian strong motion records in ITACA: overview and record processing. *Bull Earthq Eng* 9(6):1741–1759
- Paolucci R, Rovelli A, Faccioli E, Cauzzi C, Finazzi D, Vanini M, Di Alessandro C, Calderoni G (2008) On the reliability of long-period response spectral ordinates from digital accelerograms. *Earthq Eng Struct Dyn* 37(5):697–710. doi:10.1002/eqe.781
- Paolucci R, Pacor F, Puglia R, Ameri G, Cauzzi C, Massa M (2011) Record processing in ITACA, the new Italian strong-motion database, Chapter 8. In: Akkar S, Gulkan P, Van Eck T (eds) *Earthquake data in engineering seismology. Predictive models, data management and networks*, vol 14, *Geotechnical, geological and earthquake engineering series*. Springer, Dordrecht, pp 99–113
- Sawada T, Hirao K, Yamamoto H, Tsujihara O (1992) Relation between maximum amplitude ratio and spectral parameters of earthquake ground motion. In: *Proceedings of 10th world conference on earthquake engineering*, vol 2. Madrid, Spain, pp 617–22
- Smerzini C, Paolucci R (2013) SIMBAD: a database with selected input motions for displacement-based assessment and design – 3rd release. Report of the research project DPC – RELUIS (2010–2013). http://wpage.unina.it/iuniervo/SIMBAD_Database_Polimi.pdf
- Smerzini C, Galasso C, Iervolino I, Paolucci R (2013) Ground motion record selection based on broadband spectral compatibility. *Earthquake spectra*, in Press, doi:10.1193/052312EQS197M
- Trifunac MD, Brady AG (1975) A study of the duration of strong earthquake ground motion. *Bull Seismol Soc Am* 1975(65):581–626
- Trifunac MD, Todorovska M (2001) Evolution of accelerographs, data processing, strong motion arrays and amplitude and spatial resolution in recording strong earthquake motion. *Soil Dyn Earthq Eng* 21:537–555
- Tucker B, King J (1984) Dependence of sediment-filled valley response on input amplitude and valley properties. *Bull Seismol Soc Am* 74:153–165
- Zhu TJ, Heidebrecht AC, Tso WK (1988) Effect of peak ground acceleration to velocity ratio on ductility demand of inelastic systems. *Earth Eng Struct Dyn* 16:63–79

Equivalent Static Analysis of Structures Subjected to Seismic Actions

Nouredine Bourahla

Civil Engineering Department, University Saâd Dahlab, Blida, Algeria

Synonyms

Equivalent seismic force; Lateral seismic force; Seismic base shear; Seismic design; Seismic force distribution

Definition

The equivalent static lateral force method is a simplified technique to substitute the effect of dynamic loading of an expected earthquake by a static force distributed laterally on a structure for design purposes. The total applied seismic force V is generally evaluated in two horizontal directions parallel to the main axes of the building (Fig. 1). It assumes that the building responds in its fundamental lateral mode. For this to be true, the building must be low rise and must be fairly symmetric to avoid torsional movement under ground motions. The structure must be able to resist effects caused by seismic forces in either direction, but not in both directions simultaneously.

Introduction

Loading due to earthquake upon a given structure comes from ground motions, defined as a variation in the ground accelerations with time, resulting in inertial forces acting on the masses of the structure. These forces depend on the characteristics of the structure, the earthquake, and other factors such as distance from the fault and geology of the site.

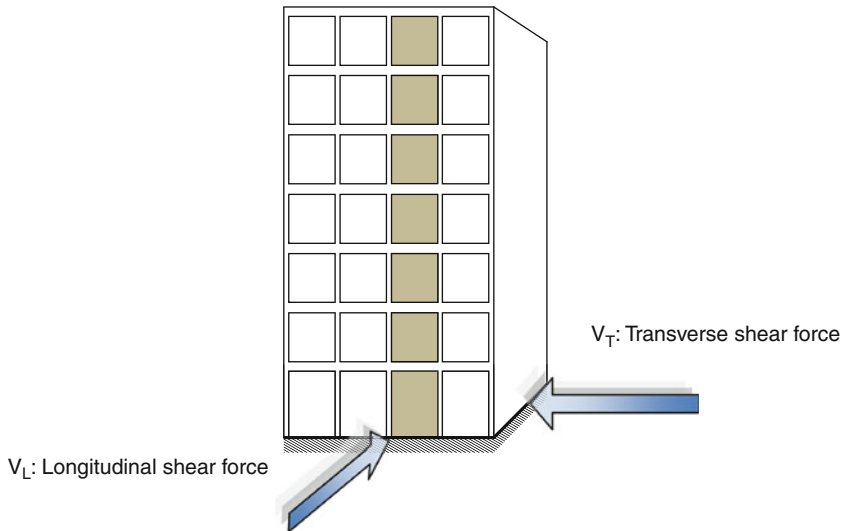
To design a structure capable to withstand the effect of an earthquake, the forces upon the structure must be first specified. However, due to the fact that each earthquake can lead to different ground motions accelerations even within a same site, the resulting forces which will affect the structure within a given prone seismic region cannot be specified with precision. As a result, a realistic estimate of such forces is important for the safety and the cost-efficiency of a structure.

The simplest way to do so is using the “equivalent static lateral force method,” even though its uses are limited by various conditions, related to the site and the structure itself and which are acknowledged and detailed below. Within this method, a seismic coefficient is applied to the mass of the structure to produce the lateral force that is approximately equivalent in effect to the dynamic loading of the expected earthquake.

Historical Background

The origin of the equivalent static lateral force method goes back to the early twentieth century, where the idea of a lateral design force V proportional to the building's weight was first developed following the Messina-Reggio earthquake of December 28, 1908. This relationship can be written as $F = CW$, where C is a lateral force coefficient, expressed as some percentage of the gravity. The first official implementation of this criterion was the specification, $C = 10\%$ of the gravity, issued as a part of the 1924 Japanese Urban Building Law Enforcement Regulations in response to the destruction caused by the great 1923 Kanto earthquake (Elnashai 2002; Reitherman 2006).

Some destructive earthquakes motivated several communities to adopt codes with different percentages for the seismic coefficient C . Later the equivalent static lateral force was based on the seismic zone factor, building period, building weight, and lateral resisting system type. These factors are still being refined and used to estimate the equivalent lateral force (Mitchell et al. 2010).



Equivalent Static Analysis of Structures Subjected to Seismic Actions, Fig. 1 Equivalent lateral shear force along two orthogonal axes

Domain of Application

The equivalent lateral static force may be applied to buildings in which the response is not significantly affected by contributions from modes of vibration higher than the fundamental mode in each principal direction and meeting the criteria for building height, horizontal and vertical regularities. The use of this method is also restricted with respect to the seismic zones. This suggests that this method is limited to structures having regular configurations with continuous seismic-force-resisting elements and a relatively uniform distribution of mass and stiffness. The main irregularities that most seismic codes use as criteria to classify the building configuration include the following aspects (FEMA 749 2010):

Horizontal Irregularities

- Torsional irregularity – This condition exists when the distribution of vertical elements of the seismic-force-resisting system within a story is such that when the building is pushed to the side by a lateral force, it will tend to twist as well as to deflect horizontally.

- Reentrant corner irregularity – This is a geometric condition that occurs when a building with an approximately rectangular plan shape has a missing corner or when a building is formed by multiple connecting wings.
- Diaphragm discontinuity irregularity – This occurs when a structure's floor or roof has a large open area as can occur in buildings with large atriums.
- Out-of-plane offset irregularity – This occurs when the vertical elements of the seismic-force-resisting system are not aligned vertically from story to story.
- Nonparallel systems irregularity – This occurs when the structure's seismic-force-resisting does not include a series of frames or walls that are oriented at approximately 90° angles with each other.

Vertical Irregularities

- Stiffness soft-story irregularity – This occurs when the stiffness of one story is substantially less than that of the stories above.
- Weight/mass irregularity – This exists when the weight of the structure at one level is

substantially in excess to the one at the levels immediately above or below it.

- In-plane discontinuity irregularity – This occurs when the vertical elements of a structure's seismic-force-resisting system such as walls or braced frames do not align vertically within a given line of framing or the frame or the wall has a significant setback.

The classification of a structure as “regular” or “irregular” affects several aspects of the design and analysis such as the permission to use the equivalent lateral force, the behavior factor, and the type of models for analysis.

Equivalent Lateral Static Force Procedure

The equivalent lateral force procedure provides a simple way to incorporate the effects of inelastic dynamic response into a linear static analysis. This procedure is useful in preliminary design of all structures and is allowed for final design of the vast majority of structures. As stated before, the procedure is valid only for structures without significant discontinuities in mass and stiffness along the height, where the dominant response to ground motions is in the horizontal direction without significant torsion (FEMA 750 2007). The equivalent static analysis procedure involves the following steps:

- Calculation of the total lateral seismic force V
- Vertical distribution of seismic forces along the height of the structure
- Horizontal distribution of the level forces across the width and breadth of the structure
- Calculation of the additional forces from the inherent and accidental torsion
- Determination of the drift, overturning moment, and P-Delta effect

Calculation of the Total Lateral Seismic Force V

Although the equivalent lateral static force is a code-dependent relationship, the general formula is expressed as the product of the effective

seismic weight W and a response coefficient C_S . It can be written as follows:

$$V = C_S W \quad (1)$$

where

W is the effective seismic weight. It includes the total *dead load* above the foundation or above the top of a rigid basement and applicable portions of other loads (*live load* and *snow*). For instance, in areas used for permanent storage such as library shelves, fixed equipments, etc., the total floor *live load* should be applicable. In areas used for housing or administrative purposes, only a percentage (20–30 %) of the floor *live load* should be applicable.

C_S is the seismic response coefficient expressed as a spectral pseudoacceleration, in g units, which depends on several factors to account for the seismic zone, the structure period, the local soil conditions, the inelastic performance of the lateral resisting system, and the nature of the structure's occupancy and use.

It might help to understand this formula to go back to the basic relationship between the force and the acceleration, which is $F = Ma$.

Seismic Response Coefficient C_S

As noted previously, the seismic response coefficient C_S depends on several factors and can be written in different manner according to the seismic codes. In all cases the controlling parameters are the same:

$$C_S = f(S_a, I, R, T, T_S, T_L) \quad (2)$$

S_a : design ground acceleration

I : importance factor

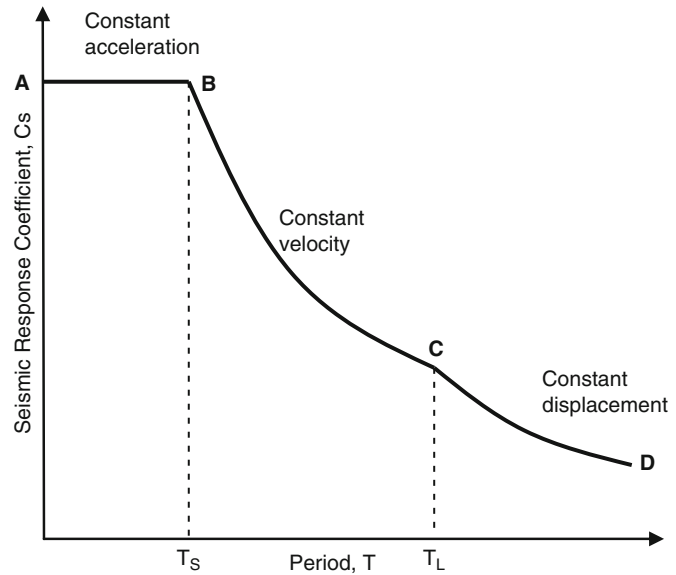
R : lateral resisting behavior factor (reduction or modification factor)

T : structure fundamental period

T_S and T_L : soil/seismicity depending factors (periods)

Before going into the details of each parameter, it is useful to know what data is necessary to obtain the seismic response coefficient. The following items refer to the information needed to obtain the seismic response coefficient:

Equivalent Static Analysis of Structures Subjected to Seismic Actions, Fig. 2 Seismic response coefficient spectral shape



E

- The location where the building will be constructed to determine the seismic zone (seismic hazard).
- The soil conditions must be known to determine the Site Class (T_S and T_L), and this requires an investigation that includes boring and removing soil samples at various depths.
- The structure fundamental period T , calculated or estimated using empirical formulae.
- The lateral resisting system type, to determine the modification factor R .
- The structure use, including consideration of the number of people who would be affected by the structure failure and the need to use the structure for its intended purpose after an earthquake. These information are necessary to determine the importance or occupancy class (I) of the building.

Design Ground Acceleration (S_a)

The design ground motion is one of the primary factors used to determine the required seismic resistance (strength) of structures and supported nonstructural components. In many seismic codes the design ground acceleration is used as a product of an acceleration factor which depends on the seismic zone and a dynamic amplification factor which depends on the soil class and the

fundamental period of the structure. Recently, the design ground motion is defined by an acceleration response spectrum characterized by three ranges as shown on Fig. 2.

AB represents the maximum constant acceleration part of the spectrum and lays in the range $0.0 < T < T_S$. T_S is a function of seismicity and site (in some codes it depends only on the site conditions). Depending on the seismic code, T_S may be as low as 0.15 s for low-hazard regions on very hard soil conditions or as high as 0.9 s in high-hazard regions on very soft soil conditions.

BC represents the constant velocity part of the spectrum in the range $T_S < T < T_L$. In this region, the seismic response coefficient is inversely proportional to the period, and the pseudovelocity (pseudoacceleration divided by circular frequency, ω) is constant. T_L the long-period transition ranges from 3 to 4 s in some codes.

CD represents the constant displacement part of the spectrum for $T > T_L$. In this region C_S is inversely proportional to T^2 . This part only affects tall and flexible structures.

Importance Factor

The importance class or factor depends on the occupancy category of the building. Hence, essential facilities such as hospitals and fire and

police stations are designed for seismic forces greater than normal. In this way, such structures are expected to resist an earthquake and be able to be functional in the post-earthquake period. Buildings containing certain highly toxic substances must also be designed to meet these higher standards. Similarly, high-occupancy buildings that represent a substantial hazard to human life (schools, health care facilities, and stadiums) must be designed for seismic forces less than vital category and greater than normal. Finally, for constructions which are not subject to human occupancy with insignificant economic loss can be designed for seismic forces less than normal. The importance factor is used either as an independent factor or combined with the seismicity and site parameters to categorize structures according to the seismic risk they could pose (seismic design category concept) (IBC 2012; FEMA 750 2007). In certain codes the importance factor may increase the seismic forces for vital facilities up to 1.5 times the normal seismic forces.

Site Class (Ground Conditions)

To take into account the site effect on the estimation of the equivalent lateral static force, the concept of Site Class is used to categorize common soil conditions into broad classes to which typical ground motion attenuation and amplification effects are assigned. Site Class is determined based on the average properties of the soil within a certain depth (30 m) from the ground surface. Most of the seismic codes specify 4–6 categories of soils ranging from hard, competent rock materials to very soft, loose materials. For each category are assigned the parameters T_S and T_L which define the transition points on the acceleration design spectrum.

Behavior (Modification) Factor R

The behavior factor or the response modification factor R , which is determined by the type of lateral load resisting system used, is a measure of the system's ability to accommodate earthquake loads and absorb energy without collapse. The resistance and energy-dissipation capacity to be assigned to the structure are related to the

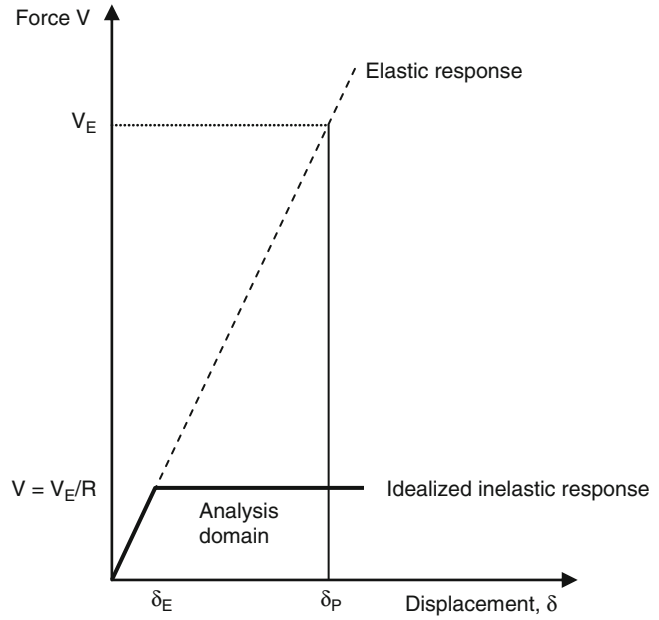
extent to which its nonlinear response is to be exploited. In operational terms, such balance between resistance and energy-dissipation capacity is characterized by the values of the behavior factor R and the associated ductility classification. As a limiting case, for the design of structures classified as non-dissipative, no account is taken of any hysteretic energy dissipation and the value of the behavior factor may not be taken, in general, as being greater than the value of 1.5. For dissipative structures, the value of the behavior factor can be taken greater than this limiting value in order to account for the hysteretic energy dissipation that mainly occurs in specifically designed zones, called dissipative zones or critical regions (EN 1998–1 2004). In general, seismic codes provide tables listing different lateral resisting systems with associated values of R .

In terms of calculation, instead of performing an explicit inelastic structural analysis in design, the capacity of the structure to dissipate energy, through mainly ductile behavior of its elements, is taken into account by performing an elastic analysis based on a lateral force reduced by a factor R (Fig. 3). Thus, the behavior factor R is an approximation of the ratio of the seismic forces that the structure would experience if its response was completely elastic to the seismic forces that may be used in the design, with a conventional elastic analysis model.

Fundamental Period Determination

The fundamental period of the structure, T , is used to determine the design ground acceleration and in some codes to establish the distribution of the shear along the height of the structure. The fundamental period T can be determined from a rational analysis using the structural properties and deformational characteristics of the resisting elements or, alternatively, using an approximate empirical relationship. In most seismic codes, the calculated period by analysis is limited by a coefficient times the approximate (empirical) period. This period limit prevents the use of unusually low base shear when using the equivalent static lateral force for design of

Equivalent Static Analysis of Structures Subjected to Seismic Actions, Fig. 3 Reduction factor R based on equivalent elastic plastic deformation



buildings (or computational models) that are overly flexible (FEMA 750 2007).

The most common empirical expression used to estimate the fundamental period is given by:

$$T = C_t h^x \quad (3)$$

where C_t and x are coefficients which depend on the type of lateral resisting system and h is the height of the building, in m, from the foundation or from the top of a rigid basement.

Vertical Distribution of the Equivalent Static Force

Once the total base shear is known, it is used to determine the forces on the various building elements. For multistory buildings, the forces must be distributed according to the fundamental mode which can be simulated by an inverted force triangle varying from zero at the base to a maximum at the top (Fig. 4). With this simplified approach, it is assumed that there is uniform mass distribution and equal floor heights, but the codes do provide for variations, where the force at a given level is proportional to the height

h times the lumped mass W at that height according to Eq. 4. The sum of the loads at each level equals the total base shear. Also note that while the greatest force is at the top of the building, the shear increases from zero at the top to its maximum at the base. Each floor shear is successively added to the sum from above.

To account for higher modes participation in tall buildings that do not have a uniform triangular distribution, some codes require that an additional force F_t be applied at the top of the building. The lateral force induced at any level h_i can be determined from the following equation:

$$F_i = \frac{W_i h_i}{\sum_{j=1}^n W_j h_j} (V - F_t) \quad (4)$$

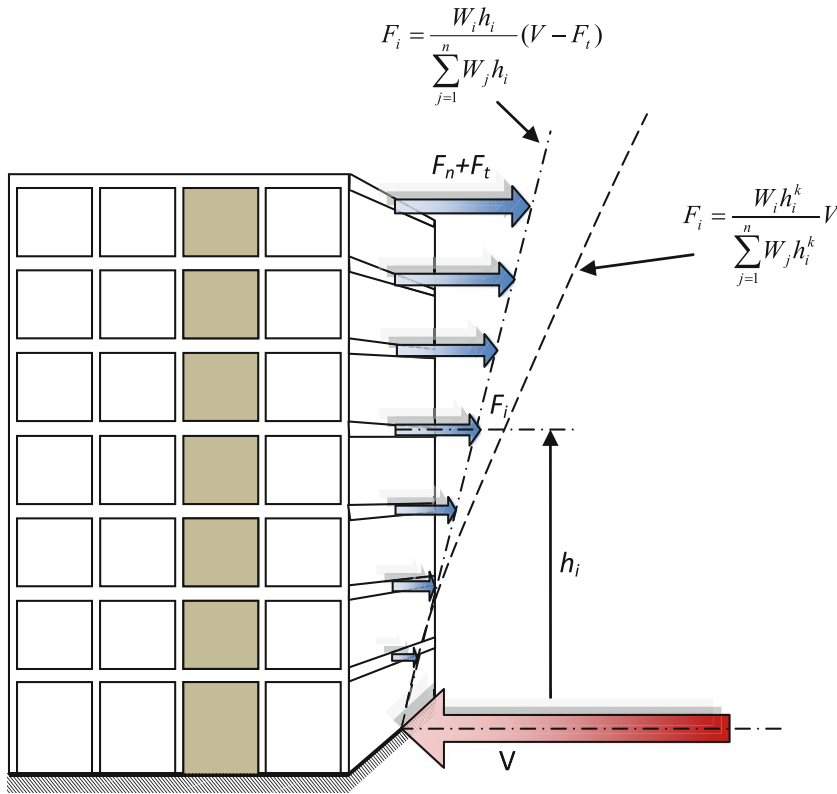
where

F_i : lateral force applied at level i

V : total lateral static force (Eq. 1)

F_t : additional force to be applied at the top of tall buildings to account for higher mode participation

W_i, W_j : lumped story masses at level i and j



Equivalent Static Analysis of Structures Subjected to Seismic Actions, Fig. 4 Vertical distribution of the lateral static force

h_i, h_j : heights of the masses $W_i W_j$ above the level of application of the seismic action (foundation or top of a rigid basement)

n : number of lumped masses along the height of the building (the number of floors in multistory buildings)

Alternatively, to take account of higher modes, some codes prescribed the following distribution formula:

$$F_i = \frac{W_i h_i^k}{\sum_{j=1}^n W_j h_j^k} V \quad (5)$$

The superscript k has a value of unity for structures with a fundamental period (T) less than or equal to 0.5 s, has a value of 2 for structures with a fundamental period greater than or equal to 2.5 s, and has a value that is linearly

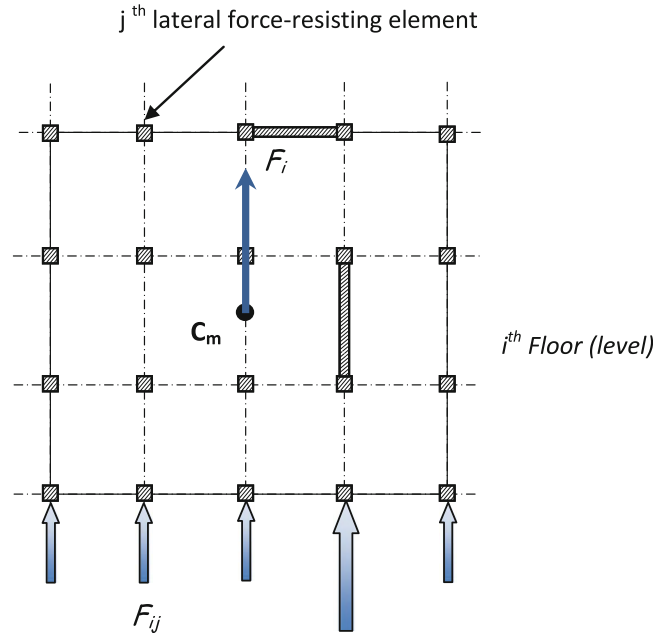
interpolated from these values for structures with a fundamental period that falls between these values (FEMA 750 2007).

Horizontal Distribution of the Equivalent Static Force

Within the context of an elastic analysis, the distribution of lateral forces to various seismic-force-resisting elements depends on the type, geometric arrangement, and vertical extents of the resisting elements and on the shape and flexibility of the floor diaphragms. In a simplified manner, the forces F_i applied at each level shall be distributed to the various elements of the vertical lateral force-resisting system in proportion to their rigidities, considering that the diaphragm is rigid and the lateral force-resisting system is

Equivalent Static Analysis of Structures Subjected to Seismic Actions,

Fig. 5 Horizontal distribution of the lateral force at floor (level) i



composed of several lines parallel to the direction of the seismic force (Fig. 5). The distribution expression can be written as follows:

$$F_{ij} = \frac{K_{ij}}{\sum_{k=1}^{nk} K_{ik}} F_i \quad (6)$$

F_{ij} : force acting on the lateral force-resisting line j at a floor level i

nk : number of lateral force-resisting elements (lines)

K_{ij} , K_{ik} : story stiffness of the lateral force-resisting element (line) k and j at level i

F_i : seismic force at floor (level) i

The design shall include the torsional moment, resulting from the eccentricities of the location of the masses and the centers of rigidities. In addition to the torsional moment, the design also should include accidental torsional moments. For rigid diagrams, the mass at each level is assumed to be displaced from the calculated center of mass in each direction a distance equal to 5 % or 10 % of the building dimension at that level perpendicular to the direction of the force under consideration. The torsional moment creates torsional displacement (twisting about the

vertical axis) and will be resisted by the lateral force-resisting elements (lines) in proportion to their rigidities and their distances from the center of rigidity (Fig. 6).

The additional force resulting from the torsional moment is given by the following expression:

$$F_{\theta j} = \frac{K_{ij} r_j}{\sum_{k=1}^{nk} K_{ik} r_k^2} F_i e \quad (7)$$

$F_{\theta j}$: force resulting from the torsional moment acting on the lateral force-resisting element (line) j

F_i : seismic force at floor (level) i

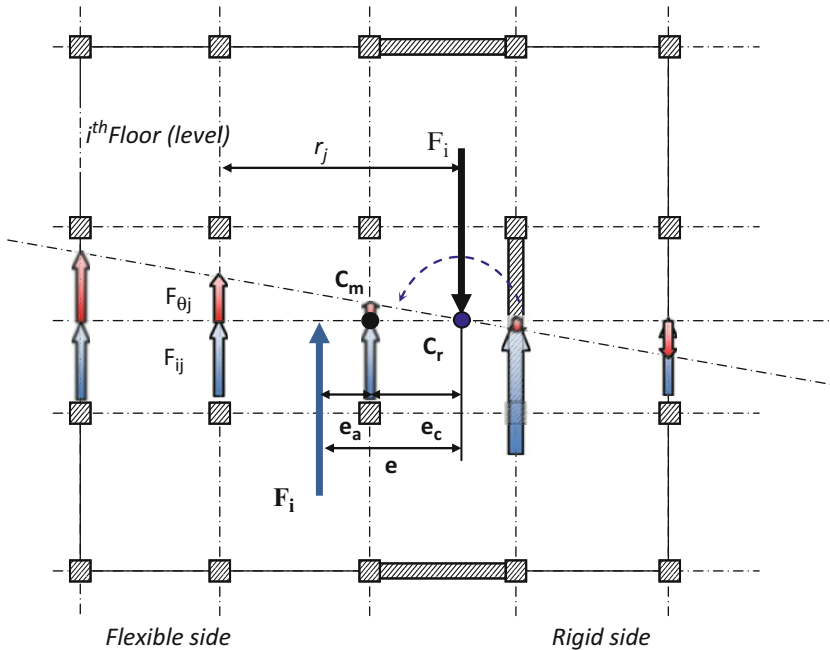
e : total eccentricity

$K_{ij} K_{ik}$: story stiffness of the lateral force-resisting element (line) k and j at level i

$r_k r_j$: distance from the axes of the force-resisting element (line) k and j to the rigidity center C_r

nk : number of lateral force-resisting elements (lines)

It should be noted that in seismic design when the force on a vertical element caused by the eccentricity acts in the same direction as that



Equivalent Static Analysis of Structures Subjected to Seismic Actions, Fig. 6 Torsional moment effect due to eccentric lateral force for rigid diaphragm

caused by the lateral load directly (lateral force-resisting elements located on the flexible side), they must be added. However, when the torsional force acts in the opposite direction (rigid side), it will not be subtracted.

Minimum Base Shear Limit

In many seismic codes, the equivalent static lateral force V is used as a reference value of the total seismic design base shear. For instance, the base shear calculated using the modal response spectrum analysis should not be less than 80 % or 90 % of the equivalent lateral base shear. This minimum base shear is provided because the computed period of vibration may be the result of an overly flexible (incorrect) analytical model.

Drift Story

Seismic design of structures must ensure that the anticipated lateral deflection in response to

earthquake shaking does not exceed acceptable levels. The lateral deflection is expressed in terms of the inter-story drift which is a measure of how much one floor or roof level displaces under the lateral force relative to the floor level immediately below. It is typically presented as a ratio of the difference in deflection between two adjacent floors divided by the height of the story that separates the floors. The determination of the story drifts has to be based on the application of the design seismic forces to a mathematical model of the physical structure. The model shall include the stiffness and strength of all elements that are significant to the distribution of forces and deformations in the structure.

Since the forces used for design include the response modification coefficient, R , the resulting displacements are too small and must be amplified. Most of the seismic codes account for this difference, by multiplying the elastic displacements by the modification coefficient R or a deflection amplification factor which is typically similar to R , but a little less than,

the R coefficient. Therefore, the deflection at a level I is given by

$$\delta_I = R.15em\delta_{ei} \quad (8)$$

The inter-story deflection becomes

$$\Delta_I = \delta_i - \delta_{i-1} \quad (9)$$

The drift is also an important consideration for structures constructed in close proximity to one another. In response to strong ground shaking, adjacent structures can hit one another, an effect known as pounding. Pounding can induce very high forces in a structure at the area of impact and has been known to be the cause of collapse of some structures. Therefore, the seismic codes require that structures be set far enough away from one another and from property lines to avoid pounding.

Overtaking Moment and P-Delta Effects

Because the inertial force created by an earthquake represented by the equivalent lateral force acts through the center of mass of a building, there is a tendency for the moment created by this force acting above the base to overturn the structure. This overturning force must be counteracted by stabilizing load. Normally, the dead weight of the building, also acting through the center of mass, is sufficient to resist the overturning force, but it must always be checked.

P-Delta or second-order effect is an instability phenomenon that results from the simultaneous actions of the lateral force and dead load. Seismic provisions require an evaluation of the structure stability under the anticipated lateral deflection by calculating a “stability coefficient” θ for each story. The value of θ is given by the following expression:

$$\theta = \frac{P_i \Delta_i}{V_i h_i} \quad (10)$$

In this formula, P_i is the weight of the structure above the story being evaluated, Δ_i is the design

story drift determined using Eq. 9, V_i is the sum of the lateral seismic design forces above the story, and h_i is the story height. If the calculated value of θ at each story is less than or equal to 0.1, the structure is considered to have adequate stiffness and strength to provide stability. If the value of θ exceeds 0.1, the lateral force analysis must include explicit consideration of P-Delta effects. These effects are an amplification of forces that occurs in structures when they undergo large lateral deflection. A limiting value for θ ($\theta_{\max}$) varying between 0.2 and 0.3 is generally prescribed. If the structure exceeds this limiting value, it is considered potentially unstable and must be redesigned unless nonlinear response history analysis is used to demonstrate that the structure is adequate.

Summary

The equivalent static lateral force method is a simplified technique to substitute the effect of dynamic loading of an expected earthquake by a static force distributed laterally on a structure for design purposes. The total applied seismic force V is generally evaluated in two horizontal directions parallel to the main axes of the building.

The equivalent static analysis procedure involves the following steps:

- Calculation of the total lateral seismic force V which depends on the seismic zone (design ground acceleration), the building importance, the soil conditions, the fundamental period of the structure, the lateral force-resisting system, and the effective seismic weight (dead and part or entire of live load)
- Determination of the vertical distribution of seismic forces along the height of the structure, where the force at a given level is proportional to the height h times the lumped mass m at that height (simplified method)
- Considering a horizontal distribution of the level forces to the various elements of the vertical lateral force-resisting system in proportion to their rigidities, considering that the diaphragm is rigid

- Calculation of the additional forces from the inherent and accidental torsion to be added to the forces resulting from the horizontal distribution of the level forces
- Determination of the drift, overturning moment, and P-Delta effect that are the direct results of the action of the lateral seismic forces

Cross-References

- Behavior Factor and Ductility
- Earthquake Response Spectra and Design Spectra
- European Structural Design Codes: Seismic Actions
- Modal Analysis
- Response Spectrum Analysis of Structures Subjected to Seismic Actions
- Structural Design Codes of Australia and New Zealand: Seismic Actions

References

- Elnashai AS (2002) A very brief history of earthquake engineering with emphasis on developments in and from the British Isles. *Chaos Soliton Fract* 13(5):967–972
- EN 1998–1 (2004) Eurocode 8: design of structures for earthquake resistance. Part 1: general rules, seismic actions and rules for building. European Committee of Standardization, Brussels
- FEMA 749 (2010) Earthquake-resistant design concepts. FEMA P-749. Building Seismic Safety Council, Federal Emergency Management Agency, Washington, DC. http://www.fema.gov/media-library-data/20130726-1759-25045-5477/fema_p_749.pdf. Accessed 25 July 2013
- FEMA 750 (2007) NEHRP recommended seismic provisions for new buildings and other structures. FEMA P-750. Building Seismic Safety Council, Federal Emergency Management Agency, Washington, DC. http://www.fema.gov/media-library-data/20130726-1730-25045-1580/femap_750.pdf. Accessed 18 July 2013
- International Code Council (2012) International Building Code (IBC), 2012th edn. ICC, Country Club Hills
- Mitchell D, Paultre P, Tinawi R, Saatcioglu M, Tremblay R, Elwood K, Adams J, DeVall R (2010) Evolution of seismic design provisions in the National building code of Canada. *Can J Civ Eng* 37:1157–1170
- Reitherman R (2006) Earthquake that initiated the development of earthquake engineering. *Bull NZSEE* 39(3):145–157

Estimation of Potential Seismic Damage in Urban Areas

Aysegul Askan, Murat Altug Erberik and Mustafa Bilal
Civil Engineering Department, Middle East Technical University, Ankara, Turkey

Synonyms

17 August Kocaeli (İzmit) earthquake; Damage probability matrices (DPM); Fragility curves; Masonry buildings; Reinforced concrete buildings; Seismic damage

Introduction

Large earthquakes are known to have significant damage potential in urban regions. Recently all over the world, efforts are made toward reduction of future probable damages. The first step in damage mitigation is the estimation of expected damage levels in buildings that are subjected to earthquakes with different intensities. Due to the inherent uncertainties involved in the analyses, estimation of seismic damage rates must be handled within a probabilistic framework.

To assess the damage rates under different shaking levels, “seismic damage” needs to be quantified and measured in a standard manner. The most common approach to quantify seismic damage rates is to perform fragility analyses. Fragility is defined as the probability of a system reaching a limit state as a function of seismic intensity levels (Kafalı and Grigouri 2004). Fragility analyses are generally carried out with one of the following five alternative approaches: empirical, analytical, subjective, experimental, and hybrid methods. The results of such damage probability assessment analyses can be expressed in terms of fragility curves or damage probability matrices. Fragility curves are continuous curves that state the cumulative damage rates whereas damage probability matrices

Estimation of Potential Seismic Damage in Urban Areas, Table 1 A typical damage probability matrix

Damage state (DS)	Central damage ratio (%)	MMI = V	MMI = VI	MMI = VII	MMI = VIII	MMI = IX
None	0	Damage state probabilities, Pr(DS,I)				
Light	5					
Moderate	30					
Heavy	70					
Collapse	100					

express the same information in terms of discrete damage probabilities for a particular intensity level.

Empirical method is based on the fundamental idea that similar type of structures experiences comparable damage rates under earthquakes. It involves analysis of empirical data collected in post-earthquake surveys. In the analytical method, in order to estimate the limit states of the structure, the seismic performance under a given ground motion level is assessed through detailed time history analysis or other simplified methods. Subjective method includes expert opinions for the assessment of the damage probabilities under various levels of shaking intensities. Experts are requested to give their opinions on the level of the damage based on their relevant experience on seismic damage assessment. It is also possible to generate fragility curves based on the experimental data that has been obtained through a series of tests. Finally, hybrid method involves a combination of the previously described approaches for assessment of damage probabilities.

An important problem is the comparison of damage probabilities obtained with these alternative methods. In this study, empirical and analytical approaches are compared in terms of damage data from a recent earthquake that occurred in Turkey. For this purpose, first, empirical damage probabilities are obtained from the damage database of the 17 August 1999 Kocaeli earthquake. Then, the results are compared with the damage rates from the analytical method. Results of this study and similar studies are important in terms of accurate estimation of seismic risks in earthquake-prone regions in the world.

Method

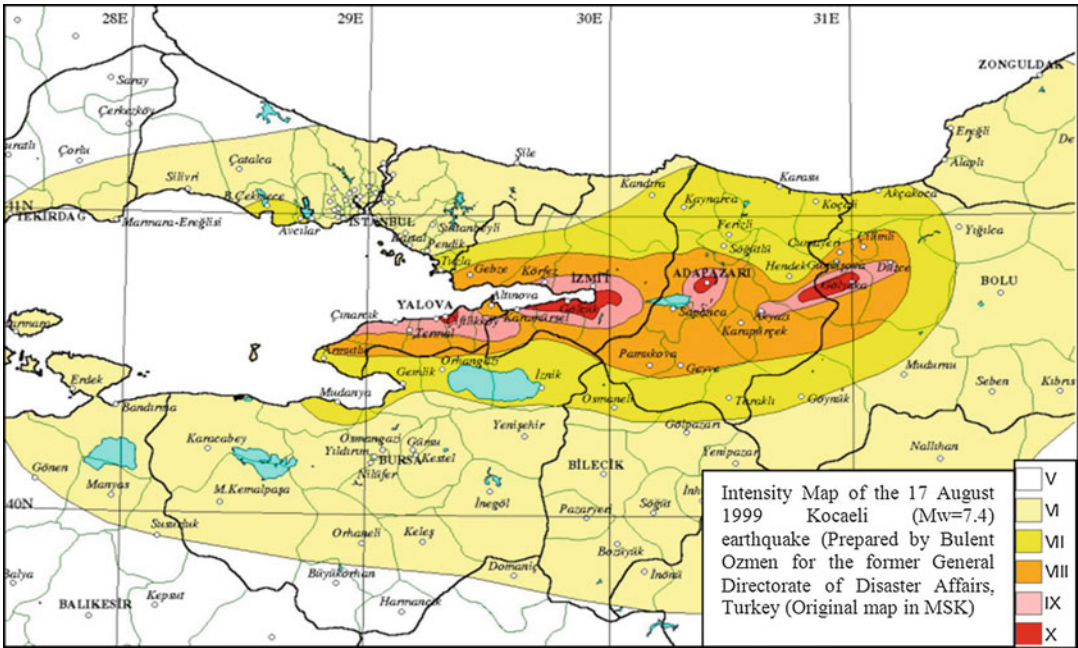
In this section, empirical and analytical methods for estimating damage state probabilities will be presented. Before that, damage probability matrices are explained briefly.

Damage Probability Matrices

Damage probability matrices (DPMs) were first introduced by Whitman (1973) for multistory buildings in the aftermath of the 1971 San Fernando earthquake. An element in this matrix, Pr(DS,I), gives the probability of occurrence of a damage state (DS) under an earthquake intensity, I. The sum of any column in this matrix is equal to unity. Table 1 shows an example damage probability matrix.

Damage state is the verbal or quantitative representation of the damage that would result in a structure under an earthquake of specific intensity. Verbal definitions are generally used in qualitative assessment of damage. A quantitative expression is used to represent the damage if the original cost of the structure is known. The ratio of the cost of repair to the cost of replacement of the structure is defined as damage ratio (DR). As the name implies, this ratio takes values that range from 0 % to 100 %. However, representative damage ratios are required in order to associate with the damage states and make the calculations easier. Therefore, the term central damage ratio (CDR) is defined for a representative value of each damage ratio. The CDR values in Table 1 are taken from Gürpınar et al. (1978).

Using a single damage ratio is preferred to summarize a full DPM. This ratio is called



Estimation of Potential Seismic Damage in Urban Areas, Fig. 1 Observed intensity map of the 17 August 1999 Kocaeli (Mw = 7.4) earthquake (Adapted from Özmen 1999)

mean damage ratio (MDR) and is obtained by the following equation:

$$MDR(I) = \sum_{DS} Pr(DS, I) * CDR(DS) \quad (1)$$

where $MDR(I)$ is the mean damage ratio corresponding to the intensity level I , $Pr(DS, I)$ is the damage state probability of the building type of interest subjected to the earthquake of intensity I , and $CDR(DS)$ is the central damage ratio corresponding to damage state DS .

In most of the previous studies, Modified Mercalli Intensity (MMI) is used as the seismic intensity scale in DPMs. Similarly, in this study MMI scale is employed to quantify the intensity of the ground shaking.

Empirical Method for Damage Probability Matrices

In the empirical approach, damage state probabilities are obtained with simple frequency analyses of different damage states in the damage database of interest. In other words, the damage

state probabilities given in Table 1 are calculated as follows:

$$Pr(DS, I) = \frac{N(DS, I)}{N(I)} \quad (2)$$

where $N(DS, I)$ is the number of buildings that are in damage state DS due to an earthquake with intensity I and $N(I)$ is the total number of buildings subjected to the same event.

In this study, regional empirical DPMs are formed using the detailed damage database of 17 August 1999 Kocaeli (Mw = 7.4) earthquake. This destructive earthquake caused significant structural damage resulting in almost 20,000 casualties and considerable economic losses. While computing empirical damage probabilities, initially the isoseismal (intensity) map of the earthquake is investigated (Fig. 1). Isoseismal maps demonstrate spatial distribution of seismic intensities in the form of contour lines of equal intensity values. These maps are mostly generated in the field in the aftermath of an earthquake. Based on the fact that different intensities are

Estimation of Potential Seismic Damage in Urban Areas, Table 2 Empirical damage probability matrix for Kocaeli city center (MMI = IX)

Damage state	CDR (%)	Masonry buildings	Reinforced concrete buildings
No damage	0	0.70	0.16
Light damage	5	0.12	0.30
Moderate damage	30	0.08	0.29
Severe damage and collapse	85	0.11	0.25
MDR (%)		11.91	31.75
Total number of flats		14.086	25.351

observed in different locations during the same event (Fig. 1), in this study city-based empirical DPMs are constructed.

While constructing the city-wise DPMs, damage assessment forms filled by the experts of the (former) General Directorate of Disaster Affairs of Turkey are used. The probability values per each damage state are calculated simply by Eq. 2. The resulting DPMs are based on assessment of households instead of buildings, since the data is collected on household basis. Converting number of households to the number of buildings is not performed since such a conversion would further introduce error into the data. The regional DPMs in this study are formed by processing the database involving approximately 170,000 flats in the central districts of Kocaeli, Sakarya, and Bolu. The DPMs are prepared in terms of two main types of buildings in Turkey: reinforced concrete and masonry buildings. Tables 2, 3, and 4 present the empirical DPMs constructed for city centers of Kocaeli, Sakarya, and Bolu, respectively, corresponding to the discrete MMI value experienced in each city. These city centers are all located in Northwestern Turkey, and they experienced significant damage in the 17 August Kocaeli earthquake.

As observed from Tables 2, 3, and 4, in Sakarya, the MDR for reinforced concrete buildings is found to be greater than those in Kocaeli and Bolu. On the other hand, the highest MDR for masonry buildings is observed in Kocaeli. The

Estimation of Potential Seismic Damage in Urban Areas, Table 3 Empirical damage probability matrix for Sakarya city center (MMI = X)

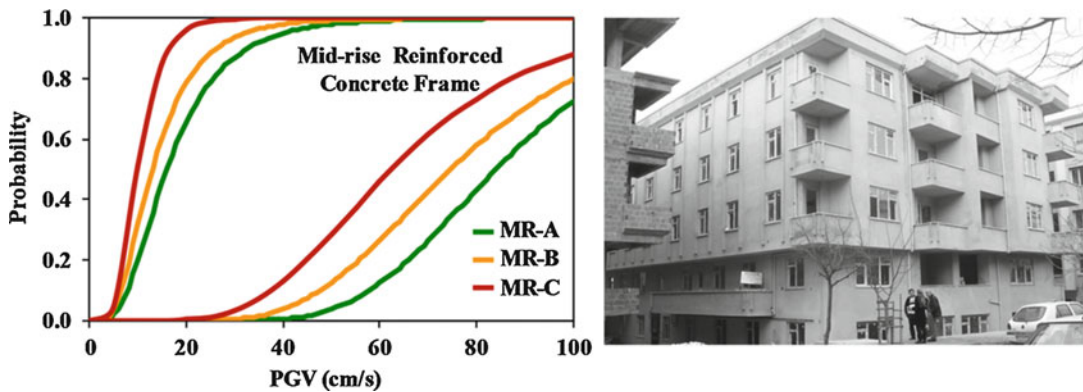
Damage state	CDR (%)	Masonry buildings	Reinforced concrete buildings
No damage	0	0.80	0.12
Light damage	5	0.08	0.22
Moderate damage	30	0.04	0.25
Severe damage and collapse	85	0.07	0.41
MDR (%)		7.51	43.28
Total number of flats		78.444	22.463

Estimation of Potential Seismic Damage in Urban Areas, Table 4 Empirical damage probability matrix for Bolu city center (MMI = VII)

Damage state	CDR (%)	Masonry buildings	Reinforced concrete buildings
No damage	0	0.86	0.20
Light damage	5	0.06	0.16
Moderate damage	30	0.03	0.36
Severe damage and collapse	85	0.05	0.27
MDR (%)		5.50	34.64
Total number of flats		12.661	7.814

latter observation is expected since Kocaeli is the closest city center to the epicenter of the earthquake. It is however interesting to observe that the highest MDR for reinforced concrete buildings is the highest in Sakarya. This result can be explained as follows: It is well known that the local soil conditions in Sakarya are worse than those in Kocaeli, and several sites in Sakarya city center indeed experienced liquefaction during the earthquake. Thus, the soil amplifications at frequencies close to the fundamental modes of reinforced concrete buildings are most probably the cause of significant damage in those buildings in Sakarya.

These empirical results are compared with the analytical ones in section “[Comparison of Results from Alternative Methods.](#)”



Estimation of Potential Seismic Damage in Urban Areas, Fig. 2 Fragility curves of mid-rise reinforced concrete buildings (Erberik 2008b)

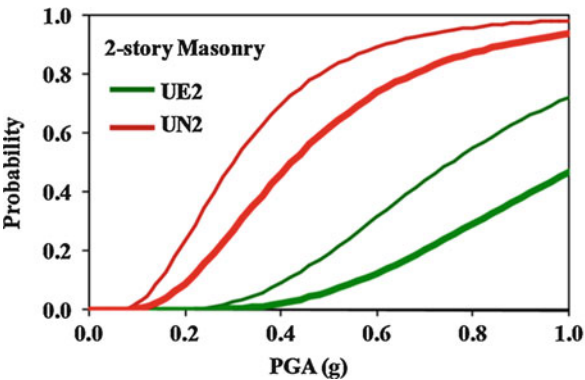
Analytical Method for Damage Probability Matrices

The first step of the analytical method involves construction of fragility curves based on the structural analyses of the building group of interest. Then, the continuous information on fragility curves is converted to discrete damage state probabilities in DPMs. There are different analysis approaches for deriving fragility curves: Static methods are simpler but yield only approximate results. Dynamic analyses are mostly more complex but they generate more accurate results in terms of structural response. In addition, nonlinear analyses should also be used in fragility analyses due to the fact that damage is directly related to nonlinear behavior. Currently, two common analytical methods for generating fragility curves are the nonlinear pushover method and nonlinear dynamic methods. In addition, it is possible to perform analyses on idealized models of buildings as single- or multi-degree-of-freedom structures.

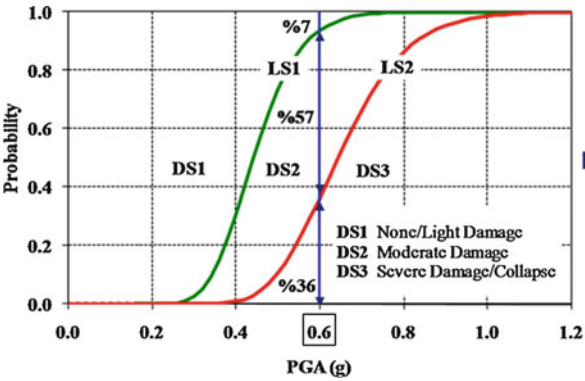
In this study, in order to compare the empirical and analytical DPMs, previously available fragility curves for masonry and reinforced concrete buildings are used (Erberik 2008a, b). These curves are derived using the analytical method described previously, and they present the local characteristics of existing reinforced concrete and masonry buildings in Turkey. They are also validated against real data from

past major earthquakes. The reinforced concrete buildings are classified as “low rise” (1–3 stories) and “mid rise” (4–9 stories). These buildings are also classified with respect to their compliance to current seismic codes and their performance levels as “high,” “moderate,” and “low.” Thus, there are six subclasses for the buildings. Typical fragility curves for the reinforced concrete buildings are shown in Fig. 2. The MR abbreviation in the figure indicates “mid-rise” buildings while the letters A, B, and C represent the high, moderate, and low seismic performance levels. The masonry buildings, on the other hand, are classified as “urban engineered,” “urban non-engineered,” and “rural.” The corresponding abbreviations, respectively, are “UE,” “UN,” and “RN.” In addition to these, number of stories is another structural parameter used in the classification of masonry buildings. Figure 3 displays an example fragility curve for two-story urban-engineered and urban-non-engineered masonry buildings (UE2, UN2). For harmonization purposes, two limit states are selected: Lower limit state (LS1) is between no damage and moderate damage states, while the upper limit state (LS2) is between moderate damage and heavy damage/collapse states.

The information on fragility curves can be easily converted to DPMs. In the example shown in Fig. 4, the fragility curves indicate the exceedance probability of two different limit states (LS1, LS2) in terms of a selected ground



Estimation of Potential Seismic Damage in Urban Areas, Fig. 3 Fragility curves of 2-story urban masonry buildings (Erberik 2008a)



Damage State (DS)		PGA=0.6g	
None/Light		0.07	
Moderate		0.57	
Severe/Collapse		0.36	

Estimation of Potential Seismic Damage in Urban Areas, Fig. 4 Transformation of the information in a fragility curve into a damage probability matrix

motion intensity parameter (peak ground acceleration, PGA in this case). The regions surrounding the fragility curves correspond to three different damage states (DS1, DS2, and DS3). To obtain the discrete probability of a damage state given an intensity level (PGA = 0.6 g in this case), a vertical line is drawn which intersects the fragility curves as shown in Fig. 4. This vertical line yields the probability of the building type of interest to experience different damage states under the corresponding PGA. In the example shown in Fig. 4, when PGA = 0.6 g, the probability of no damage (or light damage) is 7 %, the probability of moderate damage is 57 %, and the probability of severe damage (or collapse) is 36 %. This information constitutes one column of the DPM. If the same

procedure is repeated for other intensity levels, it is possible to form a full DPM.

Comparison of Results from Alternative Methods

As discussed previously, it is possible to obtain damage state probabilities with alternative approaches. In this study, in order to make a comparison with the empirical results, the analytical fragility curves complying with the characteristics of the existing building stock in Kocaeli, Sakarya, and Bolu are employed. Considering the local and structural-deficient characteristics of buildings in these city centers, fragility curves for “non-engineered” buildings are used.

Estimation of Potential Seismic Damage in Urban Areas, Table 5 Comparison of damage probabilities for masonry buildings in Kocaeli (MMI = IX)

Damage state	CDR (%)	Analytical approach	Empirical approach
No damage + light damage	2.5	0.90	0.81
Moderate damage	30	0.08	0.08
Severe damage + collapse	85	0.03	0.11
MDR (%)		6.89	13.78

Estimation of Potential Seismic Damage in Urban Areas, Table 7 Comparison of damage probabilities for masonry buildings in Sakarya (MMI = X)

Damage state	CDR (%)	Analytical approach	Empirical approach
No damage + light damage	2.5	0.75	0.89
Moderate damage	30	0.16	0.04
Severe damage + collapse	85	0.09	0.07
MDR (%)		14.60	9.38

Estimation of Potential Seismic Damage in Urban Areas, Table 6 Comparison of damage probabilities for reinforced concrete buildings in Kocaeli (MMI = IX)

Damage state	CDR (%)	Analytical approach	Empirical approach
No damage + light damage	2.5	0.15	0.46
Moderate damage	30	0.85	0.29
Severe damage + collapse	85	0.00	0.25
MDR (%)		25.9	31.1

Estimation of Potential Seismic Damage in Urban Areas, Table 8 Comparison of damage probabilities for reinforced concrete buildings in Sakarya (MMI = X)

Damage state	CDR (%)	Analytical approach	Empirical approach
No damage + light damage	2.5	0.03	0.34
Moderate damage	30	0.58	0.25
Severe damage + collapse	85	0.39	0.41
MDR (%)		50.8	43.2

The fragility curves given in Erberik (2008a, b) express the earthquake intensity in terms of PGA and peak ground velocity (PGV) for masonry and reinforced concrete buildings, respectively (Figs. 2 and 3). In order to make a comparison with the empirical damage state probabilities in Kocaeli, Sakarya, and Bolu, the PGA and PGV values recorded during the 17 August 1999 earthquake are used in the fragility curves. An important point here is the fact that there are different damage states in the analytical and empirical assessments: Thus, for a consistent comparison, the “none” and “light” damage states in the empirical DPMs are combined. Tables 5, 6, 7, 8, 9, and 10, respectively, present the comparison of two approaches in Kocaeli, Sakarya, and Bolu for masonry and reinforced concrete buildings.

According to the comparisons, the empirical and analytical DPMs yield similar results for masonry buildings in all city centers. However, the same observation is not valid for reinforced

concrete buildings: The discrete values of empirical and analytical damage state probabilities seem different. However, the MDR values are found to be consistent with each other for all building types at all locations.

The difference between empirical and analytical methods is indeed not surprising. Both methods involve inherent uncertainties: In the empirical approach, the values are directly dependent on the accuracy of the data collected in the field. Damage data collection in the aftermath of a large earthquake should be handled with care otherwise subjectivities arise in the resulting damage state probabilities. An example is conducting street surveys only in regions of severe damage while neglecting the areas with relatively “light” or “no damage” buildings, which would bias the resulting damage state probabilities.

In the analytical method, on the other hand, the accuracy of the results mostly relies on the

Estimation of Potential Seismic Damage in Urban Areas, Table 9 Comparison of damage probabilities for masonry buildings in Bolu (MMI = VII)

Damage state	CDR (%)	Analytical approach	Empirical approach
No damage + light damage	2.5	0.91	0.92
Moderate damage	30	0.06	0.03
Severe damage + collapse	85	0.02	0.05
MDR (%)		6.15	7.45

Estimation of Potential Seismic Damage in Urban Areas, Table 10 Comparison of damage probabilities for reinforced concrete buildings in Bolu (MMI = VII)

Damage state	CDR (%)	Analytical approach	Empirical approach
No damage + light damage	2.5	0.46	0.36
Moderate damage	30	0.53	0.36
Severe damage + collapse	85	0.01	0.27
MDR (%)		17.9	34.65

structural analysis approach taken or the idealizations of the exact buildings. For instance, the fragility curves for masonry buildings employed in this study are derived using three-dimensional building models; however, the curves for reinforced concrete buildings are based on equivalent single-degree-of-freedom models. Due to the equivalent idealized models, the analyses are simplified but on the other hand some accuracy is lost. This point is believed to be the main cause of difference between empirical and analytical results for the reinforced concrete buildings. This trade-off between cost of computing and cost of modeling real buildings more accurately is always a choice that a researcher must make.

Summary

In this study, observed damage state probabilities of masonry and reinforced concrete buildings in Turkey are compared with the damage state

probabilities obtained from analytical fragility curves. The damage data is collected in three city centers, which are all located in the most active seismic zone of Turkey, after the 17 August 1999 Kocaeli earthquake.

As a result of the comparisons in these city centers, in some cases very close results are obtained despite some noticeable differences in other cases. In particular, the use of equivalent simplified models of reinforced concrete buildings resulted in deviations from the empirical results. Such differences are expected as both approaches involve inherent uncertainties due to assumptions and simplifications made. However, in all cases, for all building types, the MDR values are found to be in the same order of magnitude. This observation is encouraging to use both approaches in the future.

According to the results of this case study, it is suggested that alternative methods are employed for estimating potential seismic damage in urban regions. Hybrid damage models are also recommended as they yield more accurate results. In the long run, results of this study and similar studies can be used in seismic loss estimation models, disaster management, emergency planning, as well as rapid response purposes in urban regions.

Cross-References

- [Damage to Buildings: Modeling](#)
- [Empirical Fragility](#)
- [Seismic Loss Assessment](#)

References

- Erberik MA (2008a) Generation of fragility curves for Turkish masonry buildings considering in-plane failure modes. *Earthq Eng Struct Dyn* 37:387–405
- Erberik MA (2008b) Fragility-based assessment of typical mid-rise and low-rise RC buildings in Turkey. *Eng Struct* 30:1360–1374
- Gürpınar A, Abalı M, Yüçemen MS, Yeşilçay Y (1978) Compliance with compulsory earthquake insurance. Earthquake Engineering Research Institute Report, 78–05. Middle East Technical University, Ankara (in Turkish)

- Kafali C, Grigouri M (2004) Seismic fragility analysis. In: Proceedings of the 9th ASCE specialty conference on probabilistic mechanics and structural reliability, Rome, 19–23 June 2005
- Özmen B (1999) Iso-intensity map of İzmit earthquake. Earthquake Research Department, General Directorate of Disaster Affairs, Ankara
- Whitman RV (1973) Damage probability matrices for prototype buildings, vol 380, Structures Publication. MIT, Boston

European Structural Design Codes: Seismic Actions

Michael N. Fardis

Department of Civil Engineering, School of Engineering, University of Patras, Patras, Greece

Synonyms

Design response spectra; Design seismic actions; Eurocode for seismic design, Seismic actions, Seismic response spectra

Introduction

Between 1992 and 1998, close to 60 Eurocodes were published by the European Committee for Standardization (CEN) as European pre-standards for structural design (ENVs). Their conversion to the first generation of European Standards (EN) for structural design started at the end of the 1990s and was concluded after the mid-2000s. The conversion entailed a thorough revision of the ENVs on the basis of recent scientific/technical developments and of comments from National Standards Bodies which are members of CEN.

Among the ten EN-Eurocodes, two cover the basis of structural design and the loadings (“actions”), one deals with geotechnical and foundation design, while five cover aspects specific to concrete, steel, composite (steel-concrete), timber, masonry, or aluminum construction. Instead of distributing the seismic design aspects to the EN-Eurocodes on

European Structural Design Codes: Seismic Actions, Table 1 Eurocode 8 parts

Part	EN	Title
1	EN 1998-1:2004	Design of structures for earthquake resistance – general rules, seismic actions, rules for buildings
2	EN 1998-2:2005	Design of structures for earthquake resistance – bridges
3	EN 1998-3:2005	Design of structures for earthquake resistance – assessment and retrofitting of buildings
4	EN 1998-4:2006	Design of structures for earthquake resistance – silos, tanks, pipelines
5	EN 1998-5:2004	Design of structures for earthquake resistance – foundations, retaining structures, geotechnical aspects
6	EN 1998-6:2005	Design of structures for earthquake resistance – towers, masts, chimneys

loadings, materials, or geotechnical design, all aspects of seismic design are covered by EN-Eurocode 8: “EN1998: Design of Structures for Earthquake Resistance.” This is for the convenience of countries with very low seismicity, as it gives them the option not to apply EN-Eurocode 8 at all.

EN-Eurocode 8 has six Parts, listed in Table 1.

The EN-Eurocodes are the recommended European structural design codes for civil engineering works and their parts (including prefabricated structural components), in order to facilitate the integration of the construction market in the European Union. The European Commission has published a “Recommendation on the implementation and use of Eurocodes for construction works and structural construction products” (European Commission 2003), according to which Member States should adopt the Eurocodes as a suitable tool for the structural design of construction works and refer to them in their national provisions on structural construction products. More specifically, the Eurocodes should be used as the basis for the technical specifications in the contracts for works and related engineering services in the public sector, as well as in the water, energy, transport, or telecommunications sectors.

According to (European Commission 2003), it is up to a Member State to select the level of safety and protection – which, according to (European Commission 2001), may include serviceability and durability – offered by civil engineering works on its national territory. To allow Member States to exercise this authority and accommodate geographical, climatic, and geological (including seismotectonic) differences without sacrificing harmonization of structural design codes at the European level, the notion of “Nationally Determined Parameters” (NDPs) has been adopted as the means to provide the necessary flexibility in the application of Eurocodes across (and outside) Europe. According to this approach, the EN-Eurocodes allow national choice of all key parameters or aspects that control the safety, durability, serviceability, and economy of civil engineering works designed and constructed to the EN-Eurocodes. The “Nationally Determined Parameters” in EN-Eurocodes include:

- Symbols (e.g., safety factors, the mean return period of the design seismic action, etc.)
- Technical classes (e.g., ductility classes)
- Procedures or methods (e.g., alternative models of calculation)

Alternative classes and procedures/methods considered as NDPs are identified and fully described in the normative text of the EN-Eurocode. For NDP-symbols, the EN-Eurocode may give a range of acceptable values and normally recommends in a nonnormative note a value for the symbol. It normally also recommends a class or a procedure/method among the alternatives identified and described in the EN-Eurocode text as NDPs (European Commission 2001).

Each Member State publishes its chosen values for the NDPs in a National Annex which accompanies the national version of the EN-Eurocode. According to European Commission (2003), Member States should adopt for the NDPs the choices recommended in the notes of the EN-Eurocode, in order to achieve the maximum feasible harmonization across the

EU. Member States should not diverge from the recommended choice, unless geographical, climatic, and geological differences or different levels of protection make it necessary. A National Annex may also contain country-specific data that are also considered by the EN-Eurocode as NDPs (e.g., a seismic zoning map, spectral shapes for the various types of soil profiles provided in Eurocode 8, etc.). A decision to adopt nationally or not an informative Annex of the EN-Eurocode is also stated there.

If the National Annex does not exercise national choice for some NDPs, the choice will be the responsibility of the designer, taking into account the conditions of the project and other national provisions (European Commission 2001).

A National Annex may provide supplementary information, noncontradictory to any of the rules of the EN-Eurocode. This may include references to other national documents, to assist the user in the application of the EN (European Commission 2001). What a National Annex cannot do is modify any EN-Eurocode provisions or replace them with other – e.g., national – rules. Although not encouraged, such deviations from the EN-Eurocode are allowed in national regulations other than the National Annex. However, when use is made of national regulations allowing deviation from certain Eurocode rules, the design will not be called “design according to EN-Eurocodes,” as by definition this term means compliance with all EN-Eurocode provisions, including the national choices for the NDPs (European Commission 2001).

If an EN-Eurocode is not considered relevant to a particular Member State, then a National Annex is not required. This may be the case for EN-Eurocode 8 in countries of very low seismicity (European Commission 2001).

CEN Member States have been given few years after the publication of an EN-Eurocode by CEN in English, French, and German to publish the corresponding National Annex (including the national choice for the NDPs). During that period, Member States have calibrated the NDPs so that, for the target safety level, structures designed according to the National version of the EN-Eurocodes will not cost significantly more

than those designed according to current National Standards (European Commission 2001).

CEN Member States have withdrawn those National Standards which conflict in some way with any EN-Eurocode, so that the EN-Eurocodes become the exclusive structural design standards in the European Union.

The revision of the EN-Eurocodes toward their second generation is expected to be complete by 2020, after starting in 2014.

Performance Requirements in EN-Eurocode 8 and Corresponding Seismic Hazard Levels

Seismic Design of New Buildings

Part 1 of EN-Eurocode 8 (CEN 2004a) provides for two-tier seismic design of new buildings, with the following explicit performance objective (Fardis et al. 2005):

- Under a rare seismic action: to protect life, by preventing collapse of the structure or its parts and by maintaining structural integrity and residual load capacity
- Under a frequent event: to reduce property loss by limiting structural and nonstructural damage or – for certain types of structures, such as tanks, pipelines or chimneys – to ensure integrity or a minimum operating level

The no-(local)-collapse performance level is achieved by dimensioning and detailing structural elements for a combination of strength and ductility that provides a safety factor between 1.5 and 2 against substantial loss of gravity load capacity and lateral load resistance. The damage limitation performance level is achieved by limiting the local or global deformations (normally the lateral displacements) of the building to levels acceptable for the integrity of all its parts (including nonstructural ones).

Although not explicitly stated, an additional performance objective in the design for ductility and energy dissipation is to prevent global collapse due to an extremely strong earthquake, like the “Maximum Considered Earthquake” (MCE)

of the US codes. Nonetheless, collapse may be imminent after such an earthquake (e.g., due to strong aftershocks) and repair may be unfeasible or economically prohibitive. Tolerance to extremely strong seismic actions is pursued through systematic and across-the-board application of the capacity design concept, which allows control of the inelastic response mechanism.

The two explicit performance levels – (local-) collapse prevention and damage limitation – are pursued under two different seismic actions. The seismic action under which (local-)collapse should be prevented is termed “design seismic action.” Given that issues of safety and economy are within the competence of individual Member States, the hazard levels for the seismic actions associated with these two performance levels – (local-)collapse prevention and damage limitation – are left open for national determination as NDPs. For buildings of ordinary importance, the recommendation in Part 1 of EN-Eurocode 8 is:

- A “design seismic action” (for local-collapse prevention), with recommended 10 % probability of been exceeded in 50 years (recommended mean return period: 475 years)
- A seismic action for damage limitation, with recommended 10 % probability of been exceeded (recommended mean return period: 95 years).

Seismic Assessment and Retrofitting of Existing Buildings

Part 3 of EN-Eurocode 8 (CEN 2005a) adopts a fully performance-based approach for the assessment and retrofitting of existing buildings (Fardis 2009). It defines three performance levels (called therein “limit states”):

- “Near collapse” (NC), which is similar to “collapse prevention” in the USA: The structure is heavily damaged, may have large permanent drifts, and retains little residual lateral strength or stiffness, but vertical elements can still carry the gravity loads. In the verifications, members may approach their ultimate force or deformation capacity.

- “Significant damage” (SD), corresponding to “life safety” in the USA and to the (local-) collapse prevention performance level to which new buildings are designed according to Part 1 of EN-Eurocode 8 (CEN 2004a). The structure is significantly damaged, may have moderate permanent drifts, but retains some residual lateral strength and stiffness and its full vertical load-bearing capacity. Repair may be uneconomic. The verifications should provide a margin against member ultimate capacities.
- “Damage limitation” (DL), corresponding to “immediate occupancy” in the USA. The structure has no permanent drifts; its elements have no permanent deformations, retain fully their strength and stiffness and do not need repair. Members should be verified to remain elastic.

The “seismic hazard” levels for which the three limit states are required are meant to be decided nationally as NDPs, or by the owner, if not fixed in the National Annex. Part 3 of EN-Eurocode 8 itself gives no recommendation, but mentions that the performance objective suitable to ordinary new buildings is a 225-year earthquake (20 % probability of been exceeded in 50 years), a 475-year event (10 % probability in 50 years), or a 2,475-year one (2 % in 50 years), for the DL, the SD, or the NC “limit state,” respectively. National Authorities also decide whether all three limit states need to be verified or whether checking one or two limit states at the corresponding seismic hazard level suffices. Part 3 of EN-Eurocode 8 guides National Authorities to choose the set of limit states to be checked, as well as the return period of the corresponding seismic actions, depending on whether assessment and retrofitting takes place as part of an “active” program (in which owners of certain categories of buildings are required to meet specific deadlines for completion of the assessment and – depending on its outcome – of the retrofitting) or a “passive” one (that associates seismic assessment – possibly leading to retrofitting – with discontinued use of the building). In “active” programs the relevant requirements may be less stringent than in

“passive” ones, while in “passive” programs triggered by remodelling the building, performance requirements for the assessment and the retrofitting may graduate with the extent and cost of the remodelling work undertaken. With this guidance, National Authorities will hopefully set the performance requirements for existing buildings so that, in the long run, the percentage of the existing building stock to be retrofitted is optimal for the society and the national economy as a whole.

Seismic Design of New Bridges

Part 2 of EN-Eurocode 8 (CEN 2005b) requires single-tier seismic design of new bridges with the following explicit performance objective:

- The bridge must retain its structural integrity and have sufficient residual resistance to be used for emergency traffic without any repair after a rare seismic event: the “design seismic action”; any damage due to this event must be easily repairable.

Although called “non-collapse requirement,” in reality this corresponds to life safety, rather than to near collapse, because sufficient residual resistance has to be available after the design seismic event for immediate use by emergency traffic (Kolias et al. 2012).

Part 2 of EN-Eurocode 8 (CEN 2005b) calls also for limitation of damage under a loosely defined seismic action with high probability of been exceeded in the bridge’s lifetime; such damage must be minor and limited only to secondary components and to the parts of the bridge intended for controlled damage under the “design seismic action.” However, this requirement is of no practical consequence for design: it is presumed to be implicitly met, if all the criteria for compliance with the “non-collapse requirement” are checked and met under the “design seismic action” (Kolias et al. 2012). This should be contrasted with new buildings, for which Part 1 of EN-Eurocode 8 (CEN 2004a) provides explicit checks under a well-defined “damage limitation” seismic action. However, the damage checks (inter-story drifts)

normally refer to nonstructural elements, which are not present in bridges.

Elastic Response Spectra

Introduction: Elastic Spectra

The seismic action is considered to impart concurrent translational motion in three orthogonal directions: the vertical, and two horizontal ones, at right angles to each other. The motion is taken to be applied at the interface between the structure and the ground.

The most common representation of the seismic action in structural design codes is via the response spectrum of an elastic Single-Degree-of-Freedom (SDOF) oscillator with 5 % viscous damping ratio. It gives the maximum response of this oscillator to the ground motion, as a function of its natural period T (hence the term “spectrum”). Any other alternative representation of the seismic action (e.g., in the form of acceleration time-histories) should conform to the 5 %-damped elastic response spectrum. This is indeed the representation chosen in EN-Eurocode 8 (CEN 2004a, b, 2005a, b, c, 2006).

Because:

- earthquake ground motions are traditionally recorded as acceleration time-histories.
- seismic design is still based on forces, conveniently derived from accelerations;

the pseudo-acceleration response spectrum, $S_a(T)$, is normally used. If spectral displacements, $S_d(T)$, are of interest (notably, for displacement-based design, assessment or retrofitting), they can be obtained from $S_a(T)$ assuming simple harmonic oscillation:

$$S_d(T) = (T/2\pi)^2 S_a(T) \quad (1)$$

Spectral pseudo-velocities can also be obtained from $S_a(T)$ as:

$$S_v(T) = (T/2\pi) S_a(T) \quad (2)$$

Pseudo-values do not correspond to the real peak spectral velocity or acceleration. For

damping ratio up to 10 % and for natural period T between 0.2 and 1.0 s, the pseudo-velocity spectrum approximates well the actual relative velocity spectrum.

The EN-Eurocode 8 spectra include ranges of:

- Constant spectral pseudo-acceleration for natural periods between two periods denoted by T_B and T_C
- Constant spectral pseudo-velocity between periods T_C and T_D
- Constant spectral displacement for periods longer than T_D

Design Ground Acceleration: Importance Classes

In EN-Eurocode 8, the elastic response spectrum is taken as proportional (“anchored”) to the peak acceleration of the ground, namely, to:

- The horizontal peak acceleration, a_g , for the horizontal component(s) of the seismic action or
- The vertical peak acceleration, a_{vg} , for the vertical component

Data from Europe available at the time Eurocode 8 was drafted could not support the dependence of the elastic spectrum on additional parameters. According to (Bommer and Scott 2000) and (Rey et al. 2002), the shape of elastic response spectra of European records does not show a clear dependence on distance – epicentral or to the fault.

The basis of the seismic design of new structures in EN-Eurocode 8 is the “design seismic action,” for which the no-(local-)collapse requirement should be met. The “design seismic action” is specified through the “design ground acceleration” in the horizontal direction, a_g , which is equal to the “reference peak ground acceleration” on rock from national zonation maps, times the importance factor, γ_I , of the structure:

$$a_g = \gamma_I a_{gR} \quad (3)$$

(see below; for structures of ordinary importance, $\gamma_I = 1.0$).

Within a single-tier design framework, enhanced safety of facilities that are essential or have large occupancy is normally achieved by modifying the hazard level (the mean return period) of the “design seismic action.” To this end, the seismic action is multiplied by an “importance factor,” γ_I .

By definition, $\gamma_I = 1.0$ for structures of ordinary importance (called “Importance Class II” in Parts 1, 2 and 3 of EN-Eurocode 8 (CEN 2004a, 2005a, b)).

Concerning buildings, Part 1 of EN-Eurocode 8 (CEN 2004a) recommends a value $\gamma_I = 1.2$ for those whose collapse may have unusually large social or economic consequences (large occupancy buildings, such as schools or public assembly halls) or those housing institutions of cultural importance, such as museums (buildings of “Importance Class III” in EN-Eurocode 8); $\gamma_I = 1.4$ is recommended for buildings which are essential for civil protection during the immediate post-earthquake period, categorized as “Importance Class IV” (hospitals, fire or police stations, power plants, etc.); for buildings of minor importance for public safety (i.e., belonging to “Importance Class I,” comprising agricultural and similar buildings), Part 1 of EN-Eurocode 8 recommends a value $\gamma_I = 0.8$.

Concerning bridges, Part 2 of EN-Eurocode 8 (CEN 2005b) recommends classifying under “Importance Class III” the ones which are crucial for communications especially in the immediate post-earthquake period (including access to emergency facilities), or whose downtime may have a major economic or social impact, or which, by failing, may cause large loss of life, as well as major bridges with a target design life longer than the ordinary nominal life of 50 years. For this importance class, it recommends taking $\gamma_I = 1.3$. Bridges which are not critical for communications, or considered not economically justified to design for the standard bridge design life of 50 years, should be classified under “Importance Class I” with a recommended value $\gamma_I = 0.85$, according to Part 2 of EN-Eurocode 8.

The importance classification of towers, masts, and chimneys (CEN 2005c) is similar to that of buildings: of reduced importance for

public safety, of ordinary importance, of higher importance for public safety in terms of potential casualties, or vital for civil protection; the recommended values of the NDP-importance factor γ_I are the same as for buildings.

The important classification of tanks, pipelines, or silos (CEN 2006) is more complex, taking into account not only risk to life but also the economic and social consequences of damage or failure, as well as system and redundancy aspects. The recommended values of the NDP-importance factor γ_I are equal to 0.8, 1.0, 1.2, or 1.6, if the risk to life and the economic or social consequences of failure are low, medium, high, or exceptional, respectively.

The “reference peak ground acceleration” corresponds to the reference return period, T_{NCR} , of the “design seismic action” for structures of ordinary importance (“Importance Class II”). It is reminded that under the Poisson assumption of earthquake occurrence (i.e., that the number of earthquakes in an interval of time depends only on the length of the interval in a time-invariant way), the return period, T_R , of seismic events exceeding a certain threshold is related to the probability, P , that the threshold will be exceeded in T_L years as

$$T_R = -T_L / \ln(1 - P) \quad (4)$$

So, for a given T_L (e.g., for the reference design life of $T_{LR} = 50$ years), the seismic action may equivalently be specified either via its mean return period, T_R , or its probability of been exceeded in T_L years, P_R . Values of the importance factor greater or lower than 1.0 correspond to mean return periods longer or shorter, respectively, than T_{NCR} . It is in the authority of the country to select the value of T_{NCR} that gives the appropriate trade-off between economy and public safety in its territory, as well as the importance factors for building other than ordinary, taking into account the specific regional features of the seismic hazard. Part 1 of EN-Eurocode 8 (CEN 2004a) recommends the value $T_{NCR} = 475$ years.

The mean return period, $T_R(a_g)$, of a peak ground acceleration exceeding a value a_g is the

inverse of the annual rate, $\lambda_a(a_g)$, at which this acceleration level is exceeded:

$$T_R(a_g) = 1/\lambda_a(a_g) \quad (5)$$

A functional form commonly used for $\lambda_a(a_g)$ is

$$\lambda_a(a_g) = K_o(a_g)^{-k} \quad (6)$$

If the exponent k (: slope of the “hazard curve” $\lambda_a(a_g)$ in a log-log plot) is about constant, two peak ground acceleration levels a_{g1} , a_{g2} , corresponding to two different mean return periods, $T_R(a_{g1})$, $T_R(a_{g2})$, are related as

$$\frac{a_{g1}}{a_{g2}} = \left(\frac{T_R(a_{g1})}{T_R(a_{g2})} \right)^{1/k} \quad (7)$$

The value of k characterizes the seismicity of the site. In regions where the difference in peak ground acceleration of frequent and very rare seismic excitations is very large, k is low (around 2); in such regions, full performance-based design or assessment at several performance levels with widely different hazard levels is very meaningful. Large values of k (above 4) are typical of regions where high ground acceleration levels are almost as frequent as smaller ones; one performance level (normally the one associated with the lowest among the hazard levels) will almost always govern there; performance-based design or assessment at the other levels may be redundant.

The value of the importance factor γ_I may be determined as follows: let's suppose that, for “Importance Class II” (ordinary), the “reference peak ground acceleration” on rock, a_{gR} , has a probability, P_{LR} , of been exceeded in the reference design life of $T_{LR} = 50$ years; then, the “design ground acceleration,” a_g , which has that probability, P_{LR} , of been exceeded in T_L years is obtained by multiplying a_{gR} by

$$\gamma_I = (T_L/T_{LR})^{1/k} \quad (8a)$$

Alternatively, to have a probability of exceeding the “design ground acceleration,” P_L , in T_{LR}

years, which is different from the reference probability, P_{LR} , in the same T_{LR} years, a_{gR} should be multiplied by

$$\gamma_I = (P_{LR}/P_L)^{1/k} \quad (8b)$$

Ground Acceleration of Seismic Actions Other Than the “Design Seismic Action”

Design for damage limitation under the corresponding seismic action is explicitly required for buildings (as well as for tanks, pipelines, and chimneys, but not for bridges). The ratio, v , of the PGA of the seismic action for damage limitation to the “design seismic action” (for local-collapse prevention) reflects the difference in hazard levels. It is an NDP, with recommended values for buildings, silos, tanks, and pipelines equal to 0.5 for Importance Classes I and II or 0.4 for III and IV. This gives a ground acceleration, $(v\gamma_I)a_{gR}$, equal to 40 %, 50 %, 48 %, and 56 % of a_{gR} , for Importance Classes I, II, III, and IV, respectively; in other words, the property protection is about the same for ordinary, large occupancy, or essential facilities, although the protection they offer against collapse may be very different.

The approach highlighted at the end of the last section for the estimation of the importance factor, γ_I , on the basis of the Poisson assumption of earthquake occurrence may be modified to determine the ratio, v , of the ground acceleration of the damage limitation seismic action, va_g , to the “design ground acceleration,” a_g . If the seismic action for damage limitation has a probability of been exceeded in T_{DL} years which is the same as that of exceeding the “design seismic action” in a design life of T_L years, then a_g should be multiplied by

$$v = (T_{DL}/T_L)^{1/k} \quad (9)$$

EN-Eurocode 8 recommends in a note a probability of 10 % for exceeding the damage limitation seismic action in $T_{DL} = 10$ years. For these recommended values, Eq. 4 gives a mean return period of 95 years. The value of v should vary with the exponent k ; in order for the recommended values of v for buildings (equal to

0.4 or 0.5) to be consistent with the other recommended values, k should be around 2.

If the construction phase is long and in its duration the structural system is much more vulnerable to earthquakes than after completion (as, e.g., in segmental construction), the owner may specify seismic performance requirements during construction and the associated compliance criteria. Equation 4 may then be applied, using the estimated duration of construction, T_c , as T_L , to give the mean return period of the seismic action which has the target probability of been exceeded, P (e.g., $P = 0.05$). The resulting mean return period may be used as $T_R(a_{g1})$ in Eq. 7 to find the peak ground acceleration, a_{g1} , with probability P of been exceeded during construction. In that case, $a_{g2} = a_{gR}$ and $T_R(a_{g2}) = T_{NCR}$ (Kolias et al. 2012).

EN-Eurocode 8 adopts the same spectral shape for the different seismic actions to be used at different performance levels or purposes. The difference in the hazard level is reflected only through the peak ground acceleration to which the spectrum is anchored.

Elastic Response Spectra of the Horizontal Seismic Action Components in EN-Eurocode 8

The elastic response spectral acceleration, $S_a(T)$, due to a single horizontal component of the seismic action is described in EN-Eurocode 8 by the following expressions:

In the “short-period range”

$$0 \leq T \leq T_B : S_a(T) = a_g S \left[1 + \frac{T}{T_B} (2.5\eta - 1) \right] \quad (10a)$$

In the “constant spectral pseudo-acceleration range”:

$$T_B \leq T \leq T_C : S_a(T) = a_g S \cdot 2.5\eta \quad (10b)$$

In the “constant spectral pseudo-velocity range”:

$$T_C \leq T \leq T_D : S_a(T) = a_g S \cdot 2.5\eta \left[\frac{T_C}{T} \right] \quad (10c)$$

In the “constant spectral displacement range”:

European Structural Design Codes: Seismic Actions, Table 2 Values of viscous damping, ζ , for different structural materials

Material	Damping (%)
Reinforced concrete components	5
Prestressed concrete components	2
Welded steel components	2
Bolted steel components	4

$$T_D \leq T \leq 4 \text{ sec} : S_a(T) = a_g S \cdot 2.5\eta \left[\frac{T_C T_D}{T^2} \right] \quad (10d)$$

where

a_g is the design ground acceleration on rock, from Eq. 3

S is the “soil factor”

$$\eta = \sqrt{10/(5 + \zeta)} \geq 0.55 \quad (11)$$

is the correction factor per (Bommer and Elnashai 1999) for values of the viscous damping ratio, ζ , other than the reference value of 5 %. In fact, the value $\zeta = 5\%$ is considered as representative of cracked reinforced concrete structures. The values of ζ specified in Part 2 of Eurocode 8 (CEN 2005b) for components of various structural materials are listed in Table 2.

The terms “constant spectral pseudo-velocity range” and “constant spectral displacement range” used above in Eq. 10c and 10d, respectively, are a direct consequence of Eqs. 1 and 2, alongside the functional dependence of $S_a(T)$ on T in Eq. 10c and 10d.

By definition, $S = 1$ over rock (called “type A ground”). The value $a_g S$ plays the role of an “effective ground acceleration,” as the spectral acceleration at the constant spectral acceleration plateau is always equal to $2.5a_g S$. The “soil factor,” S , amplifies uniformly the entire spectrum by S over the spectrum for rock. This feature came out of the studies that provided the background for the Eurocode 8 spectra (Rey et al. 2002).

The values of the “corner” periods T_B , T_C , and T_D (i.e., the extent of the ranges of constant spectral pseudo-acceleration, pseudo-velocity, and displacement) and of the soil factor, S , are taken

European Structural Design Codes: Seismic Actions, Table 3 Ground types in EN-Eurocode 8 for the definition of the seismic action

	Description	$v_{s,30}$ (m/s)	N_{SPT}	c_u (kPa)
A	Rock outcrop, with less than 5 m of cover by weaker material	>800	–	–
B	Very dense sand or gravel, or very stiff clay, with depth of several tens of meters and mechanical properties gradually increasing with depth	360–800	>50	>250
C	Dense to medium-dense sand or gravel, or stiff clay, several tens to many hundreds meters deep	180–360	15–50	70–250
D	Loose-to-medium sand or gravel, or soft-to-firm clay	<180	<15	<70
E	Surface alluvium layer with $v_s < 360$ m/s and thickness from 5 to 20 m, underlain by rock (with $v_s > 800$ m/s)			
S_1	≥ 10 m thick soft clay or silt, with plasticity index >40 and high water content	<100	–	10–20
S_2	Liquefiable soils; sensitive clays or any soil not of type A to E or S_1			

to depend mainly on “ground type.” In the EN-Eurocodes, the term ground includes any type of soil, as well as rock. EN-Eurocode 8 recognizes five standard ground types, as well as two special ground types, S_1 , S_2 , for which there are no recommended values for T_B , T_C , T_D , or S . Note that in reality the value of S decreases with the intensity of the motion (as measured by a_g) owing to the increased nonlinearity of the soil, but this effect has been neglected in EN-Eurocode 8.

The five standard ground types, A, B, C, D, and E and the two special ones, S_1 and S_2 , per EN-Eurocode 8 are described in Table 3. The reference approach to characterize the ground in this table is through the average value of shear wave velocity, $v_{s,30}$, at the top 30 m:

$$v_{s,30} = \frac{30}{\sum_{i=1,N} \frac{h_i}{v_i}} \quad (12)$$

where h_i and v_i are, respectively, the thickness (in m) and the shear wave velocity at small shear strains (less than 10^{-6}) of the i -th layer in N layers. If the value of $v_{s,30}$ is not available, the Standard Penetration Test (SPT) blow-count number may be used for soil types B, C, or D, according to the correspondence of SPT to $v_{s,30}$ in (Ohta and Goto 1976). If neither the SPT number nor $v_{s,30}$ is available, the undrained cohesive resistance, c_u , may be used to characterize the soil.

Ground belonging to the two special types, S_1 and S_2 , deserves a special, site-specific study to

define the seismic action. For ground-type S_1 , the special study should take into account the thickness and the v_s -value of the soft clay or silt layer and the difference with the underlying materials and should quantify their effects on the elastic response spectrum. Note that soils of type S_1 may have low internal damping and exhibit linear behavior over a large range of strains, producing peculiar amplification of the bedrock motion and unusual or abnormal soil-structure interaction effects. The scope of the site-specific study should also address the possibility of soil failure under the design seismic action (especially at ground type S_2 deposits with liquefiable soils or sensitive clays).

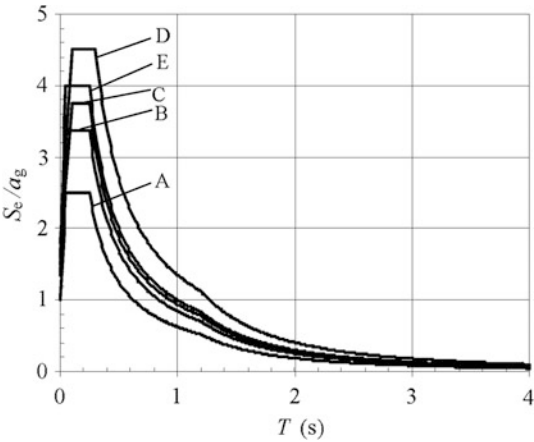
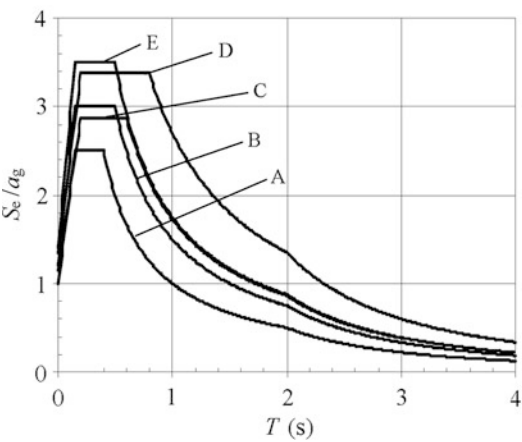
The values of T_B , T_C , T_D , and S for the five standard ground types A to E are meant to be defined by each country in the National Annex to Eurocode 8, depending on the magnitude of earthquakes contributing the most to the hazard. The geological conditions at the site may also be taken into account, in addition to the properties of the top 30 m of ground, to determine these values.

Instead of spectral amplification factors that decrease – probably through the S factor – with increasing design acceleration (spectral or ground) because of the soil nonlinearity effect (as in the US codes), the non-binding recommendation in a note in Part 1 of EN-Eurocode 8 is to have two types of spectra:

- Type 1: for moderate to large magnitude earthquakes

European Structural Design Codes: Seismic Actions, Table 4 Recommended parameter values for the standard horizontal elastic response spectra in EN-Eurocode 8

Ground type	Spectrum type 1				Spectrum type 2			
	S	T_B (s)	T_C (s)	T_D (s)	S	T_B (s)	T_C (s)	T_D (s)
A	1.00	0.15	0.40	2.00	1.00	0.05	0.25	1.20
B	1.20	0.15	0.50	2.00	1.35	0.05	0.25	1.20
C	1.15	0.20	0.60	2.00	1.50	0.10	0.25	1.20
D	1.35	0.20	0.80	2.00	1.80	0.10	0.30	1.20
E	1.40	0.15	0.50	2.00	1.60	0.05	0.25	1.20



European Structural Design Codes: Seismic Actions, Fig. 1 Five percent-damped horizontal elastic response spectra of type 1 (*left*) and 2 (*right*) recommended in EC8

for the five standard ground types of Table 3 (PGA on rock equal to 1 g)

- Type 2: for low magnitude ones (e.g., with surface magnitude less than 5.5) at close distance, producing motions rich in high frequencies over soft soils

The values of T_B , T_C , T_D , and S recommended in a non-binding note in Part 1 of EN-Eurocode 8 for the five standard ground types A to E are given in Table 4, giving the shapes depicted in Fig. 1. These values are based on (Rey et al. 2002) and European strong motion data (Ambraseys et al. 2000). There are certain regions in Europe (e.g., where the hazard is contributed mainly by strong, intermediate-depth earthquakes, as in the part of the eastern Balkans affected by the Vrancea region) where the two recommended spectral shapes may not be suitable. The lower S -values of type 1 spectra are due to the larger soil nonlinearity in the stronger ground motions

produced by moderate to large magnitude earthquakes. The recommended values of the period T_D at the outset of the constant spectral displacement region seems rather low. For flexible structures (e.g., those with seismic isolation) they may not lead to safe-sided designs. A safeguard against the rapid decay of the elastic spectrum for $T > T_D$ is provided by the lower bound of $0.2a_g S$ recommended in Part 1 of EN-Eurocode 8 for the design spectral accelerations (see Eq. 13c, 13d in section “Design Spectrum for Force-Based Design With Linear Analysis per EN-Eurocode 8”).

Simplified Design Rules, or Exemption from Applying EN-Eurocode 8, for Low Design Ground Acceleration

EN-Eurocode 8 says that its provisions need not be applied in “cases of very low seismicity,”

leaving to the National Annex to decide which combination of categories of structures, ground types, and seismic zones in a country qualify as “cases of very low seismicity.” It recommends (in a note) as a criterion for the “cases of low seismicity,” either the value of the design ground acceleration on type A ground (i.e., on rock), a_g , per Eq. 3, or the corresponding value, $a_g S$, over the ground type of the site. The recommended threshold values for the very low seismicity cases are 0.04 g for a_g or 0.05 g for $a_g S$. As the value of a_g includes the importance factor γ_I , certain structures in a region may be exempted from the application of EN-Eurocode 8, while others (those housing essential or large occupancy facilities) may not be. This is consistent with the notion that the exemption from the application of Eurocode 8 is due to the inherent lateral force resistance of any structure designed for non-seismic loadings, neglecting any contribution from the ductility and energy dissipation capacity. As EN-Eurocode 8 considers that, thanks to its overstrength, any structure is entitled to a behavior factor, q , of at least 1.5, the value of 0.05 g for $a_g S$, recommended as the threshold for very low seismicity cases, implies an assumed inherent lateral force capacity of $0.05 \times 2.5/1.5 = 0.083$ g, which is a reasonable assumption. If a National Annex states that the entire national territory is considered as a “case of very low seismicity,” then EN-Eurocode 8 does not apply at all in the country.

In cases of “low seismicity,” EN-Eurocode 8 permits simplified seismic design procedures: e.g., to design a concrete, steel, composite (steel-concrete), or masonry building to EN-Eurocodes 2, 3, 4, or 6, respectively, alone, using very few rules from Part 1 of EN-Eurocode 8 (CEN 2004a); to design a bridge for “limited ductile” behavior, with simplified rules defined in Part 2 of EN-Eurocode 8 (CEN 2005b); etc. Again it leaves to the National Annex to decide which combination of categories of structures, ground types, and seismic zones in a country belongs to the “cases of low seismicity” and recommends (in a note) employing in the criterion either the

value of the design ground acceleration on type A ground (i.e., on rock), a_g , or the corresponding value, $a_g S$, over the ground type of the site. Moreover, it recommends a value of 0.08 g for a_g , or of 0.10 g for $a_g S$, as the threshold for the low seismicity cases.

Elastic Response Spectra of the Vertical Seismic Action Component in EN-Eurocode 8

EN-Eurocode 8 requires designers to take into account the vertical component of the seismic action in design only in certain very well-prescribed situations.

- In buildings, only if the design ground acceleration in the vertical direction, a_{vg} , exceeds 0.25 g and even then only in the following cases:
 - In base-isolated buildings or
 - For (nearly) horizontal members (beams, girders, or slabs) which:
 - Have a span of at least 20 m or
 - Cantilever over at least 5 m or
 - Support directly columns or
 - Are prestressed
- In bridges, for the seismic design of:
 - Prestressed concrete decks (with the vertical component acting upwards)
 - Any bearings
 - Any seismic links
 - Piers in zones of high seismicity:
 - If the pier is already taxed by bending due to permanent actions on the deck or
 - The bridge is less than 5 km from an active seismotectonic fault (defined as one where the average historic slip rate is at least 1 mm/year and there is topographic evidence of seismic activity in the past 11,000 years); in that case, site-specific spectra which account for near-source effects should be used.
- For silos
- For tanks

Therefore, the practical importance of the vertical spectrum is less than that of the horizontal.

Consistent with the findings of (Bozorgnia and Campbell 2004), the vertical elastic response spectrum is defined independently of the horizontal response spectra, as follows:

In the “short-period range”:

$$0 \leq T \leq T_B : S_{a, \text{vert}}(T) = a_{\text{vg}} \left[1 + \frac{T}{T_B} (3\eta - 1) \right] \quad (13a)$$

In the “constant spectral pseudo-acceleration range”:

$$T_B \leq T \leq T_C : S_{a, \text{vert}}(T) = a_{\text{vg}} \cdot 3\eta \quad (13b)$$

In the “constant spectral pseudo-velocity range”:

$$T_C \leq T \leq T_D : S_{a, \text{vert}}(T) = a_{\text{vg}} \cdot 3\eta \left[\frac{T_C}{T} \right] \quad (13c)$$

In the “constant spectral displacement range”:

$$T_D \leq T \leq 4s : S_{a, \text{vert}}(T) = a_{\text{vg}} \cdot 3\eta \left[\frac{T_C T_D}{T^2} \right] \quad (13d)$$

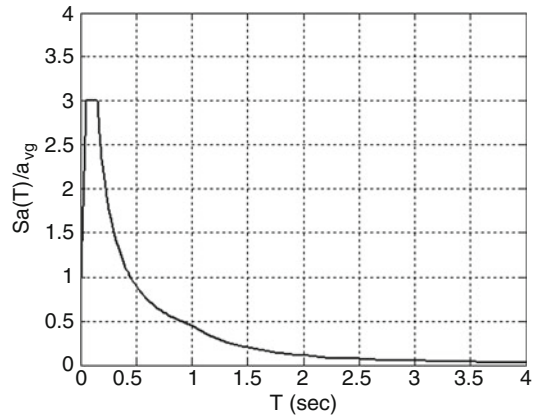
with η from Eq. 11.

The main differences between the horizontal and the vertical spectra are in:

- The amplification factor in the constant spectral pseudo-acceleration plateau, set equal to 3 instead of 2.5
- The lack of a uniform amplification of the entire spectrum due to the type of soil

Part 1 of EN-Eurocode 8 recommends in a note the following non-binding values of T_B , T_C , and T_D and of the design ground acceleration in the vertical direction, a_{vg} (see Fig. 2):

- $T_B = 0.05$ s.
- $T_C = 0.15$ s.
- $T_D = 1.0$ s.



European Structural Design Codes: Seismic Actions,
Fig. 2 Five percent-damped vertical elastic response spectrum recommended in EC8 for vertical PGA on rock $a_{\text{vg}} = 1$ g

- $a_{\text{vg}} = 0.9a_g$, if the type 1 spectrum applies to the site.
- $a_{\text{vg}} = 0.45a_g$, if the type 2 spectrum is used.

The vertical response spectrum recommended in Part 1 of EN-Eurocode 8 is based on work and data specific to Europe (Ambraseys and Simpson 1996; Elnashai and Papazoglou 1997).

The ratio a_{vg}/a_g is known to be higher at short distances (epicentral or to a causative fault). However, as distance does not enter as a parameter in the definition of the seismic action in EN-Eurocode 8, the type of spectrum has been chosen as the parameter determining this ratio, on the basis of the finding that a_{vg}/a_g increases also with magnitude (Ambraseys and Simpson 1996; Abrahamson and Litehiser 1989), which in turn determines the selection of the type of spectrum.

Design Spectrum for Force-Based Design with Linear Analysis per EN-Eurocode 8

Design Spectrum for the Horizontal Components

For the horizontal components of the seismic action, the design spectrum in Part 1 of

EN-Eurocode 8 for force-based design of new structures is:

In the “short-period range”:

$$0 \leq T \leq T_B : S_{a,d}(T) = a_g S \left[\frac{2}{3} + \frac{T}{T_B} \left(\frac{2.5}{q} - \frac{2}{3} \right) \right] \quad (14a)$$

In the “constant spectral pseudo-acceleration range,”

$$T_B \leq T \leq T_C : S_{a,d}(T) = a_g S \frac{2.5}{q} \quad (14b)$$

In the “constant spectral pseudo-velocity range,”

$$\begin{aligned} T_C \leq T \leq T_D : S_{a,d}(T) \\ = a_g S \frac{2.5}{q} \left[\frac{T_C}{T} \right], S_{a,d}(T) \geq \beta a_g \end{aligned} \quad (14c)$$

In the “constant spectral displacement range,”

$$\begin{aligned} T_D \leq T : S_{a,d}(T) \\ = a_g S \frac{2.5}{q} \left[\frac{T_C T_D}{T^2} \right], S_{a,d}(T) \geq \beta a_g \end{aligned} \quad (14d)$$

The value $2/3$ in Eq. 14a is the inverse of the overstrength factor of 1.5 considered in EN-Eurocode 8 to always be available, even without any design measures for ductility and energy dissipation. Factor β in Eq. 14c, 14d gives a lower bound for the horizontal design spectrum, acting as a safeguard against excessive reduction of the design forces due to the flexibility of the system (real or presumed in the design). Its recommended value in Part 1 of EN-Eurocode 8 is 0.2. Its practical implications may be particularly important, in view of the relatively low values recommended by Eurocode 8 for the corner period T_D at the outset of the constant spectral displacement range.

Design Spectrum in the Vertical Direction

The design spectrum in the vertical direction is obtained by substituting in Eqs. 14 the design ground acceleration in the vertical direction, a_{vg} , for the “effective ground acceleration” $a_g S$.

There is no clear, well-known energy dissipation mechanism for the response in the vertical direction. So, the “behavior factor” q in that direction is attributed to overstrength alone and is taken equal to 1.5, unless a higher value is supported by special studies and analyses.

Peak Ground Displacement and Displacement Response Spectrum

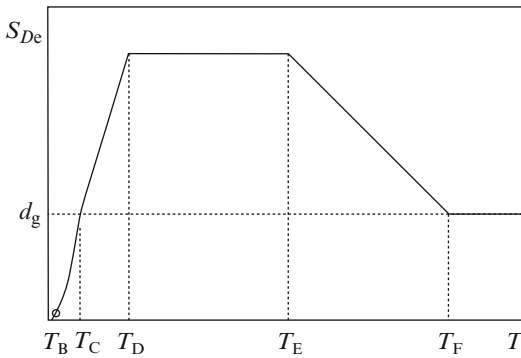
The value given in Part 1 of EN-Eurocode 8 for the design ground displacement, d_g , that corresponds to the design ground acceleration, a_g , is based on the assumption that the spectral displacement in the constant spectral displacement range (derived through Eq. 1 from the spectral acceleration at the beginning of that range at $T = T_D$, as given from Eq. 3 for the end of the constant pseudo-velocity range at $T = T_D$) is an amplification factor of 2.5 over the ground displacement that corresponds to the design ground acceleration, a_g . Taking $(T/2\pi)^2 \approx 40$, we obtain (for a_g in m/s^2 , not in g 's):

$$d_g = 0.025 a_g S T_C T_D \quad (15)$$

Equation 15 overestimates the ratio d_g/a_g , compared to more accurate predictions as a function of magnitude and distance based on (Ambraseys et al. 1996; Bommer and Elnashai 1999).

An Informative Annex in Part 1 of EN-Eurocode 8 gives also the displacement response spectrum, $S_d(T)$, including its long-period tail, beyond the upper end ($T = 4$ s) where Eq. 10d applies – alongside Eq. 1. More specifically (see Fig. 3):

- The applicability of Eq. 10d is extended beyond 4 s, up to a new control period $T_E > 4$ s.
- For $T < T_E$, $S_a(T)$, as given by Eqs. 10, may be converted to $S_d(T)$ through Eq. 1.
- Beyond T_E , $S_d(T)$ is taken to decrease linearly with T till a new control period T_F , where $S_d(T)$ becomes equal to d_g .



European Structural Design Codes: Seismic Actions,
Fig. 3 Shape of displacement response spectrum in
 EN-Eurocode 8

So, the tail of the displacement response spectrum is obtained as

$$T_E \leq T \leq T_F : S_d(T) = \left[2.5\eta + \left(\frac{T - T_E}{T_F - T_E} \right) (1 - 2.5\eta) \right] d_g \quad (16a)$$

$$T \geq T_F : S_d(T) = d_g \quad (16b)$$

with η from Eq. 11.

Values are given for the control periods, T_E and T_F , only for situations where the type 1 spectrum is considered to apply. Then, based on (Tolis and Faccioli 1999), T_E is taken as:

- $T_E = 4.5$ s for ground type A
- $T_E = 5.0$ s for ground type B
- $T_E = 6.0$ s for ground types C, D, or E

while T_F is always assumed equal to 10 s.

The above displacement response spectrum in EN-Eurocode 8 was supported by data from Europe (Bommer et al. 1998; Bommer and Elnashai 1999) at least up to periods of 3 s. However, as the European digital data from which the displacement response at longer periods can be reliably established were very limited at that time (Faccioli et al. 2004), they were supplemented with information from other regions of the world, for instance, from digital records from the 1995 Kobe earthquake, normally filtered at 20 s (Tolis and Faccioli 1999). This is one reason for not

giving values for T_E and T_F in situations where the type 2 spectrum is considered appropriate.

Topographic and Near-Source Effects on the Seismic Action

EN-Eurocode 8 provides for topographic amplification (ridge effect, etc.) of the seismic action for all types of structures. Such effects have been indeed identified in past earthquakes in Italy (Faccioli et al. 2002; Paolucci 2002, 2006). According to Part 1 of EN-Eurocode 8, topographic amplification of the full elastic spectrum is mandatory for structures of importance above ordinary. An Informative Annex in Part 5 of EN-Eurocode 8 (CEN 2004b) recommends amplification factors equal to 1.2 over isolated cliffs or long ridges with slope (to the horizontal) less than 30° or to 1.4 at ridges steeper than 30°.

EN-Eurocode 8 does not have in Part 1 general provisions for near-source effects. Indeed, directivity effects of the ground motion along the rupture propagation direction are common near the seismotectonic fault rupture, producing at the site a large velocity pulse of long period in the direction transverse to the fault. Part 2 of EC8 (“Bridges”) requires, though, the use of site-specific spectra taking into account near-source effects for bridges located within 10 km horizontally from a known active fault that may produce an event of moment magnitude higher than 6.5. In this respect, it gives a default definition of an active fault as one where the average historic slip rate is at least 1 mm/year and there is topographic evidence of seismic activity in Holocene times (i.e., during the past 11,000 years).

Representation of the Seismic Action Through Acceleration Time-Histories

A dynamic (response-history) analysis, linear or nonlinear, requires as input acceleration time-histories of the ground motion. These time-histories should conform on average with the 5 %-damped elastic response spectrum defining the seismic action.

EN-Eurocode 8 (CEN 2004a) requires as input for a response-history analysis an ensemble of at least three records (or pairs or triplets of different records, for analysis under two or three concurrent components of the action, respectively). It accepts for this purpose artificial, historic, or simulated records.

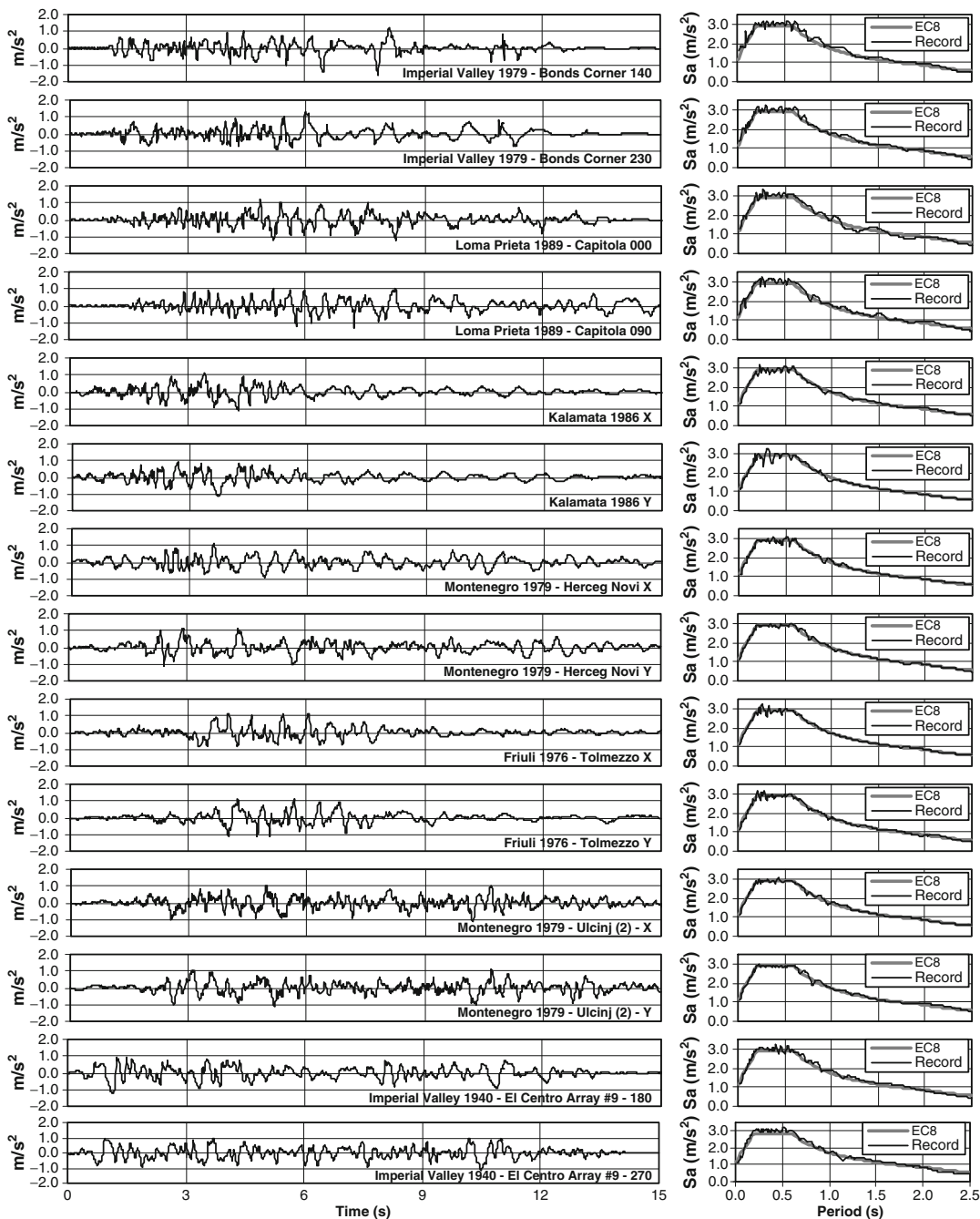
Artificial (or “synthetic”) records can be mathematically produced using random vibration theory to match almost perfectly the response spectrum defining the seismic action (Gasparini and Vanmarcke 1976). It is fairly straightforward to adjust the phases of the various sinusoidal components of the artificial waveform, as well as the time-evolution of their amplitudes (“envelope function”), so that the artificial record resembles a specific recorded motion; this type of record is called “modified historic” in Part 2 of EN-Eurocode 8. Figure 4 shows seven pairs of 15 s long “modified historic” records, modulated to match the type 1 horizontal response spectrum over ground type C for a PGA on rock of 0.10 g (0.115 g over ground type C); it also compares the resulting spectra with the target spectrum per EN-Eurocode 8.

Note, though, that records which are equally rich in all frequencies are not realistic. Moreover, a motion with a smooth response spectrum, without peaks and troughs, introduces a conservative bias in the response, as it does not let the inelastic response help the structure escape from a spectral peak to a trough at a longer period. Therefore, historic records are favored in EN-Eurocode 8. Records simulated from mathematical source models, including rupture, propagation of the motion through the bedrock to the site, and finally via the subsoil to the surface, are also preferred over artificial ones, as the final records resemble natural ones and are physically appealing. Obviously an equally good average fitting of the target spectrum requires more – appropriately selected – historic or simulated records than artificial ones. Individual recorded or simulated records should be “adequately qualified with regard to the seismogenetic features of the sources and to the soil conditions appropriate to the site” (CEN 2004a). In plain language, they should come from events with magnitude, fault

distance, and mechanism of rupture at the source consistent with those of the design seismic action. The travel path and the subsoil conditions should preferably resemble those of the site. These requirements are not only hard to meet but also may conflict with conformity (in the mean) to the target spectrum of the design seismic action. EN-Eurocode 8 also requires designers to scale individual historic or simulated records so that their peak ground acceleration (PGA) matches on average the value of $a_g S$ of the design seismic action; this requirement may also be considered against physical reality. It is more meaningful, instead, to use individual historic or simulated records with PGA values already conforming to the target value of $a_g S$. Note also that the PGA alone may be artificially increased or reduced, without affecting at all the structural response. So, it is more meaningful to select the records on the basis of conformity of the spectral values alone, as described below.

If pairs or triplets of different records are used as input for analysis under two or three concurrent seismic action components, conformity to the target 5 %-damped elastic response spectrum may be achieved by scaling the amplitude of the individual records as follows:

- For each earthquake consisting of a triplet of translational components, the records of horizontal components are checked for conformity separately from the vertical one (CEN 2005b).
- The records of the vertical component, if considered, are scaled so that the average 5 %-damped elastic spectra of their ensemble is at least 90 % of the 5 %-damped vertical spectrum at all periods between $0.2T_v$ and $2T_v$, where T_v is the period of the lowest mode with a participation factor of the vertical component higher than those of both horizontal ones (CEN 2005b).
- For analysis in 3D under both horizontal components, the 5 %-damped elastic spectra of the two horizontal components in each pair are combined by applying the Square-Root-of-Sum-of-Squares (SRSS) rule at each period value. The average of the “SRSS spectra” of the two horizontal components of the



European Structural Design Codes: Seismic Actions, Fig. 4 Historic records modified to match EC8’s type 1 horizontal response spectrum for ground type C

individual earthquakes in the ensemble should be at least $0.9\sqrt{2} \approx 1.3$ times the target 5 %-damped horizontal elastic spectrum at all periods from $0.2T_1$ up to $2T_1$, where T_1 is the

lowest natural period of the structure in any horizontal direction. If it isn’t, all individual horizontal components are scaled up, so that their final average “SRSS spectrum” exceeds

by a factor of 1.3 the target 5 %-damped horizontal elastic spectrum everywhere between $0.2T_1$ and $2T_1$.

For analysis under a single horizontal component, Part 1 of Eurocode 8 (CEN 2004a) restricts the mean 5 %-damped elastic spectrum of the applied motions not to fall below 90 % of that of the design seismic action at any period from $0.2T_1$ to $2T_1$.

If the response is obtained from at least seven nonlinear time-history analyses with (triplets or pairs of) ground motions chosen in accordance with the above, the relevant verifications may use the average of the response quantities from all these analyses as action effect. The set of seven pairs in Fig. 4 meets this condition for analysis in 3D under two simultaneous horizontal components of ground motion. If less records are used (but not less than three), the most unfavorable value of the response quantity among the analyses should be used.

Spatial Variability of the Seismic Action: Asynchronous Ground Motions

Parts 2 and 4 of EN-Eurocode 8 (CEN 2005b, 2006) require designers to consider the spatial variability of the seismic action in the design of extended structures, namely, of pipelines and multi-support bridges with a continuous deck longer than two-thirds of the distance beyond which the ground motion is taken as uncorrelated. This distance is an NDP with recommended value ranging from 600 m to 300 m for ground type A to D, respectively. Informative Annex D in (CEN 2005b) provides models of the spatial variability of the ground motion and appropriate methods of analysis, including guidance for the generation of seismic motion samples reflecting the probable spatial variability between supports. However, this theory is complex and requires special tools for its numerical implementation and special site data for the determination of the model parameters. So, the code allows using a simplified approach to take into account the spatial variability of seismic motions along

the bridge. So, Part 2 of EN-Eurocode 8 (CEN 2005b) also gives a simplified method to account for such effects in design. This may be achieved by superimposing to the seismic action effects from an analysis that neglects the spatial variability of the motion the effects of postulated static relative displacements of the supports in the same or in the opposite direction, which are a function of the peak ground displacement from Eq. 15 and the aforementioned distance beyond which the ground motion may be considered as uncorrelated.

Summary

The six parts of the first European Standard for seismic design, assessment, and retrofitting of structures were published from 2004 to 2006 as EN-Eurocode 8. In this European structural design standard, the seismic action for which new structures are designed for the Ultimate Limit State (life safety performance) is the “design seismic action”; it is determined as the product of the reference seismic action times the importance factor of the structure. The return period of the reference seismic action is chosen by the country, which retains the authority to set the safety level in its territory, taking into account economic aspects. Seismic zoning is expressed in terms of the peak ground acceleration on rock, to which the elastic acceleration response spectrum is anchored. This spectrum has regions of constant spectral pseudo-acceleration, pseudo-velocity, and displacement, the extent of which is decided by the country depending on the earthquake magnitude, as a function of the type of soil in the top 30 m and the geologic conditions underneath. Five standard types of ground are defined; for each one of them, two types of indicative spectral shapes are given: one for a surface magnitude above 5.5 and the other for below that limit. The indicative spectral shapes have been developed on the basis of strong motion data from the south of Europe. The peak ground displacement is given as a function of the peak ground acceleration, along with guidance for the construction of the displacement response spectrum on the basis of the peak ground displacement and the

acceleration response spectrum. Vertical response spectra are introduced and indicative shapes are given. Elastic response spectra are utilized to derive reduced design spectra for conventional force-based design or to generate/adopt acceleration time-histories for nonlinear dynamic analysis. For the design of new buildings, towers, chimneys, tanks, silos, or pipelines, a second seismic action is introduced for the damage limitation state (for protection of property or operation of the facility under a frequent or occasional event). Three limit states (performance levels) are introduced for the assessment and retrofitting of existing buildings: damage limitation, significant damage, and near collapse. The country may set the levels of seismic action for which these limit states are checked or may leave the decision to the owner.

Cross-References

- [Behavior Factor and Ductility](#)
- [Building Codes and Standards](#)
- [Earthquake Recurrence](#)
- [Earthquake Response Spectra and Design Spectra](#)
- [Earthquake Return Period and Its Incorporation into Seismic Actions](#)
- [EEE Catalogue: A Global Database of Earthquake Environmental Effects](#)
- [Engineering Characterization of Earthquake Ground Motions](#)
- [Nonlinear Dynamic Seismic Analysis](#)
- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- [Physics-Based Ground-Motion Simulation](#)
- [Probabilistic Seismic Hazard Models](#)
- [Random Process as Earthquake Motions](#)
- [Response-Spectrum-Compatible Ground Motion Processes](#)
- [Response Spectrum Analysis of Structures Subjected to Seismic Actions](#)
- [Review and Implications of Inputs for Seismic Hazard Analysis](#)
- [Seismic Actions due to Near-Fault Ground Motion](#)
- [Seismic Sources from Landslides and Glaciers](#)

- [Selection of Ground Motions for Response History Analysis](#)
- [Site Response for Seismic Hazard Assessment](#)
- [Spatial Variability of Ground Motion: Seismic Analysis](#)
- [Stochastic Ground Motion Simulation](#)
- [Structural Design Codes of Australia and New Zealand: Seismic Actions](#)
- [Time History Seismic Analysis](#)

References

- Abrahamson NA, Litehiser JJ (1989) Attenuation of peak vertical acceleration. *Bull Seismol Soc Am* 79:549–580
- Ambraseys N, Simpson K (1996) Prediction of vertical response spectra in Europe. *J Earthq Eng Struct Dyn* 25:401–412
- Ambraseys N, Simpson K, Bommer JJ (1996) Prediction of horizontal response spectra in Europe. *J Earthq Eng Struct Dyn* 25:371–400
- Ambraseys N, Smit P, Berardi R, Rinaldis D, Cotton F, Berge C (2000) Dissemination of European strong motion data. CD-ROM collection, European Council Environment and Climate Program
- Bommer JJ, Elnashai AS (1999) Displacement spectra for seismic design. *J Earthq Eng* 3(1):1–32
- Bommer JJ, Scott SG (2000) The feasibility of using real accelerograms for seismic design. In: Elnashai A, Antoniou S (eds) *Implications of recent earthquakes on seismic risk*. IC Press, London, pp 115–126
- Bommer JJ, Elnashai AS, Chlimintzas C, Lee D (1998) Review and development of spectra for displacement-based seismic design. Research report ESEE/98-3, engineering seismology and earthquake engineering section, Imperial College, London
- Bozorgnia Y, Campbell KW (2004) The vertical-to-horizontal spectral ratio and tentative procedures for developing simplified V/H and vertical design spectra. *J Earthq Eng* 8(2):175–207
- CEN (2004a) European standard EN 1998-1:2004. Eurocode 8: design of structures for earthquake resistance. Part 1: General rules, seismic actions and rules for buildings. Comité Européen de Normalisation, Brussels
- CEN (2004b) European standard EN 1998-5:2004. Eurocode 8: design of structures for earthquake resistance. Part 5: Foundations, retaining structures, geotechnical aspects. Comité Européen de Normalisation, Brussels
- CEN (2005a) European standard EN 1998-3:2005. Eurocode 8: design of structures for earthquake resistance. Part 3: Assessment and retrofitting of buildings. Comité Européen de Normalisation, Brussels
- CEN (2005b) European standard EN 1998-2:2005. Eurocode 8: design of structures for earthquake

- resistance. Part 2: Bridges. Comite Europeen de Normalisation, Brussels
- CEN (2005c) European Standard EN 1998-6:2005. Eurocode 8: Design of structures for earthquake resistance. Part 6: Towers, masts, chimneys. Comite Europeen de Normalisation, Brussels
- CEN (2006) European standard EN 1998-4:2006. Eurocode 8: design of structures for earthquake resistance. Part 4: Silos, tanks, pipelines. Comite Europeen de Normalisation, Brussels
- Elnashai AS, Papazoglou AJ (1997) Procedure and spectra for analysis of RC structures subjected to strong vertical earthquake loads. *J Earthq Eng* 1(1):121–155
- European Commission (2001) Guidance paper L: application and use of Eurocodes. CONSTRUCT 01/483 Rev.1, Brussels
- European Commission (2003) Commission recommendation on the implementation and use of Eurocodes for construction works and structural construction products. Document no. C(2003)4639, Brussels
- Faccioli E, Vanini M, Frassinetti L (2002) Complex site effects in earthquake ground motion, including topography. In: The 12th European conference on earthquake engineering, London, paper 844
- Faccioli E, Paolucci R, Rey J (2004) Displacement spectra for long periods. *Earthq Spectra* 20(2):347–376
- Fardis MN (2009) Seismic design, assessment and retrofitting of concrete buildings (based on EN-Eurocode 8). Springer Science+Business Media BV, Dordrecht
- Fardis MN, Carvalho EC, Elnashai A, Faccioli E, Pinto P, Plumier A (2005) Designers' guide to EN 1998-1 and EN 1998-5: Eurocode 8: design of structures for earthquake resistance. General rules, seismic actions, design rules for buildings, foundations and retaining structures. Thomas Telford Publishers, London
- Gasparini DA, Vanmarcke EH (1976) Simulated earthquake motions compatible with prescribed response spectra. Res. Report R76-4, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, MA
- Kolias B, Fardis MN, Pecker A (2012) Designers' guide to Eurocode 8: design of bridges for earthquake resistance, EN 198-2. Institution of Civil Engineers (ICE) Publishing, London
- Ohta Y, Goto N (1976) Estimation of S-wave velocity in terms of characteristic indices of soil. *Butsuri-Tanko* 29(4):34–41
- Paolucci R (2002) Amplification of earthquake ground motion by steep topographic irregularities. *Earthq Eng Struct Dyn* 31:1831–1853
- Paolucci R (2006) Numerical investigation of 3D seismic amplification by steep topographic profiles and check of the EC8 topographic amplification coefficients. In: Bouckovalas G (ed) General report proceedings of the athens workshop, ETC12. Geotechnical evaluation and application of the seismic Eurocode EC8, National Technical University of Athens, Athens, pp 187–191
- Rey J, Faccioli E, Bommer JJ (2002) Derivation of design soil coefficients (S) and response spectral shapes for Eurocode 8 using the European strong-motion database. *J Seismol* 6(4):547–555
- Tolis SV, Faccioli E (1999) Displacement design spectra. *J Earthq Eng* 3(1):107–125

Features Extraction from Satellite Data

Fabio Del Frate and Matteo Picchiani
Department of Civil Engineering and Computer
Science Engineering, University of Tor Vergata,
Rome, Italy

Synonyms

Dimensionality reduction; Image classification;
Image processing; Parameter retrieval

Introduction

Earth Observation (EO) images can be very useful in a wide range of human activities, from the understanding of global phenomena to decision-making processes. Mostly, the exploitation of the acquired images in the area and period of interest is still performed by visual interpretation. In this way, experts in EO sensing and application domains (e.g., seismology, agriculture, meteorology, forestry, etc.) add their knowledge after their analysis. When the experts' interpretation is concluded, only a minimal part of the data, the portion containing the useful information, is kept, while the remaining part can be discarded. However, this process is highly demanding in terms of costs, human effort, and time. Therefore, it cannot be performed on a regular basis, which

involves the loss of information potentially useful for research activities, service providers, and final users. The consequences of such a lack of information flow might be very significant as it may prevent operators who rely on EO data from making the right decisions, and leave relevant phenomena undetected or to be discovered too late. In this context, important support can be provided by feature extraction methods, more automated and direct procedures for the extraction of information from images. Relying on "intelligent" automatic techniques, these methods support the image interpreters in discovering the essential elements of information in an image. Sometimes the extracted feature is already the final object of interest, for example, an oil spill in the ocean or a thermal anomaly. In other cases, it consists of intermediate processing aiming at generating a new representation of the image data where particular statistical or geometrical characteristics are highlighted.

Pixel averaging is one of the simplest techniques used when extracting features in image processing. This means replacing a box of pixels with a single pixel. In particular, the value of the pixel will be obtained by computing the average of the values of all the pixels belonging to the considered box. We can say that the new pixels' values are features of the original image, representing its information content in a more compact way. From the example given, it may be understood that feature extraction can also play another crucial role in the context of image

processing and, more in general, of pattern recognition: dimensionality reduction. In fact, dimensionality reduction can be a fundamental step in the implementation of many classification or retrieval problems. If we are forced to work with a limited quantity of data, as we often are, then increasing the dimensionality of the space rapidly leads to the point at which the data is very sparse, therefore providing a very poor representation of the mapping. To describe the extraordinarily rapid growth in the difficulty of problems as the number of variables (or dimension) increases, the phrase “the curse of dimensionality” has been coined by Richard Bellman (1957). Reducing the number of input variables can lead to improved performance for a given data set, even though information is being discarded. The mapping is better specified in the lower-dimensional space, which more than compensates the loss of information.

An Overview of Feature Extraction Techniques

According to the type of satellite data to be analyzed (SAR, optical/multispectral, time series) and the considered application, different approaches can be used for feature extraction:

Spatial and Contextual Features

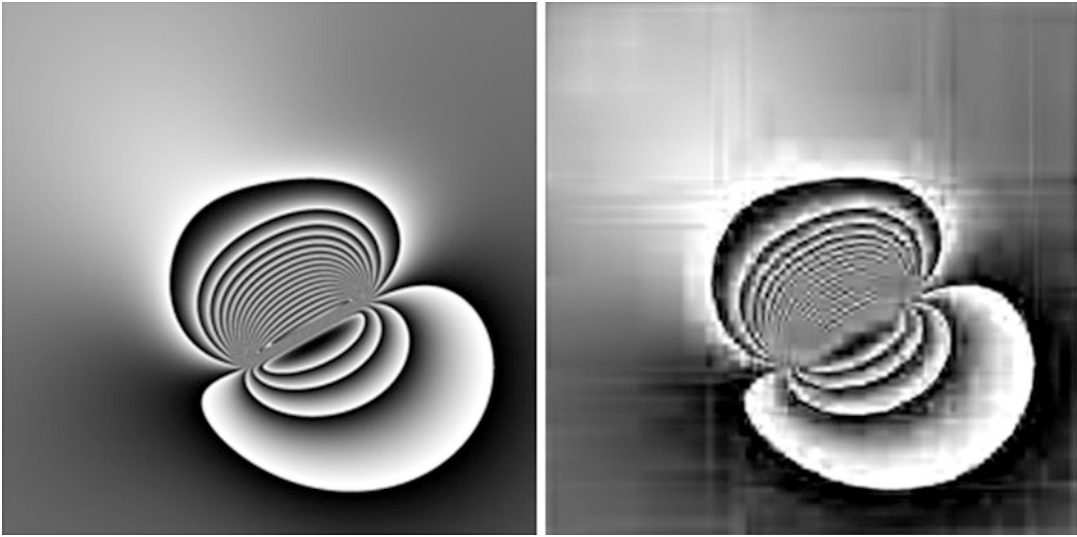
Spatial and contextual features try to synthesize the information contained in neighboring pixels with respect to a pixel under examination. Different contextual descriptors can be defined. One of the earliest and most important techniques used to express textural and tonal variations in a remotely sensed image data is the one based on the gray-level co-occurrence matrix (GLCM) (Haralick et al. 1973). The GLCM is the core of a mathematical framework capable of representing the distance and angular relationships among the pixels over an image subregion of a specified size. Each element of the GLCM is a measurement of the probability of occurrence of two gray scale values separated by a given distance in a given direction. Several quantities

can be computed from the elements of the GLCM which describe quantitatively the textural behavior of the subregion considered. Another approach for textural characterization relies on probabilistic modeling. Markov random field (MRF) models fall under this category and can be quite successful for texture modeling (Cross and Jain 1983). They capture local characteristics of an image considering regional patterns and by assuming a local conditional probability distribution. Spatial feature extraction can be based on multichannel filtering as well. The image is decomposed into a set of sub-bands which are calculated by convolving it with a bank of linear filters, nearly uniformly covering the spatial-frequency domain, followed by some nonlinear procedures. Gabor filters can be an example for this type of techniques (Jain and Farrokhnia 1991). Mathematical morphological filters should also be mentioned for the detection of textural features. In this case, the extraction is obtained by performing shape-oriented operations, like openings and closures, based on set theory (Soille 2003).

Transform Domain Features

An image transform is a mathematical operation that re-expresses in a different, and possibly more meaningful, form all, or part of, the information content of a multiband or gray scale image. One of the most used transform methods for feature extraction is the principal component analysis (PCA) (Singh and Harrison 1985) where a multispectral set of m images is re-projected in terms of a set of m principal components. Two particular properties exist: the components are extracted for decreasing variance, and the principal components are statistically unrelated or orthogonal. For multiband images, the first PCA components can represent effective features to be used in the operational scenarios to be considered, as they reveal the structure behind the correlation of the variables. Indeed no linear combination (i.e., weighted sum) of the original bands can contain more information than is present in the first principal components.

An alternative approach considers harmonic analysis, a field of mathematics that studies the



F

Features Extraction from Satellite Data, Fig. 1 $1,500 \times 1,500$ pixels synthetic interferogram's example (*left*) and reconstructed one by means of first 500 DWT components (*right*)

representation of functions as overlapping of fundamental waveforms. It is known that a multivariate function f can be well approximated by the linear combination of the elements of a given basis. This type of operation is called harmonic analysis. When harmonic analysis is applied to image data, the discrete image can be seen as a two-dimensional signal, and it is possible to consider a set of mathematical tools that perform a transformation suitable to extract some features otherwise difficult to be identified. This can be done by means of particular functional operators like the DFT (discrete Fourier transform) and the DWT (discrete wavelet transform). Both the transforms express the signal as coefficients in a function space spanned by a set of basis functions. The basis of the DFT is complex exponential functions, representing sinusoid functions in the real domain, and the multiplying coefficients are also complex numbers. The basis of the DWT is scaled and shifted versions of a mother wavelet real-valued function. In this case, the coefficients have real values. Low pass filtering of the DFT and DWT transformed images can be considered for the extraction of low spatial-frequency features (Picchiani et al. 2012). See an example of application in Fig. 1.

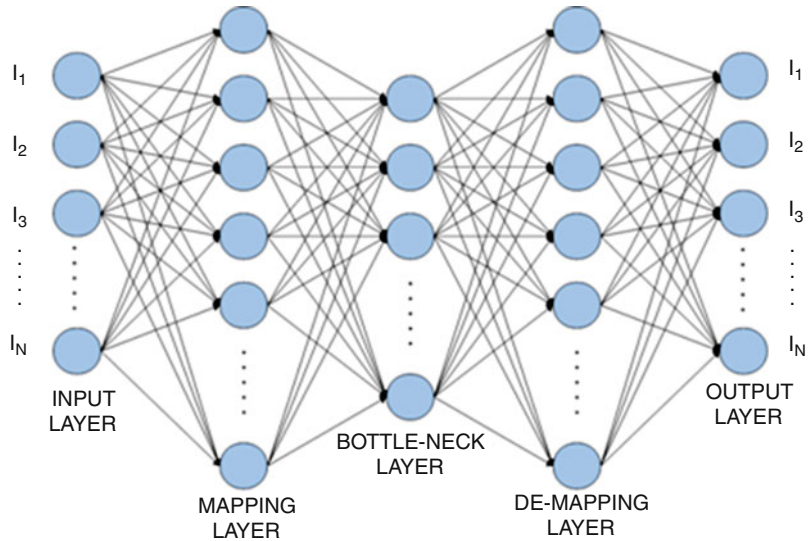
Advanced Nonlinear Techniques

Linear techniques such as PCA may be unable to capture nonlinear correlations so they may fail when dealing with highly complex data. One way of performing a nonlinear extension is to lift the input space to a higher dimensional feature space by a nonlinear feature mapping and then to find a linear dimension reduction in the feature space (Scholkopf et al. 1999). Kernel techniques allow such nonlinear extensions without explicitly forming a nonlinear mapping or a feature space, as long as the problem formulation only involves the inner products between the data points and never the data points themselves.

Neural networks can also be applied to effectively extract nonlinear features from data. For example, this can be obtained by using Autoassociative Neural Networks (AANN) (Bishop 1995). These are multilayer perceptron (MLP) networks which are trained to the purpose of reproducing the inputs in the output layer (Fig. 2). This means that the targets used in the training phase are simply the input vectors themselves. Hence, the net is said to perform autoassociative mapping (autoencoder). One of the hidden layers of the AANN, known as the bottleneck layer, is characterized by a number of units significantly lower than the number of

Features Extraction from Satellite Data,

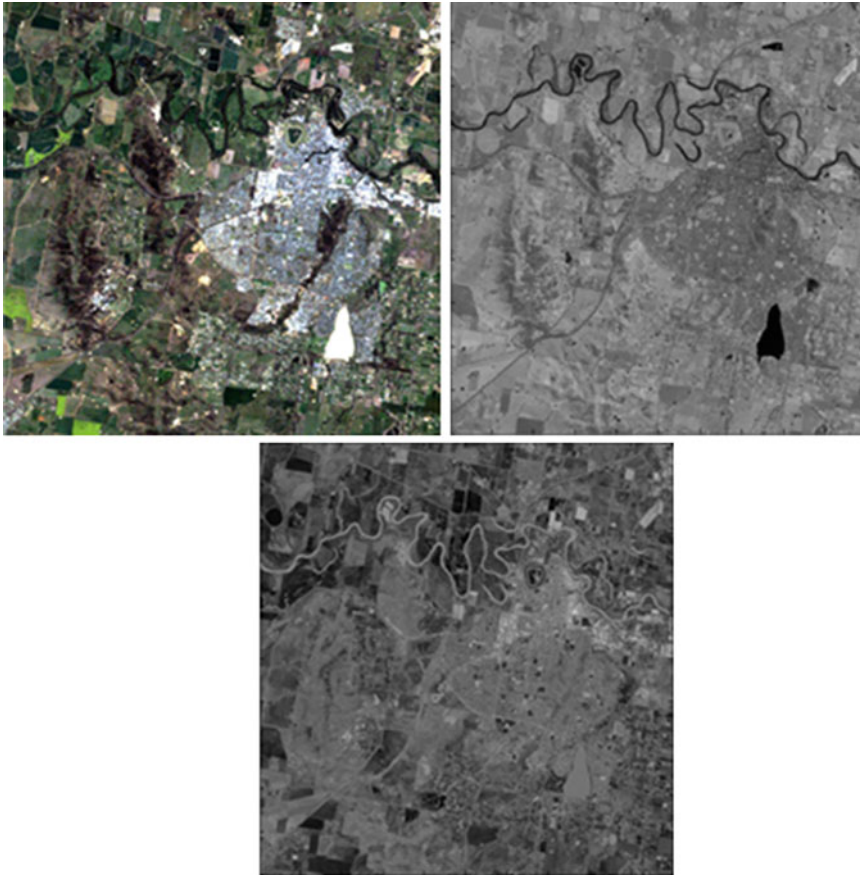
Fig. 2 Autoassociative Neural Network topology



inputs/outputs. If during the training phase the actual outputs of the network are capable of resembling those desired with a satisfactory level of error, this implies that the bottleneck layer is capable of providing, in a lower dimension, good representation of the input. In Fig. 3, two nonlinear principal components calculated from a Landsat image are shown. SOM (Self-Organizing Maps) represent another possibility of performing feature extraction using neural networks. The SOM map or Kohonen map is a two-layered network. The first layer of the network is the input layer. The second, called the competitive layer, is usually organized as a two-dimensional grid. All interconnections go from the first to the second layer. In the training phase, the data are presented in a random order to the network. All the nodes in the competitive layer compare the inputs with their weights and compete with each other to become the winning unit with the lowest difference. After that, the weights of the winning neuron and of its neighbors are adjusted towards the input vector. The magnitude of change decreases with time and with distance from the winning neuron. In this way, a SOM can automatically form 1D or multidimensional maps of the intrinsic features of the input data and reorganize them in several output clusters/classes.

Feature Extraction in Seismological Applications

There are several features of interest for seismology that can be extracted from Earth Observation data. A major category regards the thermal anomalies. Due to pressure increase and subsurface degassing activities, a sudden rise in land surface temperature (LST) can characterize the area of interest prior to an earthquake. The detection of such thermal changes can be very important in an effort to predict earthquake activities and provide time to give suitable warnings. Appropriate satellite data with thermal bands can be used to extract this kind of feature from multi-temporal images. Moreover, besides land surface, due to the release of gas emissions causing a local greenhouse effect, anomalies can also be detected in the concentration of some gases in the atmosphere. Another important application is focused on retrieving information about the surface deformation once earthquake occurs. Earthquakes produce static displacements of the ground that can be observed near the fault that slipped during a seismic rupture. The measurement of such displacements is a crucial issue in seismotectonics, as it gives insight into the geometry of the ruptured fault and the energy released by the earthquake. Such measurements are generally performed using geodetic techniques that can



F

Features Extraction from Satellite Data, Fig. 3 Original Landsat image (*left*) and two computed nonlinear principal components (*center and right*)

only provide a sparse covering, because the areas of coseismic ground displacement are not known a priori. Satellite imagery is naturally suited for such measurements, because it regularly provides comprehensive and detailed images of the ground with high radiometric and geometric quality. The displacement field can be measured by the comparison of images acquired before and after the event. Both InSAR (interferometric SAR) and optical data can be exploited for this type of investigation. Other effective uses of Earth Observation data for seismology are connected with disaster management. In the management of a seismic event to immediately draw up a draft, damage map of the stricken urban areas is of great importance in order to organize civil protection interventions. Remote sensing can provide

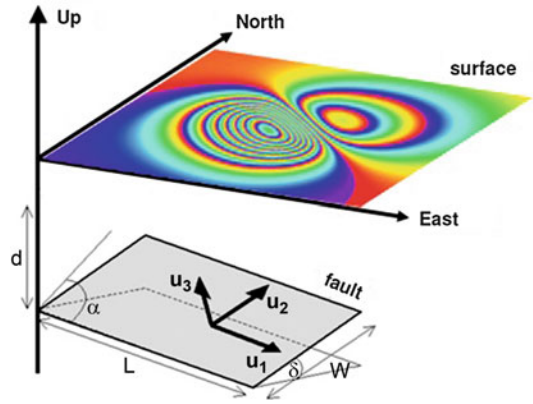
valuable items of information in this respect, thanks to its synoptic capability, in particular when the seismic event is located in remote regions or a failure affected the main communication systems. Both optical and radar sensors can be exploited for the extraction of this type of feature. Despite the great potential provided by Earth Observation data for supporting operational activities in the field seismology, it has been pointed out that the level of automatic data processing is mostly inadequate with “still too much manual labour and author arbitrariness” (Tronin 2010). This is not the case when considering the estimation of tectonic parameters from SAR interferometry where feature extraction, according to innovative technologies based on neural networks, can be fully automated.

Automatic Tectonics Parameter Retrieval from InSAR Data

Until SAR interferometry (InSAR) broke out in the beginning of the 1990s as an effective technique for analyzing a seismic event, most of the measurements used for this kind of study were based on the ground recording of the waves generated by the earthquake. In fact, such data are still very helpful in providing information on the location or the magnitude of the seismic event. However, these measurements may not be sufficient if a more accurate estimation of the earthquake parameters, such as the fault depth, is desired.

The first application of the InSAR technique for the investigation of the characteristics of a seismic event regarded the Landers Earthquake (Massonet et al. 1993). The innovative method was capable of providing effective information on the generated displacement field. After that, a few studies using the same approach were conducted on various both strong (Kobe, Japan, 1997; Izmit, Turkey, 1999; Hector Mine, CA, USA, 2000; Denali, AK, USA, 2002; Bam, Iran, 2003; Sichuan, China, 2008) and moderate (Colfiorito, Italy, 1997; Abruzzi, Italy, 2009) earthquakes. It was discovered that, thanks to its ability to generate a distributed map of the total strain release, in many cases the InSAR technique was the only one capable of providing the investigators with a higher level of information.

In fact, by comparing pre- and post-event SAR data and calculating the interferometric phase, InSAR can reconstruct the coseismic surface displacement field with a centimetric accuracy. The potential of InSAR data can also be combined with GPS measurements and with more standard seismological and geophysical data, such as strong motion records, for the assessment of normal fault models, generally stemming from the Okada formulation (Okada 1985). In particular, for modeling radar interferometric data, specific software packages can be used to calculate the displacement components (Feigl and Duprè 1999) which, in an elastic half space, are the expression of the seismic source at the surface.



Features Extraction from Satellite Data, Fig. 4 Fault geometry and symbols

In fact, the inverse problem is of great interest. This means the attempt to retrieve, also in a quantitative way, the source parameters starting from the knowledge of the InSAR surface displacement field. In particular, some useful information to define the fault geometry (dip and strike angle; width and length), the extension of the rupture, and the slip distribution on the fault plane can be obtained.

For the estimation of the desired parameters, two main steps have to be carried out after the computation of the interferogram. The first one regards the dimensionality reduction of the data. The second one is the design of a suitable retrieval algorithm. With reference to Fig. 4, besides the length and the depth of the fault, tectonic parameters that can be estimated are the following: the slip vector $U = [U_1, U_2, U_3]$, which represents the movement of the hanging wall with respect to the foot wall signed such that U_1 is positive for left lateral strike slip, U_2 is positive for thrusting dip slip, and U_3 is tensile slip; the strike (α), which is the clockwise angle between the direction of the fault trace and the North–South one; and the dip (δ), that is, the angle between the fault plane and the surface. Another possibility, optionally complementary to parameter estimation, is the classification of the fault. The main classes of normal fault, strip-like fault and reverse fault (thrust), are typically considered.

Dimensionality Reduction of the Data

Data dimensionality reduction is one of the key issues in retrieval and classification problems. In this specific case, it is certainly not appropriate to feed a classifier or a parameter estimator with all the DN numbers forming the interferometric image. In such a context, a standard technique for performing dimensionality reduction is to sample the interferogram with an adequate frequency. A better performance can be obtained considering the combination of two techniques for feature extraction. In particular, the first processing step relies on the harmonic analysis, and in the second, the AANN is applied. Such an approach leads to the determination of a considerably reduced number of components for the interferogram representation without involving the training of huge AANN topologies. An interferogram with a dimension of $1,500 \times 1,500$ pixels can effectively be represented by a vector of 50 components if the chaining of the AANN and the DWT is applied (Picchiani et al. 2012).

Classification and Retrieval Using Neural Networks

Once the data dimensionality reduction has been performed, and the interferogram is represented by means of a vector expressing a limited number of significant feature values, an algorithm is necessary to transform such a concentrated information content into a set of values to be associated with the desired quantities. Such desired quantities may either be values of tectonic parameters or probabilities of belonging to a specific fault class. In both cases, the problem is rather complex due to its inherent nonlinearities. For this reason, the use of neural networks can be considered as they are recognized as a powerful tool for inversion procedure in remote sensing applications (Stramondo et al. 2011).

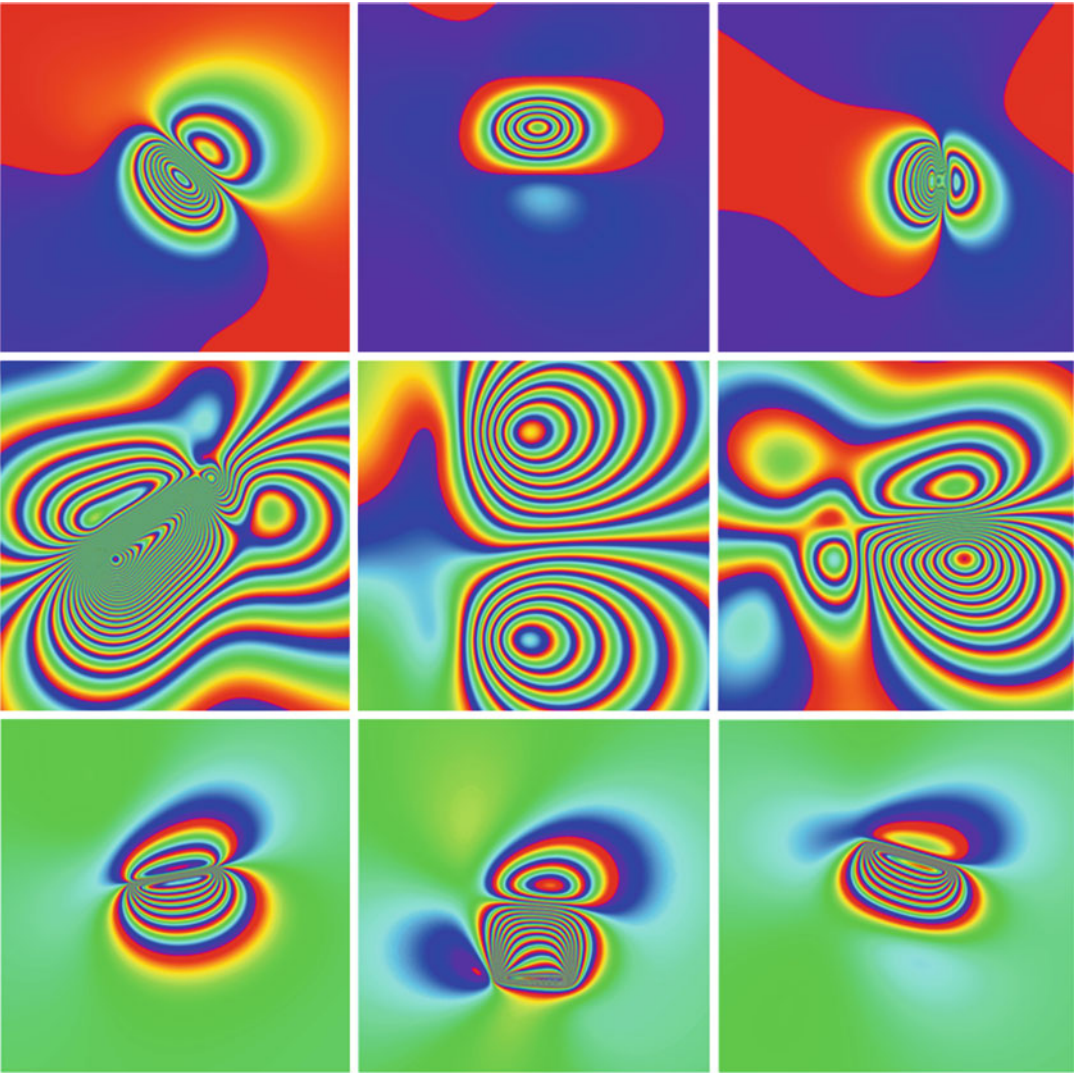
Among various topologies, multilayer perceptrons (MLPs) have been found to have the best suited topology for classification and inversion problems and can also be used in this case. These are feedforward networks where the input only flows in one direction, towards the output, and each neuron of a layer is connected

to all neurons of the successive layer but has no feedback to neurons in the previous layers.

Compared to other methods, the use of NNs is often effective because they can simultaneously address nonlinear dependences and complex physical behavior with reduced computational efforts and without requiring any a priori information (Bishop 1995). Moreover, NNs learn the input–output relationships exclusively from the training patterns; therefore, no explicit expression addressing the interactions between the parameters of interest is needed with such an approach.

In the training phase, the weights (internal parameters of the network) are progressively adjusted to enable the approximation of the needed functional relationships. The performance of the algorithm is monitored by an error (or cost) function which has a nonlinear dependence on weights. Once the error is minimized, the trained MLP performs the desired mapping of input space onto output space.

As far as the source mechanism classification is concerned, the training and validation sets can be created through the generation of synthetic differential interferograms. Such a generation can be performed using a forward model computing the surface deformation (the displacement vector) associated to specific values of the given fault parameters. For this purpose, the Okada formulation can be considered as it explains, in a close analytic form, the surface deformation due to a seismic event by a dislocation model in an elastic half space. The data set is obtained by running the forward model varying the values of its parameters appropriately all along a certain range (see Fig. 5). In the classification case, the output layer of the network consists of three units each associated with one of the three considered fault mechanisms. The network is trained in order to give the value one to the unit corresponding to the desired type of mechanism and the value zero to the other two units. In the parameter retrieval case, the output of the network is formed by the desired quantities. Once the network is trained, the technique performs in a fully automatic mode, providing the fault classification and/or the estimation of the fault parameters in real time.



Features Extraction from Satellite Data, Fig. 5 Different examples for the three considered types of faults of differential interferograms extracted

from the simulated data set generated by means of RNGCHN software. *Top*: normal fault; *middle*: strike slip fault; *bottom*: reverse fault

Summary

The contribution introduces the feature extraction techniques and explains the rationale for their use especially in the field of image processing. The main feature extraction approaches are briefly reviewed and an insight of the most important applications for the analysis of earthquakes with Earth Observation data is given. The complete methodology addressing the problem of the

estimation of tectonic parameters from SAR interferometric data is described.

Cross-References

- [Building Damage from Multi-resolution, Object-Based, Classification Techniques](#)
- [Damage Detection in Built-up Areas Using SAR Images](#)

- [Earthquake Damage Assessment from VHR Data: Case Studies](#)
- [Earthquake Magnitude Estimation](#)
- [Earthquake Mechanism Description and Inversion](#)
- [Earthquake Mechanisms and Tectonics](#)
- [Estimation of Potential Seismic Damage in Urban Areas](#)
- [Hyperspectral Data in Urban Areas](#)
- [InSAR and A-InSAR: Theory](#)
- [Land Use Planning Following an Earthquake Disaster](#)
- [Noise-Based Seismic Imaging and Monitoring of Volcanoes](#)
- [Remote Sensing in Seismology: An Overview](#)
- [SAR Images, Interpretation of](#)
- [Seismic Event Detection](#)
- [Spatial Variability of Ground Motion: Seismic Analysis](#)
- [Urban Change Monitoring: Multi-temporal SAR Images](#)

References

- Bellman R (1957) Dynamic programming. Princetown University Press, Princetown
- Bishop C (1995) Neural networks for pattern recognition. Oxford University Press, New York
- Cross GR, Jain AK (1983) Markov random field texture models. *IEEE Trans Pattern Anal Mach Intell PAMI*-5(1):25–39
- Feigl KL, Duprè E (1999) RNGCHN: a program to calculate displacement components from dislocations in an elastic half-space with applications for modeling geodetic measurements of crustal deformation. *Comput Geosci* 25(6):695–704
- Haralick RM, Shanmugam K, Dinstein I'H (1973) Textural features for image classification. *IEEE Trans Syst Man Cybern SMC* 3(6):610–621
- Jain AK, Farrokhnia F (1991) Unsupervised texture segmentation using Gabor filters. *Pattern Recogn* 24(12):1167–1186
- Massonnet D, Rossi M, Carmona C, Adragna F, Peltzer G, Feigl K, Rabaute T (1993) The displacement field of the Landers earthquake mapped by radar interferometry. *Nature* 364(6433):138–142
- Okada Y (1985) Surface deformation due to shear and tensile faults in a half space. *Bull Seismol Soc Am* 75(4):1135–1154
- Picchiani M, Del Frate F, Schiavon G, Stramondo S (2012) Features extraction from SAR interferograms for tectonic applications. *EURASIP J Adv Signal Process* 2012:155
- Schölkopf B, Mika S, Burges CJC, Knirsch P, Muller K, Ratsch G, Smola AJ (1999) Input space versus feature space in kernel-based methods. *IEEE Trans Neural Netw* 10(5):1000–1017
- Singh A, Harrison A (1985) Standardized principal components. *Int J Remote Sens* 6(6):883–896
- Soille P (2003) Morphological image analysis: principles and applications. Springer, New York
- Stramondo S, Del Frate F, Picchiani P, Schiavon G (2011) Seismic source quantitative parameters retrieval from InSAR data and neural networks. *IEEE Trans Geosci Remote Sens* 49(1):96–104
- Tronin AA (2010) Satellite remote sensing in seismology. A review. *Remote Sens* 2010(2):124–150

Frequency-Magnitude Distribution of Seismicity in Volcanic Regions

John J. Sanchez Aguilar¹, Hernán G. Alvarado² and Santiago López²

¹Departamento de Geociencias, Universidad Nacional de Colombia, Bogotá, Colombia

²Departamento de Geociencias, Universidad Nacional de Colombia-Sede Bogotá, Bogotá, Colombia

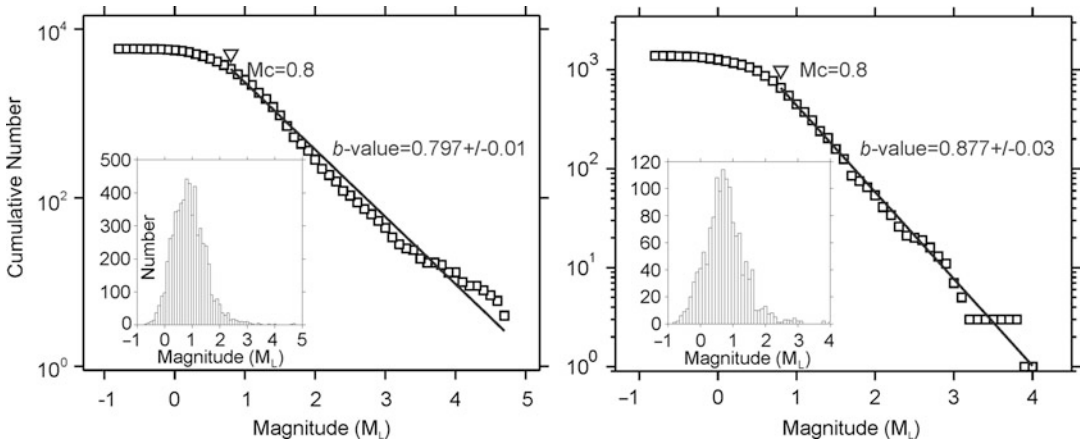
Synonyms

b-value anomalies; *b*-values; Earthquake size distribution; Seismic *b*-values around volcanoes; Volcanic seismicity; Volcano-tectonic earthquakes

Introduction

The spatial and temporal distribution of earthquakes sizes near and around volcanoes can be characterized through the well-known frequency-magnitude distribution:

$$\log N = a - bM \quad (1)$$



Frequency-Magnitude Distribution of Seismicity in Volcanic Regions, Fig. 1 Average magnitude of completeness and b -value for the data compiled by Observatorio Vulcanológico y Sismológico de Pasto (OVSP, Servicio Geológico Colombiano) at Galeras Volcano, Colombia. *Left*: Data for the period 2004–2013.

Total number of earthquakes used is 5,858. The *solid line* shows a least squares best fit for the linear part of the frequency-magnitude distribution (FMD). The *insets* show histograms of magnitudes, with bins equal to 0.1. *Right*: Data for the period 2008.3–2010.6. Total number of earthquakes is 1,384. Conventions as in *left* panel

where N is the cumulative number of earthquakes with magnitude $\geq M$ and a and b are constants representing the seismic productivity of a region and the slope of the linear part for the distribution, respectively. This is a well-known power law (Gutenberg and Richter 1944) that is applied herein.

Earthquake catalogs compiled by volcano observatories around the world are commonly used to estimate b values and map their spatial distribution. From Eq. 1, it is evident that earthquake magnitude plays a fundamental role and is important to mention that its computation may vary from one catalog to another. Firstly, different magnitudes have been traditionally calculated at different observatories (duration magnitudes, M_D , and local magnitudes, M_L , are most common) and secondly, detection capabilities vary because of the different network configurations and equipment. The magnitude of complete detection or magnitude of completeness (M_c) is the value above which the catalog is considered to be approximately complete (Fig. 1).

In a practical way M_c can be estimated by visual inspection of any of the forms of the FMD. For example, in Fig. 1 the FMD for earthquakes at Galeras Volcano, Colombia, is shown. The histogram of the number of events versus

magnitudes shows a slightly skewed distribution with a right tail and a maximum that occurs in the range $0.7 \leq M_L \leq 0.9$, a first indication of the probable value of M_c in the catalog. Another estimate of completeness may be obtained by finding the maximum curvature of the FMD using a combination of 90–95 % confidence in the curvature (derivative) estimation. Indeed, this procedure gives $M_c = 0.8$ for the whole catalog implying that earthquakes with $M_L < 0.8$ are probably not being always detected. Inspection of a curve of cumulative number of earthquakes versus time revealed that during 2004–2013 rates of seismicity at Galeras varied, most likely because of occurrence of swarms, groups of earthquakes with similar magnitudes that occur closely separated in time and not associated to a mainshock. The period 2008.3–2010.6, however, exhibited a fairly stable behavior and the corresponding FMD did not vary (Fig. 1, right panel). The data for both time periods follows fairly well a Gutenberg-Richter law and the difference in average b -values between both datasets is small. Earthquake catalogs usually are culled at the M_c value to ensure that the dataset is mostly complete. The linear part of the FMD can be fit by a least squares procedure to obtain a straight line with negative slope.

The slope is called the b -value and it represents the average size distribution of earthquakes.

Because the b -value is related to the mean magnitude in a given region, it is important to spatially map its distribution in both volcanic and tectonic environments because of the possible relationships with the state of stress in the crust. When calculating the spatial distribution of b , one must consider (1) the calculation method, which can either be maximum likelihood or weighted least squares, and (2) the parameters of the gridding scheme and the number of events to sample around each node of the grid (Wiemer and McNutt 1997).

To calculate the b -values and estimate their spatial distribution, it is necessary to create a grid in two or three dimensions. As a general rule the node separation should be proportional to the formal errors in the spatial parameters of the earthquake locations. Regarding the sampling mode around each node, there are two possibilities: either selecting a constant number of events within circles of variable radii or selecting a variable number of events inside circles of constant radii. In both methods the sampling volume has the shape of a cylinder, with its base centered on a particular node of the grid. If the calculation takes place in map view, the cylinder will be vertical, but if the calculation is on a vertical cross section, the cylinder will be horizontal, with variable height and radius depending on the sampling method. In this way the cylinders can be coin shaped or pipe shaped. If the grid is for a 3D region, the sampling volumes will be spheres.

Within each cylinder the b -value is calculated from the FMD obtained from the sample and in this case one can either assume a constant M_c (usually equal to the minimum magnitude within the sample) or calculate the FMD locally. It is usually a good practice to calculate b -values using both methods and trying several sampling schemes to test the stability of the results.

The estimation of b -values has been an active research area over the last two decades and here a description of results for volcanoes around the world is presented, then the possible implications for magma processes are summarized and the

body of accepted knowledge on seismic b -values near volcanoes is presented. Finally, a short case study on one of the most active volcanoes is shown to exemplify applications of the technique and to describe some possible implications for volcano and earthquake hazards and risk.

Seismic b -Values at Volcanoes Worldwide

The distributions of seismic b -values at depth have been studied (by spatial mapping mostly) at a number of volcanoes around the world using a variety of catalogs. A compilation of results of 2D and 3D mapping at 16 volcanoes is presented (Table 1).

Table 1 highlights that b -value anomalies are common regardless of the type of volcano, configuration of the seismograph network, magnitude scale used, or amount of earthquakes in the dataset. The pattern of the anomalies varies from place to place, however (Wyss et al. 1997), and it is established that, with no exception, volcanic areas exhibit irregularly shaped volumes of anomalously high b -values surrounded by regions of normal-to-low b -values (Wiemer and McNutt 1997; Wiemer and Wyss 2002). From Table 1, it is observed that there is notable spatial variability of b , with several volcanoes having more than one region of increased b -values. There is variability in the computed b -values as well, with b -anomalies in the range $1.0 < b < 3.0$, and although it would seem appealing to compare these values among volcanoes, conclusions drawn from such task would not be solid, because of the different magnitude scales used. For instance, the b -values of ~ 1.55 reported for shallow anomalies at Mt. St. Helens and Nevado del Ruiz are not necessarily comparable.

It is reasonable to assume that mineral stabilization (formation of crystals) in magmas takes place at depth within the magma chamber and that exsolution of gas and vesiculation might occur at shallower levels. The mineral and volatile species and types of magmas studied vary and thus the calculated depths for stabilization of minerals or gas exsolution and formation of

Frequency-Magnitude Distribution of Seismicity in Volcanic Regions, Table 1 Main features of *b*-value anomalies and data on mineral stabilization and gas processes at volcanoes around the world. Reported and calculated stabilization pressure values from various authors are converted to pascals (Pa, $\text{N/m}^2 = \text{m}^{-1} \cdot \text{kg} \cdot \text{s}^{-2}$) or megapascals ($\text{MPa} = 10^6 \text{ Pa}$) as this is the SI derived unit for pressure or stress

Volcano	Type	Number of earthquakes used	Magnitude scale	Mc	<i>b</i> -value (anomaly) ^a	Anomaly centroid depth (km)	References	Pressure (MPa)	References
St. Helens (USA)	Stratovolcano	1,674	M _L	0.4	1.5 ± 0.3	3.25 ± 0.55	Wiemer and McNutt (1997)	175	Pallister et al. (2008)
					1.6	8.35 ± 1.65			
Mt. Spurr (USA)	Stratovolcano	643	M _L	0.1	1.7	3.4 ± 1.1	Wiemer and McNutt (1997)		
					1.2	>10.0			
Mammoth Mtn. (Long Valley Caldera)	Caldera dome		M _L	1.3	1.8	5.0	Wiemer et al. (1998)	20	Burkett (2007)
					1.73 ± 0.06	8.0			
Resurgent Dome (Long Valley Caldera)	Active dome		M _L	1.3	1.9?	4.0			
					1.36 ± 0.05	10.0			
Teishi Knoll (Izu-Tobu group), located off Ito (Japan)	Submarine Crater, within pyroclastic cones field	10,000	M _L ?	1.3	1.00 ± 0.03	8.5	Wyss et al. (1997)	800	Kushiro (1991)
					1.54 ± 0.05	13.0			
Soufriere Hills (West Indies)	Stratovolcano	1,900	M _D	1.7	3.07 ± 0.07	2.0	Power et al. (1998)	150	Barclay et al. (1998)
Mt. Etna (Italy)	Stratovolcano	450	M _D	2.5	3.0 ± 0.5	7.5 ± 2.0	Murru et al. (1999)	155	Trigila et al. (1990)
						14.5 ± 2.5			
Martin (Alaska, USA)	Stratovolcano	2,500	M _L	0.7	1.6 ± 0.05 ^b	6.0 ± 1.0	Jolly and McNutt (1999)	62.5	Coombs and Gardner (2001)
Mageik (Alaska, USA)	Stratovolcano		M _L	0.7	1.8 ± 0.24 ^c	6.0 ± 1.5			

Katmai (Alaska, USA)	Stratovolcano, Caldera		M_L	0.7	1.30 ± 0.25	8.0 ± 1.0		37.5	Coombs and Gardner (2001)
Kilauea-East Rift Zone	Shield Volcano	16,963	M_p^d	1.5	1.73	12.0 ± 3.0 5 ± 2.0	Wyss et al. (2001)		
Mt. Redoubt (USA)	Stratovolcano				1.7 ± 0.1	4.0	Wiemer and Wyss (2002)	180.9	Gerlach et al. (1994)
					2.2 ± 0.2	9.0			
Nevado del Ruiz (Colombia)	Stratovolcano	2,769	M_D	0.7	1.55 ± 0.25	1.0	Rodríguez (2003)	50	Stix et al. (2003)
					1.65 ± 0.05	1.0 ± 0.5			
					1.55 ± 0.15	5.35 ± 0.35			
					1.55 ± 0.15	2.0 ± 0.2			
Pinatubo (Philippines)	Stratovolcano	1,406	M_D	0.73	1.7 ± 0.1	2.0 ± 1.0	Sánchez et al. (2004)	220	Bernard et al. (1996)
						11.0 ± 1.0			
Galeras (Colombia)	Stratovolcano	1,918	M_L^e	1.2	1.3 ± 0.05	2.5 ± 0.5	Sánchez et al. (2005)	60	Stix et al. (1997)
Miyake-jima-Tokushima (Izu Islands, Japan)	Stratovolcano	10,000	$M_L^?$	2.5	1.86 ± 0.1	8.0 ± 0.5	Rierola (2005)	224	Amma-Miyasaka et al. (2003)
						12.5 ± 2.5			
					1.3 ± 0.1	19 ± 2.0			
Makushin (Alaska, USA)	Stratovolcano	491	M_L	1.25	1.9 ± 0.1	4.0 ± 1.7	Bridges and Gao (2006) ^f	169	Goldberg (2010)

^aValues for b -anomalies and their locations are taken from results of 3D mapping and vertical cross sections, where reported

^bDuring July 27, 1995, to December 31, 1997, the Alaska Volcano Observatory reported over 2,500 earthquakes located in the general area Martin-Mageik-Trident-Katmai Caldera. The b -values reported in the table were estimated from Fig. 6 (Jolly and McNutt 1999)

^c b -values at Martin-Mageik region may include data from swarms in Sept. 1996–April 1997. Values calculated pre- and post swarm were in the range $0.92 < b < 0.98$

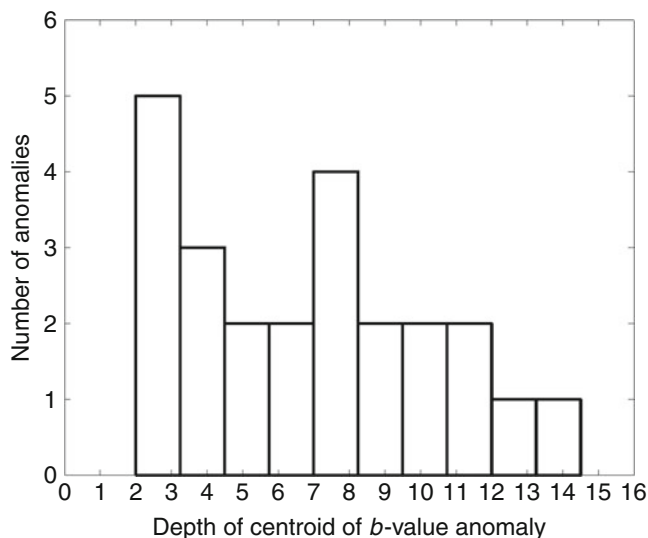
^d M_p is a “preferred magnitude” chosen hierarchically and defined in the data description prepared by F. Klein for distribution of the earthquake catalog at Hawaii. Mostly equal to M_D and in few cases equal to M_L

^eMagnitudes were recalculated by regression equation from M_D to M_L using a sample of earthquakes simultaneously recorded by broad band and short period subnetworks

^fBridges and Gao report a second, deeper b -value anomaly ($b \sim 1.6$) detected at roughly 12 km depth southeast of Makushin, but it was deemed statistically insignificant

Frequency-Magnitude Distribution of Seismicity in Volcanic Regions,

Fig. 2 Histogram of b -value anomaly centroid depths beneath volcanoes



vesicles: For Mt. Etna, Italy, Trigila et al. (1990) report a range of 85–225 Mpa pressure range for Plagioclase stabilization in basaltic lavas and it has been suggested that the feeding system for magmas at Etna may be composed of different reservoirs; at the Katmai cluster, USA, Coombs and Gardner (2001) estimated a range of 40–100 Mpa for Plagioclase stabilization in andesite-dacite-rhyolite magmas erupted during the 1912 eruption from Novarupta; at Redoubt, USA, Gerlach et al. (1994) suggested that SO_2 is stored as vapor at 6–10 km depth in the magma before the 1990 eruptions and perhaps at shallower levels (4–6 km) prior to the 2009 eruptions (Coombs and Bacon 2012); at Nevado del Ruiz, Colombia (Stix et al. 2003), it has been suggested that ascent of volatile-bearing magma took place from 9 to 15 km depth and remained at shallow residence depth of <3 km; at Mt. Pinatubo, Philippines, estimation of volatile contents in the range 5.1–6.4 % suggested a vesiculation depth of 6 km for pre-eruptive magmas and stabilization of olivine at 2.2 kbar or a depth of magma storage in the range 7–11 km was also inferred (Bernard et al. 1996); at Galeras, Colombia (Stix et al. 1997), gas exsolution might have taken place at 20–100 Mpa or 0.8–3.9 km depth; and at Makushin, USA, scoriae within mafic ignimbrite layers were analyzed (Goldberg 2010) and a stabilization pressure of pyroxenes

in a range of 1.26–2.12 kbar has been suggested. Evidence from Geology, Petrology, and Geochemistry thus indicates that several processes take place at different depths beneath volcanoes.

From Table 1 it is also evident that seismicity studies have also revealed that beneath many volcanoes, b is anomalously high at different depths. All volcanoes listed (17) exhibit b -value anomalies embedded in an otherwise region of normal-to-low b -values. With respect to the distribution of b -values at depth, most volcanoes studied (16, b -values at Yellowstone were not mapped at depth) have at least one anomaly, and 11 out of 16 volcanoes exhibit more than one anomalous region. The centroid depths for b -value anomalies vary in the range 2–19 km and their distribution are shown in Fig. 2.

From Fig. 2 it is evident that, in many cases, volcanoes exhibit two anomalies at different depths: one shallow anomaly is common in the range 2–4 km and a deeper one appears in the range 7–8 km. The first anomaly range includes not only the depth at which volatiles in magma with 4 % weight fraction may start to exsolve but also the region where the existence of open cracks is plausible. At depths similar to the range 7–8 km, alternate geophysical evidence and geology studies support the nearby presence of magma bodies. At Galeras Volcano, Colombia, the shallow plumbing magmatic system has

been conceived as a vertically elongated feature such as a reservoir extending down to 5–6 km depth or a fractured plug beneath the active cone that overpressurizes and acts as the source of material during Vulcanian explosions (Stix et al. 2003) and a deeper magma reservoir below 4 km depth.

Regarding the cause of anomalous b -values, there is general agreement that several factors may be involved; these are summarized as follows:

Accepted Knowledge on Seismic b -Values at Volcanoes

- That the frequency-magnitude distribution is related to the heterogeneity (degree of cracking) in the upper crust and can be appropriately measured by a power law expressed in the Gutenberg-Richter law. The parameters of this law are understood as the seismic productivity of the volume (“ a -value”) and the mean magnitude of the linear part (“ b -value”).
- Estimating the b -value is equivalent to estimating the mean magnitude and it is reasonable to assume the mean magnitude as proportional to mean crack length.
- Heterogeneity in volcanic areas can be pervasive and may be related not only to the nature and amount of cracking that the magma system or volcano edifice may have but also to the nature of rocks and deposits that form volcanoes, usually a combination of hydrothermally altered lava flows and pyroclastic deposits.
- Although volcanic areas were usually reported as having higher-than-normal b -values, with respect to a “global” average of $b = 1.0$, it is now known that instead, volcanoes exhibit regions of anomalously high b -values within a “normal” crust.
- That magma (molten rock inside the earth) cannot support shear stresses and therefore earthquakes cannot occur there. b values give information on faults and fractures in the vicinities of magma chambers. The resolution of earthquake locations may not be still enough to reveal details on the configuration of magma systems, although the concept of complex configurations in the magma and plumbing systems beneath volcanoes is generally accepted.
- That factors that may contribute to variations in b are:
 - (a) To increase b :
 1. Increase in material heterogeneity. At volcanoes this can be associated to increase in cracks, caused by vesiculation and fragmentation of ascending magma, processes that likely take place at relatively shallow levels (roughly 4 km depth for magmas with 4 wt. % H_2O).
 2. Increase in pore pressure caused by magma nearing groundwater (which would imply decreasing the effective stress). This process may also likely take place at relatively shallow levels.
 3. Increase in the thermal gradient. Quite possible near magma systems or intrusions, because magmas may reach temperatures near 1,200 °C.
 - (b) To decrease b :
 1. Increase in applied shear stresses. Difficult to model this near volcanoes, but plausible along the faults that are common in volcanic areas. Zones of lower-than-normal b -values have been associated to regions of high stress load and therefore areas where large earthquake nucleation could take place. In this perspective, seeking for regions of lower-than-normal b -values is just as important and may have hazard implications.
 2. Increase in effective pressure or decrease in pore pressure. Could be conceivable during or after volcanic eruptions.
 - (c) Either to increase or decrease b :
 1. Changes in the configuration of the detection network or changes in analyses procedures. These may introduce artifacts in the catalog. For example, a change in the way magnitudes are determined may cause a “stretching” of the magnitude scale, offsetting b toward lower values.

It has been suggested that processes such as vesiculation, fragmentation, and magma intrusions can affect the crack distribution beneath a volcano. Gas exsolution and subsequent bubble nucleation and growth inside magmas take place under colossal pressure conditions. Bubbles can increase in numbers and accumulate near the top of magma reservoirs increasing pressure on the host rock and facilitating cracking. Also, magmas intrude through fractures and weak zones of the crust and this process may cause new cracks and fracture propagation thus changing the homogeneity of the rock masses.

From Table 1 a scatter plot of the centroid depths for b -value anomalies versus the pressure at which mineral stabilization and/or gas exsolution might take place or the estimated depth for magma storage (from Petrology, Geochemistry, and Geology data) was built. Errors in the centroids of b -value anomalies were taken as the maximum and minimum observable extents at depth from published and authoritative material. Results are shown in Fig. 3.

Based on Fig. 3 (right panel), a positive relation between the depth of deeper b -value anomalies and the magma process beneath volcanoes is mildly revealed for a group of eight volcanoes: Mammoth Mt. (Long Valley Caldera), Nevado del Ruiz, Mt. St. Helens, Katmai-Novarupta, Redoubt, Mt. Pinatubo, Mt. Etna, and Miyakejima. These include diverse-sized volcanic systems from a variety of geological environments: six continental arc volcanoes (formed above oceanic-continental subduction zones), a hot spot volcano, and an island arc volcano (formed above oceanic-oceanic subduction zone).

b -Values at Galeras Volcano, Colombia, and Implications for Earthquake Hazard and Risk

Galeras Volcano, one of the most active volcanoes in Colombia, looms over the city of San Juan de Pasto (pop. 400,000) which lies just a few kilometers east of the active crater. Not only have repeated eruptions of Galeras affected life and property over the years and volcanic

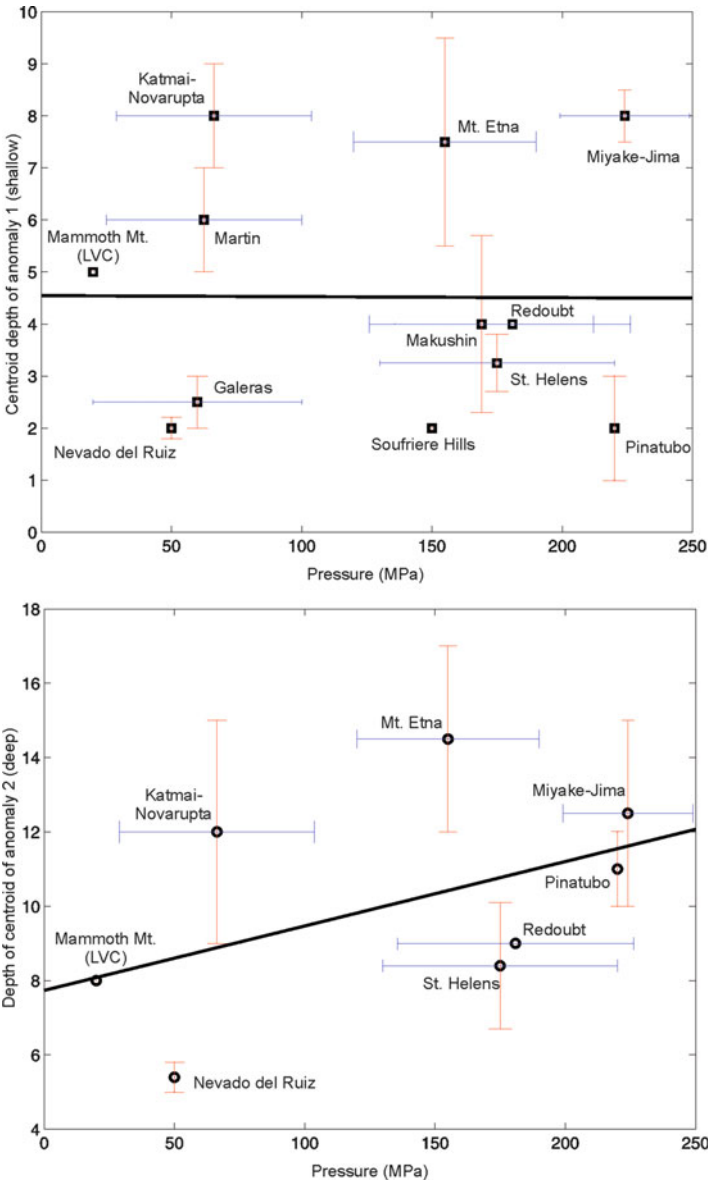
processes do represent a significant hazard for the region (INGEOMINAS 1997), but also earthquakes possibly associated to volcanic-tectonic interactions have caused damage and fatalities in recent times. Thus, seismically characterizing the region near Galeras is of relevance and mapping of b -values offers the opportunity to identify zones of smaller-than-average earthquakes as well as regions where relatively large earthquakes are typically produced.

Using preliminary earthquake locations between September 1995 and June 2002, a first approach to spatially map the earthquake size distribution around Galeras was taken (Sánchez et al. 2005). Their results of 3D mapping readily allowed to identify a prominent high b -value anomaly beneath the active crater and vertically elongated down to 5 km depth which was inferred to be the region adjacent either to the volcano's conduit, the shallow magma reservoir, or alternatively the remnants of a semi-brittle intrusion. In map view, this anomalous feature was basically surrounded by regions of normal-to-low b -values. Of particular interest here is the cluster of earthquakes located northeast of Galeras, where the source of three swarms that included damaging earthquakes ($M_L \leq 4.7$) was identified during 1993–1995 (April 1993, November–December 1993, and February 1995). This source falls within a region of low b -values detected preliminarily in the early 2005 study (Fig. 4).

Recent data for Galeras has been analyzed and the catalog of earthquakes for the period 2004–2013 was used to spatially map b -values using both the entire catalog data and a subset of it (2008–2010), during which seismicity rates were observed to be fairly stable. Results are presented in Fig. 5. In map view the region north-northeast of Galeras still appears as a region of low b -values and remains as a source of moderate-sized earthquakes (a sequence of four $4.6 < M_L < 4.7$ earthquakes occurred there during mid 2005). The active crater region remains highly productive, but it does not show as a zone of smaller-than-average earthquakes, on the contrary, falls within an area of low b -values and was the source of two separate $4.6 < M_L < 4.7$ earthquakes in 2006 and 2010.

Frequency-Magnitude Distribution of Seismicity in Volcanic Regions, Fig. 3

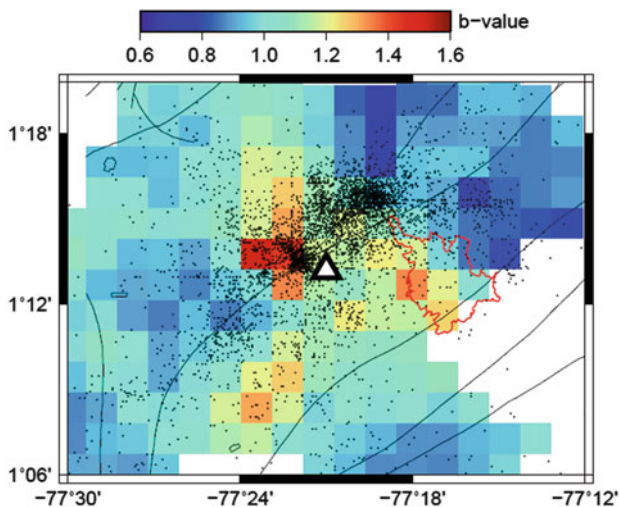
Scatter plots of the depths of *b*-value anomalies observed at volcanoes versus pressure conditions calculated for gas exolution or mineral stabilization (from Geology, Petrology and Geochemistry data). Many volcanoes from Table 1 exhibit more than one *b*-value anomaly usually at different depths. *Left*: Centroid depth of “shallower” anomalies. *Right*: Centroid depth for “deeper” anomalies. Thick black lines in both panels show a basic linear fit ($R = 0.0$ for the data on the left panel, $R = 0.5$ for data on the right panel)



It is conceivable that once a volcano or a region of the crust has acquired a given crack distribution, this is difficult to change and thus the distribution of *b*-values would be fairly constant through time. At volcanoes, however, multiple intrusions are common and therefore a much fractured host rock may be plausible. For example, the well-recorded earthquake swarms beneath the Teishi Knoll Volcano (off Ito, Japan) were associated to magma intrusions and the detected change of *b*-value with time

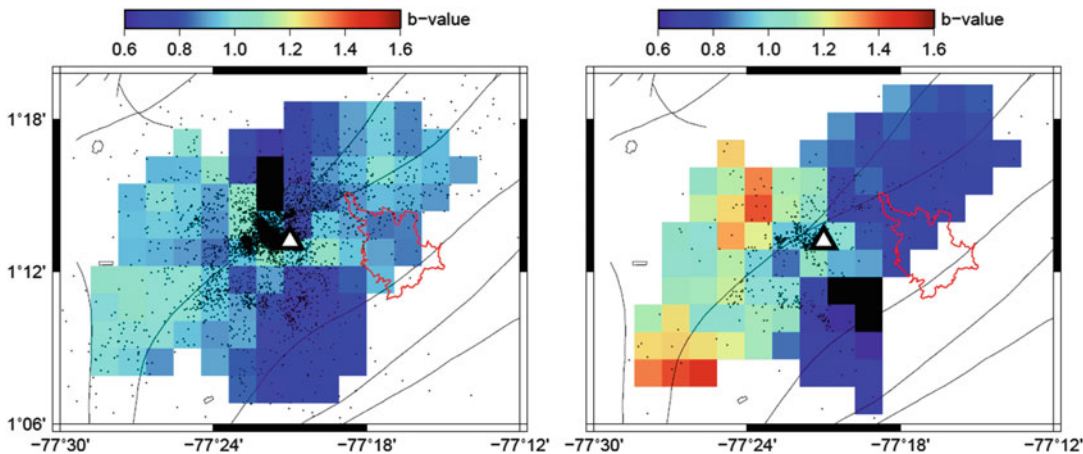
was attributed to new cracks (Wyss et al. 1997). Thus, changing the heterogeneity of the shallow crust near volcanoes, or the difficulty in causing changes in such systems, could be viewed as a matter of time scale. The results described for Galeras could signal a change in conditions of the shallow crust near the active crater.

As it can be seen in Figs. 4 and 5, the Galeras region is crossed by several faults in a general southwest-northeast direction and the volcano



Frequency-Magnitude Distribution of Seismicity in Volcanic Regions, Fig. 4 Map of b -values in the Galeras region for the period 1995–2002. High b -values are mapped as regions of hot colors (red-orange), regions of low b -values appear as areas of cold colors (blue). Dots: earthquake epicenters. Triangle: active vent. Continuous

black lines: faults. Thick red contour: limits of San Juan de Pasto. Swarms of damaging earthquakes were produced between 1993 and 1995 in the area of low b -values northeast of the volcano (Modified from Sánchez et al. 2005)



Frequency-Magnitude Distribution of Seismicity in Volcanic Regions, Fig. 5 Maps of b -values in the Galeras region. *Left*: period 2004–2013, using all earthquakes with $M_L > 0.8$. Relatively large earthquakes ($4.6 < M_L < 4.7$) occurred in 2006.16 (near the crater), 2010.64 (just east of the crater), and during 2005.64–2005.65 (north-northeast of the crater, four earthquakes). *Right*: period 2008–2010, during which the

seismicity rate was fairly constant. The different b -values in the western portions of Galeras between the different datasets may be an artifact caused by the different sampling around each node, necessary because of the lower number of earthquakes in the shorter period 2008–2010. The region of interest, northeast of the active crater remains mostly unchanged. Conventions as in Fig. 4

itself is spatially related to these structures, particularly to the splay of the two faults near the active crater. It is conceivable that magma has used faults as pathways to the surface, making

this area a particularly complex one in terms of stresses and earthquakes.

Interestingly, the area southeast of the volcano which shows persistently as a region of

very low b -values has not produced relatively large earthquakes damaging to San Juan de Pasto.

The concept that faults include both locked and unlocked portions, the former conceived as asperities that resist sliding and the latter known as creeping segments, implies that asperities are regions that accumulate stress whereas in creeping segments stresses are constantly relieved. The distribution of b -values near active faults has been successfully used to mark the sites of asperities and creeping segments along faults (Wiemer and Wyss 2002), with areas of low b -values signaling asperities and areas of higher-than-normal b -values representing creeping segments. Volcanic regions are always affected by active faults, and considering that moderate-sized earthquakes have been detected near the volcano, it is reasonable to assume that near Galeras Volcano the distribution of b -values may help in identifying areas of relatively high stress load, and from Figs. 4 and 5, it may be advisable to use a higher-quality dataset and map b -values along the faults that have been documented there. Such analyses might have implications for hazards and risk not only for the city of San Juan de Pasto but also for other towns surrounding the volcano.

Summary

The b -value (slope of the frequency-magnitude distribution of earthquakes) around active volcanoes has been mapped at a number of volcanoes around the world and, with information from published sources on types of magmas (Geology data) and pressure conditions inferred for gas exsolution, vesiculation, and mineral stabilization within magma systems (Petrology and Geochemistry data), a possible relation between the depth of b -value anomalies with magma processes is presented. The body of generally accepted knowledge on b -values at volcanoes is summarized and recent data for b -values at Galeras Volcano, Colombia, are analyzed in terms of volcano and earthquake hazards and risk.

Cross-References

- [Earthquake Location](#)
- [Earthquake Magnitude Estimation](#)
- [Seismic Anisotropy in Volcanic Regions](#)
- [Seismic Monitoring of Volcanoes](#)
- [Seismic Tomography of Volcanoes](#)
- [Volcanic Eruptions, Real-Time Forecasting of](#)
- [Volcano-Tectonic Seismicity of Soufriere Hills Volcano, Montserrat](#)

References

- Amma-Miyasaka M, Nakagawa M (2003) Evolution of Deeper Basaltic and Shallower Andesitic Magmas during the AD 1469–1983 Eruptions of Miyake-Jima Volcano, Izu-Mariana Arc: Inferences from Temporal Variations of Mineral Compositions in Crystal-Clots, *J. Petrology* 44:2113–2138.
- Barclay J, Rutherford MJ, Carroll MR, Murphy MD, Devine JD, Gardner J, Sparks RSJ (1998) Experimental phase equilibria constraints on pre-eruptive storage conditions of the Soufrière Hills magma. *Geophys Res Lett* 25:3437–3440
- Bernard A, Knittel U, Weber B, Weis D, Albrecht A, Hattori K, Klein J, Oles D (1996) Petrology and geochemistry of the 1991 eruption products of Mount Pinatubo (Luzon, Philippines). In: Newhall CG, Punongbayan RS (eds) Fire and mud: eruptions and lahars of Mount Pinatubo. Philippine Institute of Volcanology and Seismology/University of Washington Press, Quezon City/Seattle, pp 767–798
- Bridges DL, Gao SS (2006) Spatial variation of seismic b -values beneath Makushin Volcano, Unalaska Island, Alaska. *Earth Planet Sci Lett* 245:408–415
- Burkett SM (2007) Geomorphic mapping and petrography of mammoth mountain, California. MS thesis, University of New York at Buffalo, Buffalo, 114 pp
- Coombs ML, Bacon CR (2012) Using rocks to reveal the inner workings of Magma Chambers below volcanoes in Alaska's National Parks. In: Shah M (ed) Alaska Park science-volcanoes of Katmai and the Alaska Peninsula. Alaska Geographic, Anchorage, pp 27–33
- Coombs ML, Gardner JE (2001) Shallow-storage conditions for the rhyolite of the 1912 eruption at Novarupta, Alaska. *Geology* 29:775–778
- Gerlach TM, Westrich HR, Casadevall TJ, Finnegan DL (1994) Vapor saturation and accumulation in magmas of the 1989–1990 eruption of Redoubt Volcano, Alaska. *J Volcanol Geotherm Res* 62:317–337
- Goldberg AR (2010) Petrologic and volcanic history of Point Tebenkof ignimbrite, Unalaska, Alaska. In: de Wet AP (ed) Proceedings of the twenty-third annual Keck research symposium in geology. Keck Geology Consortium, Houston, pp 335–340

- Gutenberg B, Richter CF (1944) Frequency of earthquakes in California. *Bull Seismol Soc Am* 34:185–188
- INGEOMINAS (1997) Mapa de amenazas volcánicas del Galeras-tercera versión. Instituto de Investigaciones en Geociencias, Minería y Química. <http://www.ingeo Minas.gov.co/Pasto/Volcanes/Volcan-Galeras/Mapa-de-amenazas.aspx>. Accessed 4 June 2013
- Jolly AD, McNutt SR (1999) Seismicity at the volcanoes of Katmai National Park, Alaska; July 1995–December 1997. *J Volcanol Geotherm Res* 93:173–190
- Kushiro I (1991) Origin of volcanic rocks in Japanese island arcs. *Episodes* 14:258–263
- Murru M, Montuori C, Wyss M, Privitera E (1999) The location of magma chambers at Mt. oEtna, Italy, mapped by b -values. *Geophys Res Lett* 26:2553–2556
- Pallister JS, Thornber CR, Cashman KV, Clynne MA, Lowers HA, Mandeville CW, Brownfield IK, Meeker GP (2008) Petrology of the 2004–2006 Mount St. Helens lava dome – implications for magmatic plumbing and eruption triggering. In: Sherrod DR, Scott WE, Stauffer PH (eds) *A volcano rekindled: the renewed eruption of Mount St. Helens, 2004–2006*, U.S. Geological Survey professional paper 2008–1750. U.S. Geological Survey, Vancouver, pp 647–702
- Power JA, Wyss M, Latchman JL (1998) Spatial variations in the frequency-magnitude distribution of earthquakes at Soufriere Hills Volcano, Montserrat, West Indies. *Geophys Res Lett* 25:3653–3656
- Rierola M (2005) Temporal and spatial transients in b -values beneath volcanoes. Diploma thesis, ETH Zurich, Institute of Geophysics, Zurich, 77 pp
- Rodríguez SP (2003) Mapeo tridimensional del valor b en el Volcán Nevado del Ruiz. Trabajo de Grado, Universidad de Caldas, Manizales, Colombia, 72 pp
- Sánchez JJ, McNutt SR, Power JA, Wyss M (2004) Spatial variations in the frequency-magnitude distribution of earthquakes at Mount Pinatubo Volcano. *Bull Seismol Soc Am* 94:430–438
- Sánchez JJ, Gómez DM, Torres RA, Calvache ML, Ortega A, Ponce AP, Acevedo AP, Gil-Cruz F, Londoño JM, Rodríguez SP, Patiño J de J, Bohórquez OP (2005) Spatial mapping of the b -value at Galeras Volcano, Colombia, using earthquakes recorded from 1995 to 2002. *Earth Sci Res J* 9:30–36
- Stix J, Torres RA, Narváez L, Cortés GP, Raigosa J, Gómez DM, Castonguay R (1997) A model of vulcanian eruptions at Galeras volcano, Colombia. *J Volcanol Geotherm Res* 77:285–303
- Stix J, Layne GD, Williams SN (2003) Mechanisms of degassing at Nevado del Ruiz Volcano, Colombia. *J Geol Soc* 160:507–521
- Trigila R, Spera FJ, Aurisicchio C (1990) The 1983 Mount Etna eruption: thermochemical and dynamical inferences. *Contrib Miner Petrol* 104:594–608
- Wiemer S, McNutt SR (1997) Variations in the frequency-magnitude distribution with depth in two volcanic areas: Mount St. Helens, Washington, and Mt. Spurr, Alaska. *Geophys Res Lett* 24:189–192
- Wiemer S, Wyss M (2002) Mapping spatial variability of the frequency-magnitude distribution of earthquakes. *Adv Geophys* 45:259–302
- Wiemer S, McNutt SR, Wyss M (1998) Temporal and three-dimensional spatial analyses of the frequency-magnitude distribution near Long Valley Caldera, California. *Geophys J Int* 134:409–421
- Wyss M, Shimazaki K, Wiemer S (1997) Mapping active magma chambers by b values beneath the off-Ito volcano, Japan. *J Geophys Res* 102:20,413–20,422
- Wyss M, Klein F, Nagamine K, Wiemer S (2001) Anomalous high b -values in the south flank of Kilauea Volcano, Hawaii: evidence for the distribution of magma below Kilauea's east rift zone. *J Volcanol Geoth Res* 106:23–37

Friction Dampers for Seismic Protections of Steel Buildings Subjected to Earthquakes: Emphasis on Structural Design

Lucia Tirca

Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC, Canada

Synonyms

Damping; Earthquake; Energy dissipation; Friction dampers; Seismic design; Steel buildings

Introduction

To reduce the seismic demand, researchers have proposed to incorporate supplemental energy dissipation devices into the structural system of buildings. According to the primary dissipation mechanism, supplemental energy dissipation devices are grouped into two categories: hysteretic and viscoelastic. Hysteretic devices rely on the relative displacements of components within the device and are typically based on either metallic yielding or frictional sliding, while viscoelastic devices are velocity dependent. More specifically, friction devices dissipate energy

through the relative sliding developed between two solid interfaces. Depending on the type of friction devices, they could be installed in line with single-diagonal or chevron steel braces, at the intersection of X-bracing system, and in parallel with the beam located at the top of chevron bracing system. The activation of slip forces that characterize the designed friction dampers occur simultaneously with the maximum internal forces allowable to develop in the system during the ground motion excitation. The building reaches the peak interstorey drift when the available slip distance provided by friction damper devices was consumed. “The forces generated by these devices installed in the structural members are usually in phase with the internal forces resulting from ground motion shaking” (Christopoulos and Filiatrault 2006).

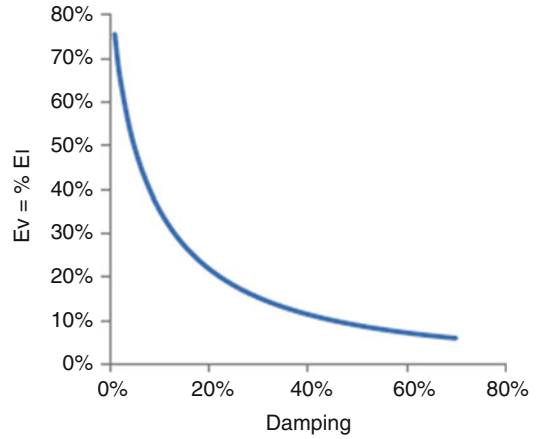
The total input energy, E_I , induced by a seismic event into a structural system can be expressed as a summation of kinetic energy, E_k , cumulative strain energy, E_S , inherent damping, E_D , and the hysteretic damping E_h of the seismic force resisting system (SFRS). In this study, E_h is the damping induced by friction devices. The energy balance equation is:

$$E_I = E_k + E_S + E_D + E_h \quad (1)$$

The kinetic and cumulative strain energy are accumulated into the primary structural system and rely on structural damage (Akiyama 2000; Tirca 2009), while the system is damped by both E_D and E_h , which are amplitude-dependent. In general, the contribution of E_D and E_h is related to the amount of post-yielding response and Eq. 1 can be rearranged as follows:

$$E_k + E_S = E_I - (E_D + E_h) \quad (2)$$

In Eq. 1, the term $(E_k + E_S)$ expresses the vibrational energy, E_v , or more specifically the potential damage energy, while $(E_D + E_h)$ is the energy dissipated by viscous damping in the structural members and supplemental devices. Thus, by adding damping into a structural system the elastic vibration energy diminishes, while structural

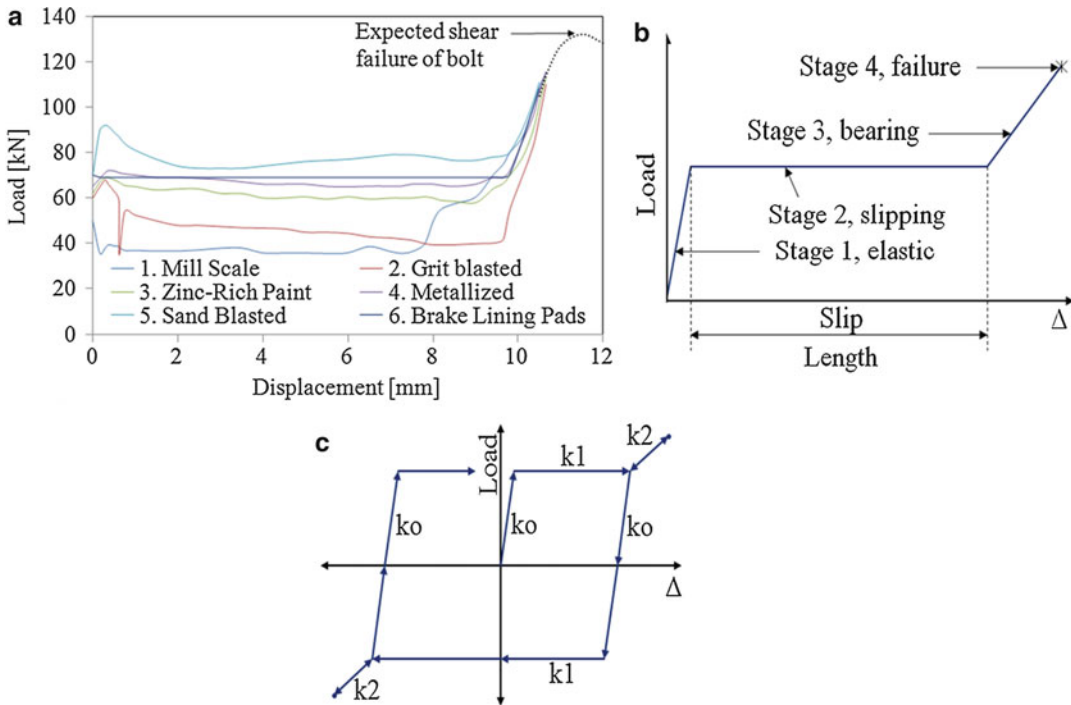


Friction Dampers for Seismic Protections of Steel Buildings Subjected to Earthquakes: Emphasis on Structural Design, Fig. 1 Vibrational energy versus supplemental damping

members are protected from damage associated with permanent deformations. In this light, the main design objective is to minimize the difference between the seismic input energy and that dissipated by the dampers (Christopoulos and Filiatrault 2006). The variation of E_v as a function of supplemental damping is depicted in Fig. 1.

Pioneering work on friction devices was conducted by Pall (1979) and Pall and Marsh (1981). Since then, several types of friction dampers have been developed and studied in the literature. These devices differ in their mechanical shape and materials used for the sliding surfaces.

Pall friction dampers dissipate energy through friction developed by the relative sliding within two surfaces in contact which are clamped by posttensioned bolts. In order to obtain stable rectangular hysteresis loops similar to that of Coulomb friction, different types of surface treatment and lining material were studied experimentally. After investigating the response of slip bolted joints under monotonic loading, Pall reported that the most stable behavior was obtained when brake lining pads in contact with mill scale surface on plate was chosen (Fig. 2a). Nevertheless, minor differences between the static and dynamic friction coefficient were observed. However,



Friction Dampers for Seismic Protections of Steel Buildings Subjected to Earthquakes: Emphasis on Structural Design, Fig. 2 Response of slip-bolted

joint: (a) monotonic test, (b) back-bone curve, and (c) hysteretic behavior (After Pall 1979)

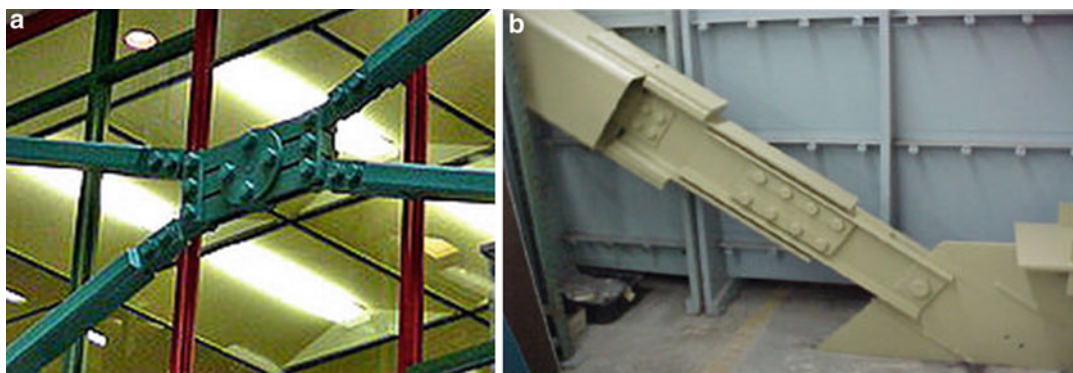
under large seismic excitations, the posttensioned bolts of friction dampers may impact into the end of slotted hole and undergo bearing or even bolt shear failure. The hysteretic behavior of friction damper shown in Fig. 2b is similar to that of an elasto-perfectly plastic system. Herein, the back-bone curve is composed of four segments: elastic, slipping, bearing, and bolt shear failure. When the demand is higher than the available slip length which is equal to the length of slotted hole, a sudden increment in storey shear forces accompanied by decreasing of Coulomb damping is encountered. The hysteresis behavior depicted in Fig. 2c in terms of slip load versus the slip length, Δ , shows rectangular symmetrical loops which are largely influenced by the fluctuation of friction coefficient during the slipping stage. The elastic stiffness, k_o , shown in Fig. 2c, is the stiffness of the attached brace member.

Due to their efficiency (Pall and Pall 2004), Pall friction dampers were employed in more than 40 new and retrofit buildings in Canada,

United States, and India (Christopoulos and Filiatrault 2006). Examples of Pall friction damper installed in an X-bracing system and single-diagonal brace are showed in Fig. 3.

In the last two decades, the following types of friction devices have been developed: slotted bolted connections (Fitzgerald et al. 1989; Grigorian et al. 1993; Tremblay 1993), Sumitomo devices developed by Sumitomo Metal Industries Ltd. of Japan and reported by Aiken and Kelly (1990), energy dissipating restraint damper (EDR) developed by Flour Daniel Inc. (Nims et al. 1993), friction variable damper developed based on the EDR damper (Zhou and Peng 2009), friction damper developed by Mualla (2000), etc.

The energy dissipated by the Sumitomo friction damper is due to friction generated when friction pads, made of cooper alloy with graphite plug inserts, slide directly on the inner surface of the outer cylinder. The precompressed internal spring, incorporated into the device, applies a



Friction Dampers for Seismic Protections of Steel Buildings Subjected to Earthquakes: Emphasis on Structural Design, Fig. 3 Pall friction dampers manufactured by Pall Dynamics Lmt. Montreal: (a)

installed in X-bracing at Concordia Library, Montreal, and (b) installed in single-diagonal brace at Boeing Commercial Airplane Factory Everett, USA (Courtesy Dr. Pall)

F

force that is converted through the action of inner and outer wedges into a normal force on the friction pads. Sumitomo device can be installed in parallel to the beam located on top of chevron bracing system or in-line with steel braces. As reported by Aiken and Kelly (1990), Sumitomo dampers were used in two high-rise buildings in Japan in order to resist small intensity earthquakes and ground floor vibrations.

The EDR damper is similar to Sumitomo device. It includes the following components: an internal spring, steel compression wedges, bronze friction wedges, stops at both spring's ends, and the outer steel cylinder. Its functioning comprising the generating slip force depends on the length of the internal spring. In addition, the normal force acting on the cylinder wall and the friction force developed between the bronze friction wedges and the inner surface of the outer cylinder determine the slip force in the device. As noted by Nims et al. (1993), the EDR device can produce a wide range of hysteretic behavior such as double-flag shape and triangular lobed shape. Both types of hysteresis loops are self-centering. By using the same principles, Zhou and Peng (2009) proposed a new friction variable damper where a sliding shaft and a friction ring were used to replace the spring and wedges. In addition, two zones with high and low friction coefficient have been defined in the internal walls of the outer cylinder.

The friction damper proposed by Mualla (2000) was added to connect the top part of chevron braces to the mid-span of a moment resisting frame beam. This friction damper, patented by DAMPTECH Denmark, consists of three steel plates that are able to rotate against each other around a pre-stressed bolt which passes through. Two circular friction pad discs inserted between the aforementioned steel plates provide dry friction lubrication and ensure stable friction forces. As reported by Liao et al. (2004), a full scale test conducted on a three-storey moment resisting frame with added chevron bracing system and Mualla's dampers was investigated. The total weight of the frame was about 34 t. The bracing system consisted of bar members that were pre-tensioned in order to avoid buckling. Due to the application of lower scaling simulated ground motions, no damage of structural members was reported during the conducted experimental tests.

Design Provisions

Since 1980, seismic response control techniques have been used as complementary solutions to the existing SFRSs. Despite of their wide-spread applications, design guidelines for structures equipped with supplemental energy dissipation devices (hysteretic and viscoelastic) are still in evolving phases. This study refers to the design

philosophy of structures with incorporated friction damper devices (FDD). As mentioned above, FDDs are hysteretic devices. For example, buckling-restrained braces are also known as hysteretic damper devices and are designed to dissipate part of vibration energy through yielding of the core plate.

In general, hysteretic devices are installed in braces and are able to undergo large inelastic response, while they dissipate most of the hysteretic energy. The purpose of installing these devices into the structural system is to maintain the main structure either elastic or within low inelastic deformations. Both, the main frame and the supplemental energy dissipation system, share the same deformation, which in turn is that of the entire system. It is important to assure stable response of these devices under dynamic loading. Thus, the added dampers installed in new or retrofit buildings should yield or slip before the shear resistance of the main structure is reached.

In North America, the first design guidelines addressing provisions for passive energy dissipated devices were introduced in FEMA 356 (2000) and FEMA 450 (2003) that refer to the seismic rehabilitation of buildings and seismic design for new buildings, respectively. Later on, FEMA 450 was replaced by FEMA-P 750 (2009). The aforementioned design guidelines were incorporated in Chapter 14 of ASCE/SEI 41-13 (Seismic evaluation and retrofit of existing buildings), as well as in Chapter 18 of ASCE/SEI 7-10 standard. It is noted that *Commentary J* of NBCC 2010 (National Building Code of Canada) includes two clauses in regard to seismic design with supplemental energy dissipated devices. However, both CSA/S16-2009 and ANSI/AISC 360-2010 standard provide design regulations for the “ductile buckling-restrained braced frames” system.

In Europe, the Eurocode 8 part 1 (CEN 2004) contains Chapter 10 entitled *Base isolation*, but this chapter does not cover regulations for passive energy dissipation devices installed into several storeys of building’s structure.

The aforementioned guidelines specify general requirements, analysis procedures, required

testing program, etc. A brief review of these guidelines with highlights on friction damper devices design is presented below.

FEMA P-750 and ASCE/SEI 7-10 Provisions

Design Philosophy

General requirements:

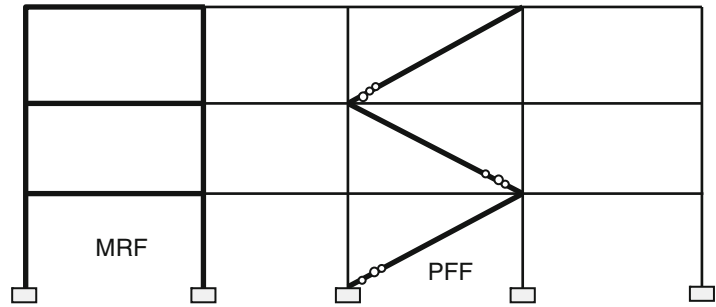
- Structures with a damping system have a seismic force resisting system that provides a complete load path. The base shear used to design the employed SFRS shall not be less than $V_{\min}$, where $V_{\min}$ is the greater of:

$$V_{\min} = V/B_{v+1} \text{ and } V_{\min} = 0.75 V.$$

Herein, V is the seismic base shear in the direction of calculation determined in accordance with the code procedure corresponding to the selected SFRS and B_{v+1} is the numerical coefficient set for effective damping reduction factor that is equal to the damping provided by supplemental dampers in addition to the inherent damping (2 % for steel structures). For irregular building structures and for damping system that has less than four damper devices per floor disposed in the direction of loading, the minimum base shear will not be less than $1.0 V$.

- Damping system may be used in addition to the SFRS in order to meet interstorey drift limits and it may be located external or internal to the structure, while it shares or not the members of SFRS. In Fig. 4 is illustrated a SFRS consisting of one bay moment resisting frame (MRF) and a damping system composed of diagonal braces with installed FDs. The FD system is located independently to the MRF system.
- Energy dissipating devices must be tested before installation for displacement and slip force corresponding to seismic demand that is required by code.
- Linear static and response spectrum analysis methods for design are accepted for regular

Friction Dampers for Seismic Protections of Steel Buildings Subjected to Earthquakes: Emphasis on Structural Design,
Fig. 4 Configuration of SFRS and damping system (no shared elements)



structures that have at least four damper devices per each floor located in one principal direction of the building. Like MRF or braced frame structures, damped structures yield during the design ground motions, while they need to behave elastically under the wind loads.

- For detailed design and seismic response investigations, nonlinear dynamic time-history analyses are recommended.
- Dampers are modeled to perform in the plastic range. The computer model should be calibrated based on experimental test results. For designs in which the SFRS is expected to yield, the postyield behavior of the structural elements must be modeled explicitly.

In the ASCE/SEI 7-10 provisions, damping system is defined as being “the collection of structural elements that includes all the individual damping devices, all structural elements or bracing required to transfer forces from damping devices to the base of the structure, and the structural elements required to transfer forces from damping devices to the seismic force-resisting system.” For details of design procedure, the reader is referred to the ASCE/SEI 7-10 document. It is noted that other design methodologies can be applied to design structures equipped with friction dampers.

Design of Structures Equipped With Pall Friction Dampers

As noted above, metallic dampers and friction dampers are hysteretic dampers. The design of structures equipped with metallic dampers or friction dampers is similar. The difference

consists of changing the yield load corresponding to metallic dampers with the slip load of friction dampers. As mentioned above, friction dampers are installed in braces and are designed not to slip under the wind forces. On the other hand, in order to increase the building performance under earthquake loads, systems composed of friction-damped braces are added to the MRF structures, while the MRF system should possess enough strength, stiffness, and ductility. Furthermore, all members of the SFRS and those attached to FDs should be designed to carry the internal forces developed when slip forces are activated. Thus, the application of capacity design checks to structural members is required. According to Christopoulos and Filiatrault (2006) the design of structures equipped with friction dampers can be divided in four steps:

- Calculate the demanded slip force and size the in-line brace by employing simplified methods (e.g., the equivalent static force procedure). Then, the demanded peak interstorey drift (e.g. linear dynamic analysis methods) should be evaluated.
- Optimize the design of friction dampers and adjacent members.
- Apply the capacity design and size all structural members to carry the designed slip forces.
- Use the nonlinear time-history dynamic analysis to check the design performance of building structure under an ensemble of selected ground motions scaled to match the design response spectrum corresponding to 2 % probability of exceedance in 50 years (NBCC 2010).

A review of different design procedures from the literature is discussed below. However, in all design procedures, the building structure is composed of a bare frame system and a supplemental damping system, whilst its response is analyzed before and after the friction dampers are activated.

Design Procedures from the Literature

The first design procedure for MRF system equipped with friction damped braces was proposed by Baktash and Marsh (1986). They considered a single storey MRF structure with friction damped braced bays. Based on this procedure, the slip force is activated when yielding in the moment resisting frame brace is reached. For a single bay MRF frame that was designed based on the principle “strong column weak beam”, the shear force resisted by the frame action, V_s is related to the plastic moment capacity of the MRF beam, M_p . Herein, $M_p = V_s h/2$, where h is the storey height. Thus, the proposed design method is based on the assumption that the lateral shear force leading the damper in-line with brace to slip must be equal to the shear force causing the MRF to yield and is expressed by the following equation:

$$V_s = 2M_p/h \quad (3)$$

The corresponding shear deflection of the storey, Δ_s is:

$$\Delta_s = (V_t - V_{br})/k_u \quad (4)$$

where V_t is the total shear force exerted by the frame and braces, V_{br} is the total shear force exerted by braces alone, and k_u is the lateral stiffness of the bare MRF.

The hysteretic energy dissipated by friction devices E_f is:

$$E_f = V_{br}\Delta_s = V_{br}(V_t - V_{br})/k_u \quad (5)$$

By differentiating Eq. 5 with respect to V_{br} and setting it equal to zero, the maximum energy dissipated by friction devices is obtained:

$$\delta E_f / \delta V_{br} = V_t/k_u - 2V_{br}/k_u = 0 \quad (6)$$

From Eq. 6, the shear force exerted by braces is:

$$V_{br} = 0.5V_t \quad (7)$$

Thus, to maximize the energy dissipated by friction devices, the total shear force must be equally shared between friction devices and the bare frame (MRF). In addition, it was reported that the optimum slip force of friction devices installed in structure depends on the structural characteristics only and not on the ground motion intensity (Baktash and Marsh 1987).

Filiatrault and Cherry (1988) proposed to determine the optimum activation of friction devices based on minimizing the *Relative Performance Index*, RPI, devised from the application of the energy concept.

$$RPI = \frac{1}{2}[SEA/SEA_0 + U_{\max}/U_{\max 0}] \quad (8)$$

where SEA is the strain energy area of all structural members of a friction damped system, SEA_0 is the strain energy area corresponding to a zero activation force, $U_{\max}$ is the maximum strain energy stored in all structural members of a friction damped system, and $U_{\max 0}$ is the maximum strain energy for a zero slip force. The values resulted for the RPI yield to three cases: (i) $RPI = 1$ corresponds to the response of a bare frame structure; (ii) $RPI < 1$ corresponds to the response of a damped structure which is smaller than the response of the bare frame structure; (iii) $RPI > 1$ corresponds to the response of a damped structure which is larger than the response of the bare frame structure. Authors recommended the selection of diagonal braces to comply to $T_b/T_u < 0.4$, where T_b and T_u are the fundamental period of the braced frame structure and bare frame structure, respectively. Herein, the fundamental periods T_b and T_u can be expressed as $T_b = 2\pi/\omega_b$ and $T_u = 2\pi/\omega_u$, where ω_b is the natural circular frequency of the fully braced frame before dampers are activated, $\omega_b = (k_b/m)^{0.5}$ and ω_u is the natural circular frequency of bare frame, $\omega_u = (k_u/m)^{0.5}$. In addition, k_b and k_u are the lateral stiffness of the

braced frame structure and bare frame structure, respectively. In these expressions, m is the total seismic mass of the system. Later on, Filiatrault and Cherry (1990) proposed the following equation to calculate the total shear force, V_o that is required to activate all friction damper devices in a structure:

$$V_o/W = (a_g/g) \times Q(T_b/T_g, T_b/T_u, N_f) \quad (9)$$

where N_f is the number of floors, a_g is the design peak ground acceleration, g is the gravity acceleration, T_g is the predominant period of the selected design ground motion, and Q is an unknown single valued function that is given in Eq. 10a for $0 \leq T_g/T_u \leq 1$ and Eq. 10b for $T_g/T_u > 1$.

$$Q = (T_g/T_u) \times [(-1.24N_f - 0.31)T_b/T_u + 1.04N_f + 0.43] \quad (10a)$$

$$Q = (T_b/T_u) \times [(0.01N_f + 0.02)T_g/T_u - 1.25N_f - 0.32] + (T_g/T_u) \times (0.002 - 0.002N_f) + 1.04N_f + 0.42 \quad (10b)$$

Equations 9 and 10a or 10b can be used directly to calculate the total base shear force V_o . It is noted that for a single diagonal brace located at floor i , the slip force is P_i and the activation shear at that floor is $V_i = P_i \cos\theta_i$, where θ_i is the angle of brace inclination reported to a horizontal line.

In the meantime, studies on hysteretic bracing systems were carried out by Ciampi et al. (1990, 1992, 1995, etc.) who considered that both MRF and bracing system are arranged in parallel. They concluded that the activation of friction damper devices should occur before yielding of moment resisting frame. In addition, Ciampi et al. (1995) referred to four key parameters such as:

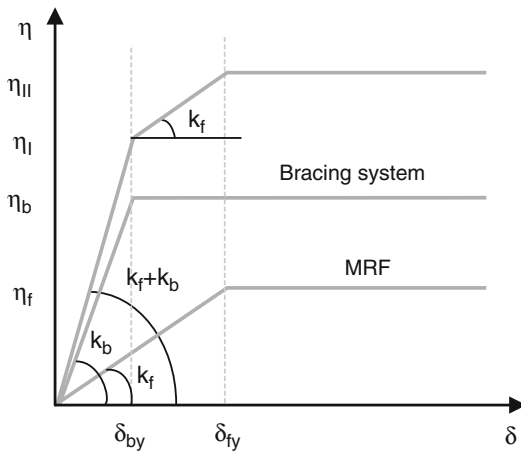
- T_f , which is the fundamental period of the bare moment resisting frame. A deductible parameter is $\alpha = T_b/T_f$ where T_b is the fundamental period of the bracing system.

- $\lambda = k_b/k_f$ which is the ratio between the stiffness of bracing system and the stiffness of the bare moment resisting frame.
- $\beta = \delta_{by}/\delta_{fy}$ which is the ratio between the displacement that cause yielding of bracing system and that that causes yielding of bare MRF system.
- $\eta_f = F_{fy}/ma_g$ which is the yield strength of the frame normalized to the mass of the structure multiplied by the design peak ground acceleration. Similarly, it can be defined $\eta_b = F_{by}/ma_g$, where F_{by} is the yield force of bracing system. The total normalized force is $\eta_{II} = \eta_f + \eta_b$.

The selection of the above parameters suggests the variation effect of β and λ on the building response. When $\beta = 1$, bracing system and MRF system yield for the same lateral displacement. When $\beta = 0$, the structural system corresponds to the MRF system only. They recommended an optimal value of β around 0.5.

The application of this design methodology to MDOF systems requires uniform distribution of stiffness and slip forces along the building height, aiming at uniform activation of friction devices in order to avoid damage concentration within a specific floor. In addition, they proposed to maintain at each storey the stiffness of braces proportional to the bare frame stiffness. Also, they recommended keeping proportionality between the horizontal projection of slip forces per floor and the corresponding static or dynamic distributed storey shear over the building height. In addition, two design assumptions were made: (i) the MRF system is designed to respond elastically while the friction damper devices behave in the non-linear range and (ii) the MRF system may yield when δ_{fy} was reached, therefore after friction damper devices were activated. The tri-linear force-displacement constitutive law is depicted in Fig. 5.

Fu and Cherry (1999, 2000) proposed a simplified design method that leads to a code compatible procedure for a *friction-damped steel braced frame system*. First, they developed the method for a SDOF system based on establishing an equivalent ductility-related force modification

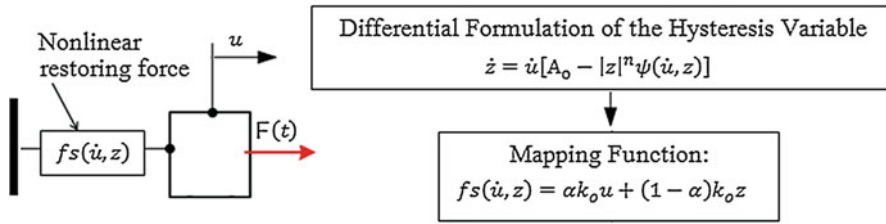


Friction Dampers for Seismic Protections of Steel Buildings Subjected to Earthquakes: Emphasis on Structural Design, Fig. 5 Trilinear force-displacement constitutive law (After Ciampi et al. 1995)

factor R , which reflects the capacity of the system to dissipate energy. They have developed a trilinear model similar with that illustrated in Fig. 5, where η parameter was replaced by the restoring force of a SDOF system, $f(t)$. In addition, η_l and η_{III} parameters were replaced by f_s and f_y , where f_s is the slip restoring force (after dampers were activated) and f_y is the yielding restoring force developed when $f(t)$ exceeds f_y and the frame members yield. Thus, to design the friction-damped steel braced frame system, the design base shear is first calculated based on the code procedure (e.g. equivalent static force procedure). In this case, the value assigned to the R factor (R_d in the current NBCC code) is evaluated based on the selected system parameters. Then, all members are designed based on forces resulted from the distribution of design base shear over the building height, and the brace stiffness and slip forces are evaluated at each floor. In this study, friction-damped braces were installed in a ductile steel MRF system. The response of building designed based on the employed R factor should be verified such that the demanded lateral frame deflection under the wind and earthquake load to be maintained below the code limits. For example, according to the NBCC 2010 requirements, the system must respond elastically under the wind load, while the interstorey drift should

not exceed $h_s/500$. Under earthquake loads, the maximum interstorey drift should be less than $2.5\%h_s$ for a building classed under the normal category. Herein, h_s is the storey's height.

Tirca (2009) proposed a design methodology to upgrade the seismic resistance of an existing steel moment frame building designed before 1970. The method complies with FEMA 356 (2000) provisions. First, the available elastic base shear, V_f , provided by the existing steel moment frame system was evaluated by using modal response spectrum method and a three-dimensional building model. Then, the required base shear carried by friction-damped braces V_{br} was computed as being the difference between the code demand, V , and V_f . Herein, V is the minimum lateral earthquake force computed based on NBCC provisions by setting $R_d = 1.0$. The effect of torsion was included. By considering the nonlinear behavior of friction dampers, the resulted base shear assigned to friction-damped braces, V_{br} was divided by B_{v+l} as per FEMA 356 and then was distributed over the structure height. According to FEMA 356 recommendation, minimum four friction-damped braces displaced in the direction of loading were installed in each floor. The slip force of each device located at storey i was computed by divided the corresponding storey shear force resulted from a dynamic analysis to the number of friction-damped braces and to the $\cos\theta_i$, where θ_i was defined above. All steel braces were designed to behave elastically while sustaining the nonlinear response of attached friction damper. Thus, braces were sized such that $130\% C_r$ to be larger than the slip force, whereas C_r is the brace compressive resistance. To preserve the existing columns of MRF system in elastic range, it was proposed to stagger the installation of friction-damped braces such that to minimize torsion. To prevent the concentration of damage at specific storeys, it was proposed to maintain a constant ratio between the horizontal projection of slip forces and dynamic distributed storey shear over the building height. However, the available slip length should be recommended in the design phase in order to avoid sudden failure. The displacement at yield, Δ_y , depends on the



Friction Dampers for Seismic Protections of Steel Buildings Subjected to Earthquakes: Emphasis on Structural Design, Fig. 6 Mechanical model of a SDOF system

stiffness of friction-damped braces, k_{bd} , as shown in Fig. 2c. As noted above, the stiffness of the frame and the stiffness of friction-damped braces are in parallel ($k_f + k_{bd}$), whereas $\Delta_y = F_s/(k_f + k_{bd})$ and F_s is the slip restoring force. In this study, the seismic response of retrofitted building equipped with friction-damped braces was verified to remain within the interstorey drift code limits and forces developed in structural members to be below the members' resistance. It was concluded that friction-damped braces installed in moment frame buildings are able to control the interstorey drift and floor acceleration.

Numerical Modeling of Friction-Damped Braces

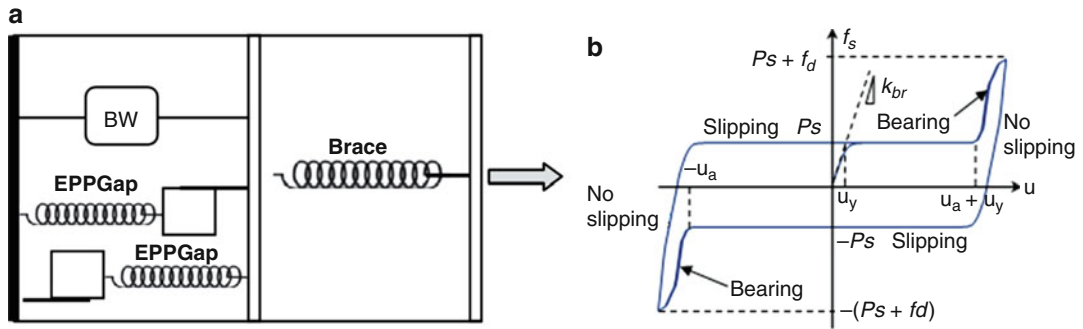
In general, researchers select the bilinear force-deformation model for simulating the behavior of friction damper devices. When performing nonlinear time-history analyses, convergence problems may arise due to the sharp transitions from the elastic to inelastic stages during the loading, unloading, and reloading cycles. Due to the large number of friction devices installed in a building structure, the bilinear model "can become computationally inefficient" when these devices are in different phases of stiffness transition (Moreschi and Singh 2003). To overpass this modeling drawback, the Bouc-Wen model is recommended. In addition, in the case of Bouc-Wen model, it is the same differential equation that governs its response (Eq. 11) and more specifically its behavior during different transition stages. Therefore, the Bouc-Wen model is able to

simulate the highly nonlinear Coulomb friction and has the ability to represent different hysteresis shapes according to the values of the parameters involved (Morales Ramirez 2011). Since the desired shape of the Coulomb dry friction law is symmetric and strength and stiffness degradation is neglected, the Bouc-Wen model is reduced to a nonlinear restoring force (Eq. 11) of a SDOF system shown in Fig. 6. The evolutionary variable z , given in Eq. 11, is defined in Eq. 12.

$$f_s(\dot{u}, z) = \alpha k_o u + (1 - \alpha) k_o z \quad (11)$$

$$\dot{z} = \dot{u} \left\{ \frac{A - |z|^n [\gamma + \beta \text{sgn}(\dot{u}z)] v}{\eta} \right\} \quad (12)$$

In the above equations, α is the participation ratio of the initial stiffness in the nonlinear response, k_o is the initial stiffness of the system, u is the displacement of the SDOF system, and z is the hysteresis variable. In Eq. 12, γ and β are parameters controlling the shape of the hysteresis cycle and the exponent n influences the sharpness of the model in the transition zones. The remaining parameters A , v , and η control the degradation process in stiffness and strength. When the degradation process is neglected, the aforementioned parameters are $A = A_o$, $v = 1$, and $\eta = 1$. The considered SDOF system is characterized by the restoring force $f_s(du/dt, z)$ that has a linear and a nonlinear component, as defined in Eq. 11. Using $n = 1$ might yield to a flexible behavior, while increasing the period of vibration and reducing the inertial forces. The smooth transition toward a rectangular hysteresis shape is obtained for a large value of variable n which might become computationally expensive, because the transient



Friction Dampers for Seismic Protections of Steel Buildings Subjected to Earthquakes: Emphasis on Structural Design, Fig. 7 Friction-damped brace: (a) Schematic model and (b) Hysteresis model plus the bearing state

analysis requires the calculation of the evolutionary variable z at each step of time for a single element. By using $n = 10$, it gives an acceptable level of prediction because the difference is quickly reduced throughout the evolution of the slipping stage.

To simulate the slip and slip-lock phases of friction-damped brace device in OpenSees (McKenna and Fenves 2004), Morales Ramirez (2011) proposed a novel equivalent uniaxial material composed of BoucWen material (BW) in parallel with gap springs made of Elastic-Perfectly Plastic Gap material (EPPGap), which are already available in the OpenSees library (Mazzoni et al. 2006). The friction damper was arranged in series with the brace member made of steel material. The brace was modeled with a truss element. The schematic representation of friction-damped brace device is shown in Fig. 7a. In the illustrated model, BoucWen material was selected to replicate the smooth hysteresis behavior of friction damper (stick-slip and slip phase), which was activated when the axial stress σ_s generated by the slip force P_s was reached ($\sigma_s = P_s/A_{br}$ and A_{br} is the gross area of the attached brace). Each friction damper is characterized by its slip force and slip length that are resulted from calculation. However, under strong ground motions, the provided slip length, u_a , may be limited and the pretensioned edge bolt may drive the friction damper into the slip-lock phase as depicted in

Fig. 7b (Morales Ramirez and Tirca 2012). Herein, u_y is the elastic axial deformation of the brace. To simulate the slip-lock phase, one equivalent EPPGap spring made of three bilinear gap springs arranged in parallel each other are defined to act in tension and another set in compression. Each uniaxial *ElasticPPGap* material (OpenSees notation) has a defined stress-strain or force-deformation relationship either in tension or in compression. Thus, one equivalent EPPGap spring is activated when the displacement demand exceeds the available slip length, $+u_a$, and the other is activated when the $-u_a$ slip length is reached. Once activated, this component is able to limit the displacement and to increase the force experienced by the friction damper. The threshold force of these gap elements is related to the maximum force that the device is able to withstand ($P_s + f_d$) without reaching failure (Fig. 7b). To control the failure phase, the *MinMax* material was set to decouple the friction device when the strain of the *ElasticPPGap* material exceeds the predefined bounds either in tension or compression. In addition, when the *MinMax* material is activated (e.g., $t = t_i$), the device is decoupled from $t = t_i$ until the end of analysis.

Detailed design examples of 4-, 8-, and 12-storey MRF building equipped with Pall friction dampers installed in-line with braces are given in Morales Ramirez (2011) reference. The building is located in Montreal, Canada, and was subjected to ten simulated and historical ground

motions scaled to match the design response spectrum over the period of interest $0.2 T_1$ – $1.5 T_1$, where T_1 is the fundamental period of the building. All analyses were performed using OpenSees.

To evaluate the seismic response of structures equipped with friction-damped braces, the following parameters should be investigated: interstorey drift, lateral deflection, and residual interstorey drift.

For an accurate OpenSees model of friction-damped brace, it is required to calibrate the model against experimental test results. Presently, experimental tests conducted on Pall friction dampers installed in-line with braces are carried out and the obtained results will be disseminated. Furthermore, the development of a code-based design method for structures equipped with friction-damped braces is required.

Summary

Since 1980, seismic response control techniques have been used as complementary solutions to the existing seismic force resisting systems and several types of friction damper devices were developed. Among them, Pall friction damper devices are the most popular. However, despite of their wide-spread applications, design guidelines for structures equipped with supplemental energy dissipation devices are still in an evolving phase. This study refers to the design philosophy of structures with incorporated friction dampers that in most cases are installed in-line with steel braces. A brief review of design guidelines and other seismic design provisions that emphasize on the design of friction-damped braces is presented. In addition, a brief discussion on the numerical modeling of friction-damped braces was conducted.

Cross-References

► [Nonlinear Dynamic Seismic Analysis](#)

References

- Aiken ID, Kelly JM (1990) Earthquake simulator testing and analytical studies of two energy-absorbing systems for multistory structures. Report No UBC/EERC-90/03, Earthquake Engineering Research Center, University of California, Berkeley
- Akiyama H (2000) Evaluation of fractural mode of failure in steel structures following Kobe lessons. *J Construct Steel Res* 55:211–227
- American Society of Civil Engineers (2010) Minimum design loads for buildings and other structures ASCE Reston. SEI 7–10
- American Society of Civil Engineers (2013) Seismic evaluation and retrofit of existing buildings ASCE Reston. SEI 41–13
- Baktash P, Marsh C (1986) Seismic behavior of friction damped braced frames. In: *Proceedings of the 3rd U.S. National Conference on Earthquake Engineering*, Charleston, pp 1099–1105
- Baktash P, Marsh C (1987) Damped moment-resistant braced frames: A comparative study. *Can J Civil Eng* 14:342–346
- CEN (2004) Eurocode 8: Design provisions for earthquake resistance structures. ENV 1998–2. Brussels, Begium
- Christopoulos C, Filiatrault A (2006) Passive supplemental damping and seismic isolation. Instituto Universitario di Studi Superiori di Pavia, IUSS Press, Pavia
- Ciampi V, Ferretti A (1990) Energy dissipation in buildings using special bracing systems. In: *Proceedings of the 9th European conference on earthquake engineering*, vol 3. Moscow, pp 9–18
- Ciampi V, Paolone A, De Angelis M (1992) On the seismic design of dissipative bracing. In: *10th world conference on earthquake engineering*, vol 7. Madrid, pp 4133–4138
- Ciampi V, De Angelis M, Paolone F (1995) Design of yielding or friction-based dissipative bracing for seismic protection of buildings. *Eng Struct* 17:381–391
- Federal Emergency Management Agency (FEMA) (2000) Prestandard and commentary for the seismic rehabilitation of buildings. FEMA-356, Washington, D.C.
- Federal Emergency Management Agency (FEMA) (2003) NEHRP Recommended seismic provisions for new buildings and other structures. FEMA-450, Washington, DC
- Federal Emergency Management Agency (FEMA) (2009) NEHRP recommended seismic provisions for new buildings and other structures. FEMA-P-750, Washington, DC
- Filiatrault A, Cherry S (1988) Seismic design of friction damped braced steel plane frames by energy methods. Earthquake Engineering Research Laboratory report UBC-EERL-88-01. Department of Civil Engineering, University of British Columbia, Vancouver

- Filiatrault A, Cherry S (1990) Seismic design spectra for friction damped structures. *ASCE J Struct Eng* 116(5):1334–1355
- Fitzgerald TF, Anagnos T, Goodson M, Zsutty T (1989) Slotted bolted connections in a seismic design of concentrically braced connections. *Earthquake Spectra*, EERI 5(2):383–91
- Fu Y, Cherry S (1999) Simplified seismic code design procedure for friction damped steel frames. *Canadian J Civil Eng* 26:55–71
- Fu Y, Cherry S (2000) Design of friction damped structures using lateral force procedure. *Earthq Eng Struct Dyn* 29:989–1010
- Grigorian CE, Yang TS, Popov EP (1993) Slotted bolted connection energy dissipators. *Earthq Spectra* 9(3):491–504
- Liao WI, Mualla I, Loh CH (2004) Shaking table test of a friction-damped frame structure in. *Struct Des Tall Spec* 13:45–54
- Mazzoni S, McKenna F, Scott M, Fenves G (2006) *OpenSees command language manual*. PEER, University of California, Berkeley
- McKenna F, Fenves GL (2004) *Open system for earthquake engineering simulation (OpenSees)*, PEER, University of California, Berkeley
- Morales Ramirez JD (2011) Numerical simulations of steel frames equipped with friction-damped diagonal-bracing devices. MASC thesis, Building, Civil and Environmental Engineering, Concordia University, Montreal
- Morales Ramirez JD, Tirca L (2012) Numerical simulation and design of friction-damped steel frame structures. In: *The 15th world conference in earthquake engineering*, Lisbon, paper #2538
- Moreschi LM, Singh MP (2003) Design of yielding metallic and friction dampers for optimal seismic performance. *Earthq Eng Struct Dyn* 32:1291–1311
- Mualla IH (2000) Experimental and computational evaluation of a novel friction damper device. PhD thesis, Department of Structural Engineering and Materials, Technical University of Denmark
- National Research Council of Canada (2010) *National building code of Canada*. National Research Council of Canada, Ottawa
- Nims D, Richter R, Bachman R (1993) The use of the energy dissipating restraint for seismic hazard mitigation. *Earthq Spectra* 9(3):467–489
- Pall A (1979) Limited slip bolted joints, a device to control the seismic response of large panel structures. PhD thesis, Building, Civil and Environmental Engineering, Concordia University, Montreal
- Pall A, Marsh C (1981) Response of friction damped braced frames. *J Struct Div ASCE* 108(ST6):1313–1323
- Pall A, Pall R (2004) Performance based design using Pall friction dampers – An economical design solution. In: *13th world conference on earthquake engineering*, Vancouver, paper #1955
- Tirca L (2009) A simple approach for the seismic retrofit of buildings with staggered friction dampers. In: Mazzolani FM (ed) *Proceedings of the first international conference on: protection of historical buildings*. CRC Press, Italy, pp 761–768
- Tremblay R (1993) Seismic behavior and design of friction concentrically braced frames for steel buildings. PhD thesis, University of British Columbia, Vancouver
- Zhou X, Peng L (2009) A new type of damper with friction. *Earthq Eng Eng Vibrat* 8(4):507–520

Geotechnical Earthquake Engineering: Damage Mechanism Observed

Ikuo Towhata

Department of Civil Engineering, University of Tokyo, Bunkyo-ku, Tokyo, Japan

Synonyms

Damage; Deformation; Earthquake; Fault; Liquefaction; Shaking

Introduction

In the modern community, there are a variety of problems caused by earthquakes, including such traditional ones as collapse of buildings and slope failures as well as lifeline problems and economic damages which became important in the recent times. Thus, new types of problem have been recognized as our culture and civilization change with time. It is also important to know that the safety demand is rising continuously with time.

Earthquake engineering addresses a large number of issues that are related with the past experiences of damage, their causative mechanisms and mitigations. Because of the complexity of the problem, the issues to be discussed range widely from geological and mechanical topics to

societal and legal issues. It is obviously impossible for one author to deal with everything. The present author was requested to write general remarks on geotechnical issues and there are many things to be picked up. Because of the page limitation and expectation that other authors will write more about technical issues, it was decided that this entry should address geotechnical damages so far experienced and their causative mechanisms with some suggestions on future directions. Because the human civilization is changing continuously, it cannot be stressed too much that more new problems will occur at any time and that all the readers of this encyclopedia make efforts to develop new solutions. To achieve this goal, it is important to visit the real earthquake damage sites and draw lessons from these real case studies.

Damage Caused by Ground Shaking

Intensity of Seismic Motion and Subsoil Conditions

Collapse of houses has been and still is the most important kind of seismic damage. Figure 1 illustrates a building crash. While this building was subjected to a complicated effect of ground shaking, it is generally agreed that the horizontal shaking is more responsible for the damage than the vertical component. The reason for this is that any structure has a safety margin under the gravitational action, and the vertical seismic action is

Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 1 Seismic collapse of RC building (2001 Gujarat earthquake of $M_w = 7.7$ in India)



resisted to a reasonable extent by this margin, whereas there is no significant safety margin against the horizontal action produced by earthquake excitations. Consequently, many efforts have been made in the recent times to improve horizontal resistance in structures.

It has been widely recognized that earthquake damage is more substantial upon soft soil deposits than on firm soil or rock outcrop. This belief is supported by an experience during the 1923 Kanto earthquake of $M_w = 7.9$ in Tokyo. Figure 2a illustrates the geomorphology of the present Tokyo at the end of the sixteenth century when the urban development was started. The eastern part of the area used to be a mouth of a big river surrounded by swamps and rice paddies underlain by soft and loose soils. In contrast, the western part of the area is composed of higher terraces covered by volcanic ash materials, associated with occasional small valleys that were produced by erosion and filled with soft soils. Figure 2b illustrates the distribution of earthquake damage percentage of traditional wooden houses. Note that this damage implies the damage by shaking and does not include those by big fires that finally destroyed most parts of the city. In this figure, a good consistency can be found between the distribution of soft soils in the eastern part and in valleys and the location of high damage rate.

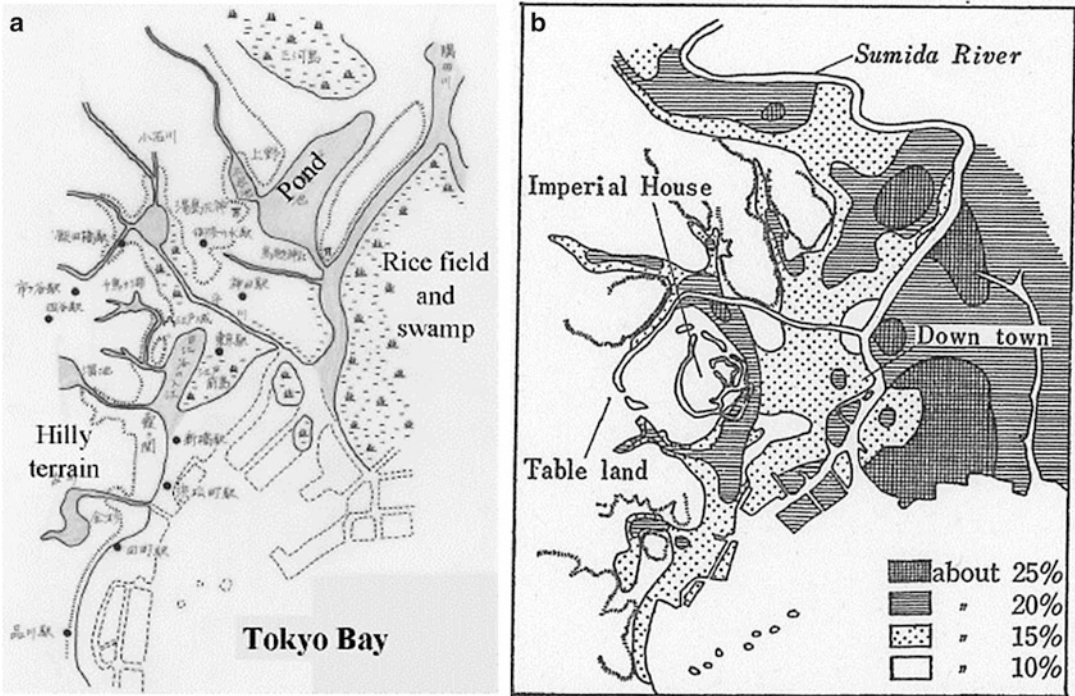
Amplification of Ground Motion in Soft Soil

The intensity of ground motion is increased by the surface soft soil deposits and this phenomenon is called amplification. It is known that the extent of amplification changes basically with the thickness of the surface soil (H) and the velocity of S-wave propagation (V_s). The S(econdary)-wave means a propagation of shear deformation and the theory of elasticity defines its velocity by

$$V_s = \sqrt{\frac{G}{\rho}} \quad (1)$$

in which G and ρ stand for the shear modulus and mass density of soil, respectively. More precisely, the extent of amplification is a function of $2\pi fH/V_s$ where f designates the number of shaking cycles per second (frequency). The most significant amplification, which is called resonance in subsoil, occurs when $2\pi fH/V_s = \pi/2$, and therefore the frequency is equal to $V_s/(4H)$. This implies that the surface ground motion is dominated by a component of this special “resonant” frequency although the motion in the base rock at a deep elevation includes many frequency components in a random manner.

Resonance occurs also in structures upon the ground surface. Generally, soft wooden houses with heavy roofs have low resonant frequency.



Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 2 Correlation between subsoil conditions and earthquake damage of wooden houses in Tokyo during the 1923 Kanto earthquake

(Drawn after Murai 1994). (a) Original geomorphology of Tokyo (Drawn after Murai 1994). (b) Distribution of house damage in 1923 (Okamoto 1973)

Therefore, if the surface ground motion is rich with (dominated by) low-frequency components, the seismic response of such house is significant and heavy damage is likely. In contrast, rigid structures have higher resonant frequencies and are prone to damage if located upon stiff soil where H is small, V_s is high, and $V_s/(4H)$ is close to the structure's resonant frequency. Figure 3a, b illustrates two wooden houses, one crushed house on soft alluvium and survival of the other on a firm gravelly fun deposit. On the contrary, Fig. 3c shows a rigid warehouse that was located on a firm gravelly fun and suffered a significant damage during the earthquake.

Seismic Inertial Force

The earthquake resistant design in the modern times was established by the use of seismic inertial force. Theoretically, this force is related with the maximum acceleration (PGA) by

Seismic inertial force for design
= (Weight of structure) $\times$ PGA/g (2)

where g stands for the gravitational acceleration ($=9.8 \text{ m/s}^2$) and PGA is affected by the aforementioned amplification of motion. Noteworthy is that this inertial force is considered to be a static force that lasts permanently despite that the real earthquake action occurs for a much shorter time. Hence, Eq. 2 overestimates the real seismic effect. To compensate for this, PGA in Eq. 2 has to be made smaller than the real maximum acceleration.

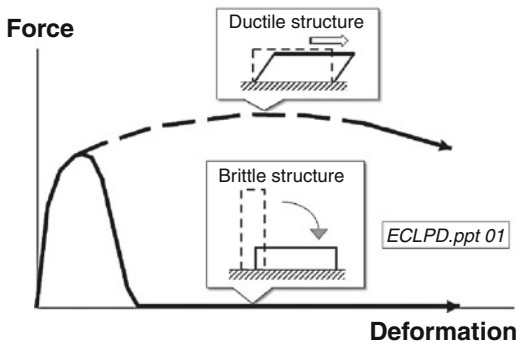
Equation 2 implies that the maximum acceleration or its equivalent inertial force controls the seismic failure of structures. This type of failure is the case of a brittle structure which collapses if the external load exceeds the resistance; see Fig. 4. Because traditional masonry buildings and wooden houses were of brittle nature, seismic design by means of this static seismic design

**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 3 Different damage extents of houses because of different resonant frequency and ground conditions (after the 1995 Kobe earthquake of $M_{jma} = 7.3$). (a) Soft house on soft soil. (b) Soft house on firm soil. (c) Rigid house on firm soil



Ductile and brittle behaviors of structure



Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 4 Schematic illustration of ductile and brittle behaviors

force achieved significant success in the first half of the twentieth century. Noteworthy is that the second half of the century developed ductile structures (Fig. 4) of which seismic deformation plays major roles in their seismic performance-based design. Moreover, underground lifelines are more affected by ground deformation than the inertial force. Therefore, *PGA* is not so important today, and the maximum velocity (*PGV*), response spectrum, duration time of shaking, and the number of loading cycles are often referred to in seismic risk assessment. Note that Mexico City experienced significantly intensified ground motion and elongated duration time of shaking due to its basin structure and soft soil deposits during the 1985 earthquake of $M_w = 8.3$.

Seismic Soil-Structure Interaction

Structures in contact with soil are always subject to the soil-structure interaction. Mononobe-Okabe theory of seismic earth pressure (Mononobe and Matsuo 1929; Okabe 1924) addresses the interaction of a rigid retaining wall and backfill soil which has yielded during shaking and exerts pressure on a wall. Note that the theory hypothesized that the wall has translated substantially to produce the active earth pressure in the backfill. Thus, the Mononobe-Okabe earth pressure is not the pressure during earthquakes but the active pressure after large deformation. Because of the long history of



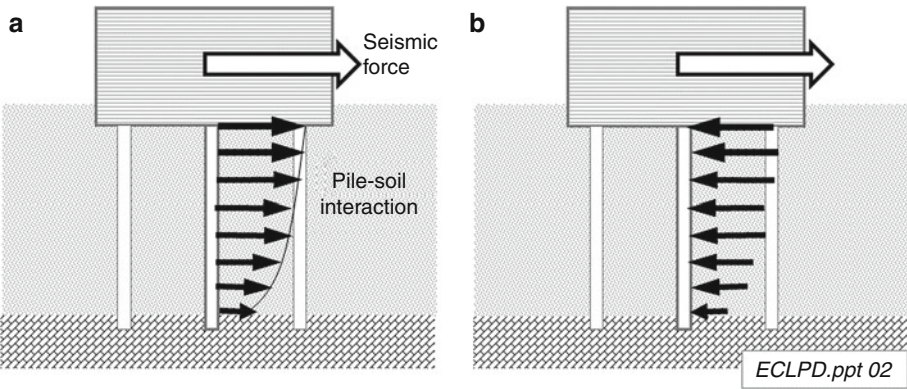
Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 5 Failure of retaining wall due to significant seismic earth pressure from backfill soil (near Hanshin Ishiyagawa Station during the 1995 Kobe earthquake; $M_{jma} = 7.3$)

practice, the Mononobe-Okabe theory has successfully reduced the number of failure of retaining walls. An exceptional case is shown in Fig. 5. More recent experiences suggest that attention should be paid to the foundation of a retaining wall. For example, Fig. 6 shows a case in which a concrete wall resting on unstable fill failed and translated substantially.

Pile foundation is subject to soil-structure interaction during earthquakes as well. As illustrated in Fig. 7, piles are first subjected to the horizontal seismic load that occurs in the superstructure (inertial effect). Second, piles are affected by the differential movement between pile and soil (kinematic pile-soil interaction). In case that soil is very soft, soil translates more than piles and hence piles are loaded in the same direction as the seismic force in the superstructure (Fig. 7a). It seems that such a difficult situation was hardly considered in the past; soil was supposed to be stable and resisted against the pile deformation (Fig. 7b). When the horizontal capacity of pile foundation is not sufficient, one

Geotechnical Earthquake Engineering: Damage Mechanism Observed,

Fig. 6 Translation of retaining wall triggered by failure of unstable fill in its foundation (in Takamachi residential area, Nagaoka, during the 2004 Niigata-Chuetsu earthquake of $M_w = 6.6$)



Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 7 Schematic illustration of pile-soil interaction during earthquakes. (a) Soil displacement > pile displacement. (b) Soil displacement < pile displacement

mitigation is to enlarge the size of foundation and increase the number of piles (Fig. 8).

Underground Structure and Earthquake Motion

The deformability of subsoil as illustrated in Fig. 7a is important because this is one of the main reasons why underground structures are destroyed during strong earthquakes. Figure 9 illustrates depression of ground surface that was caused by the collapse of a metro station in Kobe in 1995. This station collapsed because the surrounding backfill soil was not sufficiently firm

and exerted load on the station. Central columns in the station were destroyed, the ceiling fell down, and the ground surface caved in. Noteworthy is that soil becomes softer during strong shaking because shear strain in soil during strong shaking is greater and shear modulus of soil decreases as shear strain amplitude increases. Such a situation was not fully considered in the past when subsoil was optimistically assumed to reduce the deformation of underground structures. After this lesson was learned, central columns in many subway stations were reinforced to avoid future problems (Fig. 10). However, the

Geotechnical Earthquake Engineering: Damage Mechanism Observed,

Fig. 8 Increasing the number of piles to resist stronger earthquake force (Kobe)



Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 9 Depression at ground surface above the collapsed subway station (Kobe 1995, Photo by Prof. Nozomu Yoshida)

success of reinforcement depends on the intensity of earthquake shaking during future earthquakes.

Similar situation is found in the significant deformation of a tunnel in “soft” rock (Fig. 11).

Note that tunnels used to be considered stable during earthquakes because surrounding “hard” rock was expected to reduce the deformation of tunnel during earthquakes.

Geotechnical Earthquake Engineering: Damage Mechanism Observed,

Fig. 10 Seismic retrofitting of central columns in Oshiage subway station in Tokyo



Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 11 Significant seismic deformation of tunnel in soft rock (Haguro Tunnel, 2004 Niigata-Chuetsu earthquake of $M_w = 6.6$) (this tunnel did

not have an invert at the bottom). (a) Bending and compression in the top of tunnel (*crown*). (b) Uplift at the bottom of tunnel

Slope Instability

Gigantic Failure of Slopes

In addition to collapse of houses, slope failure caused by earthquake shaking is another traditional and important type of seismic disaster. If its size is huge, slope failure may kill thousands of people. For example, the failure of the slope of Mt. Huascarán in Peru triggered by the 1970 Ancash earthquake of $M_w = 7.9$ caused a huge avalanche flow that overtopped a natural protection hill, buried Yungay City (Fig. 12), and

instantaneously killed 20,000 residents (Plafker et al. 1971). After this tragedy, Yungay moved to a safer place.

The 1964 Alaska earthquake of $M_w = 9.2$ caused a submarine landslide in the Port of Valdez. The abrupt and significant change of the submarine topography triggered a huge tsunami that washed the Valdez township and claimed more than 30 lives. Figure 13 shows an aerial view of the site of former Valdez. It seems that the failed submarine slope consisted of fine nonplastic sediments that were transported

**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 12 Former site of
Yungay City (see a hill in
the back side and
Mt. Huascarán with snow)



G

**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed, Fig. 13**

Site of
former Valdez, Alaska
(Photograph taken by
Dr. K. Horikoshi in 2000)



from the glacier through the river in the figure into the sea bottom. The fine grain size of the material made the sedimentation rate very low in the seawater, resulting in loose deposit, while the lack of cohesion made the material highly prone to liquefaction. Even worse was that the fine grain size reduced the permeability and hence the dissipation of excess pore water pressure during shaking was very slow.

Consequently, this submarine slope easily liquefied and failed during the gigantic earthquake and triggered a tsunami.

Natural Dam

The 1999 Chi-Chi earthquake in Taiwan registered $M_w = 7.7$. The rock mass in the geologically young Taiwan island was not fully lithified. The complicated tectonic environment around

Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 14 Tsao-Ling slope failure in Taiwan caused by the 1999 Chi-Chi earthquake



the Taiwan island disturbed the rock mass. The high extent of annual rainfall made steep slopes in the mountainous regions. These situations were combined to develop many slope failures during the earthquake. Among them the biggest slope failure occurred in Tsao-Ling (Fig. 14), and 120 million m^3 of materials flowed down into the river to form a natural dam (Hung 2000). This slope had failed several times in the past because of heavy rains or earthquakes in 1862, 1941, 1942, 1979, and 1999 (Hung 2000). Noteworthy was that the 1999 event produced a big natural dam, similar to the previous event in 1941 (Kawata 1943), and a reservoir which was later filled with river sediments.

The 2008 Wenchuan earthquake of $M_w = 7.9$ in the Sichuan Province of China affected mountainous area where body of mountains has been deteriorated by the tectonic action. Slope failures produced natural dams (Fig. 15). The problem of the natural dams is the possibility of breaching and flooding in the downstream area. Therefore, it is important after big earthquakes in mountain regions to find the formation of natural dams and if a dam is subject to overtopping, erosion, and breaching to drain water as soon as possible. Empirically it is known that seepage and piping failure is rather rare in natural dams (Tabata et al. 2002).

The travel distance of a failed landslide mass (runoff distance) is an important issue because the longer runoff results in damage in a bigger area. Hsü (1975) proposed to study the H/L ratio of the soil flow geometry in which H stands for the height of fall of the soil mass, while L the horizontal travel distance, H/L suggesting $\tan$ (friction angle) where the friction angle is in terms of the total stress. Figure 16 was drawn by collecting data from many literatures. It is important herein, as has been pointed out by many authors, that the greater volume of soil mass is associated with the smaller H/L and, hence, the friction angle. Submarine slope failures are accompanied by even smaller friction angle.

Instability of Slope Surface

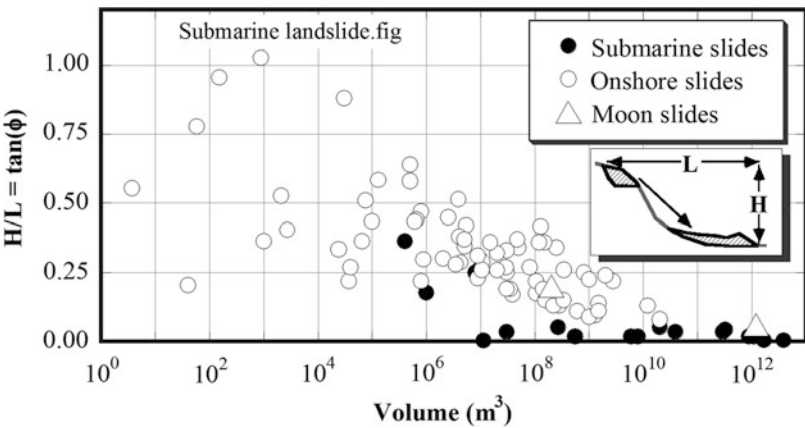
It is often the case that a mountain slope is covered by a weathered unstable material. This weak surface material may fall down during strong shaking. Although the size is small, this type of slope failure occurs at many places and often close mountain roads, making emergency activities very difficult.

Figure 17 reveals a typical seismic failure of a mountain slope. While rainfall-induced slope failures often start from the shoulder of a slope, earthquake-induced failures start from the mountaintop. It should be stressed that slopes of

Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 15 Site of natural dam of Qing zhu Jiang River near Qingchuan, Sichuan Province, China



Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 16 Empirical relationship between height of fall (H) and horizontal travel distance (L) of landslide mass (Reproduced from p. 309 of Towhata (2008))



volcanoes are more vulnerable to failure because, first, generally the volcanic slopes are made of ashes and pumices without rigid rock mass and, second, these materials are pervious and include a big amount of water. The volcanic slope in Fig. 18 started its failure from the top.

Another problem of slope instability has emerged in residential land developments in hilly areas where cut and fill is a common way of construction. Being different from public and industrial projects, individuals cannot always afford reliable but expensive safety measures. As a business practice, therefore, residential

lands have been constructed in a less costly manner. As a consequence, unfortunately, fill parts of land became unstable during strong earthquakes. See Fig. 19. Because residents do not have engineering knowledge, quality assessment of residential lands is recommended as an urgent necessity.

Long-Term Effects of Earthquake Shaking on Slope Instability

It is noteworthy that strong earthquake shaking in mountainous regions may initiate long-term instability of slopes and debris-flow hazards.

**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 17 Seismic failure in the surface of mountain slope (Sichuan Province in China)



**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 18 Failure of slope of volcanic deposits, Costa Rica, 2008, after the Cinchona earthquake of $M_w = 6.1$



The slope in Fig. 20 seismically collapsed in 1707 (120 million m^3) and, thereafter, has produced debris flows during many heavy rainfalls. It seems that the strong shaking disturbed the entire mountain body and the rock mass became subject to disintegration more easily after a huge mass of rock fell down and stress

was relieved. Similarly, the slope behind the city of Muzaffarabad in northern Pakistan became unstable after the 2005 earthquake of $M_w = 7.6$; see Fig. 21.

The second type of long-term effect is caused by a huge deposit of debris in mountain valleys. Figure 22 shows a huge mass of debris in a valley.

**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 19 Seismic instability of residential land in Sendai City after the 2011 gigantic earthquake of $M_w = 9$ in Japan



G

**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed, Fig. 20**

Ohya slide site in Shizuoka, Japan, which became unstable after gigantic collapse caused by the 1707 earthquake of magnitude = 8 or more



This debris deposit was produced by slope failures during the 2008 Wenchuan earthquake, China. Since then, the risk of debris flow and avalanche has increased significantly.

Fault-Induced Problems

Fault movement or fault rupture is an important causative mechanism of earthquake shaking and

**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 21 Seismically
disturbed slope behind
Muzaffarabad, Pakistan



**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed, Fig. 22**

Huge
debris deposit in a valley as
a consequence of slope
failures during an
earthquake in Sichuan
Province of China



is classified into normal, reverse, and strike-slip types. Those types of fault depend on the local or regional tectonic stress conditions. Traditionally fault has been a target of scientific study and has not been investigated for engineering purposes. Recently, however, it attracts more engineering concern because of the increasing demands for safety.

Damage Examples

Figure 23 illustrates the destroyed shape of a concrete gravity dam in Taiwan. The energy and

power of fault action overwhelmed the resistance of a massive concrete structure. As such, fault is generally perceived as a formidable natural disaster. In contrast to this idea, Fig. 24 shows the situation around the fault displacement in Taiwan as well. Noteworthy is that the damage was limited within the area of ground distortion, and the houses out of this distortion were not affected. This implies that the essence of fault-induced damage lies in the substantial distortion of ground, while the intensity of shaking is not so significant as may be imagined. Further, Fig. 25



Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 23 Shikang Dam in Taiwan after the 1999 Chi-Chi earthquake

indicates the ground distortion caused by an underlying strike-slip fault. The indicated distortion in the surface sandy soil was a consequence and was not the fault action. Therefore, the power of the ground distortion was not significant, and the massive concrete foundation of the house was able to resist the stress exerted by the surface soft soil.

Mitigation of Fault-Induced Damage

One of the measures to avoid the effects of fault displacement is a land-use control by which construction on a known fault line is prohibited. Thus, the area in Fig. 26 remains vacant as a park. The range of control is typically within 15 m from the fault line, as practiced in California and New Zealand. Lifelines, however, cannot sometimes avoid a known fault and may be damaged (Fig. 27). It should be recalled that the Trans-Alaska Pipeline successfully survived the effects of strike-slip action of a fault that it crossed during the 2002 Denali earthquake of $M_w = 7.9$. This success was brought about by the extremely flexible structure that absorbed the fault-induced ground distortion (USGS 2003). Thus, there are two kinds of mitigation, rigid

G

Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 24 Limited damage in a fault area (1999 Chi-Chi earthquake)



**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 25 Survival of rigid house resting on strike-slip type of ground distortion (1995 Kobe earthquake)



**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 26 Prohibition of land use along fault (Yokosuka, Japan)



foundation in Fig. 5 and this soft structure. The shield tunnel for the subway of Osaka City is made of more ductile iron segments at the crossing of the Uemachi fault (Azetori et al. 2006). The Shin-Kobe Station is situated on a normal fault, and the station building is divided into three parts without structural connection among them so that a possible fault movement may not destroy the building. The gravity-type Clyde Dam in New Zealand is divided into two parts across the fault line, and a

concrete key block is installed between them in order to prevent possible leakage of water after fault rupture (Hatton et al. 1991).

Coseismic Subsidence and Uplift

Because earthquakes are induced by rupture and distortion in the earth crust, vertical displacement often remains and affects the society. Figure 28 is the result of the subsidence of the earth crust (coseismic subsidence) after the 2011 gigantic earthquake in Japan.



Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 27 Fault-induced buckling of embedded pipeline (1990 Manjil earthquake of $M_w = 7.4$, Iran)

The lowered ground level leads to difficulty in water drainage and protection from high sea waves. The lowered ground level may or may not be recovered within years. On the contrary, the 1804 Kisagata earthquake of the magnitude being around 7.0 caused uplift and drainage of water from a formerly famous beautiful lagoon. Today, former small islands remain as small mounds in the rice field (Fig. 29).

Liquefaction of Sandy Ground

Soil liquefaction is a phenomenon in which cyclic shear deformation of soil causes high pore water pressure and reduces dramatically the shear rigidity of soil. Liquefied subsoil deforms profoundly and the function of the affected structure is lost. Liquefaction is likely to occur in loose, young, cohesionless, and water-saturated soil that is subjected to strong earthquake shaking. The high pore water pressure induces water flow toward the ground surface and this water flow transports soils. As a consequence, sandy deposits of crater shape remain (Fig. 30).

Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 28 Coseismic subsidence in Mangoku-Ura, Japan



**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 29 Coseismic uplift
in Kisagata, Japan (former
lagoon is rice field today)



**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 30 Deposit of sand
ejecta (Urayasu City on
March 13)



Typical Liquefaction Damage

The 1964 Niigata earthquake of $M_w = 7.6$ triggered liquefaction-induced damage at many places. Although many liquefaction had occurred in the human history prior to this event, they did not attract engineering concern because they occurred in rice fields, abandoned river channels, etc. Liquefaction became an important problem in the recent time because human settlement

started to expand into areas of poor soil conditions.

Liquefaction causes large deformation of ground and structures. Figure 31 shows subsidence and tilting of a building. Note that the eccentricity of the gravity force (shape of the building) induced tilting. Because consolidation settlement is quick and small in sandy ground, no pile foundation was installed under many

**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 31 Subsidence and tilting of building with eccentric center of gravity (Niigata, 1964)



**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 32 Falling of Showa Bridge in Niigata, 1964



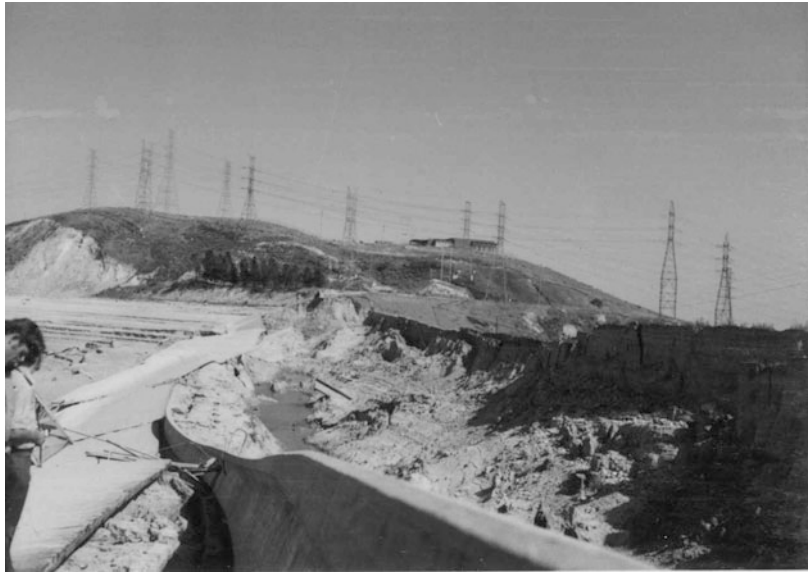
buildings, and hence subsoil liquefaction easily induced this damage. Figure 32 depicts the falling down of Showa Bridge. Although the exact cause of this damage is not yet known, liquefaction in the riverbed somehow affected the stability of the bridge foundation. After 1964, much effort was made to understand the mechanism of liquefaction damage and to develop preventive measures as well as assessment of liquefaction risk. By 1980, it had become possible to protect important

structures from liquefaction problems by densification, installing drains and grouting.

Earthfill dams with insufficient compaction are subject to liquefaction as exemplified by the collapse of the Lower San Fernando Dam in California in 1971 caused by the San Fernando earthquake of $M_w = 6.6$ (Seed et al. 1975); see Fig. 33. Moreover, tailings, which are a waste from mining industries, are prone to liquefaction because it is disposed into reservoir water, its

**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 33 Liquefaction-induced slope failure of Lower San Fernando Dam, California, USA, in 1971



**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 34 Liquefaction of mine tailings in Mochikoshi, Japan, caused by the 1978 Izu-Ohshima-Kinkai earthquake of $M_{JMA} = 7.0$



grain size is small, being of silty size, and it deposits softly in water to form loose water-saturated cohesionless subsoil. Figure 34 shows an example of liquefaction in a tailings dam.

Although end-bearing pile is an effective protection measure of buildings from subsurface liquefaction, it is not perfect. Because water and sand are ejected after liquefaction (Fig. 30), the volume of subsoil contracts, leading to

consolidation settlement in the ground around a pile-supported building (Fig. 35). Accordingly, lifeline connections are damaged.

Horizontal Displacement of Liquefied Subsoil

Those preventive measures that were developed before 1980 were suitable for important structures for which soil improvement was financially feasible. Since the 1980s, liquefaction problem of

**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 35 Elevation difference between pile-supported building and surrounding liquefied ground (Urayasu 2011)



**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 36 Buckling and floating of water pipeline in Dagupan, the Philippines



less expensive structures has attracted concern. For example, a lifeline network cannot improve soil conditions although its possible liquefaction damage affects the entire community significantly. Accordingly, an idea emerged in which the liquefaction-induced damage or deformation should be reduced to an acceptable extent. This idea has developed to the seismic performance-based design in more recent times.

Figure 36 indicates a water pipeline after extensive liquefaction caused by the 1990 Luzon earthquake of surface-wave magnitude = 7.8. The buckling of this pipeline was caused by compressional deformation of the local subsoil. Whether compression or extension, the horizontal deformation and displacement of liquefied ground exert serious effects on underground structures. Hamada et al. (1986) found a

**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 37 Devastated
gravity quay wall in
Kobe, 1995



**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 38 Cracks parallel to
quay wall in Nishinomiya
Harbor after 1995 Kobe
earthquake



liquefied gentle slope to have moved downward and destroyed gas pipelines. Kawamura et al. (1985) reported bending failure of a concrete pile foundation in Niigata where subsoil translated toward the river and the associating lateral displacement of a surface building generated significant bending moment in the pile. Furthermore, in 1995, gravity quay walls moved toward the sea in Kobe and other harbors where backfill soils liquefied (Figs. 37 and 38). The lateral expansion of backfill is evidenced by

many cracks parallel to the quay wall line in Fig. 38.

Liquefaction Problems in the Twenty-First Century

The gigantic earthquake in 2011 in Japan ($M_w = 9.0$) and the sequence of seismic events in New Zealand in 2010 and 2011 caused many liquefaction problems. They are characterized by damage in relatively inexpensive structures such as personal houses, lifelines, and river levees. It is the

**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed, Fig. 39** House
damage in Christchurch
in 2011



G

case therein that the presently available mitigation technologies have reduced the liquefaction problems of important structures for which financial resource is available for soil improvement. In contrast, individuals cannot afford the cost of perfect soil improvement, which may be too conservative for a small house. Moreover, lifelines and river levees do not have sufficient financial resources either for soil improvement over their long length. Thus, liquefaction problem is still important today.

Figure 39 illustrates a house that tilted because of subsoil liquefaction during the 2011 Christchurch earthquake of $M_w = 6.2$ in New Zealand. Although the damage may not appear serious, minor tilting can cause headache and dizziness to residents. The Christchurch officials declared liquefaction-prone area where all the residents were advised to move to safer places. Figure 40 shows liquefaction-induced tilting and subsidence of a low building resting on a shallow foundation without pile. Figure 41 demonstrates liquefaction in a recent artificial land in Itako, Japan, in 2011. Being different from Christchurch, residents in Japan wish to restore their houses in order to continue to live, while improving the subsoil condition. Apparently, soil

improvement under an existing structure is expensive and time-consuming.

River levee is prone to liquefaction because it is often situated upon water-saturated loose sand. Formerly it was supposed that earthquake resistant design is not always necessary in levees and that seismic damage of levees should be restored within a short time prior to the next flooding. Probabilistically earthquake and flooding are unlikely to occur at the same time. Importantly, levees allow a certain extent of seismic deformation and do not require a strict control of safety. Such an idea started to change in the 1990s. Figure 42 illustrates the damaged shape of Yodo River levee in Osaka City after the 1995 Kobe earthquake. See sand ejecta on the ground surface as an evidence of liquefaction. Because this site is close to the sea where high tide occurs twice a day, a significant subsidence of this levee could have led to overtopping of water and “flooding” in the back area that was lower than the average sea level. Since then, the importance of seismic design of river levees has been recognized under certain circumstances.

During the gigantic earthquake in 2011 in Japan, liquefaction damage occurred in many river levees that were located on clayey subsoil

**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 40 Liquefaction in
building foundation in
Christchurch



**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 41 Liquefaction in
residential area of
Itako City



which was unlikely to liquefy (Sasaki et al. 2012). It was interpreted that the sandy body of the levee caused consolidation settlement in the clayey subsoil, got saturated with water, and developed liquefaction (Fig. 43). Figure 44 illustrates an example of this type of liquefaction in which the original ground surface has no liquefaction in contrast to the significant distortion of the levee.

From the engineering viewpoint, this kind of liquefaction revealed such problems as identification of liquefaction-prone sites, improvement of vulnerable soils, and assessment of seismic performance (assessment of seismically induced deformation), all at reasonable costs.

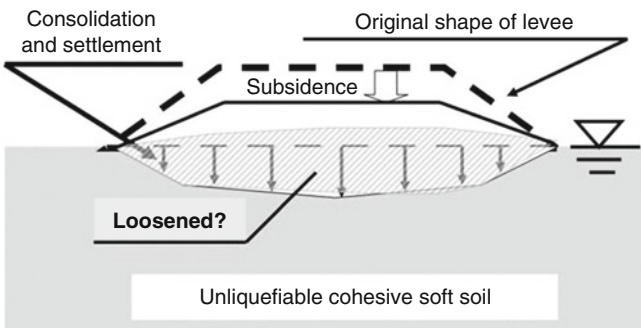
Buried lifelines are subject to liquefaction problems as well. In particular, liquefaction in

Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 42 Liquefaction effects on Yodo river levee at Torishima of Osaka, 1995



G

Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 43 Schematic illustration of mechanism of liquefaction in the body of river levee



Geotechnical Earthquake Engineering: Damage Mechanism Observed, Fig. 44 Example of liquefaction damage in a river levee resting on clayey subsoil (Naruse River, Japan, after the 2011 earthquake)

**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 45 Floating of water
pipe caused by liquefaction
(Itako, Japan, 2011; by
Prof. J. Koseki)



**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed, Fig. 46**

Two-
meter floating of manhole
caused by liquefaction
(Urayasu City, 2012)

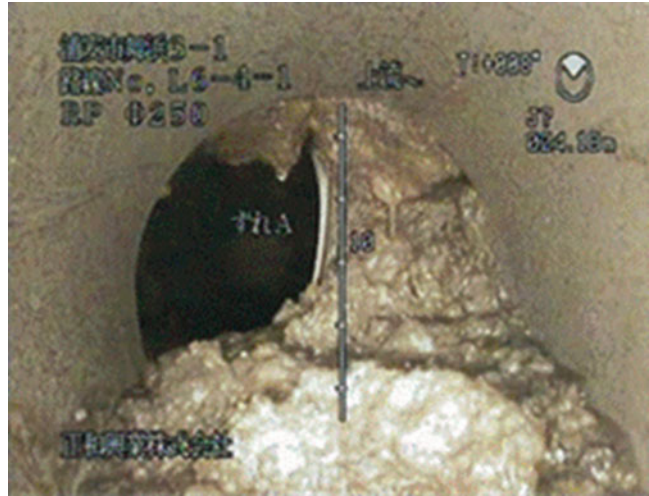


the backfill soil frequently occurs because compaction of the backfill soil around a pipe in a small trench is not easy. Figures 45 and 46 show floating of a pipe and a manhole. Liquefaction

problem of sewerage pipes is more important than that of water supply pipes because the former is embedded at deeper elevation below the groundwater table than the latter and also because

**Geotechnical
Earthquake Engineering:
Damage Mechanism
Observed,**

Fig. 47 Disconnection at sewerage pipe joint and flow-in of liquefied sand (Urayasu City, 2011)



the sewerage water is not pressurized and flows in accordance with the carefully designed slope of pipes; minor change of pipe slope makes the flow difficult. Figure 47 demonstrates one type of liquefaction damage in which a pipe joint was disconnected by large deformation, and the invasion of liquefied sand clogged the water flow.

structures as lifelines, river levees, and residential houses. Because people are demanding a higher level of seismic safety, engineers have to develop methodologies by which their lives and properties are protected from earthquake effects at affordable costs.

Summary

This section reviewed the history of earthquake-induced damage in ground and geotechnical structures, focusing on local soil conditions, intensity of ground shaking, seismic earth pressure, underground structures, slope failures, fault effects, tectonic subsidence, and liquefaction. It is therein seen that the types of damage have been changing with the change of our culture and lifestyles. In particular, liquefaction and lifeline problems were not important 50 years ago and fault problems are now seen as concerns. Another example is the ductile behavior of structures for which the conventional seismic inertial force for design is not good enough. It is therefore important for engineers to pay attention to the changing situation and, when a big earthquake occurs, conduct damage reconnaissance survey in order to detect new types of damage and develop new mitigation measures. Recent earthquakes revealed that earthquake engineering has not yet been helpful for such relatively inexpensive

Cross-References

- ▶ [Analysis and Design Issues of Geotechnical Systems: Rigid Walls](#)
- ▶ [Liquefaction: Performance of Building Foundation Systems](#)
- ▶ [Seismic Actions Due to Near-Fault Ground Motion](#)

References

- Azetori R, Nagataki M, Izumiotani T, Kitada N (2006) Protection of railway shield tunnel from fault displacement. Proc 61st Annu Conv JSCE 3:187–188 (in Japanese)
- Hamada M, Yasuda S, Isoyama R, Emoto K (1986) Generation of permanent ground displacements induced by soil liquefaction. Proc JSCE 376(III-6): 211–220
- Hatton JW, Foster PF, Thomson R (1991) The influence of foundation conditions on the design of Clyde Dam. Transactions of 17th ICOLD, Question 66, pp 157–178
- Hsü KJ (1975) Catastrophic debris streams (Sturzstroms) generated by rockfalls. Geol Soc Am Bull 86:129–140
- Hung J-J (2000) Chi-Chi earthquake induced landslides in Taiwan. Earthq Eng Eng Seismol 2(2):25–33

- Kawamura S, Nishizawa T, Wada K (1985) Seismic damage in piles detected by excavation 20 years after earthquake. *Nikkei Architecture*, Tokyo, July 29th Issue, pp 130–134 (in Japanese)
- Kawata S (1943) Study of new lake created by the earthquake in 1941 in Taiwan. *Bull Earthq Res Inst Univ Tokyo* 21:317–325 (in Japanese)
- Mononobe N, Matsuo H (1929) On the determination of earth pressure during earthquakes. In: *Proceedings world engineering conference*, Tokyo, vol 9. pp 177–185
- Murai M (1994) Edo castle and life of shogun family. *Chuko Shinsho* 45:5 (in Japanese)
- Okabe S (1924) General theory on earth pressure and seismic stability of retaining wall and dam. *Proc JSCE* 10(6):1277–1330 (in Japanese)
- Okamoto S (1973) Introduction of earthquake engineering. University of Tokyo Press, Tokyo, p 62
- Plafker G, Ericksen GE, Fernández Concha J (1971) Geological aspects of the May 31, 1970, Peru earthquake. *Bull Seism Soc Am* 61(3):543–578
- Sasaki Y, Towhata I, Miyamoto K, Shirato M, Narita A, Sasaki T, Sako S (2012) Reconnaissance report on damage in and around river levees caused by the 2011 off the Pacific coast of Tohoku Earthquake. *Soils Found* 52(5):1016–1032
- Seed HB, Lee KL, Idriss IM, Makdisi FI (1975) The slides in the San Fernando Dams during the Earthquake of February 9, 1971. *J Geotech Eng Div ASCE* 101(GT7):51–688
- Tabata S, Mizuyama T, Inoue K (2002) Natural dam and disasters. *Kokin Shoin*, Tokyo, ISBN4-7722-5065-4 C3051, pp 50–53 (in Japanese)
- Towhata I (2008) *Geotechnical earthquake engineering*. Springer-Verlag Berlin Heidelberg
- USGS (2003) Rupture in South-Central Alaska – the Denali Fault Earthquake of 2002, USGS Fact Sheet, 014-03

Global Navigation Satellite System (GNSS) in Earthquake Engineering, Usage of

Masayuki Saeki
Department of Civil Engineering, Tokyo
University of Science, Shinjuku, Japan

Synonyms

Displacement monitoring; GPS (Global Positioning System)

Overview

GNSS (Global Navigation Satellite System) is a satellite-based navigation system. Since it provides accurate position and precise time, the GNSS receiver is employed as a time synchronization device and/or a displacement monitoring sensor in earthquake engineering.

The United States' GPS (Global Positioning System) is one of the operational GNSSs. The constellation of GPS satellites consists of $24 + \alpha$ operational satellites which are deployed in six orbital planes. At least four satellites are arranged in a plane, and several spare satellites are also operated for malfunction. With this constellation, four to eight satellites will be always visible with the elevation mask of 15° . The positioning accuracy reaches to a few meters with point positioning, a few centimeters with kinematic positioning, and a few millimeters with static positioning. Since GPS was at first organized for military purpose, GPS was not perfectly accessible to civil and commercial users. However, SA (Selective Availability) was discontinued in May 2000 to promote its civil use.

In addition to GPS, other GNSS systems are in use or under development. The Russian GLONASS (Global Navigation Satellite System) is also operational GNSS. European Union Galileo positioning system, Chinese Compass navigation system, Indian Regional Navigation Satellite System, and Japanese Quasi-Zenith Satellites are under development as of April 2013.

In the following sections, at first, the concept of GNSS positioning analysis is mentioned. And after that, examples of usage of GNSSs in earthquake engineering are described, such as time synchronization, observation of quasi-static displacements of the Earth's surface, and measurement of the dynamic displacement.

Concept of GNSS Positioning Analysis

Point Positioning and Time Synchronization

A GNSS receiver collects the signals broadcast from the GNSS satellites, which include satellite

navigation information and the precise time when the GNSS satellite emits the signals. As receiving the signals, the receiver estimates the travel time by subtracting the broadcast time from the received time and calculates the range between the receiver and satellite by multiplying the travel time by the propagation velocity. Since an inexpensive crystal clock is employed in the receiver for decreasing its cost, the travel time is contaminated with receiver's clock error. Therefore, the range is called "pseudorange" instead of true range. So, in the point positioning, the receiver's position (x, y, z) and the receiver's clock error dt are treated as unknown parameters.

The receiver calculates the satellite position from the satellite navigation information. Thus, the receiver has the information about the satellites' positions and the ranges from the satellites. So the receiver is able to estimate its position and clock error with the least squares method if more than four GPS satellites are available. It is said that the accuracy of estimated position is about a few to several meters.

Since the velocity of light is about 3.0×10^{-8} m/s, for instance, 3 m error corresponds to 10 ns. So, in this case, the receiver is able to synchronize to the GPS system time with an accuracy of about 10 ns. The GPS system time is derived from the atomic clocks which are implemented in the GPS ground control stations and the GPS satellites. The GPS Navigation Message included in the GPS signals contains parameters that allow the receiver to calculate the correction time from the GPS system time to the UTC (Universal Time Coordinated) with the accuracy of sub-microsecond order (Dana 1997). Therefore, the GPS receiver for time synchronization provides the precise UTC time.

Interferometry Positioning (Static and Kinematic Positioning)

In order to achieve an accuracy of a few or sub-centimeters, interferometry positioning technique is needed. In the interferometry positioning, phase of the carrier wave is analyzed instead of the code pseudorange. The carrier phase (expressed in cycles) multiplied by the

wavelength gives the information associated with the distance between the receiver and satellite. Since the wavelength is about 19.0 cm in case of L1 carrier wave and the phase can be determined with an accuracy of 1–0.1 %, the range can be estimated less than 2 mm accuracy. This phase pseudorange enables us to estimate the receiver position with an accuracy of a few millimeters. However, how many cycles exist in the range is unknown. So there is an ambiguity in phase pseudorange. This ambiguity is called "phase ambiguity" or "integer ambiguity" and treated as unknown parameters should be estimated as an integer value.

The phase pseudorange between receiver i and satellite k is generally described as the following equation:

$$\phi_i^k = \rho_i^k + \lambda N_i^k + c\Delta t_i + c\Delta t^k + \Delta_{trop,i}^k - \Delta_{ion,i}^k + \Delta_{ant,i}^k + \varepsilon_i^k$$

where ϕ_i^k is the phase in meters, ρ_i^k is the true range, λ is wavelength, N_i^k is the phase ambiguity (or called integer ambiguity), Δt_i is clock error of receiver, Δt^k is clock error of satellite, $\Delta_{trop,i}^k$ is tropospheric delay, $\Delta_{ion,i}^k$ is ionospheric delay, and $\Delta_{ant,i}^k$ is error related to the antenna including antenna phase center variation and multipath noise. The true range ρ_i^k is expressed as

$$\rho_i^k = \|\mathbf{x}_i - \mathbf{X}^k\| = \sqrt{(x_i - X^k)^2 + (y_i - Y^k)^2 + (z_i - Z^k)^2}$$

where $\mathbf{x}_i = (x_i, y_i, z_i)$ is the GPS antenna position and $\mathbf{X}^k = (X^k, Y^k, Z^k)$ is the satellite position.

To eliminate the clock errors related to the receiver and satellite, the double-difference operation is used in the relative positioning. In this method, the relative position from a base station is accurately estimated. The DD (double-differenced) phase pseudorange is expressed as the following equation:

$$\phi_{ij}^{kl} = \rho_{ij}^{kl} + \lambda N_{ij}^{kl} + \Delta_{trop,ij}^{kl} - \Delta_{ion,ij}^{kl} + \Delta_{ant,ij}^{kl} + \varepsilon_{ij}^{kl}$$

where

$$*_{ij}^{kl} = *_{i}^k - *_{j}^k - \left(*_{i}^l - *_{j}^l \right)$$

If the baseline is short enough to assume that the measured phases are equally affected by the troposphere and ionosphere at both receivers, $\Delta_{trop, ij}^{kl}$ and $\Delta_{ion, ij}^{kl}$ are also canceled through the double-difference operation. However, such assumption is generally incorrect in case of determining the crustal deformation, because the baseline is generally over a few to tens of kilometers.

Δ_{trop} is estimated with a mathematical model such as Hopfield model, modified Hopfield models, and Saastamoinen model. In these models, the tropospheric delay in zenith direction is estimated first and is multiplied by the mapping function which represents the effect of elevation angle. For more precise application, Δ_{trop} is treated as an unknown parameter. The ionospheric delay Δ_{ion} also can be calculated with a mathematical model, but it is efficiently eliminated using the combination of L1 and L2 frequency measurements (Hofmann-Wellenhof et al. 1994). Therefore, L1/L2 GPS receiver is commonly used for determining the crustal deformation. L1 GPS receiver is used only in the case of short baseline.

In the DD phase pseudorange equation, the receiver position $\mathbf{x}_i$ and DD integer ambiguities N_{ij}^{kl} are unknown. If n satellites are available, $3+(n-1) = n+2$ unknowns should be solved although only $n-1$ equations are there at a single epoch. Therefore, in general, equations obtained at different epochs are solved together with the least squares method or Kalman filter. The solutions are called “float solution,” and the integer ambiguity should be resolved as an integer value. LAMBDA method is the most famous method for determining integer ambiguities (Teunissen et al. 1997).

In the relative positioning, relative position from a base station is estimated. Therefore, if the base station and observation point are close to each other, these points are equally displaced due to earthquake and the relative position cannot be observed. So, to measure the displacement due to earthquake, the base station is selected enough far from the observation points. It is said that the

accuracy is correlated with the baseline length and amounts to 1–0.1 ppm for baselines up to some 100 km and even better for longer baselines in case of static positioning (Hofmann-Wellenhof et al. 1994).

The static positioning is generally employed when coseismic, postseismic, and interseismic displacement is estimated. In this analysis, the unknown parameters (x, y, z) are treated as constant value and several hours measurements are analyzed together in a single batch. Since the effects of various noises are averaged, the accuracy reaches to a few millimeters. On the other hand, if the seismogram or dynamic motion is needed, kinematic positioning is applied. In this case, the unknown parameters (x, y, z) are treated as time-dependent variables and generally estimated with Kalman filter. The accuracy is said to be a few centimeters or worse. For the rapid use such as earthquake early warning, RTK (Real Time Kinematic) positioning is used. In the kinematic positioning, since the analysis is carried out in the postprocessing manner, relatively high accurate satellites’ orbit or additional constraints can be involved in the analysis. However, in the RTK GPS, the kinematic positioning is timely carried out following an earthquake event. As a result, the accuracy is relatively worse than that of the kinematic positioning.

Example of Time Synchronization

As mentioned in the previous section, the GPS receiver for time synchronization is able to output timing precisely synchronized to UTC. This enables us to observe seismogram with synchronized clocks at many stations extremely far from each other.

The National Research Institute for Earth Science and Disaster Prevention (NIED) has developed and maintained the nation-wide high-sensitivity seismograph network called “Hi-net” (High Sensitivity Seismograph Network in Japan). In this system, the seismic data originally sampled at 1 or 2 kHz is decimated to 100 Hz and created a data packet of 1 s length to which a GPS time stamp is added (Okada et al. 2004).

Examples of Static Displacement Determination with GNSS

Displacements of the Earth's surface are monitored with permanent GNSS networks. The determined displacement vectors are used for studying crustal deformation associated with earthquakes, determining the geometry of fault plane, and estimating strain accumulation.

Permanent GNSS Networks

IGS (the International GNSS Service) operates an international network of over 350 continuously operating dual-frequency GPS stations and collects GPS and GLONASS observation data sets. These data sets are analyzed and combined to form the IGS products. The primary IGS products are the GPS satellites' IGS final orbit and clock corrections. These accuracies are within 5 cm and 0.1 ns, respectively (Dow et al. 2009). Any users are freely accessible to the orbit data on line. Geocentric coordinates of IGS Tracking Stations are weekly updated and also provided (IGS HP 2013).

GEONET (GNSS Earth Observation Network System) is the largest continuous GNSS network that is operated by GSI (the Geospatial Information Authority of Japan). The GSI operates GNSS-based control stations that cover Japanese country with over 1,200 stations at an average interval of about 20 km. The GPS and GLONASS observation data are sampled at 1 s interval. The 1 Hz sampling data sets and the data sets resampled at 30 s interval are both available online for actual survey work in Japan as well as studies of earthquakes and volcanic activities (GSI HP 2013).

Coseismic, Postseismic, and Interseismic Displacement

It is considered that there are spatial variations in the degree of coupling on the fault plane (asperity model). This means that some parts on the fault plane are strongly coupled and the other parts slowly slip or rupture at relatively small earthquakes. As the plates are moved, the crustal deformation occurs and strains are accumulated around the strongly coupled zone. The

accumulated strain will be released at a large earthquake. Therefore, strain accumulation monitoring is very important to assess earthquake potential. The strain accumulation and its release are estimated by analyzing the data measured by seismograph, tiltmeter, leveling, and the permanent GPS networks.

Coseismic displacements have been measured for studying geometry of fault surface. The surface displacement of the Earth is associated with the slip on the fault surface through mathematical models. The radiated seismic wave field is dependent on the slip amplitude, rupture velocity, and source time function, whereas the quasi-static displacements depend only on the final slip amplitude. Therefore, the displacements measured with GNSS are complementary constraints on the geometry of fault surface (Segall and Davis 1997).

Hudnut et al. (2002) observe the postseismic displacements associated with the 1999 Hector Mine earthquake using the rapidly deployed continuous GPS network and SCIGN (Southern California Integrated GPS Network) stations and analyze the temporal character and spatial pattern of the postseismic transients. They report that the displacements measured at some sites display statistically significant time variation in their velocity.

Ozawa et al. (2011) investigates the spatial distribution of the coseismic and postseismic slips associated with the 2011 Tohoku earthquake (Mw9.0, March 11) by analyzing the ground displacement detected using the GEONET and reports that the detected coseismic slip area matches the area of the pre-seismic locked zone.

At subduction zones, slow-slip events have been detected by geodetic measurements in interseismic period. Hirose and Obara (2005) report the repeating occurrence of short-term and long-term slow-slip events which are accompanied by deep tremor activity around the Bungo Channel region, southwest Japan. They detect these activities using NIED Hi-net equipped with a tiltmeter and a high-sensitivity seismograph and GPS stations of GEONET. The GPS is only used to detect the long-term slow-slip events because the short-term slow-slip event is too small in magnitude to be detected by GNSS.

Examples of Dynamic Displacement Determination with GNSS

The permanent GNSS network is used to measure dynamic displacements as well as static displacements.

Strong Motion

Strong motion associated with a large earthquake is generally observed with seismograph that is inertial sensor and sensitive to acceleration. Therefore, higher frequency components can be detected accurately with the sensor, but the lower frequency components are not detectable and largely contaminated by random noise. Furthermore, if the seismograph is tilted during a large earthquake, the observed acceleration has offset due to the gravitational force. Therefore, the dynamic displacement estimated with the double integration of acceleration diverges proportional to the square of time. Sensor rotation also distorts the measurements. So it is very difficult to detect long-term displacements with inertial sensor measurements. On the other hand, GNSS sensor directly detects the displacement itself. Besides, some GPS receiver is able to sample the data up to 20–50 Hz. From these reasons, GPS sensor is also used for monitoring dynamic displacements. The accuracy (standard deviation) is said to be a few centimeters in horizontal direction and about two times larger in the vertical direction.

Nikolaidis (2001) et al. report that the dynamic ground displacement caused by the Hector Mine earthquake in southern California (Mw 7.1, October 16, 1999) is determined by analyzing the GPS data collected with SCIGN continuous GPS network. The baseline distances are from 53 to 205 km. In this research, the raw GPS data is analyzed with instantaneous positioning method to obtain displacement time series, while kinematic positioning is generally used. In the instantaneous positioning method, integer ambiguities are resolved independently for each epoch. Hence, the analysis is not affected by the cycle slips of phase measurements.

Dynamic displacements obtained from GPS network are used in the inversion for rupture processes. The dynamic displacements include

coseismic displacement as well as lower frequency components. So it is expected to merge the seismic wave inversion and geodetic observation inversion. Yue and Lay (2011) estimate the rupture process of the 2011 Tohoku earthquake (Mw 9.0, March 11) using 1 Hz sampling GPS data measured with GEONET.

Tsunami

Kato et al. (2000) develop a tsunami observation system using RTK GPS to detect a tsunami before it reaches the coast. This system consists of the Support-buoy, the Sensor-buoy, and the base station. The Sensor-buoy is designed not to react to wind wave motion and is equipped with only GPS antenna. The antenna cable is connected to the GPS receiver mounted on the Support-buoy. The Support-buoy is designed to move with wind waves and is equipped with GPS receivers and antenna, a number of sealed lead batteries, a wind generator, solar panels for power supply, and a pair of radio receiver and transmitter devices. Their experimental results show that the developed system is able to track the sea surface quite well.

Examples of Disaster Mitigation with GNSS

Early Warning System

Allen and Ziv (2011) investigate the possibility of application of RTK GPS to earthquake early warning using the data collected in the El Mayor-Cucapah earthquake in northern Baja California (Mw 7.2, 2010). Earthquake early warning is the rapid detection of a large earthquake and prediction of the expected ground shaking within seconds so that a warning can be broadcast to the nation. In general, earthquake early warning system estimates the magnitude of an earthquake with seismic waves observed around the epicenter. The problem of how accurate the magnitude is estimated with only seismic waves for a large earthquake is considered. The unique information provided by the GNSS-based displacement time series is the coseismic displacement that is hardly obtained from acceleration

time series. This information is expected to provide effective constraints to the magnitude estimation.

A GPS-based tsunami early warning component, developed by the German Research Centre for Geosciences (GFZ) within the German-Indonesian Tsunami Early Warning System (GITEWS) project, was installed in Indonesia after the disastrous Indian Ocean tsunami (December 26, 2004). This component consists of four subcomponents; the GPS real-time reference stations (GPS RTR) for ground motion detection and reference station for buoy GPS receivers; GPS sensor stations at tide gauges for ground motion detection and tide gauge data correction; GPS sensors on buoys for sea level measurements and direct tsunami detection; and external GPS stations for providing external reference frame. At a large earthquake, GPS data is transmitted from GPS stations in near real time, and displacements are determined. The ground displacements are used to select the most probable tsunami scenario from some thousands of pre-calculated scenarios, based on a matching process. The sea level height is determined by analyzing the GPS data from offshore buoys to detect tsunami directly (Falck et al. 2010).

Postseismic Damage Detection

Application of the GPS wireless sensor network (GWSN) to postseismic building-wise damage detection is studied. The details of GWSN are described in the later section.

The GWSN consists of a central server and many sensor nodes. The sensor node has a wireless communication module, microcontroller, small battery, and inexpensive L1 GPS module connected to a small patch antenna. At a large earthquake, the sensor nodes collect raw GPS data according to the command from a central server and share the collected data between the neighboring sensor nodes. After that, the sensor nodes determine the relative position from the neighboring sensor nodes onboard and transmit the estimated relative positions to the central server through wireless communication network. The central server combines the relative positions to organize the sensor mesh and detects the

displacements of each sensor. From the estimated displacements, the useful information about the building collapses and road closures due to the debris is investigated for the early stage of the rescue action after a large earthquake (Oguni et al. 2011).

Summary

GNSS is a very useful system in earthquake engineering because it enables us to determine the displacements due to earthquake at any point wherever the antenna and receiver are properly deployed. The estimated coseismic, postseismic, and interseismic displacements are used for determining the geometry of fault plane and rupture process, estimating the magnitude of earthquake, and estimating the future earthquake potential. This helps our understanding of earthquake. GNSS is also used as a time synchronization system for seismograph network.

GNSS receiver is employed as a displacement sensor in various systems for disaster mitigation. A GPS-based tsunami early warning system is installed in Indonesia. And other tsunami early warning systems are studied. Earthquake early warning system with RTK GPS and postseismic building-wise damage detection system are also studied.

As of April 2013, only the United States' GPS and Russian GLONASS are fully operational GNSSs. In the near future, European Union Galileo positioning system, Chinese Compass navigation system, Indian Regional Navigation Satellite System, and Japanese Quasi-Zenith Satellites will join in the fully operational GNSSs. Therefore, the accuracy and robustness of the system with GNSS are expected to be improved further.

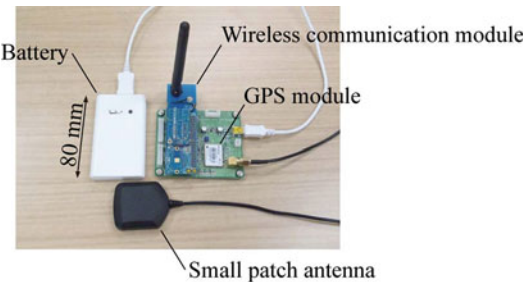
GPS Wireless Sensor Network

System Overview

Wireless sensor network is a system in which many wireless sensor nodes automatically organize and maintain a wireless network and

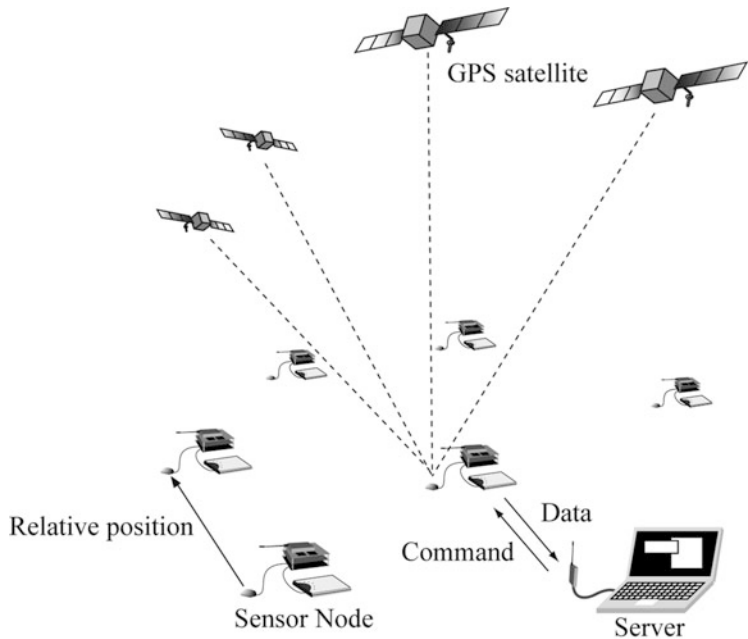
transmit their data or information to each other. The sensor node consists of a microcontroller, physical sensor, low power radio module, and small battery. In the GPS wireless sensor network (GWSN), an inexpensive L1 GPS receiver connected to a small patch antenna is employed as the physical sensor (Saeki et al. 2006). Figure 1 shows a prototype of the sensor node.

This prototype employs a wireless communication module having a 32 bit microcomputer, 128 kbyte RAM, several serial communication ports, as well as an RFIC (Radio Frequency Integrated Circuit) based on IEEE802.15.4.



Global Navigation Satellite System (GNSS) in Earthquake Engineering, Usage of, Fig. 1 Wireless sensor node equipped with an inexpensive L1 GPS module connected to a small patch antenna

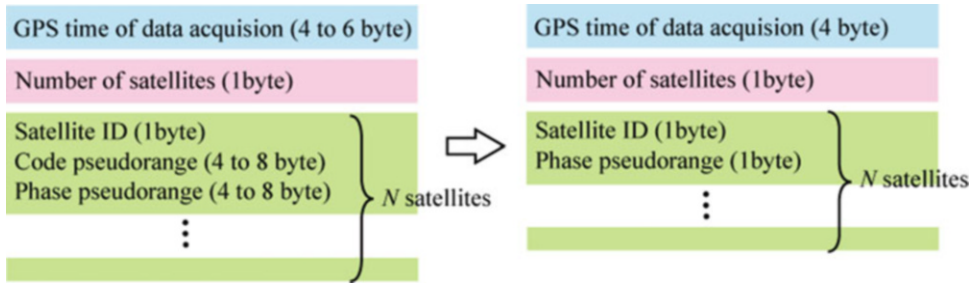
Global Navigation Satellite System (GNSS) in Earthquake Engineering, Usage of, Fig. 2 A schematic view of the GWSN (GPS wireless sensor network)



The electric current is 14.6 mA (3.0 V) at receiving data and 17.4 mA at transmitting data. This kind of wireless communication module has some sleep modes depending on usable functions, and the electric current reaches at a few μ A in the deep sleep mode.

An inexpensive L1 GPS module is connected to the wireless sensor module through a serial communication. The electric current is about 50 mA (3 V) which is highest among the electric components of the prototype. In the prototype, a small patch antenna is employed, which is generally never used for the precise displacement determination because such a small patch antenna has weak directivity and then the data is largely affected by the multipath noises. This noise contamination makes it difficult to resolve the integer ambiguities.

A schematic view of GWSN is shown in Fig. 2. This system consists of many sensor nodes and a central server. The sensor nodes are ordinarily in sleep mode for saving their battery energy. They wake up, for example, every 1 min to ask their task to the server. If the task is “sleep,” they go into sleep mode again. If the task is “observation,” they turn the GPS module on and wait a trigger signal. After receiving the



Global Navigation Satellite System (GNSS) in Earthquake Engineering, Usage of, Fig. 3 A format of compressed data of a single epoch in case of static relative positioning with short baseline

trigger signal, the sensor nodes save the required raw GPS data on their RAM during a specified time interval. After finishing data acquisition, they turn the GPS module off and return back to sleep mode. If the received task is “data collection,” the sensor node sends the saved GPS data to the server through wireless communication. The sampling interval and data length can be changed according to the command from server.

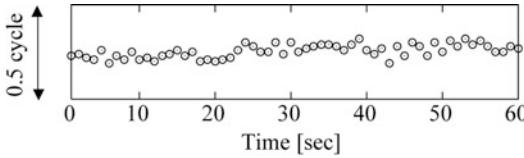
Data Compression Technique

Data compression is a very important technology for wireless sensor network, because a large amount of data requires long communication time and high energy consumption which results in short battery lifetime. For example, if a GPS module outputs about 300 bytes per epoch (practically the data size of original message depends on the type of GPS module) and a sensor node saves the whole data for 10 min, the amount of data reaches to 180 kbyte. It takes about 18 s for a sensor node to transmit the data to the server with a transmit rate of 100 kbps. If a single server covers 100 sensor nodes, it needs at least 30 min (=1,800 s) to collect data from all sensor nodes. (GWSN for the application of building-wise damage detection is specially customized so that the sensor nodes locally share their data between neighboring sensor nodes, estimate their relative position onboard, and transmit only the estimated results to the server to avoid transmitting large amount of data.) Since the original raw GPS data is large in amount, truly required data should be extracted. Figure 3 shows a format of compressed data packet.

In the GWSN, quasi-static condition is assumed, which means a relative position of sensor nodes is invariant during the short observation period (e.g., 10 min). The quasi-static assumption and dense sensor deployment enable us to omit the data of code pseudorange and integer part of phase pseudorange. As a result, the amount of data can be drastically decreased to 28 byte per epoch (Saeki et al. 2006).

The integer part of phase pseudorange is very important information especially for kinematic positioning. As mentioned above, the phase pseudorange includes phase ambiguity as unknown value. The value of phase ambiguity is constant as long as the GPS module continuously tracks the satellite’s carrier waves. If the integer part of phase pseudorange is missed, the phase ambiguity becomes different value and should be determined again. This phenomenon is called “cycle slip” and recognized as an important problem to be solved.

On the other hand, in case of static positioning, the missing integer part can be fixed by shifting the corrected DD phase pseudorange $\hat{\phi}_{ij}^{kl}(t)$ (appears in the following section) by integer value so that the difference between the successive two values is minimized. This is because the corrected DD phase pseudorange can be approximated as a linear function of time with a small slope in case that the relative position of sensor nodes is invariant and the baseline is short enough. Figure 4 shows an example of the corrected DD phase pseudorange. The fluctuation is less than 0.2 cycles after repairing the cycle slips.



Global Navigation Satellite System (GNSS) in Earthquake Engineering, Usage of, Fig. 4 An example of the corrected DD (double-differenced) phase pseudorange in case of short baseline and static condition

Static Positioning Using Linear Approximation Method

Since energy consumption of GPS module is highest among the electric components of the sensor node, it is effective to shorten the acquisition time of GPS for saving battery energy. However, in general, short data length makes it difficult to successfully determine phase ambiguities. The wrong phase ambiguities largely deteriorate the accuracy of positioning. Therefore, an algorithm improving the success rate of phase ambiguity determination with short data length is required to solve the trade-off relationship between the energy consumption and accuracy.

At first, a conventional approach is described here. The observation equation of DD phase pseudorange can be linearized by substituting $\mathbf{x}_i = \hat{\mathbf{x}}_i + \Delta\mathbf{x}$ into the equation of ϕ_{ij}^{kl} .

$$\phi_{ij}^{kl} = \hat{\rho}_{ij}^{kl} + \nabla\hat{\rho}_{ij}^{kl} \cdot \Delta\mathbf{x} + \lambda N_{ij}^{kl} + e_{ij}^{kl}$$

where $\hat{\rho}_{ij}^{kl}$ is the calculated DD range between the approximate position of GPS antenna $\hat{\mathbf{x}}_i$ and the GPS satellite position and e_{ij}^{kl} is the noise including residuals of DD ionospheric and tropospheric delay, antenna noise, and random noise. The equations obtained for available satellites provide the following simultaneous equation:

$$\mathbf{U}(t) = A(t)\Delta\mathbf{x} + \lambda\mathbf{N} + \mathbf{e}(t)$$

where $\mathbf{U}(t)$ is the vector of corrected DD phase pseudorange defined as $\hat{\phi}_{ij}^{kl} = \phi_{ij}^{kl} - \hat{\rho}_{ij}^{kl}$, $A(t)$ is the design matrix, $\mathbf{N}$ is the vector of DD phase ambiguities N_{ij}^{kl} , and $\mathbf{e}(t)$ is the vector of noise $e_{ij}^{kl}(t)$.

The above simultaneous equation is generally solved by using Kalman filter. The solution is

called “float solution” because the phase ambiguities are determined as float values. Accurate position can be estimated after resolving the phase ambiguities as integer values. The resolved integer values are called “fixed solution,” which is searched so that the following objective function J is minimized:

$$J = (\mathbf{N} - \hat{\mathbf{N}})^T Q_{\hat{\mathbf{N}}}^{-1} (\mathbf{N} - \hat{\mathbf{N}})$$

where $\hat{\mathbf{N}}$ is the float solution of phase ambiguity vector and $Q_{\hat{\mathbf{N}}}$ is the variance-covariance matrix of the float solution. The fixed solution is validated by checking the ratio J_2/J_1 where J_1 and J_2 are the minimum and the second minimum residuals, respectively. It is said that the integer ambiguity may be successfully resolved if the ratio J_2/J_1 is greater than 3–5.

As mentioned in the previous section, the corrected DD phase pseudorange $\hat{\phi}_{ij}^{kl}(t)$ can be modeled as a linear function of time. So, in the positioning analysis of GWSN, a linear function is approximately estimated by applying least squares method to the time series of corrected DD phase pseudorange. And the values corresponding to the first and last epochs are calculated using the estimated linear function. The observation equations only at the first and last epochs are solved together with least squares method to estimate the position and DD phase ambiguities. This method may slightly improve the quality of coefficient matrix of normal equation and suppress the effect of noises on the solution.

The success rate of determining the DD phase ambiguities can be improved by using the linear approximation method, especially in case of short data length. Table 1 shows the comparison between the results of the conventional approach (static positioning) and the described method. The observation data used in the analysis was collected under ideal condition in which no obstacles were over there. In this experiment, GPS antennas for survey work were fixed on tripods, and GPS data was logged over 24 h with a 1 Hz sampling rate. In the analysis, data segment with a specified data length was selected

Global Navigation Satellite System (GNSS) in Earthquake Engineering, Usage of, Table 1 Comparison between success rates of the static positioning and the linear approximation method (Saeki et al. 2006)

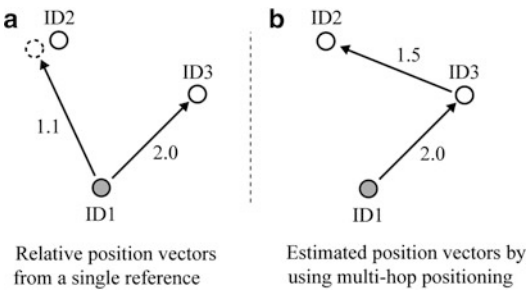
Data length (s)	Success rate (%)	
	Static positioning	Linear approximation
30	71.1	87.3
60	89.8	96.0
180	98.7	99.1
300	99.5	99.6
600	100.0	100.0

and analyzed using both the conventional and the linear approximation approaches. This estimation was performed 86,400 times by shifting the selected data by 1 s.

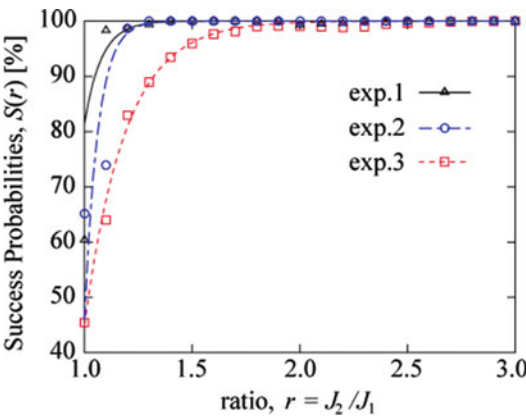
Multi-hop Positioning

The observation equation of DD phase pseudorange is solved with the assumption of white noise. This assumption may be valid for some pairs of sensor nodes existing under similar surrounding condition because such condition gives similar multipath noises and the noises are canceled out in the double-difference operation. However, for the pairs existing under different condition, multipath noises are different and not canceled out in double-difference operation. In this case, the accuracy of float solution becomes worse, and the DD phase ambiguities are not successfully resolved as integer value. Therefore, it is meaningful to select a good pair of sensor nodes whose noises are similar to each other.

Figure 5 shows a simple example of relative position determination. Suppose that three sensor nodes are there and the noises are different from each other but the noise of ID3 has some similarities to those of ID1 and ID2. This situation often happens in actual observations when many sensor nodes are deployed densely. In this case, the relative position from ID1 to ID2 tends to be estimated incorrectly, and the ratio J_2/J_1 becomes small (as described in Fig. 5a) because the noises are not canceled out in the double-difference operation. On the other hand, the vector ID1 → ID3, and also the vector ID3 → ID2, may be estimated better and the ratio becomes larger



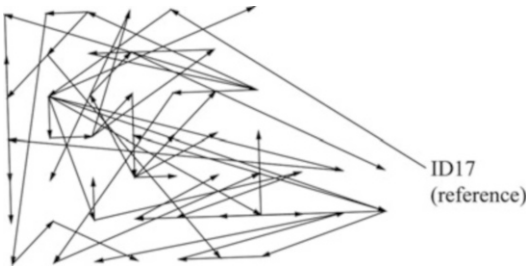
Global Navigation Satellite System (GNSS) in Earthquake Engineering, Usage of, Fig. 5 Simple example of (a) relative position vectors from a single reference point and (b) estimated position vectors by using the multi-hop positioning



Global Navigation Satellite System (GNSS) in Earthquake Engineering, Usage of, Fig. 6 Function of success probability with respect to the value of ratio test (Saeki 2012)

(Fig. 5b). In this case, the sum of vectors ID1 → ID3 and ID3 → ID2 is more likely to give better solution. This is a basic idea of multi-hop positioning (Saeki and Oguni 2012).

To select an optimum path quantitatively among the candidates (possible sum of vectors), success probability function $P(r)$ is introduced. It is empirically known that higher ratio J_2/J_1 ($= r$) gives higher success probability. Figure 6 shows the experimental results of the success probability versus the ratio J_2/J_1 . The marks represent the results obtained from an experiment, and the lines are the fitted curves using the function $P(r) = 1.0 - be^{-ar}$ where a and b are the fitting parameters. The experiment was conducted



Global Navigation Satellite System (GNSS) in Earthquake Engineering, Usage of, Fig. 7 Relative position vectors determined by Dijkstra's algorithm

under an ideal condition (no obstacles over the site), but the GPS antennas were differently fixed. In the first pair (exp.1) the antennas were both fixed on a flat concrete floor. In the second pair (exp.2) one antenna was fixed on a flat concrete floor, but another one was attached on a concrete block with a height of 15 cm. This difference gave slightly different multipath noises. In the third pair (exp.3) one antenna was fixed on a flat concrete floor, but another one was fixed on a tripod. In this case, the multipath noises were very different to each other.

The success probability of path can be evaluated by multiplying the success probabilities of the corresponding vectors.

$$P(r) = P(r_1)P(r_2) \cdots P(r_n)$$

The optimal path is searched from numerous candidates so that the success probability is maximized. In the multi-hop positioning, Dijkstra's algorithm is used as a search algorithm. This algorithm is widely used in many applications such as network routing protocols or mobile navigation systems to find the shortest (or lowest cost) path efficiently.

Figure 7 shows an experimental result of the optimum path estimated using the multi-hop positioning method. In the experiment, 53 sensor nodes were arranged on a rooftop of a building in a grid form at 2 m spaces. The raw GPS data was saved for 4 min with 1 Hz sampling rate and transmitted to the central server through wireless communication. The multi-hop positioning was

applied in postprocessing manner to estimate the best relative positions of sensor nodes which maximized the success probabilities.

Cross-References

- [Early Earthquake Warning \(EEW\) System: Overview](#)
- [Earthquake Mechanisms and Tectonics](#)
- [Remote Sensing in Seismology: An Overview](#)

References

- Allen RM, Ziv A (2011) Application of real-time GPS to earthquake early warning. *Geophys Res Lett* 38, L16310
- Dana PH (1997) Global positioning system (GPS) time, dissemination for real-time applications. *Real Time Syst* 12:9–40
- Dow JM, Neilan RE, Rizos C (2009) The international GNSS service in a changing landscape of global navigation satellite systems. *J Geod* 83:191–198
- Falck C, Ramatschi M, Subarya C, Bartsch M, Merx A, Hoeberechts J, Schmidt G (2010) Near real-time GPS applications for tsunami early warning systems. *Nat Hazard Earth Syst Sci* 10:181–189
- GSI HP (2013) <http://www.gsi.go.jp/ENGLISH/>
- Hirose H, Obara K (2005) Repeating short- and long-term slow slip events with deep tremor activity around the Bungo channel region, southwest Japan. *Earth Planets Space* 57:961–972
- Hofmann-Wellenhof B, Lichtengger H, Collins J (1994) *GPS theory and practice*, 3rd edn. Springer, New York
- Hudnut KW, King NE, Galetzka JE, Stark KF, Behr JA, Aspitoe A, van Wyk S, Moffitt R, Dockter S, Wyatt F (2002) Continuous GPS observations of post-seismic deformation following the 16 October 1999 Hector Mine, California, Earthquake (Mw 7.1). *Bull Seismol Soc Am* 92(4):1403–1422
- IGS HP (2013) <http://figs.csb.jpl.nasa.gov/>
- Kato T, Terada Y, Kinoshita M, Kakimoto H, Isshiki H, Matsuishi M, Yokoyama A, Tanno T (2000) Real-time observation of tsunami by RTK-GPS. *Earth Planets Space* 52:841–845
- Nikokadis RM, Bock Y, de Jonge PJ, Shearer P, Agnew DC, Domselaar MV (2001) Seismic wave observations with the Global Positioning System. *J Geophys Res* 106(B10):21897–21916
- Oguni K, Miyazaki T, Saeki M, Yurimoto N (2011) Wireless sensor network for post-seismic building-wise damage detection. In: *Proceedings of the 20th IEEE international workshops on enabling technologies: infrastructure for collaborative enterprises*. Paris, pp 238–243

- Okada Y, Kasahara K, Hori S, Obara K, Sekiguchi S, Fujiwara H, Yamamoto A (2004) Recent progress of seismic observation networks in Japan - Hi-net, F-net, K-NET and kiK-net -, *Earth Planets Space* 56:15–28
- Ozawa S, Nishimura T, Suito H, Kobayashi T, Tobita M, Imakiire T (2011) Coseismic and postseismic slip of the 2011 magnitude-9 Tohoku-Oki earthquake. *Nature* 475:373–376. doi:10.1038/nature10227
- Saeki M, Oguni K (2012) Multi-hop positioning, relative positioning method for GPS wireless sensor network. In: PECCS2012 – international conference on pervasive and embedded computing and communication systems, Rome, Italy 24–26
- Saeki M, Kosaka T, Kaneko S (2006) Phase ambiguity resolution based on linear modeling of DD carrier phase. In: ION GNSS 19th international technical meeting of the satellite division, Fort Worth, 26–29 Sept 2006
- Segall P, Davis JL (1997) GPS applications for geodynamics and earthquake studies. *Annu Rev Earth Planet Sci* 25:301–336
- Teunissen PJG, de Jonge PJ, Tiberius CCJM (1997) The least-squares ambiguity decorrelation adjustment: its performance on short GPS baselines and short observation spans. *J Geodesy* 71:589–602
- Yue H, Lay T (2011) Inversion of high-rate (1 sps) GPS data for rupture process of the 11 March 2011 Tohoku earthquake (Mw 9.1). *Geophys Res Lett* 38:L00G09. doi:10.1029/2011GL048700

Global View of Seismic Code and Building Practice Factors

James Edward Daniell
Geophysical Institute, Karlsruhe Institute of
Technology, Karlsruhe, Germany

Synonyms

Building design codes; Construction practice;
Earthquake construction; Seismic-resistant codes;
Worldwide earthquake standards comparison

Introduction

This entry looks at a review of seismic-resistant codes for each country in the world by examining

The online version includes code changes and code types through time.

the historical changes of codes from 1900 to 2013, the years in which updates were made, the number of buildings these codes influence, and a review of the code quality itself with respect to the hazard of that country. Over 160 countries and nations have some form of seismic code. However, the quality, extent of application, and methodologies between seismic-resistant codes differ around the world. An exploration of the location of each of these code changes, in terms of when they have been implemented, to what extent they have been implemented, and which percentage of buildings they encompass, has been undertaken.

The History of Seismic-Resistant Codes Worldwide and Their Components

Seismic-resistant building codes traditionally started being implemented during reconstruction following major earthquakes of the past, including 1755 Lisbon, 1880 Luzon, and 1908 Messina. Although no country had a formal seismic-resistant code in 1900, some countries such as Italy, Portugal, the USA, and other locations had influences from historical earthquakes, which aided the building styles of the time being more seismic resistant. This was also the case with colonial building styles in Africa. Using the components of lateral earthquake forces, ductility and drift and building height, these were slowly implemented into design, as explained in the previous chapter of Gülkan and Reitherman.

The first seismic-resistant codes simply applied a seismic coefficient and building height limits, such as those implemented within Italy (1910, 1917), Costa Rica (1914), Turkey and Japan (1923), the USA (1906, 1927), and other locations pre-1930s. It should be noted that, pre-1930, some municipalities and cities applied different seismic-resistant building codes or practices, such as building from wood after the 1848 Wellington earthquake (Beattie and Thurston 2006) or after the 1770 Haiti earthquake.

Lateral loads were often included with the wind, snow, and gravity loads in analysis.

Post-1930, the difference of the response spectrum to various structural periods was looked at, with the influence being that the lateral forces would be changed depending on the fundamental period of the building. Elastic analysis was, however, still undertaken, as opposed to elastoplastic analysis (ductile analysis).

Only in the 1960s were ductility considerations taken into account in the form of force-reduction factors, given that ductile structures were able to survive ground shaking and thus inertia forces many times greater than elastic structures. These force-reduction factors were placed slowly into the seismic-resistant codes in different countries from the 1970s to 1990s; however, greater research in the field showed the need for analysis into the effects of damage potential as a result of displacements rather than strength alone.

There were also two types of building analysis undertaken: force-based design and displacement-based design. Force-based design uses a seismic coefficient (C) and a seismic weight (W) calculated to give the design seismic forces (V):

$$V = C_s * W, \quad \text{where}$$

$$C_s = S_a / (\text{Reduction Factor} / \text{Importance Factor})$$

The seismic coefficient is made up of the design spectra of the earthquake horizontal and vertical components calculated from various parameters detailed in Table 1 below, in combination with a reduction factor corresponding to the global ductility capacity and the inherent overstrength in the lateral force-resisting system. Using this system generally works; however, it relies on the fact that strength is being used to control damage and that strength and stiffness are independent. In reality, these two are interlinked. Force-reduction factors assume that the ductility demand is the same for each type of structure unless changed; however, each building generally acts differently with respect to ductility. The distribution of strength in a building, including

the amount of displacement at different limit states, has been determined to be increasingly important, despite both methods looking at the location of plastic hinges (shear strength of members being greater than that of the shear due to flexural strength) and hence looking at the displacement ductility of the components of the building or structure. The previous chapter of Gülkan and Reitherman provides a useful background to the implementation of such parameters.

These difficulties resulted in displacement-based design, rather than force-based design, being undertaken in order to construct buildings safely (Priestley et al. 2007). Displacement-based design is still generally not used in seismic codes but can be applied as a valid alternative where wanted. A comment to this effect has been incorporated in the NZS code 1170 Part 5 (Standards New Zealand 2004).

Thus, there are now two main types of seismic codes used worldwide. The first, strength-based design, relies on a design base shear with force-based design generally based on historical earthquake hazard maps, without any view as to the performance of the structure in different events. The other, performance-based design, still uses these uniform spectra metrics from seismic zoning and traditional force-based design but adds in performance criteria that need to be satisfied at more than one earthquake level. In this way, performance levels are also defined differently for schools versus normal buildings, as a different return period can be defined for the performance, thus designating different displacements or interstory drifts required. Better levels of performance-based design are creeping slowly into revisions of seismic-resistant codes worldwide, with life safety and damage limitation of buildings at various corresponding levels of earthquake motions being implemented and even displacement-based design (Priestley 1993) being included.

Table 1 represents the key elements of a modern seismic-resistant code, which then will be summarized for the purposes of the seismic code index developed.

Global View of Seismic Code and Building Practice Factors, Table 1 The criteria examined for each of the seismic-resistant codes within the worldwide seismic code index (1900–2012) discussed in this study

Element	Sub-element	Key components	Value given out of 104
Structural design method	Material loads/strengths, dead and live load	Basic principles of design as denoted by the general building code in the country	30
Seismic actions	Horizontal components	Design response spectra (S_a) generally determined by the ground seismic acceleration, importance factor, at characteristic design periods, damping modification factor. It has different shapes for design periods, depending on these cutoffs	5
Seismic actions	Vertical components	Generally taken as a percentage of the horizontal factor to create A_v	3
Seismic actions	Ground seismic acceleration	Acceleration at which the seismic zone map coefficient will occur for a return period (475 years generally) Can be intensity or PGA based	Included as a percentage of actuality over 30
Seismic actions	Near fault	Looks at whether near-fault effects are taken into account	2
Seismic actions	Soil classification	Depending on soil type, extra factors applied to account for the increased shaking	5
Seismic actions	Importance factor	Structures generally split into various factors in order to define different levels of design in terms of life safety, e.g., hospitals given a higher level of safety	5
Seismic actions	Behavior factor (ductility factor)	Reduction of seismic loads based on post-elastic behavior – depending on the material of the structural system	5
Seismic actions	Foundation factor	Interaction between the foundation and the soil is included in some cases	5
	Existing buildings		% of buildings
	1–2-story building factor		% of buildings
Design method	Simplified spectrum method	Uses the fundamental mode of oscillation and looks at equivalent seismic forces. Simplified and does not take into account so many complexities, just using base shear multiplied by the design spectra value and distributing the lateral forces over the height of the building Also includes eccentricities, other loading details	5
Design method	Drift	Drift is used to calculate seismic forces as a boundary condition	2
Design method	Dynamic response method	Uses dynamic response spectrum method, taking into account multiple modes of oscillation, which has higher quality until the sum of effective modal masses reaches a certain percentage of the oscillating mass of the system	5
Design criteria	Nonstructural	Are nonstructural elements designed for?	2
Design criteria	Avoidance of collapse	Special notes with application to the earthquake load combination, avoidance of soft-story mechanisms, plastic hinge locations, and other combinations	5
Design criteria	Damage limitation	Load-bearing systems, pounding taken into account, drift, appendages such as chimney, facades	5

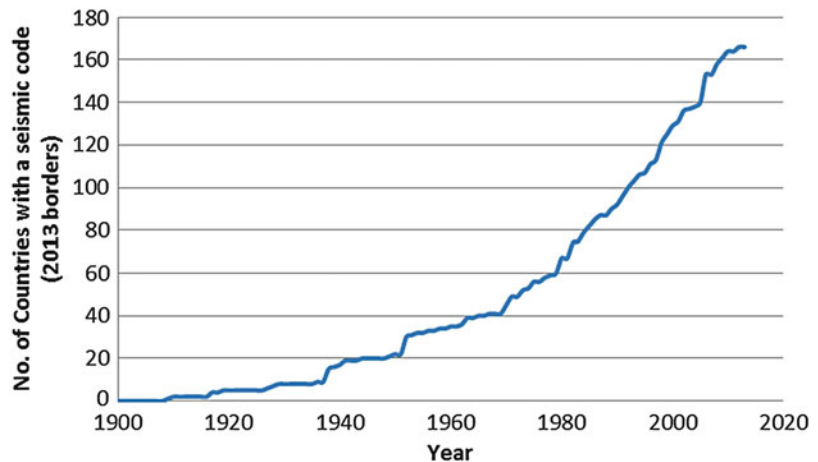
(continued)

Global View of Seismic Code and Building Practice Factors, Table 1 (continued)

Element	Sub-element	Key components	Value given out of 104
Foundation design	Foundations, retaining structures, slopes	Basic checks as to slope stability and other foundation design applications such as liquefaction hazard, shear settlement, differences for shallow and deep foundations, etc.	5
Code quality assessment	Displacement-based design/overall	Displacement-based criteria in terms of the design methodology or overengineered status, additional components and methodologies encompassing many design levels for all types of structures	15

Global View of Seismic Code and Building Practice Factors,

Fig. 1 The number of countries with a seismic-resistant code/zonation as of 2013 borders



Which Countries Have Seismic Codes?

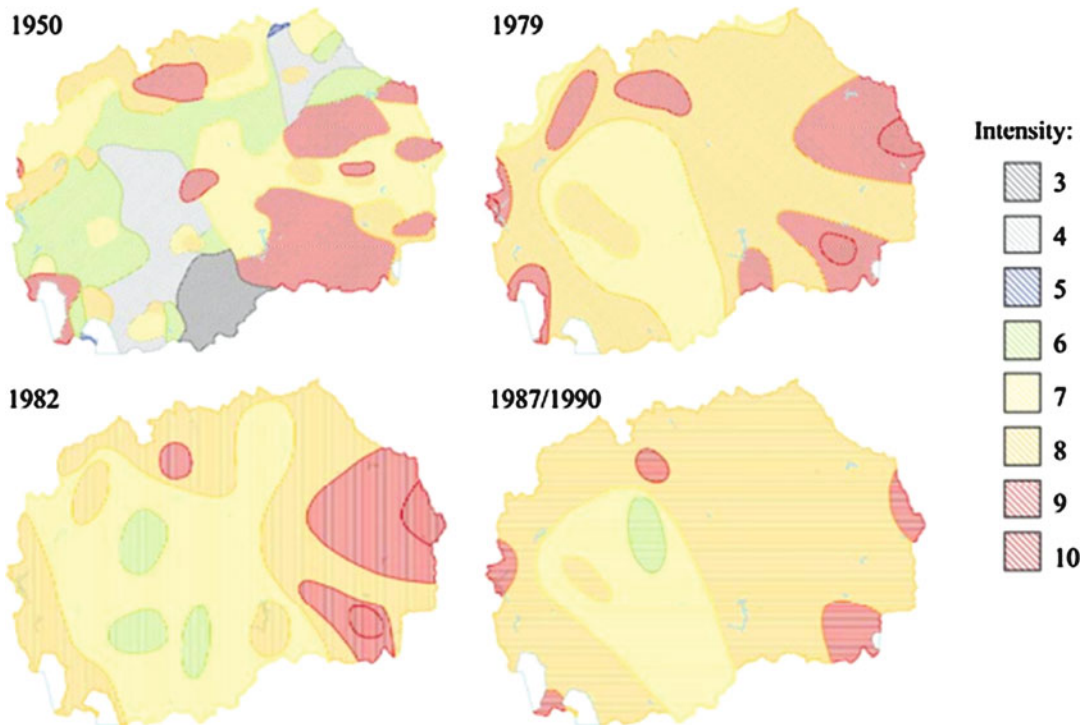
Not all countries have seismic codes; however, documentation does not exist on a global scale to see which countries have seismic-resistant codes and whether they are part of law, enforced or ignored. In addition, some countries have differences on the province-level enforcement or adherence to such codes.

In order to determine the factor by which damage is changing and the complexity of codes, a significant review and analysis was undertaken to create a harmonized up-to-date list of locations where seismic-resistant codes have been considered. There were several early initiatives which attempted to undertake this by looking at only a list of some of the countries worldwide with seismic-resistant building codes, including the World List initiative by the IAEE which includes details as to some seismic-resistant codes in various countries – IISSE

1992, 1996, 2000, 2004, 2008, and 2012 (IAEE 1996, 2000, 2013). In addition, there has been a list of countries which have some form of seismic codes in the practice of earthquake hazard assessment (92 countries) (McGuire 1993). The “Seismic Code Evaluation” work in Central America as well as South America has been collected and implemented in the database (Chin and Association of Caribbean States 2003).

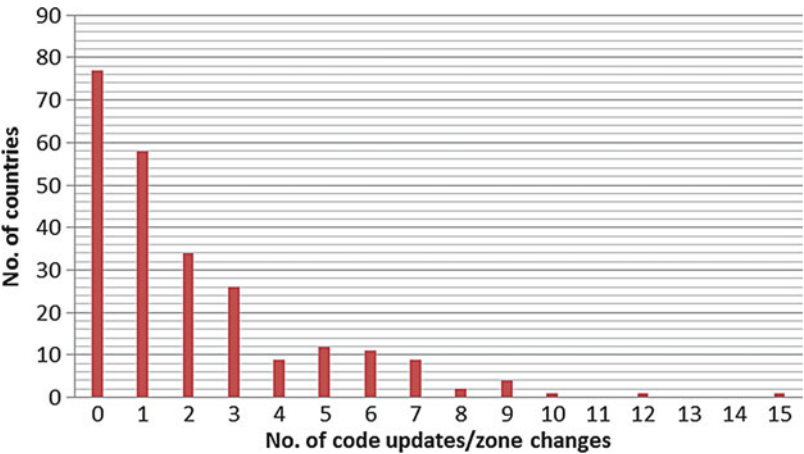
One hundred and sixty-six nations out of two hundred and forty-four nations have some form of seismic-resistant code, but they are usually not implemented for all styles of buildings, given the large number of nonengineered building styles worldwide as shown in Fig. 1.

In addition, there are nations which have some form of code; however, this has not been ratified and has not been implemented in practice. For earthquakes, a review of the various seismic-resistant codes around the world has been



Global View of Seismic Code and Building Practice Factors, Fig. 2 The seismic zonation changes for the country of Macedonia (Salic et al. 2010)

Global View of Seismic Code and Building Practice Factors, Fig. 3 The number of countries with code changes and updates since 1900

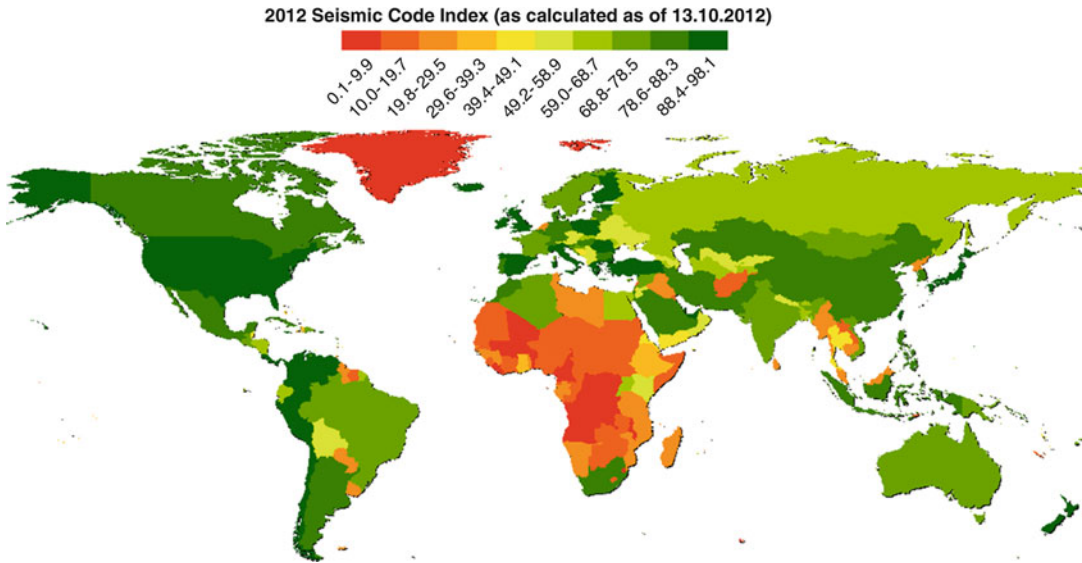


undertaken, examining the level of base shear relative to hazard, seismic zoning, and other parameters ranked in comparison to the relative hazard of a particular country versus the seismic hazard actually represented.

Many different countries have had multiple editions, adjustments, and changes to their seismic-resistant codes and zonations in the past

century. Studies such as Salic et al. (2010), as depicted in Fig. 2 for Macedonia, or Romeo (2007) for Italy show the move from seismic hazard based on historic earthquakes to probabilistic hazard with a smoothing of hazard zonations used in code through time.

Shown in Fig. 3 are the number of code changes through time for each nation.



Global View of Seismic Code and Building Practice Factors, Fig. 4 The worldwide seismic code index (level 1) as calculated for 2012, showing the quality of the

seismic-resistant and/or building code (or lack thereof) in each country regardless of enforcement and building practice factor

The seismic code index ranks the quality of seismic codes since 1900 in each nation, giving a score between 0 and 104, normalized to 100 as two extra components were added later (Daniell et al. 2011). The criteria was set that a country with a unified national code was placed at a value of 30/100, utilizing the minimum criteria set which is the use of earthquake loading in some form of design (not necessarily for all structures). The quality of this code was then objectively ranked by using the above criteria elements to make an assessment of the code quality at the time of implementation.

It can be seen from the 2012 picture of the seismic code index in Fig. 4 that much work is needed solely on the writing of seismic code indices. Despite the adherence or intention to adhere to the new Eurocodes in Africa and through Asia, many nations still have not signed or enforced codes with respect to seismic resistance.

The current state of the world seismic codes with respect to nonengineered buildings is also a key concern, given the lack of seismic codes for residential buildings and the age of buildings. Many of these buildings are largely vulnerable, use heavy or inappropriate materials, and are not built based on historical earthquake or future

earthquake influences. Bilham (2013) estimates that there are 1–3 billion family homes in this category. However, many societies have learned to build with lightweight materials such as thatch, thus protecting themselves from fatal roof and wall collapses. The age distribution of buildings thus needs to be explored.

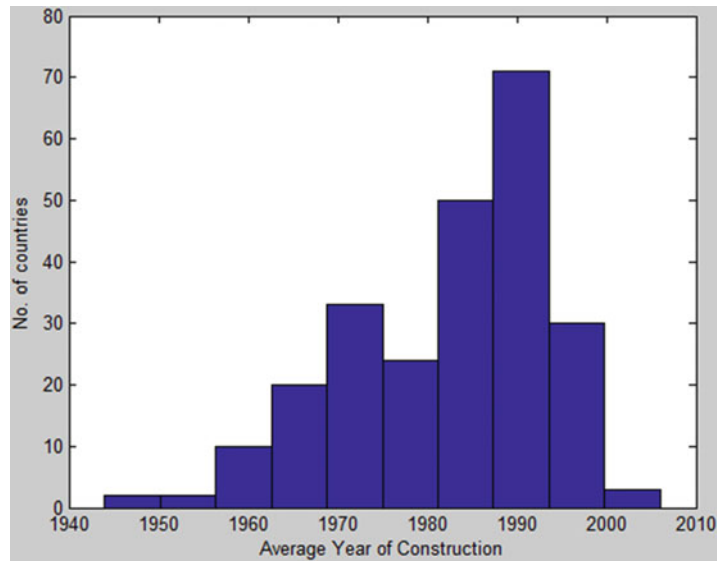
Age Distribution of Infrastructure and Its Effect on the Number of Buildings Under Seismic-Resistant Codes

A global building inventory including age and number of stories has been built (Daniell et al. 2011) using census information. The number of possible buildings that were built under some form of code was then reduced, as there are codes where not all building typologies are included under the seismic-resistant codes. The age of buildings has also been used to give a comparison as to the influence of a seismic code on the vulnerable percentage of buildings.

By using the same census data and methodology, the age distribution of infrastructure was collected, with the average shown in Fig. 5.

Global View of Seismic Code and Building Practice Factors,

Fig. 5 Average year of construction for the 244 nations worldwide using census data



G

In countries such as Australia, the average age of infrastructure has stayed fairly constant at around 18 years from 1950 to 2005 (ABS 1995–1996), with a dip toward 1971 as building increased, and is now currently at about 18 years. In terms of the average lifetime of a building, this value is quoted as approximately 75 years. By comparison, this value is around 30 years in China but in Great Britain is around 132 years. For European buildings, a study named “Housing Statistics in the European Union” used census data and studies in order to examine the age of buildings (EUROSTAT 2013). The large difference in the design life of structures leads to an interesting problem when filling in the remaining countries.

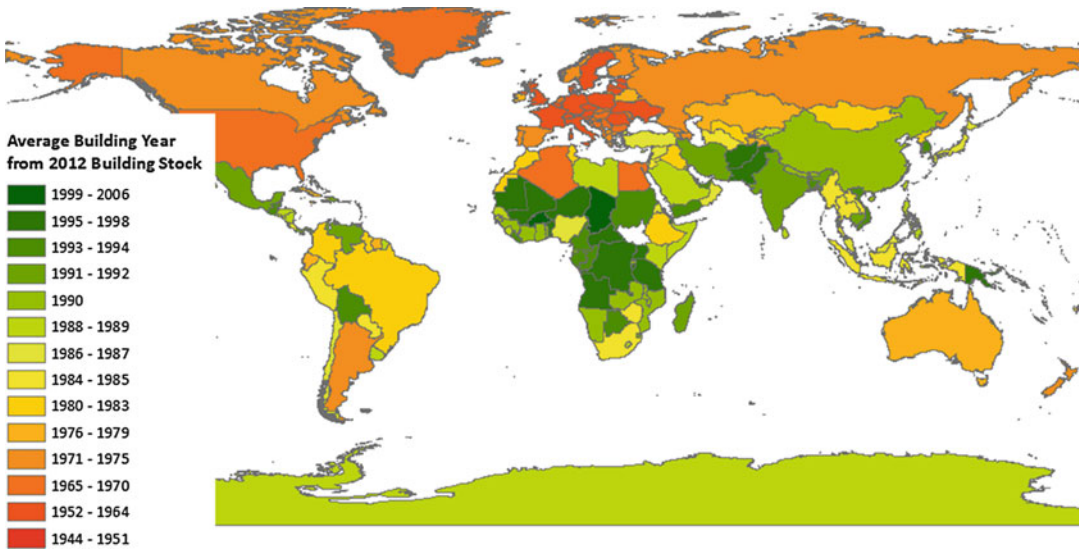
If there is not enough new investment in buildings, structures, and infrastructure, the average age of buildings will increase, if there is no new population. As an example, Germany has had very low population growth since 1991. The age of their infrastructure increased from 23.7 years to 26.9 years in 2009. Their general government buildings increased from 22.1 to 28.4 years as compared to dwellings from 25.1 to 27.8 years. Figure 6 shows the average building year in each country as a result of the analysis.

Figure 7 depicts the locations in the world that should be under seismic code as per GSHAP (1999)-derived intensity exceedances for

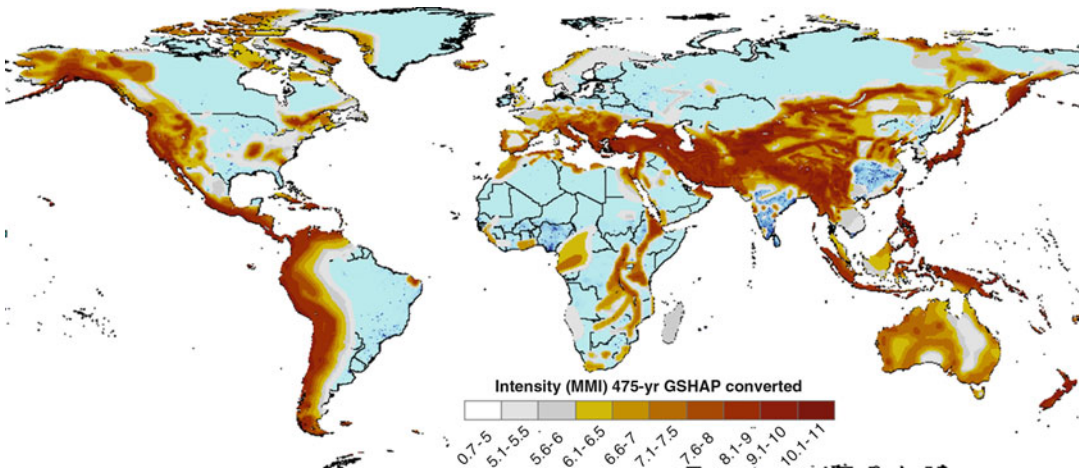
475 years globally (Giardini 1999). Different intensity-PGA relations were used in Fig. 7 to convert PGA to MMI depending on the location worldwide. It can be seen that most nations should thus have some form of seismic code, and when looking at longer return periods, even more of the globe can be exposed to damaging earthquake intensities.

The percentage of structures in each country under seismic code differs greatly when only the age percentage of structures is only looked at; however, there are usually clauses that buildings of one or two stories are not built to the earthquake code. By using household size and building counts globally, a percentage of buildings built under code has been estimated. The age of structures in each country is integrated against the seismic code index in order to create a percentage of structures that are definitely not built under code. This is, of course, assuming that the age percentage of the buildings is the same over the whole country, which is a reasonable assumption as the natural weighting is toward greater building in cities, which is represented by the entire country values.

The locations of seismic code zonation were then split from the non-code zones by calculating the percentage of buildings in these zones. It should be noted that as a proxy, the $MMI > 6$



Global View of Seismic Code and Building Practice Factors, Fig. 6 The average building year in each country



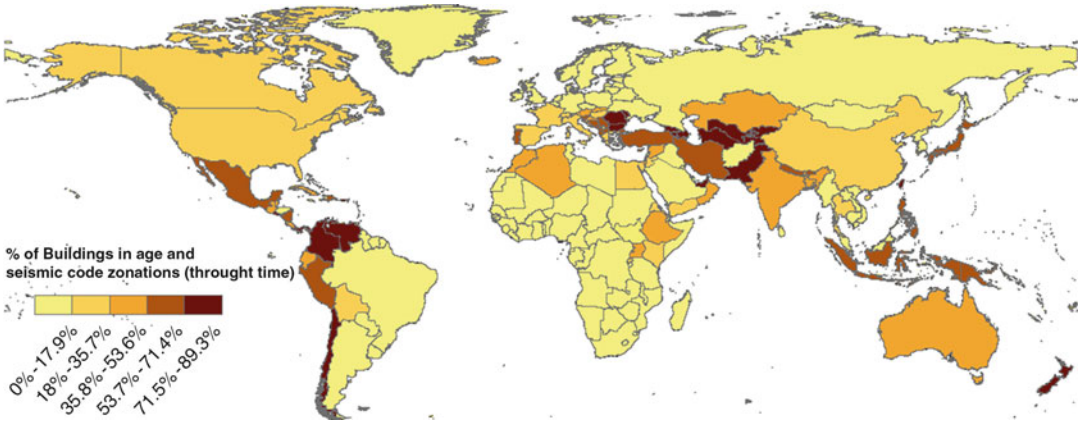
Global View of Seismic Code and Building Practice Factors, Fig. 7 The locations in the world that should be under seismic code as per GSHAP-derived intensity

exceedances for 475 years globally. Different intensity-PGA relations were used to convert PGA to MMI, depending on the location worldwide

contour was used from GSHAP, as the seismic zone maps collected have not yet been digitized. However, these maps were looked at to check the assumptions, with major changes being made in Italy (where historical codes have not covered the whole of Italy) and China.

Unfortunately, every building type is not covered by earthquake-resistant codes. The next provision that is examined is that of small buildings or those not covered under code. In order to

calculate this value, the individual building codes were looked at. It is difficult to make an assessment in some respects as to small buildings (residential 1–2 stories, or other forms) as, although these are included in some codes, there are many exemptions. The Chinese code (PRC 2001/2008/2010), for instance, uses intensity zones and story heights. Most rural housing is only 1 story and is thus not covered under earthquake-resistant codes. In addition, many



Global View of Seismic Code and Building Practice Factors, Fig. 8 The percentage of buildings in each country that should have been built under a seismic-resistant code (without small building provision)

G

urban buildings under 4 stories in the intensity zone 6 and 3 stories in the intensity zone 7 do not need tie columns for brick buildings. There is an exception for 1-story public buildings, which need to be built under code (2001 code, Chaps. 7 and 10). In Australia, all structures are included; (Standards Australia (1979, 1993, 2007)) however, for domestic buildings less than 3 stories, which are ductile, no provision is required. This code, being an objective-based code, also provides a good basis for engineers to check that the code is being adhered to. An approximate percentage of buildings coming under the provisions have been recorded. In addition, the number of informal and rural buildings is also taken into account.

Existing buildings are also included in some seismic-resistant codes in terms of requiring retrofitting, such as in Argentina, Italy, EC8 (European Union), the USA, Korea, Mexico, New Zealand, Nicaragua, and Venezuela; however, it is unsure as to what extent this has been enforced currently and thus has not been included.

By combining the age of structures which could have a seismic code and the number of nonengineered structures not built to code structures in those zones, a relative percentage of buildings in earthquake zones coming under code can be established. It should be noted that through Central Asia many buildings, although

built to former Soviet codes, have become dilapidated and thus also should be retrofitted or rebuilt.

In general, those countries which have high percentages in Fig. 8 are smaller countries with most of their building stock built in the last 30 years. Marked differences occur across nations with different zonations and building ages across nations not described by such diagrams.

Seismic Code Provisions: Germany

An example is presented below for Germany, in order to show the analysis procedure. The DIN4149 was first implemented in Germany in 1981 after the 1978 Albstadt earthquake. The update to this was given in 2005. Before this, a few other codes had been applied on regional levels, without any federal application (Abrahamczyk et al. 2005). The German code has five zones (no zone, 0, 1, 2, and 3).

Only 31 % of Germany's building stock has been produced after 1981. The code actually applied in zones 1–3 in terms of earthquake loss is for very little of the building stock (15.3 %). This means that under 5 % of possible buildings have been built under some form of earthquake-resistant code. In DIN4149, not every building needs to have had the code change applied, due to the premise of the number of stories (Table 2);

Global View of Seismic Code and Building Practice Factors, Table 2 Various limits on the requirements not to adhere to earthquake codes in zones 1–3

Earthquake zone	Importance class	Maximum story height
1	I–III	4
2	I–II	3
3	I–II	2

thus, even less than 5 % of buildings are applicable to this code. State enforcement ratification of codes is necessary in order for the DIN4149 to be used. The 2005 version has been ratified by most states; however, the new Eurocodes have only currently (as of mid-2013) been ratified by Baden-Württemberg, a state in the southwest of Germany (European Union 2006).

In zone 1, 89 % of buildings do not need to be built to code. In zone 2, around 75 % of buildings do not need to be built to code and in zone 3 around 48 %. This means that less than 0.9 % of buildings in Germany are built to a seismic-resistant code. When focusing on the population and value of the 0.9 % of buildings, this increases to approximately 3 % of capital stock, given the multi-story buildings, infrastructure, and important buildings in importance classes 3 and 4 protected by the earthquake code (hospitals, schools, industrial facilities, etc.). However, a more important problem is posed by the existing hazard, building typology, and standard building code. This formulates a similarly bleak engineered view of the world in terms of loss functions, which presents an additional difficulty in applying a single building approach.

Building Practice Versus Code Compliance

Just because a country has a seismic code in law, this does not mean that the seismic code will be enforced. Thus, building practice plays a major role in earthquake-resistant building, and not only the quality of the seismic code is important but its enforcement, adherence, and lack of corruption in the building. If a building was designed

with a high-quality code, yet the builders were not made to adhere to that code and used substandard materials, then the building would likely not stand up to the earthquake forces.

Corruption has been identified as a key component of earthquake losses, first through the Bilham and Hough (2006) paper in India and then Ambraseys and Bilham (2011) and Bilham (2013) exploring the Transparency International corruption index with respect to earthquakes. There is no doubt that engineered constructions have increased the quality of building against earthquake forces, but much of this has to do with the education of the individuals of that country. Education can be used as a proxy for the quality of understanding of building practice as well as the quality of the mathematics, engineering, and drawings that govern building construction. A study looking at the number of universities offering earthquake-related subjects is another proxy that has been previously used (Santa-Ana et al. 2012). Building error and mistakes increase with lack of education.

Summary

Seismic-resistant design codes have been implemented in many nations around the world over the past century. A summary has been provided as to their current application and to the extent of their use. Although many countries have implemented a seismic-resistant building code, the current percentage of buildings covered by these codes is generally less than 50 % of the building stock.

The enforcement and implementation of seismic-resistant codes is important as even though a country may have a very high-quality code, if it is not implemented or enforced by the government, the code is useless. Given the exponential rise in the number of seismic-resistant codes and the quality of these codes, it can be seen that more work is needed in implementing seismic-resistant codes for existing buildings and assessing residual risk, and care must be taken when referring to the use of codes without looking at exclusion clauses.

Cross-References

- [Building Codes and Standards](#)
- [Damage to Infrastructure: Modeling](#)
- [European Structural Design Codes: Seismic Actions](#)
- [Structural Design Codes of Australia and New Zealand: Seismic Actions](#)

References

- Abrahamczyk L, Langhammer T, Schwarz J (2005) Earthquake regions of the Federal Republic of Germany – a statistical analysis (in German). *Bautechnik* 82(8): 500–507
- Ambraseys NN, Bilham R (2011) Corruption kills. *Nature* 469(7329):153–155
- Australian Bureau of Statistics (ABS) (1995–1996) Australian National Accounts: estimates of capital stock, catalogue no 5221, various issues. Retrieved from <http://www.abs.gov.au/AUSSTATS/abs@.nsf/ProductsbyCatalogue/8D771B84048335BCA25722E001A7CC8?OpenDocument>
- Beattie G, Thurston SJ (2006) Changes to the seismic design of houses in New Zealand. In: *Proceedings, 2006 New Zealand society for earthquake engineering conference*, New Zealand Society for Earthquake Engineering, Wellington
- Bilham R (2013) Societal and observational problems in earthquake risk assessments and their delivery to those most at risk. *Tectonophysics* 584:166–173
- Bilham R, Hough S (2006) Future earthquakes on the Indian subcontinent: inevitable hazard, preventable risk. *South Asian J* 12:1–9
- Chin MW, Association of Caribbean States (2003) Model building codes for earthquakes and wind loads. Retrieved from <http://www.eird.org/cd/acs/English/enmodel.html>
- Daniell JE, Wenzel F, Khazai B, Vervaeck A (2011) A Country-by-Country Building Inventory and Vulnerability Index for Earthquakes in comparison to historical CATDAT Damaging Earthquakes Database losses. In: *Australian Earthquake Engineering Society 2011 conference*, Barossa Valley
- European Union (2006) EN1998-1:2004. Eurocode 8: design of structures for earthquake resistance: part 1: general rules, seismic actions and rules for buildings. Brussels: Comité Européen de Normalisation
- Giardini D (1999) The global seismic hazard assessment program (GSHAP)-1992/1999, *Annals of Geophysics*, 42(6)
- International Association of Earthquake Engineering (IAEE) (1996) Regulations for seismic design – a world list. Retrieved from <http://www.iaee.or.jp/worldlist.html>
- International Association of Earthquake Engineering (IAEE) (2000) Regulations for seismic design – a world list, Supplement 2000. Retrieved from <http://www.iaee.or.jp/worldlist.html>
- International Association of Earthquake Engineering (IAEE) (2013) Regulations for seismic design – a world list. Retrieved from <http://www.iaee.or.jp/worldlist.html>
- International Institute of Seismology and Earthquake Engineering (IISEE) (2012) Seismic design code. Retrieved from http://iisee.kenken.go.jp/net/seismic_design_code
- McGuire RK (ed) (1993) The practice of earthquake hazard assessment. U.S. Geological Survey, Denver
- Ministry of Construction (PRC) (2001/2008/2010) Chinese seismic design code: code for seismic design of buildings (GB50011-2001, 2008) (in Chinese), GB 50011-2010. China Building Industry Press, Beijing
- Priestley MJN (1993) Myths and fallacies in earthquake engineering – conflicts between design and reality. *Bull New Zeal Natl Soc Earthquake Eng* 26(3):329–341
- Priestley MJN, Calvi GM, Kowalsky MJ (2007) Displacement-based seismic design of structures. IUSS Press, Pavia; Distributed by Fondazione EUCENTRE
- Romeo RW (2007) Italian seismic hazard: experiences and new building code application. In: *Geohazards conference and EFG council meeting 2007*. Retrieved from http://www.icog.es/_portal/uploads/pub_fedgeol/12.pdf
- Salic RB, Garevski MA, Milutinovic ZV (2010) The need for advanced seismic hazard assessment of the Republic of Macedonia. In: *Proceedings of 14th European conference of earthquake engineering (ECEE) 2010*, Ohrid
- Santa-Ana PR, Santa-Ana LG, Baez Garcia J (2012) What we have forgotten about earthquake engineering, paper no 4566. In: *Proceedings of the 15th world conference on earthquake engineering*, Lisbon
- Standards Australia (1979) The design of earthquake resistant buildings. AS2121-1979. Standards Australia, Sydney
- Standards Australia (1993) Minimum design loads on structures: part 4 – earthquake loads. AS1170.4-1993. Standards Australia, Sydney
- Standards Australia (2007) Structural design actions: part 4: earthquake actions in Australia. AS 1170.4-2007, 2nd edn. Standards Australia, Sydney
- Standards New Zealand (2004) NZS 1170.5:2004 - Structural design actions. Earthquake actions., Wellington, New Zealand
- Statistical Office of the European Union (EUROSTAT) (2013) Housing statistics: tables and figures. Retrieved from http://epp.eurostat.ec.europa.eu/statistics_explained/images/5/59/Housing_statistics_YB2014.xls

Graphics Processing Unit (GPU) Technology in Earthquake Engineering, Application of

Xinzheng Lu and Zhen Xu

Key Laboratory of Civil Engineering Safety and Durability of China Education Ministry,
Department of Civil Engineering, Tsinghua University, Beijing, People's Republic of China

Synonyms

Earthquake engineering; GPU (Graphics Processing Unit); Massive parallel computing; Seismic damages simulation

Overview

The Graphics Processing Unit (GPU) located in the display card of a computer was originally designed for graphics display. Due to advances in massive parallel computing capacity, the current applications of GPUs have grown far beyond the graphics field.

General-Purpose computation on Graphics Processing Units (GPGPU) was proposed as early as the 1990s. However, such technology was not widely applied for general-purpose computation because, at the time, the application of GPGPU required a graphics application programming interface (API) (e.g., DirectX, OpenGL), which became a barrier to effective coding in general-purpose computation. In addition, many other difficulties have existed for GPGPU with respect to hardware, such as the exchange of data between the GPU and CPU. In 2006, the GPU G80 and the corresponding Compute Unified Device Architecture (CUDA) programming model were released by the NVIDIA Corporation (NVIDIA 2013), which was the first unified shader architecture and solved many of the GPGPU software and hardware problems. Since that time, GPU technology has been widely used in many fields, including many applications in earthquake engineering

involving seismic wave propagation and seismic damage prediction.

GPU Technology

Hardware Platform for GPU Computing

The term “GPU” was introduced by NVIDIA in 1999 with the release of GeForce 256, which is known as “the world’s first ‘GPU’”. Subsequently, ATI Technologies (acquired by AMD in 2006) introduced the visual processing unit (VPU) in 2002, which is not essentially different from a GPU. In 2006, the GeForce 8 series and Compute Unified Device Architecture (CUDA) were introduced by NVIDIA and constituted the earliest and subsequently most widely adopted programming model for GPU computing. In response, GPU hardware had undergone rapid development. Since 2006, the development history of the architectures for the two major types of GPUs may be depicted as follows:

NVIDIA GPUs: G80 (Nov. 2006) – GT200 (Jun. 2008) – Fermi (GF100/GF110, Mar. 2010/Nov. 2010) – Kepler (Mar. 2012).

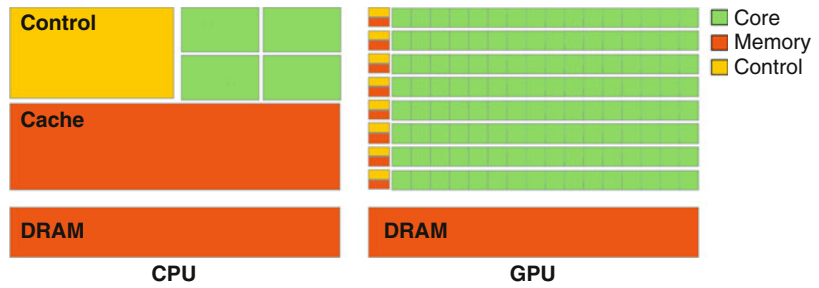
ATI/AMD GPUs: R600 (May 2007) – RV770 (Jul. 2008) – RV870 (Sep. 2009) – Cayman (Dec. 2010) – Tahiti (Jan. 2012).

At present, Intel, NVIDIA, and AMD/ATI are the market share leaders. Although over half of the market share is owned by Intel, most of the GPUs from Intel are Intel’s integrated graphics solutions, the performance of which is not as high as that of the GPUs introduced by the other two companies. Excluding these integrated GPUs, NVIDIA and AMD/ATI control over 95 % of the market. In addition, there are other companies that produce GPUs, such as VIA Technologies and Matrox.

The reason behind the discrepancy that exists in floating-point capability between the CPU and the GPU is that the GPU is specialized for compute-intensive, highly parallel computation and therefore designed in such a manner that more transistors are devoted to data processing rather than data caching and flow control

Graphics Processing Unit (GPU) Technology in Earthquake Engineering, Application of,

Fig. 1 Comparison between the CPU architecture and the GPU architecture



(NVIDIA 2013), as schematically illustrated in Fig. 1. It is obvious that the GPU has much more cores than the CPU, as shown in Fig. 1, so the GPU is well suited to problems that can be expressed as data-parallel computations with a high level of arithmetic intensity (i.e., the ratio of the demand of arithmetic operations to memory operations).

Software Platform for GPU Computing

Currently, CUDA is the most widely used software platform for GPU computing around the world (NVIDIA 2013). In addition, the Stream platforms of AMD, OpenCL, and DirectCompute are also important GPU computing platforms.

CUDA is a software environment that allows developers to use C as a high-level programming language. Other languages, application programming interfaces or directive-based approaches are also supported, such as FORTRAN, DirectCompute, and OpenACC.

CUDA is carried out by means of a hybrid computing architecture that includes a GPU and CPU. The CPU is responsible for the sequential and logical tasks, and the GPU is used to implement the massive parallel computing tasks. A typical flow chart for a CUDA program is illustrated as follows. First, data are copied from the main memory to the GPU memory. Next, the CPU instructs the main processing and the GPU executes in parallel in each core. Finally, the result is copied from the GPU memory to the main memory.

Two cases of applications of the CUDA program from the CUDA C Programming Guide (NVIDIA 2013) are provided herein to present the CUDA platform in detail.

Example 1: Vector Addition CUDA C extends C by allowing the programmer to define C functions, called *kernels*, that, when called, are executed N times in parallel by N different CUDA threads, as opposed to only once like regular C functions.

A kernel is defined using the `__global__` declaration specifier and the number of CUDA threads that execute that kernel for a given kernel call is specified using a new `<<<...>>>` execution configuration syntax. Each thread that executes the kernel is given a unique thread ID that is accessible within the kernel through the built-in `threadIdx` variable.

As an illustration, the following sample code adds the two vectors A and B of size N and stores the result into vector C :

```
// Kernel definition
__global__ void VecAdd(float* A,
float* B, float* C)
{
    int i = threadIdx.x;
    C[i] = A[i] + B[i];
}
int main()
{
    ...
    // Kernel invocation with
    N threads
    VecAdd<<<1, N>>>>(A, B, C);
}
```

Example 2: Matrix Addition The following code adds two matrices, A and B of size $N \times N$, and stores the result into matrix C :

```
// Kernel definition
```



```

__global__ void MatAdd(float A
[N] [N], float B[N] [N], float C
[N] [N])
{
int i = threadIdx.x;
int j = threadIdx.y;
C[i] [j] = A[i] [j] + B[i] [j];
}
int main()
{
...
// Kernel invocation with one
block of N * N * 1 threads
int numBlocks = 1;
dim3 threadsPerBlock(N, N);
MatAdd<<<numBlocks,
threadsPerBlock>>>
(A, B, C);
}

```

For convenience, *threadIdx* is a three-component vector, by which threads can be identified using a one-dimensional, two-dimensional, or three-dimensional thread index, thus forming a one-dimensional, two-dimensional, or three-dimensional thread block. This provides an easy means to invoke computation across the elements in a domain, such as a vector, matrix, or volume.

The index of a thread and its thread ID are related to each other in a straightforward way: For a one-dimensional block, they are the same; for a two-dimensional block of size (D_x , D_y), the thread ID of a thread of index (x , y) is ($x + y D_x$); for a three-dimensional block of size (D_x , D_y , D_z), the thread ID of a thread of index (x , y , z) is ($x + y D_x + z D_x D_y$).

Earthquake Engineering Applications

Simulation of Ground Motion

The propagation and attenuation of a seismic wave is an important issue in earthquake engineering. Abdelkhalek et al. (2012) adopted a GPU to solve the problem of seismic wave propagation. Considering both the 2D and 3D cases, this group used the GPU Tesla S1070 to perform an analysis of a seismic wave based on a finite

difference approach, obtaining an increased speed of ten times for reverse time migration and up to 30 times for a sequential simulation of seismic waves. Komatitsch et al. (2009) implemented a higher-order finite element application of seismic wave propagation based on two types of GPUs, the NVIDIA GeForce 8800 GTX and the GTX 280, and the acceleration ratios reached 15 and 25, respectively.

Okamoto et al. (2011) used the Japanese TSUBAME-1.2, supercomputer to implement a finite-difference-based simulation of seismic wave propagation. In this simulation, the computing efficiency of a single GPU (Tesla S1070s) reached 56 GFLOPS, which is 45 times the efficiency of a CPU (Intel Core i7). Unat et al. (2012) accelerated a 3D finite difference earthquake simulation with a C-to-CUDA Translator. Running on the Tesla C2050, their program was faster than a highly tuned message passing interface (MPI) implementation running on 32 Nehalem cores. Similarly, Zhou et al. (2012) designed 3D finite difference seismic wave propagation code based on the latest GPU Fermi chipset that implemented the asymmetric 3D stencil on Fermi in order to make the best use of the GPU on-chip memory and achieved aggressive parallel efficiency. The benchmark test on an NVIDIA Tesla M2090 demonstrated that a speedup factor of 10 was achieved compared with the original fully optimized AWP-ODC FORTRAN MPI code running on a single Intel Nehalem 2.4 GHz CPU socket (4 cores/CPU), and a speedup factor of 15 was achieved compared with the same MPI code running on a single AMD Istanbul 2.6 GHz CPU socket (6 cores/CPU). The sustained single-GPU performance of 143.8 GFLOPS in single precision mode was benchmarked for the test case of a $128 \times 128 \times 960$ sized mesh.

Structural Numerical Analysis

The GPU has been applied to structural numerical analysis. The finite element method is an important structural numerical analysis approach. In the elastic structural analysis based on finite element method, most of the computing time is spent on solving matrices. Many matrix computations are well suited to GPU computing. For

example, matrix multiplication can be easily processed by a GPU due to its natural parallel characteristics. Ryoo (2008) conducted a similar test for matrix multiplication with a program that attained an efficiency of 91 GFLOPS on the GPU of a GeForce 8800 GTX. Similarly, matrix decomposition, which is the foundation for solving a linear equation, can also be easily accelerated using a GPU. Volkov and Demmel (2008) optimized a matrix decomposition algorithm on the GeForce 8800 platform, and the decomposition efficiencies of the LU, QR and Cholesky methods were all greater than 180 GFLOPS.

Furthermore, Liu et al. (2008) proposed a general finite element method solution that enables dynamically fast simulation of deformation on the newly available GPU hardware with CUDA. Their test results indicate that the GPU with CUDA enables an increase speed of approximately four times for FEM deformation computation on an Intel(R) Core 2 Quad 2.0 GHz machine with GeForce 8800 GTX. Many commercial finite element analysis codes, e.g., ABAQUS, MSC. MARC, and ANSYS, have released their own GPU-based versions. The performance studies demonstrate that the GPU results in speedups in the range of 3–6 times.

In addition to the finite element simulation, other structural numerical simulations also can be greatly accelerated by the use of GPUs. For example, Durand et al. (2012) simulated rock impact on a concrete slab using the discrete element method (DEM), and the use of a GPU (Tesla C2050) reportedly showed a speedup factor of 30.

Earthquake Detection

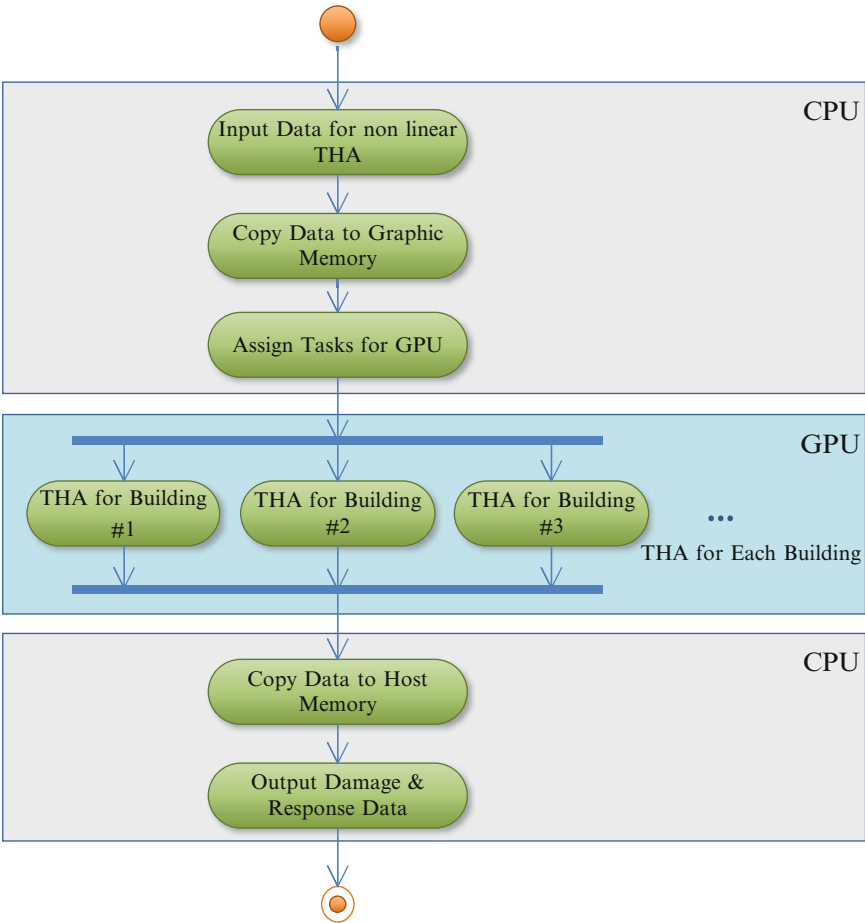
Meng et al. (2012) proposed a GPU-based computation method to accelerate the detection algorithm for earthquakes during intensive after-shocks or swarm sequences. This group applied parallel code to search the Salton Sea geothermal field for missing earthquakes in a 90-day time window around the occurrence time of the 2010 Mw7.2 El Mayor-Cucapah Earthquake. By dividing the procedure into several routines and processing them in parallel, this group achieved an approximate 40 speedup for one NVIDIA GPU card compared with the sequential CPU code.

Repeating earthquakes are occurring on the similar asperity at the plate boundary. These earthquakes have an important property; the seismic waveforms observed at the identical observation site are very similar regardless of their occurrence time. Kawakami et al. (2013) proposed a high-speed digital signal processing method for the detection of repeating earthquakes using GPGPU-acceleration. They compared the execution time between GPU (NVIDIA GeForce GTX 580) and CPU (Intel Core i7 960) processing. In the case of a band-limited phase only correlation, the obtained results indicate that a single GPU is approximately 8.0 times faster than a 4-core CPU (auto-optimization with OpenMP). In the case of the coherence function using three components, the GPU is 12.7 times as fast as the CPU. This study examines the high-speed signal processing of a huge amount of seismic data using the GPU architecture. It was found that the GPGPU-based acceleration for the temporal signal processing is very useful. In the future, multi-GPU computing will be employed to expand the GPGPU-based high-speed signal processing framework for the detection of repeating earthquakes in the future.

The Visualization of Seismic Data

The GPU can also be used for the high-performance visualization of earthquake data. Hsieh and Yang (2011) presented a method using volume rendering based on a GPU to visualize the spectral information of the 1999 Chi-Chi earthquake and the GPU-based volume-rendering system achieved a rendering speed of 30 frames per second (FPS). Xie et al. (2010) proposed a parallel visualization technique for large seismic datasets based on a GPU and CPU cooperation platform. Their experimental results showed that the method exhibited effective real-time performance for a notably large amount of seismic data (10 GB) on standard PC hardware.

Chen and Hsieh (2012) proposed a GPU-based method using volume rendering to better understand and analyze time-varying earthquake ground-motion data, including acceleration, velocity and displacement. This group used CUDA to implement volume rendering (ray



Graphics Processing Unit (GPU) Technology in Earthquake Engineering, Application of, Fig. 2 Flow chart of the seismic damage analysis using GPU

casting), numerical integration (trapezoidal quadrature) and numerical differentiation (centered finite difference), and achieved a speedup of approximately $100\times$ versus a CPU.

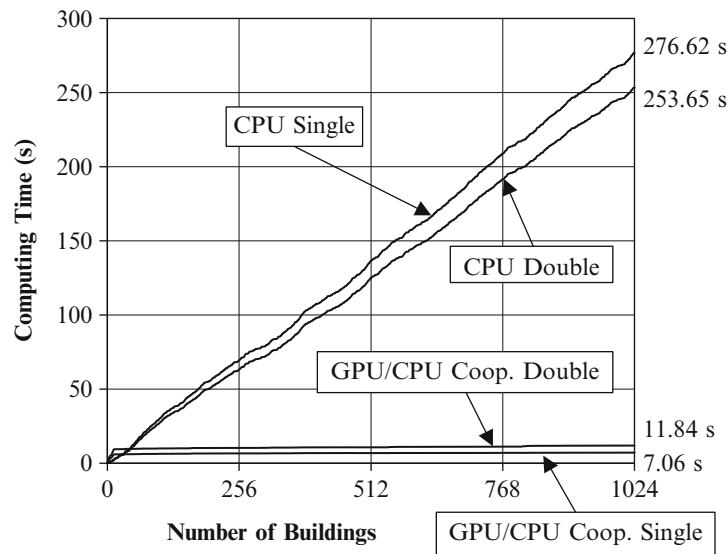
Seismic Damage Prediction

Although thousands of buildings may exist in an urban area, if each building is treated as a subtask and a proper computing model is adopted, the computing workload of each subtask is sufficiently small that it can be performed on a single GPU core. Furthermore, there is little interaction between different buildings during an earthquake, which results in there being little data exchange between different GPU cores. Because there are hundreds of cores in a GPU and each GPU core

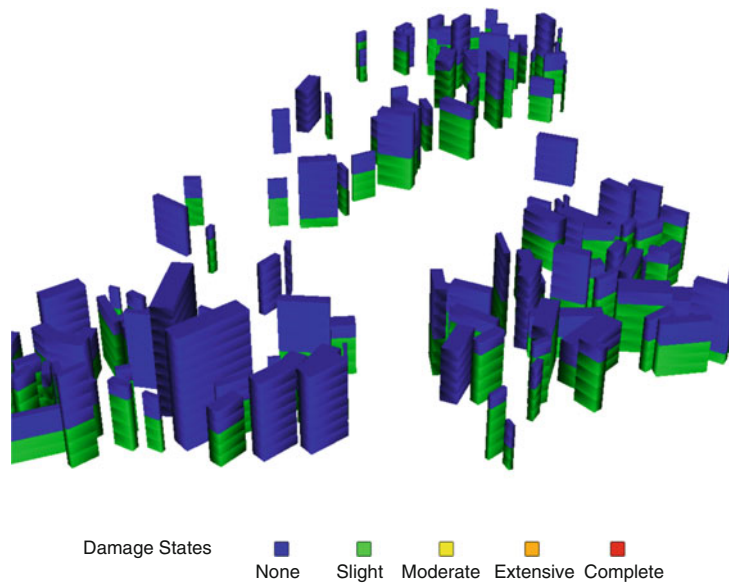
can be used for the simulation of one building, only a few task assignment rounds are required to complete the simulation of a city with thousands of buildings using GPU computing. Thus, the computational efficiency is expected to be notably high.

A novel solution proposed by Han et al. (2012) used a GPU to accelerate urban seismic damage prediction based on refined building models and nonlinear time-history analysis (THA). Figure 2 shows the flow chart of the Han’s program in which the CPU reads the data, copies the data into the GPU, and assigns the computational tasks to each GPU core, whereas the GPU implements the nonlinear THA for individual buildings and copies the results back to the host memory for output.

Graphics Processing Unit (GPU) Technology in Earthquake Engineering, Application of,
Fig. 3 Benchmark for the CPU and GPU/CPU cooperative programs



Graphics Processing Unit (GPU) Technology in Earthquake Engineering, Application of,
Fig. 4 Partial view of the damage states of each story



In the work of Han, a benchmark case of 1,024 buildings was investigated to validate the efficiency of seismic damage analysis based on GPU. The well-known El Centro record was selected as the ground motion for the benchmark. The peak ground acceleration (PGA) normalized to 200 cm/s².

Two hardware platforms with the same performance-to-price ratio were adopted, as follows:

CPU platform

Hardware: A 2.93-GHz Intel Core i3 530 processor with 4 GB of 1,333-MHz DDR3 RAM.

Compiler: Microsoft Visual C++ 2008 SP1.

GPU/CPU cooperative platform

Hardware: A 2.4-GHz Intel Celeron E3200 CPU and an NVIDIA GeForce GTX 460 with 1 GB of graphics memory.

Compiler: Microsoft Visual C++ 2008 SP1 and CUDA 4.2.

The benchmark cases demonstrated that the GPU/CPU cooperative computation time is approximately 1/39 times the CPU computational time if single precision is used, as shown in Fig. 3. This ratio increases to 1/21 if double precision is used. Seismic damage prediction based on a GPU is shown in Fig. 4, which provides an important reference for earthquake disaster prevention and mitigation.

In summary, existing studies have demonstrated that GPUs provide a highly efficient but low-cost environment for computation and visualization in earthquake engineering. Therefore, in the future, GPU technology will come to be widely used in earthquake engineering.

Acknowledgements The authors are grateful for the financial support received from the National Key Technology R&D Program (No. 2013BAJ08B02), the National Nature Science Foundation of China (Nos. 51222804, 51178249, 51378299, 51308321) and China Postdoctoral Science Foundation (2013M530632).

Cross-References

- [Earthquake Damage Assessment from VHR Data: Case Studies](#)

References

- Abdelkhalek R, Calandra H, Couland O, Roman J et al (2012) Fast seismic modeling and reverse time migration on a GPU cluster. *Concurrency Comput Pract Exp* 24:739–750
- Chen MD, Hsieh TJ (2012) Run-time GPU computing and rendering of earthquake ground-motion data. In: *Proceedings of the IEEE 9th international conference on high performance computing and communication*, pp 812–817
- Durand M, Marin P, Faure F, Raffin B (2012) DEM-based simulation of concrete structures on GPU. *Eur J Environ Civil Eng* 16:1102–1114
- Han B, Lu XZ, Xu Z (2012) Urban regional seismic damage prediction based on GPU-CPU hybrid computing. In: *Proceedings of the 15th world conference in earthquake engineering*, Lisbon
- Hsieh TJ, Yang YS (2011) Visualizing the seismic spectral response of the 1999 Chi-Chi earthquake using volume rendering technique. *J Comput Civil Eng-ASCE* 26:225–235
- Kawakami T, Okubo K, Uchida N, Takeuchi N, Matsuzawa T (2013) High-speed digital signal processing method for detection of repeating earthquakes using GPGPU-acceleration. *EGU General Assembly 2013*, Vienna
- Komatitsch D, Michéa D, Erlebacher G (2009) Porting a high-order finite-element earthquake modeling application to NVIDIA graphics cards using CUDA. *J Parallel Distrib Comput* 69:451–460
- Liu Y, Jiao S, Wu W, De S (2008) GPU accelerated fast FEM deformation simulation. In: *IEEE Asia Pacific conference on circuits and systems*, Macao, pp 606–609
- Meng XF, Yu X, Peng ZG, Hong B (2012) Detecting earthquakes around Salton Sea following the 2010 Mw7.2 El Mayor-Cucapah earthquake using GPU parallel computing. *Procedia Comput Sci* 9:937–946
- NVIDIA (2013) CUDA programming guide. http://docs.nvidia.com/cuda/pdf/CUDA_C_Programming_Guide.pdf
- Okamoto T, Takenaka H, Nakamura T, Aoki T (2011) Accelerating large-scale simulation of seismic wave propagation by multi-GPUs and three-dimensional domain decomposition. *Earth Planets Space* 62:939–942
- Ryoo S (2008) Program optimization strategies for data-parallel many-core processors. University of Illinois at Urbana-Champaign, Urbana
- Unat D, Zhou J, Cui Y, Baden SB et al (2012) Accelerating a 3D finite-difference earthquake simulation with a C-to-CUDA translator. *Comput Sci Eng* 14:48–59
- Volkov V, Demmel J (2008) LU, QR and Cholesky factorizations using vector capabilities of GPUs. Technical report UCB/EECS-2008-49. EECS Department, UC Berkeley
- Xie K, Wu P, Yang S (2010) GPU and CPU cooperation parallel visualization for large seismic data. *Electron Lett* 46:1196–1197
- Zhou J, Unat D, Choi DJ, Guest CC, Cui Y (2012) Hands-on performance tuning of 3D finite difference earthquake simulation on GPU Fermi chipset. *Procedia Comput Sci* 9:976–985

Historical Seismometer

Josep Batlló

Instituto Dom Luiz (IDL), Faculdade de Ciências da Univ. de Lisboa, Lisbon, Portugal

Synonyms

Accelerograph; Early seismic recording; Old seismographs; Seismometer; Seismoscope

Introduction

A seismometer is an instrument that detects ground motion and a seismograph is a seismometer together with a recorder that records it (a seismogram). Historical seismometers or ancient seismographs are those instruments developed previous to the design and development of the World-Wide Standardized Seismographic Network (WWSSN, Oliver and Murphy 1971). Other previous studies on this topic used to define old seismographs as those designed up to the very early years of the twentieth century because around those years the main basic features of such instruments were already the same as what our present knowledge recognizes as seismographs with analog recording. But it should be realized that the younger graduates in seismology and related topics rarely have had contact with seismic analog recording. For this

reason, some relevant instruments and topics on seismic recording developed on the first half of the twentieth century will be included here. Seismoscopes (a device that only indicates ground motion) and other devices designed and used previous to the development of seismographs will be also considered in this entry.

The ancient seismoscopes and seismographs will primarily be studied from a chronological point of view, but with specific insight into its relevance from the point of view of innovation, that is, their contribution to the evolution of the technical features and/or capabilities of seismic recording.

It is impossible to describe all the models of instruments developed. Therefore, preference will be given to those instruments showing some new development with respect to previous ones. For the readers more interested in this topic or for more details and primary references on the described instruments, there are the outstanding studies of Dewey and Byerly (1969), Howell (1990), and Ferrari (1992) where more information is available and to which this entry is highly indebted.

Recording Earthquakes

The history of early earthquake recording instruments is not a linear one or of continuous progress. It was clear from old times that earthquakes involved ground motion and, thus, recording the ground motion was an important issue to get

some insight in the earthquake phenomena. Also it cannot be ignored that rough devices to detect ground motion were known and used through antiquity and medieval times. It is confirmed that such a device was used on the Exeter siege in 1549. It was used to determine the location of enemies digging underground tunnels. The coeval document says: “hearing a noise underground, he takes a pan of water, and by removing it from place to place, came at length to the very spot where the miners were working; which he knew for certain by the shaking of the water in the pan.” But no application of this or other instruments for the recording or study of earthquakes is known from the renaissance up to the eighteenth century.

In fact, theoretical and technical issues were not developed enough at that time to allow the recording of ground motion. For these reasons the history of the development of early instruments for earthquake recording is one of accumulation of empirical knowledge and of trial and error. Only at the end of the nineteenth century were the key issues concerning ground motion recording started to be properly defined and mastered and the technology of instrumental ground motion recording started to progress steadily.

In order to clarify the following discussion, it should be remembered that, from a technical point of view, two main components can be identified in a ground motion recording instrument. The first is the transducer; this is a device that transforms ground shaking into some other variable (e.g., a voltage) that can be recorded. The second one is the recording system itself, a device allowing to visualize and to store the output of the transducer for further inspection and analysis. The third item is a signal-amplifying and preprocessing system or, in other words, a signal-conditioning system. But this was not a key issue in the early development of seismographs.

In the following sections a review will be made on the evolution of seismic recording up to the middle of the twentieth century. As it is impossible to describe all types of seismographs, the description will concentrate on the instruments that defined new characteristics or new evolution in the recording concepts. Instruments

not described here will be similar to the ones covered here.

An Overview of Seismic Recording

To better understand the descriptions of seismic instruments, some basic theoretical aspects will be described here. For more details, see e.g., Wielandt 2009 or Havskov and Alguacil 2010.

The Nature of Seismic Ground Motion

Earthquakes release energy as elastic waves. The frequency content of the generated waves depends on the size (total energy) released at the earthquake source. Larger earthquakes generate larger amplitudes at lower frequencies relative to small earthquakes. Attenuation of elastic waves propagating on Earth is larger for higher frequencies. Thus, with equal amplitude, lower-frequency waves propagate to longer distances than higher-frequency waves.

For these reasons seismographs adapted to record lower frequencies (<0.1 Hz) are needed to record large and distant earthquakes while instruments with larger amplification at higher frequencies (>1 Hz) are needed to record near, small earthquakes.

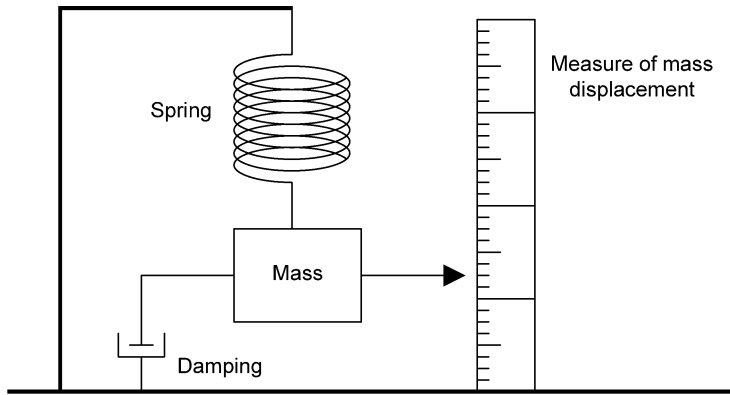
From Ground Motion to Seismic Recording

As it will be seen in the following sections, the most widely used principle for seismic recording is the inertia principle.

A device of such type, a mechanical sensor, has a mass with one degree of freedom coupled to a frame (the ground) through a linear restoring force (see Fig. 1). This can be achieved with a spring with restoring force k . A sensor will also have friction which will damp the swinging system. The equation of motion of this system can then be expressed as (Havskov and Alguacil 2010, p. 15)

$$\ddot{z} + 2h\omega_0\dot{z} + \omega_0^2 z = -\ddot{u},$$

where u is the ground displacement (the displacement of the frame); z the relative displacement between the frame and the mass and $\omega_0 = \sqrt{k/m}$,



Historical Seismometer, Fig. 1 Schematic illustration of the most simple seismometer, the inertial mechanical seismometer. A mass is held on a spiral spring. The spring connects the mass with the ground and introduces a linear restoring force. The motion of the mass is damped using

a “dash pot” so that the mass will not swing excessively near the resonance frequency of the system. A ruler is mounted on the side to measure the motion of the mass relative to the ground (From Havskov and Alguacil 2010)

H

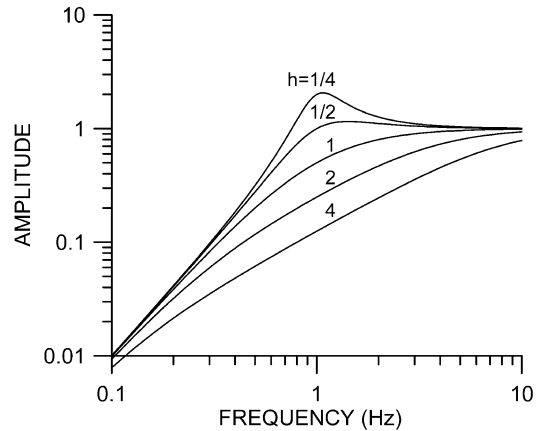
with $\omega_0 = 2\pi/T_0$; and T_0 is the sensor's natural period. The damping constant is h and it controls how much the system swings. For $h = 1$ (critical damping) there will be no overswing if the mass gets a push (see Fig. 2).

Thus, just two parameters are involved in the solution of the equation, the natural period T_0 and the damping h . If the system is submitted to a sinusoidal ground motion, the equation has an analytical solution, and the ratio between the mass displacement and ground displacement (amplitude response curve for displacement) is (Havskov and Alguacil 2010, p. 16)

$$A(\omega) = \frac{\omega^2}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4h^2\omega^2\omega_0^2}}$$

From this equation it is seen that for high frequencies $A(\omega) = 1$, so the sensor will measure pure displacement, and $A(\omega)$ diminishes rapidly for lower frequencies, as represented in Fig. 3a. Thus, it is possible to obtain a record (z) reproducing the amplitudes of the ground motion using a sensor with a natural frequency lower than the expected frequencies of the ground motion.

For low frequencies, $A(\omega) = \omega^2/\omega_0^2$ which is proportional to acceleration. Thus, the recording for frequencies lower than the sensor's natural



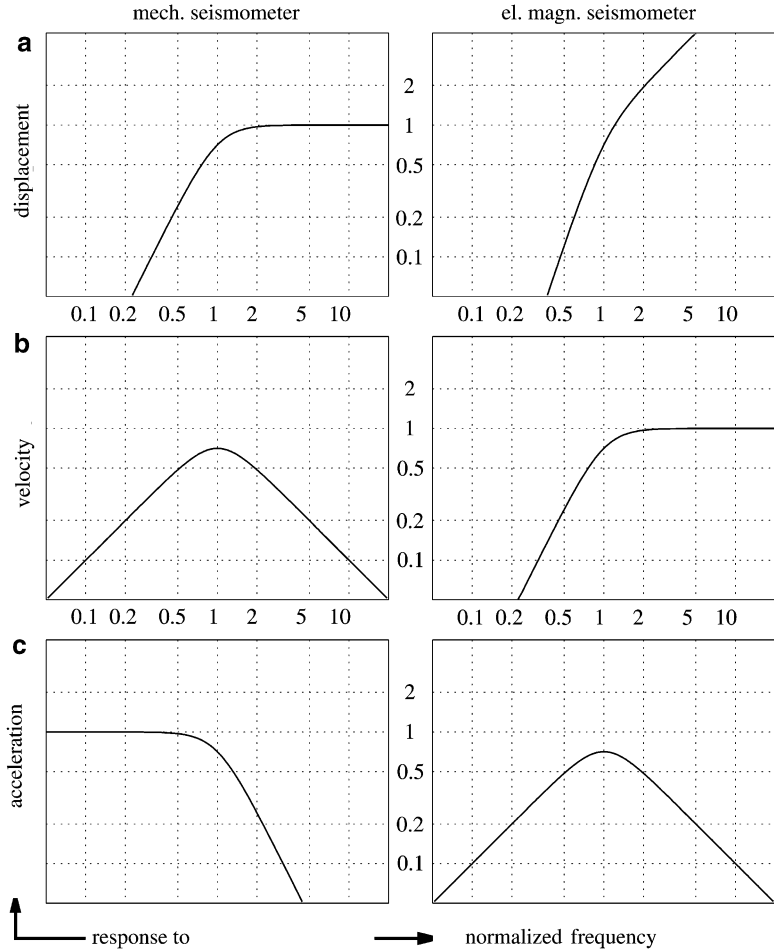
Historical Seismometer, Fig. 2 The amplitude response function $A(\omega)$ for a seismometer with a natural frequency of 1 Hz. Curves for various level of damping h are shown (From Havskov and Alguacil 2010)

frequency is directly related to the ground motion acceleration (Fig. 3c).

Such principle can be easily applied to the record of vertical and horizontal ground motion. The simplest sensor for vertical motion consists in a mass suspended by a spring (Fig. 1). However, getting a natural frequency of 0.1 Hz would require an unrealistic soft spring combined with a large mass so some tricks are needed; see below. In the case of horizontal motion, it is possible to avoid a spring by using gravity acceleration as the restoring force. The simplest way is

Historical Seismometer,

Fig. 3 Response curves of a mechanical seismometer and electromagnetic seismometer with respect to different kinds of input signals: (a) displacement, (b) velocity, and (c) acceleration. The normalized frequency is the signal frequency divided by the natural frequency (corner frequency) of the seismometer. Plots are in bi-logarithmic scale (From Wielandt 2009, Fig. 5.6)



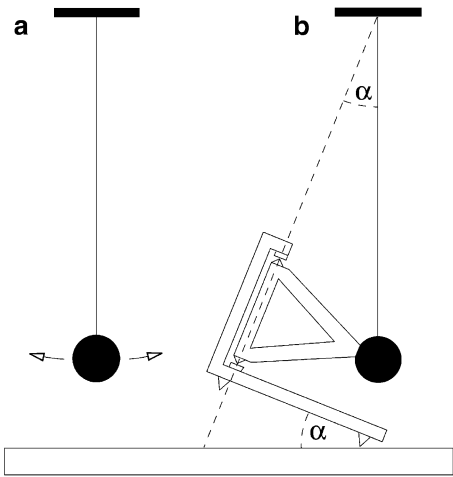
to use a common pendulum (Fig. 4a). In this case $\omega_0 = \sqrt{g/L}$, where g is the gravity acceleration and L the length of the pendulum. A problem arises when long periods are required since this would require a very large L . For a frequency of 0.1 Hz, $L = 25$ m, not very practical. This problem can be solved using the “garden-gate” sensor (Fig. 4b). In this case $\omega_0 = \sqrt{g \sin \alpha / L}$. Theoretically any period can be reached with such a configuration, but in practice, due to friction, it is hard to get a frequency lower than 0.03 Hz.

Electromagnetic sensors were introduced early in the twentieth century (Fig. 5). The output voltage of such a sensor is proportional to ground velocity for frequencies higher than the natural frequency (Fig. 3b). But good voltage recorders for electromagnetic sensors were not available

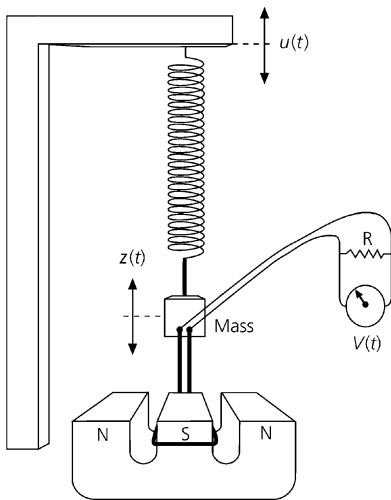
until the second half of the century. Early electromagnetic sensors were connected to galvanometers which record the output current generated by the sensor. Galvanometers have a natural period and damping as sensors so the whole response of the electromagnetic sensor-galvanometer is the result of combining both elements (see Fig. 6).

Seismoscopes: The First Known Instrument

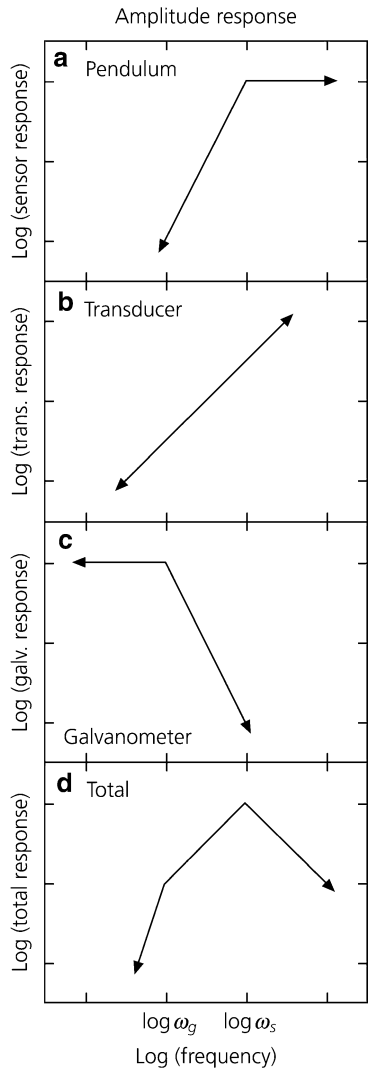
The first known devices used to detect ground motion are not seismographs but seismoscopes. These devices make it possible to observe ground shaking, but they do not give a time history of the



Historical Seismometer, Fig. 4 Schematic illustrations of (a) a common pendulum and (b) a tilted “garden gate.” For a free period of 20 s, the common pendulum must be 100 m long. The tilt angle α of a garden-gate pendulum with the same free period and a length of 30 cm is about 0.2° . The longer the period is made, the less stable it will be under the influence of small tilt changes (Adapted from Wielandt 2009, Fig. 5.8)



Historical Seismometer, Fig. 5 Schematic illustration of an electromagnetic seismograph, in which the mass is coupled to an electromagnetic transducer. $u(t)$ represents the ground motion and $z(t)$ the relative motion of the mass with respect to the frame. Motion of the coil attached to the mass through the magnetic field generates an electric current (Adapted from Stein and Wyssession 2003)



Historical Seismometer, Fig. 6 Schematic illustration of the amplitude responses of the components of an electromagnetic seismograph system using a galvanometer. ω_s represents the sensor’s natural frequency and ω_g the galvanometer’s natural frequency. The response of the electromagnetic sensor can be decomposed with the response of the mechanical sensor (a) and the response of the coil (b). The combination of this with the response of the galvanometer (c) gives the overall response of the system (d). All plots are drawn in bi-logarithmic scale (Adapted from Stein and Wyssession 2003)

motion (a seismogram). In its most basic form, they give an indication of ground shaking exceeding a threshold, but they do not give any temporal reference to when it happens or about the real shape of the motion. Others, more sophisticated,

may indicate the direction of shaking, time of the event, or maximum amplitude.

The first known device built specifically for the recording of earthquakes is a seismoscope of Chinese origin. It is acknowledged that it was designed by Zhang Heng (78–139) in 132 A.D (Zhang Heng is the name of the Chinese philosopher as written in the Pinyin transcription system. In many old articles, books, etc., it is found as Chang Hêng or also as Choko and Tyoko, Japanese adaptations of his name). It should be noted that this instrument has not been preserved, and what can be seen in books and museums are recreations produced from the descriptions of the instrument given by Fan Ye and others in the fifth century (Hsiao and Yan 2009).

From the translation of these descriptions given by Dewey and Byerly (1969) and Hsiao and Yan (2009), “The instrument was made of bronze and it resembled a wine jar with a diameter of eight chi (a Chinese measure equivalent to 2 m approx.). Inside, there was a pillar in the center and eight transmitting rods near the pillar. It carried out a switch ball and started the mechanism. On the outside of the vessel there were eight dragon-heads, facing the eight principal directions of the compass. Below each of the dragon-heads was a toad, with its mouth opened toward the dragon. The mouth of each dragon held a ball. At the occurrence of an earthquake, one of the eight dragon-mouths would release a ball into the open mouth of the toad situated below. The direction of the shaking determined which of the dragons released its ball and the instrument made a sound (probably the ball beating the toad) to warn the operator.” See Fig. 7a for a possible reconstruction. It is known that the instrument was operational because there is record of the detection of, at least, a 400-mile distant earthquake which was not felt at the location of the seismoscope (Hsiao and Yan 2009).

The Heng seismoscope is a noteworthy machine in itself. It has a place in the history of technology and it has been studied by many scholars (notice its place in time with respect to Western history realizing that it was built at the time of Roman emperor Hadrian and just half a century after the Vesuvius eruption that

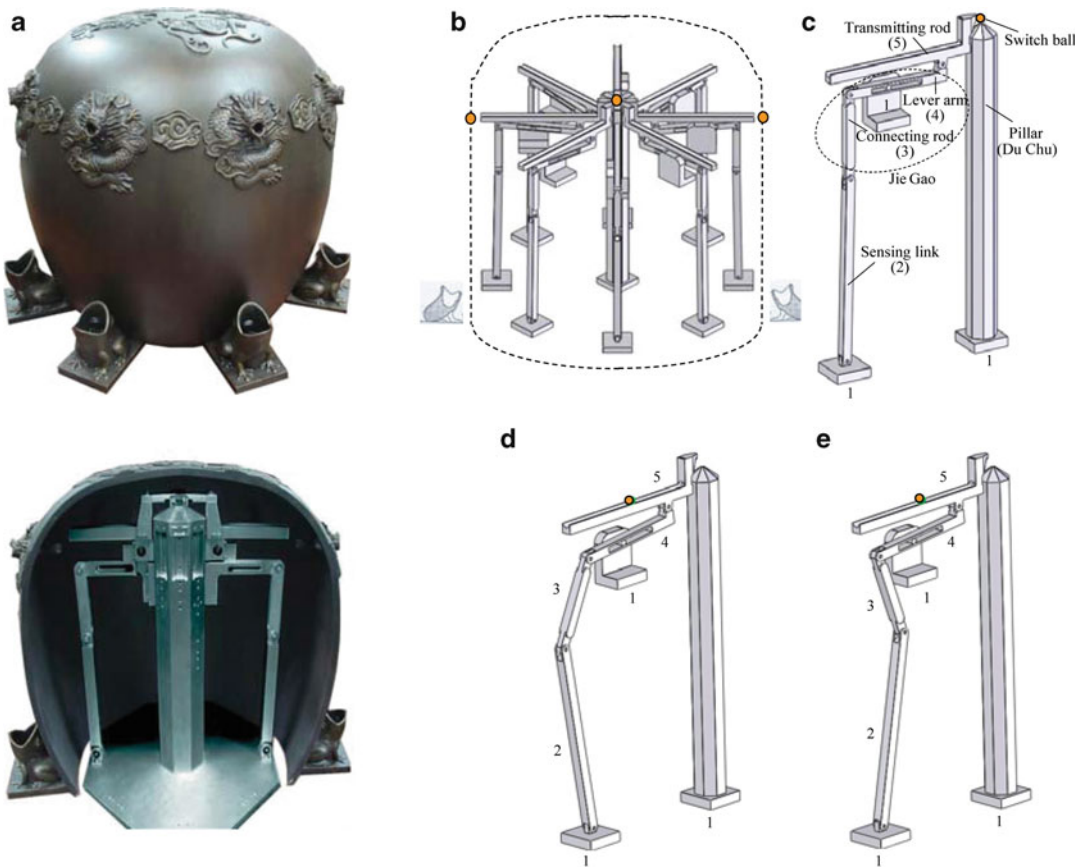
destroyed Pompeii). But it is also noteworthy from the seismological point of view because all studies on its possible detecting mechanism point to the use of the inertia principle, the same that led to the construction of most of the early seismographs. Moreover, it gave an indication not only of the occurrence of an earthquake but also of the direction of the shaking.

Many discussions have taken place about its exact internal mechanism, but a consensus has not been reached. Models with common or inverted pendulums and a model assuming that the oscillating element was the external jar suspended from the pillar have been proposed. Assuming the chronicles are correct, the most puzzling question is that only one ball was delivered and that it pointed in the direction of the epicenter. Thinking on simple mechanisms with pendulums, it looks like more than one ball can be delivered and the pointing direction may be toward or opposite to the epicenter. Anyway, the use of more complex designs, already known at that time, has shown the feasibility of such a mechanism (Fig. 7b–e).

Seismoscopes in Europe: The Eighteenth Century

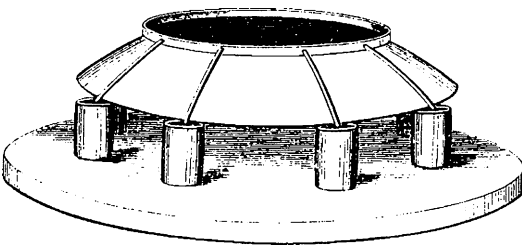
After Heng’s seismoscope, further references to such instruments disappeared from Eastern literature, and it is necessary to jump in time to the eighteenth century in Europe to find new instruments for earthquake recording. The first mention of such an instrument was written by Jean de Hautefeuille (1647–1724) who, in 1703, proposed a device to detect earthquakes. The background idea was very similar to that of Heng. He suggested to fill a bowl to the brim with mercury, so that the shaking produced by an earthquake would cause some of the liquid to spill out. In order to determine the direction of the shock, the mercury spilling out in each of the eight principal directions of the compass was to be collected in cavities (see Fig. 8 for a sketch of the instrument).

Jean de Hautefeuille was not using the inertia principle. His instrument was designed as an inclinometer because he was assuming that, in



Historical Seismometer, Fig. 7 (a) A possible reconstruction of the outside and inside parts of the Zhang Heng seismoscope. (b) and (c): the position of the eight rods and the detail of one of the rods. (d) and (e): detail showing the

feasibility of a mechanism delivering a ball pointing in the direction of the epicenter due to the compression and dilatation of the incident wave (Adapted from Hsiao and Yan (2009))



Historical Seismometer, Fig. 8 A simple sketch of the seismoscope designed by de Hautefeuille according to Hörnes (1893). A bowl is filled to the brim with mercury so that the shaking produced by an earthquake would cause some of the liquid to spill out. In order to determine the direction of the shock, the mercury spilling out in each of the eight principal directions of the compass is collected in cavities (just the front four are shown)

an earthquake, the ground would be inclined, and the direction in which the mercury spilled would be away from the origin, “where the ground began to be raised.” This is because updated Aristotelian theories about the origin of earthquakes in the eighteenth century suggested underground explosions caused by accumulations of sulfur, bituminous materials, coal, etc., causing dilatation. Thus, this earliest European instrument was expected to detect tilting of the Earth’s surface rather than horizontal displacements. Following the author, it could be possible to get an indication of the distance of the epicenter and the size of the disturbance from the amount of mercury which had sloshed out.

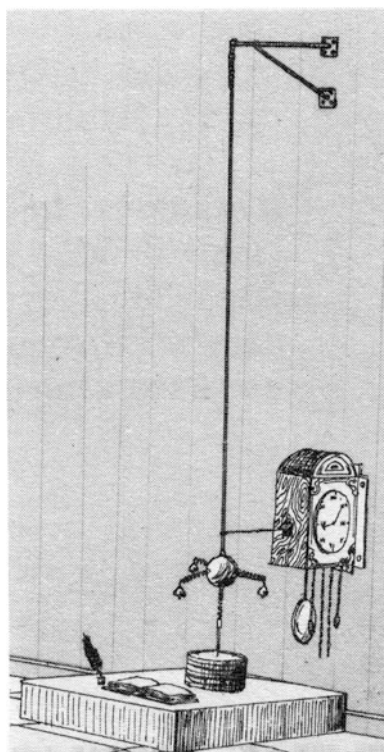
But, as far as it is known, this instrument remained as a design and it was never built.

Anyway, earthquake detection was a “hot topic” at that time, and in the following years a lot of references to seismoscopes are encountered in literature. Design of such instruments became popular among Italian scientists on the eighteenth century, and many devices were designed and developed.

Among them, Nicola Cirillo (1671–1735) is the first confirmed constructor and user of seismoscopes. It was the occasion of the Puglia earthquakes of March 1731 that made him construct the instrument. He used common pendulums of small dimensions (the length of the string was a Neapolitan palm, or 26 cm long approx.). With several instruments installed at different places, he was able to deduce that the amplitude of seismic motion attenuates with distance (Ferrari 1992).

Another instrument at that time worth mentioning is the common pendulum designed by Andrea Bina (1724–1792) in 1751. It is almost certainly the first design of a recording instrument in the history of seismology. According to Bina’s description, a sphere of lead of considerable weight was to be suspended with a rope from a beam on the ceiling to form a long common pendulum (several meters long). In the lower part of the sphere was a needle one and a half inches long penetrating slightly into a layer of sand placed in a wooden box. In the event of a tremor, the needle would leave marks in the sand that, which depending on their form, width, and depth, would have allowed the observer to draw conclusions as to the intensity, direction, and quality of the seismic movement (Ferrari 1992). As it is seen, this instrument could have given indications on the amplitudes and direction of the ground shaking, but without any timing reference. But, as in the case of de Hautefeuille seismoscope, it looks like this instrument was never built.

In 1796, Ascanio Filomarino (1751–1799) described an instrument which he called a seismograph but in fact was a seismoscope. It consisted of a long pendulum attached to a wall whose spherical mass brought down a pencil that, when shaken, marked a trace on a paper of



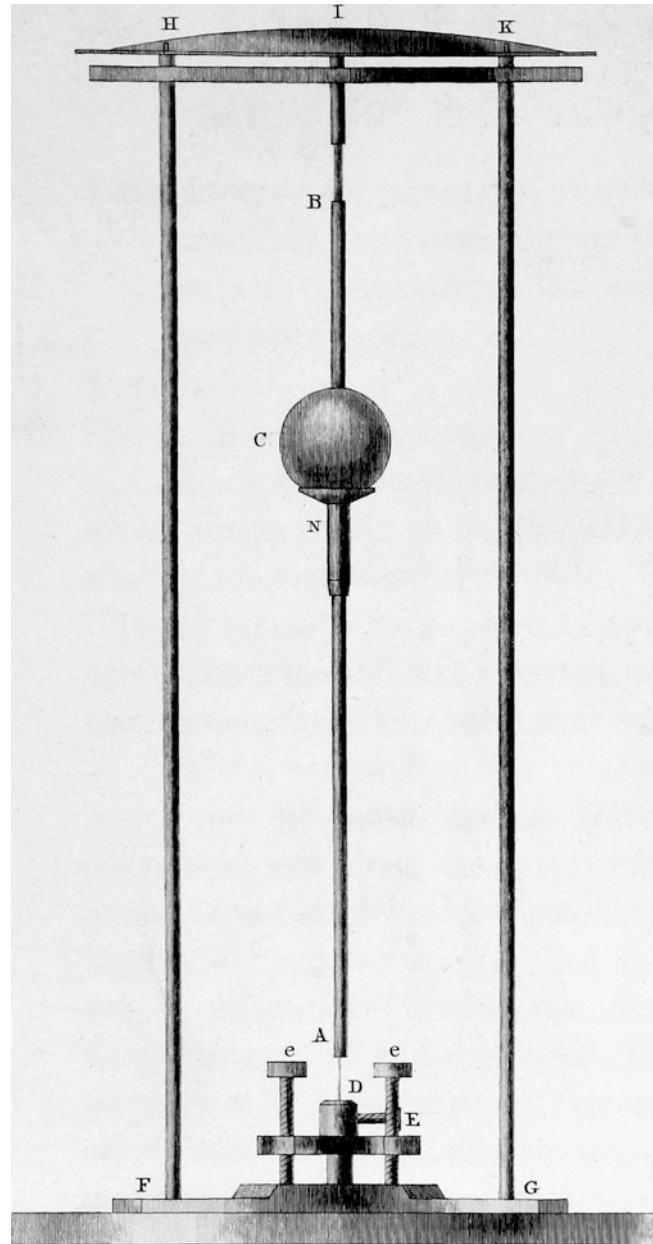
Historical Seismometer, Fig. 9 The seismoscope by Filomarino according to Ferrari (1990). See detailed description on the text

circular shape (3 in. diameter) placed on a support. In the sphere of iron there were also three bells connected which served as an alarm. Finally, and this is the innovative point, a wire of horsehair blocked the rocker of a clock. When the pendulum swung, the receding hair caused the clock to start, giving an indication on how long ago the earthquake had happened. In this case, the time of the shock is recorded (see a sketch in Fig. 9). It is known that several written records were obtained with this instrument. However, the observations given by Filomarino do not indicate that the timing device was functioning (Dewey and Byerly 1969).

Seismoscopes in the Nineteenth Century

There are references to many more seismoscopes from the middle of the eighteenth century, but it

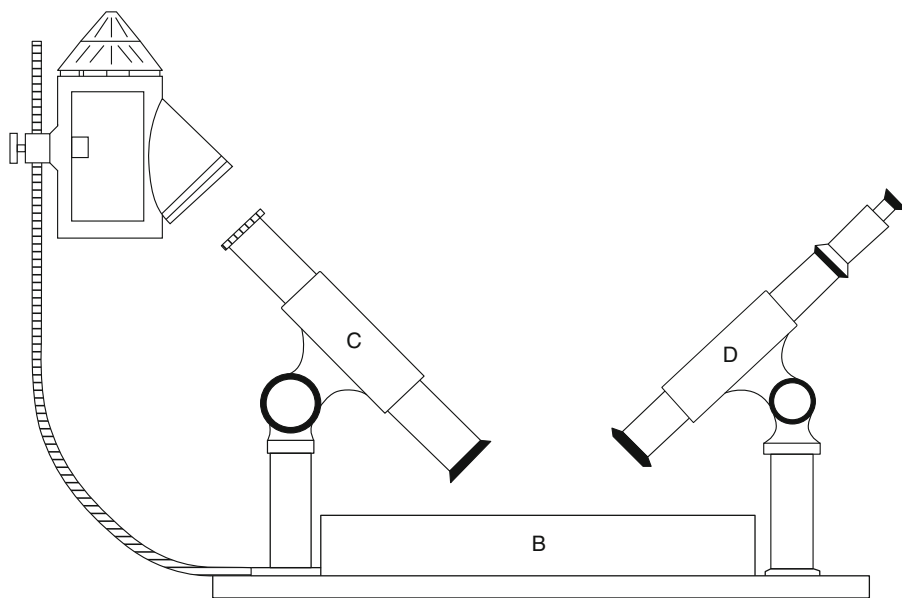
Historical Seismometer,
Fig. 10 The seismoscope
 of Forbes (1844). See
 detailed description on
 the text



H

is necessary to wait until the nineteenth century to find the first improved recording device. The most interesting instrument was designed by Scottish James D. Forbes (1809–1868). Due to a series of small earthquakes felt near Comrie, Scotland, a special committee was appointed by the British Association for the Advancement of Science being one of its goals to get instruments to record the shocks. Forbes built an inverted

pendulum (see Fig. 10). Using the same words of Forbes (1844), “a vertical metal-rod *AB*, having a ball of lead *C* moveable upon it, is supported upon a cylindrical steel-wire *D*, which is capable of being made more or less stiff by pinching it at a shorter or greater length by means of the screw *E* [...] The self-registering part of the apparatus [...] consists of a spherical segment *HIK* of copper lined with paper, against which a pencil *L*,



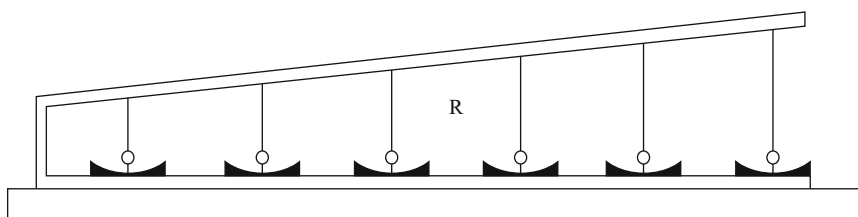
Historical Seismometer, Fig. 11 A sketch of the seismoscope of Mallet. The image of a crosshair in *C* is reflected from the surface of mercury in the basin *B* and viewed through a magnifier, *D* (From Dewey and Byerly 1969)

inserted in the top of the pendulum-rod, is gently pressed by a spiral spring. The marks thus traced on the concave surface indicate at once the direction and maximum extent of the pendulum's vibration."

He used deliberately the inverted pendulum because he knew this was an easy way to obtain long free periods with small dimensions. It is worth pointing out that Forbes was probably the first person to study, from a theoretical point of view and with the aim to calculate its response to ground motion, the effect of an earthquake on a pendulum. But Forbes assumed that an earthquake was just a lateral displacement of the ground with constant speed, and therefore, the results did not give him valuable information about the best settings for his instrument. As Dewey and Byerly (1969) pointed out, Forbes "desired a long period in order that the pendulum remained stationary as the Earth moved beneath it. He clearly wanted to measure ground displacement in an earthquake. However, [...] we can't be sure if Forbes knew that a long-period pendulum would function as a displacement meter for very-short-period oscillations of the ground."

Another interesting point about Forbes pendulum was his deliberate use of "magnification" through the increase of the distance of the writing pencil to the mass. It is the first time a reference to this issue is found. Anyway, the real performance of the instrument was somehow disappointing because it just recorded three of the sixty earthquakes felt. From our present knowledge, it is likely to assume that friction was too high.

During the same years, Robert Mallet (1810–1881), trying to calculate the velocity of seismic waves, used a totally different seismoscope consisting of a mercury container and an eleven-power magnifier. Mallet looked through at the image of a crosshair reflected at the surface of mercury (see Fig. 11). This is a sophisticated version of the "pan" used at the Exeter siege and specifically designed to measure a propagation time, there being no interest in any estimate of the amplitude of the ground motion. In the following years this instrument was to be improved with the addition of photographic recording transforming it into an almost real seismograph without using the inertia principle.



Historical Seismometer, Fig. 12 Seismoscope by Cavalleri. He studied the frequency contents of seismic waves by comparing the amplitudes of oscillation of the different pendulums with different periods (Adapted from Cavalleri 1858)

In Italy, Giovanni M. Cavalleri (1807–1874) designed and built several seismoscopes. Cavalleri constructed an instrument with six short pendulums of different periods, each of which traced the record of its motion in fine powder (Fig. 12). He used them to study the frequency contents of earthquakes and concluded (wrongly) that a range of frequencies from two to four cycles per second would be “sufficient to embrace every undulation occasioned by any earthquake.” Anyway, the Cavalleri studies showed that the ground motion during an earthquake is a composition of oscillatory motions with different frequencies, and it is the first attempt to make this kind of analysis.

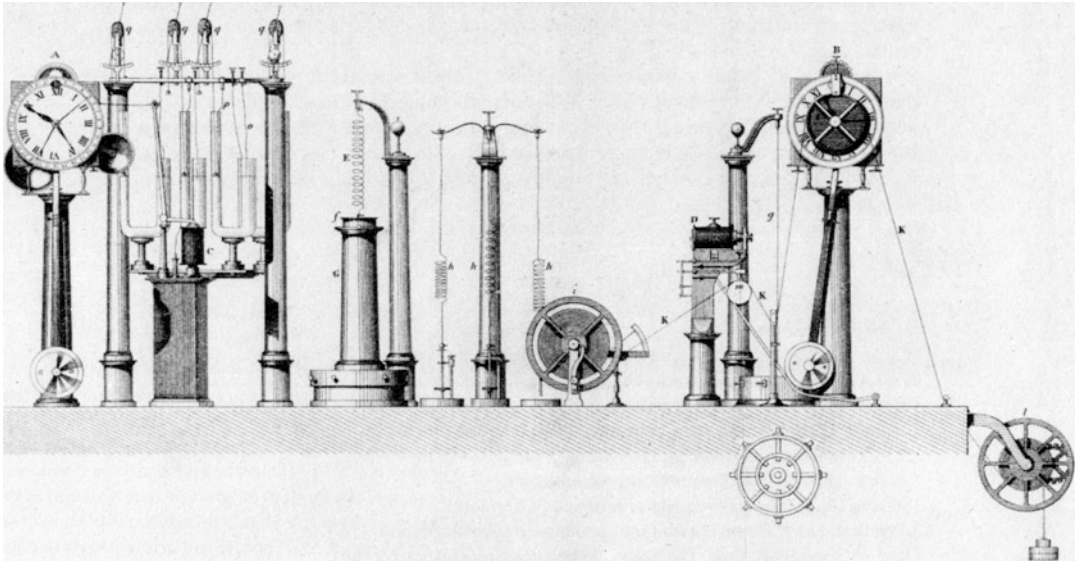
It is also around these years that the first devices to record vertical ground motion are made and that electrical circuits are used for the detection of ground motion. The most conspicuous of such instruments was the “Sismografo Elettromagnetico” designed by Luigi Palmieri (1807–1896) in 1856. This instrument, really remarkable at that time, was able to assert time, direction, intensity, and duration of an earthquake. It was not yet a seismograph, as its name may suggest, but a collection of seismoscopes, each intended to record particular parameters of an earthquake (Fig. 13). Palmieri’s instrument seems to have been an effective earthquake detector for local shocks. From the historical records, it looks like it was more sensitive to earthquakes than humans, being the first instrument to clearly attain this goal. Copies of this instrument reached Japan and were in use (and are still preserved) in Tokyo, at least from 1875 to 1885.

It is difficult to describe with words all the complexities of the amplitude indicators and

recorders of this instrument, not the least because the existing copies are different. Of specific interest, the seismoscope for detecting vertical motion consisted of a conical mass hanging on a spiral spring. The mass was suspended just over a basin of mercury. When a slight motion caused the tip of the cone to touch the mercury, an electric circuit was completed, which caused a clock to stop, indicating the time of the shock. Horizontal motion was detected with common pendulums, whose swinging completed the same circuit as that completed by the mass-spring seismoscope. In this case, a circular narrow channel filled with mercury was placed around the bottom of the oscillating mass. It was possible to reduce the distance between the mass at rest and the mercury to tens of millimeter.

These types of seismoscopes had a high sensitivity, and their design was copied in a large number of further starting/stopping devices for recording apparatus as it will be mentioned below.

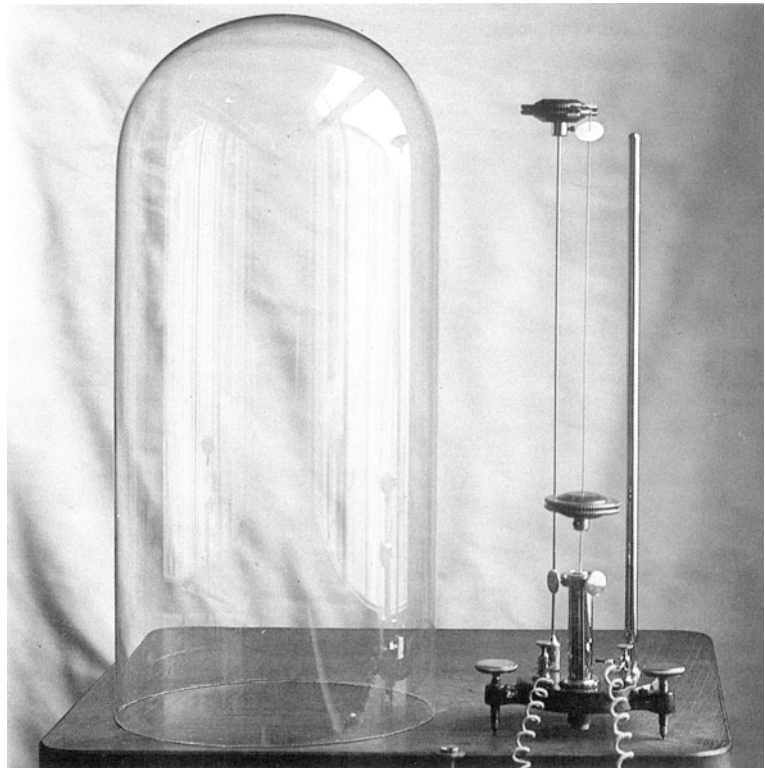
Already at the end of the century, in 1895, Giovanni Agamennone (1858–1949) introduced his double-action electric seismoscope (Fig. 14). This instrument can be considered as the final synthesis of the Italian know-how on seismoscopes, closing its evolution. It uses the oscillatory character of the earthquake ground motion in an elegant form. Two inverted pendulums of different periods are set aside each other. Their suspensions bars, connected to the poles of a battery, have an electric contact at the top. When at rest, the contact is open through a gap of few tens of a millimeter. Any slight motion of the ground causes the pendulums to oscillate (at different frequencies) and, thus, to complete the electric circuit. A bell rings and a clock, or



Historical Seismometer, Fig. 13 Palmieri seismograph. The spiral spring (*E*) was used to detect vertical motion. Horizontal motions along different axes were detected and recorded electrically by the movement of mercury in the U-tubes (on the *left* side). The closing of

an electric circuit activated electromagnets (*C* and *D*), which caused the clock to be stopped, thus recording the time of the arrival of the seismic waves, and the rotation of the drum (*i*) was started, with paper being unwound and marked at *m* (From Palmieri 1873)

Historical Seismometer, Fig. 14 Agamennone double-action electric seismoscope. Two inverted pendulums, with different free period, are set aside. The close proximity of the two upper extremities of the pendulums had the effect, in the case of a tremor, of making them come into contact with one another, completing an electric circuit (From Ferrari 1992)



any other device, may be started by the electric signal.

The First Seismograph

A fact ignored presently is that by the middle of the nineteenth century, seismoscopes became widely distributed around the world. Its presence and use other than in Italy and the UK is documented in Cincinnati, USA; on the occasion of the New Madrid earthquakes of 1811 and 1812; in Tabriz, Iran, for the October 4, 1856, Persian earthquake; in Manila, Philippines, for the June 20, 1857, event; or in several places in Germany for the June 24, 1877, Herzogenrath earthquake.

But seismoscopes do not give much information about the nature of earthquakes, and this is the most probable cause for the present low relevance of seismoscopic observations. The characteristics of ground shaking during earthquakes continued to be unknown as well as the characteristics of wave propagation at far distances. From our present knowledge about earthquake recording, it is possible to infer that sensors (seismometers) were becoming adequate for their task, but recorders were not. The lack of earthquake recordings (amplitude as a function of time, a seismogram) made it difficult to further study the earthquake phenomena.

But around 1875 Filippo Cecchi (1822–1877) built what it is now considered the first seismograph. However at the time it passed quite unnoticed. Unlike any of the instruments discussed up to this point, the Cecchi seismograph was expected to record the relative motion of a pendulum with respect to the Earth as a function of time. It was based in the inertia principle, and as described by Dewey and Byerly (1969), two common pendulums were used to record horizontal vibrations, each one oscillating in a unique direction and placed perpendicular to each other. Their free period was 1 Hz and the ground motion was magnified three times by a thread-and-pulley apparatus. A mass held on a spiral spring, as shown in Fig. 1, was used for the recording of the vertical motion. A device

designed to record rotary motions was also incorporated into the seismograph. It consisted of a crossbar with weights at both ends, much like a dumbbell, which was pivoted at its center of mass so as to rotate in a horizontal plane. Restoring force was applied to the dumbbell by springs so that it oscillated with a period of one second (Fig. 15). A more detailed description can be found in Ferrari (2006).

The instrument was not designed for continuous recording and Cecchi arranged a seismoscope to start a clock and to put into motion the recording surface, a smoked glass plate, at the time of an earthquake. The recording surface would move under the needles at a speed of one centimeter per second for 20 s. From the time on the clock, an observer arriving at the seismograph would determine how long before his arrival the earthquake had occurred.

It looks like several similar instruments were installed at Italian observatories and, certainly, one in Manila, Philippines, but the first confirmed record of this instrument was on February 23, 1887, at the occasion of the large earthquake that occurred near Nice, France (Fig. 16). This points to an insensitive instrument and explains the reason why such an achievement passed almost unnoticed. Our present interpretation is that the friction of the recording system was too high and, most likely, the starting seismoscope was not sensitive enough.

At the same time, other developments important for the future of seismometry were taking place; among them were experimental and theoretical studies about horizontal pendulums. Researchers needing to detect small gravity variations on laboratory and those working in geodesy studied the properties and behavior of horizontal pendulums (its specific characteristic was the way to hold the oscillating mass, as can be seen in Fig. 20-3) and he also developed the theory of such an instrument. A mirror was attached to the pendulum thereby magnifying its motion using a light beam reflected on it.

Another important theoretical development was done by Perry and Ayrton (1879), both British professors working at Japan. They published a paper entitled *On a neglected Principle that*

Historical Seismometer, Fig. 15 Cecchi

seismograph. Restoration of the original instrument presented at Genoa in 2000. In the front, the two common pendulums are seen. Among them, the elongated box is the place where recording plates were fixed (four plates for the three translation components and the rotational one). The rotational sensor is seen just behind to the right of the common pendulum (Photograph by the author)



may be employed in Earthquake Measurements. In this paper the authors analyzed the motion of damped and undamped mass-spring oscillators subjected to a periodic force. Now the importance of this contribution is clear, but it was almost ignored at the time of publication.

Developments in Japan

In its efforts to modernize the science and technology of the country, Japan hired, in addition to Perry and Ayrton, many foreign professors in the last quarter of the nineteenth century. Among them were John Milne (1849–1913), Alfred Ewing (1855–1935), and Thomas Gray (1850–1908). By living in one of the most earthquake-prone countries of the world, they were encouraged to study how to record earthquakes. The three, with solid engineering

background (Milne served as professor of mining and geology, Ewing of mechanical engineering, and Gray of telegraph Engineering), were in an optimum position to solve the problem of how to record ground motion; however, they were more handicapped by technical problems than by theoretical ones.

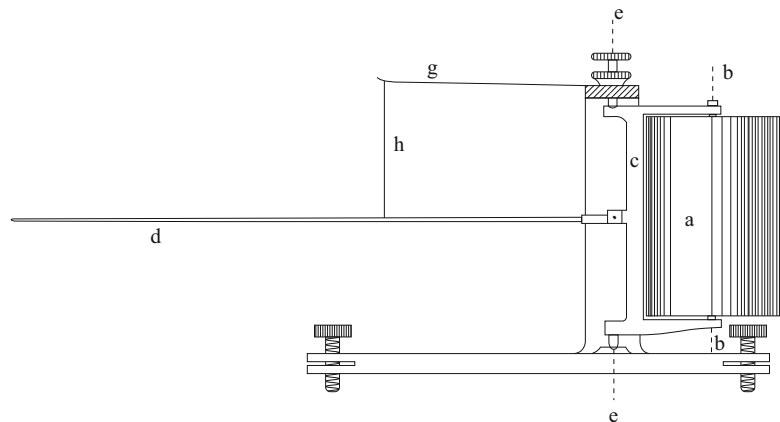
Few weeks after the earthquake on February 22, 1880, at Yokohama, and following a proposal by Milne, a seismological society was founded in Japan. Since the three were already involved in earthquake studies and have an engineering background, they gave priority to solving the technical problems of ground motion recording.

They studied different possible types of seismometers and recorders. It is known from their writings that they were looking for instruments with a free period larger than the periods of the signals they wanted to record, taking full advantage of the inertia of pendulums. The first



Historical Seismometer, Fig. 16 Cecchi seismograph record of the 1887 Ligurian earthquake. This record was obtained from one of the glass recording plates attached to one side of the elongated box described in Fig. 15 (From Ferrari 2006)

Historical Seismometer, Fig. 17 Drawing of the Ewing horizontal pendulum (From Ewing 1881)



instrument of their long production, built by Ewing, was a common pendulum of 5 s period (approximately 7 m long) because they thought this period was enough for their goal. A simple lever whose fulcrum was attached to the ground gave a six times magnification and recorded on a continuously revolving, circular smoked glass plate. The motion of the pendulum was resolved into two perpendicular components by the recording apparatus (two perpendicular levers were used). Not properly acknowledged up to now, it should be pointed that the great innovation of this instrument was the recording system, giving a long continuous record, and that it was applied successively to other models.

Almost in parallel Ewing tried recording earthquakes with a horizontal pendulum. His design was really successful, being the first to do so in the field of seismology. With small dimensions, the horizontal Ewing pendulum consisted (see Fig. 17) of a light rigid frame *c* pivoted at points *e* so as to swing like a garden gate around that axis of rotation. The cylindrical mass *a* is pivoted at the axis of percussion of the frame *b*. Again he used the revolving smoked glass and six times magnification. The weight of the oscillating mass was less than one kg.

Both of Ewing's instruments, the common-pendulum and horizontal-pendulum seismographs, recorded a small earthquake on November 3, 1880, giving the first lengthy seismic records of earthquake motion as a function of time. Finally, it was possible to have an idea of

the ground motion during an earthquake, and for the first time, seismologists could design their instruments with some knowledge of the phenomena the instruments were to record.

But an instrument for the vertical motion with a similar long period as those designed for the horizontal ones was still elusive. The solution came when Gray introduced a method of increasing the period of a mass-spring system by attaching the spiral spring to the short arm of a lever and attaching the mass to the long arm of the lever. This increased the period to few seconds by increasing the effective mass of the pendulum bob (see Figs. 20-8 and 9). Thus, even with many drawbacks with respect to our present knowledge (not a damped instrument, large dry frictions), a three-component instrument was now available.

Recording Earthquakes Globally: Teleseismic Recording

Another step was needed for the seismograph to be recognized as the key instrument not only for the study of earthquakes but also for the study of the Earth's interior. This was initiated by Ernst von Rebeur-Paschwitz (1861–1895). His main research area was geodesy and, specifically, the lateral deformations of the Earth's crust. To measure the deformations he built sensitive horizontal pendulums to measure slight changes in the direction of the vertical gravity. On April 17, 1889, two of these instruments, located in Potsdam and Wilhelmshaven, recorded simultaneously an anomalous signal that, from the newspapers information, he related to a large earthquake which had occurred in Japan.

He published a note about this observation in *Nature* that has become one of the milestones of observational seismology (Rebeur-Paschwitz 1889). The accompanying published seismic records have been reproduced many times (here also, see Fig. 18). The main result of his discovery was that, for the first time, the possibility to record distant earthquakes was confirmed. This possibility had already been proposed by others. The observation also confirmed that the whole

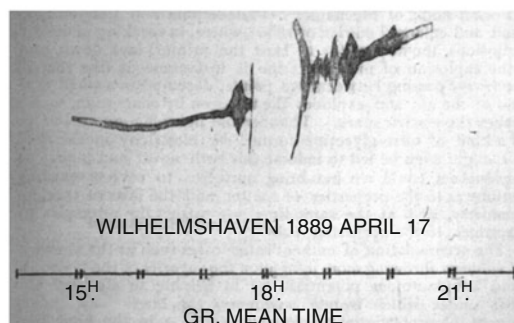
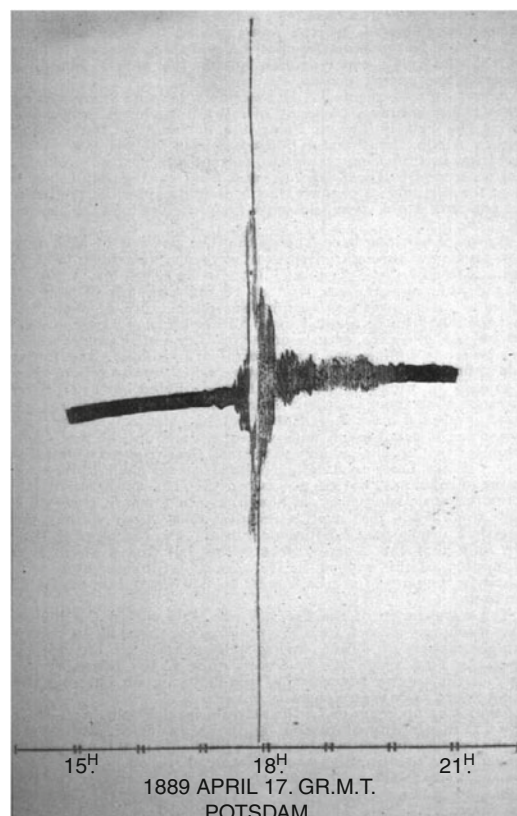
Earth is elastic and that therefore elastic wave theory can be used to analyze the behavior of seismic waves. Looking at the record, a curvature on the baseline can be seen. This curvature corresponds to the deformation caused by earth tides. The instrument was a tiltmeter but it recorded earthquakes, as gravimeters do.

The instrument consisted of a heavy base containing, inside, a horizontal pendulum sitting on a rigid frame, rotating around two bearings at the top and the bottom, each consisting of a point pressed into a socket (Figs. 19 and 20-2). He used a photographic recording. It was obtained through a mirror attached to the oscillating frame, which reflected light from a lamp through a cylindrical lens to a rotating drum which was covered with photographic paper. The drum turned 11 mm in an hour. As this system reduces friction significantly, the pendulum was only 10 cm long and carried a mass of only 42 g. It was usually used with free periods from 12 to 17 s and a static magnification of 100 (Dewey and Byerly 1969). A time reference was obtained with a second fixed light trace which wrote on the same photographic paper. Every hour, this second trace was eclipsed for 5 min. The magnification and the free period of Rebeur's pendulum were 20 times bigger than Ewing's pendulum. This instrument can be considered as the first approximation to what was needed to record teleseismic events.

The original instruments were, for many years, assumed lost, but recently original frames and other parts have been found by Fréchet and Rivera (2012).

The Seismograph Evolves

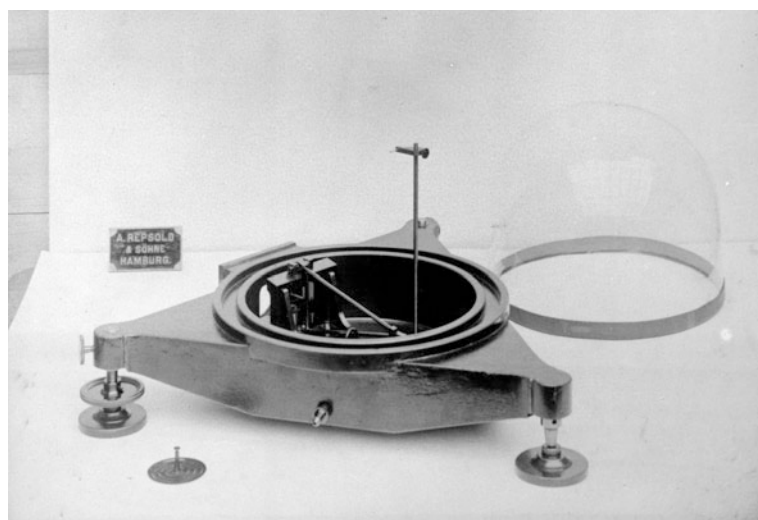
It is possible to identify, in the last quarter of the nineteenth century, three main groups or "schools" of seismograph designers and even to assign nationality to each one: the Italians, with their already long tradition of experimentation, in which Palmieri and Cecchi can be their best heralds; the Anglo-Japanese, with the Milne-Ewing-Gray trio, a rapidly progressing group characterized by its good equilibrium between the



H

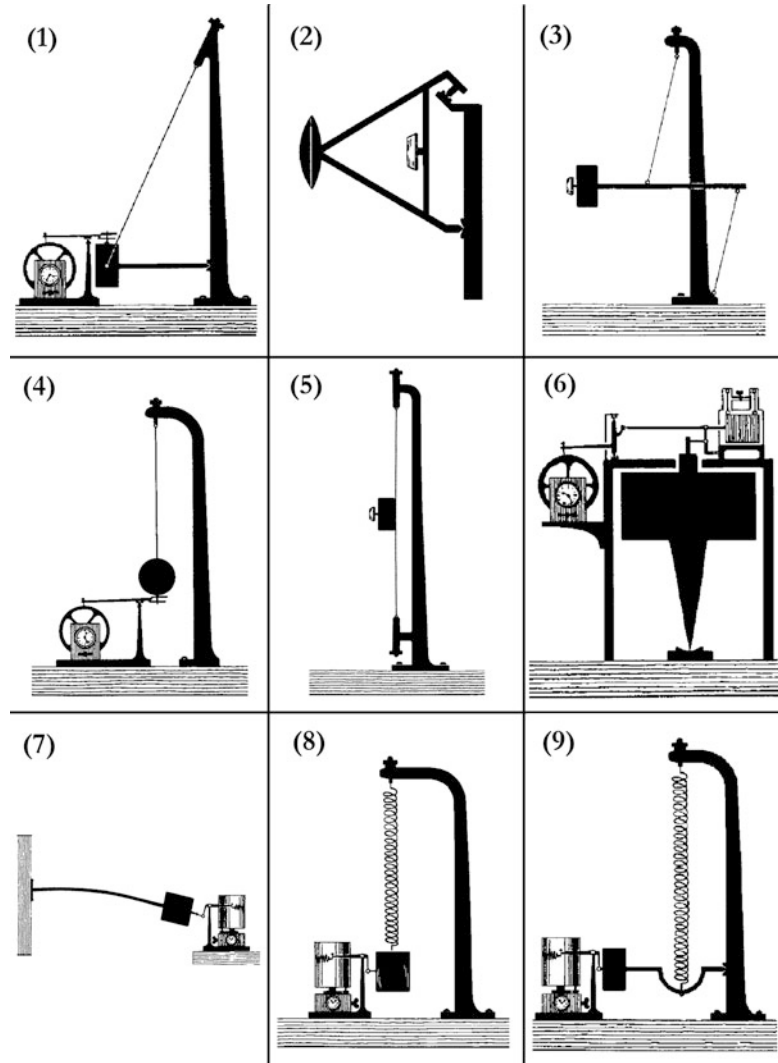
Historical Seismometer, Fig. 18 Rebeur-Paschwitz seismogram of the 1889 Japanese earthquake (From Rebeur-Paschwitz [1889](#))

Historical Seismometer, Fig. 19 Photography of the original Rebeur-Paschwitz pendulum. The top bearing is clearly seen. Also seen is a hole behind the bearing for the light beam used for photographic recording (From Fréchet and Rivera [2012](#))



Historical Seismometer,

Fig. 20 Different pendulum settings used for earthquake recording. It is possible to establish a seismograph classification according to the mechanical constitution of the transducer: (1) conical or bifilar pendulum, (2) rigid pendulum, (3) Zöllner pendulum, (4) vertical pendulum, (5) torsion pendulum, (6) inverted pendulum, (7) flexure pendulum, (8) vertical oscillator, (9) vertical rotational oscillator (From Batlló 2004)



theoretical and experimental approaches to the topic; and finally, with a really poor experience of earthquakes due to the low seismicity of its country, the Germans, Rebeur-Paschwitz at the front, with a geodetic background and a large experience in detecting very small ground motions at low frequencies.

The approach of each group to the seismic recording problem was almost independent and quite different. But when, within few years, the three arrived to obtain their first valuable records, the information among the different groups traveled fast, and the new technical procedures of one group were really soon implemented by the others.

Although other principles were tested, it is clear that all designers adopted the inertia principle for the design of the main instrument. All kinds of pendulums were tested (see Fig. 20).

Around that time (probably of Italian invention) the classical analog continuous seismic record consisting in a spiral line plotted on a paper placed on a drum was introduced. The lateral translation of a drum was obtained by just substituting its rotational axis with a screw. It was now possible to have a 24 h record on a unique sheet of paper with manageable dimensions.

In 1895, Giuseppe Vicentini (1860–1944) and Giulio Pacher (1867–1900) introduced a new

Historical Seismometer,

Fig. 21 Vicentini seismograph, with horizontal and vertical components as installed in the Alicante Observatory, Spain (Original photograph from the IGN archive)



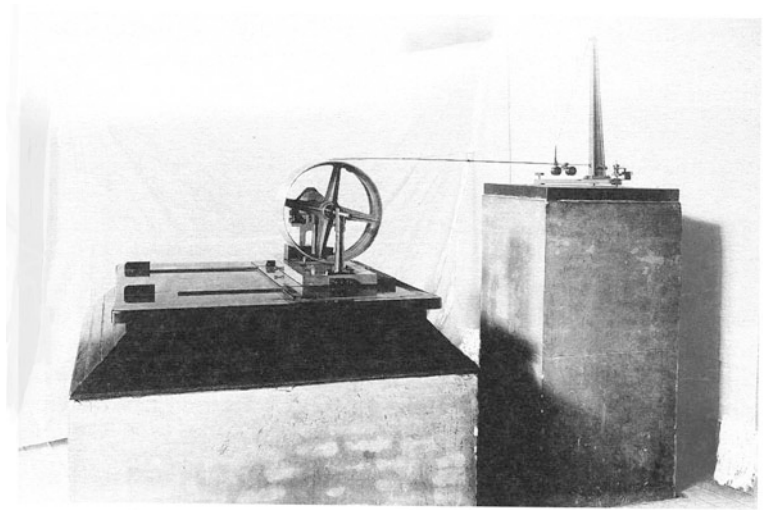
H

common pendulum specially designed to record near earthquakes. Now it was quite clear that signals of near earthquakes contained higher frequencies than those from teleseismic events. For this reason they designed an instrument optimized for near-earthquake recording. A common pendulum of 100 kg with a free period 2.3 s recorded on smoked paper with a speed of 15 mm/min. The motion of the pendulum was amplified and resolved into perpendicular components by levers. Static magnification was set to 80. The recording was on smoked paper, and following a trend common to Italian instruments at that time, they used large masses to overcome the frictions of the amplifying and inscription mechanisms. The good results obtained made them add a vertical component. It was quite original in design and consisted of a flat spring clamped to a wall from one side and with the oscillating mass (50 kg) at the other end (Fig. 20-7). The free period was 1 s approx. and magnification was 130. The resulting instrument was known as Vicentini seismograph and it was used at many places for near-earthquake recording up to the middle of the twentieth century when it was finally superseded by electromagnetic short-period seismographs (see Fig. 21). The three components were written side by side on the same continuous record.

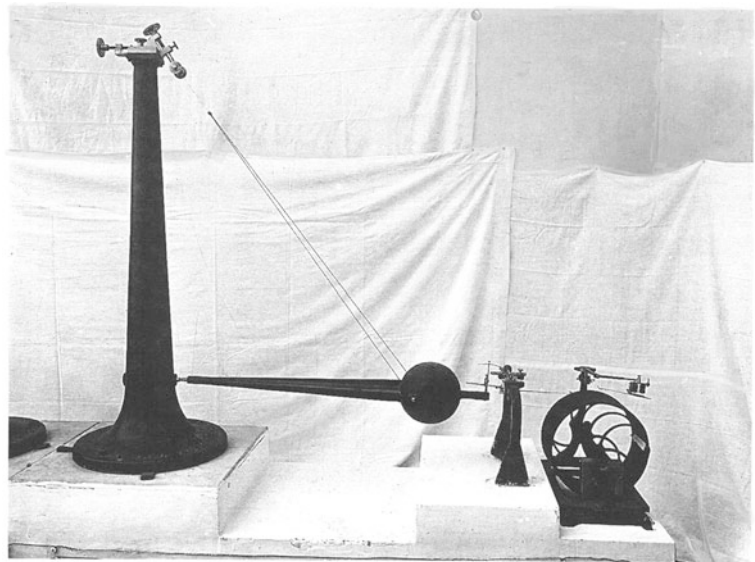
Milne, on coming back to England, centered his interest on global seismicity. For this objective he used the horizontal pendulum he designed in Japan in 1894 (Fig. 22). In this instrument he adopted the photographic record and, thus, he combined it with a light oscillating mass (less than half kilograms). The free period was set near 16–18 s and magnification was 6–9. He adopted also continuous recording, but did not adopt the laterally translating drum (he did in later versions of the instrument). For this reason the early records were long rolls of photographic paper. Paper speed was just 1 mm/min and one roll of 10 m lasted 1 week. The instrument, thus, was not convenient for immediate analysis of seismograms. Even not being an instrument of outstanding performance, it is really important in the history of seismology because, due to a proposal of Milne, it was selected by the British Association for the Advancement of Science to be deployed around the world (and, effectively, it was) in a first attempt of making a worldwide network.

By the end of the nineteenth century, the British professors in Japan returned to Europe and the Japanese took full responsibility for continuing the outstanding seismological research in Japan. Fusakichi Omori (1868–1923), a collaborator of Milne, continued the design of horizontal

Historical Seismometer,
Fig. 22 Milne
 seismograph installed at the
 Toledo Observatory, Spain.
 This instrument has a drum
 recording instead of a roll,
 which was common in the
 early instruments (From
 Batllo 2004)



Historical Seismometer,
Fig. 23 Bosch-Omori
 seismograph single
 component (the most
 common set was a pair
 recording in perpendicular
 horizontal directions)
 installed at the Toledo
 Observatory, Spain (From
 Batllo 2004)



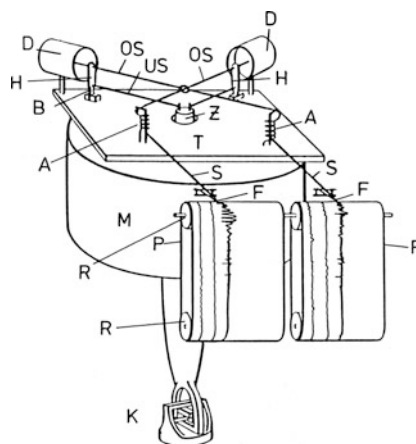
pendulums. He turned to continuous recording on smoked paper and, as needed in this case, heavier masses (10 or more kg). In its most common form, Omori's pendulum consisted of a mass on a rod, pivoting about a socket, with the mass held up by a flexible wire. With a free period of about 20 s, static magnifications about 10–15, and paper speed of 5/10 mm/min, they were rough instruments of easy use and maintenance. These qualities make them popular and were copied

elsewhere. Still in the middle of the twentieth century, they were common in secondary seismic stations. The Bosch factory, in Strasbourg, commercialized a popular version under the name "Bosch-Omori" pendulums (Fig. 23). The worldwide distribution, relatively long period, and adequate paper speed make their records still valuable for the study of large earthquakes which occurred in the early years of the twentieth century (Batlló et al. 2004, 2008).

Wiechert Inverted Pendulum

Emile Wiechert (1861–1928) began his experiments with earthquake recording in 1898, at the just created Institute for Geophysics at the University of Göttingen, with a pendulum with photographic recording and viscous damping. He introduced the damping to lessen the effects of the pendulum's free oscillations and he was the first to introduce such a device in a seismograph. But after a scientific round trip to Italian observatories, he developed already in 1900 a totally new instrument, his worldwide famous astatic pendulum. The new instrument combined characteristics of the instruments already in use in Germany and Italy and some “premiers” to be attributed to Wiechert himself. After some years of experimentation and improvements, Wiechert published it in 1904.

The first “premier” of this instrument was (recovering the idea of Forbes) the use of an inverted pendulum stabilized by springs and free to oscillate in any direction horizontally (Wiechert 1904). In this case the restoring springs were applied to the top of the inertial mass. A joint (of cardan type) at the base of the pendulum permitted motion in any horizontal direction. The 1904 pendulum (Fig. 24) had an oscillating mass M of 1,000 kg. Using such a large mass made it possible to overcome dry frictions. The free period of the pendulum was around 10 s. The relative motion of the mass with respect to the ground was resolved into perpendicular components, magnified 200 times by a mechanical lever system (H , OS , and S), and written on a smoked paper for continuous record P . Paper speed was between 10 and 15 mm/min. Time marks were put on the paper every minute by lifting the writing index off of the paper. The viscous damping was obtained by connecting the motion of the pendulum mass to pistons which fit closely inside cylinders D attached to the seismometer stand. Resistance of the air to the piston motion provided damping for the pendulum; this resistance was controlled by a valve which had the effect of regulating the amount of air space between the piston and the cylinder. The lever system and recording are indebted to Italian technology



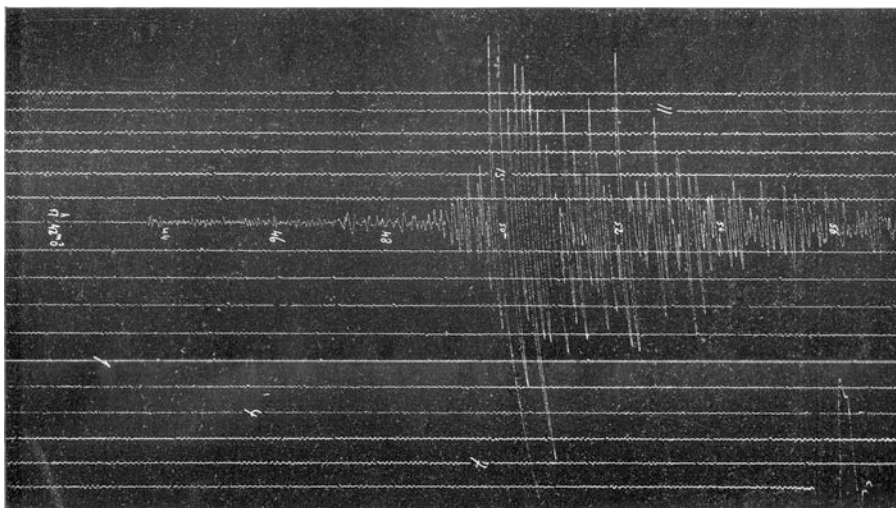
Historical Seismometer, Fig. 24 General sketch of the horizontal Wiechert astatic seismograph. See description in the text (From Ferrari 1992)

while the free mass period and high (at that time) amplification reflected German experience and the air damping device was another premier due to Wiechert. The damping device was the most important innovation because its behavior approaches the expected mathematically modeled viscous damping.

Last but not the least, in parallel with the instrument development, Wiechert studied carefully the theory of the motion of pendulums submitted to ground oscillatory motion and as a result he developed the theory of the response of mechanical seismographs to ground motion. The presentation of this problem in actual textbooks is very similar to the description by Wiechert.

Finally, in 1905 a vertical component complemented the astatic horizontal pendulum. Using as much as possible the sound design of the horizontal instrument, the overall resulting instrument was much more conventional, the large mass being held with springs, and it was impossible to obtain a free period larger than 5 s.

The Wiechert astatic instrument became a reference instrument for many years. Several manufacturers commercialized different versions of the instrument, and by the time of the second World War, more than 100 seismic stations around the world were using these kinds of instruments. Several are still in operation used as



Historical Seismometer, Fig. 25 Record of the earthquake that occurred on April 23, 1909, near Lisbon, Portugal, as recorded in the N-S component of the Wiechert

seismograph installed at Munich, Germany, at 2,000 km from the epicenter. P and S waves arrival and surface waves are clearly identified (From Batlló et al. 2008)

reference instruments. Figure 25 shows an example of a typical record obtained with a Wiechert instrument. P and S waves are clearly identified and the record is not different for our present seismograms.

The Electromagnetic Seismograph

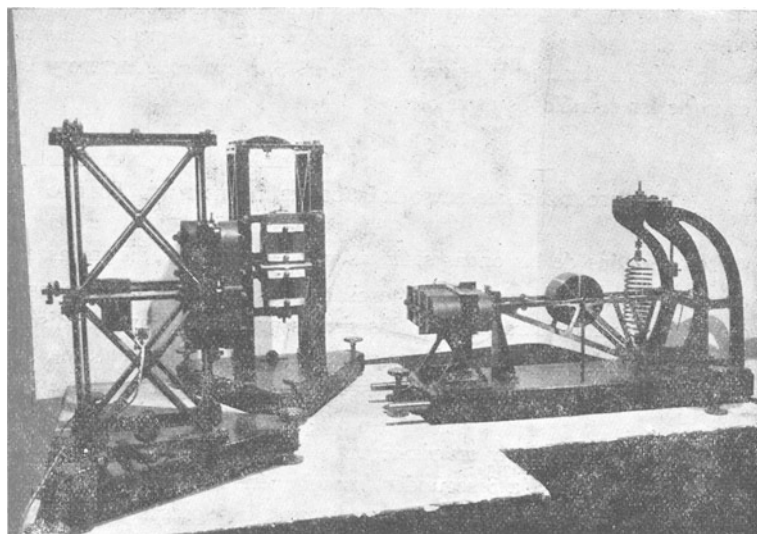
Boris B. Galitzin (1862–1916), in charge of the Pulkovo seismic station, was, as Wiechert, deeply interested in the improvement of seismic recording. In 1905 he came with a totally new instrument, the electromagnetic seismograph (Fig. 26). The new seismograph used, instead of mechanical transducers amplifying the ground motion with levers, an electromagnetic transducer. To do so coils attached to the oscillating pendulum were moving through a magnetic field created by permanent magnets. Ground motion was thus converted to an electric current and optically recorded, with the help of a galvanometer, in a photographic paper introducing, thus, a second oscillating element. Viscous damping was obtained magnetically, with a sheet of copper attached to the oscillating mass moving in the magnetic field created by a second pair of magnets. The horizontal pendulums (one

for each component) used the Zöllner suspension (Fig. 20-3) and the vertical instrument used the suspension system introduced by Ewing in Japan (Fig. 20-9). Their masses were few kilograms. Its design allowed to separate the “sensing” device from the “recording” device. From then it is called the seismometer-galvanometer system.

The new instrument was the fruit of careful experimentation but also, as in the case of the Wiechert instrument, a result of theoretical study. The horizontal pendulum’s free period was set around 24 s for the horizontal and 12 s for the vertical components. The galvanometer’s free period was set to the same frequency as the seismometers and both the seismometer and galvanometer had critical damping. It was then possible to get an analytical solution to the equation of motion for the whole coupled system seismometer-galvanometer. The peak magnification was more than one thousand times and a typical paper speed was 30 mm/min. Another advantage was that by using parallel and serial coupled resistances, both the damping and the gain of the instrument could be regulated easily.

The new instrument was much more sensitive than any previous instrument and superior to them in almost all aspects, except complexity of recording and price. At that time photographic

Historical Seismometer,
Fig. 26 Group of Galitzin
 electromagnetic
 seismometers installed at
 the Toledo Observatory,
 Spain. The two
 perpendicular horizontal
 components are placed to
 the left and the vertical
 component to the *right*. The
 size of the magnets is
 notorious (From Batlló
 2004)



H

recording was not affordable to any seismic station. While a standard mechanical seismograph (if properly damped) has a flat response to displacement for frequencies above the natural frequency, the transfer function of the seismometer-galvanometer system is no longer flat (see Fig. 6) making it a bit more difficult to analyze the seismograms. The available magnets suffered from demagnetization, and thus, the overall magnification of the system changed. For these reasons, their use expanded slowly.

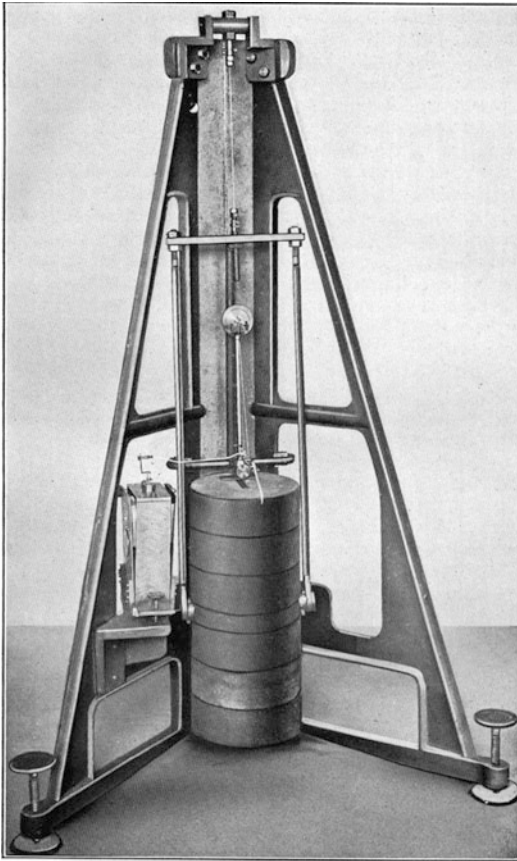
From Teleseisms to Near Earthquakes

Wiechert astatic seismograph and Galitzin's electromagnetic instruments gave a good solution to the recording of teleseisms. They were for many years the standard instruments of the most important seismic stations around the world, complemented at secondary stations with a variety of instruments, many of them horizontal pendulums of the Omori type, as the Mainka pendulums (Fig. 27), with nominal characteristics similar to the Wiechert instruments but of easier maintenance. But overall the quality of the mechanical parts of these cheaper instruments was not as good as those of the original Wiechert (and, as pointed by McComb (1936),

performance of damping oil systems as viscous damping was not satisfactory) lowering the quality of the obtained seismograms.

Nevertheless, the studies on the structure of the Earth progressed fast. But recording of small earthquakes in the near field was still deficient. To get a high amplification at frequencies near 1 Hz and higher was not an easy task. Some solutions were proposed. To improve the mechanical transducers (the main problem was to overcome the friction), large masses were used. Italian designers recognized this problem already at the end of the nineteenth century. Wiechert himself moved this way constructing a short-period Wiechert-type instrument with 17 t. Later, in 1922, the Quervain-Piccard mechanical seismograph was built in Switzerland. A mass of 21 t was "floating" on springs, and separate recording in three directions was obtained by levers. The static magnification was 1,700 and the free period was 3 s for the horizontal components and 1 s for the vertical component.

At the same time a totally different solution was adopted in California. The result was the famous Wood-Anderson instrument (Fig. 28). It takes advantage of a torsion suspension (Fig. 20-5) for the pendulum (Anderson and Wood 1925). A small mass is enough to obtain



Historical Seismometer, Fig. 27 Front view of the Mainka horizontal pendulum. The recording system has been removed to show the oscillating mass and the structure holding it. Mass is held at the top of the structure and at a second point (not seen), just behind the axis at its center, as in the Bosch-Omori instrument (From Batllo 2004)

high gain when, to avoid friction, it is combined with magnetic damping and photographic recording. In this case, the most cumbersome element is the photographic recording. With a mass of less than 1 g, it has a free period of 0.8 s and magnifications larger than 2,000 (the nominal magnification of the commercial instruments was 2,800 even though further studies by Urhammer and Collins 1990 showed it is smaller). Unfortunately, a vertical component was never produced even though its construction does not offer any further technical complication.

Benioff Reluctance Seismometer

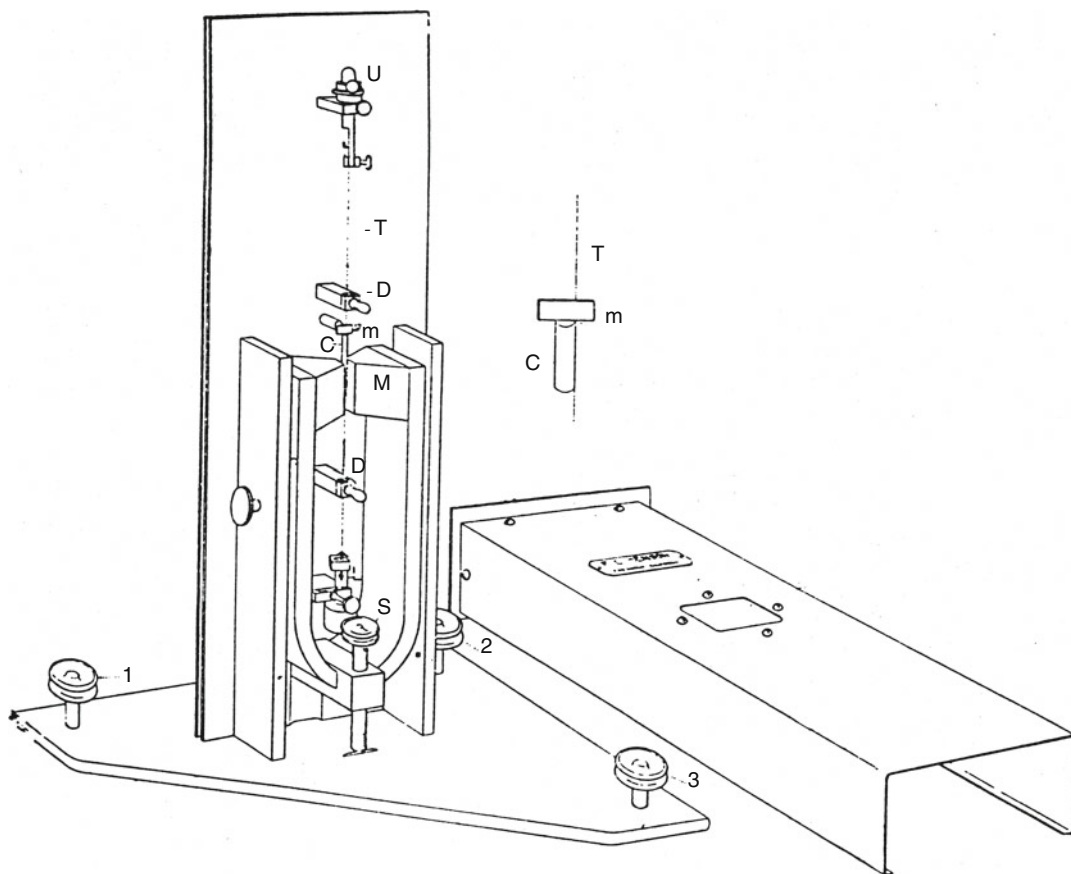
Further research of the ways to increase seismometer sensitivity to be able to detect smaller signals got Hugo Benioff (1899–1968) to introduce, in 1930, the variable reluctance seismometer (Benioff 1955). Its principle is similar to the electromagnetic sensor, but instead of a moving coil in a magnetic field, the transducer is of the same type of those used in telephone receivers. A moving magnet supplies flux across two or more air gaps to an armature around which is wound a coil of wire. Changing air gaps, caused by the motion of the magnet due to ground shaking, generates the current in the coils. The device generates much stronger signals than the classical electromagnetic sensor. Using photographic recording it then becomes easy to get amplifications of 100,000 at 1 Hz. Horizontal and vertical seismometers were produced and they became later, almost unmodified, the short-period equipment of the WWSSN (Fig. 29).

The sensitive transducer permitted to use some recording principles previously abandoned because of the lack of adequate technology. This was the case of the strainmeter developed also by Benioff in the 1930s. The strainmeter dismissed the cherished principle of pendulum inertia and measures the variation of distance between two points at the passing of seismic waves. The new variable reluctance transducer could be used to properly measure this distance and thus to use the strainmeter as seismometer.

Strong Motion Measurement

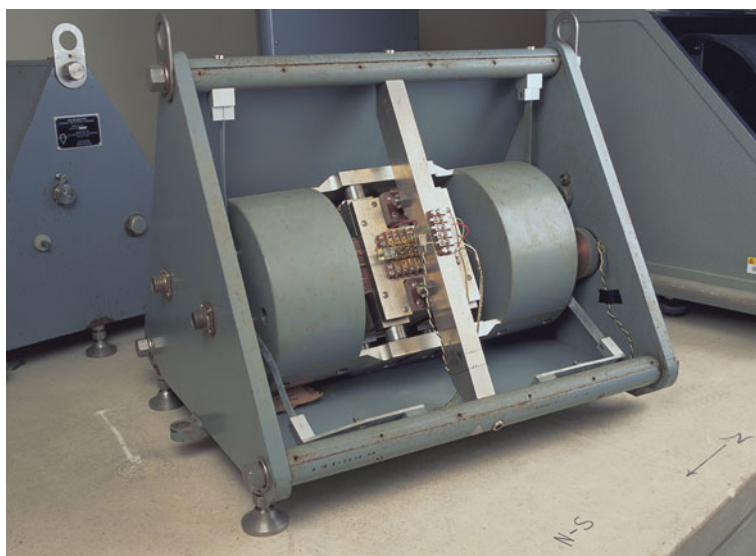
Earthquake engineering and related fields also use instruments that record ground motion. As it is seen in the previous paragraphs, shortly after the first World War, technological leadership, at least in the seismic recording, transferred to the USA. This is also the case for strong motion recording.

Since early times, there was an interest in measuring vibrations of buildings and other constructions, and attempts were made to use seismic instruments for that purpose. It is possible to give



Historical Seismometer, Fig. 28 General sketch of the Wood-Anderson seismometer. A zoom of the oscillating mass is shown in the right part (elements T , C , and m) (From Anderson and Wood 1925)

Historical Seismometer, Fig. 29 WWSSN Benioff variable reluctance seismometer for the horizontal component. The transducer is placed in the central part of the instrument between the two mass disks (Original photograph in the IGN archive)



many examples. Ewing, already in 1888 in the UK, used his double pendulum to measure the vibration of the new Tay Bridge during the passing of railway trains. Guido Alfani (1976–1940), in 1911, with a portable seismograph specially designed for use in buildings, called “trepidometer,” studied the vibration of the famous Pisa inclined tower when bells were ringing. Contemporaries Albin Belar (1864–1939) in Ljubljana and Manuel Sánchez Navarro (1867–1941) in Granada, among others, built also “tromometers” to use them in engineering works. But all of these instruments lacked enough sensibility in the needed range of frequencies to properly analyze the obtained signals (even the frequency contents of those signals were not well defined at that time). Even more, most of them were not damped instruments. As it happened in many other occasions related to the seismic recording, proper ideas already existed many years before its proper realization. It is worth to point out that already in 1915, Galitzin studied the possibility of using the piezoelectric effect to build accelerometers (Fremd 1997), but lacking the electronic technology needed to amplify signals, the device was not feasible at that time.

In the early 1930s several damaging earthquakes in the USA started a discussion about seismic safety, and the engineers realized that real progress in earthquake engineering (an engineering branch not yet defined at that time) passes through detailed knowledge of accelerations in buildings and structures, but there was no instrument capable of doing these measurements. However, theoretical and technological studies were properly developed at that time to describe the problem. So just 2 years passed since the starting of the Strong Motion Program, in 1931, up to the recording of the first accelerograms for the Long Beach earthquake of May 10, 1933.

The appointed committee for the Strong Motion Program was very fast in determining that it was important to record acceleration instead of displacement and also in the interesting range of frequencies and the expected amplitudes ($f < 10$ Hz and amplitude up to 0.5 g) to be

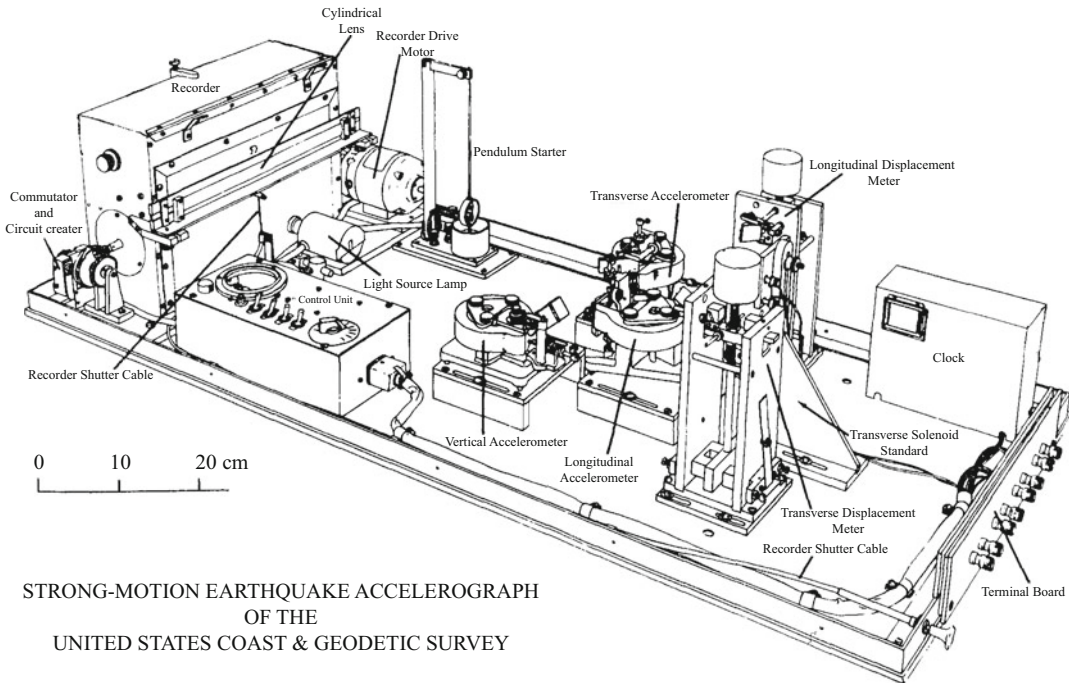
recorded. All this knowledge about expected values was already accumulated from the previously available observations as well as how to use mechanical sensors to record ground acceleration.

Once specifications were defined, a review of available technology was made. The choice was to use a Wood-Anderson torsion pendulum-type seismometer (Trifunac 2009). In this case a vertical component was also built. Problems with the proper set of the delicate suspension (in this case with four wires, as explained by Trifunac 2009) required a change for a more robust pivoted and spring-stabilized system already in 1933. Analog recording on photographic paper forced the choice for a triggered instrument. The triggering seismoscope was still of the Palmieri type. The resulting accelerometer (Fig. 30) had three components with a free period of 10 Hz (to properly record accelerations at lower frequencies). The amplitude of the trace was 4 cm for an acceleration of 0.2 g (remember that it was expected for this instrument to have a flat response to acceleration in the range of interest). The recording speed was 1 cm/s and time marks were introduced each 0.5 s.

The instrument worked quite well but was quite expensive for particulars (an indication of that is that accelerometers were not commercialized before 1963). Thus, and mainly for engineering purposes, well-calibrated seismoscopes continued to be on the first line of earthquake engineering up to the advent of electronic systems and digital recording, in the mid-1970s (Fig. 31).

Summary

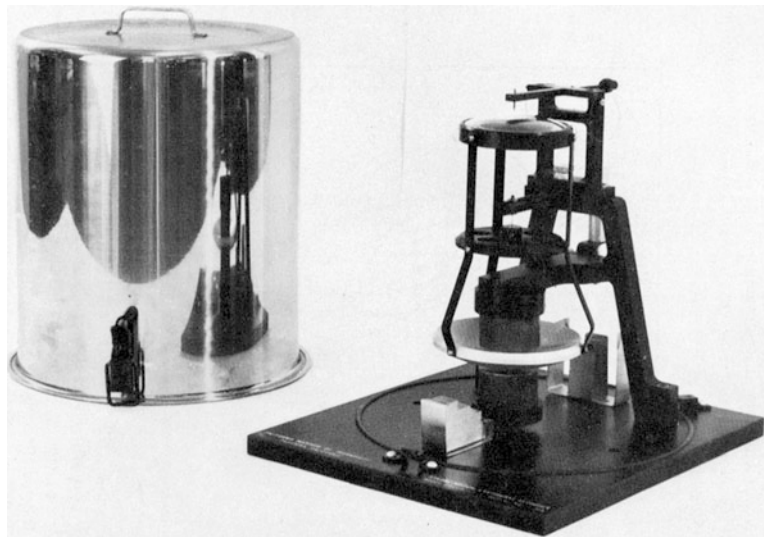
The development of the early stages of seismometry and the evolution of the early seismographs and of seismic recording have been presented. Evolution of the instruments, from the first known devices of the second century A.D. up to the devices of the nineteenth century, was slow and with many failures, the main issue being the lack of knowledge of the real ground motion during earthquakes and, thus, of a clear goal.



H

Historical Seismometer, Fig. 30 USCGS accelerograph. The three accelerometers are placed at the center of the instrument. The starter pendulum is clearly seen at the top left of the accelerometers (From Hudson 1963)

Historical Seismometer, Fig. 31 USCGS seismoscope for strong motion recording. This instrument was designed in the 1950s and continued in use up to the time of the digital accelerograph (From Hudson 1963)



This lack of knowledge was also due to the lack of proper recorders able to record continuously. This was a common problem to all sciences and it was not properly solved until the second part of the nineteenth century. Once the first records were obtained, seismometry progressed much

faster. The tools for the theoretical analysis of seismic recording (mainly Newtonian mechanics) were already available. Electromagnetic theory, useful to develop electromagnetic transducers, was developed at the same time and was soon incorporated to seismometry.

Humans experience near earthquakes (felt earthquakes) and this led to the early attempts of seismic recording. But the large development of seismic networks at the beginning of the twentieth century was mainly made with the purpose of recording teleseisms. Near-field recording was not properly addressed up to the interwar period of the twentieth century. At first specific instruments for the recording of near earthquakes were developed and later, by the demand of engineers, accelerometers were designed and deployed. In the middle of the twentieth century, with the development of the WWSSN and similar instruments (like the Kirnos seismographs), seismic instruments gave results similar to our present instruments. Nevertheless, sophisticated seismoscopes, cheap and easy to handle, were generally deployed in seismic areas up to the advent of new active seismic sensors and digital recording which solved many of the problems of the older instrumentation.

Cross-References

- [Passive Seismometers](#)
- [Recording Seismic Signals](#)
- [Seismic Instrument Response, Correction for](#)
- [Sensors, Calibration of](#)

References

- Anderson JA, Wood HO (1925) Description and theory of the torsion seismometer. *Bull Seism Soc Am* 15:1–72
- Batló J (2004) Catálogo Inventario de sismógrafos antiguos españoles. Instituto Geográfico Nacional, Madrid, 413
- Batló J, Clemente C, Pérez-Blanco F, Vidal F (2004) Reconstruction of a Bosch-Omori seismograph. *Nuncius* 19:701–711
- Batló J, Stich D, Macià R (2008) Quantitative analysis of early seismograph recording. In: Fréchet J, Meghraoui M, Stucchi M (eds) *Historical seismology: interdisciplinary studies of past and recent earthquakes*. Springer, Berlin, pp 379–396
- Benioff H (1955) Earthquake seismographs and associated instruments. *Adv Geophys* 2:219–275
- Cavalleri GM (1858) Di un nuovo sismometro collocato nel Collegio di Monza. *Atti dell'Istituto Lombardo di Scienze, Lettere ed Arti* 1:34–44
- Dewey J, Byerly P (1969) The early history of seismometry (to 1900). *Bull Seism Soc Am* 59:183–227
- Ewing JA (1881) On a new seismograph. *Proc R Soc* 31:440–446
- Ferrari G (ed) (1990) *Gli strumenti sismici storici: Italia e contesto europeo*. SGA Storia- Geofisica-Ambiente, Bologna, p 198
- Ferrari G (ed) (1992) Two hundred years of seismic instruments in Italy 1731–1940. SGA Storia- Geofisica-Ambiente, Bologna, p 156
- Ferrari G (2006) Note on the historical rotation seismographs. In: Teisseyre R, Takeo M, Majewski E (eds) *Earthquake source asymmetry, structural media and rotation effects*. Springer, Berlin, pp 367–376
- Forbes JD (1844) On the theory and construction of a seismometer, or instrument for measuring earthquake shocks, and other concussions. *Trans R Soc Edinb* 15:219–228
- Fréchet F, Rivera L (2012) Horizontal pendulum development and the legacy of Ernst von Rebeur-Paschwitz. *J Seismol* 16:315–343
- Fremd VM (1997) Historical seismometers as predecessors of modern instruments. *Cahiers du Centre Européen de Géodynamique et de Séismologie* 13:91–93
- Havskov J, Alguacil G (2010) *Instrumentation in earthquake seismology*. Springer, Dordrecht, xii + 358 pp
- Hönes R (1893) *Erdbebenkunde: die Erscheinungen und Ursachen der Erdbeben, die Methoden ihrer Beobachtung*. Veit, Leipzig, vi + 452 p
- Howell FH (1990) An Introduction to Seismological Research: History and Development (Chapter 4). Cambridge University Press, Cambridge, pp 57–72.
- Hsiao KH, Yan HS (2009) The review of reconstruction designs of Zhang Heng's seismoscope. *J Jpn Assoc Earthq Eng* 9(4):1–10
- Hudson DE (1963) The measurement of ground motion of destructive earthquakes. *Bull Seism Soc Am* 53:419–437
- McComb HE (1936) Selection, installation and operation of seismographs. Special publication no. 206. Coast and Geodetic Survey, Washington, iv + 43 p
- Oliver J, Murphy L (1971) WWNSS: seismology's global network of observing stations. *Science* 174:254–261
- Palmieri L (1873) *The eruption of Vesuvius in 1872*. Asher, London, p 148
- Perry J, Ayrton WE (1879) On a neglected principle that may be employed in earthquake measurement. *Phil Mag Ser* 5(8):30–50
- Rebeur-Paschwitz E (1889) The earthquake of Tokyo. *Nature* 40:294–295
- Stein S, Wyssession M (2003) *An introduction to seismology, earthquakes and earth structure*. Blackwell Publishing, Oxford, xi + 498 p
- Trifunac MD (2009) 75th anniversary of strong motion observation – a historical review. *Soil Dyn Earthq Eng* 29:591–606

- Urhammer RA, Collins ER (1990) Synthesis of Wood-Anderson seismograms from broadband digital records. *Bull Seism Soc Am* 80:702–716
- Wiechert E (1904) Ein astatisches Pendel höher Empfindlichkeit zur mechanischen Registrierung von Erdbeben. *Beitr Geophys* 6:435–450
- Wielandt E (2009) Seismic sensors and their calibration. In: Bormann P (ed) *New manual of seismological observatory practice (NMSOP)*. Deutsches GeoForschungsZentrum GFZ, Potsdam, p 46

Hyperspectral Data in Urban Areas

Giorgio Antonino Licciardi
GIPSA-Lab – INP Grenoble, Grenoble Institute of Technology, Saint Martin d'Hères, France

Synonyms

Change detection; Data fusion; Hyperspectral images; Hyperspectral sensors; Image processing; Material identification; Remote sensing

Introduction

The scope of this entry is to introduce hyperspectral remote sensing data and its applications in urban environment.

In the last decades imaging spectroscopy (Goetz et al. 1985), commonly referred to as hyperspectral remote sensing, has become a widely used method for identification and quantification of surface materials in different kinds of environments, such as urban, rural, and geological areas (Heiden et al. 2012).

A hyperspectral sensor is able to record a high number of spectral bands, corresponding to narrow contiguous wavelength intervals. This characteristic permits the discrimination of different materials based on their unique spectral characteristics, called spectral signature. In general, two materials can be distinguished by broad and narrow spectral reflectance features determined by

the chemical composition of the materials. However, some materials may present strong spectral similarities except for a narrow part of the spectrum. From this point of view, hyperspectral sensors have the ability to detect more materials than multispectral sensors.

In the literature, several studies have analyzed the spectral characteristics of material detected by hyperspectral instruments, usually by means of laboratory and field spectroscopic investigations (Ben-Dor 2001).

Hyperspectral remote sensing demonstrated its effectiveness in several fields of application, such as precision agriculture, security surveillance, physics and astronomy, environmental analysis, and mineralogy, as well as in hazard monitoring and management.

In urban environment, hyperspectral remote sensing can be extremely useful, in particular for the management of pre-/post-seismic events. In this optic, the ability of hyperspectral sensors to discriminate several materials can be extremely useful in the characterization of the building type, by identifying the material used for the construction. Another important use of hyperspectral imaging is the ability to produce accurate geological maps of the soil.

Hyperspectral images provide a great amount of spectral information, however, when compared with multispectral images, the main limitation of this kind of data resides in their dimensionality. Since hyperspectral data cubes are large, their dimension may have a negative impact in terms of computational processing load and also on the storage. The dimensionality problem limited also a proper development of satellite-based hyperspectral systems, since transmission and storage of acquired data become extremely difficult. From this point of view, reducing the dimensionality of hyperspectral images is one of the most important processes for the exploitation of the information provided by hyperspectral.

In the following paragraphs, different aspects of hyperspectral imaging in urban area, with a particular attention to those related to seismic events, will be addressed.

Hyperspectral Image Overview

$$O = I \otimes R$$

Hyperspectral Sensors

Both scientists and common people are becoming increasingly concerned with environmental phenomena such as the photosynthetic conditions of the vegetation, wide deforestation and fires, desertification, sea pollution, and hazard monitoring, together with the general health of the Earth. The monitoring of these events and the understanding of the impact which they could have on the fragile biophysical mechanisms are becoming more and more important than in the past. For this reason, sensors like Medium-Spectral Resolution Imaging Spectrometer (MERIS), Moderate Resolution Imaging Spectroradiometer (MODIS), and Advanced Very High Resolution Radiometer (AVHRR) have been designed and placed in orbit. These measurements are performed using several spectral bands (up to 36 for MODIS) located into the visible and the infrared range in order to collect a noteworthy dataset for every kind of global investigation (land use, ocean color, snow cover, sea ice observation, etc.). A following step has been the allocation of many contiguous and narrow bands (more than one hundred) available for the measurement. This technological evolution led to the hyperspectral imagery, which has demonstrated very high performance in several cases of material identification and urban mapping, including sub-pixel classification. Managing such dissimilar type of data is not a simple task and requires the adoption of information extraction techniques that are appropriate for each specific sensor data. From this point of view, a hyperspectral sensor can be considered as an extremely precise and sensitive optical sensor. However, due to its sensitivity, a hyperspectral sensor, compared to a multispectral one, is more prone to be affected by imprecise measurements. The precision of a measurement is determined by the instrument response R . The transformation from the input physical quantity I to the measurement O is described mathematically by the convolution:

The instrument response depends on several factors, such as optical imperfections, image motion, the nonzero spatial area of each detector, as well as the filtering produced by electronic elements. Except for the image motion, all the elements influencing the instrument response can be attributed to the acquisition mode. In particular the type of scanning system and the spectral selection mode can have a strong impact on the precision of the measurements.

In general, an optical sensor can be divided into three different scanning systems for acquiring the image: whiskbroom, pushbroom, and staring imagers. While whiskbroom scanners have a simple optical system and can acquire image in a wide swath width, if compared to the pushbroom scanners, their mechanical system is extremely complex expensive and prone to damages. For these reasons pushbroom scanners are considered as the standard for high-resolution imaging spectrometers. Another characteristic that can influence the hyperspectral measurement is the spectral selection mode. In particular, the spectral selection modes used in hyperspectral sensors can be divided in three main groups: dispersion elements, where the incoming electromagnetic radiation is separated into different angles by means of a prism or grating elements; filter-based elements, based on optical bandpass filters (tunable filters, discrete filters, and linear wedge filters); and Fourier-transform spectrometers (FTS): a Fourier-transform spectrometer is an adaption of the Michelson interferometer where a collimated beam from a light source is divided into two by a beam splitter and sent to two mirrors. These mirrors reflect the beams back along the same paths to the beam splitter, where they interfere. The signal recorded at the output depends on the wavelength of the light and the optical path difference between the beam splitter and each of the two mirrors. If the optical path difference between the two beams is zero or a multiple of the wavelength of the light, then the output will be bright; otherwise if the optical path difference is an odd multiple of

half the wavelength of the light, then the output will be dark.

Hyperspectral Sensors

Most of the developed hyperspectral sensors are designed for airborne carriers. One of the first airborne hyperspectral sensors was the AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) developed by National Aeronautics and Space Administration (NASA). AVIRIS contains 242 different detectors, each with a bandwidth of approximately 10 nm, allowing it to cover the entire range between 380 nm and 2,500 nm. Similar sensors were the Compact Airborne Spectrographic Imager (CASI), able to acquire images in 288 bands in the 380 nm–1,050 nm spectral range, and the Hyperspectral Mapper (HyMap) providing 128 bands in the wavelength region of 450–2,500 nm. The Airborne Hyperspectral Scanner (AHS) and the Multispectral Infrared and Visible Imaging Spectrometer (MIVIS) extended the spectral range up to thermal infrared, allowing a wider range of applications.

The development of hyperspectral technology for the space satellites remains difficult and very expensive in terms of payload design, maintenance, and calibration. However, these difficulties have not deterred the space agencies from finding interesting missions carrying on board hyperspectral payloads. This is the case of Hyperion developed by NASA, CHRIS (Compact High Resolution Imaging Spectrometer) PROBA-1 (Project for On-Board Autonomy) developed by a European consortium founded by European Space Agency (ESA), the upcoming PRISMA developed by ASI (Agenzia Spaziale Italiana), and EnMAP (Environmental Mapping and Analysis Program) developed by Deutschen Zentrums für Luft- und Raumfahrt (DLR).

Hyperion instrument, mounted on board of the EO-1 (Earth Observation 1) satellite, provides a high-resolution hyperspectral imager capable of resolving 220 spectral bands (from 400 to 2,500 nm) with a 30 m spatial resolution. The instrument covers a 7.5 km by 100 km land area per image and provides detailed spectral mapping

across all 220 channels with high radiometric accuracy.

CHRIS is a high-resolution hyperspectral sensor installed on board the PROBA satellite. Distinctive feature of CHRIS is its ability to observe the same area under five different angles of view in the VIS (visible)/NIR (near infra-red) bands. Another important feature of CHRIS is the ability to be reprogrammable in terms of the number of spectral bands and spectral resolution (up to 150 spectral bands with a spectral resolution of 1.25 nm). This characteristic allows the acquisition of 62 narrow and quasi-contiguous spectral bands with the spatial resolution of 34–40 m or only 18 spectral bands with an enhanced spatial resolution of 18 m.

PRISMA (PRecursore IperSpettrale della Missione Applicativa) is a new earth observation integrating a hyperspectral sensor with a panchromatic camera. The advantage of using both sensors is to integrate the classical geometric feature recognition to the capability offered by the hyperspectral sensor to identify the chemical/physical feature present in the scene. The primary applications will be the environmental monitoring, geological and agricultural mapping, atmosphere monitoring, and homeland security.

The main goal of EnMAP project is to investigate a wide range of ecosystem parameters encompassing agriculture, forestry, soil and geological environments, coastal zones, and inland waters. Thanks to the chosen sun-synchronous orbit, each point on earth can be investigated and revisited within 4 days.

Hyperspectral Data Processing

Since hyperspectral sensors are able to measure the radiance, one of the most important processing steps in hyperspectral image processing is the conversion to reflectance. Usually, the techniques for retrieving surface reflectance can be roughly divided into two major categories, relative reflectance retrieval techniques and absolute reflectance retrieval techniques. Relative reflectance retrieval is based on the knowledge of two or more target reflectances in the image. Knowing the reflectance of

a material in the image permits to define by regression a linear equation between each wavelength and the measured radiance. However, this approach is also based on the assumption that the atmosphere is uniform, resulting in significant artifacts in strong atmospheric bands if ranges of elevations are present in the scene.

Absolute approaches, on the other hand, permit the retrieval of surface reflectance based on the physical principles, in which measured radiance is typically compared to radiance generated by an atmospheric radiative transfer model, such as the MODerate resolution atmospheric TRANsmission (MODTRAN) model. Usually, the relative approach can be useful in processing airborne data, where the influence of the atmosphere is not relevant and can be considered uniform. On the other hand, absolute approaches are dominant in spaceborne images.

Another common preprocessing step is georectification, used to relate the acquired data to the ground.

In order to make maps and relate them to ground reference data, it is critical that aircraft-related and ground-related distortions are removed. The most common approach is to use a “rubber-sheet” stretching approach and numerous tie points between a base map and measured image. Recent improvements in onboard navigation information and recent software development have made it possible to georectify images to within a few pixels (~ 10 m for a 4 m GIFOV [Ground instantaneous field of view]) in a near-automated fashion.

Aside from the common processing techniques to obtain images more interpretable, there are other techniques that are more related to the dimension of hyperspectral data. If we consider the consistency of the hyperspectral data, we can easily understand the importance of finding a method which can transform the original data cube into one with reduced dimensionality and maintain, at the same time, as much information content as possible. In particular, dataset composed of hundreds of narrowband channels, besides storage and transmission, may cause problems in terms of complexity of processing and inversion phases. Therefore,

dimensionality reduction may become a key parameter to obtain a good performance.

These techniques are known under the general name of feature reduction. Besides enabling an easier storage and management of the data, feature-reduction procedures can be crucial for the implementation of optimum inversion algorithms.

Many methods have been developed to tackle the issue of high dimensionality of hyperspectral data (Serpico and Bruzzone 1994). In summary, we may say that feature-reduction methods can be divided into two classes: “feature-selection” algorithms (which suitably select a suboptimal subset of the original set of features while discarding the remaining ones) and “feature extraction” by data transformation which projects the original data space onto a lower-dimensional feature subspace that preserves most of the information, such as nonlinear principal component analysis (NLPCA; Licciardi and Del Frate 2011).

The first analysis suggests that feature selection is a more simple and direct approach compared to feature extraction and that the resulting reduced set of features is easier to interpret. Nevertheless, extraction methods can be expected to be more effective in representing the information content in lower dimensionality domain.

Feature-selection techniques can be generally considered as a combination of both a search algorithm and a criterion function. The solution to feature-selection problem is offered by the search algorithm, which generates a subset of features and compares them on the basis of the criterion function. From a computational viewpoint, an exhaustive search for the optimal solution becomes intractable even for moderate values of features (Siedlecki and Sklansky 1988). Despite these apparent difficulties, many feature-selection approaches have been developed (Serpico and Bruzzone 1994). The sequential forward selection (SFS) and the sequential backward selection (SBS) techniques are the simplest suboptimal search strategies: they can identify the best feature subset achievable by adding (to an empty set in SFS) or removing (from the complete set from SBS) one feature at a time, until the desired number of features is achieved.

The sequential forward floating selection (SFFS) and the sequential backward floating selection (SBFS) methods improve the standard SFS and SBS techniques by dynamically changing the number of features included (SFFS) or removed (SBFS) at each step.

A feature-extraction technique aims at reducing the data dimensionality by mapping the feature space onto a new lower-dimensional space. Both supervised and unsupervised methods have been developed. Unsupervised feature-extraction methods do not require any prior knowledge or training data, even though these are not directly aimed at optimizing the accuracy in a given classification task. The class comprises the “principal component analysis” (PCA, Fukunaga 1990), where a set of uncorrelated transformed features is generated; the “independent component analysis” (ICA) (Jutten and Herault 1991), a computational method for separating a multivariate signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian source signal; and the “maximum noise fraction” (MNF, Green et al. 1988), where an operator calculates a set of transformed features according to a signal-to-noise ratio optimization criterion.

Challenges in Hyperspectral Data Processing

The analysis of hyperspectral images is not an easy task. Due to the complexity of a hyperspectral sensor, hyperspectral images can be affected by several problems in terms of data uniformity. Any nonuniformity in the system generates degrading artifacts. Since pushbroom scanners mainly characterize hyperspectral sensors, in this entry we will focus on the main problems that afflict this kind of sensors.

In pushbroom imaging spectroscopy, the image is generated in the spectral and spatial dimension simultaneously. This leads to two main kinds of problems:

- Spectral nonuniformity: is the nonuniformity of the spectral response within a sensor’s spectral band and can be imaged on a detector row. This nonuniformity is typically represented by the position and shape of the spectral response

function. The related artifacts of spectral misregistration are denoted as “smile” or “frown.”

- Spatial nonuniformity: is the nonuniformity of the spatial response within an acquired spectrum and is usually imaged on a detector column. This nonuniformity is represented by the position and shape of the spatial response function in both the along-track and across-track dimensions of a spatial pixel. The related artifacts in the across-track dimension are denoted as “keystone.”

As for the spatial nonuniformity, the influence of the “keystone” effect results in a black pixel in the image that can be easily replaced by meaning the surrounding pixels with a 3×3 moving window. On the other hand, the removal of the “smile” effect is not an easy task. The consequence of this effect is that the central wavelength of a band varies with spatial position across the width of the image in a smoothly curving pattern. Very often the peak of the smooth curve tends to be in the middle of the image and gives it a shape of “smile” or “frown.” The effect of the smile is not obvious in the individual bands. Therefore, an indicator is needed to make evident whether or not a given image suffers from smile effect. A way to check for the smile effect is to look at the band difference images around atmospheric absorption (760 nm). In fact, the region of red-near infrared transition has high information content of vegetation spectra. This region is generally called “red edge” (670–780 nm) and identifies the red-edge position (REP). The smile effect is acute due to sharp absorption at 760 nm, which is within the red-edge region, and for this reason atmospheric correction of smiled data will be incorrect.

The methodologies developed so far can only reduce the intensity of smile effect but cannot remove it entirely, because during its life a detector element can change its response; therefore, the knowledge and the correction of this phenomenon became fundamental in the analysis of hyperspectral images.

Another problem that affects pushbroom scanners is the so-called “striping” noise. The striping

effect consists in strong variations in the average value in the column for every band of the image. Theoretically the elements of the detector array are identical. Thus, the detectors in a row should be well calibrated to provide the same output along the spatial domain, if illuminated by the same source. In the real world, a residual nonuniformity of the electro-optical response generates different transfer functions. Thermal fluctuations can cause additive small variations in the alignment of the detector, producing a non-predictable noise not constant during the time. Only the systematic knowledge of the gain function permits an accurate compensation of these fluctuations. The correction of the striping can be carried out by means of filtering approaches. This type of destriping involves two filters on the original image creating two new images, a low-pass filtered image and a high-pass filtered image, and combining the results. If both a low-pass and a high-pass filter of the same size are run on an image and then added together, the result will be the original image. The goal in choosing sizes for the two filters is to create an output for each without the unwanted striping. If the filter sizes are not the same, the result will not be exactly the same as the original. When the filter outputs are added together, the striping will have been removed or subdued. If the sizes of the two filters differ too much, artifacts may be introduced into the result.

From this brief overview of the principal preprocessing techniques, it is possible to understand the amount of efforts that are necessary in order to extract useful and reliable information from hyperspectral data.

Building Vulnerability Assessment

One of the aspects of pre-seismic event management is related to the assessment of the vulnerability of buildings in urban environment. Many features that can influence the seismic performance of a building can be extrapolated from the outside and then be detected by remote sensing instruments. From this point of view, urban areas can be considered as an ensemble of

different types of surfaces characterized by a continuous spatial change. This results in small urban structures mostly dominated by artificial (man-made) materials. The identification of these materials requires an accurate analysis of the spectral characteristics of these materials. Information about the material used in the roof of the buildings permits to determinate the construction technique of the building and consequently its level of vulnerability. The identification of the roof spectral characteristics can be useful also for the derivation of the type of urban structure where the building is located. For instance, the historical development of European countries leads the urban structure type to be correlated with a certain time period and construction style. Thus the knowledge of the urban structure type may give information about the type and age of the buildings. In particular, the necessary information for the evaluation of pre-seismic events scenarios can be resumed in the detection of the building type (i.e., the material) and in the detection of the identification of the contextual information, such as the position of each building in relation to the surrounding ones. Analysis of the geological information of the soil can be useful but become extremely difficult if performed by using remote sensing techniques in urban environment, since the soil tends to be covered by man-made structures.

Material Identification

Surface materials can be detected based on their spectral characteristics. In the literature there exist several works on the extraction of information about the material of the roofs by means of multispectral images. However, these techniques permit to identify a limited number of elements, due to the low spectral resolution of multispectral imager. For instance, multispectral images are not sufficient to differentiate between roofs and other surfaces having similar spectral characteristics, such as asphalt and bitumen used to seal rooftops. In this context hyperspectral imaging can be an extremely useful instrument for an effective estimation of the vulnerability, identifying more information about the material used in the construction of the investigated buildings.

In general, several tools have been developed to match reference spectra with those measured by an imaging spectrometer. The most common approach is based on the use of standard supervised classification techniques, where known spectra are used to determine the statistical properties of each class based on spectral characteristics. Examples of supervised classification approaches applied to hyperspectral data are described in McKeown et al. (1999) and Roessner et al. (2001), where the maximum likelihood classifier (MLC) was applied to map urban land cover. Other techniques are based on the use of support vector machines (SVM) (Melgani and Bruzzone 2004) and neural networks (NNs) (Licciardi et al. 2009, 2012). Other approaches have been designed explicitly for the analysis of imaging spectrometry data, such as the Spectral Angle Mapper (SAM; Kruse et al. 1993).

In order to extract as much information content as possible from hyperspectral images in urban environment, specific methods for automated information extraction have been developed (Roessner et al. 2001). In all these methods, the comprehensive mapping of the different surfaces present on urban environment can be obtained through a classification technique (Roessner et al. 2001).

For example, Clark et al. (2001), in their environmental studies on the World Trade Center area after the September 11, 2001, attack, successfully applied hyperspectral remote sensing to map material in the surrounding area. In this study, the authors used high-spatial-resolution AVIRIS data acquired over the World Trade Center to map thermal sources, asbestiform minerals, and dust and debris from the collapse, producing maps of environmental contaminants that are difficult to produce cost-effectively in any other way. McKeown et al. (1999) used data acquired by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) to classify urban and natural materials. The authors then fused the classification results with a 3D model of the buildings. Ben-Dor (2001) developed an urban spectral library using a combination of an existing spectral library developed by Price (1995) and CASI data acquired over the city of

Tel Aviv, Israel. Data acquired by the Digital Airborne Imaging Spectrometer (DAIS) have been used by Roessner et al. (2001) to obtain a map of urban materials in the city of Dresden, Germany. In this study, a maximum likelihood classifier has been used to derive a first map of pure spectral features and then used these features to unmix the other spectra.

However, since the spectral variations characterizing a certain material are caused by several factors, such as the age and deterioration of the material, or the color, and also by the illumination angles, several aspects need to be considered. In particular, aside from the spectral heterogeneity of different materials, it is important to take into account the high intra-class variability of spectra corresponding to the same material. Moreover, in urban environment there are many objects having a size smaller than the pixel. This results in mixed pixels, presenting spectra that are a combination of two or more materials that are imaged in the IFOV (instantaneous field of view) of the sensor. Moreover, the modeling of the spectral mixture becomes more complex as the pixel size increases. From this point of view, airborne hyperspectral sensors, characterized by a high spectral and/or spatial resolution, permit to obtain reliable quantitative measurements of specific absorption features of urban materials.

In terms of spectral unmixing, a number of approaches in the literature are based on the assumption that the detected spectra are linear combination of pure spectra. In particular, in linear spectral unmixing the mixed spectrum is modeled as the sum of pure spectra, each weighted by the fraction of the material within the field of view. Roberts et al. (1998) have developed a general approach for spectral unmixing in which the number and types of end-members are allowed to vary per pixel. In Roessner et al. (2001), the authors used an MLC to detect the pure spectra in the image and then use them as end-members to produce the abundance maps. An approach based on nonlinear unmixing has been proposed in Licciardi and Del Frate (2011), where a neural network (NN) has been used to quantify the abundances of previously detected end-members.

Fusion of Hyperspectral and Geometric Information

While the spectral information is extremely useful to identify different materials, it does not provide any information on the use of these materials. This means that spectral information is not sufficient to differentiate, for instance, a roof from other surfaces characterized by the same material. For this reason, additional information is required for the univocal identification of the buildings. One possible solution is to derive the shape characteristics of the buildings present in the image. Usually, when seen from above, most of the buildings present a rectangular shape. However, some exceptions exist in terms of regularity of the shape. For instance, industrial or commercial buildings present differences in terms of size and covering materials, such as bitumen or metal. Residential buildings are, on the other hand, of small size and often covered with tiles or concrete. Roads are characterized by bitumen but present oblong shapes and sometimes interrupted by the tree canopies. Parking lots are spectrally characterized by bitumen but, differently from the roads, are characterized by rectangular shapes. This requires the development of algorithms, which can handle this variety in the shape detection process. The basic approach for the extraction of objects is segmentation, consisting in labeling each pixel in the image and grouping the adjacent ones according to their label. In the literature a variety of methods for segmentation of hyperspectral image exist. Segl and Kaufmann (2001) proposed a method for the identification of buildings based on an iterative segmentation followed by a classification. In Tarabalka et al. (2010), the watershed segmentation algorithm has been applied to hyperspectral images, while in Priego et al. (2013), a segmentation technique based on the application of cellular automata is used in order to produce homogeneous regions. In both cases an support vector machine (SVM) approach is applied to label the obtained objects and classify them. The use of geodesic opening and closing operations of different sizes in order to build a morphological profile and a neural network approach for the classification of features has

been introduced in Pesaresi and Benediktsson (2001) and applied successfully to hyperspectral data in Licciardi et al. (2012).

An emerging technique in the last decades is based on the fusion of hyperspectral and Light Detection and Ranging (LiDAR) data, thanks to the increased availability of these data taken from the same area. In particular, in urban environment, the introduction of high-resolution digital surface models derived from LiDAR can effectively improve the classification and consequently the identification of different materials and surfaces.

In Heiden et al. (2012), the height information obtained from LiDAR data has been introduced into orthorectification in order to improve the geometric accuracy of urban objects in the higher-resolution hyperspectral image data. This permitted to eliminate the influence of the facades of the buildings in large field of view (FOV) of such sensors and to improve the accuracy of surface material mapping.

The fusion of LiDAR and hyperspectral data through a physical model has been used in Zhang et al. (2013) to eliminate the direct illumination component in the hyperspectral radiance data and consequently permitted to detect also materials under shadows.

Change Detection

The accurate analysis of the spectral characteristics of the materials present in urban environment covers an important role in the framework of building vulnerability assessment. However, a complete analysis of the vulnerability of buildings in urban environment and more in general environmental monitoring requires also an accurate study of the changes occurred in the time domain. From this point of view, the monitoring of urban areas using multi-temporal images has received increasingly attention in the last decades. In particular, remote sensing technology is able to provide a large amount of images on the same area over a long period of time. Land-cover change detection is a useful instrument for the description of changes in urban environments. Change detection by remote sensing has been widely used in many applications such as

land-use/land-cover monitoring, urban development, and ecosystem and disaster monitoring. Conventional change detection methods generally depend on a difference value in each individual band, such as image differencing or the image ratio. However, most of these methods are based on multispectral images and do not take into account the physical meaning of the continuous spectral signatures. In hyperspectral images, the change can be associated to change of a spectral signature from one material to another material. For this reason most of the techniques presented in the literature are based on the difference in the spectral signatures of different materials. However, the detection of changes in hyperspectral images is not a trivial task, mainly because of several factors influencing the final result. Problems that can complicate the detection of changes in urban environment are caused by differences in the illumination level, such as the presence of shadows in two different images. Geometrical differences in terms of parallax and error of misregistration may result in the detection of false changes, as well as differences in atmospheric content. Moreover, since different cities may present different patterns and characteristics, it is difficult to define a general technique that can be applied indifferently to any kind of urban pattern. In the literature the existing change detection techniques can be divided into three main groups.

Post-classification methods are easy to implement and provide “from-to” change maps but may present limits in their validity because of misclassification errors in one of the classification maps. Other methods are based on image transformation techniques such as PCA or multivariate alteration detection (MAD). These techniques project the original hyperspectral data into a feature space in order to label the changes (Nielsen et al. 1998). A third kind of hyperspectral change detection methods is based on the detection of anomalies present in the images. In particular, anomaly detection techniques are able to distinguishing unusual targets from a typical background, and this assumption can be extended to change monitoring (Eismann et al. 2008).

Summary

In this entry we provided a brief description of the different aspects of the use of hyperspectral data in urban environment. In the introduction we presented a brief overview of imaging spectrometry in urban areas. The discussion continued with a description of the characteristics of the hyperspectral sensors and the problems related to the processing of hyperspectral images. We illustrated then the different techniques for the assessment of building vulnerability through the use of hyperspectral data.

The management of pre-seismic event in urban environment can be improved with the use of remote sensing data. In recent years remote sensing becomes more and more a popular instrument in the monitoring of urban areas since many features influencing the seismic performance of a building can be extrapolated from remote sensing instruments. Hyperspectral sensors, able to provide images with unprecedented spectral detail, can have an important role in the assessment of the vulnerability of buildings in urban environment. From this point of view, urban environment is extremely challenging because of the need to discriminate as much materials as possible. Hyperspectral sensors, able to provide images with unprecedented spectral detail, can have an important role in the assessment of the vulnerability of buildings in urban environment. However, this task is often complicated by several factors, such as the different reflectance of the material, different lighting conditions, as well as different atmospheric characteristics. To solve these problems, several techniques have been developed in order to obtain efficient mapping of urban materials with the use of hyperspectral images.

Cross-References

- [Earthquake Damage Assessment from VHR Data: Case Studies](#)
- [Features Extraction from Satellite Data](#)
- [Urban Change Monitoring: Multi-temporal SAR Images](#)

References

- Ben-Dor E (2001) Imaging spectrometry for urban applications. In: Van der Meer FD, De Jong SM (eds) *Imaging spectrometry. Basic principles and prospective applications*. Kluwer, Dordrecht
- Clark RN, Green RO, Swayze GA, Meeker G, Sutley S, Hoefen TM, Livo KE, Plumlee G, Pavri B, Sarture C, Wilson S, Hageman P, Lamothe P, Vance JS, Boardman J, Brownfield I, Gent C, Morath LC, Taggart J, Theodorakos PM, Adams M (2001) Environmental studies of the world trade center area after the September 11, 2001 attack. U.S. Geological Survey, Open File Report OFR-01-0429
- Eismann MT, Meola J, Hardie RC (2008) Hyperspectral change detection in the presence of diurnal and seasonal variations. *IEEE Trans Geosci Remote Sens* 46:237–249
- Fukunaga K (1990) *Introduction to statistical pattern recognition*, 2nd edn. Academic, San Diego
- Goetz AFH, Vane G, Solomon JE, Barrett NR (1985) Imaging spectrometry for earth remote sensing. *Science* 228(4704):1147–1153
- Green AA, Berman M, Switzer P, Craig MD (1988) A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Trans Geosci Remote Sens* 26(1):65–74
- Heiden U, Heldens W, Roessner S, Segl K, Esch T, Mueller A (2012) Urban structure type characterization using hyperspectral remote sensing and height information. *Landscape Urban Plann* 105(4):361–375
- Jutten C, Herault J (1991) Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture. *Signal Processing* 1991(24):1–10
- Kruse FA, Lefkoff AB, Boradman JB, Heidebrecht KG, Shapiro AT, Barloon PJ, Goetz AFH (1993) The spectral image processing system (SIPS)- iterative visualization and analysis of imaging spectrometer data. *Remote Sens Environ* 44:145–163
- Licciardi G, Del Frate F (2011) Pixel unmixing in hyperspectral data by means of neural networks. *IEEE Trans Geosci Remote Sens* 49(11):4163–4172
- Licciardi G, Pacifici F, Tuia D, Prasad S, West T, Giacco F, Inglada J, Christophe E, Chanussot J, Gamba P (2009) Decision fusion for the classification of hyperspectral data: outcome of the 2008 GRS-S data fusion contest. *IEEE Trans Geosci Remote Sens* 47(11):3857–3865
- Licciardi G, Marpu PR, Chanussot J, Benediktsson JA (2012) Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles. *IEEE Geosci Remote Sens Lett* 9(3):447–451
- McKeown DM Jr, Cochran SD, Forel SJ, Mcglone JC, Shufelt JA, Yokum DA (1999) Fusion of HYDICE hyperspectral data with panchromatic imagery for cartographic feature extraction. *IEEE Trans Geosci Remote Sens* 37(3):1261–1277
- Melgani F, Bruzzone L (2004) Classification of hyperspectral remote sensing images with support vector machines. *Geosci Remote Sens IEEE Trans* 42(8):1778–1790
- Nielsen AA, Conradsen K, Simpson JJ (1998) Multivariate alteration detection (MAD) and MAF postprocessing in multispectral, bitemporal image data: new approaches to change detection studies. *Remote Sens Environ* 64:1–19
- Pesaresi M, Benediktsson JA (2001) A new approach for the morphological segmentation of high-resolution satellite imagery. *Geosci Remote Sens, IEEE Trans* 39(2):309–320
- Price JC (1995) Examples of high resolution visible to near-infrared reflectance and a standardize collection for remote sensing studies. *Int J Remote Sens* 16:993–1000
- Priego B, Souto D, Bellas F, Duro RJ (2013) Hyperspectral image segmentation through evolved cellular automata. *Pattern Recogn Lett* 34(14):1648–1658
- Roberts DA, Gardner M, Church R, Ustin S, Scheer G, Green RO (1998) Mapping chaparral in the Santa Monica mountains using multiple end member spectral mixture models. *Remote Sens Environ* 65:267–279
- Roessner S, Segl K, Heiden U, Kaufmann H (2001) Automated differentiation of urban surfaces based on airborne hyperspectral imagery. *IEEE Trans Geosci Remote Sens* 39:1525–1532
- Segl K, Kaufmann H (2001) Detection of small objects from high resolution panchromatic satellite imagery based on supervised image segmentation. *IEEE Trans Geosci Remote Sens* 39(9):2080–2083
- Serpico SB, Bruzzone L (1994) A new search algorithm for feature selection in hyperspectral remote sensing images. *IEEE Trans Geosci Remote Sens* 39(7):1360–1367
- Siedlecki W, Sklansky J (1988) On automatic feature selection. *Int J Pattern Recognit Artif Intell* 2(2):197–210
- Tarabalka Y, Chanussot J, Benediktsson JA (2010) Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recogn* 43(7):2367–2379
- Zhang Q, Pauca VP, Plemmons RJ, Nikic DD (2013) Detecting objects under shadows by fusion of hyperspectral and lidar data: a physical model approach. In: 5th workshop hyperspectral image and signal processing: evolution in remote sensing, WHIS-PERS2013, 2013, pp 1–4

Incremental Dynamic Analysis

Dimitrios Vamvatsikos

School of Civil Engineering, National Technical
University of Athens (N.T.U.A.), Athens, Greece

Introduction

An important issue in performance-based earthquake engineering is the estimation of structural performance under seismic loads. In particular, one is interested in estimating the probabilistic distribution of structural response in the form of engineering demand parameters (EDPs) such as peak interstory drift, peak floor acceleration, moment, or shear, given the level of seismic intensity represented by a (typically scalar) intensity measure (IM). Incremental dynamic analysis (IDA) has been developed to provide such information by employing nonlinear dynamic analyses of the structural model under a suite of ground motion records, each scaled to several intensity levels designed to force the structure all the way from elasticity to final global dynamic instability.

The use of ground motion scaling for determining the response of a structure at increasing levels of intensity is an old technique. Still, it had not been used systematically to quantify the probabilistic nature of structural response until the start of the SAC/FEMA (2000a, b) project that was conceived in the aftermath of the Northridge

1994 earthquake. Therein, a precursor to IDA was proposed in the form of the “dynamic push-over,” a method to determine the global collapse capacity of structures (Luco and Cornell 1998). Subsequently, this was recast and formalized by Vamvatsikos and Cornell (2002) as a comprehensive procedure to assess the statistics of EDP demand given the IM as well as the required (“capacity”) IM to achieve given values of EDP at any level of structural behavior. It is this format that is known as IDA and will be described in the following sections, discussing the necessary steps and concepts in executing, postprocessing, and applying IDA to solve problems of engineering significance.

Intensity Measure and Ground Motions

First and foremost for IDA, an efficient and sufficient IM (Luco and Cornell 2007) should be selected. An efficient IM will generally be well correlated with the EDPs of choice, thus showing low dispersion of demand given the IM and subsequently allowing the determination of EDP demand or IM capacity statistics using a relatively low number of ground motion records. In other words, efficiency is synonymous with economy in computational resources, becoming a nontrivial issue when considering that a single nonlinear dynamic analysis of a moderate complexity model can easily last more than 30 min. The second requirement, sufficiency, is defined

as the independence of the distribution of EDP given the IM from any other seismological parameters that may characterize the ground motion, e.g., duration, magnitude, spectral shape, or the presence of a pulse indicative of near-source forward directivity. A sufficient IM essentially captures all seismological information needed to determine the effect of a ground motion record on the structure being investigated, thus allowing unrestricted scaling in theory. Naturally, any IM for which a seismic hazard curve can be practically estimated will never be fully sufficient; thus, excessive scaling may introduce biased estimates of response (Luco and Bazzurro 2007).

A standard choice used in the literature is the 5 % damped first-mode (pseudo)spectral acceleration $S_a(T_1)$ (Shome et al 1998; Shome and Cornell 1999). This is generally adequate for first-mode-dominated structures that do not displace far into the nonlinear region, as is the case of most existing brittle or moderately ductile low/mid-rise buildings. For taller or asymmetric structures where higher modes become important or modern buildings that exhibit significant ductility, improved IM alternatives should be sought. One particularly attractive option is the geometric mean of spectral acceleration values at several periods (Cordova et al 2000; Vamvatsikos and Cornell 2005; Bianchini et al 2009) or an inelastic spectral displacement-based IM (Luco and Cornell 2007). Both can largely alleviate the effect of spectral shape, being able to capture the period elongation characterizing ductile structures and the effect of higher modes.

In terms of selecting ground motion records, the general idea is to reduce the bias that a less than ideal IM may introduce. A general rule of thumb is to utilize accelerograms that can significantly damage the investigated structure without much scaling. For modern ductile structures, this generally means records having naturally high $S_a(T_1)$ values. Still, IDA is structured in such a way that the same set of records is used throughout the entire range of structural behavior (elastic–inelastic collapse). Whenever site conditions are deemed to have different impact at

different levels of intensity, thus fundamentally changing the nature of records that one would use at low versus high levels of the IM, then methods of analysis other than IDA should probably be employed.

Structural Model

For performing IDA, a (a) realistic, (b) low-to-medium-complexity, (c) robust nonlinear structural model should be formed. Each of these three requirements comes with its own reasons.

For one, realism means incorporating all pertinent sources of material and geometric nonlinearity that are expected to arise. This should include, for example, plastic-hinge formation zones for moment-resisting frames, brace buckling for braced frames, and P-Delta effects. Nonsimulated failure modes, such as the shear failure of members or the brittle failure of beam–column joints, can be incorporated in the analysis a posteriori. Still, they essentially remove the model's ability to track structural behavior beyond their first occurrence. This means that whenever nonsimulated failures are found to have occurred, one cannot trust the model to provide estimates beyond that point.

The requirement of low-to-moderate complexity is necessary for ensuring that the computationally intensive IDA will generally be confined to reasonable amounts of time. Complex nonlinear models can complicate the execution to no end, easily forcing each nonlinear analysis to last several hours or, worse, often forcing the analyst to consider reducing the number of records or dynamic runs per record employed. This can be a tricky situation that may lead to inaccurate or biased results. Thus, it is best to strike a trade-off between model complexity and fidelity, striving to find a good balance that allows using a rich set of at least 20–30 ground motion records with six or more time history analyses for each.

Finally, robustness becomes important in tracking structural behavior in the postelastic region, especially beyond the structure's

maximum-strength point, where it starts deteriorating and approaching collapse. To be able to ensure such results, a highly reliable software platform should be used. IDA by nature will drive both the model and the analysis software to their limits, making robustness difficult to achieve. Even then, typical complex finite element models that have been used successfully for elastic or mildly inelastic analysis tend to fare poorly when driven close to global dynamic instability, being consistently less reliable than their simpler versions besides significantly raising the computational cost. For example, it is often the case that distributed plasticity models are numerically less stable and always more expensive than lumped-plasticity ones. Therefore, considerable attention should be paid to the formation and testing of the model before IDA is performed, checking its behavior and stability via nonlinear static pushover analyses.

Execution

Performing IDA is conceptually simple. One only needs to take one record at a time, incrementally scale it at constant or variable IM steps, and perform a nonlinear dynamic analysis each time. Start from a low IM value where the structure behaves elastically, and stop when global collapse is encountered. The latter is defined as the occurrence of a nonsimulated failure mode or the appearance of a global dynamic instability as a collapse mechanism showing “infinite” EDP values at a given IM level. For a well-executed analysis and robust structural model, global dynamic instability manifests itself as numerical nonconvergence.

Selecting constant IM steps is often the simplest but also the most wasteful approach to IDA. As global collapse is encountered at widely varying IM levels for each record, some ground motions will require twice or thrice the dynamic analyses than others. Using variable IM steps estimated via the hunt & fill procedure of Vamvatsikos and Cornell (2002, 2004), offers instead consistent accuracy at a predefined

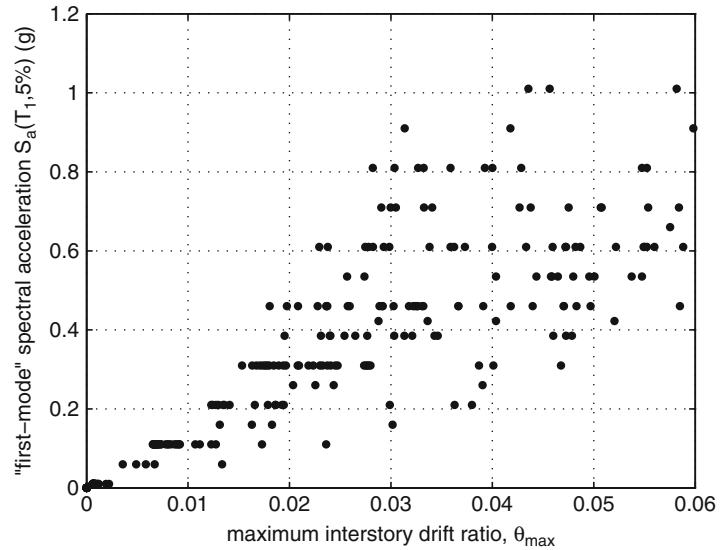
computational cost. Due to the nearly perfect computational independence of each dynamic analysis, IDA execution is an easily parallelizable problem (Vamvatsikos 2011). Hence, one can benefit from using a cluster of N computers to divide the total running time (almost) by N . Such savings make possible the use of Monte Carlo-based algorithms (Vamvatsikos 2014) to economically estimate any bias and additional variability introduced by modeling uncertainty.

Postprocessing

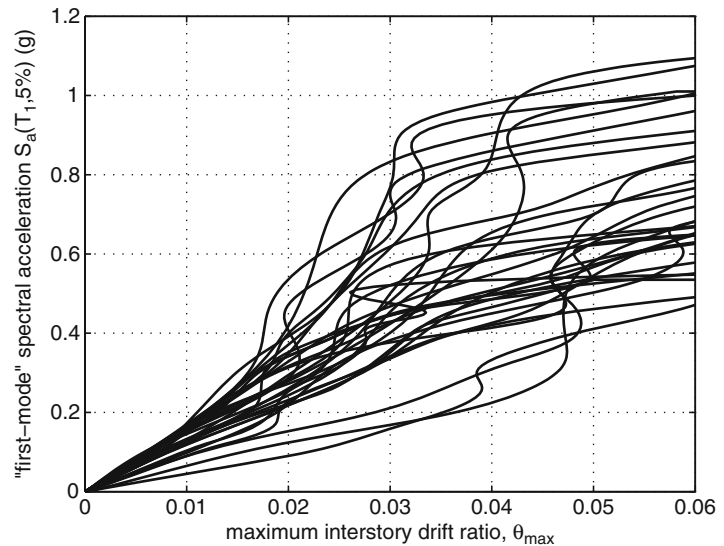
The results of IDA initially appear as distinct points, one per each dynamic analysis, in the EDP versus IM plane, as observed in Fig. 1 for a nine-story steel moment-resisting frame. Linear or spline interpolation (Vamvatsikos and Cornell 2004) is then employed to generate continuous IDA curves, one for each individual ground motion record, shown in Fig. 2. The variability offered by such curves is actually one of the eye-opening features of IDA, visually representing the probabilistic nature of seismic loading and the differences introduced even by (seemingly) similar ground motions. At the highest IM level that each IDA curve can reach, one can identify the characteristic flatline representing global collapse. In the example of Fig. 2, this is the intensity where the maximum (over all stories) peak interstory drift (the EDP of choice) increases without bound for each record, leading to global system collapse.

The complex picture of structural response shown by the IDA curves can be significantly simplified by summarizing them into the 16/50/84 % fractile curves shown in Fig. 3. This is simply the determination of the 16/50/84 % percentile values of EDP at each level of the IM, taking horizontal cross sections of the IDA curves or, equivalently, the 84/50/16 % percentiles of the IM given the EDP when vertical cross sections are assumed instead. The three curves of Fig. 3 thus provide the full characterization of the distribution of structural response via the central

Incremental Dynamic Analysis, Fig. 1 The resulting IM-EDP points for 30 ground motion records when performing IDA for a 9-story steel frame using $S_a(T_1)$ as the IM and the maximum over all stories peak interstory drift $\theta_{\max}$ as the EDP



Incremental Dynamic Analysis, Fig. 2 The 30 IDA curves derived by interpolating the IM-EDP points of Fig. 1 for the 9-story steel frame



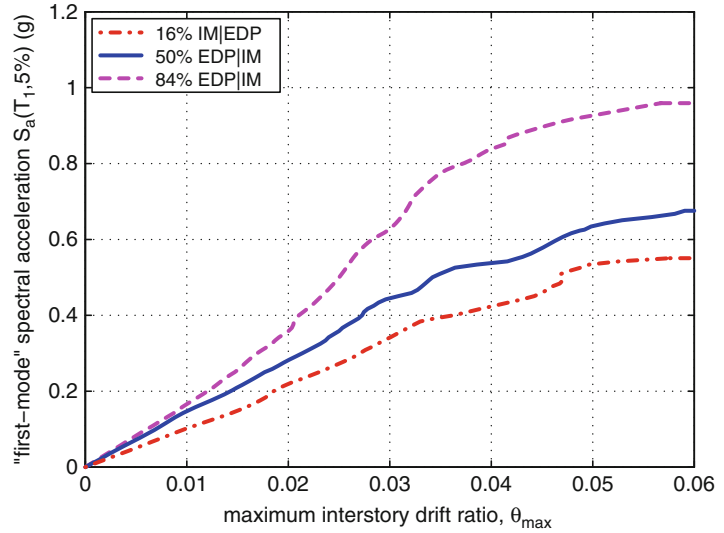
value (median) and the dispersion of EDP structural demand given the IM. For the example of the nine-story steel frame, one can observe how the median demand obeys the “equal displacement” rule, originally suggested by Veletsos and Newmark (1960) for elastoplastic single-degree-of-freedom systems: The median IDA maintains the same slope in the elastic and the inelastic range, meaning that an elastic version of this nine-story steel frame subject to the same level of ground motion would (in the median sense)

experience the same maximum interstory drift as the inelastic one. This only changes when the structure enters into the “negative stiffness” range where its strength starts degrading and the 50 % IDA deviates to the right to merge into the flatline representing the median IM collapse capacity value.

Finally, one can employ the IDA curves to define appropriate limit states and estimate the corresponding IM values of capacity. A limit state is usually tied to specific threshold EDP

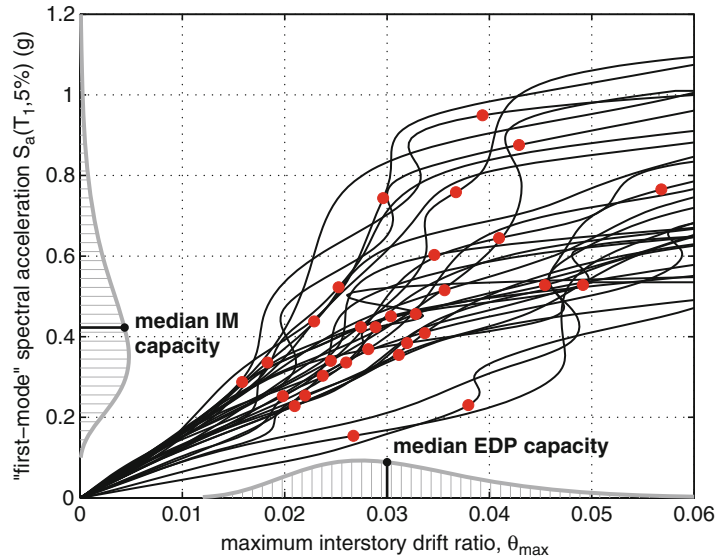
Incremental Dynamic

Analysis, Fig. 3 The 16/50/84 % fractile IDA curves obtained by summarizing the individual IDA curves of Fig. 2 for the 9-story steel frame



Incremental Dynamic

Analysis, Fig. 4 Thirty IDA curves, 30 limit-state capacity points and the corresponding EDP and IM capacity distributions for the 9-story steel frame (From Vamvatsikos 2013)



capacity values of one or more EDPs that can be either deterministic, i.e., assumed to be perfectly known, or probabilistic, the latter obviously being the more realistic option. When such threshold values are exceeded, the limit state is deemed to have been violated. The first (in terms of the lowest IM) exceedance of the limit state appears as a single point on each IDA curve (Fig. 4). The resulting distribution of IM coordinates (or IM capacities) of the limit-state points

can be used to define the so-called fragility functions, as discussed in the following section.

Use and Applications

IDA has found numerous uses within the framework performance-based earthquake engineering. The three main applications are discussed below.

First, IDA curves can be employed to evaluate fragility curves. A building-level fragility curve is a probability-valued function of the IM. It represents the probability of violating a limit state given the IM level. In terms of IDA, all that is needed is the determination of the limit-state points for each IDA curve. In the case where a single EDP is used to test for limit-state violation, these can be simply defined by simulating equiprobable values of the (random) EDP capacity. Then, one only needs to find the lowest-IM point on a given IDA curve that corresponds to each given EDP value of capacity. For example, when the EDP capacity is deterministic, all such points will align themselves along a single vertical line in the IM-EDP plane that intersects the horizontal axis at that specific value of EDP capacity. In either case, the cumulative distribution function (CDF) of the IM coordinates (or IM capacities) of the limit-state points (Fig. 4) directly provides the corresponding fragility curve. One can either employ the empirical CDF or an analytical approximation (typically via a lognormal assumption) to obtain either a point-by-point accurate estimate of the fragility curve or instead offer a simple (two-parameter) fit.

Second, the results of IDA either in their summarized form (Fig. 3) or as raw IDA curves (Fig. 2) are used within PEER-style frameworks (Cornell and Krawinkler 2000) to determine the distribution of EDP demand given the IM or P (EDP|IM) for any story or component of interest. As such, it is often one of the prominent methods employed for building seismic loss assessment, most notably in FEMA P-58 (FEMA 2012). The results of IDA are essentially convolved with the seismic hazard and appropriate loss/downtime/casualty calculations to derive the mean annual frequency of exceeding any specified level of repair cost, time to repair, or number of people injured or dead.

Finally, IDA has been adopted as the method of choice for determining strength reduction R-factors for USA, as presented in FEMA P-695 (FEMA 2009). R-factors, also known as behavior q-factors, are used in every modern seismic code to appropriately reduce the required design strength of a structure to take advantage of its

ductility. Their assessment requires a careful evaluation of the collapse capacity of a comprehensive set of archetypes to ensure that a structure designed according to the specified reduced strength can still offer the required level of safety.

Summary

Incremental dynamic analysis is a method for assessing the distribution of structural response at every level of structural behavior from elasticity to final global dynamic instability. It employs numerous nonlinear dynamic analyses of the structural model subject to a suite of ground motion records that are scaled to multiple levels of intensity. By interpolating such results and appropriately summarizing them, one can estimate any desired statistic of response given the seismic intensity. Additionally, using the IDA results, any number of limit states from minor damage to global collapse may be defined, and the associated fragility functions can be effortlessly estimated. Still, to preserve the validity of such results and ensure unbiased estimation, attention should be paid to the ground motions used and the intensity measure selected to represent them. Then, IDA becomes a powerful tool for performance-based assessment of structures, factually linking seismic hazard and structural behavior to estimate losses, downtime, and casualties.

Cross-References

- ▶ [Analytic Fragility and Limit States \[P\(EDP|IM\)\]: Nonlinear Dynamic Procedures](#)
- ▶ [Analytic Fragility and Limit States \[P\(EDP|IM\)\]: Nonlinear Static Procedures](#)
- ▶ [Assessment of Existing Structures Using Response History Analysis](#)
- ▶ [Nonlinear Dynamic Seismic Analysis](#)
- ▶ [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- ▶ [Seismic Collapse Assessment](#)
- ▶ [Seismic Fragility Analysis](#)

- [Seismic Loss Assessment](#)
- [Time History Seismic Analysis](#)

References

- Bianchini M, Diotallevi PP, Baker JW (2009) Predictions of inelastic structural response using an average of spectral accelerations. In: Proceedings of the 10th International Conference on Structural Safety & Reliability (ICOSSAR), Osaka
- Cordova PP, Deierlein GG, Mehanny SS, Cornell CA (2000) Development of a two-parameter seismic intensity measure and probabilistic assessment procedure. In: Proceedings of the 2nd U.S.-Japan workshop on performance-based earthquake engineering methodology for reinforced concrete building structures, Sapporo, pp 187–206
- Cornell CA, Krawinkler H (2000) Progress and challenges in seismic performance assessment. PEER Center News. <http://peer.berkeley.edu/news/2000spring/index.html>. Accessed May 2014
- FEMA (2009) Quantification of seismic performance factors. FEMA P-695 report, prepared by the Applied Technology Council for the Federal Emergency Management Agency, Washington, DC
- FEMA (2012) Seismic performance assessment of buildings. Methodology, vol 1. FEMA P-58-1 report, prepared by the Applied Technology Council for the Federal Emergency Management Agency, Washington, DC
- Luco N, Bazzurro P (2007) Does amplitude scaling of ground motion records result in biased nonlinear structural drift responses? *Earthq Eng Struct Dyn* 36:1813–1835
- Luco N, Cornell CA (1998) Effects of random connection fractures on demands and reliability for a 3-storey pre-Northridge SMRF structure. In: Proceedings of the 6th US national conference on earthquake engineering, Seattle
- Luco N, Cornell CA (2007) Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions. *Earthq Spectra* 23(2):357–392
- SAC/FEMA (2000a) Recommended seismic design criteria for new steel moment-frame buildings, FEMA-350, SAC Joint Venture, Federal Emergency Management Agency, Washington, DC
- SAC/FEMA (2000b) Recommended seismic evaluation and upgrade criteria for existing welded steel moment-frame buildings, FEMA-351, SAC Joint Venture, Federal Emergency Management Agency, Washington, DC
- Shome N, Cornell CA (1999) Probabilistic seismic demand analysis of nonlinear structures. RMS program, report no RMS35, PhD thesis, Stanford University
- Shome N, Cornell CA, Bazzurro P, Carballo JE (1998) Earthquakes, records and nonlinear responses. *Earthq Spectra* 14(3):469–500
- Vamvatsikos D (2011) Performing incremental dynamic analysis in parallel. *Comput Struct* 89(1–2):170–180
- Vamvatsikos D (2013) Derivation of new SAC/FEMA performance evaluation solutions with second-order hazard approximation. *Earthq Eng Struct Dyn* 42:1171–1188
- Vamvatsikos D (2014) Seismic performance uncertainty estimation via IDA with progressive accelerogram-wise latin hypercube sampling. *ASCE J Struct Eng* 140(8):A4014015
- Vamvatsikos D, Cornell CA (2002) Incremental dynamic analysis. *Earthq Eng Struct Dyn* 31(3):491–514
- Vamvatsikos D, Cornell CA (2004) Applied incremental dynamic analysis. *Earthq Spectra* 20(2):523–553
- Vamvatsikos D, Cornell CA (2005) Developing efficient scalar and vector intensity measures for IDA capacity estimation by incorporating elastic spectral shape information. *Earthq Eng Struct Dyn* 34:1573–1600
- Veletsos AS, Newmark NM (1960) Effect of inelastic behavior on the response of simple systems to earthquake motions. In: Proceedings of the 2nd world conference on earthquake engineering, vol 2. Japan, pp 895–912

InSAR and A-InSAR: Theory

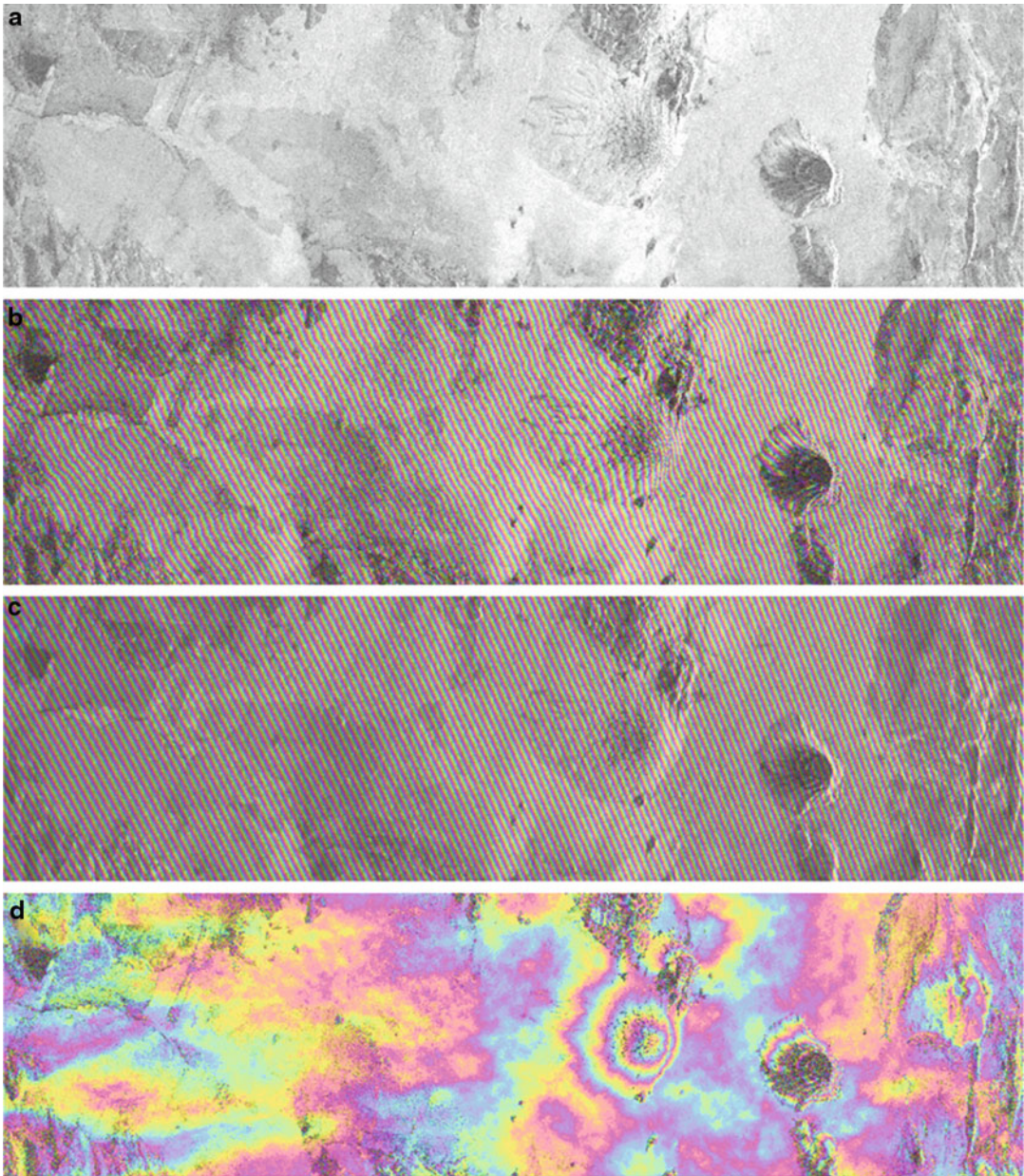
Andrew Hooper
School of Earth and Environment,
The University of Leeds, Maths/Earth and
Environment Building, Leeds, UK

Synonyms

DInSAR; Interferometric synthetic aperture radar; MT-InSAR; Multi-temporal InSAR; Persistent scatterer InSAR; PSI; PS-InSAR; SAR interferometry; SBAS; SB-INSAR; Small baseline InSAR; Time-series InSAR; TS-InSAR

Introduction

The SAR technique allows the formation of high-resolution radar images from the data acquired by side-looking instruments installed on spacecraft, aircraft, or the ground. The fundamentals underlying SAR image processing are presented in Chapter 3 “► [SAR Images, Interpretation of](#)”. Each pixel of an image corresponds to the



InSAR and A-InSAR: Theory, Fig. 1 Images formed from Envisat data acquired 35 days apart over north Iceland: (a) amplitude of master acquisition, (b) raw interferogram, (c) simulated reference phase projected onto interferometric amplitude, and (d) interferogram after reference phase is subtracted from the raw interferogram

phase. The perpendicular baseline between the two acquisition geometries is 63 m, giving an altitude of ambiguity of 145 m. In other words each fringe in (d) represents another 145 m of elevation, meaning that the a large volcano just east of center (Kollóttadyngja) rises $\sim 3 \times 145 = 435$ m above the surroundings

scattered signal from a resolution element on the ground, which is transmitted and received by the SAR. A pixel is characterized by two values: the amplitude and the phase. While the amplitude of

a single image can be interpreted in terms of the backscattering properties of the ground (Fig. 1a), the phase is not very informative because it is a pseudorandom contribution from the

configuration of all scatterers within the resolution element. However, provided that the scattering properties of the element are similar for the two acquisitions, the difference in phase between two images can be interpreted in terms of the change in range between the radar instrument and the target, which is the principle of SAR interferometry (InSAR).

For two SAR images acquired at the same moment from slightly different positions, the change in range can be interpreted in terms of the topography, allowing for the production of digital elevation models. For two images acquired at different times, movement of the ground between acquisitions can also contribute to the change in range. Thus, InSAR is also a powerful technique for measuring deformation of the ground, and this is the most common application of InSAR (sometimes referred to as differential InSAR or DInSAR). While the InSAR technique is applied to individual pairs of images, advanced InSAR (A-InSAR) techniques have been developed to analyze multiple SAR images together. These have the advantage of allowing

error terms to be better estimated and also provide a framework for looking at the temporal evolution of deformation.

Since the launch of ERS-1 by the European Space Agency (ESA) in 1991, there has always been at least one SAR satellite in operation. Therefore, in addition to the ability to use InSAR to monitor ongoing deformation, there is a long archive of data that allows us to also look at deformation that occurred in the past. In the past two decades, SAR satellites have evolved from mostly C-band sensors with ground-pixel resolutions of tens of meters to C-, X-, and L-band systems that can have spatial resolutions better than 1 m, with some satellites flying in formation or as part of a constellation. Currently, we are entering a golden age for InSAR, with a new generation of satellite missions that have been launched recently or will be launched in the next few years. These include ESA's Sentinel-1 – the first SAR mission that is operational in nature rather than purely scientific, with data being available within minutes of acquisition. A list of all satellites useful for InSAR is provided in Table 1.

InSAR and A-InSAR: Theory, Table 1 Past and present side-looking SAR satellite missions (as of 1 November 2014)

Mission	Period of operation	Wavelength (cm)	Orbit repeat (days)
SEASAT	Jun–Oct 1978	23.5	17
ERS-1	Jul 1991 to Mar 2000	5.66	3 or 35
ERS-2	Apr 1995 to Sep 2011	5.66	3 or 35
JERS-1	Feb 1992 to Oct 1998	23.5	44
SIR-C/X-SAR	9–20 Apr 1994 and 30 Sep to 11 Oct 1994	23.5, 5.8 and 3	N/A
RADARSAT-1	Nov 1995 to present	5.6	24
SRTM	11–22 Feb 2000	5.8 and 3.1	N/A
Envisat	Mar 2002 to Apr 2012	5.63	35 ^a
ALOS	Jan 2006 to Apr 2011	23.5	46
COSMO-SkyMed (constellation of four satellites)	Jun 2007 to present	3.1	16
	Dec 2007 to present	3.1	16
	Oct 2008 to present	3.1	16
	Nov 2010 to present	3.1	16
TerraSAR-X	Jun 2007 to present	3.1	11
TanDEM-X	Jun 2010 to present	3.1	11
RADARSAT-2	Dec 2007 to present	5.6	24
Sentinel-1	2014 to present	5.6	12 ^b
ALOS-2	2014 to present	23.5	14

^aFrom November 2010 to the end of its operation, Envisat operated in a 30 day orbit, which was not optimal for interferometry at high latitudes

^bIn 2016 a second Sentinel-1 satellite will be launched reducing the repeat time to 6 days

In this entry, the theory behind InSAR is described, including all the necessary steps to form and interpret an “interferogram.” Subsequently the major A-InSAR algorithms are described.

InSAR

Two SAR images acquired from approximately the same look direction can be combined to form an interferogram. The phase of an interferogram is related to the change in range between the sensor and the ground and forms the basis for interpretation in terms of ground elevation or ground deformation. This section covers the theory behind forming and interpreting a single interferogram (refer to Massonnet and Souyris (2008) for further details on processing).

Coregistration of SAR Images and Resampling

Before an interferogram can be formed from two images, the geometry of one of the images (denoted the “slave”) needs to be resampled into the geometry of the other (denoted the “master”). The reason for the difference in geometry is that the sensor acquires the images from slightly different locations. The primary difference is a translation and a linear stretch in range, although the presence of topography introduces higher-order offsets.

To carry out the resampling, the offset for each pixel must first be estimated, in a process known as “coregistration.” Coregistration can be achieved using the data themselves by cross-correlation of image amplitude, with satellite orbits being used only to estimate an initial approximate offset for the whole slave image. Range and azimuth offsets are estimated for hundreds to thousands of small subsets of the image. A 2-D polynomial is then fit to both the range and azimuth offsets, using weighting based on correlation. The polynomial, which is typically second or third order, is then used to calculate the offset for each individual pixel.

However, if the satellite orbits and the topography are accurately known, which tends to be the

case with modern satellites, it is possible to calculate the coregistration offset for each individual pixel geometrically. Amplitude cross-correlation is then used only to estimate a single offset in azimuth, due to timing errors, and a single offset in range, due to the average difference in atmospheric delay (see section “[Atmospheric Phase](#)”).

After coregistration, the slave is resampled into the master geometry by interpolation using an appropriate kernel, typically a sinc or raised cosine function, depending on the filtering that has been applied to the image. The output from the coregistration procedure is used to calculate the position of each master pixel within the slave image and the interpolation is used to calculate the value of the slave image at that point.

Raw Interferogram Formation

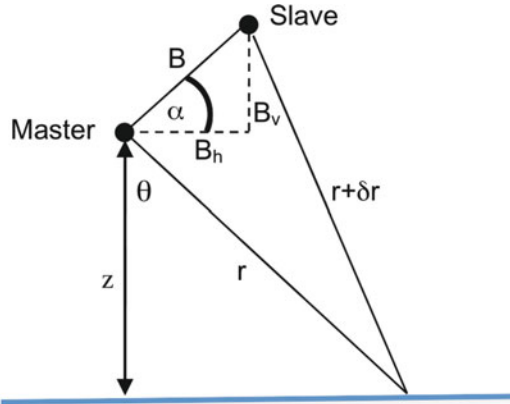
An interferogram is generated simply by multiplying the master image, pixel by pixel, by the complex conjugate of the resampled slave image. The phase of the interferogram is the phase difference between master and slave images and the amplitude is the product of the amplitude of the two images (Fig. 1b).

The phase of the raw interferogram contains contributions from several sources:

$$\phi = \phi_{\text{ref}} + \phi_{\text{topo}} + \phi_{\text{def}} + \phi_{\text{atmos}} + \phi_{\text{noise}}, \quad (1)$$

where ϕ_{ref} is the reference surface phase due to a difference in look angle of the master and slave (section “[Reference Phase](#)”), ϕ_{topo} is phase due to topography (section “[Topographic Phase](#)”), ϕ_{def} is phase due to deformation (section “[Deformation Phase](#)”), ϕ_{atmos} is phase due to propagation through the atmosphere (section “[Atmospheric Phase](#)”), and ϕ_{noise} is phase noise due to decorrelation and other effects (section “[Decorrelation](#)”).

Once an interferogram is generated, the next steps involve reducing the contributions from sources which are not of interest, in order to isolate those which are.



InSAR and A-InSAR: Theory, Fig. 2 Imaging geometry (not to scale; B is typically 10^3 – 10^4 times smaller than r). Master and slave sensors are flying into the page and looking down and to the right

Reference Phase

There is a contribution to the interferometric phase from the change in range resulting from the two images being acquired from different points in space. This can be separated into the reference phase expected if the ground surface were at a reference surface (e.g., WGS84) and phase due to modulation of the reference surface by topography. The reference phase can always be estimated and subtracted, although the accuracy of the estimate depends on the accuracy of the orbit data.

The change in range due to geometry for a reference surface can be estimated as follows (Fig. 2). From the law of cosines,

$$\frac{r^2 + B^2 + (r + \delta r)^2}{2rB} = \cos(\alpha + \pi/2 - \theta). \quad (2)$$

where r is the range from the satellite to the ground for the master acquisition, B (baseline) is the distance between the master and slave sensor, α is the angle between the horizontal and the baseline vector, and θ is the look angle. Neglecting second-order terms in δr^2 and B/r , this reduces to

$$\delta r = -B \left(\sqrt{1 - \frac{z_0^2}{r^2}} \cos \alpha - \frac{z_0}{r} \sin \alpha \right), \quad (3)$$

where z_0 is the height of the sensor above the reference surface. Change in range is converted into interferometric phase by dividing by the wavelength, λ , multiplying by -4π (-4π rather than -2π as the signal travels a distance $2\delta r$ farther), and dropping whole phase cycles. The reference phase, also known as “flat-Earth phase,” is then given by

$$\phi_{\text{ref}} = W \left\{ \frac{4\pi B}{\lambda r} \left(\sqrt{r^2 - z_0^2} \cos \alpha - z_0 \sin \alpha \right) \right\}, \quad (4)$$

where $W\{\cdot\}$ is the “wrapping” operator that drops whole phase cycles (Fig. 1c).

Topographic Phase

When generation of a DEM is the goal of the InSAR processing, the remaining phase is interpreted as the topographic contribution. To minimize noise from non-topographic contributions, simultaneous acquisitions therefore provide the best data for this application. In the case where isolation of another signal (usually deformation) is the goal of the processing, a DEM is used to simulate the topographic contribution and subtract it. If topography is considered as a perturbation to the reference surface, the relationship between elevation and phase can be determined by substituting z for z_0 in Eq. 4 and differentiating, giving

$$\frac{d\phi}{dz} = -W \left\{ \frac{4\pi B_{\perp}}{\lambda r \sin \theta} \right\}, \quad (5)$$

where $B_{\perp}$ is the perpendicular component of the baseline, or “perpendicular baseline.”

$\frac{d\phi}{dz}$ gives change in elevation with respect to the vertical vector below the satellite. To put topography in terms of the local vertical vector, the reference frame needs to be rotated, which has the effect that $\sin \theta$ becomes $\sin \theta_i$ where θ_i is the angle of incidence at the reference surface, rather than the look angle.

Considering only the first term in a Taylor expansion, the total geometric phase is given by

$$\phi \approx W \left\{ \phi_{\text{ref}} + \frac{4\pi B_{\perp}}{\lambda r \sin \theta_i} h \right\}, \quad (6)$$

where $h = -(z - z_0)$, the elevation above the reference surface. Topographic phase is then approximately proportional to elevation, which means that phase contours follow elevation contours, with one color “fringe” representing $\sim \lambda r \sin \theta_i / 2B_{\perp}$ m (Fig. 1d). This value is known as the *altitude of ambiguity*. Increasing perpendicular baseline therefore provides greater sensitivity to elevation and also leads to greater decorrelation (section “Decorrelation”).

Note that for greater accuracy, higher-order terms in the Taylor series must also be taken into account. Considering only the first-order term equates to a 40 m DEM error over 2,000 m of elevation change for ERS or typical Envisat acquisitions.

Given a DEM it is therefore relatively straightforward to project it into the radar coordinate system, calculate the topographic contribution to the phase using the formula above, and subtract it from the interferometric phase. Note that any errors in the DEM will lead to residual phase error that is proportional to the perpendicular baseline. Therefore, when the goal of the processing is to isolate a signal other than that due to topography, a small value for the perpendicular baseline is good for two separate reasons: firstly, the residual phase due to any DEM error is small and secondly, the geometric decorrelation is small (section “Decorrelation”).

When the aim is to generate a DEM from the interferometric phase, on the other hand, reversing the operation is not quite so simple, as the wrapping operator must first be inverted in a process known as *phase unwrapping*. This topic is covered in section “Phase Unwrapping.” Phase unwrapping provides the total phase difference (including whole phase cycles) between any two pixels, but not the absolute total phase for each pixel, so extra information is required to “tie” the DEM to the ground surface.

Deformation Phase

The most common application of InSAR is to measure deformation of the ground surface.

When an interferogram is formed between images acquired at different times, any change in range due to movement of the ground between the acquisitions will contribute to the interferometric phase; each displacement of half the wavelength of the carrier signal away from the satellite will increase the phase delay by one more phase cycle (Fig. 3):

$$\phi_{\text{def}} = -W \left\{ \frac{4\pi}{\lambda} \delta r \right\}, \quad (7)$$

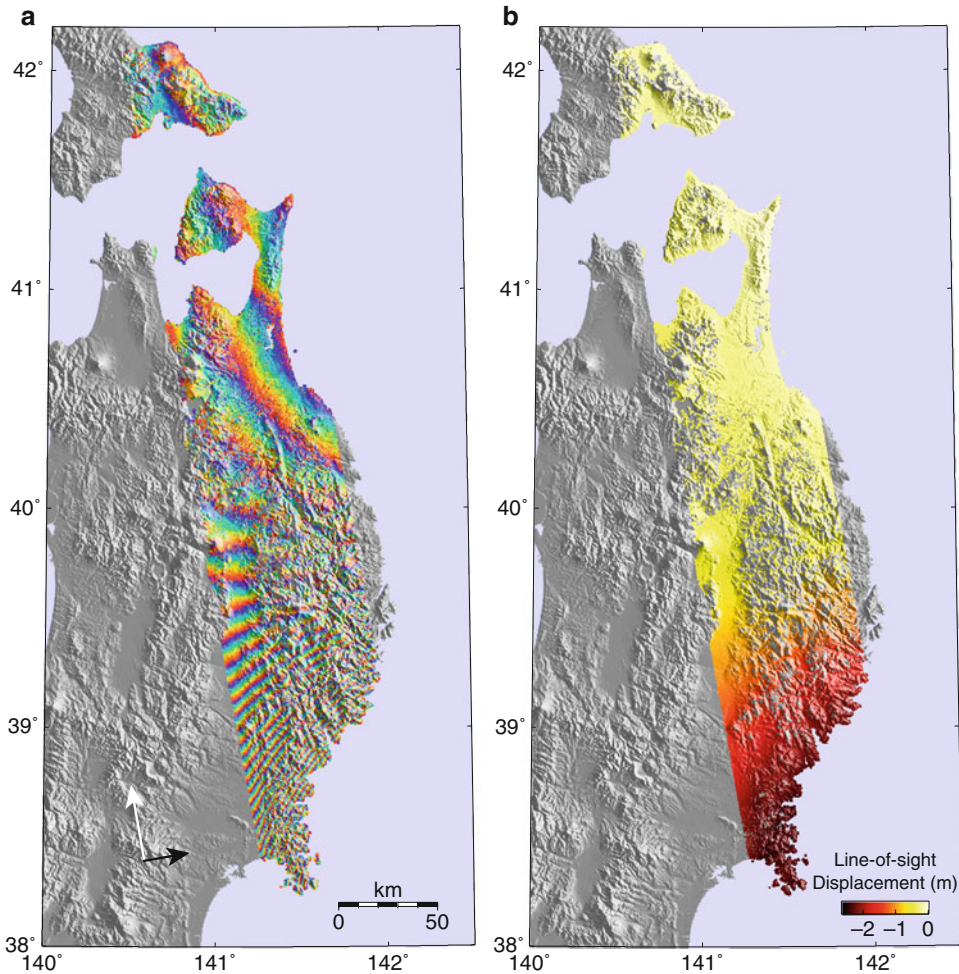
where δr is the change in range. Note that only the component of any displacement in the line-of-sight direction will result in a range change, and any displacement perpendicular to the line of sight will be invisible in the interferogram.

As in the case of DEM generation (section “Topographic Phase”), to interpret the phase as displacement, it must first be unwrapped (section “Phase Unwrapping”). The resulting unwrapped phase can then be used to derive relative line-of-sight displacements between any two points in the interferogram. To derive absolute displacements, the displacement at one or more points in the interferogram must be known a priori.

Atmospheric Phase

Propagation of the signal through the atmosphere influences the phase delay, by an amount that depends on atmospheric conditions, which are spatially variable (Hanssen 2001). Most of the spatial variability in the induced interferometric phase occurs in the ionosphere and the troposphere. The influence of the ionosphere on phase delay tends to be long wavelength (hundreds of km) and, except in the case of L-band data, is commonly ignored. There are, however, several methods for estimating the effect, if necessary (Meyer 2011). Phase delay during propagation through the troposphere, on the other hand, is equally significant at all microwave frequencies and variability is significant over short distances.

Propagation through the troposphere induces an additional phase delay which depends primarily on pressure, temperature, and humidity. These properties vary on two characteristic lateral



InSAR and A-InSAR: Theory, Fig. 3 An example interferogram displaying coseismic deformation for the $M_w = 9.0$ Tohoku-Oki earthquake, Japan, which occurred on 11 March 2011. The SAR data were acquired by the L-band ALOS satellite on 28 October 2010 and 15 March 2011. In (a) the interferometric phase is displayed with each color cycle representing 11.8 cm of displacement away from the satellite, which was moving in the direction of the *white arrow* and looking in the direction of the *black*

arrow. In (b) the phase is integrated and converted to line-of-sight displacement, indicating that the southernmost point moved more than 2.5 m away from the satellite, relative to the northern end. The orbit data provided for 15 March were preliminary and so some orbital signal remains. Signal is partially lost in the mountainous regions due to decorrelation, caused chiefly by snow (Modified after Hooper et al. (2012))

length scales: a short scale (a few km) induced by turbulent troposphere dynamics and a long scale (10s of km). The long-scale variation induces not only long-wavelength artifacts but also interferometric fringes that are strongly correlated with topography, as phase delay depends on how far through the troposphere the signal has traveled.

The turbulent variation is difficult to model and currently can only be reduced by filtering in time

and space (see section “[Phase Unwrapping](#)”). However, temporal filtering methods are less successful for reducing topographically correlated delay, especially when deformation is not linear in time, because SAR data sets typically do not sample seasonal atmospheric fluctuations evenly, resulting in biased estimates.

Alternatively, topographically correlated tropospheric artifacts can be partially corrected

using the information contained within the SAR data, based on the correlation between phase and elevation in nondeforming areas. In addition, the long-wavelength artifacts can be estimated over a whole image by consideration of how this correlation varies (Bekaert et al. 2015).

Complementary data sets, such as dense GNSS or meteorological measurements acquired at the same time as the SAR images, or global meteorological models, provide an alternative method for estimating the topographically correlated and long-wavelength tropospheric artifacts (Jolivet et al. 2011).

Decorrelation

A resolution element on the ground usually contains many scatterers, all of which contribute to the echo for that element. If the scatterers move with respect to each other, the echoes from the scatterers sum differently, a phenomenon known as *temporal decorrelation*. This induces a change to the interferometric phase. The echoes also sum differently if viewed from a different angle (*geometric* or *spatial decorrelation*) or probed with a different range of Doppler frequencies (*rotational decorrelation*). There is also *thermal decorrelation* term due to the signal-to-noise properties of the radar.

How closely the phase in one acquisition tracks that of another can be measured in terms of the complex correlation coefficient, or coherence, defined as

$$\gamma = \frac{E[u_1 u_2^*]}{\sqrt{E[|u_1|^2] E[|u_2|^2]}}, \quad (8)$$

where u_1 and u_2 are the complex signals from each image and $E[\cdot]$ represents the expected value. In practice, the expected values are usually estimated as the mean values for pixels in a small spatial window. Values of $|\gamma|$ range between 0 (no correlation) and 1 (perfect correlation).

Because the contribution from decorrelation to the interferometric phase is pseudorandom, the effects can be reduced, at the cost of resolution, by averaging the signal of all pixels within a

window of some specified size, or *multilooking*, or by spatial filtering (Goldstein and Werner 1998).

Rotational decorrelation can be lessened by only using the Doppler frequencies that were used to sample both images; frequencies outside of the overlapping frequency band are discarded for each image by bandwidth filtering. This has the effect of coarsening resolution, as resolution = $1/\text{bandwidth}$. In the case of there being no overlapping bandwidth, the result is total decorrelation.

In a similar way, geometric decorrelation can also be reduced by bandwidth filtering, at a cost of resolution (Gatelli et al. 1994). This is possible because changing the look angle slightly is approximately equivalent to shifting the chirp frequency,

$$\Delta f \approx -\frac{c}{\lambda} \frac{\Delta \theta}{\tan(\theta - \alpha)}, \quad (9)$$

where c is the speed of light, λ is the wavelength, $\Delta \theta$ is the change in look angle, and α is the angle between the horizontal and the baseline vector. $\Delta \theta$ is proportional to the perpendicular baseline; thus, if the perpendicular baseline is too large, there is no overlapping bandwidth, resulting in total decorrelation. The perpendicular baseline at which this occurs is referred to as the *critical baseline*, which is about 1,100 m for ERS and Envisat satellites as an example.

Phase Unwrapping

Phase unwrapping is the process of estimating the difference in the number of whole phase cycles between an arbitrary reference pixel and every other pixel. Pictorially, this can be considered as counting the number of color fringes from the reference pixel (Fig. 3). The basic assumption of phase-unwrapping algorithms is that the phase field is generally sampled at above the Nyquist rate, so that the phase difference between neighboring pixels is generally less than half of a phase cycle. The basis then, for most algorithms, is to calculate the phase difference between neighboring pixels, wrap this into the interval between $-1/2$ and $1/2$ of a phase cycle, and then integrate these phase differences.

However, it is rarely the case across that this assumption holds across a whole interferogram, due to steep gradients and/or discontinuities in the topographic or deformation phase or due to decorrelation noise. In this case, the path taken in the integration step matters – the algorithm needs to avoid integrating between pixels where the phase difference is more than half of a cycle. Mathematically these can be considered as branch cuts. In order to place these branch cuts, an extra constraint is required, and for most algorithms this can be framed in terms of minimization of a weighted function of the residuals between the estimated unwrapped phase differences and the wrapped phase differences:

$$\sum_{i,j} w_{i,j}^{(x)} |\Delta\phi_{i,j}^{(x)} - \Delta\psi_{i,j}^{(x)}|^p + \sum_{i,j} w_{i,j}^{(y)} |\Delta\phi_{i,j}^{(y)} - \Delta\psi_{i,j}^{(y)}|^p, \quad (10)$$

where $\Delta\phi^{(x)}$ and $\Delta\psi^{(x)}$ are the unwrapped and wrapped phase differences, respectively, between pixels in the x direction, $\Delta\phi^{(y)}$ and $\Delta\psi^{(y)}$ are the equivalent in the y direction, and w are user-defined weights (Chen 2001). The summations are carried out in both x and y directions over all i and j , respectively. This is sometimes referred to as an L^p -norm objective function, although strictly speaking, to meet the condition of positive scalability, the sum must be raised to the power of $1/p$, and p must be greater than or equal to 1. However, for $p \geq 1$, the solution that minimizes this objective function also minimizes the L^p -norm.

Two common algorithms that are applied to solve phase-unwrapping problems are the branch-cut algorithm (Goldstein et al. 1988) and the minimum cost flow algorithm (Costantini 1998). In the former case, the algorithm seeks to minimize the total number of nonzero residuals, equivalent to setting $p = 0$ and $w_{i,j}^{(x)} = w_{i,j}^{(y)} = 1$. In the latter case, the algorithm seeks to minimize the weighted sum of the absolute value of the residuals, equivalent to $p = 1$.

The third well-established algorithm implements a statistical cost flow methods and seeks to find the most probable solution overall for any

given probability distributions allocated to each phase difference (Chen 2001). The probability distributions are assigned based on the statistics of various measures, such as correlation and amplitude.

A-InSAR

Displacements can be estimated more accurately by processing many images together, rather than the two image approaches described above. While the approach to conventional InSAR is reasonably standardized, there are many variations on A-InSAR algorithms. However, an overview of the main approaches is described in this section.

The simplest approach for combining many images is to sum or “stack” the unwrapped phase of many conventionally formed interferograms. The deformation signal reinforces, whereas other signals typically do not. However, this approach is only appropriate when the deformation is episodic or purely steady state, with no seasonal deformation. Even then it is not optimal, as non-deformation signals are reduced only by unweighted averaging rather than by optimized estimation.

Algorithms for time-series analysis of SAR data have thus been developed to better address the issues facing conventional InSAR; decorrelation is tackled by using phase behavior in time to select pixels for which decorrelation noise is minimized, and non-deformation signals are estimated by a combination of modeling and filtering of the time series. These time-series algorithms fall into two categories, the first being persistent scatterer InSAR, which targets pixels whose scattering properties remain consistent both in time and from variable look directions, and the second being a more general small baseline approach. Because the two approaches are optimized for resolution elements with different scattering characteristics, they are complementary, and techniques that combine both approaches are able to extract the signal with greater coverage than either method alone (Hooper 2008).

Persistent Scatterer InSAR

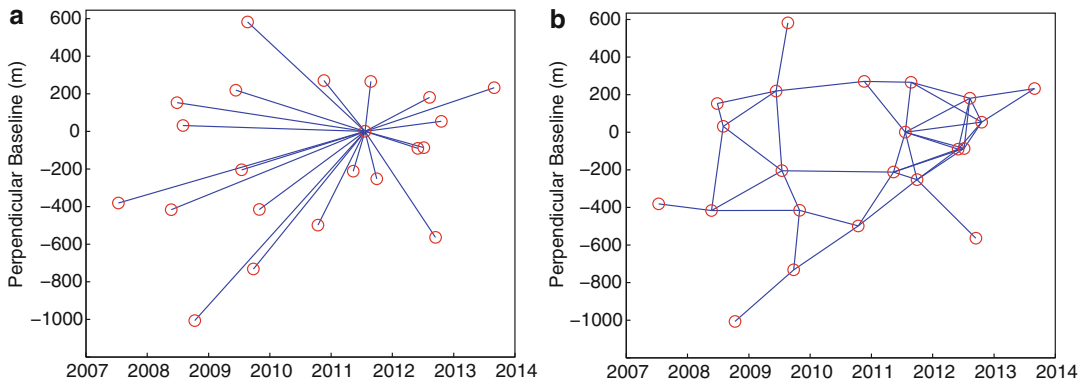
Decorrelation is caused by the contributions from scatterers within a resolution element summing differently. This can be due to a relative movement of the scatterers, a change in the looking direction of the radar platform, or the appearance or disappearance of scatterers, as in the case of snow cover. If one scatterer returns significantly more energy than other scatterers within a resolution element, however, decorrelation is reduced. This is the principle behind a “persistent scatterer” (PS) pixel (sometimes referred to as a “permanent scatterer”). Urban environments typically contain many of these dominant scatterers, for instance, roofs oriented such that they reflect energy directly back to the sensor and perpendicular structures that lead to a “double bounce,” where energy is reflected once from the structure and once from the ground, returning directly back to the sensor. Dominant scatterers can also occur in areas without man-made structures, e.g., appropriately oriented rocks, but there are fewer of them, and they tend to be less dominant.

PS algorithms operate on a time series of interferograms all formed with respect to a single “master” SAR image (Fig. 4a). The first step in the processing is the identification and selection of the usable PS pixels. There are two approaches to this; the first relies on modeling the deformation in time (Kampes 2005), and the second relies

on the spatial correlation of the deformation (Hooper et al. 2007). In the first approach, the phase is unwrapped during the selection process, by fitting a temporal model of evolution to the wrapped phase difference between pairs of nearby PS, although later enhancements to the technique allow for non-model-based improvements to the unwrapping. In the second approach a phase-unwrapping algorithm is applied to the selected pixels without assuming a particular model for the temporal evolution, other than it should be generally smooth.

In both approaches, deformation phase is then separated from atmospheric phase and noise by filtering in time and space, the assumption being that deformation is correlated in time, atmosphere is correlated in space but not in time, and noise is uncorrelated in space and time. Additional approaches to reducing atmospheric phase (section “Atmospheric Phase”) can be applied prior to the filtering. In comparative studies between the two PS approaches, estimates for the deformation estimates tend to agree quite well, but the second approach tends to result in better coverage, particularly in rural areas, which is where most volcanoes are sited.

The result of PS processing is a time series of displacement for each PS pixel, with much reduced noise terms (Fig. 5). The technique also has the advantage of being able to associate the deformation with a specific scatterer, rather than



InSAR and A-InSAR: Theory, Fig. 4 An example baseline plot for (a) the persistent scatterer method and (b) the small baseline approach. Red circles represent SAR images and blue lines indicate the interferograms that

are formed. Perpendicular baseline refers to the component of the satellite separation distance that is perpendicular to the look direction and is proportional to the difference in look angle



InSAR and A-InSAR: Theory, Fig. 5 Average displacement rates (mm/year) for Venice using a persistent scatterer approach applied to COSMO-SkyMed data from 2008 to 2011. The image background is an

aerophotograph acquired in 2009. Negative values indicate settlement, positive mean uplift (Modified after Tosi et al. (2013))

a resolution element of dimensions dictated by the radar system, usually on the order of many meters. This allows for very high-resolution monitoring of infrastructure.

Small Baseline InSAR

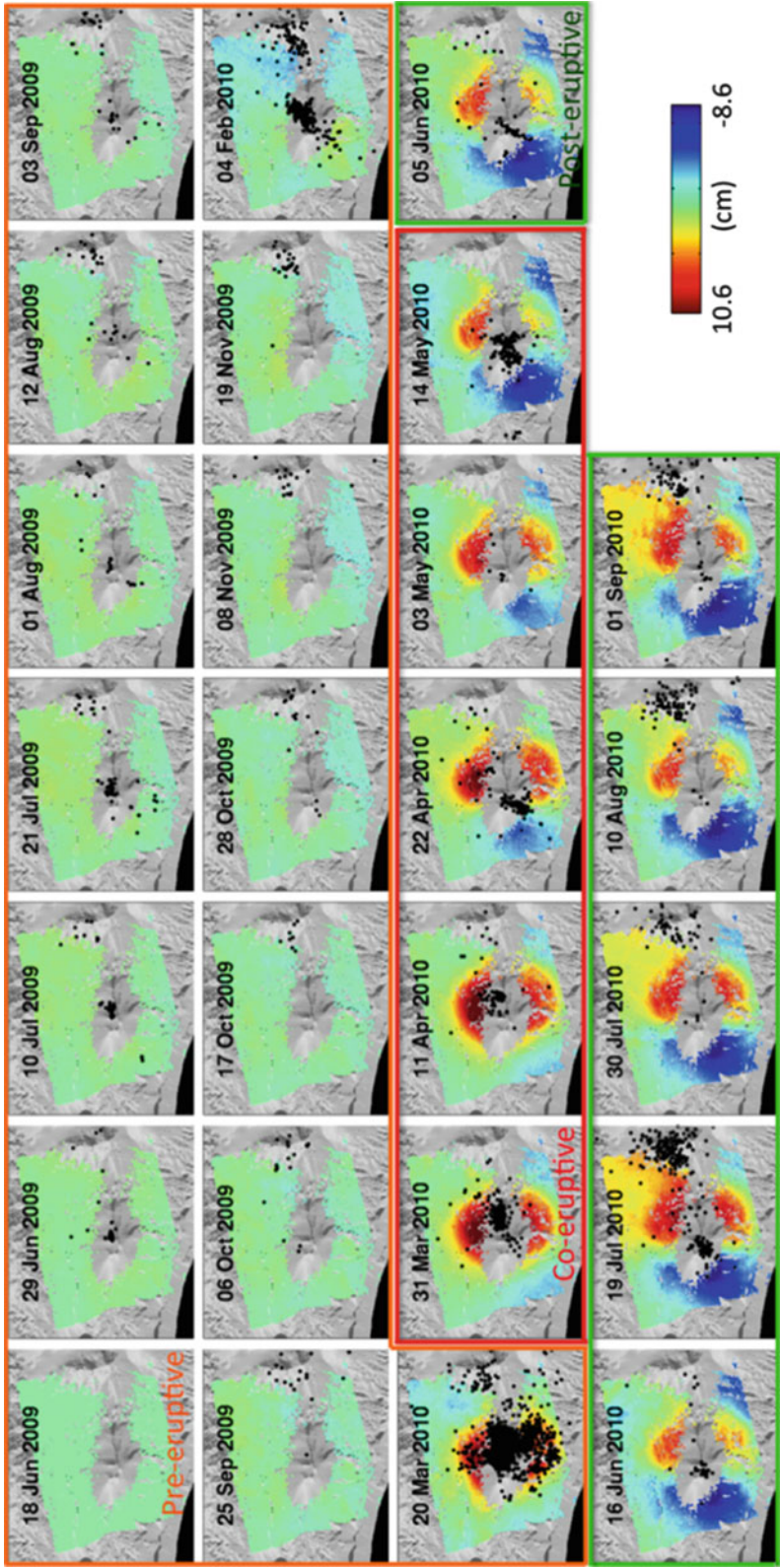
A drawback of the PS technique for many applications is that the number of PS pixels in rural environments may be limited. For non-PS pixels, containing no dominant scatterer, phase variation due to decorrelation may be large enough to obscure the underlying signal. However, by forming interferograms only between images separated by a short time interval and with a small difference in look direction, decorrelation is minimized, and for some resolution elements can be small enough that the underlying signal is still detectable. Pixels for which the phase decorrelates little over short time intervals are the targets of small baseline methods.

Interferograms are formed between SAR images that are likely to result in low

decorrelation noise, in other words, those that tend to minimize the difference in time and look direction (Fig. 4b). As it is not possible to minimize both of these at once, assumptions are made about their relative importance, based on the scattering characteristics of the area of interest.

The interferograms can then be multilooked to further decrease decorrelation noise, either using a standard windowing approach or by using amplitude to identify pixels with similar scattering characteristics (Ferretti et al. 2011). However, there may be isolated ground resolution elements with low decorrelation properties that are surrounded by elements that decorrelate highly, such as a small clearing in a forest, for which multilooking will increase the noise. Therefore, algorithms have been developed that operate at full resolution (Hooper 2008), with the option to reduce resolution later in the processing chain by selective multilooking.

Pixels are selected based on their estimated coherence in each of the interferograms, using either standard ensemble coherence estimation or enhanced techniques in the case of



InSAR and A-InSAR: Theory, Fig. 6 Cumulative line-of-sight displacement over Eyjafjallajökull Volcano, Iceland, from 18 June 2009 to 01 September 2010, processed using a small baseline approach (track 132 ascending mode). The co-eruptive interval includes a flank eruption, during which there was very little deformation, and the later summit eruption, during which the volcano deflated. The background is shaded topography. The hole in the data is due to ice and snow cover. *Black dots* are earthquake epicenters for each epoch (Icelandic Meteorological Office) (Modified after Pinel et al. (2014))

full-resolution algorithms. The phase is then unwrapped either spatially in two dimensions using standard approaches or using the additional dimension of time in three-dimensional approaches (Pepe and Lanari 2006; Hooper 2010). At this point the phase is inverted to give the phase at each acquisition time with respect to a single image, using least squares, singular value decomposition, or minimization of the L^1 -norm (Fig. 6).

Separation of deformation and atmospheric signals is achieved by filtering the resulting time series in time and space, as in the PS approach. Alternatively, if an appropriate model for the evolution of deformation in time is known, the different components can be directly estimated from the small baseline interferograms (Biggs et al. 2007).

Summary

InSAR is a powerful tool for high-resolution measurement of surface heights with meter scale accuracy, surface displacements with cm accuracy, and surface velocities with mm/year accuracy. Spatial resolution varies from under a meter to tens of meters, depending on the sensor and mode of operation.

Studies of ground deformation, in particular, have benefitted from more than 20 years of continuous satellite SAR observations. This archive is about to be enhanced by new satellite SAR missions, in addition to the growing application of ground-based and airborne SAR systems, which will provide improved temporal and spatial resolution, broader geographic coverage, and better availability to the scientific community.

The main advantage of InSAR remains its ability to collect useful imagery without regard to the time of the day or weather conditions and to record surface deformation on the scale of mm/year over broad regions with a spatial resolution on the order of meters. InSAR applications to seismology can thus be both regional in scope (entire tectonic regions) and highly focused (individual buildings).

Cross-References

- [Damage Detection in Built-Up Areas Using SAR Images](#)
- [Earthquake Magnitude Estimation](#)
- [SAR Images, Interpretation of](#)
- [SAR Tomography for 3D Reconstruction and Monitoring](#)
- [Urban Change Monitoring: Multi-temporal SAR Images](#)

References

- Bekaert DPS, Hooper A, Wright TJ (2015) A spatially-variable power-law tropospheric correction technique for InSAR data. *J Geophys Res Solid Earth*, doi:10.1002/2014JB011558
- Biggs J, Wright T, Lu Z, Parsons B (2007) Multi-interferogram method for measuring interseismic deformation: Denali fault, Alaska. *Geophys J Int* 170:1165–1179
- Chen CW (2001) Statistical-cost network-flow approaches to two-dimensional phase unwrapping for radar interferometry. PhD thesis, Stanford University
- Costantini M (1998) A novel phase unwrapping method based on network programming. *IEEE Trans Geosci Remote Sens* 36(3):813–821
- Ferretti A, Fumagalli A, Novati F, Prati C, Rocca F, Rucci A (2011) A new algorithm for processing interferometric data-stacks: SqueeSAR. *IEEE Trans Geosci Remote Sens* 49(9):3460–3470
- Gatelli F, Guarnieri AM, Parizzi F, Pasquali P, Prati C, Rocca F (1994) The wave-number shift in SAR interferometry. *IEEE Trans Geosci Remote Sens* 32(4):855–865
- Goldstein RM, Werner CL (1998) Radar interferogram filtering for geophysical applications. *Geophys Res Lett* 25(21):4035–4038
- Goldstein RM, Zebker HA, Werner CL (1988) Satellite radar interferometry: two-dimensional phase unwrapping. *Radio Sci* 23(4):713–720
- Hanssen RF (2001) Radar interferometry data interpretation and error analysis. Springer, Dordrecht
- Hooper A (2008) A multi-temporal InSAR method incorporating both persistent scatterer and small baseline approaches. *Geophys Res Lett* 35:L16302
- Hooper A (2010) A statistical-cost approach to unwrapping the phase of InSAR time series. European Space Agency (Special Publication) ESA. Noordwijk, The Netherlands SP-677
- Hooper A, Segall P, Zebker H (2007) Persistent scatterer InSAR for crustal deformation analysis, with application to Volcán Alcedo, Galápagos. *J Geophys Res* 112: B07407

- Hooper A, Bekaert D, Spaans K, Artkan M (2012) Recent advances in SAR interferometry time series analysis for measuring crustal deformation. *Tectonophysics* 514:1–13
- Jolivet R, Grandin R, Lasserre C, Doin M-P, Peltzer G (2011) Systematic InSAR tropospheric phase delay corrections from global meteorological reanalysis data. *Geophys Res Lett* 38:117311. doi:10.1029/2011GL048757
- Kampes BM (2005) Displacement parameter estimation using permanent scatterer interferometry. PhD thesis, Delft University of Technology
- Massonnet D, Souyris JC (2008) Imaging with synthetic aperture radar. EPFL-CRC Press, Lausanne, 280 p
- Meyer FJ (2011) Performance requirements for ionospheric correction of low-frequency SAR data. *IEEE Trans Geosci Remote Sens* 49(10):3694–3702
- Pepe A, Lanari R (2006) On the extension of the minimum cost flow algorithm for phase unwrapping of multitemporal differential SAR interferograms. *IEEE Trans Geosci Remote Sens* 44(9):2374–2383
- Pinel V, Poland M, Hooper A (2014) Volcanology: lessons learned from synthetic aperture radar imagery. *J Volcanol Geotherm Res* 289:81–113
- Tosi L, Teatini P, Strozzi T (2013) Natural versus anthropogenic subsidence of Venice. *Sci Rep* 3:2710

Insurance and Reinsurance Models for Earthquake

Katsuichiro Goda¹, Friedemann Wenzel² and James Edward Daniell²

¹Department of Civil Engineering, University of Bristol, Bristol, UK

²Geophysical Institute, Karlsruhe Institute of Technology, Karlsruhe, Germany

List of Acronyms

AEL	Annual expected loss
ART	Alternative risk transfer
CAT	Catastrophe
CEA	California Earthquake Authority
CNSF	Comision Nacional de Seguros y Fianzas
CVaR	Conditional value at risk
DRR	Disaster risk reduction
EP	Exceedance probability
EQC	Earthquake Commission
GDP	Gross domestic product

GP	Generalized Pareto
HNDECI	Hybrid Natural Disaster Economic Conversion Index
ICI	Iceland Catastrophe Insurance
JERC	Japan Earthquake Reinsurance Company
LP-HC	Low-probability high-consequence
PML	Probable maximum loss
POT	Peak over threshold
SEL	Scenario expected loss
SUL	Scenario upper loss
TCIP	Turkish Catastrophe Insurance Pool
TREIF	Taiwan Residential Earthquake Insurance Fund
VaR	Value at risk

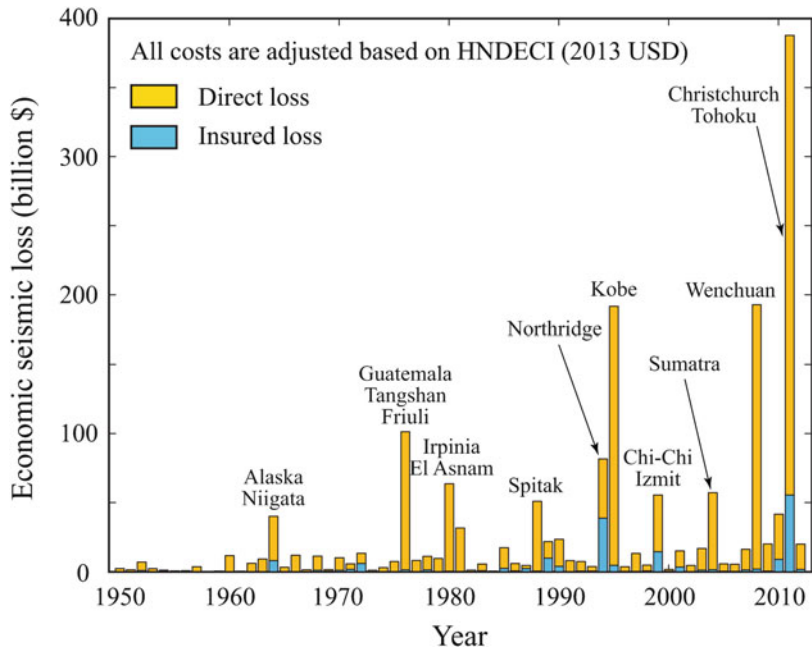
Introduction

Building sustainable and resilient communities against extremely large earthquakes is a global and urgent problem in active seismic regions. A catastrophic earthquake and its cascading events, such as tsunami and fire, affect multiple structures and infrastructure simultaneously and have far-reaching economic impact across various sectors in a complex manner. Incurred seismic damage includes loss of life and limb, direct financial loss to building properties and lifeline facilities (capital stock), and indirect loss due to the ripple effects of the direct loss across regional and national economies. All these consequences are revealed in an abrupt reduction of gross domestic product (GDP) for national economy and of stock indices and prices for companies.

One of such devastating events was the 2011 M_w 9.0 Tohoku earthquake in Japan among recent earthquake disasters around the world (e.g., 2004 Sumatra, 2008 Wenchuan, and 2010 Haiti earthquakes). For instance, a post-earthquake economic impact analysis of the 2011 Tohoku earthquake by Kajitani et al. (2013) revealed that the economic impact was significant not only in directly affected regions but also in regions outside the damaged areas due to disruptions of supply chains and domestic/international trade networks. The financial consequences of

Insurance and Reinsurance Models for Earthquake,

Fig. 1 Earthquake-related economic and insurance losses



such major earthquakes have long-term impacts (e.g., stagnant regional economy due to loss of market share). Households in disaster areas suffer from loss of asset and income; recovery from earthquakes can be extremely difficult, and financial public support/aid from governments and municipalities is not sufficient to cover the loss. Moreover, catastrophic earthquakes impose tremendous financial stress on insurers and reinsurers underwriting earthquake insurance policies (Grace et al. 2003; Muir-Wood 2011). The CATDAT damaging earthquakes database (Daniell et al. 2011) reports that the 2011 Tohoku earthquake and tsunami, the 2010–2011 Christchurch (New Zealand) sequences, and the 2010 Maule (Chile) earthquake rank in the top 10 list for the highest insured loss since 1900, exceeding \$10 billion. Figure 1 shows a history of estimated economic and insurance losses since 1950 according to the CATDAT database. The CATDAT values are adjusted based on the Hybrid Natural Disaster Economic Conversion Index (HNDECI), which takes into account temporal as well as geographical variation of consumer price indices, wage and low-wage indices, exchange rate, and other factors to facilitate the uniform and consistent comparison of the disaster

consequences (Daniell et al. 2011). The figure illustrates the significant magnitude of earthquake catastrophes, which is caused by a combination of physical hazard, exposure, vulnerability, and socioeconomic factors of the interconnected economies around the world.

Recent catastrophes have highlighted that the financial consequences to stakeholders can be significant. It is necessary to enhance the capability of dealing with such low-probability high-consequence (LP-HC) events through integrated seismic risk management by combining hard and soft disaster risk reduction (DRR) measures. Generally, hard and soft measures are complementary, and such integrated DRR solutions are robust against unforeseen and uncertain events. An important difference between hard risk mitigation measure (e.g., seismic retrofitting) and soft risk transfer measure (e.g., insurance) is that the former physically reduces the actual extent of seismic damage, while the latter transfers the incurred loss to a third party based on a pre-agreed risk-sharing scheme. Because the recovery process is (highly) nonlinear and dependent on various post-disaster situations (e.g., prolonged business interruption may result in loss of share in a competitive market), earthquake

insurance has the effects of accelerating the recovering process.

This entry is focused upon earthquake risk management using insurance and reinsurance. Earthquake insurance is a form of loss indemnity contract, where an insurer pays policyholders in the event of a major earthquake that causes damage to their properties. Insurance smoothes out the policyholder's wealth at different contingent states (i.e., no earthquake versus major earthquake), while an insurer underwrites policyholders' potential seismic risks collectively in exchange for insurance premium. Because an insurer may face an unacceptable level of insolvency risks due to catastrophic earthquakes, the risks may be transferred to a reinsurer, i.e., reinsurance refers to a risk-sharing contract between insurer and reinsurer. A reinsurer usually has greater capacity to absorb large financial shocks and helps diversify the seismic risks geographically and across different perils. It is also important to recognize the critical role of governments in national earthquake insurance programs, as they back up some layers of potential earthquake loss. An overview of the earthquake insurance and reinsurance system is given in the "[Insurance and Reinsurance System](#)" section, while a summary of existing insurance systems in major seismic countries is provided in the "[Existing Insurance Systems Around the World](#)" section.

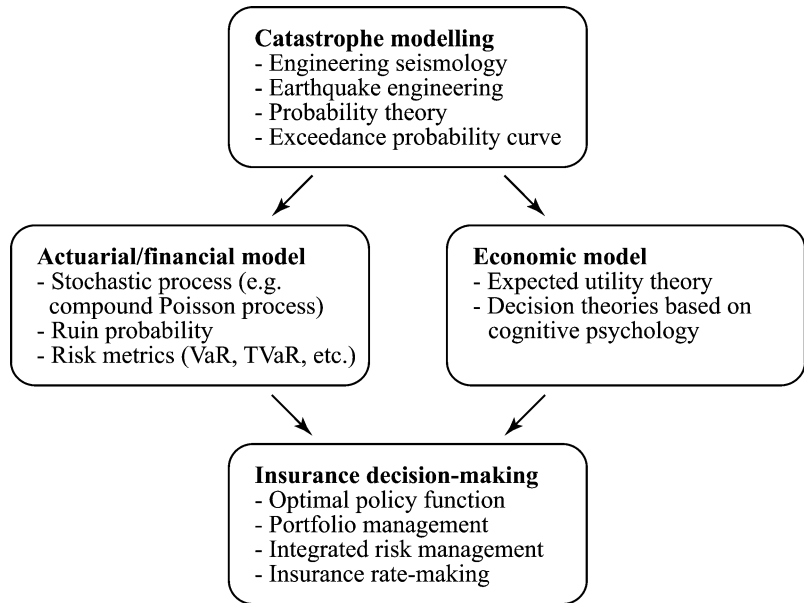
The nature of catastrophic earthquake risks, i.e., correlated loss generation, poses major challenges in achieving the effective diversification of collective seismic risk and may overwhelm the financial capacities of insurers and reinsurers. To avoid the potential insolvency and major constraint in business operation, utilizing alternative risk transfer (ART) tools, such as earthquake catastrophe (CAT) bonds, may be considered (Lalonde 2005; Cummins 2008). Another concern of providing earthquake loss coverage includes adverse selection and moral hazard (Grace et al. 2003), which are influenced by various factors, such as risk perception, asymmetrical information, (cognitive) psychology, framing, personal experience, etc. (Johnson et al. 1993;

Palm 1995; Kunreuther 1996; Kahneman 2003; Viscusi and Evans 2006). A summary of these challenges is provided in the "[Challenges in Insuring Low-Probability High-Consequence Risks](#)" section.

Approaches for modeling (catastrophe) insurance and reinsurance, adopted in different academic disciplines, are diverse, as their purposes and focuses are often different. Economists formulate insurance purchase problems as rational decision-making under uncertainty and examine arrangements for optimal insurance (Mossin 1968; Ehrlich and Becker 1972; Schlesinger 2000). On the other hand, behavioral economists and psychologists take a flexible view on decision-making problems by incorporating cognitive aspects of human beings (Johnson et al. 1993; Kahneman 2003; Viscusi and Evans 2006). In actuarial sciences, the problems are defined mathematically by capturing the key elements of the stochastic insurance risk processes (Rolski et al. 1999; Cossette et al. 2003; Goda and Yoshikawa 2012), which are not modeled in economic analyses. In earthquake engineering, earthquake CAT models that incorporate the state of the art in engineering seismology and earthquake engineering by taking into account uncertainties are used to assess the financial seismic risk quantitatively (Bommer et al. 2002; Dong and Grossi 2005; Wesson et al. 2009; Asprone et al. 2013; Yucemen 2013). Moreover, an extended approach considers the applications of actuarial/engineering methods in the context of economical decision-making problems (e.g., CAT models are embedded into the expected utility framework for insurance decision-making). Figure 2 depicts an overview of insurance models, involving various disciplines, such as seismology, earthquake engineering, statistics, economics, actuarial sciences, and psychology. Key features of different approaches are described in the "[Insurance and Reinsurance System](#)" section. Specifically, the "[Earthquake Catastrophe Model](#)" section discusses applications of the CAT models to earthquake insurance; the "[Actuarial and Financial Insurance Model](#)" section presents stochastic modeling of insurance

Insurance and Reinsurance Models for Earthquake,

Fig. 2 Overview of risk models for earthquake insurance decision-making



risk processes; and the “[Economic Insurance Model](#)” section gives a summary of methods based on economics and expected utility decision theories. Finally, in the “[Integrated Earthquake Risk Management](#)” section, the role of insurance and reinsurance in the context of integrated earthquake risk management is discussed by broadening the scope of this entry (e.g., earthquake risk premium discount as incentive to promote seismic retrofitting; Kleindorfer et al. 2005; Michel-Kerjan et al. 2013).

Insurance and Reinsurance System

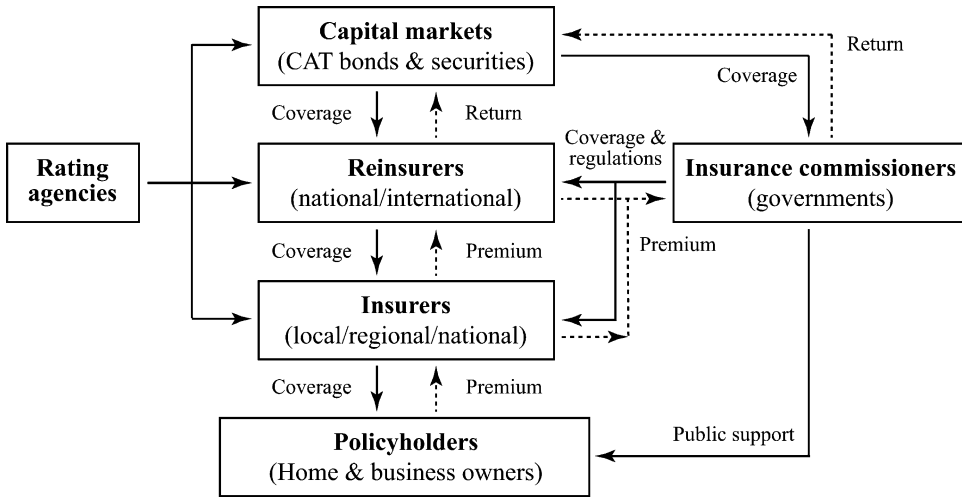
This section gives an overview of current insurance-reinsurance systems for earthquake risk coverage. In the “[Insurance and Reinsurance System](#)” section, key stakeholders as well as their interrelationships are defined. Subsequently, main features of the current insurance systems in major seismic regions/countries are described in the “[Existing Insurance Systems Around the World](#)” section. Subsequently, issues and challenges in insuring LP-HC earthquake events are discussed in the “[Challenges in Insuring Low-Probability High-Consequence Risks](#)” section.

Insurance-Reinsurance System

A generic representation of insurance-reinsurance system for managing earthquake risks is shown in Fig. 3. The identified key stakeholders are policyholders (e.g., owners of properties and enterprises), insurers operating at local/regional/national levels, reinsurers dealing with risk diversification at international scale, insurance commissioners/governments, rating agencies as auditor, and capital markets where investors gather for financial transactions.

Policyholders, Insurers, and Governments:

In the event of a major destructive earthquake, policyholders may suffer from seismic damage to their properties and assets and incur financial loss. To mitigate the potential financial impact, they may insure such risks by purchasing earthquake insurance coverage. Alternatively, they have options to mitigate the earthquake damage by adopting physical mitigation measures, such as seismic retrofitting and upgrading. Acceptance of earthquake insurance and/or physical risk mitigation measures largely depends on the risk perception of property owners and on the type of properties (e.g., home versus business). Issues related to risk perception and acceptance of earthquake protection are discussed in the “[Challenges](#)



Insurance and Reinsurance Models for Earthquake, Fig. 3 Overview of an insurance-reinsurance system

in Insuring Low-Probability High-Consequence Risks” section.

Insurers offer earthquake risk coverage, usually as part of other insurance (e.g., fire insurance and multi-peril coverage). Depending on country and property type, purchase of earthquake coverage is compulsory or voluntary. A primary insurer collects insurance premiums from policyholders and reimburses incurred seismic loss based on a pre-agreed policy upon the occurrence of a damaging earthquake. The insurance premium P_I typically consists of three components:

$$P_I = P_{\text{pure}} + P_{\text{risk}} + C_{\text{transaction}} \quad (1)$$

where P_{pure} is the pure premium and is the expected cost of the ceded risks (i.e., fair price for the risk coverage), P_{risk} is the risk premium and is determined based on various factors (e.g., insurer’s capital reserve, reinsurance price/availability, and regulatory requirements), and $C_{\text{transaction}}$ is the transaction cost related to marketing, monitoring, and claim settlement. As both P_{risk} and $C_{\text{transaction}}$ are greater than zero and not negligible (particularly for earthquake coverage, P_{risk} can be large, in comparison with P_{pure}), risk-neutral owners, who make decisions based on the expected values without considering variability of the consequences, regard insurance coverage

as overpriced and unfavorable economically. Thus only risk-averse owners, who take into account uncertainties of the consequences, may seek insurance coverage. In cases owners are affected/biased due to their risk perception (e.g., overestimation of potential consequences), earthquake insurance coverage may or may not be attractive for them. Another important aspect to note is that insurance has additional benefits of providing financial liquidity with policyholders in post-disaster situations. For example, a policyholder who has purchased a property with a mortgage may face liquidity constraints after a major earthquake that damages the property severely; the policyholder may not be able to borrow money to recover the loss immediately after the earthquake, because the collateral (i.e., property) is damaged/lost and there is an outstanding mortgage repayment. In such cases, insurance coverage is beneficial in enhancing financial resilience against major shocks.

The policy (payout) function of an insurance contract F_{payout} for the earthquake damage cost C_{EQ} takes the following form:

$$F_{\text{payout}}(C_{EQ}) = \begin{cases} 0 & C_{EQ} \leq D \\ \gamma(C_{EQ} - D) & D < C_{EQ} < L \\ \gamma(L - D) & C_{EQ} \geq L \end{cases} \quad (2)$$

where D , L , and γ are the deductible, limit (cap), and coinsurance factor, respectively. The deductible is the starting loss value of the insurance payout and is usually specified in terms of total insured value. The limit is the maximum loss value of the insurance payout, expressed in terms of total insured value. From insurers' viewpoint, the deductible is useful for reducing transaction costs related to small claims, while the upper limit is effective in avoiding very large claims (critical for their solvency). The coinsurance factor specifies the proportional share of an incurred seismic loss between a policyholder and an insurer; this helps suppress the problem related to moral hazard. For instance, consider that an owner has a house of total value equal to \$200,000 and has purchased earthquake insurance coverage with $D = 10\%$, $L = 80\%$, and $\gamma = 90\%$. In cases where the owner incurs seismic losses of \$60,000 and \$180,000, he/she will receive \$36,000 and \$126,000, respectively.

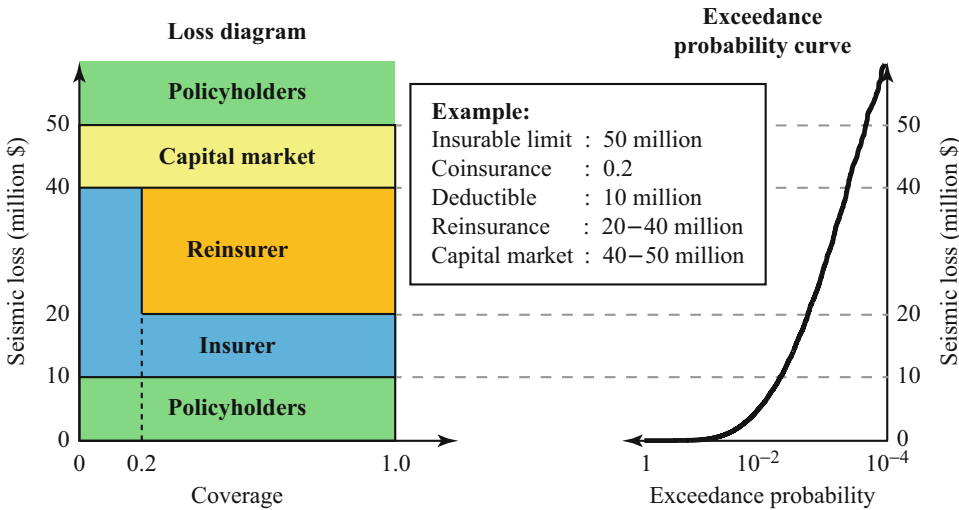
The reinsurance policy has essentially the same structure as primary insurance and is mainly classified into two types: pro rata contract and excess-of-loss contract. In pro rata contracts, an insurer and reinsurer share the risk proportionally (similar to coinsurance), while in excess-of-loss contracts, a reinsurer provides coverage in a specific layer between attachment point and exhaustion point (similar to deductible and limit).

The sales of earthquake risk coverage are often regulated by statutory bodies, such as insurance commissioners. The main reasons are that earthquake insurance, unlike other liability insurance, tends to be *public* (i.e., protection against involuntary risks) and after major earthquake disasters, governments and states may intervene to provide public support to the affected people (e.g., monetary gift and special loan programs). Although, in reality, such disaster relief activities are inevitable (which are funded by tax and borrowing), it is desirable to minimize post-disaster intervention (ex ante subsidies) as this may discourage DRR efforts prior to earthquakes (Kunreuther 1996). This is referred to as charity hazard. It has been demonstrated that DRR

measures are often cost-effective, in comparison with situations without DRR (Kleindorfer et al. 2005; Michel-Kerjan et al. 2013).

Insurers, Reinsurers, Auditors, Regulators, and Capital Markets: The basic principle of insurance portfolio management is the law of large numbers; with the increase in the number of policies, variability of aggregate claims of a portfolio becomes small, and thus actual total loss approaches the expected loss value (which increases with the number of policies). In simple terms, the financial risk becomes more predictable. It is important to recognize that this theory is applicable to *independently distributed events* (e.g., car insurance), while the earthquake risks are both spatially and temporally correlated. Insurers may experience a devastating surge of seismic loss claims in a disaster-hit region and face the possibility of insolvency. Therefore, insurers need to diversify their earthquake exposures geographically.

A conventional approach for catastrophe risk transfer is reinsurance by global reinsurers (Munich Re, Swiss Re, etc.) and governments (Fig. 3). A notable functionality of governments as reinsurer is their ability to borrow money by issuing bonds; this achieves temporal diversification of earthquake risks. Although reinsurers have greater risk-bearing capacities than insurers and achieve geographical risk diversification through their national/global portfolios, the potential size of the catastrophic earthquakes can be significant and their default risk is not zero. Because of this, reinsurers charge quite high risk premium to insurers and seek alternative means for risk transfer (Cummins 2008), which in turn affects the risk premiums for primary insurance coverage. Reinsurers also utilize retrocession, i.e., risk transfer from a reinsurer to another reinsurer to reduce the impact of peak risks. Recently, financial market instruments, so-called insurance-linked securities, become available through capital markets (Grace et al. 2003; Lalonde 2005) and have been used more frequently by insurers, reinsurers, and governments/states (Cummins 2008; see also recent market review articles/reports published by Munich Re and Swiss Re). The emerging tools



Insurance and Reinsurance Models for Earthquake, Fig. 4 Loss diagram for stakeholders, insurer, reinsurer, and capital market

take advantage of much greater financial capacity of capital markets, covering high-loss layers of insurance-reinsurance portfolios. Since earthquake insurance is an integrated part of catastrophe risk management, insurance commissioners, on behalf of governments, regulate the solvency status and insurance policy/pricing of insurers. A high insolvency risk is not acceptable for public earthquake insurance systems. For monitoring the financial status of the earthquake risk underwriters, credit rating agencies (e.g., Standard & Poor's and Moody's) send useful signals to involved parties on their catastrophe exposures. However, the amount of correlations created by retrocession is not transparent to regulators and auditors.

A useful tool for visualizing the composition of catastrophe risk diversification by different stakeholders is the loss diagram (Dong and Grossi 2005). It displays how the total loss is divided and covered by involved parties. Figure 4 shows a hypothetical loss diagram for aggregate seismic loss; in addition, a so-called exceedance probability (EP) curve of the entire portfolio is included. In the diagram, policyholders retain seismic loss up to \$10 million (i.e., deductible for the entire portfolio) and seismic loss

exceeding \$50 million (i.e., insurable limit). The loss layer between \$10 and \$40 million is covered by an insurer and a reinsurer; the first \$10 million layer is solely covered by the primary insurer, while the next \$20 million layer is shared by the excess-of-loss reinsurance with proportional rate of 0.2 (i.e., \$20 and \$40 million are the attachment and exhaustion points of the reinsurance contract, respectively). The high-loss layer between \$40 and \$50 million is ceded to capital markets via CAT bonds/securities.

Existing Insurance Systems Around the World

Earthquake risks are prevalent globally. Countries located near the plate boundaries constantly experience major seismic and volcanic activities. One notable example is the Ring of Fire, or the circum-Pacific seismic belt, which hosts the majority of large earthquakes around the world. The circum-Pacific countries include Chile, Peru, Mexico, the USA (California and Alaska), Japan, Taiwan, the Philippines, Indonesia, and New Zealand, while other major seismic countries are Iceland, India, Iran, Italy, Pakistan, and Turkey. In some of these countries, earthquake insurance systems have been established and operated (NLIRO 2008; OECD 2008). This section

provides a brief summary of existing insurance systems in California (USA), Iceland, Japan, Mexico, New Zealand, Taiwan, and Turkey. Table 1 compares the main features of different insurance systems. In addition, a brief description of natural catastrophe coverage in the EU is included.

California (USA): In California, earthquake insurance coverage had been offered privately since early twentieth century. Currently, the privately funded, state-run California Earthquake Authority (CEA), which was established in 1996 in the aftermath of the 1994 Northridge earthquake, provides earthquake insurance coverage to homeowners; policies can be obtained incidental to fire insurance (optional). Insurer's CEA membership is voluntary and member companies satisfy the mandatory offer law for selling the CEA policy. It is noteworthy that the CEA is not backed up by the State of California. The payable limit of the CEA is about US\$ 100 million; if this amount is exceeded, the claim repayments are proportionally reduced. The policy covers damage and loss due to earthquakes. The premium rates range from 0.036 % to 0.90 %, depending on zip code, structural type, built year, and number of stories. A relatively high deductible of 15 % is specified (10 % can be selected). Policyholders can insure residential properties up to the total insurance value, while they can choose the level of coverage for content (e.g., US\$ 5,000–100,000) and incidental expenses (e.g., US\$ 1,500–15,000).

Iceland: The state-owned Iceland Catastrophe Insurance (ICI), which was established in 1975 after the 1973 Mt. Heimaey volcanic eruptions, provides catastrophe coverage for natural hazards including volcanic eruptions, earthquakes, landslides, snow avalanches, and floods. The participation is compulsory, incidental to fire insurance, and covers direct loss to buildings, content, and lifelines. The deductible is 5 % for each loss. The premium rate for properties and content is 0.025 %, while that for lifelines is 0.02 %. The ICI is liable up to 1 % of the total insured capital; if this amount is exceeded, the claim repayments are proportionally reduced.

Japan: The government-endorsed Japan Earthquake Reinsurance Company (JERC), which was established in 1966 after the 1964 Niigata earthquake, plays a major role in promoting earthquake insurance for residential buildings. The purchase is incidental to fire insurance and is not mandatory (thus the penetration rate is not so high). Primary insurers sell policies and collect fees from policyholders; the covered loss is retained by the JERC, government, and private insurers. The aggregate limit of the indemnity for earthquake insurance is JPY 6.2 trillion; if this amount is exceeded, the claim repayments are proportionally reduced. The policyholder can obtain earthquake coverage up to about 30–50 % of the total insured value under fire insurance policy; in addition, the maximum insured amount is limited to JPY 50 million for buildings and JPY 10 million for movables, respectively. There is a discount mechanism of premiums to encourage seismic risk mitigation; for instance, properties with base isolation are granted for 30 % discount, and certified properties which meet the current seismic design requirements are granted for 10 % discount. Other cases include “buildings constructed after 1981 (when the new seismic design code was implemented in Japan),” and “certified buildings according to the seismic grading system.”

Mexico: Earthquake insurance in Mexico is mainly provided by private insurers, under the regulation by the Comision Nacional de Seguros y Fianzas (CNSF). The major seismic event that led to adoption of earthquake insurance was the 1985 Michoacan earthquake. The purchase is optional and is often accompanied by fire insurance. The insurance covers damage and loss of buildings and movables due to earthquakes and volcanic eruptions. The premium rates vary widely, depending on occupancy type, number of stories, and geographical region. Typical values of deductible are 2–5 %. In addition to private earthquake insurance, the government-supported Natural Disasters' Fund, which was created in 1996, provides financial support to the recovery/restoration of impaired infrastructure and to the disaster victims.

Insurance and Reinsurance Models for Earthquake, Table 1 Comparison of earthquake insurance systems in different countries (NLIRO 2008; OECD 2008)

Country	Insurance commissioner	Participation method	Coverage	Rate	Insured amount and limit	Payable claim amount	Role of primary insurers
California (USA)	California Earthquake Authority (CEA)	Incidental to fire insurance (optional)	Residential building and content; damage due to earthquakes	0.036–0.90 %; 19 classes (zip code); 8 classes (structural type and built year); 2 classes (story number)	Building: total insurance value; content, US\$5,000–100,000; incidental expense, US\$ 1,500–15,000; deductible, 15 %	US\$ 96 million per event	Sales, fee collection, certificate issuance, remittance, and assessment
Iceland	Iceland Catastrophe Insurance (ICI)	Incidental to fire insurance (compulsory)	Building, content, and lifelines; damage due to earthquakes, volcanic eruptions, landslides, avalanches, and floods	0.025 % for buildings and content; 0.02 % for lifelines	Deductible, 5 % of each loss (minimum deductible – ISK 85,000 for building, ISK 20,000 for content, and ISK 750,000 for lifelines)	1 % of the total insured capital	Sales, fee collection, and certificate issuance
Japan	Japan Earthquake Reinsurance Company (JERC)	Incidental to fire insurance (optional)	Residential building and movables for living; damage due to earthquakes, volcanic eruptions, and tsunamis	0.05–0.313 %; 8 classes (structural type and geographical region)	30–50 % of the total insured value under fire insurance policy (maximum insured amount is up to JPY 50 million for buildings and JPY 10 million for movables); deductible, 3 % (building) and 10 % (movables)	JPY 6.2 trillion	Sales, fee collection, certificate issuance, damage assessment, and insurance underwriting

Mexico	Private insurers regulated by Comision Nacional de Seguros y Fianzas (CNSF)	Optional	Residential building, commercial/ industrial building, and movables for living; damage due to earthquakes and volcanic eruptions	0.018–0.356 % for residential building; 12 classes (geographical region); 0.028–0.726 % for commercial/industrial building; 12 classes (geographical region) and 2 classes (story number)	Total insurance value (but some upper limits [70–90 %] may be applied in high-seismic regions); deductible, 2–5 % (depending on geographical region)	N/A	Sales, fee collection, certificate issuance, damage assessment, and insurance and underwriting
New Zealand	Earthquake Commission (EQC) – EQCover	Incidental to fire insurance (compulsory)	Residential building, content, and land; damage due to earthquakes, landslides, volcanic eruptions, and tsunamis	0.05 % for residential building	Building, NZ\$ 100,000; content, NZ\$ 20,000; land, total insurance value Deductible, 1 % (building), NZ\$ 200 (content), and 10 % (land, maximum deductible is set to NZ\$ 5,000)	All claims to be repaid (backed up by the government)	Sales and fee collection
Taiwan	Taiwan Residential Earthquake Insurance Fund (TREIF)	Incidental to fire insurance (compulsory)	Residential building; damage due to earthquakes	0.122 %	Building, TW\$ 1.2 million; contingent expenses, TW\$ 180,000; no deductible	TW\$ 60 billion	Sales
Turkey	Turkish Catastrophe Insurance Pool (TCIP)	Compulsory	Building; damage due to earthquakes	0.044–0.55 %; 15 classes (structural type and geographical region)	TRY110,000 (in Turkish lira [TRY]); deductible, 2 %	EUR 1 billion (from reinsurance)	Sales, fee collection, certificate issuance, and damage assessment

New Zealand: The government-owned Earthquake Commission (EQC) provides natural disaster insurance (EQCover) to homeowners in New Zealand. The EQC was established in 1944 after the devastating 1942 Wairarapa earthquakes. The participation is incidental to fire insurance and is automatic/compulsory. The contract covers damage and loss of residential properties, content, and land due to earthquakes, landslides, volcanic eruptions, and tsunamis. The premium rate is 0.05 %. The insured amount is limited to NZ\$ 100,000 for buildings and NZ\$20,000 for content, respectively. The deductible is set to NZ\$ 200 or 1 % of the loss for residential properties, NZ\$200 for content, and the greater of NZ\$ 500 or 10 % of the insured loss for land, with the maximum deductible of NZ\$ 5,000. In addition to the EQCover, private insurers offer natural disaster damage extension, which can insure commercial buildings (variable rates depending on geographical region, structural type, built year, and number of stories).

Taiwan: The government-endorsed Taiwan Residential Earthquake Insurance Fund (TREIF) operates the earthquake insurance program in Taiwan. TREIF was established in 2002 in the aftermath of the 1999 Chi-Chi earthquake. The policy is incidental to fire insurance and its purchase is compulsory. The insurance covers damage and loss of residential properties due to earthquakes. The upper limit of the repayment per household is TW\$ 1.2 million for residential properties and TW\$ 180,000 for contingent expenses, respectively. The premium rate is 0.122 %. The total amount of the payable claims is TW\$ 60 billion.

Turkey: The state-owned Turkish Catastrophe Insurance Pool (TCIP) was created after the 1999 Izmit earthquake to reduce catastrophe earthquake exposure to the government. The owners of residential properties within municipality boundaries are obliged to take earthquake insurance coverage, which insures damage and loss due to earthquakes. The premium rates range from 0.044 % to 0.55 %, depending on structural type and geographical region. The maximum

insured limit is TRY 110,000 and the deductible is 2 %. Currently, as the financial status of the TCIP is not sufficient, it cedes a large amount of its risks to international reinsurers. Note that commercial and industrial properties as well as residential properties in rural areas can be insured on a voluntary basis.

EU Countries: Porrini and Schwarze (2012) classify variable insurance schemes for natural catastrophe coverage (including earthquakes for some countries) in EU member states into five categories: (1) regional public monopoly insurer of natural hazards guided by statutory provisions and public consultation processes; (2) compulsory insurance for all natural hazards; (3) compulsory inclusion of natural hazards into general house ownership insurance; (4) free-market natural hazard insurance with ad hoc governmental relief programs; and (5) taxpayer-financed governmental relief funds. The Netherlands, Denmark, and Switzerland have adopted Schemes 1 and 2; France and Spain have used Schemes 2 and 3; the UK has adopted Schemes 3 and 4; Germany has followed Scheme 4; and Poland, Italy, and Austria have adopted Schemes 4 and 5. These schemes suppress insurance-related problems, such as adverse selection, moral hazard, and charity hazard, differently. For instance, Scheme 1 can be used to mitigate all three issues, whereas Schemes 2 and 3 can deal with adverse selection and charity hazard effectively but may be susceptible to moral hazard. Effectiveness of Schemes 4 and 5 is limited to avoiding charity hazard and adverse selection, respectively. It is noteworthy that the selection of a suitable scheme has important influence on how risk mitigation and (financial) risk management are implemented. Furthermore, insurance penetration rates for earthquake coverage in Europe are unevenly distributed among member states of the EU (Maccaferri et al. 2011). Notably, many earthquake-prone countries (e.g., Italy, Greece, Romania, Spain, and Portugal) show rates below 10 %, although high maximum losses within the past 20 years have been experienced (e.g., Greece, 2.1 % in 2010 GDP value, and Italy, 0.6 % in 2010 GDP value).

Challenges in Insuring Low-Probability High-Consequence Risks

Risk perception has major influence on decision-making processes regarding the adoption of earthquake insurance (Kunreuther 1996). Facing LP-HC risks, home and business owners may have diverse opinions on the potential magnitude of earthquake consequences. Therefore, appreciation of benefits of earthquake insurance coverage varies significantly, depending on personal characteristics (e.g., optimistic or pessimistic), proximity to physical hazards (e.g., living near a major fault), previous experience with earthquakes, and socioeconomic/demographic profiles, such as age, gender, income, and education (Palm 1995). Another important aspect influencing insurance-related decisions is cognitive limitation of human beings (Kahneman 2003). It is well known that people are affected by framing (Johnson et al. 1993) and have difficulties in estimating very low probabilities (Viscusi and Evans 2006). An example of framing is that a factory owner who is considering earthquake insurance for financial protection is likely to take insurance if he/she is told that the chance of experiencing major earthquake damage is 1 in 5 over the entire period (20 years), rather than the chance of such an event is 1 in 100 per year. Therefore, even when the overall premium is reasonably priced, not many stakeholders voluntarily purchase earthquake risk coverage.

Two known issues in offering insurance coverage are moral hazard and adverse selection (note: they are applicable to not only catastrophe earthquake but other more common types). Moral hazard refers to a case where more risky situations (i.e., increase in chances of experiencing loss and/or increase in the extent of loss) are created by the behavior of a stakeholder. An example of moral hazard is that a policyholder, after a damaging earthquake, may cause additional damage to his/her property to receive insurance repayment (when the incurred loss is relatively minor and is near the threshold of deductible). On the other hand, adverse selection refers to a situation where an insurer cannot distinguish between stakeholders having lower and

higher earthquake risk potential. In such cases, the insurer cannot charge appropriate insurance premiums that are commensurate with the risks. For instance, if an insurer charges a flat rate for all homeowners in a specific geographical region, earthquake insurance is more attractive for those who live in seismically more vulnerable houses (e.g., non-ductile old construction). This leads to higher-risk exposure for the insurer than it is originally expected and may result in increasing the premium rate (i.e., adverse selection spiral can occur). One of the main causes of moral hazard and adverse selection is the information asymmetry between an insurer and a policyholder. The private information that the stakeholder has on his/her property is not entirely accessible to the insurer; in such cases, the stakeholder may have an incentive to behave inappropriately from the insurer's perspective.

To alleviate the effects due to moral hazard and adverse selection, introducing pro rata share of the risks and differentiating the rates according to physical parameters related to seismic hazard and vulnerability of the covered properties (e.g., geographical region, structural type, built year, and number of stories) are useful (see Table 1). Other proactive ways to avoid adverse selection and encourage physical DRR measures are to implement an effective incentive scheme in determining insurance premium rates (e.g., Japanese insurance system).

Insurance and Reinsurance Models

This section gives an introduction to a wide range of earthquake insurance and reinsurance models that have been developed and applied in different academic fields (Fig. 2). The “[Earthquake Catastrophe Model](#)” section provides a brief summary of earthquake CAT modeling and quantitative risk metrics. The key requirements of viable CAT models are mentioned. Next, actuarial/financial insurance models are presented in the “[Actuarial and Financial Insurance Model](#)” section. Subsequently, economic insurance models are summarized in the “[Economic Insurance](#)

Model” section to give an overview of the decision-making processes for policyholders (demand) and for insurers (supply).

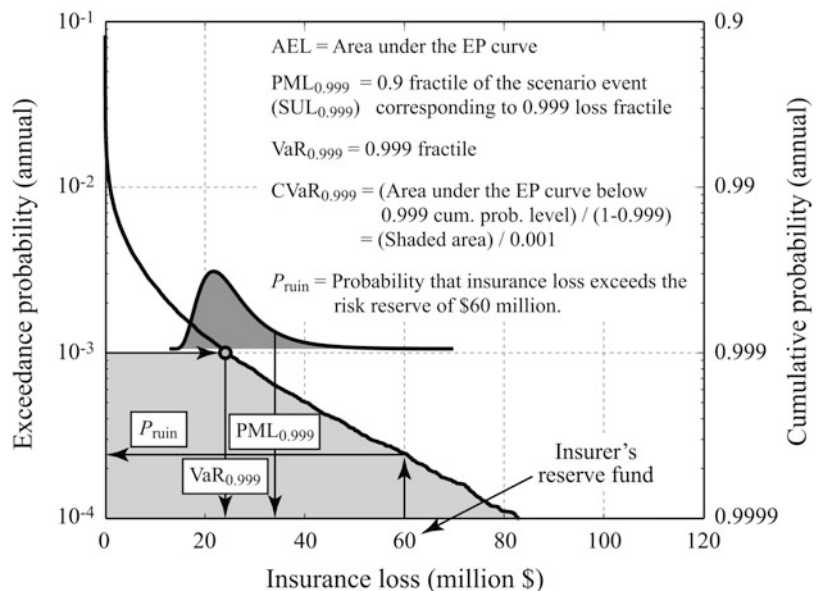
Earthquake Catastrophe Model

Earthquake CAT modeling is an integrated process of conducting numerical simulations of earthquake occurrence, ground motion prediction, damage assessment, and seismic loss calculation. It typically involves: (i) inventory/exposure database, (ii) hazard characterization, (iii) structural vulnerability assessment, and (iv) loss estimation and insurance portfolio analysis. The output for the CAT models is the probability distribution of estimated seismic loss for a building/infrastructure portfolio (i.e., EP curve). There are a variety of different methods and models used for earthquake insurance risk analysis in research and in practice (e.g., Bommer et al. 2002; Asprone et al. 2013; Yucemen 2013). Current challenges of the CAT modeling include the validation of the CAT models using scarce loss data and dealing with rapidly changing exposure data (particularly for developing countries) and the incorporation of multiple concurrent hazards and risks, such as tsunamis, aftershock, liquefaction, landslide, and fire, within a unified analytical/computational framework.

A stochastic earthquake catalog method (e.g., Goda and Yoshikawa 2012) starts with a regional probabilistic seismic hazard model and generates a synthetic earthquake catalog using Monte Carlo sampling techniques. For each earthquake in a catalog, seismic vulnerability assessment and loss estimation are carried out using fragility curves and damage-loss functions. A fragility curve evaluates conditional probability of exceeding a certain damage state (e.g., slight, moderate, and extensive), and a damage-loss function transforms damage severity into corresponding seismic loss. It is noted that uncertainty associated with every step of the computation is taken into account and is propagated. The insurance-related risk can be evaluated by applying an insurance policy function to the estimated seismic loss (either at policy level or at aggregate level). The estimated loss results can be sorted in an ascending order to develop an EP curve for the portfolio of interest. An illustrative EP curve for an insurance portfolio, which plots estimated insurance loss as a function of annual exceedance probability, is shown in Fig. 5. The EP curve is often used for insurance portfolio decision-making. Note that there are many variants of the abovementioned method, such as a multiple-event/event-loss-table method (e.g., Dong and Grossi 2005).

Insurance and Reinsurance Models for Earthquake,

Fig. 5 Exceedance probability curve and risk metrics



Insurance and Reinsurance Models for Earthquake, Table 2 Summary of risk metrics

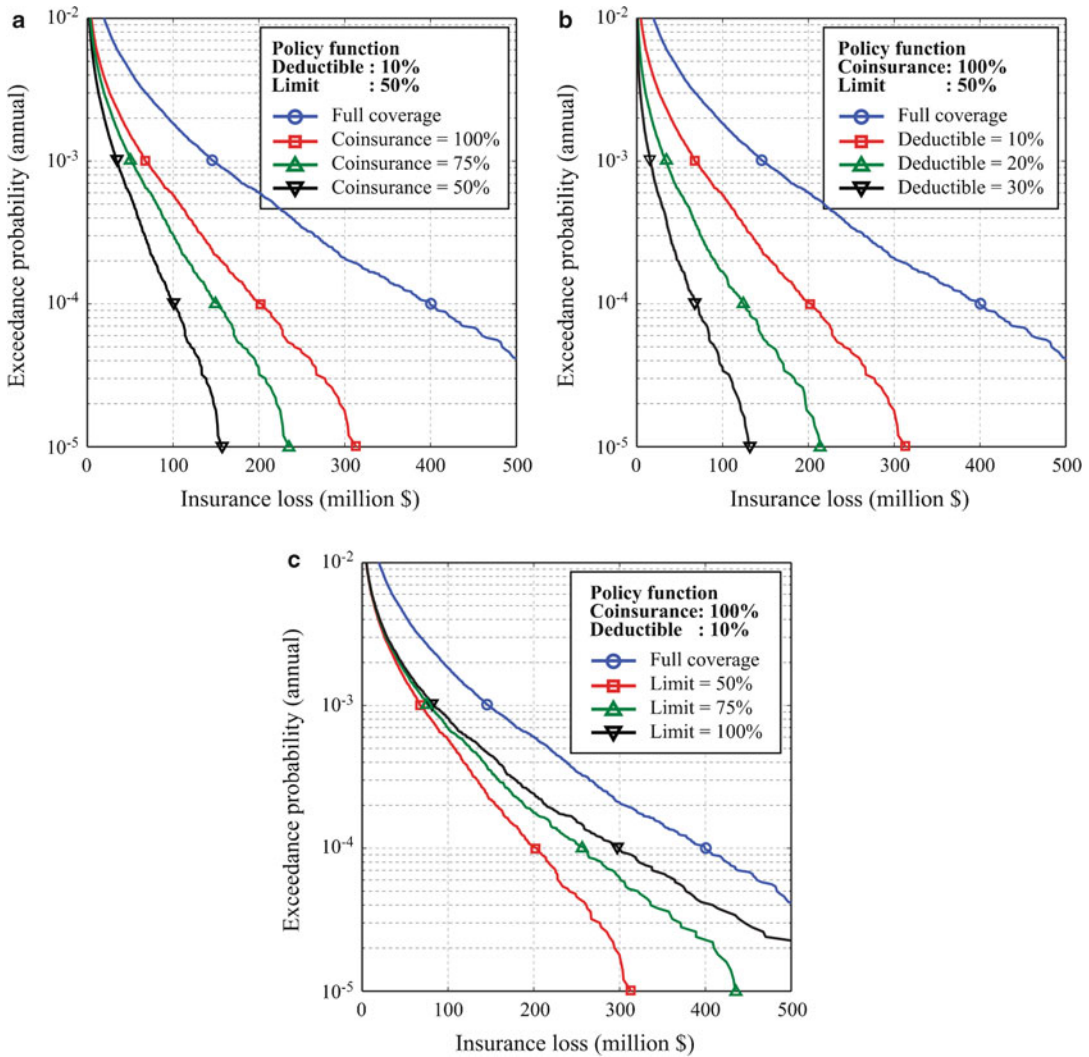
Risk metrics	Equation	Relationship with the EP curve
Annual expected loss: AEL	$AEI = E[L]$ where $E[\bullet]$ is the expectation	Area under the EP curve
Value at risk: Var_α	$Var_\alpha = \inf\{l : P(L > l) \leq 1 - \alpha\}$ where $\inf\{\bullet\}$ is the infimum and $P(L > l)$ is the probability of loss L exceeding a certain loss level l	Fractile value at a selected cumulative probability level α
Conditional value at risk: $CVaR_\alpha$	$CVaR_\alpha = \frac{1}{1-\alpha} \int_\alpha^1 Var_u du$	Area under the EP curve below a selected cumulative probability level α , normalized by the exceedance probability $1-\alpha$
Ruin probability	$P_{\text{ruin}} = P(L > R_{\text{insurer}})$ where R_{insurer} is the insurer's risk reserve	Probability when insurer's reserve fund is depleted

Note: α represents the cumulative probability; in other words, $1-\alpha$ represents the exceedance probability

To facilitate the risk-based decision-making and risk communication among different stakeholders, risk metrics can be derived from the calculated EP curve. Risk metrics that are widely used include: annual expected loss (AEL), probable maximum loss (PML), value at risk (VaR), conditional value at risk (CVaR), and ruin probability. The equations to compute these measures (except for PML) are summarized in Table 2 (see also Fig. 5 for graphical representation of the metrics). The AEL is a useful metric for ordinary risks; however, it is not suitable for catastrophic risks as it fails to capture the extent of devastating consequences due to rare events. The VaR is the fractile value on an EP curve corresponding to a selected probability level, while the CVaR, which is also referred to as expected shortfall in finance, accounts for rare events in terms of their severity and frequency by taking the conditional expectation of the EP curve. One of the crucial factors is “which probability level” to be focused upon in evaluating VaR and CVaR (also applicable to PML). The ruin probability represents the chance of insurer's insolvency (i.e., all reserve funds are depleted) and is often used for insurance regulatory purposes (e.g., an insurer needs to retain a sufficient reserve fund such that annual probability of insolvency is less than 0.005). The PML is one of the most popular metrics in financial risk management, and there are several definitions. Conventionally, it was defined as the fractile of the loss corresponding to the return period of 475 years (i.e., essentially same as VaR).

However, in different industries and countries, variants of the PML had been adopted for use. For example, in Japan, the PML is defined as the (conditional) 0.9-fractile value for a scenario that corresponds to a selected probability level (typically, return period of 475 years is considered, but this can be varied). Currently, specific nomenclatures are in use: scenario expected loss (SEL) corresponds to the original PML definition (i.e., VaR), and scenario upper loss (SUL) corresponds to the PML definition in Japan.

The CAT modeling is useful for quantifying insurers'/reinsurers' earthquake risk exposure under different situations and for deciding risk management strategies. For instance, impact of the policy function parameters (i.e., coinsurance, deductible, and limit) on insurer's earthquake risk exposure is illustrated in Fig. 6. The insurance portfolio is based on 4,000 wooden houses in Vancouver, Canada, and the EP curve is obtained from the stochastic earthquake catalog method (Goda and Yoshikawa 2012). Figure 6 shows that a high deductible and low coinsurance factor reduces earthquake risk exposure significantly at wide probability levels, whereas a low limit curtails the maximum earthquake risk exposure effectively. The limit tapers heavy right tail of the loss distribution and affects the deficit at ruin, which is one of the key decision variables in earthquake insurance management. Importantly, from insurer's viewpoint, flexibility to select a suitable policy arrangement is beneficial to enhance the performance of earthquake portfolio



Insurance and Reinsurance Models for Earthquake, Fig. 6 Effects of policy function parameters on exceedance probability curve: (a) coinsurance, (b) deductible, and (c) limit

management and to offer affordable insurance rates to the policyholders. However, such flexibility is not usually possible due to regulatory rules; in such cases, ART tools can be used.

Actuarial and Financial Insurance Model

Insurance models in actuarial sciences and financial engineering have a long history of development and wide applications in assessing underwriting risks. Popular approaches for characterizing catastrophic risk processes include a compound Poisson process (e.g., Rolski

et al. 1999) and an extreme value theory (e.g., McNeil et al. 2005). The risk process can be formulated in form of stochastic differential equations. Although analytical solutions for insurer's ruin probability are available for simple cases (e.g., compound Poisson process with exponential insurance loss), for realistic and complex problems, such solutions are not available and numerical simulations of the risk process are often required. On the other hand, statistical modeling of dependent risk processes is useful for assessing insurer's risk exposure due to

extreme events. In particular, accurate modeling of exposure dependence among different portfolios and lines of business is important in evaluating the entire risk process. This leads to advancement of dynamic financial analysis (Kaufmann et al. 2001), which integrates large-scale stochastic simulations of all related insurance and reinsurance processes (including fluctuating interest rates and catastrophic as well as non-catastrophic risks). The key challenge for both approaches is how to incorporate spatiotemporally correlated insurance claims in the risk process. In this section, two insurance models based on a compound Poisson (including diffusion-jump) process and based on extreme value theory are briefly mentioned (note: rigorous mathematical formulations are beyond the scope of this entry).

Compound Poisson and Diffusion-Jump Processes: The fluctuation of an insurer's wealth $W(t)$ subjected to both catastrophe and non-catastrophe risks, $W_{\text{noncat}}(t)$ and $W_{\text{cat}}(t)$, can be expressed as

$$W(t) = W_{\text{noncat}}(t) + W_{\text{cat}}(t) \quad (3)$$

In the context of earthquake insurance, $W_{\text{noncat}}(t)$ is related to ordinary risks (e.g., diversifiable liability insurance underwriting) and can be represented by a stochastic growth model, such as Brownian process and geometric Brownian process. The Brownian process assumes that changes of the wealth state over a short period are independent and identically distributed (*i.i.d.*) as normal variate, whereas the geometric Brownian process considers that the logarithmic ratio of two wealth states separated by short time lag is an *i.i.d.* normal variate. For instance, by adopting the geometric Brownian process, $W_{\text{noncat}}(t)$, representing incurred loss process, payout process, expense process, and investment process, can be described as

$$W_{\text{noncat}}(t) = W_0 \exp[(\alpha - 0.5\beta^2)t + \beta Z(t)] \quad (4)$$

where W_0 is the initial asset; α and β are the instantaneous growth rate and infinitesimal

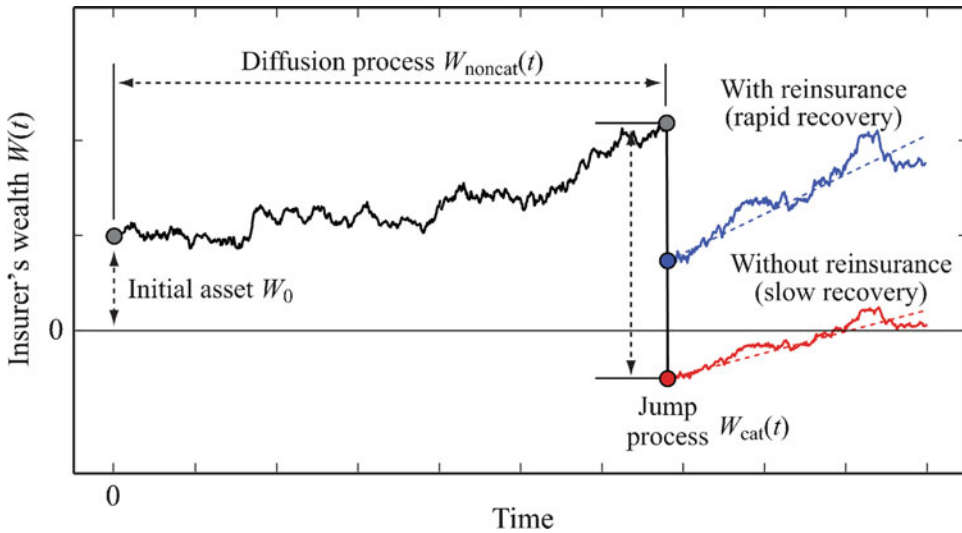
standard deviation (or volatility) of the insurer's wealth related to non-catastrophic risks, respectively; and $Z(t)$ is a standard Brownian motion due to perturbation of non-catastrophic risk processes. In practical applications, the drift and volatility parameters α and β can be estimated from relevant insurance data (see Goda and Yoshikawa 2012). The stochastic process shown in Eq. 4 can be evaluated using Monte Carlo simulation, and thus many realizations can be generated to assess the temporal evolution of the insurer's asset.

The catastrophic jump component of the risk process $W_{\text{cat}}(t)$ occurs infrequently, but when it occurs, a drop of the insurer's net worth can be extremely large. Potentially, such a downward jump causes a negative net worth, forcing an insurer to become insolvent (i.e., $W(t) < 0$). This can be mathematically represented by a compound Poisson jump process:

$$W_{\text{cat}}(t) = - \sum_{i=1}^{N(t)} L_i \quad (5)$$

where jumps (i.e., damaging earthquakes) arrive randomly according to a Poisson counting process $N(t)$ and the random size of the jumps is characterized by a probabilistic loss model. Cossette et al. (2003) presented three different formulations of the catastrophic loss model L in Eq. 5 by considering independent, fully correlated, and partially correlated cases. The degree of loss correlation from multiple policies was defined as a function of seismic intensity, which is realistic. In addition, they derived the risk-ordering relationships of the different types of portfolios with respect to loss correlation for several risk metrics, including VaR and CVaR. It is noteworthy that the jump process shown in Eq. 5 can be substituted by an earthquake CAT model, which is more rigorous and accurate than a compound Poisson process.

The abovementioned stochastic representation of insurer's wealth process is illustrated in Fig. 7. Due to the earthquake's strike, insurer's wealth state is affected by earthquake insurance claims; this may force the insurer to become insolvent or



Insurance and Reinsurance Models for Earthquake, Fig. 7 Stochastic wealth process with and without insurance

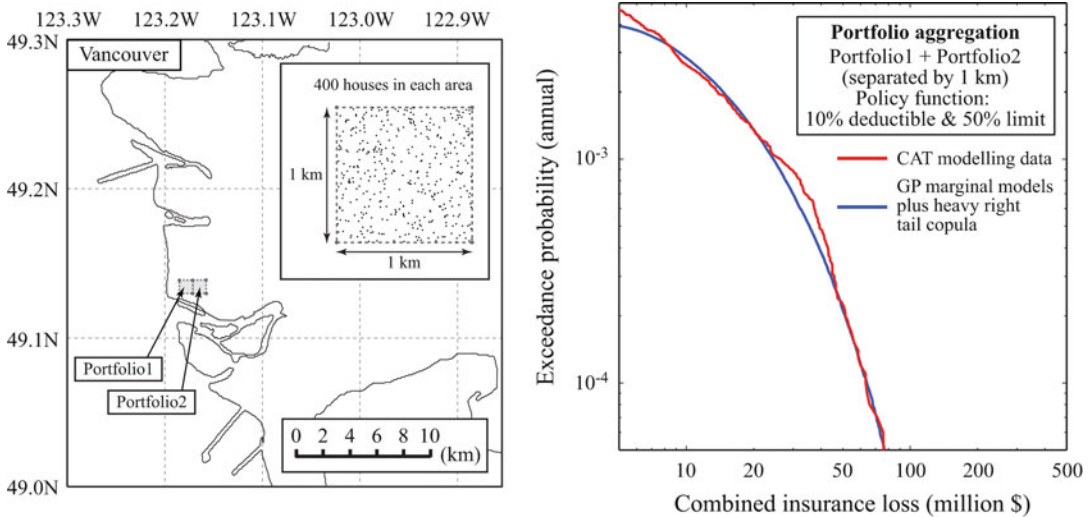
be in debt. Furthermore, financial constraints caused by the earthquake risk may result in slow recovery from the shock after the earthquake. The figure also includes a hypothetical case where the insurer has arranged reinsurance coverage for the loss; in this case, the wealth state is dropped but not to the level of insolvency, and the post-disaster recovery rate may be more rapid than the situation without reinsurance (e.g., liquidity constraints are avoided). Stochastic simulations of underwriters' risk processes are useful for investigating the effects of insurance policies and reinsurance on the portfolio and for developing effective risk management strategies.

Extreme Value Theory and Copula-Based Dependence Modeling: The statistical modeling of extreme earthquake events is one of the active research areas. Useful statistical models for capturing the key features of a highly correlated loss generation process are the extreme value distributions (McNeil et al. 2005). In particular, the peak-over-threshold (POT) model is suitable as it utilizes all data exceeding a high threshold value more efficiently than the block maxima model and is more applicable to insurance loss processes (focusing on large, rare loss events). The limiting distribution of the POT data Y that exceeds a sufficiently high threshold μ converges to the generalized Pareto (GP) distribution:

$$P(X \leq x | X > \mu) = \begin{cases} 1 - (1 + \xi[x - \mu]/\beta)^{-1/\xi} & \xi \neq 0 \\ 1 - \exp(-[x - \mu]/\beta) & \xi = 0 \end{cases} \quad (6)$$

where ξ is the shape parameter and determines the upper tail behavior of the data and β is the scale parameter. Usually, a suitable value of μ can be determined based on mean excess and shape parameter plots (McNeil et al. 2005), and given data μ , it is straightforward to estimate the parameters of the GP model based on the maximum likelihood method.

In insurance portfolio management, multivariate modeling of insurance data may be required (e.g., portfolio aggregation; if locations of two portfolios are geographically close, they are subjected to similar but not identical earthquake risks). In such a case, a dependence structure of the data needs to be characterized. To deal with insurance data having heavy right-tail and



Insurance and Reinsurance Models for Earthquake, Fig. 8 Statistical aggregation of two insurance portfolios

nonlinear dependence, a copula technique can be used (McNeil et al. 2005); the major advantage of the copula is that the dependence modeling can be carried out in isolation from the marginal distribution modeling. The copula model fitting can be performed by finding a parametric copula function that fits best to empirical copula data (transformed data such that their marginal distributions are uniformly distributed) using the maximum likelihood method. There is a wide class of copula functions (e.g., elliptical copula and Archimedean copula).

To illustrate an application of the statistical modeling approach to earthquake insurance loss, results for aggregation of two insurance portfolios at closely located locations are presented in Fig. 8 (Goda and Ren 2013). Two portfolios, each consisting of 400 houses in a 1 km by 1 km area, are considered; the two portfolios are separated by 1 km, and thus high correlation is anticipated, particularly in the upper tail (i.e., large loss in one portfolio is likely to be accompanied by large loss in the other portfolio). The insurance loss samples are generated from the stochastic earthquake catalog approach. To model the joint seismic loss distribution of the two portfolios, GP models are fitted to the marginal distributions of the two portfolios, and a suitable copula function, which is found to

be the heavy right-tail copula, is obtained. Then, using the constructed joint probability distribution, insurance loss samples are generated and plotted in Fig. 8. Comparison of the two EP curves indicates that, except for the intermediate loss range, agreement of the combined insurance loss data is good; see Goda and Ren (2013) for more comprehensive results. It is emphasized that the simulation from the statistical model is very quick, in comparison with the CAT model. Therefore, once the accuracy of the statistical method is verified, the proposed method can serve as a viable alternative to conduct various sensitivity analyses related to earthquake insurance portfolios (i.e., comparison of EP curves and risk metrics under different policy functions and reinsurance arrangements).

Economic Insurance Model

In economics, decision problems of providing insurance coverage have been studied extensively from two perspectives: (i) homeowner's demand (Mossin 1968; Schlesinger 2000) and (ii) insurer's supply (Kleindorfer et al. 2005). Other relevant problems include economic valuation and decision-making related to risk mitigation actions, such as self-insurance, self-protection, and market insurance (Ehrlich and Becker 1972). The essential tools that have been

used in such analyses are based on the expected utility framework (note: recent studies adopt descriptive decision theories to analyze the abovementioned problems). Typically, a homeowner is assumed to be a risk-averse decision maker, having a concave utility function, whereas an insurer is considered to be a risk-neutral decision maker, having a linear utility function. To simplify situations, two contingent states only, i.e., no earthquake damage and full earthquake damage, are often considered. The conclusions and theorems derived from these idealized analyses provide useful insight on rational behavior of homeowners and insurers toward insurance decision-making under uncertainty. It is important to recognize that these theoretical predictions are obtained based on many (critical) assumptions about risk attitudes of decision makers and market conditions, which may be invalid in reality and inapplicable to catastrophic earthquake risks. In the following, brief summaries of insurance decision-making models for the two cases are given.

Insurance Demand Decisions by Homeowners: A simple form of homeowner's demand problem for earthquake coverage is the following: (i) a homeowner has a utility function U (typically, concave function); (ii) earthquake damage can occur with probability of π (note: there are two contingent states only), and should this happen, he/she will lose a property valued at L ; (iii) the owner has an additional asset valued at A ; (iv) the owner pays a premium P_I . Under these assumptions, the expected utility without insurance can be expressed as $\pi U(A) + (1 - \pi)U(A + L)$, while that with (full) insurance is $U(A + L - P_I)$. The maximum insurance premium that the owner is willing to pay can be obtained by setting the derivative of the objective function:

$$O = \pi U(A) + (1 - \pi)U(A + L) - U(A + L - P_I) \quad (7)$$

with respect to an insurance parameter equal to zero and by finding a solution of the equation. A similar setup can be used for more elaborated cases to analyze the effects due to asset, policy function structure (e.g., coinsurance/deductible/

limit), risk attitude (utility function and degree of risk aversion), and risk premium (Schlesinger 2000). Recently, Asprone et al. (2013) have applied an extended version of the insurance demand model by considering multiple damage/loss levels to determine insurance premiums of building stocks in Italy.

Insurance Supply Decisions by Insurers: The supply of earthquake coverage is influenced by various factors. Insurer's main concerns in offering earthquake coverage are profitability and fear of insolvency. This insurer's decision problem can be formulated as the profit maximization subjected to survival and stability constraints. The survival criterion requires that ruin probability should be less than a threshold value (which may be given by a regulatory body exogenously), whereas the stability criterion requires that probability of operational costs (i.e., sum of claim payments and expenses) exceeding a certain threshold value should be sufficiently small. Each of these criteria produces the maximum number of policies that can be offered for a given policy-premium structure. If the maximum number of policies based on the two constraints does not exceed the minimum number of policies required to initiate such coverage, the insurer will not offer earthquake insurance. Although the above economic analysis highlights a rational basis for insurer's supply decisions, insurers tend to adopt heuristic criteria based on simpler risk metrics, such as PML and VaR.

Integrated Earthquake Risk Management

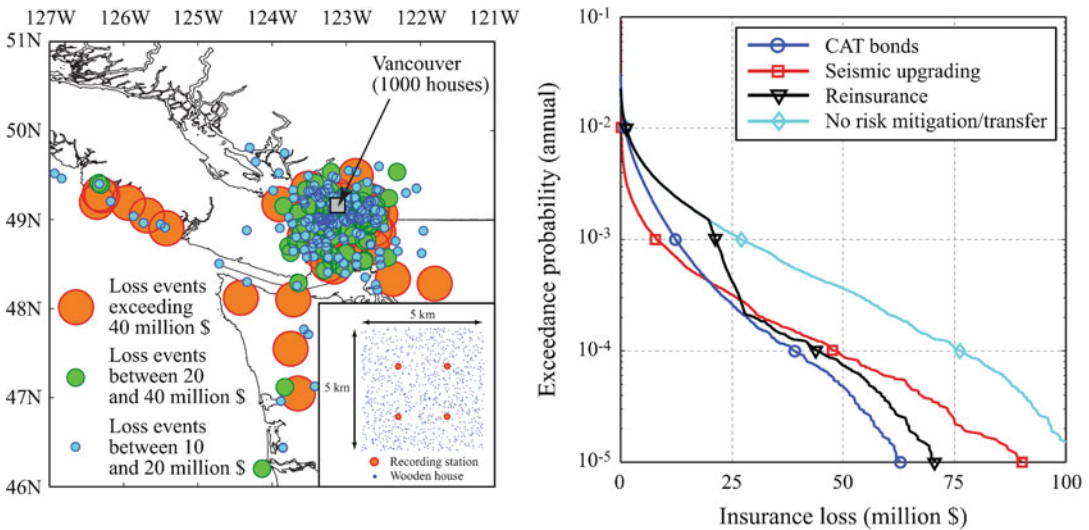
Recent catastrophes in Chile and Japan have revealed insufficiency of conventional financial risk management tools and systems, such as self-insurance, public support, and earthquake insurance. Muir-Wood (2011) suggested that for Chile, micro-insurance, targeted for low-income homeowners and small commercial enterprises, should be introduced and a new risk-pooling system should be set up to cover earthquake risks for government-owned buildings. The former issue is compounded by anticipation of ex post

subsidies from governments by Chilean people (charity hazard); however, it has been demonstrated that DRR measures are far more cost-effective than no DRR situation (Kleindorfer et al. 2005; Michel-Kerjan et al. 2013). In Japan, relatively low penetration rates of earthquake insurance (about 27 % in 2013) should be increased to achieve a nationwide risk-sharing mechanism with stable incomes; this situation was contrasted by the earthquake insurance coverage in New Zealand after the 2010–2011 Christchurch sequences. It is also important to implement risk-based insurance premiums for different structural types and seismic regions (e.g., California, Japan, and Turkey) and introduce financial incentive schemes to promote DRR investments (e.g., Japan). Such proactive earthquake risk management is the key to achieve sustainable and resilient communities against earthquake disasters.

In the envisaged risk management framework, the role of insurers, reinsurers, and governments is important. Major challenges in the current earthquake risk finance systems are related to securing sufficient funds to cover potential earthquake loss. The development of catastrophe securitization markets enables earthquake risk underwriters to have access to greater financial resources (Grace et al. 2003; Lalonde 2005; Cummins 2008). Broadly, there are three types of catastrophe insurance securities: indemnity-based securitization, index-based securitization, and parametric securitization. The indemnity-based securities are essentially the same as reinsurance contracts (but offered in form of securities), for which the loss repayment is determined based on actual security-issuer's (sponsor's) loss. Thus it suffers from the same issues as reinsurance, such as moral hazard and asymmetrical information. The index-based securities adopt actual industry-wide loss due to a damaging earthquake, rather than actual sponsor's loss. Advantages of this type of securitization include: reduced effects of moral hazard and transparency of loss information (note: individual companies are generally reluctant to reveal detailed information on their risk exposures). Difficulties arise in pricing such securities because industry loss and

company's loss are not fully correlated. In other words, the index-based securities are subjected to basis risk.

Recently, parametric securities, such as CAT bonds, have become popular. A typical setup for the CAT bonds is as follows. A single-purpose reinsurer collects funds (principal) from investors in the markets and issues CAT bonds. In cases pre-agreed/specified trigger conditions are met, the principal is paid to the sponsor; otherwise, the principal together with return higher than typical securities is paid back to the investors when the bonds mature. Triggering of the bond payments is determined based on the intensity of physical earthquake parameters, such as a combination of earthquake magnitude and location and instrumentally obtained seismic intensity measures. As estimates of the earthquake intensity parameters are promptly available from reliable third bodies (e.g., US Geological Survey and Japan Meteorological Agency) immediately after an event, bond payments become available soon after the earthquake. This is not the case for the other two types of securities; assessment and settlement of earthquake loss take a long time. Generally, in comparison with reinsurance, the CAT bonds have three main advantages: (i) relatively large total value of bonds can be issued with low default risk, because the capital markets have significantly greater risk-bearing capacity compared with reinsurance industry, (ii) principal is accessible immediately after a disaster (high liquidity), and (3) moral hazard is minimized. A main disadvantage of the CAT bonds is basis risk, i.e., discrepancy between actual loss incurred by the sponsor and payment received by the sponsor (similar to index-based securities). It consists of model risk and trigger risk. Model risk arises because the structure of the CAT bonds is often designed based on numerical CAT models, whereas trigger risk is related to inaccuracy of an adopted trigger mechanism in terms of physical seismic parameters. The basis risk can be reduced by adopting an improved version of the CAT model and by implementing an advanced trigger mechanism that uses more local seismic information (e.g., instrumental measures from nearby recording stations, rather



Insurance and Reinsurance Models for Earthquake, Fig. 9 Insurance portfolio management using seismic upgrading, reinsurance, and CAT bonds

than earthquake event characteristics; Goda (2014). In this regard, progress in the CAT modeling, contributed by all involved disciplines, has direct impact on the reliability of the CAT bonds.

Lastly, to illustrate a practical application of CAT bonds in insurance portfolio management, EP curves for four cases are presented in Fig. 9: (i) no risk mitigation/transfer; (ii) CAT bonds with trigger attachment and exhaustion points of \$20 and \$40 million, respectively; (iii) reinsurance with attachment and exhaustion points of \$20 and \$60 million and pro rata ratio of 0.2; and (iv) seismic upgrading (implementation of seismic provisions). The base portfolio is comprised of 1,000 wooden houses designed without seismic provisions, and the costs to implement the risk mitigation/transfer measures are included in the EP curves. The details of the analysis setup can be found in Goda (2014). All three risk mitigation/transfer options are effective in reducing insurer's earthquake exposure. The EP curve for the reinsurance contract has clear kinks, which correspond to the policy attachment and exhaustion points. The three options transform the original EP curve differently; reinsurance has clear influence on the specific loss layer between the attachment and exhaustion points, while seismic upgrading essentially shifts the EP curve

downward (i.e., chance of experiencing seismic loss is decreased because of enhanced seismic resistance). The influence of the CAT bonds is more widely distributed (note: the EP curve for the CAT-bond case includes the triggering errors). Importantly, the numerical CAT modeling results combined with quantitative risk metrics for different risk management strategies provide valuable insight on the earthquake risk exposure. This facilitates risk-based decision-making and cost-benefit analysis.

Summary

This entry reviewed a broad range of topics related to earthquake insurance and reinsurance. An emphasis was given to the role of insurance and reinsurance in mitigating financial consequences of catastrophic earthquakes as integrated part of disaster risk reduction. Basic terminologies and current insurance-reinsurance systems were introduced. In particular, challenges related to low-probability high-consequence risk characteristics were discussed. Various insurance and reinsurance models that have been developed and applied in different academic fields were described. The fast-evolving earthquake

catastrophe modeling techniques were summarized in the context of earthquake insurance portfolio management (including quantitative risk metrics). Subsequently, actuarial and financial insurance models, both statistical and computational, were presented. The homeowner's and insurer's decision models were then summarized. A sequential application of different insurance models (e.g., catastrophe modeling combined with dynamic financial analysis) is the state of the art in practice, requiring broad knowledge and understanding of available techniques. Finally, utilization of emerging alternative risk transfer tools, such as catastrophe bonds, was discussed to promote an integrated earthquake risk management framework.

Cross-References

- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- [Seismic Loss Assessment](#)

References

- Asprone D, Jalayer F, Simonelli S, Acconcia A, Protà A, Manfredi G (2013) Seismic insurance model for the Italian residential building stock. *Struct Saf* 44:70–79
- Bommer J, Spence R, Erdik M, Tabuchi S, Aydinoglu N, Booth E, del Re D, Peterken O (2002) Development of an earthquake loss model for Turkish catastrophe insurance. *J Seismol* 6:431–446
- Cossette H, Duchesne T, Marceau E (2003) Modeling catastrophes and their impact on insurance portfolios. *N Am Actuar J* 7:1–22
- Cummins JD (2008) CAT bonds and other risk-linked securities: state of the market and recent developments. *Risk Manag Insur Rev* 11:23–47
- Daniell JE, Khazai B, Wenzel F, Vervaeck A (2011) The CATDAT damaging earthquakes database. *Nat Hazards Earth Sys Sci* 11:2235–2251
- Dong W, Grossi P (2005) Insurance portfolio management. In: Grossi P, Kunreuther H (eds) *Catastrophe modeling: a new approach to managing risk*. Springer, New York, pp 119–133
- Ehrlich I, Becker GS (1972) Market insurance, self-insurance, and self-protection. *J Polit Econ* 80:623–648
- Goda K (2014) Seismic risk management of insurance portfolio using catastrophe bonds. *Comput Aided Civ Infrastruct Eng*. doi:10.1111/mice.12093
- Goda K, Ren J (2013) Seismic risk exposure modelling of insurance portfolio using extreme value theory and copula. In: Tesfamariam S, Goda K (eds) *Handbook of seismic risk analysis and management of civil infrastructure systems*. Woodhead Publishing, Cambridge, pp 760–786
- Goda K, Yoshikawa H (2012) Earthquake insurance portfolio analysis of wood-frame houses in south-western British Columbia, Canada. *Bull Earthq Eng* 10:615–643
- Grace MF, Klein RW, Kleindorfer PR, Murray MR (2003) *Catastrophe insurance: consumer demand, markets and regulation*. Kluwer, Boston
- Johnson EJ, Hershey J, Meszaros J, Kunreuther H (1993) Framing, probability distortions, and insurance decisions. *J Risk Uncertainty* 7:35–51
- Kahneman D (2003) Maps of bounded rationality: psychology for behavioral economics. *Am Econ Rev* 93:1449–1475
- Kajitani Y, Chang SE, Tatano H (2013) Economic impacts of the 2011 Tohoku-oki earthquake and tsunami. *Earthq Spectra* 29:S457–S478
- Kaufmann R, Gädmer A, Klett R (2001) Introduction to dynamic financial analysis. *ASTIN Bull* 31:213–249
- Kleindorfer P, Grossi P, Kunreuther H (2005) The impact of mitigation on homeowners and insurers: an analysis of model cities. In: Grossi P, Kunreuther H (eds) *Catastrophe modeling: a new approach to managing risk*. Springer, New York, pp 167–188
- Kunreuther H (1996) Mitigating disaster losses through insurance. *J Risk Uncertainty* 12:171–187
- Lalonde D (2005) Risk financing. In: Grossi P, Kunreuther H (eds) *Catastrophe modeling: a new approach to managing risk*. Springer, New York, pp 135–164
- Maccaferri S, Cariboni F, Campolongo F (2011) *Natural Catastrophes: risk relevance and insurance coverage in the EU*. European Commission, Joint Research Centre, Ispra
- McNeil AJ, Frey R, Embrechts P (2005) *Quantitative risk management: concepts, techniques, and tools*. Princeton University Press, Princeton
- Michel-Kerjan E, Hochrainer-Stigler S, Kunreuther H, Linnerooth-Bayer J, Melcher R, Muir-Wood R, Ranger N, Vaziri P, Young M (2013) *Catastrophe risk models for evaluating disaster risk reduction investments in developing countries*. *Risk Anal* 33:984–999
- Mossin J (1968) Aspects of rational insurance purchasing. *J Polit Econ* 91:304–311
- Muir-Wood R (2011) Designing optimal risk mitigation and risk transfer mechanisms to improve the management of earthquake risk in Chile. OECD working papers on Finance, Insurance and Private Pensions (No. 12). OECD, Paris
- Non-Life Insurance Rating Organization of Japan (NLIRO) (2008) *Earthquake insurance in Japan*.

- NLIRO, Tokyo. <http://www.giroj.or.jp/english/index.html>
- Organisation for Economic Co-operation and Development (OECD) (2008) Financial management of large scale catastrophes. Policy Issues in Insurance (No. 12). OECD, Paris
- Palm R (1995) Earthquake insurance – a longitudinal study of California homeowners. Westview Press, Boulder
- Porrini P, Schwarze R (2012) Insurance models and European climate change policies: an assessment. *Eur J Law Econ*. doi:10.1007/s10657-012-9376-6
- Rolski T, Schmidli H, Schmidt V, Teugels J (1999) Stochastic processes for insurance and finance. Wiley, New York
- Schlesinger H (2000) The theory of insurance demand. In: Dionne G (ed) *Handbook of insurance*. Kluwer, Boston, pp 131–151
- Viscusi WK, Evans WN (2006) Behavioral probabilities. *J Risk Uncertainty* 32:5–15
- Wesson RL, Perkins DM, Luco N, Karaca E (2009) Direct calculation of the probability distribution for earthquake losses to a portfolio. *Earthq Spectra* 25:687–706
- Yucemen MS (2013) Probabilistic assessment of earthquake insurance rates for buildings. In: Tesfamariam S, Goda K (eds) *Handbook of seismic risk analysis and management of civil infrastructure systems*. Woodhead Publishing, Cambridge, pp 787–814

losses can be prevented by implementing sound disaster mitigation plans; retrofitting the infrastructures, implementing good crisis management and recovery plans, etc. (NEES report 2004). This requires reliable predictions of damages due to anticipating earthquake scenarios. The conventional predictions are based on empirical relations which are obtained by statistical analysis of past earthquake disasters. These empirical predictions are less reliable since the conditions during the recorded major earthquakes, which are rare and have return periods of half a century or more, are significantly different from the conditions at the time of next major earthquake. During the long periods between major earthquakes, built environment and social and economic systems undergo significant changes. As an example, social and economic systems increasingly rely on lifeline networks, like electricity, communication, water, gas, sewer, transportation, etc., making it possible for any malfunction or damage to propagate way beyond what have been experienced during the past earthquake disasters. Even some of these vital lifelines may not have existed at the time of many of the past major events. Therefore, the next major earthquakes can easily outpace the earthquake mitigation efforts based on these empirical predictions.

Predictions, which take the current state of an urban system into account, enable to implement disaster mitigation and recovery plans, which must be significantly successful compared to those made solely based on the past earthquake disaster information. More reliable and detailed prediction of damages to buildings, lifelines, and other structures can be made if the enhanced materials and updated design codes used for the constructions and the current state are taken into account. Such predictions based on up-to-date information can restrain the propagation of damages and malfunctions in lifeline networks, which are becoming increasingly interdependent. With detailed, comprehensive, and reliable predictions, decision makers can prepare a sound recovery plan which minimizes the post disaster economic damages and restores the social system in a shorter time. Thus, what is necessary is

Integrated Earthquake Simulation

Lalith Wijerathne
Earthquake Research Institute, University of
Tokyo, Tokyo, Japan

Synonyms

End-to-end earthquake simulations; Large-scale earthquake simulations

Introduction

A major earthquake in a commercial metropolis can inflict catastrophic disasters, crippling economy, causing loss of lives, and damaging infrastructure and properties. Even though the earthquake itself cannot be predicted, it is a possibility that the associated catastrophic

detailed and comprehensive disaster prediction which considers the current state of built environment, social system, economy, etc. (Hori 2011).

Integrated earthquake simulation is a new paradigm which enables detailed and comprehensive analysis earthquake disasters in large urban system, taking the current state of built environment, social systems, etc. into account. The rapid progress of earth sciences, earthquake engineering, disaster mitigation sciences, and social sciences has produced well-established numerical tools for simulating many phases involved in an earthquake scenario. These numerical tools can be utilized to seamlessly simulate all the phases involved in an earthquake disaster, in high spatial and temporal resolutions. Opposing to the conventional disaster predictions, the use of such numerical simulations for disaster predictions has two main advantages: current state of built environment and socioeconomic systems can be taken into account; a large ensemble of plausible future earthquake scenarios can be taken into account. These two advantages address the major weaknesses of the conventional predictions, making the predictions based on integrated earthquake simulations more reliable.

Many of the numerical methods and tools necessary for building an integrated earthquake simulator already exist. Hence, the major task involved in developing an integrated earthquake simulator is to build a system which seamlessly integrates the necessary numerical tools for end-to-end earthquake simulations, for suitable computer hardware. There are well-developed numerical tools for earthquake hazard and disaster simulations, while the methodologies and numerical tools for simulating human responses, like evacuation, recovery, etc., are being developed. Despite the availability of some of the required numerical tools, it is not a trivial task to build a seamlessly integrated system. Several problems have to be addressed: lack of consistency between the inputs and outputs of some numerical tools, lack of numerical tools for simulating some phases, automation of analysis model constructions for large metropolis, efficient utilization of high-performance computing resources, etc.

The need for detailed, comprehensive, and reliable earthquake disaster predictions and the wide availability of high-performance computing resources have initiated several research activities on integrated earthquake simulation systems (Hori and Ichimura 2008; Hori 2011; Taborda and Bielak 2011; Muto et al. 2008). Further, the need of an integrated system for more reliable earthquake predictions has been emphasized by the Committee to Develop a Long-Term Research Agenda for the Network for Earthquake Engineering Simulation of the USA (NEES report 2004). Most of the existing integrated systems consider only the integration of earthquake hazard and seismic response of man-made structures. Taborda and Bielak (2011) have demonstrated the application of an integrated system to simulate wave propagation in large basins, amplification in surface layers, and seismic response of a small number of buildings. Muto et al. (2008) have presented a system for simulating rupture process to building damage and estimating repair costs, with small-scale seismic response analysis simulations. Hori et al. (Hori and Ichimura 2008; Hori 2011; Sobhaninejad et al. 2011) have developed a system called integrated earthquake simulator (IES) with the aim of seamlessly simulating all the involved phenomena from earthquake hazard to human responses, in large scale. They have proposed a macro-micro analysis method which not only efficiently simulate wave propagation and amplification in large domains but also treats the uncertainties due to lack of underground structure data. Also, the IES consists of an evacuation simulation module which can simulate evacuation in smaller-scale domains (Hori 2011).

The objective of this entry is to present an overview of integrated earthquake simulations. Being the most complete available integrated earthquake simulation system, most of the contents of this entry are prepared based on Hori et al.'s integrated earthquake simulator (Hori and Ichimura 2008; Hori 2011; Sobhaninejad et al. 2011). In the next section, an overview of integrated earthquake simulations is presented. The remaining sections present some details of earthquake hazard and disaster simulations;

neither tsunami nor human responses are included since these are not included in the available integrated systems. A short description of the numerical methods used for earthquake hazard and disaster simulations and automated analysis model construction based on GIS data are presented in these two sections. The last section presents a summary of the contents and a list of challenging problems which have to be addressed in the future.

Integrated System for End-to-End Earthquake Simulations

An integrated earthquake simulation system consists of a set of numerical tools, which are seamlessly integrated, for simulating the three main phases of an earthquake; earthquake and tsunami hazard, earthquake and tsunami disaster, and human actions against the earthquake disaster (see Fig. 1). Different numerical simulations that are used for simulating each of these three phases are summarized below.

1. Hazard simulations

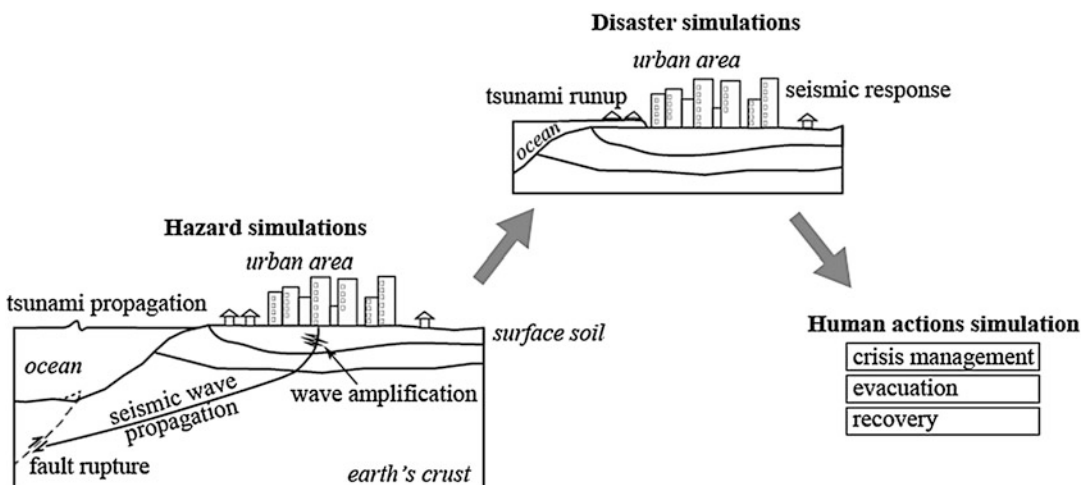
- (a) Earthquake hazard simulation: For a given fault mechanism, an earthquake wave is synthesized and propagation through the

crust is computed. Considering the 3D topographical effects and nonlinear properties of surface soil layers, wave amplification (i.e., strong ground motion) at the target area is computed. The output of the strong ground motion calculation should be consistent with the spatial and temporal resolutions required for seismic response analysis of man-made structures.

- (b) Tsunami hazard simulation: For a given fault mechanism, ocean bottom deformation and tsunami wave propagation are simulated.

2. Disaster simulations

- (a) Earthquake disaster simulation: Seismic responses of all the structures, both on the surface and underground, are computed using the strong ground motion data obtained in the hazard simulations. Depending on the type of each structure, a suitable numerical analysis method has to be used. This simulation should produce the necessary outputs, like the damage state of a building, which are required for human action simulations.
- (b) Tsunami disaster simulation: Simulation of tsunami flow in the target area considering the fluid structure interaction and damages to man-made structures due to



Integrated Earthquake Simulation, Fig. 1 Overview of integrated earthquake simulations

tsunami wave and the debris it carries. Though it is a challenging problem, simulation of damages to man-made structures is essential for the recovery process simulation.

3. Human actions against disasters

- (a) Evacuation: Generally, this involves the simulation of evacuation of buildings and moving to a safe location. In regions prone to tsunami disasters, moving to a safe high ground has to be simulated considering the human behaviors in time-critical events and the time history of inundation. The environment model of evacuation simulation should be modified, according to the results of the earthquake disaster simulation, so that unexpected delays caused by narrowing and blockages of roads are incorporated.
- (b) Crisis management: Simulation of possible emergency situations like fire, flooding, etc.
- (c) Recovery: Simulation of allocation and consumption of resource, like manpower, energy, materials, etc., to recover from the disaster, and estimate the resulting social and economic benefits.

In short, given the input data of a target area and a fault mechanism, an integrated system calculates the damages to the built environment and both the short-term and long-term human actions against the disaster. For the seismic hazard and disaster simulations, the required inputs are the data of underground structures, man-made structures, and a possible fault mechanism. The output of the disaster simulations should produce the necessary data for the simulation of human actions. In addition to the data from the disaster simulation, the human-response simulations require more data like the distribution of people at the time of the disaster and their properties, cost of repairing various damaged components, availability of various resources and manpower, etc. It would be a useful tool for decision makers, if the recovery simulation includes capabilities to find optimal way to allocate resources for a given damage state.

Applications

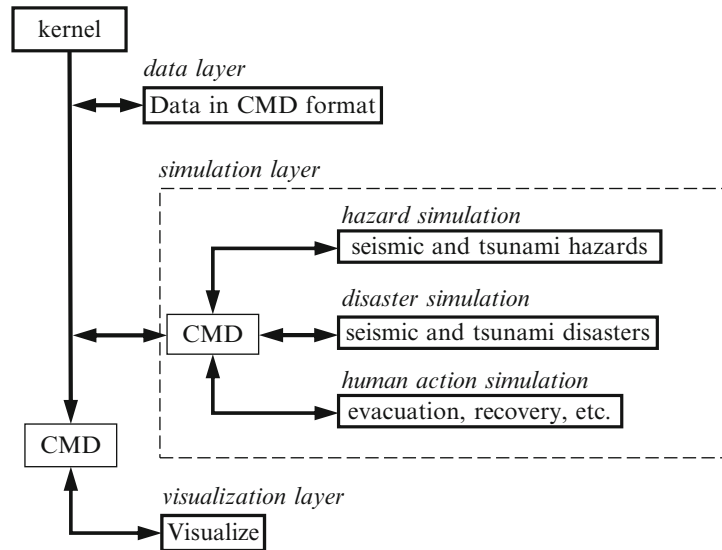
The main application of an integrated earthquake simulation system is to make detailed and comprehensive earthquake disaster predictions using the current state of built environment, social systems, and economy. This involves many uncertainties like the fault mechanism and the location of the next earthquake, lack of geological and underground structure data, etc. Unless these uncertainties are taken into consideration, the predictions of the integrated system are useless; what the decision makers need is predictions of high reliability so that they can use the predictions to implement disaster mitigation and recovery plans. In order to make reliable predictions, accounting the uncertainties of input data, some kind of stochastic approach like Monte Carlo simulation is necessary; Monte Carlo simulation is the widely used method for simulating this kind of phenomena involving significant uncertainties in input data. The detailed, comprehensive, and reliable predictions using an integrated earthquake simulator will ultimately make it possible to prevent catastrophic losses even though the anticipating earthquakes cannot be predicted.

In addition to the abovementioned main application, an integrated earthquake simulation system can be used for many other applications, some of which are listed below:

- As a tool for land-use planners to decide the required regulations or usages compatible with expected level of risk, when developing regions prone to seismic hazards. An integrated system can be used to assess the hazard level and find the construction standards and response measures which are necessary to prevent disasters in regions with seismic hazards. Then the public and policy makers can be informed of the earthquake risk, necessary planning and construction, and necessary response measures to reduce risks (NEES report 2004).
- For performance-based structural designs.
- Augmented reality and the results from a validated integrated system can be used to convince the political leaders and the public

Integrated Earthquake Simulation, Fig. 2 A

basic structure of an integrated earthquake simulator. Common modeling data (CMD) is the internal data format used by Hori et al.



that new disaster mitigation strategies are effective and economically viable risks (NEES report 2004).

- As a training tool for emergency and relief supply staffs, etc.

Not all the abovementioned numerical tools are included in present integrated earthquake simulators. Most of the existing integrated simulators include only the numerical tools for earthquake hazard and earthquake disaster simulations. The remaining sections provide detailed information of these numerical tools available in present-day integrated earthquake simulators. Though several numerical methods for tsunami hazard simulations are available (Goto et al. 1997; Mader 2004), there is a lack of numerical tools for simulating the tsunami disaster. The existing numerical tools only estimate the inundation and run-up height, and damages to man-made structures in large areas are not considered. Also, there are no well-established techniques for simulating other human responses, except the evacuation simulations. Multi-agent systems have been developed to simulate evacuation triggered by earthquakes or tsunamis. However, these tools are not sufficiently mature to simulate large urban area evacuations, which is the target domain of an integrated system.

Therefore, the remaining sections only focus on the earthquake hazard and disaster simulations.

Development of an Integrated System

Figure 2 shows a basic structure for an integrated earthquake simulator, proposed by Sobhaninejad et al. (see also Hori 2011). The kernel is the control center which controls the whole system. It controls inputs and outputs of the system and calls the individual simulation codes in proper order, providing the necessary input data and retrieving output data which are required for subsequent analysis codes. Each item shown in the simulation layer of Fig. 2 consists of a microkernel which controls the flow of each simulated phase. As an example, the seismic response analysis's microkernel controls how a large number of structures in a city is being executed, using suitable numerical tools to simulate different types of structures made of concrete, steel, wood, etc. Also, the microkernel distributes these structures among a large number of CPUs in high-performance computing (HPC) hardware such that all the CPUs have equal work loads. The use of HPC is essential to simulate millions of buildings in large metropolises.

Geographic information system (GIS) data is one of the prominent input data for an integrated earthquake simulator (Hori 2011). GISs are

widely used by many governmental and private organizations to store data. As an example, in Japan, GISs of ground elevation, borehole data, shapes of buildings, etc. are available. These data can be utilized to generate analysis model of metropolis. However, these GISs do not contain all the essential information for generating analysis models. As an example, GIS does not contain essential data for seismic response analysis like structural skeleton of buildings, material properties, etc. Although the necessary information is unavailable in current systems, the quantity and quality of GIS data are continuously improved. The next two sections present how to construct analysis models for earthquake hazard and disaster simulations based on GISs containing borehole data and configurations of buildings.

As for the development language, an object-oriented language, like C++ or python, is suitable since object-oriented features are convenient in large and complex system development: the ease of large project management with multiple developers, the ease of long-term software maintenance, ability to develop intuitive user interfaces, availability of generic programming, etc. Further, the ability to call libraries written from any of the standard languages and availability of several inter-process communication methods make it easier to integrate numerical tools developed with any language.

Being a system consisting of a number of different numerical tools, an integrated system requires a suitable internal data format to efficiently manage the flow of data among the numerical tools (Hori 2011; Sobhaninejad et al. 2011). The numerical tools for simulating various phases are usually provided by different research groups and have different data formats. Usage of multiple data formats has several problems: requires many data conversions, readability and maintainability of the source codes are poor, etc. The standard method of solving this problem is maintaining one internal data format with which the kernels interact with each other and the various numerical tools. Microkernels are responsible for translating the internal data format when interacting with the numerical tools. It would be an advantage if the basic elements of

the selected internal data format are readily exchangeable among CPUs with Message Passing Interface (MPI), which is the widely used standard library in parallel computing.

Some of the Difficulties Involved in Integrated System Development

Though it appears that integrating existing numerical tool to build an integrated earthquake simulator is simple, it involves several difficulties and challenges like:

- Lack of necessary numerical tools
- Incompatibility of inputs and outputs of some numerical tools
- Need of automated model construction
- Optimal utilization of high-performance computing resources
- Verification and validation of the integrated system

As mentioned above, there is a lack of well-developed numerical tools or methods to simulate certain phases like tsunami disaster and human responses. The incompatibility of inputs and outputs of different numerical tools is also a problem in integrating numerical tools. The outputs of most of the necessary numerical tools, which are developed in different disciplines, may not have consistent spatial and temporal resolutions required for subsequent analysis. To address this problem, close collaborations with the developers of relevant simulation tools are necessary. Being the analysis domains very large, automation of analysis model construction is essential, which is a challenging task; automated construction of suitable analysis model for each of millions of structures in a metropolis, based on available CAD or GIS data, is not trivial.

Enhancing an integrated system to optimally use high-performance computing (HPC) resources is also a major challenge. Obviously, optimal utilization of cutting-edge supercomputing resources is necessary to simulate a large metropolis. While there are HPC-enhanced numerical tools for hazard simulations, HPC-enhanced codes disaster simulation and human response have to be developed.

Further, building a seamlessly integrated system (i.e., a system which will simulate end-to-end process in a single execution) is challenging due to two main problems. The first problem is that even though each numerical tool is enhanced to optimally use supercomputing resources, it does not guarantee that a seamlessly integrated system can optimally utilize the supercomputing hardware. The main reason is that different numerical tools involve different amount of computations; hence, there is a drastic difference in the maximum number of CPUs required for the optimal performance of each numerical tool. As an example, the earthquake hazard simulation requires several hundred thousand CPUs, while evacuation simulation requires only a few thousands of CPUs. This results in waste of CPU time. The second difficulty is that different numerical tools require different parallel computing approaches, which cannot coexist in single execution. As an example, hybrid of shared and distributed memory parallelism is the best for earthquake hazard simulations, while earthquake disaster and evacuation simulations require only the distributed parallelism. These two HPC challenges, involved in end-to-end simulations in a single execution, can be easily avoided by executing each numerical tool independently, which is the same as replacing the main kernel (see Fig. 2) with human control. However, this is not an acceptable solution since simulating a large number of earthquake scenarios is going to be laborious and error prone; as mentioned, Monte Carlo simulations are necessary for improving the reliability of the earthquake disaster simulations.

Hazard Simulation

There are two main hazards involved: earthquake and tsunami. Earthquake hazard involves simulations of wave propagation through crust and amplification in the surface layers. Tsunami hazard involves ocean bottom deformation and tsunami wave propagation from source to shore. There are well-developed numerical methods and large-scale simulation programs for both the methods. However, existing integrated

earthquake simulators only include the earthquake hazard, probably because tsunami hazards are rare though cause severe destruction. In this section, the available methods of earthquake hazard simulations and model generation from GIS data is presented.

Earthquake Hazard Simulations

Earthquake hazard simulation involves three main steps: simulation of fault rupture, simulation of seismic wave propagation through the earth's crust, and wave amplification when traveling through surface soil layers (see Fig. 1). Out of these three, fault mechanism is the least understood; rupture process has complex spatial and temporal characteristics. The other two steps are mechanical processes which are well understood, and there are several numerical methods to simulate wave propagation and amplification processes (Bielak et al. 2005; Hori and Ichimura 2008; Cui et al. 2009; Käser et al. 2008). The objective of earthquake hazard simulation is, for a given fault scenario, numerical simulation of wave propagation and amplification process, at sufficiently high spatial and temporal resolutions required for earthquake disaster simulations. Though the two mechanical processes are well understood, these involve two difficulties: required high-resolution simulations involve large-scale computing and lack of accurate geological and underground structure data. This subsection presents brief descriptions of available methods to overcome these problems.

While the dramatically improving supercomputing performance provides the necessary computation power, selection of proper numerical method is essential for the optimal usage of the computing resources. The two main numerical methods used for large-scale earthquake hazard simulations are finite difference method (FDM) and finite element method (FEM) (Moczo et al. 2007). Both methods have advantages and disadvantages (Bielak et al. 2005; Hori 2011; Taborda and Bielak 2011). FDM has been the most popular due to several reasons like the ease of implementation, the wide choice of explicit and implicit time stepping methods, sufficient robustness and accuracy, efficiency, etc.

However, due to the reliance of regular grid, the conventional FDM has several drawbacks, like the involvement of over-refined grid when modeling media with high contrast in material properties, need of smaller time steps, etc. The involvement of over-refined grid and smaller time steps dramatically increases the necessary computer resources and computing time. On the other hand, FEM does not involve these problems; longer time steps can be used and mesh can be adapted to model local features. These advantages become more dramatic with increasing contrast in shear wave velocity in the heterogeneous medium. The two major drawbacks of FEM are that a considerable effort is necessary to generate a refined mesh and storing global stiffness matrix requires considerable memory. Bielak et al. (2005) have proposed an octree-based FEM method to overcome both of these problems. The near-linear scalability up to 98,000 CPUs, when simulating $180 \times 135 \times 62 \text{ km}^3$ region, shows the potential of octree-based approach (Taborda and Bielak 2011). The macro-micro analysis method, proposed by Hori et al. (Hori and Ichimura 2008; Hori 2011), overcomes the abovementioned two problems using singular perturbation expansion and bounding medium theory. A brief description of macro-micro analysis is given below.

The uncertainty of geological and underground structures is a serious problem which hinders the high-resolution earthquake hazard simulations. The presently used in situ observations and other technologies for mapping the geological and underground structures are inefficient or not sufficiently accurate for generating high-resolution analysis models, which are required for accurate and high-resolution seismic hazard simulations. This hinders accurate disaster simulations; accurate damage predictions require seismic hazard simulations of high spatial and temporal resolutions. As explained below, the macro-micro analysis method uses bounding medium theory to overcome this uncertainty problem.

Macro-Micro Analysis Method

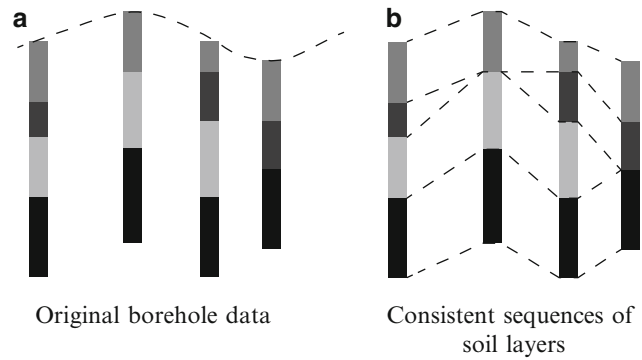
Macro-micro analysis method (Hori and Ichimura 2008; Hori 2011) is a stochastic method which

uses singular perturbation expansion and bounding medium theory to overcome the abovementioned two difficulties. In macro-micro analysis, a stochastic variational problem is posed to account the abovementioned uncertainties in geological and underground structures and material properties. In stochastic problems both the input variables with uncertainties and the output variables become random variables, which are defined both in physical space and probabilistic space. The inputs for the stochastic seismic hazard problem are the mean and variance of crust and ground structure configurations and mean and variance of uncertain material properties like density, Young's modulus, etc. Solving such stochastic problems is complex since the response of the system also becomes stochastic and not proportional to the variability of the input data. Micro-micro analysis method uses bounding medium theory to find optimistic and pessimistic estimates for the mean response of the stochastic model, without explicitly solving the stochastic problem using laborious or complex methods like Monte Carlo simulations or spectral method. The bounding medium theory determines two deterministic, but fictitious, models which provide optimistic and pessimistic estimates for the mean response.

The upper and lower bound models obtained with the bounding medium theory require large number of discretization according to the geometry and shear wave velocity of materials; hence, solving the resulting analysis models is computationally intensive. In order to efficiently solve this large-scale problem, which has wildly varying parameters, macro-micro analysis takes a multi-scale approach with singular perturbation expansion. Singular perturbation expansion is a mathematical technique which leads to multi-scale analysis. The resulting analyses in long and short scale are called macroanalysis and microanalysis. Macroanalysis considers the wave propagation in the entire domain in lower spatial resolution, while the microanalysis considers wave propagation in higher spatial resolution for each subdomain of interest. These two analyses are analogous to the wave propagation analysis in seismological scale and strong ground

Integrated Earthquake Simulation,

Fig. 3 Determination of consistent soil layers



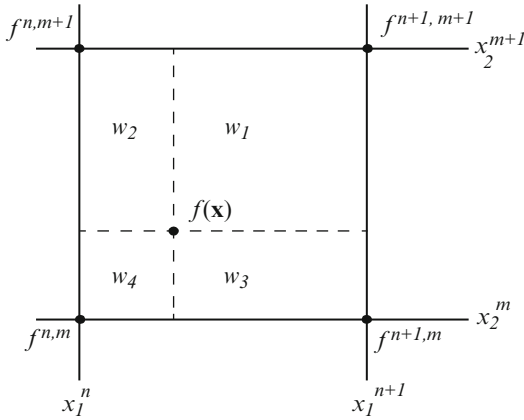
motion analysis in surface layers of urban areas. For the sake of reducing memory consumption and efficient computer implementation, macro-micro analysis uses voxel finite element method (VFEM) (Koketsu et al. 2004; Hori 2011). In summary, macro-micro uses bounding medium theory and singular perturbation expansion to numerically efficiently simulate the seismic wave propagation and amplification in large domains, which is a problem involving parameters with uncertainties, in high spatial and temporal resolutions.

Construction of Ground Structure Model from Sparse Borehole Data

As mentioned, accurate models of surface layers are necessary for high-resolution prediction of wave amplification in mediums with low shear wave velocities. However, constructing an accurate model is difficult due to limited availability of borehole data, some of which are poor in quality. Often these borehole data are coming from scattered locations and are inconsistent; borehole data is considered inconsistent if the sequence and the number of distinct layers differ from site to site (see Fig. 3a). To make a model of underground structures, based on borehole data, Hori (Hori 2011) has proposed a method which involves two steps: find a reference sequence of soil layers consistent with the data of each borehole site and interpolate the obtained consistent layers to find soil layer information at any arbitrary point. The proposed scheme, which is used in geo-information geology, for estimating a consistent set of soil layers at each site involves the following four steps:

- (i) Choose one borehole and use its layers as the reference sequence.
- (ii) Pick another borehole and compare its layers with the reference sequence.
 - (a) If the sequences are matching, choose another.
 - (b) Else make a new reference sequence consistent with the both by combining those; some layers can be of zero thickness (see Fig. 3b).
- (iii) Repeat step ii to obtain a reference layer sequence consistent with all the boreholes.
- (iv) Finalize by eliminating any insignificant layers.

Once a sequence of soil layers consistent with each borehole is found, soil layer information at any location can be evaluated using a suitable interpolation scheme. Hori (2011) has proposed the following grid algorithm to interpolate and obtain soil layer information in a regular grid. In order to explain this grid algorithm, consider the problem of interpolating the elevation of a certain soil layer interface. Assume that consistent layer information at K number of borehole sites of the domain D are available. Let the spatial location boreholes and the elevation of the specific layer interface be $\{\mathbf{x}^k\}$ and $\{f^k\}$, where $k = 1, 2, \dots, K$. In order to interpolate $\{f^k\}$, a coarse fictitious grid is introduced. Let the grid points denote by $(\mathbf{x}^{n,m})$ and the fictitious elevation of the specific layer interface at corresponding grid locations be $f^{n,m}$. The elevation of the specific layer $f(\mathbf{x})$, at borehole location $\mathbf{x}$, can be expressed by using the data of the four neighboring fictitious grid points as follows (see Fig. 4):



Integrated Earthquake Simulation, Fig. 4 Fictitious grid and the weights used for the bilinear interpolation

$$f(x) = w_1(x)f^{n,m} + w_2(x)f^{n+1,m} + w_3(x)f^{n,m+1} + w_4(x)f^{n+1,m+1},$$

where $w_1, w_2, \dots, w_4$ are bilinear weight functions. The weights are the area fractions of the rectangles formed by x and the four grid points. As an example

$$w_1 = \frac{(x_1^{n+1} - x_1)(x_2^{m+1} - x_2)}{(x_1^{n+1} - x_1^n)(x_2^{m+1} - x_2^m)}.$$

Once a coarse grid is made, the above bilinear interpolation can be repeatedly applied to obtain soil layer data in a grid, which is sufficiently fine for setting material properties of VFEM model for the microanalysis. As explained above, macro-micro analysis uses bounding medium theory to handle the uncertainties in the layer configurations and material properties.

Disaster Simulation

Out of the two main disasters involved, seismic disaster has received more attention (Hori and Ichimura 2008; Muto et al. 2008; Hori 2011; Taborda et al. 2011). There are a number of numerical methods developed for seismic

response analysis of buildings and other structures. Some of these methods are validated and have been extensively used for the purposes like structural designs, performance evaluations, structural health monitoring, etc. On the contrary, there are no well-established numerical methods for simulating large urban area tsunami disasters, incorporating the damages to structures inflicted by tsunami wave and debris. Development, verification, and validation of numerical methods capable of simulating large area tsunami inundation, including the destruction it brings, are challenging research areas which are essential for integrated earthquake simulators. In this section, the available seismic response analysis methods for buildings and approximate analysis model construction based on GIS data are briefly explained.

Available Numerical Methods for Seismic Response Analysis of Buildings

There are several numerical methods for the seismic response analysis, ranging from simple single degree of freedom (SDOF) systems to sophisticated nonlinear FEM models which consider large deformation and damage. Some well-established numerical techniques used in structural response analysis are SDOF, multi-degree of freedom systems (MDOF), one-component model (OCM), discrete element method (DEM), fiber element model, and nonlinear FEM. Though simple, nonlinear MDOF systems are capable of reproducing the complicated nonlinear structural responses (Pampanin et al. 2003). OCM models the nonlinear behavior of structural members by concentrating inelastic deformations at the ends of elements, with inelastic springs with hysteretic properties. The fiber element method (Spacone et al. 1996a, b; Haselton et al. 2008) is a moderately complex method widely used for simulating nonlinear seismic response of medium rise concrete structures. It is best to include numerical tools for all these methods in an integrated simulator. Then a user has a high flexibility in choosing a suitable analysis method depending on the characteristics and importance of each structure and available computing resources.

Automated Analysis Model Generation

It is a challenging task to automate the analysis model construction for millions of buildings in a metropolis, based on GIS data. Although GIS is a prominent source of input data for an integrated earthquake simulator, the current GIS data does not contain the necessary data like structural skeleton and material properties; only the information like location, height, shape, etc. are included in the current GIS. The difficulties in making analysis model significantly increase with the complexity of the analysis method. As an example, analysis model construction for SDOF or MDOF is way easier than analysis model construction for OCM or fiber element method. OCM and fiber element methods require detailed information of the structural skeleton, which is difficult to be generated automatically from GIS data.

Even though the present GIS does not include the necessary data for structural response analysis, it is possible to generate an approximate analysis model based on the design procedures and recommendations provided in relevant design codes and handbooks. In the absence of necessary data in GIS or other sources, such approximate models, which comply with design manuals, are useful in obtaining a probable seismic disaster of a large metropolis. An approximate method of constructing MDOF models of residential buildings using the available GIS data, which is proposed by Hori (2011), is given below.

Approximate MDOF Model Generation Based on GIS Data

The governing equation of a building, which is subjected to seismic acceleration $\ddot{z}(t)$, modeled as an MDOF system is given by

$$\mathbf{M}\ddot{\mathbf{U}}(t) + \mathbf{C}\dot{\mathbf{U}}(t) + \mathbf{K}\mathbf{U}(t) = \ddot{\mathbf{z}}(t)\mathbf{M}\mathbf{1}, \quad (1)$$

where the vector $\mathbf{1} = [1, 1, \dots, 1]^T$. The data provided by the current GIS is insufficient to define the mass, damping, and stiffness matrices,

which are represented by $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$. However, if rewritten in the form used for modal analysis, the recommendations given in design manual can be utilized to solve the above problem. To that end, the natural frequencies and corresponding modes of the undamped system is found by solving the following eigenvalue problem:

$$(\mathbf{K} - \omega_n^2 \mathbf{M})\Phi_n = 0$$

The above eigenvalue problem can be easily obtained by setting $\mathbf{U} = \Phi \exp(i\omega t)$, $\mathbf{C} = 0$ and $\ddot{\mathbf{z}}(t) = 0$ in Eq. 1. Solving this eigenvalue problem, the n^{th} natural frequency, ω_n , and the corresponding mode shape, Φ_n , can be obtained; $n = 1, 2, \dots, N$, where N is the dimension of the matrices. The N mode shapes or eigenvectors form a basis; in the presence of any repeated frequencies, the associated modes should be orthogonalized. The unknown $\mathbf{U}$ can be expressed using these N basis functions as

$$\mathbf{U} = \sum q_n \Phi_n. \quad (2)$$

Substituting Eq. 2 to Eq. 1 and following the standard modal analysis procedures, the following N set of differential equations can be obtained (Datta 2010):

$$\ddot{q}_n(t) + 2\xi_n \omega_n \dot{q}_n(t) + \omega_n^2 q_n(t) = \ddot{z}_n(t), \quad (3)$$

where ξ_n is the damping ratio and $\ddot{z}_n(t)$ is the external inertial force corresponding to the n^{th} mode. The damping ratio and inertial force due to seismic loading are given by

$$\xi_n = \frac{\Phi_n^T \mathbf{C} \Phi_n}{2\Phi_n^T \mathbf{M} \Phi_n} \text{ and } \ddot{z}_n = \frac{\Phi_n^T \mathbf{M} \mathbf{1}}{2\Phi_n^T \mathbf{M} \Phi_n}.$$

According to the standard procedure of MDOF, the N equations given by Eq. 3 are solved for q_n 's with proper boundary conditions and the unknown $\mathbf{U}$ is found using Eq. 2. It is not necessary to use all the N modes, and the solution can be approximated with a selected set of modes.

Integrated Earthquake Simulation, Table 1 Period and damping ratio of the fundamental mode for different types of structures. H is the height of a building

Building type	$T_1/(s)$	ξ_1
Wooden house	0.2–0.7	0.02
Reinforced concrete	$0.02 H$	0.03
Steel reinforced concrete	$0.03 H$	0.02

Often, only the first few modes are used for short buildings. As an example, the first three modes are used for buildings with the period of fundamental mode shorter than 0.5 s (Hori 2011).

In the absence of necessary data to define M , C , and K of a particular building, approximate analysis model can be constructed based on the standard dynamic properties given in building design codes. As an example, the standard values of the period T_1 and the damping ratio ξ_1 of the fundamental mode residential buildings recommended by the Japanese design code are given in Table 1.

Further, the periods and damping ratios of second and third modes can be obtained from the following two empirical relations:

$$\begin{bmatrix} T_2 \\ T_3 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{3}{5} \\ 1 \\ \frac{5}{5} \end{bmatrix} T_1$$

$$\xi_{n+1} = \begin{cases} 1.4 \xi_n & \text{for reinforced buildings} \\ 1.3 \xi_n & \text{for steel reinforced buildings} \end{cases}$$

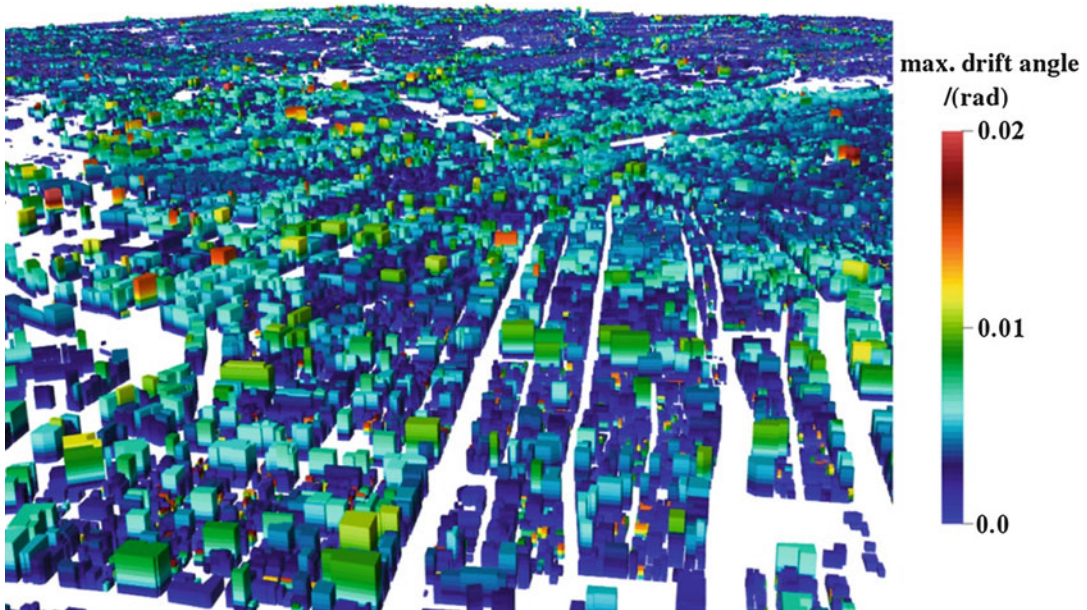
The above empirical relations make it possible to estimate the natural frequencies, $\omega_n = 2\pi/T_i$, and the damping ratios of the first three modes, which are sufficient for solving Eq. 3 for the amplitudes of the first three modes. Finally, the response can be estimated with Eq. 2. The required modal shapes can be easily estimated since the vibration modes of the dominant modes are known. The dominant modal shapes of buildings are given in standard textbooks, and even the modes of a cantilever beam are a good approximation (Hans and Boutin 2008).

Similar to the above given procedure, approximate analysis models for more complex analysis methods like OCM and fiber element model can be generated based on recommendations given in design manuals and handbooks. Unlike in the case of MDOF, automated model constructions for these methods are more challenging since these methods require structural skeleton details. The generation of structural skeletons from GIS data is straightforward when floor plan of a structure is made of multiple rectangles. However, if a building's shape changes with the floor number or has complex outer shape, the process of structural skeleton generation is complex and requires expertise in architecture and computational geometry. The spacing and the dimensions of the structural elements can be estimated based on the recommendations given in design manuals. As an example, 1/5 of floor height is often recommended as the height of beams. Obviously, automated FEM model construction is the most challenging.

Figure 5 shows the maximums of the inter-story drift angle compiled out of 400 simulations with different observed earthquake observations. The input strong ground motion for each building is computed using the 1D wave theory, using the details of underground structures and seismic waves recorded during 400 different earthquakes. The analysis models for buildings are automatically generated using the abovementioned approximate method for MDOF model generation from GIS data. Such Monte Carlo simulations can be used for more reliable earthquake disaster predictions.

Summary

In this chapter, the need of integrated earthquake simulation, a short overview of existing integrated earthquake simulators, and some remaining challenges are presented. The available integrated earthquake simulators have included only the earthquake hazard simulation



Integrated Earthquake Simulation, Fig. 5 Maximum inter-story drift angle obtained from 400 seismic response analysis simulations (Provided by Kohei Fujita, Department of Civil Engineering, University of Tokyo)

and earthquake disaster simulations. One of the remaining challenges in earthquake disaster simulation is automation of analysis model construction for complex analysis methods like fiber element model or FEM. Even though tsunami disasters are catastrophic, the present integrated systems have not included tsunami disaster simulations. It surely is challenging to develop numerical tools for tsunami disaster simulations including damages to structures. Even though tsunami disasters are rare, it is an essential item for an integrated system. The least developed is the simulations of human actions against earthquake disasters. The available numerical tools for evacuation simulations have to be further improved and enhanced with high-performance computing for simulating large domains. Out of all the numerical tools required for an integrated earthquake simulator, simulation of the recovery process is the least developed. It should be an invaluable tool for the decision makers, if a numerical method is developed to find the optimal way to allocate resources for recovering from a given earthquake disaster.

Cross-References

- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- [Physics-Based Ground-Motion Simulation](#)
- [Site Response: 1-D Time Domain Analyses](#)
- [Stochastic Ground Motion Simulation](#)

References

- Bielak J, Ghattas O, Kim EJ (2005) Parallel octree-based finite element method for large-scale earthquake ground motion simulation. *Comput Model Eng Sci* 10–2:99–112
- Committee to develop a long-term research agenda for the network for earthquake engineering simulation (NEES), National research council (2004) Preventing earthquake disasters: the grand challenge in earthquake engineering: a research agenda for the network for earthquake engineering simulation. The National Academic Press. <http://www.nap.edu/catalog/10799.html>. Accessed 15 July 2013
- Cui Y, Moore R, Olsen K, Chourasia A, Maechling P, Minster B, Day S, Hu Y, Zhu J, Jordan T (2009) Toward petascale earthquake simulations. *Acta Geotech* 4–2:79–93

- Datta TK (2010) *Seismic analysis of structures*. Wiley, Hoboken
- Goto C, Ogawa Y, Shuto N, Imamura F (1997) IUGG/IOC time project, numerical method of tsunami simulation with the leap-frog scheme, IOC manuals and guides. UNESCO, Paris
- Hans S, Boutin C (2008) Dynamics of discrete framed structures: a unified homogenized description. *J Mech Mater Struct* 3–9:1709–1739
- Haselton CB, Liel AB, Lange ST, Deierlein GG (2008) Beam-column element model calibrated for predicting flexural response leading to global collapse of RC frame buildings. PEER report 2007/03, Pacific Earthquake Engineering Research Center, College of Engineering, University of California, Berkeley
- Hori M (2011) *Introduction to computational earthquake engineering*. Imperial College Press, London
- Hori M, Ichimura T (2008) Current state of integrated earthquake simulation for earthquake hazard and disaster. *J Seismol* 12:307–321
- Käser M, Hermann V, Puente J (2008) Quantitative accuracy analysis of the discontinuous Galerkin method for seismic wave propagation. *Geophys J Int* 173(3):990–999
- Koketsu K, Fujiwara H, Ikegami Y (2004) Finite-element Simulation of Seismic Ground Motion with a Voxel Mesh. *Pure Appl Geophys* 161(11–12):2183–2198
- Mader CL (2004) *Numerical modeling of water waves*. CRC press, Boca Raton
- Moczo P, Kristeka J, Galisb M, Pazaka P, Balazovjeh M (2007) The finite-difference and finite-element modeling of seismic wave propagation and earthquake motion. *Acta Physica Slovaca* 57–2:177–406
- Muto M, Krishnan S, Beck JL, Mitrani-Reiser J (2008) Seismic loss estimation based on end-to-end simulation, life-cycle civil engineering. In: *Proceedings of the first international symposium on life-cycle civil engineering*. Varenna/Lake Como, CRC Press, Boca Raton, 10–14 June 2008. ISBN 0415468574
- Pampanin S, Christopoulos C, Priestley JN (2003) Performance-based seismic response of frame structures including residual deformations part II: multi-degree of freedom systems. *J Earthq Eng* 7(1):119–147
- Sobhaninejad G, Hori M, Kabeyasawa T (2011) Enhancing integrated earthquake simulation with high performance computing. *Adv Eng Softw* 42:286–292
- Spacone E, Filippou FC, Taucer FF (1996a) Fibre beam-column model for non-linear analysis of R/C frames: part I. Formulation. *Earthq Eng Struct Dyn* 25–7:711–725
- Spacone E, Filippou FC, Taucer FF (1996b) Fibre beam-column model for non-linear analysis of R/C frames: part II. Applications. *Earthq Eng Struct Dyn* 25–7:727–742
- Taborda R, Bielak J (2011) Large-scale earthquake simulation: computational seismology and complex engineering systems. *Comput Sci Eng* 13–4:14–27

Intensity Scale ESI 2007 for Assessing Earthquake Intensities

Pablo G. Silva¹, Alessandro Maria Michetti² and Luca Guerrieri³

¹Departamento de Geología, Universidad de Salamanca, Escuela Politécnica Superior de Ávila, Avila, Spain

²Dipartimento di Scienza e Alta Tecnologia, Università dell'Insubria, Como, Italy

³Department of Geological Survey, ISPRA, Istituto Superiore per la Protezione e la Ricerca Ambientale, Rome, Italy

Synonyms

Earthquake geological effects; Environmental seismic intensity; Macroseismic scale

Introduction

Earthquake intensity scales were introduced at the end of the nineteenth century (e.g., Rossi-Forel, Cancani, Mercalli) in order to characterize source parameters, damage distribution, and environmental impact of relevant seismic events. These intensity scales were based on a classification of earthquake effects on humans, on buildings, and on the natural environment.

Intensity provides a measure of earthquake-induced damage both at a site (local intensity) and at the epicenter (epicentral intensity). It is important to note that intensity evaluations consider the coseismic effects in the whole range of frequencies of vibratory ground motion, together with those resulting from static, finite deformations (fault ground ruptures). It is common practice to use the term “macroseismicity” to describe effects that are measurable with the intensity scales, because very small earthquakes (“microseismicity”), commonly considered intensity degree one (I), are beyond the sensitivity level of these scales.

Most common intensity scales are divided in to twelve degrees using roman numbers (I to XII).

The diagnostic effects for low intensity degrees (I to V) are those on humans, then (VI to IX) those on buildings, and for the highest degrees (X to XII) those on the natural environment. In fact, the earthquake environmental effects (EEEs) start to be noticeable from intensity $\geq IV$ as indicated in the Environmental Seismic Intensity scale (ESI-07 scale; Michetti et al. 2007), which is the subject of this chapter.

In spite of the increasing development of the instrumental record of earthquakes from the second half of the twentieth century, macroseismic intensity analyses are still the usual practice when analyzing both historical and recent earthquakes. Recent earthquakes provide the opportunity to compare earthquake damage levels and instrumental records of horizontal ground acceleration to classify and properly quantify the recorded effects on buildings and environment and consequently to refine intensity scales.

By contrast, historical earthquakes only can be explored by macroseismic analyses, and intensity is the only parameter that allows an estimate of the earthquake size to be made. Experience gained from the macroseismic analyses of instrumental events led to the development of empirical relationships in order to approximately estimate the moment magnitude (M_w) of historical events from earthquake intensity distribution. As a result, instrumental records are in the best case available from the first half of the twentieth century; however, historical intensity analyses (extended in some regions back to the sixth to seventh century B.C.) are behind most of the seismic hazard analyses, seismic-building codes, and nuclear regulatory guides worldwide. Therefore, the use and refinement of intensity scales in earthquake analysis is still, and will continue to be, imperative to the society.

ESI-07 scale (Michetti et al. 2007) classifies earthquake environmental effects (EEEs) qualitatively as described in previous macroseismic intensity scales (e.g., MCS, MSK, MM). The ESI-07 scale has been developed primarily to properly describe and quantify EEEs for different intensity levels, based on the extensive amount of observations and research accumulated in the past decades in this field (e.g., Reicherter et al. 2009;

Audemard and Michetti 2011). Moreover, the ESI-07 scale has been introduced to fill the gaps of some modern macroseismic intensity scales, such as the European Macroseismic Scale (EMS-98; Grunthal 1998; Musson et al. 2010), which basically excludes EEEs from intensity assessment. Finally, the ESI-07 scale has also the objective of incorporating paleoseismic analysis into the macroseismic studies, extending the information on earthquake intensity data to a geologic time window in the order of some tens of kyrs.

The ESI-07 scale draws upon the expertise of a group of geologists, seismologists, and engineers under the support of the International Union for Quaternary Research (INQUA). The definition of the intensity degrees has been the result of a revision of the EEEs caused by a large number of earthquakes occurring in different parts of the world, conducted by the “INQUA International Focus Group on ► Paleoseismology and Active Tectonics” since the year 2003. The scale was formally approved during the XVII INQUA Congress, held in Cairns (Australia) on August 2007. The ensuing literature includes applications of the scale with a global coverage (e.g., Ali et al. 2009; Berzhinskii et al. 2010; Di Manna et al. (2013); Fountoulis and Mavroulis 2013; Gojsar 2012; Guerrieri et al. 2008; Lalinde and Sanchez 2007; Mavroulis et al. 2013; Ota et al. 2009; Papanikolaou 2011; Papathanassiou and Pavlides 2007; Serva et al. 2007; Tatevossian 2007). More recently, Silva et al. (2014) implemented the first official Catalogue on Earthquake Environmental Effects published by the Geological Survey of Spain (IGME) based on the On-line EEE Catalogue hosted at the Geological Survey of Italy (ISPRA) (<http://www.eeecatalog.sinanet.apat.it/terremoti/index.php>).

In the following sections, we will analyze the role of earthquake environmental effects in the existing macroseismic scales, how EEEs are classified (primary and secondary effects) and incorporated to the ESI-07 scale, the guidelines for the application of the scale, and finally the use of this intensity scale in paleoseismic research. The original text of the scale after Michetti et al. (2007) can be found in Appendix 1.

Earthquake Environmental Effects in the Macroseismic Scales

As mentioned in the introduction, intensity is a description of an earthquake based on the effect to humans, to buildings, and to the natural environment. Macroseismic scales classify these effects for each intensity level. A classification is not a measurement, and there is no way to define the “energetic distance” between different degrees of an intensity scale. In fact, macroseismic classifications are based on empirical qualitative scales, which by definition include significant subjectivity or, better, expert judgment. For this reason, all the scales use integer values from I to XII. Therefore, the only mathematical property that we can use is the order (and this is why the roman numerals are used for defining the intensity degrees).

The scientific worth of intensity scales, however, is to order the coseismic effects in degrees that try to include phenomena indicative of large differences in the strength of the earthquake at a site. Intensity IX in all scales of the “Mercalli family” (or, for the sake of precision, “Cancani family,” since Cancani was the first to introduce a 12° intensity scale) describes a scenario of effects, or a level of shaking, that is completely different from (much higher than) intensity VIII and completely different from (much lower than) intensity X. This is the “power of resolution” of the scales. Even with quite a “naive” description of the earthquake, intensity is of paramount importance in seismology and geology and engineering, primarily because intensity allows us to retrieve invaluable information from the historical record of seismicity. The evolution of the scales in the past century has been an effort to keep intensity in line with new knowledge in earthquake engineering and geology and to provide classifications of the effects with increasing objectivity and repeatability but within a substantial continuity with the original intensity scales developed by Rossi-Forel in 1883, Mercalli in 1889, and Mercalli-Cancani-Sieberg (MCS; Sieberg 1930).

All the intensity scales (Rossi-Forel, Mercalli, MCS, MSK, Mercalli Modified) consider the

earthquake environmental effects as diagnostic elements for intensity evaluation. This is especially relevant from intensity degrees $\geq$ IX, where intensity evaluation based on building damage nearly saturates (Michetti et al. 2007). However, some modern practices of macroseismic investigation (e.g., Grunthal 1998) tend to only focus on the effects on humans and the anthropic environment, overlooking the earthquake environmental effects, based on the opinion that these effects are too variable and aleatory. As aforementioned, this is especially the case for the new European Macroseismic Scales (EMS-98; Grunthal 1998; Musson et al. 2010) and its impact in European building codes.

Nevertheless, other authors (e.g., Dengler and McPherson 1993; Serva 1994; Serva et al. 2007; Dowrick 1996; Hancox et al. 2002; Michetti et al. 2004; McCalpin 2009) have provided clear evidence that the characteristics of geological and environmental effects are an essential piece of information in order to understand how intensity grows with the earthquake size (for seismic events with the same hypocentral depth). In this sense, intensity combines the effects of seismic shaking with local topographic and geological conditions (topographic and site-effect amplifications or attenuations) triggering temporal or permanent environmental effects in the landscape, which nowadays are widely retrievable from historical sources and/or paleoseismic research. Intensity is therefore a way to classify the interaction between the earthquake and the environment (people, objects, buildings, lifelines, trees, soils, slopes, rivers, aquifers, coastlines, etc.).

The record of earthquake environmental effects extends the set of information even far from urban zones, offering a better and more complete picture of the seismic scenario. Additionally, environmental effects are in many cases a direct source of damage in urban areas (e.g., liquefaction processes, large slope movements), sometimes even exceeding the damage produced by seismic shaking itself (e.g., tsunamis). This constitutes a keystone for intensity assessment in some damaged zones, as it is difficult to differentiate seismic-shaking damage from that

produced by environmental effects through the application of traditional intensity scales. Most of these traditional scales consider the occurrence of observable earthquake environmental effects from intensities $\geq VI$, noticeable from intensity VIII, and destructive or devastating for intensities $\geq IX$, but with descriptions somewhat qualitative and not properly classified.

The ESI-07 scale therefore provides a classification and quantification of earthquake environmental effects for the different intensity degrees, offering quantitative numerical data for the size of primary effects (surface faulting and coseismic uplift/subsidence) and secondary effects (slope movements, ground cracks, liquefaction processes, tsunamis). Other environmental effects, with poor to no potential of preservation in the geological or geomorphological records, such as hydrological anomalies in springs and wells, tree shacking, jumping stones, and dust clouds, are also included but providing qualitative descriptions. The scale also highlights the occurrence of frequent, characteristic, and diagnostic features on the assemblage of earthquake environmental effects for the different intensity degrees (see Appendix 1).

Consequently, ESI-07 scale allows an assessment of seismic intensity based only on the earthquake environmental effects. This can be regarded as an update, reclassification, and quantification of environmental effects included in the traditional macroseismic scales (MCS, MSK, MM). Its use, alone (in unpopulated regions) or, as a rule, integrated with the macroseismic traditional scales (for instance, intensity degree IX on the ESI-07 is equivalent to degree IX on the MCS, MM, and MSK scales; Michetti et al. 2004), will offer a more comprehensive picture of the earthquake's scenario and impact and a more consolidated estimate of intensity, allowing, therefore, a better comparability among different earthquakes.

Primary and Secondary Earthquake Environmental Effects in the ESI-07 scale

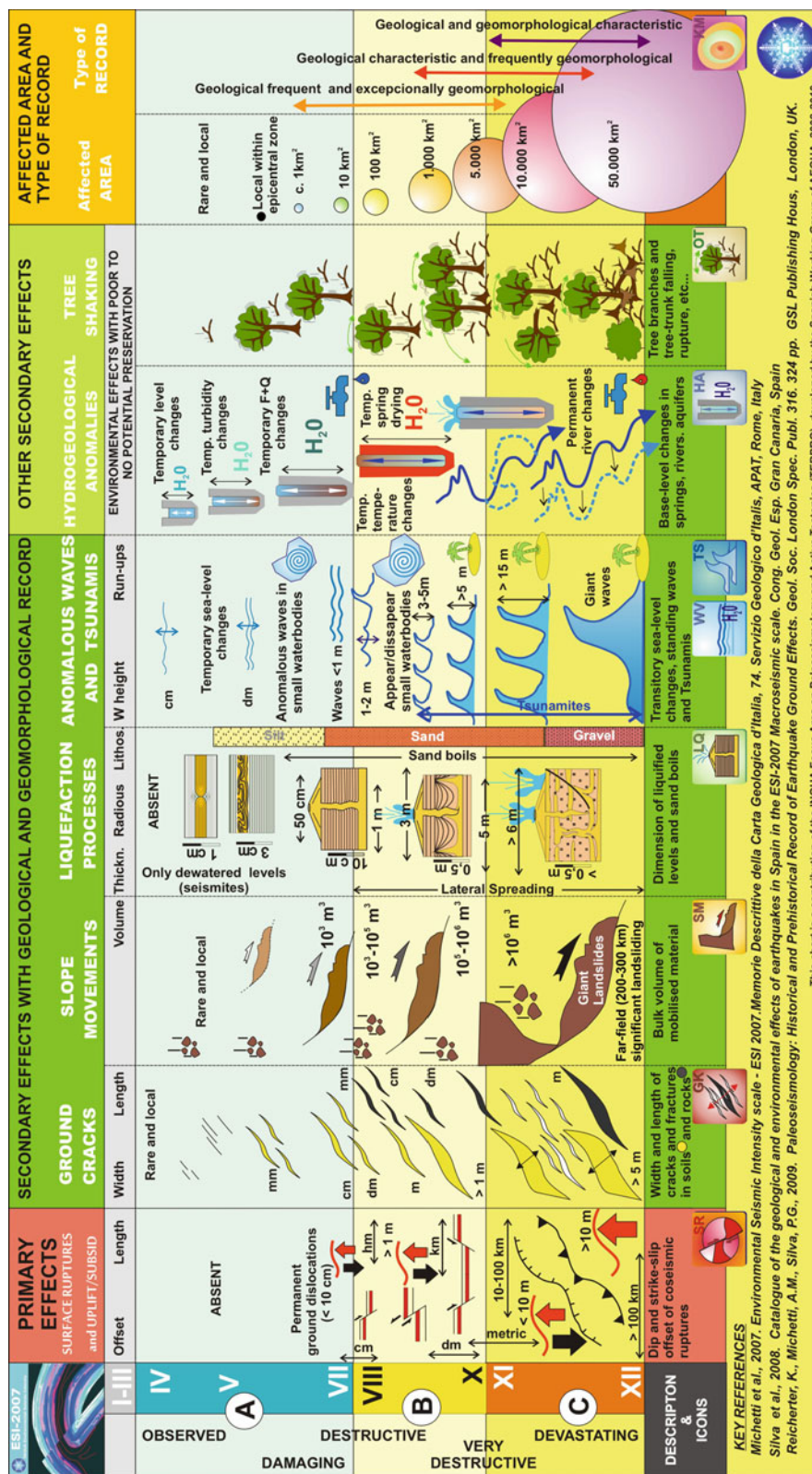
Earthquake environmental effects (EEEs) are all the phenomena generated in the natural

environment by a seismic event. These can be temporal or permanent, holding a different potential of preservation in the geological record. Typically, from intensities $\geq VIII$, EEEs can be permanently printed in the geology and landscape of a zone, being susceptible of paleoseismic analysis (Fig. 1). The ESI-07 scale classifies the EEEs in two main types: (a) primary and (b) secondary effects. While primary effects (e.g., surface faulting) are useful to directly assess the earthquake intensity, the nature, dimensions, and areal extent of secondary effects (mainly landslides and liquefaction processes) allow the complementary estimation of the epicentral intensity of an earthquake.

Primary Effects

Primary effects include any surface expression of the seismogenic tectonic source, including surface faulting, surface tectonic uplift, and subsidence. These effects are typically observed for shallow crustal earthquakes over a certain threshold value of magnitude, commonly considered to be between 6.0 and 6.5 Mw, being directly linked to the size, hence the energy of the earthquake. Primary effects in principle do not suffer saturation, although there is a physical limit to their dimensions (included in the XII degree). The size of primary effects is typically expressed in terms of two parameters: (i) total surface rupture length (SRL) and (ii) maximum displacement (MD). The amount of tectonic surface deformation (uplift, subsidence) is also considered in the assessment.

The focal depth and the stress environment of an earthquake obviously control the occurrence and the size of the observed effects. Two crustal earthquakes with the same energy but very different focal depths and stress environment can produce a different range of environmental effects, and, therefore, their respective local intensity values can significantly differ. Especially in volcanic areas, earthquakes of very shallow focus (c. 3–4 km) and low magnitude can be associated with the manifestation of primary effects. Taking these considerations into account, the ESI-07 scale considers the lower cutoff for the occurrence of primary effects



Intensity Scale ESI 2007 for Assessing Earthquake Intensities, Fig. 1 ESI-07 chart illustrating the main features and size parameters of the macroseismic scale for the different intensity degrees described in Appendix 1 (Modified from Reicherter et al. (2009))

(i.e., surface faulting) in intensity VII, whereas for typical crustal earthquakes (focal depth 5–15 km), in non-volcanic zones, primary effects start from intensity VIII.

Secondary Effects

Secondary effects include a wide range of phenomena affecting the natural environment, generally induced by the ground shaking. Their occurrence is typically observed in a specific range of intensities. The ESI-07 scale describes the characteristics and size of each type of secondary effect as a diagnostic feature in a range of intensity degrees. In some cases, it is only possible to establish a minimum intensity value. The total areal distribution of secondary effects grows with the size of the event, without saturation. Therefore, the ESI-07 scale uses this criterion as an independent tool for the assessment of the epicentral intensity I_0 .

EEEs are more commonly observed and characterized from intensity IV, which is the lower bound of the ESI-07 scale. Some types of environmental effects concerning hydrological anomalies may be observed even at lower degrees, but they cannot be considered as diagnostic for intensity assessments. The evaluation accuracy increases toward the highest degrees, particularly in the range where primary effects occur (typically for intensity VIII and above), preserving its resolution up to intensity XII. Indeed, effects on man and man-made structures saturate (i.e., all buildings in the epicentral area are completely destroyed) starting from intensity X of traditional scales, so that a reliable intensity becomes difficult to assess. Instead, in the same range, EEEs still provide reliable evidence for an intensity appraisal.

In the ESI-07 scale, EEEs are classified into eight main categories: (1) hydrological anomalies, (2) anomalous waves/tsunamis, (3) ground cracks, (4) slope movements, (5) tree shaking, (6) liquefaction, (7) dust clouds, and (8) jumping stones.

Hydrological anomalies (HA). Hydrological phenomena show up from intensity $\geq$ IV and saturate (i.e., their size does not increase) at intensity X. They can be divided in two groups: (a) surface

water effects including overflow, discharge variations, and increasing turbidity of rivers and lakes and (b) groundwater effects including drying up of springs, appearance of new springs, temperature changes, anomalies in chemical components, and turbidity of springs. Further useful information might be the amount and rates of variation in temperature and discharge, the presence of anomalous chemical element, the duration of the anomaly, and the time delay (the time interval between the seismic shock and the observation of the specific effect).

Anomalous waves/Tsunamis (AW). In this category are included seiches in closed basins, outpouring of water from pools and basins, and tsunami waves. In the case of tsunamis, more than the size of the tsunami wave itself, the effects on the shores (especially run-up, beach erosion, change of coastal morphology), without neglecting those on humans and man-made structures, are taken as diagnostic of the suffered intensity. These effects may already occur from intensity IV but are more diagnostic from intensities IX to XII. The definition of intensity degrees has taken advantage from previous tsunami intensity scales (e.g., Papadopoulos and Imamura 2001) and from many recent descriptions of the aforementioned effects worldwide (e.g., Lander et al. 2003).

Ground cracks (GK). Ground cracks appear from intensity IV and saturate (i.e., their size does not increase) at intensity X. Diagnostic parameters are lithology, strike and dip, maximum width, and areal density. Different dimensions of fractures are considered for soft materials (loose alluvial deposits and/or saturated soils) and hard rocks. This category also includes the occurrence of cracking, undulations, and/or pressure ridges in paved (asphalt or stone) roads.

Slope movements (SM). Slope movements, including underwater landslides, consider a wide variety of gravitational slope processes such as rockfalls, debris falls, and toppling; rock or debris slides, avalanches, and mudslides; debris-, earth-, and mudflows; and sackungen, complex landslides, and eventually lateral spreads. This category may show up at intensity IV and saturate (i.e., their size does not increase)

at intensity X. The total volume of mobilized material is diagnostic for intensity assessment. It can be roughly estimated on the basis of the landslide area when the depth of sliding mass can be reasonably estimated. The uncertainties introduced with this procedure do not appear to significantly influence the intensity evaluation. Additional useful information is maximum dimension of individual blocks (rockfalls), area affected by slope movements, density of slope movements, amount of slip, humidity, and time delay.

Liquefaction processes (LQ). Liquefactions generally materialize at intensity V or greater. The diagnostic features for liquefactions are the diameter of sand volcanoes and their lithology (silt, sands, and gravel). Saturation (i.e., their size does not increase) occurs at intensity X. Other useful characteristics are shape, the time delay, the depth of water table, and the occurrence of water and sand ejection.

Other environmental effects (OT) considered in the scale have a scarce to null preservation potential and a low diagnostic relevance, but of use of well-documented historical and recent events. Among them are the following:

- **Tree shaking (TR).** This phenomenon is reported to occur from a minimum intensity IV but noticeable for intensity VII and diagnostic from intensity VIII. It is important to record the occurrence of broken branches and the morphologic characteristics of the area (flat, slope). The definition of intensity degrees basically follows that provided by Dengler and McPherson (1993).
- **Dust clouds (DC)** are reported from intensity VIII, typically in dry areas.
- **Jumping stones (JS)** have been reported from minimum intensity IX. The size of stones and their imprint in soft soil are considered as diagnostic parameters for intensity evaluation.

Structure of the ESI-07 Scale

The ESI-07 scale (Fig. 1; Appendix 1) is structured in 12° and has been developed to be consistent with the modified Mercalli macroseismic scale (MM-31; Wood and Neumann 1931;

MM-56, Richter, 1958) and the MSK-64 (Medvedev-Sponheuer-Karnik Scale), since these are the most applied scales worldwide and include explicit references to environmental effects (especially in the highest degrees, X to XII). Comparing the ESI-07 with these other twelve-degree scales, we can identify four main intensity subsets as illustrated in Fig. 1:

From I to III: There are no environmental effects that can be used as diagnostic.

From IV to VII: Secondary environmental effects are easily observable starting from intensity IV, but only often permanent and diagnostic starting from intensity VII. Except for volcanic areas, this range of intensity degrees does not record primary effects. Within this range, effects on natural environment are less suitable for intensity assessment than effects on humans and man-made structures. Their use is therefore recommended especially in sparsely populated areas or in combination with damage-based intensity scales. Their use in field surveys for recent events will help to improve future versions of the scale. Environmental damage affects an area normally $\leq 10 \text{ km}^2$.

From VIII to X: In this range of intensities, primary and secondary environmental effects become characteristic and normally permanent and diagnostic. Thus, they are suitable for intensity assessment in combination with damage-based intensity scales but can be used as an independent tool, especially in uninhabited areas and for paleoseismic events. This range of intensity degrees is also a valuable tool in the analysis of well-documented historical earthquakes, refining intensity values in urban areas but mainly filling intensity data gaps among populated zones. Environmental damage typically affects areas ranging from 100 km^2 (VIII) to about $5,000 \text{ km}^2$ (X).

From X to XII: Effects on humans and anthropic structures saturate, while effects on natural environment, especially primary, become dominant and diagnostic for this range of intensity degrees; in fact, several types of

environmental effects do not suffer saturation in this range. Consequently earthquake environmental effects (EEEs) are the most effective tool to evaluate the intensity. As in the previous range of intensities, this one is fully applicable to well-documented historical events and also to paleoevents. Environmental damage can extend over areas ranging from 5,000 to 50,000 km².

Guidelines for the Application of the ESI-07 Scale

The ESI-07 scale might also provide a better estimation for intensity assessments than the traditional intensity scales, because earthquake environmental effects are independent from time (modern, historical, or paleo) and do not depend from the socioeconomic development of the region affected by the seismic event. This is true both:

(A) ***In time***, earthquake environmental effects are comparable for a time window much larger than the period of instrumental record (recent, historic, and paleoseismic events). The instrumental records cover about one century, actually much less, if we consider that the introduction of modern global seismic networks started about 50 years ago. As the impact of an earthquake on the artificial environment depends on the distribution and quality of urban buildings and infrastructures, today the comparison of a contemporary event with a historical one (i.e., seventeenth century A.D.) based only on damage indicators might reflect the level of economy in the study area more than the size of the earthquake. Furthermore, the ESI-07 scale allows the expansion of the temporal extent of earthquake catalogues (Silva et al. 2014). Evidence of surface faulting and secondary effects (i.e., Liquefaction, Landslides, Tsunamis) generated by prehistoric events can be identified via detailed paleoseismic investigations, and pertinent ESI-07 values can be derived and used for a

seismic hazard assessment over a geologic time interval.

(B) ***In different geographic areas***, earthquake environmental effects do not depend on particular socioeconomic conditions or different building practices of the damaged zones. In other words, intensity assessments based on building-damage macroseismic scales are commonly influenced by cultural and technological aspects, which may differ significantly from region to region. The ESI-07 scale is independent from these limitations. Moreover, earthquake-prone areas can be located completely or partially in sparsely populated regions, where the effects on the natural environment might be the only evidence available to estimate intensity.

In particular, the ESI-07 scale complements and/or replaces traditional macroseismic scales in the following aspects:

- (a) For earthquake intensity degrees $\geq X$, when damage-based assessments are extremely difficult because the built environment is virtually totally destroyed (so that any assessment based on it is not significant), while environmental effects are still diagnostic.
- (b) In sparsely populated areas, where man-made structures are absent or rare, so that only the environmental effects allow intensity estimates.
- (c) In paleoseismic analyses, where traditional macroseismic scales are not applicable. Additionally, the ESI-07 scale has an advantage over conventional paleoseismological research (i.e., fault-trenching), since the analysis of secondary environmental effects allows the investigation of earthquake cases where no surface faulting occurs, normally in the lower bound of the scale ($\leq VIII$). These kinds of events are commonly “lost earthquakes” for conventional fault-trenching surveys but source of significant building and environmental damage as illustrated by recent moderate seismic events such as the Lorca 2011 (Spain, 5.2 Mw) or the Emilia-Romagna 2012 (Italy, 6.1 Mw) earthquakes.

Intensity Scale ESI 2007 for Assessing Earthquake Intensities, Table 1 Ranges of surface faulting parameters (primary effects) and typical extent of total area affected by secondary effects for each intensity degree

I_0	Primary effects		Secondary effects
	Surface rupture length	Max surface displacement/deformation	Total area
IV	–	–	–
V	–	–	–
VI		–	$\leq 5 \text{ km}^2$
VII	^a	^a	10 km^2
VIII	$< 1 \text{ km}$	$< 5 \text{ cm}$	100 km^2
IX	$1\text{--}10 \text{ km}$	$5\text{--}40 \text{ cm}$	$1,000 \text{ km}^2$
X	$10\text{--}60 \text{ km}$	$40\text{--}300 \text{ cm}$	$5,000 \text{ km}^2$
XI	$60\text{--}150 \text{ km}$	$300\text{--}700 \text{ cm}$	$10,000 \text{ km}^2$
XII	$> 150 \text{ km}$	$> 700 \text{ cm}$	$> 50,000 \text{ km}^2$

^aLimited surface fault ruptures, 10–100 m long with centimeter-wide offset may occur in volcanic areas, generally associated to very shallow earthquakes (From Michetti et al. (2007))

The ESI-07 scale is an independent tool for intensity assessment. So, it can be used alone when only environmental effects are the diagnostic features present; this is the case when effects on man and man-made structures are too scarce or have saturated an area. Other than these cases, it is advisable to estimate two independent intensity fields, one for damage-based scales and one for ESI-07. Their subsequent comparison and integration, based on an expert's judgment, will lead to the best intensity scenario and the most reliable intensity estimate.

Generally, the epicentral region includes the area where the highest values of intensity are recorded. The epicentral intensity (I_0) is commonly close to, and usually corresponds with, the highest estimated value. Surface faulting parameters and the total spatial distribution of secondary effects (Landslides and/or Liquefactions) are two independent tools for assessing an ESI-07 I_0 , starting from intensity VII (Table 1). Of course, specific care has to be paid when surface faulting parameters are at the boundaries between two different degrees. In this case, the intensity value more consistent with the

characteristics and areal distribution of the secondary effects should be selected. Moreover, in the evaluation of the total area, it is recommended that isolated effects that occurred in the far field are not included. This evaluation also requires expert judgment.

Local intensity is generally evaluated through the description of secondary effects (EEEs) that have occurred in different “sites” included within a specific “locality.” In this way, a locality (i.e., a village, a cliff zone, a particular beach, etc.) may record EEE intensity data from different sites, and the eventual assessment of the local intensity is based on the EEE mean and maximum dimensions. This type of intensity has to be comparable with the corresponding traditional local intensity based on damage. It must be noted that a “locality” can be referred to an inhabited area (a village, a town) but also to natural areas without human settlements. When only primary effects are present, the local evidence of surface faulting in terms of maximum observed displacement may be used. Table 2 provides the diagnostic range of intensity in which primary and secondary EEEs can be recorded. The specific descriptions of EEEs for each intensity degree are given in Appendix 1.

Application of the ESI-07 Scale to Paleoseismic Data

Conventional paleoseismic research is focused on fault-trenching analyses, which involve the excavation of trenches of different dimensions across fault traces evidencing past coseismic surface ruptures. These kinds of analyses are focused on the identification, measurement, quantification, and geochronological dating of key stratigraphic features, indicating past surface faulting events (earthquake horizons). The final purpose is to establish the earthquake history of a fault, the seismic potential or capability to generate strong seismic events, and the evaluation of the recurrence periods of earthquakes along the fault (e.g., McCalpin 2009). Widely used empirical relations (e.g., Wells and Coppersmith 1994) among surface rupture length (SRL), maximum fault displacement (MD), and earthquake

Intensity Scale ESI 2007 for Assessing Earthquake Intensities, Table 2 Diagnostic ranges of intensity degrees for each category of environmental effects

Diagnostic range of intensity degrees of EEEs in the ESI-07 Macroseismic Scale											
Earthquake Environmental effects	IV	V	VI	VII	VIII	IX	X	XI	XII		
Surface Faulting and uplift/subsidence	Absent			VII ^a	VIII						XII
A (HA) Hydrological Anomalies	IV			VII				X			
B (AW) Anomalous Waves/Tsunamis	IV					IX					XII
C (GK) Ground Cracks	IV			X							
D (SM) Slope Movements	IV		X								
E (TR) Tree Shaking	IV				XI						
F (LQ) Liquefactions		V			X						
G (DC) Dust Clouds					VIII					XII	
H (JS) Jumping Stones	Absent					IX					XII

^aFor intensity degree VII, limited surface fault ruptures, tens to hundreds meters long with several centimeter-wide offset may occur essentially associated to very shallow earthquakes in volcanic areas. Green-shaded intensity zones represent “rarely to commonly observed” EEEs at indicated intensity levels. Orange-shaded zones represent intensity levels at which individual EEE types are “characteristic and diagnostic” to assess the intensity (Modified from Michetti et al. (2007))

moment magnitude (M_w) allow the estimation of this seismic size parameter. The application of the ESI-07 scale permits estimating the maximum epicentral intensity (I_0) from this kind of paleoseismic approach (see Table 1).

In zones where surface faulting is absent, the most helpful source of information is multiarchive paleoseismic records provided by a variety of secondary earthquake environmental effects, such as paleoliquefaction features and slope movements considered in the ESI-07 scale, but also lake seismites and speleoseismites. For example, the work of Becker et al. (2005) in the Swiss Alps provides good guidelines in the use of secondary EEEs in paleoseismic research. Accordingly, in a case where multiple paleo-earthquake evidence is available in a zone, integrated paleoseismological research is feasible from different geological archives (slope movements, liquefaction, lake, and cave geological records) to identify the seismic history of the zone. Depending on the size of the earthquake (magnitude), attenuation/amplification effects (intensities), and geographical distribution with

respect to the suspect macroseismic epicenter, a same geological archive may record evidence of a given earthquake but also of repeated events in the zone (e.g., Becker et al. 2005). The mapping and dating of paleoseismic features in different geological archives (EEEs) can provide sufficient information, making possible to apply the ESI-07 scale. In zones where surface faulting occurs, the combination of primary and secondary EEEs may offer the best scenario in order to reconstruct the seismic history of the zone, providing a reliable image of the potential seismic hazard.

Figure 2 illustrates a theoretical scenario for the paleoseismic record of EEEs in which it is possible to identify several paleo-earthquakes by means of the recognition and quantification of primary and secondary effects, allowing one to trace paleoseismic ESI-07 zones. Of course, the recognition and attribution of the observed EEEs in each site to a particular paleo-earthquake should be made on the basis of the geochronological dating of the observed features. For Holocene events, the most suitable dating technique when possible is radiocarbon dating (^{14}C). For lake seismites,



highest if the evidence from different geological archives and different sites is combined. In the absence of fault surface ruptures, the illustrated archives of secondary earthquake environmental effects can be used in a similar way (based on Becker et al. 2005)

effects in seismic hazard and seismic risk analyses. Recently, the emergence of ► [archeoseismology](#) leads to the ability to perform combined archeoseismological and paleoseismological analyses of ancient and historical earthquakes (Reicherter et al. [2009](#); Silva et al. [2011](#)).

Summary

The Environmental Seismic Intensity scale (ESI-07) is a twenty-first-century macroseismic scale based on the analysis of geological and environmental effects caused by the earthquakes. The ESI-07 scale represents a quantification of the natural effects considered in traditional

macroseismic scales such as Mercalli, MCS, MM, and MSK. It can be used alone but preferentially combined with the aforementioned macroseismic scales based on building damage. Some recent macroseismic practices in Europe (EMS-98) do not consider natural effects for intensity assessments, and consequently the application of the ESI-07 scale constitutes a relevant complementary approach. Research on earthquake environmental damage (Earthquake Environmental Effects: EEEs) is of use in populated areas and inhabited zones but also at different time windows, including those covering, instrumental, historical, ancient, and paleoseismic events. These latter effects are beyond range of the traditional macroseismic scales, and thus the ESI-07 constitutes one of the most useful tools to expand conventional seismic catalogues to prehistoric and geologic time scales, improving the knowledge on the seismic history of a region or particular tectonic structure and refine evaluations on recurrence periods. This new macroseismic scale considers both on-fault primary effects caused by surface faulting and uplift and a wide range of off-fault secondary ground effects triggered by seismic ground shaking (ground cracks, slope movements, liquefaction processes, hydrological anomalies, and tsunamis). All these primary and secondary EEEs are able to imprint permanent traces in the geological and geomorphological records, especially from intensities above VII–VIII. The comprehensive analysis of EEEs will allow a parameterization of past earthquakes as well as definition of the overall distribution of associated intensity zones, to be applied to future seismic building codes and seismic hazard analyses.

Appendix 1: Definition of Intensity Degrees of the Environmental Seismic Intensity Scale ESI-2007 (*In “italics” highlighted the diagnostic descriptions for different intensity degrees*)

From Intensities I to III

There are no environmental effects that can be used as diagnostic.

Intensity IV

Largely observed/first unequivocal effects in the environment

Primary effects: absent

Secondary effects

- (a) Rare small variations of the water level in wells and/or of the flow rate of springs are locally recorded, as well as extremely rare small variations of chemical-physical properties of water and turbidity in springs and wells, especially within large karstic spring systems, which appear to be most prone to this phenomenon.
- (b) In closed basins (lakes, even seas), seiches with height not exceeding a few centimeters may develop, commonly observed only by tidal gauges, exceptionally even by naked eye, typically in the far field of strong earthquakes. Anomalous waves are perceived by all people on small boats, few people on larger boats, and most people on the coast. Water in swimming pools swings and may sometimes overflow.
- (c) Hair-thin cracks (millimeter wide) might be occasionally seen where lithology (e.g., loose alluvial deposits, saturated soils) and/or morphology (slopes or ridge crests) are most prone to this phenomenon.
- (d) Exceptionally, rocks may fall and small landslides may be (re)activated, along slopes where the equilibrium is already near the limit state, e.g., steep slopes and cuts, with loose and generally saturated soil.
- (e) Tree limbs shake feebly.

Intensity V

Strong/marginal effects in the environment

Primary effects: absent

Secondary effects

- (a) Rare variations of the water level in wells and/or of the flow rate of springs are locally recorded, as well as small variations of chemical-physical properties of water and turbidity in lakes, springs, and wells.
- (b) In closed basins (lakes, even seas), seiches with height of decimeters may develop,

sometimes noted also by naked eye, typically in the far field of strong earthquakes. Anomalous waves up to several tens of cm high are perceived by all people on boats and on the coast. Water in swimming pools overflows.

- (c) Thin cracks (millimeter wide and several cm up to one meter long) are locally seen where lithology (e.g., loose alluvial deposits, saturated soils) and/or morphology (slopes or ridge crests) are most prone to this phenomenon.
- (d) Rare small rockfalls, rotational landslides, and slump earth flows may take place, along often but not necessarily steep slopes where equilibrium is near the limit state, mainly loose deposits and saturated soil. Underwater landslides may be triggered, which can induce small anomalous waves in coastal areas of sea and lakes.
- (e) Tree limbs and bushes shake slightly, very rare cases of fallen dead limbs and ripe fruit.
- (f) Extremely rare cases are reported of liquefaction (sand boil), small in size and in areas most prone to this phenomenon (highly susceptible, recent, alluvial and coastal deposits, near-surface water table).

Intensity VI

Slightly damaging/modest effects in the environment

Primary effects: absent

Secondary effects

- (a) Significant variations of the water level in wells and/or of the flow rate of springs are locally recorded, as well as small variations of chemical-physical properties of water and turbidity in lakes, springs, and wells.
- (b) Anomalous waves up to many tens of cm high flood in very limited areas nearshore. Water in swimming pools and small ponds and basins overflows.
- (c) *Occasionally, millimeter-centimeter wide and up to several meters long fractures are observed in loose alluvial deposits and/or saturated soils; along steep slopes or river-banks, they can be 1–2 cm wide. A few minor cracks develop in paved (either asphalt or stone) roads.*
- (d) Rockfalls and landslides with volume reaching ca. 10^3 m^3 can take place, especially where equilibrium is near the limit state, e.g., steep slopes and cuts, with loose saturated soil or highly weathered/fractured rocks. Underwater landslides can be triggered, occasionally provoking small anomalous waves in coastal areas of sea and lakes, commonly seen by instrumental records.
- (e) *Trees and bushes shake moderately to strongly; a very few treetops and unstable-dead limbs may break and fall, also depending on species, fruit load, and state of health.*
- (f) *Rare cases are reported of liquefaction (sand boil), small in size and in areas most prone to this phenomenon (highly susceptible, recent, alluvial and coastal deposits, near-surface water table).*

Intensity VII

Damaging/appreciable effects in the environment

Primary effects observed very rarely and almost exclusively in volcanic areas. Limited surface fault ruptures, tens to hundreds of meters long and with centimetric offset, may occur, essentially associated to very shallow earthquakes.

Secondary effects: *The total affected area is in the order of 10 km^2 .*

- (a) Significant temporary variations of the water level in wells and/or of the flow rate of springs are locally recorded. Seldom, small springs may temporarily run dry or appear. Weak variations of chemical-physical properties of water and turbidity in lakes, springs, and wells are locally observed.
- (b) Anomalous waves even higher than a meter may flood limited nearshore areas and damage or wash away objects of variable size. Water overflows from small basins and watercourses.
- (c) *Fractures up to 5–10 cm wide and up to hundred meters long are observed, commonly in loose alluvial deposits and/or saturated soils, rarely in dry sand, sand clay, and clay soil fractures, up to 1 cm wide.*

Centimeter-wide cracks are common in paved (asphalt or stone) roads.

- (d) Scattered landslides occur in prone areas, where equilibrium is unstable (steep slopes of loose/saturated soils), while modest rock-falls are common on steep gorges, cliffs. Their size is sometimes significant (10^3 – 10^5 m³); in dry sand, sand clay, and clay soil, the volumes are usually up to 100 m³. Ruptures, slides, and falls may affect riverbanks and artificial embankments and excavations (e.g., road cuts, quarries) in loose sediment or weathered/fractured rock. Significant underwater landslides can be triggered, provoking anomalous waves in coastal areas of sea and lakes, directly felt by people on boats and ports.
- (e) Trees and bushes shake vigorously, especially in densely forested areas, many limbs, and tops break and fall.
- (f) *Rare cases are reported of liquefaction, with sand boils up to 50 cm in diameter, in areas most prone to this phenomenon (highly susceptible, recent, alluvial and coastal deposits, near-surface water table).*

Intensity VIII

Heavily damaging/extensive effects in the environment

Primary effects: observed rarely

Ground ruptures (surface faulting) may develop, up to several hundred meters long, with offsets not exceeding a few cm, particularly for very shallow focus earthquakes such as those common in volcanic areas. Tectonic subsidence or uplift of the ground surface with maximum values on the order of a few centimeters may occur.

Secondary effects: The total affected area is in the order of 100 km².

- (a) Springs may change, generally temporarily, their flow rate, and/or elevation of outcrop. Some small springs may even run dry. Variations in water level are observed in wells. Weak variations of chemical-physical properties of water, most commonly temperature, may be observed in springs and/or wells.

Water turbidity may appear in closed basins, rivers, wells, and springs. Gas emissions, often sulfureous, are locally observed.

- (b) Anomalous waves up to 1–2 m high flood nearshore areas and may damage or wash away objects of variable size. Erosion and dumping of waste is observed along the beaches, where some bushes and even small weak-rooted trees can be eradicated and drifted away. Water violently overflows from small basins and watercourses.
- (c) *Fractures up to 50 cm wide and up to 100 m long are commonly observed in loose alluvial deposits and/or saturated soils; in rare cases, fractures up to 1 cm can be observed in competent dry rocks. Decimetric cracks common in paved (asphalt or stone) roads as well as small pressure undulations.*
- (d) Small to moderate (10^3 – 10^5 m³) landslides widespread in prone areas; rarely they can occur also on gentle slopes, where equilibrium is unstable (steep slopes of loose/saturated soils, rockfalls on steep gorges, coastal cliffs); their size is sometimes large (10^5 – 10^6 m³). Landslides can occasionally dam narrow valleys causing temporary or even permanent lakes. Ruptures, slides, and falls affect riverbanks and artificial embankments and excavations (e.g., road cuts, quarries) in loose sediment or weathered/fractured rock. Frequent occurrence of landslides under the sea level in coastal areas.
- (e) *Trees shake vigorously; branches may break and fall, even uprooted trees, especially along steep slopes.*
- (f) *Liquefaction may be frequent in the epicentral area, depending on local conditions; sand boils up to ca. 1 m in diameter; apparent water fountains in still waters; localized lateral spreading and settlements (subsidence up to ca. 30 cm), with fissuring parallel to waterfront areas (riverbanks, lakes, canals, seashores).*
- (g) *In dry areas, dust clouds may rise from the ground in the epicentral area.*
- (h) Stones and even small boulders and tree trunks may be thrown in the air, leaving typical imprints in soft soil.

Intensity IX

Destructive/effects in the environment are a widespread source of considerable hazard and become important for intensity assessment.

Primary effects: observed commonly

Ground ruptures (surface faulting) develop, up to a few km long, with offsets generally in the order of several cm. Tectonic subsidence or uplift of the ground surface with maximum values in the order of a few decimeters may occur.

Secondary effects: The total affected area is in the order of 1,000 km².

- (a) *Springs can change, generally temporarily, their flow rate and/or location to a considerable extent. Some modest springs may even run dry. Temporary variations of water level are commonly observed in wells. Water temperature often changes in springs and/or wells. Variations of chemical-physical properties of water, most commonly temperature, are observed in springs and/or wells. Water turbidity is common in closed basins, rivers, wells, and springs. Gas emissions, often sulfurous, are observed, and bushes and grass near emission zones may burn.*
- (b) *Meters high waves develop in still and running waters. In flood plains, water streams may even change their course, also because of land subsidence. Small basins may appear or be emptied. Depending on shape of sea bottom and coastline, dangerous tsunamis may reach the shores with run-ups of up to several meters flooding wide areas. Wide-spread erosion and dumping of waste is observed along the beaches, where bushes and trees can be eradicated and drifted away.*
- (c) *Fractures up to 100 cm wide and up to 100 m long are commonly observed in loose alluvial deposits and/or saturated soils; in competent rocks, they can reach up to 10 cm. Significant cracks common in paved (asphalt or stone) roads, as well as small pressure undulations.*
- (d) *Landslides are widespread in prone areas and also on gentle slopes; where equilibrium is unstable (steep slopes of loose/saturated soils; rockfalls on steep gorges, coastal cliffs), their size is frequently large (10^5 m^3),*

sometimes very large (10^6 m^3). Landslides can dam narrow valleys causing temporary or even permanent lakes. Riverbanks, artificial embankments, and excavations (e.g., road cuts, quarries) frequently collapse. Frequent large landslides under the sea level in coastal areas.

- (e) *Trees shake vigorously; branches and thin tree trunks frequently break and fall. Some trees might be uprooted and fall, especially along steep slopes.*
- (f) *Liquefaction and water upsurge are frequent; sand boils up to 3 m in diameter; apparent water fountains in still waters; frequent lateral spreading and settlements (subsidence of more than ca. 30 cm), with fissuring parallel to waterfront areas (riverbanks, lakes, canals, seashores).*
- (g) *In dry areas, dust clouds commonly rise from the ground.*
- (h) *Small boulders and tree trunks may be thrown in the air and move away from their site for meters, also depending on slope angle and roundness, leaving typical imprints in soft soil.*

Intensity X

Very destructive/effects on the environment become a leading source of hazard and are critical for intensity assessment.

Primary effects become leading.

Surface faulting can extend for few tens of km, with offsets from tens of cm up to a few meters. Gravity grabens and elongated depressions develop; for very shallow focus, earthquakes in volcanic areas rupture lengths might be much lower. Tectonic subsidence or uplift of the ground surface with maximum values in the order of few meters may occur.

Secondary effects: The total affected area is in the order of 5,000 km².

- (a) *Many springs significantly change their flow rate and/or elevation of outcrop. Some springs may run temporarily or even permanently dry. Temporary variations of water level are commonly observed in wells. Even strong variations of chemical-physical*

properties of water, most commonly temperature, are observed in springs and/or wells. Often water becomes very muddy in even large basins, rivers, wells, and springs. Gas emissions, often sulfurous, are observed, and bushes and grass near emission zones may burn.

- (b) *Meters high waves develop in even big lakes and rivers, which overflow from their beds. In flood plains, rivers may change their course, temporarily or even permanently, also because of widespread land subsidence. Basins may appear or be emptied. Depending on the shape of sea bottom and coastline, tsunamis may reach the shores with run-ups exceeding 5 m flooding flat areas for thousands of meters inland. Small boulders can be dragged for many meters. Widespread deep erosion is observed along the shores, with noteworthy changes of the coastline profile. Trees nearshore are eradicated and drifted away.*
- (c) *Open ground cracks up to more than 1 m wide and up to 100 m long are frequent, mainly in loose alluvial deposits and/or saturated soils; in competent rocks they can reach several decimeters. Wide cracks develop in paved (asphalt or stone) roads, as well as pressure undulations.*
- (d) *Large landslides and rockfalls ($>10^5$ – 10^6 m³) are frequent, practically regardless of equilibrium state of the slopes, causing temporary or permanent barrier lakes. Riverbanks, artificial embankments, and sides of excavations typically collapse. Levees and earth dams may even incur serious damage. Frequent large landslides under the sea level in coastal areas.*
- (e) *Trees shake vigorously; many branches and tree trunks break and fall. Some trees might be uprooted and fall.*
- (f) *Liquefaction, with water upsurge and soil compaction, may change the aspect of wide zones; sand volcanoes even more than 6 m in diameter; vertical subsidence even >1 m; large and long fissures due to lateral spreading are common.*
- (g) *In dry areas, dust clouds may rise from the ground.*
- (h) *Boulders (diameter in excess of 2–3 m) can be thrown in the air and move away from their site for hundreds of meters down even gentle slopes, leaving typical imprints in soil.*

Intensity XI

Devastating/effects on the environment become decisive for intensity assessment, due to saturation of structural damage.

Primary effects: dominant

Surface faulting extends from several tens of km up to more than 100 km, accompanied by offsets reaching several meters. Gravity graben, elongated depressions, and pressure ridges develop. Drainage lines can be seriously offset. Tectonic subsidence or uplift of the ground surface with maximum values in the order of numerous meters may occur.

Secondary effects: The total affected area is in the order of 10,000 km².

- (a) *Many springs significantly change their flow rate and/or elevation of outcrop. Many springs may run temporarily or even permanently dry. Temporary or permanent variations of water level are generally observed in wells. Even strong variations of chemical-physical properties of water, most commonly temperature, are observed in springs and/or wells. Often water becomes very muddy in even large basins, rivers, wells, and springs. Gas emissions, often sulfurous, are observed, and bushes and grass near emission zones may burn.*
- (b) *Large waves develop in big lakes and rivers, which overflow from their beds. In flood plains, rivers can change their course, temporarily or even permanently, also because of widespread land subsidence and landsliding. Basins may appear or be emptied. Depending on shape of sea bottom and coastline, tsunamis may reach the shores with run-ups reaching 15 m and more devastating flat areas for kilometers inland. Even meter-sized boulders can be dragged for long distances. Widespread deep erosion is observed along the shores, with noteworthy changes of the coastal morphology. Trees nearshore are eradicated and drifted away.*

- (c) Open ground cracks up to several meters wide are very frequent, mainly in loose alluvial deposits and/or saturated soils. In competent rocks, they can reach 1 m. Very wide cracks develop in paved (asphalt or stone) roads, as well as large pressure undulations.
- (d) *Large landslides and rockfalls ($>10^5$ – 10^6 m³) are frequent, practically regardless to equilibrium state of the slopes, causing many temporary or permanent barrier lakes. Riverbanks, artificial embankments, and sides of excavations typically collapse. Levees and earth dams incur serious damage. Significant landslides can occur at 200–300 km distance from the epicenter. Frequent large landslides under the sea level in coastal areas.*
- (e) *Trees shake vigorously; many branches and tree trunks break and fall. Many trees are uprooted and fall.*
- (f) *Liquefaction changes the aspect of extensive zones of lowland, determining vertical subsidence possibly exceeding several meters, numerous large sand volcanoes, and severe lateral spreading features.*
- (g) In dry areas, dust clouds arise from the ground.
- (h) *Big boulders (diameter of several meters) can be thrown in the air and move away from their site for long distances down even gentle slopes, leaving typical imprints in soil.*

Intensity XII

Completely devastating/effects in the environment are the only tool for intensity assessment.

Primary effects: dominant

Surface faulting is at least few hundreds of km long, accompanied by offsets reaching several tens of meters. Gravity graben, elongated depressions, and pressure ridges develop. Drainage lines can be seriously offset. Landscape and geomorphological changes induced by primary effects can attain extraordinary extent and size (typical examples are the uplift or subsidence of coastlines by several meters, appearance or disappearance from sight of significant landscape elements, rivers changing course, origination of waterfalls, formation or disappearance of lakes).

Secondary effects: The total affected area is in the order of 50,000 km² and more.

- (a) Many springs significantly change their flow rate and/or elevation of outcrop. Temporary or permanent variations of water level are generally observed in wells. Many springs and wells may run temporarily or even permanently dry. Strong variations of chemical-physical properties of water, most commonly temperature, are observed in springs and/or wells. Water becomes very muddy in even large basins, rivers, wells, and springs. Gas emissions, often sulfurous, are observed, and bushes and grass near emission zones may burn.
- (b) *Giant waves develop in lakes and rivers, which overflow from their beds. In flood plains, rivers change their course and even their flow direction, temporarily or even permanently, also because of widespread land subsidence and landsliding. Large basins may appear or be emptied. Depending on the shape of sea bottom and coastline, tsunamis may reach the shores with run-ups of several tens of meters devastating flat areas for many kilometers inland. Big boulders can be dragged for long distances. Widespread deep erosion is observed along the shores, with outstanding changes of the coastal morphology. Many trees are eradicated and drifted away. All boats are tore from their moorings and swept away or carried onshore even for long distances. All people outdoor are swept away.*
- (c) Ground open cracks are very frequent, up to one meter or more wide in the bedrock, up to more than 10 m wide in loose alluvial deposits and/or saturated soils. These may extend up to several kilometers in length.
- (d) *Large landslides and rockfalls ($>10^5$ – 10^6 m³) are frequent, practically regardless to equilibrium state of the slopes, causing many temporary or permanent barrier lakes. Riverbanks, artificial embankments, and sides of excavations typically collapse. Levees and earth dams incur serious damage. Significant landslides can occur at more than*

200–300 km distance from the epicenter. Frequent very large landslides under the sea level in coastal areas

- (e) Trees shake vigorously; many branches and tree trunks break and fall. Many trees are uprooted and fall.
- (f) *Liquefaction occurs over large areas and changes the morphology of extensive flat zones, determining vertical subsidence exceeding several meters, widespread large sand volcanoes, and extensive severe lateral spreading features.*
- (g) In dry areas, dust clouds arise from the ground.
- (h) *Also very big boulders can be thrown in the air and move for long distances even down very gentle slopes, leaving typical imprints in soil.*

Cross-References

- ▶ [Archeoseismology](#)
- ▶ [Earthquake Location](#)
- ▶ [European Structural Design Codes: Seismic Actions](#)
- ▶ [Luminescence Dating in Paleoseismology](#)
- ▶ [Paleoseismology and Landslides](#)
- ▶ [Paleoseismology: Integration with Seismic Hazard](#)
- ▶ [Paleoseismic Trenching](#)
- ▶ [Paleoseismology](#)
- ▶ [Probabilistic Seismic Hazard Models](#)
- ▶ [Radiocarbon Dating in Paleoseismology](#)
- ▶ [Seismic Actions Due to Near-Fault Ground Motion](#)
- ▶ [Seismic Risk Assessment, Cascading Effects](#)
- ▶ [Time History Seismic Analysis](#)
- ▶ [Tsunamis as Paleoseismic Indicators](#)

References

- Ali Z, Qaisar M, Mahmood T, Shah MA, Iqbal T, Serva L, Michetti AM, Burton PW (2009) The Muzaffarabad, Pakistan, earthquake of 8 October 2005: surface faulting, environmental effects and macroseismic intensity. *Geol Soc London Spec Publ* 316:155–172
- Audemard FA, Michetti AM (2011) Geological criteria for evaluating seismicity revisited: forty Years of paleoseismic investigations and the natural record of past earthquakes. *Geological Society of America special paper*, Washington, USA, 479. pp 1–22
- Becker A, Ferry M, Monecke K, Schnellmann M, Giardini D (2005) Multiarchive paleoseismic record of late Pleistocene and Holocene strong earthquakes in Switzerland. *Tectonophysics* 400(1–4):153–177
- Berzhinskii YA, Ordynskaya AP, Gladkov AS, Lunina OV, Berzhinskaya LP, Radziminovich NA, Radziminovich YB, Imayev VS, Chipizubov AV, Smekalin OP (2010) Application of the ESI_2007 scale for estimating the intensity of the Kultuk earthquake, August 27, 2008 (South Baikal). *Seismic Instruments* 46(4):307–324
- Dengler L, McPherson R (1993) The 17 August 1991 Honeydew earthquake north coast California: a case for revising the Modified Mercalli Scale in sparsely populated areas. *Bull Seismol Soc Am* 83(4):1081–1094
- Di Manna P, Guerrieri L, Piccardi L, Vittori E, Castaldini D, Berlusconi A, Michetti AM (2013) Ground effects induced by the 2012 seismic sequence in Emilia: implications for seismic hazard assessment in the Po Plain. *Ann Geophys* 55(4):727–733
- Dowrick DJ (1996) The modified Mercalli earthquake intensity scale- revisions arising from recent studies of New Zealand earthquakes. *Bull New Zealand Nat Soc Earthq Eng* 29(2):92–106
- Fountoulis IG, Mavroulis SD (2013) Application of the environmental seismic intensity scale (ESI 2007) and the European Macroseismic Scale (EMS-98) to the Kalamata (SW Peloponnese, Greece) earthquake ($M_s = 6.2$, September 13, 1986) and correlation with neotectonic structures and active faults. *Ann Geophys* 56(6):S0675
- Gojsar A (2012) Application of environmental seismic intensity scale (ESI 2007) to Krn Mountains 1998 $M_w = 5.6$ earthquake (NW Slovenia) with emphasis on rockfalls. *Nat Hazards Earth Syst Sci* 12:1659–1670
- Grunthal G (1998) European macroseismic scale 1998 (EMS-98). European Seismological Commission, Subcommission on Engineering Seismology, Working Group Macroseismic Scales, Conseil de l'Europe, Cahiers du Centre Européen de Géodynamique et de Séismologie, 15, Luxembourg, 99 pp
- Guerrieri L, Blumetti AM, Esposito E, Michetti AM, Porfido S, Serva L, Tondi E, Vittori E (2008) Capable faulting, environmental effects and seismic landscape in the area affected by the 1997 Umbria-Marche (Central Italy) seismic sequence. *Tectonophysics* 476(1–2):269–281
- Hancox GT, Perrin ND, Dellow GD (2002) Recent studies of historical earthquake-induced landsliding, ground damage, and MM intensity in New Zealand. *Bull New Zealand Nat Soc Earthq Eng* 35(2):59–95
- Lalinde CP, Sanchez JA (2007) Earthquake and environmental effects in Colombia in the last 35 years. INQUA Scale Project. *Bull Seism Soc Am* 97(2):646–654

- Lander JF, Whiteside LS, Lockridge PA (2003) Two decades of global Tsunamis 1982–2002. *The International Journal of the Tsunami Society* (21–1), NOAA – NGDC, Boulder, 73 pp
- Mavroulis SD, Fountoulis IG, Skourtsos EN, Lekkas EL, Papanikolaou I (2013) Seismic intensity assignments for the 2008 Andravida (NW Peloponnese, Greece) strike-slip event (June 8, Mw 6.4) based on the application of the Environmental Seismic Intensity scale (ESI 2007) and the European Macroseismic scale (EMS-98). Geological structure, active tectonics, earthquake environmental effects and damage pattern. *Ann Geophys* 56(6):S0681. doi:10.4401/ag-6239
- McCalpin JP (2009) *Paleoseismology*, 2nd edn. Elsevier, Amsterdam, 613 pp
- Michetti AM, Esposito E, Gürpınar A, Mohammadioun B, Mohammadioun J, Porfido S, Roghazin E, Serva L, Tatevossian R, Vittori E, Audemard F, Comerci V, Marco S, McCalpin J, Mörmner NA (2004) The INQUA scale. An innovative approach for assessing earthquake intensities based on seismically-induced ground effects in natural environment. In: Vittori E, Comerci V (ed) *Memorie Descrittive della Carta Geologica d'Italia*, APAT, Roma, LXVII-116
- Michetti AM, Esposito E, Guerrieri L, Porfido S, Serva L, Tatevossian R, Vittori E, Audemard F, Azuma T, Clague J, Comerci V, Gürpınar A, Mc Calpin J, Mohammadioun B, Mörmner NA, Ota Y, Roghazin E (2007) Intensity scale ESI 2007. In: Vittori E, Vittori E (ed) *Mem. Descr. Carta Geologica d'Italia*, Servizio Geologico d'Italia, Dipartimento Difesa del Suolo, APAT, Rome, 74p
- Musson RMW, Grunthal G, Stucchi M (2010) The comparison of macroseismic intensity scales. *J Seismol* 14:413–428
- Ota Y, Azuma T, Lin N (2009) Application of INQUA environmental seismic intensity scale to recent earthquakes in Japan and Taiwan. *Geol Soc London Spec Publ* 316:55–71
- Papadopoulos G, Imamura F (2001) A proposal for a new tsunami intensity scale, ITS 2001 Proceedings 20th International Tsunami Conference, Seattle, USA. Session 5–1. pp. 569–577
- Papanikolaou ID (2011) Uncertainty in intensity assignment and attenuation relationships: how seismic hazard maps can benefit from the implementation of the environmental seismic intensity scale (ESI 2007). *Quatern Int* 242:42–51
- Papathanassiou G, Pavlides S (2007) Using the INQUA scale for the assessment of intensity: case study of the 2003 Lefkada (Ionian Islands), Greece earthquake. *Quatern Int* 173–174:4–14
- Reicherter K, Michetti AM, Silva PG (eds) (2009) *Palaeoseismology: historical and prehistorical records of earthquake ground effects for seismic hazard assessment*. Geological Society of London, Special Publications, 316. London, 160 pp
- Richter CF (1958) *Elementary Seismology*. Freeman & Co. San Francisco, USA. 768p
- Serva L (1994) The effects on the ground in the intensity scales. *Terra Nova* 6:414–416
- Serva L, Esposito E, Guerrieri L, Porfido S, Vittori E, Comerci V (2007) Environmental effects from some historical earthquakes in Southern Apennines (Italy) and macroseismic intensity assessment. Contribution to INQUA EEE scale project. *Quatern Int* 173–174:30–44
- Sieberg A (1930) *Geologie der Erdbeben*. Handbuch der Geophysic 4:552–554
- Silva PG, Sintubin M, Reicherter K (eds) (2011) *Earthquake archaeology and palaeoseismology*. Quaternary International, vol 241(1). Elsevier, Amsterdam, 258 pp
- Silva PG, Rodríguez-Pascua MA, Giner JL, Pérez-López R, Lario J, Perucha MA, Bardají T, Huerta P, Roquero E, Bautista B (2014) *Catálogo de los efectos geológicos de los terremotos en España*. Publications Geological Survey of Spain (IGME) Serie Riesgos Geológicos/Geotécnia, 4. IGME, Madrid, 358 pp
- Tatevossian RE (2007) The Verny, 1887, Earthquake in Central Asia: application of the INQUA scale based on coseismic environmental effects. Dark nature: rapid environmental change and human response. *Quatern Int* 173–174:23–29
- Wells LD, Coppersmith JK (1994) New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bull Seismol Soc Am* 84:974–1002
- Wood HO, Neumann F (1931) Modified Mercalli intensity scale of 1931. *Bull Seismol Soc Am* 21(4):277–283

Interim Housing Provision Following Earthquake Disaster

Regan Potangaroa
Unitec School of Architecture, Auckland,
New Zealand

Synonyms

Core housing; Housing; Recovery phase; Shelter; Transitional housing

Introduction

The provision of interim housing following an earthquake disaster is important because it is the first substantial community step away from the effects of a sudden and unexpected seismic event and back towards some “normality.”

It is defined as housing that is “short term that can be in place for 1 month to 2 years – preferably on sites close to the damaged housing. (To the extent that displaced residents can remain near their previous homes means that economic impacts are reduced.) Interim housing will be needed because those organizations “donating” temporary shelter space need to reclaim that space for its original uses as schools – community centers – churches – that serve the community.” (ABAG 2008).

The idea of “shelter in place” is important not only from a psychosocial perspective of those affected but moreover as part of the subsequent recovery and reconstruction. Shelter in place will encourage local businesses to stay; allow families to return to local schools, friends, social networks, and places; and allow for the restoration of infrastructure based on known numbers rather than estimates or “best guesses.” Thus, it can be the trigger for an early resilient community and city wide recovery.

Interim housing will involve strategic planning and the allocation and use of significant resources, but it is the processes and priorities to be used that have been identified as “problematic” (EERI 2004; Davidson et al. 2007; Johnson 2007; Giovinazzi and Stevenson 2012).

Some of these processes and priorities are evidenced by the “philosophical-one-liners” associated with the operational side of interim housing such as “community engagement” (discussed by Davidson et al. above), “build back better” (Lyons et al. 2010), the notion of disasters as “opportunities,” and of a “one size fits all” or “cookie cutter” housing solutions (Jha et al. 2010). There are several rights-based issues relating to housing land and property that emerge post disaster (IASC 2011), and there are philosophic positions such as the “do no harm principle” (Shelter Cluster 2014) and the “low-hanging fruit” approach where you assist those that are easiest first rather than those based on need. All balanced against the SPHERE standards in humanitarian assistance (SPHERE 2011) or National codes and the political positions of Government departments in developed economies. These all serve to underline its importance but

also suggest a complexity that goes beyond “business as usual.”

The provision of interim housing following an earthquake has several similarities to other fast onset disasters such as typhoons and flooding (Johnson 2002). However, the seismic design and detailing of buildings seems to demand greater technical skills, and as noted by da Silva following the Asian Tsunami disaster in Banda Aceh, Indonesia (da Silva 2010), “. . . a key issue was confusion as to what codes and standards should be followed. . . . Several agencies complied with the Building Code for Aceh assuming that it was sufficient or that local designers and contractors knew what they were doing without realising that safe construction practices were not common practice. Local engineering consultants employed by implementing agencies to develop structural designs generally had limited experience of seismic design, which typically requires an additional post-graduate qualification. This resulted in poor design solutions which were not compliant with the Indonesian code. Recognising this, some agencies employed specialist International consultants or firms to develop or check designs, or sought advice from local and national universities. International engineers were also employed as consultants in-house. However, many of these engineers did not previously have seismic design experience and so were ascending a learning curve, trying to follow available guidance and incorporate it into the construction drawings.”

This still remains an area of confusion for aid agencies (Potangaroa 2010a); and is highlighted in the later case studies.

So how does a building professional survive, let alone remain operationally focused in such an environment? This section considers some of the theoretical context to interim housing and then discusses several case studies from recent earthquakes through the following “lens” (adapted from Davis 2010):

- Planning strategy
- Cost
- Connection to subsequent stages
- The meaning of place



Interim Housing Provision Following Earthquake Disaster, Fig. 1 Examples of interim housing after an earthquake disaster

- Seismic design
- Climatic design
- Community engagement

It then comes back to consider the potential role of building professionals in the provision of interim housing.

Background

The provision of interim housing following an earthquake seems to have the following patterns (Quarantelli 1995) or phases (MCDem 2004) as shown below in Table 1.

The above patterns are perhaps indicative of the operational level of shelter/housing provision, while the phases are more policy and strategic. And, for example, the response phase also includes rapid shelter assessments to

ascertain the “what,” the “where,” and the “who” (the so-called 3W analysis) so that any later housing assistance (including interim housing) can be more effective, can be better targeted, and have a level of accountability; all of which must be updated as the earthquake disaster response and reconstruction develops.

The recovery phase also prioritizes the economic impacts and the apparent need to get businesses back up and running. But as noted earlier, the decisions around housing and where people will stay should preempt such a priority; otherwise, there may not be an available “local” market or the required work force, whether a business has local or national markets.

The word coordination is increasingly used, but the dilemma of the recovery phase (and possibly the provision of interim housing) has been well framed by the Natural Hazards Center comments “How does one make a holistic disaster

Interim Housing Provision Following Earthquake Disaster, Table 1 The patterns and phases of interim housing following an earthquake disaster

Pattern	Measured in . . .	Objectives and locations	Phase
Spontaneous shelter ^a	Days	To provide an interim, safe haven while the situation stabilizes. This could be a school, church/mosque, stadium, park, community hall, or other public facility. Sometimes referred to as evacuation centers. It could also be the home of a friend or family member. It is often on the site of the original house such as in a self-built shelter made from debris material, a camping or basic tent, the garage, or in an undamaged part of the house. There is no outside aid such as food or shelter materials at this point	Response
Emergency or temporary shelter ^a	Months	To provide emergency shelter and contact points for Government and aid agencies supplying food, shelter materials, water, medical/health, and basic supplies such as soaps, blankets, etc. It can be a tent, a larger self-built shelter than before, a public facility, the home of family or friends, hotel, camping ground, park, or a second home. The length varies but will determine the scale of assistance and services required	Response
Interim housing ^a	Years	To provide safe and secure shelter/housing, water, and electricity so that some resumption of family life can be achieved, while efforts are under way to determine and construct permanent, sustainable housing options. This may involve repairs to original housing, the construction of new dwellings, insurance determinations, or relocation options or to find other suitable permanent housing. This can be a prefabricated temporary house, specifically built barracks, camp or motor camp area, a winterized tent, a self-built shelter, a mobile home, an apartment, or the home of family member or friend: using either a transitional or core design approach	Recovery
Permanent housing ^a	Generations	To provide long-term, permanent housing solutions for those affected. This is usually on the original house site but can be within or nearby the community. A final option would be to relocate out of the original community, and this is particularly so where other long-term services such as education/schools are required	Reconstruction

^aThe resumption of family activities is noted as the difference between shelter and housing, and where possible, the difference is noted in the text

recovery happen? How can a decision maker reshape a process that operates within an emotional, reactionary, time-sensitive, expensive, and politically charged atmosphere that is based upon incomplete information, disproportionate needs, and the worst working conditions imaginable?” (Natural Hazards Center 2001).

Davis is more specific with his comments “Finally, a question that arises in the early stages of almost every recovery programme is whether reconstruction and the wider aspects of community rehabilitation should be conducted in the same, original, disaster-prone location, or

whether the population should be encouraged or required to relocate to a new and possibly less vulnerable location.”(Davis 2007).

And it is against this background that the provision of interim housing occurs.

But it gets worse. Strangely, it is not clear who is responsible for interim housing. In a developed economy, the response phase shelter is usually taken up by the Red Cross (ARC 2012), civil society, NGOs, and faith-based organizations, but it is not completely clear who is then responsible for any subsequent interim housing in the recovery phase. Governments do set up specific

commissions, but the basis and objectives of them have been historically questioned (Hewitt 2013). This disconnection was evident in case studies 2 and 3.

A similar situation exists in developing countries and with humanitarian assistance under a cluster approach (OCHA 2012) where the response phase (tents and tarpaulins) would be taken up by the Emergency Shelter and Non-Food Items (NFI) Cluster led by either the International Federation of Red Cross and Red Crescent Societies (IFRC) for natural disasters or United Nations High Commissioner for Refugees (UNHCR) in conflict-based situations. However, the interim housing is subsequently spread across the two clusters of early recovery and camp management and is potentially involved in the WASH (water, sanitation, and hygiene promotion), Protection, and Early Education clusters. Interim housing is essentially an “orphan.”

Timelines for Interim Housing

The time it takes to provide interim housing is often/always underestimated, but that should not be the case. A model developed in the 1970s by Haas suggests the following (Haas et al. 1977):

- If the time it takes to complete the response phase is 1, then:
- The recovery phase will be 10 times.
- The reconstruction phase will be 100 times.

These are approximations, and there are certainly difficulties in ascertaining when the response phase is actually complete. But as reported by Zuo, the approach is widely accepted and quoted which he details (Zuo 2010). It does provide a useful operational and planning base that is firstly “sobering” but also one that secondly better reflects the actual times (rather than the planned ones) allowed for earthquake disaster responses.

For example, the 22 Feb 2011 Christchurch Earthquake based on a response time of around 3 months would suggest the following timeline:

- Response phase of 3 months: 22 February 2011 to 22 May 2011

- Recovery phase of 30 months: May 2011 to November 2013
- Reconstruction phase of 300 months: November 2013 to November 2037.

This timeline would be difficult for any public authority or the Government to acknowledge and especially present to its electorate. The result is the conviction that all must be OK and a filtering of information to make it fit. Consequently, the Government is communicating exactly that message, that all is well and under control. On one hand, a local respected NGO called CanCERNs has initiated the 900 Project to identify what they suspect are the remaining 900 uninhabitable houses in Christchurch 3+ years after the earthquake and hence possibly out of control (CanCERN 2014).

Such timelines suggested by Haas are unfortunately ignored by Governments and aid agencies for various reasons (Lloyd-Jones 2006). However, on the flip side, it does suggest that any reduction of the response phase will have huge reductions for the subsequent recovery and reconstruction phases. Therefore, the planning and implementation of interim housing should start immediately and not wait till after the response phase has been completed. This “lesson learnt” seems to be repeated in subsequent seismic disasters.

Costs of Interim Housing

Typically, the costs for interim housing can be comparable or at least significant against the cost of permanent housing. For this reason, attempts have been made to miss out the interim housing and go more directly to permanent housing (as in the Banda Aceh case study). Alternatively, humanitarian and aid agencies have developed transitional and core housing approaches that uses any interim house as part of a permanent housing development thus “saving” on overall cost. There are additional political and planning reasons for this approach as it allows for more permanent outcomes for those affected based on an apparent interim solution which is politically expedient when those being assisted may be settled informally and do not have any formal land

title and have been that way for some time before the earthquake.

Transitional and Core Approaches to Interim Housing

Transitional housing is where an interim house is designed to merge into a permanent one. This usually means that the structural system (and in particular the lateral load resisting part) is retained and the cladding usually changed over time. This in part seeks to save overall housing costs but more importantly to minimize the emotional impacts on families having to relocate, yet again. It can also make the shelter-in-place option more viable though it has been used in off-site as a prefabricated kitset that goes with the occupants. However, its big advance is that the occupants can upgrade as money and resources become available, and perhaps other family needs arise that require more space such as births and marriages. Typically, the transitional house could be supplied clad in plastic tarpaulins, which can be later replaced with mud bricks, mud-plastered stone, or cement-plastered concrete blocks. The design trick is to structurally ensure that the lateral load resisting as supplied has sufficient capacity for this anticipated loading of a heavier cladding.

A core house on the other hand is where some part of a permanent house is supplied which is then added and extended over time as money and resources become available.

The objective of both approaches is to “transition” the family/occupants into a more permanent or “durable” housing solution based on their priorities (rather than the aid agency) under their own direction. Such approaches are more readily found with aid agencies and in developing economies rather than Governments and developed economies despite their appropriateness for both.

Codes and Planning for Interim Housing

The level and criteria for code compliance for interim housing in post-earthquake situations are contentious. On one hand is whether there should be any/minimal code and planning compliance given the urgency of the need, the

temporaryness of the building, and the often dire living conditions of those affected, while on the other hand, the ongoing seismic activity, the relative expense of the interim house, and its possibly semi-permanence suggest that it fully complies with the stipulated country and possibly international codes. Moreover, if these are being done in large numbers (rather than as a one off), it seems to support full code compliance. There is also the issue of their disposal once they have done their job. This can be more difficult than expected (Arslan and Coşgun 2007).

Most developed economies do not have specific codes or planning procedures for interim housing, and because it usually occurs outside, any declared state of emergency will often require some type of executive order. In Christchurch, for example, specific guidance notes were compiled to facilitate the compliance and provision of post-disaster worker accommodation (MBIE 2012). FEMA on the other hand has a suite of interim house types that it can use (DHS 2013) which have had issues in the past and would be pushed to respond in a timely manner to a disaster (McIntosh 2013). There are various guidelines for short-term/interim housing for the military, for farm workers, and for the homeless (USAF 2001; QG 2010; SAMB 2009; Wilson 2013). Strangely, there is minimal design code or guidance post disaster and seemingly nothing that specifies the roles and responsibilities of local authorities or Government in terms of interim housing of affected people.

Moreover, the issue for the worker accommodation built in Christchurch will be what happens to them when they reach their use by date of 31 December 2022. Will the worker’s accommodation be demolished or simply on sold possibly creating substandard long-term housing somewhere else? Or will the date simply be extended?

In developing economies and the humanitarian aid sector, there are two “codes of compliance.” In a refugee situation, it is the UNHCR’s Handbook (UNHCR 2007), while in natural disasters, it is SPHERE (2011).

The UNHCR Handbook was developed from experience in the field and was widely accepted as the de facto standard and code even outside the

refugee situation. However, its “camp” focus meant that it increasingly did not translate well into non-refugee situations. And hence the SPHERE Project was initiated to set up universal minimum standards in core areas of humanitarian response. A lack of such standards has in the past resulted in “competition” between shelter agencies. Though it was started in 1997 (with the first trial edition in 1998), the shelter response in Banda Aceh after the 2004 Asian Tsunami is often remembered because of the competition that developed. Shelter options were seemingly raked up with one agency offering a house, then another offering a house and \$2,000USD, and the next a house fully furnished and later with an opening party for friends.

The size and specifications of temporary shelter also generated unintentional competition between shelter agencies, and, for example, the “China tent” in Pakistan following the 2005 Kashmir Earthquake was the “tent of choice” firstly because it had walls which meant that one could stand up anywhere inside it but secondly because it had a steel frame which performed better under snow loading. SPHERE has been revised in 2000, 2004, and in the current green-colored version of 2011. The key quantitative shelter points are the following:

- Site slope should neither exceed 5 %, unless extensive drainage and erosion control measures are taken, nor be less than 1 % to provide for adequate drainage. The lowest point of the site should be not less than 3 m above the estimated maximum level of the water table.
- There should be a minimum usable surface area of 45 m² for each person when including household plots/gardens or 30 m² otherwise. And where this cannot be provided, the consequences of higher-density occupation should be mitigated, for example, through ensuring adequate separation and privacy between individual households, space for the required facilities, etc.
- There should be a 30 m firebreak for every 300 m of built-up area and a minimum of 2 m (but preferably twice the overall height of any structure) between individual buildings or

shelters to prevent collapsing structures from touching adjacent buildings.

- In cold climates, household activities typically take place within the covered area, and affected populations may spend substantial time inside to ensure adequate thermal comfort. In urban settings, household activities typically occur within the covered area as there is usually less adjacent external space that can be used.
- A covered floor area in excess of 3.5 m² per person will often be required to meet these considerations.
- The internal floor-to-ceiling height should be a minimum of 2 m at the highest point. In warmer climates, the adjacent shaded external space can be used for food preparation and cooking. Where materials for a complete shelter cannot be provided, roofing materials to provide the minimum covered area should be prioritized. The resulting enclosure may not provide the necessary protection from the climate nor security, privacy, and dignity, so steps should be taken to meet these needs as soon as possible.
- Standards and guidelines on construction should be agreed with the relevant authorities to ensure that key safety and performance requirements are met. Where applicable local or national building codes have not been customarily adhered to or enforced, incremental compliance should be agreed, reflecting local housing culture, climatic conditions, resources, building and maintenance capacities, accessibility, and affordability.

These have been included because of an otherwise scarcity of guidelines.

Interim Housing and Diversity

The standardized nature of interim housing can miss out the special needs of sections of an affected population. These can include pregnant and lactating mothers, the elderly, the blind, the wheel chair bound, the bed ridden, single parent families, and those with medical or mental challenges. It can also happen in unexpected areas. For example, it was unfortunate that a Maori

family with previous gang ties was evicted from an evacuation center following the September 2010 earthquake in Christchurch (Kipa et al. 2013). This exclusion leads to the setting up of a specific Maori response when the February 2011 earthquake occurred 5 months later. The notion that New Zealand society could be fractured like this still remains deep, as it no doubt does for the Lower 9th Wards of New Orleans.

Such people are essentially invisible in a disaster, and there is a reliance on self-identifying. The sudden nature of earthquakes means that self-identification before the event is at best limited. As a result, such issues only surface (unfortunately) around the provision of interim housing, and their special needs were compounded by that time by the loss of vital connections with personal care providers and sources for medication, service animals, community liaisons, public transportation, neighbors, and their network of contacts. And the diversity challenge is whether the provision of interim housing addresses a diverse community in an inclusive manner. This is not easy or straightforward, and failures are not recorded because they are seen as failures of the wider community or society such as the above example.

Case Study 1: Sichuan, China, 2008

Seismic event: 12 May 2008, magnitude 8.0.

Deaths: 69,195 deaths and 18,392 missing (Wikipedia 2008).

People affected: the earthquake left about 4.8 million people homeless, though the number could be as high as 11 million.

Total amount of direct damage: 845.1 billion RMD (137 billion USD) (UNCRD 2008, p. 5).

Disasters of this scale are not uncommon in China, and interestingly, this was the first time that China allowed foreigners into a disaster zone (perhaps because it was hosting the Olympics in Beijing later that same year in August), and hence for many, it was a unique opportunity to observe how the Chinese disaster response works. The two key aspects relevant to interim housing seem to be their community (or village)-based

approach and partnering. This allowed the construction of around 500,000 temporary houses in 3 months following the earthquake and was one of the best disaster responses measured by the well-being of affected families (Potangaroa et al. 2009). The temporary homes or shelters had a floor area of 20 m² and were designed for three people (EERI 2008).

On 6 June 2008 (3 weeks after the disaster) the State Council published the Wenchuan Earthquake Restoration and Reconstruction Regulations and provided legal guidance for the formulation of the overall reconstruction plan (Kristin et al. 2012). The plan emphasized restoring residential housing and in particular the need for addressing the differences of multistory urban apartment blocks and low-rise rural housing. The plan also laid out the process to be adopted which consisted of the three following actions:

- Survey and assessment
- Plan drafting
- Policy measures

In response to this, the Sichuan Provincial Government published in that same month of June the “Sichuan Rural Housing Reconstruction Plan” and then in September the “Urban Housing Reconstruction Plan,” while the Central Government published the “Sichuan Rural Housing Repair and Reinforcement Plan” in December 2008.

These plans laid out six methods for reconstructing or replacing houses that were damaged in the earthquake, namely:

- Reinforcing the house
- Rebuilding the house on the original land
- Rebuilding the house on replacement land
- Renting a low-cost house
- Buying a low-cost house
- Rebuilding the house organized by “workplaces”

There were additional options in urban areas of buying a low-cost house, renting public housing, or receiving a cash subsidy.

Houses were assessed and classified according to the extent of damage into the following five categories:

- No damage
- Minor damage
- Medium damage
- Serious damage
- Collapse

Apart from houses that had “collapsed” or had sustained “no damage,” categories would receive a cash subsidy of 1,000–2,000 (\$163–\$325), 2,000–4,000 (\$325–\$650), or 4,000–5,000 (\$650–\$813) RMD (\$USD), respectively. Urban households whose houses had collapsed or were inhabitable could receive a cash subsidy, a housing tax exemption or reduction, and preferential low-cost housing. The average cash subsidy was 25,000 RMD (\$4,066USD) depending on the household’s size and economic status.

In rural areas, many houses were rebuilt on different land, and in these cases, the town and village authorities could choose to rebuild the houses collectively, with each household contributing a share. The point to note is that the whole village was relocated and not individual families. Households were entitled to a cash subsidy of RMD 20,000 (\$3,250USD) (on average) to rebuild with the exact amount determined by the number of people and their overall economic situation. Households that decided on an alternative living arrangement such as living with other family received a reduced subsidy of RMD 2,000 (\$325USD).

In October 2008, the Sichuan Government announced various preferential loan policies for households to rebuild their house or buy a new house. For households taking out a bank loan to purchase a new house, the minimum interest rate was adjusted to 0.6 of the benchmark lending interest rate, and the minimum down payment was reduced to 10 %. And for rebuilding, buying or strengthening a house received a 1 % point discount on their interest rate.

Where ever possible, communities are relocated as a community or village, and not individually (UNCRD 2008, pp. 22–24). Each

village has its own organizational structure that interfaces between the people of the village and the city authority. This provided a ready two-way information conduit with assessment and damage data being fed up and strategy and priorities being fed down. Thus, the authorities had rapid and relatively accurate data very early and hence could plan an appropriate response. This included a partnering or “twinning” arrangement whereby affected cities and towns are linked to other cities and towns in China who are then expected to assist. The partnering arrangement can be competitive as results from it are published by the Central Government. But this could also be because the workers from coastal cities working in the mountainous and colder regions of Sichuan and staying in tents and their own cars perhaps feel more motivated to get the work completed quickly.

However, despite its important role in terms of financial and resource allocation, it does seem to have three areas of concern (UNCRD 2008, p. 5):

- Variation in support from provinces.
- Inappropriate support that may not necessarily reflect the needs of a different climate, culture, and social background of the partner region. For example, facilities provided from Northern Provinces had double-panel windows and contain heat-insulating materials, while those provided from Southern Provinces have no such equipment. This is not a big problem in the phases and rescue and rehabilitation because short-term responses or restoration is required in the phases. In the reconstruction phase, however, supporting materials need to be accepted by the supported region.
- Supporting provinces tend to have a larger voice and decision-making power than supported areas.

Summary

It has been historically difficult getting access to disasters in China, and certainly while access was granted to foreigners following the earthquake in Sichuan, it was nonetheless limited and understandably constrained. Thus, reports and feedback from the field are scarce. This is a pity as

the scale and response to seismic disasters has much that other countries could learn from and adapt for their own situations. The partnering approach and the speed at which interim housing was achieved are two clear examples. While it was conceivably possible for the Government to impose price controls, this did not seem to happen during the four field trips made by the author and several others made by PhD students studying aspects outside the provision of interim housing (Chang et al. 2011). Moreover, the use of loans and other seemingly market instruments to manage the response perhaps suggested a sophistication beyond simply Government structures.

Case Study 2: Port-au-Prince, Haiti, 2010

Seismic Event: Magnitude 7.0, 12 January, 2010.

Deaths: 220,000 deaths with some variance between different estimates.

People affected: an estimated three million people were affected by the earthquake together with the collapse of 250,000 residences and 30,000 commercial buildings (Ferris 2010) (Wikipedia 2010).

This was the first large-scale humanitarian response involving a “crowded” urban context. The rural response seems to lead the urban one, and the reasons why become clear when people see photographs of the affected area in the capital of Port-au-Prince. The poorer areas effectively are 100 % covered in corrugated iron roofing, and hence, the immediate question was: Where would any interim housing be established? The planned shelter solution was to build on the site of the former house which required demolition and clearing of the site before the interim housing could be erected. Needless to say, this was problematic in itself; the “owners” had to be found, land tenure had to be identified, the rubble had to be typically carried out by hand out to the road, and the interim house had to be built. However, when this was completed, where was any permanent housing to be erected? Consequently, the interim house or T-Shelter took on a more permanent role that it was not intended, and the role of what was intended as a “transitional” house

became increasingly questioned as transition to what?

Nonetheless, providing emergency shelter support to over 1.5 million people in less than 4 months was an outstanding achievement (Rees-Gildea and Moles 2012). The Shelter Cluster’s “harmonized” emergency shelter and NFI response seemed to be a coordination success (the United Nations emergency response consists of 11 sector clusters and one of those is the Shelter Cluster. It is headed by the IFRC for natural disasters and UNHCR in conflict situations). However, the emergency shelter (E-Shelters) response was essentially based on a one-time massive delivery of tarpaulins and tents, there was some variation in quality, but more importantly, there was little or no control over the shelter outcomes that those affected were selecting. Camps both small and large blossomed throughout Port-au-Prince, and at one time, there were over 1,300 camps with 100 people or more. Camps were established in parks, on golf courses, around churches, and on street corners. The shelter agencies did not have the means to guide or direct the final outcome of the E-Shelters. The large numbers and the tight urban situation made it impossible to meet the SPHERE standards such as a minimum of 3.5 m²/person in covered area (such as a tent or tarpaulin), 30 m²/person in camp land area, and the prescribed numbers of toilets and showers. It was hoped that these perhaps could be progressively met as camp numbers decreased. This did not happen as people seemed to “settle” into camp life, and 1 year later, nearly one million people were still in camps.

Replacement of such shelters that have a service life of around 6 months was not initially planned with the intent was to move into T-Shelters by the end of 2010. That did not happen for many reasons, and consequently, there was a shelter gap as the hurricane season rolled in, and another call for replacement tarps and tents was made. This unplanned double effort put more pressure on the overall T-Shelter program and meant a higher workload for shelter agencies together with an unexpected additional budget line. This vulnerability became evident

when a relatively small depression Ivan hit Port-au-Prince destroying 9,000 E-Shelters.

The Shelter Cluster was also seeking to harmonize the T-Shelter in much the same way it had for the E-Shelter. There were several different types developed by aid agencies based around the following standardized criteria (IASC 2010):

- An area of 12–18 m² for a single-story structure
- Cost between \$1,000 and \$1,500USD (including transport and labor) later increased to \$1,800
- Designed for a 100 mph wind load
- Designed for seismic loads
- A 3-year life span but later increased to 15 depending on possible building up grading
- The lack of a diaphragm transfer system. This structural system transfers lateral loads (such as wind and seismic) through the building to vertical structure which takes it to the foundations. It is usually located in a horizontal plane such as the ceiling or roof.
- Foundations and their spatial relationship to the above wall bracing. It was noted that diagonal bracing in walls were not located above foundations and as such would be unintentionally bending and potential breaking the wall bottom plates.
- A lack of gable end bracing. This is where braces are required to prop gable ends back to a horizontal structure.
- Roof slope: pitched roofs were advised to be +30° and mono-sloped roofs 12–14°. This was done to minimize wind up lift loads.

This harmonization proved problematic for several reasons, and as a result, shelter numbers had to be lowered to fit existing budgets. For example, there were no seismic codes and, hence, no seismic zones for Haiti. In addition, the nature of the wind loading was not clear but also not questioned. Was it a 3 s gust or a sustained wind load; what was the specified terrain and at what height was this measured (Potangaroa 2010b). Consequently, agencies and their respective designers in other parts of the world using at times non-hurricane wind loading codes seemed to make their own assumptions of what this wind speed meant. For example, the ASCE 7 code produced basic wind loads of around 1.2 kPa, while the same wind speed in the British code produced a 2.5 kPa loading (it should also be noted that hurricanes are not featured in the British code). This was not apparently picked up by agencies or their designers who invariably were not based in Haiti but were working from information provided by aid agencies in the field. This resulted in significant cost over runs particularly for foundations and costs quickly blew out with the largest being in the order of \$4,500USD, nearly three times over the capped value.

In addition, there were other aspects of concern with the seismic and wind design of the T-Shelter design that included the following:

The 11 exemplar T-Shelters on the Shelter Cluster web page were examined, and these are tabulated below (Table 2).

A lack of a horizontal structure or suitable foundations would constitute a systemic issue, while the gable aspect could/would be more localized. Thus, at least six had systemic issues and nine localized issues, and given that these 11 are probably better examples underlines an issue that aid agencies need to address. This same issue of design is also raised in the Aceh case study that follows. However, aid agencies are usually reluctant to discuss these issues given the large resources and costs involved in the provision of interim housing.

Perhaps this was partly why many agencies “were inclined to work outside Port-Au-Prince, mainly funding the direct delivery of transitional shelter [T-Shelters] and not integrating a housing repair approach into their programmes. . . . This widely supported strategy [T-Shelters] was unrealistic in timing and cost-wise and was not flexible enough. Relevant progress in T-Shelter delivery started by May 2010, and the sum of the agencies capacities (125,000 units) proved to be insufficient to meet total transitional shelter need (over 180,000) within the foreseen timeline (January 2011), even after maximum delivery capacity was reached (7,300 units per month, as

Interim Housing Provision Following Earthquake Disaster, Table 2 T-shelter design aspects

	Horizontal structure	Foundations	Gable brace	Roof slope
Agency 1	No apparent structure	Do not appear to relate to diagonal braces	Partial separation for side gables?	OK
Agency 2	OK diagonal steel ties	OK concrete slab	OK diagonal timber brace	OK
Agency 3	No apparent structure	Do not appear to relate to diagonal braces	OK diagonal timber ties	OK
Agency 4	OK diagonal timber ties	Probably OK concrete block and slab foundations	No apparent brace	OK
Agency 5	OK diagonal timber ties	Probably OK because of external plywood	No apparent brace	OK
Agency 6	No apparent structure	Do not appear to relate to diagonal braces	Full height member but no apparent bracing at roof level	OK
Agency 7	No apparent structure	No apparent bracing?	No apparent brace	Not clear
Agency 8	OK diagonal steel ties	Probably OK because of external plywood	No apparent brace	OK
Agency 9	No apparent structure	Do not appear to relate to diagonal braces	No apparent brace	Less than 30°
Agency 10	OK diagonal timber ties	Probably OK because of concrete ring beam	No apparent brace	Probably OK
Agency 11	OK diagonal steel ties	Do not appear to relate across the building	No apparent brace	OK
Totals	5 with potential seismic/wind issues	6 with potential seismic/wind issues	9 with potential seismic/wind issues	1 with potential wind issue

of November 2010). All of which led to an accumulated delay, to bottlenecks in the implementation process, and to unforeseen relief responses as new emergencies arose. In addition, the approach of faster delivery of simpler transitional shelter could not be always achieved, and the shelter agencies had serious difficulties in meeting the agreed upon standards and average costs, leading upwards of a 200 % increase in the anticipated costs. . .” (EPYPSA 2011, p. 5).

That report goes on to outline the complexities of supplying transitional shelter with its comments that “The shelter sector in Haiti promoted a return to a neighbourhood strategy aimed at drawing the displaced population out of the camps through the provision of shelter (transitional, permanent, home repairs), but the shelter agencies procrastinated their engagement in housing repair.” This low engagement in housing repair and seismic strengthening may have

been caused by the following (EPYPSA 2011, p. 46):

- Lack of experience and perhaps a fear of the complexity involved (first “urban”-based seismic disaster for the humanitarian agencies with a lack of physical access, emphasis on rubble, problems of logistics, and few trained staff and labor)
- Difficulty to design big-scale funding proposals (each situation was different).
- Budget over runs (unforeseen costs that can only become clear during construction).
- Lack of funding (HQ and/or donors not willing to go into complex construction projects).
- Beneficiary selection and involvement (who and how should assistance be provided?).
- Repair or retrofit? (Repairs would fix the damage but not it’s lack of earthquake capacity. Retrofitting would address both issues but

could be expensive and would require specialist skills)

- Should this be seen as an emergency rather than a recovery/reconstruction response?
- Camp-driven response easier to manage and control (less constraints and higher visibility).
- Overall perceived as “easy to get into, but hard to get out.”

Some agencies such as the IFRC did manage to develop a seismic retrofitting program (Potangaroa et al. 2011b). This approach had the distinct advantage of being well under the \$2,200–2,400USD/family cost to relocate families out of the camps. However, as indicated earlier, this option was developed late in the response in Port-au-Prince because there was the sense that any seismic retrofitting would have to be prohibitively expensive. The approach was based on the following two factors (Potangaroa et al. 2011):

1. An acceptance of higher level of building damage in an earthquake; up to what is often termed “near collapse.” Most seismic codes are based around a “safe exit” strategy which requires a lower level of damage.
2. A realistic determination of the life span of the building. Again most codes are based around approximately 100 years, but the low building quality suggested a 25–30-year life span.

And these two factors allowed for a code compliant design to levels that were appropriate to the building culture of Port-au-Prince rather than a developed one in San Francisco or Wellington.

The seemingly final part of the interim housing in Port-au-Prince was a Government initiative to set a repair program for 16 neighborhoods and to decongest six camps associated with those neighborhoods under a proposed 16-6 Program (IOM 2012). Camp dwellers and residents in damaged neighborhoods could choose between either having their homes fixed or taking a 1-year rental subsidy of \$500USD. About 10,000 families took this option. However, what was going to happen beyond that year was not clear; again it was the “transition to what”?

Summary

The timelines required for interim housing were seemingly underestimated in Port-au-Prince. This caused several issues and, for example, the relevance of the T-Shelter decreased, “while other approaches like repairs, rental subsidies or even permanent housing became relevant and even more cost efficient as time elapsed” (EPYPSA 2011). The affected population also started to look at the T-Shelter as permanent solutions, both because they were of a higher standard than their original housing and because they were unaware of any planned transition. Shelter agencies planned to build 125,000 T-Shelters in 10 months (March 2010 to January 2011) but were unable possibly because the numbers were too ambitious to start off with. With time, material costs escalated, the T-Shelter increased its life span beyond the specified 3 years, and issues such as land tenure, rubble clearance, and urban setting resulted in lengthy delays. Hence, by November 2010, only 19,000 T-Shelters (15 % of the total goal) had been completed and delivered.

Secondly, the role of appropriate structural design was missed by aid agencies. This will be highlighted in the next case study on Banda Aceh. It would seem that this lesson has not been learnt. And while aid agencies may be unnecessarily concerned about design risks and responsibilities, the apparent situation of six systemic and nine localized failures from 11 exemplar T-Shelters in Port-au-Prince would seem to deserve decisive attention from the aid community beyond a concern with design risk and responsibilities. Moreover, there seems to be opportunities to better mainstream the social needs and concerns of communities by a reflective design approach suggested by the IFRC retrofitting approach.

Finally, it would also seem that there is currently no urban game plan for provision of interim housing. While the rural response seemed to be managed in Haiti, the urban one in Port-au-Prince was not. It would seem that the provision of interim shelter in urban areas needs more discussion and research, hopefully before the next earthquake disaster. This is supported by the case study 1 on Christchurch.

Case Study 3: Banda Aceh 2004 and Yogyakarta 2006

Seismic event: Magnitude 9.0, 26 December 2004	Seismic event: Magnitude 5.9, 27 May 2006
Deaths: 130,000 deaths	Deaths: 5,716 deaths
People affected: million houses were destroyed or made uninhabitable	People affected: 240,396 houses were destroyed (IRC 2009)

These two responses were connected firstly because of their geography and timing but also because the Banda Aceh experience is now seen as a kind of watershed for humanitarian agencies working in the interim shelter sector that flowed over into Yogyakarta. The sense at the time of Banda Aceh was to miss out any interim housing and instead go directly to permanent housing (Potangaroa 2006). The Government had indicated that they would build barrack type accommodation and that combined with people returning to their original tsunami affected sites and staying with other family members seemed to be the basis of the Government's strategy. Land tenure was an issue as much of it was oral and informally agreed between male community members. Thus, the loss of land marks and people meant that ownership was no longer certain. In response, agencies put together a multi-program such as the one proposed by UNHCR (Potangaroa 2005):

- New Housing on new or former home sites. (Four sizes of housing proposed depending on family size).
- A "Granny" flat option addition to an existing house where those affected wish to stay long term with other "host" family members (the size was based on the living/lounge area and toilet bathroom areas from a spatial survey of non affected houses in Banda Aceh).
- Shelter Kits for repair of an existing house.

However, in March 2005, the Government suggested that agencies such as UNHCR (and also IOM) should not be in the once military zone of Aceh; thus plans such as the above were

never realized at the scales planned (TIME 2005). This created a strategy and agency vacuum and consequently it was not till September 2005 (9 months after the disaster) that the IFRC launched its Temporary Shelter Plan of Action. This consisted of a 25 m² lightweight bolted steel frame, with timber cladding and a metal roof. The foundations were supposed to be anchored down but usually were leveled and left. Problems with appropriate sustainable timber supplies delayed its release but by December 2005 a prototype was in place. Consequently, many families did not receive these interim/transitional shelter kits until 2006. This interim kit was valuable and on the informal market could be purchased for \$5,000–\$6,000USD. This was also before the advent of the Shelter Cluster approach, and so agencies tended to operate independently and without any coordination.

Da Silva explains that the (da Silva 2010) "Shelter policy was not developed within the overall context of the journey from emergency shelter to durable solutions, and over the first six months there was considerable confusion and no clear policy as to the type of shelter assistance required. In June 2005, BRR [the Government's Disaster Response Agency] announced that families should be encouraged to return to their own land or to voluntarily resettle on land purchased by communities themselves or by BRR. They also stated that each affected household would be eligible for a permanent 36 m² house, with an expectation that this could be realised within a year, in the belief that in the interim affected communities were adequately housed in barracks, transitional shelter or with host families. Agencies who had already constructed 'semi-permanent' houses were faced with having to upgrade or replace housing, as these were no longer deemed adequate. This timeline proved unrealistic and did not reflect the realities on the ground, led to false expectations by the media, donors, government and beneficiaries and placed considerable pressure on implementing agencies."

Consequently, Da Silva's conclusion was that the notion that everyone who was affected should get a house led to a focus on "reconstruction rather than recovery" and an "emphasis on

providing houses rather than assistance to reconstruct.”

One further “lesson” was that agencies needed to have experienced seismic design engineers involved; otherwise, a less than adequate seismic solution could be used. And that having solely civil or structural engineers or local engineering input was no guarantee of finding an appropriate or even adequate seismic solution. Some agencies had again to come back and seismically strengthen the housing they had already constructed. Invariably this relates to “simple” structural detailing such as the spacing of column ties or beam stirrups/links. This was not the common practice amongst local consultants and builders in Aceh. In addition, the Indonesian Seismic Code excluded single-story dwellings, and the building code for Aceh issued by BRR (UNHIC 2005) did not include basic seismic design principles. This resulted in widespread confusion that compounded the view that seismic design would be considerably more expensive. BRR did not enforce the need for seismic design and themselves built houses and schools which did not conform to basic seismic guidelines. Da Silva goes on to report that “...most DEC [Disasters Emergency Committee (UK)] Member Agencies ultimately did adopt seismic design approaches but this was not always matched by quality construction and the design intent was compromised by poor quality materials and workmanship on site. In some cases remedial works and retro-fitting was needed to create a safe structure.”

Agencies tried to dodge these issues with the response in Yogyakarta; but unfortunately repeated many of the same problems.

The Indonesian Government was reported as not being able to cover the need for shelter in Yogyakarta, not simply because of a lack of finances but also due to its scale and that it was much greater than initially believed (Roychansyah 2009). To meet the need for shelter, the Indonesian Government, United Nations agencies, and other humanitarian partners developed a “roof first” interim housing strategy to be implemented from the time aid is first provided up to the stage of community rehabilitation.

The Emergency Shelter Cluster was coordinated by the IFRC who calculated that with an average household size of 4.3 that at least 1,173,742 people, and perhaps as many as 1,542,380 were left homeless by the earthquake. However, the provision of tarpaulins and tents was limited, and by the end of July 2006 (2 months after the earthquake) at the end of the Response Phase, almost 30 % of these numbers were still homeless.

Despite this short fall interim housing was achieved using emergency bamboo and tarpaulins and recycling materials (such as bricks, timber, and window, and door frames), and finally villages that had sustained less than 20 % damage accommodated those made homeless using a kinship family model within the village.

The goal during the response phase was in addition to the usual emergency tents to also provide interim housing or T-Shelter (Setiawan 2007). Most were constructed from bamboo which was selected because of its low cost, availability, and strength. Sizes varied between 18 and 36 m². There were three different levels of community participation with regard to shelters: participation in the design process, participation in contributing materials for transitional shelters, and participation in the construction process (IRP 2009; JRF 2008). Somewhere in these phases, there were two significant structural changes; the first was the inclusion of a heavy clay tile roof for cultural reasons. It seemed that no normal Javanese man would sleep under anything other than a clay tile roof. And secondly, the bamboo wall braces were lifted so that people could readily enter in the side of the interim houses. The clay tile roof rendered the house as earthquake prone, and the movement of the wall braces minimized any remaining seismic resistance. This clash between community engagement and the confusion around seismic standards that was so problematic in Aceh had not been learnt for Yogyakarta or for Haiti.

Summary

The provision of interim shelter is important because it is often the first step towards

normalizing a community following an earthquake. It is perhaps more complicated and certainly takes longer than expected, and the key appears to be the engagement with the affected people. This seems to apply whether it occurs in a developed or developing economy and whether it is provided as aid assistance or not.

The definitions for interim housing given in the literature need to be reviewed. And, for example, the one provided at the start of this chapter of between 1 month and 2 years is probably too soon on both the 1 month and the 2 years for a large-scale disaster or catastrophe.

While affected communities are the key focus for what is needed, Governments are the key influence for how that could be provided. Strangely Government departments (representing the Government) seem to have problems communicating with their affected communities when there is not an existing communication structure prior to the earthquake. Where it does not exist, departments must ensure that what is happening in the field is directly measured and observed rather than remotely from a headquarters/office location.

Interim housing is expensive and can unintentionally become more permanent housing. Its design can be problematic again in both developing and developed economies, and there are gaps for interim housing strategies for urban contexts. It is perhaps sobering to reflect upon Jean Jacques Rousseau's letter to Voltaire after the 1755 Lisbon Earthquake who is reported to have written that "... if this tragedy happened, it is not the fault of mother nature. It was not nature who gathered together twenty thousand buildings in this place." This must be a concern with over half of the world now living in cities and how the provision of interim housing after an earthquake in an urban context will be achieved.

The ability of interim housing to merge into or be the basis of any following housing strategy certainly can be an advantage. It is commonly used in the humanitarian and development sector as transitional or core shelter/housing. However, in developed economies, more reliance is placed on insurance and that those affected will seek their own solutions, and hence emergency

shelters and evacuation centers seem to be the basis for response.

Finally, it seems that shelter in place is the preferred option for those affected by earthquake, and that individual relocation is the least preferred. Yet as noted above with 50 % + of the world's people being urban based suggest that some relocation whether temporary or permanent may need to happen. Consequently, the provision of interim shelter will increasingly become more complex and involved but will still remain as one of the important normalizing steps for communities after an earthquake disaster.

Cross-References

- ▶ [Building Codes and Standards](#)
- ▶ [Community Recovery Following Earthquake Disasters](#)
- ▶ [Earthquake Disaster Recovery: Leadership and Governance](#)
- ▶ [Learning from Earthquake Disasters](#)
- ▶ [Resilience to Earthquake Disasters](#)

References

- ABAG (2008) Long term housing recovery Pub. by Earthquake and hazards program local and regional disaster recovery planning issues Association of Bay Area Governments Issues paper, 8 July 2008
- ARC (2012) Preferred sheltering practices for emergency sheltering in Australia. The application of international humanitarian best practice. Pub. by Australian Red Cross, Victoria
- Arslan H, Coşgun N (2007) The evaluation of temporary earthquake houses dismantling process in the context of building waste management. In: International earthquake symposium Koceau, Turkey 22–26 Oct 2007
- Cancern (2014) Newsletter # 113, 31 January 2014. <http://www.rebuildchristchurch.co.nz/blog/2014/1/cancern-newsletter-113-31-January-2014>. Accessed Feb 2015
- Chang Y, Wilkinson S, Potangaroa R, Seville E (2011) Identifying factors affecting resource availability for post-disaster reconstruction: a case study in China. *Construct Manage Econ* 29(1):37–48
- da Silva J (2010) Lessons from Aceh Key considerations in post-disaster reconstruction. Practical Action Publishing, Rugby/Warwickshire, p 63. ISBN 978 1 85339 700 4. www.practicalactionpublishing.org
- Davidson C, Johnson C, Lizarralde G, Sliwinski A, Dikmen N (2007) Truths and myths about community

- participation in post-disaster housing projects. *Habitat Int* 31:100–115
- Davis I (2007) Learning from disaster recovery guidance for decision makers. International Recovery Platform (IRP), Geneva and Koba, p 24
- Davis I (2010) What is the vision for sheltering and housing in Haiti? Summary observations of reconstruction progress following the Haiti earthquake of January 12th 2010 online at www.onuhabitat.org/haiti, accessed Feb 2015
- DHS (2013) Unless modified, FEMA's temporary housing plans will increase costs by an estimated \$76 million annually. Department of Homeland Security, Washinton, DC, p 3
- EERI (2004) Learning from earthquakes. Social and public policy issues following the Bam, Iran, Earthquake. EERI special earthquake report, California, Aug 2001
- EERI (2008) Reconnaissance report on the China Wenchuan Earthquake May 12, 2008. Chu-Chieh J. Lin, Associate Research Fellow of NCREE Juin-Fu Chai, Research Fellow of NCREE, Berkeley, CA
- EPYPSA (2011) Estudios Proyectos y Planificación S.A. An evaluation of the Haiti earthquake 2010 meeting shelter needs: issues, achievements and constraints. International Federation of Red Cross and Red Crescent Societies (IFRC), Geneva, p 5
- Ferris E (2010) Earthquakes and floods: comparing Haiti and Pakistan. The Brookings Institution
- Giovinazzi S, Stevenson J (2012) Temporary housing issues following the 22nd Christchurch earthquake, NZ. In: 2012 NZSEE conference University of Canterbury, Christchurch
- Haas J, Kates R, Bowden M (1977) Reconstruction following disaster. The MIT Press, Cambridge, MA/London
- Hewitt J (2013). Errors of commission. In: AUBEA 2013 conference, Auckland, 20–22 Nov 2013
- IASC (2010) Shelter Cluster Haiti transitional shelter technical guidance 19/02/10. Shelter Cluster technical working group, Port au Princes
- IASC (2011) United Nations Inter Agency Standing Committee operational guidelines on the protection of persons in situations of natural disasters. The Brookings – Bern Project on Internal Displacement, IASC, Geneva, Jan 2011
- IOM (2012) Haiti: from emergency to sustainable recovery IOM Haiti two year report (2010–2011), pp 9–11, 37–3B. http://www.com.int/Jah_19/webdav/shared/main/site/activities/countries/docs/haiti/IOM-Haiti-Two-year-Report-2010-2011-From-Emergency-to-sustainable-Recovery.pdf. Accessed Feb 2015
- IRP (2009) International Recovery Platform, recovery status report. The Yogyakarta and Central Java Earthquake 2006 with Gadjah mada uni. pub. IRP Kobe Dec 2009
- Jha A, Barenstein J, Phelps P, Pittet D, Sena S (2010) The social dimension of housing reconstruction. In: Safer homes, stronger communities: a handbook for reconstructing after natural disasters. World Bank, Washinton, DC, pp 59–76
- Johnson C (2002) What's the big deal about temporary housing? Planning considerations for temporary accommodation after disasters: example of the 1999 Turkish earthquakes. In: TIEMS disaster management conference 2002, Waterloo
- Johnson C (2007) Impacts of prefabricated temporary housing after disasters: 1999 earthquakes in Turkey. *Habitat* 31(1):36–52
- JRF (2008) Two years after the Java earthquake and tsunami: implementing community based reconstruction, increasing transparency. Java Reconstruction Fund progress report 2008, World bank, Washington, DC. <http://documents.worldbank.org/wrote/en/2008/01/97921> accessed Feb 2015
- Kipa M, Potangaroa R, Wilkinson S (2013) Iwi resilience? The Maori response following the February 22, 2011 Christchurch earthquake. In: AUBEA 2013 conference, Auckland, 20–22 Nov 2013
- Kristin D, Hedda F, Liu J, Zhang H (2012) Recovering from the Wenchuan earthquake. FAFO-Report 2012:39, Oslo. ISBN 978-82-7422-917-4, ISSN 0801-6143. www.faf.no/english
- Lloyd-Jones T (2006) Mind the gap! Post-disaster reconstruction and the transition from humanitarian relief. RICS, London
- Lyons M, Schilderman T, Boano C (2010) Building back better: delivering people-centred housing reconstruction at scale. Practical Action, Rugby; South Bank University/IFRC, London
- MBIE (2012) Guidance to assist the development of temporary accommodation to house workers in greater Christchurch. NZ Government; Ministry of Business, Innovation and Employment, Wellington, Sept 2012
- MCDEM (2004) National civil defence and emergency management CDEM strategy 2003–2006. Ministry of Civil Defence Emergency Management, NZ Govt, Wellington, pp 14–16
- McIntosh J (2013) Chapter10: The implications of post disaster recovery for affordable housing. In: Approaches to disaster management – examining the implications of hazards, emergencies and disasters. Intech, Rijeka
- Natural Hazards Center (2001) Holistic disaster recovery ideas for building local sustainability after a natural disaster. Natural Hazards Research and Applications Information Center, University of Colorado 482 UCB, Boulder
- OCHA (2012) What is the cluster approach? OCHA on message: the cluster approach. UN office for the coordination of humanitarian affairs. https://docs.unocha.org/sites/dms/Documents/120320_OOM-ClusterApproach_eng.pdf
- Potangaroa R (2005) The strategy for a shelter program. Report for UNHCR Banda Aceh, Indonesia. Technical Coordinator UNHCR Bandar Aceh, 24 Jan 2005
- Potangaroa R (2006) The development of a permanent shelter program for Aceh, North Sumatra.

- In: Proceedings of the scientific forum on the tsunami, its impact and recovery regional symposium, AIT, Thailand, 6–7th June 2005
- Potangaroa R (2010a) The seismic gap: issues of seismic design in post disaster reconstruction. i-Rec2010 participatory design and appropriate technology for post-disaster reconstruction, Ahmadabad
- Potangaroa R (2010b) Lessons learnt as part of the terms of reference deliverables. Shelter cluster technical coordinator available via the Shelter Cluster Web site. Haiti, Nov 2010
- Potangaroa R, Wang M, Chang Y (2009) Identifying resilience in those affected by the 2008 Sichuan earthquake. *J Eng Sci* (in Chinese)
- Potangaroa R, Wilkinson S, Zare M, Steinfort P (2011a) The management of portable toilets in the eastern suburbs of Christchurch after the february 22, 2011 earthquake. *Australas J Disaster Trauma Stud* 2011(2):35–48
- Potangaroa R, Mattar Neri R, Brown D (2011b) The design development and costs of several key housing issues for Delmas 19, Port Au Prince. Report for the British Red Cross Haiti, Dec 2011
- QG (2010) Form23 temporary accommodation buildings checklist.(version 1) The State of Queensland (Department of Infrastructure and Planning). Queensland Government, Brisbane, May 2010
- Quarantelli EL (1995) Patterns of shelter and housing in US disasters. *Disaster Prev Manage* 4(3):43–53
- Rees-Gildea P, Moles O (2012) Lessons learned & best practices. In: The international federation of Red Cross and Red Crescent Societies Shelter Programme in Haiti 2010–2012, IFRC, Geneva
- Roychansyah M (2009) Chapter 3: sector-specific recovery and Case Studies: 3.1 Shelter international Recovery Platform recovery status report. The Yogyakarta and Central Java Earthquake 2006, IRP Washington, DC, p 9
- SAMB (2009) Regulating temporary farm worker housing in the ALR. Discussion paper and standards. The Sustainable Agriculture Management Branch Ministry of Agriculture and Lands, Victoria
- Setiawan B (2007) Lessons from Aceh and Jogya. Towards better disaster management. Partnership for Governance Reform, Yogyakarta, Dec 2007. ISBN 978-979-26-9629-9
- Shelter Cluster Philippines (2014) Shelter and environment – an overview Typhoon Yolanda response. Philippines, Haiyan Shelter Cluster, 12 Jan 2014 download from <http://www.sheltercluster.org/>
- SPHERE (2011) The Sphere project, the humanitarian charter and minimum standards in humanitarian response. Practical Action Publishing, Rugby. ISBN 978-1-908176-00-4
- TIME (2005) After the tsunami: a time to build. Special issue, 4 Apr 2005, 165(14)
- UNCRD (2008) The disaster management planning Hyogo Office of the United Nations Centre for Regional Development (UNCRD). Report on the 2008 Great Sichuan earthquake, Mar 2009. Available at http://www.preventionweb.net/files/13114_UNCRDSichuanReport200903EN.pdf
- UNHCR (2007) Handbook for emergencies, 3rd edn. United Nations High Commissioner for Refugees, Geneva
- UNHIC (2005) Guidelines on housing repair and construction. United Nations Humanitarian Information Centre for Sumatra, Book 3 SWG, pp 1–5
- USAF (2001) Temporary lodging design guide. United States Air Force, Brooks AFD, Texas, Rec 2001
- Wikipedia (2008) 2008 Sichuan earthquake. http://en.wikipedia.org/wiki/2008_Sichuan_earthquake
- Wikipedia (2010) http://en.wikipedia.org/wiki/2010_Haiti_earthquake. Accessed Feb 2014
- Wilson W (2013) Homelessness in England – commons library standard note. Published 12 Dec 2013, Standard notes SN01164
- Zuo K (2010) Procurement and contractual arrangements for post-disaster reconstruction. PhD thesis, Auckland University School of Engineering, p 60

Land Use Planning Following an Earthquake Disaster

Wendy Saunders¹, Suzanne Vallance² and Ljubica Mamula-Seadon³

¹GNS Science, Lower Hutt, New Zealand

²Lincoln University, Christchurch, New Zealand

³Centre for Infrastructure Research at the Department of Civil and Environmental Engineering, Faculty of Engineering, The University of Auckland, Auckland, New Zealand

Synonyms

Canterbury; Governance; Post-earthquake; Pre-event; Recovery

Introduction

The term “land use planning” covers many aspects of planning: spatial, urban design, heritage, transportation, recreation, access to education, health facilities, natural hazards, landscape, biodiversity, air and water quality, affordable housing, cultural values, sustainability, etc., and this is not a complete list of the many aspects and parts to land use planning. A planner may specialize in one of these aspects, e.g., a transportation planner, or be a generalist – knowing a little about many aspects of planning.

There are many definitions of land use planning; however, March and Henry (2007, p. 17) provide a concise discussion on what planning, and more specifically land use planning, is. They state:

Land-use planning ... is focussed upon establishing the best spatial arrangements of land use, development, and management ... To do this, land-use planning is confronted with the task of establishing which of the many potential patterns and organisations of land use are likely to be the most advantageous in the future, in a particular place, and for a particular community.

Therefore, land use planning is an important mechanism for the reduction of risks in the built and natural environments (March and Henry 2007). Although planning cannot avoid all natural hazards, the thoughtfully planned location and design of structures and land uses can reduce the risks, particularly when combined with emergency management systems, i.e., early warning and response systems.

So how does land use planning contribute to recovery following an earthquake? Often touted as the key tool available for sustainable hazard mitigation, land use planning has a key role in determining how a city recovers, its long-term sustainability, and resilience. The following sections will provide an overview of the role of land use planning during recovery following a major earthquake using the 2010–2011 Canterbury earthquake sequence as a case study. This will provide a brief overview of the Canterbury

earthquake sequence and outline the role of governance in planning outcomes, the value of pre-event recovery planning for land use, and the complex post-earthquake realities of planning within the context of the Canterbury earthquakes.

The Canterbury Earthquakes

On September 4, 2010, at 4.36 am, the Canterbury (New Zealand) region experienced a 7.1 magnitude quake. Although there were no casualties, the regional center of Christchurch and the township of Kaiapoi in the Waimakariri District (to the north) suffered extensive land damage in the form of liquefaction, lateral spread (surface rupture and slippage), and rockfall. Six months later, a 6.3 magnitude earthquake occurred almost directly under the city of Christchurch. Though “smaller” in magnitude, the peak ground acceleration of 2.2 (over twice that of gravity) was one of the highest ever recorded and would have “flattened” most world cities (Anderson 2011). This time, 185 people lost their lives and thousands were injured. Almost 80% (Markham 2013) of core inner city buildings were (or will be) demolished; many schools and other major services and facilities closed either permanently or temporarily; there was significant, widespread infrastructural damage; and residents in some hillside and eastern suburbs were left homeless or living in substandard housing. Following this, a 6.3 magnitude quake on June 13, 2011, caused further damage, as did the 50 other quakes with a magnitude of 5 or more. As of March 2013 (Markham 2013):

- With 91 % of the 191,000 homes in greater Christchurch damaged, more than 450,000 residential land, building, and contents damage claims had been made to the Earthquake Commission across 200,000 South Island properties.
- Approximately 7,800 residential homes have been “Red Zoned,” i.e., tagged for demolition or removal with the land under them deemed too expensive to remediate at this time.

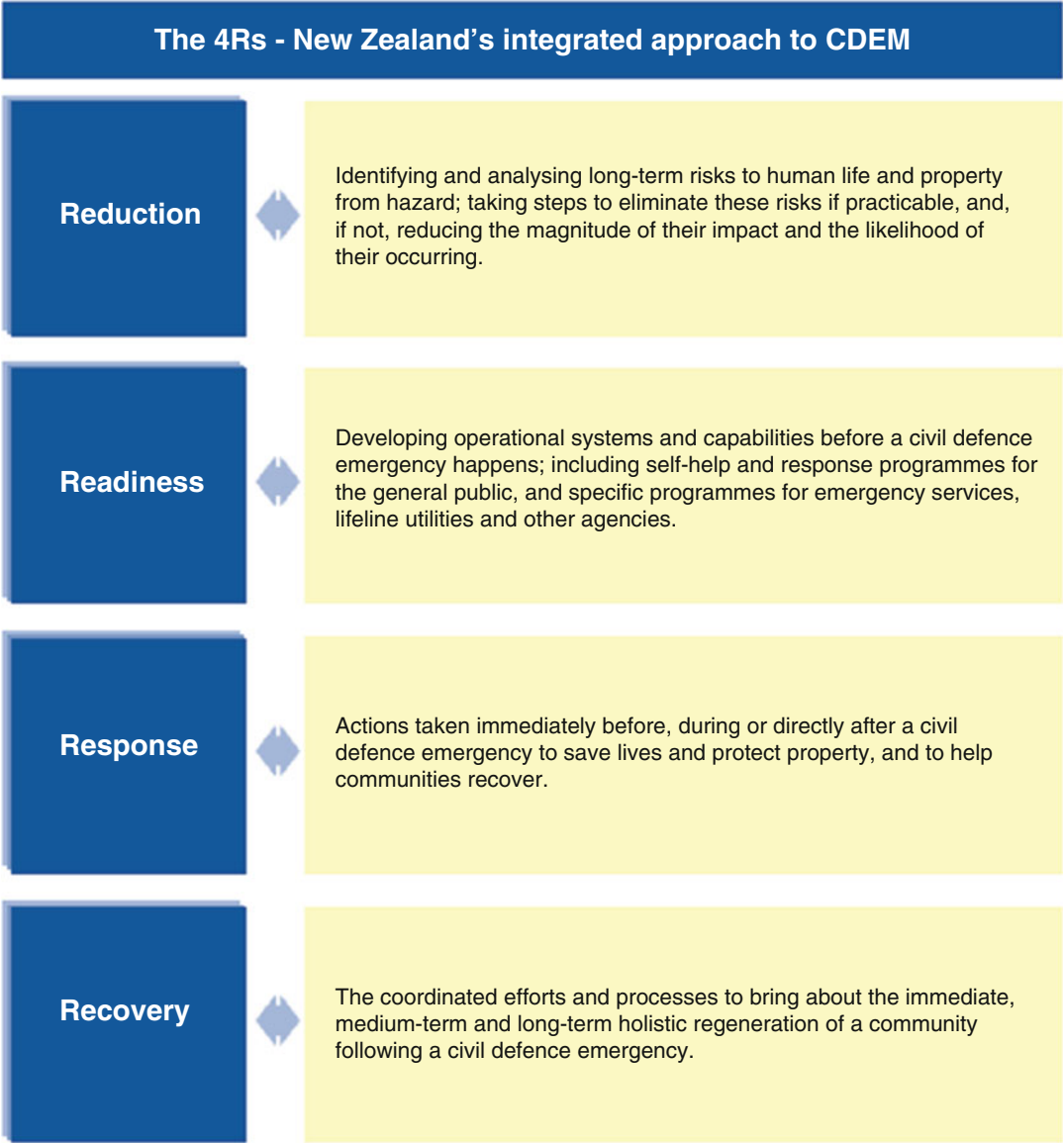
- Over 1,500 central business district (80 % of CBD building stock) and 300 suburban commercial buildings have been, or will be, demolished with over 4,100 businesses (and 30,000 employees) affected by the demolition exclusion zone (initially 390 ha).
- Loss of, or disruption to, 75% of hotel accommodation and key convention, cultural, and recreation facilities, i.e., 1,600 community facilities and more than 250 heritage buildings.
- At least 528 km of the wastewater network, 1,021 km of roads, 22 freshwater wells, and 30 bridges need to be replaced or repaired.
- An estimated rebuild cost of over \$40 billion.

These events have been referred to as the Canterbury Quakes as their effects have cut across district/city council administrative boundaries and have prompted regional migration. The extended sequence has put pressure on local authorities trying to balance “business as usual” with earthquake recovery. New Zealand also has some of the highest levels of insurance in the world; consequently, the region is undergoing extensive “demolish and rebuild” rather than “repair,” the scale of which is causing widespread disruption to residents and debris disposal issues (discussed in section [What is Governance and Why is It Important?](#)). Further, the Red Zone decision has meant the displacement and relocation of almost 8,000 households causing significant land use, infrastructure, housing, and investment changes.

It is within this context that the following sections outline the role of land use planning both pre and post event.

The Role of Land Use Planning in Earthquake Recovery

To ensure consistent outcomes are achieved during recovery, land use planning and emergency management objectives need to be integrated (Saunders et al. 2007). In New Zealand, emergency management is generally



Land Use Planning Following an Earthquake Disaster, Fig. 1 The 4R's of integrated civil defense emergency management in New Zealand (MCDEM 2008, p. 5)

divided into four complementary phases, often referred to as the “4R’s”: reduction (i.e., prevention, mitigation), readiness (i.e., preparation), response, and recovery (see Fig. 1). Land use planning has a key role in the reduction and recovery aspects: reducing risks through avoidance and mitigation of natural hazards, both pre event and during the recovery phase.

How planning and planners respond to recovery is very much dependent on the governance arrangements in place – both pre and post event – as the following discussion outlines.

Role of Governance in Pre- and Post-event Land Use Planning

Disaster management systems around the world have evolved gradually, largely in response to

disastrous events. Typically, the management of disasters, especially large-scale disasters, is a responsibility of governments, as societal institutions and life-sustaining systems are in crisis. During the last few decades, the disaster management practice has been characterized by frameworks and processes based upon a shared system of governance and policy making, with common and overlapping responsibilities apportioned among layers of government (May and Williams 1986). The approach is *ideally* structured to engage from the “bottom-up,” with local governments having primary responsibility for supplying disaster-related resources and regional, subnational, and national agencies providing support as requested. By design, the system requires extensive coordination and cooperation among all levels of government, as well as with the many NGOs involved in disaster management, often in a tiered governance structure (national, state/regional, local) (Johnson and Mamula-Seadon 2014). The related legislation, policies, guidelines, and plans together help to promote a more integrated natural hazards risk management framework and are commonly based upon sustainable development principles (Mamula-Seadon 2009; Saunders and Beban 2012).

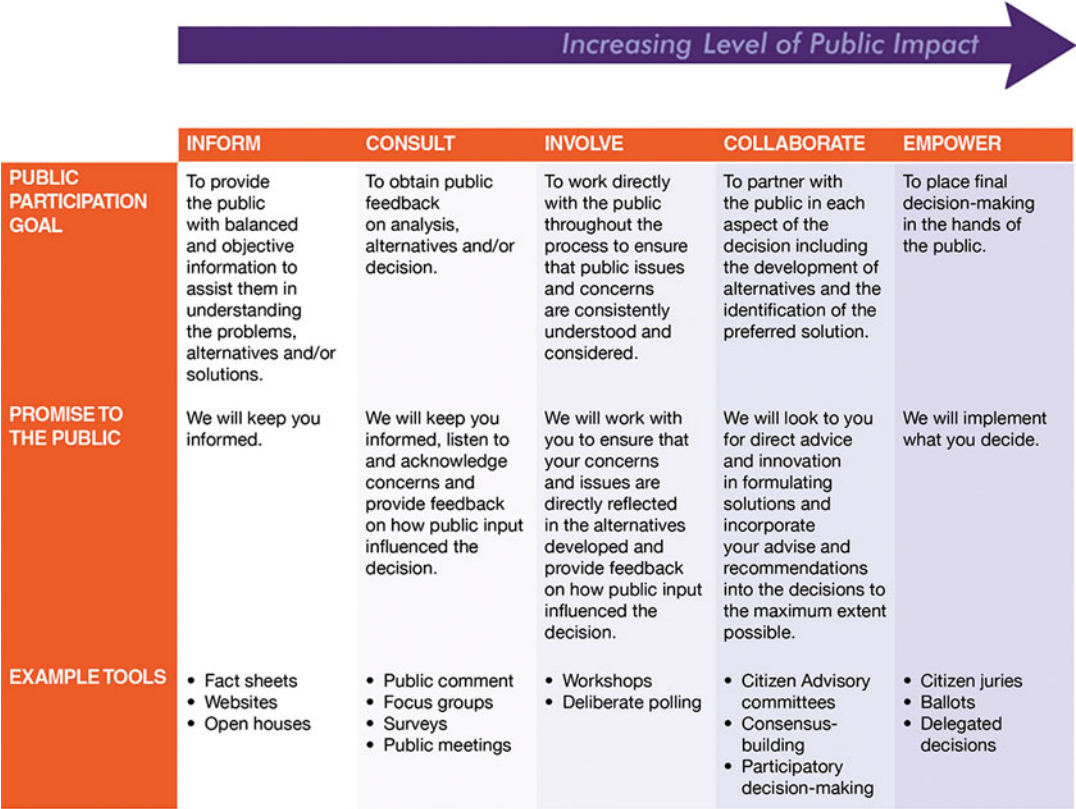
This “ideal” approach presumes integrated, decentralized, and deliberative planning and decision-making processes for land use, for development, and for comprehensive risk management, with the central government role defined as setting the direction and providing national policies, while the implementation matters were left to local government (Britton and Clark 2000; Mamula-Seadon 2009). However, recent societal and political changes, such as the emergence of public-private partnerships, outsourcing, and contracting, have meant that functions that may formerly have been carried out by public entities are now frequently dispersed among diverse sets of actors that include not only governmental institutions but also private sector and civil society entities.

What Is Governance and Why Is It Important?

Disaster governance takes a form of collaborative activities through “processes and structures of

public policy decision making and management that engage people constructively across the boundaries of public agencies, levels of government, and/or the public, private and civic spheres in order to carry out a public purpose that could not otherwise be accomplished” (Emerson et al. 2012, p. 2). Implications for land use planning have yet to be fully understood though there is no doubt that the implications are important. Recent literature has started to demonstrate important connections between “procedural” (the process) and “substantial” (the plan) issues that plague peacetime planning, but which are exacerbated by the conditions of uncertainty and urgency that accompany disasters. For example, Kweit and Kweit (2004) compared recovery processes in Grand Forks (North Dakota, USA) and East Grand Forks following severe floods in 1997. After the disaster, East Grand Forks engaged in extensive citizen participation initiatives and subsequently reported on high levels of political stability and citizen satisfaction. In contrast, Grand Forks instigated a more top-down, bureaucratic approach and has since experienced changes to their government structure, a high turnover of elected and appointed officials, and more negative citizens’ evaluations. Similarly, the faltering recovery of New Orleans post Katrina has been attributed, at least in part, to recovery planning processes that failed to engage local communities.

There is a lack of research detailing communities’ planning efforts and aspirations, or their engagement with formal state representatives *post disaster*; however, a great deal of work has been conducted on orthodox “peacetime” models based on, for example, the International Association of Public Participation 2’s spectrum (2014, see Fig. 2). This has a continuum of engagement/participatory practices that range from “token” or “passive” *informing* to consulting, involving, and collaborating, to “meaningful” or “active” *empowering* forms of participation where the agency agrees to implement the community’s decisions. Though these provide useful guidelines, the post-disaster context does add a layer of complexity to these typologies, largely because normal state processes of engagement



Land Use Planning Following an Earthquake Disaster, Fig. 2 IAP2's public participation spectrum (IAP2) (c) International Association for Public Participation www.iap2.org

may be suspended (formally under a state of national emergency or informally due to dysfunction).

Large-scale disasters deplete capital stock and services to a degree where complex rebuilding and societal activities (involving changes to governance) must happen in a compressed period of time. Governments often struggle with these and other time-compressed demands of post-disaster recovery, and as a result, preexisting governance structures – including those around the type and extent of public involvement – are modified or new structures are introduced to help provide more capacity, information, money, and other resources (Johnson and Olshansky 2013).

While these observations have been made, as yet, there has been little detailed study as to how governance structures and processes are often modified post disaster, the conditions by which

governance transformations and new or modified institutions for managing recovery are likely to occur, or what the new structures are able to achieve compared to preexisting structures (Johnson and Mamula-Seadon 2014). Unlike top-down or bottom-up approaches to disaster management, the need for flexibility calls on governments and communities to adapt to, and negotiate, different roles, responsibilities, and opportunities seeking innovative strategies for state-civil society engagement. These offer dramatically improved approaches to building better relations, stimulating more effective activities, and sustaining a flexible and adaptable organizational response (Multi-national Resilience Policy Group 2013).

It is also important to note the disaster recovery planning is embedded in more general governance structures, playing out within ideological,

socioeconomic, cultural, and environmental contexts. Examination of the Canterbury response and recovery process shows how post-disaster “time compression” can spotlight the strengths and weaknesses of disaster management processes and governance frameworks, including land use planning. For governance in general, the Canterbury transformations could be viewed as a rapid governance adaptation embedded within an overall trend in governance reforms underway in New Zealand (Johnson and Mamula-Seadon 2014). Before the earthquakes, the government had already instigated reforms to several key government policy frameworks and accompanying legislation, particularly the Resource Management Act 1991 and Local Government Act 2002, which represent two significant pieces of legislation that govern land use planning in New Zealand.

Ideally governance arrangements should allow for pre-event land use recovery planning to be undertaken. This type of planning requires political commitment and resources; however, the process can prove invaluable post event, as discussed in the following section.

Pre-event Planning for Post-disaster Recovery

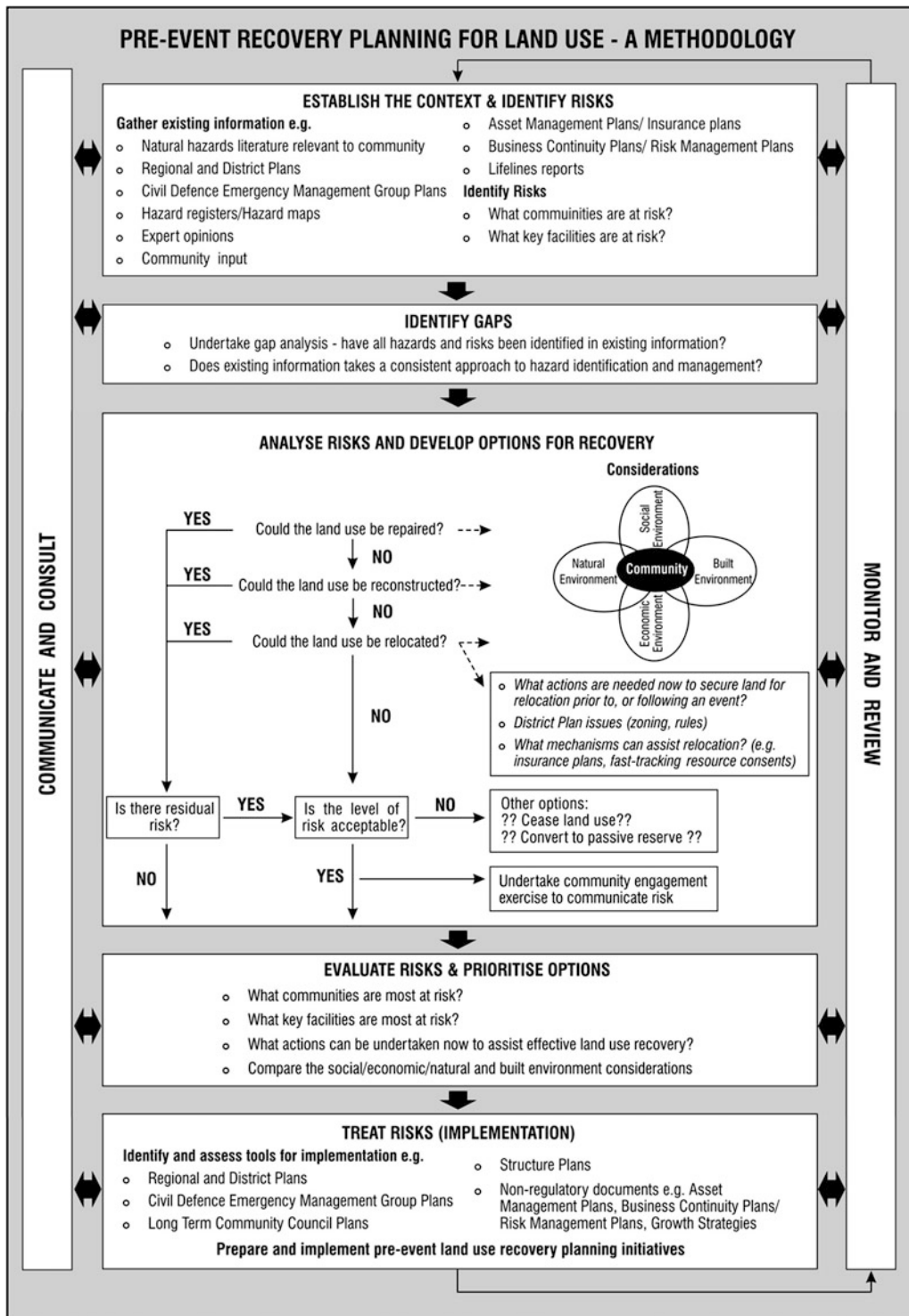
While land use planners have a critical role to play in the disaster recovery process, they are often overlooked when preplanning for disaster recovery (Smith 2011). How land use planning proceeds following an earthquake is dependent on what level of pre-event planning has been undertaken before the disaster occurs. The recovery phase can be both prolonged and political; while the response phase might demand rapid technical assessments of the structural safety of a building, the recovery phase associated with that same building may involve interwoven and subjective issues of security, weather-tightness, aesthetics, or broader contextual factors based on infrastructure provision, out-migration of residents, traffic, and access to services and facilities, including schools. Consequently, it may take years, or even decades, for something like “recovery” to occur.

Once a disaster occurs, land use planning efforts for recovery should not be something new – ideally, some planning should begin prior to an event occurring. Planning for land use recovery before an event occurs is different from emergency preparedness and response planning. However, those involved in preparedness and response planning need to be aware of any pre-event recovery planning, and vice versa, to ensure consistency in decision making (Johnson and Olshansky 2013). Where post-disaster recovery is undertaken in an environment that can be detrimental to successful outcomes (e.g., the meaningful involvement of key stakeholders, including communities), pre-event recovery planning allows key stakeholders to invest time and resources to support and encourage cooperative behavior (Smith 2011). Planning for post-disaster recovery and reconstruction requires a collaborative approach to undertake hazard identification, risk assessment, public education, consensus and partnership building, visioning, and the identification of activities that reduce risks (Johnson and Olshansky 2013). A poorly planned or inefficient response to a disaster can impose additional social, economic, and environmental burdens on a community already impacted by the disaster (Johnston et al. 2009).

An example of a framework for pre-event planning is provided in Fig. 3, from New Zealand. This framework has been based on the Standards Australia/New Zealand 31000:2009 Risk Management standard.

There are many opportunities for pre-event planning prior to an event. Questions that can be addressed include:

- Are current land use activities appropriate for their locations? What land is available for relocating particular land use activities?
- What are the social, cultural, economic, and other implications of the land use?
- What aspects of “place” do the community highly value, e.g., historic and cultural character?
- Are there any restrictions (e.g., other natural hazards, sites of cultural significance, building requirements) on the land?



Land Use Planning Following an Earthquake Disaster, Fig. 3 A framework for pre-event planning for post-disaster recovery from New Zealand (Becker et al. 2011, p. 538)

- Do current land use planning policies adequately address the potential for relocating land uses?
- Can the regulatory planning process for (re) development applications be streamlined?
- Where will debris from damaged/demolished buildings, landslides, liquefaction, etc., be disposed of? Consideration needs to be given to:
 - Contaminated material, green material, wood versus concrete, etc.
 - Overflow sites, or prearranging for debris to be transported to another location outside of disaster area
 - The location of staging and sorting sites
- Who are the key stakeholders and community members that need to be involved in preplanning and do they have a recognized role in, for example, civil defense and emergency management programs?
- How healthy are democratic and participatory processes and to what extent is decision-making distributed among different stakeholders?
- What are the short- and long-term planning requirements?
- Can any land use activities be relocated before an event occurs in order to reduce the risk?

There is a significant amount of international research and reporting on the value of pre-event planning for a range of natural hazards, including earthquake. It has also been recognized that preplanning for disaster recovery is important. An example of preplanning for debris disposal following an earthquake is that from New Zealand, where Johnston et al. (2009) have outlined 10 key steps that can be planned by both emergency managers and land use planners, ahead of time. These steps, with those key positions involved in each step, are shown in Table 1.

As shown in the example in Table 1, preplanning for a disaster requires many different parties to be involved in the development of the plan: in this case, emergency managers, planners, communication, and enforcement officers. Undertaking a multidisciplinary approach to pre-event planning, it results in a more comprehensive and robust planning process and outcome.

Pre-event planning can be valuable in building relationships between all of the involved parties, to produce a shared vision and responsibilities for completing pre-event planning tasks, rather than producing the actual plan. As shown in Table 1, building the relationships between the emergency managers, planners, and communication team pre event will mean valuable time is not required to develop these relationships post event. While pre-event planning for recovery should be undertaken before a hazard event occurs, planning can also occur after the event, either before or even while major reconstruction begins. Examples of this type of planning following earthquakes in China, Alaska, and Japan are outlined in Becker et al (2011).

Best practice around both pre-event and post-disaster land use planning shares some similarities, particularly in terms of emphasizing participatory processes. However, much of the theory around disaster-recovery land use planning either belies or underestimates the complexities involved. Recovery planning is unlike “peacetime” planning in that the landscape – geographical, social, economic, political, and conceptual – is likely to be changing rapidly even as the plans are developing. If a disaster is defined as an event which exceeds the ability of a system to cope with the disturbance, causing a loss of functionality, then “chaos,” “confusion,” and “disorder” are trademarks of the response phase. It is this dysfunction that arguably separates disaster from mere misfortune.

In restoring order – or functionality – and transitioning to the recovery phase, it can be helpful to draw upon pre-event comprehensive land use plans and strategies. This ensures that key principles and concerns embedded in “peacetime” (i.e., business as usual) documents continue to guide recovery planning, even if those plans and strategies need extensive modification.

If pre-event plans and strategies do require significant changes, recovery planners will be faced with some key challenges. The obvious need to act with urgency must be balanced against deliberative, inclusive, and participatory processes that are often seen as costly in terms of

Land Use Planning Following an Earthquake Disaster, Table 1 Ten steps for preplanning for debris disposal (Adapted from Johnston et al. 2009)

Key step	Positions involved
1. Develop debris disposal plans at a regional and local level, or make appropriate provisions in existing solid waste management plans. Plans should include a detailed strategy for debris collection, temporary storage and staging areas, recycling, disposal, hazardous waste identification and handling, administration, and communication with the public	Emergency managers
	Planners (i.e., making provisions in existing plans)
	Communication team
2. Identify potential locations or sites for the temporary and/or permanent disposal of debris. Outline a process for final site selection/confirmation in response to post-event requirements in both regional and local plans	Planners
	Emergency managers
3. Prepare a communication strategy ahead of time. You will need to tell your community when, where, and how normal rubbish collection will resume and give special instructions for reporting and sorting disaster debris	Communication team
4. Prepare for increased demands on council staff in terms of public information, operation, and enforcement of debris management systems. This may require additional staff resources to manage the increased workload	Communication team
	Enforcement officers
	Planners
	Emergency managers
5. Make arrangements for additional equipment and supplies to deal with disposal ahead of time. Identify the types of equipment and supplies that staff and contractors will need to carry out operations	Emergency managers
6. Develop mutual aid agreements with neighboring councils (and/or CDEM groups) for plant, equipment, and expertise	Emergency managers
7. Establish pre-event debris management contracts with private contractors	Emergency managers
8. Establish possible organizational structures, roles and responsibilities, and authority between various stakeholders, regulatory authorities, and decision-makers including Ministry for the Environment, Department of Labour, Ministry of Health, and within local authorities	Emergency managers
	Planners
9. Determine financing strategies at local, regional, and national level for differing sizes and types of disaster events	Emergency managers
10. Assess and develop an approach to the potential impact of disaster debris management on normal environmental processes and standards. For example, undertake a Strategic Environmental Assessment identifying practical strategies to minimize environmental and social impacts	Planners

time and other resources. An example concerns the demolition of damaged, and therefore hazardous, buildings versus the preservation of valued and meaningful heritage sites. “Participation” speaks to the notion that engineers, while experts in determining structural safety, may not be best placed to identify a community’s iconic buildings or to assess a building’s importance. There is now enormous debate about the extent to which seemingly apolitical structural or geotechnical matters impact on communities and how both urgent safety concerns and high-level strategic documents can be addressed collaboratively given the degree of trauma experienced by the affected population, the extreme need for timeliness, and, often, a lack of information and other resources.

In summary, pre-event planning for post-disaster recovery can be undertaken before an event occurs, or post-event before reconstruction begins. Pre-event planning should address both short- and long-term needs and involve key stakeholders, including community members. It should be adaptable to post-event realities – the importance of the preplanning is in relationship building and the active participation of key stakeholders.

The Reality of Planning Post-earthquake: The Canterbury Experience

Following the February 2011 earthquake, it was seen as necessary to create an administrative



body to oversee recovery in the post-quake Canterbury region. There were four key reasons for this decision:

1. The earthquake sequence was prolonged with new fault lines identified and significant aftershocks still occurring 2 years after the first event
2. The effects of the sequence were geographically widespread, cutting across four administrative boundaries: the Canterbury Regional Council (ECan), the Christchurch City Council (CCC), the Waimakariri District Council (WMK), and, to a lesser extent, Selwyn District Council (SDC).
3. With a rebuild cost of over \$40 billion, an external coordinating body was seen as necessary to manage the “the political and fiscal risks involved” (Brookie 2012, p. 28).
4. There were concerns that the Christchurch City Council was not able to cope adequately with the scale of the disaster. Councillors were publicly describing the council as “dysfunctional” and “lacking leadership,” and a divide between some of the elected representatives, the Mayor, and the bureaucracy was becoming apparent (Sachdeva 2011).

Consequently, in April 2011 the Canterbury Earthquake Recovery Authority (CERA) was created by an Act of Parliament. Essentially a government department, CERA is responsible to the Minister for Earthquake Recovery, Hon. Gerry Brownlee. CERA is a controversial entity because, through the Canterbury Earthquake Recovery (CER) Act of 2011, the Minister has extraordinary powers to expedite rapid decision-making. While this suits those who want to “get on with it,” others are concerned about the suspension of citizens’ rights and the lack of a layer of governance between CERA and the central government National Party. The CER Act (2011) gives the Minister the ability to suspend, amend, cancel, (or) delay ECan, CCC, or WMK plans or policies, and further, these local government recovery plans must be consistent with CERA’s recovery strategy and signed off by the Minister (Brookie 2012). Among other rights,

CERA’s chief executive has the authority to do the following:

- Require councils to act as directed and or to provide information on request
- Amend or revoke Resource Management Act documents and city plans (i.e., peacetime land use legislation)
- Close or otherwise restrict access to roads and other geographical areas
- Demolish buildings, or otherwise enter and deal with people’s land and property (with notice, in the case of Marae and dwelling houses)
- Require compliance of any person with a direction made under the Act

The CER Act thus forms a controversial legislative and political space where the ends justify whatever means – within which more detailed plans and strategies have been formed. We turn now to a more detailed description of five of these including:

1. The Greater Christchurch Recovery Strategy (GCRS) which is a high-level document providing *strategic direction and principles*
2. The pre-quake Greater Christchurch Urban Development Strategy (UDS) and the post-quake Land Use Recovery Plan (LURP) which has a *regional* focus
3. The Christchurch City Plan covering *metropolitan* Christchurch;
4. The Christchurch Central Recovery Plan (CCRP) which covers the Christchurch CBD;
5. The Waimakariri District Council’s Integrated and Community-based Recovery Framework which guided the infrastructure rebuild program in Kaiapoi

The Greater Christchurch Recovery Strategy (GCRS)

The GCRS was approved by the Canterbury Earthquake (Recovery Strategy Approval) Order 2012 made by the Governor-General. The strategy (Canterbury Earthquake Recovery Authority 2012):

IMMEDIATE	Repair, patch and plan • Provide basic human needs and support services. • Address health and safety issues. • Make safe or demolish unsafe and damaged buildings and structures. • Investigate, scope and initiate recovery programmes and initiatives. • Plan for land use and settlement patterns so land can be made available for displaced residents. • Conduct ongoing programme of investigation and research to understand the geotechnical issues and seismic conditions. Use this information to guide recovery activities and decisions on land suitability for rebuilding.
SHORT TERM	Begin to rebuild, replace and reconstruct • Engage both established and new communities and inform them about rebuilding and future planning. • Establish new social and health support and service delivery models. • Continue demolition of damaged buildings. • Continue repair and rebuild. • Deliver early projects to instil confidence. • Planning and supporting community resilience. • Begin replacement activity. • Begin restoration and adaptive reuse of heritage features. • Continue, monitor and review recovery.
MEDIUM TO LONGER TERM	Construct, restore and improve • Continue to build resilient communities. • Continue reconstruction. • Major construction projects are underway. • Complete restoration and adaptive reuse of heritage features. • Phase out recovery organisations. • Economy is growing and businesses are sustainable. • Labour market is active and attracting employees.

Land Use Planning Following an Earthquake Disaster, Fig. 4 Typical phases of recovery (Canterbury Earthquake Recovery Authority 2012, p. 14)

- Is a high-level document which defines what “recovery” means for greater Christchurch
- Establishes principles to guide how CERA and other agencies will work together towards recovery
- Describes in broad terms the pace and phases of recovery
- Identifies work programs and which organizations will lead specific projects
- Identifies priorities for recovery efforts
- Sets up governance structures to oversee and coordinate the work programs and links them to wider initiatives
- Commits to measuring and reporting on progress towards recovery

In particular, the plan sets out a framework and provides delivery mechanisms necessary to rebuild existing communities, develop new communities, meet the land use needs of commercial and industrial businesses, rebuild and develop the infrastructure needed to support these activities,

and take account of natural hazards and environmental constraints that may affect rebuilding and recovery.

The strategy identifies six components of recovery – cultural, social, built, natural (including geotechnical considerations), and economic. These components link together with the sixth component – community – at the center. Each requires leadership and integration.

The “recovery” process is then divided into three phases (see Fig. 4) – immediate, short-term, and long-term – with each phase assigned milestones and timelines. Immediate actions, like “repairing, patching, and planning,” for example, would ideally have been completed within a year of the first earthquake. Long-term “constructing, restoring, and improving,” on the other hand, may take until 2020 or beyond.

As an example of a recovery strategy, this document does give a good overview of six components that contribute to recovery (i.e., cultural, social, built, natural, economic, and community),

and these components should be considered by others in post-disaster recovery planning. It also provides a good indication of the sorts of recovery milestones associated with different stages of recovery that may usefully serve as a guide for other recovery agencies. On the other hand, implementation is key, and one of the main challenges has been to build both the capability and capacity necessary to deliver on the strategy. Much of the strategy's implementation has been overseen by CERA, but implemented by other local authorities (see Fig. 5) using a blend of both new and existing plans and processes.

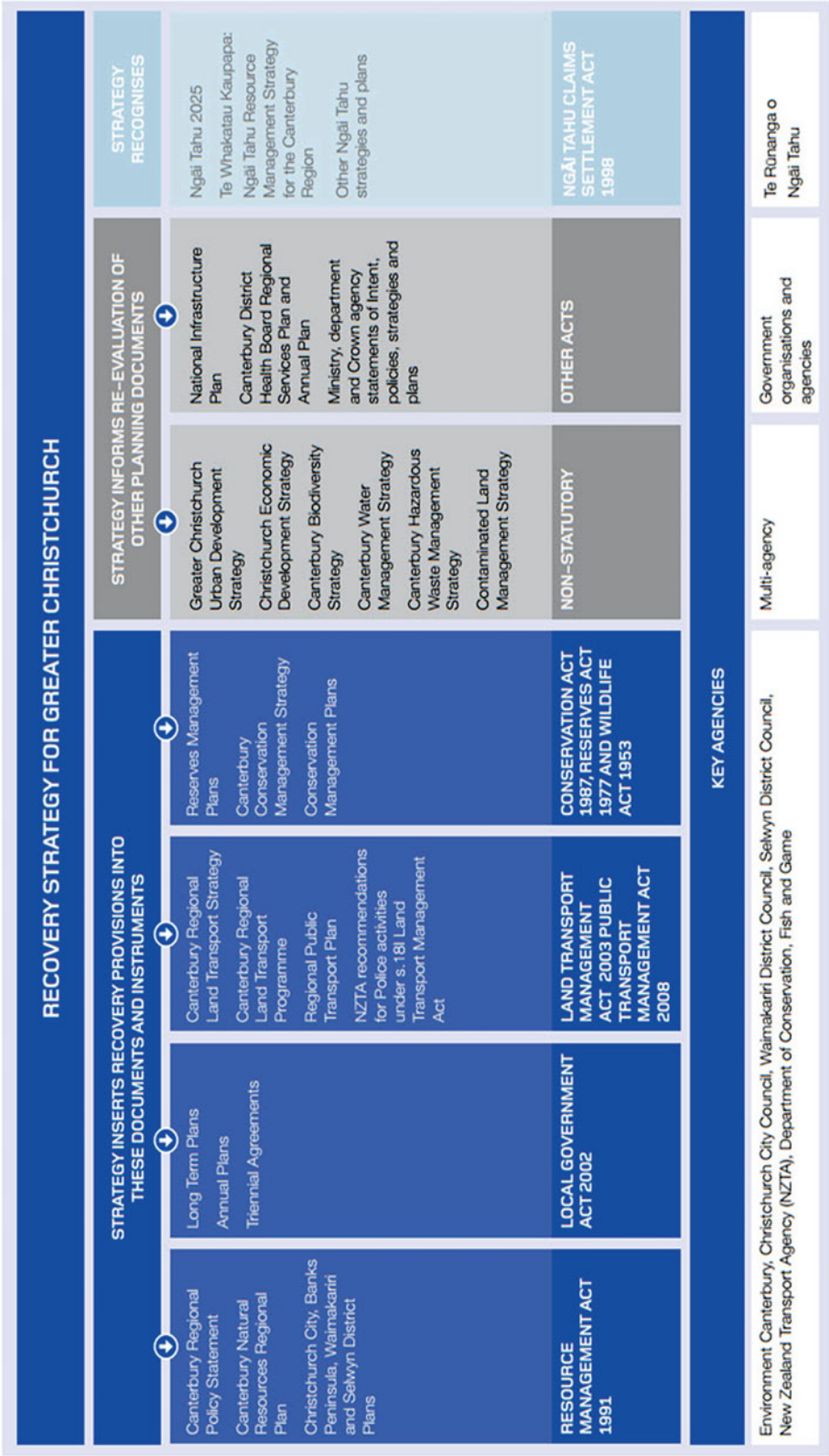
The Pre-quake Urban Development Strategy and the Land Use Recovery Plan

In a pre-quake process initiated in 2004 by the Canterbury Regional Council (ECan), the Greater Christchurch Urban Development Strategy (UDS) (Greater Christchurch 2013) was the result of 3 years of collaboration between stakeholders and communities. Based on over 3,250 submissions, the UDS sought to provide a framework for future growth management taking into account risk assessments, transportation, housing demand, and so on. Four options were presented that essentially represented unfettered sprawl at one end, to extreme containment and intensification around core suburban centers and townships in the districts at the other. Discussion around these four options resulted in a vision, guiding principles, strategic directions, and a framework for implementation that was given statutory weight in chapter 12 "Development of Greater Christchurch" of ECan's Regional Policy Statement. This, in turn, guided land use patterns – including areas that could be zoned for residential, commercial, and retail activities – in city- and district-level plans developed by the Christchurch City (CCC), Waimakariri District (WMK), and Selwyn District (SDC) Councils.

The UDS had already signaled areas in the region suitable for future growth when the earthquakes and subsequent Red Zoning decision displaced almost 8,000 households. The comprehensive planning behind the UDS allowed for a fairly rapid appraisal of land that could be

made available to accommodate the demand for new housing. However, the UDS was designed to be implemented over a period of 30 years and considered inadequate for the post-quake context. Consequently, in November 2012, the Minister for Canterbury Earthquake Recovery directed ECan (at their request) to prepare a Draft Land Use Recovery Plan (LURP) through a collaborative multiagency approach involving all of the earthquake recovery strategic partners (including the three local councils, Ngāi Tahu, and the New Zealand Transport Agency), with input from key stakeholders and the wider community.

Some of the key issues included tensions between market-led development and regulation, leadership, certainty, council processes, brown-field development, urban design, housing affordability, infrastructure, public transport, and suburban centers (AERU, 2012). Further, although the LURP had been through several consultative rounds, as the agency responsible for the city plan that will give effect to the LURP, the CCC was particularly concerned that residents had not been sufficiently informed about, for example, the new "design-led" Comprehensive Development Mechanisms (CDMs). These would see high-density housing (defined within the Canterbury cultural context as between 30 and 65 households per hectare) throughout the city. Given that even medium-density housing has a controversial history in Christchurch, the CCC argued for CDMs to be limited to Living 2 and 3 Zones (with minimum site sizes of 330 m² and 300 m², respectively). As so much of the city's public (social) housing stock has been damaged or demolished, a new Community Housing Development Mechanism (CHDM) has also been introduced to facilitate the urgent rebuilding of social housing stock in the city. These are likely to be located in areas that already accommodate such housing. This, too, is likely to be controversial, though criticism may be mitigated somewhat by the review of both mechanisms that is to take place after 5 years of implementation. Nonetheless, tension between the CCC and CERA over CDMs and CHDMs has climbed to very high levels.



Land Use Planning Following an Earthquake Disaster, Fig. 5 The relationship between the recovery strategy and other strategies, policies, and plans for greater Christchurch (Canterbury Earthquake Recovery Authority 2012, p. 22)

The LURP also identifies areas deemed suitable for greenfield development, and with the Minister's approval, statutory effect will be achieved through the insertion of chapter 6 "Recovery and Rebuilding of Greater Christchurch" into the Canterbury Regional Policy Statement. Chapter 6 will then inform district and city plans that more specifically determine the location, type, and mix of business and residential activities within specific geographical areas based on hazard identification, transportation routes, and migration patterns. Through this blend of regional, district, and city plans, the LURP is a significant document that has adopted much of the pre-quake UDS vision; however, it has been described as a "developer-friendly" document that gives the market considerable scope. It has also been criticized for promoting – through peri-urban greenfield and suburban mall development – a further hollowing out of the inner city, the rebuild of which falls under the Central City Recovery Plan (discussed further in section [The \(Christchurch\) Central City Recovery Plan](#)).

The Christchurch City Plan

City and district plans are a pre-quake requirement of all territorial local authorities in New Zealand under the Resource Management Act (1991). These local council plans provide a framework for land use and subdivision in a manner that is consistent with the Regional Council's Plans and Policy Statements (e.g., the UDS/LURP) and are reviewed every 10 years. Christchurch's City Plan was due for review when the quake sequence began and was subsequently delayed until 2013.

Given the extent of the changes that will need to be made to the plan, the council has decided to undertake the review in two stages with the first notification due in 2014 and set to address overall strategic direction; residential, commercial, and industrial land use; subdivisions (e.g., minimum site size); natural hazards; transport; and contaminated land. The second round (in 2015) will notify issues around heritage and the natural environment, the coast, public open space, and so on. Submissions can

be made as the review takes place or once the draft is notified.

While the council has signaled the process may take 3 years, some of the proposed changes – such as the CDMs, CHDMs, and greenfield development mentioned above – are likely to be controversial as are other issues around, for example, maximum height and parking restrictions, "permitted" activities such as light retail, and so on.

The (Christchurch) Central City Recovery Plan

Following the February 22, 2011, earthquake – which led to the demolition of almost 80 % of building stock in the CBD core (Markham 2013) – the Christchurch City Council began the task of developing a recovery plan for the central city. In May of 2011, the CCC initiated the 6-week *Share an Idea* campaign which, according to the CCC website (Christchurch City Council 2013a), allowed "the public to tell us their ideas about how the Central City should be redeveloped to be a great place again. ..." Combining a weekend of on-site opportunities and an online crowd-sourcing tool, the campaign generated more than 106,000 ideas from over 60,000 participants. These ideas were analyzed so as to inform the draft Central City Plan and guide inner city development for the next 10–20 years.

Share an Idea was followed by another participatory tool – the *48 h Design Challenge* – run by the CCC at Lincoln University. This provided an opportunity for the council to work with the design and architecture industry and practically test the draft Central City Plan over seven preselected sites. A total of 15 teams comprising engineers, planners, urban designers, architects and landscape architects, and students took part in the *Challenge*.

Following these exercises and other, smaller workshops, the Draft CCP was delivered to the Mayor to sign off in August of 2011 before being sent to the Minister for Earthquake Recovery for his approval. Submissions were invited and 79 comments were received. As outlined on the council's website (Christchurch City Council 2013b), the Minister's view was that "taking into account its impact, effect and funding

implications, [the Plan] could not be approved without amendment...given insufficient information...on how the Recovery Plan would be implemented and changes to the District Plan that [are] unnecessarily complex.”

Consequently, the Minister established a special unit within CERA, the Christchurch Central Development Unit (CCDU) which was to work with a “professional consortium” to deliver, within 100 days, the Central City Plan and take a lead role in its implementation alongside the CCC, Te Rūnanga o Ngāi Tahu, and other key stakeholders. The CCRP is based on a master-planned blueprint for the CBD providing street layout, the location of “anchor” projects (including a convention centre and sports facility), and various precincts. These they hope will stimulate further investment and development (though this has been hamstrung by the need to meet new building standards, current requirements for parking spaces, and so on, often with insurance pay outs that only covered the replacement value of old buildings).

Given the responsibility for the 100 Day Plan was taken from the CCC and given to a nonelected, newly created entity – the CCDU – amidst controversy around compulsory land acquisition to enable the anchor projects to go ahead, the level of public involvement may have been limited. Though the plan’s executive summary states that “The city centre [will be]...a reflection of where we have come from, and a vision of what we want to become” (Central City Development Unit 2012, p. 4), it has been suggested that, in practice, the process has been “top-down, centralised, and highly bureaucratic” (in Brookie 2012, p. 29). Drawing on the IAP2’s spectrum of participation (see Fig. 2 above), it could be argued that much of the “involvement” behind the plan’s development is limited to “informative and consultative” rather than “collaborative and empowered.”

Underlying the attempt to design a Central City Plan within 100 days is a sense that urgency both demands and justifies the suspension of peacetime processes around public engagement, transparency, and accountability. It is claimed, for example, that “The experience of other cities

after a natural disaster shows that substantial redevelopment must start within 3 years if recovery is to be successful. One year has passed. Speed is of the essence...” (Central City Development Unit 2012, p. 4). Consequently, the policies and procedures for land acquisition have been both praised as providing swifter certainty for homeowners and businesses to relocate and recover and criticized as having evolved over time with often reactive and unclear reasoning. Similarly, while the CCDU has been credited with accelerating the design and implementation of the CBD rebuild, concerns have also been raised about the acceptance of development applications without the establishment of clear criteria for screening and selecting among competing applications (McDonald 2013).

This raises questions about whether it is possible to effectively and adequately reconcile the conflict between the perceived need for urgent decision-making and action, on one hand, and robust process on the other. In exploring possible answers to this question, we turn now to an example of recovery best practice: The Waimakariri District Council’s Integrated and Community-based Recovery Framework.

The Waimakariri District Council’s Integrated and Community-Based Recovery Framework and the Kaiapoi Infrastructure Rebuild

The Waimakariri District is adjacent to the city of Christchurch with a population of about 50,000 people. It has traditionally been described as agricultural with large farms, lifestyle blocks, and smallholdings devoted to horticulture; however, a few settlements have grown rapidly and now form a commuter corridor of townships, with many residents traveling to work in Christchurch each day.

After the September 2010 earthquake, parts of the Kaiapoi township suffered extensive land damage from liquefaction and lateral spread, and over 1,000 households and the town center were severely affected. Besides serious damage to a quarter of its housing stock, Kaiapoi also lost its library, aquatic center, community halls, churches, the movie theater, bars, and cafes. Over 25% of businesses were immediately

affected and there have been significant relocations and a number of permanent closures. Much of the infrastructure was affected leaving 5,000 people without water and sewer. The damage, *proportionally*, was similar to that experienced in Christchurch.

Despite it being a Saturday, many WMK staff members came to work either to assume their post in the Emergency Operations Centre or to volunteer to help in Kaiapoi. By late afternoon, council staff members were distributing information on foot to affected residents. Other departments from “business as usual” started taking place, some of which meant literally and figuratively “stepping across a boundary.” As reported by a WMK senior manager (see Vallance 2013 for the full document):

Traditionally councils do not step across the home-owner’s boundary and any infrastructure issues between the house and the front boundary is the home-owner’s problem. But [due to demand] post-earthquake it would have been impossible to just call a plumber to get the issue fixed. So we [Waimakariri District Council] made a decision fairly early on to liaise with [and secure funding from] the Earthquake Commission and coordinate repairs *across the boundary* because there’s no point us fixing our side of the sewer and people still not being able to use [the toilet] because the pipe between the house and the boundary is broken (*italics added*).

Other councils did not take on this coordination role as quickly and, as a result, some of their residents suffered for months with poor sanitation.

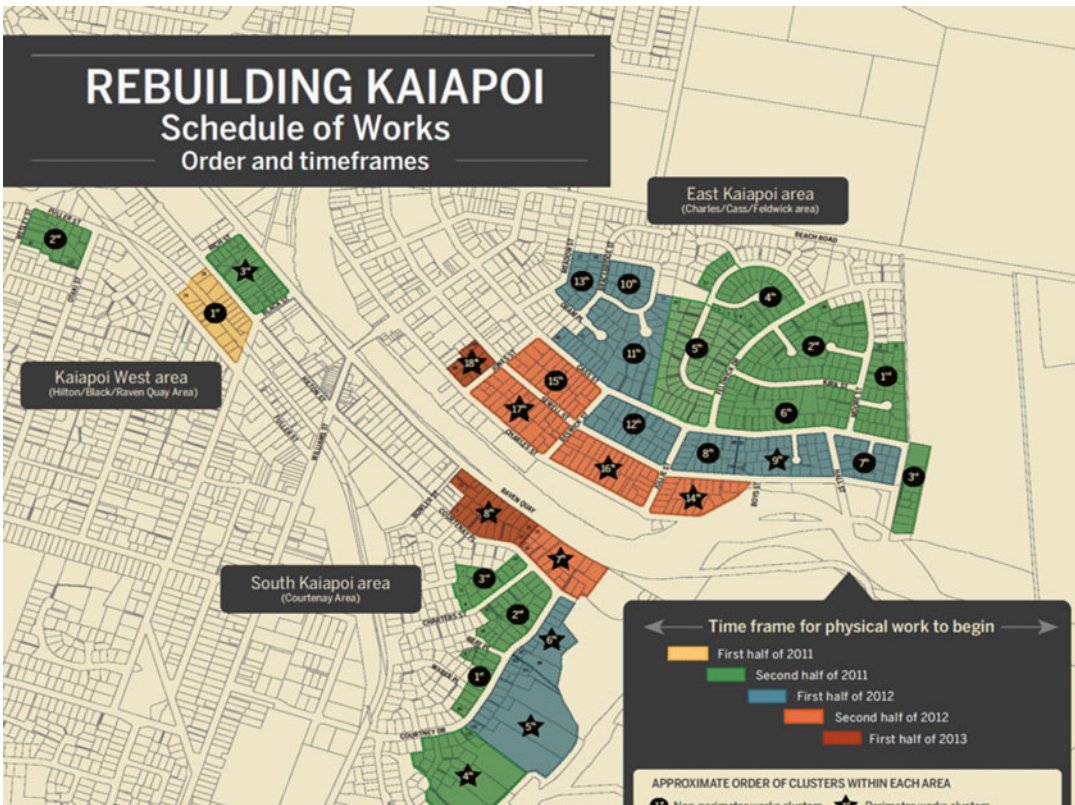
Other initiatives included three new Recovery Manager positions. The Social Recovery Manager (previously the council’s peacetime Community Team leader) was moved from the Rangiora offices to Kaiapoi to “colocate” with engineering and utilities council staff and contractors at the newly established Recovery Assistance Centre. She began acting as a liaison between engineers and contractors and Kaiapoi residents, local NGOs, and community groups. This communication between, and early involvement of, locals significantly shaped subsequent decisions to undertake the project management of what was thought to be (at the time) a neighborhood infrastructure rebuild program.

The neighborhood infrastructure rebuild program – coordinated by the newly appointed Infrastructure Recovery Manager – served about 1,200 households, broken down into smaller clusters for rebuild in different stages. It was argued that 1,200 was about the “right amount” and that if the project had been much larger, a different approach may have been required.

Importantly, the communication with residents around this staged rebuild was both deliberative and inclusive and began with numerous public meetings. These included “Q and A sessions” where, often, residents were seated according to the street in which they lived so they could get to know each other (for support) and triangulate information at a very local level. Elected councillors and senior managers from the council (often including the CEO and/or the Mayor) along with representatives from the insurance companies and geotechnical experts were also present. These meetings were well documented and the notes turned into a booklet that was then put on the council’s New Foundations website that had been established solely to disseminate earthquake-related information.

The council then published an order of works with timeframes (see Fig. 6). This was critical as it allowed affected residents to plan ahead for their temporary relocation. However, it is noted that events overtook the feasibility of that particular plan, so it was not implemented. The council also helped coordinate the provision of temporary housing located on council reserve land to accommodate residents displaced by the rebuild. Residents who were “last in the queue” of the staged rebuild were supported by the volunteer-led pastoral care team who liaised weekly with the Social Recovery Manager and who now shared a port-a-com with the Infrastructure Recovery Manager at “the Hub” in Kaiapoi.

The Hub was designed to be a “one-stop recovery shop” and was located in the council-owned community hall in Kaiapoi. It also housed the council’s pre-quake Kaiapoi Service Centre staff, some of the council’s Building Unit and Community Team usually located in Rangiora, Work and Income New Zealand, the Inland Revenue Department, local NGOs providing business



Land Use Planning Following an Earthquake Disaster, Fig. 6 The Waimakariri District Council's Schedule of Works for Kaiapoi (www.newfoundations.org.nz)

and whānau support, psychosocial and pastoral care teams, a tenancy service, Fletchers (one of the approved building repair companies), and the 15 Waimakariri Earthquake Support Service (WESS) coordinators who, in the first 2 years, assisted between 400 and 600 cases with accommodation, help with earthquake claims, legal aid, and counseling for earthquake stress. A dedicated earthquake communications manager was also appointed and colocated in the Hub. This colocation, right in the heart of Kaiapoi, enabled engineering expertise to reach residents and community intelligence to inform infrastructural recovery. The communication between the Social Recovery and Infrastructure Recovery Managers helped the engineering and utilities teams to appreciate the positive and negative impacts of their work and for residents to better understand what engineers and other experts were doing, and why.

While it is difficult to make direct comparisons between Christchurch and Waimakariri given the different contexts, types of plans, and extent of damage, satisfaction with the Waimakariri District Council's recovery framework for the first 3 years is relatively high. Indeed, as reported in Vallance (2012), a Residents' Association representative said:

There is a very good understanding on behalf of the majority of the community that the Waimakariri District Council understood the community's needs, that recovery was shared with the community so the community were engaged in different ways. If you want an example of recovery best practice, you've got it right there.

Wider support for their framework may also be reflected in the local government elections of 2013 where 9 of 10 councillors were reelected, with the Mayor unchallenged. In Christchurch, the results were quite different with only 4 of 13 representatives reelected (with the Mayor

and two others resigning). There are, therefore, some interesting parallels between these results and those documented by Kweit and Kweit (2004), reported above.

Summary

This chapter has outlined the role of land use planning during the complex and demanding recovery phase, with examples from the Canterbury region in New Zealand. Land use planning has a key role to play in the recovery phase; however, to strengthen this role, planning needs to be undertaken before an event occurs.

While any pre-event plan will need to be adapted or even rewritten post an event, the value is in the process of bringing key stakeholders together to discuss issues, agree on common goals, and form relationships that will be key during a recovery phase. Pre-event recovery planning can be useful as a means of rapidly appraising available land, as the Urban Development Strategy showed; however, such plans need to be flexible and adaptable. The Greater Christchurch Recovery Strategy highlights the value of developing an overall guiding strategy, but building capability and capacity to deliver on the strategy can be difficult and lead to high levels of tension between different recovery agencies. To support pre-event recovery planning for land use, governance arrangements need to be in place that enable pre-event planning and support the planning process post event when time compression factors influence the “business as usual” community consultation and engagement process.

As the five case studies from Canterbury have shown, the reality of recovery is complex, involving political intervention from a national level into local decision-making and governance arrangements. This in turn has directed how recovery plans have been produced, the involvement of the community in their development, and their outcomes. The Canterbury earthquake case study highlights some important lessons for those involved in recovery planning and the relationship between process and reality.

Cross-References

- ▶ [Building Disaster Resiliency Through Disaster Risk Management Master Planning](#)
- ▶ [Community Recovery Following Earthquake Disasters](#)
- ▶ [Earthquake Disaster Recovery: Leadership and Governance](#)
- ▶ [Learning from Earthquake Disasters](#)
- ▶ [Legislation Changes Following Earthquake Disasters](#)

References

- Anderson V (2011) Tuesday quake ‘no aftershock’. The Press, Christchurch
- Becker JS, Saunders WSA et al (2011) Preplanning for recovery. In: Miller DS, Rivera JD (eds) Community disaster recovery and resiliency: exploring global opportunities and challenges. Taylor & Francis, Boca Raton, pp 525–550
- Britton NR, Clark GJ (2000). Non regulatory approaches to earthquake risk reduction: the New Zealand experience. In: 12th world conference on earthquake engineering, Auckland
- Brookie R (2012) Governing the recovery from the Canterbury Earthquakes 2010–11: the debate over institutional design. Institute for Governance and Policy Studies, Victoria University, Wellington
- Canterbury Earthquake Recovery Authority (2012) Recovery strategy for Greater Christchurch Mahere Haumanutanga o Waitaha. Canterbury Earthquake Recovery Authority, Christchurch
- Central City Development Unit (2012) The plan. <http://ccdu.govt.nz/the-plan>. Retrieved Dec 2013
- Christchurch City Council (2013a) Share an idea. <http://www.ccc.govt.nz/homeliving/civildefence/chcheearthquake/ShareAnIdea.aspx>. Retrieved 15 Dec 2013
- Christchurch City Council (2013b) Central city. <http://www.ccc.govt.nz/homeliving/civildefence/chcheearthquake/centralcityplan.aspx>. Retrieved 15 Dec 2013
- Emerson K, Nabatchi T et al (2012) An integrated framework for collaborative governance. J Public Adm Res Theory 22(1):1–29
- Greater Christchurch (2013) The greater Christchurch urban development strategy. <http://www.greaterchristchurch.org.nz/>. Retrieved 1 Dec 2013
- IAP2 (2014) Public participation spectrum. <http://www.iap2.org.au/documents/item/84>. Retrieved 7 Jan 2014
- Johnson L, Mamula-Seadon L (2014) Transforming governance: organizing for and managing disaster recovery following the September 4, 2010 and February 22, 2011 Canterbury earthquakes. Earthquake Spectra

- (Special Edition – Canterbury Earthquakes) Vol 30, No 1, pp 577–605
- Johnson LA, Olshansky RB (2013) The road to recovery: governing post-disaster reconstruction. *Land Lines* (July), The Lincoln Institute of Land Policy, https://www.lincolnst.edu/pubs/2259_The-Road-to-Recovery-Governing-PostDisaster-Reconstruction, p 8
- Johnston D, Dolan L et al (2009) Disposal of debris following urban earthquakes: guiding the development of comprehensive pre-event plans, GNS science report 2009/33. GNS Science, Lower Hutt, p 27
- Kweit M, Kweit R (2004) Citizen participation and citizen evaluation in disaster recovery. *Am Rev Public Adm* 34(4):354–373
- Mamula-Seadon L (2009) CDEM, integrated planning and resilience: what is the connection? *Tephra Commun Resil: Res, Plan Civ Def Emerg Manag* 22:3–8
- March A, Henry S (2007) A better future from imagining the worst: land use planning and training responses to natural disaster. *Aust J Emerg Manag* 22(3):17–22
- Markham S (2013) The Waimakariri District Council's recover approach and programme. Otago CDEM Recovery Workshop, Alexandra
- May PJ, Williams W (1986) Disaster policy implementation: managing programs under shared governance. Plenum, New York
- MCDEM (2008) National civil defence emergency management strategy 2007. Department of Internal Affairs, Wellington
- McDonald L (2013) Call for action on mall impasse. The Press, Christchurch
- Multi-national Resilience Policy Group (2013) The New Zealand discussion – notes from meeting, Wellington/Christchurch
- Sachdeva S (2011) 'I'm off', says council identity Williams. The Press, Christchurch
- Saunders WSA, Beban JG (2012) Putting R(isk) in the RMA: technical advisory group recommendations on the resource management act 1991 and implications for natural hazards planning, GNS science miscellaneous series 48. GNS Science, Lower Hutt, p 52
- Saunders WSA, Forsyth J et al (2007) Strengthening linkages between land-use planning and emergency management in New Zealand. *Aust J Emerg Manag* 22(1):36–43
- Smith G (2011) Planning for post-disaster recovery: a review of the United States disaster assistance framework. Public Entity Risk Institute, Fairfax
- Vallance S (2012) Urban resilience: Bouncing back, coping, thriving. A paper presented at the Australian & New Zealand Disaster and Emergency Management Conference, Brisbane – 16 – 18, April 2012
- Vallance S (2013) The Waimakariri District Council's Integrated Community-based Recovery Framework. Department of Environmental Management, PO Box 84, Lincoln University, 85084

Laser-Based Structural Health Monitoring

Hoon Sohn and Byeongjin Park

Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology (KAIST), Yuseong-Gu, Daejeon, Republic of Korea

Synonyms

Damage detection; Laser Doppler vibrometer; Laser scanning; Laser thermography; Laser ultrasonics; Light detection and ranging; Nondestructive testing; Structural health monitoring

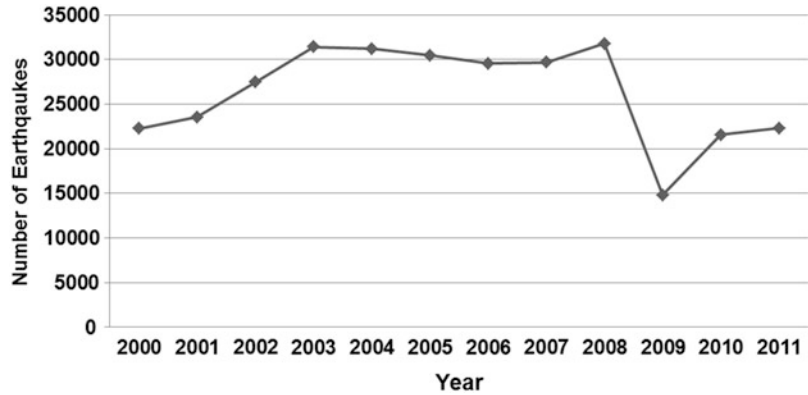
Introduction

With an average occurrence rate of 20,000 events worldwide every year, earthquakes have produced innumerable fatalities and economic loss (Fig. 1). For example, the recent Haiti earthquake (2010) was responsible for at least 300,000 injuries, 316,000 deaths, and 300,000 destroyed houses. There is a high demand for rapid and real-time health evaluation of post-earthquake structures as it is critical to distinguish which structures are inhabitable and serviceable and which are damaged and unavailable anymore. However, current evaluation methods are labor intensive, time consuming, and not cost effective.

Structural health monitoring (SHM) refers to the process of determining and tracking structural integrity and assessing the nature of damage in a structure (Chang et al. 2003). The process of SHM usually consists of four stages: (1) damage definition, (2) damage identification, (3) damage localization, and (4) damage quantification. SHM has been studied since the 1970s and primarily studied within the context of contact-type sensors. These contact sensors include accelerometers, strain gauges, acoustic emission transducers, and fiber-optic sensors, to name a few. However, it is challenging to utilize these sensors for

Laser-Based Structural Health Monitoring,

Fig. 1 Number of worldwide earthquakes during 2000–2011 (Source: United States Geological Survey)



post-earthquake monitoring as most of the installed sensors would be destroyed during the earthquake, and there would be no enough power supplies to operate these sensor systems.

In this entry, noncontact laser sensing is introduced as an alternative means for SHM. Light detection and ranging (LiDAR) and laser Doppler vibrometer (LDV) are two of the most common laser-sensing systems. A LiDAR estimates a distance from the device to a target point by measuring the time of flight (TOF) of the incident laser pulse reflected off from the target point or by measuring the phase shift of the reflected laser beam with respect to the incident continuous wave laser beam (Wehr and Lohr 1999). On the other hand, an LDV estimates the out-of-plane velocity of a vibrating target by relating the velocity to the frequency shift of the reflected laser beam with respect to the incident laser beam (Castellini et al. 2006). Furthermore, these devices can be easily deployed for long-range sensing and large area scanning. Figure 2 shows typical LiDAR and LDV devices. Riegl VZ-6000 LiDAR (Fig. 2a) can measure up to 6 km with an accuracy of 15 mm and a measurement speed of 37,000 points/s. Its measurement speed can be increased up to 222,000 points/s, but the maximum measurement range is shortened to 3.3 km. Polytec PSV-500 LDV (Fig. 2b) can measure up to 50 m/s vibration with a resolution of $2 \mu\text{m/s}/\sqrt{\text{Hz}}$ and a maximum vibration frequency of 2.5 MHz.

This entry is organized as follows: In Section “Working Principles of LiDAR and

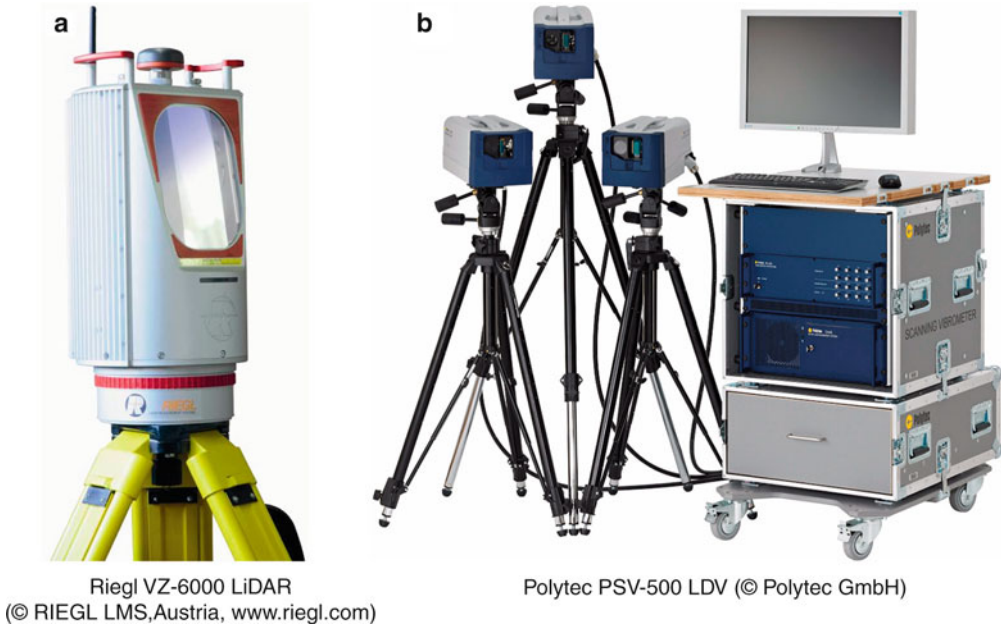
LDV,” airborne laser scanning techniques for regional damage assessment are summarized. Section “Airborne Laser Scanning for Regional Monitoring” presents ground-based laser scanning techniques for global and local damage detection, and Section “Ground-Based Laser Scanning Techniques” discusses laser ultrasonics and thermography for local damage detection. Finally, the conclusions and discussions are summarized in Section “Other Laser-Based Techniques for Local Damage Detection.”

Working Principles of LiDAR and LDV

In this entry, the working principles of LiDAR and LDV are discussed in more detail. LiDAR can work in two manners, (1) TOF measurement and (2) phase-shift measurement. The first one is more commonly used for commercial LiDAR, which is rather simple to understand (Fig. 3a). A high-power laser pulse is emitted from the LiDAR, and the laser pulse reflects back from the target object. When the reflected laser pulse is detected by the LiDAR, the distance to the target D is determined as

$$D = \frac{T}{2}c \quad (1)$$

where T and c denote the TOF, which is the time difference between laser pulse emission and reception, and the speed of light (3×10^8 m/s), respectively. As the distance measurement via LiDAR



Laser-Based Structural Health Monitoring, Fig. 2 Representative laser scanning devices: (a) light detection and ranging (LiDAR) and (b) laser Doppler vibrometer (LDV)

depends on the reception of the reflected laser pulse, the sensing ranging of LiDAR is mainly limited by the laser pulse power. Some commercial LDV devices with a very high-power pulse can measure distance up to several kilometers.

Displacement measurement based on phase shift estimates the TOF information from the phase shift of the reflected laser beam (Fig. 3b) with respect to the incident laser beam. It emits an amplitude modulated laser beam and compares it with the reference laser beam. Then the phase difference φ between two laser beams can be expressed as follows:

$$\varphi = 2\pi T f_m \quad (2)$$

where f_m refers to the modulation frequency of the laser beam. By substituting Eq. 2 into Eq. 1, the distance is estimated as

$$D = \frac{T}{2} c = \frac{\varphi}{4\pi f_m} c = \frac{\varphi}{4\pi} \lambda_m \quad (3)$$

where λ_m refers to the wavelength of the amplitude modulation which is typically ranging from

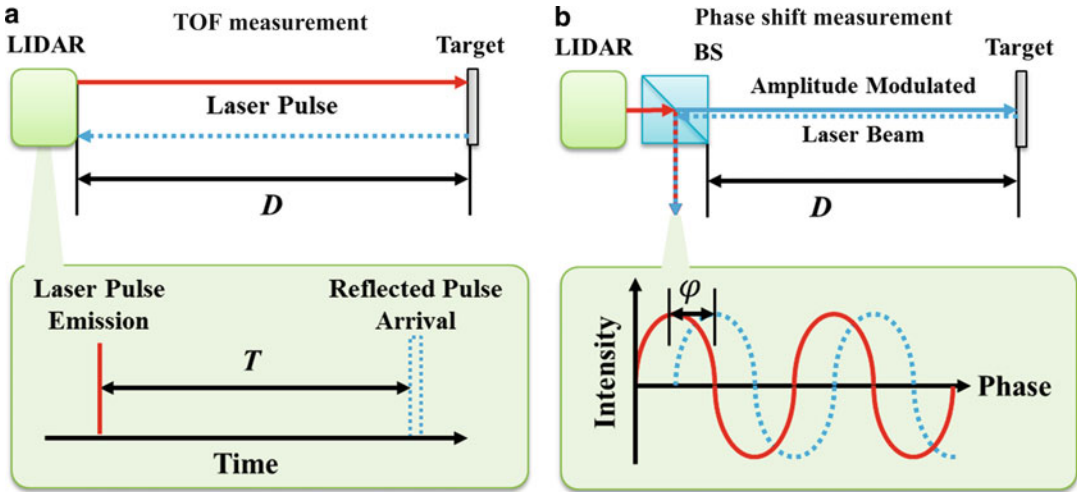
100 to 500 m for common LiDARs. The biggest drawback of the phase-shift-based displacement measurement is that its ranging capability is usually limited to several hundred meters, as $0 \leq D \leq \frac{\lambda_m}{2}$ in Eq. 3. However, its measurement speed can be up to 1 million points/s, which is much faster than the TOF-based measurement technique (usually $\sim 10,000$ points/s).

Figure 4 represents a schematic diagram of a common LDV. Assuming that a target is vibrating with a velocity of v , the frequency of the laser beam reflected from the moving target changes from its initial value of f_0 to the following value based on the Doppler effect:

$$f = \left(1 - \frac{v}{c}\right)^2 f_0 \quad (4)$$

where f denotes the frequency of the reflected laser beam. Note that v becomes negative when the target is vibrating toward the LDV. The intensity of the reflected laser beam can be expressed as

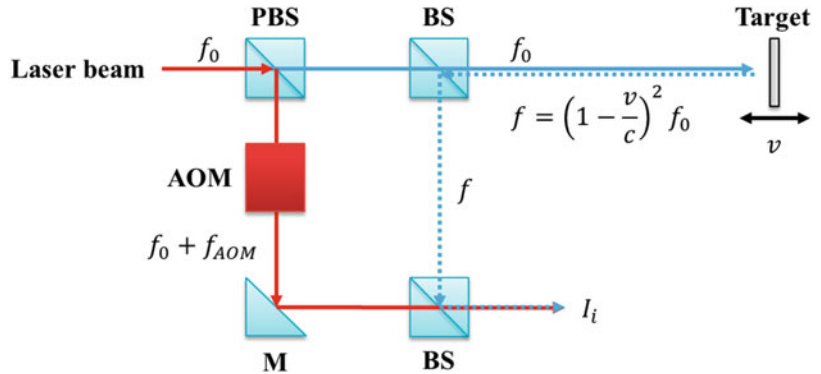
$$I_r = I \cos(2\pi f t) \quad (5)$$



Laser-Based Structural Health Monitoring, Fig. 3 Schematic diagram of LiDAR measurement working principle. Red solid line and blue dotted line denote the paths of the reference and reflected laser beam,

respectively. BS, D, T, and φ refer to a beam splitter, distance to the target, time of flight of the reflected laser pulse, and phase difference between two laser beams, respectively

Laser-Based Structural Health Monitoring, Fig. 4 Schematic diagram of a common LDV. Red solid and blue dotted lines denote the paths of the reference and reflected laser beams, respectively. PBS, BS, AOM, and M refer to a partial beam splitter, a beam splitter, an acousto-optic modulator, and a mirror, respectively



where I and t refer to the maximum intensity of the laser beam and time, respectively. This reflected laser beam interferes with the reference laser beam:

$$I_0 = I \cos(2\pi(f_0 + f_{AOM})t) \quad (6)$$

which originally has a frequency of f_0 but shifted to $f_0 + f_{AOM}$ by an acousto-optic modulator (AOM). f_{AOM} is usually set to 40 MHz for many commercial LDVs. Then the intensity of the interfered laser beam is

$$\begin{aligned} I_i &= I \cos(2\pi(f_0 + f_{AOM})t) + I \cos(2\pi f t) \\ &= \frac{I}{2} \cos(2\pi(f_0 + f_{AOM} + f)t) \\ &\quad + \frac{I}{2} \cos(2\pi(f_0 - f + f_{AOM})t) \end{aligned} \quad (7)$$

By applying a low pass filter to I_i ,

$$I_i^{LPF} = \frac{I}{2} \cos(2\pi(f_0 - f + f_{AOM})t) \quad (8)$$

where $f_0 - f$ can be expressed as

$$\begin{aligned} f_0 - f &= f_0 - \left(1 - \frac{v}{c}\right)^2 f_0 = f_0 \left(1 - \left(1 - \frac{2v}{c} + \frac{v^2}{c^2}\right)\right) \\ &= f_0 \left(\frac{2v}{c} + \frac{v^2}{c^2}\right) \end{aligned} \quad (9)$$

Since v is very small compared to c , we have

$$f_0 - f \simeq f_0 \left(\frac{2v}{c}\right) = \frac{2v}{\lambda_0} \quad (10)$$

$$I_i^{\text{LPF}} = \frac{I}{2} \cos \left(2\pi \left(\frac{2v}{\lambda_0} + f_{\text{AOM}} \right) t \right) \quad (11)$$

where λ_0 is the wavelength of the incident laser beam. Then, by measuring the frequency of the interfered laser beam, the velocity of the target v can be estimated. Note that f_{AOM} is introduced to distinguish positive and negative frequency shift from the interfered laser beam frequency. Because LDV measures velocity from the Doppler frequency shift of the reflected laser beam, the signal-to-noise ratio of LDV heavily depends on the reflectivity of the target surface. Therefore, special surface treatments such as a retroreflective tape are often required for the improved signal-to-noise ratio, but it can be rather challenging to treat surfaces of target structures in the field.

Airborne Laser Scanning for Regional Monitoring

Airborne laser scanning (ALS) techniques utilize a laser device mounted onto aircraft to monitor earthquake-affected areas. Aerial techniques with fast and extensive data acquisition availability, such as aerial and satellite imaging, have been extensively used for large area scanning. The ALS techniques have a big advantage over the 2D imaging techniques because they can gather and reconstruct height/depth information for 3D model construction (Fig. 5). Two ALS schemes are introduced in this section. In the first scheme, ALS is used to detect changes in urban features

and discern which structures are collapsed or damaged. Then, the structural damage type is identified and classified through the second ALS scheme.

Change Detection from Urban Features

Once a 3D model of the region is constructed by ALS scanning, the regional changes can be detected by comparing models before and after any event occurrence. This scheme has been applied for various purposes including forest observation and coastline monitoring. This ALS has also been applied to extracting urban features including land uses and structural identification (Priestnall et al. 2000).

Vu et al. (2004) scanned over Roppongi, Tokyo, Japan in 1999 and 2004. They compared two models and calculated the height difference at each 2D grid. Via the use of statistical analysis and filtering processes, they were able to identify outliers, which have distinctive height changes compared to the average height change of the whole scanned region. These outlier points are considered possible new constructions (height increased) and demolitions (height decreased). This concept can be applied to rapid inspection of earthquake-affected regions, but requires continuous update of the urban ALS models for comparison.

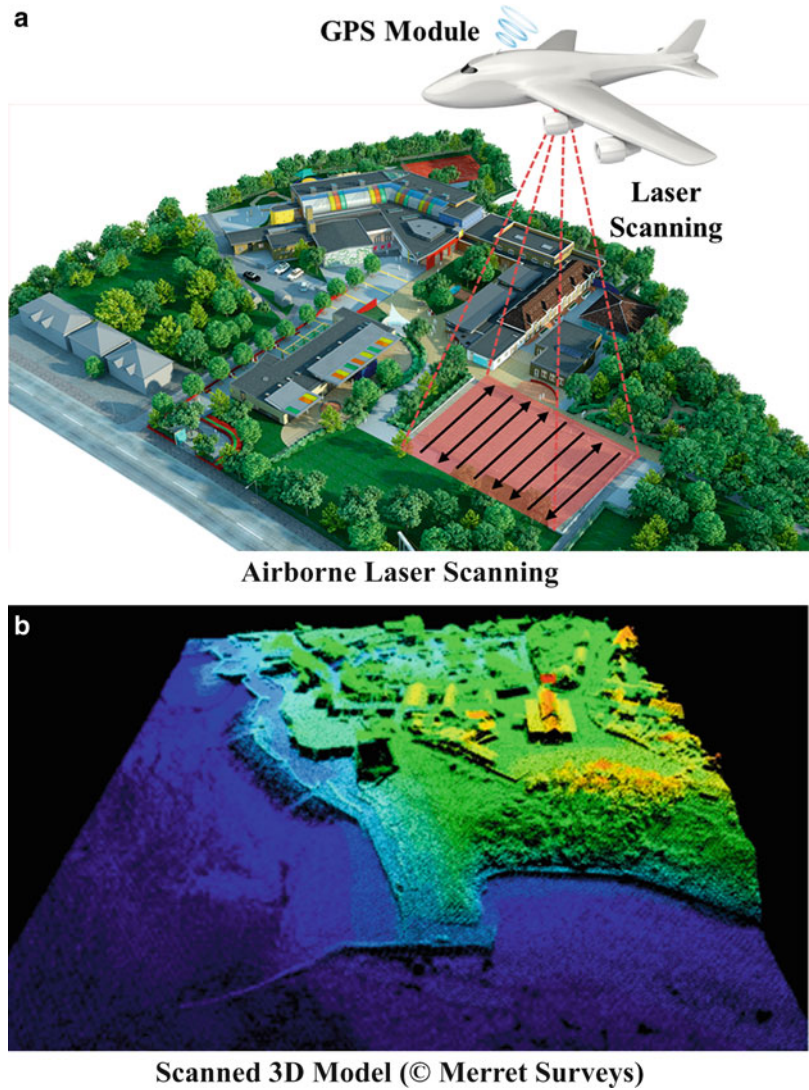
Shen et al. (2010) proposed another monitoring technique to identify demolished buildings by measuring buildings' inclinations from ALS data. They extracted a vertical axis vector of a building from its roof geometry and compared it with the ground normal vector to determine the buildings' inclination. As these two vectors can be gathered from a single ALS image, there is no requirement for previous or baseline model scanned for this technique. They applied this technique to a building affected by the Haiti earthquake and identified its inclination angle.

Classification of Structural Damage

During the rescue operation for earthquake victims, rapid assessment and classification of structural damage types is very critical to set the rescue strategy, allocate recourse, and save as many lives as possible. Schweier and Markus (2006) classified types of building damage after

Laser-Based Structural Health Monitoring,

Fig. 5 (a) Conceptual figure for the airborne laser scanning (ALS) technique and (b) an example of scanned model

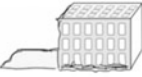


earthquakes into 10 categories including inclination, pancake collapse, and heap of debris (Fig. 6). However, it turned out to be vexing and challenging to classify them using 2D imaging data as it does not contain height information. For example, a pancake-collapsed building cannot be easily recognized with 2D imaging techniques as its identical top-view image remains intact.

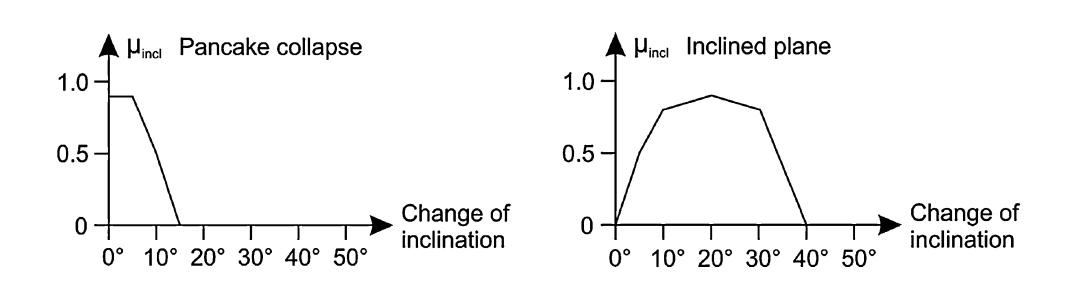
Rehor (2007) scanned a military training area in Swiss and compared the model with its original CAD blueprint to classify structural damage types of each building. Changes in building features including area size, inclination angle,

volume, and height are identified from the comparison.

The damage classification process is automated by introducing fuzzy logic membership functions for each feature. Fuzzy logic is one of the soft computing techniques, which enables the use of probability for logical analysis. For damage classification, the probabilities of all damage categories associated with each specific feature are computed using Fuzzy logic membership functions. For example, the membership functions of the feature “change of inclination” are shown for “the pancake collapse” and “inclined

 1. Inclined plane	 2. Multi layer collapse	 3. Outspread multi layer collapse	 4 a) Pancake collapse, first floor	 4b) Pancake collapse, intermediate story	 4c) Pancake collapse, upper story
 5. Pancake collapse, all stories	 5a) Pancake collapse, several lower stories	 5b) Pancake collapse, intermediate stories	 5c) Pancake collapse, upper stories	 6. Heap of debris on uncollapsed stories	 7a) Heap of debris
 7b) Heap of debris with planes	 7c) Heap of debris with vertical elements	 8. Overturn collapse, separated elements	 9a) Inclination	 9b) Overturn collapse	 10. Overhanging elements

Laser-Based Structural Health Monitoring, Fig. 6 Building damage types after an earthquake (Rehor 2007)



Laser-Based Structural Health Monitoring, Fig. 7 Fuzzy logic membership function example for pancake collapse and inclined plane (Rehor 2007)

plane” damage categories in Fig. 7. If a building had an inclination angle of 5°, this building would have a 0.9 probability for the pancake collapse and 0.25 for the inclined plane. On the other hand, if the inclination angle of another building were 20°, the probabilities for the pancake collapse and the inclined plane would be 0 and 0.9, respectively. The probabilities for other feature values are estimated in a similar manner, and the damage category with maximum probability is classified as the most probable damage category. As a result, damage types of 13 out of 17 (77 %) buildings are correctly classified. Rehor and Voegtle (2008) improved the classification accuracy to

88 % by integrating multispectral scanner data and aerial images with the above ALS model.

Ground-Based Laser Scanning Techniques

While ALS techniques are able to inspect large areas with a high speed, its damage detectability is rather limited to the detection and classification of large-scale defects. Further investigation of the structures, which are classified as intact with the previous ALS techniques, is often required as there still is a high possibility of internal damage

Laser-Based Structural Health Monitoring,

Fig. 8 (a) Scanning of the Ashdown House in UK using a ground-based laser scanner and (b) the resultant 3D model (© Halcrow Group)



Ground-based Laser Scanning



Scanned 3D Model

in these structures. In particular, this detailed inspection could be critical for post-earthquake re-occupation of buildings as the undetected internal damage may lead to the collapse of the structures and additional casualties. A number of contact-type sensors have been suggested for the detailed inspections, and recently LiDAR and LDV are gaining prominence due to their noncontact nature. In this section, ground-based laser scanning techniques for global and local damage diagnosis are presented (Fig. 8).

Serviceability Monitoring Through Deflection/Deformation Measurement

Serviceability refers to the conditions under which a building is still considered useful.

The serviceability is typically evaluated by measuring a local deflection and deformation, and the structure is considered in a proper serviceable condition when the deflection remains within a certain limit. However, deflection measurement on large-scale structure has always been a daunting task.

This problem can be readily solved with a LiDAR. As previously mentioned, LiDAR can measure 3D displacement of a structure as well as its static and dynamic deformed shapes. One of the biggest problems is that the maximum displacement measurement error of LiDAR is up to 10 mm, which may be insufficient for SHM purposes. Park et al. (2007) overcame this problem by introducing a displacement measurement

model to calibrate the error and reduced the measurement error to less than 1 mm compared to the one measured by contact-type sensors such as LVDT, strain gauge, and fiber-optic sensor.

LiDAR has been applied not only in laboratory settings but also in the field. Watson et al. (2012) scanned a highway bridge for underclearance measurement with traffic and temperature variations. They performed correlation analysis between the measured underclearance, traffic, and temperature variations and concluded that LiDAR can consistently measure the underclearance value without environmental dependence. Liu et al. (2010) carried out a static loading test on a steel girder bridge while measuring its deformation with a LiDAR and reported its behavior under truck loading.

Vibration-Based Damage Detection

Vibration-based damage detection is one of the most widely adopted SHM methods for field applications. This technique attempts to evaluate the global integrity of a structure by relating the change of modal parameters such as natural frequencies, mode shapes, and damping to the physical properties of the structure such as stiffness and mass. Conventionally, these modal parameters are measured by contact-type accelerometers, but they can be substituted with a noncontact laser scanner. Nassif et al. (2005) reported that the bridge girder vibration and displacement measured by an LDV match very well with the ones obtained by geophone sensors and LVDTs attached to the girder.

When it comes to vibration measurements, LDV is a preferred noncontact solution rather than LiDAR because modal analysis requires measurement of vibration signals with a high sampling rate. LDV can achieve a sampling rate up to several MHz, while it is usually limited up to 1 kHz in LiDAR. Siringoringo and Fujino (2009) detected damage in a structural member by performing modal analysis using an LDV. They scanned a bolted joint plate and compared modal parameters measured from different bolt-loosening conditions. Stiffness and damping

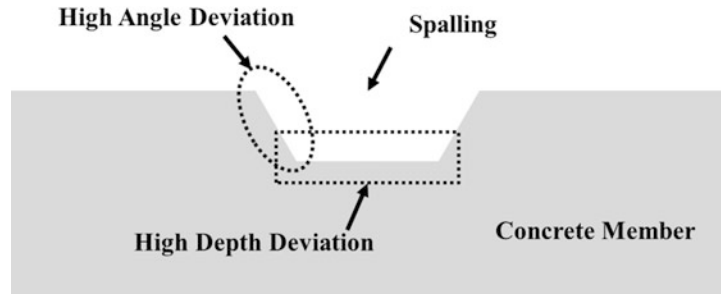
changes due to the loosened bolt torques were observed, and damage size was quantified. Khan et al. (2000) detected various cracks with an LDV, including through cracks in thin plates, narrow slots in a solid cantilever beam, and cracks in a reinforced concrete beam by comparing mode shapes and frequency response functions of the specimens before/after introducing damages.

Laser Model-Based Damage Quantification

The scope of the previously overviewed vibration-based damage detection is typically limited to damage identification and localization. For damage quantification, additional steps are warranted. The principle of the damage quantification techniques presented here is quite similar to the aforementioned ALS techniques. One of the quantification algorithms is volumetric change analysis (Olsen et al. 2010). By comparing the original and the posttest 3D models obtained from laser scanning, volumetric changes, or defects of the concrete structure, can be detected. The scanning process in Olsen et al.'s study took about 6 h to scan a concrete structure of 6.706 m³, while a conventional photographic technique required 20 h of manual photo-taking. Liu et al. (2011) suggested two other damage quantification methods, which do not require a baseline laser-scanned model. In these methods, damage is identified by measuring surface flatness and smoothness. When a local defect like spalling appears on a concrete surface, the depth and normal direction of the defect area will deviate from a reference (undamaged) plane. The depth and angle deviation of the damaged concrete surface from the intact surface are computed using the full 3D coordinate information of the target concrete structure obtained by LiDAR. While the depth information is good for detecting the deepest portion of the damage area, the angle variation is more sensitive to the edge of the defect, where a sudden gradient change of the concrete surface is expected (Fig. 9). They inspected a real concrete bridge by integrating both methods and identified mass loss in concrete members (Fig. 10).

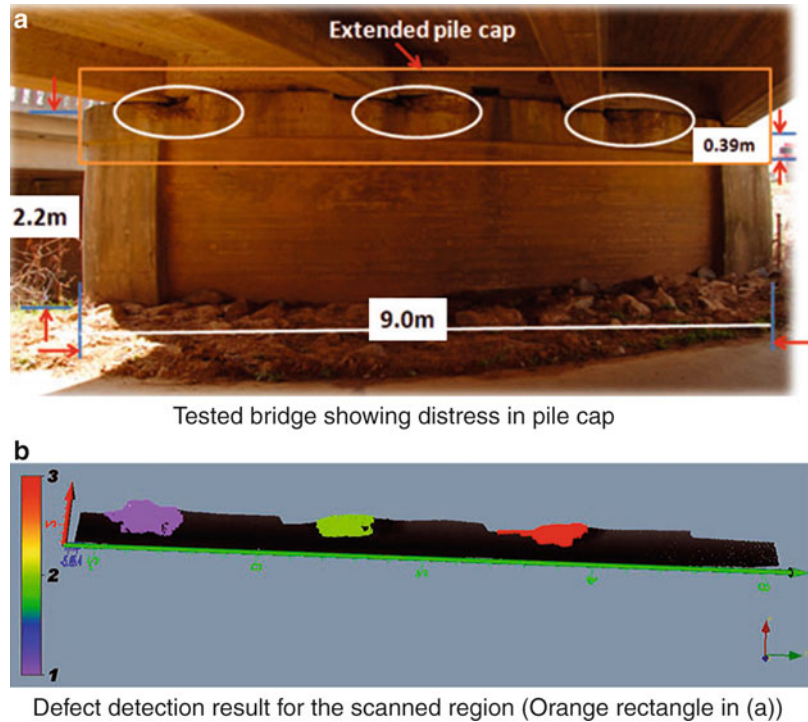
Laser-Based Structural Health Monitoring,

Fig. 9 Schematic diagram for surface damage quantification by measuring the depth and angle deviation



Laser-Based Structural Health Monitoring,

Fig. 10 Defect detection for a concrete bridge member using LiDAR scanning, presented by Liu et al. (2011)



Other Laser-Based Techniques for Local Damage Detection

After earthquakes, small defects may not pose an immediate threat to the overall structural integrity or cause a sudden collapse of the structure. However, hidden and untreated defects will continuously grow and eventually compromise the safety of the system. Therefore, there is an increasing demand to further improve damage detectability and identify defects as early as possible. Furthermore, as sensing techniques advances, the detection of small-scale defects, which was not feasible using conventional

techniques, is becoming feasible. Here, two specific state-of-the-art techniques for such micro-scale damage detection, laser ultrasonic imaging and laser thermography, are presented. Although they have been mostly applied for inspection of mechanical systems so far, their brief introduction is included here because of their great potential for SHM of earthquake-affected structures (Fig. 11).

Laser Ultrasonic Imaging

Ultrasonic techniques identify a defect by measuring interaction between ultrasonic waves and the defect. Conventionally ultrasonic waves are

Laser-Based Structural Health Monitoring,

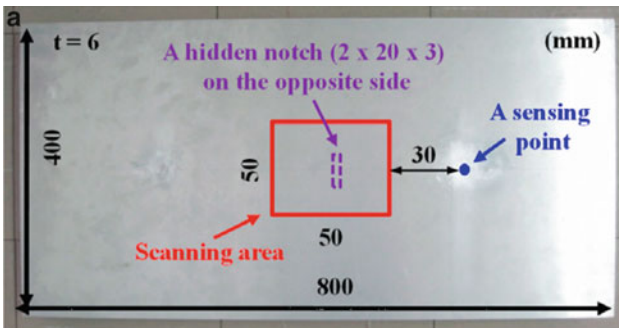
Fig. 11 Devices for laser-based SHM techniques for small-scale damage detection: (a) Nd:YAG pulse laser and (b) infrared camera



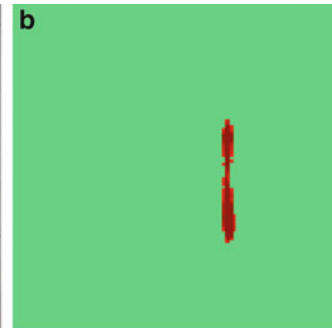
Quantel Brilliant Nd:YAG Laser
(© Quantel Laser)



VarioCAM 700 IR Camera
(© InfraTec)



Tested aluminum specimen with a hidden notch



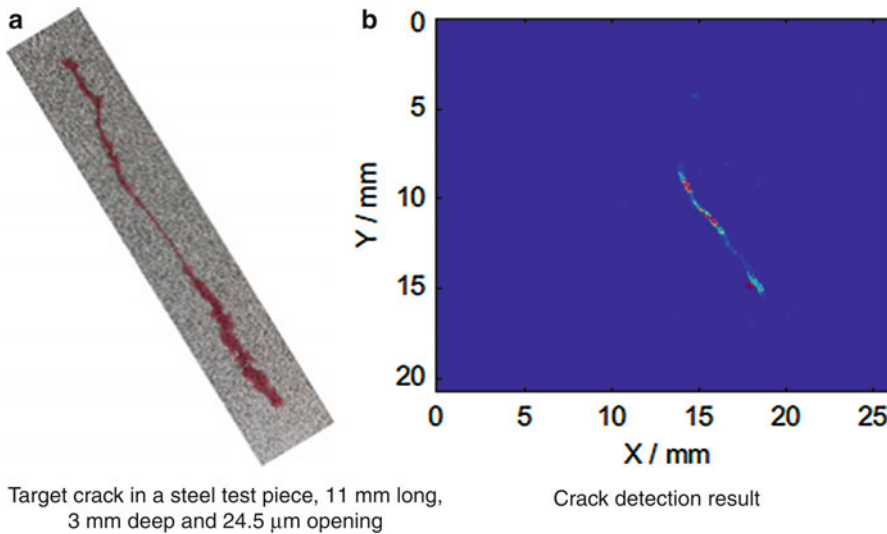
Damage visualization result

Laser-Based Structural Health Monitoring, Fig. 12 Hidden notch detection in an aluminum plate using laser ultrasonic imaging, presented by An et al. (2013). The notch location and its size are visualized with red color in (b)

generated and measured using contact-type transducers such as piezoelectric transducers (Kim and Sohn 2007), but nowadays noncontact laser solutions are available. When a high-power pulse laser beam is exerted onto a solid surface, a localized heating and corresponding elastic expansion of the specimen produce ultrasonic waves (Scruby and Drain 1990). The generated ultrasonic waves can be measured using an LDV. With its scanning capability, ultrasonic waves generated from a single point can be measured from multiple points with high spatial resolution in the order of mm. Once the ultrasonic responses are measured from multiple scan points, the wave propagation in a spatial domain and its interaction with a defect can be visualized. An et al. (2012, 2013) apply laser ultrasonic imaging techniques to the detection of a notch in a steel bridge girder and a notch in a metal plate (Fig. 12). In their study, a high-power pulse laser

beam is scanned over the inspection region for ultrasonic wave generation, and the corresponding responses are measured at a single point using an LDV. In theory, the ultrasonic image obtained by scanning the excitation laser beam with a fixed sensing point is identical to the one obtained by scanning the sensing laser beam with a fixed excitation point based on linear reciprocity. In practice, scanning the excitation laser beam is more advantageous as the performance of LDV usually varies with the surface condition and the incident angle of the sensing laser beam.

However, there are several hurdles that need to be overcome for field applications. The laser ultrasonic generation requires the use of a class IV high-power laser over 1 MW, which may harm human eyes. The scanning time for large area inspection with high spatial resolution can be prohibitively long. Currently laser ultrasonic techniques are best suited for inspection of



Laser-Based Structural Health Monitoring, Fig. 13 Crack detection in a steel test specimen using laser thermography, presented by Li et al. (2011)

metallic and composite materials, and the applicability to porous materials including concrete is limited due to rough surface conditions.

Laser Thermography

Infrared (IR) thermography techniques visualize and quantify damage by detecting changes in heat transfer characteristics near the defect. As thermal images can be taken over a large area rather quickly using the IR camera, this technique has been popular for nondestructive testing of mechanical systems. Clark et al. (2003) detected delamination in a concrete slab of a real bridge, by using natural sunlight as a passive heat source and taking thermal images. By detecting regions with abnormal temperature in comparison with the rest of the areas, possible damage locations were recognized. However, the maximum temperature differences remained below 1 °C, making it challenging to distinguish damage regions from the rest.

This leads to the development of active thermography techniques, where a controlled heat source such as a halogen lamp and acoustic transducer is used (Chatterjee et al. 2011). A laser thermography technique is relatively new in the field of active thermography and uses a high-power laser as a heat source. Compared to a

traditional heat source such as a halogen lamp, the biggest advantage of the laser is its highly localized nature of heating spot. This helps to improve the damage detection resolution and defect micro-cracks. Li et al. (2011) visualized a 50 μm wide and 5 mm long fatigue crack in a mild steel test piece (Fig. 13). However, it should be noted that the application of the laser thermography technique is usually limited to the detection of surface damages in thin structures, and it requires history baseline data collected from the intact condition to identify damage, which may show false alarms in the presence of operational and environmental variations.

Summary

Laser-based structural health monitoring is a promising technology for the monitoring of earthquake-affected structures as it does not require any sensor nor power/data cable installation. Light detection and ranging (LiDAR) and laser Doppler vibrometer (LDV) are two of the most widely adopted laser-sensing devices for civil applications due to their noncontact nature and high measurement accuracy. LiDAR estimates the distance to the target by measuring

time-of-flight or phase-shift information of the laser beam reflected off the target, and LDV measures the vibration velocity of the target by relating the frequency shift of the reflected laser beam.

Various laser-based structural health monitoring techniques applicable to earthquake-affected structures have been proposed and briefly discussed in this entry. By means of airborne laser scanning (ALS), earthquake-affected regions may be rapidly monitored for distinguishing damaged structures and further identifying their damage types. Even for intact structures, LiDAR-based displacement/deformation measurement can be used to check the serviceability of civil infrastructure. The damage location and size can be quantified with LDV-based vibration analysis and LiDAR-based 3D structural model. Additional laser techniques with potential application to post-earthquake inspection include laser ultrasonic techniques and laser thermography techniques, which are able to detect small-scale damages.

Still there are several challenges associated with field deployment of these laser devices: (1) the time-of-flight-based LiDAR cannot collect data with a high sampling rate; (2) the applicability of the phase-shift-based LiDAR is limited to a short range below several hundred meters; (3) special surface treatment is often necessary for LDV; and (4) the use of high-power pulse laser may have implications associated with eye safety. However, the developments and improvements of not only hardware devices but also damage detection algorithms will inevitably contribute to the fast and accurate health assessment of earthquake-affected structures and prevent further catastrophic failures.

Cross-References

- [Post-Earthquake Diagnosis of Partially Instrumented Building Structures](#)
- [Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring](#)
- [System and Damage Identification of Civil Structures](#)

References

- An YK, Song HM, Park HJ, Sohn H, Yun CB (2012) Remote guided wave imaging using wireless PZT excitation and laser vibrometer scanning for local bridge monitoring. In: *Bridge Maintenance, Safety, Management, Resilience and Sustainability: Proceedings of the Sixth International IABMAS Conference*: 173, <http://koasas.kaist.ac.kr/handle/10203/171481>
- An YK, Park B, Sohn H (2013) Complete noncontact laser ultrasonic imaging for automated crack visualization in a plate. *Smart Mater Struct* 22(2):025022
- Castellini P, Maratarelli M, Tomasini EP (2006) Laser Doppler Vibrometry: development of advanced solutions answering to technology's needs. *Mech Syst Signal Process* 20:1265–1285
- Chang PC, Flatau A, Liu SC (2003) Review paper: health monitoring of civil infrastructure. *Struct Health Monit Int J* 2(3):257–267
- Chatterjee K, Tuli S, Pickering SG, Almond DP (2011) A comparison of the pulsed, lock-in and frequency modulated thermography nondestructive evaluation techniques. *NDT & E Int* 44(7):655–667
- Clark MR, McCann DM, Forde MC (2003) Application of infrared thermography to the non-destructive testing of concrete and masonry bridges. *NDT&E International* 36(4):265–275
- Khan AZ, Stanbridge AB, Ewins DJ (2000) Detecting damage in vibrating structures with a scanning LDV. *Opt Lasers Eng* 32(6):583–592
- Kim SB, Sohn H (2007) Instantaneous reference-free crack detection based on polarization characteristics of piezoelectric materials. *Smart Mater Struct* 16(6):2375
- Li T, Almond DP, Rees DAS (2011) Crack imaging by scanning pulsed laser spot thermography. *NDT & E Int* 44(2):216–225
- Liu W, Chen S, Boyajian D, Hauser E (2010) Application of 3D LiDAR scan of bridge under static loading testing. *Mater Eval* 68(12):1359–1367
- Liu W, Chen S, Hauser E (2011) LiDAR-based bridge structure defect detection. *Exp Tech* 35(6):27–34
- Nassif HH, Gindy M, Davis J (2005) Comparison of laser Doppler vibrometer with contact sensors for monitoring bridge deflection and vibration. *NDT & E Int* 38(3):213–218
- Olsen MJ, Kuester F, Chang BJ, Hutchinson TC (2010) Terrestrial laser scanning-based structural damage assessment. *J Comput Civ Eng* 24(3):264–272
- Park HS, Lee HM, Adeli H, Lee I (2007) A new approach for health monitoring of structures: terrestrial laser scanning. *Comput Aided Civ Infrastruct Eng* 22:19–30
- Priestnall G, Jaafar J, Duncan A (2000) Extracting urban features from LiDAR digital surface models. *Comput Environ Urban Syst* 24(2):65–78
- Rehor M (2007) Classification of building damage based on laser scanning data. *Photogramm J Finl* 20(2):54–63

- Rehor M, Voegtle T (2008) Improvement of building damage detection and classification based on laser scanning data by integrating spectral information. *Int Arch Photogramm Remote Sens Spat Inf Sci XXXVII (Part B7)*. Beijing, pp 1599–1606
- Schweier C, Markus M (2006) Classification of collapsed buildings for fast damage and loss assessment. *Bull Earthq Eng* 4(2):177–192
- Scrubby CB, Drain LE (1990) *Laser ultrasonics: techniques and applications*. Adam Hilgher, Norfolk
- Shen Y, Lixin W, Zhi W (2010) Identification of inclined buildings from aerial LIDAR data for disaster management. In: *Proceedings of 18th international conference on geoinformatics*, pp 1–5, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5567852&tag=1
- Siringoringo DM, Fujino Y (2009) Noncontact operational modal analysis of structural members by laser Doppler vibrometer. *Comput Aided Civ Infrastruct Eng* 24:249–265
- Vu TT, Matsuoka M, Yamazaki F (2004) LIDAR-based change detection of buildings in dense urban areas. In: *Proceedings of IEEE international geoscience and remote sensing symposium*, vol 5, pp 3413–3416, http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1370438
- Watson C, Chen S, Bian H, Hauser E (2012) Three-dimensional terrestrial LIDAR for operational bridge clearance measurements. *J Perform Constr Facil* 26(6):803–811
- Wehr A, Lohr U (1999) Airborne laser scanning—an introduction and overview. *ISPRS J Photogramm Remote Sens* 54:68–82

(nonlinear) response in selected components of the structural frame. Such response is associated with structural damage that produces direct (capital) loss repair cost, indirect loss (possible closure, rerouting, business interruption), and perhaps casualties (injuries, loss of life) (Constantinou et al. 2007b).

There are various mechanisms/devices that have been invented in the past for the protection of structures from the damaging effects of earthquakes. A lot of these devices are based on the idea of uncoupling the structures from the ground. Rollers, layers of sand, or similar materials are among the ideas proposed in the past to allow a building to slide. Such mechanisms are early versions of the now widely accepted and applied concept of base/seismic isolation (Kelly 2004).

Lead-rubber bearings are one of various existing types of seismic isolation devices and have been used for years for the protection of structures from the damaging effect of earthquakes. This entry presents summarized information regarding the use, benefits, properties, and applications of lead-rubber bearings along with literature sources for further reading on the subject.

Lead-Rubber Bearings with Emphasis on Their Implementation to Structural Design

Ioannis Kalpakidis
KBR, Houston, TX, USA

Synonyms

Base isolation; Bridge bearings; Elastomeric bearings; Lead-plug bearings; Lead-rubber bearings; Rubber bearings; Seismic isolation

Introduction

The seismic design of conventionally framed bridges and buildings relies on the dissipation of earthquake-induced energy through inelastic

Seismic Isolation

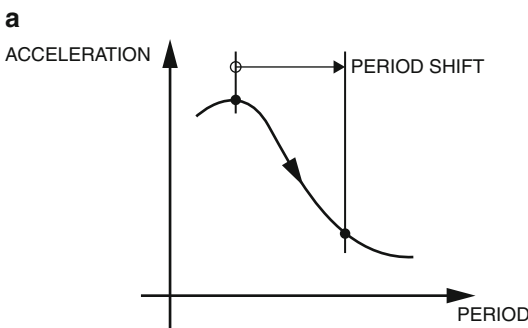
The concept of base isolation is quite simple. The seismic isolation system is a group of low-stiffness elements that is placed between the foundation and the structure and reduces the effects of the earthquake on the latter. Use of these elements causes a decrease of the fundamental frequency of the structure shifting it away from both its fixed-base frequency and the predominant frequencies of the ground motion (Kelly 2004).

Seismic isolators are typically installed between the girders and bent caps (abutments) in bridges and the foundation and first suspended level in a building. Fig. 1 (Constantinou et al. 2007b) shows a typical bridge installation of seismic isolators; the isolators were installed at the top of the pier as shown in Fig. 1b (the steel plate on the side of the isolator is part of the lateral wind-restraint system). For bridge

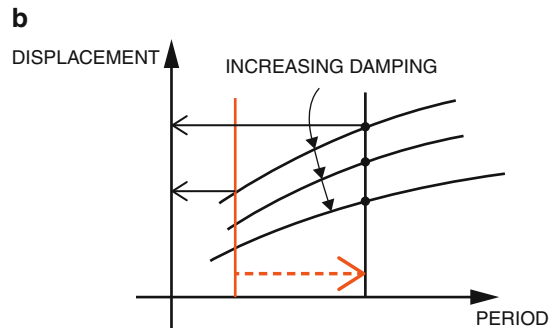


Lead-Rubber Bearings with Emphasis on Their Implementation to Structural Design, Fig. 1 Typical installation of a seismic isolator in a

bridge (Constantinou et al. 2007b). (a) Bridge elevation. (b) Friction pendulum seismic isolator



Lead-Rubber Bearings with Emphasis on Their Implementation to Structural Design, Fig. 2 Concept of seismic isolation (Constantinou



et al. 2007b). (a) Decrease in spectral accelerations. (b) Increase in spectral displacements

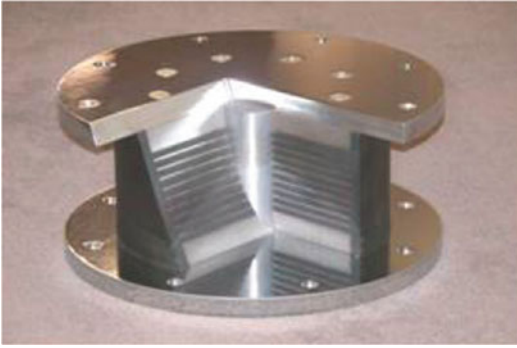
construction, the typical design goals associated with the use of seismic isolation are (a) reduction of forces (accelerations) in the superstructure and substructure and (b) force redistribution between the piers and the abutments (Constantinou et al. 2007b). Reduction of forces is achieved through the aforementioned structural frequency shift as shown in Fig. 2a (Constantinou et al. 2007b). Control of the increase in displacements that results from the use of seismic isolation is achieved through introducing damping (energy dissipation) in the isolator as shown in Fig. 2b (Constantinou et al. 2007b).

There are various types of seismic isolation devices, including but not limited to: (a) elastomeric-based systems, (b) low-damping natural and synthetic rubber bearings, (c) high-damping natural rubber (HDNR) systems,

(d) isolation systems based on sliding, (e) friction pendulum system, (f) double and triple concave friction pendulum sliding bearings, (g) spring-type systems, and (h) lead-rubber bearings (Kelly 2004, Constantinou et al. 2007b). Elastomeric bearings (low- and high-damping rubber, lead rubber) and sliding bearings (spherical sliding or friction pendulum bearings, flat sliding or EradiQuake bearings) are the two common types of seismic isolation bearings used in the United States (Constantinou et al. 2007b).

Lead-Rubber Bearings: Short Description and Historical Background

The lead-rubber bearing (LRB) was invented in 1975 by W. H. Robinson in New Zealand (Skinner et al. 1993; Kelly 2004; Robinson and Tucker 1977, 1983). It has been extensively used



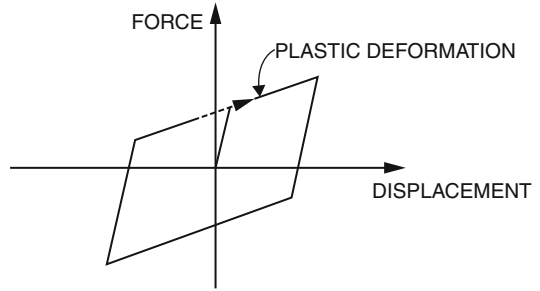
Lead-Rubber Bearings with Emphasis on Their Implementation to Structural Design, Fig. 3 Lead-rubber bearing construction details (Constantinou et al. 2007b)

in New Zealand, Japan, and the United States (Kelly 2004). Such bearings have been used to isolate many buildings, and buildings using them performed well during the 1994 Northridge and 1995 Kobe earthquakes (Kelly 2004). A short summary of the response of one of these buildings during the 1994 Northridge earthquake is provided in the next section.

Generally, a lead-rubber bearing is composed of a low-damping (unfilled) elastomer and one (typically) or more (less common) lead cores. A cutaway view of a lead-rubber bearing is illustrated in Fig. 3 (Constantinou et al. 2007b). In this illustration, the separate components of the bearing can be seen clearly: (a) the lead plug in the core of the bearing, (b) the end and flange plates on top and bottom of the bearing, and (c) the low-damping elastomer with alternating rubber layers and steel shim plates.

The elastomer rubber layers serve as the “isolation” (i.e., stiffness and fundamental structural frequency reduction) component; lead also contributes to the stiffness of the bearing, but this contribution is relatively minor. The steel shim plates assist with vertical load support. The lead core mainly serves as the “damping” (i.e., energy dissipation) component. A typical hysteretic (force-displacement) loop for a lead-rubber bearing is shown in Fig. 4 (Constantinou et al. 2007b).

The lead core provides excellent energy dissipation capability under confined conditions. Such conditions can be achieved by constructing the



Lead-Rubber Bearings with Emphasis on Their Implementation to Structural Design, Fig. 4 Hysteretic energy dissipation in lead-rubber bearings (Constantinou et al. 2007b)

plug slightly (less than 5 %) longer than the height of the rubber bearing and connecting the top and bottom flange plates to the rubber bearing end plates through countersunk bolts. Then the plug is compressed upon bolting the flange plates to the end plates; it expands laterally and wedges into the rubber layers between the shim plates. The magnitude of energy dissipation capacity depends on the lead plug diameter (Constantinou et al. 2007b).

Example: Performance of a LRB-Isolated Building During the 1994 Northridge Earthquake

As an example of the benefits from the use of lead-rubber bearings, a brief summary of the behavior of the LRB-isolated University of Southern California Teaching Hospital during the 1994 Northridge earthquake is provided in this section. A very detailed description has been outlined by Kelly (2004). Design and construction details of this building have been described by Asher and Van Volkinburg (1989), Asher et al. (1990), and KPFF Consulting Engineers (1991).

The hospital is an eight-story concentrically braced steel frame with significant irregularities in its plan and elevation. The isolation system consists of 68 lead-rubber isolators and 81 elastomeric isolators. The lead-rubber bearings are distributed in the perimeter of the building under the braced frames to minimize torsional response in the isolated structure. Elastomeric bearings are used in the interior.

The fixed-base period of the superstructure is approximately 0.7 s. The design base shear of the hospital is 0.15 g, while the design corner bearing displacement including torsion is 26.0 cm. The effective period of the isolation system at this displacement is approximately 2.3 s.

The strongest motions recorded at the site were in the north–south direction. The pseudo-acceleration response spectrum of the north–south foundation motion (Clark et al. 1996) shows significant short-period energy in the range of 0.2–0.6 s, but relatively little energy above a period of approximately 1.3 s. At periods greater than 2 s, the corresponding 10 % damped spectral displacement is less than 3 cm.

The isolation frequency due to the north–south acceleration was approximately 0.7 Hz and the damping ratio approximately 14 %. In the east–west direction, the isolation frequency was 0.75 Hz and the damping ratio was 13 %. The isolation system performed well, its damping capacity was good, and the structure was effectively isolated from the earthquake ground motions. It is noted that the Northridge earthquake caused significant damage to other buildings in the medical center (Kelly 2004).

Entry Roadmap

Section [Introduction](#) of this entry has presented a brief introduction, description, historical background, and an application of lead-rubber isolators. The following sections of this entry are outlined below.

Section [Lead-Rubber Bearing Properties and Structural Design](#) provides a more detailed description of the properties of lead-rubber

bearings along with their significance in structural design.

Section [Applications](#) provides a few more examples of lead-rubber bearing applications in buildings, bridges, and structure retrofit.

Section [Summary](#) is a summary of this entry.

Section [Cross-References](#) provides a list of related entries from this Encyclopedia.

Section [References](#) provides a list of references used in this entry.

Lead-Rubber Bearing Properties and Structural Design

A brief description of lead-rubber bearings and their components has been provided in section [Introduction](#). This section provides further details regarding the properties of lead-rubber bearings.

An idealized force-displacement loop for a lead-rubber bearing is shown in Fig. 5 (Constantinou et al. 2007b) with its significant parameters/mechanical properties being the characteristic strength, Q_d , and the post-elastic stiffness, K_d , defined as follows:

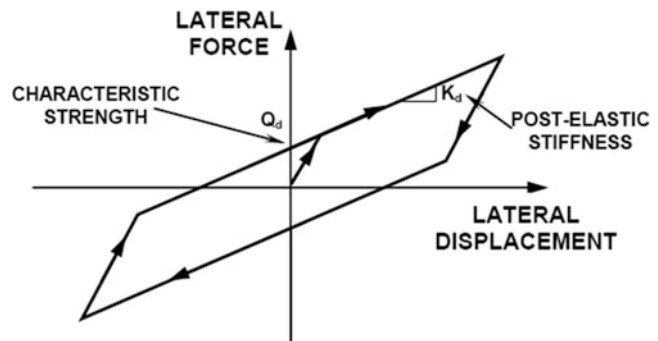
$$Q_d = A_L \sigma_L \quad (1)$$

$$K_d = \frac{f_L G A_r}{T_r} \quad (2)$$

where f_L is a parameter that represents the effect of the lead core on the post-yield stiffness (typically close to unity), A_L is the cross-sectional area of the lead core, T_r is the total thickness of the rubber layers, G is the shear modulus of

Lead-Rubber Bearings with Emphasis on Their Implementation to Structural Design,

Fig. 5 Idealized force-displacement relation of a typical seismic isolation system (Constantinou et al. 2007b)



Lead-Rubber Bearings with Emphasis on Their Implementation to Structural Design, Table 1 Lead-rubber bearing components/properties and environmental/other effects

Property	Controlling component	Component significance
Post-elastic stiffness, K_d	Elastomer (typically low-damping rubber)	Structure stiffness reduction Structure fundamental frequency reduction
Characteristic strength, Q_d	Lead core	Energy dissipation
Environmental or other factor	Component/property	Effect
Aging	Elastomer/post-elastic stiffness, K_d	Minor increase
	Lead/characteristic strength, Q_d	Insignificant (lead yield strength not affected by aging); confinement changes may affect strength slightly
Temperature	Elastomer/post-elastic stiffness, K_d	Increase for low temperatures; effect depends on exposure duration; elastomer becomes brittle below glass transition temperature (compound dependent, -55°C for natural rubber)
	Lead/characteristic strength, Q_d	Increase/decrease for temperature decrease/increase (the same happens to lead yield strength)
Heating	Elastomer/post-elastic stiffness, K_d	Insignificant
	Lead/characteristic strength, Q_d	Strength drops significantly, especially during the first few cycles and for high-frequency and large-amplitude motion due to lead core temperature increase; strength recovers its original value after lead core is allowed to cool down
Cumulative travel	Elastomer/post-elastic stiffness, K_d	None
	Lead/characteristic strength, Q_d	Low speed strength increase after high-amplitude cumulative travel; none in all other cases

rubber, A_r is the bonded rubber area (i.e., excluding the outer rubber cover beyond the steel shims), and σ_L is the effective yield stress of lead (Constantinou et al. 2007b).

The elastomer has a shear modulus between 65 and 100 psi at 100 % rubber shear strain. The lead core has a diameter ranging between 15 % and 33 % of the bearing bonded (i.e., distance between the outer edges of the shim plates) diameter. The maximum shear strain range for LR bearings varies depending on the manufacturer but is generally between 125 % and 200 % (Constantinou et al. 2007b). The effective yield stress of lead typically ranges between 10 and 20 MPa (Skinner et al. 1993; Constantinou et al. 2007b).

Numerous tests and studies have been conducted to investigate and determine how these properties are affected by environmental and other factors such as aging, temperature, heating, history of loading, etc., and modification factors for these properties have been proposed

(Constantinou et al. 1999, 2007a, b; Kalpakidis and Constantinou 2009a, b; Kalpakidis et al. 2010). Some of these effects and a summary of lead-rubber bearing properties and components along with the significance of these components are outlined in Table 1.

A theoretical interpretation of the effects of seismic isolation on the dynamic properties and, consequently, on the dynamic behavior of a structure has been performed by Kelly (1990; 2004) through the use of a simplified two-degree-of-freedom (2-DOF) model. This simplified model treats the superstructure as a single-degree-of-freedom (SDOF) system that has appropriate period, stiffness, damping, and mass values. This SDOF system representing the superstructure is on top of a mass (equal to the base mass) which in turn is on top of the isolation system (defined by appropriate stiffness and damping values).

Assuming a very wide separation (order of 10^{-2} to 10^{-1}) between the square of the isolation

frequency (i.e., frequency of the combined isolation-superstructure system when the superstructure is rigid) and the square of the fixed-base frequency of the superstructure, as well as a value with a 10^{-2} to 10^{-1} order of magnitude for both of the respective damping ratios, the resulting two modes for the combined superstructure-isolation system (simplified 2-DOF model) have the following characteristics (Kelly 1990, 2004):

- Isolation frequency/mode 1: The frequency is affected very slightly by the flexibility of the superstructure. The mode shape involves very little structural deformation (the structure is nearly rigid), while the isolation system picks up much more deformation.
- Structural frequency/mode 2: The mode shape involves much more structural deformation than the first mode, but its effect is insignificant due to the very small participation factor associated with this mode.

In short, as mentioned earlier in this entry, the presence of a seismic isolation system (with either lead-rubber bearings or other types of isolators) is beneficial because it increases the fundamental period of the superstructure (i.e., decreases the corresponding spectral acceleration) while at the same time it adds an energy dissipation mechanism to the system.

Applications

This section contains a brief summary of example applications of lead-rubber bearings (a building application has been briefly described in section [Example: Performance of a LRB-Isolated Building during the 1994 Northridge Earthquake](#)). It is noted that each example herein is merely a high-level summary of an application. For further details, the reader is referred to the respective sources provided for each example.

Richmond–San Rafael Bridge (Retrofit)

In general, use of seismic isolation can reduce forces in bridge foundations by up to 70 %. Isolated bridges are superior in performance, and

their cost is lower than that of non-isolated bridges ([DIS, Inc.](#)).

The Richmond–San Rafael Bridge is the northernmost east–west crossing of the San Francisco Bay in California. It connects Richmond (east end) to San Rafael (west end). It opened in 1956, at which time it was one of the longest bridges in the world. Its length is 5.5 miles and it spans two principal ship channels. In the fall of 2001, extensive seismic retrofit work was performed along with necessary age-related maintenance. The retrofit was completed in 2005.

The bridge structure has a characteristic shape shown in Fig. 6 that has been described as “roller coaster span” and “bent coat hanger.” Without the use of seismic isolators, shorter and stiffer piers would attract most of the lateral load. 55-in. diameter lead-rubber isolators were used for the retrofit. The benefit from using isolators was lateral load redistribution throughout the structure ([DIS, Inc.](#)).

San Francisco City Hall (Retrofit)

The City Hall was built in the late 1800s and was destroyed in the 1906 San Francisco earthquake. It was subsequently rebuilt at a nearby location and reopened in 1915. The City Hall was again damaged in the 1989 Loma Prieta earthquake. Subsequently, seismic retrofit work was performed and the building was reopened in 1999.

For the seismic retrofit, the two-block-long building was cut from its foundation and made to “float” on more than 500 lead-rubber isolators, designed to dissipate seismic energy and to allow the building to sway horizontally up to 26 in. without shaking apart ([MCEER](#)). A photo of the building can be seen in Fig. 7.

Oakland City Hall (Retrofit)

Construction of the Oakland City Hall was completed in 1914. A photo of the building can be seen in Fig. 8. The structure was the first high-rise government office building in the United States (Walters 2003). It is an 18-story building with full basement. Its podium contains a central rotunda, council chambers, and administration offices. An office tower is above the podium, and a clock tower is above the office tower ([MCEER](#); Walters 2003). The structure suffered extensive damage in



Lead-Rubber Bearings with Emphasis on Their Implementation to Structural Design, Fig. 6 Richmond-San Rafael Bridge (MTC CA) Photo: William Hall, Caltrans

Lead-Rubber Bearings with Emphasis on Their Implementation to Structural Design, Fig. 7 San Francisco City Hall (Wikimedia-a) Photo by Andreas Praefcke is licensed under CC BY 3.0



the 1989 Loma Prieta earthquake, losing about 20 % of its strength in the north–south direction and 30 % in the east–west direction (Walters 2003).

Subsequently, several studies were performed resulting in seismic isolation retrofit work that was completed in 1995 (Walters 2003). 42 lead-rubber bearings and 69 natural rubber bearings were utilized for the retrofit (MCEER).

The main reason for selecting seismic isolation as the retrofit method (instead of other

conventional strengthening methods) was the fact that the existing structural system of the City Hall (riveted steel frames with infill masonry) could provide sufficient lateral strength and stiffness only until the development of cracks. After cracks develop, the response would degrade under repeated cyclic loading (Benjamin and Williams 1958, Walters 2003). The result would be lack of ductility and of reliable post-cracking strength, which would require

Lead-Rubber Bearings with Emphasis on Their Implementation to Structural Design,

Fig. 8 Oakland City Hall
(Wikimedia-b) Photo by Sanfranman59 is licensed under CC BY-SA 3.0



L

a nearly elastic resistance of maximum earthquake forces (Walters 2003). The most effective way to satisfy this demand would be the use of seismic isolation that would significantly reduce the seismic forces acting on the structure.

Emergency Headquarters at Fukushima Daiichi Nuclear Power Plant (New Construction)

During the July 2007 Niigataken Chuetsu-oki earthquake (M6.8), there was no damage to the safety of nuclear power plant facilities at TEPCO Kashiwazaki-Kariwa Nuclear Power Plant. The reactors were shut down safely. In addition, the Emergency Headquarters located on the first floor of the office building suffered little structural damage. However, minor damage (such as door lock failure, overturning of racks, sanitary system trouble, etc.) disrupted the emergency response (Hijikata et al. 2012; TEPCO 2007).

Subsequently, TEPCO decided to use base isolation in “Emergency Headquarters” buildings at every nuclear power plant site. One such building

was constructed at the Fukushima Daiichi site in March 2010. The isolation system consisted of four lead-rubber bearings at the corners, 10 natural rubber bearings at the perimeter, 31 sliding bearings, and 16 oil dampers (Hijikata et al. 2012).

The building suffered no damage from the Great East Japan earthquake of March 2011 (Hijikata et al. 2012).

Summary

This entry has presented an overview of the behavior, properties, and applications of lead-rubber seismic isolation bearings.

Seismic hazard mitigation resulting from the use of lead-rubber bearings is twofold. First and foremost, these devices lengthen the structural period and thus decrease the seismic demand on the structure. In addition, they absorb a portion of the seismic energy.

Lead-rubber bearings have been used to isolate many buildings, and buildings using them

performed well during strong earthquakes including the 1994 Northridge and 1995 Kobe earthquakes (Kelly 2004) and the 2011 Great East Japan earthquake (Hijikata et al. 2012). Benefits from their use are numerous. Applications include newly constructed buildings and bridges as well as retrofit of existing buildings and bridges. A very significant benefit from the use of lead-rubber bearings for seismic retrofit, other than the (obvious) mitigation of seismic hazard, is the cost-effective preservation and seismic upgrade of historic buildings and other old architecturally and culturally significant structures.

Additionally, benefits from using lead-rubber bearings in high-importance structures are tremendous. Such structures (e.g., bridges, hospitals, government buildings) are expected to resume their normal operation immediately after an extremely strong earthquake. Seismic isolation serves as an additional means to that scope. This benefit is even more evident in high-importance structures the destruction of which would mean not only disruption of crucial service (e.g., transportation, health care, energy production) but also potential health hazard for whole communities (nuclear power plants, chemical plants, etc.).

Cross-References

- [Code-Based Design: Seismic Isolation of Buildings](#)
- [Earthquake Protection of Essential Facilities](#)
- [Friction Dampers for Seismic Protections of Steel Buildings Subjected to Earthquakes: Emphasis on Structural Design](#)
- [Passive Control Techniques for Retrofitting of Existing Structures](#)
- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)

References

- Asher JW, Van Volkinburg DR (1989) Seismic isolation of the USC University Hospital. In: *Proceedings of ASCE structures congress*, vol 1, San Francisco, pp 605–614
- Asher JW et al (1990) Seismic isolation design of the USC University Hospital. In: *Proceedings of 4th U.S. National conference on earthquake engineering, Earthquake Engineering Research Institute*, vol 3, Palm Springs, pp 529–538
- Benjamin JR, Williams HA (1958) The behavior of one-story brick shear walls. *American Society of Civil Engineers, Journal of the Structural Engineering Division, Proceedings of the American Society of Civil Engineers*, Vol. 84, No. ST4, Paper 1723
- Clark PW, Hashino M, Kelly JM (1996) Performance of seismically isolated structures in the January 17, 1994 Northridge Earthquake. In: *Proceedings of sixth U.S.-Japan workshop on the improvement of building structural design and construction practices in the United States and Japan*. Applied Technology Council and Japan Structural Consultants Association, ATC-15–5, Victoria
- Constantinou MC, Tsopelas P, Kasalanati A, Wolff ED (1999) Property modification factors for seismic isolation bearings. Technical report MCEER-99-0012, Multidisciplinary center for earthquake engineering research, University at Buffalo, State University of New York, Department of Civil, Structural and Environmental Engineering, Ketter Hall, Buffalo, NY 14260
- Constantinou MC, Whittaker AS, Fenz DM, Apostolakis G (2007a) Seismic isolation of bridges. Department of Civil, Structural and Environmental Engineering, University at Buffalo, State University of New York, Buffalo, NY 14260
- Constantinou MC, Whittaker AS, Kalpakidis Y, Fenz DM, Warn GP (2007b) Performance of seismic isolation hardware under service and seismic loading. Technical Report MCEER-07-0012, Multidisciplinary Center for Earthquake Engineering Research, University at Buffalo, State University of New York, Department of Civil, Structural and Environmental Engineering, Ketter Hall, Buffalo, NY 14260
- KPFF Consulting Engineers (1991) USC University Hospital, Project summary. Santa Monica
- Dynamic Isolation Systems, Incorporated (DIS, Inc.). Portfolio. Online resource. <http://www.dis-inc.com/portfolio.html>. Accessed 1 June 2014
- Hijikata K, Takahashi M, Aoyagi T, Mashimo M (2012) Behavior of a base-isolated building at Fukushima Dai-Ichi nuclear power plant during the Great East Japan Earthquake. In: *Proceedings of the international symposium on engineering lessons learned from the 2011 Great East Japan Earthquake*, 1–4 Mar 2012, Tokyo
- Kalpakidis IV, Constantinou MC (2009a) Effects of heating on the behavior of lead-rubber bearings. I: theory. *ASCE J Struct Eng* 135(12):1440–1449
- Kalpakidis IV, Constantinou MC (2009b) Effects of heating on the behavior of lead-rubber bearings. II: verification of theory. *ASCE J Struct Eng* 135(12):1450–1461
- Kalpakidis IV, Constantinou MC, Whittaker AS (2010) Effects of large cumulative travel on the behavior of lead-rubber seismic isolation bearings. *ASCE J Struct Eng* 136(5):491–501
- Kelly JM (1990) Base isolation: linear theory and design. *J Earthq Spectra* 6:223–244

- Kelly JM (2004) Seismic isolation. In: Bozorgnia Y, Bertero VV (eds) *Earthquake engineering from engineering seismology to performance-based engineering*. CRC Press, Boca Raton, pp 676–715
- Multidisciplinary Center for Earthquake Engineering Research (MCEER) Online resource. http://mceer.buffalo.edu/info/service/reference_services/baseIsolation2.asp. Accessed 1 June 2014
- Robinson WH, Tucker AG (1977) A lead-rubber shear damper. *Bull NZ Nat Soc Earthq Eng* 10:151–153
- Robinson WH, Tucker AG (1983) Test results for lead-rubber bearings for the William M. Clayton building, Toe Toe bridge, and Waitotukupuna Bridge. *Bull NZ Nat Soc Earthq Eng* 14:21–33
- Skinner RI, Robinson WH, McVerry GH (1993) *An introduction to seismic isolation*. Wiley, Chichester
- TEPCO (2007) Damage of Kashiwazaki-Kariwa nuclear power plant site in the Niigataken Chuetsu-oki Earthquake in 2007. Resource document. <http://www.tepco.co.jp/nu/kk-np/chuetsu/another-j.html>. Accessed 1 June 2014
- Walters M (2003) The seismic retrofit of the Oakland City Hall. In: Huang M (ed) *Proceedings of SMIP03 seminar on utilization of strong-motion data*, Oakland, 22 May 2003, pp 149–164

Learning from Earthquake Disasters

Sarah Beaven^{1,3}, Thomas Wilson^{1,3},
Lucy Johnston², David Johnston^{4,5} and
Richard Smith⁶

¹Department of Geological Sciences, University of Canterbury, Christchurch, New Zealand

²University of Canterbury, Christchurch, New Zealand

³Natural Hazards Research Group, University of Canterbury, Christchurch, New Zealand

⁴Joint Centre for Disaster Research, GNS Science/Massey University, Wellington, New Zealand

⁵Risk and Society, GNS Science, Lower Hutt, New Zealand

⁶Science and Education, Earthquake Commission, Wellington, New Zealand

Synonyms

All-hazards risk mitigation; Co-creation of knowledge; Disaster data management; Earthquake risk reduction; Global disaster rates; Policy learning

Introduction

The United Nations International Strategy for Disaster Reduction (UNISDR) defines disaster as the “serious disruption of the functioning of a community or society involving widespread human, material economic or environmental losses and impacts, which exceeds the ability of the affected community or society to cope using its own resources” (UNISDR 2014). The idea of learning, from earthquake disasters, to reduce the risks associated with future earthquakes is widely held around the world and across a range of sectors. The United Nations defines disaster risk reduction as the “concept and practice of reducing disaster risks through systematic efforts to analyse and manage the causal factors of disasters, including through reduced exposure to hazards, lessened vulnerability of people and property, wise management of land and the environment, and improved preparedness for adverse events” (UNISDR 2014). The research sector has traditionally contributed to these efforts according to discipline. The physical sciences identify the causes and probability of earthquake hazards and relevant branches of engineering focus on mitigating the impact of earthquakes on the built environment, while the medical and social sciences research a range of human impacts, in socioeconomic, political, and cultural spheres and measures to address them.

The growing emphasis on risk reduction, however, is creating a demand for integrated research approaches that focus directly on contributing to disaster mitigation initiatives. The US Earthquake Engineering Research Institute (EERI), for example, aims to integrate engineering with physical and social science initiatives and so reduce earthquake risk by (1) advancing the science and practice of earthquake engineering; (2) improving understanding of the impact of earthquakes on the physical, social, economic, political, and cultural environment; and (3) advocating comprehensive and realistic measures for reducing the harmful effects of earthquakes (EERI 2014).

UNISDR has also stressed the importance of more multidisciplinary, problem-solving

approaches to disaster risk reduction, warning that such initiatives will require better mechanisms “for integrating science and technology into policy processes” and “for bridging the gaps between environmental, humanitarian, development and governmental actors” (UNISDR 2014).

This emphasis on increased coordination between disciplines, scientific and policy domains, and between global sectors is informed, in turn, by a growing perception that existing knowledge is *not* being translated into effective disaster risk reduction measures. The 2009 Science Plan for Integrated Research into Disaster Risk (IRDR), for example, is a response to concern that “the widening gap between advancing science and technology and society’s ability to capture and use them” is a major factor in rising global disaster losses (ICSU 2008). Jointly sponsored by UNISDR, the International Council for Science (ICSU), and the International Social Science Council (ISSC), this plan aims to foster a better understanding of the processes involved in the translation of “research findings about natural hazards and human behavior” into practices and policies that are “effective in minimizing the human and economic costs of hazards” (ICSU 2008, p. 5, 7).

This Science Plan draws on White et al.’s seminal 2001 review of trends in the hazard and disaster field. Explicitly couched as an investigation of the relationship between increasing disaster losses and the science that could have been expected to slow, arrest, and reverse that rate of increase, this review began with a series of questions: “Is human knowledge and understanding of the causes of the losses inadequate despite the research effort, or is it that existing knowledge is not applied, or not used in an effective fashion? Could there be other explanations that lie outside our scholastic assumption that knowledge has a key role to play?” (White et al. 2001, p. 81).

Finding that the preceding decades had featured increasing convergence in the literature concerning broader causal factors, White et al. concluded that the lack of knowledge explanation was unlikely. Instead the evidence they

draw from this field suggested that existing scientific knowledge had either not been applied at all or had not been applied effectively. In addition, they found that global patterns of unsustainable development had also contributed to rising disaster losses (White et al. 2001).

This encyclopedia chapter is largely concerned with research into the broader, *collective* learning processes and pathways found to have been involved in the translation of scientific knowledge into earthquake disaster risk reduction policy, as well as into challenges to such learning. The first section begins with examples taken from US disaster policy research that are directly concerned with the translation of research findings into policy and practice after disaster events. Using social learning policy approaches, this work has confirmed that earthquake and other disasters have had a focusing effect on policy agendas in the US. Higher-level support, policy frameworks that facilitate risk reduction, increasingly networked global communities, and multilateral learning networks have been found to be key factors in learning disaster risk reduction lessons after major events. These US examples are followed by an example from New Zealand, in which learning from an earthquake disaster has informed an all-hazards risk reduction approach and a policy framework that has devolved responsibility for risk reduction to local levels.

Where the first section is concerned with research focused on the opportunities that contemporary developments offer for disaster risk reduction, the second section turns to consider the inhibiting effects the same developments have had on disaster risk reduction. Outlining evidence of barriers to consensus building both within and across disciplines, sectors, and domains, it also touches on data management issues associated with recent technological and social developments. The article concludes with a brief discussion of the larger global issues that have been found to be associated with the distribution of vulnerability to disasters, unsustainable development, the emergence of subgovernment policy communities, and the emerging focus on major disasters and risk.

Largely limited to considerations of *broader collective learning processes*, the focus is schematic, rather than comprehensive, and so leaves significant research outside the immediate frame. For a comprehensive review of the current wider hazard and disaster research landscape, readers are referred to White et al.'s elegant 2001 account. The rich body of work concerned with historical and other understandings of lesson learning after earthquake disasters also falls largely outside scope, as does extensive recent progress in understanding the psychological and other processes involved in decision-making by individuals.

Finally, a brief clarification concerning the tone of this entry, and expectations as to audience. The expectation for the encyclopedia entry is that the majority of its readers are unlikely to be familiar with the field. Accordingly current findings in the relevant topic area are conveyed as clearly as possible for a wider disciplinary audience. This particular field is often divided, particularly over the significance of the extension of governance to include social actors and current emphases on consensus and risk reduction. Such debates are understood to be integral to the production of knowledge and, where relevant to the focus of this entry, are conveyed as evenhandedly as possible. The field is also characterized by densely logical argument, the comprehensive referencing of influences, carefully defined terminology, and painstaking attention to the detail, diversity, and complexity of subject matter. For reasons of scope, and in the interests of clarity, the following comparatively brusque reduction of this diverse wider body of work to a few broad findings runs counter to these hallmarks. References are provided for those who wish to pursue the arguments in more depth.

Translating Science into Policy and Practice

It is widely acknowledged that disaster risk reduction is the responsibility of national governments and needs to be effected at local levels (UNISDR 2014; Global Network 2013). Recent

global developments, however, have impacted on political authority at both national and local levels. Economic integration and market liberalism have limited the authority of governments over multinational economic actors, global bodies are able to impose political requirements on nation states that often conflict with the demands of their constituents, and information and communication technologies have facilitated both global economic transactions and transnational networks of civil society actors (Skogstad 2003). This has contributed to an increase in expert and market-based authority, at the expense of democratic process, as governments respond to an increasingly complex and fragmented policy-making environment by turning to these and other sectors for resources and cooperation (Skogstad 2003, 2005). But at the same time, increased appetite for public participation – fuelled by transnational networks of civil society actors – has also contributed to an international trend of greater public involvement in policy formation, so offering the potential, at least, for new forms of democratic process (Rowe and Frewer 2005; Skogstad 2003).

As a result, a large part of policy making in industrial societies is now carried out collaboratively by government officials, scientific and other experts, and public representatives, through “decentralized, and more or less regularized and coordinated, interactions between state and societal actors” (Skogstad 2005). As a means “to coordinate resources of information, support and authority across state and non-state actors,” this subgovernment sphere has been recognized to have the potential to provide “more effective government, and arguably... more legitimate government” (Skogstad 2005). This section is concerned with research that has focused on the *opportunities* for earthquake risk reduction in the United States and New Zealand that have been provided by these changes.

Higher-Level Coordination in the US Earthquake Policy Domain

US policy theorists interested in the role of technical information in policy making have responded to these changes in the political

environment and to emerging complexity theory, by focusing on the wider influences on the policy process over time. In particular, this expanded focus includes all actors who regularly and actively attempt to influence public policy concerning a particular issue or problem. Collectively described as policy subsystems, communities, networks, or domains, these larger groupings have been found to typically manifest (one or more) advocacy coalitions (Skogstad 2005; Birkland 1998; Sabatier 1998). Usually including actors from a range of government and other organizations and interest groups, this kind of coalition has been found to be held together by shared beliefs about the relevant policy issue. By engaging “in coordinated activity over time,” advocacy coalitions have been found to influence policy outcomes (Sabatier 1998). It has been widely recognized, however, that major policy changes have not been caused solely by the activity of such coalitions. Instead, these policy changes only happen when external factors create a much wider perception that the existing policy has *failed*. Equally, although necessary, this wider consensus has not been found sufficient in itself to trigger major shifts in policy (Birkland 1998; Sabatier 1998).

The Focusing Effect of Earthquake Disasters

The only significant application of this theory in the earthquake risk reduction context is Birkland’s (1998) influential comparative study of four US disaster policy domains between 1960 and 1990. This project compares all congressional hearing testimony concerning earthquake, hurricane, oil spill, and nuclear accident disasters over this 30-year period. It was designed to test the idea that “sudden, attention-grabbing events, known as focusing events,” could advance issues on the relevant policy agenda and – by contributing to the perception that existing policy had failed – potentially trigger major changes in risk reduction policy (Birkland 1998). Birkland established that major US earthquake events between 1960 and 1990 had had a significant focusing effect on the congressional earthquake policy agenda. After each major event, the dominant topic on the policy agenda changed, the

volume of earthquake testimony increased, and much more of this post-event testimony was critical of existing policy. Testimony that referred to specific earthquake events was also much more likely to be critical of existing policy than other earthquake testimonies (Birkland 1998). This contributed to the much wider consensus concerning the need for major change to earthquake risk reduction policy that led to the 1977 National Earthquake Hazard Reduction (NEHR) Act (Birkland 1998).

Birkland’s (1998) research also established that the focusing effect of major events was much greater in the earthquake domain than in the other three disaster domains. The comparison between congressional testimony in the oil spill and nuclear power plant accident policy domains indicated that such differences were likely to be related to disaster coverage (Birkland 1998). Major nuclear accidents had had little impact on the volume or tone of congressional testimony, which continued to feature a deadlock between the powerful industry lobby and those opposed to it. By contrast, the high-profile Exxon Valdez disaster triggered the resolution of a similar 14-year deadlock between the oil industry and environmental advocacy groups, by way of the major policy change enacted through the 1990 Oil Pollution Act (Birkland 1998). Birkland (1998) found that the ambiguity of the harm caused by nuclear accidents contributed to their lack of focusing effect. Unable to be visually conveyed to the public, this harm was open to scientific dispute, unlike the dramatic footage of oil-soaked beaches and struggling wildlife following the Exxon Valdez disaster.

Coordinated Scientific/Political Advocacy

Testimony in the earthquake policy domain revealed that the 1977 NEHR Act was drafted and enacted by a very active coalition of senior scientists, academics, and government officials, who had come together over the preceding 15 years to lobby congress for changes to the existing earthquake policy (Birkland 1998). Coalescing around a shared commitment to science-based earthquake risk reduction, this group dominated the wider policy community involved in

congressional earthquake hearings and policy formation between 1960 and 1990. After the 1964 9.2 Alaskan and the 1971 6.6 San Fernando events in particular, this coalition was responsible for authoritative critiques of existing policy in congressional science committee hearings (Birkland 1998; Hamilton 2003).

This distinguished the earthquake policy domain from the hurricane domain over the same period, which continued to be dominated by local governmental interests pushing for disaster relief and engineered solutions to hurricanes. Although prominent researchers and officials had warned of the need for hurricane mitigation over this period, they had not succeeded in activating a multi-sectoral coalition that lobbied “for improved federal policy to deal with hurricanes in *testimony before the Congress*” (Birkland 1998; emphasis original). Where the earthquake agenda featured a mix of mitigation and disaster relief policy and was usually heard in congressional science committees, the majority of hurricane testimony concerned disaster relief and was heard in congressional public works committees, where mitigation was subordinate to local distributive spending on disaster relief and public works projects (Birkland 1998).

Higher-Level Coordination of Disciplinary and Agency Fragmentation and Competition

The translation of science into policy in this US earthquake disaster domain thus required that the scientific community became politically active and forged alliances within the government sector, first in pushing for changes to policy and then in drafting and jointly enacting these changes. This was a complex and extensive process, however, which involved as much competition as it did consensus building, both within and between scientific and government spheres.

After the 9.2 1964 Alaskan earthquake drew attention to potential earthquake risk in the US, for example, several major earthquake research centers were established which then “competed vigorously to stake out a lead-agency position for earthquake research” (Hamilton 2003). Over the same period three high-level reports proposing

competing national earthquake research programs were presented to congressional committees in 1965, 1968, and 1969 by agencies representing seismological, geological, and engineering interests, respectively. By 1970, “competition among the disciplines and between the agencies, combined with the waning concern after the Alaskan earthquake, contributed to a lack of budgetary attention to the earthquake threat.” (Hamilton 2003).

Then on 9 February 1971, the 6.6 San Fernando earthquake caused 65 fatalities and spectacular footage of collapsed freeway overpasses and hospitals. Renewing both public and high-level political interest in the issue, this event led to federal reforms merging relevant agencies, at the same time as renewed congressional activity led to the 1977 NEHR Act (Hamilton 2003; Birkland 1998). A sequence of further changes to the Act established the NEHR Program (NEHRP), formalizing the role of the advocacy coalition and addressing the fragmentation and competition that had characterized this domain by integrating all earthquake-related disciplines, policy, and practice into a single coordinated program (Hamilton 2003). Responsible for policy formation and for funding and coordinating a range of research programs, this high-level multilateral initiative continues to include US federal agencies responsible for emergency response (Federal Emergency Management Association [FEMA]), standards and technology (National Institute of Standards and Technology [NIST]), and for both academic (National Academy of Sciences [NAS]) and government research (United States Geological Survey [USGS]). The directors of these agencies, together with those of the White House Office of Science and Technology Policy (OSTP), and the Office of Management and Budget (OMB) make up the NEHRP Interagency Coordinating Committee (ICC) (NEHRP 2014).

This research into the US earthquake policy domain established that scientific findings concerning earthquake risk reduction have been translated into policies and practices through complex, high-level processes that have extended well beyond decision-making interactions

between individual scientists and policy makers. Key factors included:

- A political environment in which governance structures were being extended to include non-state actors, through emerging subgovernment networks responsible for drafting and implementing policy
- High-profile earthquake disasters, which triggered a widespread consensus that policy change was required – lessons really have been learned *from* earthquake events
- A politically active scientific community, able and open to forming the ongoing coalitions required in the formation of such networks and committed to using the focusing effect of disasters to push for policy change
- Ongoing high-level political intervention to reduce disciplinary and government sector fragmentation and competition, by creating an integrated national earthquake risk reduction coordinating program.

Disaster Risk Reduction Policy and Practice: Coordination at Local Levels

Research into the role of policy subsystems in disaster risk reduction has not been restricted to higher-level changes to policy. Others have focused on such activity at the local level. Busenberg (2000), for example, has studied the local policy subsystem that emerged in Alaska after the 1989 Exxon Valdez disaster. The 1990 Oil Pollution Act (OPA) provided for a Regional Citizens Advisory Council (RCAC) to oversee environmental risk management in the region, legislating a permanent role for the advocacy coalition involved in its design and implementation (as did the 1977 NEHR Act) (Birkland 1998). Guaranteeing the independence of the RCAC, the 1990 OPA requires the oil industry to fund the operations of this body for as long as the Alaskan Pipeline remains in operation. The RCAC has 19 member organizations. These include a range of local interest and local government groups, each with a representative on the Board of Directors; five issue-specific committees include board members, experts, and local citizens.

Busenberg focused on the innovative learning arrangements that brought the RCAC into the wider, multilateral policy network involved in significant policy decisions concerning the environmental management of the industry in the region (Busenberg 2000). Between 1990 and 2000, these decisions resulted in policy changes that significantly reduced the risk of future oil spill disasters. This included legislation requiring a sophisticated local weather forecasting system, an integrated regional marine firefighting service, and supported tanker navigation (including redundant tug capacity) within Prince William Sound. The first of these initiatives was immediately enacted, while the latter two required supporting evidence established by research programs partially or completely funded by the RCAC (Busenberg 2000).

Social Learning in Relation to Other Theories of Learning and Participation

Busenberg and Birkland are both working within the social learning policy theory tradition. For both, the term “learning” does not simply refer to an increase in knowledge but is rather defined in relation to policy change. Busenberg has distinguished the incremental, ongoing local-level policy learning that he traces over 10 years in Prince William Sound from the larger changes in policy after major events, which interests Birkland. The coordination and consensus building process is, however, very similar, and both researchers define the social learning process in relation to policy outcome. Busenberg finds that local, incremental learning within the relevant networks of organizations and individuals occurred only when initiatives gained enough support from key stakeholders to be *enacted as risk-reducing policy change*. In the case of supported tanker navigation, this learning had required a higher-level intervention from the state governor. By contrast, initiatives that were researched and promoted but were not enacted in policy are interpreted as evidence that, in these cases, social learning did *not* occur. Similarly, Birkland defines social learning with reference

to evidence of policy outcome: “Policy learning can be identified if there is *prima facie* evidence of policy changes that are reasonably linked to the causal factors that connected the event under consideration to its harms, and if addressing these factors would be likely to mitigate the problem” (Birkland 2009).

The social learning process explored by these and other policy theorists is described as social because it is not restricted to individuals, or organizations, or even sectors, but relies on a *much wider social consensus* concerning the failure of existing policy. It is understood to be a learning process in part because it draws from other learning theories that also feature this negatively driven sequence, in which previously unexamined assumptions are understood to have failed, and so triggered a new set of assumptions. Informed by both Kuhn’s scientific theory of paradigm shifts and so-called double-loop organizational learning theory, for example, social learning theory also relies on the punctuated equilibrium pattern identified by complexity theory as that in which complex human, social, ecological, and biological systems learn by adapting to disruption. Equally, for most of these theorists, the activities of policy subsystems are described as learning because they *increase knowledge* concerning both the policy problem and likely outcomes of policy alternatives (Sabatier 1998). The emphasis on improved policy outcomes, however, is distinct to Busenberg and Birkland’s application of this approach to disaster risk reduction policy.

It is useful to clarify the relationship between this social or policy learning framework and similar concepts current in the hazard and disaster context. A variety of closely related learning arrangements have been recently reviewed by Voorberg et al. (2013). Covering a variety of disciplines, this review finds the terms social innovation, knowledge co-creation, or knowledge coproduction are all used to describe what appears to be the same broad learning process (note that their review does *not* include constructivist uses of the term knowledge coproduction, which refer to the mutually constitutive relationship between

science and society). This embedded process takes place in a specific local and institutional context, in which relevant stakeholders participate and collaborate across organizational boundaries and jurisdictions, engaging in interactive exchanges of knowledge, information, and experiences through intra- and interorganizational networks (Voorberg et al. 2013). To this extent, the social learning process constitutes a particular variant of this collaborative learning process. However, Voorberg et al. (2013) found that irrespective of discipline, the focus is almost entirely restricted to considering the process, and value, of such participation, rather than its outcomes, if any. By contrast, the social learning process researched by Birkland and Busenberg is defined in relation to policy outcome.

Secondly, it is also useful to relate the social learning policy process to the international trend of greater public and community involvement in policy formation. Such participation is recognized to be an essential component of any strategy designed to increase community resilience to disasters. In a review of relevant literature, Rowe and Frewer (2005) found more than 120 types of public participation described. They distinguish between communication, consultation, and participation, in order to develop a useful typology (Table 1).

According to this typology the RCAC studied by Busenberg can be seen as a *participatory mechanism* of public engagement – although established as a council, the focus on oil spill disaster risk reduction puts it in the task force category: the use of “a small group of participants (public representatives) with ready access to all pertinent information, to solve specific problems” (Rowe and Frewer 2005). Congressional hearings, however, largely constitute a public *communication mechanism*. The social learning that concerns Busenberg and Birkland does thus rely on forms of public engagement. Their focus, however, is not on public engagement per se, but rather on the larger policy process, including all events and factors that influence policy design and enactment; again, their emphasis is on the policy outcome.

Learning from Earthquake Disasters, Table 1 Mechanisms of public engagement (Adapted from Rowe and Frewer 2005)

Engagement mechanism	Examples	Characteristics
Communication		Information to -> public
Communication type 1	Traditional “publicity”	Controlled selection; targets particular population with set information
Communication type 2	Public hearings, seminars, etc.	Rely on public initiative, provide flexible information
Communication type 3	Drop-in centers, website information	Rely on public initiative; provide set information (even if FTF)
Communication type 4	Hotlines	Rely on public initiative; uncontrolled selection/flexible information
Consultation		Information from <- public
Consultation type 1	Opinion poll, survey, referendum, etc.	Large samples – specific questions; controlled selection/ closed response
Consultation type 2	Consultation document	Open responses on specific issues – controlled selection
Consultation type 3	Electronic consultation – interactive website	Open responses on specific issue – uncontrolled selection
Consultation type 4	Focus group	Emphasis on information quality; open responses, controlled selection
Consultation type 5	Study circle	Emphasis on information quality; open responses, uncontrolled selection
Consultation type 6	Citizen panel – e.g., health panel	Ongoing debate of topics; controlled selection, open responses
Participation		Information to/from <-> public
Participation type 1	Planning workshop, citizens jury	Controlled selection, facilitated expert and other inputs, open responses
Participation type 2	Negotiated rule making; task force	Groups access pertinent information to solve problems; open responses controlled selection, inputs unfacilitated
Participation type 3	Deliberative polling or planning cell	Polled before/after deliberation; controlled selection, open responses
Participation type 4	Town meeting with voting	Voting after debate; uncontrolled selection; open response mode

All-Hazard, Risk-Based, Problem-Solving Approaches to Disaster Mitigation

The third example of this kind of social learning from earthquake events is taken from New Zealand and occurred in the aftermath of the 1987 M_w 6.3 Edgumbe earthquake. An earlier history of seismic activity, including the destructive M_w 7.8 Napier earthquake in 1931, had contributed to well-enforced New Zealand building codes. These codes, and the fact that the Edgumbe event occurred in a largely rural area, contributed to a lack of fatalities. These factors had not, however, been able to prevent total losses of NZ\$430 million. Drawing political and international attention to the potential for

much greater earthquake losses in New Zealand and the Pacific region, the Edgumbe event also highlighted the potential for losses associated with damage to critical facilities, including supply of water, electricity, and communications, as well as sewerage disposal and transport (Shephard 1997).

Largely in response to this event, major national policy changes “designed to encourage more effective safety and loss prevention strategies” were enacted in a series of subsequent government reforms, including the 1987 Recovery Plan for Natural Disasters, and reforms to the Local Government Act 1989, the Resource Management Act 1991 and the Building Act 1991

(Shephard 1997). The “essential idea” of these reforms: “was that central government would accept shared responsibility for the restoration of damage from natural disasters *only if the local authority concerned had done its part to minimize, mitigate and manage the risk to its assets...* they were expected to... *put in place protection and damage limitation measures that would reduce the consequences.*” (Helm 1996; cited in CAE 1997, emphasis added).

Where the focusing effect of US earthquake and oil spill disaster events had led to *hazard-specific policy reforms*, the New Zealand government emphasis on risk and decentralized responsibility effectively meant that the lesson learned from the Edgecumbe earthquake was to *prepare locally for all hazards*.

This political climate facilitated the collaborative Christchurch Engineering Lifelines Project, conducted in 1993 and 1994 by task groups that included physical scientists, academic and professional engineers, local and national government agencies, and representatives from relevant private sectors, including insurance and utility providers. The project was run out of the University of Canterbury’s Center for Advanced Engineering (CAE), which was founded in May 1987, 2 months after the Edgecumbe earthquake. The Christchurch Engineering Lifelines Project aimed to identify the vulnerability of critical facilities to damage from a *range of natural hazards* and to arrive at practical strategies to mitigate potential damage; it also aimed to communicate relevant issues to the people involved in managing the services and the public (CAE 1997). Creating detailed hazard and critical facility maps, multi-sector task groups came together to superimpose these maps, going over critical facilities in detail, network by network, to assess the likely vulnerability and importance of elements, discuss often inexpensive mitigation measures, and assess risks associated with network interdependence.

Rather than simply translating scientific findings into risk reduction policy and practice, this project involved the collective, cogeneration of risk reduction knowledge through the interactions of scientists, engineers, local government,

and those with day-to-day experience of the relevant lifelines. In addition to being deeply local and enabled by a strong national policy framework, this project was also informed by globally networked international research and industry communities.

The initial assessment phase was conducted during 1993 and 1994. In collaboration with the New Zealand Society for Earthquake Engineering (NZSEE) and with financial support from New Zealand’s Earthquake Commission (EQC), project members also visited Los Angeles in 1994 following the M_w 7.2 Northridge earthquake and Kobe after the 1995 Great Hanshin earthquake. The Christchurch project was detailed and disseminated in a 1997 report (CAE 1997). In addition to summarizing the mitigation measures already carried out and ongoing project work, this report noted that “a very significant benefit” had been the formation of the multilateral working group, which had brought together those with “dissimilar working interests,” and so established useful ongoing relationships “in preparation, as it were, for a major emergency requiring the various utilities to work together” (CAE 1997). The subsequent Civil Defence and Emergency Management Act (2002) built on this legacy. Devolving hazard and emergency management to the local level and supporting the integration of regional lifelines groups in emergency management training programs, this act reinforced the culture of hazard mitigation, contributing to a community of practice involving regional providers of critical facilities.

In 2010 the M_w 7.2 Greendale earthquake began the destructive “Canterbury earthquake sequence” that included the M_w 6.2 Christchurch earthquake on 22 February 2011. Leading to 185 deaths in Christchurch, widespread liquefaction and lateral spreading in the central and eastern suburbs of the city, this event caused extensive damage to buildings and critical facilities. In a recent study, the Lifelines project and its outcomes have been estimated to have significantly reduced the impacts of the earthquake sequence on the city (Fenwick 2012).

Risk reduction initiatives developed through the project are estimated to have greatly reduced

the overall community losses associated with critical facility outages. Investment in alternative, redundant supply routes for electricity and telecommunications allowed for rapid restoration of supply in all but the most badly damaged regions. Telecommunications companies with “strong node” and alternative “hot spot” buildings were able to maintain coverage in the city. Seismic strengthening in the city’s port, directly over the epicenter of the damaging 22 February 2011 event, made it possible for a New Zealand naval frigate to dock and unload aid and supplies within 4 h of this event; the port was operational again within days (Fenwick 2012).

Utility companies that had adopted a consistent focus on risk mitigation also saved significantly in direct asset replacement and repair costs. Electricity provider Orion Energy, for example, spent \$NZ6 million on seismic strengthening and is estimated to have saved \$NZ60–65 million, while the total repair costs after the February event paid by another provider, Transpower, amounted to only \$NZ150,000 (Fenwick 2012). By contrast, much of the aging water and wastewater network the project had identified as a seismic risk had proved too expensive and difficult to upgrade. The 2014 estimate of the cost of repairing these seismically damaged networks is NZ\$2 billion (SCIRT 2014).

At least as valuable as the financial saving was the networked community of professional and academic engineers, infrastructure companies, and local and national authorities created by the project and further developed by EQC, NZSEE, and MCDEM initiatives. This collaborative critical facilities culture greatly facilitated the restoration of lifeline function after each of the four major events in the 2010–2011 Christchurch earthquake sequence. It has also informed the “Stronger Christchurch Infrastructure Rebuild Team” (SCIRT), a formal alliance of national and local government agencies and private organizations tasked with rebuilding the city’s horizontal infrastructure (SCIRT 2014). Designed to “harness the expertise of public and private sectors, promote information sharing and

Learning from Earthquake Disasters, Table 2 Comparative policy learning from earthquake disasters in the United States and New Zealand

<i>In the United States and New Zealand:</i>	
<i>Major earthquakes caused extensive damage</i>	
At a time when governance structures were extending to include non-state actors	
Creating a widespread perception that existing legislation had failed	
Leading to a range of legislation designed to address issues exposed by the event(s)	
Generating increased research activity	
Motivating ongoing coalitions between state and non-state actors in learning networks	
<i>Differences:</i>	
United States	New Zealand
Hazard-specific legislation	All-hazard legislative focus
Hazard-specific research and coalitions	All-hazard research and coalitions
Higher-level initiatives and organizational coalitions	Devolution risk management to local-level initiatives and organizational coalitions

collaboration, and manage risks and opportunities” this work program includes seismic strengthening (SCIRT 2014).

As in the US examples, it is clear that the 1987 Edgecumbe earthquake occurred at a time in which governance structures were being extended; the earthquake thus contributed to a widespread consensus that policy change was required, a range of legislation devolving responsibility for risk reduction to the local level, increased research activity, and the motivation to form the ongoing coalitions holding the relevant learning network together. The 2010–2011 Canterbury earthquake sequence confirmed the value of this learning, while also focusing attention again on the need for seismic risk reduction and so informing the Stronger Christchurch Infrastructure Rebuild Team initiative.

However, the multi-hazard approach and the devolution of responsibility for all hazards to the local level differ significantly from the *hazard-specific* US policy responses to earthquake disasters (see Table 2).

To the extent that it has required that local authorities, communities, researchers, and the private sector work together to reduce the risks of a wide range of hazards in their own local region, the focusing effect of an earthquake disaster has been used to reduce the risks associated with a much wider range of disaster types. This can be seen as an example, then, of the kind of “all-hazard, risk-based, problem-solving approach” to disaster risk reduction recommended by UNISDR, which requires “collaboration and communication across the scientific disciplines and with all stakeholders, including representatives of governmental institutions, scientific and technical specialists and members of the communities at risk to guide scientific research, set research agendas, [and] bridge the various gaps between risks and between stakeholders” (UNISDR 2014). On the other hand, however, this process remained limited in focus to the vulnerability of city infrastructure – it did not include social vulnerability or involve the significant public participation constituted by the RCAC; so to this extent, it did not involve the collaboration with “all stakeholders” specified by the UNISDR.

Challenges to Learning from Earthquake Disasters

Considering the challenges to this kind of learning after earthquakes means turning from work focused on the opportunities offered by the focusing effects of earthquake events, increasingly networked global communities, and the extension of governance to include social actors to work that is focused on the *risks* posed by the same developments. This section starts with challenges to the formation and operation of collaborative, “innovative” learning or policy communities, and a brief discussion of identified barriers to higher-level responses to earthquake disasters. Data management issues and wider implications of the role of focusing disaster events and unsustainable global development conclude the section.

Challenges to Multidisciplinary, Multi-sector Learning After Earthquakes

The processes that bring actors together from across sectors and disciplines to coalesce around the shared goal of risk reduction are also at work around the sets of ideas or paradigms that are shared within individual disciplines, organizations, and sectors. At every level, incompatibility between these idea sets, or shared norms and values, have been found to have the potential to inhibit knowledge sharing and so impede the translation of scientific knowledge into policy and practice.

This is apparent in Birkland’s (1998) comparative study, for example, which revealed that a relative lack of focusing effect effectively left the hurricane policy domain dominated by local government interests, who prioritized policies that were *relevant to their constituents*. This meant they preferred engineered solutions like levees nonstructural mitigation techniques, such as stricter enforcement of flood plain regulation or coastal development, considered by Birkland and other researchers to be a more *scientifically credible* mitigation approach. Similarly, the “ambiguity” of the harms caused by nuclear accidents had left *commercial* industrial interests dominating the nuclear disaster policy domain. Focusing events had been required in the oil spill and earthquake domains to generate enough social and political pressure to trigger higher-level commitment to risk-reducing policy change. Without this shift in the political environment, the natural disaster domain remained dominated by the drive for *political relevance*, while the industrial disaster domain remained dominated by the commercial drive for *financial profit*; neither of these drivers was compatible with the drive for *scientific credibility* of concern to Birkland and Busenberg.

Incompatibility between these drivers has also been found to significantly inhibit the translation of scientific findings into policy and practice at the level of individual decision-making. In a recent study, for example, policy makers and practitioners working in disaster risk reduction and environmental management domains, in

both developed and developing countries, did not access relevant academic publications unless they had been directly involved in their production (Weichselgartner and Kasperson 2010). They were in any case reluctant to use this research – assessing the vulnerability and resilience of populations in their own geographic jurisdictions – to inform decisions about disaster risk reduction and environmental policy and practice. Although conceding broad relevance, policy makers and practitioners did not consider these research findings to be relevant to their specific knowledge needs (Weichselgartner and Kasperson 2010). At the same time, even in this domain, where researchers are typically highly motivated to contribute to disaster risk mitigation, the authors of the relevant publications acknowledged that their work was largely driven by considerations of *scientific credibility* and so designed, conducted, and disseminated with an academic audience in mind, rather than the needs of potential users (Weichselgartner and Kasperson 2010).

By creating incompatible cultures and values, these drivers have also been found to contribute to value clashes between academic and other spheres that inhibit cross-sector collaborative activity (Cash et al. 2003). Academic incentive regimes are based on publications and peer recognition, which have been described as academic “status competitions,” sometimes taking “the form of winner takes all” when it comes to publishing first or winning research grants (Bruneel et al. 2010). Contesting the credibility of findings is an essential part of the peer review process, informing a value set in which contest is more usual than that in private and government sectors, where individuals are required to subordinate their own interests to those of their organization and where commercial and political incentive regimes can require that knowledge is kept private, rather than contested in the public domain (Bruneel et al. 2010).

In a range of collaborations involving research and private sectors, these differences in cultures, norms, and values have been found to inhibit collaborative knowledge production. Conversely, ongoing cross-sector engagement over time has

been found to *significantly reduce* this inhibiting effect (Bruneel et al. 2010; Cash et al. 2003). The balance of contestation and consensus building required to initiate and continue such cross-sector engagements, however, has been found to be delicate; major value clashes have been found to be particularly compromising.

Challenges to Higher-Level Learning from Earthquake Disasters

The tensions between potentially incompatible scientific, economic, and political drivers that have worked against coordinated disaster risk reduction initiatives at local and national levels are also in evidence in the recent global developments that have reduced political authority at these levels. Economic integration and market liberalism have limited the political authority of governments over multinational economic actors; bodies like the WTO are able to impose global political requirements on nation states that often conflict with the demands of their constituents, and information and communication technologies have facilitated both global economic transactions and transnational networks of civil society actors (Skogstad 2003). These limitations on political authority at the national level, moreover, have not been evenly distributed. Less affluent, developing nations have been much less well positioned than developed nations when it comes to asserting political sovereignty over the activities of multinational companies and against the political strictures of global organizations. Consequently the nations *most exposed* to disaster risk have also been those nations least able to shoulder the responsibility for disaster risk reduction that continues to fall to national government level.

This has meant that developed countries that experience earthquake disasters are more likely to benefit from the distribution of political, public, and international attention and have more to gain from the resultant focusing effect than developing countries. A sequence of Global Network reports has found that networked disaster risk reduction activities involving “state and non-state actors” working “together to ensure the safety and well-being of their communities”

have been much more effective “in countries where capable, accountable and responsive local governments worked collaboratively with civil society and at-risk communities” (Global Network 2013). However, there has been very little progress in developing countries across a range of risk governance indicators, including “lack of political authority; inadequate capacities and financial resources; and minimal support from central government,” which have all been identified as “significant barriers” to the implementation of disaster risk reduction policies at the local level (Global Network 2013).

In many countries, this can mean that rather than triggering risk-reducing policy changes, the focusing effect of a major earthquake disaster can instead generate a tidal wave of international aid and more or less fragmented and competitive activity among aid organizations. Dombrowsky, for example, has found US\$4 billion in aid donated worldwide after the 2004 Sumatran earthquake and subsequent tsunami provided an “excessive” volume of food and clothing that “wrecked local agricultural, market and production systems,” while the “accelerated” building of new housing created “extreme shortages of materials, harmful deforestation, profiteering and, more recently, inappropriate construction methods” (Dombrowsky 2007). The “tens of thousands of projects” carried out by competing international organizations in Southeast Asia after the Sumatran earthquake and tsunami “were not networked and there was no co-ordination aimed at developing a long-term, preventive protection and sustainability strategy” (Dombrowsky 2007).

Conversely, Christoplos (2006) has argued that although scientifically credible, disaster risk reduction initiatives can be incompatible with local and global *political* imperatives after disasters. Informing the humanitarian focus on “addressing acute human suffering,” these political drivers also inform the immediate priorities for local “politicians, donors and disaster affected people,” who need to get “people out of tents and into houses and off food aid and into jobs” (Christoplos 2006). In addition, he noted that the 2004 tsunami decimated many of the local

governments and communities required to develop networked community disaster risk reduction initiatives. He has called for compromise, rather than blame casting and identifying “villains,” from those working toward disaster risk reduction, in order to achieve a consensus concerning larger, shared goals (Christoplos 2006).

Data Management Challenges to Learning from Earthquake Disasters

The focusing effect of disasters has also been found to contribute to the enormous volume of post-disaster reviews, investigations, and evaluations, most entitled “lessons learned,” which are conducted and documented at different levels, in a range of government, academic, NGO, and private sectors. Again, this material is largely produced in response to *political* pressures and so conducted to reassure relevant authorities and the public that an attempt to learn lessons has been made, rather than to generate scientifically credible evidence for policy change (Birkland 2009). International aid agencies are also under intense political pressure to prove that their money is being well spent. This pressure can lead NGOs to compete for prestige projects and the associated positive publicity. It also contributes to considerable information gathering, largely for internal consumption, within organizations (Dombrowsky 2007). Much of this material may contain risk reduction lessons of scientific and practical value. It cannot, however, be easily accessed, since it is not coordinated or collated, but remains fragmented, dispersed within agencies and organizations, and across a diverse range of academic journals (Birkland 2009).

The same issue occurs at the global scale. As the pace with which information technology evolves outstrips the capacity to *coordinate* data management across local, national, and global levels, it becomes increasingly difficult to establish global disaster data. Organizations like Munich Re and the Center for Research on the Epidemiology of Disasters (CRED) have been set up to address this issue by gathering credible, comparable global data and providing it in an

accessible form. Even these organizations, however, are struggling with inconsistencies created by “the use of different data sources, different frameworks, different metrics, or different scales” and because of “a lack of consistency in the management of the data on which the assessments are based” (ICSU 2009).

The management and accessibility of *research data* in this area has also been identified as an urgent issue. The 2013 OECD Global Science Forum report on Data and Research Infrastructure for the Social Sciences, for example, addressed “rapid growth in new forms of data collected in conjunction with commercial transactions, internet searches, social networking, and the like, and by technological advances in the capacity to access and link existing survey, census, and administrative data sets.” It was tasked with weighing the opportunities offered by this growth for evidence-based research concerning disasters and other global challenges related to the health and well-being of populations around the world (OECD 2013). Issues identified in this report as imperative included the difficulty of accessing the large volumes of data collected and held by the government and private sectors; ethical issues associated with the use and sharing of data concerning human subjects; the need to standardize, manage, and archive data in accessible formats; and the urgent need for more international comparability and collaboration. All these issues reflected a larger problem: the need for a more integrated, coordinated approach to the collection, management, and archiving of data, at every level. The report concluded that the “the need for global coordination of activities designed to address the challenges identified in this report is paramount” (OECD 2013).

This has been identified as the paradox of the twenty-first century (Dombrowsky 2007). More data is available than ever before, in a huge range of areas. Existing national structures and systems, however, are no longer able to manage the complexity and volume of multinational data flows, and there are as yet no global structures and guidelines through which such data can be managed, coordinated, and applied for disaster risk reduction purposes.

Challenges to Learning from Disasters Arising from Focusing Effects

The lack of coordinating structures at the global level is also an issue when it comes to coordinating disaster risk reduction efforts. Incompatible economic, political, and scientific drivers have been found to be best coordinated, at the national level, through policy frameworks that facilitate an integrated focus on the issue of disaster risk reduction. At the global level, however, where similar pressures are contributing to rising disaster losses, there are as yet no structures with the authority or means to establish equivalent global frameworks.

Disaster sociologists point to the extent to which the social issues created by global tensions, rather than being foregrounded by rising disaster rates, have instead been *obscured* by the focusing effect of major natural disasters like earthquakes and tsunami. By drawing attention to the drama associated with such natural forces, in other words, earthquakes and other natural disasters have distracted attention from the extent to which the capacity to cope with and recover from such extreme events can be reduced by lack of resources, including access to information, financial credit, social support, and services. The focus on major natural disasters can thus obscure the fact that populations constrained in this way are more vulnerable to death, injury, and loss during disasters and less able to recover (Gaillard 2010).

On the larger scale, the focusing effect of dramatic earthquake disasters is understood to work in a similar way, grabbing attention from the impacts of climate-related disasters, which are particularly common in less developed global regions. Climate-related events have the most destructive global impact; cause the greatest losses, the most deaths; and are also the hazard event category that is increasing in frequency (Global Network 2013). Moreover, this class of disaster includes the less dramatic, ongoing “everyday disasters” such as droughts and floods, which are collectively responsible for more disaster losses than the more dramatic major events. The hazards associated with extreme poverty have also been found to fall into the category of

“everyday disaster” – responsible for the vast majority of harms and deaths affecting the global human population, but obscured by the focus on so-called “natural” disasters (Global Network 2013).

These issues have had a fragmenting effect within the scientific community. It has been argued, for example, that hazard and vulnerability disaster risk reduction traditions are informed by inverse and fundamentally incompatible paradigms. Such arguments have contrasted a hazard mitigation emphasis on infrequent, major natural disasters, and reliance on expensive technical solutions, designed and implemented at high levels by expert groups and delivered through top-down mechanisms, with the emphasis of vulnerability-based approaches on the social causes of disasters, and on bottom-up, participatory approaches designed to counteract these social constraints (Gaillard 2010). Related constructivist critiques have emphasized the risks to democratic process associated with the delegation of policy formation to subgovernment networks and have been similarly critical of disaster risk reduction policy and initiatives developed by experts that fund technical risk reduction measures, rather than public engagement mechanisms, and measures to address the social issues that contribute to disaster impacts (Hayward 2013).

Summary

There is general agreement within research and policy communities when it comes to some broad principles concerning learning from disasters. Governments are responsible for putting policy frameworks in place to support risk reduction initiatives. Implemented at both national and local levels, the most effective initiatives have reduced fragmentation and competition by adopting collaborative, cross-sector participatory approaches in an attempt to include as many stakeholders as possible, including those most exposed to disaster. Earthquake disasters have had a focusing effect on policy agendas in the US. Higher-level support, policy frameworks that

facilitate risk reduction, increasingly networked global communities, and multilateral learning networks have been found to be key factors in the translation of science into disaster risk reduction policy after major events. In New Zealand, learning from earthquake disasters has informed an all-hazards risk reduction approach and a policy framework that has devolved responsibility for risk reduction to local levels. There is broad agreement that this kind of current extension of the government sphere to include both state and non-state actors in policy formation and implementation can facilitate such approaches and can also facilitate the translation of science into policy and practice, when researchers take the opportunity to engage with all end-users, at both national and local levels.

White et al. have argued that this kind of consensus reflects an increasing convergence in the hazard and disaster research community around the common goal of disaster risk reduction, notwithstanding cross-disciplinary tensions concerning disaster risk reduction priorities and the risks and opportunities associated with the extension of the government sphere. This convergence is also indicated in a cross-disciplinary agreement concerning the importance of working through such tensions collaboratively, both within the scientific community and with other sectors involved in disaster risk reduction.

Fragmentation, competition, and lack of coordination, however, have been found to inhibit the development of multilateral disaster risk reduction initiatives at local, national, and global levels and within and across disciplines, sectors, and domains. At both local and national levels, *coordinating measures* introduced after disasters have been found to reduce fragmentation and balance competition, both within and between sectors. Such measures have facilitated the compromises between economic, political, and scientific drivers required by collaborative risk reduction initiatives. Global disaster risk reduction is challenged by the absence of similarly authoritative coordinating mechanisms at this level and related issues arising from the rapid and uncoordinated development of information technology, the distribution of vulnerability to disasters, the

emergence of subgovernment policy communities, and aspects of current emphasis on risk and disasters. Unsustainable global development, widely acknowledged to be a causal factor in rising disaster losses, also continues to significantly inhibit disaster risk reduction at both national and local levels and so the wider global disaster risk reduction effort.

Cross-References

- ▶ “Build Back Better” Principles for Reconstruction
- ▶ Community Recovery Following Earthquake Disasters
- ▶ Earthquake Disaster Recovery: Leadership and Governance
- ▶ Earthquake Risk Mitigation of Lifelines and Critical Facilities
- ▶ Legislation Changes Following Earthquake Disasters
- ▶ Resilience to Earthquake Disasters
- ▶ Resiliency of Water, Wastewater, and Inundation Protection Systems
- ▶ Seismic Resilience
- ▶ Structural Design Codes of Australia and New Zealand: Seismic Actions

References

- Birkland T (1998) Focusing events, mobilisation, and agenda setting. *J Public Policy* 18(1):53–74
- Birkland T (2009) Disasters, lessons learned, and fantasy documents. *J Conting Crisis Manag* 17(3):146–156
- Bruneel J, D’Este P, Salter A (2010) Investigating the factors that diminish the barriers to university-industry collaboration. *Res Policy* 39:858–868
- Busenberg G (2000) Resources, political support, and citizen participation in environmental policy: a reexamination of conventional wisdom. *Soc Nat Res* 13:579–587
- Cash DW, Clark WC, Alcock F, Dickson NM, Eckley N, Guston DH, Mitchell RB (2003) Knowledge systems for sustainable development. *Proc Natl Acad Sci* 100(14):8086–8091
- Center for Advanced Engineering (1997) Risks and realities: a multi-disciplinary approach to the vulnerability of lifelines to natural hazards. Center for Advanced Engineering, Christchurch
- Christoplos I (2006) The elusive “window of opportunity” for risk reduction in post-disaster recovery. ProVention Consortium Forum, 2–3 Feb 2006, Bangkok
- Dombrowsky WR (2007) Lessons learned? Disasters, rapid change and globalization. *International Review of the Red Cross*, 89(866)
- EERI (2014) Earthquake engineering research institute. <https://www.eeri.org/>. Retrieved 20 Jan 2014
- Fenwick T (2012) The value of lifeline seismic risk mitigation in Christchurch. NZ Earthquake Commission: Wellington, New Zealand
- Gaillard JC (2010) Vulnerability, capacity and resilience: perspectives for climate and development policy. *J Int Dev* 22:218–232
- Global network of Civil Society Organisations For Disaster Reduction (2013) Views from the frontline: beyond 2015. <http://www.globalnetwork-dr.org/views-from-the-frontline/vfl-2013.html>. Retrieved 2 January 2014
- Hamilton RM (2003) Milestones in earthquake research. *Geotimes*, March. www.agiweb.org/geotimes/mar03/comment.html. Retrieved 08 Jan 2014
- Hayward BM (2013) Rethinking resilience: reflections on the earthquakes in Christchurch, New Zealand, 2010 and 2011. *Ecol Soc* 18(4):37. <http://www.ecologyandsociety.org/vol18/iss4/art37/>. Retrieved 01 Jan 2014
- ICSU (2008) A science plan for integrated research on disaster risk. International Council for Science, Paris. <http://www.icsu.org/publications/reports-and-reviews/IRDR-science-plan>. Retrieved 24 Oct 2013
- NEHRP (2014) US National earthquake hazard reduction program. www.nehrp.gov. Retrieved 09 Jan 2014
- OECD Global Science Forum (2013) New data for understanding the human condition: international perspectives. <http://www.oecd.org/sti/sci-tech/new-data-for-understanding-the-human-condition.htm>. Retrieved 10 December 2013
- Rowe G, Frewer LJ (2005) A typology of public engagement mechanisms. *Sci Technol Hum Values* 30(2):252–290
- Sabatier P (1998) The advocacy coalition framework: revisions and relevance for Europe. *J Eur Public Policy* 5(1):98–130
- SCIRT (2014) Stronger Christchurch infrastructure rebuild team
- Shephard RB (1997) A decade of progress since the Edgecumbe earthquake: risk and insurance. *Bull N Z Soc Earthquake Eng* 30(2):185–193
- Skogstad G (2003) Who governs? Who should govern?: political authority and legitimacy in Canada in the twenty first century. *Can J Polit Sci* 36(5):955–973
- Skogstad G (2005) Policy networks and policy communities: conceptual evolution and governing realities. In: Annual meeting of the Canadian political science association. London, pp 1–19
- UNISDR (2014) United Nations international strategy for disaster risk reduction. www.unisdr.org. Retrieved 09 Jan 2014

- Voorberg W, Bekkers V, Tummers L (2013) Co-creation and co-production in social innovation: a systematic review and future research agenda. In: European group for public administration conference 11–13 Sept. Edinburgh
- Weichselgartner J, Kaspersen R (2010) Barriers in the science-policy-practice interface: toward a knowledge-action-system in global environmental change research. *Glob Environ Chang* 20:266–277
- White GF, Kates RW, Burton I (2001) Knowing better and losing even more: the use of knowledge in hazard management. *Glob Environ Chang* 3(3–4):81–92

tends to slow recovery, but often they facilitate in building a better, more resilient, post-disaster environment. In this entry, the focus will be on the common post-disaster legislative changes encountered for reconstruction, such as those affecting building acts and building codes. Examples and effects of the changes are discussed using a variety of cases from earthquakes and other disasters. This entry provides an account of the ways in which legislation changes post-disaster can help and/or hinder reconstruction programs.

Legislation Changes Following Earthquake Disasters

Suzanne Wilkinson¹, James Olabode Bamidele Rotimi² and Sandeeka Mannakarra¹

¹Department of Civil and Environmental Engineering, The University of Auckland, Auckland, New Zealand

²Auckland University of Technology, Auckland, New Zealand

Synonyms

Disaster legislation; Legislation changes post-disaster

Introduction

Most disasters are followed by legislative changes, emergency legislation, and new disaster legislative measures. Some of the changes are a reaction to the need to build back safer and to be seen to be facilitating a better future environment. For instance, the Royal Commission into the Canterbury Earthquakes in New Zealand recommended changes to the Building Act, a rewrite of building codes and building standards. Other disasters, such as the Wenchuan earthquake in China, Black Saturday bushfires in Australia, and Hurricane Katrina and Northridge earthquake both in the USA, had all occasioned changes in legislative requirements for reconstruction and code changes for buildings. The impact of legislative changes

Legislation Challenges Post-disaster

The organization and coordination of recovery is usually complex because a wide range of activities occur simultaneously after a significant disaster. There is an equally wide range of needs that would have to be met, some of which may be conflicting and adds to the complexity of post-disaster recovery. Experiences from past disaster recovery arrangements show a struggle to meet recovery needs and changes to the legislative, and regulatory recovery environments seem common place. Several contributory factors account for the success or otherwise of disaster management goals and objectives, irrespective of whether they are in legislation or in recovery management plans. These include:

- Pre-disaster trends and levels of preparedness which are linked to vulnerability
- The extent of damage resulting from the disaster
- Availability and accessibility to the required resources for both response and recovery
- The prevailing political will and governmental interests in disaster management activities

In terms of legislation and regulation, there is often little provision in legislation to facilitate large-scale reconstruction programs. Prior to the Canterbury Earthquakes in New Zealand, Feast (1995) had identified several issues in relation to planning and construction legislation that could impede reconstruction of Wellington, following a major earthquake. Feast's study suggested that

much of the legislation (in particular the Resource Management Act (RMA) and the Building Act (BA) that existed during the period was neither drafted to cope with an emergency situation nor developed to operate under the conditions that would prevail in the aftermath of a severe seismic event. Specifically commenting on the RMA, Feast had explained that its consultation procedure could prevent meeting the reconstruction requirements of a devastated city within a reasonable time period (Feast 1995). Decades after in 2012 (following the Canterbury earthquakes), some of Feast's suggestions were relived, and changes to both the RMA and BA were, and are, still being made based on recommendations by the Royal Commission set up to review the Canterbury recovery program. Changes around seismic codes, strengthening of existing buildings, changes to land use, and planning parameters are being made as a response to the Canterbury disaster.

Rolfe and Britton (1995) are of the opinion that the pace of reconstruction is severely impacted by political and cultural conflicts over recovery plans, which means the successful achievement of disaster management goals will depend on the political environment. Once a disaster happens, there is a continuous tension between strictly applying reconstruction regulations which aim at preventing a recurrence of the previous community vulnerabilities and allowing an affected community to move back to their former habitation quickly. Clearly, the quicker communities return to habitability of as many of their homes as possible, the better it will be for restoring a sense of normality. However, to reduce future vulnerabilities, it is rational to approach rebuilding programs cautiously to build in resilience and build back better than before the disaster. In New Zealand, the rebuilding of Napier following the 1931 earthquake did not reduce vulnerabilities to future disasters as the city was rebuilt in the same location mainly producing unreinforced masonry buildings. Retrofitting of Napier buildings was subsequently required under new seismic codes because, as has been repeatedly demonstrated, unreinforced masonry buildings are vulnerable to earthquakes.

When considering the effects of legislation changes after disasters, it can be seen that legislation changes post-disaster can be broadly classified into two categories: (1) legislation for compliance and (2) legislation for facilitation. *Legislation for compliance* entails using legislation to enforce recovery initiatives, such as those that would require buildings to be upgraded or rebuilt to different standards and codes. The lack of enforcement of hazard-related laws and adequate risk-based building controls contributed to the large-scale devastation caused by the 2004 Indian Ocean tsunami (Mulligan and Shaw 2007). The same was witnessed in countries such as Pakistan, Turkey, Samoa, and Haiti. In reconstruction programs, there is a need to enforce the new building design codes which inevitably appear after a disaster. Inspection of the reconstruction, including buildings using the new codes, requires clear guidelines to maintain the required standards.

Legislation for Facilitation denotes legislation being used to simplify and assist recovery activities to speed up the recovery process. These changes can usually be seen in changed processes, such as in building and development consent processing. However, legislation that is customarily used to impose security and safety controls (such as building consents) can become an obstacle to rebuilding programs. Time-consuming procedures, insufficient resources to process permits, and lack of fast-tracked methods could delay reconstruction. For example, delays in the issuance of permits were a major reason for the holdup in housing repair and rebuilding following the 2005 Bay of Plenty storm in New Zealand (Middleton 2008). Fast-tracked consenting procedures, collaboration between councils, open access to information between stakeholders, and using legislation to remove bureaucracy are viable options to speed up recovery. During the 1994 Northridge earthquake, legislation was suspended, and emergency powers were used with the consequence of reducing highway reconstruction time. Similarly, during recovery from the Australian bushfires, planning and building permits were exempted for temporary accommodation so they could be put up

quickly. For permanent dwellings, planning permits were exempted, and only building permits were needed after the bushfire (DPCD 2013). Also property falling under the Wildlife Management Overlay which would normally be subject to more rigorous planning and building permit requirements had to be relaxed to only require a simplified planning consent and building permit (DPCD 2013).

From the forgoing, it is clear that legislation and regulatory requirements can have significant influence on the rate of recovery after a disaster event. The overall desire is for legislation to enhance the recovery and reconstruction process so that it improves the functioning of an affected community and reduces risks from future events. Any legislation changes that need to be made post-disaster may be better considered before a disaster so that its implementation could be facilitated early on during recovery. Often the opportunities to introduce mitigating measures become limited over the course of recovery because of the desire to return to normalcy and thus rebuild quickly after disasters. Menoni (2001) notes that “market forces put pressures to reconstruct as quickly as possible . . . hampering efforts to implement lessons learnt from the disaster in the attempt to reduce pre-earthquake vulnerability.”

Pressures to rebuild critical infrastructure quickly are borne by national and local administration with the possible implication of reduced quality of delivery. Rushed rebuilding programs have led to increased vulnerability of poorly planned and designed built environments to future disasters (Jigyasu 2010). Ingram et al. (2006) also explain that the clamor to rebuild quickly amplifies the underlying social, economic, and environmental weaknesses that result in large-scale disasters (Ingram et al. 2006).

Improving Recovery Through Legislation

Well-articulated and implemented legislation should not only provide an effective means of reducing and containing vulnerabilities (disaster

mitigation) but also become a means of facilitating better thought out and designed reconstruction programs.

After damage assessments and evaluations, building and environmental legislation should not present impediments to reconstruction and rebuilding programs. Martín (2005) suggests that there is a relationship between building/environmental regulations and rehabilitation works (Martín 2005) and that regulations could become burdensome in rehabilitation and reconstruction projects and are worthy of considerations (Martín 2005). Martín (2005) describes burdensome regulations as those which incorporate excessive rules and regulations and red tapes (statutory procedures) that add unnecessarily to costs. Listokin and Hattis (2004) provide useful analysis on two kinds of barriers that building codes could pose to rehabilitation works. They are “hard” and “soft” barriers to rehabilitation. The hard barriers are impediments to rehabilitation as a result of overregulation, which would not add appreciably to building value or public safety (Burby et al. 2006) and could discourage housing development or rehabilitation because they are added burdens (May 2004). For instance, building and environmental regulations that do not reduce the vulnerability of built assets to a hazard event are unnecessary. Also to insist on expensive structural solutions in a highly hazardous zone, where a simple alternative will be to restrict development in that zone, is another example of regulation that could fall under Listokin and Hattis’ hard category.

Soft impediments, on the other hand, are administrative requirements that require extra time, money, and effort to accomplish rehabilitation and reconstruction projects (Listokin and Hattis 2004). These are red tapes (bureaucratic procedures) that could delay new construction and rehabilitation/reconstruction of physical facilities (May 2004). Such soft impediments are the focus in this entry.

Bureaucratic procedures must be supportive of emergency management under different emergency scenarios whether routine or chaotic. However, research suggests that bureaucracies have been less supportive of the

expediency that is desired in disaster response and recovery. May (2004) suggests three sources of impediments by regulatory processes:

- Regulatory approvals. These are delays associated with consent processes and approvals that arise from cumbersome decision-making processes and duplication of regulations. These types of delays are inherent in building and environmental legislation.
- Regulatory enforcement strategies and practices. These are overly rigid practices that foster an unsupportive regulatory environment for the development and rehabilitation of the built stock. In post-disaster situations, rigid enforcement strategies discourage genuine recovery efforts.
- Patchwork of administrative arrangements. This could result from duplication of administrative structures (as in layers or hierarchies of control) and gaps in regulatory decision processes. May (2004) explains that patchwork frustrates regulatory implementation and adds to complexities in regulatory processes.

Regulatory process barriers could also result from *administrative conflicts* in and among disaster agencies (Listokin and Hattis 2004). For example, rivalry between responding agencies is not foreign to emergency services and is an obstacle to effective emergency management (McEntire 2002; Quarantelli 1998). Rotimi (2010) explains that rivalry may result from existing silos, from the absence of a coordinating agency, and from the ability of a coordinating agency incorporating other agencies perceived responsibilities. Hence, a broad cooperative effort is needed for the success of post-disaster reconstruction activities. Rotimi (2010) therefore contends that organizations have to coalesce to plan for resource utilization in the restoration of physical assets. Coordination is therefore central to multi-organizational response and recovery programs (McEntire 2002). This multi-organizational coordination is often embedded into emergency legislation post-disaster and usually involves the creation of new recovery coordinating agencies such as Victorian

Bushfire Reconstruction and Recovery Authority (VBRRA) in Australia or Canterbury Earthquake Recovery Authority (CERA) in New Zealand.

Another useful dimension to the problems with burdensome regulations is provided by Listokin and Hattis (2004). It is that regulatory procedures could become too rigid forcing implementers to “go by the book” even though variations may be warranted. This places implementers in a state of continuous fear of liability should things go wrong. Rotimi (2010) explains that some latitude of control and discretion is often required to aid decision-making as long as such decisions are pragmatic. Commenting on the rebuilding program after the flooding incident in New Orleans, Stackhouse (2006) says “removing democratic processes from the rebuilding process has the advantage of expediting decision making by allowing politically dangerous but practical outcomes.” This statement suggests that greater freedom in decision-making by officials of coordinating agencies could increase the speed of rebuilding programs after significant disaster events. Evidence from literature on recovery often shows slow recovery. For example, on the Bay of Plenty storm in New Zealand in 2005, Middleton (2008) showed that at 300 days after the event, 35 households still required permanent rehousing out of a total of 300 compulsory evacuations. At the same time, nine households were still occupying temporary accommodation. Middleton (2008) suggests that the slow pace of recovery was attributable to the inadequacy of personnel to carry out the stipulated building safety evaluations and to process consents for reconstruction work. Both of these factors have legislative connotations.

Considering New Zealand before the Canterbury earthquakes, there was an emphasis on readiness and response activities, with little consideration given to planning for sustained recovery activities. Where recovery was considered, it was for the short term, as evident in emergency awareness campaigns that encouraged communities to prepare for 3–7 days after an event. Recent emergency events clearly show that longer-term recovery plans and more robust legislation were required given the complexities

associated with the rebuilding of damaged built assets. Warnings about the inadequacies of routine construction processes being modified on an ad hoc basis during the small disaster recovery phases in previous hazard events were unheeded in New Zealand (Rotimi et al. 2006). While such an approach works reasonably well for small-scale emergencies, the effectiveness of reconstruction in a large disaster required large-scale legislative changes (Rotimi et al. 2009). Post-Canterbury, an extensive review of legislation to cope with future disasters is now underway.

Post-disaster Legislation Changes in Action: Three Case Studies

Despite the type of disaster, there are often commonalities in the post-disaster legislative changes enacted and the impacts of these changes. Starting with an earthquake, and ending with a fire, this section demonstrates through three cases, the impact of legislation on recovery and some good and bad practices.

The Northridge Earthquake, USA, 1994

The Northridge earthquake provides an example of a disaster situation where legislative changes helped to facilitate reconstruction projects. The moderate earthquake struck Southern California in the early hours of 17 January 1994 with a magnitude of 6.8 on the Richter scale, small compared to other earthquakes, but causing significant damage.

Comerio (1996) gave an insight into the extent of damage. There were damage to 27 bridges and a collapse of sections of six freeways. Four hundred and fifty public buildings suffered significant damage, 6,000 commercial buildings, 49,000 housing units in 10,200 buildings had serious structural damages, while 388,000 housing units in 85,000 buildings experienced minor damages. The total value of damage to houses in Los Angeles was estimated to be about \$1.5 billion.

The Northridge earthquake caused a shift in emphasis from disaster preparation and relief to

recovery (Comerio 1996), and this shift largely resulted in the success of emergency management programs for the restoration of the affected areas. Reconstruction contributed to the economic revitalization of the affected area, and bureaucratic requirements were suspended to encourage rapid rebuilding of damaged infrastructure. Marano and Fraser (2006) conclude that “identifying and easing regulations and statutes that inhibit reconstruction can mean a dramatically faster and less costly recovery.”

Wu and Lindell (2004) provide an insight into some of the actions that were taken to increase the speed of housing reconstruction in Los Angeles after the earthquake showing that it is possible to expedite procedural requirements by establishing fast-tracked processes that would operate after a disaster would benefit recovery.

While the rapid recovery experienced after the Northridge earthquake can be attributed to other factors, such as political will, public policy changes and enabling emergency management legislation played a substantial role in the rebuilding programs after the earthquake (Comerio 2004).

Hurricane Katrina, New Orleans, USA, 2005

The disaster that followed Hurricane Katrina provides lessons on legislative changes either proposed or already implemented to enable its recovery from the event. The Hurricane was a category 3 storm when it struck New Orleans and the Gulf Coast in the morning of August 29, 2005. The storm surge caused severe destruction along the Gulf Coast from central Florida to Texas in the USA. The most severe damage occurred in New Orleans, Louisiana, because of the failure of the levee system that was designed to contain the resulting storm surges. The worst damage was caused by floods with an estimated 1.2 million people evacuated before the incident and another 100–120 thousand afterward. About 350,000 houses were destroyed and over 200,000 persons required temporary shelters scattered around 16 states in the USA. By all accounts, Hurricane Katrina was a catastrophic event with economic loss estimates of about \$200 billion (Burby et al. 2006).

A brief description of some of the policy changes and legislative reviews that occurred as a result of Katrina is provided by Rotimi (2010) including:

1. Changes in building codes and standards. There were changes made to the building codes in New Orleans with a view in improving the resilience of built spaces in New Orleans. For example, there were revisions made to the base flood elevation levels for new construction to 3 ft or higher (Colton et al. 2008). This is a risk mitigation strategy which has been tied to flood insurance cover so that only buildings that meet these new guidelines can qualify for flood insurance and subsequent compensations. Overall, funding sources and budget priorities have been developed for reconstructing flood protection in New Orleans (Colton et al. 2008).
2. Changes in emergency management regulations and guidelines. Colten et al. (2008) explain that the Katrina event necessitated the review and updating of Louisiana and New Orleans' response strategies and their emergency operations plans. The legislative reviews included the adoption of an all-hazards approach thus expanding the scope and magnitude of anticipated hazards and allowing greater involvement of non-agency actors who proved crucial to response and recovery after the event. Colton et al. (2008) noted that partnering with nongovernmental stakeholders was a paradigm shift that emerged out of the Katrina experience.
3. Changes in land development regulations. Changes to land and development regulations are largely seen as a veritable tool for mitigating disaster risk in disaster management (Ingram et al. 2006). After Hurricane Katrina, changes in land use planning and zoning systems have been proposed to reduce the vulnerability of the New Orleans region from future flooding disasters (Burby et al. 2006).

The legislative changes in New Orleans show the importance placed on built asset

reinstatements as a major input to holistic recovery. Similar to Northridge, building reconstruction in New Orleans aimed to stimulate development and growth. The rate at which recovery is achieved is therefore tied to the speed of reconstruction, and any recovery is underpinned by well-thought-out and implemented legislative and regulatory changes.

Victorian Bushfires, Australia, 2009

The "Black Saturday" disaster was one of the most damaging disasters in Australian history, leaving 173 people dead and many seriously injured, affecting 6,000 households, destroying more than 2,000 homes, and damaging around 430,000 ha of land. Despite the institutions and procedures set up for expediting community recovery, reconstruction proceeded slowly. Shortly after the bushfires in March 2009, the Victorian Government introduced a new residential bushfire construction building standard AS3959-2009 in response to the need to better protect the bushfire-affected communities from future fire events. Under the new codes, there are increased construction requirements depending where a building is situated, ranging from ember protection to direct flame contact protection.

Changes have been documented by Mannakkara and Wilkinson (2013) which showed that one of the first steps taken in Australia was to publish a revised edition of the Australian building code for bushfire-prone areas (AS 3959) on March 11, 2009 (VBBRA 2009). The revisions introduced Bushfire Attack Levels (BAL) to identify the bushfire risk of properties. Stringent design and construction requirements were specified for each BAL to provide greater fire protection. Another key change in legislation was regarding land use. Soon after the fires, the entire state of Victoria was declared bushfire prone and placed under the Wildfire Management Overlay (WMO) which imposed stricter planning regulations (VBBRA 2009). The introduction of the buy-back scheme posed a solution for people on high-risk lands (Victorian Government 2012).

As with earthquake recovery, constructing bushfire-resistant buildings contributes to saving

lives and properties in the event of a bushfire. Mannakkara and Wilkinson (2013) report on the bushfire legislative and code change impacts. The more stringent construction requirements stipulated in the new Australian building standards are mainly concerned with the use of noncombustible materials for housing reconstruction. Bushfire housing reconstruction proceeded slowly because the advanced resources for direct flame contact protection such as the window systems, roof systems, shutters, and external cladding materials as required in the new AS3959-2009 were not available in the market. The main reason for the unavailability of these resources was that it required a considerable amount of time for manufacturers to undertake the research and development needed to test and release these new materials onto the market. Only in March 2010, a year after the bushfires, for instance, a combined window and screen system manufactured for use in direct flame contact protection zones was ready for release onto the market. The delays with production of compliant materials for the bushfire recovery, combined with growing demand on building services in the local construction market, created a series of scarcities, which greatly hindered housing recovery in the fire-affected areas. During the bushfire recovery, there were extra costs for the construction requirements of bushfire houses; these extra costs were significantly underrepresented (i.e., officially given assessments were from AUD 10,000 to 40,000 depending on the protection levels required, whereas the extra real cost of a house to be rebuilt to new codes was somewhere up to AUD 100,000). The increase in costs caused financial pressure for the affected house owners who already struggled to procure suitable resources for rebuilding their houses. The uncertainty about the number of houses in the direct flame contact protection zone was another major concern for the building product manufacturers.

Given few incentives from the government and the low likelihood of profitability, material producers were reluctant to put effort into developing materials for houses in the direct flame zone, which they believed would account for

only a small fraction of their market. Lack of training and understanding of the new standards introduced in Australia slowed recovery; even one and a half years after the bushfires, designers and builders were still trying to come to terms with the application of the standard.

General Implications of Legislation on Recovery

Having highlighted some of the issues that are connected to the appropriateness of legislation and regulatory provisions in the previous sections, this section presents a summary of their implication on recovery. The summary is in line with Rotimi (2010) who describes the effect that poor legislative provisions could have on post-disaster reconstruction activities, thus:

1. Loss of vital momentum of action. The efficiency of post-disaster reconstruction activities is impacted as a result of delays caused by poor planning and implementation, restrictive legislation and regulatory provisions, and lack of government commitment in reconstruction programs.
2. Loss of commitment to the reconstruction process. There is a tendency for poor commitment to recovery programs by responsible authority because disaster practitioners are unable to apply pragmatic solutions to real-time reconstruction problems for fear of being held liable for their decisions.
3. Difficulties in achieving reconstruction deliverables and inability to accelerate the process of reinstatement. Introduce measures for risk and vulnerability reduction and aid planning for sustainable developments.
4. Impairment of overall community recovery and quality of life. Of essence, reconstruction should become a tool for empowerment till a level of functioning is reached where communities are self-sustaining and require no external interventions and also a therapeutic process for overall community recovery.

Summary

This entry has shown common practices relating to changes in legislation for disaster recovery. Changing legislation after a disaster is a common response to the disasters but takes on different forms. Most disasters are followed by legislative changes, emergency legislation, and new disaster legislative, and these measures aim to encourage better building practices and future resilience. Putting in place legislation which enforces revised building codes to increase disaster resilience in the built environment is recommended as is that which speeds up slow processes or offers flexibility. Putting in place rigid rules and bureaucracies slows recovery and creates confusion. As cautioned by Ingram et al. (2006), legislative and regulatory changes need scrutiny to avoid issues such as resource constraints, high costs, and impacts on livelihood that can unnecessarily hinder recovery progress.

Cross-References

- “Build Back Better” Principles for Reconstruction
- Economic Recovery Following Earthquakes Disasters
- Reconstruction Following Earthquake Disasters
- Reconstruction in Indonesia Post-2004 Tsunami: Lessons Learnt
- Resilience to Earthquake Disasters
- Resourcing Issues Following Earthquake Disaster

References

- Burby RJ, Salvesen D, Creed M (2006) Encouraging residential rehabilitation with building codes: New Jersey experience. *J Am Plann Assoc* 72(2):183–196
- Colton CE, Kates RW, Laska SB (2008) Community resilience: lessons from New Orleans and Hurricane Katrina, CARRI research report 3, Community and Regional Resilience Initiative of the Oak Ridge National Laboratory. www.resilientUS.org
- Comerio MC (2004) Public policy for reducing earthquake risks: a US perspective. *Build Res Inform* 32(5):403–413
- Comerio MC (1996) The impact of housing losses in the northridge earthquake: Recovery and Reconstruction issues, Center for environmental design research, University of California, US
- DPCD (2013) List of amendments to the victoria planning provisions. In: Department of Planning and Community Development (ed). *A Guide to the Planning System* Victoria State Government, Victoria
- Feast J (1995) Current planning and construction law: the practical consequences for rebuilding Wellington after the quake. Paper presented at the Wellington after the Quake: the challenges of rebuilding cities, Wellington
- Ingram JC, Franco G, del Rio CR, Khazai B (2006) Post-disaster recovery dilemmas: challenges in balancing short-term and long-term needs for vulnerability reduction. *Environ Sci Policy* 9(7–8):607–613
- Jigyasu R (2010) Appropriate technology for reconstruction. In: Lizarralde G, Davidson C, Johnson C (eds) *Rebuilding after disasters: from emergency to sustainability*. Taylor and Francis, London
- Listokin D, Hattis D (2004) Building codes and housing. Paper presented at the workshop on regulatory barriers to affordable housing, US Department of Housing and Urban Development, Washington, DC
- Mannakkara S, Wilkinson S (2013) The impact of post-disaster legislative and regulatory changes on the recovery of the built environment. The Society of Construction Law Australia, Sydney
- Marano N, Fraser AA (2006) Speeding reconstruction by cutting red tape. Retrieved 20 Oct 2006 from www.heritage.org/
- Martín C (2005) Response to “building codes and housing” by David Listokin and David B. Hattis. *Cityscape J Policy Devel Res* 8(1):253–259
- May PJ (2004) Regulatory implementation: examining barriers from regulatory processes. Paper presented at the workshop on regulatory barriers to affordable housing, US Department of Housing and Urban Development from www.2004nationalconference.com
- McEntire DA (2002) Coordinating multi-organisational responses to disaster: lessons from the March 28, 2000, Fort Worth tornado. *Disaster Prev Manag* 11(5):369–379
- Menoni S (2001) Chains of damages and failures in a metropolitan environment: some observations on the Kobe earthquake in 1995. *J Hazard Mater* 86(1–3):101–119
- Middleton D (2008) Habitability of homes after a disaster. Paper presented at the 4th international i-REC conference on building resilience: achieving effective post-disaster reconstruction, Christchurch New Zealand
- Mulligan M, Shaw J (2007) What the world can learn from Sri Lanka’s Post-Tsunami experiences. *Int J Asia Pac Stud* 3(2):65–91
- Quarantelli EL (1998) *What is a disaster? Perspectives on the question*. London: Routledge

- Rolfe J, Britton NR (1995) Organisation, government and legislation: who coordinates recovery? Paper presented at the Wellington after the quake: the challenge of rebuilding cities, Wellington
- Rotimi JOB (2010) An examination of improvements required to legislative provisions for post disaster reconstruction in New Zealand. Doctoral thesis, University of Canterbury, Christchurch
- Rotimi JOB, Le Masurier J, Wilkinson S (2006) The regulatory framework for effective post-disaster reconstruction in New Zealand. Paper presented at the 3rd international conference on post-disaster reconstruction: meeting stakeholder interests, Florence
- Rotimi JO, Myburgh D, Wilkinson S, Zuo K (2009) Legislation for effective post-disaster reconstruction. *Int J Strateg Prop Manag* 13(2):143
- VBBRA (2009) 100 day report. Victorian Bushfire Reconstruction and Recovery Authority, Victorian Bushfire Reconstruction and Recovery Authority, Melbourne
- Victorian Government (2012) Victorian bushfire recovery three year report. Victorian Government, Melbourne
- Wu J, Lindell MK (2004) Housing reconstruction after two major earthquakes: the 1994 Northridge earthquakes in the U.S. and the Chi-Chi earthquake in Taiwan. *Disasters* 28(1):63–81

employed. In order to implement a successful remediation work for a target structure, a thorough understanding of the following are required: the liquefaction hazard at the site, the potential consequences of liquefaction for the structure, the performance requirements of the work, and the available construction materials and methods. Several alternative approaches can be taken if liquefaction poses as threat to existing or proposed structures. For existing structures, the choices are:

- (1) Retrofitting the structure and/or site to reduce the potential failure
- (2) Abandoning the structure if the retrofit costs exceed the potential benefits derived from maintaining the structure
- (3) Accepting the risk and maintaining the existing use

Additional options would be to continue the use of the structure but to change the operation or to alter the use of the structure such that the hazard becomes tolerable even if failure occurs.

For new construction, moving the project assumes that there are alternative sites, where liquefaction hazard is not a concern, but is equally appropriate for the proposed project. The costs of necessary mitigation actions may, in some cases, make selection of an alternative site a more cost-effective alternative than the use of the primary site. The choice of accepting the risk associated with the hazard, and the potential loss of life, social disruption, economic loss, and political ramifications that might result from structural failure, must be based on engineering evaluation of the site.

Thus, the overall goal of any remediation work is to investigate alternatives to mitigate the liquefaction hazard so that relative benefits versus costs for different levels of mitigation efforts can be assessed. In investigating potential remedial measures, the following factors should be considered:

- Technical adequacy
- Field verifiability
- Costs

Liquefaction: Countermeasures to Mitigate Risk

Rolando P. Orense
Department of Civil and Environmental
Engineering, University of Auckland,
Auckland, New Zealand

Synonyms

Ground improvement; Liquefaction mitigation; Liquefaction prevention; Liquefaction remediation

Introduction

Past several earthquakes have vividly demonstrated the impact of soil liquefaction and the associated ground deformations to buildings, bridges, buried lifelines, and other civil infrastructure. To reduce the risk of damage to the structures, remediation measures are usually

- Maintenance
- Long-term performance
- Environmental impact

In this section, general descriptions of the principles involved in ground improvement methods, i.e., densification, solidification, drainage, reinforcement, containment, and soil replacement, as well as some representative methods, are presented. Verification techniques and design issues associated with general ground improvement methods are also explained.

General Description of Countermeasures

Countermeasures to mitigate the damaging effects of liquefaction can be classified into two categories: (1) ground improvement (to remediate liquefiable soils) and (2) structural enhancement (to relieve the effects of liquefaction). Some projects may require the combination of the two methods.

Ground improvement, whose goal is to limit soil displacements and settlements to acceptable levels, is being used increasingly for remediating liquefiable soils due to the wide variety of methods that are available. These methods can be modified and customized to specific site settings, including space/operation restrictions and the presence of existing structures. In addition, one or more methods can often provide economical solutions for liquefaction remediation problems.

Ground improvement methods are generally based on one or more of the following principles:

- Densification
- Solidification
- Drainage
- Reinforcement and containment
- Soil replacement

Descriptions of the principles behind the improvement mechanism associated with each category are provided below, along with some particular treatment techniques in the category

that can potentially be used at existing structures and for future sites. In addition, some design considerations are also briefly described. Details concerning the specific treatment methods can be found in references on ground improvement (e.g., Cooke and Mitchell 1999; Andrus and Chung 1995; Xanthakos et al. 1994; Hausmann 1990; JGS 1998; PHRI 1997). Note that there is no attempt herein to provide a comprehensive discussion of all available liquefaction countermeasures; rather, only representative methods are discussed for each improvement mechanism. Moreover, new techniques are being developed when difficult circumstances are encountered in the field, while available techniques are being refined to suit the requirement of the problem. Thus, more and more ground improvement techniques are becoming part of the expanding set of available mitigation measures.

Structural enhancement can reduce the damage to the structure while allowing liquefaction to occur. Examples include strengthening the foundations of the structures and the ground supporting the structures to avoid reduction of bearing capacity or possible uplift due to soil liquefaction. The use of flexible structures, which can deform together with the ground deformation, is a viable option for underground structures.

It should be mentioned that mitigation of liquefaction risk is an area subject to considerable controversy and that the current understanding of the efficacy of some of these methods is still evolving (Seed et al. 2003). Moreover, while the case histories clearly indicate that ground improvement leads to a significant decrease and/or elimination of large liquefaction-induced ground deformations, the data is not yet sufficient to be able to predict the ground deformations for a given set of site conditions and earthquake motion (Hausler and Sitar 2001). Furthermore, large earthquakes which can verify the effectiveness of these techniques are infrequent and current database of observed field performance has marginal quality. Fortunately, this is being addressed by more detailed studies of the field case histories (e.g., Yasuda et al. 1996; Ishii et al. 2013) coupled with carefully designed experimental testing and numerical analyses.

Densification Methods

Densification or compaction methods involve rearranging the soil particles into tighter configuration, resulting in increased density. This increases the shear strength and liquefaction resistance of the soil and encourages a dilative instead of a contractive dynamic soil response. Densifying loose sandy deposit with vibration and/or impact has been used extensively, making it the most popular liquefaction countermeasure.

An increase in soil density can be achieved through a variety of means (JGS 1998). These include (1) compaction by penetration (penetration of sandy material into the liquefiable deposit will push the surrounding soil and result in reduced void ratio and therefore increase in soil strength), (2) compaction by vibration (subjecting the loose sandy deposit to vibration energy will compact the soil and increase its strength), and (3) compaction by impact energy (impact energy can densify loose granular deposit).

The suite of methods which involve compaction by displacement and/or vibration involves a vibrating probe repetitively penetrating the ground to densify the liquefiable soils and is generally referred to as *vibromethod*. These methods involve various kinds of equipment and procedures and are known by some generic or proprietary names. The simplest categorization of vibromethods is as follows:

- Vibro-replacement: the cavity formed by probe penetration is filled with imported materials, such as crushed stone gravel and sand.
- Vibro-flotation: the surrounding soil is compacted with a horizontally vibrating motor attached to the end of the probe, sometimes aided by water sprayed from the nozzle tip end.
- Vibro-rod: compaction is achieved using vertical vibratory penetration to the top of the probe or rod.
- Vibratory tamper method: a rigid steel plate (tamper) attached to a vibrator at the surface of the ground to compact the surface soil.

Vibro-replacement methods can improve liquefiable soil deposits not only by densification

but also by increasing the in situ lateral stress, replacing the liquefiable in situ soil with non-liquefiable material, reinforcing the original ground with stiffer columns of fill materials, and providing drainage paths for excess pore water pressure. Two of the most widely implemented methods are the sand compaction pile (SCP) method (popular in Japan) and stone column (SC) method (common in the United States). Sand compaction piles are constructed by driving a sand-filled steel casing into the ground after which the casing is then slowly withdrawn and the deposited sand is compacted to form a high density sand column (Schaefer 1997; JGS 1998). The sand backfill must have specific properties and sometimes may be difficult to source. Recent case histories indicate that metal slag, crushed oyster shell, granulated fly ash, or recycled crushed concrete and asphalt are feasible alternatives to sand (Kitazume 2005). A variation on the sand compaction pile technique is the use of gravel backfill, to create stone columns (Munfakh et al. 1987; Sondermann and Wehr 2004). Generally the fill material consists of crushed coarse aggregates of various sizes, with the ratio in which the stones of different sizes will be mixed decided by design criteria.

Vibro-flotation method compacts loose soil by inserting a horizontally vibrating vibroflot into the ground while jetting water from the bottom end of the vibroflot, followed by gradually extracting it while forcing sand or gravel from the ground surface into the voids created adjacent to the vibroflot. On the other hand, vibro-rod method involves driving a vibrating rod or pile into the liquefiable soil, thereby causing it to locally and temporarily liquefy, after which new sand is placed along the rod or pile to densify the in situ soil. In vibro-flotation and vibro-probe methods, sand or gravel fill is injected from the ground surface into the voids created in the original ground during vibration.

Dynamic compaction involves repetitively dropping a large weight from a significant height onto the ground causing the soil grains to rearrange and form a denser arrangement. Additionally, the impact of the dropped weight on the ground surface produces dynamic stress waves,

which can be large enough to generate significant excess pore water pressure in the soils beneath the point of impact (Idriss and Boulanger 2008). The dissipation of these excess pore water pressures results in densification, accompanied by surface settlement. The drop height, weight, and spacing vary, depending on ground and groundwater conditions. Dynamic compaction is known to be fast and economic, especially in treating large areas. However, it has obvious disadvantages due to the noise and vibration that are produced. Moreover, it is often effective only in the upper 10 m of the deposit and it is less effective for soils with high fines content.

In compaction grouting, a very stiff grout is injected into the soil such that it does not permeate the native soil, but results in coordinated growth of the bulb-shaped grout that pushes and displaces the surrounding soil. Typically the grout consists of soil–cement–water mixture with sufficient silt sizes to provide plasticity, together with sand and gravel sizes to develop internal friction (Welsh 1992). Note that the strength of the grout is unimportant because the purpose of the technique is to densify the surrounding soil by displacement. Since the technique involves the pressurized injection of grout into the soil deposit using small-scale, maneuverable, and vibration-free equipment, there is minimum disturbance to the structure and surrounding ground during implementation. Additional advantages include relatively little site disruption and the ability to work in constricted space, resulting in greater economy. However, it has some disadvantages; for example, stabilization of near-surface soils is generally ineffective due to the fact that the overlying restraint is small and the grouting pressures can heave the ground surface rather than densify the soil.

Compaction by man-made explosions can also be employed to densify loose liquefiable deposits. By detonating explosive charges placed at various depths in boreholes across the site in a controlled manner, the explosions propagate dynamic shear stresses through the ground which can induce liquefaction. The post-liquefaction consolidation following the

dissipation of excess pore water pressure results in densification of the soil, together with ground settlement (Narin van Court and Mitchell 1994; Maeda et al. 2006). This method is effective in sites where cleaner loose sands, which will undergo relatively large and rapid reconsolidation, are present at greater depths. Conversely, it may be less effective in densifying shallow deposits.

In general, densification or compaction methods are more suitable for use in saturated, cohesionless soils with a limited percentage of fines. They cause noise and vibration during installation and also increase horizontal earth pressures against adjacent structures. This increase in pressure is the major disadvantage of using compaction methods in close proximity to retaining walls and pile foundations. The major advantage of compaction methods is the relatively low cost/benefit ratio. The necessary degree of compaction can be evaluated using penetration resistances that have been back-calculated from an acceptable factor of safety against liquefaction.

Solidification Methods

Increase in liquefaction resistance can also be obtained through a more stable skeleton of soil particles. This can be achieved by solidification methods, which involve filling the voids with cementitious materials resulting in the soil particles being bound together. This will prevent the development of excess pore water pressure, preventing the occurrence of liquefaction. Typical methods include deep mixing method, premixing method, and grouting techniques.

The deep mixing method achieves liquefaction remediation by agitating and mixing stabilizing material such as cement in sandy soil and solidifying the soil. The soil–cement materials in the mixed columns can have a wide range of unconfined compressive strengths, depending on the amount of cementitious material used and the in situ soil characteristics (Kitazume and Terashi 2013). This technique can be applied over a whole area of liquefiable soil (block type) or over partial sections (wall type, pile type, or lattice type). In block type, the mix columns

overlap each other, and the whole layer is improved due to the increased stiffness. In partial improvement, unimproved sections remain and may liquefy; however, depending on the shape of installation, the mix columns are effective in preventing liquefaction by restraining the shear deformation of the ground through reinforcement and confinement, as described below.

The premixing method, on the other hand, is done by adding stabilizing material to the soil in advance and placing the treated soil at the site (JGS 1998). This is recommended for constructing new reclaimed land or in filling behind caissons, allowing both reclamation work and anti-liquefaction measures to be implemented at the same time to develop stable ground.

Chemical grouting, sometimes called permeation grouting, is a technique that transforms granular soils into hardened soil mass by injecting cement or other grouting materials that permeate and fill the pore space. The hardened grout improves the native soil by cementing the soil particles together and filling the voids in between (minimizing the tendency of the soil to contract during shearing). The treated soil would have increased stiffness and strength and decreased permeability. Because of its minimal disturbance to the in situ soil, it is an effective method in treating liquefiable deposits adjacent to existing foundations or buried structures.

In jet grouting, high-pressure jets of air, water, and/or grout are injected into the native soil in order to break up and loosen the ground and mix it with thin slurry of cementitious materials. In essence, it is not truly grouting but rather a mix-in-place technique to produce a soil-cement material. Depending on the application and soils to be improved, different kinds of jet are combined by using single fluid system (slurry grout jet), double fluid system (slurry grout jet surrounded by an air jet), and triple fluid system (water jet surrounded by an air jet, with a lower grout jet). The process can construct grout panels, full columns, or anything in between (partial columns) with a specified strength and permeability.

Solidification techniques can generally be used in a wide range of soil types, including those with high fines content. They are

advantageous because the installation methods are relatively quiet and induce relatively small vibrations as compared to compaction methods. The induced horizontal earth pressures are smaller than with compaction methods and are larger than with drainage methods. Their disadvantage is the relatively high cost/benefit ratio as compared to compaction and drainage methods.

Drainage Methods

Excess pore water pressure generated by cyclic loading can be reduced and/or dissipated by installing permeable drain within the deposit. The methods employed can be divided into two categories depending on the material used as drain: grave drain and artificial drain methods. These methods rely on two mechanisms to reduce damage due to liquefaction: (1) delaying the development of excess pore water pressure due to earthquake shaking and (2) preventing the migration of high excess pore water pressure from untreated liquefied zones into non-liquefied areas (say underneath the structure) to prevent secondary liquefaction caused by pore water pressure redistribution.

Gravel drains are typically installed either as column-like drains in a closely spaced grid pattern or as backfill around underground structures to control the levels of maximum excess pore water pressure ratio during earthquake shaking. They can also be installed as wall-like or column-like perimeter drains at both sides of densified (treated) zones or to isolate the high excess pore water pressure from liquefied areas. In the installation of gravel drains, casing with an auger inside is drilled into the ground down to the specified depth. Crushed stone is then discharged into the casing and the gravel drain is formed by lifting the casing pipe.

Artificial drains can be made of geosynthetic composites or piles with drainage functions. They have not only drainage effect but also reinforcing effect. For example, plastic drain consists of a plastic perforated pipe drainage skeleton wrapped in geofabric to prevent clogging from soil particles. These can be easily installed; however, close spacing is usually required due to the limited capacity of each drain.

Design charts for drains were initially developed by Seed and Booker (1977) to control the maximum excess pore water pressure levels, but more recent design charts and analytical methods (e.g., Iai and Koizumi 1986; Pestana et al. 1997) provide better methods of taking into account various factors affecting the drain performance, such as the hydraulic properties of the drain and permeability and volumetric compressibility of the native soil.

Drainage remediation methods are suitable for use in sands, silts, or clays. One of the greatest advantages of drains is that they induce relatively small horizontal earth pressures during installation. Therefore, they are suitable for use adjacent to sensitive structures. In the design of drains, it is necessary to select a suitable drain material that has a coefficient of permeability substantially larger than the in situ soils. Installation can be carried out with low noise and vibration. However, since the in situ soils improved by this method remain in a loose condition, the method has obvious disadvantages when compared to compacted deposits, such as negligible ductility and significant residual settlement of the treated soils. It is effective only if it successfully promotes sufficiently rapid dissipation of pore pressures as to prevent the occurrence of liquefaction; if pore pressure dissipation is not sufficiently rapid during the relatively few critical seconds of the earthquake, this method does relatively little to improve post-liquefaction performance (Seed et al. 2003). Thus, the method is usually combined with densification method, i.e., the surrounding ground is compacted to some extent during the drain installation.

Replacement Methods

The remove and replace method is another liquefaction mitigation technique. This involves the removal of the in situ liquefiable material and replacement with a non-liquefiable material such as clay, gravel, or dense sand. Its advantage is that it provides high degree of confidence in the final product and uses construction equipment and practices that are widely available and easily tested (Idriss and Boulanger 2008). However, this method is only feasible when relatively small

volumes of liquefiable material are involved at shallow depths.

Reinforcement and Containment Methods

When saturated sand deposits are sheared during seismic loading, excess pore water pressure is generated. To prevent the occurrence of liquefaction, the soil should undergo smaller shear deformation during earthquake. This can be achieved by underground diaphragm walls, sheet piles, or lattice-shaped walls using deep mixing techniques. This method simply surrounds the liquefiable soil with continuous underground wall, and thus, it can be used for the soil at the bottom of existing structures where liquefaction remediation is difficult with other methods. The shear deformation in the ground is reduced during an earthquake to mitigate liquefaction and at the same time provide support to the overlying structure. It can also be used to prevent excessive deformation of structures by restraining the lateral flow of ground after liquefaction.

The presence of stone columns, timber piles, and other stiff elements provide additional shear resistance against earthquake shaking. These reinforcements would take much of the earthquake load, thereby minimizing the shear deformation of the soil.

The key advantages of the method include (1) less vibration and noise, (2) applicable to existing structures, (3) no restriction to soil type and depth, and (4) effective use of land due to smaller treatment area. On the other hand, disadvantages of the method are the (1) lack of simplified methods to evaluate effectiveness and (2) generally higher construction costs.

Increasing In Situ Stress

In situ effective stresses within the soil mass can be increased, resulting in an increase in shear resistance. Pre-loading involves overconsolidating the soil with surcharge embankment by initially inducing increased vertical and horizontal effective stresses, and when the surcharge is removed, the resulting overconsolidation leaves the in situ soil somewhat more resistant against liquefaction. This method is applicable to soils with large fines content and can also be applied

even below existing structures since the soil can be overconsolidated.

Lowering the groundwater table makes the soil above the water table unsaturated, resulting in increased effective stress for soils below the water table. In addition, if the water table is reduced to a level below the liquefiable soil layer, liquefaction is prevented because the absence of water makes the buildup of excess pore water pressure impossible (Cox and Griffiths 2010). For this method, it is always necessary to maintain the lower groundwater level, and this may result in higher operating cost and maintenance cost associated with continuous pumping. Thus, this technique is typically limited to short-term applications.

Structural Strengthening Techniques

There are many types of measures which can be adopted to minimize damage to structures while allowing the occurrence of liquefaction itself. The selection of the appropriate remedial measure depends on the type of structure and the results of assessment of the required stability of the structure considering the effect of liquefaction, e.g., loss of bearing capacity, increase in earth pressure, etc. Several structural strengthening techniques implemented to various structures are described below.

Deep foundations, such as piles or piers whose lengths extend below the occurrence of liquefaction (or significant cyclic softening due to partial liquefaction), can provide reliable vertical support and so can reduce or eliminate the risk of unacceptable liquefaction-induced settlements. However, pile or pier foundations do not necessarily prevent damages that may occur as a result of differential lateral structural displacements, so piles and/or piers must be coupled with sufficient lateral structural connectivity at the foundation as to safely resist unacceptable differential lateral displacements. Moreover, pile foundations are likely to be damaged by lateral forces caused by a decrease in subgrade reaction accompanying liquefaction and associated lateral spreading of liquefied soil. To reduce the damage from such ground deformations, inclined piles or large-diameter vertical piles have frequently been

used, or the number of piles has been increased. Moreover, significant research efforts over the past 20 years have led to the development of a range of methods to address the effect of laterally spreading ground on pile foundations, ranging from simplified pseudo-dynamic approach (e.g., Hamada et al. 1986; Cubrinovski et al. 2012) to fully nonlinear, time-domain, fully coupled effective stress analysis of soil/pile/superstructure interaction analyses (e.g., Iai et al. 1998; Bowen and Cubrinovski 2008). These types of methods, complemented with appropriate conservatism, can provide a suitable basis for analysis of this issue and for the design and detailing of piles (or piers) and pile/cap connections.

In structures with spread foundations, differential settlement and cracks in foundation can occur as a result of loss in bearing capacity of foundation ground following liquefaction. To address this, spread foundations can be supported by piles. In addition, stiff, reinforced shallow foundations are used to resist both differential lateral and vertical displacements. Japanese practice has increasingly employed both grade beams and continuous reinforced foundations for low to moderate height structures, and performance of these types of systems in earthquakes has been good. Note that although stiff, shallow foundations can undergo differential settlements resulting in rotational tilting of the structure; these can be easily re-leveled after earthquake-induced settlements through several techniques, including jacking, under-excavation, or compaction grouting.

Quay walls can become unstable because of increase in active earth pressure and reduction in passive earth pressure. To prevent damage to such waterfront structures, strong quay walls such as pile-supported quay walls have been employed.

Buried pipelines may be bent, compressed, or extended as a result of large-ground displacement induced by liquefaction. Underground tanks are likely to be uplifted as a result of high excess pore water pressure developing underneath. For buried pipes, flexible joints have been used to absorb the large-ground deformation while some pipes have been supported by pile foundations. To prevent floating of underground structures,

counterweights are usually placed on top of the structure or the structure is supported by piles.

When the foundation ground liquefies or softens, embankments built on top may slide or settle due to the decrease in bearing capacity. Possible countermeasures that can be used include installation of sheet piles on both edges of the embankment connected by tie rod or placement of berms and counterweight fills to minimize damage to the embankments.

Verification of Effectiveness

After the appropriate remediation technique has been selected and applied to the target site, attempts should be made to verify that the desired level of improvement has been achieved. It is no longer an acceptable practice to simply implement mitigation; the adequacy of the mitigation must also be evaluated.

The most direct way of effectiveness verification is to compare the soil characteristic which was deemed deficient before and after the improvement. For example, if the remediation was performed to increase the strength of the soil, then strength measurements before and after the improvement would be necessary to confirm the effectiveness of the chosen technique. In some cases, direct verification of the target characteristic may not be feasible and checking of related characteristics can be done more easily.

Laboratory Testing Techniques

Verification of the effectiveness of improvement technique through laboratory tests, such as triaxial tests, simple shear tests, and torsional shear tests, offers several advantages and disadvantages. One of the merits of these element tests is that by obtaining soil samples before and after improvement, a more direct inspection can be made. In addition, more accurate measurement of stress, strain, and other parameters can be performed, and therefore, a more accurate characterization of the improved soil is possible.

On the other hand, since soil sampling has to be done only at several representative locations, verification of effectiveness only at these discrete

points can be made and not throughout the whole target area. Moreover, the effect of sample disturbance during sampling, handling, and transportation can have adverse effects on the characteristics of the sample which could affect the verification process.

In Situ Testing Techniques

By far, the most popular approach of checking the level of improvement is by performing in situ tests before and after the treatment process. Procedures like the standard penetration tests (SPT), cone penetration tests (CPT), flat plate dilatometer tests, and pressure meter tests are quite common. This approach has the advantage that the in situ tests can be done quickly, especially the SPT and CPT, and therefore, the level of improvement can be immediately assessed.

However, the results of these in situ tests should be interpreted carefully. For example, the SPT N-value is affected not only by density and confining pressure but also by lateral stress. Therefore, evaluating liquefaction potential using simplified approaches based on SPT N-values alone should be done with caution. These semiempirical methodologies are typically defined for normally consolidated soils, while ground improvement techniques can undoubtedly alter the soil's density, lateral stress, stress and strain history, fabric, and prior effect of ageing. Moreover, the time-dependent nature of the improvement should be considered. Results obtained immediately after the implementation of ground improvement technique may not represent the long-term effect of the technique due to factors like stress redistribution, creep-induced relaxation in lateral stress, etc. In addition, the location of in situ tests and location of treatment points should be considered. For example, to evaluate the effectiveness of sand compaction pile method, it has been customary to perform the penetration test at the center of the grid of sand piles, deemed as the point where the effectiveness of sand compaction is weakest.

Geophysical Testing Techniques

Another approach to verification is through direct transmission seismic testing, such as downhole

and cross-hole seismic testing as well as spectral analysis of surface waves (SASW), which provides calculation of improvement values with some accuracy. In this procedure, the velocity of seismic waves, which is a function of basic soil properties such as modulus of elasticity, density, and Poisson's ratio, is measured between two points and this can give a more accurate picture of the level of improvement as compared to traditional penetration tests. Seismic reflection and refraction tests can be performed to verify the effectiveness of the technique for a wide area.

Design Issues in Ground Improvement

Several factors influence the stability and deformation of improved ground zones and supported structures during and after an earthquake (PHRI 1997; JGS 1998; Mitchell et al. 1998). Some of the more important factors are discussed below.

Size, Location, and Type of Treated Zone

Even when a wide area of soil undergoes liquefaction, the zone requiring soil improvement is generally limited to the area which directly affects the stability of the structure. For example, the portion of the subsoil which provides stability to structures on spread foundation is that directly underneath and on the perimeter of the structure; the portion of the subsoil at considerable distance from the structure has negligible effect on the stability of the said structure. The question then boils down to how wide the soil improvement should be.

In practice, the region to be remediated should be determined in both horizontal and vertical directions. For the vertical direction, it is usual practice to improve the ground to the deepest part of the liquefiable layer. If only the shallow portion of the liquefiable layer is improved, the excess pore water pressure generated at the bottom of the unimproved soil, if it liquefies, may induce upward seepage towards the surface and may result in potential instability at the upper improved zone. As for the lateral extent of treatment necessary, this is dependent on the type of remediation measure employed and the allowable

deformation the structure can undergo. This can be investigated by performing detailed analysis and, if possible, supplemented by model tests. In many cases, the lateral extent of treatment beyond the edge of the structure is a distance equal to the depth of treatment beneath the structure.

Pore Pressure Migration

The difference in stiffness between a compacted zone of material and the surrounding loose liquefiable material could result in seepage between these two materials during and after earthquakes. Higher excess pore water pressure is developed in the untreated material, resulting in propagation into the adjacent improved soil. Although the effect of pore pressure migration into a dense material is very complex, it may nevertheless lead to strength loss of the densified soil.

Ground Motion Amplification

Due to the increase in its stiffness, remediated ground may induce an increase in ground motion, and this may be disadvantageous to the structure on top of it. Although little information is available regarding the influence of size of improved ground and its stiffness on ground motions, engineers should consider a balance in design such that the size and stiffness of the remediated ground would result in acceptable deformations and ground accelerations of the structure during an earthquake.

Dynamic Fluid Pressure

At the boundary between the improved and unimproved ground, both dynamic fluid pressure and static pressure corresponding to earth pressure coefficient equal to 1.0 are applied due to the liquefaction of the untreated ground. These forces must be taken into account when investigating the stability and deformation of structure supported in the improved zone.

Inertia Force

The dynamic response of the improved ground may change as a result of the liquefaction of the unimproved soil and the migration of excess pore

water pressure, leading to a change in the inertia force on the improved ground. This inertia force should be included when evaluating the stability of an improved ground zone supporting a structure. Note however, that the phasing between inertia forces acting on the improved ground and supported structure may be different and this can have significant impact on the results of simplified stability and deformation analyses.

Influence of Structure

The existence of a structure either on top of or adjacent to an improved ground will change the stress state within the said zone and can affect the dynamic response and stress–strain behavior of the improved soil. The effects would depend on the soil and structure conditions, as well as on the installation process.

Forces Exerted by Laterally Spreading Soil

Liquefaction-induced lateral ground movements have been observed to occur at gently sloping sites, with gradients less than 1 %. Therefore, the improved ground should be sufficiently designed to resist the forces induced by the laterally moving ground. There are two force components that should be considered: first, the load exerted by the moving liquefied ground itself, and second, the force exerted by the unliquefied surficial crust riding on top of the liquefied ground.

Summary

Various remediation techniques are available to mitigate the liquefaction risk to structures constructed on loose saturated sandy deposits. These techniques can be roughly classified into two groups: the first is to improve the soil so that liquefaction may not occur, while the second is to provide measures for structures to prevent or minimize damage even if liquefaction occurs. Remedial treatment of soils to prevent liquefaction is based either on the principle of improving the properties of soils or the principle of changing stress or deformation conditions. In some cases, liquefaction is allowed to occur, but the damage

is prevented or minimized by strengthening the structure or foundations.

Ground improvement is being used increasingly for remediating liquefiable soils due to the wide variety of methods that are available. Backed by several decades of research and experience, these methods have been utilized in many soil conditions and have been developed by both the construction industry and research institutions. These techniques can be adapted to site-specific conditions and, in some cases, one or more methods being combined to provide economical solutions for liquefaction remediation problems. Often, new ground improvement techniques are developed when difficult circumstances are encountered and these new techniques are then refined and become part of the expanding set of available mitigation measures.

A few of the remediation methods which have been implemented at many sites all over the world have been tested by few large-scale earthquakes, and most of them performed well, with negligible or minor ground damage. Thus, these remediation techniques will continue to play an important role in the mitigation of seismic risk to existing and future structures.

Cross-References

- [Liquefaction: Performance of Building Foundation Systems](#)

References

- Andrus RD, Chung RM (1995) Ground improvement techniques for liquefaction remediation near existing lifelines. National Institute of Standards and Technology report NISTIR 5714, Gaithersburg, p 74
- Bowen HJ, Cubrinovski M (2008) Effective stress analysis of piles in liquefiable soil: a case study of a bridge foundation. *Bull NZ Soc Earthq Eng* 41(4):247–262
- Cooke HG, Mitchell JK (1999) Guide to remedial measures for liquefaction mitigation at existing highway bridge sites. Technical report MCEER-99-015, Multidisciplinary Center for Earthquake Engineering Research, Buffalo, p 176
- Cox BR, Griffiths SC (2010) Practical recommendations for evaluation and mitigation of soil liquefaction in Arkansas. Project report MBTC 3017, p 175

- Cubrinovski M, Haskell JJM, Bradley BA (2012) Analysis of piles in liquefying soils by the pseudo-static approach. In: Sakr MA, Ansal A (eds) *Special topics in earthquake geotechnical engineering*, vol 16, Geotechnical, geological and earthquake engineering. Springer, Dordrecht, pp 147–174
- Hamada M, Yasuda S, Isoyama R, Emoto K (1986) Study on liquefaction-induced permanent ground displacements. Association for the Development of Earthquake Prediction, Tokai University, Shimizu, Japan
- Hausler EA, Sitar N (2001) Performance of soil improvement techniques in earthquakes. In: *Fourth international conference on recent advances in geotechnical earthquake engineering and soil dynamics*, San Diego, CA, Paper 10.15
- Hausmann MR (1990) *Engineering principles of ground modification*. McGraw-Hill, New York, p 632
- Iai S, Koizumi K (1986) Estimation of earthquake induced excess pore pressure for gravel drains. In: *Proceedings, 7th Japan earthquake engineering symposium*, Tokyo, Japan, pp 679–684
- Iai S, Ichii K, Li H, Morita T (1998) Effective stress analyses of port structures. *Soils Found Special issue on Geotechnical Aspects of the January 17, 1995 Hyogoken-Nambu earthquake* 2:97–114
- Idriss IM, Boulanger RW (2008) *Soil liquefaction during earthquakes*, vol MNO-12, Monograph series. Earthquake Engineering Research Institute, Oakland
- Ishii H, Funahara H, Matsui H, Horikoshi K (2013) Effectiveness of liquefaction countermeasures in the 2011 M = 9 gigantic earthquake, and an innovative soil improvement method. *Indian Geotech J* 43(2):153–160
- Japanese Geotechnical Society, JGS (1998) *Remedial measures against soil liquefaction*. A.A. Balkema, Rotterdam
- Kitazume M (2005) *The sand compaction pile method*. A. A. Balkema, London
- Kitazume M, Terashi M (2013) *The deep mixing method*. CRC Press/Balkema, Leiden
- Maeda S, Nagao K, Tsujino S (2006) Shock wave densification method – countermeasures against soil liquefaction by controlled blasting. Report of Satokogyo Engineering Research Institute, no. 31, pp 21–27 (in Japanese)
- Mitchell JK, Cooke HG, Schaeffer J (1998) Design considerations in ground improvement for seismic risk mitigation. In: *Proceedings, geotechnical earthquake engineering and soil dynamics III*, ASCE geotechnical publication no. 75, vol 1, American Society of Civil Engineers, Reston, VA, pp 580–613
- Munfakh GA, Abramson LW, Barksdale RD, Juran I (1987) In situ ground reinforcement. In: *Soil improvement – a ten year update*. ASCE geotechnical special publication, vol 12. American Society of Civil Engineers, Reston, VA, pp 1–17
- Narin Court W, Mitchell JK (1994) Explosive compaction: densification of loose, saturated cohesionless soils by blasting. Geotechnical engineering report UCB/GT/94-03. University of California, Berkeley
- Pestana JM, Hunt CE, Goughnour RR (1997) FEQDrain: a finite element computer program for the analysis of the earthquake generation and dissipation of pore water pressure in layered sand deposits with vertical drains. Report no. EERC 97–17. Earthquake Engineering Research Center, UC-Berkeley
- Port and Harbour Research Institute, PHRI (1997) *Handbook on liquefaction remediation of reclaimed land*. A.A. Balkema, Rotterdam
- Schaefer V (ed) (1997) *Ground improvement, ground reinforcement, ground treatment: developments 1987–1997*. ASCE geotechnical special publication, vol 69. American Society of Civil Engineers, Reston, VA, p 618
- Seed HB, Booker JR (1977) Stabilization of potentially liquefiable sand deposits using gravel drains. *J Geotech Eng Div ASCE* 103(GT7):757–768
- Seed RB, Cetin KO, Moss RES, Kammerer A, Wu J, Pestana J, Riemer M, Sancio RB, Bray JD, Kayen RE, Faris A (2003) *Recent advances in soil liquefaction engineering: a unified and consistent framework*. Geotechnical report no. UCB/EERC-2003/06. Earthquake Engineering Research Center, University of California, Berkeley
- Sondermann W, Wehr W (2004) Chapter 2: deep vibro techniques. In: Moseley MP, Kirsch K (eds) *Ground improvement* 2nd edn. Spon Press, Taylor & Francis Group, London, UK
- Welsh JP (1992) *Grouting techniques for excavation support*. Excavation Support for the Urban Infrastructure. ASCE, New York, pp 240–261
- Xanthakos PP, Abramson LW, Bruce DA (1994) *Ground control and improvement*. Wiley, New York, p 910
- Yasuda S, Ishihara K, Harada K, Shinkawa N (1996) Effect of soil improvement on ground subsidence due to liquefaction. *Spec Issue Soils Found* 99–107

Liquefaction: Performance of Building Foundation Systems

Shideh Dashti

Department of Civil, Environmental, and Architectural Engineering, University of Colorado at Boulder, Boulder, CO, USA

Synonyms

Building settlement on softened ground; Shallow-founded structures on liquefiable soils; Soil-structure-interaction on softened ground

Definition

When founded on liquefiable ground, building response is commonly evaluated by the geotechnical engineer and structural engineer in a decoupled manner. The geotechnical engineer evaluates the potential for liquefaction triggering under a likely earthquake scenario (in most cases ignoring the presence of the structure), assesses the resulting building settlements, and typically designs a mitigation strategy to prevent liquefaction from occurring or to reduce settlements. On the other side, the structural engineer typically evaluates the seismic performance of the structure under fixed-base conditions (no soil-structure interaction considered), assuming that the liquefaction hazard is mitigated or bypassed. In performance-based structural design, the building *performance* is evaluated in terms of critical engineering demand parameters that correlate well with its damage potential (e.g., inter-story drift). In this article, the *performance* of a building on liquefiable ground is mainly discussed from the geotechnical engineer's viewpoint: the settlement potential.

Introduction

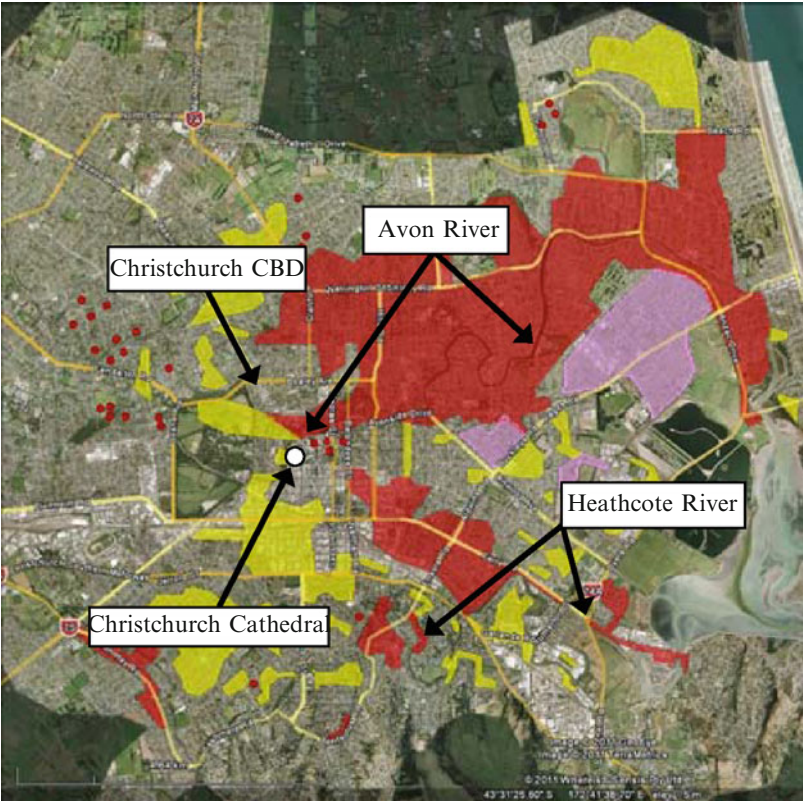
Liquefaction continues to pose a significant risk of damage to the built environment, as observed in both New Zealand and Japan in 2011 (e.g., Fig. 1). More than 50 % of the Christchurch area in New Zealand was affected by soil liquefaction during the 2011 Christchurch earthquake (Fig. 2). Damage caused by settlement, tilt, and lateral sliding, which mostly affected buildings on shallow foundations and their surrounding lifelines, caused severe economic losses (Green et al. 2011; Cubrinovski and McCahon 2012). In many cases, the structures were uneconomical to repair and were demolished. Future earthquakes in major cities around the world are expected to continue causing liquefaction-related damage.

Effective mitigation of the soil liquefaction hazard requires a thorough understanding of the potential consequences of liquefaction and the building performance objectives. The consequences of liquefaction in terms of ground displacement, in turn, depend on site conditions, earthquake loading characteristics, and building properties. This article provides an overview of



Liquefaction: Performance of Building Foundation Systems, Fig. 1 Settlement and tilting of residential buildings in Kamisu City, Japan, following the 2011 Tohoku-Kanto earthquake (Ashford et al. 2011)

**Liquefaction:
Performance of Building
Foundation Systems,
Fig. 2** Liquefaction map
of eastern Christchurch
during the 2011
Christchurch, New
Zealand, earthquake (*red*,
moderate to severe; *yellow*,
low to moderate; *pink*,
liquefaction of
roads – Green et al. 2011)

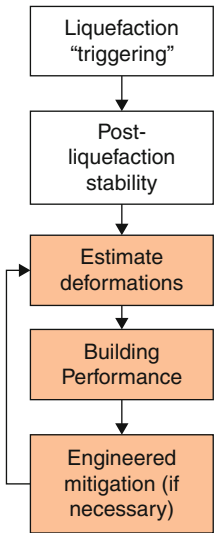


the state of practice in evaluating liquefaction-induced building settlements in geotechnical engineering. A summary of observations from recent case histories, physical model studies, and numerical analyses is presented to provide recommendations for practice and future research.

**Background and Developments in
Liquefaction Engineering**

State of Practice in Evaluating Liquefaction-Induced Building Displacements
Figure 3 provides an overview of the various steps used to analyze the likelihood and impact of liquefaction in design and building assessment (Seed et al. 2003). The process starts by assessing the likelihood of liquefaction triggering (e.g., Youd et al. 2001; Seed et al. 2003). If the soil is

**Liquefaction:
Performance of Building
Foundation Systems,
Fig. 3** Steps in
liquefaction engineering
(Adapted from Seed
et al. 2003)



judged to liquefy during the expected earthquake scenario, the overall stability of the foundation is evaluated using the available post-liquefaction,

residual soil strength to avoid failure and large deformations. If overall stability is not a concern, the consequences of liquefaction need to be evaluated in terms of building displacement. The evaluation of liquefaction-induced building settlements and the impact of those settlements on building performance (highlighted boxes in Fig. 3, which are discussed in this article) are currently steeped in empiricism with significant uncertainties (Seed et al. 2003).

Liquefaction-induced settlements are commonly estimated using simplified empirical procedures (e.g., Tokimatsu and Seed 1987; Ishihara and Yoshimine 1992), which were based on undrained cyclic laboratory tests on samples of saturated, clean sand. These empirical procedures evaluate post-liquefaction, volumetric reconsolidation settlements due to the dissipation of excess pore water pressures after cyclic loading. These methods have generally compared well with previous case histories and physical model studies in the free field or away from buildings. They, however, ignore the presence of the structure and the displacement mechanisms that are active under its foundation. Therefore, these procedures may be misleading in assessing the extent of building settlement and subsequent damage.

To bring in the influence of a structure, practicing engineers often use a combination of (1) empirical methods (e.g., Tokimatsu and Seed 1987; Ishihara and Yoshimine 1992) with the added confining pressure of the structure and (2) case history observations of the effect of foundation width on foundation settlement both normalized by liquefiable layer thickness (Liu and Dobry 1997). In current practice, if the estimated settlements obtained in this manner are judged to be excessive, mitigation techniques are considered to reduce the settlements to an acceptable level or to avoid soil liquefaction altogether. The same empirical procedures as above are then used to evaluate the effectiveness of the proposed mitigation techniques (e.g., ground improvement) in terms of reduction in settlement.

There are presently no well-calibrated, simplified design procedures for estimating the combined and complex effects of deviatoric and

volumetric settlements due to cyclic softening under the static and dynamic loads of structures. This is in contrast with those procedures available for evaluating liquefaction triggering and post-liquefaction reconsolidation settlements in the free field. The lack of a proper understanding of the dominant displacement mechanisms near a building and a reliable analytical tool for estimating liquefaction-induced building settlements has sometimes resulted in the implementation of inadequate hazard mitigation measures (as observed during the 1999 Kocaeli earthquake in Turkey).

Case Histories of Liquefaction-Induced Building Damage

Observations of building performance on liquefied sites during previous earthquakes showed punching settlement, bearing failure, tilt, and lateral shifting of buildings. In the 1964 Niigata (Japan) and the 1990 Luzon (Philippines) earthquakes, most of the damaged buildings were two to four stories, founded on shallow foundations and relatively thick and uniform deposits of clean sand. The confining pressure and shear stress imposed by the buildings and their adjacent structures on the soil affected building movements (e.g., Tokimatsu et al. 1994). Contrary to previous case histories, in the 1999 Kocaeli (Turkey) earthquake, many of the damaged structures were influenced by the liquefaction of thin deposits of silt and silty sand (Sancio et al. 2004; Bray et al. 2000; Bird and Bommer 2004). The building settlement was directly proportional to its contact pressure, and the building's height/width (H/B) aspect ratio greatly affected the degree of tilt (Sancio et al. 2004), showing the importance of the building's dynamic properties. Some buildings translated laterally up to 1 m.

More recently, liquefaction-induced settlements of 1–2 m and tilts exceeding 2° were observed in low- to mid-rise structures in the M_w 6.1, 2011 Christchurch, New Zealand, earthquake. Ground motions recorded on liquefiable sites showed amplified spectral content for periods greater than 2 s. The uplift forces from groundwater pressures caused floors to bulge upward and the foundations to damage and tilt

due to lateral spreading (Cubrinovski and McCahon 2012). Similarly, during the M_w 9, 2011 Tohoku-Kanto earthquake in Japan, the building damage in the Kanto region was dominated by liquefaction, not ground shaking alone (Ashford et al. 2011).

Despite these well-documented case histories, the relation between key ground motion characteristics, soil properties, and the response of building (in terms of settlement, tilt, lateral displacement, and subsequently building performance and damage potential) associated with soil liquefaction, with or without remediation, is not well understood. Buildings that are significantly tilted may need to be demolished and rebuilt, representing a complete loss, despite having a rigid body rotation and intact structural components. On the other hand, there may also be cases in which the liquefied soil, serving as a base isolator, reduces shaking-induced damage to structures by dissipating energy and altering motion frequency content.

Physical Model Studies of Shallow-Founded Structures on Liquefiable Sand

Because of the uncertainties involved in interpreting case histories and limited instrumental recordings at key locations, physical modeling under controlled and simplified conditions provides valuable insights to improve the understanding of liquefaction and its effects. Several researchers have used reduced-scale shaking table and centrifuge tests in the past to study the response of rigid, shallow model foundations situated atop deposits of saturated, loose to medium dense, clean sand (e.g., Yoshimi and Tokimatsu 1977; Liu and Dobry 1997; Hausler 2002; Madabhushi and Haigh 2010).

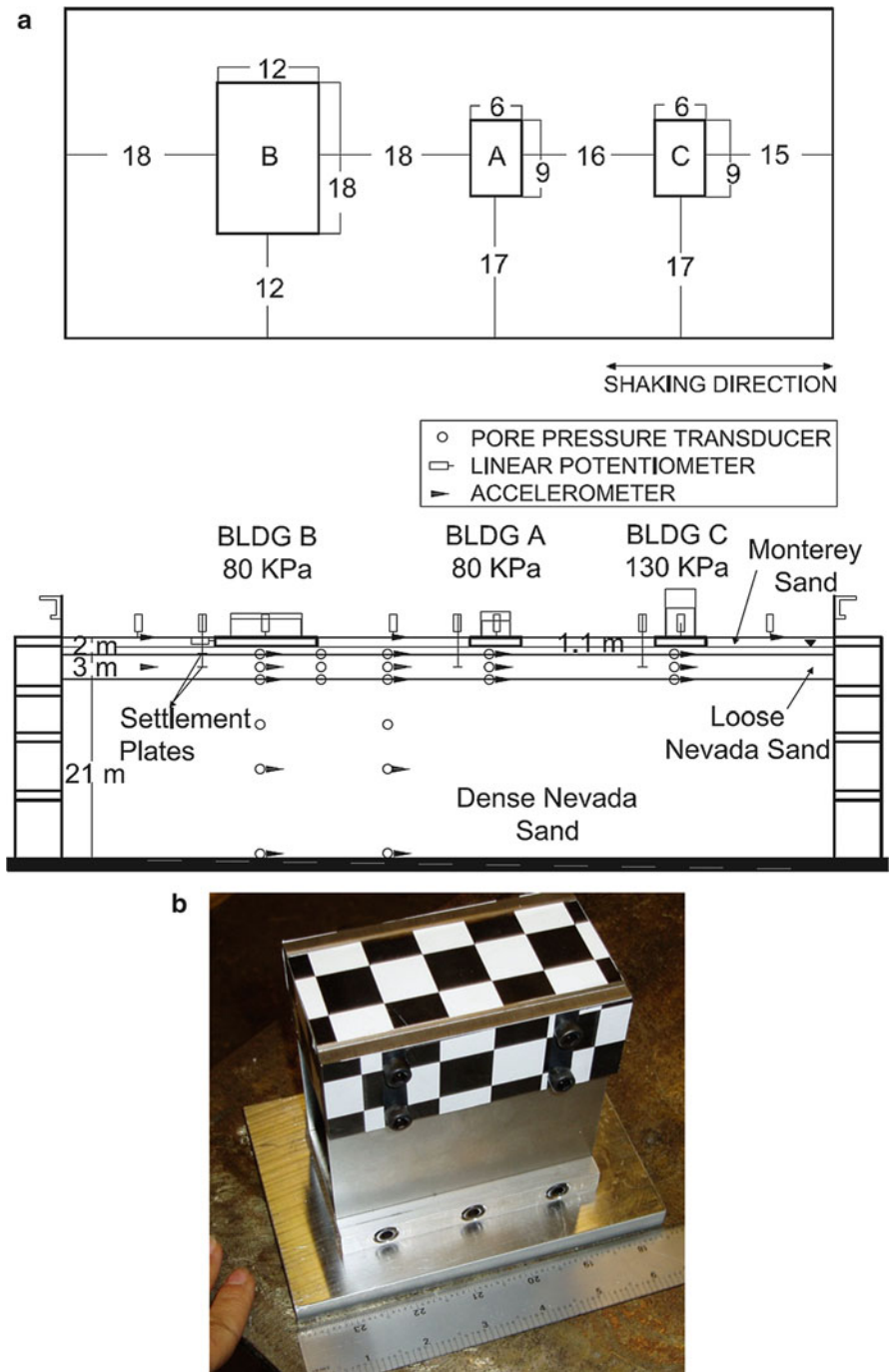
In a recent study, centrifuge experiments were performed to identify the dominant mechanisms of building settlement on relatively thin and shallow deposits of liquefiable, clean sand (Dashti 2009; Dashti et al. 2010a). These tests employed elastic, single-degree-of-freedom (SDOF) structural models with realistic fundamental frequencies (as opposed to a rigid foundation) on liquefiable ground (as shown in Fig. 4). Structures and the soil adjacent to the structures settled

significantly more than the soil in the free field (soil away from the structures) in most cases, as shown in Fig. 5 for a representative test. These comparisons demonstrate the inadequacy of using empirical, free-field procedures to evaluate building settlements.

A strong tendency for flow away from the structures toward the free field was observed during stronger shaking events, which likely led to localized volumetric strains due to partially drained cyclic loading (Dashti et al. 2010a). The relative influence of various testing parameters and structural properties was investigated. Conceptually, the study classified the primary settlement mechanisms as (a) volumetric types, rapid drainage during cyclic loading (ϵ_{p-DR}), sedimentation (ϵ_{p-SED}), and consolidation due to excess pore pressure dissipation (ϵ_{p-CON}) and (b) deviatoric types, partial bearing capacity loss under the static load of the structure (ϵ_{q-BC}) and soil-structure interaction (SSI)-induced building ratcheting under the dynamic load of the building (ϵ_{q-SSI}).

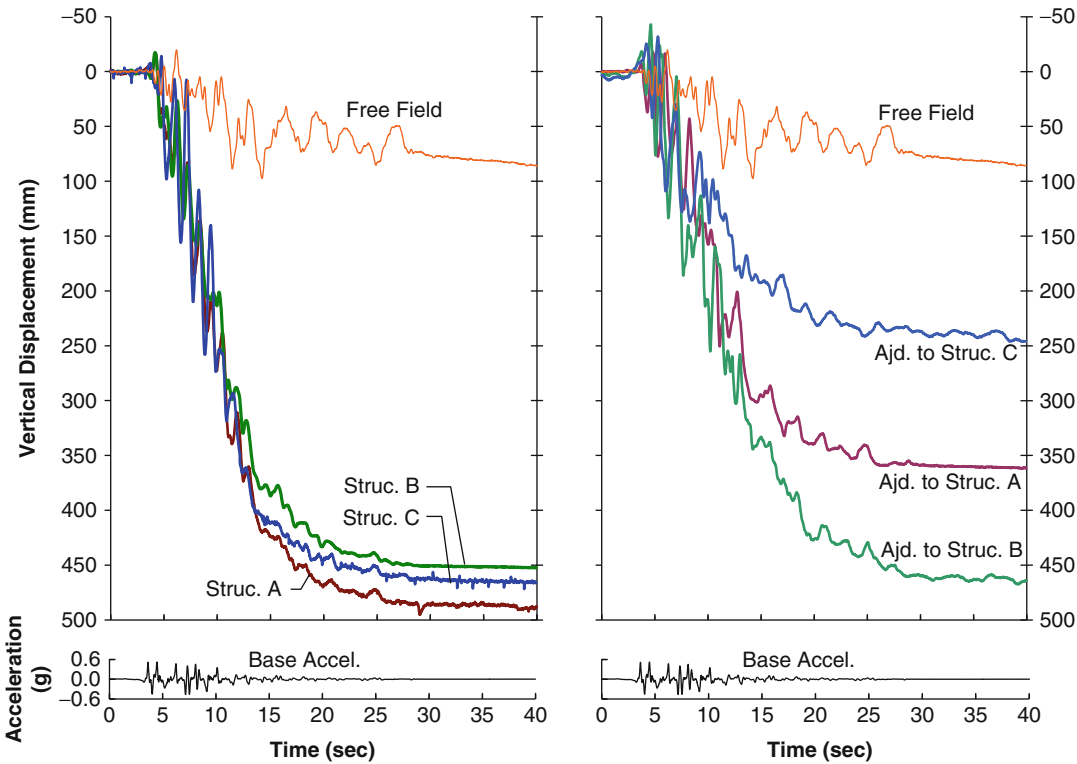
Dashti et al. (2010a) showed that building settlement is not proportional to the thickness of the liquefiable material, as volumetric settlements may not be dominant underneath a structure. As long as there is sufficient thickness of liquefiable soils present under the building, significant settlements may be observed due to deviatoric or shear-induced strains. Hence, the available empirical charts of building settlement that are normalized by the thickness of liquefiable layer may be misleading and should not be used in the engineering practice.

The relative importance of the identified displacement mechanisms, which affects the choice of an effective remediation strategy, depended strongly on several parameters: the liquefiable soil's initial relative density (D_r) and thickness, flow boundary conditions, and ground shaking intensity and the building's fundamental frequency, foundation geometry, and contact pressure. Further, centrifuge experiments revealed that the settlement time history of buildings during each earthquake followed the shape of the Arias Intensity time history of the motion, as shown by the representative results in Fig. 6.



Liquefaction: Performance of Building Foundation Systems, Fig. 4 Centrifuge experiment by Dashti (2009): (a) schematic drawings (plan and elevation views) of a representative test with three SDOF structures

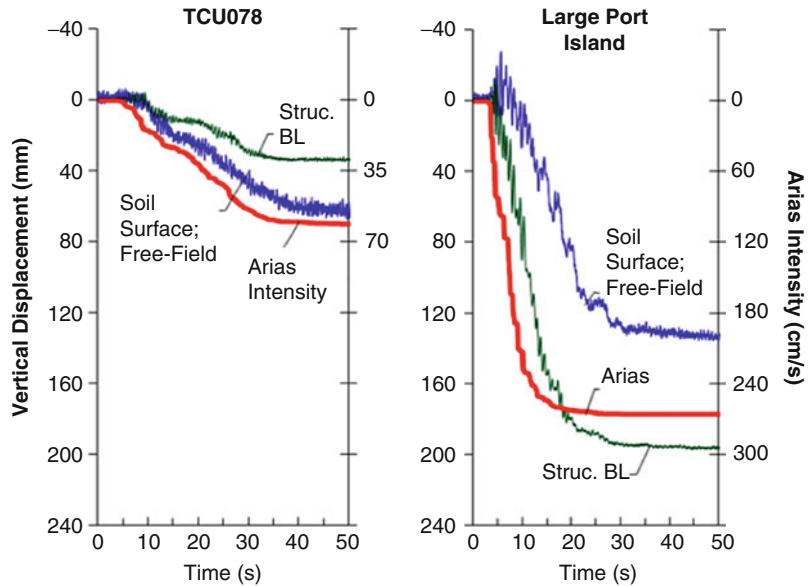
with different properties and (b) picture of model building C used in centrifuge testing. Dimensions shown in prototype scale meters (when the model was spun to 55 g of centrifugal acceleration)



Liquefaction: Performance of Building Foundation Systems, Fig. 5 Settlement of different structures compared to the free field and the adjacent soil during a

representative centrifuge experiment T3-30, with liquefiable layer thickness of 3 m and relative density of 30 % (Adapted from Dashti 2009)

Liquefaction: Performance of Building Foundation Systems, Fig. 6 Settlement time history of the baseline structure (BL) and free field soil surface compared to the corresponding Arias Intensity time history of the base motion during a representative centrifuge experiment T3-50, with liquefiable layer thickness of 3 m and relative density of 50 % (Adapted from Dashti et al. 2010b)



Arias Intensity (I_a) is an index representing the energy of the ground motion in units of L/T (Arias 1970) and defined as

$$I_a(T) = \frac{\pi}{2 \cdot g} \int_0^T a^2(t) \cdot dt \quad (1)$$

over the time period from 0 to T , where a = the measured acceleration value. The Arias Intensity of an earthquake motion depends on the intensity, frequency content, and duration of the ground motion. Its rate represents roughly the rate of earthquake energy buildup. These preliminary experimental results show the importance of taking into account holistic ground motion intensity parameters (e.g., I_a or its slope) in the selection of ground motions for design.

Numerical Studies of Liquefaction Effects

In analytical studies, Soil-Foundation-Structure Interaction (SFSI) effects are commonly evaluated employing equivalent linear, visco-elastic soil models. In reality, the soil response is nonlinear, particularly when it undergoes a significant strength loss and softening as in liquefaction. The insight gained from the previous case histories and physical model studies (summarized in the previous sections) points to the need for performing fully coupled, effective stress, nonlinear dynamic analysis of the soil-structure system, to capture the interacting influences of the key factors on soil and structural response. Popescu and Prevost (1993), Elgamal et al. (2005), Popescu et al. (2006), Andrianopoulos et al. (2006), Lopez-Caballero and Modaressi Farahmand-Razavi (2008), Andrianopoulos et al. (2010), Shahir and Pak (2010), Dashti and Bray (2013), Lopez-Caballero and Modaressi Farahmand-Razavi (2013), and Karamitros et al. (2013) conducted 2-D (plane strain) and 3-D, fully coupled, nonlinear, finite element and finite difference analyses to study the dynamic interaction between homogeneous and layered liquefiable sand and a structure. In these analyses, the structure was simulated either as a surface load, rigid structure, or an elastic SDOF model. The foundation was mostly

fixed to the soil mesh, assuming no relative movement at the interface.

Numerical simulations of SDOF structures on liquefiable ground have shown that liquefaction increases the building's settlement potential, but reduces the seismic demand on the structure and the peak inter-story drift ratio (Lopez-Caballero and Farahmand-Razavi 2013). This observation agrees with previous case histories and may influence the choice and design of remediation techniques in the future. Further, the characteristics of input ground motion used in numerical simulations have been shown to govern the effectiveness of the mitigation technique in terms of settlement and inter-story drift.

Structural engineers employ high-fidelity analytical models of structures, often with no consideration of soil response or soil-structure interaction. In addition, most work in numerical modeling of earthquake effects on structures has focused on the effects of ground shaking alone. Only a few studies have addressed the combined influence of shaking and liquefaction-induced permanent ground movements on the performance of a structure (e.g., Fotopoulou and Pitilakis 2012; Negulescu and Foerster 2010; Aygun et al. 2011). At this time, the influence of ground settlement on the building's engineering demand parameters and damage potential is not clearly understood for a wide range of structures, ground motions, and soil conditions. Complex, nonlinear, dynamic simulations of the soil-structure system are expensive and time consuming and often not justifiable in most projects.

In summary, most analyses of soil liquefaction have ignored structural performance and nonlinearities. On the other hand, nonlinear analyses of structures typically ignore soil nonlinearities or bring in ground shaking and permanent soil displacements in a decoupled manner. It is advantageous, if the project resources allow, that the triggering of liquefaction, post-liquefaction instability, the resulting displacements and ground shaking, and the nonlinear response of the structure be considered simultaneously in a coupled analysis. This approach brings in the complexities of both soil and structure, in order to accurately evaluate the performance of structures on

liquefied ground and the effectiveness of mitigation methods in terms of improved performance. At this time, the cost and expertise required for reliably running such analyses are often prohibitive for most projects.

Summary

The profession has largely addressed the problem of liquefaction triggering assessment. However, there is still much uncertainty in evaluating the consequences of liquefaction in terms of building settlements and the effect of those settlements on building performance. Reliable and simplified procedures for estimating liquefaction-induced building movements that take into account the key displacement mechanisms are currently lacking. As such, observations of ground and building response in the previous case histories, physical model studies, and numerical analyses are combined in this article to provide an overview of the complexities of the problem and guidance for practicing engineers.

The governing displacement mechanisms near a structure are primarily deviatoric induced when the liquefiable soil layer is sufficiently thick and shallow. Methods that estimate free-field, post-liquefaction, volumetric reconsolidation settlement (e.g., Tokimatsu and Seed 1987; Ishihara and Yoshimine 1992) cannot be used to estimate building settlement, because they cannot estimate deviatoric displacement. These procedures either completely ignore the presence of the structure or simplistically bring in its influence through an added foundation load.

It is not appropriate to normalize building settlements by the thickness of the liquefied soil layer. This type of normalization implies that volumetric-induced displacement mechanisms govern building settlement. This is not the case for shallow liquefiable layers under a building, where deviatoric displacement mechanisms dominate (ϵ_{q-BC} and ϵ_{q-SSI}). Localized volumetric strains resulting from partial drainage in response to intense transient hydraulic gradients (ϵ_{p-DR}) are also important.

If liquefaction is triggered in the free field, it is likely that it will also occur under the edges of a building's shallow foundation and in the adjacent soil (Travasrou et al. 2006; Dashti et al. 2010a, b). If significant excess pore water pressures are expected to be generated, liquefaction-induced building movements should be evaluated. A well-calibrated, fully coupled, nonlinear, effective stress dynamic analysis may be performed to provide insight into the consequences of liquefaction in terms of displacements and building performance. This type of advanced analysis may be justifiable for sensitive projects and should only be performed by well-trained and experienced engineers with calibrated and validated constitutive models.

For many regular projects, performing sophisticated, nonlinear, effective stress, dynamic analyses of the soil-structure system may not be practical. First, as an initial check, the seismic bearing capacity of the building must be evaluated taking into account the available post-liquefaction, residual strength of the foundation soil, and the inertial loading of the building. If bearing capacity failure is a concern (factor of safety of near 1), large differential building settlements are expected with a possibility of failure. Hence, remediation is necessary. If the building passes this first instability check, small to moderate settlements are expected typically, for which there are currently no reliable, simplified analytical methods available. In these cases, the available empirical methods for estimating volumetric settlement in the free field must be used with caution and engineering judgment. Considerable localized volumetric strains due to partial drainage as well as static and dynamic deviatoric strains may produce large building settlements and tilt, none of which are taken into account by the available procedures. Special attention should be given to taller buildings with higher aspect ratios (i.e., $H/B > 1.5$), which are more prone to tilting or extreme differential settlements. Previous case histories and physical model studies have shown that a building's tilting potential and overall settlements increase proportionally with its aspect ratio (i.e., H/B) and contact pressure, respectively (Sancio et al. 2004; Dashti

et al. 2010a). In designing a mitigation measure, the dominating mechanisms of displacement should be considered and minimized.

The liquefaction hazard facing existing structures in many parts of the world demands more reliable, performance-based engineering procedures to design remediation techniques. These methods should focus on reducing displacements under and around the foundation as well as structural nonlinearities. Soil, foundation, and building response need to be explicitly linked to damage in developing effective and economic remediation strategies in the future.

Cross-References

- [Dynamic Soil Properties: In Situ Characterization Using Penetration Tests](#)
- [Geotechnical Earthquake Engineering: Damage Mechanism Observed](#)
- [Liquefaction: Countermeasures to Mitigate Risk](#)
- [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)

References

- Andrianopoulos KI, Bouckovalas GD, Karamitros DK, Papadimitriou AG (2006) Effective stress analyses for the seismic response of shallow foundations on liquefiable sand. In: Proceedings of the 6th European conference on numerical methods in geotechnical engineering, Graz
- Andrianopoulos KI, Papadimitriou AG, Bouckovalas GD (2010) Bounding surface plasticity model for the seismic liquefaction analysis of geostructures. *J Soil Dyn EQ Eng* 30:895–911
- Arias A (1970) A measure of earthquake intensity. In: Hansen RJ (ed) *Seismic design for nuclear power plants*. MIT Press, Cambridge, Mass
- Ashford SA, Boulanger RW, Donahue JL, Stewart JP (2011) Geotechnical quick report on the Kanto plain region during the March 11, 2011, Off pacific coast of Tohoku earthquake, Japan. In: NSF supported geotechnical extreme events reconnaissance (GEER) report, 5 April 2011
- Aygun B, Dueas-Osorio L, Padgett JE, Desroches R (2011) Efficient longitudinal seismic fragility assessment of a multispan continuous steel bridge on liquefiable soils. *J Bridge Eng* 16(1):93–107
- Bird JF, Bommer JJ (2004) Earthquake losses due to ground failure. *J Eng Geol* 74(2):147–179
- Bray JD, Stewart JP, Baturay MB, Durgunoglu T, Onalp A, Sancio RB, Ural D, Ansal A, Bardet JB, Barka A, Boulanger R, Cetin O, Erten D (2000) Damage patterns and foundation performance in Adapazari. *Earthq Spectra* 16(S1):163–189
- Cubrinovski M, McCahon I (2012) Short term recovery project 7: CBD foundation damage. Natural hazards research platform. University of Canterbury, Christchurch
- Dashti S (2009) Toward evaluating building performance on softened ground. PhD dissertation, University of California, Berkeley
- Dashti S, Bray JD (2013) Numerical simulation of building response on liquefiable sand. *J Geotechnic Geoenviron Eng, ASCE* 139(8):1235–1249
- Dashti S, Bray JD, Pestana JM, Riemer MR, Wilson D (2010a) Mechanisms of seismically-induced settlement of buildings with shallow foundations on liquefiable soil. *J Geotechnic Geoenviron Eng, ASCE* 136(1):151–164
- Dashti S, Bray JD, Pestana J, Riemer MR, Wilson D (2010b) Centrifuge testing to evaluate and mitigate liquefaction-induced building settlement mechanisms. *J Geotechnic Geoenviron Eng, ASCE* 136(7):918–929
- Elgamal A, Lu J, Yang Z (2005) Liquefaction-induced settlement of shallow foundations and remediation: 3D numerical simulation. *J Earthq Eng* 9(1):17–45
- Fotopoulou SD, Pitilakis KD (2012) Vulnerability assessment of reinforced concrete buildings subjected to seismically triggered slow-moving earth slides. *Landslides* (published online)
- Green RA, Cubrinovski M, Wotherspoon L, Allen J, Bradley B, Bradshaw A, Bray J, De Pascale G, Orense R, Orense R, O'Rourke T, Pender M, Rix G, Wells D, Wood C, Wotherspoon L, Cox B, Henderson D, Hogan L, Kailey P, Lasley S, Robinson K, Taylor M, Winkley A, and Zupan J (2011) Geotechnical reconnaissance of the 2011 Christchurch, New Zealand earthquake. NSF supported geotechnical extreme events reconnaissance (GEER) report, vol 1, 8 Nov 2001
- Hausler EA (2002) Influence of ground improvement on settlement and liquefaction: a study based on field case history evidence and dynamic geotechnical centrifuge tests. PhD dissertation, University of California, Berkeley
- Ishihara K, Yoshimine M (1992) Evaluation of settlements in sand deposits following liquefaction during earthquakes. *J Soils Found* 32(1):173–188
- Karamitros DK, Bouckovalas GD, Chaloulos YK (2013) Seismic settlement of shallow foundations on liquefiable soil with a clay crust. *Soil Dyn Earthq Eng* 46:64–76
- Liu L, Dobry R (1997) Seismic response of shallow foundation on liquefiable sand. *J Geotech Geoenviron Eng* 123(6):557–567

- Lopez-Caballero F, Modaressi Farahmand-Razavi A (2008) Numerical simulation of liquefaction effects on seismic SSI. *Soil Dyn Earthq Eng* 28:85–98
- Lopez-Caballero F, Modaressi Farahmand-Razavi A (2013) Numerical simulation of mitigation of liquefaction seismic risk by preloading and its effects on the performance of structures. *Soil Dyn Earthq Eng* 49:27–38
- Madabhushi SPG, Haigh SK (2010) Effect of superstructure stiffness on liquefaction-induced failure mechanisms. *Int J Geotech Earthq Eng* 1:71–87, ISSN 1947–8488
- Negulescu C, Foerster E (2010) Parametric studies and quantitative assessment of the vulnerability of an RC frame building exposed to differential settlements. *Nat Hazards Earth Syst Sci* 10: 1781–1792
- Popescu R, Prevost JH (1993) Centrifuge validation of a numerical model for dynamic soil liquefaction. *Soil Dyn Earthq Eng* 12:73–90
- Popescu R, Prevost JH, Deodatis G, Chakraborty P (2006) Dynamics of nonlinear porous media with applications to soil liquefaction. *Soil Dyn Earthq Eng* 26:648–65
- Sancio R, Bray JD, Durgunoglu T, Onalp A (2004) Performance of buildings over liquefiable ground in Adapazari, Turkey. In: *Proceedings, 13th world conference on earthquake engineering, Vancouver, Aug 2004, No. 935*
- Seed RB, Cetin KO, Moss RES, Kammerer AM, Wu J, Pestana JM, Riemer MF, Sancio RB, Bray JD, Kayen RE, Faris A (2003) Recent advances in soil liquefaction engineering: a unified and consistent framework. In: *Proceedings, 26th annual ASCE Los Angeles geotechnical spring seminar, Keynote Presentation, H.M.S. Queen Mary, Long Beach*
- Shahir H, Pak A (2010) Estimating liquefaction-induced settlement of shallow foundations by numerical approach. *J Comput Geotech* 37:267–279
- Tokimatsu K, Seed HB (1987) Evaluation of settlements in sands due to earthquake shaking. *J Geotech Eng, ASCE* 113(8):861–878
- Tokimatsu K, Kojima J, Kuwayama AA, Midorikawa S (1994) Liquefaction-induced damage to buildings I 1990 Luzon earthquake. *J Geotech Eng, ASCE* 120(2):290–307
- Travasarou T, Bray JD, Sancio RB (2006) Soil-Structure Interaction Analyses of Building Responses During the 1999 Kocaeli Earthquake. In: *Proceedings, 8th US Nat. Conf. EQ Engrg., 100th Anniv. EQ Conf. Comm. the 1906 San Francisco Earthquake, EERI, April, Paper 1877*
- Yoshimi Y, Tokimatsu K (1977) Settlement of buildings on saturated sand during earthquakes. *Soils Found* 17(1):23–38
- Youd TL, Idriss IM, Andrus RD, Arango I, Castro G, Christian JT, Dobry R, Finn WDL, Harder LF, Hynes ME, Ishihara K, Koester JP, Liao SSC, Marcuson WF,

Martin GR, Mitchell JK, Moriwaki Y, Power MS, Robertson PK, Seed RB, Stokoe KH (2001) Liquefaction resistance of soils: summary report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of liquefaction resistance of soils. *J Geotech Geoenv Eng, ASCE* 127(10):817–833

Liquid Storage Tanks: Seismic Analysis

Maria Vathi and Spyros A. Karamanos
Department of Mechanical Engineering,
University of Thessaly, Volos, Greece

Synonyms

Dynamic analysis of liquid storage tanks; Earthquake analysis of liquid containers

Introduction

Liquid-containing tanks are used in water distribution systems and in industrial plants for the storage and/or processing of a variety of liquids and liquid-like materials, including oil, liquefied natural gas, chemical fluids, and wastes of different forms. These tanks are mainly categorized in two types: ground-supported tanks and elevated tanks. In the present chapter, ground-supported tanks of upright cylindrical shape are presented.

As the necessity and importance of liquid storage tanks has increased over the years, mainly because of their association with the need for continuous supply of energy and water resources, so has the need to understand better their structural behavior and performance. In particular, the analysis and design of liquid storage tanks against earthquake-induced actions has been recognized as an important issue towards safeguarding their structural integrity and has been the subject of numerous analytical, numerical, and experimental works. A pioneering work of Housner (1957), motivated by the need for accurate determination of seismic actions in tanks, presented a solution for the hydrodynamic effects in non-deformable

vertical cylinders and rectangles. For the first time, the total seismic action was split in two parts, namely, the “impulsive” part and the “convective” part. This concept constitutes the basis for the American Petroleum Institute (2003, 2007) Standard provisions (Appendix E) for vertical cylindrical tanks. Veletsos and Yang (1977) and Haroun (1983) have extended this formulation to include the effects of shell deformation and its interaction with hydrodynamic effects. More recently, the case of uplifting of unanchored tanks as well as soil-structure interaction effects have been studied extensively, in the papers by Peek (1988), Natsiavas (1988), Veletsos and Tang (1990), and Malhotra (1995). Notable contributions on the seismic response of anchored and unanchored liquid storage tanks have been presented by Fischer (1979) and Rammerstorfer et al. (1988), with particular emphasis on design implications. Apparently, those papers constitute the basis for the seismic design provisions concerning vertical cylindrical tanks in Part 4 of Eurocode 8, Annex A (European Committee for Standardization 2006b). In addition to the numerous analytical/numerical works, notable experimental contributions on this subject have been reported (Niwa and Clough 1982; Manos and Clough 1982). The reader is referred to the review paper of Rammerstorfer et al. (1990) for a literature review of liquid storage tank response under seismic loads, including fluid-structure and soil-structure interaction effects. Recently, the European research project INDUSE (Pappa et al. 2012) has been completed on the seismic analysis and design of industrial facilities, with particular emphasis on liquid storage tanks. Results from this study have been incorporated in specific design guidelines.

During a strong seismic event, liquid storage tanks may exhibit significant damages. Typical tank damage can be in the following form: (a) buckling of the shell, precipitated by excessive axial compression due to overall bending or beamlike action of the structure; (b) damage to the roof, caused by sloshing of the upper part of the contained liquid when there is insufficient freeboard between the liquid surface and the roof; and (c) failure of attached piping and other

accessories, due to the inability of these elements to accommodate the deformations of the flexible shell. Shell buckling usually takes the form of “diamond shaped” or “elephant’s foot” buckling that appears at a short distance above the base and usually extends around most or all of the circumferences. In previous earthquake events, damages of various severity have been reported, e.g., from fracture of a pipe that created only slight leaks on minor repairs (damage to an overflow pipe) to complete loss of tank content.

Seismic analysis of liquid-containing tanks differs from typical civil engineering structures (i.e., buildings and bridges) in two ways: First, during seismic excitation, liquid inside the tank exerts hydrodynamic force on tank walls and base due to liquid-tank interaction. Second, liquid storage tanks are generally very thin walled and this results in relatively low ductility and low redundancy.

The present chapter outlines principles and methodologies for the seismic analysis of above-ground ground-supported liquid storage tanks of upright cylindrical shape without restrictions on their size and their use. The contents of this chapter are applicable to tanks for both hydrocarbon (oil) and water storage. However, they do not cover containers of different shape, such as horizontal-cylindrical and spherical vessels. For the seismic design of such vessels, the reader is referred to the publication of Karamanos et al. (2006). In addition, they do not refer to elevated tanks, where the tank is supported on braced or unbraced legs, on skirt, or on a more complex structural system. Moreover, the present chapter refers mainly to steel tanks, which are widely used in industrial applications. However, the basic principles of this chapter can be applicable to tanks made of different material.

In the present chapter, following a short presentation of current design standards in “[Relevant Standards and Guidelines](#)”, and a short reference to the seismic input in the course of spectral analysis in “[Seismic Input](#)”, seismic actions in liquid storage tanks are discussed in “[Seismic Loading on Liquid Storage Tanks](#)” in a concise manner, with special reference to hydrodynamic loading. In “[Special Issues on Seismic Action](#)”,

some special issues including the response of unanchored tanks are discussed, whereas in “[Seismic Resistance of Liquid Storage Tanks](#)”, the resistance of liquid storage tanks is outlined. Finally, in “[Summary](#)”, some concluding remarks are stated.

Relevant Standards and Guidelines

There exist several relevant standards and guidelines for the seismic analysis of liquid storage tanks. In this paragraph, the most important European and American Standards as well as the New Zealand Recommendations (Priestley et al. [1986](#)) are briefly presented.

A basic standard on tank seismic design is the 4th part of Eurocode 8 (European Committee for Standardization [2006b](#)). It is a standard that concerns the seismic design of silos, tanks, and pipelines for earthquake resistance and is part of the CEN/TC250 standard group, often referred to as “Structural Eurocodes.” It specifies principles and application rules for the seismic design of structural components including aboveground and buried pipeline systems, silos, and storage tanks of different types and uses. This standard contains significant information on the calculation of seismic action based mainly on the work of Scharf ([1990](#)). Furthermore, seismic resistance in terms of shell buckling is based mainly on the work of Rotter ([1990](#)). It should be noted that EN 1998-4 rules are quite general and may refer to tanks of different material, i.e., steel, concrete, or plastic. It should be noted that this standard should be used together with the other relevant Structural Eurocodes. More specifically, Part 1 of Eurocode 8 (European Committee for Standardization [2004a](#)) should be used for the seismic input in terms of the design spectrum. In addition, Part 1–6 of Eurocode 3 (European Committee for Standardization [2007](#)) gives basic design rules for the strength and structural stability of steel tanks, covering the basic failure modes of steel shells, namely, plasticity, cyclic plasticity, buckling, and fatigue, with emphasis on buckling. Finally, Part 4.2 of Eurocode 3 (European Committee for Standardization

[2006a](#)) provides principles and application rules for the general structural design of vertical cylindrical aboveground steel tanks for the storage of liquid products that should be taken into account.

API 650 (American Petroleum Institute [2007](#)) is a standard dedicated to the general design of liquid storage tanks, developed by the American Petroleum Institute and widely used by the petrochemical industry for the design and construction of tanks in petrochemical facilities. In particular, Appendix E of API 650 refers exclusively to seismic design and contains provisions for both determining seismic actions on tanks and calculating the strength of the tank. Appendix E of the new 11th edition of API 650 (American Petroleum Institute [2007](#)) has significant revisions with respect to the previous 2003 edition (American Petroleum Institute [2003](#)) and includes provisions for site-specific seismic input, calculation of hoop hydrodynamic stresses, distinction between ringwall and slab overturning moments, freeboard requirements, and consideration of vertical excitation effects. It should be noted that the seismic provisions of the updated Appendix E are in accordance with the provisions of ASCE 7 Standard (American Society of Civil Engineers [2006](#)), which refers to structural loading.

EN 14015 (European Committee for Standardization [2004b](#)) is a European standard outside the Structural Eurocode framework, i.e., independent of EN 199x Standards, which specifies the requirements for the materials, design, fabrication, erection, testing, and inspection of site built, vertical, cylindrical, flat-bottomed, aboveground, welded, steel tanks for the storage of liquids at ambient temperatures and above. It constitutes an updated version of the classical British Standard BS 2654 (British Standards Institution [1989](#)), which is a specification for the manufacture of vertical steel welded non-refrigerated storage tanks with butt-welded shells for the petroleum industry. It is a complete standard for tank design analogous to API 650 (American Petroleum Institute [2007](#)). Recommendations for seismic design and analysis of storage tanks are stated in Annex G of EN 14015,

which are quite similar with the requirements of Annex E of API 650 (American Petroleum Institute 2003).

The New Zealand Seismic Tank Design Recommendations (Priestley et al. 1986) is a document that, when published in 1986, contained pioneering recommendations for the seismic design of storage tanks, developed by a study group of the New Zealand National Society for Earthquake Engineering. The intention of this study group was to collate existing information, available in research papers and codes, and to produce uniform recommendations that would cover a wide range of tank configurations and contained materials. These recommendations attempted to produce a unified approach for the seismic design of storage tanks, regardless of material or function, and to provide additional information to that already available in alternative sources. Several decades after their publication, these recommendations have been widely referenced and adopted internationally, including EN 1998-4 (European Committee for Standardization 2006b).

Seismic Input

In the framework of a pseudo-dynamic spectral analysis, the earthquake motion is represented by an appropriate design spectrum, $S_a(T)$. As an example, EN 1998-1 (European Committee for Standardization 2004a) provides a design spectrum for general use. Similar design spectra are provided by all seismic design codes. The design spectrum for horizontal earthquake excitation is usually expressed through an analytical equation for the spectral acceleration S_a in terms of the natural period T of the structural system, the design ground acceleration a_g , the damping coefficient ξ , the behavior factor q , the importance factor γ_I , and the soil factor S .

In addition, provisions for the vertical component of the seismic action are given in terms of the peak vertical ground acceleration a_{gv} usually as a percentage of a_g . In this section, two main issues, the behavior factor and the importance factor, are discussed in more detail.

Behavior Factor

The behavior factor, q , accounts for the capacity of the structural system to dissipate seismic energy. This results in a reduction of the forces obtained from a linear elastic analysis, in order to account for the nonlinear response of the structure, associated with the material, the structural system, and the design procedure. However, the structural system should be able to sustain such nonelastic behavior and dissipate seismic energy.

Upright cylindrical steel tanks (non-base-isolated) supported directly on the ground or on the foundation may be designed with a behavior factor q that represents a quasi-elastic behavior, i.e., $q \leq 1.5$. More specifically, a value of $q = 1$ is recommended for the convective motion where negligible dissipation of energy occurs, whereas a value of $q = 1.5$ (quasi-elastic behavior) is recommended for the impulsive motion.

The relatively low values of q for the impulsive motion are justified by the unstable nature of shell buckling primary mode of failure. This philosophy is adopted by European Standard EN 1998-4 (European Committee for Standardization 2006b). According to this standard, values of behavior factor greater than 1.5 for the impulsive motion can be used (up to 2.5); in special cases when the tank or its foundation is appropriately designed to allow uplift and the localization of deformations in the shell wall, the bottom plate or their intersection does not cause failure.

American Standard API 650 (American Petroleum Institute 2007) uses a similar concept to account for the capacity of the tank to dissipate seismic energy through the so-called response modification factor or simply reduction factor, denoted by “ R .” The response modification factor for ground-supported, liquid storage tanks designed and detailed to these provisions shall be less than or equal to the values shown in Table E-4 of API 650.

Nevertheless, there exist significant differences between API 650 (American Petroleum Institute 2007) and EN 1998-4 (European Committee for Standardization 2006b) for the value of the behavior factor. In particular, API

650 proposes larger values for the behavior factor for both the impulsive and convective actions than EN 1998-4. It is important to note that in both standards those values are significantly higher than unity which implies the presence of a dissipating mechanism.

In conclusion, the behavior (reduction) factor for the seismic design of liquid storage tanks is a controversial issue that has not been examined thoroughly so far.

Importance Factor

Reliability differentiation in liquid storage tank design is implemented by classifying tanks into different “importance classes.” The importance classes depend on the potential loss of life due to the failure of the particular tank and on the economic and social consequences of failure. Clearly, the importance of a specific tank should be assessed considering the facility or plant it belongs to and its proximity to urban areas. In EN 1998-4 (European Committee for Standardization 2006b), the following classes for liquid storage tanks are specified:

- *Class I* refers to situations where the risk to life is low and the economic and social consequences of failure are small or negligible.
- Situations with medium risk to life and local economic or social consequences of failure belong to *Class II*.
- *Class III* refers to situations with a high risk to life and large economic and social consequences of failure.
- *Class IV* refers to situations with exceptional risk to life and extreme economic and social consequences of failure.

An importance factor γ_I is assigned to each importance class, ranging from 0.8 to 1.6 for the four importance classes unless otherwise specified in National Annexes, depending on the local seismic hazard conditions or by the owner. An analogous classification (with three important classes) is also specified in API 650 (American Petroleum Institute 2007), Appendix E, Section E.5.1.2, Table E-5.

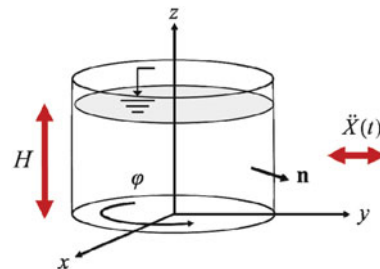
Seismic Loading on Liquid Storage Tanks

Horizontal Seismic Action

General

A cylindrical coordinate system is usually used to describe the tank geometry with orthogonal coordinates: r, φ, z where the origin at the centre of the tank bottom and the z axis is vertical (Fig. 1). The height of the tank to the original of the free surface of the fluid and its radius are denoted by H and R , respectively; ρ is the mass density of the fluid, while the nondimensional coordinates $\xi = r/R$ and $\zeta = z/H$ in the radial direction and along the height of the tank, respectively, are usually considered.

The liquid contained in an upright cylinder subjected to horizontal seismic excitation $\ddot{X}(t)$ may be considered as ideal fluid (potential flow) resulting in a hydrodynamic problem (Ibrahim 2005) associated with the motion of the liquid free surface and the development of standing waves. The solution of the hydrodynamic problem shows that the liquid motion can be expressed as the sum of two separate contributions, called impulsive and convective, respectively. The impulsive component of the motion satisfies exactly the boundary conditions at the tank wall and the bottom of the tank and corresponds to zero pressure at the free surface of the fluid. It represents the motion of the fluid part that “follows” the motion of the container. The convective term is associated with liquid sloshing of



Liquid Storage Tanks: Seismic Analysis, Fig. 1 Cylindrical coordinate system describing tank geometry

its free surface and satisfies the kinematic and dynamic conditions at the free surface.

As a result, the hydrodynamic pressure is decomposed into an impulsive part and a convective part. In an analogous manner, the total mass of the liquid, m_L , can be considered as the sum of two parts, an impulsive part, m_I , and a convective part, m_C , so that $m_L = m_I + m_C$.

The convective motion is associated with an infinite number of modes and the corresponding masses are computed as follows:

$$\frac{m_{Cn}}{m_L} = \frac{2 \tanh(k_n R \gamma)}{k_n R \gamma (k_n^2 R^2 - 1)}, \quad n = 1, 2, 3, \dots \quad (1)$$

where n is the number of sloshing mode, R is the radius of the tank, γ is the tank aspect ratio defined as the ratio of the liquid height over the tank radius ($= H/R$), and $k_n R = \lambda_n$ are the roots of equation $J'_1(\cdot) = 0$. More specifically the first three roots are $k_1 R = \lambda_1 = 1.841$, $k_2 R = \lambda_2 = 5.331$, and $k_3 R = \lambda_3 = 8.536$ corresponding to the first three sloshing modes. Note that the total convective mass is the sum of the individual convective masses, $m_C = \sum_{n=1}^{\infty} m_{Cn}$. Usually, consideration of only the first one ($n = 1$), or the first two modes ($n = 1, 2$) in some special cases, is adequate for design purposes.

Masses m_I and m_{Cn} for $n = 1, 2$ are plotted in Fig. 2 as functions of the tank aspect ratio $\gamma = H/R$. The impulsive mass is expressed as the ratio m_I/m_L of the impulsive mass over the total liquid mass. This ratio increases with

increasing values of γ , tending asymptotically to unity. On the other hand, the convective mass, expressed in normalized form as m_{C1}/m_L , decreases with increasing values of γ . The value of m_{C2} is quite small but may have some effect on the dynamic response in broad tanks where $\gamma \rightarrow 0$.

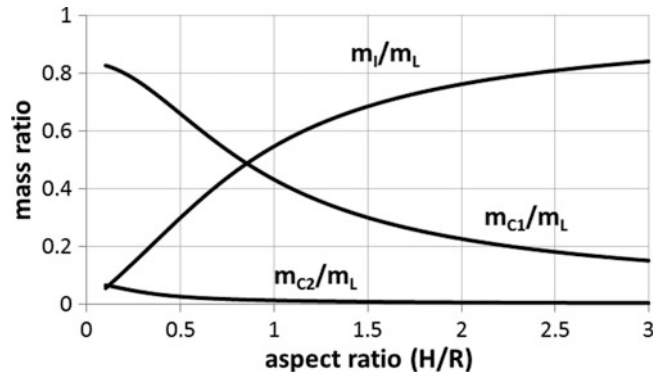
In the course of a design procedure, the impulsive mass m_I should consist of the above impulsive mass of the liquid, plus the mass of the moving tank wall, m_{SH} , and the mass of the tank roof, m_r .

Vibration Periods and Damping Coefficients

In case of rigid (non-deformable) tank walls, the impulsive motion follows exactly the ground motion. In such a case the impulsive circular frequency has an infinite value ($\omega_I \rightarrow \infty$) and the impulsive period is zero ($T_I = 0$). This is not applicable in the case of flexible (deformable) tank walls. To simplify the analysis, it is reasonable to assume that the flexibility of tank wall affects almost exclusively the impulsive component of the response, and therefore, the impulsive hydrodynamic effects for flexible tanks are generally significantly larger than those for rigid tanks. On the other hand, the convective components of the response can be considered insensitive to variations in wall flexibility, because they are associated with natural periods of vibration that are significantly longer than the dominant periods of the ground motion or pipe wall vibration motion. Therefore, the convective components of seismic action for tank proportions normally encountered in practice can be

Liquid Storage Tanks: Seismic Analysis,

Fig. 2 Mass ratio as a function of the aspect ratio



considered the same as those obtained for rigid tanks.

Analytical expressions are available for the impulsive vibration period and frequency accounting for tank flexibility. From hydrodynamics, the natural circular frequency and natural period of the impulsive response of flexible tanks are calculated by the expressions:

$$\omega_I = \frac{2\pi}{Y_I H \sqrt{\frac{\rho R}{t_{eq} E}}} \quad (2)$$

$$T_I = Y_I H \sqrt{\frac{\rho R}{t_{eq} E}} \quad (3)$$

where ρ is the mass density of the fluid, E is Young's modulus, t_{eq} is the equivalent uniform thickness of the tank wall (can be taken as the weighted average over the wetted height of the tank wall, the weight may be taken proportional to the strain in the wall of the tank, which is maximum at the base of the tank), and coefficient Y_I is given in Table 1 for specific values of the aspect ratio γ of the tank ($\gamma = H/R$). For intermediate values of γ , a simple interpolation may be used.

Furthermore, the circular frequency (Fig. 3) and natural period of the convective response,

denoted as ω_{Cn} and T_{Cn} , respectively, for the n -th sloshing mode can be calculated by the following expressions:

$$\omega_{Cn} = \sqrt{g \frac{\lambda_n}{R} \tanh(\lambda_n \gamma)}, \quad n = 1, 2, 3, \dots \quad (4)$$

$$T_{Cn} = \frac{2\pi}{\omega_{Cn}}, \quad n = 1, 2, 3, \dots \quad (5)$$

where λ_n are the roots of $J'_1(\cdot) = 0$, $\lambda_1 = 1.841$, $\lambda_2 = 5.331$, and $\lambda_3 = 8.536$ for the first three sloshing modes).

The impulsive and convective damping coefficients can be taken equal to $\xi_I = 5\%$ and $\xi_C = 0.5\%$ for the impulsive and the convective motion, respectively.

Impulsive Motion and Pressure

The impulsive motion can be expressed in terms of the impulsive acceleration of the liquid storage tank, $\ddot{u}_I(t)$, which can be calculated from the following linear oscillator equation:

$$\ddot{u}_I + 2\xi_I \omega_I (\dot{u}_I - \dot{X}) + \omega_I^2 (u_I - X) = 0 \quad (6)$$

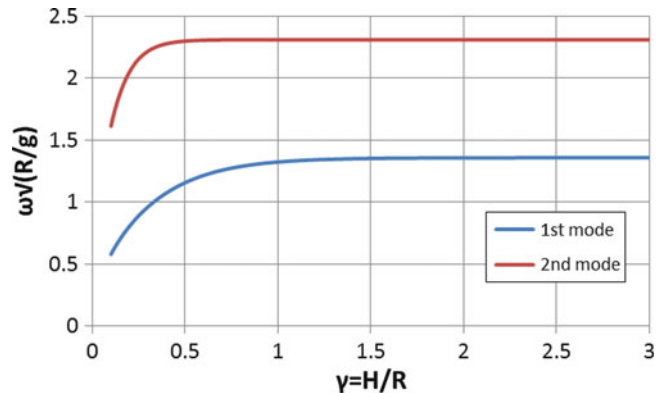
where $u_I(t)$ is the generalized coordinate representing the total impulsive displacement

Liquid Storage Tanks: Seismic Analysis, Table 1 Coefficient Y_I in terms of the aspect ratio

H/R	0.3	0.5	0.7	1.0	1.5	2.0	2.5	3.0
Y_I	9.28	7.74	6.97	6.36	6.06	6.21	6.56	7.03

Liquid Storage Tanks: Seismic Analysis,

Fig. 3 Values of the first two sloshing frequencies as functions of γ



motion of the tank, ξ_I the viscous damping ratio of the impulsive motion, and ω_I the impulsive frequency defined in Eq. 2. Setting

$$a_I(t) = u_I(t) - X \quad (7)$$

one can obtain

$$\ddot{a}_I + 2\xi_I\omega_I\dot{a}_I + \omega_I^2 a_I = -\ddot{X}(t) \quad (8)$$

In the above expressions, function $a_I(t)$ can be considered as the generalized coordinate representing the relative impulsive displacement of the tank with respect to its base.

The spatial-time variation of the *impulsive* component of hydrodynamic pressure within the liquid domain is given by the following expression:

$$p_I(\xi, \zeta, \varphi, t) = C_I(\xi, \zeta) \rho H \cos \varphi \ddot{u}_I(t) \quad (9)$$

where the spatial coefficient C_I is defined:

$$C_I(\xi, \zeta) = 2 \sum_{n=0}^{\infty} \frac{(-1)^n}{I_1'(v_n/\gamma) v_n^2} \cos(v_n \zeta) I_1\left(\frac{v_n}{\gamma} \xi\right) \quad (10)$$

in which $v_n = \frac{2n+1}{2}\pi$, $\gamma = H/R$ is the aspect ratio of the tank, whereas $I_1(\cdot)$ and $I_1'(\cdot)$ denote the modified Bessel function of order 1 and its derivative, respectively.

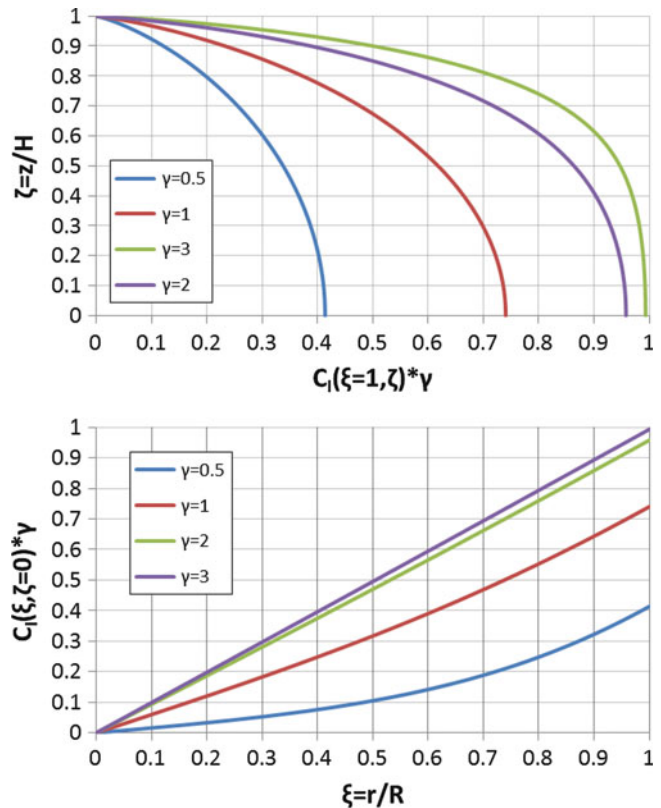
For seismic design purposes, the distribution of the impulsive hydrodynamic pressure on the pipe walls is assumed as follows:

$$p_I(\xi, \zeta, \varphi) = C_I(\xi, \zeta) \rho H \cos \varphi S_a(T_I) \quad (11)$$

where ξ is taken equal to 1 for computing the pressure on the lateral surface of tank shell, whereas ζ should be considered equal to zero for calculating the pressure on the tank bottom plate (Fig. 4).

Liquid Storage Tanks: Seismic Analysis,

Fig. 4 Variation of the impulsive pressure (normalized to $\rho R a_g$) for three values of $\gamma = H/R$, (a) variation along the height, (b) radial variation on the tank bottom. a_g is the maximum ground acceleration



Convective Motion and Pressure

The convective acceleration, $\ddot{u}_{Cn}(t)$, is calculated from the following linear oscillator equation:

$$\ddot{u}_{Cn} + 2\xi_{Cn}\omega_{Cn}(\dot{u}_{Cn} - \dot{X}) + \omega_{Cn}^2(u_{Cn} - X) = 0 \quad (12)$$

where $u_{Cn}(t)$ is the generalized coordinate representing the total convective motion of the liquid associated with the n -th mode, ξ_{Cn} the viscous damping ratio of the convective motion for the n -th sloshing mode, and ω_{Cn} the convective frequency. Setting

$$a_{Cn}(t) = u_{Cn}(t) - X \quad (13)$$

one can write

$$\ddot{a}_{Cn} + 2\xi_{Cn}\omega_{Cn}\dot{a}_{Cn} + \omega_{Cn}^2 a_{Cn} = -\ddot{X}(t) \quad (14)$$

where $a_{Cn}(t)$ is the generalized coordinate representing the relative convective motion of the liquid associated with the n -th sloshing mode.

The spatial distribution within the liquid domain and time variation of the “convective” component of hydrodynamic pressure associated with the n -th convective mode is given by the following expression:

$$p_{Cn}(\xi, \zeta, \varphi, t) = \rho\psi_n \cosh(\lambda_n \gamma \zeta) J_1(\lambda_n \xi) \cos \varphi \ddot{u}_{Cn}(t) \quad (15)$$

where

$$\psi_n = \frac{2R}{(\lambda_n^2 - 1)J_1(\lambda_n) \cosh(\lambda_n \gamma)} \quad (16)$$

J_1 is the Bessel function of the first order and $\lambda_1 = 1.841$, $\lambda_2 = 5.331$, and $\lambda_3 = 8.536$ are the roots of $J_1'(\lambda) = 0$. In Eq. 15, r is taken equal to $R(\xi = 1)$ for the pressure on the lateral surface, whereas z is taken equal to zero ($\zeta = 0$) for the pressure on the bottom plate. In the majority of practical applications, only the first convective mode ($n = 1$) of the fluid is necessary (Pappa et al. 2011) and the distribution of the pressure is as follows:

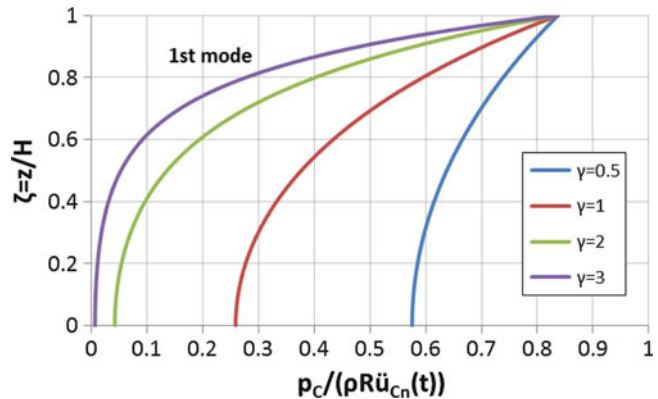
$$p_C(\xi, \zeta, \varphi) = 1.146R\rho \frac{\cosh(1.841\gamma\zeta)J_1(1.841\xi)}{\cosh(1.841\gamma)} \cos \varphi S_a(T_{C1}) \quad (17)$$

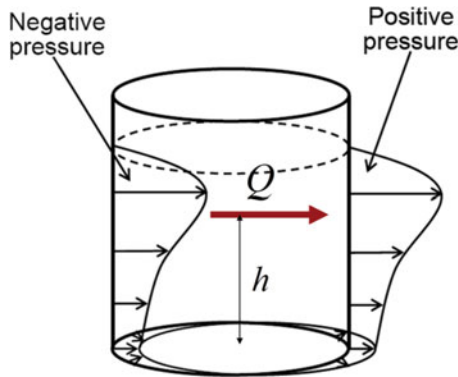
where J_1 is the Bessel function of first order. In the above expressions, r is taken equal to R ($\xi=1$) for determining the hydrodynamic pressure on the lateral tank (shell) surface (Fig. 5),

whereas z is taken equal to zero ($\zeta = 0$) for the hydrodynamic pressure on the bottom plate. In addition, 1.841 is the first root of $J_1'(0) = 0$.

Liquid Storage Tanks: Seismic Analysis,

Fig. 5 Variation of first-mode sloshing pressures on tank wall along the tank height





Liquid Storage Tanks: Seismic Analysis,
Fig. 6 Horizontal seismic force Q acting at a height h

In broad tanks (i.e., tanks with large values of H/R ratio), the sloshing pressures may be significant near the bottom, while in slender (tall) tanks, the sloshing effect is limited to the vicinity of liquid surface.

Horizontal Seismic Forces

Integrating the pressure distribution on the tank wall, the impulsive base shear force at the base of the wall can be calculated as follows:

$$Q_I(t) = \int_{B_q} p_I(\mathbf{n} \cdot \mathbf{e}_x) dB_q \quad (18)$$

where $\mathbf{e}_x$ is the unit vector in the direction of the earthquake, $\mathbf{n}$ is the unit normal vector to the liquid surface, and B_q is the boundary surface of the liquid (Fig. 6).

The result of this integration can be readily written as follows:

$$Q_I(t) = m_I \ddot{u}_I(t) \quad (19)$$

where m_I is the impulsive mass defined in Fig. 2 and $\ddot{u}_I(t)$ is the impulsive acceleration parameter, defined in Eq. 6.

The convective base shear can be obtained by integrating the convective pressure (p_C) components on the lateral shell of the tank:

$$Q_C(t) = \int_{B_q} p_C(\mathbf{n} \cdot \mathbf{e}_x) dB_q \quad (20)$$

The result of this integration can be readily written as follows:

$$Q_C(t) = \sum_{n=1}^{\infty} m_{Cn} \ddot{u}_{Cn}(t) \quad (21)$$

where m_{Cn} is the convective mass for the n -th mode defined in Fig. 2 and $\ddot{u}_{Cn}(t)$ is the convective acceleration parameter for the n -th sloshing mode, defined in Eq. 12.

In the course of a spectral seismic analysis of the tank, the impulsive base shear Q_I and convective base shear Q_{Cn} for the n -th mode at the base of the wall can be computed in terms of the corresponding spectral acceleration as follows:

$$Q_I = m_I S_a(T_I) \quad (22)$$

$$Q_{Cn} = m_{Cn} S_a(T_{Cn}), \quad n = 1, 2, 3, \dots \quad (23)$$

Ringwall and Slab Overturning Moment/Heights

The ringwall moment, M , is the overturning moment that acts immediately above the bottom plate of the tank. It includes only the contribution of pressure on the wall and should be used for the calculation of stresses and stress resultants in the tank wall and at its connection to the base, for tank structural verification. On the other hand, the slab moment, M' , is the overturning moment that acts immediately below the bottom plate of the tank. It includes the contribution of pressure on the wall together with the pressure on the tank bottom and should be used for the verification of its support structure, base anchors, or foundation.

The impulsive ringwall base moment is calculated through the following integration:

$$M_I(t) = \int_{B_q} p_I z(\mathbf{n} \cdot \mathbf{e}_x) dB_q \quad (24)$$

The result of this integration can be readily written as follows:

$$M_I(t) = m_I h_I \ddot{u}_I(t) \quad (25)$$

or

$$M_I(t) = Q_I(t) h_I \quad (26)$$

where m_I is the impulsive mass defined in Fig. 2; h_I is the height at which acts the horizontal seismic force, $Q_I(t)$; and $\ddot{u}_I(t)$ is the impulsive acceleration parameter, defined in Eq. 6.

The convective ringwall base moment is calculated in an analogous manner:

$$M_C(t) = \int_{B_q} p_{Cz}(\mathbf{n} \cdot \mathbf{e}_x) dB_q \quad (27)$$

The result of this integration can be readily written as follows:

$$\begin{aligned} M_C(t) &= \sum_{n=1}^{\infty} m_{Cn} \ddot{u}_{Cn}(t) h_{Cn} \\ &= \sum_{n=1}^{\infty} Q_{Cn}(t) h_{Cn} \end{aligned} \quad (28)$$

where m_{Cn} is the convective mass for the n -th mode defined in Fig. 2; h_{Cn} is the height at which acts the horizontal seismic force for the n -th sloshing mode, $Q_{Cn}(t)$; and $\ddot{u}_{Cn}(t)$ is the

convective acceleration parameter for the n -th sloshing mode, defined in Eq. 12.

Analytical expressions for the heights at which impulsive and convective forces are acting are available from hydrodynamics, represented also in graphical form in Fig. 7.

For the impulsive moment:

$$h_I = H \frac{\sum_{n=0}^{\infty} \frac{(-1)^n I_1(v_n/\gamma)}{v_n^4 I_1'(v_n/\gamma)} (v_n(-1)^n - 1)}{\sum_{n=0}^{\infty} \frac{I_1(v_n/\gamma)}{v_n^3 I_1'(v_n/\gamma)}} \quad (29)$$

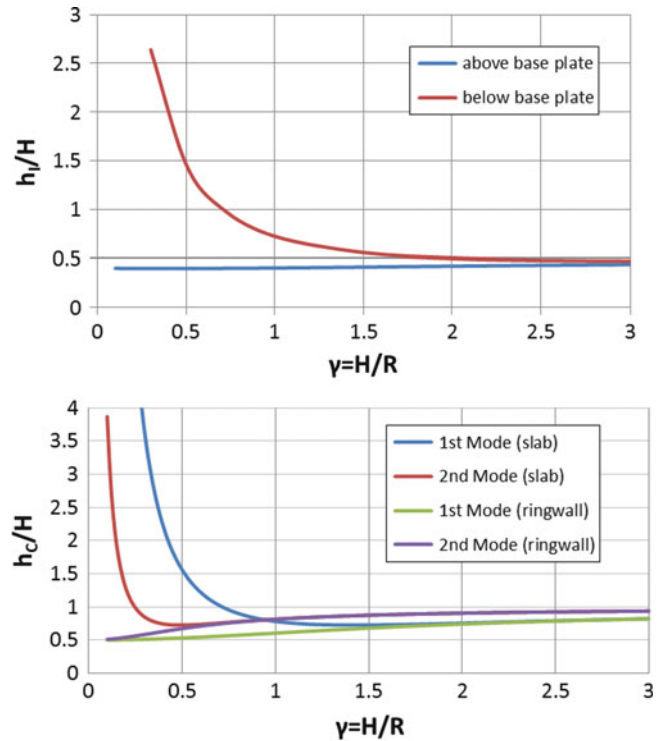
$$h'_I = H \frac{\frac{1}{2} + 2\gamma \sum_{n=0}^{\infty} \frac{v_n + 2(-1)^{n+1} I_1(v_n/\gamma)}{v_n^4 I_1'(v_n/\gamma)}}{2\gamma \sum_{n=0}^{\infty} \frac{I_1(v_n/\gamma)}{v_n^3 I_1'(v_n/\gamma)}} \quad (30)$$

For the convective moment:

$$h_{Cn} = H \left(1 + \frac{1 - \cosh(\lambda_n \gamma)}{\lambda_n \gamma \sinh(\lambda_n \gamma)} \right) \quad (31)$$

Liquid Storage Tanks: Seismic Analysis,

Fig. 7 Ratios h_I/H , h'_I/H , h_C/H , and h'_C/H as functions of the aspect ratio



$$h'_{Cn} = H \left(1 + \frac{2 - \cosh(\lambda_n \gamma)}{\lambda_n \gamma \sinh(\lambda_n \gamma)} \right) \quad (32)$$

In the course of a spectral design and analysis procedure, the impulsive and convective ringwall moments are given by the following expressions:

$$M_I = m_I h_I S_a(T_I) \quad (33)$$

$$M_C = m_{Cn} h_C S_a(T_{Cn}) \quad (34)$$

Furthermore, simplified equations for the height h_I are stated in Section 6.1.2 of API 650 (American Petroleum Institute 2007) as alternative to Eq. 29:

when $D/H \geq 1.333$:

$$h_I = 0.375H \quad (35)$$

when $D/H < 1.333$:

$$h_I = \left[0.5 - 0.094 \frac{D}{H} \right] H \quad (36)$$

and

$$h_C = \left[1 - \frac{\cosh\left(\frac{3.67H}{D}\right) - 1}{\frac{3.67H}{D} \sinh\left(\frac{3.67H}{D}\right)} \right] H \quad (37)$$

Similarly, the impulsive and convective slab moments are given by the following expressions:

$$M'_I = m_I h'_I S_a(T_I) \quad (38)$$

$$M'_C = m_{Cn} h'_C S_a(T_{Cn}) \quad (39)$$

and an alternative equation for the height h'_I can be found in Section 6.1.2 of API 650 (American Petroleum Institute 2007):

when $D/H \geq 1.333$:

$$h'_I = 0.375 \left[1 + 1.333 \left(\frac{0.866 \frac{D}{H}}{\tanh\left(0.866 \frac{D}{H}\right)} - 1 \right) \right] H \quad (40)$$

when $D/H < 1.333$:

$$h'_I = \left[0.5 - 0.06 \frac{D}{H} \right] H \quad (41)$$

Combination of Impulsive/Convective Action

In the course of a spectral analysis and design procedure, the impulsive and convective action should be combined in order to compute the total horizontal hydrodynamic force. In particular it is necessary to combine (a) the sloshing modes to obtain the resultant convective action and (b) the impulsive and convective actions for the total seismic force.

Both combinations should be conducted according to the SRSS rule, rather than the addition of the maximum convective and maximum impulsive forces as noted in a recent publication (Pappa et al. 2011).

More specifically, the total horizontal seismic force is

$$Q_T = \sqrt{[m_I S_a(T_I)]^2 + \sum_{n=1}^{N_1} [m_{Cn} S_a(T_{Cn})]^2} \quad (42)$$

where N_1 is the number of sloshing modes considered.

The above SRSS combination rule applies also to the total “ringwall” and “slab” moments:

$$M_T = \sqrt{[m_I h_I S_a(T_I)]^2 + \sum_{n=1}^{\infty} [m_{Cn} h_C S_a(T_{Cn})]^2} \quad (43)$$

$$M'_T = \sqrt{[m_I h'_I S_a(T_I)]^2 + \sum_{n=1}^{\infty} [m_{Cn} h'_C S_a(T_{Cn})]^2} \quad (44)$$

One should note that the impulsive force is significantly higher than the convective force. Therefore, in most of the cases, refinements on the value of the convective force and the combination of the corresponding modes are of little importance on the calculation of the total seismic force (Pappa et al. 2011).

Meridional Stresses

The meridional stress, σ_z , in mechanically anchored tanks is related to meridional membrane force, N , per unit circumferential length:

$$\sigma_z = N/t_i \quad (45)$$

where t_i is the thickness of the tank shell course under consideration and the axial force N per unit circumferential length is given by the following equations:

- On the compression side of the tank (Fig. 8 – location 1):

$$N = -1.273 \frac{M_T}{D^2} - w_t(1 + cA_v) \quad (46)$$

- On the tension side of the tank (Fig. 8 – location 2):

$$N = 1.273 \frac{M_T}{D^2} - w_t(1 - cA_v) \quad (47)$$

where A_v is the spectral acceleration coefficient for the vertical motion in [% g] ($A_v = 100a_{gv}/g$), w_t is the load per unit circumferential length because of shell and roof weight acting at the base of shell [N/m] calculated by

$$w_t = \frac{W_s}{\pi D} + w_{rs} \quad (48)$$

and c is a factor that reflects the contribution of the vertical component. In Eq. 48 W_s is the total weight of the tank shell and appurtenances and w_{rs} is the load acting on the shell roof. Note that

the first term in Eqs. 46 and 47 is directly obtained from simple mechanics of materials considering bending stresses of a beam of circular cross section.

In API 650 (American Petroleum Institute 2007), the value of c is taken equal to 0.40, whereas EN 1998-4 (European Committee for Standardization 2006b) does not give an explicit expression for the effect of the vertical ground acceleration component on the meridional stresses.

Hoop Hydrodynamic Stresses

In the course of seismic loading, the hydrodynamic pressure loading on the tank wall, apart from the development of significant meridional stresses, results also in the development of additional stresses in the circumferential direction, denoted as σ_h , called hoop hydrodynamic stresses (Fig. 9).

Explicit equations for computing hoop hydrodynamic stresses for the convective and impulsive component of liquid motion (denoted as σ_{hc} and σ_{hl} respectively), in terms of the z coordinate of the tank and the D/H (aspect) ratio of the tank, are reported by Wozniak and Mitchell (1978), also adopted by API 650 (American Petroleum Institute 2007), Appendix E, Section E.6.1.4, as follows:

For tanks with $D/H \geq 1.333$:

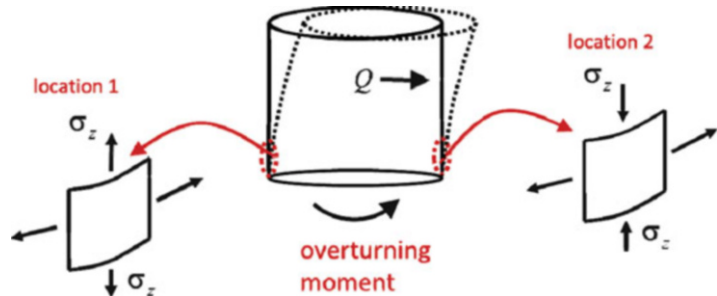
$$\sigma_{hl} = \frac{8.48A_fGDH}{t_i} \left[\frac{z_1}{H} - 0.5 \left(\frac{z_1}{H} \right)^2 \right] \tanh \left(0.866 \frac{D}{H} \right) \quad (49)$$

For tanks with $D/H < 1.33$ and $z_1 < 0.75D$:

Liquid Storage Tanks:

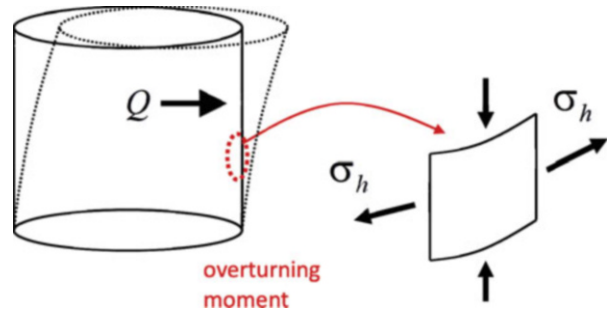
Seismic Analysis,

Fig. 8 Meridional stresses σ_z on the tension side (location 1) and the compression side (location 2) of the tank under seismic loading



Liquid Storage Tanks: Seismic Analysis,

Fig. 9 Hoop hydrodynamic stresses σ_h



$$\sigma_{hl} = \frac{5.22A_I G D^2}{t_i} \left[\frac{z_1}{0.75D} - 0.5 \left(\frac{z_1}{0.75D} \right)^2 \right] \quad (50)$$

For tanks with $D/H < 1.33$ and $z_1 \geq 0.75D$:

$$\sigma_{hl} = \frac{2.6A_I G D^2}{t_i} \quad (51)$$

For all proportions of D/H :

$$\sigma_{hc} = \frac{1.85A_C G D^2 \cosh \left[\frac{3.68(H - z_1)}{D} \right]}{t_i \cosh \left[\frac{3.68H}{D} \right]} \quad (52)$$

In the above equations, stresses are expressed in [MPa], D is the nominal tank diameter in [m], H is the maximum design product level in [m], t_i is the thickness of the shell ring under consideration in [mm], $z_1 = H - z$ is the distance from the liquid surface in [m], G is the specific gravity of tank contents, and A_I and A_C are the impulsive and convective spectral acceleration coefficients in [% g] defined as $A_I = 100S_a(T_I)/g$ and $A_C = 100S_a(T_C)/g$.

The hoop stresses due to impulsive and convective motion should be combined, using the SRSS rule, and superimposed to the hydrostatic stress σ_{hs} to obtain the total hoop stress as follows:

$$\sigma_h = \sigma_{hs} + \sqrt{\sigma_{hl}^2 + \sigma_{hc}^2} \quad (53)$$

where σ_h is the total hoop stress in the shell, to be used for hoop strength design.

When vertical acceleration is taken into account, the above equation is written as follows:

$$\sigma_h = \sigma_{hs} + \sqrt{\sigma_{hl}^2 + \sigma_{hc}^2 + (A_v \sigma_{hs})^2} \quad (54)$$

where A_v is the spectral acceleration coefficient for the vertical motion in [% g].

Sloshing Wave Height

From hydrodynamics, the elevation of the liquid free surface (sloshing wave height) is given by the following expression:

$$\eta(r, \theta, t) = \sum_{n=1,2,3,\dots}^{\infty} \frac{2\lambda_n R J_1(\lambda_n \xi) \tanh(\lambda_n \gamma) \cos \theta a_{Cn}}{J_1(\lambda_n)(\lambda_n^2 - 1)} \quad (55)$$

In this expression, $\lambda_1 = 1.841$, $\lambda_2 = 5.331$, and $\lambda_3 = 8.536$ are the roots of $J'_1(\lambda) = 0$, and a_{Cn} are the generalized coordinates of the convective motion with respect to the tank base given by Eq. 14. Considering only 1 term ($n = 1$), $r = R$, one obtains Eq. 56 for the maximum elevation $\eta_{\max}$:

$$\eta_{\max} = 0.84RS_a(T_C)/g \quad (56)$$

that can be used for design purposes.

Vertical Seismic Action

The hydrodynamic pressure on the walls of a rigid tank due to vertical ground acceleration $A_v(t)$ can be computed by the following equation:

$$p_v(\zeta, t) = \rho H(1 - \zeta) \ddot{X}_v(t) \quad (57)$$

where $\ddot{X}_v(t)$ is the vertical ground acceleration input.

The corresponding hydrodynamic pressure distribution is axisymmetric and does not produce a shear force or moment resultant at any horizontal level of the tank or immediately above or below the base.

In the course of a spectral dynamic analysis, the peak vertical ground acceleration a_{gv} can be used for the calculation of p_v in the above equation replacing $\ddot{X}_v(t)$.

If one assumes that the tank is not moving rigidly with vertical acceleration $\ddot{X}_v(t)$, it may be reasonable to assume a contribution to the hydrodynamic pressure, $p_{vf}(\zeta, t)$, due to the deformability (radial “breathing”) of the shell as suggested by Fischer and Seeber (1988). This additional term may be calculated as

$$p_{vf}(\zeta, t) = 0.815f(\gamma)\rho H \cos\left(\frac{\pi}{2}\zeta\right)a_{vf}(t) \quad (58)$$

where

$$f(\gamma) = 1.078 + 0.274\ln\gamma \quad \text{for } 0.8 \leq \gamma < 4 \quad (59)$$

$$f(\gamma) = 1.0 \quad \text{for } \gamma < 0.8 \quad (60)$$

and $a_{vf}(t)$ is the acceleration response of a simple oscillator having a frequency equal to the fundamental frequency ω_{vd} of the axisymmetric vibration of the tank containing with the liquid. This fundamental frequency ω_{vd} can be calculated by the corresponding fundamental period, estimated by the expression

$$T_{vd} = 4R \left[\frac{\pi\rho H(1 - \gamma^2)I_0(\gamma_1)}{2EI_1(\gamma_1)t_{eq}} \right]^{1/2} \quad (\text{for } \zeta = 1/3) \quad (61)$$

where $\gamma_1 = \frac{\pi}{2\gamma}$ and $I_0(\cdot)$, $I_1(\cdot)$ denote the modified Bessel function of order 0 and 1, respectively. In the course of a spectral dynamic analysis, the maximum value of $a_{vf}(t)$ is obtained from the vertical acceleration response spectrum for the appropriate values of period and damping. If soil flexibility is neglected, the applicable

damping values are those of the material of the shell. Furthermore, the behavior factor value, q , adopted for the response due to the impulsive component of the pressure and the tank wall inertia may be used for the response to the vertical component of the seismic action. The maximum value of the pressure due to the combined effect of p_v and p_{vf} may be obtained by applying the SRSS combination rule to the individual maxima.

Special Issues on Seismic Action

Unanchored Tanks

Numerous aboveground tanks are not anchored to the foundation and, therefore, in the course of a strong earthquake event, uplift of tank bottom from the ground may occur due to seismic overturning moment. Uplift may cause significant plastic deformations in the tank, especially in its base plate or in the area of nozzles and piping attachments, leading to shell wall fracture and leakage of the liquid.

Base plate uplifting in those tanks referred to as unanchored or self-anchored, apart from its effect on the base plate connection may lead to an increase of the compressive vertical stress in the shell, which is critical for buckling-related modes of failure. At the wall side opposite to the uplifted area, vertical compression is maximum and hoop compressive stresses are generated in the shell, due to the membrane action of the base plate. The uplifting parameters of the tank can be seen in (Fig. 10).

The vertical uplift of two tanks with aspect ratios $\gamma = 1.131$ and 0.783 at the edge of the tank base, w , derived from a nonlinear finite element model, is given in Fig. 11 as a function of the normalized overturning moment M/WH (Vathi and Karamanos 2014).

API 650 (American Petroleum Institute 2007) provisions for tank uplifting effects are based on the value of the anchorage ratio J , defined as follows:

$$J = \frac{M_T}{D^2[w_t(1 - 0.4a_{gv}) + w_a - 0.4w_{int}]} \quad (62)$$

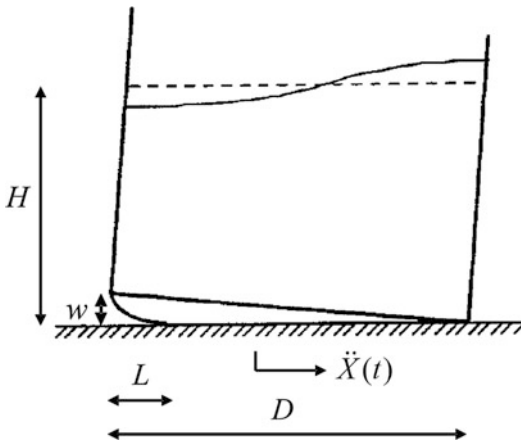
where w_a is the force per unit circumferential length resisting uplift in the annular region and w_{int} the uplift load per unit circumferential length due to product pressure. The values of w_a and w_{int} can be computed by appropriate formulae in API 650.

Tanks with $J < 0.785$ do not uplift, whereas if $J > 1.54$, the tank is unstable and mechanical anchorage is necessary.

An estimate of the membrane stress σ_{rb} in the base plate due to uplift can be obtained by Eq. 63:

$$\sigma_{rb} = 0.568E \left(\frac{pL(w)}{Et} \right)^{2/3} \quad (63)$$

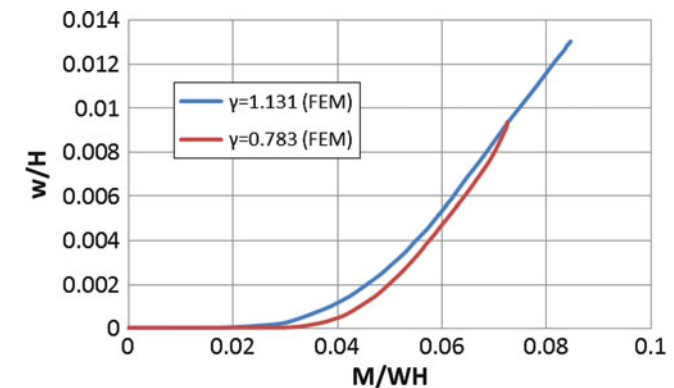
proposed in EN 1998-4 (European Committee for Standardization 2006b) and written herein in



Liquid Storage Tanks: Seismic Analysis,
Fig. 10 Uplifting parameters of an unanchored tank

Liquid Storage Tanks: Seismic Analysis,

Fig. 11 Maximum vertical uplift of the tank w versus overturning moment M of two fixed-roof unanchored cylindrical tanks; finite element results (Vathi and Karamanos 2014)



more convenient terms. In this equation, L is the uplifted length of the tank which is a function of the uplifting size w , so that $L = L(w)$. The corresponding strain can be computed as follows:

$$\varepsilon_{rb} = \frac{\sigma_{rb}}{E} = 0.568 \left(\frac{pL(w)}{Et} \right)^{2/3} \quad (64)$$

When significant uplift takes place in large-diameter tanks, the state of stresses in the uplifted part of the base plate at the ultimate limit state is dominated by plate bending (including the effect of the pressure acting on the tank base), not by membrane stresses. In such cases, a finite element analysis method should be used for the calculation of the state of stresses.

As far as the plastic rotation of the base plate connection is concerned, it is recommended to design the bottom annular ring with a thickness less than the wall thickness, so as to avoid flexural yielding at the base of the wall.

Finally, the plastic rotation associated to uplift w at the edge of the tank and a base separation of L can be computed as follows:

$$\theta_p = \left(\frac{2w}{L} - \frac{w}{2R} \right) \quad (65)$$

which should be less than the rotation capacity of the welded connection.

Forces on Anchors

In the case of mechanically anchored tanks, the seismic design load, P_{AB} , at each anchor

necessary to design the anchor can be calculated by the following simple equation:

$$P_{AB} = N \left(\frac{\pi D}{n_A} \right) \quad (66)$$

where N is the maximum tensile force per unit circumferential length given in Eq. 47 and n_A is the number of anchors around the tank circumference.

Soil-Tank Interaction

For tanks on relatively deformable ground, the base motion can be significantly different from the free-field motion; in general the translational component is modified and there is also a rocking component. Moreover, for the same input motion, as the flexibility of the ground increases, both the fundamental period of the tank-fluid system and the total damping increase, reducing the peak seismic force. The increase in the fundamental period is more pronounced for tall, slender tanks, because the contribution of the rocking component is greater. The reduction of the peak seismic force, however, is less pronounced for tall tanks, since the damping associated with rocking is smaller than that associated with horizontal translation.

The convective periods and pressures are assumed not to be affected by soil-structure interaction. A simple procedure, initially proposed for buildings and consisting of an increase of the fundamental period and of the damping of the structure on a rigid soil, has been extended to the impulsive (rigid and flexible) components of the response of tanks (Habenberger and Schwarz 2002). A good approximation can be obtained through the use of an equivalent simple oscillator with parameters adjusted to match frequency and peak response of the actual system. The properties of this “substitute oscillator” are given in the form of graphs, as functions of the ratio H/R described in the above publications.

An alternative procedure to account for soil flexibility has been suggested by Priestley et al. (1986). The procedure suggests the modification of the frequency and the damping of the impulsive motion, based on the work of

Gazetas (1983), and the corresponding equations for the modified natural periods and damping values can be found in Annex A, Section A.7.2.2 of EN 1998-4 (European Committee for Standardization 2006b).

Seismic Action on Nozzles (Pipe Attachments)

The region of the tank shell where piping is attached to should be designed to resist the forces transmitted by the piping amplified by an appropriate load factor. If reliable information is not available for the attached piping system or an accurate analysis cannot be performed, a relative displacement Δ between the first anchoring point of the piping and the tank should be assumed to take place in the most adverse direction and used for the analysis of the nozzle, with a value, as suggested in EN 1998-4 (European Committee for Standardization 2006b):

$$\Delta = \frac{x}{x_0} d_g \quad (67)$$

In the above expression, x_0 is a reference length equal to 500 m, x the distance between the anchoring point of the piping and the point of connection with the tank, and d_g the design ground displacement.

An alternative methodology for calculating actions on piping attachments is stated by API 650 (American Petroleum Institute 2007). Piping systems shall provide a minimum imposed displacement in the piping, supports, and tank (Table E-8 of API 650) in lieu of a more detailed analysis of the piping system and its connection.

Seismic Resistance of Liquid Storage Tanks

Limit States (General)

The failures against which liquid storage tanks have to be verified when subjected to seismic loading are:

- Development of elephant's foot buckling at the lower course of the tank

- Buckling at the top of the tank due to the alternating sign of the hydrostatic and hydrodynamic pressures
- Rupture or leakage because of the increased induced hoop hydrodynamic stresses
- Plastification of the base plate in unanchored tanks due to uplifting
- Failure of the anchor bolts
- Failure at the nozzles and piping system, especially in the case of unanchored tanks
- Sliding of the tank
- The connection area of hydraulic/piping systems with the tank (nozzles) shall be verified to accommodate stresses and distortions due to relative displacements between tanks or between tanks and soil, without their functions being impaired (see also provisions for attached piping).

The above modes are examined below in more detail. It is noted that a thorough presentation of each failure mode and its modeling in the course of a seismic design and analysis procedure is out of the scope of the present chapter.

Shell Buckling Mode

The maximum actions (e.g., membrane forces and bending moments, circumferential or meridional) induced in the seismic design shall be less or equal to the resistance of the shell evaluated as in the persistent or transient design situations. For the cylindrical steel shell buckling by vertical (meridional) compression with simultaneous action of hoop tension due to hydrostatic loading (“elephant’s foot mode of failure”) is a critical situation.

Steel tanks with very thin shells have also displayed another type of shell buckling mode involving diamond-shaped buckles at a distance above the base of the tank, usually with low internal pressure.

Observations from previous seismic events have indicated that buckling of the lower courses has occasionally resulted in the loss of tank contents due to weld or attached piping fracture and, in some cases, total collapse of the tank. Therefore, in the course of a seismic design procedure, it should be verified that the tank is safe against local buckling. Thus, to ensure “integrity” of the tank under the seismic action against elephant’s foot buckling:

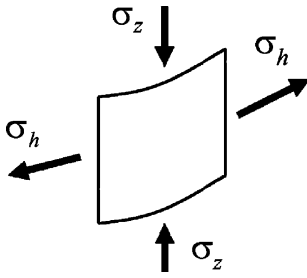
- The geometry (thickness) of the tank should be such that it can withstand significant compressive stresses without buckling in the case of an earthquake.

Section 8.5 and Annex D of EN 1993-1-6 (European Committee for Standardization 2007), paragraphs D.1.2 and D.1.5, give a systematic methodology for the calculation of cylindrical shell resistance against meridional axial compression in the presence of hoop tension due to internal pressure. Moreover, paragraph A.10 of EN 1998-4 (European Committee for Standardization 2006b) provides alternative rules for internally pressurized shell buckling, based on the work by Rotter (1990).

Roof Damage

Sloshing motion of tank liquid contents occurs during earthquake motion. Tank inspections following strong earthquakes indicated that the actual amplitude of liquid surface elevation may be of several meters in some cases. For full or near-full tanks, resistance of the roof to the impact of sloshing waves results in an upward (internal) loading on the roof which may cause buckling of the upper courses of the shell walls. Damage may also occur at the frangible joints between the walls and the ribs of the roof, with accompanying spillage of tank contents over the top of the tank wall.

In practice, a freeboard shall be provided having a height not less than the calculated height of the slosh waves. Freeboard at least equal to the calculated height of slosh waves shall be provided, if the contents are toxic or if spilling could cause damage to piping or scouring of the foundation. Freeboard less than the calculated height of the slosh waves may be sufficient, if the roof is designed for the associated upward pressure or if an overflow spillway is provided to control spilling of contents. Damping devices against sloshing, e.g., grillages or vertical partitions, can be used to reduce sloshing wave height. If the tank has no rigid roof, adequate freeboard is



Liquid Storage Tanks: Seismic Analysis, Fig. 12 Hoop hydrodynamic stresses, σ_h , and meridional stresses, σ_z , on the tank wall

needed to prevent undesirable effects of liquid spilling.

Hoop Stresses (Hydrostatic and Hydrodynamic)

Tensile hoop stresses, σ_h , may increase due to hydrodynamically induced pressures between the fluid and the tank and exceeding allowable hoop stress, leading to fracture and leakage, as shown in Fig. 12. This phenomenon has occurred extensively in old riveted tanks, where leakage at the riveted joints resulted from seismic pressure-induced yielding. Welded steel tanks with butt-welded joints are less vulnerable in such a rupture because of better ductility of the welded joints; however, large tensile hoop stresses, σ_h , can contribute to the premature development of elephant foot buckling near the tank base due to excessive meridional compression, σ_z , and its interaction with the hoop stress σ_h .

Relevant shell design provisions for tensile stress failure exist in all specifications, where allowable stresses and efficiency factors for welds are specified. If overmatched butt-welded joints are used, efficiency factor is usually equal to unity.

Base Plate Connection Failure

In the case of unanchored tanks, uplifting may occur and, instantaneously, the welded connection between the tank bottom plate and the tank shell at the uplifting side of the tank is subjected to significant tensile stresses that may lead to weld failure. This may result in fracture at the base plate welded connection and immediate loss

of tank content. To avoid this failure the plastic rotation at the welded connection, calculated through Eq. 65, should be limited to a certain value. Such limit values are stated in Section A.9 of EN 1998-4 (European Committee for Standardization 2006b). In addition, strong cyclic loading at the weld, associated with plastic rotation less than the allowable plastic rotation of the connection, may result in fracture due to low-cycle fatigue (Prinz and Nussbaumer 2012).

Resistance of Anchor Bolts

Numerous steel tanks, especially tall (slender) tanks, are anchored with bolts or straps. However, these anchors may be insufficient to withstand the total imposed load in strong earthquake events and can be damaged, in the form of anchor pull-out, stretching, or rupture. One should consider that failure of an anchor does not necessarily lead to tank collapse or loss of tank contents. On the other hand, failure of a significant number of bolts may lead to failure of the tank supporting system and eventually to tank overturning and loss of contents.

The resistance of an anchor bolt subjected to tension can be considered as a fraction of the ultimate force, defined as the product of the bolt net area with the ultimate stress of the bolt material. Bolts should also be designed so that their attachment plate to the shell tank does not fail in shear. Finally, the bolts should be adequately anchored and designed against pullout or punching shear failure.

Failure at Attached Piping Locations

Piping systems connected to tanks should be designed in a way that during earthquakes the potential deformation of the connection location with piping is accommodated. If not, fracture may occur at this area, at the vicinity of the weld leading to loss of content (Wieschollek, Hoffmeister, and Feldmann 2013). To minimize the risk of such a failure, piping systems should provide sufficient flexibility to avoid release of the product by failure of the piping system. To achieve this, the piping system and supports shall be designed so as to not induce significant mechanical loading on the tank shell attachment.

Local loads at piping connections shall be considered in the design of the tank shell. Moreover, the piping shall be designed to minimize unfavorable effects of interaction between tanks and other structures. Mechanical devices which add flexibility to the piping connection, such as bellows, expansion joints, and other flexible components, may be used when they are designed for seismic loads and displacements.

If there is no flexibility loop between the point where the pipe is independently (rigidly) supported and the nozzle, and there are relative anchor motions, the tank nozzle and tank shell should be appropriately checked. Construction of a flexibility loop in the pipe between the tank nozzle and the independent piping supports is always beneficial.

Stability Against Sliding

In general, sliding of the tank should be prevented during a seismic event. The resisting shear force at the interface of the base of the structure and the foundation shall be evaluated taking into account the effects of the vertical component of the seismic action.

For the evaluation of the sliding resistance, the transfer of the total lateral shear force between the tank and the subgrade shall be considered. For self-anchored (unanchored) flat-bottom steel tanks, the overall horizontal seismic shear force shall be resisted by the friction between the tank bottom and the foundation or subgrade. Self-anchored storage tanks shall be proportioned such that the seismic base shear, Q , does not exceed a limit value Q_s , defined as follows:

$$Q_s = \mu(W_s + W_r + W_f + W)(1 - 0.4a_{gv}) \quad (68)$$

where μ is an appropriate friction coefficient.

Limited sliding in unanchored tanks may be acceptable, only if the implications of sliding for the connections between the various parts of the structure and between the structure and any piping are taken into account in the tank analysis and the corresponding verifications.

Summary

In this chapter, important information concerning the seismic analysis of circular cylindrical liquid storage tanks is stated. This information includes a short presentation of the current design standards, a short reference to the seismic input in the course of spectral analysis, the seismic actions that occur during a seismic event, and the resistance of the tank to these actions. Both rigid and flexible tanks that are anchored or unanchored may be considered. It is important to remember that liquid-tank interaction is a key issue for tank seismic response. Furthermore, the total motion can be decomposed in an impulsive part that follows the motion of the container and a convective part, associated with free surface liquid sloshing.

Using the analysis described in the present chapter, for determining the seismic action and considering all possible failure modes, the seismic response of a liquid storage tank can be evaluated rationally and effectively towards safeguarding its structural integrity.

Cross-References

- [Earthquake Response Spectra and Design Spectra](#)
- [European Structural Design Codes: Seismic Actions](#)
- [Modal Analysis](#)
- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)

References

- American Petroleum Institute (2003) Seismic design of storage tanks – Appendix E, welded steel tanks for oil storage, API 650, 10th edn. American Petroleum Institute, Washington, DC
- American Petroleum Institute (2007) Seismic design of storage tanks – Appendix E, welded steel tanks for oil storage, API 650, 11th edn. American Petroleum Institute, Washington, DC

- American Society of Civil Engineers (2006) Minimum design loads for buildings and other structures, ASCE 7-05. American Society of Civil Engineers, Reston
- British Standards Institution (1989) Specification for manufacture of vertical steel welded non-refrigerated storage tanks with butt-welded shells for the petroleum industry, BS 2654. British Standards Institution, London, UK
- European Committee for Standardization (2004a) General rules, seismic actions and rules for buildings. Eurocode 8. Part 1, CEN/TC 250, EN 1998-1. European Committee for Standardization, Brussels
- European Committee for Standardization (2004b) Specification for the design and manufacture of site built, vertical, cylindrical, flat-bottomed, above ground, welded, steel tanks for the storage of liquids at ambient temperature and above, EN 14015. European Committee for Standardization, Brussels
- European Committee for Standardization (2006a) Tanks. Eurocode 3. Part 4-2. CEN/TC 250, EN 1993-4-2. European Committee for Standardization, Brussels
- European Committee for Standardization (2006b) Silos, tanks and pipelines. Eurocode 8. Part 4. CEN/TC 250, EN 1998-4. European Committee for Standardization, Brussels
- European Committee for Standardization (2007) Strength and stability of shell structures. Eurocode 3. Part 1-6. CEN/TC 250, EN 1993-1-6. European Committee for Standardization, Brussels
- Fischer FD (1979) Dynamic fluid effects in liquid-filled flexible cylindrical tanks. *Earthq Eng Struct Dyn* 7:587–601
- Fischer FD, Seeber R (1988) Dynamic response of vertically excited liquid storage tanks considering liquid-soil-interaction. *Earthq Eng Struct Dyn* 16:329–342
- Gazetas G (1983) Analysis of machine foundation vibrations: state-of-the-art. *Soil Dyn Earthq Eng* 2(1):2–43
- Habenberger J, Schwarz J (2002) Seismic response of flexibly supported anchored liquid storage tanks. In: *Proceedings of the 13th European conference on earthquake engineering*, London
- Haroun MA (1983) Vibration studies and tests of liquid storage tanks. *Earthq Eng Struct Dyn* 11:179–206
- Housner GW (1957) Dynamic pressures on accelerated fluid containers. *Bull Seismol Soc Am* 47:15–35
- Ibrahim RA (2005) *Liquid sloshing dynamics: theory and applications*. Cambridge University Press, Cambridge, UK
- Karamanos SA, Patkas LA, Platyrrachos MA (2006) Sloshing effects on the seismic design of horizontal-cylindrical and spherical industrial vessels. *J Press Vessel Technol ASME* 128(3):328–340
- Malhotra PK (1995) Base uplifting analysis of flexibly supported liquid-storage tanks. *Earthq Eng Struct Dyn* 24(12):1591–1607
- Manos GC, Clough RW (1982) Further study of the earthquake response of a broad cylindrical liquid-storage tank model, Report No. UCB/EERC-82/7. University of California, Berkeley
- Natsiavas S (1988) An analytical model for unanchored fluid-filled tanks under base excitation. *J Appl Mech ASME* 55:648–653
- Niwa A, Clough RW (1982) Buckling of cylindrical liquid-storage tanks under earthquake excitation. *Earthq Eng Struct Dyn* 10:107–122
- Pappa P, Vasilikis D, Vazouras P, Karamanos SA (2011) On the seismic behaviour and design of liquid storage tanks. In: *III ECCOMAS thematic conference on computational methods in structural dynamics and earthquake engineering*, COMPDYN 2011, Corfu
- Pappa P et al (2012) Structural safety of industrial steel tanks, pressure vessels and piping systems under seismic loading (INDUSE). RFCS project final report, Brussels. www.mie.uth.gr/induse
- Peek R (1988) Analysis of unanchored liquid storage tanks under lateral loads. *Earthq Eng Struct Dyn* 16:1087–1100
- Priestley MJN, Davidson BJ, Honey GD, Hopkins DC, Martin RJ, Ramsey G, Vessey JV, Wood JH (1986) Seismic design of storage tanks, recommendations of a study group of the New Zealand National Society for Earthquake Engineering, Wellington, New Zealand
- Prinz GS, Nussbaumer A (2012) Fatigue analysis of liquid-storage tank shell-to-base connections under multi-axial loading. *Eng Struct* 40:77–82
- Rammerstorfer FG, Fischer FD, Scharf K (1988) A proposal for the earthquake resistant design of tanks – results from the Austrian project. In: *Proceedings of the 9th world conference on earthquake engineering*, vol 6. Tokyo, pp 715–720
- Rammerstorfer FG, Fischer FD, Scharf K (1990) Storage tanks under earthquake loading. *Appl Mech Rev ASME* 43(11):261–283
- Rotter JM (1990) Local inelastic collapse of pressurized thin cylindrical steel shells under axial compression. *ASCE J Struc Eng* 116(7):1955–1970
- Scharf K (1990) Beiträge zur Erfassung des Verhaltens von erdbebenereigten, ober-irdischen Tankbauwerken. *Fortschritt-Berichte VDI. Reihe 4. Bauingenieurwesen*. Nr. 97. VDI Verlag, Düsseldorf
- Vathi M, Karamanos SA (2014) Modelling of uplifting mechanism in unanchored liquid storage tanks subjected to seismic loading. In: *Second European conference on earthquake engineering and seismology (2ECEES)*, Istanbul
- Veletsos AS, Tang Y (1990) Soil-structure interaction effects for laterally excited liquid storage tanks. *Earthq Eng Struct Dyn* 19:473–496
- Veletsos AS, Yang JY (1977) Earthquake response of liquid storage tanks. In: *2nd engineering mechanics conference*, ASCE, Raleigh, pp 1–24
- Wieschollek M, Hoffmeister B, Feldmann M (2013) Experimental and numerical investigations on nozzle reinforcements. In: *Proceedings of the ASME 2013 pressure vessels & piping division conference (PVP2013)*, Paris, 14–18 July 2013
- Wozniak RS, Mitchell WW (1978) Basis of seismic design provisions for welded steel oil storage tanks. *Advances in storage tank design*, API 43rd mid-year meeting, Toronto

Long-Period Moment-Tensor Inversion: The Global CMT Project

Göran Ekström and Meredith Nettles
Department of Earth and Environmental
Sciences, Lamont-Doherty Earth Observatory,
Columbia University, Palisades, NY, USA

Synonyms

Global centroid-moment-tensor inversion; Long-period earthquake analysis

Introduction

Long-period moment-tensor inversion is a methodology to estimate first-order earthquake parameters by inverse modeling of long-period seismograms, typically recorded at teleseismic distances. Data from large numbers of globally distributed digital broadband seismographs and knowledge of elastic wave propagation through the heterogeneous Earth currently make possible the uniform determination of moment tensors of earthquakes with $M \geq 5.0$. The Global CMT Project is an ongoing systematic research effort to study all significant earthquakes occurring worldwide using long-period moment-tensor inversion and to maintain and curate a catalog of the resulting earthquake parameters.

Earthquakes are complex geophysical events with the potential to cause vast destruction. When and where earthquakes occur and a characterization of their size and effects form knowledge of scientific and societal value. For example, systematic analysis of earthquakes led to the mapping of seismic belts and, subsequently, the understanding that most earthquakes occur at the boundaries of tectonic plates. The monitoring and cataloging of earthquake characteristics and their geographical distribution were key ingredients for the development and acceptance of the theory of plate tectonics.

The size of an earthquake is a fundamental descriptive parameter. It was traditionally assessed using the maximum intensity of ground shaking experienced near the earthquake or, following the development and global deployment of seismographs, by measurement of the logarithmic amplitude of weak ground motion recorded at large distances, corrected for distance and attenuation effects – the quantity known as the earthquake magnitude. Since most earthquakes are not felt and the determination of a magnitude is an analysis step that often accompanies earthquake detection and location, the seismic magnitude is the most commonly reported measure of earthquake size. While useful in determining the relative sizes of earthquakes and for estimating and predicting the frequency of earthquakes of a given magnitude using observed frequency–magnitude relationships, traditional magnitude scales are phenomenological, empirical, and imprecise and cannot be directly related to physical properties of the earthquake.

The fundamental insight that earthquakes reflect the sudden relaxation of tectonic stresses across faults dates back more than a century. A deeper physical and quantitative understanding of the relationship between fault motion and seismic-wave generation developed in the 1960s and 1970s. In particular, the seismic moment M_0 was identified as the central physical parameter that describes the size of an earthquake. An earthquake that involves average relative motion $\bar{d}$ of the two sides of a fault over an area A has a seismic moment $M_0 = \mu A \bar{d}$, where μ is the rigidity of the rock surrounding the fault. The seismic moment carries units of force $\times$ distance and represents the strength of two perpendicular force couples with no net rotation acting at the hypocenter. The double-couple model of forces is consistent with the notion that the elastic rebound in an earthquake is the response to sudden unrestrained stress at the earthquake fault. This “stress glut” is a tensor quantity closely related to the moment tensor of the earthquake. It reflects the strength and three-dimensional geometry of fault motion. From the seismic moment given in Nm, a

moment magnitude M_W (sometimes simply M) can be calculated using a standard formula,

$$M_W = \frac{2}{3}(\log_{10} M_0 - 9.1), \quad (1)$$

designed to agree, on average, with traditional earthquake magnitudes for moderate earthquakes (Kanamori 1977). The seismic moment and the moment magnitude are the preferred modern measures of earthquake size. The moment-tensor elements are directly related to the strike and dip angles of the earthquake fault and the rake angle of fault motion in the earthquake.

The development of detailed models of the layered elastic and density structure of the Earth in the 1970s and 1980s made possible the accurate forward calculation of the seismic wavefield caused by forces acting in the Earth. In particular, in the formalism of the Earth's long-period elastic normal modes, the excitation of seismic waves by body forces and moment-tensor sources is straightforward (Gilbert 1971). Linear elastic wave theory thus provides the framework in which earthquake source characteristics can be deduced from observed seismic waves. At long periods, the seismic wavefield is simple and can be explained by a small number of fundamental earthquake parameters.

Improvements in the digital recording of broadband ground motion on a global scale during the 1980s provided the observations needed for systematic application of inverse modeling methods to characterize global earthquake activity. At present, several groups routinely determine moment tensors and seismic moments for earthquakes worldwide. Knowledge about global seismic activity collected and derived from these efforts is important for improved understanding of earthquake sources, active tectonics, and global seismic hazard.

Long-Period Earthquake Analysis

Earthquake fault rupture is complex. The slip on the fault is distributed heterogeneously over a

surface, and the seismic rupture and slip take time to complete. Seismic waves excited by the earthquake reflect the full complexity of the earthquake forces, but the observed vibrations average out the details in a way that depends on period and wavelength. At long periods and long wavelengths, the effect of the entire earthquake is well represented by a small number of point-source parameters. Analogously to distributions of other physical quantities, such as masses or forces, it is useful to expand the description of the earthquake source using spatial and temporal moments. Only the lowest moments of the source influence the long-period source radiation. At the same time, the lowest moments carry fundamental information about the earthquake source including the seismic moment and the fault geometry.

The Moment Tensor and Its Centroid

The spatial and temporal complexity of an earthquake can be described by a distribution of stress glut in space and time, $\dot{\Gamma}_{ij}(\mathbf{r}, t)$. The zeroth-order moment tensor is the time and space integral of $\dot{\Gamma}_{ij}(\mathbf{r}, t)$,

$$M_{ij} = M_{ij}^{(0,0)} = \int_{t_1}^{t_2} \int_V \dot{\Gamma}_{ij}(\mathbf{r}, t) dV dt, \quad (2)$$

where the zeros (0,0) in the superscript refer to the zeroth spatial and temporal moments, respectively. In spherical coordinates, the elements of the moment tensor are

$$\mathbf{M} = \begin{bmatrix} M_{rr} & M_{r\theta} & M_{r\varphi} \\ M_{\theta r} & M_{\theta\theta} & M_{\theta\varphi} \\ M_{\varphi r} & M_{\varphi\theta} & M_{\varphi\varphi} \end{bmatrix}, \quad (3)$$

with (r, θ, φ) denoting (up, south, east). Just as the stress-glut-rate tensor, the moment tensor is symmetric:

$$M_{r\theta} = M_{\theta r}, M_{r\varphi} = M_{\varphi r}, M_{\theta\varphi} = M_{\varphi\theta}.$$

For a planar fault with a strike angle ϕ and dip angle δ and with earthquake fault motion

described by the rake angle λ , the moment-tensor elements are

$$\begin{aligned} M_{rr} &= M_0 \sin 2\delta \sin \lambda \\ M_{\theta\theta} &= -M_0(\sin \delta \cos \lambda \sin 2\phi + \sin 2\delta \sin \lambda \sin^2 \phi) \\ M_{\varphi\varphi} &= M_0(\sin \delta \cos \lambda \sin 2\phi - \sin 2\delta \sin \lambda \cos^2 \phi) \\ M_{r\theta} &= -M_0(\cos \delta \cos \lambda \cos \phi + \cos 2\delta \sin \lambda \sin \phi) \\ M_{r\varphi} &= M_0(\cos \delta \cos \lambda \sin \phi - \cos 2\delta \sin \lambda \cos \phi) \\ M_{\theta\varphi} &= -M_0\left(\sin \delta \cos \lambda \cos 2\phi + \frac{1}{2} \sin 2\delta \sin \lambda \sin 2\phi\right), \end{aligned} \quad (4)$$

where M_0 denotes the seismic moment of the earthquake.

A point-source representation of a seismic source can be expressed as

$$\dot{\Gamma}_{ij}(\mathbf{r}, t) = M_{ij} \delta^3(\mathbf{r} - \mathbf{r}_s) \delta(t - t_s), \quad (5)$$

where $(\mathbf{r}_s, t_s)$ is the location and time of the instantaneous earthquake. The location of the point source that is most representative of a given stress-glut distribution can be defined by considering the first moments of the stress-glut-rate distribution. The first spatial moment of $\dot{\Gamma}_{ij}(\mathbf{r}, t)$ around some reference location $\mathbf{r}_h$ is defined as

$$M_{ijp}^{(1,0)} = \int_{t_1}^{t_2} \int_V \dot{\Gamma}_{ij}(\mathbf{r}, t) (r_p - r_{hp}) dV dt, \quad (6)$$

where the subscript p represents one of the three spatial coordinates (r, θ, φ) . The first temporal moment around some reference time t_h is

$$M_{ij}^{(0,1)} = \int_{t_1}^{t_2} \int_V \dot{\Gamma}_{ij}(\mathbf{r}, t) (t - t_h) dV dt. \quad (7)$$

The time and location for which the sum of the squares of these four moments is minimum can be considered the center of the stress-glut-rate distribution and defines the moment-tensor centroid. A point-source representation of an earthquake is therefore given by the following ten parameters: the six independent elements of M_{ij} , the centroid location $\mathbf{r}_c$ (latitude, longitude, depth), and the centroid time t_c .

A common extension to the point-source representation in Eq. 5 is to incorporate a duration of faulting or, more generally, source activity. In a moment expansion of the stress glut, this corresponds to a nonzero second temporal moment of the stress glut, $M_{ij}^{(0,2)}$. A convenient way to incorporate this complexity is to introduce a source-time function $S(t)$ and write the source representation as

$$\dot{\Gamma}_{ij}(\mathbf{r}, t) = M_{ij} \delta^3(\mathbf{r} - \mathbf{r}_s) \delta(t - t_s) * S(t; \tau_h), \quad (8)$$

where $*$ indicates convolution and $S(t; \tau_h)$ is a function of unit area, symmetric around $t = 0$, with a total duration $2\tau_h$.

Synthetic Seismograms and Inversion

The relationship between an instantaneous moment-tensor point source and the ground motion $u(\mathbf{r}, t)$ at an arbitrary point on the Earth is illustrated by the expression describing the excitation and summation of the Earth's normal modes,

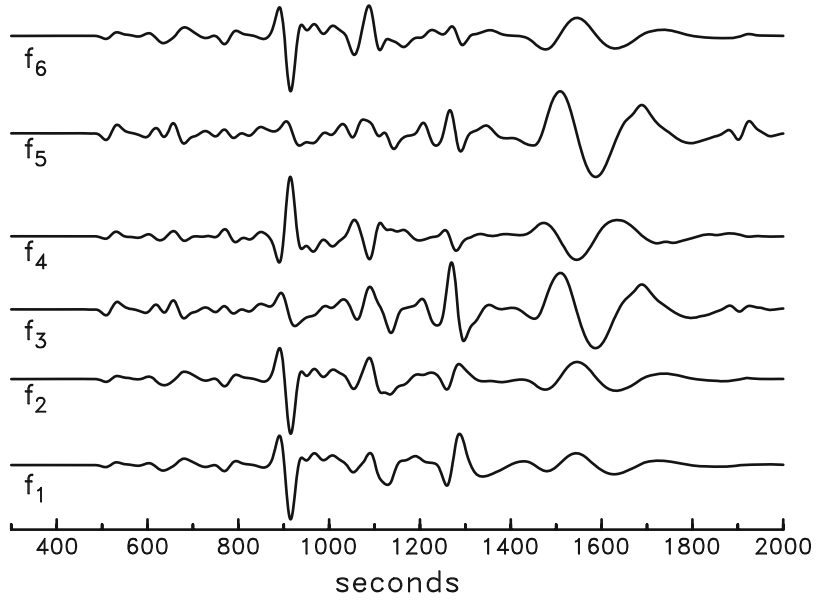
$$u(\mathbf{r}, t) = \sum_k \left[1 - \cos \omega_k(t - t_s) e^{-\alpha_k(t - t_s)} \right] \mathbf{M} : \mathbf{e}_k(\mathbf{r}_s) \mathbf{s}_k(\mathbf{r}), \quad (9)$$

where the summation is over modes k , ω_k and α_k are the normal-mode frequency and decay constants, $\mathbf{e}_k(\mathbf{r}_s)$ is the strain of the k th mode at the source location, and $\mathbf{s}_k(\mathbf{r})$ is the displacement of the k th mode at the receiver location. From this expression, it is clear that there is a linear relationship between the moment-tensor elements and the resulting ground motion. The linear relationship is not limited to the normal-mode formulation but is a consequence of linear elastic theory.

For convenience, the six independent elements of the moment tensor can be written as a vector f_i , with $f_1 = M_{rr}$, $f_2 = M_{\theta\theta}$, $f_3 = M_{\varphi\varphi}$, $f_4 = M_{r\theta}$, $f_5 = M_{r\varphi}$, and $f_6 = M_{\theta\varphi}$. Ground motion u_k at a location $\mathbf{r}$ in a direction given by the subscript k and caused by a moment-tensor

Long-Period Moment-Tensor Inversion: The Global CMT Project,

Fig. 1 Kernel east–west seismograms for a hypothetical deep earthquake ($h = 500$ km) in Bolivia recorded at a station in Los Angeles, California. Each trace corresponds to the motion associated with a single moment-tensor element f_i . The ground motion has been filtered between 30 and 300 s



point source at $\mathbf{r}_s$ with a time history $S(t; \tau_h)$ can then generally be written as

$$u_k(\mathbf{r}, t) = \sum_{i=1}^6 \psi_{ik}(\mathbf{r}, \mathbf{r}_s, t) * S(t; \tau_h) \cdot f_i, \quad (10)$$

where the kernel seismograms $\psi_{ik}(\mathbf{r}, \mathbf{r}_s, t)$ represent the individual contributions to the seismogram from each moment-tensor component i . Figure 1 shows the six moment-tensor kernels for a hypothetical deep earthquake in Bolivia recorded on the east–west component of a seismometer in Los Angeles, California. The different shapes of the kernels illustrate the linear independence of the different moment-tensor contributions to any recorded ground motion.

Equation 10 and Fig. 1 suggest the least-squares minimization problem

$$\min \left(\int_{t_1}^{t_2} [o_k(t) - u_k(t)]^2 dt \right) \quad (11)$$

to estimate the six moment-tensor elements f_i , where $o_k(t)$ represents the observed seismogram, t_1 and t_2 define a signal time window, and it is understood that the observed and synthetic waveforms have been filtered in the same way. Expressed as in Eqs. 10 and 11, the inverse

problem for f_i is strictly linear and can be solved in a single step. However, this formulation is based on a fixed source location ($\mathbf{r}_s, t_s$). In practice, this location is a reported hypocenter that may not correspond closely to the optimal moment-tensor point-source location and time, the centroid ($\mathbf{r}_c, t_c$). To estimate the centroid as well as the moment-tensor elements, additional steps are needed.

When an initial estimate of the moment-tensor elements has been obtained by solving the inverse problem in Eq. 11, the change in the synthetic seismograms with respect to small changes in source location and the moment-tensor elements can be written as

$$\begin{aligned} \delta u_k(t) = u_k(t) - u_k^{(0)}(t) = & b_k(t) \delta r_s \\ & + c_k(t) \delta \theta_s + d_k(t) \delta \varphi_s + e_k(t) \delta t_s \\ & + \sum_{i=1}^6 \psi_{ik}^{(0)}(t) * S(t) \cdot \delta f_i, \end{aligned} \quad (12)$$

where the superscript (0) refers to the synthetic seismogram and moment-tensor kernels calculated for the initial source location. The quantities b_k , c_k , d_k , and e_k represent linear-perturbation kernels,

$$\begin{aligned}
b_k(t) &= \sum_{i=1}^6 \frac{\partial \psi_{ki}^{(0)}}{\partial r_s}(t) * S(t) \cdot f_i^{(0)} \\
c_k(t) &= \sum_{i=1}^6 \frac{\partial \varphi_{ki}^{(0)}}{\partial \theta_s}(t) * S(t) \cdot f_i^{(0)} \\
d_k(t) &= \sum_{i=1}^6 \frac{\partial \varphi_{ki}^{(0)}}{\partial \varphi_s}(t) * S(t) \cdot f_i^{(0)} \\
e_k(t) &= \sum_{i=1}^6 \frac{\partial \varphi_{ki}^{(0)}}{\partial t_s}(t) * S(t) \cdot f_i^{(0)},
\end{aligned} \tag{13}$$

reflecting changes in the seismogram due to a small perturbation in the source location. A second step in the inversion can then be formalized by minimizing the remaining misfit between observed and synthetic waveforms with respect to perturbations in the moment-tensor elements as well as the centroid parameters,

$$\min \left(\int_{t_1}^{t_2} \left[\left(o_k(t) - u_k^{(0)}(t) \right) - \delta u_k(t) \right]^2 dt \right). \tag{14}$$

By iteratively updating the moment-tensor and centroid estimates, as well as the corresponding synthetic seismograms and kernels, and inverting the residual signal for additional perturbations in the source parameters, an optimal estimate of the point-source parameters is obtained.

The preceding paragraphs summarize the theoretical developments and technical approaches described by Gilbert (1971) and Dziewonski et al. (1981). Dziewonski et al. (1981) introduced the full formalism for the simultaneous estimation of the moment tensor and the source centroid, giving rise to the concept of the centroid-moment tensor, or the CMT. The CMT concept is general and can in principle be used with any recorded seismograms, provided that the point-source approximation remains valid. Conditions for success are that high-quality, well-calibrated seismograms are available and that synthetic waveforms corresponding to the observed seismograms can be generated with sufficient fidelity to allow a quantitative comparison.

The Global CMT Project

A systematic effort to use long-period centroid-moment-tensor inversion for the study of global seismicity was initiated by A. M. Dziewoński and J. H. Woodhouse at Harvard University in the early 1980s. This sustained analysis effort was known as “the Harvard CMT Project” and the catalog of earthquake parameters that resulted as “the Harvard CMT Catalog.” In 2006, the research effort moved with G. Ekström and M. Nettles to Lamont–Doherty Earth Observatory of Columbia University, where it continues under the name “the Global CMT Project,” with the primary research goal of curating, improving, and extending the CMT catalog of global earthquake focal mechanisms (Ekström et al. 2012).

Proof of Concept

In the first demonstration of the CMT approach (Dziewonski et al. 1981), it was shown that CMTs could be calculated for globally distributed earthquakes as small as $M_w = 5.5$ by least-squares inversion of long-period ($T > 45$ s) body-wave seismograms recorded on a sparse network of stations. The synthetic seismograms were calculated by summation of normal modes for the spherically symmetric Preliminary Reference Earth Model (PREM) (Dziewonski and Anderson 1981). Three specific circumstances explain the success of this approach. First, the general characteristics of the seismic source are such that a point-source approximation is appropriate at the periods considered. For example, a typical duration of an $M = 6.0$ earthquake is 5 s, and a typical fault dimension is 10 km, both of which are small compared with a wave period of ~ 50 s and a wavelength of ~ 200 km. Second, the background seismic noise in the Earth is low at periods of 30–200 s. Even relatively small earthquakes generate ground motion in this period range that is observable at large distances with modern digital instrumentation. Third, while the three-dimensional (3D) elastic heterogeneity of the Earth is significant, the selected part of the long-period seismic wavefield can be simulated with high accuracy using a one-dimensional (1D) Earth model. In particular, seismic body-wave

phases diving deep into the Earth's mantle and core typically do not exhibit travel-time residuals greater than several seconds, which is small with respect to the periods used in the CMT waveform inversion.

Project Evolution

Early applications of the CMT approach were constrained by the very limited availability of high-quality digital seismograms. Only 10–20 stations provided digital data on a global scale in the early 1980s. In addition, although the importance of the Earth's shallow elastic heterogeneity was appreciated, the 3D elastic structure of the Earth was unknown. Methods for synthesizing seismograms in heterogeneous 3D Earth structure were also limited and not appropriate for routine calculations on then-available computers. Rapid progress occurred during the 1980s and 1990s. New networks of globally distributed seismographic stations were deployed, soon providing broadband data in near-real time that could be incorporated in CMT analysis. Low-degree tomographic models of the Earth's upper mantle were developed, and new approximate methods for synthesizing normal-mode seismograms for these models in a computationally efficient manner made possible the inclusion of very long-period ($T > 135$ s) surface waves in the CMT analysis. These waves, often called mantle waves because their primary sensitivity is to elastic structure in the mantle, are particularly useful for the analysis of large earthquakes since their long period and wavelength make the point-source approximation valid even for very large events.

Since 1984, the methodology used in the CMT project for synthesizing normal-mode seismograms in an elliptical and heterogeneous 3D Earth has been based on the path-average approximation. In this approach, the average 1D structure for the minor arc and the great circle connecting the earthquake and each recording station are calculated from an existing 3D Earth model. Small adjustments in the normal-mode frequencies corresponding to deviations from the global-average 1D Earth structure are calculated for each path and applied in the mode sum.

To account for the different average structures along the great-circle and minor-arc paths, a fictitious shift in the geographical location of the earthquake is also applied for each mode, using an approach introduced by Woodhouse and Dziewonski (1984). This method for seismogram synthesis corrects the phase of long-period seismic waves for the effect of large-wavelength 3D heterogeneity to first order and retains the computational efficiency of normal-mode summation.

New generations of tomographic models of the Earth's mantle have led to an improved ability to match long-period body waves and very long-period surface waves (mantle waves). These 3D models have not, however, been adequate for matching the large propagation delays observed for fundamental-mode Love and Rayleigh surface waves at periods shorter than 100 s. In addition, the linear-perturbation approach to correcting the mode sum is inappropriate for heterogeneity that causes travel-time delays of the same order as, or longer than, the period of the wave. The value of including intermediate-period surface waves in the CMT analysis is that these waves frequently are the largest phases in long-period seismograms. Following the development in the late 1990s of high-resolution global dispersion models for Love and Rayleigh waves, the possibility arose to incorporate propagation delays from such maps in the calculation of the fundamental-mode contribution to the synthetic seismograms. In 2004, a hybrid approach to the calculation of synthetic seismograms was introduced in the CMT analysis. The body-wave portion and very long-period mantle-wave portion of the wavefield are synthesized using the path-average normal-mode approach, while intermediate-period minor-arc fundamental-mode surface waves are calculated using a surface-wave approach, with phase delays derived from surface-wave phase-velocity maps (Ekström et al. 1997). The consideration of intermediate-period surface waves allows for the analysis of smaller earthquakes than was previously possible and offers improved constraints on the moment tensors for shallow earthquakes.

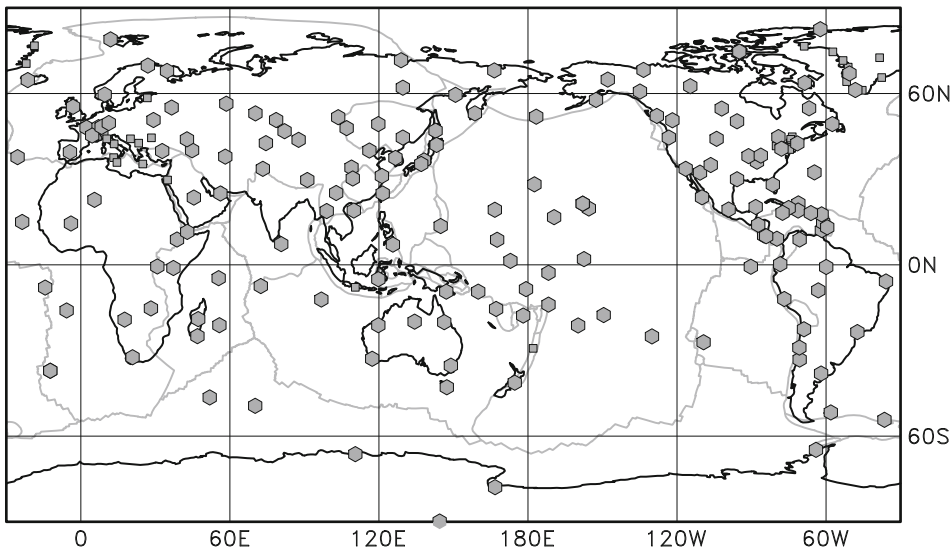
Current Practice

Analysis of seismicity in the Global CMT Project is divided into calendar months. The analysis is initiated 2–3 months after real time and is completed before the end of 4 months. Earthquakes that have the potential to yield robust CMT results are selected for analysis, primarily from the hypocenter catalog of the National Earthquake Information Center (NEIC) of the United States Geological Survey (USGS). The current objective is to attempt analysis for all earthquakes $M_W \geq 5.0$. Because of the large uncertainties in traditional magnitudes, all events with m_b or $M_S \geq 4.8$ are selected for analysis. This typically leads to a monthly list of ~300 candidate earthquakes, of which more than half are successfully analyzed.

Seismograms are collected from the IRIS-USGS Global Seismographic Network (seismic network codes II and IU) with additional data from stations of the China Digital Seismograph Network (IC), the Geoscope (G), Mednet (MN), Geofon (GE), and Caribbean (CU) networks. Normally, more than 150 stations are available for any earthquake. Figure 2 shows a map with the stations contributing to CMT analysis for earthquakes in 2012.

Three-component data from all stations are extracted for each event, and the responses are equalized by deconvolution of the instrument response and convolution with standard band-pass filters that depend on the data type. Rotated horizontal components (longitudinal and transverse) are constructed and considered in the analysis. Three types of data are used: (1) body waves, which are the signals that arrive in the time window before the arrival of the minor-arc surface waves; (2) mantle waves, which consist of up to 4.5 h of data, including several very long-period Love and Rayleigh wave arrivals (e.g., G1–G4 and R1–R4); and (3) intermediate-period surface waves, which consist of a time window centered on the minor-arc arrival times for surface waves.

Filtering strategies and data selection depend on the earthquake magnitude. Body-wave data for earthquakes $5.5 \leq M_W \leq 6.5$ are filtered to displacement between 150 and 40 s. For earthquakes $M_W < 5.5$, the seismograms are filtered to velocity. For earthquakes $M_W \geq 6.5$, the waves are filtered to displacement between 150 and 50 s. Mantle-wave data are considered only for earthquakes $M_W \geq 5.5$, since smaller earthquakes do not produce observable signals in this band.



Long-Period Moment-Tensor Inversion: The Global CMT Project, Fig. 2 Map showing the locations of 206 stations that contributed seismograms to the GCMT

analyses in 2012. Stations that contributed data for more than 200 earthquakes are shown with *hexagons*; other stations are shown with *squares*

For earthquakes $M_W \leq 7.0$, the waves are filtered to displacement between 350 and 125 s. For earthquakes $M_W > 7.0$, the filter is shifted in period with earthquake size so that for earthquakes $M_W > 8.0$, the range is 450–200 s. Intermediate-period surface waves for earthquakes $M_W \geq 5.5$ are filtered to displacement between 150 and 50 s. For smaller earthquakes, the response is modified to velocity, to emphasize the shorter-period signals, which are less obscured by noise. Intermediate-period surface waves are not considered for earthquakes deeper than 300 km.

The three data types provide complementary constraints on the seismic source. Body waves sample the focal sphere of the earthquake in a relatively uniform fashion. Intermediate-period surface waves provide good constraints for smaller earthquakes. The very long-period mantle waves are less sensitive to the spatial extent and temporal history of the seismic source, an advantage for large earthquakes. Seismograms of the three wave types are weighted in the CMT inversion in such a way that the mantle waves do not contribute for earthquakes $M_W < 5.5$, have a relative weight that increases linearly with magnitude for $5.5 \leq M_W \leq 6.5$, and are weighted fully for earthquakes $M_W > 6.5$. Body and surface waves are weighted fully for $M_W < 6.5$, have a decreasing relative weight for $6.5 \leq M_W \leq 7.5$, and do not contribute for earthquakes with $M_W > 7.5$ since, for earthquakes of this size, the signal at shorter periods is dominated by spatial and temporal source complexities that are not accounted for in the CMT inversion.

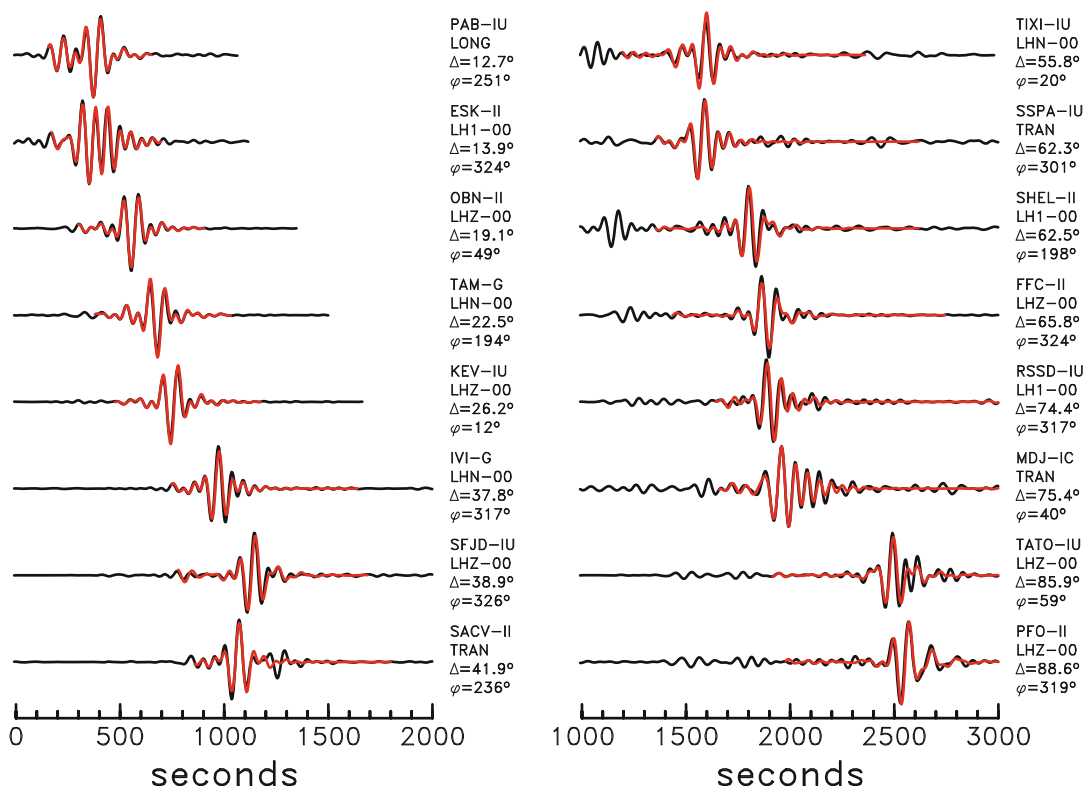
Data selection is necessary to eliminate erroneous seismograms and to exclude traces that are dominated by noise or signals not associated with the earthquake. An automated editing program is used to assist with the data-selection tasks, including assessing observed signal levels and waveform fits in predicted time windows across the network and eliminating faulty and outlying traces. The final inversion results and corresponding fit to the data undergo human review. The validity and quality of the result is assessed based on the total number of seismograms that were fit in the inversion and

the reduction in data variance obtained. The convergence and stability of the results are used as additional criteria. For small earthquakes and occurrences of overlapping signals from two or more earthquakes, careful human editing of the waveforms is required for the generation of stable and reliable results.

An example of the quality of fit achieved between observed and model waveforms is shown in Fig. 3, with surface-wave seismograms observed at a number of stations at different distances and azimuths from the $M_W = 6.1$ Emilia-Romagna, Italy, earthquake of May 20, 2012.

While in theory all six elements of the moment tensor are resolvable from the data, in practice, trade-offs among the diagonal elements of the moment tensor make the isotropic component of the moment tensor, $M_{rr} + M_{\theta\theta} + M_{\varphi\varphi}$, poorly constrained, except for large and deep earthquakes. However, because the stress glut associated with tectonic faulting is not expected to have a significant isotropic component, there is no expectation that the isotropic component for standard earthquakes should be nonzero. In the CMT inversions, the trace of the moment tensor is therefore constrained to equal zero, $M_{rr} + M_{\theta\theta} + M_{\varphi\varphi} = 0$. In addition, for some earthquakes, the focal depth of the source centroid is held fixed in the inversion. A minimum source depth of 12 km is used to reduce instabilities that can occur in long-period moment-tensor inversions for sources located at or near the Earth's surface. When the depth exhibits persistent instability in the inversions, the depth is fixed at the reported hypocentral depth or at a depth deemed appropriate for the source region.

The source half duration is a parameter assumed in the inversion based on the scalar moment of the event, with an initial estimate derived from the reported magnitude. The empirical relationship $\tau_h = 2.26 \times 10^{-6} M_0^{1/3}$ is used to estimate source half duration, where τ_h is the half duration measured in seconds and M_0 is the scalar moment measured in Nm. Before 2004, the moment-rate function used in the CMT inversions was modeled as a boxcar. Since 2004, the moment-rate function has been modeled as an isosceles triangle with half duration τ_h .



Long-Period Moment-Tensor Inversion: The Global CMT Project, Fig. 3 A selection of observed (*black*) and modeled (*red*) waveforms for the May 20, 2012, $M_W = 6.1$ Emilia-Romagna, Italy, earthquake (875 waveforms were used in the CMT inversion). The station, network, and channel of motion are given to the right of each pair of

waveforms, along with the distance Δ and azimuth ϕ from the earthquake epicenter. In the channel designation, LHZ refers to the vertical component, LHN, LHE, and LH1 refer to recorded horizontal components, and TRANS and LONG refer to components rotated into the directions of transverse and longitudinal motion at the station

The Global CMT Catalog

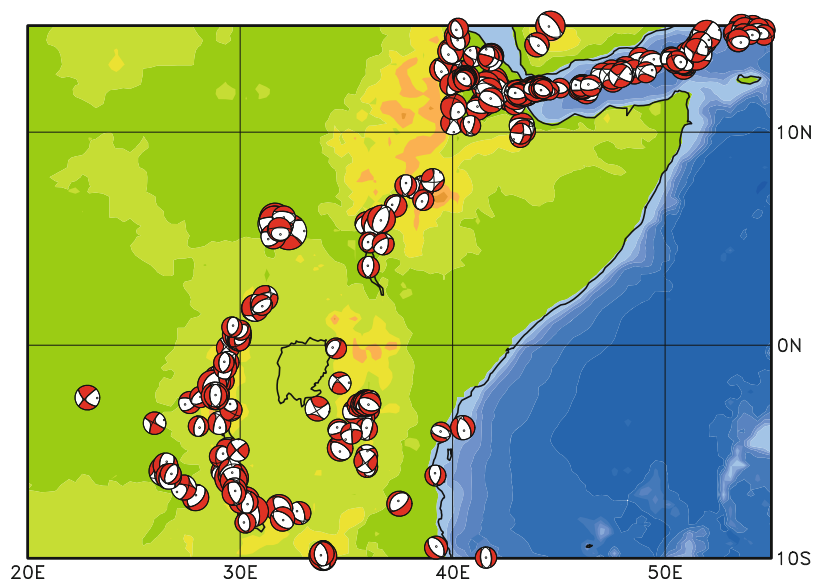
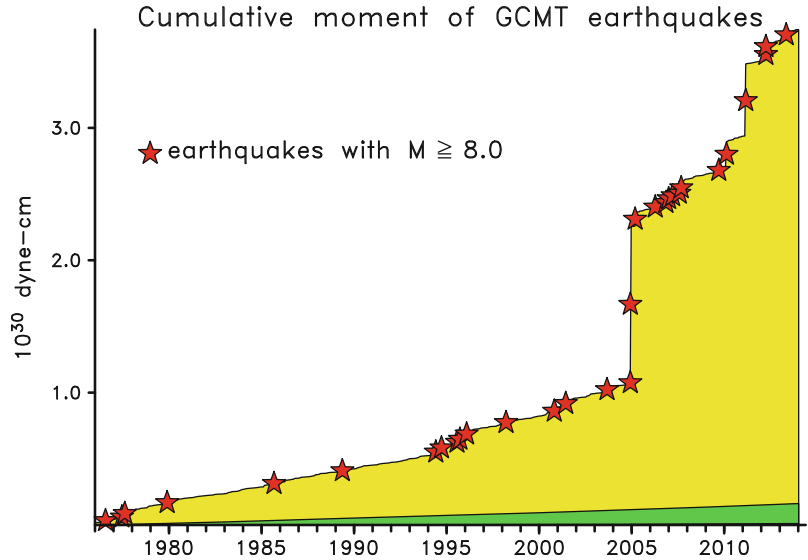
The catalog of earthquake parameters that has resulted from the research effort described in the preceding paragraphs now contains more than 40,000 earthquakes for the period 1976–2013. It is the most comprehensive and internally consistent global collection of moment tensors. Typical (one-sigma) uncertainties in the estimates of seismic moment are probably no larger than 20 %, implying uncertainties in M_W of approximately 0.05 magnitude units. The catalog is uniform in its methodology, and for large earthquakes, it can be considered uniform and complete, with exception for a small number of large earthquakes that have occurred in the coda of even larger mainshocks. The cumulative moment of earthquakes occurring globally since

1976 is shown in Fig. 4. The figure illustrates the well-known observation that most of the moment is associated with large earthquakes. The largest steps in the graph correspond to the 2004 $M_W = 9.3$ Sumatra earthquake, the 2010 $M_W = 8.8$ Chile earthquake, and the 2011 $M_W = 9.1$ Tohoku earthquake.

Figure 5 shows focal mechanisms from the Global CMT catalog for Eastern Africa, illustrating the use of the catalog for characterizing seismicity and the mode of seismic deformation in areas with limited regional earthquake-monitoring capabilities. The dominant normal-faulting focal mechanisms reflect the process of continental rifting and extension in the area.

For moderate and small earthquakes, general improvements in station coverage and the 2004

Long-Period Moment-Tensor Inversion: The Global CMT Project, Fig. 4 Cumulative moment of all earthquakes in the GCMT catalog through the end of 2013. Red stars indicate times of earthquakes of $M_W \geq 8.0$. The area shaded green reflects the moment of earthquakes with $M_W \leq 6.5$

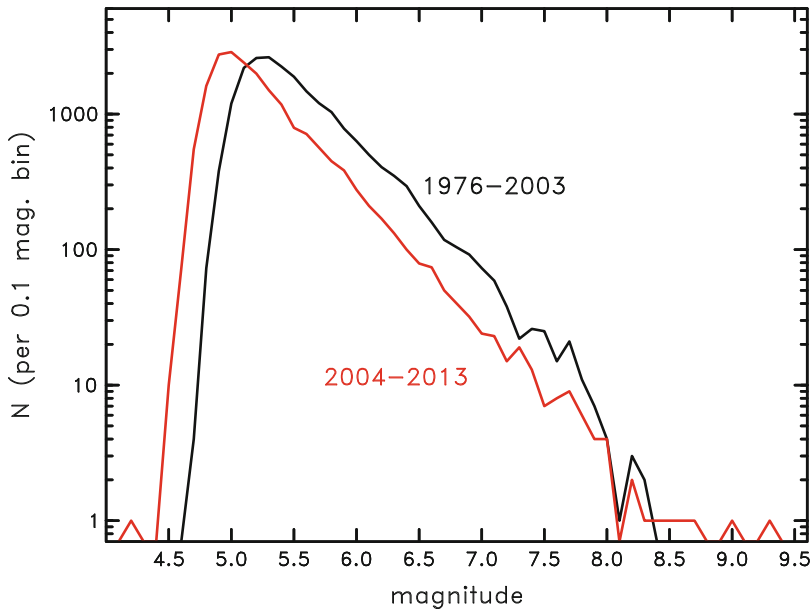


Long-Period Moment-Tensor Inversion: The Global CMT Project, Fig. 5 Map showing the Horn of Africa and surrounding area, with focal mechanisms for earthquakes from the GCMT catalog for the period 1976–2013. Background map shows topography. The quadrants

associated with compressional P -wave motion in the lower-hemisphere focal-mechanism plots are shaded red. The radii of the focal mechanisms increase linearly with M_W . The magnitudes of the earthquakes shown range from 4.7 to 7.1

incorporation of intermediate-period surface waves in the CMT analysis have extended the completeness of the catalog to smaller magnitudes. Figure 6 shows a magnitude–frequency diagram for the catalog divided in two parts:

1976–2003 and 2004–2013. The data for 2004–2013 indicate linearity of the slope of the magnitude–frequency relation (and thus completeness) to $M_W = 5.0$, while the earlier data show completeness to about $M_W = 5.4$.



Long-Period Moment-Tensor Inversion: The Global CMT Project, Fig. 6 Magnitude–frequency diagram for earthquakes in the GCMT catalog for two time periods:

(black line) 1976–2003, (red line) 2004–2013. The number of earthquakes is counted in bins of 0.1-magnitude-unit width

The Global CMT catalog is updated routinely and is available from a number of sources. The primary distribution point is the Global CMT website www.globalcmt.org. The full catalog is also available via the website of the Incorporated Research Institutions for Seismology (IRIS), as well as from the International Seismological Centre (ISC) and the NEIC.

Summary

Long-period moment-tensor inversion is a robust method to characterize the geometry and size of earthquakes worldwide. The linear relationship between the amplitude of long-period seismograms and the seismic moment leads to precise and accurate measurements of earthquake size. Catalogs of moment tensors provide primary data on active faults and the geometry of regional seismotectonic deformation.

Global moment-tensor determination is a mature field. Further advances are, however, anticipated as improved 3D models of the Earth's elastic structure are developed and as methods for

calculating synthetic seismograms in a heterogeneous Earth in a routine manner become available.

Cross-References

- [Earthquake Magnitude Estimation](#)
- [Earthquake Mechanism Description and Inversion](#)
- [Earthquake Mechanisms and Tectonics](#)
- [Moment Tensors: Decomposition and Visualization](#)
- [Regional Moment Tensor Review: An Example from the European–Mediterranean Region](#)
- [Reliable Moment Tensor Inversion for Regional- to Local-Distance Earthquakes](#)

References

- Dziewonski AM, Anderson DL (1981) Preliminary reference Earth model. *Phys Earth Planet Int* 25:297–356
- Dziewonski AM, Chou T-A, Woodhouse JH (1981) Determination of earthquake source parameters from

- waveform data for studies of global and regional seismicity. *J Geophys Res* 86:2825–2852
- Ekström G, Tromp J, Larson EWF (1997) Measurements and global models of surface wave propagation. *J Geophys Res* 102:8137–8157
- Ekström G, Nettles M, Dziewonski AM (2012) The global CMT project 2004–2010: centroid-moment tensors for 13,017 earthquakes. *Phys Earth Planet Int* 200–201:1–9
- Gilbert F (1971) Excitation of the normal modes of the Earth by earthquake sources. *Geophys J Roy Astron Soc* 22(2):223–226
- Kanamori H (1977) The energy release in great earthquakes. *J Geophys Res* 82:2981–2987
- Woodhouse JH, Dziewonski AM (1984) Mapping the upper mantle: three-dimensional modeling of Earth structure by inversion of seismic waveforms. *J Geophys Res* 89:5953–5986

Luminescence Dating in Paleoseismology

Steven L. Forman
Department of Geology, Baylor University,
Waco, TX, USA

Synonyms

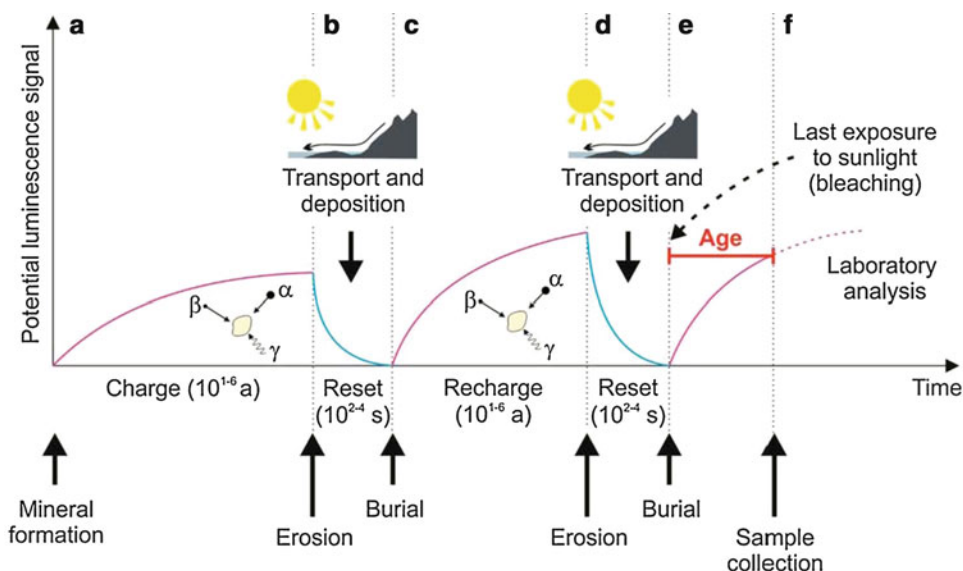
Optical dating; Optically stimulated dating; OSL; Photostimulation

Introduction

Optically stimulated luminescence (OSL) dating or optical dating provides a measure of time since sediment grains were deposited and shielded from further light or heat exposure, which often effectively resets the luminescence signal (Fig. 1). This technique, as thermoluminescence, was originally developed in the 1950s and 1960s to date fired archeological materials, like ceramics (Aitken 1985). Ensuing research in the 1970s documented that marine and other sediments with a prior sunlight exposure of hours to days were suitable for thermoluminescence dating (Wintle and Huntley 1980). Discoveries in the 1980s and 1990s that exposure of quartz and

feldspar grains to a tunable light source, initially with lasers and later by light-emitting diodes, yields luminescence components that are solar reset within seconds to minutes expanded greatly the utility of the method (Huntley et al. 1985; Hutt et al. 1988; Aitken 1998). In the past 15 years, there have been significant advances in luminescence dating with the advent of single aliquot and grain analysis, and associated protocols with blue/green diodes that can effectively compensate for laboratory-induced sensitivity changes (Murray and Wintle 2003; Wintle and Murray 2006; Duller 2012) and render accurate ages for the past ca. 100,000 years. Most recently, the development of protocols for inducing the thermal transfer of deeply trapped electrons has extended potentially OSL dating to the 10^6 year timescale for well solar-reset quartz and potassium feldspar grains from eolian and littoral environments (Duller and Wintle 2012).

Common silicate minerals like quartz and potassium feldspar contain lattice-charge defects formed during crystallization and from subsequent exposure to ionizing radiation. These charge defects are potential sites of electron storage with a variety of trap-depth energies. A subpopulation of stored electrons with trap depths of ~ 1.3 – 3 meV is a subsequent source for time-diagnostic luminescence emissions. Free electrons are generated within the mineral matrix by exposure to ionizing radiation from the radioactive decay of daughter isotopes in the ^{235}U , ^{238}U , and ^{232}Th decay series and a radioactive isotope of potassium, ^{40}K , with lesser contributions from the decay of ^{85}Rb and cosmic sources. The radioactive decay of ^{40}K releases beta and gamma radiation, whereas the decay in the U and Th series generates mostly alpha particles and some beta and gamma radiation. Alpha particles are about 90–95 % less effective in inducing luminescence compared to beta and gamma radiation. Thus, the population of stored electrons in lattice-charge defects increases with prolonged exposure to ionizing radiation, and the resolved luminescence emission increases with time. Exposure of mineral grains to light or heat (at least 300°C) reduces the luminescence to a low and definable residual level. Often this



Luminescence Dating in Paleoseismology,

Fig. 1 Processes associated with OSL dating. (a) Luminescence is acquired in mineral grains with exposure to ionizing radiation and trapping of electrons. (b) The luminescence for grains is zeroed by exposure to sunlight with erosion and transport. (c) With burial and exposure to ionizing radiation, free electrons are stored in charge defects within grains crystal lattice. (d) Further light

exposure of grains with erosion and transport zeros the luminescence. (e) The grains are buried again and luminescence is acquired with exposure to ionizing radiation. (f) Careful sampling without light exposure and measuring of the natural luminescence, followed by a normalizing test dose (L_n/T_n) compared to the regenerative dose to yield an equivalent dose (D_e) (From Mellett 2013)

luminescence “cycle” occurs repeatedly in many depositional environments with signal acquisition of mineral grains by exposure to ionizing radiation during the burial period and signal resetting (“zeroing”) with light exposure concurrent to sediment erosion and transportation. Often mineral grains that are fresh from bedrock sources have significantly lower luminescence emissions per radiation dose in comparison to grains that have cycled repeatedly. OSL dating provides an estimate of the time elapsed since latest period of burial and, thus, yields a depositional age (Fig. 1).

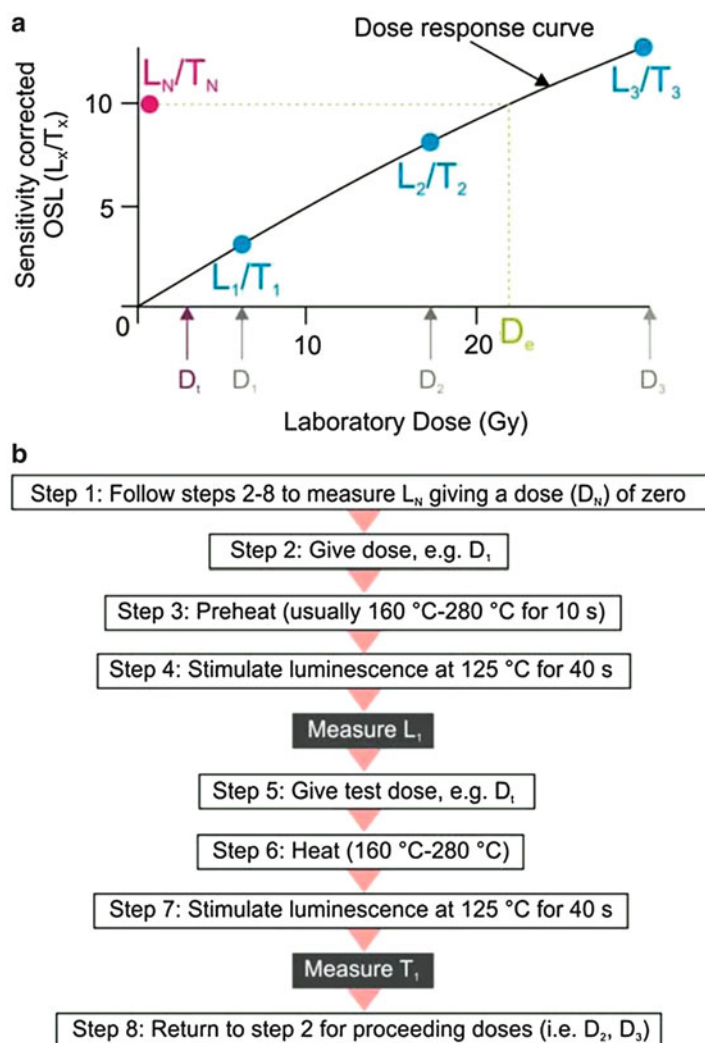
Principles of Luminescence Dating

The exposure of quartz and feldspar grains to sunlight for >60 s effectively diminishes the time-stored OSL signal to a low definable level. This residual level is the point from which the geological OSL signal accumulates post burial

(Fig. 2). Many types of sediment receive prolonged (>1 h) light exposure with transport and deposition, particularly in eolian, littoral, and sublittoral sedimentary environments. In addition, the inherent residual level is influenced by the susceptibility of the luminescence signal of a specific mineral to solar resetting. The OSL signal of potassium feldspar is usually more resistant to solar resetting than most quartz. There is significant variability in the luminescence properties of quartz and potassium feldspar grains related to crystalline structure, minor and rare-earth impurities, solid-solution relations, number of luminescence cycles (Fig. 1), and radiation history (Mejdahl 1986). Thus, because of this inherent variability in dose sensitivity of quartz and feldspar, analytical procedures for dating often need to be tailored for a specific geologic provenance. The advent of single aliquot regenerative (SAR) dose procedures for quartz (Murray and Wintle 2003; Wintle and Murray 2006) has provided the needed analytical flexibility to compensate for

Luminescence Dating in Paleoseismology,

Fig. 2 (a) Determination of equivalent dose (in grays) using the single aliquot regenerative (SAR) protocols, where the natural luminescence emission is L_n/T_n and the regenerative dose is L_x/T_x ; sensitivity changes are corrected by the administration of a small test dose (e.g., 5 grays). (b) Generalized SAR protocol (From Mellett 2013)



variable luminescence properties of quartz and feldspar grains and laboratory-induced sensitivity changes, particularly associated with preheat treatments and with laboratory beta irradiation (Fig. 2).

OSL dating is predicated on measurements of a specific mineral and particle size, usually quartz or potassium feldspar. Mineral separations are performed by standard techniques using heavy liquids and hydrofluoric acid (HF) to digest non-quartz minerals and etch the outer 10–20 μm of quartz grains, which are affected by alpha radiation. The purity of the quartz extract is primal for effective dating because a small amount contamination (1 %) by potassium

feldspar and other minerals can dominate the luminescence emissions. Multiple soaks in HF may be needed to obtain a pure quartz separate. Purity of the separate is accessed through microscopic inspection and point counting of grain mineralogy. Spectral purity of quartz is often determined by excitation by infrared light from a diode array with subsequent light emissions associated presumably with feldspar contaminants. However, some quartz grains yield considerable emissions with infrared excitation and may host feldspathic or other mineral inclusions; such grains should be analyzed as feldspar grains.

Sediment grains act as long-term radiation dosimeters when shielded from further light

exposure with the luminescence signal a measure of radiation exposure during the burial period. The radiation dose that

$$\text{OSL age (year)} = \frac{\text{Equivalent dose (D}_e\text{; grays)}}{\alpha D_\alpha w + D_\beta w + D_\gamma w + D_c} \quad (1)$$

α = *alpha efficiency* (0.03–0.15) for grains
>50 μm = 0

D_α = alpha dose

D_β = beta dose

D_γ = gamma dose

D_c = cosmic dose

w = attenuation by water

is equivalent to the natural luminescence emission of isolated quartz and feldspar grains is referred to as the equivalent dose (D_e , measured in grays: 100 rads = 1 gray) and is one half of the OSL age equation (Eq. 1). In most dating applications, quartz is often the favored mineral because of its abundance in sediments, ease of physical separation, and known stability of luminescence emissions. In contrast, feldspar minerals are often less abundant and have a troubling signal instability (anomalous fading), though yield considerably brighter OSL emissions. The recent development of charge transfer techniques for potassium feldspar (e.g., post IR290) that use elevated preheats (~290 °C) to transfer electrons from stable deeper to shallower traps for ease of measurement has extended dating possibilities to 10^5 – 10^6 timescales for well solar-reset grains (Duller and Wintle 2012). Similar protocols have been also developed for quartz that has been particularly useful for dating Pleistocene loess deposits (e.g., Brown and Forman 2012).

A common approach in OSL dating is to use SAR protocols on quartz aliquots with the protocols customized for a specific sample, study site, or area (Fig. 2). The SAR approach is predicated on a number of assumptions. First, the fast component of luminescence emissions, light released within the first 4 s, is the dominant

signal, usually >90 % of the total light output above background levels. Second, the zero dose ratio is <5 % of the natural ratio. Furthermore, the natural ratio is not at dose saturation that errors in calculating aliquot or grain equivalent doses are ≤ 10 % and the recycling ratio is between 1.1 and 0.9, a measure of the coincident of the first and the last same regenerative dose and indicative that the test dose compensates well for sensitivity changes (Murray and Wintle 2003). A critical practice in luminescence dating is the application of preheats (160–260 °C) to isolate a stable and a time-indicative luminescence emission (Fig. 2). The accuracy of the SAR protocols is evaluated by applying a known dose and testing for specific heat treatments within a sequence of analysis (Fig. 2) if the known dose can be recovered. Often such tests indicate that the known dose can be recovered using a variety of preheat treatments. In practice, using the SAR protocols to determine an equivalent dose involves calculation of a ratio between the natural luminescence and the luminescence from a known test dose (L_n/T_n ratio), which is compared to the luminescence emissions for regenerative doses (L_x/T_x) and also divided by the luminescence from the same test dose (Fig. 2). Applying a known test dose with each measurement cycle provides a metric to correct for changes in the sensitivity of the dated quartz grain(s) to acquire luminescence. Often with each successive measurement cycle, the luminescence output increases by 5–15 % or more from the same radiation dose.

The equivalent dose using the SAR protocols is determined often on >30 aliquots of quartz or feldspar grains. Each aliquot often contains 10–100s of quartz grains, the total number dependent on grain size (e.g., 100–150 μm), plate area, and luminescence yield. Statistical analyses of equivalent dose distributions are critical to render accurate OSL ages with specific age models (Galbraith and Roberts 2012). A common metric used is an overdispersion percentage of a D_e distribution and is an estimate of the relative standard deviation from a central D_e value in context of a statistical estimate of errors. A zero

overdispersion percentage indicates high internal consistency in D_e values with 95 % of the D_e values within 2σ errors, though rarely, if ever is a naught value calculated with equivalent dose data. Overdispersion values $\sim \leq 20$ % (at two sigma errors) are routinely assessed for quartz grains that are well solar reset, like eolian sands (e.g., Olley et al. 2004; Wright et al. 2011), and this value is considered a threshold metric for calculation of a D_e value using the central age model of Galbraith and Roberts (2012). Overdispersion values >20 % may indicate mixing of grains of various ages or the partial solar resetting of grains, though overdispersion values up to 32 % have been associated with a single equivalent dose population, particularly if the equivalent dose distribution is symmetrical (Arnold and Roberts 2009). The minimum or maximum age model may be an appropriate statistical treatment for equivalent dose data that is positively or negatively skewed (Galbraith and Roberts 2012), depending on the pedologic and the sedimentologic context. The net effect of pedogenesis and bioturbation is often the mixing in of younger grains, and thus, a maximum age model may be appropriate (Ahr et al. 2013). In contrast, high-energy fluvial and colluvial depositional environments often incorporate partially solar-reset grains; thus, the minimum age model may be appropriate. Other numeric treatments such as finite mixture modeling may also be applicable for separating multiple populations of equivalent dose (Galbraith and Green 1990). The efficacy of these numeric analyses for isolating grain populations should be questioned, and independent tests of accuracy are advised.

Another advantageous approach in OSL dating is single-grain dating using the SAR protocols (e.g., Duller 2008, 2012). Dating single grains of quartz and feldspar is particularly suitable with a mixture of grain populations in sediments, a common occurrence in some fluvial and colluvial sedimentary environments. A significant challenge with single-grain dating is that often 1000s of quartz grains need to be analyzed because 20–90 % of the grains yield little or no luminescence emissions. Thus, a single age

determination can occupy a luminescence reader for weeks to months to render sufficient number of analyses to statistically separate equivalent dose populations. Though this approach is time consuming, highly credible data is generated for identifying different equivalent dose grain populations and ultimately accurate ages.

A determination of the environmental dose rate (grays/ka) is needed to render an OSL age (Eq. 1), which is an estimate of the exposure of mineral grains to ionizing radiation from the decay of U and Th series, ^{40}K , and cosmic sources during the burial period. The U, Th, and ^{40}K content of the sediments can be determined by elemental analyses by neutron activation analyses or inductively coupled plasma-mass spectrometry, though such analyses assume secular equilibrium in the U and Th decay series. A suitable alternative is either gamma spectrometry in the field or in laboratory; the laboratory variant (germanium gamma spectrometry) can detect isotopic disequilibrium in the U and Th decay series. If disequilibrium is detected, the dose rate should be suitably modified. The beta and gamma doses should be adjusted according to grain diameter to compensate for mass attenuation (Fain et al. 1999). There is no appreciable alpha dose for coarse grains ($>50 \mu\text{m}$) with the outer 10–20 μm of grains etched by soaking in HF during preparation. Fine grain sediments ($<40 \mu\text{m}$) are fully affected by alpha radiation. A cosmic ray component must be included in dose rate calculations which compensates for geographic position, elevation, depth of burial, and density of burial material (Prescott and Hutton 1994). An estimate of the moisture content (by weight) during the burial period is also needed for age calculation. Often estimating moisture is difficult because there are few sedimentologic or diagenetic criteria to quantify changes in moisture content during the burial period. Thus, realistic moisture contents are usually derived from particle size considerations, measured moisture contents in an undisturbed setting, evaluation of sediment compaction, level of water table to sampling site, and climatic conditions; and appropriate uncertainty (10–20 %) should be included.

Sedimentologic Context for OSL Dating

There are a number of sedimentary facies in tectonic settings that can be or have the potential to be dated by luminescence. A luminescence age for clastic sediment is a measure of the time since the last sunlight exposure. The zeroing of luminescence in mineral grains, usually by sunlight, must be related to a significant tectonic event, like burial of soil with sand-blow emplacement or colluviation post faulting, for the OSL age to be meaningful. In turn, dating sediments that were displaced or deformed provides maximum limiting ages on tectonic activity (e.g., Chen et al. 2013). Eolian sediments, like loess, sand sheet deposits, dune sands, and cover loams, are the most preferred sediments for OSL dating because of the long (hours) light exposure prior to deposition. Also, littoral and sublittoral sediments often receive long light exposure within the swash zone (Argyilan et al. 2005), though storm beach deposits may be variably solar reset.

Mineral grains in water-lain environments such as glacial-marine, certain lacustrine, and fluvial can be incompletely solar reset reflecting attenuation of spectra as light penetrates a turbid water column. Low-energy fluvial sediment, like shallow channel fills with millimeter-scale, horizontally bedded, medium sands, is the preferred facies for OSL dating (Schaetzl and Forman 2008). Higher-energy facies with erosion of previous deposited sediments and formation of new grains with clast percussion during saltation and creep at the base of the water column often yield a mixture of apparent grain ages. Colluvial and alluvial sediment, often associated with normal and reverse faulting, may be amenable for OSL dating, though care is needed to avoid sampling intraclasts in proximal colluvium, closest to the fault zone. Proximal colluvium is difficult to date by OSL and yields often highly dispersed equivalent dose values by aliquots and single grains, necessitating the use of finite mixture models (e.g., Cortes et al. 2012). Distal colluvium, deposited by grain transport with overland flow downslope of the fault offset, can be well solar reset and is more suitable for OSL dating (Forman

et al. 1989; Sohbati et al. 2012). There is some promise in OSL dating of minerals from fault gouge to constrain the timing of tectonic activity, though luminescence characteristics are highly variable (Spencer et al. 2012; Banerjee et al. 1999).

Weathered horizons exhibiting pedogenic accumulations of clay, silt, carbonate, or silica should be avoided scrupulously for OSL dating. Postdepositional pedogenesis can affect detrimentally the luminescence time signal by altering the radionuclide concentration, disrupting crystal structure, and mixing in fully or partially solar-reset grains and thus increasing overdispersion values (Ahr et al. 2013).

A detailed study of luminescence characteristics of quartz grains in soils indicates that the upper (<5 cm) part of A horizons in well-developed soils (e.g., alfisols) is well solar reset and is another target for OSL dating, particularly in the buried context (Ahr et al. 2013).

Sampling of Sediment for OSL Dating

Sampling sediment for luminescence dating is relatively straightforward, though care should be exercised that the appropriate sedimentary facies is sampled. It is critical that the sedimentologic, pedologic, stratigraphic, and tectonic context is well known prior to sampling, so a judicious decision is made on sample selection that weighs the likelihood of solar resetting with original deposition, minimizes pedogenic alterations, and maximizes that the sediment-depositional (OSL) age closely constrains tectonic activity. Ideally, at the sampling site, at least 20 cm of homogenous sediment should surround the collected sediment to maximize uniformity in the dose rate environment during the burial period. Sampling within 20 cm of boulders or major lithologic contacts should be avoided to obviate potential inhomogeneities in radioactivity. If such inhomogeneities cannot be avoided, field gamma spectrometry or collection of sediment from the disparate units is advised to document the in situ dose rate. Also, care should be

exercised in collection samples within 1–1.5 m of the surface, with likely pedogenic alterations. Approximately 30 g of sediment should be collected, though more or less sediment may be adequate depending on the concentration of the chosen particle size for dating. The geological luminescence signal of sediment is reduced rapidly with exposure to sunlight; thus, care must be taken not to expose the sediment to any light during sampling. The OSL signal of quartz is significantly reduced with a few seconds of light exposure. Prior to sampling, the section should be excavated back at least 20 cm to expose a fresh face. Immediately prior to sampling, the face should be scraped free of light-exposed grains on the surface. It is best to take the sediment intact; though the outer mineral grains may have been light exposed, the internal grains have been shielded from sunlight. The most efficacious approach for sampling for OSL dating is to use light-tight sediment coring tubes, composed of aluminum, copper, steel, or black, stiff plastic, like ABS. The tube should be 10–20 cm long, 2–4 cm diameter, and with an end cap, the tube can be hammered or pushed into the desired sediment for dating. For clay-rich or indurated sediments, blocks can be carved from the section and placed in Kubiena tins or other metallic containers. A secondary bag sample of sediment (~50 g) that can be light exposed should also be taken from a 20 cm radius from the tube/block collection point for elemental (dose rate) and particle size analyses.

Summary

Optically stimulating luminescence is a technique that dates the burial time of quartz and feldspar grains. The zeroing mechanism is the exposure of grains to >1 min of sunlight, and there is subsequent signal acquisition with burial, shielding from sunlight, and exposure to ionizing radiation, mostly from radioactive decay during the burial period. Specific minerals, like quartz and potassium feldspar, are dated, with a known particle size range (e.g., 100–150 μm) either as

aliquots (10–100s of grains) or single grains. In many sedimentary environments, like eolian, sublittoral, and littoral, mineral grains are well solar reset and are most amenable for OSL dating. However, in certain fluvial, colluvial, and alluvial environments, solar resetting of mineral grains is partial, or older grains are incorporated, and thus, there are often multiple populations of grain ages. The collection of unsuitable samples for OSL dating can be obviated by a careful sedimentary facies analysis that maximizes for light exposure of grains with transportation and deposition. In turn, there are a variety of computational solutions for single aliquot data to isolate grain populations that may be temporally significant, though the efficaciousness of such solutions should be questioned. Single-grain analyses, though laborious requiring the analyses of 100–1000s of grains, provide the needed analytical data to define grain populations and accurate ages.

The effective age range of OSL dating of quartz is ca. 100,000 years, though this limit will vary by ca. ± 50 ka with environmental dose rate and electron trap, stability density, and energy. Thermal transfer techniques may extend the dating of quartz and feldspar grains from well solar-reset sediments to 10^5 – 10^6 time-scales. OSL has wide application in tectonic settings in dating fault offset or uplifted sediments or sediment deposition that closely precede or succeed a faulting event. Judicious application of OSL dating is hinged on documenting the complex sedimentologic association with fault movement and understanding equally if not more complex the luminescence emissions from grains of quartz and feldspar which yield a time signature.

Cross-References

- [Archeoseismology](#)
- [Paleoseismic Trenching](#)
- [Paleoseismology](#)
- [Radiocarbon Dating in Paleoseismology](#)
- [Tsunamis as Paleoseismic Indicators](#)

References

- Ahr SW, Nordt LC, Forman SL (2013) Soil genesis, optical dating, and geoarchaeological evaluation of two upland Alfisol pedons within the Tertiary Gulf Coastal Plain. *Geoderma* 192:211–226
- Aitken MJ (1985) Thermoluminescence dating. Academic, New York
- Aitken MJ (1998) An introduction to optical dating: the dating of quaternary sediments by the use of photon-stimulated luminescence. Oxford University Press, New York
- Argyilan EP, Forman SL, Johnston JW, Wilcox DA (2005) Optically stimulated luminescence dating of late holocene raised strandplain sequences adjacent to Lakes Michigan and Superior, Upper Peninsula, Michigan, USA. *Quatern Res* 63(2):122–135
- Arnold LJ, Roberts RG (2009) Stochastic modelling of multi-grain equivalent dose (D-e) distributions: implications for OSL dating of sediment mixtures. *Quat Geochronol* 4(3):204–230
- Banerjee D, Singhvi AK, Pande K, Gogte VD, Chandra BP (1999) Towards a direct dating of fault gouges using luminescence dating techniques – methodological aspects. *Curr Sci* 77(2):256–268
- Brown ND, Forman SL (2012) Evaluating a SAR TT-OSL protocol for dating fine-grained quartz within Late Pleistocene loess deposits in the Missouri and Mississippi river valleys, United States. *Quat Geochronol* 12:87–97
- Chen Y, Li S-H, Sun J, Fu B (2013) OSL dating of offset streams across the Altyn Tagh Fault: channel deflection, loess deposition and implication for the slip rate. *Tectonophysics* 594:182–194
- Cortes AJ, Gonzalez LG, Binnie SA, Robinson R, Freeman SPHT, Vargas EG (2012) Paleoseismology of the Mejillones Fault, northern Chile: insights from cosmogenic Be-10 and optically stimulated luminescence determinations. *Tectonics* 31, TC2017
- Duller G (2008) Single-grain optical dating of quaternary sediments: why aliquot size matters in luminescence dating. *Boreas* 37(4):589–612
- Duller G (2012) Improving the accuracy and precision of equivalent doses determined using the optically stimulated luminescence signal from single grains of quartz. *Radiat Meas* 47:770–777
- Duller G, Wintle AG (2012) A review of the thermally transferred optically stimulated luminescence signal from quartz for dating sediments. *Quat Geochronol* 7:6–20
- Fain J, Soumana S, Montret M, Miallier D, Pilleyre T, Sanzelle S (1999) Luminescence and ESR dating-Beta-dose attenuation for various grain shapes calculated by a Monte-Carlo method. *Quat Sci Rev* 18:231–234
- Forman SL, Machette MN, Jackson ME, Matt P (1989) Evaluation of thermoluminescence dating of paleoearthquakes on the American Fork segment, Wasatch fault zone, Utah. *J Geophys Res* 94:1622–1630
- Galbraith RF, Green PF (1990) Estimating the component ages in a finite mixture. *Nucl Tracks Radiat Meas* 17(3):197–206
- Galbraith RF, Roberts RG (2012) Statistical aspects of equivalent dose and error calculation and display in OSL dating: an overview and some recommendations. *Quat Geochronol* 11:1–27
- Huntley DW, Godfrey-Smith DI, Thewalt MLW (1985) Optical dating of sediments. *Nature* 313:105–107
- Hutt G, Jaek I, Tchonka J (1988) Optical dating: K-feldspars optical response stimulation spectra. *Quat Sci Rev* 7:381–385
- Mejdahl V (1986) Thermoluminescence dating of sediments. *Radiat Prot Dosimetry* 17:219–227
- Mellet CL (2013) Luminescence dating. In: Clarke LE (ed) *Geomorphical techniques* (online edition). British Society for Geomorphology, London, pp 1–11. http://www.geomorphology.org.uk/assets/publications/subsections/pdfs/OnsitePublicationSubsection/90/4.2.6_luminescencedating.pdf
- Murray AS, Wintle AG (2003) The single aliquot regenerative dose protocol: potential for improvements in reliability. *Radiat Meas* 37(4–5):377–381
- Olley JM, Pietsch T, Roberts RG (2004) Optical dating of Holocene sediments from a variety of geomorphic settings using single grains of quartz. *Geomorphology* 60(3–4):337–358
- Prescott JR, Hutton JT (1994) Cosmic ray contributions to dose rates for luminescence and ESR dating: large depths and long-term time variations. *Radiat Meas* 23:497–500
- Schaetzl RJ, Forman SL (2008) OSL ages on glaciofluvial sediment in northern Lower Michigan constrain expansion of the Laurentide ice sheet. *Quatern Res* 70(1):81–90
- Sohbati R, Murray AS, Buylaert J-P, Ortuno M, Cunha PP, Masana E (2012) Luminescence dating of Pleistocene alluvial sediments affected by the Alhama de Murcia fault (eastern Betics, Spain) – a comparison between OSL, IRSL and post-IR IRSL ages. *Boreas* 41(2):250–262
- Spencer JQG, Hadizadeh J, Gratier J-P, Doan M-L (2012) Dating deep? Luminescence studies of fault gouge from the San Andreas Fault zone 2.6 km beneath Earth's surface. *Quat Geochronol* 10:280–284
- Wintle AG, Huntley DJ (1980) Thermoluminescence dating of ocean sediments. *Can J Earth Sci* 17:348–360
- Wintle AG, Murray AS (2006) A review of quartz optically stimulated luminescence characteristics and their relevance in single-aliquot regeneration dating protocols. *Radiat Meas* 41(4):369–391
- Wright DK, Forman SL, Waters MR, Ravesloot JC (2011) Holocene eolian activation as a proxy for broad-scale landscape change on the Gila River Indian Community, Arizona. *Quatern Res* 76(1):10–21

Masonry Box Behavior

Rui Marques
Engineering Department, Civil Engineering
Section, Pontifical Catholic University of Peru,
Lima, Peru

Synonyms

Box action; Building global behavior; Diaphragm effect; Structural connectivity

Introduction

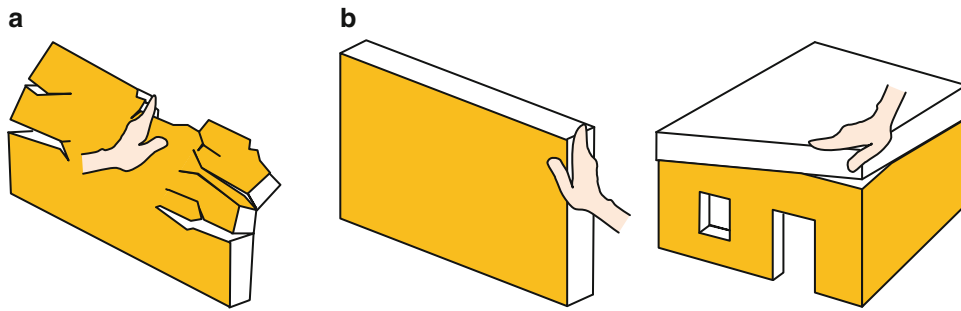
This entry deals with the box behavior of masonry structures, which is a major hypothesis for the application of seismic assessment procedures based on the in-plane response of the walls, when using performance-based approaches for seismic safety. A general description is also made of the unreinforced and confined masonry building typologies, regarding the constructive technique, industrial development, and applicability. Methods and procedures are presented for the seismic assessment and safety verification of masonry buildings, which are applied to the case of a dwelling.

Masonry has been continually used for building, employing very different materials and bond patterns, but the main distinction between masonry constructions probably is the presence

or absence of rigid floor diaphragms well connected to the walls. According to this aspect, masonry constructions can be divided in two categories: structures with and without box behaviour, which present a very dissimilar response when subjected to seismic actions. For the case without box behavior, the walls of the building behave independently and out of phase, with combined in-plane and out-of-plane deformations, and presenting mainly out-of-plane damage (Fig. 1a). On the other hand, when with box behavior the building acts as a jointly assemblage of walls and roofs, with mainly in-plane response of the walls (Fig. 1b).

The concept of box behavior was a key assumption of methods developed for the seismic analysis of masonry structures, e.g., the POR (Tomažević 1978). In this case, a macroelement discretization has been adopted for the walls through a frame-type assemblage of pier, spandrel, and connection elements (Marques and Lourenço 2011). For the POR, it was initially considered that the walls deform jointly and without plane rotation, thus the global response of a masonry structure being computed as the sum of the in-plane response of individual walls. Within this hypothesis, even if the masonry is a brittle material, the variety of wall sections (deformation capacities) in a building allows a significant structural ductility.

These methods consider generally an elastic–perfectly plastic load–displacement response for the walls according to given strength



Masonry Box Behavior, Fig. 1 Illustration of box behavior concepts (Adapted from Touliatos 1996): (a) out-of-plane failure, (b) in-plane response due to box behavior

criterion and ultimate drift, depending on the formed collapse mechanism by shear or flexure. The response of the building is then evaluated by a pushover loading analysis, which allows identifying the sequence of damage events on panels, reflecting in a graphical representation of the base shear force against the displacement of a control point, e.g., the centroid of the roof slab, which is called the capacity curve. This curve is an intrinsic feature of the building that reflects, beyond the base shear strength and deformation capacity, the dissipated energy.

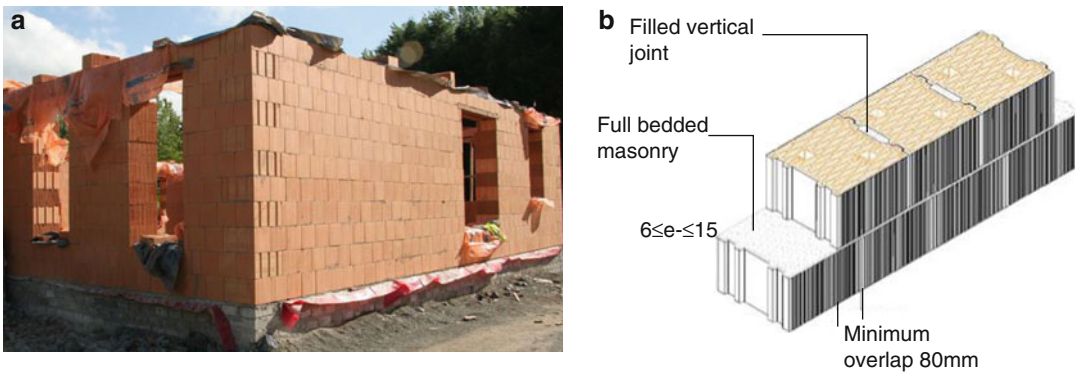
The concept of box behavior and the current performance-based approaches for seismic assessment specified in codes (e.g., EN 1998-1 2004; Sullivan et al. 2012) are explained here referring to cases of unreinforced and confined masonry building structures. In the next two sections, these typologies are presented in general terms.

Unreinforced Masonry

Unreinforced masonry is an ancient, traditional, and ongoing construction typology, whose system consists in the juxtaposition and superposition of masonry pieces through joints. In the antiquity, stone pieces and mud bricks were used for the construction of unreinforced masonry monumental buildings, but the great revolution of this construction technique was with the production of burnt clay bricks. In recent years, a large development was verified in the industry of ceramic bricks allowing for high

quality and efficiency, regarding mechanical properties and functional aspects of the brick itself and of the masonry components as an assemblage. At the same time, concrete and calcium silicate blocks were also developed, which are an alternative to ceramic bricks. Masonry requires also for a good quality of remaining constituents, particularly the joint mortar. Block masonry systems have been also developed with mortar pocket and/or tongue-and-groove head joints, e.g., in Fig. 2. Aiming to control cracking due to creep and shrinkage movements, bed-joint steel truss reinforcement is in many cases recommended.

Unreinforced masonry is widely used in Germany (mainly ceramic brick masonry, e.g., Jäger et al. 2010) and in Brazil (mainly concrete block masonry, e.g., Parsekian et al. 2012), which demonstrates the wide applicability of unreinforced masonry around the world, both in cases of strong-economy nations and developing countries. In these countries, great development was observed with the industrialization of, more than a masonry unit, a masonry system allowing for a modular design with use of complementary pieces, such as half bricks, corner blocks, lintel bricks, etc. Masonry systems need to consider also for functionality regarding the installation of water supplies and cable networks, namely, through driving of shafts into the walls. Another aspect to consider is the adequacy of used solutions for slabs, particularly concerning the condition of box behavior. Regarding this point, solutions need to be considered that allow an effective connection of the



Masonry Box Behavior, Fig. 2 Example of an unreinforced masonry system: (a) building view, (b) scheme of the bond

walls to the slab system, namely, through the use of a bond beam.

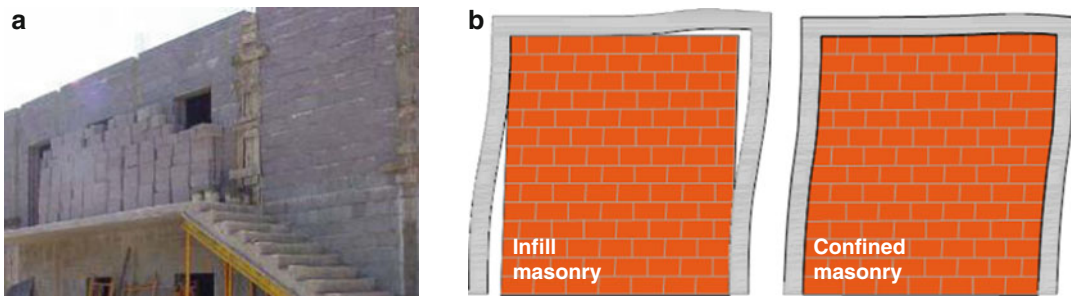
Even if the constructive process of masonry is very simple, skilled labor is a major aspect for unreinforced masonry construction, since the practices for infill walls have caused a great degradation of the masonry works, which requires for a re-specialization. Once more, the architectural trends and requested esthetics for buildings are a great challenge for masonry as a material and for engineers, when requiring for asymmetric, complex, and irregular configurations. For this reason, more than other construction typology, unreinforced masonry requires a much synchronized planning involving both architects and engineers to allow the feasibility of the construction. However, unreinforced masonry presents a restricted applicability in high seismicity regions due to limited earthquake resistance, in which cases an improvement to the masonry system is necessary, namely, by using the confined masonry typology presented in the next section. Reinforced masonry is also an alternative, which is largely disseminated in the United States and Canada (Parsekian et al. 2012), but which implementation seems not easily and directly adaptable to other countries.

Confined Masonry

Confined masonry was historically introduced as a reaction to destructive earthquakes in Italy and

Chile (1908 Messina and 1928 Talca quakes, respectively), which almost completely destroyed traditional unreinforced masonry buildings. The improvement of this technique was reflected on the good performance of confined masonry buildings subjected to the 1939 Chilean earthquake, which was probably the main reason for the large dissemination of this technique in all Latin America (Brzev 2007). The construction with confined masonry spread widely to all continents, in countries with moderate-to-high seismicity such as Slovenia, India, New Zealand, Japan, and Canada.

Conceptually, the confined masonry system is based on embracing masonry panels with frame elements, similarly to reinforced concrete frames, but with the difference that in the confined masonry the reinforced concrete elements are cast only after the masonry construction (Fig. 3a). For this reason, contrarily to reinforced concrete structures where infill masonry is built after concrete hardening, in the case of confined masonry most of the building weight rests on the masonry panels. In addition, due to concrete shrinkage, the connection between masonry and concrete is very effective (Fig. 3b). The interaction between the confining elements and the masonry panel allows a confined masonry wall under lateral loading to behave as a whole up to large deformation levels, allowing improved strength and ductility (e.g., Gouveia and Lourenço 2007). A typical phenomenon in confined masonry is the dowel



Masonry Box Behavior, Fig. 3 Illustrations of confined masonry typology: (a) construction example (Courtesy of Paulo B. Lourenço), (b) difference between infill and confined masonry regarding the frame-to-masonry connection

action of the longitudinal reinforcement bars of the concrete columns, which balances the loss of shear strength in the masonry panel, allowing to keep the strength necessary for a ductile response.

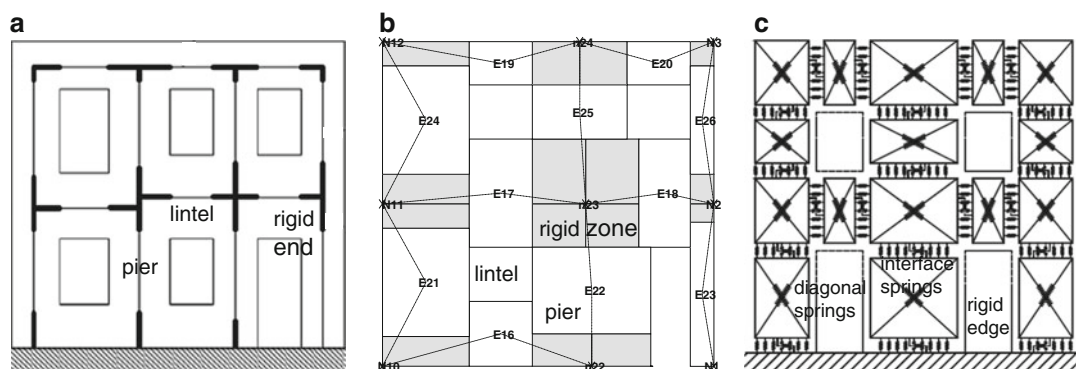
In practice, and relatively to the reinforced concrete system, confined masonry allows normally for economical savings due to the required smaller sections of concrete and steel. However, even if extensive experimental studies have been performed to evaluate the shear response of confined masonry structures (e.g., San Bartolomé 1994), the practice of self-construction and the use of prescriptive rules have led to poor earthquake resistance. Additionally, the use in many cases of poor-quality units leads to its break and to an inadequate behavior of the confined walls. The observed seismic damage in confined masonry buildings occurs in some cases at upper stories of the buildings, with associated out-of-plane damage, and is mostly due to the absence of box behavior in the affected stories. Brzev et al. (2010) attribute the seismic damage on confined masonry buildings mainly to inadequate wall density, poor quality of masonry and construction, deficiencies in detailing of reinforced concrete confining elements, absence of confinements at openings, and geotechnical issues.

In Europe, industrialized systems for confined masonry have been developed, namely, in Spain and Portugal, which allow for a modular design (e.g., Lourenço et al. 2008). An important aspect in the case of confined masonry is the formation of thermal bridges due to masonry discontinuity,

whose problem can normally be solved by coating the columns with thermal insulation plates or by casting the piers within special blocks. This construction typology is also a potential solution for the problem of housing in developing countries due to its quick construction and economy. Confined masonry was, for example, used in the reconstruction of Haiti after the 2010 earthquake.

Macroelement Models

Masonry structures present specific and diverse bond typologies, for which several modeling approaches have been adopted. In the academic-research field, the modeling of masonry buildings has been applied using two different scales, namely, the finite element and macroelement approaches (Lourenço 2002). The concept of using structural component models for masonry structures, designated by macroelement modeling, was introduced by Tomažević (1978) and applied to perform seismic assessment. This concept is the one addressed next, given the easy implementation of material laws and the formulation of structural equilibrium. In addition, the adopted structural component discretization largely reduces the number of degrees of freedom in relation to the traditional finite element modeling approaches, allowing for more resource- and time-efficient computations and making them attractive to practitioners. In the following, the available models are briefly described for unreinforced and confined masonry.



Masonry Box Behavior, Fig. 4 Schematization of wall models in (a) SAM II, (b) TreMuri, and (c) 3DMacro

Recently, and mainly in Italy, several user-friendly computer codes based on macroelements have been developed for assessing the seismic safety of unreinforced masonry buildings. Marques and Lourenço (2011) benchmarked the ANDILWall/SAM II (Calliari et al. 2010), the TreMuri (Lagomarsino et al. 2009), and the 3DMacro (Gruppo Sismica 2013) software codes and provided the basic description of the macroelement formulation and assemblage used in these methods. Briefly, SAM II and TreMuri are based on frame-type modeling by using one-dimensional macroelements, while the 3DMacro is based on a discretization with two-dimensional discrete elements, as shown in Fig. 4. The seismic assessment is made through performance-based approaches, i.e., procedures for nonlinear static (pushover) analysis, according to recent design codes (EN 1998-1 2004; Sullivan et al. 2012).

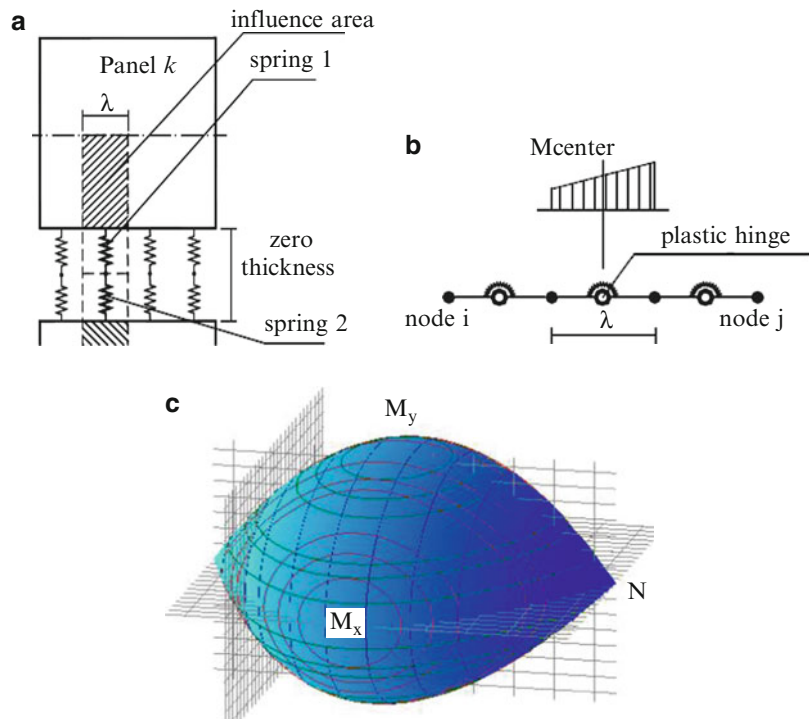
The methods above were developed throughout a sequential process, including first the idealization of the macroelement, then the definition of the wall assemblage, and finally the simulation of the full building model. The validation process of these methods was made by referring to experimental testing and more advanced computations, for individual masonry panels, full plane masonry walls, and three-dimensional structures, e.g., for the building tested by Magenes et al. (1995). This validation process is not easy due to the large variability of masonry materials and due to the importance that the complex structural organization in full buildings might assume.

On the other hand, rather few quasi-static tests on unreinforced masonry buildings have been carried out allowing this validation. Many unreinforced masonry building models have been tested in shaking tables, but, in general, the dynamic experimental behavior is difficult to compare with the pushover response.

Confined masonry is a particular case of masonry structures, even if it presents some similarity with reinforced concrete structures because of the presence of a frame. Confined masonry is characterized by casting of the reinforced concrete elements only after the masonry work, which provides a good connection between the confining elements and the masonry panels due to the combination of bond effects, shrinkage of the reinforced concrete elements, and the fact that the vertical dead load is transferred to the walls. The interaction behavior between the confining elements and masonry through the existing interface is a specific aspect that needs to be considered in the response of confined masonry walls under lateral loading. Some models have been implemented for confined masonry structures based on a wide-column approach (e.g., (Marques and Lourenço 2013), considering the interaction behavior between the confining elements and the masonry implicitly in the wall shear response.

Micro-modeling strategies can also be used for confined masonry, namely, based on the finite element method, to model explicitly the concrete–masonry interface (e.g., Calderini et al. 2008). Alternatively, a discrete element

Masonry Box Behavior,
Fig. 5 Discrete
 macroelement: (a) interface
 model, (b) frame element,
 and (c) example of
 interaction domain
 N – M_x – M_y



approach is also applicable, such as that idealized by Calìo et al. (2012) originally for unreinforced masonry and which has been extended in the 3DMacro software (Gruppo Sismica 2013) to model reinforced concrete/steel/masonry mixed structures. This last approach uses an interface connection (constituted by nonlinear springs) between the masonry panels (Fig. 5a), which in the case of two neighboring confined masonry panels is interposed by a frame modeled through nonlinear beam finite elements with concentrated plasticity (Fig. 5b). For the beam elements and in agreement to a given type of interaction (axial, flexural, or axial-flexural), the corresponding hinges are considered according to the respective N – M_x – M_y domain (such as in Fig. 5c).

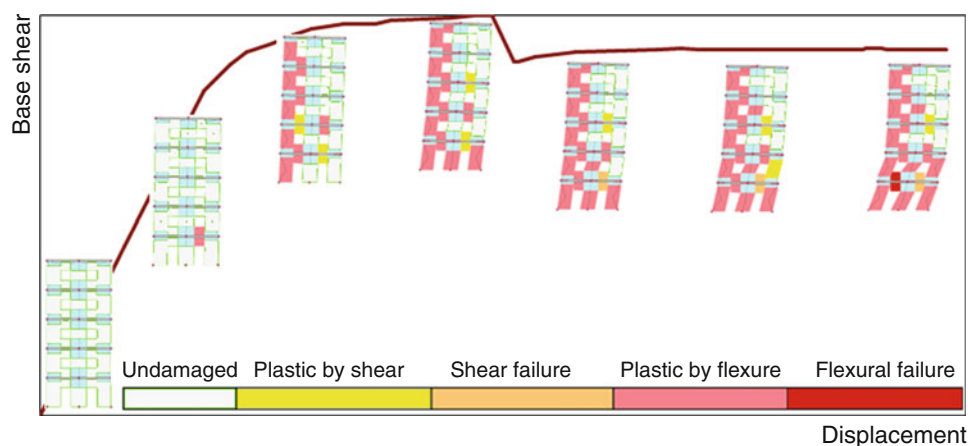
Nonlinear Static (Pushover) Analysis

The first approaches for seismic analysis were based on the concept of seismic coefficient, which represents the proportion of the building weight that is applied as a lateral force to the structure. Later, approaches were introduced

considering the response in displacements of the masonry walls, accounting for the greater sensitivity of the damage to the imposed displacements (e.g., Turnšek and Čačovič 1970).

The masonry structures can present significant ductility when subjected to lateral actions, which is the greater the higher the ductility of sections and the larger the variety of wall geometries (deformation capacities). This concept is early evidenced from damage patterns observed in masonry buildings subjected to earthquakes (e.g., 1963 Skopje, 1965 Oaxaca and 1976 Friuli earthquakes) and also based on experimental evidence (Tomažević 1999). The masonry walls show a great capacity of inelastic cyclic deformation and energy dissipation due to progressive activation of damage mechanisms. Furthermore, the walls are capable to support the vertical loading even if significantly damaged (e.g., with diagonal cracks).

The study of the nonlinear behavior of structures was started in the 1960s for reinforced concrete, particularly based on the observation of earthquake-damaged buildings. Later, numerical models were developed for the nonlinear static



Masonry Box Behavior, Fig. 6 Illustration of the pushover analysis on a masonry building

analysis of reinforced concrete structures (Saiidi and Sozen 1981). A similar investigation sequence and line time was observed for masonry structures (Tomažević 1978). However, if for the case of reinforced concrete structures advanced models were proposed in the background of extensive resources and studies, the proposed methods for masonry were simpler. This was due to the secondary role of masonry structures in the construction sector and particularly to the very complex behavior of masonry.

The nonlinear static analysis, or pushover analysis as popularized, is focused on the evaluation of the seismic performance of structures based on the control of damage mechanisms and deformation. Several procedures for pushover analysis have been developed, which allow considering the inelastic reserve of capacity in the design and safety verification of structures as a simplified alternative to the nonlinear dynamic time history analysis, e.g., the N2 method (Fajfar and Fischinger 1988) specified in seismic codes (EN 1998-1 2004; Sullivan et al. 2012). The procedures for pushover analysis are based on the control of deformation, where the displacement demands due to the design earthquakes are obtained through a spectral analysis of the response of an equivalent single-degree-of-freedom system, which are after compared with the displacement capacities corresponding to given performance levels.

The displacement capacities are computed from the capacity curve obtained for the building, which presents the evolution of the base shear force (horizontal force representative of the seismic action) against the displacement of a control point (significant point of the structure, usually the centroid of the roof slab), e.g., in Fig. 6. This curve is computed by applying an incremental lateral static loading on the structure, which is processed with force and/or displacement control, and for which two load distribution patterns are usually assumed: proportional to the inertial masses multiplied by the displacements of the first vibrational mode of the structure (modal distribution) and proportional to the inertial masses (uniform distribution).

Seismic Safety Assessment

Traditionally, the seismic design of buildings is based on computing a solicitation for the structure in terms of a base shear force through the consideration of a seismic coefficient or spectrum, which is reduced by a behavior factor and distributed after to the structure. However, the behavior factor values recommended in Eurocode 8 (EN 1998-1 2004) for masonry structures (1.5–2.0) turned out to be very conservative, and often the buildings are considered unsafe. In this case, the consideration of an overstrength

ratio that multiplies the basic value of the behavior factor is mandatory, which reflects the possibility of force redistribution between elements beyond the elastic limit (Magenes 2006). The values for the overstrength ratio are specified in the Italian code (IBC 2008) based on experimental and analytical evidences.

Currently, a transition stage from force-based to performance-based design is verified through the use of approaches that consider the ductility or displacement capacity of the structure. The displacement-based approach, which in effect provides a safety verification rather than an explicit design, is presently the more developed level of seismic safety assessment. In this case, performance is related to acceptable damage and damage to displacement. The N2 method (Fajfar and Fischinger 1988) specified in Eurocode 8 (EN 1998-1 2004) is the method considered here.

For the safety evaluation, the N2 method defines a bilinear representation of the capacity curve corresponding to an equivalent single-degree-of-freedom system. According to the Italian code (IBC 2008), this representation includes a first straight line that passes by the origin and intercepts the capacity curve of the actual system for a base shear force of 70 % of the maximum and a second line, horizontal and defined in mode that the area below the envelope of the actual system equalizes the area under the bilinear idealized response. The bilinear representation allows to obtain the reference period of vibration for the computation of the target displacement.

The procedure for safety verification is based on the evaluation of the seismic performance of the building, in terms of deformation, verifying that the target displacement is lower than the displacement capacity of the building, obtained on the capacity curve in correspondence with the established limit states. Commonly, damage and ultimate limit states are considered, which are, respectively, associated with the peak base shear and with a post-peak remaining base shear force of 80 %. A factor is also considered as the ratio between the elastic response force and the yield force of the equivalent system, which value

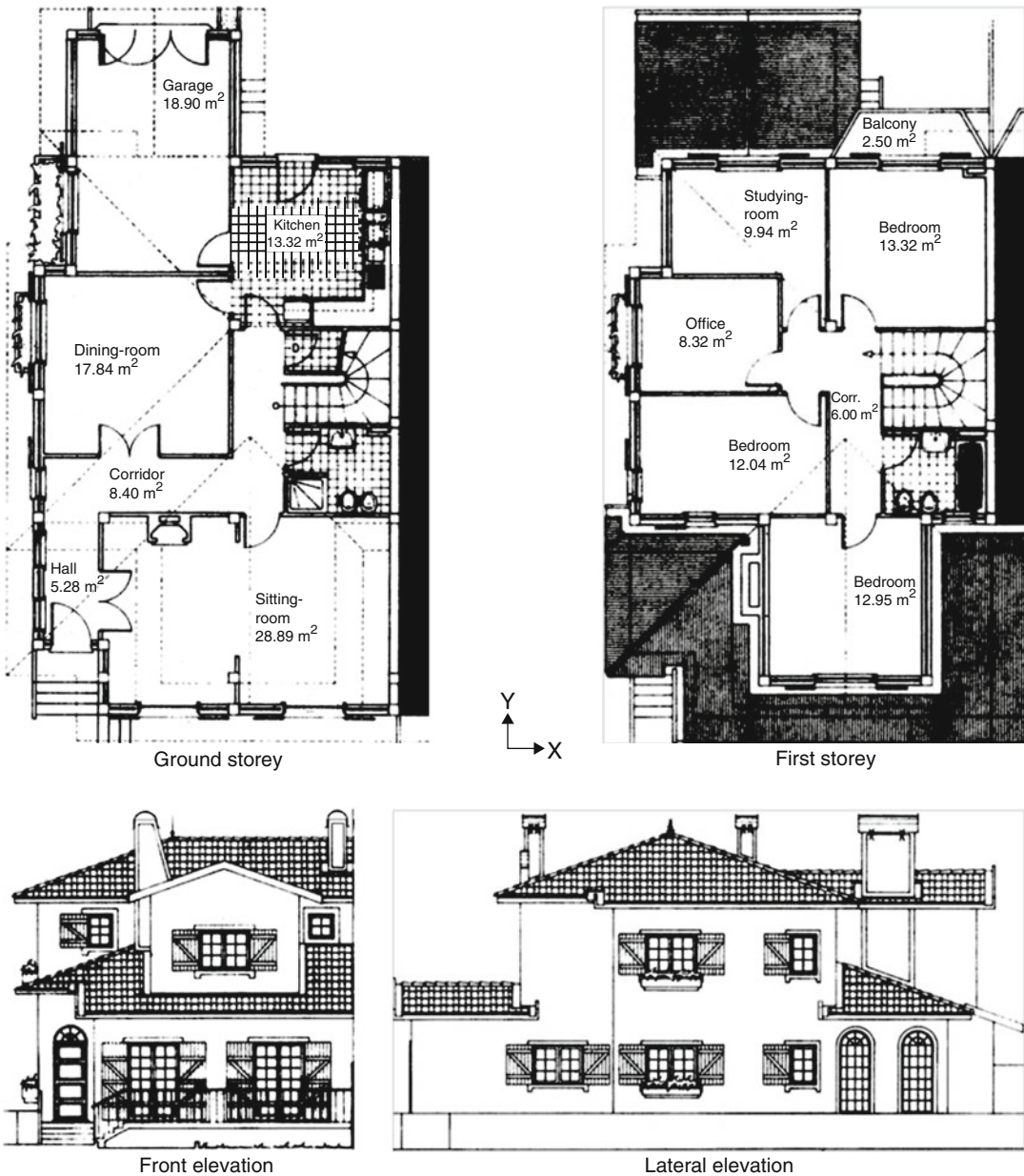
(imposed to be less than 3.0) represents a limitation to the ductility of the structural system as a whole.

The out-of-plane assessment is an issue not clearly evidenced in seismic codes, which implies a personal interpretation of safety from the engineer. However, in general, it can be considered that the control of slenderness limitations, minimum thickness requirements, and appropriate structural conception and detailing (rigid diaphragms and efficient floor-to-wall connection) avoid out-of-plane driven failures. A possibility to check the out-of-plane safety is the application of a kinematic limit analysis, which in effect has been implemented in several commercial software codes, additionally to the global displacement-based safety verification. In this case, the user needs to define the potential kinematic blocks (portions of connected walls), the kind of mechanisms, and respective hinges and constraints.

Application to Dwelling House

Dwelling houses up to two stories are the majority of the building stock in Europe, in terms of both existing and newly constructed buildings. The loss of masonry as a structural solution, due to an ungrounded perception of its lack of capacity to resist earthquakes and also to the dissemination of reinforced concrete structures, caused a strong reduction of the use of structural masonry in new buildings. On the other hand, a large development occurred in the masonry industry, namely, with the introduction of high-quality masonry systems regarding functional and mechanical features. In the academic field important efforts have been also made to develop adequate tools to account for the intrinsic nonlinear capacity of masonry structures when subjected to earthquakes. A first comparison of these tools was made by Marques and Lourenço (2011) referring to a simple building configuration. In this work the comparison is extended, concerning a real and more complex structure.

The studied two-story semidetached dwelling is representative of typical housing in southern

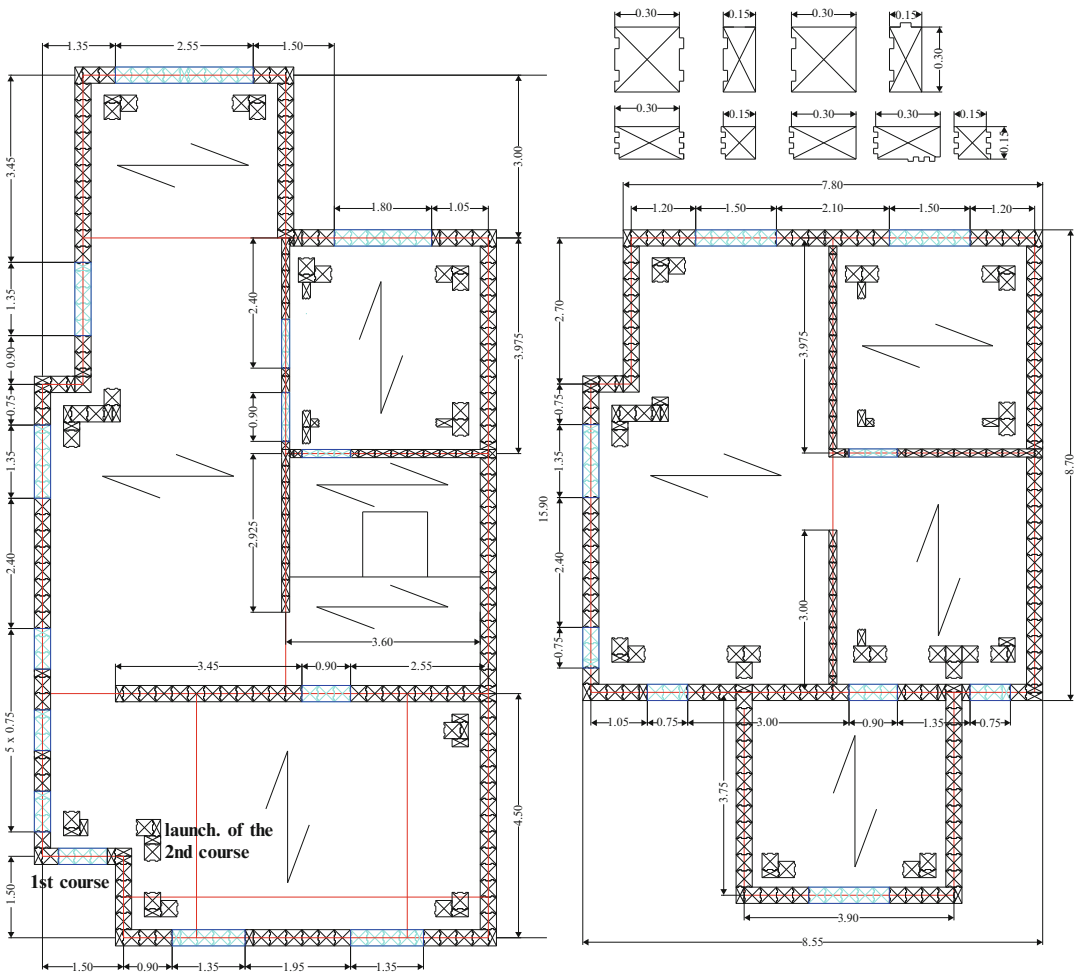


Masonry Box Behavior, Fig. 7 Architectural drawings of the dwelling: plans and elevation views

Europe, with rooms and kitchen in the ground story and bedrooms and an office in the first story. The house presents also a garage and a multilevel roof, as shown in Fig. 7. The structure of the building was originally designed in reinforced concrete. In the following, alternative to the original structure, solutions of unreinforced masonry and confined masonry structures are

presented and compared. The building is assumed to be constructed on type B ground (deposits of very dense sand, gravel, or very stiff clay).

Unreinforced masonry, considering that is a simple construction technique which allows an energy-efficient enclosure with no thermal bridges concern, is the first adopted option. In this case,

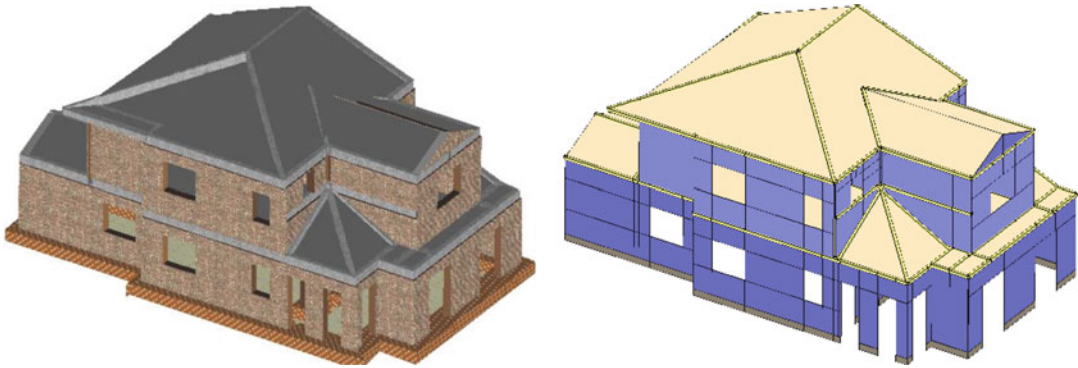


Masonry Box Behavior, Fig. 8 Structural plans of the constructive solution in unreinforced masonry

a clay block masonry system is used with tongue-and-groove head joints and with shell-bedded horizontal joints. The system uses a vertically perforated block of $0.30 \text{ m} \times 0.30 \text{ m} \times 0.19 \text{ m}$ and complementary pieces (half block, end piece, corner block, adjusting piece, and lintel block) allowing a geometry in plan with a module of 0.15 m . A M10 mortar pre-batched, according to Eurocode 6 (EN 1996-1-1 2005), is used for the joints. The plans for the proposed structure are presented in Fig. 8, which show the bond in the first course and the second course arrangement around the wall crossings.

The floors of the building are made with prestressed ribbed slabs, which have a thickness of 0.19 and 0.15 m in the accessible zones and for

the ceiling/roof, respectively. The slabs are similar to those used in the original reinforced concrete structure and are subjected to dead loads of 4.5 and 3.0 kN/m^2 and to live loads of 2.0 and 0.4 kN/m^2 for the same areas. A bond beam is made with the lintel block that serves as a formwork, with a concrete section of $0.20 \times 0.35 \text{ m}^2$ and reinforcement of $4\phi 10 \text{ mm}$ with stirrups $\phi 6 \text{ mm} @ 0.20 \text{ m}$ (0.10 m at the ends). The properties for the masonry were considered according to Eurocode 6 and the producer-declared values and selected materials, resulting in a compressive strength of 3.25 MPa , a pure shear strength of 0.1 MPa , an elastic modulus of $3,250 \text{ MPa}$, and a shear modulus of $1,300 \text{ MPa}$. Additionally, the in-plane ultimate



Masonry Box Behavior, Fig. 9 Geometrical and computational models of an unreinforced masonry building

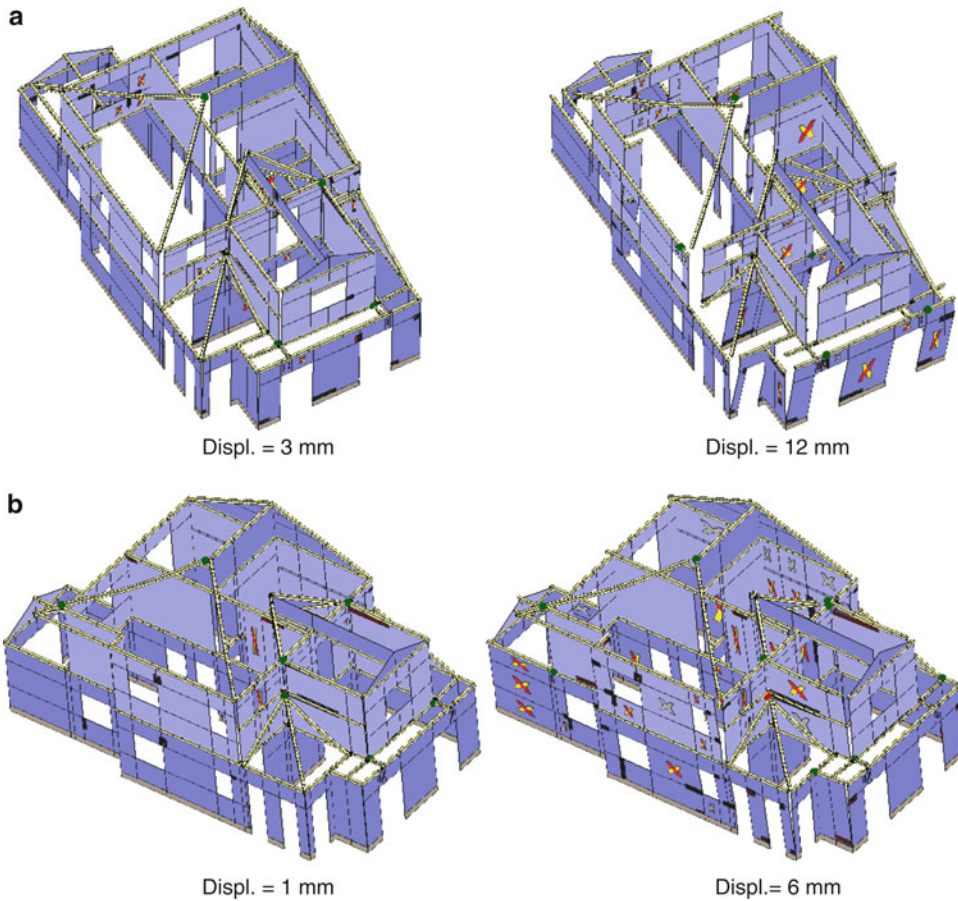
drifts for the masonry panels were assumed, according to Eurocode 8 (EN 1998-1 2004) and the Italian code (IBC 2008) with values of 0.4 % and 0.8 %, respectively, for the shear and flexural failure modes. The weight for the masonry with plastering is 12.0 kN/m^3 .

The building was modeled in the macroelement software code developed by Gruppo Sismica (2013), according to the geometrical and computational models presented in Fig. 9. The floors are simulated as polygonal diaphragms elastically deformable, considering an orthotropic slab element. The model was analyzed under pushover loading in the principal directions, and considering “mass” and “first-mode” load distributions. Concerning the +X and +Y analysis with first-mode load distribution, the damage sequence for the building is shown in Fig. 10. For +X and regarding the front façade, a trend is captured in the ground story with shear failure of the central pier and rocking of the extreme piers and with concentration of displacements at this story. A similar deformation mechanism and substantial shear damage are observed in Fig. 10a for the other wall alignments. In +Y, rocking and shear failures are predicted for slender and squat panels, respectively, and identifying a first story mechanism.

A confined masonry solution was also considered following the prescriptive rules from the European regulations. For the confining columns, according to Eurocode 6 (EN 1996-1-1 2005), a longitudinal steel reinforcement must be included equivalent to 0.8 % of the column’s

cross-section area, with a minimum of $4\phi 8 \text{ mm}$. The transversal reinforcement consists of 6 mm diameter stirrups spaced in general of 0.20 m and of 0.10 m at the column ends. The solution for the horizontal confining element is the same adopted for the unreinforced masonry structure. The structural plans for the confined masonry solution are presented in Fig. 11. The confined masonry structure was also modeled in the software by Gruppo Sismica (2013) with the geometrical and computational models shown in Fig. 12. Note that the main change with respect to the unreinforced masonry model is the inclusion of the reinforced concrete confining columns, which are simulated as beam finite elements with concentrated plasticity and feature a tridimensional constitutive behavior $N-M_x-M_y$. The model was then analyzed under pushover loading in the main directions and assuming load distributions proportional to the mass and to the first vibration mode.

The damage sequence for the confined masonry model is illustrated in Figs. 13 and 14 for the +X and +Y analysis with “first-mode” load distribution, corresponding to different displacement levels. For both analyses, damage in the building starts to be relevant at 1 mm displacement, when several masonry panels develop tensile cracks in the interfaces and also diagonal shear cracking. For a displacement of 4 mm, damage spreads widely to all walls of the building. Then, damage evolves with increasing deformation due to the progressive formation of flexural plastic hinges at the ends of the confining elements.

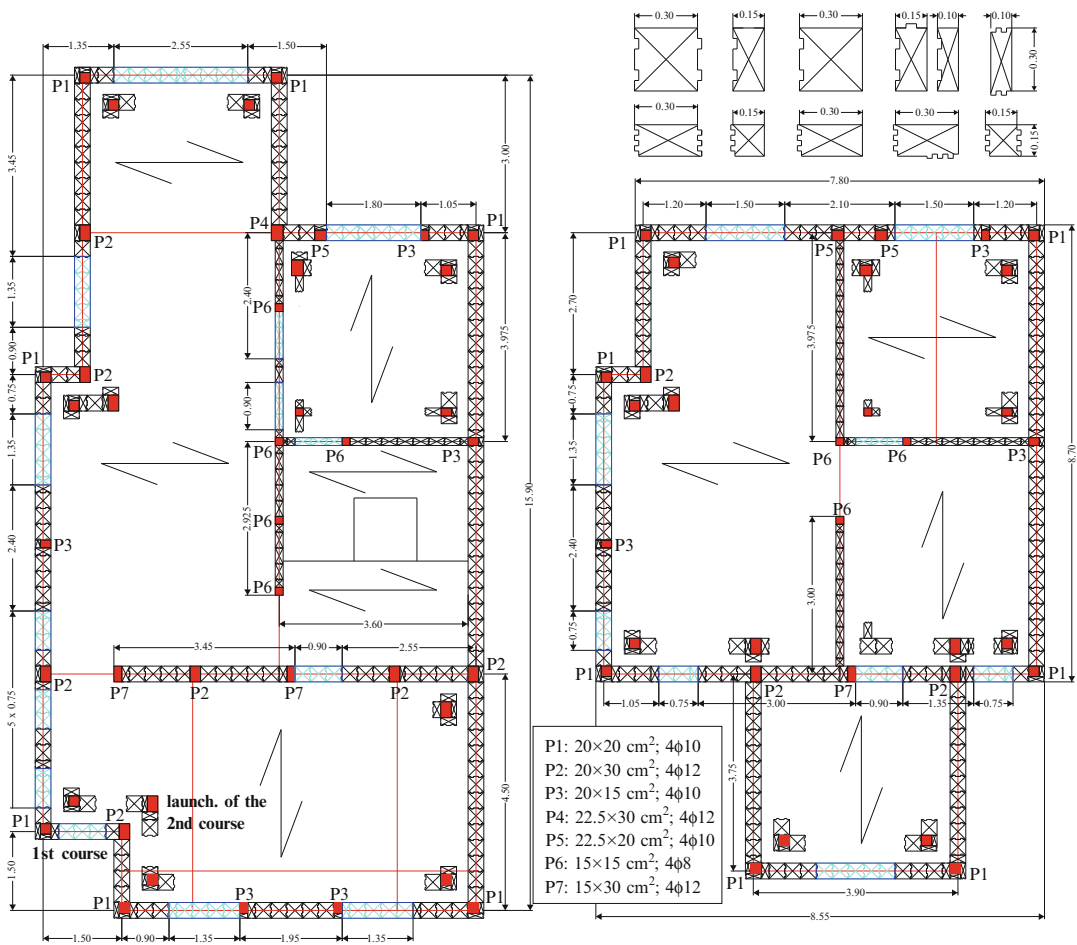


Masonry Box Behavior, Fig. 10 Predicted damage sequence corresponding to the (a) +X and (b) +Y analysis (X diagonal shear, = tension cracks, • flexural plastic hinge)

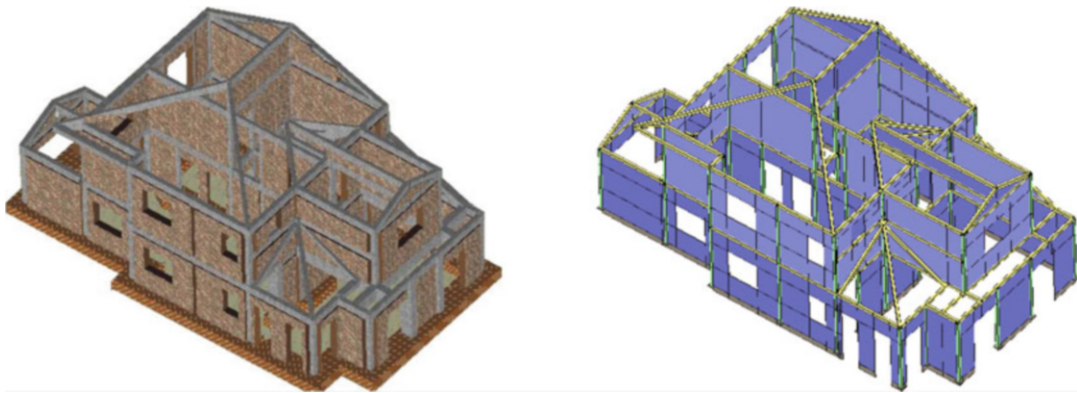
The capacity curves for both unreinforced masonry and confined masonry structural solutions are compared in Figs. 15 and 16, where the base shear is expressed as a percentage of the building weight (%g). In X, the weaker direction, the unreinforced masonry building presents a base shear strength of about 23 % of the weight and a deformation capacity higher than 10 mm. Regarding the confined masonry solution, an increase of the base shear capacity is clearly observed for the X direction, comparatively to the unreinforced masonry solution. Still, a sharp decrease of capacity is found after peak due to a small number of confined panels and the absence of walls at the first story in the left part of the building. The high degradation of the

confined walls at peak load generates a ground story mechanism. Concerning the Y direction, a remarkable improvement of the base shear strength and ductility is observed.

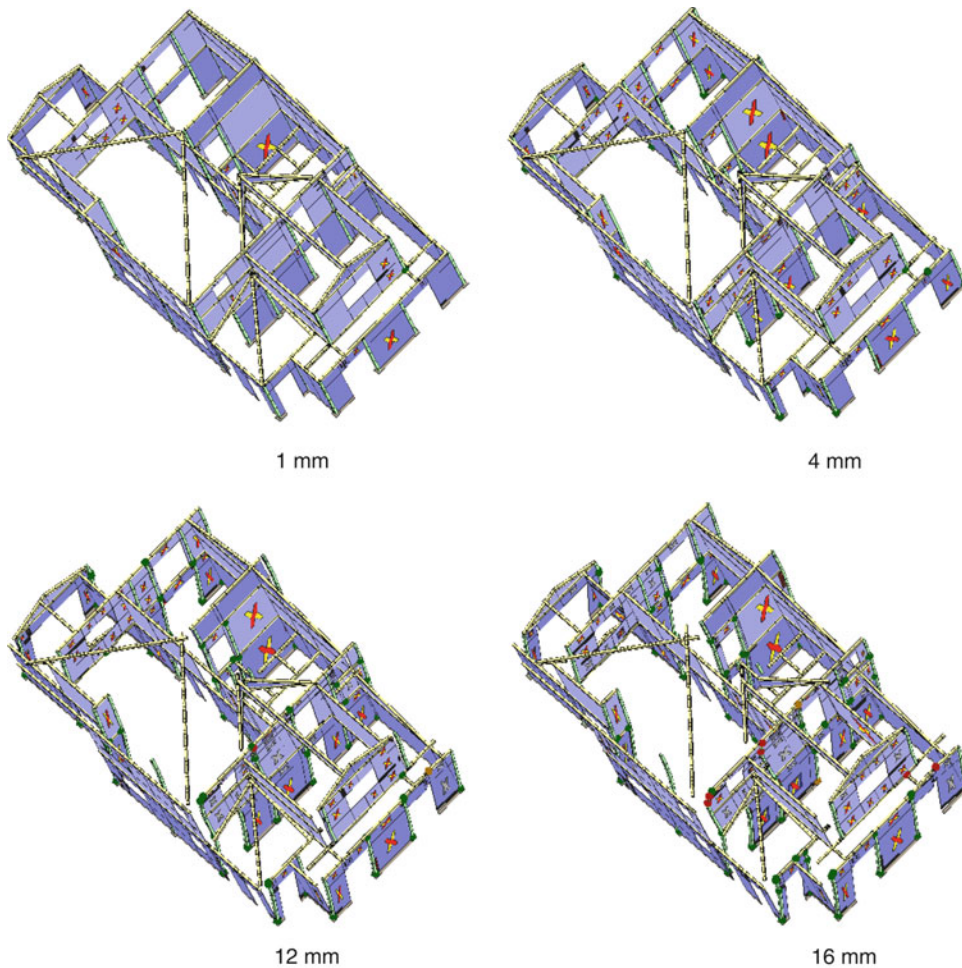
Finally, considering that the building is subjected to a type 2 spectrum (“near-field earthquake”) in Eurocode 8 (EN 1998-1 2004), and from using the N2 method (Fajfar and Fischinger 1988) in the software code by Gruppo Sismica (2013), an evaluation is made of the values of the reference peak ground acceleration on type A ground (rock formations), $a_{g,u}$, corresponding to the ultimate displacement allowed for the building. These values are presented in Fig. 17, which can be considered as an indicator of the seismic performance of the structure. The $a_{g,u}$



Masonry Box Behavior, Fig. 11 Structural plans of the constructive solution in confined masonry



Masonry Box Behavior, Fig. 12 Geometrical and computational models of a confined masonry building



Masonry Box Behavior, Fig. 13 Damage sequence for the +X analysis with first-mode load distribution (X diagonal shear, = tension cracks, • flexural plastic hinge)

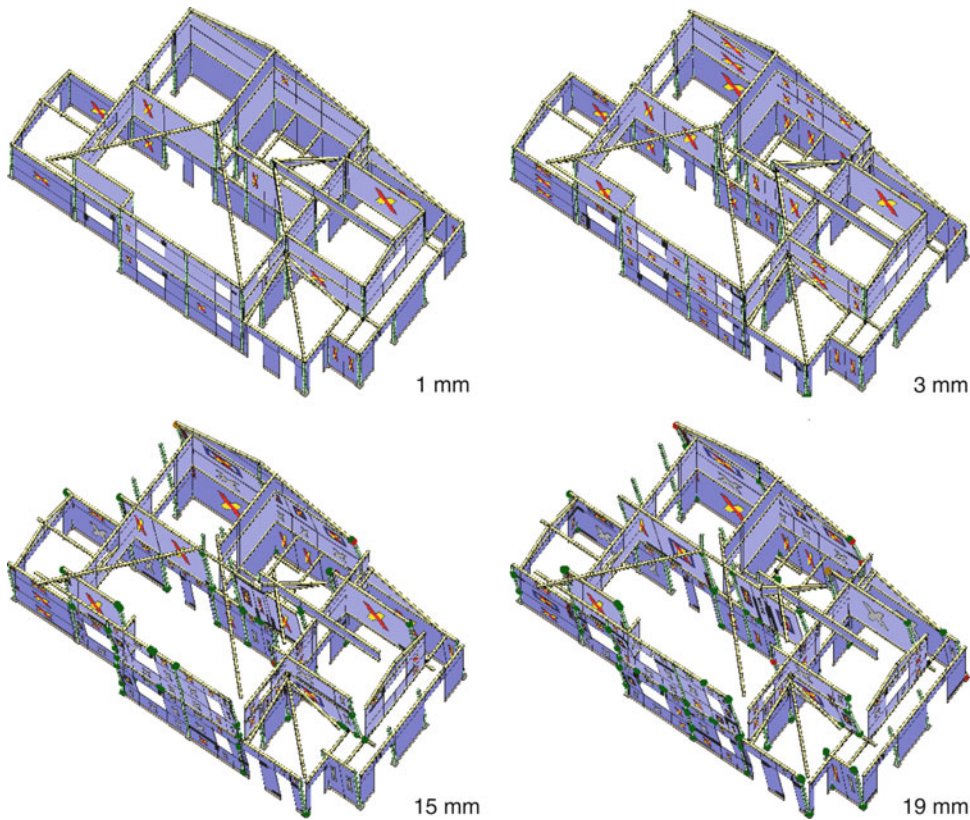
can be interpreted as the maximum ground acceleration supported by the building for a condition of near collapse or ultimate state, which can be directly compared with the design ground acceleration for a global safety verification.

Conclusions and Future Directions

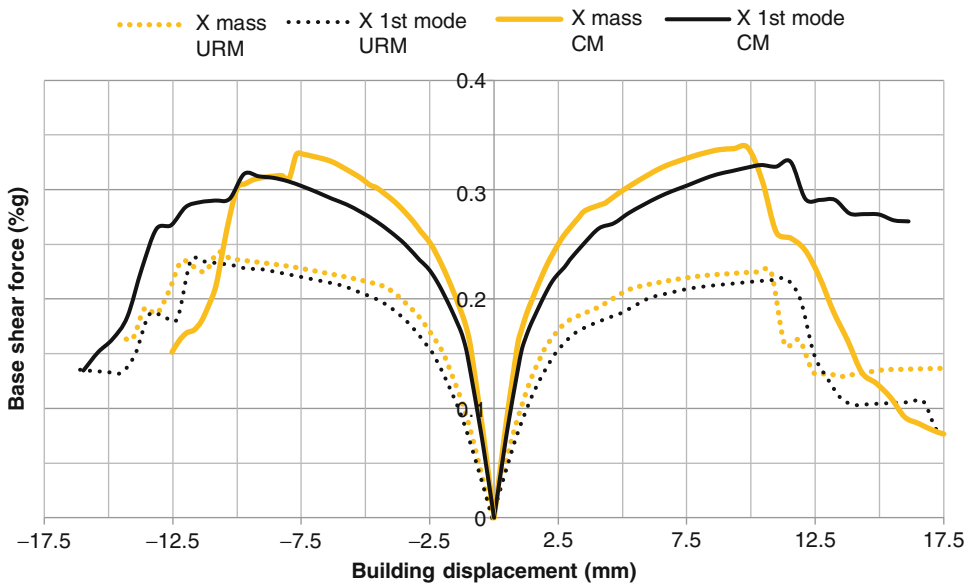
Masonry constructions are, more than a legacy from the past, a solution for the future, given the existing development in the sector. The box behavior is a fundamental feature of masonry buildings, which allows the structures to present a significant ductility when subjected to

earthquakes. This ductility is reflected in the capacity of inelastic displacements for the masonry buildings, allowing them to support significant ground accelerations. Using a typical house in southern Europe as a case study, structural masonry typologies featuring box behavior allow ensuring seismic safety up to large ground acceleration levels, namely, 0.15 and 0.20 g for unreinforced masonry and confined masonry solutions, respectively.

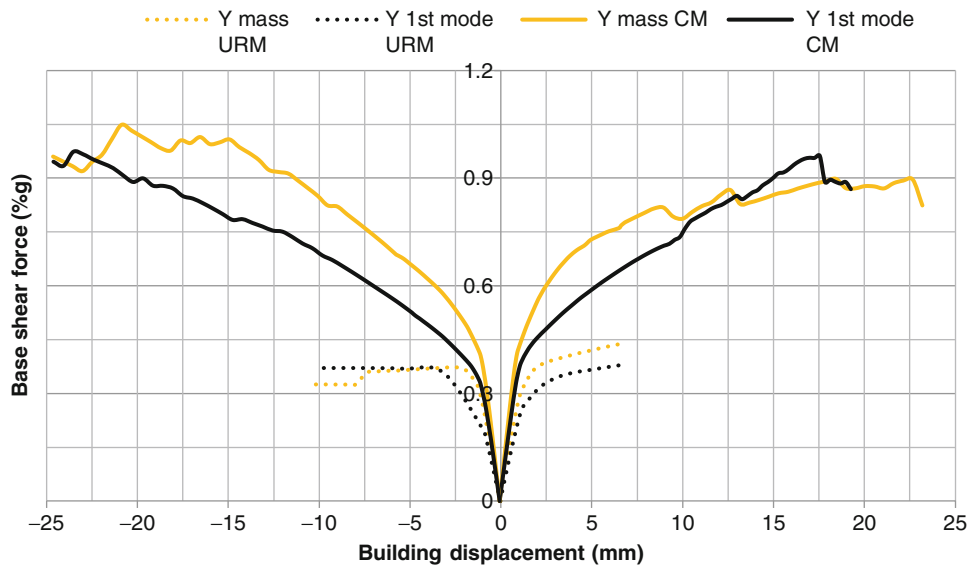
The box behavior is then a request for earthquake-resistant masonry construction, which can be obtained through the adoption of slab solutions allowing for an enough diaphragm effect and good floor-to-wall connection.



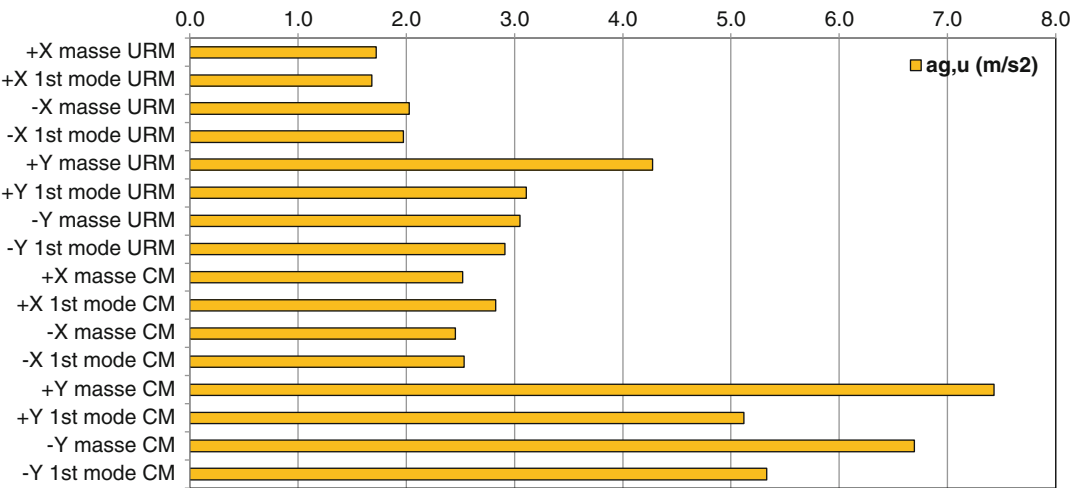
Masonry Box Behavior, Fig. 14 Damage sequence for the +Y analysis with first-mode load distribution (X diagonal shear, = tension cracks, • flexural plastic hinge)



Masonry Box Behavior, Fig. 15 Capacity curves in X direction for the unreinforced masonry and confined masonry models



Masonry Box Behavior, Fig. 16 Capacity curves in Y direction for the unreinforced masonry and confined masonry models



Masonry Box Behavior, Fig. 17 Evaluation of the maximum acceleration capacity of the building

The supply of box behavior needs also to be considered when retrofitting existing masonry buildings with flexible timber floors, e.g., through the adoption of a multi-plank system for the floors. The development of solutions to provide box behavior to social housing buildings in developing countries is also a problem that requires study. This is, for example, the case of traditional adobe masonry constructions in mountainous regions of Peru, which are usually covered with

lightweight wooden roofs poorly connected to the adobe walls.

Summary

Masonry box behavior is a feature of masonry structures that present floors significantly rigid and well connected to the walls. This is the case of new masonry buildings constructed with

a floor system of ribbed or reinforced concrete slab bonded to the walls by a ring beam, unlike existing buildings with timber floors. However, for this last case, a retrofit solution by stiffening the floor system and providing adequate wall-to-floor connections allows also a box behavior.

Nevertheless, the concept of box behavior is mainly applicable to current unreinforced and confined masonry buildings. The box action is a key assumption of macroelement methods developed for the seismic assessment of masonry buildings, when considering that the walls in a given direction respond jointly under lateral loading. Under this hypothesis, the seismic response of a building can be obtained through nonlinear static (pushover) analysis, i.e., submitting the building to an incremental static loading and considering the inelastic stage of behavior.

The pushover analysis provides the capacity curve (base shear force vs. displacement of the centroid of the roof slab) of the building, whose evolution depends on the sequence of cracking and failure events verified for the structure. The capacity curve is the reference to proceed with a performance-based seismic assessment, which is commonly based on the comparison of the displacement capacity and demand. Under this approach, the masonry box behavior normally allows the building to perform better when subjected to earthquakes.

This entry provides a review of current masonry typologies featuring a box behavior, and is explained with application to the seismic assessment of a dwelling house.

Cross-References

- [Behavior Factor and Ductility](#)
- [European Structural Design Codes: Seismic Actions](#)
- [Masonry Components](#)
- [Masonry Modeling](#)
- [Masonry Structures: Overview](#)
- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- [Response Spectrum Analysis of Structures Subjected to Seismic Actions](#)

References

- Brzev S (2007) Earthquake-resistant confined masonry construction. National Information Center of Earthquake Engineering (NICEE), Kanpur
- Brzev S, Astroza M, Moroni MO (2010) Performance of confined masonry buildings in the February 27, 2010 Chile earthquake. EERI report, Earthquake Engineering Research Institute, Oakland
- Calderini C, Cattari S, Lagomarsino S (2008) Numerical investigations on the seismic behaviour of confined masonry walls. In: Proceedings of the 2008 seismic engineering conference commemorating the 1908 Messina and Reggio Calabria earthquake. AIP Conference Proceedings 1020, New York, pp 816–823
- Calìò I, Marletta M, Pantò B (2012) A new discrete element model for the evaluation of the seismic behaviour of unreinforced masonry buildings. *Eng Struct* 40:327–338
- Calliari R, Manzini CF, Morandi P, Magenes G, Remino M (2010) User manual of ANDILWall, version 2.5. ANDIL Assolaterizi, Rome
- EN 1996-1-1 (2005) Eurocode 6: design of masonry structures – part 1-1: general rules for reinforced and unreinforced masonry structures. European Committee for Standardization, Brussels
- EN 1998-1 (2004) Eurocode 8: design of structures for earthquake resistance – part 1: general rules, seismic actions and rules for buildings. European Committee for Standardization, Brussels
- Fajfar P, Fischinger M (1988) N2 – a method for nonlinear seismic analysis of regular buildings. In: Proceedings of the 9th world conference on earthquake engineering, vol 5, Tokyo-Kyoto, pp 111–116
- Gouveia JP, Lourenço PB (2007) Masonry shear walls subjected to cyclic loading: influence of confinement and horizontal reinforcement. In: Proceedings of the 10th North American Masonry conference (paper 042), St. Louis
- Gruppo Sismica (2013) Theoretical manual of the 3DMacro software, beta version. Gruppo Sismica, Catania
- IBC (2008) Italian building code, Ministerial Decree dated of 14-01-2008. Ministero delle Infrastrutture e dei Trasporti, Rome
- Jäger W, Hirsch R, Masou R (2010) Product and system development in masonry construction under requirements of sustainable construction. In: Proceedings of the 8th international Masonry conference (keynote speech), Dresden (CD-ROM)
- Lagomarsino S, Penna A, Galasco A, Cattari S (2009) User guide of TreMuri (Seismic analysis program for 3D masonry buildings), version 1.7.34. University of Genoa
- Lourenço PB (2002) Computations of historical masonry constructions. *Prog Struct Eng Mater* 4(3):301–319
- Lourenço PB, Vasconcelos G, Gouveia JP, Medeiros P, Marques R (2008) CBloco: handbook of structural design. Cerâmica Vale da Gândara SA, Viseu
- Magenes G (2006) Masonry building design in seismic areas: recent experiences and prospects from

- a European standpoint. In: *Proceedings of the 1st European conference on earthquake engineering and seismology (keynote k9)*, Geneva
- Magenes G, Calvi GM, Kingsley R (1995) Seismic testing of a full-scale, two-story masonry building: test procedure and measured experimental response. In: *Experimental and numerical investigation on a brick masonry building prototype: numerical prediction of the experiment*. Report 3.0. Gruppo Nazionale per la Difesa dai Terremoti, Pavia
- Marques R, Lourenço PB (2011) Possibilities and comparison of structural component models for the seismic assessment of modern unreinforced masonry buildings. *Comput Struct* 89(21–22):2079–2091
- Marques R, Lourenço PB (2013) A model for pushover analysis of confined masonry structures: implementation and validation. *Bull Earthquake Eng.* 11(6): 2133–2150 doi:10.1007/s10518-013-9497-5
- Parsekian GA, Hamid AA, Drysdale RG (2012) Behavior and design of structural masonry. EdUFSCar-Editora da Universidade Federal de São Carlos, São Paulo
- Saiidi M, Sozen MA (1981) Simple nonlinear seismic analysis of R/C structures. *ASCE J Struct Div* 107(5):937–953
- San Bartolomé A (1994) *Masonry buildings: seismic behavior and structural design*. Fondo Editorial, Pontificia Universidad Católica del Perú, Lima
- Sullivan TJ, Priestley MJN, Calvi GM (eds) (2012) *A model code for the displacement-based seismic design of structures*, DBD12. IUSS Press, Pavia
- Tomažević M (1978) The computer program POR. Report ZRMK. Institute for Testing and Research in Materials and Structures, Ljubljana
- Tomažević M (1999) *Earthquake-resistant design of masonry buildings*, vol 1, Series on innovation in structures and construction. Imperial College Press, London
- Touliatos PG (1996) Seismic behaviour of traditionally-built constructions: repair and strengthening. In: *Courses and lectures-international centre for mechanical sciences*. Springer, New York, pp 57–70
- Turnšek V, Čačovič F (1970) Some experimental results on the strength of brick masonry walls. In: *Proceedings of the 2nd international brick Masonry conference*. British Masonry Society, Stoke-on-Trent, pp 149–156

Masonry Components

Graça Vasconcelos
Department of Civil Engineering, ISISE,
University of Minho, Guimarães, Portugal

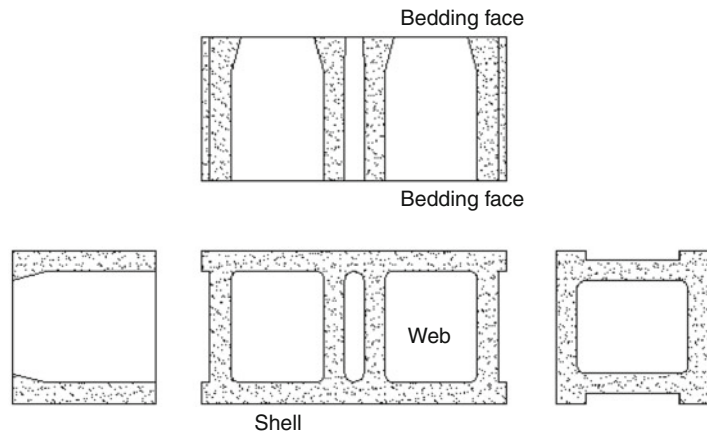
Synonyms

Masonry materials; Mechanical behavior

Introduction

Masonry is a nonhomogeneous material, composed of units and mortar, which can be of different types, with distinct mechanical properties. The design of both masonry units and mortar is based on the role of the walls in the building. Load-bearing walls relate to structural elements that bear mainly vertical loads but can serve also to resist to horizontal loads. When a structural masonry building is submitted to in-plane and out-of-plane loadings induced by an earthquake, for example, the masonry walls are the structural elements that ensure the global stability of the building. This means that the walls should have adequate mechanical properties that enable them to resist to different combinations of compressive, shear, and tensile stresses. The boundary conditions influence the resisting mechanisms of the structural walls under in-plane loading, and in a building, the connection at the intersection walls is of paramount importance for the out-of-plane resisting mechanism. However, it is well established that the masonry mechanical properties are also relevant for the global mechanical performance of the structural masonry walls. Masonry units for load-bearing walls are usually laid so that their perforations are vertically oriented, whereas for partition walls, brick units with horizontal perforation are mostly adopted.

As a composite material, the mechanical properties of masonry under different loading configurations depend on the properties of masonry components. The unit has an important contribution for the resisting mechanisms of masonry particularly in case of resisting mechanisms that are associated to crushing and tensile cracking of masonry. On the other hand, mortar joints acting as a plane of weakness on the composite behavior of masonry can control the shear behavior, which is particularly relevant in case of strong unit–weak mortar joint combinations. The mortar joints have also a central role on the flexural behavior in the direction perpendicular to the bed joints, as it is tightly controlled by the tensile bond adherence. In case of experimental characterization, flexural testing of masonry in the direction perpendicular to bed joints can even

Masonry Components,**Fig. 1** Example of a masonry unit: concrete block

be used as a means of obtaining the tensile bond strength.

Besides the review on masonry material components, it is important to present and discuss the most important mechanical properties of masonry assemblages, together with a discussion of the main parameters affecting the seismic performance of masonry under distinct loading configurations. Understanding the response of the basic masonry materials is essential for interpreting the seismic response of masonry structural systems.

Masonry Units

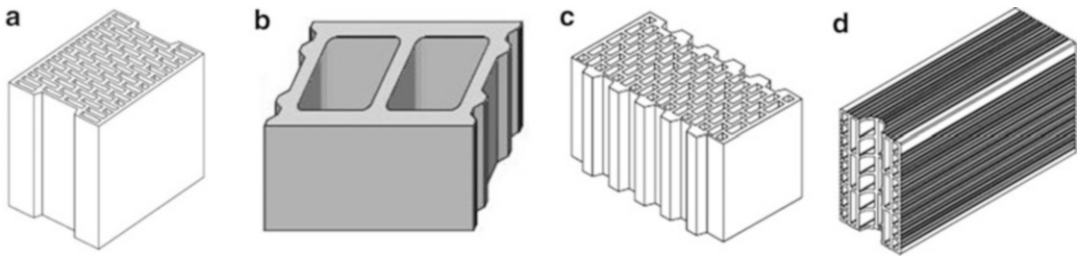
Masonry is a composite material composed of masonry units with a regular arrangement that are connected with mortar commonly at horizontal bed and vertical head joints. The interface between units and mortar represents in general an important role on the mechanical behavior of the composite material submitted to distinct types of loading.

The masonry units represent the fundamental material for the formation of the main body of the masonry structural element and can be made of distinct raw materials, namely, clay, mud, concrete, calcium silicate, and stone. However, the clay and concrete are far the most common raw materials used in structural masonry units.

The masonry units have commonly a rough rectangular shape, and the dimensions are defined

generally by the (length) $\times$ (height) $\times$ (width) and are laid usually according to the larger dimension (length). The length and height of the masonry units are usually multiples of 200 mm (nominal dimensions), including the 10 mm for the mortar thickness so that modularity of the structural elements can be achieved. The modularity is an important characteristic of masonry to make the construction technology and geometrical implementation of the structural elements (walls) with openings easier. The external vertical surface of the masonry units is known as the shell of the unit, and the walls perpendicular to the face are the webs of the units (Fig. 1). The top and bottom faces of the masonry units are known as the bedding areas.

The masonry units can be solid or have vertical and horizontal voids/perforations known as cores with smaller dimension and cells. Generally, solid units can have up to 25 % of perforations in relation to their gross area. For structural purposes, it is more common that concrete or clay bricks should have vertical perforations along their full height. The horizontally perforated brick units are more common for nonstructural purposes, namely, for masonry infill walls typically used in some south European countries. Examples of common masonry units are shown in Fig. 2. Clay units have generally a set of vertical cores with reduced area, whereas the concrete blocks commonly have large hollow cells, as seen in Fig. 2a. In vertical perforated units, the faces are called as face shells connected



Masonry Components, Fig. 2 Examples on different types of masonry units: (a) clay brick unit, (b) concrete block unit, (c) masonry unit with tongue and groove vertical joints, and (d) masonry unit with mortar pocket

by the internal solid parts called as webs. When the perforation does not go through the entire height of the unit, it is called a frog. This has usually a conical shape. Sometimes, the units have indented ends over the full height and are called as frogged ends. When it is intended to have dry vertical joints, without the addition of mortar, tongue and groove or interlocking systems are designed (see Fig. 2c). In this case, the out-of-plane resistance should rely on the combination of the bed joint resistance and on the tongue and groove system. Other times, the geometry of the units foresees the addition of mortar into vertical pockets (see Fig. 2d), where, depending on the geometry, different reinforcing systems can be added to improve the resistance of masonry to in-plane and out-of-plane loads.

The raw materials used in the manufacturing process of clay units are commonly surface clays (recent sedimentary formations) but shales formed from clays under pressure or fire clay, mined at deeper levels. All of these clays are equivalent in terms of silica and alumina compounds with different types of metallic oxides. The surface clays present a great variability, and in some cases, a mixture of clay of distinct locations can be used to reduce the variability. The material used in the concrete units is a dry concrete composed of Portland cement, stone aggregates, and water. It is also common to use other blended cements including blast-furnace cement and fly ash and inert fillers, considered commonly as by-products, aiming at reducing the percentage of Portland cement. In other instances, expanded clay aggregates are used to reduce the weight of

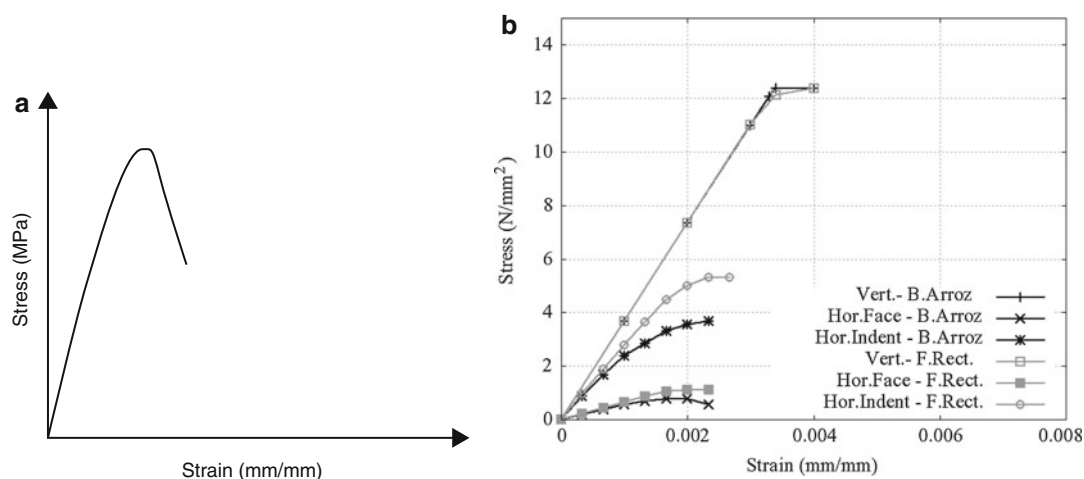
the concrete units. Additive such as pozzolanic materials and other workability agents can be also used. The calcium silicate units are composed of sand and hydrated lime.

Besides the raw materials, the production technologies used to produce the clay, concrete masonry, and calcium silicate units are very different. The clay units are normally extruded and fired at different temperatures, whereas the concrete units are produced in molds with a required geometry through a vibration and pressing process. The calcium silicate units are manufactured by pressing the mixture of sand and hydrated lime and then autoclaving them in order to produce a tightly grained unit.

Mechanical Properties

The most relevant mechanical properties of masonry units consist of the compressive strength, elastic modulus, and tensile strength. The mechanical behavior of the masonry assemblages depends greatly on the mechanical properties of the masonry units.

The compressive strength of the unit can be seen as a measure of its quality, and it is important for predicting the compressive strength of masonry assemblages. The compressive strength and elastic modulus can be obtained experimentally from uniaxial compressive tests according to the European standards. The compressive strength is calculated from the loaded area, which is the gross area (length) $\times$ (width) of the unit when the units are oriented in the same way as they are intended to be laid in a bed of mortar. In general, an average value is obtained from the experimental results, being possible to calculate

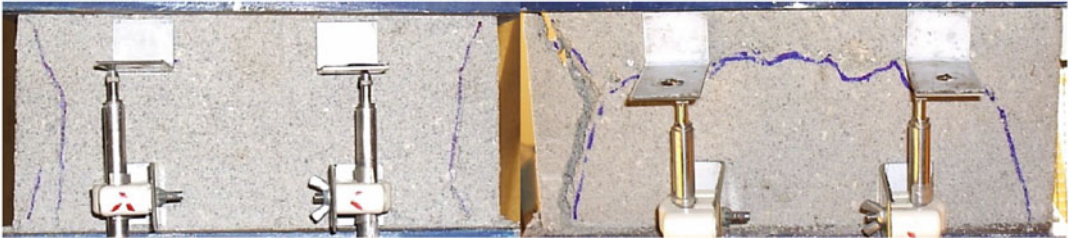


Masonry Components, Fig. 3 Details about compression behavior of masonry units: (a) typical stress–strain diagrams and (b) effect of loading direction on the compressive strength of units

the characteristic value and the corresponding normalized compressive strength of the masonry unit, f_b , by multiplying the average values by a coefficient taking into account the moisture environment of the curing conditions (oven dry as a reference) and also by the shape factor, accounting for the dimensions of the width and the length. The modulus of elasticity can be calculated as a secant modulus of elasticity between 0 % and 33 % of the compressive strength of masonry unit in the stress–strain diagram obtained from the uniaxial compressive tests. This property can be important if advanced modeling of masonry is required.

In case of modern clay and concrete masonry units, considerably high values of compressive strength can be achieved, being generally higher than the strength requirements for units to be used in seismic zones. It is common to have an average compressive strength higher than 10 MPa. It should be noted that the compressive strength of masonry units is different from the compressive strength of the raw material due to the effect of the shape and geometry of the units. In spite of attempts that have been made to obtain the complete stress–strain diagram of masonry units in compression, it is hard to obtain the post-peak branch of the stress–strain diagrams describing the high-rate crack damage progress of the units after the maximum load is attained (see Fig. 3a).

It should be also noticed that the compressive strength of the masonry units can differ significantly according to the loading direction, namely, in the directions perpendicular and parallel to the laying and in the direction perpendicular to the face. According to the work carried out by Lourenço et al. (2010), it was observed that compressive strength is considerably higher in the direction perpendicular to the bed joints, due to the orientation of vertical perforations, in comparison with when the direction is parallel to the bed joints. A reduction of more than 30 % in the normalized compressive strength obtained in the parallel direction to the bed joints was also found experimentally for concrete units, as seen in Fig. 3b. This difference is attributed mainly to the geometry of the masonry units with distinct arrangements of the internal perforations and cells. This results naturally in the different compressive behavior of masonry under compression for the different loading directions. The failure modes recorded in clay and concrete masonry units confirm its brittle character. In clay units with vertical perforation, it is common to observe cracking and splitting of the internal webs and shells. In case of concrete masonry units, the failure mode has a commonly pyramidal trunk (Gihad et al. 2007; Haach 2009) (see Fig. 4). The first cracks appear vertically in corners of the units.



Masonry Components, Fig. 4 Crack patterns of concrete blocks under uniaxial compression

With increase of the load, there was a tendency of the connection of vertical cracks by a horizontal crack in the upper region of the unit. This behavior can be explained by the lateral restrictions caused by the steel plates in top and bottom of the specimen, generating friction forces. This horizontal crack occurs because the upper part of the units slides over the pyramidal-trunk surface of rupture. In some specimens near the collapse limit, a vertical crack also appeared in the central region of the unit. The brittle failure mode of the masonry units can influence the seismic performance of the masonry under shear walls due to local failures determining the failure mode of the walls (Tomažević et al. 2006).

Mortars for Masonry

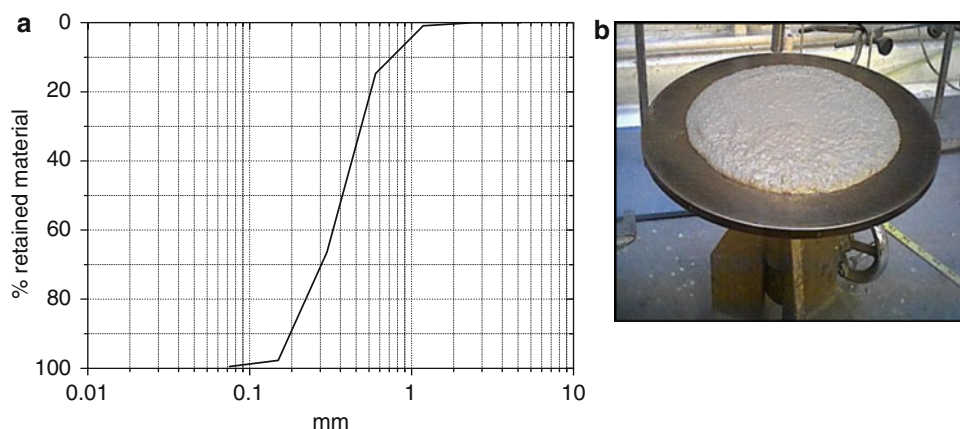
Mortar is a component of masonry, and it is used to bond individual masonry units into a composite assemblage. It has a central role in the stress transfer among units when masonry is loaded by promoting and more uniform bearing and avoiding stress concentrations that can result in the premature collapse of masonry. The mortar has also the role of smoothing the irregularities of blocks and accommodating deformations associated with thermal expansion and shrinkage. As pointed out by Vasconcelos and Lourenço (2009), the deformability of masonry is clearly influenced by the material at the bed joints. Very distinct pre-peak behavior was found by considering dry saw unit–mortar interfaces, rough dry joints, lime mortar, or dry clay resulting from sieving granitic soil. The mortar also influences the bond strength (tensile and shear) of the joints.

The mortars for laying masonry units and filling the vertical joints are commonly a combination of Portland cement, lime, sand, and water in specified proportions. The strength of mortars is controlled by the cement, and the workability is controlled mainly by lime and by the grading of the sand, as shown in Fig. 5a. The sand can be natural or artificial resulting from crushing stone. The size and grading of the sand particles influence both the plastic and hardened properties. More graded sand (increased variation of the size and distribution of particles) contributes to improve workability of mortars, which play a major role on the laying process of masonry units. The workability can also be improved through the use of additives like clay fillers and air entrainment. Mortar mixes can be defined by specific proportions of the compounds in volume or in weight of the cement or binder (cement and lime) content. For example, the mortar mix defined by the trace 1:2:9 (cement:lime:sand) by volume means that it has double the volume of lime in relation to cement and has sand with a volume nine times the volume of cement. It can be considered also that the mortar mix has three times more sand in volume than the total binder of the mixture (cement and lime).

Properties of Fresh Mortar

The knowledge about the fresh and hardened properties of mortar is fundamental in ensuring a good performance of masonry walls. The most important properties in the fresh state of mortars are the workability, air content, and setting time (rate of hardening).

The workability of mortars plays an important role on the construction process of masonry



Masonry Components, Fig. 5 (a) Grading curve for sand and (b) measurement of mortar flow

structures. Workability may be considered as one of the most important properties of mortar, and it is related to the process of laying masonry units, and, thus, it influences directly the bricklayer's work. On the other hand, it is important to mention that the quality of the workmanship can influence considerably the mechanical properties of masonry. A workable mortar is easy to adhere to the surface of the trowel, slides off easily, spreads readily, and adheres easily to vertical surfaces. The workability can be improved by the addition of air entrainment agents to the cementitious materials, enhancing also the durability. The addition of lime and the use of an appropriate curve grade for the sand influence also positively the workability of mortar (Haach et al. 2011). The workability is an outcome of several properties such as consistency, plasticity, and cohesion. Given that plasticity and cohesion are difficult to measure, consistency is frequently used as the measure of workability. The consistency is obtained experimentally by the flow table test, as shown in Fig. 5b. An acceptable value for workability for masonry construction ranges from 150 to 180 mm.

The water retention is the property of the mortar that avoids the rapid loss of the mixing water in the masonry units and to the air, and it plays a major role on the bond adherence of the mortar to the masonry units. The ability of the mortar to retain water is important to prevent the excessive stiffening of the mortar before it is used in the

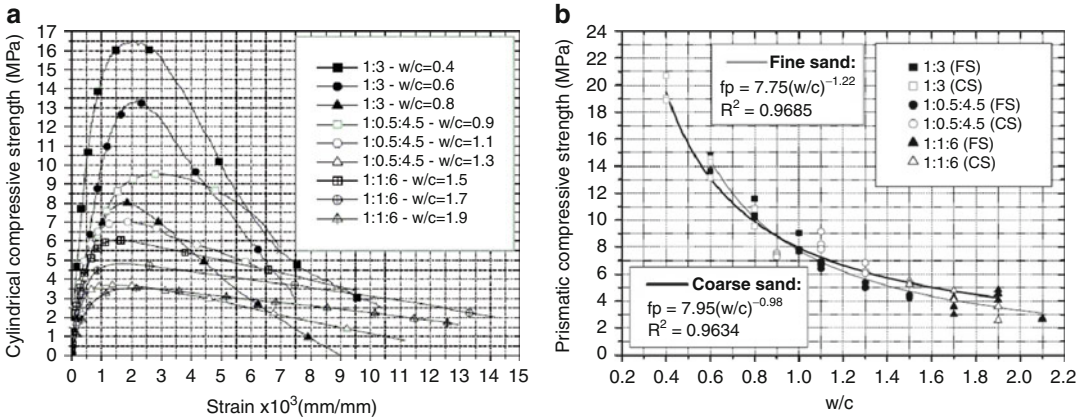
laying of masonry units to ensure an adequate hydration of cement and to prevent the water from bleeding out of the mortar. The ability of the mortar to retain water is related to the masonry unit's absorption and should be higher for higher absorption units.

The setting time of fresh mortar relates to the hardening process of mortar. If the setting time is low, the mortar can extrude out of the joints as laying is carried out. If the setting time is high, the mortar placing on the joints can be difficult. The proper hardening rate of the mortar contributes for the adequate bond to the masonry units.

Properties of Hardened Mortar

The most important mechanical properties of hardened mortar are compressive strength and bond. The bond presents a central role, not only in the mechanical performance of masonry under different loading configurations (shear, tension), but it is also important for the durability of masonry.

The bond between mortar and masonry units develops through mechanical interlocking resulting from its adherence. The bond can in certain extent result also from chemical adhesion. Several factors influence the bond between mortar and masonry units, such as properties of masonry units, type of mortar, water–cement ratio, air content, workmanship, workability, and curing conditions. The initial water absorption of the masonry units should be compatible



Masonry Components, Fig. 6 Behavior of hardened mortar: (a) stress–strain diagrams and influence of the addition of lime in the compressive strength and (b) influence of the w/c ratio in the compressive strength

with a good workability and appropriate water retention of the mortar to avoid rapid absorption of the water by the units from the mortar, reducing considerably the water availability for the hardening. The roughness of the masonry units is also important, being enhanced by rougher surfaces of the masonry units. More workable mortars result in better penetration through the voids of masonry units, improving the mechanical interlocking between mortar and masonry units. The additives that are placed in the mortar mix to enhance the workability also contribute to the enhancement of the bond.

Although it is known that the mortar plays a major role in the deformation of masonry under compression, it has also some influence on its compressive strength. More deformable mortar, with lower modulus of elasticity, increases the deformability of masonry under compression. The compressive strength of mortar is also used as an indicator of the workmanship quality, being common to take some mortar specimens during the construction for posterior testing and comparison with the compressive strength required. The compressive strength of mortar is affected by several factors, such as the cement content, the addition of lime, and water–cement ratio. The cement gives the strength to mortar, and if it is replaced by a certain quantity of lime, the compressive strength reduces, as shown in Fig. 6a. Additionally, the compressive strength of mortar

is also strongly reduced by the increase in the water–cement ratio (w/c) (Fig. 6b).

Masonry as a Composite Material

The masonry is considered as a composite material composed of units and mortar and unit–mortar interfaces, and its mechanical behavior depends on the mechanical characteristics of the elements and also on its arrangements. The loading configurations to which masonry is subjected depend on the structural element to which it belongs.

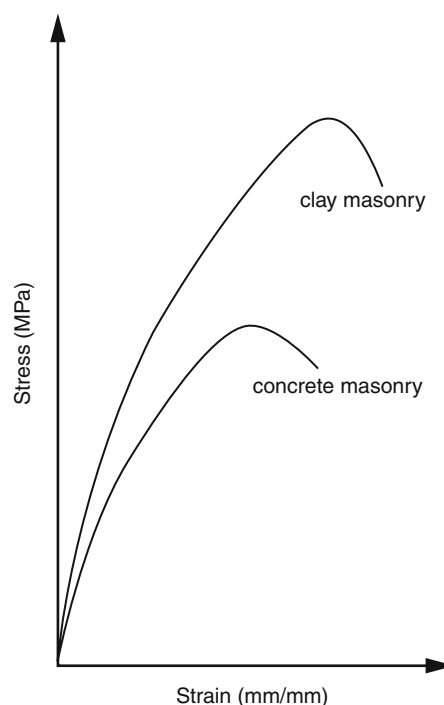
Compressive Behavior

The compressive strength of masonry is the primary mechanical property characterizing its structural quality and is fundamental for structural stability in case of load-bearing masonry walls. Compressive behavior is also important when masonry is subjected to lateral loading because the in-plane behavior depends on the compressive properties of masonry, especially if flexural resistance mechanisms predominate (Haach et al. 2011). The finite element numerical analysis of masonry walls based on macro-modeling also requires the data regarding the mechanical behavior of masonry under compression and the key mechanical properties, namely, the compressive strength, elastic modulus, and

fracture energy. Masonry is a composite material made of units and mortar, so it has been largely accepted that its failure mechanism and resistance is governed by the interaction between the different components.

In case of hollow or solid units with full mortar bedding and when the mortar has lower compressive strength than the masonry units, the cracking paths and overall behavior are considerably controlled by mechanical properties of mortar and masonry units. As the mortar has generally lower modulus of elasticity than the masonry unit, it exhibits a trend to expand laterally within the mortar joints more than the masonry, being restrained by the masonry units. This interaction results in a triaxial compression state of the mortar and on a compression-lateral tensile state on the masonry units. This stress state results in the vertical cracking of the units, as seen commonly in the experimental testing of masonry. In case of masonry units with face shell mortar bedding, it is common to find vertical cracking along the webs of the units. This is related to the nonuniform stress distribution of stresses along the height of the unit and along the thickness of the masonry and to the principal tensile stresses mainly at the top and bottom of the units (Lourenço et al. 2010).

The typical stress–strain diagram describing the compressive behavior of masonry is shown in Fig. 7. The pre-peak behavior is characterized by nonlinearity beyond approximately 60 % of the peak stress, particularly in concrete masonry. The clay brick masonry tends to present a more linear elastic behavior with nonlinearly close to the peak load, achieving also higher values of the compressive strength. Almost no post-peak is usually recorded, which is associated with the brittle character of unreinforced masonry under compression. However, the post-peak behavior of masonry is dependent on the type of units and also dependent on the mortar used. Concrete masonry specimens built with lime-based mortar present slight lower strength than masonry prisms built with cement-based mortar. Additionally, higher deformations characterized the compressive behavior of masonry built with lower strength mortar. In this case, the ability



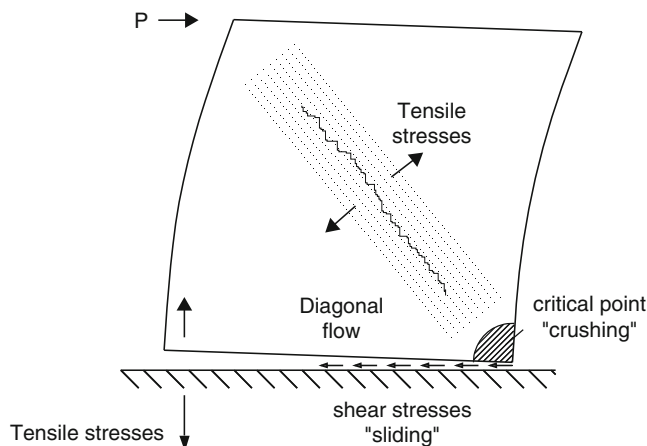
Masonry Components, Fig. 7 Typical stress–strain diagram of concrete masonry under compression

deformation after peak load is higher, enabling more ductile response of masonry under compression (Haach et al. 2014). The compressive strength of masonry is always higher than the compressive strength of mortar and lower than the compressive strength of masonry unit. However, the compressive strength of masonry units takes a central role on the compressive strength of masonry. Higher strength masonry units lead to higher strength of masonry. The relation appears to be linear in concrete block masonry (Drysdale and Hamid 2005).

The direction of compression of masonry units is also an important factor to take into account. Even if, in general, masonry is load in the direction normal to bed joints, in case of masonry beams and flexural walls (out-of-plane loading), the compression in the parallel direction to the bed joints is relevant for their mechanical behavior. The compressive strength in the parallel direction to the bed joints reduces considerably in relation to the compressive strength in the normal direction to the bed joints, particularly in

Masonry Components,

Fig. 8 Typical failure patterns that can develop in masonry walls under shear



case of hollow units as it depends on the geometry of perforations. In case of clay brick units with vertical perforation, the compressive strength in the parallel direction to the bed joints is about 30 % of the compressive strength of the units in the normal direction to the bed joints, resulting in the lowering of the compressive strength of masonry in this direction. The failure modes are also distinct, being the failure in the parallel direction more ductile.

Masonry Under Shear

The main resisting mechanisms that are characteristic of the response of the masonry walls submitted to combined in-plane loading are shear and flexure, which results in distinct failure modes (Fig. 8). In general, in squat walls, shear resisting mechanism predominates, and in slender walls, the flexural resistance mechanism plays the major role. Low pre-compression load levels are associated to flexural resisting mechanisms, and high pre-compression load levels are in general associated with the development of more dominant shear resisting mechanism. The shear resisting mechanism is associated with diagonal cracks in the alignment of the compressive strut related to the tensile stresses developed in the perpendicular direction of the strut. Diagonal shear cracking can occur with distinct patterns, namely, cracks developing along the unit-mortar interfaces or through both unit and mortar interfaces and through masonry units as a combination of joint failure or brick shear-tension

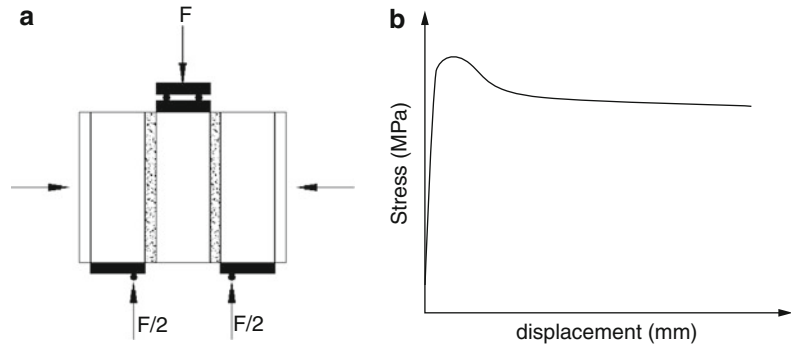
splitting. In diagonal cracking along the unit-mortar interfaces, the shear behavior of the bed joints plays an important role on the response of the walls.

Shear Behavior Along the Bed Joints

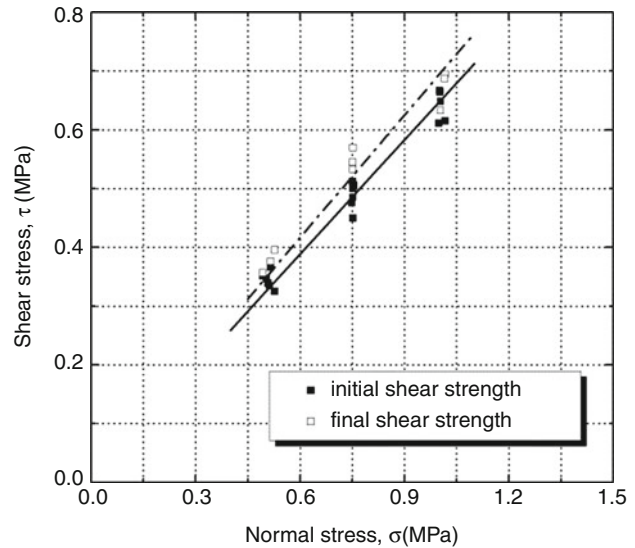
The influence of mortar joints acting as a plane of weakness on the composite behavior of masonry is particularly relevant in case of strong unit-weak mortar joint combinations. Two basic failure modes can occur at the level of the unit-mortar interface: tensile failure (mode I) associated with stresses acting normal to joints and leading to the separation of the interface and shear failure (mode II) corresponding to a sliding mechanism of the units or shear failure of the mortar joint. The preponderance of one failure mode over another or the combination of various failure modes is essentially related to the orientation of the bed joints with respect to the principal stresses and to the ratio between the principal stresses.

The shear behavior of mortar masonry joints is characterized experimentally based on direct shear tests by following the typical shear test configuration as shown in Fig. 9a. The typical behavior of mortar masonry joints under increasing shear load and constant pre-compression load is presented in Fig. 9b. The general shape of the shear stress-shear displacement is characterized by a sharp initial linear stretch. The peak load is rapidly attained for very small shear displacements. Nonlinear deformations develop in the

Masonry Components,
Fig. 9 Typical shear stress–slipping diagram of mortar masonry joints under shear



Masonry Components,
Fig. 10 Relation between shear stress and normal stress



pre-peak regime only very close to the peak stress. After peak load is attained, there is a softening branch corresponding to progressive reduction of the cohesion of the joint, until reaching a constant dry-friction value. This stabilization is followed by the development of large plastic deformations.

After peak load is attained, there is a softening branch corresponding to progressive reduction of the cohesion of the joint, until reaching a constant dry-friction value. This stabilization is followed by the development of large plastic deformations.

In case of moderate pre-compression stresses, for which the nonlinear behavior of mortar is negligible and the friction resistance takes the central role, the shear resistance of masonry bed joints is linearly dependent on the compressive

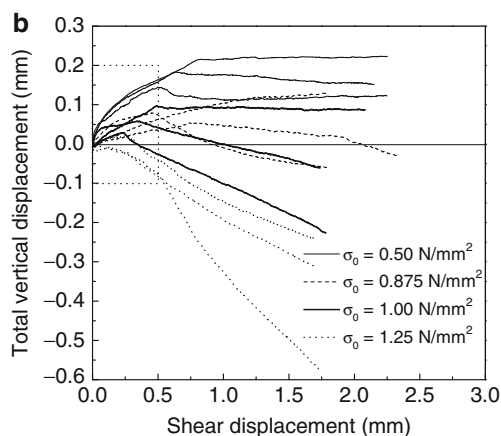
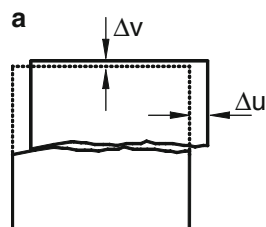
stress (see Fig. 10) and is given by the Coulomb friction criterion:

$$\tau = c + \mu\sigma \quad (1)$$

where c is the shear strength at zero compressive stress (usually denoted by cohesion) and μ is the friction coefficient or tangent of the friction angle. For dry joints, the cohesion is obviously zero. It should be kept in mind that the failure envelope given by Eq. 1 describes only a local failure of masonry joints and cannot be directly assumed as the shear strength of the walls submitted to in-plane horizontal loads (Mann and Müller 1982; Calvi et al. 1996). In any case, the shear bond strength of masonry joints assumes a major role on the shear resistance when it can be

Masonry Components,

Fig. 11 Dilation of shear mortar joints: (a) definition and (b) effect of vertical pre-compression on the dilation



described by the Coulomb friction criterion (diagonal cracking along the unit–mortar interfaces). The shear bond strength of masonry units can be seen as the initial shear strength used to calculate the shear strength according to the Coulomb friction criterion, as suggested by Eurocode 6 (2005).

The strength values, particularly the cohesion, are greatly dependent on the moisture content, porosity of the units, initial rate of absorption of the units, the strength and composition of mortar, as well as the nature of the interface (Amadio and Rajgelj 1991). More plastic mortar enhances shear behavior of joints by promoting better adherence. Binda et al. (1994) pointed out that when strong mortar is considered, the strength of the units can also determine the shear behavior of the joints. In case of hollow concrete masonry, the mortar placed on the internal webs contributes considerably to the increase of the shear strength as it increases the mechanical interlocking. This implies that a wide range of shear strength values may be pointed out for various combinations of units and mortar. Mann and Müller (1982) indicated a mean friction coefficient of approximately 0.65 on brick–mortar assemblages and a cohesion ranging from 0.15 MPa up to 0.25 MPa, depending on the mortar grade. From the results of direct shear tests carried out by Pluijm (1999), the coefficient of internal friction ranges between 0.61 and 1.17, whereas cohesion varies from 0.28 MPa up to 4.76 MPa, depending on different types of units and mortar.

Another important issue regarding shear tests is the dilatant behavior of masonry joints. The dilatancy represents the difference between the variation on the normal displacements of the upper and the lower unit, Δv , as a result of the variation of the shear displacement Δu (see Fig. 11). The opening of the joint is associated with positive dilation, whereas negative values of dilatancy represent the compaction of the joint. The dilatancy of rock joints is mostly controlled by the joint roughness. In conjunction with the cohesion and the friction angle, the dilatancy is also required as a parameter for micro-modeling of masonry. As pointed out by Lourenço (1996), dilatancy in masonry wall structures leads to a significant increase of the shear strength in case of confinement.

In-Plane Tensile Strength

The other approach for the shear resistance of masonry shear walls is based on the Turnšek and Sheppard (1980) criterion, which is based on the assumption that diagonal cracking occurs when the maximum principal stress at the center of the wall reaches the tensile strength of the masonry.

The stress state is calculated by assuming that masonry is an isotropic and homogeneous material, which does not correspond to its actual behavior, since tensile strength is dependent on the orientation of the principal stress regarding the mortar bed joints. For height-to-width ratios (h/l) higher than 1.5, from which walls can be

considered as a solid in the Saint-Venant sense, the tensile principal stress can be calculated by Eq. 2:

$$\sigma_t = \sqrt{\left(\frac{\sigma_0}{2}\right)^2 + \tau_{\max}^2} - \frac{\sigma_0}{2} \quad (2)$$

being the vertical stress considered as the average stress, σ_0 , calculated as the ratio between the compressive load (N) and the area (txl) of the walls $N/(txl)$, and the horizontal stress is negligible. This assumption was confirmed by using photoelastic analysis as reported by Turnšek and Čačovič (1971).

Considering that the maximum shear stress, $\tau_{\max}$, assumes a parabolic distribution, the horizontal shear force corresponding to the opening of shear cracks, H_s , is derived by Eq. 2, presenting the following expression:

$$H_s = \frac{f_t}{b} \frac{lt}{b} \sqrt{1 + \frac{\sigma_0}{f_t}} \quad (3)$$

where f_t is taken as the tensile strength of masonry. The variable b takes the value of 1.5 for walls with height-to-length ratios larger than 1.5. In case of height-to-width ratios (h/l) ranging between 1.0 and 1.5, the shear stress distribution deviates from the parabolic shape, and the horizontal normal stress becomes different from zero. In case of unreinforced masonry, this force is considered as the shear resistance when the failure mode corresponds to the cracking involving the cracking of units along the diagonal compression strut.

The in-plane tensile strength of masonry, f_t , can be obtained experimentally through diagonal compression tests by following the recommendation of standard ASTM E519 (2002). The tensile strength of masonry is calculated through Eq. 4, assuming that in the center of the panel, a pure shear stress state develops corresponding to the tensile strength to the maximum principal stress given by Eq. 4:

$$f_t = 0.707 \times \frac{P}{A} \quad (4)$$

where P is the vertical load applied and A is the horizontal cross section of the specimens. The shear deformation is calculated based on Eq. 5, where ΔH and ΔV are the deformation measured along the compression and tension diagonals and g is the width of the diagonal of the panel. The shear modulus is calculated by the ratio between the shear stress and the shear deformation (Eq. 6):

$$\gamma = \frac{\Delta V + \Delta H}{g} \quad (5)$$

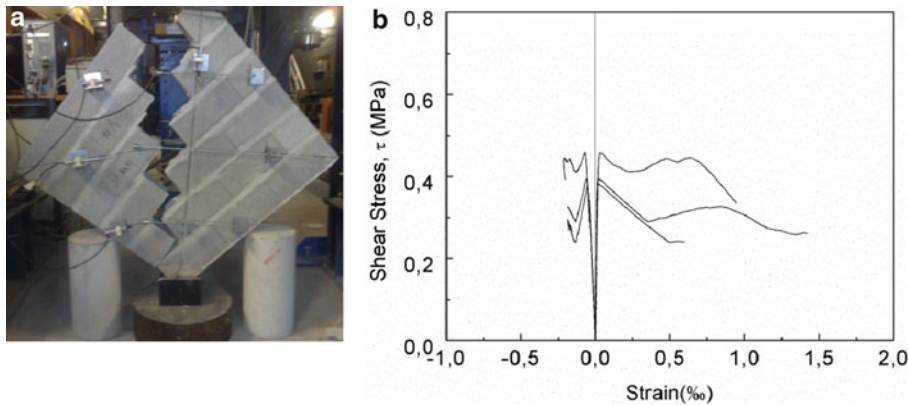
$$G = \frac{\tau}{\gamma} \quad (6)$$

The typical failure mode found in current modern unreinforced masonry composed of regular units and submitted to diagonal compression load results from the opening of a stair-stepped crack along the unit–mortar interface developing in the direction of load. The crack is developed in the perpendicular direction to the tensile stresses, which means that it appears when the tensile stress in masonry is reached. The failure of unreinforced masonry occurs suddenly in very brittle style (see Fig. 12a).

According to Haach (2009), the non-filling of vertical joints appears not to significantly influence the crack patterns and failure modes of unreinforced masonry, even if it can clearly influence the shear strength of masonry. The mortar type also influences the tensile strength as it influences the tensile and shear bond strengths of masonry joints, particularly in case of cracks that develop along the unit–mortar interfaces. The tensile strength of units influences the values of the tensile strength when the crack passes through the units.

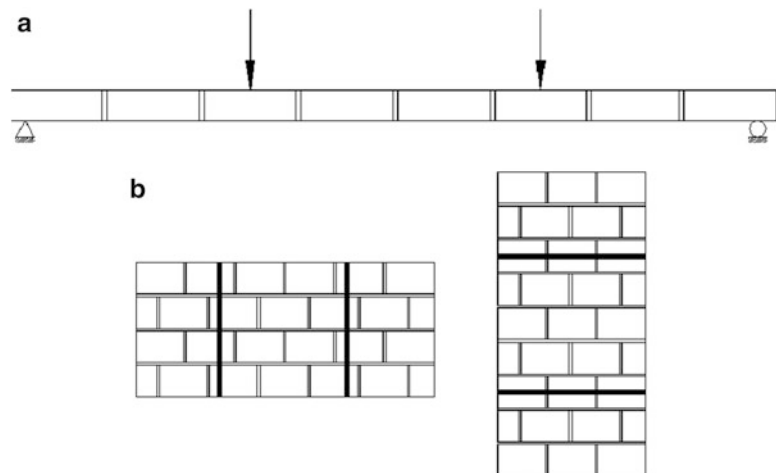
Flexural Strength

The flexural strength of masonry assemblages subjected to out-of-plane bending relates to the resistance of walls submitted to lateral loads from wind, earthquakes, or earth pressures. Depending on the boundary conditions and wall geometry, the bending can develop about vertical axis, about the horizontal axis, or about both directions. Thus, the tensile strength is referred to the



Masonry Components, Fig. 12 Details of diagonal compression tests on unreinforced masonry: (a) failure patterns and (b) typical shear stress–strain diagrams (negative values correspond to vertical strains)

Masonry Components, Fig. 13 Loading configuration for the experimental determination of the flexural strength of masonry: (a) geral scheme and (b) bending parallel and in the normal direction to the bed joints

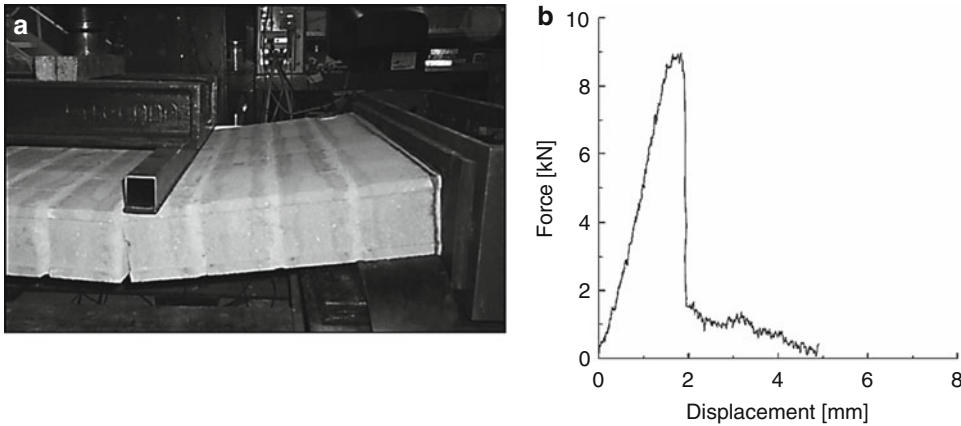


direction of the tension that can develop in the direction normal to the bed joints, f_{tn} , or in the direction parallel to the bed joints, f_{tp} .

The flexural strength of masonry units can be obtained experimentally according to EN 1052–2 (1999), by considering a four-point load testing configuration (see Fig. 13a) being the load applied typically according to the scheme as shown in Fig. 13b to obtain the flexural strength in the parallel and perpendicular direction to the bed joints.

The unreinforced masonry under flexure is characterized by a very brittle behavior, which is associated with the localized central crack involving the failure of the unit–mortar interface and the units (see Fig. 14a). When the flexure

develops in the normal direction to the bed joints, usually the crack patterns develop along a bed joint (de-bonding of the mortar from the masonry unit). In flexure parallel to the bed joints, the usually observed crack patterns are (1) stepped cracks along the unit–mortar interface, when masonry is made with strong units and weak mortar joints, and (2) cracks passing through head joints and masonry units. The flexural strength in the perpendicular direction to the bed joints can be also taken as the tensile bond strength of mortar joints. The typical force–displacement diagram relating to the vertical load applied and the maximum displacement measured at midspan, presented in Fig. 14b, confirms the brittle nature of masonry under flexure.



Masonry Components, Fig. 14 Details about the flexure behavior of unreinforced masonry: (a) crack pattern in flexure in the normal direction to the bed joints and (b)

typical force–displacement diagram (direction parallel to bed joints)

After peak load has been attained, there is an abrupt reduction of the bearing capacity, meaning that the masonry loses almost all resistance with no increment of displacement. The pre-peak regime is characterized by an elastic range with only a small nonlinearity very close to the peak load.

The flexural strength depends on the type of mortar, especially on the resistance and tensile bond strength. Also here the workability of mortar plays a major role as the tensile bond strength depends on the adequate adherence of the mortar to the unit. The flexural strength depends also on the tensile strength of the masonry units, particularly when flexure develops in the parallel direction to the bed joints.

Brief Code Considerations

The compressive strength of masonry can be estimated through empirical formulas generally based on the results of experimental tests. European masonry code (Eurocode 6 2005) proposes Eq. 7 to estimate the compressive strength of masonry:

$$f_k = k_b^{0.7} f_m^{0.3} \quad (7)$$

where k depends on the type and shape of units and mortar at bed joints, f_b is the normalized compressive strength of the unit, and f_m is the characteristic compressive strength of mortar.

For hollow clay units of group 2 and general-purpose mortar, k is 0.45.

The modulus of elasticity of masonry can be determined based on the experimental results, generally by taking the tangent value at 1/3 of the compressive strength of masonry in the stress–strain diagrams or by considering the secant values in a range between 0.1 and 0.4 of the compressive strength. It can be also estimated from the compressive strength of masonry. According to Eurocode 6 (2005), the elastic modulus can be obtained from Eq. 8:

$$E = k_E f_k \quad (8)$$

where k_E is recommended to be 1,000.

On the other hand, the values of shear modulus, G , used, for example, in the calculation of the lateral stiffness of masonry walls, can be estimated by multiplying the modulus of elasticity by 0.4 (Eurocode 6 2005).

In terms of shear, the Eurocode 6 (2005) suggests that the shear strength of masonry is calculated through Eq. 9:

$$f_{vk} = f_{vk0} + 0.4\sigma_d \quad (9)$$

where f_{vk0} is the characteristic initial strength of masonry, obtained for zero compressive stress, and σ_d is the average normal stress.

The value of the characteristic shear stress should not be higher than $0.065f_b$ (f_b is the normalized compressive strength of masonry units) neither exceed f_{vlt} , which should be defined in the National Annex.

The values of the initial shear strength given in Eurocode 6 (2005) depend on the type of unit (clay, calcium silicate, concrete, stone) and on the type of mortar (general-purpose mortar, light-weight mortar, or thin-layer mortar).

The characteristic flexural strength of masonry can be obtained by experimental tests or alternatively through the values suggested in Eurocode 6 (2005), depending on the type of unit and type of mortar. Typically, the characteristic flexural strength in the direction normal to the bed joints ranges from 0.1 to 0.2 MPa, and the characteristic flexural strength in the direction parallel to the bed joints ranges from 0.2 to 0.4 MPa.

There are requirements for the masonry units and mortar to be used in earthquake-prone regions. In case of masonry units, the normalized compressive strength should be higher than 5 MPa in the normal direction to the bed joints and 2.0 MPa in the parallel direction to the bed joints. The recommended values for the compressive strength of mortar are 5.0 MPa for unreinforced masonry and 10.0 MPa for reinforced masonry (Eurocode 8 2004).

Summary

In this section, a review on the masonry components and on mechanical properties of masonry under distinct loading configurations has been made. Additionally, some code considerations about the design mechanical properties of masonry are provided.

The masonry units have a wide range of possibilities either from the viewpoint of raw materials or from geometrical configurations. The geometrical configuration should comply with thermal and mechanical requirements to optimize the performance both from the mechanical and physical point of view. Besides the strength, it is required that the mortar should present an

adequate workability and water retention ability so that an adequate mechanical behavior of masonry is attained. These properties play a major role on the bond strength of masonry (tensile and shear bond strength), which in turn have an important contribution on the shear and flexural resistance of masonry.

The compressive strength of masonry is clearly dependent on the mechanical properties of the components, the masonry unit being more important than mortar, which contributes mainly to the deformability of masonry. Depending on the failure patterns, which are dependent on the level of vertical load applied and boundary conditions, the shear response of masonry walls under in-plane loading can be largely dependent on the shear bond strength (failure load described by the Coulomb friction criterion) or on the in-plane tensile strength of masonry (failure load described by the Turnšek and Čačovič criterion). The bond strength of the masonry joints and the in-plane tensile strength of masonry play an important role on the flexural resistance of masonry. The tensile strength is more important than the flexural strength in the direction parallel to the bed joints and the tensile bond strength of primary importance in case of the flexural strength in the normal direction to the bed joints.

Cross-References

► [Masonry Structures: Overview](#)

References

- Amadio C, Rajgelj S (1991) *Shear behavior of brick-mortar joints*. Masonry Int 5(1):19–22
- ASTM E 519–02 (2002) Standard test method for diagonal tension (shear) in masonry assemblages. Annual book of ASTM standards, American Society for Testing and Materials, West Conshohocken, USA
- Binda L, Fontana A, Mirabella G (1994) Mechanical behavior and stress distribution in multiple-leaf stone walls. In: Proceedings of 10th international brick block masonry conference, Calgary, pp 51–59
- Calvi GM, Kingsley GR, Magenes G (1996) Testing masonry structures for seismic assessment,

- Earthquake Spectra. *J Earthquake Eng Res Inst* 12(1): 145–162
- Drysdale R, Hamid AA (2005) *Masonry structures: behavior and design*, Canadian edition. Canada Masonry Design Centre, Mississauga. ISBN 0-9737209-0-5
- EN 1052–2 (1999) *Methods of test for masonry – Part 2: determination of flexural strength*, CEN - European Committee for Standardization, Brussels
- EN 1996-1-1 (2005) *Eurocode 6: design of masonry structures – Part 1–1: general rules for reinforced and unreinforced masonry structures*, CEN - European Committee for Standardization, Brussels
- EN 1998-1-1 (2004) *Eurocode 8: design of structures for earthquake resistance – Part 1: general rules, seismic actions and rules for buildings*, CEN - European Committee for Standardization, Brussels
- Gihad M, Lourenco PB, Roman HR (2007) *Mechanics of hollow concrete block masonry prisms under compression: review and prospects*. *Cem Concr Compos* 29(3):181–192
- Haach VG (2009) *Development of a design method for reinforced masonry subjected to in-plane loading based on experimental and numerical analysis*. PhD thesis, University of Minho, Guimarães, 367 pp
- Haach VG, Vasconcelos G, Lourenço PB (2011) *Parametric study of masonry walls subjected to in-plane loading through numerical modeling*. *Eng Struct* 33(4):1377–1389
- Haach VG, Vasconcelos G, Lourenço PB (2014) *Assessment of compressive behavior of concrete masonry prisms partially filled by general purpose mortar*. *J Mater Civ Eng* 26(10):04014068
- Lourenço PB (1996) *Computational strategies for masonry structures*. PhD thesis, Delft University of technology, Delft. ISBN 90-407-1221-2
- Lourenço PB, Vasconcelos G, Medeiros P, Gouveia J (2010) *Vertically perforated clay brick masonry for load-bearing and non-load-bearing masonry walls*. *Construct Build Mater* 24(11):2317–2330
- Mann W, Müller H (1982) *Failure shear-stressed masonry – an enlarged theory, tests and application to shear walls*, *Proc Br Ceram Soc* 30:223–235
- Pluijm RVD (1999) *Out-of-Plane bending of masonry, behavior and strength*. PhD thesis, Eindhoven University of Technology
- Tomažević M, Lutman M, Bosiljkov V (2006) *Robustness of hollow clay masonry units and seismic behavior of masonry walls*. *Construct Build Mater* 20(10): 1028–1039
- Turnšek V, Čačovič F (1971) *Some experimental results on the strength of brick masonry walls*. In: *SIBMAC proceedings*, London, pp 149–156
- Turnšek V, Sheppard P (1980) *The shear and flexural resistance of masonry walls*. In: *International research conference on earthquake engineering*, Sophie
- Vasconcelos G, Lourenço PB (2009) *Experimental characterization of stone masonry in shear and compression*. *Construct Build Mater* 23(11):3337–3345

Masonry Macro-block Analysis

Nuno Mendes

ISISE, University of Minho, Guimarães, Portugal

Synonyms

Macro-block; Macro-element

Introduction

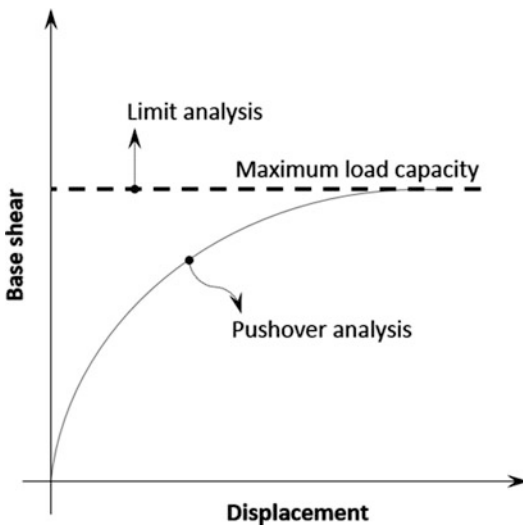
The structural analysis involves the definition of the model and selection of the analysis type. The model should represent the stiffness, the mass, and the loads of the structure. The structures can be represented using simplified models, such as the lumped mass models, and advanced models resorting the finite element method (FEM) and discrete element method (DEM). Depending on the characteristics of the structure, different types of analysis can be used such as limit analysis, linear and nonlinear static analysis, and linear and nonlinear dynamic analysis.

Unreinforced masonry structures present low tensile strength, and the linear analyses seem to not be adequate for assessing their structural behavior. On the other hand, the static and dynamic nonlinear analyses are complex, since they involve large time computational requirements and advanced knowledge of the practitioner. The nonlinear analysis requires advanced knowledge on the material properties, analysis tools, and interpretation of results. The limit analysis with macro-blocks can be assumed as a more practical method in the estimation of maximum load capacity of structure. Furthermore, the limit analysis requires a reduced number of parameters, which is an advantage for the assessment of ancient and historical masonry structures, due to the difficulty in obtaining reliable data.

The observation of the damage in masonry buildings caused by earthquakes in the past has been shown that the masonry structures can be discretized into several macro-blocks and interfaces.

The macro-blocks are portions of a structure with similar material properties and structural behavior, to which the mechanical properties of the material can be assigned or, by simplification, they can be assumed to be infinitely rigid. The interfaces in a macro-modelling represent, in general, the cracks associated to the failure mechanisms. However, in a micro-modelling strategy applied to masonry, in which the units and the joints are individually considered, the interfaces simulate the behavior of the joints. Furthermore, different criteria for the strength parameters of macro-blocks and interfaces can be considered.

The limit analysis (Fig. 1) with macro-blocks is a simplified and powerful structural analysis



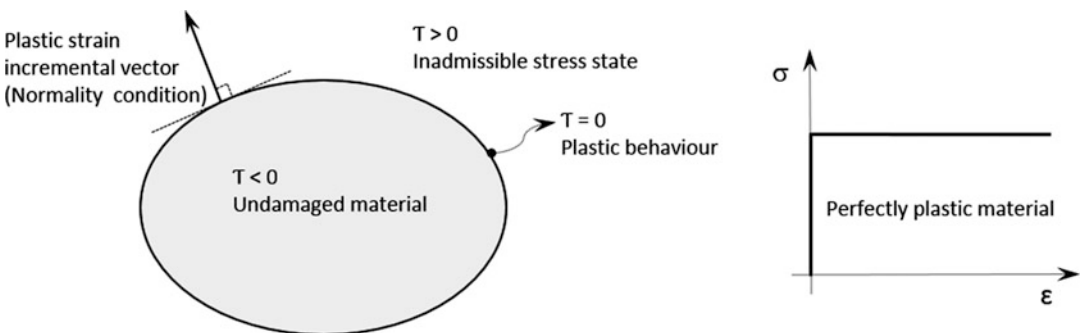
Masonry Macro-block Analysis, Fig. 1 Maximum load capacity obtained from the pushover analysis and limit analysis

tool to evaluate the ultimate capacity of masonry structures by static models, involving the equilibrium of the macro-blocks through the limit analysis basic concepts.

Classic Limit Analysis Concepts

In the classic limit analysis using rigid-plastic material behavior, a yield function τ based on stresses is defined, and three states are considered (Fig. 2): (i) if $\tau < 0$, the material presents rigid behavior and the material remains undamaged; (ii) if $\tau = 0$, the material becomes plastic and corresponds to the yield surface; (iii) if $\tau > 0$, the stress state is inadmissible. The points inside or on the yield function correspond to the states in which the stress can be sustained by the material, and the points outside correspond to stress states inadmissible for the material. The classic limit analysis considers that when the stress states are on the yield surface, the material becomes plastic and the flow direction is normal to the yield surface. This assumption is called as associated flow or normality condition. The associated flow implies that the yield surface must be convex and provides the maximum energy dissipation (Nielsen 1999).

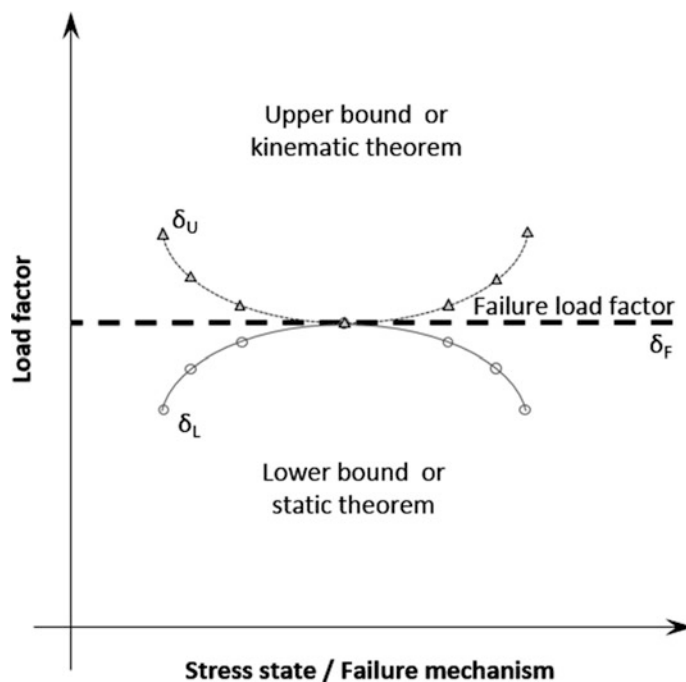
A general load on a structure can be multiplied by a load factor δ and increased from zero to a limit (failure load factor δ_F) for which the structure remains the equilibrium condition (equilibrium between internal and external forces) and the yield condition (the stresses must be less than or equal to the material strength),



Masonry Macro-block Analysis, Fig. 2 Classic limit analysis theory

Masonry Macro-block Analysis,

Fig. 3 Graphical representation of the limit analysis theorems



i.e., the structure remains safe. This approach is associated to the static or lower bound theorem of the classic limit analysis, in which the safety factor is the largest of all the statically admissible load factors (δ_L). Furthermore, the classic limit analysis presents the kinematic approach, in which a sufficient number of releases (e.g., plastic hinges) have to occur to transform the structure into a mechanism. Each kinematically admissible mechanism presents an associated load factor, which is, in general, larger or equal to the safety factor. This approach corresponds to the kinematic or upper bound theorem of the classic limit analysis, in which the safety factor is the smallest of all kinematically admissible load factors (δ_U). The largest load factor obtained from the static theorem and the smallest load factor obtained from the kinematic theorem are the same, stating that a structure can present a statically admissible state and be unsafe if a sufficient number of sections reach the yield surface to transform the structure into a mechanism. This particular case corresponds to the uniqueness theorem of the classic limit analysis. The graphical representation of the theorems of classic

limit analysis is presented in Fig. 3 and is summarized as follows (Kamenjarzh 1996; Nielsen 1999):

- Static or lower bound theorem: If a load presents a magnitude such that the stress state satisfies the equilibrium and yield conditions, then the structure will not collapse and the load factor (δ_L) is less than or equal to failure load factor (δ_F). In this static approach the failure load factor is determined by searching for the maximum load factor.
- Kinematic or upper bound theorem: If a kinematic admissible mechanism can be found for which the work of the external loads exceeds the internal work, then the structure will collapse and the load factor (δ_U) is greater than or equal to the failure load factor (δ_F). In the kinematic approach the failure load factor is determined by searching for the minimum load factor.
- Uniqueness theorem: If the equilibrium, mechanism, and yield conditions are satisfied, then the load factor obtained from the static and kinematic approach is the same and is equal to the failure load factor (δ_F). Thus, the

failure load factor is determined by equating the load factors of both approaches and using, for example, optimization techniques.

Application of Limit Analysis in Masonry Structures

Heyman (1966) was among the first and the most famous promoters to use the limit analysis based on plasticity theory on arches and other masonry structures. Although the masonry is a quasi-brittle material, the plastic limit analysis was used to assess the structural behavior of masonry structures, assuming the following assumptions:

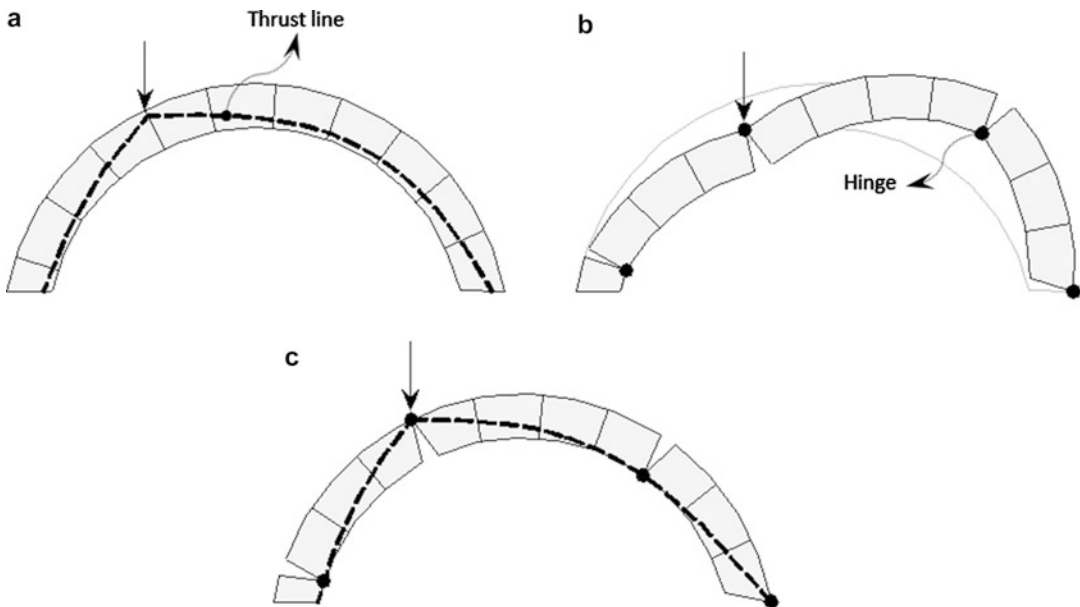
- Masonry has no tensile strength: Masonry presents low tensile strength and quasi-brittle tensile behavior, mainly in ancient masonry structures built with weak mortar, justifying this assumption.
- Masonry has infinite compressive strength: In general, masonry presents a compressive strength higher than the compressive stresses present in the structure. For this reason, the most cases of collapse of masonry buildings

are related to tensile and shear failures. Thus, this assumption is in general accepted and the crushing failure of masonry is not considered.

- Sliding failure is not permitted: Although the sliding between units can sometimes occur at the joints, in general the sliding failure is not the most relevant failure mechanism and this assumption can be accepted.

The failure load assessment of masonry structures based on the limit analysis and the assumptions defined by Heyman (1966) can be carried out using two approaches: (a) static approach and (b) kinematic approach.

In the static approach the resultant of the compressive stresses is plotted at each cross section of the structure. The line containing all the positions of the compressive stress resultants is called by thrust line (Fig. 4a). The thrust line can be determined by analytical methods or graphical methods, such as funicular polygon (Kooharian 1952). The structure is safe when the thrust line is located totally inside of the geometry of structure and the equilibrium with the external loads can be found. In this conditions and according to the lower bound theorem, the applied load is less



Masonry Macro-block Analysis, Fig. 4 Application of the limit analysis to a masonry arch: (a) static approach (lower bound theorem); (b) kinematic approach (upper bound theorem); (c) uniqueness theorem

than or equal to the failure load. Nevertheless, several thrust lines contained in the geometry of the structure can be found, meaning that the thrust line definition is an indeterminate problem (hyperstatic structure), where a solution can be obtained by using an iterative method. Thus, if the thrust line is found to be totally inside of the geometry of structure that was found, the structure is safe, which does not mean that the structure necessarily works according to this solution. However, nothing can be concluded with respect to the stability if the thrust line is outside of the geometry of structure, unless the infinite possibilities were tried.

With regard to kinematic approach, when the structure presents an enough number of hinges to transform it into a mechanism (Fig. 4b), the load factor obtained by equating the work of the external loads to zero corresponds to an upper bound limit (upper bound theorem). The load obtained from the kinematic approach is greater than or equal to the failure load. Different positions for the hinges on the structure can be adopted, and several mechanisms, and respective load factors, can be found. The mechanism that presents the lowest load factor corresponds to the collapse mechanism. Thus, the kinematic approach presents also several solutions.

Taking into account the results obtained from both approaches, the static approach presents a safe solution of the maximum load capacity of masonry structures. However, for complex masonry structures it can be difficult to predict the position of the thrust line. On the other hand, the kinematic approach can lead to an unsafe estimation of the ultimate load. This problem can be solved by implementing computational tools of optimization and searching by the failure load (minimum load factor).

A particular case can occur when the thrust line becomes tangent to the boundaries of the geometry at an enough number of cross sections to create a mechanism (Fig. 4c). In this condition, the mechanism corresponds to the collapse mechanism, the thrust line is unique, and the load corresponds to the failure load (uniqueness theorem).

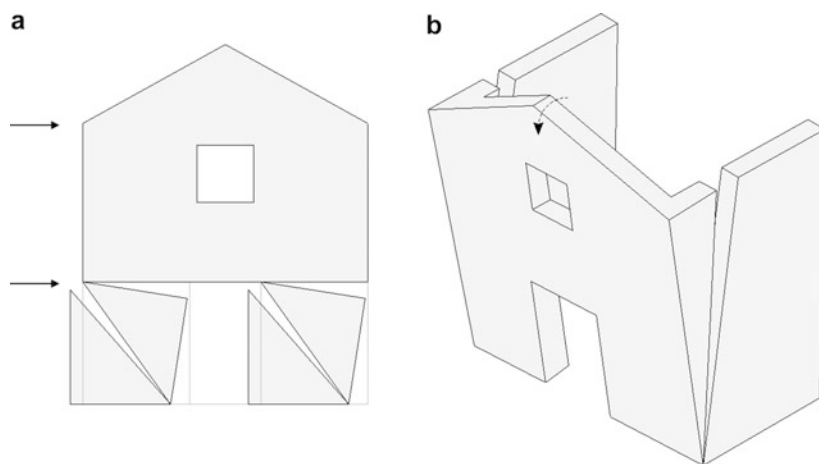
The assumptions adopted for the limit analysis of masonry structures are simplifications of its real and complex structural behavior. In fact,

the compressive strength of masonry is not infinite, and the structures can present high concentrations of compressive stress. The use of finite compressive strength for masonry is more realistic and can provide a small reduction of the load factor at the collapse. Furthermore and as previously mentioned, the sliding failure can occur, which can be a nonconservative assumption for some particular construction techniques of masonry (e.g., sliding at the interface between the rings of a multi-ring masonry bridge). The consideration of the sliding failure implies the lack of the normality condition (nonassociated flow). Thus, several researchers have been conducting efforts to develop limit analysis considering the crushing, the sliding, and the twisting failures of masonry. For more information on innovate limit analyses of masonry structures, see, e.g., Orduña and Lourenço (2003), Gilbert et al. (2006), and Portioli et al. (2013).

Macro-block Approach

The limit analysis with macro-blocks allows to determine the maximum load capacity of structures by using simple procedures based on kinematic approach and involving the equilibrium of the macro-blocks. A macro-block corresponds to a portion of structures with similar material properties and structural behavior, which can represent the structural element (e.g., piers) or a set of structural elements (e.g., a façade). In this approach, the structure is discretized in several macro-blocks with independent structural behaviors. This type of analysis is a practical tool for assessing of the structural behavior of structures and does not require a high number of parameters for the material properties.

The existing masonry structures present in general local modes under an earthquake, due to the loss of equilibrium of parts of the structure. Thus, the limit analysis using macro-blocks is particularly suitable for evaluating the seismic performance of existing masonry buildings, considering even the in-plane or the out-of plane collapse mechanisms (Fig. 5). It should be noted that the out-of-plane behavior of masonry



Masonry Macro-block Analysis, Fig. 5 Examples of collapse mechanisms using macro-blocks: (a) in-plane collapse of façade; (b) out-of-plane collapse of façade

structures is complex and that in general this typology of buildings does not satisfy the assumptions of the typical static nonlinear analysis imposing lateral forces proposed for reinforced concrete buildings (rigid floors, regularity in plan and in elevation, and response governed by the first mode). Furthermore, this type of analysis is also a useful tool for evaluating the efficiency of strengthening techniques applied to existing masonry buildings.

In general, the application of limit analysis with rigid macro-blocks for masonry structures is based on the assumptions proposed by Heyman (1966) (no tension strength, infinite compressive strength, and the sliding cannot occur). However, more realist assumptions can be considered, such as a finite compressive strength.

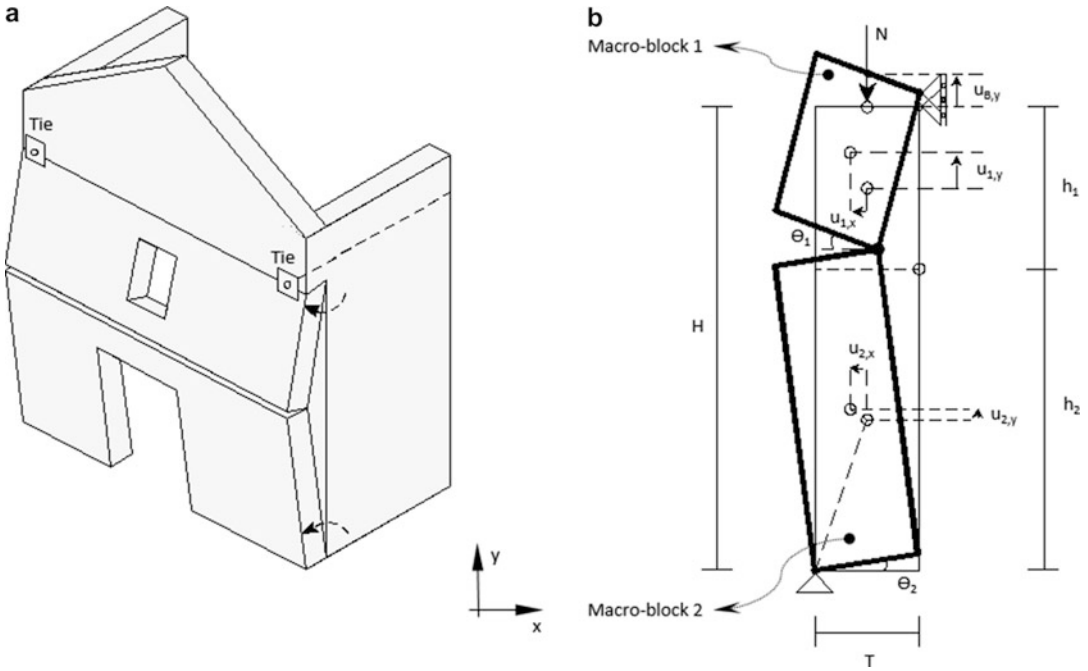
The macro-block approach involves a first step for selection of mechanisms. The mechanisms can be proposed on the basis of the knowledge obtained from the post-earthquake survey of similar buildings, using the crack patterns obtained from the experimental research, and on the basis of the practitioner experience. The selection of the mechanisms to consider is a fundamental step in this type of analysis and requires a detailed knowledge of the features of the structure, such as the quality of the connections between masonry and the connections between masonry walls and floors, the presence

of cracks, and the interaction with adjacent buildings. In fact, a bad evaluation of the possible mechanisms can lead to the non-consideration of the mechanism with the lowest load factor and, consequently, can lead to a failure load higher than the real maximum capacity of the structure. However, the main local collapse mechanisms of masonry building typologies have been compiled in abacus on the form of graphical interpretations schemes, such as the abacus for religious and urban masonry buildings (D'Ayala and Speranza 2002; Lagomarsino and Podesta 2004). Next, the load factor of each considered mechanism is determined by applying the principle of virtual works, and the mechanism that presents the lowest load factor is assumed as the collapse mechanism. Finally, the stability of the structure is analyzed, taking into account the load factor of the collapse mechanism and the requirements defined in the codes (e.g., by comparing the maximum capacity in terms of horizontal acceleration on the structural and the demands in terms of peak ground acceleration (PGA)).

Subsequently, an example of application of the limit analysis with rigid macro-blocks for masonry structures is presented.

Example of Application

Figure 6a presents a possible out-of-plane mechanism of a façade from masonry buildings with



Masonry Macro-block Analysis, Fig. 6 Out-of-plane collapse of tied masonry wall: (a) general view; (b) scheme of the mechanism

ties at the top and subjected to a seismic action with uniform acceleration acting in the orthogonal direction to the plan of the façade. The structure presents three horizontal cracks at the façade and two cracks at the corners. This local mechanism is composed by three hinges, in which the hinge associated to the crack within the wall is not initially known. Figure 6b shows the scheme of the mechanism in analysis. The height H and thickness T of the masonry wall are equal to 3.50 m and 0.30 m, respectively. The specific mass of the masonry γ is equal to 20 kN/m³. Furthermore, it is assumed that the vertical force at the top of the masonry wall N is equal to 10 kN/m.

The first step corresponds to the determination of the variables of rotation (θ) and displacements of the macro-blocks at the center of mass (u) in the vertical and horizontal directions. Since the position of the crack within the wall is not known, the heights of the macro-blocks are defined as a function of the variable X :

$$h_1 = XH = 3.5H \quad (1)$$

$$h_2 = (1 - X)H = (1 - 3.5)H \quad (2)$$

If a rotation $\theta_1 = 1$ is given to macro-block 1, then the macro-block 2 rotation is:

$$\begin{aligned} \theta_1 h_1 &= \theta_2 h_2 \\ \theta_2 &= \frac{h_1}{h_2} \theta_1 \\ \theta_2 &= \left(\frac{X}{1 - X} \right) \theta_1 = \frac{X}{1 - X} \end{aligned} \quad (3)$$

and the displacements of the macro-blocks and of application of the vertical force N are:

$$u_{1,x} = \frac{h_1}{2} \theta_1 = \frac{XH}{2} = \frac{3.5}{2} X \quad (4)$$

$$\begin{aligned} u_{1,y} &= \frac{T}{2} \theta_1 + T \theta_2 = \frac{T}{2} + T \frac{X}{1 - X} \\ &= 0.15 + 0.30 \frac{X}{1 - X} \end{aligned} \quad (5)$$

$$u_{2,x} = \frac{h_2}{2} \theta_2 = \frac{(1-X)}{2} H \frac{X}{(1-X)} = \frac{XH}{2} = \frac{3.5}{2} X \quad (6)$$

$$u_{2,y} = \frac{T}{2} \theta_2 = \frac{T}{2} \frac{X}{(1-X)} = 0.15 \frac{X}{(1-X)} \quad (7)$$

$$u_{B,y} = u_{1,y} = \frac{T}{2} = 0.15 + 0.30 \frac{X}{1-X} \quad (8)$$

The forces applied at the macro-blocks due to the self-weight (P) and the inertial forces proportional to mass (F) are:

$$P = \gamma TH = 21 \text{ kN/m} \quad (9)$$

$$P_1 = PX = 21X \quad (10)$$

$$P_2 = P(1-X) = 21(1-X) \quad (11)$$

$$F = \delta P = 21\delta \quad (12)$$

$$F_1 = \delta P_1 = 21X\delta \quad (13)$$

$$F_2 = \delta P_2 = 21(1-X)\delta \quad (14)$$

where δ is the load factor. P and F are forces by meter of wall width.

The application of the principal of virtual works provides:

$$\begin{aligned} & -P_1 u_{1,y} - P_2 u_{2,y} - N u_{B,y} + F_1 u_{1,x} + F_2 u_{2,x} = 0 \\ & - \left(3.15X + \frac{6.3X^2}{1-X} \right) - (3.15X) - \left(1.5 + \frac{3X}{1-X} \right) \\ & + (36.75X^2\delta) + (36.75X\delta - 36.75X^2\delta) \\ & = 0 - 6.3X - \frac{(3X + 6.3X^2)}{1-X} - 1.5 + 36.75X\delta = 0 \end{aligned} \quad (15)$$

Thus, it is possible to obtain the load factor as a function of X :

$$\delta = \frac{6.3X + \frac{(3X + 6.3X^2)}{1-X} + 1.5}{36.75X} \quad (16)$$

The minimum load factor δ can be computed by equating the:

$$\frac{d\delta}{dx} = 0 \quad (17)$$

and a value of X equal to 0.287 is obtained. Replacing the value of X in the Eq. 16, the minimum load factor is equal to 0.50, leading to the conclusion that position of the crack within of façade that cause the minimum load factor is equal to 2.50 m with respect to the base of structure ($h_2 = 2.50$ m).

Summary

The limit analysis with macro-blocks is a simple tool for evaluating the maximum load capacity of masonry structures. The possible mechanisms are proposed and then the respective load factors are determined. The mechanism that presents the minimum load factor corresponds to the collapse, and its load factor is assumed as the failure load. Thus, a careful evaluation of the possible mechanisms is needed, aiming at not excluding any important mechanism and predicting correctly the maximum load capacity of the structure.

The macro-block analysis applied to masonry structures is based on the first assumptions used for the classic limit analysis of arches (no tension strength, infinite compressive strength, the sliding cannot occur), which is acceptable for the majority of the masonry buildings. However, several approaches considering more realistic assumptions have been proposed and implemented in computational tools, aiming at improving the capabilities and the applicability of this type of analysis.

As final conclusion, the limit analysis with macro-blocks is a powerful tool for the seismic assessment of masonry structures, including the historical buildings and the out-of-plane behavior. Furthermore, it allows evaluating the efficiency of strengthening techniques.

Cross-References

- [Masonry Structures: Overview](#)
- [Nonlinear Dynamic Seismic Analysis](#)

References

- D'Ayala D, Speranza E (2002) An integrated procedure for the assessment of the seismic vulnerability of historic buildings. In: 12th European conference on earthquake engineering, paper no. 561, London, UK
- Gilbert M, Casapulla C, Ahmed HM (2006) Limit analysis of masonry block structures with non-associative frictional joints using linear programming. *Comput Struct* 84:873–887. doi:10.1016/j.compstruc.2006.02.005
- Heyman J (1966) The stone skeleton. *Int J Solids Struct* 2(2):249–279
- Kamenjarzh J (1996) Limit analysis of solids and structures. CRC Press, Boca Raton
- Kooharian A (1952) Limit analysis of voussoir (segmental) and concrete arches. *J Am Concr Inst* 24(4):317–328
- Lagomarsino S, Podesta S (2004) Seismic vulnerability of ancient churches: I. Damage assessment and emergency planning. *Earthquake Spectra* 20(2):377–394, <http://dx.doi.org/10.1193/1.1737735>
- Nielsen M (1999) Limit analysis and concrete plasticity, 2nd edn. CRC Press, Boca Raton
- Orduña A, Lourenço PB (2003) Cap model for limit analysis and strengthening of masonry structures. *J Struct Eng* 129(10):1367–1375, [http://dx.doi.org/10.1061/\(ASCE\)0733-9445\(2003\)129:10\(1367\)](http://dx.doi.org/10.1061/(ASCE)0733-9445(2003)129:10(1367))
- Portioli F, Casapulla C, Cascini L, D'Aniello M, Landolfo R (2013) Limit analysis by linear programming of 3D masonry structures with associative friction laws and torsion interaction effects. *Arch Appl Mech* 83(10):1415–1438. doi:10.1007/s00419-013-0755-4

Masonry Modeling

Paulo B. Lourenço

Department of Civil Engineering, University of Minho, ISE, Guimarães, Portugal

Synonyms

Computational strategies; Finite element modeling; Numerical modeling; Structural analysis

Introduction

This entry deals with modeling of masonry structures. In general, the approach towards the numerical representation of masonry can focus on the micro-modeling of the individual

components, units and mortar, or the macro-modeling of masonry as a composite (Rots 1991). Much effort is made today in the link between the micro- and macro-modeling approaches using homogenization techniques.

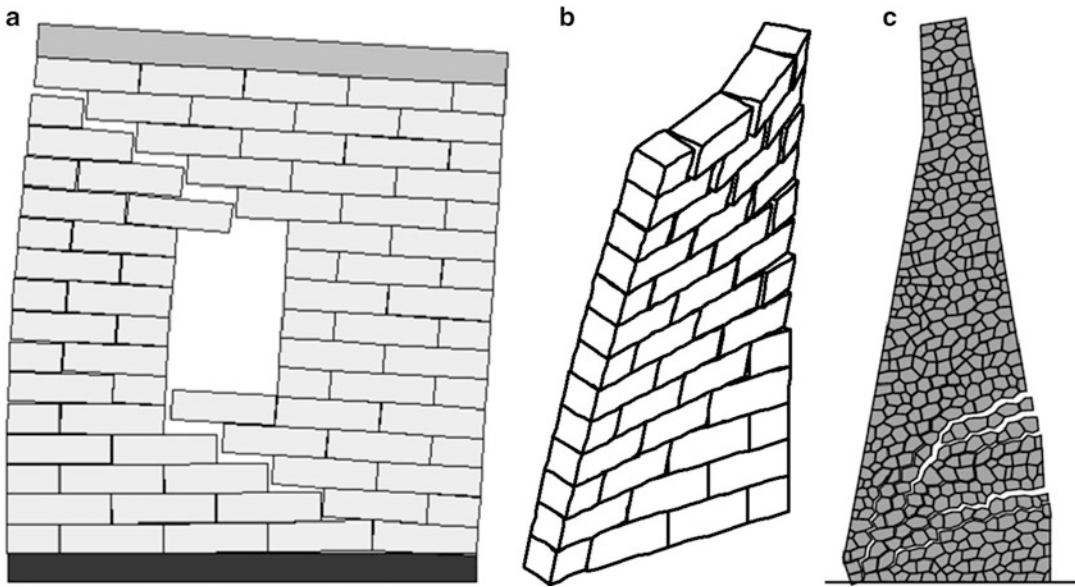
Masonry joints act as planes of weakness, and the explicit representation of the joints and units in a numerical model seems a logical step towards a rigorous analysis tool (Fig. 1). This kind of analysis is particularly adequate for small structures, subjected to states of stress and strain strongly heterogeneous, and demands the knowledge of each of the constituents of masonry (unit and mortar) as well as the interface. In terms of modeling, all the nonlinear behavior can be concentrated in the joints and in straight potential vertical cracks in the centerline of all units. In general, a higher computational effort ensues, so this approach still has a wider application in research and in small models for localized analysis.

The approach based on the use of averaged constitutive equations seems to be the only one suitable to be employed a large-scale finite element analyses. Two different approaches are illustrated in Fig. 2, one collating experimental data at average level and another from homogenization techniques. A major difference is that homogenization techniques provide continuum average results as a mathematical process that includes the information on the microstructure.

These three different approaches, namely, micro-modeling, macro-modeling, and homogenization, are reviewed next.

Micro-Modeling

Different approaches are possible to represent heterogeneous media, namely, the discrete element method, the discontinuous finite element method (FEM), and limit analysis. The finite element method remains the most used tool for numerical analysis in solid mechanics, and an extension from standard continuum finite elements to represent discrete joints was developed in the early days of nonlinear mechanics.



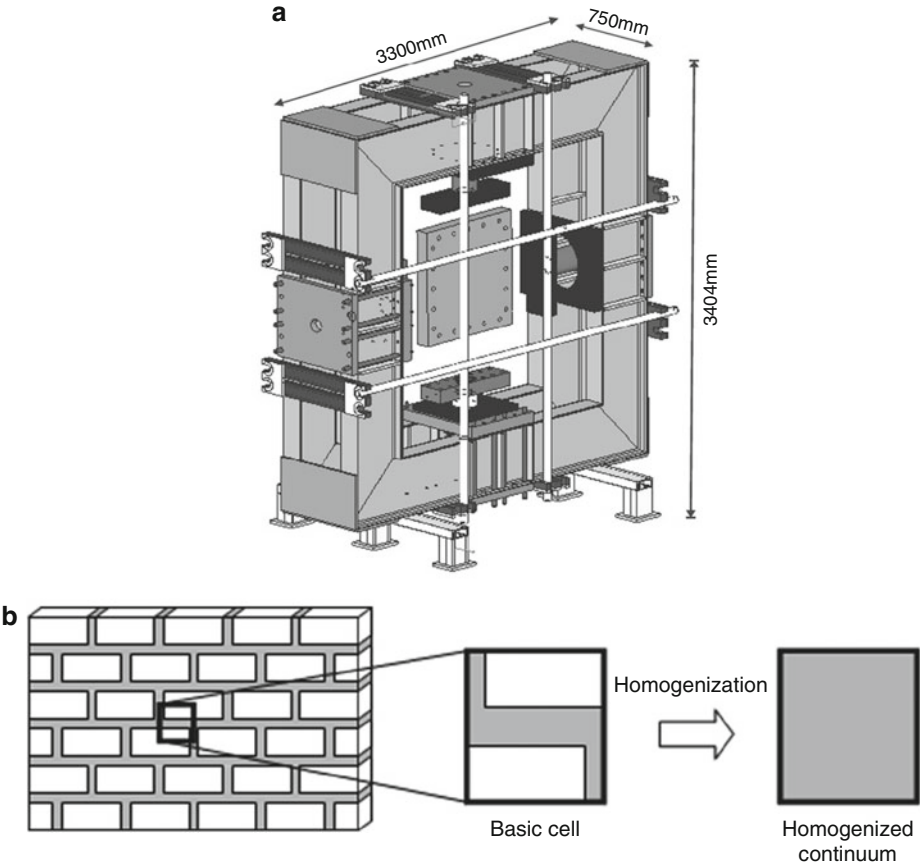
Masonry Modeling, Fig. 1 Examples of cracking at failure using structural micro-modeling: (a) shear wall with opening; (b) wall subjected to out-of-plane loading; (c) retaining wall

Interface elements were initially employed in concrete and rock mechanics, being used since then in a great variety of structural problems. A complete micro-model must include all the failure mechanisms of masonry, namely, cracking of joints, sliding over one head or bed joint, cracking of the units, and crushing of masonry, as done in (Lourenço and Rots 1997) for monotonic loading and in (Oliveira and Lourenço 2004) for cyclic loading.

The typical characteristics of discrete element methods are (a) consideration of rigid or FEM deformable blocks, (b) connection between vertices and sides/faces, (c) interpenetration usually possible, and (d) explicit integration of the equations of motion for the blocks using the real damping coefficient (dynamics) or artificially large (statics). The main advantages are the formulation for large displacements, including contact update, and an independent mesh for each block, in case of deformable blocks. The main disadvantages are the large number of contact points required for accurate representation of interface stresses and a time-consuming analysis, especially for 3D problems. Masonry applications can be found in (Lemos 2007).

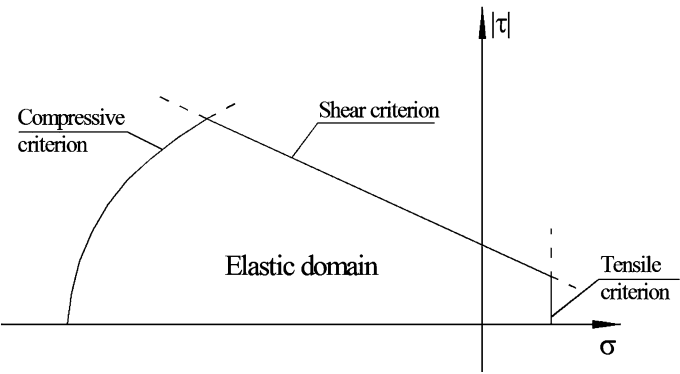
Computational limit analysis received far less attention from the technical and scientific community for blocky structures. Still, limit analysis has the advantage of being a simple tool, while having the disadvantages that only collapse load and collapse mechanism can be obtained and loading history can hardly be included. A limit analysis constitutive model for masonry that incorporates nonassociated flow at the joints, tensile, shear, and compressive failure and a novel formulation for torsion is given (Orduña and Lourenço 2005a, b). The salient characteristics of limit analysis are (a) rigid blocks, (b) interpenetration not allowed, and (c) mathematical formulation that leads to an optimization problem (linear or nonlinear). The main advantages of the technique are adequate formulation for design problems (requires a low number of parameters) and fast analysis. The main advantages are that only the collapse load and mechanism can be obtained, tensile strength cannot be included in the analysis, and the introduction of the loading history remains a challenge.

As an example of the possibilities that can be achieved with micro-modeling, a powerful



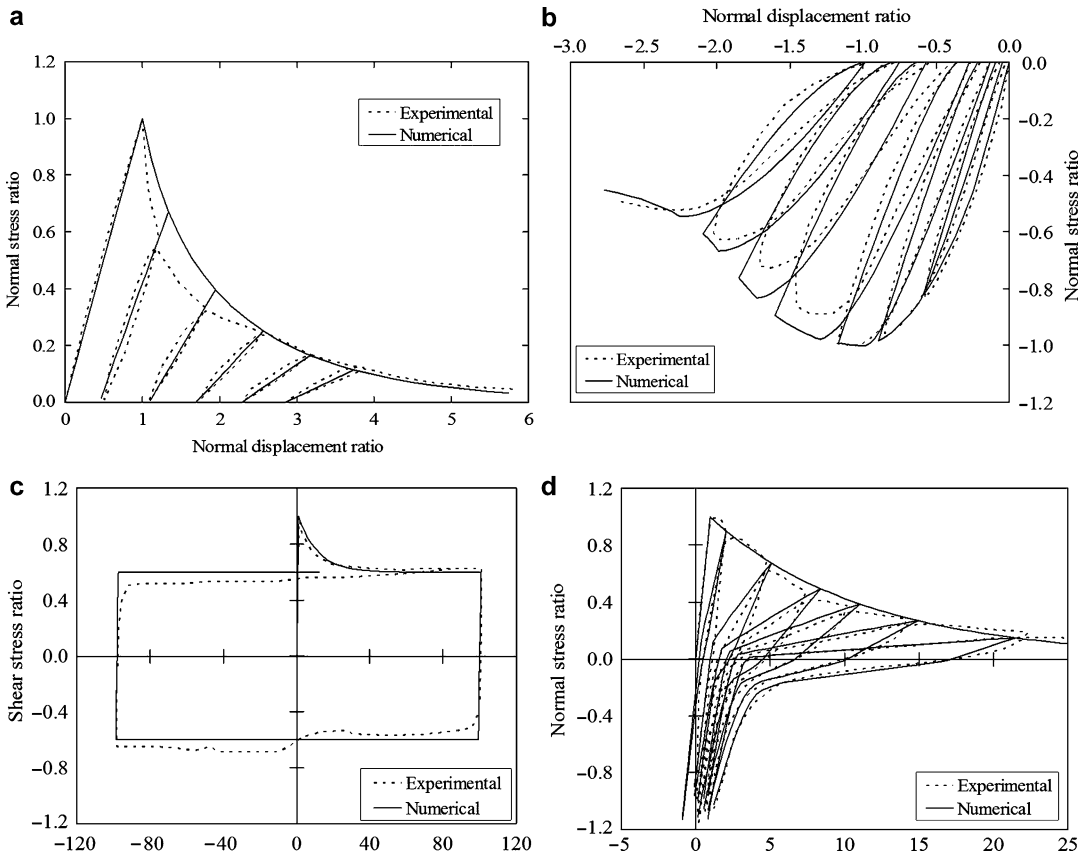
Masonry Modeling, Fig. 2 Constitutive behavior of materials with microstructure: (a) collating experimental data; (b) mathematical process using geometry and mechanics of components

Masonry Modeling, Fig. 3 Multi-surface interface constitutive model



interface model is applied to illustrative examples. A constitutive interface model can be defined by a convex composite yield criterion (see Fig. 3), composed by three individual yield functions, usually with softening included for all

modes so that experimental observations can be replicated. The three functions include the usual Coulomb friction criterion for shear, combined with tension and compression cutoffs. The adoption of appropriate inelastic constitutive relations

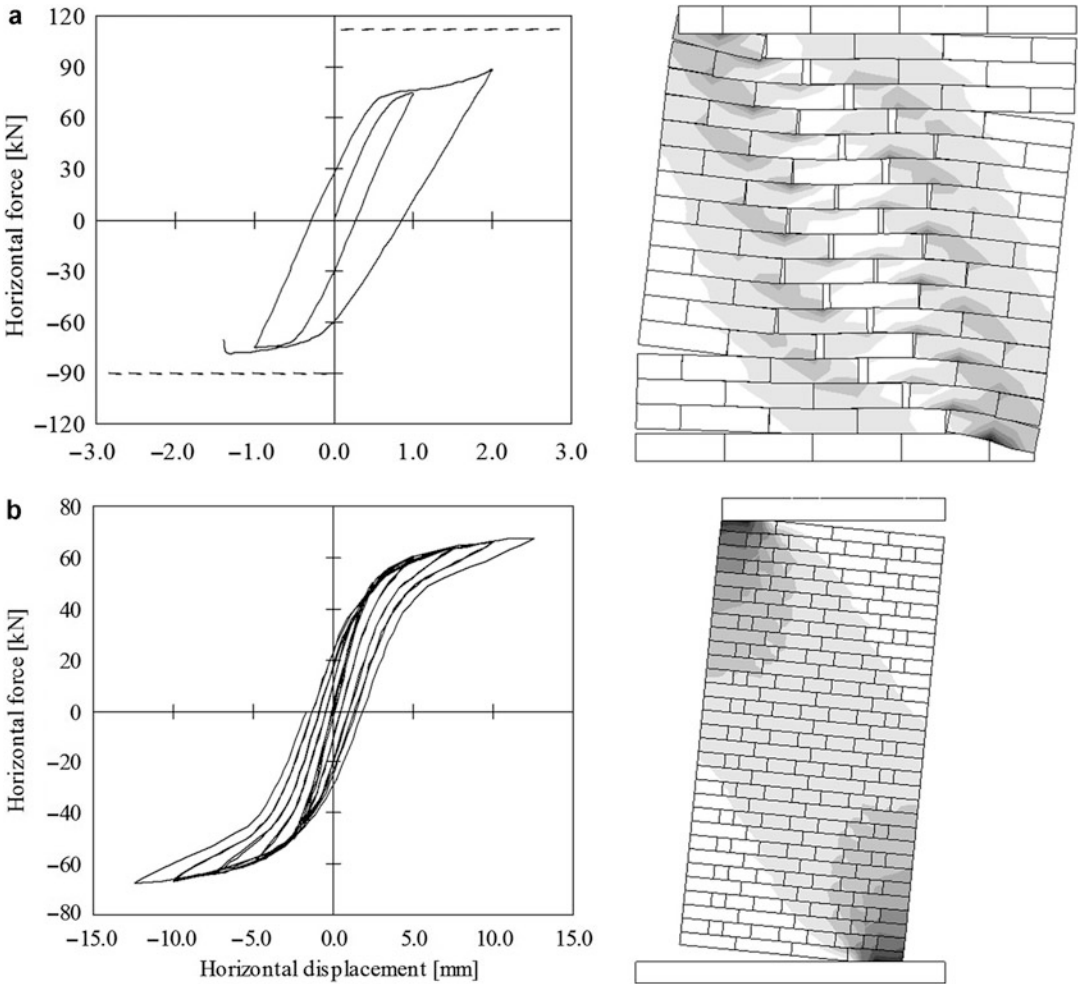


Masonry Modeling, Fig. 4 Comparison between experimental and numerical results for uniaxial testing: (a) tension; (b) compression; (c) shear; (d) tension-compression

and evolution rules makes possible to reproduce nonlinear behavior during loading, unloading, and reloading, consisting of exponential softening for tension and shear, and parabolic hardening followed by exponential softening in compression (see Fig. 4). In this case, isotropic and kinematic hardening laws are used, and aspects related to the algorithm can be found in (Oliveira and Lourenço 2004).

The ability of the model to reproduce the main features of structural masonry elements is shown in Fig. 5. Shear walls have been used traditionally to validate constitutive models and to characterize masonry structural solutions, due to the fact that the most demanding demand for a masonry building with box behavior usually leads to a combination of vertical and horizontal loading.

For the first wall considered, it was found that the geometric asymmetry in the microstructure (arrangement of the units) influenced significantly its structural behavior. Note that, depending in the loading direction, the masonry course starts either with a full unit or only with half unit. Figure 5a shows that the monotonic collapse load is 112.0 kN in the LR direction and 90.8 kN in the RL direction, where L indicates left and R indicates right. The cyclic collapse load is 78.7 kN, which represents a loss of about 13 % with respect to the minimum monotonic value but a loss of about 30 % with respect to the maximum monotonic value. This demonstrates the importance of cyclic loading but also the importance of taking into account the microstructure. It is also clear from these analyses that masonry shear walls with diagonal zigzag cracks



Masonry Modeling, Fig. 5 Results obtained with interface cyclic loading model for shear walls, in terms of force–displacement diagram and failure mode: **(a)** wall failing in shear; **(b)** wall failing in bending

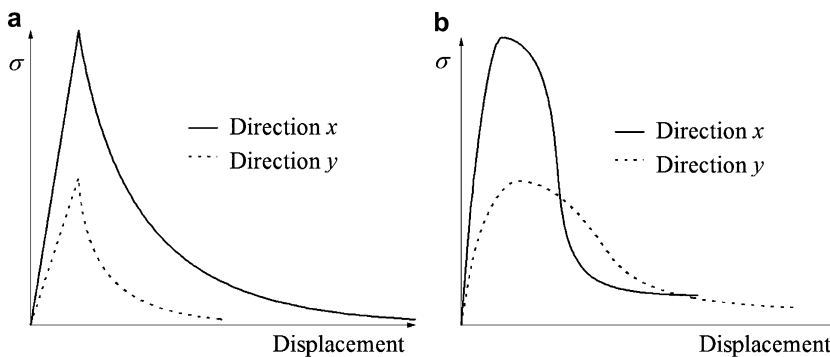
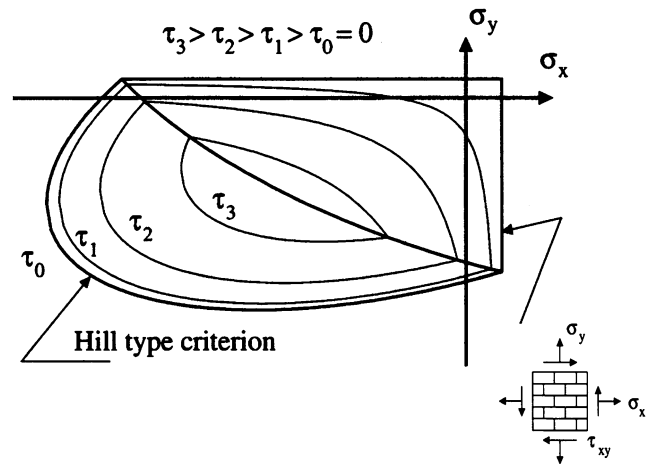
possess an appropriate seismic behavior with respect to energy dissipation.

The second wall considered (Fig. 5b) is a tall wall that simply rocks in both ways. The highly nonlinear shape of the load–displacement curve is essentially due to the opening and subsequent closing, under load reversal, of the top and bottom bed joints. Similar deformed patterns, involving the opening of extreme bed joints, were observed during the experimental test. Numerical results show that the cyclic behavior of the wall is controlled by the opening and closing of the top and bottom bed joints, where damage is mainly concentrated. The model also

shows low (static) energy dissipation, which is a consequence of the activated nonlinear mechanism (opening–closing of joints and the absence of a diagonal shear crack). As shown, the failure of this wall is much different from the previous wall, stressing the relevance of the internal structure.

Macro-Modeling

Difficulties of conceiving and implementing macro-models for the analysis of masonry structures arise especially due to the fact that few

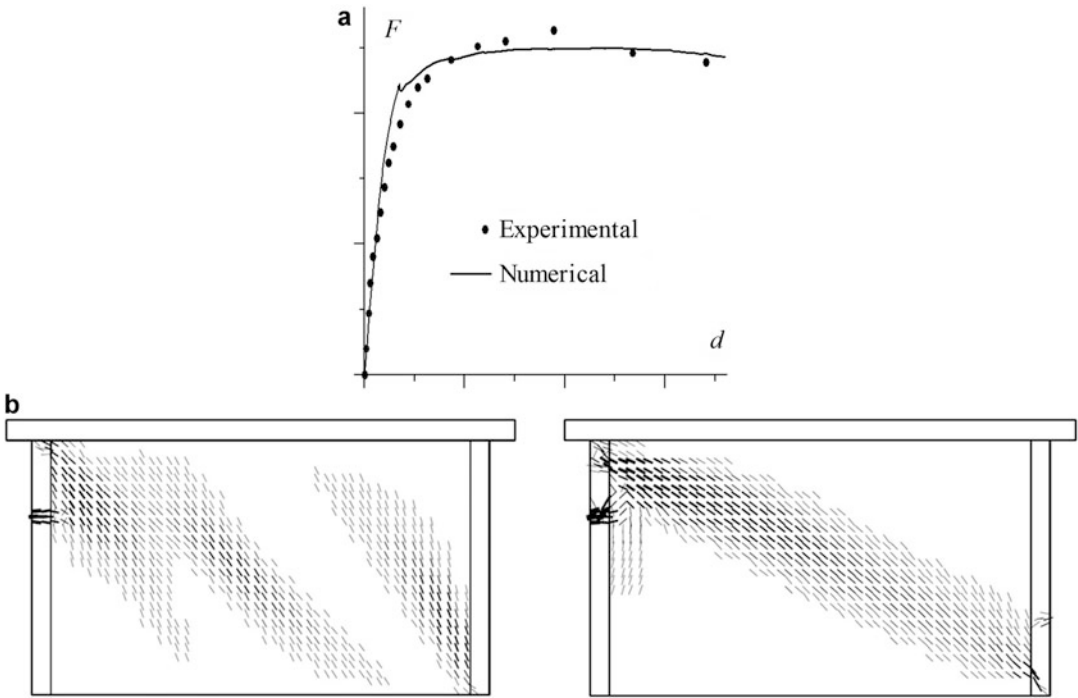
Masonry Modeling,**Fig. 6** Continuum failure surface for masonry (plane stress representation)**Masonry Modeling, Fig. 7** Behavior of the model for (a) tension and (b) compression, along two orthogonal directions

experimental results are available (either for pre- and post-peak behavior), but also due to the intrinsic complexity of formulating anisotropic inelastic behavior. Only a reduced number of authors tried to develop specific models for the analysis of masonry structures (Dhanasekar et al. 1985; Seim 1994; Lourenço et al. 1998; Berto et al. 2002; Pelà et al. 2011), always using the finite element method.

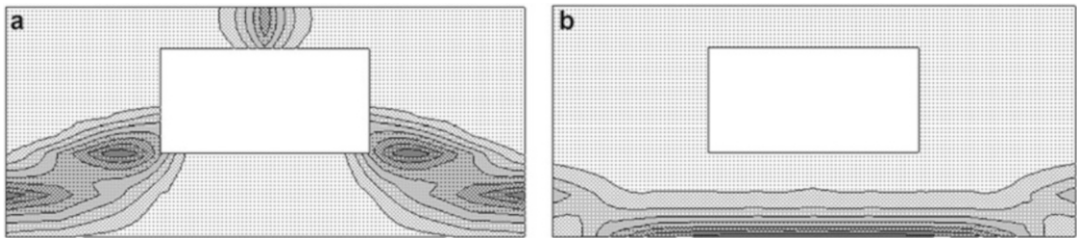
Formulations of isotropic quasi-brittle materials behavior consider, generally, different inelastic criteria for tension and compression. The new model introduced in (Lourenço et al. 1997), extended to accommodate shell masonry behavior (Lourenço 2000), combines the advantages of modern plasticity concepts with a powerful representation of anisotropic material behavior, which includes different

hardening/softening behaviors along each material axis. The model includes the combination of a Rankine-like yield surface for tension and a Hill-like yield surface in compression (see Fig. 6). The behavior of the model in uniaxial tension and uniaxial compression, along two orthogonal directions, is given in Fig. 7. This composite surface permits to reproduce the results obtained in uniaxial tests, in which different behaviors are obtained along different directions. The application of the model in structural modeling of masonry structures leads to excellent results, both in terms of collapse loads and in terms of reproduced behavior (see Lourenço et al. 1998; Lourenço 2000).

Figure 8 shows the results of modeling another shear wall with an initial vertical pre-compression pressure. The horizontal force



Masonry Modeling, Fig. 8 Results for masonry shear wall: (a) load-displacement diagram; (b) predicted cracking pattern at peak and ultimate load



Masonry Modeling, Fig. 9 Results for a panel subjected to uniform out-of-plane loading: predicted cracking pattern at (a) bottom and (b) top face of the panel

F drives the wall to failure and produces a horizontal displacement d at top. The wall is confined by two concrete slabs (top and bottom) and two masonry flanges (left and right). This confinement and the large size of the wall make it appropriate for continuum modeling. Initially, cracking occurs well distributed in the panel and finally concentrates in a single shear band from one corner of the panel to the other. The compressive stresses are well below the crushing strength of masonry, i.e., failure is dominated by tension. A complete discussion of the numerical

results has been given in (Lourenço et al. 1998). Figure 9 shows the results of modeling a panel with out-of-plane pressure. The panel is simply supported on two sides (left and right), fully clamped on one side (bottom), and free on the other (top). The central opening simulates a window, and the panel was loaded with an air bag with a uniformly distributed load. The predicted form of collapse includes diagonal cracks from each lower corner of the panel up to the opening, which were also observed in the experiments. This form of yield line collapse

does not mean that yield line design is safe due to the quasi-brittle behavior of the material. A complete discussion of the numerical results has been given in Lourenço (2000). A discussion on the application of these models to large historic constructions is given in Lourenço (2002).

Homogenization Approaches

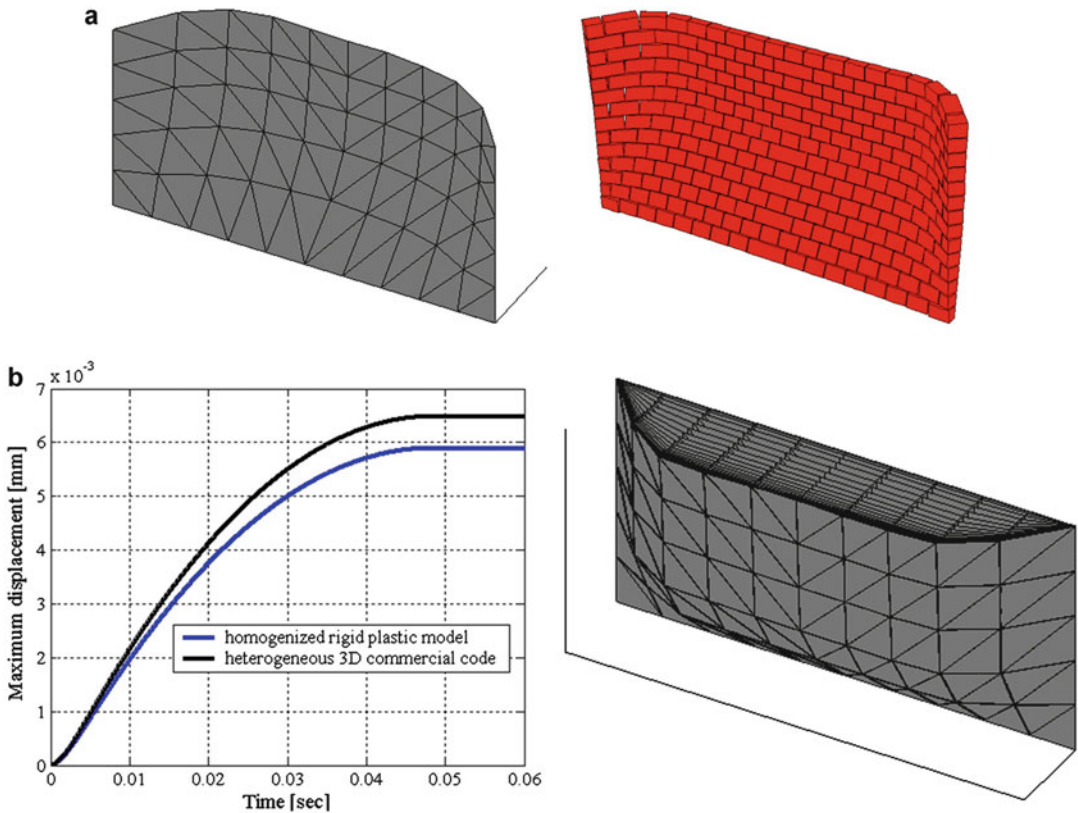
Homogenization techniques permit the establishment of constitutive relations in terms of average stresses and strains from the geometry and constitutive relations of the individual components. This can represent a step forward in masonry modeling because of the possibility to use standard isotropic material models and data for masonry components, instead of the rather expensive approach of testing large masonry specimens under homogeneous loading conditions.

The complex geometry of the masonry representative volume, i.e., the geometrical pattern that repeats periodically in space, means that no closed-form solution of the problems exists for running bond masonry. One of the first ideas presented was to substitute the complex geometry of the basic cell with a simplified geometry, so that a closed-form solution for the homogenization problem was possible. This approach, rooted in geotechnical engineering applications, assumed masonry as a layered material (Lourenço 1996). This simplification does not allow including information on the arrangement of the masonry units and provides significant errors in the case of nonlinear analysis. Moreover, the results depend on the sequence of homogenization steps. To overcome the limitation, micro-mechanical homogenization approaches that consider additional internal deformation mechanisms have been derived (Zucchini and Lourenço 2002, 2009). The implementation of these approaches in standard macroscopic finite element nonlinear codes is relatively simple, and the approaches can compete with macroscopic approaches (see Lourenço et al. (2007) for a detailed review on homogenization approaches).

A powerful micro-mechanical model for the limit analysis for masonry has been recently developed (Milani et al. 2006a, b). In this model, the elementary cell is subdivided along its thickness in several layers. For each layer, fully equilibrated stress fields are assumed, adopting polynomial expressions for the stress tensor components in a finite number of sub-domains. The continuity of the stress vector on the interfaces between adjacent sub-domains and suitable anti-periodicity conditions on the boundary surface are further imposed. In this way, linearized homogenized surfaces in six dimensions for masonry in- and out-of-plane loaded are obtained. Such surfaces are then implemented in a finite element limit analysis code for simulation of 3D structures and including, as recent advances, blast analysis, quasiperiodic masonry internal structure, and FRP strengthening. Figure 10 shows a masonry wall constituted by a periodic arrangement of bricks and mortar arranged in running bond. For a general rigid-plastic heterogeneous material, homogenization techniques combined with limit analysis can be applied for the evaluation of the homogenized in- and out-of-plane strength domain.

The homogenized failure surface obtained has been coupled with finite element limit analysis. Both upper- and lower-bound approaches have been developed, with the aim to provide a complete set of numerical data for the design and/or the structural assessment of complex structures. The finite element lower-bound analysis is based on an equilibrated triangular element, while the upper bound is based on a triangular element with discontinuities of the velocity field in the interfaces. Recent developments include the extension of the model to blast analysis (Milani et al. 2009), to quasiperiodic masonry (Milani et al. 2010), and to FRP strengthening (Milani and Lourenço 2013).

An enclosure running bond masonry wall subjected to a distributed blast pressure is considered first (Milani et al. 2010). The wall is supposed simply supported at the base and on vertical edges and free on top due to the typical imperfect connection between infill wall and RC



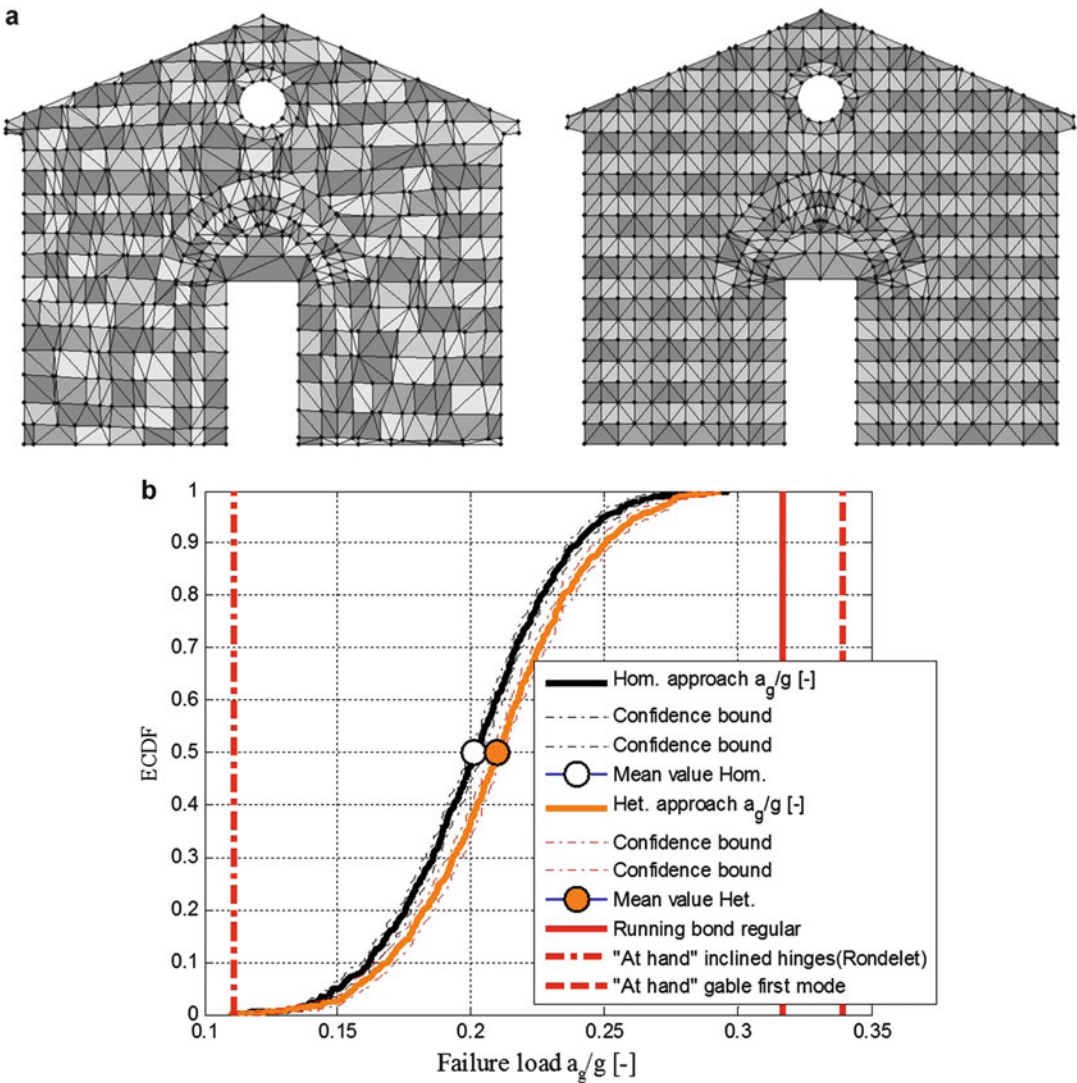
Masonry Modeling, Fig. 11 Masonry infill wall subjected to blast pressure: comparison among deformed shapes at $t = 400 \mu\text{s}$ for (a) homogenized limit analysis

and heterogeneous 3D elastic-plastic FE approach; (b) comparison between maximum out of plane displacements and limit analysis failure mode

beam. A full 3D FE heterogeneous elastic-plastic dynamic analysis has been also conducted, in order to have a deep insight into the problem and to collect alternative data to compare with. For the 3D model, a rigid infinitely resistant behavior for bricks was assumed, whereas for joints, a Mohr-Coulomb failure criterion with the same tensile strength and friction angle used in the homogenized approach for joints was adopted. Eight-noded brick elements were utilized both for joints and bricks, with a double row of elements along wall thickness. A comparison between the deformed shapes at $t = 400 \mu\text{s}$ obtained with the present model and the commercial software is schematically depicted in Fig. 11a. As it is possible to notice, the models give almost the same response in terms of deformed shape for the particular instant

time inspected, confirming that reliable results may be obtained with the model proposed. On the other hand, it is worth underlining that the homogenized rigid-plastic model required only 101 s to be performed on a standard PC Intel Celeron 1.40 GHz equipped with 1Gb RAM, a processing time around 10–3 lower than the 3D case. Comparisons of time-maximum displacement diagrams provided by the two models analyzed are reported in Fig. 11b, together with the evolution of the deformation provided by the homogenized model proposed.

Recently, two different classes of problems have been investigated (Milani et al. 2010), the first consisting of full stochastic representative element of volume (REV) assemblages without horizontal and vertical alignment of joints and the second assuming the presence of a horizontal



Masonry Modeling, Fig. 12 Church of Gondar (Portugal), FE discretization adopted: (a) heterogeneous random mesh vs. mesh for running bond regular heterogeneous and homogenized random analysis; (b) ECDF of

the failure load provide through a direct heterogeneous approach and homogenized limit analysis simulations

alignment along bed joints, i.e., allowing block height variability only row by row. The model is characterized by a few material parameters, and it is therefore particularly suited to perform large-scale Monte Carlo simulations. Masonry strength domains are obtained equating the power dissipated in the heterogeneous model with the power dissipated by a fictitious homogeneous

macroscopic plate. A stochastic estimation of out-of-plane masonry strength domains (both bending moments and torsion are considered) accounting for the geometrical statistical variability of blocks dimensions is obtained with the proposed model. The case of deterministic block height (quasiperiodic texture) can be obtained as a subclass of this latter case. As an important

benchmark, the case in which joints obey a Mohr–Coulomb failure criterion is also tested and compared with results obtained assuming a more complex interfacial behavior for mortar. In order to show the capabilities of the approach proposed when dealing with large-scale structures, the ultimate behavior prediction of a Romanesque masonry church facade located in Portugal are presented. Comparisons with finite element heterogeneous approaches and “at hand” calculations show that reliable predictions of the load bearing capacity of real large-scale structures may be obtained with a very limited computational effort (see Fig. 12).

Conclusions

Constraints to be considered in the use of advanced modeling are the cost, the need of an experienced user/engineer, the level of accuracy required, the availability of input data, the need for validation, and the use of the results. As a rule, advanced modeling is a necessary means for understanding the behavior and damage of masonry constructions. For the study of isolated structural elements and problems controlled by the geometry of a few units, micro-modeling strategies are available. For large-scale applications, average continuum mechanics is usually adopted, and homogenization techniques represent a popular and active field in masonry research. The assessment and design of unreinforced masonry structures subjected to seismic loading is particularly challenging when masonry box behavior is not present. In this case, macro-block analysis seems adequate for simplified assessment and design of strengthening measures.

Summary

Masonry joints act as planes of weakness, and the explicit representation of the joints and units in a numerical model seems a logical step towards a rigorous analysis tool (Fig. 1). This kind of

analysis is particularly adequate for small structures, subjected to states of stress and strain strongly heterogeneous. In general, a higher computational effort ensues, so this approach still has a wider application in research and in small models for localized analysis. The approach based on the use of averaged constitutive equations seems to be the most suitable to be employed a large-scale finite element analyses, either by collating experimental data at average level (macro-modeling) or from homogenization techniques.

A micro-model for masonry joints is usually composed by three individual yield functions, consisting of the usual Coulomb friction criterion for shear, combined with tension and compression cutoffs. A macro-model for the masonry composite must permit to reproduce the results obtained in uniaxial tests, with distinct behavior obtained along different directions. Finally, homogenization techniques require as input the geometry of the units and joints, as well as their mechanical properties. The most successful applications are based on micromechanical models that perform a simplified single-step homogenization because a close form solution is not available.

This entry provides a review of different modeling approaches for masonry structures with illustrative applications.

References

- Berto L, Saelens A, Scotta R, Vitaliani R (2002) Orthotropic damage model for masonry structures. *Int J Numer Methods Eng* 55(2):127–157
- Dhanasekar M, Page AW, Kleeman PW (1985) The failure of brick masonry under biaxial stresses. *Proc Inst Civ Eng, Part 2* 79:295–313
- Lemos JV (2007) Discrete element modeling of masonry structures. *Int J Archit Herit* 1(2):190–213
- Lourenço PB (1996) A matrix formulation for the elastoplastic homogenisation of layered materials. *Mech Cohes-Frict Mater* 1:273–294
- Lourenço PB (2000) Anisotropic softening model for masonry plates and shells. *J Struct Eng* 126(9):1008–1016
- Lourenço PB (2002) Computations of historical masonry constructions. *Prog Struct Eng Mater* 4(3):301–319

- Lourenço PB, Rots JG (1997) Multisurface interface model for the analysis of masonry structures. *J Eng Mech ASCE* 123(7):660–668
- Lourenço PB, de Borst R, Rots JG (1997) A plane stress softening plasticity model for orthotropic materials. *Int J Numer Methods Eng* 40:4033–4057
- Lourenço PB, Rots JG, Blaauwendraad J (1998) Continuum model for masonry: parameter estimation and validation. *J Struct Eng* 124(6):642–652
- Lourenço PB, Milani G, Tralli A, Zucchini A (2007) Analysis of masonry structures: review of and recent trends of homogenization techniques. *Can J Civ Eng* 34(11):1443–1457
- Milani G, Lourenço PB (2013) Simple homogenized model for the non-linear analysis of FRP strengthened masonry structures. Part I: theory. *J Comput Const ASCE* 139(1):59–76
- Milani G, Lourenço PB, Tralli A (2006a) Homogenised limit analysis of masonry walls. Part I: failure surfaces. *Comput Struct* 84(3–4):166–180
- Milani G, Lourenço PB, Tralli A (2006b) A homogenization approach for the limit analysis of out-of-plane loaded masonry walls. *J Struct Eng ASCE* 132(10):1650–1663
- Milani G, Lourenço PB, Tralli A (2009) Homogenized rigid-plastic model for masonry walls subjected to impact. *Int J Solids Struct* 46(22–23):4133–4149
- Milani G, Lourenço PB, Tralli A (2010) A simplified homogenized limit analysis model for randomly assembled blocks out-of-plane loaded. *Comput Struct* 88(11–12):690–717
- Oliveira DV, Lourenço PB (2004) Implementation and validation of a constitutive model for the cyclic behaviour of interface elements. *Comput Struct* 82(17–19):1451–1461
- Orduña A, Lourenço PB (2005a) Three-dimensional limit analysis of rigid blocks assemblages. Part I: torsion failure on frictional joints and limit analysis formulation. *Int J Solids Struct* 42(18–19):5140–5160
- Orduña A, Lourenço PB (2005b) Three-dimensional limit analysis of rigid blocks assemblages. Part II: load-path following solution procedure and validation. *Int J Solids Struct* 42(18–19):5161–5180
- Pelà L, Cervera M, Roca P (2011) Continuum damage model for orthotropic materials: application to masonry. *Comput Methods Appl Mech Eng* 200(9–12):917–930
- Rots JG (1991) Numerical simulation of cracking in structural masonry. *Heron* 36(2):49–63
- Seim W (1994) Numerical modeling of the failure of biaxially loaded masonry walls with consideration of anisotropy (in German). PhD thesis, University of Karlsruhe
- Zucchini A, Lourenço PB (2002) A micromechanical model for the homogenisation of masonry. *Int J Solids Struct* 39(12):3233–3255
- Zucchini A, Lourenço PB (2009) Validation of a micro-mechanical homogenisation model: application to shear walls. *Int J Solids Struct* 46(3–4):871–886

Masonry Structures: Overview

Paulo B. Lourenço

Department of Civil Engineering, University of Minho, ISE, Guimarães, Portugal

Synonyms

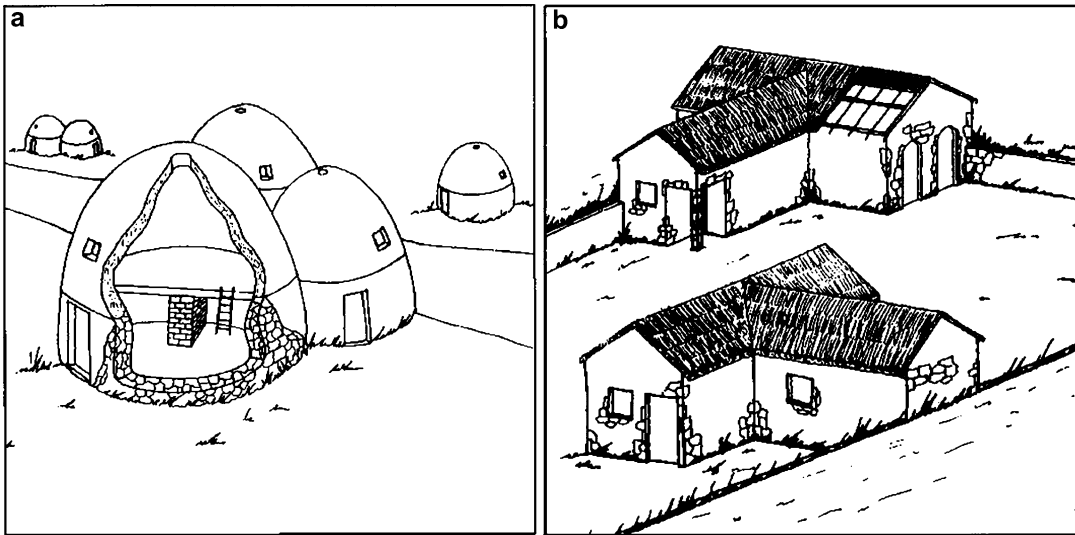
Brick; Building; Concrete; Housing; Mortar; Stone; Structural analysis

Introduction

Masonry is the oldest building material that still finds wide use in today's building industries. The most important characteristic of masonry construction is its simplicity. Laying pieces of stone, bricks, or blocks on top of each other, either with or without cohesion via mortar, is a simple, though adequate, technique that has been successfully used ever since remote ages. Naturally, innumerable variations of masonry materials, techniques, and applications occurred during the course of time. The influence factors were mainly the local culture and wealth, the knowledge of materials and tools, the availability of material, and architectural reasons.

The primitive savage endeavors of mankind to secure protection against the elements and from attack included seeking shelter in rock caves, learning how to build tents of bark, skins, turfs, or brushwood and huts of wattle and daub. Some of such types crystallized into houses of stone, clay, or timber. The evolution of mankind is thus linked to the history of building materials.

The first masonry material to be used was probably stone. In the ancient Near East, evolution of housing was from huts, to apsidal houses (Fig. 1a), and finally to rectangular houses (Fig. 1b). The earliest examples of the first permanent houses can be found near Lake Hullen, Israel (c. 9000–8000 BC), where dry-stone huts, circular and semisubterranean, were found. Several other legacies survived until present



Masonry Structures: Overview, Fig. 1 Examples of prehistoric architecture of masonry in the ancient Near East: (a) beehive houses from a village in Cyprus (c. 5650 BC); (b) rectangular dwellings from a village in Iraq

as testimonies of ancient and medieval cultures (Fletcher and Cruikshank 1996), often as a stone skeleton (Heyman 1997).

The first assault to the use of structural masonry happened at the middle of the nineteenth century, when cast-iron beams and columns started to be produced. By the end of the century, skyscraper constructions methods had eliminated the necessity of massive ground-level piers of masonry. Nevertheless, the collapse of masonry as a structural material started in the beginning of the twentieth century, with the introduction of German, French, and British regulations for design of reinforced concrete structures. Concrete was used in constructions of walls as early as the fourth-century BC around Rome. But, only in 1854, a system for reinforced concrete was patented in Britain by W.B. Wilkinson. By the beginning of the twentieth century, it was clear that reinforced concrete was a durable, strong, moldable, and inexpensive material, and masonry was practically forgotten as a structural material in several developed countries.

In Europe, the building solutions using unreinforced structural masonry represent about 15 % to more than 50 % of the new housing construction, taking as reference countries with low seismicity (e.g., Germany, Netherlands, or

Norway) but also countries with high seismicity (e.g., Italy). A usual solution is the adoption of masonry units with large thickness in the building envelope to fulfill thermal requirements. It is stressed that an integrated and complete building technology is needed, including units with different shapes and solutions for floors. Reinforced masonry was developed in different countries as a response to the lower performance of unreinforced masonry buildings under large horizontal loading, but no unified solution was found. Diverse solutions with different levels of success coexist, together with recent innovative solutions.

It is common practice to combine prefabricated slabs with load-resisting walls so that formwork, scaffolding, and execution times can be significantly reduced. In the USA, in the last 30–40 years, reinforced masonry became an attractive and efficient solution from a perspective of cost-benefit analysis for buildings in regions of low to high seismicity, e.g., hotels, residential buildings, office buildings, schools, commercial buildings, or warehouses. The American standard solution includes reinforced concrete horizontal bond beams, two-cell blocks filled with grout, and vertical reinforcement. Besides other reinforced masonry

approaches, confined masonry is popular in developing countries, as the changes with respect to unreinforced masonry construction are small. In this system vertical and horizontal reinforced concrete elements of small section are included in the masonry. These elements aim at providing an increase of shear and flexural strength, together with a larger energy dissipation capacity and larger ductility with respect to horizontal actions.

The last decades have witnessed an enormous development in the characterization of masonry materials and in numerical methods for structural analysis. Today, with the help of a computer, it is possible to analyze structures with a high level of accuracy. The material science (directly dependent on the observation and analysis of the experimental behavior) has suffered a slower evolution, but important advances in constitutive models and structural component models have occurred in various fields. For a long time, it was believed that the decline of masonry as a structural material was not only due to economic reasons but also to underdeveloped masonry codes and lack of insight in the behavior of this type of structures. This is not presently the case, as modern codes are available, e.g., Eurocode 6 or EN 1996-1-1:2005 (CEN 2005), but also modern engineering books, e.g., Drysdale and Hamid (2008), Hendry (1998), Pfefferman (1999), or Sahlin (1971), which provide general overviews of design. More specific books, with a focus on earthquake performance are also available, e.g., Tassios (1988) or Tomaževič (1999). It is stressed that the design of unreinforced masonry structures under seismic loading has not yet received general consensus.

Description

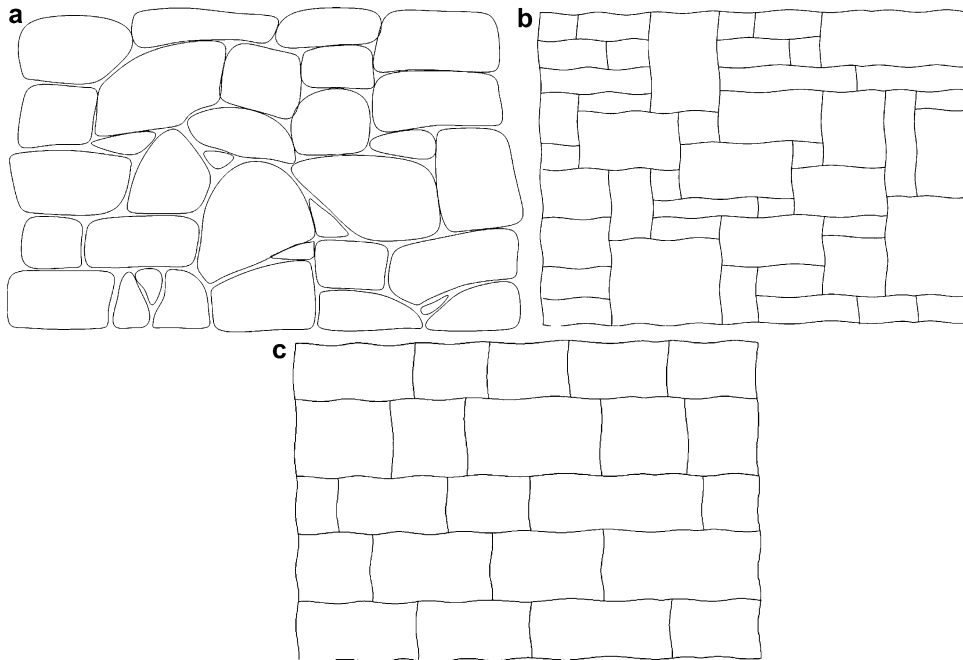
Masonry is a heterogeneous material that consists of units and joints. Units are such as bricks, blocks, ashlar, adobe, irregular stones, and others. Mortar can be clay, bitumen, chalk, lime/cement-based mortar, glue, or others. A (very) simple classification of stone masonry

is shown in Fig. 2. The huge number of possible combinations generated by the geometry, nature, and arrangement of units as well as the characteristics of mortars raises doubts about the accuracy of the term “masonry.” Just for brick masonry, some usual combinations are shown in Fig. 3.

When the walls of ancient constructions were of small width, stone units could be of the full width (bond stone or through stone). If the walls were very thick, ashlar would only be used for the outer leafs and the inside would be filled with irregular stones or rubble, or more than one leaf of masonry would be used. Indeed, physical evidence shows us that ancient masonry is a very complex material with three-dimensional internal arrangement, usually unreinforced, but which can include some form of traditional reinforcement, see Fig. 4. Moreover, these materials are associated with complex structural systems, where the separation between architectural features and structural elements is not always clear.

Modern masonry can also exhibit significant variations, not only of materials but also of building technology, see Fig. 5. The choice of materials and the thermal solution, particularly for the enclosure walls, which is a matter of growing concern, is mostly due to tradition and local availability of the materials. Also, the use of reinforcement is associated with tradition and local technological developments, with different approaches from one country to the other.

Nevertheless, the mechanical behavior of the different types of masonry has generally a common feature: a very low tensile strength. This property is so important that it has determined the shape of ancient constructions. The difficulties in performing advanced testing of ancient structures are quite large due to the innumerable variations of masonry, the variability of the masonry itself in a specific structure, and the impossibility of reproducing it all in a specimen. Therefore, most of the advanced experimental research carried out in the last decade has concentrated in brick/block masonry and its relevance for design. The masonry components are detailed in another essay.



Masonry Structures: Overview, Fig. 2 Different kinds of stone masonry: (a) rubble masonry; (b) ashlar masonry; (c) coursed ashlar masonry

Mechanics

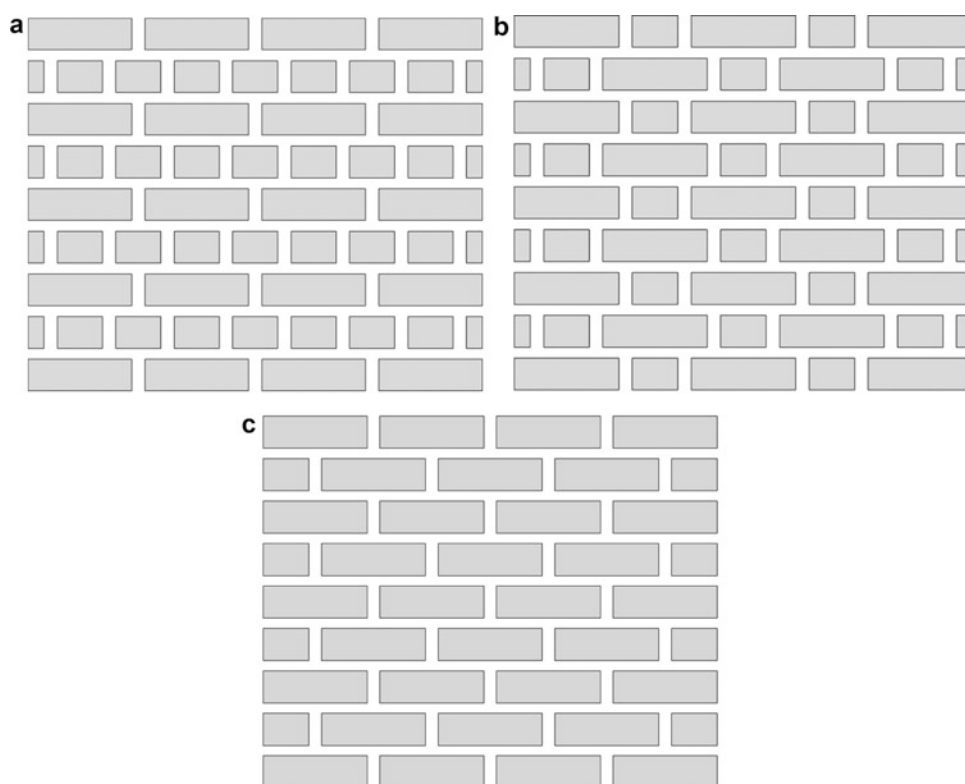
The fact that ancient and modern masonry have so much variability in materials and technology make the task of structural analysis of these structures particularly complex. From a very simplified perspective, it is possible to distinguish masonry as reinforced and unreinforced. The presence of (distributed) reinforcement provides masonry with tensile strength and renders masonry closer to reinforced concrete. In such a case, the orthotropic behavior of masonry and the nonlinear constitutive behavior become less relevant, and the techniques normally used for the design and analysis of reinforced concrete structures can possibly be used. Conversely, in the case of unreinforced masonry structures, the very low tensile strength of the material renders the use of nonlinear constitutive behavior more obvious. This is particularly true in the assessment of existing structures and in seismic analysis.

Masonry is usually described as a material that exhibits distinct directional properties due to the

mortar joints, which act as planes of weakness. This description is associated mostly with the material, whereas a different description can be given at structural level.

Description at material level: In general, the approach toward the numerical representation of masonry, i.e., masonry modeling, can address the micro-modeling of the individual components, viz., unit (brick, block, etc.) and mortar, or the macro-modeling of masonry as a composite (CUR 1997). Depending on the level of accuracy and the simplicity desired, it is possible to use the following modeling strategies (see Fig. 6):

1. Detailed micro-modeling – units and mortar in the joints are represented by continuum elements, whereas the unit-mortar interface is represented by discontinuum elements.
2. Simplified micro-modeling – expanded units are represented by continuum elements, whereas the behavior of the mortar joints and unit-mortar interface is lumped in discontinuum elements.



Masonry Structures: Overview, Fig. 3 Different arrangements for brick masonry: (a) English (or cross) bond; (b) Flemish bond; (c) stretcher bond

M

3. Macro-modeling – units, mortar, and unit-mortar interface are smeared out in a homogeneous continuum.

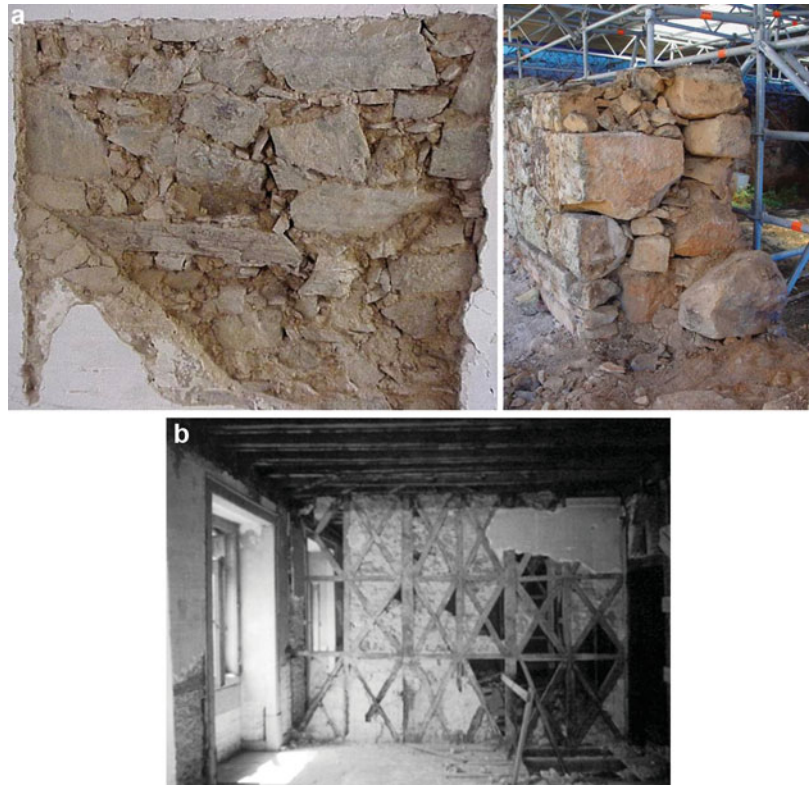
In fact, the term “micro-modeling” is probably not the most adequate and the term “meso-modeling” would be more reasonable, leaving the former designation for approaches at a lower scale. But the terms macro- and micro-modeling are now widely accepted by the masonry community. A major step in the computational representation using modern analysis techniques is provided in Lourenço (1996).

Description at structural level: The simplest approach related to the modeling of masonry buildings is given by the application of different structural elements resorting to, e.g., truss, beam, panel, plate, or shell elements to represent columns, piers, arches, and vaults, with the assumption of homogeneous (macro) material behavior. Fig. 7 illustrates various possibilities.

The lumped approach or mass-spring-dashpot model of Fig. 7a is at best a crude approximation of the actual geometry of the structure, using floor levels and lumped parameters as structural components. The simplicity of the geometric model allows increased complexity on the loading side and in the nonlinear dynamic response. The structural component model in Fig. 7b approximates the actual structural geometry more accurately by using beams and joints as structural components. This approach allows the assessment of the system behavior in more detail. In particular, it is possible to determine the sequential formation of local, predefined failure mechanisms and overall collapse, both statically and dynamically. Finally, the structural model in Fig. 7c approximates the actual structural geometry using macro-blocks with a discrete set of failure lines. Most of these efforts address seismic design and assessment. Models such as the ones shown in Fig. 7a, b are adequate in case that masonry box

Masonry Structures: Overview,

Fig. 4 Examples of different masonry types: (a) irregular stone wall with a complex transverse cross section, from the eighteenth century in Northern Portugal; (b) timber-braced “Pombalino” system emerging after the 1755 earthquake in Lisbon



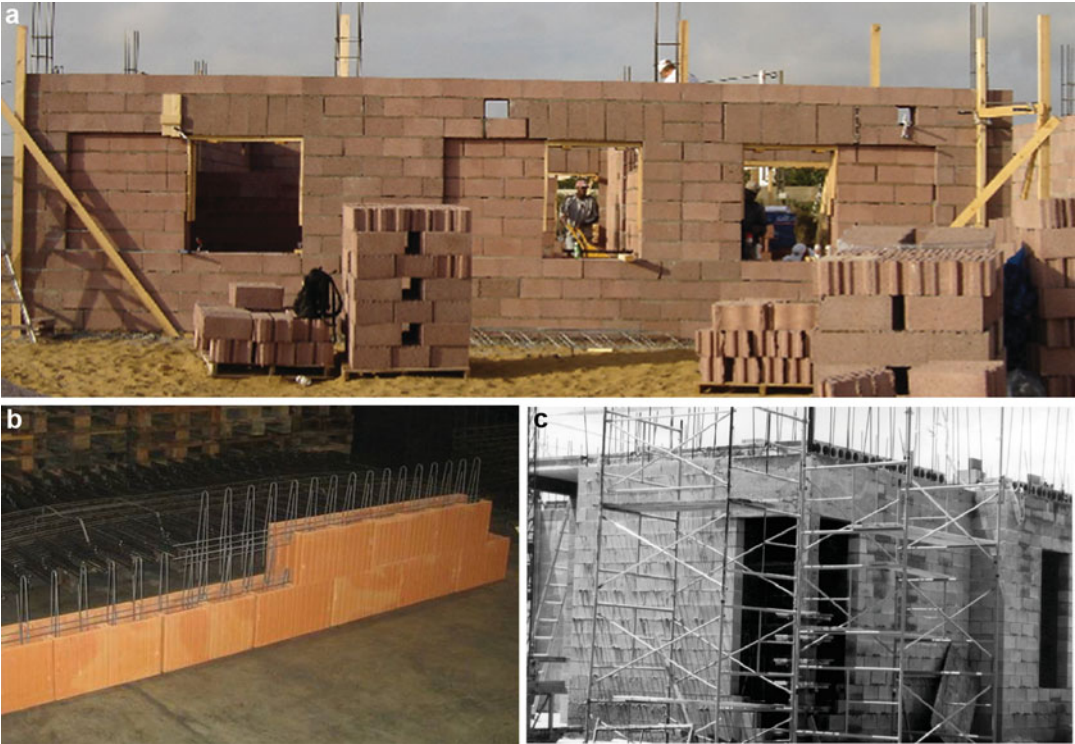
behavior is exhibited by the building, whereas in Fig. 7c, the so-called macro-block analysis is adequate in case that the masonry building does not exhibit box behavior.

Accurate modeling of masonry structures requires a thorough experimental description of the material. Obtaining experimental data, which is reliable and useful for numerical models, has been hindered by the lack of communication between analysts and experimentalists. The use of different testing methods, test parameters, and materials preclude comparisons and conclusions between most experimental results. It is also current practice to report and measure only strength values and to disregard deformation characteristics. Only recently, information in the post-peak or softening regime became more widely available.

Masonry Behavior

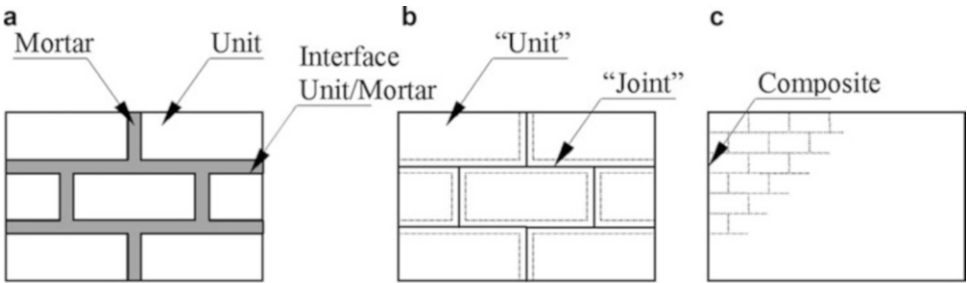
Softening is a gradual decrease of mechanical resistance under a continuous increase of

deformation forced upon a material specimen or structure. It is a salient feature of quasi-brittle materials like clay brick, mortar, ceramics, rock, or concrete, which fail due to a process of progressive internal crack growth. Such mechanical behavior is commonly attributed to the heterogeneity of the material, due to the presence of different phases and material defects, like flaws and voids. Even prior to loading, mortar contains micro-cracks due to the shrinkage during curing and the presence of the aggregate. The clay brick contains inclusions and micro-cracks due to the shrinkage during the burning process. Stone also, usually, contains inclusions and micro-cracks. The initial stresses and cracks as well as variations of internal stiffness and strength cause progressive crack growth when the material is subjected to progressive deformation. Initially, the micro-cracks are stable which means that they grow only when the load is increased. Around peak load an acceleration of crack formation takes place and the formation of macro-cracks starts. The macro-cracks are unstable,



Masonry Structures: Overview, Fig. 5 Examples of modern masonry: (a) confined masonry in areas of moderate to high seismicity, with thick blocks; (b, c) different reinforced masonry solutions, adopted in Switzerland (left) and the USA (right)

M

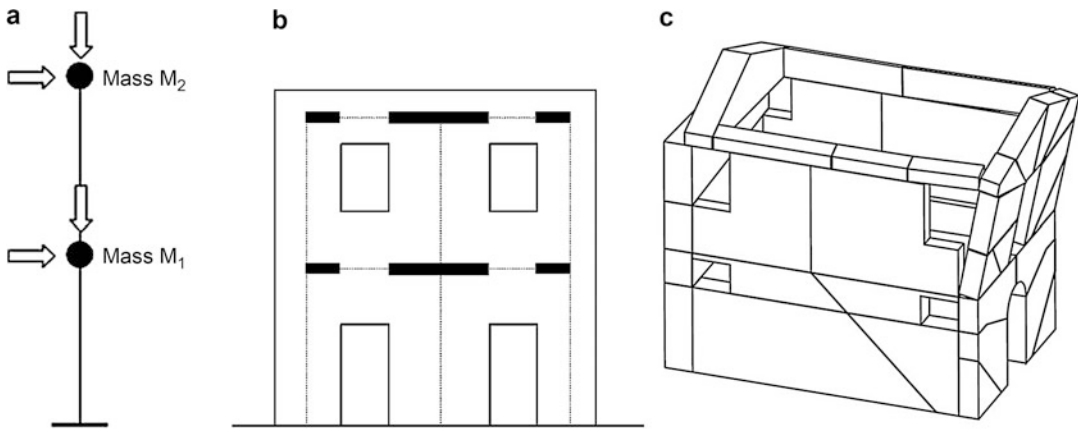


Masonry Structures: Overview, Fig. 6 Modeling strategies for masonry structures: (a) detailed micro-modeling; (b) simplified micro-modeling; (c) macro-modeling

which means that the load has to decrease to avoid an uncontrolled growth. In a deformation-controlled test, the macro-crack growth results in softening and localization of cracking in a small zone while the rest of the specimen unloads.

The properties of masonry are strongly dependent upon the properties of its constituents.

Compressive strength tests are easy to perform and give a good indication of the general quality of the materials used. It is difficult to relate the tensile strength to the compressive strength due to the different shapes, materials, manufacture processes, and volume of perforations in the units. The bond between the unit and mortar is often the weakest link in masonry assemblages.



Masonry Structures: Overview, Fig. 7 Examples of structural component models: (a) lumped parameters for a complete building with 3 degrees of freedom per story;

(b) beam elements for wall with openings; (c) macro-elements for seismic assessment

The nonlinear response of the joints, which is then controlled by the unit-mortar interface, is one of the most relevant features of masonry behavior. Two different phenomena occur in the unit-mortar interface, one associated with tensile failure (mode I) and the other associated with shear failure (mode II).

Different test setups have been used for the characterization of the tensile behavior of the unit-mortar interface. These include (three-point, four-point, bond-wrench) flexural testing, diametral compression (splitting test), and direct tension testing. An important aspect in the determination of the shear response of masonry joints is the ability of the test setup to generate a uniform state of stress in the joints. This objective is difficult because the equilibrium constraints introduce nonuniform normal stresses in the joint.

Finally, the uniaxial behavior of the composite material is anisotropic, i.e., depending on the loading direction with respect to the material axes, namely, the directions parallel and normal to the bed joints. The compressive strength of masonry in the direction normal to the bed joints has been traditionally regarded as the sole relevant structural material property, at least until the recent introduction of numerical methods for masonry structures. A test frequently used to obtain this uniaxial compressive strength is the

stacked bond prism, but it is still somewhat unclear what the consequences are of using this type of specimens in the masonry strength. It has been accepted by the masonry community that the difference in elastic properties of the unit and mortar is the precursor of failure in compression, with uniaxial compression of masonry leading to a state of triaxial compression in the mortar and of compression/biaxial tension in the unit. Uniaxial compression tests in the direction parallel to the bed joints have received substantially less attention from the masonry community. However, regular masonry is an anisotropic material and, particularly in the case of low longitudinal compressive strength of the units due to high or unfavorable perforation, the resistance to compressive loads parallel to the bed joints can have a decisive effect on the load-bearing capacity.

The constitutive behavior of masonry under biaxial states of stress cannot be completely described from the constitutive behavior under uniaxial loading conditions. The influence of the biaxial stress state has been investigated up to peak stress to provide a biaxial strength envelope, which cannot be described solely in terms of principal stresses because masonry is an anisotropic material. Therefore, the biaxial strength envelope of masonry must be described either in terms of the full stress vector in a fixed set of

material axes or in terms of principal stresses and the rotation angle between the principal stresses and the material axes.

Summary

Masonry is a building material with many variations and different levels of success in modern construction worldwide. Masonry is also widely available in the built heritage, as most of the pre-twentieth-century buildings are made of masonry. Masonry consists of units laid with a certain bond, usually including mortar in the joints. Masonry can be typically unreinforced or reinforced and an understanding of this composite material requires the characterization of its components. Because masonry is a composite material, different modeling strategies are available either with the discretization of the bond or not. These approaches are denoted by micro- and macro-modeling, usually combined with advanced nonlinear analysis tools. Adequate seismic design and simple assessment of unreinforced masonry buildings are a challenging task and the applicability of the methods depends on the existence of box behavior for the building. In the absence of box behavior, macro-block analysis is a popular and powerful approach. Finally, it is noted that adequate characterization of the experimental behavior is needed for an effective analysis and design of masonry structures, and this is currently available in the literature.

References

- CEN (2005) EN 1996-1-1:2005, General rules for reinforced and unreinforced masonry structures. Committee for Standardization, European
- CUR (1997) Structural masonry: an experimental/numerical basis for practical design rules (CUR report 171). A.A. Balkema, Rotterdam. ISBN 9789054106807
- Drysdale R, Hamid A (2008) Masonry structures: behavior and design, 3rd edn. Boulder, Colorado. ISBN 1929081332
- Fletcher B, Cruikshank D (1996) Sir Banister Fletcher's history of architecture: centenary edition. Butterworth-Heinemann, Oxford. ISBN 9780750622677

- Hendry AW (1998) Structural masonry. Macmillan Press, Houndmills. ISBN 0333733096
- Heyman J (1997) The stone skeleton: structural engineering of masonry architecture. Cambridge University Press, Cambridge. ISBN 0521629632
- Lourenço PB (1996) Computational strategies for masonry structures. PhD dissertation, Delft University of Technology. ISBN 9040712212
- Pfefferman O (1999) Maçonnerie portante. Kluwer Academic
- Sahlin S (1971) Structural masonry. Prentice Hall, Englewood. ISBN 0138539375
- Tassios TP (1988) Meccanica delle murature. Liguori Editore. ISBN 8820716380
- Tomažević M (1999) Earthquake-resistant design of masonry buildings. Imperial College Press, London. ISBN 1860940668

Mechanisms of Earthquakes in Aegean

Anastasia A. Kiratzi

Department of Geophysics, Aristotle University of Thessaloniki, Thessaloniki, Greece

Synonyms

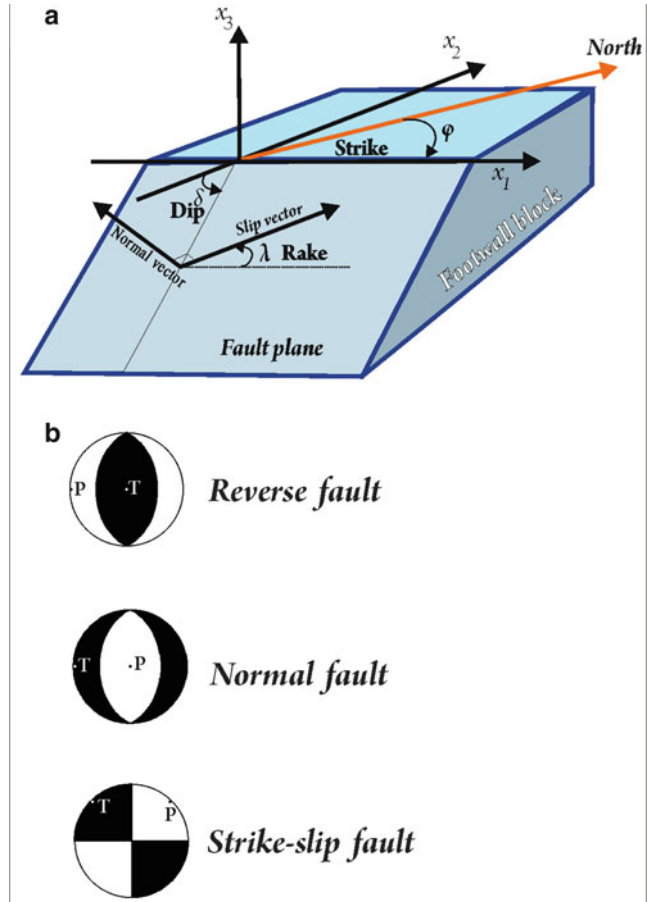
Fault-plane solutions; Focal mechanisms; Moment-tensor solutions

Introduction

Definition of Basic Concepts

The **focal mechanism** illustrates the geometry of faulting during an earthquake. For a fault-related earthquake, the focal mechanism describes the orientation of the causative fault (its strike and dip angle) and the direction of the slip on its plane (rake angle). The pattern of the radiated seismic waves depends on the fault geometry, thus the focal mechanism is determined from seismograms recorded at various distances and azimuths from the seismic source. The simplest method relies on the first-motion polarity of body waves, and the more sophisticated methods are based on waveform modeling of body and surface waves. Fault-plane solutions are described by two

Mechanisms of Earthquakes in Aegean,
Fig. 1 (a) Fault geometry used in earthquake studies, characterized by the strike, dip, and rake angles of the fault plane; (b) beach-ball presentation of focal mechanisms for the three major fault types



mutually orthogonal planes in space, one of which is the fault plane and the other is the auxiliary plane, to which the slip vector is normal. There is an ambiguity in identifying the fault plane on which the slip occurred, from the orthogonal, mathematically equivalent, auxiliary plane. The ambiguity can be resolved using additional information, for example, if the fault reaches the surface, if the aftershocks define a clear pattern which is aligned with one of the two nodal planes, and if directivity effects could be identified.

The fault geometry is shown in Fig. 1a. The fault plane is characterized by its normal vector. The direction of motion is given by the slip vector which is supposed to lie in the fault plane, so it is perpendicular to the normal vector. The slip vector indicates the direction in which the upper block of the fault, e.g., the hanging wall block,

moved with respect to the lower block, e.g., the footwall block. In the coordinate system adopted (Fig. 1a), the parameters used to describe the focal mechanism are: (a) the strike, ϕ , which is the angle between the trace of the fault (the intersection of the fault plane with the horizontal) and North ($0^\circ \leq \phi \leq 360^\circ$) (the strike angle is measured in such a manner so that the fault plane dips to the right-hand side); (b) the dip, δ , which is the angle between the fault plane and the horizontal at a right angle to the trace ($0^\circ \leq \delta \leq 90^\circ$); and (c) the rake, λ , which is the angle between the direction of relative displacement or slip and the horizontal measured on the fault plane ($0^\circ \leq \lambda \leq 360^\circ$).

The three major fault types e.g., reverse dip slip (reverse faults that dip at shallow angles $\sim <45^\circ$ are called thrusts), normal dip slip, and strike-slip are described by their corresponding

focal mechanisms, providing the values of these three angles, e.g., the strike, the dip, and the rake. The focal mechanism solution is typically displayed graphically using the so-called beach-ball diagram (Fig. 1b). The beach ball hints the stress-field orientation at the time of rupture, and the basic fault types can be related to the orientation of the principal stress directions. The tension axis (T) and the pressure axis (P) tend to approach the minimum and the maximum principal compressive stress directions, respectively, of the stress tensor. They coincide in the case the fault surface corresponds to the plane of maximum shear. The principal stress directions can be obtained by the inversion of fault-plane solutions. The T- and P-axes are determined from an eigenvalue and eigenvector analysis of the moment tensor. Since this tensor is symmetric, its eigenvalues are real, and its eigenvectors are mutually orthogonal, forming the set of T, P, and N (null) axes.

The Aegean

The Aegean Sea and the surrounding lands, mainland of Greece, southern Balkans, and western Anatolia, hereafter referred to as the Aegean, exhibit active deformation as manifested by the record of intense earthquake activity (the area hosts more than 60 % of the total European seismicity). Active tectonics apart from forming magnificent landscapes also provide a wealth of data to the geoscientists. To this end, the Aegean, even though it covers a small area in eastern Mediterranean, roughly $700 \text{ km} \times 700 \text{ km}$, is listed among the best studied regions of continental deformation and is often called a *natural geophysical laboratory*.

The major kinematic boundary conditions within the eastern Mediterranean are sketched in Fig. 2 and outline the following:

- The northward movement of the African (Nubia) plate relative to the Eurasian plate



Mechanisms of Earthquakes in Aegean, Fig. 2 A sketch of the major kinematic boundaries within eastern Mediterranean, with the rates of motion relative to

Eurasia. CTF Cephalonia transform fault zone, AFZ Achaia fault zone, PTF Pafos transform fault zone (Figure modified from Papazachos et al. 1998)

(convergence rate between 5 and 10 mm/year) and the subduction of its frontal part beneath the overriding Aegean lithosphere. The subduction rate exceeds the convergence rate, because of the rapid SW motion of the southern Aegean relative to Eurasia which leads to a subduction rate of ~ 35 mm/year; in this setting, the overriding Aegean plate must extend to compensate for the difference between the subduction rate and the convergence rate.

- (b) The westward motion of Anatolia relative to Eurasia, toward the Aegean, driven by the N-S convergence (at a rate of 30–40 mm/year), between the Arabian plate and Eurasia. This lateral extrusion of Anatolia, as a coherent rigid block, is facilitated by the operation of two strike-slip faults: the right-lateral North Anatolian Fault (NAF), with a slip rate ~ 24 mm/year and the more diffuse left-lateral East Anatolian Fault (EAF), with a slip rate ~ 9 mm/year (McClusky et al. 2000). The presence of the Hellenic subduction allows the Aegean plate to easily override the retreating subduction boundary.
- (c) The collision of the Adria–Apulia block with Eurasia along the western coastal region of Albania and western Greece.

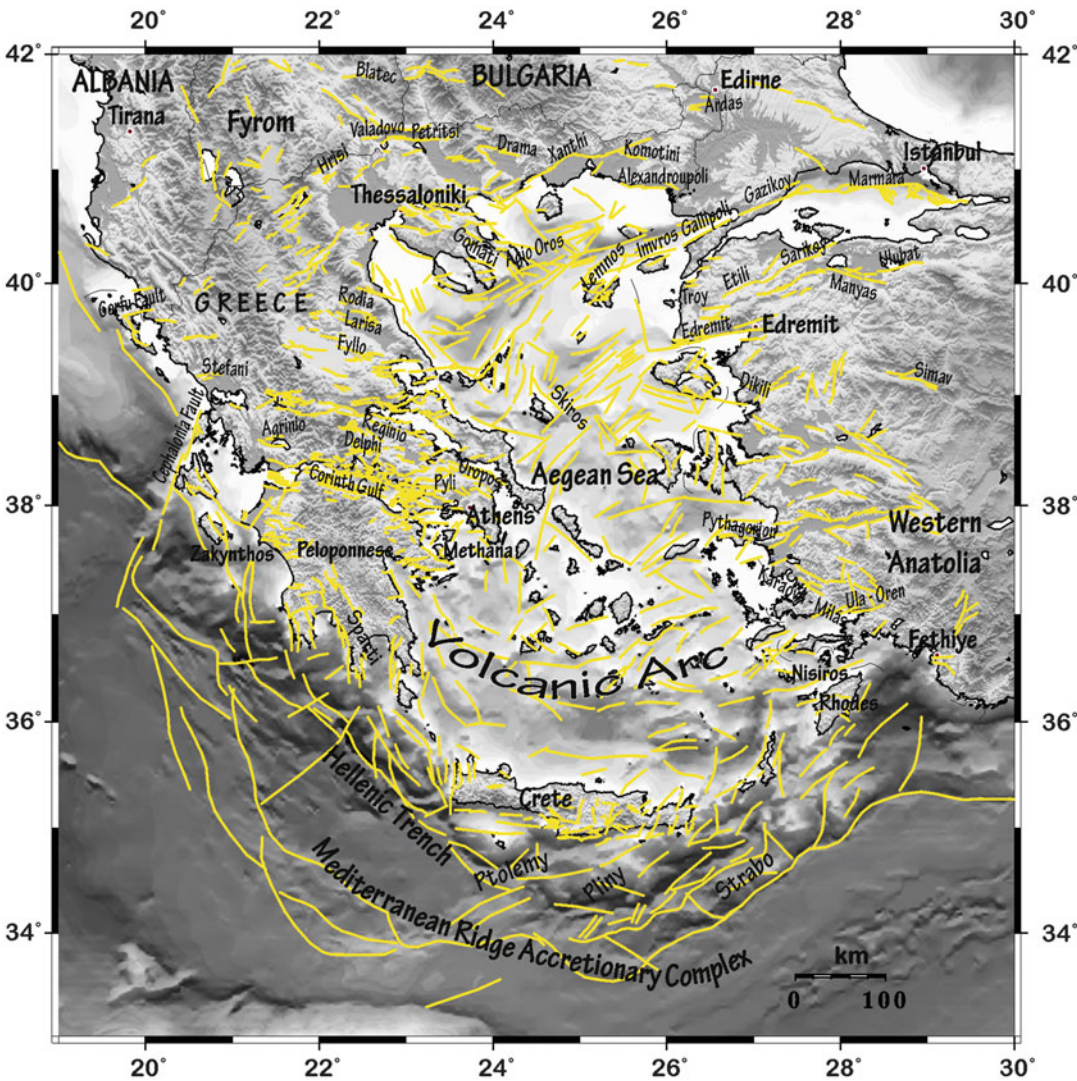
In this geodynamic context, the deformation is distributed and is taken up by a number of faults, of variable length and orientation, on land and offshore, as illustrated in Fig. 3. Major fault segments are populated by families of minor faults forming wide zones of distributed faulting.

The Hellenic Subduction Zone

The Hellenic subduction zone (Jolivet et al. 2013; Ring et al. 2010) extends (Fig. 3) from approximately south of Zakynthos in the west up to Rhodes Island in the east. South of Crete, there is the *Mediterranean Ridge* which is the accretionary prism formed by the pile of sediments, up to 10 km thick that overlies the oceanic crust. The *Hellenic Trench* is a scarp in the bathymetry 2–3 km high, which is more clearly developed

in the western part of the subduction zone, between Peloponnese and Crete, while it splits into branches toward the east, the best developed of which are known as the *Ptolemy*, *Pliny*, and *Strabo Trenches*. After the recognition that the Mediterranean Ridge is an accretionary complex, it became apparent that the Hellenic Trench can be better understood as sediment-loaded fore-arc basin or else as a backstop to the deformed accretionary complex. The Hellenic Trench does not define the plate boundary, as it is common in other ocean trenches; it is not the surface projection of the subduction interface. The African–Aegean plate boundary itself lies south of the Mediterranean Ridge or very likely is buried beneath it. This can be easily understood from an assumption and simple geometry. Assuming that the low-angle thrusts, underlying the Hellenic Trench at depths up to 40–50 km, represent the interface between the subducting African lithosphere and the overriding Aegean plate, then their surface projection should be sought at least 150 km south of Crete at the Mediterranean Ridge. The degree of coupling on the Hellenic subduction interface, i.e., the fraction of the motion across the plate boundary accommodated by elastic strain accumulation, is estimated to be low (10–20 % of the full African–Aegean convergence rate).

The *Aegean Volcanic Arc* (Fig. 3) extends from Methana on the eastern coast of Peloponnese to the Nisyros Island in the east, close to coast of western Anatolia. The Hellenic Trench, the south Aegean Volcanic Arc, and the back-arc sea are the tectonic elements of the *Hellenic Arc*, which exhibits a strong curvature. It must be noted that in the present work, the back-arc area refers to the region behind the south Aegean Volcanic Arc and the fore-arc area refers to the region in front of the south Aegean Volcanic Arc (toward Crete). The identification of the Wadati–Benioff zone in the Hellenic Arc has been accomplished since the early 1970s (Papazachos and Comninakis 1971) and was subsequently better defined (Papazachos et al. 2000). This zone is relatively flat ($\sim 16^\circ$) beneath Peloponnese, has an average dip of 30° beneath Crete, and is steeper (50°) in the eastern part, near Rhodes Island.



Mechanisms of Earthquakes in Aegean, Fig. 3 Major tectonic structures in the Aegean, which are discussed in the text. The fault lines (available from the database of

Basili et al. 2013) are plotted aiming to show that in this tectonic setting, the deformation is distributed along wide zones including major and minor fault segments

The present work is focused on the earthquake focal mechanisms of the Aegean and the information they provide on the faulting characteristics and the regional stress field. Thus, in the sections that follow, a brief background on the sources of focal mechanisms for earthquakes in the Aegean is presented. Then, the available focal mechanisms are classified depending on the faulting type, and the general pattern of the spatial distribution of each category is presented. A number of features of specific interest are discussed in more detail, for

the region between the meridians 19–28°E and the parallels 34–41°N, even though in a number of cases broader regions are included in the figures.

Published Earthquake Focal Mechanisms in the Aegean: Historical Background

The published focal mechanisms for the Aegean follow the worldwide advances in instrumentation

and computational methods. In a global scale, the period 1962–1976 marks the era of analogue records, and from 1976 onward, the digital era takes place. In a regional scale, the first seismograph in Greece was installed in Athens in 1911 and marks the era of instrumental seismology for the Aegean. In mid-1960s, the World-Wide Standardized Seismographic Network (WWSSN) was established which provided high-quality analogue seismograms to the scientific community. These data were collected at Lamont–Doherty Observatory and at the United States Geological Survey (USGS) in Denver and Golden, Colorado.

By the 1980s, WWSSN was upgraded to form the 150+ station Global Seismographic Network (GSN), which currently has substantial capability in recording a broad spectrum of motions from Earth tides to local ground noise. Nowadays, free access to three-component broadband time series of ground motion is provided by the International Federation of Digital Seismograph Networks (FDSN), through the IRIS Data Management System (iris.edu/data). The advent of digital seismology together with the subsequent theoretical and computational advances in calculating synthetic seismograms increased the number of the published earthquake focal mechanisms worldwide as well as for the Aegean area. Thus, since 1976, nearly all earthquakes with $M_w > 5.5$ have a published solution based on waveform modeling.

Over the years, the published earthquake mechanisms for the Aegean have been determined using different methods and waveforms, which are briefly presented in the following section.

Focal Mechanisms Determined from First-Motion Polarities

These are focal mechanisms of Aegean earthquakes determined from the first-motion polarity of P-waves. From the published first-motion solutions, more reliable are those for which (a) the first P-wave polarity was read on long-period instruments, preferably on the WWSSN 15–100 s instruments, and (b) the observed polarity distributions are also published. Long-period records are generally less noisy, so the polarity is clearer to read, while the publication of the

observed polarities provides the means to evaluate the solution, judging from the number of the polarities, the azimuthal distribution of the stations used, and the number of nodal readings that constrain the orientation of the nodal planes, to mention a few criteria. Depending on the distribution and quality of first-motion data, more than one focal mechanism solution may fit the data equally well (see, e.g., the compilation by Udias et al. 1989). From the first-motion solutions which are available for Greece, a significant number include focal mechanisms for small and moderate magnitude ($M < 5$) earthquakes and aftershock sequences, recorded by temporary networks. These were installed in mainland Greece and the islands, in the framework of a series of projects, mainly in the decades from 1980 to 2000 (e.g., Hatzfeld et al. 1990). First-motion polarity solutions for earthquakes in the Aegean remain an important database even though their significance was later surpassed by the solutions obtained using waveform modeling.

Focal Mechanisms Determined from Waveform Modeling at Teleseismic Distances

These are focal mechanisms of Aegean earthquakes determined by the computation of synthetic seismograms that best fit the observed seismograms at teleseismic distances in the range of 30° – 90° . The technique is known as waveform modeling, and synthetics are calculated using a ray theory approach, a well-known code is MT5 (Zwick et al. 1994), usually referred to as *body-wave analysis*, or on a normal-mode approach commonly referred to as *centroid moment tensor* or CMT analysis. The main tenet in both approaches is to use wavelengths of seismic waves that are longer compared to the earthquake source dimensions (i.e., the fault length). If this is fulfilled and if the source-station separation is larger than the fault length, then the source appears as a point in space (the so-called centroid) even though it may have finite time duration. In a number of cases, the point-source approximation is not valid. For example, when the observed waveforms support the existence of discrete sub-events, then the contribution of these sub-events can be modeled introducing a number

of separate sources appropriately lagged in space and time. Well-known source for CMT solutions is the Global CMT Project (globalCMT.org), formerly known as Harvard CMT catalog (Ekström et al. 2012). Other sources include the body-wave moment-tensor solutions, which have been reported from the United States Geological Survey since 1981 (Sipkin 1994).

There are differences in the previous inversion schemes. In the body-wave analysis, the focal mechanisms are constrained to be double-couple solutions (i.e., to result from slip on a fault). The window of data examined in the inversion is short; it includes the first part of the P and SH (horizontal component of S-wave) waveforms. For example, for shallow earthquake sources, the data window examined usually ranges from ~ 20 to 40 s and allows the inspection of relatively high frequencies. Focal depths constrained by the so-called depth phases, pP and sP, which leave the source upward, are reflected at the free surface and continue further as P-waves. To identify the separation between P and pP or sP is more difficult the shallower the source. For example, for depths in the range 10–20 km, the time separation between P and pP and between P and sP is of the order of 4–10 s. Commonly, during the processing stage of digital waveforms from the modern global networks, after removing the instrument response, the waveforms are convolved with the response of the old long-period WWSSN 15–100 instruments (with a peak response at ~ 15 s), because it provides a frequency range suitable for the inversion.

The CMT method is based on the linear relationship that exists between the six independent elements of a zeroth-order moment-tensor representation of an earthquake and the ground motion that the earthquake generates. The “centroid” of the CMT method refers to the center of the earthquake moment distribution in time and space, defined by four additional parameters. Ten parameters (six moment-tensor elements and the centroid latitude, longitude, depth, and centroid time) thus provide the point-source CMT representation of an earthquake. No volumetric component is allowed in the moment tensor, and the published focal mechanisms are constrained to be

purely deviatoric. Longer time windows are used compared to the time windows used in the classic body-wave analysis. For the typical $5.5 \leq M_w \leq 6.5$ Aegean earthquakes analyzed, the body waves are filtered to displacement between 150 and 40 s, with a flat response between 100 and 50 s. In a number of cases, the CMT inversions cannot sufficiently resolve the spatial centroid of the source, in particular the focal depth for shallow sources. Thus, in the GCMT catalog, sometimes the focal depth is fixed, when the inversion attempts to place the source above the minimum allowed depth of 12 km or when the depth exhibits persistent instability in the inversions.

Focal Mechanisms Determined from Waveform Modeling at Regional Distances

These are focal mechanisms of Aegean earthquakes determined by waveform analysis, moment-tensor inversion from body waves (Dreger 2003; Sokos and Zahradnik 2008), and centroid moment-tensor inversion, using ground motion records from broadband sensors at regional distances. The trick in this methodology is to use long-period regional-distance seismic waves. The source process can be reduced to a simple delta function “spike” in space and time. The wave propagation is also simplified because filtering regional seismograms to long periods (or low frequencies) results in waves that have only propagated in a few wave length cycles and can be predicted using relatively simple one-dimensional layered Earth models. Canceling out all of these complex effects leaves the robust extraction of the radiation pattern. A more generalized moment tensor can be solved with simple constraints added e.g., that the source has no isotropic component, although this constraint is not valid when estimating the moment tensor of volcanic sources.

The Era of Real-Time Focal Mechanisms in the Aegean

Seismic instrumentation in Greece was gradually upgraded to obtain its present-day configuration which counts 200+ broadband sensors in operation. The funding to the Greek seismological

Mechanisms of Earthquakes in Aegean, Table 1 Classification of the Aegean focal mechanisms based on the plunge of P, T, and N stress axes (After Zoback 1992)

Plunge (pl) of axes			Faulting type
P-axis	T-axis	N-axis	
$pl \geq 52^\circ$	$pl \leq 35^\circ$		Normal faulting [NF]
$40^\circ \leq pl < 52^\circ$	$pl \leq 20^\circ$		Normal oblique faulting [NS]
$pl \leq 35^\circ$	$pl \geq 52^\circ$		Thrust faulting [TF]
$pl \leq 20^\circ$	$40^\circ \leq pl < 52^\circ$		Thrust oblique faulting [TS]
$pl < 40^\circ$	$pl < 40^\circ$	$pl \geq 45^\circ$	Strike-slip faulting [SS]

community usually increases in the aftermath of a strong event, thus the seismological networks were mainly upgraded after the Athens 1999 earthquake. The establishment of the Hellenic Unified Seismological Network (HUSN) marks the beginning of a new era for instrumentation coverage in Greece (see <http://bbnet.gein.noa.gr/HL/> for network geometry). Four centers constitute the core members of HUSN: (a) the Geodynamic Institute of the National Observatory of Athens (NOA), (b) the Seismological Laboratory of the Aristotle University of Thessaloniki (AUTH), (c) the Seismological Laboratory of Patras University (UPSL), and (d) the Seismological Laboratory of Athens University (UOA).

In late 2005, these Greek seismological agencies began a systematic determination of moment-tensor solutions for strong earthquakes, and these solutions are transmitted to the European-Mediterranean Seismological Centre (EMSC-CSEM/). They can be retrieved from the following website using the code of each institute (<http://emsc-csem.org>).

Gross Features of the Spatial Organization of the Aegean Mechanisms

In the following sections, the spatial organization of earthquake focal mechanisms in the Aegean is discussed, using data from published sources and for earthquakes after mid-1950s mainly with $M_w \geq 5.5$. For the period 1950–1959, the focal mechanisms are mainly from first-motion polarities. Even though smaller magnitude events are usually not a good proxy for regional tectonics, in

a number of cases, they are included in the plots for emphasis to some patterns. To classify the focal mechanisms, the criteria listed in Table 1 were used. Focal mechanisms from shallow-focus ($h \leq 50$ km) earthquakes and deeper than 50 km are examined separately. This cutoff depth, even though arbitrarily chosen, provides a useful depth range, because the deeper than 50 km earthquake mechanisms are all confined within the subducting African lithosphere.

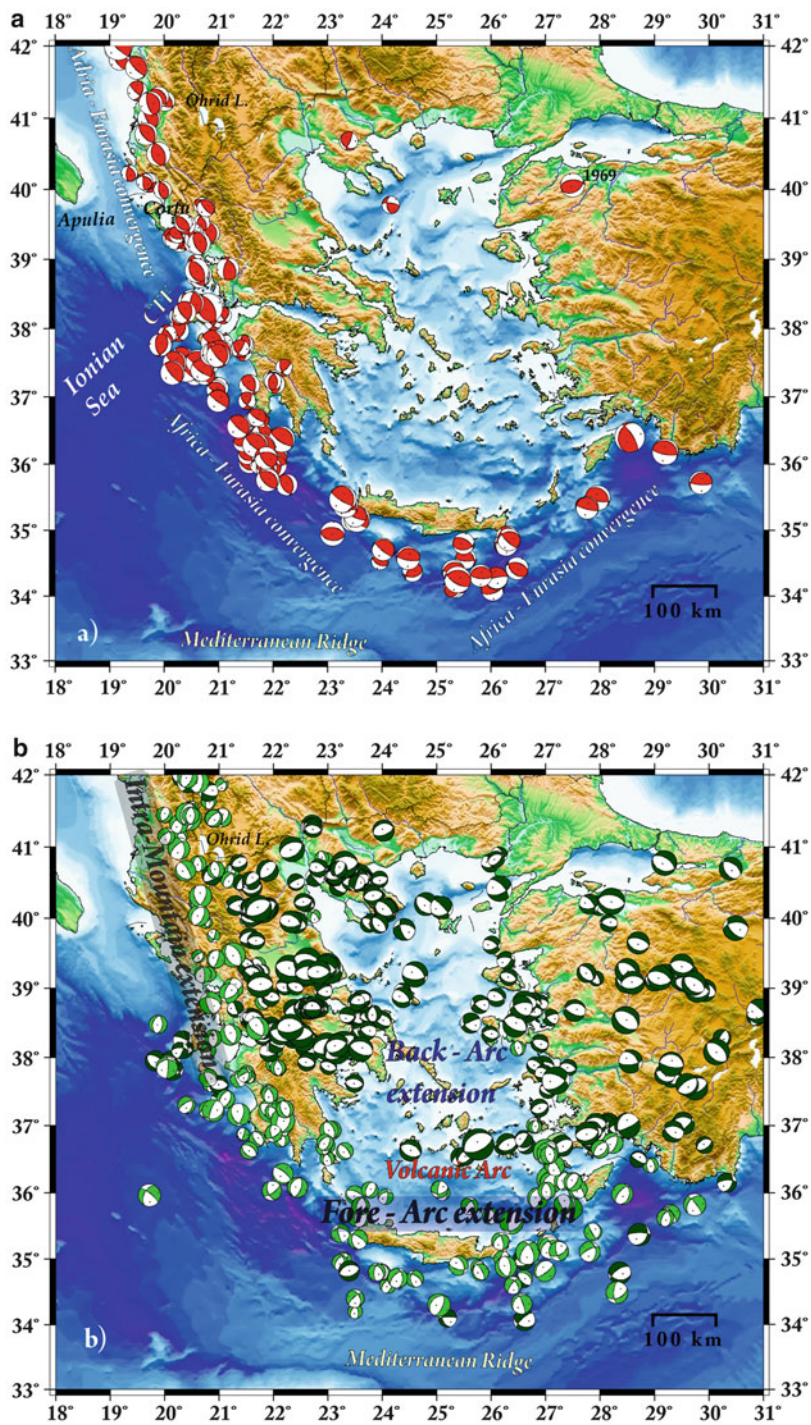
It should be noted that the maximum depth of earthquakes within the lithosphere, which defines the thickness of the seismogenic layer, is of the order of 10–20 km in the Aegean. This thickness provides a scale for the dimensions of active structures, as well. Earthquakes which are strong enough to rupture the entire thickness of the seismogenic layer are produced from large faults. Moderate-size events with ruptures confined within the seismogenic layer are produced from smaller faults (Jackson 2001).

Figure 4a–d summarize the classification of the shallow ($h \leq 50$ km) focus and intermediate ($h > 50$ km) focus earthquake focal mechanisms, respectively, into three major categories: reverse/thrust faulting, normal faulting, and strike-slip faulting. Some general features are summarized below, while selective features of the spatial organization of the focal mechanisms, in all cases for shallow-focus events, are discussed in more detail in the following sections.

The Zone of Thrust/Reverse Faulting

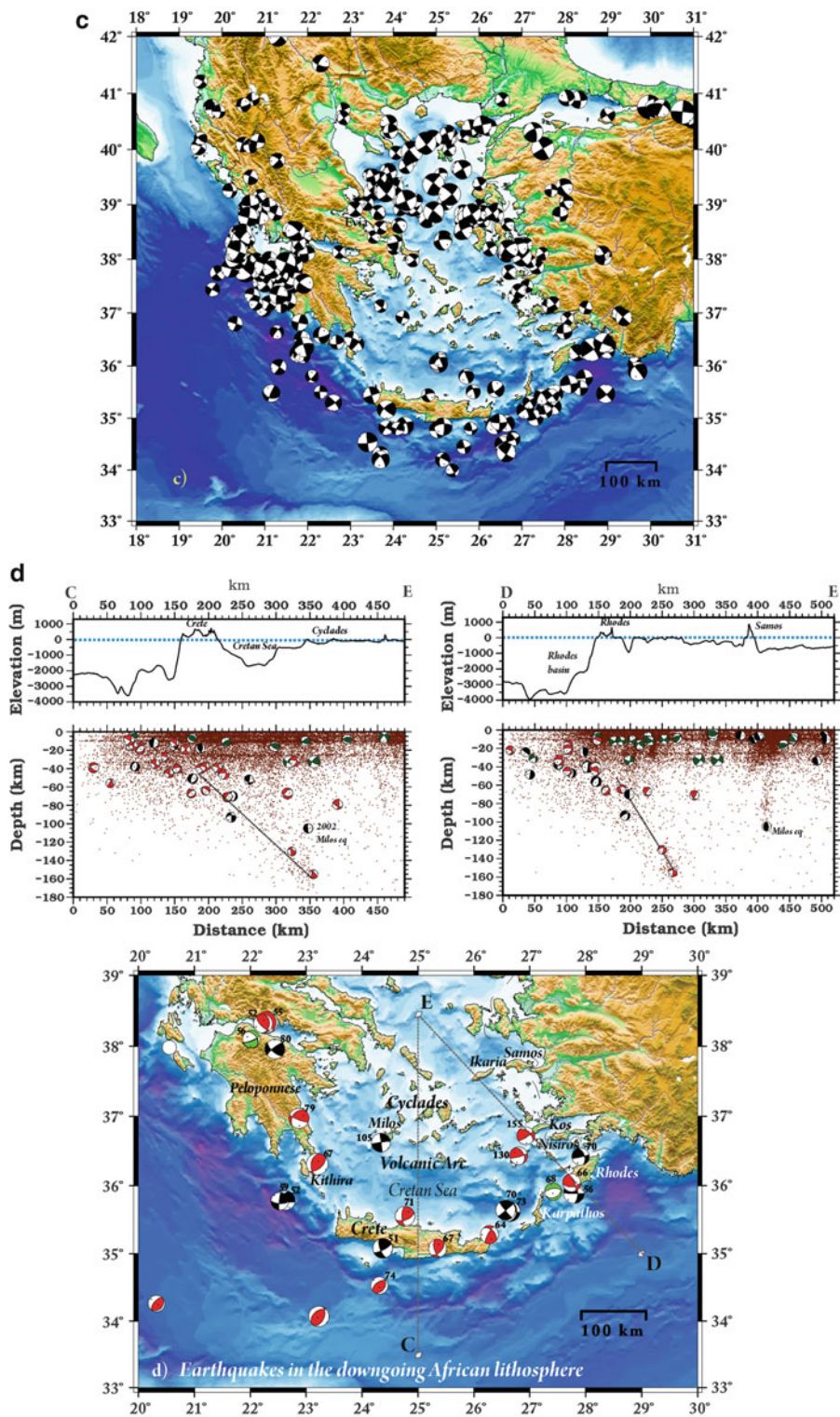
The Thrust Zone Along the Adria: Eurasia Convergence

Thrust faulting (Fig. 4a) dominates the foreland and hinterland flanks of Albania, where the



Mechanisms of Earthquakes in Aegean, Fig. 4 (continued)

M



Mechanisms of Earthquakes in Aegean, Fig. 4 (continued)

collision of Adria–Eurasia occurs. The active thrust front is related with (a) low-angle thrusting ($<35^\circ$) toward the coastal region, which extends to depths up to 30 km, and (b) steeply dipping ($>35^\circ$) reverse type faulting, probably blind, which is overlain by anticlines, clearly visible in the topography. In the anticline geometry, the large thickness of salt formations plays a significant role (Copley et al. 2009). The majority of earthquakes are connected with the low-angle thrust faults, and fewer are connected with the steeper reverse faults (Kiritzi and Louvari 2003; Tselentis et al. 2006).

The zone of thrusting from Albania continues southward (Fig. 4a), through the coast of western Greece, along the inner east coast of the Ionian Islands and connects south of Zakynthos Island with the western termination of the Hellenic subduction zone. This thrusting is mainly along N-S trending planes, with slip vectors trending E-W. High-resolution seismic images of P- and S-wave velocity perturbations, across a line that starts from Corfu and crosses western and central Greece, provide new evidence that continental crust is subducting below northern Greece (Pearce et al. 2012) in contrast to the oceanic crust subducting between Peloponnese in the south. The images show a ~ 20 km thick low

velocity layer (LVL) dipping at $\sim 17^\circ$ within the upper mantle beneath northern Greece. It is interpreted as subducted continental crust. This interpretation requires that ~ 8 km of sediments have been scraped from the subducting slab and have been accreted into the thrust belt of the overriding plate, in accordance with previous results (Royden and Papanikolaou 2011). This is required because marine seismic data indicate that the thickness of the continental crust within the foreland is ~ 28 km, consisting of ~ 20 km of crystalline crust overlain by ~ 8 km of sediments. The low seismicity level between Africa and Apulia and the very similar pattern of the subducted slab north and south of the Cephalonia transform fault support the model of a continuous slab from north to south (Pearce et al. 2012). This continuity and the offset induced by the Cephalonia transform fault would provide a lateral ramp geometry for this fault (Shaw 2012) instead of a slab tear.

The Thrust Zone Along the Africa–Eurasia Convergence Zone

South of Zakynthos Island and up to Rhodes Island and the southern coast of western Anatolia, the thrust and reverse faulting reflects the African–Aegean collision (Fig. 4a). The

M

Mechanisms of Earthquakes in Aegean, Fig. 4 Spatial organization of earthquake focal mechanisms in the Aegean (the size of the beach-ball scales with magnitude). Data were classified into the three major fault types, based on the plunges of the principal axes, and are presented for shallow ($h \leq 50$ km) and deeper sources, separately. The data included in the figures are for earthquakes with $M_w > 4$ which occurred in the period 1950–2013, with the smaller magnitude events to span the most recent years. From 1960 onward, most of the mechanisms have been determined with waveform modeling. (a) Reverse/thrust faulting in the Aegean along the subduction zones. In western Albania and Greece, the compression reflects the Adria–Eurasia collision zone; south of Zakynthos Island, the compression reflects the African–Aegean convergence. Sparse thrusting in northern Greece is also evident together with the focal mechanism for the event of 3 March 1969 M_w 5.8, south of the North Anatolian Fault, which probably reflects

contractional irregularities on strike-slip faults. (b) Normal faulting in the Aegean. Dark-shaded beach balls denote the N-S extension in the back-arc region (as defined here), whereas the light-shaded beach balls denote the $\sim$ E-W intra-mountain extension (Hellenides Mts) which connects with the along-arc extension that the overriding Aegean plate undergoes, south of the volcanic arc. (c) Strike-slip faulting in the Aegean discussed in more detail in the text. (d) Distribution of intermediate-depth ($h > 50$ km) earthquake focal mechanisms ($M_w > 5.5$), whose depths are also plotted. They are all confined within the subducting African slab which undergoes along-arc compression. The majority of the data show strike-slip and reverse/thrust faulting, with the deepest events in the eastern part of the subduction. The two cross sections depict the steeper dip angle of the subduction in the eastern edge compared to the central part. The seismicity plotted (dots) is from HypoDD relocated events and from field experiments

earthquakes along this zone follow the bathymetry of the Hellenic Trench and show a tendency, at least during the years of observation, to form dense discrete clusters, separated by areas with no strong earthquakes, as, for example, south of Peloponnese and Crete. The low-angle landward dipping planes are the fault planes, and the P-axis is orthogonal to the subduction interface in its western margin and parallel to it, due to arc-oblique convergence in its eastern termination.

The Zone of Normal Faulting

E-W Extension Along the Albanides: Hellenides Mountainous Chains

There is a distinct zone of $\sim$ N-S normal faulting (Fig. 4b – light-shaded beach balls) that runs along the mountainous area of eastern Albania and continues in the mainland of Greece, along the backbone of Hellenides Mountains. The $\sim$ E-W extension along this zone is perpendicular to the long axis of the mountains. There is evidence that this N-S normal faulting reaches the surface. This can be deduced from the shape of a number of intra-mountain basins, as, for example, Lake Ohrid, which has the characteristic shape of a normal fault-bounded basin, and from the surface expression of the faults at its southern end in NW Greece (Goldsworthy et al. 2002).

E-W Extension Along the Southern Part of the Overriding Aegean Plate (Fore-Arc)

South of the Volcanic Arc in southern Aegean, the distribution of the focal mechanisms reveals a broad zone of N-S normal faulting (Fig. 4b – light-shaded beach balls). It was first identified in western Crete and Peloponnese (Lyon-Caen et al. 1988), but as the focal mechanism database enriches, it is evident that this E-W extension is the dominant mode of deformation within the entire southern part of the overriding Aegean plate.

N-S Extension Along Mainland Greece and Western Anatolia (Back-Arc)

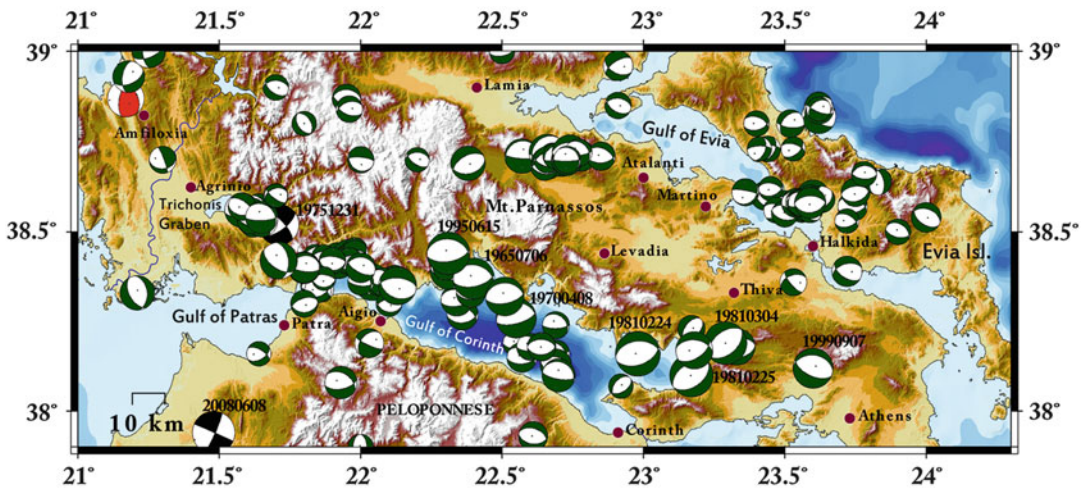
The back-arc Aegean and western Anatolia are dominated by normal faulting along roughly E-W

trending planes. This is the dominant mode of faulting (Fig. 4b – dark-shaded beach balls), which is frequently combined with a considerable strike-slip component. The dip angle of the nodal planes (both planes considered) is in the range of 30 – 65° with an average value of $\sim 50^\circ$. The age of the onset of the back-arc Aegean extension is from 15 to 45 Ma, depending on the author and the data used to define it (as, e.g., normal faulting bounding sedimentary basins, metamorphic core complexes in central Cyclades islands). For ~ 30 Ma, only back-arc extension driven by slab rollback was dominant. The interaction between the back-arc Aegean extension and the westward extrusion of Anatolia initiated around 13 Ma ago (middle Miocene) and dominates the Aegean tectonics ever since. It is worth mentioning that since the 1970s, the extension in the Aegean has been connected to the sedimentary basin formation (McKenzie 1978) and to the subduction process itself (Le Pichon and Angelier 1979).

The Zone of Strike-Slip Faulting

The Shallow-Depth Strike-Slip Fault Zones

Figure 4c shows all earthquakes with strike-slip focal mechanisms. It is evident that there are several narrow zones which accommodate shear motions, and in general, the NE-SW trending structures are associated with dextral strike-slip motions, whereas the NW-SE trending ones are associated with left-lateral strike-slip motions. In eastern Aegean Sea, the strike-slip zones reflect the shear transmitted from the east, and the major and minor segments of the North Anatolian Fault are well outlined. The strike-slip faulting reaches Evia Island but it does not cross central Greece. In western Greece, the most striking feature is the wide strike-slip zone which covers the Ionian Islands and western Peloponnese. North of $\sim 40^\circ$ parallel, the sparse strike-slip faulting is mainly confined at the edges of major normal faults. It is worth noting here the abundant strike-slip faulting along the Hellenic arc, which follows bathymetric lineaments. It is more evident at its eastern termination from the occurrence of a few strong events.



Mechanisms of Earthquakes in Aegean, Fig. 5 Development and continuation of ~E-W normal faulting bounding graben structures and basins in the region extending from the Gulf of Corinth to the Gulf of

Evia, as discussed in the text. In this and the following figures, the dates for all earthquakes with $M_w \geq 5.9$ are shown next to the beach balls

The Intermediate-Depth Mechanisms Within the Subducting African Lithosphere

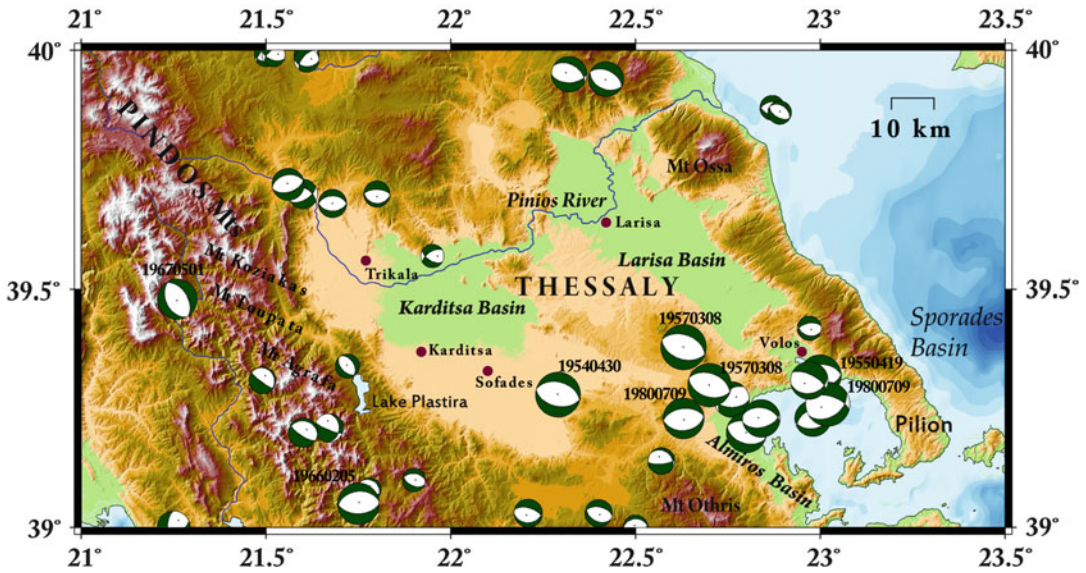
Figure 4d summarizes the focal mechanisms for the intermediate-depth ($h > 50$ km) earthquakes ($M_w > 5.5$), which are all confined within the subducting African plate. These focal mechanisms indicate reverse faulting with considerable strike-slip component or pure strike-slip faulting. Evidently, the subducting African slab undergoes along-arc compression. The downdip length of the slab is approximately 300 km and reaches to a depth of about 160 km, as deduced from the results of field experiments (Hatzfeld and Martin 1992). There is a pronounced consistency regarding the P-axes, which, in general, are parallel to the local curvature of the subduction zone, whereas the T-axes are aligned downdip within the subducting slab (Benetatos et al. 2004). The deepest events are concentrated at the eastern part of the subducting slab, NW of Rhodes, where the slab has a steeper dip (see the cross sections in Fig. 4d), as was mentioned in the introduction.

In the following sections, special features of the earthquake focal mechanisms in the Aegean are discussed in more detail.

Characteristics of Normal Faulting in Mainland Greece

The prevalence of normal faulting in the mainland of Greece is manifested in a number of grabens and basins in central Greece, whose edges are bounded by large segmented normal faults, which in many cases are associated with impressive topographic escarpments. The maximum typical lengths of these fault segments are of the order of 15–25 km. In the following section, starting from south toward north, a number of these regions of significant normal faulting concentration are discussed.

The best studied region is the Gulf of Corinth (Fig. 5), an asymmetric graben (or rift) 100 km long, which ranks first among the rapidly extending regions in Greece. The geodetic measurements indicate a fast rate of 12–14 mm/year in the west part of the Gulf of Corinth and a slower rate of 6–8 mm/year in the eastern part. The present-day structure and kinematics of the Corinth Rift can be explained with a series of décollements relayed by steeper ramps that altogether formed a mechanically weak, crustal-scale detachment (Jolivet et al. 2010). The E-W normal



Mechanisms of Earthquakes in Aegean, Fig. 6 Development of WNW-ESE and E-W normal faulting in Thessaly, as deduced from focal mechanisms, mainly along the southern margins of the basins,

extending from Sofades to the Pelion Peninsula. Along this zone, a series of strong events occurred in the 1950s and 1980s (dates designated in the figure)

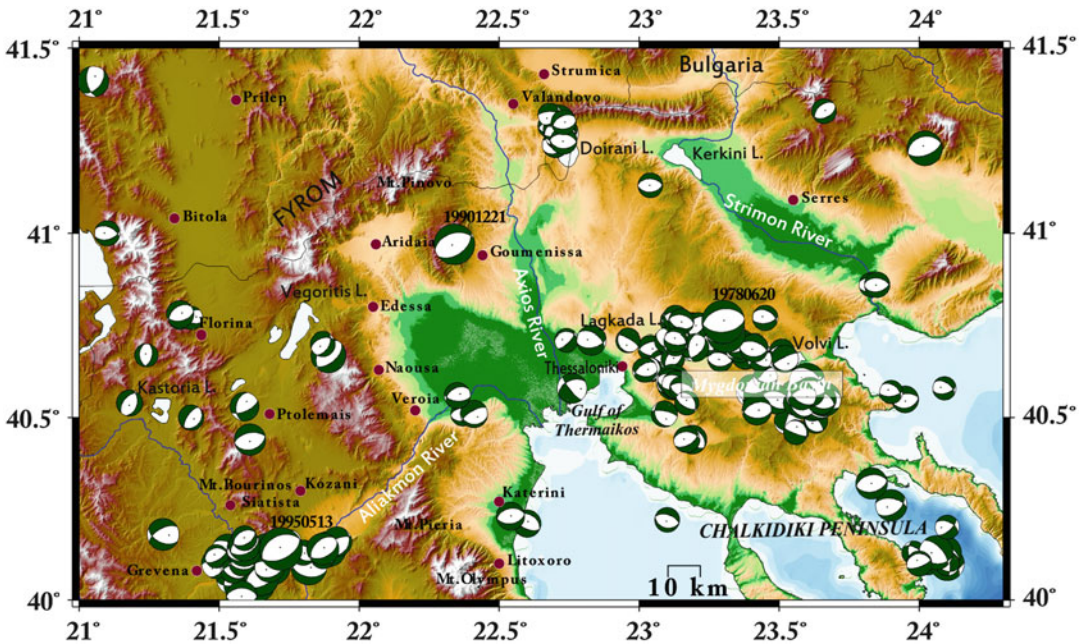
faulting observed in the Gulf of Corinth continues eastward through the Thiva basin and into the Gulf of Evia. In the family of faults within the broader region of this basin, the most recent earthquake activity occurred during February–March 1981 and on September 1999 earthquake sequence. Few mechanisms are available for strong earthquakes on Evia Island itself; however from a number of moderate-size events, there are indications of normal faulting along mainly E-W trending planes.

Further to the north in central Greece (Fig. 6), most of the available focal mechanisms are from earthquakes with epicenters along the southern side of the Karditsa and Larisa basins, between Sofades and Volos. These basins are bounded by normal faulting striking NW-SE or E-W.

In northern Greece, a region of pronounced normal faulting is the Mygdonian basin (Fig. 7) close to Thessaloniki, the second metropolitan city in Greece, after the capital, Athens. The basin between the Lagkada and Volvi Lakes shows the operation of nearly E-W trending normal faulting (Tranos and Lacombe 2014), which is considered more active and more recent and is

optimally oriented to the present-day stress field. One of these faults moved during the 1978 Mw 6.5 earthquake which produced significant losses to the city of Thessaloniki. The zone of E-W normal faulting close to eastern Thessaloniki is not clear if it continues westward, crossing the Gulf of Thermaikos. In this area, there is only moderate-size recent seismicity, with focal mechanisms predominantly showing strike-slip motions; however it is doubtful whether they reflect the regional tectonics. The epicenters of this modern seismicity are in the flood plain of the Axios and Aliakmon rivers, and the thick pile of sediments obscures field observations.

Further to the west, (Fig. 7) along the Ptolemais plains, the focal mechanisms indicate the operation of normal faulting along NE-SW trending planes, and the most well-studied event in this area is the 15, 1995, sequence between the cities of Grevena and Kozani. There are indications, from field observations and moderate-size seismicity, on the existence of gradually evolving network of distributed active normal faulting, between the Kastoria Lake and Vegoritis Lake (see also Fig. 3).



Mechanisms of Earthquakes in Aegean, Fig. 7 Development and continuation of normal faulting in northern Greece. Along the Mygdonian basin, there are abundant focal mechanisms that show development of

mainly E-W normal faulting, whereas in other parts of this region, the number of the available mechanisms is limited. Nevertheless, they do depict the acting stress field and the large-scale topography

M

Characteristics of Strike-Slip Motions in the Aegean

Strike-Slip Motions in Central and Eastern Aegean Sea

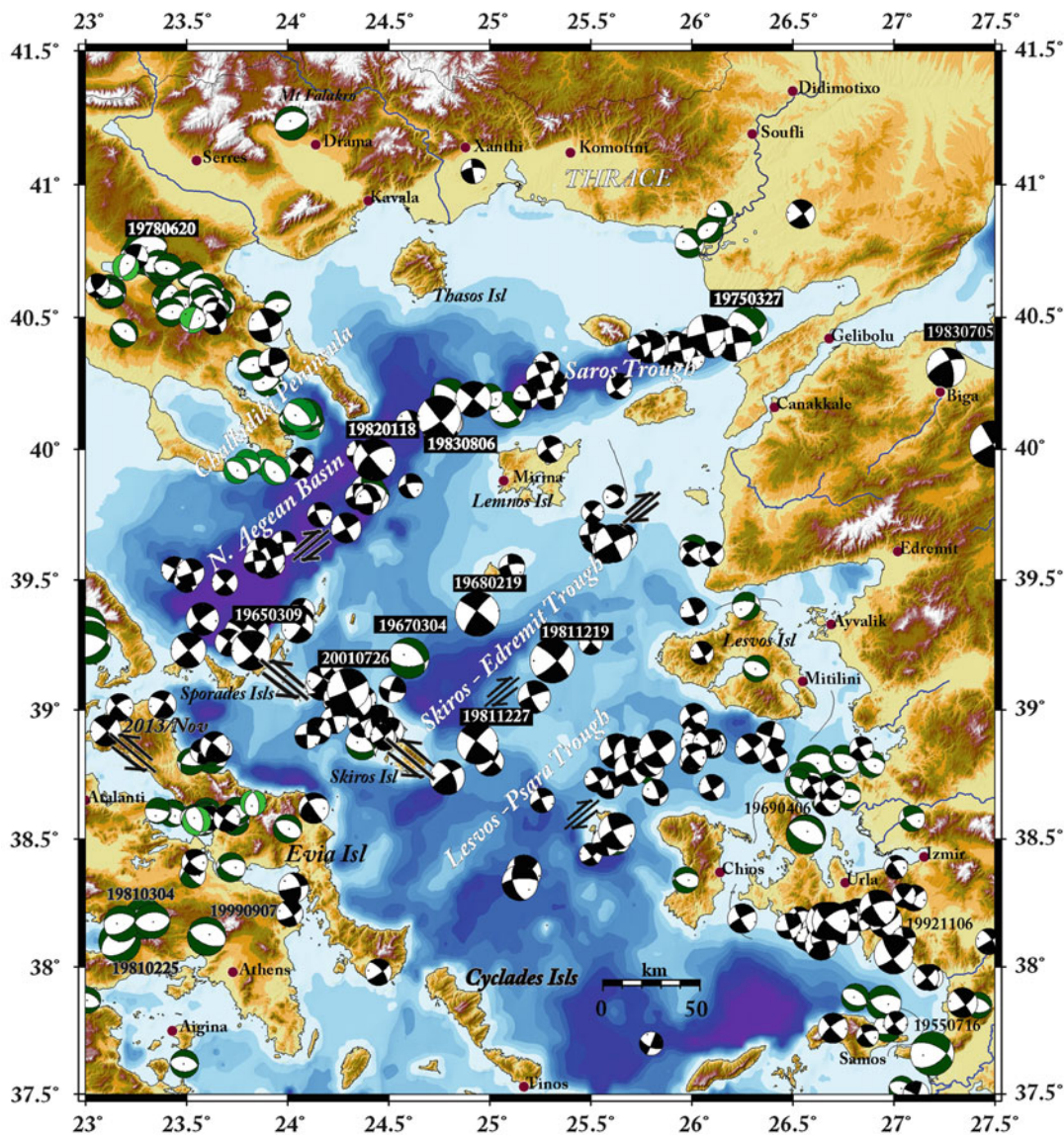
The main characteristics of the strike-slip faulting in the eastern Aegean Sea and Thrace are (Fig. 8) (a) its connection with topographic and bathymetric escarpments that run for several hundreds of kilometers and (b) its concurrent operation with normal faulting, which differs in orientation from the classic E-W trending faults characterizing the Greek mainland. The most prominent example for the latter is the normal faulting event of 19670304 Mw 6.6 in Skiros basin; this ENE-WSW extension orientation characterizes most of the normal faulting events in central Aegean Sea (Chatzipetros et al. 2013). It can be traced up to the three-leg Chalkidiki peninsula as deduced from a few moderate earthquake focal mechanisms.

In northern-eastern Greece, the Kavala-Xanthi-Komotini fault zone has a prominent

escarpment in the morphology, which suggests normal oblique faulting as the prevailing mechanism. Unfortunately, the instrumental seismicity and the available focal mechanisms are not sufficient to describe the type of faulting.

Along the eastern Aegean Sea, strike-slip faulting forms distinctive stands, which bound a number of basins/troughs, the most prominent of which are: (a) The North Aegean Basin (NAB) which consists a number of fault-bounded basins which extend from Saros bay in the east to the Sporades islands in the west. Moreover, along the NAB which accommodates a large part of the westward motion of Anatolia relative to Eurasia, the occurrence of strong events is abundant. (b) The Skiros- Edremit Trough and the Lesvos-Psara Trough, which are both the sites of several strong events. There is evidence from the bathymetry and the focal mechanisms of small magnitude events that the strike-slip faulting reaches the Cyclades Islands in the west.

The dextral strike-slip motions in the north and central Aegean Sea extend westward and



Mechanisms of Earthquakes in Aegean, Fig. 8 Distribution of strike-slip faulting in central and eastern Aegean Sea and regions of localized normal faulting within it, as discussed in the text. The arrows

indicate the polarity of shear motions, right-lateral in the central and eastern part and left-lateral near the interface with the mainland of Greece

terminate close to the mainly E-W striking normal faulting system that controls the eastern coastline of the Greek mainland. This transition from dextral NE-SW shear motion to E-W normal faulting is reflected in the azimuth of the slip vectors, which changes from NE-SW to N-S.

Strike-Slip Motions Near the Intersection with the Eastern Greek Mainland

The linkage of the strike-slip motions in eastern Aegean Sea with the normal faulting in mainland Greece has been an interesting field of research and debate. In this transition zone, the Sporades Islands and Evia Island have a key location (see Fig. 8).

One point that is worth mentioning is the presence of NW-SE trending structures, close to this transition zone, that are activated as strike-slip faults with left-lateral component and not dextral as in the eastern Aegean Sea.

The first well-recorded information for left-lateral strike-slip motions was from the 20010726 Skiros Island earthquake sequence (Fig. 8). A number of previous strong earthquakes, for example, the event of 19650309 in Sporades Islands, could be also connected with left-lateral strike-slip motions. A well-located seismic sequence in northern Evia, which burst in November 2013, provided additional evidence for left-lateral strike-slip motions.

In view of the above, the operation of these NNW-SSE trending structures as left-lateral strike-slip faults should be also taken into account in the discussions regarding the change in the direction of the slip vectors. It is also worth mentioning that this left-lateral strike-slip component along the western shoreline of the North Aegean Sea is well resolved in the GPS-derived strain field (Müller et al. 2013). The change of strike of the active structures from east to west, e.g., from NE-SW to east Aegean to NW-SE at the edge of the Greek mainland to E-W on the mainland itself, and the subsequent change in the direction of the slip vectors can be accommodated by clockwise rotations of the fault-bounded blocks in central Greece relative to Eurasia (broken-slats model of Taymaz et al. 1991).

Strike-Slip Motions in Western Greece

As previously mentioned, western Greece lies in the crossroads of continent–continent collision in the north and ocean–continent subduction in the south (Fig. 9). These two regimes are offset by ~100 km (Pearce et al. 2012) across the Cephalonia–Lefkada Transform Fault, whose existence has been postulated by McKenzie (1978). It is expressed as an abrupt lateral ramp in the bathymetry (Shaw 2012) approximately 4,000 m deep, between the Adriatic Sea and the Ionian Islands. Two segments have been identified along this fault zone, the 40 km long Lefkada segment and the 90 km long Cephalonia segment

(Louvari et al. 1999). This fault zone is very active, and this is the reason why the Hellenic Seismic Code is designed for the highest resistant constructions. Recently, on 26 January 2014, an Mw 6.1 seismic sequence burst on Cephalonia Island, followed by another Mw 6.0 event on 3 February 2014. The first event was connected with a N24°E and steeply dipping strike-slip fault, whereas the second event was connected with a N297°E and shallow dipping thrust fault. These thrust faulting events are not unusual in the Ionian Islands and are usually associated with severe ground shaking. This zone of thrusting follows the Ionian Thrust which is a major structural feature, a crustal-scale thrust fault (Kokkalas et al. 2013). It is worth noting that the 1953 earthquakes that destroyed the islands of Cephalonia and Zakynthos are associated with coastal uplift and thrusting mechanisms.

The strike-slip motions in western Greece cover a wide zone between the Ionian Islands and western Peloponnese. Strike-slip motions along the northwestern coast of Peloponnese have been already noticed in Louvari et al. (1999) and were confirmed by the occurrence of the 8 June 2008 Mw 6.3 Achaia earthquake. This event, which had a source depth of ~22 km, was the first strong event which provided evidence for a gradually developing right-lateral strike-slip zone, the Achaia strike-slip fault zone.

From a number of moderate-size earthquake focal mechanisms, there is evidence for left-lateral strike-slip motions, oblique to the main branches of the right-lateral shear. For example, such a left-lateral shear zone extends from Lake Trichonis toward Amfiloxia and the Gulf of Amvrakikos (Fig. 9), and its operation is supported by the morphology and a characteristic drainage pattern (Vassilakis et al. 2011).

This tectonic setting with distinct parallel branches of NE-SW right-lateral strike-slip faulting and oblique to it NW-SE left-lateral strike-slip faulting resembles the tectonic setting in central and eastern Aegean Sea, as it was previously discussed.

Mechanisms of Earthquakes in Aegean, Table 2 Typical focal mechanisms for normal/reverse/strike-slip type faulting in the Aegean, derived by summing magnitude-normalized moment-tensor components (for $M_w > 4$ and $h \leq 50$ km) of each category. The average depths provide a scale of the thickness of the seismogenic layer

Typical focal mechanisms for shallow earthquakes in the Aegean										
Nodal plane 1			Nodal plane 2			P-axis	T-axis	N-axis	Depth	Fault
Strike°	Dip°	Rake°	Strike°	Dip°	Rake°	az°/pl°	az°/pl°	az°/pl°	(km)	type
89	45	−94	275	45	−86	276/87	182/0	92/3	13 ± 9	Normal
331	33	102	136	58	82	232/13	23/76	141/7	16 ± 11	Reverse
45	84	177	135	87	6	270/2	0/6	166/83	15 ± 9	Strike-slip

The available focal mechanisms for earthquakes with $M > 4$ and for depths less than 50 km are not balanced in terms of the faulting type. The majority is for normal (40 %) and strike-slip (40 %) faulting and the remaining (20 %) is for reverse/thrust faulting. This crude quantification also reflects the dominant modes of deformation within the Aegean area. Taking this gross statistics a step further, Table 2 summarizes the typical focal mechanisms for each fault type, deduced by summing the moment-tensor components and normalizing for the magnitude.

Summarizing, the compression revealed by the distribution of the reverse/thrust faulting (Fig. 4a) is well confined in two regimes: (a) in the north, the collision of Adria and Eurasia takes place, and the active thrust front is confined in a narrow belt in the lowlands of Albania and western Greece; new evidence suggest that continental crust is subducting below northwestern Greece; (b) in the south, at the Ionian Islands, and southward the zone of low-angle thrusting follows the curvature of the Hellenic Trench and continues up to western Anatolia; this low-angle land dipping thrusting reflects the African–Aegean convergence; it does not mark the subduction interface itself, as this is projected south of the Mediterranean Ridge (very likely below the thick pile of sediments). These two subduction zones are separated by the Cephalonia–Lefkada Transform Fault.

~N-S extension (Fig. 4b) is confined in continental Greece and western Anatolia. Within the Aegean Sea itself, the extension is localized, and its orientation is mainly ENE-WSW. In central Greece, the extension is accommodated by slip along families of normal fault segments.

The major and most prominent faults bound the southern part of a series of grabens. The Gulf of Corinth and the Gulf of Evia are good examples of this geometry.

~E-W extension (Fig. 4b) occurs along a distinct belt which can be traced all the way from the mountainous inner Albania, through the backbone of the Hellenides Mountains, spreading along the southern Aegean and ending in western Anatolia. In Albania and northwestern Greece, the zone of E-W extension is subparallel to the outer zone of thrusting and is attributed to contrasts of the gravitational potential energy, between the high mountains in the west and the lowlands of western Albania and the Adriatic Sea. E-W extension occurs south of the Volcanic Arc, in the southern Aegean Sea and Crete itself, and reflects the deformation within the overriding Aegean plate. The boundary, if any, between these two extensional regimes of similar orientation should be sought somewhere south of the 38° parallel where the signature of the subduction process begins to emerge.

Strike-slip faulting (Fig. 4c) is abundant in the Aegean especially in its eastern and western edge. Shear motions tend to develop in parallel strands. For example, in eastern and central Aegean Sea (Fig. 8), the best developed and more clearly related to the prolongation of the North Anatolian Fault into the Aegean is the strike-slip faulting along the North Aegean Basin (or else North Aegean Trough). The right-lateral shear ends abruptly at the western edge of the Greek mainland, near the Sporades Islands and Evia Island. New evidence from modern strong events suggests that along this interface, left-lateral strike-slip faulting is taking place,

which is further supported by updated versions of the GPS-derived strain field (Müller et al. 2013). In brief, as previously suggested depending on the fault strike, the present deformation field allows for both right-lateral and left-lateral shear motions (Kiritzi 2002).

The shear imposed from the Anatolia extrusion passes central Greece as it is depicted by the orientation of the earthquake slip vectors, which alter from being parallel to the shear motions to become approximately N-S. Clockwise rotations of the fault-bounded blocks in central Greece relative to Eurasia (broken-slats model of Taymaz et al. 1991) provide a suitable kinematic mechanism to explain this change in the orientation of the slip vectors.

The shear motions which pass the Greek mainland are accommodated in the western Aegean margin in a similar pattern as in central and eastern Aegean Sea (Fig. 9). This is evidenced by the wide zones of strike-slip faulting which develop, again in parallel strands, along the western coasts of the Ionian Islands and western Peloponnese. These strike-slip faults are mechanically necessary in order to transport the strain and act as planes of weakness. The right-lateral Cephalonia-Lefkada and the more immature Achaia strike-slip fault zones are well resolved from modern strong events.

There are indications, from moderate-size events, for left-lateral strike-slip faulting along two zones: (a) one extending from the Gulf of Patras and Lake Trichonis up to Amfiloxia, following the NNE-trending Amfiloxia fault zone (Kiritzi et al. 2008) and (b) one extending from Pirgos in western Peloponnese up to Zakynthos Island (see Fig. 9 for locations). These two strands are not yet well documented from the occurrence of strong events.

Regarding the intermediate-depth ($h > 50$ km) earthquake focal mechanisms, these are all confined within the subducting Africa slab (see Fig. 4d). The azimuths of the P-axes are consistently parallel to the local trend of the Hellenic arc showing along-arc shortening. The azimuths of the T-axes are radial to the arc and are aligned with the dipping slab (Benetatos et al. 2004).

In conclusion, the earthquake focal mechanisms in the Aegean most of the times reveal remarkable uniformity, order, and kinematic links. They provide the most significant tool to study present-day tectonics and the key to study past tectonics.

Data Sources

The data used in this work have been all retrieved from published sources, especially for the stronger earthquakes. A number of focal mechanisms from smaller magnitude events are from the author's unpublished work. Most of the figures were produced using the GMT software (Wessel and Smith 1998).

Cross-References

- [Earthquake Mechanisms and Stress Field](#)
- [Earthquake Mechanism Description and Inversion](#)
- [Earthquake Mechanisms and Tectonics](#)
- [Long-Period Moment-Tensor Inversion: The Global CMT Project](#)
- [Moment Tensors: Decomposition and Visualization](#)
- [Non-Double-Couple Earthquakes](#)

References

- Basili R, Kastelic V, Demircioglu MB, Garcia Moreno D, Nemser ES, Petricca P, Sboras SP, Besana-Ostman GM, Cabral J, Camelbeeck T, Caputo R, Danciu L, Domac H, Fonseca J, García-Mayordomo J, Giardini D, Glavatovic B, Gulen L, Ince Y, Pavlides S, Sesetyan K, Tarabusi G, Tiberti MM, Utkucu M, Valensise G, Vanneste K, Vilanova S, Wössner J (2013) The European database of seismogenic faults (EDSF) compiled in the framework of the project SHARE. <http://diss.rm.ingv.it/share-edsf/>. doi:10.6092/INGV.IT-SHARE-EDSF
- Benetatos C, Kiritzi A, Papazachos C, Karakaisis G (2004) Focal mechanisms of shallow and intermediate depth earthquakes along the Hellenic arc. *J Geophys Res* 37(253–296):2004
- Chatzipetros A, Kiritzi A, Sboras S, Pavlides S (2013) Active faulting in the north-eastern Aegean Sea

- Islands. *Tectonophysics* 597:106–122. doi:10.1016/j.tecto.2012.11.026
- Copley A, Boait F, Hollingsworth J, Jackson J, McKenzie D (2009) Subparallel thrust and normal faulting in Albania and the roles of gravitational potential energy and rheology contrasts in mountain belts. *J Geophys Res* 114, B05407. doi:10.1029/2008JB005931
- Dreger DS (2003) TDMT_INV: time domain seismic moment tensor INVersion. *Int Handb Earthq Eng Seismol* 81B:1627
- Ekström G, Nettles M, Dziewonski AM (2012) The global CMT project 2004–2010: centroid-moment tensors for 13,017 earthquakes. *Phys Earth Planet Inter* 200–201:1–9. doi:10.1016/j.pepi.2012.04.002
- Goldsworthy M, Jackson J, Haines J (2002) The continuity of active fault systems in Greece. *Geophys J Int* 148(3):596–618. doi:10.1046/j.1365-246X.2002.01609.x
- Hatzfeld D, Martin C (1992) Intermediate depth seismicity in the Aegean defined by teleseismic data. *Earth Planet Sci Lett* 113:267–275
- Hatzfeld D, Pedotti G, Hatzidimitriou P, Makropoulos K (1990) The strain pattern in the western Hellenic arc deduced from a microearthquake survey. *Geophys J Int* 101:181–202
- Jackson J (2001) Living with earthquakes: know your faults. *J Earthq Eng* 5:5–123
- Jolivet L, Labrousse L, Agard P, Lacombe O, Bailly V, Lecomte E, Mouthereau F, Mehl C (2010) Rifting and shallow-dipping detachments, clues from the Corinth Rift and the Aegean. *Tectonophysics* 483:287–304
- Jolivet L, Faccenna C, Huet B, Labrousse L, Le Pourhiet L, Lacombe O, Lecomte E, Burov E, Denèle Y, Brun J-P, Philippon M, Paul A, Salaün G, Karabulut H, Piromallo C, Monié P, Gueydan F, Okay A, Oberhänsli R, Pourteau A, Augier R, Gadenne L, Driussi O (2013) Aegean tectonics: strain localisation, slab tearing and trench retreat. *Tectonophysics* 597–598:1–33
- Kiratzí A (2002) Stress tensor inversions along the westernmost north and central Aegean Sea. *Geophys J Int* 106:433–490
- Kiratzí A, Louvari E (2003) Focal mechanisms of shallow earthquakes in the Aegean Sea and the surrounding lands determined by waveform modelling: a new database. *J Geophys Res* 36:251–274
- Kiratzí A, Sokos E, Ganas A, Tselentis A, Benetatos C, Roumelioti Z, Serpetsidaki A, Andriopoulos G, Galanis O, Petrou P (2008) The April 2007 earthquake swarm near Lake Trichonis and implications for active tectonics in western Greece. *Tectonophysics* 452:51–65
- Kokkalas S, Kamberis E, Xypolias P, Sotiropoulos S, Koukouvelas I (2013) Coexistence of thin- and thick-skinned tectonics in Zakynthos area (western Greece): insights from seismic sections and regional seismicity. *Tectonophysics* 597–598:73–84
- Le Pichon X, Angelier J (1979) The Hellenic arc and trench system: a key to the neotectonic evolution of the eastern Mediterranean area. *Tectonophysics* 60(1–2):1–42. doi:10.1016/0040-1951(79)90131-8
- Louvari E, Kiratzí A and Papazachos B (1999) The Cephallonia Transform Fault and its continuation to western Lefkada Island. *Tectonophysics* 308:223–236
- Lyon-Caen H, Armijo R, Drakopoulos J, Baskoutas J, Delibasis N, Gaulon R, Kouskouna V, Latoussakis J, Makropoulos N, Papadimitriou P, Papanastasiou D, Pedotti G (1988) The 1986 Kalamata (South Peloponnese) earthquake: detailed study of a normal fault, evidences for east–west extension in the Hellenic arc. *J Geophys Res* 93:14967–15000
- McClusky S, Balassanian S, Barka A, Demir C, Georgiev I, Hamburg M, Hurst K, Kahle H, Kastens K, Kekelidze G, King R, Kotzev V, Lenk O, Mahmoud S, Mishin A, Nadariya M, Ouzounis A, Paradissis D, Peter Y, Prilepin M, Reilinger R, Sanli I, Seeger H, Tealeb A, Toksoz MN, Veis G (2000) Global positioning system constraints on plate kinematics and dynamics in the eastern Mediterranean and Caucasus. *J Geophys Res* 105:5695–5720
- McKenzie D (1978) Active tectonics of the Alpine-Himalayan belt: the Aegean Sea and surrounding regions. *Geophys J Roy Astron Soc* 55:217–254
- Müller M, Geiger A, Kahle H-G, Veis G, Billiris H, Paradissis D, Felekis S (2013) Velocity and deformation fields in the North Aegean domain, Greece, and implications for fault kinematics, derived from GPS data 1993–2009. *Tectonophysics* 597–598:34–49
- Papazachos BC, Comninakis PE (1971) Geophysical and tectonic features of the Aegean arc. *J Geophys Res* 76:8517–8533
- Papazachos BC, Papadimitriou EE, Kiratzí AA, Papazachos CB, Louvari E (1998) Fault plane solutions in the Aegean Sea and the surrounding area and their tectonic implication. *Boll Geof Teor App* 39:199–218
- Papazachos BC, Karakostas VG, Papazachos CB, Scordilis EM (2000) The geometry of the Wadati-Benioff zone and lithospheric kinematics in the Hellenic arc. *Tectonophysics* 319:275–300
- Pearce FD, Rondenay S, Sachpazi M, Charalampakis M, Royden LH (2012) Seismic investigation of the transition from continental to oceanic subduction along the western Hellenic subduction zone. *J Geophys Res* 117, B07306. doi:10.1029/2011JB009023
- Ring U, Glodny J, Will T, Thomson S (2010) The Hellenic subduction system: high-pressure metamorphism, exhumation, normal faulting, and large-scale extension. *Annu Rev Earth Planet Sci* 38:45–76. doi:10.1146/annurev.earth.050708.170910
- Royden LH, Papanikolaou DJ (2011) Slab segmentation and late Cenozoic disruption of the Hellenic arc. *Geochem Geophys Geosyst* 12:Q03010. doi:10.1029/2010GC003280
- Shaw B (2012) Active tectonics of the Hellenic subduction zone. Springer theses, vol 1. doi:10.1007/978-3-642-20804-1_1. © Springer, Berlin

- Sipkin SA (1994) Rapid determination of global moment-tensor solutions. *Geophys Res Lett* 21: 1667–1670
- Sokos E, Zahradnik J (2008) ISOLA a Fortran code and a Matlab GUI to perform multiple-point source inversion of seismic data. *Computers & Geosciences* 34:967–977. doi: 10.1016/j.cageo.2007.07.005
- Taymaz T, Jackson J, McKenzie D (1991) Active tectonics of the north and central Aegean Sea. *Geophys J Int* 106(2):433–490. doi:10.1111/j.1365-246X.1991.tb03906.x
- Tranos MD, Lacombe O (2014) Late Cenozoic faulting in SW Bulgaria: fault geometry, kinematics and driving stress regimes. Implications for late orogenic processes in the Hellenic hinterland. *J Geodyn* 74:32–55
- Tselentis G-A, Sokos E, Martakis N, Serpetsidaki A (2006) Seismicity and seismotectonics in Epirus, western Greece: results from a microearthquake survey. *Bull Seismol Soc Am* 96(5):1706–1717. doi:10.1785/0120020086
- Udias A, Buforn E, de Gauna R (1989) Catalogue of focal mechanisms of European earthquakes. Universidad Complutense de Madrid, Madrid, p 274, ISBN 84-600-721903
- Vassilakis E, Royden L, Papanikolaou D (2011) Kinematic links between subduction along the Hellenic trench and extension in the Gulf of Corinth, Greece: a multidisciplinary analysis. *Earth Planet Sci Lett* 303:108–120
- Wessel P, Smith WHF (1998) New improved version of the generic mapping tools released. *Eos Trans AGU* 79:579
- Zoback ML (1992) First- and second-order patterns of stress in the lithosphere: the world stress map project. *J Geophys Res* 97:11703–11728
- Zwack P, McCaffrey R, Abers G (1994) MT5 Program, IASPEI Software Library, 4

Mechanisms of Earthquakes in Iceland

Páll Einarsson

Institute of Earth Sciences, University of Iceland, Reykjavík, Iceland

Synonyms

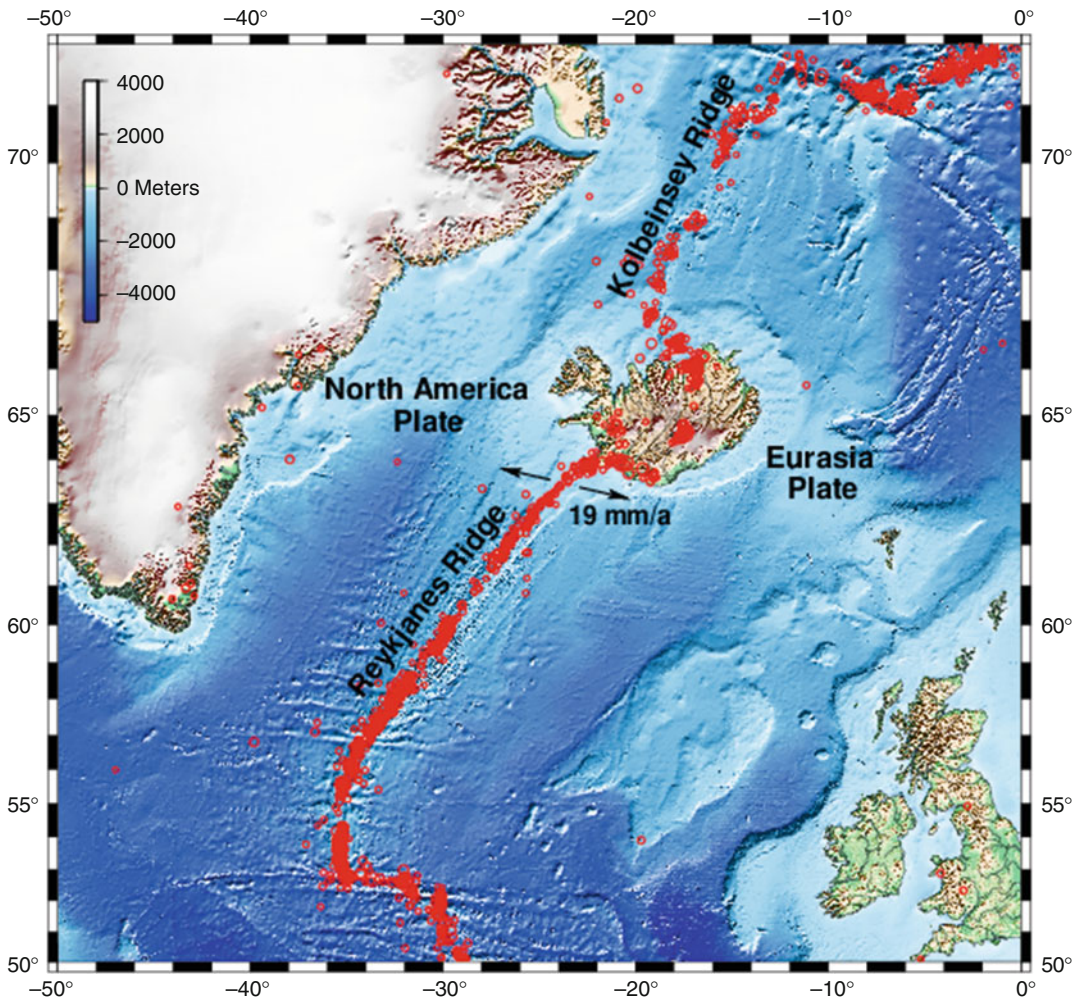
Bookshelf faulting; Divergent plate boundaries; Hreppar microplate; Seismicity of Iceland; South Iceland Seismic Zone; Tjörnes fracture zone; Transform zones; Volcanic earthquakes

Introduction

Most of the seismicity of Iceland is due to processes related to the mid-Atlantic plate boundary that crosses the country (Fig. 1). These processes include transform faulting, rifting, dike injection, inflation and deflation of magma chambers, cooling of magma bodies, and volume changes of geothermal fluids (Einarsson 1986, 1991, 2008; Sigmundsson 2006).

South of Iceland the seismically active belt follows the axis of the Reykjanes Ridge, and north of the country it follows the Kolbeinsey Ridge. On both ridges the seismicity decreases toward Iceland. This may be due to the existence of the Iceland hotspot, an area of high topography and heat flow, a result of excessive volcanism. The excessive volcanism produces a relatively thick crust of elevated temperature gradient. Combined with the spreading of the plates away from the plate boundary, the thick crust forms a complex of submarine highs extending between Scotland and Greenland (Fig. 1). The spreading velocity along the plate boundary is relatively low, less than 2 cm/year. Yet the topography of the ridge segments near Iceland, Reykjanes and Kolbeinsey Ridges, shows characteristics of fast-spreading ridges. Near its southern end, 53–59°N, on the other hand, the Reykjanes Ridge has the characteristics of a slow-spreading ridge, rough topography, deep rift valley, and relatively high seismicity. North of 59°, the fast-spreading characteristics become more prominent (Einarsson 2008). There is no direct evidence for faster spreading, however. Magnetic anomalies show spreading rates of 18–20 mm per year in a direction of 105°, corresponding to the distance to the Euler pole between the North America and Eurasia plates that is located in Siberia (Einarsson 2008; Sigmundsson 2006).

The plate boundary on Iceland is expressed by a series of seismic and volcanic zones. Two transform zones connect the presently active Northern and Eastern Volcanic Zones to the ridges offshore (Fig. 2). The transform zones host the largest earthquakes in the area, as large as magnitude 7, whereas the volcanic rift zones produce only smaller earthquakes, rarely exceeding magnitude 5.5.



Mechanisms of Earthquakes in Iceland, Fig. 1 Epicenters in Iceland and along the Mid-Atlantic Ridge system 1964–2004. Data are from the catalogs of the International Seismological Centre. Bathymetry is

from General Bathymetric Chart of the Oceans (Modified from Sólmes et al. (2013) with help from Gunnar B. Guðmundsson at the Icelandic Meteorological Office)

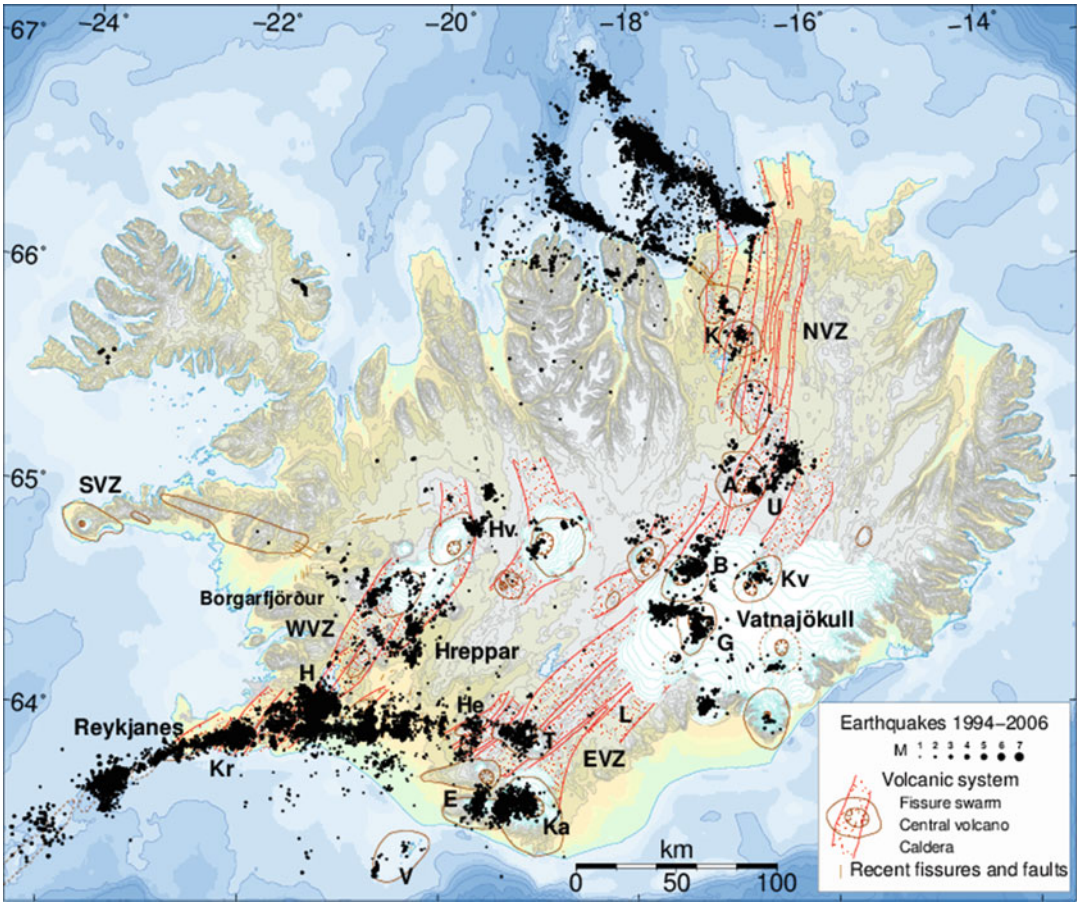
The structure of the plate boundary is strongly influenced by the Iceland mantle plume presumed to be centered beneath central Iceland. The plume is thought to have a deep root in the mantle, and its upwelling leads to the production of the magma that feeds the volcanoes of the hotspot. The relative motion of the Mid-Atlantic Ridge with respect to the plume leads to ridge jumps, propagating rifts and other complexities (e.g., Einarsson 1991, 2008). A new rift zone is initiated when the old rift has migrated some critical distance off the plume center. The transform zones are also

unstable and respond to changes in the configuration of the rifts.

Seismic and Tectonic Characteristics of Seismic Areas

Plates, Microplates, and Boundaries

The mid-Atlantic plate boundary is relatively simple in most parts of the NE Atlantic consisting of rifting and transform segments separating the two major plates, Eurasia and North America.



Mechanisms of Earthquakes in Iceland, Fig. 2 Volcanic systems of Iceland and epicenters located by the local seismic network in the period 1994–2006. The volcanic systems are from Einarsson and Sæmundsson (1987) and the seismic data from the Icelandic Meteorological Office. The fissure swarms of the volcanic systems delineate the divergent segments of the plate boundary, the Northern, Western, and Eastern Volcanic Zones (NVZ, WVZ, and EVZ), the last two of

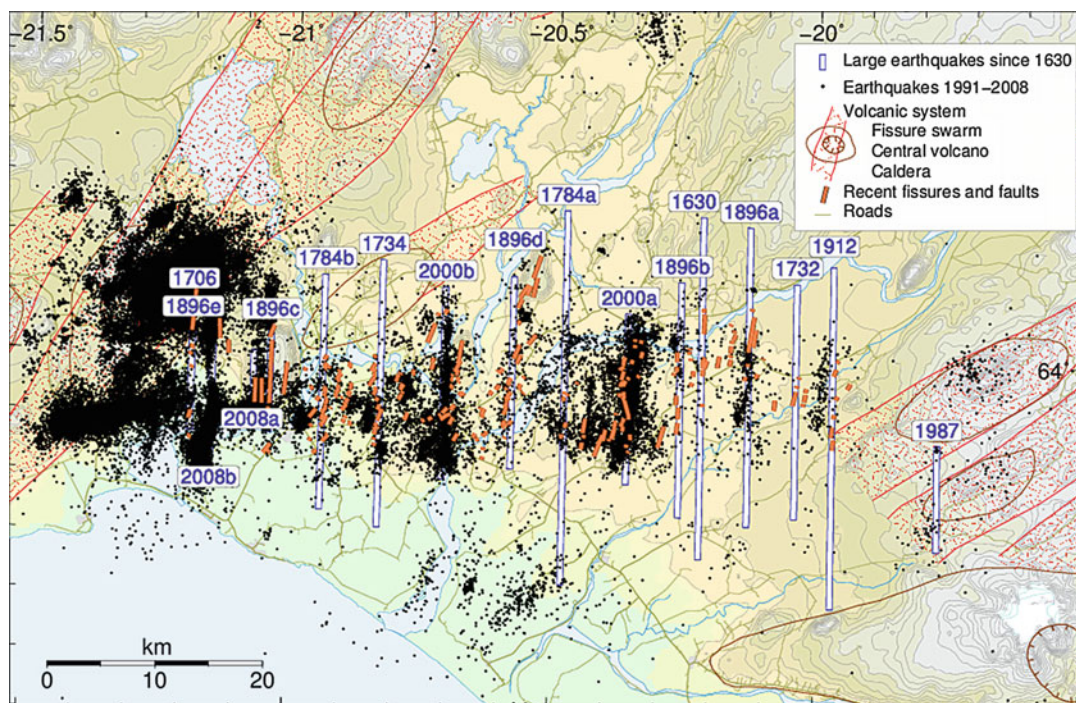
which outline the Hreppar Microplate. SVZ marks the Snæfellsnes Volcanic Zone and K, A, B, Kv, G, T, He, Ka, E, V, H, and Kr mark the Krafla, Askja, Bárðarbunga, Kverkfjöll, Grímsvötn, Torfajökull, Hekla, Katla, Eyjafjallajökull, Vestmannaeyjar, Hengill, and Krísuvík volcanic systems. Hveravellir, Laki, and Upptýppingar are shown with Hv, L, and U (Modified from Sólnes et al. (2013) with the help of GBG at IMO)

The boundary is clearly defined by the epicenters of earthquakes that show a narrow zone of deformation (Fig. 1). As it crosses the Iceland platform, however, the deformation zone becomes wider, as shown by the distribution of earthquakes and volcanism in Fig. 2. The boundary breaks up into a series of more or less oblique segments. Sometimes the boundary branches out, and small microplates or blocks are formed that may move independently of the large plates. The largest of these is the Hreppar Microplate that is

bounded by the two subparallel volcanic rift zones in South Iceland (Einarsson 2008; Sigmundsson et al. 1995; La Femina et al. 2005), the Western and Eastern Volcanic Zones (Fig. 2). Each branch or segment of the plate boundary has its own characteristics, both regarding tectonic style and seismicity.

South Iceland Seismic Zone

The southern boundary of the Hreppar Microplate, near 64°N, is a zone of high seismic



Mechanisms of Earthquakes in Iceland, Fig. 3 South Iceland Seismic Zone, epicenters, and suggested source faults of historical earthquakes. Epicenters for the time period 1991–2006 are from the Icelandic Meteorological Office, volcanic systems are from Einarsson and

Sæmundsson (1987), recent fissures and faults are from Einarsson (2010), and source faults of historical earthquakes (1630–1912) are based on fault mapping and investigation of historical documents (Modified from Sólnes et al. (2013) with the help of GBG at IMO)

M

activity, the South Iceland Seismic Zone (SISZ), which takes up most of the transform motion between the Reykjanes Ridge and the Eastern Volcanic Zone (Fig. 3). It has been argued that rifting is dying out in the Western Volcanic Zone and is being taken over by the Eastern Volcanic Zone (Einarsson 2008).

The majority of destructive, historical earthquakes in Iceland have taken place within the South Iceland Seismic Zone. The zone has been defined by destruction areas of historical earthquakes, Holocene surface ruptures, and instrumentally determined epicenters (Figs. 2 and 3). It is oriented E-W and is 10–15 km wide. Destruction areas of individual earthquakes and surface faulting show, however, that each event is associated with faulting on N-S striking planes, perpendicular to the main zone. The overall left-lateral transform motion across the zone thus appears to be accommodated by right-lateral

faulting on many parallel, transverse faults and counterclockwise rotation of the blocks between them, “bookshelf faulting” (Einarsson 1991, 2010).

Earthquakes in South Iceland tend to occur in major sequences in which most of the zone is affected (Einarsson et al. 1981). These sequences tend to last from a few days to about 3 years or even more. Each sequence typically begins with a magnitude 7 event in the eastern part of the zone, followed by smaller events farther west. Sequences of this type occurred in 1896, 1784, 1732–1734, 1630–1633, 1389–1391, 1339, and 1294. The sequences thus occur at intervals that range between 45 and 112 years, and it has been argued that a complete strain release of the whole zone is accomplished in about 140 years (Stefánsson et al. 1993). In addition to the sequences, single events occasionally occur near the ends of the zone, e.g., in 1912 at the eastern end.

The 1896 and 1912 earthquakes were followed by a long quiet interval, which in 1985 led to a long-term forecast of a major earthquake sequence within the next decades (Stefánsson et al. 1993). The forecast was fulfilled in June 2000 when two magnitude 6.5 events occurred in the central part of the zone. The sequence started on June 17 with a magnitude 6.5 event in the eastern part of the zone which was immediately followed by triggered activity along at least a 80 km long stretch of the plate boundary to the west (Árnadóttir et al. 2001, 2004). The second mainshock of about the same magnitude occurred 21 km west of the first one on June 21. It was clearly preceded by clustering of micro-earthquakes near the coming hypocenter (Stefánsson 2011). The mainshocks of the sequence occurred on N-S striking faults, transverse to the zone itself. The sense of faulting was right-lateral strike-slip consistent with the model of bookshelf faulting for the SISZ. The two mainshocks occurred on preexisting faults and were accompanied by surface ruptures expressed by en echelon tension gashes and push-up structures typical for strike-slip faults (Clifton and Einarsson 2005). The main zones of rupture were 15 km long, had a N-S trend, and coincided with the epicentral distributions. Faulting along conjugate faults was also observed, but was less pronounced than the main rupture zones. Faulting extended to a depth of about 10 km as shown by aftershock distribution and modeling of displacement fields at the surface determined by GPS and InSAR measurements (Árnadóttir et al. 2001; Pedersen et al. 2003). The maximum fault displacements according to these models were 2.5 m for both earthquakes.

The observed faulting structures were clearly smaller than those observed in association with earlier historical earthquakes such as those of 1630, 1784, 1896, and 1912. These earlier earthquakes were therefore larger than the earthquakes of June 2000, which is consistent with the measured magnitude of 7 for the 1912 earthquake.

High acceleration was recorded by a network of strong-motion accelerographs. The highest recorded acceleration was 0.83 g, recorded by an instrument in a bridge abutment within 3 km

of the source fault of the June 21 event. Indications of acceleration in excess of 1 g, such as overturned stones, were abundant in the source areas. In light of the high accelerations, the damage to man-made structures was surprisingly low. Only 20 houses were deemed unusable, none of them collapsed, however. Damage was strongly correlated with age of the buildings. There were no casualties and very little injury to humans or animals (Sigbjörnsson and Ólafsson 2004).

The earthquakes of 2000 only released about one fourth of the strain accumulated across the SISZ since the major sequence of 1896–1912. Further activity was therefore anticipated. A strong earthquake occurred near the western end of the zone, in the Ölfus district, on May 29, 2008, between the towns of Selfoss and Hveragerði (Fig. 3). The moment magnitude of the event was about 6.3, but it turned out to be a double event (Decriem et al. 2010; Sigbjörnsson et al. 2009). The first event took place on a fault at the W-foot of Mt. Ingólfssjall. It immediately triggered another slip event on a parallel fault, 4 km to the west. Both sub-events were associated with right-lateral strike-slip on N-S striking faults and had moment magnitudes (M_w) of 5.8 and 5.9, respectively. The surface effects of the earthquakes were locally very strong. High acceleration was seen in the epicentral areas, rock shattering occurred along many topographic protrusions, rockfalls were common on steep slopes, and sandblows were observed in the alluvium of Ölfusá river that crosses the affected area. Damage to houses was widespread even though no house collapsed. Injuries to people were minor.

Combining field knowledge and mapping of surface ruptures, historical documents, and experience from the 2000 and 2008 earthquakes, it is possible to identify with some certainty the source faults of some of the historical earthquakes since 1630 (Fig. 3, Table 1). Magnitudes of events prior to 1912 are estimated by Stefánsson and Halldórsson (1988). It is noteworthy that each earthquake appears to have occurred on separate fault, except possibly those of 1706 and 1896 in the western part of the zone. The repeat time of events on a particular fault therefore is generally longer than 400 years.

Mechanisms of Earthquakes in Iceland, Table 1 Earthquakes in the South Iceland Seismic Zone

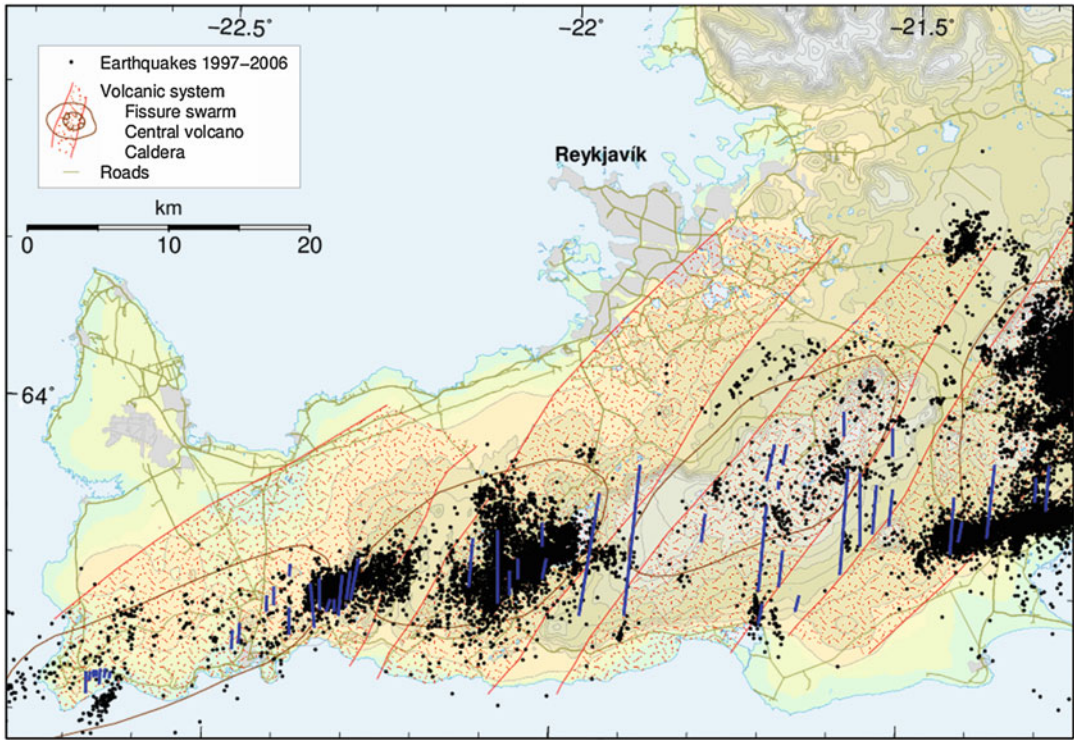
Year	Day	Latitude	Longitude	Magnitude	Area
1630	02 21	64.0	−20.21	~7	Land, Minnivellir
1633					Ölfus
1706	04 20	63.9	−21.20	6.0	Ölfus, Hveragerði
1732	09 07	64.0	−20.04	6.7	Land, Leirubakki ?
1734	03 21	63.9	−20.83	6.8	Flói, Litlureykir ?
1766	09 09	63.9	−21.16	6.0	Ölfus, Gljúfur – Kross – Kaldaðarnes
1784	08 14	63.9	−20.47	7.1	Holt – Gíslholtsvatn
1784	08 16	63.9	−20.95	6.7	Flói – Laugardælir
1829	02 21	63.9	−20.0	6.0	Rangárvellir – Hekla
1896	08 26	64.0	−20.13	6.9	Skarðsfjall – Fellsmúli
1896	08 27	64.0	−20.26	6.7	Flagbjarnarholt – Lækjarbotnar
1896	09 05	63.9	−21.04	6.0	Selfoss – Ingólfssfjall
1896	09 05	64.0	−20.57	6.5	Skeið – Arakot – Borgarkot
1896	09 06	63.9	−21.20	6.0	Ölfus – Hveragerði
1912	05 06	63.9	−19.95	7.0 M_S	Selsund – Galtalækur
1935	10 09	64.0	−21.37	6.0 M_L	Hellisheiði
1987	05 25	63.9	−19.78	5.9 M_w	Vatnafjöll
2000	06 17	63.97	−20.35	6.5 M_w	Holt, Skammbeinsstaðir
2000	06 21	63.97	−20.70	6.4 M_w	Flói, Grímsnes, Hestvatn
2008	05 29	63.9	−21.09	5.8 M_w	Ölfus, Ingólfssfjall
2008	0529	63.9	−21.16	5.9 M_w	Ölfus, Kross

Reykjanes Peninsula Oblique Rift

The mid-Atlantic plate boundary goes onshore at the SW tip of the Reykjanes Peninsula and extends from there with a trend of 70° (azimuth) along the whole peninsula (Figs. 2 and 4). The rift zone is highly oblique and is relatively homogeneous until it joins with the Western Volcanic Zone and the South Iceland Seismic Zone at the Hengill triple junction near 64°N and 21.4°W. The most prominent structural elements are extensional, i.e., hyaloclastite ridges (tindar) formed by fissure eruptions beneath the Pleistocene glacier, postglacial eruptive fissures, normal faults, and open fissures. These structures trend mostly highly obliquely to both the plate boundary and the plate velocity vector (Fig. 4). The average trend of several of the most prominent features is 35° (Einarsson 2008). The fissures and normal faults are grouped into swarms, 4–5 of which can be identified and are shown in Fig. 4. Less conspicuous, but probably equally important, are N-S strike-slip faults that cut across the plate boundary at a high angle. They are expressed as N-S trending arrays of left-stepping,

en echelon fissures with push-up hillocks between them. Typical spacing between the faults is 1 km but in some areas it may be as small as 0.4 km. The N-S strike-slip faults may act to accommodate the transcurrent component along the plate boundary in the same way as is explained for the South Iceland Seismic Zone by the bookshelf faulting model.

Seismicity on the Reykjanes Peninsula is high and appears to be episodic in nature with active episodes occurring every 30 years or so. Recent active periods were in the beginning of the 20th century, 1929–1935, 1967–1973, and in 2000. The largest earthquakes in the latest two episodes (1967–1973 and since 2000) are known to have been associated with strike-slip faulting (e.g., Einarsson 1991; Árnadóttir et al. 2004). At the present time, therefore, deformation along the plate boundary appears to be accommodated by strike-slip faulting, possibly with some contribution of crustal stretching. No evidence of magmatic contribution has been detected in the last decades, such as inflation sources, volcanic tremor, or earthquake swarms propagating along



Mechanisms of Earthquakes in Iceland, Fig. 4 Reykjanes Peninsula Oblique Rift, fissure swarms, epicenters of 1997–2006, and identified strike-slip faults shown in blue. Fissure swarms are from Einarsson and Sæmundsson (1987), seismic data are from the Icelandic

Meteorological Office, and the strike-slip faults are based on Clifton and Kattenhorn (2006) to a large extent. The four volcanic systems shown are Reykjanes, Krisuvík, Brennisteinsfjöll, and Hengill (Modified from Sólness et al. (2013) with the help of GBG at IMO)

the fissure swarms. Magmatic activity within this segment also appears to be highly episodic, but with a much longer period, of the order of 1000 years. The latest eruptive period ended in 1240 AD. During this most recent episode, all the volcanic systems of the peninsula were active with several lava eruptions issuing from their fissure swarms (see, e.g., Sólness et al. 2013, pp. 379–401).

Crustal deformation along the plate boundary on the Reykjanes Peninsula appears to occur in two different modes: (1) dry or seismic mode and (2) wet or magmatic mode (Hreinsdóttir et al. 2001). Deformation in the dry mode occurs mostly by strike-slip faulting during periods when magma is not available to the crust in any appreciable quantity. The mode of deformation changes when magma enters the crust. Dikes are opened up against the local minimum

compressive stress, and they propagate into the plates on both sides of the plate boundary and the fissure swarms are activated.

Known significant earthquakes of the Reykjanes Peninsula Oblique Rift are summarized in Table 2. It is noteworthy that following the eruptive period that ended in 1240 AD, no earthquakes are mentioned until 1724. This could be the result of poor documentation, but it is also possible that diking during the magmatic episode left this plate boundary segment completely de-stressed. It then took 500 years of plate movements to build up sufficient stress to produce earthquakes again.

Tjörnes Fracture Zone

The transform motion between the offshore Kolbeinsey Ridge and the Northern Volcanic Rift Zone of Iceland is taken up by the Tjörnes

Mechanisms of Earthquakes in Iceland, Table 2 Earthquakes in the Reykjanes Peninsula Oblique Rift

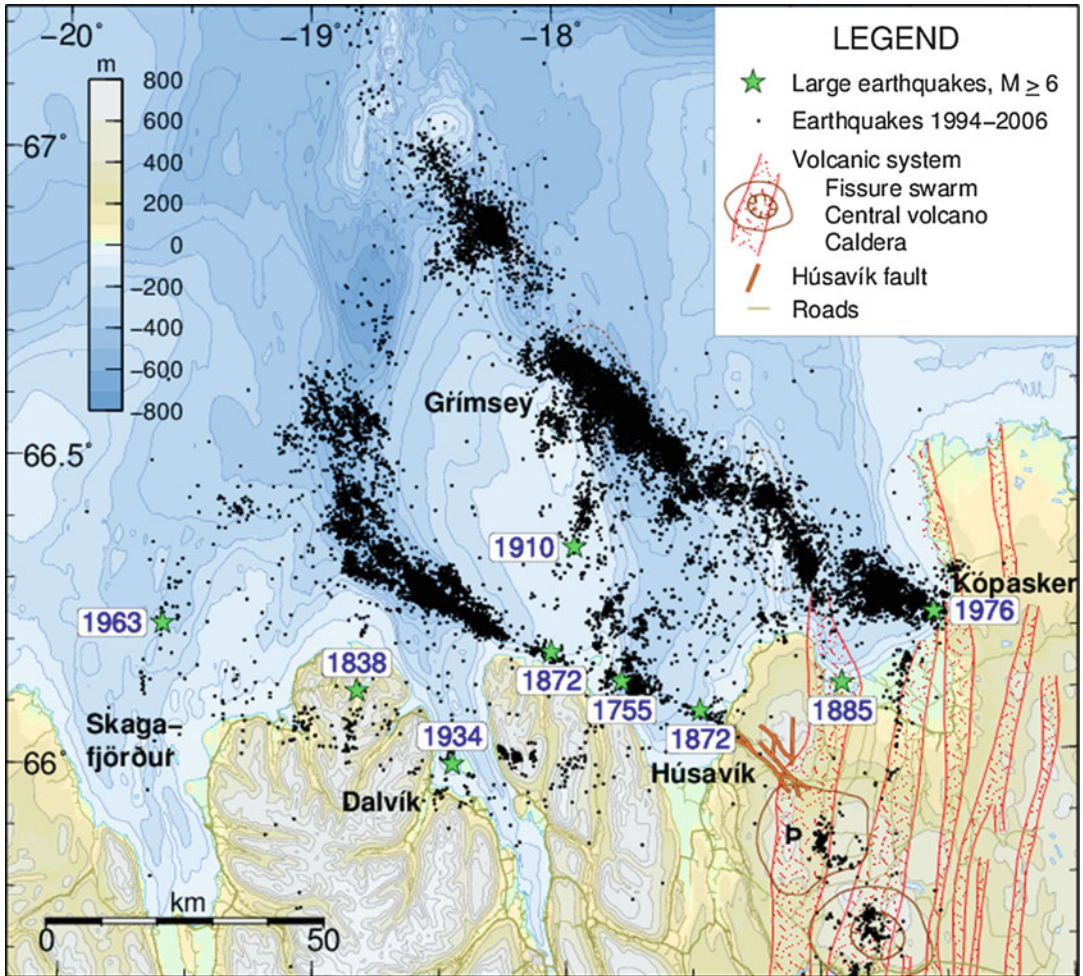
Year	Day	Latitude	Longitude	Magnitude	Area
1211		63.8	−22.8		Reykjanes
1724	08	63.9	−22.0		Krísuvík – Geitafell ?
1929	07 23	63.9	−21.70	6.3 M _S	Heiðin há, Hvalhnúkur
1933	06 10	63.9	−22.2	6.0 M _S	Núpslíðarháls
1935	10 09	64.0	−21.37	6.0 M _S	Hellisheiði
1968	12 05	63.9	−21.70	6.0 M _S	Heiðin há, Hvalhnúkur
1973	09 15	63.9	−22.1	5.4 M _S	Móhalsadalur
1973	09 16	63.9	−22.3	5.2 m _b	Fagradalsfjall
1998	06 04	64.0	−21.28	5.4 M _w	Hellisheiði
1998	11 13	63.95	−21.34	5.1 M _w	Ölfus, Hjalli
1998	11 14	63.96	−21.24	4.5 M _S	Ölfus, Hjalli
2000	06 17	63.96	−21.70	5.5 M _w	Hvalhnúkur
2000	06 17	63.95	−21.95	5.9 M _w	Kleifarvatn
2000	06 17	63.90	−22.12	5.3 M _w	Núpslíðarháls
2003	08 23	63.9	−22.06	5.2 M _w	Krísuvík
2013	10 13	63.81	−22.66	5.2 M _w	Reykjanes

Fracture Zone, a broad zone of seismicity, strike-slip faulting, and crustal extension (Einarsson 1991; Stefánsson et al. 2008). The zone consists of several subparallel NW-trending seismic zones, termed the Grímsey Oblique Rift, the Húsavík-Flatey Faults, and the Dalvík Seismic Zone (Einarsson 2008), and a few northerly trending rift or graben structures (Fig. 5). Recent GPS measurements indicate that about $\frac{3}{4}$ of the plate transform motion is taken up by the Grímsey Oblique Rift, $\frac{1}{4}$ by the Húsavík-Flatey Faults, and an insignificant part by the Dalvík Zone (Metzger et al. 2013). Yet, all the zones have produced significant and damaging earthquakes in the last 150 years.

The Grímsey Oblique Rift is the northernmost seismic zone and is entirely offshore. It has an overall NW-SE trend but is composed of several volcanic systems with N-S trend. Evidence for Holocene volcanism is abundant on the sea-floor. At its SE end, the zone connects to the Krafla fissure swarm, but the NW end joins the Kolbeinsey Ridge just south of the rapidly disappearing Kolbeinsey Island. The larger earthquakes of the Grímsey Zone seem to be associated with left-lateral strike-slip faulting on NNE-striking faults (Rögnvaldsson et al. 1998). The zone has the characteristics of an oblique rift. There is a striking similarity between the

Reykjanes Peninsula Oblique Rift and the Grímsey Zone. The two zones are symmetrical with respect to the plate separation vector. They are nearly perfect mirror images of each other, with the axis of symmetry parallel to the plate separation vector. The similarity is seen in the overall trend, the en echelon fissure swarms, the transverse bookshelf faults, and the occurrence of geothermal areas. Earthquake swarms are common in both zones.

The second seismic zone, the **Húsavík-Flatey Faults**, is about 40 km south of the Grímsey Zone and is well defined by the seismicity near its western end, where it joins with the Eyjafjarðaráll Rift (E in Fig. 5). This transform zone can be traced on the ocean bottom to the coast in the Húsavík town, continuing on land into the volcanic zone, where it merges into one of its fissure swarms (Fig. 5). This is a highly active seismic zone. The last major earthquakes occurred in 1872 when the town of Húsavík suffered heavy damage during an earthquake sequence, including at least two M 6.5 events. Well-developed faults can be seen along the whole length of this zone. They show evidence of right-lateral faulting and are subparallel to the zone itself. No evidence of bookshelf faulting has been found so far. Volcanism only plays a subordinate role if any, in this zone.



Mechanisms of Earthquakes in Iceland, Fig. 5 Tjörnes Fracture Zone, epicenters 1994–2006, and suggested sources of historical earthquakes since 1755. Fissure swarms are from Einarsson and Sæmundsson (1987), seismic data are from the Icelandic

Meteorological Office, and the sources of the historical events are from Stefánsson et al. (2008). E marks the Eyjafjarðaráll Graben and P the Peistareykir volcanic system (Modified from Sólnes et al. (2013) with the help of GBG at IMO)

A third zone, the **Dalvík Zone**, is indicated by scattered seismicity about 30 km south of the Húsavík-Flatey Faults (Fig. 5). Earthquakes as large as M_s 7 have occurred in this zone, but it lacks clear topographic expression. In spite of rather clear alignment of epicenters, the Dalvík Zone is not seen as a throughgoing fault on the surface. A common feature of all three seismic zones is the occurrence of earthquakes on transverse structures, similar to that observed in the South Iceland Seismic Zone and the Reykjanes

Peninsula. This is clearly seen in the Dalvík Zone (Fig. 5). In spite of apparently slow strain accumulation in this zone, it has produced two significant earthquakes in the last century, the Dalvík earthquake of 1934 (M 6.3) that caused considerable damage in the town of Dalvík and the 1963 Skagafjörður earthquake (M 7.0) offshore in the mouth of Skagafjörður, one of the largest earthquakes in Iceland in the last century.

Large earthquakes of the Tjörnes Fracture Zone during the last centuries are summarized

Mechanisms of Earthquakes in Iceland, Table 3 Earthquakes in the Tjörnes Fracture Zone

Year	Day	Latitude	Longitude	Magnitude	Area
1755	09 11	66.13	–17.76	~7	Húsavík-Flatey
1838	06 12	66.12	–18.82	6.5	Fljót, Dalvík Zone
1872	04 18	66.08	–17.45	6.5	Húsavík
1872	04 18	66.18	–18.04	6.5	Flatey
1885	01 25	66.12	–16.88	6.3	Kelduhverfi
1910	01 22	66.35	–17.94	7.0	Grímsey
1921	08 23	67	–18	6.3	
1934	06 02	66.00	–18.44	6.2	Dalvík
1963	03 27	66.23	–19.60	7.0 M _S	Skagafjörður, Dalvík Zone
1976	01 13	66.23	–16.50	6.4 M _S	Kópasker, Grímsey Oblique Rift

in Table 3, based mainly on Stefánsson et al. (2008). For the pre-instrumental era, i.e., prior to 1960, the location of the epicenters is uncertain. Most of the zone is offshore and the felt areas are therefore poorly constrained. In addition, some of the events may belong to a series of triggered events, similar to recent events in South Iceland. The felt reports may therefore belong to different events within the sequence that will further confuse the issue. In addition to the large historical earthquakes in Table 3, several sequences and swarms containing events as large as magnitude 5.6 have occurred in recent decades, along the GOR in 1967–1969 and 2013, and in Eyjafjarðaráll-HFF in 2012.

The focal mechanisms have only been instrumentally determined for the 1963 and 1976 events. Both are strike-slip mechanisms on ESE-striking (right-lateral) or NNE-striking (left-lateral) fault planes. The aftershock distributions (see Fig. 5) would favor the ESE plane as the fault plane for the 1976 event, but the NNE plane for the 1963 event. It may be inferred that the other earthquakes are also caused by strike-slip faulting, except possibly the 1885 earthquake. It occurred within the fissure swarm of the Peistareykir volcanic system and may have been related to rifting activity. From the distribution of background seismicity, it may be inferred that the 1755 and 1872 earthquakes took place on the main Húsavík-Flatey Faults, whereas the 1838, 1910, 1934, and 1963 earthquakes occurred on bookshelf-type faults, transverse to the seismic zones.

Volcanic Zones

Earthquakes occur in the volcanic zones of Iceland but are rarely large (e.g., Pedersen et al. 2007). Events larger than magnitude 5 are only known in connection with unusually large events such as during the Krafla rifting episode in 1975–1984 and at Bárðarbunga volcano prior to the Gjálp eruption of 1996. Events of magnitude 6 and larger have so far not been recorded. It should be noted, however, that large earthquakes have sometimes occurred in the transform zones of South and North Iceland following or in connection with major rifting events in the adjacent divergent parts of the plate boundary. The best documented case is the Kópasker earthquake of TFZ in 1976 during the first rifting event of Krafla (Einarsson 1991). Other cases are the SISZ earthquakes of 1784 following the Laki eruption of 1783, the 1872 earthquakes of the Húsavík-Flatey Faults in connection with an eruption 1867 offshore and the Askja rifting episode of 1874–1875, and the SISZ earthquakes of 1732–1734 following an unusually active period in the rift zones.

Inspection of Fig. 2 reveals a low-magnitude background seismicity of several of the central volcanoes. Apparent spatial correlation with high-temperature geothermal areas suggests that this activity is the expression of heat mining of magmatic sources within these volcanoes. The cooling of hot rock bodies leads to contraction, stress changes, and fracturing. The resulting earthquakes are expected to have a large non-double-couple component in their source mechanism, which has been confirmed in the

case of the Hengill and Krafla volcanoes (Miller et al. 1998). Other low-level seismicity sources that may have similar origin include Peistareykir, Askja (SE caldera part), Kverkfjöll, Grímsvötn, Hveravellir, Torfajökull, Katla (caldera part), Krísuvík, and Reykjanes.

Some of the volcanoes have been shown to go through inflation-deflation cycles during limited times, i.e., a magma chamber is inflated by inflowing magma and subsequently deflates, either by eruption to the surface or by intrusion of a dike. The best documented case is that of Krafla during the rifting episode of 1975–1984, but also Grímsvötn has shown this behavior since 1997 and Hekla during the last decades (Sturkell et al. 2006). Inflation and deflation of a magma chamber leads to stress changes in the chamber roof and eventually to earthquakes if the stress change is large enough. Krafla and Grímsvötn have shown seismicity that conforms with this mechanism, but the Hekla chamber is deep-seated. It is apparently below the brittle-ductile transition and has not produced stress changes large enough to cause brittle failure during recent cycles. The largest earthquakes of this origin have reached magnitude 5, during deflation of Krafla in 1976 (Einarsson 1991).

The brittle-ductile transition is shown to be at about 8–9 km depth at the constructive plate boundaries in Iceland but becomes deeper with increasing age (e.g., Stefánsson et al. 1993). Occasionally earthquakes have been detected at considerably larger depth in the volcanic zones, indicating brittle failure of crustal material that would be ductile at normal strain rates, i.e., strain rates in the crust associated with plate movements. In the case of the 1973 Heimaey (Vestmannaeyjar) and the 2010 Eyjafjallajökull eruptions, these “deep” events occurred in connection with obvious magmatic movements and high strain rates. In other cases the magmatic connection is inferred, mainly because high strain rate is the only known mechanism to cause brittle failure at abnormal depth. A dike intrusion at 15–30 km depth at Upptyppingar in the Northern Volcanic Zone in 2007–2008 was accompanied by an earthquake swarm and verified by surface deformation measurements (e.g., Hooper

et al. 2011). Persistent deep earthquake activity in the Askja volcano area at depths of 15–34 km has been interpreted as the expression of magma movements in the lower crust (Soosalu et al. 2009), similarly deep events in the Katla area (Jakobsdóttir 2008).

A large sequence of magmatic rifting events occurred in the Northern Volcanic Zone, beginning in 1974 and lasting until 1989, with diking and eruptions in the interval 1975–1984. The activity was mainly concentrated on the Krafla volcanic system but also extended into the adjacent Grímsey Oblique Rift, a part of the Tjörnes Fracture Zone. Most of this time, magma apparently ascended from depth and accumulated in a magma chamber at about 3 km depth beneath the Krafla volcano. The inflation periods were punctuated by sudden deflation events lasting from several hours to 3 months when the walls of the chamber were breached and magma was injected into the adjacent fissure swarm accompanied by propagating earthquake swarms (Wright et al. 2012). Twenty deflation events were documented. Large-scale rifting was observed in the fissure swarm during these deflation events, and in nine of them, magma reached the surface in basaltic fissure eruptions lasting from a half hour to 14 days. The earthquakes accompanying the diking were smaller than magnitude 4.5, except the triggered events in the adjacent Tjörnes Fracture Zone (Einarsson 1986, 1991).

A unique earthquake sequence occurred at the Bárðarbunga volcano in Central Iceland during the period 1974–1996. Fifteen earthquakes of magnitude 5 and larger occurred in the caldera region at fairly regular intervals during this period. The focal mechanisms were characterized by reverse faulting but with a significant non-double-couple component (Einarsson 1991; Ekström 1994). Each one of the large events was preceded by a few months of increasing seismic activity, but followed by quiescence or only very few earthquakes. This behavior is highly anomalous. Most other large earthquakes are followed by months or years of high aftershock activity that gradually decays with time. The last large event of the sequence occurred on September

29, 1996 ($M_w = 5.6$). In contrast to previous events, it was immediately followed by an intense earthquake swarm in the summit region of Bárðarbunga. The earthquakes migrated to the SE and S, and in the evening of September 30, they were replaced on the seismic records by continuous volcanic tremor signifying the beginning of an eruption. The eruption is termed the Gjálp eruption and was one of the larger eruptions of the twentieth century in Iceland (Einarsson et al. 1997; Guðmundsson et al. 1997). It occurred on a 7 km long fissure SE of the caldera and beneath the Vatnajökull glacier. It resulted in a catastrophic flood of about 4 km³ of meltwater.

Several interpretations of the cause of the Bárðarbunga earthquakes have been proposed. The one favored here assumes that they are the result of pressure drop in a magma chamber beneath the caldera of Bárðarbunga (Einarsson 1991). This drop began prior to 1974 and continued until the plumbing of the volcano changed in 1996. The earthquakes were, according to this interpretation, generated by piecemeal collapse of the caldera, by reverse faulting along the curved caldera main fault. The curvature produced the non-double-couple component of the focal mechanism. In the final event in 1996, a pocket of magma was cut by the fault, leading to the formation of a ring dike and triggering a dike propagating out of the caldera to feed the Gjálp eruption.

Intraplate Earthquakes

Even though earthquakes are rare outside the plate boundary areas of Iceland, the class of intraplate earthquakes is of considerable importance for the assessment of seismic risk. The main reason is that the population of Iceland is highly clustered, with half the population concentrated in the Reykjavík metropolitan area, 10–30 km off the plate boundary. Estimating correctly the probability of a large earthquake within the city is a fundamental issue.

Considering Fig. 2 and an earlier analysis of intraplate earthquakes in Iceland (Einarsson 1989), three classes of intraplate earthquakes can be identified:

1. *Earthquakes in West Iceland.* Earthquakes occur west of the main rift zone through Iceland, within the tongue of lithosphere that is bounded by the two transform zones in North and South Iceland. This seismicity is associated with internal deformation of the North American Plate. Best known are the Borgarfjörður events of 1974 in Central West Iceland (Fig. 2), an extended sequence that culminated with an earthquake of m_b 5.5. They were associated with normal faulting on NE- or E-striking faults. Crustal extension in this area may be linked with the center of the Iceland hotspot and its track into the North America Plate. The Snæfellsnes Volcanic Zone may be another expression of this extension (Fig. 2). It is a zone of recent volcanic activity laying unconformably on basalts of 6 Ma age. Distributed, low-level seismicity is seen in Fig. 2 connecting the Snæfellsnes Volcanic Zone to the Western Volcanic Zone.
2. *Earthquakes at the insular shelf edge off Eastern and SE Iceland.* Iceland sits on a platform bounded by a relatively sharp edge that resembles a continental shelf edge (see Fig. 1). It is particularly sharp off the E and SE coast, and there it is seismically active. The edge in this area represents a discontinuity in crustal and mantle structure related to eastward jumps in the plate boundary. There is probably an age jump of 15–25 m.y. and a resulting thermal discontinuity across the shelf edge. The present seismicity is thus interpreted as the result of differential lithospheric response to loading or different cooling rate on either side of the edge.
3. *Earthquakes under and around the Vatnajökull glacier.* Small earthquakes are occasionally seen in the area of Vatnajökull glacier, the largest ice mass of Iceland. The ice is shrinking due to the global warming of the climate. This leads to crustal uplift to restore isostatic equilibrium. The area around the glacier has been shown to rise 10–25 mm/year in recent years (Árnadóttir et al. 2009), and it is a probable cause of this low-level seismic activity.

Summary

Most earthquakes in Iceland are causally linked with the mid-Atlantic plate boundary that crosses the country. The largest earthquakes occur in the two major transform zones, one in South Iceland and the other in North Iceland, that link the subaerial volcanic rift zones with the spreading ridge segments offshore. These earthquakes are caused by strike-slip faulting and may reach magnitude of 7. The transform zones are rather complicated structures, and in many cases the faulting takes place on several parallel strike-slip faults that strike almost perpendicularly to the plate boundary, a faulting mode termed “bookshelf faulting.” Triggering of earthquakes has been observed, both by other earthquakes and by rifting in an adjacent zone. Earthquakes in the volcanic and divergent zones of Iceland are generally much smaller. Events of magnitude 5–6 have been found to occur within volcanic calderas. In the case of Krafla in 1975–1976, they were associated with deflation of the volcano’s magma chamber. Deflation is also suspected to be the cause of a persistent series of events in the Bárðarbunga caldera in 1974–1996. The focal mechanisms were of the reverse faulting type with a significant non-double-couple component. Dike intrusions and other magmatic processes are responsible for some earthquake activity, mostly at lower magnitudes still. Intraplate earthquakes have been documented in several cases. In West Iceland normal faulting was found to be responsible for a sequence of events in 1974, culminating in an earthquake of magnitude 5.5. The shelf edge off the SE, E, and NE coast is found to be seismically active, and the present isostatic uplift around the shrinking Icelandic glaciers is probably responsible for occasional events there.

Cross-References

- [Earthquake Magnitude Estimation](#)
- [Earthquake Mechanisms and Stress Field](#)
- [Earthquake Mechanisms and Tectonics](#)
- [Earthquake Swarms](#)
- [Frequency-Magnitude Distribution of Seismicity in Volcanic Regions](#)

- [InSAR and A-InSAR: Theory](#)
- [Non-Double-Couple Earthquakes](#)
- [Paleoseismology](#)
- [Seismic Monitoring of Volcanoes](#)

References

- Árnadóttir P, Hreinsdóttir S, Guðmundsson G, Einarsson P, Heinert M, Völkens C (2001) Crustal deformation measured by GPS in the South Iceland Seismic Zone due to two large earthquakes in June 2000. *Geophys Res Lett* 28:4031–4033
- Árnadóttir P, Geirsson H, Einarsson P (2004) Coseismic stress changes and crustal deformation on the Reykjanes Peninsula due to triggered earthquakes on June 17, 2000. *J Geophys Res* 109, B09307. doi:10.1029/2004JB003130
- Árnadóttir P, Lund B, Jiang W, Geirsson H, Björnsson H, Einarsson P, Sigurdsson P (2009) Glacial rebound and plate spreading: results from the first countrywide GPS observations in Iceland. *Geophys J Int* 177(2):691–716. doi:10.1111/j.1365-246X.2008.04059.x
- Clifton A, Einarsson P (2005) Styles of surface rupture accompanying the June 17 and 21, 2000 earthquakes in the South Iceland Seismic Zone. *Tectonophysics* 396:141–159
- Clifton A, Kattenhorn S (2006) Structural architecture of a highly oblique divergent plate boundary segment. *Tectonophysics* 419:27–40
- Decrie J, Árnadóttir P, Hooper A, Geirsson H, Sigmundsson F, Keiding M, Ófeigsson BG, Hreinsdóttir S, Einarsson P, LaFemina P, Bennett RA (2010) The 2008 May 29 earthquake doublet in SW Iceland. *Geophys J Int*. doi:10.1111/j.1365-246X.2010.04565.x
- Einarsson P (1986) Seismicity along the eastern margin of the North American Plate. In: Vogt PR, Tucholke BE (eds) *The geology of North America, vol M. The Western North Atlantic Region*. Geological Society of America, Boulder, Colorado, pp 99–116
- Einarsson P (1989) Intraplate earthquakes in Iceland. In: Gregersen S, Basham PW (eds) *Earthquakes at North-Atlantic passive margins: neotectonics and postglacial rebound*. Kluwer Academic, Dordrecht, pp 329–341
- Einarsson P (1991) Earthquakes and present-day tectonism in Iceland. *Tectonophysics* 189:261–279
- Einarsson P (2008) Plate boundaries, rifts and transforms in Iceland. *Jökull* 58:35–58
- Einarsson P (2010) Mapping of Holocene surface ruptures in the South Iceland Seismic Zone. *Jökull* 60:121–138
- Einarsson P, Björnsson S, Foulger GR, Stefánsson R, Skaftadóttir P (1981) Seismicity pattern in the South Iceland seismic zone. In: Simpson D, Richards P (eds) *Earthquake prediction – an international review*. American Geophysical Union, Maurice Ewing Series, vol 4, pp 141–151

- Einarsson P, Sæmundsson K (1987) Earthquake epicenters 1982–1985 and volcanic systems in Iceland. In: Sigfússon ÞÍ (ed) Í hlutarins eðli, Festschrift for Þorbjörn Sigurgeirsson (Map). Menningarsjóður, Reykjavík
- Einarsson P, Brandsdóttir B, Guðmundsson MT, Björnsson H, Grönvold K, Sigmundsson F (1997) Center of the Iceland hotspot experiences volcanic unrest. *Eos* 78:369–375
- Ekström G (1994) Anomalous earthquakes on volcano ring-fault structures. *Earth Planet Sci Lett* 128:707–712
- Guðmundsson MT, Sigmundsson F, Björnsson H (1997) Ice-volcano interaction of the 1996 Gjálp subglacial eruption, Vatnajökull, Iceland. *Nature* 389:954–957
- Hooper A, Ofeigsson BG, Sigmundsson F, Lund B, Geirsson H, Einarsson P, Sturkell E (2011) Increased capture of magma in the crust promoted by ice-cap retreat in Iceland. *Nat Geosci*. doi:10.1038/NCEO1269
- Hreinsdóttir S, Einarsson P, Sigmundsson F (2001) Crustal deformation at the oblique spreading Reykjanes Peninsula, SW Iceland: GPS measurements from 1993 to 1998. *J Geophys Res* 106:13803–13816
- Jakobsdóttir S (2008) Seismicity in Iceland: 1994–2007. *Jökull* 58:75–100
- La Femina PC, Dixon TH, Malservisi R, Árnadóttir Þ, Sturkell E, Sigmundsson F, Einarsson P (2005) Geodetic GPS measurements in south Iceland: strain accumulation and partitioning in a propagating ridge system. *J Geophys Res* 110, B11405. doi:10.1029/2005JB003675
- Metzger S, Jónsson S, Danielsen G, Hreinsdóttir S, Jouanne F, Giardini D, Villemin T (2013) Present kinematics of the Tjörnes Fracture Zone, North Iceland, from campaign and continuous GPS measurements. *Geophys J Int* 192:441–455
- Miller A, Foulger GR, Julian B (1998) Non-double-couple earthquakes. 2. Observations. *Rev Geophys* 36:551–568
- Pedersen R, Jónsson S, Árnadóttir Þ, Sigmundsson F, Feigl K (2003) Fault slip distribution of two June 2000 M_w 6.4 earthquakes in South Iceland estimated from joint inversion of InSAR and GPS measurements. *Earth Planet Sci Lett* 213:487–502
- Pedersen R, Sigmundsson F, Einarsson P (2007) Controlling factors on earthquake swarms associated with magmatic intrusions; constraints from Iceland. *J Volcanol Geotherm Res* 162:73–80. doi:10.1016/j.jvolgeores.2006.12.010
- Rögnvaldsson S, Guðmundsson A, Slunga R (1998) Seismotectonic analysis of the Tjörnes Fracture Zone, an active transform fault in north Iceland. *J Geophys Res* 103:30117–30129
- Sigbjörnsson R, Ólafsson S (2004) On the South Iceland earthquakes in June 2000: strong-motion effects and damage. *Boll Geof Teor Appl* 45:131–152
- Sigbjörnsson R, Snæbjörnsson JP, Higgins SM, Halldórsson B, Ólafsson S (2009) A note on the M_w 6.3 earthquake in Iceland on 29 May 2008 at 15:45 UTC. *Bull Earthq Eng* 7:113–126. doi:10.1007/s10518-008-9087-0
- Sigmundsson F (2006) Iceland geodynamics, crustal deformation and divergent plate tectonics. Springer – Praxis, Chichester
- Sigmundsson F, Einarsson P, Bilham R, Sturkell R (1995) Rift-transform kinematics in south Iceland: deformation from Global Positioning System measurements, 1986 to 1992. *J Geophys Res* 100:6235–6248
- Sólnes J, Sigmundsson F, Bessason B (eds) (2013) Natural hazards in Iceland: volcanic and seismic hazard (in Icelandic). University of Iceland Press, Reykjavík
- Soosalu H, Key AJ, White RS, Knox C, Einarsson P, Jakobsdóttir SS (2009) Lower-crustal earthquakes caused by magma movement beneath Askja volcano on the north Iceland rift. *Bull Volcanol* 72:55–62. doi:10.1007/s00445-009-0297-3
- Stefánsson R (2011) Advances in earthquake prediction, research and risk mitigation. Springer, Berlin/Heidelberg
- Stefánsson R, Böðvarsson R, Slunga R, Einarsson P, Jakobsdóttir S, Bungum H, Gregersen S, Havskov J, Hjelme J, Korhonen H (1993) Earthquake prediction research in the South Iceland seismic zone and the SIL project. *Bull Seismol Soc Am* 83:696–716
- Stefánsson R, Guðmundsson GB, Halldórsson P (2008) Tjörnes fracture zone, new and old seismic evidences for the link between the North Iceland rift zone and the mid-Atlantic ridge. *Tectonophysics* 447:117–126. doi:10.1016/j.tecto.2006.09.019
- Stefánsson R, Halldórsson P (1988) Strain release and strain build-up in the South Iceland Seismic Zone. *Tectonophysics* 152: 267–276
- Sturkell E, Einarsson P, Sigmundsson F, Geirsson H, Pedersen R, Van Dalfsen E, Linde A, Sacks S, Stefánsson R (2006) Volcano geodesy and magma dynamics in Iceland. *J Volcanol Geotherm Res* 150:14–34
- Wright TJ, Sigmundsson F, Ayele A, Belachew M, Brandsdóttir B, Calais E, Ebinger C, Einarsson P, Hamling I, Keir D, Lewi E, Pagli C, Pedersen R (2012) Geophysical constraints on the dynamics of spreading centres from rifting episodes on land. *Nat Geosci*. doi:10.1038/NCEO1428

Mechanisms of Earthquakes in Vrancea

Mircea Radulian

National Institute for Earth Physics, Măgurele, Romania

Synonyms

Earthquake; Focal mechanism; Vrancea

Introduction

The Vrancea seismic region is defined as an earthquake nest, meaning a particular kind of intermediate-depth earthquake concentration or clustering. First, an earthquake nest implies a volume of intense activity that is isolated from nearby activity. Second, this activity is persistent over time (at least for decades). From this point of view, nests are distinct from aftershock sequences or earthquake swarms. The definition of what represents an earthquake nest is somewhat arbitrary, and there are many regions in the world where clustering of seismic activity can be categorized as earthquake nests. Nevertheless, earthquake nests are associated to three most famous and intensively studied areas, Bucaramanga (Colombia), Hindu Kush (Afghanistan), and Vrancea (Romania).

Earthquake nests provide perhaps the best setting for understanding the physical mechanism responsible for intermediate-depth earthquakes (waveform data related to a well-defined seismogenic zone are practically available continuously). Mechanisms involving dehydration embrittlement or thermal shear runaway are mostly invoked to explain brittle or brittle-like failure at intermediate depths.

The Vrancea nest occurred in the last stage of the Carpathians tectonic history. Formation of Carpathians comprises first an extension phase between the European and Apulian Plates in Triassic to Early Cretaceous, then a continental collision phase of the European Eurasian and African Plates and finally a post-collision phase (11–0 Ma), when a set of mechanisms were developed such as slab detachment, oceanic or continental lithosphere delamination, thermal or lithospheric instabilities, or combination of them. The collision led to the closure of the Alpine Tethys and to the lateral extrusion of Intra-Alpine plate. The extrusion led to a fan-shaped migration of the Carpathian collision front, accommodated by subduction of the last remnant of the Tethys Ocean. The Vrancea, located at the southeastern bending of Carpathians, represents the only segment of the mountain belt where significant seismicity is recorded at present, concentrated in a

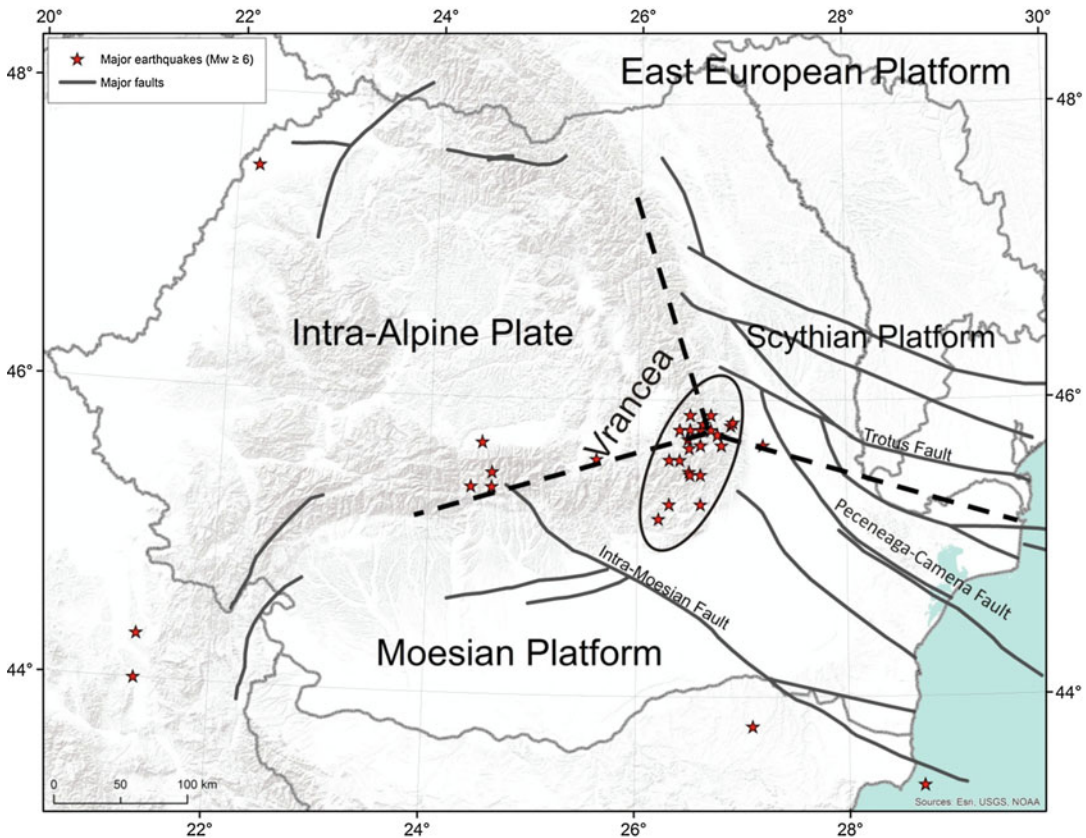
finger-shaped pattern descending almost vertically in the mantle.

Several geodynamic models have been proposed to explain the nature of the seismogenic body beneath the Vrancea region (Ismail-Zadeh et al. 2012). Basically, they belong to two categories: (1) a relic oceanic lithosphere, either attached or already detached from the continental crust, and (2) continental lithosphere that has been delaminated, after continental collision and orogenic thickening. There are arguments in favor and against these two types of models. Oceanic-type models involve break-off process (Fuchs et al. 1979; Wortel and Spakman 2000; Sperner et al. 2001), lateral migration of an oceanic slab (Girbacea and Frisch 1998; Gvirtzman 2002), or subduction and lateral tearing of a slab (Martin and Rietbrock 2006; Wenzel et al. 1998; Wortel and Spakman 2000). Continental-type models involve active delamination (Knapp et al. 2005; Koulakov et al. 2010) or gravitational instability (Lorinczi and Houseman 2009; Ren et al. 2012). The recent results of Bokelman et al. (2014) regarding the dispersion of P waves and the hypothesis of a “double seismic zone” (Bonjer et al. 2005; Radulian et al. 2007; Cărbunar and Radulian 2011) bring evidence in favor of an oceanic nature of the subducting lithosphere.

Tectonic Setting

The earthquake-prone Vrancea region is situated at the bend of the Southeast Carpathians at the contact of East European Platform (EEP) to the north and northeast, Scythian platform to the east, North Dobrogea orogeny to the southeast, Moesian Platform to the south and southwest, and Carpathian orogen and Transylvanian basin (Intra-Alpine plate) to the west and northwest (Fig. 1).

The region is characterized by the frequent occurrence of strong intermediate-depth earthquakes ($M \sim 7$), confined in a narrow zone ($\sim 60 \times 20 \text{ km}^2$ on a horizontal section) within a cold vertically descending lithosphere. The part of the descending lithosphere which is



Mechanisms of Earthquakes in Vrancea, Fig. 1 Location of the Vrancea region. The North Dobrogea zone is located between the Scythian and Moesian Platforms, bounded by Trotuș and Peceneaga-Camena Faults

seismically active is confined in the depths between 70 and 170 km, with no seismicity practically detected below this range. This volume fits well within a high seismic wave velocity body, evidenced by all tomographic studies (e.g., Oncescu 1984; Koch 1985; Wortel and Spakman 2000; Sperner et al. 2001; Martin and Rietbrock 2006; Koulakov et al. 2010). The mantle seismogenic volume is separated from the sub-Carpathian crust by a relatively aseismic zone between 40 and 70 km depth.

The seismicity in the Vrancea overlying crust, spread mostly to the southeast between Trotuș and Intra-Moesian faults, is significantly less important as magnitude (below 5.5) and energy release rate (about four orders of magnitude smaller than in the mantle; Ismail-Zadeh et al. 2012). There was debate on a possible causative relation between mantle and shallow

events, as was suggested by the simple geometry of seismicity alignments (Răileanu et al. 2009; Bonjer et al. 2008; Cărbunar and Radulian 2011). Mitrofan et al. (2014) showed that the triggering effect of the Vrancea intermediate-depth events upon the shallow events in different segments of the Carpathians foredeep region is not noticeable, except one lineament located to the east, close to Danube River, in the Mărăști-Galați-Brăila area. They explain this triggering mechanism by a process of exhumation of the material originating in a slab slicing progression that is consistent with the reverse-faulting mechanism predominating in the Vrancea intermediate-depth domain (see text below).

The intermediate-depth earthquakes along Vrancea strongly affect extended areas in the Southeast and Central Europe. For this reason, the Vrancea source is controlling seismic hazard

level in almost half of the territory of Romania and in cross-border areas as well (in Bulgaria, Republic of Moldova, Serbia, Ukraine). At the same time, the particular pattern of the macroseismic effects provides an efficient clue to identify a major earthquake that occurred in the Vrancea zone in historical times (24 events with $M_w > 7$ since 1,000 according to SHEEC catalogue; Stucchi et al. 2013). The four large events recorded during the last century (10 November 1940, 4 March 1977, 30 August 1986, 30 May 1990) had a significant impact on a large geographical area (Kronrod et al. 2013). The event of 4 March 1977 was the most disastrous, especially toward SW of the epicenter: 1,570 people died, 11,300 were injured, and 32,500 residential and 763 industrial units were destroyed or were seriously damaged in Romania (Sandi 2001).

Analysis of the macroseismic and instrumental data from the intermediate-depth Vrancea earthquakes revealed several peculiarities of the earthquake effects (e.g., Mândrescu and Radulian 1999; Ismail-Zadeh et al. 2012; Kronrod et al. 2013) that can be summarized as follows:

- Extended areas, N-W elongated, are strongly affected.
- NE-SW enhancement of effects coincides with the geometry of seismicity and of the fault-plane solutions.
- Source directivity caused by particular rupture orientation can play a significant role in distributing asymmetrically the intensity effects at regional scale.
- Local and regional geological conditions can control the amplitudes of earthquake ground motion to a larger degree than magnitude or distance.
- Apparently the focal depth can shape to some extent specific patterns in the ground motion distribution.

Focal Mechanisms in the Vrancea Source

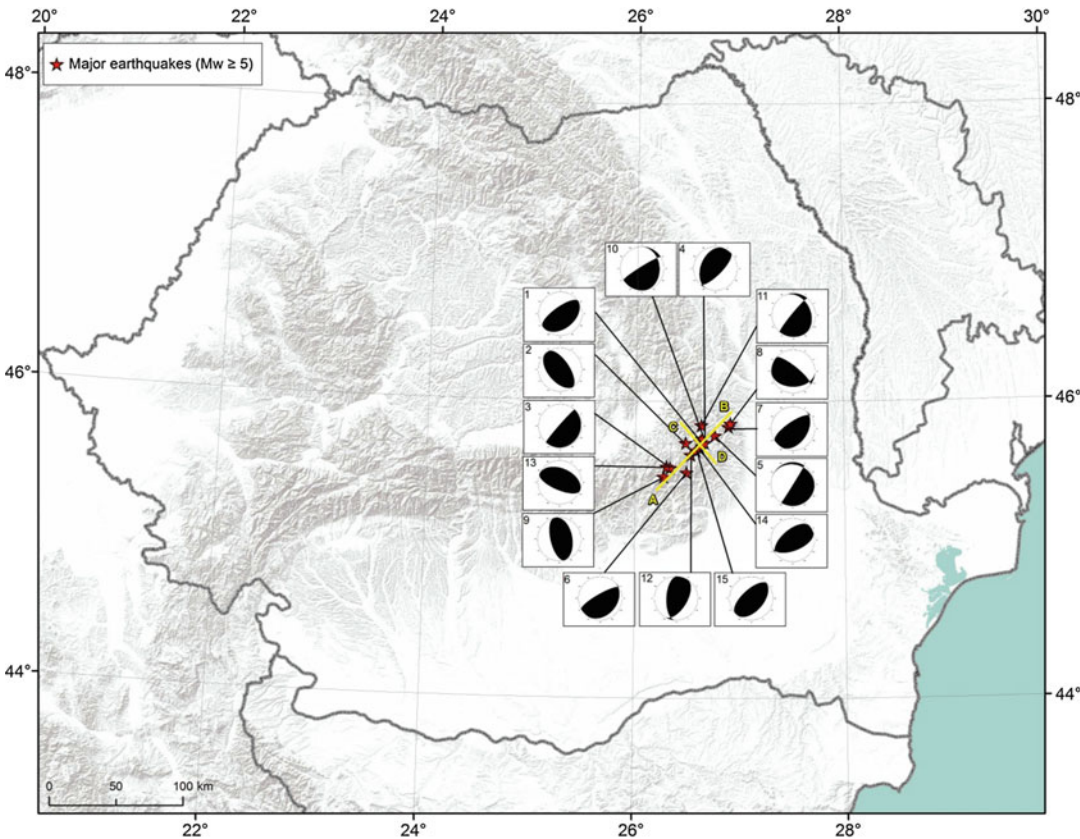
All the previous studies on the subcrustal seismic activity in the Vrancea region outline a predominant dip-slip, reverse faulting, characterizing

both the moderate and strong earthquakes (Enescu 1980; Oncescu 1987; Oncescu and Trifu 1987). In most of the cases, the principal T axis tends to be vertical, the principal P axis tends to be horizontal and oriented perpendicular to the Carpathians arc, and one nodal plane is dipping about 70° toward NW, while the other nodal plane is dipping SE.

The focal mechanisms of the largest Vrancea earthquakes that occurred since 1977 up to the present ($M_w \geq 5.0$) from the CMT catalogue (Dziewonski et al. 1981; <http://www.globalcmt.org/>, accessed on February 2014), represented in Figs. 2 and 3, indicate reverse faulting, and in a number of cases, the focal mechanisms are connected with a strong strike-slip component. The parameters of the nodal planes and P and T principal axes are given in Table 1. These mechanisms are connected with the sinking of the lithosphere in the mantle beneath the Vrancea region. The rate of seismic moment per volume, $\sim 0.8 \times 10^{19}$ Nm/year, in Vrancea is comparable to southern California (Wenzel et al. 1998) since the seismic active volume is extremely concentrated in space.

The fault-plane solutions of the instrumentally recorded large earthquakes are remarkably similar (Fig. 4). The individual solutions for the strong earthquakes of 10 November 1940 ($M_w = 7.7$), 4 March 1977 ($M_w = 7.4$), 30 August 1986 ($M_w = 7.1$), and 30 May 1990 ($M_w = 6.9$) show one another only slight variations. These results suggest that the large earthquakes in the Vrancea subcrustal region are governed by the same geodynamic process.

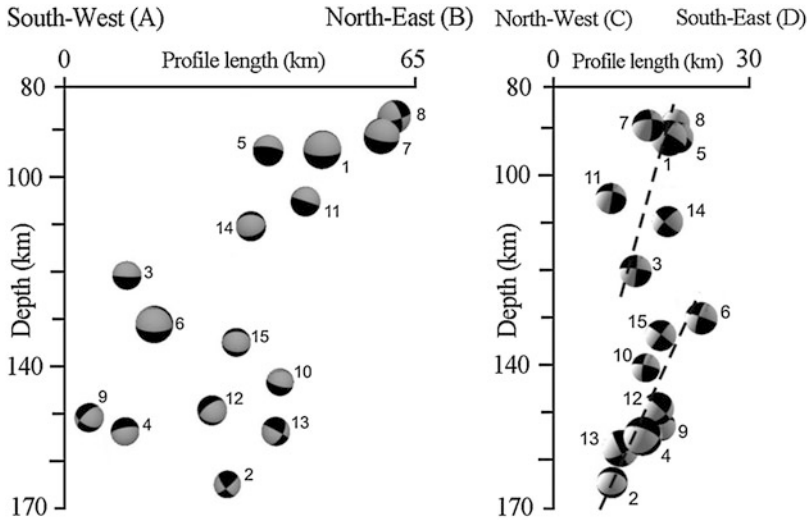
The nodal plane typically striking NE-SW ($\sim 220^\circ$) and dipping 60 – 70° to the NW is ascribed to be the rupture plane (Oncescu and Bonjer 1997; Sperner et al. 2001; Wenzel et al. 2002). This plane coincides with the location of the aftershocks of the major events (Fuchs et al. 1979; Oncescu and Trifu 1987; Trifu et al. 1992). For some of the moderate-size events ($M_w < 7$), the rupture plane is rotated about 90° and dips to the SE (events of 02 October 1978 M_w 5.2, 31 May 1990 M_w 6.4, 28 April 1999 M_w 5.3, and 18 June 2005 M_w 5.2 in Fig. 2). In all cases, the rake angle is close to 90° (i.e., up dip).



Mechanisms of Earthquakes in Vrancea, Fig. 2 Focal mechanisms of the Vrancea largest earthquakes ($M_w \geq 5.0$) recorded since 1977 from the CMT catalogue

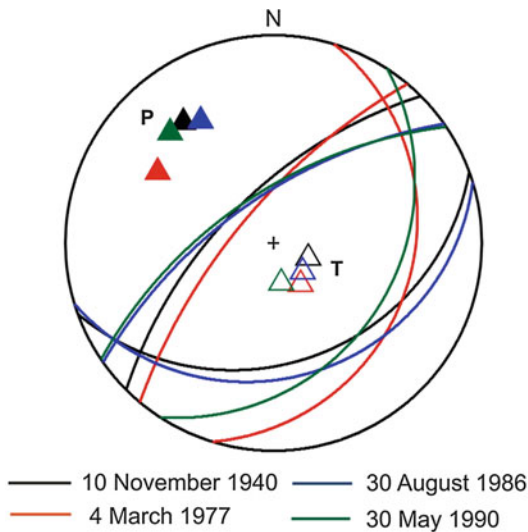
(Harvard, USA) at surface. The events are numbered according to Table 1. Profiles A–B and C–D point the position of the vertical cross sections in Fig. 3

Mechanisms of Earthquakes in Vrancea, Fig. 3 Representation of the focal mechanisms of the earthquakes in Fig. 2 on two vertical cross sections NE-SW (A–B) and NW-SE (C–D). All the events are located at intermediate depths beneath the bend joining Eastern and Southern Carpathians



Mechanisms of Earthquakes in Vrancea, Table 1 CMT fault-plane solutions for Vrancea earthquakes with $M_w \geq 5.0$ recorded since 1977 to the present

No.	Date day/month/year	Time UTC	Latitude °N	Longitude °E	h km	M_w	Plane 1			Plane 2			P axis		T axis	
							az.	dip	slip	az.	dip	slip	az.	pl.	az.	pl.
1.	04/03/1977	19:21	45.77	26.76	94	7.4	50	28	86	235	62	92	323	17	151	73
2.	02/10/1978	20:28	45.72	26.48	164	5.2	140	39	85	326	51	94	53	6	263	83
3.	31/05/1979	07:20	45.55	26.33	120	5.3	23	6	72	221	85	92	309	40	133	50
4.	11/09/1979	15:36	45.56	26.30	154	5.3	202	29	70	45	63	101	127	17	338	70
5.	01/08/1985	14:35	45.73	26.62	94	5.8	288	9	-14	32	88	-98	293	46	130	42
6.	30/08/1986	21:28	45.52	26.49	131	7.0	39	19	70	240	72	97	324	27	160	63
7.	30/05/1990	10:40	45.83	26.89	91	6.9	33	29	70	236	63	101	318	17	168	70
8.	31/05/1990	00:17	45.85	26.91	87	6.3	90	26	54	309	69	106	27	22	244	63
9.	28/04/1999	08:47	45.49	26.27	151	5.3	350	36	93	166	54	88	258	9	66	80
10.	06/04/2000	00:10	45.75	26.64	143	5.0	356	18	29	238	81	106	315	34	167	51
11.	27/10/2004	20:34	45.84	26.63	105	6.0	335	19	27	219	81	107	294	34	149	51
12.	14/05/2005	01:53	45.64	26.53	149	5.5	31	44	111	183	50	71	286	3	29	76
13.	18/06/2005	15:16	45.72	26.66	154	5.2	293	32	90	112	58	90	203	13	22	77
14.	25/04/2009	17:18	45.68	26.62	110	5.4	230	46	77	68	46	102	149	0	58	81
15.	06/10/2013	01:37	45.67	26.58	135	5.2	40	37	84	228	53	95	315	8	160	81



Mechanisms of Earthquakes in Vrancea, Fig. 4 Fault-plane solutions for the major Vrancea earthquakes instrumentally recorded (equal-area lower-hemisphere projection)

Reverse faulting with down-dip extension is predominant along the entire depth range of the seismogenic volume except some variations noticed around 100 km depth (Oncescu and Trifu 1987; Oncescu and Bonjer 1997; Radulian et al. 2002).

The projection of the earthquake foci on a vertical plane crossing the Vrancea area perpendicularly to the Carpathians bend (projection C-D in Fig. 2) suggests the presence of a double seismic zone (Bonjer et al. 2008; Cărbunar and Radulian 2011), which is a characteristic of many subduction zones. Only that in this case the shift between the two nearly parallel “layers” is of the order of location errors (10 km). According to the earthquake triggering scenario suggested by Ganas et al. (2010), there are two NW-dipping nearly parallel planes responsible for triggering successive major events. They are optimally oriented planes of weakness inside the slab to drive the release of the accumulated strain. The foci of the large Vrancea shocks instrumentally recorded are clustered on these two faults accommodating the deformation of the slab (down-dip extension) by shear failure. One cluster is located around 90–100 km depth and the other around 130–140 km depth. A third,

subparallel, major, NW-dipping plane of weakness at depth of about 150–160 km cannot be excluded if the hypothesis of seismic gap at this depth range (Oncescu and Bonjer 1997 and Hurukawa et al. 2008) is valid.

Summary

The Vrancea seismic region is a particularly active source in Europe, embedded in the mantle beneath the Southeastern Carpathians arc bend in a continental tectonic setting. Despite the extreme concentration of the focal volume, frequent earthquakes of magnitude above 7 are generated here in an isolated high-velocity lithospheric body. The focal mechanisms show predominant reverse faulting with the T axis along down dip. The distribution of the strong ground motion is characterized by specific asymmetric patterns, affecting extended areas elongated along NE-SW direction. The elongation coincides with the pattern of the seismicity and the geometry of the rupture plane dipping toward NW. Recent higher-resolution investigations reveal the existence of possible parallel alignments of weakness crossing the seismogenic volume along planes oriented NE-SW and dipping nearly vertically toward NW. These weakness planes are developed separately in two or eventually three segments on depth (centered at 90–100, 130–140, and 150–160 km depth, respectively) where the seismicity is clustered and major shocks nucleate alternatively in time. Repeated generation of large events along parallel sub-faults separated in depth suggests a combination of dehydration process that preferentially weakens preexisting sub-faults in the subducting crust (Kiser et al. 2011) and thermal shear runaway process (Wiens and Sniders 2001).

Cross-References

- [Earthquake Mechanism and Seafloor Deformation for Tsunami Generation](#)
- [Earthquake Mechanisms and Stress Field](#)

- ▶ [Earthquake Mechanism Description and Inversion](#)
- ▶ [Earthquake Mechanisms and Tectonics](#)
- ▶ [Long-Period Moment-Tensor Inversion: The Global CMT Project](#)
- ▶ [Mechanisms of Earthquakes in Aegean](#)
- ▶ [Mechanisms of Earthquakes in Iceland](#)
- ▶ [Regional Moment Tensor Review: An Example from the European–Mediterranean Region](#)

References

- Bokelmann G, Rodler FA (2014) Nature of the Vrancea seismic zone (Eastern Carpathians) – New constraints from dispersion of the first-arriving P-waves. *Earth and Planetary Science Letters* 390:59–68
- Bonjer KP, Ionescu C, Sokolov V, Radulian M, Grecu B, Popa M, Popescu E (2005) Source parameters and ground motion pattern of the October 27, 2004 intermediate depth Vrancea earthquake. In: EGU General Assembly 2005, Vienna (abstracts)
- Bonjer KP, Ionescu C, Sokolov V, Radulian M, Grecu B, Popa M, Popescu E (2008) Ground motion patterns of intermediate-depth Vrancea earthquakes: the October 27, 2004 event. In: Zaiceanu A, Craifaleanu I, Paskaleva I (eds) *Harmonization of seismic hazard in Vrancea zone*, NATO Science for Peace and Security Series – C. Springer, Dordrecht, pp 47–62
- Cărbunar OF, Radulian M (2011) Geometrical constraints for the configuration of the Vrancea (Romania) intermediate-depth seismicity nest. *J Seismol* 15:579–598
- Dziewonski AM, Chou TA, Woodhouse JH (1981) Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *J Geophys Res* 86(2825–2852):1981
- Enescu D (1980) Contributions to the knowledge of the focal mechanism of the Vrancea strong earthquakes of March 4, 1977. *Rev Roum Géol Géophys Géogr Ser Géophys* 24(1):3–18
- Fuchs K, Bonjer K-P, Bock G, Comea I, Radu C, Enescu D, Jianu D, Nurescu A, Merkle G, Moldoveanu T, Tudorache G (1979) The Romanian earthquake of March 4, 1977. II. Aftershocks and migration of seismic activity. *Tectonophysics* 53:225–247
- Ganas A, Grecu B, Batsi E, Radulian M (2010) Vrancea slab earthquakes triggered by static stress transfer. *Nat Hazards Earth Syst Sci* 10:2565–2577
- Girbacea R, Frisch W (1998) Slab in the wrong place: lower lithospheric mantle delamination in the last stage of the Eastern Carpathian subduction retreat. *Geology* 26:611–614
- Gvirtzman Z (2002) Partial detachment of a lithospheric root under the southeast Carpathians: toward a better definition of the detachment concept. *Geology* 30(1):51–54
- Hurukawa N, Popa M, Radulian M (2008) Relocation of large intermediate-depth earthquakes in the Vrancea region, Romania, since 1934 and a seismic gap. *Earth Planets Space* 60:565–572
- Ismail-Zadeh A, Maţenco L, Radulian M, Cloetingh S, Panza GF (2012) Geodynamics and intermediate-depth seismicity in Vrancea (The South-Eastern Carpathians): current state-of-the art. *Tectonophysics* 530–531:50–79
- Kiser E, Ishii M, Langmuir CH, Shearer PM, Hirose H (2011) Insights into the mechanism of intermediate-depth earthquakes from source properties as imaged by back projection of multiple seismic phases. *J Geophys Res* 116, B06310. doi:10.1029/2010JB007831
- Knapp JH, Knapp CC, Raileanu V, Matenco L, Mocanu V, Dinu C (2005) Crustal constraints on the origin of mantle seismicity in the Vrancea Zone, Romania: the case for active continental lithospheric delamination. *Tectonophysics* 410:311–323
- Koch M (1985) Nonlinear inversion of local seismic travel times for the simultaneous determination of the 3D-velocity structure and hypocentres – application to the seismic zone Vrancea. *J Geophys* 56:160–173
- Koulakov I, Zaharia B, Enescu B, Radulian M, Popa M, Parolai S, Zschau J (2010) Delamination or slab detachment beneath Vrancea? New arguments from local earthquake tomography. *Geochem Geophys Geosyst* 11(G3):Q03002. doi:10.1029/2009GC002811
- Kronrod T, Radulian M, Panza GF, Popa M, Paskaleva I, Radovanovich S, Grijbovski K, Sandu I, Pekevski L (2013) Integrated transnational macroseismic data set for the strongest earthquakes of Vrancea (Romania). *Tectonophysics* 590:1–23
- Lorinczi P, Houseman GA (2009) Lithospheric gravitational instability beneath the Southeast Carpathians. *Tectonophysics* 474:322–336
- Măndrescu N, Radulian M (1999) Macroseismic field of the Romanian intermediate-depth earthquakes. In: Wenzel F, Lungu D, Novak O (eds) *Vrancea earthquakes: tectonics, hazard and risk mitigation*. Kluwer, Dordrecht, pp 163–174
- Martin S, Rietbrock A (2006) Guided waves at subduction zones: dependencies on slab geometry, receiver locations and earthquake sources. *Geophys J Int* 167:693–704
- Mitrofan H, Chitea F, Anghelache M-A, Visan M (2014) Possible triggered seismicity signatures associated with the Vrancea intermediate-depth strong earthquakes (Southeast Carpathians, Romania). *Seismol Res Lett* 85(2):314–323
- Onescu MC (1984) Deep structure of Vrancea region, Romania, inferred from simultaneous inversion for hypocenters and 3-D velocity structure. *Ann Geophys* 2:23–28
- Onescu MC (1987) On the stress tensor in Vrancea region. *J Geophys Res* 62:62–65

- Oncescu MC, Bonjer KP (1997) A note on the depth recurrence and strain release of large Vrancea earthquakes. *Tectonophysics* 272:291–302
- Oncescu MC, Trifu CI (1987) Depth variation of the moment tensor principal axes in Vrancea (Romania) seismic region. *Ann Geophysicae* 5B:149–154
- Radulian M, Băllă A, Popescu E (2002) Earthquakes distribution and their focal mechanism in seismogenic zones of Romania. *Roman Report Phys* 47:945–963
- Radulian M, Bonjer K-P, Popescu E, Popa M, Ionescu C, Grecu B (2007) The October 27th, 2004 Vrancea (Romania) earthquake. *ORFEUS Newsletter* 1(7), Jan 2007
- Răileanu V, Dinu C, Ardeleanu L, Diaconescu V, Popescu E, Băllă A (2009) Crustal seismicity and associated fault systems in Romania. In: *Proceedings of the 27th ECGS workshop: seismicity patterns in the euromed region*, Luxembourg, 17–19 Nov.2008, in “Cahiers du Centre Europeen de Geodynamique et de Seismologie”, pp 153–159
- Ren Y, Stuart GW, Houseman GA, Dando B, Ionescu C, Hegedüs E, Radovanović S, Shen Y, South Carpathian Project Working Group (2012) Upper mantle structures beneath the Carpathian–Pannonian region: implications for the geodynamics of continental collision. *Earth Planet Sci Lett* 349–350:139–152
- Sandi H (2001) Obstacles to earthquake risk reduction encountered in Romania. In: Lungu D, Saito T (eds) *Earthquake hazard and countermeasures for existing fragile buildings*. Independent Film, Bucharest, pp 261–266
- Sperner B, Lorenz F, Bonjer K, Hettel S, Muller B, Wenzel F (2001) Slab break-off – abrupt cut or gradual detachment? New insights from the Vrancea Region (SE Carpathians, Romania). *Terra Nova* 13:172–179
- Stucchi M, Rovida A, Gomez Capera AA, Alexandre P, Camelbeeck T, Demircioglu MB, Kouskouna V, Gasperini P, Musson RMW, Radulian M, Sesetyan K, Vilanova S, Baumont D, Faeh D, Lenhardt W, Martinez Solares JM, Scotti O, Zivcic M, Albin P, Batllo J, Papaioannou C, Tatevossian R, Locati M, Meletti C, Viganò D, Giardini D (2013) The SHARE European Earthquake Catalogue (SHEEC) 1000–1899. *J Seismol* 17:523–544
- Trifu CI, Deschamps A, Radulian M, Lyon-Caen H (1992) The Vrancea earthquake of May 30, 1990: an estimate of the source parameters. In: *Proceedings of the XXIIInd ESC Gen. Ass.*, Barcelona, 1990, vol 1, pp 449–454
- Wenzel F, Achauer U, Enescu D, Kissling E, Mocanu V (1998) Detailed look at final stage of plate break-off is target of study in Romania. *EOS Trans Am Geophys Union* 79(48):589–594
- Wenzel F, Sperner B, Lorenz F, Mocanu V (2002) Geodynamics, tomographic images and seismicity of the Vrancea region (SE-Carpathians, Romania), vol 3, EGU Stephan Mueller Special Publication Series. Katlenburg-Lindau, Germany, pp 95–104
- Wiens DA, Snider NO (2001) Repeating deep earthquakes: evidence for fault reactivation at great depth. *Science* 293:1463–1466
- Wortel MJR, Spakman W (2000) Subduction and slab detachment in the Mediterranean-Carpathian region. *Science* 290:1910–1917

MEMS Sensors for Measurement of Structure Seismic Response and Their Application

Kenji Oguni

Department of System Design Engineering,
Keio University, Yokohama, Japan

Synonyms

Accelerometer; Displacement sensor; MEMS (microelectromechanical systems); Restoring force characteristics; SDOF (single degree of freedom) mass-spring system; Sensor network; Velocimeter

Introduction

MEMS is an abbreviation for microelectromechanical systems. MEMS sensor is a measurement device consisting of mechanical parts for picking up events or environmental changes (i.e., sensors) and electric circuit for amplifying and filtering the signals from sensors. MEMS sensor is usually integrated on a small circuit board.

The major advantages of MEMS sensors for measurement of structure seismic response are their small size, low power consumption, and low price in contrast with their high accuracy in measurement. For example, a typical price of MEMS accelerometers is less than 10 USD. In spite of this low cost, a MEMS accelerometer with 3 axes, ± 2 g full scale, 5 mm $\times$ 5 mm $\times$ 2 mm size, 1.1 mA current consumption at the input voltage of 2.7 V, and noise density of 175 $\mu\text{g}/\sqrt{\text{Hz}}$ is available off the shelf.

MEMS sensors are basically implemented as fine patterns printed on silicon wafers. These patterns are the combination of the mechanical parts corresponding to the sensor pickup and the electric circuit to convert, amplify, and filter the signal from sensor pickup. The size and the energy consumption of the MEMS can be suppressed by implementing the patterns on the silicon wafers as small as possible. The micro-fabrication technology made mass production and distribution of low-cost MEMS sensors possible.

A typical micro-fabrication process of the MEMS sensors is a photolithograph method. In photolithography, photosensitive material is painted on a silicon wafer. Then, this layer is masked by the patterns of the electric circuit. After the exposure of this wafer to the light, unmasked area is scraped off by reactive gas or fluid (gas, dry etching; fluid, wet etching). Finally, the masked area remains as fine patterns of the electric circuit. This photolithographic method enables us to automatically produce more than 1,000 circuit boards (each of them corresponds to a MEMS sensor) from one silicon wafer plate (30 cm in diameter). This is the key production technology for low-cost MEMS sensors.

Among many types of MEMS sensors, MEMS accelerometers are mainly used in the field of earthquake engineering. The measurement principle of most MEMS accelerometers is that of SDOF (single degree of freedom) mass-spring system. If some condition (mentioned below in this manuscript) is satisfied, the relative displacement between the mass and the support is proportional to the absolute acceleration of the support.

This is the overview of the MEMS technology and the MEMS sensors related to the field of earthquake engineering. This manuscript is organized as follows. First, the difference between a displacement sensor, a velocimeter, and an accelerometer is illustrated by the analysis of SDOF mass-spring system. Through this analysis, the reason for (i) many MEMS accelerometers are available with low cost, and (ii) no MEMS velocimeters and MEMS displacement sensors are available. Also, the limitation of the MEMS accelerometers and the point to be noted for usage of the MEMS accelerometers are discussed. This discussion leads us to a possible application of the MEMS accelerometers in the field of earthquake engineering.

Vibration Measurement Device

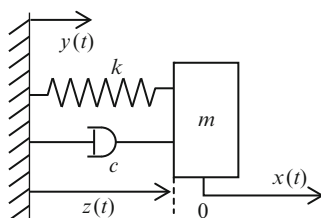
Consider an SDOF (single degree of freedom) mass-spring system shown in Fig. 1. The mass is m , spring constant is k , and the dashpot is c . The governing equation of this system is

$$m(\ddot{y}(t) + \ddot{z}(t)) + c\dot{z}(t) + kz(t) = 0 \quad (1)$$

where t is time. Suppose that the displacement of the support is $y(t) = Y_0 \sin \omega t$. Then Eq. 1 becomes

$$\ddot{z}(t) + 2\zeta\omega_0\dot{z}(t) + \omega_0^2 z(t) = \omega^2 Y_0 \sin \omega t \quad (2)$$

where $\omega_0^2 = k/m$ and $\zeta = c/\sqrt{4mk}$. The following expression for $z(t)$ can be obtained by solving Eq. 2.



$x(t)$: absolute displacement of the mass
 $y(t)$: absolute displacement of the support
 $z(t)$: relative displacement of the mass to the support

MEMS Sensors for Measurement of Structure Seismic Response and Their Application, Fig. 1 SDOF mass-spring system (simplified model for vibration measurement device)

$$z(t) = \frac{(\omega/\omega_0)^2}{\sqrt{[1 - (\omega/\omega_0)^2]^2 + (2\xi\omega/\omega_0)^2}} Y_0 \sin(\omega t - \phi),$$

$$\phi = \tan^{-1} \left(\frac{2\xi\omega/\omega_0}{1 - (\omega/\omega_0)^2} \right) \quad (3)$$

Consider the following three cases for the angular frequency of the absolute displacement of the support ω and the natural angular frequency of the mass-spring system ω_0 :

Case 1: $\omega_0 \ll \omega$

Case 2: $\omega_0 \cong \omega$

Case 3: $\omega_0 \gg \omega$

In Case 1, since $[1 - (\omega/\omega_0)^2]^2 \cong (\omega/\omega_0)^4 \gg (2\xi\omega/\omega_0)^2$ and $\phi = \pi$,

$$z(t) \cong -Y_0 \sin \omega t = -y(t). \quad (4)$$

Equation 4 means that for $\omega_0 \ll \omega$, the absolute displacement of the support $y(t)$ is almost equal to $-z(t)$, where $z(t)$, the relative displacement of the mass to the support, is the measurable quantity. Therefore, for the case of $\omega_0 \ll \omega$, this SDOF mass-spring system can be used as a displacement sensor by converting the relative displacement between the mass and the support to some electric signal.

In Case 2, $\omega_0 \cong \omega$ and $\phi = \pi/2$ result in

$$z(t) \cong \frac{1}{2\xi\omega_0} Y_0 \cos \omega t = \frac{1}{2\xi\omega_0} \dot{y}(t). \quad (5)$$

Equation 5 means that for $\omega_0 \cong \omega$, the relative displacement of the mass to the support, $z(t)$, is almost proportional to the absolute velocity of the support, $\dot{y}(t)$. Therefore, when the condition for Case 2 is satisfied, this SDOF mass-spring system can be used as a velocimeter by converting the relative displacement between the mass and the support to some electric signal.

In Case 3, since $\omega_0 \gg \omega$, $\omega/\omega_0 \cong 0$ and $\phi = 0$. Therefore,

$$z(t) \cong \frac{1}{\omega_0^2} \omega^2 Y_0 \sin \omega t = -\frac{1}{\omega_0^2} \ddot{y}(t). \quad (6)$$

Equation 6 means that in Case 3, the relative displacement of the mass to the support, $z(t)$, is almost proportional to the absolute acceleration of the support, $\ddot{y}(t)$. Therefore, for the case of $\omega_0 \gg \omega$, this SDOF mass-spring system can be used as an accelerometer by converting the relative displacement between the mass and the support to some electric signal.

Here, the physical meaning of each case for the SDOF mass-spring system shown in Fig. 1 should be discussed. In Case 1, this SDOF mass-spring system works as a displacement sensor. The condition $\omega_0 \ll \omega$ requires relatively small ω_0 . Since $\omega_0 = \sqrt{k/m}$, small ω_0 means large m and small k . Thus, an SDOF displacement sensor consists of a very heavy mass supported by an extremely soft spring which results in a bulky and fragile device. This is the reason for nonexistence of a MEMS displacement meter based on the SDOF mass-spring system.

In Case 2, $\omega_0 \cong \omega$ requires a subtle tuning of the device. Thus, a velocimeter becomes a precision instrument and this is not suitable for the mass production on a silicon wafer (typical production method for MEMS devices). This is why there exists no MEMS velocimeter based on the SDOF mass-spring system. One comment should be added for an SDOF velocimeter. As shown in Eq. 5, the proportional constant between the absolute velocity of the support $\dot{y}(t)$ and the measured relative displacement of the mass $z(t)$ is $1/(2\xi\omega_0)$. Ordinary materials have small ξ in the order of 0.05. Therefore, $\dot{y}(t)$ is amplified a lot. The gain of the SDOF velocimeter is large enough for using this as a precise seismometer.

In Case 3, $\omega_0 \gg \omega$ means relatively large ω_0 . Since $\omega_0 = \sqrt{k/m}$, large ω_0 means small m and large k . Thus, an SDOF accelerometer consists of a light mass supported by an extremely hard spring which results in a small and tough device. The smaller and the tougher, the device works as a better accelerometer. This nature of the SDOF accelerometer is suitable for the mass production on a silicon wafer. This is why MEMS accelerometers with reasonable cost are widely supplied in the market.

There is an additional explanation for MEMS accelerometers. As shown in Eq. 6, the proportional constant between the absolute acceleration of the support $\ddot{y}(t)$ and the measured relative displacement of the mass $z(t)$ is $-1/\omega_0^2$. Since large ω_0 is the requirement for an SDOF accelerometer, the amplification factor $1/\omega_0^2$ becomes very small. Therefore, a MEMS accelerometer should have an internal electric circuit for amplifying the signal. The measuring accuracy of the MEMS accelerometer is limited by this amplifier circuit. The accuracy of this amplifier circuit is the noise density mentioned at the beginning of this manuscript. The lowest noise density of currently available MEMS accelerometers is in the order of several tens of $\mu g/\sqrt{\text{Hz}}$. Thus, acceleration measurement using a MEMS accelerometer with the bandwidth of 100 Hz results in the noise in the order of 1 gal. Compared with non-MEMS precise accelerometers whose measurement noise is in the order of a few milligal, the measurement accuracy of the MEMS accelerometer is not high. This implies limitation of the application of the MEMS accelerometer in the field of earthquake engineering.

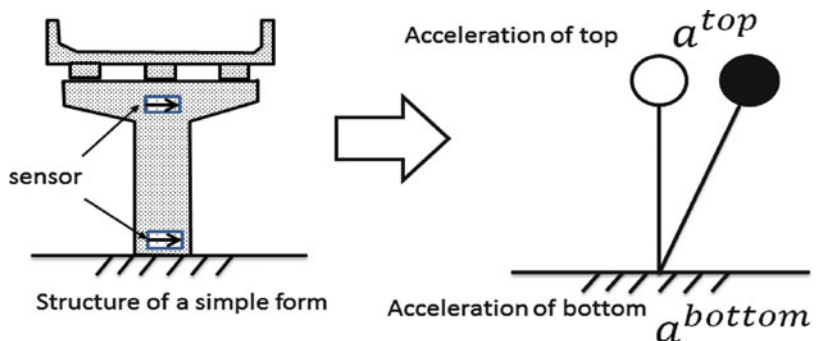
Application of MEMS Accelerometers for Earthquake Engineering

Sensor network has been started by Smart Dust project (Kahn et al. 1999) in the late 1990s. Ever since, sensor network technology has been rapidly advanced. This is mainly because of the

advancement of MEMS technology. This sensor network technology has been also imported to the field of earthquake engineering. Measurements of the ambient vibration of the structures and the structure seismic response to weak ground motion are performed by using MEMS accelerometer sensor network. These measurements are intended for structural health monitoring. However, no successful example has been observed in these structural health monitoring applications by densely deployed MEMS accelerometers (Farrar et al. 2001; Farrar and Worden 2007). The major reason for these unsuccessful attempts for MEMS accelerometer-based structural health monitoring is the insufficient accuracy of MEMS accelerometers. Structural health monitoring using the ambient vibration of the structures and the structure seismic response to weak ground motion requires the change in the structural response to weak input (less than 10 gal). Therefore, the accuracy of MEMS accelerometers mentioned before (1 gal for 100 Hz measurement bandwidth) is obviously insufficient. MEMS accelerometer-based application for earthquake engineering has to be the measurement of structure seismic response to strong (at least 100 gal) ground motion.

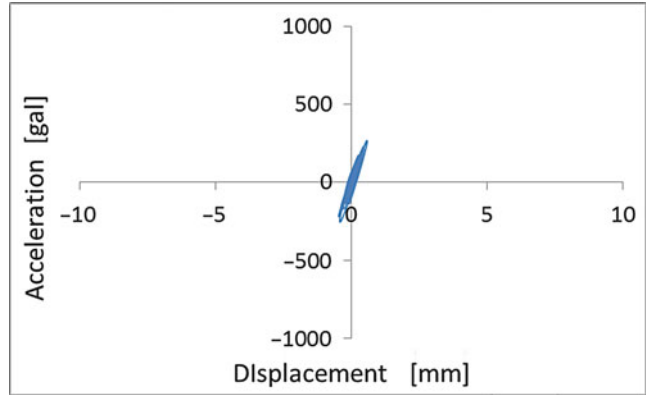
An example application of a sensor network system equipped with MEMS accelerometers is the measurement of restoring force characteristics of structures subjected to strong ground motion. Right after a severe earthquake, the damage level of the structures in disaster-stricken region has to be quickly identified. For this

MEMS Sensors for Measurement of Structure Seismic Response and Their Application,
Fig. 2 Restoring force characteristics measurement using MEMS accelerometer sensor network



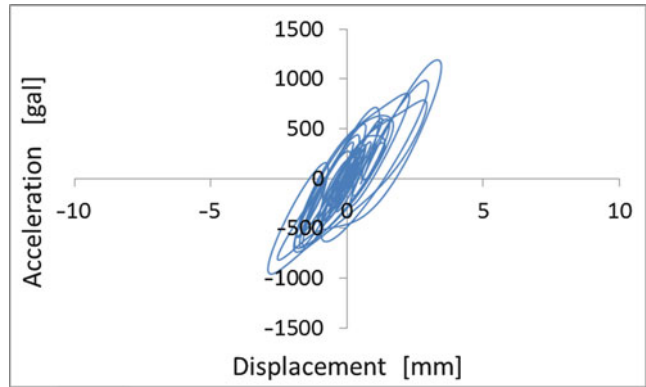
**MEMS Sensors for
Measurement of
Structure Seismic
Response and Their
Application,**

Fig. 3 Typical restoring force characteristics under weak seismic ground motion



**MEMS Sensors for
Measurement of
Structure Seismic
Response and Their
Application,**

Fig. 4 Typical restoring force characteristics under strong seismic ground motion



purpose, MEMS accelerometer sensor network system can be used.

Structures of a simple shape such as the standard viaducts for railways and highways can be safely approximated as an SDOF mass-spring system as shown in Fig. 2. Then, the acceleration at the top of the structure is proportional to the inertia force acting on the mass and the relative displacement between the top and the bottom of the structure corresponds to the relative displacement between the mass and the support. Therefore, if the acceleration at the top of the structure and the relative displacement between the top and the bottom of the structure subjected to ground motion can be observed, the property of the spring, i.e., the restoring force characteristics of the structure, can be obtained.

At the top and the bottom of the structure shown in Fig. 2, sensor nodes equipped with a MEMS accelerometer are installed. These sensor nodes are also equipped with wireless communication device and a synchronized measurement can be performed between these sensor nodes (Ikeda et al. 2012).

Typical restoring force characteristics are shown in Figs. 3 and 4. The vertical axis is the acceleration at the top of the structure and the horizontal axis is the relative displacement between the top and the bottom of the structure. Subjected to weak ground motion shown in Fig. 3, the structure remains elastic and the restoring force characteristics are linear. On the other hand, Figure 4 corresponding to the case of strong ground motion shows

hysteresis curves which indicate the plastic response of the structure.

These restoring force characteristic can be obtained by a synchronized acceleration measurement using a pair of sensor nodes installed at the top and the bottom of the structure. The relative displacement can be obtained through time integration of the acceleration data using a band-pass filter which includes the lowest natural frequency of the structure and excludes the higher natural frequencies.

The advantage of this application over unsuccessful attempts for MEMS accelerometer-based structural health monitoring is the required accuracy in the acceleration measurement. The measurement accuracy of the commercially available MEMS accelerometers is good enough for constructing restoring force characteristics.

Summary

MEMS sensors for measurement of structure seismic response and their application are discussed. The principle of the measurement of displacement, velocity, and acceleration is illustrated through an analysis of an SDOF mass-spring system. Based on this analysis, the reason for the popularity of the MEMS accelerometers and their limitations are discussed. An example of a possible application of MEMS accelerometer sensor network is presented.

References

- Farrar CR, Worden K (2007) An introduction to structural health monitoring. *Phil Trans R Soc Lond A* 365(1851):303–315
- Farrar CR, Doebling SW, Nix DA (2001) Vibration-based structural damage identification. *Phil Trans R Soc Lond A* 359(1778):131–149
- Ikeda H, Kawaguchi T, Oguni K (2012) Automatic detection of damage level of structures under the severe earthquake using sensor network. *Appl Mech Mater* 166–169:2216–2220
- Kahn JM, Katz RH, Pister KSJ (1999) Mobile networking for smart dust. In: *Proceedings of MobiCom 99*, online publication

Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis

Thanasis Papageorgiou¹ and Charis J. Gantes²

¹Department of Structural Engineering, School of Civil Engineering, National Technical University of Athens (N.T.U.A.), Athens, Greece

²Institute of Steel Structures, School of Civil Engineering, National Technical University of Athens, Zografou Campus, Athens, Greece

Synonyms

Complex response; Decoupling; Equivalent damping; Irregular structures; Primary-secondary systems

Introduction

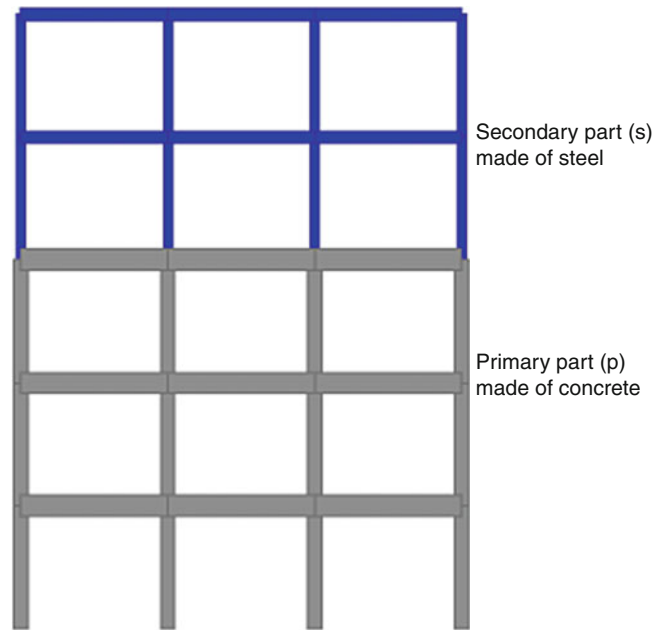
Issues pertaining to the dynamic behavior of mixed in-height structures, and more specifically concrete-steel buildings, are presented. The term “mixed structures” denotes here buildings that consist of a lower and an upper part with different dynamic response characteristics. When such structures are subjected to a dynamic excitation, e.g., an earthquake, they respond in a different way than more uniform structures; thus, their dynamic analysis and seismic design must account for these peculiarities.

The dissimilarity in the dynamic response characteristics of the two parts can be due to several factors. It may occur due to different lateral force resisting systems in the upper and lower part, or it might be the result of the action of seismic isolation or energy dissipation devices. One of the most common causes, which is the coexistence of different construction materials leading to nonuniform damping and elastoplastic behavior over the height of the building, is studied here.

More specifically, this section is devoted to modeling and analysis issues of structures consisting of a lower part made of concrete that

**Mixed In-Height
Concrete-Steel Buildings
Under Seismic Actions:
Modeling and Analysis,**

Fig. 1 Schematic
representation of mixed
in-height structure



**Mixed In-Height
Concrete-Steel Buildings
Under Seismic Actions:
Modeling and Analysis,**

Fig. 2 Mixed in-height
concrete-steel building



M

rests on the ground and supports an upper part made of steel. A schematic representation of such a structure is shown in Fig. 1. The lower part in the pertinent bibliography is usually referred to as primary structure and is denoted with the letter *p*, while the upper part is called secondary structure and is denoted with the letter *s*. Emphasis is directed here towards the different damping characteristics of the two parts, as reinforced concrete structures are known to exhibit higher damping than steel structures. Without loss of generality, it

is assumed that the concrete and steel parts have damping ratios equal to 5 % and 2 %, respectively, which are values commonly encountered in design codes for these two materials.

This configuration is very common in the case that an addition of floors occurs sometime in the life span of a building originally made of concrete, where, for reasons of speed of construction and reduction of inertia forces and floor mass or merely for architectural reasons, the added floors are made of steel (Fig. 2). Other typical examples



Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 3 Telecommunications tower made of steel mounted on concrete building

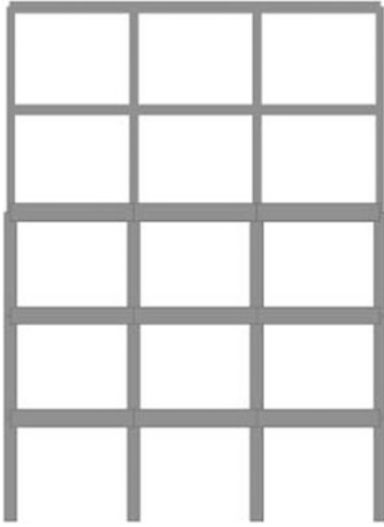
Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 4 Typical stadium configuration with concrete seats and steel cover



of mixed in-height structures are the cases of concrete buildings with telecommunication towers mounted on top of them (Fig. 3) and stadiums with the structures supporting the spectator's seats made of concrete and covered by a steel truss canopy (Fig. 4). Thus, mixed

concrete-steel structural systems are widespread and offer in many occasions desirable solutions due to architectural, structural, and functional reasons.

As a result, there are numerous structures with mixed concrete-steel configuration, having parts



Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 5 Typical MDoF structure configured in one material

that, during a seismic event, exhibit different damping and dissipate energy in different ways. Given that the pertinent seismic design regulations refer to structures consisting of only one material and specify accordingly analysis and design rules, the modeling, calculation of seismic actions, analysis, and design of mixed in-height concrete-steel buildings require a different treatment. The aim of this section is to investigate analysis possibilities of such mixed structures in ways similar to the ones followed for typical homogeneous structures, e.g., with methods that engineers are well acquainted with.

Fundamental Seismic Engineering Concepts

A typical multi-degree of freedom (MDoF) structure consisting of only one structural material, like the one shown in Fig. 5, is the case usually encountered in engineering practice and fully covered by pertinent seismic codes in terms of analysis and design specifications. The seismic oscillation of a structure with n degrees of freedom is described by Eq. 1, where the ground motion is denoted with $\ddot{x}_g$; the displacement

with x ; M , C , K are the mass, damping, and stiffness matrices, respectively; and r is a unit vector. The damping matrix here represents a typical Rayleigh type damping and can be easily produced using the mass and stiffness matrices.

$$M\{\ddot{x}\} + C\{\dot{x}\} + K\{x\} = -M\{r\}\ddot{x}_g \quad (1)$$

This equation is easy to handle and may be solved to obtain the response of the structure in a manner of ways, widely implemented and incorporated in commercial software. For example, the equation itself can be solved in the time or in the frequency domain, yielding thus the time history of the response in terms of total accelerations or displacements at each degree of freedom. Furthermore, a choice closer to engineering practice is to perform an eigenvalue analysis using the mass and stiffness matrices of Eq. 1 and thus obtain the modal quantities of the structure. Next, each mode is analyzed separately as a single degree of freedom (SDoF) structure as in Eq. 2, where q , ω , Γ are the response, eigenfrequency, and participation factor of each mode.

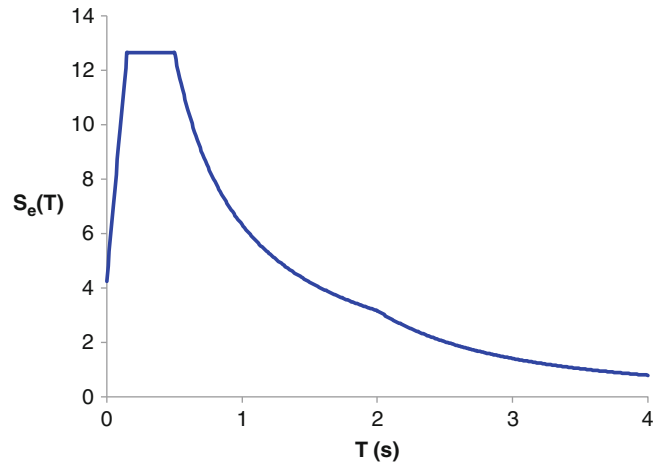
$$\ddot{q}_i + 2\zeta_i\omega_i\dot{q}_i + \omega_i^2q_i = -\Gamma_i\ddot{x}_g, \quad i = 1, 2, \dots, n \quad (2)$$

After formulating Eq. 2, it is possible either to analyze it also in the time or frequency domain and thus obtain the time history of the response of each mode or to estimate the maxima of each modal response via the use of a response spectrum, like the ones defined by EC8 (European Committee of Standardization 2004) and shown in Fig. 6. Then, by means of superimposing the modal maxima, it is possible to obtain the maxima of the overall response.

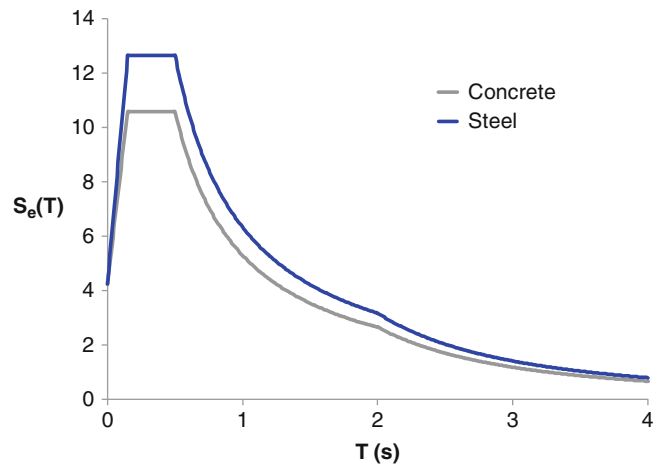
Peculiarities in the Seismic Response of Mixed Structures

In contrast to the case of typical structures consisting of only one material, the dynamic analysis of irregular structures made of concrete

**Mixed In-Height
Concrete-Steel Buildings
Under Seismic Actions:
Modeling and Analysis,
Fig. 6** EC8 elastic
response spectrum



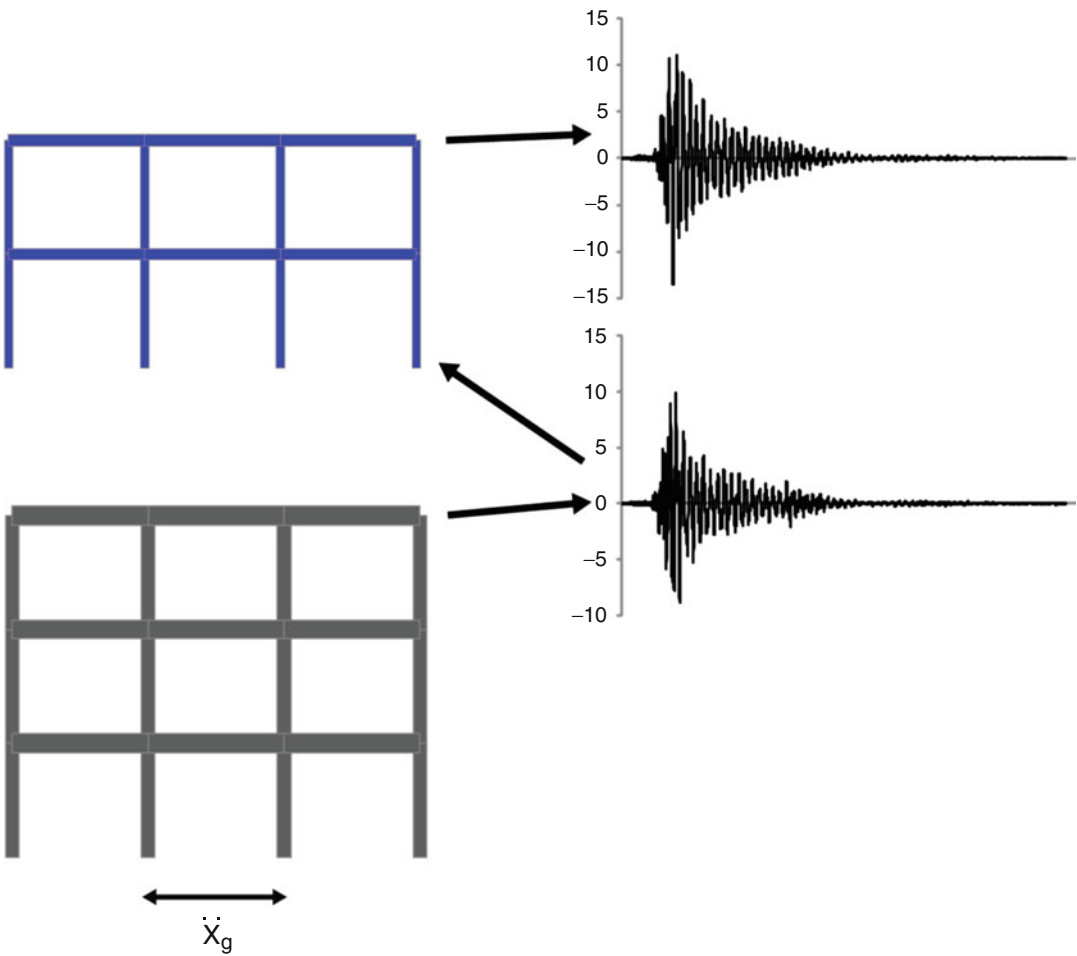
**Mixed In-Height
Concrete-Steel Buildings
Under Seismic Actions:
Modeling and Analysis,
Fig. 7** Comparison
between the concrete and
steel, compatible to EC8,
elastic spectra



and steel as the ones schematically shown in Fig. 1 is not straightforward. The difficulties occur already from the beginning, when forming the equation of motion. Unlike the case of uniform structures, where the damping matrix of Eq. 1 can be easily produced by the mass and stiffness matrices, now a rather cumbersome procedure must be followed in order to produce a damping matrix that manages to represent the actual damping distribution of the structure. During this approach, which is not supported by readily available commercial software, two separate damping matrices of Rayleigh type are formed, one for each part of the structure, which are then merged into a single matrix to be used in Eq. 1.

Then, the equation of motion itself cannot be decoupled in real-valued modal equations of motion like in Eq. 2. Instead, complex-valued modes occur, leading thus to complex-valued eigenfrequencies and modal damping ratios as well as modal participation factors. Moreover, should someone neglect the above and try to implement a typical spectral analysis using code compatible or even seismic spectra, the dilemma of choosing a single universal damping ratio comes up, which may alter significantly the response, as shown in the comparison of the EC8 compatible elastic spectra for damping values of 5 % and 2 %, shown in Fig. 7.

Thus, a common engineering analysis problem turns into a very complex one. As a result,



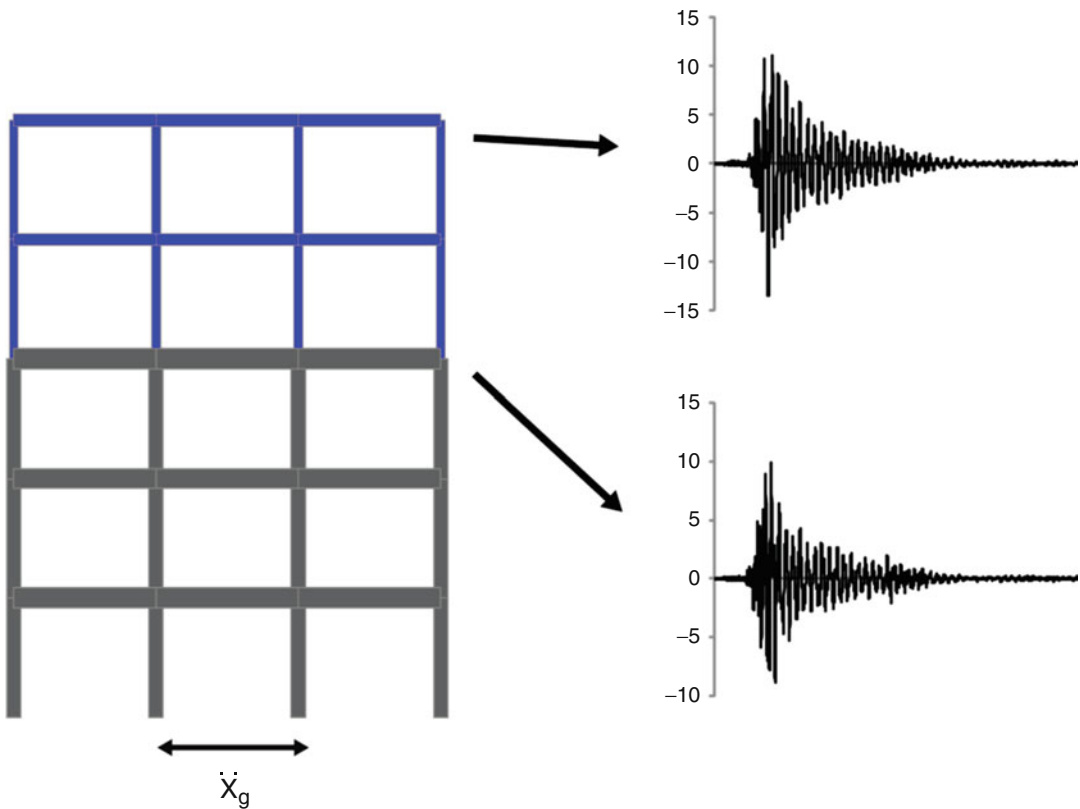
Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 8 Decoupled analysis procedure

and in order to avoid out of the ordinary analysis procedures, the common practice for such cases of irregularity is to make conservative assumptions for the damping ratios, which then leads to overestimation of the response. Due to the frequency with which such irregular structures are encountered, a significant amount of research has been carried out in order to provide analysis solutions that are compatible with common engineering practice.

An interesting approach that has been developed in that direction is the decoupled analysis of such structures. In this procedure the irregular structure is actually divided in two homogeneous parts, and each of them is analyzed separately, as

shown in Fig. 8, in contrast to the typical coupled analysis procedure shown in Fig. 9. In the decoupled procedure, the ground acceleration is induced in the concrete part, and its response is then used as a fictitious excitation for the steel part of the structure. The advantage of this method is that in each step the structural part analyzed is homogeneous and therefore typical analysis techniques are applicable. The disadvantage is that the so-called decoupling error is introduced, due to the fact that the analysis is not carried out in the actual structure.

An analytical prediction of this decoupling error for the case of harmonic excitations has been attempted (Chen and Wu 1999), by



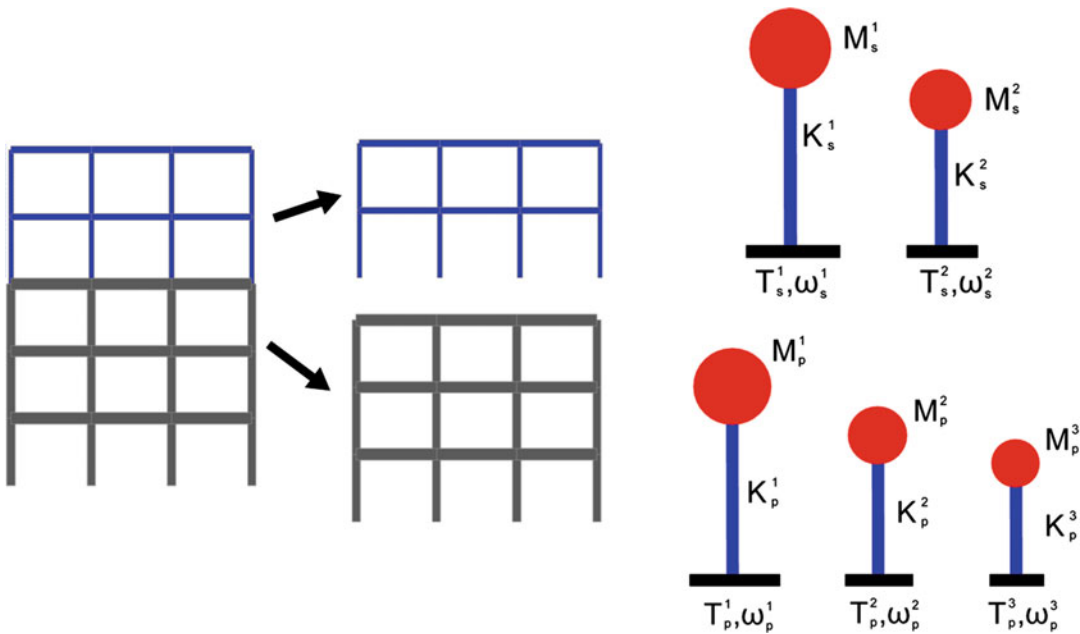
Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 9 Coupled analysis procedure

introducing for that purpose a 2-DoF oscillator and thus managing a limited connection to MDoF structures, while analytical solutions for the decoupling error based on a complex-valued process are also provided (Gupta and Tembulkar 1984a, b).

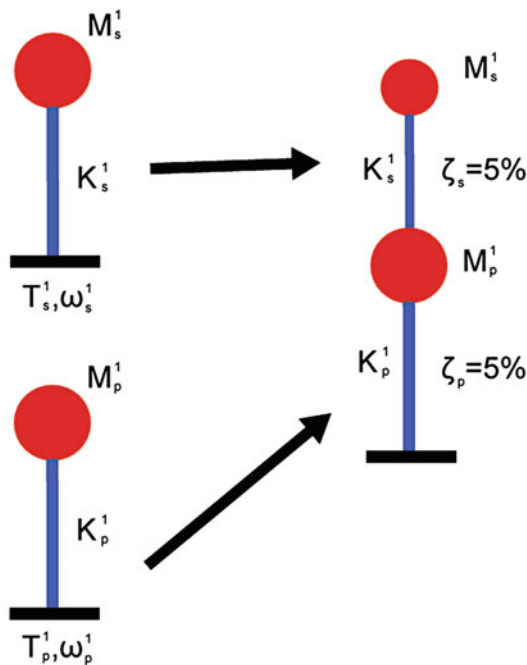
In a modification of these approaches (Papageorgiou 2010a; Papageorgiou and Gantes 2005), the decoupling error is investigated numerically, by performing time history analyses of the 2-DoF model, so that other issues can be taken into account, such as inelastic material behavior as well as nonharmonic excitations including earthquake accelerograms. The analyses are also based on 2-DoF oscillators, but the results are efficiently correlated to MDoF structures by means of assigning the first mode characteristics of each part, i.e., modal mass and eigenfrequency, to the dynamic characteristics of each degree of freedom of the 2-DoF model, as

shown in Figs. 10 and 11. Moreover, graphical tools are produced allowing the prediction of the decoupling error at a preliminary design level, as will be shown next.

In the case that the errors arising from the decoupling procedure are above acceptable limits, a coupled analysis of the irregular structure is the proper solution. In this case and in order to avoid the adoption of a conservative damping ratio, the possibility of using equivalent damping ratios has been thoroughly investigated. By means of such ratios, the exact analysis of the structure, which is difficult to handle with readily available commercial software, is substituted with approximate analyses (Bilbao et al. 2006; Papagiannopoulos and Beskos 2009, 2006), even for the specific case of composite concrete-steel structures (Huang et al. 1996). More specifically, for the case of irregular concrete-steel structures, the results of numerical time history analyses



Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 10 Partition of the mixed structure and modal analysis of each part



Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 11 Formation of the equivalent 2-DoF oscillator

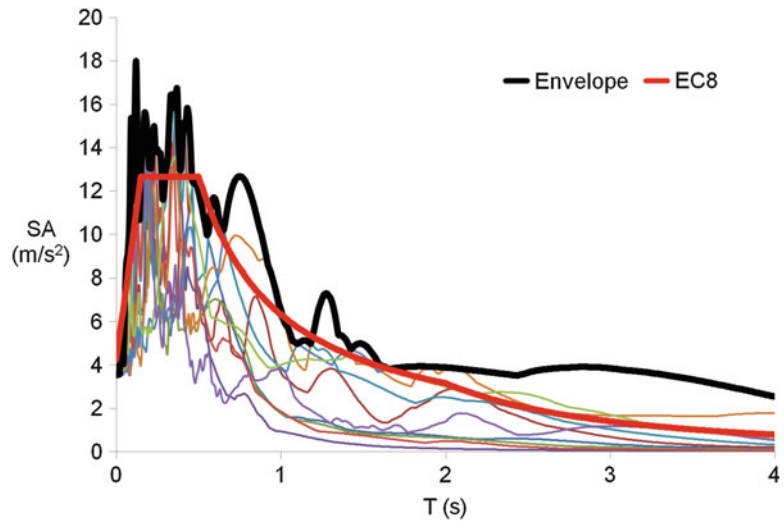
over a wide range of dynamic characteristics of the two parts permitted the development of graphical tools proposing equivalent damping ratios, both modal and uniform for the entire structure (Papageorgiou and Gantes 2008a, b, 2010b, 2011), as will be presented in the following.

Decoupling

As explained above, a common way to overcome the difficulties in the analysis and design of irregular concrete-steel structures is to divide them in two homogeneous parts and analyze each part separately. The concrete primary structure is analyzed first and its response at the support level of the secondary structure is then used as a fictitious excitation for the steel secondary structure. This decoupled procedure resembles the provisions of current seismic codes relevant to appendages, with the drawback that these provisions cannot be implemented in the case of superstructures with significant masses like the added steel floors

Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis,

Fig. 12 Spectra of the seismic records used and comparison of their envelope to EC8 compatible spectrum



of structures schematically represented in Figs. 1, 2, 3, and 4. This division also allows different engineering teams to undertake the design of each part, thus allowing offices specialized in constructions of each material to undertake the analysis and design of each relevant part of the structure.

The disadvantage of the decoupled procedure is that in either of its two analysis stages, the structure analyzed is a part of the real one and not the actual structure with its full mass, damping, and stiffness matrices. Thus, the interaction of the two parts is neglected and an error is introduced, in the prediction of the response of both the primary and the secondary structure. In order to have a reliable and safe, i.e., conservative, prediction of the decoupling error, a thorough investigation has been carried out in order to obtain its distribution for a wide range of dynamic characteristics of the two parts of the structure.

For that purpose, the 2DoF representation illustrated in Fig. 11 has been implemented. From the fundamental eigenfrequencies and the corresponding modal masses of the two parts, two ratios are defined, namely, the eigenfrequency and the mass ratios of secondary to primary structure, R_ω and R_m , respectively, as in Eq. 3. Thus, an (R_ω, R_m) plane can be defined containing a wide range of ratio pairs correlated through

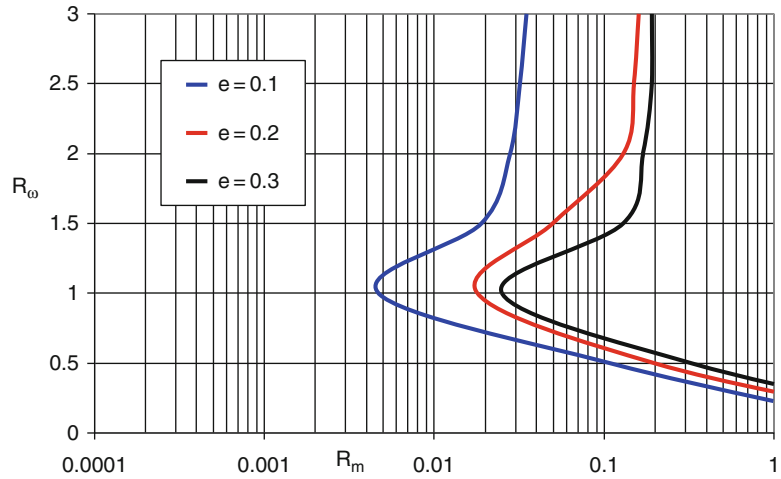
the first mode dynamic quantities to a wide range of irregular structures.

$$R_m = \frac{M_s}{M_p}, R_\omega = \frac{\omega_s}{\omega_p} \quad (3)$$

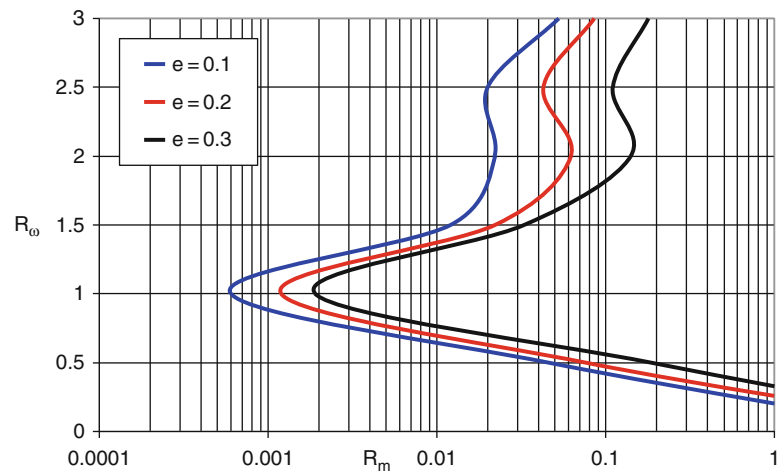
For the specific case where earthquake excitations are considered, a range of seismic records have been chosen based on the dispersion of the frequencies they resonate the most, and they have been used in the analyses. More specifically, the records were chosen from seismic events in Greece as well as other seismic active areas of the world, and they were selected so as to produce a group of spectra enveloping the code prescribed ones, as shown in Fig. 12.

For each point in the (R_ω, R_m) plane, the corresponding structure is analyzed both approximately, with a decoupled analysis, and “exactly,” with a coupled analysis. Thus, the decoupling error is calculated for all points of the (R_ω, R_m) plane, i.e., for the chosen wide range of dynamic characteristics of the irregular structures, as described in Eq. 4 for all levels of the structure. Three reference error levels are chosen, namely, 10 %, 20 %, and 30 %, and the points of the (R_ω, R_m) plane at which these error levels appear are noted. Finally, the envelope of these points is plotted in a set of curves named decoupling curves, one for each error level, as

Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis,
Fig. 13 Seismic decoupling curves for the primary structure



Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis,
Fig. 14 Seismic decoupling curves for the secondary structure



shown in Fig. 13 for the concrete part and in Fig. 14 for the steel part of the irregular structure.

$$e = \left| \frac{\max(|\ddot{x}^{\text{dec}}|) - \max(|\ddot{x}^{\text{coup}}|)}{\max(|\ddot{x}^{\text{coup}}|)} \right| \quad (4)$$

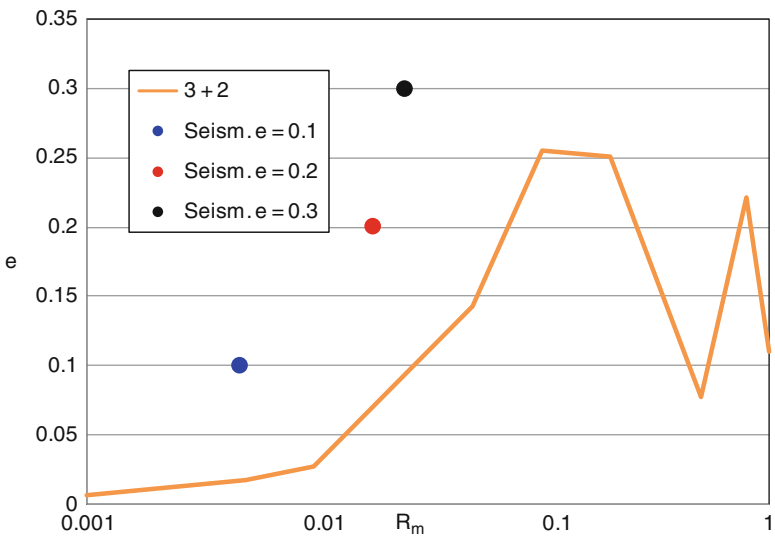
These decoupling curves may be of use at a preliminary design stage, when the decision has to be taken whether to proceed with the decoupled or the coupled analysis procedure. At this stage and after the conceptual design of the structure has been concluded, modal analyses of each part of the structure can be carried out. Thus, the R_ω and R_m ratios can be defined so that the

structure can be positioned as a point in the (R_ω, R_m) plane. If the point representing the structure lies to the left of a curve, it means that the expected decoupling error is smaller than the one predicted by this specific curve, while if it is on the right, it means larger expected error. Hence, with a maximum error tolerance defined and with the decoupling error prediction by the curves of Figs. 13 and 14, the decision whether to proceed with decoupling can be taken.

The fact that the decoupling curves of Figs. 13 and 14 have been constructed as envelopes of the points arising from each seismic record makes them a conservative, and thus safe, prediction of the decoupling error. The veracity of this claim is

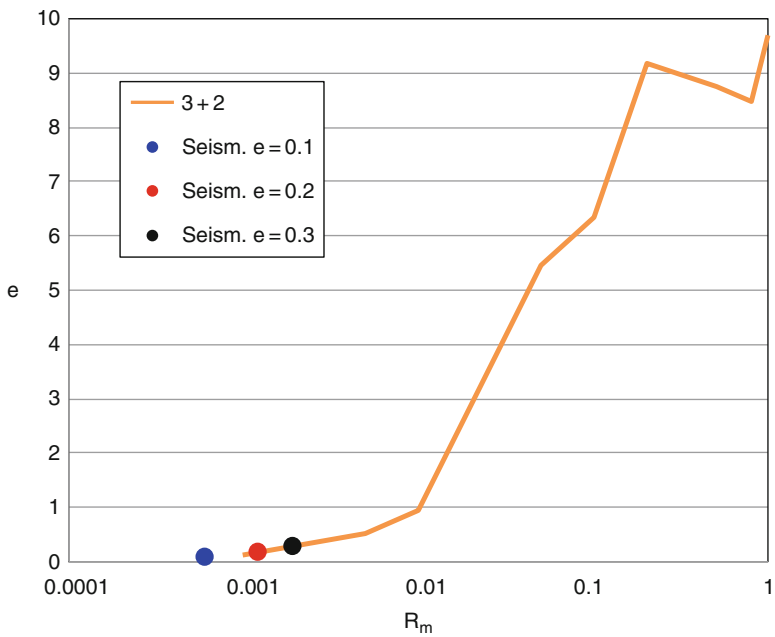
Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis,

Fig. 15 Comparison between actual error and prediction from seismic decoupling curves for the primary structure



Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis,

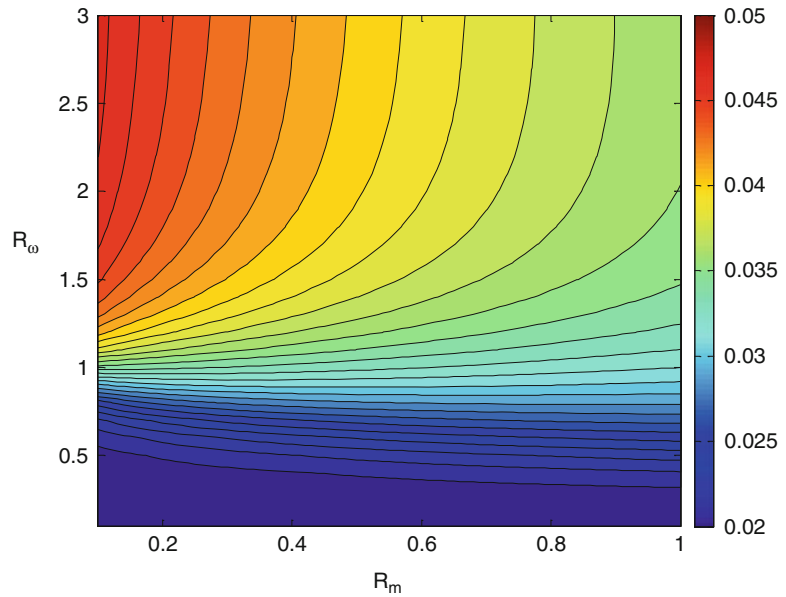
Fig. 16 Comparison between actual error and prediction from seismic decoupling curves for the secondary structure



tested over an irregular structure comprised by a two-floor steel superstructure mounted atop of a three-story concrete substructure, as the example shown in Fig. 1. In order to test the efficiency of the decoupling curves towards a conservative prediction of the decoupling error, a series of analyses is performed with the first modes of the two parts being in resonance. This is the worst-case scenario for the decoupling error, and the

structure of Fig. 1 is modified accordingly in order to be tested over a range of mass ratios. The decoupling error is monitored and shown in Fig. 15 for the primary structure and in Fig. 16 for the secondary structure. In order to provide an assessment of the efficacy of the decoupling curves, the points of the decoupling curves of Figs. 13 and 14 for $R_\omega = 1$ are shown in the respective colors.

**Mixed In-Height
Concrete-Steel Buildings
Under Seismic Actions:
Modeling and Analysis,**
Fig. 17 First mode
damping ratio



As seen from Figs. 15 and 16, the decoupling curves manage a conservative prediction of the decoupling error for the specific mass ratios. Even considering the worst-case scenario exhibited above, i.e., that of resonance between primary and secondary structures, the error prediction actually coincides with the actual error of the superstructure (Fig. 16), while it is very conservative as far as the primary structure is concerned (Fig. 15). Therefore, the developed decoupling curves can be a useful and at the same time safe tool at a preliminary design stage of analysis in order to enable decision taking whether to proceed with coupled or decoupled analysis.

Equivalent Modal Damping Ratios

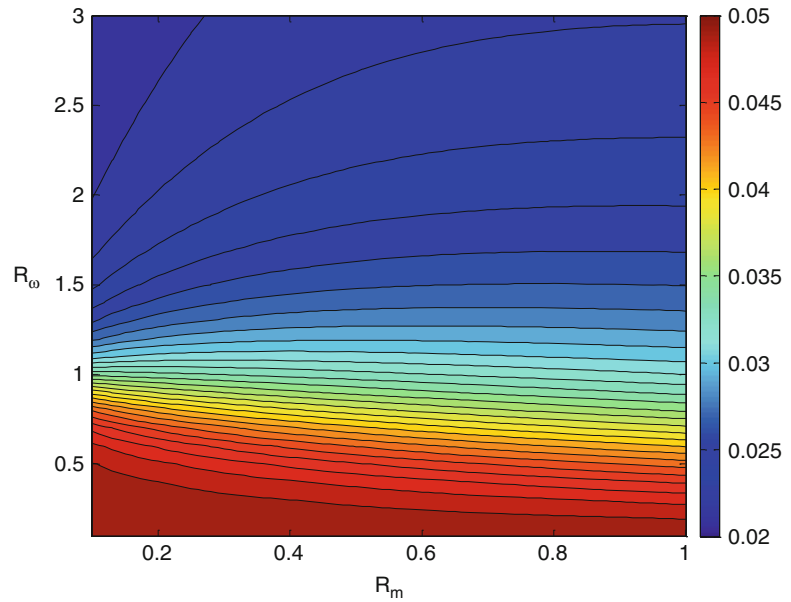
It is evident from Figs. 13 and 14 that there are areas of the (R_ω, R_m) plane, i.e., irregular structures with certain combination of dynamic characteristics of the two parts, for which the decoupling error is larger than 30 %. Furthermore, as shown in Fig. 16, this error can be disproportionately large for $R_\omega = 1$ and large R_m values. In such cases, the decoupled analysis procedure is not acceptable. Thus, in order to

overcome the coupled analysis difficulties of an irregular structure consisting of two parts with different damping ratios, in this section equivalent damping ratios are proposed, suitably defined so as to effectively approximate the actual response by means of typical analysis that can be performed using commercially available software.

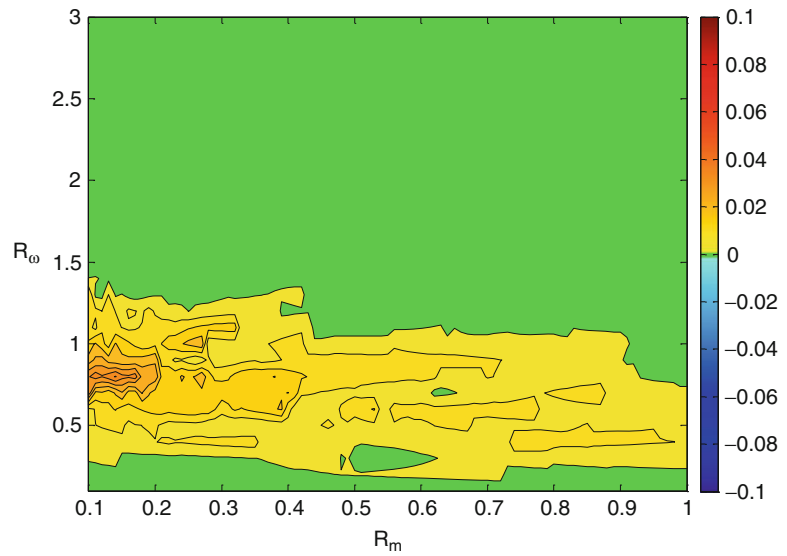
Initially, for each point in the (R_ω, R_m) plane, i.e., for each corresponding irregular structure, the complex damping ratios that normally occur from the complex-valued modal analysis are modified to real-valued ones, by simply adopting their absolute value. The distribution of the modal damping ratios for the first two modes of the structure over the (R_ω, R_m) plane is shown in Figs. 17 and 18.

As in the case of the decoupled analysis, the use of the absolute value of the complex damping ratios in a common real-valued modal analysis described in Eq. 2 also introduces an error, since the correct approach requires an analysis procedure that is complex valued in all steps. This error is again quantified for structures covering all points in the (R_ω, R_m) plane, as shown in Eq. 5 in terms of displacements, for all structure levels. For every point of the (R_ω, R_m) plane, i.e., for every structure investigated, the absolute value of

**Mixed In-Height
Concrete-Steel Buildings
Under Seismic Actions:
Modeling and Analysis,
Fig. 18** Second mode
damping ratio



**Mixed In-Height
Concrete-Steel Buildings
Under Seismic Actions:
Modeling and Analysis,
Fig. 19** Error envelopes
for the primary structure



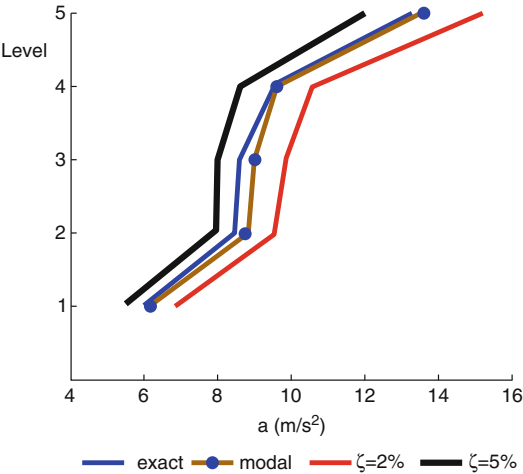
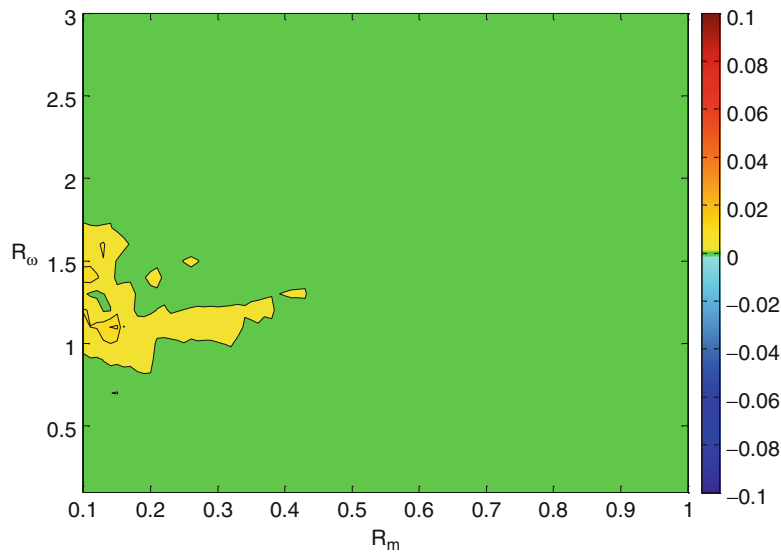
these errors is obtained for each seismic record used, and the envelopes of the maxima are presented in Figs. 19 and 20.

$$e = \frac{\max(x^{\text{appr}}) - \max(x^{\text{exact}})}{\max(x^{\text{exact}})} \quad (5)$$

The fact that the error distribution both for the primary structure and for the secondary one is

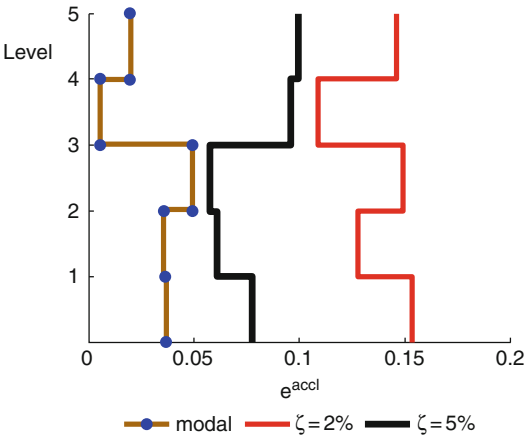
sufficiently small, at least for a preliminary design level, encourages the adoption of the equivalent damping ratios in the analysis of actual irregular structures. For the structure with two steel floors mounted atop of three concrete levels, schematically shown in Fig. 1, with such dynamic characteristics of the two parts that lead to $R_\omega = 1$ and $R_m = 0.5$, i.e., a case that the two parts are in resonance, the maximum total

Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 20 Error envelopes for the secondary structure



Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 21 Total acceleration profile for a five-story irregular structure

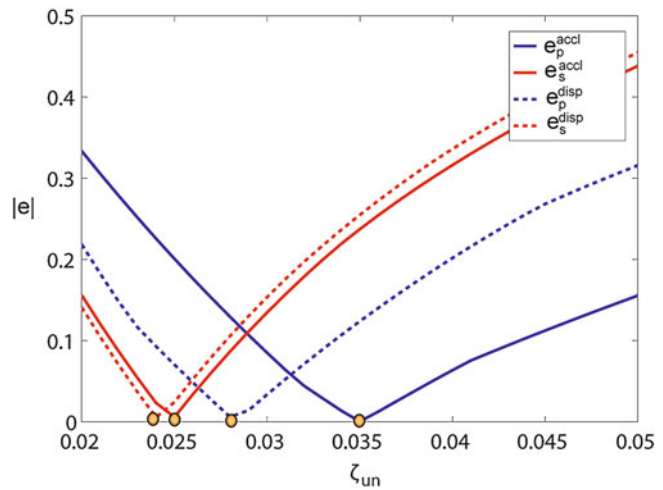
acceleration distribution has been compared between the actual response and the one arising from the adoption of the equivalent modal damping ratios, and the results are shown in Fig. 21. Moreover, in order to allow comparison with common engineering practice, the distribution obtained by the adoption of overall damping ratios equal to 2 % and 5 % is presented as well. The good performance of the proposed equivalent damping ratios is evident.



Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 22 Error profile for a five-story irregular structure

Furthermore, in order to demonstrate the effectiveness of the proposed equivalent damping ratios, the error distribution as defined in Eq. 6, in absolute values, over the profile of the structure, is presented in Fig. 22. Along with the error occurring from the modification of the complex-valued damping ratios in order to be used in a real-valued modal analysis, the error that would occur should the damping ratio be assumed to have overall values equal to 2 % or 5 %, again in absolute values, is presented also. It is made clear that the complex-valued modal

**Mixed In-Height
Concrete-Steel Buildings
Under Seismic Actions:
Modeling and Analysis,**
Fig. 23 Error variation
with respect to the uniform
damping ratio



damping ratios, suitably modified in order to be used in a common, real-valued, modal analysis procedure, give totally acceptable results, and thus, the whole process is deemed suitable to be adopted in engineering practice. Moreover, the fact that there is readily available commercial software capable of performing structural analysis allowing the specification of modal damping ratios enhances the application prospects of the above-described procedure.

$$e = \left| \frac{\max(\ddot{x}^{\text{appr}}) - \max(\ddot{x}^{\text{exact}})}{\max(\ddot{x}^{\text{exact}})} \right| \quad (6)$$

Equivalent Uniform Damping Ratios

In an effort to provide engineers with equivalent damping ratios that are even closer to engineering practice compared to the modal ones, uniform damping ratios that apply to the whole structure and not just to certain modes are proposed next. In order to come up with these uniform equivalent damping ratios for each structure of the (R_ω, R_m) plane, a trial and error procedure is followed during which the set of possible uniform damping ratios between 0.02 and 0.05 with a step of 0.001 is tested and the response at each level of the structure is compared to the actual one.

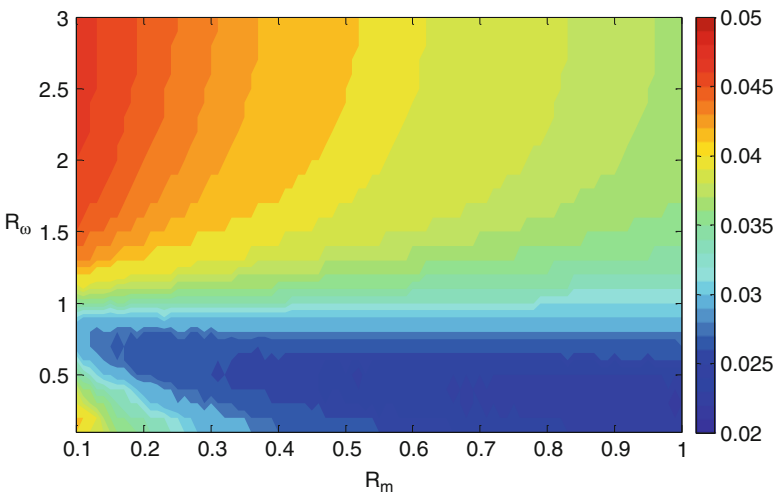
In order to increase the accuracy of this procedure, the efficacy of the proposed uniform

damping ratios is tested independently for the primary and the secondary structure, in terms of total accelerations and displacements, and the dependence of the error on the damping ratios is shown in Fig. 23. The value of equivalent damping ratio within the above range that minimizes the error in the prediction of the response of the structure, whether in terms of accelerations or in terms of displacements, is chosen as optimum and used as equivalent uniform damping ratio for obtaining the corresponding response quantity.

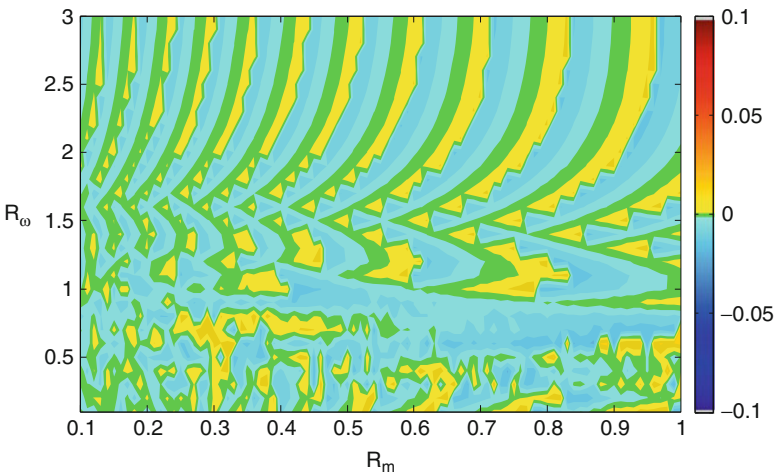
This procedure is carried out for all points of the (R_ω, R_m) plane, i.e., for irregular structures with a wide range of dynamic characteristics, and for each point the optimum damping ratio in terms of total accelerations and displacements is noted for the concrete and the steel part of the structure. Thus, a distribution of proposed equivalent damping ratios is obtained in terms of total accelerations and displacements in the primary and the secondary part of the structure, as shown in Figs. 24, 25, 26, 27, 28, 29, 30, and 31, along with the corresponding errors.

The error distributions are sufficiently small in this case as well, and thus, their adoption instead of the actual damping distribution is recommended. For the irregular structure of Fig. 1, and for the specific case $R_\omega = 1, R_m = 0.5$, i.e., a case that the two parts are in resonance, the response of the structure using the proposed equivalent damping ratios is tested against the actual response.

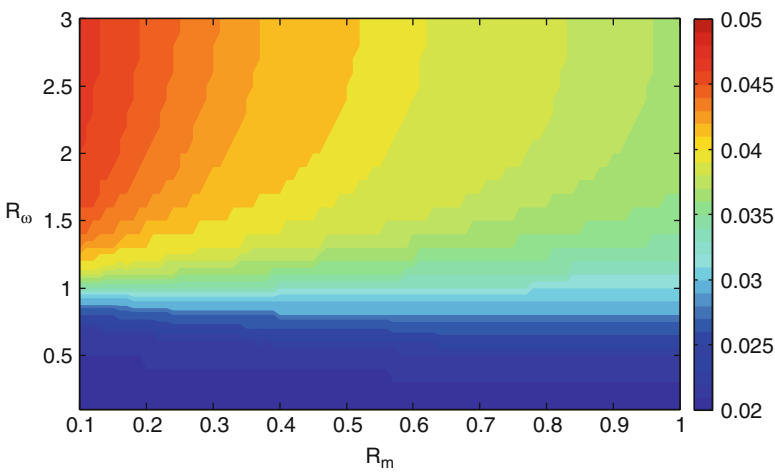
Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 24 Equivalent damping ratio minimizing the displacement error in the primary structure



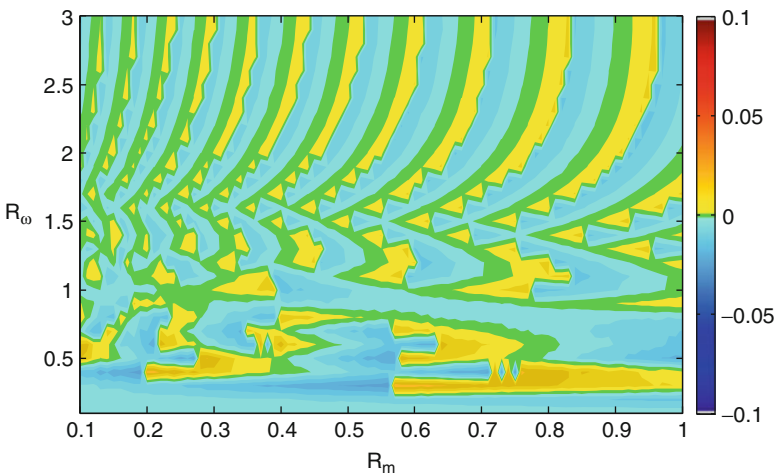
Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 25 Minimized displacement error in the primary structure



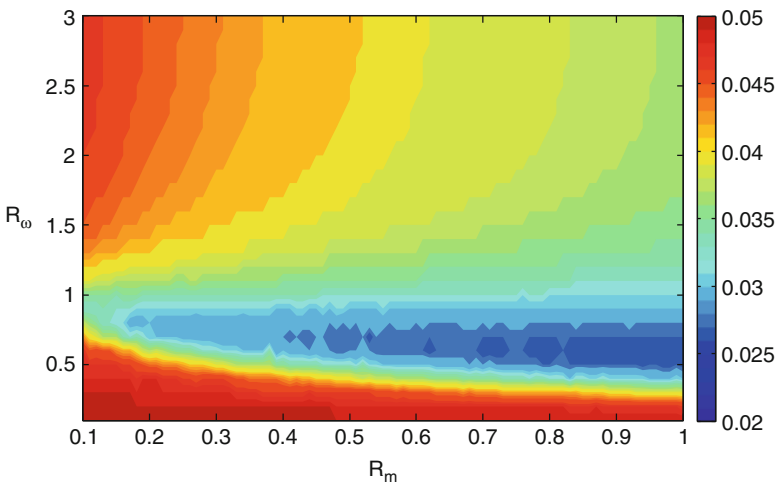
Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 26 Equivalent damping ratio minimizing the displacement error in the secondary structure



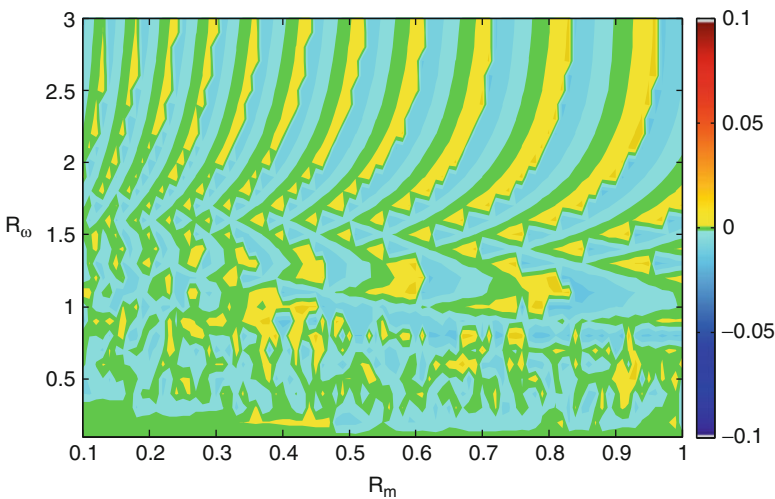
**Mixed In-Height
Concrete-Steel Buildings
Under Seismic Actions:
Modeling and Analysis,
Fig. 27** Minimized
displacement error in the
secondary structure



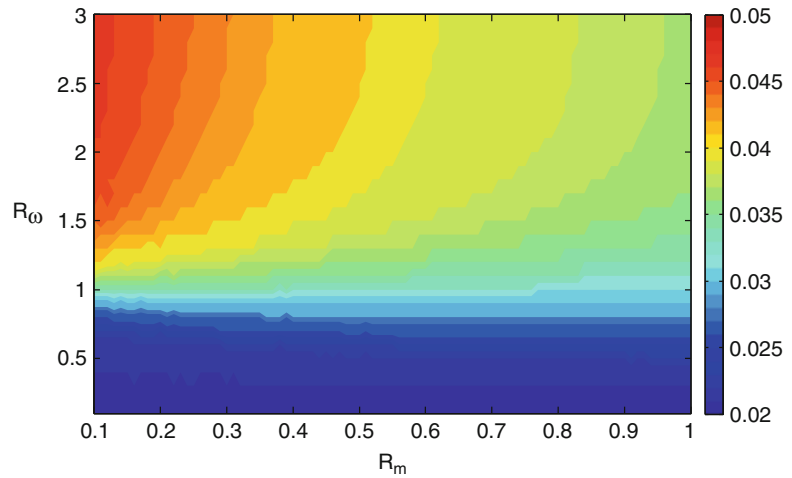
**Mixed In-Height
Concrete-Steel Buildings
Under Seismic Actions:
Modeling and Analysis,
Fig. 28** Equivalent
damping ratio minimizing
the acceleration error in the
primary structure



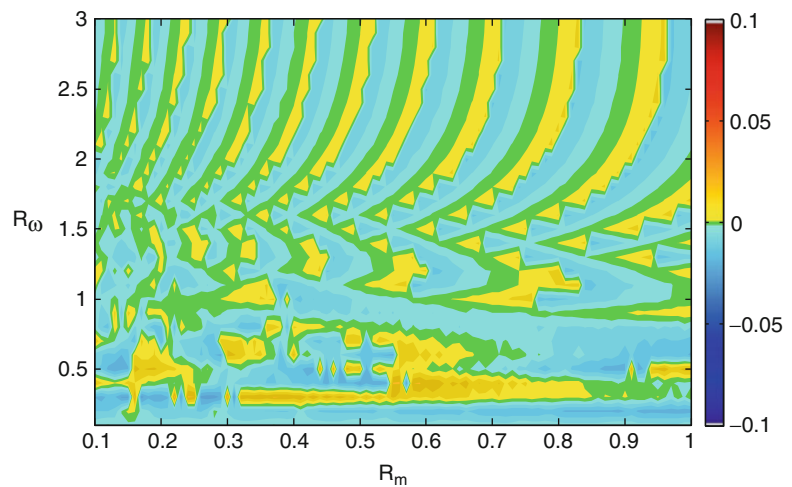
**Mixed In-Height
Concrete-Steel Buildings
Under Seismic Actions:
Modeling and Analysis,
Fig. 29** Minimized
acceleration error in the
primary structure



Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 30 Equivalent damping ratio minimizing the acceleration error in the secondary structure



Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 31 Minimized acceleration error in the secondary structure



M

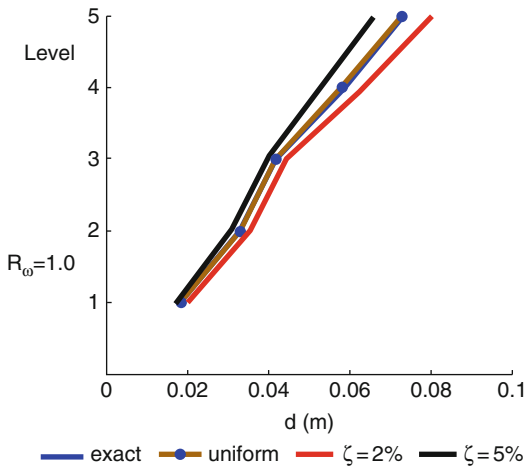
More specifically, the structure is analyzed two times, each time with the proposed damping ratios for each part, and the displacement profile is constructed as a synthesis of the occurring profiles from each corresponding part. Thus, the total response is obtained and also compared with the response that would occur should overall damping ratios equal to 2 % and 5 % be selected in Fig. 32, while in Fig. 33 the absolute error profiles are compared. A similar approach would be needed to compute accelerations.

Thus, it is evident that even the uniform damping ratios occurring from a trial and error procedure can be applied instead of the adoption of the actual damping distribution. Furthermore,

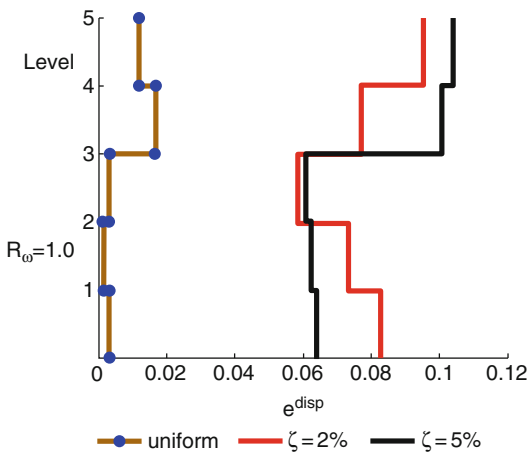
the uniform damping ratios are even closer to engineering practice and may be applied in a response spectrum analysis procedure, where the response occurs by using accordingly produced spectra taking into consideration the proposed damping ratios.

Summary

Despite the frequency with which irregular concrete-steel structures are encountered in engineering practice, there are still no code provisions regarding their seismic design. As a result, intense research efforts have been directed towards



Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 32 Displacements profile for a five-story irregular structure



Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis, Fig. 33 Error profile for a five-story irregular structure

supplying answers to the problems arising in such cases. These efforts, as presented here, have started providing results that are very useful for engineering practice, aiming at developing tools that can be used by structural analysis and design teams using available commercial software.

These results come in the form of decoupling criteria, aiming at helping engineers to decide whether they can proceed with a convenient decoupled analysis or not, while in the case that

the decoupling error is intolerable, equivalent damping ratios are proposed, modal as well as uniform ones. Still, further research is necessary in order to cover various issues, regarding, for example, structures that are torsionally sensitive or have nonuniform concrete and steel parts, so that higher vibration modes are significant to the overall response, or aiming to address the differences in the elastoplastic behavior of the two parts.

Cross-References

- [Classically and Nonclassically Damped Multi-degree of Freedom \(MDOF\) Structural Systems, Dynamic Response Characterization of](#)
- [European Structural Design Codes: Seismic Actions](#)
- [Incremental Dynamic Analysis](#)
- [Modal Analysis](#)
- [Nonlinear Dynamic Seismic Analysis](#)
- [Response Spectrum Analysis of Structures Subjected to Seismic Actions](#)
- [Seismic Analysis of Steel Buildings: Numerical Modeling](#)
- [Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling](#)
- [Structures with Nonviscous Damping, Modeling, and Analysis](#)
- [Time History Seismic Analysis](#)

References

- Bilbao A, Avilés R, Agirrebeitia J, Ajuria G (2006) Proportional damping approximation for structures with added viscoelastic dampers. *Finite Elem Anal Des* 42:492–502
- Chen G, Wu J (1999) Transfer-function-based criteria for decoupling of secondary systems. *J Eng Mech* 125(3):340–346
- Eurocode 8 (EC8) (2004) Design of structures for earthquake resistance
- Gupta AK, Tembulkar JM (1984a) Dynamic decoupling of secondary systems. *Nucl Eng Des* 81:359–373
- Gupta AK, Tembulkar JM (1984b) Dynamic decoupling of multiply connected MDOF secondary systems. *Nucl Eng Des* 81:375–383
- Huang BC, Leung AYT, Lam KL, Cheung VK (1996) Analytical determination of equivalent modal

- damping ratios of a composite tower in wind-induced vibrations. *Comput Struct* 59(2):311–316
- Papageorgiou AV, Gantes CJ (2005) Decoupling criteria for the dynamic response of primary/secondary structural systems. In: 4th European workshop on irregular and complex structures (4EWICS), Thessaloniki
- Papageorgiou AV, Gantes CJ (2008) Equivalent uniform damping ratios for irregular in height concrete/steel structural systems. In: 5th European conference on steel and composite structures (Eurosteel 2008), Graz, 3–5 Sept 2008, pp 1485–1490
- Papageorgiou AV, Gantes CJ (2008) Modal damping ratios for irregular in height concrete/steel structures. In: 5th European workshop on the seismic behavior of irregular and complex structures (5EWICS), Catania, 16–17 Sept 2008, pp 157–168
- Papageorgiou AV, Gantes CJ (2010a) Decoupling criteria for inelastic irregular primary/secondary structural systems subject to seismic excitation. *J Eng Mech* 136(10):1234–1247
- Papageorgiou AV, Gantes CJ (2010b) Equivalent modal damping ratios for concrete/steel mixed structures. *Comput Struct* 88(19–20):1124–1136
- Papageorgiou AV, Gantes CJ (2011) Equivalent uniform damping ratios for linear irregularly damped concrete/steel mixed structures. *Soil Dyn Earthquake Eng* 31(3):418–430
- Papagiannopoulos GA, Beskos DE (2006) On a modal damping identification model of building structures. *Arch Appl Mech* 76:443–463
- Papagiannopoulos GA, Beskos DE (2009) On a modal damping identification modal for non-classically damped linear building structures subjected to earthquakes. *Soil Dyn Earthquake Eng* 29:583–589

Modal Analysis

Nikolaos Nikitas¹, John Hugh George Macdonald² and Konstantinos Daniel Tsavdaridis¹

¹School of Civil Engineering, University of Leeds, Leeds, UK

²Civil Engineering Department, University of Bristol, Bristol, UK

Synonyms

Generalized coordinate analysis; Modal expansion; Mode acceleration method; Mode displacement method; Mode superposition method; Multi-degree-of-freedom (MDOF) system decoupling; Order reduction; Response series truncation

Introduction

The current entry attempts to synoptically guide the reader through the merits and critical details of modal analysis having a particular focus on earthquake engineering, which is the main subject of this encyclopedia. Tracking the history of modal analysis, the conception of vibration modes dates back to the eighteenth century and the pioneering studies and debates of Daniel and John Bernoulli, Euler, and d'Alembert (see Kline 1990) who while studying the problem of the vibration of a taut string came up with the notion of the modal shape contributions that build up the total observed oscillations. Such revolutionary for the time vibration knowledge along with a systematic treatment and practical extensions of modal analysis topics, more focused on continuous systems, appeared first in a very concerted way in the historical "Theory of Sound" by Lord Rayleigh (1877). Much later, today, the term modal analysis is used with different context in engineering language depending on the engineering stream and application ranging from the simple eigenvalue analysis, which recovers the normal mode frequencies and shapes, to the complex experimental dynamic properties identification. Its typical and closest to earthquake engineering definition refers to the ensemble of theoretical tools and operations, which are employed in simplifying the dynamic response calculation of a very large extended structural system under generic time-dependent loading. A much wider picture of modal analysis encompassing apart from the previous side a vast experimental background that has an invaluable contribution in understanding, enhancing, and controlling structures pertains primarily in mechanical engineering. This latter broad view was captured for the first time in a very holistic manner by a seminal book publication by Ewins (1984), which was later succeeded by similar influential titles like the ones by his students Maia et al. (1997) and He and Fu (2001).

The description that follows does not intend to be an alike broad exhaustive review of the so-called *experimental* or *operational modal analysis*, and the reader is suggested to refer to

the quoted titles for this purpose. The definition assumed herein is closer to the first solely analytical approach, and as such it was previously very effectively included in the fundamental dynamics handbooks by Biggs (1964), Clough and Penzien (1993), and Chopra (1995) or in latest educational dynamics literature like Rajasekaran (2009). There the main attributes, limitations, and benefits of expressing a discretized structure's response to some random or deterministic dynamic excitation (earthquakes being an ideal generic example) as the superposition of modal responses are discussed and exemplified through practical exercises. The influence of damping in the modal decomposition process along with the crucial question of how many modes are needed in accurately reproducing the full dynamic information hidden in response metrics are meticulously addressed for becoming a tool in the hands of practicing structural engineers. The presentation that follows collects and reproduces a synopsis of all this essential digested knowledge, which is of high priority for seismic analysis and design. This information directly links to other entries of the encyclopedia and allows their better understanding and connection. Although it aims to be a self-contained entry, it is founded on some previous knowledge of structural modeling and the dynamics of single-degree-of-freedom (SDOF) and multi-degree-of-freedom (MDOF) systems, which are presented assuming the reader was previously exposed to basic vibration textbooks.

The entry begins with the study of a typical linear undamped n -degree-of-freedom system (with the application of multistory buildings always in mind) under a common form of dynamic loading. This system, in a straightforward manner, is transformed to n independent/uncoupled generalized SDOF systems acted upon by certain fractions of the initially defined loading, and two different approaches are introduced for performing any further practical analysis. Subsequently, findings are extrapolated to a variously damped viscous system (the proportional or classical and the generic or nonclassical), which is a very useful and effective approximation of the energy dissipation mechanism even in

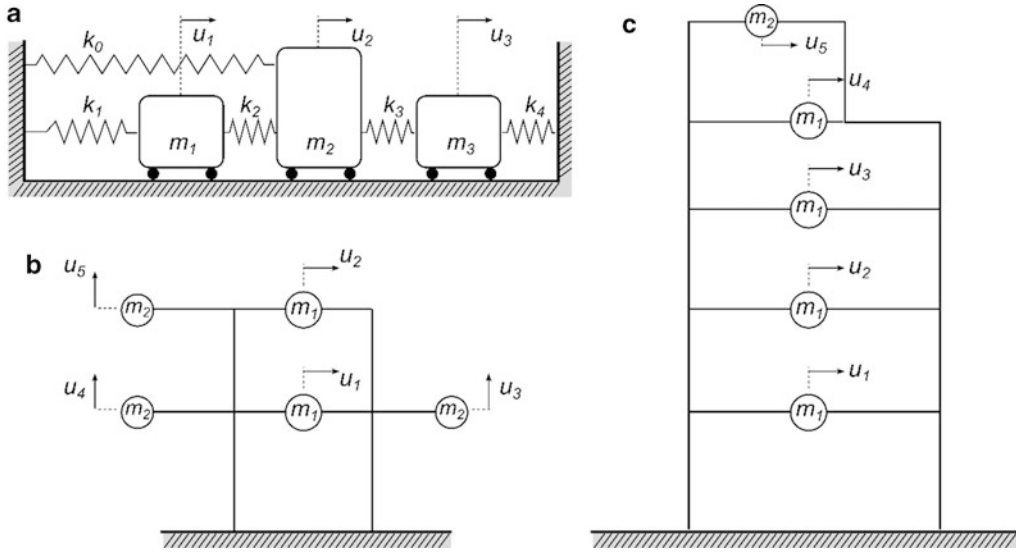
complex structures consisting of many interconnecting nonuniform parts. The alternatives of working either in the time or frequency domain are both discussed in brief. For the subsequent main topic of modal truncation that emerges, where one has to decide the minimum number of modes r ($r < n = \text{number of degrees of freedom}$) that suffice for a sufficiently accurate reduced representation of the dynamic solutions (which may be displacements, rotations, cumulative shear forces, moments, etc.), all main different approaches and practices are presented. Miscellaneous details on various alterations of the analysis (e.g., mode acceleration or mode displacement form) practiced for increasing its numerical efficacy together with connections to aseismic structural design and brief essential reference to valuable extensions (e.g., nonlinear systems, health monitoring, and control) are provided throughout.

Modal Analysis Basics

The Undamped/Conservative System

The topic of dynamic-degree-of-freedom (DOF) selection, which is typical when transforming continuous systems to discrete models, along with reference to *nonholonomic* systems (i.e., where independent motion coordinates are essentially accompanied by additional constraints for fully determining motion) is not pursued herein. For the presentation in hand, the simple form of the already discrete engineering models assumed yield such discussions unnecessary. Thus, for this point, imagine a simplified n -degree-of-freedom discrete undamped structural model relieved by any *gyroscopic effects*. Such can be the translational toy mechanical system in Fig. 1a or the sway multistory shear frames in Fig. 1b, c.

In all cases depicted in Fig. 1, the envisaged finite degrees of freedom (DOFs) necessary are clearly designated. For all systems additional assumptions (e.g., rigid beams, massless columns, and springs) were employed for reducing the required DOFs. The set of governing equations of motion may be put, by either following the force equilibrium approach on n free-body



Modal Analysis, Fig. 1 MDOF structural models (a) toy mechanical discrete system (b, c) multistorey sway shear frames

diagrams or the energy approach, in the matrix form

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{F}(t), \quad (1)$$

where $\mathbf{M}$ and $\mathbf{K}$ are the constant symmetric mass and stiffness matrices ($n \times n$) $\ddot{\mathbf{u}}(t)$ and $\mathbf{u}(t)$ are the column matrices or vectors ($n \times 1$) of accelerations and displacements expressed in physical coordinates varying with time t and $\mathbf{F}(t)$ the vector ($n \times 1$) of the externally applied (not motion-dependent) dynamic loads. As a convention, all matrices are herein denoted by bold letters, while common letters are used to denote scalar magnitudes. Regarding the stiffness matrix $\mathbf{K}$, we must notice that it encompasses both the elastic and geometric stiffness information of the structural model, e.g., forces causing buckling are assigned to geometric stiffness. Also the constant nature of $\mathbf{K}$ translates to a linear elastic operational regime. As it can be easily verified through all Fig. 1 examples, for $\mathbf{M}$ it is straightforward to make it diagonal using the so-called *lumped mass* form, where translational DOFs are defined on each discrete mass. Yet this, which is not always possible or practical, will generally result in a non-diagonal $\mathbf{K}$ and, thus, in statically coupled Eq. 1, with the term *statically coupled* meaning

that each of the constituting linear equations of motion includes components of the displacement vector $\mathbf{u}(t)$ corresponding to various DOFs and not only the DOF referred to the existing component of the acceleration vector $\ddot{\mathbf{u}}(t)$. This coupling is in fact a coordinate coupling that perplexes the solution of Eq. 1 especially in the common case of large n . Alternatively, a different formulation practiced in finite element (FE) derivations for either $\mathbf{M}$ or $\mathbf{K}$, entitled *consistent form*, leads in general to non-diagonal forms for both these matrices, thereby leading to *statically and dynamically* coupled equations of motion. This means that each one of these equations includes components of the vectors $\mathbf{u}(t)$ and $\ddot{\mathbf{u}}(t)$ corresponding to various DOFs and not only to one DOF. And exactly, modal analysis aims at decoupling these equations. If Eq. 1 gets uncoupled, this will yield n independent equations of motion, i.e., each one of which refers to only one DOF, with the result that the MDOF system described by Eq. 1 can be treated as n equivalent uncoupled SDOF systems. Each of them corresponds to a so-called natural mode of vibration, which is characterized by a particular frequency and shape (modal/eigen frequency ω_i and modal/eigen vector $\boldsymbol{\varphi}_i$, ($n \times 1$)). Such an advantageous simplification can be established

by transforming the vector $\mathbf{u}(t)$ into a new vector $\mathbf{q}(t)$, ($n \times 1$), whose components are called normal or principal coordinates, as well as generalized displacements. This transformation is equivalent to the separation of variables technique, exercised in partial differential equations (PDEs), and is described as below

$$\mathbf{u}(t) = \Phi \mathbf{q}(t) = \sum_{i=1}^n \boldsymbol{\varphi}_i q_i(t), \quad (2)$$

or alternatively for the displacement vector $\mathbf{u}_i(t)$ corresponding to each individual mode i

$$\mathbf{u}_i(t) = \boldsymbol{\varphi}_i q_i(t). \quad (3)$$

In Eqs. 2 and 3, Φ is the modal matrix ($n \times n$) consisting of all n mode shapes $\boldsymbol{\varphi}_i$ arranged as its columns, which includes the entire modal shape information. And $\mathbf{q}(t)$ is a vector (i.e., column matrix) of n components, with the arbitrarily chosen i th component $q_i(t)$ representing the unknown time-functioning generalized displacement for the i th mode.

The modal parameters of frequency ω_i and shape $\boldsymbol{\varphi}_i$ are deduced by solving the so-called eigenvalue problem. This comes up when replacing Eq. 3 inside the free/unforced counterpart of Eq. 1 resulting from Eq. 1 after setting $\mathbf{F}(t) = 0$ and recognizing the pure harmonic nature of the associated $\mathbf{q}(t)$. Specifically speaking, considering the free/unforced counterpart of Eq. 1, i.e., the free-vibration equations

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = 0, \quad (4)$$

and assuming that each modal response $\mathbf{u}_i(t) = \boldsymbol{\varphi}_i q_i(t)$ is a solution to Eq. 4, it follows

$$\mathbf{M}\boldsymbol{\varphi}_i \ddot{q}_i(t) + \mathbf{K}\boldsymbol{\varphi}_i q_i(t) = 0, \quad (5)$$

which after premultiplying with the transposed mode shape vector $\boldsymbol{\varphi}_i^T$, i.e., a row matrix with the same elements as the column matrix $\boldsymbol{\varphi}_i$, gives

$$\boldsymbol{\varphi}_i^T \mathbf{M} \boldsymbol{\varphi}_i \ddot{q}_i(t) + \boldsymbol{\varphi}_i^T \mathbf{K} \boldsymbol{\varphi}_i q_i(t) = 0. \quad (6)$$

Now, defining the frequency ω_i as equal to

$$\omega_i^2 = \frac{\boldsymbol{\varphi}_i^T \mathbf{K} \boldsymbol{\varphi}_i}{\boldsymbol{\varphi}_i^T \mathbf{M} \boldsymbol{\varphi}_i}, \quad (7)$$

Equation 6 becomes equal to the harmonic vibration in $q_i(t)$

$$\ddot{q}_i(t) + \omega_i^2 q_i(t) = 0, \quad (8)$$

and combining with Eq. 5 results in

$$(\mathbf{M} - \omega_i^2 \mathbf{K}) \boldsymbol{\varphi}_i = 0, \quad (9)$$

which has a nonzero solution $\boldsymbol{\varphi}_i$ only for the zero determinant

$$|\mathbf{M} - \omega_i^2 \mathbf{K}| = 0. \quad (10)$$

Equation 10 is also termed the characteristic or frequency equation of the system. As expected due to the problem's indeterminate nature, the resulting vectors $\boldsymbol{\varphi}_i$ are not unique solutions but represent families of solutions of the form $\alpha \boldsymbol{\varphi}_i$, with α being an arbitrary scalar constant. The computational effort/time required for a solution is proportional to n^3 . It is noticed that deriving such sets of ω_i , $\boldsymbol{\varphi}_i$ solutions for large structural systems have introduced a whole new field for devising efficient solvers of Eqs. 9 and 10 (i.e., eigensolvers); the curious reader may consult §8 (Rades 2010). One of the most renowned equivalent numerical techniques of great application belongs to Rayleigh and Ritz (for some historical notes and disputes on its origins, one can read Leissa (2005)). Further, it is vital to refer to the important orthogonality properties, viz.,

$$\boldsymbol{\varphi}_i^T \mathbf{K} \boldsymbol{\varphi}_j = \boldsymbol{\varphi}_i^T \mathbf{M} \boldsymbol{\varphi}_j = 0 \text{ for all } i \neq j, \quad (11)$$

which result from the symmetrical form of the mass and stiffness matrices $\mathbf{M}$ and $\mathbf{K}$, respectively, and allow Eq. 1 to be uncoupled into n SDOF equations. Indeed, premultiplying Eq. 1 with $\boldsymbol{\varphi}_i^T$ and replacing $\mathbf{u}(t)$ with $\Phi \mathbf{q}(t)$ according to Eq. 2, one gets

$$\boldsymbol{\varphi}_i^T \mathbf{M} \ddot{\mathbf{q}}(t) + \boldsymbol{\varphi}_i^T \mathbf{K} \mathbf{q}(t) = \boldsymbol{\varphi}_i^T \mathbf{F}(t). \quad (12)$$

Owing to the orthogonality conditions, Eq. 12 becomes

$$\boldsymbol{\varphi}_i^T \mathbf{M} \ddot{q}_i(t) + \boldsymbol{\varphi}_i^T \mathbf{K} \boldsymbol{\varphi}_i q_i(t) = \boldsymbol{\varphi}_i^T \mathbf{F}(t), \quad (12a)$$

or equally

$$M_i \ddot{q}_i(t) + K_i q_i(t) = P_i(t), \quad (12b)$$

which after division with M_i becomes

$$\ddot{q}_i(t) + \omega_i q_i(t) = \frac{P_i(t)}{M_i}, \quad (12c)$$

where

$$K_i = \boldsymbol{\varphi}_i^T \mathbf{K} \boldsymbol{\varphi}_i, \quad M_i = \boldsymbol{\varphi}_i^T \mathbf{M} \boldsymbol{\varphi}_i, \quad (13)$$

$$\omega_i = K_i/M_i \text{ and } P_i(t) = \boldsymbol{\varphi}_i^T \mathbf{F}(t)$$

with K_i , M_i , ω_i and $P_i(t)$ being the so-called scalar modal stiffness, mass, frequency, and force, respectively. Note that combining orthogonality with Eq. 9 the typical SDOF relation, $K_i = \omega_i^2 M_i$ comes up. The values of K_i and M_i are subject to the scaling opted for the modal vectors $\boldsymbol{\varphi}_i$ and as such on their own are not very informative of the whole system. A common scaling, or else referred to as normalization, used for $\boldsymbol{\varphi}_i$ is the one that results in unit modal masses. Namely, if each $\boldsymbol{\varphi}_i$ is divided by $1/\sqrt{M_i}$, then evidently the second of Eq. 13 yields

$$\frac{\boldsymbol{\varphi}_i^T}{\sqrt{M_i}} \mathbf{M} \frac{\boldsymbol{\varphi}_i}{\sqrt{M_i}} = I. \quad (14)$$

With all the previous data in hand, one can now proceed with solving the initial problem of determining the response of the undamped system to some generic forcing $\mathbf{F}(t)$. The complete solution to Eq. 12b, which is a generalized SDOF equation, provides the part of the total response that owes exclusively to the i th mode and can be solved independently of all other modes. As is well known from the solution to an SDOF equation, the modal solution $q_i(t)$ of Eq. 12c consists

of a free-vibration solution $q_i^c(t)$ due to exclusively the initial conditions $q_i(0)$ and $\dot{q}_i(0)$ and a particular forced-vibration solution $q_i^p(t)$ due to exclusively the imposed modal force $P_i(t)$. Namely, the resultant solution $q_i(t)$ in the time domain equals

$$q_i(t) = q_i^c(t) + q_i^p(t), \quad (15)$$

with

$$q_i^c(t) = q_i(0) \cos \omega_i t + \frac{\dot{q}_i(0)}{\omega_i} \sin \omega_i t, \quad (16)$$

and

$$q_i^p(t) = \frac{1}{\omega_i} \int_0^t \frac{P_i(\tau)}{M_i} \sin \omega_i(t - \tau) d\tau. \quad (17)$$

The forced-vibration solution $q_i^p(t)$ described in the time domain by Eq. 17 is the Duhamel integral expression, for the calculation of which a number of numerical techniques (e.g., Newton-Cotes formulas, Simpson's rule, Romberg integration, etc.) may be employed. It is worth pointing out that recovering the modal initial conditions $q_i(0)$ and $\dot{q}_i(0)$ from the initial conditions $\mathbf{u}(0)$ and $\dot{\mathbf{u}}(0)$ can be a straightforward result of Eq. 2 and the orthogonality properties (11). Namely, by multiplying Eq. 2 with $\boldsymbol{\varphi}_i^T \mathbf{M}$ and taking into account the orthogonality properties (11), it follows that

$$\begin{aligned} \boldsymbol{\varphi}_i^T \mathbf{M} \mathbf{u}(0) &= \boldsymbol{\varphi}_i^T \mathbf{M} \Phi \mathbf{q}(0) \\ &= \boldsymbol{\varphi}_i^T \mathbf{M} \boldsymbol{\varphi}_i q_i(0) = M_i q_i(0), \end{aligned} \quad (18a)$$

$$\boldsymbol{\varphi}_i^T \mathbf{M} \dot{\mathbf{u}}(0) = \boldsymbol{\varphi}_i^T \mathbf{M} \Phi \dot{\mathbf{q}}(0) = \boldsymbol{\varphi}_i^T \mathbf{M} \boldsymbol{\varphi}_i \dot{q}_i(0) = M_i \dot{q}_i(0), \quad (18b)$$

from which the magnitudes $q_i(0) = \boldsymbol{\varphi}_i^T \mathbf{M} \mathbf{u}(0)/M_i$ and $\dot{q}_i(0) = \boldsymbol{\varphi}_i^T \mathbf{M} \dot{\mathbf{u}}(0)/M_i$ are derived.

Finally, after solving Eqs. 16 and 17 for the unknowns $q_i^c(t)$ and $q_i^p(t)$, and then Eq. 15 for the unknown $q_i(t)$, Equation 2 will be used for synthesizing the total response time histories $\mathbf{u}(t)$ from the individual synchronous modal

contributions $\mathbf{u}_i(t)$, which have been determined as the products $\boldsymbol{\varphi}_i q_i(t)$ of the evaluated modal shapes $\boldsymbol{\varphi}_i$ and generalized displacements $q_i(t)$. This writes as

$$\mathbf{u}(t) = \sum_{i=1}^n \mathbf{u}_i(t) = \sum_{i=1}^n \boldsymbol{\varphi}_i q_i(t). \quad (19)$$

The most valuable feature of Eq. 19 is that not all n terms of the sum are required for accurately representing the total response. A large-scale structure like a tall building has hundreds of DOFs, which will produce hundreds of natural modes. Yet, it is common that for typical operational loading conditions, the lower (i.e., of lower frequencies) few modes, whose number r is far smaller than the number n of all the DOFs of the structure, are enough to compute the vector $\mathbf{u}(t)$ by using Eq. 19, for the higher (i.e., of higher frequencies) modes of a structure are not excited considerably. The value of r , as well as to what extent it can be reduced with respect to n , is a function not only of the initial loading $\mathbf{F}(t)$'s characteristics (i.e., shape distribution and frequency content) and their match to specific natural modes but also of the response metric that needs to be considered. For instance, representation of dynamic moment, shear, and stress variables in general requires more natural modes than representing the displacement solutions. This finding that was meticulously exemplified in Chopra (1995, 1996) and Clough and Penzien (1993) owes primarily to the fact that higher-order response attributes need additional information for their description (analogous to the added information embedded in derivatives of a function compared to the actual function).

The Classical/Proportional Damping Addition

The conservative (i.e., undamped, i.e., without energy losses) system treated earlier is fundamental for any dynamic study. Yet, it is not completely accurate in representing reality. All structures have the inherent ability of damping with which they can dissipate energy inputs. As a matter of fact, apart from being a structural property, effective damping can also surface from the

interaction of the structure with environmental or other types of loads. Distinctive examples of such cases, also described as *self-excitation*, are the Tacoma Narrows Bridge collapse (with the associated pure torsional flutter) and the large lateral vibrations of London Millennium Bridge (with pedestrians acting as negative dampers). The damping property is almost impossible to faithfully model ab initio. This is a result of a multitude of reasons. There is little information on inherent material damping, and further energy dissipation can originate from sources like connections, crack openings, or inter-element friction that are really hard to capture entirely. The practical and plausible solution is to model damping as a linear viscous term $\mathbf{C}\dot{\mathbf{u}}(t)$, i.e., function of velocity, which efficiently provides a route toward changing the energy balance in the equations of motion. This will lead to modifying the set of Eq. 1 as follows:

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{F}(t), \quad (20)$$

where $\mathbf{C}$ stands for the damping matrix ($n \times n$). This is in general non-diagonal. A work-around toward approximating the many unknown elements of $\mathbf{C}$ is to connect it to the straightforwardly derived structural matrices $\mathbf{K}$ and $\mathbf{M}$. The approximation of stiffness-proportional damping, which is actually appealing to intuition, can easily be shown to result in damping increasing with frequency. This is contrary to many experimental findings, which in general refute such a pattern. Mass-proportional damping on the other hand, which is simplistically linked to air/medium damping, also contradicts experience. However, combining the two proportionality approaches yields the so-called Rayleigh damping, which seems to constitute a good and much-used idealization for the energy dissipation mechanism of evenly distributed structures. A typical example of the latter is a multistory building of similar material and structural system throughout its height. According to the Rayleigh approach, $\mathbf{C}$ is given the form

$$\mathbf{C} = a_0 \mathbf{M} + a_1 \mathbf{K}, \quad (21)$$

where a_0 and a_1 are unknown scalar constants. Employing Rayleigh damping allows the classical modal decomposition (uncoupling) process earlier performed for the undamped case to apply to the damped case as well. Accordingly, one can firstly insert Eq. 2 in Eq. 20 and then multiply with the transposed mode shape vector $\boldsymbol{\varphi}_i^T$ derived from the undamped free-vibration Eq. 4 to get

$$\boldsymbol{\varphi}_i^T \mathbf{M} \Phi \ddot{\mathbf{q}}(t) + \boldsymbol{\varphi}_i^T (a_0 \mathbf{M} + a_1 \mathbf{K}) \Phi \dot{\mathbf{q}}(t) + \boldsymbol{\varphi}_i^T \mathbf{K} \Phi \mathbf{q}(t) = \boldsymbol{\varphi}_i^T \mathbf{F}(t). \quad (22)$$

Equation 22 due to Eqs. 11 and 13 becomes alike to Eq. 12

$$\boldsymbol{\varphi}_i^T \mathbf{M} \boldsymbol{\varphi}_i \ddot{q}_i(t) + \boldsymbol{\varphi}_i^T (a_0 \mathbf{M} + a_1 \mathbf{K}) \boldsymbol{\varphi}_i \dot{q}_i(t) + \boldsymbol{\varphi}_i^T \mathbf{K} \boldsymbol{\varphi}_i q_i(t) = \boldsymbol{\varphi}_i^T \mathbf{F}(t). \quad (22a)$$

Or alternatively

$$M_i \ddot{q}_i(t) + (a_0 M_i + a_1 K_i) \dot{q}_i(t) + K_i q_i(t) = P_i(t). \quad (22b)$$

In the last pair of equations, it should be noted that exclusively due to the fact that $\mathbf{C}$ was assumed to be a weighted sum of $\mathbf{M}$ and $\mathbf{K}$, orthogonality still applies and results in uncoupled modal damping expressions C_i for each mode i . Namely,

$$\begin{aligned} \boldsymbol{\varphi}_i^T \mathbf{C} \boldsymbol{\varphi}_j &= 0 \text{ for all } i \neq j \text{ and} \\ C_i &= a_0 M_i + a_1 K_i = \boldsymbol{\varphi}_i^T \mathbf{C} \boldsymbol{\varphi}_i. \end{aligned} \quad (23)$$

Equation 23 allows Eq. 22b to be rewritten as

$$M_i \ddot{q}_i(t) + C_i \dot{q}_i(t) + K_i q_i(t) = P_i(t), \quad (22c)$$

which after division with M_i becomes equal to

$$\ddot{q}_i(t) + 2\xi_i \omega_i \dot{q}_i(t) + \omega_i^2 q_i(t) = \frac{P_i(t)}{M_i}, \quad (22d)$$

with the modal mass and stiffness M_i and K_i , the modal frequency ω_i , and the modal force $P_i(t)$ defined by Eq. 13. As for the so-called

modal damping ratio (or fraction of the critical damping) ξ_i , it stands for the ratio $C_i/2\omega_i M_i$, i.e.,

$$\xi_i = C_i/2\omega_i M_i. \quad (24)$$

It is worth noting that Eq. 23 characterizes not only the Rayleigh damping model but also any damping model that offers the possibility for the classical modes of the undamped free vibration to be used in order to uncouple the equations of damped motion of the structure under consideration. Any such damping model is grouped as classical damping and substantially simplifies all subsequent dynamic calculations. Obviously, the subcases of mass-proportional and stiffness-proportional damping brought up earlier must also be categorized as classical damping.

Subsequently the process of calculating modal responses is identical to the previous description, where someone solves independently each modal equation and then synthesizes the total response by summing up the few lower modal contributions up to a certain order. The modal solution $q_i(t)$ of the Eq. 22 is now also affected by the modal damping as expressed by the modal damping ratio ξ_i . In analogy with Eqs. 15 up to 17, now the modal solution $q_i(t)$ generalizes to

$$\begin{aligned} q_i(t) &= q_i^c(t) + q_i^p(t) = \\ &= e^{-\xi_i \omega_i t} \left(q_i(0) \cos \tilde{\omega}_i t + \frac{\dot{q}_i(0) + \xi_i \omega_i q_i(0)}{\tilde{\omega}_i} \sin \tilde{\omega}_i t \right) \\ &\quad + \int_0^t \frac{P_i(\tau)}{M_i} h_i(t-\tau) d\tau, \end{aligned} \quad (25)$$

where $\tilde{\omega}_i$ is the so-called damped modal frequency, which stands for the multiple $\omega_i \sqrt{1 - \xi_i^2}$ of the undamped modal frequency ω_i , i.e.,

$$\tilde{\omega}_i = \omega_i \sqrt{1 - \xi_i^2}, \quad (26)$$

and $h_i(t - \tau)$ is the modal function known as unit-impulse response function given by

Modal Analysis, Table 1 Recommended damping values from Chopra (1995)

Type of structure	Damping ratio % (stress below ½ of yield)	Damping ratio % (stress at or just below yield)
Welded steel, prestressed concrete	2–3	5–7
Reinforced concrete (with considerable cracking)	3–5	7–10
Bolted steel, timber structures with bolted joints	5–7	10–15

$$h_i(t - \tau) = \frac{1}{\tilde{\omega}_i} e^{-\xi_i \omega_i (t - \tau)} \sin \tilde{\omega}_i (t - \tau). \quad (27)$$

In the above analysis, one step was tacitly skipped. When adopting the Rayleigh damping expression of Eq. 21, the unknown constants a_0 and a_1 need to be determined for the damping matrix to be fully described. Indicatively this can be done as follows:

The modal damping can be determined experimentally through analyzing response measurements. This is the exact inverse problem of what is dealt in this entry. In short an answer is enabled through different techniques, e.g., the half-power bandwidth method, the free decays of isolated modes, or some other more elaborate modal identification technique. In the absence of such experimental data, figures that have developed from long structural experience can be used. Such information is given in Table 1 noting the stress amplitude dependence.

With values for the modal damping ratios in hand, one can write Eq. 23 for two modes, customarily the first and the last one considered (i.e., r th), viz.,

$$\left. \begin{aligned} C_1 &= a_0 M_1 + a_1 K_1 = 2\omega_1 M_1 \xi_1 \\ C_r &= a_0 M_r + a_1 K_r = 2\omega_r M_r \xi_r \end{aligned} \right\}. \quad (28)$$

The latter system of two algebraic equations is then to be solved simultaneously for a_0 and a_1 . A further generalization of classical damping can

also be considered when more than two modal damping values are taken into account in capturing the damping distribution. Then we have the Caughey damping approach according to which C is expressed as

$$C = M \sum_{j=0}^{l-1} a_j (M^{-1} K)^j, \quad (29)$$

where l is the number of modal damping ratios to be used. Evidently for $l = 2$ this reduces to Rayleigh damping, while for $l = 3$ Eq. 29 results in

$$C = a_0 M + a_1 K + a_2 K M^{-1} K. \quad (30)$$

The l unknown constants are found again in a straightforward manner by solving a $l \times l$ system of algebraic equations.

The Nonclassical Damping Case

The classical damping approach although not physically realizable has found extensive use in engineering practice and particularly in standard earthquake engineering design. However, for structures that are heavily damped or unevenly distributed or even when they possess closely spaced modes, the couplings induced by the damping matrix may produce significant shortcomings and complications in the earlier presented modal analysis. An example of such a case, where nonclassical damping is to be expected, would appear in a nuclear reactor containment vessel founded on a soft soil. The quite different damping properties between the stiff structure and the foundation will induce a highly damped, very complex dynamic soil-structure interaction that establishes itself through a very generically shaped damping matrix (Clough and Mojtahedi 1976). For making things clear, in these instances the solution again starts from writing the damped equation of motion

$$M\ddot{u}(t) + C\dot{u}(t) + Ku(t) = F(t). \quad (20)$$

Still, what changes is that after substituting in the latter the modal decomposition indicated in Eq. 2

and premultiplying with the undamped mode vector $\boldsymbol{\varphi}_i^T$, one gets

$$\boldsymbol{\varphi}_i^T \mathbf{M} \boldsymbol{\varphi}_i \ddot{q}_i(t) + \boldsymbol{\varphi}_i^T \mathbf{C} \boldsymbol{\Phi} \dot{\mathbf{q}}(t) + \boldsymbol{\varphi}_i^T \mathbf{K} \boldsymbol{\varphi}_i q_i(t) = \boldsymbol{\varphi}_i^T \mathbf{F}(t) \quad (31a)$$

or alternatively

$$M_i \ddot{q}_i(t) + \boldsymbol{\varphi}_i^T \mathbf{C} \boldsymbol{\Phi} \dot{\mathbf{q}}(t) + K_i q_i(t) = P_i(t). \quad (31b)$$

Equation 31b obviously is a system of n -coupled equations, since the row vector $\tilde{\mathbf{C}} = \boldsymbol{\varphi}_i^T \mathbf{C} \boldsymbol{\Phi}$, ($1 \times n$), has in general nonzero elements that introduce in Eq. 31b apart from $\dot{q}_i(t)$ also terms $\dot{q}_j(t)$ with $j \neq i$. Thus, contrary to the principle and target of modal analysis, it is not possible to deal with n -independent generalized SDOF modal systems. It actually seems that there is no benefit in transforming the initial coupled equation of motion to its modal form using the undamped modes. There are a number of approaches to tackle this problem:

The first more simplistic approach disregards all the coupled contributions of Eq. 31b by essentially zeroing in the row vector $\tilde{\mathbf{C}}$ all elements apart from C_i , i.e., $\tilde{\mathbf{C}} = [0 \dots C_i \dots 0]$. Subsequently, the remaining analysis becomes identical to the above described for the classically damped case.

A different also approximate approach would be for one to keep all terms inside Eq. 31b and attempt to solve the system of coupled equations using numerical techniques, that is, direct integration (e.g., Newmark's method). More on the latter will appear in the entry of nonlinear systems. The relative benefit of employing Eq. 31b against applying direct integration to the initial n equations of damped motion described by the matrix Eq. 20 lies in the fact that due to generally few modes dominating response, a much reduced number of r equations can be considered. This is of great importance in large structural systems with thousands of DOFs.

Yet, both of the above analytical approaches may introduce errors of undefined magnitude. The most rigorous treatment of the problem is to uncouple the equations of motion using damped

modes instead of the undamped ones. Their derivation, indicatively provided herein, proceeds by first writing the free-vibration (unforced) counterpart of the forced-vibration matrix Eq. 20

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{0}. \quad (32)$$

And then, assuming a solution of the form

$$\mathbf{u}_i(t) = \boldsymbol{\phi}_i e^{\lambda_i t}, \quad (33)$$

Equation 32 leads to

$$(\lambda_i^2 \mathbf{M} + \lambda_i \mathbf{C} + \mathbf{K}) \boldsymbol{\phi}_i = \mathbf{0}, \quad (34)$$

where λ_i and $\boldsymbol{\phi}_i$ are the damped eigenvalues and eigenvectors, respectively.

The methodology to solve Eq. 34 relies on the so-called state-space formulation. According to it, Eq. 34 may be rewritten as

$$\begin{bmatrix} -\mathbf{K} & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{bmatrix} \begin{bmatrix} \lambda_i \boldsymbol{\phi}_i \\ \lambda_i^2 \boldsymbol{\phi}_i \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{K} \\ \mathbf{K} & \mathbf{C} \end{bmatrix} \begin{bmatrix} \boldsymbol{\phi}_i \\ \lambda_i \boldsymbol{\phi}_i \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad (35)$$

and introducing the block matrices $\mathbf{A}$ and $\mathbf{B}$, ($2n \times 2n$), and column block vector $\mathbf{v}_i$, ($2n \times 1$), i.e.,

$$\mathbf{A} = \begin{bmatrix} -\mathbf{K} & \mathbf{0} \\ \mathbf{0} & \mathbf{M} \end{bmatrix}, \mathbf{B} = \begin{bmatrix} \mathbf{0} & \mathbf{K} \\ \mathbf{K} & \mathbf{C} \end{bmatrix} \text{ and } \mathbf{v}_i = \begin{bmatrix} \boldsymbol{\phi}_i \\ \lambda_i \boldsymbol{\phi}_i \end{bmatrix}, \quad (36)$$

reshapes Eq. 35 into

$$(\lambda_i \mathbf{A} + \mathbf{B}) \mathbf{v}_i = \mathbf{0}, \quad (37)$$

which has a nonzero solution $\mathbf{v}_i$ only for the zero determinant

$$|\lambda_i \mathbf{A} + \mathbf{B}| = 0. \quad (38)$$

The latter is an eigenvalue problem of order $2n$, which is known as the complex eigenvalue problem owing to the fact that both λ_i and $\boldsymbol{\phi}_i$ are in general complex valued. Before proceeding any further, it is worth noticing that the increased

order $2n$ of the complex eigenvalue problem in comparison with the order n of the classical eigenvalue problem requires an increased computational effort/time proportional to $(2n)^3 = 8n^3$ for a solution. This is the reason for this approach to be outside the main pursuits of applied structural dynamics. Now, returning to Eq. 38, if λ_i is a solution, then its conjugate is also one. Typically, λ_i is expressed as

$$\lambda_i = \omega_i \left(-\zeta_i + j\sqrt{1 - \zeta_i^2} \right), \quad (39)$$

where $j = \sqrt{-1}$ is the imaginary unit and ω_i and ζ_i are the modal frequency and modal damping ratio, respectively. The complex nature of the eigenvectors ϕ_i accounts for phase differences other than 0° and 180° between their components. Orthogonality relations now become

$$\mathbf{v}_i^T \mathbf{A} \mathbf{v}_j = \mathbf{v}_i^T \mathbf{B} \mathbf{v}_j = 0 \text{ for all } i \neq j, \quad (40)$$

which allow the uncoupling of Eq. 35 or of its forced counterpart. Having laid the ground for subsequent analysis (e.g., time response solutions), no additional details are presented herein, and the reader is suggested to consult relevant fundamental literature on this special topic for additional resources and worked examples, e.g., (Crandall and McCalley 2002; Itoh 1973; Hansen et al. 2012).

Member Forces

The procedures described above yielded the displacement time response, both modal and total, of a linear elastic structure. Yet, it is of great relevance for design purposes to derive also the force and stress developing on individual members/elements. There are two distinct ways to achieve this:

1. Having in hand the modal displacements $\mathbf{u}_i(t)$ along the structure, one can use the member stiffness properties in order to derive the modal force $R_i(t)$ on the member and subsequently the required stresses. Considering all the modal contributions, one can write

$$R(t) = \sum_{i=1}^r R_i(t), \quad (41)$$

which apart from forces also applies to the resultant stresses.

2. Alternatively the modal displacements $\mathbf{u}_i(t)$ can be used to define equivalent modal static forces

$$\mathbf{F}_i^s(t) = \mathbf{K} \mathbf{u}_i(t) = \omega_i^2 \mathbf{M} \mathbf{u}_i(t). \quad (42)$$

Solving the structure statically for the obtained $\mathbf{F}_i^s(t)$, which are now applied as external forces at each time instant, will give again the modal member force $R_i(t)$. All individual modal contributions again have to be added together in accordance with Eq. 41.

The Earthquake Problem

The general modal analysis background presented above will now shift the focus to the earthquake-loading scenario of a classically damped n -DOF structure. Such a structure could be a typical multistory frame, e.g., Fig. 1c, which is customarily used as a benchmark in similar studies. Earthquake loading is the result of seismic action of the ground on the supports of the structure, that is, the result of a vibration motion of the supports of the structure caused by an earthquake. Specifically speaking, the seismic action translates to a ground acceleration input $\ddot{u}_g(t)$ at the foundation level (this is in most instances horizontal but can also be vertical), whose transmission to the upper structure creates an equivalent dynamic force distribution that constitutes what is called earthquake loading. Disregarding the special topic of multiple support excitation (for such an extension, consult Clough and Penzien (1993) or Chopra (1995)), the earthquake (or seismic) load vector that enters the matrix equation of motion, Eq. 20, can be expressed as

$$\mathbf{F}(t) = -\mathbf{M} \mathbf{u} \ddot{u}_g(t). \quad (43)$$

The latter implies a synchronous though spatially varying loading of the different lumped masses. The vector $\mathbf{u}$, ($n \times 1$), is an influence vector of the ground motion on the structure and represents the displacements that will result for each DOF when a static unit ground displacement $u_g(t) = 1$, of the same direction as $\ddot{u}_g(t)$, is applied to an undeformed (absolutely rigid) model of the structure. For example, in the frame of Fig. 1b, considering a horizontal earthquake input, $\mathbf{u}$ becomes $[1 \ 1 \ 0 \ 0 \ 0]^T$, while for the frame of Fig. 1c, it turns to $[1 \ 1 \ 1 \ 1 \ 1]^T$. For proof purposes, Eq. 43 and the definition of $\mathbf{u}$ can naturally be derived when expressing the equilibrium at each mass of the MDOF structure in terms of the total displacement $\mathbf{u}_t(t)$, which evidently is the vector sum of the relative-to-base displacement $\mathbf{u}(t)$ and the ground/base displacement $\mathbf{u}_g(t)$, i.e.,

$$\mathbf{u}_t(t) = \mathbf{u}(t) + \mathbf{u}_g(t). \quad (44)$$

Indeed, the equilibrium of the masses of a seismically excited MDOF structure is expressed by

$$\mathbf{M}\ddot{\mathbf{u}}_t(t) + \mathbf{C}\dot{\mathbf{u}}_t(t) + \mathbf{K}\mathbf{u}_t(t) = 0, \quad (45)$$

and substituting $\ddot{\mathbf{u}}_t(t)$ in accordance with Eq. 24, it follows

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = -\mathbf{M}\mathbf{u}\ddot{u}_g(t), \quad (46)$$

which proves that the seismic load vector $\mathbf{F}(t)$ equals $-\mathbf{M}\mathbf{u}\ddot{u}_g(t)$, thereby proving Eq. 32.

In a broader perspective, the expression for the seismic load can be generalized to describe a wider family of loading cases for which

$$\mathbf{F}(t) = \mathbf{s}f(t), \quad (47)$$

with the vector $\mathbf{s}$, ($n \times 1$), embodying all spatial variations and the scalar function $f(t)$ embodying all synchronous time variations. Trivially for $\mathbf{s} = -\mathbf{M}\mathbf{u}$ and $f(t) = \ddot{u}_g(t)$, one reverts to the earthquake-related problem.

This is actually the starting point for any subsequent analysis intending to obtain response histories. Yet, it must be noted that for most usual

practices in seismic calculations, a different characterization for the earthquake motion is opted. This is the instrumental approach of the so-called response spectra and attempts to remove the tedious numerical burdens that are related to obtaining all the history of the response of a specific structure to the long time sequence of the ground acceleration of a specific earthquake excitation in order that the maximum value of the response is evaluated. Namely, with response spectra, one only has to consider the maximum response output (this may be in terms of displacement, S_d , or velocity, S_v , or acceleration, S_a) that an SDOF system produces under a specific earthquake excitation. Thus, an earthquake is no more an explicit time function $\ddot{u}_g(t)$. It is described through its maximum effect on SDOF systems with specific natural frequency and damping. This approach best serves the engineering need for evaluating the strength of a structure against the maximum value of its response to a given seismic excitation.

Whatever the earthquake definition assumed and the analysis opted (time histories or response maxima), there are substantial benefits associated with the use of modal analysis. Yet, the inherent in response spectra need for a decomposition of an initial MDOF structure to equivalent SDOF ones makes modal analysis in that instance more than a beneficial feature an irreplaceable prerequisite.

Excitation and Participation Factors

In what follows, some essential notions relevant to the earthquake analysis of structures are developed through exemplifying the modal analysis application. Combining Eqs. 13 and 47 yields

$$\begin{aligned} P_i(t) &= \boldsymbol{\varphi}_i^T \mathbf{F}(t) = \boldsymbol{\varphi}_i^T \mathbf{s}f(t) \\ &= L_i f(t) \quad \text{with } L_i = \boldsymbol{\varphi}_i^T \mathbf{s}, \end{aligned} \quad (48)$$

as well as

$$\begin{aligned} \frac{P_i(t)}{M_i} &= \frac{\boldsymbol{\varphi}_i^T \mathbf{s}}{M_i} f(t) = \frac{L_i}{M_i} f(t) \\ &= \Gamma_i f(t) \quad \text{with } \Gamma_i = \frac{L_i}{M_i}, \end{aligned} \quad (49)$$

and, hence,

$$\Gamma_i = \frac{L_i}{M_i} = \frac{\boldsymbol{\varphi}_i^T \mathbf{s}}{M_i}. \quad (50)$$

The modal factors L_i and Γ_i are termed the *modal excitation factor* and the *modal participation factor*, respectively. It is easy to verify that the values of both the factors Γ_i and L_i depend on the normalization used for $\boldsymbol{\varphi}_i$ (since the ratio L_i/Γ_i equals the modal mass $M_i = \boldsymbol{\varphi}_i^T \mathbf{M} \boldsymbol{\varphi}_i$).

In view of Eq. 50 and the orthogonality conditions, it holds true

$$\boldsymbol{\varphi}_i^T \mathbf{s} = \Gamma_i M_i = \Gamma_i \boldsymbol{\varphi}_i^T \mathbf{M} \boldsymbol{\varphi}_i = \boldsymbol{\varphi}_i^T \sum_{j=1}^n \Gamma_j \mathbf{M} \boldsymbol{\varphi}_j, \quad (51)$$

which allows expanding of the vector $\mathbf{s}$ in terms of the modal contributions $\mathbf{s}_i$ as follows:

$$\mathbf{s} = \sum_{i=1}^n \mathbf{s}_i = \sum_{i=1}^n \Gamma_i \mathbf{M} \boldsymbol{\varphi}_i. \quad (52)$$

Evidently, the vector $\mathbf{s}$ and its modal components $\mathbf{s}_i$ are independent of the absolute magnitudes of $\boldsymbol{\varphi}_i$. The modal factors L_i and Γ_i naturally emerge when performing the modal analysis steps indicated in Eqs. 22c and 22d. Namely, when substituting for the load form adopted in Eq. 47 and writing the i modal matrix equation of motion, one gets

$$M_i \ddot{q}_i(t) + C_i \dot{q}_i(t) + K_i q_i(t) = L_i f(t), \quad (53a)$$

or equally

$$\ddot{q}_i(t) + 2\omega_i \xi_i \dot{q}_i(t) + \omega_i^2 q_i(t) = \Gamma_i f(t). \quad (53b)$$

The rationale behind the modal decomposition of the load distribution $\mathbf{s}$ lies in introducing inertial modal load contributions. To explain this, the inertial force in the i th mode is given as $\mathbf{M}\ddot{\mathbf{u}}_i$ or, taking into account Eq. 3, as $\mathbf{M}\boldsymbol{\varphi}_i \ddot{q}_i(t)$. The $\mathbf{M}\boldsymbol{\varphi}_i$ pre-factor, thus, makes apparent the underlying inertial concept linking to Eq. 52. Γ_i evidently from Eq. 53b is a weighting factor that scales the

load fraction to be assigned to each specific mode. The obscurity of the normalization dependence for Γ_i hinders it, becoming a direct clear indicator of the relative significance of the i th mode.

Relevant to this, the most important and useful feature of the expansion employed in Eq. 52 is that the load distribution $\mathbf{s}_i$ produces response exclusive to mode i or else that the response at each mode i owes solely to $\mathbf{s}_i$. This can be easily derived by observation of Eq. 52. When multiplying it with $\boldsymbol{\varphi}_j^T$ and $f(t)$ to create what is defined as the modal load $P_j(t)$ one gets

$$\begin{aligned} P_j(t) &= \boldsymbol{\varphi}_j^T \mathbf{s} f(t) \\ &= \sum_{i=1}^n \boldsymbol{\varphi}_j^T \mathbf{s}_i f(t) = \sum_{i=1}^n \Gamma_i \boldsymbol{\varphi}_j^T \mathbf{M} \boldsymbol{\varphi}_i f(t). \end{aligned} \quad (54)$$

This expression due to orthogonality yields a nonzero value only for $j = i$.

Contribution Factors (Chopra's Physical Interpretation)

Subsequently, if one follows the above procedures, the ensuing earthquake response problem can be solved. Yet, as first developed by Chopra (1995), a transformation is put forward that can further assist in a physical interpretation of modal analysis relevant to earthquake engineering. Namely, due to the linearity of the structural system, if one assumes

$$q_i(t) = \Gamma_i D_i(t), \quad (55)$$

then the modal Eq. 53b can be replaced by the equivalent

$$\ddot{D}_i(t) + 2\omega_i \xi_i \dot{D}_i(t) + \omega_i^2 D_i(t) = f(t), \quad (56)$$

which describes a unit mass SDOF system, with frequency ω_i and damping ξ_i loaded by $f(t)$; the latter can be either a force or a ground acceleration without any distinction. Accordingly, $D_i(t)$ is but the i th modal response of the system to the loading $f(t)$. The contribution of the i th mode to response can then be given by Eq. 3 as

$$\mathbf{u}_i(t) = \boldsymbol{\varphi}_i q_i(t) = \Gamma_i \boldsymbol{\varphi}_i D_i(t). \quad (57)$$

Defining equivalent modal static forces $\mathbf{F}_i^s(t)$ as in Eq. 42

$$\begin{aligned} \mathbf{F}_i^s(t) &= \mathbf{K} \mathbf{u}_i(t) = \omega_i^2 \mathbf{M} \mathbf{u}_i(t) \\ &= \omega_i^2 (\mathbf{M} \Gamma_i \boldsymbol{\varphi}_i) D_i(t) = \mathbf{s}_i \omega_i^2 D_i(t), \end{aligned} \quad (58)$$

one can write for any dynamic modal response quantity $R_i(t)$

$$R_i(t) = R_i^s \omega_i^2 D_i(t). \quad (59)$$

In the latter R_i^s is the modal response quantity that results from static analysis of the structure under the influence of the vector $\mathbf{s}_i$ of static external forces. This clearly shows that within the realm of modal decomposition, any complicated dynamic analysis of a n -DOF structure reduces to the product of two simple steps. In the first one performs static calculations for deriving all the n different R_i^s and then multiplies with the response solutions $D_i(t)$ of the n different SDOF systems loaded by $f(t)$. The total response is evidently given by the modal summation illustrated in Eq. 41.

The above modal analysis allows the introduction of another set of quantities, named the *contribution factors* $\tilde{R}_i$. These similarly to the participation factors indicate the significance of each mode in the total response. Still, their added merit lies in the fact that they are dimensionless and relieved by the abstractness of normalization. To produce them, one needs to write Eq. 59 as

$$R_i(t) = \tilde{R}_i R^s \omega_i^2 D_i(t), \quad (60)$$

in which R^s is the total response to the vector $\mathbf{s}$ of static external forces, and $\tilde{R}_i$ is the contribution factor of the i th mode defined as

$$\tilde{R}_i = \frac{R_i^s}{R^s}. \quad (61)$$

It can easily be proved that

$$\sum_{i=1}^n \tilde{R}_i = 1, \quad (62)$$

and, hence, if all modes beyond the r first modes are truncated, the corresponding error e_r is

$$e_r = 1 - \sum_{j=1}^r \tilde{R}_j, \quad (63)$$

which is usually very small for the truncation beyond the first three or four modes.

Extensions and Miscellanea

There are a number of additional topics that were intentionally excluded from the above described backbone of modal analysis. A handful of them is further elaborated in what follows. The particular intention is not to fully capture the extensive available knowledge but rather motivate the reader to further pursue the vast capabilities and multifaceted merits associated with this unique dynamics tool.

Mode Acceleration Method

The entirety of the above approach is based on estimating modal displacements, which subsequently will combine to form the total response. This displacement-based approach that naturally acquires the characterization mode displacement method was physically shown to extend into the calculation of any response variable.

Although the mode displacement method was developed typically first, it is not the only option available. One of the most useful alternatives is the so-called mode acceleration method along with its implementation variations. In identical form, though founded on a different basis, this may also be found under the name of static correction method, and the origin of this naming will below become clear. It was first devised for an undamped system, and its main attribute is that instead of modal displacements, one can seek solutions in terms of only modal accelerations for an undamped system or in terms of modal accelerations and velocities for a damped system. A brief presentation is herein put forward following the derivations and critique proposed by Cornwell et al. (1983).

Starting from the classically damped modal equation Eq. 22d, one can rewrite it as

$$\begin{aligned} q_i(t) &= \frac{1}{\omega_i^2} \left(\frac{P_i(t)}{M_i} - \ddot{q}_i(t) - 2\xi_i \omega_i \dot{q}_i(t) \right) \\ &= \frac{P_i(t)}{K_i} - \frac{1}{\omega_i^2} \ddot{q}_i(t) - \frac{2\xi_i}{\omega_i} \dot{q}_i(t), \end{aligned} \quad (64)$$

which substituting in the modal superposition Eq. 19 yields

$$\mathbf{u}(t) = \sum_{i=1}^n \boldsymbol{\varphi}_i \left(\frac{P_i(t)}{K_i} - \frac{1}{\omega_i^2} \ddot{q}_i(t) - \frac{2\xi_i}{\omega_i} \dot{q}_i(t) \right) \quad (65a)$$

or equally

$$\begin{aligned} \mathbf{u}(t) &= \sum_{i=1}^n \boldsymbol{\varphi}_i \frac{\boldsymbol{\varphi}_i^T \mathbf{F}(t)}{K_i} \\ &\quad - \sum_{i=1}^n \boldsymbol{\varphi}_i \left(\frac{1}{\omega_i^2} \ddot{q}_i(t) + \frac{2\xi_i}{\omega_i} \dot{q}_i(t) \right). \end{aligned} \quad (65b)$$

A plausible replacement is now sought for the first term in Eq. 65b. Namely, taking into account the definition of the equivalent static displacement of the structure $\mathbf{u}^s(t)$ as $\mathbf{u}^s(t) = \mathbf{K}^{-1} \mathbf{F}(t)$, as well as the definition of the equivalent static generalized displacements $q_i^s(t)$ as $q_i^s(t) = P_i(t)/K_i$, and applying Eq. 19, one gets

$$\begin{aligned} \mathbf{K}^{-1} \mathbf{F}(t) &= \mathbf{u}^s(t) = \sum_{i=1}^n \mathbf{u}_i^s(t) \\ &= \sum_{i=1}^n \boldsymbol{\varphi}_i \frac{P_i(t)}{K_i} = \sum_{i=1}^n \boldsymbol{\varphi}_i \frac{\boldsymbol{\varphi}_i^T \mathbf{F}(t)}{K_i}. \end{aligned} \quad (66)$$

Thus, the first term of the right-hand member of Eq. 65b only accounts for the total equivalent static displacement, as shown by Eq. 66. This fact evidently transforms the total displacement solution expressed by Eq. 65b into

$$\begin{aligned} \mathbf{u}(t) &= \mathbf{K}^{-1} \mathbf{F}(t) \\ &\quad - \sum_{i=1}^n \boldsymbol{\varphi}_i \left(\frac{1}{\omega_i^2} \ddot{q}_i(t) + \frac{2\xi_i}{\omega_i} \dot{q}_i(t) \right). \end{aligned} \quad (67)$$

The latter relation gives this whole alternative solution process a unique physical meaning. To obtain the dynamic response of a structure, one needs to consider first the easy to derive equivalent static (or pseudostatic) displacement and, subsequently, fine-tune it by adding the dynamic effects as introduced by the modal acceleration and velocity responses of the n individual modes. The higher mode, which trivially translates to higher frequency, would impose a relatively faster quadratic convergence for the second term in the right-hand side of Eq. 67 (see ω_i^2 denominator). Thus, this would allow less modes to be considered when truncating for economy and practicality the full series expansion. The method actually even when modes higher than $r (\ll n)$ are discarded clearly considers partly their influence through their static function.

Evidently, for an undamped system, Eq. 67 reduces to

$$\mathbf{u}(t) = \mathbf{K}^{-1} \mathbf{F}(t) - \sum_{i=1}^n \boldsymbol{\varphi}_i \frac{1}{\omega_i^2} \ddot{q}_i(t) \text{ with } \xi_i = 0, \quad (67a)$$

where the dynamic contribution of the modal velocities has been canceled out and only remains the dynamic effect of modal accelerations, which justifies the name mode acceleration method.

It is worth noting that even for a damped system, what matters in structural analysis is the maximum displacement $\max \mathbf{u}$, which causes the maximum strain and, hence, the maximum stress, required for proportioning the structural members. As a rule of thumb, this $\max \mathbf{u}$ is taken for equal to a combination of the maximum displacements $\max \mathbf{u}_i$ of the three or four first modes of the structure, which necessitates that the corresponding synchronous modal velocities become zero, $\dot{\mathbf{u}}_i = 0$. Thus, the dynamic contribution of the modal accelerations to account for

the maximum value of the dynamic response of the damped system only remains.

The method is greatly superior in terms of numerical efficacy outperforming the conventional mode displacement method in all cases by consistently requiring reduced number of modes for accurate, similar minimal error, solutions. Further the mode acceleration solution was found to be more sensitive to damping by means that increasing uniformly the damping of all modes would lead to faster convergence rates of associated dynamic solutions on reduced modal information.

Nonlinearity

The most critical assumption for applying modal analysis lies in the validity of the superposition of modes adopted in Eq. 2. This in simple terms translates to linearity of the structural system. With nonlinear behavior, developing the pursuit for a response-independent stationary modal reference system is not straightforward. Thus, the economy in the analysis of linear structures achieved through the modal order reduction cannot be realized in the analysis of nonlinear structures. Interestingly, in most real-life cases, structures move substantially, yield, or just interact with their loading environment. In all such instances, nonlinear descriptions are needed to capture the underlying phenomena waiving the validity of most of the above analysis.

A generalized manifold-type similar concept of nonlinear normal modes (NNM) has recently developed within the nonlinear dynamics field. Yet, this should be seen by engineers as more of a mathematical intricacy rather than a practical tool credible to be used in ordinary analysis of typical high-order systems. In generic nonlinear structures (this is to distinguish from cases of linear subsystems connecting through nonlinear links), the strict approach consists of numerically integrating the n -coupled equations of motion concurrently. To this purpose the time-continuous matrix Eq. 20 is written in a variational form to enable a time-stepping approach to be used for calculating purposes. Namely, Eq. 20 changes to

$$\mathbf{M}\Delta\ddot{\mathbf{u}}_i + \mathbf{C}_i\Delta\dot{\mathbf{u}}_i + \mathbf{K}_i\Delta\mathbf{u}_i = \Delta\mathbf{F}_i, \quad (68)$$

where

$$\begin{aligned} \Delta\ddot{\mathbf{u}}_i &= \ddot{\mathbf{u}}_{i+1} - \ddot{\mathbf{u}}_i, \quad \Delta\dot{\mathbf{u}}_i = \dot{\mathbf{u}}_{i+1} - \dot{\mathbf{u}}_i, \\ \Delta\mathbf{u}_i &= \mathbf{u}_{i+1} - \mathbf{u}_i, \quad \Delta\mathbf{F}_i = \mathbf{F}_{i+1} - \mathbf{F}_i, \end{aligned} \quad (69)$$

with i representing the time-step index contrary to its earlier modal meaning, while $\mathbf{C}_i$ and $\mathbf{K}_i$ are nonlinear damping and stiffness matrices, respectively, whose elements are functions of the displacement $\mathbf{u}(t)$ and the velocity $\dot{\mathbf{u}}(t)$ and take on values depending on the considered time step. The index i would translate in terms of absolute time to $i\Delta t$, where Δt denotes the time step. Alternatively, the time index $i + 1$ can be used (i.e., this is the implicit vs the explicit method when i is used). The overall response calculation problem includes three unknowns $\ddot{\mathbf{u}}_i$, $\dot{\mathbf{u}}_i$, and $\mathbf{u}_i$ for which apart from Eq. 68 two more equations could be derived from different approximations for $\ddot{\mathbf{u}}_i$ and $\dot{\mathbf{u}}_i$. As a matter of fact, depending on these approximations, which link to an assumed variance of the $\ddot{\mathbf{u}}_i$ and/or $\dot{\mathbf{u}}_i$ variables during the time-step duration, a number of different methods have developed. The most broadly used for earthquake-related purposes is Newmark's β -method.

It is noteworthy that although methods like the latter would easily produce bounded solutions (i.e., stable, meaning results will not “blow up”), whilst the time-step Δt is accurately prescribed. As a matter of fact, this is the most important parameter for any numerical integration scheme and must be sufficiently small. A first approximation for the Δt value can come from the associated modal analysis pertaining in the linear operation regime of the structure. Namely, the highest linear mode would correspond to the lowest natural period T_n . Δt should be sufficiently smaller than this T_n value. It should be reminded that this imposes large numerical limitations and costs and that the inherent softening associated with yielding would increase T_n (on the other hand, hardening would have an adverse effect by further reducing T_n).

Frequency Domain Solutions

After bringing the structural system to its modal description equivalent, the solutions pursued whether in terms of modal displacements or in terms of modal accelerations and velocities were always expressed in the time domain. Considering the case of the classically damped system with periodic loading and focusing on the probably most significant part of the response, the forced or else for this case steady state, one may suggest some alternatives to Eqs. 17 and 25. The reason is that the Duhamel's integral that provide the steady-state time response involves the convolution operation between the applied load and the unit-impulse response function. This term tends to perplex calculations.

Without detailing proofs, the alternative representation is in the frequency domain and can easily surface when one assigns inside the SDOF a periodic loading of the form $P_i(t) = P_i^o e^{j\omega t}$, where P_i^o is a scalar amplitude and again $j = \sqrt{-1}$. Then by rearranging the generalized SDOF response becomes

$$q_i(t) = \frac{P_i^o e^{j\omega t}}{(K_i - \omega_i^2 M_i) + j\omega C_i}. \quad (70)$$

Dividing Eq. 70 with $P_i(t)$ in order to create an output over input ratio, one gets

$$\begin{aligned} H(\omega) &= \frac{q_i(t)}{P_i(t)} = \frac{\tilde{q}_i(\omega)}{\tilde{P}_i(\omega)} \\ &= \frac{1}{(K_i - \omega^2 M_i) + j\omega C_i}, \end{aligned} \quad (71)$$

where over dash stands for the Fourier transform. The $H(\omega)$ function depending only on ω is called the receptance frequency response function. If, instead of displacement response, acceleration or velocity were used for the numerator, one would have the accelerance and mobility frequency response function, respectively. $H(\omega)$ is the Fourier transform of the unit-impulse response function given in Eq. 27. Generalizing the concept for an MDOF system, this instead of a scalar becomes a matrix $\mathbf{H}(\omega)$ with components

$h_{kl}(\omega)$, denoting the ratio of displacement at position k and force at position l

$$h_{kl}(\omega) = \sum_{i=1}^n \frac{\varphi_i(k)\varphi_i(j)}{M_i(\omega_i^2 - \omega^2) + j\omega C_i}. \quad (72)$$

Subsequently the convolution in the time domain is replaced in the frequency domain from a simple product of the frequency response function with the input force.

Operational Modal Analysis

To this point the presentation was focused on the so-called direct problem, which is concerned with obtaining the dynamic response of a fully known system to a given loading. By fully known it is meant that all the necessary properties to describe its behavior through a model, mainly an FE model, are in hand. Such can be the modal properties instead of the full stiffness, damping, and mass matrices. This in fact can bring a vast simplification and data economy having merits even more exaggerated than the ones modal truncation introduced before. Indicatively, for an n -DOF system, one, instead of determining $3n^2$ matrix terms, could obtain similar results by employing only a reduced number of modal frequency, damping, and shape sets. Still, this a priori assumed knowledge is not always the case. There are many uncertainties assigned to modeling (e.g., links' behavior, foundations, damping, etc.) which can bring up substantial errors in any associated model parameters and subsequently response predictions, used either for design, assessment, or control of the structure. To deal with this significant engineering issue, a major part of modal analysis nowadays has been devoted to the so-called inverse analysis. This translates to a mapping of the measured response to the system modal characteristics, and it first appeared with the development of the space program of the USA back in the 1950s. Interestingly, although the direct problem is a single-valued one, the inverse is not. The latter brings up the need for optimizing any of the prediction/analysis routines that will eventually

produce modal parameter outputs from measured response data.

The equation-free description that follows leaves aside the very interesting and demanding experimental (including signal preprocessing) techniques devised through the years to obtain the measured response from a structure under different forms of loading. The main scope of this entry is to only refer briefly to some of the most typical available tools able to obtain modal estimates once measured data are available. The mathematical background necessary to fully explain these methods are superseding the space limitations of this entry. Thus, the reader is motivated to also direct to the very insightful and broad §4 of Maia et al. (1997) as well as to a true encyclopedia on the topic by Ljung (1999) for further studying. Excellent software can be found to practice inverse analysis, some even in freeware form, which encompasses the full information provided herein and beyond. Such availability is a true late year's addition.

Practically all useful inverse analysis methods refer to MDOF systems. For them one may have information on their structural excitation or not. Their difficulty ranges from the simple peak picking method, where one observes peaks in the frequency response functions, to much more demanding alternatives employing complex algebra (Hankel matrices, Markov parameters, etc.) together with the earlier introduced state-space formulation. Most of the MDOF methods break down the structure to a superposition of independent-uncoupled SDOF systems, making even more profound the linearity and modal decomposition demands. In general, the strict target is to fit an analytical function of the form introduced in Eq. 72 to the measured data, which are customarily transformed to frequency response functions. This fitting operation inherently involved in the process lends the name curve fitting to the concept. Depending on the specific algorithm opted for the fitting, one may get the circle fitting, rational function polynomial, or many other refined approaches along the same lines.

Alternatively, one, instead of working in the frequency domain, may work employing the

time counterpart of frequency response functions, the impulse response functions. Typically the complex exponential method is used for such fittings. A similar very widespread approach fitting free decays instead of impulse response functions is named after Ibrahim, i.e., Ibrahim time domain method. Alternatively time-based methods like the option of the eigen-realization algorithm which will recover families of systems with identical eigenvalues to the measured one or generic time series tools like the autoregressive moving average can further be employed.

In all cases there is a major parameter affecting results. This is the order of the assumed model in the identification analysis. Namely, earlier treating the direct problem, one had to decide the order to which the modal solutions were to be truncated. The order that should be decided for the number of modes to be identified here is the direct equivalent.

Summary

Although every effort was made to synoptically present all the critical points that could not be missed of any educational work addressing modal analysis, the reader may still find omissions. To this reason, the wealth of references provided throughout is an excellent resource to complement the current entry as a further reading that could cover practically every aspect relating to the subject.

To recapitulate the main attributes presented herein, a short mind map is put forward that can illustrate the simple steps involved in modal analysis when employed in practical earthquake engineering purposes. This consists of:

- Determining the structural matrices $\mathbf{K}$, $\mathbf{M}$, and $\mathbf{C}$
- Determining the modal natural frequencies ω_i and shapes $\boldsymbol{\varphi}_i$
- Computing each modal-generalized response q_i (or $\ddot{q}_i$ and $\dot{q}_i$) and turning it to modal displacement $\mathbf{u}_i$

- Calculating the total displacement vector $\mathbf{u}$ after deciding the minimum number of modes r needed for an accurate solution

References

- Biggs JM (1964) Introduction to structural dynamics. McGraw Hill, New York
- Chopra AK (1995) Dynamics of structures: theory and applications to earthquake engineering. Prentice Hall, Englewood Cliffs
- Chopra AK (1996) Modal analysis of linear dynamics systems: physical interpretation. *J Struct Eng* 122(5):517–527
- Clough RW, Mojtahedi S (1976) Earthquake response analysis considering non-proportional damping. *Earthq Eng Struct Dyn* 4:489–496
- Clough RW, Penzien J (1993) Dynamics of structures, 2nd edn. McGraw Hill, New York
- Cornwell RE, Craig RR Jr, Johnson CP (1983) On the application of the mode-acceleration method to structural engineering problems. *Earthq Eng Struct Dyn* 11:679–688
- Crandall SH, McCalley RB Jr (2002) Chapter 28, Part I, matrix methods of analysis. In: Harris CM, Piersol AG (eds) Harris' shock and vibration handbook. McGraw-Hill, New York
- Ewins DJ (1984) Modal testing: theory and practice. Research Studies Press, Somerset
- Hansen C, Snyder S, Qiu X, Brooks L, Moreau D (2012) Active control of noise and vibration. CRC Press, Boca Raton
- He J, Fu ZF (2001) Modal analysis. Butterworth Heinemann, Oxford
- Itoh T (1973) Damped vibration mode superposition method for dynamic response analysis. *Earthq Eng Struct Dyn* 2:47–57
- Kline M (1990) Mathematical thought from ancient to modern times, vol II. Oxford University Press, Oxford, UK
- Leissa AW (2005) The historical bases of the Rayleigh and Ritz methods. *J Sound Vib* 287:961–978
- Ljung L (1999) System identification – theory for the user, 2nd edn. Prentice Hall, New Jersey
- Lord R (1877) The theory of sound, vol I. Macmillan, UK (reprinted 1945 by Dover Publications, New York)
- Maia NMM, Silva JMM, He J, Lieven NAJ, Lin RM, Skingle GW, To WM, Urgueira APV (1997) Theoretical and experimental modal analysis. Research Studies Press, Somerset
- Rades M (2010) Mechanical vibrations II, structural dynamic modelling. Editura Printech, Bucharest
- Rajasekaran S (2009) Structural dynamics of earthquake engineering: theory and application using MATHEMATICA and MATLAB. Woodhead, Cambridge, UK

Model Class Selection for Prediction Error Estimation

E. Simoen¹, C. Papadimitriou² and Geert Lombaert¹

¹Department of Civil Engineering, KU Leuven, Leuven, Belgium

²Department of Mechanical Engineering, University of Thessaly, Volos, Greece

Synonyms

Bayesian error estimation; Bayesian model class updating; Likelihood function

Definitions

Model class selection Determining the most suitable model class based on the available observed data, by applying Bayesian inference at the model class level.

Prediction error The discrepancy between the observed data and the predictions made by a numerical model. In Bayesian model updating, the likelihood function is constructed as the probability density function of the prediction error.

Introduction

In Bayesian model updating, probability density functions (PDFs) of model parameters are updated based on prior knowledge and information contained in experimental data (Beck and Katafygiotis 1998). This requires the construction of the *likelihood function*, i.e., the PDF of the experimental data, or, equivalently, the PDF of the error between model predictions and observed data. Most often, a fixed zero-mean uncorrelated Gaussian prediction error (i.e., “white noise”) is assumed or presumed given; however, in many engineering applications, the magnitude of the true prediction error is unknown. Moreover, a white noise model may

not always be appropriate, for instance, when the errors show spatial or temporal correlations or a systematic component. In such cases, one can make use of the available observed data to try and estimate (characteristics of) the prediction error, provided sufficient data are at hand. Bayesian model class selection (MCS) (Beck and Yuen 2004) constitutes one of the most effective tools to this end.

Bayesian Model Updating and Prediction Error

The objective of model updating (often also referred to as *parameter estimation*) is to calibrate unknown system properties which appear as parameters in numerical models, based on actually observed behavior of the system of interest. In Bayesian model updating, this is performed in a probabilistic uncertainty quantification framework: PDFs representing the uncertainty on the model parameters are updated through the experimental data; this procedure is described briefly below.

Uncertainty in Model Updating

In general terms, a model $\mathcal{M}_M(\boldsymbol{\theta}_M)$ belonging to the model class $\mathcal{M}_M$ provides a mapping from the parameters $\boldsymbol{\theta}_M \in \mathbb{R}^{N_M}$ to an output vector $\mathbf{G}_M(\boldsymbol{\theta}_M) \in \mathbb{R}^N$ through the transfer operator $\mathbf{G}_M$:

$$\mathcal{M}_M(\boldsymbol{\theta}_M) : \mathbb{R}^{N_M} \rightarrow \mathbb{R}^N : \boldsymbol{\theta}_M \mapsto \mathbf{G}_M(\boldsymbol{\theta}_M) \quad (1)$$

In structural engineering, the prediction model class usually involves a mechanical model such as a finite element model. The model parameters $\boldsymbol{\theta}_M$ are typically properties directly connected to (sub)structure stiffness or mass (e.g., Young's modulus, moment of inertia, mass density), material properties, or parameters describing connections and boundary conditions. The model outputs $\mathbf{G}_M(\boldsymbol{\theta}_M)$ should correspond to the observed system behavior, which can consist of directly measured structural responses (displacements, accelerations) or derived system properties such as modal data (natural frequencies, mode shapes, modal flexibilities, modal

curvatures, modal strain energies). Most often, the latter are preferred: modal data are rich in information content and can be obtained in an *operational* state of the structure, meaning that no external excitation needs to be applied or known. This also explains why, for the cases considered in this text, no input vector $\mathbf{x}$ is present in Eq. 1.

In the ideal case, with a perfect model and perfect measurements, the model output $\mathbf{G}_M(\boldsymbol{\theta}_M)$ corresponds perfectly to the true system output $\mathbf{d}$, so that $\mathbf{G}_M(\boldsymbol{\theta}_M) = \mathbf{d}$. This is the underlying assumption in deterministic parameter identification, where the objective is to determine the model parameters $\boldsymbol{\theta}_M$ for a given set of observed system outputs $\mathbf{d}$. In real applications, however, the numerical model is never capable of perfectly representing the behavior of the true physical system; in other words, a modeling error is always present. This error can be described as the discrepancy between the model predictions $\mathbf{G}_M(\boldsymbol{\theta}_M)$ and the true system output $\mathbf{d}$, i.e., $\boldsymbol{\eta}_G = \mathbf{d} - \mathbf{G}_M(\boldsymbol{\theta}_M)$.

As the true system output has to be measured and processed experimentally, the data $\mathbf{d}$ are also imperfect: they are always subject to measurement error, resulting in a discrepancy between the true system output $\mathbf{d}$ and the actually observed data $\bar{\mathbf{d}}$. This difference is defined as the measurement error $\boldsymbol{\eta}_D = \bar{\mathbf{d}} - \mathbf{d}$. Eliminating the unknown true system output $\mathbf{d}$ from the error equations and collecting both errors on the right-hand side of the equation yield

$$\bar{\mathbf{d}} - \mathbf{G}_M(\boldsymbol{\theta}_M) = \boldsymbol{\eta}_G + \boldsymbol{\eta}_D = \boldsymbol{\eta} \quad (2)$$

The sum of both errors is the difference between the model predictions and the observed quantities and is defined as the total observed prediction error $\boldsymbol{\eta}$. The above expression serves as a starting point for the Bayesian uncertainty quantification method.

Bayesian Model Updating Methodology

The general principle of Bayesian parameter estimation (Beck and Katafygiotis 1998) is that uncertainties in the model parameters

$\theta_M \in \mathbb{R}^{N_M}$ that parameterize the model class $\mathcal{M}_M$ are quantified by probability density functions (PDFs), which are updated in an inference scheme based on the available information. Measurement and modeling uncertainty are taken into account by modeling the respective errors as random variables: PDFs are appointed to η_G and η_D , which are parameterized by parameters $\theta_G \in \mathbb{R}^{N_G}$ and $\theta_D \in \mathbb{R}^{N_D}$. These parameters are added to the structural model parameters θ_M to form the general model parameter set $\theta = \{\theta_M, \theta_G, \theta_D\}^T \in \mathbb{R}^{N_\theta}$. This in fact corresponds to adding two probabilistic model classes to the structural model class $\mathcal{M}_M$ to form a *joint* model class $\mathcal{M} = \mathcal{M}_M \times \mathcal{M}_G \times \mathcal{M}_D$, parameterized by θ . This procedure of probabilistic modeling is commonly referred to as *stochastic embedding* (Beck 2010).

It is important to note that in Bayesian statistics, the *Bayesian* interpretation of probability is pertained as opposed to the classical *frequentist* interpretation. In the frequentist interpretation, the probability attributed to a random variable is seen as a measure for the long-term frequency of occurrence of that variable. The Bayesian interpretation is a subjective interpretation where probability reflects a measure of plausibility or degree of belief attributed to a variable, given the current state of information. It is clear that only the Bayesian interpretation is meaningful in the context of forming inferences on model parameters using observed data.

To express the updated probabilities of the unknown parameters θ , given some observations $\bar{\mathbf{d}}$ and a certain joint model class $\mathcal{M}$, Bayes' theorem is used:

$$p(\theta|\bar{\mathbf{d}}, \mathcal{M}) = c p(\bar{\mathbf{d}}|\theta, \mathcal{M}) p(\theta|\mathcal{M}) \quad (3)$$

where $p(\theta|\bar{\mathbf{d}}, \mathcal{M})$ is the updated or posterior PDF of the model parameters given the measured data $\bar{\mathbf{d}}$ and the assumed model class $\mathcal{M}$; c is a normalizing constant that ensures the posterior PDF integrates to one; $p(\bar{\mathbf{d}}|\theta, \mathcal{M})$ is the PDF of the observed data given the parameters θ ; and $p(\theta|\mathcal{M})$ is the initial or prior PDF of the parameters. In the following, the explicit dependence on the model class $\mathcal{M}$ is omitted in order to simplify the notations.

The prior PDF $p(\theta)$ quantifies the uncertainty associated with model parameters θ in the absence of measurement results. In many cases, the prior PDF is chosen based on engineering judgment or on computational tractability (e.g., conjugate priors (Diaconis and Ylvisaker 1979)). Except in cases where a large amount of data is at hand, the prior PDF has a significant influence on the Bayesian updating results. A wide range of methods has been developed to obtain "objective" prior PDFs based on the given prior information. One of the most commonly used approaches in this respect is the method based on the maximum entropy (ME) principle (Jaynes 1957; Soize 2008), which determines the prior PDF that, given the current state of prior information, results in maximum entropy.

The PDF of the experimental data $p(\bar{\mathbf{d}}|\theta)$ can be interpreted as a measure of how good a model succeeds in explaining the observations $\bar{\mathbf{d}}$. As this PDF reflects the likelihood of observing the data $\bar{\mathbf{d}}$ when the model is parameterized by θ , it is also referred to as the *likelihood* function $L(\theta|\bar{\mathbf{d}})$. Since the data set $\bar{\mathbf{d}}$ is fixed, this function in fact no longer represents a conditional PDF and can be denoted as $L(\theta; \bar{\mathbf{d}})$; in the following, however, the common notation of $L(\theta|\bar{\mathbf{d}})$ is pertained. The likelihood function is determined according to the total probability theorem in terms of the probabilistic models of the measurement and modeling errors:

$$p(\bar{\mathbf{d}}|\theta) \equiv L(\theta|\bar{\mathbf{d}}) = \int_{\mathbb{R}^N} p_{\bar{\mathbf{d}}}(\bar{\mathbf{d}}|\theta, \mathbf{d}) p_{\mathbf{d}}(\mathbf{d}|\theta) d\mathbf{d} \quad (4)$$

$$= \int_{\mathbb{R}^N} p_{\eta_D}(\bar{\mathbf{d}} - \mathbf{d}|\theta_D) p_{\mathbf{d}}(\mathbf{d}|\theta) d\mathbf{d} \quad (5)$$

$$= \int_{\mathbb{R}^N} p_{\eta_D}(\bar{\mathbf{d}} - \mathbf{d}|\theta_D) p_{\eta_G}(\mathbf{d} - \mathbf{G}_M(\theta_M)|\theta_G) d\mathbf{d} \quad (6)$$

where $p_{\eta_D}(\bar{\mathbf{d}} - \mathbf{d}|\theta_D)$ corresponds to the probability of obtaining a measurement error η_D , given the PDF of η_D parameterized by θ_D , and where $p_{\eta_G}(\mathbf{d} - \mathbf{G}_M(\theta_M)|\theta_G)$ represents the probability of obtaining a modeling error η_G when the PDF

of $\boldsymbol{\eta}_G$ is known and parameterized by $\boldsymbol{\theta}_G$. Here, it is implicitly assumed that the modeling error and measurement error are independent variables.

The above equations show that the likelihood function can be computed as the convolution of the PDFs of the measurement and modeling error. In most realistic applications, however, no distinction can be made between measurement and modeling error. In those cases, the likelihood function can be constructed using the probabilistic model of the total prediction error $\boldsymbol{\eta}$, parameterized by $\boldsymbol{\theta}_\eta$:

$$p(\bar{\mathbf{d}}|\boldsymbol{\theta}) \equiv L(\boldsymbol{\theta}|\bar{\mathbf{d}}) = p(\boldsymbol{\eta}|\boldsymbol{\theta}_\eta) \quad (7)$$

Even though the likelihood function has a significant influence on the model updating results, its construction has received much less attention than the construction of the prior PDF. This is mainly due to the fact that most often very little or no information is at hand regarding the characteristics of the error(s); only in selected cases, a realistic estimate can be made concerning the probabilistic model representing the prediction error, for instance, based on the analysis of measurement results. Most often, it is simply assumed that the probabilistic model of the prediction error is known and fixed, so that the parameter set reduces to $\boldsymbol{\theta} = \{\boldsymbol{\theta}_M\} \in \mathbb{R}^{N_M}$.

In structural mechanics applications, an uncorrelated zero-mean Gaussian model (i.e., “white noise”) with a fixed and constant variance is typically selected to represent the prediction error. However, this common assumption is not necessarily justified or realistic. For instance, in most practical applications, the magnitude (or variance) of the errors is not (precisely) known. Moreover, a white noise model might not be appropriate to represent the true prediction error, as significant spatial or temporal correlations between the modeling error components are often likely to be present. Consider, for instance, the case where mode shapes need to be predicted along a densely populated sensor grid. Then, it can be expected that the modeling errors for two nearby mode shape components are related. In the case where time-domain data are employed,

it can similarly be suspected that modeling errors at consecutive time steps are correlated at high sampling frequencies. Besides correlations, the prediction error can show a systematic component, e.g., due to incorrect modeling assumptions or inexact measurement equipment or setup, which means assuming a zero-mean error is also not always realistic.

In order to avoid incorrect or unsuitable assumptions and thereby influencing the Bayesian updating results in an unfounded way, one can make use of the available observed data to try and estimate (characteristics of) the prediction error, for instance, using Bayesian model class selection.

Bayesian Model Class Selection

Bayesian inference can be applied at model class level to assess the plausibility of several alternative model classes based on the available observations $\bar{\mathbf{d}}$; this is referred to as Bayesian model class selection or MCS (Beck and Yuen 2004; Yuen 2010). The set of alternative model classes commonly concern (mechanical) prediction model classes $\mathcal{M}_M$, but here the method will be used to distinguish between alternative probabilistic prediction error models $\mathcal{M}_\eta$. The following, however, is elaborated for a set of general model classes $\mathcal{M}_i$.

Suppose there is a set $\mathcal{M}$ of N_C candidate model classes $\mathcal{M}_i$:

$$\mathcal{M} = \{\mathcal{M}_i\} \quad \text{where } i = 1, \dots, N_C \quad (8)$$

Then, the posterior probability of each model class $\mathcal{M}_i$ is given by Bayes’ theorem as

$$P(\mathcal{M}_i|\bar{\mathbf{d}}, \mathcal{M}) = \frac{p(\bar{\mathbf{d}}|\mathcal{M}_i)P(\mathcal{M}_i|\mathcal{M})}{p(\bar{\mathbf{d}}|\mathcal{M})} \quad (9)$$

where $P(\mathcal{M}_i|\mathcal{M})$ is the prior probability of each model class $\mathcal{M}_i$ (e.g., taken equal to $1/N_C$ when the model classes are considered equally plausible a priori), where usually the sum of all prior probabilities over all model classes is taken to be equal to 1.

The term $p(\bar{\mathbf{d}}|\mathcal{M}_i)$ in the numerator on the right-hand side of Eq. 9 is the *evidence* (sometimes also referred to as the *model class likelihood*) for the model class $\mathcal{M}_i$ provided by the data $\bar{\mathbf{d}}$. The evidence, hereafter denoted with ϵ_i , is a very important quantity in Bayesian model class selection and can be determined based on the law of total probability as

$$\epsilon_i = p(\bar{\mathbf{d}}|\mathcal{M}_i) = \int_{D_i} p(\bar{\mathbf{d}}|\boldsymbol{\theta}_i, \mathcal{M}_i) p(\boldsymbol{\theta}_i|\mathcal{M}_i) d\boldsymbol{\theta}_i \quad (10)$$

where $\boldsymbol{\theta}_i$ is the parameter vector in a parameter space D_i that defines each model in $\mathcal{M}_i$. By definition, the evidence ϵ_i gives the probability of the data according to model class $\mathcal{M}_i$. Following Eq. 10, it is determined as the weighted average of the probability of the data according to each model in the model class $\mathcal{M}_i$, where the weights are equal to the prior probability of the parameter values corresponding to each model. In fact, the evidence is equal to the a priori expected value of the likelihood. Note that it is also equal to the reciprocal of the normalizing constant c in the Bayes' theorem described by Eq. 3.

The denominator in Eq. 9, $p(\bar{\mathbf{d}}|\mathcal{M})$, is the probability of the data according to all model classes in $\mathcal{M}$ and is determined by applying the law of total probability as

$$p(\bar{\mathbf{d}}|\mathcal{M}) = \sum_{i=1}^{N_c} p(\bar{\mathbf{d}}|\mathcal{M}_i) P(\mathcal{M}_i|\mathcal{M}) \quad (11)$$

As this value is a constant, it is clear that in order to establish the most probable model class, the numerator term $p(\bar{\mathbf{d}}|\mathcal{M}_i)P(\mathcal{M}_i|\mathcal{M}) = \epsilon_i P(\mathcal{M}_i|\mathcal{M})$ in Eq. 9 has to be maximized with respect to i . Usually, the process of model class selection consists of determining the posterior probabilities of all model classes in the set $\mathcal{M}$ and ranking them accordingly, where the best model class is the one that results in the highest value of the quantity $\epsilon_i P(\mathcal{M}_i|\mathcal{M})$. Since most often equal prior model class probabilities are adopted, it suffices to compute the evidence values ϵ_i for all

model classes and to rank them accordingly. Then, the most probable model class – according to the available data – corresponds to the model class with the highest evidence value.

The actual computation of the evidence values according to Eq. 10 poses a challenging problem in most practical applications, as it entails computing complex and usually high-dimensional integrals. To overcome this problem, asymptotic approximations can be applied (Papadimitriou et al. 1997; Beck 2010), or, alternatively, methods based on stochastic simulation such as MCMC sampling (Gelfand and Dey 1994; Cheung and Beck 2010) may be used.

Model Parsimony Through Bayesian MCS

In the context of model class selection, the principle of model parsimony expresses that simpler models that are reasonably consistent with the observed data are preferable to more complicated models (e.g., with more modeling parameters), even though the more complex models might result in a better data fit. In other words, the selected class of models should be able to reproduce the observed system behavior closely, but should otherwise be as simple as possible. This principle of model simplicity is also often referred to as “Occam’s razor” and is pursued in MCS for two reasons: firstly, the achieved data fit improvement is in most cases only marginal, and secondly, an over-parameterized model may lead to an over-fit of the data. This often results in relatively poor predictions made by the model, due to the high sensitivity of the model to small changes in the data.

Many authors have suggested adding penalty terms to the log likelihood to discourage highly parameterized model classes. However, it can be shown that applying a Bayesian inference scheme at the model class level automatically enforces model parsimony, thereby avoiding the need for ad hoc penalty terms (Beck 2009). Using Bayes’ theorem and the fact that the posterior PDF integrates to one, the logarithm of the evidence in Eq. 10 can be reformulated and expanded for a general model class $\mathcal{M}$ as

$$\begin{aligned} \log \epsilon &= \log p(\bar{\mathbf{d}}|\mathcal{M}) = \mathbb{E}\{\log p(\bar{\mathbf{d}}|\boldsymbol{\theta}, \mathcal{M})\}_{\text{po}} \\ &\quad - \mathbb{E}\left\{\log \frac{p(\boldsymbol{\theta}|\bar{\mathbf{d}}, \mathcal{M})}{p(\boldsymbol{\theta}|\mathcal{M})}\right\}_{\text{po}} \end{aligned} \quad (12)$$

The first term in the above log evidence expression is the posterior mean value of the log likelihood function, termed $\log \epsilon_{\text{data}}$, which gives a measure for the average data fit for the model class $\mathcal{M}$. It is easily verified that the second term equals the Kullback–Leibler divergence D_{KL} between the prior and the posterior PDF. This term gives a measure for the difference between the prior and posterior PDF: it can be interpreted as a measure for the information that is gained from the observations $\bar{\mathbf{d}}$. It can be shown that this term is nonnegative and increases with the number of model parameters in the model class (Beck and Yuen 2004; Beck 2007); therefore, this term penalizes more complex models that attempt to extract more information from the data, thus eliminating the need for ad hoc penalty terms. In the following, the negative of this term is denoted as the Occam term $\log \epsilon_{\text{occam}}$, as it enforces the Occam simplicity principle. As such, the log evidence can be rewritten as

$$\log \epsilon = \log \epsilon_{\text{data}} + \log \epsilon_{\text{occam}} \quad (13)$$

where

$$\log \epsilon_{\text{data}} = \mathbb{E}\{\log p(\bar{\mathbf{d}}|\boldsymbol{\theta}, \mathcal{M})\}_{\text{po}} \quad (14)$$

$$\log \epsilon_{\text{occam}} = -\mathbb{E}\left\{\log \frac{p(\boldsymbol{\theta}|\bar{\mathbf{d}}, \mathcal{M})}{p(\boldsymbol{\theta}|\mathcal{M})}\right\}_{\text{po}} \quad (15)$$

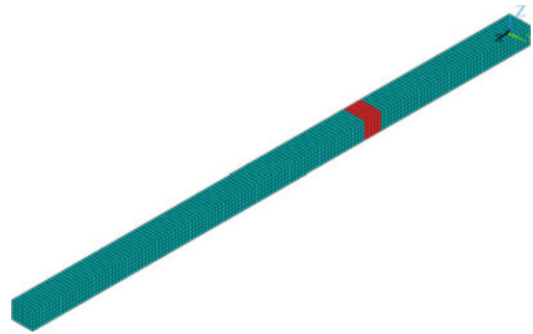
Thereby, it is shown that the log evidence consists of a data fit term and a term which provides a penalty for more complex models that attempt to extract more information from the data. In other words, by applying the Bayesian inference scheme for model class selection, an explicit trade-off is made between the data fit of the model class and its complexity, without introducing ad hoc concepts.

Application: Estimation of Prediction Error Correlation Model Class

This example treats a common application of model updating in structural mechanics, namely, vibration-based structural health monitoring (SHM). In this technique, modal data extracted from vibration experiments are used to update stiffness values along a structure; as such, structural damage (i.e., loss in stiffness) can be identified, located, and quantified. In this context, Bayesian model updating provides an effective tool to assess the influence of measurement and modeling error on the SHM results.

The practical application of the prediction error estimation procedure explained above is demonstrated in a small-scale example, where a vibration test on a damaged reinforced concrete (RC) beam is simulated. This is done by modeling the beam using a detailed 3D FE model consisting of 5250 volume elements (Fig. 1) where the steel reinforcement is taken into account through an adapted material density. Structural damage is simulated by reducing the Young's modulus of a small area around one third of the length of the beam to 18.75 GPa; the rest of the beam is appointed a stiffness modulus of 37.5 GPa.

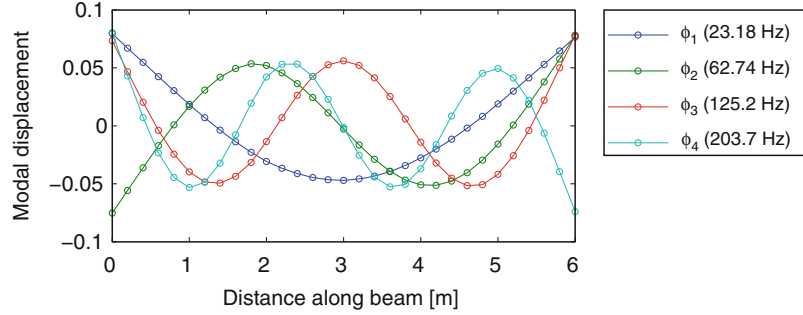
The simulated modal data consists of the first four bending modes $\bar{\boldsymbol{\phi}}_r \in \mathbb{R}^{N_s}$ at N_s degrees of freedom (DOFs) or sensors and their associated natural frequencies $\bar{f}_r$ with eigenvalues



Model Class Selection for Prediction Error Estimation, Fig. 1 3D solid FE model of the RC beam; the area with reduced stiffness is shown in red

Model Class Selection for Prediction Error Estimation,

Fig. 2 Simulated mode shapes $\bar{\phi}_r$ at $N_s = 31$ sensors and corresponding natural frequencies $\bar{f}_r = \sqrt{\bar{\lambda}}/2\pi$



$\bar{\lambda}_r = (2\pi\bar{f}_r)^2$, where free-free boundary conditions are adopted in the 3D model. To simulate a measurement error, an uncorrelated zero-mean Gaussian error with a standard deviation of 0.1 % is superimposed on these modal data. In order to assess the influence of the sensor density, three sensor configurations are considered, with $N_s = 6, 31$, and 151 equidistant sensors, and where a denser configuration includes the previous one. The final simulated experimental data set $\bar{\mathbf{d}} = \{\bar{\lambda}_1, \dots, \bar{\lambda}_4, \bar{\phi}_1^T, \dots, \bar{\phi}_4^T\}^T$ for 31 sensors is shown in Fig. 2.

The model of the RC beam that will be updated using a Bayesian inference approach is a 2D finite element model which consists of 150 beam elements and 151 nodes, with two DOFs per node, resulting in a total of $N_d = 302$ DOFs in the FE model. This prediction model is parameterized by a single parameter θ_M representing the Young's modulus of the damaged region of elements at around one third of the length of the beam. It is expected that this model combined with the simulated data will lead to prediction errors that are correlated in space; the objective is therefore to effectively account for these correlations through the proposed Bayesian MCS approach.

A zero-mean Gaussian prediction error is adopted so that $\boldsymbol{\eta} \sim \mathcal{N}(0, \Sigma_\eta)$, where it is assumed that the eigenvalue errors $\eta_{\lambda,r} = \bar{\lambda}_r - \lambda_r(\theta_M)$ are independent from the mode shape errors $\boldsymbol{\eta}_{\phi,r} = \bar{\phi}_r - \phi_r(\theta_M)$, meaning that the covariance matrix Σ_η can be constructed as

$\Sigma_\eta = \text{blkdiag}(\Sigma_\lambda, \Sigma_\phi)$. The eigenvalue covariance matrix is assumed to be diagonal and parameterized as $\Sigma_\lambda = \theta_\lambda^2 \text{diag}(\bar{\lambda}_1^2, \dots, \bar{\lambda}_4^2)$. For the mode shapes, it is suspected that the correlation length scale depends on the considered mode; therefore, it is assumed that the covariance matrix for the mode shape components can be constructed as the block diagonal matrix of 4 individual covariance matrices: $\Sigma_\phi = \text{blkdiag}(\Sigma_\phi^1, \dots, \Sigma_\phi^4)$.

Because the spatial correlation structure is unknown, a set of three alternative prediction error model classes is determined for Σ_ϕ : an uncorrelated model class A, a model class B with an exponential correlation function, and a model class C with an exponentially damped cosine correlation function. Each of these model classes is parameterized by a number of prediction error parameters as follows:

$$\Sigma_\phi^{A,r} = (\theta_{\phi,1}^{A,r})^2 \mathbf{I}_{N_s} \quad (16)$$

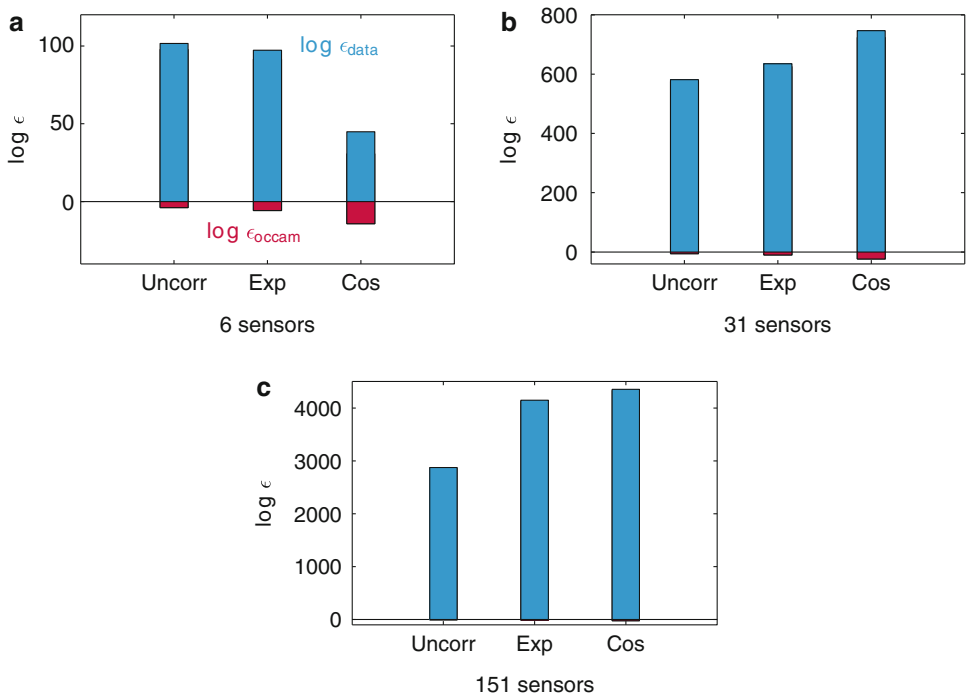
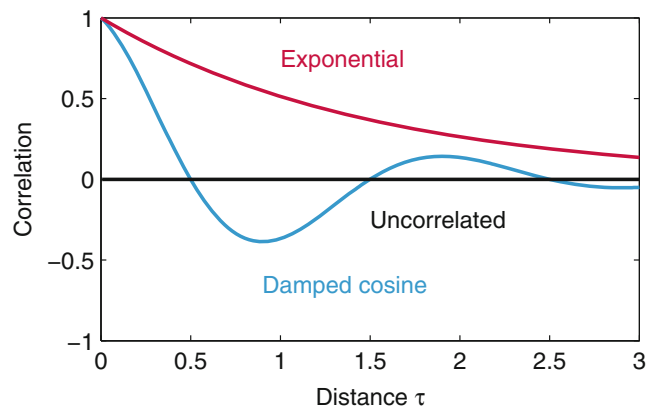
$$[\Sigma_\phi^{B,r}]_{ij} = (\theta_{\phi,1}^{B,r})^2 \exp\left(-\frac{\tau_{ij}}{\theta_{\phi,2}^{B,r}}\right) \quad (17)$$

$$[\Sigma_\phi^{C,r}]_{ij} = (\theta_{\phi,1}^{C,r})^2 \exp\left(-\frac{\tau_{ij}}{\theta_{\phi,2}^{C,r}}\right) \cos(\theta_{\phi,3}^{C,r} \tau_{ij}) \quad (18)$$

where τ_{ij} represents the distance between sensors i and j . These correlation models of increasing

Model Class Selection for Prediction Error Estimation,

Fig. 3 Illustration of the three employed correlation models



Model Class Selection for Prediction Error Estimation, Fig. 4 Log data fit values $\log \epsilon_{data}$ and log Occam values $\log \epsilon_{occam}$ for 6, 31, and 151 sensors

complexity (i.e., with more parameters) are illustrated in Fig. 3.

Based on the ME principle, gamma-distributed prior PDFs are assumed for all parameters (Soize 2003), and the Bayesian updating scheme is performed for all three alternative model classes using Markov Chain Monte Carlo

(MCMC) sampling where a Metropolis–Hastings sampling algorithm is used to obtain 50,000 samples of the posterior PDF. The evidence values are computed using asymptotic expressions (Simoen et al. 2013) and are shown in Fig. 4 for the three sensor configurations. Several observations can be made. Firstly, the log Occam value

clearly increases in absolute value as the model class complexity increases, confirming the statements made above regarding model parsimony. Secondly, for a low number of sensors, the uncorrelated model class is selected, indicating that correlation only becomes important for relatively dense sensor grids. However, as the number of sensors increases, the uncorrelated model class is clearly rejected, and the cosine correlation model class is eventually distinctly preferred.

The PDFs of the parameters characterizing this correlation model are estimated as well as the PDF of the stiffness parameter θ_M . For 151 sensors (and the selected cosine prediction error model class), a posterior mean value of $\mu(\theta_M) = 18.95$ GPa is found, with a standard deviation of $\sigma(\theta_M) = 0.80$ GPa.

This application shows that the Bayesian MCS approach can effectively be used to estimate a suitable correlation structure. Moreover, the MCS approach can also be applied to distinguish between different types of prediction error models (e.g., Gaussian vs. non-Gaussian) or between alternative descriptions for systematic prediction errors.

Summary

This reference entry describes the use of Bayesian model class selection to determine, among a set of possible model classes, the prediction error model class most suited for Bayesian model updating, according to the available experimental data. It is demonstrated that, provided sufficient information is available, the Bayesian MCS approach is an effective tool to this end, ensuring a more realistic joint structural-probabilistic model and corresponding Bayesian model updating results.

Cross-References

- [Bayesian Statistics: Applications to Earthquake Engineering](#)

- [Uncertainty Quantification in Structural Health Monitoring](#)
- [Vibration-Based Damage Identification: The Z24 Bridge Benchmark](#)

References

- Beck J (2007) Bayesian updating and model class selection of deteriorating hysteretic structural models using seismic response data. In: Proceedings of the ECCOMAS thematic conference on computational methods in structural dynamics and earthquake engineering, Rethymno
- Beck J (2009) Using model classes in system identification for robust response predictions. In: Proceedings of COMPDYN 2009, ECCOMAS thematic conference on computational methods in structural dynamics and earthquake engineering, Rhodos
- Beck JL (2010) Bayesian system identification based on probability logic. *Struct Control Health Monit* 17(7):825–847
- Beck J, Katafygiotis L (1998) Updating models and their uncertainties. I: Bayesian statistical framework. *ASCE J Eng Mech* 124(4):455–461
- Beck J, Yuen K-V (2004) Model selection using response measurements: Bayesian probabilistic approach. *ASCE J Eng Mech* 130(2):192–203
- Cheung SH, Beck JL (2010) Calculation of posterior probabilities for bayesian model class assessment and averaging from posterior samples based on dynamic system data. *Comput-Aided Civ Infrastruct Eng* 25(5):304–321
- Diaconis P, Ylvisaker D (1979) Conjugate priors for exponential families. *Ann Stat* 7(2):269–281
- Gelfand A, Dey D (1994) Bayesian model choice: asymptotics and exact calculations. *J R Stat Soc* 56(3):501–514
- Jaynes E (1957) Information theory and statistical mechanics. *Phys Rev* 106(4):620–630
- Papadimitriou C, Beck J, Katafygiotis L (1997) Asymptotic expansions for reliability and moments of uncertain systems. *ASCE J Eng Mech* 123(12):1219–1229
- Simoen E, Papadimitriou C, Lombaert G (2013) On prediction error correlation in Bayesian model updating. *J Sound Vib* 332(18):4136–4152
- Soize C (2003) Probabilités et modélisation des incertitudes: éléments de base et concepts fondamentaux. Handed out at the séminaire de formation de l'école doctorale MODES, Paris
- Soize C (2008) Construction of probability distributions in high dimensions using the maximum entropy principle: applications to stochastic processes, random fields and random matrices. *Int J Numer Methods Eng* 75:1583–1611
- Yuen K-V (2010) Recent developments of Bayesian model class selection and applications in civil engineering. *Struct Saf* 32(5):338–346

Model-Form Uncertainty Quantification for Structural Design

Ramana V. Grandhi¹ and Christopher Corey Fischer²

¹210 Russ Engineering Center, Wright State University, Dayton, OH, USA

²Wright State University, Dayton, OH, USA

Synonyms

Additive Adjustment Factor; Adjustment Factor Approach; Bayesian Model Averaging; Bayes' Theorem; Model Combination; Model-Form Uncertainty; Multiplicative Adjustment Factor; Probabilistic Adjustment Factor; Uncertainty; Uncertainty Quantification

Introduction

Multiple forms of uncertainty exist in the prediction of a system response through the use of any type of modeling process. These uncertainties can be thought of presenting in one of three forms: parametric, predictive, and model-form uncertainty (Kennedy and O'Hagan 2001). The first of these three forms, parametric uncertainty, refers to the natural variability present within any input parameter, parameter in which a given model is reliant for predicting the response of interest. The latter two forms of uncertainty, predictive and model-form uncertainty, refer to the natural variability present within the modeling process itself. Uncertainty quantification work in the literature has primarily focused on the quantification of parametric uncertainty through the exploration, adaptation, and application of multiple approaches and methodologies. However, the majority of these approaches and methodologies fail to account for the presence of uncertainties that arise as a result of the modeling process.

Multiple models often arise as possibilities for representing the same physical situation when the solution approach for solving an engineering

problem utilizes physics-based simulations. This is often a result of differing assumptions on governing physics, physical representation of boundary or loading conditions, or solution approaches. The availability of multiple models in simulating a given physical scenario can also result from the use of various modeling packages utilizing various fidelities or even operating on differing assumptions and/or mesh sizes within the same fidelity level. In very rare situations, the phenomena being modeled have been explored extensively and are well enough understood that a "best" model can emerge from the set of possible models, where in this work the term "best" model refers to the model that most accurately represents the true physical scenario being modeled. However, in most physical scenarios in which multiple sets of differing, complex physics can couple in various ways, there exists uncertainty in the selection of this "best" model as a result of a lack of knowledge of the true physics governing the physical scenario. As a result of this uncertainty, a single "best" model will not emerge from the model set, and thus, multiple models operating on assumptions of differing physics can produce inconsistent results for the prediction of the same physical problem. Therefore, in order to completely quantify the uncertainty present in the modeling process, it is necessary to not only quantify parametric uncertainties, but also the uncertainty inherent in selecting the "best" model – the model-form uncertainty.

Multiple approaches for the quantification of model-form uncertainties are presented, discussed, and applied to a nonlinear transient concrete creep engineering problem. This entry presents the application bounds of each approach such as requirements on the availability of experimental data and whether or not the approach can account for probabilistic model predictions – predictions including the quantification of parametric and/or predictive uncertainties. These approaches include the traditional adjustment factor approaches, probabilistic adjustment factor approach (an adaptation of the prior), and the Bayesian model averaging (BMA) approach.

Uncertainty in Modeling Process

The process involved in modeling a physical scenario requires the construction of a mathematical model which entails discretization of the physical scenario as well as making assumptions on the physics governing the true scenario. It is however, often beyond the capability of the designer to fully understand the true engineering problem at hand and thus capture the full complexity and all aspects of the physical scenario. Therefore, as a result of the assumptions made during development of physics-based models, discrepancy inevitably exists between the physical scenario being modeled and the prediction of the response by the computational model. This discrepancy is referred to as the predictive uncertainty of a particular model (Droguett and Mosleh 2008), and is unique to each individual model within a set of models considered for predicting a system response.

As mentioned previously, in the solution of an engineering problem, it is common practice for multiple models to be constructed for the purpose of representing the same physical scenario. Such examples include models of varying fidelities, such as assumptions on linear, quasilinear, or nonlinearity within a model, or models that account for complex phenomenon or boundary conditions in different manners, using different simplification techniques. As a result, the predictions of each of these models, for the particular output of interest, can be different from one another. Guedes-Soares states that in situations where multiple models yield different responses, that there can exist only one correct model (Soares 1997). However, as stated earlier, it is often beyond the capability of the designer to select the “best” model within the design space, or even a subset of the design space, and arises as a result of the inability to completely understand the full complexity of the physics governing the problem. As is the case in many engineering problems, due to the physical complexity and multidisciplinary interactions, a correct model will often not exist, instead one merely seeks to identify the “best” model among the model set. For reasons mentioned above, there exists uncertainty in the selection of this “best” model.

This uncertainty in the identification of the model most accurately predicting the correct outcome is referred to as model-form uncertainty (Zhang and Mahadevan 2000).

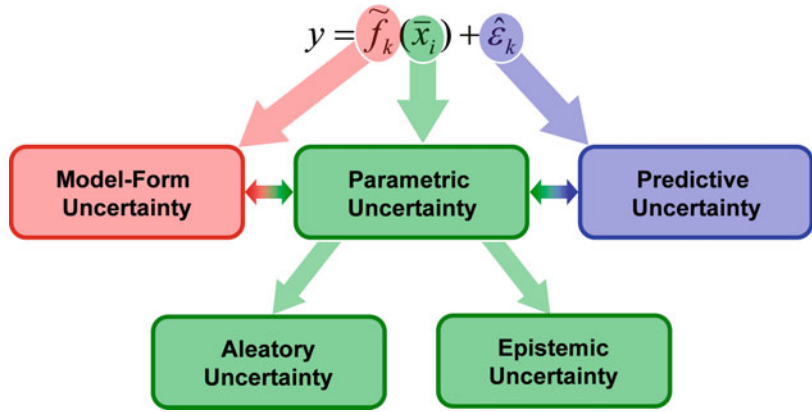
The definition of input parameters to each model within a model set, such as dimensions, material properties, environmental conditions, or modeling constants, are often defined as being deterministic. However, in nature, the parameters are rarely deterministic, or seldom able to be accurately represented as deterministic values, within the true physical scenario. As a result of this fact, there exists a third type of uncertainty in the modeling process – parametric uncertainty – which refers to the natural variability present in the values of parameters that a mathematical model is dependent upon (Droguett and Mosleh 2008). Parametric uncertainties are often classified as either aleatory or epistemic uncertainty, dependent upon the degree of information known in regards to the form of their uncertainty (Paté-Cornell 1996). Aleatory uncertainty refers to the form of uncertainty that arises from the natural unpredictable variation in the performance of a system (Hacking 1984), and can often be represented through distribution functions of a parameter’s variability. Epistemic uncertainty, on the other hand, is defined as uncertainty due to a lack of knowledge regarding the performance of a system that can, in theory, be reduced through the introduction of additional data or information (Chernoff and Moses 2012).

All of the aforementioned uncertainties are present in nearly every physics-based modeling problem. A general representation for the modeling of an output of interest, y , is given in Eq. 1. As can be seen, this output of interest is a function of three terms: $\tilde{f}_k$, $\bar{x}_i$, and $\hat{e}_k$.

$$y = \tilde{f}_k(\bar{x}_i) + \hat{e}_k \quad (1)$$

Here, $\tilde{f}_k$ represents the result/prediction obtained using model k , to the set of input parameters, $\bar{x}_i$. The second term, $\hat{e}_k$, represents the discrepancy between the result/prediction obtained from model k , and the true physical value of the output of interest. The selection of the appropriate model, $\tilde{f}_k(\bar{x}_i)$, given this

Model-Form Uncertainty Quantification for Structural Design,
Fig. 1 Modeling uncertainty breakdown



representation of a modeling problem, is representative of the inherent model-form uncertainty whereas, the variation in $\tilde{f}_k(\bar{x}_i)$ due to the uncertainty present in the set of input parameters, $\bar{x}_i$, is representative of parametric uncertainty. Finally, the determination of a discrepancy between response/prediction of model k and the true physical value of the output of interest, $\hat{\epsilon}_k$, is representative of the predictive uncertainty present in model k . The breakdown of these three distinct types of uncertainty is demonstrated in Fig. 1.

Bayes' Theorem for Quantifying Model Probability

The first step for quantifying model-form uncertainty is to define model probabilities. Model probability is defined as the degree of belief that a model is the best approximating model from within a model set (Park et al. 2010). Original model probabilities can be assigned on one of two different concepts. Either expert opinion, if available, can be used to assign model probabilities based upon some prior knowledge of each model within the model set, or the more likely basis is to assign these probabilities based upon a uniform distribution of the probabilities assigned to each model within the model set. In the case of the uniformly distributed model probabilities where there exists K models within the model set, each model is assigned a probability of $\Pr(M_k) = 1/K$. Note that the sum of model probabilities, as shown in Eq. 2, for a model set, must be equal to 1.

$$\sum_{k=1}^K \Pr(M_k) = 1, \quad k = 1, 2, \dots, K \quad (2)$$

In the event that experimental data is available, Bayes' theorem can then be used to update these original (prior) probabilities into posterior probabilities. Applying Eq. 3 updates the model probability of model k .

$$\Pr(M_k|D) = \frac{\Pr(D|M_k)\Pr(M_k)}{\sum_{k=1}^K \Pr(D|M_k)\Pr(M_k)} \quad (3)$$

In Eq. 3, $\Pr(D|M_k)$, which is broadly known as $L(M_k|D)$, is called the likelihood of model M_k given experimental data D , or the likeliness that model M_k will predict the known data D , because the first argument D is known, while the second argument M_k is not held constant (Robert 2007). The numerical value of the likelihood, $\Pr(D|M_k)$, is not needed in the uncertainty quantification process; however, it is of importance in the evaluation of model probabilities given experimental data. Note that the denominator in Eq. 3 is common to all models, and this model likelihood value is calculated using Eq. 4 as shown by Park et al. in (Park et al. 2010).

$$\Pr(D|M_k) = \left(\frac{1}{2\pi(\hat{\sigma})_{mle}^2} \right)^{N/2} \exp\left(-\frac{N}{2}\right) \quad (4)$$

Model likelihood then becomes a function of the maximum likelihood estimator, $(\hat{\sigma})_{mle}^2$, a

measure of the variation within model k , as calculated in Eq. 5, where N in Eqs. 4 and 5 represents the number of experimental data points available for calculating the error term, ϵ . This maximum likelihood estimator provides an estimate for the model's variance based upon given experimental data in which the error term is the difference between the given experimental data point i and the prediction for that data point from model k . In Eq. 5 this error term is squared so as to eliminate the possibility of a positive and negative error canceling each other out.

$$(\hat{\sigma})_{mle}^2 = \frac{\sum_{i=1}^N \epsilon_{k,i}^2}{N} \quad (5)$$

Model-Form Uncertainty Quantification Techniques

Many methods for the quantification of model-form uncertainty have recently arisen within the literature. Each one of these methods exhibit unique advantages and disadvantages. Original work in model-form uncertainty quantification gave birth to methods requiring the presence of experimental data, such as Bayesian model averaging which requires the presence of experimental data in order to develop a maximum likelihood estimator of variance within a model prediction as is used in updating model probabilities through Bayes' theorem. Riley and Grandhi quantified the model-form and predictive uncertainty in the calculation of the flutter velocity of the AGARD 445.6 wing using BMA in (Riley and Grandhi 2011b). Other methods requiring the availability of experimental data include the methods of Continuous Model Expansion explored by Draper, which showed a reliance on experimental data points, as well as difficulty in handling asymmetric distributions of parametric uncertainties (Draper 1995), and work done recently by Allaire and Willcox, which require a maximum entropy representation of the modeling uncertainty, an extra step that could be cost intensive for a high simulation-cost model (Allaire and Willcox 2010). As previously mentioned, the common thread among these three methods is the

dependence on the availability of experimental data. The adjustment factor approach, however, was demonstrated by Mosleh and Apostolakis as a possible method for model-form uncertainty quantification which operates on the utilization of expert opinion to assign model probabilities in the absence of empirical data (Mosleh and Apostolakis 1986). Riley and Grandhi present an adaptation of this approach for use with nondeterministic model predictions, called the probabilistic adjustment actor approach (Riley and Grandhi 2011a). This work will explore the adjustment factor approaches, probabilistic adjustment factor approach, and the Bayesian Model Averaging approach and apply these methods to a multiphysics problem.

Additive Adjustment Factor Approach

As mentioned previously, the adjustment factor approach was first demonstrated by Mosleh and Apostolakis as a method for quantifying model-form uncertainty in the absence of experimental data by using an adaptation of Bayes' Theorem. This approach modifies the result of the "best" model-model within the model set being considered for use in solving the problem that obtains the highest model probability—by applying an adjustment factor to account for the uncertainty that exist in selection of the "best" model. The applicability of this approach to engineering problems has been well demonstrated in the literature. Zio and Apostolakis utilized an adjustment factor approach to quantify the uncertainty present in the selection of a "best" radioactive waste repository model in (Zio and Apostolakis 1996). Reinert and Apostolakis also used this approach in the assessment of risk for decision-making processes in (Reinert and Apostolakis 2006).

Multiple derivations of the adjustment factor approach exist in the literature. These derivatives all employ a similar technique of quantifying the model-form uncertainty through the use of expert opinion regarding a model's accuracy with respect to other models in the model set by assigning model probabilities, and updating those probabilities through the use of Bayes'

theorem upon the availability of experimental data. In this approach, $\Pr(M_k)$ represents the model probability assigned to model k . Recall that this model probability is the probability that model k is the “best” model among model set M being considered, where Eq. 6 defines the model set M .

$$M = \{M_1, M_2, \dots, M_K\} \quad (6)$$

Note that model probabilities for each of the K individual models within the model set M remain bounded by the laws of probability theory. Thus, constraints are applied to the model probability values, as shown in Eq. 7.

$$\sum_{k=1}^K \Pr(M_k) = 1 \quad \text{such that} \quad 0 \leq \Pr(M_k) \leq 1 \quad (7)$$

The various derivations of the adjustment factor approach differ here, in the form of the adjustment factor being applied—used to adjust—the “best” model. In the additive adjustment factor approach, the adjusted model, y , is formed by adding an additive adjustment factor, E_a^* , to the “best” model from the model set being considered, as shown in Eq. 8.

$$y = y^* + E_a^* \quad (8)$$

This additive adjustment factor, E_a^* , is assumed to be a normally distributed factor representing the uncertainty present in the selection of the “best” model as being most accurate at predicting the true physical response. In assuming a Gaussian form for this factor, the first and second moments—expected value and variance—of the adjustment factor are calculated as shown in Eqs. 9 and 10.

$$E[E_a^*] = \sum_{k=1}^K \Pr(M_k)(y_k - y^*) \quad (9)$$

$$\text{Var}[E_a^*] = \sum_{k=1}^K \Pr(M_k)(y_k - E[y])^2 \quad (10)$$

It is important to note that use of the adjustment factor approach is reliant on models within

the model set to be deterministic; this is to say that the models cannot incorporate parametric uncertainty. Therefore, the expected value of the adjusted model, $E[y]$, is calculated as shown in Eq. 11, as the sum of the prediction of the “best” model and the expected value of the additive adjustment factor, $E[E_a^*]$, found using Eq. 9.

$$E[y] = y^* + E[E_a^*] \quad (11)$$

Similarly, due to each individual model being deterministic, the variance of the adjusted model, $\text{Var}[y]$, assumes only the variance of the additive adjustment factor, $\text{Var}[E_a^*]$, as shown in Eq. 12.

$$\text{Var}[y] = \text{Var}[E_a^*] \quad (12)$$

The first and second moments—expected value and variance—of the adjusted prediction, y , are calculated by Eqs. 11 and 12 which fully define a normal distribution representing the uncertainty inherent in the prediction of a system response as a result of model-form uncertainty.

Multiplicative Adjustment Factor

Another derivative of the adjustment factor approach is that of a multiplicative adjustment factor. In the multiplicative adjustment factor approach, the adjusted model, y , is formed by multiplying the prediction of the “best” model, y^* from the model set being considered by a multiplicative adjustment factor, E_m^* , as shown in Eq. 13.

$$y = y^* * E_m^* \quad (13)$$

This multiplicative adjustment factor, E_m^* , is assumed to be a log normally distributed factor representing the uncertainty present in the selection of the “best” model as being most accurate at predicting the true physical response. In assuming a log normally distributed form for this factor, the first and second moments—expected value and variance—of the adjustment factor are calculated as shown in Eqs. 14 and 15.

$$E[\ln(E_m^*)] = \sum_{k=1}^K \Pr(M_k)(\ln(y_k) - \ln(y^*)) \quad (14)$$

$$\text{Var}[\ln(E_m^*)] = \sum_{k=1}^K \Pr(M_k)(\ln(y_k) - E[\ln(y)])^2 \quad (15)$$

As was the case with the additive adjustment factor approach, it is important to note that use of the adjustment factor approach is reliant on models within the model set to be deterministic. Therefore, the expected value of the adjusted model, $E[\ln(y)]$, is calculated as shown in Eq. 16, as the sum of the natural logarithm of the prediction of the “best” model and the expected value of the multiplicative adjustment factor, $E[\ln(E_m^*)]$, from Eq. 14.

$$E[\ln(y)] = \ln(y^*) + E[\ln(E_m^*)] \quad (16)$$

Much like the additive adjustment factor approach, due to each individual model being deterministic, the variance of the adjusted model, $\text{Var}[\ln(y)]$, assumes only the variance of the multiplicative adjustment factor, $\text{Var}[\ln(E_m^*)]$, as shown in Eq. 17.

$$\text{Var}[\ln(y)] = \text{Var}[\ln(E_m^*)] \quad (17)$$

While these adjustment factor approaches have the benefit of being able to quantify model-form uncertainty in the absence of experimental data, they lack the ability to handle probabilistic model predictions, this is to say that the adjustment factor approach cannot quantify the model-form uncertainty present in the selection of a “best” model, from a model set, when the models within said model set contain quantified parametric uncertainty. This is a result of the assumption made during derivation of the approach that individual models are deterministic. Therefore, a rederivation of the approach, working on the assumption that individual models are probabilistic/stochastic in nature, would allow for the approach to be adapted to handle parametric uncertainty.

Probabilistic Adjustment Factor Approach

The probabilistic adjustment factor approach is an adaptation of the traditional adjustment factor approach in that it is capable of handling stochastic models, and thus parametric uncertainty. Note that this approach does not quantify the parametric uncertainty within each model of a model set, but it can handle parametrically uncertain models in its analysis (Riley and Grandhi 2011a). As with the traditional adjustment factor approach, a distribution must first be assumed for each of the individual models within the model set for derivation of the probabilistic adjustment factor approach. In general, there is no restriction on the form of this distribution; however, as was the case with the additive adjustment factor approach, this work will focus on the approach derived from assumption of a normal distribution.

The first step to quantifying model-form uncertainty through the use of the probabilistic adjustment factor approach is the same as that of the traditional adjustment factor approach. Model probabilities are applied to each individual model within the model set using either expert opinion or uniformly distributed probabilities, and then updating through the application of Bayes’ theorem given the availability of experimental data, such that constraints of Eq. 7 are satisfied. The adjusted model for the probabilistic adjustment factor approach can then be computed as shown in Eq. 18.

$$y = E[y^*] + E_{pafa}^* \quad (18)$$

Equation 18 is similar to Eq. 8 from the additive adjustment factor approach with the slight difference in the utilization of the expected value of the “best” model, $E[y^*]$, rather than the deterministic “best” model prediction. Calculating the first and second moments—expected value and variance—of the probabilistic adjustment factor is also different for this new approach, as the approach had to be rederived to handle the stochastic model set. The calculation of these

two moments can be performed using Eqs. 19 and 20.

$$E[E_{pafa}^*] = \sum_{k=1}^K \Pr(M_k)(E[y_k] - E[y^*]) \quad (19)$$

$$\text{Var}[E_{pafa}^*] = \sum_{k=1}^K \Pr(M_k)(E[y_k] - E[y])^2 \quad (20)$$

After calculating the first and second moments of the probabilistic adjustment factor, the expected value and variance of the adjusted model can then be calculated using Eqs. 21 and 22.

$$E[y] = E[y^*] + E[E_{pafa}^*] \quad (21)$$

$$\text{Var}[y] = \text{Var}[E_{pafa}^*] + \sum_{k=1}^K \Pr(M_k)(\text{Var}[y_k])^2 \quad (22)$$

As can be seen, there is only a slight difference in the formulations of the expected value equations for the traditional adjustment factor approach and the probabilistic adjustment factor approach, only in that of operating on the expected value of the “best” model prediction rather than the deterministic model prediction, as is the case for the traditional approach. The key difference in the two derivations comes in the formulation of the variance equation. The variance, shown in Eq. 22, includes the addition of the summation of weighted individual model variances—weighted by model predictions—to the variance of the adjustment factor calculated using Eq. 20, known as the between-model variance. The second term in the equation represents the variance in the adjusted model due to variances within each of the individual models—the within-model variance. Therefore, the first term in Eq. 22 can be thought of as representing the model-form uncertainty within the problem, whereas the second term represents the parametric uncertainty inherent to each model within the model set, and quantified using parametric uncertainty quantification techniques. In the event that

individual models within the model set are deterministic, do not account for parametric uncertainties, and experimental data is available, the probabilistic adjustment factor approach can still be utilized by assuming the maximum likelihood estimator of variance for each individual model to be said model’s variance in the evaluation of Eq. 22.

Bayesian Model Averaging

The Bayesian model averaging (BMA) technique is one that requires the availability of experimental data as mentioned previously; however, this technique can handle both stochastic and deterministic models. BMA is a technique for quantifying model-form uncertainty by averaging the predictions of each individual model within a model set using each model’s corresponding model probability as a weighting factor. As is the case with the previous approaches, a distribution must be assumed for the averaged model, and the individual models, if stochastic, must follow this same distribution. For the purpose of this work, each model, if stochastic, must follow a normal distribution—defined by an expected value and variance. Thus, the averaged model assumes the form of a Gaussian distribution. The expected value of the averaged model, $E[y|D]$, is calculated by Eq. 23 as the summation of the model probabilities, $\Pr(M_k|D)$, multiplied by the corresponding model’s expected value, $E[y|M_k, D]$.

$$E[y|D] = \sum_{k=1}^K \Pr(M_k|D)E[y|M_k, D] \quad (23)$$

Similarly, the equation for calculating the variance of the averaged model is shown in Eq. 24. This equation states that the variance of the averaged model, $\text{Var}[y|D]$, is the summation of the model probabilities, $\Pr(M_k|D)$, multiplied by the corresponding model’s variance, $\text{Var}[y|M_k, D]$, plus the summation of the model probabilities multiplied by the squared difference between

the corresponding model's expected value, $E[y|M_k, D]$, and the averaged model's expected value, $E[y|D]$.

$$\begin{aligned} \text{Var}[y|D] &= \sum_{k=1}^K \Pr(M_k|D) \text{Var}[y|M_k, D] \\ &+ \sum_{k=1}^K \Pr(M_k|D) (E[y|M_k, D] - E[y|D])^2 \end{aligned} \quad (24)$$

In the event that BMA is being applied to a deterministic model set, case in which uncertainties due to input parameters are not included, Eqs. 25 and 26 are used to calculate the averaged model expected value and variance. The difference in the formulation between Eqs. 23 and 25 is that the latter used the individual model deterministic predictions, y_k , rather than their expected values. The key difference in these two derivations is seen in the formulation of the equation for calculation of averaged model variance, where the first term of Eq. 24 changes by using the maximum likelihood estimator of variance, $(\hat{\sigma})_{mle}^2$ calculated using Eq. 5, in place of the individual stochastic model variances as well as substituting the individual deterministic model predictions, y_k , for the stochastic model expected values in the first term of Eq. 26, corresponding to the second term of Eq. 24 (Park and Grandhi 2011).

$$E[y|D] = \sum_{k=1}^K \Pr(M_k|D) y_k \quad (25)$$

$$\begin{aligned} \text{Var}[y|D] &= \sum_{k=1}^K \Pr(M_k|D) (y_k - E[y|D])^2 \\ &+ \sum_{k=1}^K \Pr(M_k|D) (\hat{\sigma}_k)_{mle}^2 \end{aligned} \quad (26)$$

Application

The outlined methodology is demonstrated through application to a concrete creep engineering problem. Creep is defined as the slow

deformation a solid material undergoes under the influence of sustained stresses generated by external loads over an extended period of time. The creep deformation experienced by concrete may be three or four times as large as the initial elastic deformation caused by externally applied loads. It is widely accepted that creep deformation is largely attributed to shearing forces acting on material particles which cause them to slip against each other. Water within concrete has a large influence on the amount of slip occurring between these particles as a result of the action of weakening attractive forces binding those particles. Creep in concrete is not fully understood because it involves many factors that influence the total amount of creep a specimen experiences. The factors that considerably influence concrete creep are the aggregate, cement type, water-cement ratio, member size, curing condition, temperature and relative humidity, age of loading, and stress intensity.

The aspects of the aggregate that influence creep are the volume and the material properties. The aggregate serves to restrain creep from occurring; therefore, the magnitude of creep largely depends on the quantity and properties of the aggregate added into the mixture. Creep is inversely related to the aggregate. The content and type of cement paste are also key affecters on concrete creep. This is because creep mainly occurs in the hydrated cement paste that surrounds the aggregate. The amount of cement and amount of creep are directly correlated. Type of cement affects the strength of the concrete mixture at the time of loading. Therefore, when rapid hardening cements are used, the concrete matrix stiffness is increased at the time of loading allowing for the concrete to be more resistant to creep. Water-cement ratio is also directly related to creep. Lower water content results in higher concrete strength and thus fewer pores in the mature cement. This leads to a decrease in the amount of creep. It is thought that creep deformation decreases with an increase in member size or concrete specimen which is highly related to mobility of moisture in the concrete mixture. It has been shown that curing conditions have a substantial effect on the maturity of

concrete, and thus concrete strength. Longer curing durations tend to lead to an increase in strength and consequently decrease in creep. Ambient conditions such as temperature and relative humidity are also sources that influence creep. Higher temperatures traditionally result in higher creep until a maximum is reached in the vicinity of 71 °C at which point further rise in temperature results in a decrease in creep. Humidity results in moisture within the concrete to be diffused to the ambient, thus allowing for more creep. Due to the phenomenon that concrete strength increases as it matures, age of loading has an inverse correlation with creep. Finally, it is general acceptance that the amount of creep is approximately proportional to the applied stress; however, this proportionality is only valid up to 0.2 to 0.5 times the ultimate strength.

Several mathematical models have been suggested to estimate the amount of creep in concrete. Most of the creep-prediction models developed were empirically derived based on the outcomes of experiments carried out on concrete specimens. Here, four empirical models recommended by different committees are used to predict the amount of creep that a concrete specimen experiences under a particular environment. The four creep-prediction models are the equations contained in the ACI 209 (1997) (209 1997), the AASHTO (2007) (AASHTO 2007), the CEB-FIP (1990) (CEB 1990), and the JSCE (1996) (Sakata and Shimomura 2004) design codes. For a given condition, the empirical models for predicting creep strain, given by these four codes as a function of time (number of days after onset of loading), are given in Eqs. (27–30) respectively.

$$M_1 = 1297 \cdot 10^{-1} \frac{t^{0.6}}{10 + t^{0.6}} \quad (27)$$

$$M_2 = 671 \cdot 10^{-6} \frac{t}{24.7 + t} \quad (28)$$

$$M_3 = 1527 \cdot 10^{-6} \left(\frac{t}{307.16 + t} \right)^{0.3} \quad (29)$$

$$M_4 = 1039 \cdot 10^{-6} \left(1 - \exp \left(-0.09(1.14(t + 28) - 31.8)^{0.6} \right) \right) \quad (30)$$

The transient prediction of creep strain as predicted by each of the four models is plotted in Fig. 2. As can be seen, there is some discrepancy among the four different models. Consequently, this discrepancy is the source of model-form uncertainty. The uncertainty that arises as a result of not knowing with perfect certainty the true system response and thus which model among those available is most accurate at predicting said result.

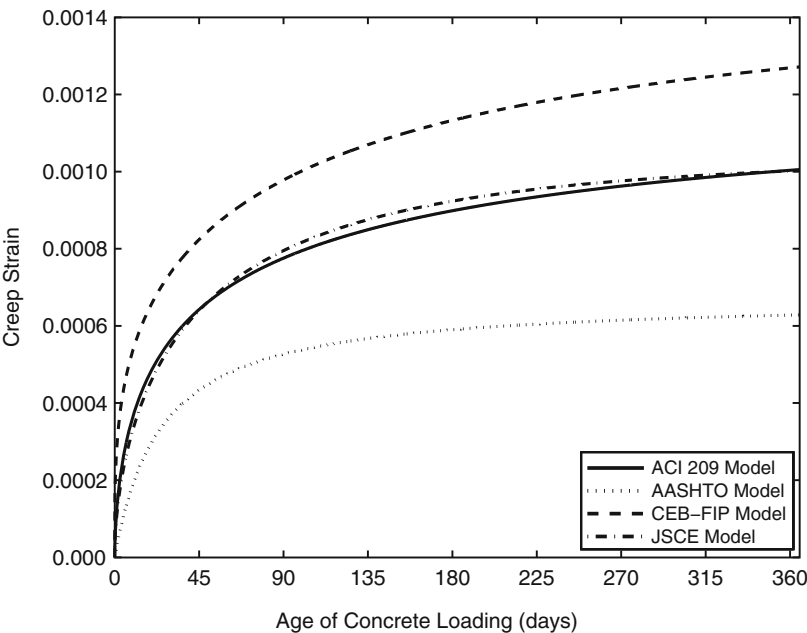
Utilizing the additive adjustment factor approach to quantify the aforementioned model-form uncertainty, model probabilities must first be assigned using expert opinion in the absence of experimental data. The model probabilities were taken to be 0.5, 0.15, 0.05, and 0.3 for models M_1 , M_2 , M_3 , and M_4 respectively. Note that this approach is only applicable to deterministic model predictions. Application of this technique at each point in the time domain is plotted in Fig. 3, where the upper and lower 95 % confidence bounds are determined using Eqs. 31 and 32 respectively.

$$95\% \text{ Upper Bound} = E(y|D) - 1.96 \sqrt{\sum_{k=1}^K \Pr(M_k|D) \sigma_k^2 + \sum_{k=1}^K \Pr(M_k|D) (f_k - E(y|D))^2} \quad (31)$$

$$95\% \text{ Lower Bound} = E(y|D) - 1.96 \sqrt{\sum_{k=1}^K \Pr(M_k|D) \sigma_k^2 + \sum_{k=1}^K \Pr(M_k|D) (f_k - E(y|D))^2} \quad (32)$$

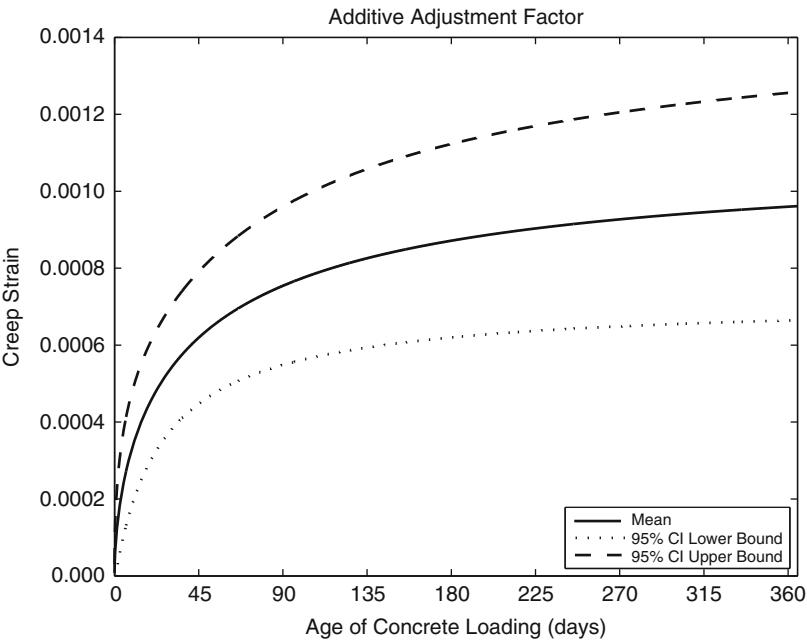
Model-Form Uncertainty Quantification for Structural Design,

Fig. 2 Plot of each of the four models used for prediction of creep strain as a function of time



Model-Form Uncertainty Quantification for Structural Design,

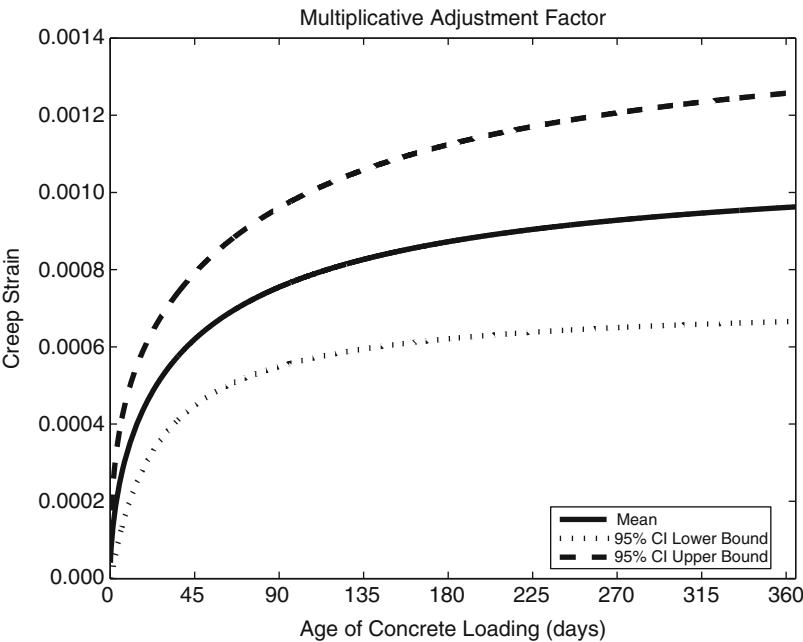
Fig. 3 Transient response obtained utilizing an additive adjustment factor approach



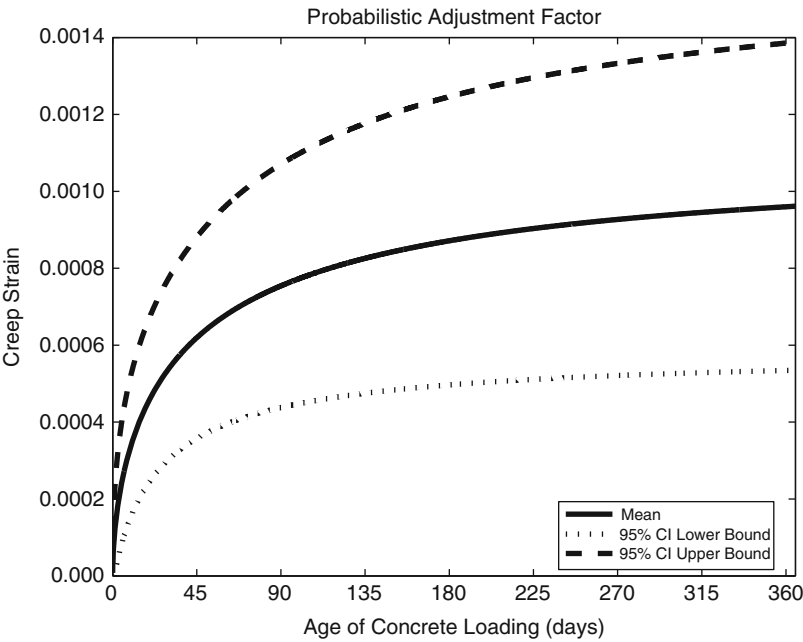
Utilizing the multiplicative adjustment factor approach to quantify the aforementioned model-form uncertainty, model probabilities in the same manner as with the additive adjustment factor approach. Note that this approach is also only

applicable to deterministic model predictions. Application of this technique at each point in the time domain is plotted in Fig. 4. It can be seen that the latter two model-form uncertainty quantification techniques yield nearly identical results.

Model-Form Uncertainty Quantification for Structural Design,
Fig. 4 Transient response obtained utilizing a multiplicative adjustment factor approach



Model-Form Uncertainty Quantification for Structural Design,
Fig. 5 Transient response obtained utilizing a probabilistic adjustment factor approach



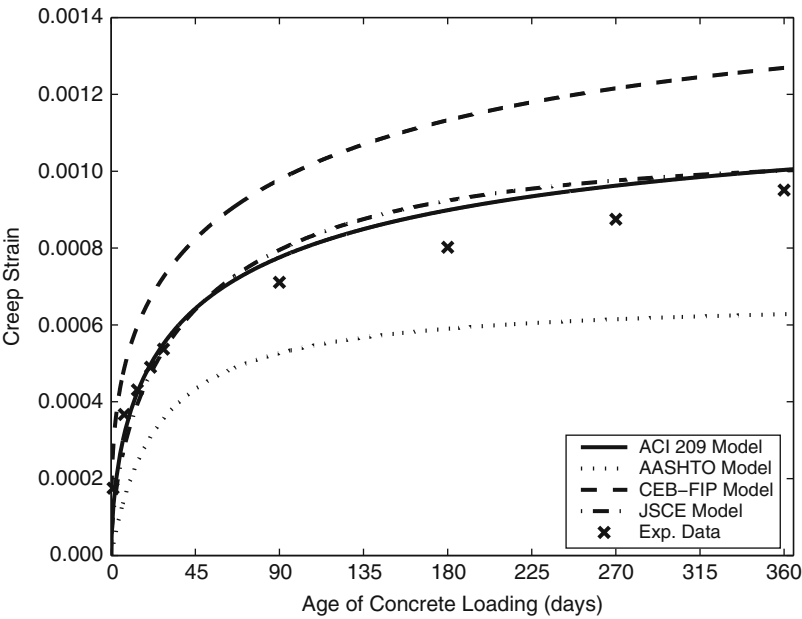
Utilization of the probabilistic adjustment factor approach to quantify the aforementioned model-form uncertainty is performed in a nearly identical manner as the previous two approaches.

Model probabilities defined for the previous two techniques were also applied in this approach. Note that this approach is applicable to probabilistic model predictions, thus allowing for the

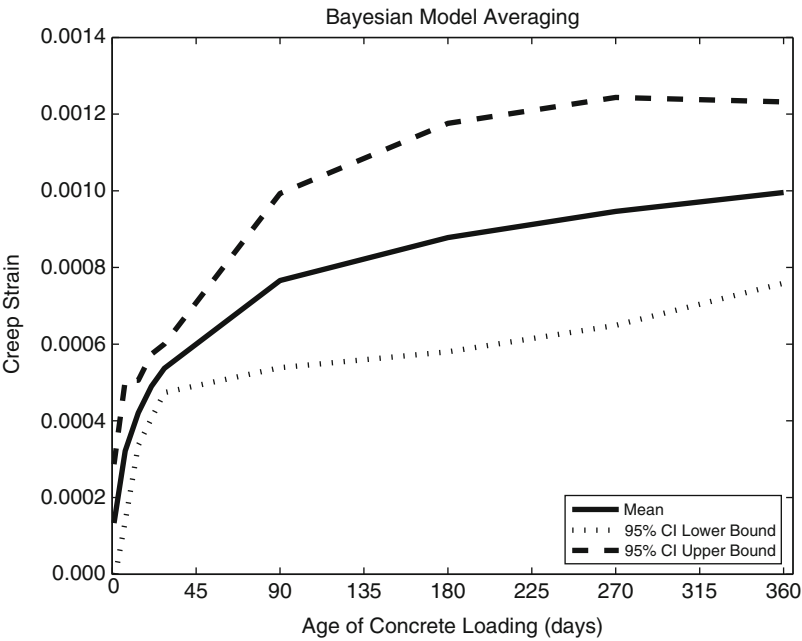
Model-Form Uncertainty Quantification for Structural Design, Table 1 Measured creep strains of tested con-
creted specimen

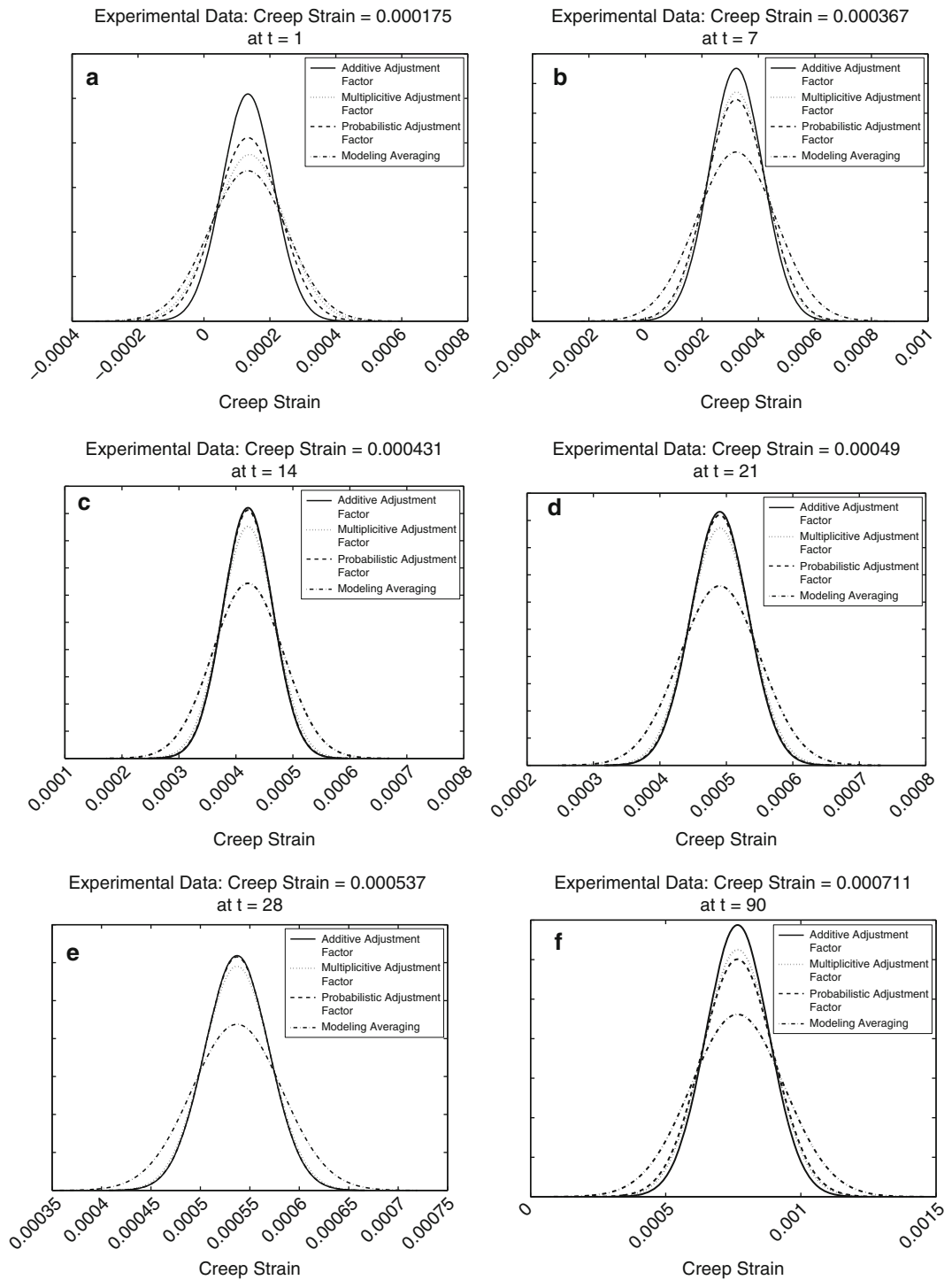
Age of Concrete after loading	1	7	14	21	28	90	180	270	360
Measured Creep Strain (10^{-6})	175	367	431	490	537	711	802	875	951

**Model-Form Uncertainty
Quantification for
Structural Design,
Fig. 6** Plot of each of the
four models used for
prediction of creep strain as
a function of time with
measured creep strains

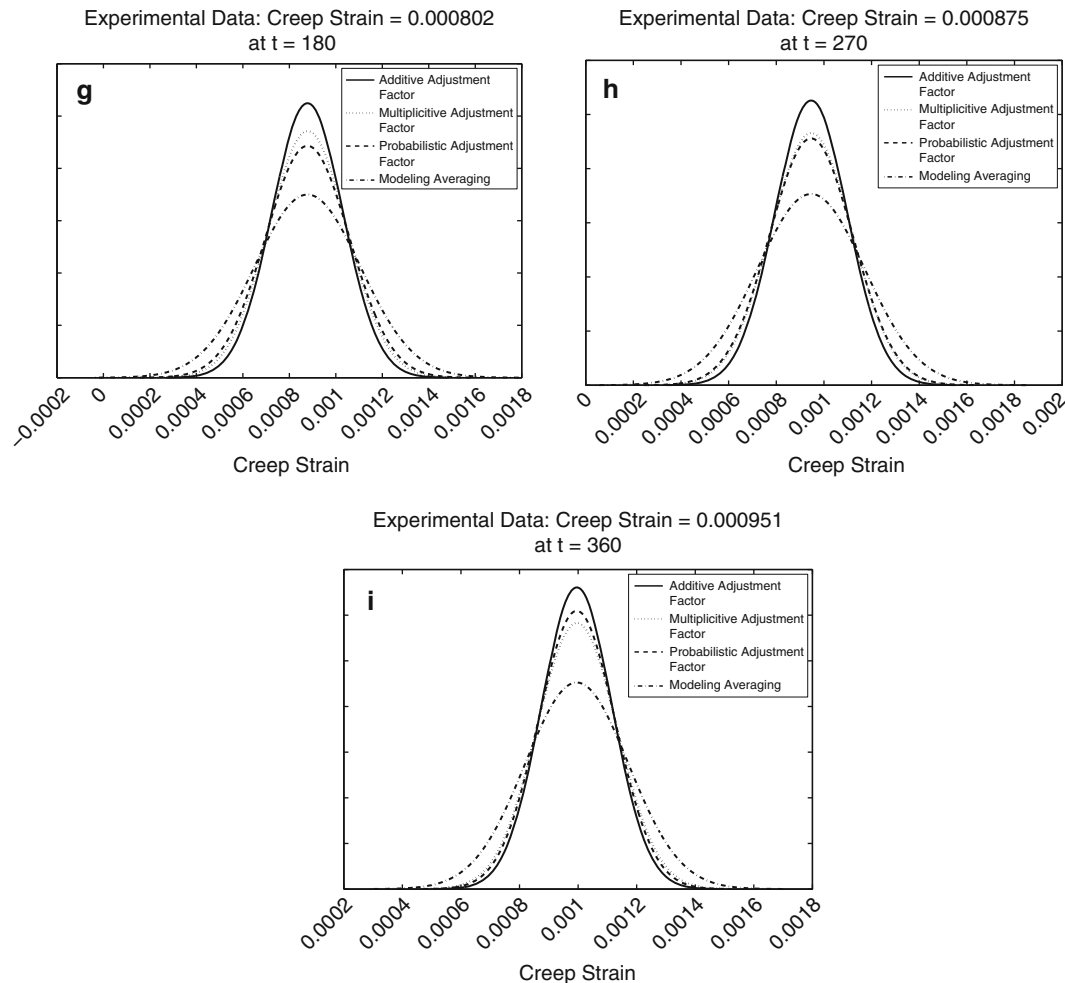


**Model-Form Uncertainty
Quantification for
Structural Design,
Fig. 7** Transient response
obtained utilizing Bayesian
model averaging approach





Model-Form Uncertainty Quantification for Structural Design, Fig. 8 (continued)



Model-Form Uncertainty Quantification for Structural Design, Fig. 8 Model-form uncertainty quantification results

incorporation of parametric or predictive uncertainties. Here, predictive uncertainties of 15, 19, 20, and 16 % of the model prediction are accounted for in models M_1 , M_2 , M_3 , and M_4 respectively. Application of this technique at each point in the time domain is plotted in Fig. 5. It can be seen that this model-form uncertainty quantification technique yields wider confidence intervals than the previous two approaches due to incorporation of predictive uncertainties.

As mentioned previously, the Bayesian model averaging technique for quantifying model-form uncertainty requires the availability of

experimental data for proper application. The data given in Table 1 contains experimental data of measured creep strain to be used in applying the Bayesian model averaging approach. This data is plotted along with the transient responses of Fig. 2 in Fig. 6. Bayes' theorem is applied using this measured data and uniformly distributed prior model probabilities (therefore each model assumes a probability of 0.25). Upon updating prior probabilities using Bayes' theorem and obtaining maximum likelihood estimators, at which point Bayesian model averaging is carried out. The transient response is plotted in Fig. 7. It can be seen that this response is not as

seamless as the adjustment factor approaches. This is a result of the fact that this approach can only be implemented at the points that experimental data is available.

Figure 8 shows the normal distribution generated by each model-form uncertainty quantification technique, at each point an experimental data point is available. The results in this figure were all applied on model probabilities updated from uniformly distributed model probabilities using Bayes' theorem. The model variances required for applying the probabilistic adjustment factor approach are assumed to be the maximum likelihood estimators found in performing Bayes' theorem. It can be seen that the additive and multiplicative adjustment factor approaches yield nearly identical results; whereas the probabilistic adjustment factor approach yields slightly different results. The Bayesian model averaging approach on the other hand yields the most conservative distribution. Note that the adjustment factor approaches can be applied under conditions necessary for the BMA approach, as shown here, but the opposite is not possible.

Summary

Model-form uncertainty arises as a result of lack of knowledge or certainty about the models available for predicting a given response. Four methodologies were presented for the quantification of model-form uncertainty in structural design. These methodologies were the additive adjustment factor, multiplicative adjustment factor, probabilistic adjustment factor, and Bayesian model averaging approaches. Limitations on these techniques include a constraint that experimental data points be available for Bayesian model averaging and that model predictions be deterministic for the additive and multiplicative adjustment factor approaches. The probabilistic adjustment factor approach, a novel derivation of the additive adjustment factor approach, can handle stochastic model predictions in contrast to that from which it was derived.

Cross-References

- [Bayesian Statistics: Applications to Earthquake Engineering](#)
- [Model Class Selection for Prediction Error Estimation](#)
- [Physics-Based Ground-Motion Simulation](#)
- [Probabilistic Seismic Hazard Models](#)
- [Reliability Estimation and Analysis](#)
- [Site Response for Seismic Hazard Assessment](#)
- [Structural Reliability Estimation for Seismic Loading](#)
- [Uncertainty Quantification in Structural Health Monitoring](#)
- [Uncertainty Theories: Overview](#)

References

- 209 AC (1997) Aci 209r-92: prediction of creep, shrinkage, and temperature effects in concrete structures (reapproved 1997). Technical report, American Concrete Institute, Farmington Hills
- AASHTO (2007) Aashto lfrd bridge design specifications: customary u.s. units, 3rd edn. Technical report, American Association of State Highway and Transportation Officials (AASHTO), Washington, DC
- Allaire D, Willcox K (2010) Surrogate modeling for uncertainty assessment with application to aviation environmental system models. *AIAA J* 48(8): 1791–1803
- CEB (1990) Creep and shrinkage, in cebp model code 1990. Technical report, Comité Euro-International du Béton (CEB), Lausanne
- Chernoff H, Moses LE (2012) Elementary decision theory. Courier Dover Publications, New York
- Draper D (1995) Assessment and propagation of model uncertainty. *J R Stat Soc Ser B (Methodol)* 57:45–97
- Droguett EL, Mosleh A (2008) Bayesian methodology for model uncertainty using model performance data. *Risk Anal* 28(5):1457–1476
- Hacking I (1984) The emergence of probability: a philosophical study of early ideas about probability, induction and statistical inference. Cambridge University Press, London
- Kennedy MC, O'Hagan A (2001) Bayesian calibration of computer models. *J R Stat Soc Ser B (Stat Method)* 63(3):425–464
- Mosleh A, Apostolakis G (1986) The assessment of probability distributions from expert opinions with an application to seismic fragility curves. *Risk Anal* 6(4):447–461

- Park I, Amarchinta HK, Grandhi RV (2010) A Bayesian approach for quantification of model uncertainty. *Reliab Eng Syst Saf* 95(7):777–785
- Park I, Grandhi RV (2011) Quantifying multiple types of uncertainty in physics-based simulation using Bayesian model averaging. *AIAA J* 49(5):1038–1045
- Paté-Cornell ME (1996) Uncertainties in risk analysis: six levels of treatment. *Reliab Eng Syst Saf* 54(2):95–111
- Reinert JM, Apostolakis GE (2006) Including model uncertainty in risk-informed decision making. *Ann Nucl Energy* 33(4):354–369
- Riley ME, Grandhi RV (2011a) A method for the quantification of model-form and parametric uncertainties in physics-based simulations. In: 52nd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference, Denver. AIAA
- Riley ME, Grandhi RV (2011b) Quantification of model-form and predictive uncertainty for multi-physics simulation. *Comput Struct* 89(11):1206–1213
- Robert C (2007) *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer, New York
- Sakata K, Shimomura T (2004) Recent progress in research on and code evaluation of concrete creep and shrink-age in Japan. *J Adv Concr Technol* 2:133–140
- Soares CG (1997) Quantification of model uncertainty in structural reliability. In: *Probabilistic methods for structural design*. Springer, Norwell, MA, pp 17–37
- Zhang R, Mahadevan S (2000) Model uncertainty and Bayesian updating in reliability-based inspection. *Struct Saf* 22(2):145–160
- Zio E, Apostolakis G (1996) Two methods for the structured assessment of model uncertainty by experts in performance assessments of radioactive waste repositories. *Reliab Eng Syst Saf* 54(2):225–241

Moment Tensors: Decomposition and Visualization

Václav Vavryčuk

Institute of Geophysics, Czech Academy of Sciences, Prague, Czech Republic

Synonyms

Dynamics and mechanics of faulting; Earthquake source observations; Focal mechanism; Seismic anisotropy; Theoretical seismology

Introduction

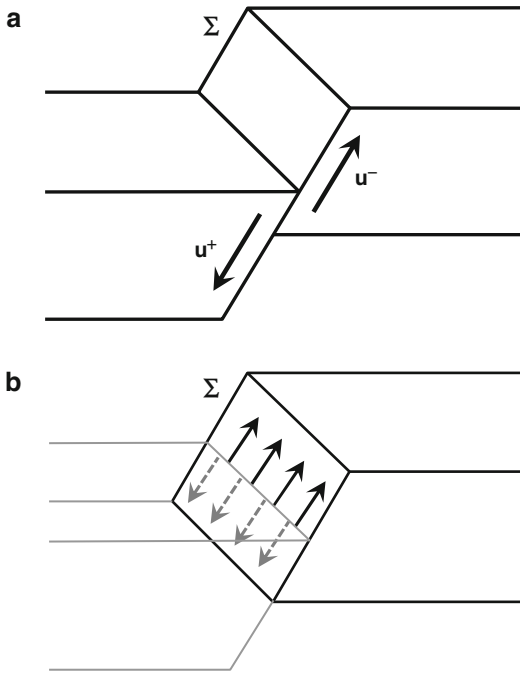
Elastic waves are generated by forces acting at the source and affected by the response of a medium to these forces. Mathematically, they are expressed using the representation theorem (Aki and Richards 2002, Eq. 2.41):

$$u_i(\mathbf{x}, t) = \int_{-\infty}^{\infty} \iiint_V f_k(\boldsymbol{\xi}, \tau) G_{ik}(\mathbf{x}, t; \boldsymbol{\xi}, \tau) d\tau dV(\boldsymbol{\xi}), \quad (1)$$

where $u_i = u_i(\mathbf{x}, t)$ is the observed displacement field, $\mathbf{x}$ is the position vector of an observer, t is time, and Green's tensor $G_{ik} = G_{ik}(\mathbf{x}, t; \boldsymbol{\xi}, \tau)$ is the solution of the equation of motion for a point single force with time dependence of the Dirac delta function. The Green's tensor is defined as the i th component of the displacement produced by a force in the x_k direction and describes propagation effects on waves along a path from the source to a receiver. Vector $f_k = f_k(\boldsymbol{\xi}, \tau)$ is the density of the body forces acting at the source being a function of the position vector $\boldsymbol{\xi}$ and time τ at the source. The integration in Eq. 1 is over focal volume V and time τ . For simplicity, an infinite medium is assumed in Eq. 1.

Seismic waves generated at an earthquake source and propagating in the Earth have some specific properties. Firstly, the body forces associated with the earthquake source are not distributed in a volume but along a fault. Secondly, the earthquake source is not represented by single forces but rather by dipole forces. The dipole forces cause the two blocks at opposite sides of the fault to mutually move (Fig. 1a). They are described by moment tensor density $m_{kl} = m_{kl}(\boldsymbol{\xi}, \tau)$ defined along fault Σ (Fig. 1b). A substitution of single forces by dipole forces leads to modifying Eq. 1 as follows (Burridge and Knopoff 1964; Aki and Richards 2002, Eq. 3.20):

$$u_i(\mathbf{x}, t) = \int_{-\infty}^{\infty} \iint_{\Sigma} m_{kl}(\boldsymbol{\xi}, \tau) \frac{\partial}{\partial \xi_l} G_{ik}(\mathbf{x}, t; \boldsymbol{\xi}, \tau) d\tau d\Sigma(\boldsymbol{\xi}). \quad (2)$$



Moment Tensors: Decomposition and Visualization, Fig. 1 (a) Example of motion (*normal faulting*) and (b) distribution of equivalent dipole forces along fault Σ

If the size of the fault is small with respect to distance between the source and the receiver, the representation theorem can be simplified by introducing the point source approximation:

$$u_i(\mathbf{x}, t) = \int_{-\infty}^{\infty} M_{kl}(\tau) \frac{\partial}{\partial \xi_l} G_{ik}(\mathbf{x}, t; \xi, \tau) d\tau, \quad (3)$$

or simply

$$u_i = M_{kl} * G_{ik,l}, \quad (4)$$

where $M_{kl} = M_{kl}(t)$ is the seismic moment tensor

$$M_{kl} = \iint_{\Sigma} m_{kl} d\Sigma, \quad (5)$$

and symbol “*” means the time convolution. Integration in Eq. 5 is performed over fault Σ . Moment tensor $\mathbf{M}$ is a symmetric tensor describing nine couples of equivalent dipole forces which can act at the earthquakes source (Fig. 2).

It is a basic quantity evaluated for earthquakes on all scales from acoustic emissions to large devastating earthquakes (see entries “► [Long-Period Moment-Tensor Inversion: The Global CMT Project](#),” “► [Reliable Moment Tensor Inversion for Regional- to Local-Distance Earthquakes](#),” and “► [Regional Moment Tensor Review: An Example from the European–Mediterranean Region](#)”).

The most common type of the moment tensor is the double-couple (DC) source which represents the force equivalent of shear faulting on a planar fault in isotropic media. However, many studies reveal that seismic sources often display more general moment tensors with significant non-double-couple (non-DC) components (Julian et al. 1998; Miller et al. 1998; see entry “► [Non-Double-Couple Earthquakes](#)”). An explosion is an obvious example of a non-DC source, but non-DC components can also be produced by a collapse of a cavity in mines (Rudajev and Šílený 1985), by inflation or deflation of magma chambers in volcanic areas (Mori and McKee 1987), by shear faulting on a nonplanar (curved or irregular) fault, by tensile faulting induced by fluid injection when the slip vector is inclined from the fault and causes its opening (Vavryčuk 2001, 2011), or by shear faulting in anisotropic media (Kawasaki and Tanimoto 1981; Vavryčuk 2005).

Moment Tensor Decomposition

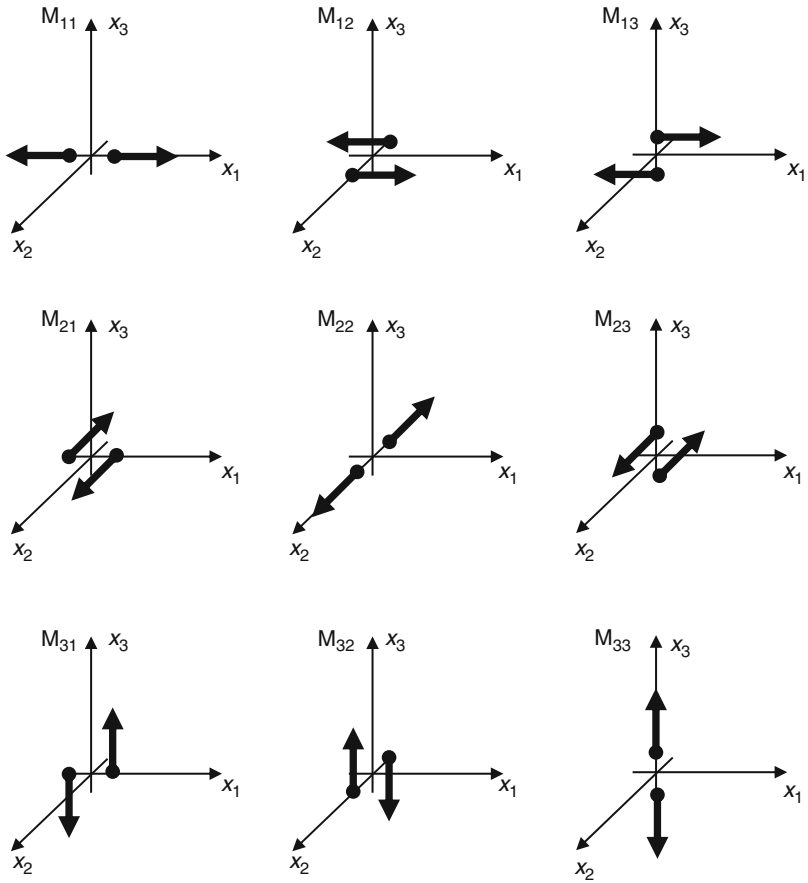
In order to identify which type of seismic source is physically represented by a retrieved moment tensor $\mathbf{M}$, the moment tensor is usually diagonalized and decomposed into some elementary parts. The first step is the decomposition into its isotropic (ISO) and deviatoric (DEV) parts:

$$\mathbf{M} = \mathbf{M}_{\text{ISO}} + \mathbf{M}_{\text{DEV}}, \quad (6)$$

where

$$\mathbf{M}_{\text{ISO}} = \frac{1}{3} \text{Tr}(\mathbf{M}) \cdot \mathbf{I}, \quad (7)$$

**Moment Tensors:
Decomposition and
Visualization, Fig. 2** Set
of nine couples of
equivalent dipole forces
forming the moment tensor



and matrix $\mathbf{I}$ is the identity matrix. The second step is the decomposition of the deviatoric part of $\mathbf{M}$. This step is more ambiguous and can be performed in many alternative ways. The deviatoric part can be decomposed into three double couples (Jost and Herrmann 1989), into major and minor double couples (Kanamori and Given 1981; Wallace 1985), into double couples with the same T axis (Wallace 1985), or into a double couple and a compensated linear vector dipole (CLVD) component (Knopoff and Randall 1970). The last decomposition into the DC and CLVD components proved to be useful for physical interpretations and became widely accepted. This decomposition was further developed and applied by Sipkin (1986), Jost and Herrmann (1989), Kuge and Lay (1994), Vavryčuk (2015), and others, and it will be treated here in detail.

Definition of ISO, CLVD, and DC

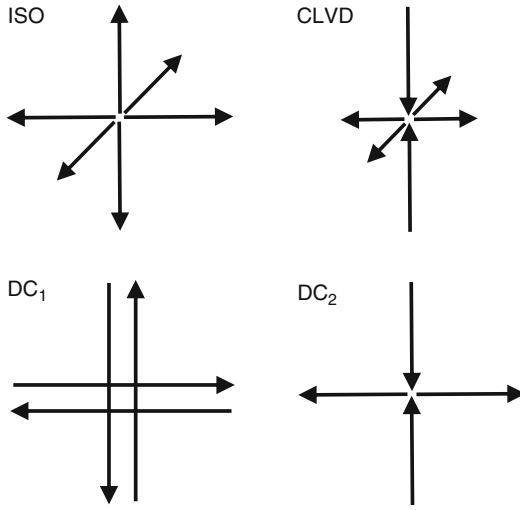
The seismic moment tensor $\mathbf{M}$ can be decomposed using eigenvalues and an orthonormal basis of eigenvectors in the following way:

$$\mathbf{M} = M_1 \mathbf{e}_1 \otimes \mathbf{e}_1 + M_2 \mathbf{e}_2 \otimes \mathbf{e}_2 + M_3 \mathbf{e}_3 \otimes \mathbf{e}_3, \quad (8)$$

where

$$M_1 \geq M_2 \geq M_3, \quad (9)$$

and vectors $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ define the T (tension), N (intermediate or neutral), and P (pressure) axes, respectively. Symbol “ $\otimes$ ” in Eq. 8 means the dyadic product of two vectors. Two basic properties of the moment tensor are separated in Eq. 8: (1) the orientation of the source defined by three eigenvectors and (2) the type and size of the source defined by three eigenvalues of $\mathbf{M}$.



Moment Tensors: Decomposition and Visualization, Fig. 3 The ISO, DC, and CLVD⁺ base tensors of the moment tensor. The DC part is plotted in the original coordinate system associated with the fault (DC₁) and after its diagonalization (DC₂)

The eigenvalues can be represented as a point in three-dimensional (3-D) space (Riedesel and Jordan 1989):

$$\mathbf{m} = M_1 \hat{\mathbf{e}}_1 + M_2 \hat{\mathbf{e}}_2 + M_3 \hat{\mathbf{e}}_3, \quad (10)$$

where vectors $\hat{\mathbf{e}}_1$, $\hat{\mathbf{e}}_2$, and $\hat{\mathbf{e}}_3$ define the coordinate system in this space. In order to get a unique representation, the eigenvalues must be ordered according to Eq. 9. Consequently, the points representing the source type cannot cover the whole 3-D space but only its wedge called the “source-type space.” The choice of the coordinate system and the metric for parameterizing the source-type space differ for individual moment tensor decompositions.

For physical reasons, the three terms in Eq. 8 are further restructured to form isotropic (ISO), double-couple (DC), and compensated linear vector dipole (CLVD) parts (Fig. 3) in the following way (Knopoff and Randall 1970; Dziewonski et al. 1981; Sipkin 1986; Jost and Herrmann 1989):

$$\begin{aligned} \mathbf{M} &= \mathbf{M}_{\text{ISO}} + \mathbf{M}_{\text{DC}} + \mathbf{M}_{\text{CLVD}} \\ &= M_{\text{ISO}} \mathbf{E}_{\text{ISO}} + M_{\text{DC}} \mathbf{E}_{\text{DC}} + M_{\text{CLVD}} \mathbf{E}_{\text{CLVD}}, \end{aligned} \quad (11)$$

where $\mathbf{E}_{\text{ISO}}$, $\mathbf{E}_{\text{DC}}$, and $\mathbf{E}_{\text{CLVD}}$ are the ISO, DC, and CLVD elementary (base) tensors and M_{ISO} , M_{DC} , and M_{CLVD} are the ISO, DC, and CLVD moments. The base tensors read

$$\begin{aligned} \mathbf{E}_{\text{ISO}} &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{E}_{\text{DC}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \\ \mathbf{E}_{\text{CLVD}}^+ &= \frac{1}{2} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad \mathbf{E}_{\text{CLVD}}^- = \frac{1}{2} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}, \end{aligned} \quad (12)$$

where $\mathbf{E}_{\text{CLVD}}^+$ or $\mathbf{E}_{\text{CLVD}}^-$ is used if $M_1 + M_3 - 2M_2 \geq 0$ or $M_1 + M_3 - 2M_2 < 0$, respectively. Hence, the CLVD tensor is aligned along the axis with the largest magnitude deviatoric eigenvalue. The eigenvalues of the base tensors are ordered according to Eq. 9 in order to lie in the source-type space. The norm of all base tensors, calculated as the largest magnitude eigenvalue (i.e., the maximum of $|M_i|$, $i = 1, 2, 3$), is equal to 1. This condition is called the unit “spectral norm” and physically means that the maximum dipole force of the base tensors is unity (Fig. 3). Alternative alignments of the CLVD tensor and other normalizations of the base tensors are also admissible (Chapman and Leaney 2012; Tape and Tape 2012) but have less straightforward physical interpretations.

Equations 11 and 12 uniquely define values M_{ISO} , M_{DC} , and M_{CLVD} expressed as follows:

$$M_{\text{ISO}} = \frac{1}{3}(M_1 + M_2 + M_3), \quad (13)$$

$$M_{\text{CLVD}} = \frac{2}{3}(M_1 + M_3 - 2M_2), \quad (14)$$

$$M_{\text{DC}} = \frac{1}{2}(M_1 - M_3 - |M_1 + M_3 - 2M_2|), \quad (15)$$

where M_{CLVD} includes also the sign of the elementary CLVD tensor. If the elementary CLVD tensor $\mathbf{E}_{\text{CLVD}}$ is considered with its sign as in Eq. 11, then M_{CLVD} should be calculated as the absolute value of Eq. 14. Values M_{ISO} , M_{DC} , and

M_{CLVD} in Eqs. 13, 14, and 15 are usually further normalized and expressed using scalar seismic moment M and relative scale factors C_{ISO} , C_{DC} , and C_{CLVD} :

$$\begin{bmatrix} C_{\text{ISO}} \\ C_{\text{CLVD}} \\ C_{\text{DC}} \end{bmatrix} = \frac{1}{M} \begin{bmatrix} M_{\text{ISO}} \\ M_{\text{CLVD}} \\ M_{\text{DC}} \end{bmatrix}, \quad (16)$$

where M reads

$$M = |M_{\text{ISO}}| + |M_{\text{CLVD}}| + M_{\text{DC}}, \quad (17)$$

or equivalently (Bowers and Hudson 1999)

$$M = ||\mathbf{M}_{\text{ISO}}|| + ||\mathbf{M}_{\text{DEV}}||, \quad (18)$$

where $||\mathbf{M}_{\text{ISO}}||$ and $||\mathbf{M}_{\text{DEV}}||$ are the spectral norms of the isotropic and deviatoric parts of moment tensor $\mathbf{M}$, respectively. Scale factors C_{ISO} , C_{DC} , and C_{CLVD} satisfy the following equation:

$$|C_{\text{ISO}}| + |C_{\text{CLVD}}| + C_{\text{DC}} = 1. \quad (19)$$

Equations 13, 14, 15, 16, and 17 imply that C_{DC} is always positive and in the range from 0 to 1; C_{CLVD} and C_{ISO} are in the range from -1 to 1. Consequently, the decomposition of $\mathbf{M}$ can be expressed as

$$\mathbf{M} = M(C_{\text{ISO}}\mathbf{E}_{\text{ISO}} + C_{\text{DC}}\mathbf{E}_{\text{DC}} + |C_{\text{CLVD}}|\mathbf{E}_{\text{CLVD}}), \quad (20)$$

where M is the norm of $\mathbf{M}$ calculated using Eq. 17 and represents a scalar seismic moment for a general seismic source. The absolute value of the CLVD term in Eq. 20 is used because the sign of CLVD is included in the elementary tensor $\mathbf{E}_{\text{CLVD}}$.

Physical Properties of the Decomposition

The above decomposition of the moment tensor is performed in order to physically interpret a set of nine dipole forces representing a general point seismic source and to identify easily some basic types of the source in isotropic media:

- The explosion/implosion is an isotropic source, and thus, it is characterized by $C_{\text{ISO}} = \pm 1$ and by zero C_{CLVD} and C_{DC} .
- Shear faulting is represented by the double-couple force and characterized by $C_{\text{DC}} = 1$ and by zero C_{ISO} and C_{CLVD} .
- Pure tensile or compressive faulting is free of shearing and thus characterized by zero C_{DC} . However, the non-DC components contain both ISO and CLVD. The ISO and CLVD components are of the same sign: they are positive for tensile faulting but negative for compressive faulting.
- Shear-tensile (dislocation) source defined as the source, which combines both shear and tensile faulting (Vavryčuk 2001, 2011), is characterized by nonzero ISO, CLVD, and DC components. The positive values of C_{ISO} and C_{CLVD} correspond to tensile mechanisms when fault is opening during rupturing. The negative values of C_{ISO} and C_{CLVD} correspond to compressive mechanisms when a fault is closing during rupturing. The ratio between non-DC and DC components defines the angle between the slip and the fault.
- Shear faulting on a nonplanar fault is characterized generally by a nonzero C_{DC} and C_{CLVD} . The C_{ISO} is zero, because no volumetric changes are associated with this type of source.

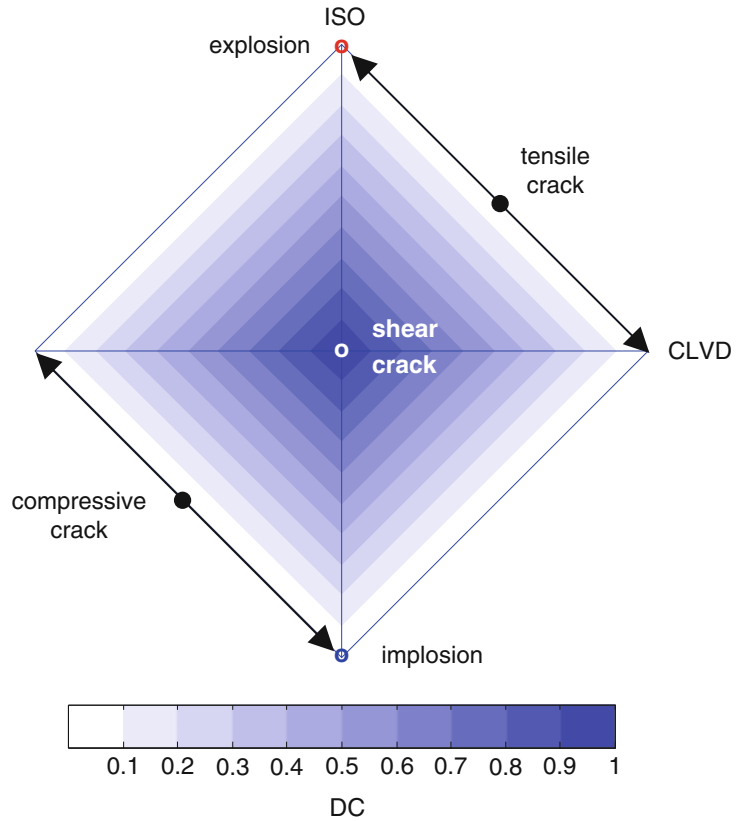
Source-Type Plots

For physical interpretations, it is advantageous to visualize the retrieved moment tensors graphically. Double-couple components of moment tensors are displayed using the well-known “beach balls” which show orientations of the fault together with the slip vector defining the shear motion along the fault (see entry “► Earthquake Mechanism Description and Inversion”). The non-double-couple components of moment tensors are displayed in the so-called source-type plots.

All moment tensors fill a source-type space which is a wedge in the full 3-D space. The magnitude of the vector in this space is the scalar

Moment Tensors: Decomposition and Visualization,

Fig. 4 Diamond CLVD-ISO source-type plot with positions of basic types of seismic sources. The arrows indicate the range of possible positions of moment tensors for pure tensile or compressive cracks



M

moment, and its direction defines the type of the source. In order to identify the type of the source visually, it is convenient to plot all unit vectors of the source-type space in a 2-D figure using some projections. Here, three basic plots are mentioned: diamond CLVD-ISO plot (Vavryčuk 2015), Hudson's skewed diamond plot (Hudson et al. 1989), and the Riedesel-Jordan lune plot (Riedesel and Jordan 1989).

Diamond CLVD-ISO Plot

The diamond CLVD-ISO plot shows the position of the source in the CLVD-ISO coordinate system in which the DC component is represented by the color intensity (Fig. 4). Since the sum of absolute values of the CLVD and ISO cannot exceed 1, moment tensors must lie inside a diamond. A source with pure or predominant shear faulting is located at the origin of coordinates or close to it. An explosion or implosion source is located at the top or bottom vertex of the

diamond, respectively. Motion on a pure tensile or compressive crack is plotted at the margin of the diamond. Points along the CLVD axis correspond to faulting on nonplanar faults, and points in the first and third quadrants of the diamond correspond to shear-tensile sources.

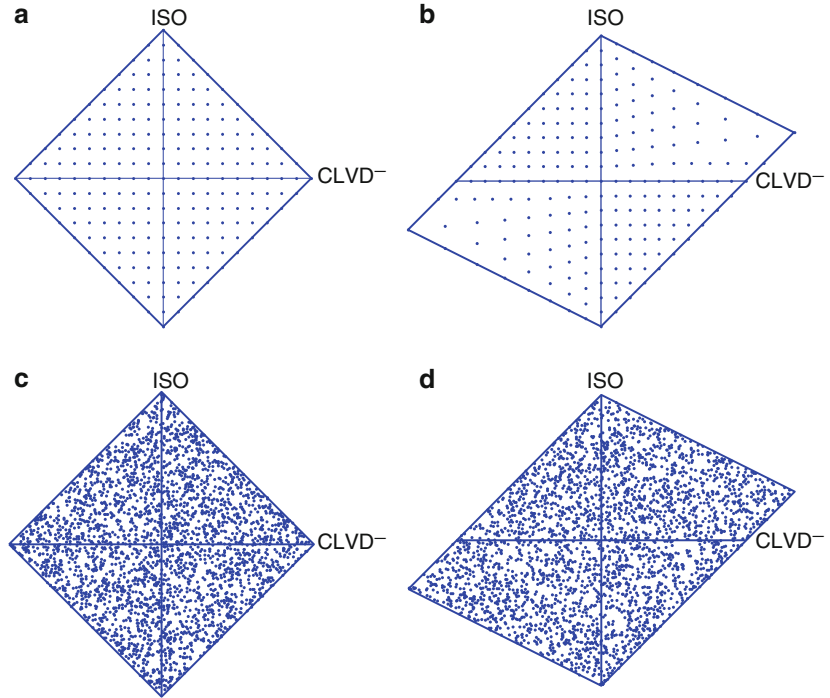
For pure tensile and shear-tensile sources, the ISO/CLVD ratio depends on the elastic properties of the medium surrounding the source. In isotropic media, this ratio is (Vavryčuk 2001, 2011)

$$\frac{C_{\text{ISO}}}{C_{\text{CLVD}}} = \frac{3}{4} \left(\frac{v_P}{v_S} \right)^2 - 1. \quad (21)$$

Hence, the point representing the pure tensile faulting in Fig. 4 (black dot) can be close to $C_{\text{ISO}} = 1$ (corresponding to an explosion) for high values of v_P/v_S but also close to $C_{\text{CLVD}} = 1$ for low values of v_P/v_S . The limiting cases are

Moment Tensors: Decomposition and Visualization, Fig. 5

The Hudson's diamond τ - k plot (a, c) and the skewed diamond u - v plot (b, d). The CLVD⁻ means the reversed CLVD axis. The dots in (a, b) show a regular grid in C_{ISO} and C_{CLVD} from -1 to $+1$ with step of 0.1 . The dots in (c, d) show 3,000 sources defined by moment tensors with randomly generated eigenvalues. For scaling of the eigenvalues, see the text. Plots (c) and (d) indicate that the distribution of random sources is uniform in both projections



$$\frac{v_P}{v_S} \rightarrow \infty \text{ and } \frac{v_P}{v_S} = \frac{2}{\sqrt{3}}, \quad (22)$$

describing fluids and the lower limit of stable solids ($\lambda = -2/3 \mu$), respectively. Similar conclusions can be drawn for pure compressive faulting (see Fig. 4).

Note that the abovementioned basic types of sources cannot be located in the second or fourth quadrants of the diamond source-type plot in Fig. 4. Moment tensors located in this area indicate numerical errors of the moment tensor inversion, a more complicated source model or faulting in anisotropic media.

As mentioned above, values $C_{\text{ISO}} = \pm 1$ and $C_{\text{DC}} = 1$ correspond to an explosion/implosion and to shear faulting in isotropic media, respectively. Their physical meaning is thus straightforward. However, the moment tensor with $C_{\text{CLVD}} = 1$ does not correspond to any simple seismic source, and the presence of CLVD in moment tensors often causes confusions and poses questions whether it is necessary to introduce the CLVD. The decomposition described above indicates that the CLVD component is required to render the decomposition

mathematically complete, and the CLVD component cannot thus be avoided. Although, it has no physical meaning itself, it can be interpreted physically in combination with the ISO component as a product of tensile faulting. In the case of a pure tensile crack, the CLVD component's major dipole is aligned with the normal to the crack surface and the volume change associated with the opening crack is described by the ISO component.

Hudson's Skewed Diamond Plot

Hudson et al. (1989) introduced two source-type plots: a diamond τ - k plot which is the diamond CLVD-ISO plot described in the previous section but with the opposite direction of the CLVD axis (Fig. 5a) and a skewed diamond u - v plot (Fig. 5b). The latter plot is introduced in order to conserve the uniform probability of moment tensor eigenvalues. If eigenvalues M_1 , M_2 , and M_3 have a uniform probability distribution between -1 and $+1$ and satisfy the ordering condition (9), then all points fill uniformly the skewed diamond plot. Axis u defines the deviatoric sources and axis v connects the pure explosive and implosive sources.

The moment tensor with arbitrary (but ordered) eigenvalues M_1, M_2 , and M_3 is projected into the u - v plot using the following equations:

$$\begin{aligned} u &= -\frac{2}{3M}(M_1 + M_3 - 2M_2), \\ v &= \frac{1}{3M}(M_1 + M_2 + M_3), \end{aligned} \quad (23)$$

where M is the scalar seismic moment calculated as the spectral norm of complete moment tensor $\mathbf{M}$

$$M = \max(|M_1|, |M_2|, |M_3|). \quad (24)$$

Equation 23 is similar to Eqs. 13 and 14 in the ISO-CLVD-DC decomposition except for scaling.

Figure 5a, b show mapping of a regular grid in C_{ISO} and C_{CLVD} calculated using Eqs. 13, 14, 15, 16, and 17 into the diamond CLVD-ISO plot and into the skewed diamond u - v plot. Figure 5b shows that the CLVD-ISO grid is deformed in the first and third quadrants of the u - v plot. Figure 5c, d demonstrate that sources with randomly generated eigenvalues cover uniformly the source-type plots. The uniform probability distribution function (PDF) is produced by the Hudson's skewed diamond plot (Fig. 5d) but also for the diamond CLVD-ISO plot. In this respect, the Hudson's skewed diamond plot does not provide any particular advantage compared to the standard CLVD-ISO plot (for details, see Tape and Tape 2012; Vavryčuk 2015).

Riedesel-Jordan Plot

A completely different approach is suggested by Riedesel and Jordan (1989) who introduce a compact plot displaying both the orientation and type of source on the focal sphere. Apparently, this plot looks simple and mathematically elegant but introduces difficulties. The moment tensor is represented by a vector defined in Eq. 10, and the coordinate axes $\hat{\mathbf{e}}_1$, $\hat{\mathbf{e}}_2$, and $\hat{\mathbf{e}}_3$ are identified with the T , N , and P axes of $\mathbf{M}$: $\mathbf{e}_1$, $\mathbf{e}_2$, and $\mathbf{e}_3$ defined in Eq. 8. The vector is normalized using the Euclidean norm

$$M = \sqrt{\frac{1}{2}(M_1^2 + M_2^2 + M_3^2)} \quad (25)$$

and projected on the sphere using a lower-hemisphere equal-area projection (see Fig. 6a).

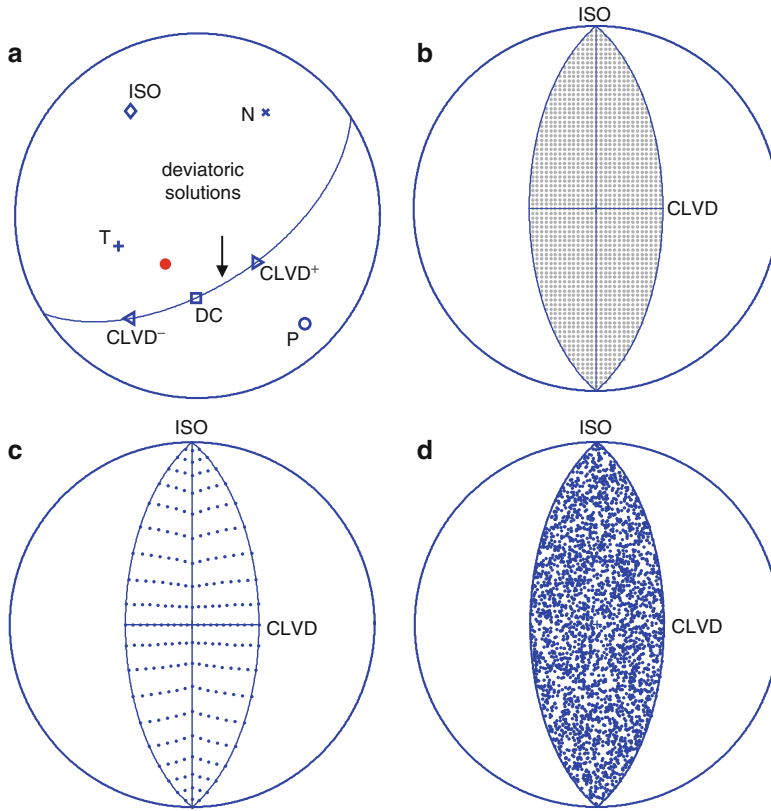
Chapman and Leaney (2012) pointed out, however, that this representation is not optimum for several reasons. Firstly, vector $\mathbf{m}$ cannot lie everywhere on the focal sphere but inside its small part called the “lune” (Tape and Tape 2012). The lune covers only one sixth of the whole sphere (see Fig. 6b). Secondly, vectors characterizing positive and negative isotropic sources (explosion and implosion) are physically quite different, but they are displayed in the same area on the focal sphere in this projection. Thirdly, analysis of uncertainties of a moment tensor solution by plotting a cluster of vectors $\mathbf{m}$ includes both effects – uncertainties in the orientation and in the source type. This is fine if the moment tensor is nondegenerate. However, difficulties arise when the moment tensor is degenerate or nearly degenerate, because small perturbations cause significant changes of eigenvectors.

Some of the mentioned difficulties can be avoided by fixing the eigenvectors and analyzing the size of clusters produced by a varying source type only. If we fix the eigenvectors in the form

$$\begin{aligned} \mathbf{e}_1 &= \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{6}}, \frac{1}{\sqrt{2}} \right)^T, \\ \mathbf{e}_2 &= \left(\frac{1}{\sqrt{3}}, -\frac{2}{\sqrt{6}}, 0 \right)^T, \\ \mathbf{e}_3 &= \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{2}} \right)^T, \end{aligned} \quad (26)$$

in the north-east-down coordinate system, we obtain a plot shown in Fig. 6b. This plot resembles the diamond CLVD-ISO plot (Fig. 4) but adapted to a spherical metric. The basic source types are characterized by the following unit vectors:

$$\mathbf{e}_{\text{ISO}} = \frac{1}{\sqrt{3}}(\mathbf{e}_1 + \mathbf{e}_2 + \mathbf{e}_3) = (1, 0, 0)^T, \quad (27)$$



Moment Tensors: Decomposition and Visualization, Fig. 6 Riedesel-Jordan source-type plot. (a) The original compact plot proposed by Riedesel and Jordan (1989) displaying the orientation of the moment tensor eigenvectors (P , T , and N axes), basic source types (ISO, CLVD, DC), and the position of the studied moment tensor (*red dot*). (b, c, d) A modified Riedesel-Jordan plot proposed

by Chapman and Leaney (2012). The *dashed area* in (b) shows the area of admissible positions of sources. The *dots* in (c) show a regular grid in C_{ISO} and C_{CLVD} from -1 to $+1$ with step of 0.1 . The *dots* in (d) show 3,000 sources defined by moment tensors with randomly generated eigenvalues. Plot (d) indicates that the distribution of random sources is uniform in this projection

$$\mathbf{e}_{\text{DC}} = \frac{1}{\sqrt{2}}(\mathbf{e}_1 - \mathbf{e}_3) = (0, 0, 1)^T, \quad (28)$$

$$\begin{aligned} \mathbf{e}_{\text{CLVD}}^+ &= \sqrt{\frac{2}{3}} \left(\mathbf{e}_1 - \frac{1}{2} \mathbf{e}_2 - \frac{1}{2} \mathbf{e}_3 \right) \\ &= \left(0, \frac{1}{2}, \frac{\sqrt{3}}{2} \right)^T, \end{aligned} \quad (29)$$

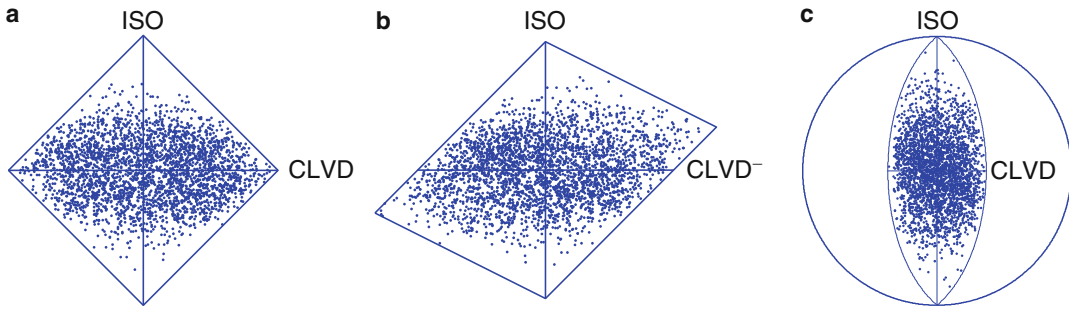
$$\begin{aligned} \mathbf{e}_{\text{CLVD}}^- &= \sqrt{\frac{2}{3}} \left(\frac{1}{2} \mathbf{e}_1 + \frac{1}{2} \mathbf{e}_2 - \mathbf{e}_3 \right) \\ &= \left(0, -\frac{1}{2}, \frac{\sqrt{3}}{2} \right)^T. \end{aligned} \quad (30)$$

Basic properties of the Riedesel-Jordan projection are exemplified in Fig. 6c, d. Figure 6c shows

mapping of a regular grid in C_{ISO} and C_{CLVD} calculated using Eqs. 13, 14, 15, 16, and 17, and Fig. 6d indicates that the PDF of sources with randomly distributed eigenvalues M_1 , M_2 , and M_3 is uniform. For a detailed analysis on the probability of eigenvalues in the spherical projection, see Tape and Tape (2012).

Analysis of Moment Tensor Uncertainties Using Source-Type Plots

The source-type plots are often used for assessing uncertainties of the ISO, CLVD, and DC components of moment tensors. The reason for using the source-type plots for assessing the errors is



Moment Tensors: Decomposition and Visualization, Fig. 7 Distribution of random sources displayed in three different source-type plots. (a) The diamond CLVD-ISO plot, (b) the Hudson's skewed diamond plot, and (c) the

Riedesel-Jordan plot. The *dots* show 3,000 sources defined by moment tensors with randomly generated components in the interval from -1 to $+1$. The distribution of sources is nonuniform for all three source-type plots

simple. The moment tensor is usually plotted as a cluster of acceptable solutions, and the size of the cluster reflects uncertainties of the solution. Such approach is, however, simplistic and rough because the same uncertainties produce differently large clusters in dependence of the position of the cluster. Although the source-type plots display a uniform PDF for randomly generated eigenvalues (see section “Source-Type Plots”), the behavior of moment tensor uncertainties is not simple. When uncertainties of moment tensor components are analyzed, the moment tensor is not in the diagonal form. After diagonalizing the moment tensor, the errors are projected into the errors of eigenvalues in a rather complicated way. This is demonstrated in Fig. 7. Moment tensors in this figure have all components random and distributed with a uniform probability in the interval from -1 to 1 . Nevertheless, some source types are quite rare. In particular, sources with a high explosive or implosive component are almost missing. This observation is common for all source-type plots.

More realistic sources are modeled in Fig. 8: the pure DC and ISO sources defined by tensors $\mathbf{E}_{\text{DC}}$ and $\mathbf{E}_{\text{ISO}}$ from Eq. 12 are contaminated by random noise with a uniform distribution in the interval from -0.25 to 0.25 . The noise is superimposed to all tensor components and 1,000 random moment tensors are generated. As expected, the randomly generated source tensors form clusters, but their shape is different for different projections and their size depends also

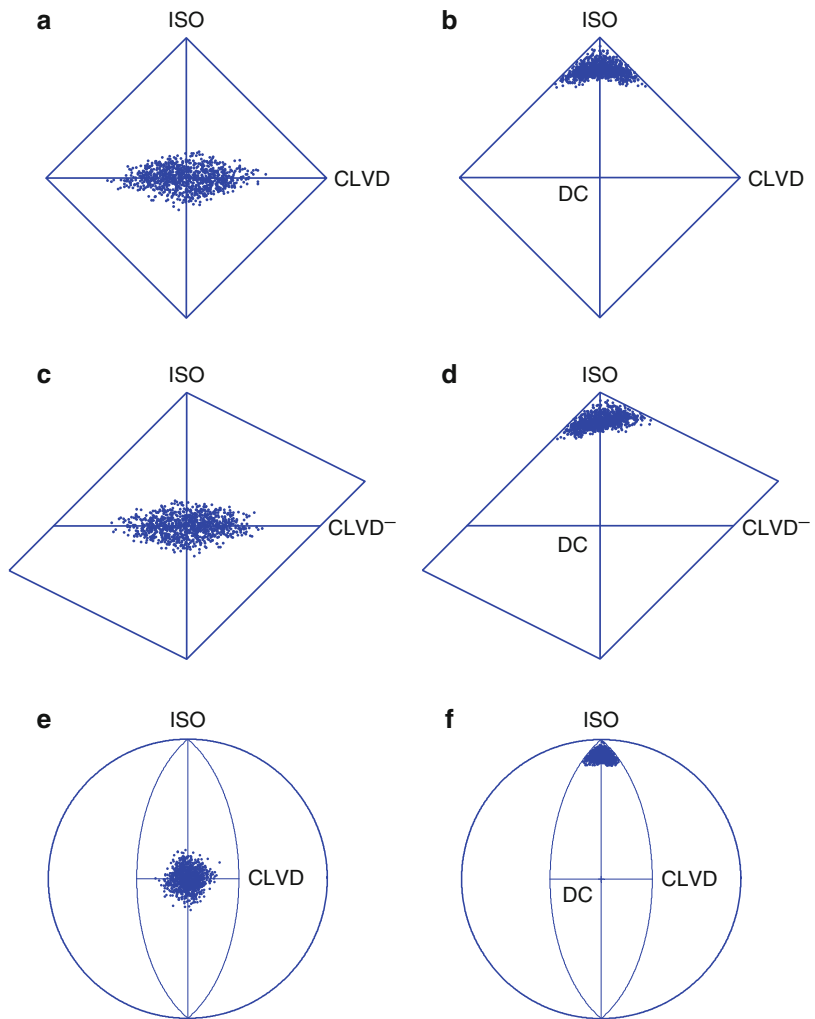
on the type of the source. For the DC source (Fig. 8, left-hand plots), the maximum PDF is in the center of the cluster which coincides with the position of the uncontaminated source. In the diamond CLVD-ISO plot and in the skewed diamond plot, the cluster is asymmetric being stretched along the CLVD axis. A more symmetric shape of the cluster is produced in the Riedesel-Jordan plot. However, the symmetry of the cluster is apparent because the CLVD and ISO axes are of different lengths. A significantly higher scatter of the CLVD components compared to the ISO components in moment tensor inversions has been observed and discussed also in Vavryčuk (2011). For the pure ISO source (Fig. 8, right-hand plots), the clusters are smaller than for the DC source, and the maximum PDF is out of the position of the uncontaminated source. This means that the ISO percentage is systematically underestimated due to errors of the inversion for highly explosive or implosive sources.

Source Tensor Decomposition

A simple classification of sources based on the moment tensor decomposition is possible in isotropic media only. In anisotropic media, the problem is more complicated. The moment tensor is affected not only by the geometry of faulting but also by the elastic properties of the focal zone. Depending on these properties, the moment tensors can take a general form with nonzero DC,

Moment Tensors: Decomposition and Visualization,

Fig. 8 Distribution of pure DC (left-hand plots) and pure explosive (right-hand plots) sources contaminated by random noise and displayed in three different source-type plots. (a, b) The diamond CLVD-ISO plot, (c, d) the Hudson's skewed diamond plot, and (e, f) the Riedesel-Jordan plot. The dots show 1,000 DC sources defined by the elementary tensor $\mathbf{E}_{DC}$ (see Eq. (12)) and contaminated by noise with a uniform distribution from -0.25 to $+0.25$. The noise is superimposed to all tensor components



CLVD, and ISO components even for simple shear faulting on a planar fault (Vavryčuk 2005). For this reason, physical interpretations of shear or tensile dislocation sources in anisotropic media should be based on the decomposition of the source tensor, which is directly related to geometry of faulting.

The source tensor $\mathbf{D}$ (also called the potency tensor) is a symmetric dyadic tensor defined as (Ben-Zion 2003; Vavryčuk 2005)

$$D_{kl} = \frac{uS}{2} (s_k n_l + s_l n_k), \quad (31)$$

where vectors $\mathbf{n}$ and $\mathbf{s}$ denote the fault normal and the direction of the slip vector, respectively, u is

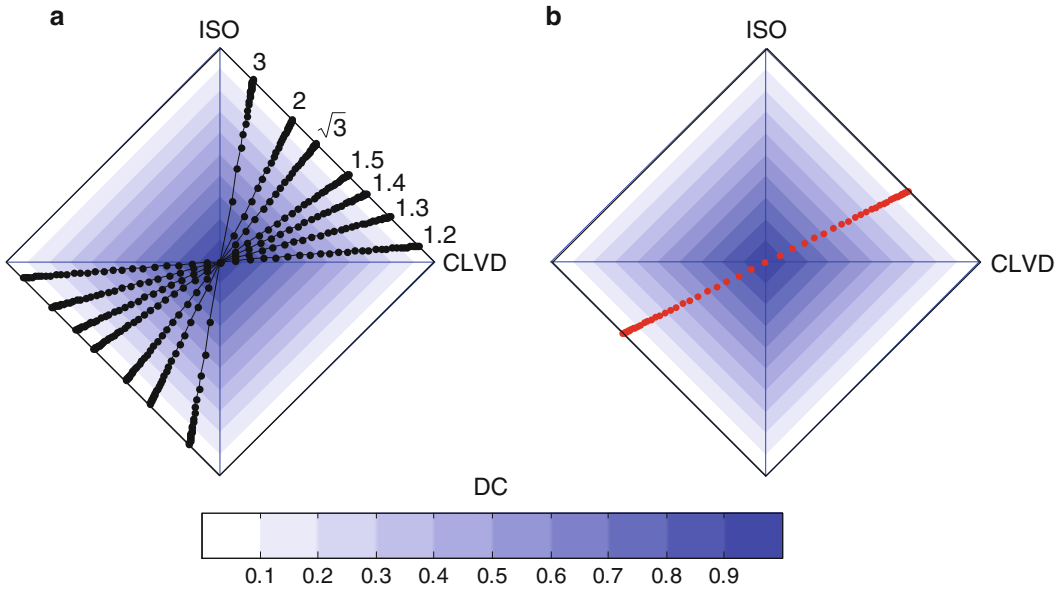
the slip and S is the fault size. The relation between the source and moment tensors reads in anisotropic media (Vavryčuk 2005, his Eq. 4)

$$M_{ij} = c_{ijkl} D_{kl}, \quad (32)$$

and in isotropic media

$$M_{ij} = \lambda D_{kk} \delta_{ij} + 2\mu D_{ij}, \quad (33)$$

where c_{ijkl} is the tensor of elastic parameters and λ and μ are the Lamé's parameters. While the moment and source tensors diagonalize in anisotropic media in different systems of eigenvectors and thus their relation is complicated, the eigenvectors of the moment and source tensors are the



Moment Tensors: Decomposition and Visualization, Fig. 9 Diamond source-type plots for the shear-tensile source model in an isotropic medium characterized by various values of the v_p/v_s ratios (the values are indicated in the plot). Red dots, source tensors; black dots, moment

tensors. The dots correspond to the individual sources. The slope angle (i.e., the deviation of the slip vector from the fault) ranges from -90° (pure compressive crack) to $+90^\circ$ (pure tensile crack) in steps of 3° .

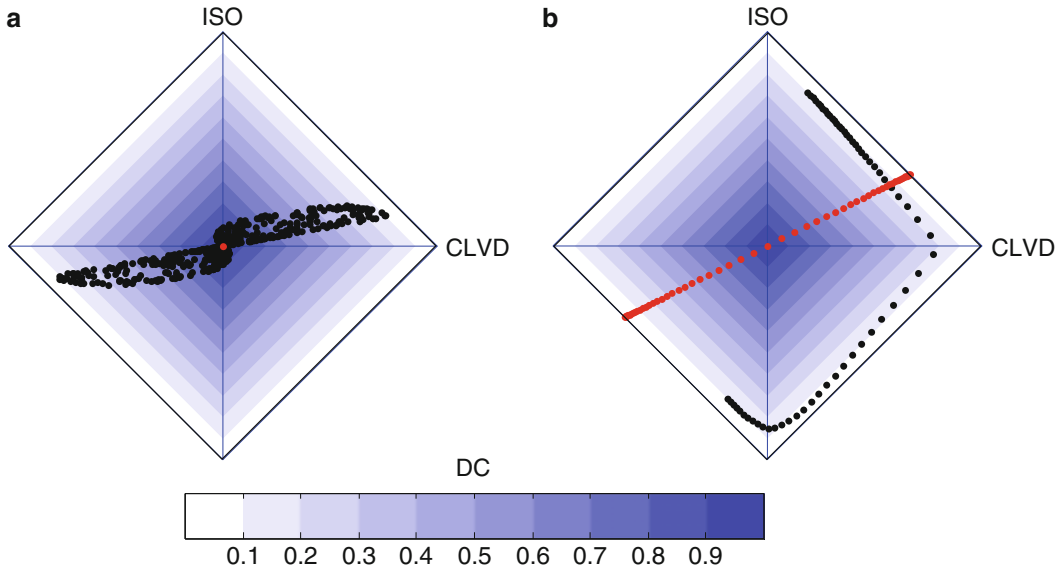
same in isotropic media and their decomposition according to formulas in section “Definition of ISO, CLVD, and DC” yields similar results.

Properties of the moment and source tensor decompositions for shear and tensile sources in isotropic and anisotropic media are illustrated in Figs. 9 and 10. Figure 9 shows the source-type plots for tensile sources with a variable slope angle (i.e., the deviation of the slip vector from the fault) situated in an isotropic medium. The plot shows that the ISO and CLVD components are linearly dependent for both moment and source tensors. For the moment tensors, the line direction depends on the v_p/v_s ratio (Fig. 9a). For the source tensors, the line is independent of the properties of the elastic medium, and the C_{ISO}/C_{CLVD} ratio is always 1/2 (Fig. 9b). The differences between the behavior of the source and moment tensors are even more visible in anisotropic media. Figure 10 indicates that the ISO and CLVD components of the moment tensors of shear faulting (Fig. 10a, black dots) or tensile faulting (Fig. 10b, black dots) may behave in a

complicated way. For example, shear faulting in anisotropic media can produce strongly non-DC moment tensors (Vavryčuk 2005). This prevents a straightforward interpretation of moment tensors in terms of the physical faulting parameters. Therefore, first, the source tensors must be calculated from moment tensors and then interpreted (Fig. 10, red dots). If elastic properties of the medium in the focal zone needed for calculating the source tensors are not known, they can be inverted from the non-DC components of the moment tensors (Vavryčuk 2004, 2011; Vavryčuk et al. 2008). Note that the retrieved medium parameters do not refer to local material properties of the fault, but to the medium surrounding the fault.

Summary

The moment tensor represents equivalent body forces of a seismic source. The forces described by the moment tensor are not the actual forces



Moment Tensors: Decomposition and Visualization, Fig. 10 Diamond source-type plots for the shear (a) and tensile (b) source models in an anisotropic medium. The black dots in (a) correspond to 500 moment tensors of shear sources with randomly oriented fault and slip. The black dots in (b) correspond to moment tensors of tensile sources with strike = 0°, dip = 20°, and rake = -90° (normal faulting). The slope angle ranges from -90° (pure

compressive crack) to +90° (pure tensile crack) in steps of 3°. The red dots in (a) and (b) show the corresponding source tensors. The medium is transversely isotropic with the following elastic parameters (in $10^9 \text{ kg m}^{-1} \text{ s}^{-2}$): $c_{11} = 58.81$, $c_{33} = 27.23$, $c_{44} = 13.23$, $c_{66} = 23.54$, and $c_{13} = 23.64$. The medium density is $2,500 \text{ kg/m}^3$. The parameters are taken from Vernik and Liu (1997) and describe the Bazhenov shale (depth of 12,507 ft)

acting at the source because the moment tensor description assumes elastic behavior of the medium and ignores nonlinear rheology at the focal area. Nevertheless, the moment tensor proved to be a useful quantity and became widely accepted in seismological practice for studying seismic sources. The moment tensor is evaluated for earthquakes on all scales from acoustic emissions to large devastating earthquakes.

In order to understand physical processes at the earthquake source, the moment tensor is commonly decomposed into double-couple (DC), isotropic (ISO), and compensated linear vector dipole (CLVD) components. High percentage of DC indicates a source with shear faulting in isotropic media, and high percentage of ISO indicates an explosive or implosive source. A combination of positive (negative) ISO and CLVD is produced by tensile (compressive) faulting. The type of the source can be visualized using the so-called source-type plots; among

them, the diamond CLVD-ISO plot, the Hudson's skewed diamond plot, and the Riedesel-Jordan lune plot are in common use. In anisotropic media, the physical interpretation of the DC, ISO, and CLVD percentages is not straightforward, and the decomposition of the moment tensor must be substituted by that of the source tensor.

Cross-References

- ▶ [Earthquake Mechanism Description and Inversion](#)
- ▶ [Long-Period Moment-Tensor Inversion: The Global CMT Project](#)
- ▶ [Non-Double-Couple Earthquakes](#)
- ▶ [Reliable Moment Tensor Inversion for Regional- to Local-Distance Earthquakes](#)
- ▶ [Regional Moment Tensor Review: An Example from the European-Mediterranean Region](#)

References

- Aki K, Richards PG (2002) Quantitative seismology. University Science, Sausalito
- Ben-Zion Y (2003) Appendix 2: key formulas in earthquake seismology. In: Lee WHK, Kanamori H, Jennings PC, Kisslinger C (eds) International handbook of earthquake and engineering seismology, Part B. Academic, London, p 1857
- Bowers D, Hudson JA (1999) Defining the scalar moment of a seismic source with a general moment tensor. *Bull Seismol Soc Am* 89(5):1390–1394
- Burridge R, Knopoff L (1964) Body force equivalents for seismic dislocations. *Bull Seismol Soc Am* 54:1875–1888
- Chapman CH, Leaney WS (2012) A new moment-tensor decomposition for seismic events in anisotropic media. *Geophys J Int* 188:343–370
- Dziewonski AM, Chou T-A, Woodhouse JH (1981) Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *J Geophys Res* 86:2825–2852
- Hudson JA, Pearce RG, Rogers RM (1989) Source type plot for inversion of the moment tensor. *J Geophys Res* 94:765–774
- Jost ML, Herrmann RB (1989) A student's guide to and review of moment tensors. *Seismol Res Lett* 60:37–57
- Julian BR, Miller AD, Foulger GR (1998) Non-double-couple earthquakes 1. Theory *Rev Geophys* 36:525–549
- Kanamori H, Given JW (1981) Use of long-period surface waves for rapid determination of earthquake-source parameters. *Phys Earth Planet In* 27:8–31
- Kawasaki I, Tanimoto T (1981) Radiation patterns of body waves due to the seismic dislocation occurring in an anisotropic source medium. *Bull Seismol Soc Am* 71:37–50
- Knopoff L, Randall MJ (1970) The compensated linear-vector dipole: a possible mechanism for deep earthquakes. *J Geophys Res* 75:4957–4963
- Kuge K, Lay T (1994) Data-dependent non-double-couple components of shallow earthquake source mechanisms: effects of waveform inversion instability. *Geophys Res Lett* 21:9–12
- Miller AD, Foulger GR, Julian BR (1998) Non-double-couple earthquakes 2. Observations. *Rev Geophys* 36:551–568
- Mori J, McKee C (1987) Outward-dipping ring-fault structure at Rabaul caldera as shown by earthquake locations. *Science* 235:193–195
- Riedesel MA, Jordan TH (1989) Display and assessment of seismic moment tensors. *Bull Seismol Soc Am* 79:85–100
- Rudajev V, Šílený J (1985) Seismic events with non-shear components. II. Rockbursts with implosive source component. *Pure Appl Geophys* 123:17–25
- Sipkin SA (1986) Interpretation of non-double-couple earthquake mechanisms derived from moment tensor inversion. *J Geophys Res* 91:531–547
- Tape W, Tape C (2012) A geometric comparison of source-type plots for moment tensors. *Geophys J Int* 190:499–510
- Vavryčuk V (2001) Inversion for parameters of tensile earthquakes. *J Geophys Res* 106(B8):16.339–16.355. doi:10.1029/2001JB000372
- Vavryčuk V (2004) Inversion for anisotropy from non-double-couple components of moment tensors. *J Geophys Res* 109:B07306. doi:10.1029/2003JB002926
- Vavryčuk V (2005) Focal mechanisms in anisotropic media. *Geophys J Int* 161:334–346. doi:10.1111/j.1365-246X.2005.02585.x
- Vavryčuk V (2011) Tensile earthquakes: theory, modeling, and inversion. *J Geophys Res* 116:B12320. doi:10.1029/2011JB008770
- Vavryčuk V (2015) Moment tensor decompositions revisited. *J Seismol* 19:231–252. doi:10.1007/s10950-014-9463-y
- Vavryčuk V, Bohnhoff M, Jechumtálová Z, Kolář P, Šílený J (2008) Non-double-couple mechanisms of micro-earthquakes induced during the 2000 injection experiment at the KTB site, Germany: a result of tensile faulting or anisotropy of a rock? *Tectonophysics* 456:74–93. doi:10.1016/j.tecto.2007.08.019
- Vernik L, Liu X (1997) Velocity anisotropy in shales: a petrological study. *Geophysics* 62:521–532
- Wallace TC (1985) A reexamination of the moment tensor solutions of the 1980 Mammoth Lakes earthquakes. *J Geophys Res* 90:11,171–11,176

Noise-Based Seismic Imaging and Monitoring of Volcanoes

Florent Brenguier

Institut des Sciences de la Terre, University of Grenoble, Grenoble, France

Synonyms

Eruption forecasting; Piton de la Fournaise Volcano; Seismic imaging; Seismic noise; Volcano monitoring

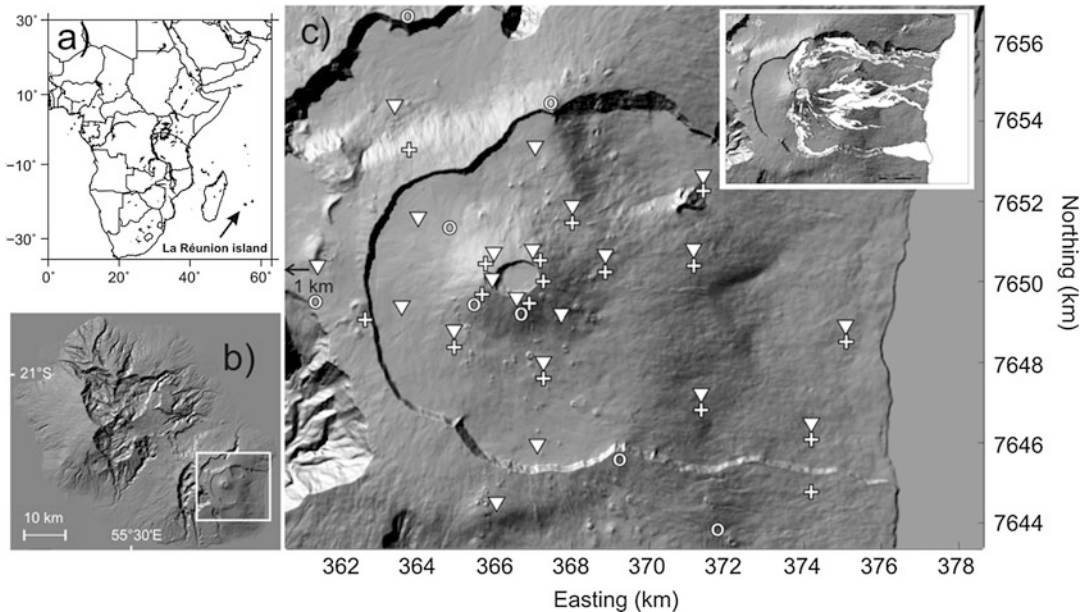
Introduction

Volcanic eruptions represent a great hazard to population leaving nearby volcanoes. In order to better understand volcanic activity and thus to improve eruption forecasting, it is of primary importance to better image and monitor the structure of volcanic edifices.

In the field of ultrasonics, it has been shown both theoretically and experimentally that a random wavefield has correlations which, on average, take the form of the Green's function of the media (Weaver and Lobkis 2001). In seismology, recent studies also showed that the Green's function between pairs of seismographs can be extracted from cross-correlations of coda waves and ambient noise (Campillo 2006). Moreover, some authors used correlations of long ambient

seismic noise sequences from regional networks in order to extract and to invert Rayleigh waves to produce high-resolution seismic images of the shallow crustal layers under these networks (Shapiro et al. 2005). The first part of this entry will show how noise-based seismic imaging can be employed to image volcano interiors.

The early detection of volcanic unrest mainly relies on the monitoring of volcanic seismicity and ground deformation. Those methods provide insights into the dynamics of magma pressurization and transport. However, despite considerable effort, the precise forecasting of eruptions and their intensity has proven to be difficult. Therefore, there is a constant need for novel observational methods to obtain information about ongoing volcanic processes. Pressurized volcanic fluids (magma, water) or gases induce deformation and thus perturbations of the elastic properties of volcanic edifices. These small perturbations can be detected as changes of seismic wave properties using repetitive seismic sources (Ratdomopurbo and Poupinet 1995; Grêt et al. 2005; Wegler et al. 2006). However, none of these approaches were apt to provide a continuous monitoring of volcano elastic properties. In a pioneering work, Sens-Schoenfelder and Wegler (2006) proposed to use the repetitive waveforms of seismic noise cross-correlations to track for subsurface volcanic edifice velocity changes. In this manner, the continuous recording of ambient seismic noise allows continuous monitoring of volcano interiors. The second part of this paper will show how



Noise-Based Seismic Imaging and Monitoring of Volcanoes, Fig. 1 Location of (a) La Réunion island, (b) Piton de la Fournaise Volcano, and (c) the UnderVolc and observatory seismic and GPS networks. UnderVolc broadband seismic stations are shown as inverted

triangles, GPS as crosses, and observatory short-period seismic stations as circles. The *inset* map shows lava flows from 2000 to 2010 (Courtesy of OVPF, T. Staudacher, Z. Servadio)

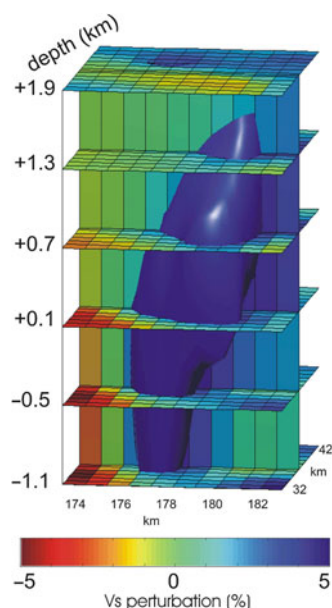
noise-based seismic monitoring can be employed to monitor volcano interiors.

Noise-Based Seismic Imaging of Piton de la Fournaise Volcano

Piton de la Fournaise Volcano (PdF) is a hot spot, shield volcano located on La Réunion island in the Indian Ocean (Fig. 1). It erupted more than 30 times between 2000 and 2010. These eruptions lasted from a few hours to a few months and were associated with the emission of mainly basaltic lava with volume ranging from less than one to tens of million cubic meters (Peltier et al. 2009). The time period that is considered here (1999–2011) started and ended with two major eruptions, namely, the March 1998 eruption (60 million of cubic meter of lava emitted) and the April 2007 eruption associated to the 300 m high collapse of the main Dolomieu crater (130 million of cubic meter of lava emitted) (Staudacher et al. 2009).

Piton de la Fournaise Volcano is monitored by OVPF/IPGP volcano observatory. In 2009, in the framework of UnderVolc project (Brenguier et al. 2012), 15 new broadband seismic stations complemented the observatory seismic network that was composed of about 20 seismic stations. The intense eruptive activity together with a weak tectonic activity makes Piton de la Fournaise Volcano well suited for studies focused on the processes of magma pressurization and injection and for the development of innovative imaging and monitoring methods.

By using continuous records of ambient seismic noise from the PdF Volcano observatory short-period seismic network, Brenguier et al. (2007) were able to measure Rayleigh wave velocity dispersion curves obtained from the cross-correlations of 18 months of ambient seismic noise. The tomography of these measurements as well as the depth inversion led to a high-resolution image of the shallow structure of PdF Volcano (Fig. 2). The results show a high seismic velocity body located in the central part of the



Noise-Based Seismic Imaging and Monitoring of Volcanoes, Fig. 2 3D tomographic image of the shallow central part of Piton de la Fournaise Volcano using cross-correlations of ambient seismic noise (Breguier et al. 2007)

volcano. This body is being interpreted as a preferential zone of magma injection within the edifice. The high velocities are associated with intrusive magma slowly cooling within the surrounding effusive low velocity material. This information is crucial in order to define a more precise structural model for the volcano edifice and thus to better model and then understand the response of the volcano to magmatic activity.

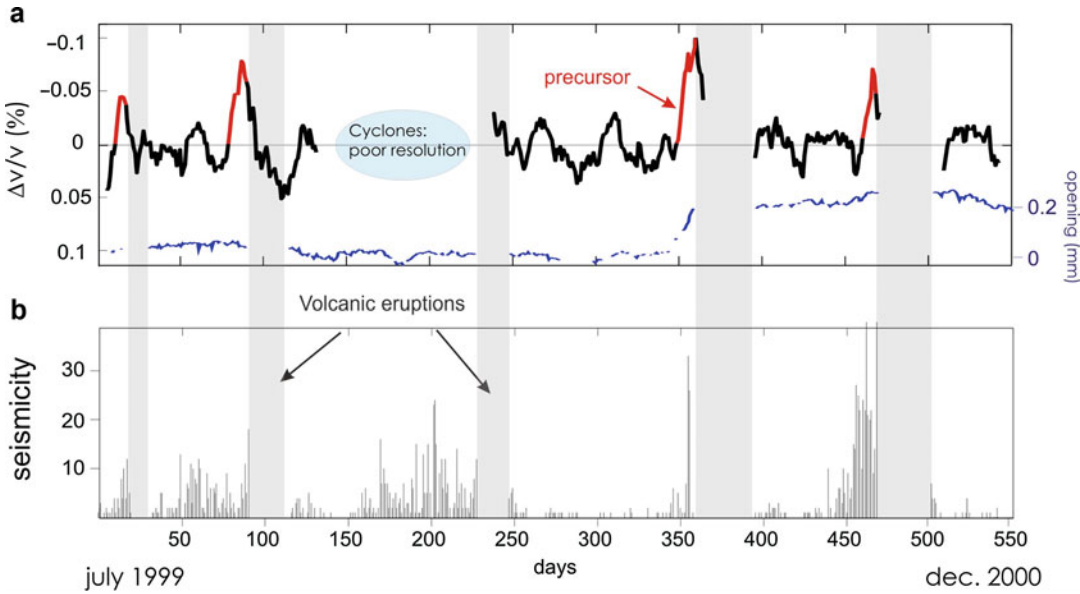
Noise-Based Seismic Monitoring of Piton de la Fournaise Volcano

The approach consists in measuring very small waveform time delays in the coda of noise cross-correlations. This method is described as the so-called Moving Window Cross Spectrum (Ratdompurbo and Poupinet 1995; Clarke et al. 2011) or Coda Wave Interferometry (Snieder et al. 2002) techniques. Coda waves (late part of seismograms) are scattered waves that travel long distances and thus accumulate

time delays as a consequence of a seismic velocity change in the propagating medium. Measuring travel time perturbations in the coda thus allows detecting very small velocity changes that would not be detectable by a classical measure of first arrival time delays. The drawback of that approach is that it is difficult to estimate the travel paths of the scattered waves constituting the coda. However, recent promising results suggest it may be possible to produce refined 3D maps of small changes in a near future (Larose et al. 2010).

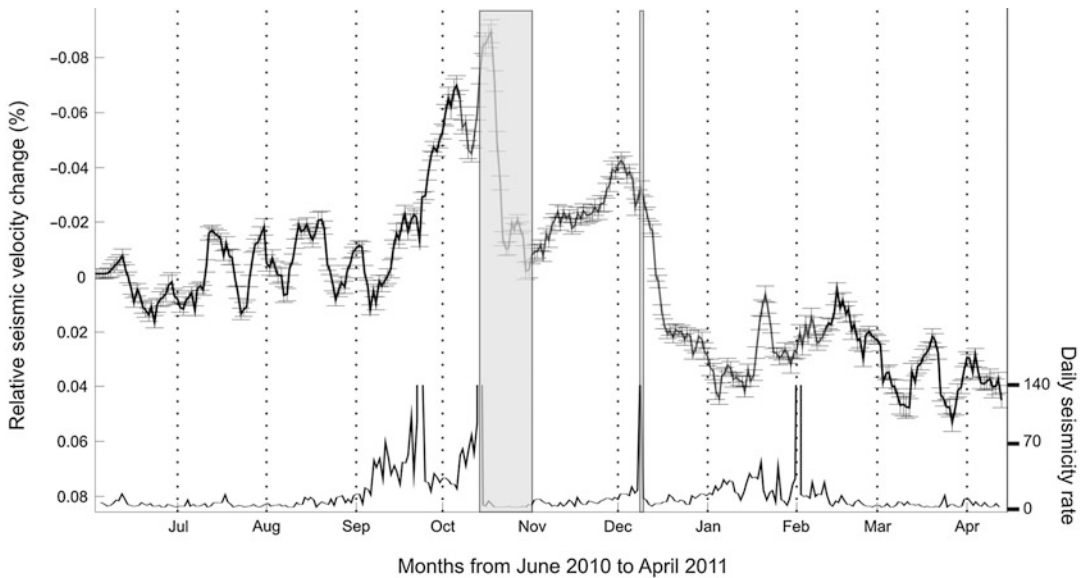
In a previous work Breguier et al. (2012) measured seismic velocity variations within the Piton de la Fournaise edifice from June 2010 to April 2011 following Breguier et al. (2008). Continuous seismic velocity changes as well as the daily seismicity rate are shown on Figs. 3 and 4 for two different time periods at PdF volcano. Figure 3 shows small (0.1 %) seismic velocity drops preceding eruptions at PdF volcano. On Fig. 4, the authors observe a velocity drop starting about 1 month before the 2010 October 14th eruption. The velocity decrease correlates in time with an increase of seismicity and has been interpreted as the deformation of the edifice induced by magma pressure buildup and injection. These velocity drops are interpreted as being associated with the deformation of the edifice induced by magma pressure buildup. Interestingly, seismic velocity increases during the eruption of October 2010 and after the short eruption of December 2010 indicating a possible mechanism of stress relaxation induced by the emptying of the magmatic reservoir. Furthermore, to improve the accuracy and thus time resolution of velocity measurements, Baig et al. (2009) developed a cross-correlation filtering method based on time-frequency transforms and phase coherence filtering.

Noise-based seismic velocity monitoring is thus a unique method that allows to precisely monitor volcanic activity. It must however be mentioned that other phenomena perturb seismic velocities of the rock mass such as the presence of water in the medium, temperature and barometric atmospheric pressure changes, and solid and oceanic tides. It is thus important



Noise-Based Seismic Imaging and Monitoring of Volcanoes, Fig. 3 (a) Relative velocity changes compared to extensometer (FORX) located on PdF volcano. The travel time shifts are measured in the frequency range

0.1–0.9 Hz. For details, see Brenguier et al. (2008) (b) Inter-eruptive seismicity (pre-eruptive swarms are excluded)



Noise-Based Seismic Imaging and Monitoring of Volcanoes, Fig. 4 Relative seismic velocity changes with error bars and daily seismicity rate between June 2010 and April 2011. Eruption periods are shown as gray rectangles

to correctly model these effects in order to correct them from the observed seismic velocity changes in order to extract the volcanic-related signal.

In April 2007, a major eruption occurred at Piton de la Fournaise Volcano ejecting more than 250 million cubic meters of lava and located a few kilometers east of the main active central part of the volcano. Clarke et al. (2013) observed an unusual high seismic velocity drop associated with this volcanic episode that could not be explained by the pre-eruptive edifice inflation due to magma pressure buildup as described by (Brennguier et al. 2008). Also, (Clarke et al. 2013) showed surface displacements images obtained from InSAR data inversion by J-L. Froger (OPGC, France). These results show that between March and May 2007, a widespread volcanic edifice flank movement occurred with a maximum of 1.4 m of eastward displacement. However, the timing of this movement and, in particular, its relation with the occurrence of the April 2007 eruption could not be identified by a lack of temporal resolution of the InSAR images. Clarke et al. (2013) proved using high temporal resolution seismic velocity change measurements that the strong drop of seismic velocities started at the time of a small eruption preceding the main April eruption by a few days. Inferring the association between this high seismic velocity drop and the widespread volcanic flank movement, the authors concluded that the volcanic flank movement started at the time of the small eruption preceding the main April 2007 eruption by a few days. Considering new simulation of volcanic edifice deformation from Got et al. (2013), it is likely that the small eruption preceding the main one triggered a large elastoplastic deformation of the edifice flank that released stresses and favored the horizontal migration of magma to a long distance through the April 2007 eruptive fissure. This study showed how noise-based monitoring has been used to explain the origin of the unusual April 2007 eruption at Piton de la Fournaise Volcano and that this method can also be used to detect volcanic flank movements and possible instabilities on volcanoes.

This method is now used routinely at the Piton de la Fournaise Observatory in order to improve eruption forecasting together with earthquake and deformation observations. Lecocq et al. (2013) developed an integrated package that includes all processing steps required to obtain continuous seismic velocity changes from continuous seismic records. This package is available for free at www.msnoise.org.

Summary

Ambient seismic noise continuously travels along the surface and the interiors of the Earth. These seismic waves thus carry crucial information about the medium they propagate through. When crossing volcanoes, these waves are affected by the mechanical heterogeneities of the volcanic edifice and are also perturbed during unrest periods by temporal changes of volcano interiors. Noise-based seismic imaging and monitoring allows imaging the structure and temporal evolution of volcanic edifices. In particular, it proved success in detecting very small changes of seismic velocities of volcanic edifices (0.1 %) that have been shown, on Piton de la Fournaise Volcano (La Réunion island), to be precursors of volcanic eruptions.

Cross-References

- [Seismic Monitoring of Volcanoes](#)
- [Seismic Noise](#)
- [Seismic Tomography of Volcanoes](#)
- [Tracking Changes in Volcanic Systems with Seismic Interferometry](#)
- [Volcanic Eruptions, Real-Time Forecasting of](#)

References

- Baig A, Campillo M, Brennguier F (2009) Denoising seismic noise cross correlations. *J Geophys Res Solid Earth* (1978–2012), 114(B8)
- Brennguier F, Shapiro N, Campillo M, Nercessian A, Ferrazzini V (2007) 3-D surface wave tomography of

- the Piton de la Fournaise volcano using seismic noise correlations. *Geophys Res Lett* 34(2):2305
- Brenguier F, Shapiro N, Campillo M, Ferrazzini V, Duputel Z, Coutant O, Nercissian A (2008) Towards forecasting volcanic eruptions using seismic noise. *Nat Geosci* 1(2):126–130
- Brenguier F, Kowalski P, Staudacher T, Ferrazzini V, Lauret F, Boissier P, Di Muro A (2012) First results from the UnderVolc high resolution seismic and GPS network deployed on Piton de la Fournaise Volcano. *Seismological Research Letters* 83(1):97–102
- Campillo M (2006) Phase and correlation in random seismic fields and the reconstruction of the green function. *Pure Appl Geophys* 163(2):475–502
- Clarke D, Zaccarelli L, Shapiro NM, Brenguier F (2011) Assessment of resolution and accuracy of the Moving Window Cross Spectral technique for monitoring crustal temporal variations using ambient seismic noise. *Geophys J Int* 186(2):867–882
- Clarke D, Brenguier F, Froger J-L, Shapiro N, Peltier A, Staudacher T (2013) Timing of a large volcanic flank movement at Piton de la Fournaise volcano using noise-based seismic monitoring and ground deformation measurements. *Geophys J Int* 195(2):1132–1140
- Got JL, Peltier A, Staudacher T, Kowalski P, Boissier P (2013) Edifice strength and magma transfer modulation at Piton de la Fournaise volcano. *J Geophys Res Solid Earth* 118:1–18. doi:10.1002/jgrb.50350
- Grêt A, Snieder R, Aster R, Kyle P (2005) Monitoring rapid temporal changes in a volcano with coda wave interferometry. *Geophys Res Lett* 32:1–4
- Larose E, Planes T, Rossetto V, Margerin L (2010) Locating a small change in a multiple scattering environment. *Appl Phys Lett* 96:204101
- Lecocq T, Caudron C, Brenguier F (2013) MSNoise, a Python package for monitoring seismic velocity changes using ambient seismic noise. *Seismol Res Lett* 85(3), 715–726.
- Peltier A, Bachèlery P, Staudacher T (2009) Magma transport and storage at Piton de La Fournaise (La Réunion) between 1972 and 2007: a review of geophysical and geochemical data. *J Volcanol Geotherm Res* 184(1–2):93–108
- Ratdomopurbo A, Poupinet G (1995) Monitoring a temporal change of seismic velocity in a volcano: application to the 1992 eruption of Mt. Merapi (Indonesia). *Geophys Res Lett* 22(7):775–778
- Sens-Schoenfelder C, Wegler U (2006) Passive image interferometry and seasonal variations of seismic velocities at Merapi Volcano. *Indones Geophys Res Lett* 33:1–5
- Shapiro N, Campillo M, Stehly L, Ritzwoller M (2005) High-resolution surface-wave tomography from ambient seismic noise. *Science* 307(5715):1615
- Snieder R, Grêt A, Douma H, Scales J (2002) Coda wave interferometry for estimating nonlinear behavior in seismic velocity. *Science* 295(5563):2253
- Staudacher T, Ferrazzini V, Peltier A, Kowalski P, Boissier P, Catherine P, Lauret F, Massin F (2009) The April 2007 eruption and the Dolomieu crater collapse, two major events at Piton de la Fournaise (La Réunion Island, Indian Ocean). *J Volcanol Geotherm Res* 184(1–2):126–137
- Weaver RL, Lobkis OI (2001) Ultrasonics without a source: thermal fluctuation correlations at MHz frequencies. *Phys Rev Lett* 87(13). doi:10.1103/PhysRevLett.87.134301
- Wegler U, Lühr B, Snieder R, Ratdomopurbo A (2006) Increase of shear wave velocity before the 1998 eruption of Merapi volcano (Indonesia). *Geophys Res Lett* 33:1–4

Non-Double-Couple Earthquakes

G. R. Foulger and Bruce R. Julian
Department of Geological Sciences, Durham
University, Durham, UK

Introduction

Non-double-couple (“non-DC”) earthquake mechanisms differ from what is expected for pure shear faulting in a homogeneous, isotropic, elastic medium. Until the mid-1980s, the DC assumption underlay nearly all seismological analysis, and was highly successful in advancing our understanding of tectonic processes and of seismology in general. In recent years, though, many earthquakes have been found that do not fit the DC model. Earthquakes that depart strongly from DC theory range in size over many orders of magnitude and occur in many environments, but are particularly common in volcanic and geothermal areas. Moreover, minor departures from the DC model are detected increasingly frequently in studies using high-quality data. These observations probably reflect departures from idealized models, caused by effects such as rock anisotropy or fault curvature.

At the same time, it has become clear that industrial activities such as oil and gas production and storage, hydrofracturing, geothermal energy exploitation, CO₂ sequestration, and waste

disposal can induce earthquakes, and that these events often have non-DC mechanisms. Induced earthquakes pose legal hazards because of their destructive potential, but they can also be beneficial, because they can be used to monitoring physical processes that accompany industrial activity.

Non-DC earthquakes are therefore important for improving our understanding of how faults and volcanoes work, perhaps leading some day to an ability to predict earthquakes and eruptions, for avoiding nuisance seismicity in connection with industrial activity, and for monitoring such activity in detail.

Non-DC earthquake mechanisms, by definition, require for their description a more general mathematical formalism than the DC model. The most widely used such formalism is the expansion of the elastodynamic field in terms of the spatial moments of the equivalent-force system (Gilbert 1970). Usually, attention is restricted to second moments, and thus to second-rank moment tensors, and moreover these tensors are usually assumed to be symmetric, so that they have six independent components (A DC has four independent components.). This restriction is not always justified, however. Sources involving net forces or torques are theoretically possible, and phenomena such as landslides and volcanic eruptions provide clear examples of them.

Because it has two extra adjustable parameters, a moment tensor can describe volume changes and general kinds of shear deformation. This increased generality allows the moment tensor to encompass physical processes such as geometrically complex faulting, tensile faulting, faulting in anisotropic media, faulting in heterogeneous media (e.g., near an interface), and polymorphic phase transformations.

The moment tensor also has an important computational property: Mathematical expressions for static and dynamic displacement fields are linear in the moment-tensor components. This property facilitates the evaluation of these fields and enormously simplifies the inverse problem of deducing source mechanisms from observations.

Representation of Earthquake Mechanisms

The Equivalent Force System

Physically, an earthquake involves a nonlinear failure process occurring within a limited region. The equivalent force system, acting in an intact (unfaulted) “model” medium, would produce the same displacement field outside the source region. Any physical source has a unique equivalent force system in a given model medium, but the converse statement is not true. Different physical processes can have identical force systems and therefore identical static and dynamic displacement fields.

The equivalent force system is all that can be deduced from observations of displacement fields, so it constitutes a phenomenological description of the source. There is a one-to-one correspondence between equivalent force systems and elastodynamic fields outside the source region, so we can, in principle at least, determine force systems from observations. On the other hand, the correspondence between force systems and physical source processes is one-to-many, so equivalent force systems (and therefore seismic and geodetic observations) cannot uniquely diagnose physical source processes.

Failure can be regarded as a sudden localized change in the constitutive relation (stress–strain law) in the Earth (Backus 1977). Before an earthquake, the stress field satisfies the equations of equilibrium. At the time of failure, a rapid change in the constitutive relation causes the stress field to change. The resulting disequilibrium causes dynamic motions that radiate elastic waves. We disregard the effect of gravity in the following discussion. In the absence of external forces, the equation of motion is

$$\rho \ddot{u}_i = \sigma_{ij,j}, \quad (1)$$

where $\rho(\mathbf{x})$ is density, $\mathbf{u}(\mathbf{x}, t)$ is the particle displacement vector, $\sigma(\mathbf{x}, t)$ is the physical stress tensor, $\mathbf{x}$ is position, and t is time. Dots indicate differentiation with respect to t , ordinary subscripts indicate Cartesian components of vectors or tensors, the subscript, j indicates

differentiation with respect to the j th Cartesian spatial coordinate x_j , and duplicated indices indicate summation. The true stress is unknown, however, so in theoretical calculations we use the stress $\mathbf{s}(\mathbf{x}, t)$, given by the constitutive law of the model medium (usually Hooke's law). If we replace σ_{ij} by s_{ij} in the equations of motion, though, we must also introduce a correction term, $\mathbf{f}(\mathbf{x}, t)$:

$$\rho \ddot{u}_i = s_{ij,j} + f_i, \quad (2)$$

$$f_i \stackrel{\text{def}}{=} (\sigma_{ij} - s_{ij})_{,j}. \quad (3)$$

This term has the form of a body-force density and is the equivalent force system of the earthquake. It differs from zero only within the source region.

Net Forces and Torques

Nearly all analyses of earthquake source mechanisms explicitly exclude net forces and torques. The equivalent forces given by Eq. 3, which arise from the imbalance between true physical stresses and those in the model, are consistent with these restrictions. The stress glut $\sigma_{ij} - s_{ij}$ is symmetric, so $\mathbf{f}$ exerts no net torque at any point. Furthermore, $\sigma_{ij} - s_{ij}$ vanishes outside the source region, so Gauss's theorem implies that the total force vanishes at each instant.

More complete analysis, however, including the effects of gravitation and mass advection, shows that Eq. 3 is based on overly restrictive assumptions and that net force and torque components are possible for realistic sources within the Earth (Takei and Kumazawa 1994). These forces and torques transfer linear and angular momentum between the source region and the rest of the Earth, with both types of momentum conserved for the entire Earth. An easily understood example is the collapse of a cavity, in which rocks fall from the ceiling to the floor. While the rocks are falling, the Earth outside the cavity experiences a net upward force, relative to the state before and after the event.

The net force component in any source is constrained by the principle of conservation of

momentum; if the source region is at rest before and after the earthquake, the total impulse of the equivalent force (its time integral) must vanish. No such requirement holds for the torque. The total torque exerted by gravitational forces need not vanish even after the earthquake. Horizontal displacement of the center of mass of the source region leads to a gravitational torque, which must be balanced by stresses on the boundary of the source and causes the radiation of elastic waves. Because gravity acts vertically, there can be no net torque about a vertical axis.

The Moment Tensor

We cannot use the equivalent force system $\mathbf{f}(\mathbf{x}, t)$ and the elastodynamic Eq. 2 to determine the displacement field for a hypothetical source. The equivalent force system itself depends on the displacement field that is being sought. Two different approaches are commonly used: (i) In the kinematic approach we assume some mathematically tractable displacement field in the source region (e.g., suddenly imposed slip, constant over a rectangular fault plane), derive the equivalent force system from Eq. 3, and solve Eq. 2 for the resulting displacement field outside the source region; and (ii) in the inverse approach, we use Eq. 2 to determine the force system $\mathbf{f}(\mathbf{x}, t)$ from the observed displacement field and compare the result with force systems predicted theoretically for hypothesized source processes. The most useful way to parameterize the force system in this approach is to use its spatial moments.

The Moment-Tensor Expansion for the Response
Given the equivalent force system $\mathbf{f}(\mathbf{x}, t)$, computing the response of the Earth is a linear problem, and its solution can be expressed as an integral over the source volume V (Aki and Richards 2002, Eq. 3.1, omitting displacement and traction discontinuities for simplicity):

$$u_i(\mathbf{x}, t) = \iiint_V G_{ij}(\mathbf{x}, \boldsymbol{\xi}, t) * f_j(\boldsymbol{\xi}, t) d^3\boldsymbol{\xi}, \quad (4)$$

where $G_{ij}(\mathbf{x}, \boldsymbol{\xi}, t)$ is the Green's function, which gives the i th component of displacement at

position $\mathbf{x}$ and time t caused by an impulsive force in the j direction applied at position ξ and time 0, and the symbol $*$ indicates temporal convolution. If we expand the Green's function in a Taylor series in the source position ξ ,

$$G_{ij}(\mathbf{x}, \xi, t) = G_{ij}(\mathbf{x}, \mathbf{0}, t) + G_{ij,k}(\mathbf{x}, \mathbf{0}, t)\xi_k + \dots, \quad (5)$$

Eq. 4 for the response becomes

$$u_i(\mathbf{x}, t) = G_{ij}(\mathbf{x}, \mathbf{0}, t) * F_j(t) + G_{ij,k}(\mathbf{x}, \mathbf{0}, t) * M_{jk}(t) + \dots, \quad (6)$$

where

$$F_j(t) \stackrel{\text{def}}{=} \iiint_V f_j(\xi, t) d^3\xi \quad (7)$$

is the total force exerted by the source and

$$M_{jk}(t) \stackrel{\text{def}}{=} \iiint_V \xi_k f_j(\xi, t) d^3\xi \quad (8)$$

is the moment tensor. If the equivalent force is derivable from a stress glut via Eq. 3, then the moment tensor is the negative of the volume integral of the stress glut:

$$M_{jk}(t) = - \iiint_V (\sigma_{ij} - s_{ij}) d^3\xi \quad (9)$$

The moment tensor is a second-rank tensor, which describes a superposition of nine elementary force systems, with each component of the tensor giving the strength (moment) of one force system. The diagonal components M_{11} , M_{22} , and M_{33} correspond to linear dipoles that exert no torque, and the off-diagonal elements M_{12} , M_{13} , M_{21} , M_{23} , M_{31} , and M_{32} correspond to force couples. It is usually assumed that the moment tensor is symmetric, ($M_{12} = M_{21}$, $M_{13} = M_{31}$, $M_{23} = M_{32}$), so that the force couples exert no net torque (see above), in which case only six moment-tensor components are independent. In this case, the off-diagonal components correspond to three pairs of force couples, each exerting no net torque ("double couples").

The magnitudes of the six (or nine) elementary force systems (the moment-tensor components) transform according to standard tensor laws under rotations of the coordinate system, so there exist many different combinations of elementary forces that are equivalent. In particular, for a symmetric (six-element) moment tensor one can always choose a coordinate system in which the force system consists of three orthogonal linear dipoles, so that the moment tensor is diagonal. In other words, a general point source can be described by three values (the principal moments) that describe its physics and three values that specify its orientation.

The moment tensor has three important properties that make it useful for representing seismic sources. (1) It makes the "forward problem" of computing theoretical seismic-wave excitation linear. A general source is represented as a weighted sum of elementary force systems, so any seismic wave is just the same weighted sum of the waves excited by the elementary sources. The linearity of the forward problem in turn makes much more tractable the inverse problem of determining source mechanisms from observations. (2) It simplifies the computation of wave excitation. By transforming the moment tensor into an appropriately oriented coordinate system, the angles defining the observation direction can be made to take on special values such as 0 and $\pi/2$. Thus radiation by the elementary sources must be computed not for a general direction, but for only a few directions for which the computation is easier. In a laterally homogeneous medium, for example, radiation must be computed for only a single azimuth. (3) The moment tensor is more general than the DC representation. It includes DCs as special cases, but has two more free parameters than a DC (six vs. four), which enable it to represent sources involving volume changes and more general types of shear.

Higher-Rank Moment Tensors

As Eq. 6 shows, "the" moment tensor described above is only one of an infinite sequence of spatial moments that appear in the expansion of the Earth's response to an earthquake. The later terms involve higher-rank moment tensors,

which contain information about the spatial and temporal distribution of failure in an earthquake, and have great potential value for studying source finiteness and rupture propagation.

Surface Sources (Faults)

A fault is a surface across which there is a discontinuity in displacement. The equivalent force distribution, $\mathbf{f}$, for a generally oriented fault in an elastic medium is, from Eq. 2,

$$f_k(\eta, t) = - \iint_A [u_i(\xi, t)] c_{ijkl} v_j \frac{\partial}{\partial \eta_l} \delta(\eta - \xi) dA, \quad (10)$$

where η is the position where the force is evaluated, ξ is the position of the element of area dA , and the integration extends over the fault surface (Aki and Richards 2002, Eq. 3.5). The unit vector normal to the fault surface is $v(\xi)$ and $[u(\xi, t)]$ is the displacement discontinuity across the fault in the direction of v . The components of the elastic modulus tensor are c_{ijkl} and $\delta(\mathbf{x})$ is the three-dimensional Dirac delta function.

Substituting the force distribution from Eq. 10 into Eq. 8 gives the moment tensor of a general fault,

$$M_{ij} = -c_{ijkl} A \overline{v_k [u_l]}, \quad (11)$$

where A is the total fault area and the overbar indicates the average value over the fault.

Shear Faults

For a planar shear fault (with normal v in the x_3 direction and displacement discontinuity $[u]$ in the x_1 direction, say, so that $v_1 = v_2 = 0$ and $[u_2] = [u_3] = 0$) in a homogeneous isotropic medium ($c_{ijkl} = \lambda \delta_{ij} \delta_{kl} + 2\mu \delta_{ik} \delta_{jl}$), Eq. 8 gives a moment tensor with two nonzero components $M_{13} = M_{31} = \mu A \bar{u}$, where $\bar{u}$ is the average slip. This corresponds to a pair of force couples, one with forces in the x_1 direction and moment arm in the x_3 direction, and the other with these directions interchanged.

We get the same DC moment tensor for a fault with normal in the x_1 direction and displacement

discontinuity in the x_3 direction. This ambiguity between “conjugate” faults is an example of the fundamental limitations on the information that can be deduced from equivalent force systems.

Tensile Faults

For a tensile fault in a homogeneous isotropic medium, with the fault lying in the x_1 - x_2 plane and opening in the x_3 direction, $v_1 = v_2 = 0$ and $[u_1] = [u_2] = 0$. Equation 8 gives a moment tensor with three non-zero components: $M_{11} = M_{22} = \lambda A \bar{u}$ and $M_{33} = (\lambda + 2\mu) A \bar{u}$, corresponding to two dipoles in the fault plane with moments of $\lambda A \bar{u}$, and a third dipole oriented normal to the fault and with a moment of $(\lambda + 2\mu) A \bar{u}$.

Volume Sources

Some possible non-DC source processes, such as polymorphic phase transformation, occur throughout a finite volume rather than on a surface. The equivalent force system often can be expressed in terms of the “stress-free strain” $\Delta_\sigma e_{ij}$, which is the strain that would occur in the source volume if the tractions on its boundary were held constant by externally applied artificial forces (It might more accurately be called the “fixed-stress strain.”). By reasoning that involves a sequence of imaginary cutting, straining, and welding operations (e.g., Aki and Richards 2002, Sect. 3.4), the moment tensor of such a volume source is found to be

$$M_{ij} = \iiint_V c_{ijkl} \Delta_\sigma e_{kl} dV. \quad (12)$$

For a stress-free volume change $\Delta_\sigma V$ in an isotropic medium, for example,

$$M_{ij} = K \Delta_\sigma V \delta_{ij}, \quad (13)$$

where $K = \lambda + (2/3)\mu$ is the bulk modulus.

The stress-free strain is *not*, in general, the strain that actually occurs in a seismic event (Richards and Kim 2005). Because the source is imbedded in the Earth, its deformation is resisted by the stiffness of the surrounding medium, making the true strain changes smaller than the

stress-free values. For example, in terms of the true volume change ΔV , the moment tensor of an isotropic source in an infinite isotropic elastic medium is

$$M_{ij} = (\lambda + 2\mu)\Delta_r V \delta_{ij}, \quad (14)$$

so

$$\Delta_r V = \frac{\lambda + (2/3)\mu}{\lambda + 2\mu} \Delta_\sigma V. \quad (15)$$

The distinction is quantitatively significant; for an infinite homogeneous Poisson solid, $\Delta_r V = (5/9)\Delta_\sigma V$. Furthermore, the stiffness seen by the source, and thus the true strain, depends on the elasticity structure outside the source volume.

Decomposing Moment Tensors

To make a general moment tensor easier to understand, it helps to decompose it into simpler force systems. First, we express the moment tensor in its principal axis coordinate system. Three values (the Euler angles, for example) are needed to specify the orientation of this system, and three other values specify the moments of three orthogonal dipoles oriented parallel to the coordinate axes. Writing these three principal moments as a column vector, we first decompose the moment tensor into an isotropic force system and a deviatoric remainder,

$$\begin{bmatrix} M_1 \\ M_2 \\ M_3 \end{bmatrix} = M^{(V)} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \begin{bmatrix} M'_1 \\ M'_2 \\ M'_3 \end{bmatrix}, \quad (16)$$

with $M^{(V)} \stackrel{\text{def}}{=} (M_1 + M_2 + M_3)/3$, and then decompose the deviatoric part into a DC (principal moments in the ratio 1 : -1 : 0) and a “compensated linear vector dipole” (CLVD), a source with principal moments in the ratio 1 : -1/2 : -1/2 (Knopoff and Randall 1970). Figure 1 illustrates the three elementary force systems used in this decomposition, showing their compressional-wave radiation patterns and the distributions of compressional-wave polarities on the focal sphere (an imaginary sphere surrounding the earthquake hypocenter, to which observations are often referred). Many other

decompositions of the deviatoric part are possible, including many decompositions into two DCs or into two CLVDs. The method of Knopoff and Randall (1970), which makes the largest axis of the CLVD coincide with the corresponding axis of the DC, avoids pathological behavior and is the method most widely used in seismological research:

$$\begin{bmatrix} M'_1 \\ M'_2 \\ M'_3 \end{bmatrix} = M^{(DC)} \begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} + M^{(CLVD)} \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{bmatrix}, \quad (17)$$

with $M^{(DC)} = M'_1 - M'_2$, and $M^{(CLVD)} = -2M'_1$. The moment-tensor elements are arranged so that $|M'_1| \leq |M'_2| \leq |M'_3|$.

The quantity

$$\varepsilon \stackrel{\text{def}}{=} \frac{-M'_1}{|M'_3|} \equiv \frac{1}{2} \frac{M^{(CLVD)}}{|M^{(DC)} + M^{(CLVD)}|} \quad (18)$$

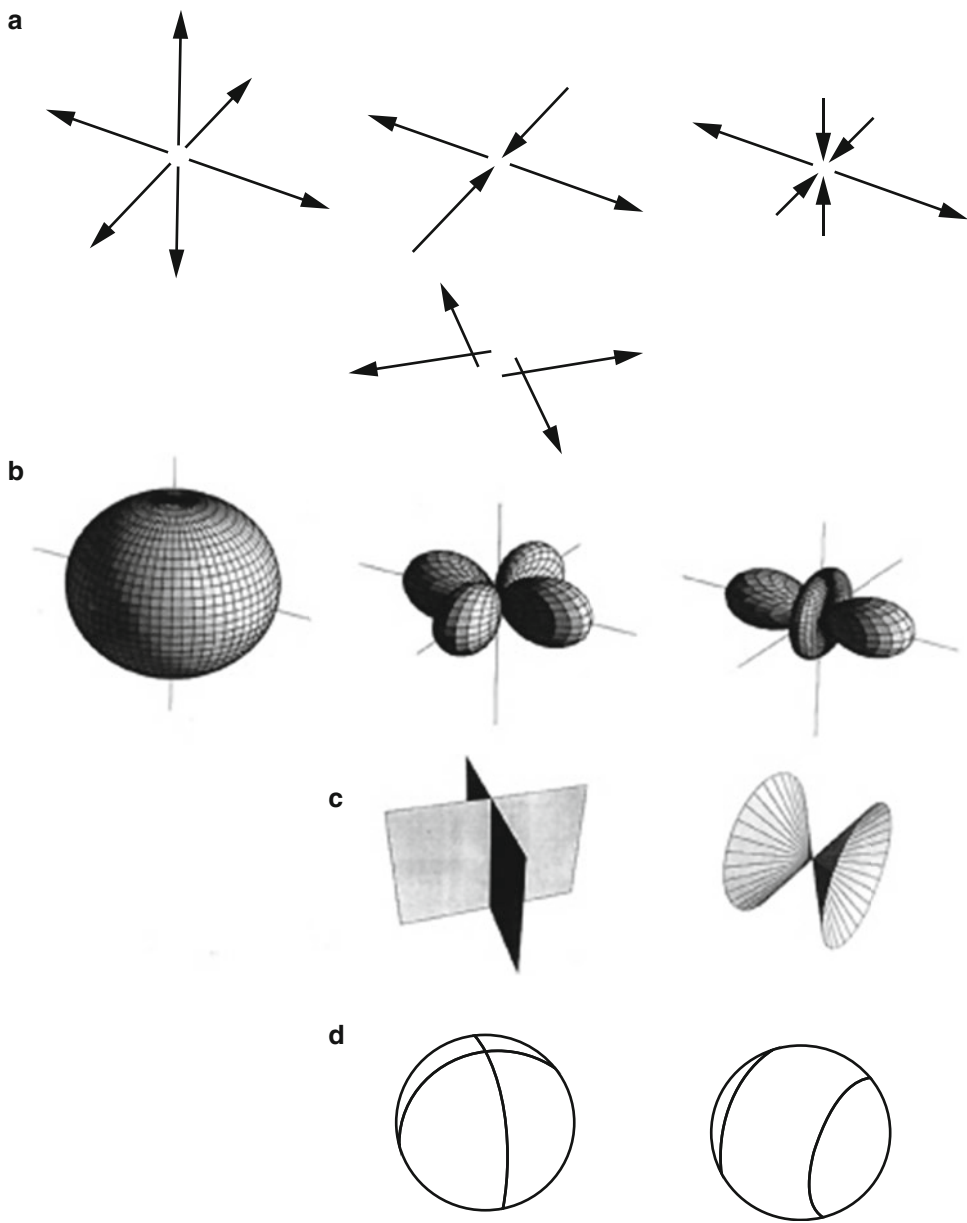
is sometimes used as a measure of the departure of the deviatoric part of a moment tensor from a pure DC. It ranges in value from zero for a pure DC to $\pm 1/2$ for a pure CLVD. Positive ε corresponds to extensional polarity for the major dipole of the CLVD component.

Displaying Focal Mechanisms

Focal-Sphere Polarity Plots

Since the 1960s, common practice has been to display DC mechanisms by showing the compressional-wave polarity fields on maps of a focal hemisphere, and this method is now also widely used for general symmetric moment tensors. Both azimuthal-equal-area (Lambert) and stereographic projections are used, with teleseismic data usually plotted on lower hemispheres and local data on upper hemispheres. Shaded fields usually represent positive polarities (outward first motions). Figures 2, 4, 6, 8, 9, 11, 12, and 13 below contain examples of focal-sphere polarity plots.

A disadvantage of focal-sphere polarity plots is that different mechanisms can lead to

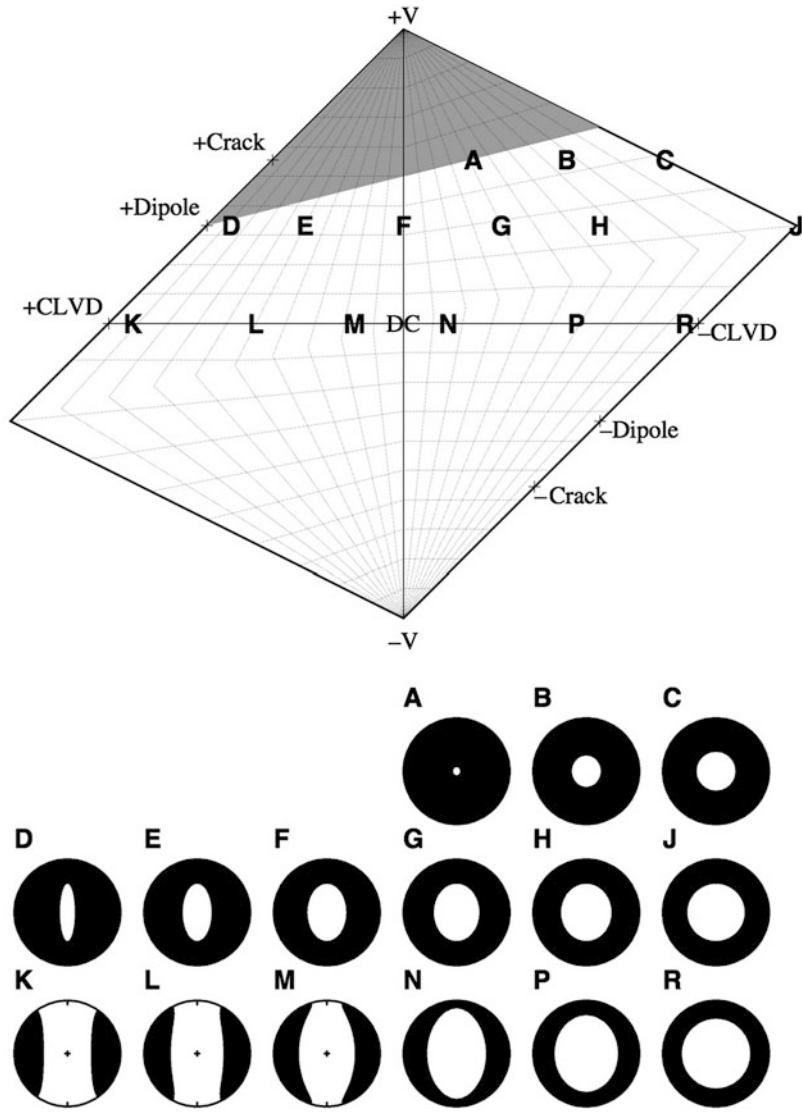


Non-Double-Couple Earthquakes, Fig. 1 Three source types commonly used in decomposing moment tensors: (from left to right) isotropic, double couple (DC), and compensated linear-vector dipole (CLVD). *Top row:* equivalent force systems, in principal-axis

coordinates. For the DC, the force system in a fault-oriented coordinate system is shown underneath. *Second row:* compressional-wave radiation patterns. *Third row:* compressional-wave nodal surfaces. *Fourth row:* curves of intersection of nodal surfaces with the focal sphere

identical plots. Figure 2 shows, for example, that a rather wide range of mechanisms can have positive (or negative) compressional-wave polarities over the entire focal sphere.

Source-Type Plots
Focal-sphere polarity plots display information about source orientation as well as “source type” (the relative values of the principal



Non-Double-Couple Earthquakes, Fig. 2 *Top*: “Source-type plot” of Hudson et al. (1989), which displays earthquake mechanisms (symmetric moment tensors) without regard to their orientations. The quantity k (Eq. 19), which measures volume change, ranges from -1 at the *bottom* of the plot to $+1$ at the *top*, and is constant along the sub-horizontal grid lines. The quantity -2ε (Eq. 18), which describes the deviatoric part of the moment tensor, ranges from -1 on the *left-hand side* of the plot to $+1$ on the *right-hand side*, and is constant along the grid lines that run from *top* to *bottom*. DC: double-couple mechanisms; +Crack: Opening tensile faults;

+Dipole: Force dipoles with forces directed outward; +CLVD: “Compensated linear-vector dipoles,” with dominant dipoles directed outward. –Crack, –Dipole, and –CLVD: The same mechanisms with opposite polarities. *Shaded area*: Region in which all compressional waves have outward polarities. A similar region of inward polarities occurs at the *bottom* of the plot. *A–R*: Some representative mechanisms. *Bottom*: Conventional equal-area focal-hemisphere plots of compressional-wave polarities for 15 representative mechanisms shown on the source-type plot

moments). In fact, conventional DC polarity plots contain information about *only* source orientation. When considering non-DC mechanisms, however, it is useful to display the source type without regard to orientation. The normalized principal moments contain two independent degrees of freedom, and there are many possible ways to display such information in two dimensions. Figure 2 shows the “source-type plot” (Hudson et al. 1989), which gives $T^{\text{def}} = -2\varepsilon$ (Eq. 18) vs.

$$k \stackrel{\text{def}}{=} \frac{M^{(V)}}{|M^{(V)}| + |M'_3|}, \quad (19)$$

a measure of the volume change.

Determining Earthquake Mechanisms

Many types of seismic and geodetic data can be used to determine earthquake focal mechanisms. These range from simple polarities (signs) of observable quantities, through measurements of their amplitudes, to complete time histories of their evolution.

Wave Polarities

First-Motion Methods

P-phase polarities are the most commonly used data in focal-mechanism studies, because they can be determined easily, even using recordings of low dynamic range and accuracy from single-component seismometers. Such polarities alone are of very limited use, however, for studying (or even for identifying) non-DC earthquakes. Even with observations well distributed on the focal sphere, it usually is difficult to rule out DC mechanisms in practice. This difficulty is especially severe if the earthquake lacks a large isotropic component.

In analyzing *P*-wave polarities, seismologists usually constrain earthquake mechanism to be DCs. Finding a DC mechanism amounts to finding two orthogonal nodal planes (great circles on

the focal sphere) that separate the compressional and dilatational polarities into four equal quadrants. Three independent quantities (fault-plane strike and dip angles and the rake angle of the slip vector, say) are needed to specify the orientations of the nodal planes. Nodal planes can be sought either manually, by plotting data on maps of the focal sphere and using graphical methods to find suitable nodal planes, or by using computerized methods, most of which systematically search through the space of possible solutions.

Moment-Tensor Methods

For general moment-tensor mechanisms, hand-fitting mechanisms to observed polarities is impractical. The number of unknown parameters increases from three to five (the six moment tensor components, normalized in some arbitrary way), and furthermore the theoretical nodal surfaces become general ellipses, rather than great circles, on the focal sphere. Searching methods still work, but the addition of two more unknown parameters makes them costly in terms of computer time. Linear programming (sections “[Wave Amplitudes](#)” and “[Amplitude Ratios](#),” below), an analytical method that treats linear inequalities, is well suited to determining moment tensors from observed wave polarities (Julian 1986).

Near-Field Polarities

Near-field observations (made within a few wavelengths of the source) provide a simple and elegant method of detecting an isotropic source component using a single radial-component seismogram (McGarr 1992). The method is not guaranteed to detect isotropic components whenever they exist, but it involves few assumptions, so detected isotropic components are comparatively reliable.

Assume that all the moment-tensor components are proportional to the unit step function, $U(t)$ (any monotonic function of time will work). Place the origin of the coordinate system at the source, with the x_1 axis directed toward the

observer. Then from Eq. 4.29 of Aki and Richards (2002), in an infinite homogeneous medium the radial displacement (the x_1 component observed on the x_1 axis) is

$$u_1(t) = \frac{3}{4\pi\rho r^4} (3M_{11} - Tr\mathbf{M})w(t) \quad (20a)$$

$$+ \frac{1}{4\pi\rho V_P^2 r^2} (4M_{11} - Tr\mathbf{M})U(t - r/V_P) \quad (20b)$$

$$- \frac{1}{4\pi\rho V_S^2 r^2} (3M_{11} - Tr\mathbf{M})U(t - r/V_S) \quad (20c)$$

$$+ \frac{1}{4\pi\rho V_P^3 r} M_{11}\delta(t - r/V_P), \quad (20d)$$

where

$$w(t) \stackrel{\text{def}}{=} \int_{r/V_P}^{r/V_S} \tau U(t - \tau) d\tau \quad (21)$$

vanishes for $t \leq r/V_P$, increases monotonically for $r/V_P \leq t \leq r/V_S$, and is constant for $r/V_S \leq t$. Here $r \equiv x_1$ is the source-observer distance and V_P and V_S are the compressional- and shear-wave speeds. The near-field term Eq. 20a produces a monotonic trend on the seismogram between the compressional and shear waves. For a purely deviatoric source ($Tr \mathbf{M} = 0$), the polarity of this near-field term must be the same as that of the compressional wave Eqs. 20b and 20d. Opposite polarities on any radial seismogram imply that the source mechanism has an isotropic component. Similarly, the near-field shear wave Eq. 20c must have the opposite polarity to the compressional wave, and any observations to the contrary indicate an isotropic source component.

Polarities of Other Seismic Waves

The polarities of shear waves can furnish valuable information to complement compressional-wave first motions, but because shear seismic phases arrive later in the seismogram, where scattered energy (“signal-generated noise”) is also arriving, their polarities often are difficult

to measure reliably. An additional severe difficulty arises for vertically polarized shear (SV) waves incident at the surface beyond the critical angle $\sin^{-1}(V_S/V_P)$ (at epicentral distances between about 0.7 times the focal depth and several thousand kilometers). Outside the “shear-wave windows” at small and large distances, incident shear waves excite evanescent compressional waves that complicate signals and render them practically useless for analysis (Booth and Crampin 1985). Still another problem arises because wave scattering by heterogeneities near the surface excites compressional waves that arrive slightly before the direct shear wave, making the true S onset difficult to identify. These latter two difficulties are caused by conditions at or near the free surface and can be partially alleviated by installing seismometers in deep boreholes.

None of these problems affect horizontally polarized shear waves in spherically symmetric media, and in practice useful *SH* polarities can usually be determined reliably on transverse-component seismograms obtained by numerical rotation.

Wave Amplitudes

The amplitude of a radiated seismic wave contains far more information about the earthquake mechanism than does its polarity alone, so amplitude data can be valuable in studies of non-DC earthquakes. Moreover, because seismic-wave amplitudes are linear functions of the moment-tensor components, determining moment tensors from observed amplitudes is a linear inverse problem, which can be solved by standard methods such as least squares. Conventional least-squares methods, however, cannot invert polarity observations such as first motions, which typically are the most abundant data available. Linear programming methods, which can treat linear inequalities, are well suited to inverting observations that include both amplitudes and polarities (Julian 1986). In this approach, bounds are placed on observed amplitudes, so that they can be expressed as linear inequality constraints. Polarities are already in

the form of linear inequality constraints if the moment-tensor representation is used.

Linear programming methods seek solutions by attempting to minimize the L1 norm (the sum of the absolute values) of the residuals between the constraints and the theoretical predictions.

Amplitude Ratios

Seismic-wave amplitudes are subject to distortion during propagation, particularly because of focusing and de-focusing by structural heterogeneities. A simple way to reduce the effect of this distortion when deriving earthquake mechanisms is to use as data the ratios of amplitudes of waves that have followed similar paths, such as $P:SV$, $P:SH$, or $SH:SV$. If the ratios of the wave speeds is constant in the Earth, then the amplitudes of the waves are affected similarly and the ratio is relatively unaffected.

Using amplitude ratios makes inverting for source mechanisms more difficult, however, because a ratio is a nonlinear function of its denominator. Systematic searching methods still work, but because the dimensionality of the model space is increased by two over that for a DC mechanism, the computational labor is greatly increased (typically by a factor of more than 100).

The efficient linear programming method described above is easily extended to treat amplitude-ratio data in addition to polarities and amplitudes (Julian and Foulger 1996). An observed ratio is expressed as a pair of bounding values, each of which gives a linear inequality that is mathematically equivalent to a polarity observation with a suitably modified Green's function.

Waveforms

A digitized waveform is just a series of amplitude measurements, so waveform inversion may be regarded as an extension of amplitude inversion.

Mathematical Formulation

A theoretical seismogram can be written as a sum of terms, each of which is the temporal convolution of a Green's function and the time function of one source component (Eq. 6).

The source components can include components of the equivalent force and moment-tensor components of any order. If there are n such source components (six, in the usual case of a symmetric second-rank moment tensor), which we arrange in a column vector $\phi(t)$, then a set of m seismograms corresponds to the system of simultaneous equations

$$\begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_m(t) \end{bmatrix} = \begin{bmatrix} A_{11}(t) & \cdots & A_{1n}(t) \\ A_{12}(t) & \cdots & A_{2n}(t) \\ \vdots & \ddots & \vdots \\ A_{m1}(t) & \cdots & A_{mn}(t) \end{bmatrix} * \begin{bmatrix} \phi_1(t) \\ \phi_2(t) \\ \vdots \\ \phi_n(t) \end{bmatrix}, \quad (22)$$

where each matrix element A_{ij} is a Green's function giving the i th seismogram generated by a source whose j th component is the impulse $\delta(t)$, and whose other components are zero. The symbol $*$ indicates temporal convolution. The Green's functions can be thought of as a multichannel filter, which takes the n source time functions $\phi_j(t)$ as inputs and generates as output the m synthetic seismograms $u_i(t)$.

Estimating the source time functions from a set of observed seismograms and assumed Green's functions is thus a multichannel inverse filtering, or deconvolution, problem. It can be solved, for example, by transforming to the frequency domain (so that the convolutions become multiplications) and then applying standard least-squares methods to solve for each spectral component. Alternatively, the output seismograms in Eq. 22 can be concatenated into a single time series, and the Green functions in each column of the matrix $\mathbf{A}$ similarly concatenated. Then, when each time series is expressed explicitly in terms of its samples, the least-squares normal equations turn out to have a "block-Toeplitz" structure, so that they can be solved efficiently by Levinson Recursion (Sipkin 1982).

Limitations

Imperfect Earth Models

If the Earth model used to analyze the radiation from an earthquake differs from the true structure

of the Earth, systematic errors will be introduced into the Green functions and thus into inferred source mechanisms. Near-source anisotropy (section “[Shear Faulting in an Anisotropic Medium](#)”) is one cause of such errors.

Deficient Combinations of Modes

Even if the Green’s functions are correct, it may be impossible to determine the source mechanism completely in some circumstances. For particular types of seismic waves, there may be certain source characteristics that cannot be determined. For example, shear waves alone cannot detect isotropic source components, because purely isotropic sources do not excite shear waves. Similarly, sources with vertical symmetry axes (those whose only nonzero components are $M_{11} = M_{22}$ and M_{33} , with the x_3 axis vertical) excite no horizontally polarized shear (*SH*) or Love waves, and cannot be detected using such waves alone. For any Rayleigh mode, the component M_{33} occurs only in the combination $M_{11} + M_{22} + f(\omega)M_{33}$, where the frequency function $f(\omega)$ depends on the mode and the Earth model. The M_{33} component can be traded off against $M_{11} + M_{22}$ without changing this combination, so isotropic sources are unresolvable by Rayleigh waves of a single frequency and mode (Mendiguren 1977), and even with multimode observations, determining M_{33} requires a priori assumptions about its spectrum. In most studies, enough different modes and/or frequencies are used so that none of these degenerate situations arises. Furthermore, all general inversion methods in widespread use provide objective information about uncertainty and nonuniqueness in derived values, so degeneracies can be detected if they happen to occur.

Shallow Earthquakes

If an earthquake is effectively at the free surface (shallow compared to the seismic wavelengths used), then it is impossible to determine its full moment tensor. Only three moment-tensor components can be determined, and these are not enough even under the a priori assumption of a DC mechanism (which requires four parameters). This degeneracy follows from the

proportionality between the coefficient C_{ij} giving the amplitude of a seismic mode excited by the moment-tensor component M_{ij} and the displacement derivative $u_{i,j}$ for the mode. (This proportionality follows from the principle of reciprocity.) Vanishing of the traction on the free surface,

$$\begin{aligned}\sigma_{13} &= \mu(u_{1,3} + u_{3,1}) = 0 \\ \sigma_{23} &= \mu(u_{2,3} + u_{3,2}) = 0 \\ \sigma_{33} &= \lambda(u_{1,1} + u_{2,2}) + (\lambda + 2\mu)u_{3,3} = 0,\end{aligned}\tag{23}$$

implies that three linear combinations of the excitation coefficients vanish:

$$\begin{aligned}C_{13} &= 0 \\ C_{23} &= 0 \\ \lambda C_{11} + \lambda C_{22} + (\lambda + 2\mu)C_{33} &= 0.\end{aligned}\tag{24}$$

(Here we restrict ourselves to symmetric moment tensors.) Therefore a moment tensor whose only nonzero components are M_{13} and M_{23} radiates no seismic waves (to this order of approximation), and moreover a diagonal moment tensor with elements in the ratio $\lambda : \lambda : (\lambda + 2\mu)$ likewise radiates no seismic waves. The first case corresponds to a vertical dip-slip shear fault or a horizontal shear fault, and the second corresponds to a horizontal tensile fault. This second undeterminable source type has an isotropic component, so isotropic components cannot be determined for shallow sources.

This degeneracy is most often important in studies using surface waves or normal modes, because these are usually observed at frequencies below 0.05 Hz, for which the wavelengths are greater than the focal depths of many earthquakes.

Non-DC Earthquake Processes

Processes Involving Net Forces

Most experimental investigations of earthquake source mechanisms have excluded net forces and torques from consideration a priori. As discussed

above in section “[Net Forces and Torques](#),” the laws of physics do not require such restrictions. Net forces are possible for an internal source because momentum can be transferred between the source region and the rest of the Earth. Momentum conservation does, however, require that the impulse (time integral) of the net force component must vanish if the source is at rest before and after the event.

Landslides

Among sources that involve net forces, landslides have received the most attention. Modeling a landslide as a block of mass M sliding down a ramp gives an equivalent force of $-M\mathbf{a}$, where $\mathbf{a}$ is the acceleration of the block.

The gravitational forces on a landslide also produce a torque of magnitude $Mg\Delta x$, where g is the acceleration of gravity and Δx is the horizontal distance the slide travels.

Volcanic Eruptions

The eruption of material by a volcano applies a net force to the Earth, much as an upward-directed rocket exhaust would. Of course, the total impulse imparted to the Earth-atmosphere system is zero, as with any internal source, but the spatially and temporally concentrated force at the volcanic vent can generate observable seismic waves, whereas the balancing forces transmitted from the ejected material through the atmosphere back to the Earth’s surface excite waves that probably are unobservable in practice. Therefore a volcanic eruption may be modeled as a point force $S\Delta P$, where S is the area of the vent and ΔP is the pressure difference between the source reservoir within the volcano and the atmosphere (Kanamori et al. 1984).

Other processes accompanying volcanic eruptions might act as seismic-wave sources. A change in pressure in a deep spherically symmetric reservoir acts as an isotropic source with a moment tensor given by Eq. 14 (section “[Volume Sources](#)”). For a tabular or crack-shaped reservoir, the force system is the same as that for a tensile fault, discussed in sections “[Tensile Faults](#)” and “[Tensile Faulting](#).”

Unsteady Fluid Flow

If the speed, and thus the momentum, of fluid flowing in a conduit varies with time, a time-varying net force,

$$\mathbf{F} = - \iiint_V \rho \mathbf{a} dV, \quad (25)$$

is exerted on the surrounding rocks, where ρ is the density of the fluid and $\mathbf{a}$ is its acceleration. This process may cause “long-period” volcanic earthquakes and the closely related phenomenon of volcanic tremor. Time variations in the flow speed might be caused by the breaking of barriers to flow, or be self-excited by nonlinear interaction between the flowing fluid and deformable channel walls (Julian 1994).

Complex Shear Faulting

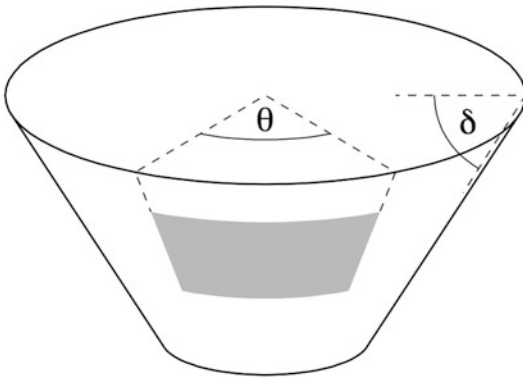
Multiple Shear Events

If earthquakes occur close together in space and time, observed seismic waves may not be able to resolve them, and they may be misinterpreted as a single event. The apparent moment tensor of the composite event is then the sum of the true moment tensors of the earthquakes, and because the sum of two DCs is not, in general, a DC, shear faulting can produce non-DC mechanisms in this way. Combining DCs cannot ever produce mechanisms with isotropic (volume change) components, because the trace of the moment tensor is a linear function of its components, so multiple shear-faulting mechanisms lie on the horizontal axis of the source-type plot.

By analyzing complete seismic waveforms, it is often possible to resolve a multiple event into subevents with different mechanisms. Doing this requires use of algorithms that allow the moment tensor to vary with time in a general way. Algorithms that assume that all the moment-tensor components have identical time functions can produce spurious non-DC mechanisms, even if the subevents have identical DC mechanisms.

Volcanic Ring Faults

Dikes intruded along conical surfaces with both outward and inward dips are often found in



Non-Double-Couple Earthquakes, Fig. 3 Geometry of volcanic ring faulting used in computing mechanisms shown in Fig. 4. Dip-slip motion on fault of dip δ is uniformly distributed over azimuth range θ

exhumed extinct volcanoes, and are expected consequences of the stresses caused by inflation and deflation of magma chambers (Anderson 1936). These dikes are of two types: “cone sheets,” which dip inward at about $30\text{--}70^\circ$, form as tensile faults during inflation, and nearly vertical or steeply outward-dipping “ring dikes” form through shear failure accommodating subsidence following deflation or eruption of magma. For both types, the axes of the cones are approximately vertical. If dip-slip shear faulting occurs on a conical fault, and the rupture in an earthquake spans a significant azimuth range (Fig. 3), the resulting mechanism, considered as a point source, can have a non-DC component (Ekström 1994). (Strike-slip motion on such a surface always gives pure DC mechanisms.) Figure 4 shows a suite of theoretical source mechanisms corresponding to dip-slip ruptures spanning various azimuth ranges on conical faults of different dips. For steeply dipping faults the non-DC components are small (for vertical faults they vanish), so cone sheets are more efficient than ring dikes as a non-DC process.

Tensile Faulting

Opening Tensile Faults in the Earth

An obvious candidate mechanism for geothermal and volcanic earthquakes is tensile faulting, in which the displacement discontinuity is normal,

rather than parallel, to a fault surface. The equivalent force system of a tensile fault consists of three orthogonal linear dipoles with moments in the ratio $(\lambda + 2\mu):\lambda:\lambda$. It is equivalent to an isotropic source of moment $(\lambda + 2\mu/3)A\bar{u}$ plus a CLVD of moment $(4\mu/3)A\bar{u}$ (section “[Tensile Faults](#)”). The far-field compressional waves have all first motions outward, with amplitudes largest (by a factor of $1 + 2\mu/\lambda$) in the direction normal to the fault. Figure 2 shows the position of tensile faults on a source-type plot.

Compressive stress tends to prevent voids from forming at depth in the Earth, but high fluid pressure can overcome this effect and allow tensile failure to occur. The situation is conveniently analyzed using Mohr’s circle diagrams (Fig. 5). The effect of interstitial fluid at pressure p in a polycrystalline medium such as a rock is to lower the effective principal stresses by the amount p . Thus fluid pressure, if high enough, can cancel out much of the compressive stress caused by the overburden. Fluid pressures comparable to the lithostatic load are found surprisingly often in deep boreholes.

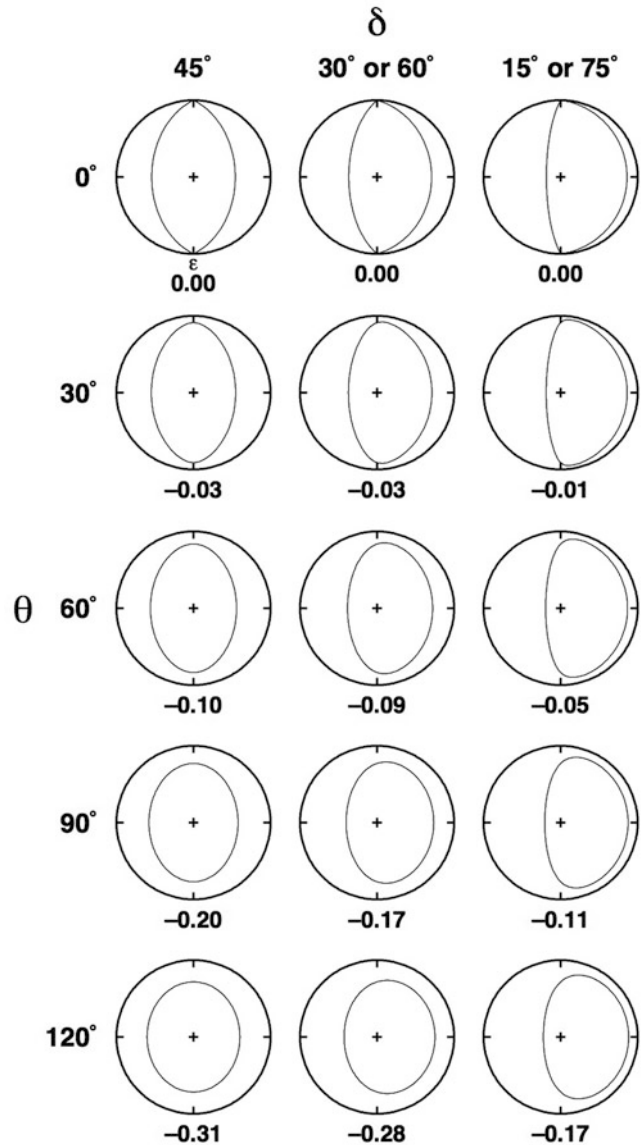
A second prerequisite for tensile failure is that the shear stresses be small, or equivalently that the principal stresses be nearly equal. The diameter of the Mohr’s circle in Fig. 5 is equal to the maximum shear stress (difference between the extreme principal stresses). If this diameter is too large, the circle can touch the failure envelope only along its straight portion, which corresponds to shear failure. Only if the shear stress, and thus the diameter, is small will the circle first touch the failure envelope in the tensile field to the left of the τ axis.

Combined Tensile and Shear Faulting

Although tensile faults could cause earthquakes that involve volume increases, they cannot explain non-DC earthquakes whose isotropic components indicate volume decreases. Tensile faults can open suddenly for a variety of reasons, but they would be expected to close gradually, and not to radiate elastic waves. If a tensile fault and a shear fault intersect, however, then stick-slip instability could cause sudden episodes of either opening or closing, with volume increases or decreases. The stresses around the

Non-Double-Couple Earthquakes,

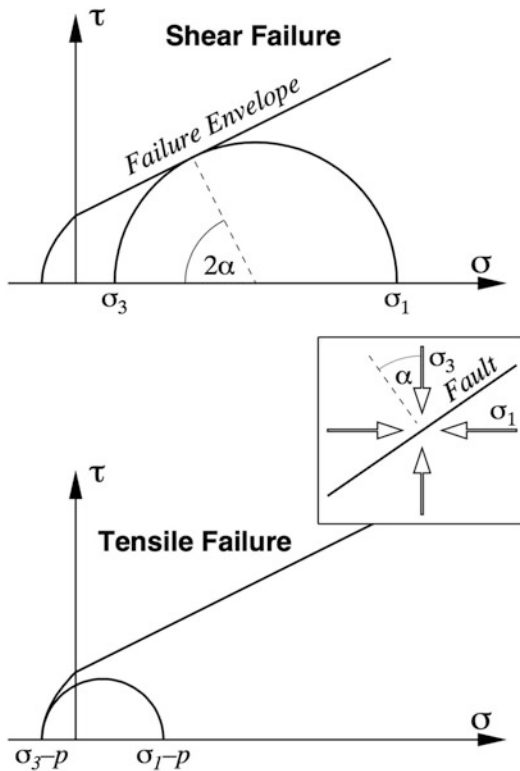
Fig. 4 Non-DC mechanisms for volcanic ring faulting (Fig. 3). Theoretical compressional-wave nodal surfaces for an arcuate dip-slip fault whose strike spans a range θ and averages north-south. Each focal sphere corresponds to two situations: a fault dipping to the west by the smaller angle δ given or to the east by the larger angle. All mechanisms are purely deviatoric. Numbers below each mechanism give values of ε (Eq. 18), which describe the deviatoric parts of the moment tensors. Upper focal hemispheres are shown in equal-area projection. Lower hemisphere plots are *left-right* mirror images. For normal faulting, central fields have dilatational polarity



tips of both shear and tensile faults favor this type of fault pairing, and faulting of this kind occurs in rocks in the laboratory (Brace et al. 1966). A similar situation can occur in the case of shear faulting near mines, with the tunnel playing the role of the tensile fault.

Figure 6 shows the theoretical source mechanisms for combined tensile and shear faults of different geometries and relative seismic moments. When the tensile fault opens or closes in the direction normal to its face, the

mechanisms have large isotropic components, with most of the focal sphere having the same polarity as the tensile fault and two unconnected fields having the opposite polarity. The symmetry of the moment tensors makes it impossible to determine the angle between the two faults. An angle of $45^\circ - \alpha$ is equivalent to an angle of $45^\circ + \alpha$. When the tensile fault opens or closes obliquely, in the direction parallel to the shear fault, then the mechanisms are closer to DCs, and less sensitive to the relative seismic moments.



Non-Double-Couple Earthquakes, Fig. 5 Conditions for shear and tensile failure. Mohr's circle diagram shows the relationship between shear traction τ and normal traction σ across a plane in a stressed medium. Locus of (σ, τ) points for different orientations of the plane is a circle of diameter $\sigma_1 - \sigma_3$, centered at $((\sigma_1 + \sigma_3)/2)$, where σ_1 and σ_3 are the extreme principal stresses. Failure occurs when circle touches the "failure envelope," and the point of tangency determines the orientation of the resulting fault (see *inset*). (Theoretical failure envelope shown corresponds to Griffith theory of failure, as modified by F. A. McClintock and J. B. Walsh (Price 1966).) Straight portion of failure envelope in compressional field ($\sigma > 0$) represents the Navier-Coulomb criterion for shear failure. (Top) At high confining stress with no fluid pressure, only shear failure occurs. (Bottom) High fluid pressure lowers the effective confining stress, and tensile failure occurs for low stress differences

If the dominant principal axis of the tensile fault lies in the same plane as the P and T axes of the shear fault (i.e., if the null axis of the shear fault lies in the tensile-fault plane), then the composite mechanism lies on the line between the DC and Crack positions on a focal-sphere plot. For more general (and less physically plausible) geometrical arrangements,

the composite mechanism lies within a region consisting of two triangles (Fig. 7).

Shear Faulting in an Anisotropic Medium

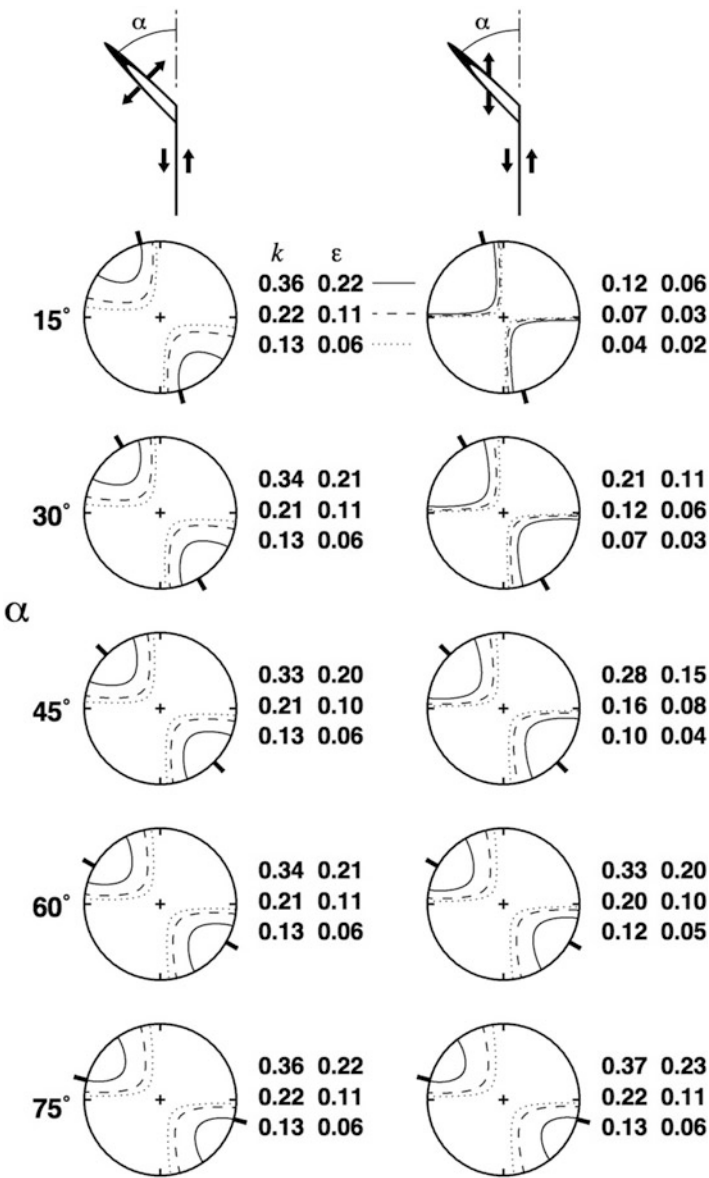
The equivalent force system of an earthquake depends on the constitutive law used to compute the model stress s_{ij} in Eq. 2. This means that a fault in an anisotropic elastic medium has a different equivalent force system than it would if the medium were isotropic, and in particular that a shear fault in an anisotropic medium generally has a non-DC moment tensor, which can be determined, for example, from Eq. 11 (Vavryčuk 2005). Most rocks are anisotropic, because of layering on a scale smaller than seismic wavelengths, preferential orientation of crystals, and the presence of cracks and inclusions, so most earthquakes are affected by anisotropy.

Elastic wave propagation in an anisotropic medium is more complicated in several ways than in an isotropic medium. The particle motion in body waves is no longer either longitudinal or transverse to the direction of propagation, but is generally oblique, so body-wave modes are referred to as "quasi-compressional" and "quasi-shear." The "direction of propagation," in fact is no longer a single direction, but rather two directions for each mode: the normal to the wavefront (the "phase velocity" direction) and the direction of energy transport (the "group velocity" direction, from the source to the observer).

Green's functions for the anisotropic medium must be used to compute the radiated seismic waves for the force system given by Eq. 11 (Gajewski 1993) and to solve the inverse problem of determining the force system from observed seismograms. If enough information is available about source-region anisotropy to determine such Green's functions, then it will be possible to recognize when non-DC force systems are consistent with shear faulting. In practice, however, information about anisotropy in earthquake focal regions is seldom available, so Green's functions appropriate for isotropic constitutive laws are used instead. In this case, the non-DC force system of most interest is not the one given by Eq. 11, but rather the one that would be derived

Non-Double-Couple Earthquakes,

Fig. 6 Non-DC mechanisms for combined tensile and shear faulting with different geometries. Both faults are vertical, with the shear fault striking north-south and the tensile fault striking west of north at the angle α , indicated by bold ticks. *Left column:* tensile fault opening normal to its face. *Right column:* tensile fault opening obliquely, parallel to the shear fault. Compressional-wave nodal surfaces are shown for different relative moments of the tensile (M^T) and shear (M^S) components. *Solid curves:* $M^T = 0.5M^S$. *Dashed curves:* $M^T = 0.2M^S$. *Dotted curves:* $M^T = 0.1M^S$. Numbers to the right of each plot give values of k and ϵ (Eqs. 18 and 19), for each mechanism. Focal hemispheres (either upper or lower, because of symmetry) are shown in equal-area projection



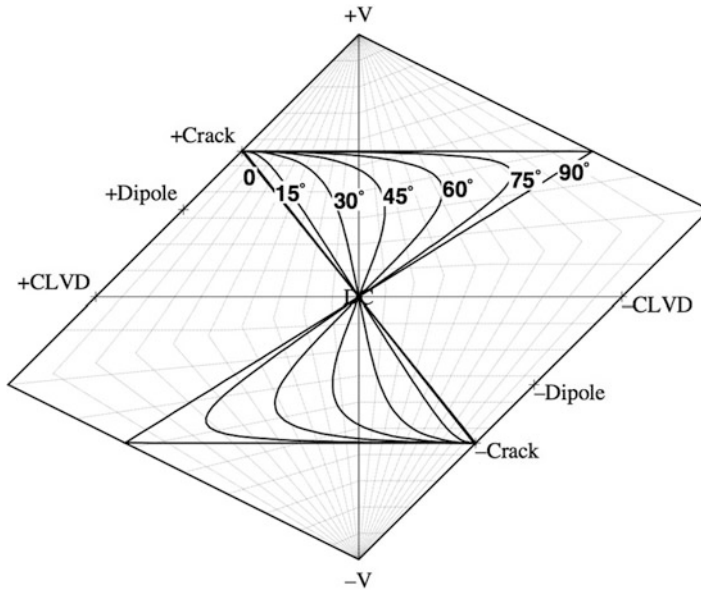
from seismic waves using an isotropic constitutive law when the focal region is actually anisotropic.

Shear Faulting in a Heterogeneous Medium

If an earthquake occurs in a place where the elastic moduli vary spatially, its apparent mechanism will be distorted, and a DC earthquake may appear to have non-DC components. This occurs when the spatial derivatives of Green's functions (strains) appearing in the second term on the right side of

the moment expansion Eq. 6 vary significantly over the source region, so that the values at $\xi = 0$ are inappropriate for a portion of the moment release. In effect, neglected higher moments are contaminating estimates of the lower moments.

This effect is *not* a consequence of using an incorrect Earth model to compute the Green function. In this discussion, we assume that the Earth model and Green's function are exact. The distortion of the focal mechanism is caused by the finiteness of the source region. Errors in the



Non-Double-Couple Earthquakes, Fig. 7 Source types for combined tensile and shear faulting. Numbers give angles between the tensile-fault planes and the intermediate principal (null) axes of the shear faults. Small angles are physically most plausible (For the mechanisms shown in Fig. 6 this angle is zero.). For all possible relative

orientations and moments, the source type lies between the corresponding curve and the *straight line* from +Crack to -Crack. The upper half of the plot corresponds to opening faults and the lower half to closing faults. For an explanation of the plotting method, see Fig. 2

Green function due to our incomplete knowledge of Earth structure and to the mathematical complexity of elastodynamics can cause severe errors in derived earthquake mechanisms, but that is a different phenomenon.

Consider an earthquake near an interface across which the elastic moduli change discontinuously (Woodhouse 1981), so that the truncated Taylor series Eq. 5 is a particularly poor approximation. Then the inferred mechanism of an earthquake that is assumed to occur on one side of the interface will be distorted if the earthquake actually is on the other side. If the source region includes both sides of the interface, it is unavoidable that a portion of the moment release will be distorted in this manner.

If we arrange the independent components of the moment tensor in a column vector,

$$\mathbf{m}^{\text{def}} [M_{11} \ M_{12} \ M_{22} \ M_{13} \ M_{23} \ M_{33}]^T, \quad (26)$$

then the seismic waves excited can be written $\mathbf{g}^T \mathbf{m}$, where $\mathbf{g}$ is a column vector whose

components are spatial derivatives of Green's functions appearing in the second term on the right side of Eq. 6. Because displacement and stress are continuous at the interface, the elements of $\mathbf{g}$ on one side of the interface are related to those on the other side by a relation that can be written

$$\mathbf{A}^+ \mathbf{g}^+ = \mathbf{A}^- \mathbf{g}^-, \quad (27)$$

where $\mathbf{A}$ is a matrix that depends on the orientation of the interface and the elastic moduli adjacent to it, and the superscripts $+$ and $-$ indicate values on the two sides of the interface. It follows that

$$\mathbf{g}^{+T} \mathbf{m} = \mathbf{g}^{-T} [\mathbf{A}^{+-1} \mathbf{A}^-]^T \mathbf{m}, \quad (28)$$

or in other words that an earthquake with moment tensor $\mathbf{m}$ occurring on the $+$ side of the interface excites the same waves as an earthquake with moment tensor $[\mathbf{A}^{+-1} \mathbf{A}^-]^T \mathbf{m}$ occurring on the opposite ($-$) side. In a coordinate

system with the x_3 axis normal to the interface, the matrix connecting the true and apparent moment tensors is

$$[\mathbf{A}^{+-1}\mathbf{A}^-]^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & R_1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & R_2 & 0 & 0 \\ 0 & 0 & 0 & 0 & R_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & R_3 \end{bmatrix}, \quad (29)$$

where

$$R_1 \stackrel{\text{def}}{=} \frac{\lambda^- - \lambda^+}{\lambda^+ + 2\mu^+}, \quad (30)$$

$$R_2 \stackrel{\text{def}}{=} \frac{\mu^-}{\mu^+}, \quad (31)$$

and

$$R_3 \stackrel{\text{def}}{=} \frac{\lambda^- + 2\mu^-}{\lambda^+ + 2\mu^+}. \quad (32)$$

Figure 8 shows the distortion of the apparent mechanisms of DCs of various orientations occurring adjacent to a horizontal interface across which the elastic-wave speeds change by 20 %. If the fault plane is parallel to the interface (or if the interface *is* the fault plane), then shear faulting does not lead to apparent non-DC mechanisms, although the scalar seismic moment is distorted. This case is not illustrated in the figure, but is clear from the structure of the matrix in Eq. 29. Only M_{13} and M_{23} are nonzero and matrix multiplication merely multiplies these elements by R_2 , producing a DC of the same orientation, but with its moment multiplied by the contrast in rigidity modulus. If the fault is perpendicular to the interface, then the apparent mechanism is still a DC for all slip directions, but its orientation and seismic moment are distorted, as the first column of the figure illustrates. For general fault orientations, the apparent mechanism has artificial isotropic and CLVD components.

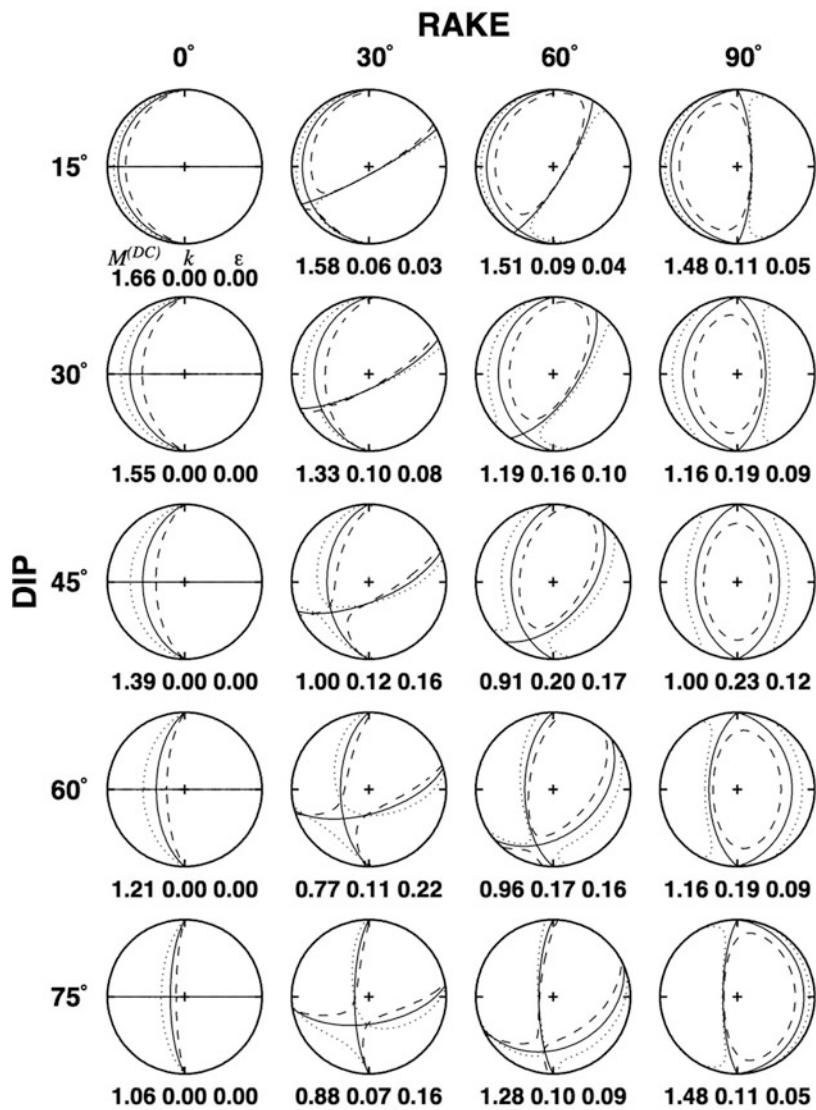
Rapid Polymorphic Phase Changes

Except in the shallow crust, compressional stresses in the Earth greatly exceed shear stresses. Therefore, earthquake processes that involve even relatively small volume changes could release large amounts of energy. For this reason, it has long been speculated that polymorphic phase transformations in minerals might cause deep earthquakes. Such speculation has been stimulated also by consideration of the simple theory of frictional sliding, which seems to require impossibly large shear tractions when the confining pressure is high, and by the theory of plate tectonics, which involves large-scale vertical motions in the upper mantle. Many common minerals undergo polymorphic changes in crystal structure in response to changes in pressure and temperature, and some of the major structural features in the Earth, most notably the “transition region” at depths between about 400 and 800 km in the upper mantle, are attributed to such phase changes (in this case, involving the mineral olivine, $(\text{Fe, Mg})_2\text{SiO}_4$, transforming to the spinel and then perovskite crystal structures).

As slabs of lithosphere subduct into the mantle, olivine and other minerals are carried out of their stability fields and into the stability fields of denser phases, into which they transform. If these changes occur rapidly enough to radiate seismic waves, they constitute earthquakes, and their mechanisms will have isotropic components. They probably will also have deviatoric components, because the process of phase transformation will release shear strain, much as explosions are often observed to release tectonic shear strain. There seems to be no reason, however, why such a deviatoric component should be a DC rather than a CLVD, and in fact the CLVD force system was first invented as a possible mechanism for deep earthquakes caused by phase transformations (Knopoff and Randall 1970).

Observations

Non-DC earthquakes are observed in many environments, including volcanic and geothermal areas, mines, and deep subduction zones.



Non-Double-Couple Earthquakes, Fig. 8 Apparent non-DC mechanisms caused by shear faulting with unit moment near a horizontal interface, for various dip and rake angles. *Solid curves*: compressional-wave nodal planes corresponding to true (DC) mechanisms. *Dotted curves*: nodal surfaces for apparent mechanisms obtained if DC moment release on the low-speed side of interface is assumed to be on the high-speed side. *Dashed curves*: same, with sides interchanged. Both media are Poisson solids, and the ratio of elastic moduli across the interface

is 1.7:1, corresponding to a wave-speed contrast of about 20 % if density is proportional to wave speed. Numbers below each mechanism give the DC moment and the values of k and ϵ (Eqs. 18 and 19) corresponding to the *dotted curves*. Dips of 0 are not shown because the apparent mechanisms are DCs (although with moments distorted by the ratio of the rigidity moduli). Dips of 90° correspond to cases with rake of 0 (first column). Focal hemispheres (which may be considered either upper or lower) are shown in equal-area projection

They include both natural earthquakes and events induced by human activity. An in-depth review of many observed examples is given by Miller et al. (1998b).

Landslides and Volcanic Eruptions

Landslides and volcanic eruptions have equivalent force systems that include net forces (section “Landslides”). An example is the Mount

St. Helens, Washington, eruption of May 18, 1980, which began with a landslide with a mass of about 5×10^{13} kg that traveled about 10 km from the volcano. The seismic observations are well explained by two forces, a near-horizontal, southward directed force representing the landslide, and a vertical force representing the eruption (Kanamori et al. 1984).

The gravitational forces on a landslide also exert a torque, which can be substantial, on the source volume (section “[Landslides](#)”). The dimensions of the landslide that accompanied the May 18, 1980, eruption of Mt. St. Helens, for example, correspond to a net torque of about 5×10^{18} N m. By comparison, the two largest earthquakes accompanying the eruption had surface-wave magnitudes of about 5.3, which correspond to seismic moments of about 2.6×10^{17} N m. Apparently, no seismological analyses of landslides to date have included torques in the source mechanism.

Long-Period Volcanic Earthquakes

Long-period volcanic earthquakes have dominant seismic frequencies an order of magnitude lower than those of most earthquakes with comparable magnitudes. They are caused by unsteady flow of underground magmatic fluids and are expected to have mechanisms involving net forces (section “[Unsteady Fluid Flow](#)”). An example is a 33-km-deep, long-period earthquake that occurred a year before the 1986 eruption of Izu Ōshima volcano, Japan. The seismograms are characterized by nearly monochromatic shear-wave trains that have dominant frequencies of about 1 Hz and last for more than a minute. The shear-wave polarization directions are consistent with a net force mechanism.

Several types of non-DC earthquakes occur at the intensively monitored active Sakurajima volcano in southern Kyushu, Japan. It is a rich source of low-frequency earthquakes and explosion earthquakes, which accompany crater eruptions that radiate spectacular visible shock waves into the atmosphere. The moment tensors of the explosion earthquakes have been explained as either deflation of cracks or other cavities that might rapidly expel gas into the

atmosphere (Uhira and Takeo 1994), or as sources dominated by vertical dipoles involving, for example, the opening of horizontal cracks (Iguchi 1994). The discrepancies between the results of these two studies reflect the difficulty of distinguishing between vertical forces and vertical dipoles.

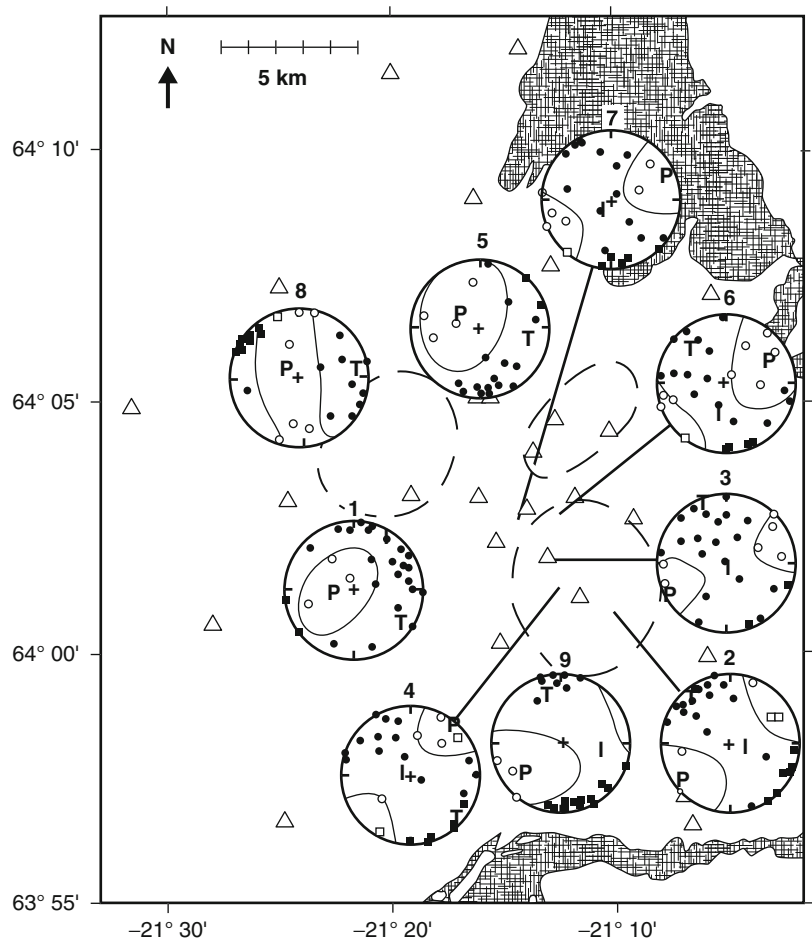
Short-Period Volcanic and Geothermal Earthquakes

Observations from dense local seismic networks show that earthquakes in volcanic areas commonly have non-DC mechanisms. In many cases, the data have been subjected to careful analysis, including waveform modeling (e.g., Dreger et al. 2000; Julian and Sipkin 1985), corrections for wave propagation through three-dimensional crustal structure determined from tomography (e.g., Miller et al. 1998a; Ross et al. 1996), and the use of multiple seismic phases. The clearest cases are from Iceland, California, and Japan.

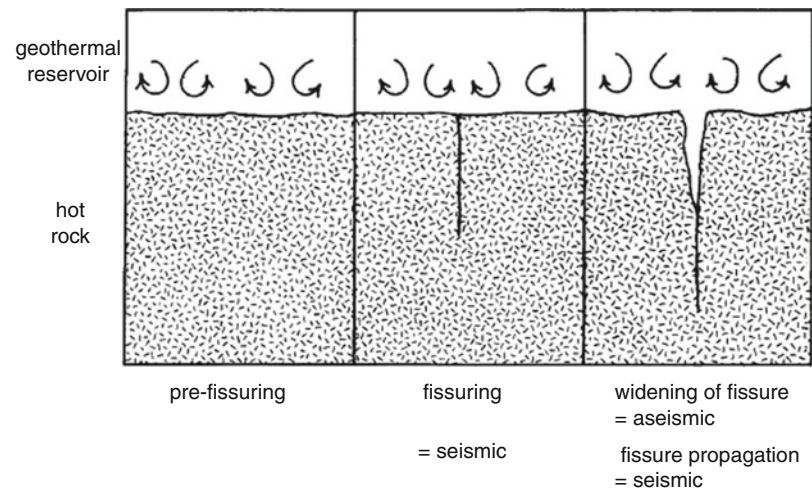
Three volcano-geothermal areas on the spreading plate boundary in Iceland are rich sources of short-period, non-DC earthquakes. These are the Reykjanes, Hengill-Greisdalur, and Krafla volcanic-geothermal systems. At all three areas, small earthquakes occur that have unequal dilatational and compressional areas on the focal sphere, suggestive of isotropic components in their mechanisms. The non-DC earthquakes are intermingled spatially with DC events, indicating that the non-DC mechanisms are not artifacts of instrumental errors or propagation through heterogeneous structures.

The Hengill-Greisdalur triple junction is the best studied of these areas. It was there that experiments were first performed in sufficient detail to demonstrate beyond reasonable doubt that non-DC earthquakes with net volume changes occur in nature (Foulger 1988; Miller 1998a) (Fig. 9). The area is currently the richest source of extreme non-DC earthquakes known. Approximately 70 % of the earthquakes involve volume increases, i.e., their radiation patterns are dominated by compressions, and they are thought to be caused by thermal contraction in the heat source of the geothermal area (Fig. 10).

Non-Double-Couple Earthquakes, Fig. 9 Map of the Hengill-Greisdalur volcanic complex, Iceland, showing focal mechanisms of nine representative earthquakes. Events 5, 6, 8, and 9 are indistinguishable from DCs. Compressional-wave polarities (*solid circles* indicate compressions; *open circles* indicate dilatations) and *P*-wave nodes are shown in upper focal hemisphere equal area projection. *Squares* denote downgoing rays plotted on the upper hemisphere. *T*, *I*, and *P* are positions of principal axes. *Lines* indicate epicentral locations. Where *lines* are absent, mechanisms are centered on epicenters. *Triangles* indicate seismometer locations; *dashed lines* show outlines of the main volcanic centers

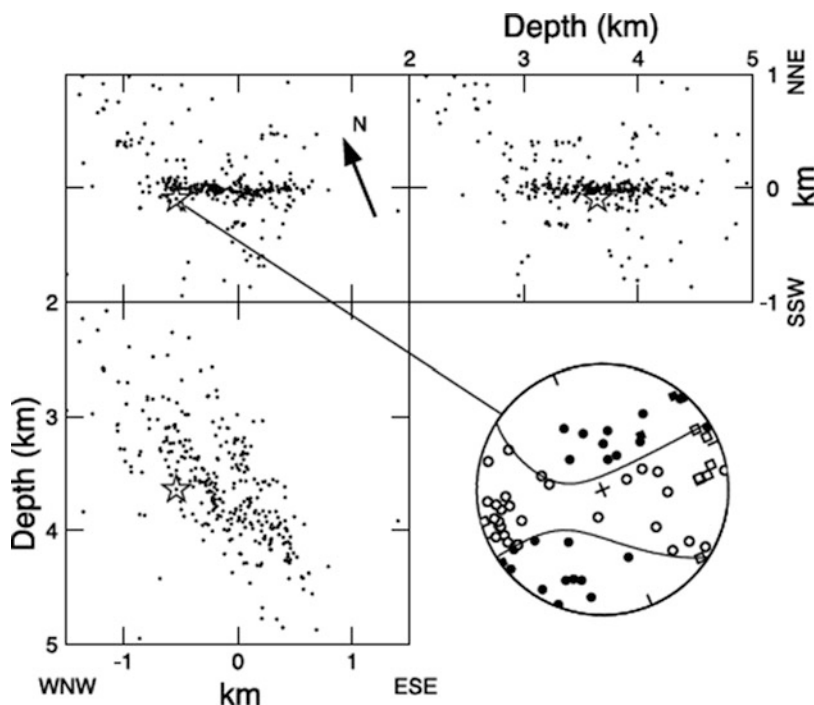


Non-Double-Couple Earthquakes, Fig. 10 Schematic illustration of seismogenic tensile cracking by thermal stresses caused by convective cooling of rocks at the heat source of a geothermal system (From Foulger 1988)



Non-Double-Couple Earthquakes,

Fig. 11 Non-shear fault plane delineated by earthquake hypocenters. Views from three orthogonal directions showing the locations of 314 earthquakes (*black dots*) that lie between 2 and 5 km depth. *Upper left*: map view; *upper right*: SSW-NNE vertical cross-section; *lower left*: WNW-ESE vertical cross-section. *Star*: earthquake whose focal mechanism is shown at *lower right* as an upper-hemisphere plot of compressional-wave polarities. The mechanism for this earthquake is compatible with compensated tensile failure



The source mechanisms require more than simple tensile cracking, because the radiation patterns include both compressions and dilatations. Fluid flow and/or shear faulting probably also contribute components.

Non-DC geothermal earthquakes with volumetric mechanisms occur also at several volcanic-geothermal areas in California, including the Geysers steam field (Ross et al. 1996), the Coso Hot Springs (Julian et al. 2010), and Long Valley caldera (Foulger et al. 2004). Industrial steam extraction and water reinjection at the Geysers induces thousands of small earthquakes per month in the reservoir. The mechanisms are typically closer to DCs than those of Icelandic geothermal earthquakes, and the volumetric non-DC components range from explosive to implosive, with approximately equal numbers of each type.

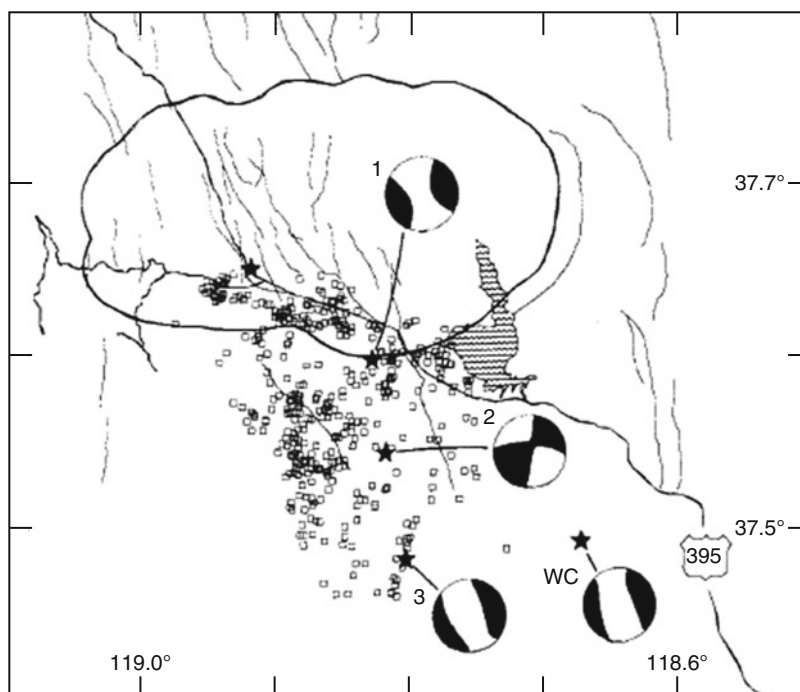
Non-DC earthquake focal mechanisms were observed at the Coso Hot Springs geothermal field, California, in March 2005. There, a fluid injection activated a fault plane on which earthquakes with non-DC mechanisms of uniform type occurred. These involved volume increases. The orientation of the mechanisms relative to the

fault plane implied a process dominated by tensile failure with subsidiary shear faulting.

Abundant non-DC earthquakes were observed in a detailed experiment at Long Valley caldera in 1997. More than 80 % of the mechanisms had explosive volumetric components and most had compensated linear-vector dipole components with outward-directed major dipoles. The simplest interpretation of these mechanisms is combined shear and extensional faulting accompanied by a volume-compensating process such as the rapid flow of water, steam, or CO₂ into opening tensile cracks. A particularly well-recorded swarm was consistent with extensional faulting on an ESE-striking subvertical plane, an orientation consistent with the activated zone defined by the earthquake hypocenters. The orientations of non-DC earthquakes beneath Mammoth Mountain, a volcanic cone on the southwestern caldera rim, varied systematically with location, reflecting a variable local stress field. Events in a spasmodic burst indicated nearly pure compensated tensile failure and high fluid mobility (Foulger et al. 2004) (Fig. 11). Earthquakes with M_W 4.6–4.9 observed at

Non-Double-Couple Earthquakes,

Fig. 12 Long Valley caldera, California, and vicinity, showing the best located earthquakes in 1980 and mechanisms for the largest earthquakes of 1978 and 1980 as lower-hemisphere, equal-area projections, with fields of compressional-wave polarity in black



regional distances on broadband sensors also had non-DC moment tensors with significant volumetric components and were probably caused by hydrothermal or magmatic processes (Dreger et al. 2000).

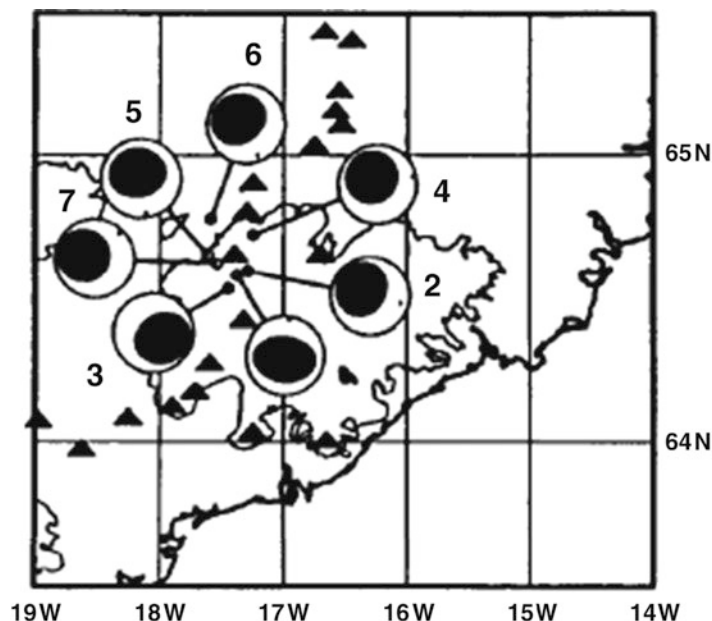
Earthquakes with non-DC radiation patterns, some of them moderately large, accompany magmatic and volcanic activity. Many non-DC earthquakes with P -wave polarities that are either all dilatational or all compressional accompanied the 1983 eruption of Miyakejima volcano, Japan. The earthquakes radiated significant shear waves, indicating that their mechanisms were not purely isotropic. The P -wave polarities and P - and SV -wave amplitudes are compatible with combined tensile and shear faulting (section “Combined Tensile and Shear Faulting”), an interpretation supported by the numerous open cracks that formed prior to the eruption. A similar but substantially larger earthquake, of magnitude 3.2, occurred 10 km beneath the Unzen volcanic region, western Kyushu, in 1987 (Shimizu et al. 1987).

Long Valley caldera, in eastern California, has experienced some of the largest clearly non-DC

volcanic earthquakes. Four events with M_L greater than six occurred there in May 1980 (Fig. 12), and at least two had non-DC mechanisms. These events followed 2 years of volcanic unrest characterized by escalating seismic activity and surface deformation indicating magma-chamber inflation. Moment tensor inversions of different subsets of the available data conducted independently and using different methods all gave mechanisms that were similar and close to CLVDs. Any volumetric components were unresolvably small. The source processes probably involved tensile faulting at high fluid pressure, perhaps associated with dike intrusion. A M_S 5.6 earthquake similar in mechanism to the Long Valley earthquakes, though differently orientated, occurred near Tori Shima island, in the Izu-Bonin arc, on June 13, 1984. Because the earthquake was shallow, its full moment tensor could not be determined well, but the data were consistent with a mechanism close to a CLVD with a vertical symmetry axis. The seismic moment and source duration required intrusion of approximately 0.02 km^3 within 10–40 s (Kanamori et al. 1993).

Non-Double-Couple Earthquakes,

Fig. 13 Harvard centroid moment tensor mechanisms of earthquakes at Bardarbunga volcano, southeast Iceland, from Ekström (1994)



Bardarbunga volcano, beneath the Vatnajökull ice cap in Iceland, regularly generates non-DC earthquakes with M_W 5.2–5.6 ($M_0 = 8$ to 30×10^{16} Nm) (Ekström 1994). They typically have nearly vertical CLVD-like mechanisms. A particularly well-studied example that occurred in 1996 yielded a non-DC mechanism with a 67 % vertically oriented compensated linear vector dipole component, a 32 % DC component, and a statistically insignificant (2 %) volumetric contraction. They may result from slip on a ring fault and can be explained by activation of less than 55° of such a fault. Cone sheets, which dip inward, have the most favorable geometry (Fig. 13).

Earthquakes at Mines

Deep mine excavations perturb the stresses in the surrounding rocks, reducing some components from values initially of the order of 100 MPa practically to zero. The resulting stress differences can exceed the strength of competent rock and cause earthquakes (often called “rock bursts,” “coal bumps,” etc.). Seismic data show clearly that many earthquakes at mines have non-DC mechanisms, usually with predominantly dilatational radiation patterns, that are incompatible with shear faulting.

An M_L 3.5 earthquake on May 14, 1981, at the Gentry Mountain mine, Utah, coincided with the collapse of a large cavity at 200-m depth in the mine.

Well-recorded earthquakes occurring at the Coeur d’Alene mining district, Idaho, had dilatational P -wave polarities, consistent with cavity collapse.

A M_W 3.9 event occurred in the Crandall Canyon coal mine, Utah, August 6, 2007, in association with a gallery collapse that killed six miners and, later, three rescuers. Its focal mechanism is dominated by an implosive component, consistent with a shallow underground collapse (Dreger et al. 2008).

Many non-DC earthquakes occurred in association with the excavation of a 3.5-m-wide tunnel in unfractured, homogeneous granite at the Underground Research Laboratory in Manitoba, Canada (Feignier and Young 1992). Moment tensors computed for 33 earthquakes with M_W -2 to -4 include tensile, implosive, and shear mechanisms. Some were associated with breakout on the tunnel roof, and others occurred ahead of the active face.

Earthquakes located within 150 m of deep gold mines in the Witwatersrand, South Africa, had large volume decreases consistent with the partial collapse of mine galleries (McGarr 1992).

Fluid Injection

Fluids are injected into the shallow crust increasingly often for reasons such as increasing permeability within geothermal and hydrocarbon reservoirs by hydrofracturing, sequestration of CO₂, storing gas, and disposing of waste. Such injection often induces earthquakes, but they are not always monitored in detail. Nevertheless, in a number of well-studied cases non-DC source mechanisms have been observed (e.g., Baig and Urbancic 2010; Julian et al. 2010). Source types typically range from explosive to implosive dipoles. As underground fluid injections increase in importance over the coming decades, more detailed studies of the source mechanisms of the earthquakes induced may prove to be important for understanding the response of the Earth to such operations, project optimization and hazard mitigation.

Other Shallow Earthquakes

Real fault surfaces are not perfectly planar, and earthquakes on curved faults can, for some geometries, have non-DC mechanisms (Frohlich et al. 1989). Effects of this sort may account for the non-DC components observed in some large earthquakes, such as the 28-km deep M_S 7.8 event that occurred near Taiwan on November 14, 1986 (Zheng et al. 1995). The focal mechanism is inconsistent with shear faulting, but requires an implosive component. A similar, M 4.6 earthquake that occurred February 10, 1987, beneath the Kanto district, Japan. It had primarily dilatational compressional-wave polarities and was consistent with a conical nodal plane.

Deep-Focus Earthquakes

Earthquakes up to 700 hundred km deep occur in the mantle beneath subduction zones. The physical causes of deep earthquakes are not well understood. This is because minerals are expected to flow plastically at depths greater than about 30 km in normal areas. Processes that have been suggested to explain deep earthquakes include plastic instabilities, shear-induced melting, polymorphic phase transformations, and transformational faulting. Focal mechanisms can potentially shed light on this problem.

Volume changes in deep earthquakes are statistically unresolvable (less than 10 % of the seismic moment) (Kawakatsu 1991), indicating that polymorphic phase changes (section “[Rapid Polymorphic Phase Changes](#)”), play an insignificant role in deep earthquakes. Deep earthquakes do, however, have larger CLVD components than shallow earthquakes, and in some cases the sizes of the CLVD components increases systematically with depth and magnitude. The lack of large volume changes in deep earthquakes is compatible with complex shear faulting, and detailed waveform analysis resolves some deep non-DC earthquakes into DC subevents.

Glacial Earthquakes

Glaciers are subject to several kinds of sudden transient phenomena, apparently caused by diverse physical processes. Application of seismological analysis techniques to these events promises to provide fundamentally new types of data bearing on glacier dynamics and on understanding glaciers’ response to influences such as climate change (Ekström et al. 2006).

Earthquakes apparently caused by glacier surges or calving, even though they can be large ($M \geq 4.5$), were recognized on seismograms only recently, because they have source durations of the order of a minute or more and excite high-frequency seismic waves only weakly. The application of array-processing techniques to digital seismic data from the global network now enables us to detect dozens of large glacial earthquakes per year that previously went unnoticed. A particularly well-recorded event, of magnitude about 5, occurred in 1999 near the Dall glacier in the Denali range of Alaska, within a regional seismometer network and near a second, temporarily deployed, network. The low-frequency (~ 0.01 – 0.2 Hz) seismic waves are inconsistent with any moment-tensor mechanism, but agree well with those predicted for an approximately horizontal force directed in the direction of the glacier flow (Ekström et al. 2003). Seismic data from numerous other events, most of them near the margins of the Greenland icecap, are consistent with similar single-force mechanisms.

Most small ($M \geq -2.5$) icequakes in mountain glaciers have mechanisms describable by moment tensors, but they often differ greatly from DCs. Walter et al. (2009, 2010) recorded tens of thousands of events on Gornergletscher, in the Swiss Alps, on dense seismometer networks, and extended Dreger's waveform-inversion method, originally developed for earthquakes above magnitude 4, to frequencies approaching 1 KHz, in order to study the mechanisms of these microearthquakes. Most events had large explosive volumetric components, with source types close to those for opening tensile cracks. These included shallow events, probably representing crevasse opening, and events 100 m below the surface. Many shallow events were of a different type, equivalent to a DC combined with a volume decrease. A later study (Walter et al. 2010) of events near the base of the glacier found opening-crack mechanisms, but with inferred crack planes oriented horizontally.

Summary

A wide variety of processes can cause earthquake mechanisms to depart from the ideal DC force system that characterizes planar shear faulting in a homogeneous isotropic medium. These departures can be as extreme as unbalanced forces or torques associated with major landslides, glacial surges, or volcanic eruptions, or can be minor anomalies that are barely resolvable with the best data currently available. Studying non-DC earthquake mechanisms can be valuable for refining our knowledge of how faults work, for learning to understand and predict volcanic activity, for prospecting and exploiting geothermal energy and hydrocarbons, and for avoiding damage to civil and industrial infrastructure. Moreover, non-DC mechanisms are of direct value for monitoring mining and other industrial activities, particularly ones that involve fluid injection.

References

Aki K, Richards PG (2002) Quantitative seismology, 2nd edn. University Science, Sausalito

- Anderson EM (1936) The dynamics of the formation of cone-sheets, ring-dykes, and cauldron-subsidences. *Proc R Soc Edinb* 56:128–156
- Backus G (1977) Interpreting the seismic glut moments of total degree two or less. *Geophys J R Astron Soc* 51:1–25
- Baig A, Urbancic T (2010) Microseismic moment tensors: a path to understanding frac growth. *Lead Edge* 29:320–324
- Booth DC, Crampin S (1985) Shear-wave polarizations on a curved wavefront at an isotropic free surface. *Geophys J R Astron Soc* 83:31–45
- Brace WF, Paulding BW Jr, Scholz CH (1966) Dilatancy in the fracture of crystalline rocks. *J Geophys Res* 71:3939–3953
- Dreger DS, Tkalčić H, Johnston M (2000) Dilational processes accompanying earthquakes in the Long Valley caldera. *Science* 288:122–125
- Dreger D, Ford SR, Walter WR (2008) Source analysis of the Crandall Canyon, Utah mine collapse. *Science* 321:217
- Ekström G (1994) Anomalous earthquakes on volcano ring-fault structures. *Earth Planet Sci Lett* 128:707–712
- Ekström G, Nettles M, Abers GA (2003) Glacial earthquakes. *Science* 302:622–624
- Ekström G, Nettles M, Tsai VC (2006) Seasonality and increasing frequency of Greenland Glacial earthquakes. *Science* 311:1756–1758
- Feignier B, Young RP (1992) Moment tensor inversion of induced microseismic events: evidence of non-shear failures in the $-4 < M < -2$ moment magnitude range. *Geophys Res Lett* 19:1503–1506
- Foulger GR (1988) Hengill triple junction, SW Iceland; 2. Anomalous earthquake focal mechanisms and implications for process within the geothermal reservoir and at accretionary plate boundaries. *J Geophys Res* 93:13507–13523
- Foulger GR, Julian BR, Hill DP, Pitt AM, Malin P, Shalev E (2004) Non-double-couple microearthquakes at Long Valley caldera, California, provide evidence for hydraulic fracturing. *J Volcanol Geotherm Res* 132:45–71
- Frohlich C, Riedesel MA, Apperson KD (1989) Note concerning possible mechanisms for non-double-couple earthquake sources. *Geophys Res Lett* 16:523–526
- Gajewski D (1993) Radiation from point sources in general anisotropic media. *Geophys J Int* 113:299–317
- Gilbert F (1970) Excitation of the normal modes of the Earth by earthquake sources. *Geophys J R Astron Soc* 22:223–226
- Hudson JA, Pearce RG, Rogers RM (1989) Source type plot for inversion of the moment tensor. *J Geophys Res* 94:765–774
- Iguchi M (1994) A vertical expansion source model for the mechanisms of earthquakes originated in the magma conduit of an andesitic volcano: Sakurajima, Japan. *Bull Volcanol Soc Jpn* 39:49–67

- Julian BR (1986) Analysing seismic-source mechanisms by linear-programming methods. *Geophys J R Astron Soc* 84:431–443
- Julian BR (1994) Volcanic tremor: nonlinear excitation by fluid flow. *J Geophys Res* 99:11859–11877
- Julian BR, Foulger GR (1996) Earthquake mechanisms from linear-programming inversion of seismic-wave amplitude ratios. *Bull Seismol Soc Am* 86: 972–980
- Julian BR, Sipkin SA (1985) Earthquake processes in the Long Valley caldera area, California. *J Geophys Res* 90:11155–11169
- Julian BR, Foulger GR, Monastero FC, Bjornstad S (2010) Imaging hydraulic fractures in a geothermal reservoir. *Geophys Res Lett* 37:L07305. doi:10.1029/2009GL040933
- Kanamori H, Given JW, Lay T (1984) Analysis of seismic body waves excited by the Mount St. Helens eruption of May 18, 1980. *J Geophys Res* 89:1856–1866
- Kanamori H, Ekström G, Dziewonski A, Barker JS, Sipkin SA (1993) Seismic radiation by magma injection – an anomalous seismic event near Tori Shima, Japan. *J Geophys Res* 98:6511–6522
- Kawakatsu H (1991) Insignificant isotropic component in the moment tensor of deep earthquakes. *Nature* 351:50–53
- Knopoff L, Randall MJ (1970) The compensated linear vector dipole: a possible mechanism for deep earthquakes. *J Geophys Res* 75:4957–4963
- McGarr A (1992) Moment tensors of ten Witwatersrand mine tremors. *Pure Appl Geophys* 139:781–800
- Mendiguren JA (1977) Inversion of surface wave data in source mechanism studies. *J Geophys Res* 82:889–894
- Miller AD, Julian BR, Foulger GR (1998a) Three-dimensional seismic structure and moment tensors of non-double-couple earthquakes at the Hengill-Grensadalur volcanic complex, Iceland. *Geophys J Int* 133:309–325
- Miller AD, Foulger GR, Julian BR (1998b) Non-double-couple earthquakes II. Observations. *Rev Geophys* 36:551–568
- Price NJ (1966) *Fault and joint development in brittle and semi-brittle rock*. Pergamon Press, New York
- Richards PG, Kim WY, (2005) Equivalent volume sources for explosions at depth: Theory and observations. *Bull Seismol Soc Am* 95:401–407
- Ross A, Foulger GR, Julian BR (1996) Non-double-couple earthquake mechanisms at The Geysers geothermal area, California. *Geophys Res Lett* 23:877–880
- Shimizu H, Ueki S, Koyama J (1987) A tensile-shear crack model for the mechanism of volcanic earthquakes. *Tectonophysics* 144:287–300
- Sipkin SA (1982) Estimation of earthquake source parameters by the inversion of waveform data: synthetic waveforms. *Phys Earth Planet Inter* 30:242–259
- Takei Y, Kumazawa M (1994) Why have the single force and torque been excluded from seismic source models? *Geophys J Int* 118:20–30
- Uhira K, Takeo M (1994) The source of explosive eruptions of Sakurajima volcano, Japan. *J Geophys Res* 99:17775–17789
- Vavryčuk V (2005) Focal mechanisms in anisotropic media. *Geophys J Int* 161:334–346
- Walter F, Clinton JF, Deichmann N, Dreger DS, Minson SE, Funk M (2009) Moment tensor inversions of icequakes on Gornergletscher, Switzerland. *Bull Seismol Soc Am* 99:852–870
- Walter F, Dreger DS, Clinton JF, Deichmann N, Funk M (2010) Evidence for near-horizontal tensile faulting at the base of Gornergletscher, a Swiss Alpine glacier. *Bull Seismol Soc Am* 100:458–472
- Woodhouse JH (1981) The excitation of long period seismic waves by a source spanning a structural discontinuity. *Geophys Res Lett* 8:1129–1131
- Zheng T, Yao Z, Liu P (1995) The 14 November 1986 Taiwan earthquake – an event with isotropic component. *Phys Earth Planet Inter* 91:285–298

Nonlinear Analysis and Collapse Simulation Using Serial Computation

Xinzheng Lu¹ and Xiao Lu²

¹Key Laboratory of Civil Engineering Safety and Durability of China Education Ministry, Department of Civil Engineering, Tsinghua University, Beijing, People's Republic of China

²School of Civil Engineering, Beijing Jiaotong University, Beijing, People's Republic of China

Synonyms

Collapse simulation; Elemental nonlinearity; Geometric nonlinearity; Material nonlinearity; Nonlinear analysis

Introduction

Due to the extensive use of performance-based seismic design in earthquake engineering, nonlinear analysis has become an important method for evaluating the seismic performances of structures. In comparison with elastic analysis, nonlinear analysis is usually more complex and expensive. Nonlinear analysis generally involves

nonlinear static analysis (pushover) and nonlinear dynamic analysis in earthquake engineering. The nonlinearities mainly come from material nonlinearity, geometric nonlinearity, elemental nonlinearity, and contact nonlinearity. Only when these nonlinearities are considered correctly in the analysis models can the structural responses be reasonably predicted.

Material Nonlinearity

Most materials used in structural engineering exhibit nonlinear behavior, such as concrete, steel, and soil. For example, the deformation is not proportional to the load, and permanent plastic deformation remains after being unloaded. In addition to the elastic material constants (Young's modulus and Poisson's ratio), the yield criterion, the work hardening, and the flow rules are very essential to describe the nonlinear material behavior (Chen 2005).

The yield stress of a material is a measured stress level that separates the elastic and inelastic behavior of the material. The magnitude of the yield stress is generally obtained from the results of a uniaxial test. However, the stresses in a structure are usually multiaxial. A measurement of yielding for the multiaxial state of stress is called the yield criterion. For example, the Mohr-Coulomb criterion, the Buyukozturk criterion, and the Drucker-Prager criterion are common yield criteria for concrete, and the von Mises criterion is the widely used yield criterion for steel.

Isotropic hardening, kinematic hardening, and combined hardening are three typical work hardening rule types in civil engineering (Chen 2005). The isotropic work hardening rule assumes that, due to work hardening, the center of the yield surface remains stationary in the stress space and the size (radius) of the yield surface expands. In contrast, for the kinematic hardening rule, the yield surface does not change in size or shape, and the center of the yield surface can move in the stress space.

Yield stress and work hardening rules are two experimentally related phenomena that

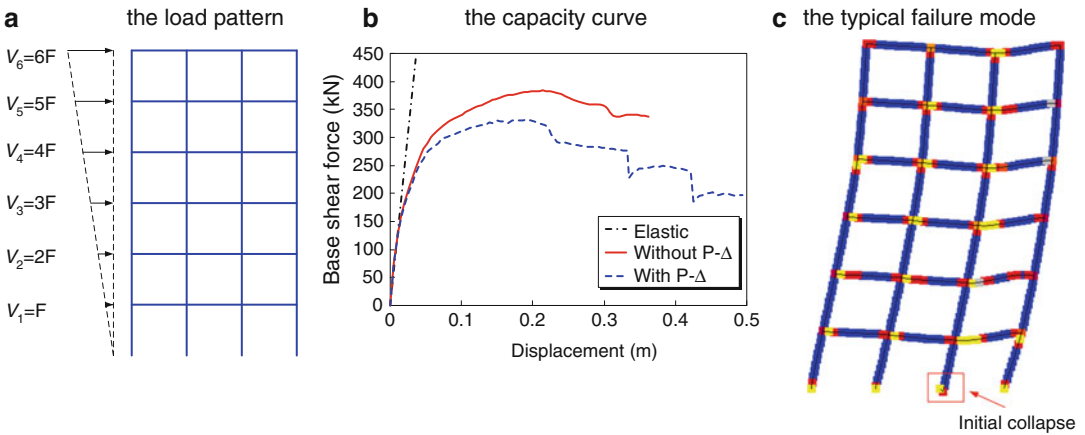
characterize plastic material behavior. The flow rule is also essential in establishing the incremental stress-strain relationships for plastic material. The flow rule describes the differential changes in the plastic strain components as a function of the current stress state (Chen 2005). In most nonlinear analyses, the associated flow rule in which the direction of inelastic straining is normal to the yield surface is adopted.

Geometric Nonlinearity

There are two natural classes of geometric nonlinearities: the large displacement, small strain problem and the large displacement, large strain problem. For the large displacement, small strain problem, changes in the stress-strain law can be neglected, but the contributions from the nonlinear terms in the strain-displacement relations cannot be neglected. For the large displacement, large strain problem, the constitutive relation, which must be defined in the correct frame of reference, is transformed from this frame of reference to the one in which the equilibrium equations are written. Most of the geometric nonlinearities in structural engineering correspond to the large displacement, small strain problem; two typical examples are the $P-\Delta$ effect of a structure and the $P-\delta$ effect of a component. The total Lagrangian procedure can be used to describe the geometric nonlinearity in both the static and dynamic analyses. The equilibrium is expressed with the original undeformed state as the reference. The nonlinear analysis method should have the ability to address this geometric nonlinearity; if it does not have the ability, the structural capacity will be overestimated, leading to an unsafe design.

Elemental Nonlinearity

During the nonlinear analysis, the structural components may reach their load-carrying capacities and be out of work, at which point, the elemental nonlinearity occurs. To handle



Nonlinear Analysis and Collapse Simulation Using Serial Computation, Fig. 1 Example of a pushover analysis: (a) the load pattern, (b) the capacity curve, and (c) the typical failure mode

this problem, the elemental deactivation technology, in which the failed elements are deactivated when a specified elemental-failure criterion is satisfied, can be used (Lu et al. 2011, 2013). For example, the material-related failure criterion should be used to monitor the failure of structural elements. If the strain at any integration point exceeds the material failure criterion, the stress and the stiffness of this element are deactivated, which means that this element no longer contributes to the stiffness computation of the entire structure. The nodes that are connected to the deactivated element are simultaneously checked. A node is considered “isolated” when all its constituent elements are deactivated, which implies that this node contains zero stiffness in relation to the degrees of freedom (DOFs). Because the “isolated” nodes are more likely to cause convergence problems, all of the DOFs of the “isolated” nodes are removed from the global stiffness matrix.

Nonlinear Static Analysis

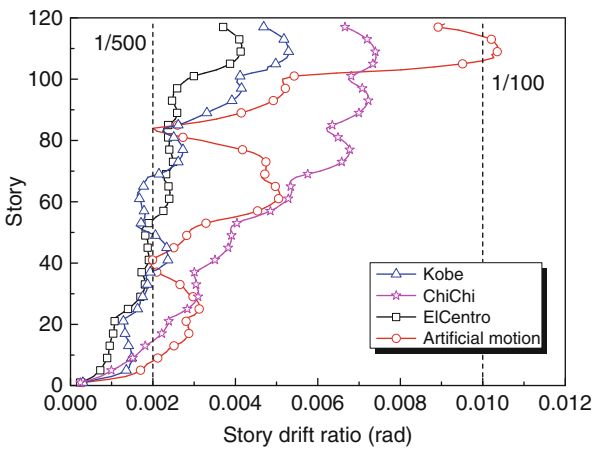
Nonlinear static analysis, which is commonly referred to as pushover analysis in earthquake engineering, enables the simple evaluation of the bearing of structural loads and the deformation capacities, as well as the evaluation of

the internal forces and deformations of components. This analysis method was first proposed by Freeman et al. (1975) to approximately assess the structural seismic performance using the seismic demand spectrum. During the pushover analysis, a given lateral load pattern is applied to the building and increases gradually to obtain the structural internal forces and displacements. Combined with the use of the seismic demand spectrum or the target displacement performance, the approximate seismic performance under the expected earthquake can be evaluated. As a simplified performance evaluation method, the following assumption should be complied: the deformation of the building is controlled by the first vibration mode, and this deformation pattern remains unchanged along the building height during the analysis process. Correspondingly, four types of load patterns are widely used in the pushover analysis, including the height-based load pattern, the fundamental mode-dominated load pattern, the SRSS-based load pattern, and the uniform load pattern.

Taking a simple frame structure (shown in Fig. 1a) as an example, the fiber-beam element (Lu et al. 2013) or plastic hinge model (Ibarra et al. 2005) can be used to simulate the nonlinear behavior of beams and columns. The inverted-triangle load pattern, which is a particular

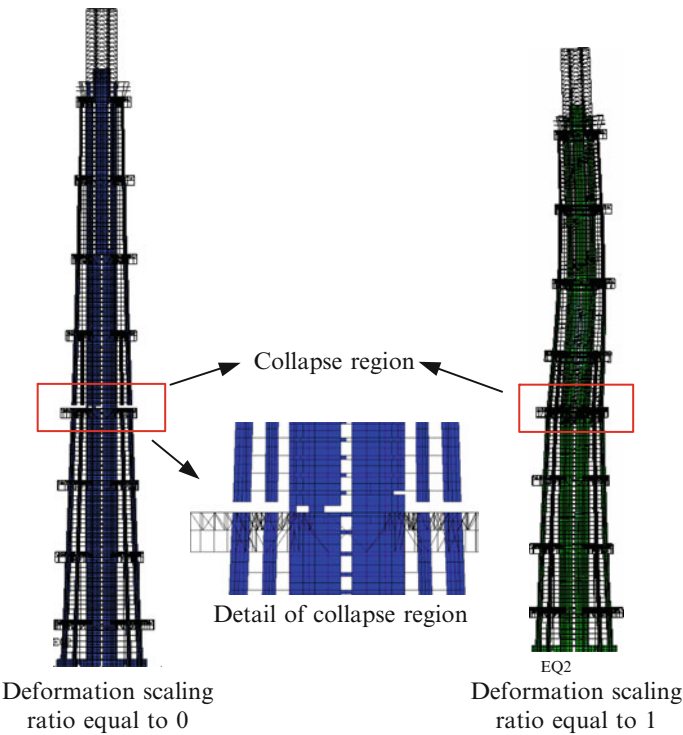
Nonlinear Analysis and Collapse Simulation Using Serial Computation,

Fig. 2 Displacement responses at the MCE level



Nonlinear Analysis and Collapse Simulation Using Serial Computation,

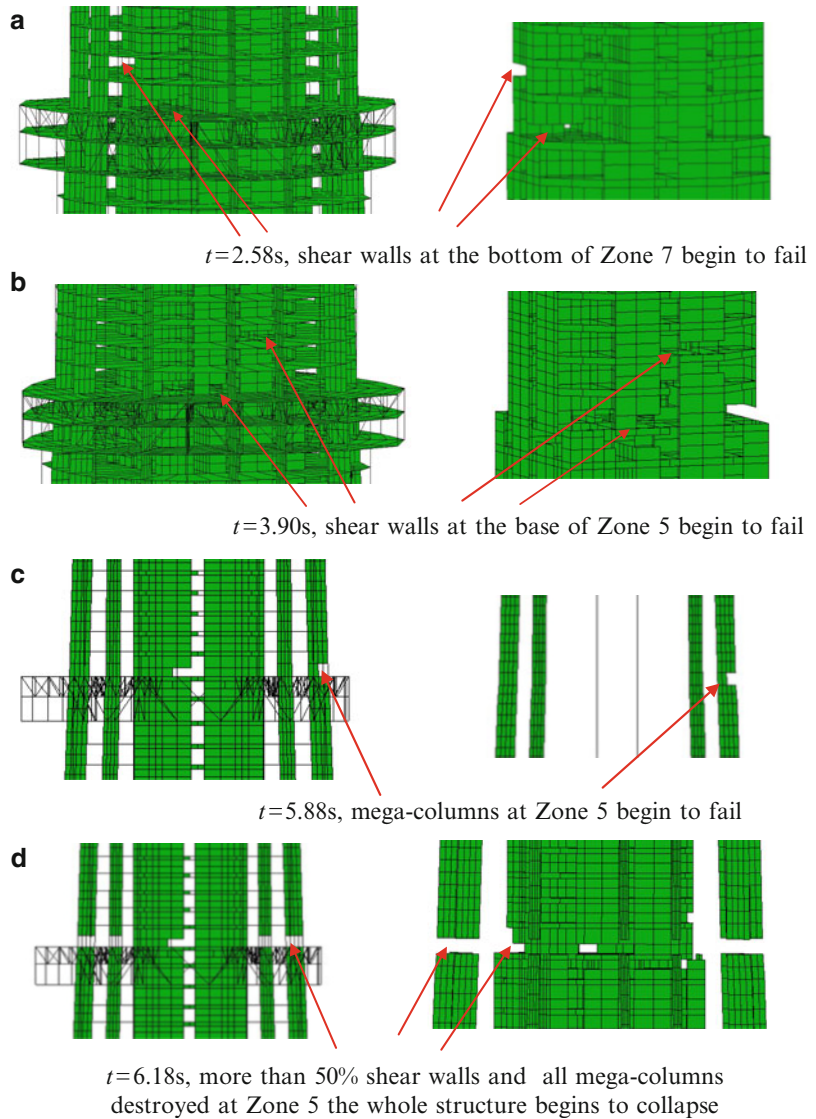
Fig. 3 Typical collapse mode subjected to the El-Centro ground motion



pattern of the height-based load pattern, is applied to this building, resulting in the capacity curve. If the linear analysis is conducted, the capacity curve is the straight line in Fig. 1b. In addition, if only material nonlinearity is considered, the capacity curve is the solid line in Fig. 1b, which overestimates the load-carrying

capacity of the building. Correspondingly, if both the material and geometric nonlinearities are considered, the capacity curve is the dashed line in Fig. 1b, which better reflects the real seismic performance of the building. In addition, the initial collapse is predicted to occur in the middle column (shown in Fig. 1c).

Nonlinear Analysis and Collapse Simulation Using Serial Computation,
Fig. 4 Details of the collapse process



Nonlinear Dynamic Analysis

In comparison with the nonlinear static analysis, the nonlinear dynamic analysis, which has attracted more attention in recent years, exhibits a broader range of applications. Because nonlinear dynamic analysis is directly based on the structural dynamic equation, the displacement, velocity, and acceleration of each story of a building and the internal forces, cracking, and yielding of the individual structural components of a building can be obtained

at any moment. The analysis also reflects the characteristics of ground motion, including amplitude and duration. Therefore, this method has been widely used in many building codes for the seismic evaluation of buildings with unusual configurations or of special importance. However, the structural response may be very sensitive to the characteristics of the individual ground motion used as the seismic input; therefore, seven pairs of ground motion records, at least three pairs, are suggested in most building codes as the basic seismic input to

achieve a reliable estimation of the mean structural responses. In addition, the damping effect must be considered in the dynamic analysis. Rayleigh damping is one of the most effective means of deriving a suitable damping matrix, which can be expressed in proportion to the structural mass matrix and the initial or current tangent stiffness matrix (Clough and Penzien 1975). Subsequently, based on the nonlinear detailed finite element model and appropriate ground motion records, the seismic performance of the building at various seismic hazard levels can be predicted and evaluated, as well as the potential collapse modes.

Taking an actual super-tall building, for example, the nonlinear dynamic analysis of the Shanghai Tower, with the total height of 632 m, is described in the following. The beams and columns are simulated by fiber-beam elements, the shear walls are simulated by multilayer shell elements, and elemental deactivation technology is used to simulate the components failure (Lu et al. 2011). The fundamental period is approximately 9.8 s. The displacement responses under four widely used ground motion records at the maximum considered earthquake (MCE) level and the typical potential collapse mode under the extreme earthquake are shown in Figs. 2 and 3, respectively.

The details of the collapse process are illustrated in Fig. 4. When $t = 2.58$ s, some coupling beams in the core tube begin to fail, and the flange wall of the core tube at the bottom of Zone 7 is crushed. This crushing is caused by the changes in the layout of the openings in the core tube between Zones 6 and 7, which results in sudden changes in the stiffness and the stress concentration. When $t = 3.90$ s, the shear wall at the bottom of Zone 5 begins to fail because the cross section of the core tube changes from Zone 4 to Zone 5, as shown in Fig. 4b. When $t = 5.88$ s, over 50 % of the shear walls at the bottom of Zone 5 fail, and the internal forces are redistributed to the remaining components. The vertical and horizontal loads in the mega-columns gradually increase and reach their load capacities. Subsequently, the mega-columns begin to fail. When $t = 6.18$ s, the core tube and mega-columns in Zone 5 are completely destroyed, and the collapse begins to propagate

throughout the entire structure. Obviously, Zone 5 is a typical potentially weak part.

Summary

To improve the structural seismic safety of new and existing buildings is always the most important objective in earthquake engineering. Based on the appropriate numerical models, including the models of the material nonlinearity, geometric nonlinearity, and elemental nonlinearity, the nonlinear analysis, especially the nonlinear dynamic analysis, is currently one of the best tools for the prediction of structural responses at varying seismic hazard levels. In addition, the potential collapse modes and the corresponding weak parts can be predicted using a collapse simulation, which provides a better understanding of the collapse mechanism of buildings and contributes to the further development of both the optimum design and the structural health monitoring methodologies.

Cross-References

- [Nonlinear Dynamic Seismic Analysis](#)
- [Seismic Collapse Assessment](#)

References

- Chen HF (2005) Elasticity and plasticity. China Architecture & Building Press, Beijing
- Clough RW, Penzien J (1975) Dynamics of structures. McGraw Hill, New York
- Freeman SA, Nicoletti JP, Tyrell JV (1975) Evaluation of existing buildings for seismic risk—a case study of Puget Sound Naval Shipyard, Bremerton, Washington. In: Proceedings of the 1st U.S. national conference earthquake engineering, EERI, Berkeley, pp 113–122
- Ibarra LF, Medina RA, Krawinkler H (2005) Hysteretic models that incorporate strength and stiffness deterioration. *Earthq Eng Struct Dyn* 34(12):1489–1511
- Lu X, Lu XZ, Zhang WK, Ye LP (2011) Collapse simulation of a super high-rise building subjected to extremely strong earthquakes. *Sci China Technol Sci* 54(10):2549–2560
- Lu X, Lu XZ, Guan H, Ye LP (2013) Collapse simulation of reinforced concrete high-rise building induced by extreme earthquakes. *Earthq Eng Struct Dyn* 42(5):705–723

Nonlinear Dynamic Seismic Analysis

Evangelos Sapountzakis

Department of Structural Engineering, School of Civil Engineering, National Technical University of Athens (N.T.U.A.), Athens, Greece

Synonyms

Flexural–torsional analysis; Geometrically nonlinear dynamic analysis; Secondary torsional moment; Shear deformation warping

Introduction

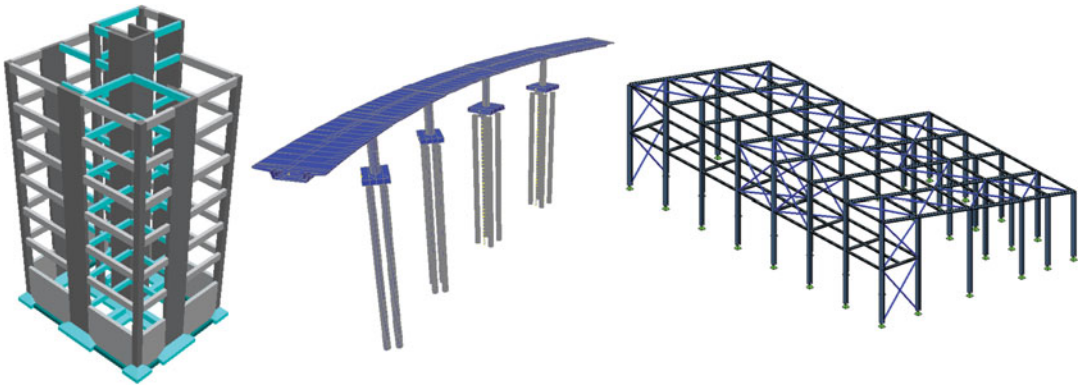
In engineering practice beam-like continuous systems are encountered in the vast majority of structural engineering projects (Fig. 1). The term “beam-like” refers to these structural members, one dimension of which is significantly larger than the other two. As far as the analysis of such members is concerned, many modeling strategies can be applied (based mainly on finite elements); however, it is common knowledge that, especially in cases of complicated beam assemblages, beam theories are a very attractive modeling approach. Beam elements are practical and computationally efficient and offer better insight of the structural phenomena since they permit their isolation and their independent investigation. Furthermore, due to easy parameterization of all necessary data, beam elements are more convenient for parametric analyses than more refined models which, in most cases, require the construction of multiple models.

Apart from static external loading, structures are very often subjected to external loading, the magnitude, the direction, and the position of which vary with time, i.e., dynamic external loading. Such cases of loading include seismic excitations, excitations due to wind conditions, vibrations induced by large mechanical devices, impact loading due to explosions, etc. The aforementioned arbitrary external dynamic loading leads to the formulation of the flexural–torsional

vibration problem. The complexity of this problem increases significantly in the case the cross section’s centroid does not coincide with its shear center (asymmetric beams). Furthermore, when arbitrary torsional boundary conditions are applied either at the edges or at any other interior point of a beam due to construction requirements, the beam under the action of general twisting loading is led to nonuniform torsion. Moreover, since requirement of weight saving is a major aspect in the design of structures, thin-walled elements of arbitrary cross section and low flexural and/or torsional stiffness are extensively used. Treating displacements and angles of rotation of these elements as being small leads in many cases to inadequate prediction of the dynamic response; hence, the occurring nonlinear effects should be taken into account. This can be achieved by retaining the nonlinear terms in the strain–displacement relations (finite displacement–small strain theory).

When the displacement components of a member are small, a wide range of linear analysis tools, such as modal analysis, can be used, and some analytical results are possible. As these components become larger, the induced geometrical nonlinearities result in effects that are not observed in linear systems. When finite displacements are considered, the flexural–torsional dynamic analysis of beams becomes much more complicated, leading to the formulation of coupled and nonlinear flexural, torsional, and axial equations of motion. The analysis of these systems becomes even more complicated when shear deformation effect in flexure and secondary torsional moment deformation effect (STMDE) in torsion, which are significant in many cases (e.g., short beams, beams of box-shaped cross sections, folded structural members, beams made of materials weak in shear, etc.), are taken into account.

In section “[Nonlinear Flexural Dynamic Analysis of Beams with Shear Deformation Effect](#)” of this chapter, the geometrically nonlinear dynamic flexural analysis of homogeneous prismatic beam members taking into account shear deformation and rotary inertia effects (Timoshenko beam theory) is presented. The differential equations of



Nonlinear Dynamic Seismic Analysis, Fig. 1 Structural engineering projects composed by beam structural members

motion governing free or forced vibrations of Timoshenko beams are formulated employing the principle of virtual work. In order to be able to examine the flexural problem independently (i.e., excluding torsional response) and without loss of generality, the case of beams of arbitrary doubly symmetric cross sections subjected to torsionless bending (external transverse loading passes through the shear center of the cross section which coincides with its centroid) is examined. For a detailed review of the relevant literature, the reader is referred to the study of Sapountzakis and Dourakopoulos (2009a, b).

Since very often beam members under arbitrary external dynamic loading develop torsional response, the rigorous dynamic analysis of nonuniform torsion of beams in the nonlinear range is necessary. In section “[Nonlinear Torsional Dynamic Analysis of Beams with Secondary Torsional Moment Deformation Effect \(STMDE\)](#)” of this chapter, the geometrically nonlinear dynamic torsional analysis of homogeneous prismatic beam members taking into account STMDE (analogy with Timoshenko beam theory) is presented. The differential equations of motion governing free or forced vibrations of beams under nonuniform torsion are formulated employing the principle of virtual work. Similarly with section “[Nonlinear Flexural Dynamic Analysis of Beams with Shear Deformation Effect](#),” the case of beams of arbitrary doubly symmetric cross sections subjected to twisting and/or warping moments is examined.

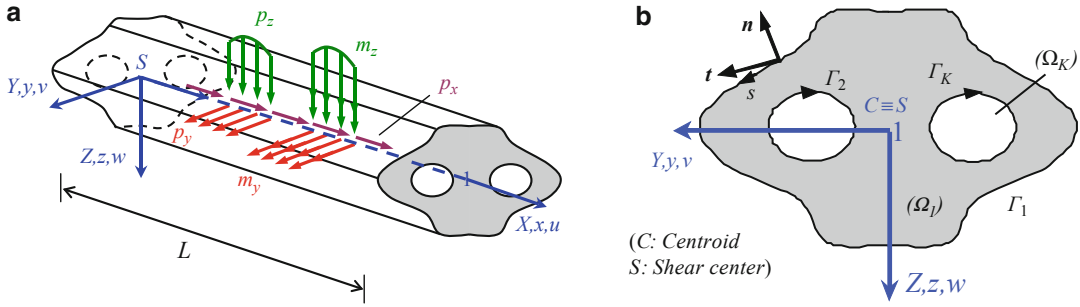
For a detailed review of the relevant literature, the reader is referred to the studies of Sapountzakis and Tsipiras (2010a, b, c).

Finally, in order to highlight the significant coupling between flexure and torsion when displacements and rotations become large, in section “[Nonlinear Flexural–Torsional Dynamic Analysis of Beams of Arbitrary Cross Section](#),” the formulation for the geometrically nonlinear flexural–torsional dynamic analysis of prismatic beams is analyzed. In this section, the most general case is examined by considering arbitrary thin- or thick-walled cross sections possessing no axis of symmetry. The beam may undergo moderately large deflections and twisting rotations, under the most general boundary conditions. The beam is subjected to arbitrarily distributed or concentrated transverse loading, which can be applied on any arbitrary point of the lateral surface of the beam, to bending moments, as well as to twisting and/or axial loading. For a detailed review of the relevant literature, the reader is referred to the study of Sapountzakis and Dikaros (2011).

Nonlinear Flexural Dynamic Analysis of Beams with Shear Deformation Effect

Statement of the Problem

Let us consider a prismatic beam of length L of doubly symmetric cross section of area A , occupying the multiply connected region Ω of the y, z



Nonlinear Dynamic Seismic Analysis, Fig. 2 Prismatic beam in axial–flexural loading (a) with an arbitrary doubly symmetric cross section occupying the two-dimensional region Ω (b)

plane and is bounded by the curve $\Gamma = \cup_{j=1}^K \Gamma_j$, which is piecewise smooth, i.e., it may have a finite number of corners (Fig. 2). The material of the beam is considered to be homogeneous, isotropic, and linearly elastic with modulus of elasticity E , shear modulus G , Poisson ratio ν , and mass density ρ . In Fig. 2b, $Cxyz \equiv CXYZ$ is the principal bending system of axes passing through the cross section's centroid C (coinciding with the shear center S). The beam is subjected to the combined action of the arbitrarily distributed or concentrated time-dependent axial loading $p_x = p_x(x, t)$; transverse loading $p_y = p_y(x, t)$, $p_z = p_z(x, t)$ acting in the y, z directions, respectively; and bending moments $m_y = m_y(x, t)$, $m_z = m_z(x, t)$ along y, z axes, respectively.

Under the action of the aforementioned loading, the beam does not develop twisting rotation, and the displacement field taking into account shear deformation effect can be written as

$$\bar{u}(x, y, z, t) = u(x, t) - y\theta_z(x, t) + z\theta_y(x, t) \quad (1a)$$

$$\bar{v}(x, t) = v(x, t) \quad (1b)$$

$$\bar{w}(x, t) = w(x, t) \quad (1c)$$

where $\bar{u}, \bar{v}, \bar{w}$ are the displacement components of an arbitrary point of the cross section with respect to $Cxyz$; $u(x, t)$, $v(x, t)$, $w(x, t)$ are the corresponding components of the centroid, and $\theta_y(x, t)$, $\theta_z(x, t)$ are the angles of rotation due to bending of the cross section which do not

coincide with the deflection slopes (w' , v'). Employing the strain–displacement relations of the Green–Lagrange strain tensor (Rothert and Gensichen 1987; Ramm and Hofmann 1995), the strain components can be obtained as

$$\begin{aligned} \varepsilon_{xx} &= \frac{\partial \bar{u}}{\partial x} + \frac{1}{2} \left[\left(\frac{\partial \bar{u}}{\partial x} \right)^2 + \left(\frac{\partial \bar{v}}{\partial x} \right)^2 + \left(\frac{\partial \bar{w}}{\partial x} \right)^2 \right] \\ &\approx \frac{\partial \bar{u}}{\partial x} + \frac{1}{2} \left[\left(\frac{\partial \bar{v}}{\partial x} \right)^2 + \left(\frac{\partial \bar{w}}{\partial x} \right)^2 \right] \end{aligned} \quad (2a)$$

$$\begin{aligned} \gamma_{xz} &= \frac{\partial \bar{w}}{\partial x} + \frac{\partial \bar{u}}{\partial z} + \left(\frac{\partial \bar{u}}{\partial x} \frac{\partial \bar{u}}{\partial z} + \frac{\partial \bar{v}}{\partial x} \frac{\partial \bar{v}}{\partial z} + \frac{\partial \bar{w}}{\partial x} \frac{\partial \bar{w}}{\partial z} \right) \\ &\approx \frac{\partial \bar{w}}{\partial x} + \frac{\partial \bar{u}}{\partial z} + \left(\frac{\partial \bar{v}}{\partial x} \frac{\partial \bar{v}}{\partial z} + \frac{\partial \bar{w}}{\partial x} \frac{\partial \bar{w}}{\partial z} \right) \end{aligned} \quad (2b)$$

$$\begin{aligned} \gamma_{xy} &= \frac{\partial \bar{v}}{\partial x} + \frac{\partial \bar{u}}{\partial y} + \left(\frac{\partial \bar{u}}{\partial x} \frac{\partial \bar{u}}{\partial y} + \frac{\partial \bar{v}}{\partial x} \frac{\partial \bar{v}}{\partial y} + \frac{\partial \bar{w}}{\partial x} \frac{\partial \bar{w}}{\partial y} \right) \\ &\approx \frac{\partial \bar{v}}{\partial x} + \frac{\partial \bar{u}}{\partial y} + \left(\frac{\partial \bar{v}}{\partial x} \frac{\partial \bar{v}}{\partial y} + \frac{\partial \bar{w}}{\partial x} \frac{\partial \bar{w}}{\partial y} \right) \end{aligned} \quad (2c)$$

$$\varepsilon_{yy} = \varepsilon_{zz} = \gamma_{yz} = 0 \quad (2d)$$

where it has been assumed that for moderate displacements $(\partial \bar{u} / \partial x)^2 \ll (\partial \bar{u} / \partial x)$, $(\partial \bar{u} / \partial x)(\partial \bar{u} / \partial z) \ll (\partial \bar{u} / \partial x) + (\partial \bar{u} / \partial z)$, $(\partial \bar{u} / \partial y) \ll (\partial \bar{u} / \partial x) + (\partial \bar{u} / \partial y)$. Substituting the displacement components Eq. 1 to the

strain–displacement relations Eq. 2, the strain components are written as

$$\varepsilon_{xx} = u' + z\theta'_y - y\theta'_z + \frac{1}{2}(v'^2 + w'^2) \quad (3a)$$

$$\gamma_{xy} = v' - \theta_z \quad (3b)$$

$$\gamma_{xz} = w' + \theta_y \quad (3c)$$

It is noted that within the Euler–Bernoulli theory context, it holds that $\theta_z = v'$, $\theta_y = -w'$ (the cross section remains perpendicular to the deformed axis); hence, shear strain components Eqs. 3b and 3c vanish. Considering strains to be small and employing the second Piola–Kirchhoff stress tensor, the nonvanishing stress components are defined in terms of the strain ones as

$$\begin{Bmatrix} S_{xx} \\ S_{xy} \\ S_{xz} \end{Bmatrix} = \begin{bmatrix} E^* & 0 & 0 \\ 0 & G & 0 \\ 0 & 0 & G \end{bmatrix} \begin{Bmatrix} \varepsilon_{xx} \\ \gamma_{xy} \\ \gamma_{xz} \end{Bmatrix} \quad (4)$$

where E^* is obtained from Hooke's stress–strain law as $E^* = E(1 - \nu)/(1 + \nu)(1 - 2\nu)$. If the assumption of plane stress condition is made, the above expression is reduced in $E^* = E/(1 - \nu^2)$ (Vlasov 1963), while in beam formulations, E is frequently considered instead of E^* ($E^* \approx E$). This last consideration has been followed throughout this study, while any other reasonable value of E^* could be used without any difficulty. Substituting Eq. 3 into Eq. 4, the nonvanishing stress components are obtained as

$$S_{xx} = E \left[u' + z\theta'_y - y\theta'_z + \frac{1}{2}(v'^2 + w'^2) \right] \quad (5a)$$

$$S_{xy} = G \cdot (v' - \theta_z) \quad (5b)$$

$$S_{xz} = G \cdot (w' + \theta_y) \quad (5c)$$

In order to establish the nonlinear equations of motion, the principle of virtual work

$$\delta W_{\text{int}} + \delta W_{\text{mass}} = \delta W_{\text{ext}} \quad (6)$$

where

$$\delta W_{\text{int}} = \int_V (S_{xx}\delta\varepsilon_{xx} + S_{xy}\delta\gamma_{xy} + S_{xz}\delta\gamma_{xz})dV \quad (7a)$$

$$\delta W_{\text{mass}} = \int_V \rho(\ddot{u}\delta u + \ddot{v}\delta v + \ddot{w}\delta w)dV \quad (7b)$$

$$\delta W_{\text{ext}} = \int_F (t_x\delta\bar{u} + t_y\delta\bar{v} + t_z\delta\bar{w})dF \quad (7c)$$

under a total Lagrangian formulation is employed. In the above equations, $\delta(\cdot)$ and $(\cdot)$ denote virtual quantities and derivatives with respect to time, respectively, V is the volume of the beam in the initial (undeformed) configuration, F is the lateral surface of the beam in the initial configuration including the end cross sections, and t_x , t_y , t_z are the components of the traction vector with respect to the undeformed surface of the beam. Moreover, the stress resultants of the beam are given as

$$N = \int_{\Omega} S_{xx}dx \quad (8a)$$

$$M_y = \int_{\Omega} S_{xx}zdx \quad (8b)$$

$$M_z = - \int_{\Omega} S_{xx}ydx \quad (8c)$$

$$Q_y = \int_{\Omega} S_{xy}dx \quad (8d)$$

$$Q_z = \int_{\Omega} S_{xz}dx \quad (8e)$$

Substituting the expressions of the stress components Eq. 5 into Eq. 8, the stress resultants are obtained as

$$N = EA \left[u' + \frac{1}{2}(v'^2 + w'^2) \right] \quad (9a)$$

$$M_y = EI_y \theta'_y \quad (9b)$$

$$M_z = EI_z \theta'_z \quad (9c)$$

$$Q_y = GA_y \gamma_{xy} = GA_y (v' - \theta_z) \quad (9d)$$

$$Q_z = GA_z \gamma_{xz} = GA_z (w' + \theta_y) \quad (9e)$$

where A , I_y , I_z are the cross-sectional area and the principal moments of inertia of the cross section given as

$$A = \int_{\Omega} dx \quad (10a)$$

$$I_y = \int_{\Omega} z^2 dx \quad (10b)$$

$$I_z = \int_{\Omega} y^2 dx \quad (10c)$$

and GA_y , GA_z are the shear rigidities of the Timoshenko theory, where

$$A_y = \kappa_y A = \frac{1}{a_y} A \quad (11a)$$

$$A_z = \kappa_z A = \frac{1}{a_z} A \quad (11b)$$

are the shear areas with respect to y , z axes, respectively, with κ_y , κ_z being the shear correction factors and a_y , a_z being the shear deformation coefficients.

Employing Eq. 8 and substituting Eqs. 1, 3, and 9 to 6, the differential equations of motion are derived as

$$-EA(u'' + w'w'' + v'v'') + \rho \zeta \ddot{u} = p_x \quad (12a)$$

$$-(Nv')' + \rho \zeta \ddot{v} - GA_y(v'' - \theta'_z) = p_y \quad (12b)$$

$$-EI_z \theta''_z + \rho I_z \ddot{\theta}_z - GA_y(v' - \theta_z) = m_z \quad (12c)$$

$$-(Nw')' + \rho \zeta \ddot{w} - GA_z(w'' + \theta'_y) = p_z \quad (12d)$$

$$-EI_y \theta''_y + \rho I_y \ddot{\theta}_y + GA_z(w' + \theta_y) = m_y \quad (12e)$$

Combining Eqs. 12b and 12c with 12d and 12e in order to eliminate θ_y , θ_z , the following differential equations are derived

$$\begin{aligned} EI_z v'''' - \left(\frac{EI_z}{G \zeta_y} \rho \zeta + \rho I_z \right) \ddot{v}'' + \rho \zeta \ddot{v} + \frac{EI_z}{GA_y} (Nv')''' - (Nv')' \\ - \frac{\rho I_z}{GA_y} \left(\frac{\partial^2 (Nv')'}{\partial t^2} - \rho A \ddot{v} \right) = p_y - \frac{EI_z}{GA_y} p_y'' + \frac{\rho I_z}{GA_y} \ddot{p}_y - m'_z \end{aligned} \quad (13a)$$

$$\begin{aligned} EI_y w'''' - \left(\frac{EI_y}{G \zeta_z} \rho A + \rho I_y \right) \ddot{w}'' + \rho \zeta \ddot{w} + \frac{EI_y}{GA_z} (Nw')''' - (Nw')' \\ - \frac{\rho I_y}{GA_z} \left(\frac{\partial^2 (Nw')'}{\partial t^2} - \rho A \ddot{w} \right) = p_z - \frac{EI_y}{GA_z} p_z'' + \frac{\rho I_y}{GA_z} \ddot{p}_z + m'_y \end{aligned} \quad (13b)$$

In Eqs. 12a and 13, the quantities p_x , p_y , p_z , m_z , m_y (external axial loading, transverse loading, and bending moments along each direction, respectively) constitute the externally applied time-dependent loading along the beam (Fig. 2a) and arise (through the principle of virtual work) from the externally applied traction vectors t_x , t_y , t_z . Ignoring time derivatives of fourth order from the above equations (Thomson 1981), the following governing differential equations are derived as

$$-EA(u'' + w'w'' + v'v'') + \rho \zeta \ddot{u} = p_x \quad (14a)$$

$$\begin{aligned} EI_z v'''' - \rho I_z \left(\frac{Ea_y}{G} + 1 \right) \ddot{v}'' + \rho \zeta \ddot{v} + \frac{EI_z}{GA_y} (Nv')''' - (Nv')' \\ - \frac{\rho I_z}{GA_y} \frac{\partial^2 (Nv')'}{\partial t^2} = p_y - \frac{EI_z}{GA_y} p_y'' + \frac{\rho I_z}{GA_y} \ddot{p}_y - m'_z \end{aligned} \quad (14b)$$

$$\begin{aligned} EI_y w'''' - \rho I_y \left(\frac{Ea_z}{G} + 1 \right) \ddot{w}'' + \rho \zeta \ddot{w} + \frac{EI_y}{GA_z} (Nw')''' - (Nw')' \\ - \frac{\rho I_y}{GA_z} \frac{\partial^2 (Nw')'}{\partial t^2} = p_z - \frac{EI_y}{GA_z} p_z'' + \frac{\rho I_y}{GA_z} \ddot{p}_z + m'_y \end{aligned} \quad (14c)$$

The stress resultants and bending rotations are expressed as

$$M_z = EI_z v'' + \frac{EI_z}{GA_y} [p_y - \rho A \ddot{v} + (Nv')'] \quad (15a)$$

$$M_y = -EI_y w'' + \frac{EI_y}{GA_z} [\rho A \ddot{w} - p_z - (Nw')'] \quad (15b)$$

$$Q_y = -EI_z v''' - \frac{EI_z}{GA_y} \left[p'_y - \rho A \ddot{v}' + (Nv')'' \right] + \rho I_z \ddot{\theta}_z - m_z \quad (15c)$$

$$Q_z = -EI_y w''' + \frac{EI_y}{GA_z} \left[\rho A \ddot{w}' - p'_z - (Nw')'' \right] - \rho I_y \ddot{\theta}_y + m_y \quad (15d)$$

$$\theta_z = v' - \frac{Q_y}{GA_y} \quad (15e)$$

$$\theta_y = -w' + \frac{Q_z}{GA_z} \quad (15f)$$

The above equations are also subjected to the pertinent boundary conditions of the problem which are given as

$$a_1 u(x, t) + \alpha_2 N(x) = \alpha_3 \quad (16a)$$

$$\beta_1 v(x, t) + \beta_2 R_y^b(x) = \beta_3 \quad (16b)$$

$$\bar{\beta}_1 \theta_z(x, t) + \bar{\beta}_2 M_z^b(x) = \bar{\beta}_3 \quad (16c)$$

$$\gamma_1 w(x, t) + \gamma_2 R_z^b(x) = \gamma_3 \quad (16d)$$

$$\bar{\gamma}_1 \theta_y(x, t) + \bar{\gamma}_2 M_y^b(x) = \bar{\gamma}_3 \quad (16e)$$

at the beam ends $x = 0, L$, where $a_i, \beta_i, \gamma_i, \bar{\beta}_i, \bar{\gamma}_i$ $i = 1, 2, 3$ are known time-dependent functions. The boundary conditions Eq. 16 are the most general boundary conditions for the problem at hand, including also the elastic support. It is apparent that all types of the conventional boundary conditions (clamped, simply supported, free, or guided edge) can be derived from Eq. 16 by specifying appropriately the functions $a_i, \beta_i, \gamma_i, \bar{\beta}_i, \bar{\gamma}_i$ (e.g., for a clamped edge, it is $\alpha_1 = \beta_1 = \bar{\beta}_1 = \gamma_1 = \bar{\gamma}_1 = 1, \alpha_2 = \alpha_3 = \beta_2 = \beta_3 = \gamma_2 = \gamma_3 = \bar{\beta}_2 = \bar{\beta}_3 = \bar{\gamma}_2 = \bar{\gamma}_3 = 0$). In Eq. 16, the reaction forces R_y^b, R_z^b ; bending moments M_z^b, M_y^b ; and the angles of bending rotations θ_y, θ_z at the beam ends ignoring the inertia terms are written as

$$R_y^b = Nv' - EI_z v''' - \frac{EI_z}{GA_y} Nv''' \quad (17a)$$

$$R_z^b = Nw' - EI_y w''' - \frac{EI_y}{GA_z} Nw''' \quad (17b)$$

$$M_z^b = EI_z v'' + \frac{EI_z}{GA_y} Nv'' \quad (17c)$$

$$M_y^b = -EI_y w'' - \frac{EI_y}{GA_z} Nw'' \quad (17d)$$

$$\theta_y^b = -w' - \frac{EI_y}{G^2 A_z^2} Nw''' - \frac{EI_y}{GA_z} w''' \quad (17e)$$

$$\theta_z^b = v' + \frac{EI_z}{G^2 A_y^2} Nv''' + \frac{EI_z}{GA_y} v''' \quad (17f)$$

In case of nonzero initial conditions, the following relations are valid for $t = 0$

$$u(x, 0) = \bar{u}_0(x) \quad (18a)$$

$$\dot{u}(x, 0) = \dot{\bar{u}}_0(x) \quad (18b)$$

$$v(x, 0) = \bar{v}_0(x) \quad (18c)$$

$$\dot{v}(x, 0) = \dot{\bar{v}}_0(x) \quad (18d)$$

$$w(x, 0) = \bar{w}_0(x) \quad (18e)$$

$$\dot{w}(x, 0) = \dot{\bar{w}}_0(x) \quad (18f)$$

$$\theta_z(x, 0) = \bar{\theta}_{z0}(x) \quad (18g)$$

$$\dot{\theta}_z(x, 0) = \dot{\bar{\theta}}_{z0}(x) \quad (18h)$$

$$\theta_y(x, 0) = \bar{\theta}_{y0}(x) \quad (18i)$$

$$\dot{\theta}_y(x, 0) = \dot{\bar{\theta}}_{y0}(x) \quad (18j)$$

Applications

Free Vibrations of Hinged Beam of Rectangular Cross Section

In order to examine the influence of shear deformation on the free vibrations of beams in the

Nonlinear Dynamic Seismic Analysis, Table 1 Material properties, geometric, inertia constants, and shear deformation coefficients of the rectangular cross section of application section “Free Vibrations of Hinged Beam of Rectangular Cross Section”

$E = 70 \text{ GPa}$	$G = 27 \text{ GPa}$
$\rho = 2.7 \text{ tn/m}^3$	$A = 150 \times 10^{-6} \text{ m}^2$
$I_y = 7,812.5 \times 10^{-12} \text{ m}^4$	$I_z = 450 \times 10^{-12} \text{ m}^4$
$a_y = 1.74$	$a_z = 1.20$

geometrically nonlinear range, a hinged beam of a $25 \times 6 \text{ mm}$ rectangular cross section subjected to free vibrations is studied. It is worth here mentioning that the axially immovable ends lead to the development of axial loading along the beam as deflections increase. The employed initial conditions are represented by the following expression $w(x, 0) = w_{\max} \sin(\pi x/L)$ where $w_{\max}$ is the central beam deflection. In Table 1, the material properties, the geometric and inertia constants, and the shear deformation coefficients of the examined cross section are presented.

In Fig. 3, the obtained results of the period ratios T/T_0 , at various amplitude ratios $w_{\max}/r$ for three different cases of beam slenderness L/r , as obtained from the numerical solution of the initial-boundary value problem Eqs. 14, 16, and 18 using analog equation method (AEM) and boundary element method (BEM)-based techniques (Sapountzakis and Dourakopoulos 2009a, b), are presented taking into account or ignoring shear deformation effect. T denotes the period of free vibration obtained by nonlinear analysis including rotary inertia and taking into account or ignoring shear deformation effect, T_0 is the corresponding period of the linear vibration ignoring both rotary inertia and shear deformation effect, and $r = \sqrt{I_y/A}$ is the radius of gyration about y axis. The obtained results are compared with the corresponding ones available in the study of Foda (1999) obtained by a multiple scales procedure. In Table 2 the ratios of the nonlinear frequency ω_{NL} over the linear one ω_L at various amplitude ratios $w_{\max}/r$ and slenderness L/r , obtained by AEM, are presented as compared with those obtained from a finite element solution employing polynomial expressions for the displacement components (Rao et al. 1976).

From this table the convergence of both methods can be observed. Moreover, from the above figure and table, the increment of vibration frequency with the increment of deflections can be observed due to the fact that axially immovable ends induce tensile forces to the beam, thus increasing its stiffness.

Free Vibrations of Beam with Very Flexible Boundary Conditions

In this application the free vibrations of Timoshenko beams with very flexible boundary conditions is examined, since in this case the natural frequencies and the corresponding modeshapes are highly sensitive to the effects of shear deformation and rotary inertia (Aristizabal-Ochoa 2004). Thus, the linear free vibration analysis of pinned-free beams of length $L = 5.0 \text{ m}$ of various cross sections is examined ($\rho = 2.40 \text{ tn/m}^3$, $E = 25.42 \text{ GPa}$, $G = 11.05 \text{ GPa}$). In Table 3 the geometric and inertia constants as well as the shear correction factors of the examined cross sections are presented ($\kappa = 1/\alpha$). The differential equations for the linear free vibrations of this special case can be obtained by Eq. 14 for zero axial stress resultant and external force as

$$EI_z v'''' - \rho I_z \left(\frac{Ea_y}{G} + 1 \right) \ddot{v}'' + \rho \zeta \ddot{v} = 0 \quad (19a)$$

$$EI_y w'''' - \rho I_y \left(\frac{Ea_z}{G} + 1 \right) \ddot{w}'' + \rho A \ddot{w} = 0 \quad (19b)$$

The boundary conditions for the examined case can be derived by Eq. 16 as

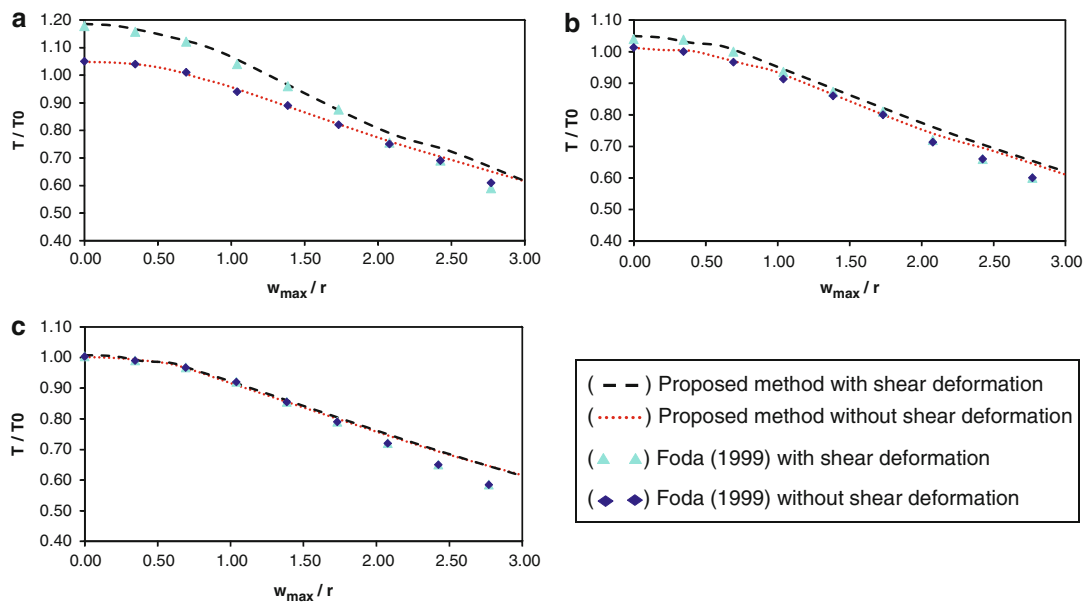
$$v(0, t) = w(0, t) = 0 \quad (20a)$$

$$M_y(0, t) = M_z(0, t) = 0 \quad (20b)$$

$$V_y(L, t) = V_z(L, t) = 0 \quad (20c)$$

$$M_y(L, t) = M_z(L, t) = 0 \quad (20d)$$

In Table 4 the first three natural frequencies of the pinned-free Timoshenko beams of various cross sections and in Fig. 4 the corresponding modeshapes for the cross section with web



Nonlinear Dynamic Seismic Analysis, Fig. 3 Free vibration period ratios T/T_0 at various amplitude ratios w_{max}/r for beam slenderness $l/r = 50$ of the simply

supported beam of application section “Free Vibrations of Hinged Beam of Rectangular Cross Section”

Nonlinear Dynamic Seismic Analysis, Table 2 Frequency ratios ω_{NL}/ω_L at various amplitude ratios w_{max}/r of the simply supported beam of application section “Free Vibrations of Hinged Beam of Rectangular Cross Section”

w_{max}/r	$L/r = 10$		$L/r = 50$		$L/r = 100$	
	AEM	Rao et al. (1976)	AEM	Rao et al. (1976)	AEM	Rao et al. (1976)
0.4	1.0214	1.0193	1.0198	1.0154	1.0103	1.0149
0.8	1.0694	1.0737	1.0587	1.0585	1.0551	1.0581
1.0	1.1116	1.1116	1.0928	1.0897	1.0920	1.0890
2.0	1.4296	1.3701	1.3230	1.3143	1.3227	1.3125
3.0	1.9268	1.6884	1.6432	1.6052	1.6410	1.6030

thickness $t_w = 0.254$ mm are presented as obtained from an AEM solution (Sapountzakis and Dourakopoulos 2009a, b) and semi-analytical methods employing sinusoidal functions (Aristizabal-Ochoa 2004; Kausel 2002).

Forced Vibrations or Beam of Hollow Rectangular Cross Section Under Biaxial Bending
In order to examine the influence of shear deformation on the free vibrations of beams in the geometrically nonlinear range, a clamped beam

Nonlinear Dynamic Seismic Analysis, Table 3 Geometric constants of the pinned-free beam of application section “Free Vibrations or Beam with Very Flexible Boundary Conditions”

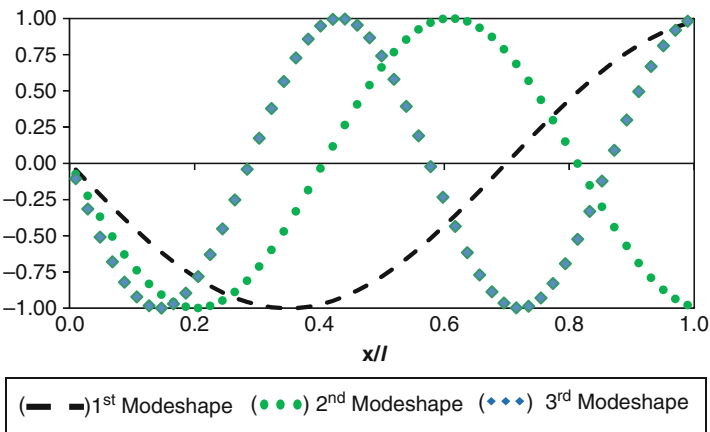
Web thickness t_w (mm)	Cross-sectional area A (m ²)	Moment of inertia I (m ⁴)	Shear correction factor $\kappa = 1/a$
101.6	0.359	0.193	0.54
25.4	0.229	0.162	0.21
7.62	0.199	0.154	0.07
0.254	0.186	0.151	0.002

Nonlinear Dynamic Seismic Analysis, Table 4 Natural frequencies $f = \omega/2\pi$ (Hz) of the beam of application section “Free Vibrations or Beam with Very Flexible Boundary Conditions,” for various cross sections

t_w (mm)	f_1			f_2	f_3
	Aristizabal-Ochoa (2004)	Kausel (2002)	AEM	AEM	AEM
101.6	148.2	218	153.9	315.0	467.6
25.4	116.8	133.4	120.6	222.9	120.6
7.62	73.4	76.4	77.0	135.8	193.0
0.254	12.8	12.8	13.7	23.6	33.3

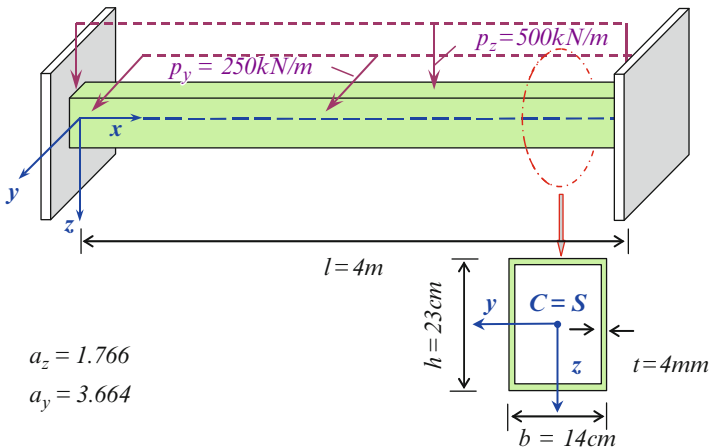
Nonlinear Dynamic Seismic Analysis,

Fig. 4 Three first vibration modeshapes of the beam of application section “Free Vibrations or Beam with Very Flexible Boundary Conditions,” for $t_w = 0.254$ mm



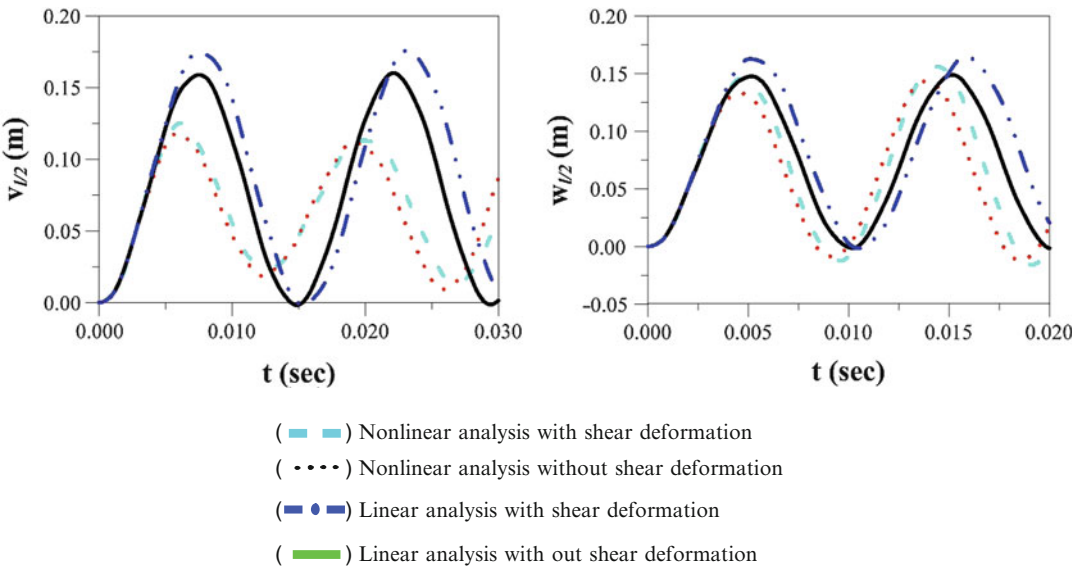
Nonlinear Dynamic Seismic Analysis,

Fig. 5 Clamped beam of hollow rectangular cross section of application section “Forced Vibrations or Beam of Hollow Rectangular Cross Section Under Biaxial Bending”



($E = 210$ GPa, $\nu = 0.3$, $\rho = 7.85$ tn/m³) of length $L = 4.0$ m and of a hollow rectangular cross section, subjected to the suddenly applied uniformly distributed transverse loadings $p_y(t) = 250$ kN/m, $p_z(t) = 500$ kN/m, $t \geq 0$,

(Fig. 5) is examined. In Fig. 6, the time histories of central displacements $v_{L/2}$, $w_{L/2}$, taking into account or ignoring shear deformation effect, are presented. Moreover, in Table 5, the maximum values of the central deflections $v_{l/2}$, $w_{l/2}$



Nonlinear Dynamic Seismic Analysis, Fig. 6 Time histories of central deflections along directions $y(v_{l/2})$, $z(w_{l/2})$ of the beam of application section “[Forced Vibrations or](#)

[Beam of Hollow Rectangular Cross Section Under Biaxial Bending](#)”

Nonlinear Dynamic Seismic Analysis, Table 5 Maximum central deflections $v_{\max}(\text{m})$, $w_{\max}(\text{m})$ and periods $T_y, T_z(\text{s})$ of the first cycle of the clamped beam of application section “[Forced Vibrations or Beam of Hollow Rectangular Cross Section Under Biaxial Bending](#)”

	Without shear deformation		With shear deformation	
	Linear analysis	Nonlinear analysis	Linear analysis	Nonlinear analysis
$v_{\max}$	0.1588	0.1180	0.1739	0.1252
$w_{\max}$	0.1476	0.1330	0.1626	0.1460
T_y	0.01483	0.01229	0.01532	0.01240
T_z	0.01019	0.00911	0.01051	0.00958

and the periods T_y, T_z of the first cycle are presented for the aforementioned cases of analysis, taking into account or ignoring shear deformation effect. From this figure and table, the influence of geometrical nonlinearity can be observed reducing the magnitude of deflections and the period of vibrations due to axially immovable ends which lead to the development of axial forces, thus increasing the stiffness of the beam. Shear deformation effect is also

apparent increasing both the maximum central transverse displacements (especially in z direction) and the calculated periods of the first cycle of motion.

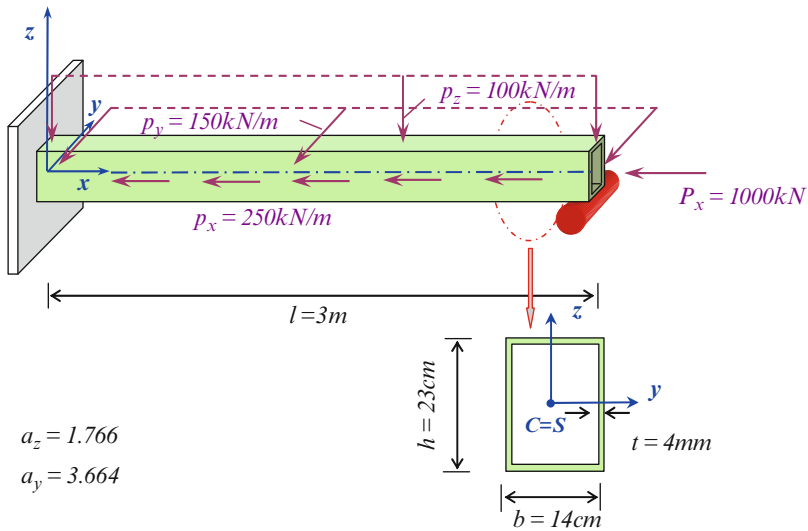
Forced Vibrations or Beam of Hollow Rectangular Cross Section Subjected to Axial Compressive Load

In order to further investigate the influence of geometrical nonlinearity on the response of Timoshenko beams, the nonlinear dynamic analysis of a clamped-rolled beam of length $l = 3.0$ m, having a hollow rectangular cross section ($E = 210$ GPa, $\nu = 0.3$, $\rho = 7.85 \text{ tn/m}^3$) subjected to the suddenly applied uniformly distributed loads $p_x(t) = -250$ kN/m, $p_y(t) = -150$ kN/m, $p_z(t) = -150$ kN/m and concentrated axial load $P_x(t) = -500$ kN at the rolled end ($t \geq 0$), as is shown in Fig. 7, has been studied.

In Fig. 8, the time histories of the central transverse displacements $v_{l/2}, w_{l/2}$ of the beam performing either a linear or a nonlinear analysis are presented taking into account or ignoring shear deformation effect. Moreover, in Table 6 the maximum values of the central deflections

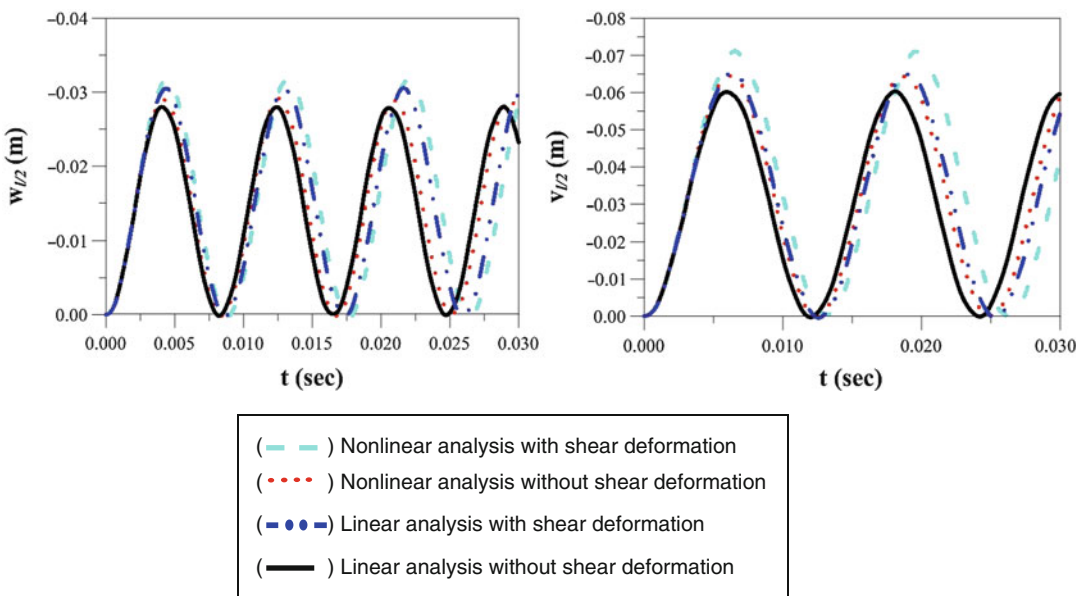
Nonlinear Dynamic Seismic Analysis,

Fig. 7 Clamped-rolled beam of hollow rectangular cross section of application section “Forced Vibrations or Beam of Hollow Rectangular Cross Section Subjected to Axial Compressive Load”



$a_z = 1.766$

$a_y = 3.664$



Nonlinear Dynamic Seismic Analysis, Fig. 8 Time histories of central deflections along directions $y(v_{L/2})$, $z(w_{L/2})$ of the beam of application section “Forced Vibrations or

Beam of Hollow Rectangular Cross Section Subjected to Axial Compressive Load”

$w_{\max}$, $w_{\min}$ and the periods T_y , T_z of the first cycle are presented for the aforementioned cases of analysis, including or ignoring shear deformation effect. In all of the aforementioned cases, rotary inertia has been taken into account. From the obtained results, the influence of

geometrical nonlinearity is apparent, increasing the magnitude of kinematical components due to compressive loading, while the influence of shear deformation effect is also intense leading to further increment of the kinematical components.

Nonlinear Torsional Dynamic Analysis of Beams with Secondary Torsional Moment Deformation Effect (STMDE)

Statement of the Problem

The prismatic beam of doubly symmetric cross section described in section “Statement of the Problem” is examined and is subjected to the combined action of the arbitrarily distributed or concentrated time-dependent conservative axial load $p_x(x, t)$, twisting $m_t = m_t(x, t)$, and warping $m_w = m_w(x, t)$ moments acting in the x direction (Fig. 9a), while its ends are subjected to the most general boundary conditions with respect to the axial displacement, the angle of twist, and warping intensity. The beam can be twisted freely

about the longitudinal axis and its flexural behavior is not examined. The analysis is carried out with respect to the coordinate system $Sxyz$, the longitudinal axis Sx of which passes through the cross sections’ shear center. It is noted that due to the fact that the cross section is a doubly symmetric one, the shear center of the cross section coincides with the centroid (Fig. 9b).

Assuming that the cross section maintains its shape during deformation (no distortional effects), considering that the points of the cross section follow a circular trajectory around axis Sx , and considering twisting rotations to be large and taking into account STMDE effect, the displacement field of the beam is written as

$$\begin{aligned} \bar{u}(x, y, z, t) = & u(x, t) + (\theta_x^P(x, t))' \phi_S^P(y, z) \\ & - [(\theta_x^P(x, t))' - \theta_x'(x, t)] [\phi_S^P(y, z) + \phi_S^S(x, y, z, t)] \end{aligned} \quad (21a)$$

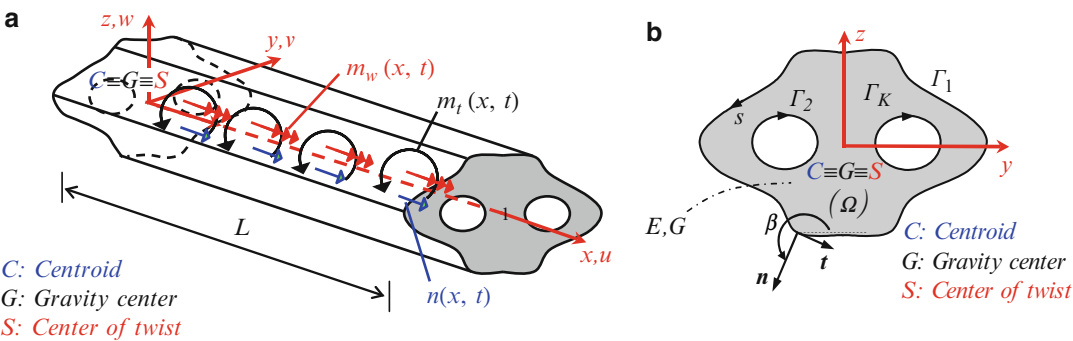
$$\bar{v}(x, y, z, t) = -z \sin \theta_x(x, t) - y(1 - \cos \theta_x(x, t)) \quad (21b)$$

$$\bar{w}(x, y, z, t) = y \sin \theta_x(x, t) - z(1 - \cos \theta_x(x, t)) \quad (21c)$$

In Eq. 21, apart from the “average” axial displacement $u(x, t)$ of the cross section defined in section “Nonlinear Flexural Dynamic Analysis of Beams with Shear Deformation Effect,” $\theta_x^P(x, t)$, $\theta_x(x, t)$ denote the primary part and the total angle

Nonlinear Dynamic Seismic Analysis, Table 6 Periods $T_y, T_z(s)$ of the first cycle and maximum central displacements $v_{\max}(m), w_{\max}(m)$ of the beam of application section “Forced Vibrations of Beam of Hollow Rectangular Cross Section Subjected to Axial Compressive Load”

	Without shear deformation		With shear deformation	
	Linear analysis	Nonlinear analysis	Linear analysis	Nonlinear analysis
$v_{\max}$	-2.90×10^{-2}	-3.23×10^{-2}	-3.25×10^{-2}	-3.68×10^{-2}
$w_{\max}$	-1.35×10^{-2}	-1.41×10^{-2}	-1.51×10^{-2}	-1.60×10^{-2}
T_y	8.17×10^{-3}	8.75×10^{-3}	9.03×10^{-3}	9.46×10^{-3}
T_z	5.74×10^{-3}	5.86×10^{-3}	6.20×10^{-3}	6.36×10^{-2}



Nonlinear Dynamic Seismic Analysis, Fig. 9 Prismatic beam of an arbitrarily shaped doubly symmetric constant cross section occupying region Ω (a) subjected to axial and torsional loading (b)

of twist, respectively; ϕ_S^P , ϕ_S^S are the primary and secondary warping functions with respect to the shear center S . Considering the “average” axial displacement and the warping displacements of the cross section to be small, strain components Eq. 2 and stress–strain relationship Eq. 4 are once again employed to derive the nonvanishing stress components of the second Piola–Kirchhoff stress tensor as

$$S_{xx} = E \left[u' + (\theta_x^P)'' \phi_S^P + \frac{1}{2} (y^2 + z^2) (\theta_x')^2 \right] \quad (22a)$$

$$S_{xy} = \underbrace{G \theta_x' \left(\frac{\partial \phi_S^P}{\partial y} - z \right)}_{\text{primary}} + \underbrace{G \left((\theta_x^P)' - \theta_x' \right) \left(-\frac{\partial \phi_S^S}{\partial y} \right)}_{\text{secondary}} \quad (22b)$$

$$S_{xz} = \underbrace{G \theta_x' \left(\frac{\partial \phi_S^P}{\partial z} + y \right)}_{\text{primary}} + \underbrace{G \left((\theta_x^P)' - \theta_x' \right) \left(-\frac{\partial \phi_S^S}{\partial z} \right)}_{\text{secondary}} \quad (22c)$$

In Eq. 22a, the term $(y^2 + z^2) \cdot (\theta_x')^2/2$ is often described as the “Wagner strain” (Sapountzakis and Tsipiras 2010a).

Local Equilibrium Equation, Primary Torsional Warping Function

In order to establish the local equilibrium equation in the longitudinal direction, the calculus of variations is applied. The principle of virtual work under a total Lagrangian formulation is applied (Eq. 6). Taking into account the primary warping function’s ϕ_S^P virtual components and applying the procedure analyzed in the study of Sapountzakis and Tsipiras (2010a), the following local equilibrium equation:

$$\frac{\partial S_{xx}}{\partial x} + \frac{\partial S_{xy}}{\partial y} + \frac{\partial S_{xz}}{\partial z} - \rho \ddot{u} = 0 \quad \sigma \tau \Omega, \quad \forall x \in (0, L) \quad (23)$$

along with its corresponding boundary condition

$$S_{xy} n_y + S_{xz} n_z = t_x \quad \sigma \tau \Omega \quad \Gamma_j, \quad \forall x \in (0, L) \quad (24)$$

Subsequently, taking into account the secondary warping function’s ϕ_S^S virtual components and applying the same procedure, Eq. 23 can be once again derived. Hence, the two warping functions introduced in the displacement field of the beam should fulfill the local equilibrium Eq. 23 and the corresponding boundary condition Eq. 24. Employing the stress expressions (Eq. 22) and after some algebraic manipulations, it can be easily concluded that the warping functions can be established by the following boundary value problems as

$$\nabla^2 \phi_S^P = 0 \quad \sigma \tau \Omega \quad (25a)$$

$$\frac{\partial \phi_S^P}{\partial n} = (z n_y - y n_z) \quad \sigma \tau \Omega \quad \Gamma_j \quad (25b)$$

$$G \left((\theta_x^P)' - \theta_x' \right) \nabla^2 \phi_S^S - E \left[u'' + (\theta_x^P)''' \phi_S^P + (y^2 + z^2) \theta_x' \theta_x'' \right] + \rho \left[\ddot{u} + (\theta_x^P)' \phi_S^P \right]$$

$$- \left[\left((\theta_x^P)' - \theta_x' \right) (\phi_S^P + \phi_S^S) - \left((\theta_x^P)' - \theta_x' \right) \phi_S^S \right] = 0 \quad \sigma \tau \Omega, \quad \forall x \in [0, L] \quad (26a)$$

$$-G \left((\theta_x^P)' - \theta_x' \right) \frac{\partial \phi_S^S}{\partial n} = t_x \quad \sigma \tau \Omega \quad \Gamma_j, \quad \forall x \in [0, L] \quad (26b)$$

which are considered to be valid at the beam ends as well. In the above equations, $\nabla^2 = \partial^2/\partial y^2 + \partial^2/\partial z^2$ is the Laplace operator, while $\partial/\partial n = (\partial/\partial y)n_y + (\partial/\partial z)n_z$ denotes the derivative with respect to the outward normal vector $\mathbf{n}$ to the boundary

Γ_j (directional derivative). In order to simplify the boundary value problem yielding ϕ_S^S , the usual assumption of ignoring the axial traction component ($t_x \approx 0$) is adopted; hence, relation Eq. 26b is modified as

$$\frac{\partial \phi_S^S}{\partial n} = 0 \quad \sigma_{\tau\theta} \quad \Gamma_j, \quad \forall x \in [0, L] \quad (27)$$

Boundary value problem Eq. 26 can be further simplified, by ignoring the influence of geometrical nonlinearity terms and inertia terms of the secondary warping function and taking advantage of the global equation of motion for the beam which will be formulated in the next section. It can be also proved that ϕ_S^S is a two-dimensional function. Finally, in order to compute unique solutions of Eqs. 25 and 26, ϕ_S^P, ϕ_S^S are considered to fulfill the following orthogonality conditions:

$$\int_{\Omega} \phi_S^P d\Omega = 0 \quad (28a)$$

$$\int_{\Omega} \phi_S^P y d\Omega = 0 \quad (28b)$$

$$\int_{\Omega} \phi_S^P z d\Omega = 0 \quad (28c)$$

$$\int_{\Omega} \phi_S^S d\Omega = 0 \quad (28d)$$

Stress Resultants, Global Equations of Motion

The kinematical unknowns of the problem depend only on the variable x ($u, \theta_x, (\theta_x^P)'$) and will be established from the solution of global differential equations of motion, which will be formulated employing the calculus of variations. In order to facilitate the analysis, the following stress resultants are defined as

$$N = \int_{\Omega} S_{xx} d\Omega \quad (29a)$$

$$M_t^P = \int_{\Omega} \left[S_{xy} \left(\frac{\partial \phi_S^P}{\partial y} - z \right) + S_{xz} \left(\frac{\partial \phi_S^P}{\partial z} + y \right) \right] d\Omega \quad (29b)$$

$$M_t^S = \int_{\Omega} \left(S_{xy} \frac{\partial \phi_S^S}{\partial y} + S_{xz} \frac{\partial \phi_S^S}{\partial z} \right) d\Omega \quad (29c)$$

$$M_w = \int_{\Omega} S_{xx} \phi_S^P d\Omega \quad (29d)$$

where N is the axial force, M_t^P, M_t^S are the primary and secondary twisting moment (Tsipiras and Sapountzakis 2012) resulting from the primary (S_{xy}^P, S_{xz}^P) and secondary (S_{xy}^S, S_{xz}^S) shear stress distribution, respectively, and M_w is the warping moment due to the torsional curvature. Substituting Eqs. 22 into Eq. 29, the stress resultants are obtained as

$$N = EA \left[u' + \frac{1}{2} \frac{I_P}{A} (\theta_x')^2 \right] \quad (30a)$$

$$M_t^P = GI_t^P \theta_x' \quad (30b)$$

$$M_t^S = -GI_t^S \left((\theta_x^P)' - \theta_x' \right) \quad (30c)$$

$$M_w = EC_S (\theta_x^P)'' \quad (30d)$$

where $A = \int_{\Omega} d\Omega$ is the area of the cross section. The geometric constants I_P (polar moment of inertia with respect to the shear center S), I_t^P (primary torsion constant), I_t^S (secondary torsion constant), and C_S (warping constant) are given by the following definitions:

$$I_P = \int_{\Omega} (y^2 + z^2) d\Omega \quad (31a)$$

$$I_t^P = \int_{\Omega} \left(y^2 + z^2 + y \frac{\partial \phi_S^P}{\partial z} - z \frac{\partial \phi_S^P}{\partial y} \right) d\Omega \quad (31b)$$

$$I_t^S = \int_{\Omega} \left[\left(\frac{\partial \phi_S^S}{\partial y} \right)^2 + \left(\frac{\partial \phi_S^S}{\partial z} \right)^2 \right] d\Omega \quad (31c)$$

$$C_S = \int_{\Omega} (\phi_S^P)^2 d\Omega \quad (31d)$$

Applying the principle of virtual work under a total Lagrangian formulation (Eq. 6), taking into account the virtual quantities of $u, \theta_x, (\theta_x^P)'$ (and

their derivatives) and the stress components given by Eq. 22, the Euler–Lagrange equations governing the global dynamic equilibrium of the beam, together with the pertinent boundary conditions at the beam ends $x = 0, L$, are expressed as

$$-\rho A \ddot{u} + \frac{dN}{dx} = -p_x(x, t) \quad (32a)$$

$$\begin{aligned} -\rho I_P \ddot{\theta}_x + \frac{dM_t^P}{dx} + \frac{dM_t^S}{dx} + \frac{d}{dx} \left[\frac{1}{2} EI_n (\theta'_x)^3 + N \frac{I_P}{A} \theta'_x \right] \\ = -m_t(x, t) \end{aligned} \quad (32b)$$

$$-\rho C_S (\ddot{\theta}_x^P)' + \frac{dM_w}{dx} + M_t^S = -m_w(x, t) \quad (32c)$$

$$(N + \bar{N}_0) \delta u(0) = 0 \quad (33a)$$

$$(N - \bar{N}_L) \delta u(L) = 0 \quad (33b)$$

$$\begin{aligned} \left[M_t^P + M_t^S + \frac{1}{2} EI_n (\theta'_x)^3 \right. \\ \left. + N \frac{I_P}{A} \theta'_x + \bar{M}_{t0} \right] \delta \theta_x(0) = 0 \end{aligned} \quad (33c)$$

$$\begin{aligned} \left[M_t^P + M_t^S + \frac{1}{2} EI_n (\theta'_x)^3 \right. \\ \left. + N \frac{I_P}{A} \theta'_x - \bar{M}_{tL} \right] \delta \theta_x(L) = 0 \end{aligned} \quad (33d)$$

$$(M_w + \bar{M}_{w0}) \delta (\theta_x^P)'(0) = 0 \quad (33e)$$

$$(M_w - \bar{M}_{wL}) \delta (\theta_x^P)'(L) = 0 \quad (33f)$$

where $I_n = I_{PP} - I_P^2/A$ and $I_{PP} = \int_{\Omega} (y^2 + z^2)^2 d\Omega$ is the fourth moment of inertia with respect to the shear center S (coinciding with the centroid). In all expressions of Eq. 33, the quantities p_x, m_t, m_w (external axial loading, twisting, and warping moment, respectively) constitute the externally applied time-dependent loading along the beam (Fig. 9a) and arise (through the principle of virtual work) from the externally applied traction vectors t_x, t_y, t_z . The quantities $\bar{N}_0, \bar{N}_L, \bar{M}_{t0}, \bar{M}_{tL}, \bar{M}_{w0}, \bar{M}_{wL}$ (appearing in boundary conditions Eqs. 33a and 33b) constitute externally applied reactions

(axial force, twisting moment, and warping moment, respectively) at the beam end cross sections and are given as

$$\bar{N}_0 = \int_{\Omega_0} t_x d\Omega \quad (34a)$$

$$\bar{N}_L = \int_{\Omega_L} t_x d\Omega \quad (34b)$$

$$\begin{aligned} \bar{M}_{t0} = \int_{\Omega_0} [t_y(-z \cos \theta_x - y \sin \theta_x) \\ + t_z(y \cos \theta_x - z \sin \theta_x)] d\Omega \end{aligned} \quad (34c)$$

$$\begin{aligned} \bar{M}_{tL} = \int_{\Omega_L} [t_y(-z \cos \theta_x - y \sin \theta_x) \\ + t_z(y \cos \theta_x - z \sin \theta_x)] d\Omega \end{aligned} \quad (34d)$$

$$\bar{M}_{w0} = \int_{\Omega_0} t_x \phi_S^P d\Omega \quad (34e)$$

$$\bar{M}_{wL} = \int_{\Omega_L} t_x \phi_S^P d\Omega \quad (34f)$$

From Eqs. 32b, 33c, and 33d, it can be concluded that axial and torsional stress resultants are coupled and cannot be studied independently, which is the case in linear analysis. Introducing the expressions of stress resultants Eq. 30 into Eqs. 32 and 33, the following governing differential equations of motion are obtained as

$$-\rho A \ddot{u} + EA u'' + EI_P \theta'_x \theta''_x = -p_x(x, t) \quad (35a)$$

$$\begin{aligned} -\rho I_P \ddot{\theta}_x + G(I_t^P + I_t^S) \theta''_x - GI_t^S (\theta_x^P)'' \\ + \frac{3}{2} EI_{PP} (\theta'_x)^2 \theta''_x + EI_P u' \theta''_x \\ + EI_P u'' \theta'_x = -m_t(x, t) \end{aligned} \quad (35b)$$

$$\begin{aligned} -\rho C_S (\ddot{\theta}_x^P)' + EC_S (\theta_x^P)''' - GI_t^S ((\theta_x^P)' - \theta'_x) \\ = -m_w(x, t) \end{aligned} \quad (35c)$$

subjected to the initial conditions ($x \in (0, L)$)

$$u(x, 0) = \bar{u}_0(x) \quad (36a)$$

$$\dot{u}(x, 0) = \dot{\bar{u}}_0(x) \quad (36b)$$

$$\theta_x(x, 0) = \bar{\theta}_{x0}(x) \quad (36c)$$

$$\dot{\theta}_x(x, 0) = \dot{\bar{\theta}}_{x0}(x) \quad (36d)$$

$$(\theta_x^P)'(x, 0) = (\bar{\theta}_{x0}^P)'(x) \quad (36e)$$

$$(\dot{\theta}_x^P)'(x, 0) = (\dot{\bar{\theta}}_{x0}^P)'(x) \quad (36f)$$

and to the most general boundary conditions

$$\alpha_1 N + \alpha_2 u = \alpha_3 \quad (37a)$$

$$\beta_1 M_t + \beta_2 \theta_x = \beta_3 \quad (37b)$$

$$\bar{\beta}_1 M_w + \bar{\beta}_2 (\theta_x^P)' = \bar{\beta}_3 \quad (37c)$$

where N , M_w are the axial force and the warping moment at the beam ends given by relations Eqs. 30a and 30d, respectively, while M_t is the twisting moment at the beam ends given as

$$M_t = G(I_t^P + I_t^S)\theta_x' - GI_t^S(\theta_x^P)' + \frac{1}{2}EI_{PP}(\theta_x')^3 + EI_P u' \theta_x' \quad (38)$$

and a_i , β_i , $\bar{\beta}_i$ ($i = 1, 2, 3$) are time-dependent functions specified at the boundary of the beam defining any possible support condition as described in section “Statement of the Problem.” In the above differential equations, viscous damping can be also added without any difficulty. Finally, it is noted that in Eq. 35b, the derivatives of secondary torsion constant with respect to x have been ignored since it can be proved that it is independent of the axial coordinate.

In Eq. 35c, the third derivative of the primary angle of twisting rotation can be observed, thus making the solution of the system more complicated. In order to remove this term, the independent warping parameter $\eta_x = (\theta_x^P)'$ is introduced, and Eqs. 35, 36, 37, and 38 are expressed as

$$-\rho A \ddot{u} + EA u'' + EI_P \theta_x' \theta_x'' = -p_x(x, t) \quad (39a)$$

$$\begin{aligned} & -\rho I_P \ddot{\theta}_x + G(I_t^P + I_t^S)\theta_x'' - GI_t^S \eta_x' \\ & + \frac{3}{2}EI_{PP}(\theta_x')^2 \theta_x'' + EI_P u' \theta_x'' + EI_P u'' \theta_x' \\ & = -m_t(x, t) \end{aligned} \quad (39b)$$

$$-\rho C_S \ddot{\eta}_x + EC_S \eta_x'' - GI_t^S(\eta_x - \theta_x') = -m_w(x, t) \quad (39c)$$

$$u(x, 0) = \bar{u}_0(x) \quad (40a)$$

$$\dot{u}(x, 0) = \dot{\bar{u}}_0(x) \quad (40b)$$

$$\theta_x(x, 0) = \bar{\theta}_{x0}(x) \quad (40c)$$

$$\dot{\theta}_x(x, 0) = \dot{\bar{\theta}}_{x0}(x) \quad (40d)$$

$$\eta_x(x, 0) = \bar{\eta}_{x0}(x) \quad (40e)$$

$$\dot{\eta}_x(x, 0) = \dot{\bar{\eta}}_{x0}(x) \quad (40f)$$

$$\alpha_1 N + \alpha_2 u = \alpha_3 \quad (41a)$$

$$\beta_1 M_t + \beta_2 \theta_x = \beta_3 \quad (41b)$$

$$\bar{\beta}_1 M_w + \bar{\beta}_2 \eta_x = \bar{\beta}_3 \quad (41c)$$

Ignoring all the nonlinear terms of the stress resultants, the initial-boundary value problem of nonuniform torsion problem with STMDE within the range of linear elastic dynamic analysis can be derived. It can be easily concluded that in this simplified problem, the axial and torsional deformation are uncoupled; hence, they can be investigated independently. Moreover, the geometrically nonlinear dynamic problem of nonuniform torsion without STMDE can be described by setting $(\theta_x^P)' = \theta_x'$ in Eq. 21 and following the procedure presented in this section. The differential equations and the boundary conditions of the corresponding simplified initial-boundary value problem are given as

$$-\rho A \ddot{u} + EA u'' + EI_P \theta_x' \theta_x'' = -p_x \quad (42a)$$

$$\begin{aligned} & \rho I_P \ddot{\theta}_x - \rho C_S \ddot{\eta}_x + EC_S \theta_x''' - GI_t^P \theta_x'' \\ & - \frac{3}{2}EI_{PP}(\theta_x')^2 \theta_x'' - EI_P u' \theta_x'' - EI_P u'' \theta_x' = m_t - m_w' \end{aligned} \quad (42b)$$

$$u(x, 0) = \bar{u}_0(x) \quad (43a)$$

$$\dot{u}(x, 0) = \dot{\bar{u}}_0(x) \quad (43b)$$

$$\theta_x(x, 0) = \bar{\theta}_{x0}(x) \quad (43c)$$

$$\dot{\theta}_x(x, 0) = \dot{\bar{\theta}}_{x0}(x) \quad (43d)$$

$$\alpha_1 N + \alpha_2 u = \alpha_3 \quad (44a)$$

$$\beta_1 M_t + \beta_2 \theta_x = \beta_3 \quad (44b)$$

$$\bar{\beta}_1 M_w + \bar{\beta}_2 \theta'_x = \bar{\beta}_3 \quad (44c)$$

where M_t, M_w at the beam ends are given as

$$M_t = \rho C_S \ddot{\theta}'_x + G I_t^P \theta'_x - E C_S \theta'''_x + \frac{1}{2} E I_{PP} (\theta'_x)^3 + E I_P u' \theta'_x - m_w \quad (45a)$$

$$M_w = E C_S \theta''_x \quad (45b)$$

Finally, in order to neglect the influence of nonuniform torsion, it is considered that $C_S = 0$, $m_w = 0$ in Eq. 42b, 45a and in the corresponding simplified initial-boundary value problem, the boundary condition Eq. 44c with respect to warping is neglected.

Secondary Warping Function

The boundary value problem for the evaluation of ϕ_S^S has been defined in section “Local Equilibrium Equation, Primary Torsional Warping Function” (Eqs. 26a and 27). With the aid of the global equilibrium equation of motion Eq. 39c and ignoring the influence of m_w , Eq. 26a is written as

$$\begin{aligned} & (-\rho \ddot{\eta}_x + E \eta''_x) \left(\frac{C_S}{I_t^S} \nabla^2 \phi_S^S - \rho \frac{C_S}{G I_t^S} \ddot{\phi}_S^S - \phi_S^P \right) \\ & - E [u'' + (y^2 + z^2) \theta'_x \theta''_x] \\ & + \rho \left[\ddot{u} - (\ddot{\eta}_x - \ddot{\theta}'_x) (\phi_S^P + \phi_S^S) \right] \\ & = 0 \quad \sigma \tau \Omega, \quad \forall x \in [0, L] \end{aligned} \quad (46)$$

Removing m_w constitutes an assumption similar to $t_x \approx 0$ which has been adopted in section

“Local Equilibrium Equation, Primary Torsional Warping Function.” In order to further simplify the computation of secondary warping function, nonlinear terms $(E u'', E(y^2 + z^2) \theta'_x \theta''_x, \rho \ddot{u})$ as well as inertia terms $(-\ddot{\phi}_S^S \rho C_S / (G I_t^S), \rho (\ddot{\eta}_x - \ddot{\theta}'_x) (\phi_S^P + \phi_S^S))$ appearing in Eq. 26a are neglected. As a consequence, the longitudinal local equilibrium in geometrically nonlinear analysis is partially violated; however, in the corresponding problem in the linear range, local equilibrium is fulfilled, if the inertia terms of the secondary warping function are neglected. Since $\eta_x \neq \theta'_x$ along the beam length, it holds that $(-\rho \ddot{\eta}_x + E \eta''_x) \neq 0$. Consequently, Eq. 46 is written as

$$\nabla^2 \phi_S^S = \frac{I_t^S}{C_S} \phi_S^P \sigma \tau \Omega, \quad \forall x \in [0, L], \quad (47)$$

from which it can be concluded that ϕ_S^S is a two-dimensional time-independent warping function.

Applications

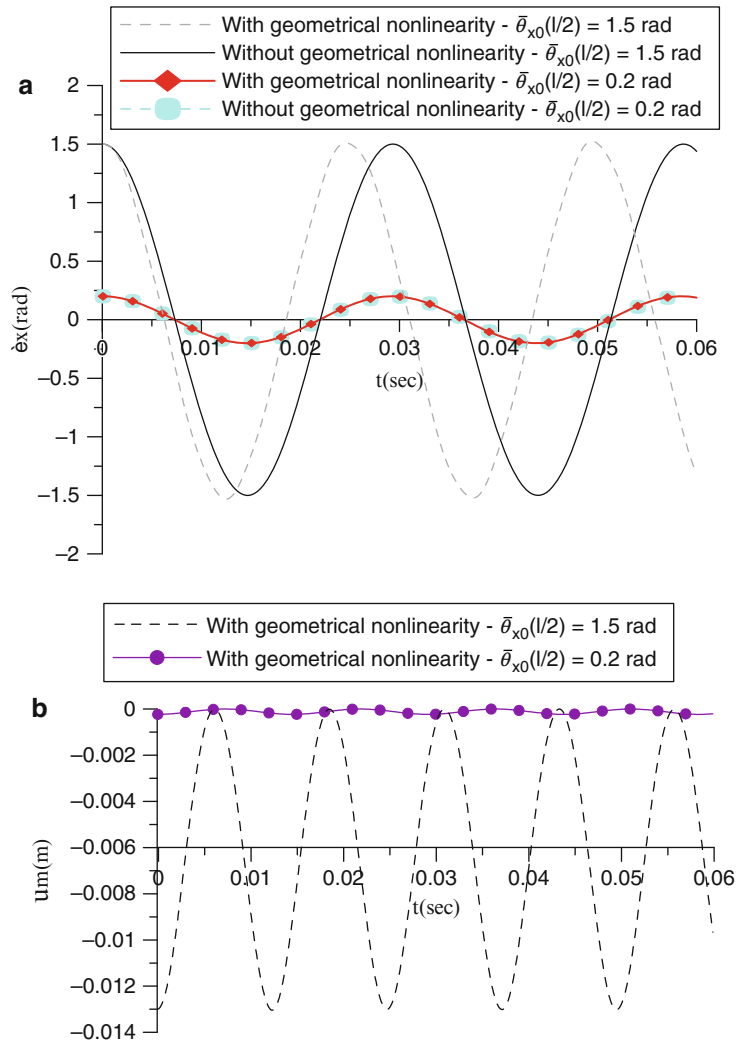
Free Vibrations of Beam of Thin-Walled I-Section Without STMDE

In order to investigate the influence of geometrical nonlinearity on the free vibrations of beams under nonuniform torsion, a beam of an I-shaped cross section ($E = 2.1 \times 10^8 \text{ kN/m}^2$, $G = 8.1 \times 10^7 \text{ kN/m}^2$, $\rho = 8.002 \text{ kNs}^2/\text{m}^4$) of length $L = 4.0 \text{ m}$, having flange and web width $t_f = t_w = 0.01 \text{ m}$ and total height and total width $h = b = 0.20 \text{ m}$, has been studied. Two cases of boundary conditions are studied, while the influence of STMDE is neglected. The ends of the beam are simply supported (according to the torsional boundary conditions), the left end is axially immovable, while the right one is subjected to a constant force $\bar{N}(L, t)$. The initial-boundary value problems described by Eqs. 42, 43, and 44 are solved, in order to obtain the response of the beam in the time domain taking into account or ignoring the influence of axial inertia.

The linear fundamental modeshape of the angle of twist is firstly considered as initial

Nonlinear Dynamic Seismic Analysis,

Fig. 10 Time histories of the angle of twist at the midpoint (a) and axial displacement at the right end point (b) of the beam of application section “Free Vibrations of Beam of Thin-Walled I-Section Without STMDE,” for the case of initial twisting rotations $\bar{\theta}_{x0}(x)$

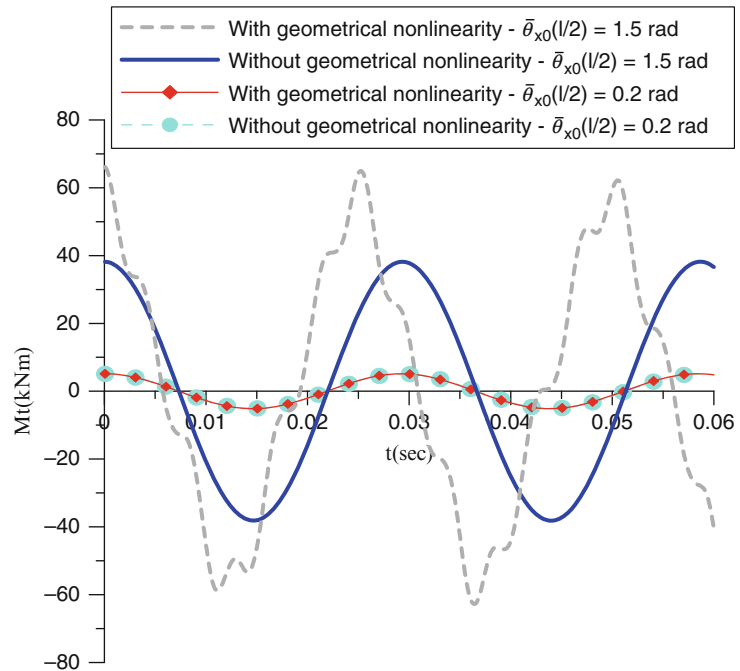


twisting rotation $\bar{\theta}_{x0}(x)$ along with zero initial twisting velocity $\dot{\bar{\theta}}_{x0}(x)$, while the axial force at the right end is considered to be zero. The normalization of the modeshape is performed by specifying the amplitude of the angle of twist at the midpoint of the beam $\bar{\theta}_{x0}(L/2)$. In Figs. 10 and 11, the time histories of characteristic kinematical components (angle of twist $\theta_x(l/2, t)$ at the midpoint of the beam, axial displacement $u(l, t)$ at the beam's right end), and the twisting moment at the beam's left end $M_t(0, t)$, respectively, are presented for two values of the initial midpoint angle of twist amplitude ($\bar{\theta}_{x0}(l/2) = 1.5$ rad, $\bar{\theta}_{x0}(l/2) = 0.2$ rad) performing two cases of analysis (linear analysis and nonlinear analysis

ignoring axial inertia term). From these figures, the effects of both the nonlinear terms and the initial midpoint angle of twist amplitude are observed on both the kinematical components and stress resultants. It is worth here noting that the pronounced alteration of the twisting moment's time history pattern and a slight change of the amplitude of the angle of twist (which is not clearly shown in Fig. 10a) are attributed to the fact that the fundamental nonlinear modeshape does not coincide with the linear one. Moreover, as it is easily verified from Fig. 10, the frequency of the axial displacement's response is twice as much as the one of the twisting rotation. Moreover, the case of constant along the beam initial

Nonlinear Dynamic Seismic Analysis,

Fig. 11 Time history of the twisting moment at the left end of the beam of application section “Free Vibrations of Beam of Thin-Walled I-Section Without STMDE,” for the case of initial twisting rotations $\bar{\theta}_{x0}(x)$



twisting velocities $\dot{\bar{\theta}}_{x0}(x) = 50$ rad/s, variable along the beam initial axial displacements given from $\bar{u}_0(x) = [\bar{N}(l, 0)/(EA)]x$, along with vanishing initial twisting rotations $\bar{\theta}_{x0}(x) = 0$ and initial axial velocities $\dot{\bar{u}}_0(x) = 0$, applied to the beam with immovable left and free right end subjected to constant axial load $\bar{N}(l, t) = -3,000$ kN according to the axial boundary conditions, has been also studied. In Figs. 12 and 13, the time history of the angle of twist $\theta_x(l/2, t)$ at the midpoint of the beam and $u(l, t)$ at the beam end, respectively, is presented for three cases of analysis, namely, a linear one, a nonlinear complete, and a nonlinear one ignoring the influence of axial inertia. From these figures the influence of geometrical nonlinearity can be once again observed, while the influence of axial inertia proves to be negligible.

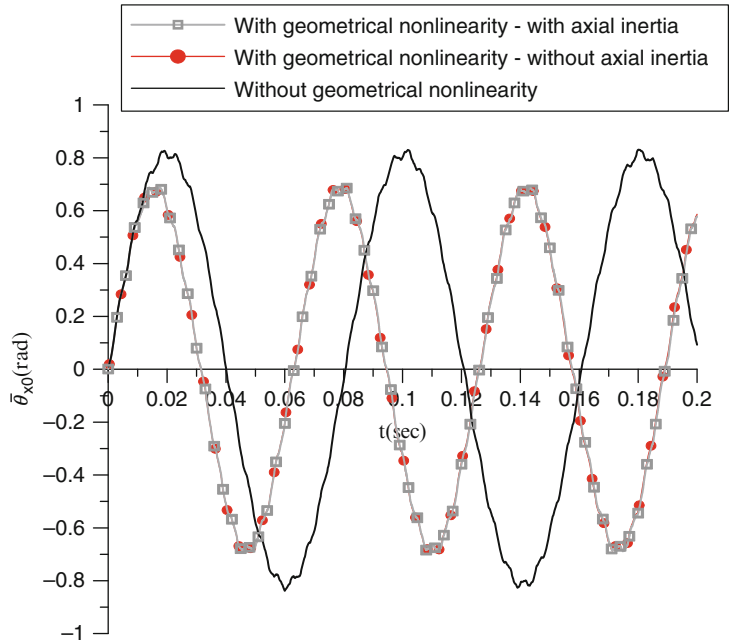
Forced Vibrations of Beam of Thin-Walled I-Section Without STMDE

In order to investigate the influence of geometrical nonlinearity on the forced vibrations of beams under nonuniform torsion, the beam of application section “Free Vibrations of Beam of Thin-Walled I-Section Without STMDE” has been

studied ignoring the influence of STMDE. The studied beam is considered to be simply supported (according to the torsional boundary conditions) and subjected to a moving concentrated twisting moment. All of the initial conditions are zero except from the initial axial displacements, which are given from $\bar{u}_0(x) = [\bar{N}(l, 0)/(EA)]x$, while the beam has an immovable left and free right end subjected to constant axial load $\bar{N}(l, t) = -2,500$ kN according to the axial boundary conditions. The concentrated twisting moment has a constant numerical value $M_t = 20.0$ kNm and “travels” with a constant velocity $v = 40$ m/s; thus, the beam is subjected to free vibrations after $t = 0.1$ s. In Fig. 14, the time history of the angle of twist $\theta_x(l/2, t)$ at the midpoint of the beam is presented for the aforementioned cases of analysis (linear, nonlinear) demonstrating the discrepancy of the nonlinear response of the beam compared with that of the linear one in both its forced and free vibrating phase. In Fig. 15, the time history of the axial displacement at the beam’s right end $u(l, t)$ (which is constant in the linear case) is presented for whichever of the aforementioned cases of analysis, the results are not trivial.

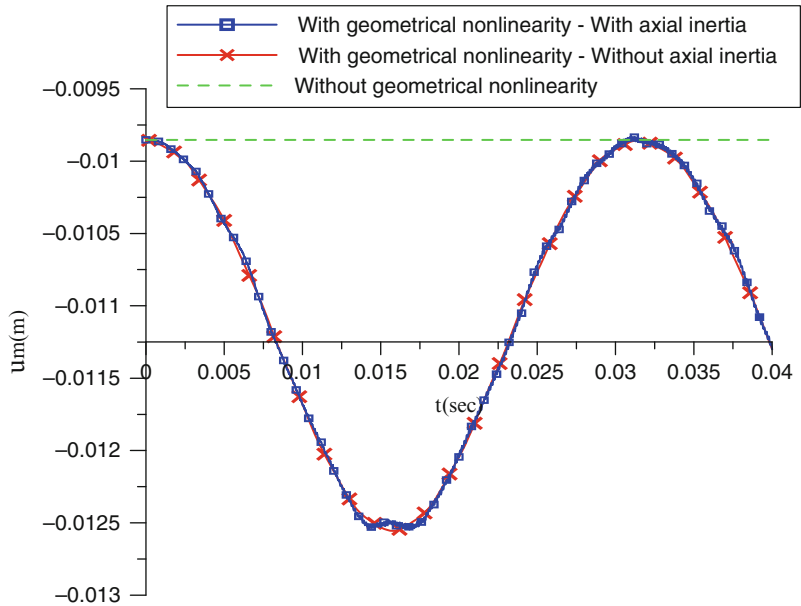
Nonlinear Dynamic Seismic Analysis,

Fig. 12 Time history of the angle of twist at the midpoint of the beam of application section “Free Vibrations of Beam of Thin-walled I-Section Without STMDE,” for the case of initial twisting velocities $\dot{\theta}_{x0}(x)$ and initial axial displacements $\bar{u}_0(x)$



Nonlinear Dynamic Seismic Analysis,

Fig. 13 Time histories of the axial stress resultant at the left end (a) and the axial displacement at the right end (b) of the beam of application section “Free Vibrations of Beam of Thin-walled I-Section Without STMDE,” for the case of initial twisting velocities $\dot{\theta}_{x0}(x)$ and initial axial displacements $\bar{u}_0(x)$



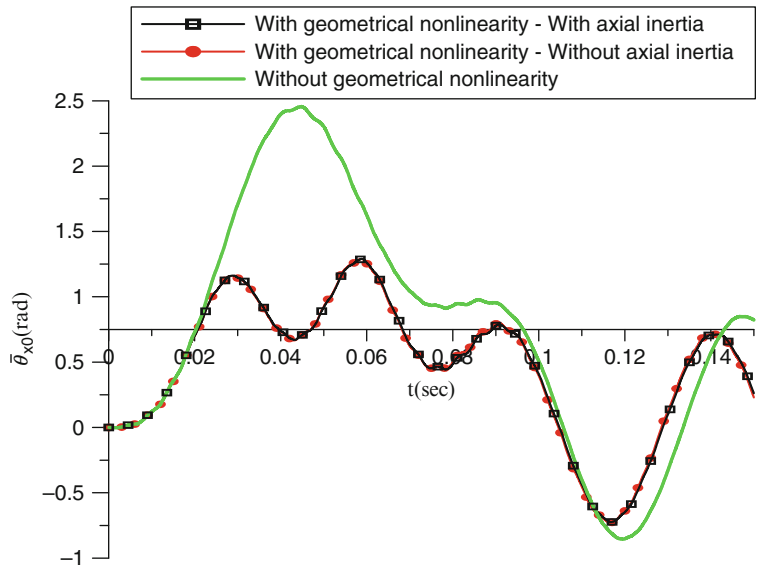
Free Vibrations of Beam of Thin-Walled I-Section with STMDE

In order to investigate the influence of STMDE in linear free vibrations of a beam of an open cross section under nonuniform torsion, the beam of application section “Free Vibrations of Beam of

Thin-Walled I-Section Without STMDE” ($I_t^S = 3.048 \times 10^{-4} \text{m}^4$) is examined. The beam’s ends are simply supported according to its torsional boundary conditions, while the left end is immovable and the right end is subjected to a compressive axial load according to its axial

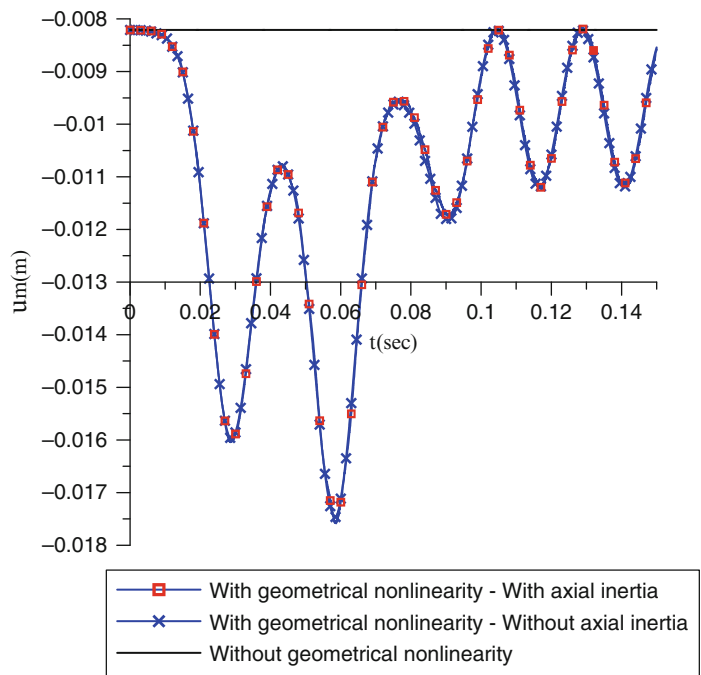
Nonlinear Dynamic Seismic Analysis,

Fig. 14 Time history of the angle of twist at the midpoint of the beam of application section “[Forced Vibrations of Beam of Thin-Walled I-Section Without STMDE](#)”



Nonlinear Dynamic Seismic Analysis,

Fig. 15 Time history of the axial displacement at the right end of the beam of application section “[Forced Vibrations of Beam of Thin-Walled I-Section Without STMDE](#)”



boundary conditions. For comparison reasons, the axial load–torsional fundamental natural frequency relation of the aforementioned beam has been investigated. Mohri et al. (2004), without considering the secondary twisting moment deformation effect, proposed analytical load–frequency relations by dropping the

higher-order warping inertia term and assuming that (i) the fundamental modeshape of vibration follows a sinusoidal form $\sin(\pi x/l)$ both in the pre- and post-buckling region, (ii) the beam vibrates harmonically, and (iii) vibrations of small amplitude around a static equilibrium state are performed. These load ($\bar{N}$)–frequency

(ω_f) relations for the pre- and post-buckling regions are given, respectively, as

$$\omega_f^2 = \omega_\theta^2 \left(1 - \frac{\bar{N}}{\bar{N}_{cr,\theta}} \right) \quad (48a)$$

$$\omega_f^2 = 2\omega_\theta^2 \left(\frac{\bar{N}}{\bar{N}_{cr,\theta}} - 1 \right) \quad (48b)$$

where ω_θ is the natural frequency for vanishing axial load and $\bar{N}_{cr,\theta}$ is the torsional buckling load of the beam, respectively, given as (Mohri et al. 2004)

$$\omega_\theta^2 = \frac{\pi^2}{\rho l^2 I_p} \left(\frac{\pi^2 E C_s}{l^2} + G I_t \right) \quad (49a)$$

$$\bar{N}_{cr,\theta} = \frac{A}{I_p} \left(\frac{\pi^2 E C_s}{l^2} + G I_t \right) \quad (49b)$$

According to the results of the presented methodology, the nonlinear initial-boundary value problem (Eqs. 39, 40, and 41) has been solved for several values of constant axial load $\bar{N}$ to obtain the response of the beam in the time domain, taking into account or neglecting the STMDE. The fundamental natural frequency ω_f is then computed by exploiting the first few full cycles of vibration. The computation is based on the fast Fourier transform (FFT) of the time history, while a Hanning data window is employed (Brigham 1988). The free vibrations are initiated by dropping all the torsional loading terms and by employing the linear fundamental modeshape of the angle of twist as initial twisting rotations $\bar{\theta}_{x0}(x)$ along with zero initial twisting velocities $\dot{\bar{\theta}}_{x0}(x)$. In all cases, the STMDE is neglected in the evaluation of $\bar{\theta}_{x0}(x)$. $\bar{\theta}_{x0}(x)$ is computed by employing the methodology presented in the study of Sapountzakis and Tsipiras (2010a) and corresponds to the sinusoidal form assumed by Mohri et al. (2004). In order to perform vibrations of small amplitude, the initial midpoint angle of twist amplitude $\bar{\theta}_{x0}(l/2)$ is chosen closely to the midpoint angle of twist amplitude $\theta_0(l/2)$ corresponding to the static equilibrium state. Mohri et al. (2004) assuming that the twisting deformation mode follows a sinusoidal form

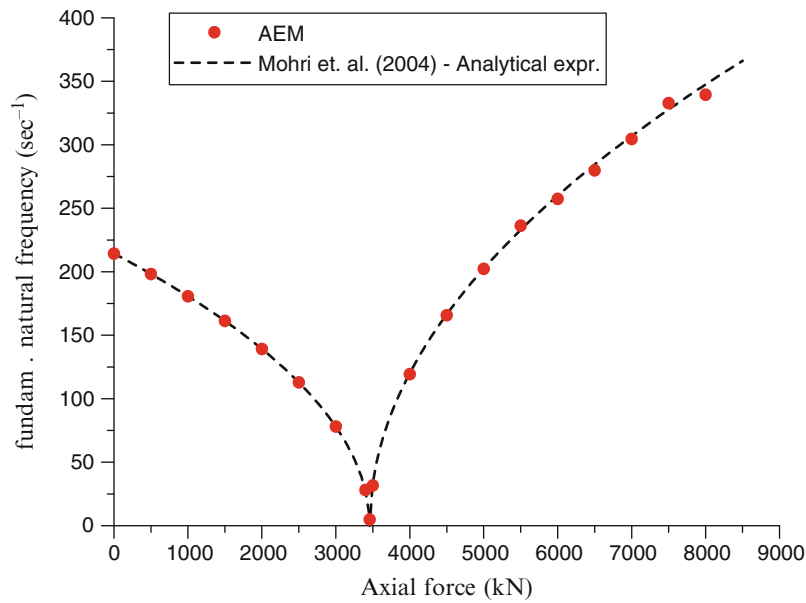
$\sin(\pi x/l)$ proposed the following analytical expression to evaluate $\theta_0(l/2)$

$$\theta_0(l/2) = \begin{cases} 0, & \text{prebuckling region} \\ \pm \sqrt{\frac{8l^2}{3\pi^2 E A I_n} (\bar{N} - \bar{N}_{cr,\theta})}, & \text{postbuckling region} \end{cases} \quad (50)$$

In Fig. 16, the load–frequency relations of Eq. 50 are presented along with pairs of values ($\omega_f, \bar{N}$) obtained from the proposed method, ignoring STMDE. The pairs ($\omega_f, \bar{N}$) are computed by using the values of initial midpoint angle of twist amplitude $\bar{\theta}_{x0}(l/2)$, which are given in Table 7 along with the corresponding values of $\theta_0(l/2)$ computed from Eq. 50. From Fig. 16, the validity of the developed method in predicting $\bar{N} - \omega_f$ pairs of values of beams undergoing small-amplitude torsional vibrations is concluded. The higher discrepancies between the two sets of solutions in the post-buckling region as compared to the ones in the pre-buckling region are attributed to the fact that the previously mentioned assumptions (i), (ii) are valid only in the pre-buckling region. To illustrate this point better, in Fig. 17, the obtained time histories of the angle of twist $\theta_x(l/2, t)$ at the midpoint of the beam for a pre-buckling ($\bar{N} = -1,000$ kN) and a post-buckling ($\bar{N} = -5,000$ kN) axial loading are presented, respectively, demonstrating that only the buckled beam undergoes multifrequency vibrations. For comparison purposes, in Fig. 17b, the time history of $\theta_x(l/2, t)$ employing the nonlinear fundamental modeshape of the angle of twist as initial twisting rotations $\bar{\theta}_{x0}(x)$ (along with zero initial twisting velocities $\dot{\bar{\theta}}_{x0}(x)$) is also included ($\bar{N} = -5,000$ kN), showing that the initiation of free vibrations with the nonlinear modeshape does not induce higher harmonics in the response of the beam. Finally, in Table 8, the obtained results of pairs of values ($\omega_f, \bar{N}$) are presented taking into account or ignoring STMDE and employing either the linear or the nonlinear fundamental modeshape to initiate the free vibrations, showing that in both pre- and post-buckling regions, STMDE does not affect the fundamental natural frequency of beams of open thin-walled cross sections undergoing

Nonlinear Dynamic Seismic Analysis,

Fig. 16 Axial load $\bar{N}$ –torsional fundamental natural frequency ω_f relation of Eq. 50 (Mohri et al. 2004) along with pairs of values $(\omega_f, \bar{N})$ obtained for application section “Free Vibrations of Beam of Thin-walled I-Section with STMDE” (AEM)



Nonlinear Dynamic Seismic Analysis, Table 7 Axial load $\bar{N}$, midpoint angle of twist amplitude $\theta_0(l/2)$ corresponding to the static equilibrium state (Mohri et al. 2004), along with values of initial midpoint angle of twist amplitude $\bar{\theta}_{x0}(l/2)$ used to initiate free vibrations of application section “Free Vibrations of Beam of Thin-Walled I-Section with STMDE”

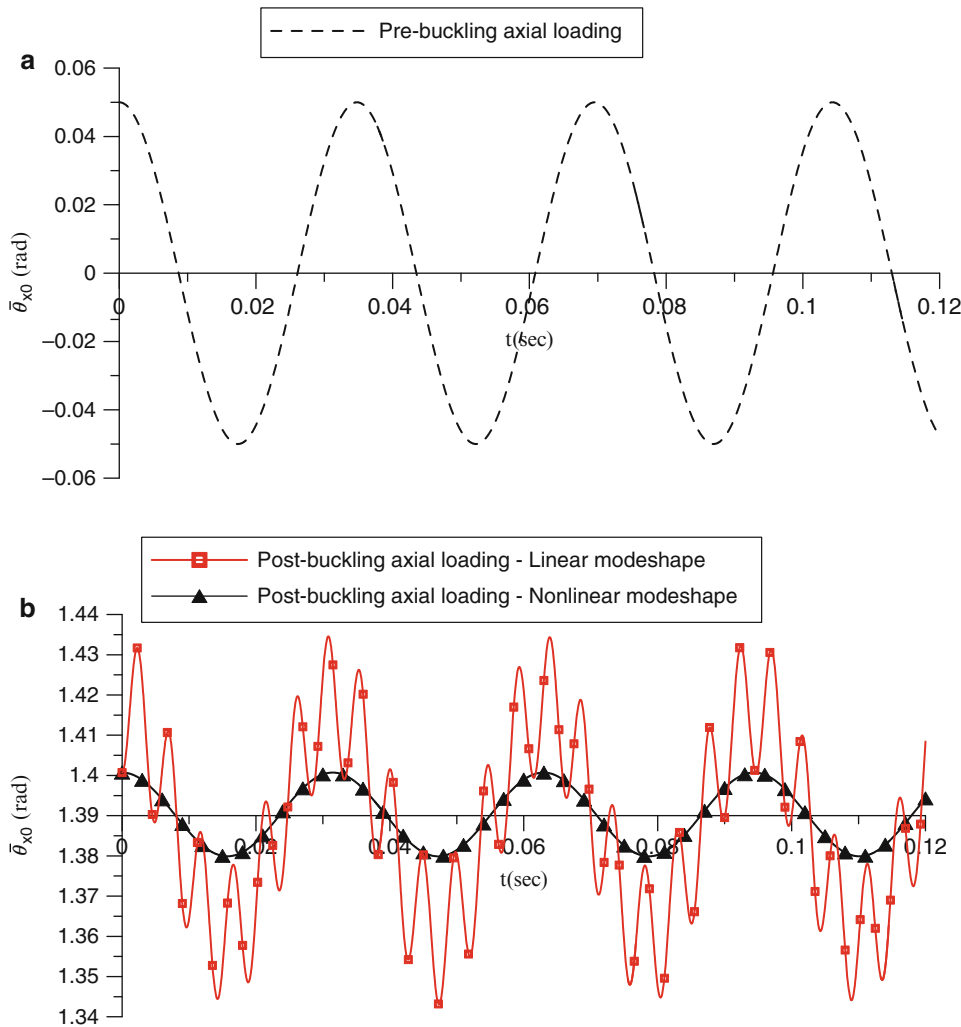
$ \bar{N} $ (kN)	$\theta_0(L/2)$ (rad)	$\bar{\theta}_{x0}(L/2)$
0–3,458	0.00	0.05
3,500	0.22	0.27
4,000	0.80	0.85
4,500	1.11	1.16
5,000	1.35	1.40
5,500	1.55	1.60
6,000	1.73	1.85
6,500	1.90	2.00
7,000	2.05	2.20
7,500	2.19	2.35
8,000	2.32	2.50

small-amplitude vibrations. The insignificance of STMDE for linear static loading conditions of beams of such cross sections has already been reported in the literature (Murín and Kutis 2008).

Primary Resonance of Beam of Thin-Walled I-Section with STMDE

In this application, in order to investigate the influence of geometrical nonlinearity and

STMDE on the response of beams of open cross section under the action of external twisting loading (forced vibrations), the beam of application section “Free Vibrations of Beam of Thin-walled I-Section Without STMDE” is examined under nonuniform torsion in its pre-buckled state. More specifically, the primary resonance of the beam is studied by applying a concentrated twisting moment $M_{t,ext}$ at its midpoint given as $M_{t,ext}(t) = M_{t0} \sin(\omega_{f,lin}t)$, where $M_{t0} = 5$ kNm and $\omega_{f,lin} = 214.23 \text{ s}^{-1}$ (initial conditions $\bar{\theta}_{x0}(x) = 0$, $\bar{\theta}_{x0}(x) = 0$). $\omega_{f,lin}$ is the fundamental natural frequency of the beam undergoing linear torsional vibrations, ignoring STMDE, and is numerically evaluated by performing modal analysis employing Eqs. 39, 40, and 41 and ignoring geometrical nonlinearity. In Figs. 18 and 19, the time histories of the angle of twist $\theta_x(l/2, t)$ at the midpoint of the beam are presented (with or without STMDE) for two cases of analysis, namely, ignoring or considering geometrical nonlinearity, respectively. The linear analyses are carried out by dropping all the nonlinear terms of Eqs. 39, 40, and 41. As expected, only in the linear cases (Fig. 18), deformations continue to increase with time. The beating phenomenon observed in the nonlinear response



Nonlinear Dynamic Seismic Analysis, Fig. 17 Time histories of the angle of twist at the midpoint of the beam of application “Free Vibrations of Beam of Thin-walled

I-Section with STMDE,” for a pre-buckling ($\bar{N} = -1,000$ kN) (a) and a post-buckling ($\bar{N} = -5,000$ kN) axial loading

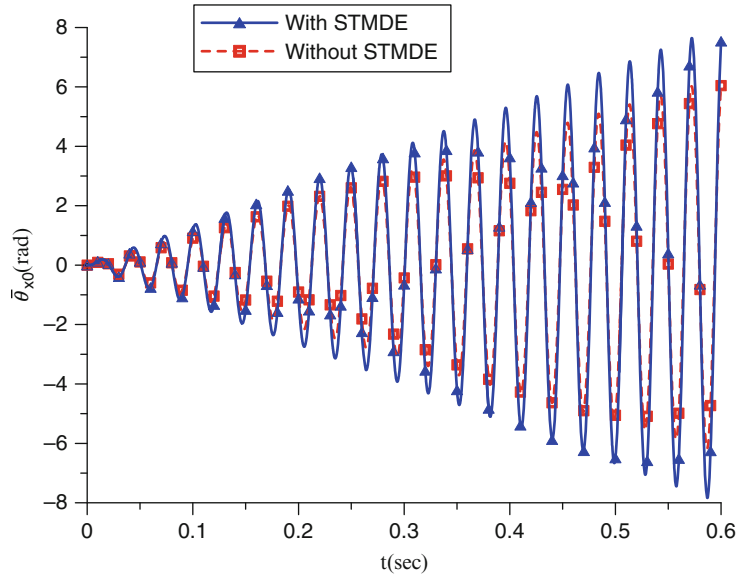
Nonlinear Dynamic Seismic Analysis, Table 8 Pairs of values $\bar{N}$ (kN), ω_f (s⁻¹) of the beam of application section “Free Vibrations of Beam of Thin-Walled I-Section with STMDE”

$ \bar{N} $ (kN)	ω_f (s ⁻¹)			
	$\bar{\theta}_{x0}(x)$ from linear analysis		$\bar{\theta}_{x0}(x)$ from nonlinear analysis	
	With STMDE	Without STMDE	With STMDE	Without STMDE
1,000	180.48	180.70	—	—
5,000	199.77	199.46	199.98	199.67

(Fig. 19) is explained from the fact that large twisting rotations increase the beam’s fundamental natural frequency ω_f (by increasing the stiffness of the beam), thereby causing a detuning of ω_f with the frequency of the external loading ($\omega_{f,lin}$). After the angle of twist reaches its maximum value, the amplitude of twisting deformations decreases leading to the reversal of the previously mentioned effects.

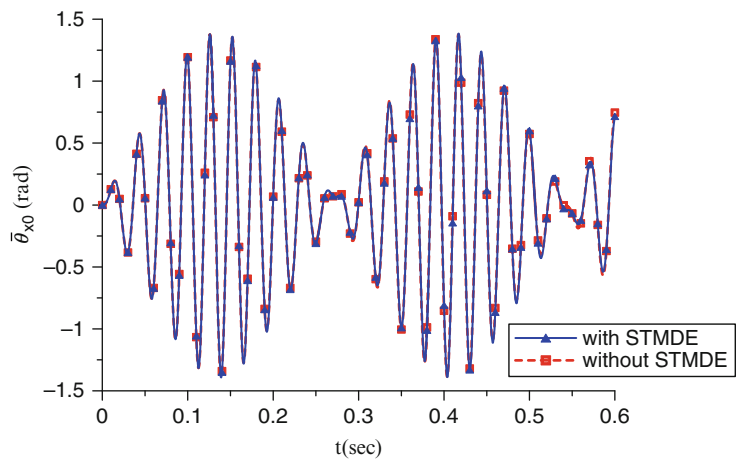
Nonlinear Dynamic Seismic Analysis,

Fig. 18 Time history of the angle of twist at the midpoint of the beam of application section “Primary Resonance of Beam of Thin-Walled I-Section with STMDE” taking into account or ignoring secondary twisting moment deformation effect – geometrically linear case



Nonlinear Dynamic Seismic Analysis,

Fig. 19 Time history of the angle of twist at the midpoint of the beam of application section “Primary Resonance of Beam of Thin-Walled I-Section with STMDE” taking into account or ignoring secondary twisting moment deformation effect – geometrically nonlinear case



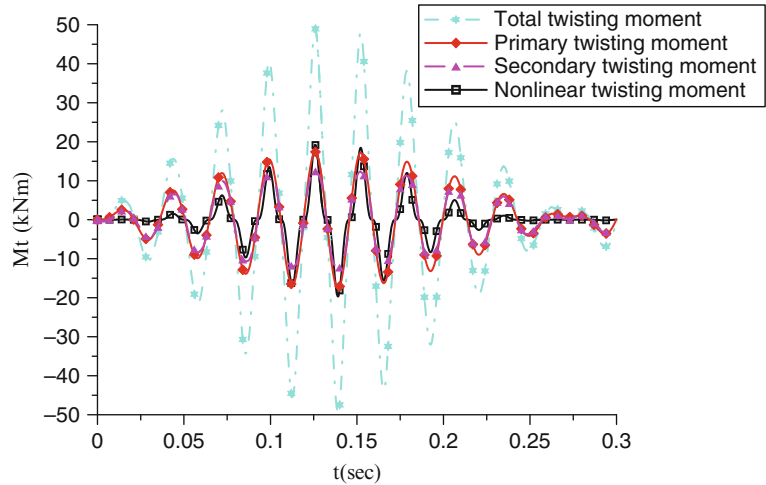
Moreover, in Fig. 20, the time histories of the primary twisting moment M_t^P , the linear secondary twisting moment $M_{t,lin}^S$, the nonlinear twisting moment $M_{t,nl}$, and the total twisting moment $M_t = M_t^P + M_{t,lin}^S + M_{t,nl}$ at the beam's left end are presented taking into account STMDE and performing geometrically nonlinear analysis. It is deduced that the STMDE influences negligibly both kinematical and stress components of beams of open thin-walled cross sections undergoing forced vibrations.

Primary Resonance of Beam of Closed Cross Section with STMDE

In this application, in order to investigate the influence of STMDE on the forced linear vibrations of a beam of closed cross section under nonuniform torsion, a steel beam ($E = 2.1 \times 10^8 \text{ kN/m}^2$, $G = 8.0769 \times 10^7 \text{ kN/m}^2$) of length $L = 5.0 \text{ m}$ and of a hollow rectangular cross section (total height $H = 1.64 \text{ m}$, total width $B = 1.05 \text{ m}$, web thickness $t_w = 0.05 \text{ m}$, and flange thickness $t_f = 0.04 \text{ m}$) is examined.

Nonlinear Dynamic Seismic Analysis,

Fig. 20 Time histories of various twisting moment components at the left end of the beam of application section “Primary Resonance of Beam of Thin-Walled I-Section with STMDE” taking into account the secondary twisting moment deformation effect (geometrically nonlinear analysis)



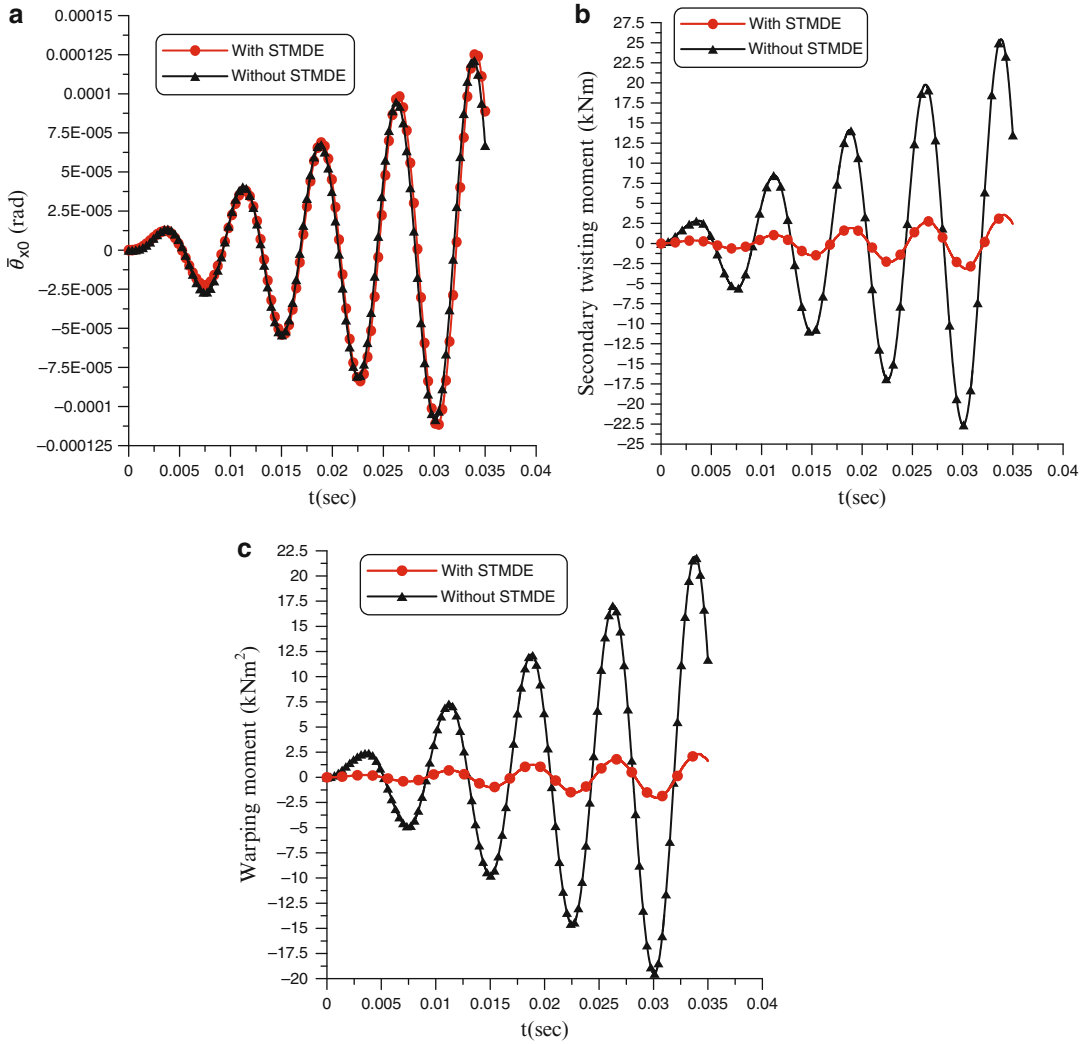
The geometric constants of the beam are assumed to take the values presented in the study of Murín and Kutiš (2008) ($A = 0.240\text{m}^2$, $I_t = 0.089824\text{m}^4$, $I_t^S = 0.001107\text{m}^4$, $C_S = 0.000193\text{m}^6$). The beam's left end is clamped, and its resonance is studied by applying a distributed twisting moment $m_{t,\text{ext}}$ at $0 < x < 5$ (m) given as $m_{t,\text{ext}}(x, t) = m_{t0} \sin(\omega_{f,\text{lin}} t)$, where $m_{t0} = 5$ kNm/m and $\omega_{f,\text{lin}} = 835, 793\text{s}^{-1}$ ($M_t(l, t) = 0$, vanishing initial conditions $\bar{\theta}_{x0}(x) = 0, \dot{\bar{\theta}}_{x0}(x) = 0$). $\omega_{f,\text{lin}}$ is the fundamental natural frequency of the beam undergoing linear torsional vibrations, ignoring STMDE, and is numerically evaluated by performing modal analysis similarly with applications sections “Free Vibrations of Beam of Thin-Walled I-Section Without STMDE” and “Free Vibrations of Beam of Thin-Walled I-Section with STMDE.”

In Fig. 21, the time histories of the angle of twist $\theta_x(L, t)$, the secondary twisting moment $M_t^S(0, t)$ (scaled quantity $M_t^S(0, t) \times 0, 1$), and the warping moment $M_w(0, t)$, respectively, are presented taking into account or ignoring STMDE. It can be observed that the influence of STMDE on the magnitude of the angle of twist and the torsional stiffness of the beam is negligible, while its influence on the stress resultants is significant and cannot be neglected.

Nonlinear Flexural–Torsional Dynamic Analysis of Beams of Arbitrary Cross Section

Statement of the Problem

Let us consider a prismatic beam of length L (Fig. 22), of constant arbitrary cross section of area A . The homogeneous isotropic and linearly elastic material of the beam's cross section, with modulus of elasticity E , shear modulus G , and Poisson's ratio ν , occupies the two-dimensional multiply connected region Ω of the y, z plane and is bounded by the $\Gamma_j (j = 1, 2, \dots, K)$ boundary curves, which are piecewise smooth, i.e., they may have a finite number of corners. In Fig. 22, CYZ is the principal bending coordinate system through the cross section's centroid C , while y_C, z_C are its coordinates with respect to the Syz shear system of axes through the cross section's shear center S , with axes parallel to those of the CYZ system. The beam is subjected to the combined action of the arbitrarily distributed or concentrated, time-dependent, and conservative axial loading $p_X = p_X(X, t)$ along the X direction, twisting moment $m_x = m_x(x, t)$ along the x direction, and transverse loading $p_y = p_y(x, t)$, $p_z = p_z(x, t)$ acting along the y and z directions, applied at distances y_{p_y}, z_{p_y} and y_{p_z}, z_{p_z} , with respect to the Syz shear system of axes, respectively (Fig. 22b).



Nonlinear Dynamic Seismic Analysis, Fig. 21 Time history of the angle of twist $\theta_x(L, t)$ (a), the secondary twisting moment $M_t^S(0, t)$ (b), the warping moment $M_w(0, t)$ (c) of

the beam of application section “[Primary Resonance of Beam of Closed Cross Section with STMDE](#)”

Under the action of the aforementioned loading, the displacement field of an arbitrary point of the cross section can be derived with respect to those of the shear center as

$$\begin{aligned} \bar{u}(x, y, z, t) = & u(x, t) - (y - y_C)\theta_z(x, t) \\ & + (z - z_C)\theta_y(x, t) \\ & + \theta'_x(x, t)\varphi_S^P(y, z) \end{aligned} \quad (51a)$$

$$\bar{v}(x, y, z, t) = v(x, t) - z \sin(\theta_x(x, t)) - y[1 - \cos(\theta_x(x, t))] \quad (51b)$$

$$\bar{w}(x, y, z, t) = w(x, t) + y \sin(\theta_x(x, t)) - z[1 - \cos(\theta_x(x, t))] \quad (51c)$$

$$\theta_y(x, t) = v'(x, t) \sin(\theta_x(x, t)) - w'(x, t) \cos(\theta_x(x, t)) \quad (51d)$$

$$M_R = \int_{\Omega} S_{xx}(y^2 + z^2) d\Omega \quad (53)$$

which is a higher-order stress resultant. Using the expressions of the stress components (Eq. 52), the stress resultants are obtained as

$$N = EA \left\{ u' + \frac{1}{2} \left(v'^2 + w'^2 + \frac{I_p}{A} \theta_x'^2 \right) - \theta_x' [z_C(v' \cos \theta_x + w' \sin \theta_x) + y_C(v' \sin \theta_x - w' \cos \theta_x)] \right\} \quad (54a)$$

$$M_Y = -EI_Y (w'' \cos \theta_x - v'' \sin \theta_x - \beta_Z \theta_x'^2) \quad (54b)$$

$$M_Z = EI_Z (v'' \cos \theta_x + w'' \sin \theta_x - \beta_Y \theta_x'^2) \quad (54c)$$

$$M_t = GI_t \theta_x' \quad (54d)$$

$$M_w = -EC_S \left(\theta_x'' + \beta_\omega \theta_x'^2 \right) \quad (54e)$$

$$\begin{aligned} M_R = & N \frac{I_p}{A} - 2EI_Z \beta_Y (v'' \cos \theta_x + w'' \sin \theta_x) \\ & - 2EI_Y \beta_Z (w'' \cos \theta_x - v'' \sin \theta_x) \\ & + 2EC_S \beta_\omega \theta_x'' + \frac{1}{2} E \left(I_{pp} - \frac{I_p^2}{A} \right) \theta_x'^2 \end{aligned} \quad (54f, g)$$

where the area A , the principal moments of inertia I_Y, I_Z with respect to the cross section's centroid, the polar moment of inertia I_p , the torsion constant I_t , and the warping constant C_S with respect to the shear center S are given by relations Eqs. 10, 31a, 31b, and 31d, respectively, while $I_{pp} = \int_{\Omega} (y^2 + z^2)^2 d\Omega$ is the fourth moment of inertia with respect to the shear center S defined in section “[Stress Resultants, Global Equations of Motion](#)” as well. Finally, the Wagner's coefficients β_Z, β_Y , and β_ω are given as

$$\beta_Z = \frac{1}{2I_Y} \int_{\Omega} (z - z_C)(y^2 + z^2) d\Omega \quad (55a)$$

$$\beta_Y = \frac{1}{2I_Z} \int_{\Omega} (y - y_C)(y^2 + z^2) d\Omega \quad (55b)$$

$$\beta_\omega = \frac{1}{2C_S} \int_{\Omega} (y^2 + z^2) \phi_S^p d\Omega \quad (55c)$$

Similarly with the previous sections, applying the principle of virtual work (Eq. 6), the equations of motion of the beam can be derived as

$$-N' + \rho A \ddot{u} = p_X \quad (56a)$$

$$\begin{aligned} & [-N(v' - y_C \theta_x' \sin \theta_x - z_C \theta_x' \cos \theta_x)]' \\ & + (M_Z \cos \theta_x)'' + (M_Y \sin \theta_x)'' \end{aligned}$$

$$- \rho A \ddot{v} - \rho A (y_C \sin \theta_x + z_C \cos \theta_x) \ddot{\theta}_x$$

$$+ \rho A (z_C \sin \theta_x - y_C \cos \theta_x) \dot{\theta}_x^2$$

$$- (\rho \{ (I_Y - I_Z) \sin \theta_x \cos \theta_x v' \}$$

$$- [(I_Y - I_Z) \cos^2 \theta_x - I_Y] w' \} \ddot{\theta}_x)'$$

$$- (\rho \{ (I_Y - I_Z) \sin \theta_x \cos \theta_x w' \}$$

$$+ [(I_Y - I_Z) \cos^2 \theta_x - I_Y] v' \} \dot{\theta}_x^2)'$$

$$- [\rho (I_Y - I_Z) \cos \theta_x \sin \theta_x (2\dot{\theta}_x \dot{v}' - \ddot{w}')]'$$

$$+ \left\{ \rho [(I_Y - I_Z) \cos^2 \theta_x - I_Y] (2\dot{\theta}_x \dot{w}' + \ddot{v}') \right\}' = p_Y \quad (56b)$$

$$[-N(w' + y_C \theta_x' \cos \theta_x - z_C \theta_x' \sin \theta_x)]'$$

$$- (M_Y \cos \theta_x)'' + (M_Z \sin \theta_x)''$$

$$+ \rho A \ddot{w} - \rho A (-y_C \cos \theta_x + z_C \sin \theta_x) \ddot{\theta}_x$$

$$- \rho A (z_C \cos \theta_x + y_C \sin \theta_x) \dot{\theta}_x^2$$

$$- (\rho \{ (I_Z - I_Y) \cos \theta_x \sin \theta_x w' \}$$

$$+ [(I_Z - I_Y) \cos^2 \theta_x - I_Z] v' \} \ddot{\theta}_x)'$$

$$+ (\rho \{ (I_Z - I_Y) \cos \theta_x \sin \theta_x v' \}$$

$$- [(I_Z - I_Y) \cos^2 \theta_x - I_Z] w' \} \dot{\theta}_x^2)'$$

$$- [\rho (I_Z - I_Y) \cos \theta_x \sin \theta_x (2\dot{\theta}_x \dot{w}' + \ddot{v}')]'$$

$$+ \left\{ \rho [(I_Z - I_Y) \cos^2 \theta_x - I_Z] (\ddot{w} - 2\dot{\theta}_x \dot{v}' \dot{w}') \right\}' = p_Z \quad (56c)$$

$$\begin{aligned}
& N(-y_C v' \theta'_x \cos \theta_x - y_C w' \theta'_x \sin \theta_x - z_C w' \theta'_x \cos \theta_x \\
& \quad + z_C v' \theta'_x \sin \theta_x) + \quad - p_y z_{p_y} \cos \theta_x - p_z z_{p_z} \sin \theta_x \\
& \quad - p_y y_{p_y} \sin \theta_x \quad (56d)
\end{aligned}$$

$$[N(z_C(v' \cos \theta_x + w' \sin \theta_x))]'$$

$$+ [N(y_C(v' \sin \theta_x - w' \cos \theta_x))]'$$

$$M_Y(w'' \sin \theta_x + v'' \cos \theta_x)$$

$$+ M_Z(w'' \cos \theta_x - v'' \sin \theta_x) - M'_t - M''_w$$

$$- (M_R \theta'_x)' - \rho A [(z_C \cos \theta_x + y_C \sin \theta_x) \ddot{v}$$

$$+ (-z_C \sin \theta_x + y_C \cos \theta_x) \ddot{w}]$$

$$+ \rho \left\{ [(I_Y - I_Z) \cos^2 \theta_x + I_Z] v'^2 \right.$$

$$- 2(I_Z - I_Y) v' w' \sin \theta_x \cos \theta_x + I_s$$

$$+ [(I_Z - I_Y) \cos^2 \theta_x + I_Y] w'^2 \left. \right\} \ddot{\theta}_x$$

$$+ \rho \left\{ [(I_Z - I_Y) \cos^2 \theta_x + I_Y] w' \right.$$

$$+ (I_Y - I_Z) \sin \theta_x \cos \theta_x v' \left. \right\} (\ddot{v} + 2\dot{\theta}_x \dot{v}')$$

$$+ \rho \left\{ [(I_Z - I_Y) \cos^2 \theta_x - I_Z] v' \right.$$

$$+ (I_Z - I_Y) \sin \theta_x \cos \theta_x w' \left. \right\} (\ddot{w} - 2\dot{\theta}_x \dot{v}' v')$$

$$- \rho \left\{ (I_Y - I_Z) \sin \theta_x \cos \theta_x v'^2 \right.$$

$$+ (I_Z - I_Y) \sin \theta_x \cos \theta_x w'^2$$

$$+ [2(I_Z - I_Y) \cos^2 \theta_x - I_Z + I_Y] v' w' \left. \right\} \dot{\theta}_x^2$$

$$- (\rho C_s \ddot{\theta}_x)' = m_x + p_z y_{p_z} \cos \theta_x$$

where the expressions of the stress resultants are given from Eq. 54. Equation 56 is coupled and highly complicated. This set of equations can be simplified if the approximate expressions (Sapountzakis and Dikaros 2011)

$$\cos \theta_x = 1 - \frac{\theta_x^2}{2!} = 1 - \frac{\theta_x^2}{2} \quad (57a)$$

$$\sin \theta_x = \theta_x - \frac{\theta_x^3}{3!} = \theta_x - \frac{\theta_x^3}{6} \quad (57b)$$

are employed. Thus, using the aforementioned approximations, neglecting the term $\rho A \ddot{u}$ of Eq. 56a denoting the axial inertia of the beam, employing the expressions of the stress resultants (Eq. 54), and ignoring the nonlinear terms of the fourth or greater order (Sapountzakis and Dikaros 2011), the governing partial differential equations of motion for the beam at hand can be written as

$$\begin{aligned}
& -EA \left[u'' + w' w'' + v' v'' + \frac{I_p}{A} \theta'_x \theta''_x - z_C (\theta_x \theta'_x w'' + \theta_x \theta''_x w' + \theta'_x v' + \theta'_x v'' + \theta_x^2 w') \right. \\
& \quad \left. - y_C (\theta_x \theta''_x v' + \theta_x \theta'_x v'' - \theta''_x w' - \theta'_x w'' + \theta_x^2 v') \right] = p_x \quad (58a)
\end{aligned}$$

$$\begin{aligned}
& EI_Z v'''' - N \left[-z_C \theta''_x - y_C (\theta_x'^2 + \theta''_x \theta_x) + v'' \right] \\
& \quad + (EI_Z - EI_Y) (w'''' \theta_x + 2w''' \theta'_x + w'' \theta''_x \\
& \quad - v'''' \theta_x^2 - 4v''' \theta'_x \theta'_x - 2v'' \theta''_x \theta'_x - 2v' \theta_x'^2) \\
& \quad + EI_Z \beta_Y (-2\theta'_x \theta_x''' - 2\theta_x''^2) \\
& \quad + EI_Y \beta_Z (2\theta'_x \theta_x''' \theta_x + 2\theta_x''^2 \theta_x + 5\theta_x'^2 \theta''_x) + \rho A \ddot{v} + \\
& \quad + \rho [(I_Z - I_Y) (\theta_x v'' + \theta'_x v') - I_Z w''] \\
& \quad - A \left(y_C \theta_x + z_C - \frac{1}{2} z_C \theta_x^2 \right) \ddot{\theta}_x + \rho (I_Z - I_Y) \\
& \quad \times [2\theta_x \theta'_x \ddot{v}' + \theta_x^2 \ddot{v}'' + 2(\theta'_x \dot{\theta}_x + \theta_x \dot{\theta}'_x) \dot{v}' + 2\theta_x \dot{\theta}_x \ddot{v}'' \\
& \quad - \theta_x' \ddot{w}' - \theta_x \ddot{w}'' + v' \theta_x \ddot{\theta}'_x] + \rho I_Z \\
& \quad \times [-w' \ddot{\theta}'_x - \ddot{v}'' - 2\dot{\theta}_x \dot{w}'' + v'' \dot{\theta}_x^2 + 2v' \dot{\theta}_x \dot{\theta}'_x - 2\dot{\theta}'_x \dot{w}'] \\
& \quad - \rho A (y_C - z_C \theta_x) \dot{\theta}_x^2 = p_y \\
& \quad - p_x (v' - y_C \theta'_x \theta_x - z_C \theta'_x) \quad (58b) \\
& EI_Y w'''' - N [w'' + y_C \theta''_x - z_C (\theta_x'^2 + \theta_x \theta''_x)]
\end{aligned}$$

$$\begin{aligned}
& + (EI_Z - EI_Y)(v''''\theta_x + 2v''''\theta_x' + v''\theta_x'' \\
& + w''''\theta_x'^2 + 4w''''\theta_x\theta_x' + 2w''\theta_x\theta_x'' + 2w''\theta_x'^2) \\
& + EI_Z\beta_Y(-2\theta_x'\theta_x'''\theta_x - 5\theta_x'^2\theta_x'' - 2\theta_x''^2\theta_x) \\
& + EI_Y\beta_Z(-2\theta_x'\theta_x''' - 2\theta_x''^2) + \rho A\ddot{w} \\
& + \rho \left[(I_Z - I_Y)(-\theta_x w'' - \theta_x' w') + I_Y v'' \right. \\
& + A \left(-z_C \theta_x + y_C - \frac{1}{2} y_C \theta_x^2 \right) \left. \right] \ddot{\theta}_x - \rho(I_Z - I_Y) \\
& \cdot \left[2\theta_x \theta_x' \ddot{w}' + \theta_x^2 \ddot{w}'' + 2 \left(\theta_x' \dot{\theta}_x + \theta_x \dot{\theta}_x' \right) \dot{w}' \right. \\
& + 2\theta_x \dot{\theta}_x \ddot{w}'' + \theta_x' \ddot{v}' + \theta_x \ddot{v}'' + w' \theta_x \ddot{\theta}_x' \left. \right] - \rho I_Y \\
& \cdot \left[-v' \ddot{\theta}_x' + \ddot{w}'' - 2\dot{\theta}_x v'' - w'' \dot{\theta}_x^2 - 2w' \dot{\theta}_x \dot{\theta}_x' - 2\dot{\theta}_x' v' \right] \\
& - \rho A(z_C + y_C \theta_x) \dot{\theta}_x^2 = p_z \\
& - p_X(w' + y_C \theta_x' - z_C \theta_x' \theta_x) \quad (58c)
\end{aligned}$$

$$\begin{aligned}
& EC_S \theta_x'''' - GI_t \theta_x'' - \frac{3}{2} E \left(I_{pp} - \frac{I_p^2}{A} \right) \theta_x'^2 \theta_x'' \\
& - N \left[\frac{I_p}{A} \theta_x'' + y_C(w'' - v''\theta_x) - z_C(v'' + w''\theta_x) \right] \\
& + (EI_Z - EI_Y)(v''w'' - v''^2\theta_x + w''^2\theta_x) \\
& + EI_Z\beta_Y(2\theta_x''w''\theta_x + w''\theta_x'^2 \\
& + 2\theta_x'v'''' + 2\theta_x''v'' + 2\theta_x'w'''\theta_x) \\
& + EI_Y\beta_Z(-2\theta_x''v''\theta_x + 2\theta_x''w'' + 2\theta_x'w''') \\
& - v''\theta_x'^2 - 2\theta_x'v''''\theta_x) + \rho(v'^2 I_Y + w'^2 I_Z + I_S) \ddot{\theta}_x \\
& - \rho A \left(z_C + y_C \theta_x - \frac{1}{2} z_C \theta_x^2 \right) \ddot{v} \\
& + \rho A \left(y_C - z_C \theta_x - \frac{1}{2} y_C \theta_x^2 \right) \ddot{w} \\
& + \rho(I_Z - I_Y)\theta_x(w'\ddot{w}' - v'\ddot{v}') \\
& + \rho I_Z(w'\ddot{v}' + 2\dot{\theta}_x \dot{w}' w') \\
& + \rho I_Y(-v'\ddot{w}' + 2\dot{\theta}_x \dot{v}' v') - \rho C_S \ddot{\theta}_x'' \\
& = m_x + p_z y_{p_z} - p_y z_{p_z}
\end{aligned}$$

$$\begin{aligned}
& + \left(\frac{1}{2} \theta_x^2 z_{p_y} + \frac{1}{6} \theta_x^3 y_{p_y} - \theta_x y_{p_y} \right) p_y \\
& + \left(\frac{1}{6} \theta_x^3 z_{p_z} - \frac{1}{2} \theta_x^2 y_{p_z} - z_{p_z} \theta_x \right) p_z \\
& - p_X \left(\frac{I_p}{A} \theta_x' - y_C v' \theta_x - z_C w' \theta_x - z_C v' + y_C w' \right) \quad (58d)
\end{aligned}$$

while the expression of the axial stress resultant N is given as

$$\begin{aligned}
N = EA \left[u' + \frac{1}{2} \left(v'^2 + w'^2 + \frac{I_p}{A} \theta_x'^2 \right) \right. \\
\left. - \theta_x'(z_C(w'\theta_x + v') + y_C(v'\theta_x - w')) \right] \quad (59)
\end{aligned}$$

The above governing differential equations (Eq. 58) are also subjected to the initial conditions ($x \in (0, l)$)

$$u(x, 0) = u_0(x) \quad (60a)$$

$$u_{,t}(x, 0) = u_{0,t}(x) \quad (60b)$$

$$v(x, 0) = v_0(x) \quad (61a)$$

$$v_{,t}(x, 0) = v_{0,t}(x) \quad (61b)$$

$$w(x, 0) = w_0(x) \quad (62a)$$

$$w_{,t}(x, 0) = w_{0,t}(x) \quad (62b)$$

$$\theta_x(x, 0) = \theta_{x0}(x) \quad (63a)$$

$$\theta_{x,t}(x, 0) = \theta_{x0,t}(x) \quad (63b)$$

together with the corresponding boundary conditions of the problem at hand, which are given as

$$a_1 u(x, t) + \alpha_2 N_b(x, t) = \alpha_3 \quad (64)$$

$$\beta_1 v(x, t) + \beta_2 R_{by}(x, t) = \beta_3 \quad (65a)$$

$$\bar{\beta}_1 \theta_Z(x, t) + \bar{\beta}_2 M_{bz}(x, t) = \bar{\beta}_3 \quad (65b)$$

$$\gamma_1 w(x, t) + \gamma_2 R_{bz}(x, t) = \gamma_3 \quad (66a)$$

$$\bar{\gamma}_1 \theta_Y(x, t) + \bar{\gamma}_2 M_{bY}(x, t) = \bar{\gamma}_3 \quad (66b)$$

$$\delta_1 \theta_x(x, t) + \delta_2 M_{bt}(x, t) = \delta_3 \quad (67a)$$

$$\bar{\delta}_1 \theta_{x,x}(x, t) + \bar{\delta}_2 M_{bw}(x, t) = \bar{\delta}_3 \quad (67b)$$

at the beam ends $x = 0, l$, where R_{by} , R_{bz} and M_{bZ} , M_{bY} are the reactions and bending moments with respect to y, z or to Y, Z axes, respectively, given by the following relations (ignoring again the nonlinear terms of the fourth or greater order)

$$\begin{aligned} R_{by} = & N(v' - z_C \theta_x' - y_C \theta_x \theta_x') \\ & + EI_Z(v'''' \theta_x^2 - w'' \theta_x' - w'''' \theta_x - v'''' + 2v'' \theta_x \theta_x' + 2\beta_Y \theta_x' \theta_x'') \\ & + EI_Y(w''' \theta_x + w'''' \theta_x' - 2v'' \theta_x \theta_x' - v''' \theta_x^2 - \beta_Z \theta_x'^3 - 2\beta_Z \theta_x \theta_x' \theta_x'') \end{aligned} \quad (68a)$$

$$\begin{aligned} R_{bz} = & N(w' + y_C \theta_x' - z_C \theta_x' \theta_x) \\ & + EI_Y(w'''' \theta_x^2 + v'''' \theta_x - w'''' + v'' \theta_x' + 2w'' \theta_x \theta_x' + 2\beta_Z \theta_x' \theta_x'') \\ & - EI_Z(v'' \theta_x' + w''' \theta_x^2 + v''' \theta_x + 2w'' \theta_x \theta_x' - 2\beta_Y \theta_x \theta_x' \theta_x' - \beta_Y \theta_x'^3) \end{aligned} \quad (68b)$$

$$\begin{aligned} M_{bZ} = & EI_Z(w'' \theta_x - \beta_Y \theta_x'^2 + v'' - v'' \theta_x^2) \\ & + EI_Y(-w'' \theta_x + v'' \theta_x^2 + \beta_Z \theta_x'^2 \theta_x) \end{aligned} \quad (68c)$$

$$\begin{aligned} M_{bY} = & EI_Z(-w'' \theta_x^2 + \beta_Y \theta_x'^2 \theta_x - v'' \theta_x) \\ & + EI_Y(\beta_Z \theta_x'^2 - w'' + w'' \theta_x^2 + v'' \theta_x) \end{aligned} \quad (68d)$$

while M_t and M_w are the torsional and warping moments at the boundaries of the beam, respectively, given as

$$\begin{aligned} M_{bt} = & GI_t \theta_x' - EC_S \theta_x''' \\ & + N(-z_C w' \theta_x + y_C w' - y_C v' \theta_x - z_C v' + \frac{I_p}{A} \theta_x') \\ & + EI_Z \beta_Y \cdot (-2\theta_x' v'' - 2\theta_x' w'' \theta_x) \\ & + EI_Y \beta_Z (-2\theta_x' w'' + 2\theta_x' v'' \theta_x) + \frac{1}{2} E \left(I_{pp} - \frac{I_p^2}{A} \right) \theta_x'^3 \end{aligned} \quad (69a)$$

$$M_{bw} = -EC_S (\theta_x'' + \beta_\omega \theta_x'^2) \quad (69b)$$

Finally, α_k , β_k , $\bar{\beta}_k$, γ_k , $\bar{\gamma}_k$, δ_k , $\bar{\delta}_k$ ($k = 1, 2, 3$) are time-dependent functions specified at the end cross sections of the beam ($x = 0, l$) as described in section “[Statement of the Problem](#).”

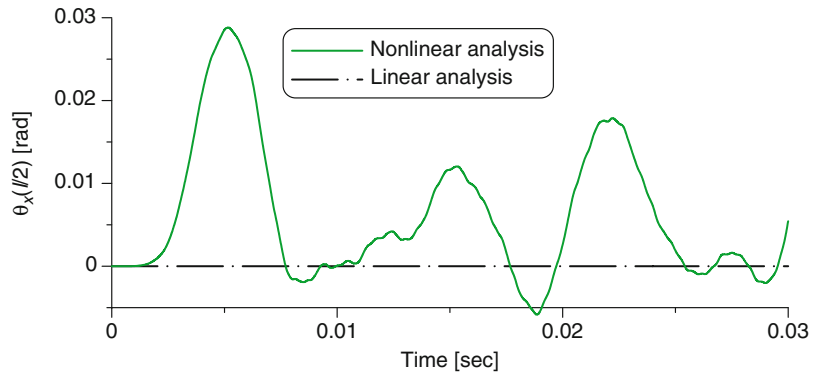
Applications

Forced Vibrations or Beam of Hollow Rectangular Cross Section Under Biaxial Bending

In the first application, in order to demonstrate the strong coupling between flexure and torsion, the beam of application section “[Forced Vibrations or Beam of Hollow Rectangular Cross Section Under Biaxial Bending](#)” ($I_p = 3.16407 \times 10^{-5} \text{m}^4$, $I_{pp} = 3.94264 \times 10^{-7} \text{m}^6$, $I_t = 2.10844 \times 10^{-5} \text{m}^4$, $C_S = 3.59634 \times 10^{-9} \text{m}^6$) is reexamined employing the governing differential equations derived in this section for the most general case (Eqs. 58, 59, 60, 61, 62, 63, 64, 65, 66, and 67). The time histories of $v(l/2, t)$, $w(l/2, t)$ evaluated employing AEM to solve the differential equations of motion (Sapountzakis and Dikaros 2011) occurred identical with the ones presented in Fig. 6 (case without shear deformation). However, what is of interest in this case

Nonlinear Dynamic Seismic Analysis,

Fig. 23 Time history of the angle of twist θ_x at the midpoint of the clamped beam of application section “Forced Vibrations or Beam of Hollow Rectangular Cross Section Under Biaxial Bending”



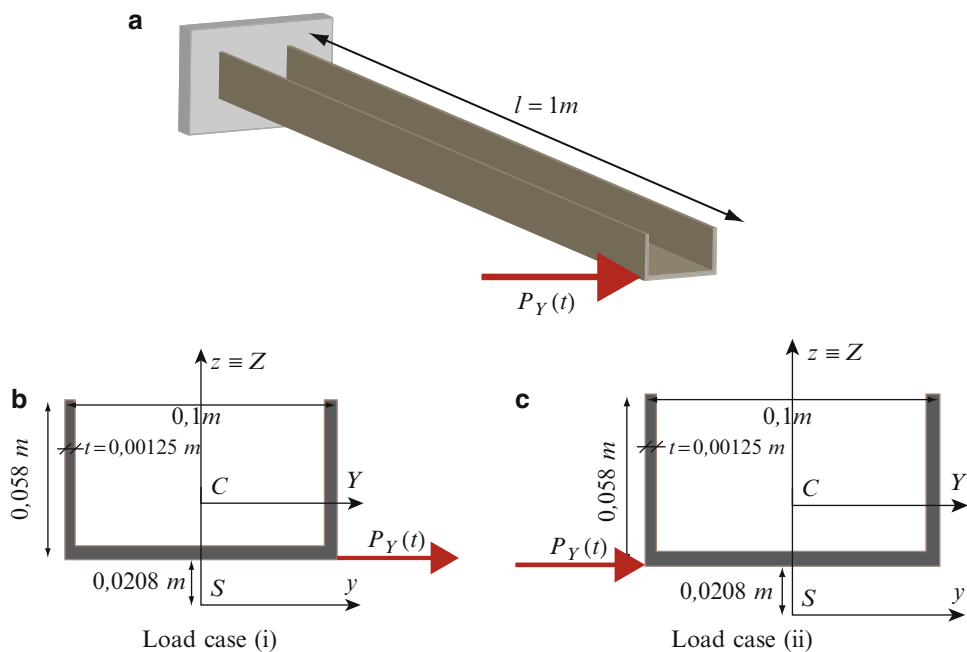
is the fact that even though the beam is of a doubly symmetric cross section and loaded on its centroid (i.e., no twisting loading is applied), torsional deformation is developed. In Fig. 23, the time history of the angle of twist $\theta_x(l/2, t)$ is presented. As it can be observed, this phenomenon cannot be predicted from linear analysis.

Forced Vibrations or Beam of Monosymmetric Cross Section Under Eccentric Transverse Loading
In this application, in order to investigate the response of a monosymmetric beam and the influence of the loading point upon the cross section, in nonlinear flexural–torsional vibrations, the forced vibrations of a cantilever beam ($E = 2.164 \times 10^8 \text{ kN/m}^2$, $G = 8.0148 \times 10^7 \text{ kN/m}^2$, $\rho = 7.85 \text{ tn/m}^3$, $l = 1 \text{ m}$) of a thin-walled open-shaped cross section (Fig. 24), under two load cases, have been studied (its geometric constants are given in Table 9). More specifically, the beam is subjected to a suddenly applied concentrated force $P_Y(t) = 5 \text{ kN}$ either on the right (load case (i), Fig. 24b) or on the left (load case (ii), Fig. 24c) flange. In Fig. 25, the time histories of the axial displacement $u(l, t)$, the transverse displacements $v(l, t)$, $w(l, t)$, and the angle of twist $\theta_x(l, t)$ of the cantilever beam, respectively, and in Table 10 the maximum values of these kinematical components are presented. From the aforementioned figures and table, it can easily be observed that geometrical nonlinearity affects substantially

the dynamic response of the beam inducing non-vanishing axial displacement and displacement with respect to the z axis, while in linear analysis, these kinematical components vanish. From the obtained results, it can also be verified that the loading position has significant influence on the response, altering substantially the magnitude of the kinematical components. This discrepancy can be explained by the fact that in load case (ii), the change of eccentricity of the transverse load during torsional rotation increases the magnitude of the twisting moment, acting adversely compared with load case (i), where the applied twisting moment is reduced during torsional rotation.

Prismatic Beam of Asymmetric Cross Section Under Harmonic Excitation

In this application, the forced vibrations of a hinged steel L-shaped beam of unequal legs (asymmetric cross section) (Fig. 26a) ($E = 2.1 \times 10^8 \text{ kN/m}^2$, $\rho = 7.85 \text{ tn/m}^3$, $G = 8.0769 \times 10^7 \text{ kN/m}^2$, $l = 1.0 \text{ m}$), having the geometric constants presented in Table 11, are studied. The beam is subjected to a uniformly distributed harmonic loading $p_z(x, t) = p_0(x) \sin(\omega_{f_{lin}} t)$, $p_0(x) = 10 \text{ kN/m}$ as this is shown in Fig. 26. The frequency $\omega_{f_{lin}}$ is considered as $\omega_{f_{lin}} = 2\pi f_{lin}$, where $f_{lin} = 118.2 \text{ Hz}$ is the first natural frequency of the examined beam, performing a linear analysis (Sapountzakis and Dourakopoulos 2009c). Due to lack of



Nonlinear Dynamic Seismic Analysis, Fig. 24 Cantilever beam of application section “**Forced Vibrations or Beam of Monosymmetric Cross**”

Section Under Eccentric Transverse Loading” (a). Transverse force applied on the *right* (b) or on the *left* (c) flange

Nonlinear Dynamic Seismic Analysis, Table 9 Geometric constants of the beam of example 2

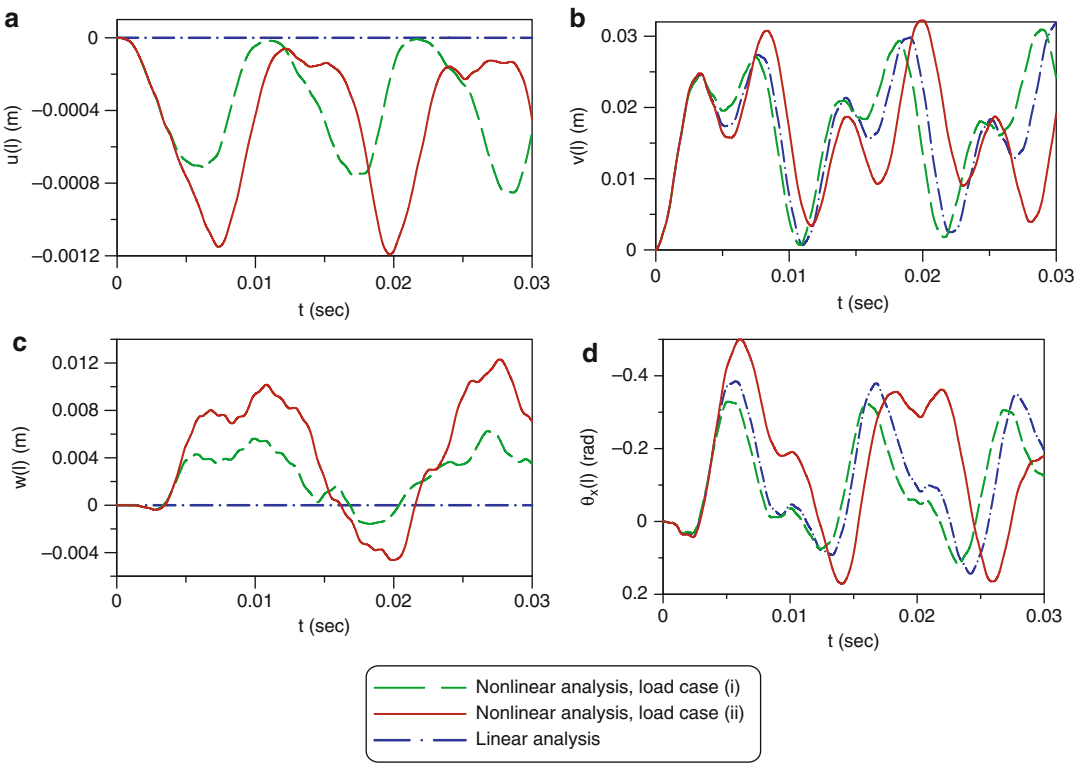
$A = 2.66875 \times 10^{-4} \text{ m}^2$	$I_Y = 9.39789 \times 10^{-8} \text{ m}^4$	$I_Z = 4.50061 \times 10^{-7} \text{ m}^4$
$I_p = 9.06833 \times 10^{-7} \text{ m}^4$	$I_t = 9.10243 \times 10^{-9} \text{ m}^4$	$C_s = 1.31047 \times 10^{-10} \text{ m}^6$
$I_{pp} = 4.58807 \times 10^{-9} \text{ m}^6$	$\beta_z = 6.10287 \times 10^{-2} \text{ m}$	$z_c = 3.687 \times 10^{-2} \text{ m}$

symmetry, the beam is expected to develop transverse displacements as well as angle of twist. In Fig. 27, the time histories of the displacements $\tilde{v}(l/2, t)$, $\tilde{w}(l/2, t)$, with respect to the $S\tilde{x}\tilde{y}\tilde{z}$ system of axes and the angle of twist $\theta_x(l/2, t)$, are presented, noting the significant difference in response between linear and nonlinear analyses. More specifically, it is observed that only in the linear response deformations continue to increase with time, while the beating phenomenon observed in the nonlinear one is explained from the fact that large kinematical components increase the beam’s fundamental natural frequency ω_f (by increasing the stiffness of

the beam due to the tensile axial force induced by the axially immovable ends), thereby causing a detuning of ω_f with the frequency of the external loading (ω_{fin}). After the kinematical components reach their maximum values, the amplitude of deformations decreases, leading to the reversal of the previously mentioned effects.

Summary

The main conclusions that can be drawn from this investigation are:



Nonlinear Dynamic Seismic Analysis, Fig. 25 Time history of the axial u (a), transverse v (b), w (c), displacements and angle of twist θ_x (d) at the tip of the cantilever beam of application section “[Forced Vibrations or Beam of Monosymmetric Cross Section Under Eccentric Transverse Loading](#)”

Nonlinear Dynamic Seismic Analysis, Table 10 Maximum values of the kinematical components $u(l, t)$ (m), $v(l, t)$ (m), $w(l, t)$ (m), and $\theta_x(l, t)$ (rad) of the cantilever beam of application section “[Forced Vibrations or Beam of Monosymmetric Cross Section Under Eccentric Transverse Loading](#)” for load cases (i), (ii)

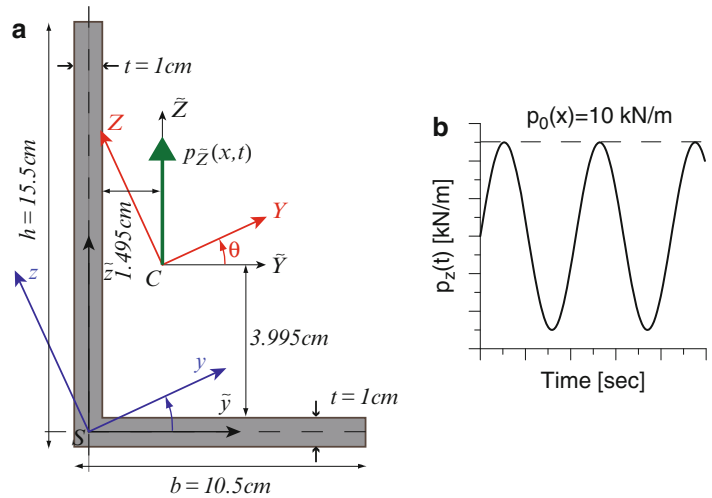
	Linear analysis	Nonlinear analysis	
		Load case (i)	Load case (ii)
$u(l)_{\max}$	0.00000	−0.00085	−0.00119
$v(l)_{\max}$	0.03190	0.03091	0.03217
$w(l)_{\max}$	0.00000	0.00626	0.01230
$\theta_x(l)_{\max}$	−0.38479	−0.32897	−0.50086

- a. In some cases, the effect of shear deformation is significant, especially for low beam slenderness values, increasing both the maximum transverse displacements and the calculated periods of the first cycle of motion.
- b. The discrepancy between the results of the linear and the nonlinear analysis is remarkable.
- c. The geometrical nonlinearity leads to coupling between the torsional and axial equilibrium equations and alters the modeshapes of vibration. Consequently, the initiation of

The coupling effect of the transverse displacements in both directions in the nonlinear analysis influences these displacements.

Nonlinear Dynamic Seismic Analysis,

Fig. 26 L-shaped cross section of unequal legs of application section “Prismatic Beam of Asymmetric Cross Section Under Harmonic Excitation” (a) and transverse harmonic excitation (b)



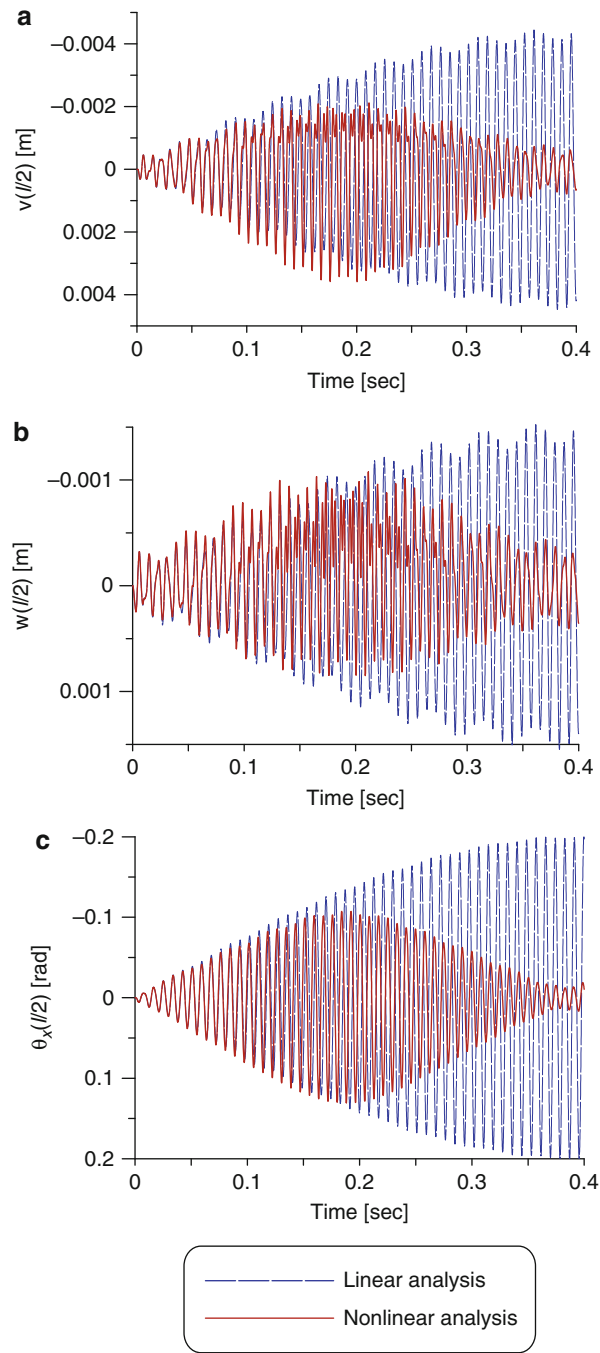
Nonlinear Dynamic Seismic Analysis, Table 11 Geometric constants of the asymmetric beam of application section “Prismatic Beam of Asymmetric Cross Section Under Harmonic Excitation”

$A = 2.500\text{E} - 3\text{ m}^2$	$I_t = 8.3903\text{E} - 8\text{ m}^4$	$y_C = 3.6620\text{E} - 2\text{ m}$
$I_Y = 7.2359\text{E} - 6\text{ m}^4$	$C_S = 1.1937\text{E} - 10\text{ m}^6$	$z_C = 3.1900\text{E} - 2\text{ m}$
$I_Z = 1.3220\text{E} - 6\text{ m}^4$	$\beta_Y = 8.1915\text{E} - 2\text{ m}$	$\theta = 4.3000\text{E} - 1\text{ rad}$
$I_p = 1.4552\text{E} - 5\text{ m}^4$	$\beta_Z = 3.8866\text{E} - 2\text{ m}$	
$I_{pp} = 1.6873\text{E} - 7\text{ m}^6$	$\beta_\omega = -1.6050\text{E} - 1$	

- small-amplitude free vibrations of buckled beams with the linear fundamental modeshape as initial twisting rotations induces higher harmonics in the beam’s response, but its fundamental natural frequency is only slightly affected.
- d. Large twisting rotations have a profound effect on both the positions around which vibrations are performed and the fundamental natural frequency of buckled beams undergoing large-amplitude free vibrations. The computation of dynamic characteristics of buckled beams differs from the one of beams at a pre-buckled state where the increase of initial amplitude of vibration always leads to an increase of the fundamental natural frequency.
- e. The natural frequency of the axial displacement’s response of buckled beams undergoing large-amplitude free vibrations is twice as much as the one of the twisting rotation, when vibrations are performed around the static equilibrium position of the pre-buckling configuration.
- f. As expected, geometrical nonlinearity bounds the (twisting) deformations of beams at a pre-buckled state subjected to primary resonance excitations. A beating phenomenon is observed in the time histories of kinematical and stress components.
- g. STMDE affects negligibly the kinematical and stress components of beams of open-shaped thin-walled cross sections undergoing free or forced vibrations of small or large amplitude.
- h. STMDE affects the kinematical components of beams of closed-shaped cross sections

**Nonlinear Dynamic
Seismic Analysis,**

Fig. 27 Time history of the displacement $\tilde{v}$ (a), $\tilde{w}$ (b) and of the angle of twist θ_x (c) at the midpoint of the hinged–hinged beam of application section “Prismatic Beam of Asymmetric Cross Section Under Harmonic Excitation”



undergoing linear vibrations. Its effect is much more pronounced on stress components, concluding that it cannot be neglected in linear dynamic analysis of beams of such cross sections.

i. The strong coupling between flexure and torsion may lead beams of doubly symmetric cross sections, undergoing biaxial transverse loading applied on their centroid, to develop torsional rotation.

- j. The eccentricity change of the transverse loading during the torsional beam motion, resulting in additional torsional moment influences the beam response.

Cross-References

- [Nonlinear Finite Element Analysis](#)
- [Seismic Analysis of Concrete Bridges: Numerical Modeling](#)
- [Seismic Analysis of Steel and Composite Bridges: Numerical Modeling](#)
- [Seismic Analysis of Steel Buildings: Numerical Modeling](#)
- [Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling](#)

References

- Aristizabal-Ochoa JD (2004) Timoshenko beam-column with generalized end conditions and nonclassical modes of vibration of shear beams. *ASCE J Eng Mech* 130(10):1151–1159
- Brigham E (1988) *Fast Fourier transform and its applications*. Prentice Hall, Englewood Cliffs
- Foda MA (1999) Influence of shear deformation and rotary inertia on nonlinear free vibration of a beam with pinned ends. *Comput Struct* 71:663–670
- Kausel E (2002) Nonclassical modes of unrestrained shear beams. *ASCE J Eng Mech* 128(6):663–667
- Mohri F, Azrar L, Potier-Ferry M (2004) Vibration analysis of buckled thin-walled beams with open sections. *J Sound Vib* 275:434–446
- Murín J, Kutis V (2008) An effective finite element for torsion of constant cross-sections including warping with secondary torsion moment deformation effect. *Eng Struct* 30:2716–2723
- Ramm E, Hofmann TJ (1995) *Stabtragwerke, Der Ingenieurbau*. In: Mehlhorn G (ed) *Band Baustatik/Baudynamik*. Ernst & Sohn, Berlin
- Rao GV, Raju IS, Kanaka Raju K (1976) Nonlinear vibrations of beams considering shear deformation and rotary inertia. *AIAA J* 14(5):685–687
- Rothert H, Gensichen V (1987) *Nichtlineare Stabstatik*. Springer, Berlin
- Sapountzakis EJ, Dikaros IC (2011) Non-linear flexural-torsional dynamic analysis of beams of arbitrary cross section by BEM. *Int J Non-Linear Mech* 46:782–794
- Sapountzakis EJ, Dourakopoulos JA (2009a) Nonlinear dynamic analysis of Timoshenko beams by BEM. Part I: Theory and numerical implementation. *Nonlinear Dyn* 58:295–306
- Sapountzakis EJ, Dourakopoulos JA (2009b) Nonlinear dynamic analysis of Timoshenko beams by BEM. Part II: Applications and validation. *Nonlinear Dyn* 58:307–318
- Sapountzakis EJ, Dourakopoulos JA (2009c) Shear deformation effect in flexural-torsional vibrations of beams by BEM. *Acta Mech* 203:197–221
- Sapountzakis EJ, Tsipiras VJ (2010a) Nonlinear nonuniform torsional vibrations of bars by the boundary element method. *J Sound Vib* 329:1853–1874
- Sapountzakis EJ, Tsipiras VJ (2010b) Warping shear stresses in nonlinear no uniform torsional vibrations of bars by BEM. *Eng Struct* 32:741–752
- Sapountzakis EJ, Tsipiras VJ (2010c) Shear deformable bars of doubly symmetrical cross section under nonlinear nonuniform torsional vibrations—application to torsional postbuckling configurations and primary resonance excitations. *Nonlinear Dyn* 62:967–987
- Thomson WT (1981) *Theory of vibration with applications*. Prentice Hall, Englewood Cliffs
- Tsipiras VJ, Sapountzakis EJ (2012) Secondary torsional moment deformation effect in inelastic nonuniform torsion of bars of doubly symmetric cross section by BEM. *Int J Non-Linear Mech* 47(4):68–84
- Vlasov V (1963) *Thin-walled elastic beams*. Israel Program for Scientific Translations, Jerusalem

Nonlinear Finite Element Analysis

Charis J. Gantes

Institute of Steel Structures, School of Civil Engineering, National Technical University of Athens, Zografou Campus, Athens, Greece

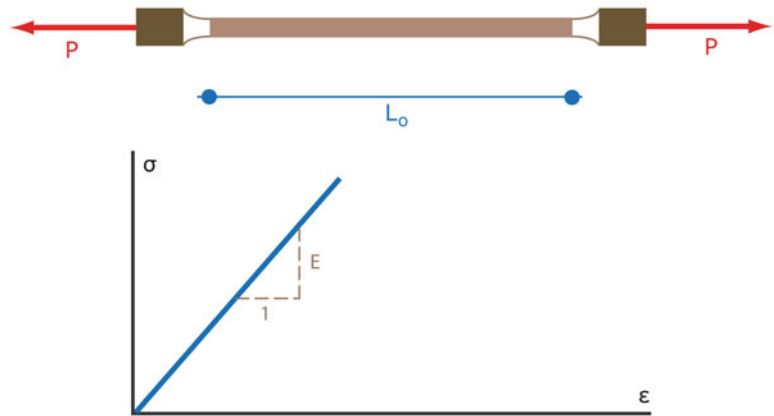
Synonyms

Advanced numerical analysis; Geometrically and material nonlinear analysis with imperfections; Numerical collapse prediction

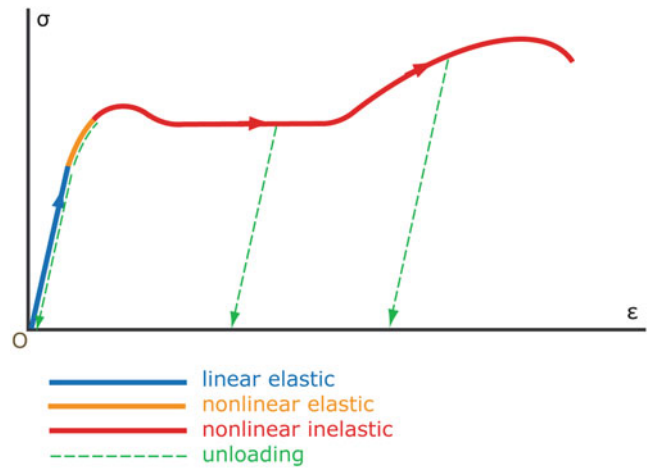
Introduction into Basic Concepts of Linear and Nonlinear Behavior

The aim of this contribution is to present the state of the art of nonlinear finite element analysis tools for predicting the collapse load of structures, hence their strength. Reliable evaluation

Nonlinear Finite Element Analysis, Fig. 1 Tension test and linear material behavior



Nonlinear Finite Element Analysis, Fig. 2 Nonlinear material behavior



of collapse load is a powerful tool in structural design in general and in earthquake-resistant design in particular. Collapse may occur due to material nonlinearity, geometric nonlinearity, or a combination of both types of nonlinearity. Some basic concepts of linear and nonlinear structural behavior are therefore briefly reviewed first. For a more in-depth understanding of the associated concepts, the reader should consult pertinent text books (e.g., Bazant and Cedolin 1991).

In classical structural analysis materials are assumed to behave in a linear manner. This behavior is commonly described by a linear stress-strain relation obtained by means of a tension test, as shown in Fig. 1. Stress σ , defined as the applied tensile force P divided by the cross-sectional area A , is then proportional to strain ϵ ,

which is equal to the elongation ΔL of the bar divided by its undeformed length L_0 . This is commonly known as Hooke's law and the proportionality ratio is the modulus of elasticity E of the material:

$$\sigma = E \cdot \epsilon \tag{1}$$

In case the applied load is increased beyond a certain level, the stress-strain relation is no longer proportional. The material behavior then becomes nonlinear, as illustrated in Fig. 2, which provides an out-of-scale description of a tension test for a steel specimen by means of the stress-strain curve, frequently referred also as the material's constitutive law. The green lines denote unloading and may coincide with the corresponding loading curve, in which case the

response is called elastic, or not, and then the response is inelastic. Thus, the behavior is linear in the blue part of the curve in Fig. 2, nonlinear elastic in the orange part, and nonlinear inelastic in the red part. The nonlinear response of common structural materials is characterized by a sharp reduction of stiffness, corresponding to a decrease of the slope of the stress-strain curve.

The ability of a material to deform well into the inelastic range without rupture is called ductility. It is directly related to the material's ability to absorb energy when it is subjected to extreme actions and is a very desirable feature for good seismic behavior. Steel is such a material; thus, a key objective of structural design in seismic regions is to configure layout and detailing of steel and reinforced concrete structures in such a way that material failure rather than other undesirable failure modes is critical, so that proper advantage of its ductility is taken.

In engineering practice, the nonlinear material behavior is commonly described in a simplified manner. Two common such idealizations are the bilinear ones, either elastic-perfectly plastic or elastic-plastic with linear hardening, shown in Fig. 3. The stress corresponding to the end of the linear region is called yield stress of the material and is denoted by f_y , while the maximum stress, also called ultimate stress, is denoted by f_u . The corresponding strains are ϵ_y and ϵ_u , and the difference between them is a measure of the material's ductility. In the elastic-plastic idealization, the higher value of f_u compared to f_y is conservatively ignored.

Thus, by material nonlinearity, we refer to cases where a structure is loaded to such levels that at certain cross sections of certain members, the strains exceed ϵ_y and Hooke's law no longer applies. Material nonlinearity is commonly a critical failure mode for "stocky" structures (Fig. 4), consisting of short members with respect to their cross section and thick sections compared to their overall dimensions.

In order to explain the concept of geometric nonlinearity, it is reminded that in classical structural analysis it is commonly assumed that equilibrium equations can be written in the structure's undeformed configuration. This is theoretically

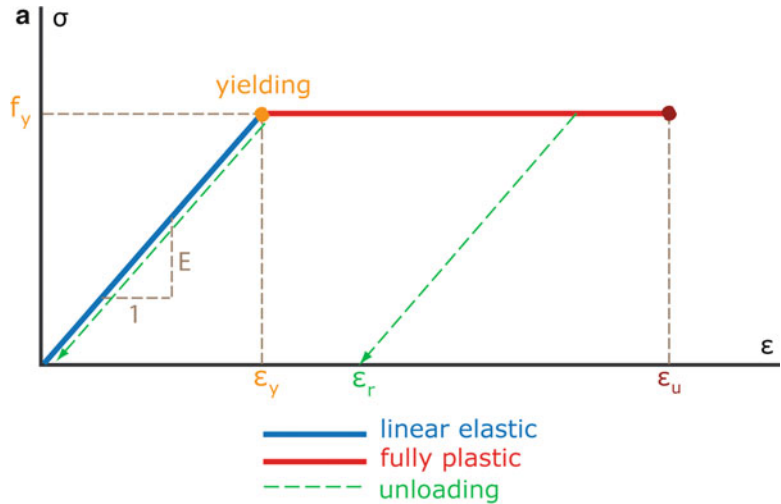
wrong, as equilibrium of a loaded structure is actually achieved at its deformed configuration. However, in most cases engineering structures possess sufficient stiffness, so that the difference between undeformed and deformed configuration is small and can be safely ignored. As a result, action (usually external loads) and effect (e.g., deformations or internal forces and moments) are proportionally related, as shown in Fig. 5, for a simply supported beam. This type of response is known as geometrically linear. The above assumption simplifies structural analysis significantly and is a prerequisite for the applicability of the principle of superposition, which is at the core of most widely used analysis methods for everyday structures.

Some structures, however, are more flexible, so that the difference between their undeformed and deformed configuration is no longer negligible. One such example is the shallow, two-member truss of Fig. 6, known in the literature as von Mises truss, which exhibits gradual stiffness reduction as applied load increases, eventually leading to a so-called limit point or critical point corresponding to maximum load and zero stiffness.

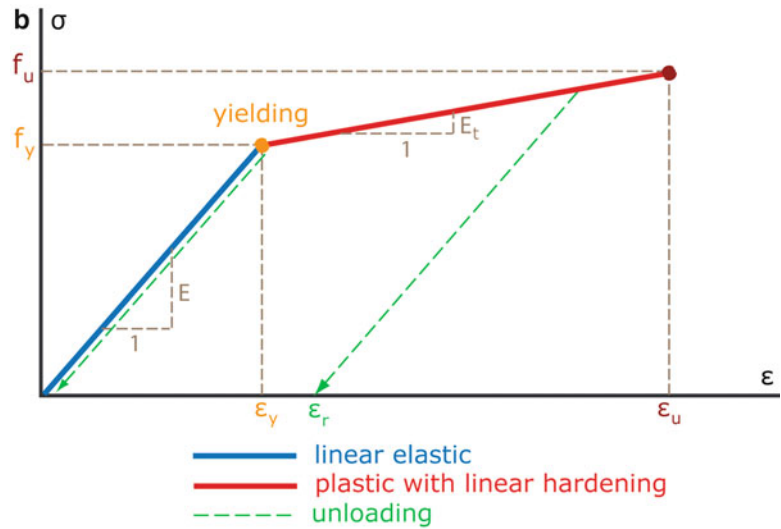
This type of structural behavior is called geometrically nonlinear and invalidates the use of the principle of superposition, thus of all commonly used structural analysis methods. The so-called nonlinear analysis methods are required, often involving iterations to achieve convergence, as equilibrium equations must now be formulated in the deformed geometry, which however is unknown. Thus, nonlinear analysis is a lot more time and computationally demanding than linear ones. Typical examples of structures that are prone to geometric nonlinearity are cables subjected to transverse loads, shallow arches and domes, slender bars that buckle under axial compression, and thin-walled shells (Fig. 7).

It is noted here that almost all structures behave linearly for low load levels and nonlinearly above a certain load level. Moreover, collapse of most structures is associated with combined material and geometric nonlinearity. As noted above, stocky structures (structures of small slenderness) are more susceptible to

Nonlinear Finite Element Analysis, Fig. 3 Common idealizations of nonlinear material behavior



Elastic-perfectly plastic behavior



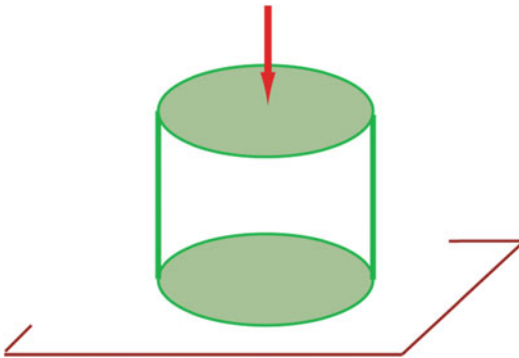
Elastic-plastic behavior with linear hardening

material nonlinearity; therefore, collapse initiation is governed by material nonlinearity. Then, as collapse progresses, stiffness decreases very rapidly, large deformations and strains occur, and therefore geometric nonlinearity develops as well, as illustrated in Fig. 8, where linear behavior is denoted with blue color, material nonlinearity with orange, and combined material and geometric nonlinearity with red.

Slender structures on the other hand are flexible and exhibit large deformations; thus, they are more susceptible to geometric nonlinearity;

therefore, collapse initiation is due to geometric nonlinearity. As collapse progresses, stiffness decreases further very rapidly, even larger deformations and strains occur, and therefore material nonlinearity develops as well. This type of behavior is described in Fig. 9, where linear behavior is denoted with blue color, geometric nonlinearity with green, and combined material and geometric nonlinearity with red.

Moreover, as structures enter into their nonlinear range, their behavior is affected by the unavoidable presence of initial imperfections,



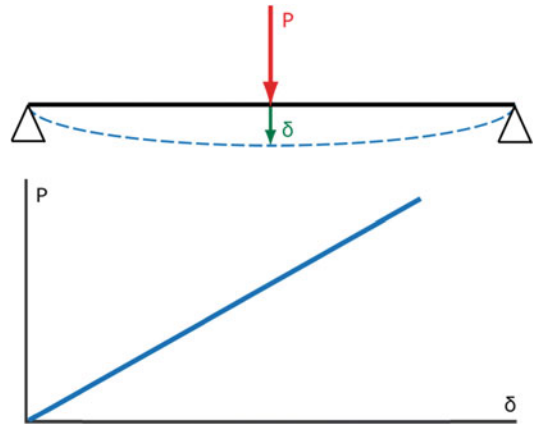
Nonlinear Finite Element Analysis, Fig. 4 Stocky structure that is prone to material nonlinearity

including geometric irregularities, material and cross section inhomogeneities, load eccentricities, etc. This is illustrated in Fig. 10 for the example of a shallow arch subjected to a vertical concentrated load at its crown, where the presence of imperfections drastically reduces the arch's strength, depending also on their magnitude w_0 .

It is thus concluded that the capacity of a structure depends on its sensitivity to material nonlinearity, geometric nonlinearity, and initial imperfections. Therefore, all these factors should be taken into account in order to predict the structure's collapse, thus evaluate its strength.

Treatment of Nonlinear Behavior in Engineering Practice

In recent years structural design codes have adopted limit state design (LSD), replacing the older concept of allowable stress design (ASD). Limit state design requires the structure to satisfy two principal criteria: the ultimate limit state (ULS) and the serviceability limit state (SLS). A limit state is a set of performance criteria that must be met when the structure is subject to loads. To satisfy the ultimate limit state, the structure must not collapse when subjected to any of the pertinent combinations of its design loads. To satisfy the serviceability limit state criteria, a structure must remain functional for its intended use subject to routine, everyday loading.



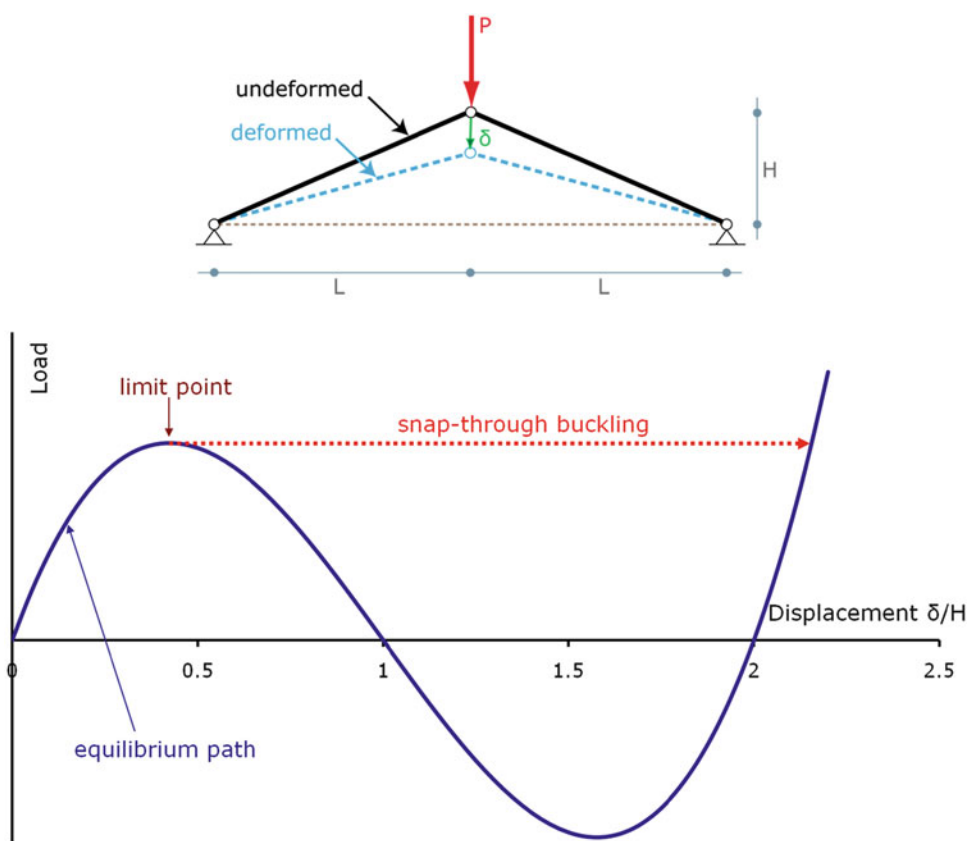
Nonlinear Finite Element Analysis, Fig. 5 Linear response of a simply supported beam

The prevailing approach proposed in modern codes for carrying out ultimate limit state checks consists of obtaining action effects by means of linear elastic analysis of the structure subjected to design loads and comparing them to resistances that account for both types of eventual nonlinearity, material, and geometric. Thus, prediction of collapse, a highly nonlinear phenomenon, is accomplished by means of linear elastic analysis.

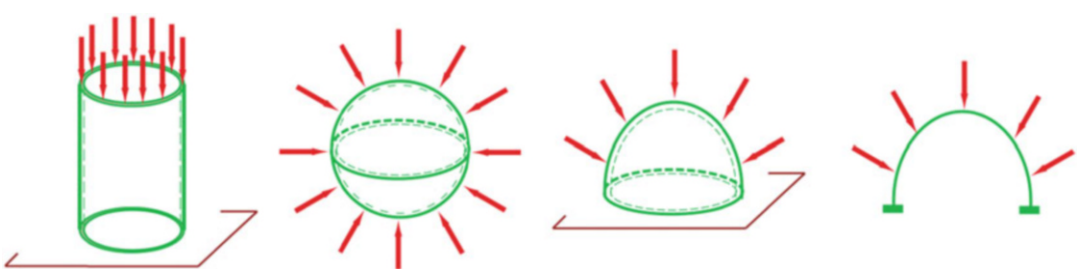
In order to evaluate strength, modern codes assume full exploitation of material capacity and appropriate reduction factors, by means of which geometric nonlinearities and imperfections are taken into account. For example, the strength of a steel bar subjected to compression is given by Eurocode 3 (European Committee for Standardisation 2004) as:

$$N_{b,Rd} = \frac{\chi \cdot \beta \cdot A \cdot f_y}{\gamma_{M1}} \quad (2)$$

In this relation, A is the cross-sectional area and f_y the material's yield strength; thus, the product $A \cdot f_y$ represents the cross section's axial strength and addresses the issue of material nonlinearity, adopting an elastic-perfectly plastic idealization of the constitutive law. On the other hand, χ is a reduction factor accounting for global buckling, obtained from the so-called buckling curves, which have been calibrated for



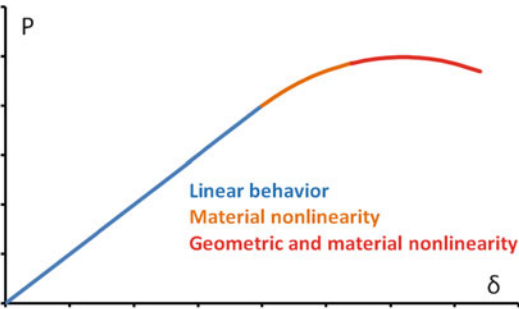
Nonlinear Finite Element Analysis, Fig. 6 Shallow two-bar truss and its geometrically nonlinear response



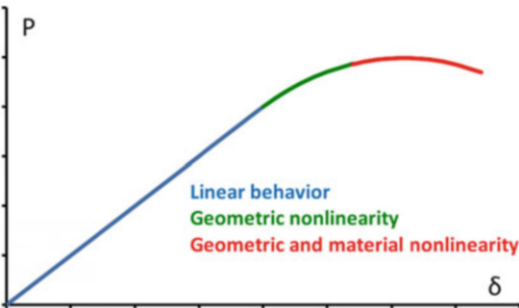
Nonlinear Finite Element Analysis, Fig. 7 Slender structures that are prone to geometric nonlinearity

specific types of cross sections by means of a wide range of experimental results, and incorporates the effects of initial imperfections and residual stresses, while β is a reduction factor accounting for local buckling, depending on the classification of the cross section into appropriate slenderness categories. Thus, the product $\chi \cdot \beta$ addresses the issues of geometric nonlinearity and imperfections.

The above approach offers significant advantages, as it is straightforward and the prediction of collapse is accomplished by means of linear elastic analysis. This means that it can be carried out using widely available commercial software, is computationally inexpensive, and does not require special expertise of the engineer in setting up the structural model, selecting the analysis parameters, and interpreting the results.



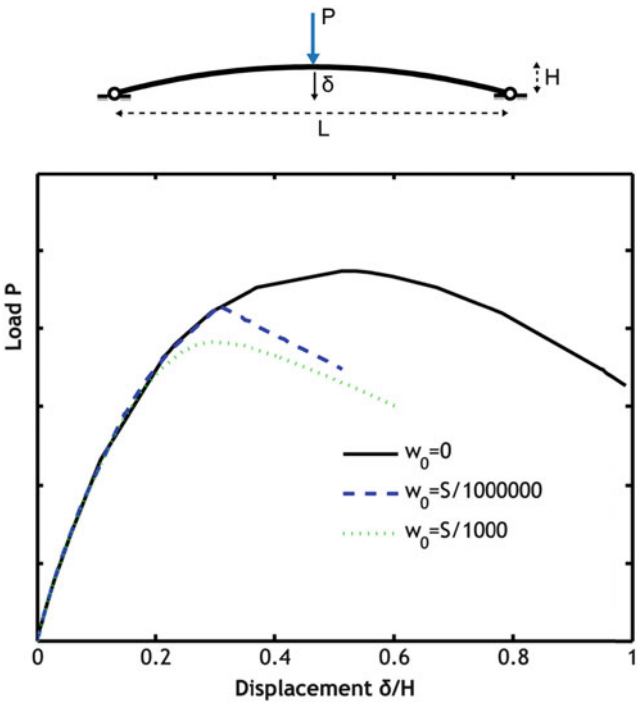
Nonlinear Finite Element Analysis, Fig. 8 Typical load-displacement curve of stocky structures



Nonlinear Finite Element Analysis, Fig. 9 Typical load-displacement curve of slender structures

Nonlinear Finite Element Analysis,

Fig. 10 Influence of imperfections on strength of shallow arch



Furthermore, the results are quite reliable for ordinary structural systems.

Although this approach is appropriate for the design of ordinary structural systems, more complex structures with unusual geometry or cross sections require more accurate methods of their strength evaluation, as such cases are not directly covered by buckling curves and interaction equations between different action components, while simplifying assumptions that are made in an effort to approximate their behavior may lead to significant inaccuracies. To address such cases, modern structural design codes allow predicting collapse by means of more elaborate, nonlinear numerical analysis, taking also initial imperfections into account.

Prediction of collapse of structures can thus be accomplished by means of such nonlinear numerical analysis, making use of commercially available finite element software, setting up an appropriate finite element model, obtaining critical buckling modes from Linearized Buckling Analysis (LBA), and then using a linear combination of these modes as imperfection pattern for a Geometrically and Material Nonlinear

Imperfection Analysis (GMNIA). Equilibrium paths accompanied by snapshots of deformation and stress distribution at characteristic points are then used for evaluating the results of the GMNIA analysis. This approach is particularly applicable to seismic design, as all major seismic design codes allow structures to behave inelastically during major earthquakes, defining “no collapse” as the design objective.

In the following, practical details of the implementation of this approach are discussed, and pertinent case studies are presented. It is thus aimed to contribute towards making advanced numerical analysis methods accessible to practicing structural engineers, by providing guidelines for using such methods for structural design and by demonstrating some of the capabilities afforded by such methods to the structural engineering community.

Overview of Linearized Buckling and Nonlinear Numerical Algorithms

The geometric and material nonlinear phenomena characterizing structures near collapse are very complicated in nature and require, in most cases, highly sophisticated numerical tools for their simulation. A common feature of most nonlinear analysis algorithms is that the load is applied in successive steps and iterations are performed within each step to achieve convergence (Stricklin et al. 1973).

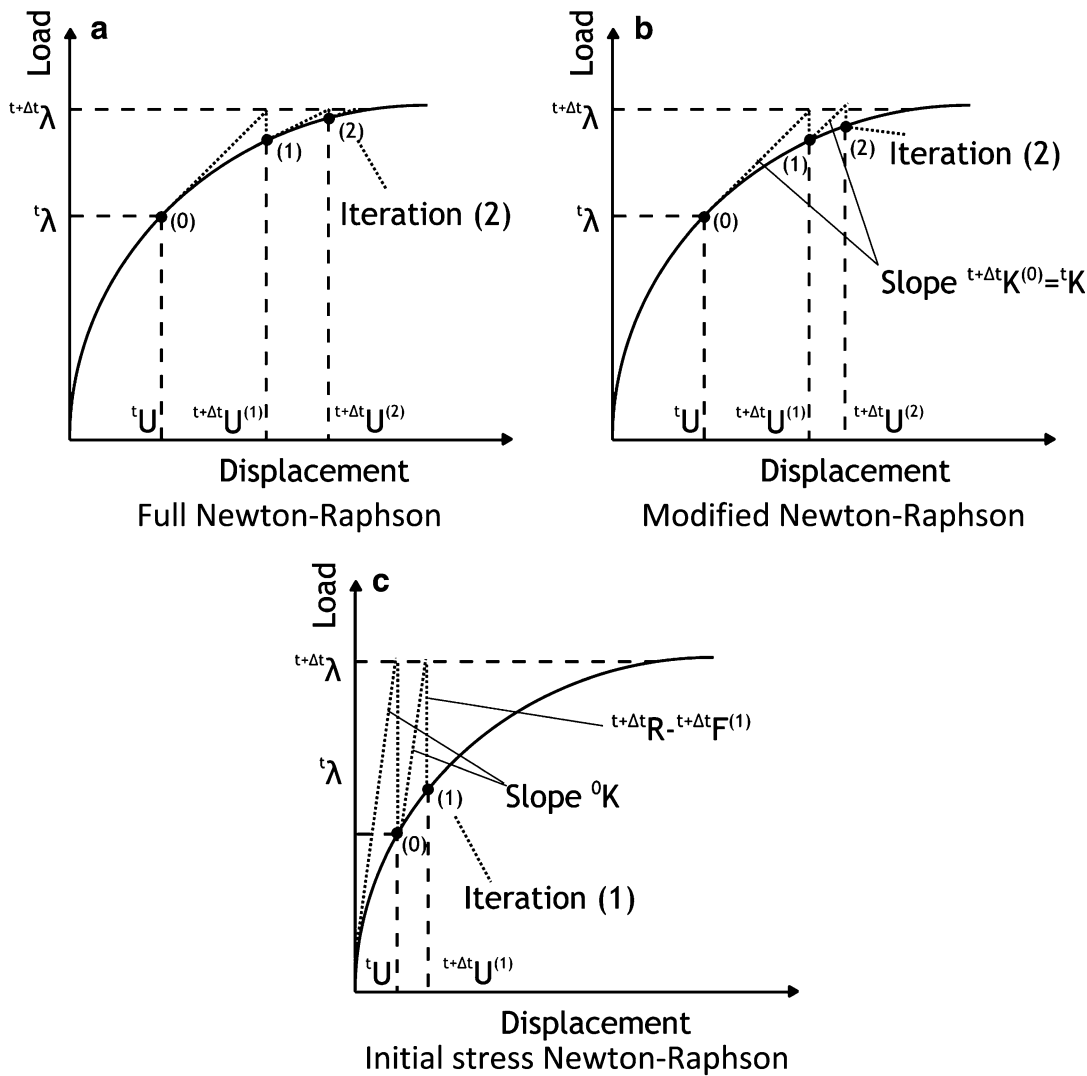
The simplest numerical method for a detailed geometrically and material nonlinear (GMN) analysis is the Newton-Raphson scheme (Crisfield 1979; Bathe 1995), which can be found in three forms: (i) the full Newton-Raphson, which is the most accurate, but also the most time consuming, since the tangent stiffness of the structure has to be calculated and factorized within each iteration in the solution procedure; (ii) the modified Newton-Raphson, which differs from the full Newton-Raphson in that the calculation and the factorization of the tangent stiffness matrix take place only in some iterations within each step, thus requiring in most cases a larger number of iterations per step but

a smaller computational effort per iteration; and (iii) the “initial stress” Newton-Raphson procedure, where the calculation of the tangent stiffness matrix is performed once, at the beginning of the analysis, and then kept the same throughout the analysis (Fig. 11).

Applying the Newton-Raphson method in a nonlinear finite element system will yield results only in the pre-collapse range, but it will fail to give information about the post-collapse response. To circumvent this limitation, a constraint can be added into the finite element system, which relates the load increment and the incremental displacements within each iteration (Fig. 12). This technique allows the calculation of the whole equilibrium path, even beyond the critical limit points. A number of different solution algorithms have been proposed in the literature (Riks 1979; Crisfield 1981; Ramm 1981; Bathe and Dvorkin 1983).

In case initial imperfections are also included in the numerical model, the above GMN analysis is denoted as GMNIA. Such analyses are often preceded by linearized buckling analyses (LBA) (Bathe 1995), which seeks deformed configurations of the structure that correspond to zero determinant of a characteristic stiffness matrix, assuming linear response beyond an initial configuration. Such deformed configurations are known as buckling modes and the corresponding loads as linear critical buckling loads. Critical buckling loads obtained by means of LBA are, in most cases, not safe predictions of strength. However, it is advisable that this type of analysis precedes the subsequent, more exact, nonlinear analyses, for several reasons (Dimopoulos and Gantes 2012):

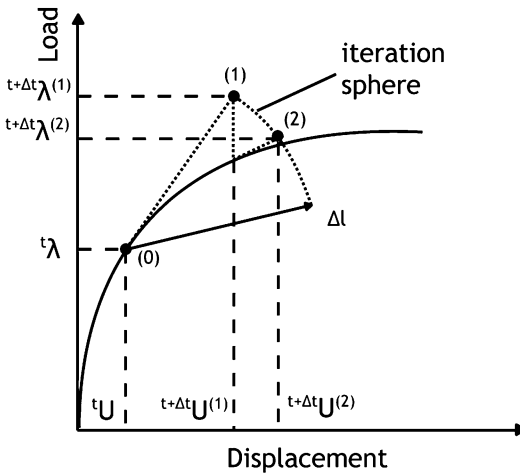
- (i) Critical buckling loads obtained by means of linearized buckling analyses are an initial indication, and in most cases an upper bound, of actual strength. Taking also into account the fact that this is a very fast and inexpensive type of analysis, it may be very useful as a tool for evaluating alternative solutions during preliminary design, before resorting to the much more time-consuming and expensive nonlinear analyses.



Nonlinear Finite Element Analysis, Fig. 11 Variations of the Newton-Raphson method

- (ii) Critical buckling loads are often needed for calculating nondimensional slenderness ratios to be used with buckling curves in the framework of code-based design procedures. Even though analytical expressions for these buckling loads are available for simple cases of geometry, loading patterns, and boundary conditions, in the majority of other cases, buckling loads must be obtained numerically.
- (iii) It has been shown that buckling analysis can be used in place of a geometrically

- nonlinear analysis for the calculation of the nonlinear stability limit of a number of structures by performing a number of successive buckling analyses on the unstressed and some stressed configurations of the structures (Chang and Chen 1986; Dimopoulos and Gantes 2012).
- (iv) Buckling modes obtained by means of linearized buckling analyses are commonly used as initial imperfections for geometrically and material nonlinear imperfection analyses. Such imperfections are necessary



in order to trigger all possible failure mechanisms and to make sure that the critical failure mechanism is captured by the nonlinear numerical analysis algorithm.

It should be noted that the choice of buckling modes as imperfection patterns is not unique and has in many cases been found to lead to lower compliance to experimental results than other shapes of initial imperfections. For the case of cylindrical shell structures, a type of structures that is particularly sensitive to imperfections, imperfections affine to the collapse mode of the perfect shell were found to be more unfavorable than ideal buckling modes (Schneider and Brede 2005; Schneider et al. 2005). Furthermore, fabrication processes are responsible for imperfection patterns, creating residual stress profiles in the body of the cross sections, which can have significant consequences on the ultimate bearing capacity of steel structures (Berry et al. 2000; Herynk et al. 2007). However, the choice of manufacturing-related imperfection patterns is usually dependent upon the structural type as well as the specific manufacturing method, cannot be easily generalized, and often lacks adequate data to be properly implemented. In case such information is available, it is, in general, preferable to use these imperfection patterns, as they more accurately represent real strength.

In the absence of such data, however, the use of buckling modes as imperfection shapes is considered to be the next best available choice.

Details of the formulation and implementation of these algorithms are beyond the scope of this presentation, and the reader should consult other specialized texts for that purpose. The engineers applying this methodology should have good knowledge of theoretical and practical aspects of the finite element method for linear and nonlinear structural applications (as outlined, e.g., by Crisfield et al. 1997; Bathe and Cimento 1980; Bathe et al. 1980; Bathe 1995; Belytschko et al. 2000; Kojic and Bathe 2004). Emphasis here is placed on the application of such methods to determine structural collapse, thus strength. For that purpose, the successive steps of a proposed numerical methodology are presented, followed by two case studies, referring to steel members with unconventional geometry that are part of two long-span roofs, which have been recently designed in Greece.

Numerical Methodology for Collapse Prediction

A state-of-the-art strategy for understanding the behavior, predicting all possible failure mechanisms, evaluating the strength, and assessing the vulnerability of structures based on advanced nonlinear numerical analyses consists of the following steps (Gantes and Fragkopoulos 2010; Gantes 2011; Gantes and Koulatsou 2013):

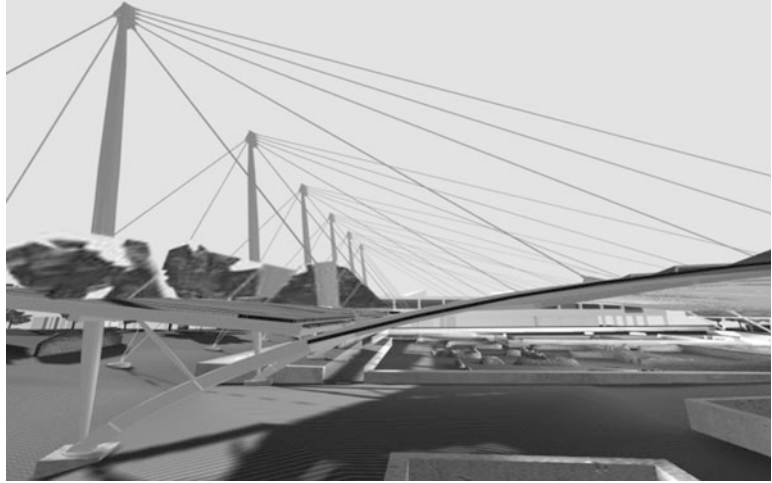
Step 1. Setting up an appropriate finite element model

In this step special attention should be paid to the following:

- (i) The finite element software, the specific elements used to model the structure, as well as the numerical solution algorithm employed for numerical analysis should be verified, by means of comparisons to analytical solutions and/or numerical solutions from the literature.
- (ii) The model should be able to predict all anticipated failure mechanisms.

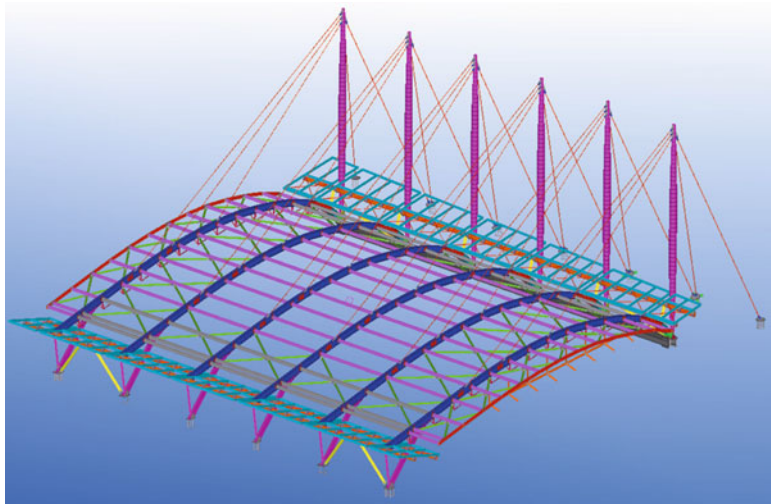
Nonlinear Finite Element Analysis,

Fig. 13 Architectural proposal for the steel roof over Aristotle's Lyceum



Nonlinear Finite Element Analysis,

Fig. 14 Perspective view of the suspended steel roof over Aristotle's Lyceum



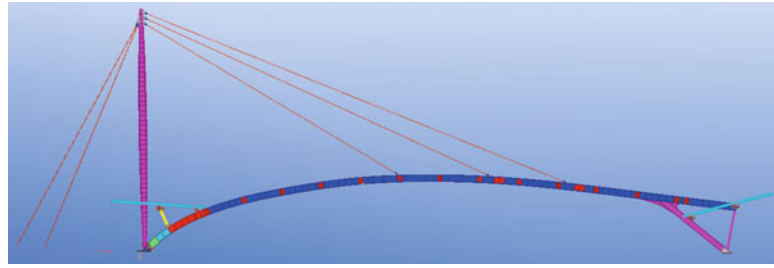
For example, modeling a steel column with beam elements is not appropriate for predicting failure where local or lateral buckling may be predominant. Shell or plate elements should be used instead, and the mesh should be fine enough to capture the curvature part of the cross section where local or lateral buckling occurs.

- (iii) Basic features of the model, such as element connectivity, boundary conditions, and loads, should be checked by means of simple, linear, static, and modal analyses. Sum of support reactions can be used to check load application. Mode

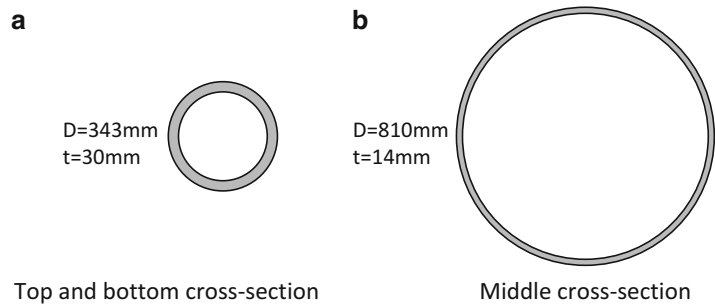
shapes and periods of vibration can be compared to expected values, in order to identify possible errors in mass or stiffness.

- (iv) The mesh density should be verified by performing a convergence study. Element size should be successively divided in half, and representative quantities of the response should be plotted against mesh density until sufficient convergence is achieved. It should be noted that a finer mesh may be required for subsequent nonlinear analyses than the one obtained from linear static

Nonlinear Finite Element Analysis, Fig. 15 View of a suspended main arch and the corresponding pylon of the steel roof over Aristotle's Lyceum



Nonlinear Finite Element Analysis, Fig. 16 Cross sections of pylon



convergence study. However, the latter one should be considered as a minimum.

Step 2. Carrying out linearized buckling analysis

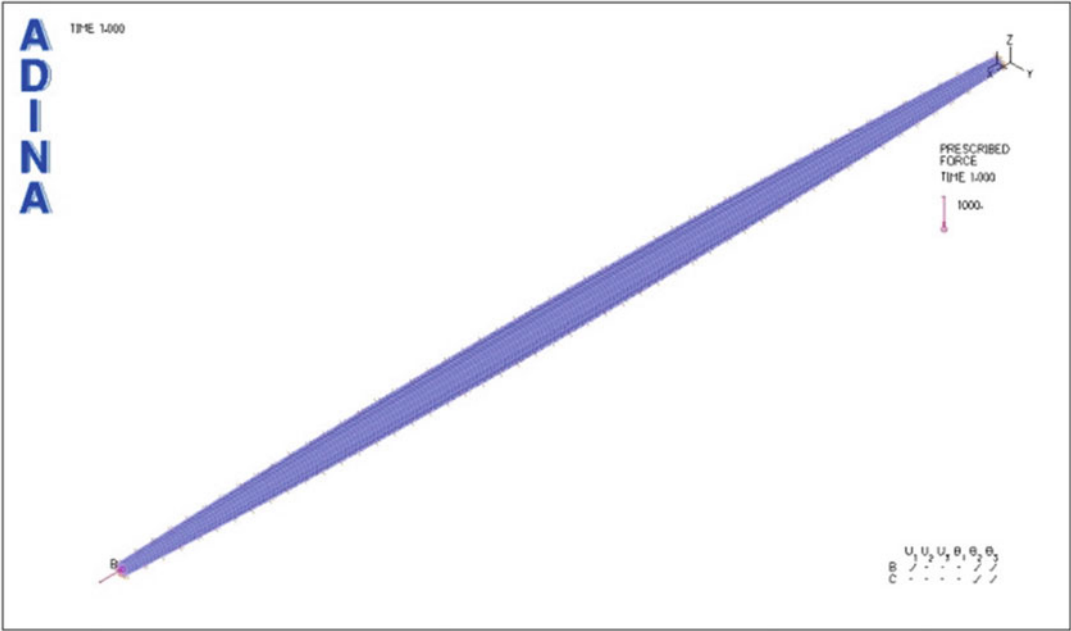
One question arising when carrying out such an analysis concerns the number of modes that should be evaluated. Even though the first buckling mode is the one corresponding to the lowest critical buckling load, it is, in general, not sufficient, for several reasons. A higher mode associated with an unstable post-buckling equilibrium path may be more critical than a lower one with a stable post-buckling equilibrium path, due to its sensitivity to imperfections. Even two modes with stable post-buckling equilibrium paths, but with nearly coinciding critical buckling loads, may interact in the presence of imperfections and lead to unstable post-buckling behavior (e.g., Bazant and Cedolin 1991; Livanou et al. 2013). Therefore, several buckling modes should be obtained. There is no clear-cut criterion as to their number; however, it is recommended to obtain sufficient number of modes, so that the critical buckling load of the last mode is sufficiently larger than that of the first mode, in order to eliminate any possibility of mode interactions and the

associated detrimental effects for the bearing capacity of the structure.

Step 3. Carrying out nonlinear analysis with imperfections

Taking all above discussion into account, a geometrically and material nonlinear imperfection analysis (GMNIA) is recommended for reliably understanding the behavior, predicting all possible failure mechanisms, evaluating the strength, and assessing the vulnerability of structures. In many cases it is useful to carry out first separate analyses accounting only for material nonlinearity (MNIA) and geometric nonlinearity (GNIA), which help identify the prevailing failure mechanism of the structure in question. But the ultimate strength evaluation should always be based on GMNI analyses, accounting for both types of nonlinearity. Several decisions, related to the details of this type of analysis, arise and must be considered by the engineer:

- (i) *Shape of initial imperfections:* As stated above, in the absence of experimental data on manufacturing-related imperfection patterns, it is widely recognized as appropriate to use a linear combination of the critical buckling modes obtained from

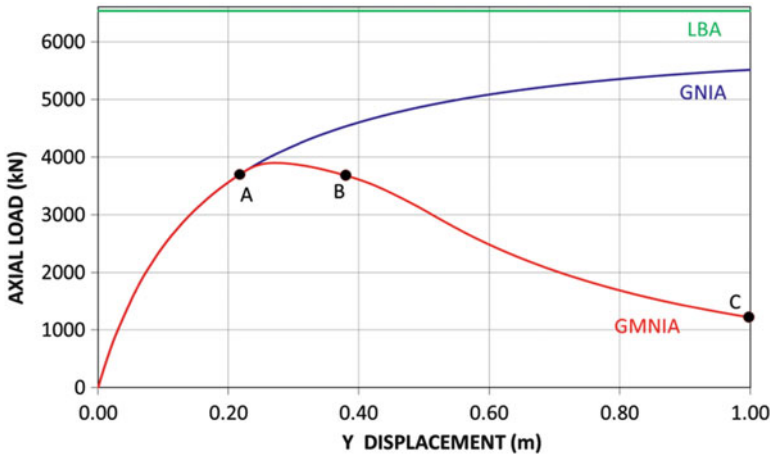


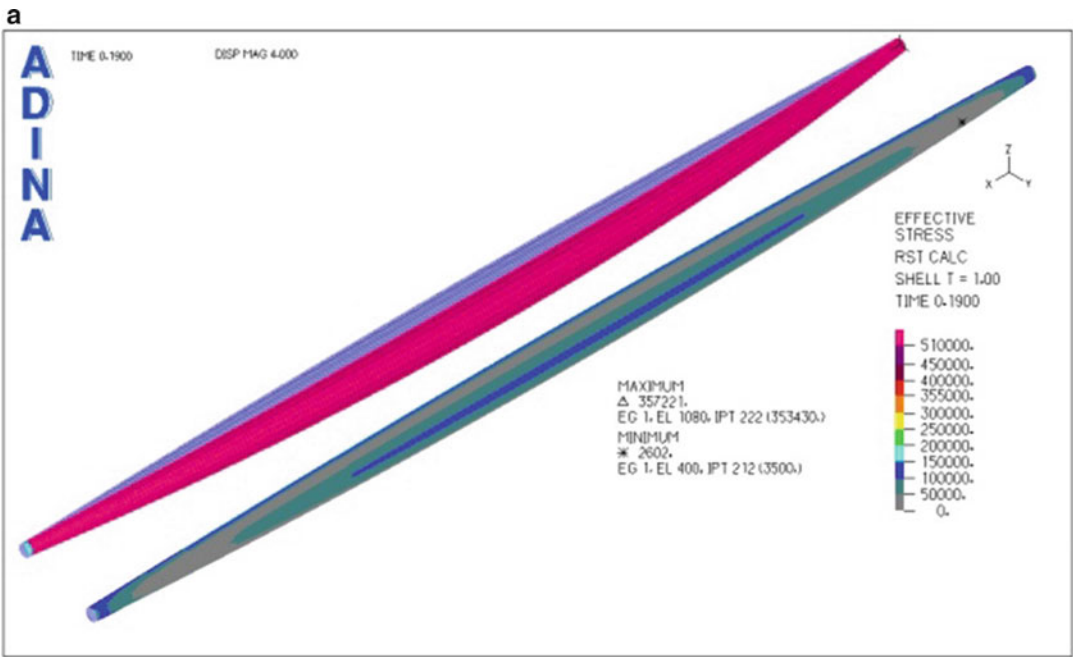
Nonlinear Finite Element Analysis, Fig. 17 Pylon shell element model of steel roof over Aristotle’s Lyceum



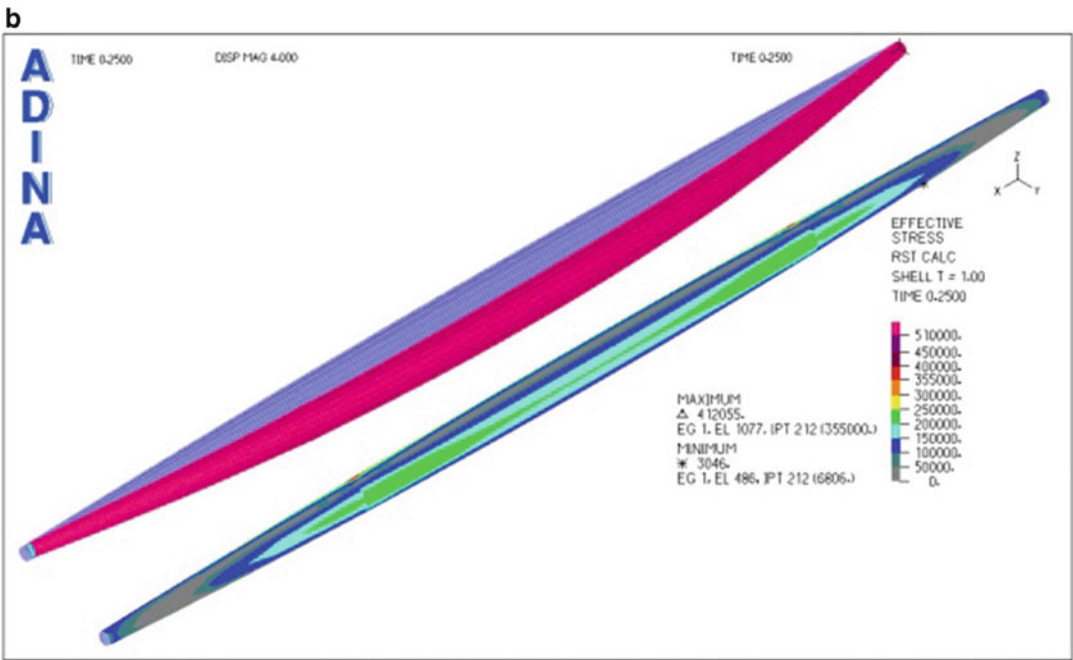
Nonlinear Finite Element Analysis, Fig. 18 First global flexural buckling mode of pylon of the steel roof over Aristotle’s Lyceum

Nonlinear Finite Element Analysis,
Fig. 19 Equilibrium paths of pylon of the steel roof over Aristotle’s Lyceum





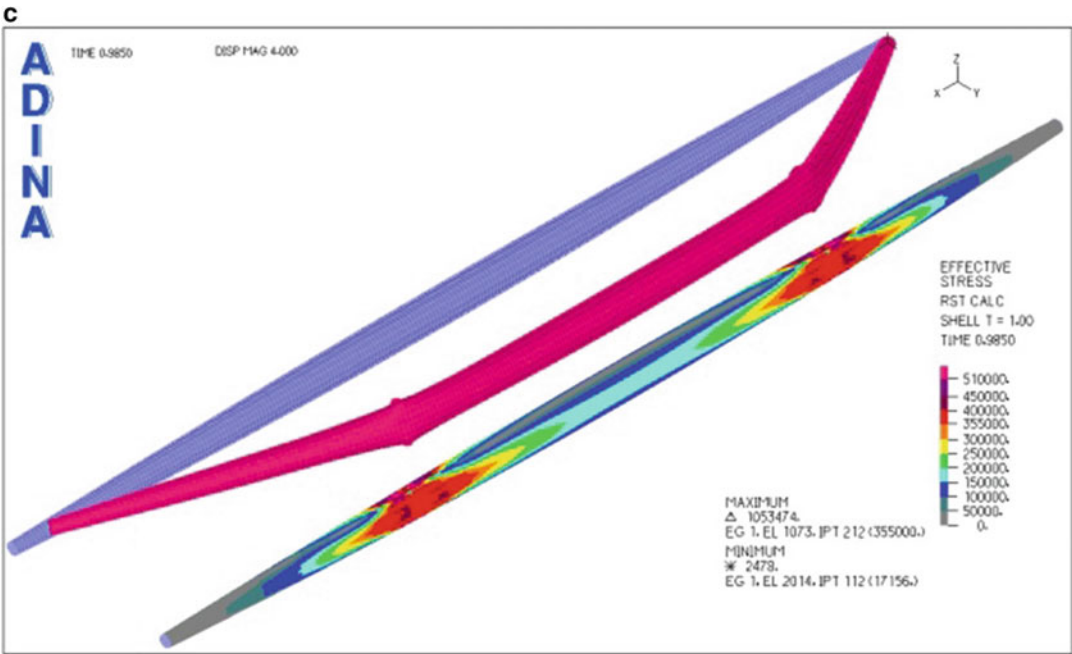
Point A



Point B

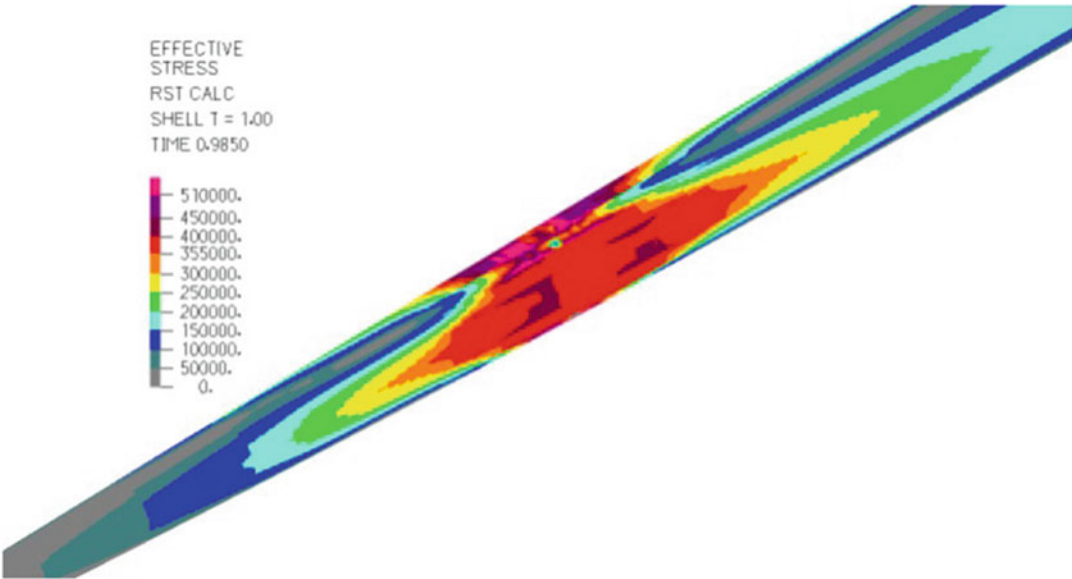
Nonlinear Finite Element Analysis, Fig. 20 (continued)

N

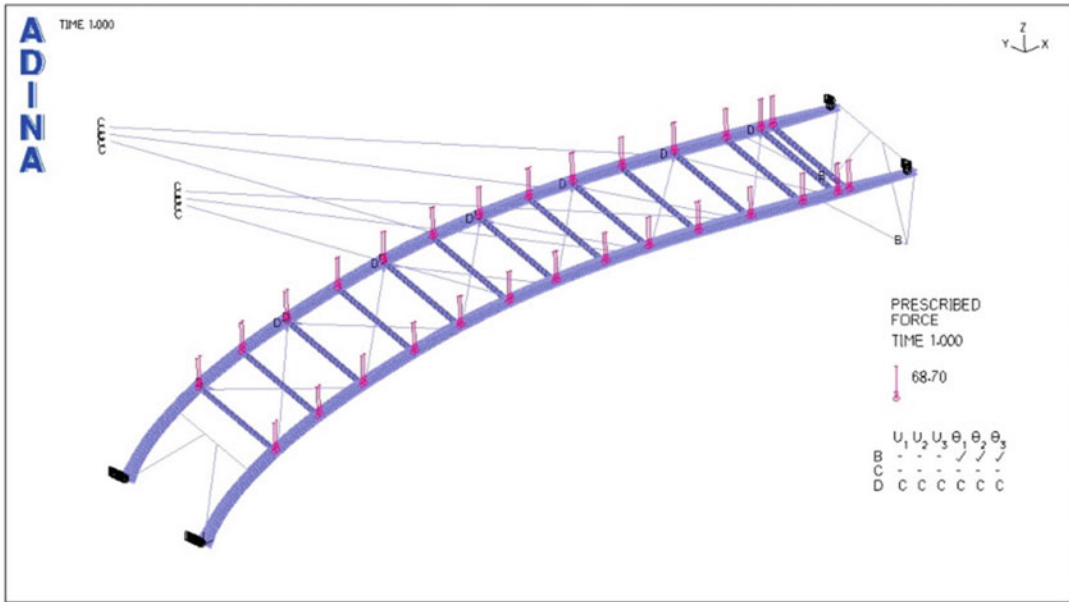


Point C

Nonlinear Finite Element Analysis, Fig. 20 Deformation and von Mises stress distribution at characteristic points A, B, and C along the GMNIA equilibrium path of pylon of steel roof over Aristotle’s Lyceum (kN/m²)



Nonlinear Finite Element Analysis, Fig. 21 Detail of von Mises stress distribution at characteristic point C along the GMNIA equilibrium path of pylon of steel roof over Aristotle’s Lyceum (kN/m²)



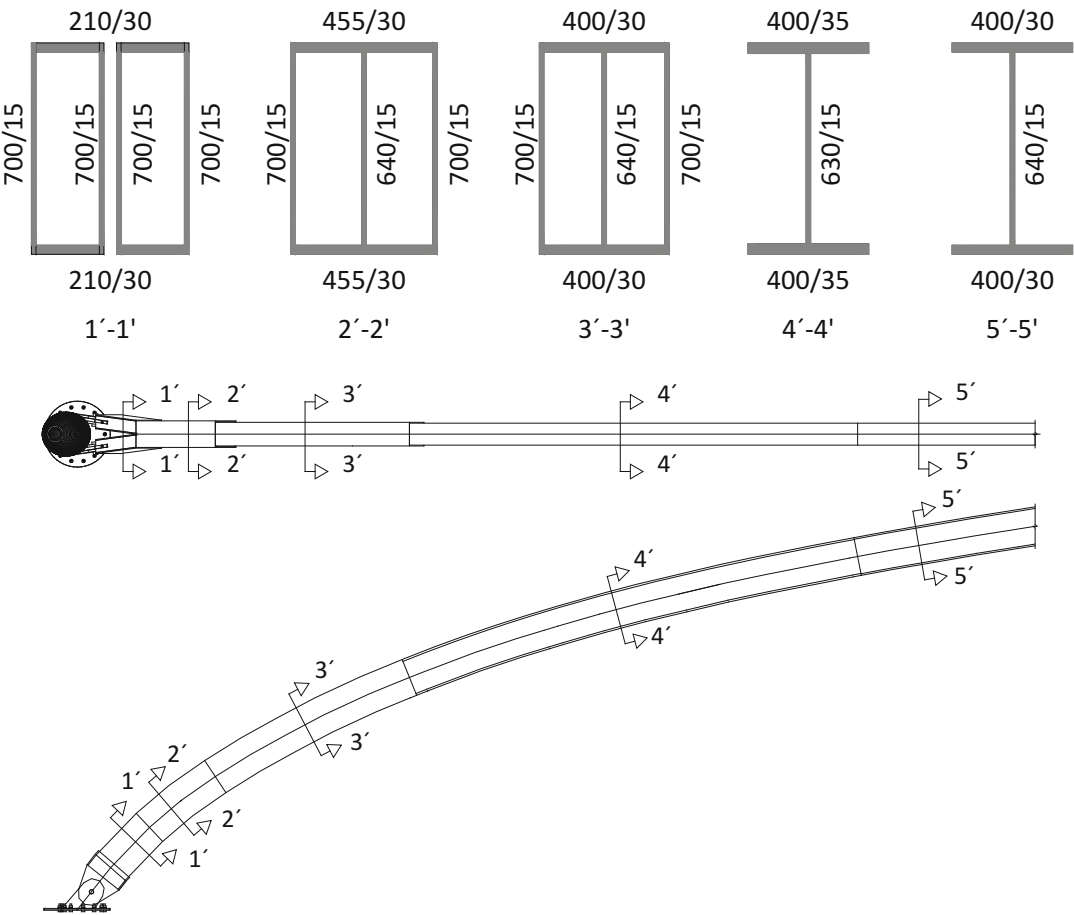
Nonlinear Finite Element Analysis, Fig. 22 Finite element model of two adjacent main arches of steel roof over Aristotle's Lyceum

linearized buckling analysis as the shape of initial imperfections. Regarding the number of critical buckling modes, reference is made to the discussion in step 2 above. In the absence of any available information about the relative importance of individual modes, equal weights may be assumed for their linear combination.

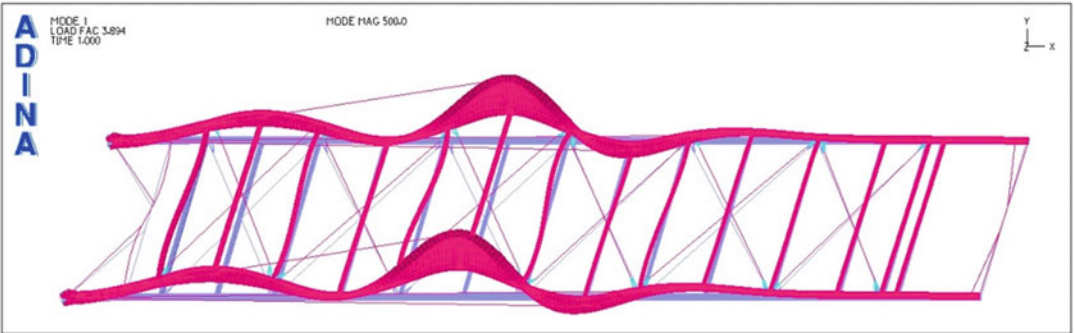
- (ii) *Amplitude of initial imperfections:* With the exception of cases with steep unstable post-buckling equilibrium paths, for example, frequently encountered for thin-walled shells, the size of initial imperfections is not very important for the type of response and the values of critical response quantities obtained from the analysis. For certain types of structures, pertinent codes specify recommended magnitudes of imperfections or manufacturing and erection tolerances that can be regarded as upper bounds of imperfection magnitudes (European Committee for Standardisation 2004). In some cases, a parametric study for different sizes

within the range of realistic construction tolerances may provide valuable insight into the effect of imperfection magnitude on ultimate strength.

- (iii) *Details of numerical algorithm:* As mentioned above, all numerical algorithms for nonlinear structural analysis rely on applying the loads in small steps, performing iterations of linear analyses within each step until a predefined convergence criterion is met, and using the final configuration of the structure at each step as a starting configuration for the next one. The choice of an arc-length algorithm with either MNR (as a first alternative) or NR (in cases of convergence difficulties) iterations is recommended for GMNI analyses for practical applications. If convergence difficulties are still encountered, the engineer may have to increase the default number of maximum iterations per step or relax the convergence tolerances, but in the second case, special attention is recommended, as erroneous



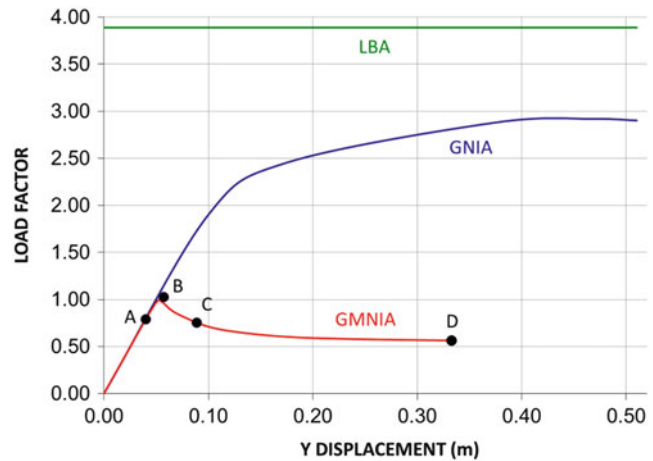
Nonlinear Finite Element Analysis, Fig. 23 Cross sections of main arch



Nonlinear Finite Element Analysis, Fig. 24 First buckling mode of two adjacent main arches of steel roof over Aristotle's Lyceum

Nonlinear Finite Element Analysis,

Fig. 25 Equilibrium paths of two adjacent main arches of steel roof over Aristotle's Lyceum



convergence to configurations without physical meaning may be obtained.

- (iv) *Evaluation of results:* It is recommended to evaluate the results of GMNI analyses by using a plot of the equilibrium path, accompanied by characteristic snapshots of the structure's deformation and stress distribution along this path. The equilibrium path is a plot of a characteristic response quantity (usually on the horizontal axis) with respect to the external action (usually on the vertical axis). Very often the external action is represented by a load multiplier, λ , which acts upon the required design loads. Particular attention is needed for the selection of the characteristic response quantity of the horizontal axis, which should be representative of the overall structural response and not of some local phenomena, depending also on the critical failure mode. Multiple equilibrium paths for more than one response quantity may be useful in some cases.

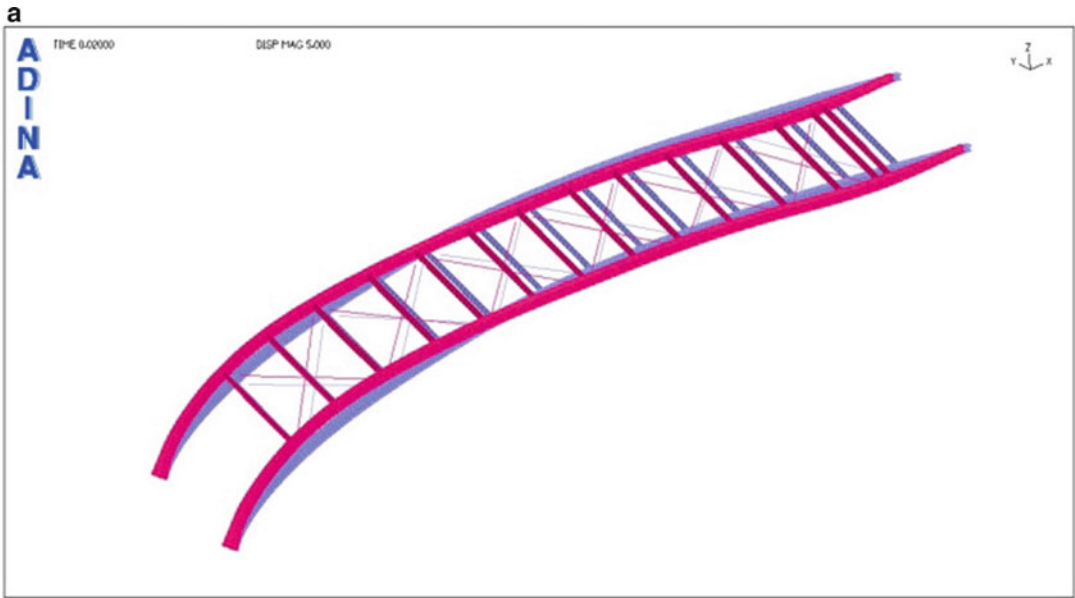
Snapshots of the structure's deformation and stress distribution at characteristic points (e.g., before and after change of slope or change of curvature, at the ultimate limit point, at the end of the curve) are very useful in appreciating the physical phenomena occurring with progressing external loading.

Important quantities of an equilibrium path include: (a) the slope of its initial, usually straight, part, representing the structure's initial stiffness; (b) the maximum value of the external action, representing the structure's strength; (c) the amplitude of displacements corresponding to the maximum value of the external action, pointing out whether serviceability problems may be encountered before failure; and (d) the maximum deformation at failure, as well as the area between the equilibrium path and the horizontal axis, particularly in the post-linear range, representing the structure's ductility.

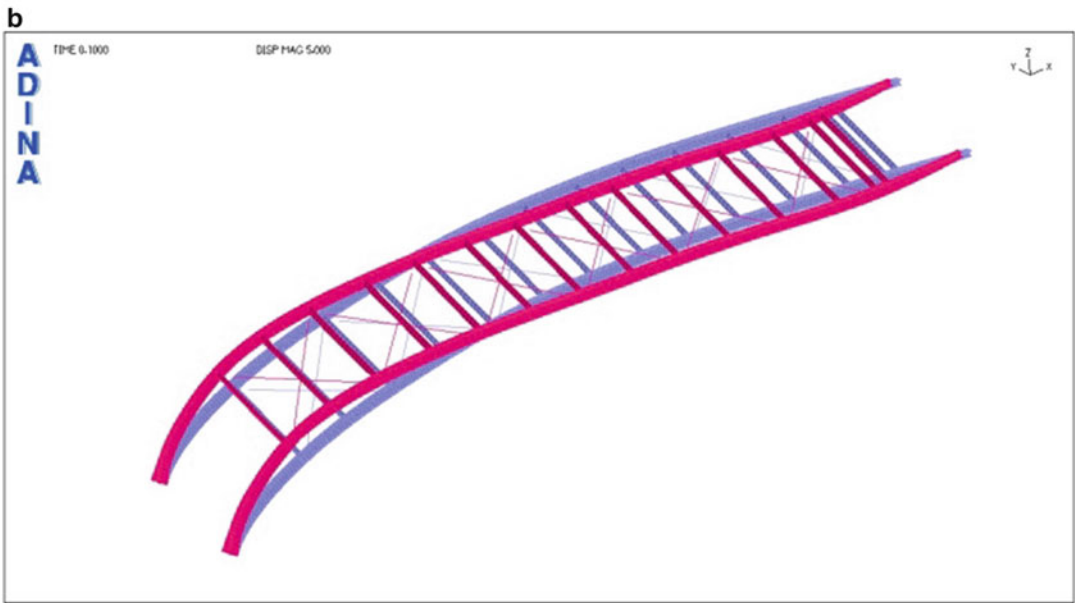
Comparison of representative numerical results with experimental measurements, if available, is strongly encouraged as a means for calibrating and verifying the numerical results. Numerical analysis can then, with increased reliability, be extended for other cases of geometry, cross sections, material, boundary conditions, or loads, thus performing "numerical experiments" that may help optimize the design.

Case Studies

The abovementioned approach, making use of advanced nonlinear numerical analyses, is

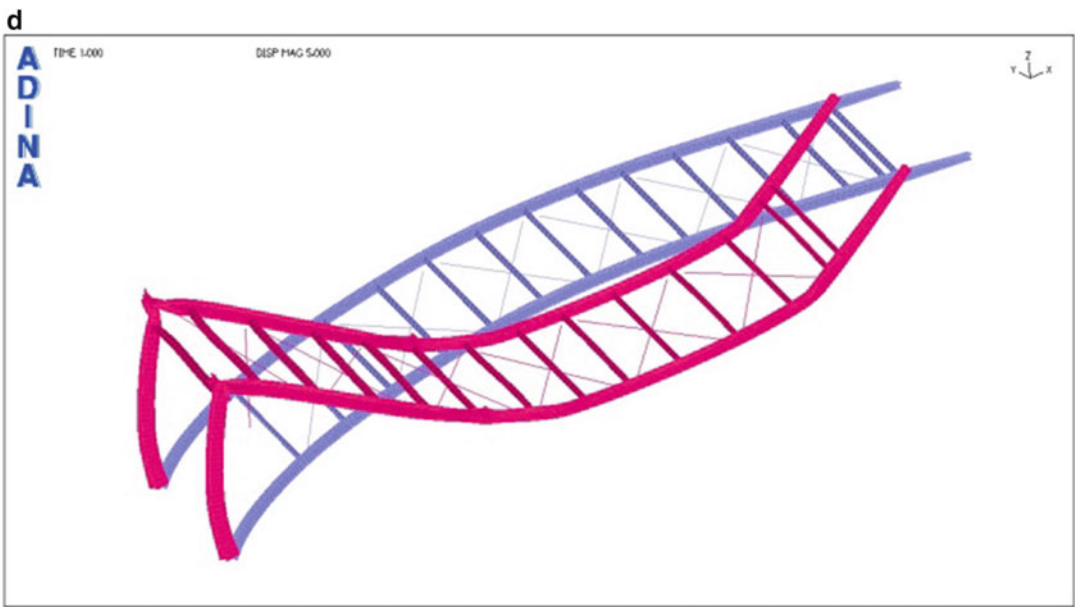
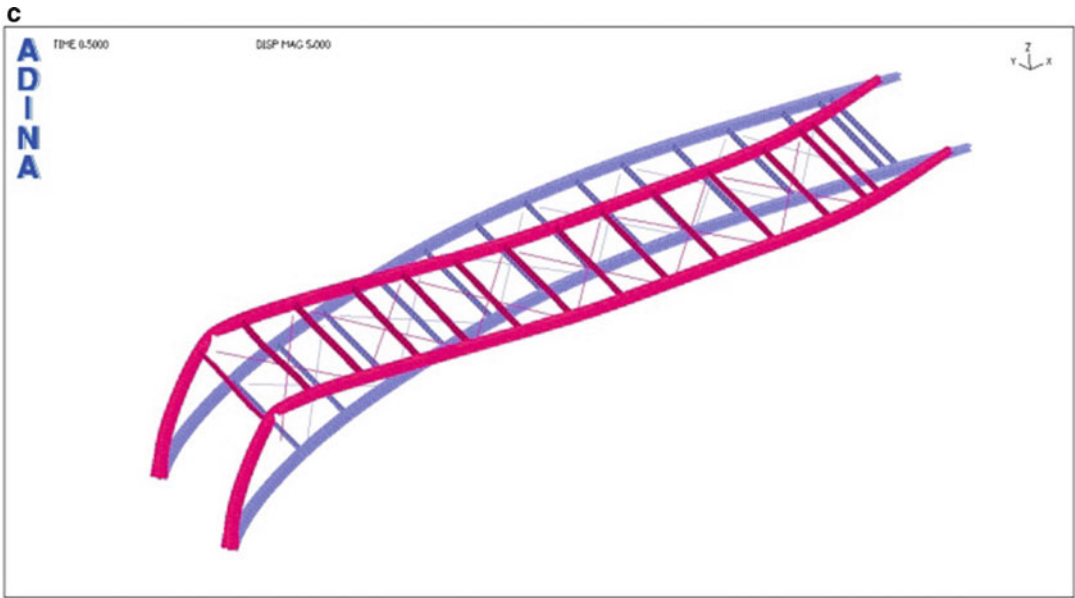


Point A

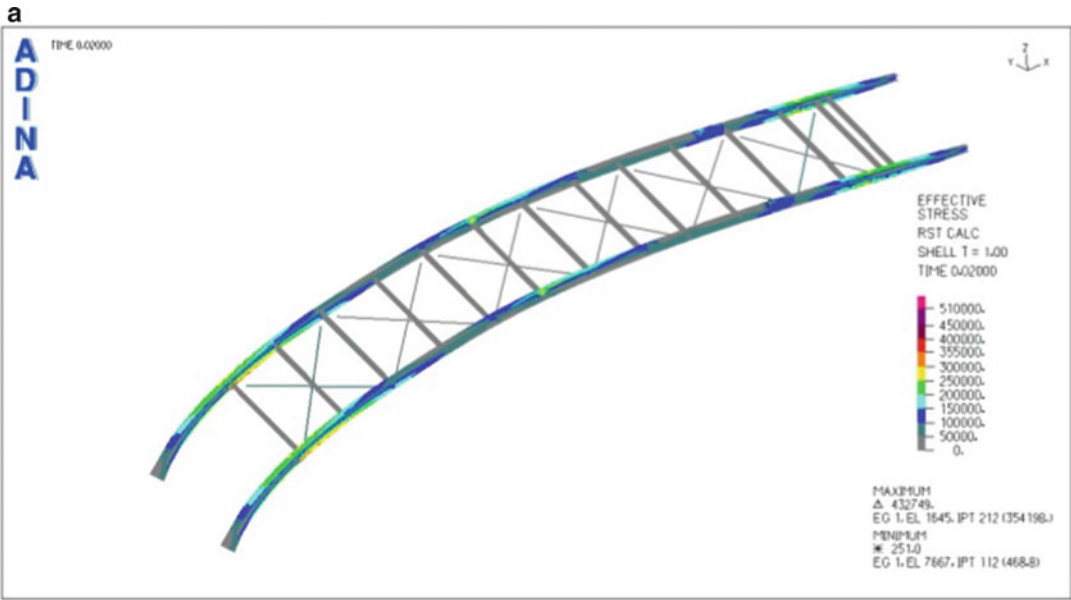


Point B

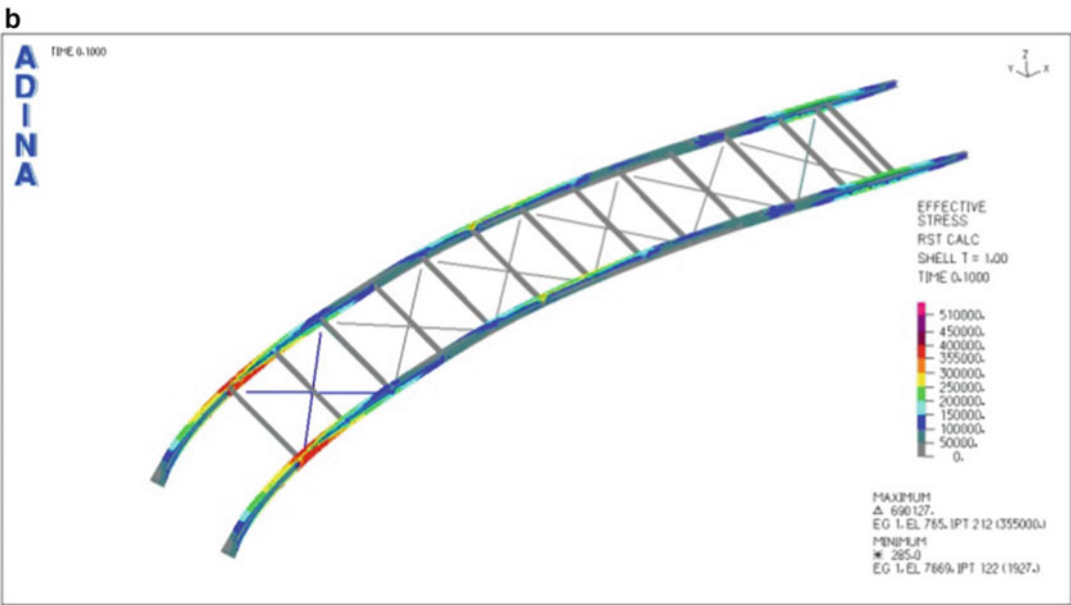
Nonlinear Finite Element Analysis, Fig. 26 (continued)



Nonlinear Finite Element Analysis, Fig. 26 Deformation at characteristic points along the GMNIA equilibrium path of two adjacent main arches of steel roof over Aristotle’s Lyceum

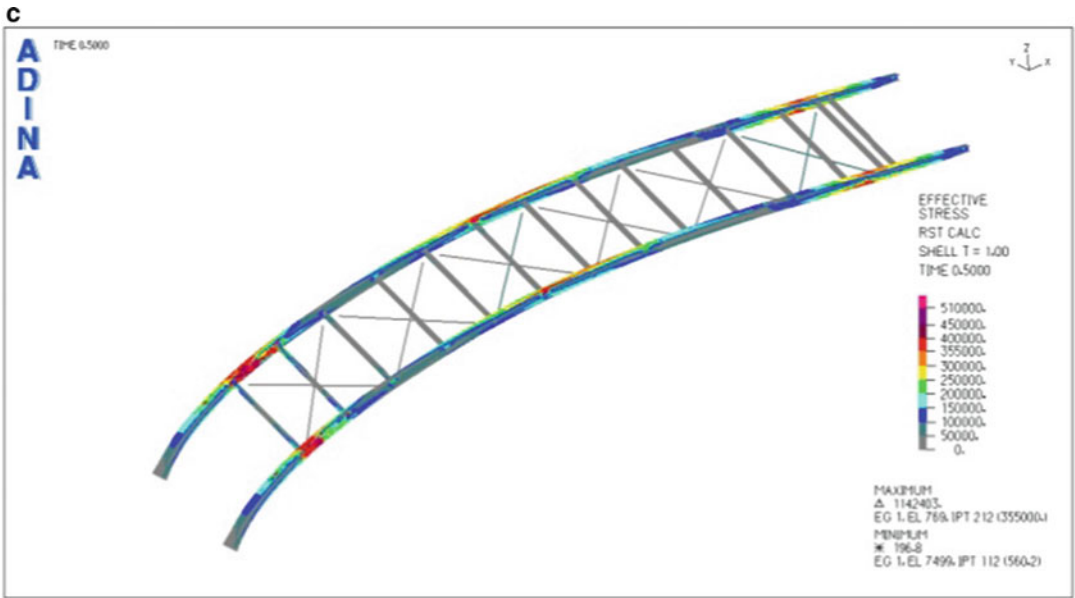


Point A

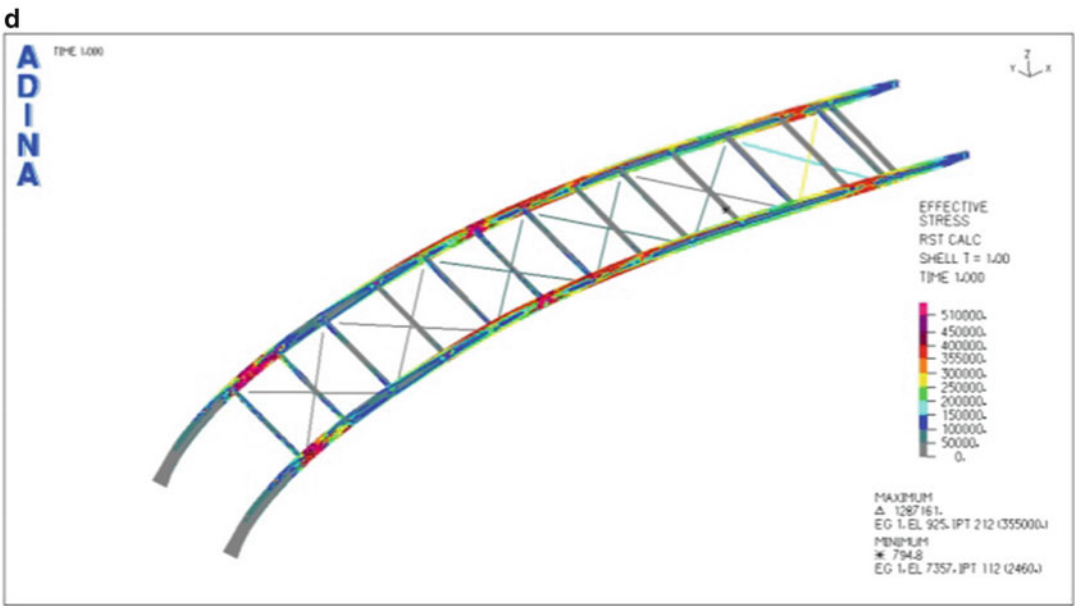


Point B

Nonlinear Finite Element Analysis, Fig. 27 (continued)

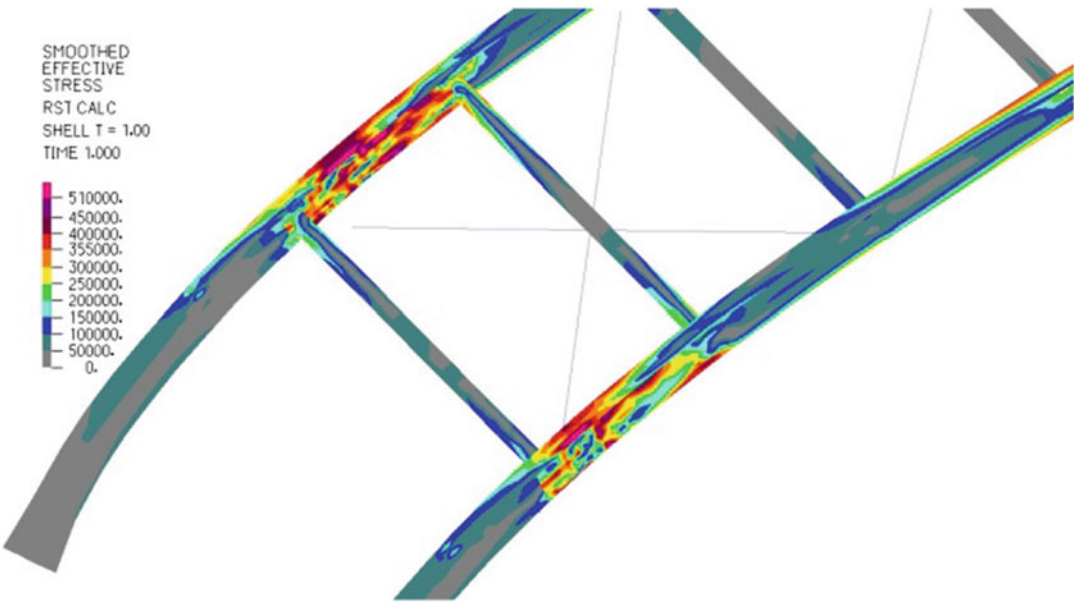


Point C

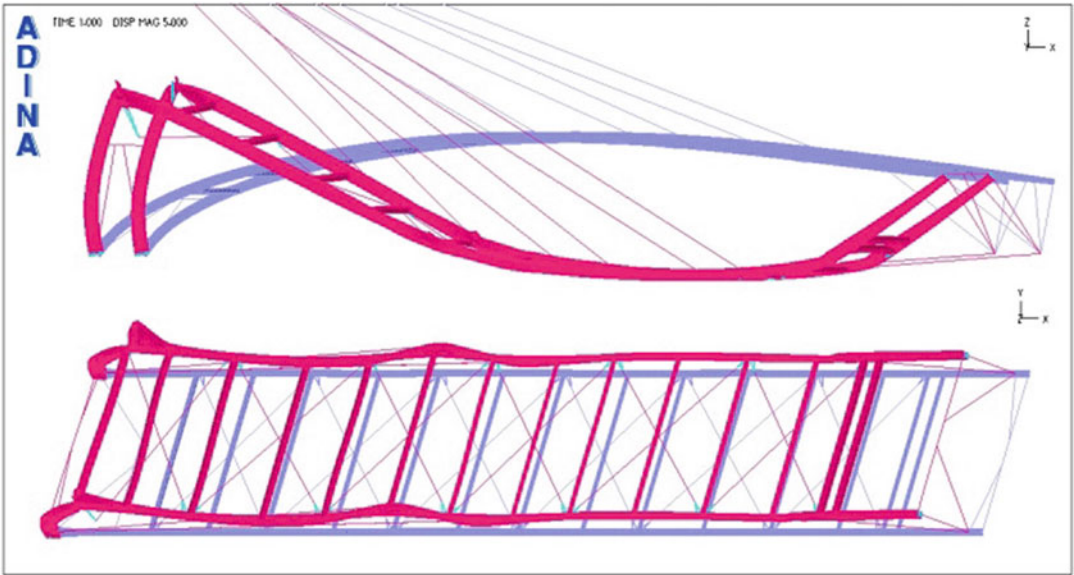


Point D

Nonlinear Finite Element Analysis, Fig. 27 Von Mises stress distribution at characteristic points along the GMNIA equilibrium path of two adjacent main arches of steel roof over Aristotle’s Lyceum (kN/m^2)



Nonlinear Finite Element Analysis, Fig. 28 Detail of von Mises stresses distribution of characteristic point D along the GMNIA equilibrium path of two adjacent main arches of steel roof over Aristotle’s Lyceum (kN/m²)



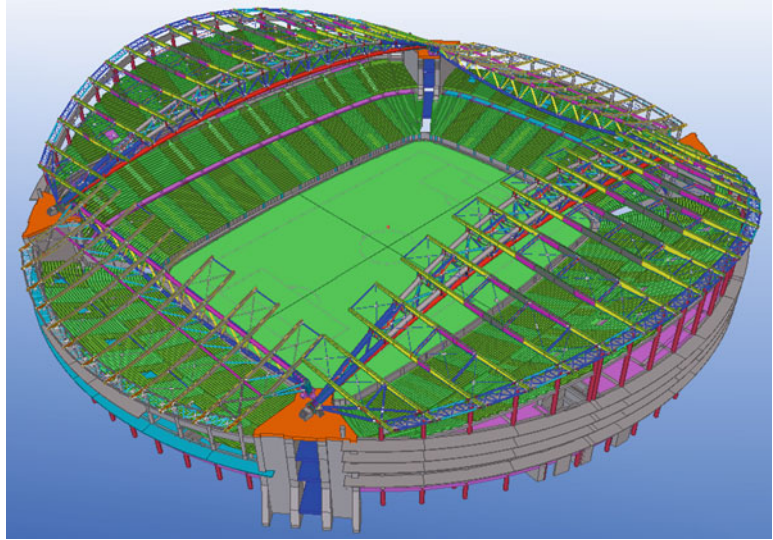
Nonlinear Finite Element Analysis, Fig. 29 Elevation and plan view of main arches at characteristic point D along the GMNIA equilibrium path (same scale of deformations in both views)

next illustrated by presenting the results of evaluation of the ultimate strength of members with unconventional geometry and/or cross section belonging to two long-span roofs that have been

recently designed in Greece (Gantes and Koulatsou 2013). Due to the unconventional nature of these members, their design according to design codes had to rely upon simplifying

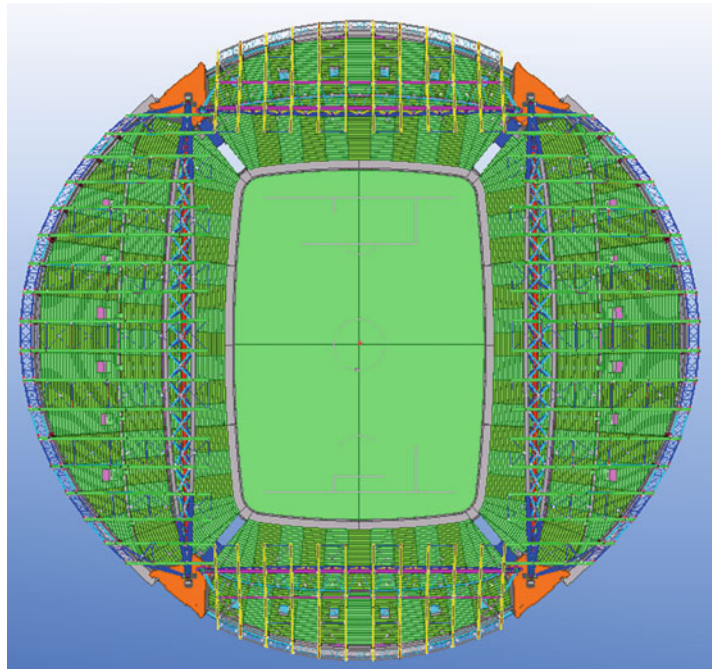
Nonlinear Finite Element Analysis,

Fig. 30 Perspective view of football stadium of Panathinaikos FC



Nonlinear Finite Element Analysis, Fig. 31

Plan view of football stadium of Panathinaikos FC

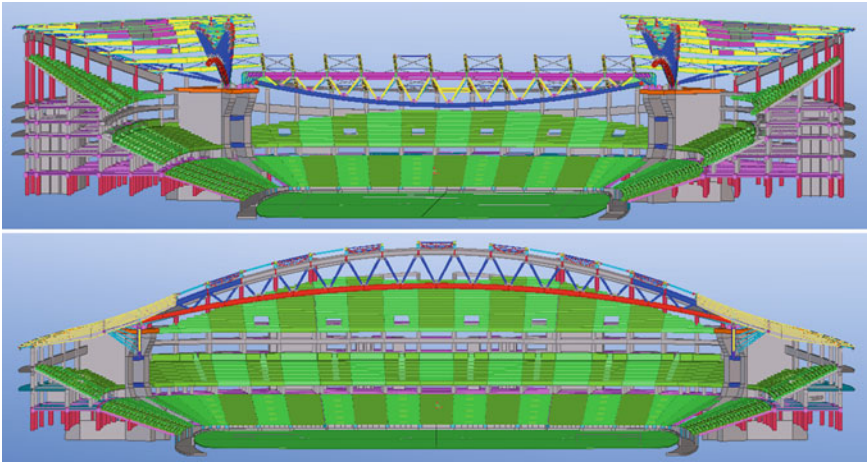


N

assumptions of questionable validity, thus application of this approach was meaningful and appropriate. The analyses have been carried out using the well-known and extensively tested finite element software Adina [ADINA R&D 2012].

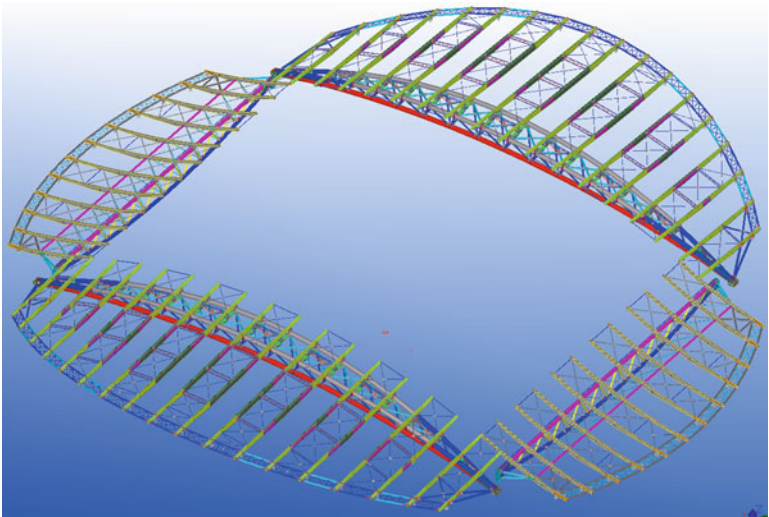
Suspended Steel Roof Over Aristotle's Lyceum

The archeological findings of the School of Aristotle are located in the center of Athens, Greece. For the protection of the archeological site, an architectural competition was held by the Greek

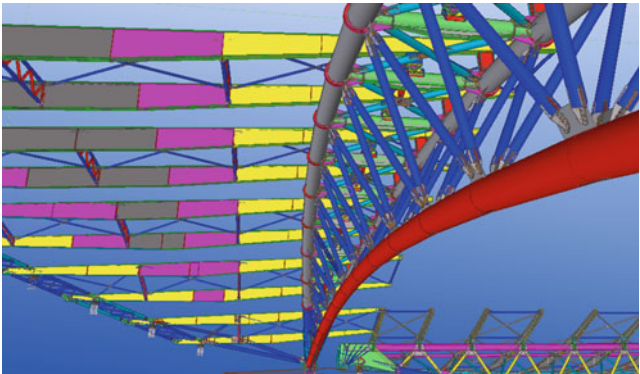


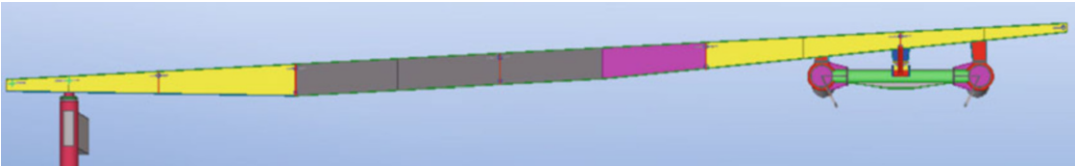
Nonlinear Finite Element Analysis, Fig. 32 Sections of the stadium and main truss girders

Nonlinear Finite Element Analysis, Fig. 33 Secondary beams and bracings of stadium's roof

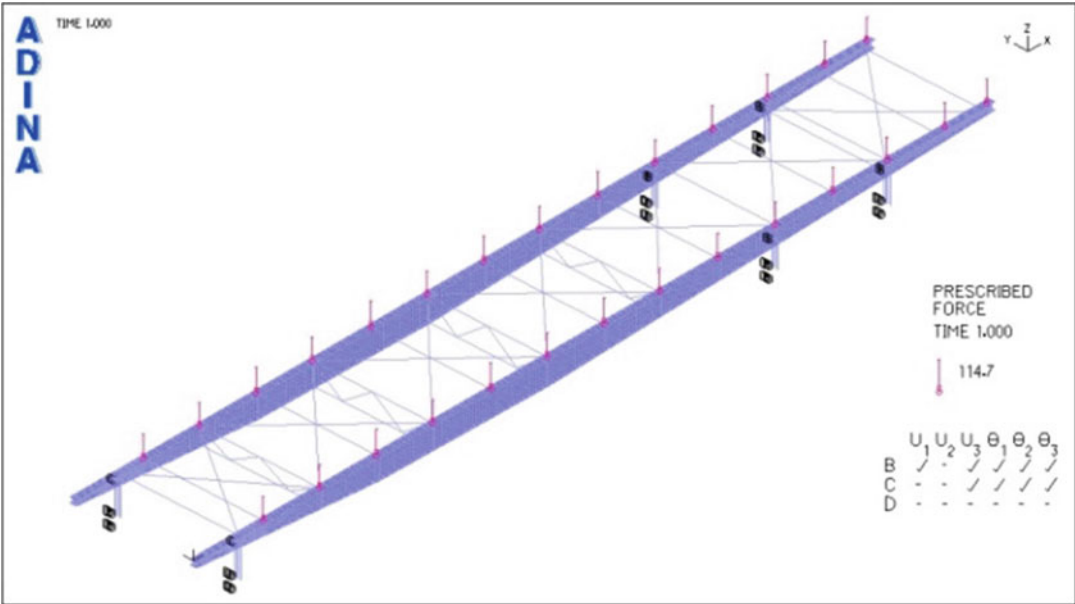


Nonlinear Finite Element Analysis, Fig. 34 Support of secondary beams on main truss girder





Nonlinear Finite Element Analysis, Fig. 35 Cross sections of secondary beams



Nonlinear Finite Element Analysis, Fig. 36 Finite element model of two secondary beams and their cross-bracing in the steel roof of Panathinaikos stadium

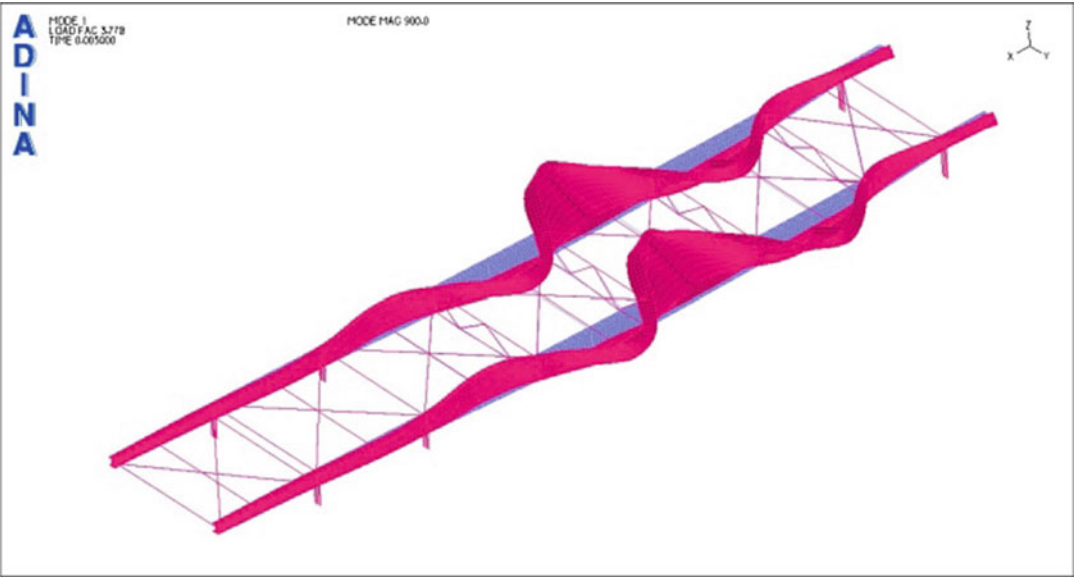
Ministry of Culture, and the preliminary architectural proposal (Fig. 13) of the present suspended steel roof was awarded the first prize.

The roof steel structure (Fig. 14) consists of six parallel main arches having a span of 60 m and spaced at a distance of 10.5 m that are interconnected by means of purlins and horizontal bracings. Each main arch is pinned on the ground on one edge, while the other edge is supported by a V-shaped column. The arches are suspended by prestressed cables from a 25 m tall pylon, which has a slight inclination with respect to the vertical plane. The stability in and out of plane of the pylons is ensured by two back-stayed prestressed cables.

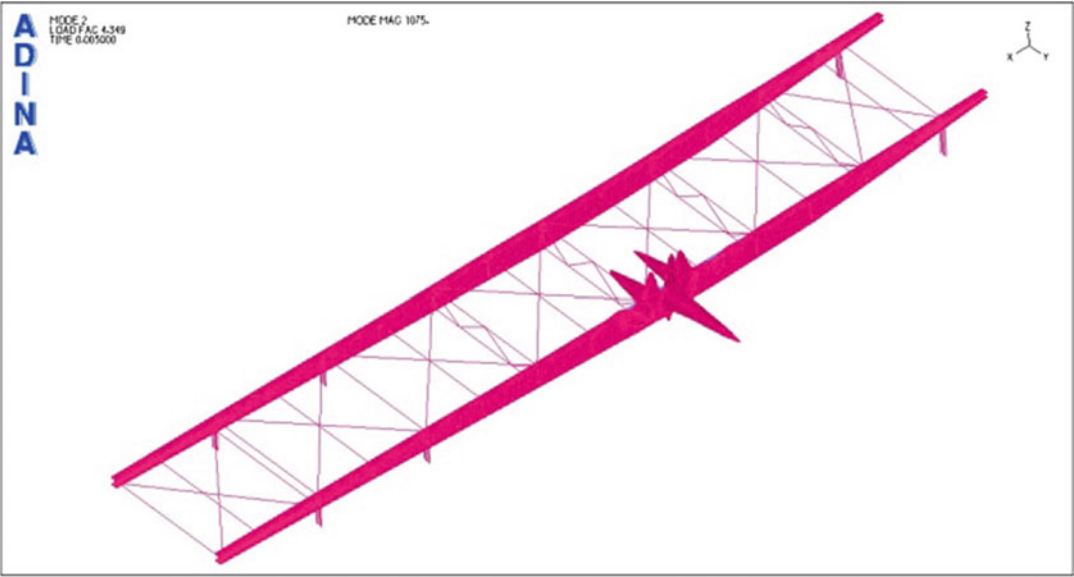
On the exterior of the pylons and V-shaped columns, independent steel structures for the

extension of the cladding as well as eccentric vertical bracings for contributing to the resistance against lateral loads transversely to the main arches are provided.

The main arches (Fig. 15) consist of bending and compression members, which are made of built-up cross sections, varying along the length, having the shape of a rectangular box (green, cyan, and red parts in Fig. 15) and an I shape in the remaining parts (blue parts in Fig. 15). The 25 m tall pylons (Fig. 15) are compression members with hollow circular cross section, varying over the height in both diameter and wall thickness. For the design of these two members, main arches and pylons, numerical models have been created and analyzed linearly and nonlinearly.



Nonlinear Finite Element Analysis, Fig. 37 First buckling mode (lateral buckling) of two secondary beams and their cross-bracing in the steel roof of Panathinaikos stadium



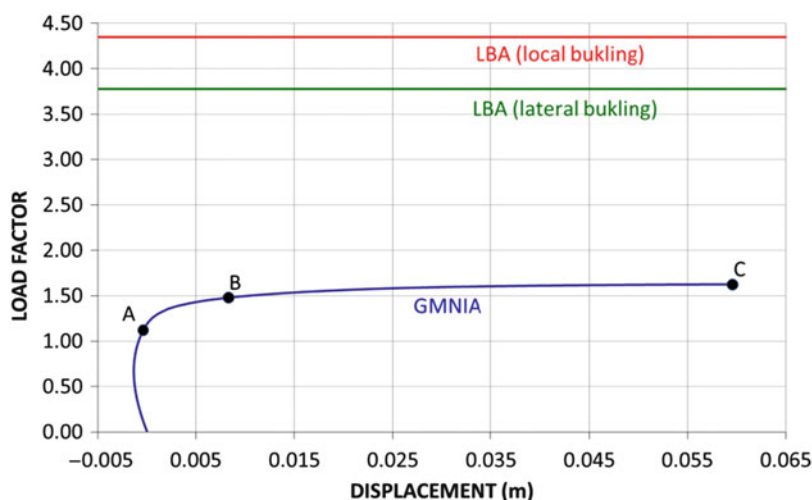
Nonlinear Finite Element Analysis, Fig. 38 Second buckling mode (local buckling) of two secondary beams and their cross-bracing in the steel roof of Panathinaikos stadium

With respect to the resistance of the pylon in compression, the applicability of buckling curves and associated reduction factors is questionable due to the varying cross section (Fig. 16). In order

to obtain the pylon's strength against compression, a shell finite element model has been created, shown in Fig. 17. A conceptual design decision was taken, to use cross sections of class

Nonlinear Finite Element Analysis,

Fig. 39 Equilibrium paths of secondary beams of Panathinaikos stadium



1 or 2 for the pylon, to avoid local buckling as a possible failure mechanism. This was confirmed by the linearized buckling analysis, which showed no local buckling modes among the first ten, and thus, only the shape of the first global flexural buckling mode, presented in Fig. 18, was used as initial imperfection for nonlinear analyses.

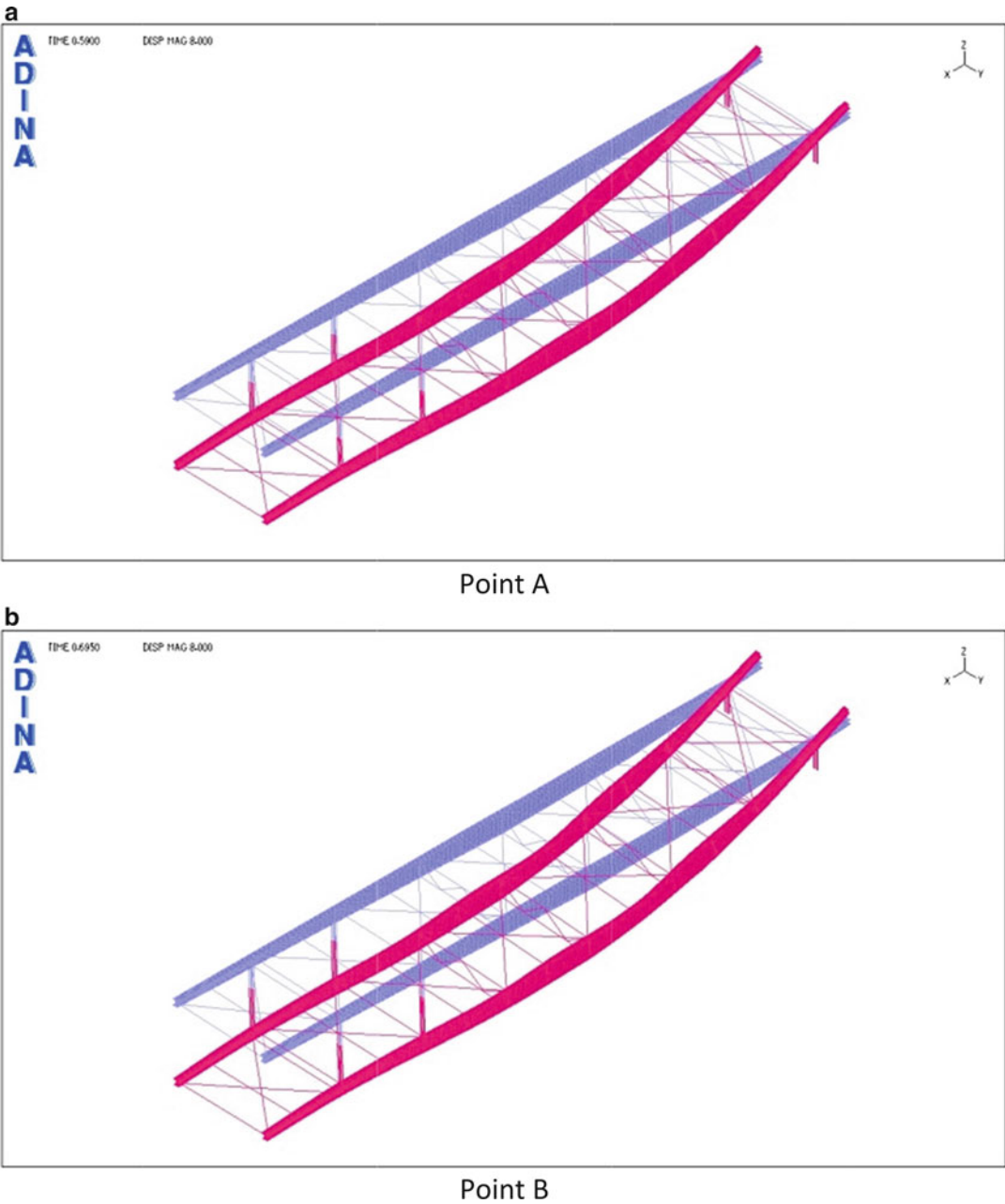
Then, GNI and GMNI analyses were carried out for this imperfection shape having the magnitude prescribed by Eurocode 3 specifications, and the corresponding equilibrium paths are shown in Fig. 19. It is observed that the presence of imperfections reduces significantly the elastic critical load with respect to the linear one, as expected. Failure is encountered on the ascending branch of the equilibrium path, and it is mainly due to material yielding, accompanied by moderate global flexural buckling. This is verified by the deformation and stress snapshots on three characteristic points A, B, and C, before, on, and beyond the critical point, shown in Figs. 20 and 21. The clear formation of plastic hinges at two locations, highlighted in Fig. 21, helps identify these cross sections as weak. These locations are characterized by a decrease of cross-sectional thickness, so that increased thickness at these areas would be the most appropriate way for improving overall pylon strength.

With respect to calculation of the ultimate strength of the main arch, the use of code provisions is also of questionable validity due to its unusual shape and cross sections, thus rendering evaluation of the resistance to lateral buckling with analytical tools unreliable. For that purpose, a finite element model of two arches with their interconnecting purlins and bracings has been created, illustrated in Fig. 22. Similarly to the pylon, cross sections (Fig. 23) of class 1 or 2 were used for the main arches, so that local buckling was not critical. This model is able to simulate the behavior of the system regarding lateral buckling in a reliable manner, since bracings provide realistic lateral support to the system. Main arches and purlins are modeled with shell elements, whereas horizontal bracings are modeled with beam elements and cables with nonlinear, tension-only prestressed truss elements.

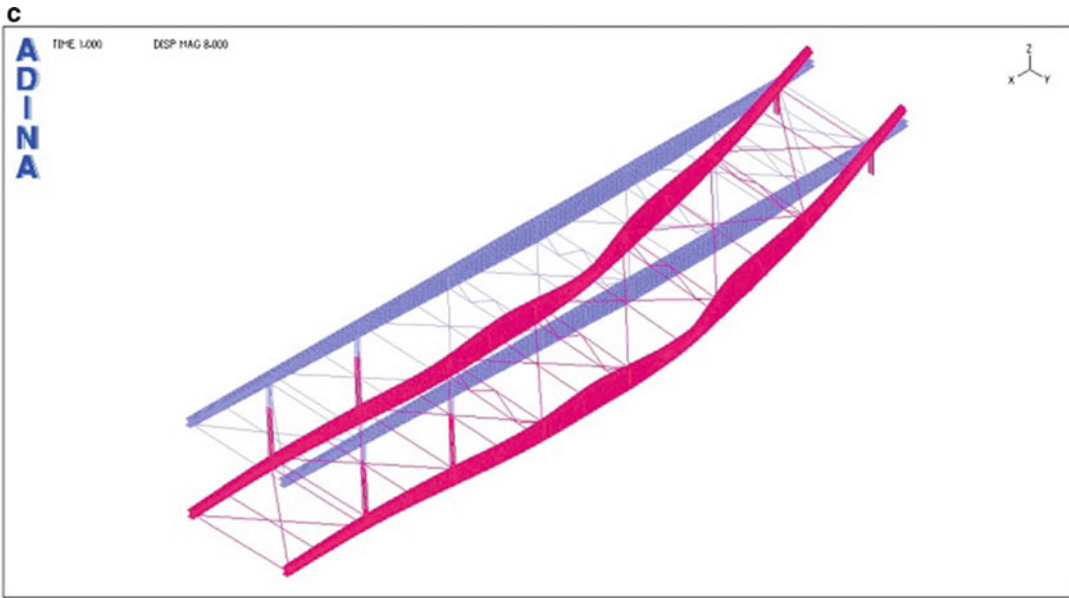
First, linearized buckling analysis was performed in order to obtain the first critical buckling mode (Fig. 24), which was used as shape of initial imperfection for the subsequent nonlinear analyses. GNI and GMNI analyses were then carried out, and their equilibrium paths are displayed in Fig. 25, accompanied by corresponding snapshots of deformed geometry and stresses for characteristic points A, B, C, and D along the GMNIA paths, illustrated in Figs. 26

and 27, respectively. The reference loads applied for the nonlinear analyses were chosen equal to the design loads corresponding to the most onerous load combinations. The multiplying load factor reported on the vertical axis of the equilibrium path (Fig. 25) denotes the percentage of reference

loads acting on the structure at each load step. As observed, the maximum load factor for the main arches is approximately equal to 1.1, which represents a rather low factor of safety against collapse and signifies a possible need for strengthening.



Nonlinear Finite Element Analysis, Fig. 40 (continued)



Point C

Nonlinear Finite Element Analysis, Fig. 40 Deformation at characteristic points along the GMNIA equilibrium path of the secondary beams of Panathinaikos stadium

From the above discussion and the deformed shapes and stress contours of Figs. 26 and 27, it is concluded that the box-shaped cross sections at the lower end, combined with the lateral support provided by the bracing system, are sufficient for preventing lateral buckling; thus, failure is primarily due to material yielding, with the formation of plastic hinges becoming particularly evident from the deformation and stress pictures on the descending branch of the equilibrium path. This is also verified by the stress contours detail shown in Fig. 28, corresponding to the equilibrium path of GMNI analysis point D. From the deformed shape of main arches (Fig. 29), it is also observed that the system has little lateral deformation, while vertical deflections dominate.

Figures 28 and 29 are also very useful in identifying as weak the locations of cross section change from double box section to I section (sections 3'-3' and 4'-4' in Fig. 23). An extension of the box section or a more gradual transition to the I section would thus be an appropriate intervention for improving the collapse behavior of the arch and enhancing its ultimate strength.

Moreover, it is verified for this example also that the elastic critical load derived from linearized buckling analysis is reduced significantly due to geometric imperfections and material yielding.

Roof of Panathinaikos Football Stadium

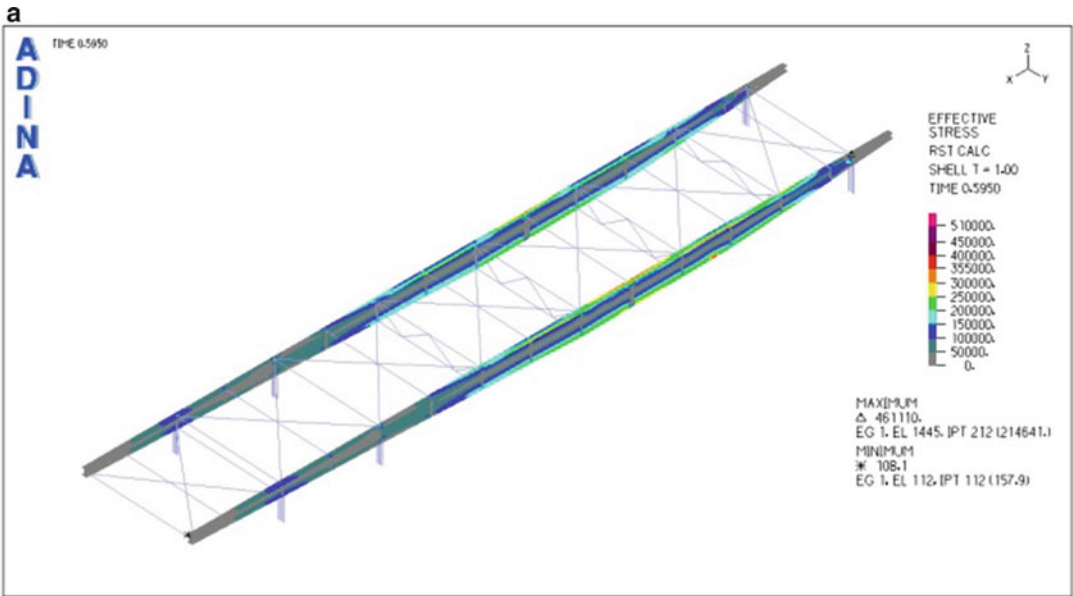
The second case study refers to the secondary beams of the steel roof in the new football stadium of Panathinaikos FC, which will be constructed in Votanikos, Athens, Greece, with a maximum capacity of approximately 40,000 spectators. A perspective view of the structural parts of the stadium is shown in Fig. 30 and its plan view in Fig. 31.

The roof has the shape of a cylindrical surface, and it consists of four structurally independent parts. The basic element of each part is a main truss girder, simply supported on reinforced concrete pylons, arranged at the four corners of the stadium, as shown also in the sections of Fig. 32. Secondary beams are supported on these main girders and on the exterior peripheral reinforced concrete columns of the grandstands (Figs. 33

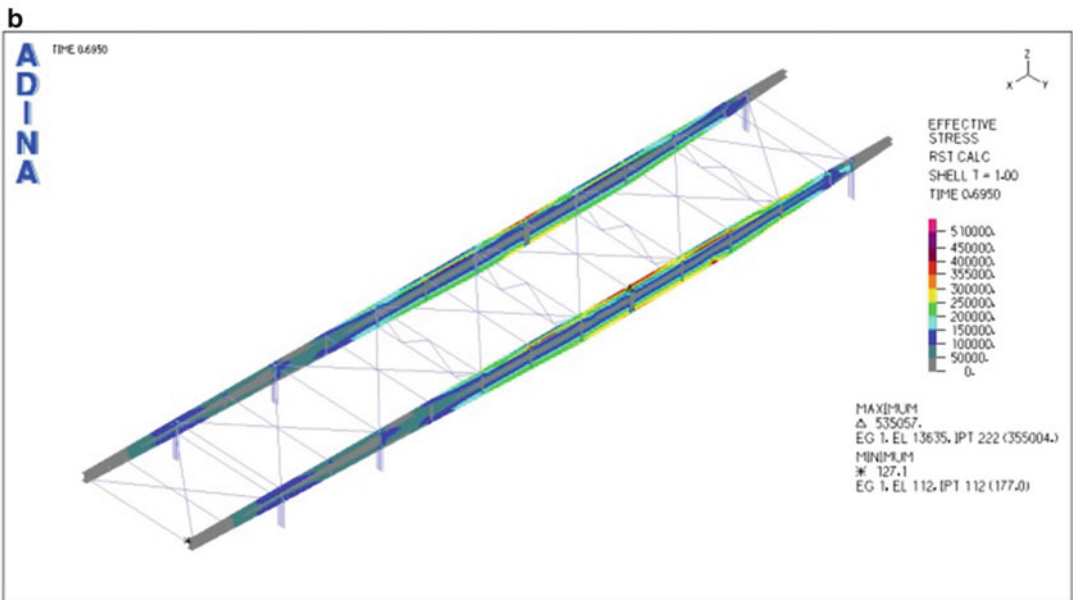
and 34). These beams carry the roof cladding. Appropriately arranged cross-bracings contribute to the overall stability.

Aim of the numerical investigation presented here is the evaluation of the ultimate capacity of secondary beams. These beams have different

lengths, depending on their location, so that they adjust to the overall roof geometry. In the initial structural design, which is described here, they were made of welded I-shaped cross sections with variable height over their length, as illustrated in Fig. 35. The unusual geometry and cross

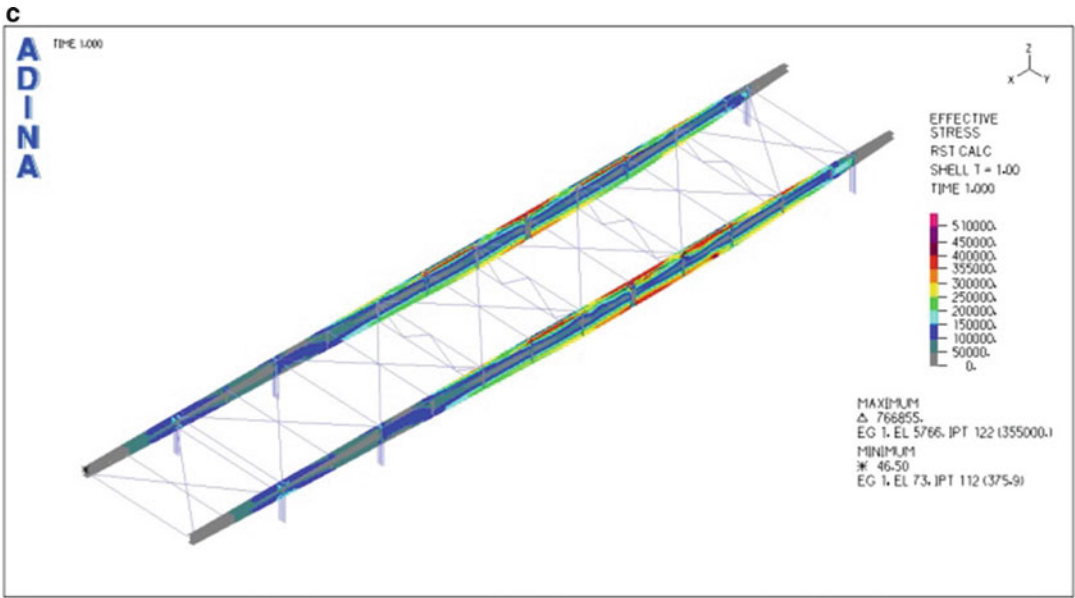


Point A



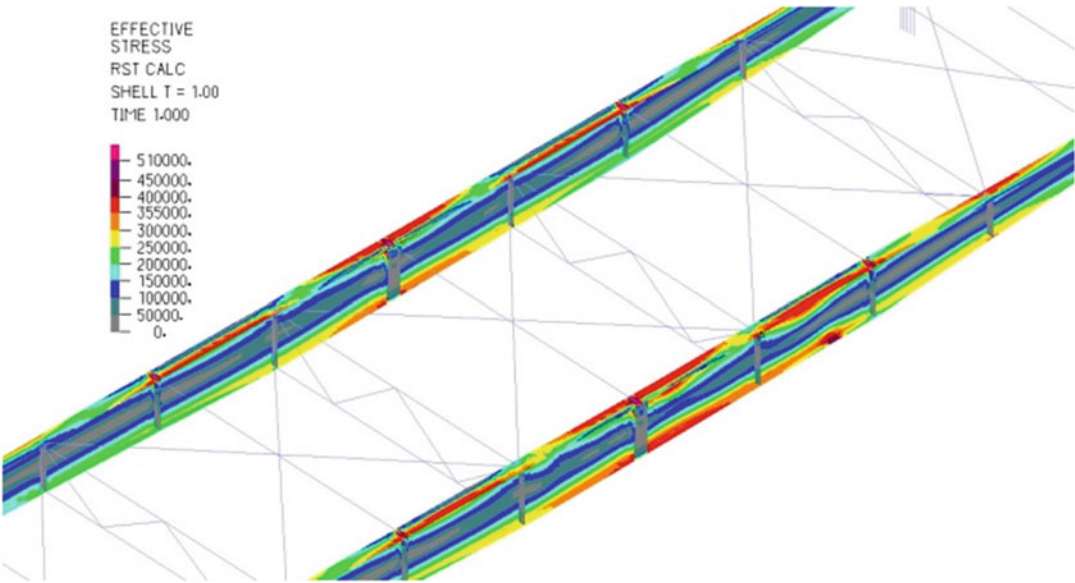
Point B

Nonlinear Finite Element Analysis, Fig. 41 (continued)



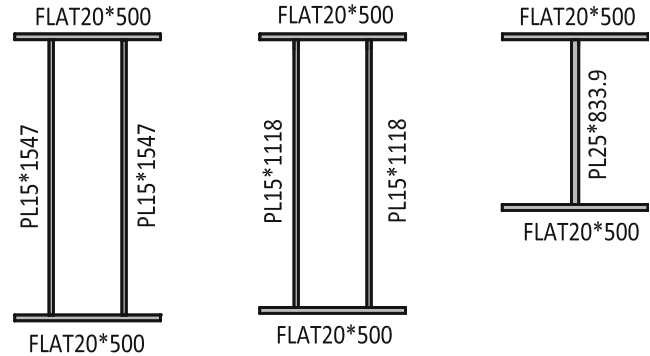
Point C

Nonlinear Finite Element Analysis, Fig. 41 Von Mises stresses at characteristic points along the GMNIA equilibrium path of secondary beams of Panathinaikos stadium (kN/m^2)



Nonlinear Finite Element Analysis, Fig. 42 Detail of von Mises stresses distribution of characteristic point C along the GMNIA equilibrium path of two adjacent secondary beams of Panathinaikos stadium (kN/m^2)

Nonlinear Finite Element Analysis, Fig. 43 Final cross sections of secondary beams of Panathinaikos stadium



sections of the beams highlighted the need for advanced numerical modeling and analyses, in order to address in a reliable manner the lateral and local buckling issues that were encountered.

In order to evaluate the behavior of secondary beams against lateral and local buckling and their strength, a finite element model consisting of two adjacent secondary beams and their cross-bracing members was created, shown in Fig. 36. Secondary beams were modeled with shell elements and cross-bracing members with beam elements. Secondary beam support on the main truss girders was modeled with spring elements.

First, linearized buckling analysis was performed in order to estimate the buckling modes and corresponding critical loads for both lateral and local buckling. The first two buckling modes are presented in Figs. 37 and 38, which illustrate the first lateral buckling mode and the first local buckling mode, respectively. A linear combination of these two first buckling modes was used as initial imperfection for the GMNI analysis.

GNI and GMNI analysis were performed, and the results are displayed by means of the equilibrium paths, shown in Fig. 39, and the snapshots of deformation and stress distribution for characteristic points A, B, and C along the GMNIA equilibrium path (Figs. 40 and 41). As in the previous example, the multiplying load factor reported on the vertical axis of the equilibrium path (Fig. 39) denotes the percentage of reference loads acting on the structure at each load step. In this case the maximum load factor for the main arches is approximately equal to

1.5, however, a significant reduction of stiffness is observed for much lower loads.

It is observed that failure was due to combined lateral buckling and material yielding, with small contribution from local buckling, as indicated by the equilibrium path and the corresponding deformation and stress distribution (Fig. 42). As in the first case study, here also the elastic critical buckling load derived by linearized buckling analysis decreases due to geometric imperfections and material yielding.

Due to the stiffness reduction occurring at low load levels, as mentioned above, associated with onset of lateral buckling, it was decided to change the beams' cross section from I-shaped to narrow box cross sections with protruding flanges that vary over their length (Fig. 43), in order to improve their torsional inertia, thus increasing their stiffness against lateral buckling.

Summary

For structures with unusual shape and variable cross sections, for which the use of pertinent codes and recommendations requires simplifying assumptions of questionable validity, an alternative approach has been presented for understanding the behavior, predicting all possible failure mechanisms, and evaluating the ultimate strength by means of linearized buckling and geometrically and material nonlinear imperfection analyses with commercially available finite element software. Steps include setting up an appropriate finite element model, obtaining critical buckling

modes from linearized buckling analysis (LBA), and then using a linear combination of these modes as imperfection pattern for a geometrically and material nonlinear imperfection analysis (GMNIA).

Failure dominated by either material yielding or instability can thus be addressed, as well as interaction of failure modes. Equilibrium paths accompanied by snapshots of deformation and stress distribution at characteristic points are used for evaluating the results of the GMNI analysis, identifying the dominant failure modes as well as proposing and evaluating alternative strengthening measures.

Results of two case studies have been used to demonstrate this methodology, involving tall compression members and long-span beams with varying cross section, encountered in two steel roofs that have been recently designed in Greece.

Cross-References

- [Assessment of Existing Structures Using Inelastic Static Analysis](#)
- [Behavior Factor and Ductility](#)
- [Nonlinear Analysis and Collapse Simulation Using Serial Computation](#)
- [Plastic Hinge and Plastic Zone Seismic Analysis of Frames](#)
- [Seismic Collapse Assessment](#)

References

- ADINA R&D (2012) Report ARD 12-8: Theory and Modeling Guide – volume I: ADINA Solids & Structures
- Bathe KJ (1995) Finite element procedures. Prentice-Hall, Englewood Cliffs
- Bathe KJ, Cimento AP (1980) Some practical procedures for the solution of nonlinear finite element equations. *Comput Methods Appl Mech Eng* 22(1):59–85
- Bathe KJ, Dvorkin EN (1983) On the automatic solution of nonlinear finite element equations. *Comput Struct* 17(5–6):871–879
- Bathe KJ, Snyder MD, Cimento AP, Rolph WD (1980) On some current procedures and difficulties in finite element analysis of elastic-plastic response. *Comput Struct* 12(4):607–624
- Bazant ZP, Cedolin L (1991) Stability of structures: elastic, inelastic, fracture, and damage theories. Oxford University Press, Oxford
- Belytschko T, Liu WK, Moran B (2000) Nonlinear finite elements for continua and structures. Wiley, New Jersey
- Berry PA, Rotter JM, Bridge RQ (2000) Compression tests on cylinders with circumferential weld depressions. *J Eng Mech (ASCE)* 126(4):405–413
- Chang SC, Chen JJ (1986) Effectiveness of linear bifurcation analysis for predicting the nonlinear stability limits of structures. *Int J Numer Methods Eng* 23:831–846
- Crisfield MA (1979) A faster modified Newton-Raphson iteration. *Comput Methods Appl Mech Eng* 20(3):267–278
- Crisfield MA (1981) A fast incremental/iterative solution procedure that handles snap-through. *Comput Struct* 13(1–3):55–62
- Crisfield MA, Jelenic G, Mi Y, Zhong H-G, Fan Z (1997) Some aspects of the non-linear finite element method. *Finite Elem Anal Des* 27(1):19–40
- Dimopoulos CA, Gantes CJ (2012) Comparison of alternative algorithms for buckling analysis of slender steel structures. *Struct Eng Mech* 44(2):219–238
- European Committee for Standardisation (2005) Eurocode 3: Design of steel structures, Part 1-1: General rules and rules for buildings
- Gantes CJ (2011) Numerical evaluation of ultimate bearing capacity of steel structures. In: Topping BHV, Tsompanakis Y (eds) Civil and structural engineering computational technology. Saxe-Coburg Publications, Stirlingshire
- Gantes CJ, Fragkopoulos Kç (2010) Strategy for numerical verification of steel structures at the ultimate limit state. *Struct Infrastruct Eng* 6(1–2):225–255
- Gantes CJ, Koulatsoy K (2013) Methodology for nonlinear finite element analyses to evaluate strength of steel structures. In: Obrębski JB, Tarczewski R (eds) Beyond the limits of man. Paper presented at International Association for Shell and Spatial Structures (IASS), headquartered in Madrid, Spain, symposium 2013, Wrocław, 23–27 Sept 2013
- Herynk MD, Kyriakides S, Onoufriou A, Yun HD (2007) Effects of the UOE/UOC pipe manufacturing processes on pipe collapse pressure. *Int J Mech Sci* 49:533–553
- Kojic M, Bathe KJ (2004) Inelastic analysis of solids and structures, Series in computational fluid and solid mechanics. Springer, Berlin
- Livanou MA, Gantes CJ, Avraam TP (2013) Revisiting the problem of buckling mode interaction in 2-DOF models and built-up columns. In: Obrębski JB, Tarczewski R (eds) Beyond the limits of man. Paper presented at International Association for Shell and Spatial Structures (IASS), headquartered in Madrid, Spain, symposium 2013, Wrocław, 23–27 Sept 2013
- Ramm E (1981) Strategies for tracing nonlinear responses near limit points. In: Wunderlich W, Stein E, Bathe KJ

- (eds) Nonlinear finite element analysis in structural mechanics. Springer, New York, pp 63–89
- Riks E (1979) An incremental approach to the solution of snapping and buckling problems. *Int J Solids Struct* 15(7):529–551
- Schneider W, Brede A (2005) Consistent equivalent geometric imperfections for the numerical buckling strength verification of cylindrical shells under uniform external pressure. *Thin-Walled Struct* 43(2):175–188
- Schneider W, Timmel I, Höhn K (2005) The conception of quasi-collapse-affine imperfections: a new approach to unfavourable imperfections of thin-walled shell structures. *Thin-Walled Struct* 43(8):1202–1224
- Stricklin JA, Haisler WE, von Riesenmann WA (1973) Evaluation of solution procedures for material and/or geometrically nonlinear structural analysis. *AIAA J* 11(3):292–299

Nonlinear Seismic Ground Response Analysis of Local Site Effects with Three-Dimensional High-Fidelity Model

Tsuyoshi Ichimura¹, Kohei Fujita^{2,3} and Muneo Hori¹

¹Earthquake Research Institute, University of Tokyo, Bunkyo, Tokyo, Japan

²Department of Architecture and Architectural Engineering, Graduate School of Engineering, Kyoto University, Kyoto, Japan

³RIKEN Advanced Institute for Computational Science, Kobe, Japan

Synonyms

Characterization and spatial variability of earthquake ground motions; Nonlinear finite element analysis; Site response

Introduction

An accurate estimation of seismic ground motion plays an important role, in order to realize rational countermeasures against an earthquake disaster. This review focuses on seismic ground response analysis above engineering bedrock. As has been pointed out, in a large earthquake, a local concentration of seismic ground motion due to complex

ground structure often results in inducing severe damage to structures (e.g., see Liang and Sun (2000) for a review of relationships between complex ground structures and damage to structures). Although many countermeasures have been implemented, there remain sites in which local concentration of ground motion takes place; for example, a concentration of earthquake ground motion due to the complex ground structure caused severe damage to a buried gas pipeline network in the 2011 Tohoku Earthquake (Advisory Committee for Natural Resources and Energy et al. 2012). Because the impedance contrast between the engineering bedrock and the sedimentary layer is sharp, seismic ground motion is greatly amplified in sedimentary layers of complicated configuration showing nonlinear behavior. As a result, remarkably different seismic ground motion may be observed at even very close points. To prevent damage to structures due to locally concentrated seismic ground motion, a method is needed which can estimate ground motion amplification due to complex ground structure with high resolution and high accuracy.

A candidate for such an estimation method is observation that uses a dense network of seismographs. For example, a superdense seismograph network is operated in Tokyo (Shimizu et al. 2000), where the network of average distance between seismographs being 3 km has been installed for the real-time monitoring of earthquakes. However, it is difficult to estimate a local concentration of earthquake ground motion even with such a superdense seismograph network because the complex ground structure varies over distances of 10^{1-2} m. Also, it is difficult to estimate strain distribution because a seismograph can only observe point-wise wave profiles.

Another candidate is numerical simulation. The increase in ground structure data now enables the construction of a high-resolution three-dimensional (3D) numerical model that considers surface geometries and heterogeneities of the ground structure as faithfully as possible; from now on, this model is referred to as a “high-fidelity model.” Although 3D nonlinear seismic ground response analysis with the high-fidelity model can be expected to accurately estimate 3D local

site effects, it is not an easy task to conduct such numerical analyses because of the huge computation and modeling costs. As a result, numerical simulation with a simplified ground model is conventionally used. A representative analysis is a one-dimensional (1D) amplification analysis (e.g., SHAKE (Schnabel et al. 1972)). Because the 3D ground structure is approximated to a layered medium in the 1D amplification analysis, the 3D problem is reduced to the 1D problem. Stiffness and damping change momentarily, depending on shear strain due to 1D wave propagation. Although the 1D amplification analysis can estimate nonlinear seismic ground responses with less computational cost, it is not possible to estimate the local concentration which is caused by a 3D complex ground structure, by using the stratified-layer approximation. Recently, a two-dimensional (2D) nonlinear seismic ground response analysis has often been conducted for complex ground structure. It is quite natural that the 2D analysis gives more accurate estimations than the 1D analysis. However, a serious problem remains for the 2D analysis, with regard to selection of a suitable 2D cross section of the ground structure. Thus, 3D nonlinear seismic ground response analysis with a high-fidelity model is desirable.

Because numerical simulation performance and ground structure data have increased with recent advances in Information and Communications Technology (ICT), nonlinear seismic ground response analysis which computes a 3D high-fidelity model is becoming possible (e.g., Ichimura et al. (2014a, b)). Such an analysis enables estimation of the local concentration and strain distribution, which are difficult to observe. This review explains the detail and utility of the nonlinear seismic ground response analysis with a 3D high-fidelity model (Ichimura et al. 2014a).

Three-Dimensional Nonlinear Seismic Ground Response Analysis Method with a High-Fidelity Model

Because the finite element method (FEM) can accurately generate a 3D high-fidelity numerical model for complex ground structures and

accurately satisfy traction-free boundary conditions on the surface, the following nonlinear dynamic FEM with unstructured finite elements is used for the nonlinear seismic ground response analysis:

$$\left(\frac{4}{dt^2}\mathbf{M} + \frac{2}{dt}\mathbf{C}^n + \mathbf{K}^n\right)\delta\mathbf{u}^n = \mathbf{F}^n - \mathbf{Q}^{n-1} + \mathbf{C}^n\mathbf{v}^{n-1} + \mathbf{M}\left(\mathbf{a}^{n-1} + \frac{4}{dt}\mathbf{v}^{n-1}\right), \quad (1)$$

with

$$\begin{cases} \mathbf{Q}^n = \mathbf{Q}^{n-1} + \mathbf{K}^n\delta\mathbf{u}^n, \\ \mathbf{u}^n = \mathbf{u}^{n-1} + \delta\mathbf{u}^n, \\ \mathbf{v}^n = -\mathbf{v}^{n-1} + \frac{2}{dt}\delta\mathbf{u}^n, \\ \mathbf{a}^n = -\mathbf{a}^{n-1} - \frac{4}{dt}\mathbf{v}^{n-1} + \frac{4}{dt^2}\delta\mathbf{u}^n. \end{cases} \quad (2)$$

Here, $\delta\mathbf{u}$, $\mathbf{u}$, $\mathbf{v}$, $\mathbf{a}$, and $\mathbf{F}$ indicate the incremental displacement, displacement, velocity, acceleration, and external force vectors, respectively. $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ indicate mass, damping, and stiffness matrices, respectively. dt and n indicate time increment and time step number, respectively. The Rayleigh damping matrix is used for $\mathbf{C}$; the element damping matrix $\mathbf{C}_e^n$ is calculated using the element mass matrix $\mathbf{M}_e$ and the element stiffness matrix $\mathbf{K}_e^n$ as

$$\mathbf{C}_e^n = \alpha\mathbf{M}_e + \beta\mathbf{K}_e^n.$$

Here, α and β are determined by solving the least squares equation,

$$\text{minimize} \left[\int_{f_{\min}}^{f_{\max}} \left(h^n - \frac{1}{2} \left(\frac{\alpha}{2\pi f} + 2\pi f \beta \right) \right)^2 df \right],$$

where $f_{\max}$, $f_{\min}$, and h^n indicate the maximum and minimum target frequencies and the damping ratio at time step n , respectively. Note that the damping matrix is calculated for every time step because the stiffness and damping ratio change by time step number, according to the constitutive relation of the ground material. Small elements are generated locally when modeling complex geometry with solid elements. Satisfying the

Courant condition when using an explicit time integration method (e.g., central difference method) for such small elements leads to small time increments and a huge computational cost. To resolve this problem, the Newmark- β method ($\beta = 1/4$, $\delta = 1/2$) is used for the time integration. Note that an explicit scheme is selected to compute the response robustly; an implicit scheme sometimes does not reach an accurate solution because of complex nonlinear constitutive relation. Thus, to verify the convergence of a solution, attention must be paid to the spatial and temporal discretization.

The most significant difference between the ordinary FEM analysis and the present nonlinear seismic ground response FEM analysis is the treatment of boundary conditions. Because the target domain is cut off from a half-infinite ground structure, half-infinite and absorbing conditions must be added on the bottom and side boundaries of the domain that describe the half-infinite continuity of the ground structure in the nonlinear seismic ground response FEM analysis. Several methods have been proposed, and one simple and effective method is based on damper and free motion ($\mathbf{u}^f$) on the side and bottom boundaries; $\mathbf{u}^f$ is computed based on the 1D wave theory, by applying the stress caused by $\mathbf{u}^f$ onto the side and bottom boundaries and by absorbing the scattered wave ($\dot{\mathbf{u}} - \dot{\mathbf{u}}^f$) on the side and bottom boundaries by the damper. This method can describe the half-infinite continuity with the assumption that the ground structure on the side boundaries is extended infinitely.

The computation is performed by repeating the following steps:

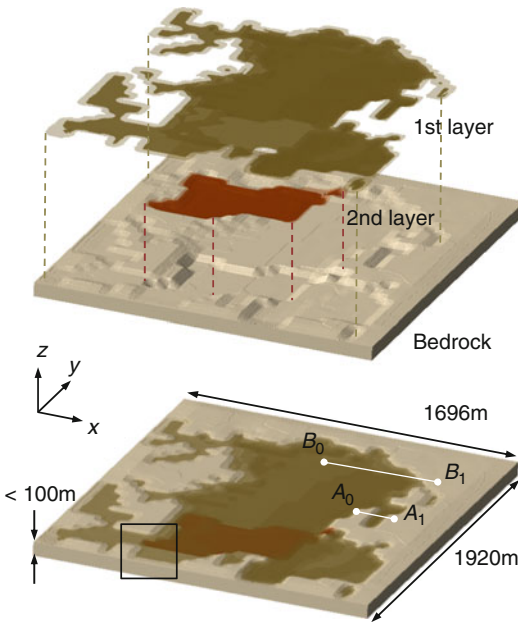
1. Compute stiffness and damping ratio for n -th time step with nonlinear constitutive relation using $\mathbf{u}^{n-1}$.
2. Compute $\mathbf{K}^n$ and $\mathbf{C}^n$ using stiffness and damping ratio for n th time step.
3. Compute $\delta\mathbf{u}^n$ by solving Eq. 1, and update values in Eq. 2 using $\delta\mathbf{u}^n$.

The bottlenecks of the nonlinear seismic ground response FEM analysis with a 3D high-fidelity model are the modeling cost and the

computation cost. For example, with the size of the target region of the order of $10^3 \times 10^3 \times 10^2$ m and a spatial resolution in the FE model of the order of $10^{-1\sim 0}$ m to satisfy the temporal resolution of a few Hz, the degrees of freedom (DOF) of the FE model become of the order of $10^{7\sim 9}$ and the time steps needed in the analysis would be of the order of $10^{3\sim 4}$. Because constructing such a large 3D FE model is difficult and the computational cost becomes enormous, it is still difficult to make a nonlinear, dynamic analysis for such a large problem. A fully automated modeling method that can generate elements of high quality is needed for constructing a 3D FE model of ground structures with large DOF. Here, the elements of high quality mean that the aspect ratio of the elements is small; an element with large aspect ratio behaves more stiffly, leading to a significant loss of accuracy in the analysis results. Although it is common to perform manual tuning to avoid elements of low quality, manual tuning of a huge model is extremely costly. Thus, a suitable method for the automated generation of an FE model is desirable. The remaining bottleneck is the large computational cost. The unknown in Eq. 1 is $\delta\mathbf{u}^n$; this is solved for each time step to obtain the time-history response of $\mathbf{u}$. Most of the computation time is spent solving the matrix equation of Eq. 1. Thus, a fast solver is also desirable.

Seismic Ground Motion Computed with a 3D High-Fidelity Model

Seismic ground motion is computed by nonlinear seismic ground response FEM analysis with a 3D high-fidelity model (see Ichimura et al. (2014a for details)). This example was an effort to reproduce the earthquake ground motion distribution in the 2011 Tohoku-Oki earthquake. The size of the target region modeled was 1,696 m in the east to west direction and 1,920 m in the north to south direction, where the soft sedimentary layers form a complex soil structure (see Fig. 1 for the 3D high-fidelity model of the target ground structures). The site is known to have had significantly larger



Nonlinear Seismic Ground Response Analysis of Local Site Effects with Three-Dimensional High-Fidelity Model, Fig. 1 Overall view of 3D ground structure model

observed seismic ground motion compared with that of nearby observation sites in past earthquake events. The modified Ramberg-Osgood model (Idriss et al. 1978) and the Masing rule (Masing 1926) were used as the nonlinear constitutive relation of soil material, and semi-infinite domain absorbing boundary conditions for the bottom and side boundaries of the domain were implemented, as explained above.

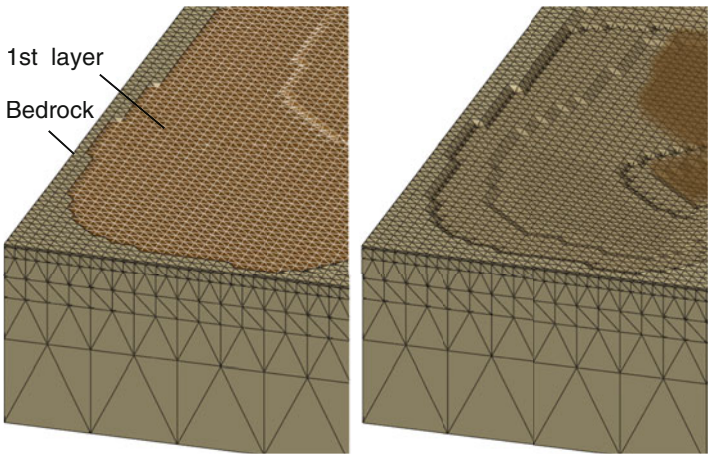
Here, a fully automated modeling method of Ichimura et al. (2009) with modifications was applied to constructing a 3D FE model, with the underground 3D soil structure being defined as a layered medium. A structured grid was used as background cells covering the target domain, and elements were made for each background cell. Here, the elements were generated in each cell so that they have low aspect ratios. This leads to a fully automated, robust method for making elements of high quality. Because the generation of elements in each cell is performed individually, it is suitable for parallelization, leading to fast element generation using parallel computation. First,

elements with linear-tetrahedral elements and cubic elements are generated, and the both types of elements are converted to second-order tetrahedral elements, leading to a 3D FE soil structure model consisting of only second-order tetrahedral elements. Also, a fast solver was implemented with the Block-Jacobi method, mixed-precision arithmetic, a geometric multigrid method, a predictor-corrector method, and parallel computing using hybrid parallelization MPI-OpenMP to reduce computational cost and shorten computation time.

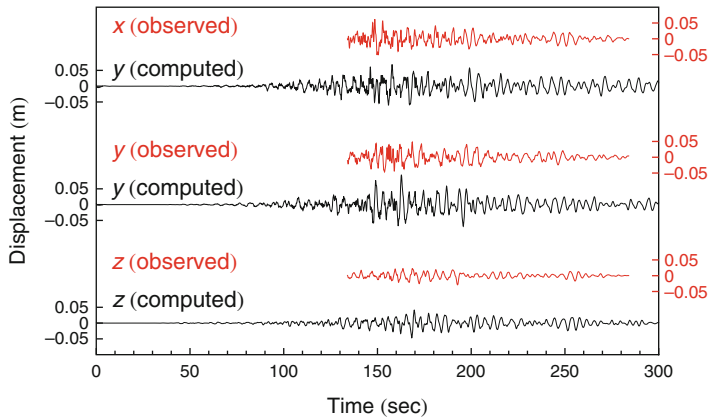
Figure 2 shows a close-up view of the 3D FE soil structure model. It can be seen that the complex 3D geometry consisting of three soil layers is modeled faithfully. The minimum size of the second-order tetrahedral elements is 4 m. With 32,509,107 DOF, the number of tetrahedral elements and nodes are 7,779,048 and 10,836,369, respectively. Using this 3D FE model, a seismic response analysis with nonlinear dynamic FE analysis of 60,000 time steps with 0.005 s time increments was conducted. Here, Eq. 1 is solved for each time step with relative error less than 10×10^{-6} . A PC cluster with eight computing nodes was used, each computing node with dual hexa-core Intel Xeon X5680 CPUs, connected with the InfiniBand QDR network. Using this small system, an analysis can be conducted in only 1,107,495 s. On average, one step is solved in 18.46 s, which is fast enough, considering that a 32 million DOF matrix equation is solved. This analysis time would be expected to become much shorter as a faster computer or a larger computer system is used. Due to recent technological progress, a dynamic nonlinear FE analysis is becoming feasible for application to nonlinear soil response analyses.

The analysis results and observed data are compared at observation points shown in Fig. 3. The computed and observed waves matched well, although no tuning was performed on the input wave, soil parameters, or soil model. It is also possible to obtain better conformity using dense observation networks to tune models based on data obtained from such dense networks. The seismic ground motion distribution is evaluated at the surface. Figure 4a shows the maximum

Nonlinear Seismic Ground Response Analysis of Local Site Effects with Three-Dimensional High-Fidelity Model, Fig. 2 3D ground structure FE model. The close-up view of the region in the rectangle in Fig. 1

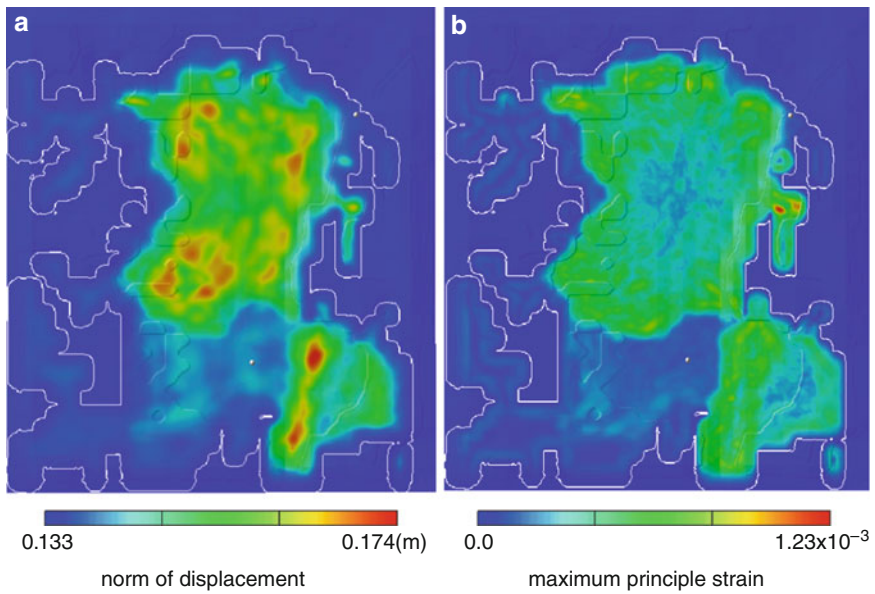


Nonlinear Seismic Ground Response Analysis of Local Site Effects with Three-Dimensional High-Fidelity Model, Fig. 3 Comparison of observed and computed waves



values of magnitude of displacement. It can be seen that the maximum displacement distribution has a high correlation with the depth of layer 1. To evaluate the strain in the axial direction (one of the damage indices for pipeline structure), the maximum values of principal strain in the whole time history are shown in Fig. 4b. Compared with the distribution of maximum displacement, it can be seen that estimating the distribution of strain from the ground structure is difficult; using numerical results of the seismic ground response analysis is effective for evaluating the distribution of strain. Two observation lines (A_0A_1 and B_0B_1) are set that imitate pipelines in parts of the model with nonuniform ground, and the axial strain is analyzed (see Fig. 1). Line A_0A_1 has a

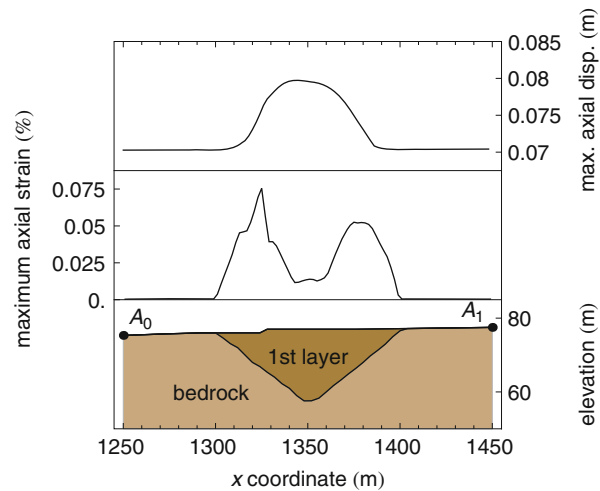
V-shaped depth distribution in layer 1, which is similar to that reported in Tsukamoto et al. (1984), and Line B_0B_1 goes through the ground structure with a constantly increasing depth of layer 1, followed by a region with constant depth of layer 1. Figures 5 and 6 show the maximum axial strain and the soil structure under each line. The axial strain along line A_0A_1 is similar to the observed strain on buried pipelines in V-shaped surface soil layers reported in Figs. 1 and 2 of Tsukamoto et al. (1984). The response in the bedrock was small but became larger for the soft layer, with a nearly constant response in the center of the V-shaped valley. This leads to small axial strain in the center of the V-shaped valley and large strain between the bedrock and the soft



Nonlinear Seismic Ground Response Analysis of Local Site Effects with Three-Dimensional High-Fidelity Model, Fig. 4 Distribution of time-history maximum values of norm of displacement and maximum principal strain

Nonlinear Seismic Ground Response Analysis of Local Site Effects with Three-Dimensional High-Fidelity Model,

Fig. 5 Maximum axial displacement, maximum axial strain, and underground structure of line A_0A_1

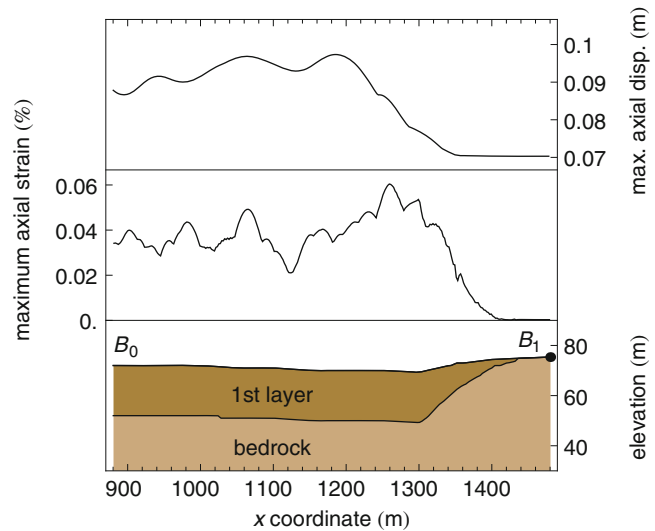


soil layer. This phenomenon is the cause of the pipeline damage reported in Liang and Sun (2000) and Tsukamoto et al. (1984). The response of line B_0B_1 is more complex than that of A_0A_1 . Similar to line A_0A_1 , high axial strain occurs between the bedrock and the soft layers. On the other hand, a complex distribution of axial strain

can be observed from $x=900\text{--}1,200\text{ m}$, with a horizontally layered structure. This is because the soil structure around line B_0B_1 is complex and cannot be approximated by a two-dimensional structure, as shown in Fig. 6; the bedrock leans out from the northern side (see Fig. 1). From these observations, it can be seen

Nonlinear Seismic Ground Response Analysis of Local Site Effects with Three-Dimensional High-Fidelity Model,

Fig. 6 Maximum axial displacement, maximum axial strain, and underground structure of line B_0B_1



that nonuniform ground causes a complex ground motion distribution, suggesting that even 2D analyses could not be suitable to evaluating ground motion in some cases.

Summary

Increased ground structure data, computer performance, and numerical simulation abilities with advances in ICT now enable the execution of nonlinear seismic ground response analysis with a 3D high-fidelity model. Such analysis can accurately estimate earthquake ground motion in a complex ground structure model which is constructed from observed data. Even today, strain distribution can be estimated in a small computer environment; in the near future, such analyses are expected to become commonplace.

Cross-References

- ▶ [Engineering Characterization of Earthquake Ground Motions](#)
- ▶ [Geotechnical Earthquake Engineering: Damage Mechanism Observed](#)
- ▶ [Nonlinear Dynamic Seismic Analysis](#)
- ▶ [Nonlinear Finite Element Analysis](#)
- ▶ [Physics-Based Ground-Motion Simulation](#)

- ▶ [Seismic Actions Due to Near-Fault Ground Motion](#)
- ▶ [Seismic Design of Pipelines](#)
- ▶ [Site Response: 1-D Time Domain Analyses](#)
- ▶ [Site Response: Comparison Between Theory and Observation](#)
- ▶ [Spatial Variability of Ground Motion: Seismic Analysis](#)

References

- Advisory Committee for Natural Resources and Energy, Urban Area Thermal Energy Committee Gas Safety Subcommittee, Working Group for Earthquake Disaster Prevention (2012) Report on disaster mitigation for gas supply in view of great east Japan earthquake 2011 (in Japanese)
- Ichimura T, Hori M, Bielak J (2009) A hybrid multiresolution meshing technique for finite element three-dimensional earthquake ground motion modeling in basins including topography. *Geophys J Int* 177:1221–1232
- Ichimura T, Fujita K, Hori M, Sakanoue T, Hamanaka R (2014a) Three-dimensional nonlinear seismic ground response analysis of local site effects for estimating seismic behavior of buried pipelines. *J Press Vessel Technol, ASME* 136:041702
- Ichimura T, Fujita K, Tanaka S, Hori M, Lalith M, Shizawa Y, Kobayashi H (2014b) Physics-based urban earthquake simulation enhanced by 10.7 BlnDOF $\times$ 30 K time-step unstructured FE Non-linear seismic wave simulation. In: *Proceedings of the international conference for high performance*

- computing, networking, storage and analysis, New Orleans, Nov 2014
- Idriss IM, Singh RD, Dobry R (1978) Nonlinear behavior of soft clays during cyclic loading. *J Geotech Eng* 104:1427–1447
- Liang J, Sun S (2000) Site effects on seismic behavior of pipelines: a review. *J Press Vessel Technol* 122:469–475
- Masing G (1926) Eigenspannungen und verfestigung beim messing. In: *Proceedings of the 2nd international congress of applied mechanics*, pp 332–335 (in German)
- Schnabel PB, Lysmer J, Seed HB (1972) SHAKE: a computer program for earthquake response analysis of horizontally layered sites, UCB/EERC-72/12. University of California, Berkeley
- Shimizu Y, Watanabe A, Koganemaru K, Nakayama W, Yamazaki F (2000) Super high-density realtime disaster mitigation system. In: *12th world conference on earthquake engineering*, New Zealand
- Tsukamoto K, Nishio N, Satake M, Asano T (1984) Observation of pipeline behavior at geographically complex site during earthquake. In: *8th world conference on earthquake engineering*, San Francisco, vol 7, pp 247–254

Nonlinear System Identification: Particle-Based Methods

Eleni N. Chatzi¹ and Andrew W. Smyth²

¹Department of Civil, Environmental and Geomatic Engineering, ETH Zurich, Institute of Structural Engineering, Zurich, CH, Switzerland

²Department of Civil Engineering and Engineering Mechanics, Columbia University, New York, NY, USA

Synonyms

Joint state and parameter estimation; Nonlinear systems; Online identification; Particle filter; Structural identification; Unscented Kalman filter

Introduction

The presence of nonlinearity in structural systems is inevitably linked to the dissipation of energy occurring either within particular components of the structure, such as joints or damper elements,

or as a result of exceeding material capacity which in turn leads to response that lies outside the usual assumption of the elastic range. Therefore, nonlinearity in the form of hysteretic behavior is commonly observed in structures subjected to significant levels of excitation such as earthquake, wind, or sea waves. Typically, the description of the restoring force that is associated with the behavior of such systems involves the use of nonlinear laws which render the governing system of equations a highly nonlinear one.

The task of joint parameter and state estimation for such systems is nontrivial as it requires the implementation of methods able to account for nonlinearity. The complexity of the problem is further increased when a real-time estimation is sought. The latter is quite commonly the case in structural health monitoring (SHM) implementations and in vibration control applications where the real-time feedback of the restoring force is necessary. In serving such a purpose, time domain identification methods constitute a class of methods particularly well suited for the online monitoring of dynamic systems. Numerous techniques have been proposed for time domain identification in civil engineering, including the more simplified autoregressive-moving average (ARMA) models, the least squares estimation (LSE) method, the Eigensystem realization algorithm (ERA), as well as the Bayesian approximation techniques, such as the observer/Kalman filter identification (OKID), the extended Kalman filter (EKF), the unscented Kalman filter (UKF), and the sequential Monte Carlo methods. Some of these techniques lend themselves more readily than others to problems where the system dynamics or the measurement equations are expressed as nonlinear functions of the states.

The EKF has by far been the most extensively used identification algorithm, for the case of nonlinear systems, over the past 30 years, and has been applied for a number of civil engineering applications, such as structural damage identification, parameter identification of inelastic structures, and so forth. It is based on the propagation of a Gaussian random variable (GRV) through the first-order linearization of the state-space model of the system. Despite

its wide use, the EKF is only reliable for systems that are almost linear on the time scale of the updating intervals. In what is presented herein, alternative techniques are presented that prove more efficient in handling higher-order nonlinearities. The UKF is an extension of the Kalman filter that does not require the calculation of Jacobians in order to linearize the state equations. Instead, the state is approximated by a GRV which is now represented by a set of carefully chosen points. These sample points completely capture the true mean and covariance of the GRV when propagated through the actual nonlinear system.

The sequential Monte Carlo method, also known as the particle filter (PF), is especially useful since it is capable of handling nonlinear systems described by non-Gaussian posterior probability of the state, as well as systems with non-Gaussian noise. This attribute makes the particular method quite attractive for SHM applications. The implementation concept is the approximation of the posterior probability of the state through the generation of a large number of samples that are appropriately weighted. Particle filters are essentially an extension to point-mass filters with the difference that the particles are no longer uniformly distributed over the state but instead concentrated in regions of high probability. A significant drawback is the fact that depending on the problem, a large number of samples may be required, thus rendering the PF analysis computationally expensive. Additionally, this method is faced with what is referred to as the sample impoverishment problem, which essentially refers to the loss of diversity among particles as the analysis progresses and can be quite pronounced when dealing with parameter identification problems, as is typically the case in SHM. In what follows, the workings of the methods are described and enhancements are presented to effectively treat issues arising in SHM applications.

As a literature note, a number of authors have implemented time domain modeling approaches within a monitoring framework (Juang and Pappa 1985; Juang et al. 1993; Smyth et al.

2002; Zhang et al. 2002; Nagarajaiah and Li 2004; Yang et al. 2006; Fraraccio et al. 2008). Additionally, several authors have experimented with the use of the EKF on nonlinear identification problems (Corigliano and Mariani 2004; Mariani and Corigliano 2005; Yun and Shinozuka 1980). The alternative nonlinear KF extension, i.e., the unscented Kalman filter, was first introduced in 1994 when J. Uhlmann suggested that rather than approximating the known system function through linearization, it might be beneficial to apply the exact nonlinear function to an approximating probability distribution (Julier and Uhlmann 1997). Wan and Van der Merwe (2000) further verified the validity of this proposal. In the case of SHM applications, Wu and Smyth have shown that the UKF is superior to the EKF for joint state and parameter estimation problems while proving more robust for increased measurement noise levels for high-dimensional systems (Wu and Smyth 2007). Chatzi and Smyth further demonstrated the potential of the UKF in balancing rapid estimation with high precision (Chatzi et al. 2010; Chatzi and Smyth 2009). The particle filter methods have also been widely applied for the case of nonlinear identification problems (Arulampalam et al. 2002; Chen et al. 2005; Ching et al. 2006; Nasrellah and Manohar 2011; Eftekhari et al. 2012).

The Dynamic Problem Formulation and the Optimal Bayesian Solution

Within a structural health monitoring framework, the target commonly is the tracking of dynamic response, described by the general continuous state-space formulation

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) + \mathbf{v}(t) \quad (1)$$

or by its discretized equivalent which is better suited for SHM implementations where information is provided via sampling at an interval Δt :

$$\mathbf{x}_k = \mathbf{F}(\mathbf{x}_{k-1}) + \mathbf{v}_{k-1} \quad (2)$$

The information on the system is provided through noisy observations of the form

$$\mathbf{y}_k = H(\mathbf{x}_k) + \boldsymbol{\eta}_k \quad (3)$$

where $\mathbf{x}_k \in \mathbb{R}^n$ is the **state variable vector** at time $t = k\Delta t$, $\mathbf{v}_k$ is the zero mean **process noise** vector with covariance matrix $\mathbf{Q}_k$, $\mathbf{y}_k \in \mathbb{R}^m$ is the current **observation vector**, and $\boldsymbol{\eta}_k$ is the **observation noise** vector with corresponding covariance matrix $\mathbf{R}_k$. It is worth noting that the regimes described herein do not require feedback from all of the system's degrees of freedom (dof's). Instead, we might have limited information drawn from few sensors on selected locations of the structure/system.

From a Bayesian perspective, the problem of determining the estimate of $\mathbf{x}_k$ based on the sequence of all available measurements up to time step k , $\mathbf{y}_k$, is a recursive one, involving the construction of a posterior PDF $p(\mathbf{x}_k|\mathbf{y}_{1:k})$. Assuming the prior distribution, $\mathbf{y}_0$, is known and that the required PDF $p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})$ at time $k-1$ is available, the prior probability $p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ can be obtained sequentially through:

Step 1 Prediction (Chapman-Kolmogorov Equation)

$$p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1})d\mathbf{x}_{k-1} \quad (4)$$

The probabilistic model of the state evolution $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, also referred to as transitional density, is defined by the process Eq. 2. Consequently, the measurement $\mathbf{y}_k$ is taken into account in formulating a posterior estimate through:

Step 2 Update (Bayesian Theorem)

$$p(\mathbf{x}_k|\mathbf{y}_{1:k}) = p(\mathbf{x}_k|\mathbf{y}_k, \mathbf{y}_{1:k-1}) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_{1:k-1})}{p(\mathbf{y}_k|\mathbf{y}_{1:k-1})} \quad (5)$$

where the likelihood function $p(\mathbf{y}_k|\mathbf{x}_k)$ is completely defined by the observation Eq. 3.

Once the posterior PDF is known, the optimal estimate can be computed using different criteria, one of which is the minimum mean square error (MMSE) estimate. Although, the Bayesian solution is hard to compute analytically, the particle-based methods described herein in fact are approximations of this concept. For further details, the interested reader is pointed to the work of Ristic and Arulampalam (2004).

The Joint State and Parameter Estimation Problem

Within the context of SHM implementations, it is often the case that the parameters describing the model of the system are either unknown or known with some uncertainty. To this end, it is often desirable to additionally identify the time-invariant parameters of the system model. In a joint state and parameter estimation problem, the state vector of Eq. 1 is augmented in order to include the parameters to be identified, i.e.,

$$\tilde{\mathbf{x}} = \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\theta} \end{bmatrix}, \quad \mathbf{x} \in \mathbb{R}^n, \quad \boldsymbol{\theta} \in \mathbb{R}^l \quad (6)$$

where $\boldsymbol{\theta}$ stands for the unknown parameter vector of dimension l . Given that this component is time invariant, Eq. 2 is modified as follows:

$$\tilde{\mathbf{x}}_k = \begin{bmatrix} \mathbf{x}_k \\ \boldsymbol{\theta}_k \end{bmatrix} = \begin{bmatrix} F(\mathbf{x}_{k-1}) \\ \boldsymbol{\theta}_{k-1} \end{bmatrix} + \begin{bmatrix} \mathbf{v}_{k-1} \\ \mathbf{v}_{\theta_{k-1}}^0 \end{bmatrix} \quad (7)$$

where $\mathbf{v}_{\theta_{k-1}}^0$ is an added process noise vector, relating to the time-invariant parameter components.

The Unscented Kalman Filter Method

The UKF approximates the state as a GRV, and therefore, assuming $N(\mathbf{x}; \boldsymbol{\mu}, \mathbf{P})$ is a Gaussian density with argument $\mathbf{x}$, mean $\boldsymbol{\mu}$, and covariance $\mathbf{P}$, the UKF relates to the two steps of the Bayesian approach Eqs. 4 and 5 through the following relationships:

$$\begin{aligned} p(\mathbf{x}_{k-1}|\mathbf{y}_{1:k-1}) &= N(\mathbf{x}_{k-1}; \hat{\mathbf{x}}_{k-1|k-1}, \mathbf{P}_{k-1|k-1}) \\ p(\mathbf{x}_k|\mathbf{y}_{1:k-1}) &= N(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k-1}, \mathbf{P}_{k|k-1}) \\ p(\mathbf{x}_k|\mathbf{y}_{1:k}) &= N(\mathbf{x}_k; \hat{\mathbf{x}}_{k|k}, \mathbf{P}_{k|k}) \end{aligned} \quad (8)$$

In the classic Kalman filter formulation, outlined in the seminal paper by Kalman (1960), the prediction and update steps are formalized in a set of equations derived on the basis of linearity. In the case examined herein, the governing equations are nonlinear and a discretized equivalent is utilized for the propagation of the statistics of $\mathbf{x}_k \in \mathbb{R}^n$ through the system. The unscented transformation is used for this purpose which involves the computation of a set of discrete points, termed the sigma points, χ_k^i , with corresponding weights w_i . In order to do so, the original state vector is redefined as the concatenation of the original state vector and noise variables as $\mathbf{x}_{k-1}^\alpha = [\mathbf{x}_{k-1}^T \mathbf{v}_{k-1}^T \boldsymbol{\eta}_{k-1}^T]^T$. Then, the sigma points are distributed symmetrically around the mean $\hat{\mathbf{x}}_{k-1}^\alpha \in \mathbb{R}^L$ as follows:

$$\chi_{k-1} = \begin{bmatrix} \hat{\mathbf{x}}_{k-1}^\alpha & \hat{\mathbf{x}}_{k-1}^\alpha + \sqrt{(L+\lambda)\mathbf{P}_k^\alpha} \\ \hat{\mathbf{x}}_{k-1}^\alpha - \sqrt{(L+\lambda)\mathbf{P}_k^\alpha} \end{bmatrix} \quad (9)$$

Each of the columns of the above matrix corresponds to a sigma point vector $\chi_{k-1}^i \in \mathbb{R}^L$ with components corresponding to the state, process, and measurement noise, respectively, and $\mathbf{P}_k^\alpha$ is the corresponding augmented covariance vector, incorporating the process and observation noise components:

$$\chi_{k-1}^i = \begin{bmatrix} \chi_{k-1}^{x,i} \\ \chi_{k-1}^{v,i} \\ \chi_{k-1}^{\eta,i} \end{bmatrix}, \quad \mathbf{P}_{k-1}^\alpha = \begin{bmatrix} \mathbf{P}_{k-1} & 0 & 0 \\ 0 & \mathbf{Q}_{k-1} & 0 \\ 0 & 0 & \mathbf{R}_{k-1} \end{bmatrix} \quad (10)$$

The weights for the state and covariance components of the sigma point vector are calculated as

$$\begin{aligned} w_0^x &= \frac{\lambda}{L+\lambda}, w_0^p = \left[\frac{\lambda}{L+\lambda} + (1-\alpha^2+\beta) \right] \\ w_i^x &= w_i^p = \frac{1}{2(L+\lambda)}, \quad i = 1, \dots, 2L \end{aligned} \quad (11)$$

where α, κ, β are the KF parameters, $\lambda = \alpha^2(L+\kappa) - L$ is a scaling parameter. α and κ are scaling

factors (for a two-dimensional state vector, $\kappa = 0$ is optimal), and β is related to the distribution (equals to 2 for Gaussian priors). Finally, L is the dimension of the augmented state vector.

Step 1 UKF Prediction

The state and process noise components of the sigma point vectors are propagated through the nonlinear function $F(\mathbf{x}_k)$ without any need for linearization:

$$\chi_{k|k-1}^{x,i} = F\left(\chi_{k-1}^{x,i}, \chi_{k-1}^{v,i}\right), \quad i = 0, \dots, 2L \quad (12)$$

The discrete set of the sample points, $\chi_{k|k-1}^{x,i} \in \mathbb{R}^n$, is utilized to derive a representation of the prior density $p(\mathbf{x}_k | \mathbf{y}_{1:k-1})$. The prior mean and covariance are approximated as the weighted sample mean and covariance of the prior sigma points and the time update step is continued as follows:

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{i=0}^{2L} w_i^x \chi_{k|k-1}^{x,i} \quad (13)$$

$$\mathbf{P}_{k|k-1} = \sum_{i=0}^{2L} w_i^p \left[\chi_{k|k-1}^{x,i} - \hat{\mathbf{x}}_{k|k-1} \right] \left[\chi_{k|k-1}^{x,i} - \hat{\mathbf{x}}_{k|k-1} \right]^T \quad (14)$$

Step 2 UKF Update

An estimate of the observation vector (reflecting the measured quantities) may then be approximated as

$$\hat{\mathbf{y}}_{k|k-1} = \sum_{i=0}^{2L} w_i^x H\left(\chi_{k|k-1}^{x,i}, \chi_{k-1}^{\eta,i}\right) \quad (15)$$

where the state and observation noise components of the sigma point vectors are propagated through the nonlinear observation function $H(\mathbf{x}_k)$. Finally, the posterior estimate $\hat{\mathbf{x}}_{k|k}$ is obtained as the weighted sum of the prior $\hat{\mathbf{x}}_{k|k-1}$ and the measurement residual, $\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}$, which reflects the discrepancy between the predicted measurement and the actual measurement. The Kalman gain matrix at time step k , $\mathbf{K}_k$, acts as the weighing factor:

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{y}_k - \hat{\mathbf{y}}_{k|k-1}) \text{ where } \mathbf{K}_k = \mathbf{P}_k^{xy} (\mathbf{P}_k^{yy})^{-1}$$

$$\text{with } \begin{cases} \mathbf{P}_k^{yy} = \sum_{i=0}^{2L} w_i^p \left[H(\mathbf{x}_{k|k-1}^{x,i}) - \hat{\mathbf{y}}_{k|k-1} \right] \left[H(\mathbf{x}_{k|k-1}^{x,i}) - \hat{\mathbf{y}}_{k|k-1} \right]^T \\ \mathbf{P}_k^{xy} = \sum_{i=0}^{2L} w_i^p \left[\mathbf{x}_{k|k-1}^{x,i} - \hat{\mathbf{x}}_{k|k-1} \right] \left[H(\mathbf{x}_{k|k-1}^{x,i}) - \hat{\mathbf{y}}_{k|k-1} \right]^T \end{cases} \quad (16)$$

The posterior state covariance is also calculated as

$$\mathbf{P}_k = \mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{P}_k^{yy} \mathbf{K}_k^T \quad (17)$$

Figure 1 illustrates the workings of the unscented Kalman filter.

The unscented Kalman filter, as its linear equivalent, is quite sensitive to the selection of the initial state covariance $\mathbf{P}_0$ as well as the noise covariance matrices $\mathbf{Q}_k$, $\mathbf{R}_k$. As in the linear KF case, the autocovariance least-squares (ALS) technique (Rajamani and Rawlings 2009) can be utilized which employs autocovariances of routinely operating data for estimating the noise covariance matrices.

Gaussianity is made for the state and noise components. The general particle filter, however, does not make any prior assumption on the state distribution. Instead, the posterior probability density function (PDF), $p(\mathbf{x}_k | \mathbf{y}_{1:k})$, is approximated via a set of random samples, also known as support points $\mathbf{x}_k^i$, $i = 1, \dots, N$, with associated weights w_k^i . This means that the probability density function at time k can be approximated as follows:

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) = \sum_{i=1}^N w_k^i \delta(\mathbf{x}_k - \mathbf{x}_k^i) \quad (18)$$

where

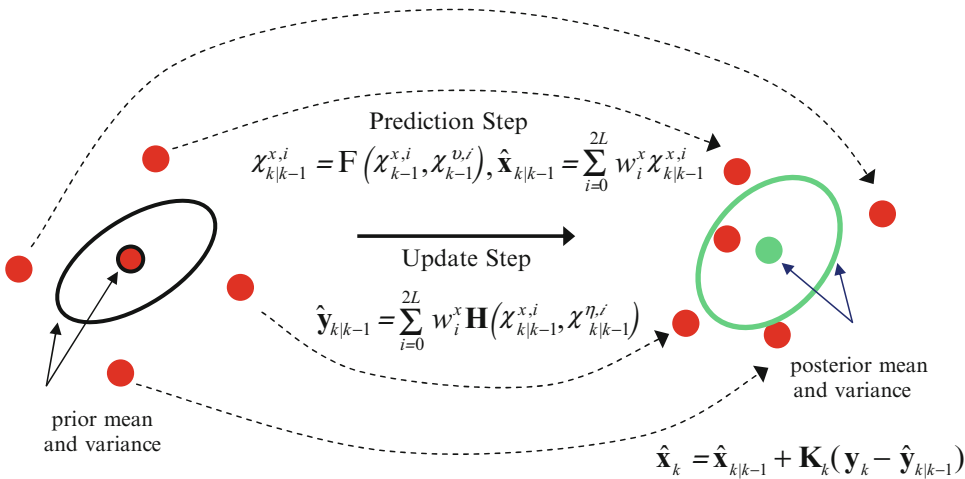
$$w_k^i \propto \frac{p(\mathbf{x}_k^i | \mathbf{y}_{1:k})}{q(\mathbf{x}_k^i | \mathbf{y}_{1:k})} \quad (19)$$

N

The Particle Filter Method

The unscented Kalman filter is in fact a particular case of the particle filter where the assumption of

where $\mathbf{x}_k^i$ are the N samples drawn at time step k from the **importance density function** $q(\mathbf{x}_k^i | \mathbf{y}_{1:k})$. The appropriate selection of this



Nonlinear System Identification: Particle-Based Methods, Fig. 1 The unscented Kalman filter process for a two-dimensional state

function constitutes a fundamental consideration in this class of methods as explained next.

Using the state-space assumptions (1st-order Markov process/observational independence for a given state), the importance weights can be estimated recursively by

$$w_k^i \propto w_{k-1}^i \frac{p(\mathbf{y}_k | \mathbf{x}_k^i) p(\mathbf{x}_k^i | \mathbf{x}_{k-1}^i)}{q(\mathbf{x}_k^i | \mathbf{x}_{k-1}^i, \mathbf{y}_k)} \quad (20)$$

where $p(\mathbf{x}_k^i | \mathbf{x}_{k-1}^i)$ is the transitional density, defined by the process Eq. 2, and $p(\mathbf{y}_k | \mathbf{x}_k)$ is the likelihood function, defined by the observation Eq. 3. As the number of samples increases, this sequential Monte Carlo approach becomes an equivalent representation to the function description of the PDF and the solution approaches the optimal Bayesian estimate.

The Importance Density Function

An important issue in the implementation of particle filters is the selection of the importance density. It has been proved that the optimal importance density function that minimizes the variance of the true weights is given by

$$\begin{aligned} q(\mathbf{x}_k | \mathbf{x}_{k-1}^i, \mathbf{y}_{1:k})_{\text{opt}} &= p(\mathbf{x}_k | \mathbf{x}_{k-1}^i, \mathbf{y}_k) \\ &= \frac{p(\mathbf{y}_k | \mathbf{x}_k, \mathbf{x}_{k-1}^i) p(\mathbf{x}_k | \mathbf{x}_{k-1}^i)}{p(\mathbf{y}_k | \mathbf{x}_{k-1}^i)} \end{aligned} \quad (21)$$

However, sampling from $p(\mathbf{x}_k | \mathbf{x}_{k-1}^i, \mathbf{y}_k)$ might not be straightforward, leading to the use of the transitional prior as the importance density function which greatly simplifies the analysis process:

$$q(\mathbf{x}_k | \mathbf{x}_{k-1}^i, \mathbf{y}_{1:k}) = p(\mathbf{x}_k | \mathbf{x}_{k-1}^i) \quad (22)$$

This essentially means that at time step k , the samples $\mathbf{x}_k^i$ are drawn from the transitional density, which is entirely defined by the process equation. Additionally, Eq. 20 yields

$$w_k^i = w_{k-1}^i p(\mathbf{y}_k | \mathbf{x}_k^i) \quad (23)$$

Therefore, the selection of the importance weights is essentially dependent on the likelihood

of the error between the estimate and the actual measurement, as this is defined by the observation function H and the assumed properties of the observation noise $\boldsymbol{\eta}_k$.

The weights are then normalized so that their sum equals unity. In relating this process to the prediction and update steps referenced earlier, a two-stage methodology is once again derived where:

Step 1 PF Prediction

Given the value of the N discrete state vectors (particles), at the previous time step, $\mathbf{x}_{k-1}^i$, these are propagated through the process equation of the dynamic system in order to yield a prior estimate:

$$\mathbf{x}_{k|k-1}^i = F(\mathbf{x}_{k-1}^i) + \mathbf{v}_{k-1}, \quad i = 1, \dots, N \quad (24)$$

Step 2 PF Update

The evaluation of the importance weights through the use of the likelihood function essentially constitutes the measurement update step, leading to the calculation of the posterior estimate through the weighted mean of the sample points:

$$\hat{\mathbf{x}}_k = \sum_{i=1}^N w_k^i \mathbf{x}_{k|k-1}^i \quad (25)$$

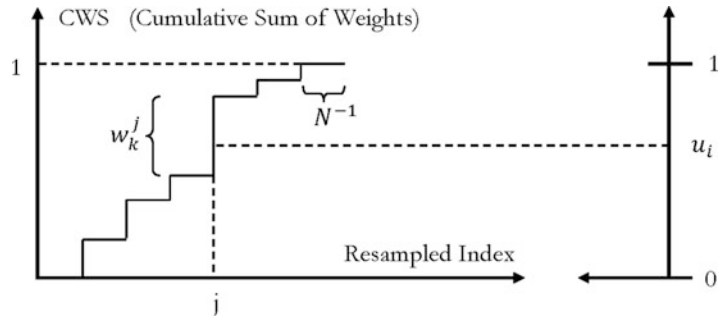
The Problem of Degeneracy and the Resampling Process

A second issue pertaining to the implementation of particle filters (PFs) is that of **degeneracy**, meaning that after some time steps, importance weights are unevenly distributed; thus, considerable computational effort is spent on updating particles with “trivial” contribution to the approximation of $p(\mathbf{x}_k | \mathbf{y}_{1:k})$. This may create numerical instabilities and calls for appropriate treatment. In quantifying this divergence, a measure of degeneracy has been specified, also known as the effective sample size:

$$N_{\text{eff}} = \frac{1}{\sum_{i=1}^N (w_k^i)^2} \quad (26)$$

Nonlinear System Identification: Particle-Based Methods,

Fig. 2 The process of resampling: the uniformly distributed random variable u_i maps into index j ; thus, the corresponding particle $\mathbf{x}_k^j$ is likely to be selected due to its considerable weight w_k^j



Resampling is a technique developed for the tackling of degeneracy. It discards those particles with negligible weights and enhances the ones with more significant weights. Resampling takes place when N_{eff} falls below some user-defined threshold N_T . A particle is more likely to be selected as a member of the remaining set, if it corresponds to a higher weight as schematically shown in Fig. 2.

The workings of the various stages of the particle filter algorithm are illustrated in Fig. 3.

The Problem of Sample Impoverishment

The use of the resampling technique however is known to lead to other problems, relating to the loss of diversity of particles. As the higher-weight particles are duplicated, diversity is lost leading to the **sample impoverishment (or particle depletion) phenomenon**. Such a phenomenon is more likely to occur when the process noise levels are low and is particularly problematic in the case of joint state and parameter estimation. As demonstrated in section “[The Joint State and Parameter Estimation Problem](#)” and in the example that follows, in such a case the state vector is augmented to include the unknown, but time-invariant (constant) system parameters yielding state components with minimal variability throughout the analysis process. Since the addition of a significant amount of process noise would lead to algorithm convergence problems and instability issues, one typically needs to resort to the use of a large number of particles. Although, both the UKF and PF present the advantage of parallel implementation, the use of an excessive number of particles inevitably incurs

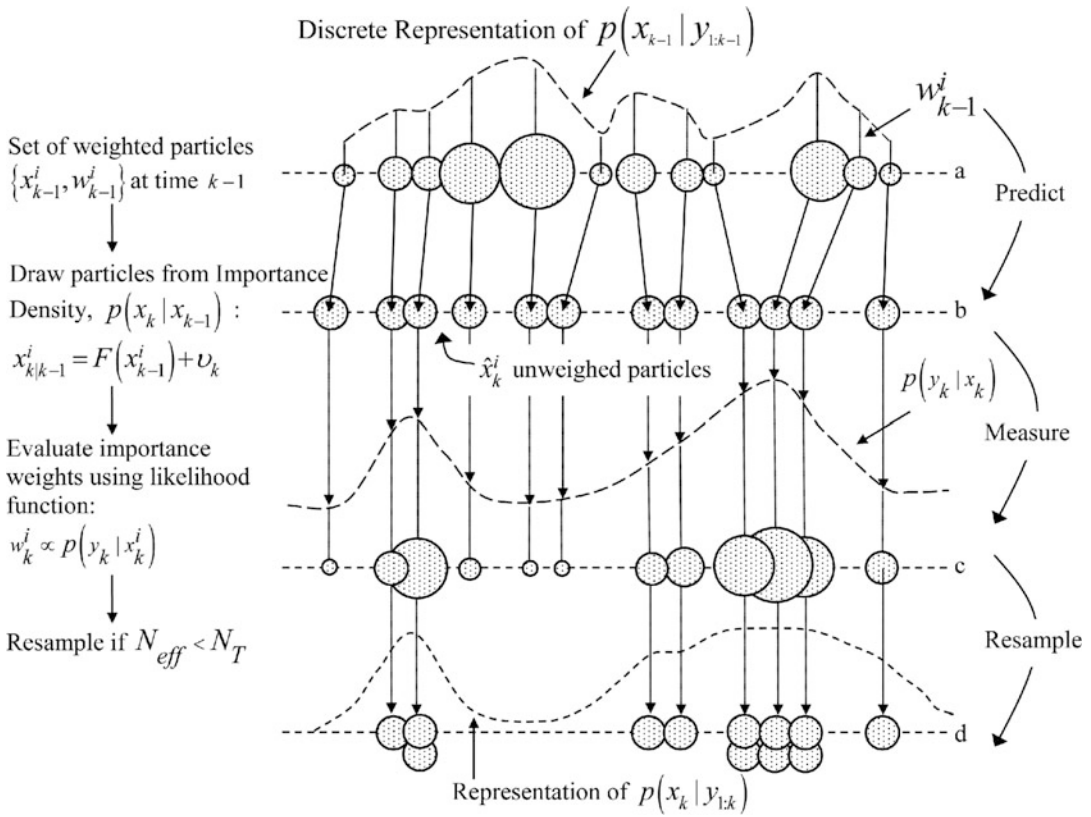
a significant computational cost which can be a major disadvantage. The UKF is free of such a problem as the conditioning of the state via the Kalman gain induces a variability in the evolution of the parameter components $\boldsymbol{\theta}$ of the augmented state vector $\tilde{\mathbf{x}}$ of Eq. 7.

More information on the selection of the importance weights based on importance sampling can be found in Doucet et al. (2001), Bergman (1999), and Bergman et al. (2001).

The Particle Filter with Mutation

In order to tackle the sample impoverishment problem, Chatzi and Smyth (2013) proposed an enhancement of the PF, termed the particle filter with mutation (MPF), which incorporates a mutation operator in the resampling process. **Mutation** is a process typically used in genetic algorithms (GAs) where it serves as a means of maintaining diversity among the members of a population. Within the framework of GAs, mutation is typically enforced under two regimes, the *creep mutation* and the *jump mutation*. Jump mutations involve random modifications in the binary encoding of the system’s variables, whereas creep mutation takes place in the real number representation of the variables (i.e., the phenotype). The operation implemented in the MPA scheme resembles the creep mutation process.

In SHM implementations or system identification problems where the estimation of system parameters is required, the state vector is commonly augmented in order to include these

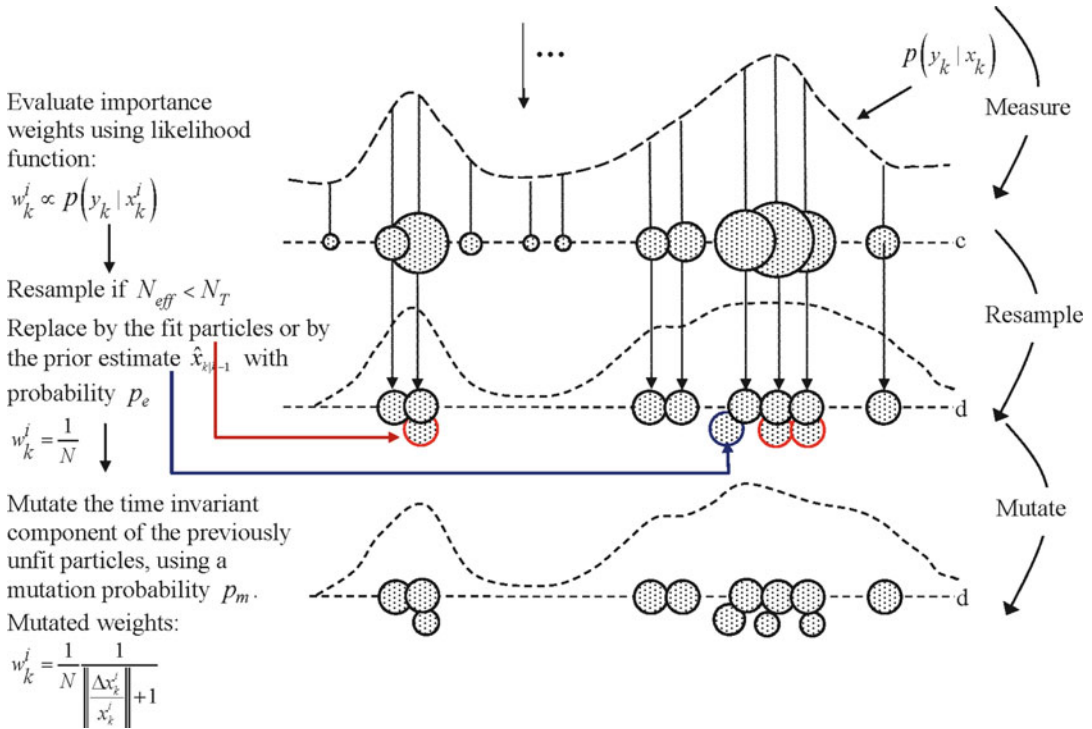


Nonlinear System Identification: Particle-Based Methods, Fig. 3 The particle filter implementation steps

parameters as in Eq. 6. Consequently, the state-space equations are also expanded to include the zeroth derivative equations that enforce the time invariance of those parameters, which are considered as constants. Equation 7 reflects with the discrete time equivalent, indicating the time invariance of the parameter components. Hence, especially in the case of higher complexity problems, the initial particle selection space needs to be appropriately spanned and quite densely sampled in order to achieve an efficient parameter estimate. In other words the initial sample space needs to include seeds sufficiently close to the true parameter values for a successful prediction. In order to cope with this restriction, the MPF algorithm features a twofold innovation, enforced during the resampling step, i.e., when the effective sample size, N_{eff} , drops below a certain threshold.

Propagating the Weighted Estimate of the State

Firstly, part of the formerly unfit particles is replaced in step k by the prior estimate of the weighted mean of the state, obtained as $\hat{\mathbf{x}}_{k|k-1} = F(\hat{\mathbf{x}}_{k-1})$. Replacement is performed under a uniform probability, p_e , i.e., if $\text{rand} < p_e$, where rand is a uniformly distributed random number, then replacement takes place; otherwise, the particles remain unchanged. Function F is the state-space function of the process Eq. 2. In this manner, the actual value of the weighted estimate is incorporated in the particles and propagated through the nonlinear system. Formerly, the optimal estimate would only appear as the weighted result and the actual fitness for that particle was not explicitly evaluated at any step of the algorithm. For strongly nonlinear systems, this may lead to imprecise estimations.



Nonlinear System Identification: Particle-Based Methods, Fig. 4 The particle filter with mutation implementation steps

Mutation of the Time Invariant Components

Secondly, after the formerly unfit particles have been replaced by either particles of more significant weight (standard resampling) or by the propagated value of $\hat{x}_{k-1}$, their time-invariant components are mutated. Mutation takes place by shifting the parameter components by a random amount using a mutation probability p_m . Both the mutation probability and the extent of the shifting interval can be user specified. The mutated particles are assigned a weight that is inversely proportional to the relative difference between the mutated vector and the original one ("parent" vector) according to the following relationship:

$$w_k^i = \frac{1}{N} \frac{1}{\left\| \frac{\Delta x_k^i}{x_k^i} \right\| + 1} \quad (27)$$

whereas the weights of the non-mutated resampled particles remain equal to $w_k^i = \frac{1}{N}$.

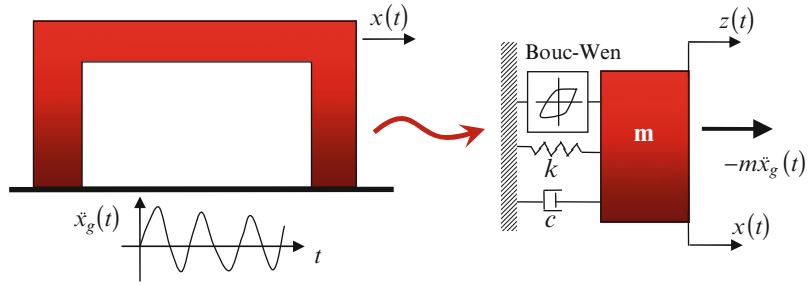
In expression Eq. 27, Δx_k^i is the difference between the mutated vector and the parent one and $\| \cdot \|$ is the L_2 norm. The full set of weights is then once again normalized before proceeding to the next time step. A graphical representation of the MPF, relating it to the stages previously outlined for the standard PF, is displayed in Fig. 4.

As a literature note, the blending of the standard PF with evolutionary concepts has only very recently been explored in the literature. Akhtar et al. (2011) propose a Particle Swarm Optimization accelerated Immune Particle Filter (PSO-ac-IPF). Park et al. (2007) propose the so-called genetic filter which involves a standard genetic algorithm step (with all three GA operators, i.e., crossover, mutation, and selection) in place of standard resampling. Similarly, Kwok et al. (2005) employ the crossover operator to what they call the evolutionary particle filter.

The drawback of the previous approaches is that they more or less incorporate a GA or PSO

Nonlinear System Identification: Particle-Based Methods,

Fig. 5 One floor shear-frame structure with a Bouc-Wen-type material nonlinearity



(Particle Swarm Optimization) step in the process, often requiring several loops that might significantly delay what is already a relatively lengthy estimation procedure. More recently, Yu et al. (2010) have suggested the use of simply the mutation operator for an adaptive mutation PF. The scheme proposed by Yu et al. quite resembles the earlier-mentioned notion of adding artificial process noise to the state vector.

Application: Joint State and Parameter Estimation for a Nonlinear Shear-Frame Structure

The approaches outlined above are tested on the real-time system identification of a nonlinear single degree of freedom (sdf) system. The system is assumed to describe the hysteretic response of a one storey shear-type frame building. This type of structure is described by a single translational degree of freedom in the elastic case. In the present example, the structure exhibits material nonlinearity when subjected to a ground motion, simulating earthquake excitation. As noted in

Fig. 5, this type of system is equivalent to a mass-spring-dashpot-hysteretic element system. The Bouc-Wen model is used in order to account for nonlinear hysteretic response, which involves an additional hysteretic degree of freedom, henceforth denoted as $z(t)$.

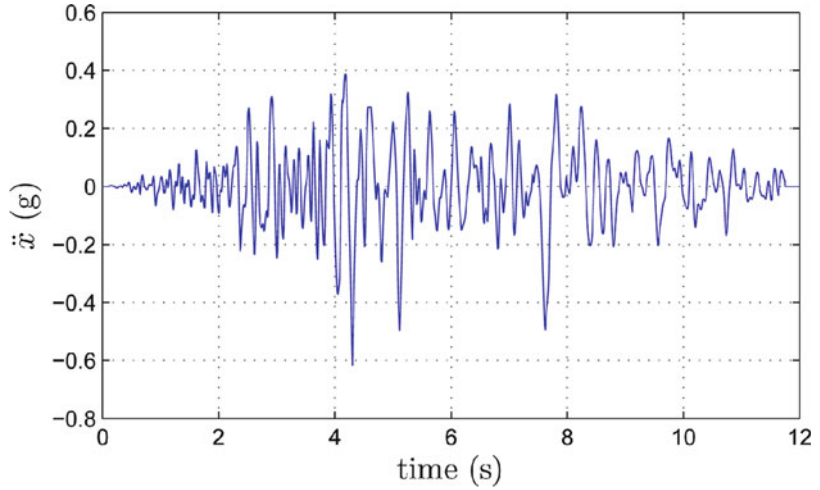
Using the Bouc-Wen model, it is quite straightforward to compile the system's governing equations of motion. In fact the Bouc-Wen enables an elegant and concise formulation of the nonlinear phenomenon of hysteresis. For reasons of simplicity, dependence on time is not explicitly denoted in the expressions that follow:

$$\begin{aligned} m\ddot{x} + c\dot{x} + \alpha kx + (1 - \alpha)kz &= -m\ddot{x}_g \\ \dot{z} &= \dot{x} - \beta|\dot{u}||z|^{n-1}z - \gamma\dot{u}|z|^n \end{aligned} \quad (28)$$

Additionally, we assume that we have measurements of the system's absolute acceleration, via the use of appropriate accelerometer sensors. Since the aim is the joint state and parameter identification, as the nonlinear dof $z(t)$ cannot be measured, the system is cast into the following state-space form:

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} \dot{x}(t) \\ -\frac{c}{m}\dot{x}(t) - \alpha\frac{k}{m}x(t) - (1 - \alpha)\frac{k}{m}z(t) - \ddot{x}_g(t) \\ \dot{x}(t) - \beta|\dot{u}(t)||z(t)|^{n-1}z(t) - \gamma\dot{u}(t)|z(t)|^n \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \mathbf{v}(t) \quad (29)$$

Nonlinear System Identification: Particle-Based Methods,
Fig. 6 The input ground acceleration: Northridge earthquake



where $\mathbf{v} \in \mathbb{R}^8$ is the process noise vector, and $\tilde{\mathbf{x}} \in \mathbb{R}^8$ is the augmented vector that includes the displacement x , velocity $\dot{x}$, hysteretic dof z , as well as the stiffness k , damping c , and Bouc-Wen parameter constants α, β, γ . Hence, the parameter vector $\boldsymbol{\theta} \in \mathbb{R}^5$ is defined as $\boldsymbol{\theta} = [k \ c \ \alpha \ \beta \ \gamma \ n]^T$ and

$$\tilde{\mathbf{x}} = [x \ \dot{x} \ z \ k \ c \ \beta \ \gamma \ n]^T \quad (30)$$

The state-space (process) equation is complimented by the accompanying measurement (observation) equation:

$$\begin{aligned} y(t) = \ddot{x}_{tot}(t) = & -\frac{c}{m}\dot{x}(t) - \alpha\frac{k}{m}x(t) \\ & - (1 - \alpha)\frac{k}{m}z(t) + \boldsymbol{\eta}(t) \end{aligned} \quad (31)$$

where $\boldsymbol{\eta} \in \mathbb{R}$ is the measurement noise vector. The process equation can be brought into a discrete form by implementing a simple integration scheme. In order to keep the process online, a simple Euler scheme is implemented herein and proves to be sufficient for our purposes.

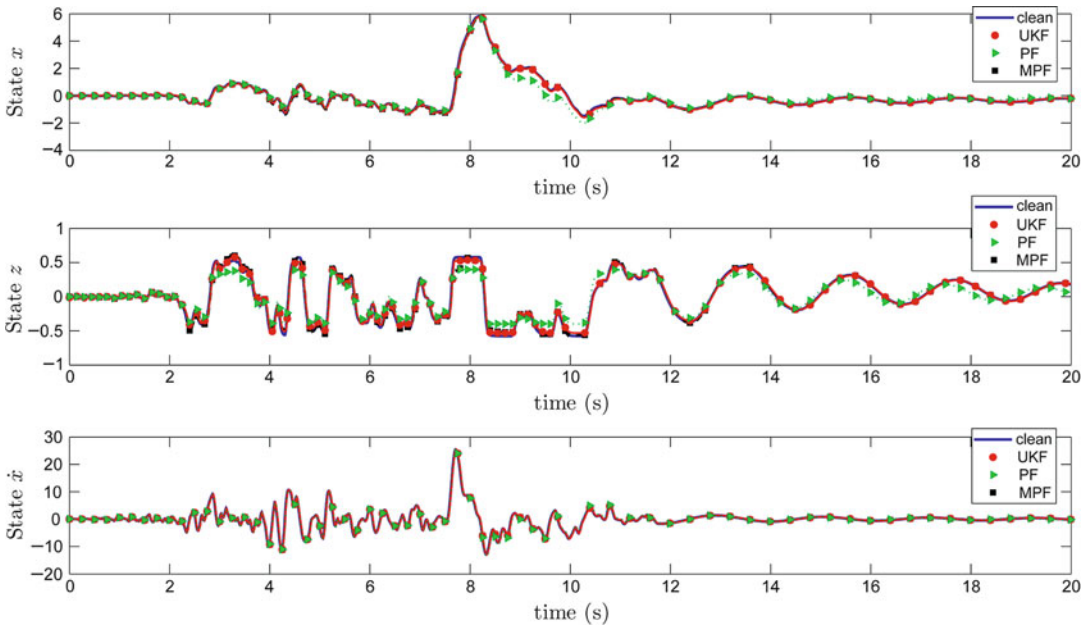
The actual response of the system is numerically generated using $k = 9$, $m = 1$, $c = .25$, $\alpha = 0.2$, $\beta = 2$, $\gamma = 1$ & $n = 2$ and the implemented ground motion is a scaled

record of the Northridge earthquake (1994) shown in Fig. 6. Furthermore, the discretization is performed using a sampling frequency of 100 Hz which is also the sampling frequency of the input ground motion. The analysis is performed for a total of 20 s.

The particle-based algorithms outlined in the previous sections are implemented, namely, the UKF, the PF, and the MPF algorithm. The UKF employs 21 particles ($=2*10 + 1$), the PF requires 5,000 particles for achieving admissible accuracy, and the MPF is run using 800 particles. The tuning parameters of the mutation operator for the PF are selected as $p_e = 0.2$, $p_m = 0.1$.

A Gaussian, process noise level of 0.1 % RMS noise-to-signal ratio is assumed, while the observation noise corresponds to approximately 10 % RMS. No process noise is added for the time-invariant parameter components. The addition of some minor noise could improve parameter estimation for the PF; however, it might also lead to instabilities and non-converging behavior. The observation noise is chosen so as to reflect a realistic instrumentation noise level, whereas the process noise is kept to a low level indicating confidence that the observed system can be described by this type of a model formulation.

Figure 7 illustrates the estimated state evolution for the three filters. It is already obvious that the standard particle filter (PF) underperforms



Nonlinear System Identification: Particle-Based Methods, Fig. 7 Estimated versus actual state evolution

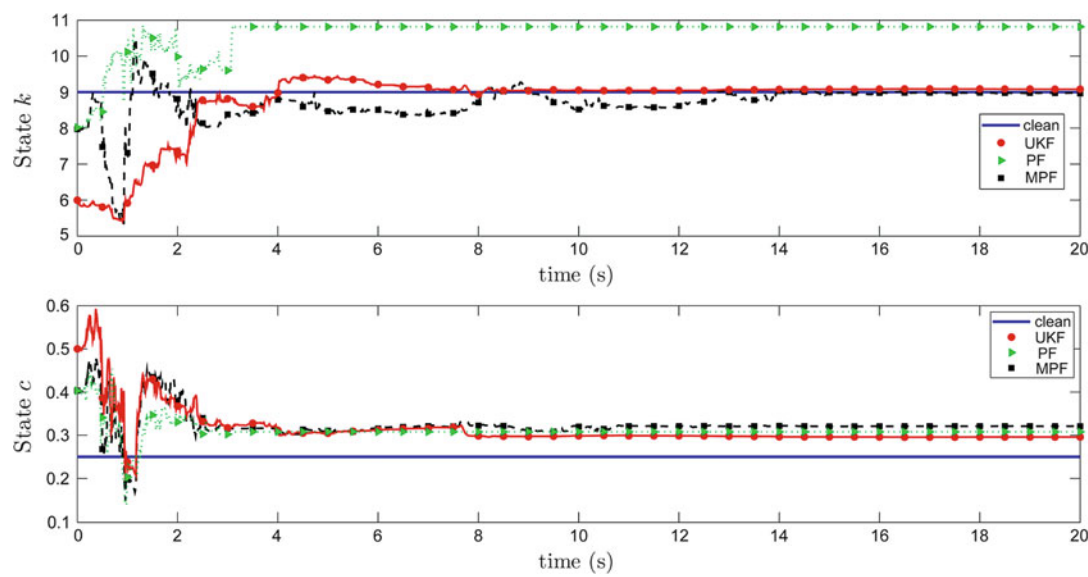
and is not able to accurately track the system's states with a relative error of $\sim 20\%$ for the displacement estimate. The accuracy of the other two filters is almost identical and very close to the actual time history, with a relative error of $\sim 3.5\%$ for the UKF and $\sim 3\%$ for the MPF. As verified by the parameter estimation results, summarized in Figs. 8 and 9, the PF fails to accurately identify the true parameter values, as a result of the sample impoverishment problem outlined in section "The Problem of Sample Impoverishment." The use of the mutation operator for the particle filter with mutation (MPF) on the other hand alleviates this problem. The latter is schematically presented in Figs. 10 and 11 where it is obvious that, for the PF, the particle group eventually degenerates to a single parameter set for the time-invariant parameters. The use of the mutation operator for the MPF succeeds in maintaining the diversity of the population with a small scatter around the finally identified value, which prevents the algorithm from reaching a premature convergence.

From a computational cost perspective, in this quite simple sdof problem, all of the suggested

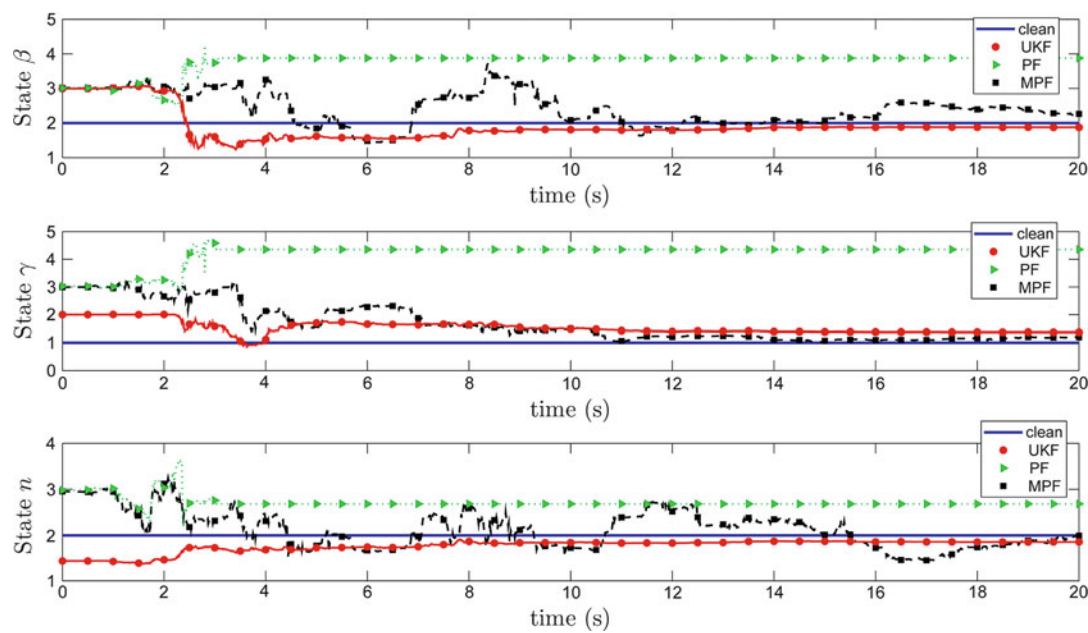
methods can be implemented in real time, i.e., on the fly, as data is acquired. The computational time required for the whole analysis on a 4 core CPU is of the order of 1.0 s for the UKF, 5.7 s for the PF (5,000 particles), and 1.5 s for the MPF. The required time would of course increase for a higher dimensionality; nonetheless, one of the benefits of particle-based regimes is the fact that they can be implemented in parallel, as the particle evaluations are mutually independent.

Summary

This reference entry describes the use of particle-based methods for joint state and parameter identification of a structural system for SHM purposes. This class of methods is adopted in order to tackle the difficulties arising due to the nonlinear nature of the physical system and the uncertainty related to our knowledge of the system characteristics. The workings of each method are described, and the advantages, limitations, and enhancements of the presented approaches



Nonlinear System Identification: Particle-Based Methods, Fig. 8 Estimated versus Actual Material Parameter evolution



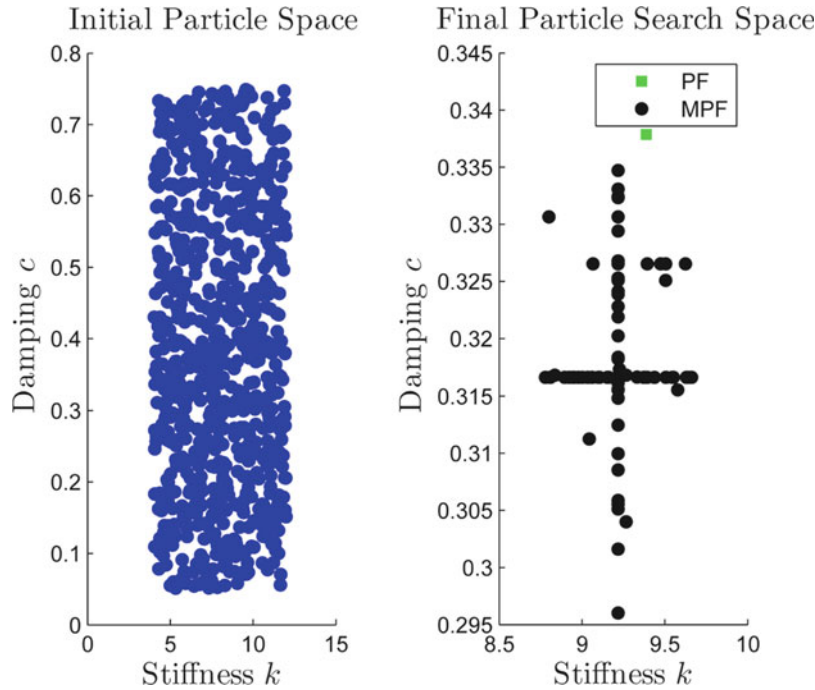
Nonlinear System Identification: Particle-Based Methods, Fig. 9 Estimated versus actual Bouc-Wen parameter evolution

are presented and discussed. As demonstrated, the use of particle-based techniques enables the real-time tracking of state evolution and the accurate identification of unknown system parameters

in a robust and reliable manner. The benefits would be even more pronounced if a parallel processing regime were adopted, drastically cutting down the required computational time.

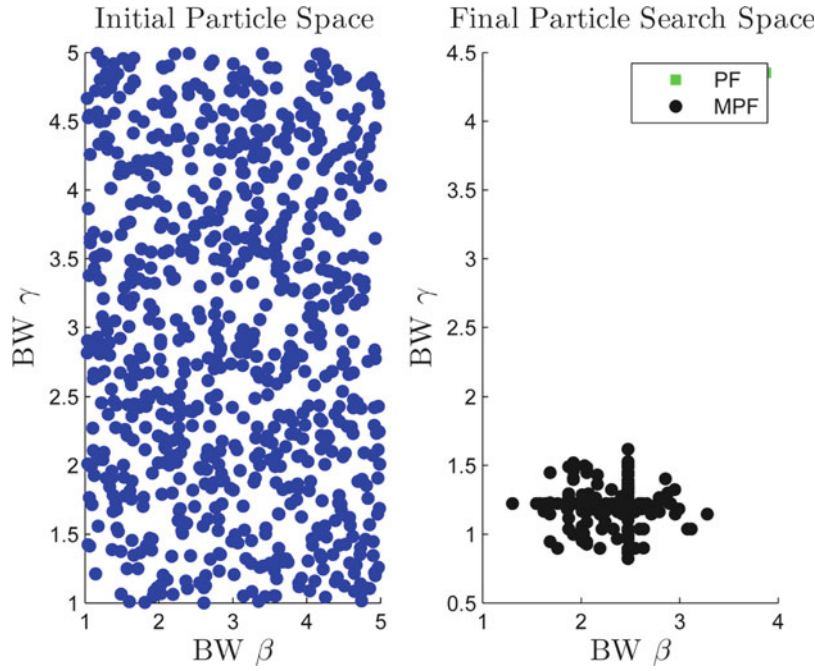
Nonlinear System Identification: Particle-Based Methods,

Fig. 10 Initial versus final particle space for the material parameters



Nonlinear System Identification: Particle-Based Methods,

Fig. 11 Initial versus final particle space for the BW parameters



Cross-References

- [Advances in Online Structural Identification](#)
- [Parametric Nonstationary Random Vibration Modeling with SHM Applications](#)
- [System and Damage Identification of Civil Structures](#)
- [Vibration-Based Damage Identification: The Z24 Bridge Benchmark](#)

References

- Akhtar S, Ahmad AR, Abdel-Rahman EM, Naqvi T (2011) A PSO accelerated immune particle filter for dynamic state estimation. In: Canadian Conference on Computer and Robot Vision, St. John's, Newfoundland, 25–27 May 2011. pp 72–79
- Arulampalam S, Maskell S, Gordon N, Clapp T (2002) A tutorial on particle filters for on-line non-linear/non-Gaussian Bayesian tracking. *IEEE T Signal Process* 50(2):174–188
- Bergman N (1999) Recursive Bayesian estimation: navigation and tracking applications. PhD thesis, Linköping University
- Bergman N, Doucet A, Gordon N (2001) Optimal estimation and Cramer Rao bounds for partial non-Gaussian state space models. *Ann I Stat Math* 53(1):97–112
- Chatzi EN, Smyth AW (2009) The unscented Kalman filter and particle filter methods for nonlinear structural system identification with non-collocated heterogeneous sensing. *Struct Contr Health Monit* 16–1:99–123
- Chatzi EN, Smyth AW (2013) Particle filter scheme with mutation for the estimation of time-invariant parameters in structural health monitoring applications. *J Struct Contr Health Monit* 20:1081–1095
- Chatzi EN, Smyth AW, Masri SF (2010) Experimental application of on-line parametric identification for nonlinear hysteretic systems with model uncertainty. *J Struct Safety* 32(5):326–337
- Chen T, Morris J, Martin E (2005) Particle filters for state and parameter estimation in batch processes. *J Process Contr* 15(6):665–673
- Ching J, Beck JL, Porter KA, Shaikhutdinov R (2006) Bayesian state estimation method for nonlinear systems and its application to recorded seismic response. *J Eng Mech* 132(4):396–410
- Corigliano A, Mariani S (2004) Parameter identification in explicit structural dynamics: performance of the extended Kalman filter. *Comput Method Appl Mech Eng* 193(36–38):3807–3835
- Doucet A, De Freitas JFG, Gordon NJ (eds) (2001) Sequential Monte Carlo methods in practice, Springer series in statistics for engineering and information science. Springer, New York
- Eftekhari AS, Ghisi A, Mariani S (2012) Parallelized sigma-point Kalman filtering for structural dynamics. *Comput Struct* 92–93:193–205
- Faraccio G, Brügger A, Betti R (2008) Identification and damage detection in structures subjected to base excitation. *Exp Mech* 48(4):521–528
- Juang JN, Pappa RS (1985) An eigensystem realization algorithm for modal parameter identification and model reduction. *J Guid Control Dynam* 8(5):620–627
- Juang JN, Phan MQ, Horta LG, Longman RW (1993) Identification of observer/Kalman filter Markov parameters: theory and experiments. *J Guid Control Dynam* 16(2):320–329
- Julier SJ, Uhlmann JK (1997) A new extension of the Kalman filter to nonlinear systems. In: Proceedings of AeroSense: the 11th international symposium on aerospace/defense sensing, simulation and controls, Orlando
- Kalman RE (1960) A new approach to linear filtering and prediction problems. *Trans ASME J Basic Eng* 82–(Series D):35–45
- Kwok N, Fang G, Zhou W (2005) Evolutionary particle filter: re-sampling from the genetic algorithm perspective. In: IEEE/RSJ international conference on intelligent robots and systems (IROS 2005), Edmonton, pp 2935–2940
- Mariani S, Corigliano A (2005) Impact induced composite delamination: state and parameter identification via joint and dual extended Kalman filters. *Comput Method Appl Mech Eng* 194(50–52):5242–5272
- Nagarajaiah S, Li Z (2004) Time segmented least squares identification of base isolated buildings. *Soil Dyn Earthq Eng J* 24(8):577–586
- Nasrellah HA, Manohar CS (2011) Particle filters for structural system identification using multiple test and sensor data: a combined computational and experimental study. *J Struct Contr Health Monit* 18(1):99–120
- Park S, Hwang J, Rou K, Kim E (2007) A new particle filter inspired by biological evolution: genetic filter. *Proc World Acad Sci Eng Technol* 21:459
- Rajamani MR, Rawlings JB (2009) Estimation of the disturbance structure from data using semidefinite programming and optimal weighting. *Automatica* 45:142–148
- Ristic NGB, Arulampalam S (2004) Beyond the Kalman filter, particle filters for tracking applications. Artech House Publishers, Boston
- Smyth AW, Masri SF, Kosmatopoulos EB, Chassiakos AG, Caughey TK (2002) Development of adaptive modeling techniques for non-linear hysteretic systems. *Int J Non-Linear Mech* 37(8):1435–1451
- Wan E, Van Der Merwe R (2000) The unscented Kalman filter for nonlinear estimation. In: Adaptive systems for signal processing, communications, and control symposium 2000, AS-SPCC. The IEEE 2000, Lake Louise, pp 153–158

- Wu M, Smyth AW (2007) Application of the unscented Kalman filter for real-time nonlinear structural system identification. *J Struct Contr Health Monit* 14(7): 971–990
- Yang JN, Lin S, Huang H, Zhou L (2006) An adaptive extended Kalman filter for structural damage identification. *J Struct Contr Health Monit* 13(4):849–867
- Yu JX, Tang YL, Liu WJ (2010) Adaptive mutation particle filter based on diversity guidance. In: 2010 international conference on machine learning and cybernetics (ICMLC), TBD Qingdao, China. vol 1, 11–14 July 2010, pp 369–374
- Yun CB, Shinozuka M (1980) Identification of nonlinear structural dynamics systems. *J Struct Mech* 8(2):187–203
- Zhang H, Foliente GC, Yang Y, Ma F (2002) Parameter identification of inelastic structures under dynamic loads. *Earthq Eng Struct Dynam* 31(5):1113–1130

- Dynamic loading due to wind gusts
- Any kind of irregular train of impacts or shocks

As the occurrence (arrival) times of the loads are random, their mathematically sound characterization is in terms of stochastic point (random counting) processes. The most fundamental and the simplest of such processes is a Poisson process. However, it is known that the Poisson process is an adequate model of a random train of events if the events are rare. Otherwise, other models, for example, the renewal processes, are more suitable. The impulse process excitation (or the random train of impulses) may be driven by different processes. If it is driven by a Poisson process, the state vector of the dynamic system is a nondiffusive, so-called Poisson-driven, Markov process. Pertinent mathematical tools such as generalized Itô's differential rule or Kolmogorov–Feller integrodifferential equation may be used. However, if the impulse process is non-Poisson (i.e., it is driven by a counting process other than Poisson), the state vector of the dynamic system is not a Markov process, and no mathematical tools may be directly applied. Then the problem has to be converted into a Markov one. Here-with it is explained how it may be achieved with the aid of two exact methods. The first one is the method of augmentation of state vector by additional state variables, which are pure-jump stochastic processes. In this approach the augmented state vector becomes a nondiffusive, Poisson-driven, Markov process, and the differential equations for response moments may be derived. In the second method, the state vector of the dynamic system is augmented by Markov states of an auxiliary, pure-jump stochastic process. This approach allows to obtain the set of integrodifferential equations governing the joint probability density of the state vector and also the differential equations for response moments. Two example non-Poisson impulse processes are considered: an Erlang renewal impulse process and the process driven by two independent Poisson processes.

Non-Poisson Impulse Processes

Radoslaw Iwankiewicz

Institute of Mechanics and Ocean Engineering,
Hamburg University of Technology, Hamburg,
Germany

Synonyms

Counting process; Equations for response statistical moments; Jump processes; Markov processes; Non-Poisson processes; Point process; Probability density; Random impulses; Random vibrations; Renewal processes

Introduction

Dynamic loads are adequately idealized as discontinuous stochastic processes, or random trains of loading events, in the following problems:

- Behavior of a vehicle traveling over a very rough road: impacts or shocks, due to sudden humps or holes
- Impact loads on booms of bucket-wheel excavators used in opencast mining engineering
- Moving loads on a bridge due to highway traffic

Stochastic Point Processes

Specification of a Random Counting Process

Let $N(t)$ denote a random counting process, or a random variable, specifying the number of events (or time points) in an interval $[0, t]$, i.e., excluding the event that possibly occurs at the time instant t . Consequently the sample paths of $N(t)$ are ever-increasing step (pure-jump) functions, which are left continuous with right limits. Strictly speaking, an additional assumption $\Pr\{N(0) = 0\} = 1$ should also be imposed. The occurrences of events are not, in general, assumed to be independent. The increment $dN(t)$ of the counting process during an infinitesimal time interval $[t, t + dt]$, denoted conventionally as $dN(t)$, is defined as $dN(t) = N(t + dt) - N(t)$.

The point process is *regular* or *orderly*, if the probability governing the counting measure satisfies the following condition:

$$\sum_{k \geq 1} \Pr\{dN(t) = k\} = O(dt^2), \quad (1)$$

which means that in the infinitesimal time interval, there can only occur, with nonzero probability, one event or no event.

Let the disjoint infinitesimal time intervals $[t_i, t_i + dt_i]$, $i = 1, 2, \dots, n$ be chosen from the interval $[0, t]$. Product density functions are defined as follows (Srinivasan 1974):

$$f_n(t_1, t_2, \dots, t_n) dt_1 \dots dt_n = E \left[\prod_{i=1}^n dN(t_i) \right]. \quad (2)$$

Equivalently, if the point process is regular, the n th degree product density function $f_n(t_1, \dots, t_n)$ represents the probability that one event occurs in each of disjoint intervals $[t_i, t_i + dt_i]$, irrespective of other events in the interval $[0, t]$; thus,

$$f_n(t_1, t_2, \dots, t_n) dt_1 \dots dt_n = \Pr \left\{ \bigwedge_{i=1}^n dN(t_i) = 1 \right\}. \quad (3)$$

$t_1 \neq t_2 \neq \dots \neq t_n.$

In what follows the attention is confined to regular or orderly point processes. In particular,

$$\Pr\{dN(t) = 1\} = f_1(t)dt, \quad (4)$$

where $f_1(t)$ is the product density of degree one. The regularity assumption Eq. 1 implies that

$$\Pr\{dN(t) = 0\} = 1 - f_1(t)dt + O(dt^2), \quad (5)$$

$$E[dN(t)] = f_1(t)dt + O(dt^2), \quad (6)$$

and

$$E[\{dN(t)\}^n] = f_1(t)dt + O(dt^2), \quad (7)$$

for arbitrary n .

Product density of degree one $f_1(t)$ represents the mean rate of occurrence of events (mean arrival rate). It should be noted that $f_1(t)$ is not a probability density; its integration over the whole time interval $[0, t]$ yields an expected number of events in this interval, which usually is not equal to one:

$$\begin{aligned} \int_0^t f_1(\tau) d\tau &= \int_0^t E[dN(\tau)] = E \left[\int_0^t dN(\tau) \right] \\ &= E[N(t)]. \end{aligned} \quad (8)$$

Product density of degree two, satisfying the relationship

$$f_2(t_1, t_2) dt_1 dt_2 = E[dN(t_1) dN(t_2)], \quad t_1 \neq t_2, \quad (9)$$

specifies the correlation between arrival rates at two different time instants t_1, t_2 (or the correlation of increments of the counting measure $N(t)$ on disjoint infinitesimal time intervals).

If k out of n time instants are set equal, i.e., $t_{j_1} = t_{j_2} = \dots = t_{j_k}$, or k out of n infinitesimal intervals all overlap, the product density of degree n degenerates to $(n - k + 1)$ th degree product density; thus,

$$\begin{aligned}
 E \left[\prod_{i=1}^n dN(t_i) \right] \Big|_{t_{i_1}=\dots=t_{i_k}} &= E \left[\prod_{i=1}^n dN(t_i) \{dN(t_{j_1})\}^k \right] \\
 i \neq j_r, \quad r &= 1, \dots, k \\
 &= f_{n-k+1}(t_1, \dots, t_n, t_{j_1}) dt_1 \dots dt_n dt_{j_1}.
 \end{aligned} \tag{10}$$

For example,

$$\begin{aligned}
 E[dN(t_1)dN(t_2)]|_{t_1=t_2} &= E \left[\{dN(t_1)\}^2 \right] \\
 &= f_1(t)dt.
 \end{aligned} \tag{11}$$

Joint density function, defined as

$$f_k(t_1, \dots, t_k) = \sum_{n=k}^{\infty} \frac{1}{(n-k)!} \underbrace{\int_0^t \dots \int_0^t}_{(n-k)\text{-fold}} \pi_n(t_1, \dots, t_k, t_{k+1}, \dots, t_n) dt_{k+1} \dots dt_n \tag{13}$$

$$\pi_k(t_1, \dots, t_k) = \sum_{n=k}^{\infty} \frac{(-1)^{n-k}}{(n-k)!} \underbrace{\int_0^t \dots \int_0^t}_{(n-k)\text{-fold}} f_n(t_1, \dots, t_k, t_{k+1}, \dots, t_n) dt_{k+1} \dots dt_n. \tag{14}$$

The probability that exactly n events occur in the time interval $[0, t]$ is evaluated as (Srinivasan 1974)

$$\Pr\{N(t) = n\} = \frac{1}{n!} \underbrace{\int_0^t \dots \int_0^t}_{n\text{-fold}} \pi_n(t_1, t_2, \dots, t_n) dt_1 dt_2 \dots dt_n. \tag{15}$$

Moreover, the correlation functions of the n th degree are defined in terms of product densities as (Stratonovich 1963)

$$\begin{aligned}
 g_1(t) &= f_1(t), \\
 g_2(t_1, t_2) &= f_2(t_1, t_2) - f_1(t_1)f_1(t_2), \\
 g_3(t_1, t_2, t_3) &= f_3(t_1, t_2, t_3) - 3\{f_1(t_1)f_2(t_2, t_3)\}_s \\
 &\quad + 2f_1(t_1)f_1(t_2)f_1(t_3),
 \end{aligned} \tag{16}$$

$$\begin{aligned}
 &\pi_n(t_1, t_2, \dots, t_n) dt_1 dt_2 \dots dt_n \\
 &= \Pr \left\{ \bigwedge_{i=1}^n dN(t_i) = 1 \wedge N(t) = n \right\},
 \end{aligned} \tag{12}$$

specifies the probability that one event occurs in each of disjoint intervals $[t_i, t_i + dt_i]$ and there are no other events in the whole time interval $[0, t]$, i.e., that there are exactly $N(t) = n$ events.

The following relationships between the product density and the joint density functions hold (Srinivasan 1974):

where $\{\cdot \cdot \cdot\}_s$ denotes the symmetrizing operation, i.e., the arithmetic mean of all terms similar to the one in brackets and obtained by all possible permutations of t_1, t_2, t_3 . For example,

$$\begin{aligned}
 \{f_1(t_1)f_2(t_2, t_3)\}_s &= \frac{1}{3} (f_1(t_1)f_2(t_2, t_3) \\
 &\quad + f_1(t_2)f_2(t_1, t_3) + f_1(t_3)f_2(t_1, t_2)).
 \end{aligned} \tag{17}$$

Poisson Process

Poisson process is a special case of a point process, whose increments $dN(t)$ defined on disjoint time intervals dt are independent. The nonhomogeneous Poisson process is completely characterized by its first-order product density:

$$f_1(t) = v(t), \tag{18}$$

(16) which is the *intensity* of the Poisson process.

Higher-order correlation functions are equal to zero:

$$g_n(t_1, \dots, t_n) = 0 \quad (19)$$

for $n > 1$.

Due to the independence of events, Eq. 2 becomes

$$f_n(t_1, \dots, t_n) = \prod_{i=1}^n v(t_i). \quad (20)$$

Substitution of Eq. 20 into Eq. 14 and into Eq. 15 yields, respectively (Snyder and Miller 1991),

$$\pi_n(t_1, \dots, t_n) = \prod_{i=1}^n v(t_i) \exp\left(-\int_0^t v(\tau) d\tau\right), \quad (21)$$

$$\Pr\{N(t) = n\} = \frac{1}{n!} \left(\int_0^t v(\tau) d\tau\right)^n \exp\left(-\int_0^t v(\tau) d\tau\right). \quad (22)$$

For a homogeneous Poisson process ($v(t) = v = \text{const.}$), the following expressions are obtained:

$$f_n(t_1, \dots, t_n) = v^n, \quad (23)$$

$$\pi_n(t_1, \dots, t_n) = v^n \exp(-vt), \quad (24)$$

$$\Pr\{N(t) = n\} = \frac{(vt)^n}{n!} \exp(-vt). \quad (25)$$

Renewal Processes

The renewal process can be defined as a sequence of random time points $t_1, t_2, \dots, t_n$ on the positive real line, such that

$$t_i - t_{i-1} = T_i, \quad i = 2, 3, \dots, t_1 = T_1, \quad (26)$$

where the time intervals $\{T_i, i = 2, 3, \dots\}$ between the successive points, called *inter-arrival times*, are the positive, independent, and identically distributed random variables. The point process is called an *ordinary renewal process* if the time T_1 measured from the origin to

the first event has the same distribution as other time intervals T_i . This means that the time origin is placed at the instant of 0th, or initial, event which is not counted. If T_1 has another distribution than other time intervals T_i , the point process is called a *general* or *delayed renewal process*. In that case the time origin is placed arbitrarily.

An ordinary renewal process can be defined equivalently as the sequence of positive, independent, and identically distributed random variables $\{T_i, i = 1, 2, \dots\}$.

Consider an interval $[0, t]$ of the time axis. An *ordinary renewal density* $h_o(t)$ (Cox 1962; Cox and Isham 1980) represents the probability that a random point (not necessarily the first) occurs in $[t, t + dt]$, given that a random point occurs at the origin. A *modified renewal density* $h_m(t)$ (Cox 1962) represents the probability that a random point (not necessarily the first) occurs in $[t, t + dt]$, with arbitrarily chosen time origin. A modified renewal density is the first-order product density of the renewal point process:

$$h_m(t)dt = \Pr\{dN(t) = 1\} = f_1(t)dt. \quad (27)$$

If this probability is irrespective of the position of the interval $[t, t + dt]$ on the time axis, the renewal process is stationary.

Product densities of higher degrees of a renewal process appear to split into a product form or get factorized (Srinivasan 1974):

$$\begin{aligned} f_n(t_1, \dots, t_n) dt_1 \dots dt_n &= E[dN(t_1) \dots dN(t_n)] \\ &= h_m(t_1) h_o(t_2 - t_1) h_o(t_3 - t_2) \dots \\ &\quad h_o(t_n - t_{n-1}) dt_1 dt_2 \dots dt_n, \\ &\quad (t_1 < t_2 < \dots < t_n), \end{aligned} \quad (28)$$

hence

$$f_n(t_1, \dots, t_n) = h_m(t_1) h_o(t_2 - t_1) h_o(t_3 - t_2) \dots h_o(t_n - t_{n-1}) \quad (t_1 < t_2 < \dots < t_n). \quad (29)$$

Let the probability density of the random variable X_1 be denoted as $g_1(t)$ and the probability density of each of the variables $\{T_i, i = 2, 3, \dots\}$ as $g(t)$. It can be shown that the renewal densities $h_m(t)$ and $h_o(t)$ satisfy, respectively,

inhomogeneous Volterra integral equations of the second kind, called *renewal equations*. These equations are derived (Srinivasan 1974) by considering the fact that the occurrence of the point in $[t, t + dt]$ is due to two mutually exclusive events: either it is the first point or it is the subsequent point. If it is the first point, the probability of its occurrence is just $g_1(t)dt$ (in the case of a delayed renewal process) or $g(t)dt$ (in the case of an ordinary renewal process). If it is the subsequent point, the preceding one has occurred at an arbitrary $t - u$, $u \in [0, t]$, u being the time interval between those two points. This leads to the following integral equations:

$$h_m(t) = g_1(t) + \int_0^t h_m(t-u)g(u)du, \quad (30)$$

$$h_o(t) = g(t) + \int_0^t h_o(t-u)g(u)du. \quad (31)$$

The renewal densities can be evaluated by taking the Laplace transforms of the equations Eqs. 30 and 31, which finally yields (Cox 1962; Cox and Isham 1980)

$$h_m(t) = \mathcal{L}^{-1} \left\{ \frac{g_1^*(s)}{1 - g^*(s)} \right\}, \quad (32)$$

$$h_o(t) = \mathcal{L}^{-1} \left\{ \frac{g^*(s)}{1 - g^*(s)} \right\}, \quad (33)$$

where $\mathcal{L}^{-1} \{ \dots \}$ denotes an inverse Laplace transform.

A class of the renewal processes which is important in applications are Erlang processes, where the time intervals between events have gamma (or Pearson type III) probability distribution with the density function

$$g(t) = v^k t^{k-1} \exp(-vt) / (k-1)!, \quad t > 0, \quad (34)$$

where $k = 1, 2, 3, \dots$. A homogeneous Poisson process is a special case of such a process, specified by letting $k = 1$, in which case the time intervals have the negative exponential distribution characterized by the density function:

$$g(t) = v \exp(-vt), \quad t > 0. \quad (35)$$

An important property of the gamma distribution with density function given by Eq. 34 is that it is the distribution of the sum of k independent random variables, each of whose distribution is negative exponential with parameter v . Hence, the events driven by an Erlang process with parameter k can be regarded as every k th event of the generating Poisson process with the mean arrival rate v .

The renewal densities of the Erlang process are

$$h_o(t) = \frac{v}{2} (1 - \exp(-2vt)), \quad k = 2, \quad (36)$$

$$h_o(t) = \frac{v}{3} \left[1 - \left(\sqrt{3} \sin \frac{\sqrt{3}}{2} vt + \cos \frac{\sqrt{3}}{2} vt \right) \exp \left(-\frac{3}{2} vt \right) \right], \quad k = 3, \quad (37)$$

$$h_o(t) = \frac{v}{4} [1 - 2 \sin(vt) \exp(-vt) - \exp(-2vt)], \quad k = 4. \quad (38)$$

It is interesting to note that, although the Erlang events are every k th Poisson events, the renewal densities, which are the mean arrival rates of Erlang events, only asymptotically (as $t \rightarrow \infty$) tend to v/k .

Random Pulse Trains Driven by Different Stochastic Point Processes

Random Trains of Overlapping Pulses: Filtered Stochastic Point Processes

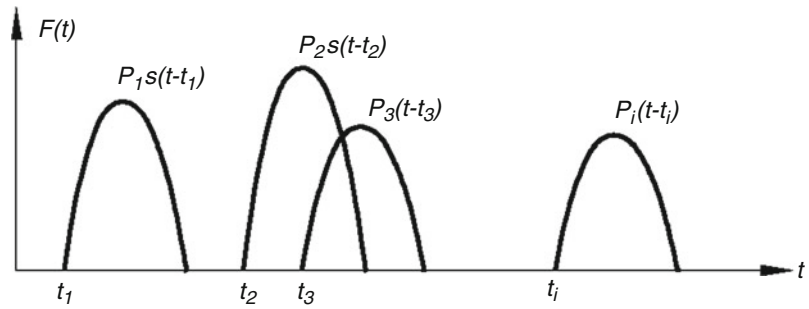
General Case

A filtered stochastic point process $\{X(t), t \in [0, t]\}$ is defined as

$$X(t) = \sum_{i=1}^{N(t)} s(t, t_i \mathbf{P}_i), \quad (39)$$

where $\{N(t), t \in [0, \infty]\}$ is a general counting process and $\mathbf{P}_i$ is a vector random variable

Non-Poisson Impulse Processes, Fig. 1 Train of overlapping pulses



attributed to a random point t_i . In a general case, the random variables which are the components of the vector $\mathbf{P}_i$ do not have to be independent, they can be correlated, and they can be characterized by different probability distributions, nor have the vector random variables $\mathbf{P}_i$ attributed to different points to be mutually independent or identically distributed. The only assumption made at present is that these random variables are statistically independent of the counting process $N(t)$. The nonrandom function $s(t, t_i, \cdot)$, called the *filter function*, represents the effect at the time t of an event occurring at the random instant t_i , the event being characterized by a vector random variable $\mathbf{P}_i$. For causality reasons, it is assumed that $s(t, \tau, \mathbf{P}_i) = 0$ for $\tau > t$. Hence, the process of Eq. 39 represents the cumulative effect of a train of point events occurring at random instants t_i belonging to the interval $[0, t]$, described by a general stochastic point process. The process $X(t)$, $t \in [0, \infty]$ defined by the formula Eq. 39 can be interpreted as a random train of general pulses $F(t)$ (Fig. 1), or signals, with origins at the random times t_i ; $s(t, t_i, \mathbf{P}_i)$ being the *pulse shape function*.

After the division of the interval $[0, t]$ onto disjoint, contiguous subintervals, $X(t)$ can be written down as the Riemann–Stieltjes sum. The limit, in the mean-square sense, of the sequence of such sums, is the mean-square Riemann–Stieltjes integral with respect to the counting process $N(t)$ or the stochastic integral:

$$X(t) = \int_0^t s(t, \tau, \mathbf{P}(\tau)) dN(\tau), \quad (40)$$

where $\mathbf{P}(\tau)$ is the vector random variable assigned to the point occurring in the interval $[\tau, \tau + d\tau]$.

The expected value of the process $X(t)$ is obtained just by averaging the expression Eq. 40, which yields (Iwankiewicz 1995)

$$\begin{aligned} E[X(t)] &= \int_0^t E[s(t, \tau, \mathbf{P}(\tau))] f_1(\tau) d\tau \\ &= \int_0^t \int_{\mathcal{P}_\tau} s(t, \tau, \mathbf{p}) f_{\mathbf{P}}(\mathbf{p}, \tau) f_1(\tau) d\mathbf{p} d\tau, \quad (41) \end{aligned}$$

where $f_{\mathbf{P}}(\mathbf{p}, \tau)$ is the joint probability density of the vector random variable $\mathbf{P}(\tau)$, which may be time variant, and $\mathcal{P}_\tau$ is the sample space of this vector random variable.

Subsequent moments of the process $X(t)$ are evaluated by averaging of the pertinent multifold integrals obtained based on Eq. 40.

For example, the second-order moment (the mean square) is obtained as

$$\begin{aligned} E[X^2(t)] &= \int_0^t \int_0^t E[s(t, \tau_1, \mathbf{P}(\tau_1)) s(t, \tau_2, \mathbf{P}(\tau_2))] \\ &\quad E[dN(\tau_1) dN(\tau_2)]. \quad (42) \end{aligned}$$

In order to evaluate this integral, the degeneracy property of the second-degree product density must be taken into account, which takes place within the integration domain, for $\tau_1 = \tau_2$. This yields

$$\begin{aligned}
 E[X^2(t)] &= \int_0^t E[s^2(t, \tau, \mathbf{P}_\tau)] f_1(\tau) d\tau \\
 &+ \iint_{00}^t E[s(t, \tau_1, \mathbf{P}_{\tau_1}) s(t, \tau_2, \mathbf{P}_{\tau_2})] f_2(\tau_1, \tau_2) d\tau_1 d\tau_2. \\
 E[X(t_1)X(t_2)] &= \int_0^{\min(t_1, t_2)} E[s(t_1, \tau, \mathbf{P}(\tau)) s(t_2, \tau, \mathbf{P}(\tau))] f_1(\tau) d\tau \\
 &+ \iint_{00}^{t_1 t_2} E[s(t_1, \tau_1, \mathbf{P}(\tau_1)) s(t_2, \tau_2, \mathbf{P}(\tau_2))] f_2(\tau_1, \tau_2) d\tau_1 d\tau_2.
 \end{aligned}
 \tag{43}$$

Likewise the correlation function of the process $X(t)$ is obtained as

The general expression for the n th order moment is

$$E[X^n(t)] = \underbrace{\int_0^t \cdots \int_0^t}_{n\text{-fold}} E\left[\prod_{k=1}^n s(t, \tau_k, \mathbf{P}(\tau_k))\right] E\left[\prod_{k=1}^n dN(\tau_k)\right]. \tag{45}$$

Of course, as $\tau_1 \neq \tau_2 \neq \dots \neq \tau_n$, then

$$E\left[\prod_{k=1}^n dN(\tau_k)\right] = f_n(\tau_1, \dots, \tau_n) d\tau_1 \cdots d\tau_n. \tag{46}$$

In the multidimensional integration domain, any possible equations of the arguments τ_k take place. Therefore, in order to evaluate the integral

Eq. 45, all possible degeneracies of n th degree product density $f_n(\tau_1, \dots, \tau_n)$ must be taken into account. Moreover, in general case, the integration is performed with respect to the joint probability density of n vector random variables $\mathbf{P}_k, k = 1, 2, \dots, n$.

In particular the expression for the third-order moment is obtained as

$$\begin{aligned}
 E[X^3(t)] &= \int_0^t E[s^3(t, \tau, \mathbf{P}_\tau)] f_1(\tau) d\tau \\
 &+ 3 \iint_{00}^t E[s^2(t, \tau_1, \mathbf{P}(\tau_1)) s(t, \tau_2, \mathbf{P}(\tau_2))] f_2(\tau_1, \tau_2) d\tau_1 d\tau_2 \\
 &+ \iiint_{000}^t E[s(t, \tau_1, \mathbf{P}(\tau_1)) s(t, \tau_2, \mathbf{P}(\tau_2)) s(t, \tau_3, \mathbf{P}(\tau_3))] f_3(\tau_1, \tau_2, \tau_3) d\tau_1 d\tau_2 d\tau_3.
 \end{aligned}
 \tag{47}$$

However, in general, the evaluation of the above integrals becomes cumbersome, especially in the case of higher-order moments. Then, in the case of a Poisson process, it is much easier to handle the cumulants, which can be obtained directly from the log-characteristic function, called also a cumulant-generating function. In the case of a filtered renewal process, the

recursive expressions for the moments can be obtained from the integral equations governing the characteristic function.

Filtered Renewal Process

Consider a filtered process $\{X(t), t \in [0, \infty]\}$, driven by an ordinary renewal counting process $\{X(t), t \in [0, \infty]\}$, in the form of

$$X(t) = \sum_{j=1}^{N(t)} s(t - t_j, P_j), \quad (48)$$

where the filter function is assumed to be causal, i.e., $s(\tau, P_j) = 0$ for $\tau < 0$. The random variables P_j are assumed to be independent and to have identical probability distributions characterized by the common density function $f_P(p)$. The probability distributions of the inter-arrival times are characterized by the probability distribution function $G(t)$ and by the probability density function $g(t)$.

The characteristic function $\Phi_X(\theta, t)$ of the filtered renewal process $X(t)$ defined by Eq. 48 appears to satisfy the inhomogeneous Volterra integral equation of the second kind. The following derivation is due to Takacs (1956).

The filtered process $X(t)$ given by Eq. 48 may be regarded as a sum of the first pulse occurring after first inter-arrival time T_1 and the filtered process with the origin shifted by T_1 , i.e., $X(t - T_1)$; thus,

$$X(t) = s(t - T_1, P_1) + X(t - T_1). \quad (49)$$

Suppose that $T_1 = \tau$. In view of independence of random variables P_j , the conditional characteristic function, given that the first pulse occurs at the time $T_1 = \tau$, is expressed as

$$\begin{aligned} \Phi_X(\theta, t | T_1 = \tau) &= E[\exp(i\theta X(t) | T_1 = \tau)] \\ &= E[\exp(is(t - \tau, P_1)) \exp(i\theta X(t - \tau))] \\ &= \Gamma(\theta, t - \tau) \Phi_X(\theta, t - \tau), \end{aligned} \quad (50)$$

where

$$\begin{aligned} \Gamma(\theta, t - \tau) &= E[\exp\{i\theta s(t - \tau, P)\}] \\ &= \int_{-\infty}^{\infty} \exp[i\theta s(t - \tau, p)] f_P(p) dp \end{aligned} \quad (51)$$

is the characteristic function of a single general pulse $s(t - \tau, P)$.

By unconditioning one obtains

$$\Phi_X(\theta, t) = \int_0^t \Phi_X(\theta, t - \tau) \Gamma(\theta, t - \tau) g(\tau) d\tau + C, \quad (52)$$

where the integration constant C is evaluated from the obvious condition for $\theta = 0$

$$\Phi_X(0, t) = 1 = \int_0^t g(\tau) d\tau + C, \quad (53)$$

thus $C = 1 - G(t)$.

Hence, the final result is

$$\Phi_X(\theta, t) = \int_0^t \Phi_X(\theta, t - \tau) \Gamma(\theta, t - \tau) g(\tau) d\tau + 1 - G(t). \quad (54)$$

After differentiating the Eq. 54 r times with respect to θ and after substituting $\theta = 0$, the following integral equation governing the moment $E[X^r(t)]$ is arrived at

$$E[X^r(t)] = \sum_{i=0}^r \binom{r}{i} \int_0^t \phi_{r-i}(t - \tau) E[X^i(t - \tau)] g(\tau) d\tau, \quad (55)$$

where

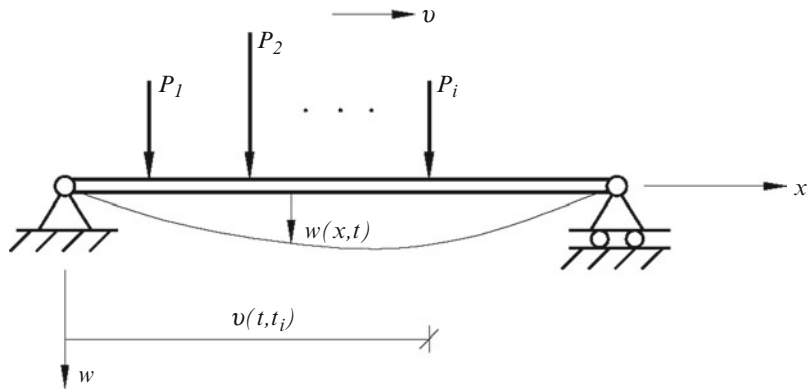
$$\begin{aligned} \phi_k(t - \tau) &= E[s^k(t - \tau, P)] \\ &= \int_{-\infty}^{\infty} s^k(t - \tau, p) f_P(p) dp. \end{aligned} \quad (56)$$

Taking Laplace transforms of the both sides of Eq. 55, solving the obtained algebraic equation for the transform of the r th order moment, and taking the inverse transform, one obtains the following recursive expression (Takacs 1956):

$$E[X^r(t)] = \sum_{i=0}^{r-1} \binom{r}{i} \int_0^t \phi_{r-i}(t - \tau) E[X^i(t - \tau)] h_o(\tau) d\tau, \quad (57)$$

where $h_o(\tau)$ is the ordinary renewal density.

Non-Poisson Impulse Processes, Fig. 2 Beam under a random train of moving loads



The specific formulae for the mean value function $E[X(t)]$ and the mean-square function $E[X^2(t)]$ are obtained as

$$E[X(t)] = \int_0^t \phi_1(t - \tau) h_o(\tau) d\tau, \quad (58)$$

$$E[X^2(t)] = \int_0^t \phi_2(t - \tau) h_o(\tau) d\tau + 2 \int_0^t \int_0^t \phi_1(t - \tau) \phi_1(t - u) h_o(u - \tau) h_o(\tau) du d\tau, \quad (59)$$

where

$$\phi_1(t - \tau) = E[s(t - \tau, P)], \quad (60)$$

$$\phi_2(t - \tau) = E[s^2(t - \tau, P)]. \quad (61)$$

The same result for $E[X^2(t)]$ as Eq. 59 is obtained if the expression Eq. 29 for the second-order product density for the ordinary renewal process is inserted into the general expression Eq. 43. In the integration domain $\tau_1 \in (0, t)$, $\tau_2 \in (0, t)$, it must be assumed

$$f_2(\tau_1, \tau_2) = \begin{cases} h_o(\tau_1) h_o(\tau_2 - \tau_1), & \tau_2 > \tau_1, \\ h_o(\tau_1) h_o(\tau_2 - \tau_1), & \tau_2 > \tau_1. \end{cases} \quad (62)$$

More results for the moments of the filtered renewal process (e.g., the response of a linear

dynamic system to a renewal pulse train) may be found in Iwankiewicz (1995).

Example Problem: Dynamic Response of a Bridge (Beam) to a Random Train of Moving Loads

Consider the well known in structural engineering problem of the dynamic response of a highway bridge to the moving load due to the vehicular traffic. It is known that if the bridge has a long span, the coupling of the motion of the bridge and of the vehicle as well as the inertia of the vehicle may be neglected and the vehicles may be adequately idealized by point forces. The times of occurrence of vehicles at the bridge are random, and in the simplest case, all the vehicles may be assumed to travel with the same, constant velocity. The bridge, idealized as a beam, is then subjected to a random train of moving forces with random magnitudes P_j and at a given time instant t randomly located at the beam, as shown in Fig. 2.

The equation governing the transverse motion of the beam has the form

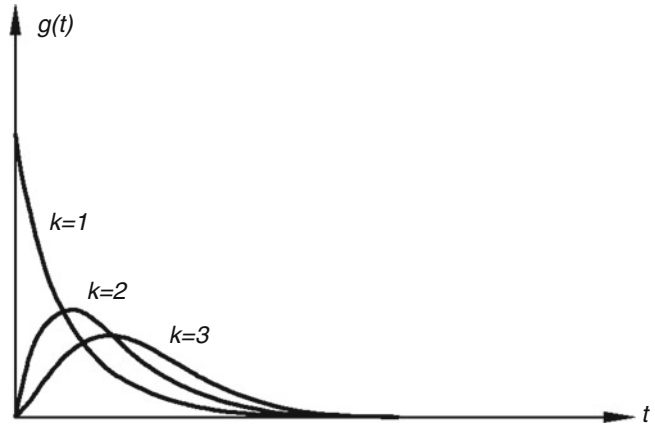
$$EI \frac{\partial^4 w(x, t)}{\partial x^4} + c \frac{\partial w(x, t)}{\partial t} + \mu \frac{\partial^2 w(x, t)}{\partial t^2} = \sum_{j=1}^{N(t)} P_j \delta(x - v(t - t_j)). \quad (63)$$

Application of the normal mode approach

$$w(x, t) = \sum_{i=1}^n q_i(t) \phi_i(x), \quad (64)$$

Non-Poisson Impulse Processes,

Fig. 3 Different gamma probability density plots



where $\varphi_i(x)$ are the normal modes, yields the equations

$$\ddot{q}_i(t) + 2\zeta_i\omega_i\dot{q}_i(t) + \omega_i^2q_i(t) = \beta_i \sum_{j=1}^{\mathcal{N}(t)} P_j \phi_i(v(t-t_j)), \quad i = 1, 2, \dots, n \quad (65)$$

where $\phi_i(v(t-t_j)) = s(t-t_j)$ plays the role of the pulse shape.

Thus, the problem is converted into the problem of the random train of general pulses. The underlying counting process $\mathcal{N}(t)$ may be any counting process. It is known from the highway traffic engineering that the inter-arrival times between the vehicles are adequately idealized as positive random variables with a unimodal probability density function. If the inter-arrival times are independent and identically distributed, the underlying counting process is a renewal process.

For example, if the renewal process is an Erlang process, the probability density function of the inter-arrival times is

$$g(t) = \frac{v^k t^{k-1}}{(k-1)!} \exp(-vt), \quad t > 0. \quad (66)$$

Some example probability density plots are shown in Fig. 3.

Example Problem: Dynamic Response of Linear System to a Random Train of Impulses

A random train of impulses, or an impulse process excitation, $F(t)$ is shown in Fig. 4.

Vibrations of a linear oscillator under a random train of impulses are governed by

$$\ddot{Y}(t) + 2\zeta\omega\dot{Y}(t) + \omega^2Y(t) = F(t) = \sum_{i=1}^{\mathcal{N}(t)} P_i \delta(t-t_i), \quad (67)$$

where $\mathcal{N}(t)$ is a stochastic point (random counting) process, giving the random number of time points t_i in the time interval $[0, t]$, i.e., excluding the one that may occur at t . The impulse magnitudes P_i are independent random variables, identically distributed as a random variable P . The variables P_i are also statistically independent of the random times t_i or of the counting process $\mathcal{N}(t)$. The counting process may be, e.g., a Poisson or a renewal process.

From the impulse-momentum principle, it follows that

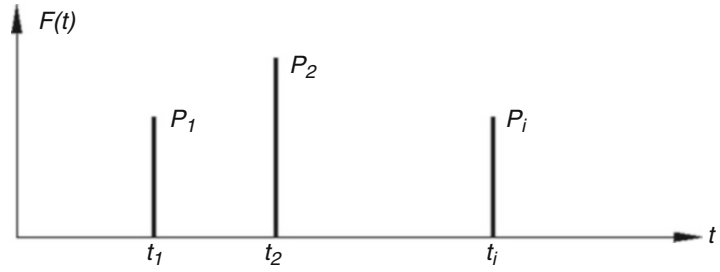
$$\dot{Y}(t_i)_+ = \dot{Y}(t_i)_- + P_i.$$

Hence, the velocity response process $\dot{Y}(t)$ changes by jumps; it is piecewise continuous.

Consequently the displacement response process $Y(t)$ is continuous, but it is only piecewise continuously differentiable.

Non-Poisson Impulse Processes,

Fig. 4 Random train of impulses



The jump change in the velocity may occur at any point; hence, the usual rules of calculus do not apply.

The usual notation of the differential equation of motion is **not mathematically meaningful**.

Stochastic counterparts of the usual differential equations are the **stochastic differential equations**:

$$dY(t) = \dot{Y}(t)dt, d\dot{Y}(t) = -2\zeta\omega\dot{Y}(t)dt - \omega^2 Y(t)dt + P(t)d\mathcal{N}(t), \quad (68)$$

where $P(t)$ is the magnitude of the impulse which occurs in the time interval $[t, t + dt)$.

Nevertheless, the response of **linear dynamical systems** may be directly analyzed in **time domain**. The explicit expression for the response to a random train of impulses, based on the **linear superposition principle**, is obtained as

$$Y(t) = \sum_{i=1}^{\mathcal{N}(t)} P_i h(t - t_i) = \int_0^t P(\tau) h(t - \tau) d\mathcal{N}(\tau), \quad (69)$$

where $h(t) = \frac{1}{\omega_d} \exp(-\zeta\omega t) \sin \omega_d t$ is the **impulse response function** and $\omega_d = \omega \sqrt{1 - \zeta^2}$. The integral with respect to the increments of the stochastic point process is the stochastic counterpart of the usual Duhamel convolution integral.

Statistical moments of the response process may be evaluated by averaging of the pertinent multifold integrals Eq. 45.

Response of Dynamic Systems to Non-Poisson Impulse Processes. State-Space Formulation

Itô's differential rule for nondiffusive, Poisson-driven Markov processes

The stochastic equations of motion of the dynamic system subjected to a Poisson impulse process excitation are

$$d\mathbf{Y}(t) = c(\mathbf{Y}(t), t)dt + b(\mathbf{Y}(t), t)P(t)d\mathcal{N}(t), \mathbf{Y}(0) = y_0. \quad (70)$$

If the excitation process is statistically independent of the initial conditions, the vector process $\mathbf{Y}(t) = [Y_1(t), Y_2(t), \dots, Y_n(t)]^T$ is a **nondiffusive, Poisson-driven, Markov process**.

If the function $V(\mathbf{Y}(t), t)$ is bounded for t and $\mathbf{Y}(t)$ finite and is once continuously differentiable with respect to all its arguments, the differential of a compound function $V(t, \mathbf{Y}(t))$ of the state variables $\mathbf{Y}(t)$ is expressed by the following **generalized Itô's differential rule** (Snyder and Miller 1991; Iwankiewicz and Nielsen 1999):

$$\begin{aligned} dV(t, \mathbf{Y}(t)) &= \frac{\partial V(t, \mathbf{Y}(t))}{\partial t} dt \\ &+ \sum_{i=1}^n \frac{\partial V(t, \mathbf{Y}(t))}{\partial Y_i} c_i(\mathbf{Y}(t), t) dt \\ &+ [V(t, \mathbf{Y}(t) + b(\mathbf{Y}(t), t, P(t))) - V(t, \mathbf{Y}(t))] d\mathcal{N}(t). \end{aligned} \quad (71)$$

Differential equations governing the response joint statistical moments $\mu_{ij}(t) = E[Y_i(t)Y_j(t)]$, $\mu_{ijk}(t) = E[Y_i(t)Y_j(t)Y_k(t)]$, $\mu_{ijkl}(t) = E[Y_i(t)Y_j(t)Y_k(t)Y_l(t)]$,

etc. are derived from this rule by assuming the function $V(t, \mathbf{Y}(t))$ in form of different products of the state variables, for example, $V(t, \mathbf{Y}(t)) = Y_i(t)Y_j(t)$, $V(t, \mathbf{Y}(t)) = Y_i(t)Y_j(t)Y_k(t)$, and $V(t, \mathbf{Y}(t)) = Y_i(t)Y_j(t)Y_k(t)Y_l(t)$.

Next, the expectation of both sides of the above differential rule must be taken and

$$E[dV(t, \mathbf{Y}(t))] = dE[V(t, \mathbf{Y}(t))].$$

The effective averaging is possible if the driving process is a Poisson process, because then the increment $dN(t) = N(t+dt) - N(t)$ is statistically independent of $V(t, \mathbf{Y}(t))$ and

$$\begin{aligned} E[V(t, \mathbf{Y}(t))dN(t)] &= E[V(t, \mathbf{Y}(t))]E[dN(t)] \\ &= E[V(t, \mathbf{Y}(t))]v(t)dt. \end{aligned}$$

For a general Poisson-driven pulse problem, the equations for the mean values, the second-, third-, and fourth-order joint central moments of the response, are obtained as

$$\dot{\mu}_i(t) = E[c_i(\mathbf{Y}(t), t)] + v(t)E[P]E[b_i(\mathbf{Y}(t), t)] \quad (72)$$

$$\begin{aligned} \dot{\kappa}_{ij}(t) &= 2 \left\{ E \left[Y_i^0 \left(c_j^0(\mathbf{Y}^0(t), t) + v(t)b_j(\mathbf{Y}(t), t)P \right) \right] \right\}_s \\ &\quad + v(t)E[P^2]E[b_i(\mathbf{Y}(t), t)b_j(\mathbf{Y}(t), t)] \end{aligned} \quad (73)$$

$$\begin{aligned} \dot{\kappa}_{ijk}(t) &= 3 \left\{ E \left[Y_i^0 Y_j^0 \left(c_k^0(\mathbf{Y}^0(t), t) + v(t)b_k(\mathbf{Y}(t), t)P \right) \right] \right\}_s \\ &\quad + 3v(t)E[P^2] \left\{ E \left[Y_i^0 b_j(\mathbf{Y}(t), t)b_k(\mathbf{Y}(t), t) \right] \right\}_s \\ &\quad + v(t)E[P^3]E[b_i(\mathbf{Y}(t), t)b_j(\mathbf{Y}(t), t)b_k(\mathbf{Y}(t), t)] \end{aligned} \quad (74)$$

$$\begin{aligned} \dot{\kappa}_{ijkl}(t) &= 4 \left\{ E \left[Y_i^0 Y_j^0 Y_k^0 \left(c_l^0(\mathbf{Y}^0(t), t) + v(t)b_l(\mathbf{Y}(t), t)P \right) \right] \right\}_s \\ &\quad + 6v(t)E[P^2] \left\{ E \left[Y_i^0 Y_j^0 b_k(\mathbf{Y}(t), t)b_l(\mathbf{Y}(t), t) \right] \right\}_s \\ &\quad + 4v(t)E[P^3] \left\{ E \left[Y_i^0 b_j(\mathbf{Y}(t), t)b_k(\mathbf{Y}(t), t)b_l(\mathbf{Y}(t), t) \right] \right\}_s \\ &\quad + v(t)E[P^4]E[b_i(\mathbf{Y}(t), t)b_j(\mathbf{Y}(t), t)b_k(\mathbf{Y}(t), t)b_l(\mathbf{Y}(t), t)] \end{aligned} \quad (75)$$

N

where $Y_i^0(t) = Y_i(t) - \mu_i(t)$ and $c_j^0(\mathbf{Y}^0(t), t) = c_j(\mathbf{Y}(t), t) - E[c_j(\mathbf{Y}(t), t)]$ denote the components of the zero-mean (centralized) state vector and drift vector, respectively; $E[P^r]$ denotes the r th moment of the random variable P , i.e., $E[P^r] = \int_{\mathbf{P}} p^r f_P(p) dp$; and $\{\dots\}_s$ denotes the

Stratonovich symmetrizing operation, e.g.,

$$\{Y_i Y_j c_k\}_s = \frac{1}{3} (Y_i Y_j c_k + Y_i Y_k c_j + Y_j Y_k c_i). \quad (76)$$

Equations for response moments of a linear system form always a closed set and can be directly solved numerically. Equations for response moments of a nonlinear system with polynomial nonlinearity form an infinite hierarchy.

Then special closure approximations (truncation procedures) must be used. For other types of nonlinearity, the equations involve unknown expectations of nonlinear functions of state variables, and some tentative forms of the joint probability density must be used (Iwankiewicz et al. 1990).

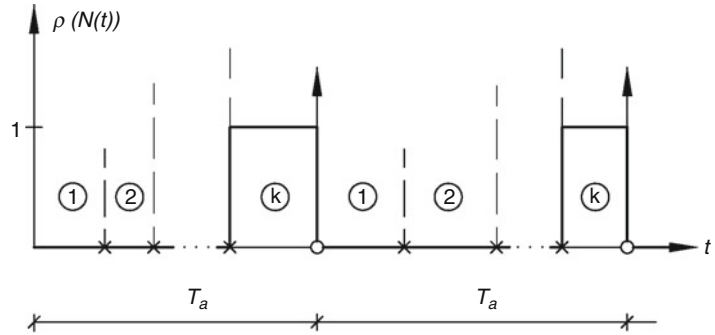
Conversion of Non-Markov Impulse Problems to Markov Ones

Non-Markov Nature of the State Vector of the Dynamic System

A renewal impulse process is

$$F(t) = \sum_{i, R=1}^{R(t)} P_{i, R} \delta(t - t_{i, R}) \quad (77)$$

Non-Poisson Impulse Processes, Fig. 5 Train of Erlang impulses. $\times$: Poisson-driven points, $\circ$: Erlang-driven points



where the occurrence times $t_{i,R}$ of impulses are driven by a renewal process. $R(t)$ and $P_{i,R}$ are statistically independent and identically distributed random magnitudes of the impulses.

The stochastic equations of motion in a general case of a nonlinear oscillator and a parametric excitation are

$$\begin{aligned} dY(t) &= \dot{Y}(t)dt, d\dot{Y}(t) \\ &= f(Y(t), \dot{Y}(t))dt + b(Y(t), \dot{Y}(t))P(t)dR(t), \end{aligned} \quad (78)$$

where $f(Y(t), \dot{Y}(t))$ is a nonlinear function of the instantaneous values of $Y(t)$ and $\dot{Y}(t)$ which represents all restoring force together with damping terms of the equation of motion and $b(Y(t), \dot{Y}(t))$ is the parametric excitation term. As the increments $dR(t)$ of the renewal counting process $R(t)$ are not independent, the state vector $\mathbf{Y}(t) = [Y(t), \dot{Y}(t)]^T$ is **not a Markov process**.

Augmentation of a State Vector by Additional Variables

A basic idea of conversion of a non-Markov impulse problem to a Markov one is to replace the original train of impulses by an equivalent one for which the response process becomes a Markov process.

If the impulse process is driven by an Erlang renewal process, the following exact replacement is valid with probability 1 (Iwankiewicz and Nielsen 1999):

$$\sum_{i,R=1}^{R(t)} P_{i,R} \delta(t - t_{i,R}) = \sum_{i=1}^{N(t)} \rho(N(t_i)) P_i \delta(t - t_i). \quad (79)$$

Based on the fact that the events of the Erlang renewal process are every second, or every third, or every fourth, etc. Poisson events, a zero-memory transformation $\rho(N(t))$ of the homogeneous Poisson process $N(t)$ is introduced which assumes values $\rho(N(t_i)) = 1$ for every k th Poisson event and $\rho(N(t_i)) = 0$ for all other Poisson events. The sample paths of $\rho(N(t))$ are assumed to be left continuous with right limits. The arrival times t_i are driven by a homogeneous Poisson process $N(t)$ with a mean arrival rate ν , and the impulse magnitudes P_i are compounded with $\rho(N(t_i))$. Thus, the actual impulse process is obtained by selecting, with the aid of an auxiliary stochastic variable $p(N(t))$, every k th impulse from the train driven by a Poisson process $N(t)$. The replacement Eq. 79 is illustrated in Fig. 5, where the dashed-line spikes represent these of the impulses driven by $N(t)$, whose magnitudes are multiplied by $\rho(N(t_i)) = 0$, hence are excluded, and the solid-line spikes represent the remaining $N(t)$ -driven impulses, whose magnitudes are multiplied by $\rho(N(t_i)) = 1$; hence, these are the impulses driven by the underlying Erlang renewal process $R(t)$.

Consequently the increments of the Erlang renewal process can be expressed in terms of the Poisson counting process:

$$dR(t) = \rho(N(t))dN(t). \quad (80)$$

The transformation $\rho(N(t))$ is expressed in terms of the auxiliary variables $C_j(t)$ and $S_j(t)$ (Iwankiewicz and Nielsen 1999):

$$\rho(N(t)) = \rho(C_j(N(t)), S_j(N(t))) \quad (81)$$

where

$$\begin{aligned} C_j(N(t)) &= \cos\left(\frac{j\pi N(t)}{k}\right), \quad S_j(N(t)) \\ &= \sin\left(\frac{j\pi N(t)}{k}\right) \end{aligned} \quad (82)$$

hence

$$dR(t) = \rho(C_j(N(t)), S_j(N(t))) dN(t). \quad (83)$$

For example, if $k = 2$, $k = 3$, and $k = 4$, the required transformations of the Poisson counting process $N(t)$, such that $\rho(N(t)) = 1$ for every second, third, and fourth Poisson event and $\rho(N(t)) = 0$ for all other Poisson events, are, respectively,

$$\begin{aligned} \rho(N(t)) &= \frac{1}{2} (1 - \cos(\pi N(t))) \\ &= \frac{1}{2} (1 - (-1)^{N(t)}), \end{aligned} \quad (84)$$

$$\rho(N(t)) = \frac{1}{3} \left(1 - \sqrt{3} \sin\left(\frac{2}{3}\pi N(t)\right) - \cos\left(\frac{2}{3}\pi N(t)\right) \right), \quad (85)$$

$$\rho(N(t)) = \frac{1}{4} \left(1 - 2 \sin\left(\frac{1}{2}\pi N(t)\right) - \cos(\pi N(t)) \right). \quad (86)$$

For all auxiliary variables $C_j(t)$ and $S_j(t)$, the stochastic differential equations are written down, which are all driven by the Poisson process $N(t)$.

It should be noted that as the increment $dN(t)$ is independent of $N(t)$, then $dN(t)$ is independent of $\rho(N(t))$.

The original state vector $\mathbf{Y}(t) = [Y(t), \dot{Y}(t)]^T$ is augmented by the auxiliary variables $C_j(t)$ and $S_j(t)$ to make up a vector $\mathbf{Z}(t)$:

$$\begin{aligned} d\mathbf{Z}(t) &= \mathbf{c}(\mathbf{Z}(t), t) dt \\ &+ \mathbf{b}(\mathbf{Z}(t), t,) P(t) dN(t), \quad \mathbf{Z}(0) = \mathbf{z}_0. \end{aligned} \quad (87)$$

The augmented state vector $\mathbf{Z}(t)$ is driven by a Poisson process; hence, it is a **nondiffusive (Poisson-driven) Markov process**. Equations for

moments may be derived from the Itô's differential rule for Poisson-driven Markov processes.

Now the renewal, impulse process excitation is considered where the inter-arrival times T_a are the sum of two independent, negative exponential distributed variables T_r and T_d , with probability density functions given, respectively, by (for $t > 0$)

$$\begin{aligned} g_{T_r}(t) &= v \exp(-vt), \quad g_{T_d}(t) \\ &= \mu \exp(-\mu t). \end{aligned} \quad (88)$$

The following replacement holds with probability 1 for the renewal driven train of impulses (Iwankiewicz 2003):

$$\sum_{i,R=1}^{R(t)} P_{i,R} \delta(t - t_{i,R}) = \sum_{i=1}^{N_\mu(t)} Z(t_i) P_i \delta(t - t_i), \quad (89)$$

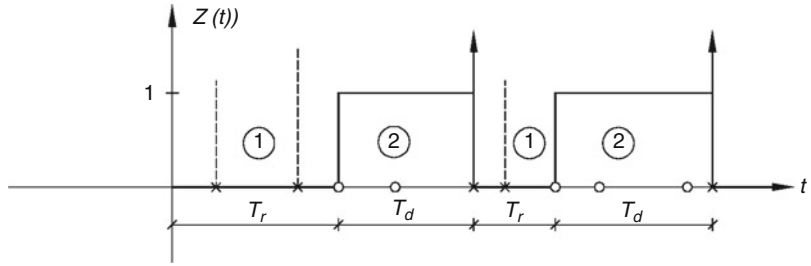
where the arrival times $t_{i,R}$ are driven by the underlying renewal process $R(t)$, the arrival times t_i are driven by a homogeneous Poisson process $N_\mu(t)$ with a mean arrival rate μ , and $Z(t_i)$ is a value at t_{i-} of the zero-one auxiliary stochastic jump process $Z(t)$ governed by (Iwankiewicz 2002)

$$dZ(t) = (1 - Z) dN_v(t) - Z dN_\mu(t). \quad (90)$$

The processes $N_v(t)$ and $N_\mu(t)$ are independent Poisson processes, with parameters v and μ , respectively, and $dN_v(t)$ and $dN_\mu(t)$ are the increments during the infinitesimal time interval $[t, t + dt)$. The process $Z(t)$ is zero-one valued, and $dZ(t) = Z(t + dt) - Z(t)$ denotes the jump increment, which may be the jump from 0 to 1, i.e., $dZ(t) = 1$, when $Z(t) = Z(t_-) = 0$ and $dN_v(t) = 1$, or the jump from 1 to 0, i.e., $dZ(t) = -1$, when $Z(t) = Z(t_-) = 1$ and $dN_\mu(t) = 1$. The sample paths of $N(t)$ and of $Z(t)$ are assumed to be left continuous with right limits. Thus, the actual impulse process is obtained by selecting, with the aid of an auxiliary stochastic variable $Z(t)$, some impulses from the train driven by a Poisson process $N_\mu(t)$. The replacement Eq. 89 is illustrated in Fig. 6, where the dashed-line spikes represent these of

Non-Poisson Impulse Processes, Fig. 6

Train of impulses expressed in terms of the jump process $Z(t)$ driven by two Poisson processes, $\times$: N_μ -driven points, $\circ$: N_ν -driven points



the impulses driven by $N_\mu(t)$, whose magnitudes are multiplied by $Z(t_i) = 0$, hence are excluded, and the solid-line spikes represent the remaining $N_\mu(t)$ -driven impulses, whose magnitudes are multiplied by $Z(t_i) = Z(t_{i-}) = 1$; hence, these are the impulses driven by the underlying counting process $R(t)$. The inter-arrival time between the solid-line spikes (actual impulses) is exactly equal to the inter-arrival time $T_a = T_r + T_d$, as defined above. The replacement implies the equivalence of the increments $dR(t) = Z(t) dN_\mu(t)$, which holds with probability 1.

The augmented state vector $\mathbf{Z}(t) = [Y(t), \dot{Y}(t), Z(t)]^T = [Z_1(t), Z_2(t), Z_3(t)]^T$ is governed by stochastic equations:

$$\begin{aligned} dZ_1 &= Z_2 dt, \\ dZ_2 &= f(Z_1, Z_2) dt + b(Z_1, Z_2) P(t) Z_3 dN_\mu(t), \\ dZ_3 &= (1 - Z_3) dN_\nu(t) - Z_3 dN_\mu(t). \end{aligned} \quad (91)$$

The state vector $\mathbf{Z}(t) = [Z_1(t), Z_2(t), Z_3(t)]^T$ is driven by two independent Poisson processes, and hence, it is a **nondiffusive Markov process**. Equations for moments may be derived from the generalized Itô's differential rule for Poisson-driven Markov processes.

Augmentation of a State Vector by Auxiliary Markov States

General Integro-differential Equations for the Joint Probability Density of the State Vector As the explicitly introduced, pure-jump stochastic processes $\rho(N(t))$ and $Z(t)$ are Poisson driven, they are characterized by negative exponential distributed phases, hence by a chain of

Markov states. Consequently the original state variables and the states of the auxiliary pure-jump stochastic process are jointly Markovian. The jumps have to be defined in such a way that the actual impulse (i.e., the jump in the velocity response $Z_2(t)$) only occurs if there is a jump between some particular Markov states.

The problem is characterized by the set of joint probability density – discrete distribution functions $q_j(z_1, z_2, t)$ of the response state variables – the displacement $Z_1(t)$ and the velocity $Z_2(t)$, and the m states $S(t)$ of a pertinent Markov chain, defined as

$$\begin{aligned} q_j(z_1, z_2, t) dz_1 dz_2 &= \Pr \{ Z_1(t) \in (z_1, z_1 + dz_1) \\ &\wedge Z_2(t) \in (z_2, z_2 + dz_2) \wedge S(t) = j \}, \end{aligned} \quad (92)$$

where $j = 1, 2, \dots, m$. The fundamental equation for such a continuous-jump Markov process is the general forward integrodifferential Chapman–Kolmogorov equation (Gardiner 1985; Iwankiewicz and Nielsen 1999):

$$\begin{aligned} \frac{\partial}{\partial t} q_j(\mathbf{z}, t) &= - \sum_{r=1}^2 \frac{\partial}{\partial z_r} [c_r(\mathbf{z}, t) q_j(\mathbf{z}, t)] \\ &+ \sum_{i=1}^m \int_{-\infty}^{\infty} [J_{\{z\}}(\mathbf{z}, j | \mathbf{x}, i, t) q_i(\mathbf{x}, t) \\ &- J_{\{z\}}(\mathbf{x}, i | \mathbf{z}, j, t) q_j(\mathbf{z}, t)] d\mathbf{x} \end{aligned} \quad (93)$$

where in the present problem $\mathbf{q}(\mathbf{z}, t) = [q_1(\mathbf{z}, t), q_2(\mathbf{z}, t), \dots, q_m(\mathbf{z}, t)]$, $c_r(\mathbf{z}, t)$ are the drift terms of the equation of motion written down in the state space form, i.e., $c_1(\mathbf{z}, t) = z_2$, $c_2(\mathbf{z}, t) = f(\mathbf{z}, t)$, $j = 1, 2, \dots, m$ and $J_{\{z\}}(z_1, z_2, j | x_1, x_2, i, t) = J_{\{z\}}(\mathbf{z}, j | \mathbf{x}, i, t)$ is the **jump probability intensity function** defined as

$$J_{\{Z\}}(z, j | x, i, t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr\{Z_1(t + \Delta t) = z_1, Z_2(t + \Delta t) = z_2, S(t + \Delta t) = j | Z_1(t) = x_1, Z_2(t) = x_2, S(t) = i\}}{\Delta t} \quad (94)$$

which is determined from the pertinent chain of Markov states as follows. When $i = j$, the Markov chain remains in the same state, and no actual impulse occurs; hence, both the displacement and the velocity state variables are continuous. The nonzero jump probability intensity functions are only defined for $i \neq j$, such that there is a transition in a Markov chain (jump in the auxiliary process). Only some of those transitions are associated with the occurrence of the actual impulse, or the jump in the velocity process $Z_2(t)$. Hence, if there is a transition from $S(t) = i$ to $S(t + \Delta t) = j$, but no actual impulse occurs (no jump in the velocity process), the jump probability intensity function is

$$J_{\{Z\}}(z_1, z_2, j | x_1, x_2, i, t) = \pi(j|i) \delta(z_1 - x_1) \delta(z_2 - x_2), \quad (95)$$

where

$$\pi(j|i) = \frac{\Pr\{S(t + \Delta t) = j | S(t) = i\}}{\Delta t} \quad (96)$$

is determined from the pertinent chain of Markov states. If the transition from $S(t) = i$ to $S(t + \Delta t) = j$ is associated with the actual impulse (the jump in the velocity process $Z_2(t)$), the jump probability intensity function is expressed as

$$J_{\{Z\}}(z_1, z_2, j | x_1, x_2, i, t) = \pi(j|i) \delta(z_1 - x_1) \int_{\mathbf{P}} \delta(z_2 - (x_2 + b(z_1, z_2)p)) f_P(p) dp \quad (97)$$

where $\mathbf{P}$ denotes the sample space and $f_P(p)$ the probability density function of the random impulse magnitude P .

Summation of the joint probability density-distribution functions $q_Z(z_1, z_2, j, t)$ over all Markov states yields the joint probability density of the original state variables:

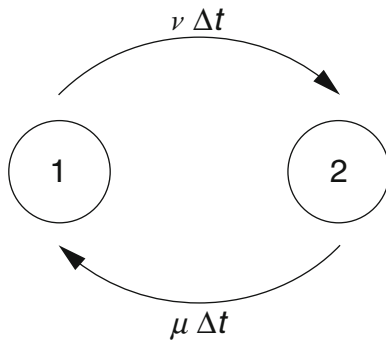
$$q_Z(z_1, z_2, t) = \sum_{j=1}^m q_j(z_1, z_2, t). \quad (98)$$

Detailed Integrodifferential Equations for the Joint Probability Density of the State Vector for Renewal Impulse Process Driven by Two Independent Poisson Processes The jump process $Z(t)$ driven by two independent Poisson processes Eq. 90 is tantamount to a two-state Markov chain $S(t)$, such that $S(t) = 1$ when $Z(t) = 0$ and $S(t) = 2$ when $Z(t) = 1$ (Figs. 6 and 7).

If the excitation is multiplicative to the displacement process, z_1 , i.e., $b(Z_1, Z_2) = b(Z_1)$, or if it is external (additive), i.e., $b(Z_1, Z_2) = \text{const.} = b$, then when there is an impulse of magnitude p , according to the equation of motion, there is a jump in the velocity by $b(z_1)p$, i.e., a jump from x_2 to $z_2 = x_2 + b(z_1)p$.

The jump probability intensity function $J_Z(z_1, z_2, j | x_1, x_2, i, t)$ determined with the aid of the Markov chain shown in Fig. 7 equals (Iwankiewicz 2008)

$$J_Z(z_1, z_2, j | x_1, x_2, i, t) = \begin{cases} 0, & j = i, \\ \mu \delta(z_1 - x_1) \int_{\mathbf{P}} \delta(z_2 - (x_2 + b(z_1)p)) f_P(p) dp, & j = 1, i = 2, \\ v \delta(z_1 - x_1) \delta(z_2 - x_2), & j = 2, i = 1, \end{cases} \quad (99)$$



Non-Poisson Impulse Processes, Fig. 7 Markov chain for a two-state jump process driven by two independent Poisson processes

where $f_P(p)$ is the probability density function of the random impulse magnitude and P denotes the sample space of the impulse magnitude.

The explicit equations for $j = 1$ and $j = 2$ are obtained after the insertion of the jump probability intensity function Eq. 99 into Eq. 93 and integration with respect to x_1, x_2 , respectively, as (Iwankiewicz 2008)

$$\begin{aligned} \frac{\partial}{\partial t} q_1(z_1, z_2, t) = & - \sum_{r=1}^2 \frac{\partial}{\partial z_r} [c_r(z_1, z_2, t) q_1(z_1, z_2, t)] \\ & + \mu \int_P q_2(z_1, z_2 - b(z_1)p, t) f_P(p) dp - \nu q_1(z_1, z_2, t) \end{aligned} \quad (100)$$

$$\begin{aligned} \frac{\partial}{\partial t} q_2(z_1, z_2, t) = & - \sum_{r=1}^2 \frac{\partial}{\partial z_r} [c_r(z_1, z_2, t) q_2(z_1, z_2, t)] \\ & + \nu q_1(z_1, z_2, t) - \mu q_2(z_1, z_2, t). \end{aligned} \quad (101)$$

At the initial time instant $t = 0$, the continuous processes $Z_1(t) = Y(t)$, $Z_2(t) = Y(t)$ are statistically independent of the jump process, or of the Markov states. It may be assumed that the jump process starts with probability 1 from the first (“off”) state, i.e., $\Pr\{S(0) = 1\} = P_1(0) = 1$ and $\Pr\{S(0) = 2\} = P_2(0) = 0$. Consequently the random initial conditions are written as

$$\begin{aligned} q_Z(z_1, z_2, 1, 0) &= q_1(z_1, z_2, 0) = p(z_1, z_2) \cdot P_1(0) = p(z_1, z_2) \cdot 1 = p(z_1, z_2) \\ q_Z(z_1, z_2, 2, 0) &= q_2(z_1, z_2, 0) = p(z_1, z_2) \cdot P_2(0) = p(z_1, z_2) \cdot 0 = 0, \end{aligned} \quad (102)$$

where $p(z_1, z_2)$ is the joint probability density of $Z_1(0) = Y(0)$, $Z_2(0) = \dot{Y}(0)$.

If the system starts from rest, i.e., $Z_1(0) = Y(0) = 0$, $Z_2(0) = \dot{Y}(0) = 0$, then

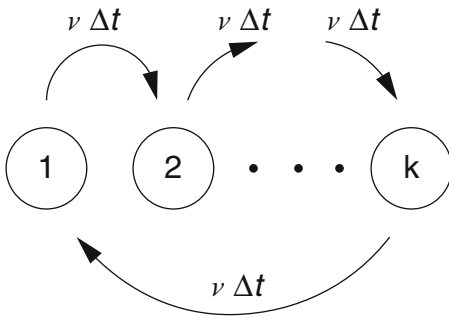
$$\begin{aligned} q_Z(z_1, z_2, 1, 0) &= q_1(z_1, z_2, 0) = \delta(z_1) \delta(z_2) P_1(0) = \delta(z_1) \delta(z_2) \cdot 1 = \delta(z_1) \delta(z_2) \\ q_Z(z_1, z_2, 2, 0) &= q_2(z_1, z_2, 0) = \delta(z_1) \delta(z_2) P_2(0) = \delta(z_1) \delta(z_2) \cdot 0 = 0. \end{aligned} \quad (103)$$

Detailed Integrodifferential Equations for the Joint Probability Density of the State Vector for Renewal Impulse Process Driven by an Erlang Renewal Process The jump process $\rho(N(t))$ is tantamount to a k -state Markov chain $S(t)$ shown in Fig. 8.

In this model, the actual impulse, and hence the jump in the velocity variable, occurs when the jump is from the state k to 1 (cf. Fig. 5). All other jumps in the auxiliary process occur from $j-1$ to j , for $j = 2, \dots, k$, but there are no corresponding impulses.

The jump probability intensity function is expressed as (Iwankiewicz 2006)

$$\begin{aligned} J_Z(z_1, z_2, j | x_1, x_2, i, t) &= \begin{cases} \nu \delta(z_1 - x_1) \int_P \delta(z_2 - (x_2 + b(z_1)p)) f_P(p) dp, & j = 1, i = k. \\ \nu \delta(z_1 - x_1) \delta(z_2 - x_2), & j = 2, 3, \dots, k, i = j - 1 \end{cases} \end{aligned} \quad (104)$$



Non-Poisson Impulse Processes, Fig. 8 Markov chain for a jump process driven by an Erlang renewal process

The governing equations for $j = 1, 2, \dots, k$ are obtained after the insertion of the jump probability intensity function Eq. 104 into Eq. 93 and integration, as (Iwankiewicz 2006)

$$\begin{aligned}
 \frac{\partial}{\partial t} q_1(z_1, z_2, t) &= - \sum_{r=1}^n \frac{\partial}{\partial z_r} [c_r(z_1, z_2, t) q_1(z_1, z_2, t)] \\
 &\quad + v \int_{\mathbf{P}} q_k(z_1, z_2 - b(z_1)p, t) f_P(p) dp \\
 &\quad - v q_1(z_1, z_2, t) \\
 &\quad \vdots \\
 \frac{\partial}{\partial t} q_j(z_1, z_2, t) &= - \sum_{r=1}^n \frac{\partial}{\partial z_r} [c_r(z_1, z_2, t) q_j(z_1, z_2, t)] \\
 &\quad + v q_{j-1}(z_1, z_2, t) - v q_j(z_1, z_2, t), \\
 j &= 2, \dots, k.
 \end{aligned} \tag{105}$$

Generating Equation for Moments

General Equation The original state vector of the dynamic system $\mathbf{Z}(t)$ is not a Markov process,

but the original state variables together with the states of the auxiliary jump process are jointly Markovian. Therefore, the generating equation for moments must be derived for the expectations:

$$E_j[V(\mathbf{Z}(t), t)] = \int_{-\infty}^{\infty} V(\mathbf{z}(t), t) q_j(\mathbf{z}, t) d\mathbf{z}, \quad j = 1, 2, \dots, m. \tag{106}$$

$$\begin{aligned}
 \frac{d}{dt} E_j[V(\mathbf{Z}(t), t)] &= \frac{\partial}{\partial t} \int_{-\infty}^{\infty} V(\mathbf{z}(t), t) q_j(\mathbf{z}, t) d\mathbf{z} \\
 &= E_j \left[\frac{\partial}{\partial t} V(\mathbf{Z}(t), t) \right] + \int_{-\infty}^{\infty} V(\mathbf{z}(t), t) \mathcal{K}_j[q(\mathbf{z}, t)] d\mathbf{z},
 \end{aligned} \tag{107}$$

where $\mathcal{K}_j[\dots]$ is the forward integrodifferential Chapman–Kolmogorov operator (Gardiner 1985; Iwankiewicz and Nielsen 1999):

$$\begin{aligned}
 \mathcal{K}_j[q(\mathbf{z}, t)] &= - \sum_{r=1}^2 \frac{\partial}{\partial z_r} [c_r(\mathbf{z}, t) q_j(\mathbf{z}, t)] \\
 &\quad + \sum_{i=1}^m \int_{-\infty}^{\infty} [J_{\{\mathbf{Z}\}}(\mathbf{z}, j | \mathbf{x}, i, t) q_i(\mathbf{x}, t) \\
 &\quad - J_{\{\mathbf{Z}\}}(\mathbf{x}, j | \mathbf{z}, j, t) q_j(\mathbf{z}, t)] d\mathbf{x}.
 \end{aligned} \tag{108}$$

After the integration by parts of the last term and some rearrangements, the **generating equation for moments** is arrived at (Iwankiewicz 2014)

$$\begin{aligned}
 \frac{d}{dt} E_j[V(\mathbf{Z}(t), t)] &= E_j \left[\frac{\partial}{\partial t} V(\mathbf{Z}(t), t) \right] + \sum_{r=1}^2 E_j \left[\frac{\partial V(\mathbf{Z}(t), t)}{\partial z_r} c_r(\mathbf{Z}(t), t) \right] + \\
 &\quad \sum_{i=1}^m \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [V(\mathbf{y}(t), t) J_{\{\mathbf{Z}\}}(\mathbf{y}, j | \mathbf{z}, i, t) q_i(\mathbf{z}, t) - V(\mathbf{z}(t), t) J_{\{\mathbf{Z}\}}(\mathbf{y}, i | \mathbf{z}, j, t) q_j(\mathbf{z}, t)] d\mathbf{y} d\mathbf{z}, \\
 j &= 1, 2, \dots, m.
 \end{aligned} \tag{109}$$

As the usual-sense marginal joint probability density function $q(\mathbf{z}, t)$ of the state variables is obtained by summation, so is the usual-sense expectation:

$$\begin{aligned} q(\mathbf{z}, t) &= \sum_{j=1}^m q_j(\mathbf{z}, t), \Rightarrow E[V(\mathbf{Z}(t), t)] \\ &= \sum_{j=1}^m E_j[V(\mathbf{Z}(t), t)]. \end{aligned} \quad (110)$$

Detailed Equations for the Renewal Impulse Process Driven by Two Poisson Processes The insertion of the jump probability intensity function Eq. 99 into the generating equation for moments Eq. 109 followed by the integration with respect to y yields the problem-specific generating equations for moments (Iwankiewicz 2014):

$$\begin{aligned} \frac{d}{dt} E_1[V(\mathbf{Z}(t), t)] &= E_1 \left[\frac{\partial}{\partial t} V(\mathbf{Z}(t), t) \right] + \sum_{r=1}^2 E_1 \left[\frac{\partial V(\mathbf{Z}(t), t)}{\partial Z_r} c_r(\mathbf{Z}(t), t) \right] \\ &\quad + \mu E_2 \left[\int_{\mathbf{p}} V(\mathbf{Z}(t) + \mathbf{b}(\mathbf{Z})\mathbf{p}, t) f_p(p) dp \right] - v E_1[V(\mathbf{Z}(t), t)], \\ \frac{d}{dt} E_2[V(\mathbf{Z}(t), t)] &= E_2 \left[\frac{\partial}{\partial t} V(\mathbf{Z}(t), t) \right] + \sum_{r=1}^2 E_2 \left[\frac{\partial V(\mathbf{Z}(t), t)}{\partial Z_r} c_r(\mathbf{Z}(t), t) \right] \\ &\quad + v E_1[V(\mathbf{Z}(t), t)] - \mu E_2[V(\mathbf{Z}(t), t)], \end{aligned} \quad (111)$$

where $\mathbf{b}(\mathbf{Z}) = [0, b(Z_1, Z_2)]^T$. As a result of integration with respect to $\mathbf{z}$, the Markov state probabilities may appear

$$\int_{-\infty}^{\infty} q_j(\mathbf{z}, t) d\mathbf{z} = \mathcal{P}_j(t) = \Pr\{S(t) = j\}. \quad (112)$$

Differential equations governing the time evolution of Markov states probabilities $\mathcal{P}_j(t)$ are derived from the general expression for m -state Markov chain:

$$\mathcal{P}_j(t + \Delta t) = \sum_{i=1}^m \mathcal{P}_{ji}(\Delta t) \mathcal{P}_i(t), \quad j = 1, 2, \dots, m. \quad (113)$$

Detailed Equations for the Impulse Process Driven an Erlang Renewal Process The insertion of the jump probability intensity function Eq. 104 into the generating equation for moments Eq. 109 followed by the integration with respect to y yields the problem-specific generating equations for moments (Iwankiewicz 2014):

$$\begin{aligned} \frac{d}{dt} E_1[V(\mathbf{Z}(t), t)] &= E_1 \left[\frac{\partial}{\partial t} V(\mathbf{Z}(t), t) \right] + \sum_{r=1}^2 E_1 \left[\frac{\partial V(\mathbf{Z}(t), t)}{\partial Z_r} c_r(\mathbf{Z}(t), t) \right] \\ &\quad + v E_k \left[\int_{\mathbf{p}} V(\mathbf{Z}(t) + \mathbf{b}(\mathbf{Z})\mathbf{p}, t) f_p(p) dp \right] - v E_1[V(\mathbf{Z}(t), t)], \\ \frac{d}{dt} E_j[V(\mathbf{Z}(t), t)] &= E_j \left[\frac{\partial}{\partial t} V(\mathbf{Z}(t), t) \right] + \sum_{r=1}^2 E_j \left[\frac{\partial V(\mathbf{Z}(t), t)}{\partial Z_r} c_r(\mathbf{Z}(t), t) \right] \\ &\quad + v E_{j-1}[V(\mathbf{Z}(t), t)] - v E_j[V(\mathbf{Z}(t), t)], \quad j = 2, 3, \dots, k. \end{aligned} \quad (114)$$

Cross-References

- [Probability Density Evolution Method in Stochastic Dynamics](#)
- [Stochastic Analysis of Linear Systems](#)
- [Stochastic Analysis of Nonlinear Systems](#)

References

- Cox DR (1962) *Renewal theory*. Methuen, London
- Cox DR, Isham V (1980) *Point processes*. Chapman and Hall, London
- Gardiner CW (1985) *Handbook of stochastic methods for physics, chemistry and the natural sciences*. Springer, New York
- Iwankiewicz R (1995) *Dynamical mechanical systems under random impulses*. World Scientific, series on advances in mathematics for applied sciences, vol 36. World Scientific. Singapore, New Jersey, London, Hong Kong
- Iwankiewicz R (2002) Dynamic response of non-linear systems to random trains of non-overlapping pulses. *Meccanica* 37:167–178
- Iwankiewicz R. (2003). Dynamic systems under random impulses driven by a generalized Erlang renewal process. In: Furuta H, Dogaki M, Sakano M (eds) *Proceedings of the 10th IFIP WG 7.5 working conference on reliability and optimization of structural systems*, 25–27 March 2002, Kansai University/Balkema, Osaka, pp 103–110
- Iwankiewicz R (2006) Equation for probability density of the response of a dynamic system to Erlang renewal random impulse processes. In: Sørensen JD, Frangopol DM (eds) *Proceedings of the 12th IFIP WG 7.5 working conference on reliability and optimization of structural systems*, 22–25 May 2005, Taylor and Francis, Aalborg, pp 107–113
- Iwankiewicz R (2008) Equations for probability density of response of dynamic systems to a class of non-Poisson random impulse process excitations. *Probab Eng Mech* 23:198–207
- Iwankiewicz R (2014) Response of dynamic systems to renewal impulse processes: generating equation for moments based on the integro-differential Chapman-Kolmogorov equations. *Probab Eng Mech* 35:52–66
- Iwankiewicz R, Nielsen SRK (1999) *Advanced methods in stochastic dynamics of non-linear systems*. Aalborg University Press, Denmark
- Iwankiewicz R, Nielsen SRK, Thoft-Christensen P (1990) Dynamic response of non-linear systems to Poisson-distributed pulse trains: Markov approach. *Struct Saf* 8:223–238
- Snyder DL, Miller MI (1991) *Random point processes in time and space*. Springer, New York

- Srinivasan SK (1974) *Stochastic point processes and their applications*. Griffin, London
- Stratonovich RL (1963) *Topics in the theory of random noise*. Gordon and Breach, New York/London
- Takacs L (1956) On secondary stochastic processes generated by recurrent processes. *Acta Math Acad Sci Hung* 7:17–29

Novel Bio-Inspired Sensor Network for Condition Assessment

Simon Laflamme

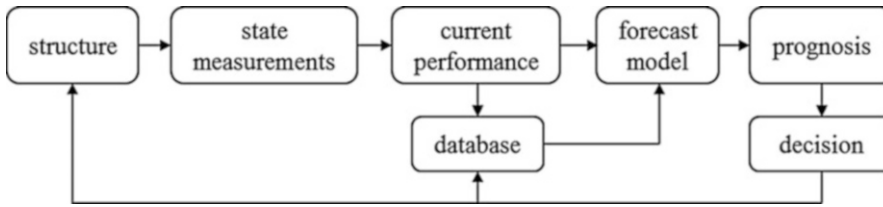
Department of Civil, Construction, and Environmental Engineering, Iowa State University, Ames, IA, USA

Synonyms

Bio-inspired sensor; Condition assessment; Flexible strain gauge; Sensing skin; Sensor network; Shape reconstruction; Soft elastomeric capacitor; Structural health monitoring

Introduction

Condition assessment of civil structures is a task dedicated to forecasting future structural performances based on current states and past performances and events. The concept of condition assessment is often integrated within a closed-loop decision, where structural conditions can be adapted based on system prognosis. Figure 1 illustrates a particular way to conduct condition assessment. In the process, various structural states are measured, which may include excitations (e.g., wind, vehicles) and responses (e.g., strain, acceleration). These measurements are processed to extract indicators (e.g., maximum strain, fundamental frequencies) of current structural performance. These indicators are stored in a database, and also used within a forecast model (e.g., time-dependent reliability, Markov decision process) that will lead to a prognosis on the structural system, enabling optimization of



Novel Bio-Inspired Sensor Network for Condition Assessment, Fig. 1 Condition assessment process

structural management decisions (e.g., inspection, repairs, maintenance). The forecast model itself may query information from the database that include past performance indicators and events. One of the fundamental benefits of condition assessment is the availability of condition-based maintenance decisions (CBM), which enables maintenance in function of current and expected conditions rather than based on usage rate (preventive-based maintenance) or breakdown (breakdown-based maintenance). CBM has the potential to significantly improve life-cycle costs and structural resiliency by optimizing maintenance and inspection schedules and forecasting structural behavior (Jardine et al. 2006). It could also be used for automatic assessment of structural conditions following natural hazards (e.g., earthquakes).

The literature on structural prognosis and condition assessment has fundamentally focused on the specialized task of damage diagnosis, including sensing hardware and signal processing methods, but much remains to be done on developing integrated sensing solution that could lead to condition assessment (Jardine et al. 2006). The vast majority of cited work in condition assessment is in the field of machinery, and applications in civil engineering are mostly limited to bridges. See References (Frangopol et al. 2004; Teughels and De Roeck 2004; Perera and Ruiz 2008) for examples. Other specialized civil engineering applications include buildings (Savadkoohi et al. 2011), pipes (Dilena et al. 2011), and wind turbine blades (Abounhnik and Albarbar 2012). While a literature survey indicates a growth in research on condition assessment methods for civil structures, the full potential of these methods is yet to be realized (Farrar and Lieven

2007). Some challenges impeding broad applicability include the needs to develop (1) algorithms enabling real-time decision making; (2) robust sensing systems for online data acquisition; (3) methods for collecting event information; and (4) more accurate forecasting models (Frangopol et al. 2004; Jardine et al. 2006; Farrar and Lieven 2007).

The author has developed a novel sensing method designed for condition assessment of civil structures. The method consists of an array of soft elastomeric capacitors (SECs) acting as flexible strain gages. Arranged in a network, SECs are capable of covering very large areas at low costs. It was demonstrated that a network was capable of covering an area of $70 \times 280 \text{ mm}^2$ with only four sensors (Laflamme et al. 2013a). Analogous to biological skin, the SEC network can localize strain over a global area. The technology is an alternative to fiber optics technologies. With both fiber optic sensors and the SEC technology, strain data can be measured over large systems. Others have proposed alternatives to conventional strain sensing, including conducting cement mixes (Materazzi et al. 2013) and piezoelectric networks (Giurgiutiu 2009). Conducting polymers, such as soft resistors and capacitors, have also gained popularity for structural health monitoring applications (Tata et al. 2009; Loh et al. 2009; Gao et al. 2010). The proposed SEC differs from existing literature in that it combines both a large physical size and high initial capacitance, resulting in a larger surface coverage and higher sensitivity. Also, it combines the advantages of being cost-effective, easy to install, robust with respect to mechanical tampering, and customizable in shapes and sizes, and low powered.

Deployment of the sensor network over large areas allows the measurement of strains over large surfaces. These measurements can be used to reconstruct physics-based features associated with the structural behavior. For instance, the analysis of deflection shapes can give insights on structural performance, whether it is by studying changes in curvature through time, or simply by counting the number of cycles and/or over-stresses. These physics-based features can be integrated into a forecast model to establish the severity of the problem and enable decision making.

In this article, the promise of the SEC network for condition assessment applications is presented. The next section describes the SEC used in the network setup. The description includes a discussion of the fabrication process, the electromechanical model used for converting signal into strain and shows a comparison with off-the-shelf resistance-based strain gauges (RSGs). The subsequent section discusses the application of the sensor in a sensor network for conducting condition assessment. It also describes an algorithm used for extracting physics-based features from the network signal and demonstrates the application. The last section concludes the article.

Soft Elastomeric Capacitors

The proposed sensor network for condition assessment applications has been developed by the author (Laflamme et al. 2012, 2013a). It consists of an array of SECs, a type of conducting polymers. The field of conducting polymers has been pioneered in the 1970s, when it was discovered that polymers can not only be used as insulators but also as conducting mediums (Shirakawa et al. 1977). They have since then been used for various purposes, including flexible sensors and actuators (Osada and De Rossi 2000). These synthetic metals typically originate from the constitution of a nanocomposite mix of organic and inorganic particles, which can be obtained via chemical and electromechanical preparations, as discussed in Reference

(Gangopadhyay and De 2000). Figure 2 shows the principle using scanning electron microscope (SEM) photos, in which an organic material (poly-styrene-co-ethylene-co-butylene-co-styrene (SEBS), Fig. 2a) is mixed with inorganic particles (titanium dioxide (TiO_2), Fig. 2b) to form a nanocomposite mix SEBS + TiO_2 (Fig. 2c).

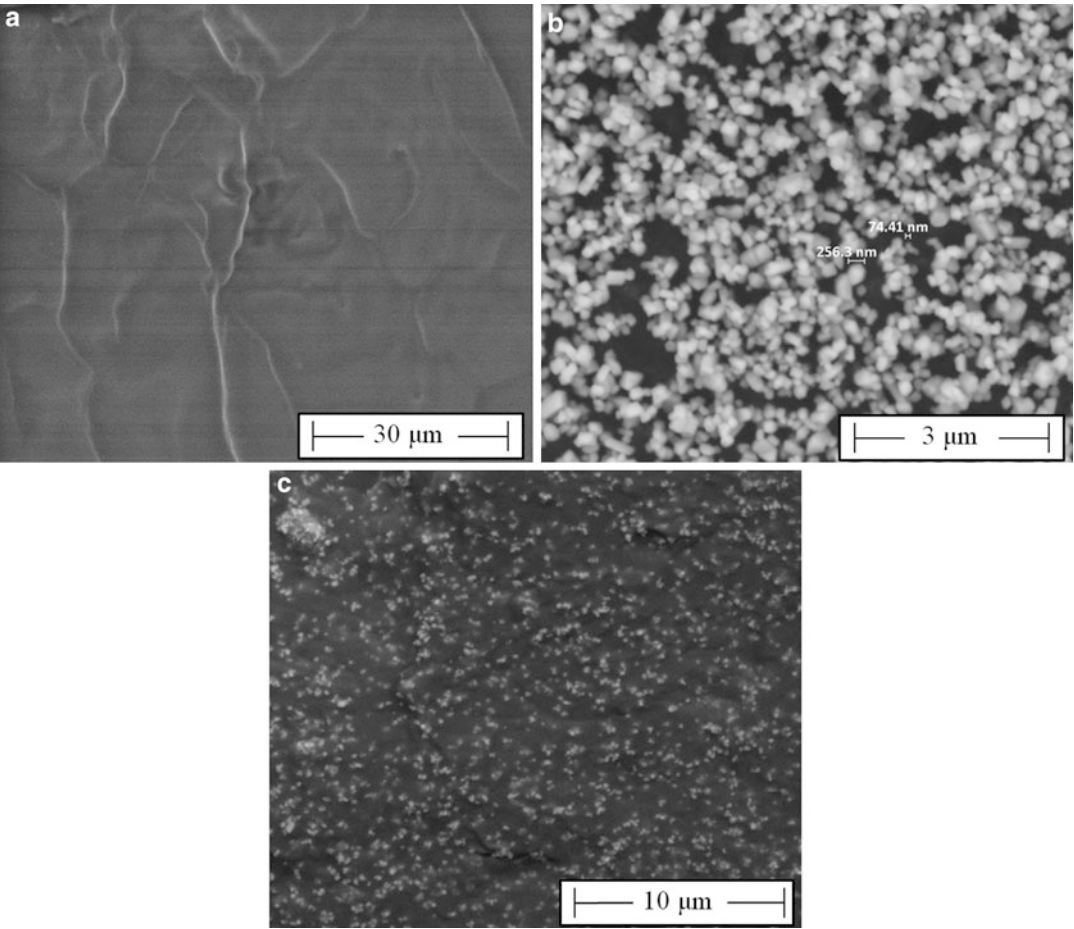
The SEC is fabricated using the principles of conducting polymers. Figure 3a shows the schematic of a capacitor, constituted from a dielectric sandwiched between two conducting plates that can be connected to an electric charge. It follows that a soft capacitor can be built from an elastomeric dielectric layer sandwiched between compliant electrodes. Here, the dielectric is a nanocomposite mix of SEBS doped with TiO_2 , the same materials showed in Fig. 2. The compliant electrodes are fabricated from a nanocomposite of SEBS and carbon black (CB). Figure 3b shows a picture of a single SEC. In this section, the fabrication process of the SEC is described, its electromechanical model derived, and its performance versus off-the-shelf RSGs compared.

Fabrication

Figure 4 illustrates the fabrication process of a SEC. First, a solution of SEBS dissolved in toluene (solvent) is created. Then, TiO_2 particles are dispersed in part of this solution using a sonication process. The resulting SEBS + TiO_2 mix is spread over a glass slide and allowed to dry, during which phase the solution becomes solid and the toluene evaporates. Meanwhile, the CB particles are dispersed in the remaining SEBS-toluene solution, also using sonication. Finally, this electrode mix is painted or sprayed onto the top and bottom surfaces of the dielectric and allowed to dry.

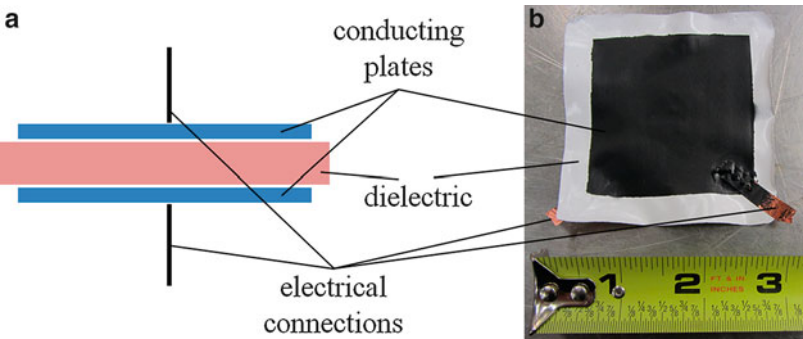
The SEC's dielectric layer is doped with TiO_2 to improve the sensor's mechanical properties, specifically its dielectric properties. The capacitance value C of a SEC is written

$$C = \epsilon_0 \epsilon_r \frac{A}{h}$$



Novel Bio-Inspired Sensor Network for Condition Assessment, Fig. 2 SEM photos of (a) SEBS, (b) TiO₂, and (c) SEBS + TiO₂

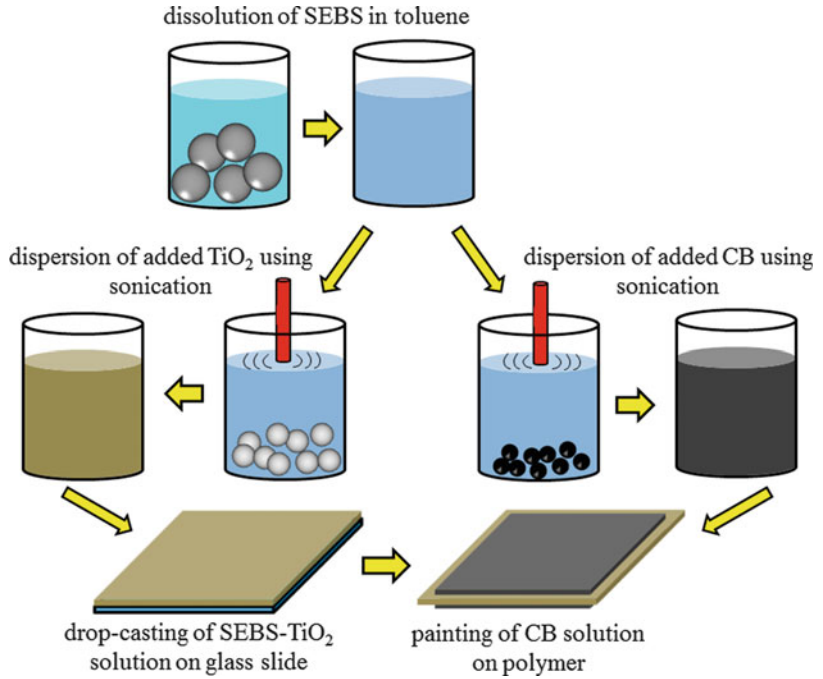
Novel Bio-Inspired Sensor Network for Condition Assessment, Fig. 3 (a) Schematic of a capacitor; and (b) a single SEC (bottom electrode not shown)



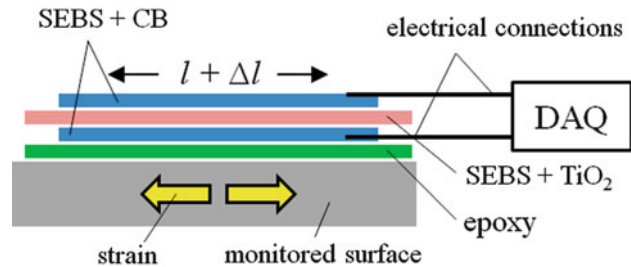
where $\epsilon_0 = 8.854 \text{ pF/m}$ is the vacuum permittivity, ϵ_r the dimensionless polymer relative permittivity, $A = wl$ the sensor area with width w and length l , and h the height of the

dielectric. Altering the nanocomposition of the dielectric enables a customization of ϵ_r . For instance, the author has shown that it was possible to dramatically increase the relative

Novel Bio-Inspired Sensor Network for Condition Assessment, Fig. 4 Fabrication process



Novel Bio-Inspired Sensor Network for Condition Assessment, Fig. 5 Sensing principle (layers not scaled)



permittivity of the SEBS by grafting polyaniline (PANI) on the polymer backbone (Kollosche et al. 2011). Here, TiO₂ is used due to the low cost and high stability of the particles.

Electromechanical Model

The sensing principle of the electromechanical sensor consists in measuring strain via changes in capacitance. Figure 5 illustrates the sensing principle for a SEC glued onto a monitored surface using an epoxy. In the example, a strain in the monitored surface provokes a change in the sensor geometry $\hat{l}$. This strain is linearly transduced by a change in the capacitance $\hat{l}^nC$ of the

sensor. This change is measured by a data acquisition system (DAQ).

The SEC materials can be considered as incompressible (the Poisson ratio of SEBS $\nu \approx 0.49$ (Wilkinson et al. 2004)). It follows that the sensor volume V is preserved ($\Delta V = 0$):

$$V = V + \Delta V$$

$$w \cdot l \cdot h = (w + \Delta w) (w + \Delta l) (h + \Delta h)$$

Ignoring higher order terms, the last equation can be written:

$$-\frac{\Delta h}{h} \approx \frac{\Delta l}{l} + \frac{\Delta w}{w}$$

$$-\varepsilon_z \approx \varepsilon_x + \varepsilon_y$$

Also, for small changes in C , the differential of the equation governing capacitance is taken as:

$$\Delta C = \left(\frac{\Delta l}{l} + \frac{\Delta w}{w} - \frac{\Delta h}{h} \right) C$$

$$\frac{\Delta C}{C} = \varepsilon_x + \varepsilon_y - \varepsilon_z$$

Substituting ε_z , the last equation becomes:

$$\frac{\Delta C}{C} = 2(\varepsilon_x + \varepsilon_y)$$

which results in a gauge factor $\lambda = 2$. The equation above shows that the sensor measures additive strain. The principal strain components and magnitudes can be decomposed by leveraging network applications of the SEC. This is out-of-the-scope of this chapter.

While the nanocomposite mix does not influence the gauge factor, it plays an important role in the sensor sensitivity $\frac{\Delta C}{\varepsilon_x + \varepsilon_y} = 2C$. It results that the sensitivity can be improved by increasing the materials permittivity ε_r , resulting in a better resolution. The sensitivity can also be increased by altering the sensor geometry.

Comparison Versus Off-the-Shelf Strain Gauge

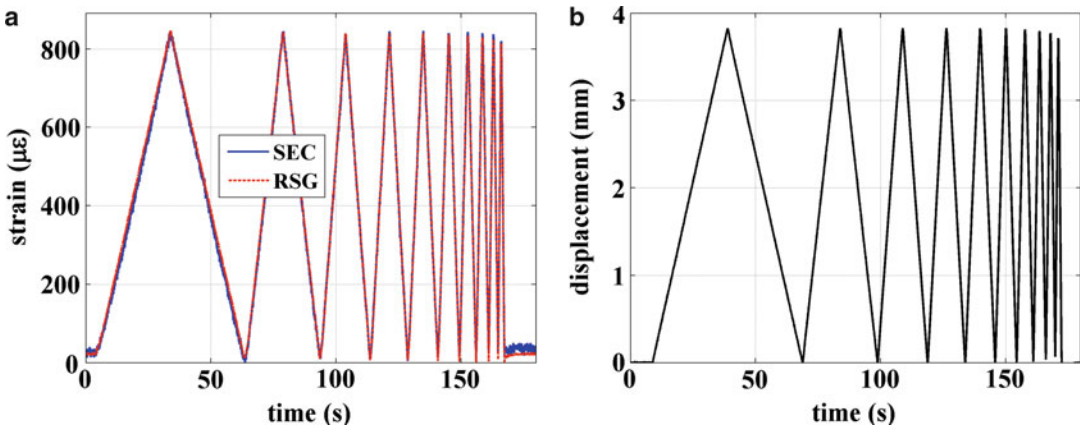
In this subsection, a performance comparison between an SEC and an off-the-shelf RSG is presented. Additional details on this comparison can be found in Reference (Laflamme et al. 2013a). The test setup consists of a three-point load setup on a simply supported aluminum beam of support-to-support dimensions $406.4 \times 101.6 \times 6.35 \text{ mm}^3$ ($16 \times 4 \times 0.25 \text{ in}^3$). A SEC and a RSG (Vishay Micro-Measurements, CEA-06-500UW-120, resolution of $1 \mu\varepsilon$) are installed centered onto the bottom surface of the beam. The setup is similar to Fig. 8, except with one sensor of each type, centered. Both sensors are installed following a similar procedure. The monitored surface is sanded, painted with a primer, and a thin layer

of an off-the-shelf epoxy (JB Kwik) is applied on which the sensors are adhered. Data from the SECs are acquired using an inexpensive off-the-shelf data acquisition system (ACAM PCap01) sampled at 48 Hz. RSG data are acquired using a Hewlett-Packard 3852 data acquisition system, and data sampled at 55 Hz. The excitation history consists of a displacement-based triangular wave loads with increasing frequencies from 0.0167 to 0.40 Hz to remain in a quasi-static range.

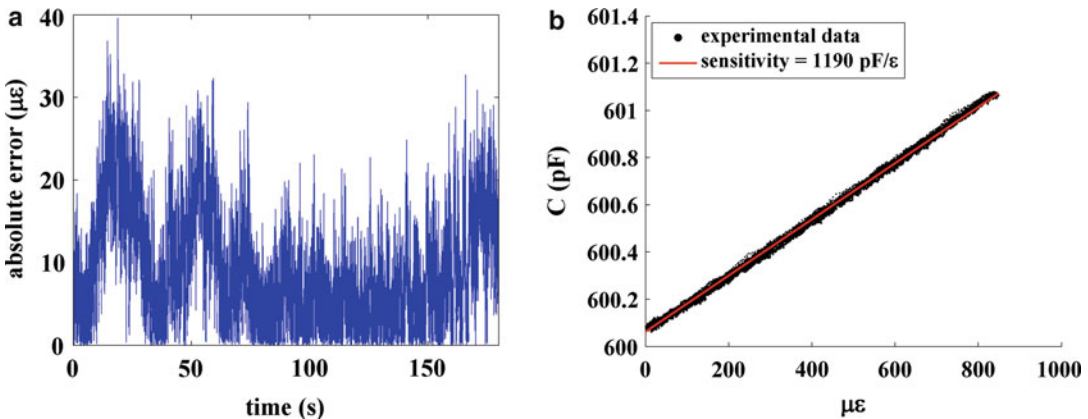
Figure 6 shows the results from the test. The time series responses from the SEC and RSG are shown in Fig. 6a, along with the loading history in Fig. 6b. Results from the SEC compares well against readings from the off-the-shelf RSG. Figure 7a shows the absolute error between the SEC and RSG readings. The SEC can track the time history with a resolution of 25–30 $\mu\varepsilon$. This resolution could be improved with the fabrication of a dedicated DAQ system (Laflamme et al. 2012). Figure 7b studies the sensitivity of the SEC obtained experimentally versus the theoretical value. The experimental sensitivity of 1190 pF/ ε is close to the theoretical value of $2 \times 600 = 1,200 \text{ pF}/\varepsilon$, a 0.84 % difference. Also, results from Fig. 7b exhibit linearity, consistent with theory.

Sensor Network for Condition Assessment

The previous section described the theory for SECs and showed that it can be used effectively as a large-scale strain gauges. In this section, the SEC concept is extended to multiple sensors installed in an array form, which allows monitoring of large surface areas. In the context of condition assessment, spatial and temporal surface strain data can be assembled, from which physics-based features can be extracted. As discussed in the introduction, these features can be stored, compared, and analyzed to evaluate structural usage, or to detect changes and/or anomalies in the structural behavior. This study of structural behavior can be incorporated in a forecast model to enable prognosis



Novel Bio-Inspired Sensor Network for Condition Assessment, Fig. 6 Strain gauges comparison: (a) time series response; and (b) loading history



Novel Bio-Inspired Sensor Network for Condition Assessment, Fig. 7 (a) Absolute error between the SEC and RSG; and (b) capacitance versus strain for the SEC

of the structural system, with the overarching objective to optimize structural management decisions.

Here, the promise of SEC networks at conducting condition assessment is shown by extracting deflection shapes from the measurements. Remark that the design of accurate forecast models based on signal features is a challenging problem and constitutes an active field of research. Reference (Frangopol et al. 2004) reviews some fundamental forecast models, and constitutes a good introduction to the research problem for the interested reader.

Feature Extraction

Spatial and temporal strain data contain rich information about the monitored system. However, the task of condition assessment cannot be successfully conducted without the extraction of meaningful features from data. The problem is analogous to signals from accelerometers from which, in most cases, frequency features (e.g., fundamental frequencies, mode shapes) need to be extracted to enable the analysis of the vibration signature. Several types of features can be extracted from an array of strain gauges,

including local plastic deformations, overstrains, cycles, and deflection shapes. Here, the concept of the sensor network for condition assessment is demonstrated by extracting deflection shapes from a monitored surface.

The problem of real-time reconstruction of deflection shapes from position and curvature measurements from sensor networks has been widely studied, with applications to condition assessment, structural health monitoring, and shape control. See references (Jones et al. 1999; Glaser et al. 2012); for instance. Here, the algorithm consists of fitting the curvature data using a polynomial interpolation and double-integrating to obtain the deflection shape. In the case of a two-dimensional beam equipped with four sensors, the fitting function is taken as a third degree polynomial to avoid possible over-fitting and also allows some additional filtering on the measured strain data (Jones et al. 1999; Glaser et al. 2012):

$$\hat{\epsilon}_{m,i} = a_0 + a_1 x_i + a_2 x_i^2 + a_3 x_i^3$$

where the *hat* denotes an estimation for the *i*th sensor, a_1 to a_4 are constants, and x is the Cartesian location $0 \leq x \leq L$ along the beam of length L . Minimizing the error J for n sensors:

$$J = \sum_i^n (\epsilon_{m,i} - \hat{\epsilon}_{m,i})^2$$

leads to the expression:

$$\mathbf{A} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{\Xi}_m$$

with:

$$\mathbf{A} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} \quad \mathbf{\Xi}_m = \begin{bmatrix} \epsilon_{m,1} \\ \epsilon_{m,2} \\ \dots \\ \epsilon_{m,n} \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}$$

The deflection shape $y(x)$ is obtained by integrating the curvature twice:

$$\begin{aligned} y(x) &= \int_0^L \int_0^L \frac{\delta^2 y}{\delta x^2} dx^2 \\ &= \int_0^L \int_0^L -\frac{\epsilon_m}{c} dx^2 \\ &= \int_0^L \int_0^L -\frac{1}{c} (a_0 + a_1 x_j + a_2 x_j^2 + a_3 x_j^3) dx^2 \\ &= -\frac{1}{c} \left(a_0 \frac{x^2}{2} + a_1 \frac{x^3}{6} + a_2 \frac{x^4}{12} + a_3 \frac{x^5}{20} \right) + b_1 x + b_2 \end{aligned}$$

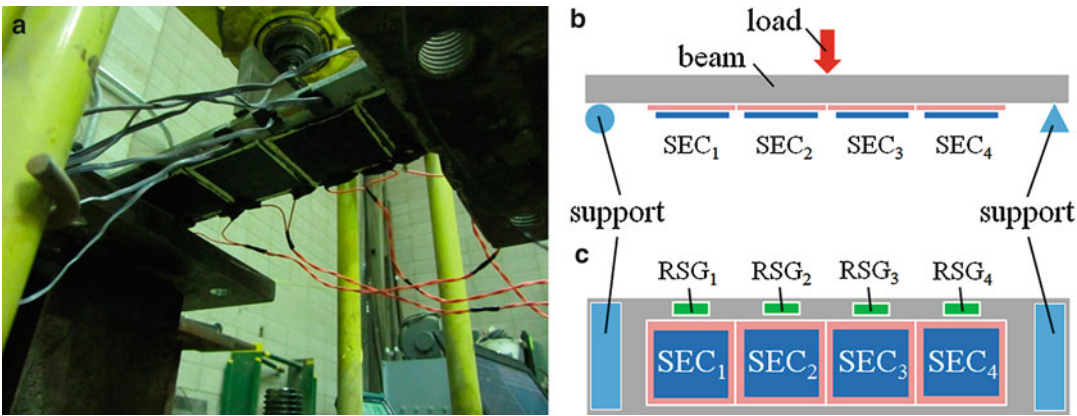
where c is the distance from the surface to the centroid of the beam, and b_1 and b_2 , are constants that can be determined by enforcing boundary conditions. For example, in the case of a simply-supported beam ($y(0) = y(L) = 0$):

$$\begin{aligned} b_1 &= \frac{1}{c} \left(a_0 \frac{L}{2} + a_1 \frac{L^2}{6} + a_2 \frac{L^3}{12} + a_3 \frac{L^4}{20} \right) \\ b_2 &= 0 \end{aligned}$$

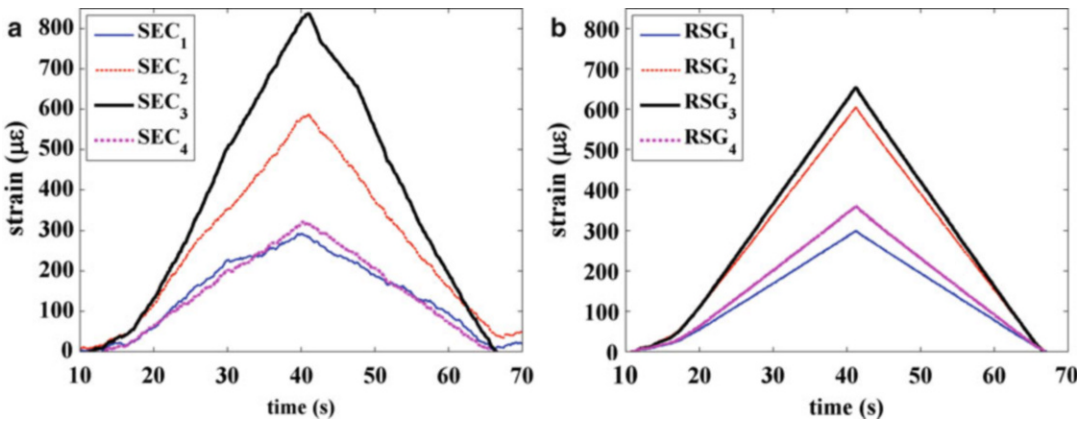
Laboratory Demonstration

The method for extracting the deflection shape feature explained in the previous subsection is demonstrated in what follows. Additional details on the test can be found in Reference (Laflamme et al. 2013a). The test setup, shown in Fig. 8, consists of the same aluminum specimen used for the comparison against the RSG, with the number of SECs extended to four in order to create a sensor network. The SECs and RSGs are located under the beam at $x = \{0.20, 0.40, 0.60, 0.80\}L$, and a three-point load setup is used. A similar experimental procedure as in the previous section is used. The objective of this laboratory verification is to extract deflection shapes from the sensor network.

Figure 9 plots the sensors signals during the first triangular load for the SECs (Fig. 9a) and for the RSGs (Fig. 9b). SEC₁ and SEC₄ have closely spaced signals, as it would be expected for strain gauges symmetrically placed, while SEC₃ shows a substantial difference with respect to SEC₂. The



Novel Bio-Inspired Sensor Network for Condition Assessment, Fig. 8 Laboratory setup. (a) Picture of the setup; (b) setup schematic, elevation view; and (c) setup schematic, bottom plan view



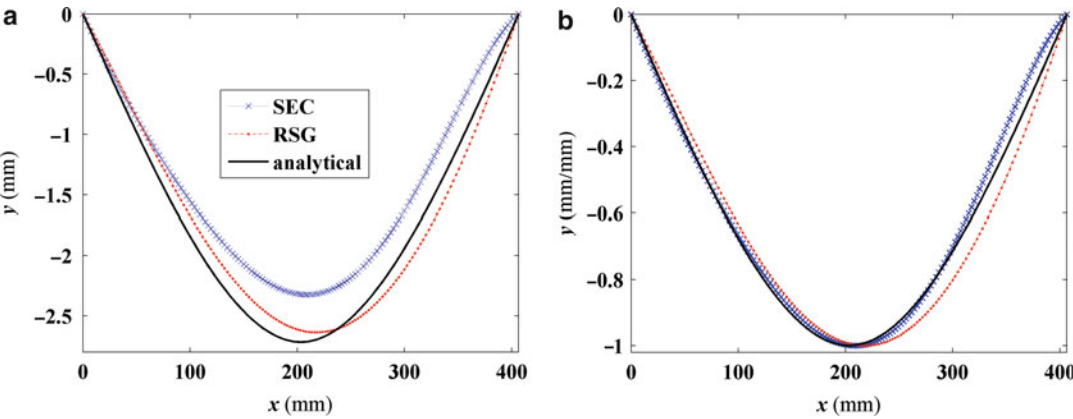
Novel Bio-Inspired Sensor Network for Condition Assessment, Fig. 9 Sensors signals for the first triangular load. (a) SECs; and (b) RSGs

signals for the RSGs are coupled, with a small difference that can be explained by small asymmetries in the sensors placement. All SECs, except for SEC₃, underestimate strain with respect to RSGs.

Figure 10a shows the deflection shapes taken at time $t = 40$ s extracted using the methodology previously described. Results are benchmarked against the analytical solution obtained from the Euler-Bernoulli beam theory. The SECs underestimate the deflection shape as a result of the underestimation of strain from SEC₁ SEC₂, and SEC₄. The deflection shape from the RSG is closer to the analytical result, but with a shift

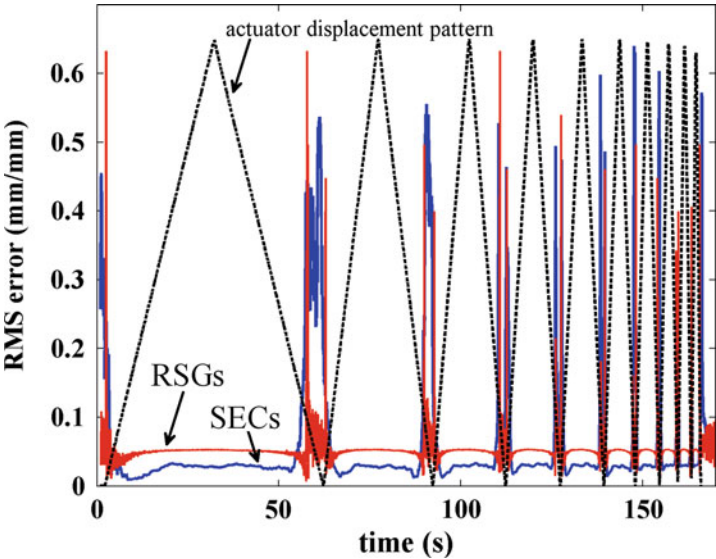
of the maximum deflection to the right. Figure 10b shows the deflection shapes normalized to their maximum unit deflection. Results show that the SECs give a better estimate of the normalized deflection shapes compared to RSGs. The shift of the maximum deflection point can be explained by the slightly higher strain readings obtained with RSG₃ and RSG₄, both located to the right-hand-side of the beam.

The promise of the SEC network for extracting deflection shapes is further investigated by comparing the root mean square (RMS) error of the normalized deflection shapes with respect to the analytical solution. Results



Novel Bio-Inspired Sensor Network for Condition Assessment, Fig. 10 Deflection shapes extracted from the sensors signals. (a) Non-normalized; and (b) normalized

Novel Bio-Inspired Sensor Network for Condition Assessment, Fig. 11 RMS error of the normalized deflection shapes with respect to the analytical solution



are shown in Fig.11 for the entire loading history. The SEC network obtains a more accurate shape than the RSG network beyond an initial level of loading. When the load is around zero, the noise in the sensors signals results in highly inaccurate deflection shapes. The significant difference in performance between both sensors can be attributed to the SECs averaging strain over a large area, while the RSGs measure a localized strain. SECs are less sensitive to placement errors.

Summary

A novel bio-inspired sensing solution has been presented for condition assessment of civil structures. The technology consists of a SEC transducing changes in strain into changes in capacitance. Arranged in a network setup, the technology can be used to extract physics-based features for condition assessment. Comparisons against an off-the-shelf RSG showed that the SEC is capable of tracking a quasi-static strain history at a

resolution in the range of 25–30 $\mu\epsilon$. This resolution is limited by existing off-the-shelf DAQ systems dedicated to capacitance measurements. Laboratory verifications demonstrated the performance of the SEC network at extracting deflection shapes. The study of the RMS error showed that the SEC network provided accurate normalized deflection shapes, with performance levels beyond the RSG network. This performance can be explained partly by the capacity of the SEC to average strain over a large area, unlike RSGs that measure localized strain. Thus, slight misplacement of SECs has minimum consequences on the shape extraction. Given the results from this laboratory demonstration, the SEC network offers great promise as a sensing method for condition assessment of civil structures.

References

- Abouhnik A, Albarbar A (2012) Wind turbine blades condition assessment based on vibration measurements and the level of an empirically decomposed feature. *Energy Conversion and Management* 64(2012):606–613
- Dilena M, DellOste M, Morassi A (2011) Detecting cracks in pipes filled with fluid from changes in natural frequencies. *Mech Syst Signal Process* 25(8):3186–3197
- Farrar C, Lieven N (2007) Damage prognosis: the future of structural health monitoring. *Philos Trans R Soc A Math Phys Eng Sci* 365:623–632
- Frangopol D, Kallen M, Noortwijk J (2004) Probabilistic models for life-cycle performance of deteriorating structures: review and future directions. *Prog Struct Eng Mater* 6(4):197–212
- Gangopadhyay R, De A (2000) Conducting polymer nanocomposites: a brief overview. *Chem Mater* 12(3):608–622
- Gao L, Thostenson E, Zhang Z, Byun J, Chou T (2010) Damage monitoring in fiber-reinforced composites under fatigue loading using carbon nanotube networks. *Philos Mag* 90(31–32):4085–4099
- Giurgiutiu V (2009) Piezoelectricity principles and materials. In: *Encyclopedia of structural health monitoring*. Wiley, pp 981–991
- Glaser R, Caccese V, Shahinpoor M (2012) Shape monitoring of a beam structure from measured strain or curvature. *Exp Mech* 52(6):591–606
- Jardine A, Lin D, Banjevic D (2006) A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mech Syst Signal Process* 20(7):1483–1510
- Jones R, Bellemore D, Berko T, Sirkis J, Davis M, Putnam M, Friebele E, Kersey A (1999) Determination of cantilever plate shapes using wavelength division multiplexed fiber bragg grating sensors and a least-squares strain-fitting algorithm. *Smart Mater Struct* 7(2):178
- Kollosche M, Stoyanov H, Laflamme S, Kofod G (2011) Strongly enhanced sensitivity in elastic capacitive strain sensors. *J Mater Chem* 21:8292–8294
- Laflamme S, Kollosche M, Connor J, Kofod G (2012) Soft capacitive sensor for structural health monitoring of large-scale systems. *Struct Control Health Monitor* 19.1 (2012):70–81
- Laflamme S, Kollosche M, Kollipara VD, Saleem HS, Kofod G (2012) Large-scale surface strain gauge for health monitoring of civil structures. In: *SPIE smart structures and materials/nondestructive evaluation and health monitoring*. International Society for Optics and Photonics, pp 83471P–83471P
- Laflamme S, Saleem HS, Vasan BK, Geiger RL, Chen D, Kessler MR, Rajan K (2013a) Soft elastomeric capacitor network for strain sensing over large surfaces. *IEEE/ASME Trans Mech*
- Laflamme S, Kollosche M, Connor J, Kofod G (2013b) Robust flexible capacitive surface sensor for structural health monitoring applications. *ASCE J Eng Mech* 139(7):879–885
- Loh K, Hou T, Lynch J, Kotov N (2009) Carbon nanotube sensing skins for spatial strain and impact damage identification. *J Nondestruct Eval* 28(1):9–25
- Materazzi A, Ubertini F, D'Alessandro A (2013) Carbon nanotube cement-based transducers for dynamic sensing of strain. *Cem Concr Compos* 37:2–11
- Osada Y, De Rossi DE (2000) *Polymer sensors and actuators*. Springer, Germany
- Perera R, Ruiz A (2008) A multistage fe updating procedure for damage identification in large-scale structures based on multiobjective evolutionary optimization. *Mech Syst Signal Process* 22(4):970–991
- Savadkoobi A, Molinari M, Bursi O, Friswell M (2011) Finite element model updating of a semi-rigid moment resisting structure. *Struct Control Health Monit* 18(2):149–168
- Shirakawa H, Louis EJ, MacDiarmid AG, Chiang CK, Heeger AJ (1977) Synthesis of electrically conducting organic polymers: halogen derivatives of polyacetylene, (CH) x . *J Chem Soc Chem Commun* 16:578–580
- Tata U, Deshmukh S, Chiao J, Carter R, Huang H (2009) Bio-inspired sensor skins for structural health monitoring. *Smart Mater Struct* 18:104026
- Teughels A, De Roeck G (2004) Structural damage identification of the highway bridge Zz24 by FE model updating. *J Sound Vib* 278(3):589–610
- Wilkinson A, Clemens M, Harding V (2004) The effects of sebs-g-maleic anhydride reaction on the morphology and properties of polypropylene/PA6/sebs ternary blends. *Polymer* 45(15):5239–5249

Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings

Ioannis Koutromanos¹ and P. Benson Shing²

¹Department of Civil and Environmental Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

²Department of Structural Engineering, University of California, San Diego, La Jolla, CA, USA

Synonyms

Creep; Equivalent strut model; Finite element model; Infilled frame; Masonry; Nonlinear analysis; Reinforced concrete

Introduction

Masonry-infilled reinforced concrete frames constitute a significant portion of the building inventory in seismically active regions around the world. Even though early efforts to analyze the behavior of infilled frames (e.g., Polyakov 1960) date back to more than half a century ago, the modeling of the interaction between the frame members and the masonry infill walls is still an active research area.

It is well known that the resistance of an infilled frame is not a simple sum of the resistance of a bare frame and that of the infill wall due to the fact that the load-resistance mechanism of a frame can change as a result of its interaction with the infill. As described in ASCE/SEI 41 (ASCE/SEI 2007) and shown in Fig. 1, when the system is subjected to a horizontal force acting toward the right, the frame tends to separate from the infill wall at the bottom left and top right corners. Compressive contact stresses develop at the other two corners.

A numerical model for the analysis of infilled frames must be able to capture the effect of the frame–infill separation and the development of compressive contact stresses at two of the four

corners of the infill wall. The damage and nonlinear response of the RC frame members and of the infill walls also need to be accounted for. A masonry infill wall can fail by corner crushing, as shown in Fig. 2a, or shear sliding along bed joints, as shown in Fig. 2b. Shear sliding is expected to be important for cases when the mortar bed joints are relatively weak as compared to the masonry units. Another damage mode, which is common for older structures with relatively weak frame members, has diagonal/sliding cracks developed in the infill wall and shear cracks in the reinforced concrete columns, as shown in Fig. 2c.

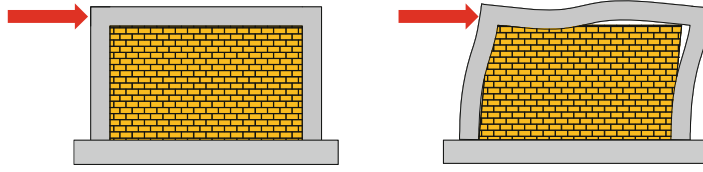
This entry is aimed to provide an overview of analytical tools that can be used to study the behavior of masonry-infilled RC frames under earthquake loading. It is not intended to be an exhaustive summary of the literature; rather, it will focus on some of the most common and representative analysis methods developed for such structures including their advantages and limitations. In this respect, two types of analysis methods will be considered: (i) simplified, design-oriented analysis tools, and (ii) refined tools based on the finite element method.

Simplified Modeling Using Equivalent Strut Concept

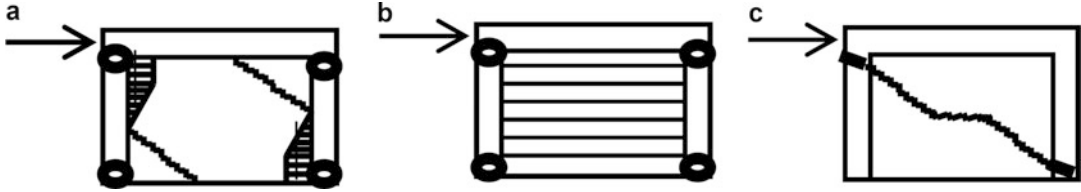
The vast majority of the simplified analysis methods proposed for infilled frames are based on the *equivalent strut* concept, with which the effect of the infill wall is modeled with diagonal struts. An example of this is shown in Fig. 3. The use of diagonal struts can approximately reproduce the frame–infill contact condition and the stress field in the infill walls. Since earthquake ground motions introduce cyclic lateral loading to infilled frames, at least two diagonal truss elements (one for each loading direction) are required to model an infill wall.

Strut Calibration

The use of a single strut for each direction of loading is expected to be adequate when the infill wall is relatively weak as compared to the frame so that failure is expected to occur in the infill

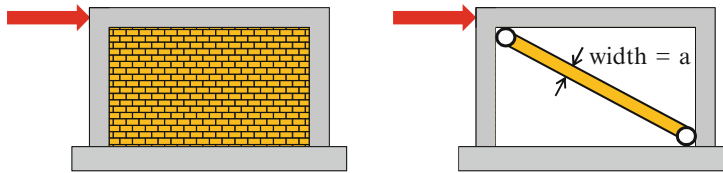


Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings, Fig. 1 Deformation of infilled frame subjected to a horizontal force



Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings, Fig. 2 Typical damage patterns for infilled frames (Mehrabi et al. 1994). (a)

Corner crushing (with flexural hinges in columns). (b) Sliding along bed joints (with flexural hinges in columns). (c) Diagonal/sliding cracks (with shear cracks in columns)



Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings, Fig. 3 Substitution of infill wall with a diagonal strut member for monotonic loading

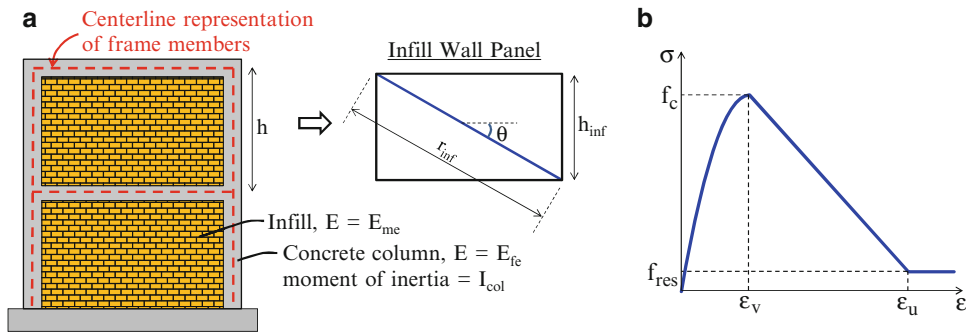
wall. The width of the equivalent strut depends on the contact length between the infill wall and the adjacent columns, which, in turn, depends on the stiffness of the infill as compared to that of the columns (Stafford Smith 1967; Mainstone 1972). However, the contact length between the infill and the columns and, thereby, the width of the equivalent strut are expected to change as the load experienced by an infilled frame increases and inelastic behavior develops in the masonry wall. Hence, strictly speaking, the effective strut width determined to represent the stiffness of an infilled frame will not be the same as that required to calculate the strength. For determining the lateral stiffness of an infilled frame, ASCE/SEI 41 (ASCE/SEI 2007) recommends the following expression based on the work of Mainstone (1972):

$$a = 0.175(\lambda_1 \cdot h)^{-0.4} \cdot r_{\text{inf}} \quad (1)$$

where h is the height of the column in a centerline representation of the frame geometry, as shown in Fig. 4a, r_{inf} is the length of the diagonal of the infill wall, and $(\lambda_1 \cdot h)$ is a dimensionless parameter representing a relative stiffness coefficient for the masonry infill and the frame, with λ_1 given by the following expression:

$$\lambda_1 = \left[\frac{E_{\text{me}} \cdot t_{\text{inf}} \cdot \sin(2\theta)}{4E_{\text{fc}} \cdot I_{\text{col}} \cdot h_{\text{inf}}} \right]^{\frac{1}{4}} \quad (2)$$

In Eq. 2, E_{me} is the modulus of elasticity of the masonry, t_{inf} is the thickness of the infill wall, E_{fc} is the modulus of elasticity of the concrete in the frame, I_{col} is the moment of inertia of the cross



Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings, Fig. 4 Material model and parameters for an equivalent strut. (a) Infilled frame properties and dimensions. (b) Uniaxial model for masonry

section of the frame columns, and h_{inf} is the height of the infill wall. A number of other expressions have been proposed in the literature to determine the strut width, as summarized in Asteris et al. (2011). The cross-sectional area of the diagonal strut is the product of the width, a , and the thickness of the infill, t_{inf} .

Mainstone's work suggested different effective strut widths for strength and stiffness for the reason mentioned above. However, for computer-based models, it is convenient to have a constant strut width. To this end, one can determine the effective compressive strength of a diagonal strut rather than the actual compressive strength of the masonry. Unfortunately, no general guidelines are available for this purpose. It is not difficult to convince oneself that this strength depends on the failure mechanism of the infill wall, which could be governed by the sliding of the bed joints, corner crushing, or diagonal/sliding failure. The guidelines in ASCE 41 recommend that the resistance of an infill wall be equal to the product of the shear strength of the masonry bed joints and the cross-sectional area of the wall. A different set of guidelines on determining the effective strut width and the shear strength of masonry infill is given in Appendix B of the MSJC code (2011). However, it should be mentioned that the latter is more for design and for performance assessment.

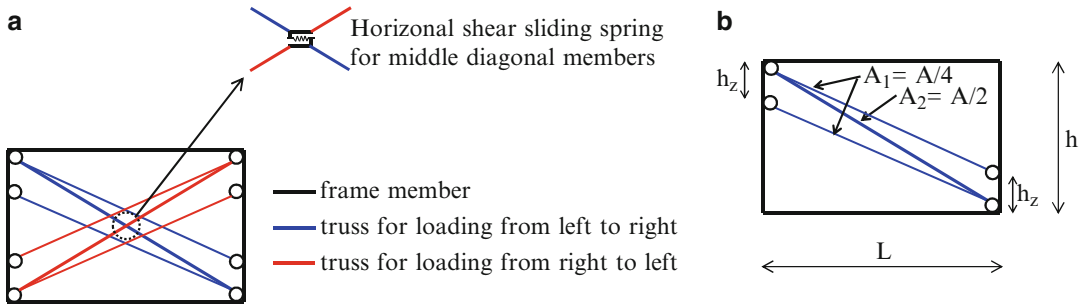
Klingner and Bertero (1976) used strut-based models to simulate the behavior of infilled frames under cyclic loading. They considered

infill walls constructed of reinforced masonry. They established a nonlinear force–deformation law for the struts to capture the strength degradation due to masonry crushing, the tensile resistance contributed by reinforcing steel in the wall, and the stiffness degradation of the infill wall due to damage. They used the prism compressive strength for the determination of the strut peak resistance, assuming that the area of the diagonal struts remains constant. Their analytical models provided satisfactory estimates of the cyclic response of experimentally tested infilled frames.

Ideally, two struts should be used to represent an unreinforced masonry infill wall, with each strut carrying only compression. A material model that can be used for this purpose is the Concrete01 material in OpenSees (McKenna et al. 2000), whose compressive stress–strain law is shown in Fig. 4b. The peak strength of the material model can be calibrated such that the horizontal component of the strut force is equal to a target resistance (e.g., the value stipulated in ASCE 41), $V_{u, inf}$, using the following equation:

$$f'_m = \frac{V_{u, inf}}{a \cdot t_{inf} \cdot \cos \theta} \quad (3)$$

Other stress–strain laws for infills that have been proposed in the literature (e.g., Crisafulli and Carr 2007) can also be used. A study by El-Dakhakni et al. (2003) has suggested that the



Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings, Fig. 5 Multiple diagonal strut model of an infilled frame under lateral loading. (a) Modeling approach. (b) Distribution of total area into multiple struts

material anisotropy of masonry should be accounted for when determining the stress–strain law for the equivalent strut. However, there has been no systematic study to determine whether accounting for the anisotropy will significantly increase the accuracy of an equivalent strut analogy.

The equivalent strut method is an oversimplification of the actual behavior of an infill wall and fails to capture some key failure mechanisms, such as the one depicted in Fig. 2c. A strut model will not account for the possible shear failure of a column that could be induced by the frame–wall interaction. There is no simple solution to overcome this problem. A study by Stavridis (2009) based on detailed nonlinear finite element models has demonstrated that the compressive stress field in a masonry infill wall may not be accurately represented by a single diagonal strut and that a strut model ignores the shear transfer between the beam and the infill. Hence, replacing a wall by a diagonal strut will not lead to a realistic representation of the load transfer from the frame to the wall. Moreover, as mentioned previously, it is not possible to have a single strut width to capture both the initial stiffness and load capacity of an infilled frame.

To represent the load transfer mechanism in a more accurate manner, multi-strut approaches have been proposed in a number of studies (El-Dakhakni et al. 2003; Crisafulli and Carr 2007). Crisafulli and Carr (2007) have proposed a multi-strut approach as shown in Fig. 5. The figure shows that several of the struts are

connected to the columns with an eccentricity, h_z , which is to be calculated using the following equation:

$$h_z = \kappa \cdot \frac{\pi}{2\lambda_1} \quad (4)$$

where κ is a constant which can take a value between 0.33 and 0.5.

The total strut area, A , needs to be distributed among the various truss elements as shown in Fig. 5b. The middle truss element is assigned half of the cross-sectional area of the equivalent strut, and each of the remaining two struts are assigned one fourth of the cross-sectional area. To account for the possibility for sliding along the bed joints, the middle strut is subdivided into two members, which are connected by a horizontal sliding “spring” and also by a vertical spring with large stiffness to enforce the continuity of vertical displacement at the location of the sliding connections. The sliding spring must have a large, penalty stiffness, and the frictional resistance can be estimated based on the amount of normal (gravity) stresses carried by the infill wall before the application of lateral loads.

The multi-strut model described above has been implemented by Crisafulli and Carr (2007) in a panel element, which only has four corner nodes and can be connected to the frame only at the points representing the beam–column joints. Thus, the effect of the frame–infill contact forces on the bending moments of the frame members

cannot be captured by a panel element. For this reason, it seems preferable to avoid the use of a panel element and simply connect the multiple truss elements to a model representing the RC frame. However, this will necessarily require that each column be represented by two beam–column elements.

While a multi-strut approach may better represent the load transfer mechanism in an infilled frame, it is not certain that it will lead to a significant improvement. A validation analysis by Crisafulli (1997) has shown that excellent results can be obtained by the proposed multi-strut; however, the same reference reported that a very careful adjustment of the properties of the model was required to achieve a good agreement. Details of this adjustment were not discussed. In view of the approximate nature of the equivalent strut approach, the additional complications introduced by a multi-strut model may not warrant such efforts.

In view of the aforementioned issues, a reasonable approach is to treat diagonal struts as purely phenomenological models. They can be calibrated in such a way that they represent not only the behavior of infill walls but that of an infilled frame as a whole. Such calibration can rely on experimental data, or in the absence of experimental data, on refined finite element models. Stavridis (2009) has used experimental data and finite element analysis results to derive a set of simple rules to determine ASCE 41-type pushover curves for infilled frames. Such a curve can be used to determine the load–displacement relation for an equivalent diagonal strut so that the overall load–displacement relation of an infilled frame can be captured. However, that study focused on non-ductile RC frames with relatively strong solid brick infill. More studies are needed to develop pushover curves for other infilled frame configurations.

Modeling of RC Frame Members

The RC frame members in a simplified analysis can be modeled using nonlinear beam elements. A variety of force-based and displacement-based beam elements, with lumped plasticity (having inelastic deformations only at end plastic hinges)

or with distributed plasticity, are available in analysis programs (Filippou and Fenves 2004). Many options also exist for modeling the cross-sectional behavior of the beam elements, which relates the stress resultants, namely, the axial forces and bending moments, to the corresponding generalized strains, i.e., the axial strain along the reference axis and the curvature. The most accurate and efficient formulation is the one based on the discretization of the cross section into fibers of concrete and steel reinforcement, with each fiber having an appropriate uniaxial constitutive law.

Shear failure can occur in the columns due to the forces developed from the frame–infill interaction, especially for non-ductile frames with strong masonry infill. If column shear failure is deemed probable, it can be accounted for in the analytical models through, e.g., nonlinear springs representing the shear force – shear deformation relation for the columns. However, this approach must be used with caution, because a strut model neglects many important aspects of the frame–infill interaction, such as the frictional shear transfer along the beam–infill interface and the variation in the axial forces of the columns due to the friction along the column–infill interfaces (Shing and Stavridis 2014). The introduction of shear springs in the columns may lead to unexpected results.

Infilled Frames Under Combined In- and Out-of-Plane Loading

The simplified strut-modeling concept can be extended to the analysis of infilled frames subjected to combined in- and out-of-plane loading. The simplest possible approach is to reduce the in-plane resistance of the strut elements using a reduction coefficient accounting for the out-of-plane force using the following expression (Al-Chaar 2002):

$$R_{i-o} = 1 + \frac{1}{4} \frac{OP_{\text{demand}}}{OP_{\text{capacity}}} - \frac{5}{4} \left(\frac{OP_{\text{demand}}}{OP_{\text{capacity}}} \right)^2 \quad (4)$$

where OP_{demand} is the applied out-of-plane pressure and OP_{capacity} is the out-of-plane capacity of

Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings, Table 1 Values of coefficient λ_2 for determination of out-of-plane capacity of infill walls according to ASCE 41

h_{inf}/t_{inf}	5	10	15	20
λ_2	0.129	0.060	0.034	0.013

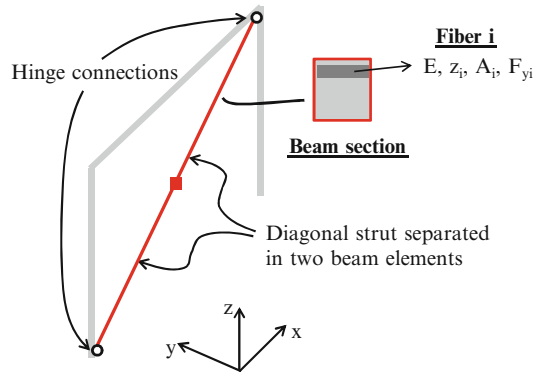
the wall, which can be estimated using the following relation proposed in ASCE 41:

$$OP_{capacity} = \frac{0.7f'_m \lambda_2}{h_{inf}/t_{inf}} \quad (5)$$

where λ_2 depends on the height-to-thickness ratio of the infill wall in accordance with Table 1. Equation (5) can only be used if several criteria, which allow the development of arching action in the infill walls, are satisfied. More specifically, the infill wall must be in full contact with the surrounding frame, the ratio h_{inf}/t_{inf} must not exceed 25 and the frame members must be sufficiently stiff and strong to allow the development of the thrusts from arching action. While this approach is conceptually simple, it requires the estimation of the out-of-plane load demand prior to the analysis.

The interaction of in-plane and out-of-plane loading can also be captured by accounting for the out-of-plane flexure of the masonry infill walls in strut-based models. A method proposed by Kadysiewski and Mosalam (2009) uses beam elements to account for the interaction of the axial force and out-of-plane bending. As shown in Fig. 6, each infill wall is represented with a diagonal beam member whose material can develop both tensile and compressive resistance. The member is subdivided into two beam elements which use a fiber section model. Each fiber in the section has the same elastic modulus, but the strength, sectional area, and distance of each fiber from the reference axis of the beam need to be determined so that the simplified model provides the in- and out-of-plane strength values established in ASCE 41 and can also reproduce a target interaction relation between in- and out-of plane resistance.

While the method by Kadysiewski and Mosalam (2009) provides a reasonable generalization of the ASCE 41 recommendations, it still has some issues. First of all, the determination of



Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings, Fig. 6 Modeling approach for infilled frame under combined in- and out-of-plane loading (Kadysiewski and Mosalam 2009)

the properties of each fiber is an underconstrained problem and has no unique solution. While Kadysiewski and Mosalam have proposed a procedure to circumvent this difficulty, there is no sufficient justification for the method they are proposing. Furthermore, other issues exist with the specific model pertaining to the behavior after yielding and also when collapse is expected to occur, because the model cannot capture the strength degradation occurring at the collapse limit state. Thus, further research is required to ensure a more sound calibration process and also capture the strength degradation effect.

Infilled Frames with Wall Openings

The simplified strut approach based on the representation of infill walls with truss elements can also be used for the analysis of structures where the infill walls have openings. A relatively small opening will not have a significant effect on the behavior of an infilled frame, and for this reason strut-based models can still be used with minor modifications. Al-Chaar (2002) has recommended that the cross-sectional area of the diagonal struts be multiplied by a reduction factor, R_i , which accounts for the existence of openings and is given by:

$$R_i = 0.6 \left(\frac{A_{op}}{A_{inf}} \right)^2 - 1.6 \frac{A_{op}}{A_{inf}} + 1 \quad (5)$$

where A_{op} is the area of the openings in an infill wall and A_{inf} is the area of the infill wall assuming

that it has no openings. The above expression can be used for $A_{op} < 0.60A_{inf}$. If the area of the openings in a wall is greater than 60 % of A_{inf} , then the effect of the infill wall can be neglected. Stavridis (2009) has also proposed formulas to account for the effects of an opening on the initial stiffness and strength of an infilled frame.

The current state of knowledge does not allow the establishment of general guidelines for the strut-based modeling of infill walls with openings. Several efforts have been made, especially for infilled steel frames (e.g., Mosalam et al. 1998), but they are far from conclusive. This is the reason why modern evaluation documents such as ASCE/SEI 41 (2007) state that the use of strut-based models for the analysis of infill walls with openings requires judgment and should be conducted on a case-by-case basis. Additional studies are needed to establish appropriate guidelines for truss models that can be applied to perforated infill walls.

Refined Finite Element Models

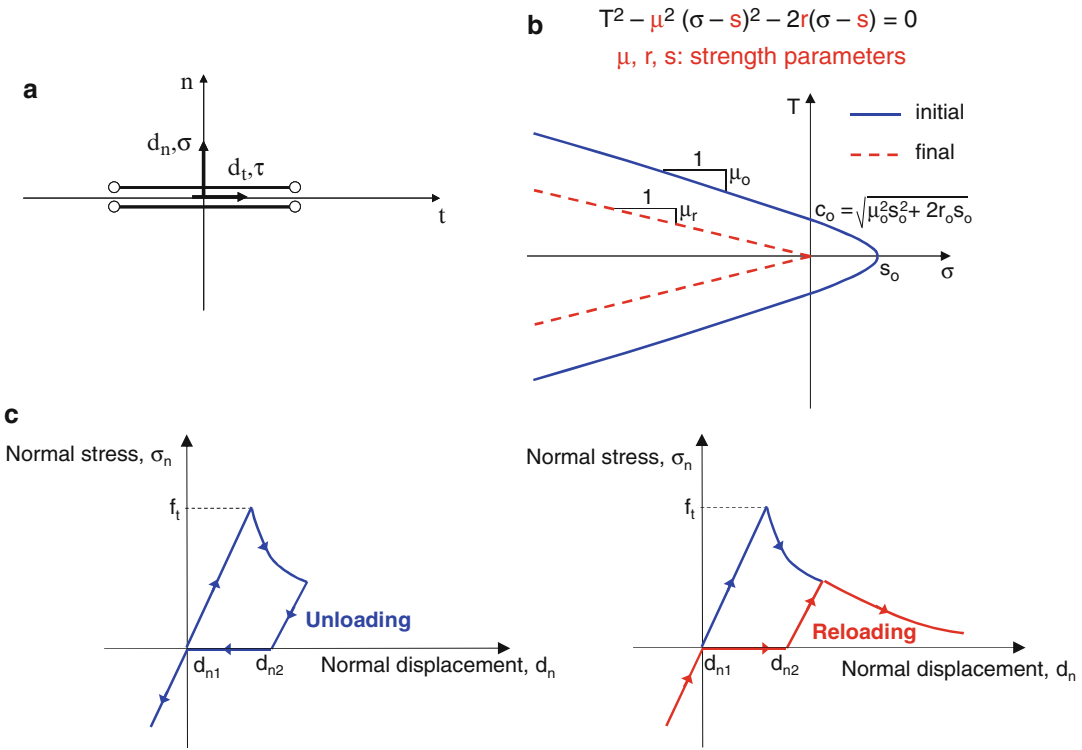
More refined models, based on the nonlinear finite element method, can be employed for the simulation of infilled frames (Mehrabi and Shing 1997; Hashemi and Mosalam 2007; Stavridis and Shing 2010; Koutromanos et al. 2011). The main advantage of nonlinear finite element analyses is that they can provide detailed information on the initial stiffness, stiffness and strength degradation, and damage pattern of a structure. The analysis of frames with openings in the infill walls poses no additional difficulty, and it can be conducted using the same types of elements and material laws as for solid infill walls.

Analysis for In-Plane Loading

For refined finite element analysis, the fracture behavior of concrete and masonry can be simulated with continuum elements based on the nonlinear fracture mechanics concept. In these elements, cracks are modeled in a smeared fashion, i.e., with a material stress–strain law representing distributed crack development in

a continuum rather than a traction–separation law for individual cracks. The constitutive models for these elements must account for the effect of cracking-induced damage and compressive crushing in concrete and masonry. Various formulations, based on plasticity, damage mechanics, or simplified nonlinear orthotropic laws, are available in a number of analysis programs. The reinforcing steel can be represented with truss elements using uniaxial constitutive laws. Appropriate interface or spring elements can be added in a model to capture the bond–slip behavior of the reinforcing steel; however, Mehrabi et al. (1994) have shown that the influence of the bond–slip effect is normally insignificant for infilled frames.

The use of continuum elements alone to model the behavior of the concrete and masonry materials is expected to provide accurate estimates of the response when the damage is dominated by the crushing of the infill. Special care is required for cases where cracks are dominated by mode II fracture or when cracks are not aligned with the element boundaries, which can be the case for diagonal/sliding cracks in infill walls or diagonal shear cracks in RC frame members. In such a case, cohesive crack interface elements must be added in the model to represent cracks in a discrete manner. These elements use “traction–separation” (stress–displacement) laws capable of describing the mixed-mode fracture behavior of cracks and mortar joints. Different elastic–plastic cohesive crack interface constitutive laws have been formulated and are available in analysis programs. A cohesive crack model formulated by Koutromanos and Shing is presented in Fig. 7. The displacement and stress vectors of the model include a normal and tangential (shear) component, as shown in Fig. 7a. The failure surface of the model, shown in Fig. 7b, is characterized by three key strength quantities, namely, the tensile strength, s ; the cohesive strength, c (which is the sliding resistance of an interface when there is a zero normal compressive stress); and the asymptotic frictional coefficient, μ . It represents a generalized Mohr–Coulomb law. The failure surface translates



Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings, Fig. 7 Formulation of discrete-cohesive crack interface element for finite

element analysis of infilled frames (Koutromanos and Shing 2012). (a) Interface element. (b) Failure surface. (c) Normal tensile unloading–reloading law

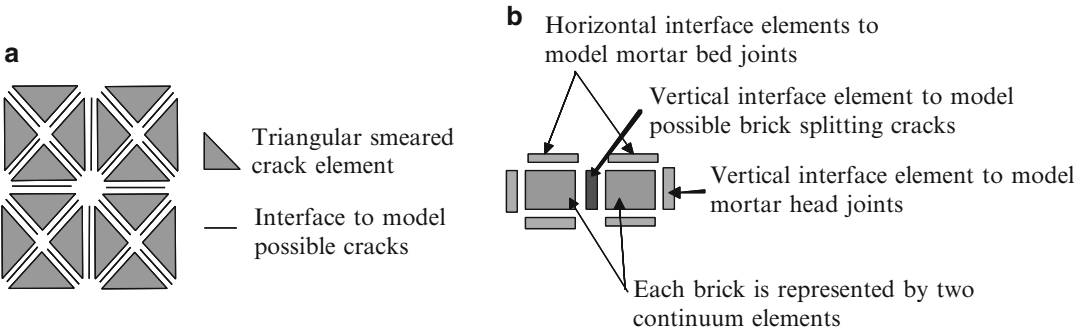
N

and shrinks when fracture occurs, so that the effect of strength degradation is captured. The hysteretic behavior of the cohesive element captures the effect of crack opening and closing as shown in Fig. 7c.

The contact condition between the frame members and the infill walls can also be captured in an analysis using interface elements. Alternatively, if the cohesive strength of the frame–infill interfaces is relatively small and if the frictional resistance along the interface can be entirely attributed to Coulomb friction, standard contact formulations included in many commercial programs can be used instead.

Stavridis and Shing (2010) have used the aforementioned scheme to model masonry-infilled RC frames. They have used both triangular smeared-crack elements and interface elements to simulate the behavior of concrete frame members, as shown in Fig. 8a, and quadrilateral

smeared-crack elements and interface elements to model the unreinforced masonry walls, as shown in Fig. 8b. The zero-thickness interface elements used for the mortar joints are only meant to capture the localized mixed-mode fracture along the brick–mortar interface and cannot account for compressive crushing in the mortar. If the compressive crushing of the mortar were to be accounted for, the interaction between the mortar joints and brick units would need to be explicitly accounted for in the analysis. This interaction can strengthen the mortar layer and weaken the brick units and result in a masonry assembly whose compressive strength is between those of the two constituents. Since this interaction is not captured with the use of the zero-thickness interface, it should be accounted for indirectly by adjusting the properties of the continuum elements in compression to represent the compressive behavior of the

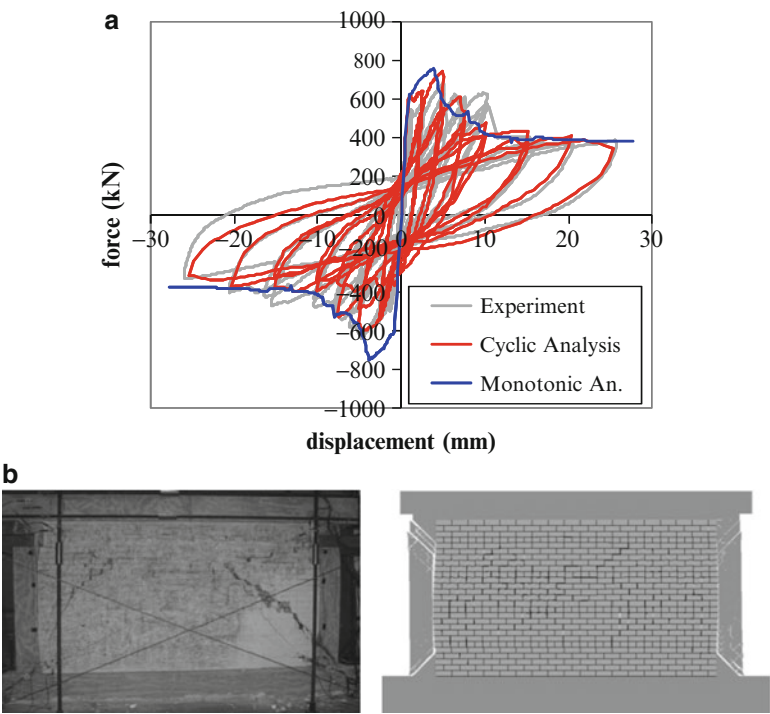


Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings, Fig. 8 Discretization scheme to capture strongly localized cracks in the refined

finite element analysis of infilled RC frames. **(a)** Reinforced concrete members. **(b)** Unreinforced masonry panels

Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings, Fig. 9

Comparison of refined finite element analysis with experimental tests on a single-story, single-bay infilled RC frame. **(a)** Load-displacement curves. **(b)** Crack patterns



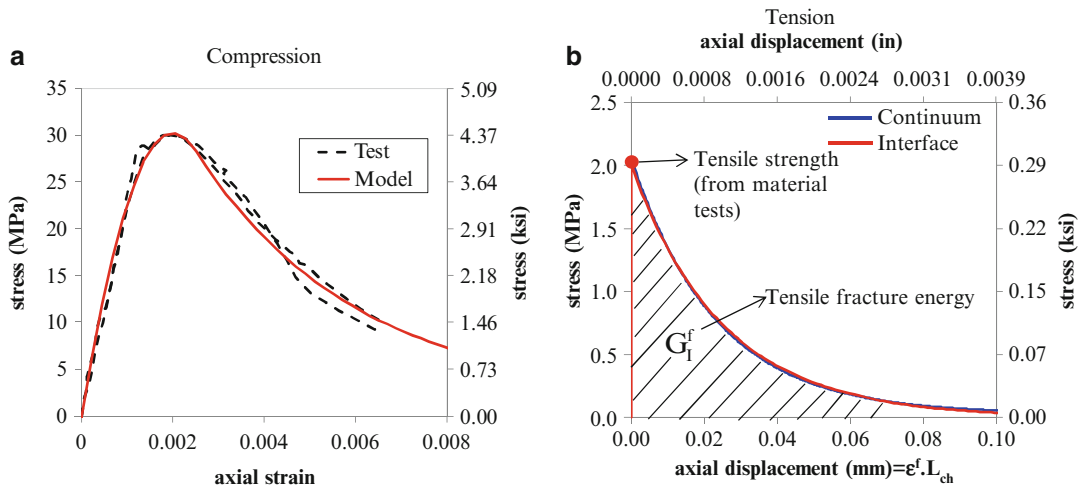
masonry assembly. For the RC columns, each reinforcing bar in the columns has been divided into multiple truss elements so that each discrete crack will cross the right quantity of reinforcement.

Koutromanos et al. (2011) have used aforementioned meshing scheme with novel constitutive models to successfully capture the global and local response and damage pattern of infilled

frames under static and dynamic loads. One example is shown in Fig. 9.

Analysis for Combined In- and Out-of-Plane Loading

Refined finite element analysis can also consider combined in- and out-of-plane loading. For this purpose, three-dimensional interface elements need to be used with three-dimensional



Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings, Fig. 10 Calibration of refined constitutive models for concrete. (a) Uniaxial compression (calibration with uniaxial tests on concrete

cylinders or masonry prisms). (b) Uniaxial tension (calibration with splitting tension tests on concrete cylinders or brick units)

continuum elements or shell elements with smeared-crack formulations. While the current state of the knowledge in material models and element formulations allows the use of three-dimensional finite element analysis for infilled frames under combined in- and out-of-plane loads, there have been relatively few such studies. Hashemi and Mosalam (2007) have used shell elements to model the infill walls subjected to combined in- and out-of-plane loading.

Calibration of Constitutive Models for Nonlinear Finite Element Analysis

Refined constitutive models for continuum elements and cohesive crack interface elements typically include many parameters, which require calibration with data from material tests. Data from uniaxial compression tests on concrete cylinders can be used for the calibration of the continuum material models for the concrete, as shown in Fig. 10a, while data from masonry prism tests (uniaxial compression and bond wrench) can be used for the calibration of the constitutive models for the masonry. Interface elements typically include parameters pertaining to mixed-mode fracture, and data from mixed-mode fracture tests on concrete or masonry mortar joints need to be used for the calibration (e.g.,

Hassanzadeh 1990). For mortar joints in masonry infill walls, the tensile (bond) strength typically ranges between 275 kPa (40 psi) and 690 kPa (100 psi), while the frictional coefficient ranges between 0.65 and 0.90.

An important aspect regarding the calibration of continuum elements that have material laws with strain softening is to avoid the spurious mesh-size sensitivity, which will lead to the loss of objectivity of the numerical results. This is caused by strain localization, as explained, e.g., in Bazant and Planas (1998). To remedy this problem, the softening portions of the stress-strain laws will require regularization, i.e., adjustment to account for the element size. In addition, one needs to consider the fact that in the meshing scheme shown in Fig. 8a, tensile cracking can occur in both the continuum elements and the discrete-cohesive crack interface elements. Obviously, the two types of elements should be calibrated to give identical tensile stress-versus-fracturing displacement behavior, as shown in Fig. 10b. For the continuum elements, the fracturing displacement is equal to the product of the inelastic strain, ϵ^f , times the characteristic element length, L_{ch} . The area under the stress-fracturing displacement curve is a material constant called the mode I fracture

Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings, Table 2 Values of constant G_{fo}^I for determination of tensile fracture energy of concrete (Fib 1999)

D_{max} , mm (in)	8 (0.31)	16 (0.63)	32 (1.26)
G_{fo}^I , N/mm (lb/in)	0.025 (0.14)	0.030 (0.17)	0.058 (0.33)

energy or tensile fracture energy, G_f^I . The following expression, which is proposed in FIB (1999), can be used for the determination of G_f^I :

$$\begin{aligned} G_f^I &= G_{fo}^I \left(\frac{f_{cm}}{f_{cmo}} \right)^{0.7} & \text{for } f_{cm} \leq 80 \text{ MPa} \\ G_f^I &= 4.3 G_{fo}^I & \text{for } f_{cm} > 80 \text{ MPa} \end{aligned} \quad (6)$$

where f_{cm} is the mean compressive strength of the concrete, f_{cmo} is equal to 10 MPa (1.46 ksi), and G_{fo}^I is a reference value of the fracture energy, which represents the fracture energy for concrete with a compressive strength equal to 10 MPa, and it depends on the maximum aggregate size, d_{max} , as shown in Table 2. The fracture energy for the masonry units and for the mortar joints can be calibrated using data from experimental tests (e.g., van De Pluijm 1997).

Often, there may not be sufficient material test data to calibrate all the material parameters in a finite element model. In such cases, a sensitivity analysis is required to determine the sensitivity of the numerical results to the values of parameters that cannot be determined from material test data. A parametric study by Stavridis and Shing (2010) with finite element models has indicated that the shear strength parameters for mortar joints are most influential on the load–displacement response of an infilled frame.

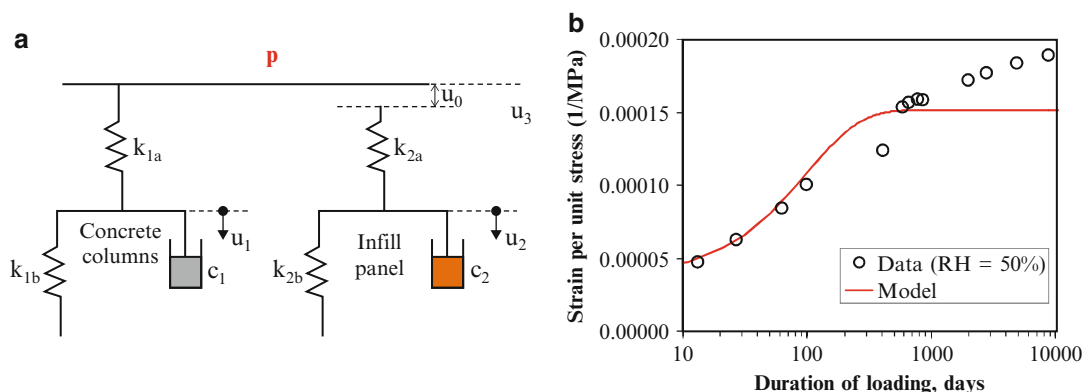
Determination of Gravity Load Distribution Between Frame Columns and Infill Walls

The behavior of an infill wall strongly depends on the compressive stress in the bed joints. An increased gravity load will increase compressive stress and, thereby, the resistance of the bed

joints, thus increasing the stiffness and strength of a wall. An accurate estimate of the gravity load distribution between the infill wall and the surrounding frame is necessary to ensure that the analysis provides accurate estimates of the strength and stiffness of the system. The most straightforward approach to estimate the gravity load distribution for existing structures is to use in situ tests with flat-jacks. Since it may not always be feasible to conduct such tests, analytical models may be used to estimate the fraction of gravity loads carried by the frame columns and infill walls, respectively.

The analytical determination of the gravity load distribution between the frame and the infill wall requires several considerations. First, a part of the gravity loads may be applied onto the RC columns before the construction of the infill walls because these walls could be constructed after the frame has been completed. Second, long-term effects such as concrete and masonry creep, concrete shrinkage, and brick masonry expansion with time due to water absorption can significantly affect the gravity load distribution. While refined finite element models with viscoelastic material properties can be employed for the determination of the gravity load distribution, the increased computational burden of such analyses may not necessarily produce results of increased accuracy due to lack of experimental data to allow the calibration of multiaxial viscoelastic constitutive models.

Based on the above consideration, simplified models are deemed preferable for the determination of the gravity load distribution, since they are easy to calibrate and their reliability is not necessarily inferior to that of more refined models. Such a simplified physical model is shown in Fig. 11a. The model consists of springs and dashpots to represent the instantaneous axial stiffness and the viscoelastic (creep) properties of the RC columns and the masonry infill. The “gap” u_o , shown in Fig. 11a, corresponds to the short-term deformation of the columns due to the gravity loads that are applied before the construction of the infill walls. The spring-and-dashpot assemblages representing the concrete columns and masonry walls can be calibrated with data from



Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings, Fig. 11 Simplified physical model for the estimation of the gravity load

creep tests on concrete and masonry; Fig. 11b shows an example of the calibration of a creep model for concrete. The effect of the brick expansion can also be easily added in the model. A detailed explanation of the simplified modeling approach and of how the gravity load distribution can be correctly modeled in a finite element model is given in Koutromanos (2011).

Summary: Concluding Remarks

The modeling of infilled frames is a challenging task because a variety of failure mechanisms may affect the load-resistance properties of the system. The analysis can be conducted using simplified models based on the equivalent strut concept or refined finite element models using appropriate constitutive models for the materials. Strut-based models are conceptually simple and easy to calibrate and implement, but their accuracy is inherently limited, especially if walls with openings are to be analyzed. It is not certain whether the use of complicated multi-strut models is meaningful, because such models are bound to misrepresent some important mechanisms that can develop in infilled frames. On the other hand, refined finite element models are more general and realistic, but the efforts on the calibration of the constitutive models and preparation of such analyses are much more significant. Furthermore, refined finite element analyses are computationally

demanding. Thus, the selection of the modeling approach depends on the desired level of accuracy, the expertise of the analyst, and the time and computational resources available for the analysis.

References

- Al-Chaar G (2002) Evaluating strength and stiffness of unreinforced masonry infill structures. Research report, US Army Corps of Engineering
- ASCE/SEI (2007) Seismic rehabilitation of existing buildings, ASCE/SEI 41–06. American Society of Civil Engineers
- Asteris PG, Antoniou ST, Sofianopoulos DS, Chrysostomou CZ (2011) Mathematical macro-modeling of infilled frames: state-of-the-art. ASCE J Struct Eng 137(12):1508–1517
- Bazant Z, Planas H (1998) Fracture and size effect in concrete and other quasibrittle materials. CRC Press, Boca Raton, 640 p
- Crisafulli FJ (1997) Seismic behaviour of reinforced concrete structures with masonry infills. PhD dissertation, University of Canterbury, Christchurch, New Zealand. Available online at <http://ir.canterbury.ac.nz/handle/10092/1221>
- Crisafulli FJ, Carr AJ (2007) Proposed macro-model for the analysis of infilled frame structures. Bull NZ Soc Earthq Eng 40(2):69–77
- El-Dakhakni WW, Elgaaly M, Hamid AA (2003) Three-strut model for concrete masonry-infilled steel frames. ASCE J Struct Eng 129(2):177–185
- Federation Internationale Beton du (1999) Structural concrete: textbook on behaviour, design and performance, vol. 1: Introduction – design process – materials. International Federation for Structural Concrete, Lausanne, 224 p

- Filippou FC, Fenves GL (2004) Methods of analysis for earthquake-resistant structures. In: Bozorgnia Y, Bertero VV (eds) *Earthquake engineering – from engineering seismology to performance-based earthquake engineering*. CRC Press, Boca Raton
- Hashemi A, Mosalam KM (2007) Seismic evaluation of reinforced concrete buildings including effects of masonry infill walls. In: Report PEER 2007/100. Pacific Earthquake Engineering Research Center, Berkeley
- Hassanzadeh M (1990) Determination of fracture zone properties in mixed mode I and II. *Eng Fract Mech* 35(4/5):845–853
- Kadysiewski S, Mosalam KM (2009) Modeling of unreinforced masonry infill walls. Considering in-plane and out-of-plane interaction. In: Report PEER 2008/102. Pacific Earthquake Engineering Research Center, Berkeley
- Klingner RE, Bertero VV (1976) Infilled frames in earthquake-resistant construction. Report UCB/EERC-76/32, Earthquake Engineering Research Center, University of California, Berkeley. Available online at <http://nisee.berkeley.edu/elibrary/Text/61000515>
- Koutromanos I (2011) Numerical analysis of masonry-infilled reinforced concrete frames subjected to seismic loads and experimental evaluation of retrofit techniques. PhD dissertation, University of California, San Diego, La Jolla. Available online at http://nees.org/resources/3579/download/Ioannis_Koutromanos-PhD_Dissertation.pdf
- Koutromanos I, Shing PB (2012) A cohesive crack model to simulate cyclic response of concrete and masonry structures. *ACI Structural Journal* 109: 349–358
- Koutromanos I, Stavridis A, Shing PB, Willam K (2011) Numerical modeling of masonry-infilled RC frames subjected to seismic loads. *Comput Struct* 89(3–4): 1026–1037
- Mainstone B (1972) On the stiffnesses and strengths of infilled frames. CP 2/72, Garston [Eng.], Gt. Brit. Building Research [Establishment] Current paper, 90 p
- Masonry Standards Joint Committee (2011) TMS 402-11/ACI 530-11/ASCE 5–11: building code requirements and specification for masonry structures, The Masonry Society, American Concrete Institute, and American Society of Civil Engineers, Boulder, CO, Farmington Hills, MI, and Reston, VA
- McKenna F, Fenves GL, Scott MH, Jeremic B (2000) Open system for earthquake engineering simulation. <http://opensees.berkeley.edu>
- Mehrabi AB, Shing PB (1997) Finite element modeling of masonry-infilled RC frames. *ASCE J Struct Eng* 123(5):604–613
- Mehrabi AB, Shing PB, Schuller MP, Noland JL (1994) Performance of masonry-infilled R/C frames under in-plane lateral loads. In: Report CU/SR-94/6. Department of civil, environmental & architectural engineering, University of Colorado at Boulder
- Mosalam KM, White RN, Ayala G (1998) Response of infilled frames using pseudodynamic experimentation. *Earthq Eng Struct Dyn* 27(6):589–608
- Polyakov SV (1960) On the interaction between masonry filler walls and enclosing frame when loaded in the plane of the wall. In: Polyakov SV (ed) *Translation in earthquake engineering*. EERI, San Francisco, pp 36–42
- Shing PB, Stavridis A (2014) Analysis of seismic response masonry-infilled RC frames through collapse. *ACI Special Publication* 297
- Stafford Smith B (1967) Methods for predicting the lateral stiffness and strength of multi-storey infilled frames. *Build Sci* 2:247–257
- Stavridis A (2009) Analytical and experimental study of seismic performance of reinforced concrete frames infilled with masonry walls. PhD dissertation, University of California, San Diego
- Stavridis A, Shing PB (2010) Finite element modeling of nonlinear behavior of masonry-infilled RC frames. *ASCE J Struct Eng* 136(3):285–296
- Van der Pluijm R (1997) Non-linear behaviour of masonry under tension. *Heron* 42(1):25–54

Ocean-Bottom Seismometer

Mechita C. Schmidt-Aursch¹ and
Wayne C. Crawford²

¹Alfred-Wegener-Institut, Helmholtz-Zentrum
für Polar- und Meeresforschung, Bremerhaven,
Germany

²Institut de Physique du Globe de Paris, Sorbonne
Paris Cité, Univ Paris Diderot, UMR 7154
CNRS, Paris, France

Synonyms

OBS

Introduction

About 70 % of the world is covered by oceans. Because of the difficulty of accessing the ocean floor, most of the seafloor and the crust below was unexplored for a long time. In the early 1930s, the seismic refraction method was developed and geoscientists tried to develop techniques to use this method offshore. They experimented with cabled sources and geophones but also with free-fall instruments. The first layout of a stand-alone ocean-bottom seismometer (OBS) was published in 1938 (Ewing and Vine 1938) and tested in the years 1939–1940. This OBS used a gasoline-filled rubber balloon for buoyancy, which floats approx. 3 m above the seafloor.

An aluminum housing containing an automatic oscillograph was mounted below the balloon. The iron ballast and the external geophone for recording man-made explosive seismic signals were located on the ocean bottom (Ewing et al. 1946).

After these first experiments, there were only intermittent OBS deployments until the late 1950s and early 1960s, where seismic monitoring of nuclear explosions became suddenly important. For this purpose, a uniform monitoring station distribution around the globe was desired, so seismic stations in the oceans were necessary. Programs like the “Vela Uniform project” promoted advancements in OBS technology (VESIAC 1965). Since then, the usage of OBS for passive earthquake recording and active-source experiments has become more and more common, as many scientific targets are located offshore, including continental margins, mid-ocean ridges, gas and oil reservoirs, and potential tsunami-generating areas.

Today, a great variety of OBS exist for different scientific purposes: short-period instruments mainly for active-source experiments, broadband seismometers for passive earthquake recording, small-sized OBS for short-term deployments, and larger platforms with independent operating times of 1 year or more. Technology developed rapidly since the first OBS deployments, but the principle remains the same: an instrument carrier, equipped with one or more sensors, a data logger with batteries, and a release unit, is weighted with some ballast anchor and sinks freely down to the



Ocean-Bottom Seismometer, Fig. 1 Stacked OBS frames with floatation consisting of glass spheres (*left*) and syntactic foam (*right*) (Photographs W. Crawford, Paris (*left*) and M. Schmidt-Aursch, Bremerhaven (*right*))

seafloor to record autonomously natural or man-made seismic signals. After a certain time or after receiving a hydroacoustic signal, a release unit detaches the anchor and the OBS ascends to the sea surface using some kind of buoyancy. The instrument can then be recovered to retrieve the data for further processing and interpretation.

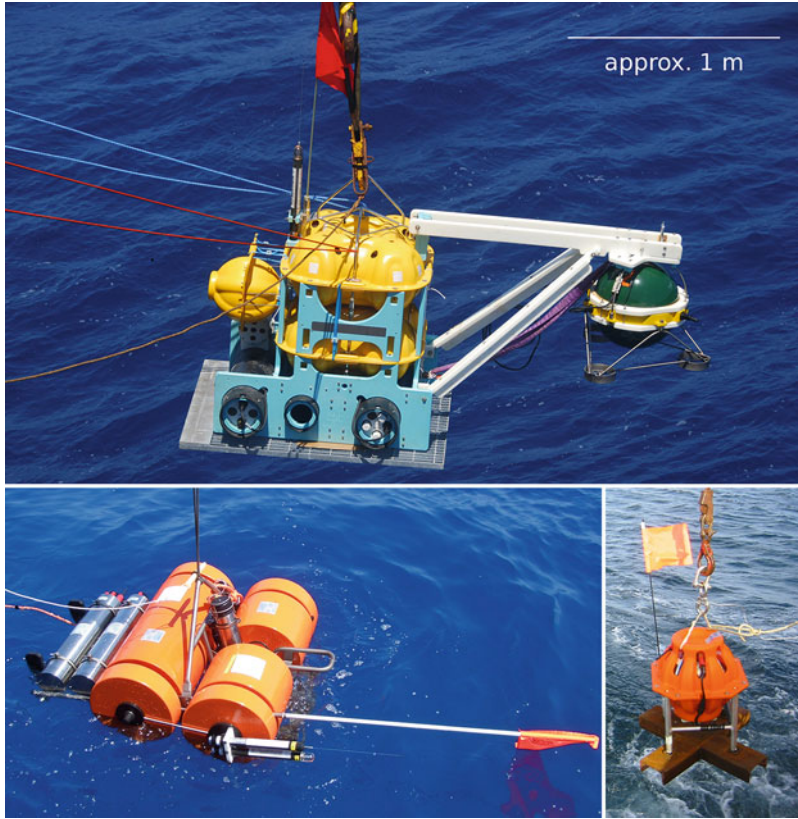
Besides these free-fall and pop-up ocean-bottom seismometers, there also exist systems with real-time data transfer using moorings or cables. These are sometimes also called ocean-bottom seismometers, although they belong to the category of ocean-bottom observatories. This entry gives an overview of the principles of self-contained OBS without cable connection or any other telemetry, introduces the state of the art of the main component technologies, and explains the special demands of OBS data processing.

General Hardware Requirements

Ocean-bottom seismometers are deployed in a rather hostile environment: the seafloor. Ambient temperatures are low, the ambient pressure is high, and water salinity causes metal parts to

corrode. The seismometer should be well coupled to the seafloor and withstand water currents, and the instrument should cause as little noise on the seismometer as possible. The stations might face high accelerations and shocks during handling on the ship's deck, during deployment and recovery, and on touchdown at the seafloor. An OBS must therefore be mechanically robust and solidly built. Pressure casings for water depths of several thousand meters – a water pressure of tens of MPa – must have thick and homogenous walls, and the floatation must be highly incompressible. To resist shocks, the frame, buoyancy, and all other parts must be of rugged design.

On the other hand, the instruments should be as compact as possible, because transport costs are high and deck space is limited on most vessels. Most OBS designs allow stacking of the frames and the floatation (Fig. 1) and storing the other parts in standard transport boxes. As ship time is limited and expensive, the instruments should be easy to mount so that a large number of units can be prepared in a short time. The mounted devices must be able to be handled on deck and by crane during bad weather conditions. For recovery, easy sighting of the instruments in the waves and a safe pick up from the sea surface is desirable. There is no ideal OBS design to



Ocean-Bottom Seismometer, Fig. 2 Examples of ocean-bottom seismometers. *Top:* OBS with a mechanically isolated broadband seismometer and differential pressure gauge for long-term deployments. Floatation: glass spheres; frame: high-density polypropylene (PEHD); pressure tubes: anodized aluminum (Photograph W. Crawford, Paris). *Bottom left:* OBS with integrated wideband seismometer and hydrophone for

long-term deployments. Floatation: syntactic foam; frame and pressure tubes: titanium alloy (Photograph M. Schmidt-Aursch, Bremerhaven). *Bottom right:* compact OBS with integrated short-period seismometer and hydrophone for short-term deployments. Floatation and housing for seismometer; data logger; batteries; strobe and radio beacon: glass sphere; frame: aluminum (Photograph courtesy of F. Klingelhöfer, IFREMER, Brest)

satisfy these specifications; therefore, various designs have been developed that are optimized for specific purposes, like especially compact and light-weighted OBS for active-source experiments, using large numbers of instruments or particularly solid units with larger internal volumes for long-term deployments (Fig. 2).

The harsh conditions offshore also require enhanced specifications for the electronics. All electronic parts must be shock resistant: this is a challenge for sensitive instruments like broadband seismometers or the most common data storage devices. Cables and connectors outside of the pressure tubes must be watertight, saltwater resistant, and designed for high pressures.

Placed on deck of a ship, OBS are exposed to a wide temperature range as the vessel might operate in the sunlit tropics or in polar regions where even the water temperatures are below zero. Therefore, the electronics must be able to operate over a large temperature range. Batteries are so far the only power supply at the seafloor for untethered systems; hence, all components must have low power consumption. In contrast to onshore stations, limited or no maintenance is possible on the ocean bottom. Most OBS can receive some remote-control commands, but for the acoustic link, the presence of a vessel is necessary. The instruments must therefore be highly reliable during long-term deployments.

Instrument Carrier

Frame and Floatation

The base of every OBS is some kind of frame, on which all the other components are mounted. The frame must resist high pressures at the deep sea and the chemically aggressive subsea environment. It must be rigid and stable enough to carry all attaching parts, but on the other hand, it should be as compact and easy to handle as possible. These structures are mostly built of aluminum or titanium alloy or some synthetic material like, e.g., polypropylene. Aluminum alloy is lightweight and cheap, but not as resistant against corrosion as the more expensive titanium alloy. Synthetic materials are rust-free but the frames need larger diameters or thicknesses to reach the same mechanical strength as the slim metal carriers. Figures 2 and 3 present three different OBS types with structures built of aluminum, high-density polypropylene, and titanium alloy.

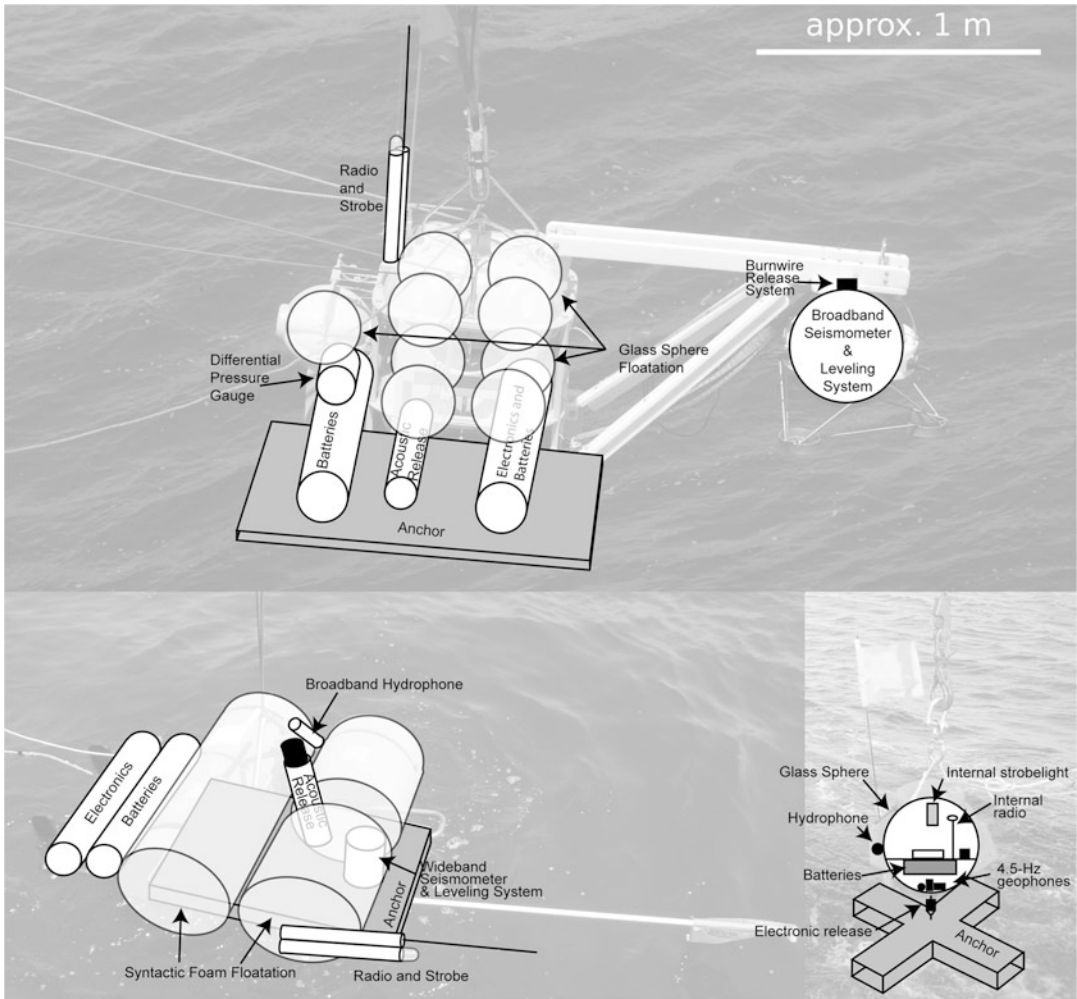
The frame needs some buoyancy to ascend to the surface after termination of the seismic measurements. The floatation must withstand very high pressures up to tens of MPa in a couple of thousand-meter water depth. The first OBS prototypes used oil- or gasoline-filled rubber balloons for floatation. The densities of oil and gasoline are lower than the density of water, and in addition, oil and gasoline are nearly incompressible. Therefore, the balloons kept their volume and hence their static buoyancy also in deep water. Soon the application of evacuated glass spheres as floatation was developed; their usage is still very common today. They are lightweight and can simultaneously be used as housing for seismometers and other electronic parts. Figures 2 and 3 (right) show a very compact short-period OBS mainly designed for active-source experiments, where the seismometer, data logger, batteries and recovery strobe light, and radio beacon are integrated into the buoyancy glass sphere. Larger OBS for long-term deployments need a couple of spheres to achieve enough buoyancy to carry two or more pressure cylinders full of electronics and batteries back to the surface (Figs. 2 and 3, top). Glass spheres are very shock sensitive; therefore, an additional plastic casing

(e.g., polyethylene) is necessary. If a glass sphere contains electronics, the casing must be perforated to allow access to connectors or to provide visibility of internal flash lights. Nevertheless, they risk implosion at the seafloor and hence the loss of the instrument.

Since the early 1960s, syntactic foam has also been used for floatation. Syntactic foam is a synthetic compound of hollow particles in a background matrix of ceramic, metal, or polymer. An epoxy matrix with hollow glass microballoons with diameters less than 1/100 mm is widely used for deep-sea floatation. This foam is highly incompressible, showing a small reduction in buoyancy only after several years of deep-sea deployments. Syntactic foam can easily be customized to various shapes like plates, blocks, or barrels, so customized OBS designs are possible (Figs. 2 and 3, bottom left). Floatation units made of syntactic foam are robust and easy to handle but much heavier than the glass spheres. The majority of OBS floatation is painted in signal colors to help in spotting the instrument after emerging at the sea surface.

Anchor Weight and Release Unit

An anchor weight is necessary for the OBS to sink to the seafloor. Most anchors are made of untreated metals like iron or steel; some models also use concrete blocks. The entire system of the frame, sensors and batteries, floatation, and anchor must be carefully balanced in air as well as in water. The weight or buoyancy is different in salt water, freshwater, and air, so a good calculation of buoyancy is necessary. The OBS should sit solidly on the deck, hang well balanced beneath a crane, and should sink and rise in the ocean without swinging or rotating. It must have enough weight and a low enough center of gravity in water to be stable at the seafloor, resisting bottom currents and providing a good coupling of the seismometer to the ground. On the other hand, the impact of the OBS on the sea bottom at deployment should be smooth to avoid damage, especially to the sensitive seismometer. If the instrument weighs too much in water or sinks too fast in the water column, it might sink deep enough into soft sediments to be stuck.



Ocean-Bottom Seismometer, Fig. 3 Sketch of ocean-bottom seismometers. *Top:* OBS with a mechanically isolated broadband seismometer and differential pressure gauge for long-term deployments. *Left:* OBS with integrated wideband seismometer and hydrophone for

long-term deployments. *Right:* Compact OBS with integrated short-period seismometer and hydrophone for short-term deployments (Drawing adapted from Auffret et al. (2004))

A central element of an OBS is the release unit, which connects the frame with the anchor weight. After a certain time (timed release) or on receiving an acoustic command (acoustic release), the release unit disconnects the anchor from the OBS and the OBS can rise freely to the sea surface. The ballast remains at the bottom where it degrades naturally. There are two main release mechanisms: a burn wire or a motor-driven turning hook. A motor-driven releaser clamps the anchor with a hook that is held by a

bolt (Fig. 4, right). The bolt turns mechanically, the hook is released, and finally the anchor is decoupled. A burn wire release holds the anchor using a wire that is partially exposed to seawater. A second larger wire or metal post is also exposed to seawater. When a positive voltage is applied between the wire and the post, electrolysis corrodes the wire, which then breaks, releasing the anchor (Fig. 4, left). The “burn” time depends on the length and thickness of the exposed wire: 10–15 min is typical for many systems.



Ocean-Bottom Seismometer, Fig. 4 Example of release units. *Left:* motor-driven release unit with integrated transducer (Photograph M. Schmidt-Aursch, Bremerhaven). *Right:* burn wire mechanism (Photograph courtesy of A. Ndiaye, INSU, Paris)

Most release units can also send back a response, allowing a two-way hydroacoustic link to the OBS, e.g., to determine the distance to the system, to ask for state of health, or to send simple commands. Some research vessels are equipped with permanently installed transducers, but more often, mobile transducer systems are used on board. For the mobile units, the ship has to stop and the transducer lowered into the water to achieve a coupling of the acoustic signal to the water column. The hydroacoustic signal uses frequencies between 5 kHz and 20 kHz with different modulation techniques. Normally, the signal is coded; each unit uses unique codes for release and other commands. Because the release unit is the central component to recall the OBS, its power supply is usually completely independent from the main system. Some models even use

independent batteries for communication and for activating the release.

Flag, Radio Beacon, and Flasher

After an OBS reaches the surface, the next challenge is to recover it. The OBS is small compared to the area of the ocean in which it may surface and local currents might make the OBS drift away from the deployment position on its way down and up through the water column. After emerging at the surface, currents and wind may also move the OBS away. Waves and decreased visibility due to rain or fog can hinder the search. Painting the OBS with a signal color (yellow, orange, or red) helps observers to spot the instruments. A flag can also help sighting OBS at large distances or in tall waves, but high-profile flags can also generate noise on the data, so not all OBS have flags.

However, almost all OBS are equipped with a VHF radio beacon and a Xenon flasher (Fig. 5, middle and left). The casing of the beacons is usually made of aluminum and the radio antenna is made of steel, so sacrificial anodes are necessary to avoid electrolytic corrosion of the aluminum. The radio beacon and strobe both contain a pressure switch that turns the units off at 1–10-m water depth to save battery power while the instruments are at the seafloor. The strobe light additionally uses a light sensor that only allows flashing in the dark, also for battery saving purposes. The flashers are especially useful during night, the light can be spotted over larger distances than the flags, but the estimation of the station distance is more difficult than during daylight.

Radio beacons are available with different frequencies or different transmitting cycles to distinguish individual OBS in case that several instruments are “on the air” at the same time. The majority of VHF beacons operate in nautical frequency bands, so besides handheld direction finders, also shipborne cross bearing receivers can be used to determine the direction of the radio signals. The signal range at the sea surface is quasi-optical, which means approximately to the optical horizon, from elevated locations or helicopters, the OBS can be located over

Ocean-Bottom Seismometer,

Fig. 5 Example of a Xenon flasher (*left*), VHF radio beacon (*middle*), and a satellite transmitter (*right*) (Photograph M. Schmidt-Aursch, Bremerhaven)



distances of several tens of kilometers. Some models integrate a GPS system and send position data using the “Automatic Identification System” (AIS) that is mandatory for many vessel classes. There are also beacons that send an identifying signal to a satellite after emerging at the sea surface (Fig. 5, right). The operator then receives an email containing the position of the OBS. This remote surveillance technique is especially useful, if the station rises ahead of time or is caught by a trawler, but the transponders are bulky and costly, require additional license fees, and are therefore not widespread.

Sensors and Data Logger

Seismometer

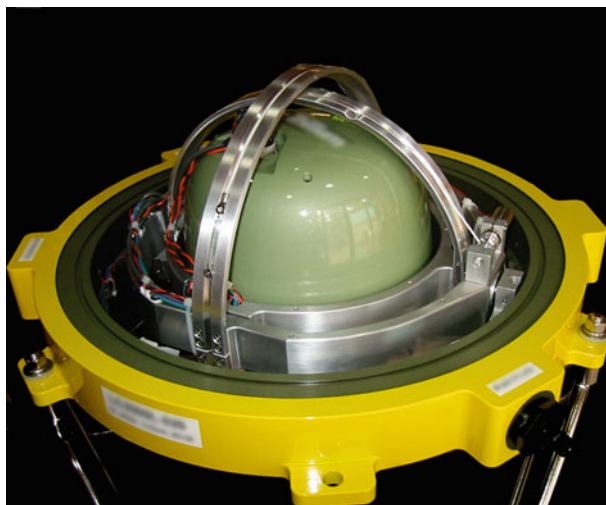
The main purpose of an OBS is to record seismic signals on the sea bottom; therefore, a seismometer specially adapted to the marine environment is necessary. For active-source experiments (using man-made sources) or local seismicity

studies, a short-period seismometer, sometimes also called geophone (Greek for “earth sound”), is sufficient. The most common are the three-component seismometers with a corner frequency of 4.5 Hz. These passive analog sensors consume no power and are relatively compact. Either they are mounted in a small pressure casing on the frame of the instrument or they can even be integrated into the main casing with other electronics (Fig. 3, bottom right). Some models are equipped with detachable seismometers, which are normally released by a burn wire or corrosion mechanism after arriving at the seafloor. High-frequency geophones are often leveled, usually with a passive, gravity-based system, a so-called gimballed system.

To study lower-frequency signals, a wideband or broadband seismometer is required. For offshore deployments, broadband seismometers with a pass band out to 120–240 s are available, but wideband seismometers with a pass band out to 30–120 s are more common because of their smaller size and lesser power consumption. These active wide- or broadband seismometers output one vertical and two horizontal components. Similar to onshore stations, the seismometers must be perfectly leveled in order to work properly: in fact, it is even more important at the seafloor because seafloor currents can create noise on the horizontal component that rotates on to the vertical component if the seismometer is not perfectly leveled (Crawford and Webb 2000). Because the tilt of the OBS on the seafloor is unknown, all wideband and broadband seismometers contain an automatic electromechanical leveling system (Fig. 6). Leveling is activated some hours after the OBS arrives at the seafloor and is usually repeated from time to time for long-term deployments. Care must be taken not to level (or even check the level) too frequently, as the switching on of the leveling circuitry can introduce a noise spike in the data. There are several ways to level either the entire seismometer package or each component separately. Some models are gimbal mounted in two directions (Fig. 6); other models contain only one gimbal but additionally turn the seismometer until all three components are leveled.

Ocean-Bottom Seismometer,

Fig. 6 Example of a broadband seismometer (*green sphere*) mounted in a two-axis gimballing system. The pressure case was removed (Photograph W. Crawford, Paris)



For all seismometers, a good coupling to the seafloor is essential. Two main approaches are in use today: integrated and mechanically isolated seismometers. Integrated seismometers are fixed directly to the instrument carrier and coupling is provided by the entire system including the anchor weight (Figs. 2 and 3, bottom). These systems are tightly arranged and easy to handle during deployment and recovery. Mechanically isolated seismometers have their own pressure case and are normally mounted on a deployment arm (Figs. 2 and 3, top). The sensor package is automatically detached when the station reaches the ocean bottom. These models are more complex, but movements or tilting of the instrument carrier will not be transmitted to the seismometer. The best coupling would be achieved by burying the seismometer into the sediment (e.g., Duennebie and Sutton 2007), but unfortunately all these instruments still need support from remote-operating vehicles (ROV) for deployment and recovery, which makes such experiments very expensive and time-consuming.

Pressure Sensor

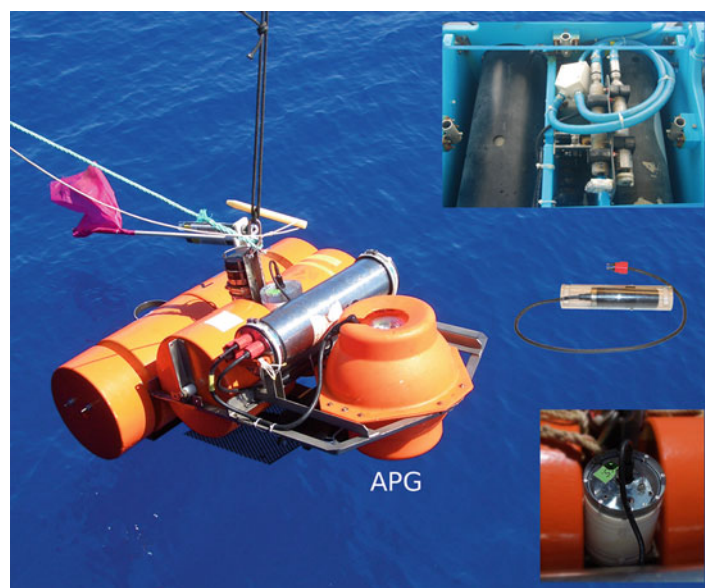
Besides the seismic channels, the majority of OBS types provide a fourth channel recording the signals of a pressure sensor. In water, pressure is much easier to measure than displacement or acceleration. Pressure is omnidirectional, so a single channel is sufficient and no special coupling to the surrounding water is necessary.

For applications only interpreting compressional waves (P-waves), like some active-source experiments, the so-called ocean-bottom hydrophones (OBH) can be used. This kind of instrument is equipped solely with a hydrophone. When combined with three-component seismological data, the additional pressure channel can be used to remove water-bounce phases from seismic signals (the pressure and seismic signals from water bounces have different polarities; see Blackman et al. 1995) and at low frequencies (less than approximately 0.1 Hz) remove low-frequency noise on the vertical seismometer channel (Webb and Crawford 1999) and determine the physical properties of the seafloor (“seafloor compliance,” e.g., Crawford et al. 1991).

The overall bandwidth of hydrostatic pressure in the oceans is large; it starts at 100 KPa at the sea surface and increases by 10 Mpa per thousand-meter water depth. Compared to this, pressure changes caused by earthquakes are very small (generally less than 1 Pa). An instrument that is able to adapt to the large pressure range as well as to detect small pressure changes over the entire seismic band would be very complex and power consuming. Therefore, various sensors exist, which are optimized for specific purposes. There are absolute gauges quantifying the entire hydrostatic pressure, or differential units, which adapt to the mean surrounding pressure and record only the small changes.

Ocean-Bottom Seismometer,

Fig. 7 Examples of pressure sensors. *Left:* instrument carrier with a differential pressure gauge (DPG) and an absolute pressure gauge (APG). *Right: APG (top),* hydrophone (*middle*), and DPG (*bottom*) (Photographs M. Schmidt-Aursch, Bremerhaven and W. Crawford, Paris (*top right*))

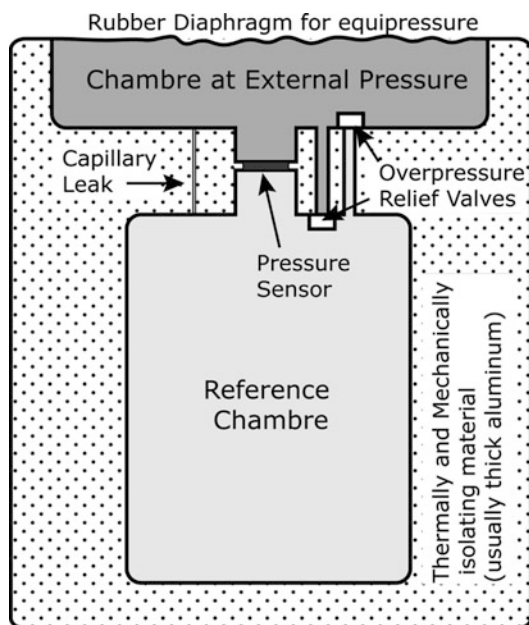


The simplest differential pressure sensor is an analog hydrophone (Fig. 7, middle). The first hydrophones were utilized during World War I, mainly to detect submarines. There are several ways to convert pressure changes into electricity, including moving coils in solenoids actuated by membranes and interferometric fiber-optic coils, but most OBS are equipped with a pressure-compensated hydrophone using the piezoelectric effect. It consists mainly of a piezoelectric ceramic cylinder, which linearly generates a voltage when opposed to mechanical stress. The entire system is encapsulated in an elastomer, e.g., polyurethane. Hydrophones are compact, passive sensors that need no additional power supply unless they integrate a preamplifier. Similar to seismometers, hydrophones are available in various short- and long-period versions with frequencies from approx. 50 kHz down to 0.01 Hz. The hydrophone acts electronically as a capacitor; therefore, the high-impedance input of the data logger must carefully be adapted to achieve both short settling times at the seafloor and linear sensitivity down to very low frequencies.

Differential pressure gauges (DPGs, Cox et al. 1984) can measure pressure signals from approximately 0.0005 Hz to 40 Hz and are

popular on broadband OBS. Differential pressure is measured between a reference chamber and the ocean using a strain gauge with small dynamic range (Fig. 7, left and bottom, and Fig. 8). The reference chamber and the gauge around it are filled with silicon oil, and the gauge is left open to ocean pressures by a soft membrane. The pressure in the reference chamber is kept close to that of the ocean using a capillary leak and, for deployment and recovery, overpressure relief valves. The capillary leak creates a high-pass filter whose corner frequency depends on the tube's diameter and length as well as the viscosity of the oil within the gauge; typically, a value of 0.002–0.004 Hz is sought. These gauges have proven difficult to calibrate accurately. In principle, changes in the viscosity of the oil with temperature and pressure should change the high-pass corner frequency, but in practice the biggest uncertainty has proven to be the absolute gain which appears to be 15–20% lower than that measured by an absolute gauge, even for DPGs that were carefully calibrated in the laboratory.

Absolute pressure gauges (APGs, Fig. 7, left and top) can be used to measure pressures from approximately 1 Hz to DC. The most commonly available absolute pressure gauges for ocean floor experiments had, until recently, a noise floor of



Ocean-Bottom Seismometer, Fig. 8 Sketch of a differential pressure gauge (DPG)

about 10 Pa, but sensitivities are now down to about 0.1 Pa. APGs are expensive and power hungry, but they are usually very well calibrated and they can also measure tsunamis and seafloor vertical motions. The pressure is measured by counting the oscillations of a quartz crystal resonator, which is different from the voltage-based signals from hydrophones, DPG, and seismometers. Several commercial companies sell absolute pressure gauges designed for full ocean depth: one possible configuration would be to deploy an APG sampling at a relatively low rate strapped on to an OBS containing a hydrophone.

Data Logger

The analog signals of the sensors are recorded by a data logger. The technology has evolved a long way from the early photographic oscillographs to the present-day 24-bit digitizers with solid-state-disk storage media, and development continues rapidly. Many efforts have been made to provide the same technical specifications – like dynamic range or size of data storage – for OBS as for onshore stations. Nevertheless, this has not been fully possible because of the added requirements

of seafloor stations: besides stringent requirements for mechanical and electric robustness, seafloor operations demand small dimensions (to fit into the pressure cases), low power consumption (to enable fully autonomous long-term deployments), and a high-precision clock. Size and power consumption are decreasing and data quality and storage are increasing continually as technology advances, but timing is still an issue.

Onshore stations can receive wireless time signals (e.g., DCF77 or GPS) to synchronize the internal clock. On the seafloor, this is unfortunately not possible, so great efforts have been made to develop high-precision clocks. Temperature-stabilized clocks using miniature ovens have been used but they are generally power hungry. Most loggers currently use as their time base temperature-compensated, microprocessor-controlled crystal oscillators (MCXO) with a frequency deviation in order of 10^{-7} to 10^{-8} . Chip-scale atomic clocks (CSAC) with much higher accuracy up to 10^{-10} may be used in the future, but they consume currently more power than crystal oscillators and there are some reliability/aging issues. The internal clock of the data logger is synchronized to an external time signal (e.g., GPS) before deployment and after recovery to determine the drift (“skew”) of the clock. Modern standard clocks show drift rates less than 500 ms/year. The clocks should not only feature a small absolute drift, drift rates should be constant over time to enable a linear time correction of the data afterward. Constant seafloor temperatures contribute to this, but large temperature contrasts and mechanical stress during deployment and recovery must be balanced by the system. For safety reasons, a backup battery independent from the main power supply is desirable for the internal clock.

Specific Requirements of Data Processing

Location of OBS

In contrast to onshore stations, for which the exact location can easily be determined, only the deployment position of the OBS is known.

Most OBS sink relatively slowly to the seafloor (less than 1 m/s) to prevent a hard impact on the ground, so, especially for large water depths, there is much time for currents to push the station away from the direct path. Depending on instrument design, current speed, and sinking time, an OBS might drift several hundreds of meters before reaching the bottom. This discrepancy is too high for experiments mapping small-scale subsurface structures with active sources and specially designed passive arrays that need exact station spacing.

Deployed OBS can be located with the help of the acoustic release unit. Most release units can respond to the onboard unit and the onboard unit can calculate the two-way travel time. Using the sound speed in water enables a calculation of the distance to the station, and ranging from different positions allows a triangulation of the OBS location at known water depth. For good constraints, the instrument should be ranged at from at least three points at approximately equal distance on a circle around the OBS-presumed position. The farther the interrogation, the better the horizontal resolution, but the instrument stops responding beyond some distance, often similar to the water depth. If the water depth is not exactly known, a ranging directly over the instrument is required. Some onboard units are equipped with more than one transducer, and small travel time differences of the response at the transducers can be used to determine the bearing to the station, allowing positioning (in theory) from a single point. Together with the ranged distance and known water depth, the location of the instrument can be determined.

If the OBS experiment is combined with multichannel or wide-angle seismics, the recorded airgun shots of the seismic profiles can be used to determine the position of the instrument on the seafloor. The calculation will be done after recovery; no additional ranging with the onboard unit is necessary. The procedure is similar to the direct ranging via the acoustic release unit; at least three shots evenly distributed around the OBS are necessary. Knowing the exact time and position of the shots, the sound velocity of water, and the water depth, the distances between

shots and OBS and hence the location of the instrument can be computed from the first arrivals of the water wave in the seismograms.

Orienting Horizontal Components

In a free-fall deployment, no method is known to assure that the OBS seismometers' north axis is aligned with geographic north. For active-source experiments, the orientation of the horizontal components is not always important, but it is essential for many passive seismological methods (e.g., receiver functions, shear wave splitting). There are two main approaches to identify the OBS orientation: direct determination by an additional sensor or indirect estimation by analyzing the seismological data.

The simplest direct determination of the orientation would be an electronic compass measuring the three components of the Earth's magnetic field and subsequent calculation of the seismometer's orientation with respect to magnetic north. Unfortunately, magnetic compasses are strongly affected by large metal parts (e.g., anchors) and local variations of the declination, which can be large, especially in volcanic areas. Fiber-optic gyros could be an alternative as they compute the true north direction by measuring the rotation of the Earth, but these sensors are still too expensive and power consuming to be installed on OBS as a standard. Microelectromechanical systems (MEMS) could offer an efficient solution in the future. MEMS accelerometers and gyroscopes, which are widespread in inertial navigation and portable electronic devices like mobile phones, would record accelerations and rotations of the instrument on its way through the water column. The spatial orientation of the instrument could then be calculated for each moment, from the initial point on deck, where the orientation of the OBS is well known, to the final arrival at the seafloor. With this method, the known orientation of the station on deck could be transferred to the sea bottom, but the MEMS systems are not yet stable and precise enough to cope with the intense movements of an OBS during deployment.

The horizontal orientation of the seismometer can also be estimated by analyzing the recorded data. The polarity of P-wave arrivals from

explosives or airgun shots can be used to determine the orientation of the sensor (e.g., Anderson et al. 1987), but in most cases no sources with equally distributed azimuths are available. Instead of man-made signals, the polarization of P- and Rayleigh waves from teleseismic events can be analyzed (Stachnik et al. 2012), but this also requires a good azimuthal distribution of the signals in order to average out anomalies due to structure. Another problem is the naturally high noise level on the horizontal OBS components in combination with the small amplitudes of teleseismic events. A new approach is to use ambient noise instead of teleseismic events (Zha et al. 2013). For this method, virtual Rayleigh waves are calculated from the cross-correlations of vertical and horizontal components of various station pairs. These virtual Rayleigh waves can be used for a similar polarization analysis as that applied to teleseismic data.

Timing

Although the internal clocks in OBS data loggers are synchronized to an external time signal (e.g., GPS) before launching and after retrieving the instrument, the clock can stop before the final synchronization or an incorrect reference time can be recorded during one of the synchronizations. The offset between the internal clock and the external signal after recovery (“skew”) must be corrected, as many applications (e.g., active seismics and event localization) rely on a precise common time base of all stations.

There are two main methods used to correct/verify drift rates in the seismological data, and both procedures assume a constant linear clock drift over the entire recording time. Using the first method, earthquakes or shots are located using data from the OBS, and then the time residuals for each station are plotted as a function of time. A bad timing or non-drift-corrected instrument can be identified by a linear trend in the time residuals. A bad clock drift is often caused by misreading the synchronization time by a multiple of 1 s, so the correction is quite simple to calculate. A non-drift-corrected instrument will have a nonintegral seconds offset at the end of the experiment, allowing a less precise correction.

If several instruments are badly corrected, it can be challenging to determine which one is faulty as several time residuals will exhibit interrelated drift. Ambient noise analysis can also be used to determine the clock drift (Hannemann et al. 2013). Cross-correlations of the vertical seismometer components of station pairs show time shifts if the stations are not properly time corrected. This method also tests the synchronizations and the linearity of the clock drift.

Passive seismological data are normally divided into daily files to keep the huge data amounts manageable. The start time for each daily file is corrected to account for the time drift. Most seismological data formats do not have an accurate enough sample rate specification to account for the drift, which is generally smaller than 1 part in 10^8 (e.g., the true sampling rate for nominally 100 sps data could be 100.000001 sps), so the time at the end of the daily file will be off by the drift rate times the 86,400 s in 1 day. Normally, this offset is smaller than one sample, but this can produce gaps in the data stream. To avoid these gaps, the entire data set can be resampled to a new, skew-corrected sampling rate. Now, the data will be continuous, but the shape of the signal may be slightly changed. Whether resampling is used or not depends on the further usage of the data and analyses to be applied.

Alternatively, some data formats (e.g., miniSEED) allow the time to be specified several times throughout the day. In the miniSEED format, the start time, the time correction, and whether this time correction was applied to the start time can be specified in every record header, which typically occurs every 500–2,000 samples (for a 4,096-byte record length and depending on compression). Although this allows an accurate specification of the clock drift, the software reading the data must still be able to stitch the data together properly when assembling multi-day records. How to do this does not yet appear to have been defined.

Noise Reduction

Noise levels of OBS data are generally higher than of onshore data, because ocean-bottom

seismometers are less coupled to the ground and the oceans are a very noisy environment. At low frequencies (<0.1 Hz), the two major sources of noise are seafloor currents and ocean waves, which have the strongest effect on, respectively, the horizontal channels and the pressure channel. Much of this noise can be removed from the vertical channel by removing the correlated signal on the pressure and horizontal seismometer channels (Webb and Crawford 1999; Crawford and Webb 2000).

The current noise is removed by calculating the transfer functions (the amplitude ratio between the two in the frequency domain) between the horizontal channels and the vertical channel in the absence of earthquakes and then subtracting the horizontal channels times this transfer function from the vertical channel. Next, the transfer function between the pressure channel and the vertical channel is calculated, and the pressure data times this transfer function is subtracted from the vertical channel. If $G_{AA}(f)$ is the power spectral density (mean squared fast Fourier transform (FFT) over several windows) of the channel we want to remove noise from, $G_{SS}(f)$ is the power spectral density of the channel containing the noise source, and $C_{AS}(f)$ is the coherency between them, then the transfer function is defined as

$$T_{AS}(f) = C_{AS}(f) \sqrt{\frac{G_{AA}(f)}{G_{SS}(f)}}$$

The transfer function is then a measure of how much of the noise on the “source” channel is induced into the vertical channel. Channel A is then “cleaned” by calculating its FFT, $A(f)$, and that of the source channel, $S(f)$, applying the formula

$$A'(f) = A(f) - T_{AS}^*(f)S(f)$$

and taking the inverse FFT of $A'(f)$ to get the time domain signal. Note that the power spectral densities and coherence should be calculated on an earthquake-free section of data, so that the only relation between them is the noise to be removed.

If removing the noise recorded on multiple channels, either a multi-coherence method must be used or each channel must be cleaned of the other before being sequentially used to clean the vertical channel. Calling the vertical channel “Z,” the horizontal channels “X” and “Y,” and the pressure channel “P,” the standard sequence is:

1. Calculate Z', Y' and P' by subtracting the coherent part of X.
2. Calculate Z'' and P'' by subtracting the coherent part of Y' .
3. Calculate Z''' by subtracting the coherent part of P'' .

This correction does not distort seismological signals because $T_{ZS}(f) \ll 1$ for the noise sources, whereas $T_{ZS}(f) = O(1)$ for seismological signals. This method cannot be used to improve the horizontal channels, for two reasons. First, the horizontal channel noise is only reduced by a few dB when the correlated noise from the other horizontal channel is removed (compared to 20+ dB for the vertical channel). Second, $T_{XY}(f) = O(1)$ for both the noise and seismological signals, so applying this correction would greatly distort the seismological signal.

These techniques should only be applied at frequencies below the microseism band. In the microseism band, the dominant “noise” signals are seismic surface waves, so removing the correlated pressure or horizontal channel signal would also eliminate, or strongly distort, seismic signals. Removing the horizontal signal times a transfer function from the vertical channel does not significantly alter the shape of any seismic waves that are also recorded on the horizontal, as the seafloor current transfer function is typically on the order of 10^{-3} or less, much smaller than the transfer function for seismic waves.

Summary

Ocean-bottom seismometers are indispensable for exploring offshore structures like mid-ocean ridges, continental margins, hydrocarbon reservoirs, or potential tsunami-generating areas, as

well as “mixed” areas such as hot-spot volcanoes, island chains, and many plate interfaces. Since the first OBS deployments in 1939, technology has advanced rapidly and modern autonomous OBS contain state-of-the art mechanics and electronics. All elements are of a rugged design to cope with the harsh conditions on board and on the sea bottom like high pressure and saltwater environment. An OBS consists of an instrument carrier with floatation made of one or more evacuated glass spheres or syntactic foam and an anchor weight coupled to a release unit. After sinking freely to the seafloor, a seismometer, a hydrophone or an absolute/differential pressure gauge, and a data logger record seismic signals continuously. Receiving a hydroacoustic signal or waiting a prespecified time, the anchor is detached from the instrument and the OBS rises back to the sea surface. Signal color painting, radio beacon, flasher, and a flag facilitate the recovery of the instrument. In contrast to onshore stations, ocean-bottom stations need some additional data processing to enhance the data quality like to orient the seismometer, correction of the drift of the internal clock, and reduction of noise on the vertical component.

Cross-References

- [MEMS Sensors for Measurement of Structure Seismic Response and Their Application](#)
- [Principles of Broadband Seismometry](#)
- [Recording Seismic Signals](#)
- [Seismometer Arrays](#)
- [Seismic Instrument Response, Correction for](#)
- [Seismic Network and Data Quality](#)
- [Seismic Noise](#)

References

- Anderson PN, Duennebie FK, Cessaro RK (1987) Ocean bore-hole horizontal seismic sensor orientation determined from explosive charges. *J Geophys Res Solid Earth* 92(B5):3573–3579
- Auffret Y, Pelleau P, Klingelhoefer F et al (2004) MicroOBS: a new generation of ocean bottom seismometer. *First Break* 22:41–47

- Blackman DK, Orcutt JA, Forsyth DW (1995) Recording teleseismic earthquakes using ocean-bottom seismographs at mid-ocean ridges. *Bull Seis Soc Am* 85(6):1648–1664
- Cox CS, Deaton T, Webb SC (1984) A deep sea differential pressure gauge. *J Atmos Oceanic Tech* 1:237–246
- Crawford WC, Webb SC, Hildebrand JA (1991) Seafloor compliance observed by long-period pressure and displacement measurements. *J Geophys Res* 96(10):16151–16160
- Crawford W, Webb S (2000) Identifying and removing tilt noise from low-frequency (<0.1 Hz) seafloor vertical seismic data. *Bull Seismol Soc Am* 90(4):952–963
- Duennebie FK, Sutton GH (2007) Why bury ocean bottom seismometers? *Geochem Geophys Geosyst* 8: Q02010. doi:10.1029/2006GC001428
- Ewing M, Vine AC (1938) Deep sea measurements without wires and cables. *Transactions of American Geophysical Union*, 19th annual meeting, pp 248–251
- Ewing M, Woollard GP, Vine AC, Worzel JL (1946) Recent results in submarine geophysics. *Geol Soc Am Bull* 57(10):909–934
- Hannemann K, Krüger F, Dahm T (2013) Measuring of clock drift rates and static time offsets of ocean bottom stations by means of ambient noise. *Geophys J Int*. doi:10.1093/gji/ggt434
- Stachnik JC, Sheehan AF, Zietlow DW, Yang Z, Collins J, Ferris A (2012) Determination of New Zealand ocean bottom seismometer orientation via Rayleigh-wave polarization. *Seismol Res Lett* 83(4):704–713. doi:10.1785/0220110128
- VESIAC (VELA Seismic Information Analysis Center, University of Michigan) (1965) A bibliography of seismology for the VELA uniform program: report of VESIAC. Geophysics Laboratory, Institute for Science and Technology, University of Michigan
- Webb SC, Crawford WC (1999) Long period seafloor seismology and deformation under ocean waves. *Bull Seis Soc Am* 89(6):1535–1542
- Zha Y, Webb SC, Menke W (2013) Determining the orientations of ocean bottom seismometers using ambient noise correlation. *Geophys Res Lett* 40. doi:10.1002/grl.50698

Online Response Estimation in Structural Dynamics

Kristof Maes and Geert Lombaert
Department of Civil Engineering, KU Leuven,
Leuven, Belgium

Synonyms

Real-time response extrapolation

Introduction

The dynamic forces acting on a structure and the corresponding system response are of great importance to many engineering applications. Often, however, the dynamic forces can hardly be obtained directly by measurements, e.g., for wind loads, and the number of response measurements during operation is in most cases limited due to practical and economical considerations. Therefore, the dynamic forces and corresponding system response have to be determined indirectly from measurements using system inversion techniques. These techniques allow estimating the dynamic forces acting on a structure from the response in a limited number of sensors, while at the same time the system response can be extrapolated to unmeasured locations from the identified forces and system states.

Originally, force identification and state estimation problems were treated separately. Force identification problems were solved offline in a deterministic setting. Many methods were proposed, most of them based on the inversion of the frequency response function (Guillaume et al. 2002; Parloo et al. 2003) or making use of a time-domain approach (Klinkov and Fritzen 2007; Nordström and Nordberg 2002). Several state estimation algorithms have been proposed for linear as well as for nonlinear systems (Wu and Smyth 2007; Hernandez 2011). Currently, the attention is shifted to joint input-state estimation, resulting in numerous recursive combined deterministic-stochastic approaches (Hwang et al. 2009; Ma et al. 2003; Gillijns and De Moor 2007). These methods allow for online response estimation in many engineering fields. This contribution focuses on recursive state estimation and joint input-state estimation algorithms applied for response estimation in structural dynamics.

System Model

In structural dynamics, first principle models, e.g., finite element (FE) models, are widely used. In many cases, modally reduced-order

models are applied, constructed from a limited number of structural modes. When proportional damping is assumed, the continuous-time decoupled equations of motion for modally reduced-order models are given by

$$\ddot{\mathbf{z}}(t) + \mathbf{\Gamma}\dot{\mathbf{z}}(t) + \mathbf{\Omega}^2\mathbf{z}(t) = \mathbf{\Phi}^T\mathbf{S}_p(t)\mathbf{p}(t) \quad (1)$$

where $\mathbf{z}(t) \in \mathbb{R}^{n_m}$ is the vector of modal coordinates, with n_m the number of modes taken into account in the model. The excitation force is written as the product of a selection matrix $\mathbf{S}_p(t) \in \mathbb{R}^{n_{\text{dof}} \times n_p}$, specifying the force locations, and a time history vector $\mathbf{p}(t) \in \mathbb{R}^{n_p}$, with n_{dof} the number of degrees of freedom of the FE model and n_p the number of forces. $\mathbf{\Gamma} \in \mathbb{R}^{n_m \times n_m}$ is a diagonal matrix containing the terms $2\xi_j\omega_j$ on its diagonal, where ω_j and ξ_j are the natural frequency and modal damping ratio corresponding to mode j , respectively. $\mathbf{\Omega} \in \mathbb{R}^{n_m \times n_m}$ is a diagonal matrix as well, containing the natural frequencies ω_j on its diagonal, and $\mathbf{\Phi} \in \mathbb{R}^{n_{\text{dof}} \times n_m}$ is a matrix containing the mass normalized mode shapes $\mathbf{\Phi}_j$ as columns.

The output vector is generally written as

$$\mathbf{d}(t) = \mathbf{S}_{d,a}\mathbf{\Phi}\ddot{\mathbf{z}}(t) + \mathbf{S}_{d,v}\mathbf{\Phi}\dot{\mathbf{z}}(t) + \mathbf{S}_{d,d}\mathbf{\Phi}\mathbf{z}(t) \quad (2)$$

where $\mathbf{S}_{d,a}$, $\mathbf{S}_{d,v}$, and $\mathbf{S}_{d,d} \in \mathbb{R}^{n_d \times n_{\text{dof}}}$ are selection matrices indicating the degrees of freedom corresponding to the acceleration, velocity, and displacement or strain measurements, respectively. The output vector is composed of $n_{d,d}$ displacement or strain measurements, $n_{d,v}$ velocity measurements, and $n_{d,a}$ acceleration measurements, where n_d is the sum of $n_{d,d}$, $n_{d,v}$, and $n_{d,a}$.

Equations 1 and 2 can be written into state-space form. After time discretization and adding noise, the following discrete-time combined deterministic-stochastic state-space description of the system is obtained:

$$\mathbf{x}_{[k+1]} = \mathbf{A}\mathbf{x}_{[k]} + \mathbf{B}\mathbf{p}_{[k]} + \mathbf{w}_{[k]} \quad (3)$$

$$\mathbf{d}_{[k]} = \mathbf{G}\mathbf{x}_{[k]} = \mathbf{J}\mathbf{p}_{[k]} + \mathbf{v}_{[k]} \quad (4)$$

where $\mathbf{x}_{[k]} = \mathbf{x}(k\Delta t)$, $\mathbf{p}_{[k]} = \mathbf{p}(k\Delta t)$, and $\mathbf{d}_{[k]} = \mathbf{d}(k\Delta t)$, $k = 1, \dots, N$, Δt is the sampling

time step, and N is the total number of samples. $\mathbf{d}_{[k]} \in \mathbb{R}^{n_d}$ is the output vector, assumed to be measured, and $\mathbf{p}_{[k]} \in \mathbb{R}^{n_p}$ is the input vector. The state vector $\mathbf{x}_{[k]} \in \mathbb{R}^{2n_m}$ consists of the modal displacements and velocities:

$$\mathbf{x}_{[k]} = \begin{bmatrix} \mathbf{z}_{[k]} \\ \dot{\mathbf{z}}_{[k]} \end{bmatrix} \quad (5)$$

The deterministic system behavior is described by the four system matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{G}$, and $\mathbf{J}$. The process noise vector $\mathbf{w}_{[k]} \in \mathbb{R}^{2n_m}$ and measurement noise vector $\mathbf{v}_{[k]} \in \mathbb{R}^{n_d}$ account for unknown excitation sources and modeling errors. In addition, the measurement noise vector $\mathbf{v}_{[k]}$ accounts for measurement errors. As an alternative to models based on first principles, models can be directly identified from experimental vibration data using system identification techniques; see, e.g., Van Overschee and De Moor (1993) and Reynders and De Roeck (2008).

State Estimation

The objective of state estimation is to estimate the system states from a limited number of response measurements (e.g., acceleration measurements) and a system model. The state vector can be determined using a deterministic approach (e.g., least squares estimation of the state vector in the frequency domain) or using a combined deterministic-stochastic approach, which mostly results in recursive time-domain algorithms which can be applied for online state estimation. The best-known recursive state estimation algorithm for linear systems is the Kalman filter algorithm (Kalman 1960), which is outlined next.

In the derivation of the Kalman filter, the unknown system input is assumed to be zero-mean white noise and is included in the vectors $\mathbf{w}_{[k]}$ and $\mathbf{v}_{[k]}$. Under this assumption, the system described by Eqs. 3 and 4 becomes

$$\mathbf{x}_{[k+1]} = \mathbf{A}\mathbf{x}_{[k]} + \mathbf{w}_{[k]} \quad (6)$$

$$\mathbf{d}_{[k]} = \mathbf{G}\mathbf{x}_{[k]} + \mathbf{v}_{[k]} \quad (7)$$

The noise processes $\mathbf{w}_{[k]}$ and $\mathbf{v}_{[k]}$, which now account for all excitation sources, are assumed to be zero mean and white, with known covariance matrices $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{S}$:

$$\mathbb{E} \left[\begin{pmatrix} \mathbf{w}_{[k]} \\ \mathbf{v}_{[k]} \end{pmatrix} \begin{pmatrix} \mathbf{w}_{[l]}^T & \mathbf{v}_{[l]}^T \end{pmatrix} \right] = \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^T & \mathbf{R} \end{bmatrix} \delta_{[k-l]} \quad (8)$$

with $\mathbf{R} > 0$, $\begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^T & \mathbf{R} \end{bmatrix} \geq 0$, and $\delta_{[k]} = 1$ for $k = 0$ and 0 otherwise.

If the covariance matrix of the unknown system input $\mathbf{C}_p \in \mathbb{R}^{n_{\text{dof}} \times n_{\text{dof}}}$ and the measurement error covariance matrix $\mathbf{R}_m \in \mathbb{R}^{n_d \times n_d}$ are known, the noise covariance matrices are calculated as

$$\begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^T & \mathbf{R} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_{\text{dof}} \\ \mathbf{J}_{\text{dof}} \end{bmatrix} \mathbf{C}_p \begin{bmatrix} \mathbf{B}_{\text{dof}}^T & \mathbf{J}_{\text{dof}}^T \end{bmatrix} + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_m \end{bmatrix} \quad (9)$$

where the system matrices $\mathbf{B}_{\text{dof}} \in \mathbb{R}^{2n_m \times n_{\text{dof}}}$ and $\mathbf{J}_{\text{dof}} \in \mathbb{R}^{n_d \times n_{\text{dof}}}$ are obtained by assuming $n_p = n_{\text{dof}}$ in Eqs. 3 and 4, respectively, i.e., excitation acting at every degree of freedom in the model.

State estimation consists of estimating the system states $\mathbf{x}_{[k]}$ from a set of response measurements $\mathbf{d}_{[k]}$. A state estimate $\hat{\mathbf{x}}_{[k,l]}$ is defined as an estimate of $\mathbf{x}_{[k]}$, given the output sequence $\mathbf{d}_{[k]}$, with $n = 0, 1, \dots, l$. The corresponding error covariance matrix, denoted as $\mathbf{P}_{[kl]}$, is defined as

$$\mathbf{P}_{[kl]} = \mathbb{E} \left\{ (\mathbf{x}_{[k]} - \hat{\mathbf{x}}_{[k|k-l]}) (\mathbf{x}_{[k]} - \hat{\mathbf{x}}_{[k|k-l]})^T \right\} \quad (10)$$

The Kalman filter algorithm is initialized using an initial state estimate vector $\hat{\mathbf{x}}_{[0|-1]}$ and its error covariance matrix $\mathbf{P}_{[0|-1]}$, both assumed known. Hereafter, it propagates by computing state estimates $\hat{\mathbf{x}}_{[k+1|k]}$ recursively from the following equation:

$$\hat{\mathbf{x}}_{[k+1|k]} = \mathbf{A}\hat{\mathbf{x}}_{[k|k-1]} + \mathbf{K}_{[k]} (\mathbf{d}_{[k]} - \mathbf{G}\hat{\mathbf{x}}_{[k|k-1]}) \quad (11)$$

The gain matrix $\mathbf{K}_{[k]}$ is determined such that the state estimates are minimum variance

and unbiased (i.e., $\mathbf{K}_{[k]} = \operatorname{argmin}_{\mathbf{K}_{[k]}} \operatorname{tr}\{\mathbf{P}_{[k|k]}\}$, $\mathbb{E}\{\mathbf{x}_{[k]} - \hat{\mathbf{x}}_{[k|k-1]}\} = \mathbf{0}$). The recursive state estimator of the form (11) can be split in two steps, i.e., the measurement update and the time update:

Measurement update

$$\mathbf{L}_{[k]} = \mathbf{P}_{[k|k-1]} \mathbf{G}^T (\mathbf{G} \mathbf{P}_{[k|k-1]} \mathbf{G}^T + \mathbf{R})^{-1} \quad (12)$$

$$\hat{\mathbf{x}}_{[k|k]} = \hat{\mathbf{x}}_{[k|k-1]} + \mathbf{L}_{[k]} (\mathbf{d}_{[k]} - \mathbf{G} \hat{\mathbf{x}}_{[k|k-1]}) \quad (13)$$

$$\mathbf{P}_{[k|k]} = \mathbf{P}_{[k|k-1]} - \mathbf{P}_{[k|k-1]} \mathbf{G}^T (\mathbf{G} \mathbf{P}_{[k|k-1]} \mathbf{G}^T + \mathbf{R})^{-1} \mathbf{G} \mathbf{P}_{[k|k-1]} \quad (14)$$

with $\mathbf{K}_{[k]} = \mathbf{A} \mathbf{L}_{[k]}$

Time update

$$\hat{\mathbf{x}}_{[k+1|k]} = \mathbf{A} \hat{\mathbf{x}}_{[k|k]} \quad (15)$$

$$\mathbf{P}_{[k+1|k]} = \mathbf{A} \mathbf{P}_{[k|k]} \mathbf{A}^T + \mathbf{Q} - \mathbf{A} \mathbf{L}_{[k]} \mathbf{S}^T - \mathbf{S} \mathbf{L}_{[k]}^T \mathbf{A}^T \quad (16)$$

The algorithm given in Eqs. 12, 13, 14, 15, and 16 distinguishes itself from the classical Kalman filter by including the correlation between the process noise vector $\mathbf{w}_{[k]}$ and the measurement noise vector $\mathbf{v}_{[k]}$, i.e., $\mathbf{S} \neq \mathbf{0}$. In the equations above, the system is assumed to be time invariant. The algorithm can, however, also be applied to time-variant systems, resulting in system matrices $\mathbf{A}_{[k]}$ and $\mathbf{G}_{[k]}$ depending on the time step k .

In general, the system input does not satisfy the white noise assumption. The Kalman filter, however, yields good results as long as the applied forces are broadband and approximately white. If the forces acting to the structure do not meet this assumption, the quality of the estimated states can be improved by estimating both the system input and states, i.e., joint input-state estimation.

Joint Input-State Estimation

The objective of joint input-state estimation is to jointly estimate the forces applied to the system and the corresponding states from a limited

number of response measurements and a system model. Similar to state estimation, the applied forces can be determined using a deterministic approach, e.g., Guillaume et al. (2002), or alternatively using a combined deterministic-stochastic approach, which mostly results in recursive time-domain algorithms which can be applied for online joint input-state estimation. The algorithm outlined next is a joint input-state estimation algorithm which was originally proposed by Gillijns and De Moor (2007). The algorithm was applied for force identification and response estimation in structural dynamics by Lourens et al. (2012) and was extended for applications in structural dynamics where ambient excitation becomes important by Maes et al. (2013).

The system under consideration is described by Eqs. 3 and 4. The noise processes $\mathbf{w}_{[k]}$ and $\mathbf{v}_{[k]}$ are assumed to be zero mean and white, with known covariance matrices $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{S}$, defined by Eq. 8. Joint input-state estimation consists of estimating the forces $\mathbf{p}_{[k]}$ and states $\mathbf{x}_{[k]}$ from a set of response measurements $\mathbf{d}_{[k]}$. A state estimate $\hat{\mathbf{x}}_{[k|l]}$ is defined as an estimate of $\mathbf{x}_{[k]}$, given the output sequence $\mathbf{d}_{[n]}$, with $n = 0, 1, \dots, l$. The corresponding error covariance matrix, denoted as $\mathbf{P}_{[k|l]}$, is defined in Eq. 10. An input estimate $\hat{\mathbf{p}}_{[k|l]}$ and its error covariance matrix $\mathbf{P}_{p[k|l]}$ are defined similarly. The filtering algorithm is initialized using an initial state estimate vector $\hat{\mathbf{x}}_{[0|-1]}$ and its error covariance matrix $\mathbf{P}_{[0|-1]}$, both assumed known. Hereafter, it propagates by computing the force and state estimates recursively in three steps, i.e., the input estimation step, the measurement update, and the time update:

Input estimation

$$\tilde{\mathbf{R}}_{[k]} = \mathbf{G} \mathbf{P}_{[k|k-1]} \mathbf{G}^T + \mathbf{R} \quad (17)$$

$$\mathbf{M}_{[k]} = \left(\mathbf{J}^T \tilde{\mathbf{R}}_{[k]}^{-1} \mathbf{J} \right)^{-1} \mathbf{J}^T \tilde{\mathbf{R}}_{[k]}^{-1} \quad (18)$$

$$\hat{\mathbf{p}}_{[k|k]} = \mathbf{M}_{[k]} (\mathbf{d}_{[k]} - \mathbf{G} \hat{\mathbf{x}}_{[k|k-1]}) \quad (19)$$

$$\mathbf{P}_{p[k|k]} = \left(\mathbf{J}^T \tilde{\mathbf{R}}_{[k]}^{-1} \mathbf{J} \right)^{-1} \quad (20)$$

Measurement update

$$\mathbf{L}_{[k]} = \mathbf{P}_{[k|k-1]} \mathbf{G}^T \tilde{\mathbf{R}}_{[k]}^{-1} \quad (21)$$

$$\hat{\mathbf{x}}_{[k|k]} = \hat{\mathbf{x}}_{[k|k-1]} + \mathbf{L}_{[k]} \left(\mathbf{d}_{[k]} - \mathbf{G} \hat{\mathbf{x}}_{[k|k-1]} - \mathbf{J} \hat{\mathbf{p}}_{[k|k]} \right) \quad (22)$$

$$\mathbf{P}_{[k|k]} = \mathbf{P}_{[k|k-1]} + \mathbf{L}_{[k]} \left(\tilde{\mathbf{R}}_{[k]} - \mathbf{J} \mathbf{P}_{[k|k]} \mathbf{J}^T \right) \mathbf{L}_{[k]}^T \quad (23)$$

$$\mathbf{P}_{xp[k|k]} = \mathbf{P}_{px[k|k]}^T - \mathbf{L}_{[k]} \mathbf{J} \mathbf{P}_{[k|k]} \quad (24)$$

Time update

$$\hat{\mathbf{x}}_{[k+1|k]} = \mathbf{A} \hat{\mathbf{x}}_{[k|k]} + \mathbf{B} \hat{\mathbf{p}}_{[k|k]} \quad (25)$$

$$\mathbf{N}_{[k]} = \mathbf{A} \mathbf{L}_{[k]} (\mathbf{I} - \mathbf{J} \mathbf{M}_{[k]}) + \mathbf{B} \mathbf{M}_{[k]} \quad (26)$$

$$\begin{aligned} \mathbf{P}_{[k+1|k]} = & \begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{P}_{[k|k]} & \mathbf{P}_{xp[k|k]} \\ \mathbf{P}_{px[k|k]} & \mathbf{P}_{[k|k]} \end{bmatrix} \begin{bmatrix} \mathbf{A}^T \\ \mathbf{B}^T \end{bmatrix} \\ & + \mathbf{Q} - \mathbf{N}_{[k]} \mathbf{S}^T - \mathbf{S} \mathbf{N}_{[k]}^T \end{aligned} \quad (27)$$

The gain matrices $\mathbf{M}_{[k]}$ and $\mathbf{L}_{[k]}$ are determined such that both the input estimates $\hat{\mathbf{p}}_{[k|k]}$ and the state estimates $\hat{\mathbf{x}}_{[k|k-1]}$ are minimum variance and unbiased (i.e., $\mathbf{M}_{[k]} = \operatorname{argmin}_{\mathbf{M}_{[k]}} \operatorname{tr} \{ \mathbf{P}_{p[k|k]} \}$, $\mathbb{E} \{ \mathbf{p}_{[k]} - \hat{\mathbf{p}}_{[k|k]} \} = 0$, $\mathbf{L}_{[k]} = \operatorname{argmin}_{\mathbf{L}_{[k]}} \operatorname{tr} \{ \mathbf{P}_{[k|k]} \}$, and $\mathbb{E} \{ \mathbf{x}_{[k]} - \hat{\mathbf{x}}_{[k|k]} \} = 0$). In the equations above, the system is assumed to be time invariant. The algorithm can, however, also be applied to time-variant systems, resulting in system matrices $\mathbf{A}_{[k]}$, $\mathbf{B}_{[k]}$, $\mathbf{G}_{[k]}$, and $\mathbf{J}_{[k]}$ depending on the time step k .

Very often unknown ambient forces such as wind loads are acting on the structure. For these loads, the force locations or spatial distributions of the forces are not well known. In this case, the estimated forces $\mathbf{p}_{[k]}$ compensate for any unknown source of vibration. They are not the true forces acting on the structure but equivalent forces that act at predefined locations (Lourens et al. 2012).

Response Estimation

After applying the Kalman filter algorithm or the joint input-state estimation algorithm, the estimated state vector $\hat{\mathbf{x}}_{[k|k]}$ and force vector $\hat{\mathbf{p}}_{[k|k]}$ (in the case of joint input-state estimation) can be used to estimate the output at any arbitrary location in the structure, using the following modified output equation:

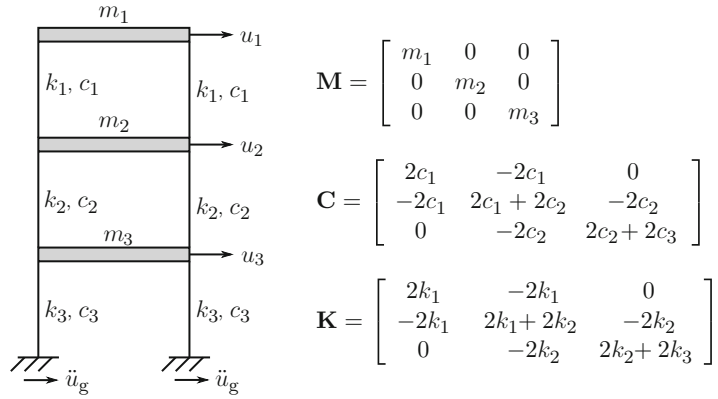
$$\hat{\mathbf{d}}_{e[k|k]} = \mathbf{G}_e \hat{\mathbf{x}}_{[k|k]} + \mathbf{J}_e \hat{\mathbf{p}}_{[k|k]} \quad (28)$$

where $\hat{\mathbf{d}}_{e[k|k]}$ is the estimated response. The matrices $\mathbf{G}_e$ and $\mathbf{J}_e$ are related to the extrapolated output quantities $\hat{\mathbf{d}}_{e[k|k]}$ and, therefore, do not equal the original matrices $\mathbf{G}$ and $\mathbf{J}$. Note that when Eq. 28 is used after applying the Kalman filter algorithm, the excitation $\hat{\mathbf{p}}_{[k|k]}$ is contained in the noise vectors $\mathbf{w}_{[k]}$ and $\mathbf{v}_{[k]}$ (cf. Eq. 7) and therefore disregarded. In this way, an error is introduced for acceleration estimation. For displacement, velocity, or strain estimation, this is not the case, as there is no direct feedthrough from the ambient forces $\hat{\mathbf{p}}_{[k|k]}$ to the output $\hat{\mathbf{d}}_{e[k|k]}$ which has to be accounted for (i.e., the rows of the matrix $\mathbf{J}_e$ corresponding to the displacement, velocity, or strain measurements are all zero).

Application: Response Estimation in a Three-Story Shear Building Subject to an Earthquake

The response estimation procedure is illustrated using numerical simulations for a three-story building subjected to horizontal earthquake excitation. The building is modeled as an idealized shear building (Fig. 1). The story masses are lumped to the floor levels, the floors are presumed infinitely rigid, and axial member deformations are neglected. This system is described by three degrees of freedom (DOFs), u_1 , u_2 , and u_3 . Under the assumption that the structure is subjected to a uniform horizontal base motion $\ddot{u}_g(t)$, the system is governed by the following set of equations of motion:

Online Response Estimation in Structural Dynamics, Fig. 1 A three-story shear building model and its mass, damping, and stiffness matrices



$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{S}_p(t)\mathbf{p}(t) \quad (29)$$

where $\mathbf{u}(t) = [u_1(t), u_2(t), u_3(t)]^T$ is the vector of displacements relative to the displacements of the building basement and $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are the mass, damping, and stiffness matrix of the

structure, respectively, defined in Fig. 1. $\mathbf{S}_p(t) = \mathbf{I}_3$ and $\mathbf{p}(t) = [-m_1, -m_2, -m_3]^T \ddot{u}_g(t)$, where $\mathbf{I}_3 \in \mathbb{R}^{3 \times 3}$ is an identity matrix.

The story masses m_i , damping values c_i , and stiffnesses k_i are assumed known and are given by

$$\begin{array}{lll} m_1 = 10 \times 10^3 \text{ kg} & c_1 = 6 \times 10^3 \text{ Ns/m} & k_1 = 6 \times 10^6 \text{ N/m} \\ m_2 = 10 \times 10^3 \text{ kg} & c_2 = 7 \times 10^3 \text{ Ns/m} & k_2 = 7 \times 10^6 \text{ N/m} \\ m_3 = 10 \times 10^3 \text{ kg} & c_3 = 8 \times 10^3 \text{ Ns/m} & k_3 = 8 \times 10^6 \text{ N/m} \end{array} \quad (30)$$

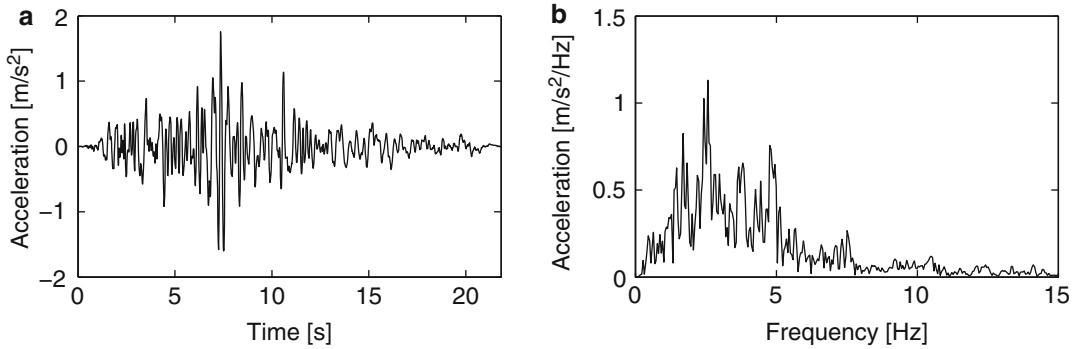
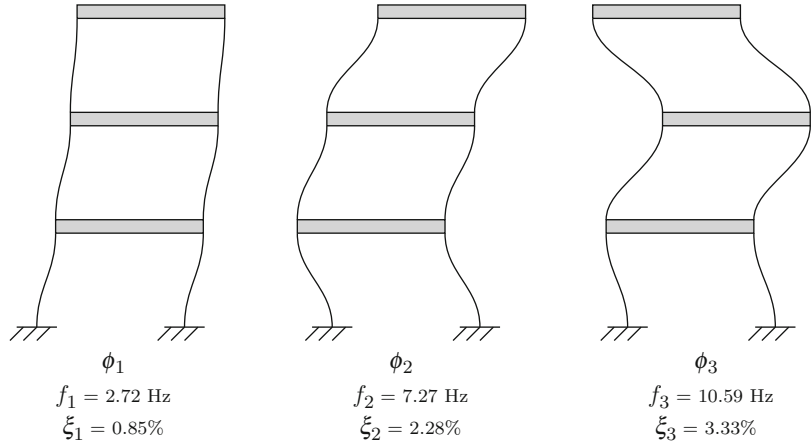
The modal properties of the system are found as the solution of the undamped eigenvalue equation $\mathbf{K}\Phi = \mathbf{M}\Phi\Lambda$, where $\Phi \in \mathbb{R}^{3 \times 3}$ collects the three eigenvectors $\phi_j \in \mathbb{R}^3$ that correspond to the eigenvalues $\lambda_j = \omega_j^2 = (2\pi f_j)^2$ located on the diagonal of Λ . The first three natural frequencies f_j of the shear building are found as 2.72, 7.27, and 10.59 Hz. The corresponding mode shapes are shown in Fig. 2. The modal damping ratios ξ_j are calculated as $\gamma_j/(2\omega_j)$, where γ_j is a diagonal element of the matrix Γ , with $\Gamma = \Phi^T \mathbf{C} \Phi$. Note that $\mathbf{C} = 10^{-3} \text{ s} \times \mathbf{K}$, so proportional damping was implicitly assumed. The first three modal damping ratios of the shear building are 0.85 %, 2.28 %, and 3.33 %.

The building is subject to the horizontal ground motion recorded during the Northridge earthquake that occurred on January 17, 1994. The epicenter of the earthquake was located in Northridge, California. The ground motion considered was recorded in the

northwest direction in Montebello, California (http://peer.berkeley.edu/peer_ground_motion_database). The time history and narrowband frequency spectrum of the recorded ground acceleration $\ddot{u}_g$ are shown in Fig. 3.

A reduced-order state-space model is constructed from the first three eigenmodes of the shear building, applying a zero-order hold assumption on the force. The data assumed for the response estimation consists of the ground acceleration time history $\ddot{u}_g$, three absolute horizontal acceleration time histories, $\ddot{u}_{a,1}$, $\ddot{u}_{a,2}$, and $\ddot{u}_{a,3}$, and one relative horizontal displacement time history, u_3 . The response is calculated at a sampling rate of 100 Hz. The initial state $\mathbf{x}_{[0]}$ is assumed zero. The output vector $\mathbf{d}_{[k]}$ considered for response estimation is composed of the relative acceleration time histories $\ddot{u}_1$, $\ddot{u}_2$, and $\ddot{u}_3$ and the relative displacement time history u_3 , where $\ddot{u}_j = \ddot{u}_{a,j} - \ddot{u}_g$. Note that the ground acceleration time history $\ddot{u}_g$ can also be used directly as input

Online Response Estimation in Structural Dynamics, Fig. 2 Three natural modes of the three-story shear building



Online Response Estimation in Structural Dynamics, Fig. 3 (a) Time history and (b) narrowband frequency spectrum up to 15 Hz of the recorded ground accelerations

in northwest direction due to the Northridge earthquake, January 17, 1994, 12:31 UTC, Montebello, California

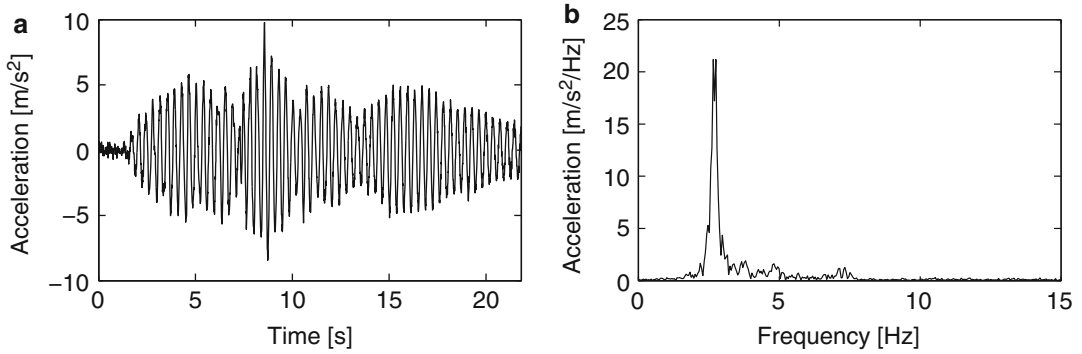
to a forward calculation, hereby obtaining the response in the structure at predefined locations. The proposed filtering algorithms, however, have the advantage of canceling out measurement errors and modeling errors. A small error on the natural frequencies of the system, for example, hardly influences the quality of the predicted response obtained by applying the Kalman filter or the joint input-state estimation algorithm, whereas for the forward calculation, this is mostly not the case.

Gaussian white noise is added to the ground acceleration time history $\ddot{u}_g$ and to the calculated time histories ($\ddot{u}_{a,1}$, $\ddot{u}_{g,2}$, $\ddot{u}_{a,3}$, and u_3), in order to represent the measurement error on the data. A choice is made to assume the same noise level γ_a for all acceleration signals and a noise

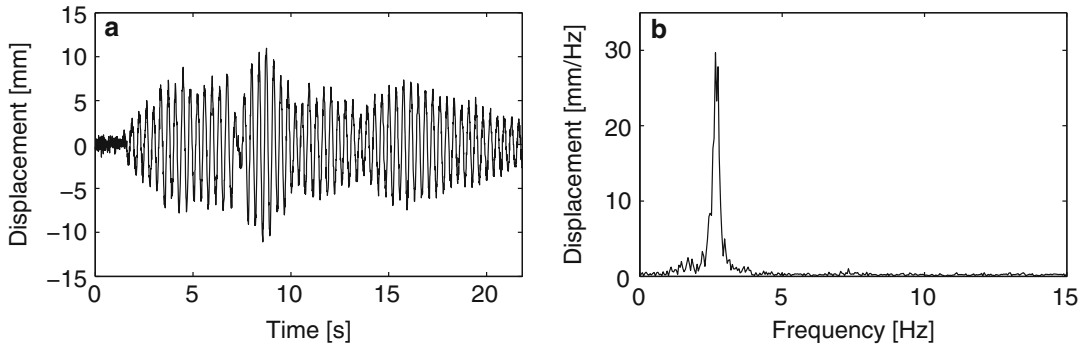
level γ_d for the displacement signal. The noise level γ_a is 5 % of the maximal acceleration $\ddot{u}_{a,3}$, and the noise level γ_d is 5 % of the maximal displacement u_3 ($\gamma_a = 0.17 \text{ m/s}^2$, $\gamma_d = 0.54 \text{ mm}$). The polluted output time histories ($\tilde{u}_{1[k]}$, $\tilde{u}_{2[k]}$, $\tilde{u}_{3[k]}$, and $\tilde{u}_{3[k]}$) are calculated according to Eq. 31:

$$\begin{aligned}
 \tilde{u}_{1[k]} &= \ddot{u}_{a,1[k]} - \ddot{u}_g[k] + \gamma_a(r_{a,1[k]} + r_{g[k]}) \\
 \tilde{u}_{2[k]} &= \ddot{u}_{a,2[k]} - \ddot{u}_g[k] + \gamma_a(r_{a,2[k]} + r_{g[k]}) \\
 \tilde{u}_{3[k]} &= \ddot{u}_{a,3[k]} - \ddot{u}_g[k] + \gamma_a(r_{a,3[k]} + r_{g[k]}) \\
 \tilde{u}_{3[k]} &= u_{3[k]} + \gamma_d r_{d,3[k]}
 \end{aligned} \quad (31)$$

where $r_{a,1[k]}$, $r_{a,2[k]}$, $r_{a,3[k]}$, $r_{g[k]}$, and $r_{d,3[k]}$ are random numbers drawn independently from a normal distribution with zero mean and unit



Online Response Estimation in Structural Dynamics, Fig. 4 (a) Time history and (b) narrowband frequency spectrum up to 15 Hz of the simulated acceleration $\ddot{u}_1$, measurement noise included



Online Response Estimation in Structural Dynamics, Fig. 5 (a) Time history and (b) narrowband frequency spectrum up to 15 Hz of the simulated displacement $\ddot{u}_3$, measurement noise included

standard deviation. The time history and narrowband frequency spectrum of the simulated acceleration $\ddot{u}_1$ and the displacement $\ddot{u}_3$ are shown in Figs. 4 and 5, respectively.

The response signals are now used as input to the response estimation procedure, hereby comparing the Kalman filter and the joint input-state estimation algorithm presented in sections “[State Estimation](#)” and “[Joint Input-State Estimation](#),” respectively. The response to be estimated consists of the horizontal displacement u_1 of the top floor.

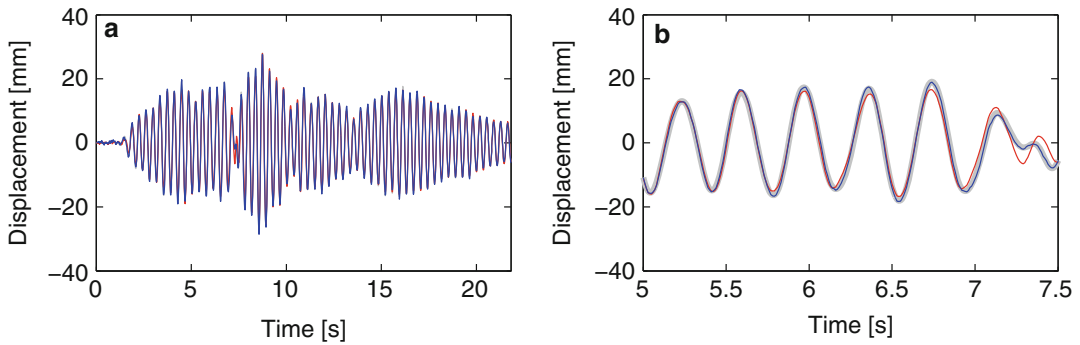
The Kalman filter is applied to estimate the system states $\hat{\mathbf{x}}_{[k|k]}$, and, subsequently, the response of the top floor $\hat{\mathbf{d}}_{e[k|k]}$ is estimated according to Eq. 28, where the term $\mathbf{J}_e \hat{\mathbf{p}}_{[k|k]}$ is disregarded. The noise covariance matrices $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{S}$ are calculated from Eq. 9, for $\mathbf{C}_p$ and $\mathbf{R}_m$ given by the following equations:

$$\mathbf{C}_p = \begin{bmatrix} m_1^2 & m_1 m_2 & m_1 m_3 \\ m_2 m_1 & m_2^2 & m_2 m_3 \\ m_3 m_1 & m_3 m_2 & m_3^2 \end{bmatrix} \sigma_{\ddot{u}_g}^2 \quad (32)$$

$$\mathbf{R}_m = \begin{bmatrix} 2\gamma_a^2 & \gamma_a^2 & \gamma_a^2 & 0 \\ \gamma_a^2 & 2\gamma_a^2 & \gamma_a^2 & 0 \\ \gamma_a^2 & \gamma_a^2 & 2\gamma_a^2 & 0 \\ 0 & 0 & 0 & \gamma_d^2 \end{bmatrix} \quad (33)$$

where $\sigma_{\ddot{u}_g}$ is the standard deviation of the ground acceleration $\ddot{u}_g$ imposed to the building, assumed 0.2 m/s² for this example. The initial state estimate vector $\hat{\mathbf{x}}_{[0|-1]}$ and its error covariance matrix $\mathbf{P}_{[0|-1]}$ are both assumed zero.

The joint input-state estimation algorithm is applied to estimate the system states $\hat{\mathbf{x}}_{[k|k]}$ and the applied forces $\hat{\mathbf{p}}_{[k|k]}$, hereafter estimating the response of the top floor $\hat{\mathbf{d}}_{e[k|k]}$, according to Eq. 28. A distribution of the equivalent force is



Online Response Estimation in Structural Dynamics, Fig. 6 (a) Complete and (b) detail of the simulated (gray) and estimated displacement time history u_1 (red, Kalman filter; blue, joint input-state estimation algorithm)

assumed by putting $\mathbf{S}_p(t) = [1, m_2/m_1, m_3/m_1]^T$ in Eq. 29. The estimated force $\hat{\mathbf{p}}_{[k|k]} \in \mathbb{R}$ has a physical meaning, i.e., the inertial force resulting from the ground motion applied at the top of the shear building. The noise covariance matrices $\mathbf{Q}$, $\mathbf{R}$, and $\mathbf{S}$ are calculated from Eq. 9, for $\mathbf{C}_p = \mathbf{0}$ and $\mathbf{R}_m$ given by Eq. 33. The initial state estimate vector $\hat{\mathbf{x}}_{[0|-1]}$ and its error covariance matrix $\mathbf{P}_{[0|-1]}$ are both assumed zero.

The results of the response estimation procedure are shown in Fig. 6. The detail in Fig. 6b reveals that the results obtained by applying the joint input-state estimation algorithm are more accurate than the results from the Kalman filter. The inertial forces applied to the structure, which result from ground motion, do not perfectly fulfill the white noise assumption, which explains the errors on the results from the Kalman filter. The equivalent forces assumed when applying the joint input-state estimation algorithm successfully compensate for the excitation applied to the structure. As a result, only very small errors on the estimated displacement due to measurement errors remain.

Summary

This reference entry focuses on two recursive time-domain algorithms that can be applied for response estimation in structural dynamics. The first algorithm is the Kalman filter, a well-known state estimation algorithm that has been used for many applications in various engineering

disciplines. The second algorithm is a recursive joint input-state estimation algorithm. Both algorithms are briefly outlined and thereafter compared using simulations for a three-story building subject to an earthquake.

Cross-References

- [Site Response: Comparison Between Theory and Observation](#)

References

- Gillijns S, De Moor B (2007) Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough. *Automatica* 43(5):934–937
- Guillaume P, Parloo E, De Sitter G (2002) Source identification from noisy response measurements using an iterative weighted pseudo-inverse approach. In: *Proceedings of ISMA2002 international conference on noise and vibration engineering*, Leuven, pp 1817–1824
- Hernandez E (2011) A natural observer for optimal state estimation in second order linear structural systems. *Mech Syst Signal Process* 25(8):2938–2947
- Hwang J, Kareem A, Kim W (2009) Estimation of modal loads using structural response. *J Sound Vib* 326:522–539
- Kalman RE (1960) On the general theory of control systems. In: *Proceedings of the first international congress of IFAC, Moscow*
- Klinkov M, Fritzen C (2007) An updated comparison of the force reconstruction methods. *Key Eng Mater* 347:461–466
- Lourens E, Papadimitriou C, Gillijns S, Reynders E, De Roeck G, Lombaert G (2012) Joint input-response

- estimation for structural systems based on reduced-order models and vibration data from a limited number of sensors. *Mech Syst Signal Process* 29:310–327
- Ma C, Chang J, Lin D (2003) Input forces estimation of beam structures by an inverse method. *J Sound Vib* 259(2):387–407
- Maes K, Lourens E, Van Nimmen K, Reynders E, Van den Broeck P, Guillaume P, De Roeck G, Lombaert G (2013) Verification of joint input-state estimation by in situ measurements on a footbridge. In: Chang F-K (ed) *Proceedings of the 9th international workshop on structural health monitoring*. Stanford, vol 1, pp 343–350
- Nordström L, Nordberg T (2002) A critical comparison of time domain load identification methods. In: *Proceedings of the 6th international conference on motion and vibration control*, Saitama, pp 1151–1156
- Parloo E, Verboven P, Guillaume P, Van Overmeire M (2003) Force identification by means of in-operational modal models. *J Sound Vib* 262(1):161–173
- Reynders E, De Roeck G (2008) Reference-based combined deterministic-stochastic subspace identification for experimental and operational modal analysis. *Mech Syst Signal Process* 22(3):617–637
- Van Overschee P, De Moor B (1993) Subspace algorithm for the stochastic identification problem. *Automatica* 29(3):649–660
- Wu M, Smyth A (2007) Application of the unscented Kalman filter for real-time nonlinear structural system identification. *Struct Control Health Monit* 14:971–990

Operational Modal Analysis in Civil Engineering: An Overview

Edwin Reynders and Guido De Roeck
Department of Civil Engineering, KU Leuven,
Leuven, Belgium

Synonyms

Ambient vibration testing; System identification

Introduction

The susceptibility of civil structures and buildings to vibrations is the major concern when designing earthquake-resistant structures. The experimental verification of the dynamic design

values, in particular the relevant standing wave or modal characteristics (natural frequencies, damping ratios, and mode shapes), is essential for validating the dynamic structural design, hence for guaranteeing the safety of the structure under earthquake loading. Furthermore, the modal characteristics depend on the stiffness of the structure, so monitoring their evolution over time allows, in principle, to identify structural damage. For example, comparing modal characteristics that have been measured before and after the occurrence of an earthquake can help in detecting structural damage that is not directly visible. However, modal characteristics may also be sensitive to environmental conditions, such as temperature or relative humidity, and to loading conditions such as excitation amplitude or added mass. These influences need to be properly taken into account when employing the results of a modal test for design validation, model calibration, structural health monitoring, etc.

Depending on the type of excitation that is employed to perform a modal test, a distinction is made between forced vibration testing, ambient vibration testing, and combined or hybrid vibration testing, also termed experimental modal analysis (EMA), operational modal analysis (OMA), and operational modal analysis with exogenous inputs (OMAX), respectively. In EMA, which was historically the first testing approach that was developed, the structure is excited by one or several measured dynamic forces, the response of the structure to these forces is recorded, and the modal characteristics in the frequency range of interest are identified from the measured data. The first EMA techniques were single degree of freedom methods such as peak picking and circle fitting, in which it is assumed that at resonance, only a single mode dominates. Consequently, these techniques are not useful when some modes of interest have close natural frequencies. This disadvantage was later removed with the introduction of multiple degree of freedom methods for EMA. By now, EMA is a well-established and often-used approach in mechanical and aerospace engineering, as documented by,

e.g., Maia and Silva (1997), Heylen et al. (1997), Allemang (1999), and Ewins (2000).

EMA methods are in general not well suited for the modal testing of civil structures and buildings, because these structures cannot be tested in laboratory conditions. When employing compact and practical actuators, the contribution of the measured excitation to the total structural response is rather low. This implies that the ever-present ambient or operational excitation – due to wind or traffic loading, for example – can most often not be neglected. OMA techniques have therefore been developed. They extract the modal characteristics from the measured operational response only. The unmeasured, ambient forces are usually modeled as white noise sequences, i.e., as zero-mean time sequences of which all samples are statistically uncorrelated and have the same variance. In civil engineering, OMA has become the primary modal testing method, especially for structures whose performance under earthquake loading is critical, such as bridges, high-rise buildings, and dams. Peeters and De Roeck (2001) review some established OMA techniques, while Magalhaes and Cunha (2011) provide an extensive tutorial introduction to OMA using bridge vibration data.

Nevertheless, the OMA approach has two disadvantages: the mode shapes cannot be scaled in an absolute way (e.g., mass normalized), and there is no control over the ambient excitation signal, which may be narrow banded in the frequency domain. For these reasons, there has been an increasing interest during the last few years toward combined modal testing or OMAX techniques. The main difference between OMAX and the traditional EMA approach is that the operational loads are included in the identified system model: they are not considered as noise but as useful excitation. Consequently, the amplitude of the artificial forces can be equal to, or even lower than, the amplitude of the operational forces. This is of crucial importance for the modal testing of large structures, since it allows using actuators that are small and practical when compared to the ones needed for EMA testing of such structures. Reynders (2012) provides an in-depth review of the most relevant and powerful EMA, OMA, and

OMAX techniques and compares their performance in an extensive simulation study.

Whether an EMA, OMA, or OMAX approach is followed, all modal tests consist of three distinct stages:

1. Designing the experiment and collecting the vibration response data (for OMA) or the force and vibration response data (for EMA or OMAX)
2. Identifying a set of dynamic structural models belonging to a given model class (e.g., linear time-invariant (LTI)) from the measured data
3. Extracting and validating a set of modal characteristics based on the identified models

Each of these stages is discussed in more detail in the following sections.

Experiment Design and Data Collection

The first stage of an operational modal analysis consists of designing the experiment and gathering the necessary data. For a pure OMA test, only response data are measured. For an OMAX test, the artificial excitation is also measured.

Modes are standing waves. The largest frequency and the shortest wavelength of all standing waves of interest have a prime influence on the experiment design. The largest frequency of interest determines the sampling rate and the necessary synchronization accuracy among the different sensors, which is especially important for wireless sensor networks. Components of the signal with a frequency that is larger than half of the sampling frequency should be filtered out before sampling in order to avoid aliasing. The shortest wavelength of interest determines the spatial resolution of the test: the shorter the wavelength, the closer the response sensors need to be in order to capture the short-wavelength mode shapes in sufficient detail.

For measuring the response of the structure, many types of sensor are available. Figure 1 provides a few examples. Sensor types that are often used include accelerometers which measure uniaxial or triaxial acceleration, velocity or



Operational Modal Analysis in Civil Engineering: An Overview, Fig. 1 Some sensors that are employed for operational modal analysis: two uniaxial wired accelerometers (*left*), a triaxial wireless accelerometer (*middle*), and two optical fiber sensors (*right*) (Right photo

reproduced with permission from: Reynders et al., Damage identification on the Tilff bridge by vibration monitoring using optical fibre strain sensors, *ASCE Journal of Engineering Mechanics*, 133(2):185–193, 2007)

displacement sensors such as laser vibrometers, LVDTs (linear variable differential transformers) and capacitive sensors, and strain sensors such as strain gages or fiber optic sensors. Most sensors only measure at a single location. Some quasi-distributed sensing techniques, based on, e.g., laser scanning or photogrammetry, are available, but their use on large and/or complex structures can be challenging. The development of wireless sensing technology has greatly improved the ease and speed of occasional modal testing.

The number of sensors employed in a test can be smaller than the total number of locations at which the response needs to be measured. In such case, the test is performed in different setups, each of which yields partial mode shape estimates. They can be assembled when sensor locations overlap between different setups in such a way that there is at least one common nonzero modal displacement among both setups for every mode of interest.

For OMAX testing (and also for traditional EMA testing), actuators are also needed. Figure 2 provides a few examples. Whenever the cost of testing and the ease of installation are a concern, the use of shakers can be excluded since they are not very cost-effective. Due to the low maximum force, the classic impact hammer is only feasible for the testing of relatively small structures. A larger and more controllable version consists of a drop weight system. When specific excitation signals are needed (multisine, swept sine, etc.), pneumatic artificial muscles can in some cases provide a good alternative for shakers (Deckers et al. 2008).

When an a priori estimate of the modal characteristics that one wishes to measure is available, e.g., from a preliminary finite element model, it is possible to estimate the optimal sensor (and actuator) configuration according to a given performance measure. As shown recently by Papadimitriou and Lombaert (2012), it is



Operational Modal Analysis in Civil Engineering: An Overview, Fig. 2 Some actuators that are employed for OMAX testing: drop weight system (*left*), impact hammer (*center*), and pneumatic artificial muscle (*right*)

(Reproduced with permission from: Reynders et al., combined experimental-operational modal testing of footbridges, *ASCE Journal of Engineering Mechanics*, 136(6):687–696, 2010)

important to take the effect of spatially correlated prediction errors into account when computing the optimal sensor placement.

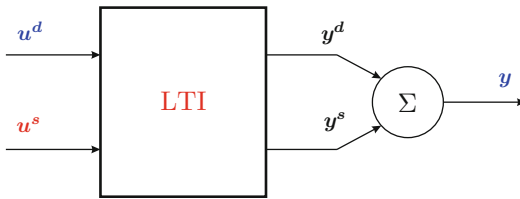
System Identification

Once the data have been gathered, a mathematical model that compactly describes the measured signals can be identified. Since the modal characteristics are classically defined as the eigenparameters of a dynamic, linear time-invariant (LTI) system description with general viscous damping, it is such a description that is identified from the OMA or OMAX data (Fig. 3). A very large number of system identification algorithms are available from the literature, but, as shown by Ljung (1999), they can be considered as particular implementations of just a few general ideas. This section therefore provides a concise overview of the different approaches rather than providing a complete list of algorithms. Most identification methods can be applied to time-domain as well as to frequency-domain data. This means that most time-domain system identification algorithms have a

frequency-domain counterpart which often differs from it only by Fourier transforms, and vice versa.

First, a distinction can be made between non-parametric and parametric identification. Non-parametric system identification involves the estimation of an impulse response function, frequency response function (FRF), correlation function, or power spectral density (PSD), not as a mathematical function depending on a few parameters, but as a set of tabulated values for each considered time lag or frequency. Although nonparametric models are sometimes directly used for modal analysis, they are most often used as preprocessed data for parametric identification since the estimation accuracy of parametric approaches is much higher than that of nonparametric approaches (Peeters and De Roeck 2001; Reynders 2012).

Parametric linear system identification as a research discipline originated in the late 1960s and developed along two lines that are still dominant today. The first line is the prediction error framework, where a system model is identified by minimizing the difference between the measured system response and the response predicted by



Operational Modal Analysis in Civil Engineering: An Overview, Fig. 3 A linear time-invariant (*LTI*) system model is identified from measured data. The total response y is measured, and part of the excitation u^d may also be measured. The operational excitation u^s is not measured. The subscript d stands for deterministic, while the subscript s stands for stochastic

the model, most often using the maximum likelihood (ML) principle (Ljung 1999; Pintelon and Schoukens 2001). The second line seeks at identifying the system model by exploiting the correlation structure (over time) of the measured response signals (and, if relevant, the measured excitation signals). This line started with the seminal work of Ho and Kalman (1966) and Akaike (1974) on system realization, which was later generalized into the so-called subspace approach (Van Overschee and De Moor 1996; Katayama 2005). Both approaches have distinct advantages: maximum likelihood methods have optimal asymptotic statistical properties under fairly general assumptions, while subspace methods are very robust, statistically accurate, and computationally much less demanding.

Estimating and Validating a Set of Modal Characteristics

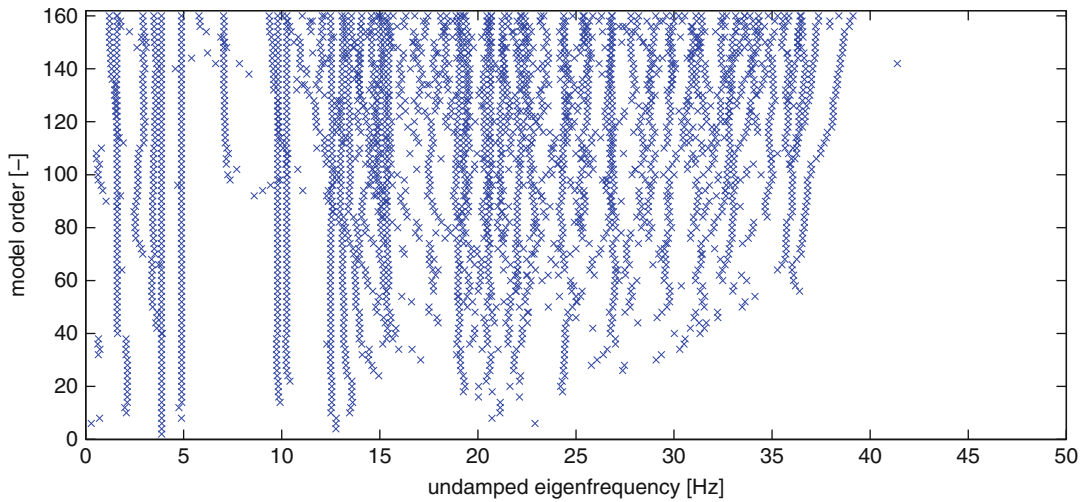
The modal characteristics corresponding to an identified system description can be obtained by exploiting its eigenstructure, that is, by decomposing the original description in terms of its free vibration solutions.

When the identified system description is nonparametric and consists of, e.g., tabulated frequency response data, rough estimates of the modal characteristics can be obtained when the damping is low. One of the most popular techniques, because of its simplicity and intuitiveness, is peak picking. Under the assumptions

that the modes have low damping and well-separated natural frequencies and that the estimated frequency response functions (in case of input-output measurements) or power spectral densities (in case of output-only measurements) are noiseless and have infinite frequency resolution, the natural frequencies can be obtained as the peaks of the frequency response function or power spectral density. The deflection shapes at those frequencies are estimates of the corresponding mode shapes, which converge to the true mode shapes when the damping goes to zero or the natural frequency spacing goes to infinity. The assumption that the natural frequency spacing is large can be relaxed by extending the peak picking method with singular value decomposition (Shih et al. 1988). The corresponding approach is called complex mode indicator function when applied to FRF data and frequency domain decomposition when applied to PSD data. However, peak picking or its extensions often lead to inaccurate estimates because most of its assumptions are not tenable in practical situations, especially in operational conditions. They are therefore mainly used for the estimation of modal characteristics in the preliminary phase of a modal test, even in laboratory conditions (Ewins 2000, p. 304).

When the identified system description is parametric, e.g., in state-space form, no assumptions – other than that the structure's response can be described with the identified description – are needed for estimating the modal characteristics. The assumed damping model is most often general viscous damping, which contains proportional damping as a special case. This means that the approach can also be used when localized dampers are present and the mode shapes are complex.

When compared to nonparametric methods, parametric system identification methods need at least one additional user-defined integer: the model order, which equals the number of eigenvalues present in the model, hence, in theory, twice the number of natural frequencies. From control theory, several model validation techniques are available for choosing the model order in an automated way, so that the prediction



Operational Modal Analysis in Civil Engineering: An Overview, Fig. 4 Stabilization diagram for one setup of a modal test on a three-span prestressed concrete bridge. The modes that are stable in all setups are shown in Fig. 5

(Reproduced with permission from: Reynders et al., Fully automated (operational) modal analysis, *Mechanical Systems and Signal Processing*, 29:228–250, 2012)

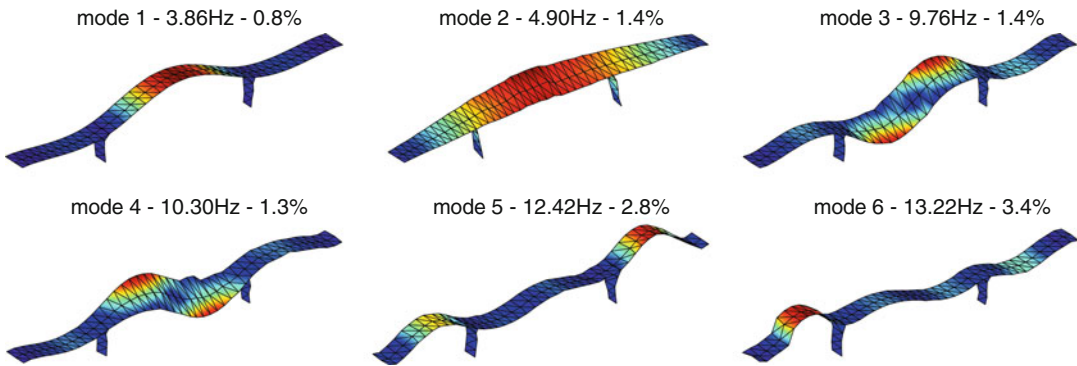
capacity of the identified models is maximized (Ljung 1999). However, in modal testing applications, one is not primarily interested in the prediction capacity of an identified model as such, but rather in the physical relevance of the individual modes that constitute the model. An alternative approach has therefore been developed, based on the empirical observation that in a very large number of modal identification problems, the physical modes of the structure appear at nearly the same eigenfrequency when the model order is over-specified, while other spurious modes do not. In this approach, parametric models are estimated for a wide range of model orders, and the modes corresponding to all these models are plotted in a model order versus natural frequency diagram, called stabilization diagram (Heylen et al. 1997; Allemang 1999). The physical modes then show up as vertical lines in the diagram, and one of the modes of each line can be chosen as a representative of the corresponding physical mode. Figure 4 shows an example of a stabilization diagram, and Fig. 5 presents the corresponding stable modes.

Finally, the obtained set of modes can be validated according to relevant validation criteria. A criterion that is often used is the so-called

modal assurance criterion (MAC) which is the dimensionless correlation coefficient between two mode shapes ϕ_j and ϕ_l :

$$MAC(\phi_j, \phi_l) = \frac{|\phi_j^H \phi_l|^2}{\|\phi_j\|^2 \|\phi_l\|^2},$$

where the subscript H denotes complex conjugate (or Hermitian) transpose, $\|\cdot\|$ denotes amplitude, and $\|\cdot\|$ denotes the Euclidian norm. When the mass of the structure is approximately uniformly distributed and the spatial resolution of the identified mode shapes is sufficiently large, the MAC value between different mode shapes should be zero (Allemang and Brown 1982). When real normal modes are expected, i.e., when proportional damping can be assumed, the modal phase collinearity (MPC) can be used for assessing the quality of the identified mode shapes (Pappa et al. 1992). The MPC value of a real normal mode shape, which is collinear in the complex plane, equals unity, while the MPC value of a complex mode shape is significantly lower than unity. Finally, for some of the most powerful parametric identification methods, the statistical accuracy of the estimated modal



Operational Modal Analysis in Civil Engineering: An Overview, Fig. 5 The final set of validated modal characteristics for an operational modal analysis of a three-span prestressed concrete bridge (Reproduced with

permission from: Reynders et al., Fully automated (operational) modal analysis, *Mechanical Systems and Signal Processing*, 29:228–250, 2012)

characteristics can be assessed. This is, for instance, the case for the stochastic subspace identification method, for which the complete joint probability distribution of the estimated natural frequencies, damping ratios, and mode shape components is available (Reynders et al. 2008; Reynders 2012).

diagram technique for estimating a final set of modal characteristics. Finally, a range of criteria is available for validating the estimated modal characteristics, from assessing the statistical accuracy of the estimates to assessing the complexity of the mode shapes.

Summary

The identification of modal characteristics from vibration data is very important in earthquake engineering, not only for design validation but also for structural health monitoring. Operational modal analysis, which identifies the modal characteristics from measured response data only, is often the most effective modal testing approach in earthquake engineering. Many system identification algorithms are available for estimating modal characteristics from the measured (force-) response data. Nonparametric algorithms are simple and intuitive, but they often yield inaccurate results. Parametric algorithms lead generally to much more accurate estimates; in particular, subspace identification combines a good statistical accuracy with a low computation cost and a high computational robustness. Since parametric algorithms need the model order as a user-defined integer, they are often combined with the stabilization

Cross-References

- ▶ [Advances in Online Structural Identification](#)
- ▶ [Ambient Vibration Testing of Cultural Heritage Structures](#)
- ▶ [Bayesian Operational Modal Analysis](#)
- ▶ [Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations](#)
- ▶ [Laser-Based Structural Health Monitoring](#)
- ▶ [Nonlinear System Identification: Particle-Based Methods](#)
- ▶ [Post-Earthquake Diagnosis of Partially Instrumented Building Structures](#)
- ▶ [Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring](#)
- ▶ [Stochastic Structural Identification from Vibrational and Environmental Data](#)
- ▶ [System and Damage Identification of Civil Structures](#)
- ▶ [Vibration-Based Damage Identification: The Z24 Bridge Benchmark](#)

References

- Akaike H (1974) Stochastic theory of minimal realization. *IEEE Trans Automat Control* 19(6):667–674
- Allemang RJ (1999) *Vibrations: experimental modal analysis*, 7th edn. University of Cincinnati, Cincinnati
- Allemang RJ, Brown DL (1982) A correlation coefficient for modal vector analysis. In: *Proceedings of the 1st international modal analysis conference*, Orlando, pp 110–116
- Deckers K, De Troyer T, Reynders E, Guillaume P, Lefeber D, De Roeck G (2008) Applicability of low-weight pneumatic artificial muscle actuators in an OMAX framework. In: Sas P, Bergen B (eds) *Proceedings of ISMA2008 international conference on noise and vibration engineering*, KU Leuven, Leuven, Sept 2008, pp 2445–2456
- Ewins DJ (2000) *Modal testing*, 2nd edn. Research Studies Press, Baldock
- Heylen W, Lammens S, Sas P (1997) *Modal analysis theory and testing*. Department of Mechanical Engineering, KU Leuven, Leuven
- Ho BL, Kalman RE (1966) Effective reconstruction of linear state-variable models from input/output functions. *Regelungstechnik* 14(12):545–548
- Katayama T (2005) *Subspace methods for system identification*. Springer, London
- Ljung L (1999) *System identification: theory for the user*, 2nd edn. Prentice Hall, Upper Saddle River
- Magalhaes F, Cunha Á (2011) Explaining operational modal analysis with data from an arch bridge. *Mech Syst Signal Process* 25(5):1431–1450
- Maia NMM, Silva JMM (1997) *Theoretical and experimental modal analysis*. Research Studies Press, Taunton
- Papadimitriou C, Lombaert G (2012) The effect of prediction error correlation on optimal sensor placement in structural dynamics. *Mech Syst Signal Process* 28:105–127
- Pappa RS, Elliott KB, Schenk A (1992) A consistent-mode indicator for the eigensystem realization algorithm. Report NASA TM-107607. National Aeronautics and Space Administration
- Peeters B, De Roeck G (2001) Stochastic system identification for operational modal analysis: a review. *ASME J Dyn Syst Meas Control* 123(4):659–667
- Pintelon R, Schoukens J (2001) *System identification*. IEEE Press, New York
- Reynders E (2012) System identification methods for (operational) modal analysis: review and comparison. *Arch Comput Method Eng* 19(1):51–124
- Reynders E, Pintelon R, De Roeck G (2008) Uncertainty bounds on modal parameters obtained from stochastic subspace identification. *Mech Syst Signal Process* 22(4):948–969
- Shih CY, Tsuei YG, Allemang RJ, Brown D (1988) Complex mode indicator function and its applications to spatial domain parameter estimation. *Mech Syst Signal Process* 2(4):367–377
- Van Overschee P, De Moor B (1996) *Subspace identification for linear systems*. Kluwer, Dordrecht

Paleoseismology of Glaciated Terrain

Björn Lund
Department of Earth Sciences, Uppsala
University, Uppsala, Sweden

Synonyms

Deglaciation seismotectonics; Early Holocene faulting; Endglacial faulting; Glacially induced faulting; Glacio-isostatic faulting; Glacio-seismotectonics; Lateglacial faulting; Postglacial faulting

Introduction

Earthquakes generally occur in response to tectonic loading of the crust. However, once the crust has reached a state of stress close to frictional equilibrium other processes that affect the crustal stress field, such as the filling of a reservoir or water injection in a borehole, may trigger seismicity. The discovery of large fault scarps in northern Fennoscandia, and their subsequent dating to the end of the latest glaciation, the Weichselian, showed that also the load of an ice sheet can have a significant impact on earthquake occurrence.

As an ice sheet grows the weight of the ice will make the lithosphere subside into the softer

asthenosphere below, a process known as glacial isostatic adjustment (GIA). The load increases both vertical and horizontal stresses in the lithosphere, and additional horizontal stresses develop due to the flexure of the elastic part of the lithosphere. Since variations to the size of the ice sheet takes place on much shorter time scales than asthenospheric flow, the flexural stresses develop at a slower pace. How these glacially-induced stresses affect earthquake occurrence depend on the prevailing stress field prior to the start of the glaciation, but generally in continental areas an ice sheet will tend to decrease differential stress under the load and therefore stabilize faults. As the ice retreats the vertical stress will decrease faster than the horizontal flexural stress due to the delayed response of the asthenosphere. In addition, horizontal tectonic stresses may have accumulated under the stabilizing weight of the ice sheet. The combination of deglaciation stresses, accumulated tectonic stress and the pre-existing stress field can potentially destabilize faults and cause large “postglacial” earthquakes, as was the case in Fennoscandia approximately 10,000 years ago.

Since the discovery of the Fennoscandian faults much effort in searching for and understanding postglacial faulting has been concentrated in the previously glaciated regions of northern Europe and North America. These are generally intraplate, continental regions which are not very seismically active at present. As many of the countries in these regions: Canada, Sweden, Finland and Russia, are investigating the

feasibility of long term storage of high-level nuclear waste, which require safety assessments spanning hundreds of thousands to a million years, the possibility of postglacial faulting contributes significantly to the seismic hazard of such repositories. The connection between postglacial faults and nuclear waste repositories have made the identification of these faults a controversial, and highly debated subject.

Over the years, as more workers have been identifying and analyzing faults related to the latest glaciation in Fennoscandia and North America, these faults have become known as *postglacial*, *endglacial* or *lateglacial* faults, with the name indicating both cause and timing. Other terms emphasize either the cause (*glacially induced* or *glacio-isostatic faults*) or occurrence time (*early Holocene faults*). In the broader context of the effect of glaciations on crustal deformation and seismicity, terms such as *deglaciation seismotectonics* and *glacio-seismotectonics* have been used. The general discussion in this article will use the term *glacially induced* fault to emphasize the effect of glaciers on faulting both during their growth and decay, and to avoid the association to the end of the latest glacial period. The term postglacial fault is the one most commonly used and it will be used here to refer to faults that occurred at the end of, or after a glaciation. For specific faults, when the occurrence time has been associated with the last stage of deglaciation, the term endglacial will be used.

This article will review where glacially induced faults have been observed and much of the focus will be on the faults that occurred at the end of the latest glacial period. Identification of the faults will be discussed, their mechanism of faulting and finally a brief look at the current effect of deglaciation on fault stability and seismicity.

Observations of Glacially Induced Faulting

This section briefly reviews the identification and analysis of glacially induced faults in various regions of the world.

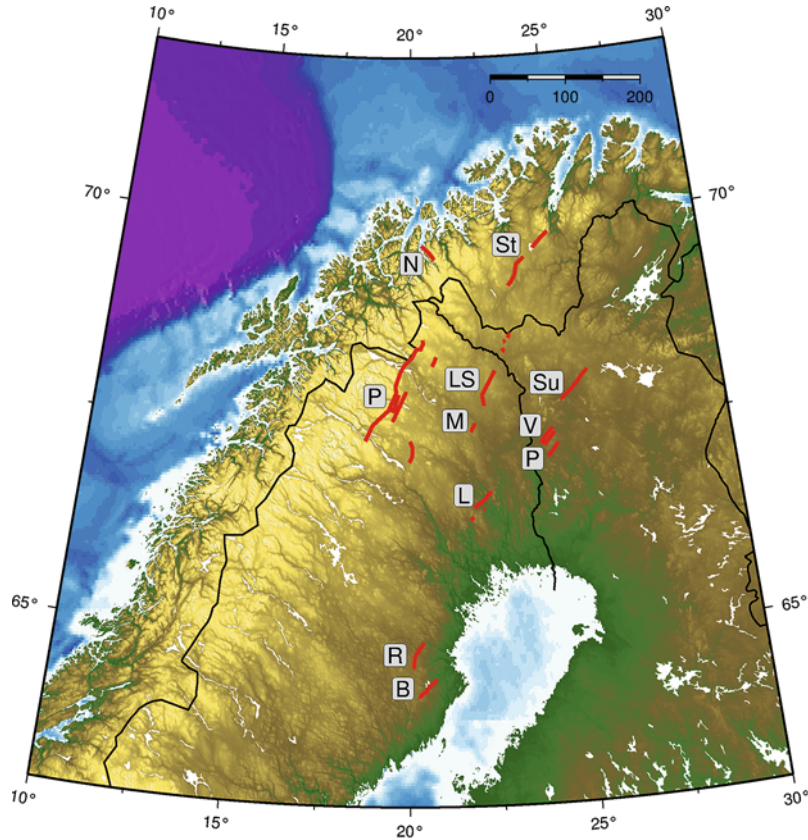
Fennoscandia

Northern Fennoscandia's spectacular endglacial faults (EGFs), in a region often referred to as the Lapland fault province (Fig. 1a), have long been known to the local Sami populations as prominent features in the landscape. The name of the 155 km long Pärvie fault scarp translates from Sami along the lines of "Like a breaking wave," an apt description of how the fault disrupts the gentle mountain terrain in northernmost Sweden (Fig. 1b). First described as late- or postglacial in Finland in the 1960s, subsequent investigations in Finland, Sweden and Norway identified approximately a dozen kilometer-scale fault scarps by the end of the 1980s (reviewed in Kuivamäki et al. 1998; Lagerbäck and Sundh 2008; Olesen et al. 2004). Most faults run in a north-northeasterly direction (Fig. 1a) and have reverse motion offsets with downthrow to the northwest, and inferred dips to the southeast. Generally, the faults have reactivated old, Proterozoic or Archean structures, although not necessarily the most prominent of these. The largest known EGF is the 155 km long Pärvie fault, which has a throw of up to 10 m and which runs just east of the Caledonian mountain front in northernmost Sweden. The Lapland fault province extends into Finnmark in northern Norway and into northwestern Finland. To the south, large EGFs had until recently only been detected down to Burträsk, south of Skellefteå in northern Sweden. However, soft-sediment deformation in varved sequences in south and central Sweden has been attributed to strong seismic shaking during end- and postglacial periods, but without identification of a causative fault (Mörner 2004). Recent LiDAR imagery (Fig. 2) have disclosed a number of new, smaller faults in Lapland, more complex fault systems around some of the larger faults, notably on the central Pärvie fault, and a 5 km long scarp in central Sweden of endglacial origin, albeit yet unconfirmed in bedrock (Smith et al. 2014).

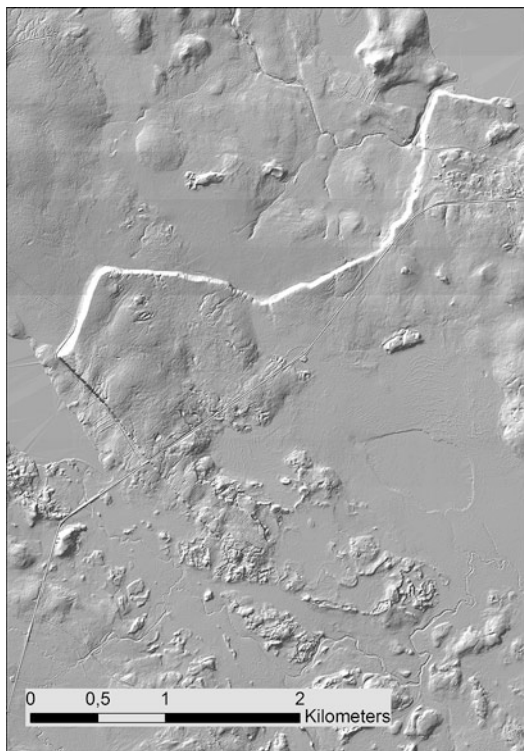
The fault scarps are inferred to have been produced by large earthquakes, evidenced by the ubiquitous occurrences of faulting-related phenomena such as low angle landslides and

**Paleoseismology of
Glaciated Terrain,**

Fig. 1 (*Upper*) Map of northern Fennoscandia showing the Lapland fault province. Faults are: in Norway St: Stuoragurra, N: Nordmannvikkalden; in Sweden P: Pärvie, M: Merasjärv, LS: Lainio-Suijavaara, L: Lansjärv, R: Röjnoret, B: Burträsk; in Finland P: Pasmajärvi, V: Venejärvi, Su: Suasselkä. (*Lower*) The Pärvie fault north of Tjuonajokk, Sweden (Photo: B. Lund)



P



Paleoseismology of Glaciated Terrain, Fig. 2 LiDAR image of the Risträskkölén section of the Lansjärv endglacial fault in northern Fennoscandia. The fault scarp is up to 25 m high in the western corner, where the fault abruptly changes strike and continues to the southeast. Further to the southeast the fault can be seen to displace hummocky moraines just before it disappears (Processed LiDAR image courtesy of Henrik Mikko, Geological Survey of Sweden)

soft-sediment disturbances (seismites). Extensive trenching has been carried out on, and in the vicinity of, some of the more accessible faults and shows both liquefaction phenomena and that faulting occurred as single events rather than repeated movements on the faults. The glacial sediments aid in determining the timing of the earthquakes relative to local deglaciation. Using morphological features associated with ice-flow, glacial melting and marine erosion and deposition (many of the events occurred in areas which at that time were covered by the sea), rupture has been observed to occur just before local deglaciation (the Pärvie fault), during deglaciation, just after local deglaciation (the Lansjärv fault) or later in postglacial times (the

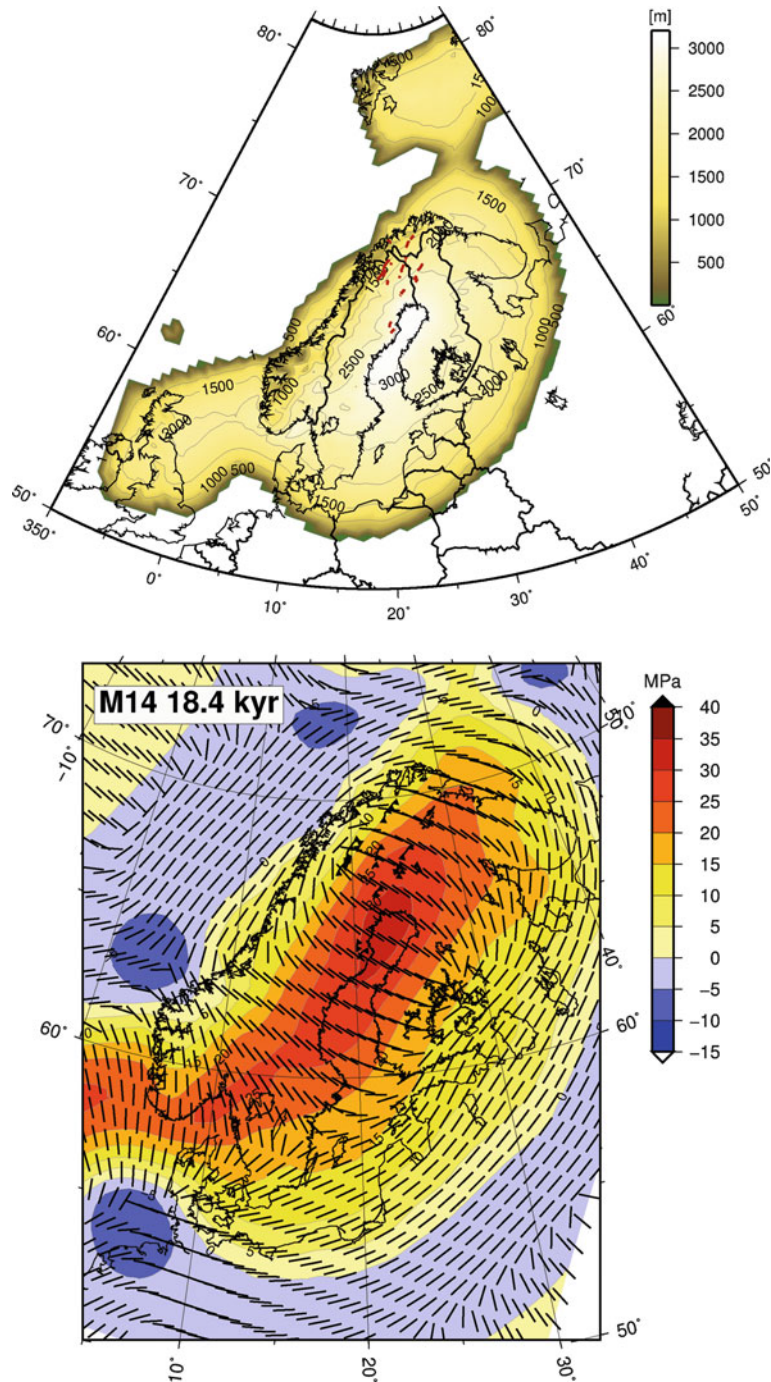
Stuoragurra fault). Deglaciation in northern Fennoscandia occurred approximately 10,000 years ago but absolute dating of the ruptures, and also relative dating of the different scarps, has proven difficult. This is due both to the scarcity, and uncertain origins, of organic material in the ruptured sediments and because the course and chronology of the latest Fennoscandian deglaciation is not well known (Lagerbäck and Sundh 2008).

The Lapland fault province occupies a region mostly between the highest mountains in northern Scandinavia and the center of the Weichselian ice sheet, which is inferred to have been located in the Bay of Bothnia where it reached a thickness of approximately 3 km (Fig. 3a). Although the Weichselian glacial period consisted of a number of stadials, the whole region is well within the subsidence bowl of the maximum extent of the Weichselian ice sheet at the last glacial maximum (LGM), as shown clearly even today by the current glacial rebound field. As such, the horizontal flexural stresses in the region have been compressional, and most likely of significant magnitude, during much of the glacial period (see the example of one possible model realization in Fig. 3b).

Earthquake moment magnitude is calculated from the rupture area, the amount of slip and the shear modulus of the rock. First-order estimates of the endglacial earthquake magnitudes can be obtained from empirical relationships between the surface rupture length and magnitude. More accurate estimates need better constraints on the width of the fault plane, on the slip distribution in the earthquake and the shear modulus. Current microearthquake activity on the faults suggests that they are seismogenic down to at least 35 km depth (Arvidsson 1996; Juhlin and Lund 2011; Lindblom et al. 2014). The dip of the fault planes at depth is poorly constrained, but recent reflection seismic and microearthquake investigations indicate that the dip of the Pärvie and Burträsk faults are around 60° to the southeast (Juhlin et al. 2010; Juhlin and Lund 2011; Lindblom et al. 2014). Assuming that the current seismicity reflects the downdip extent of the rupture planes, the width of the fault planes can then be

**Paleoseismology of
Glaciated Terrain,**

Fig. 3 (Upper) Map of the UMISM model of the Weichselian ice sheet at the last glacial maximum (LGM) 18,400 years BP (Schmidt et al. 2014). Endglacial faults indicated with red lines. (Lower) Modelled glacially induced maximum horizontal stress at LGM (Lund and Schmidt 2011). The model has elastic lithospheric thickness 120 km, upper mantle viscosity $5 \cdot 10^{20}$ Pa s and lower mantle viscosity $3 \cdot 10^{21}$ Pa s. Stress directions indicated by black bars, magnitudes with the color scale. Endglacial faults indicated in black



calculated. The slip distributions, or even the average slip, on the faults cannot be estimated with any certainty. One available option is to use relations of surface rupture offset to slip

distribution estimates for similarly large reverse faulting earthquakes today. It is, however, still unclear if rupture of endglacial faults is similar to that of “regular” tectonic earthquakes, or if the

relatively higher stresses produced at shallow depths by the ice produce relatively more shallow slip, thereby making such comparisons less relevant. With these limitations in mind, the large endglacial earthquakes in Lapland have estimated magnitudes between 7 and 8.

As alluded to above, the endglacial faults of northern Scandinavia are still seismically active (Arvidsson 1996; Lindblom et al. 2014). In fact, most of the earthquakes in northern Sweden occur in the vicinity of an endglacial fault, and all the major endglacial faults in northern Sweden and Norway are seismically active. Recent installations of seismic stations in the vicinity of the Suasselkä and Pasmajärvi faults in Finland indicate that these also host microearthquakes, albeit at a lower rate of activity. The magnitudes of the current earthquakes on the faults reach 3.6 for events that have well determined locations. There are indications that historic events of magnitude approximately 4.5 can be associated with some of the endglacial faults. The cause of the current activity is debated and includes hypotheses such as current tectonic strain release in significantly weakened zones or aftershocks to the endglacial events. In spite of the pervasive microseismicity, it has not been possible to detect any fault movement at the surface since the time of rupture, neither using geological indicators, GPS nor InSAR (Mantovani and Scherneck 2013).

Scotland

The ice sheet over Great Britain was significantly smaller and thinner than the Fennoscandian ice sheet, although it was at times connected to the Fennoscandian ice through an ice bridge across the North Sea. Early investigations of postglacial tectonics in Scotland identified displaced late- and postglacial shorelines in the Firth of Forth and the former ice-dammed lake in Glen Roy. Further work in Scotland indicated that a number of fault offsets were spatially and temporally associated with seismically induced features such as landslides and liquefied sediments (Fenton 1992; Ringrose 1989) and reported

10–100 m of horizontal displacements on 1–14 km long faults. Later investigations in the northwest Highlands have questioned some of the earlier work, especially the evidence for significant strike-slip faulting (Firth and Stewart 2000). Although the study still considered the lateglacial and early Holocene period as significantly seismically active, postglacial faulting was found to be limited to meter-scale vertical movements along pre-existing faults. Interestingly, all examples of postglacial seismicity in Scotland have been attributed to the disappearance of the smaller ice cap developed during the Younger Dryas, and not the deglaciation of the main British ice sheet (Firth and Stewart 2000). Current seismicity in Scotland does not seem to concentrate at inferred postglacial faults, nor does it correlate well with the center of uplift. The seismicity is highly variable, but have been suggested to cluster at the limits of the former ice-sheet, or in areas of local uplift anomalies (Firth and Stewart 2000).

Other European Observations

Smaller-scale structures inferred to be glacially induced have been identified in a number of areas around Europe. Close to the Fennoscandian ice sheet there are reports of faults in Russian Karelia with postglacial movement. Initially inferred to have repeated earthquakes, later analysis favor a “one-off” deformation phase approximately 11 kyr BP but found a seismic origin difficult to confirm (Kuivamäki et al. 1998).

In Germany, shallow meter-scale faults and folds were identified in a sand-pit in Upper Pleistocene alluvial-aeolian deposits 1 km away from the Osning Thrust, a major fault system in central Europe (Brandes et al. 2012). The faults have experienced normal faulting motion some 16–13 kyr ago, when the area was in the forebulge of the Fennoscandian ice sheet. Later, during deglaciation reverse motion occurred, inferred to be due to the N-S directed compressional stress field in northern Germany. Glaciotectonics has been ruled out since the faults are south of the maximum extent of the ice sheet.

In Siekierki near the Polish/German border, soft-sediment deformation structures have recently been interpreted as seismites due to glacially induced earthquakes during the Saalian deglaciation 120 kyr ago (Van Loon and Pisarska-Jamroży 2014). However, no causative fault was identified.

In Northern Iceland, the Kerlingar normal fault runs subparallel to the Northern Volcanic Zone (NVZ), but east of the zone itself. It is an unusually long normal fault with throw down to the east and it is not parallel to the fissure swarms in the NVZ (Hjartardóttir et al. 2010). Hjartardóttir et al. (2010) suggests that the Kerlingar fault formed shortly after the late Pleistocene deglaciation due to isostatic rebound with differential movements between two adjacent crustal domains.

North America

Postglacial faults were first described in 1843, in New York State (Mather 1843), although at that time neither the theory of global glacial periods, nor the concept of postglacial faulting were known. The offset grooves and scratches observed by Mather (1843) were in fact glacial striations. Several other observations of small-scale postglacial faults with offsets of a few millimeters were reported from northeastern USA and eastern Canada during the early twentieth century, and many more during the latter part of the century (Fenton 1994). To date, only two large scale phenomena in Canada have been proposed as large post- or endglacial faults: the Aspy fault on Cape Breton Island and the differential uplift at Peel Sound in Arctic Canada. The Aspy Fault offsets an 125,000 year old intertidal rock platform by 15 m, but has no exposed fault scarp, and the surficial deposits show no evidence of Holocene movement (Adams 1996). At Peel Sound, the marine beaches on opposite sides of the sound are tilted and have differential uplift of over 60 m, which is inferred to have occurred during a short period of time 9,000 years ago. The feature controlling this movement is apparently beneath the sound, but it has not been

possible to identify a causative fault, or to show that the uplift is related to earthquakes at all (Adams 1996).

In the more tectonically active western parts of North America there have only been a few reports indicative of postglacial faulting: a spectacular fault scarp with suspected postglacial movements in the eastern Sierra Nevada, California, postglacial offsets in Archean basement rock in Montana and indications that glacial loading and unloading in Puget Sound may have controlled the timing of movements on the Seattle fault, above the Cascadia subduction zone (see review in Munier and Fenton 2004).

The Wasatch and Teton faults in the Basin-and-Range province in western USA show an increase in slip rate in the late Pleistocene and in the Holocene (see review in Hampel et al. 2010). This has been interpreted as a response to unloading due to the disappearance of Lake Bonneville and the Yellowstone ice cap, respectively, at the end of the latest glacial period (Hampel et al. 2010).

Grollimund and Zoback (2001) suggested that stresses induced by the Laurentide ice sheet affected the Reelfoot Rift in Missouri, USA, and triggered the New Madrid earthquakes of 1811–1812. Later models indicate, however, that such triggering is unlikely (Wu and Johnston 2000).

South America

Investigations from a bog at Puerto del Hambre in the southern Strait of Magellan indicate that the ground surface was downfaulted by at least 30 m sometime between the occupation of the strait by a proglacial lake (12,640–10,315 ^{14}C years BP) and the marine incursion of the site at 8,265 ^{14}C years BP (Bentley and McCulloch 2005). This inferred postglacial faulting seems not to have been associated with one or more large earthquakes, as sediment cores from the area do not show any indication of soft-sediment deformation, slumping, liquefaction or other disturbances to the overall stratigraphy.

Identification of Glacially Induced Faults

As discussed in the Introduction, the identification of postglacial faults has been an intensively debated issue both in Scandinavia and in Canada. Identifying glacially induced faulting and earthquake activity is non-trivial as the surface expressions produced must be distinguished from similar phenomena created by other processes such as glacial movement, ice-water interactions, gravity and temperature. It is also important, not least from a seismic hazard point of view, to distinguish “one-off” postglacial faulting from other neotectonic faulting which may “just” have long recurrence times.

As the physical characteristics of postglacial faulting, tectonic faulting and glaciotectionic deformation overlap, no single criterion can uniquely distinguish them. Instead, multiple lines of evidence need to be assembled for a correct classification. Several identification and classification schemes have been proposed and discussed over the years (see review in Munier and Fenton 2004), the scheme by Muir-Wood (1993) has for example been widely used in Fennoscandia and Canada to differentiate between postglacial faulting and glaciotectionic deformation. It contains the following key criteria, which should all be fulfilled:

1. The surface or material that appears to be offset has to have originally formed as a continuous, unbroken unit. Can the surface be dated? Is it the same age?
2. Can the apparent evidence of an offset be shown to be related directly to a fault?
3. Is the ratio of displacement to overall length of the feature less than 1:1,000? For most faults this ratio, a function of the strength of the rock prior to fault rupture, is between 1:10,000 and 1:100,000.
4. Is the displacement reasonably consistent along the length of the feature?
5. Can the movement be shown to be synchronous along its entire length?

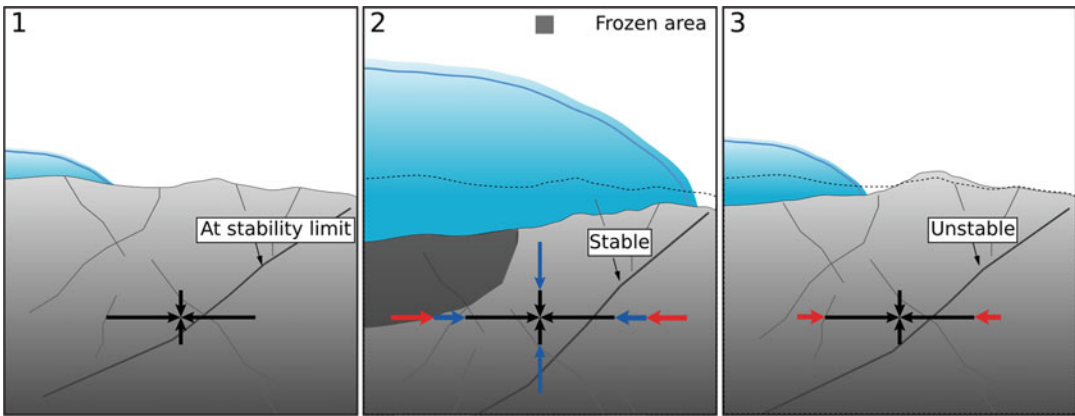
However, a simple list like the above rarely covers all the variations in scales and postglacial

deformation styles in the various glaciated or previously glaciated regions of the world (Munier and Fenton 2004). Detailed investigations at a variety of scales and with multiple techniques are essential in identifying the two over-arching attributes of postglacial faults; their age and their underlying driving mechanism.

The techniques used to investigate postglacial faults are similar to those employed in most regular paleoseismic investigations. In searching for possible candidates in the vast, sparsely inhabited areas of Fennoscandia and Canada, the analysis of stereoscopic aerial photographs has been very successful. As referred to above, LiDAR based digital elevation models nowadays provide unprecedented resolution of features on the Earth's surface (Fig. 2), even through dense vegetation. The LiDAR images not only provide a tool to identify faults and landslides, they are also used to analyze glacial landforms and their spatial relation to the faults. Detailed field mapping is essential to understand regional and local stratigraphy and geomorphic development, in order to aid in determining both the formation process and the age. Geophysical investigations such as magnetics, gravity, ground penetrating radar, sonar profiling (in lakes) and reflection seismic profiling, geodesy and local earthquake seismology reveal various aspects of fault geometry and conditions. Trenching is vital to gain higher confidence in the style of faulting and the age determination, and to investigate fault reactivation patterns. Finally, drilling is used to investigate shallow fault geometry and to determine mechanical, chemical and hydrological properties of the fault.

Mechanics of Glacially Induced Faulting

Although the effects of glaciers and ice sheets on fault stability and seismicity had already received some attention, interest in the topic increased significantly with the observation that there are very few earthquakes below Greenland and Antarctica (Johnston 1987). The analysis indicated that as the ice load increases on the crust, both



Paleoseismology of Glaciated Terrain, Fig. 4 Schematic stress evolution at a fixed location in the upper crust during a glacial period. Pre-existing (e.g., tectonic) stress field (black), glacially induced stresses: direct elastic load (blue), flexural (red). (1) Shortly before

glaciation. (2) During glaciation. (3) Shortly after deglaciation. Note that neither the vertical displacement nor the arrows are to scale (Adapted from a figure courtesy of B. Fälvh, Uppsala University)

vertical and horizontal stresses increase and thereby stabilize faults (Fig. 4). In Mohr-Coulomb terms, the increase in average stress moves the Mohr-circle to the right, away from the failure envelope. Subsequent progress in modelling glacial isostatic adjustment (GIA) made it possible to refine the analysis of fault stability by including flexural stresses and temporal stress variations due to the viscosity of the mantle (e.g., Wu and Hasegawa 1996). Glacially induced stresses under large ice sheets can amount to 30–40 MPa both horizontally and vertically, depending on the size of the ice sheet and the rheology of the Earth. Although seemingly large, they are only 10–15 % of the weight of the overburden at 10 km depth, appropriate for earthquake nucleation. In addition, there are tectonic, topographic, sediment induced and other stresses that can be significant in a specific area (Fig. 4). It is important that all these stresses, the full stress field, is included in the analysis of fault stability as it is the total stress that determines the stress state, which fault orientations that are most easily activated and the stability of these (e.g., Lund et al. 2009).

Not generally included in the GIA based models is the accumulation of tectonic stress

during glaciation. Although much smaller than the glacially induced stress variations, at least in continental interiors, there can be enough stress accumulated during the cause of a glacial period to significantly affect the stability of faults once the stabilizing effect of the ice is disappearing (Johnston 1989). Over even longer time periods, such as the duration of the Greenland or Antarctic ice sheets, tectonic loading may even be able to bring faults back towards frictional equilibrium, and hence destabilize them again in spite of the overriding ice sheet (Stewart et al. 2000).

Modelling Glacially Induced Stresses

GIA models have primarily been concerned with the rates of vertical uplift through time in order to fit observations of Holocene sea-level rise, historical tide-gauge data and recent satellite based data, e.g., GPS, InSAR and GRACE gravity. Such studies have provided a wealth of information on climate, ice sheet configuration and evolution, and the rheology of the Earth. The emergence of suitable finite element models provided a convenient tool to investigate the stress field induced in the crust by a glaciation, and thus

the mechanisms of glacially induced faulting (Wu and Hasegawa 1996). The stress field that develops in the Earth due to the ice load depends both on the ice load itself and the rheological and elastic structure of the Earth. Even during glacial periods the large ice sheets evolve constantly, it is nowadays well established that during the latest glaciation, in the last 100,000 years, the ice sheets grew and declined several times and therefore had very varying extent and thickness. Crustal stresses evolve, to first order, with the ice load implying that a good ice model is vital in GIA modelling (e.g., Schmidt et al. 2014).

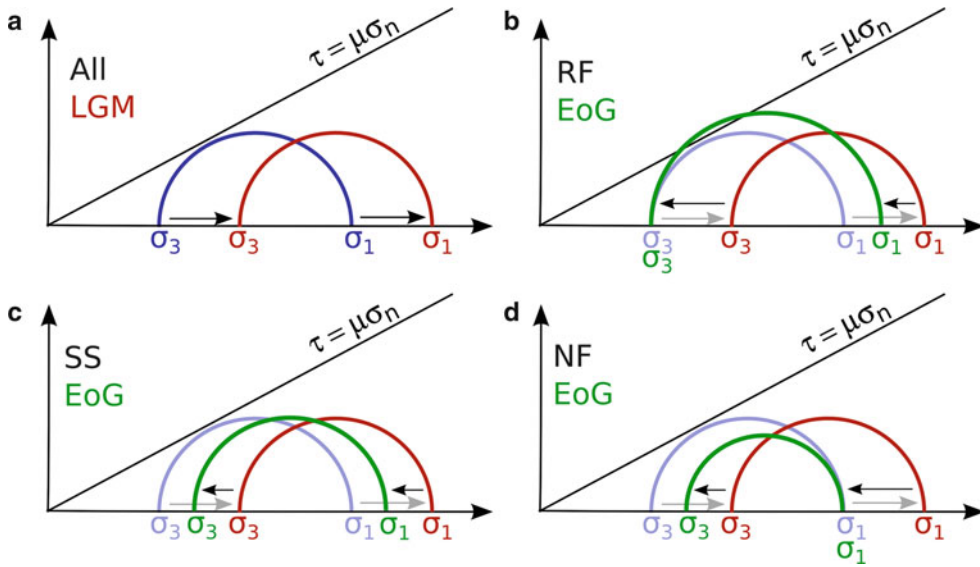
As outlined above, the growth of an ice sheet and the subsequent down-warping of the Earth's elastic lithosphere induces vertical stress in the Earth proportional to the ice load. It also induces horizontal stresses which are a combination of the Poisson effect of the load and the flexural stresses that develop as the lithosphere bends (Fig. 4). The horizontal stresses will be compressive in the upper part of the lithosphere under the load, and tensional outside the load in the so called fore-bulge. Under the load, the vertical stresses will initially be larger than the horizontal as flexure develops slowly due to the slow flow of the mantle. This produces an induced normal faulting stress regime. As the ice sheet stops growing, horizontal stress magnitudes will generally catch-up, and surpass, the vertical, inducing a strike-slip state of stress. During deglaciation the vertical stress decreases much more rapidly than the horizontal stresses, again due to the slower response of the mantle, such that the stress field will evolve to a reverse state. During and after deglaciation the uplift of the formerly glaciated region will decrease the compression in the upper crust, as the collapse of the fore-bulge will decrease the induced tension in the crust. The inward migration of the fore-bulge may produce induced tensional stresses in areas that were previously subjected to compressional glacial stresses.

Second only to the ice load itself, the rheology and elasticity of the Earth model has a major influence on the glacially induced stresses. For a certain ice load, the elastic structure determines

the magnitude of the stresses, which is especially important in the upper crust, where the glacial stresses are the largest. A high resolution layered, or three-dimensional, structure is vital in order to not significantly over-estimate the stress magnitudes in the uppermost crust. A laterally varying thickness of the elastic part of the lithosphere, and lateral variations of the parameters within the lithosphere, also affects the magnitude of the stress field and its spatial variation (Lund et al. 2009). The viscosity of the mantle, including lateral variations and non-linear rheology, strongly affects the temporal evolution of stress magnitudes outside the LGM ice margin, but has little effect inside the margin (e.g., Wu et al. 1999). The interaction of the ice load with Earth structure produces some interesting effects. The horizontal stress magnitudes induced by flexure of the elastic lithosphere depend on the spatial extent of the ice sheet relative to the thickness of the elastic lithosphere (Johnston et al. 1998). Accordingly, for commonly inferred lithospheric thicknesses, a smaller ice sheet, the size of the former British ice sheet, would produce maximum amplification of the horizontal stresses. The Fennoscandian ice sheet would similarly produce higher flexural stresses than the North American ice sheet, consistent with the observation of large endglacial fault in northern Fennoscandia.

Fault Stability

The GIA models discussed above have included neither a pre-existing stress field nor pre-defined faults. Instead, fault stability has been evaluated in post-processing, where the modelled glacially induced stresses have been added to a pre-existing stress field and areas and times susceptible to faulting have been studied. Faulting potential is commonly investigated using Mohr-Coulomb theory, through changes either in the Coulomb Failure Stress ($CFS = \tau - \mu(\sigma_n - P_p)$), where τ is the shear stress on a fault, σ_n the normal stress, μ the coefficient of friction and P_p the pore pressure



Paleoseismology of Glaciated Terrain, Fig. 5 Variation in fault stability at a location centrally under the ice sheet, from prior to the glaciation (blue lines), to LGM (a, red lines) and to the end of glaciation

(EoG, green lines) for pre-existing reverse (b), strike-slip (c) and normal (d) stress fields. (a) applies to all pre-existing stress fields as it is assumed here that the glacially induced stress field is isotropic ($\sigma_1^G = \sigma_2^G = \sigma_3^G$)

(e.g., Lund 2005), or the similar fault stability margin (Wu and Hasegawa 1996).

The models generally show that for a stable continental interior, with a preexisting stress field in a state of reverse faulting, the region within the LGM ice margin is stabilized by the growth of an ice sheet. Towards the end of deglaciation, as the vertical stress diminishes much faster than the horizontal stresses, optimally oriented faults in the region centrally under the ice sheet tend to become unstable (Fig. 5 and Lund et al. 2009; Wu et al. 1999). This model agrees well with the occurrence of the endglacial earthquakes in northern Fennoscandia, both in timing and strike direction (Lund et al. 2009). For strike-slip or normal faulting stress fields it is not the interior (Fig. 5c, d) but the areas outside the LGM ice margin that are affected by fault instability, not only at the time of deglaciation but also during times of glacier advance and stand still. The stability of the fore-bulge region depends very much on the details of ice and Earth models, and is as such less well constrained.

Fault stability and earthquake occurrence depend on more factors than just the state of stress. Pore pressure is a vital parameter in the analysis of faulting, and pore pressures in the crust during deglaciation are notoriously difficult to estimate due to the lack of data. The Fennoscandian ice sheet was cold based for much of the latest glacial period (Lagerbäck and Sundh 2008) and it is likely that permafrost could have existed in the crust below the ice for long periods of time. Without a hydraulic head, it is uncertain how pore pressure in the crust would be affected by the ice and permafrost. If there was hydraulic contact between water in the ice sheet and crustal pore fluids, high pore pressures would have been able to penetrate deep into the crust during the long duration of the ice sheet (e.g., Lönnqvist and Hökmark 2013). The combination of high pore pressures and permafrost could potentially have maintained higher than hydrostatic pressures in the crust during deglaciation, adding to the instability of faults.

In a further advancement of finite element based GIA models it has become possible to

explicitly include faults and pre-existing stresses. Early models imposed velocity boundary conditions to have faults slip at a constant rate prior to glaciation, thereby including tectonic stressing rates, but did not include the deep mantle (see Hampel et al. 2010). The models agree with previous results but more explicitly show the inhibition of slip during glaciation and rapid slip increase at deglaciation. Later models include a proper mantle description and come to similar conclusions (Steffen et al. 2014), although still in a state of early development.

Effects of Current Deglaciation

The current global warming trend increases melting of ice sheets and glaciers worldwide. Does this deglaciation influence seismicity rates in the affected regions? Areas that are glaciated and tectonically active, such as Iceland and Alaska, are good candidates for investigations of the process but few studies have been attempted. Although glacier thinning rates may have increased, the time scales involved, compared to the time of good seismic network coverage, make it difficult to establish solid background rates to compare to. A glacier surge in Alaska affected local seismicity by increasing it in the surge reservoir region and decreasing it in the receiving region, in agreement with predictions for a compressive tectonic stress state (Sauber and Molnia 2004). Also in Alaska, stress changes due to retreating ice in the Icy Bay area in the 80 years between the 1899 Yakataga (M_W 8.1) and Yakutat (M_W 8.1) earthquakes and the 1979 St. Elias (M_S 7.2) earthquake are inferred to have decreased the stability of the St. Elias fault system by 2–3 MPa (Sauber and Molnia 2004). In Iceland, the current thinning of the Vatnajökull glacier has been suggested to contribute to high seismicity and unusual focal mechanisms (Pagli and Sigmundsson 2008). Assessing how the glacially induced stress change affects earthquake activity in Iceland is, however, non-trivial as time

varying stresses from volcanic activity and rifting are active in the glaciated region.

In the areas of the two remaining continental ice sheets, Greenland and Antarctica, significantly increased seismic detection capabilities have not seriously challenged the more than 25 year old hypothesis of ice sheets as inhibitors of seismic activity. There is significant earthquake activity along the margins of the Greenland ice sheet but very little in the interior (i.e., Olivieri and Spada 2014). There has been suggestions that glacial isostatic adjustment from the Pleistocene deglaciation still affects the occurrence and mechanisms of earthquakes off the coast of Greenland (Chung 2002), and that the seismicity is affected by elastic rebound stresses from the current thinning. The data is currently too sparse to distinguish between these processes and to validate any of them with numerical models (Olivieri and Spada 2014). What is clear, however, is that the current rapid thinning of outlet glaciers in Greenland cause significant ice-related events, so called glacial earthquakes or ice-quakes (e.g. Ekström et al. 2006). Similarly, in Antarctica there is significant ice-quake activity in ice streams and outlet glaciers, but very few earthquakes detected in the continent itself. Recent investigations have, however, found swarms of seismic activity inferred to be caused by magmatic activity in West Antarctica (Lough et al. 2013).

Summary

Loads on the crust caused by glaciers and ice sheets can have a significant effect on earthquake occurrence, as indicated by the large fault scarps in northern Fennoscandia caused by earthquakes at the end of the latest glaciation, and the current seismic quiescence under the Greenland and Antarctic ice sheets. Therefore, the assessment of long term seismic hazard in regions affected by glaciation need to take this effect into account.

A large number of observations of faulting in North America and Northern Europe indicate that

these intraplate, continental regions experienced a substantial increase in seismicity as the ice sheets of the latest glaciation disappeared. In the Lapland fault province in northern Fennoscandia a dozen fault scarps, ranging in length from a few tens of kilometers to 155 km, have been associated with one-step rupture at the end of the deglaciation phase. The association to earthquake rupture comes from detailed investigations of soft-sediment disturbances such as landslides, seismites and stratigraphic offsets in trenches. These investigations also determine the age of faulting relative to the glacial sediments. The magnitudes of the largest endglacial earthquakes have been estimated as M_w 7–8 using fault scarp lengths, fault planes delineated by current microseismicity and empirical scaling relations. Postglacial fault observations in Scotland and Canada generally show shorter lengths and smaller offsets, although there are a few observations of larger scale postglacial phenomena where the cause has not been properly identified. Outside the margin of the former maximum extent of the ice sheets, such as in Germany and USA, glacially induced faulting has been identified or suggested, not necessarily contemporaneous with the latest deglaciation.

Identification of glacially induced faulting is non-trivial as the faults, or faulting related disturbances, first have to be distinguished from structures related to glaciotectionic processes. Then, identified faults have to be shown to have a glacially related origin and not be due to other neotectonic processes. Several classification schemes exist, which all suggest that a wide variety of data should be collected prior to decision making. Areal, satellite, ground mapping, geophysics, trenching and drilling data all provide complementary evidence of fault age and driving mechanisms.

Rapid model development and increase in computing power has enabled more complex models of glacial isostatic adjustment (GIA) and fault stability. A glacier load bends the elastic upper lithosphere down into the softer asthenosphere, inducing both directly load related

stresses and flexural stress in the crust. The load tends to stabilize faults inside the load margin, but may destabilize faults outside the margin, depending on the pre-existing stress field. During deglaciation, the vertical stress decreases more rapidly than the horizontal stresses due to the delayed response of the asthenosphere. If the pre-existing stress field inside the former ice margin is suitably oriented, faults may become destabilized. Outside the margin the fore-bulge will collapse and move inward, destabilizing faults if the pre-existing stress conditions combine favorably with the glacial stresses. The latest model generation includes actual faults and can estimate fault offsets in simulated earthquakes. In addition to direct load and flexural stresses, tectonic horizontal stresses may accumulate under the stabilizing weight of the ice sheets. Stress accumulation is generally slow in cratonic interiors and uncertainties in the accumulation rates makes the relative importance of the process debatable. The GIA models can currently to some degree predict location, orientation and onset times of glacially induced fault instability. However, there is still need for considerable progress in understanding the details of endglacial fault mechanisms in terms of location, geometry and size.

Improved models of glacially induced seismicity will allow better prediction of what may happen as glaciers and ice sheets decrease in the current global warming trend. The evidence for current crustal scale deglaciation related seismicity is scarce but it may increase in the next century, emphasizing the need for better seismic hazard estimates in the vicinity of thinning ice sheets.

Cross-References

- [Earthquake Magnitude Estimation](#)
- [Earthquake Mechanisms and Stress Field](#)
- [Earthquake Mechanisms and Tectonics](#)
- [Earthquake Recurrence](#)

- [Earthquake Return Period and Its Incorporation into Seismic Actions](#)
- [Paleoseismology](#)
- [Paleoseismology and Landslides](#)
- [Paleoseismology: Integration with Seismic Hazard](#)
- [Paleoseismic Trenching](#)
- [Remote Sensing in Seismology: An Overview](#)

References

- Adams J (1996) Paleoseismology in Canada: a dozen years of progress. *J Geophys Res* 101:6193–6207
- Arvidsson R (1996) Fennoscandian earthquakes: whole crustal rupturing related to postglacial rebound. *Science* 274:744–746
- Bentley MJ, McCulloch RD (2005) Impact of neotectonics on the record of glacier and sea level fluctuations, Strait of Magellan, southern Chile. *Geogr Ann* 87:393–402
- Brandes C, Winsemann J, Roskosch J, Meinsen J, Tanner DC, Frechen M, ... Wu P (2012) Activity along the Osning Thrust in Central Europe during the Lateglacial: ice-sheet and lithosphere interactions. *Quat Sci Rev* 38:49–62. doi:10.1016/j.quascirev.2012.01.021
- Chung W-Y (2002) Earthquakes along the passive margin of Greenland: evidence for postglacial rebound control. *Pure Appl Geophys* 159:2567–2584
- Ekström G, Nettles M, Tsai VC (2006) Seasonality and increasing frequency of Greenland glacial earthquakes. *Science* 311:1756–1758
- Fenton C (1992) Late quaternary fault activity in north west Scotland. In: Möner N-A, Owen LA, Stewart I, Vita-Finzi C (eds) *Neotectonics – recent advances*. Quaternary Research Association, London, pp 62
- Fenton C (1994) Postglacial faulting in eastern Canada (Open file no 2774). Geological Survey of Canada, Ottawa
- Firth CR, Stewart IS (2000) Postglacial tectonics of the Scottish glacio-isostatic uplift centre. *Quat Sci Rev* 19:1469–1493
- Grollmund B, Zoback MD (2001) Did deglaciation trigger intraplate seismicity in the New Madrid seismic zone? *Geology* 29:175–178
- Hampel A, Hetzel R, Maniatis G (2010) Response of faults to climate-driven changes in ice and water volumes on Earth's surface. *Phil Trans R Soc A* 368:2501–2517
- Hjartardóttir AR, Einarsson P, Brandsdóttir B (2010) The Kerlingar fault, northeast Iceland: a Holocene normal fault east of the divergent plate boundary. *Jökull* 60:103–116
- Johnston AC (1987) Suppression of earthquakes by large continental ice sheets. *Nature* 330:467–469
- Johnston AC (1989) The seismicity of 'stable continental interiors'. In: *Earthquakes at North-Atlantic passive margins: neotectonics and postglacial rebound*. Kluwer, Dordrecht, pp 299–327
- Johnston P, Wu P, Lambeck K (1998) Dependence of horizontal stress magnitude on load dimension in glacial rebound models. *Geophys J Int* 132:41–60
- Juhlin C, Lund B (2011) Reflection seismic studies over the Burträsk fault, Skellefteå, Sweden. *Solid Earth* 2:9–16. doi:10.5194/se-2-9-2011
- Juhlin C, Dehghannejad M, Lund B, Malehmir A, Pratt G (2010) Reflection seismic imaging of the end-glacial Pärvie Fault system, northern Sweden. *J Appl Geophys* 70:307–316. doi:10.1016/j.jappgeo.2009.06.004
- Kuivamäki A, Vuorela P, Paananen M (1998) Indications of postglacial and recent bedrock movements in Finland and Russian Karelia (Technical report no YST-99). Geological Survey of Finland, Helsinki
- Lagerbäck R, Sundh M (2008) Early Holocene faulting and paleoseismicity in northern Sweden (Technical report no C 836). Geological Survey of Sweden, Uppsala
- Lindblom E, Lund B, Tryggvason A, Uski M, Bödvarsson R, Juhlin C, Roberts R (2014) Microearthquakes illuminate the deep structure of the endglacial Pärvie fault, northern Sweden. *Geophys J Int* (Submitted)
- Lönnqvist M, Hökmark H (2013) Approach to estimating the maximum depth for glacially induced hydraulic jacking in crystalline rock at Forsmark, Sweden. *J Geophys Res* 118:1–15
- Lough AC, Wiens DA, Barcheck CG, Anandakrishnan S, Aster RC, Blankenship DD, ... Wilson TJ (2013) Seismic detection of an active subglacial magmatic complex in Marie Byrd Land, Antarctica. *Nat Geosci* 6:1031–1035. doi:10.1038/NGEO1992
- Lund B (2005) Effects of deglaciation on the crustal stress field and implications for endglacial faulting: a parametric study of simple Earth and ice models (Technical report no TR-05-04). Swedish Nuclear Fuel and Waste Management Co. (SKB), Stockholm. www.skb.se. Accessed 10 Feb 2015
- Lund B, Schmidt P (2011) Stress evolution and fault stability at Olikiluoto during the Weichselian glaciation (Technical report no 2011-14). Posiva Oy, Eurajoki. (92 p)
- Lund B, Schmidt P, Hieronymus C (2009) Stress evolution and fault stability during the Weichselian glacial cycle (Technical report no TR-09-15). Swedish Nuclear Fuel and Waste Management Co. (SKB), Stockholm. www.skb.se. Accessed 10 Feb 2015
- Mantovani M, Scherneck H-G (2013) DInSAR investigation in the Pärvie end-glacial fault region, Lapland,

- Sweden. *Int J Remote Sens* 34:8491–8502. doi:10.1080/01431161.2013.843871
- Mather WW (1843) *Geology of New York, Part 1, comprising the geology of the first geological district.* Carroll and Cook, Albany. (653 p)
- Mörner N-A (2004) Active faults and paleoseismicity in Fennoscandia, especially Sweden. Primary structures and secondary effects. *Tectonophysics* 380:139–157
- Muir-Wood R (1993) A review of the seismotectonics of Sweden (Technical report no TR-93-13). Swedish Nuclear Fuel and Waste Management Co. (SKB), Stockholm
- Munier R, Fenton C (2004) Review of postglacial faulting. In: Munier R, Hökmark H (eds) *Respect distances* (Technical report no TR-04-17). Swedish Nuclear Fuel and Waste Management Co. (SKB), Stockholm, pp 157–218
- Olesen O, Blikra L, Braathen A, Dehls J, Olsen L, Rise L, ... Anda E (2004) Neotectonic deformation in Norway and its implications: a review. *Nor J Geol* 84:3–34
- Olivieri M, Spada G (2014) Ice melting and earthquake suppression in Greenland. *Polar Sci.* doi:10.1016/j.polar.2014.09.004
- Pagli C, Sigmundsson F (2008) Will present day glacier retreat increase volcanic activity? Stress induced by recent glacier retreat and its effect on magmatism at the Vatnajökull ice cap, Iceland. *Geophys Res Lett* 35. doi:10.1029/2008GL033510
- Ringrose PS (1989) Recent fault movement and palaeoseismicity in western Scotland. In: Mörner N-A, Adams J (eds) *Paleoseismicity and neotectonics.* Elsevier, Amsterdam, pp 305–314
- Sauber JM, Molnia BF (2004) Glacier ice mass fluctuations and fault instability in tectonically active southern Alaska. *Global Planet Change* 42:279–293
- Schmidt P, Lund B, Näslund J-O, Fastook J (2014) Comparing a thermomechanical Weichselian Ice Sheet reconstruction to reconstructions based on the sea level equation: aspects of ice configurations and glacial isostatic adjustment. *Solid Earth* 5:371–388. doi:10.5194/se-5-371-2014
- Smith CA, Sundh M, Mikko H (2014) Surficial geology indicates early Holocene faulting and seismicity, central Sweden. *Int J Earth Sci* 103:1711–1724. doi:10.1007/s00531-014-1025-6
- Steffen R, Wu P, Steffen H, Eaton DW (2014) On the implementation of faults in finite-element glacial isostatic adjustment models. *Comput Geosci* 62:150–159
- Stewart IS, Sauber J, Rose J (2000) Glacio-seismotectonics: ice sheets, crustal deformation and seismicity. *Quat Sci Rev* 19:1367–1389
- Van Loon AJ, Pisarska-Jamroz M (2014) Sedimentological evidence of Pleistocene earthquakes in NW Poland induced by glacio-isostatic rebound. *Sediment Geol* 300:1–10. doi:10.1016/j.sedgeo.2013.11.006
- Wu P, Hasegawa HS (1996) Induced stresses and fault potential in eastern Canada due to a disc load: a preliminary analysis. *Geophys J Int* 125:415–430
- Wu P, Johnston P (2000) Can deglaciation trigger earthquakes in N. America? *Geophys Res Lett* 27:1323–1326
- Wu P, Johnston P, Lambeck K (1999) Postglacial rebound and fault instability in Fennoscandia. *Geophys J Int* 139:657–670

Paleoseismic Trenching

H. Serdar Akyüz¹, Volkan Karabacak² and Cengiz Zabcı¹

¹İstanbul Teknik Üniversitesi, Maden Fakültesi, Jeoloji Müh. Bölümü, Ayazağa, İstanbul, Turkey

²Eskişehir Osmangazi Üniversitesi, Mühendislik-Mimarlık Fakültesi, Jeoloji Müh. Bölümü, Meşelik, Eskişehir, Turkey

Synonyms

Active fault; Earthquake history; Paleoseismology; Trenching

Introduction: The Importance of Trenching in Paleoseismology

Trenching on a fault trace is a direct way to understand the historical (mainly Holocene) evolution of a fault segment. What are signatures of past earthquakes at the Earth's surface? How do we recognize and identify past (paleo-) earthquakes? One of the most famous and fundamental principles of modern geology, “the present is the key to the past,” provides the answer, from which historical seismic events are accepted to have the same effect as modern ones do. However, the incompleteness of the historical record makes it hard to determine and identify

paleo-earthquake-related structures. The discrimination of paleoseismic evidences from geological and geomorphological processes and the attribution of them to certain historical events make an important contribution to the seismic risk assessment of any region.

Paleoseismology is the art of the identification of past earthquakes in terms of location, timing, and size by using geological and geomorphological evidence (McCalpin 2009). This branch of Earth science is mainly based on the interpretation of features recorded and preserved by depositional and erosional processes. Paleoseismology addresses a specific timescale that connects neotectonics and instrumental seismology, mostly from the Late Pleistocene to the present day. A detailed paleoseismic study can reveal a seismic history even going back several thousand years. Characterizing past earthquake parameters (e.g., fault location, earthquake magnitude, and recurrence interval) helps construct a strong database, which can be used in the modeling of potential seismic hazards.

However, there are no fixed or standard techniques for the study of past earthquake records. Accordingly, different methods and approaches must be specifically applied to each individual case. The most widely used and effective method is the trenching of recent deposits across active faults. The aim of trenching is to expose the stratigraphic and structural relationships of buried modern deposits and to discriminate the signature of paleo-events. A paleoseismologist determines the timing of identified earthquakes by using various Quaternary-dating methods allowing correlation with historical documents. The main target of a paleoseismic study is to understand the history of studied fault segment(s) and to develop a knowledge base for establishing the location and magnitude of future earthquakes for the same region. Paleoseismic results play a vital role for seismic risk assessments. This chapter aims to summarize the main steps undertaken during a paleoseismic trenching study. The most commonly used dating methods in paleoseismology are also briefly outlined.

Flowchart in Trenching Studies

The identification of a paleo-event, its dating, and its correlation with a historical earthquake is the main objective of a paleoseismic study. However, to reach that final goal, there are several steps that need to be undertaken, such as site location, logistic settings, selection of the trench type, taking of safety precautions, preparation of trench walls for logging, sampling for dating, and post-field studies, which will all be briefly mentioned in the following sections.

Site Selection: Geoinformatics, In Situ Observations, Shallow Geophysical Methods, and the Micro-Topographic Surveys

The determination of a suitable site plays a fundamental role in paleoseismology. Not only finding the fault trace but also exposing the suitable sedimentation is crucial for the recovery of a paleo-event in a trench study. Although the techniques may differ from region to region, the general sequence of investigations start with geoinformatics, continue with in situ observations, and usually end with the application of shallow geophysical and micro-topographic surveys.

Geoinformatics: Remote Sensing and Geographic Information Systems (GIS)

Aerial photographs are one of the main sources used in the identification of faults and other tectonic landforms. Analyses of aerial photograph stereopairs allow easily visualization of pseudo-3D reflections of the study region in a 2D environment. Skilled researchers can easily interpret fault scarps and tectonic landforms by using aerial photographs of actively deformed regions. Detailed topographic datasets also assist in the interpretation of physiographic features. Technological progress has greatly advanced remote sensing abilities. For example, one of the latest commercial satellites, Geoeye-2, is planned to start collecting satellite images of the Earth with a ground resolution up to 0.34 m by the second half of 2013 (GeoEye Elevating Insight 2013).

Moreover, advances in LiDAR (Light Detection and Ranging) and high computational technologies make the analysis of high-resolution digital elevation models (DEM) possible. These topographic data with submeter pixel resolution not only provide the easy identification of fault scarps but also yield auto-quantification of displacements (Zielke and Arrowsmith 2012) or precise mapping of seismogenic faults even in highly forested mountainous terrains (e.g., Cunningham et al. 2006). It is possible to merge all these products in a GIS database and to make sophisticated spatial analyses of tectonic landforms in site selection for trenching. These modern tools can focus attention on small study areas; thereby they significantly save time and reduce budgets in paleoseismological projects.

In Situ Observations: Geological and Geomorphological Mapping

To understand geometrical relations among recent (late Quaternary) geological units, and sometimes between older rocks and recent deposits, geological mapping of units and landforms of a trench site should be undertaken. Detailed mapping can also identify piercing points (i.e., offset features) along or across faults. Remote sensing analysis on the region of interest is controlled with field observations during the mapping of geological and geomorphological features. Mapping usually includes fault scarps and detailed subdivisions of Quaternary deposits across the fault. The paleoseismic history is interpreted from the constructed relationship of the mapped Quaternary deposits and the faulting pattern. The displacement measurements are often indicated on these maps. Beside the mapped features, the depositional conditions and sedimentary sources of the site are studied to gather the maximum information for understanding the local stratigraphy before trenching.

Modern depositional and erosional processes at a paleoseismic site supply data so that the trench stratigraphy can be interpreted. Paleoseismologists mostly prefer fine-grained sediments, where slow sedimentation rates allow an older

history to be recorded at relatively shallower depths. However, boulder- and cobble-rich sediments can be deposited abruptly and can instantaneously cover paleo-events. Furthermore, it is also difficult to recognize deformational structures within these coarse-grained sediments. As a result, mixed sedimentary environments such as fluvial and alluvial rivers (mostly floodplains and point bars), seasonal swamps, marshes, sag ponds, deltas, alluvial fans (especially the distal part), piedmonts, and deserts are the ideal for paleoseismic trenching.

Colluvium in front of the fault scarp is one of the most important depositional features recording the material shed from the uplifted fault block. The clast type, size, and shape that form colluvial deposits depend on the lithology of uplifted block but may also change laterally from proximal to distal parts.

Shallow Geophysical Methods

Geophysical exploration methods have a critical role for detecting the exact location of main faults and any subsidiary branches. The application of geophysical techniques can be useful in definition of the stratigraphy and the deeper structure of the fault zone, aspects that cannot be determined by trenching or drilling. Geophysical methods are often applied in regions where tectonic landforms are buried and cannot be identified by surficial observations. The most commonly used geophysical methods are seismic reflection/refraction, ground-penetrating radar (GPR), and electric, magnetic (aero and electro), and gravity techniques in paleoseismological studies. For example, along the 1999 İzmit earthquake rupture (North Anatolian Fault, Turkey), GPR profiles clearly show not only the fault zone but also displaced fluvial channels of different ages (Ferry et al. 2004). In addition to 2D representations, for faults with a more complicated tectonic setting, 3D GPR images are needed to constrain fault structure, such as demonstrated on the hidden faults in the transpressional setting of the Alpine Fault (South Island, New Zealand) (Carpentier et al. 2012).

Micro-Topographic Surveys: Construction of Very High-Detailed Morphology

Before the excavation of a natural surface on a trenching site, detailed surveying of the morphology of the field can reveal landscape anomalies, which cannot be easily recognized with the naked eye. Interpretation of high-resolution topographic data can be used to analyze the effect of tectonic activity in the formation of such morphological features. A quantitative assessment of morphological structures provides detailed information especially about the likely characteristics of the faulting. Moreover, well-defined and well-analyzed detailed morphology also decreases the excavation length (and time) at a probable trench site. Toward this goal, micro-topography and the related physiographic features can be measured by using instruments such as the electronic theodolite (total station), d-GPS (or RTK GPS), and the terrestrial LiDAR (TLS) in the field. With the introduction of the high-resolution topographic data collection of the TLS, it is even now possible to constrain the 3D-slip vectors recorded by displaced landforms (Gold et al. 2012). Not only detailed topographic maps but also topographic profiles across fault zones provide invaluable information about the history of deformation.

Final Decision: Site Selection

The best location for a trench is highly site dependent. The site selection is undertaken through the synthesis of all datasets and studies, which are briefly listed above. All combined data are interpreted to identify two (or alternatively three) targets in a trench study in order to expose (a) the full rupture geometry and the style of faulting; (b) the stratigraphy, which reflects the most complete depositional history; and (c) any buried displaced feature, if present. For the first goal, it is important to stay on the main fault strand, which records the most complete seismic history of the deformation zone. However, it is also common to target trench studies on secondary faults. Pantosti et al. (2008) preferred to excavate one of their trenches on an antithetic fault scarp instead of the main displacement zone, because of logistic reasons. Secondly, a paleoseismologist always tries to find the most

suitable place with a continuous sedimentation for the most complete stratigraphic record. Even depositional environments with very slow sedimentation rates can be chosen for slow-moving faults, like the Alhama de Murcia Fault (Eastern Betic shear zone, Spain) to expose a longer paleoseismic record (Ortuño et al. 2012). These first two objectives mostly yield the recovery of location and timing of a faulting event by fault-perpendicular trenches. However, on strike-slip faults, the magnitude of displacement can only be obtained with fault-parallel trenches, where buried and laterally displaced features are exposed.

Logistics: Administrative and Environmental Issues

Administrative Issues: Local and Governmental Permissions

Following the site selection, one of the mandatory issues is getting excavation permissions both from local (land owners) and administrative authorities. If the site is located in an archeological site or is close to a military base, then additional permissions may be required.

On cultivated lands, firstly, it is important to get the permission of the landowner but the local authorities must also be informed. The contact also provides logistic information for researchers about the existence of any possible human-made structures, such as waterworks, electricity cables, and oil/gas pipelines, which can easily be destroyed during a trench excavation. Moreover, these authorities, like local municipalities, may help in finding the work machines, which are used in trenching.

Environmental Issues: Water Table, Geography, and Transportation

The high water table in a site can cause the collapse of trench walls or it can make almost impossible for researchers to move inside the trench. The dewatering for a more pleasant working space is done in two general ways: (a) periodically pumping the water out of the trench and releasing it some distance away, and (b) draining the water by digging a shallow pit downslope from the toe of the trench. Also, the

drier season is selected to have a lower water table as much possible in most cases.

It ought to be possible to find a proper work machine for excavation (e.g., backhoe, trackhoe, or bulldozer) in most study areas. But the trench location may be located too far from a settlement, or there may be no excavation machine around. In such cases, trenches can be dug by hand. Prentice et al. (2002) dug several hand-made trenches even up to several meters in length and depth in Mongolia, where there are no settlements nearby.

The excavation of a site also means the destruction of the local stratigraphy and structures. Therefore, any trenching should be done in the most preservative way to avoid damage both for the ongoing and any future studies at the same site. Moreover, excavation material should be backfilled carefully, especially on cultivated lands. For example, preservation of the natural soil can be critical and can be done by separating it from other excavated materials and putting it back at the end of the trench study.

Excavation: Choosing the Trench Orientation, Size, and Pattern

The choices for trench excavation depend on several factors such as the kind of material being trenched, the width of the deformation zone, the fault type, the relief of the site, the target depth, the stability of the trench walls, and the style of working.

Trench Orientation: Fault-Perpendicular and Fault-Parallel Trenches

The sense of fault displacement is the major criteria in positioning trenches before starting the excavation. Fault-perpendicular trenches are often used to locate fault zones and identify the recurrence patterns. Dip-slip paleo-events can be characterized in terms of recurrence and displacement by a single fault-perpendicular trench, especially where the deformation is localized in a narrow zone. However, multiple trenches, both fault-perpendicular and fault-parallel, are needed to measure the horizontal slip along strike-slip faults. The concept of 3D trenching was developed for this type of faulting, where both

recurrence and displacement history could be inferred. To reveal the slip history, several approaches could be used, such as (a) multiple trenches closely spaced; (b) successive exposures, which are dug orthogonally to the fault zone; or (c) two fault-parallel trenches on each side of the fault, exposing the offset features (piercing lines). In all of these approaches, at least one fault-perpendicular trench is excavated to locate the fault zone. For example, Marco et al. (2005) used buried stream channels as piercing lines to restore the slip history of individual and cumulative events at the Jordan Gorge Segment of the Dead Sea Fault.

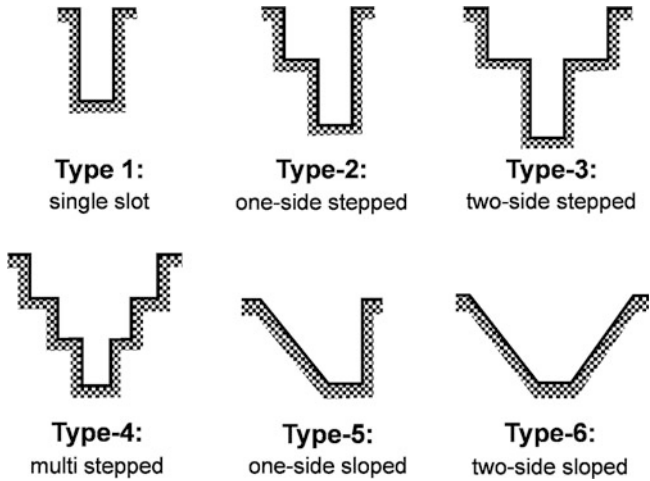
Trench Size

The main goal in trench studies is to expose maximum information with the most efficient excavation. Thus, determination of excavation dimensions plays a crucial role in the planning of the study time and budget. Depending on the width of the fault zone, the trench length may vary from a few meters to several hundred meters. Excavations are often started from foot-wall side and extended toward the hanging wall across dip-slip faults. The trench length is relatively shorter across strike-slip faults. The width is preferred to be large enough for the maintenance of a safe and comfortable working space, especially for taking photos of the trench walls. Depth mostly depends on the stability of walls and water table. Deeper trenches increase the probability of exposing a longer seismic history. Safety precautions are also very important in planning of the trench dimensions. (Please see the relevant section in this article). On the other hand, the geometry of trench walls, either oriented vertically or inclined, should have a planar shape as much as possible for a perfect projection surface to log.

Trench Patterns

Trench arrangement usually differs according to the structural and depositional properties of sites. The most appropriate arrangement is mostly selected according to the water table, stability of trench walls, type of deposits, logistic conditions, and the budget. Figure 1 shows cross sections of

Paleoseismic Trenching, Fig. 1 Cross sections showing different types of trench patterns. Type 1, single slot; Type 2, one-side stepped; Type 3, two-side stepped; Type 4, multisteped; Type 5, one-side sloped; Type 6, two-side sloped (or open pit) (Modified and redrawn after McCalpin 1989)



Paleoseismic Trenching, Fig. 2 Sample photos from different trench types: (a) single slot (Photo by Dr. Aynur Dikbaş) and (b) two-side stepped (Photo by Dr. Taylan Sançar)

different trench types (Type 1 to Type 6). The most often used pattern is single-slot (California-style, Type 1) trench, which is easily shored with a minimum material excavation and is least time consuming (Fig. 2a). This style of trenching is also the cheapest one, but also may acquire hydraulic or wooden shores against collapse of trench walls, especially in loose sediments. On the other hand, it is obvious that deeper sections increase the possibility for the recovery of the longer seismic history. Thus, paleoseismologists tend to excavate deeper trenches in suitable

sites (Fig. 2b). Deeper trenches, Type 2 to Type 6, expose more faulting events, but they are expensive and time consuming. Multisteped or sloped patterns are mainly precautions against trench collapse.

Safety Precautions

Trench walls tend to collapse easily under various geological and morphological conditions. Fault zones easily reduce the cohesion of the material by crushing and creating open voids. High water table can evoke the collapse of unconsolidated

young sediments. Cohesion differences between stratigraphic units in a section may also cause caving inside the cohesionless unit. Precautions, which are listed under two subtitles below, include only general and most common suggestions. Local regulations must always be consulted and followed. BS 5930 “the code of practice for site investigations” in the UK, for example, gives detailed guidance on legal, environmental, and technical matters relating to site investigation. It should be well understood that trenching is dangerous and loss of life can occur if undertaken unsafely. All available and mandatory risk assessments should be undertaken prior to work commencing and then reviewed once the trench is open to take into account any unexpected ground conditions.

Safety Precautions Inside the Trench

Trenching in loose materials or high water table may cause collapsing or caving of trench walls. Stepped- or sloped-type trenches are relatively safer with respect to single-slot type based on local conditions. For example, according to safety regulations in the USA, the vertical walls exceeding 1.5 m high in Type A or 1.2 m in Type B and Type C soils must contain another 1.5 m-wide horizontal bench on each side (OSHA 1989). In addition to benching, the stability of walls can be supported by hydraulic or wooden shores (Fig. 2a). If trench walls contain blocky materials, fallen boulders or cobbles can easily harm or even kill researches who work inside. Hard hats should always be worn in any trench over 1 m deep, which can be deadly under the right circumstances.

Safety Precautions Outside the Trench

Trenches with considerable depths are dangerous spots not only for domestic animals but also for the people who live nearby. An excavation area can attract especially kids and curious adults, and closing to the trench edges can be dangerous both for them and researchers inside. Paleoseismologists usually use safety tapes or simple fences to avoid such accidents. One member of the study team should also stay outside not only to inform cautious visitors but also to monitor any

progressive cracks on the ground following the excavation of the site.

Preparing Trench for the Study

Following the excavation and implementation of safety precautions, the walls must be perfectly cleaned enough to expose all stratigraphic and structural relationships. The shovel of work machines usually leaves many scars, clay packs, or pseudo-corrugations. Tools such as scrapers, masonry towels, hammers, brushes, and water sprays are used for the cleaning of different material types on the trench walls.

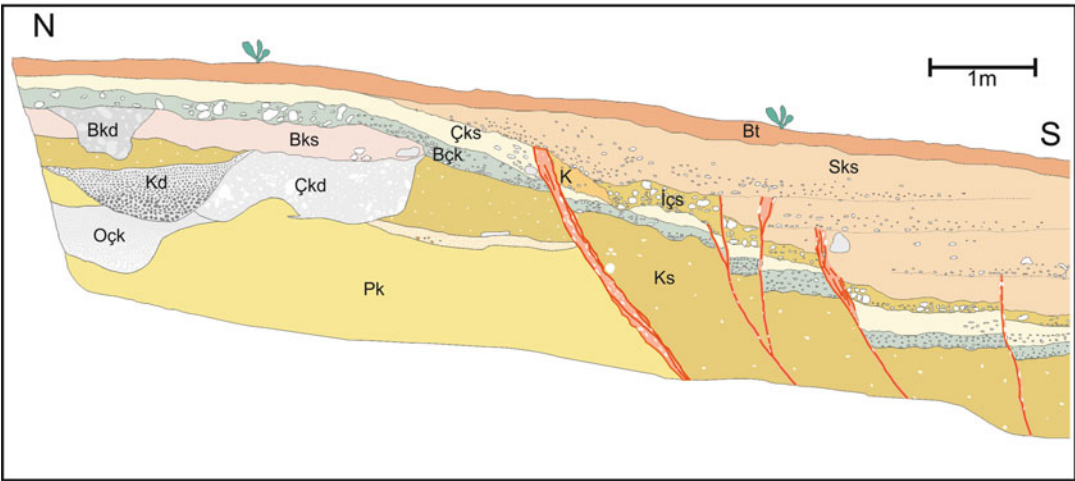
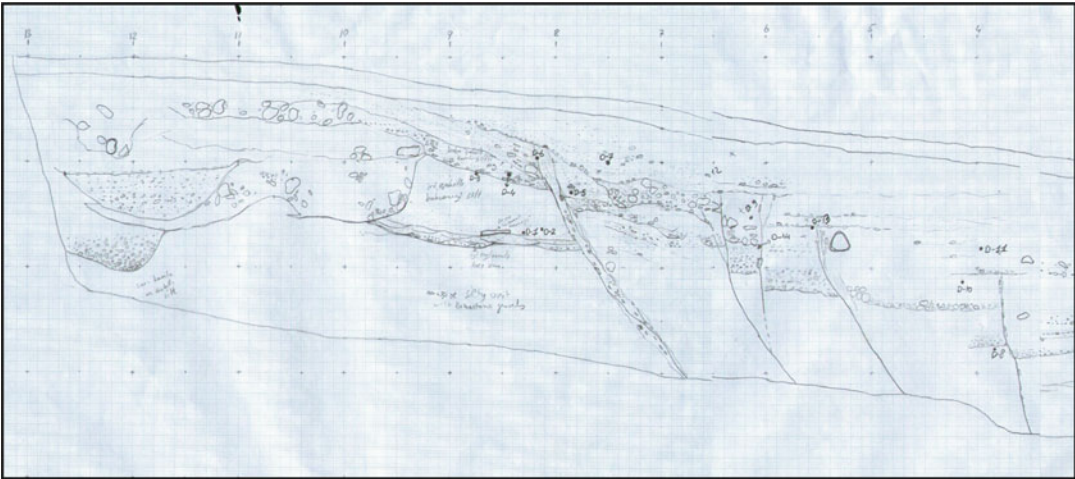
After a careful cleaning process, a reference grid is constructed for logging. A typical grid is composed of perfect vertical and horizontal lines, which are generally spaced 1 m (or 0.5 m) apart. Horizontality and verticality can be constructed and checked using a spirit level. Laser levelers, which increase the precision and save lots of time, are getting more common each day in trench studies. At the end of gridding, every horizontal and vertical reference point is labeled successively every 1 or 0.5 m (Fig. 3).

Logging: Recording the Stratigraphic and Structural Relationships

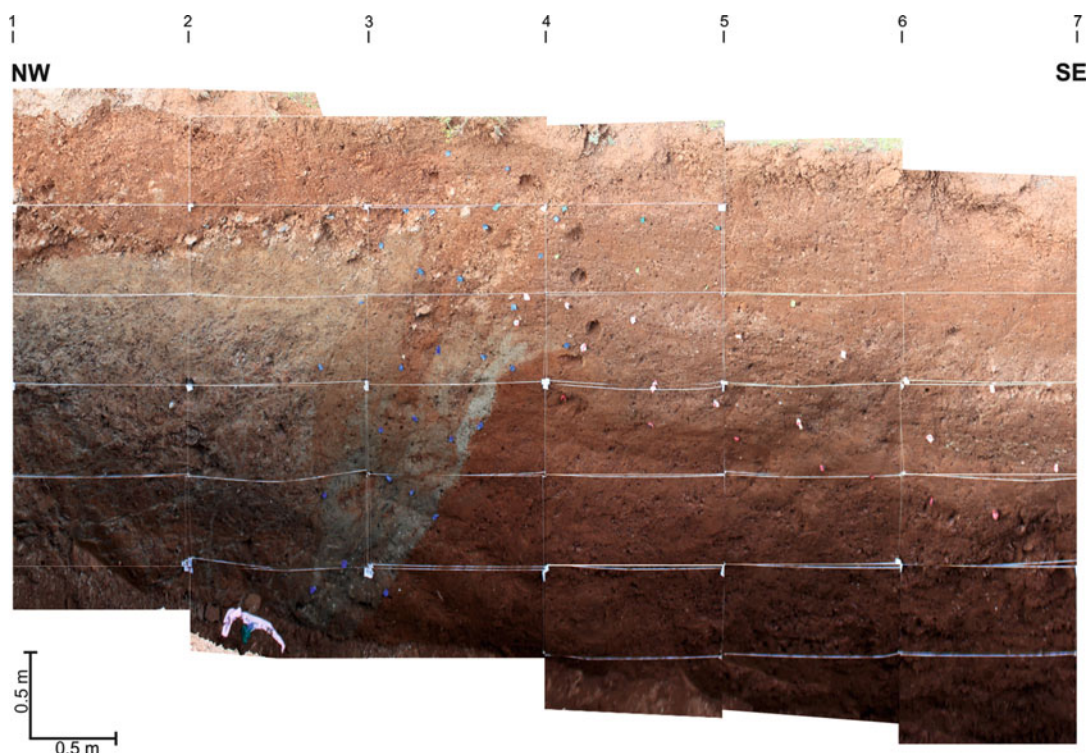
Trench logs are the key records, which contain almost all information regarding the stratigraphy and structures in a trench study. These logs not only document primary and/or secondary earthquake evidence, but they also show the precise coordinates of sampling points. In many trench studies, usually a single wall is logged. However, logging of both walls provides confirmation of structural features and also helps the 3D reconstruction of the distribution of stratigraphic units. There are two common logging methods: manual and photomosaic.

Squared paper is used in the traditional manual method, for which each critical stratigraphic horizon and structure is logged with plotting the measurements of critical points with respect to the reference grid (Fig. 4). An ordinary tape measure or a measuring rod with a spirit level is used in taking these coordinates. In manual logging, there must be at least two people working simultaneously for an efficient study.

Paleoseismic Trenching, Fig. 3 An ideal reference grid (Photo by Dr. Volkan Karabacak)



Paleoseismic Trenching, Fig. 4 Traditional manual log of a trench wall (Taken from Dr. Volkan Karabacak)



Paleoseismic Trenching, Fig. 5 Photomosaic of a trench wall (Photo by Dr. Volkan Karabacak)

While one person measures the horizontal and vertical distances from trench features to the nearest grid line, a second person plots the positions of these measurements on the squared paper with a scale.

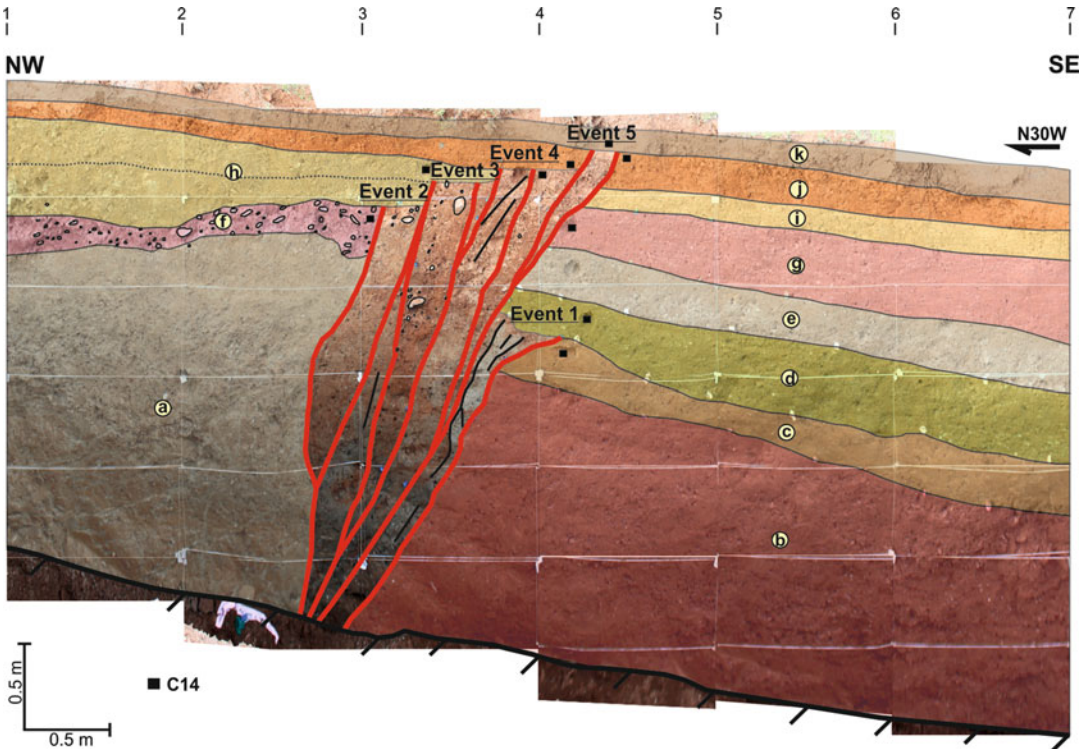
Especially after the development of the digital photography techniques, logging on the photo-mosaic of walls has almost become a standard in trench studies. The trench walls are photographed and combined into a mosaic covering the whole wall (Fig. 5). Fast orthorectification of photos provide precise representation of trench exposures.

In addition to these two common techniques, there are different methodological attempts in logging of a trench wall. These relatively new methods aim to provide more quantitative information about structures and the distribution of the depositional units by using such as spectral imaging (Ragona et al. 2006) or the magnetic susceptibility measurements of the trench walls (Fraser et al. 2009).

Irrelevant to which method is chosen during the logging, the final record is digitized by using vector-based drawing software in post-field studies. All stratigraphic units are labeled and shown with proper symbols and colors. In addition to the clear imaging of faults and event horizons, the reference grid is provided to give a true spatial sense (Fig. 6).

Types of Sediment-Structure Relations in the Identification of Paleoearthquakes

Destructive earthquakes, mainly with magnitude >6, often produce surface deformation, which depends on parameters such as the physical structure of the crust, the type of faulting, and the focal depth. In a properly selected trench site, it is highly probable to see surface faulting within the young deposits. A key objective of paleoseismologists is to identify and define event horizon(s) and recover



Paleoseismic Trenching, Fig. 6 A final log of a trench study (Taken from Dr. Volkan Karabacak)

the seismic history of the studied fault segment with dating of the relevant stratigraphic units.

Each trench has its own story. The historical evolution of the studied fault segment is hidden on the trench wall exposures. Even though there are some definite event structures for different fault types, all fault-related structures can be classified under two major titles: (a) on-fault (primary) and (b) off-fault (secondary or indirect) structures (Fig. 7). On-fault structures are mostly preferred in trench studies, due to the ambiguous nature of off-fault structures.

On-Fault Structures

The most commonly seen evidence for a past earthquake on the trench wall is upward termination of a fault against overlying strata (Fig. 7a). The thickness change of a layer on both sides of the fault reflects an erosional and depositional stage after an earthquake (Fig. 7b). A colluvial wedge is another clear evidence for a paleo-

surface rupture, where a vertical separation or offset occurs between faulted blocks (this is also possible in lateral fault movements in rugged areas) (Fig. 7c). A monoclinical-like structure develops above the faulted zone in environments with high sedimentation rates (Fig. 7d). Open fissures can often be seen both on extensional and strike-slip faults. They can be filled by younger infills, which mark an event on the trench wall (Fig. 7e). Disoriented pebble- or cobble-sized clasts, seen in coarse-grained layers or in shear zones within fine-grained sediments, usually indicate a surface deformation of an earthquake (Fig. 7f, g).

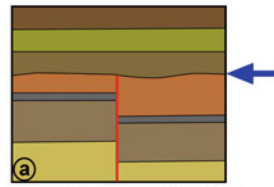
Off-Fault Structures

Where it is not possible to open a trench on the main fault zone, for a variety of reason, paleoseismologists will then look to examine secondary (or off-fault) structures, to reveal the paleoseismic history. Angular unconformities

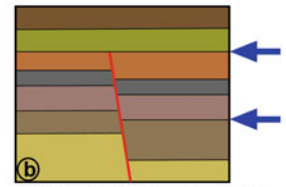
Paleoseismic Trenching,

Fig. 7 Earthquake indicators used in the identification of paleo-events. On-fault structures: (a) upward termination of a fault, which is overlain by undeformed stratum; (b) difference in thickness of strata at each side of the fault and downward growth of the displacement along the fault trace; (c) scarp-derived colluvium; (d) formation of monoclinical folding, where fast sedimentation covers the fault scarp; (e) open surface cracks, infilled with material of the overlying unit; (f) disordered pebbles covered by undeformed layer; and (g) sheared layer covered with unshered one. Off-fault structures: (h) angular unconformity within modern sediments; (i) sand boil, sand dike, and liquefied sand; (j) minor cracks and fissures with no offset; and (k) soft-sediment deformation in general that blankets with undeformed one (Modified and redrawn after Allen (1986); reprinted with permission from Active Tectonics (1986) by the National Academy of Sciences, Courtesy of the National Academy Press, Washington, DC)

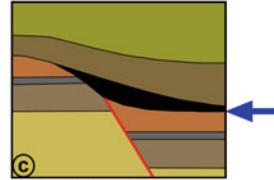
ON - FAULT



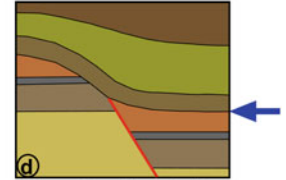
Upward fault termination



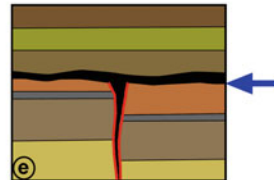
Different thickness of strata on both side of the fault



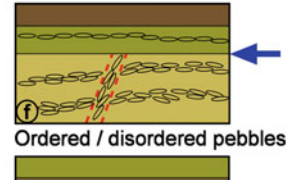
Colluvial wedge in erosional conditions



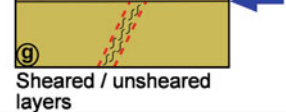
Warped sedimentation on fault scarp (no colluvium)



Infill to open crack

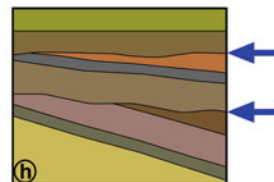


Ordered / disordered pebbles

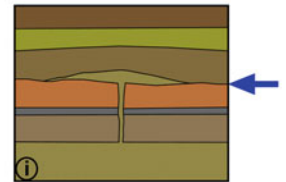


Sheared / unshered layers

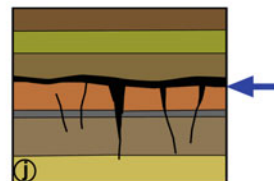
OFF - FAULT



Angular unconformity



Sand boil (blow), dike and liquefied sand



Minor (secondary) fractures and fissures



Soft-sediment deformation covered undeformed strata

within the modern sediments are important markers, indicating a nearby fault-related deformation (Fig. 7h). Sandblow (or boil), mud diapirism, liquefied sand, minor fractures, and

cracks are all soft-sediment deformational structures that are related to ground shaking and can be interpreted as indicators of paleo-earthquakes (Fig. 7i-k).

Dating Paleo-earthquakes

In addition to identification of paleo-earthquakes on trench walls, it is crucial to date important seismic horizons to model paleo-earthquakes and future risk. During or after the logging of stratigraphic horizons and structural features on the trench wall, horizons are sampled for dating to construct the temporal relationships. Especially event horizons, which cover earthquake-related structures, are the key levels in the recovery of the seismic history in a trench study. Late Quaternary deposits can be dated by a wide range of correlated, relative, numerical, and calibrated age methods. Among all, radiocarbon dating is the most common applied method in paleoseismological studies. The main materials are in situ organic compounds, such as bones, teeth, shells, charcoals, peats, chunks, and organic soils for radiocarbon dating. Advances in luminescence dating also introduced the application of this method in trench studies. Other dating methods of young sediments such as varve chronology, electron spin resonance, amino acid racemization, etc. can be used in suitable environments. Prior to sampling for dating, the most proper method should be determined, and the careful sampling techniques should be applied to avoid any contamination.

Advantages and Disadvantages of Trenching Studies

Paleoseismic trenching plays an important role in the reconstruction of a seismic history and makes an invaluable contribution to seismic risk assessments of a region (Gürpınar 2005). However, paleoseismology has advantages and limits like any other applied techniques. It is important to know all these aspects of paleoseismic studies to construct a sound project and solve a scientific problem.

The Advantages of Paleoseismic Trenching

- Paleoseismic trenching can provide the exact location of the fault (or deformational structures), which is very important information in urban or industrial planning.

- The reconstruction of the seismic history from well-chosen trench sites provides reliable information for seismic risk assessments.
- In addition to the location and timing of a paleo-event, well-located fault-perpendicular trenches for dip-slip faults or fault-parallel trenches for strike-slip faults provide the slip history (magnitude of displacement).

The Disadvantages of Paleoseismic Trenching

- Sometimes the destructive earthquakes do not produce any surface rupture along the source fault. These kinds of paleo-earthquakes do not leave their surface expressions on the Earth's surface, which make it impossible to identify anything in a trench study. Thus, paleoseismic trenching may only give information for large earthquakes that produce surface deformations.
- The distributed deformation, especially in thick water-saturated sediments, can be reflected as departed surface faulting after different events. Studies, especially with only a single trench under these circumstances, are resulted with missing recovery of total history.
- The thickness of stratigraphic units, which records the event history in a trench study, is directly affected by the sedimentation rate. Continuous sedimentation is always desired in trench studies; however, very fast sedimentation can cover event horizons with a thick pile of sediments, which prevent reaching the paleo-events.
- High water table usually prevents reaching deeper (older) stratigraphy.
- Logistics (i.e., permissions or environmental issues) may prevent excavating a perfect trench site, which is not ideally located or positioned.
- There may not be enough and suitable sampling material for dating based on environmental and/or sedimentary conditions.
- Measured displacements may also include post-seismic and/or creeping motions in addition to the coseismic slip.
- It is hard to identify all individual earthquakes along fast-moving faults (especially at overlapping sections of neighboring segments) due to limitation of dating methods.

Summary

In the developing world, mankind needs better life conditions, new settlements, and new industrial areas. It is well known that most part of the Earth's crust has a perpetual motion and that earthquakes are one of the most destructive natural hazards. Therefore, scientists have to find safe zones in which to live, especially at actively deformed regions, or provide information to mitigate against the seismic hazard. To achieve this, they have to know the exact location of the deformation zone and understand the characteristic behavior of the fault zone. Then, they can predict the future seismic risk of the area. Paleoseismological trenching is the most important tool to understand the nature of paleo-earthquakes and to predict the present and future behavioral characteristics of studied fault segments. The famous phrase of James Hutton "the present is the key to the past," then, can be improved to "the past is the key to the future" in the paleoseismic point of view. Of course, the term "past" covers mainly Holocene or Late Pleistocene, but not millions of years, while the term "future" embraces tens to thousands of years.

Trenching is a multidisciplinary study, including geology, geophysics, geomorphology, geoinformatics, geochronology, history, even mythology, etc. Pre-field, remote sensing studies provide important information on the location of the deformation zone and possible trench sites, which saves time, manpower, and budget. After the completion of desk and preliminary field studies, the first step is to find a convenient site for trenching. Important issues in proper location are continuous sedimentation, a low water table, presence of dating material, and sufficient logistic conditions (transportation, providing a work machine, etc.). Detailed geomorphological and geological maps are prepared to understand depositional or erosional conditions around trench sites. High-resolution micro-topographic maps give invaluable and precise information about effects of deformation and offset structures. Shallow geophysical methods are also important to

define exact deformation zone and the continuation of the fault in depth.

After deciding on the placement of the trench site, the orientation, size, and pattern of the excavation should be chosen based on scientific and logistic conditions. Safety precautions during and after excavation are very important both inside and outside of the trench. These precautions must be strictly applied until the closure of the trench. After excavation, the trench is prepared for a detailed study within the following steps: (1) The trench walls are cleaned and the traces of the work machine are removed; (2) the walls are put in a reference frame with vertical and horizontal lines (mainly by using a plastic string) in 1×1 or 0.5×0.5 m grids; and (3) stratigraphic horizons and critical structures are mapped onto squared paper scaled appropriately. Alternatively, systematic photographs and their photo-mosaics of trench walls can also be used for logging. Note that the identification of past (paleo-)earthquakes on trench walls needs careful, detailed, and accurate analyses of both exposed trench stratigraphy and structure. On-fault or off-fault deformation and structures (primary and/or secondary) are used to reveal seismic history of the studied fault segment. (4) Sample collection for dating of critical (event) horizons should be undertaken and is important in trench studies to model the timing of events and recurrence intervals. Finally, after the collection of all available data at the trench site, the excavated material is filled back carefully to restore the original environmental conditions as much as possible.

In conclusion, paleoseismological trenching, the direct observation method of historical earthquakes, provides both the exact location and history of the fault. Exposing faulted modern (or Holocene) sediments in suitable environments gives invaluable information about surface-rupturing earthquakes for a few hundreds to thousands of years. These paleoseismic data are very important in the planning of urban and industrial settlements and, most importantly, in providing a safer life for mankind.

Cross-References

- Archeoseismology
- Earthquake Recurrence
- Earthquake Return Period and Its Incorporation into Seismic Actions
- Luminescence Dating in Paleoseismology
- Paleoseismology
- Paleoseismology: Integration with Seismic Hazard
- Radiocarbon Dating in Paleoseismology
- Seismic Event Detection
- Tsunamis as Paleoseismic Indicators

References

- Allen CR (1986) Seismological and paleoseismological techniques of research in active tectonics. In: Wallace Chairman RE (ed) *Active tectonics: studies in geophysics*. National Academy Press, Washington, DC, pp 148–154
- Carpentier SFA, Green AG, Langridge R, Boschetti S, Doetsch J, Abächerli AN, Horstmeyer H, Finnemore M (2012) Flower structures and Riedel shears at a step over zone along the Alpine Fault (New Zealand) inferred from 2-D and 3-D GPR images. *J Geophys Res Solid Earth* 117(B2):B02406
- Cunningham D, Grebbly S, Tansey K, Gosar A, Kastelic V (2006) Application of airborne LiDAR to mapping seismogenic faults in forested mountainous terrain, southeastern Alps, Slovenia. *Geophys Res Lett* 33(20), L20308
- Ferry M, Meghraoui M, Girard J-F, Rockwell TK, Kozacı Ö, Akyüz S, Barka A (2004) Ground-penetrating radar investigations along the North Anatolian Fault near Izmit, Turkey; constraints on the right-lateral movement and slip history. *Geol Boulder* 32(1):85–88
- Fraser J, Pigati JS, Hubert-Ferrari A, Vanneste K, Avşar U, Altinok S (2009) A 3000-year record of ground-rupturing earthquakes along the Central North Anatolian Fault near Lake Ladik, Turkey. *Bull Seismol Soc Am* 99(5):2681–2703
- Geo Eye Elevating Insight (2013) About Geoeye-2. <http://launch.geoeeye.com/LaunchSite/about/Default.aspx>. Accessed 03 Aug 2013
- Gold PO, Cowgill E, Kreylos O, Gold RD (2012) A terrestrial lidar-based workflow for determining three-dimensional slip vectors and associated uncertainties. *Geosphere* 8(2):431–442
- Gürpınar A (2005) The importance of paleoseismology in seismic hazard studies for critical facilities. *Tectonophysics* 408(1–4):23–28
- Marco S, Rockwell TK, Heimann A, Frieslander U, Agnon A (2005) Late Holocene activity of the Dead Sea Transform revealed in 3D palaeoseismic trenches on the Jordan Gorge segment. *Earth Planet Sci Lett* 234(1–2):189–205
- McCalpin JP (1989) Current investigative techniques and interpretive models for trenching active dip-slip faults. In: Watters RC (ed) *Engineering geology and geotechnical engineering. Proceedings of the 25th symposium on engineering geology and geotechnical engineering*. A.A. Balkema, Rotterdam, pp 249–258
- McCalpin JP (ed) (2009) *Paleoseismology*, 2nd edn. Academic, Amsterdam
- Ortuño M, Masana E, García-Meléndez E, Martínez-Díaz J, Štěpančíková P, Cunha PP, Sohbati R, Canora C, Buylaert J-P, Murray AS (2012) An exceptionally long paleoseismic record of a slow-moving fault: the Alhama de Murcia fault (Eastern Betic shear zone, Spain). *Geol Soc Am Bull* 124(9–10):1474–1494
- Pantosti D, Pucci S, Palyvos N, Martini PMD, D'Addezio G, Collins PEF, Zabci C (2008) Paleoeearthquakes of the Duzce fault (North Anatolian Fault Zone): insights for large surface faulting earthquake recurrence. *J Geophys Res* 113:B01309
- Prentice CS, Kendrick K, Berryman K, Bayasgalan A, Ritz JF, Spencer JQ (2002) Prehistoric ruptures of the Gurvan Bulag fault, Gobi Altay, Mongolia. *J Geophys Res Solid Earth* 107(B12):2321
- Ragona D, Minster B, Rockwell T, Jussila J (2006) Field imaging spectroscopy: a new methodology to assist the description, interpretation, and archiving of paleoseismological information from faulted exposures. *J Geophys Res Solid Earth* 111(B10):B10309
- U.S. Occupational Safety and Health Administration (OSHA) (1989) Occupational safety and health standards – excavations: final rule. *Fed Reg* 29 CFR 1926, 54(209):45894–45991
- Zielke O, Arrowsmith JR (2012) LaDiCaoz and LiDARimager – MATLAB GUIs for LiDAR data handling and lateral displacement measurement. *Geosphere* 8(1):206–221

Paleoseismology

Sarah J. Boulton
Plymouth University, Centre for Research in
Earth Sciences, Plymouth University, Devon, UK

Synonyms

Ancient earthquakes; Earthquake geology;
Paleoseismology; Prehistoric earthquakes

Introduction

The aim of paleoseismology is to generate a record of past earthquakes (i.e., magnitude, recurrence interval, timing, etc.) from a range of geological observables preserved within a landscape. This is achieved by identifying features associated with a single paleoearthquake as opposed to the long-term deformation along a fault or within a basin. As such paleoseismology has been defined as “the study of prehistoric earthquakes, their timing, location and size” (McCalpin and Nelson 2009) and as “the study of the ground effects from past earthquakes as preserved in the geologic and geomorphic record” (Michetti et al. 2005). The paleoseismic record provides information on the local and regional deformation accommodated by slip on faults as well as data on patterns of seismicity, which can inform hazard assessments for future earthquakes. In addition to providing information on past earthquakes, paleoseismology can also constrain models of fault behavior and enhance our understanding of the influence of active faulting in the landscape. So although paleoseismology is part of the broader field of earthquake geology, it is distinct and separate from modern instrumental seismology, tectonic and structural geology (including neotectonics), historical seismology, geodesy, and tectonic geomorphology. However, paleoseismology does involve aspects of many of these disciplines as well as elements from stratigraphy, sedimentology, quaternary geology, archeology, remote sensing, and geochronometry.

Paleoseismology began to be recognized as a scientific discipline during the 1970s in the United States and Japan. Although pioneers such as G. K. Gilbert in North America, A. McKay in New Zealand, and B. Koto in Japan made important early descriptions of fault scarps and earthquake ruptures, these studies made little attempt to relate features observed directly along faults to determine specific number or magnitude of earthquakes that had taken place. Prior to the 1970s, historical records represented the sole evidence used in the assessment of past earthquake activity. However, the use of

historical records to investigate past earthquakes is problematic in areas where records only exist for a couple of centuries, such as in North America, and where the return period of the fault is much larger than the period of data collection. Even in Europe and Asia where historical records can go back thousands of years, ambiguity can exist on the identity and location of the actual seismogenic fault and/or the earthquake rupture, as fault damage can be spread over large geographic areas. Historical records may also be incomplete due to low population densities in affected areas or political unrest resulting in events going unreported and can suffer a range of biases, such as conflation, amalgamation, or exaggeration of earthquakes, and deliberate or accidental misinformation (Rucker and Niemi 2010).

Paleoseismology, by contrast, is a scientific discipline whose studies identify and document a range of landscape features and excavate paleoseismic trenches, in order to understand and determine reliable earthquake magnitudes and recurrence intervals of prehistoric earthquakes, as well as correlating historical records to geological features along specific structures. It is important to note that the period of time that prehistory (prior to contemporary written accounts) encompasses will vary by location. The associated discipline of archeoseismology that seeks to supplement geological observations with archeological remains and evidence in the geoarcheological record can also complement the understanding of individual earthquakes in regions with long histories of human inhabitancy. However, circular reasoning needs to be avoided in these situations; for example, geological or archeological evidence is used to interpret an earthquake, which is dated with reference to an earthquake catalog and then ends up in such a catalog to be used to date future discoveries (Rucker and Niemi 2010).

Evidence for Paleoeearthquakes

Paleoseismic methods (paleoseismology) are used to investigate surface deformation and,

thus, independently increase the understanding of historical earthquakes. These methods include the application of the Environmental Seismic Intensity scale (ESI 2007), a scheme developed to quantify the intensity distribution of modern and historic earthquakes using the dimensions of the rupture (length, width, area, volume) and secondary features in the natural landscape (Michetti et al. 2007). However, for an earthquake to generate evidence that can be preserved at the surface, the magnitude of the earthquake must be large, generally considered to be $>6.5 M_w$ (moment magnitude) (McCalpin and Nelson 2009) or intensity VIII on the ESI 2007 scale (Michetti et al. 2007). Smaller earthquakes do not often cause sufficient surface deformation that can subsequently be preserved. This is not always the case though; for example, some large earthquakes that nucleate near the brittle-ductile transition [i.e., 1989 Loma Prieta $M 6.9$ (California) and 2001 $M_w 7.6$ Bhuj (India) earthquakes] have little or no surface expression (Yeats 2007). In contrast, some smaller earthquakes occasionally do cause surface rupture, such as in Sicily (Italy) where faults that generate $M_w 3.0$ – 5.0 earthquakes can produce surface faulting even during shallow events (Azzaro et al. 2000). Therefore, the local tectonic setting and regional stress regime should always be taken into account when undertaking paleoseismic studies.

The initial identification of potentially seismogenic faults through the identification of geomorphic features that form the “seismic landscape” is the first stage of a paleoseismic investigation (Michetti et al. 2005). A seismic landscape will have its topography, geology, and stratigraphy determined in part by the style of faulting, rate of tectonic activity, thickness, and rheology of the seismogenic layer; the prevailing climate is also critical since this controls rates of deposition and erosion that progressively shape geomorphic landforms over many seismic cycles. Thus, landscapes dominated by extensional faulting will exhibit features such as horsts and grabens and wind gaps on uplifted footwalls with triangular facets along the fault trace, whereas strike-slip faults will be characterized by features such as

shutter ridges, sag ponds, and offset or beheaded streams. Such assorted features can aid in the identification of faults both in the field and through the interrogation of satellite imagery, digital elevation models (DEMs), or aerial photographs. In turn, these features are often the focus of subsequent localized investigations such as geophysical surveys, paleoseismic ground investigation, or paleoseismic trenching.

A range of primary and secondary features can be associated with individual earthquakes and are used in paleoseismologic investigation once the active fault has been identified. Primary evidence forms instantaneously at the time of the earthquake due to slip along the fault plane and is therefore termed coseismic deformation. Secondary evidence can be caused by coseismic shaking or by subsequent (postseismic) erosion or deposition taking place at some time after the earthquake. Paleoseismic features can also be considered to be “on-fault” if along or near the fault or “off-fault” if they occur at some distance from the rupture. A further distinction can be made between geomorphic evidence and stratigraphic evidence of fault motion as observed in natural exposures and artificial paleoseismic trenches (McAlpin and Nelson 2009).

Typical examples of primary, on-fault evidence include: fault scarps, fissures, and folded, sheared, or displaced stratigraphic horizons (Table 1). Surface ruptures can be further divided into three categories: primary, subordinate, and sympathetic (Michetti 1994). Primary off-fault evidence can include tilted, uplifted, or submerged surfaces such as terraces (fluvial or coastal) or planation surfaces, drainage anomalies, change in coastal landforms, broken speleothems, and damage to man-made structures (both damage to ancient buildings and modern structures).

Secondary paleoseismic evidence consists of a great range of geomorphic and stratigraphic features including soft-sediment deformation, rock-falls, landslides, mud/sand volcanoes (also known as sand blows or boils), disturbed trees, and turbidite deposits, with little differentiation between the features that form on-fault and off-fault.

Paleoseismology, Table 1 Hierarchical classification of paleoseismic evidence with examples of stratigraphic and geomorphic features, modified from McCalpin and Nelson (2009)

Location	On-fault		Off-fault	
Timing	Coseismic	Postseismic	Coseismic	Postseismic
Primary paleoseismic evidence				
<i>Geomorphic features</i>	Surface ruptures	Colluvial aprons	Titled surfaces	Tectonic alluvial terraces
	1. Primary (i.e., fault scarps on main fault)	Afterslip modification to primary on-fault coseismic features	Uplifted shorelines	Afterslip modification to primary off-fault coseismic features
	2. Subordinate (ruptures branching from main fault)		Subsided shorelines	
	3. Sympathetic (ruptures on isolated faults)			
	Fissures			
	Folds			
	Mole tracks			
	Pressure ridges			
<i>Stratigraphic features</i>	Faulted strata	Scarp-derived colluvial wedges	Tsunami deposits and erosional unconformities caused by tsunamis	Erosional unconformities and deposits induced by uplift, subsidence, and tilting
	Folded strata	Fissure fills		
	Unconformities			
	Disconformities			
<i>Abundance of similar nonseismic features</i>	Few	Few	Some	Common
Secondary paleoseismic evidence				
<i>Geomorphic features</i>	Sand/mud volcanoes	Retrogressive landslides in the fault zone	Sand/mud volcanoes	Retrogressive landslides beyond the fault zone
	Landslides and lateral spreads in the fault zones		Landslides and lateral spreads beyond the fault zone	
	Disturbed trees and tree-throw craters		Disturbed trees and tree-throw craters	
			Fissures and Sackungen	
<i>Stratigraphic features</i>	Sand dykes and sills	Sediments deposited from retrogressive landslides	Sand dykes	Erosion or deposition in response to landscape disturbance (i.e., landslides, fissures, etc.)
	Soft-sediment deformation		Filled craters	
	Landslide toe thrusts		Soft-sediment deformation	
			Turbidites	
			Tsunamiites	
<i>Abundance of similar nonseismic features</i>	Some	Very common	Some	Very common

It should be noted that the distinction between primary and secondary off-fault evidence can be difficult to determine as seismic shaking may generate features that continue to evolve after the cessation of the earthquake. Furthermore, some features could represent both primary and secondary evidence. Tsunami deposits, for example, could be considered as primary evidence where the tsunami is produced directly by primary fault displacement, but could also be considered secondary evidence if say a landslide is triggered by earthquake shaking in a coastal location subsequently resulting in a tsunami. Secondary features are also more ambiguous than primary evidence as other geological mechanisms can result in their formation. McCalpin and Nelson (2009) used these classifications to define 16 categories (Table 1) of paleoseismic evidence that can commonly be distinguished in either geomorphic landforms or in the stratigraphic record. Therefore, using a combination of different lines of paleoseismic evidence tends to produce the most robust results in paleoseismic studies. As on-fault evidence often is clearer but provides little evidence of the strength of shaking or spatial extent of ground deformation, by contrast, geomorphic evidence can be preserved over wide areas, providing this spatial dimension to earthquake studies, yet it can be difficult to find the source of these off-fault features and age control may be lacking due to an absence of datable material.

As a result many paleoseismic studies focus on the excavation of trenches across faults, exposing offset and deformed strata resulting from seismogenic faulting. However, paleoseismic trenching studies commonly only take place at a few sites restricting the lateral applicability of the results. Crucially, these ground investigations can produce critical age control mainly through the use of radiocarbon dating or increasingly the application of optically stimulated luminescence (OSL) dating from material in offset stratigraphic horizons. It must also be noted that geological processes unrelated to earthquake faulting can produce features similar to coseismic and postseismic landforms and deposits and that erosion and sedimentation

subsequent to an earthquake will act to modify and conceal the effects of surface faulting.

Seismic Hazard Assessments from Paleoseismology

The main goal of paleoseismic studies is to determine the dimensions, magnitude, and timing of past earthquakes and the slip rate on seismic faults to inform deterministic or probabilistic seismic hazard assessments (SHAs) and seismic risk assessments. The most common way of determining the magnitude of past earthquakes is to use primary and secondary evidence to determine the surface rupture length of the past event and the amount of displacement (maximum or average slip) that occurred. These parameters can then be compared to worldwide catalogs of historical earthquakes and rupture lengths to calculate the probable magnitude of the paleoearthquake (Wells and Coppersmith 1994). Similarly, evidence for the area of displacement on a fault, or the seismic moment, can be used to estimate the magnitude of paleoearthquakes. However, these methods produce uncertain results due to the complex nature of faulting and the widespread use of global relationships that might not hold true on a regional scale. By contrast, spatial distributions of secondary evidence may reveal patterns of ground motion intensity that can be used to infer earthquake magnitudes. These observations can be used in a framework to determine the relative intensity of an earthquake, such as the ESI 2007 scale that combines observations of primary and secondary evidence on a scale of I–XII, with earthquakes intensity IV and above producing observable ground deformation of increasing severity (Michetti et al. 2007).

There are a number of problems with this approach. One is that primary and secondary features of the same age (within error) observed in a stratigraphic or geomorphic context are assumed to be the result of a single earthquake. However, earthquakes are observed to often cluster in time and space, and such earthquake sequences cannot be easily separated in the geological record due to the uncertainties in dating

methods resulting in an overestimation of the earthquake magnitude. Displacement is also not constant along the strike of the fault, going from zero displacement at the fault tips to a maximum near the center, which is often modeled as a symmetrical profile in seismic hazard assessment but could be an irregular shape – this could lead to bias in displacement reconstructions.

The earthquake recurrence interval (the time elapsed between any two paleoearthquakes) on a fault is also a key component of characterizing the seismic hazard. Faults that rupture over long recurrence intervals in larger events can be less hazardous than faults that rupture more frequently in moderate events, as the likelihood of impacting a critical facility is lower for the long return earthquake. Recurrence intervals can be calculated in one of two ways. The less preferred way is to calculate the average recurrence interval over multiple paleoearthquakes, a method commonly used in neotectonic studies. The preferred alternative is to individually date each paleoearthquake “event horizon” to determine the actual time between each earthquake. It is important to appreciate that the time since the last earthquake on a fault (the elapsed time) does not represent a recurrence interval because it does not bound two earthquakes. Also, when determining recurrence intervals, it is important to assess the degree of underrepresentation or overrepresentation in the seismic record (McCalpin and Nelson 2009). Underrepresentation results from an incomplete preservation record as a result of numerous factors that can affect the potential for preservation in a site as well as subsequent tectonic activity obscuring earlier events. The contrasting issue of overrepresentation is less often considered but can occur when nonseismic features are interpreted as coseismic or postseismic evidence. Overrepresentation can also occur if the paleoseismic record is well preserved during a period of earthquake clustering, which in turn is extrapolated over longer periods and thereby leads to an erroneously large number of past earthquakes being inferred (McCalpin and Nelson 2009). Therefore, it is important to consider the length of the window of observation and compare short-term slip

rates with the long-term average to assess the applicability of results.

Fault slip rate is the displacement of the fault over time and, as implied above, can apparently differ when considered at different timescales. The true slip rate is that considered over one or more closed (i.e., complete) seismic cycles, whereas an apparent slip rate is obtained if the slip rate is calculated over a time period containing complete and incomplete seismic cycles. However, due to variations in the displacement and time span between earthquakes, these measurements normally indicate that the slip rate varies between earthquake cycles and can be much higher than the long-term average during earthquake clustering or much lower during periods of fault quiescence.

These various parameters of fault and earthquake behavior can then be used to undertake seismic hazard assessments (SHAs); there are two main types of SHAs, deterministic seismic hazard assessment and probabilistic seismic hazard assessment.

Deterministic Seismic Hazard Assessment

Deterministic seismic hazard assessments (DSHAs) calculate the worst-case scenario for a site using the maximum potential magnitude for the largest active fault (the “controlling fault”) and the distance from that fault to the site in question as two parameters that combine (using an attenuation equation) to determine the potential ground motion at any point from the fault. The method also requires knowledge of the rock or soil properties at the site. Yet, DSHAs do not take into account the time probability of an earthquake occurring during the life time of the facility, so data on the recurrence interval and slip rate on the controlling fault is irrelevant.

Probabilistic Seismic Hazard Assessment

Probabilistic seismic hazard assessments (PSHAs), by contrast, calculate the worst seismic effects for a site using both the time-independent factors (magnitude and distance) used in DSHA combined with the time-dependent factors of slip rate and recurrence interval for the controlling fault. Therefore, this method determines both

the probability of an earthquake or earthquakes occurring near the critical facility within a defined time period (normally related to the design life of the facility) and also what the effects of these earthquakes would be on the facility. Currently, PSHA is the most commonly used method having generally replaced DSHA for most facilities.

Summary

Paleoseismology seeks to identify and appraise prehistoric or pre-instrumental earthquakes through the observation, documentation, and interpretation of geomorphic and stratigraphic evidence formed on, or near, earthquake ruptures as a result of primary rupture or secondary shaking. Such investigations have greatly improved our understanding of seismogenic fault behavior, revealing the segmentation of major faults (i.e., Schwartz and Coppersmith 1984), the temporal clustering of earthquakes in a nonperiodic fashion, the scaling relations between earthquake recurrence intervals and fault slip rates, and the tendency for faults to have characteristic rupture patterns (i.e., amount of slip: rupture length, magnitude). Of course, paleoseismological studies have also revealed periodic, nonscalar, and noncharacteristic earthquake faults and rigorous field studies are needed to test these apparent conflicting relationships. Variables determined from paleoearthquake studies (magnitude, slip, recurrence intervals) can be used to undertake seismic hazard assessments (SHA), either as a deterministic approach to calculate the worst-case scenario or a probabilistic approach to calculate the probability of an earthquake of a given size occurring within a given time period. However, since many paleoseismic data sets are too incomplete to precisely estimate fault slip rates or recurrence intervals, the contribution of paleoseismology to seismic hazard remains inaccurate (Grant 2002). Refining the record of past earthquakes is crucial for the improvement of earthquake-related hazard (seismic, tsunami, landslide, etc.) assessments for urban planning and critical facilities (i.e., nuclear power stations,

dams, chemical/petroleum facilities). Because even in well-constrained areas, lessons can still be learned by refining historical data with geological observations, as witnessed by the 2011 Tohoku (Japan) earthquake, where geological evidence of past megathrust earthquakes was available but not integrated into the existing hazard planning (Silva et al. 2011).

Cross-References

- ▶ [Archeoseismology](#)
- ▶ [Earthquake Magnitude Estimation](#)
- ▶ [Earthquake Mechanisms and Tectonics](#)
- ▶ [Earthquake Recurrence](#)
- ▶ [Intensity Scale ESI 2007 for Assessing Earthquake Intensities](#)
- ▶ [Luminescence Dating in Paleoseismology](#)
- ▶ [Paleoseismic Trenching](#)
- ▶ [Paleoseismology and Landslides](#)
- ▶ [Paleoseismology of Glaciated Terrain](#)
- ▶ [Paleoseismology of Rocky Coasts](#)
- ▶ [Paleoseismology: Integration with Seismic Hazard](#)
- ▶ [Radiocarbon Dating in Paleoseismology](#)
- ▶ [Remote Sensing in Seismology: An Overview](#)
- ▶ [Seismic Risk Assessment, Cascading Effects](#)
- ▶ [Site Response for Seismic Hazard Assessment](#)
- ▶ [Tsunamis as Paleoseismic Indicators](#)

References

- Azzaro R, Bella D, Ferreli L, Maria Michetti A, Santagati F, Serva L, & Vittori E (2000) First study of fault trench stratigraphy at Mt. Etna volcano, Southern Italy: understanding Holocene surface faulting along the Moscarello fault. *Journal of Geodynamics*, 29(3):187–210
- Grant LB (2002) Paleoseismology. In: Lee WHK, Kanamori H, Jennings PC, Kisslinger C (eds) *International handbook of earthquake and engineering seismology*, vol 81. AIASPEI, Elsevier, Amsterdam pp 475–490
- McCalpin JP, Nelson AR (2009) Introduction to paleoseismology. In: McCalpin JP (ed) *Paleoseismology*. vol 95, 2nd edn, International geophysics series. Elsevier, Amsterdam, pp 1–29

- Michetti AM (1994) Coseismic surface displacement vs. magnitude: relationships from paleoseismological analyses in the Central Apennines (Italy). In: Special issue 'proceedings of the CRCM'93', Journal Geodetic Society of Japan, Kyoto, pp 375, 380
- Michetti AM, Audemard FA, Marco S (2005) Future trends in paleoseismology: integrated study of the seismic landscape as a vital tool in seismic hazard analyses. *Tectonophysics* 408:3–21
- Michetti AM, Esposito E, Guerrieri L, Porfido S, Serva L, Tatevossian R, Vittori E, Audemard F, Azuma T, Clague J, Commerci V, Gurpinar A, McCalpin JP, Mohammadioun B, Morner NA, Ota Y, Roghazin E (2007) Environmental seismic intensity scale-ESI 2007. *Memorie Descrittive della Carta Geologica d'Italia* 74:7–54
- Rucker JD, Niemi TM (2010) Historical earthquake catalogues and archaeological data: achieving synthesis without circular reasoning. In: Sintubin M, Stewart IS, Niemi TM, Altunel E (eds) *Ancient earthquakes*. Geological Society of America special paper, vol 471. Geological Society of America, Boulder, pp 97–106
- Schwartz DP, Coppersmith KJ (1984) Fault behavior and characteristic earthquakes: examples from the Wasatch and San Andreas fault zones. *J Geophys Res Solid Earth* 89(B7):5681–5698
- Silva PG, Sintubin M, Reicherter K (2011) New advances in studies of earthquake archaeology and palaeoseismology. *Quat Int* 242:1–3
- Wells DL, Coppersmith KJ (1994) New empirical relationships among magnitude, rupture length, rupture width, rupture area and surface displacement. *Bull Seismol Soc Am* 4(84):975–1002
- Yeats RS (2007) Paleoseismology: why can't earthquakes keep on schedule? *Geology* 35:863–864

Paleoseismology and Landslides

Randall W. Jibson
U.S. Geological Survey, Golden, CO, USA

Synonyms

Earthquake-triggered landslides; Seismic ground failure; Slope-stability analysis

Introduction

Most moderate to large earthquakes trigger landslides (Fig. 1). In many environments, landslides preserved in the geologic record can be analyzed



Paleoseismology and Landslides, Fig. 1 Madison Canyon landslide, triggered by the 1959 Hebgen Lake, Montana, earthquake (M_w 7.1). Strong shaking caused $28 \times 10^6 \text{ m}^3$ of rock to slide into the canyon, which dammed the river and created a lake more than 60 m deep. Slide scar at left is 400 m high, debris is as thick as 67 m in valley axis, and slide debris traveled 130 m up the right valley wall. Twenty-eight people were killed by the slide (Photograph courtesy of J.R. Stacy, U.S. Geological Survey Photographic Library, photo no. 209a)

to determine the likelihood of seismic triggering. If evidence indicates that a seismic origin is likely for a landslide or group of landslides, and if the landslides can be dated, then a paleoearthquake can be inferred, and some of its characteristics can be estimated. Such paleoseismic landslide studies thus can help reconstruct the seismic shaking history of a site or region (Jibson 2009).

Paleoseismic landslide studies differ fundamentally from paleoseismic fault studies. Whereas fault studies seek to characterize the movement history of a specific fault, landslide studies characterize the shaking history of a site or region irrespective of the earthquake source. In regions that contain multiple seismic sources and in regions where surface faulting (► [Seismic Actions due to Near-Fault Ground Motion](#)) is absent, paleoseismic ground-failure studies thus can be valuable tools in hazard and risk studies that are more concerned with shaking hazards

than with interpretation of the movement histories of individual faults. In fact, paleoseismic studies in some parts of the world typically rely more on ground failure than on surface fault ruptures.

The practical lower bound earthquake that can be interpreted from paleoseismic landslide investigations is about magnitude 5–6 (► [Earthquake Magnitude Estimation](#)). This range is comparable or perhaps slightly lower than that for paleoseismic fault studies. Obviously, however, larger earthquakes tend to leave much more abundant and widespread evidence of landsliding than smaller earthquakes; thus, available evidence and confidence in interpretation increase with earthquake size.

Paleoseismic landslide analysis involves three steps: (1) identify a feature as a landslide, (2) date the landslide, and (3) determine if the landslide was triggered by earthquake shaking. This article addresses each of these steps and discusses methods for interpreting the results of such studies. Only subaerial landslides are discussed here; submarine landslides are analyzed using different methods.

In this article, *landslide* is used as a generic term to include all types of downslope movement of earth material, including types of movement that involve little or no true sliding. Thus, rock falls, debris flows, etc., are considered types of landslides. The classification system of Varnes (1978) is used, which categorizes landslides by the type of material involved (soil or rock) and by the type of movement (falls, topples, slides, slumps, flows, or spreads). Other modifiers commonly are used to indicate velocity of movement, degree of internal disruption, state of activity, and moisture content.

Identifying Landslides

Identifying surface features as landslides can be relatively easy for fairly recent, well-developed, simple landslides. Older, more degraded landslides or those having complex or unusual morphologies can be more difficult to identify. In general, landslides are identified by anomalous

Paleoseismology and Landslides, Table 1 Relative abundance of earthquake-induced landslides

Abundance	Landslide type
Very abundant	Rock falls
	Disrupted soil slides
	Rock slides
Abundant	Soil lateral spreads
	Soil slumps
	Soil block slides
	Soil avalanches
Moderately common	Soil falls
	Rapid soil flows
	Rock slumps
Uncommon	Subaqueous landslides
	Slow earth flows
	Rock block slides
	Rock avalanches

Note: Data from Keefer (1984). Landslide types use nomenclature of Varnes (1978) and are listed in decreasing order of abundance

topography, including arcuate or linear scarps, backward-rotated masses, benched or hummocky topography, bulging toes, and ponded or deranged drainage. Abnormal vegetation type and age also are common.

Earthquakes can trigger all types of landslides, and all types of landslides triggered by earthquakes also can occur without seismic triggering. Therefore, an earthquake origin cannot be determined solely on the basis of landslide type. However, some types of landslides tend to be much more abundant in earthquakes than other types. Table 1 shows the relative abundance of various types of earthquake-triggered landslides. Overall, the more disrupted types of landslides are much more abundant than the more coherent types of landslides. Also, most earthquake-induced landslides occur in intact materials rather than in preexisting landslide deposits; thus, the number of reactivated landslides is small compared to the total number of landslides triggered by earthquakes. Earthquake-triggered landslides most commonly occur in materials that are weathered, sheared, intensely fractured or jointed, or saturated.

Sackungen (ridge-crest troughs) are a somewhat controversial type of ground failure that has been related, in some cases, to seismic

shaking. Sackungen are identified by one or more of the following: (1) grabens or troughs near and parallel to ridge crests of high mountains, (2) uphill-facing scarps a few meters high that parallel the topography, (3) double-crested ridges, and (4) bulging lower parts of slopes.

Determining Landslide Ages

Paleoseismic interpretation requires establishing the numerical age of a paleoearthquake. In the case of earthquake-triggered landslides, this means that dating landslide movement is required. Several methods for dating landslide movement can be used; some are similar or identical to those used for dating fault scarps, while others are unique to landslides.

Different types of landslides could be datable by different methods, depending on a variety of factors such as distance of movement, degree of internal disruption, landslide geometry, type of landslide material, type and density of vegetation, and local climate. Ideally, multiple, independent dating methods should be used to increase the level of certainty of the age of landslide movement.

Historical Methods

Some old landslides might have been noted by local inhabitants or could have damaged or destroyed human works or natural features. In some parts of the world, potentially useful historical records or human works extend back several hundreds or thousands of years. For fairly recent events, comparing successive generations of topographic maps or aerial photographs can bracket the time period in which mappable landslides first appeared.

Dendrochronology

Dendrochronology can be applied to date landslide movement in several ways. At the simplest level, the oldest undisturbed trees on disrupted or rotated parts of landslides should yield reasonable minimum ages for movement. On rotational slides that remained fairly coherent, preexisting trees that survived the sliding will have been

tilted because of headward rotation of the ground surface; if both tilted and straight trees are present on such landslides, the age of slide movement is bracketed between the age of the oldest straight trees and the youngest tilted trees. Using this simple application of dendrochronology to date coherent translational slides is more difficult because trees can remain upright and intact even after landslide movement. On all types of landslides, trees growing from the surface of the scarp will yield minimum ages of scarp formation, from which the age of slide movement can be interpreted. In some cases, trees killed by landslide movement will be preserved and can thus yield the exact date of movement.

A more sophisticated application of dendrochronology involves quantitative analysis of growth rings. For trees that have survived one or more episodes of landslide movement, such analysis can be used to identify and date reaction wood (eccentric growth rings), growth suppression, and corrosion scars, which might be evidence of landslide movement. Some landslides block stream drainages and form dams that impound ponds or lakes. Inundation of areas upstream from landslide dams can drown trees that can be dated dendrochronologically.

Radiometric and Cosmogenic Dating

Radiometric dating ([► Radiocarbon Dating in Paleoseismology](#)) (most commonly using ^{14}C) can be used in a variety of ways to date organic material buried by landslide movement. Landslide scarps degrade similarly to fault scarps, and so colluvial wedges at the bases of landslide scarps might contain organic material that can be retrieved by trenching or coring and dated radiometrically. Fissures on the body of a landslide, particularly near the head where extension can take place, also can trap and preserve organic matter. If the landslide mass is highly disrupted, as in rock or soil falls or avalanches, then some vegetation from the original ground surface might have become mixed with the slide debris; such organic material excavated from slide debris can be dated radiometrically. At the toes of landslides, slide material commonly is deposited onto undisturbed ground; if this original ground

surface can be excavated beneath the toe of a slide, buried organic material from this surface can be dated to indicate the age of initial movement.

Sag ponds commonly form on landslides, and organic material deposited in such ponds can be dated radiometrically. Organics at the base of the pond deposits should yield reliable dates of pond formation.

Vegetation submerged from inundation of areas upstream from landslide dams also can be dated radiometrically. Similarly, landslides into lakes can submerge and kill vegetation that can be dated.

Rock-fall and rock-avalanche deposits can be dated cosmogenically (► [Luminescence Dating in Paleoseismology](#)) if the surface of the deposit has been relatively stable since the time of emplacement. This method measures the amount of time that specific types of mineral grains have been exposed to cosmic radiation. The assumption is that a significant proportion of the material on the surface of a rock-fall or rock-avalanche deposit was newly fractured and exposed when the landslide occurred.

Lichenometry

Lichenometry – analysis of the age of lichens based on their size – can be used to date rock-fall and rock-avalanche deposits. By measuring lichen diameters on rock faces freshly exposed at the time of failure, numerical ages can be estimated by assuming that lichens colonized the rock face in the first year after exposure. Because rock-fall and rock-avalanche deposits typically include abundant rocks having freshly exposed faces, numerous samples generally can be taken to create a database for the statistical analysis required by lichenometry. Lichenometric ages must be calibrated at sites of known historical age or by comparison with other numerical dating techniques. Lichenometric dating is subject to considerable uncertainty, however, because several decades can elapse before lichens colonize a fresh rock exposure, and lichens might never colonize unstable landslide deposits on very steep slopes.

Weathering Rinds

For a given climate and rock type, measuring the thickness of weathering rinds can be used to date when rocks were first exposed at the ground surface. For rock falls and rock avalanches and for other landslides whose movement exposed rock fragments at the ground surface, measuring the thickness of weathering rinds can be used to date landslide movement. Determining which rock surfaces were initially exposed at the time of landsliding can be difficult, but if a sufficiently large number of samples can be measured, consistent statistical results of predominant ages that relate to landslide movement can be obtained.

Pollen Analysis

Analysis of pollen in deposits filling depressions on landslides can yield both an estimated age of initial movement and, in some cases, a movement history through time. Such analyses assume that sediment deposition and incorporation of pollen occur immediately following landslide movement and that local climatic and vegetation variations can be accounted for. Pollen samples from the buried ground surface beneath the toes of landslides also have potential for use in dating landslide movement.

Geomorphic Analysis

Landslides are disequilibrium landforms that change through time more rapidly than surrounding terrain. By analyzing the degree of degradation of landslide features such as scarps, ridges, sags, and toes, relative ages can be assigned to various landslides. Criteria for such relative age classification might include degree of definition of landslide features, soil development, tephra cover, stream dissection, preservation of vegetation killed by movement, and drainage integration.

Models of fault-scarp degradation also have potential application in landslide dating because landslide scarps should behave similarly to fault scarps. Several approaches to morphologic fault-scarp dating have been proposed, all of which require calibration for various parameters such as climate and scarp material. Scarp degradation commonly is modeled as a diffusion process, in

which degradation rate varies in time and is a function of slope angle, which represents the degree to which the scarp is out of equilibrium with the surrounding landscape.

Analysis of soil-profile development also is a potential tool for dating landslides. New soil profiles will begin to develop on disrupted landslide surfaces. If such surfaces can be identified, dating the newly developed soil profile will indicate the age of movement.

Interpreting an Earthquake Origin for Landslides

Interpreting an earthquake origin for a landslide or group of landslides is by far the most difficult step in the process, and methods and levels of confidence in the resulting interpretation vary widely. This section summarizes several basic approaches that can be used to interpret the seismic origin of landslides.

Regional Analysis of Landslides

Many paleoseismic landslide studies involve analysis of large groups of landslides rather than individual features. The premise of these regional analyses is that a group of landslides of the same age, scattered across a discrete area, probably was triggered by a single event of regional extent. In an active seismic zone, that event commonly is inferred to be an earthquake. Such an interpretation could be justified in areas where landslide types and distributions from historical earthquakes have been documented and can be used as a standard for comparison. In areas where such historical observations are absent, assuming an earthquake origin for landslides of synchronous age is much more tenuous, primarily because large storms also can trigger widespread landslides having identical ages and spatial distributions.

Differentiating between such groups of storm- and earthquake-triggered landslides might be possible using a statistical approach to characterize the distribution of steep slopes in a region. Storm-triggered landslides form most commonly near the bases of slopes and thus tend to form

landscapes characterized by steep inner gorges. Earthquake-triggered landslides, on the other hand, tend to form either near ridge crests or more uniformly across the entire reach of slopes; corresponding landscapes generally lack well-developed inner gorges. These landscape patterns provide supportive evidence of the possible seismic origin of landslides in a region, but they cannot be used to definitively determine the origin of any specific landslide.

Some criteria to support a seismic origin for landslides include (1) ongoing seismicity in the region that has triggered landslides; (2) coincidence of landslide distribution with an active fault or seismic zone; (3) geotechnical slope-stability analyses showing that earthquake shaking would have been required to induce slope failure (discussed in detail subsequently); (4) presence of liquefaction features associated with the landslides; (5) correlation between the elongation of a landslide distribution and the location and dimensions of the seismogenic faults in a region; and (6) landslide distribution that cannot be explained solely on the basis of geologic or geomorphic conditions. Obviously, the more of these criteria that are satisfied, the stronger the case for seismic origin.

Landslide Morphology

Some landslides have morphologies that strongly suggest triggering by earthquake shaking. For example, stability analyses of landslides on low-angle basal shear surfaces show that they generally form much more readily under the influence of earthquake shaking than in other conditions. Landslides that formed as a result of liquefaction of subsurface layers also are much more likely to have formed seismically than aseismically. Slides that form as a result of intense rainfall are more fluid and tend to spread out more across a depositional area, whereas seismically induced landslides tend to have a blockier appearance and a more limited depositional extent in some cases. None of these criteria is definitive, but the types and characteristics of landslides described previously do suggest seismic triggering and can be used as corroborative evidence of earthquake triggering.

Landslide size is considered evidence of seismic triggering in some cases. In areas where large landslides have been documented in historical time to occur only during earthquakes, the large size of prehistoric landslides could suggest seismic origin and could even be used to infer the relative size of the triggering earthquake; very large landslides commonly are triggered by longer duration and longer period shaking, which generally relate to larger magnitude earthquakes.

Multiple lines of evidence strengthen an argument for seismic triggering. For example, a large, ancient landslide near an active fault and having a low-angle basal shear surface might be considered a strong candidate for having been seismically triggered, particularly if similar landslides have been documented in recent earthquakes.

Sackungen

Sackungen are geomorphic features in mountainous areas that are characterized by ridge-parallel, uphill-facing scarps; double ridge lines; and troughs or closed depressions along ridge crests. While topography and gravity clearly influence the ridge-parallel geometry of sackungen, several different processes for their origin have been proposed, including gravitational spreading due to long-term creep, stress relief due to deglaciation, faulting, strong shaking, or a combination of factors. It appears that sackungen can form under a variety of conditions or combinations of conditions because sackungen have been documented to have formed in different ways in different tectonic and geologic settings.

Sackungen have been documented in several historical earthquakes, but the specific mechanism by which they form appears to be complex. Strong shaking certainly plays a role, but in many cases sackungen formed in the immediate vicinity of, and parallel to, the seismogenic faults, which suggests that fault-related tectonic deformation as well as strong shaking might contribute to their formation. And some sackungen appear to have multiple episodes of movement, some related to seismic shaking and some to periods of climatically induced increased groundwater levels.

Because sackungen can form in a variety of tectonic and geologic environments and can form by several different processes, paleoseismic interpretation is difficult and commonly tenuous. Criteria for establishing the seismic or nonseismic origin of sackungen have been proposed with the aim of differentiating between features indicating abrupt, episodic movement versus those indicating gradual, continuous movement. McCalpin (1999) proposed seven criteria, including stratigraphic, geomorphic, and structural evidence, to differentiate between seismic and nonseismic movement: (1) evidence of continuous deformation of sediments suggests a nonseismic origin; (2) sackung deformation events that are contemporaneous with other regional paleoseismic features could be coseismic; (3) if sackungen overlie a steeply dipping crustal fault zone that has a net displacement much larger than the scarp height, the fault could be active; (4) gravity-driven sackungen tend to occur in swarms and be shorter, less continuous, and arcuate, whereas tectonic scarps tend to be longer, more continuous, singular, and straighter; (5) height-to-length ratios of gravity-driven sackungen are much greater than those of tectonic faults; (6) an asymmetrical fault zone having a sharp upper boundary and transitional brecciated lower boundary is more likely to be a sackung than a tectonic fault; and (7) subsurface deformation zones of tectonic faults can occur in any spatial relation with the modern topography, whereas subsurface deformation zones of sackungen are closely related to modern topography.

No single criterion is sufficient to unequivocally prove the seismic or aseismic origin of a sackung feature. And a seismic origin could have resulted from strong shaking, primary tectonic faulting, sympathetic faulting on a feature other than the seismogenic fault, or a combination of these factors. Also, a single sackung feature could have had both seismic and aseismic episodes of movement. Therefore, paleoseismic interpretation of sackungen is generally quite challenging and in many cases impossible. In some cases, however, evidence for abrupt episodes of movement that can be linked to seismic event can provide valuable paleoseismic evidence.

Sediment from Earthquake-Triggered Landslides

Earthquake-triggered landslides can profoundly affect alluvial systems by denuding slopes, which generates large amounts of disrupted sediment that will move into the alluvial system and physically disrupt drainage systems. The commonest types of landslides triggered by earthquakes are shallow, highly disrupted slides in unconsolidated surficial material, and the deposits of these types of landslides tend to move quickly into stream drainages. Thus, earthquakes can deposit large pulses of sediment into alluvial systems, which can (1) create new alluvial fans, (2) cause widespread aggradation of channels, (3) provide material for subsequent deposition on fan surfaces by debris flows and hyperconcentrated flows, and (4) affect the overall development of the fan surface on the long term. Thus, landslides triggered by earthquakes can leave evidence in the depositional record of alluvial systems.

Large earthquakes can cause a spectrum of ground-failure effects including abundant landslides, pervasive ground cracking, microfracturing of surficial hillslope materials, collapse of drainage banks over long stretches, widening of hillside rills, and lengthening of first-order tributary channels. Such widespread disruption increases the capacity of channels to carry runoff by enlarging upstream channels and detaches large amounts of loose slope material, which increases the amount of sediment available for transport and deposition.

Comparison of normal debris-flow deposits to those deposited soon after major earthquakes shows that the post-earthquake deposits tend to (1) be abnormally thicker, (2) contain larger clasts, (3) contain a higher percentage of coarse clasts, and (4) have more angular clasts.

Landslides that Straddle Faults

Landslides sometimes occur on slopes immediately above fault traces, and the slide mass can extend across the trace. Subsequent surface movement of such a fault would offset the landslide mass and allow estimation of fault slip rates if the slide could be dated. This approach does not

require that the landslide be seismically triggered because the paleoseismic interpretation is based on post-landslide fault offset of the landslide mass. However, landslides triggered in the immediate vicinity of active faults commonly are seismically triggered.

Precariously Balanced Rocks

Precariously balanced rocks have been used as crude paleoseismoscopes. The premise of this approach is that areas containing precariously balanced rocks indicate the absence of strong earthquake motions since the precarious rocks developed; paleoseismic interpretations can be made by estimating the peak accelerations required to cause toppling and the length of time the rocks have been precarious. The shaking required to topple precarious rocks has been estimated using analytical and numerical modeling, physical modeling using shaking-table tests, and field experiments on actual precarious rocks. Precarious rocks have been defined as being capable of being toppled by peak accelerations of 0.1–0.3 g; rocks requiring 0.3–0.5 g are commonly defined as semi-precarious. Precarious rocks can be dated cosmogenically and by analysis of rock-varnish microlaminations.

Results from precarious-rock analyses are not always consistent with other lines of paleoseismic evidence. There are several possible reasons for these inconsistencies, and no consensus currently exists regarding the validity of the results of the precarious-rock studies. Therefore, several caveats regarding precarious-rock studies should be kept in mind: (1) Large uncertainties exist in the required toppling accelerations owing both to the geometric complexity of the rocks and the complexities of 3-D ground motion. (2) Not all precarious rocks in a given area will be toppled by the estimated threshold ground shaking. Studies of overturning of tombstones in Japan have shown that a given threshold acceleration will overturn only a fraction of a group of seemingly identical tombstones. Therefore, finding some precarious rocks in an area does not necessarily mean that the area has not experienced the threshold ground shaking. (3) Establishing the age of precarious rocks does

not necessarily determine the minimum time since a certain level of shaking has occurred because the toppling acceleration of a given rock will have been continuously changing as the rock has evolved into a precarious state. For example, a precarious rock with an estimated age of 20 ka and a present-day toppling acceleration of 0.2 g might have had a toppling acceleration of 1.0 g at 20 ka, 0.5 g 10 ka, etc.; therefore, it cannot be concluded that the present-day toppling acceleration has not been exceeded in 20 ka.

Speleoseismology

Speleoseismology is the investigation of earthquake records in caves. Such records can include broken speleothems (stalactites, stalagmites, soda straws, etc.), cave-sediment deformation structures, offset along fractures and bedding planes, simple rock falls (incision), and coseismic fault displacement. Before an earthquake origin can be inferred, all other possible causes of the disturbance must be ruled out. Such causes include human or animal disturbance, water flow, ice movement, debris flow, and sediment creep.

By measuring and dating the tilting and collapse of many stalagmites in a region, it is possible to differentiate sudden (seismic) versus gradual movements and local versus regional causes. Tilting and collapse events can be dated by analysis of radiometrically determined speleothem growth rates; uranium-series isotopes can be analyzed to date speleothems precisely within the 0–500 ka range. By modeling stalagmites as simple inverted pendulums, it is possible to estimate the minimum ground shaking necessary to cause collapse using pseudostatic engineering analysis.

Numerical and physical models have been developed to examine the ground shaking that would be required to break and topple various types of speleothems. These models have been used to measure the natural frequencies and damping characteristics of speleothems and the peak ground accelerations necessary to break them. The natural frequencies of most speleothems are between 50 and 700 Hz, well above the range of seismically generated ground

motion (0.1–30 Hz). The only exceptions are so-called soda straws: long, slender speleothems that can have natural frequencies as low as 20 Hz. Most speleothems would require ground accelerations in excess of 1 g to cause breakage; some very long, thin soda straws, however, could be broken at accelerations as low as 0.1–0.2 g. Thus it appears that only exceptionally long, thin speleothems having weak sections are likely to break during earthquakes, and only about 2 % of such structures have been observed to have broken in recent, well-documented earthquakes.

Summary

Many methods for interpreting the seismic origin of landslides have been developed and, in some cases, successfully applied to paleoseismic analysis. Virtually all of the methods summarized in this section have one aspect in common, which is stated explicitly in most papers: the seismic origin of the features being interpreted remains tentative and cannot be proven, because in each case a nonseismic process could have produced the observed features. Circumstantial evidence for seismic triggering ranges from very strong to extremely tenuous. Indeed, on the latter end of the spectrum, the reasoning can be rather circular: an earthquake origin for a feature is assumed, and then an earthquake origin is interpreted and concluded from analysis of that feature. Any paleoseismic interpretation of a feature is limited primarily by the certainty with which seismic triggering can be established.

Using Stability Analysis to Determine Seismic Landslide Origin

The most direct way to assess the relative likelihood of seismic versus aseismic triggering of an individual landslide is to apply established methods of static and dynamic slope-stability analysis (Jibson and Keefer 1993). Such an analysis involves constructing a detailed slope-stability model of static conditions to determine if failure is likely to have occurred in any reasonable set of groundwater and shear-strength conditions in the absence of earthquake shaking.

All potential nonseismic factors must be considered; these might include processes such as fluvial or coastal erosion that oversteepens the slope or undrained failure resulting from rapid draw-down (for slopes subject to submersion). If aseismic failure can reasonably be excluded even in worst-case conditions (minimum shear strength, maximum piezometric head), then an earthquake origin can be inferred. Dynamic slope-stability analyses can then be used to estimate the minimum shaking conditions that would have been required to cause failure.

This approach is by far the most involved but also yields quantitative results that can be used to assess landslide origins (Jibson and Keefer 1993). Steps involved in a typical stability analysis are summarized in the following sections.

Geotechnical Investigation

Accurately modeling the stability of a slope requires detailed investigation to determine the geotechnical properties of the slope materials. The key properties required for a stability analysis include the material shear strength (friction angle and cohesion), unit weight, and moisture content. Investigating these properties might involve (1) in situ approaches such as cone-penetration testing or (2) acquiring samples that can be tested in the laboratory.

Shear strength can be characterized in different ways to model different types of failure conditions. In aseismic conditions, effective (drained) shear strengths are used because pore-water pressures are assumed to be in static equilibrium. During earthquakes, many soils behave in a so-called undrained manner because excess pore pressures induced by the transient ground deformation cannot dissipate during the brief duration of the shaking; therefore, total (undrained) shear strengths are used to model seismic failure conditions. Effective shear strengths can be measured in the laboratory using various methods: (1) direct shear in which the strain rate is slow enough to allow full drainage and (2) consolidated-undrained triaxial (CUTX) shear in which pore pressure is measured to allow modeling of drained conditions. Total (undrained) shear strength can be measured

using CUTX shear tests or simpler methods such as vane shear.

Static (Aseismic) Slope-Stability Analysis

Static slope-stability analysis models the stability of slopes in the absence of earthquake shaking. A stability model using effective shear strengths can be constructed, and the worst possible groundwater conditions can be modeled to determine the likelihood of aseismic failure. If a slope is stable even in worst-case aseismic conditions, then it is likely that seismic shaking was necessary to induce failure.

Slope stability is quantified using the factor of safety (FS), the ratio of the sum of the resisting forces or moments that act to inhibit slope movement to the sum of the driving forces or moments that tend to cause movement. Slopes having factors of safety greater than 1.0 are thus stable; those having factors of safety less than 1.0 should move. Of course, input parameters have uncertainties, and so determining the stability of slopes from the factor of safety requires judgment. A good rule of thumb for interpreting factors of safety for slopes that have well-constrained input parameters is as follows:

FS < 1 is considered unstable.

FS = 1.00–1.25 is considered marginally stable.

FS = 1.25–1.50 is considered stable.

FS > 1.50 is considered very stable.

Dynamic (Seismic) Slope-Stability Analysis

Analysis of slope stability during earthquake shaking is best modeled using sliding-block analysis. This type of analysis was first introduced by Newmark (1965) and is used widely in engineering practice. Newmark's method models a landslide as a rigid-plastic friction block that slides on an inclined plane. The block begins to slide when a given critical (or yield) base acceleration is exceeded; thus, critical acceleration is defined as the base acceleration required to overcome basal shear resistance and initiate sliding. The analysis calculates the cumulative permanent displacement of the block as it is subjected to the effects of an earthquake acceleration-time history, and the user judges the significance of the

displacement. Laboratory model tests and analysis of actual earthquake-induced landslides have confirmed that Newmark's method can fairly accurately predict landslide displacements if slope geometry and soil properties are known accurately and if earthquake ground accelerations can be estimated using real or artificial acceleration-time histories. More sophisticated forms of sliding-block analysis have been developed that allow modeling the landslide block as a flexible rather than a rigid mass, which yields more accurate results for deeper, larger landslides.

The critical acceleration is a simple function of the static factor of safety and the landslide geometry; it can be expressed as

$$a_c = (FS - 1)g \sin \alpha, \quad (1)$$

where a_c is the critical acceleration in terms of g , the acceleration of Earth's gravity; FS is the static factor of safety; and α is the thrust angle, the angle from the horizontal that the center of mass of the potential landslide block first moves.

Calculation of the estimated landslide displacement consists of a two-part integration with respect to time: (1) the parts of the selected acceleration-time history that lie above the critical acceleration of the landslide block are integrated to yield the velocity of the block with respect to its base and (2) the velocity curve is then integrated to determine the cumulative permanent displacement of the block.

Conducting a rigorous sliding-block analysis requires knowing the critical acceleration of the landslide and selecting one or more earthquake acceleration-time histories ([► Selection of Ground Motions for Response History Analysis](#), [► Time History Seismic Analysis](#)) to approximate the earthquake shaking at the site. The critical acceleration of a potential landslide can be determined in two ways: (1) For relatively simple slope models where material properties do not differ significantly between layers, Eq. 1 can be used to estimate the critical acceleration. (2) For more complex slope models that include layers having complex geometries or widely differing material properties, the critical acceleration should be determined using iterative pseudostatic

analysis, where different seismic coefficients are used until the static factor of safety reaches 1.0. The seismic coefficient yielding a factor of safety of 1.0 is the yield or critical acceleration.

Selecting an appropriate suite of strong-motion records ([► Selection of Ground Motions for Response History Analysis](#), [► Time History Seismic Analysis](#)) for the dynamic analysis can be challenging. Some key properties to consider in estimating ground motions and selecting records include earthquake magnitude, source distance, peak ground acceleration, Arias intensity, and shaking duration.

The significance of the Newmark displacements must be judged in terms of the probable effect on the potential landslide mass. For shallower landslides in brittle surficial rock and soil, estimated displacements in the 5–10-cm range commonly correlate with failure. For deeper landslides in more compliant material, estimated displacements in the 10–30-cm range more commonly correlate with landslide initiation. When displacements in this range occur, previously undisturbed materials can lose some of their strength and be in a residual-strength condition. Static factors of safety using residual shear strengths can then be calculated to determine the stability of the landslide after earthquake shaking (and consequent inertial landslide displacement) ceases.

Interpreting Minimum Ground Motions Required to Cause Slope Failure

If static stability analysis clearly indicates that failure of a landslide in aseismic conditions is highly unlikely, then an earthquake origin can be hypothesized. A dynamic analysis can then be used to estimate the minimum shaking necessary to have caused failure. Such an approach requires a general relationship between critical acceleration, shaking intensity (which can be characterized in various ways), and Newmark displacement. Several such relations have been published.

For example, consider a hypothetical landslide that is stable in aseismic conditions and that has a critical acceleration of 0.15 g . The following equation (Jibson 2007) could be used to estimate the minimum peak ground acceleration

(► **Selection of Ground Motions for Response History Analysis**) needed to cause failure:

$$\log D_N = 0.215 + \log \left[\left(1 - \frac{a_c}{a_{\max}} \right)^{2.341} \left(\frac{a_c}{a_{\max}} \right)^{-1.438} \right] \quad (2)$$

where D_N is Newmark displacement in centimeters, a_c is critical acceleration, and $a_{\max}$ is peak ground acceleration; the ratio of a_c to $a_{\max}$ is commonly referred to as the critical acceleration ratio. Applying this equation requires judgment regarding the critical amount of Newmark displacement that would reduce shear strength on the failure surface to residual levels and lead to continuing failure. The general guidelines stated previously suggest that a reasonable estimate of critical displacement might be 10 cm. For the hypothetical example under consideration, insertion of a displacement value (D_N) of 10 cm into Eq. 2 yields a critical acceleration ratio of 0.2. For a critical acceleration (a_c) of 0.15 g for the hypothetical landslide, this would yield a peak ground acceleration ($a_{\max}$) of 0.75 g as a minimum ground acceleration required to trigger enough displacement to cause general failure.

The peak ground acceleration from such an analysis could be used by itself as a basis for hazard assessment, or it could be used to estimate various magnitude/distance combinations of possible triggering earthquakes. If more than one landslide of identical age were similarly analyzed in an area, iterative magnitude and distance combinations could be optimized to estimate likely earthquake characteristics.

Equations that use other parameters are available and could be applied similarly. For example, a minimum threshold Arias (1970) intensity leading to slope failure can be estimated using the following equation (Jibson 2007) if a reasonable critical displacement can be specified:

$$\log D_N = 2.401 \log I_A - 3.481 \log a_c - 3.230 \quad (3)$$

where D_N is Newmark displacement in centimeters, I_A is Arias intensity in meters per second, and a_c is critical acceleration in terms of g.

Another approach for estimating ground motions from the results of slope-stability analyses uses a quantity referred to as $(A_c)_{10}$, which is the critical acceleration of a landslide that will yield 10 cm of displacement (the estimated critical displacement leading to catastrophic failure) in a given level of earthquake shaking. The following regression model relates Arias intensity to $(A_c)_{10}$:

$$\log (A_c)_{10} = 0.79 \log I_A - 1.095, \quad (4)$$

where $(A_c)_{10}$ is in g's and I_A is in meters per second (Crozier 1992). If the critical acceleration of a landslide can be determined, then this value can be used as the threshold value of $(A_c)_{10}$ in Eq. 4, and the Arias intensity that would trigger the critical displacement of 10 cm can be calculated.

Interpreting Results of Paleoseismic Landslide Studies

Once a landslide or group of landslides has been identified, dated, and linked to earthquake shaking, interpretations regarding the magnitude and location of the triggering earthquake can be made. The previous section outlined a method for detailed geotechnical analysis to address this issue, but in many cases such an analysis will be impossible owing to lack of data or the unsuitability of the landslide for detailed modeling. Several other approaches to this last level of paleoseismic interpretation are possible; in most cases, multiple lines of evidence will be required to make reasonable estimates of magnitude and location. Perhaps the most important aspect of such interpretation is a thorough understanding of the characteristics of landslides triggered by recent, well-documented earthquakes.

Characteristics of Landslides Triggered by Earthquakes

Comprehensive studies of landslides caused by historical earthquakes have allowed documentation of minimum earthquake magnitudes and intensities that have triggered landslides of



Paleoseismology and Landslides, Fig. 2 Dense concentration of disrupted slides and falls triggered by the 1994 Northridge, California, earthquake (M_w 6.7).

Virtually all of the *light-colored* areas in the photo are triggered landslides (Photograph by R.W. Jibson, U.S. Geological Survey)

various types, average and maximum areas affected by landslides as a function of magnitude, and maximum distances of landslides from earthquake sources as a function of magnitude (Keefer 1984, 2002). For these comparisons, landslides were grouped into three categories: disrupted slides and falls, including falls, slides, and avalanches in rock and soil (Fig. 2); coherent slides, including slumps and block slides in rock and soil and slow earth flows (Fig. 3); and lateral spreads and flows, including lateral spreads and rapid flows in soil and subaqueous landslides (Fig. 4).

Minimum Earthquake Magnitudes that Trigger Landslides

Table 2 shows the minimum magnitudes of earthquakes that have triggered various types of landslides. Landslides of various types have threshold magnitudes ranging from 4.0 to 6.5; the more disrupted types of landslides have lower threshold magnitudes than the more coherent types of slides. Although smaller earthquakes could conceivably trigger landslides, such triggering by

very weak shaking probably would occur on slopes where failure was imminent before the earthquake.

Minimum Shaking Intensities that Trigger Landslides

Table 3 shows the lowest Modified Mercalli Intensity (MMI) values and the predominant minimum MMI values reported where the three categories of landslides occurred. The data show that landslides of various types are triggered one to five levels lower than indicated in the current language of the MMI scale.

Areas Affected by Earthquake-Triggered Landslides

Drawing boundaries around all reported landslide locations in historical earthquakes and calculating the areas enclosed yields a plot of area versus earthquake magnitude (Fig. 5); a well-defined upper bound curve represents the maximum area that can be affected for a given magnitude (Keefer 1984). Average area affected by landslides as a function of earthquake magnitude



Paleoseismology and Landslides, Fig. 3 Rotational slump that moved as a coherent landslide in the 2004 Niigata Ken Chuetsu, Japan, earthquake (M_w 6.8) (Photograph by D.S. Kieffer, Graz University of Technology, Austria)



Paleoseismology and Landslides, Fig. 4 Lateral-spread landslide triggered by the 1980 Mammoth Lakes, California, earthquake (M_w 6.2) (Photograph by E.L. Harp, U.S. Geological Survey)

can be predicted using the following regression equation (Keefer and Wilson 1989):

$$\log A = M - 3.46 \pm 0.47 \tag{5}$$

where *A* is area affected by landslides in square kilometers and *M* is a composite magnitude term, which generally indicates surface-wave

Paleoseismology and Landslides, Table 2 Minimum earthquake magnitude required to trigger landslides

Earthquake magnitude	Type of landslide
4.0	Rock falls, rock slides, soil falls, disrupted soil slides
4.5	Soil slumps, soil block slides
5.0	Rock slumps, rock block slides, slow earth flows, soil lateral spreads, rapid soil flows, subaqueous landslides
6.0	Rock avalanches
6.5	Soil avalanches

Note: Data from Keefer (1984)

Paleoseismology and Landslides, Table 3 Minimum modified Mercalli intensity required to trigger landslides

Landslide type	Lowest modified Mercalli intensity	Predominant modified Mercalli intensity
Disrupted slides and falls	IV	VI
Coherent slides	V	VII
Lateral spreads and flows	V	VII

Note: Data from Keefer (1984)

Paleoseismology and Landslides, Fig. 5 Area affected by seismically triggered landslides plotted as a function of earthquake magnitude. *Solid line* is upper bound of Keefer (1984); *dashed line* is from Rodriguez et al. (1999); *dotted line* is regression line from Keefer and Wilson (1989)

magnitudes below 7.5 and moment magnitudes above 7.5. Area affected by landslides also is influenced by the geologic conditions that control the distribution of susceptible slopes and by the focal depth of the earthquake.

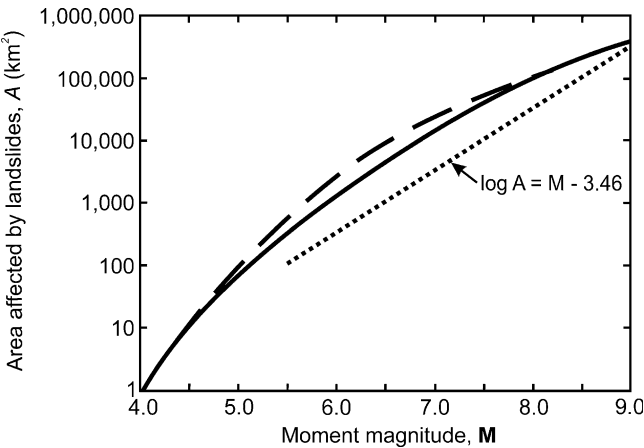
Maximum Distance of Landslides from Earthquake Sources

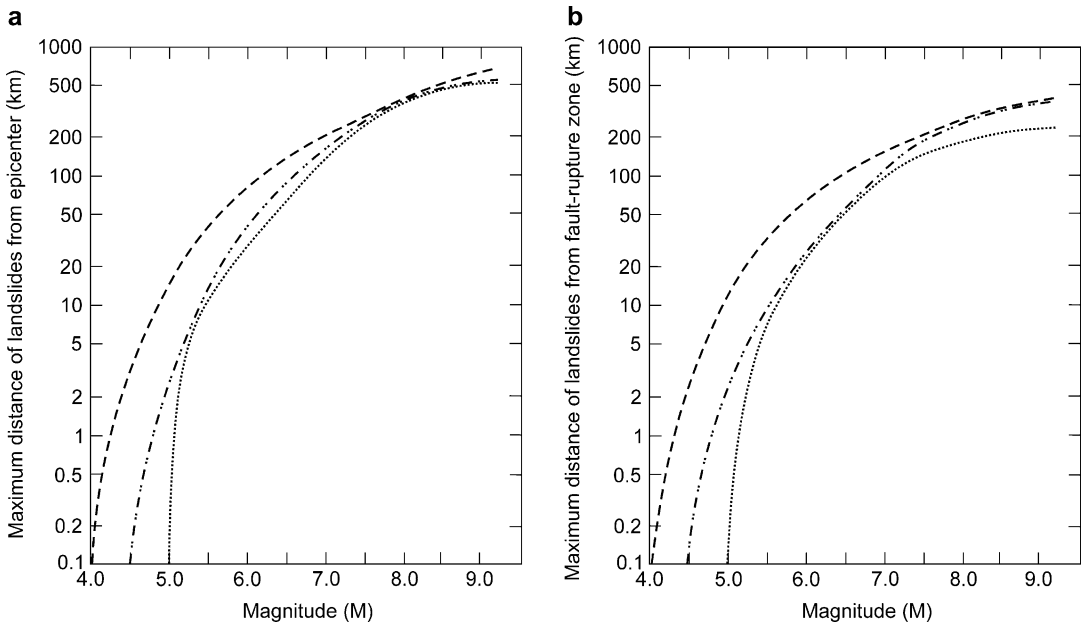
The maximum distance of the three categories of landslides from the earthquake epicenter and from the closest point on the fault-rupture surface relates closely to earthquake magnitude (Fig. 6). Upper bound curves are well defined and are constrained to pass through the minimum threshold magnitudes shown in Table 2 as distance approaches zero. Although the upper bounds shown have been exceeded a few times in subsequent earthquakes, they remain fairly reliable indicators of the maximum possible distances at which the three classes of landslides could be triggered in earthquakes of various magnitudes.

Figure 6 indicates that disrupted slides and falls have the lowest shaking threshold and that lateral spreads and flows have the highest shaking threshold. As with area, earthquakes having focal depths greater than 30 km generally triggered landslides at greater distances than shallower earthquakes of similar magnitude.

Interpreting Earthquake Magnitude and Location

Figures 5 and 6 and Tables 2 and 3 allow interpretation of earthquake magnitude and location in





Paleoseismology and Landslides, Fig. 6 Maximum distance to landslides from (a) epicenter and (b) fault-rupture zone for earthquakes of different magnitudes. *Dashed line* is upper bound for disrupted slides and falls;

dash-double-dot line is upper bound for coherent slides; and *dotted line* is upper bound for lateral spreads and flows (Modified from Keefer 1984)

a variety of ways. If a single landslide is identified as being seismically triggered, then a minimum magnitude and MMI can be estimated based on the landslide type. If several landslides in an area are identified as being seismically induced, then application of the magnitude-area and magnitude-distance relationships can yield minimum magnitude estimates. As the area in which landslides documented to have been triggered by the same earthquake increases, the estimated magnitude will increase toward the actual magnitude of the triggering earthquake. Therefore, documentation and analysis of landslides over a large area will produce more accurate magnitude estimates. If seismic source zones are well documented, then the distance from the closest source zone to the farthest landslide will yield a reasonable minimum magnitude estimate. The observation that greater source depth relates to greater areas affected and greater source distances for landslides of all types further complicates estimation of earthquake magnitude.

For a specific region, earthquake magnitude can be estimated based on comparison of

paleoseismic landslide distribution with landslide distributions from recent, well-documented earthquakes in the region.

Earthquake locations generally are estimated based on the distribution of synchronous landslides attributed to a single seismic event. In a broad area of roughly similar susceptibility to landsliding, the earthquake epicenter probably will coincide fairly closely with the centroid of the landslide distribution. In areas of highly variable or asymmetrical landslide susceptibility, epicentral estimation is much more difficult and subject to error. In areas where seismic source zones are well defined, the epicentral location is best defined as the point in a known seismic source zone (or along a known seismogenic fault) closest to the centroid of the landslide distribution.

Additional Considerations

The primary limitation of paleoseismic analysis of landslides is the inherent uncertainty in

interpreting a seismic origin. Unlike liquefaction, which can occur aseismically only in relatively rare conditions, landslides of all types form readily in the absence of earthquake shaking as a result of many different triggering mechanisms. In many cases, ruling out aseismic triggering will be impossible, and the level of confidence in any resulting paleoseismic interpretation will be limited. For this reason, paleoseismic landslide analysis should include, so far as possible, multiple lines of evidence to constrain a seismic origin. In this way, a strong case can be built for seismic triggering of one or more landslides, even if no single line of evidence is unequivocal. Where independent paleoseismic evidence from fault or liquefaction studies is available, paleoseismic landslide evidence can provide useful corroboration.

Detailed slope-stability analyses generally can be performed only on certain types of landslides. Failure conditions of falls, avalanches, and disrupted slides cannot easily be modeled using Newmark's method, and even static stability analyses of these types of slides can be very problematic. Also, the pre-landslide geometry of slides in very steep terrain can be difficult or impossible to reconstruct. Thus, detailed dynamic slope-stability analysis can be applied only to fairly coherent landslides where pre-landslide geometry can be reconstructed with confidence, where groundwater conditions can be modeled reasonably, and where the geotechnical properties of the materials can be accurately measured.

Even allowing for these limitations, paleoseismic landslide studies have been extremely useful where applied successfully, and they hold great potential in the field of paleoseismology. Dating landslide deposits is, in many cases, easier than dating movement along faults because many different dating methods can be used on the same slide to produce redundant results. In addition, landslides have the potential for preserving large amounts of datable material in the various parts of the slide (scarp, body, toe, etc.). In areas containing multiple or poorly defined seismic sources, paleoseismic ground-failure analysis might be preferable to

fault studies because landslides preserve a record of the shaking history of a site or region from all seismic sources. Knowing the frequency of strong shaking events could, in many cases, be more critical than knowing the behavior of any individual fault.

Paleoseismic landslide analysis could have greatest utility in assessing earthquake hazards in stable continental interiors where fault exposures are rare or absent but where earthquakes are known to have occurred. In such areas, analysis of earthquake-triggered ground failure, both landslides and liquefaction, might be one of the few paleoseismic tools available.

Another advantage of paleoseismic landslide analysis is that it gets directly at the effects of the earthquakes being studied. Ultimately, most paleoseismic studies are aimed at assessing earthquake hazards. Fault studies can be used to estimate slip rates, recurrence intervals, and, indirectly, magnitudes. From these findings, the effects of a possible earthquake on such a fault are extrapolated. In paleoseismic landslide studies, the effects are observed directly. Thus, if a seismic origin can be established, a landslide shows directly the effects of some previous earthquake. Even if magnitude and location are poorly constrained, at least a partial picture of the actual effects of seismic shaking in a locale or region can be estimated.

In conclusion, paleoseismic landslide analysis can be applied in a variety of ways and can yield many different types of results. Although interpretations are limited by the certainty with which a seismic origin can be established, paleoseismic landslide studies can play a vital role in the paleoseismic interpretation of many areas, particularly those lacking fault exposures.

Summary

In many environments, landslides preserved in the geologic record can be analyzed to determine the likelihood of seismic triggering. If evidence indicates that a seismic origin is likely for a landslide or group of landslides, and if the landslides can be dated, then a paleoearthquake

can be inferred, and some of its characteristics can be estimated. Such paleoseismic landslide studies thus can help reconstruct the seismic shaking history of a site or region. In regions that contain multiple seismic sources and in regions where surface faulting is absent, paleoseismic ground-failure studies are valuable tools in hazard and risk studies that are more concerned with shaking hazards than with interpretation of the movement histories of individual faults. Paleoseismic landslide analysis involves three steps: (1) identifying a feature as a landslide, (2) dating the landslide, and (3) showing that the landslide was triggered by earthquake shaking. Showing that a landslide was triggered by seismic shaking can be challenging, but some types of landslides can be analyzed using established static and dynamic methods of slope-stability analysis to determine the likelihood of seismic triggering and to estimate minimum shaking levels required to initiate landslide movement.

Cross-References

- ▶ [Earthquake Magnitude Estimation](#)
- ▶ [Luminescence Dating in Paleoseismology](#)
- ▶ [Radiocarbon Dating in Paleoseismology](#)
- ▶ [Seismic Actions due to Near-Fault Ground Motion](#)
- ▶ [Selection of Ground Motions for Response History Analysis](#)
- ▶ [Time History Seismic Analysis](#)

References

- Arias A (1970) A measure of earthquake intensity. In: Hansen RJ (ed) *Seismic design for nuclear power plants*. MIT Press, Cambridge, MA, pp 438–483
- Crozier MJ (1992) Determination of paleoseismicity from landslides. In: Bell DH (ed) *Landslides (Glissements de terrain)*, Proceedings of the 6th international symposium, vol 2. A. A. Balkema, Christchurch/Rotterdam, pp 1173–1180
- Jibson RW (2007) Regression models for estimating coseismic landslide displacement. *Eng Geol* 91:209–281

- Jibson RW (2009) Using landslides for paleoseismic analysis. In: McCalpin JP (ed) *Paleoseismology*, 2nd edn. Academic, New York, pp 565–601
- Jibson RW, Keefer DK (1993) Analysis of the seismic origin of landslides – examples from the New Madrid seismic zone. *Geol Soc Am Bull* 105:521–536
- Keefer DK (1984) Landslides caused by earthquakes. *Geol Soc Am Bull* 95:406–421
- Keefer DK (2002) Investigating landslides caused by earthquakes – a historical review. *Surv Geophys* 23:473–510
- Keefer DK, Wilson RC (1989) Predicting earthquake-induced landslides, with emphasis on arid and semi-arid environments. In: Sadler PM, Morton DM (eds) *Landslides in a semi-arid environment*, vol 2. Inland Geological Society, Riverside, pp 118–149
- McCalpin JP (1999) Criteria for determining the seismic significance of sackungen and other scarplike landforms in mountainous regions. In: *Techniques for identifying faults and determining their origins*. U.S. Nuclear Regulatory Commission, NUREG/CR-5503, Washington, DC pp A-122–A-142
- Newmark NM (1965) Effects of earthquakes on dams and embankments. *Geotechnique* 15:139–160
- Rodríguez CE, Bommer JJ, Chandler RJ (1999) Earthquake-induced landslides: 1980–1997. *Soil Dyn Earthq Eng* 18:325–346
- Varnes DJ (1978) Slope movement types and processes. In: Schuster RL, Krizek RJ (eds) *Landslides – analysis and control*. National Academy of Sciences, Washington, DC, pp 11–33, Transportation Research Board Special report 176

Paleoseismology of Rocky Coasts

Niki Evelpidou¹ and Paolo Pirazzoli²

¹Faculty of Geology and Geoenvironment/
National and Kapodistrian University of Athens,
Athens, Greece

²Laboratoire de Géographie Physique, Paris,
France

Synonyms

Littoral; Sea level; Shoreline; Subsidence; Uplift

Introduction

Important earthquakes are often accompanied by vertical-land displacements. Therefore in coastal

areas they may result in rapid changes of the relative sea level. An essential tool for the study of coastal paleoseismicity is the identification of fossil paleoshorelines, paying special attention to sea-level indicators that are consistent or provide evidence of rapid relative sea-level change.

Here, different types of sea-level indicators that are often used in literature in order to determine changes in fossil shorelines are summarized. Information is also provided regarding four case studies of important earthquakes that occurred in Greece (in AD 365, 1953, and 1956) and in Japan (in 1923).

How Can Fossil Paleoshorelines Be Identified?

Fossil paleoshorelines can be identified and traced from geomorphological, biological, sedimentological, stratigraphical, or archeological sea-level indicators.

The coastal geomorphological features that are used as sea-level indicators are the result of either erosional or depositional processes. Erosional features can only be preserved on hard, solid rocks, and in some cases they constitute indicators of sea-level change. Such indicators are marine notches, potholes, abrasion platforms, etc. Among the depositional formations, marine terraces and beachrocks stand out as the most important sea-level indicators.

Erosional Geomorphological Sea-Level Indicators

Tidal notches are well known as precise sea-level indicators that usually undercut limestone cliffs in the midlittoral zone (e.g., Pirazzoli 1986), which renders them the most important erosional geomorphological sea-level indicators. In microtidal areas sheltered from wave action, elevated or submerged notches are used to indicate former sea-level positions, with up to a decimeter confidence.

Bioerosion by endolithic organisms and surface feeders grazing upon epi- and endolithic algae are generally acknowledged to play an important role in tidal-notch development.

The erosion rate is generally highest near mean sea level (MSL) and decreases gradually toward the upper and lower limits of the intertidal range. Accordingly, in places sheltered from continuous wave action, if MSL remains stable, notch profiles will be typically reclined U shaped or V shaped, with a vertex located near MSL, the base of the notch near the lowest-tide, and the top near the highest-tide level. In a moderately exposed site, continuous wave action may splash seawater onto the roof, thus shifting the top of the notch upward, above the highest-tide level.

The rate of maximum undercutting (near MSL) varies with the rock type and the local climate and has been roughly estimated to be of the order of 1 mm/year (Laborel et al. 1999). However this is only a first-order value, while lower rates are generally observed in hard limestones, especially in nontropical areas. More detailed estimations show a range varying from 0.2 to 5 mm/year, depending on lithology, location, and probably duration of bioerosion (for references, see Pirazzoli 1986, Table 1 and Laborel et al. 1999, Table 1).

When a tidal notch is uplifted or submerged, its profile may provide valuable information concerning the type of the paleoseismic event, e.g., whether one or more events took place, whether the movement was coseismic or gradual, etc.

Depositional Geomorphological Sea-Level Indicators

Marine Terraces

Marine terraces are quite common coastal features, owing their presence to a combination of eustatic and crustal processes (Pirazzoli 1994).

The presence of a sequence of uplifted, stepped marine terraces usually corresponds to the superimposition of eustatic change in sea level and of a tectonic uplifting trend that may include sequences of paleoseismic events. The presence of marine terraces several hundred meters above sea level indicates tectonic uplift and allows the estimation of the long-term rate of tectonic deformation.

Marine terraces' age usually increases with altitude, while conservation quality decreases. When uplifted marine terraces are used for the study of sea-level changes, the uplift rate of each section is considered to have remained constant and the eustatic sea-level position corresponding to at least one uplifted terrace has to be known. Marine terraces usually provide evidence of tectonic trends rather than of single paleoseismic events.

Terms often used by geomorphologists for marine terraces are the "beach angle" or "shoreline angle," expected to indicate a frequent sea-level (and wave) position during the period of terrace formation. However, in some cases it can be difficult to precisely determine shoreline angles for Pleistocene marine terraces, during which relative sea-level changes are not known in great detail. This approach can be more useful for Holocene marine terraces and beach ridges formed between two successive coseismic uplifts, like near the Wairarapa Fault in New Zealand where the mean earthquake recurrence time is expected to be of $1,230 \pm 190$ years (Little et al. 2009).

Beachrocks

Beachrock formation is a diachronic and wide-ranging sedimentary process, and since lithification takes place at the coastline, beachrocks have been used as indicators in Quaternary sea-level and neotectonic studies (e.g., Kelletat 2006).

There are, however, several problems in the use of beachrocks as sea-level indicators; the most important are mentioned next. It is difficult to determine the upper level of beachrock cementation, which may correspond to the mean high water springs, and in regions with high tidal range, this limit may be quite large. Another insurmountable problem arises from fluctuations in the tidal range, originating from local changes in the coastal configuration.

Within a sedimentary body, the cementation zone rises and falls following the perpetual tidal cycle of sea-level rise and fall; thus the lower level of cementation zone will correspond to the lowest sea level, at any time during this sedimentary body's existence. Therefore, the cementation

of the lower level of an exposed formation may have occurred at an earlier stage, during a rising sea level, more recently during a falling sea level, or during any variation. Consequently, only the upper limit of beachrocks constitutes an accurate indicator of past tidal level.

Additionally, there is the possibility of confusion with other cemented materials in the intertidal zone, although a careful study of the cement, the composition, and the microstructure of the sample can accurately determine the type of formation.

Submerged beachrocks or raised deposits may give evidence of tectonic activity in the area, but it is not possible to provide information on whether this displacement is based on one or more paleoseismic events and whether this movement was rapid or slow.

Biological Sea-Level Indicators

The biological zonation depends essentially on wave exposure. One may distinguish a supralittoral zone, a midlittoral zone, and an infralittoral zone.

The biomass of the supralittoral zone is very low (presence of lichens such as *Verrucaria* or *Lichina* and/or of associated endolithic *Cyanobacteria*) and bioconstruction is uncommon but may occur in the form of incrustations by *Chthamalus* shells in shaded rock crevices exposed to wave splash. Limited erosion may take place, mainly due to boring activity in carbonate substrate. According to Le Campion-Alsumard (1979), endolithic microorganisms may create many microtunnels, with their roots, that are easily preserved and fossilized; these marks may be suitable as paleobathymetric indicators. Other algae, which may occur in the lower supralittoral zone, are the green *Endoderma* and the brown *Chrysophyceae*, which may be outgrown by various other algae. The most widespread animals in the lower supralittoral zone are the gastropods *Littorina* and *Melanerita*, which graze the algae, some crustaceans, and occasionally opportunistic visitors that do not normally reside there, such as hermit crabs, insects, spiders, birds, and small rodents (Pirazzoli 1996).

The midlittoral zone is submerged at close intervals by waves and tide. The upper part of this zone is never rich in species and even barren for most of the year on the Arctic and Antarctic shores, due to the grinding action of ice and may become vegetated by benthic diatoms and ephemeral algae only during the summer. *Fucus* and *Pelvetia* algae are frequent in North Atlantic and North Pacific coasts. Farther south a greater amount of annual species is found, and the importance of *Cyanobacteria* increases at the tropics. Lithothamnium species have been observed on exposed shores of the Indian and Pacific Ocean, as well as *Cladophora*, *Ectocarpus*, and *Ulva*. The lower part of the midlittoral zone is in general densely covered by furoid and turf algae, which are grazed by several herbivores (littorinoids, limpets).

Barnacles (*Balanus*, *Elminius*, *Tetraclita*, and especially *Chthamalus*) live in the upper part of the midlittoral zone, while mussels (*Mytilus*) and oysters tend to occupy lower levels. Barnacles, mussels, oysters, and *Lithophyllum* form a rim just above sea level in the western Mediterranean and may be fossilized in situ after death.

Erosive agents include many *Cyanobacteria* in the upper part of this zone and limpets (*Patella*) and Chitons in the lower part.

The infralittoral (or sublittoral) zone extends from the biological mean sea level (BMSL) (Laborel and Laborel-Deguen 1994), located at the base of the midlittoral zone, to a depth of 25–50 m (depending on water transparency) but may be absent in very turbid coastal waters. The position of the BMSL in relation to the tide-gauge MSL varies with exposure.

This zone is densely populated by brown algae, coralline encrusting algae (*Porolithon*, *Neogoniolithon*, *Lithophyllum*), fixed vermetid gastropods (e.g., *Dendropoma*, *Vermetus*, *Serpulorbis*), cirripedes like *Balanus*, or coral reefs in warm waters. They may be accompanied by turf-forming algae and furoid vegetation (*Fucus*, *Cystoseira*, *Sargassum*) that are grazed by herbivorous fish and sea urchins; the latter like clionid sponges and other borers (e.g., *Lithophaga*) may also attack the rocky substrate.

All bioconstructions (encrusting algae, vermetids, reef-building corals and associates, barnacles, and oysters) may fossilize in situ in the sublittoral zone. The most accurate altimetric indications for the estimation of past sea levels derive from the comparison of the fossil bioconstructions with their present-day counterpart. The most useful sea-level indicators are those with the narrowest vertical zonation, e.g., the upper level of coral microatolls and rims made by *Dendropoma* or *Neogoniolithon*. Fossil *Lithophaga*, which are frequently used to indicate uplift on carbonate shores, deserve special mention. They are generally a poor sea-level indicator, because they live between the upper limit of the sublittoral zone and depths greater than 30 m. However, when the upper limit of their population is well marked and forms a horizontal line, it corresponds to the BMSL. Nonetheless, as reported by Shaw et al. (2010), *Lithophaga* shells tend to incorporate host-rock carbon, thus giving older apparent radiocarbon ages. More details on the vertical range, altitudinal accuracy, and resistance to erosion of various Mediterranean midlittoral and infralittoral sea-level indicators are given by Laborel and Laborel-Deguen (1994).

Although biological indicators are very useful for dating an uplift, they are unlikely to be preserved when they are submerged because of the subsequent bioerosion. An additional indicator is useful for diagnosing the type of the paleoseismic event, e.g., coseismic or gradual. In any case, if the shoreline has quickly shifted from the tidal zone, the biota will be well preserved, while slow emergence will expose it leading to gradual destruction in the supralittoral wave zone.

Sedimentological/Stratigraphical Sea-Level Indicators

The most common method for locating material that will provide information regarding sea-level change is the use of a vibrating sampler, provided that samples are slightly moderately disturbed. The method of excavation trenches is also often used for smaller depths.

In order to acquire accurate results from coring, it is necessary to obtain a solid and undisturbed sample, so that the conditions are similar to the ones prevailing at the sampling location. There are several biomarkers, present in sampling cores, which can be dated and associated to paleo-sea levels. An indispensable requirement is the recording of the precise location and height of the sampling with differential GPS.

A stratigraphic column may provide evidence of the paleogeographic/paleoenvironmental changes that originated from paleoseismic events, while it may provide clear evidence of paleo-tsunami in the area. In a stratigraphic column history is fossilized making clear whether transition was slow or abrupt. If datable material is available, the event could be dated as well.

Archeological Sea-Level Indicators

Most archeological remains provide no evidence on how far from sea level they were constructed. A human civilization may not be linked to the sea, and in most cases it is impossible to distinguish a house or pottery found near the sea from an equivalent found some kilometers inland. On the other hand, the civilization may have developed specific activities, closely related to the sea, requiring sailors, fishermen, boat builders, and salt gatherers. In this case the settlement was probably built close to the shore.

Apart from the undisputable evidence of minimum relative sea-level rise, provided by submerged structures that had to have their foundations on dry land (houses, tombs, mosaic floors, passageways, storage tanks, tells, middens), many archeological findings refer to structures that were partly in the sea and may be considered as reliable sea-level indicators (slipways, breakwaters, jetties, quays, docks, channels, drains, salt pans, the lowest part of certain flights of steps or of coastal quarries that used a wood splitting wedge for cutting the lowest level of stones, or finally installations for fishing or fish farming). For a reliable estimation of the ancient sea level, a good understanding of the functionality of the structure and the local

hydrographic and climatic constraints is always essential.

In any case the ruins may provide valuable information concerning the paleoseismic events in an area. For example, around 3300 BP an earthquake destroyed House B in Grotta (Naxos Island, Cyclades). The archeologists who excavated Grotta and Aplomata on Naxos spoke of two seismic events: one at an early phase of the LH IIIA2 and another one in LH IIIC period, based on the ruins of the archeological sites. The destruction of the site known as "Kolona," literally "Column," at Heraion (Samos Island, Eastern Aegean), has been assigned to seismic damage. The drums of the only surviving column are offset laterally and tilted. This effect was clearly the result of rocking produced by a strong earthquake. The coeval destruction of at least one other temple nearby supports the possibility of a destructive earthquake circa 530 BC.

Structures Partly in the Sea

Most known examples (e.g., Flemming 1979) come from the tideless Mediterranean, where much well-preserved material exists. However, similar methods and techniques have also been occasionally adopted in fully tidal waters.

Gradients for ancient slipways may vary from 4° to 15°, and the original position of the sea level can be calculated approximately by making assumptions about the water depth needed at the foot of the structure to take the bow of the boat and the length of slipway needed to be dry to support the boat while work was done on it. Well-preserved slipways in the Mediterranean can provide an accuracy of sea-level estimate of ± 0.25 m.

Breakwaters, jetties, and moles are generally walls, built to create a barrier between the open sea and a calm sheltered area of water, where ships may safely moor. Mooring quays are built on the inner side. The base of the breakwater was in the water, while the top must have been out of the water, high enough to prevent waves from climbing over. Protection moles have often been constructed around Roman fish tanks. Since the upper part of the majority of these structures has

been damaged by countless storms, the identification of the upper level is generally problematic. However, when the upper surface is intact, the ancient sea level may be estimated within ± 1 m.

Quays are well-squared structures built to provide mooring for ships, access for men, and storage space for cargoes. They were often associated with steps leading down to the vessels, mooring stones, bollards, and warehouses. Quay surfaces varied generally around 1.0–2.0 m. Through the combination of the water depth at the foot of the quay wall, the level of the mooring stones, surfaces reached by steps, and the main working surfaces of the quay, it may be possible, according to Flemming (1979), to determine the original sea level to about ± 0.5 m.

Docks are small rectangular basins in which a ship can be berthed and the basin can be pumped dry. The interpretation of docks is similar to that for quays.

Fish Tanks

Though stone fish traps and turtle pens have been reported in many cultures (e.g., Caribbean Indian and Australia aborigine), Roman fish tanks (*piscinae*) are worth special mention. Such installations were especially frequent along the Tyrrhenian coasts of central Italy between the first century BC and the first century AD, when they came into fashion for wealthy Romans. When certain elements are well preserved and the functioning of the structure is well understood, fish tank remains allow estimating with excellent accuracy (± 0.1 m) the sea-level position of about 2,000 year ago. The main sea-level indicators are the top of closing gates (*cataractae*), generally slightly above the high-tide level. In several examples they were located near the top of the walls (*crepidines*) delimiting the basins from where they could be easily worked, by using the *crepidines* as a footwalk. For several other coastal structures, like salt pans, foundations of towers and lighthouses, roads, seaside villas, wells, etc., the exact relation to the sea level may vary between sites.

How Can Fossil Paleoshorelines Be Dated? Radiocarbon, Archeological Remains, Coastal Cores, and Historical Information

Fossil paleoshorelines can be dated through sea-level indicators, using different methods, depending on the type of indicators available. For example, radiocarbon dating of biological indicators is one of the easiest and most trustworthy methods. On the other hand, this is not possible with submerged paleoshorelines since any biological indicator is destroyed by bioerosion occurring after submergence. In this case other, indirect ways of dating fossil paleoshorelines are used. In some cases it is possible to date a relative sea-level change, based on an ancient site in the nearby area.

Emerged or submerged beachrocks may be dated, but their accurate dating is a very difficult task, even though biogenic materials such as shells or corals can easily be dated by radiocarbon. The age acquired corresponds to the time of the organism's death. After death, the organism was transported and deposited on a beach and later was cemented into the beachrock. Although in many cases this sequence may be complete within a few years, it is likely that the time interval between the death of the organism and the cementation of the beachrock may last hundreds or thousands of years. For this reason, the age acquired by dating a constituent organism may be regarded as a maximum for the cementation of beachrock. The minimum age for this process is only attainable by dating the cement. However, several difficulties may arise with extracting the adequate amount of cement, especially in submerged beachrocks. Furthermore, it is possible that after exposure, water passing through beachrock may cause carbonate exchange resulting in continuous renewal of the apparent cementing age.

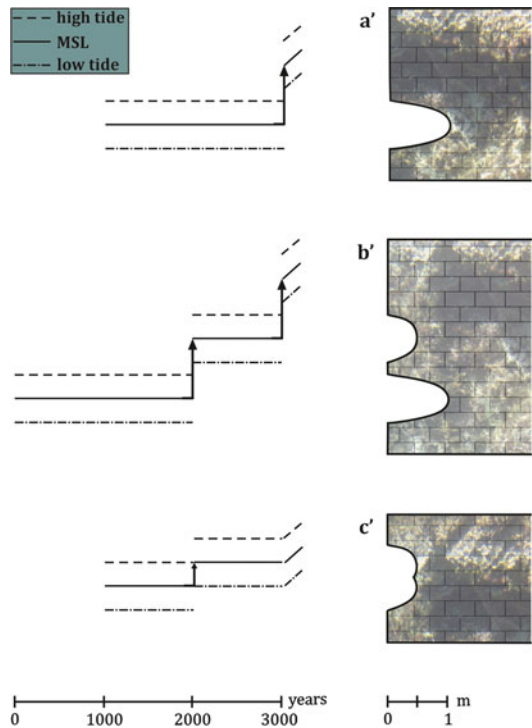
Dating material provided by coastal cores could often be the only solution, especially in cases where biological, geomorphological, and archeological sea-level indicators are absent from a study area, or their dating is impossible. In relatively recent sediments, which were

deposited in various geoenvironments, a variety of information that determines the paleogeography of a region and the various changes in sea level is recorded, through datable material found at a certain depth of the stratigraphic column. In these cases, depth precision is determined by the accuracy of the relationship of the biological marker to mean sea level. However, an uncertainty is added in all estimations, especially if the cores come from marsh deposits where compaction can be significant and the positional uncertainty of the biological markers increases considerably.

How Can Coseismic Sea-Level Changes Be Distinguished from Gradual Changes?

The best indicators to distinguish rapid relative sea-level changes from gradual ones are tidal notches. The shape of the notch profiles may provide qualitative information on the rate of sea-level change and on tectonic movements. In the case of a rapid (e.g., coseismic) emergence greater than the tidal range, the notch that is completely emerged will be preserved from further marine bioerosion, while a new notch will develop in the new lower intertidal zone (Pirazzoli 1986). On the other hand, in the case of rapid submergence greater than the tidal range, the whole notch will be submerged. It will not be completely preserved, due to further bioerosion as in the case of emergence, and its profile will be uniformly deepened by the rate of bioerosion that predominates in the infralittoral zone, while a new tidal notch will develop in the intertidal range. However, the infralittoral bioerosion rate is generally estimated to be one order of magnitude less than the intertidal one.

If the rapid submergence is smaller than the tidal range, the height of the notch will increase, with further deepening in the new intertidal zone, while the part of the notch below the low tide level will continue to deepen at a much slower rate, and the notch profile will be marked by an undulation at the level of the roof of the former notch. Finally, in the case of gradual



Paleoseismology of Rocky Coasts, Fig. 1 Different profiles of notches that give evidence of a rapid movement caused by earthquakes with magnitude larger than 6.0 Mw, commonly associated to morphogenic faults producing direct surface faulting (Ambraseys and Jackson 1990) and ground deformation (Pavlidis and Caputo 2004)

submergence, the height of the notch will gradually increase, the lower part of its profile becoming higher than the upper part, and no marked undulation will appear in the profile. Many combinations of vertical movements leading to emergence or submergence are possible, as well as different erosion rates, tidal ranges, and periods of time, but even so, notch profiles would be similar to those of Fig. 1.

Case Studies

AD 365, Crete Earthquake

On 21 July AD 365, a strong underwater earthquake occurred near the southwestern part of Crete. Nearly all towns in the island were



Paleoseismology of Rocky Coasts, Fig. 2 Remains of at least three fossil ripple notches have been preserved between +4.00 and +4.75 m below Afrata (1.5 km N of Gonias, near the eastern foot of Rodopos peninsula).

These tidal notches have probably developed a few centuries before the general uplift of AD 365 (Photo P.A.P. n° 3904, Sept. 1977)

destroyed, and widespread damage extended to central and southern Greece, northern Libya, Egypt, Cyprus, and Sicily. The earthquake was followed by a tsunami that devastated the southern and eastern coasts of the Mediterranean, particularly Libya, Alexandria, and the Nile Delta, hurling ships nearly two miles inland. The Roman historian Ammianus Marcellinus described in detail the tsunami hitting Alexandria.

In Crete, systematic geomorphological surveys carried out in the late 1970s (Pirazzoli et al. 1982) showed that a block of lithosphere approximately 200 km long, extending from central Crete to the small Antikythera island, was uplifted and inclined northeastward. The uplift reached a maximum of 9 m in southwestern Crete, leaving along the coast a very visible emerged shoreline that has been dated by radiocarbon in many positions and also remains of Roman harbors, like in Kisamos and Falasarna (where the coseismic uplift reached about 6 m).

If the AD 365 tsunami is well dated historically, direct historical evidence confirming that such uplift in Crete was synchronous to the earthquake that produced the tsunami is generally missing, and many radiocarbon dates converge

only by a few decades toward the critical period. A closer connection to a major destructive earthquake that occurred shortly after AD 355/361 in Crete was established archeologically by a number of coins found in Kisamos among skeletons of people killed and buried by fallen debris (Stiros and Papageorgiou 2001).

If the type and geometry of the fault, which produced the block movement, are still a matter of debate, some geomorphological elements provide information on the preparatory phase that permitted the necessary strain accumulation of the major earthquake. These elements consist of nine fossil shorelines, preserved as ripple notches in Antikythera and as bioconstructed vermetid rims at Moni Khrisoskalitisas, remnants of which have been reported in several sites (e.g., Plaka, Falasarna, Koutsounari, Afrata, Damnoni, Sougia) (Fig. 2) showing that the western part of the block uplifted in AD 365 underwent between 4,000 and 1,700 year BP, along a length of at least 150 km, ten small coseismic subsidences (from 10 to 25 cm each time), apparently due to gravitational forces, without any noticeable tilting (Pirazzoli et al. 1981, 1982). Shaw et al. (2010) attempted to narrow the chronology of the subsidence movement having preceded the AD

Paleoseismology of Rocky Coasts,

Fig. 3 Submerged fossil shoreline in Keros island, which has been also located along the rocky coasts of Sifnos, Antiparos, Paros, Naxos, and Iraklia, reveals a widespread evidence of a recent subsidence, part of which according to Evelpidou et al. (2012b) is due to Amorgos 1956 earthquake



365 uplift event by using *Lithophaga* shells and showing that these shells may incorporate host-rock carbon. If the reality of such incorporation by *Lithophaga* shells has been confirmed by Evelpidou et al. (2012a) for tsunami deposits in the Gulf of Euboea, however the chronological narrowing proposed by Shaw et al. (2010) cannot be applied to Crete before AD 365, because the time of the subsidence events dated by Pirazzoli et al. (1981, 1982) were not deduced from lithophagid dates, but from organic accretions of *Dendropoma*, *Neogoniolithon*, *Lithophyllum*, other calcareous algae, and *vermetids*, which are excellent sea-level indicators providing reliable radiocarbon-age estimations.

1956, Cyclades: Amorgos Earthquake

The Amorgos earthquake, on 9 July 1956 in the south central Aegean Sea, was one of the largest ($M_S = 7.4$) and most destructive crustal earthquakes in the twentieth century in the Aegean and was followed by a more destructive aftershock ($M_S = 7.2$) (Makropoulos et al. 1989).

The whole event resulted in 53 deaths and considerable damage notably on the island of Santorini and generated a local tsunami, which

affected the shores of the Cyclades and Dodecanese Islands, Crete, and the coast of Asia Minor, with run-up values of 30, 20, and 10 m reported on the southern coast of Amorgos, Astypalaia, and Folegandros, respectively (Papastamatiou et al. 1956; Galanopoulos 1960; Soloviev et al. 2000).

According to Evelpidou et al. (2012b), an extended underwater geomorphological survey located a well-developed submerged notch along the rocky coasts of Sifnos, Antiparos, Paros, Naxos, Iraklia, and Keros (Fig. 3), revealing widespread evidence of a recent 30–40 cm submergence, part of which may have seismic origin.

Comparison with information reported from earthquakes having affected the area suggests that at least part of the recent submergence might be an effect of the 1956 Amorgos earthquake. Modeling of the coseismic and short-term postseismic effects of the earthquake revealed that part of the observed subsidence may be explained in some of the islands by a fast postseismic relaxation of a low-viscosity layer underlying the seismogenic zone. However far-field observations are underestimated by

Evlepidou et al. (2012b) model and may be affected by a wider deformation field induced by the largest aftershock of the Amorgos sequence or by other earthquakes.

1953, Ionian: Cephalonia Earthquake

The region of Cephalonia and Ithaca islands is characterized by the frequent occurrence of shallow seismic events with magnitudes up to 7.2. Based on historical documents, large earthquake events took place in 1469, 1636, 1767, 1867, and 1953, while moderate-to-large ones occurred in 1658, 1723, 1742, 1759, 1766, 1862, 1912, 1915, 1925, 1932, and 1939 (Papagiannopoulos et al. 2012). The most destructive earthquake in the region hit the area in 1953. The 1953 series of earthquakes greatly damaged or destroyed 91 % and 70 % of all houses in Cephalonia and Ithaca, respectively. The first earthquake of the series occurred on 9 August and hit Ithaca and the town of Sami (Cephalonia). The second one occurred on 11 August and was greater than the first, destroying many towns around Cephalonia: Argostoli, Lixouri, Agia Efimia, and Valsamata. The third one on 12 August had a magnitude of 7.2 and completely destroyed Cephalonia, Ithaca, and Zakynthos. The number of human casualties as well as curves that illustrate the distribution of

seismic intensity has been given by Grandazzi (1954).

According to the map of seismic intensity by Grandazzi (1954, Fig. 3), in Cephalonia a seismic intensity of IX is observed for most part of the island, with the exception of Fiscardo peninsula, where the intensity was VII. In addition, in Ithaca island an intensity of IX was measured for the southern part, while in the northern one the intensity was smaller (VIII). This is also illustrated on the map of destruction by Grandazzi (1954, Fig. 4) (Poros, Cephalonia), where the percentage of destructions both in Ithaca and Cephalonia are increased toward the south.

The observations of Galanopoulos (1955) and Mueller-Miny (1957) about the uplift of the shorelines of Cephalonia during 12 August 1953 earthquake were also verified by Stiros et al. (1994). In fact, the 1953 earthquake left recognizable uplifted marks in several places around Cephalonia island, which reach a maximum elevation of +0.7 m in Poros area (Fig. 3). In fact, the uplift is ranging from 30 to 70 cm and has been evident in the central part of the island. According to Stiros et al. (1994), the uplifted part is bounded by two subparallel and homothetic major thrusts and can be better explained if assumed that the surface deformation reflects a

Paleoseismology of Rocky Coasts,

Fig. 4 Holes made by *Lithophaga* shells in a tufa layer of Miura Peninsula uplifted by the 1923 earthquake (Photo P.A.P. n° 595, 1974)





Paleoseismology of Rocky Coasts, Fig. 5 Uplifted fossil shorelines in Poros area (Cephalonia). The lower one at +0.7 m is ascribed to the 1953 earthquake

halotectonic deformation, indicating that the seismic deformation of the uppermost crustal strata in the area may mimic the style of the long-term halotectonic deformation.

The study of Stiros et al. (1994) shows that no postseismic displacement occurred after the 1953 coseismic uplift (Fig. 4) (Poros, Cephalonia).

1923, The Great Japan Earthquake

On 1 September 1923, shortly before noon (at 2 h 58' 44" GMT), a violent shock of earthquake occurred, shaking houses and other buildings intensely. A second and a third quake followed shortly after and then many aftershocks, spreading disaster all over. The afflicted zone covered seven prefectures: Tokyo, Kanagawa, Shizuoka, Chiba, Saitama, Yamanashi, and Ibaraki. The fire that followed reduced to ashes a great part of Tokyo and of the port of Yokohama.

According to the Bureau of Social Affairs (1926), the number of houses having been

damaged in Tokyo and Kanagawa Prefectures is of:

- 381.090 (54.9 %) entirely burnt
- 517 (0.1 %) partially burnt
- 83.319 (12.1 %) entirely collapsed
- 91.233 (13.1 %) partially collapsed
- 1.390 (0.2 %) swept away
- 136.572 (19.6 %) damaged

This makes a total of 558.049 (80.4 %) houses entirely or partially burnt, entirely or partially collapsed, or swept away due to the earthquake or the fire that followed.

According to <http://earthquake.usgs.gov/learn.today/index.php?r>, the magnitude M_L of this earthquake was 8.3 and it produced a death toll of 142,800.

The uplift of the ground in the disturbed zone was generally (Fig. 4) (Miura Peninsula) along 15 km of coast, especially in Tokyo Bay and Sagami Bay, reaching a maximum of 1.8 m in



Paleoseismology of Rocky Coasts, Fig. 6 Abrasion coastal platform uplifted in 1923 at Miura Peninsula. Water level (at Aburatsubo tidal station) is at -76 cm. Spring tidal amplitude is ± 54 cm (Photo P.A.P. 592, May 1974)

Chiba Prefecture, while more to the south (e.g., Oshima (Vries Island)) the ground sunk (Fig. 6).

Summary

Paleo-sea levels estimated through indicators such as geomorphological, biological, sedimentological, stratigraphical, or archeological may give evidence of coastal paleoseismicity. After the indicators' identification comes the important step of interpretation and dating of the event. Different types of indicators provide different amount of confidence and precision in determining the respective tectonic events, mainly due to the variety of their formation process and position as well as their subsequent evolution. Especially tidal notches may additionally provide information about whether a sea-level change has been gradual or coseismic, based on their profiles.

Four case studies of significant paleoseismic events are presented, all of which have greatly affected the coastal zone.

Cross-References

- ▶ [Archeoseismology](#)
- ▶ [Early Earthquake Warning \(EEW\) System: Overview](#)
- ▶ [EEE Catalogue: A Global Database of Earthquake Environmental Effects](#)
- ▶ [Luminescence Dating in Paleoseismology](#)
- ▶ [Paleoseismic Trenching](#)
- ▶ [Paleoseismology](#)
- ▶ [Radiocarbon Dating in Paleoseismology](#)
- ▶ [Tsunamis as Paleoseismic Indicators](#)

References

- Ambraseys NN, Jackson JA (1990) Seismicity and associated strain of central Greece between 1890 and 1988. *Geophys J Int* 101:663–708
- Evelpidou N, Vassilopoulos A, Pirazzoli PA (2012a) Holocene emergence in Euboea island (Greece). *Mar Geol* 295–298:14–19
- Evelpidou N, Melini D, Pirazzoli P, Vassilopoulos A (2012b) Evidence of a recent rapid subsidence in the S–E Cyclades (Greece): an effect of the 1956 Amorgos earthquake? *Cont Shelf Res* 39–40:27–40

- Flemming NC (1979) Archaeological indicators of sea level. In: NIVMER (ed) *Les Indicateurs de niveaux marins*. pp 149–166. *Oceanis*, 5, hors-série
- Galanopoulos A (1955) Seismic geography of Greece. *Ann Géol Pays Hell* 6:83–121
- Galanopoulos AG (1960) Tsunamis observed on the coasts of Greece from antiquity to present time. *Ann Geofisica* 13:369–386
- Grandazzi M (1954) Le tremblement de terre des Iles Ioniennes (aout 1953). *Ann Géogr* 63(340):431–453
- Kelletat D (2006) Beachrock as sea-level indicator? Remarks from a geomorphological point of view. *J Coast Res* 22:1558–1564
- Laborel J, Laborel-Deguen F (1994) Biological indicators of relative sea level variations and of co-seismic displacements in the Mediterranean region. *J Coast Res* 10(2):395–415
- Laborel J, Morhange C, Collina-Girard J, Laborel-Deguen F (1999) Littoral bioerosion, a tool for the study of sea-level variations during the Holocene. *Bull Geol Soc Den* 45:164–168
- Le Campion-Alsumard T (1979) Les végétaux perforants en tant qu'indicateurs paleobathymétriques. In: NIVMER (ed) *Les Indicateurs de niveaux marins*. pp 259–264. *Oceanis*, 5, hors-série
- Little TA, Van Dissen R, Schermer E, Came R (2009) Late Holocene surface ruptures on the southern Wairarapa fault, New Zealand: link between earthquakes and the uplifting of beach ridges on a rocky coast. *Lithosphere* 1(1):4–28
- Makropoulos K, Drakopoulos J, Latousakis J (1989) A revised and extended catalogue for Greece since 1900. *Geophys J Int* 98:391–394
- Mueller-Miny H (1957) Beiträge zur Morphologie der mittleren jonischen Inseln. *Ann Géol Pays Hell* 8:1–28
- Papagiannopoulos GA, Hatzigeorgiou GD, Beskos DE (2012) An assessment of seismic hazard and risk in the islands of Cephalonia and Ithaca, Greece. *Soil Dyn Earthq Eng* 32:15–25
- Papastamatiou J, Zachos K, Voutetakis S (1956) The earthquake of Santorini of 9 July 1956. Institute of Geology and Subsurface Research (I.G.S.R.). Athens, Greece
- Pavlidis S, Caputo R (2004) Magnitude versus faults' surface parameters: quantitative relationships from the Aegean. *Tectonophysics* 380(3–4):159–188
- Pirazzoli PA (1986) Marine notches. In: van de Plassche O (ed) *Sea-level research: a manual for the collection and evaluation of data*. Geo Books, Norwich, pp 361–400
- Pirazzoli PA (1994) Tectonic shorelines. In: Carter RWG, Woodroffe CD (eds) *Coastal evolution*. University Press, Cambridge, pp 451–476
- Pirazzoli PA (1996) *Sea-level changes – the last 20000 years*. Wiley, Chichester
- Pirazzoli PA, Thommeret J, Thommeret Y, Laborel J, Montaggioni LF (1981) Les rivages émergés d'Antikythira (Cerigotto): corrélations avec la Crète Occidentale et implications cinématiques et géodynamiques. In: *Actes du Colloque "Niveaux marins et tectonique quaternaires dans l'aire méditerranéenne"*. CNRS et University Paris I, Paris, pp 49–65
- Pirazzoli PA, Thommeret J, Thommeret Y, Laborel J, Montaggioni LF (1982) Crustal block movements from Holocene shorelines: Crete and Antikythira (Greece). *Tectonophysics* 86:27–43
- Shaw B, Jackson JA, Higham TFG, England PC, Thomas AL (2010) Radiometric dates of uplifted marine fauna in Greece: implications for the interpretation of recent earthquake and tectonic histories using lithophagid dates. *Earth Planet Sci Lett* 297:395–404
- Soloviev SL, Solovieva ON, Go CN, Kim KS, Shchetnikov NA (2000) *Tsunamis in the Mediterranean sea 2000 B.C.–2000 A.D.* Kluwer Academics, Dordrecht
- Stiros SC, Papageorgiou S (2001) Seismicity of Western Crete and the destruction of the town of Kisamos at AD 365; archaeological evidence. *J Seismol* 5:381–397
- Stiros SC, Pirazzoli PA, Laborel J, Laborel-Deguen F (1994) The 1953 earthquake in Cephalonia (Western Hellenic Arc): coastal uplift and halotectonic faulting. *Geophys J Int* 117:834–849

Paleoseismology: Integration with Seismic Hazard

Kuvvet Atakan

Department of Earth Science, University of Bergen, Bergen, Norway

Synonyms

Deterministic seismic hazard analysis; Earthquake geology; Earthquake hazard analysis; Paleoseismology; Probabilistic seismic hazard analysis; Seismic hazard analysis; Seismic hazard assessment

Introduction

Paleoseismology is the branch of science that aims to understand the earthquakes and their effects that have occurred in the geological past – i.e., during the Quaternary Period

(2.588 Ma to present), most commonly during the Holocene. Contributions from these past earthquakes turn out to be very critical not only to the understanding of the deformational processes within the Earth's crust but to a large extent also to develop better models for assessing seismic hazard. The main scope of this entry is to elaborate on the latter, the use of paleoseismological data in seismic hazard assessment.

Paleoseismology, also used as synonymous with "earthquake geology," has become an important and integrated part of the seismic hazard assessment during the late 1970s, following systematic work done on the San Andreas Fault in California, USA. Several studies in California (e.g., Sieh 1978; Wallace 1981, 1990; Sieh and Jahns 1984; Sieh et al. 1989; Grant and Sieh 1994; Grant and Lettis 2002) have been influential in making paleoseismology visible to other branches of Earth science. Researchers working with seismic hazard assessment became especially interested in these studies as they could see the importance of the paleoseismological data in their work. Later in the 1980s and 1990s, paleoseismological studies have spread in many other countries, most notably in Japan, Canada, New Zealand, Italy, Turkey, Greece, and Spain. In the late 1980s several monographs have contributed to a wider recognition of paleoseismology as an important field of geology (e.g., Wallace 1986; Vita-Finzi 1986; Crone and Omdahl 1987). In the late 1990s the two books published on paleoseismology by McCalpin (1996) and Yeats et al. (1997) were influential in introduction paleoseismological studies to a wider community.

Main Elements of a Paleoseismological Study

Large earthquakes occurring on crustal scale faults leave visible traces on the Earth's surface during their rupture process. This evidence on the surface is subject to both erosional and depositional processes controlled by the climatic conditions. As a consequence these evidences may either be destroyed or buried.

Paleoseismological studies are usually conducted on these faults with the aim of identifying individual paleoearthquakes with quantitative constraints on their age and size. Final aim is usually try to establish a maximum earthquake magnitude and a recurrence interval for a given fault.

Paleoseismological studies contain a number of distinct elements. They usually start with a regional analysis on the tectonic setting and structural geology of the area of interest. This is done to find out the regional deformation rates and includes geodetic strain rates obtained through the time-lap analysis of GPS data. Once the regional strain rate is known, it is possible to look for evidence at a local scale that can be expected from that rate of deformation. These regional scale studies are then followed by various types of analyses of the fault segment(s), through geological, geophysical, geodetic techniques, as well as detailed analyses of the local stratigraphy in order to select a suitable site(s) for detailed analysis. Trenching is done at the selected site(s) for identifying the paleoearthquakes. Various types of evidence are then used to identify the individual paleoearthquakes. In order to constrain the ages of these events, various dating techniques are applied on available material on event horizons within the trench stratigraphy. This is followed by establishing the slip per event and the rupture length. Based on these, finally a maximum magnitude and a recurrence interval for the fault are established. These parameters are especially important in probabilistic seismic hazard assessment (PSHA). In the following section how these parameters are integrated in seismic hazard assessment are explained.

Integrating Paleoseismological Data in Seismic Hazard Assessment

Integrating paleoseismological data in Seismic Hazard Assessment (SHA) is dependent on the type of methodology used, i.e., probabilistic (PSHA) or deterministic seismic hazard assessment (DSHA).

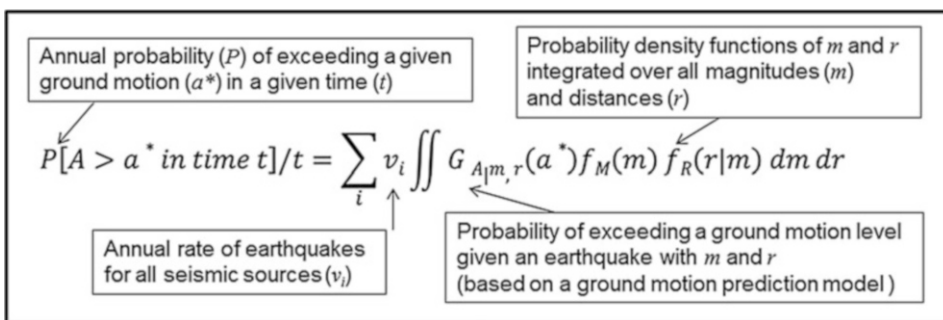
Paleoseismic Data and Probabilistic Seismic Hazard Assessment

Standard probabilistic seismic hazard analysis (Cornell 1968; McGuire 1993) is based on statistical treatment of earthquakes using probability density functions assuming different types of earthquake recurrence (Fig. 1). One of the most common applications of PSHA is based on a Poissonian assumption for earthquake recurrence, which assumes that the earthquake occurrence has not have memory, i.e., occurrence of earthquakes in a given area in the future has not have relation to the occurrence of previous earthquakes in the same region. Such a “random” occurrence of earthquakes is applied, not because the earthquakes do not follow a certain pattern of occurrence governed by the physical processes, but because of recognition of the fact that it is simply not always possible to assess these physical parameters that lead to stress accumulation and release process. Therefore the simple assumption of a “random” earthquake occurrence is valid, and hence it is an attempt to cover the large uncertainties associated with the complex physical processes that lead to rupture along a fault. One of the important implications of applying a recurrence relation, let it be Poissonian or other types such as conditional, exponential, Weibull, Brownian, gamma, or log-normal (e.g., Main 1995; Mathews et al. 2002), is that the complete earthquake history is known and that it represents a statistically valid sample to develop a reliable estimate of the frequency of occurrence of various magnitude

levels in a given area, the so-called Gutenberg-Richter relation.

The completeness criterion in earthquake catalogs is not always met. This is especially true in areas of low seismicity where the instrumental and historical earthquake catalogs do not cover the occurrence of large and destructive earthquakes when the recurrence time of these is beyond the limits of the total catalog time span. Using such incomplete catalogs in developing the Gutenberg-Richter relation (Fig. 2) may lead to erroneous estimate of the seismic hazard in an area. Important parameters, such as the maximum expected earthquake magnitude and its frequency of occurrence are obtained, based on the maximum observed earthquake. This may be subject to interpretation as, in many cases, simple extrapolation of the frequency-magnitude relation curve will not give realistic estimates of the maximum expected magnitude (m_2 in Fig. 2). What is usually adopted is then a truncation to the curve (m_1 in Fig. 2), whereas in reality paleoseismological data may imply a larger (or lower) level on the maximum expected magnitude (m_3 in Fig. 2). In these cases, it is essential to extend the catalog time span by investigating the earthquake occurrence in the geological past through paleoseismological studies. Extending the time span of the catalogs with the occurrence of large earthquakes in the geological past brings important constraints in establishing the maximum expected magnitude in an area.

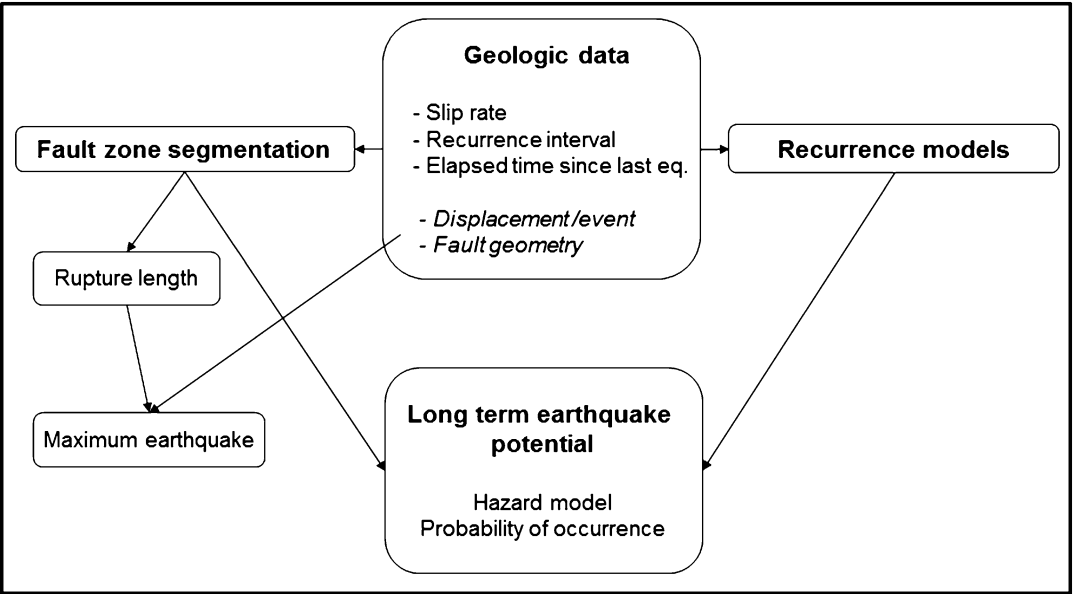
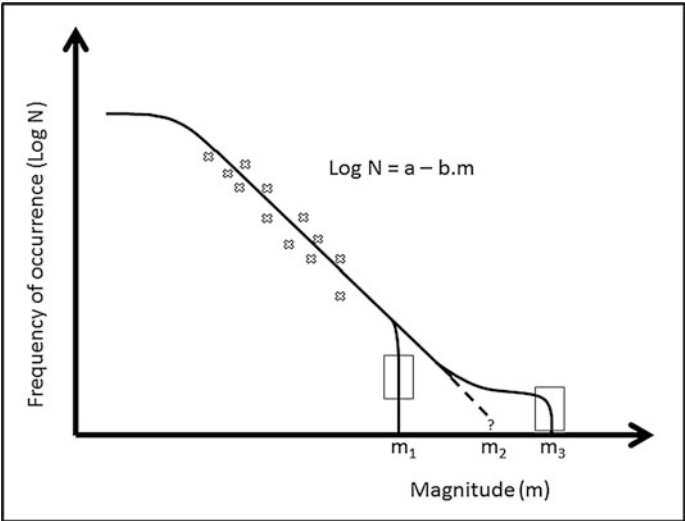
Applying paleoseismological data to probabilistic seismic hazard assessment (Fig. 3) using



Paleoseismology: Integration with Seismic Hazard, Fig. 1 Standard probabilistic seismic hazard computations (Cornell 1968; McGuire 1993)

**Paleoseismology:
Integration with Seismic
Hazard, Fig. 2**

A hypothetical case of frequency of occurrence of earthquake magnitudes (Gutenberg-Richter relation). Crosses refer to the individual data points of the cumulative number of frequency in the earthquake catalog which is usually limited in incomplete earthquake catalogs. The different possibilities then exist in estimating the maximum expected magnitude (e.g., m_1 , m_2 , or m_3)



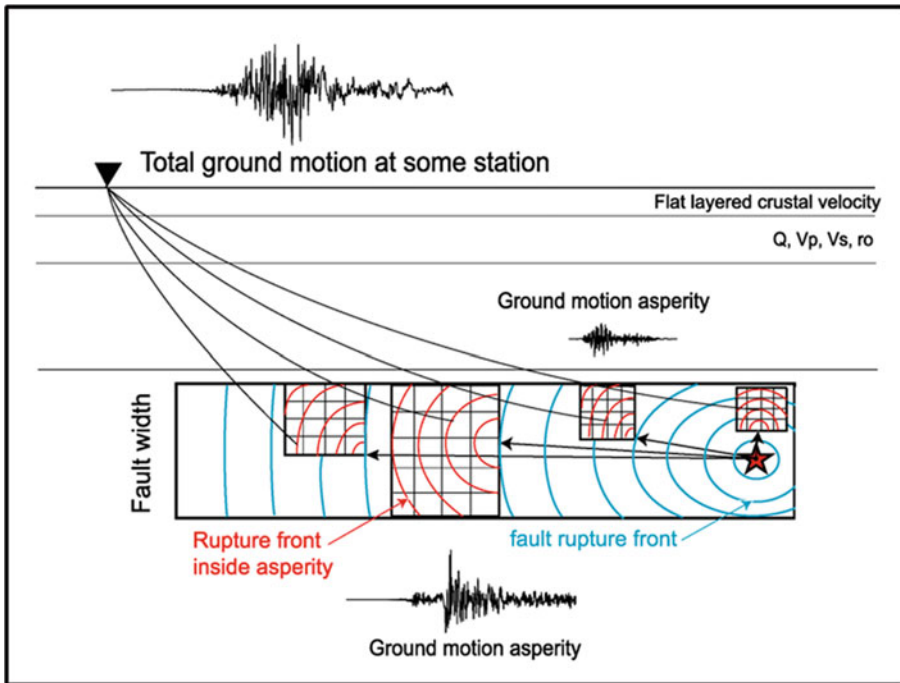
Paleoseismology: Integration with Seismic Hazard, Fig. 3 The use of geologic data in assessing the long-term earthquake potential (Redrawn from Schwartz and Coppersmith 1986)

time-dependent recurrence models on the other hand has typically been related to establishing the last occurrence of a large earthquake along a given fault and its recurrence in time. Together with slip rate this information is then used to develop recurrence models for the probabilistic computations. The maximum expected magnitude of a future earthquake is also used based on

the slip per event and the fault geometry together with the study of the fault segmentation and the rupture length (Schwartz and Coppersmith 1986).

Paleoseismic Data and Deterministic Seismic Hazard Assessment

While PSHA provides a useful tool for defining earthquake design loads for noncritical



Paleoseismology: Integration with Seismic Hazard, Fig. 4 A simplified sketch showing the main principles of ground motion simulations based on earthquake scenarios

including the fault rupture complexity (Courtesy of N. Pulido, 2004)

structures, especially for relatively low hazard levels (Bommer et al. 2000), in cases where the hazard is dominated by a single large earthquake from a nearby fault (e.g., Pulido et al. 2004; Sørensen et al. 2007) or if the engineering design of a critical structure requires realistic ground motions, it is preferable to perform deterministic seismic hazard assessment (DSHA). This is because in DSHA, the earthquake scenarios are defined unambiguously.

During the last decade or so, there has been a growing interest in assessing seismic hazard using ground motion simulations based on deterministic earthquake scenarios. These scenarios are usually built upon a fault rupture model where the rupture complexity is addressed either kinematically or dynamically. In kinematic models (Fig. 4) the variation of slip along the fault is modeled through predefined asperities where dynamic rupture parameters such as the rupture velocity and rise time are kept constant or varied within a given range. The obtained

ground motions for each element of the sub-faults are then propagated through a crustal velocity structure either for a simple flat-layered model or more complex 3-D structures to obtain the ground motions on the surface. The simulations can be done either for a broad frequency band using full waveform modeling for the low-frequency part of the ground motion and stochastic simulations for the high-frequency part (e.g., Pulido et al. 2004; Sørensen et al. 2007) or only using stochastic simulations (Boore 2009; Gofrani et al. 2013). Other more complicated uses of dynamic rupture models are only conducted for a few cases where data to constrain the dynamic rupture parameters as well as the three-dimensional velocity structure of the wave propagation path and the local site conditions are available (e.g., Olsen et al. 1997; Olsen 2000).

The use of paleoseismological data in developing better earthquake rupture scenarios in ground motion simulations has significant

potential which is not currently well exploited. Although most of the information regarding the fault geometry, its kinematics, rupture length, maximum magnitude, etc., is based on paleoseismological data, other constraints such as the slip distribution from past earthquakes along the fault, the location, and the size of the asperities could be better utilized in the computations. It is therefore important that paleoseismological studies focus not only on the standard parameters relevant for PSHA (e.g., maximum magnitude and recurrence interval) but also focus on other parameters, such as the location of the rupture initiation point and rupture propagation direction with its possible effects on the surface (e.g., Dor et al. 2008), as well as the details of the fault segmentation, asperities, and slip distribution.

Uncertainties in Paleoseismological Studies

Using paleoseismological data in seismic hazard assessment requires that the associated uncertainties are quantified and can be accounted for in the hazard computations. Although there are well-established methodologies that exist accounting for the various types of uncertainties (aleatory or epistemic) in seismic hazard analysis (e.g., Abrahamson and Bommer 2005), quantifying uncertainties in paleoseismic studies are usually only restricted to the analytical uncertainties associated with the age determinations. A few exceptions exist, systematically accounting for uncertainties associated with the paleoseismological (e.g., Atakan et al. 2000) and archeoseismological (Sintubin and Stewart 2008) studies.

Summary

Importance of paleoseismic data in seismic hazard assessment has been recognized since the late 1970s. There are well-established methodologies developed for integrating the paleoseismic data in probabilistic seismic hazard analysis, which usually involve extending the earthquake catalog

time span back in the geological past. This is especially true for areas of low deformation rate with sporadic seismic activity, such as plate interiors or stable continental regions, or in areas where the historical earthquake records are incomplete and scarce. The typical application of paleoseismic data in this case is associated with establishing the maximum magnitude based on the paleoearthquakes on a given fault and the recurrence in time. In seismic hazard analysis using renewal models where conditional probabilities are calculated, the occurrence of the last earthquake along a given fault as well as the magnitude and the recurrence interval are parameters that are typically used. There is however an unexploited potential of paleoseismic data in the deterministic seismic hazard studies using ground motion simulations based on earthquake rupture scenarios. The details of the slip distribution, repeated directivity effects, location and size of the asperities, as well as the rupture initiation point can be assessed by paleoseismic investigations. The use of such data in ground motion simulations requires a systematic analysis of the uncertainties associated with the paleoseismological data, not only those related to the analytical uncertainties of the age determinations but also those that are related to the interpretation of the various indices used in identifying the paleoearthquakes and their source parameters.

Cross-References

- [Archeoseismology](#)
- [Earthquake Location](#)
- [Earthquake Magnitude Estimation](#)
- [Earthquake Mechanisms and Stress Field](#)
- [Earthquake Mechanisms and Tectonics](#)
- [Earthquake Recurrence](#)
- [Earthquake Recurrence Law and the Weibull Distribution](#)
- [Earthquake Return Period and Its Incorporation into Seismic Actions](#)
- [Integrated Earthquake Simulation](#)
- [Luminescence Dating in Paleoseismology](#)
- [Paleoseismology of Glaciated Terrain](#)
- [Paleoseismic Trenching](#)

- [Paleoseismology](#)
- [Paleoseismology and Landslides](#)
- [Paleoseismology of Rocky Coasts](#)
- [Physics-Based Ground-Motion Simulation](#)
- [Probabilistic Seismic Hazard Models](#)
- [Radiocarbon Dating in Paleoseismology](#)
- [Random Process as Earthquake Motions](#)
- [Remote Sensing in Seismology: An Overview](#)
- [Seismic Actions Due to Near-Fault Ground Motion](#)
- [Seismic Risk Assessment, Cascading Effects](#)
- [Site Response for Seismic Hazard Assessment](#)
- [Spatial Variability of Ground Motion: Seismic Analysis](#)
- [Tsunamis as Paleoseismic Indicators](#)
- [Uncertainty Theories: Overview](#)

References

- Abrahamson NA, Bommer JJ (2005) Probability and uncertainty in seismic hazard analysis. *Earthq Spectra* 21(2):603–607. doi:10.1193/1.1899158
- Atakan K, Midzi V, Toirán BM, Vanneste K, Camelbeeck T, Meghraoui M (2000) Seismic hazard in regions of present day low seismic activity: uncertainties in paleoseismic investigations in the Bree fault scarp (Roer Graben, Belgium). *Soil Dyn Earthq Eng* 20(5–8):415–427
- Bommer JJ, Scott SG, Sarva SK (2000) Hazard-consistent earthquake scenarios. *Soil Dyn Earthq Eng* 19:219–231
- Boore DM (2009) Comparing stochastic point-source and finite-source ground-motion simulations: SMSIM and EXSIM. *Bull Seismol Soc Am* 99(6):3202–3216. doi:10.1785/0120090056
- Cornell CA (1968) Engineering seismic risk analysis. *Bull Seismol Soc Am* 58:1583–1906
- Crone AJ, Omdahl EM (eds) (1987) Directions in paleoseismology. US Geological survey open file report 87-673, 456 p
- Dor O, Yildirim C, Rockwell TK, Ben-Zion Y, Emre O, Sisk M, Duman TY (2008) Geological and geomorphologic asymmetry across the rupture zones of the 1943 and 1944 earthquakes on the North Anatolian Fault: possible signals for preferred earthquake propagation direction. *Geophys J Int* 173:483–504. doi:10.1111/j.1365-246X.2008.03709.x
- Ghofrani H, Atkinson GM, Goda K, Assatourinas K (2013) Stochastic finite fault simulations of the 2011 Tohoku, Japan, earthquake. *Bull Seismol Soc Am* 103(28):1307–1320. doi:10.1785/0120120228
- Grant LB, Lettis WR (2002) Introduction to the special issue on paleoseismology of the San Andreas Fault system. *Bull Seismol Soc Am* 92(7):2551–2554. doi:10.1785/0120000600
- Grant LB, Sieh K (1994) Paleoseismic evidence of clustered earthquakes on the San Andreas Fault in the Carrizo Plain, California. *J Geophys Res Solid Earth* 99(B4):6819–6841. doi:10.1029/94JB00125
- Main IG (1995) Earthquakes as critical phenomena: implications for probabilistic seismic hazard analysis. *Bull Seismol Soc Am* 85(5):1299–1308
- Mathews MV, Ellsworth WL, Reasenber PA (2002) A Brownian model for recurrent earthquakes. *Bull Seismol Soc Am* 92(6):2233–2250. doi:10.1785/0120010267
- McCalpin J (ed) (1996) *Paleoseismology*. Academic, London, 553 p. ISBN 0-12-481826-9
- McGuire R (1993) Computations of seismic hazard. *Anal di Geofisica XXXVI*(3–4):181–200
- Olsen KB (2000) Site amplification in the Los Angeles basin from three-dimensional modeling of ground motion. *Bull Seismol Soc Am* 90(6B):S77–S94. doi:10.1785/0120000506
- Olsen KB, Madariaga R, Archuleta RJ (1997) Three-dimensional dynamic simulation of the 1992 landers earthquake. *Science* 278(5339):834–838. doi:10.1126/science.278.5339.834
- Pulido N, Ojeda A, Atakan K, Kubo T (2004) Strong ground motion estimation in the Marmara Sea region (Turkey) based on a scenario earthquake. *Tectonophysics* 391:357–374
- Schwartz DP, Coppersmith KJ (1986) Seismic hazards: new trends in analysis using geologic data. In: *Active tectonics*. National Academy Press, Washington, DC, pp 215–230
- Sieh KE (1978) Prehistoric large earthquakes produced by slip on the San Andreas Fault at Palmett Creek, California. *J Geophys Res Solid Earth* 83(B8):3907–3939. doi:10.1029/JB083iB08p03907
- Sieh KE, Jahns RH (1984) Holocene activity of the San Andreas Fault at Wallace Creek, California. *Geol Soc Am Bull* 95(8):883–896. doi:10.1130/0016-7606
- Sieh KE, Stuiver M, Brillinger D (1989) A more precise chronology of earthquakes produced by the San Andreas Fault in southern California. *J Geophys Res Solid Earth* 94(B1):603–623. doi:10.1029/JB094iB01p00603
- Sintubin M, Stewart I (2008) A logical methodology for archaeoseismology: a proof of concept at the archaeological site of Sagalassos, Southwest Turkey. *Bull Seismol Soc Am* 98(5):2209–2230. doi:10.1785/0120070178
- Sørensen MB, Pulido N, Atakan K (2007) Sensitivity of ground motion simulations to earthquake source parameters: a case study for Istanbul, Turkey. *Bull Seismol Soc Am* 97(3):881–900. doi:10.1785/0120060044
- Vita-Finzi (1986) *Recent earth movements: an introduction to neotectonics*. Academic, London, 226p. ISBN 0127223703

- Wallace RE (1981) Active faults, paleoseismology, and earthquake hazards in the western United States. In: Simpson DW, Richards PG, Wallace RE (eds) Earthquake prediction. American Geophysical Union (AGU) Publications, Washington, DC, USA, doi:10.1029/ME004p0209
- Wallace RE (1986) Active tectonics, Studies in geophysics. National Academy of Sciences, Washington, DC
- Wallace RE (ed) (1990) The San Andreas Fault system, California. US Geological survey professional paper 1515
- Yeats RS, Sieh KE, Allen CR (1997) The geology of earthquakes. Oxford University Press, New York, 568p. ISBN 0195078276

Parametric Nonstationary Random Vibration Modeling with SHM Applications

Luis David Avendaño-Valencia and
Spilios D. Fassois
Department of Mechanical & Aeronautical
Engineering, Stochastic Mechanical Systems and
Automation (SMSA) Laboratory, University of
Patras, Patras, Greece

Synonyms

Fault diagnosis; Nonstationary random vibration; Signal-based modeling (identification); Structural Health Monitoring; Time-dependent ARMA modeling; Time-frequency analysis

Introduction

Nonstationary random vibration is characterized by *time-dependent (evolutionary)* characteristics (Priestley 1988; Roberts and Spanos 1990, Chapter 7; Bendat and Piersol 2000, Chapter 12; Newland 1993, pp. 211–219; Preumont 1994, Chapter 8; Hammond and White 1996; Kitagawa and Gersch 1996). Typical examples of nonstationary random vibration include earthquake ground motion and resulting structural vibration response, as well as the vibration of surface vehicles, flying aircraft, mechanisms,

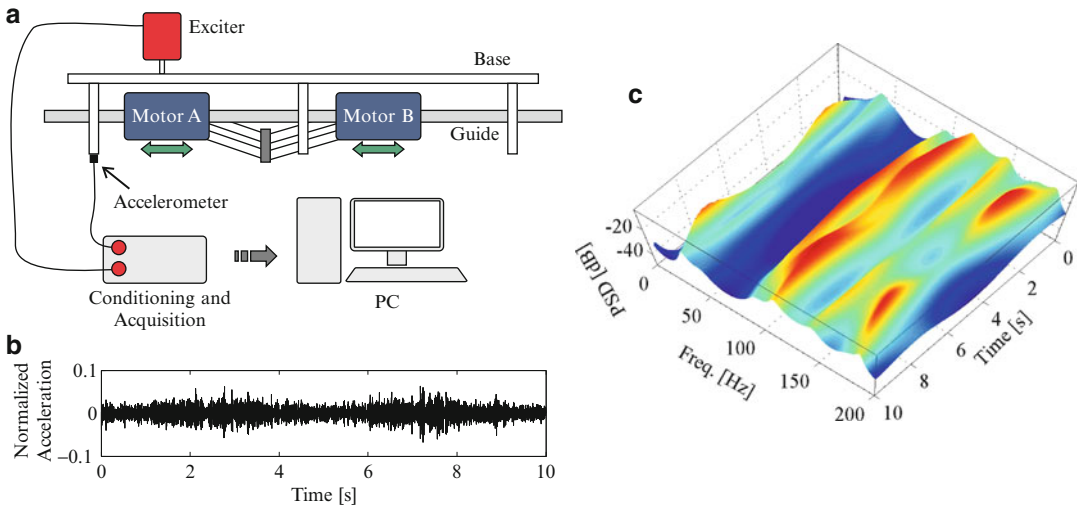
rotating machinery, cranes, bridges with passing vehicles, and so on. Nonstationary random vibration typically originates from *time-varying* dynamics or the linearization of *nonlinear* dynamics.

An example of a system exhibiting nonstationary random vibration is the mechanism of Fig. 1a. It is a 2-DOF pick-and-place mechanism consisting of two coaxially aligned linear motors carrying prismatic links (arms) connected to their ends. The mechanism is clamped on an aluminum base and is excited by a zero-mean Gaussian stationary random excitation force, applied vertically with respect to the base by means of an electromechanical shaker, while the linear motors are following predetermined trajectories. The nonstationary nature of the resulting vibration (see Fig. 1b) is due to the time-varying position of the linear motors, thus controlling the position of the links, and is evident in the time-varying power spectral density (TV-PSD) parametric estimate of Fig. 1c. Further details on this example may be found in (Spiridonakos and Fassois 2013).

From a mathematical point of view, nonstationary random vibration is characterized by time-dependent statistical moments. In the presently assumed *Gaussian* case, this means that the mean is a function of time t and the autocovariance function (ACF) a function of two time arguments t_1 and t_2 . That is, for a random vibration signal $x(t)$, one has $E\{\cdot\}$ designates statistical expectation)

$$\begin{aligned}\text{Mean : } \mu(t) &= E\{x(t)\}, \\ \text{ACF : } \gamma(t_1, t_2) &= E\{(x(t_1) - \mu(t_1)) \cdot (x(t_2) - \mu(t_2))\} \\ &\quad (1)\end{aligned}$$

In many random vibration problems, the mean is constant and is thus easily estimated and removed from the signal (sample-mean adjusted signal). The case of a time-dependent mean (also referred to as deterministic trend) may be treated via proper techniques, such as curve fitting, high-pass filtering, or special parametric models (such as integrated models with a deterministic trend parameter (Box et al. 1994)).



Parametric Nonstationary Random Vibration Modeling with SHM Applications, Fig. 1 Example of a laboratory pick-and-place mechanism exhibiting nonstationary random vibration: (a) schematic diagram

of the mechanism and the laboratory setup; (b) measured vibration acceleration response signal (normalized); (c) estimate of the TV-PSD (time-frequency spectrum)

This entry focuses on the following two subjects: (i) The signal-based modeling (identification) of nonstationary random vibration based on a uniformly sampled (with sampling period T) signal realization $x[t]$, for $t = 1, 2, \dots, N$. Absolute time is $t \cdot T$, N is the signal length in samples, while the use of brackets indicates function of an integer variable. (ii) The use of an identified random vibration model (or in fact a set of such models) for Structural Health Monitoring (SHM), where the objective is the *detection* of potential structural damage and its *characterization (identification)* (Fassois and Sakellariou 2009). Other important uses of an identified model – not discussed in this entry – include model-based analysis (like the extraction of time-dependent power spectral density (PSD) and time-dependent vibration modes (Newland 1993, p. 218; Preumont 1994, Chapter 8; Poulimenos and Fassois 2006)), prediction, classification, and control.

Nonstationary random vibration signal-based modeling (identification) has received significant attention in recent years. The available methods may be broadly classified as *parametric* or *nonparametric* (Poulimenos and Fassois 2006).

Nonparametric methods have received most of the attention and are based upon nonparameterized representations of the vibration energy as a simultaneous function of time and frequency (time-frequency representations). They include the classical spectrogram (based upon the short-time Fourier transform – STFT) and its ramifications (Newland 1993, p. 218; Hammond and White 1996; Bendat and Piersol 2000, p. 504), Mark's physical spectrum (Preumont 1994, Section 8.3), the Cohen class of distributions (Hammond and White 1996), Priestley's evolutionary spectrum (Priestley 1988, Section 6.3; Preumont 1994, Section 8.4), as well as wavelet-based methods (Newland 1993, Chapter 17). On the other hand, *parametric methods* are based upon parameterized representations, usually of the *time-dependent autoregressive moving average (TARMA)* type. These have an apparently similar form to their conventional (stationary) counterparts, but are characterized by *time-dependent* parameters and innovations variance. Thus, a $TARMA(n_a, n_c)$ model, with n_a and n_c designating its AR and MA orders, respectively, is defined as follows (Poulimenos and Fassois 2006):

TARMA(n_a, n_c) model :

$$x[t] = -\sum_{i=1}^{n_a} a_i[t]x[t-i] + \sum_{i=1}^{n_c} c_i[t]w[t-i] + w[t];$$

$$w[t] \sim \text{NID}(0, \sigma_w^2[t])$$
(2)

where $x[t]$ represents the nonstationary random signal model; $w[t]$ an unobservable *normally and independently distributed* (NID) (thus *white*) nonstationary *innovations* sequence with zero-mean and time-dependent variance $\sigma_w^2[t]$; n_a, n_c the autoregressive (AR) and moving average (MA) orders, respectively; and $a_i[t]$ and $c_i[t]$ the corresponding AR and MA *time-dependent parameters*.

Parametric TARMA models are of three main families, according to the form of “structure” imposed upon the evolution of the time-dependent parameters and innovations variance (Poulimenos and Fassois 2006):

- (a) *Unstructured parameter evolution (UPE)* models, in which no particular “structure” is imposed on the parameter evolution. Prime models in this family include short-time ARMA (ST-ARMA) and recursive models (such as recursive ARMA or in short RARMA) models.
- (b) *Stochastic parameter evolution (SPE)* models, in which *stochastic* “structure” is imposed on the parameter evolution via stochastic smoothness constraints. Prime models in this family include smoothness-priors ARMA (SP-ARMA) models.
- (c) *Deterministic parameter evolution (DPE)* models, in which deterministic “structure” is imposed on the parameter evolution. Prime models in this family include the so-called functional series TARMA (FS-TARMA) models in which the parameters are projected on properly selected functional subspaces.

Parametric models, and their respective signal-based (identification) methods, are known to be characterized by a number of important advantages, such as representation parsimony, improved accuracy and resolution,

improved tracking of the TV dynamics, flexibility in analysis, synthesis (simulation), prediction, diagnosis, and control. For instance, once a TARMA model is available, the corresponding “frozen”-type TV-PSD may be readily obtained as

“Frozen” TV-PSD :

$$S_F(\omega, t) = \frac{\left| 1 + \sum_{i=1}^{n_c} c_i[t] \cdot e^{-j\omega T_{si}} \right|^2}{\left| 1 + \sum_{i=1}^{n_a} a_i[t] \cdot e^{-j\omega T_{si}} \right|^2} \cdot \sigma_w^2[t]$$
(3)

with ω representing frequency in *rad/s*, j the imaginary unit, and $|\cdot|$ complex magnitude. Notice that this would be the PSD of the vibration signal if the system were “frozen” (made stationary) at each time instant t . This entry focuses on parametric models and methods and in particular on the SPE and DPE families.

The problem of Structural Health Monitoring (SHM) for structures exhibiting nonstationary random vibration responses may be treated in a statistical time series (STS) framework (Fassois and Sakellariou 2009), using either nonparametric or parametric models and statistical decision-making schemes. In the majority of applications thus far, nonparametric models (nonparametric time-frequency-type representations) are used, and damage detection and identification are based on potential discrepancies observed between those obtained in a *baseline* (healthy) phase and an *inspection* (current) phase (Feng et al. 2013). Although the use of parametric models and methods may potentially lead to performance improvements, it has thus far received limited attention (Poulimenos and Fassois 2004; Spiridonakos and Fassois 2013).

Brief Historical Notes. Among parametric models, the unstructured parameter evolution (UPE) family of methods was initially developed. Prime methods in this area include the short-time ARMA (ST-ARMA) method (Niedzwieki 2000, pp. 79–82; Owen et al. 2001) and the class of recursive (or adaptive) methods (Ljung 1999, Chapter 11; Niedzwieki 2000, Chapters 4 and 5).

The stochastic parameter evolution (SPE) family of methods was developed primarily for the modeling of earthquake ground motion signals (Kitagawa and Gersch 1996; Gersch and Akaike 1988). The deterministic parameter evolution (DPE) family was introduced in a broader context in Rao (1970) and later in Kozin (1977). In Kozin (1988) and Fouskitakis and Fassois (2002) it was employed for earthquake ground motion modeling. In Poulimenos and Fassois (2009b) it was applied to the vibration of a bridge-like structure with a moving mass. In a broader context the reader is also referred to (Niedzwieki 2000, Chapter 4).

Article Roadmap. This entry is organized as follows: SPE and DPE nonstationary random vibration modeling is discussed in section “[Parametric TARMA Modeling of Nonstationary Random Vibration](#),” where specific model forms and identification schemes are briefly reviewed. Structural Health Monitoring (SHM) based on nonstationary random vibration parametric modeling is presented in section “[Structural Health Monitoring \(SHM\) Based on Nonstationary Random Vibration](#).” The application of these concepts to random vibration modeling and SHM for the pick-and-place mechanism of Fig. 1 is outlined in section “[Illustrative Example: Nonstationary Random Vibration Modeling and SHM for a Pick-and-Place Mechanism](#),” while a summary is provided in section “[Summary](#).”

Parametric TARMA Modeling of Nonstationary Random Vibration

Stochastic Parameter Evolution (SPE) TARMA Modeling

In the context of *stochastic parameter evolution (SPE)* TARMA models, the parameters are assumed to follow *stochastic smoothness constraints* in the form of linear integrated autoregressive (IAR) models with integration order q (in this context referred to as *smoothness-priors order*). A *smoothness-priors* TARMA (SP-TARMA) model thus has parameters that obey the relations (Kitagawa and Gersch 1996)

$$(1 - B)^q \cdot a_i[t] = v_{ai}[t], \quad v_{ai}[t] \sim \text{NID}(0, \sigma_v^2) \quad (4a)$$

$$(1 - B)^q \cdot c_i[t] = v_{ci}[t], \quad v_{ci}[t] \sim \text{NID}(0, \sigma_v^2) \quad (4b)$$

where the signal innovations $w[t]$ (see Eq. 2) and the parameter innovations $v_{ai}[t]$ and $v_{ci}[t]$ are all mutually independent and normally and identically distributed (NID) sequences, each being zero mean and with variance $\sigma_w^2[t]$ and σ_v^2 , respectively. These smoothness constraints are characterized by unit roots that represent integrated stochastic models describing homogeneously nonstationary evolutions (Box et al. 1994, Chapter 4).

Generalization of SP-TARMA models, in the form of *generalized stochastic constraint time-dependent autoregressive moving average* TARMA (GSC-TARMA) models, was recently introduced (Avendaño-Valencia and Fassois 2013). In this, the model parameters are allowed to follow more general autoregressive (AR) models of the forms

$$a_i[t] = -\sum_{k=1}^q \mu_k \cdot a_i[t-k] + v_{ai}[t], \quad (5a)$$

$$v_{ai}[t] \sim \text{NID}(0, \sigma_v^2)$$

$$c_i[t] = -\sum_{k=1}^q \mu_k \cdot c_i[t-k] + v_{ci}[t], \quad (5b)$$

$$v_{ci}[t] \sim \text{NID}(0, \sigma_v^2),$$

where, as in the SP-TARMA case, $w[t]$, $v_{ai}[t]$, and $v_{ci}[t]$ are mutually independent and normally and identically distributed (NID) innovation sequences. The coefficients μ_k are referred to as the *stochastic constraint parameters*. These are collected in the vector $\boldsymbol{\mu} = [\mu_1 \cdot \dots \cdot \mu_q]^T$, which is along with the covariance $\boldsymbol{\Sigma}_v = \sigma_v^2 \cdot \mathbf{I}_{n_a+nc}$ (where $\mathbf{I}_{n_a+nc}$ is the identity matrix with the indicated dimensions), and the innovations variance $\sigma_w^2[t]$ defines the model *hyperparameters*. Then the time-dependent AR/MA parameters $\boldsymbol{\theta}[t]$, hyperparameters $\mathcal{P}$, and structural parameters $\mathcal{M}$ of a GSC-TARMA model are

$$\begin{aligned}\boldsymbol{\theta}[t] &= [a_1[t] \cdots a_{n_a}[t] \vdots c_1[t] \cdots c_{n_c}[t]]^T, \\ \mathcal{P} &= \{\boldsymbol{\mu}, \sigma_w^2[t], \boldsymbol{\Sigma}_v\}, \quad \mathcal{M} = \{n_a, n_c, q\}\end{aligned}\quad (6)$$

It should be noted that the stochastic constraint parameters and parameter innovations variances may be – more generally – different for each AR/MA parameter, but the above simple form is adopted here for purposes of presentation simplicity. Comparing the GSC-TARMA and SP-TARMA model forms, one sees that in the latter case the stochastic constraint parameters are essentially prefixed, limiting the types of trajectories that each AR/MA parameter would be capable of following.

The identification of an SP-TARMA or GSC-TARMA model consists of the selection of the model structure $\mathcal{M}$ and the estimation of the time-dependent parameter vector $\boldsymbol{\theta}[t]$ and hyperparameters $\mathcal{P}$. The estimation of the model parameters and hyperparameters may be posed as the maximization of the *joint a posteriori probability* density function of $\boldsymbol{\theta}_1^N = \{\boldsymbol{\theta}[1], \dots, \boldsymbol{\theta}[N]\}$ given the available observations $x_1^N = \{x[1], \dots, x[N]\}$, namely, $p(\boldsymbol{\theta}_1^N, \mathcal{P} | x_1^N)$. These, combined with the Gaussianity assumption for $w[t]$ and

$\mathbf{v}[t] = [v_{a_1}[t] v_{a_2}[t] \cdots v_{c_{n_c}}[t]]^T$, lead to the following cost function:

$$\begin{aligned}\mathcal{J}(\boldsymbol{\theta}_1^N, \mathcal{P}) &= \frac{N \cdot q \cdot (n_a + n_c)}{2} \ln |\sigma_v^2| \\ &\quad + \frac{1}{2} \sum_{t=1}^N \left(\ln \sigma_w^2[t] + \frac{w^2[t]}{\sigma_w^2[t]} + \frac{\mathbf{v}^T[t] \cdot \mathbf{v}[t]}{\sigma_v^2} \right)\end{aligned}\quad (7)$$

which must be minimized in order to provide the optimal *maximum a posteriori* (MAP) estimate of $\boldsymbol{\theta}[t]$ and $\mathcal{P}$.

An estimate of $\boldsymbol{\theta}[t]$ based on fixed values of $\mathcal{P}$ may be obtained recursively using the Kalman filter (or a proper nonlinear approximation filter in the full TARMA case) based on the following state-space representation of the SP/GSC-TARMA model (Poulimenos and Fassois 2006):

$$\mathbf{z}[t] = \mathbf{F}(\boldsymbol{\mu}) \cdot \mathbf{z}[t-1] + \mathbf{G} \cdot \mathbf{v}[t] \quad (8a)$$

$$x[t] = \mathbf{h}^T[t] \cdot \mathbf{z}[t] + w[t] \quad (8b)$$

with:

$$\begin{aligned}\mathbf{z}[t-1] &= \begin{bmatrix} \boldsymbol{\theta}[t-1] \\ \boldsymbol{\theta}[t-2] \\ \vdots \\ \boldsymbol{\theta}[t-q] \end{bmatrix}, \quad \mathbf{F}(\boldsymbol{\mu}) = \begin{bmatrix} \mu_1 & \mu_2 & \cdots & \mu_{q-1} & \mu_q \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \otimes \mathbf{I}_{n_a+n_c}, \\ \mathbf{G} &= \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \otimes \mathbf{I}_{n_a+n_c}, \quad \mathbf{h}[t] = \begin{bmatrix} \mathbf{x}[t-1] \\ \mathbf{w}[t-1] \\ 0 \\ \vdots \\ 0 \end{bmatrix}\end{aligned}$$

where $\mathbf{x}[t-1] = [x[t-1] x[t-2] \cdots x[t-n_a]]^T$ and $\mathbf{w}[t-1] = [w[t-1] w[t-2] \cdots w[t-n_c]]^T$. Notice that in estimation and in the full TARMA case, the innovations $w[t]$ may be replaced by their respective a posteriori estimates $\hat{w}[t] = x[t] - \mathbf{h}^T[t] \cdot \hat{\mathbf{z}}[t]$,

where $\hat{\mathbf{z}}[t]$ is the a posteriori state estimate (Niedzwieki 2000, p. 263). After initial estimation with the Kalman filter, refined parameter estimates may be obtained by using the Kalman smoother (Poulimenos and Fassois 2006).

The estimation of an SP-TARMA model is performed by computing θ_1^N as described above, using fixed values of σ_w^2 and σ_v^2 . Since the values of these variances are generally unavailable, a *normalized* form of the Kalman filter may be used, where the prediction/update equations are divided by σ_w^2 . In this way, a single parameter $\lambda = \sigma_v^2/\sigma_w^2$ is left as a design (user selected) parameter that adjusts the “tracking speed” versus “smoothness of the estimates” in the algorithm (Poulimenos and Fassois 2006).

In the GSC-TARMA case, an expectation-maximization scheme can be used as follows (Avenidaño-Valencia and Fassois 2013): (i) *expectation step*: obtain $\theta[t]$ using the Kalman filter as previously described, based on fixed values of $\mathcal{P}$; (ii) *maximization step*: estimate μ and $\sigma_w^2[t]$ based on the estimated values of $\theta[t]$. The E and M steps are sequentially repeated until convergence of the cost function $\mathcal{J}(\theta_1^N, \mathcal{P})$ is achieved. The estimation of σ_v^2 is avoided within the M step, as it destabilizes the algorithm. Thus, as in the SP-TARMA case, σ_v^2 is left as a design parameter to adjust the tracking speed and the parameter smoothness.

For both SP and GSC-TARMA models, the selection of the value of σ_v^2 may not be straightforward, since a very low value may over-smooth the estimated parameter trajectories, while a high value may lead to “noisy” trajectories. The selection of σ_v^2 may be guided by comparing the innovations (prediction residuals) with the parameter innovations (prediction error of the parameters) – that is, the *residual sum of squares* (RSS) to the *parameter prediction error sum of squares* (PESS):

$$RSS = \sum_{t=1}^N \hat{w}^2[t], \quad PESS = \sum_{t=1}^N \hat{v}^T[t] \cdot \hat{v}[t] \quad (9)$$

where $\hat{w}[t] = x[t] - \mathbf{h}^T[t] \cdot \hat{\mathbf{z}}[t|t-1]$ stands for the one-step-ahead prediction error (residual) at time t and $\hat{v}[t] = \hat{\theta}[t|t] - \hat{\theta}[t|t-1]$ is an estimate of the parameter innovations with $\hat{\theta}[t|t-1]$ being the a priori and $\hat{\theta}[t|t]$ the a posteriori Kalman filter estimates of $\theta[t]$. Both RSS and PESS are computed from the Kalman filter predictions of

$w[t]$ and $v[t]$ obtained for a specific value of σ_v^2 . A high RSS indicates poor modeling accuracy, whereas high PESS indicates noisy parameter estimates. A curve displaying the PESS versus the RSS (parameterized in terms of σ_v^2) may be constructed and used for selecting a good compromise (and hence a proper σ_v^2).

Remarks: (i) In the definition of the GSC-TARMA model, it is also possible to include a stochastically time-dependent innovations variance, as in Kitagawa and Gersch (1996). (ii) Additional definitions are possible for the general SPE-TARMA model class. For instance non-Gaussian signals or nonlinear stochastic parameter evolution dynamics may be included, which may be appropriate for some rapidly evolving processes (Kitagawa and Gersch 1996).

Deterministic Parameter Evolution (DPE) TARMA Modeling

DPE-TARMA models are typically defined in terms of the *functional series TARMA* (FS-TARMA) model form, for which the temporal evolution of the parameters is expressed via projections in proper functional subspaces. Thus, for an FS-TARMA $(n_a, n_c)_{[p_a, p_c, p_s]}$ model, with p_a, p_c, p_s designating its AR, MA, and innovations variance functional subspace dimensionalities, the evolution of the parameters is as follows (Poulimenos and Fassois 2006):

$$\begin{aligned} a_i[t] &= \sum_{k=1}^{p_a} a_{i,k} \cdot G_{b_{a(k)}}[t], \quad c_i[t] = \sum_{k=1}^{p_c} c_{i,k} \cdot G_{b_{c(k)}}[t], \\ \sigma_w^2[t] &= \sum_{k=1}^{p_s} s_k \cdot G_{b_{s(k)}}[t] \end{aligned} \quad (10a)$$

$$\begin{aligned} \mathcal{F}_{AR} &= \{G_{b_{a(1)}}[t], \dots, G_{b_{a(p_a)}}[t]\}, \\ \mathcal{F}_{MA} &= \{G_{b_{c(1)}}[t], \dots, G_{b_{c(p_c)}}[t]\}, \\ \mathcal{F}_{\sigma_w^2[t]} &= \{G_{b_{s(1)}}[t], \dots, G_{b_{s(p_s)}}[t]\} \end{aligned} \quad (10b)$$

with “ $\mathcal{F}$ ” designating the functional subspace of the indicated quantity; $b_{a(k)}$ ($k = 1 \dots, p_a$), $b_{c(k)}$

($k = 1 \dots, p_c$), and $b_{s(k)}$ ($k = 1 \dots, p_s$) indices indicating the specific basis functions included in each subspace; and $a_{i,k}$, $c_{i,k}$, and s_k the AR, MA, and innovations variance, respectively, *coefficients of projection*. Thus, for an FS-TARMA model the model parameter vector is $\boldsymbol{\vartheta} = [\boldsymbol{\vartheta}_a^T \boldsymbol{\vartheta}_c^T \boldsymbol{\vartheta}_s^T]^T$, while model structure is specified by the AR and MA orders n_a, n_c and the AR, MA, and innovations variance basis function index vectors $\mathbf{b}_a = [b_{a(1)} \dots b_{a(p_a)}]^T$, $\mathbf{b}_c = [b_{c(1)} \dots b_{c(p_c)}]^T$, and $\mathbf{b}_s = [b_{s(1)} \dots b_{s(p_s)}]^T$, that is,

$$\begin{aligned} \boldsymbol{\vartheta} &= [\boldsymbol{\vartheta}_a^T \boldsymbol{\vartheta}_c^T \boldsymbol{\vartheta}_s^T]^T \\ &= [a_{1,1} \dots a_{n_a,p_c} | c_{1,1} \dots c_{n_c,p_c} | s_1 \dots s_{p_s}]^T \end{aligned} \quad (11a)$$

$$\mathcal{M} = \{n_a, n_c, \mathbf{b}_a, \mathbf{b}_c, \mathbf{b}_s\} \quad (11b)$$

In a recent extension, adaptable FS-TARMA (AFS-TARMA) models that employ functional subspaces parameterized by a parameter vector $\boldsymbol{\delta}$ were introduced (Spiridonakos and Fassois 2014a). The model definition is as in Eq. 10, but the functional bases are parameterized according to the following forms: $\mathcal{F}_{AR} = \{G_{b_{a(1)}}[t, \boldsymbol{\delta}_a], \dots, G_{b_{a(p_a)}}[t, \boldsymbol{\delta}_a]\}$, $\mathcal{F}_{MA} = \{G_{b_{c(1)}}[t, \boldsymbol{\delta}_c], \dots, G_{b_{c(p_c)}}[t, \boldsymbol{\delta}_c]\}$, and $\mathcal{F}_{\sigma_w^2[t]} = \{G_{b_{s(1)}}[t, \boldsymbol{\delta}_s], \dots, G_{b_{s(p_s)}}[t, \boldsymbol{\delta}_s]\}$, where $\boldsymbol{\delta}_a$, $\boldsymbol{\delta}_c$, and $\boldsymbol{\delta}_s$ indicate the AR, MA, and innovations variance functional subspace parameter vector, respectively. The model structure is in this case defined by just the model orders and functional subspace dimensionalities, while the complete parameter vector includes the functional subspace parameters as well:

$$\begin{aligned} \boldsymbol{\theta} &= [\boldsymbol{\vartheta}^T \boldsymbol{\delta}^T]^T, \quad \boldsymbol{\delta} = [\boldsymbol{\delta}_a^T \boldsymbol{\delta}_c^T \boldsymbol{\delta}_s^T]^T, \\ \mathcal{M} &= \{n_a, n_c, p_a, p_c, p_s\} \end{aligned} \quad (12)$$

The advantage of the AFS-TARMA model structure is that the selection of the basis functions, which is a structural problem in conventional FS-TARMA models, becomes part of the parameter estimation problem and is thus

significantly simplified. The adaptable models may thus better “adapt” to a given nonstationary signal, while the modeling procedure is easier for the user.

The estimation of AFS/FS-TARMA models is typically accomplished within a *maximum likelihood* (ML) framework. In the FS-TARMA case, and under Gaussian innovations, the log-likelihood function is (Spiridonakos and Fassois 2014b)

$$\begin{aligned} \ln \mathcal{L}(\boldsymbol{\vartheta} | x_1^N) &= -\frac{N}{2} \cdot \ln 2\pi \\ &\quad - \frac{1}{2} \sum_{t=1}^N \left(\ln \sigma_w^2[t] + \frac{w^2[t]}{\sigma_w^2[t]} \right) \end{aligned} \quad (13)$$

As the likelihood function in Eq. 13 is non-quadratic in terms of the unknown parameter vector, the maximization is based on iterative schemes and rather accurate initial parameter estimates are required for avoiding potential local extrema.

In the simpler case of FS-TAR models, initial parameter values may be obtained by estimating $\boldsymbol{\vartheta}_a$ via *ordinary least squares* (OLS) and subsequently estimating $\boldsymbol{\vartheta}_s$ via an overdetermined set of equations after estimating $\sigma_w^2[t]$ via a sliding window approach (Poulimenos and Fassois 2006). In the FS-TARMA case more elaborate techniques are necessary, as the prediction error is a nonlinear function of the projection coefficient vector $\boldsymbol{\vartheta}_c$. These include *linear multistage* or *recursive* methods (Poulimenos and Fassois 2006). Linear multistage methods first obtain an initial *estimate* of the prediction error sequence based on high-order TAR models and subsequently employ the obtained values to estimate the FS-TARMA model coefficients of projection (the two-stage least-squares (2SLS) approach). Recursive methods use the recursive extended least squares (RELS) or the recursive maximum likelihood (RML) algorithms to obtain initial coefficient of projection estimates (Poulimenos and Fassois 2006). Several runs over the data are typically recommended to avoid the influence of the unknown initial conditions and ensure convergence of the algorithm. For adaptable

(AFS-TARMA) models, an estimation scheme based on separable nonlinear least squares (SNLS), in which the vectors $\boldsymbol{\vartheta}$ and $\boldsymbol{\delta}$ are estimated separately in an sequential fashion, has been suggested (Spiridonakos and Fassois 2014a, b). The reader is referred to (Poulimenos and Fassois 2006; Spiridonakos and Fassois 2014b) for further details on AFS/FS-TARMA models and their estimation.

Remarks: (i) For conventional FS-TARMA models, the functional subspaces include linearly independent basis functions selected from an ordered set, such as Chebyshev, trigonometric, b-splines, wavelets, and other functions. For simplicity a functional subspace is often selected to include consecutive basis functions up to a maximum index. Yet, for purposes of model parsimony (economy) and effective estimation, some functions may not be necessary and may be dropped. (ii) An FS-TARMA $(n_a, n_c)_{[p_a, p_c, p_s]}$ model of the form (10) is referred to as a *fully parametric* FS-TARMA model. The term *semi-parametric* FS-TARMA $(n_a, n_c)_{[p_a, p_c]}$ model implies that the innovations variance is not parameterized, that is, it is not projected on a functional subspace.

Model Structure Selection

Model structure selection is the process by which the structural parameters of the model are obtained. This is typically an iterative and tedious procedure, in which models corresponding to various candidate structures are first estimated, and the one providing the *best fitness* is selected. This procedure may be facilitated via integer optimization schemes or backward/forward regression schemes (Spiridonakos and Fassois 2014b). Model fitness may be judged in terms of a number of criteria, which may include the residual sum of squares (RSS) (often normalized by the series sum of squares, SSS), the likelihood function, and the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). The latter two are typically preferred as they maintain a balance between model fit (model accuracy) and model size (thus discouraging overfitting). The RSS/SSS and BIC criteria are defined as

$$\begin{aligned} \text{RSS/SSS} &= \sum_{t=1}^N \hat{w}^2[t] / \sum_{t=1}^N x^2[t], \\ \text{BIC} &= -\ln \mathcal{L}(\cdot) + \frac{\ln N}{2} \cdot d \end{aligned} \quad (14)$$

where $\hat{w}[t]$ is the obtained one-step-ahead prediction error at time t , $\mathcal{L}(\cdot)$ is the likelihood of the respective model family, and d is the number of estimated parameters in the model (FS-TARMA, $d = \dim \boldsymbol{\vartheta}$; AFS-TARMA, $d = \dim \boldsymbol{\theta}$; SPE-TARMA, $d = \dim z[t] = (n_a + n_c) \cdot q$) (Poulimenos and Fassois 2006; Kitagawa and Gersch 1996, Chapter 2). Notice that in the DPE-TARMA case, the likelihood is defined by Eq. 13, while in the SPE-TARMA case the log-likelihood function (w.r.t. the hyperparameters) is equivalent to $\mathcal{J}(\hat{\boldsymbol{\theta}}_1^N, \hat{\mathcal{P}})$ (Avendaño-Valencia and Fassois 2013).

It should be noted that model estimation always requires attention on part of the user in order to detect numerical problems (for instance, those due to inverting an ill-conditioned matrix) or estimating a number of parameters that is not commensurate with the signal length. As a rough guide, the number of signal samples per estimated parameter (SPP) should be at least 15 (Spiridonakos and Fassois 2014b). Finally formal *model validation* – which examines the validity of the model assumptions (such as innovations whiteness and Gaussianity) – should be performed before final model acceptance (Box et al. 1994, Chapter 8; Poulimenos and Fassois 2006; Spiridonakos and Fassois 2014b).

Structural Health Monitoring (SHM) Based on Nonstationary Random Vibration

Let s_v designate a given structure in one of several potential health states. $v = o$ designates the healthy state, while any other v from the set $V = \{a, b, \dots\}$ designates the structure in a damaged (faulty) state of a distinct *type* $a, b, \dots$ and so forth (for instance, damage in a particular region, or of a particular nature).

In general, each damage type may include a continuum of damages, each being characterized by its own damage *magnitude*.

The SHM problem may be then posed as follows: Given the structure in a currently *unknown* state u , first determine whether or not the structure is damaged ($u = o$ or $u \neq o$) (the *damage detection subproblem*). In case the structure is found to be damaged, determine which one of $\{a, b, \dots\}$ is the current damage type (the *damage identification subproblem*). The *damage magnitude estimation subproblem*, which focuses on estimating the magnitude of the current damage, will not be treated in this entry.

When the main information available for solving this problem is in the form of measured structural vibration response, then the problem is classified as a *vibration-based* SHM problem. If additional information – such as an analytical structural dynamics model – is used for its solution, then the method is classified as *analytical model based*, otherwise as a *data based*. This entry focuses in the latter case where analytical models are not available or are hard to obtain – the reader is referred to Fassois and Sakellariou (2009) and Farrar and Worden (2013) for a broader overview and details. In the context of this entry, the important – but obviously more difficult and scarcely studied – case of *nonstationary* random vibration is considered.

From an operational viewpoint, vibration-based SHM is organized into two distinct phases: First, an initial *baseline phase* in which a set of vibration response signals $x_v[t]$ ($t = 1, \dots, N$), possibly for all $v \in \{o, a, b, \dots\}$, are obtained and properly processed (this phase is carried out only once). Second, an *inspection phase* in which (typically) a single vibration response signal $x_u[t]$ is obtained and decisions on the presence and type of damage need to be made (the aforementioned damage detection and identification subproblems). This phase is typically carried out continuously or periodically, each time using a fresh vibration signal.

In this entry two SHM approaches are presented within a nonstationary random vibration context: a *parameter-based* one and a *residual-based* one. Both are based on the modeling of the nonstationary random vibration signals via the earlier discussed FS-TARMA representations. For purposes of simplicity, the damage identification subproblem is treated via successive binary hypothesis testing (instead of a single multiple hypothesis testing). In this (former) context, once the presence of damage is detected, its type is determined via successive (pairwise) comparisons with each potential damage type. An implicit assumption behind both approaches is that the operating and measurement conditions in the baseline and inspection phases are *identical*. Hence, the various signals correspond to each other in a proper way – in particular in their time duration they describe the exact same motion or operational cycle.

A Parameter-Based Approach

The essence of this approach is on the use of the projection coefficient vector $\boldsymbol{\vartheta}$ of an FS-TARMA model of the nonstationary random vibration as the *characteristic quantity (or feature)* in the decision-making mechanism. The underlying thesis is that each distinct health state is characterized by its own projection coefficient vector, thus “comparing” that of the current state u to that of the healthy state o leads to damage detection. In the positive (damage) case, “comparing” that of the current state u sequentially (pairwise) to that of each damage state $\{a, b, \dots\}$ leads to damage identification.

Hence, in the (initial) baseline phase FS-TARMA models corresponding to the healthy o and each damage state $\{a, b, \dots\}$ are estimated. Then, in the (current) inspection phase, an (identical in structure) FS-TARMA model corresponding to the current structural state u is estimated based on the currently available vibration signal. Then damage detection may be treated via a hypothesis test of the form

$$\begin{aligned} H_0 : \boldsymbol{\vartheta}_o - \boldsymbol{\vartheta}_u &= \mathbf{0} && \text{Null hypothesis – healthy state} \\ H_1 : \boldsymbol{\vartheta}_o - \boldsymbol{\vartheta}_u &\neq \mathbf{0} && \text{Alternative hypothesis – damaged state} \end{aligned}$$

As the true coefficient of projection vectors are not available, decision making is based on corresponding estimates $\hat{\boldsymbol{\vartheta}}_o$ and $\hat{\boldsymbol{\vartheta}}_u$, and since these are random quantities, on their distributions. In this context, under mild assumptions, the estimators are shown to be asymptotically (for “long” data records, i.e., $N \rightarrow \infty$) Gaussian distributed, with mean equal to the true coefficient of projection vector and covariance Σ that

may be estimated (Poulimenos and Fassois 2009a; Spiridonakos and Fassois 2013). Under these conditions, the quantity d_M^2 , below, follows (assuming negligible variability for the covariance estimator) chi-square distribution with d degrees of freedom ($d = \dim \boldsymbol{\vartheta}$). Hence, decision making may be made as follows at the α risk level (i.e., false alarm probability equal to α):

$$d_M^2 = \left(\hat{\boldsymbol{\vartheta}}_u - \hat{\boldsymbol{\vartheta}}_o \right)^T \cdot \hat{\Sigma}_o^{-1} \cdot \left(\hat{\boldsymbol{\vartheta}}_u - \hat{\boldsymbol{\vartheta}}_o \right) \leq \chi_d^2(1 - \alpha) \Rightarrow \text{Accept } H_0 \quad (15)$$

otherwise $\Rightarrow \text{Accept } H_1$

with $\chi_d^2(1 - \alpha)$ designating the $1 - \alpha$ critical point of the chi-square distribution (d degrees of freedom) and $\hat{\Sigma}_o$ the estimator covariance in the healthy case. As already indicated, damage identification may be treated via similar pairwise tests in which the current true coefficient of projection vector $\boldsymbol{\vartheta}_u$ is compared to the true vector $\boldsymbol{\vartheta}_v$, corresponding to each damage state (for $v \in \{a, b, \dots\}$). Of course, ramifications of the method are possible, for instance, by only including specific elements of the coefficient of projection vector, or a properly transformed (for instance, via principal component analysis) version. An obvious disadvantage of the general parameter-based approach is that model estimation needs to be carried out in the inspection phase as well.

A Residual-Based Approach

The essence of this approach is on the use of the FS-TARMA model residual signal $\hat{w}[t]$, and more specifically its time-dependent variance $\sigma_w^2[t]$, as the *characteristic quantity* (or *feature*) in the decision-making mechanism. The underlying thesis is that for each distinct health state (for instance,

the healthy state), the model residual signal (then $\hat{w}_o[t]$) is characterized by its own residual variance (then $\sigma_{w_o}^2[t]$) – this may be obtained in the baseline phase. Then, under the hypothesis that the structure is still in the same health state during inspection, the new residual signal ($\hat{w}_u[t]$) obtained by driving the *current* (fresh) vibration signal ($x_u[t]$) through the *same baseline model* (no model reestimation involved) should be characterized by the same time-dependent variance (then $\sigma_{w_o}^2[t]$) *if and only if* the hypothesis of the structure being in the same health state (for instance, healthy state) is correct (as a change would result in increased variance). Then, a decision on the health state of the structure may be made based on comparing the current time-dependent residual variance ($\sigma_{w_u}^2[t]$) to the baseline time-dependent variance (then $\sigma_{w_o}^2[t]$) at each time instant. Of course, this procedure may be repeated for any other health state of the structure in the baseline phase for binary *damage identification*.

Then damage detection may be treated via a hypothesis test of the form (Poulimenos and Fassois 2004):

$$\begin{array}{ll} H_0 : \sigma_{w_u}^2[t] \leq \sigma_{w_o}^2[t] & \text{Null hypothesis – healthy state} \\ H_1 : \sigma_{w_u}^2[t] > \sigma_{w_o}^2[t] & \text{Alternative hypothesis – damaged state} \end{array}$$

As the theoretical variances $\sigma_{w_u}^2[t]$ and $\sigma_{w_o}^2[t]$ are unavailable, they need to be estimated from the

obtained residual series $\hat{w}_u[t]$ and $\hat{w}_o[t]$ using a moving average filter (sliding window) as follows:

$$\begin{aligned}\hat{\sigma}_{w_u}^2[t] &= \frac{1}{\ell} \sum_{\tau=t-(\ell-1)/2}^{t+(\ell-1)/2} \hat{w}_u^2[\tau], \\ \hat{\sigma}_{w_o}^2[t] &= \frac{1}{\ell_o} \sum_{\tau=t-(\ell_o-1)/2}^{t+(\ell_o-1)/2} \hat{w}_o^2[\tau]\end{aligned}\quad (16)$$

with ℓ , ℓ_o designating the corresponding window lengths. Under the null hypothesis (H_0), given the residual normality and uncorrelatedness, the statistic defined as the ratio of the two variance estimators follows an F distribution with $(\ell - 1, \ell_o - 1)$ degrees of freedom, that is,

$$F[t] = \frac{\hat{\sigma}_{w_u}^2[t]}{\hat{\sigma}_{w_o}^2[t]} \sim F_{\ell-1, \ell_o-1} \quad (17)$$

This leads to the following sequential F-test (at the α risk level, i.e., false alarm probability equal to α):

$$\begin{aligned}F[t] \leq F_{1-\alpha} &\Rightarrow \text{Accept } H_0 \\ \text{otherwise} &\Rightarrow \text{Accept } H_1\end{aligned}\quad (18)$$

with $F_{1-\alpha} = F_{\ell-1, \ell_o-1}(1 - \alpha)$ indicating the distribution's $(1 - \alpha)$ critical point. An obvious advantage of this approach is that no model reestimation is required in the inspection phase.

Illustrative Example: Nonstationary Random Vibration Modeling and SHM for a Pick-and-Place Mechanism

The Structure and its Nonstationary Random Vibration Response

The system studied in this example is the 2-DOF pick-and-place mechanism mentioned earlier (Fig. 1a). The random vibration response is measured in the same direction as the excitation, using lightweight piezoelectric accelerometers. During a single experiment a single cycle is performed, in which the linear motors move from their rightmost to their leftmost position and back. The measured vibration response is conditioned and driven into a data acquisition

module, which digitizes the signal with a sampling frequency $f_s = 512$ Hz – signal length 10 s ($N = 5,120$ samples). Each signal is subsequently sample mean corrected and normalized (scaled). The frequency range of interest is 5–200 Hz, with the lower limit set in order to avoid instrument dynamics and rigid body modes.

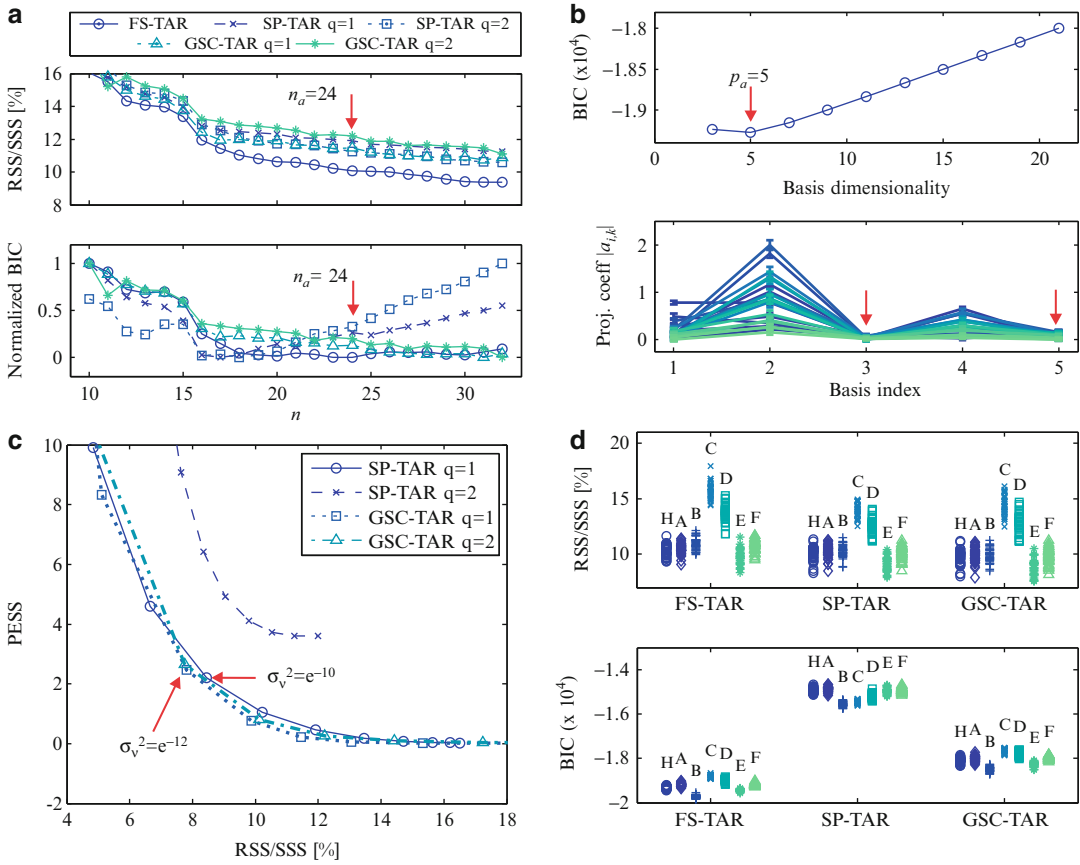
For the SHM problem six damage scenarios are considered, which correspond to the loosening or removal of various bolts at different points of the mechanism (damages A to C and E), loosening the slider of motor B (damage D), and adding a mass at the free end of the slider of motor A (damage F). For each health state, a set of 40 random vibration responses are recorded – see (Spiridonakos and Fassois 2013) for details.

Nonstationary Random Vibration Parametric Modeling (Healthy Structure)

The nonstationary random vibration response (see Fig. 1b) of the healthy structure is now modeled via the stochastic parameter evolution (both SP-TAR and GSC-TAR models) and the deterministic parameter evolution (FS-TAR models) methods using a single data record obtained from the healthy structure. The details for each method are summarized as follows:

Stochastic parameter evolution modeling: SP-TAR models are estimated via the Kalman filter – smoother method (Poulimenos and Fassois 2006) – and GSC-TAR models via the expectation-maximization method (Avendaño-Valencia and Fassois 2013). The innovations variance is in each case estimated via a moving rectangular (600 sample long) window. The following structural parameters are considered: AR order $n_a \in \{3, \dots, 32\}$, with $\{q = 1, \sigma_v^2 = \exp(-11)\}$ and $\{q = 2, \sigma_v^2 = \exp(-24)\}$. The value of σ_v^2 is further optimized by estimating SP/GSC-TAR models with the selected model order and $\sigma_v^2 = \exp(v)$, $v = \{-10, -11, \dots, -30\}$, $q = \{1, 2\}$. The optimal σ_v^2 is determined by considering the RSS and PESS.

Deterministic parameter evolution modeling: Fully parameterized FS-TAR models with trigonometric basis functions of the form



Parametric Nonstationary Random Vibration Modeling with SHM Applications, Fig. 2 Model structure selection for SP-TAR, GSC-TAR, and FS-TAR models: (a) model order selection – RSS/SSS and BIC (the latter normalized between 0 and 1) curves for SP/GSC/FS-TAR models with orders $n \in \{10, \dots, 32\}$. (b) Selection of the subspace dimensionality for FS-TAR models: *top*, BIC of FS-TAR(24) models with $p_a = p_s \in \{3, 5, \dots, 21\}$;

bottom, estimated projection coefficients for each basis dimensionality (continuous lines connect the point estimates and bars indicate $\pm$ one standard deviation). (c) Optimization of σ_v^2 for SP/GSC-TAR models – PESS versus RSS/SSS for SP-TAR(24) and GSC-TAR(24) models. (d) Sample distribution of RSS/SSS and BIC for the selected model structure reestimated for each healthy (H) and damage (A ... F) data record (40 models per health state)

$$\begin{aligned} G_0[t] &= 1, \quad G_{2k-1}[t] \\ &= \sin(2\pi kt/N), \quad G_{2k}[t] \\ &= \cos(2\pi kt/N) \quad k = 1, 2, \dots \end{aligned} \quad (19)$$

are considered. The functional subspace dimensionality p_a corresponds to the number of sine and cosine functions used by the FS-TAR model plus one (the constant component), $t = 1, 2, \dots, N$ is the normalized discrete time, and N the signal length. Parameter estimation is based on ordinary least squares, while the innovations variance is

estimated via the instantaneous method (Poulimenos and Fassois 2006). The following structural parameters are considered: AR order $n_a \in \{3, \dots, 32\}$, $p_a = p_s \in \{3, 5, \dots, 21\}$.

Modeling results: The results of the model structure selection procedure are depicted in Fig. 2a–c). Figure 2a shows the RSS/SSS and normalized BIC for models from each considered class (FS/SP/GSC-TAR), with the selected order ($n_a = 24$) being indicated by an arrow. Figure 2b shows the functional subspace dimensionality

selection for FS models. The top plot shows the BIC of FS-TAR(24) models versus functional basis dimensionality; the selected dimensionality ($p_a = 5$) is indicated by an arrow. The bottom plot shows the absolute value of the estimated coefficients of projection with their corresponding ± 1 standard deviation interval indicated by the bars. It is evident that the standard deviation interval of the estimated coefficients of projection for the basis indices 2 and 4, indicated by the arrows in the plot, consistently contains zero and may be thus removed from the model. An FS-TAR(24)_[3,5] model with functional basis indices $b_a = [0,1,3]$ and $b_s = [0,1,2,3,4]$ is finally selected (i.e., including the functions $G_0[t], G_1[t], G_3[t]$ in the AR subspace and the functions $G_0[t], G_1[t], G_3[t], G_4[t]$ in the innovations variance subspace).

Figure 2c shows the PESS versus RSS/SSS plot for the selection of σ_v^2 for SP-TAR and GSC-TAR models. An arrow indicates the model selected, which corresponds to a “good” compromise between low PESS and RSS/SSS. The selected model structures are SP-TAR(24) with $q = 1$, $\sigma_v^2 = e^{-10}$, and GSC-TAR(24) model with $q = 1$, $\sigma_v^2 = e^{-12}$.

Models of the *selected* (under the healthy condition) structure are subsequently fitted (estimated) for each data record corresponding to the healthy and each damaged state of the structure (40 models per health state, each one based on a distinct data record). In Fig. 2d the sample distribution of RSS/SSS and BIC of the above selected models are presented for the various health states. The FS-TAR model structure uniformly (for all health states and data records) achieves the lowest BIC; although its RSS/SSS values are not minimal.

The “frozen”-type TV-PSDs of the healthy structure, as obtained by the aforementioned three model types and a single data record, are presented in Fig. 3. For purposes of comparison, the nonparametric spectrogram (Gaussian window $\sigma = 8$, $N_{fft} = 1,024$ samples, 51 samples advance ($\sim 5\%$ of N_{fft})) estimate is also shown. While all TV-PSDs are in rough overall agreement, it is obvious that the parametric model-

based ones are much cleaner and informative than their nonparametric (spectrogram) counterpart. This is an important feature of parametric methods.

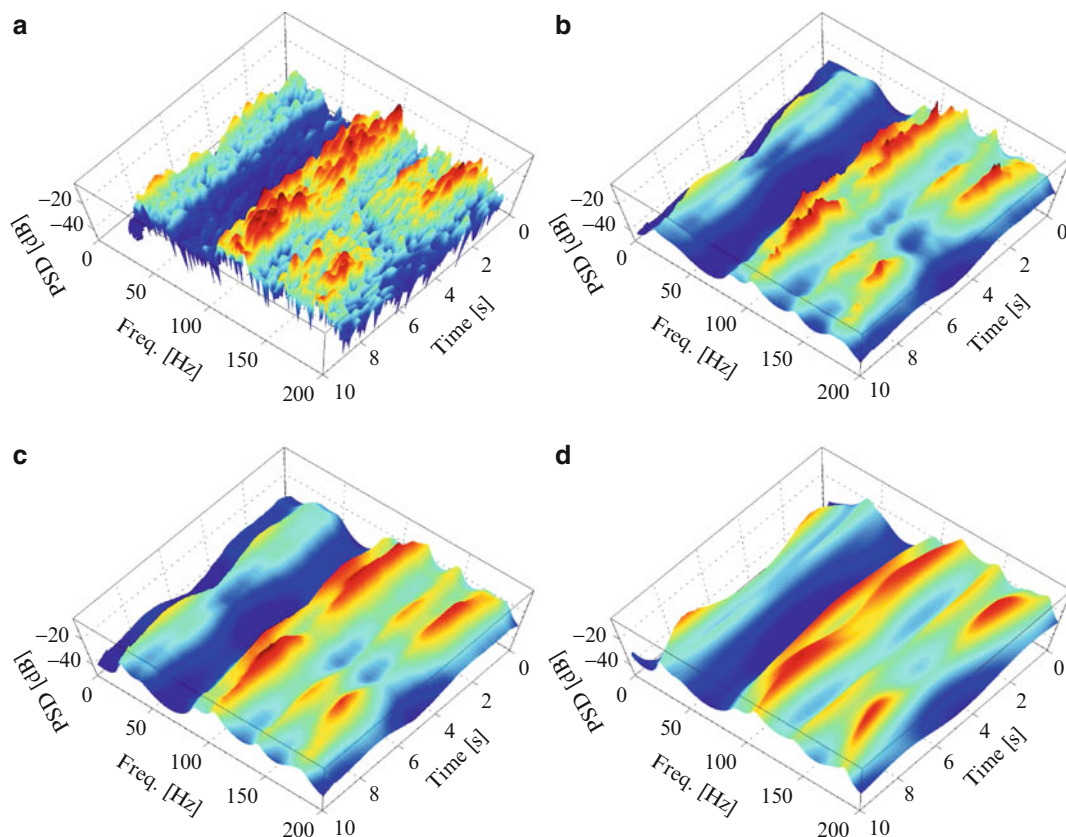
Nonstationary Vibration Response-Based SHM

As already mentioned the healthy (H) and six damage scenarios (A to F) are considered in SHM, with 40 data records used in each health state.

Two versions of the FS-TAR model parameter-based approach are used: (a) the original version in which the coefficient of projection covariance matrix Σ_o is estimated based on a single data record and (b) an alternative version in which the covariance matrix is estimated based on several (presently 35) data records (using the sample covariance estimator).

Damage detection results using version (a) are depicted in Fig. 4a. All 40 cases corresponding to the healthy structure provide d_M^2 values lying below the selected detection threshold (dashed horizontal line: $\alpha = 10^{-14}$), thus correctly detecting the current healthy state. Also, in all cases corresponding to damages A...D, the obtained d_M^2 values are above the detection threshold, thus correctly detecting damage. Only certain cases corresponding to damages E and F are not properly detected. As indicated by the ROC curves (receiver operating characteristic curves which depict the true positives, TPs, versus false positives, FPs, as the threshold varies) of Fig. 4b, the two versions of the parameter-based approach perform quite adequately and similarly (the performance is almost ideal if damages E and F are excluded).

Next, the FS-TAR residual-based approach is employed using variance estimates obtained via a sliding window of length $\ell = \ell_o = 200$. Figure 4c provides a comparison of the obtained $F[t]$ statistic for a healthy and a damaged (damage D) case. $F[t]$ is, at all times, under the threshold F_u ($\alpha = 10^{-4}$) indicating healthy structure, or above it (for at least some times) indicating damaged structure. As indicated by the ROC curves of Fig. 4d, the residual-based approach



Parametric Nonstationary Random Vibration Modeling with SHM Applications, Fig. 3 Spectrogram and “frozen”-type TV-PSD estimates obtained from estimated

TAR models using a single vibration response of the healthy structure: (a) spectrogram; (b) SP-TAR(24) ($q = 1$); (c) GSC-TAR(24) ($q = 1$); (d) FS-TAR(24)_[3,5]

P

provides, in this application, inferior performance, which is somewhat improved when only two types of damage, namely, C and D, are considered.

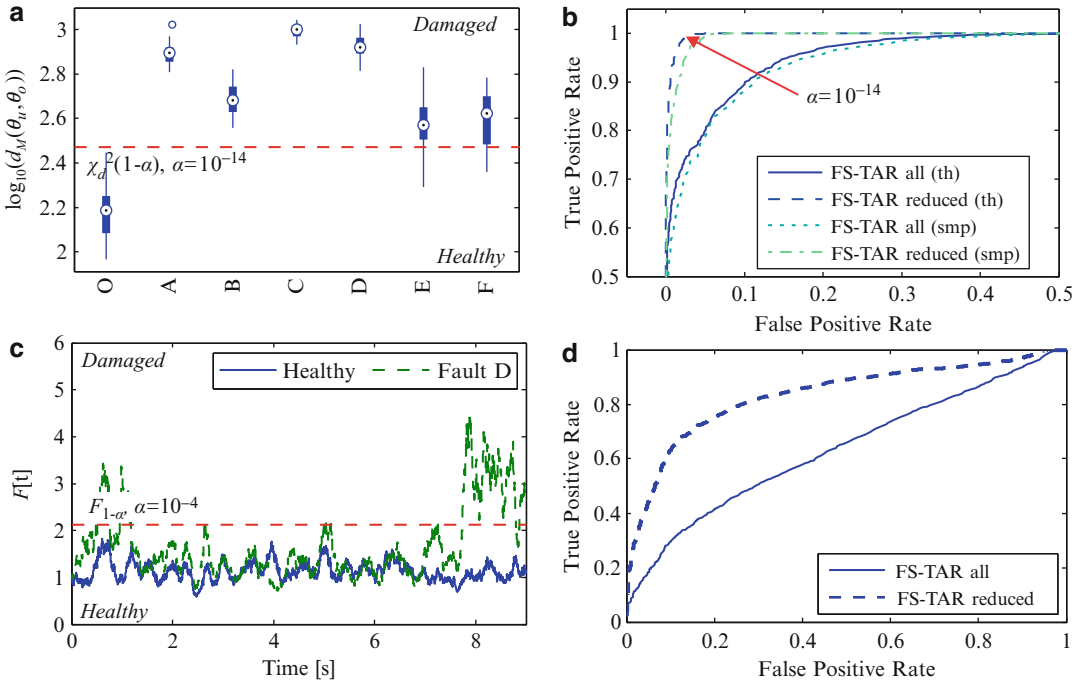
Summary

Models and methods for nonstationary random vibration parametric modeling have been presented, with focus on the stochastic parameter evolution (SP-TAR and GSC-TAR models) and deterministic parameter evolution (FS-TAR models) methods. Approaches for nonstationary random vibration-based Structural Health Monitoring (SHM) employing

these models have been also discussed. An illustrative application of the methods to the vibration response modeling and SHM for a laboratory pick-and-place mechanism has been also presented.

Some key points are summarized below:

- Parametric SPE and DPE modelings of nonstationary random vibration are more involved than their nonparametric counterparts but offer unique opportunities for more accurate and compact representations, improved time-frequency resolution, analysis, and SHM. Other areas such as



Parametric Nonstationary Random Vibration Modeling with SHM Applications, Fig. 4 Summary of damage detection results: (a) boxplots of d_M^2 for the parameter-based approach (original version; 40 experiments per health state) – a damage is detected if d_M^2 exceeds the threshold (dashed horizontal line; $\alpha = 10^{-14}$); (b) ROC curves for the original (“th”) and alternative (“smp”) versions of the approach when all damage types are

included (“all”) and when damage types E and F are excluded (“reduced”); (c) performance of the residual-based approach – $F[t]$ versus the threshold ($\ell = \ell_o = 200$, $\alpha = 10^{-4}$) for a single healthy and a damaged state; (d) ROC curves for the residual-based approach when all damage types are included (“all”) and when only damage types C and D are included (“reduced”)

prediction and control may significantly benefit as well.

- SPE modeling is better suited to “slow” or “medium” variations in the nonstationary dynamics, whereas DPE modeling is also suited (with proper functional subspace selection) for “fast” variations.
- The problem of model structure selection is important for both SPE and DPE modelings. Yet, in the latter case, it may be significantly alleviated via the new class of adaptable models (Spiridonakos and Fassois 2014a).
- The problem of local extrema in the (non-convex) estimation criterion is important for both SPE and DPE modeling. Good initial values and careful use of model validation and diagnostic tools are thus necessary (Poulimenos and Fassois 2009b).

- SHM in nonstationary random vibration environments is a rather recent but important area with many potential applications. The presented approaches should be viewed as initial attempts to address the problem.

Cross-References

- [Model Class Selection for Prediction Error Estimation](#)
- [Operational Modal Analysis in Civil Engineering: an Overview](#)
- [Stochastic Structural Identification from Vibrational and Environmental Data](#)
- [System and Damage Identification of Civil Structures](#)

References

- Avendaño-Valencia L, Fassois S (2013) Generalized stochastic constraint TARMA models for in-operation identification of wind turbine non-stationary dynamics. In: Basu B (ed) Key engineering materials (volumes 569–570) – damage assessment of structures X. Trans Tech Publications, Switzerland, pp 587–594
- Bendat J, Piersol A (2000) Random data analysis and measurement procedures. Wiley, New York
- Box G, Jenkins G, Reinsel G (1994) Time series analysis, forecasting and control. Prentice-Hall, Englewood Cliffs, Third edition
- Farrar C, Worden K (2013) Structural health monitoring, a machine learning perspective. Wiley, Chichester
- Fassois S, Sakellariou J (2009) Statistical time series methods for structural health monitoring. In: Encyclopedia of structural health monitoring. Wiley, Chichester, pp 443–472
- Feng Z, Liang M, Chu F (2013) Recent advances in time-frequency analysis methods for machinery fault diagnosis: a review with application examples. *Mech Syst Signal Pr* 38(1):165–205
- Fouskitakis G, Fassois S (2002) Functional series TARMA modeling and simulation of earthquake ground motion. *Earthq Eng Struct Dyn* 31:399–420
- Gersch W, Akaike H (1988) Smoothness priors in time series. In: Spall J (ed) Bayesian analysis of time series and dynamic models. Marcel Dekker, New York, pp 431–476
- Hammond J, White P (1996) The analysis of non-stationary signals using time-frequency methods. *J Sound Vib* 190:419–447
- Kitagawa G, Gersch W (1996) Smoothness priors analysis of time series. Springer, New York
- Kozin F (1977) Estimation and modeling of non-stationary time series. In: Proceedings of the symposium on computational mechanics in engineering, California
- Kozin F (1988) Autoregressive moving average models of earthquake records. *Probabilist Eng Mech* 3(2):58–63
- Ljung L (1999) System identification: theory for the user. Prentice Hall PTR, Upper Saddle River
- Newland D (1993) An introduction to random vibrations, spectral and wavelet analysis. Dover, New York
- Niedzwieki M (2000) Identification of time-varying processes. Wiley, England
- Owen J, Eccles B, Choo B, Woodings M (2001) The application of auto-regressive time series modelling for the time-frequency analysis of civil engineering structures. *Eng Struct* 23:521–536
- Poulimenos AG, Fassois SD (2004) Vibration-based on-line fault detection in non-stationary structural systems via statistical model based method. In: Proceedings of the 2nd European workshop on structural health monitoring, Munich
- Poulimenos AG, Fassois SD (2006) Parametric time-domain methods for non-stationary random vibration modeling and analysis: a critical survey and comparison. *Mech Syst Signal Pr* 20(4):763–816
- Poulimenos AG, Fassois SD (2009a) Asymptotic analysis of non-stationary functional series TARMA estimators. In: Proceedings of the 15th symposium on system identification, Saint-Malo
- Poulimenos AG, Fassois SD (2009b) Output-only stochastic identification of a time-varying structure via functional series TARMA models. *Mech Syst Signal Pr* 23(4):1180–1204
- Preumont A (1994) Random vibration and spectral analysis. Kluwer, Dordrecht
- Priestley M (1988) Non-linear and non-stationary time series analysis. Academic Press, London
- Rao T (1970) The fitting of non-stationary time-series models with time-dependent parameters. *J R Stat Soc B Met* 32(2):312–322
- Roberts J, Spanos P (1990) Random vibration and statistical linearization. Wiley, Chichester
- Spiridonakos M, Fassois S (2013) An FS-TAR based method for vibration-response-based fault diagnosis in stochastic time-varying structures: experimental application to a pick-and-place mechanism. *Mech Syst Signal Pr* 38:206–222
- Spiridonakos M, Fassois S (2014a) Adaptable functional series TARMA models for non-stationary signal representation and their application to mechanical random vibration modeling. *Signal Process* 96:63–79
- Spiridonakos M, Fassois S (2014b) Non-stationary random vibration modelling and analysis via functional series time dependent ARMA (FS-TARMA) models – a critical survey. *Mech Syst Signal Pr* 47(1–3):175–224

Passive Control Techniques for Retrofitting of Existing Structures

Alessandro Palmeri

School of Civil and Building Engineering,
Loughborough University, Loughborough,
Leicestershire, UK

Introduction

There are many reasons why existing structures may have insufficient capacity to resist the expected seismic events, or their performance may be considered as unsatisfactory by designers and stakeholders. Examples include (but are not limited to):

- Updated hazard assessment for the site, based on new seismological data and/or improved analyses, that increases the intensity of the design earthquake for the various limit states;
- Change of use of an existing structure that increases the overall risk in case of an earthquake (e.g., an office building which now headquarters strategic departments of civil protection or an industrial building which now contains hazardous substances);
- Design and construction of the structure carried out with obsolete methods (e.g., with insufficient ductility of members and joints).

In such cases, two options are available, namely, building a new structure once the existing one has been knocked down or retrofitting the existing structure, in such a way that enhanced/updated performance criteria are satisfied. In many cases, the second option is more advantageous, reducing direct and indirect costs.

Traditional retrofitting of building structures can be achieved by locally increasing stiffness and strength of structural members (e.g., Wu et al. 2006; Hueste and Bai 2007; Ozcan et al. 2008). This can affect both the demand (as the modal frequencies would increase and may have either a beneficial or a detrimental effect) and the capacity (as the overall resistance will be higher). However, traditional retrofitting (e.g., adding steel or FRP jackets and plates to increase the axial and bending performance of RC columns and beams) may be very expensive due to the large number of local interventions and may also prevent the use of the structure for many months (with the associate indirect costs for the loss of use).

A different retrofitting approach consists of equipping the existing structure with passive control devices, i.e., dampers and isolators (e.g., Soong and Spencer 2002; Dargush and Sant 2005; Di Sarno and Elnashai 2005). Such devices are said to be “passive” because they do not require any power supply to work (unlike “active” and “semi-active” devices). Dampers,

commonly placed in the superstructure of buildings, allow reducing the seismic demand by increasing the overall energy dissipation capability, although they may also increase the stiffness of the structure. Viscous, viscoelastic and elastoplastic mechanisms of energy dissipation are typically exploited in this type of devices. Conversely, isolators placed between the superstructure and the foundation of a building improve the seismic performance by increasing the fundamental period of vibration of the overall structural system, in such a way that most of the energy of the earthquake is filtered as the seismic motion of the superstructure is decoupled (i.e., isolated) from the ground. High-damping rubber bearings (HDBRs) and friction pendulum bearings (FPBs) are among the most popular types of isolators.

In the following, the effectiveness of different passive control techniques for earthquake protection will be demonstrated by considering the seismic response of SDoF (single-degree-of-freedom) oscillators, with and without dampers, namely:

- Fluid viscous dampers (FVD);
- Elastomeric viscoelastic dampers (EVDs);
- Steel hysteretic dampers (SHDs).

Other control techniques can be adopted in the professional practice, which exploit different dynamic phenomena. This is the case of tuned mass dampers (TMDs), including sloshing liquid dampers (SLDs), in which a secondary mass is tuned to the mode of vibration of the existing structure, allowing part of the seismic energy to be transferred to the TMD, where this energy can be dissipated (e.g., Hoang et al. 2008; Lin et al. 2011). Another possibility is to exploit rocking mechanisms, in which part of the seismic energy is transformed into potential energy, as the center of mass of the structure raises when the structure uplifts, and then this energy is dissipated through impacts (e.g., Marriott et al. 2008; Palmeri and Makris 2008). These alternative control techniques are not discussed in the following, as they are not as effective as dampers and isolators in the seismic protection of

building structures (e.g., TMDs are very effective against wild loads, while rocking is particular effective for bridge structures).

Similarly, seismic isolators are not treated in the following, as this technique is more appealing and economically advantageous for new structures rather than for the seismic retrofitting of existing structures. It must be said, however, that in some situations the installation of dampers within an existing building can be very difficult, e.g., because of the large forces that such devices need to transfer to the old structural members, and in this case base isolation system may become the only viable solution.

Governing Equations

The dynamic equilibrium at a generic time t for a SDoF oscillator subjected to seismic input can be written as:

$$m\ddot{u}(t) + f_S(t) + f_D(t) = -m\ddot{u}_g(t), \quad (1)$$

where m is the mass of the oscillator; $u(t)$ is the time history of the relative displacement of the mass with respect to the ground; $\ddot{u}_g(t)$ is the absolute acceleration of the ground; $f_S(t)$ and $f_D(t)$ are the time-varying reaction forces due to stiffness and damping mechanisms. In the above expression, the first term $m\ddot{u}(t) = f_I(t)$ can be interpreted by a ground observer as the inertial force experienced by the mass, while $-m\ddot{u}_g(t) = F(t)$ is the dynamic force induced by the ground shaking. The total inertial force, within a Galilean reference frame, is proportional to the absolute acceleration of the mass m , that is: $f_I^{(G)}(t) = m(\ddot{u}(t) + \ddot{u}_g(t)) = f_I(t) - F(t)$. For a linear system, $f_S(t) = k u(t)$ and $f_D(t) = c\dot{u}(t)$ are proportional to the relative displacement and velocity of the mass through the elastic stiffness k and the viscous damping coefficient c .

Multiplying both sides of Eq. 1 by the infinitesimal relative displacement $du(t)$, and integrating from 0 to the generic time instant t , one obtains:

$$\begin{aligned} \int_0^t m\ddot{u}(\tau)du(\tau) + \int_0^t f_S(\tau)du(\tau) \\ + \int_0^t f_D(\tau)du(\tau) = \int_0^t F(\tau)du(\tau), \end{aligned} \quad (2)$$

which gives the energy balance of the dynamic system over the time interval $[0, t]$ (Uang and Bertero 1990).

Assuming that the oscillator is initially at rest (i.e., $u(0) = 0$ and $\dot{u}(0) = 0$), it can be easily shown that the first integral in the left-hand side is the kinetic energy of the system at the generic time t :

$$T(t) = \frac{1}{2}m\dot{u}(t)^2 = \int_0^t m\ddot{u}(\tau)du(\tau), \quad (3)$$

while the integral in the right-hand side gives the cumulative work done by the seismic input:

$$W_{in}(t) = \int_0^t F(\tau)du(\tau). \quad (4)$$

Additionally, the integral of the linear-elastic stiffness term gives the potential energy of the oscillator:

$$V(t) = \frac{1}{2}ku(t)^2 = \int_0^t ku(\tau)du(\tau), \quad (5)$$

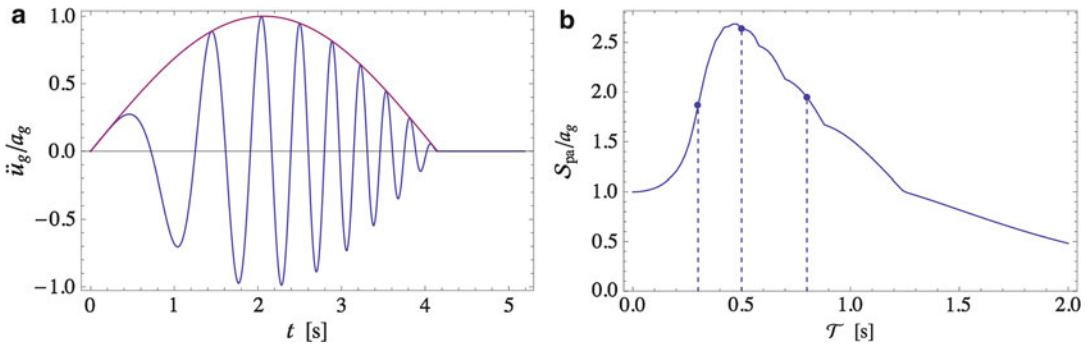
and the linear-viscous damping term provides the energy dissipation:

$$W_D(t) = \int_0^t f_D(\tau)du(\tau) = c \int_0^t \dot{u}(\tau)^2 d\tau. \quad (6)$$

Taking into account Eqs. 3, 4, 5, and 6, it follows that for a linear SDoF oscillator the energy balance at a generic time t can be expressed as (Uang and Bertero 1990):

$$E(t) = T(t) + V(t) = W_{in}(t) - W_D(t), \quad (7)$$

where $E(t)$ is the total mechanical energy stored within the system at time t . Since the potential for damaging the structure increases with this quantity, passive control techniques aim to reduce $E(t)$ either by reducing the input energy $W_{in}(t)$ (e.g.,



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 1 Seismic input. Time history of ground acceleration (a) and response spectrum in terms of pseudo-acceleration (b)

changing the stiffness) or increasing the energy dissipation capacities of the structure (e.g., using dampers to increase $W_D(t)$).

For illustration purposes, specific examples will be provided in the next three sections for different types of dampers. The method used for comparing the different systems will be briefly discussed in the next subsection.

Seismic Excitation and Performance Analysis

The seismic performance of any control technique is greatly influenced by the characteristics of the ground motion, e.g., the peak ground acceleration (PGA), the duration of the event, the variation with time of both amplitude and frequency content, as well as by the characteristics of the existing structure, e.g., modal frequencies, available ductility, overstrength, etc. In order to make a fully comprehensive comparison between different techniques, all these aspects should be considered and the results carefully analyzed. This is far beyond the scope and the limits of the present contribution, and for this reason, a very simple signal has been used throughout as seismic input for our numerical investigations, namely, a harmonic function in which both amplitude and frequency vary with time, i.e., an amplitude-modulated cosine-sweep function. The mathematical expression of the envelope of the signal is given by

$$e_g(t) = \begin{cases} a_g \sin\left(\frac{\pi t}{t_g}\right), & \text{if } 0 \leq t \leq t_g; \\ 0, & \text{otherwise;} \end{cases} \quad (8)$$

and the ground acceleration is

$$\ddot{u}_g(t) = e_g(t) \cos\left(\frac{\omega_g^2 t^2}{2N_g\pi}\right), \quad (9)$$

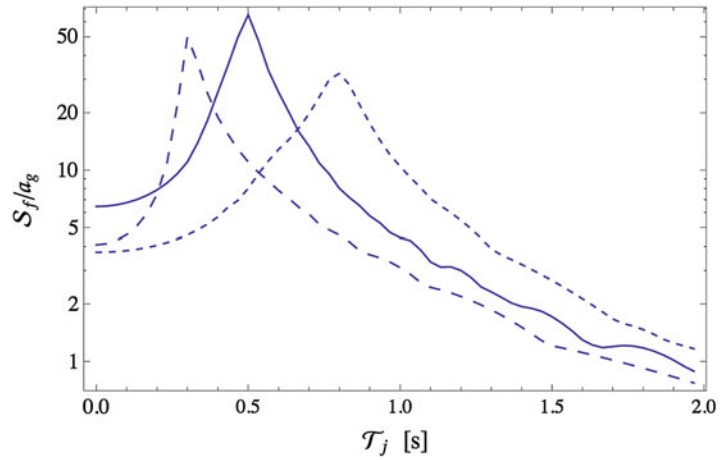
where a_g is the PGA and $t_g = 2N_g\pi/\omega_g$ is the duration of the signal, which in turn depends on the mean circular frequency ω_g and number of cycles N_g of the ground motion. As an example, for $\omega_g = 12.5$ rad/s and $N_g = 8.25$, the duration of the signal is $t_g = 4.15$ s. The accelerogram obtained with this set of parameters is shown within Fig. 1a, in which the red half sine gives the envelope.

This signal has then been used in the following to investigate the effects of different protection systems, and Fig. 1b shows the corresponding elastic response spectrum in terms of pseudo-acceleration, $S_{pa}(T_i)$, which is mathematically defined as:

$$S_{pa}(T_i) = \left(\frac{2\pi}{T_i}\right)^2 \max_{0 \leq t \leq t_g} \{|u(t)|\}, \quad (10)$$

and represents a measure of the maximum force experienced by the SDoF oscillator of natural period T_i during the seismic event. The three vertical dashed lines denote the three natural periods $T_i = 0.3, 0.5$, and 0.8 s considered in the analysis, which corresponds to frequency ratios $\beta_i = \omega_g/\omega_i$ equal to $0.60, 0.99$, and 1.59 , i.e., “stiff,” “resonant,” and “flexible” test structures.

Passive Control Techniques for Retrofitting of Existing Structures, Fig. 2 Floor spectra. Stiff (dashed line), Resonant (solid line) and Flexible (dotted line) existing (primary) structure



The value assumed for the equivalent viscous damping ratio is $\zeta_0 = c/c_{\text{crit}} = 0.02$, being $c_{\text{crit}} = 2\sqrt{km}$ the viscous damping coefficient for a critically damped system. Even though quite small, this value of ζ_0 is reasonable for an undamaged structure within the linear-elastic range.

The semilogarithmic graph of Fig. 2 compares the so-called floor spectra for the three SDoF oscillators with the selected natural periods T_i . Each floor spectrum is obtained by using the absolute acceleration of the i th (primary) oscillator, $\ddot{u}^{(\text{abs})}(t) = \ddot{u}(t) + \ddot{u}_g(t)$, as seismic input for a (secondary) oscillator with natural period T_j and viscous damping ratio ζ_0 . The floor spectrum $S_f(T_i, T_j)$ then gives the maximum pseudo-acceleration for a light attachment (of natural period $T_j = 2\pi/\omega_j$) to the primary oscillator (of natural period $T_i = 2\pi/\omega_i$). The governing equations are:

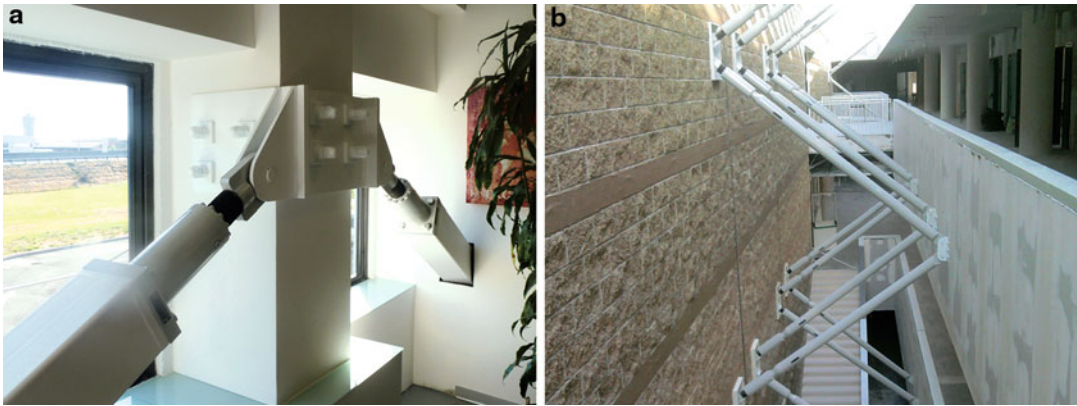
$$\begin{cases} \ddot{u}(t) + 2\zeta_0 \omega_i \dot{u}(t) + \omega_i^2 u(t) = -\ddot{u}_g(t); \\ \ddot{y}(t) + 2\zeta_0 \omega_j \dot{y}(t) + \omega_j^2 y(t) = -(\ddot{u}(t) + \ddot{u}_g(t)), \end{cases} \quad (11)$$

where $y(t)$ is the relative displacement of the secondary mass with respect to the primary one, and the assumption is made that the secondary mass is light enough for its feedback force on the primary mass to be negligible. The reason why the floor spectra are considered as part of this

study is because they allow quantifying the maximum force that a seismic event induces in secondary substructures attached to the primary structural system and then can give an indication about the effectiveness of different retrofitting strategies in terms of building content, including building services (e.g., Sackman and Kelly 1979; Muscolino and Palmeri 2007). As expected, (i) each floor spectrum of Fig. 2 takes the maximum value for $T_j = T_i$, i.e., when primary and secondary oscillators are tuned, and (ii) the maximum amplification is observed for $T_i = 0.5$ s (solid line), i.e., for the resonant primary oscillator.

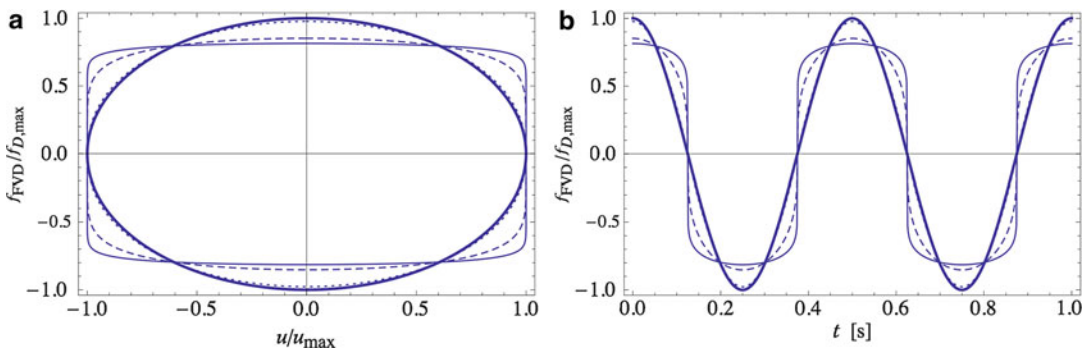
Fluid Viscous Dampers (FVDs)

Viscous fluids have the potential to dissipate large amount of energy, as friction forces arise when they are forced to flow through orifices (e.g., Lee and Taylor 2001; Martinez-Rodrigo and Romero 2003; Molina et al. 2004; Marano et al. 2013). Typically, a fluid viscous damper (FVD) consists of a stainless steel piston traveling through two adjacent chambers filled with silicone oil (often chosen because it is fire resistant, nontoxic, and stable for long periods of time). The movement of the piston causes the silicone oil to flow through an orifice in the piston head, transforming part of the mechanical energy into heat, which is then dissipated into the atmosphere.



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 3 Fluid viscous dampers. Seismic retrofit of buildings in Imola (Courtesy of FIP

Industriale, Franco Baroni and Stefano Silvestri) (a) and in L'Aquila (Courtesy of FIP Industriale and Prof Vincenzo Gattulli) (b)



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 4 Fluid viscous dampers. Hysteresis loop (a) and time history (b) of the reaction

force under harmonic cycles for different values of the velocity exponent: $\alpha = 1.00$, solid thick line; $\alpha = 0.90$, dotted line; $\alpha = 0.30$, dashed line; $\alpha = 0.10$, solid thin line

Figure 3 shows two recent applications of FVDs for the seismic retrofitting of existing structures in Italy, namely, a prefabricated industrial building in Imola (Fig. 3a), necessary after the 2012 Northern Italy earthquakes, and building “A” of the Faculty of Engineering of the University of L'Aquila (Fig. 3b), required after the 2009 L'Aquila earthquake. In both cases, the presence of the control devices increases the damping capacity of the structure and improves their seismic behavior.

The reaction force $f_{\text{FVD}}(t)$ exerted by FVDs depends on the relative velocity $\dot{u}(t)$ between the opposite ends of the device, and their nonlinear force–velocity relationship is generally expressed as (e.g., Lin and Chopra 2002)

$$f_{\text{FVD}}(t) = c|\dot{u}(t)|^\alpha \text{sgn}[\dot{u}(t)], \quad (12)$$

where c is the damping coefficient (measured in $\text{kN s}^\alpha/\text{m}^\alpha$); α is a velocity exponent, which usually takes a value between 0.3 and 0.9; $|\cdot|$ stands for the absolute value; and $\text{sgn}[\cdot]$ is the signum function, which gives -1 if the argument within square brackets is negative, $+1$ if positive, and 0 if the argument is zero.

Figure 4 illustrates the effect of the velocity exponent α in terms of hysteresis loop $f_{\text{FVD}}(t) - u(t)$ (Fig. 4b) and time history of the reaction force $f_{\text{FVD}}(t)$ (Fig. 4a). Assuming a harmonic displacement $u(t) = u_{\text{max}} \sin(2\pi t/T)$ with period $T = 0.5$ s, it is shown that reducing the exponent α changes the hysteresis loop, from the ellipse

typical of the linear behavior ($\alpha = 1$, thick solid line) to a rectangle ($\alpha \rightarrow 0$). In both graphs, the maximum force $f_{D,\max} = 2\pi c u_{\max}/T$ experienced by the linear device ($\alpha = 1$) is used to make dimensionless the vertical axis. This reveals that, if the damping coefficient c and the period of oscillation T are kept constant, the peak force in the FVD decreases with α ; on the other hand, the less α , the larger the reaction force for small velocities.

More precisely, one can prove that linear ($\alpha = 1$) and nonlinear ($0 < \alpha < 1$) devices experience the same reaction force if the velocity takes the value:

$$|\dot{u}(t)| = v_D = \left(\frac{c_{\text{nonlin}}}{c_{\text{lin}}} \right)^{\frac{1}{1-\alpha}}; \quad (13)$$

for $|\dot{u}(t)| > v_D$ the linear device reacts with higher forces, and the opposite happens for $|\dot{u}(t)| < v_D$. Given that the maximum velocity tends to increase with the amplitude of the oscillatory motion, it follows that nonlinear VFDs, with a small velocity exponent (say $\alpha < 0.50$), are more efficient in the seismic retrofit of existing structures, as comparatively they require lesser deformations to apply the control forces.

When the nonlinear FVD is added to the SDoF oscillator, the equation of seismic motion becomes:

$$\ddot{u}(t) + 2\zeta_0 \omega_i \dot{u}(t) + \omega_i^2 u(t) + \frac{1}{m} f_{\text{FVD}}(t) = -\ddot{u}_g(t), \quad (14)$$

and any reliable scheme of numerical integration (e.g., the Newmark- β method) can be used to evaluate the dynamic response.

In order to allow assessing the performance of FVDs in improving the seismic behavior of an existing structure, Fig. 5 compares the seismic response of three pairs of SDoF oscillators, with and without the additional FVD, for the three periods of vibration T_i selected in the previous section, namely, $T_i = 0.3$ s (stiff oscillator) for the top row, $T_i = 0.5$ s (resonant oscillator) for the central row, and $T_i = 0.8$ s (flexible oscillator) for the bottom row.

The accelerogram $\ddot{u}_g(t)$ is the sweep function of Eq. 9 (see Fig. 1a), and the PGA is $a_g = 0.25$ g, which is representative of a moderate seismic event. The viscous damping ratio of the existing structure is assumed to be $\zeta_0 = 0.02$, while the additional FVD is characterized by the mechanical parameters $c/m = 1.59 \text{ m}^{0.7} \text{ s}^{-1.7}$ and $\alpha = 0.3$.

The time histories of the mass displacement, $u(t)$, are plotted in the three graphs in the first column (Fig. 5a–c), using lighter dashed lines for the seismic responses without control devices and thick solid lines when the nonlinear FVD is added. In all the three cases, a similar reduction in the maximum displacements is observed.

The second column from the left (Fig. 5d–f) presents the comparison in terms of force–displacement hysteresis loops: namely, the narrow pseudo-ellipses show the energy dissipated through the inherent viscous damping of the SDoF oscillator, while the pseudo-rectangles are the hysteresis loops associated with the additional nonlinear FVD. In all the cases, the latter source of energy dissipation is larger, which explains why the seismic displacements are reduced. This is confirmed by the graphs in the third column (Fig. 5g–i), in which the comparison is presented in terms of time histories of various forms of energy and the external work. The two dashed lines are used for the oscillator without FVD, and specifically to plot the cumulative work done by the seismic input, $W_{\text{in}}(t)$, and the total mechanical energy $E(t)$. Solid lines are used for the oscillator with the FVD: in this case, the $W_{\text{in}}(t)$ is the highest among the solid curves, while the total mechanical energy is the envelope of potential energy $V(t)$ and kinetic energy $T(t)$. The other two solid lines are used to plot the energy dissipated by the two damping mechanics, i.e., the additional FVD (black, high curve) and the inherent damping of the existing structure (gray, low curve).

Additionally, the floor spectra $S_f(T_j)$ depicted in the fourth column (Fig. 5j–l) demonstrate that the additional FVD allows reducing the maximum accelerations experienced by any attachment to the building structure (as in previous cases, solid and dashed lines are used for the seismic response with and without additional

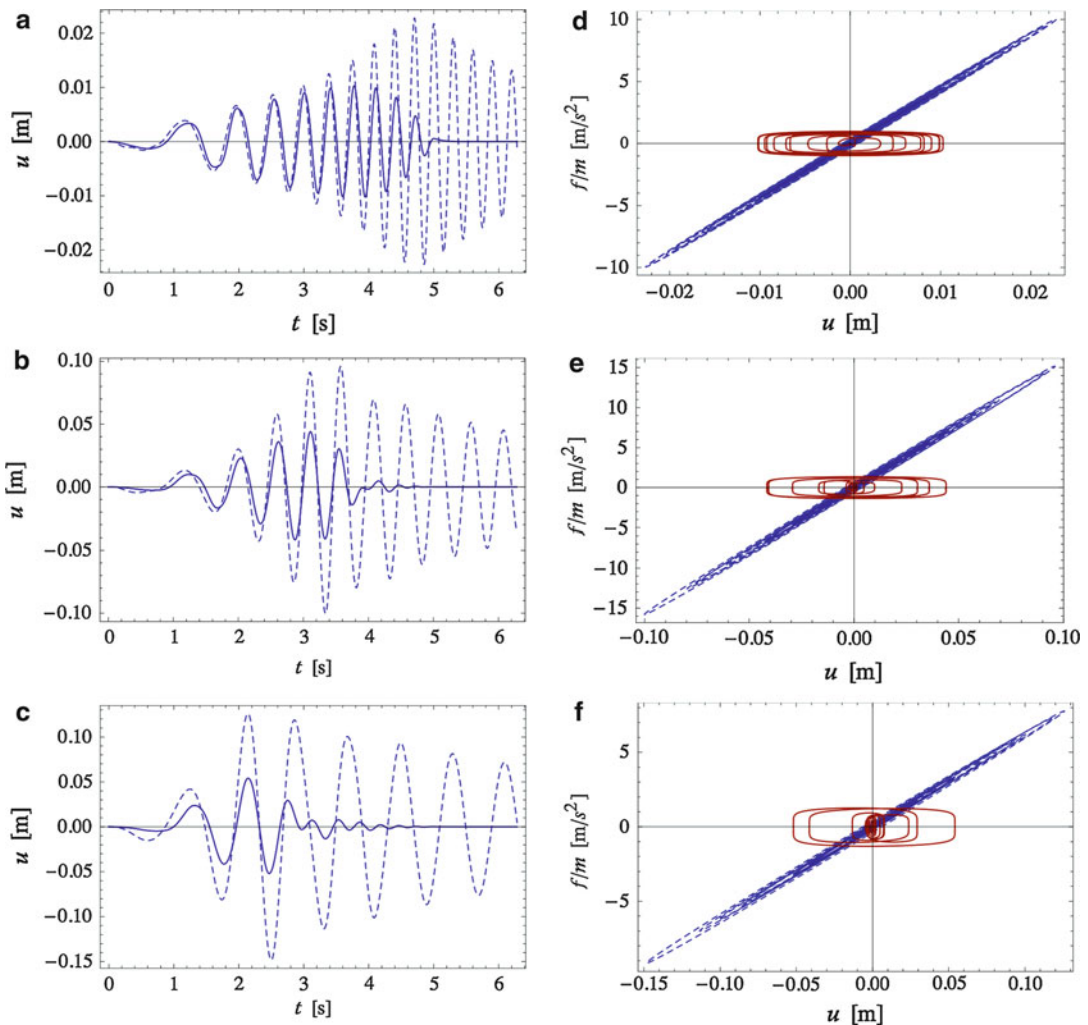
control device). This is particularly significant if primary and secondary structure are tuned (i.e., $T_i = T_j$).

Figure 6 quantifies the performance of the additional FVD for increasing levels of the seismic input, i.e., PGA $a_g = 0.10\text{ g}$ (low intensity), $a_g = 0.25\text{ g}$ (moderate earthquake), and $a_g = 0.40\text{ g}$ (strong motion). It is evident that in all the three cases, the FVD allows mitigating the seismic response. Due to the highly nonlinear behavior of the selected device, however (i.e., the small velocity exponent $\alpha = 0.3$), this retrofitting technique is more efficient in case of low intensity, that is, the lower the PGA, the higher the

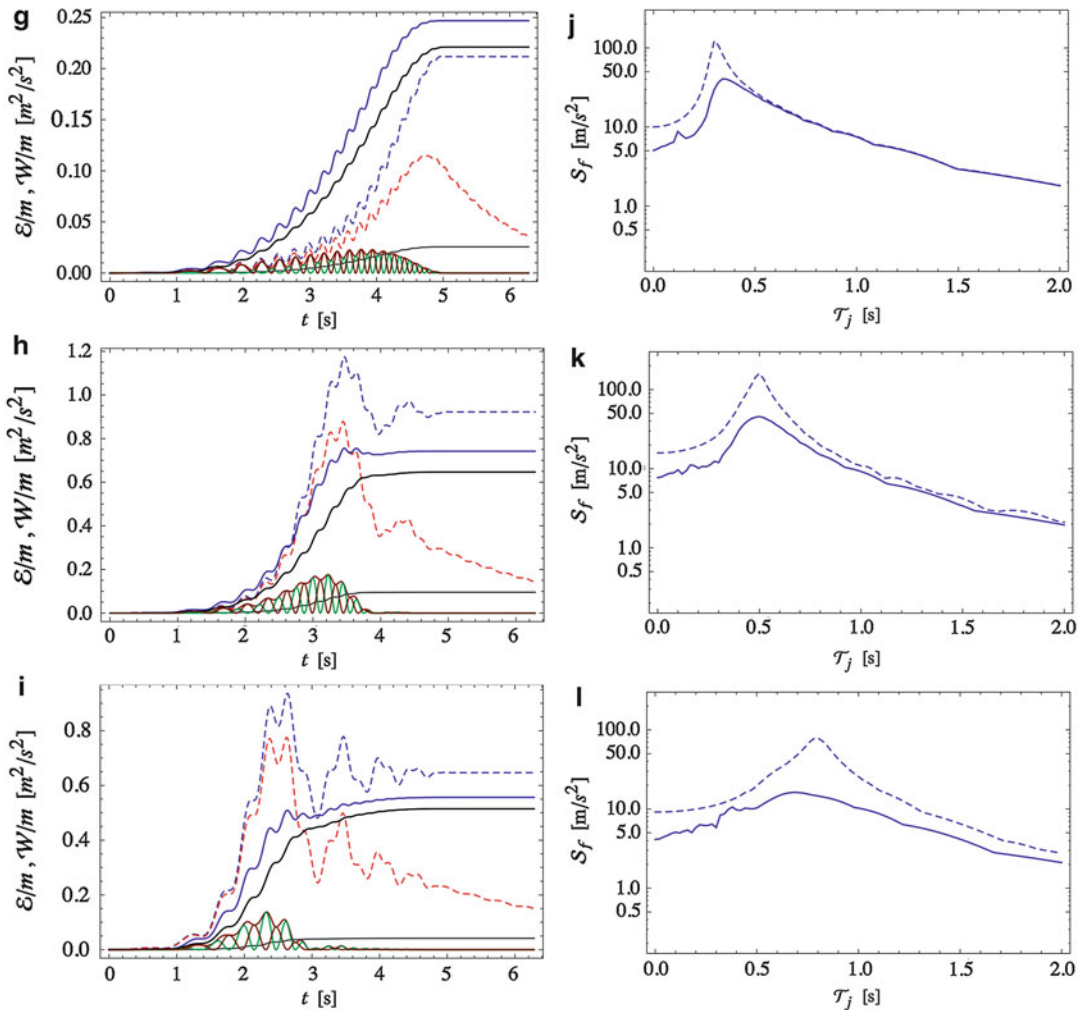
percentage reduction in displacements and accelerations.

Elastomeric Viscoelastic Dampers (EVDs)

Viscoelastic damping is another mechanism of energy dissipation which has been successfully used to improve the dynamic behavior of structures (e.g., Zhang and Soong 1992; Chang et al. 1995; Singh and Moreschi 2002; Park et al. 2004). Historically, one of the first and most impressive applications is the use of



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 5 (continued)



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 5 Effect of the period of vibration. Seismic response of SDOF oscillators with different periods of vibration $T_i = 0.3$ s (top row), 0.5 s

(central row), and 0.8 s (bottom row), in terms of time history of displacement (first column), hysteresis loop (second column), energy dissipated (third column) and floor spectra (fourth column)

10,000 elastomeric viscoelastic dampers (EVDs) in each of the Twin Towers of the World Trade Center (1968–2001) for mitigating the wind-induced vibrations, which otherwise would have been excessive. Because the amount of energy that can be dissipated through viscoelastic damping is comparatively smaller than for fluid viscous dampers and steel hysteretic dampers, EVDs are mainly used for wind engineering applications, although they can also be effective for the seismic retrofit of existing structures.

Figure 7, for instance, shows an application of EVDs, in which the devices are mounted as part of chevron steel braces externally connected to the main structural frame. In this way, energy is dissipated through the shear deformations within each EVD (Fig. 7a), which in turn are caused by the relative movements of the stories.

From a mathematical point of view, the reaction force experienced by a linear viscoelastic device at rest for $t \leq 0$ can be expressed in the time domain through the following convolution integral:

$$f_{\text{EVD}}(t) = \int_0^t \varphi_{\text{EVD}}(t-s) \dot{u}(s) ds, \quad (15)$$

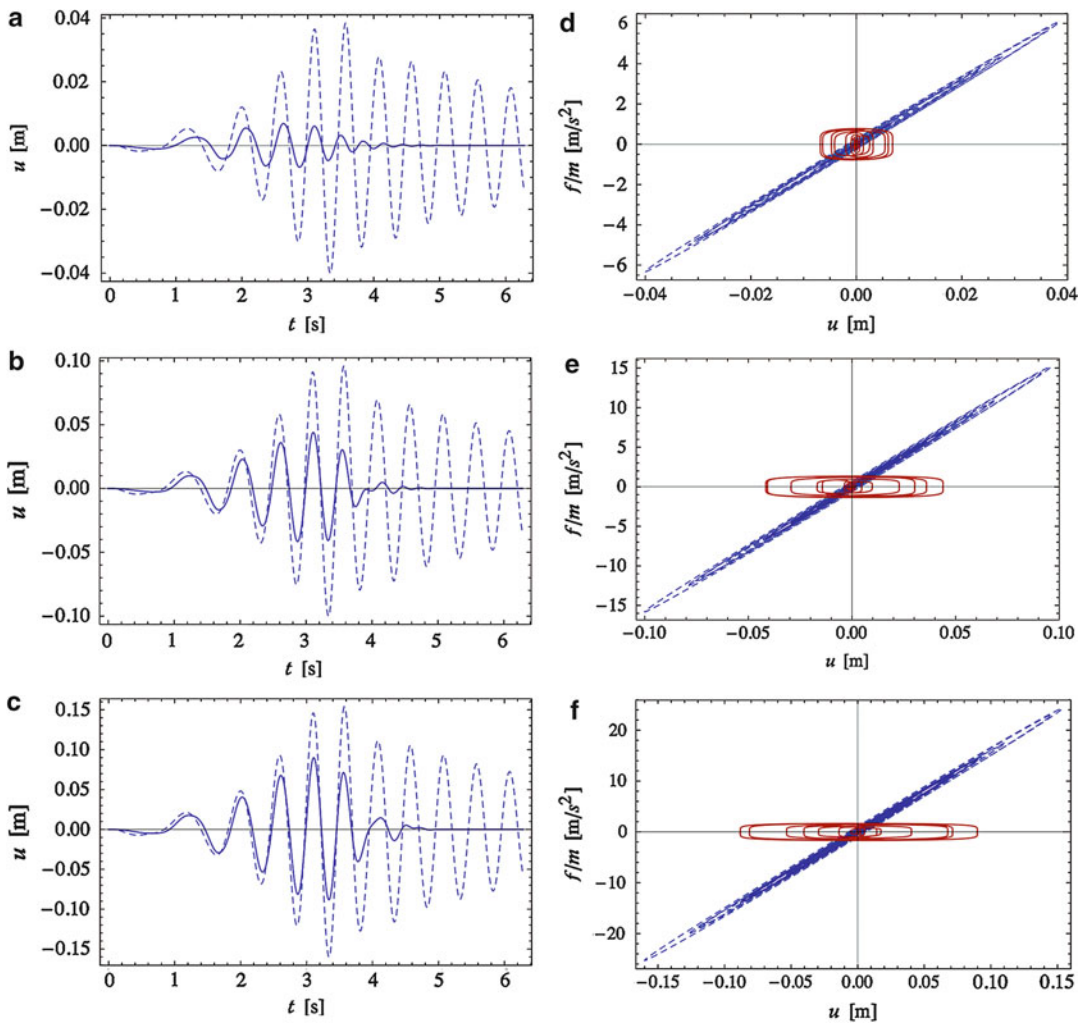
where $\varphi_{\text{EVD}}(t)$ is the relaxation function of the damper, representing the time history of the reaction force due to a unit-step displacement; and $\dot{u}(t)$ is the relative velocity between the two ends of the device. One can easily prove that elastic stiffness and viscous damping are particular cases of the viscoelastic behavior. If the relaxation function for the elastic behavior is a Heaviside's step function of amplitude k

$$\varphi_{\text{EVD}}(t) = k\Theta(t), \quad (16)$$

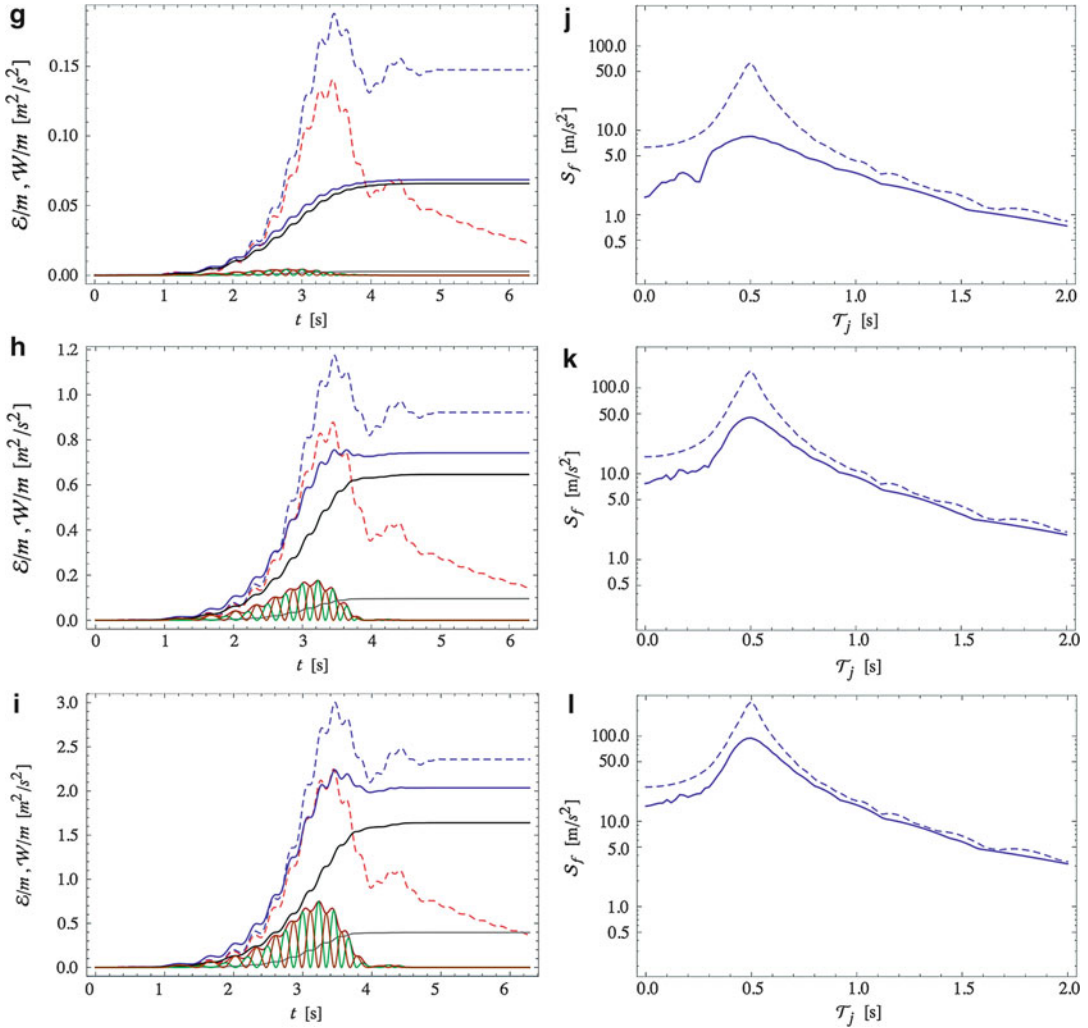
where $\Theta(t) = 0$ for $t < 0$; $1/2$ for $t = 0$; 1 for $t > 0$, then substituting Eq. 16 into Eq. 15 gives

$$\begin{aligned} f_{\text{EVD}}(t) &= \int_0^t k\Theta(t-s) \dot{u}(s) ds \\ &= k \int_0^t \dot{u}(s) ds = ku(t), \end{aligned} \quad (17)$$

which is the constitutive law of a linear-elastic spring of stiffness k . Similarly, if the relaxation



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 6 (continued)



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 6 Effect of the intensity of the ground shaking. Seismic response of SDOF oscillators with different values of PGA $a_g = 0.10 \text{ m/s}^2$ (top

row), 0.25 m/s^2 (central row), and 0.40 m/s^2 (bottom row), in terms of time history of displacement (first column), hysteresis loop (second column), energy dissipated (third column) and floor spectra (fourth column)

function is taken as a Dirac's delta function of intensity c

$$\varphi_{\text{EVD}}(t) = c\delta(t), \quad (18)$$

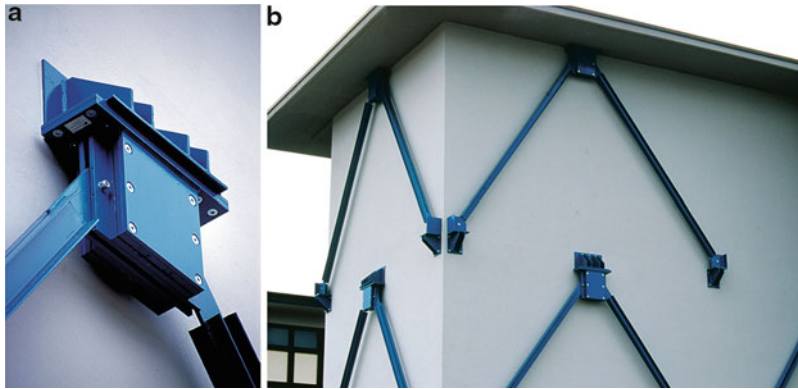
where $\delta(t) = \dot{\Theta}(t)$, Eq. 15 then simplifies as:

$$f_{\text{EVD}}(t) = \int_0^t c\delta(t-s)\dot{u}(s)ds = c\dot{u}(t), \quad (19)$$

which is the constitutive law of a linear-viscous dashpot of damping coefficient c .

In general, the relaxation function is zero for $t < 0$ (this ensures the causality of the constitutive law) and monotonically decreasing for $t \geq 0$. For structural dampers made of natural rubber or other elastomeric materials, the relaxation function can be accurately represented as the superposition of a certain number N of exponential functions with various rates of decay, that is:

$$\varphi_{\text{EVD}}(t) = \left[\sum_{\ell=1}^N R_{\ell} \exp(-t/\tau_{\ell}) \right] \Theta(t), \quad (20)$$



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 7 Elastomeric viscoelastic dampers. Devices as installed in the seismic retrofit of the Gentile-Fermi School in Fabriano, Italy (Courtesy of

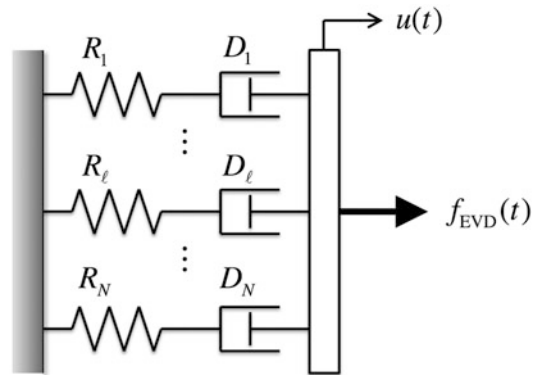
FIP Industriale, Prof Rodolfo Antonucci and Francesco Balducci); detail of the damper (a) and installation with external chevron braces (b)

where the N pairs $\{R_\ell, \tau_\ell\}$ define the discrete relaxation spectrum of the EVD, in which the rigidity coefficient R_ℓ gives the amplitude of the ℓ th contribution, while the relaxation time τ_ℓ provides a measure of the time required by such term to vanish (i.e., the larger the relaxation time, the slower the relaxation process). It is worth noting here that for $\tau_\ell \rightarrow +\infty$, the ℓ th contribution in the r.h.s. of Eq. 20 becomes purely elastic; similarly, the pure viscous behavior is recovered for $\tau_\ell \rightarrow 0$. Any intermediate value of the relaxation time determines a truly viscoelastic behavior, which combines stiffness and energy dissipation.

Interestingly, one can prove that Eq. 20 is the relaxation function of N Maxwell's elements in parallel, i.e., a set of N parallel elastic springs R_ℓ each one connected in series with a viscous dashpot $D_\ell = R_\ell \tau_\ell$ (see Fig. 8). Accordingly, the reaction force can be expressed as (Palmeri et al. 2003; Adhikari and Wagner 2004)

$$f_{\text{EVD}}(t) = \sum_{\ell=1}^N R_\ell \lambda_\ell(t), \quad (21)$$

where $\lambda_\ell(t)$ is the ℓ th additional internal variable associated with the dynamic response of the viscoelastic device, which corresponds to the elastic deformation within the ℓ th Maxwell element and is ruled by



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 8 Generalized Maxwell's model. Spring-dashpot model for EVDs

$$\dot{\lambda}_\ell(t) = \dot{u}(t) - \frac{\lambda_\ell(t)}{\tau_\ell}. \quad (22)$$

Although EVDs used in the engineering practice usually require two or more Maxwell's elements to capture their dynamic behavior, the simplest case of a single Maxwell's element (i.e., $N = 1$) can be effectively used to assess the effects of the relaxation time τ_1 on the seismic performance of a building structure retrofitted with viscoelastic devices. The resulting governing equations for a SDoF oscillator equipped with additional viscoelastic damping then become

$$\begin{cases} \ddot{u}(t) + 2\zeta_0\omega_i\dot{u}(t) + \omega_i^2 u(t) + \underbrace{\frac{R_1}{m}\lambda_1(t)}_{f_{\text{EVD}}(t)/m} = -\ddot{u}_g(t); \\ \dot{\lambda}_1(t) = \dot{u}(t) - \frac{\lambda_1(t)}{\tau_1}, \end{cases} \quad (23)$$

in which the two linear differential equations have to be solved simultaneously.

For this parametric study, ground acceleration $\ddot{u}_g(t)$, undamped periods of vibration $T_i = 2\pi/\omega_i$, and viscous damping ratio ζ_0 are the same as in the previous section; the rigidity coefficient is assumed to be $R_1/m = 0.5 \text{ s}^{-2}$, where m is the mass of the oscillator; three values of the relaxation time have been chosen, namely, $\tau_1 = 0.0136$, 0.136 , and 1.36 s .

In a first stage, the intermediate value of the relaxation time has been used, i.e., $\tau_1 = 0.136 \text{ s}$, and the beneficial effects of the additional viscoelastic damping have been assessed for three periods of vibration, namely, $T_i = 0.3$, 0.5 , and 0.8 s (corresponding to “stiff,” “resonant,” and “flexible” structure, respectively). The comparison between their seismic responses with (solid lines) and without (dashed lines) EVDs is ordinally shown in the three rows of Fig. 9.

Figure 9a–c reveals that in all the three cases, the EVDs are effective in reducing the maximum displacements induced by the ground shaking. Their hysteresis loops (Fig. 9d–f) are ellipses, in which the slope of the major axis increases with the effective stiffness of the device. It is worth stressing here that, while the FVDs considered in the previous section only increase the damping capabilities of the structure, without affecting its stiffness, EVDs have both elastic stiffness and viscous damping. As a result, the retrofitted structure not only is capable of dissipating more energy but is also stiffer. This is confirmed by the floor spectra of Fig. 9j–l, in which the peak of the solid curves (retrofitted system) is always lower (because of the damping) and occurs at a reduced period T_j (because of the increased stiffness).

In a second stage, the effects of the relaxation time τ_1 have been investigated, while keeping

constant the undamped period of vibration ($T_i = 0.5 \text{ s}$). The results of these numerical analyses are offered within Fig. 10, which shows reduced peak displacements in all cases (Fig. 10a–c). Depending on the relaxation time, however, the type of control mechanism provided by the additional EVD can be quite different. In the top row ($\tau_1 = 0.0136 \text{ s}$, Fig. 10a, d, g, j), for instance, the elastomeric device behaves essentially as a linear-viscous damper, while in the bottom row ($\tau_1 = 1.36 \text{ s}$, Fig. 10c, f, i, l), the control force is more similar to an elastic force. A better performance is clearly seen in the central row ($\tau_1 = 0.136 \text{ s}$, Fig. 10b, e, h, k), where the control force is truly viscoelastic: this happens because in this case the timescale for the relaxation of the EVD is comparable to the period of the oscillations, while the relaxation is too fast in the top row ($\tau_1/T_i < 1/20$), too slow in the bottom row ($\tau_1/T_i > 2$).

There are interesting consequences for the different types of control forces exerted by the elastomeric device. In terms of hysteresis loops, for instance, the slope of the major axis of the viscoelastic ellipses in the $f_{\text{EVD}}(t) - u(t)$ diagram increases with the relaxation time τ_1 . In the top row (Fig. 10d), in which the viscous damping prevails, the major axis is substantially horizontal (similar to the linear hysteresis loop for $\alpha = 1$ within Fig. 4a); on the contrary, the steepest slope is seen in the bottom row (Fig. 10f), in which the elastic stiffness prevails, and therefore the ellipse is very narrow (meaning that little energy is dissipated in each cycle). In the central row, where the EVD provides both stiffness and damping, the hysteresis loop has the major axis inclined and shows that a significant amount of energy is dissipated.

It is also interesting to note the differences between the floor spectra for the three cases. While the proper viscoelastic behavior in the central row (Fig. 10k) allows the peak of the floor spectrum to reduce (because of the additional damping) and shift to the left (because of the additional stiffness), only the reduction of the peak is seen in the top row (Fig. 10j) and the peak shift in the bottom row (Fig. 10l).

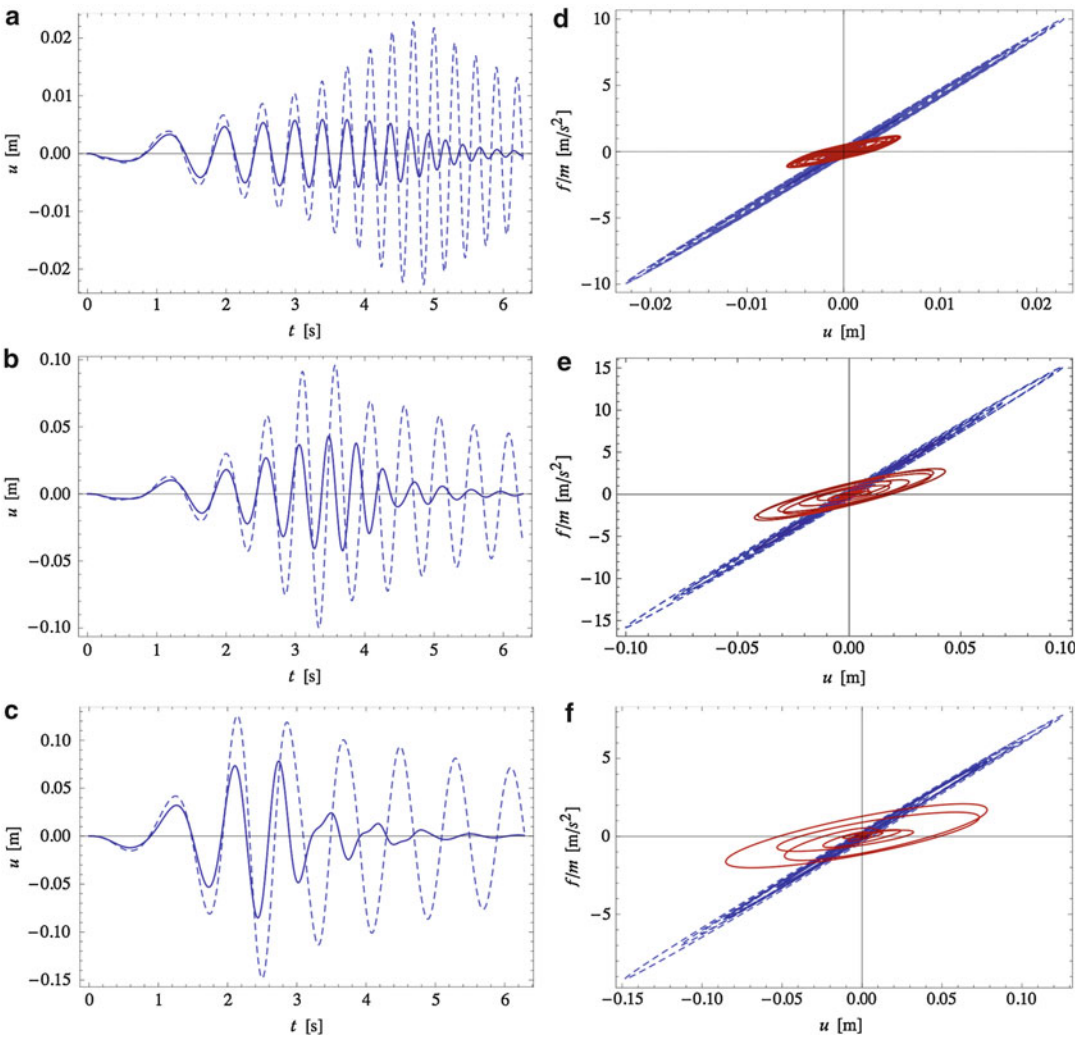
From the above observations, it follows that the relaxation times of the devices (also in relationship to the periods of vibration) are the key parameters to be considered while designing the seismic retrofitting of an existing structure with EVDs.

Steel Hysteretic Dampers (SHDs)

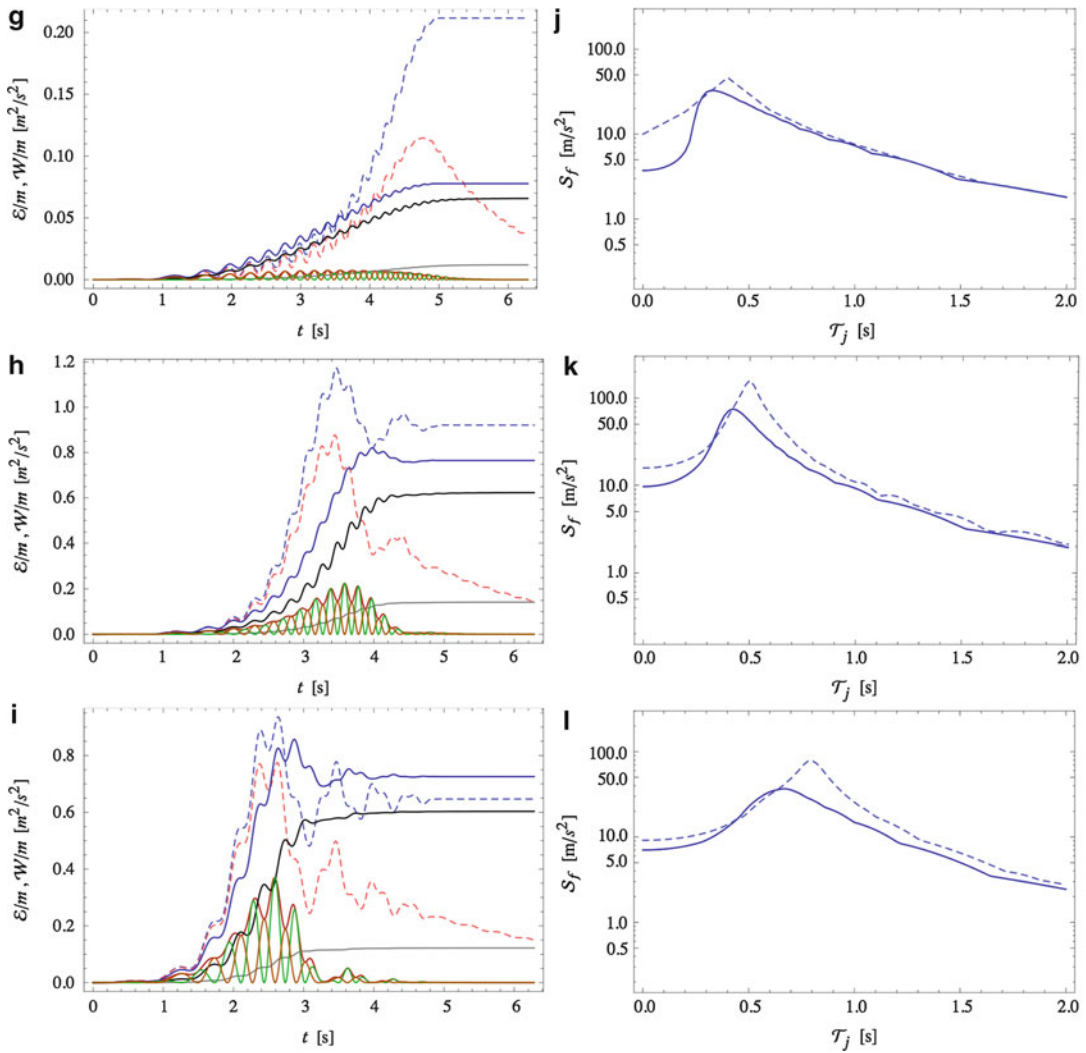
Plastic deformations in metals can be exploited to dissipate energy and therefore mitigate the

effects of seismic forces. In this case, the device has to be designed in order to maximize the plastic work and ensure stable hysteretic cycles (e.g., Nakashima et al. 1996).

Low-yield steel plates with hourglass and triangular shapes have been proposed, and they are best known with the acronyms ADAS (added viscous and stiffness) and TADAS (triangular ADAS), respectively. Such devices are usually installed at the top of stiff chevron steel braces, so that the relative lateral movement of two consecutive stories induces bending about the weak



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 9 (continued)



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 9 Effect of the period of vibration. Seismic response of SDOF oscillators with different periods of vibration $T_i = 0.3$ s (top row), 0.5 s

(central row), and 0.8 s (bottom row), in terms of time history of displacement (first column), hysteresis loop (second column), energy dissipated (third column) and floor spectra (bottom column)

axis of the steel plate (Xia and Hanson 1992; Chou and Tsai 2002; Alehashem et al. 2008; Bayat and Abdollahzadeh 2011).

Figure 11 (reproduced from Alehashem et al. 2008) shows the typical geometry of ADAS and TADAS devices. In the first case (Fig. 11a), the steel plates are clamped, and their hourglass shape allows maximizing the energy dissipation when the plates experience plastic bending. In the second case (Fig. 11b), the steel plates are clamped at the top and pinned

at the bottom: the bending moment then increases linearly with the height, and the triangular shape then becomes the most efficient one.

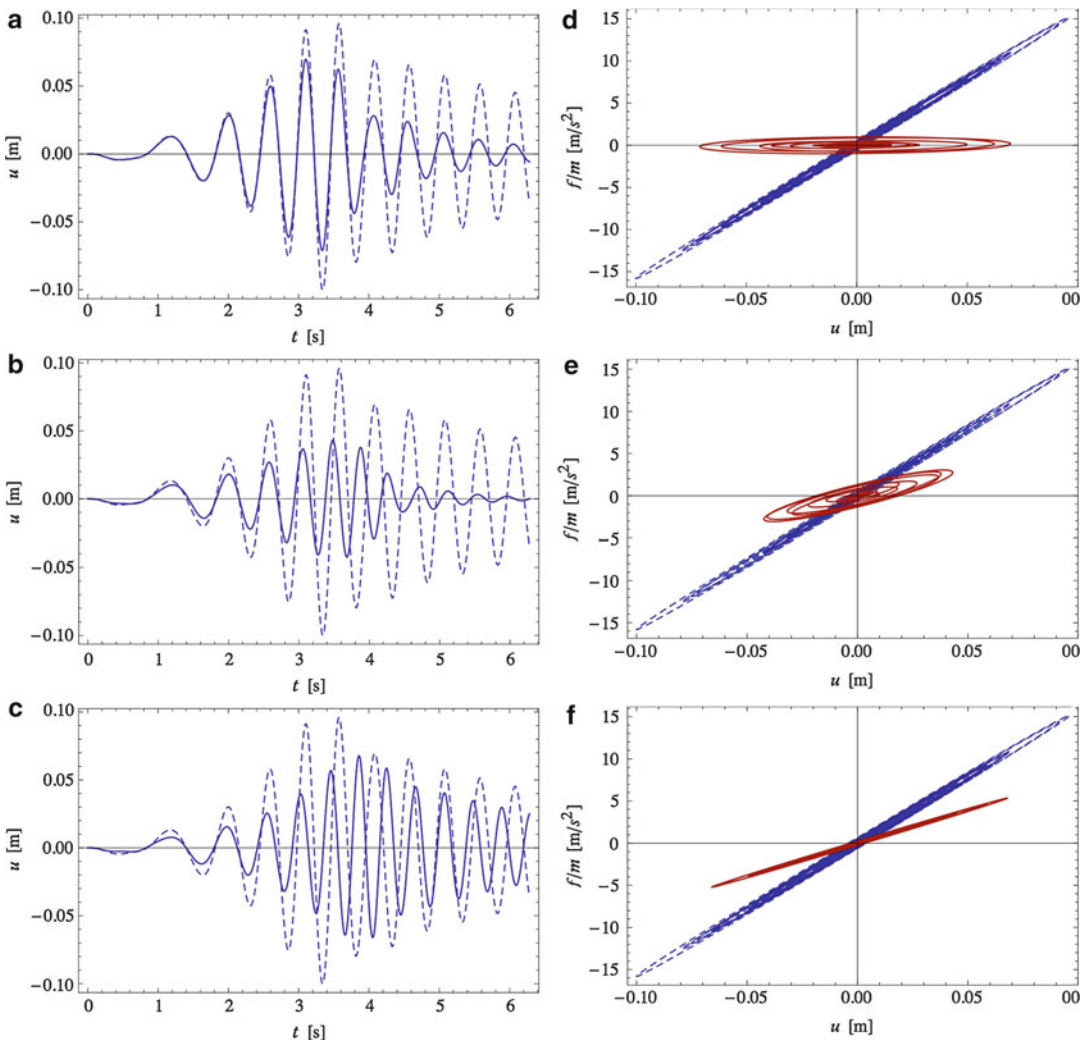
Buckling-restrained braces (BRBs), also known as BRADs (buckling-restrained axial dampers), are another type of metal dampers, in which axial rather than bending stresses in low-yield steel are exploited. The system typically consists of a cruciform low-yield pinned steel brace, whose buckling is prevented by encasing the brace within a concrete-filled steel

tube, and a special coat inhibits the bond between inner steel and concrete (so that the brace can freely elongate and shorten in each cycle). This solution is particularly efficient as the brace is subjected to uniaxial stress/strain, and therefore all the volume of material can contribute to energy dissipation (e.g., Sabelli et al. 2003; Black et al. 2004; Di Sarno and Manfredi 2010).

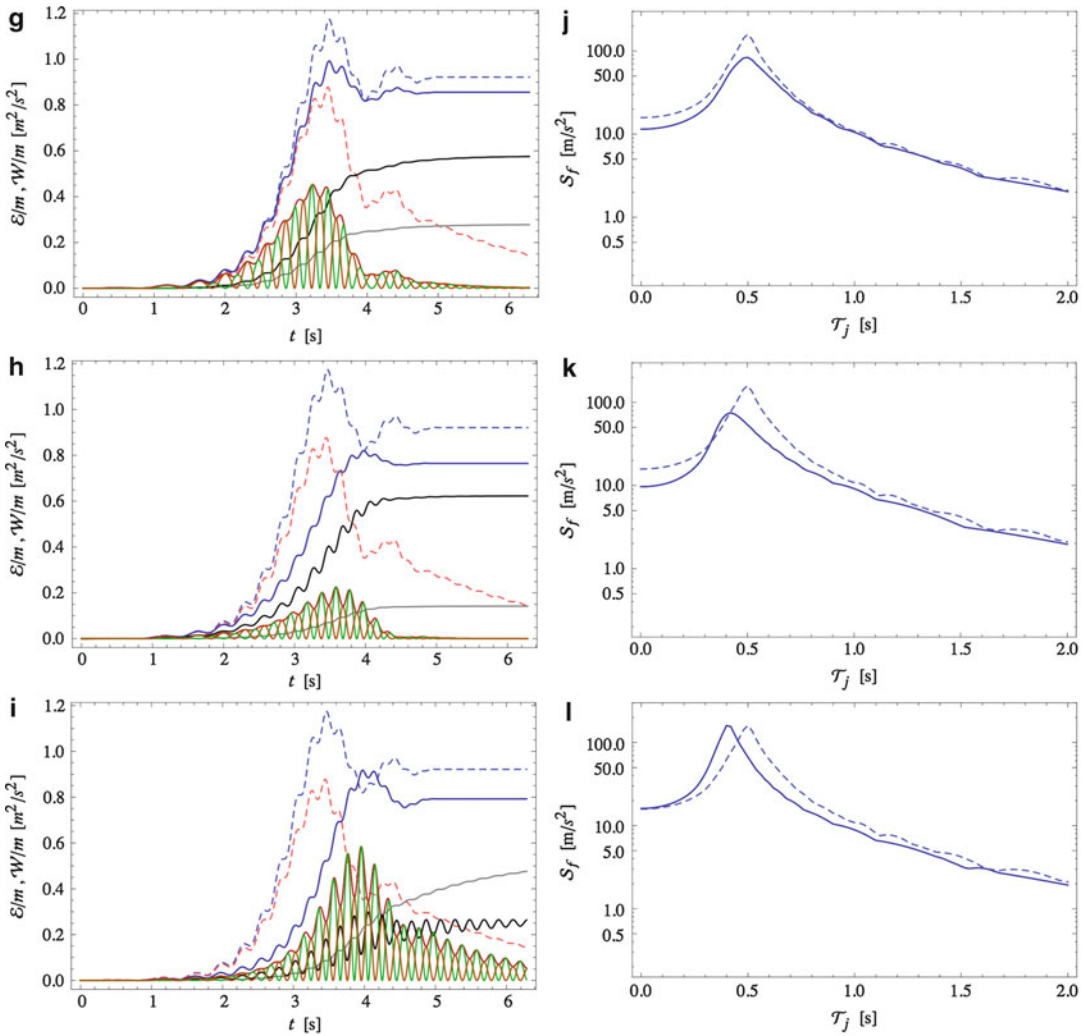
Figure 12 shows the application of this type of devices for the seismic retrofit of two schools with RC framed structure in Italy. In the first case (Fig. 12a), the devices are left exposed (increasing the sense of safety for the occupants),

while in the second case (Fig. 12b), the choice has been to cover the devices.

Many mathematical models have been proposed for the representation of the elastoplastic force experienced by SHDs (steel hysteretic dampers) such as ADAS and BRB devices. The Bouc–Wen model is probably the most popular one, as it can be easily implemented to perform time-history analyses (Bouc 1971; Wen 1976; Ismail et al. 2009). The model requires the introduction of a dimensionless hysteretic variable $z(t)$, which is proportional to the elastoplastic force:



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 10 (continued)



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 10 Effect of the relaxation time. Seismic response of SDoF oscillators with different values of the relaxation time $\tau_i = 0.0136$ s (*top row*),

0.136 s (*central row*), and 1.36 s (*bottom row*), in terms of time history of displacement (*first column*), hysteresis loop (*second column*), energy dissipated (*third column*) and floor spectra (*fourth column*)

$$f_{\text{SHD}}(t) = f_y z(t), \quad (24)$$

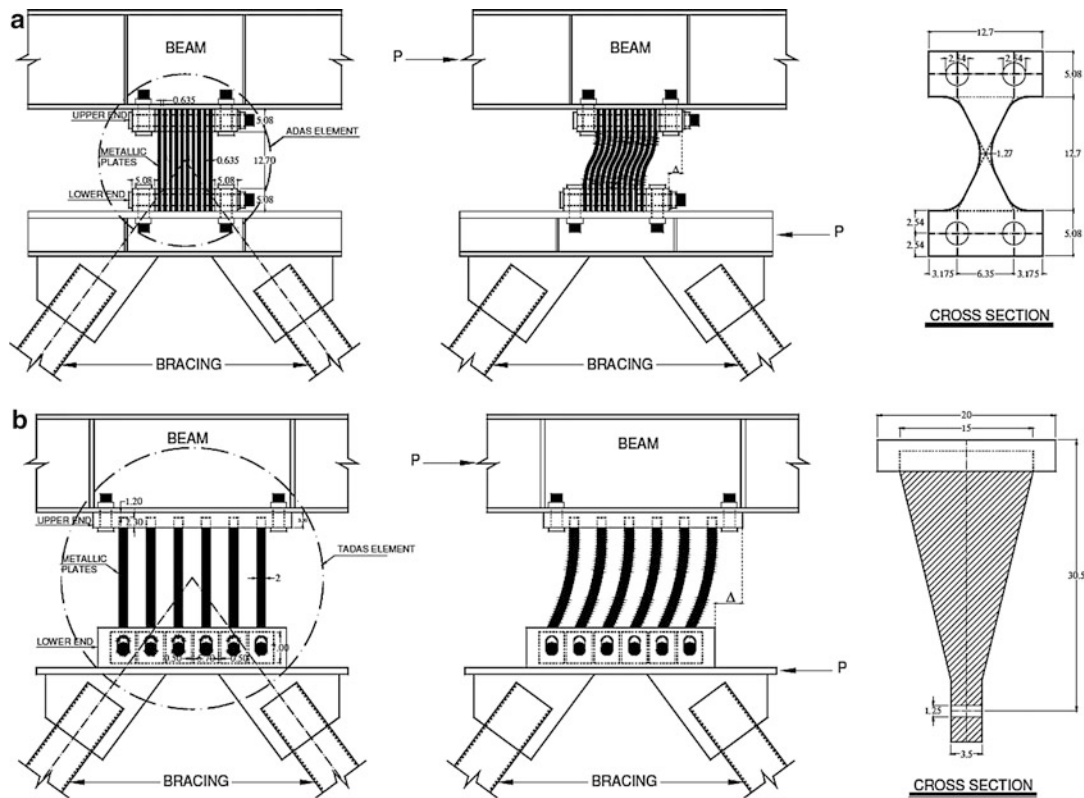
and is ruled by the following nonlinear differential equation:

$$\begin{aligned} \ddot{z}(t) = & \frac{1}{\delta_y} \left(\dot{u}(t) - \sigma |\dot{u}(t)| |\dot{z}(t)|^{n-1} \dot{z}(t) \right. \\ & \left. - (1 - \sigma) \dot{u}(t) |\dot{z}(t)|^n \right), \end{aligned} \quad (25)$$

where f_y and δ_y are the yield force and yield displacement of the device, while σ and n are

two dimensionless parameters which control the shape of the hysteretic cycles. That is, the larger $n > 1$, the quicker the transition between elastic and plastic branches, and vice versa (see left column of Fig. 13); the larger $\sigma > 0$, the steeper the slope of the hysteretic loop $f_{\text{SHD}}(t) - u(t)$ when the device reenters the elastic branch (for $\sigma = 0.5$ one obtains the same slope f_y/δ_y as the elastic branch; see right column of Fig. 13).

For the sake of simplicity, the model of Eqs. 24 and 25 assumes that once the plastic



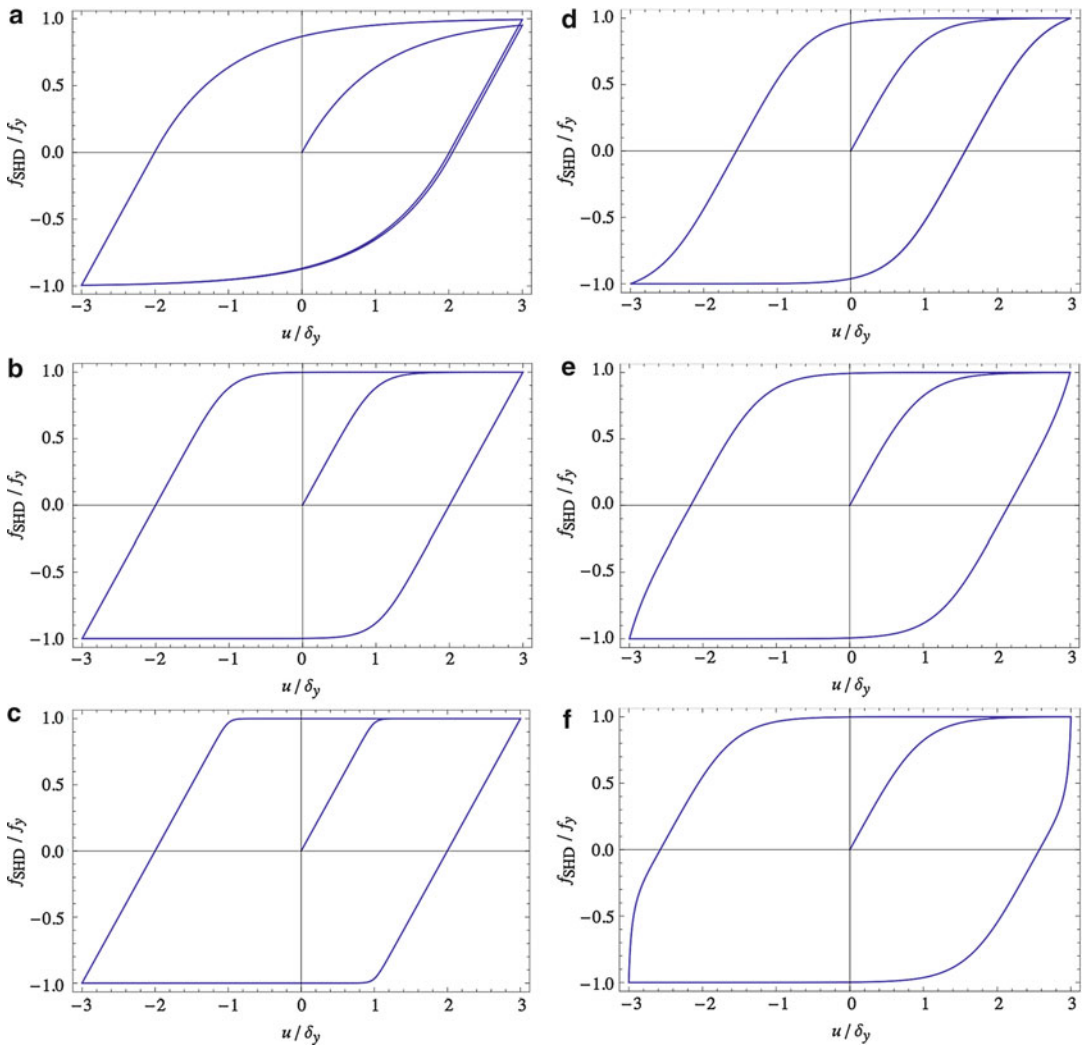
Passive Control Techniques for Retrofitting of Existing Structures, Fig. 11 Steel hysteretic dampers. Schematics of two different devices: ADAS, with

hourglass shape (a); TABAS, with triangular shape (b) (Reproduced from Alehashem et al. 2008)



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 12 Steel hysteretic dampers. Devices as installed in the seismic retrofit of two schools in Italy, the Cappuccini School in Ramacca (a) (Courtesy

of FIP Industriale and Dr Fabio Neri) and the Giulio Perticari School in Senigallia (b) (Courtesy of FIP Industriale and Prof Rodolfo Antonucci)



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 13 Bouc–Wen hysteretic loops. Effects of the parameters α and n on the shape of

the force–displacement loop: $\sigma = 0.5$, $n = 1.01$ (a); $\sigma = 0.5$, $n = 5$ (b); $\sigma = 0.5$, $n = 15$ (c); $\sigma = 0.1$, $n = 3$ (d); $\sigma = 1.0$, $n = 3$ (e); $\sigma = 10$, $n = 3$ (f)

branch has been reached, there is no further increase in the elastoplastic force, i.e., $|f_{\text{SHD}}(t)| \leq f_y$ (elastic-perfectly plastic behavior). If the device shows a residual post-yielding stiffness $k_{\text{d,res}}$, an elastic term can be added to the r.h.s. of Eq. 24, which then becomes:

$$f_{\text{SHD}}(t) = (f_y - k_{\text{d,res}}\delta_y)z(t) + k_{\text{d,res}}u(t). \quad (26)$$

Additionally, degradation effects can also be accounted for through a reduction of the yield

force f_y and the elastic stiffness f_y/δ_y , which in turn can be expressed as functions of the absorbed hysteretic energy:

$$W_{\text{SHD}}(t) = \int_0^t f_{\text{SHD}}(\tau)\dot{u}(\tau)d\tau, \quad (27)$$

i.e., the more hysteretic energy is dissipated, the more damage is accumulated in the device, and the larger is the reduction in stiffness and strength. It must be said, however, that the majority of the devices used for the seismic retrofitting

of existing structures show stable hysteretic loops for many cycles, without any appreciable degradation.

It is also worth noting here that, while FVDs and EVDs are frequency-dependent, i.e., the rate of change of displacements and forces applied to such devices changes their dynamic response, SHDs are frequency-independent, so their performance is not affected by the energy content of the dynamic excitation. Another peculiar aspect of SHDs is that they require relatively large displacements to dissipate energy, and for this reason they are specifically used for earthquake engineering applications

(while both FVDs and EVDs can also be used for other dynamic loads, e.g., to mitigate the effects of wind forces in tall buildings).

In order to perform the dynamic analysis, the equation of motion for a SDoF oscillator equipped with SHD can then be posed in the following state-space form:

$$\dot{\mathbf{y}}(t) = \mathbf{F}(\mathbf{y}(t), t), \quad (28)$$

where $\mathbf{y}(t) = \{u(t), \dot{u}(t), z(t)\}^T$ is the array of the three state variables of the system (displacement, velocity, and hysteretic variable), and $\mathbf{F}$ collects the differential equation for each state variable:

$$\mathbf{F}(\mathbf{y}(t), t) = \left\{ \begin{array}{l} \dot{u}(t) \\ -\omega_0^2 u(t) - 2\zeta_0 \omega_0 \dot{u}(t) - \beta z(t) - \ddot{u}_g(t) \\ \frac{\kappa \omega_0^2}{\beta} \left[\dot{u}(t) - \sigma |\dot{u}(t)| |\dot{z}(t)|^{n-1} \dot{z}(t) - (1 - \sigma) \dot{u}(t) |\dot{z}(t)|^n \right] \end{array} \right\}, \quad (29)$$

in which $\beta = f_y/m$ is a measure of the ground acceleration needed to reach the yielding point of the SHD and $\kappa = f_y/(k \delta_y)$ is the dimensionless stiffness of the device, normalized with respect to the elastic stiffness of the structure without damper; the superscripted T stands for the transpose operator. Different schemes of numerical integration can be used for Eqs. 28 and 29, including the 4th-order Runge–Kutta method.

As for the other types of damping mechanics considered in the previous sections, it is interesting to assess the sensitivity of their performance to the two governing parameters for SHDs, i.e., β and κ (or, alternatively, f_y and δ_y). The value of β determines the intensity of the seismic event that mobilizes the devices: if β is too small, even modest earthquakes will result in plastic deformations in the dampers; if β is too large, the dampers would only be effective for the most intense events. The selection of κ is also very important too, as it affects the fundamental period of the retrofitted structure. Indeed, if T_1 is the fundamental period of the existing structure, $T_1^* = T_1/(1 + \kappa) < T_1$ is the new period with the SHDs installed. A poor choice of κ may result in

an overall increase in the seismic forces (which would then reduce the effectiveness of the SHDs). Additionally, if κ is too small, then the existing structure would still take most of the seismic forces in the retrofitted structure, with the SHDs then being unable to dissipate enough energy during the seismic event.

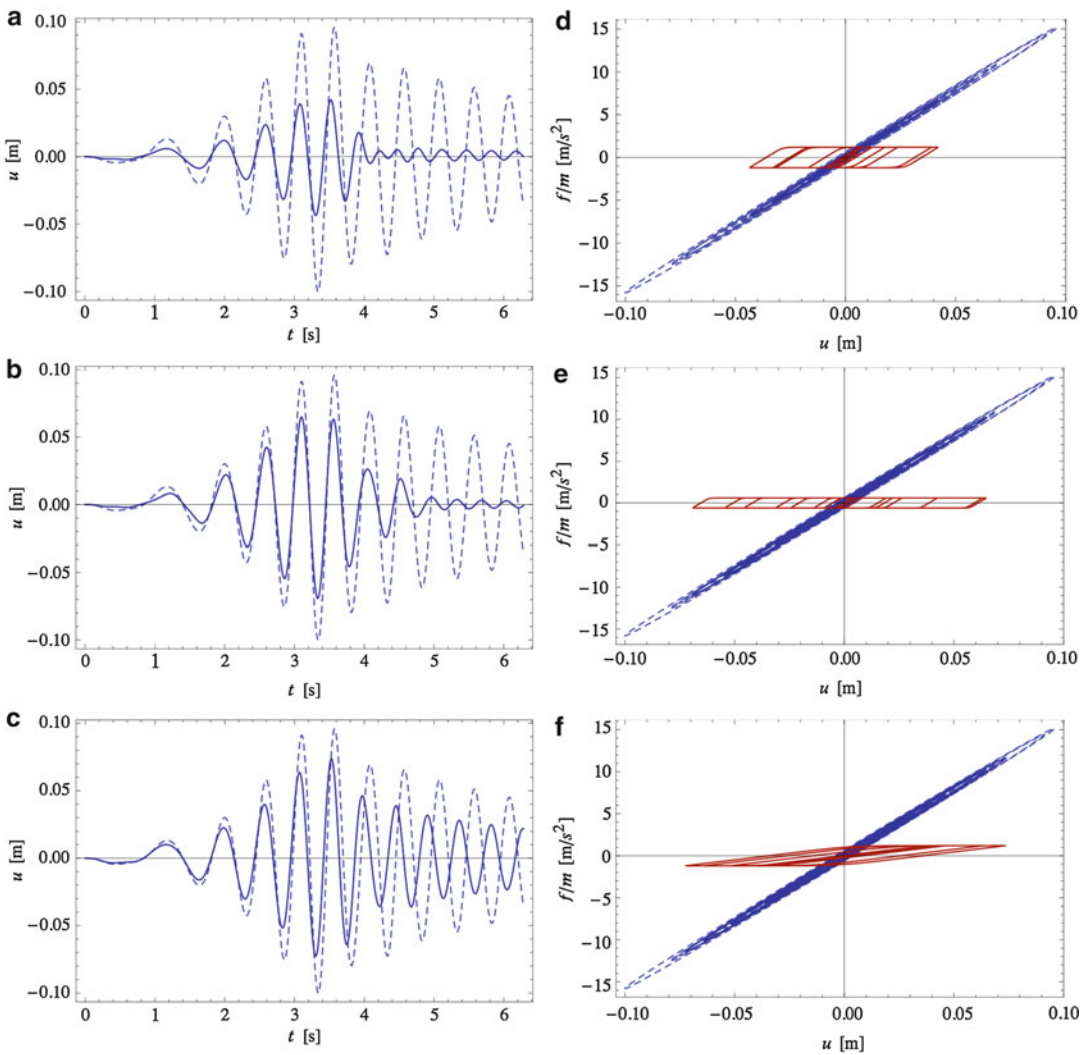
Figure 14 confirms that a careful selection of the design parameters is required in order to maximize the performance of SHDs. In the top row $\beta = 1.2 \text{ m/s}^2$, which is about half of the pseudo-spectral acceleration for $T_0 = 0.5 \text{ s}$ (see Fig. 1b), and $\kappa = 1$, meaning that existing structure and retrofitting device will take the same seismic forces in the elastic range. As a result, a significant reduction in the maximum displacement is achieved (Fig. 14a), a large amount of energy is dissipated (Fig. 14d), and the ordinates of the floor spectrum are considerably mitigated (Fig. 14j). The other two rows of Fig. 14 show that lower values of yielding force ($\beta = 0.6 \text{ m/s}^2$ in the second row) and stiffness ($\kappa = 0.2$ in the third row) diminish the effectiveness of SHDs, as in both cases large displacements are then experienced by the structure.

Conclusions

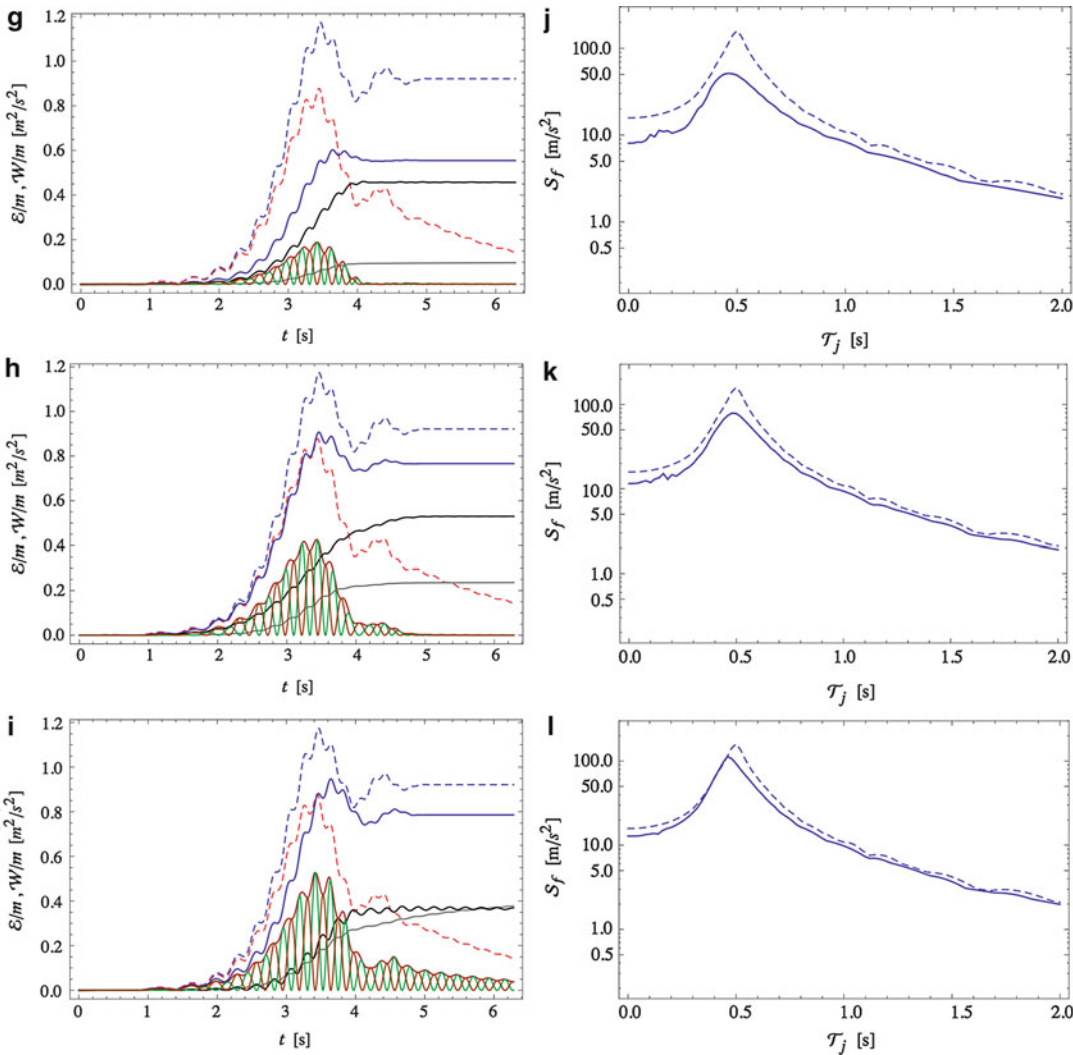
In this contribution, three among the most popular passive control techniques for the seismic retrofitting of existing building structures have been reviewed. Governing equations and mathematical models have been presented for viscous fluid dampers (VFDs), elastomeric viscoelastic dampers (EVDs), and steel hysteretic dampers (SHDs), along with some examples of recent applications. A simple seismic signal, a sinusoidal function with both amplitude and frequency

varying with time, has been used to investigate the effects of the governing parameters for each type of device.

It has been shown that, despite the different damping mechanisms (viscosity of fluids, relaxation of elastomers, plasticity of metals), all these devices can significantly improve the seismic response of an existing structure, in terms of maximum displacements and floor spectra. In all cases, however, the design parameters of the added dampers must be carefully selected in order to maximize the amount of energy



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 14 (continued)



Passive Control Techniques for Retrofitting of Existing Structures, Fig. 14 Effect of the yield force and stiffness. Seismic response of SDoF oscillators with different values of design parameters β and κ , in terms of

time history of displacement (*first column*), hysteresis loop (*second column*), energy dissipated (*third column*) and floor spectra (*fourth column*)

dissipation, and nonlinear time-history analyses are instrumental to measure the expected benefit of the control devices.

For the sake of simplicity, all the numerical results presented have been obtained considering a single-degree-of-freedom (SDoF) oscillator. The extension of these results to multi-degree-of-freedom (MDoF) structures is straightforward if the existing building is regular both in plan and in elevation, while further

design considerations are needed to handle situations where torsional effects and soft stories are present, as larger inter-story drifts are expected at certain critical locations. Another important design constraint is the available ductility of the existing structural members. Indeed, the additional energy dissipation capacity can only be exploited if the existing structure can accommodate sufficiently large deformations.

References

- Adhikari S, Wagner N (2004) Direct time-domain integration method for exponentially damped linear systems. *Comput Struct* 82:2453–2461
- Alehashem SMS, Keyhani A, Pourmohammad H (2008) Behavior and performance of structures equipped with ADAS and TADAS dampers (a comparison with conventional structures). 14th World Conference on Earthquake Engineering, 12–17 Oct 2008, Beijing
- Bayat M, Abdollahzadeh G (2011) Analysis of the steel braced frames equipped with ADAS devices under the far field records. *Latin Am J Solids Struct* 8:163–181
- Black C, Makris N, Aiken ID (2004) Component testing, seismic evaluation and characterization of buckling-restrained braces. *J Struct Eng ASCE* 130:880–894
- Bouc R (1971) Modèle mathématique d'hystérésis: application aux systèmes à un degré de liberté (in French). *Acustica* 24:16–25
- Chang KC, Soong TT, Oh S-T, Lai ML (1995) Seismic behavior of steel frame with added viscoelastic dampers. *J Struct Eng ASCE* 121:1418–1426
- Chou C-C, Tsai K-C (2002) Plasticity-fibre model for steel triangular plate energy dissipating devices. *Earthq Eng Struct Dyn* 31:1643–1655
- Dargush GF, Sant RS (2005) Evolutionary aseismic design and retrofit of structures with passive energy dissipation. *Earthq Eng Struct Dyn* 34:1601–1626
- Di Sarno L, Elnashai AS (2005) Innovative strategies for seismic retrofitting of steel and composite structures. *Prog Struct Eng Mater* 7:115–135
- Di Sarno L, Manfredi G (2010) Seismic retrofitting with buckling restrained braces: application to an existing non-ductile RC framed building. *Soil Dyn Earthq Eng* 30:1279–1297
- Hoang N, Fujino Y, Warnitchai P (2008) Optimal tuned mass damper for seismic applications and practical design formulas. *Eng Struct* 30:707–715
- Hueste MBD, Bai J-W (2007) Seismic retrofit of a reinforced concrete flat-slab structure: part I – seismic performance evaluation. *Eng Struct* 29:1165–1177
- Ismail M, Ikhouane F, Rodellar J (2009) The hysteresis Bouc-Wen model, A survey. *Arch Comput Meth Eng* 16:161–188
- Lee D, Taylor DP (2001) Viscous damper development and future trends. *Struct Des Tall Build* 10:311–320
- Lin W-H, Chopra AK (2002) Earthquake response of elastic SDF systems with non-linear fluid viscous dampers. *Earthq Eng Struct Dyn* 31:1623–1642
- Lin J-L, Tsai K-C, Yu Y-J (2011) Bi-directional coupled tuned mass dampers for the seismic response control of two-way asymmetric-plan buildings. *Earthq Eng Struct Dyn* 40:675–690
- Marano GC, Quaranta G, Avakian J, Palmeri A (2013) Identification of passive devices for vibration control by evolutionary algorithms, Ch. 15. In: Gandomi AH, Yang X-S, Marand ST, Alavi AH (eds) *Metaheuristic applications in structures and infrastructures*. Elsevier, pp 373–387. <http://www.sciencedirect.com/science/article/pii/B9780123983640000152>
- Marriott D, Pampanin S, Bull D, Palermo A (2008) Dynamic testing of precast, post-tensioned rocking wall systems with alternative dissipating solutions. *Bull N Z Soc Earthq Eng* 41:90–103
- Martinez-Rodrigo M, Romero ML (2003) An optimum retrofit strategy for moment resisting frames with nonlinear viscous dampers for seismic applications. *Eng Struct* 25:913–925
- Molina FJ, Sorace S, Terenzi G, Magonette G, Viacoz B (2004) Seismic tests on reinforced concrete and steel frames retrofitted with dissipative braces. *Earthq Eng Struct Dyn* 33:1373–1394
- Muscolino G, Palmeri A (2007) An earthquake response spectrum method for linear light secondary substructures. *ISET J Earthq Technol* 44:193–211
- Nakashima M, Saburi K, Tsuji B (1996) Energy input and dissipation behaviour of structures with hysteretic dampers. *Earthq Eng Struct Dyn* 25:483–496
- Ozcan O, Binici B, Ozcebe G (2008) Improving seismic performance of deficient reinforced concrete columns using carbon fiber-reinforced polymers. *Eng Struct* 30:1632–1646
- Palmeri A, Makris N (2008) Response analysis of rigid structures rocking on viscoelastic foundation. *Earthq Eng Struct Dyn* 37:1039–1063
- Palmeri A, Ricciardelli F, De Luca A, Muscolino G (2003) State space formulation for linear viscoelastic dynamic systems with memory. *J Eng Mech ASCE* 129:715–724
- Park J-H, Kim J, Min K-W (2004) Optimal design of added viscoelastic dampers and supporting braces. *Earthq Eng Struct Dyn* 33:465–484
- Sabelli R, Mahin S, Chang C (2003) Seismic demands on steel braced frame buildings with buckling-restrained braces. *Eng Struct* 25:655–666
- Sackman JL, Kelly JM (1979) Seismic analysis of internal equipment and components in structures. *Eng Struct* 1:179–190
- Singh MP, Moreshchi LM (2002) Optimal placement of dampers for passive response control. *Earthq Eng Struct Dyn* 31:955–976
- Soong TT, Spencer BF Jr (2002) Supplemental energy dissipation: state-of-the-art and state-of-the-practice. *Eng Struct* 24:243–259
- Uang C-M, Bertero VV (1990) Evaluation of seismic energy in structures. *Earthq Eng Struct Dyn* 19:77–90
- Wen YK (1976) Method for random vibration of hysteretic systems. *J Eng Mech ASCE* 102:246–263
- Wu Y-F, Liu T, Oehlers DJ (2006) Fundamental principles that govern retrofitting of reinforced concrete columns by steel and FRP jacketing. *Adv Struct Eng* 9:507–532
- Xia C, Hanson RD (1992) Influence of ADAS element parameters on building seismic response. *J Struct Eng ASCE* 118:1903–1918
- Zhang R-H, Soong TT (1992) Seismic design of viscoelastic dampers for structural applications. *J Struct Eng ASCE* 118:1375–1392

Passive Seismometers

Gerardo Alguacil¹ and Jens Havskov²

¹Instituto Andaluz de Geofísica, University of Granada, Granada, Spain

²Department of Earth Science, University of Bergen, Bergen, Norway

Synonyms

Accelerometer; Amplitude response; Phase response; Response function; Seismometer

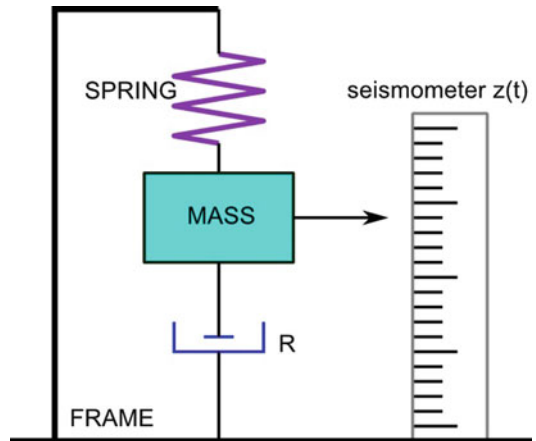
Introduction

A seismic sensor measures the ground motion and outputs a voltage, usually proportional with ground velocity. Earlier purely mechanical sensors measured the displacement of a stylus representing the amplified ground motion while newer sensors pick up the motion using a coil moving in a magnetic field. Common for these two types of sensors is that there is no electronics involved and they are therefore called passive sensors in contrast to the many new sensors with active electric circuits as an integrated part of the sensor, the so-called active sensors.

Standard Inertia Seismometer

The objective is to measure the ground motion at a point with respect to this same point undisturbed. The main difficulty is that the measurement is done in moving reference frame. So displacement cannot be measured directly and, according to the inertia principle, an inertial force will appear on a mass only if the reference frame (in this case the ground) has an acceleration, so the seismometer can only measure velocities or displacements associated with nonzero values of ground acceleration.

Since the measurements are done in a moving reference frame (the Earth's surface), almost all seismic sensors are based on the inertia of



Passive Seismometers, Fig. 1 A mechanical inertial seismometer. *R* is a dash pot (Figure from Havskov and Alguacil 2010)

a suspended mass, which will tend to remain stationary in response to external motion. The relative motion between the suspended mass and the ground will then be a function of the ground's motion.

Figure 1 shows a simple seismometer that will detect vertical ground motion. It consists of a mass suspended from a spring. The resonance angular frequency of the mass-spring system is $\omega_0 = \sqrt{k/m}$, where $\omega_0 = 2\pi/T_0$ and T_0 is the corresponding natural period (s) of the swinging system. The motion of the mass is damped using a “dash pot” so that the mass will not swing excessively near the resonance frequency of the system. A ruler is mounted on the side to measure the motion of the mass relative to the ground.

If the ground moves with a very fast sinusoidal motion, it would be expected that the mass remains stationary and thus the ground sinusoidal motion can be measured directly as the relative mass-frame motion. The amplitude of the measurement would also be the ground's displacement amplitude and the seismometer would have a gain of 1. It is also seen that if the ground moves up impulsively, the mass moves down relative to the frame, represented by the ruler, so there is a phase shift of π (or 180°) in the measure of ground displacement. In general, a sinusoidal ground motion will produce a sinusoidal motion of the mass with the same frequency but with

a frequency-dependent phase shift and amplitude. With the ground moving very slowly, the mass would have time to follow the ground motion; in other words, there would be little relative motion, the gain would be low, and there would be less phase shift. At the resonance frequency, with a low damping, the mass could get a new push at the exact right time, so the mass would move with a larger and larger amplitude, thus the gain would be larger than 1. In order to get the exact motion of the seismometer mass relative to the ground motion including the phase shift, the equation for the swinging system must be solved.

If $u(t)$ is the ground's vertical motion and $z(t)$ the displacement of the mass relative to the ground, both positive upwards, there are two real forces acting on the mass m : the force of the deformed spring and the damping.

Spring force. $-kz$, negative since the spring opposes the mass displacement, k is the spring constant. This linear relation is strictly valid for small deformations only.

Damping force. $-d\dot{z}$, where d is the friction constant. Thus the damping force is proportional to the mass times the velocity and is negative since it also opposes the motion.

The acceleration of the mass relative to an inertial reference frame will be the sum of the acceleration $\ddot{z}$ with respect to the frame (or the ground) and the ground acceleration $\ddot{u}$.

Since the sum of forces must be equal to the mass times the acceleration, we have

$$-kz - d\dot{z} = m\ddot{z} + m\ddot{u} \quad (1)$$

For practical reasons, it is convenient to use ω_0 and the seismometer damping constant, $h = \frac{d}{2m\omega_0}$, instead of k and d , since both parameters are directly related to measurable quantities (see Havskov and Alguacil 2010). Equation 1 can then be written

$$\ddot{z} + 2h\omega_0\dot{z} + \omega_0^2 z = -\ddot{u} \quad (2)$$

This equation shows that the acceleration of the ground can be obtained by measuring the relative

displacement of the mass, z , and its time derivatives.

In the general case, there is no simple relationship between the sensor motion and the ground motion, and Eq. 2 will have to be solved so that the input and output signals can be related. This is most simply done assuming a harmonic ground motion

$$u(t) = U(\omega)e^{i\omega t} \quad (3)$$

where $U(\omega)$ is the complex amplitude and ω is the angular frequency. Equation 3 is written in complex form for simplicity of solving the equations and the real part represents the actual ground motion. Since a seismometer is assumed to represent a linear system, the seismometer mass motion is also a harmonic motion with the same frequency, and its amplitude is $Z(\omega)$

$$z(t) = Z(\omega)e^{i\omega t} \quad (4)$$

then

$$\begin{aligned} \ddot{u} &= -\omega^2 U(\omega)e^{i\omega t} \\ \dot{z} &= i\omega Z(\omega)e^{i\omega t} \\ \ddot{z} &= -\omega^2 Z(\omega)e^{i\omega t} \end{aligned} \quad (5)$$

Inserting in Eq. 2 and dividing by the common factor $e^{i\omega t}$, the relationship between the output and input complex amplitudes can be calculated as $T(\omega) = Z(\omega)/U(\omega)$, the so-called displacement frequency response function or transfer function:

$$T_d(\omega) = \frac{Z(\omega)}{U(\omega)} = \frac{\omega^2}{\omega_0^2 - \omega^2 + 2i\omega\omega_0 h} \quad (6)$$

From this expression, the amplitude displacement response $A_d(\omega)$ and phase response $\Phi_d(\omega)$ can be calculated as the modulus and phase of the complex amplitude response:

$$A_d(\omega) = |T_d(\omega)| = \frac{\omega^2}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4h^2\omega^2\omega_0^2}} \quad (7)$$

$$\Phi_d(\omega) = a \tan \left(\frac{\text{Im}(T_d(\omega))}{\text{Re}(T_d(\omega))} \right) = a \tan \left(\frac{-2h\omega\omega_0}{\omega_0^2 - \omega^2} \right) \quad (8)$$

and $T_d(\omega)$ can be written in polar form as

$$T_d(\omega) = A_d(\omega)e^{i\Phi_d(\omega)} \quad (9)$$

From Eq. 7, it can be seen what happens in the extreme cases. For high frequencies

$$A_d(\omega) \rightarrow 1 \quad (10)$$

This is a constant gain of one and the sensor behaves as a pure displacement sensor.

For low frequencies,

$$A_d(\omega) \rightarrow \frac{\omega^2}{\omega_0^2} \quad (11)$$

which is proportional to acceleration and the output from the sensor is proportional to acceleration. For a high damping,

$$A_d(\omega) \approx \frac{\omega}{2h\omega_0} \quad (12)$$

and the output is proportional to ground velocity; however, the gain is low since h is high.

Figure 2 shows the amplitude and phase response of a sensor with a natural period of 1 s and damping from 0.1 to 4. As it can be seen, a low damping ($h < 1$) results in a peak in the response function. If $h = 1$, the seismometer mass will return to its rest position in the least possible time without overshooting, and the seismometer is said to be critically damped. From the shape of the curve and Eq. 7, it is seen that the seismometer can be considered a second-order high-pass filter for ground displacement. Seismometers perform optimally at damping close to critical. The most common value to use is $h = 1/\sqrt{2} = 0.707$. Why exactly this value and not 0.6 or 0.8? In practice it does not make much difference, but the value 0.707 is a convenient value to use when describing the amplitude response function. Inserting $h = 0.707$ in Eq. 7, the value for $A_d(\omega_0) = 0.707$. This is the amplitude value

used to define the corner frequency of a filter or the -3 dB point. So using $h = 0.707$ means that the response can be described as a second-order high-pass Butterworth filter with a corner frequency of ω_0 . This filter has the flattest possible response in its passband.

When the damping increases above 1, the sensitivity decreases, as described in Eq. 7, and the response approaches that of a velocity sensor (mass motion is proportional to ground velocity). From Fig. 2, it can be seen that for $h = 4$, the response approaches a straight line indicating a pure velocity response within a limited frequency band.

The Velocity Transducer

Nearly all traditional seismometers use a velocity transducer to measure the motion of the mass (example in Fig. 3). The principle is to have a moving coil within a magnetic field. This can be implemented by having a fixed coil and a magnet that moves with the mass or a fixed magnet and the coil moving with the mass. The output from the coil is proportional to the velocity of the mass relative to the frame, and this kind of electromagnetic seismometer is therefore called a velocity transducer. Two new constants are brought into the system:

Generator constant G . This constant relates the velocity of the mass to the output of the coil. It has a unit of V/ms^{-1} . Typical values are in the range of 30–500 V/ms^{-1} .

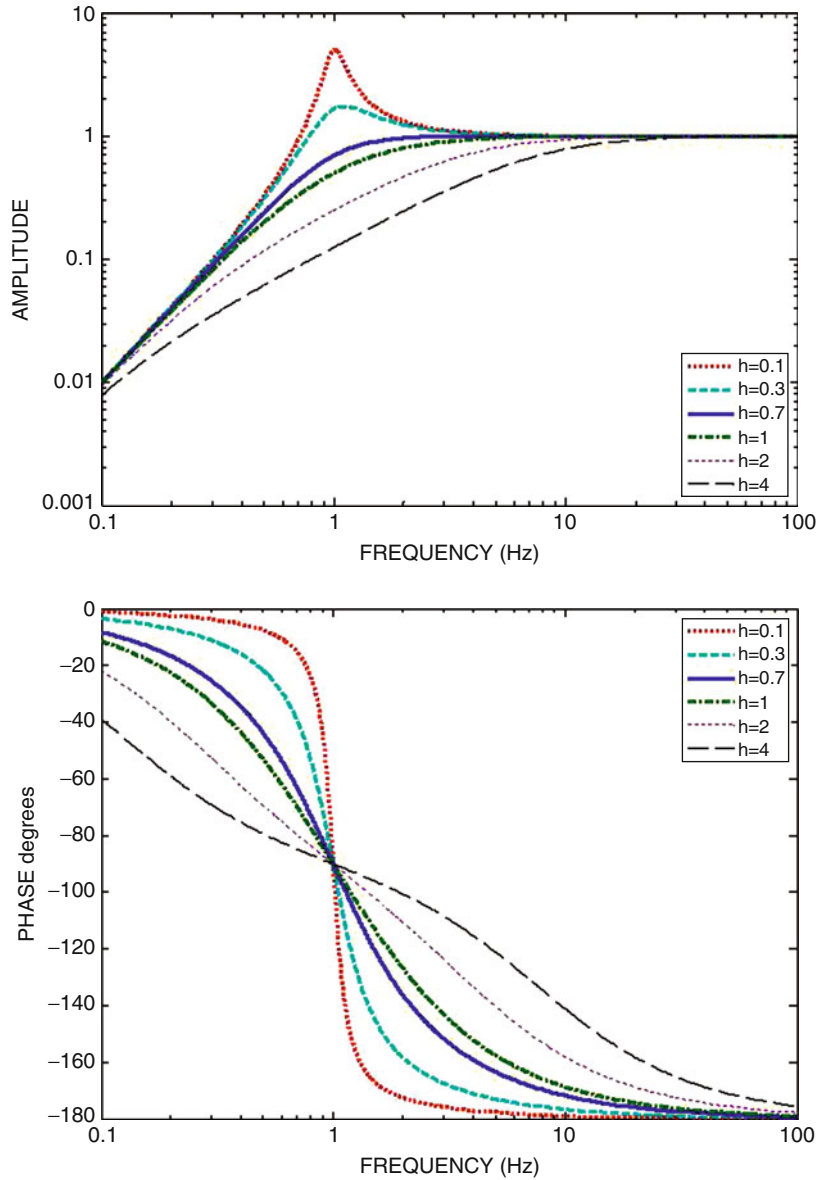
Generator coil resistance R_g . The resistance of the generator coil (also called signal coil) in Ohms (the coil is built as hundreds or thousands of turns of thin wire winding).

The signal coil makes it possible to damp the seismometer in a very simple way by loading the signal coil with a resistor. When a current is generated by the signal coil, it will oppose to the mass motion with a proportional magnetic force (see next section).

The frequency response function for the velocity transducer is different than for the mechanical sensor. With the velocity transducer, the observed output signal is now a voltage proportional to the mass-frame velocity $\dot{Z}(\omega) = i\omega Z(\omega)$ and G ,

Passive Seismometers,

Fig. 2 The amplitude and phase response functions for a seismometer with a natural frequency of 1 Hz. Curves for various levels of damping h are shown. Note that the phase shift goes toward 0 at low frequencies and toward 180° at high frequencies as qualitatively deduced



instead of $Z(\omega)$. The displacement response for the velocity sensor is then

$$\begin{aligned} T_d^v(\omega) &= \frac{\dot{Z}(\omega)}{U(\omega)} G = \frac{i\omega\omega^2 G}{\omega_0^2 - \omega^2 + i2\omega\omega_0 h} \\ &= \frac{i\omega^3 G}{\omega_0^2 - \omega^2 + i2\omega\omega_0 h} \end{aligned} \quad (13)$$

and it is seen that the only difference compared to the mechanical sensor is the factors G and $i\omega$.

For the response curve Eq. 13, the unit is $(\text{ms}^{-1}/\text{m})$ $(\text{V}/\text{ms}^{-1}) = \text{V}/\text{m}$. It is assumed in Eq. 13 that a positive velocity gives a positive voltage. Most often the response for the velocity transducer is shown for input velocity, which is Eq. 13 divided by $i\omega$:

$$T_v^v(\omega) = \frac{G\omega^2}{\omega_0^2 - \omega^2 + i2\omega\omega_0 h} \quad (14)$$

and the response looks like the displacement response for a mechanical sensor (Fig. 2). It is

Passive Seismometers,

Fig. 3 A model of an electromagnetic sensor. The coil resistance is R_g , the damping resistor is R , and the voltage output is V_{out} . The dashpot damping has been replaced by the damping from the coil moving in the magnetic field (Figure from Havskov and Alguacil 2010)

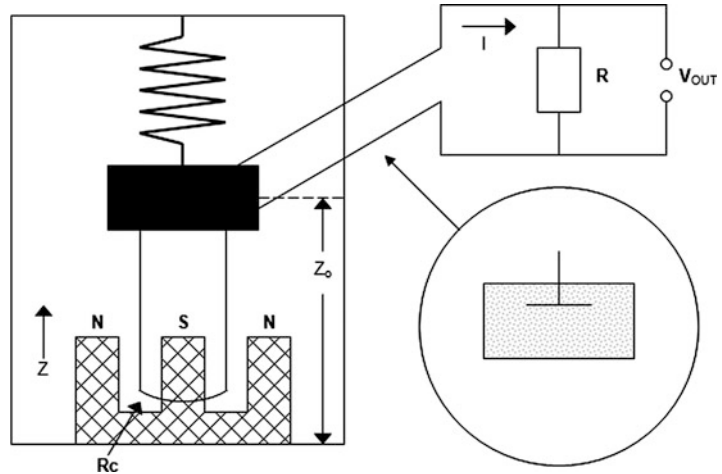
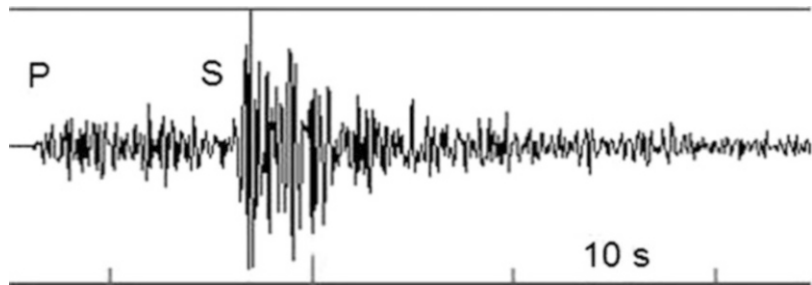
**Passive Seismometers,**

Fig. 4 Recording of a small earthquake from a 1 Hz passive sensor. The first signal recorded is the primary wave (P) and the second is the shear wave (S)



seen that the response to ground velocity is constant for $\omega > \omega_0$ and the sensor is therefore often called a velocity sensor.

In practice, the sensor output is always connected to an external resistor R_e (for damping control and because amplifiers have a finite input impedance). This forms a voltage divider. Thus the effective generator constant (or loaded generator constant) G_e becomes

$$G_e = G \cdot \frac{R_e}{R_e + R_g} \quad (15)$$

and this must be used in Eqs. 13 and 14 instead of G . All available passive seismometers are now using a velocity transducer and the most common way to use the is to digitize it and record it in a computer. The signal can then be plotted to show the so-called seismogram, an example in Fig. 4.

Other passive sensors, seldom used for earthquake recording, but quite common in structural dynamics monitoring, are piezoelectric accelerometers. An inertial mass is fixed on a piezoelectric material (normally in a multilayer arrangement), which has an elastic behavior. Due to the high elastic compliance, the resulting natural frequency of the system is also high and the strain is proportional to ground acceleration up to this frequency. Because of piezoelectricity, the system output is an electric charge proportional to the strain, so a charge amplifier (current integrator) is used to yield a voltage proportional to acceleration. In practice, the charge amplifier cannot perform a perfect integration up to DC ($\omega = 0$), so there is some low-frequency limit for the flat response to acceleration.

The piezoelectric properties of some materials are quite temperature dependent and this has to be somehow compensated for.

Damping

For a purely mechanical seismometer there is a damping h_m due to friction (mainly air friction and spring internal elastic dissipation) of the mechanical motion. In a sensor with electromagnetic transducer, a voltage E is induced in the coil proportional to velocity. If the signal coil is shunted with an external resistance, let R_T be the total circuit resistance, then a current $I = E/R_T$ (neglecting self-induction) will flow through the circuit. This will cause a force on the mass proportional to this current and in the sense opposed to its motion, as is given by Lenz law. This will introduce an additional electrical or electromagnetic damping h_e .

Thus the total damping is the sum of electromagnetic and mechanical contributions:

$$h = h_e + h_m \quad (16)$$

h_m is also called open-circuit damping, since this is the damping of the seismometer with no electrical connection. The mechanical damping, hereafter called open-circuit damping, cannot be changed and range from a very low value of 0.01 to 0.3 while the electrical damping can be regulated with the value of the external resistor to obtain a desired total damping. The electrical damping can be calculated as (Havskov and Alguacil 2010)

$$h_e = \frac{G^2}{2M\omega_0 R_T} \quad (17)$$

where M is the seismometer mass and R_T is the total resistance of the generator coil and the external damping resistor. Seismometer specifications often give the critical damping resistance CDR , which is the total resistance $CDR = R_T$ required to get a damping of 1.0.

From Eqs. 17 and 16, it is seen that if the total damping h_1 is known for one value of R_T , R_{T1} , the required resistance R_{T2} for another required total damping h_2 can be calculated as

$$R_{T2} = R_{T1} \frac{h_1 - h_m}{h_2 - h_m} \quad (18)$$

If the mechanical damping is low ($h_m \approx 0$), Eq. 18 can be written in terms of CDR ($h_2 = 1$, $CDR = R_{T1}$) as

$$R_{T2} = \frac{CDR}{h_2} \quad (19)$$

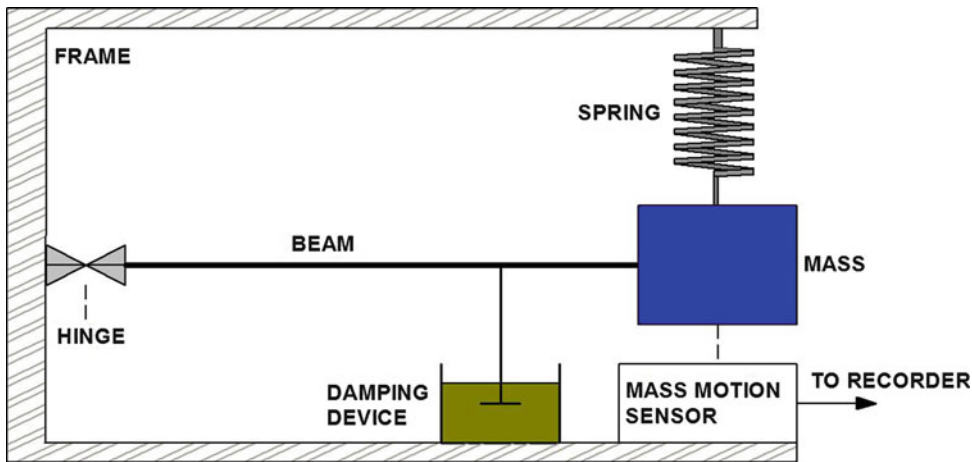
and thus, the desired total resistance for any required damping can easily be calculated from CDR .

As an example, consider the classical 1 Hz sensor, the Geotech S13 (Fig. 15). The coil resistance is 3,600 Ω and $CDR = 6,300 \Omega$. Since the open-circuit damping is low, it can be ignored. The total resistance to get a damping of 0.7 would then, from Eq. 19, be $R = 6,300/0.7 = 9,000 \Omega$, and the external damping resistor to connect would have a value of $9,000 - 3,600 = 5,400 \Omega$.

Construction of Passive Sensors

The mass-spring system of the vertical seismometer serves as a very useful model for understanding the basics of seismometry. However, in practical design, this system is too simple, since the mass can move in all directions as well as rotate. So, nearly all seismometers have some mechanical device which will restrict the motion to one translational axis. Figure 5 shows how this can be done in principle for a vertical seismometer.

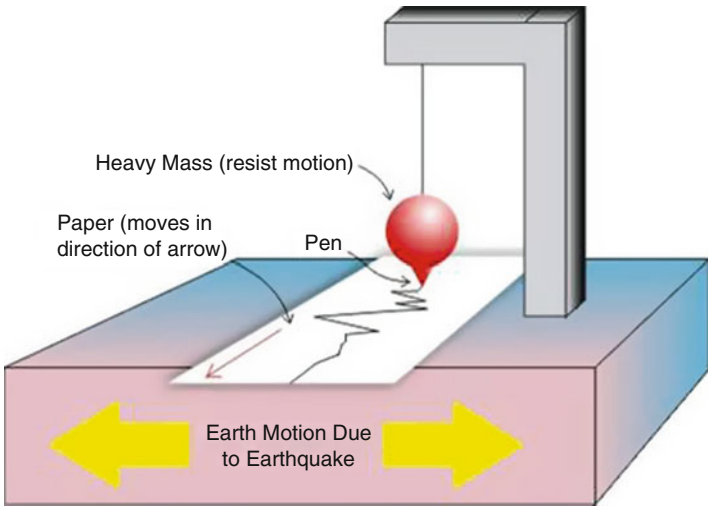
It can be seen that due to the hinged mass, the sensor is restricted to move vertically. The mass does not only move in the z direction but in a circular motion with the tangent to the vertical direction. However, for small displacements, the motion is sufficiently linear. The above pendulum arrangement is in principle the most common way to restrict motion and can also be used for horizontal seismometers. Pendulums are also sensitive to angular motion in seismic waves, which normally is so small that it has no practical importance.



Passive Seismometers, Fig. 5 A vertical mass-spring seismometer where the horizontal motion has been restricted by a horizontal hinged rod. Usually, the hinge

is a thin flexible leaf to avoid friction (Figure from Havskov and Alguacil 2010)

Passive Seismometers, Fig. 6 Horizontal seismometer made with a pendulum (From <http://www.azosensors.com/>)



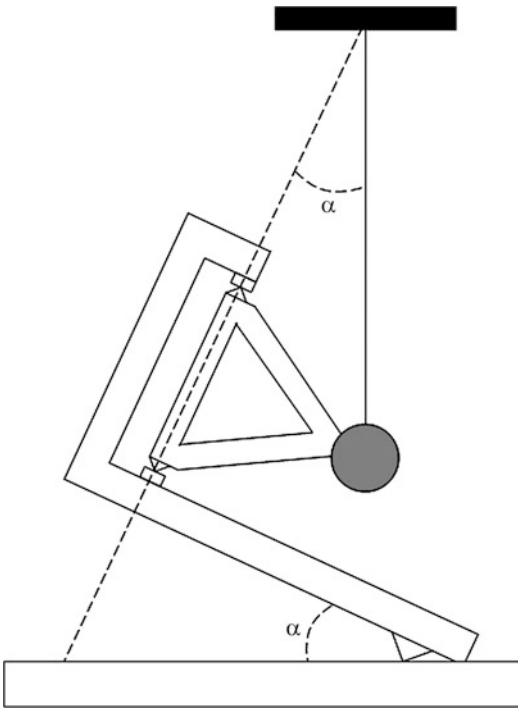
So far only a vertical seismometer has been described; however, the ground motion must be measured in all three directions, x , y , and z . Normally z is positive up, x is to the east, and y to the north. A horizontal seismometer is in its simplest form a pendulum (see Fig. 6).

For a small-size mass m compared to the length L of the string, the natural frequency

$$\omega_0 = \sqrt{g/L} \quad (20)$$

where g is the gravitational constant. For small translational ground motions, the equation of motion is identical to Eq. 1 with z replaced with the angle of rotation. Note that ω_0 is independent of the mass.

It is relatively easy to make sensors with a natural frequency down to 1 Hz (so-called short-period (SP) seismometers), but in seismology it is desirable to measure much lower frequencies using so-called long-period (LP)



Passive Seismometers, Fig. 7 Principle of the garden-gate pendulum. The tilt angle is exaggerated. A string pendulum must have length L to have the same period, so at a small angle α , L becomes very large (Figure from Havskov and Alguacil 2010)

seismometers, and a good station with a passive sensor should be able to measure down to ideally 0.01 Hz. For the vertical seismometer, this would require a very soft spring combined with a heavy mass, which is not possible in practice. For the pendulum, L would have to be very long. For a 1 Hz seismometer, it would be 9.8 m. So there are various ways of making the natural frequency smaller without using a very large design; however, in practice it is hard to achieve less than 0.03 Hz with a passive seismometer, while active seismometers can go down to 0.003 Hz.

For a horizontal seismometer, the simplest solution is the “garden-gate” pendulum (Fig. 7). The mass moves in a nearly horizontal plane around a nearly vertical axis. The restoring force is now $g \sin(\alpha)$ where α is the angle between the vertical and the rotation axis, so the natural frequency becomes

$$\omega_0 = \sqrt{g \sin(\alpha)/L} \quad (21)$$

where L is the vertical distance from the mass to the point where the rotation axis intersects the vertical above the mass (see Fig. 7).

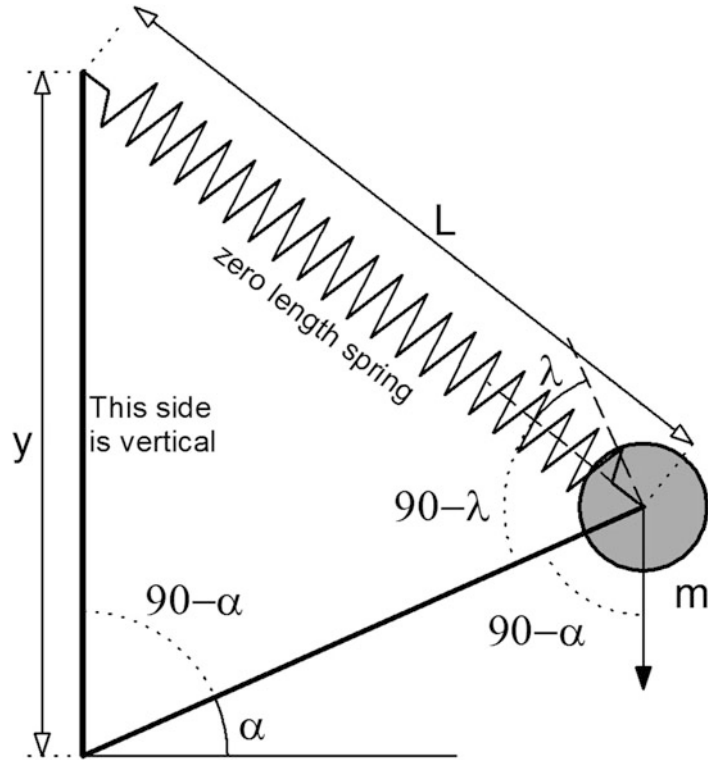
To obtain a natural frequency of 0.05 Hz with a pendulum length of 20 cm will require a tilt of 0.1° . This is close to the lowest stable period that has been obtained in practice with these instruments. Making the angle smaller makes the instrument very sensitive to small tilt changes. The “garden gate” was one of the earliest designs for long-period horizontal seismometers. It is still in use in a few places but no longer produced. Until the new broadband sensors were installed in the GSN (Global Seismographic Network), this kind of sensors was used in the WWSSN (World-Wide Standardized Seismograph Network) (see, e.g., Peterson and Orsini 1976).

The *astatic* spring geometry for vertical seismometers was invented by LaCoste (1934). The principle of the sensor is shown in Fig. 8. The sensor uses a “zero-length spring” which is designed such that the force $F = k \cdot L$, where L is the *total* length of the spring. Normal springs used in seismometers do not behave as zero-length springs, since $F = k \cdot \Delta L$, where ΔL is the change in length relative to the unstressed length of the spring. However, it is possible to construct a “zero-length spring” by, e.g., twisting the wire as it is wound into a spring. The physical setup is to have a mass on a beam and supported by the spring so that the mass is free to pivot around the lower left hand point.

This system can be constructed to have an infinite free period, which means that the vertical restoring force must cancel the gravity force at any mass position. Therefore, if the mass is at equilibrium at one angle, it will also be at equilibrium at another angle, which is similar to what was obtained for the garden-gate horizontal seismometer. Qualitatively, what happens is that if, e.g., the mass moves up, the spring force lessens, but it turns out the force from gravity in direction of mass motion is reduced by the exact same amount, due to the change in angle as will be shown below.

Passive Seismometers,

Fig. 8 The principle behind the LaCoste suspension. The mass m is sitting on a hinge, which has an angle α with the horizontal and suspended by a spring of length L (Figure from Havskov and Alguacil 2010)



The gravity force F_g acting on the mass in direction of rotation can be written as

$$F_g = mg \cos(\alpha) \quad (22)$$

while the spring restoring force F_s acting in the opposite sense is

$$F_s = kL \cos(\lambda) \quad (23)$$

λ can be replaced by α using the law of sines:

$$\begin{aligned} \frac{L}{\sin(90 - \alpha)} &= \frac{y}{\sin(90 - \lambda)} \text{ or } \cos(\lambda) \\ &= \frac{y}{L} \cos(\alpha) \end{aligned} \quad (24)$$

Equating F_s and F_g and including the expression for $\cos(\lambda)$ gives

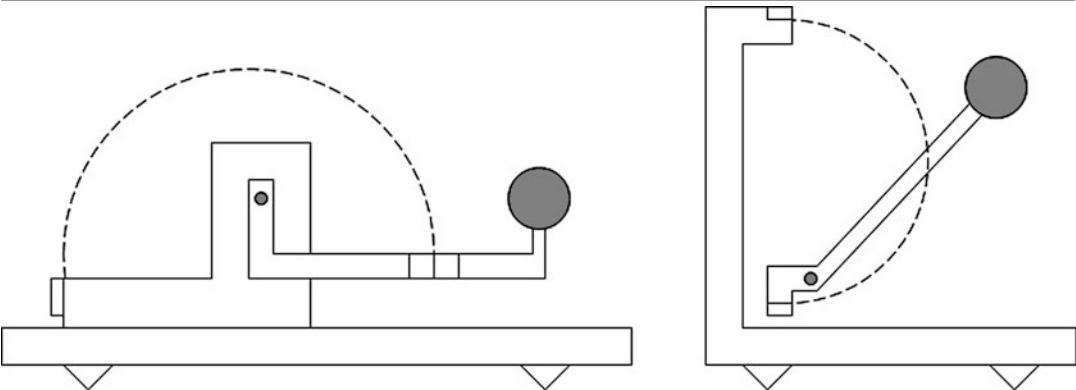
$$ky = mg \text{ or } y = mg/k \quad (25)$$

As long as this condition holds, the total force is zero independent of the angle. As with the garden-

gate seismometer, this will not work in practice and, by inclining the vertical axis slightly, any desired period within mechanical stabilization limits can be achieved. In practice, it is difficult to use free periods larger than 20–30 s.

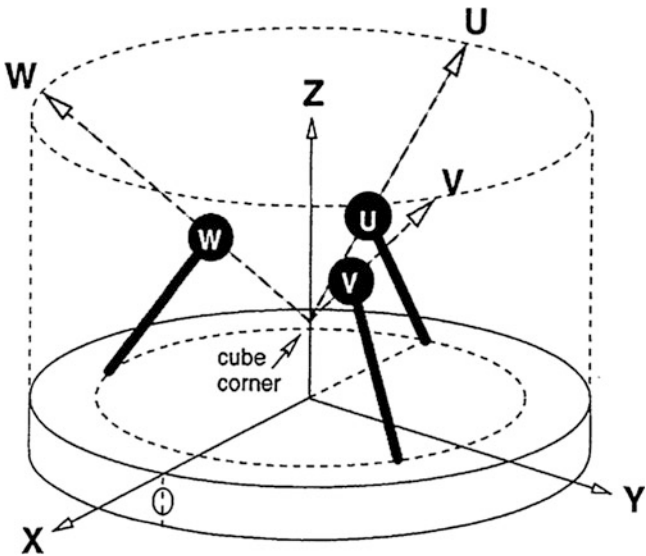
The astatic leaf-spring suspension (Wielandt and Streckeisen 1982) is comparable to the LaCoste suspension, but simpler to make (Fig. 9). The delicate equilibrium of forces in astatic suspensions makes them sensitive to external disturbances, so they are difficult to operate without a stabilizing feedback system.

The LaCoste pendulums can be made to operate as vertical seismometers, as has been the main goal or as sensors with an oblique axis. Sensors with an oblique axis are used to obtain both vertical and horizontal motions. The normal arrangement for a 3-component sensor is to have 3 sensors oriented in the Z , N , and E directions. Since horizontal and vertical seismometers differ in their construction, it is difficult to make them equal. A three-component sensor can be constructed by using three identical sensors whose axes U , V , and W are inclined against the vertical like the edge of



Passive Seismometers, Fig. 9 Leaf-spring astatic suspensions. The figure to the *left* shows a vertical seismometer and to the *right* an oblique axis seismometer (Figure from <http://jclahr.com/science/psn/wielandt/node15.html>)

Passive Seismometers, Fig. 10 The triaxial geometry of the STS-2 seismometer. The oblique components are *W*, *V*, and *U* (Figure from Havskov and Alguacil 2010)



a cube standing on its corner (Fig. 10). Each sensor is made with an *astatic* leaf-spring suspension (Fig. 9). The angle of inclination is $\tan^{-1}\sqrt{2} = 54.7$ degrees, which makes it possible to electronically recombine the oblique components to *X*, *Y*, and *Z* simply

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \frac{1}{\sqrt{6}} \cdot \begin{pmatrix} -2 & 1 & 1 \\ 0 & \sqrt{3} & -\sqrt{3} \\ \sqrt{2} & \sqrt{2} & \sqrt{2} \end{pmatrix} \begin{pmatrix} U \\ V \\ W \end{pmatrix} \tag{26}$$

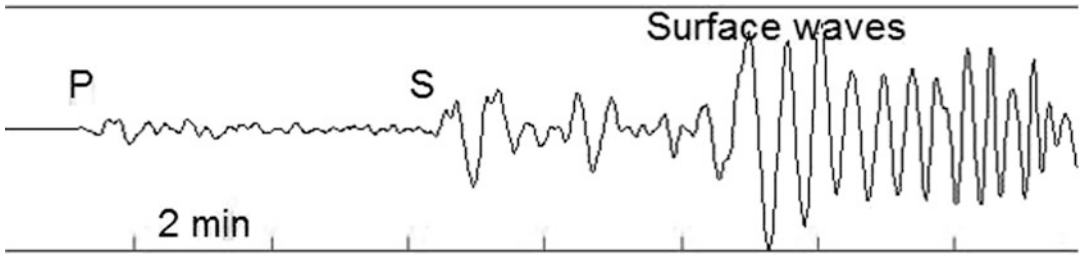
This arrangement is now the heart of some modern active seismic sensors with a natural

frequency below 0.1 Hz. Figure 11 shows an example of a long-period recording.

For more details on passive seismic sensors, see also the New Manual of Seismological Observatory Practice (NMSOP-2), Bormann (2012), Aki and Richards (1980), and the old Manual of Seismological Observatory Practice, Wilmore (1979).

Sensitivity of Passive Sensors

The output from a passive sensor V_{out} is, above its natural frequency,

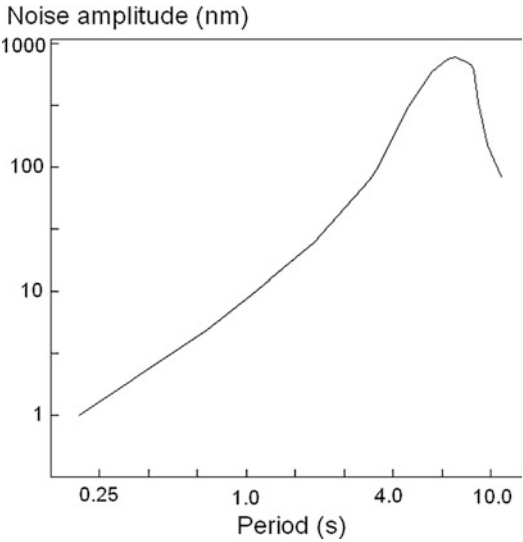


Passive Seismometers, Fig. 11 Example of recording an earthquake using a long-period sensor. Note the low frequency of the signal. The distance from the location of

the earthquake (epicenter) to the recording station is 3,500 km. The harmonic waves at the end of the record are surface waves traveling along the Earth’s surface

$$V_{\text{out}} = GV \tag{27}$$

where V is the ground velocity. At a low-noise site, the ground velocity amplitude is typically 10 nm/s and a typical value of the generator constant is 100 V/ms⁻¹ so the output would be 1 μV. This is the ground seismic signal. The sensor itself will, since it has no active electronics, usually produce less than 10 times this in electrical noise from the Brownian motion of the mass and the current flowing through the coil. So in general, passive seismic sensors can easily be made sensitive enough for most sites. The low-level output must be amplified to be recorded. For analog recording, this means a sensitive amplifier (usually with a noise level of 0.1 μV), while modern recording systems use digital recording. Many digital recorders will have a noise level also of 0.1 microvolt, meaning the number 1 (one count) corresponds to 0.1 μV and the seismic noise signal would then be recorded with 10 counts of resolution. Active sensors usually have higher electrical outputs, so not all digital recorders have enough sensitivity for passive sensors. For frequencies f below the natural frequency, the response decreases proportional to f^2 . So, for a 1 Hz sensor, the sensitivity is down by a factor of 100 at 0.1 Hz. Since the Earth’s natural background noise also increases for frequencies below 1 Hz until about 0.2 Hz (Fig. 12), a 1 Hz seismometer will in practice have sufficient sensitivity for most seismological applications down to 0.1 Hz.



Passive Seismometers, Fig. 12 Typical amplitude of the microseismic background noise as a function of period(s) (The data for the plot are taken from Brune and Oliver 1959)

Examples of Some Passive Sensors

Passive sensors are today only sold for sensors with a natural frequency at 1 Hz and higher. For sensors with a lower natural frequency, active sensors are now both smaller, better, and cheaper than the old passive sensors and therefore no longer constructed. For earthquake seismology, it is particularly sensors with natural frequencies of 1, 2 and 4.5 Hz that are used, with the 1 Hz sensor the natural choice and there are still hundreds of seismic stations with 1 Hz sensors.

Passive Seismometers,

Fig. 13 Typical construction of a geophone. Note the leaf-spring suspension. The magnet is fixed to the case and the coil moves with the mass <http://vibration.desy.de/equipment/geophones/>

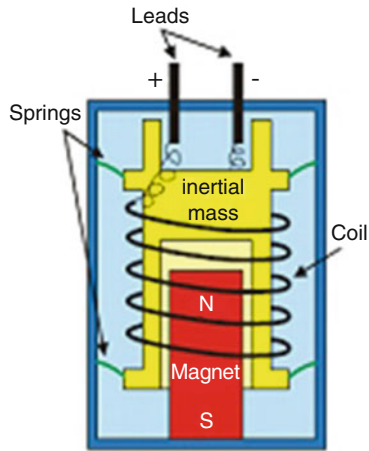
**Passive Seismometers,**

Fig. 14 The Geospace GSX recorder (box below) connected to a small battery and a Geospace three-component 2 Hz geophone (green cylinder) (Figure from www.geospace.com)



P

Sensors with a frequency up to 4.5 Hz and above are constructed in the thousands and mostly used for exploration purposes and they are usually called geophones, in contrast to the 1 Hz sensor which is usually called a seismometer.

Geophone: An example is seen in Fig. 13. It is small in size (about 3×5 cm) and only costs about \$100, since it is produced in large numbers for seismic exploration. It typically has a mass of 20 g and the generator constant is often around 30 V/ms^{-1} .

The sensor is very simple and robust to use. Due to its low sensitivity, it requires more amplification than the standard sensors.

It has traditionally not been used for earthquake recording as much as the 1 Hz sensor; however, with more sensitive and higher dynamic range recorders, it is now possible to use it directly without any special filtering and obtain good recordings down to 0.3 Hz by post processing.

A modern 2 Hz, three-component geophone (Fig. 14). This is a very compact 2 Hz sensor used both for earthquake recording and seismic exploration. The figure shows how compact it is possible to make a complete seismograph.

A classical 1 Hz seismometer, the S13 from Geotech (see Fig. 15): This 1 Hz seismometer is

one of the most sensitive passive 1 Hz seismometers produced and it has been the standard by which other 1 Hz seismometers are measured. It is suspended by both leaf and helical springs



Passive Seismometers, Fig. 15 Geotech S13 seismometer (The picture is from www.bgr.bund.de/EN/Themen/Seismologie/Seismologie/Seismometer_Stationen/stationen_node.html)

Passive Seismometers, Fig. 16 The ranger 1 Hz seismometer from Kinemetrics (Figure from www.kinemetrics.com)



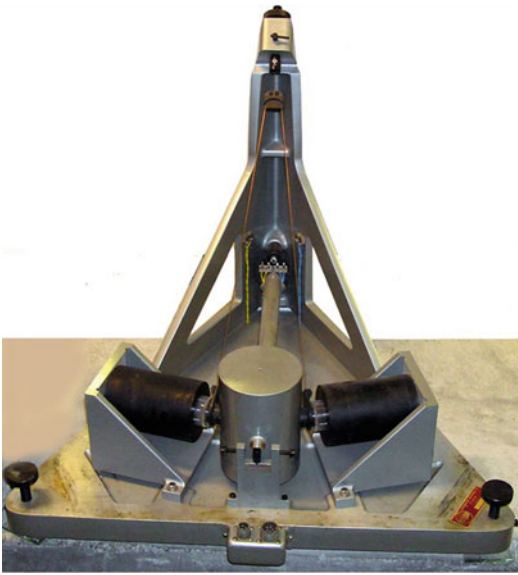
in such a way that by an internal adjustment, the sensor can be used both as a vertical and horizontal seismometer. It is still produced.

Another classical 1 Hz seismometer from Kinemetrics (Fig. 16). This seismometer was designed for the Moon and then later produced in a terrestrial version and is still produced. It can work both as a horizontal and vertical seismometer by an external adjustment.

A long-period horizontal seismometer (Fig. 17). This seismometer was used by the WWSSN and was thus produced in large numbers. The longest practical free period was 30 s and it was mostly used with a 25 s period. At very stable sites, periods larger than 30 s have been used.

Summary

Passive seismic sensors function without internal electronics, so no power supply is required, and they have been in operation for more than 100 years. All modern passive sensors are of the so-called velocity types, meaning that the voltage output is proportional to the ground velocity for frequencies above the sensor's natural frequency. Passive LP sensors are no longer sold. Compared to active sensors, passive SP sensors are low cost, very stable, and very sensitive but cannot easily resolve signals for frequencies below 1 Hz, although a good 1 Hz sensor can provide useful signals down to 0.1 Hz. Passive sensors (both 1 and



Passive Seismometers, Fig. 17 A typical horizontal long-period seismometer which was made by Sprengnether (no longer exists). The two *black* cylinders seen at each side are the magnets and the coils are moving inside with the mass in the *center*. Leveling is done with the three screws at the corners of the base plate. Once the horizontal leveling is done (front screws), the back screw is used to adjust the period. The base plate side length is 63 cm and the weight 45 kg

4.5 Hz) are still a good choice for recording small earthquakes, while the higher-frequency geophones are mostly used for seismic exploration.

Cross-References

- [Principles of Broadband Seismometry](#)
- [Seismic Accelerometers](#)
- [Seismic Instrument Response, Correction for](#)
- [Sensors, Calibration of](#)

References

- Aki K, Richards PG (1980) Chapter 10: Principles of seismometry. In: Quantitative seismology – theory and methods, vol 1. W. H. Freeman, San Francisco, pp 477–524
- Bormann P (ed) (2012) New manual of seismological observatory practice (NMSOP-2). IASPEI, GFZ German Research Centre for Geosciences, Potsdam; nmsop.gfz-potsdam.de

- Brune JN, Oliver J (1959) The seismic noise at the earth's surface. *Bull Seismol Soc Am* 49:349–353
- Havskov J, Alguacil G (2010) Instrumentation in earthquake seismology. Springer, Dordrecht, 358 pp
- LaCoste LJB (1934) A new type long period seismograph. *Physics* 5:178–180
- Peterson J, Orsini NA (1976) Seismic research observatories: upgrading the world wide seismic data network. *Eos Trans Am Geophys Union* 57:548–546
- Wielandt E, Streckeisen G (1982) The leaf-spring seismometer: design and performance. *Bull Seismol Soc Am* 72:2349–2367
- Wilmore PL (ed) (1979) Manual of seismological observatory practice, report SE-20. World Data Center A for Solid Earth Geophysics/US Department of Commerce/NOAA, Boulder

Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers

Yunbyeong Chae¹, Baiping Dong², James M. Ricles³ and Richard Sause⁴

¹Department of Civil and Environmental Engineering, Old Dominion University, Norfolk, VA, USA

²Department of Civil and Environmental Engineering, Lehigh University, Bethlehem, PA, USA

³Bruce G. Johnston Professor of Structural Engineering, Department of Civil and Environmental Engineering, Lehigh University, Bethlehem, PA, USA

⁴Joseph T. Stuart Professor of Structural Engineering, Department of Civil and Environmental Engineering, Lehigh University, Bethlehem, PA, USA

Synonyms

Magneto-rheological damper; Performance-based seismic design; Simplified design procedure

Introduction

It is well-known that supplemental damping devices increase the energy dissipation capacity

of structures, reducing the seismic demand on the primary structure (Constantinou et al. 1998; Soong and Dargush 1997). A structural system with supplemental dampers is often represented by an equivalent linear system. Kwan and Billington (2003) derived optimal equations for the equivalent period and damping ratio of SDOF systems with various nonlinear hysteresis loops based on time-history analysis and regression analysis. Symans and Constantinou (1998) studied the dynamic behavior of SDOF systems with linear or nonlinear viscous fluid dampers and derived an equation for the equivalent damping ratio of the nonlinear viscous fluid damper. Ramirez et al. (2002) proposed a simplified method to estimate displacement, velocity, and acceleration for yielding structures with linear or nonlinear viscous dampers. Lin and Chopra (2003) investigated the behavior of SDOF systems with a diagonal brace in series with a nonlinear viscous damper by transforming the system to an equivalent linear Kelvin model.

Fan (1998) investigated the behavior of nonductile reinforced concrete frame buildings with viscoelastic dampers. He derived an equivalent elastic-viscous model based on the complex stiffness and energy dissipation of the viscoelastic system and proposed a simplified design procedure for a structure with viscoelastic dampers. Lee et al. (2005, 2009) applied this method to structures with elastomeric dampers and validated the simplified design procedure by comparing the design demand with the results from nonlinear time-history analysis.

In this paper, a systematic procedure for the design of structures with MR dampers, referred to as the *Simplified Design Procedure* (SDP), is developed. The procedure is similar to that developed by Lee et al. (2005, 2009), but with modifications to account for the characteristics of the MR dampers. A quasistatic MR damper model for determining the loss factor and the effective stiffness of an MR damper is introduced and incorporated into the procedure to calculate the design demand for the structure with MR dampers. The procedure is evaluated by comparing the predicted design demand to the seismic

response determined from nonlinear time-history analysis.

Simplified Design Procedure (SDP)

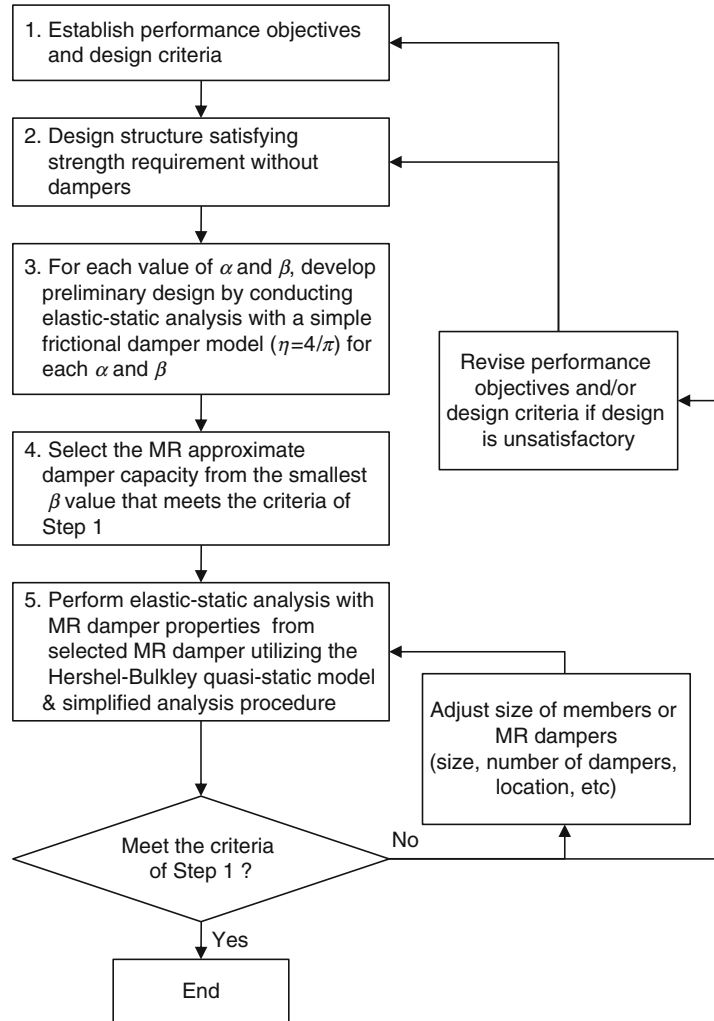
In the SDP developed by Lee et al. (2005, 2009), the supplemental damper properties are represented by β , which is the ratio of the damper stiffness per story in the global direction to the lateral load resisting frame story stiffness, k_0 , without dampers and braces of the structural system. The structural system with dampers is converted into a linear elastic system characterized by the initial stiffness of the structure, α (the ratio of brace stiffness per story in the global direction to the lateral load resisting frame story stiffness k_0), β , and η . α , β , and η may vary among the stories of the structure. By conducting an elastic-static analysis with the RSA method, the design demand for the structure is determined.

Since the loss factor of an MR damper depends on the displacement of the structure, as can be shown below in Eq. 8, the SDP for an elastomeric damper developed by Lee et al. (2005, 2009) needs to be modified for structures with MR dampers. The loss factor η is associated with the energy dissipation of the damper over a cycle. For the purpose of determining the energy dissipation over a cycle of displacement, the properties of the MR damper are assumed to remain constant. The linearization and energy dissipation of an MR damper are discussed later.

Figure 1 summarizes the SDP for structures with MR dampers. In Step 1, the seismic performance objectives and associated design criteria are established for the design of the structure. In Step 2, the structure is designed without MR dampers in accordance with the design code selected in Step 1 to satisfy the strength requirement for the members in the structure. In Step 3, the MR dampers are incorporated into the design of the structure to satisfy the specified performance objectives. The design demand for the structure is estimated for a range of selected values for α , β , and a constant loss factor $\eta = 4/\pi$

Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers,

Fig. 1 Generalized simplified design procedure (SDP) for structures with MR dampers



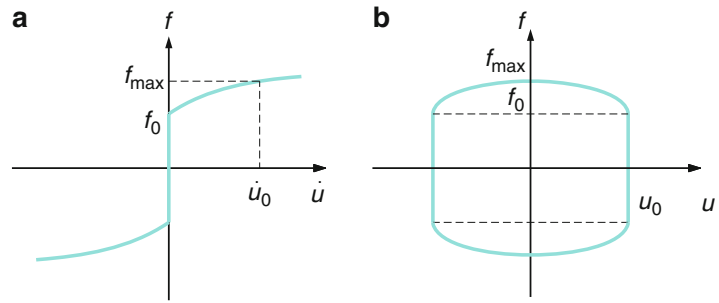
using a *simplified analysis procedure*, discussed later. In the simplified analysis procedure, the MR damper behavior is based on the simple frictional MR damper model (Chae 2011). The required MR damper sizes are then selected in Step 4 based on the smallest β value that meets the design criteria and performance objectives in Step 1. Since the simple frictional damper model does not account for the velocity dependent behavior of an MR damper, a more accurate determination of the design demand is determined in Step 5 using a more sophisticated MR damper model (i.e., Hershel-Bulkley model) in the simplified analysis procedure. The design is then revised with final member sizes and the MR

damper sizes are selected (location, number, force capacity, etc.). If the performance objectives cannot be met in an economical manner, then the performance objectives and/or structural system design need to be revised as indicated in Fig. 1.

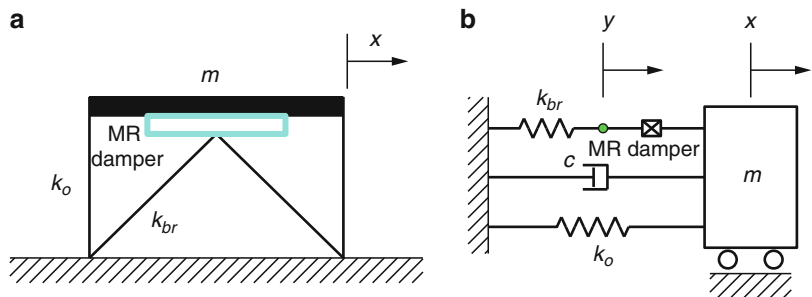
Equivalent Linear System for an MR Damper

The SDP requires that the structure with the nonlinear MR dampers be linearized. In order to linearize the system for estimating the response of structures with MR dampers, the

Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers, Fig. 2 Hershel-Bulkley visco-plasticity MR damper model: (a) force-velocity relationship; (b) force-displacement relationship



Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers, Fig. 3 SDOF system: (a) schematic of equivalent SDOF system with MR damper and brace; (b) mechanical model



Hershel-Bulkley quasi-static MR damper model is used. Figure 2 shows the force-velocity and force-displacement relationships for the model where the damper force f is given as

$$f = \text{sign}(\dot{u})\{f_0 + C|\dot{u}|^n\} \quad (1)$$

In Eq. 1 u is the damper displacement relative to the initial position of the damper and $\dot{u}$ is the damper velocity. f_0 is the frictional force. C and n are the coefficients of the nonlinear dashpot. Suppose that the MR damper is subjected to a harmonic displacement motion

$$u(t) = u_0 \sin(\omega t) \quad (2)$$

where u_0 is the amplitude of displacement and ω is the excitation frequency of the damper. The energy dissipated by the damper over one cycle of the harmonic motion is equal to

$$E_{MRD} = \int_0^{2\pi/\omega} f(t)\dot{u}(t)dt = 4f_0u_0 + 2^{n+2}C\gamma(n)u_0^{1+n}\omega^n \quad (3)$$

where

$$\gamma(n) = \frac{\Gamma^2\left(1 + \frac{n}{2}\right)}{\Gamma(2 + n)} \quad (4)$$

In Eq. 4, $\Gamma()$ is the gamma function (Soong and Dargush 1997). In general, diagonal bracing is installed in the building in series with the dampers. Therefore, the energy dissipation of an MR damper needs to be studied considering the stiffness of the diagonal bracing. Figure 3 shows an SDOF system with an MR damper and diagonal bracing. Under the harmonic motion $x = x_0 \sin(\omega t)$, the maximum damper displacement, u_{d0} , and velocity, $\dot{u}_{d0}$, of the MR damper occurs when the displacement x and velocity $\dot{x}$ are a maximum, respectively, where u_{d0} and $\dot{u}_{d0}$ can be calculated as (Chae 2011)

$$u_{d0} = x_0 - f_0/k_{br} \quad (5a)$$

$$\dot{u}_{d0} = \dot{x}_0 = x_0\omega \quad (5b)$$

In Eq. 5a k_{br} is the stiffness of the diagonal bracing. Substitution of Eq. 5a into Eq. 3 results

in the expression for the energy dissipation of the MR damper for a SDOF system

$$E_{MRD} = 4f_0 u_{d0} + 2^{n+2} C \gamma(n) u_{d0}^{n+1} \omega^n \quad (6)$$

The strain energy of the MR damper, E_S , is calculated from the equivalent stiffness of the damper and the maximum damper displacement:

$$E_S = \frac{1}{2} k_{eq} u_{d0}^2 \quad (7)$$

where k_{eq} is the equivalent stiffness of the MR damper, defined as $k_{eq} = f_0 / u_{d0}$ and based on the secant stiffness from the force-displacement relationship of the Hershel-Bulkley model. The loss factor η of the MR damper by definition is

$$\eta = \frac{1}{2\pi} \frac{E_{MRD}}{E_S} = \frac{4\{f_0 + C\gamma(n)(2u_{d0}\omega)^n\}}{\pi k_{eq} u_{d0}} \quad (8)$$

Simplified Analysis Procedure

The simplified analysis procedure provides an elastic-static method for calculation of the design demand of an MDOF system with MR dampers. The simplified analysis procedure utilizing the response spectrum analysis (RSA) method is summarized in Figs. 4 and 5.

In the simplified analysis procedure, the maximum structural displacements are determined by the well-known equal displacement rule. In the equal displacement rule, the maximum displacement of the nonlinear structure, whose lateral stiffness is based on its initial tangent stiffness, is assumed to be equal to that of a linear structure. The equal displacement rule is only applicable to structures that lie in the low-frequency and medium-frequency spectral regions (Newmark and Hall 1973).

In order to obtain the equivalent period for an MDOF structure with MR dampers, the combined stiffness of the MR dampers and diagonal

bracing needs to be added to the stiffness of the structure. Thus, the global effective stiffness of the MDOF system is given as

$$\mathbf{K}_{\text{eff}} = \mathbf{K}_0 + \mathbf{K}_{\text{brsystem}} \quad (9)$$

where $\mathbf{K}_0$ is the stiffness of the structure without diagonal braces and MR dampers, and $\mathbf{K}_{\text{brsystem}}$ is the stiffness associated with the braces and MR dampers. The structure is assumed to have N DOF, thus the dimension of $\mathbf{K}_{\text{eff}}$ is $N \times N$. The combined stiffness K_{brsystem}^i of the diagonal bracing and MR damper at the i -th MR damper location is

$$K_{\text{brsystem}}^i = \frac{k_{br}^i k_{eq}^i}{k_{br}^i + k_{eq}^i} \quad (10)$$

where k_{br}^i and k_{eq}^i are the horizontal stiffness of the diagonal bracing and the MR damper associated with the i th MR damper, respectively. k_{eq}^i can be calculated utilizing the secant stiffness method as noted above. The individual combined stiffnesses based on Eq. 10 are appropriately assembled to form $\mathbf{K}_{\text{brsystem}}$. The effective periods and mode shapes of the structure are then obtained by performing an eigenvalue analysis of the structure, considering the seismic mass of the structure.

The equivalent damping ratio ξ_{eq} of an MDOF system is determined using the lateral force energy method proposed by Sause et al. (1994), where

$$\xi_{eq} = \frac{1}{2} \frac{\sum_{i=1}^L \eta_i F_d^i u_d^i}{\mathbf{F}^T \mathbf{x}_0} + \xi_{in} \quad (11)$$

In Eq. 11 η_i and u_d^i are the loss factor and maximum damper displacement of the i th MR damper, respectively, and L is the number of MR dampers. Since the damper displacement is unique for each MR damper, the loss factor of each MR damper, which is a function of damper displacement, is also unique for each damper. For each damper η_i is obtained from Eq. 8. ξ_{in} in

Given:

- MR damper properties: f_0^i , C_i , n_i (i : index for MR damper location)
- Structural properties: $\mathbf{M}$, $\mathbf{K}_0$, k_{br}^i , ξ_{in}^j (inherent damping ratio of the j -th mode)

Step 1. Assume $\mathbf{x}_0$ and set $\omega_{eff}^1 = \omega_n^1$ (fundamental frequency of structure without MR dampers)

Step 2. Determine maximum damper displacements

$$u_{d0}^i = u_0^i - f_0^i / k_{br}^i$$

u_0^i : maximum deformation of damper and bracing of the i^{th} MR damper

Step 3. Calculate equivalent stiffness of each MR damper

$$k_{eq}^i = f_0^i / u_{d0}^i$$

Step 4. Determine $K_{brsystem}^i = \frac{k_{br}^i k_{eq}^i}{k_{br}^i + k_{eq}^i}$ for each MR damper and update $\mathbf{K}_{eff}$

Step 5. Update modal frequency ω_{eff}^j and modal vector ϕ_j ($j = 1, \dots, N$)

$$\omega_{eff}^j, \phi_j \leftarrow \text{eig}(\mathbf{K}_{eff}, \mathbf{M})$$

where $\mathbf{M}$ is the mass matrix of the structure

Step 6. Calculate loss factor of each MR damper

$$\eta_i = \frac{4\{f_0^i + C_i \gamma(n_i)(2u_{d0}^i \omega_{eff}^1)^{n_i}\}}{\pi k_{eq}^i u_{d0}^i}$$

where ω_{eff}^1 is the fundamental modal frequency

Step 7. Perform modal analysis described in Figure 5

Step 8. Apply the modal combination rule (e.g., SRSS or CQC) to get the final displacement $\mathbf{x}_0$ and velocity of each MR damper $\dot{u}_{d0}^i$

$$\mathbf{x}_0 = \text{function of } (\mathbf{x}_0^1, \dots, \mathbf{x}_0^N),$$

$$\dot{u}_{d0}^i = \text{function of } (\dot{u}_{d0}^{i1}, \dots, \dot{u}_{d0}^{iN})$$

Step 9. Repeat **Step 2** ~ **Step 8** until $\mathbf{x}_0$ convergence is achieved.

Step 10. Calculate maximum damper force for each MR damper

$$f_{max}^i = f_0 + C(\dot{u}_{d0}^i)^n$$

Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers,
Fig. 4 Simplified analysis procedure used to design

MDOF structures with passive MR dampers utilizing response spectrum analysis (RSA) method

Eq. 11 is the inherent damping ratio and $\mathbf{x}_0$ is the vector of the displacements of the structure that develop under the lateral force $\mathbf{F}$. The individual damper force F_d^i and the lateral force vector $\mathbf{F}$ are defined as

$$F_d^i = k_{eq}^i u_d^i, \quad \mathbf{F} = \mathbf{K}_{eff} \mathbf{x}_0 \quad (12)$$

In the simplified analysis procedure using the RSA, the relationships in Eq. 12 are substituted into Eq. 11, and the inherent damping ξ_{in} and $\mathbf{x}_0$

from each mode are considered, as indicated in Substep 3 in Fig. 5.

Performance-Based Design of a Three-Story Building with MR Dampers

Prototype Building Structure

Based on the proposed SDP, a three-story building with MR dampers is designed. The floor plan and elevation of the prototype structure is shown

Step 7. For $j=1$ to N^{th} mode

Substep 1. Assume modal displacement vector $\mathbf{x}_0^j$

Substep 2. Determine maximum displacement for each damper u_{d0}^{ij}

$$u_{d0}^{ij} = u_0^{ij} - f_0^i / k_{br}^i$$

u_0^{ij} : maximum deformation of damper and bracing at the i^{th} MR damper in mode j

Substep 3. Calculate equivalent modal damping ratio

$$\xi_{eq}^j = \frac{1}{2} \frac{\sum_{i=1}^L \eta_i k_{eq}^i (u_{d0}^{ij})^2}{\mathbf{x}_0^{jT} \mathbf{K}_{eff} \mathbf{x}_0^j} + \xi_{in}^j$$

Substep 4. Find maximum modal displacement (y_0^j) from response spectrum

$$y_0^j = S_d(T_{eff}^j, \xi_{eq}^j) \quad \text{where } T_{eff}^j = 2\pi / \omega_{eff}^j$$

Substep 5. Update modal displacement vector $\mathbf{x}_0^j$

$$\mathbf{x}_0^j = \Gamma_j^T \boldsymbol{\phi}_j y_0^j$$

$\boldsymbol{\phi}_j$: mode vector; $\Gamma_j = \boldsymbol{\phi}_j^T \mathbf{M} \mathbf{1} / M_j$; $\mathbf{1}$: unit vector; M_j : modal mass ($= \boldsymbol{\phi}_j^T \mathbf{M} \boldsymbol{\phi}_j$)

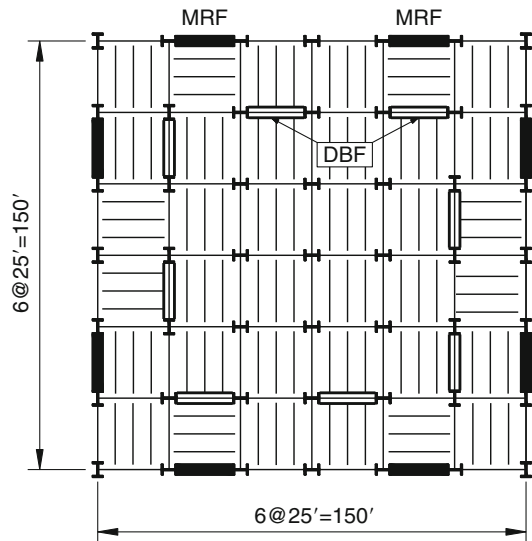
Substep 6. Repeat Substep 2 ~ 5 until convergence in $\mathbf{x}_0^j$ is achieved

Substep 7. Calculate maximum velocity $\dot{u}_{d0}^{ij}$ for each MR damper, where for the i^{th} MR damper

$$\dot{u}_{d0}^{ij} = u_0^{ij} \omega_{eff}^j$$

Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers, Fig. 5 Modal analysis method for the simplified analysis procedure utilizing response spectrum analysis (RSA) method

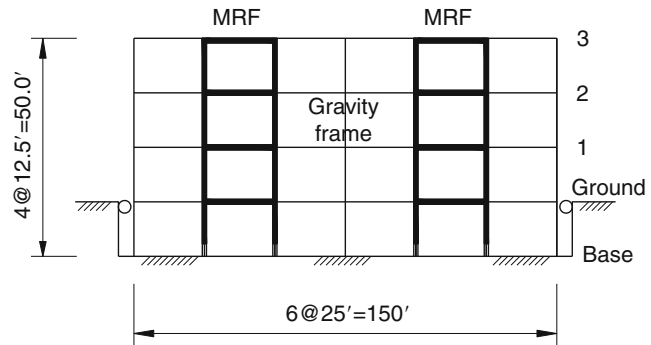
in Figs. 6 and 7, respectively. It consists of a three-story, six-bay building and represents a typical office building located in Southern California. Lateral loads are resisted by a four perimeter moment resisting frames (MRFs) and four damped braced frames (DBFs) in the two orthogonal principle directions of the building's floor plan. MR dampers are installed in the DBFs to control the drift of the building, adding supplemental damping to the structure. The DBFs have continuous columns, with pin connections at the beam-to-column connections and at the ends of the diagonal bracing. A rigid diaphragm system is assumed to exist at each floor level and the roof of the building to transfer the floor inertia loads to the MRFs and DBFs. The building has a base-moment where a point of inflection is assigned at one third of the height of the column from the column base in the analysis model.



Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers, Fig. 6 Floor plan of prototype building

Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers,

Fig. 7 Elevation of prototype building



Performance Objectives

In this design, three different performance objectives for the building are considered:

1. Limit the story drift to 1.5 % under the DBE ground motion
2. Limit the story drift to 3.0 % under the MCE ground motion
3. Design strength of members in the DBF shall not be exceeded by the demand imposed by the DBE ground motion

The MCE ground motion is represented by a response spectra that has a 2 % probability of exceedance in 50 years, and the DBE ground motion is two third the intensity of the MCE ground motion (FEMA 2000a). The performance objectives of 1.5 % story drift satisfies the life safety performance level under the DBE, and the 3 % story drift satisfies the collapse prevention level under the MCE. The performance levels are defined in FEMA-356 (2000b). To minimize the damage and repair cost to the DBF structure, the third performance objective is adopted to have the DBF structure remain elastic under the DBE.

Performance-Based Design

The prototype building structure is intended to provide the basis for an MRF and DBF with MR dampers which can be constructed in the laboratory for future tests. Due to laboratory constraints, the prototype building structure and resulting MRF and DBF were designed at 0.6-scale. The MRFs are designed to satisfy the

strength requirement of the current building seismic code of ICC (2006); the member design criteria is based on the AISC steel design provisions (2005b). The design response spectrum is based on a site in Southern California where the spectral acceleration for the short period, S_s , and for a 1 s period, S_1 , are equal to 1.5 and 0.6, respectively. The strength contribution from the DBFs and MR dampers is not considered when the MRFs are designed since, as noted above, the DBFs and MR dampers are intended only to control the story drift of the building system. More detailed information on the design of the MRFs and gravity frames can be found in Chae (2011).

Once the MRFs and gravity frames are designed for strength, the required capacity of the MR dampers to control the drift is determined. The DBF members are then designed by imposing the displacement and damper force demands on the DBF, which are obtained from the simplified analysis procedure and the required MR damper capacity. The maximum displacements and the maximum MR damper forces are assumed to occur concurrently in the SDP. The design of the three-story building is revised until the performance objectives and strength requirements are satisfied.

Large-scale MR dampers were used for the study which can generate a 200 kN damper force at a velocity of 0.1 m/s (Chae et al. 2010). The parameters for the Hershel-Bulkley model associated with the large-scale MR damper are: $f_0 = 138.5$ kN, $C = 161.8$ kNsec/m, and $n = 0.46$.

The optimal damper location which satisfies the performance objectives is determined by using the simplified analysis procedure, resulting in one large-scale MR damper in the second and third stories, respectively, with $\alpha = 10$ and $\beta = 0.3$ (Chae 2011). Tables 1 and 2 summarize the member sizes for the MRFs, gravity frames, and DBFs. Table 3 summarizes the calculated design demand associated with maximum story drift and maximum damper forces. As can be observed, the design demand for the story drift under the DBE and MCE are less than 1.5 % and 3.0 %, respectively, in order to satisfy the performance objectives. Table 4 shows the DBE demand-to-capacity ratios for the DBF members. The demand-to-capacity ratio for each member is less than 1.0, which means the members

are designed to remain elastic under the DBE. The design of the braces was controlled by stiffness, $\alpha = 10$, and not strength, hence, the demand-to-capacity ratios for the braces are small in Table 4.

Assessment of Simplified Design Procedure

The SDP is assessed by comparing the design demand from the SDP with results from a series of nonlinear time-history analyses (NTHA) of the three-story building using the nonlinear finite element program OpenSees (2009).

OpenSees Model

Symmetry in the floor plan and ground motions along only one principal axis of the building were considered in the analysis. Hence, only one-quarter of the building was modeled consisting of one MRF, one DBF, and the gravity frames that are within the tributary area of the MRF and DBF. The OpenSees model is shown in Fig. 8. The beams and columns of the MRF structure are modeled with a nonlinear distributed plasticity force-based beam-column element with five fiber sections along the element length. The cross section of the element is discretized into 18 fibers, including 12 fibers for the web and 3 fibers each for the top and bottom flanges. Each fiber is modeled with a bilinear stress-strain relationship with a post-yielding stiffness that is 0.01 times the elastic stiffness. The beam-column joints in the MRF are modeled using a panel zone element, where shear and symmetric column bending deformations are considered (Seo et al. 2009). The doubler plates in the panel zones of the MRF are included in the model.

Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers, Table 1 Member sizes for MRFs and gravity frames

Story (or floor level)	MRFs		Gravity frames	
	Column	Beam	Column	Beam
1	W8X67	W18X46	W8X48	W10X30
2	W8X67	W14X38	W8X48	W10X30
3	W8X67	W10X17	W8X48	W10X30

Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers, Table 2 Member sizes for DBFs

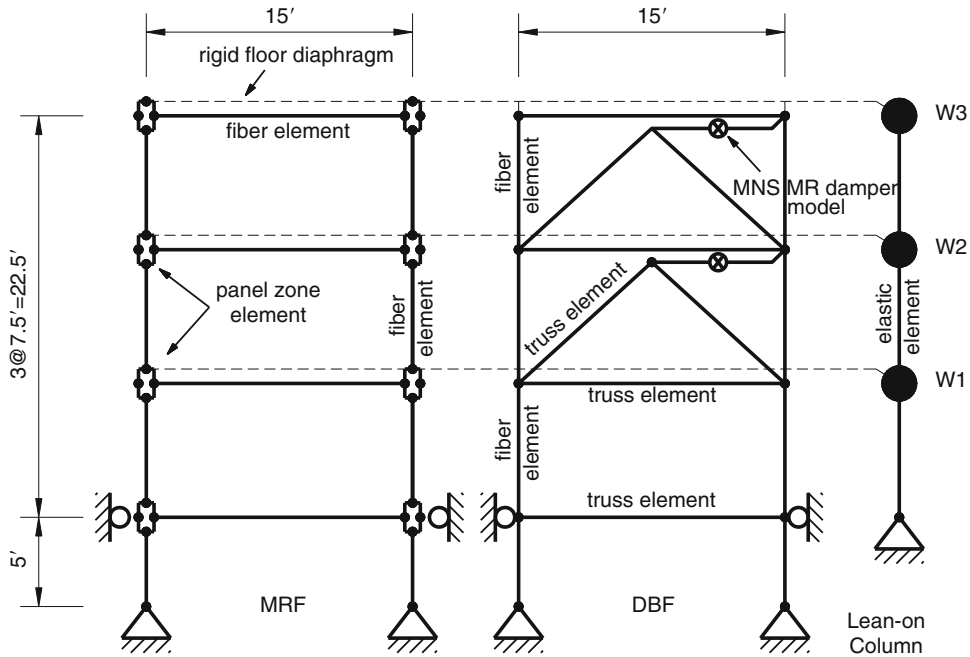
Story (or floor level)	Column	Beam	Diagonal bracing
1	W10X33	W10X30	—
2	W10X33	W10X30	W6X20
3	W10X33	W10X30	W6X20

Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers, Table 3 Calculated design demand associated with maximum story drift and maximum damper force from SDP

Story	Maximum story drift (%)		Maximum damper force (kN)	
	DBE	MCE	DBE	MCE
1	1.18	1.91	—	—
2	1.35	2.32	222.9	244.4
3	1.41	2.57	233.6	261.6

Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers, Table 4 Demand-to-capacity ratio for DBF frames, DBE

Story (or floor level)	Column (W10X33)	Beam (W10X30)	Brace (W6X20)
1	0.955	0.521	—
2	0.303	0.576	0.270
3	0.079	0.354	0.283



Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers, Fig. 8 OpenSees model for 0.6-scale building

The nonlinear force-based fiber element is also used to model the columns of the DBF. The beams and braces of the DBF are modeled using linear elastic truss elements. The gravity frames are idealized using the concept of a lean-on column, where an elastic beam-column element with geometric stiffness is used to model the lean-on column. The section properties of the lean-on column are obtained by taking the sum of the section properties of each gravity column within the tributary area (i.e., one quarter of the floor plan) of the MRF and the DBF. The MR dampers are modeled using the MNS MR damper model implemented into OpenSees by Chae et al. (2010). The MR damper is assumed to be located between the top of the brace and the adjacent beam-column joint, as shown in Fig. 8. The results reported in this paper are for MR dampers that are passive controlled with a constant current input of 2.5 A. Studies with the MR dampers in semi-active control mode are presented in Chae (2011).

The gravity loads from the tributary gravity frames are applied to the lean-on column

to account for the $P-\Delta$ effect of the building. To model the effect of the rigid floor diaphragm, the top node of the panel zone element in the MRF and the beam-column joint in the DBF are horizontally constrained to the node of the lean-on column at each floor level, while the vertical and rotational degrees of freedom of these nodes are unconstrained.

Rayleigh damping is used to model the inherent damping of the building with a 5 % damping ratio for the first and second modes.

Comparison of Response

An ensemble of 44 ground motions listed in FEMA P695 (ATC 2009) is scaled to the DBE and MCE levels using the procedure by Somerville et al. (1997) for the NTHA.

A summary of the median and standard deviation of maximum story drift and residual story drift from the NTHA is given in Table 5. Figures 9 and 10 compare the calculated design demand for drift from the SDP with the median values for maximum story drift from the NTHA under the DBE and MCE ground motions. The story drift

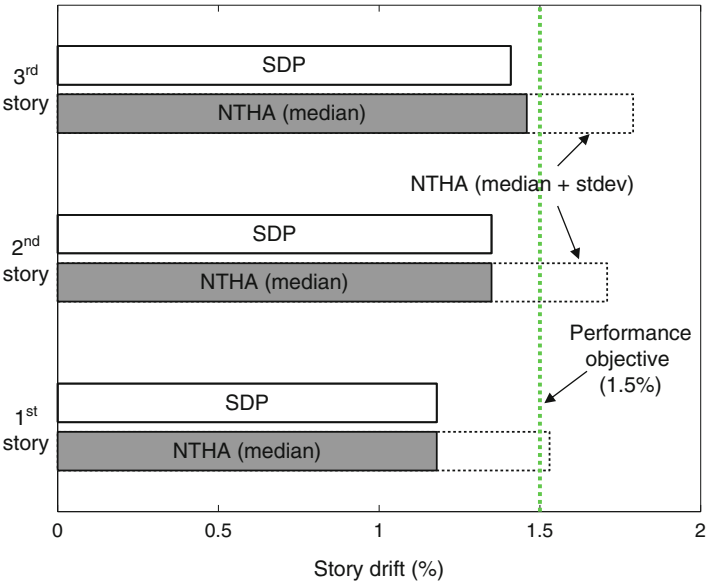
Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers, Table 5 Median and standard deviation of maximum and residual story drift from nonlinear time history analysis

Story	DBE level		MCE level	
	Max story drift (%)	Residual drift (%)	Max story drift (%)	Residual drift (%)
1	1.18 (0.35) ^a	0.11 (0.21)	1.86 (0.85)	0.42 (0.62)
2	1.35 (0.36)	0.17 (0.26)	2.10 (0.85)	0.57 (0.66)
3	1.46 (0.33)	0.22 (0.27)	2.32 (0.84)	0.63 (0.69)

^aValue in () indicates standard deviation response

Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers,

Fig. 9 Comparison of story drift between SDP and NTHA, DBE ground motions



design demand calculated by the SDP shows good agreement in Fig. 9 with the median maximum story drifts from the NTHA. The calculated design demand for story drift from the SDP also shows good agreement with the median values from the NTHA under the MCE, see Fig. 10. Figures 9 and 10 show that the median values of the maximum story drift from the NTHA satisfies the performance objectives of 1.5 % and 3.0 % under the DBE and MCE levels, respectively. The residual story drift of the building after the DBE has a maximum median value and standard deviation of 0.22 % and 0.27 %, respectively, which occurred in the third story as summarized in Table 5. The residual drift is small.

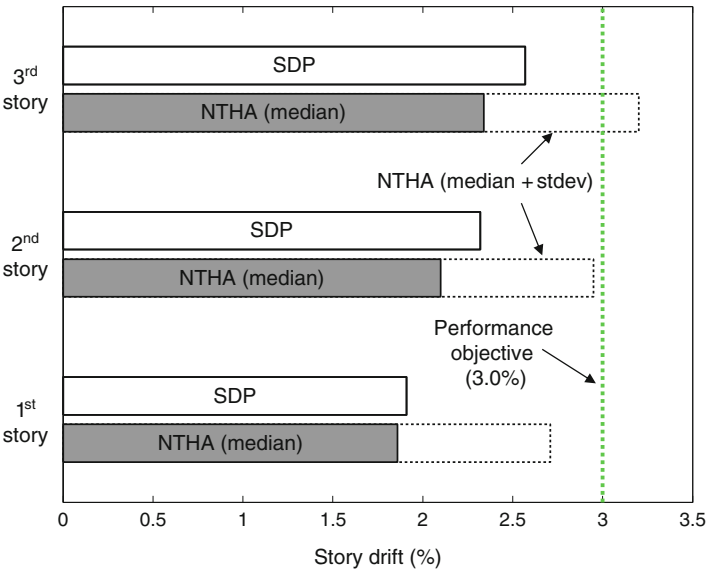
Table 6 compares the design demand for maximum MR damper forces calculated by the SDP with the median maximum MR damper forces from the NTHA. The MR damper force design

demands from the SDP are slightly smaller than the median NTHA results for the DBE. However, the differences between the SDP and the NTHA are only 3.9 % and 3.3 % for the MR dampers in the second and third stories, respectively. For the MCE, the differences between the median NTHA results and the SDP for the MR damper forces in the second and third stories are 1.6 % and 0.5 %, respectively. The design demand calculated by the SDP shows reasonably good agreement with the median results from the NTHA for the maximum MR damper forces.

The linear elastic behavior of the DBF columns under the DBE is confirmed by checking the plastic rotation developed in the columns. Summarized in Table 7 are the median and standard deviation of the DBF maximum magnitude of column plastic rotation from the NTHA for the DBE ground motions. In the first story, some

Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers,

Fig. 10 Comparison of story drift between SDP and NTHA, MCE ground motions



Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers, Table 6 Comparison of maximum damper forces

Story	DBE level		MCE level	
	SDP	NTHA	SDP	NTHA
1	—	—	—	—
2	222.9	231.9 (6.4) ^a	244.4	248.4 (7.3)
3	233.6	241.5 (8.3)	261.6	260.2 (9.4)

^aValue in () indicates standard deviation response

plastic rotation developed at the base of the column under the DBE. However, the median is zero and the standard deviation is 0.0005 rad for the maximum plastic rotation. The median and standard deviation of the maximum plastic rotation in the second and third stories columns are zero under the DBE ground motion, as given in Table 7, which indicates linear elastic behavior of the columns at these stories. The median of the residual plastic rotation at the base of the column is zero and the standard deviation for the residual plastic rotation is 0.0004 rad.

The building’s response under the DBE and MCE determined by the NTHA appears to have met the performance objectives for the structure. Moreover, the performance under the DBE meets the immediate occupancy level (FEMA 2000b), where the maximum residual drift is practically equal to the allowable construction tolerance of 0.2 % for steel structures (AISC 2005a), with

Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers, Table 7 Median and standard deviation of DBF maximum magnitude of column plastic rotation from nonlinear time-history analysis and DBE ground motions

Story	Location along column	Max plastic rotation (rad %)	Residual plastic rotation (rad %)
1	Bottom	0.00 (0.05) ^a	0.00 (0.04)
	Top	0.00 (0.00)	0.00 (0.00)
2	Bottom	0.00 (0.00)	0.00 (0.00)
	Top	0.00 (0.00)	0.00 (0.00)
3	Bottom	0.00 (0.00)	0.00 (0.00)
	Top	0.00 (0.00)	0.00 (0.00)

^aValue in () indicates standard deviation response

minimal damage occurring in the building (the median of the maximum plastic rotations in the beams and columns of the MRF are 0.37 % and 0.07 % radians, respectively, with the MRF having the same residual drift statistics as the DBF due to the rigid diaphragm (Chae 2011)).

Summary

A simplified design procedure was developed to enable the performance-based design of structures with MR dampers. The design procedure utilizes a systematic approach to calculate the design demand of structures with MR dampers.

The simplified analysis procedure enables the design demand to be determined without performing a nonlinear time-history analysis by linearizing the structure and utilizing the response spectrum analysis method.

To linear the nonlinear MR dampers, the simplified analysis procedure is based on an equivalent linear MR damper model using the Hershel-Bulkley quasi-static MR damper model. The energy dissipated by the MR damper over one cycle of a harmonic motion is calculated and the equivalent stiffness determined based on the secant stiffness method from the damper force-displacement relationship. The loss factor of the MR damper is obtained from the energy dissipated by the damper and the strain energy calculated from the equivalent stiffness of the damper. Both the equivalent stiffness and loss factor of the MR damper are dependent on the maximum displacement of the damper. A three-story building was designed using the SDP, where three performance objectives associated with two seismic hazard levels were selected. The SDP was assessed by comparing the design demand calculated by the SDP with the response determined from nonlinear time-history analyses. The MNS MR damper model was implemented into the OpenSees computer program and statistics for the response to DBE, and MCE ground motions were obtained from a series of nonlinear time-history analyses using 44 different ground motions. The performance of the building from the nonlinear time-history analysis indicated that building design satisfies the three performance objectives. The design demand associated with story drift and maximum MR damper forces from the SDP showed good agreement with the median values from the nonlinear time-history analyses, confirming the robustness of the SDP.

References

- AISC (2005a) Code of standard practice for steel buildings and bridges. American Institute of Steel Construction, Chicago
- AISC (2005b) Seismic provisions for structural steel buildings. American Institute of Steel Construction, Chicago
- Applied Technology Council (2009) Quantification of building seismic performance factors. ATC-63 project report (FEMA P695), Redwood City
- Chae Y (2011) Seismic Hazard mitigation of building structures using magneto-rheological dampers, PhD dissertation, Lehigh University
- Chae Y, Ricles JM, Sause R (2010) Development of a large-scale MR Damper model for seismic hazard mitigation assessment of structures. In: 9th US National and 10th Canadian conference on earthquake engineering, Toronto
- Constantinou MC, Soong TT, Dargush GF (1998) Passive energy dissipation systems for structural design and retrofit. Monograph series, MCEER. State University of New York at Buffalo, Buffalo
- Fan CP (1998) Seismic analysis, behavior, and retrofit of non-ductile reinforced concrete frame buildings with viscoelastic dampers. PhD dissertation, Lehigh University, Bethlehem
- Federal Emergency Management Agency (2000a) Recommended seismic design criteria for new steel moment-frame buildings. Report no. FEMA-350, Washington, DC
- Federal Emergency Management Agency (2000b) Prestandard and commentary for the seismic rehabilitation of buildings. Report no. FEMA-356, Washington, DC
- International Code Council (2006) International building code. International Code Council, Falls Church
- Kwan WP, Billington SL (2003) Influence of hysteretic behavior on equivalent period and damping of structural systems. *ASCE J Struct Eng* 129(5):576–585
- Lee K-S, Fan C-P, Sause R, Ricles J (2005) Simplified design procedure for frame buildings with viscoelastic and elastomeric structural dampers. *Earthq Eng Struct Dyn* 34:1271–1284
- Lee K-S, Ricles J, Sause R (2009) Performance-based seismic design of steel MRFs with elastomeric dampers. *J Struct Eng* 135(5):489–498
- Lin WH, Chopra AK (2003) Earthquake response of elastic Single-degree-of-freedom systems with nonlinear viscoelastic dampers. *ASCE J Eng Mech* 129(6):597–606
- Newmark NM, Hall WJ (1973) Seismic design criteria for nuclear reactor facilities. Report no. 46, Building Practices for Disaster Mitigation, National Bureau of Standards, U.S. Department of Commerce, pp 209–236
- OpenSees (2009) Open system for earthquake engineering simulation. Pacific Earthquake Engineering Research Center, University of California, Berkeley
- Ramirez OM, Constantinou MC, Gomez JD, Whittaker AS (2002) Evaluation of simplified methods of analysis of yielding structures with damping systems. *Earthq Spectra* 18(3):501–530
- Sause R, Hemingway GJ, Kasai K (1994) Simplified seismic response analysis of viscoelastic-damped frame structures. In: 5th U.S. National conference on earthquake engineering, Vol. I, pp 839–848

- Seo CY, Lin YC, Sause R, Ricles JM (2009) Development of analytical models for 0.6 scale self-centering MRF with beam web friction devices. In: 6th International conference for steel structures in seismic area (STESSA), Philadelphia
- Somerville P, Smith N, Punyamurthula S, Sun J (1997) Development of ground motion time histories for Phase 2 of the FEMA/SAC steel project. Report no. SAC/BD-97/04, SAC Joint Venture, Sacramento
- Soong TT, Dargush GF (1997) Passive energy dissipation systems in structural engineering. Wiley, West Sussex
- Symans MD, Constantinou MC (1998) Passive and fluid viscous damping systems for seismic energy dissipation. *ISET J Earthq Technol* 35(4):185–206

Physics-Based Ground-Motion Simulation

Ricardo Taborda¹ and Daniel Roten²

¹Department of Civil Engineering, and Center for Earthquake Research and Information, University of Memphis, Memphis, TN, USA

²San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA, USA

Synonyms

Deterministic earthquake simulation; Deterministic ground-motion simulation; Physics-based earthquake simulation

Introduction

Physics-based earthquake ground-motion simulation, also referred to as deterministic earthquake ground-motion simulation, can be defined as the prediction of the ground motion generated by earthquakes by means of numerical methods and models that incorporate explicitly the physics of the earthquake source and the resulting propagation of seismic waves. Other approaches such as the stochastic ground-motion simulation method or ground-motion prediction equations (i.e., attenuation relationships) integrate the

physics and specific characteristics of the earthquake source, directivity, path, attenuation and scattering, basin, and site effects by means of indirect, approximate, or statistical approaches. These methods are valid representations of the physics of earthquakes, but do not necessarily solve the accepted mathematical abstractions that describe the physics of source dynamics and wave propagation. Physics-based ground-motion simulation, on the other hand, seeks to account for most or all of these aspects explicitly, thus the distinction given to its name.

This chapter reviews the background of physics-based ground-motion simulation, describes some of the most popular methods used in this field, and presents application examples. The first section focuses on the background of the method. The second section presents the concept of a physics-based simulation framework. Subsequent sections review the basic methods applied for the solution of seismic waves traveling in solids, including elastic, anelastic, and elastoplastic media, and provide details about the input data used in simulations (source and velocity models) while addressing other important complementary aspects in numerical modeling and computer simulation along the way. The final sections present recent examples and applications and a summary with a perspective on the future use of physics-based simulation in earthquake hazard analysis and engineering.

Background

The foundation for a physics-based approach to earthquake ground-motion simulation was laid down in the late 1960s and early 1970s when the finite-difference (FD) and the finite-element (FE) methods were first introduced in seismology (Alterman and Karal 1968; Lysmer and Drake 1972). Before numerical methods started to be more broadly used, only a few problems under restricted conditions could be solved analytically. Classical examples include the surface response of semi-cylindrical and semi-elliptical canyons and alluvial valleys under incident plane SH

waves (e.g., Wong and Trifunac 1974). Analytical approaches were, however, useful only to characterize the ground motion under idealized settings that neglected the geometrical irregularity of the geology and the heterogeneity of the media and were typically considered for sites away from the source, or with simplified source representations.

By contrast, FD and FE approaches offered more flexibility. They could be used to describe the propagation of waves in (irregular) stratified media, alluvial valleys, and sedimentary basins, assess ground-motion amplification, and study source dynamics (e.g., Boore 1972; Smith 1975). Initially, due to computational limitations, numerical methods were used mostly for solving two-dimensional (2D) problems. With time, advances in numerical methods and algorithms and the growth and increased availability of computing power and memory capacity allowed scientists to model larger and more complex 2D and small three-dimensional (3D) problems. Virieux (1984), in particular, introduced the staggered-grid FD scheme in the context of seismology, which would later become a preferred approach for modeling seismic wave propagation problems.

The first 3D simulations done at scales large enough to synthesize source, path, basin, and site effects were published in the early 1990s. Frankel and Vidale (1992), for instance, simulated the ground motion generated by an M_L 4.4 aftershock of the 1989 Loma Prieta, California, earthquake using a point source model and a FD representation of the Santa Clara Valley. The model had about four million nodes in a simulation domain of 30 km by 22 km and 6 km in depth. The synthetic seismograms obtained were valid for a maximum frequency, $f_{\max} = 1$ Hz, and the minimum shear-wave velocity ($V_{S_{\min}}$) considered was 600 m/s. Although the synthetics lacked the level of fidelity expected from direct comparisons with data, they were, in general, comparable in amplitude and duration. More important, the results of Frankel and Vidale (1992) showed that simulations offered a plausible means to understanding the characteristics of 3D wave propagation in highly heterogeneous media.

Following Frankel and Vidale (1992), FD applications gained significant traction (e.g., Olsen et al. 1995; Graves 1996). Some of these works were done using parallel computers. Olsen et al. (1995), for instance, simulated the ground motion of an M 7.75 scenario earthquake on the San Andreas Fault. The model covered the entire Greater Los Angeles metropolitan region in a simulation domain of 230 km by 140.4 km and 46 km in depth. The simulation was designed for $f_{\max} = 0.4$ Hz and $V_{S_{\min}} = 1$ km/s, using a grid spacing of 0.4 km. The resulting model had over 23.5 million grid points. This simulation was one of the first to use parallel computers for simulating the ground motion at a regional scale. It required nearly 23 h to complete 4,800 time steps on a 512-processor nCUBE-2 computer. Despite some limitations, Olsen et al. (1995) made important observations about the significance of 3D basin and edge effects that could not have been described at the time using only equivalent 1D or 2D models.

Similarly, FE applications were also developed for both source dynamics and wave propagation problems (e.g., Bao et al. 1998; Aagaard et al. 2001). Bao et al. (1998), for example, used a parallel computer FE application to simulate the ground response of the San Fernando Valley, California, to an aftershock of the 1994 Northridge earthquake. The FE mesh consisted of 76.8 million elements, with parameters $V_{S_{\min}} = 220$ m/s and $f_{\max} = 1.6$ Hz. The simulation required 7.2 h to execute 16,667 time steps (40 s of ground motion) on 256 processors in a Cray T3D machine. Bao et al. (1998) observed that, besides the larger amplitudes and longer durations associated with the deeper parts of the basin beneath the San Fernando Valley, the ground motion was also significantly amplified by the constructive interference of surface and trapped body waves in the shallower soft-material deposits.

Pseudo-spectral, high-order FE, spectral element (SE), and discontinuous Galerkin (DG) methods have also been used (e.g., Seriani 1998; Komatitsch and Vilotte 1998; Dumbser and Käser 2006). Some SE applications have been particularly successful in simulations at the continental and global scales. Komatitsch

et al. (2003), for instance, show results using a SE approach to model seismic waves propagating through the Earth's globe, including the full complexity of a 3D crustal model with a mesh of 5.5 billion grid points, for seismic periods as low as 5 s (i.e., $f_{\max} = 0.2$ Hz). By contrast, problems of small and moderate sizes continue to be addressed using analytical or semi-analytical methods with approaches such as the boundary element, coupled boundary-domain element, discrete wave-number, and hybrid methods (e.g., Bouchon 1979; Mossessian and Dravinski 1987; Bielak et al. 1991; Sánchez-Sesma and Luzón 1995).

More recently, the increased availability and power of parallel computers have facilitated enormously the advance of 3D physics-based earthquake ground-motion simulation. Chaljub et al. (2010) and Bielak et al. (2010), for example, show the tremendous advance by both Europe- and US-based research groups to conduct verification of simulation codes used for modeling the wave propagation characteristics of historic and scenario earthquakes. It is now commonplace to see simulation domains ranging in the order of tens to hundreds of kilometers, with total number of elements or cells in the order of tens to hundreds of billions. Maximum frequencies modeled today vary between 2 and 5 Hz for regional-scale ground-motion simulations and are as high as 10 Hz for smaller-scale rupture dynamics and local wave propagation problems (e.g., Cui et al. 2010; Taborda and Bielak 2013; Shi and Day 2013). In addition, the advent of newer technologies such as general purpose graphic processing units (GPGPU), hybrid CPU and GPU systems, and many integrated core (MIC) architectures are further helping to accelerate forward and inverse, regional and global wave propagation simulations (e.g., Komatitsch et al. 2010; Zhou et al. 2012; Rietmann et al. 2012).

Altogether, the progress shown over the past few years has opened the possibility of using physics-based ground-motion simulation in earthquake engineering applications. The framework for using simulations for a physics-based approach to regional seismic hazard mapping, for

instance, has already been put in place and is being used at low frequencies (up to 0.5 Hz) in Southern California (Graves et al. 2011). Some obstacles, however, still need to be sorted out. The present knowledge of the earthquake source, crustal structure, material properties, and local site effects is still far from ideal. All these aspects will require much research in the years to come, but the trajectory indicates that physics-based earthquake ground-motion simulation will play a significant role in future seismic hazard estimation, risk assessment, and earthquake engineering analysis and design.

The Physics-Based Simulation Workflow

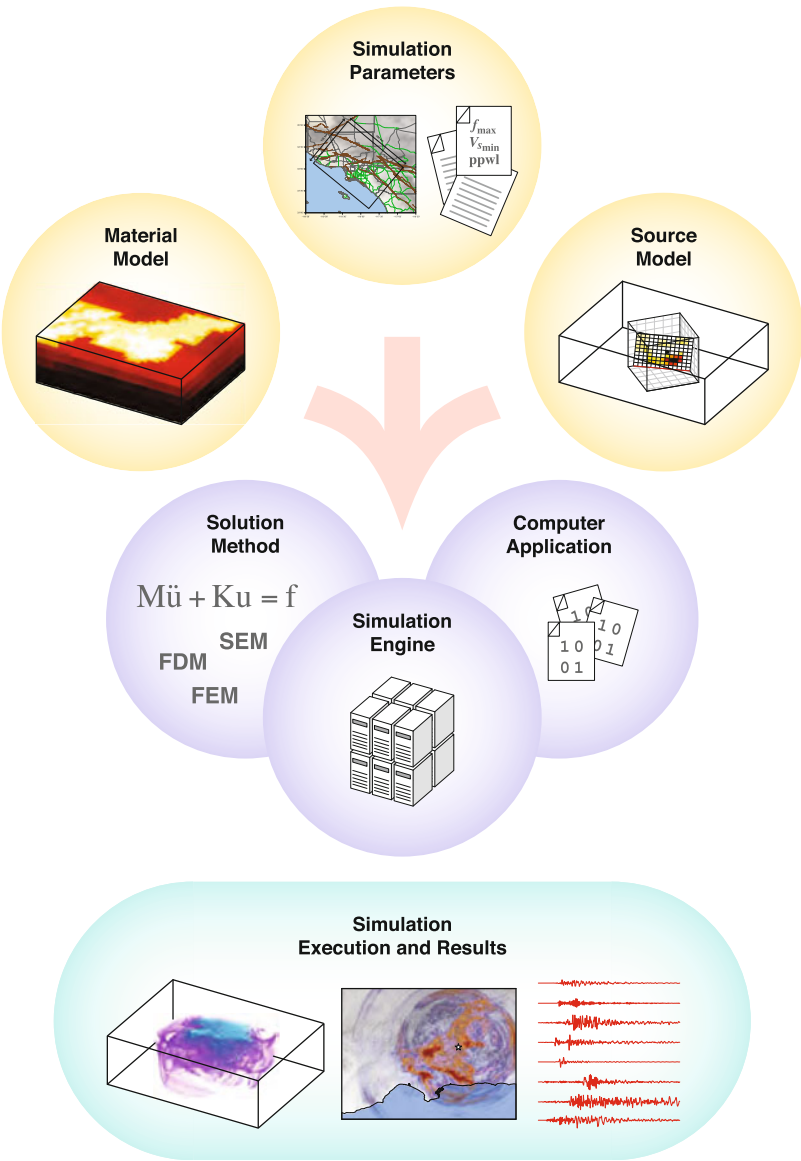
Physics-based earthquake simulations operate within a basic general workflow that consists of the following elements or steps:

- The selection of a region of interest and simulation domain
- The selection of a source model and a material model
- The definition of the modeling parameters (maximum frequency, minimum velocity, etc.)
- The implementation of solution methods and operation of a simulation engine
- The execution of the simulation and collection of results

This simulation workflow is illustrated in Fig. 1. The top section of the workflow refers to the input data that is required for the simulation, that is, the selection of the simulation domain, the source model, the material model, and the simulation parameters. The source model provides information about the fault rupture characteristics in the form of its location, orientation, and slip history. Most physics-based earthquake simulations use kinematic models to represent the source as will be explained later. However, dynamic rupture simulations that fully solve the rupture evolution on the fault plane and the triggered wave propagation problem can also be combined with the simulation of the ground motion. The material model provides information

Physics-Based Ground-Motion Simulation,

Fig. 1 Typical workflow in physics-based earthquake ground-motion simulations. The *top* section (in yellow) refers to the input models and parameters; the *middle* section (in purple) refers to the solution method and implementation into the simulation engine; and the *bottom* section (in green) refers to the simulation execution and results



about the properties of the material contained in the chosen simulation domain. Material models are often referred to simply as seismic velocity models, as most of these models only provide information about the *P*- and *S*-wave velocities and the density of the medium. However, a complete description of the material also provides information about its dissipation properties and capacity. A later section is dedicated to the description of some publicly accessible velocity models. The last pieces of information at the input data level are the simulation parameters.

At the most basic level, these consist of the maximum targeted frequency in the simulation ($f_{\max}$, usually defined in Hz) and the minimum shear-wave propagation velocity ($V_{s\min}$, usually defined in m/s). Also relevant at this point is the definition of the number of points per wavelength, which is the integer number of points that will be used to discretize a complete wave cycle. Together, these parameters define the level of refinement or resolution necessary to solve the problem with a certain acceptable level of accuracy to the extent possible.

The second section in the workflow shown in Fig. 1 refers to the solution method and its implementation in a computer code application. Together, they provide a simulation engine that can be used in personal computers, clusters, or supercomputers. There are various simulation computer codes used in research today, some of which provide open-access distribution. A well-documented example of a code available to users is the SPECFEM software family, distributed by the Computational Infrastructure for Geodynamics project (<http://www.geodynamics.org>). Given the computing resources necessary for the solution of forward wave propagation simulations at regional scales, the implementation of such simulation engines in parallel computer codes and the use of computer clusters and supercomputers have become a commonplace in physics-based earthquake simulation. Until now, this has restricted the use of physics-based earthquake simulations to research-oriented activities. However, as the projected capacity of personal computers over the next 50 years will make parallel computing applications massively accessible to the public, it is expected that the use of these simulation codes and that of physics-based simulations overall will increase within the seismology and earthquake engineering communities in the near future. The last portion of the workflow shown in Fig. 1 makes reference to the execution of the simulation in itself and the gathering of results. Typical simulation output datasets come in three basic flavors: individual station records, plane wave fields, and volumetric wave fields. Station records are usually output in the form of text or binary files with velocities (or displacements) ordered sequentially at every time step. Plane and volumetric wave fields are commonly delivered as binary files indexed in such a way that they can be sliced in 2D and 3D arrays with the ground response for every time step. Simulators usually produced these outputs at a (decimated) time step that is larger than that actually used internally in the computation of the solution. Station records and plane wave field outputs are the most common of the three because their file sizes make them more manageable in terms of memory and disk space capacity.

A station record file is typically of the order of a few hundreds of kilobytes to a couple megabytes, and a plane file ranges between tens and hundreds of gigabytes. Volumetric wave fields, on the other hand, can reach the order of terabytes and thus become onerous to transfer and store.

The following sections expand on some of these aspects, including the most common solution methods, typical representations of source models, and seismic velocity models used in regional simulations, and provide examples about the use given to physics-based earthquake simulations today. Specific computer code applications are not covered here, although some basic principles about their numerical implementation in computer codes are covered in the sections dedicated to the solution methods used in physics-based simulation.

Wave Propagation in Elastic Media

As mentioned in the previous sections, there are various analytical and numerical methods used for solving wave propagation problems. The simulation of the ground motion at scale, that is, the simulation in domains large enough for synthesizing the earthquake source and the local and regional response of the ground, has, however, been primarily approached using FD, FE, SE, and, lately, DG methods. Since the formulation of both the SE and DG methods share some of the basic concepts of FE method, special attention is given here only to the FE and FD methods. This section covers these two methods applied to elastic media first. Subsequent sections address the problem of anelasticity and plasticity separately.

The Finite-Element Method

Earthquake ground-motion simulation entails obtaining the solution of the linear momentum equation, which can be written in Cartesian coordinates and indicial notation as

$$\sigma_{ij,j} + f_i = \rho \ddot{u}_i. \quad (1)$$

Here, σ_{ij} represents the Cauchy stress tensor, ρ is the mass density, and f_i and u_i are the body forces and displacements in the i direction within a bounded domain Ω . The two dots over the displacements indicate second derivative in time. The indices i and j in the subscripts represent the Cartesian coordinates x , y , and z . When a subscript follows a comma, this indicates a partial derivative in space with respect to the corresponding index. For the special case of elastic isotropic solids, the stress tensor can be expressed in terms of strains following Hooke's law of elasticity, and the strains, in turn, can be expressed in terms of displacements. The resulting expression for the stress tensor is

$$\sigma_{ij} = \lambda u_{k,k} \delta_{ij} + \mu (u_{i,j} + u_{j,i}) \quad (2)$$

where λ and μ are the Lamé parameters and δ_{ij} is Kronecker's delta. In general, the Lamé parameters in Eq. 2 and the density in Eq. 1 are assumed to be locally constant. Substituting Eq. 2 into Eq. 1 leads to

$$\lambda u_{k,kj} \delta_{ij} + \mu (u_{i,jj} + u_{j,ji}) + f_i = \rho \ddot{u}_i. \quad (3)$$

This is Navier's equation of elastodynamics. Using the standard Galerkin method, one can obtain the weak form of this equation and then discretize the problem in space. This procedure entails the introduction of set of arbitrary functions v , known as the test functions. The test functions are auxiliary functions which help formulate an approximate solution $\hat{u}$ to the displacements u , called the trial functions. The domain Ω is then discretized in space using a set of global piecewise linear basis functions ϕ , which divide the domain into discrete elements Ω_e . As a result, both the test and trial functions become linear combinations of the global basis functions,

$$v_h(x, y) = \sum_{i=1}^N \phi_i(x, y) v_i^h, \text{ and} \quad (4)$$

$$\hat{u}_h(x, y) = \sum_{i=1}^N \phi_i(x, y) \hat{u}_i^h, \quad (5)$$

respectively. Here, h is used to indicate that the domain Ω has been approximated to a discrete version of it, Ω_h , composed of all the elements with domain Ω_e , where h is a discretization parameter (e.g., element size). These elements Ω_e are connected to each other along their edges and at their vertices. The vertices of the elements are called nodes. Both the nodes and the elements constitute a FE mesh, where N is the total number of nodes. The index i indicates that the values of v and $\hat{u}$ are evaluated at the nodes using the associate global function ϕ_i . It can be shown that substituting Eqs. 4 and 5 into Eq. 3 leads to

$$\mathbf{M}\mathbf{u} + \mathbf{K}\mathbf{u} = \mathbf{f}, \quad (6)$$

where $\mathbf{M}$ and $\mathbf{K}$ are the assembled global mass and stiffness matrices of the system's discrete FE mesh representation, $\mathbf{f}$ is the assembled vector of body forces (which is determined based on the kinematic representation of the source – see the section “[Source Models](#)” below), and $\mathbf{u}$ is the assembled vector of displacements at the nodes in the FE mesh. Once again, the double dots over the displacement vector mean second derivative in time (i.e., acceleration). Not shown here for brevity is the fact that in following the FE method, the test function terms vanish. The matrices $\mathbf{M}$ and $\mathbf{K}$ and the vector $\mathbf{f}$ in Eq. 6 are composed of terms corresponding to the nodes of the elements in the FE mesh. The i th row and j th column terms in these matrices and vector are given by

$$\mathbf{M}_{ij} = \int_{\Omega} \rho \phi_i \phi_j d\Omega, \quad (7)$$

$$\mathbf{K}_{ij} = \int_{\Omega} (\mu + \lambda) \phi_i \phi_j^T d\Omega + \int_{\Omega} \mu \phi_i^T \phi_j d\Omega, \quad (8)$$

$$\mathbf{f}_i = \int_{\Omega} \phi_i f d\Omega. \quad (9)$$

However, the global matrices $\mathbf{M}$ and $\mathbf{K}$ are seldom constructed explicitly for the full simulation domain Ω . A common practice is, instead, to perform the products $\mathbf{M}\ddot{\mathbf{u}}$ and $\mathbf{K}\mathbf{u}$ at the element level Ω_e using local basis functions ψ_i instead of

the global basis functions ϕ_i . This leads to performing the products $\sum_e \mathbf{M}^e \ddot{\mathbf{u}}^e$ and $\sum_e \mathbf{K}^e \mathbf{u}^e$, where $\mathbf{M}^e$ and $\mathbf{K}^e$ are the mass and stiffness matrices of each finite element in the mesh, built using the local basis functions, and $\ddot{\mathbf{u}}^e$ and $\mathbf{u}^e$ are the corresponding acceleration and displacement vectors that have the nodes associated with each element.

When operations are done at the element level using local basis functions as just described, the system of ordinary differential equations in Eq. 6 can be rewritten as

$$\sum_e \mathbf{M}^e \ddot{\mathbf{u}}^e + \sum_e \mathbf{K}^e \mathbf{u}^e = \sum_e \mathbf{f}^e. \quad (10)$$

Here, the summation symbol means assembling of all elements e in the FE mesh. At any time step n , the acceleration $\ddot{\mathbf{u}}_n$ can then be expressed in terms of the displacements by applying second-order central differences. Then, for the n th time step in the simulation, Eq. 10 becomes

$$\sum_e \mathbf{M}^e \left(\frac{\mathbf{u}_{n-1}^e - 2\mathbf{u}_n^e + \mathbf{u}_{n+1}^e}{\Delta t^2} \right) + \sum_e \mathbf{K}^e \mathbf{u}_n^e = \sum_e \mathbf{f}_n^e, \quad (11)$$

where Δt is the size of the time step. Furthermore, the system can be uncoupled using a diagonally lumped mass matrix. In that case, the elements of the mass matrix are such that $m_{ij} = m_i$ for $i = j$ and $m_{ij} = 0$ for $i \neq j$. This allows the forward step-by-step explicit solution of the displacements at time step $n + 1$. The solution for each node i in the mesh can be written as

$$u_{n+1}^i = \frac{\Delta t^2}{m_i} f_n^i - (u_{n-1}^i - 2u_n^i) - \frac{\Delta t^2}{m_i} \left(\sum_e \mathbf{K}^e \mathbf{u}_n^e \right)_i, \quad (12)$$

where m_i is the mass lumped at node i , u_n^i is the displacement at node i and step n , and the summation within the parenthesis corresponds to the assembling of the stiffness contributions of all the elements that share node i .

Note that in the formulation presented in Eqs. 1 through 9, no mention was made about

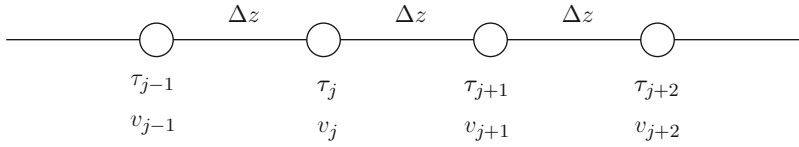
the boundary conditions. In FE, the traction-free conditions at the free surface are naturally met and no special treatment is needed. For the lateral and the bottom faces of the domain, however, appropriate measures need to be taken to effectively diminish or vanish the occurrence of spurious reflections at the finite boundaries of the simulation domain. There are several alternatives to implement absorbing boundary conditions in FE applications. Perhaps the simplest of them all consists on placing dampers at the nodes on the boundaries designed to absorb compression and shear plane waves locally. This approximation, while far from ideal, has been used in large-scale simulations with minimum reflections and acceptable performance (e.g., Bielak et al. 2010). Other more accurate absorbing boundary conditions can be satisfied with the implementation, for instance, of the perfectly matching layers (PML) method (e.g., Ma and Liu 2006).

Equations 6 through 12 above provide the basic formulation for a forward wave propagation problem in an elastic medium in which the conditions imposed by the geometrical irregularities or the material's heterogeneity are approximated by means of an appropriate discretization of the simulation domain. In earthquake ground-motion modeling, the meshing criteria are determined based on the maximum simulation frequency ($f_{\max}$), the desired number of points per wavelength (p), and the local material properties defined by the shear-wave velocity (V_S). The size e of each element is set so that it satisfies the rule

$$e_{\max} \leq \frac{V_S}{p f_{\max}}. \quad (13)$$

The adequate numbers of points per wavelength depends on the type of finite element used and the level of accuracy sought in the simulation. Acceptable minimum values for p in the case of first-order (linear) elements vary between 8 and 12, but a minimum of ten points per wavelength is recommended, unless higher-order (quadratic) elements are used.

Adequately setting the size of elements, however, does not necessarily cover all the



Physics-Based Ground-Motion Simulation, Fig. 2 Conventional spatial grid for the velocity–stress FD scheme (Modified from Moczo et al. 2004)

geometrical and material irregularities present in simulation models, such as surface topography, material inhomogeneities, and internal structural interfaces. From a meshing point of view, the best alternative in FE to handle most of these problems is to work with conforming meshes. These are meshes that discretize the simulation domain adjusting both the shape and size of the elements – and thus the location of nodes – to conform to the specific characteristics in the geometry and material properties of the medium. Though powerful in this sense, conforming meshes are more difficult to build and require more memory (if the stiffness matrices of the elements are to be stored) or more computing time (if the stiffness matrices are to be rebuilt for each element at every time step). Nonconforming meshes, on the other hand, rely on the size of the element in order to capture, to a certain extent, the changes in the geometry and material properties of the media. While nonconforming meshes are very efficient because they can be tailored to use template elements whose stiffness matrices need to be computed only once and then scaled up based on material properties (e.g., Tu et al. 2006), this approach requires additional attention when it comes to handling sharp contrasts in material properties, discontinuities, or strongly irregular geometries. Some of these problems can be overcome using hybrid approaches that combine multiple element types or mixed meshing strategies (e.g., Hermann et al. 2011) or by means of special elements with extended or fictitious domains (e.g., Restrepo and Bielak 2014). A more detailed description of these alternatives, however, is out of the scope of this chapter.

Additional aspects pertaining to modeling the earthquake source and the effects of attenuation are discussed in subsequent sections.

The Finite-Difference Method

Attention is now given to the basic concepts and approach used for modeling wave propagation problems using the FD method.

In FD solutions, the derivatives in differential equations are approximated with finite differences computed over a discrete grid. Consider the case of a vertically propagating, planar *SH* wave in a horizontally layered medium. The plane strain approximation reduces the wave Eq. 3 to one dimension:

$$\mu \frac{\partial^2 u_x}{\partial z^2} + f_x = \rho \ddot{u}_x. \quad (14)$$

To eliminate the double derivatives, Eq. 14 is often expressed using the velocity–stress formulation:

$$\rho \frac{\partial v}{\partial t} = \frac{\partial \tau}{\partial z} + f_x, \quad (15)$$

$$\frac{\partial \tau}{\partial t} = \mu \frac{\partial v}{\partial z}, \quad (16)$$

where the velocity in the *x*-direction, $\dot{u}_x$, has been replaced with *v* and the shear stress component σ_{xz} with τ . FD solutions require discretization of both space and time on a numerical grid (Fig. 2). Partitioning the domain space using a mesh of $z_0, z_1, z_2, \dots, z_j$ with uniform increment Δz and the time using a mesh of $t_0, t_1, t_2, \dots, t_n$ with uniform increment Δt , the partial derivatives in Eq. 15 at point *j* and time step *n* may be approximated using

$$\frac{\partial v_j^n}{\partial t} \approx \frac{v_j^{n+1} - v_j^n}{\Delta t}, \quad (17)$$

$$\frac{\partial \tau_j^n}{\partial z} \approx \frac{\tau_{j+1}^n - \tau_j^n}{\Delta z}. \quad (18)$$

The approximations in Eqs. 17 and 18 use the forward difference formula. Substituting Eqs. 17 and 18 into Eq. 15 and omitting the body force term yield

$$\rho \frac{v_j^{n+1} - v_j^n}{\Delta t} = \frac{\tau_{j+1}^n - \tau_j^n}{\Delta z}. \quad (19)$$

Using analogous approximations for the temporal and spatial derivatives in Eq. 16 yields

$$\frac{\tau_j^{n+1} - \tau_j^n}{\Delta t} = \mu \frac{v_{j+1}^n - v_j^n}{\Delta z}. \quad (20)$$

By solving Eqs. 19 and 20 for v_j^{n+1} and τ_j^{n+1} , velocities and stresses at time $n + 1$ can be determined from the values at time n :

$$v_j^{n+1} = v_j^n + \frac{\Delta t}{\rho} \frac{\tau_{j+1}^n - \tau_j^n}{\Delta z}, \quad (21)$$

$$v_j^{n+1} = v_j^n + \frac{\Delta t}{\rho} \frac{\tau_{j+1}^n - \tau_j^n}{\Delta z}. \quad (22)$$

Solutions of the wave equation require knowledge of the initial velocities and stresses at time $n = 0$, v_j^0 and τ_j^0 , respectively. By using Eqs. 21 and 22 for all j , velocities and stresses can iteratively be determined for $n = 1, 2, 3, \dots$ until the desired time step.

The finite-difference scheme in Eqs. 21 and 22 is *conditionally stable*. The solution converges only if the Courant number

$$C = \frac{\beta \Delta t}{\Delta z} \leq C_{\max}, \quad (23)$$

where $\beta = \sqrt{\mu/\rho}$ is the shear-wave velocity and $C_{\max} = 1$ for the FD scheme in Eqs. 21 and 22. Equation 23 is called the *Courant–Friedrichs–Lewy* condition.

A disadvantage of the forward difference formula is that the approximations to the temporal Eq. 17 and spatial Eq. 18 derivatives are not

symmetric with respect to the grid point of interest. A more accurate approximation of the partial derivatives in Eq. 15 can be obtained using the *central difference formula*:

$$\frac{\partial v_j^n}{\partial t} \approx \frac{v_j^{n+1} - v_j^{n-1}}{2\Delta t}, \quad (24)$$

$$\frac{\partial \tau_j^n}{\partial z} \approx \frac{\tau_{j+1}^n - \tau_{j-1}^n}{2\Delta z}. \quad (25)$$

This leads to a numerical scheme that depends both on values from the current time step n and the previous time step $n - 1$:

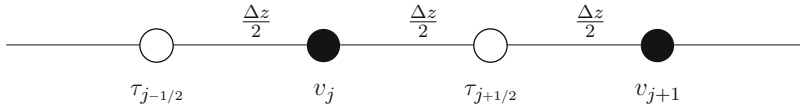
$$v_j^{n+1} = v_j^{n-1} + \frac{2\Delta t}{\rho} \frac{\tau_{j+1}^n - \tau_{j-1}^n}{2\Delta z}, \quad (26)$$

$$\tau_j^{n+1} = \tau_j^{n-1} + 2\Delta t \mu \frac{v_{j+1}^n - v_{j-1}^n}{2\Delta z}. \quad (27)$$

To solve Eqs. 26 and 27, the initial values at $n = 0$ and $n = 1$ must be known. Additionally, storing the velocities and stresses from both the current and the previous time step increases memory requirements. Approximating the spatial derivatives with the central difference formula and the temporal derivatives with the forward difference formula may seem a convenient alternative. However, such a combination leads to a FD scheme that is *unconditionally unstable*, i.e., it will not converge regardless of the value of the Courant number C . The stability of a numerical scheme is often analyzed using the *von Neumann method*, which is based on Fourier decomposition of the numerical solution.

A more efficient numerical scheme is obtained by staggering the position of stresses and velocities on the temporal and spatial grid as shown in Fig. 3 (Virieux 1984). By shifting the grid position for the velocities v by $\frac{1}{2}$ grid point in time and also shifting the grid position of the stresses τ by $\frac{1}{2}$ grid point in space, Eqs. 15 and 16 can be approximated using

$$\rho \frac{v_j^{n+\frac{1}{2}} - v_j^{n-\frac{1}{2}}}{\Delta t} = \frac{\tau_{j+\frac{1}{2}}^n - \tau_{j-\frac{1}{2}}^n}{\Delta z}, \quad (28)$$



Physics-Based Ground-Motion Simulation, Fig. 3 Staggered spatial grid for the velocity–stress FD scheme (Modified from Moczo et al. 2004)

$$\frac{\tau_{j+1/2}^n - \tau_{j-1/2}^{n-1}}{\Delta t} = \mu \frac{v_{j+1}^{n-1/2} - v_j^{n-1/2}}{\Delta z}. \quad (29)$$

The required initial conditions for this scheme are the velocities at time $1/2$, $v_{1/2}$, and stresses at time 0, τ_j^0 . Moving forward, the stresses at time n are computed first using

$$\frac{\tau_{j+1/2}^n - \tau_{j+1/2}^{n-1}}{\Delta t} = \mu \frac{v_{j+1}^{n-1/2} - v_j^{n-1/2}}{\Delta z}, \quad (30)$$

and then the velocities at time $n + 1/2$ are determined using

$$v_j^{n+1/2} = v_j^{n-1/2} + \frac{\Delta t}{\rho} \frac{\tau_{j+1/2}^n - \tau_{j-1/2}^n}{\Delta z}. \quad (31)$$

The forward difference formula in Eq. 18 represents the first-order approximation to the first spatial derivative, while the central difference formula (25) represents the second-order approximation. Both schemes can be derived by replacing the functional values at $f(z_0 \pm \Delta z)$ with a Taylor expansion:

$$\begin{aligned} f(z_0 \pm \Delta z) &= f(z_0) \pm f'(z_0)\Delta z \\ &\pm f''(z_0) \frac{\Delta z^2}{2} \pm f'''(z_0) \frac{\Delta z^3}{3!} \\ &+ O(\Delta z^4). \end{aligned} \quad (32)$$

The order of a numerical scheme is determined by the order of the truncation error, which is defined as the difference between the exact solution and the finite-difference approximation. Higher-order approximations may also be derived. For example, the fourth-order approximation in space combined with the second-order approximation in time has become popular for simulating ground motion (e.g., Olsen 1994; Graves 1996).

All the FD schemes discussed above Eqs. 21, 22, 26, 27, 30, and 31 are explicit, i.e., the velocities (or stresses) at a given grid point and time step are derived only from stresses and velocities of the previous time step(s). In implicit FD schemes, the velocities (stresses) at a given time step depend on velocities and stresses from both the current and previous time step. Implicit FD schemes are more difficult to solve and not frequently used in ground-motion modeling.

FD solutions to the wave Eq. 3 in two and three dimensions approximate the spatial and temporal derivatives in the same way. In computational grids used for 2D and 3D FD solutions, each element in the velocity vector and the stress tensor may be staggered with respect to the other elements (e.g., Graves 1996).

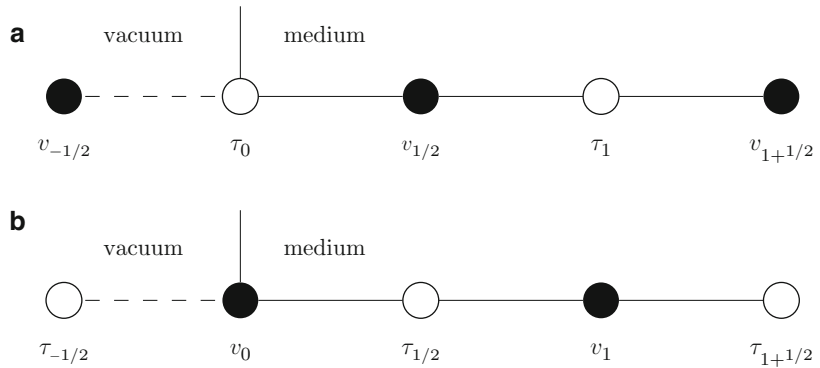
The FD solution derived above discusses only wave propagation in a continuum. In a real-world situation, discontinuities will be encountered at the free surface, at the boundaries of the computational domain, and at the contact between two different media in a heterogeneous medium.

Such internal material discontinuities can be treated using a homogeneous or a heterogeneous approach. In a homogeneous approach, boundary conditions at or near interfaces are explicitly discretized using a separate FD scheme, which is not a trivial problem. Therefore, most ground-motion prediction applications use a heterogeneous approach, where only one FD scheme is used for all internal grid points regardless of their distance to internal interfaces (Moczo et al. 2004). Such heterogeneous approaches typically define effective material parameters to improve accuracy near interfaces (e.g., Zahradnik et al. 1993).

In the above example of a vertically propagating *SH* wave, consider the interface generated by a horizontally layered sedimentary deposit resting on top of a denser, higher-velocity bedrock.

Physics-Based Ground-Motion Simulation,

Fig. 4 Free surface defined to coincide with position of (a) shear stress and (b) velocity (Modified from Moczo et al. 2004)



Assume that the interface coincides with the grid point defining the stress $\tau_{j+1/2}$ in Fig. 3. It can be shown (Moczo et al. 2002) that the boundary conditions at the interface can be fulfilled by defining the shear modulus at the interface as the harmonic average of the shear moduli from the two connected media:

$$\mu_{j+1/2}^H = \frac{2}{\frac{1}{\mu_j} + \frac{1}{\mu_{j+1}}}. \quad (33)$$

If the interface coincides with the grid point defining the velocity v_j (Fig. 3), on the other hand, the density ρ at the interface Eq. 31 must be replaced with the arithmetic average of the two materials (Moczo et al. 2002),

$$\rho_j^A = \frac{1}{2} \left(\rho_j - 1/2 + \rho_j + 1/2 \right). \quad (34)$$

Similar averaging methods are implemented in most heterogeneous 2D and 3D FD codes used in research today (e.g., Graves 1996; Cui et al. 2010).

In contrast to FE methods, the free surface requires special attention in FD methods. The boundary conditions at the free surface specify that the shear stress vanishes, i.e., $\tau = 0$. In the simple 1D case described earlier, the location of the free surface can be defined such that it intersects the position of the shear stress (Fig. 4a) or the position of the velocity (Fig. 4b). In the former case, the shear stress at the surface is explicitly set to $\tau_0 = 0$ during each iteration, and the velocity half a grid point below the surface is

calculated using Eq. 31. If the free surface coincides with the position of the velocity, v_0 (Fig. 4b), antisymmetry is used to ensure the traction-free boundary condition at the free surface (Levander 1988):

$$\tau_{-1/2} = -\tau_{+1/2}. \quad (35)$$

When staggered finite-difference grids are defined in two and three dimensions, a planar free surface will intersect both stresses and velocities. In that case, both approaches are employed to ensure that stresses vanish at the free surface (e.g., Graves 1996; Gottschämer and Olsen 2001; Moczo et al. 2004).

Irregular (nonplanar) free-surface boundary conditions are far more difficult to implement in FD methods. Typically, such methods require a much finer sampling of the wave field for accurate results (e.g., Robertsson 1996; Ohminato and Chouet 1997).

Similar to FE methods, absorbing boundary conditions are required at the bottom faces of the computational domain. Both damping zones (e.g., Cerjan et al. 1985) and perfectly matched layers (e.g., Marcinkovich and Olsen 2003) have been implemented in FD codes for simulation of strong ground motion.

Intrinsic Attenuation and Plasticity

The methods for solving wave propagation problems just described apply only to linear elastic conditions. However, the accurate representation

of seismic waves requires the consideration of energy losses due to internal friction or intrinsic attenuation and, when earthquake induced deformations are large enough, those due to plastic deformation as well. Both these losses are important because their omission may lead to the overestimation of the amplification and duration of seismic waves in regions with high dissipative properties, for attenuation effects in general, and in regions with soft materials that have low yielding limits or are exposed to large-magnitude earthquakes, in the case of plastic deformation. This section deals with the basic most common approaches used to include realistic attenuation and plasticity in simulations.

Intrinsic Attenuation

The general formulation of the evolution of linear isotropic viscoelastic material in time is governed by the stress–strain relation, in which the stress can be expressed as a convolution of the strain rate with a relaxation function as in

$$\sigma(\mathbf{x}, t) = \int_0^t \varphi(\mathbf{x}, t - \tau) \dot{\varepsilon}(\mathbf{x}, \tau) d\tau, \quad (36)$$

which is equivalent to

$$\sigma(\mathbf{x}, t) = \varphi(\mathbf{x}, t) * \dot{\varepsilon}(\mathbf{x}, t) \quad (37)$$

or

$$\sigma(\mathbf{x}, t) = \dot{\varphi}(\mathbf{x}, t) * \varepsilon(\mathbf{x}, t), \quad (38)$$

where the symbol $*$ is used to represent the convolution integral. $\sigma(\mathbf{x}, t)$ and $\varepsilon(\mathbf{x}, t)$ are the stress and strain states at a point $\mathbf{x}$ in time t , and $\varphi(\mathbf{x}, t)$ is the corresponding stress relaxation function. Dots on top indicate derivatives in time. Equations 37 and 38 are the same because of the differentiation properties of convolution.

The solution of the wave propagation problem in viscoelastic media in the time domain which follows from the direct substitution of Eq. 36 in Eq. 1 is, however, inconvenient because that would entail the computation of a convolution term at every time step. Such an approach would require the storage of the complete strain

history, making the computational implementation of the viscoelastic problem practically intractable.

It is clear from Eq. 38 that the formulation of the viscoelastic problem in the frequency domain is, on the other hand, straightforward. Applying the Fourier transform, Eq. 38 becomes

$$\sigma(\omega) = M(\omega) \varepsilon(\omega), \quad (39)$$

where

$$M(\omega) = \dot{\varphi}(\omega) \quad (40)$$

is understood as a frequency-dependent viscoelastic modulus. Note that, for simplicity, the spatial variable ($\mathbf{x}$) has been dropped. In general, $M(\omega)$ is defined as a complex quantity and it is such that

$$\lim_{\omega \rightarrow 0} M(\omega) = M_R, \quad (41)$$

$$\lim_{\omega \rightarrow \infty} M(\omega) = M_U, \quad (42)$$

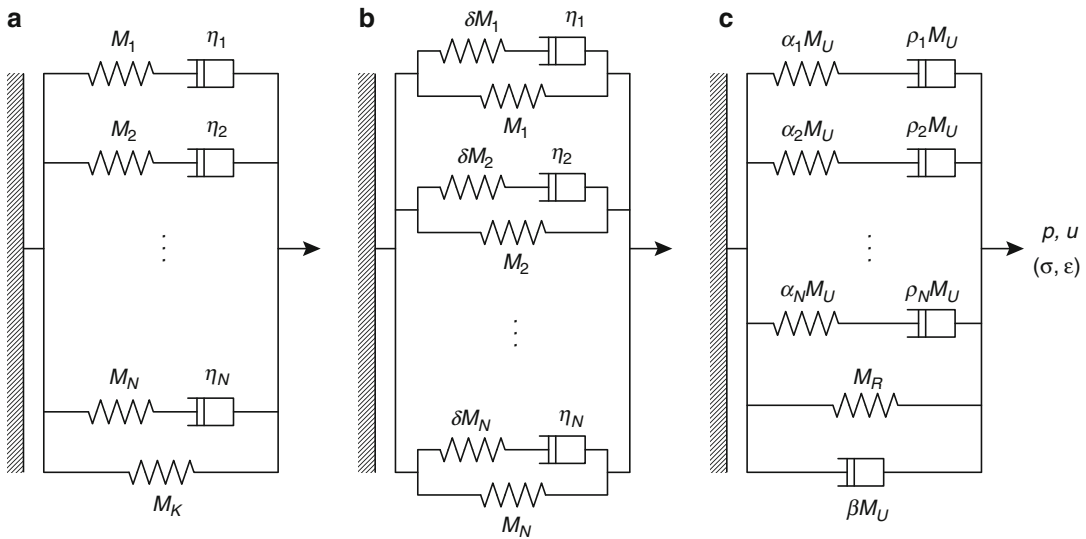
where M_R and M_U are defined as the relaxed and unrelaxed material viscoelastic moduli. They correspond to the long-term equilibrium and instantaneous elastic response of the material and together define the relaxation modulus:

$$\delta M = M_U - M_R. \quad (43)$$

In practice, the viscoelastic modulus is expressed in terms of the material's quality factor $Q(\omega)$, which is defined as

$$Q(\omega) = \frac{\Re[M(\omega)]}{\Im[M(\omega)]}. \quad (44)$$

Understanding the formulation of anelasticity in the frequency domain facilitates its implementation in the time domain – where the system does not have to be fully assembled as in the frequency domain. The challenges in the formulation of the anelastic wave propagation problem in the time domain are (i) to solve the convolution term efficiently and (ii) to satisfy the behavior of the



Physics-Based Ground-Motion Simulation, Fig. 5 Examples of rheological models used to incorporate the effect of attenuation. (a) Generalized Maxwell model (Emmerich and Korn 1987), (b) generalized

Zener body (Carcione et al. 1988), and (c) a generalized Maxwell model augmented with a viscous damper (Bielak et al. 2011)

quality factor Q in the frequency domain – which for the range of frequencies typically covered by most simulations is considered to be constant in the low frequencies and frequency dependent in the higher frequencies.

Two particular works established the ground for most current approaches used to model the effects of anelasticity in time-stepped solutions (for frequency-independent Q): Liu et al. (1976) and Day and Minster (1984). Liu et al. (1976) were the first to use a rheological model made of a set of mechanical bodies (Zener mechanisms) in the context of seismic problems; and Day and Minster (1984) showed that if $M(\omega)$ is expressed as a rational function, then its inverse form in the time domain can be solved numerically – which then led to the formulation of efficient algorithms to account for anelastic losses.

Following the ideas set forth by Day and Minster (1984), in which a set of internal memory variables are used to represent the relaxation process, others have formulated memory-efficient approaches to anelastic wave propagation simulation using FD and FE approaches (e.g., Day and Bradley 2001; Ma and Liu 2006).

Emmerich and Korn (1987) proposed the use of a rational function $M(\omega)$ corresponding to a rheological model composed of a set of Maxwell bodies in parallel and a Hooke element also in parallel. They defined this model as the generalized Maxwell body (GMB) (Fig. 5a). Similarly, Carcione et al. (1988) employed a generalized Zener body (GZB) (Fig. 5b). Various separate implementations have been inspired in these two models, which were later shown to be equivalent (Moczo and Kristek 2005). Some implementation examples using FD, DG, and SE methods are those described in Chaljub et al. (2010). Most of these are formulated in terms of stresses and strains. More recently, Bielak et al. (2011) introduced a memory-efficient internal friction approach solely based on displacements more suitable for FE. The model proposed by Bielak et al. (2011) uses a set of Maxwell elements in parallel and a Voigt element also in parallel (Fig. 5c).

To correlate the problem of anelastic attenuation with the solution of the wave propagation problem in elastic media shown in the previous section, the model proposed by Bielak et al. (2011) is summarized next for the case of

a FE approach. Consider a semidiscretized version of the equations of elastodynamics where a mapping operator T is introduced to modify the displacements associated with the internal body forces as in

$$\mathbf{M}\ddot{\mathbf{u}} + \sum_e [\mathbf{K}^e(T\mathbf{u})^e] = \mathbf{f}. \quad (45)$$

Here, T is such that it represents the convolution term used in the stress–strain relation, which after applying finite elements becomes the stiffness–displacement product $\mathbf{K}^e(T\mathbf{u})^e$ given by

$$\mathbf{K}^e(T\mathbf{u})^e = \mathbf{K}^e \left[\mathbf{u}^e + \beta^e \dot{\mathbf{u}}^e - \sum_j \alpha_j^e \beta_j^e \exp(-\gamma_j^e t) * \mathbf{u}^e \right]. \quad (46)$$

Note that the convolution term is still present in Eq. 46. This term is replaced by the auxiliary memory variable φ , such that at any time step n with a discrete time-width Δt , the product $\mathbf{K}^e(T\mathbf{u})^e$ is given by

$$\mathbf{K}^e(T\mathbf{u})^e = \mathbf{K}^e \left[\mathbf{u}^e + \beta^e \frac{\mathbf{u}_n^e - \mathbf{u}_{n-1}^e}{\Delta t} - \sum_j \alpha_j^e \beta_j^e (\varphi_j^e)_n \right], \quad (47)$$

where

$$\begin{aligned} (\varphi_j^e)_n = & \frac{\Delta t}{2} \left[(1 - \Delta t \gamma_j^e) \mathbf{u}_n^e + \mathbf{u}_{n-1}^e \right] \\ & + \exp(-\gamma_j^e \Delta t) (\varphi_j^e)_{n-1}. \end{aligned} \quad (48)$$

α , β , and γ are constants associated with the mechanical elements of the adopted model (i.e., spring and dashpot constants) that are determined for each finite element e based on the quality factor Q of the material contained in the element. These constants can be derived separately for the quality factors associated with the propagation of P and S waves, Q_P and Q_S . Details about the particular implementation of this model can be found in Bielak et al. (2011). Information about the derivation of the values of Q_P and Q_S is provided in the section on “Material Models”.

Plasticity

The problem of wave propagation in elastoplastic and elasto-visco-plastic media has also been considered in 3D physics-based earthquake simulations, but has not yet reached the same level of maturity as anelastic wave propagation. This is mainly due to the computational complexity involved in incorporating full 3D nonlinear soil behavior in the solution scheme and the entailed computational overhead both in memory and processing time – as well as the difficulty to accurately reproduce ground motions in the frequency ranges where plastic behavior is believed to be more relevant (above 0.5 Hz).

Numerical modeling of the response of nonlinear sedimentary deposits also dates back to the 1960s and 1970s. Initial simulations considered 1D and 2D models and used linear equivalent methods to approximate the stress–strain relationship (e.g., Idriss and Seed 1968). Although the linear equivalent method continues to be used extensively in engineering research and practice, it has long been understood that it does not capture all the characteristics of nonlinear plastic behavior. Alternatively, there exist an abundant number of models based on more rigorous methods (e.g., Prevost 1978) to describe the cyclic stress–strain behavior of geomaterials. However, such models are usually defined in terms of material parameters upon which there is no consensus. In addition, they tend to be computationally expensive. This has made it difficult to apply rigorous plastic models in 3D regional earthquake simulations where the computational aspects are critical and the knowledge about the material properties at regional scales is limited – especially for the near-surface layers, where nonlinear soil effects are more significant.

An approach used to overcome the computational difficulties of incorporating near-surface plastic effects in physics-based simulations has been the combination of 3D linear elastic or anelastic simulation results with 1D nonlinear analyses of the incident motion at the interface between the bedrock and the sedimentary deposits (e.g., Roten et al. 2012). Hybrid

simulations, however, cannot completely describe the 3D aspects present when off-fault and near-surface plasticity is combined with the source, path, and basin effects. Alternatively, there have been successful first approximations to obtain full 3D regional-scale simulations that consider plastic deformations using computationally tractable material models such as the classical Drucker–Prager yield criterion (e.g., Dupros et al. 2010; Taborda et al. 2012; Roten et al. 2014). Although the material models used in these simulations do not accurately reproduce the elasto-visco-plastic behavior of most geomaterials, they are useful to understand the 3D nature of nonlinear wave propagation in heterogeneous media.

Following the formulations presented before for the case of the FE method, the solution of the wave propagation problem in plastic media can be described as follows. Consider the semidiscretized version of Navier’s Eqs. 1 through 6, but in a manner in which the stresses are preserved explicitly. In that case, Eq. 6 becomes

$$\mathbf{M}\ddot{\mathbf{u}} + \sum_e \int_{\Omega_e} \mathbf{B}^T \boldsymbol{\sigma} d\Omega_e = \mathbf{f}, \quad (49)$$

where $\mathbf{B}$ is the strain matrix and $\boldsymbol{\sigma}$ is the stress tensor over element Ω_e . The summation, again, means assembling of elements.

Some important points need to be noticed about Eq. 49. See that in order to obtain the contribution of the internal forces given by the product $\mathbf{B}^T \boldsymbol{\sigma}$ in the integral term, one must know the state of stresses. The stresses, however, depend on the state of total strain in the material ($\boldsymbol{\varepsilon}$), and the strains, in turn, need to be compatible with the stresses themselves according to the constitutive model chosen to represent the material’s plastic behavior. In the elastic problem, this relationship between the stress and strain tensors is linear. In plasticity, on the contrary and in general, it is not. Therefore, it follows from this that, embedded within the time integration of Eq. 49, there is a nonlinear problem that requires the implementation of an implicit solution

scheme, which in most cases requires additional computational effort.

In general, following the classical theory of plasticity, the total strain ($\boldsymbol{\varepsilon}$) can be expressed in terms of the sum of the elastic ($\boldsymbol{\varepsilon}^e$) and plastic ($\boldsymbol{\varepsilon}^p$) deformation components,

$$\boldsymbol{\varepsilon}_{ij} = \boldsymbol{\varepsilon}_{ij}^e + \boldsymbol{\varepsilon}_{ij}^p. \quad (50)$$

This is useful because one can then express the associated admissible stress in terms of the product of the elastic stiffness tensor (D) and the elastic strain,

$$\sigma_{ij} = D_{ijkl} \varepsilon_{kl}^e. \quad (51)$$

Equations 50 and 51 also need to be thought carefully. In solving Eq. 49, at any given time step, one can obtain the total strain from the current state of displacements. The objective then becomes finding its plastic and elastic components, so that the stress given by Eq. 51 remains compatible with the constitutive model and the total strain. Considering that by definition the elastic deformation component is bound to the stresses by the elastic stiffness tensor – as opposed to the plastic deformation which is unbounded in general – the critical point becomes finding the corresponding plastic deformation $\boldsymbol{\varepsilon}^p$.

This is where the choice of the material model plays its role. Constitutive models provide the means for tracking the progress of the plastic deformation through the combination of a yielding potential function (g) and the rate of the plastic strain ($\dot{\boldsymbol{\varepsilon}}_{ij}^p$) – that is, a function of how the plastic strain changes in time. Yielding potential functions are of the form

$$g(\sigma_{ij}, k) = F(\sigma_{ij}) - k(\sigma_{ij}, k_n) < 0, \quad (52)$$

where F represents the current state of stresses and k defines the hardening characteristics of the material. At the upper limit of Eq. 52, $F - k = 0$ represents a yielding surface which defines the plastic state of the material. In other words, the yielding surface provides a limiting state on which a material’s particle must remain while in

a plastic deformation condition. Below this surface, the particle behaves elastically, and on the surface, plastically. Here, the hardening characteristics controlled by k refer to the ability of some materials to regain strength after they have gone over the plastic domain. In perfectly elastoplastic materials, for instance, the yielding surface can be thought of as being “flat” and k is null. In materials with hardening characteristics, on the other hand, the yielding surface changes depending on the state of deformation. That is why k is also a function of σ . k_n is used here generically to represent the material parameters controlling hardening.

As noted in Eq. 52, both F and k are functions of the stresses and therefore depend on the plastic state of deformation. It follows from Eqs. 50, 51, and 52 that the key component toward finding the solution of Eq. 49 is determining the state of plastic deformation, which is given by the plastic potential strain rate:

$$\dot{\epsilon}_{ij}^p = \dot{\lambda} \frac{\partial h(\sigma_{ij}, k)}{\partial \sigma_{ij}} = \dot{\lambda} \frac{\partial g(\sigma_{ij}, k)}{\partial \sigma_{ij}}, \quad (53)$$

where $\dot{\lambda}$ is a plastic multiplier and h is a function which defines the plastic potential. As noted in the right-hand side of Eq. 53, for simplicity, h is a function similar to and often assumed to be the same as the yielding potential function g explained before. The plastic multiplier $\dot{\lambda}$ sets the magnitude of the plastic deformation in the current state of stresses as a material’s particle moves over the yielding surface.

As it can be inferred from this description, the solution of Eq. 49 leads to additional computations to help ensure that the stress–strain relationship is maintained according to the nonlinear constitutive model of choice. This requires the use of implicit solution schemes, as opposed to the explicit approach used in the elastic and viscoelastic problems. This additional effort explains why considering the material’s plastic behavior in physics-based ground-motion simulations in 3D has often been ignored or simplified using hybrid and indirect approaches. Future developments, however, are likely to revert this

trend, and progress is expected to be done in this area in the near future.

Source Models

As seen in the simulation workflow section, physics-based earthquake simulations depend on two basic models used to represent the earthquake itself and the propagation media, these are the source model and the material model, respectively. They, in turn, define the characteristics of the applied and internal body forces in the formulation of the solution of the wave propagation problem. This section deals with the basic concepts of source models used as input to earthquake simulations.

The fault’s rupture and the resulting wave propagation problem are seldom solved together in a single simulation, especially at regional scales. This is mainly due to the complexity of the rupturing process on the fault, which entails a multi-physics problem with plastic deformation and/or the use of dislocation models to represent the loss of friction/contact on the fault’s surface. Source models are, instead, resolved prior to performing the simulation of the ground motion and then treated in the simulation as basic input data as seen in the simulation workflow section. Some common alternatives from where source models are derived include:

- Source inversion studies
- Dynamic rupture simulations
- Pseudo-dynamic rupture generators
- Kinematic source model generators
- Basic geologic and seismogenic information about the fault

Regardless of the method employed to obtain the source model, the most common approach used in simulations is to convert the source model into a kinematic source representation, that is, a model that represents the source as a set of equivalent forces (or stresses) that are applied to the forward simulation model to trigger the propagation of seismic waves. These equivalent body forces are such that they produce

a displacement field away from the source that is equivalent to that one would have obtained using a more rigorous solution for the source if embedded in the simulation.

In the case of small-magnitude earthquakes ($M < 5$) or sources small enough compared to the wavelength of the radiated energy, the effect of the earthquake rupture and the discontinuity of displacements that occurs at the fault can be modeled using a single set of self-balanced (double-couple) forces acting on a point, that is, a *point source* model. In the case of large-magnitude earthquakes, on the other hand, the rupture occurs over extended areas on the fault's surface and thus cannot be treated as a point source. In such cases, the earthquake is modeled as the sum of many point sources. The approach relies on the idea that it is possible to discretize the fault's ruptured area into a collection of smaller subfaults, each with an assigned point source model. The subfaults are such that they adjust to the geometry of the entire fault and the collective action of the point sources adds up the right amount of energy release, equivalent to that of the complete earthquake model. Extended fault models composed of multiple subfaults are called *finite slip* or *finite source* models.

In both point and extended fault models, the point source at each (sub)fault is defined in terms of its geometry and rupture characteristics. These are given by:

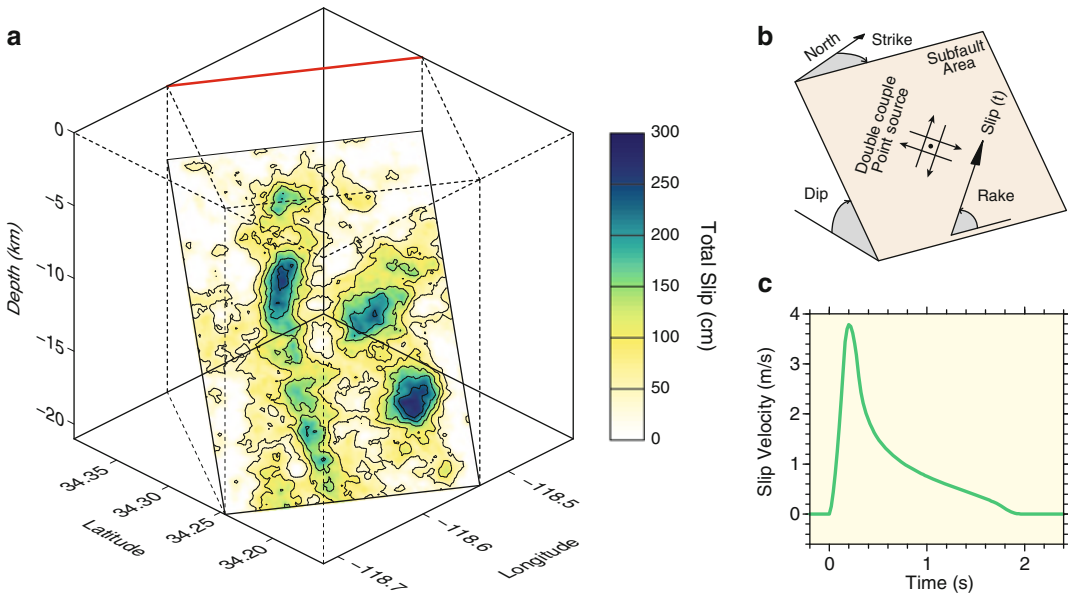
- The location given in latitude, longitude, and depth
- The orientation given by the strike, dip, and rake angles of the (sub)fault
- The (sub)fault's area and average shear modulus of that area
- The evolution of the slip on the (sub)fault with time

Figure 6 shows an example of a kinematic source model. Part (a) shows the total slip distribution on the fault plane of the source model by Graves and Pitarka (2010) for the 1994 M_w 6.7 Northridge, California, earthquake. This model is composed of 140×140 subfaults with strike and dip angles of 122° and 40° , respectively, and

variable rake angle with an average of 101° . Part (b) shows the concept of the discrete subfault as used in a kinematic finite slip model in which each rectangular (or triangular) patch on the fault plane has an independent geometry (area, strike, dip, rake) and slip (as a function of time). And part (c) shows a typical slip velocity function, which is the function that defines the history of slip associated with the point source in each subfault area.

Point source models used in physics-based simulations can be simply built based on the earthquake focal mechanism and the selection of an appropriate slip function such as that shown in Fig. 6c. Extended fault models, on the other hand, are available from source inversions and rupture generators, and distributed in various formats. The US Geological Survey Earthquake Hazards Program, for instance, offers finite fault models for significant earthquakes in its Historical Earthquake Information database Web portal. Another useful source of fault models is the Finite-Source Rupture Model Database (SRCMOD) maintained by the eQuake-RC project Web site (<http://equake-rc.info/>). Models distributed by SRCMOD are contributed by researchers from all over the world and distributed in a set of common data files that include basic metadata and simple single-rupture-plane source-model representations. Another popular distribution format among modelers, especially in the USA, is the Standard Rupture Format (SRF) used by Graves and Pitarka (2010). The SRF encapsulates the rupture process in a single (ascii) text file that can have multiple fault planes and subfaults with variable geometry.

One important aspect to note as simulations advance toward higher frequencies is the influence that the source type and source model description have on the characteristics of the ground motion. If seen in the frequency domain, for instance, the slip-rate function of a single point source (as that shown in Fig. 6c) will reveal that the energy of the slip is mostly contained below a certain frequency. This, together with the seismic velocities represented in the model, will influence the energy distribution of the ground motions in the frequency domain,



Physics-Based Ground-Motion Simulation, Fig. 6 Example of a kinematic source model: (a) finite slip representation of the 1994 M 6.7 Northridge earthquake composed of 140×140 subfaults (Modified after

Graves and Pitarka 2010), (b) subfault model concept for a double-couple point source with independent geometry and source-time function, and (c) a typical slip-rate function in time

provided models are built with appropriate accuracy. In other words, and it should come as no surprise, the source spectrum is important in determining the frequency content of the ground motion.

Furthermore, in extended source models, the smoothness and homogeneity of the subfault characteristics also influence the outcome. If the distribution of total slip on the subfaults is fairly homogeneous, or if the evolution of the slip through the fault plane follows a certain pattern in space and time, then the wave field will evolve more smoothly as it travels away from the source plane. It will be more coherent. On the other hand, if each subfault slips randomly, the ground motion will be less coherent, and waves will be more likely to interfere with each other as they travel away from the fault. Similarly, the orientation of the slip on the plane, given by the rake angle, will dictate the strength of directivity effects. Additional complexity can be introduced if we also consider the fact that subfaults do not necessarily have to be aligned. Figure 5a depicts extended source models as a collection of point sources with variable rake angle and slip, but

constant strike and dip angles, with all the subfaults being part of a single fault plane. In reality, however, faults are not smooth planes, but irregular contact areas. This increases the variability of the ground motion considerably, as shown by recent models developed by Shi and Day (2013) to account for the geometrical heterogeneity (roughness) of the fault.

As mentioned above, these are all important factors that modelers need to consider, especially when simulating ground motions at high frequencies (>1 Hz), because short wavelengths can better capture the variations of the fault structure and rupture characteristics.

Material Models

Provided a source model, the second necessary component for a simulation is the material model, which defines the mechanical properties of the propagating media in the modeling domain. The most basic 3D material model used for elastic wave propagation simulations defines the material density (ρ) and seismic velocities of P and S waves

(V_P and V_S , respectively) at any arbitrary point within the simulation domain. Anelastic simulations require, in addition, the attenuation properties of the propagating media. The material's attenuation properties are defined in terms of the quality factors Q_P and Q_S associated with the attenuation characteristics of P and S waves, respectively. These quality factors are defined using attenuation rules or attenuation relationships which, in the context of physics-based ground-motion simulation, are empirical functions based on the values of V_P and/or V_S . For elastoplastic simulations, material models must also provide the parameters that define the characteristics of the adopted nonlinear constitutive model. Since elastoplastic material properties are unique to the adopted constitutive model, they are not described here; besides, complete definitions of all the mechanical properties of the propagating media in a single model are uncommon. Instead, for the case of anelastic ground-motion simulation, material models are divided in two parts: (a) seismic velocity models and (b) attenuation relationships. These are described next.

Seismic Velocity Models

Seismic velocity models, also known as community velocity models (CVMs), are datasets or computer programs that define the values of V_P , V_S , and ρ at any point in a particular region of interest. The point's location is usually expressed in terms of latitude, longitude, and depth (with respect to the free surface) or elevation (with respect to the sea level). Internally, CVMs are built using different types of datasets, which may include but are not limited to:

- Source inversion studies
- Surface topography or digital elevation maps
- Subsurface topography or geological horizons
- Gravity observations and refraction surveys
- Teleseismic and 3D tomographic inversions
- Mantle and Moho 1D background models
- Empirical rules correlating V_P , V_S , and ρ
- Shallow and deep boreholes and observations from oil wells

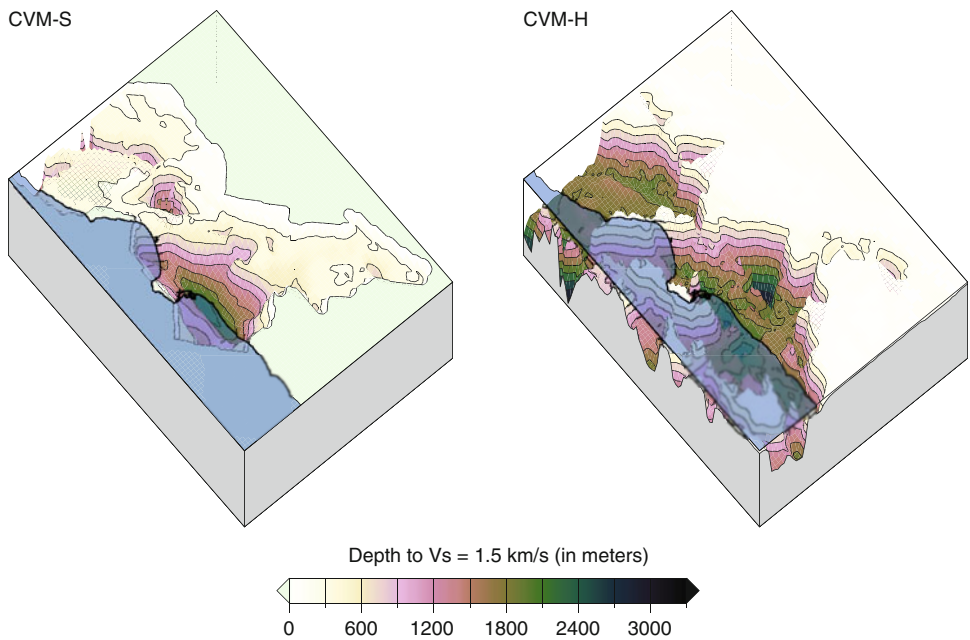
- Geotechnical layer models based on indirect measures (e.g., V_{S30} data)
- Random media representations

Some velocity models are built using voxels (volume elements) that define homogeneous 3D regions. Other models combine different datasets using interpolation rules that assign weights based on the location of the point of interest relative to the data points.

Most CVM computer codes are not necessarily optimized to work seamlessly with earthquake simulation codes. Modelers use additional tools to convert the data retrieved from a given CVM into grids or meshes. One particular tool of the sort is the unified community velocity model (UCVM) software framework (Small et al. 2015). UCVM is a collection of software tools developed and maintained by the Southern California Earthquake Center (SCEC) to provide an efficient and standard access to multiple, alternative velocity models. Although UCVM was primarily built to manage the SCEC community velocity models CVM-S and CVM-H for Southern California (Fig. 7), it supports and can be used to register other models as well.

There exist a good number of velocity models available to the community. It is, however, not possible to cover them all here in detail. A selection of some of the most relevant models used in simulations in the USA, Europe, and Japan is summarized in Table 1.

As simulations aim to produce more realistic ground motions comparable to observations, an important aspect in the construction and use of velocity models is that of the representation of the geotechnical layers and the variability of the material properties at small scales. For the most part, despite their level of detail, seismic velocity models tend to be smooth representations of the crustal structure. This is primarily due to the fact that geologic, exploration, and other survey data are only available at coarse resolutions. Modelers need then to resort to other methods to represent the presence of near-surface soft-soil deposits and the random characteristics of geomaterials. The presence of soil deposits is typically handled by introducing some kind of geotechnical layer



Physics-Based Ground-Motion Simulation, Fig. 7 Comparison of the two Southern California community velocity models CVM-S and CVM-H. Depth of

the major basins in the Greater Los Angeles region (in meters) as inferred from the isosurface for $V_S = 1.5$ km/s

Physics-Based Ground-Motion Simulation, Table 1 Summary of velocity models

Model	Region	Available at
CVM-S	Southern California (SCEC Model)	http://sceec.usc.edu/scecpedia/
CVM-H	Southern California (Harvard Model)	http://sceec.usc.edu/scecpedia/
CenCalVM	Central California (San Francisco Bay Area)	http://earthquake.usgs.gov/research/structure/3dgeologic/
WFCVM	Wasatch Front (Salt Lake City, Utah)	http://geology.utah.gov/ghp/consultants/geophysical_data/cvm.htm
CUSVM	Central U.S. (New Madrid Seismic Zone)	http://earthquake.usgs.gov/research/cus_seisvelmodel/
J-SHIS	Japan Substructure	http://www.j-shis.bosai.go.jp
JIVSM	Japan Integrated Velocity Structure Model	See Koketsu et al. (2009)
Grenoble Basin	Grenoble Basin, France	See Chaljub et al. (2010)
Mygdonian Basin	Northern Greece (Euroseistest)	http://euroeisdb.civil.auth.gr/

model (GTL), which softens the transition from the rock basement to the surface. GTLs use empirical rules to estimate the near-surface (V_{S30}) seismic velocities and then interpolate the material properties from surface to depth. The interpolation is parameterized using existing

borehole profiles (Ely et al. 2010). The variability of the medium, on the other hand, is introduced by means of random spatially correlated perturbations to the model at small-scale resolutions (Hartzell et al. 2010; Withers et al. 2013). Both these elements have shown to have a significant

effect on the ground motion, especially at higher frequencies, thus their importance in the future of physics-based simulation.

Attenuation (Q) Relationships

As mentioned before, the quality factors associated with P and S waves, Q_P and Q_S , are most commonly defined based on the values of the seismic velocities V_P and V_S . The value of Q_S is usually defined from rules that depend directly on the value of V_S . Typical forms of Q_S – V_S relationships are piecewise linear and continuous polynomial functions. The value of Q_P , on the other hand, is usually defined in terms of Q_S and, in some cases, in terms of the velocity contrast V_P/V_S . Table 2 shows a collection of different Q_S – V_S and Q_P – Q_S relationships used in simulations. Each relationship is listed along with a publication, the earthquake or scenario event for which they were employed, and the minimum shear-wave velocity $V_{S_{\min}}$ and maximum frequency ($f_{\max}$) that was used in the associated simulations. Some of the references provided in Table 2 include notes (see superscripts) to clarify the nature, region, or scope within which the relationship was introduced or used. The FD and FE superscripts, in particular, indicate when a reference is being cited specifically in reference to the results published therein using one of these methods. Aagaard et al. (2008) or Bielak et al. (2010), for instance, include simulations using several methods; thus the superscripts indicate which method was used with the relationship associated with these studies.

As it can be seen from this table, there is no consensus on the most appropriate set of relationships between seismic velocities and quality factors. Notice also that the majority of the relationships are independent of depth (z) and all are independent of the frequency, even though it is known that these relationships are both depth and frequency dependent.

Here, the frequency dependence of Q is particularly important as it is understood to be of greater significance at frequencies above 1 Hz. Up until recently, this was not a major factor in physics-based ground-motion simulation because

simulations were typically done for maximum frequencies no greater than 1 Hz. However, with the increasing capacity of supercomputers, modelers are now more often being able to simulate ground motions at higher frequencies, thus its relevance in future efforts. In this case, the main ideas of the viscoelastic methods described in the “[Intrinsic Attenuation](#)” section will still apply, as they are formulated in frequency. Their current implementation in computer codes used in simulations, however, had been usually calibrated to adjust to constant target values of Q . Therefore, in years to come, the implementation of these viscoelastic models and the relationships shown in Table 2 will need to be revised to offer a consistent approach to modeling high-frequency (0–10 Hz) ground motions. This is a topic of current research and ongoing efforts such as those reported by Withers et al. (2014) offer a glimpse of the future in this regard.

Recent Examples and Applications

There is an ample spectrum of physics-based simulations available in the literature. Most of them have been published over the last two decades after supercomputing centers open for public research became available in the mid- to late 1990s and early 2000s, which boosted the capacity of modelers to conduct regional-scale simulations at resolutions not possible before parallel computer codes were developed and tested. Some of these simulations have already been cited to illustrate the various aspects involved in physics-based simulation. However, they cannot all possibly be covered here, thus only a small selection is addressed next. The selection includes examples of scenario and real earthquake simulations that have been used in verification and validation studies; those are the simulations of the Great Southern California ShakeOut and the 2008 Chino Hills, California, earthquake. Also covered here to a lesser extent is the application of physics-based simulation as a tool to construct a physics-based framework for probabilistic seismic hazard analysis, as it is done in the CyberShake project of the Southern California Earthquake Center.

Physics-Based Ground-Motion Simulation, Table 2 Examples of Q_S – V_S and Q_P – Q_S relationships used in past physics-based simulations. Most of these correspond to the region of Southern California and to past earthquakes, unless noted differently. FD and FE superscripts indicate when a publication is cited specifically in reference to results published therein using a particular method

Publication	Simulation	V_{Smin} (m/s)	f_{max} (Hz)	$Q_S = g(V_S)$ (V_S in km/s, z in km)	$Q_P = h(Q_S)$
Olsen et al. (2003)	1994 Northridge	500	0.5	$20V_S$ $100V_S$	$V_S < 1.5$ $V_S \geq 1.5$ $1.5Q_S$
Bielak et al. (2010) ^{FD}	ShakeOut ^a	500	0.5	$50V_S$	$2Q_S$
Cui et al. (2010)	M8 ^a	400	2.0		
Graves et al. (2011)	CyberShake ^a	500	0.5		
Komatitsch et al. (2004)	2001 Hollywood 2002 Yorba Linda	670	0.5	90 ∞	Sediments Bedrock ∞
Taborda et al. (2007) ^b	ShakeOut ^a	200	1.0	$50V_S$	
Bielak et al. (2010) ^{FE,b}	ShakeOut ^a	500	0.5		
Graves (2008)	2001 Big Bear	250	1.0	$60V_S$ $50V_S$	$V_S < 0.9$
Aagaard et al. (2008) ^{FD}	1989 Loma Prieta	330–760	0.5–1.0	$60V_S^{1.5}$	$0.9 \leq V_S < 3.4$
				500	$V_S \geq 3.4$
				13	$V_S < 0.3$
Brocher (2008) ^c	–	–	–	$-16 + 104.13V_S$	
				$-25.225V_S^2$	$0.3 \leq V_S < 5$
				$+8.2184V_S^3$	
Chaljub et al. (2010) ^d	2003 Lancey, and Event S1 ^a	300	2.0	50 ∞	$z < 1$ $z \geq 1$ $3/4(V_P/V_S)^2Q_S$ ∞
Taborda and Bielak (2013)	2008 Chino Hills	200	4.0	$10.5 - 1.6V_S + 153V_S^2 - 103V_S^3$ $+ 34.7V_S^4 - 5.29V_S^5 + 0.31V_S^6$	$3/4(V_P/V_S)^2Q_S$

^aDenotes scenario events

^bRayleigh damping instead of a viscoelastic model

^cEmpirical relations (no simulation) for Northern California

^dSimulations for Grenoble Valley, France

^{FE}Finite-element simulation therein

^{FD}Finite-difference simulation therein

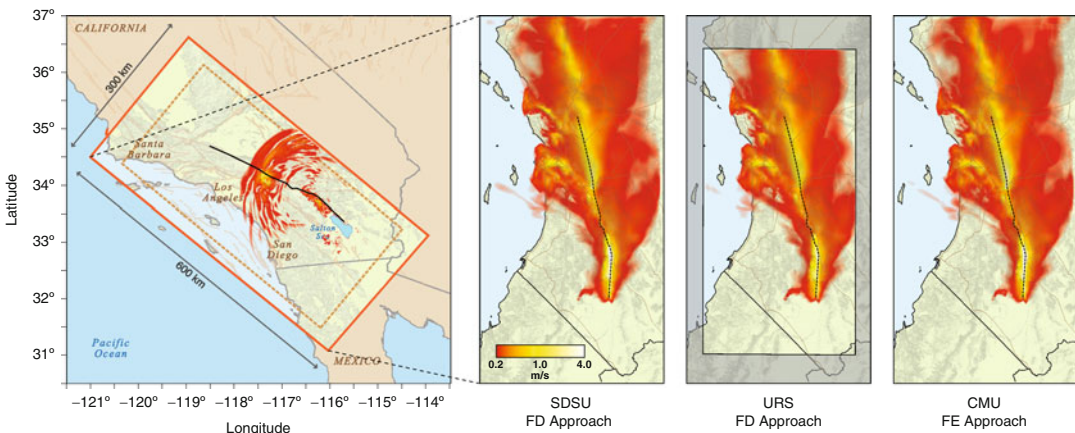
ShakeOut

The 2008 Great Southern California ShakeOut was a multidisciplinary earthquake preparedness and emergency management exercise involving Earth science, engineering, and social sciences which has now transcended to an annually repeated drill held in many other US and worldwide seismic regions. In its first edition, the ShakeOut included the definition of an M 7.8 scenario earthquake rupturing the southern segment of the San Andreas Fault. The regional ground motion over an area of 600×300 km covering all major cities in Southern California for the ShakeOut scenario was computed using physics-based ground-motion simulation. A physics-based approach was chosen over empirical ground-motion prediction equations because the latter were considered poorly constrained for such an event. Details about the scenario and results can be found in Jones et al. (2008). The predicted ground motions helped estimate the dynamic response of buildings, losses, casualties, and the socioeconomic impacts of such an earthquake and are described in a special issue of *Earthquake Spectra* (Porter et al. 2011).

The simulations of ground motions for the ShakeOut were independently carried out by three different groups. Two groups employed

FD codes and one group used a FE code. The results were presented in Bielak et al. (2010) and verified using different comparison methods. Ground motions predicted by the three codes (Fig. 8) were found to be in good agreement with each other independently of the differences between the three code implementations. The analysis of the ground motion revealed very strong (>3 m/s) shaking near the fault and damaging ground motions (>0.5 m/s) over large areas of Los Angeles, San Bernardino, and Riverside (Jones et al. 2008). Strong long-period ground motions (>1 m/s) were also anticipated for the Los Angeles Basin, where significant damage and potential collapse was interred for older high-rise buildings. These strong amplifications were attributed to surface waves channeled along a chain of sedimentary basins acting as a waveguide, which had been previously observed during a similar simulation for the TeraShake scenario earthquake (Olsen et al. 2006).

Also based on ShakeOut simulation results, Graves (2008) reported large variations in Los Angeles Basin ground-motion levels resulting from small ($\sim 15\%$) adjustments to the rupture velocity. Olsen et al. (2008) had demonstrated that replacing the kinematic rupture models with spontaneous rupture models reduces PGV extremes in the region by a factor of 2–3 and attributed these



Physics-Based Ground-Motion Simulation, Fig. 8 Region of interest and surface projection of the simulation domain used in the ShakeOut scenario with a superimposed still image of the ground velocity halfway the rupturing segment of the San Andreas Fault (left), and

comparison of horizontal magnitudes of peak ground velocities obtained from the three simulation sets corresponding to the two finite-difference and one finite-element codes (After Bielak et al. 2010)

reductions to the less coherent wave field excited by dynamic rupture models. Day et al. (2012) identified the segment between the Cajon Pass and the northern Coachella Valley as the main contributor to amplification in the Los Angeles Basin, with the highest excitation resulting from super-shear (or energetically forbidden sub-shear, super-Rayleigh) rupture speeds. These studies have suggested that the level of shaking in the Los Angeles region during a large ShakeOut-type earthquake will depend strongly on the details of the source and warrant more research into the physics of the rupture process and the wave propagation, yet they are excellent examples of the reach of physics-based earthquake simulation as a tool to gain insight about the seismic conditions of regions prone to large-magnitude earthquakes.

2008 M_w 5.4 Chino Hills Earthquake

In contrast to the simulation of scenario earthquakes such as the ShakeOut, the simulation of the wave propagation from past events offers the additional possibility of validating deterministic ground-motion prediction results against recorded data. Various examples of validation of past earthquakes are available for low frequencies ($f < 0.5$ Hz), where physics-based simulations have shown to perform best (e.g., Komatitsch et al. 2004). On the other hand, there are efforts to advance the simulation of earthquakes using fully deterministic simulations alone or in hybrid approaches to predict broadband ground motions at the higher frequencies of engineering interest (up to and above 10 Hz). A series of simulations oriented toward that goal are those done for the 2008 M_w 5.4 Chino Hills, California, earthquake.

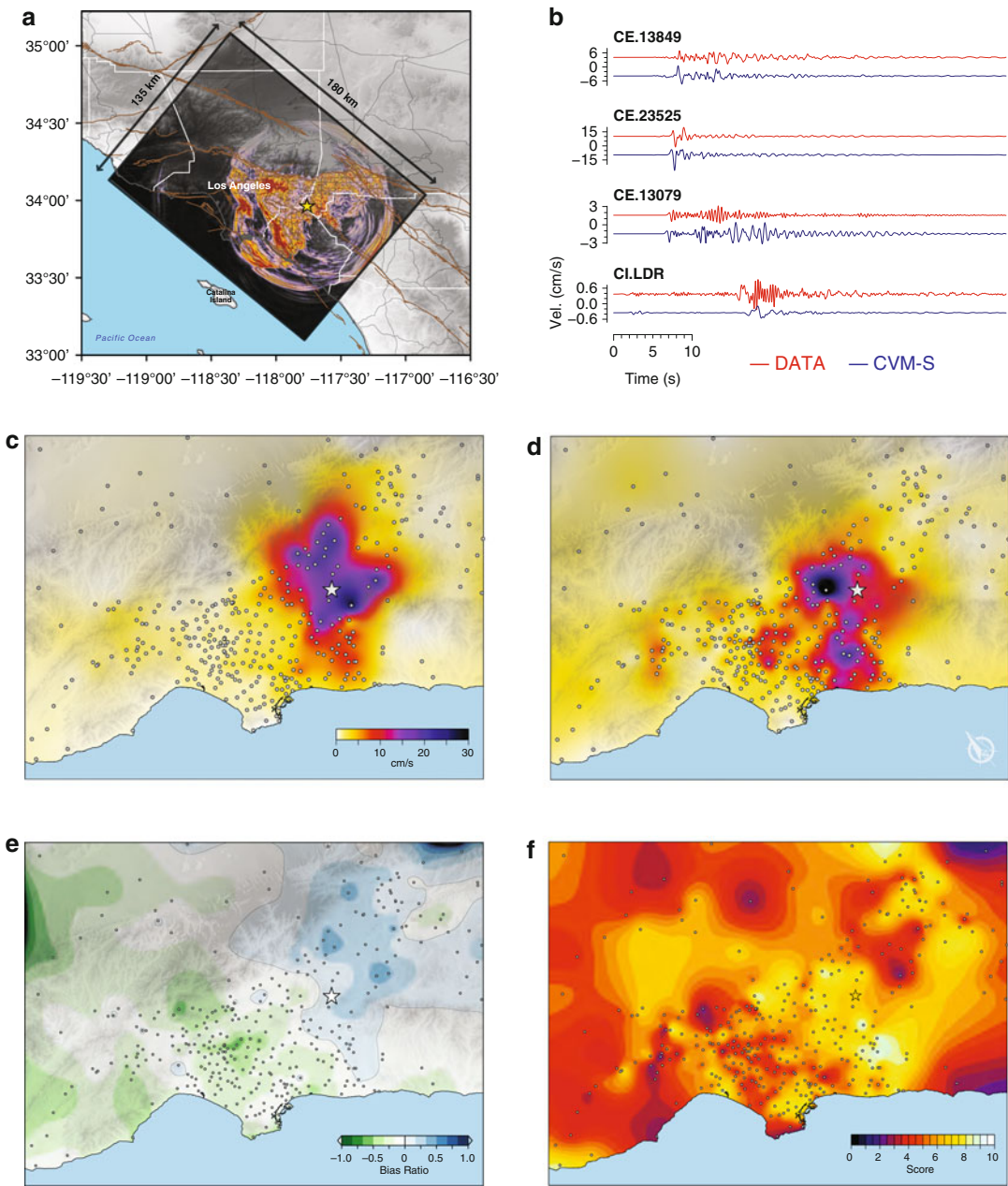
Olsen and Mayhew (2010) generated broadband (0–10 Hz) synthetics for the Chino Hills earthquake using a hybrid method based on the combination of low-frequency (<1.6 Hz) FD signals with high-frequency scattering operators. Olsen and Mayhew (2010) modeled the rupture as a point source with a combined strike-slip/thrust fault mechanism and used a minimum S -wave velocity of 500 m/s. Because visual

inspection was not deemed suitable for evaluation of the quality of fit between observed and synthetic data at shorter periods, Olsen and Mayhew (2010) used goodness-of-fit criteria to validate the broadband synthetics at 33 selected stations in the Los Angeles Basin. The goodness-of-fit results suggested that the threshold for acceptance was reached at two thirds of all the sites at short periods and at all the analyzed sites at moderate to long periods.

More recently, Taborda and Bielak (2013) used an entirely deterministic FE approach to model the wave propagation in the Greater Los Angeles region during the Chino Hills earthquake for frequencies up to 4 Hz and shear-wave velocities as low as 200 m/s. Their simulations used a finite source model derived from an independent inversion study and performed validations at over 300 stations between simulation synthetics and actual records. Taborda and Bielak (2013) also used a modified version of the goodness-of-fit criterion defined by Anderson (2004). Similar to Olsen and Mayhew (2010), they obtained a good agreement over the entire region at low frequencies (<0.5 Hz), but observed a decay in the goodness of fit with frequency, specially at the higher end (2–4 Hz). Their analysis indicated that this was mainly due to inaccuracies in the velocity model used (CVM-S), and in particular in the near-surface sedimentary layers, which have a stronger influence on the quality of the fit at the higher frequencies. Nonetheless, Taborda and Bielak (2013) found that the synthetics were overall realistic and concluded that future improvements to the material model will eventually render better fits. Figure 9 summarizes the results obtained by Taborda and Bielak (2013) and exemplifies the level of agreement (and lack thereof) that is possible when employing physics-based earthquake simulation to reproduce (well-constrained models of) past earthquakes.

CyberShake

The SCEC CyberShake project is an effort to develop a framework for probabilistic seismic hazard analysis that incorporates explicitly the



Physics-Based Ground-Motion Simulation, Fig. 9 Summary of results of simulation and validation of the 2008 M_w 5.4 Chino Hills earthquake. (a) Region of interest and simulation domain surface projection, (b) comparison of data and synthetics at selected locations, (c) spatial distribution of PGV interpolated from data at

recording stations, (d) spatial distribution of PGV interpolated from synthetics at the same stations, (e) bias ration between data and synthetic values of PGV, and (f) spatial distribution of the goodness-of-fit scores obtained for PGV (After Taborda and Bielak 2013)

source and wave propagation effects by means of physics-based simulations to estimate the ground motions of the expected earthquakes in a region (Graves et al. 2011). CyberShake has been initially tested for the case of the Greater Los Angeles area and will be extended to the Southern and Central California regions in the future. It consists of conducting hundreds of thousands of forward simulations for as many variations of fault ruptures as possible to consider for all active faults in the region of interest. However, in order to avoid the overload of conducting as many forward simulations as possible rupture variations for all the active faults and hypocenter locations, CyberShake precomputes strain Green tensors (SGT) between a collection of receivers (or regular grid of stations), and each of the subfault patches in each fault plane for all the finite fault models in the region. Then, CyberShake uses reciprocity to compute the synthetic seismograms for any given earthquake scenario by combining the corresponding contribution of each subfault's slip to the earthquake's finite fault model. The combined results of all the ground-motion synthetics obtained from the reciprocity computations for all the intended forward simulations offer a unique physics-based approach to the long-term seismic hazard of the region considered because the intensity measures produced by the simulations naturally incorporate 3D source, path, basin, and site effects and spatial variability of the region (as inferred from the employed velocity and source models), thus avoiding other assumptions for these effects implicitly embedded in empirical ground-motion relationships used in traditional probabilistic seismic hazard analysis. While significant further research will be required before CyberShake seismic hazard maps can be adopted in earthquake engineering analysis and design, its operational framework offers a perspective to the potential future use of physics-based earthquake ground-motion simulation.

Summary

This chapter presented the background and basic operational framework of physics-based

(deterministic) ground-motion simulation, its most commonly employed methods, and some notable recent application examples. The generation of ground-motion synthetics using a physics-based approach requires a (kinematic) finite source model and a material (seismic velocity and attenuation) model, which are used as input data for solving a forward wave propagation problem. Simulations are resolved using numerical techniques such as the finite-difference, finite-element, spectral element, or discontinuous Galerkin methods. Current uses of physics-based ground-motion simulation include the simulation of large earthquake scenarios for the assessment of the potential impact of seismic events and the evaluation of appropriate emergency management and response strategies and the construction of physics-based probabilistic seismic hazard maps (at long periods) based on local and regional fault models. Physics-based ground-motion simulation methods have been thoroughly tested through verification of scenario events and validation of small- and moderate-magnitude past earthquakes. Validation results, in particular, show that accurate seismograms (comparable to real records) can be obtained through physics-based ground-motion simulation, especially at low frequencies where both the source and the material models are better constrained. It is expected that future advances in all related areas (source models, material models, characterization of site effects, simulation methods) will permit the use of physics-based ground-motion simulations in broadband seismic hazard studies and engineering analysis and design, especially in regions where there is insufficient historical data and where simulations can complement the regional perspective of earthquake hazards.

Cross-References

- [Earthquake Mechanisms and Tectonics](#)
- [Earthquake Mechanism Description and Inversion](#)
- [Engineering Characterization of Earthquake Ground Motions](#)

- ▶ Integrated Earthquake Simulation
- ▶ Nonlinear Seismic Ground Response Analysis of Local Site Effects with Three-Dimensional High-Fidelity Model
- ▶ Probabilistic Seismic Hazard Models
- ▶ Seismic Actions Due to Near-Fault Ground Motion
- ▶ Site Response for Seismic Hazard Assessment
- ▶ Stochastic Ground Motion Simulation

References

- Aagaard BT, Hall JF, Heaton TH (2001) Characterization of near source ground motions with earthquake simulations. *Earthq Spectra* 17(2):177–207
- Aagaard BT, Brocher TM, Dolenc D, Dreger D, Graves RW, Harmsen S, Hartzell S, Larsen S, Zoback ML (2008) Ground-motion modeling of the 1906 San Francisco earthquake, part I: validation using the 1989 Loma Prieta earthquake. *Bull Seismol Soc Am* 98(2):989–1011
- Alterman Z, Karal FC (1968) Propagation of elastic waves in layered media by finite difference methods. *Bull Seismol Soc Am* 58(1):367–398
- Anderson JG (2004) Quantitative measure of the goodness-of-fit of synthetic seismograms. In: *Proceedings of the 13th world conference on earthquake engineering*. International Association for Earthquake Engineering Paper, 243. Vancouver
- Bao H, Bielak J, Ghattas O, Kallivokas LF, O'Hallaron DR, Shewchuk JR, Xu J (1998) Large-scale simulation of elastic wave propagation in heterogeneous media on parallel computers. *Comput Methods Appl Mech Eng* 152(1–2):85–102
- Bielak J, MacCamy RC, McGhee DS, Barry A (1991) Unified symmetric BEM-FEM for site effects on ground motion – SH waves. *J Eng Mech ASCE* 117(10):2265–2285
- Bielak J, Graves RW, Olsen KB, Taborda R, Ramírez-Guzmán L, Day SM, Ely GP, Roten D, Jordan TH, Maechling PJ, Urbanic J, Cui Y, Juve G (2010) The ShakeOut earthquake scenario: verification of three simulation sets. *Geophys J Int* 180(1):375–404
- Bielak J, Karaoglu H, Taborda R (2011) Memory-efficient displacement-based internal friction for wave propagation simulation. *Geophysics* 76(6):T131–T145
- Boore DM (1972) Finite difference methods for seismic wave propagation in heterogeneous materials. In: Bolt BA (ed) *Methods in computational physics*, vol 11. Academic Press, New York
- Bouchon M (1979) Discrete wave number representation of elastic wave fields in three-space dimensions. *J Geophys Res* 84(B7):3609–3614
- Brocher TM (2008) Compressional and shear-wave velocity versus depth relations for common rock types in northern California. *Bull Seismol Soc Am* 98(2):950–968
- Carcione JM, Kosloff D, Kosloff R (1988) Wave propagation simulation in a linear viscoelastic medium. *Geophys J Int* 95(3):597–611
- Cerjan C, Kosloff D, Reshef M (1985) A nonreflecting boundary condition for discrete acoustic and elastic wave equations. *Geophysics* 50:705–708
- Chaljub E, Moczo P, Tsuno S, Bard P-Y, Kristek J, Kaser M, Stupazzini M, Kristekova M (2010) Quantitative comparison of four numerical predictions of 3D ground motion in the Grenoble Valley, France. *Bull Seismol Soc Am* 100(4):1427–1455
- Cui Y, Olsen K, Jordan T, Lee K, Zhou J, Small P, Roten D, Ely G, Panda D, Chourasia A, Levesque J, Day S, Maechling P (2010) Scalable earthquake simulation on petascale supercomputers. In: *SC'10 Proceedings of the 2010 ACM/IEEE international conference for high performance computing, networking, storage and analysis*, New Orleans, LA, November 13–19, pp 1–20
- Day SM, Bradley CR (2001) Memory-efficient simulation of anelastic wave propagation. *Bull Seismol Soc Am* 91(3):520–531
- Day SM, Minster JB (1984) Numerical simulation of attenuated wavefields using a Padé approximant method. *Geophys J Int* 78(1):105–118
- Day SM, Roten D, Olsen KB (2012) Adjoint analysis of the source and path sensitivities of basin-guided waves. *Geophys J Int* 189(2):1103–1124
- Dumbser M, Käser M (2006) An arbitrary high-order discontinuous Galerkin method for elastic waves on unstructured meshes – II. The three-dimensional isotropic case. *Geophys J Int* 167(1):319–336
- Dupros F, de Martin F, Foerster E, Komatitsch D, Roman J (2010) High-performance finite-element simulations of seismic wave propagation in three-dimensional nonlinear inelastic geological media. *Parallel Comput* 36(5–6):308–325
- Ely GP, Jordan TH, Small P, Maechling PJ (2010) A Vs30-derived near-surface seismic velocity model. In: *Abstract AGU fall meeting*, no S51A-1907, San Francisco, 13–17 Dec 2010
- Emmerich H, Korn M (1987) Incorporation of attenuation into time-domain computations of seismic wave fields. *Geophysics* 52(9):1252–1264
- Frankel A, Vidale J (1992) A three-dimensional simulation of seismic waves in the Santa Clara Valley, California, from a Loma Prieta aftershock. *Bull Seismol Soc Am* 82(5):2045–2074
- Gottschämer E, Olsen KB (2001) Accuracy of the explicit planar free-surface boundary condition implemented in a fourth-order staggered-grid velocity-stress finite-difference scheme. *Bull Seismol Soc Am* 91(3):617–623
- Graves RW (1996) Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences. *Bull Seismol Soc Am* 86(4):1091–1106

- Graves RW (2008) The seismic response of the San Bernardino basin region during the 2001 Big Bear lake earthquake. *Bull Seismol Soc Am* 98(1):241–252
- Graves RW, Pitarka A (2010) Broadband ground-motion simulation using a hybrid approach. *Bull Seismol Soc Am* 100(5A):2095–2123
- Graves R, Jordan T, Callaghan S, Deelman E, Field E, Juve G, Kesselman C, Maechling P, Mehta G, Milner K, Okaya D, Small P, Vahi K (2011) CyberShake: a physics-based seismic hazard model for Southern California. *Pure Appl Geophys* 168(3–4):367–381
- Hartzell S, Harmsen S, Frankel A (2010) Effects of 3D random correlated velocity perturbations on predicted ground motions. *Bull Seismol Soc Am* 100(4):1415–1426
- Herrmann V, Käser M, Castro CE (2011) Non-conforming hybrid meshes for efficient 2-D wave propagation using the Discontinuous Galerkin Method. *Geophys J Int* 184(2):746–758
- Idriss IM, Seed HB (1968) Seismic response of horizontal soil layers. *J Soil Mech Found Div ASCE* 94(SM4):1003–1031
- Jones LM, Bernknopf R, Cox D, Goltz J, Hudnut K, Mileti D, Perry S, Ponti D, Porter K, Reichle M, Seligson H, Shofak K, Treiman J, Wein A (2008) The ShakeOut scenario, Technical report USGS-R1150, CGS-P25. U.S. Geological Survey, Reston, Virginia
- Koketsu K, Miyake H, Fujiwara H, Hashimoto T (2009) Progress towards a japan integrated velocity structure model and long-period ground motion hazard map. In: *Proceedings of the 14th world conference on earthquake engineering*, paper no S10–038, Beijing
- Komatitsch D, Vilotte J-P (1998) The spectral element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures. *Bull Seismol Soc Am* 88(2):368–392
- Komatitsch D, Tsuboi S, Ji C, Tromp J (2003) A 14.6 billion degrees of freedom, 5 teraflops, 2.5 terabyte earthquake simulation on the Earth Simulator. In: *SC'03 Proceedings of the ACM/IEEE conference for high performance computing and networking*. IEEE Computer Society, Phoenix, p 8
- Komatitsch D, Liu Q, Tromp J, Suss P, Stidham C, Shaw JH (2004) Simulations of ground motion in the Los Angeles basin based upon the spectral-element method. *Bull Seismol Soc Am* 94(1):187–206
- Komatitsch D, Erlebacher G, Göddeke D, Michéa D (2010) High-order finite-element seismic wave propagation modeling with MPI on a large GPU cluster. *J Comput Phys* 229(20):7692–7714
- Levander AR (1988) Fourth-order finite-difference P - SV seismograms. *Geophysics* 53(11):1425–1436
- Liu H-P, Anderson DL, Kanamori H (1976) Velocity dispersion due to anelasticity; implications for seismology and mantle composition. *Geophys J R Astron Soc* 47(1):41–58
- Lysmer J, Drake LA (1972) A finite element method for seismology, Chapter 6. In: Alder B, Fernbach S, Bolt B (eds) *Methods in computational physics*, vol 11. Academic, New York
- Ma S, Liu P (2006) Modeling of the perfectly matched layer absorbing boundaries and intrinsic attenuation in explicit finite-element methods. *Bull Seismol Soc Am* 96(5):1779–1794
- Marcinkovich C, Olsen K (2003) On the implementation of perfectly matched layers in a three-dimensional fourth-order velocity-stress finite difference scheme. *J Geophys Res* 108(B5):2276
- Moczo P, Kristek J (2005) On the rheological models used for time-domain methods of seismic wave attenuation. *Geophys Res Lett* 32(L01306):5
- Moczo P, Kristek J, Vavryuk V, Archuleta RJ, Halada L (2002) 3D heterogeneous staggered-grid finite-difference modeling of seismic motion with volume harmonic and arithmetic averaging of elastic moduli and densities. *Bull Seismol Soc Am* 92(8):3042–3066
- Moczo P, Kristek J, Halada L (2004) The finite difference method for seismologists – an introduction. Comenius University, Bratislava
- Mossessian TK, Dravinski M (1987) Application of a hybrid method for scattering of P , SV , and Rayleigh waves by near-surface irregularities. *Bull Seismol Soc Am* 77(5):1784–1803
- Ohminato T, Chouet BA (1997) A free-surface boundary condition for including 3D topography in the finite-difference method. *Bull Seismol Soc Am* 87(2):494–515
- Olsen KB (1994) Simulation of three-dimensional wave propagation in the Salt Lake basin. PhD thesis, University of Utah, Salt Lake City
- Olsen KB, Mayhew JE (2010) Goodness-of-fit criteria for broadband synthetic seismograms, with application to the 2008 M_w 5.4 Chino Hills, California, earthquake. *Seismol Res Lett* 81(5):715–723
- Olsen KB, Pechmann JC, Schuster GT (1995) Simulation of 3D elastic wave propagation in the Salt Lake basin. *Bull Seismol Soc Am* 85(6):1688–1710
- Olsen KB, Day SM, Bradley CR (2003) Estimation of Q for long-period (>2 sec) waves in the Los Angeles basins. *Bull Seismol Soc Am* 93(2):627–638
- Olsen KB, Day SM, Minster JB, Cui Y, Chourasia A, Faerman M, Moore R, Maechling P, Jordan T (2006) Strong shaking in Los Angeles expected from southern San Andreas earthquake. *Geophys Res Lett* 33(L07305):1–4
- Olsen KB, Day SM, Minster JB, Cui Y, Chourasia A, Okaya D, Maechling P, Jordan T (2008) TeraShake2: spontaneous rupture simulations of M_w 7.7 earthquakes on the southern San Andreas fault. *Bull Seismol Soc Am* 98(3):1162–1185
- Porter K, Hudnut K, Perry S, Reichle M, Scawthorn C, Wein A (2011) Foreword. *Earthq Spectra* 27(2):235–237
- Prevost J-H (1978) Plasticity theory for soil stress–strain behavior. *J Eng Mech Div ASCE* 104(5):1177–1194
- Restrepo D, Bielak J (2014) Virtual topography: a fictitious domain approach for analyzing free-surface

- irregularities in large-scale earthquake ground motion simulation. *Int J Numer Methods Eng* 100(7):504–533
- Rietmann M, Messmer P, Nissen-Meyer T, Peter D, Basini P, Komatitsch D, Schenk O, Tromp J, Boschi L, Giardini D (2012) Forward and adjoint simulations of seismic wave propagation on emerging large-scale gpu architectures. In: SC'12 Proceedings of the ACM/IEEE international conference on high performance computing, networking, storage and analysis, Salt Lake City, pp 38:1–38:11
- Robertsson JO (1996) A numerical free-surface condition for elastic/viscoelastic finite-difference modeling in the presence of topography. *Geophysics* 61(6):1921–1934
- Roten D, Olsen KB, Pechmann JC (2012) 3D simulations of *M* 7 earthquakes on the Wasatch fault, Utah, part II: broadband (0–10 Hz) ground motions and nonlinear soil behavior. *Bull Seismol Soc Am* 92(5):2008–2030
- Roten D, Olsen KB, Day SM, Cui Y, Fäh D (2014) Expected seismic shaking in Los Angeles reduced by San Andreas fault zone plasticity. *Geophys Res Lett* 41(8):2769–2777
- Sánchez-Sesma FJ, Luzón F (1995) Seismic response of three-dimensional alluvial valleys for incident *P*, *S*, and Rayleigh waves. *Bull Seismol Soc Am* 85(1):269–284
- Seriani G (1998) 3-D large-scale wave propagation modeling by spectral element method on Cray T3E multiprocessor. *Comput Methods Appl Mech Eng* 164(1–2):235–247
- Shi Z, Day SM (2013) Rupture dynamics and ground motion from 3-D rough-fault simulations. *J Geophys Res* 118(3):1122–1141
- Small P, Gill D, Maechling PJ, Taborda R, Callaghan S, Jordan TH, Olsen KB, Ely G (2015) The unified community velocity model software framework. *Comput Geosci* (Submitted)
- Smith WD (1975) The application of finite element analysis to body wave propagation problems. *Geophys J Int* 42(2):747–768
- Taborda R, Bielak J (2013) Ground-motion simulation and validation of the 2008 Chino Hills, California, earthquake. *Bull Seismol Soc Am* 103(1):131–156
- Taborda R, Ramírez-Guzmán L, López J, Urbanic J, Bielak J, O'Hallaron D (2007) Shake-Out and its effects in Los Angeles and Oxnard areas. *Eos Trans AGU* 88(52): Fall meeting supplement, abstract IN21B-0477
- Taborda R, Bielak J, Restrepo D (2012) Earthquake ground motion simulation including nonlinear soil effects under idealized conditions with application to two case studies. *Seismol Res Lett* 83(6):1047–1060
- Tu T, Yu H, Ramírez-Guzmán L, Bielak J, Ghattas O, Ma K-L, O'Hallaron DR (2006) From mesh generation to scientific visualization: an end-to-end approach to parallel supercomputing. In: SC'06 Proceedings of the ACM/IEEE international conference on high performance computing, networking, storage and analysis. IEEE Computer Society, Tampa, p 15
- Virieux J (1984) *SH*-wave propagation in heterogeneous media: velocity-stress finite-difference method. *Geophysics* 49(11):1933–1957
- Withers KB, Olsen KB, Shi S, Day SM, Takedatsu R (2013) Deterministic high-frequency ground motions from simulations of dynamic rupture along rough faults. In: Abstract SSA annual meeting, Salt Lake City, 17–19 Apr 2013
- Withers KB, Olsen KB, Shi Z, Day SM (2014) High-complexity deterministic *Q(f)* simulation of the 1994 Northridge *M_w* 6.7 earthquake. In: Proceedings of the SCEC annual meeting, no GMP-066, Palm Springs, 6–10 Sept 2014
- Wong HL, Trifunac MD (1974) Surface motion of a semi-elliptical alluvial valley for incident plane *SH* waves. *Bull Seismol Soc Am* 64(5):1389–1408
- Zahradnik J, Moczo P, Hron F (1993) Testing four elastic finite-difference schemes for behavior at discontinuities. *Bull Seismol Soc Am* 83(1):107–129
- Zhou J, Unat D, Choi DJ, Guest CC, Cui Y (2012) Hands-on performance tuning of 3D finite difference earthquake simulation on GPU fermi chipset. *Procedia Comput Sci* 9:976–985

Plastic Hinge and Plastic Zone Seismic Analysis of Frames

Vissarion Papadopoulos¹ and
Michalis Fragiadakis²

¹Institute of Structural Analysis and Seismic Research, School of Civil Engineering, National Technical University of Athens (N.T.U.A.), Athens, Greece

²School of Civil Engineering, Laboratory for Earthquake Engineering, National Technical University of Athens (N.T.U.A.), Athens, Greece

Synonyms

Beam–column elements; Concentrated plasticity; Distributed plasticity; Fiber elements; Lumped plasticity

Introduction

The most common approach to perform seismic design of structures is the traditional one based on a linear elastic analysis assumption. On the other

hand, nonlinear analysis methods due to their higher complexity and computational cost are mainly applied as a verification tool for the assessment of existing structures. The use of nonlinear analysis methods provides a robust and reliable framework for seismic design since they allow the adoption of more elaborate design criteria, based on less simplifying assumptions. For this reason the application of nonlinear analysis approaches for the seismic design of structures as discussed in detail in recent guidelines, e.g., ASCE (2007), is gaining ground among practicing engineers. This development is also attributed to the rapid increase of computing power which makes the implementation of such advance approaches more attractive for real-world applications.

Regarding finite element (FE) modeling assumptions, building structures are primarily modeled with one-dimensional beam–column finite elements (e.g., beams or rods). These models are used to represent the linear skeleton of such structures, while more detailed two- and/or three-dimensional FE models such as the ones proposed by Spiliopoulos and Lykidis 2006 are only rarely utilized. The FE structural model should also be able to adequately capture the response of other structural components that may considerably affect the overall capacity, e.g., infill walls, shear walls, and other nonstructural components. A discussion on these modeling issues can be found in the NIST 2010 document. In addition, soil–foundation–structure interaction may also play a significant role on the overall structural response. A detailed discussion on the effect of soil models of different complexity on the structural response can be found in Assimaki et al. 2012.

Considering the structural nonlinear response, there are two major sources of nonlinearity: material and geometric nonlinearity. Material nonlinearity is considered the primary source of damage for low- and medium-rise building structures, while geometrical nonlinearities should be taken into account in high-rise buildings with small aspect ratios that suffer from large horizontal deflections that introduce $P-\Delta$ effects. For the nonlinear material response, the FE simulation

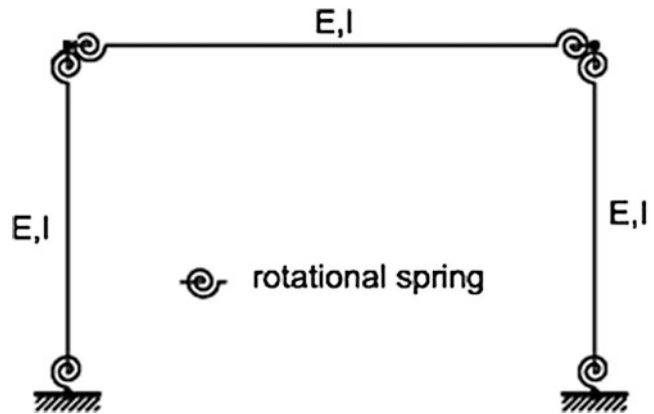
with beam–column members falls into two categories: concentrated (or lumped) and distributed plasticity approach. In concentrated plasticity, also called briefly as plastic hinge approach, the plastic deformations are “lumped” at the ends of a linear elastic element and are based on the moment–rotation relationships of the end sections for a given axial force. In distributed plasticity, beam–column elements allow for the formation of plastic zones along the member, while plastic hinges may form anywhere along the member. In the latest case, the constitutive inelastic behavior is monitored at section integration points of the beam–column elements, therefore accounting for the axial–moment interaction in a straightforward manner. Several commercial software packages are available, and usually each follows a different formulation that the user/engineer must be aware in order to obtain reliable predictions of the nonlinear structural response (Fragiadakis and Papadrakakis 2008).

Material Nonlinear Approaches

Concentrated (Lumped) Plasticity Approach

Concentrated plasticity is the approach commonly suggested in most design codes and guidelines (FEMA 2009). The first concentrated plasticity beam elements were the parallel model of Clough and Johnston 1966 and the series model of Giberson (1967). Preceding developments were based on the series model aiming at including the interaction of axial force and bending moment in the form of a nonlinear rotational spring at the element ends (Fig. 1) (Powell and Chen 1986). Further developments on lumped plasticity elements were focused on the implementation of cyclic laws. Models that introduce cyclic stiffness degradation and pinching have been proposed. Such models either modify the path of the reloading branch (Clough and Johnston 1966) or introduce “pinching” (Ibarra et al. 2005). Apart from linear or piecewise linear models, smooth hysteretic models have also been developed (Wen 1976) in order to provide a continuous change of stiffness for the nonlinear springs. Huang and Foutch (2009)

Plastic Hinge and Plastic Zone Seismic Analysis of Frames, Fig. 1 Lumped plasticity model of a portal frame: three beam–column FE and four rotational springs



performed a comparative study investigating the effect of different cyclic laws. An efficient beam–column element is also the force-based lumped plasticity element proposed in Scott and Fenves (2006). This approach combines the benefits of concentrated and distributed plasticity elements since it uses a minimum number of integration sections maintaining all the advantages of force-based distributed plasticity elements which are discussed in the next section.

In lumped plasticity approaches, the constitutive laws are expressed in terms of stress resultants (i.e., moment–rotations). This allows for the adoption of more complex phenomenological relationships with respect to distributed plasticity elements in which direct constitutive stress–strain relationships are used instead. In addition, lumped plasticity beam–column elements are more robust and reduce the computational cost and the memory requirements. This allows simulating complicated responses provided that the springs are appropriately calibrated. However, the concentrated plasticity approach relies on simplifying assumptions, restricting the inelastic deformations to prespecified regions of small length. Furthermore, there is no axial–moment interaction, and thus the springs are calibrated assuming constant axial force during dynamic analysis. Keeping constant the properties of the springs during analysis is another restriction of this modeling approach. Anagnostopoulos (1981) presented a parametric investigation showing that under monotonic loading the response is very sensitive

to the parameters of the springs, while Dides and de la Llera (2005) showed that for typical multistory structures, the response is influenced by the ratio of the beam-to-column stiffness. It has to be mentioned that such models usually lead to better response estimates for steel (Papadrakakis and Papadopoulos 1995) rather than for concrete structures. Prior to adopting lumped plasticity modeling, one must also be aware of the inability of lumped plasticity to capture the response under softening behavior (e.g., reinforced concrete structures).

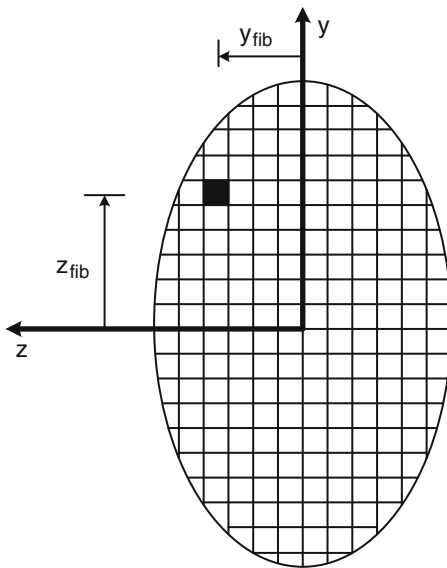
End Section Moment–Rotation

As mentioned above, lumped plasticity elements receive as input the moment–rotation relationships of the cross sections of interest, assuming that it is already known. The section moment–rotation ($M-\theta$) relationship can be defined assuming a backbone whose parameters are determined combining first principles and default values or performing section (fiber) analysis with appropriate software. This backbone is discussed in the FEMA P695 (2009) guidelines and is shown in Fig. 2. As shown in this figure, the curve essentially consists of three branches, an initial elastic (K_e), a hardening branch (K_h) and a descending branch with slope K_c that is used to introduce degradation. Degradation can be either “in-cycle,” i.e., defined with a negative stiffness branch and, or “cyclic.” Cyclic degradation usually refers to strength degradation (but not exclusively) at given displacement cycles due to capacity reduction during cyclic loading.

The sensitivity of a building's response to the parameters that define the backbone is demonstrated in Vamvatsikos and Fragiadakis (2010), while a comprehensive discussion on cyclic and in-cycle degradation is also offered in NIST (2010) document.

Distributed Plasticity Approach

Distributed plasticity beam-column elements offer a more accurate description of the inelastic behavior since they allow inelastic deformations to be developed anywhere within the member. These elements are also known as “fiber”

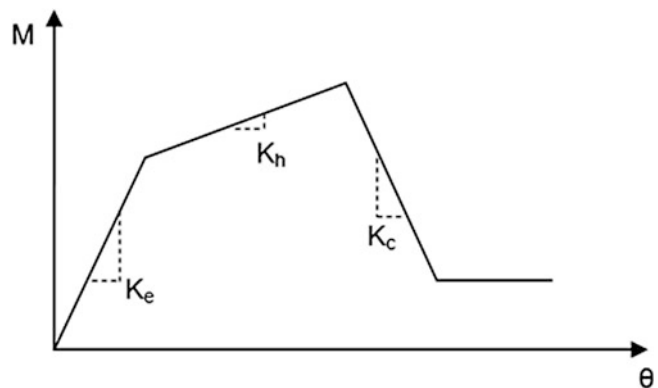


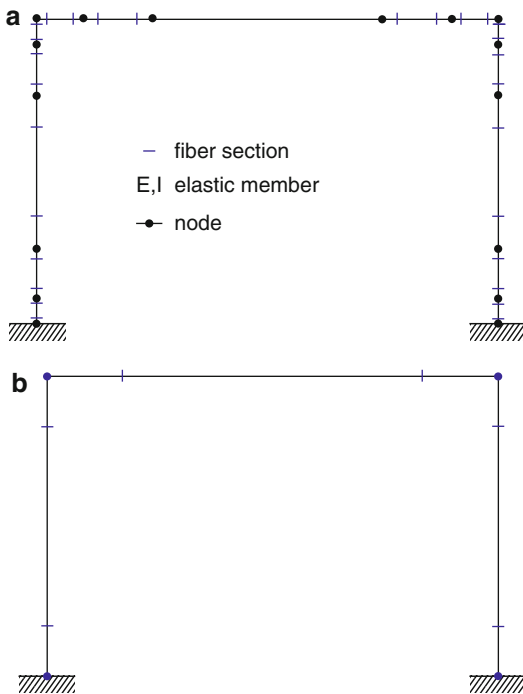
Plastic Hinge and Plastic Zone Seismic Analysis of Frames, Fig. 2 Discretization of an arbitrary cross section to fibers

elements, since the sections are divided to horizontal and vertical layers forming small areas, known as “fibers,” where the strain and stiffness parameters are evaluated (Fig. 3). Fiber elements are based on the classical FE theory with cubic Hermitian shape functions (Hellesland and Scordelis 1981). These elements are known in the literature as “displacement-based” or “stiffness-based” elements, and their shortcoming is that they require a fine mesh of beam-column elements in areas near the sections in which inelastic deformations are expected to be high (Fig. 4a). This is because the cubic shape function assumption leads to a linear distribution of the curvature along the element, which is inaccurate when the element end sections have yielded.

To overcome the problem stemming from the unknown distribution of curvature along the element, flexibility-dependent shape functions can be used instead (Fig. 3b). In this case, it is the distribution of forces (moments) that matters, while the element stiffness matrix can be easily calculated by summing the section flexibility matrices and inverting the resulting element flexibility matrix. This concept was extended by Kaba and Mahin (1984) and later was improved by Zeris and Mahin (1988) and Ciampi and Carlesimo (1986). These elements assume that the element is divided to equally spaced sections that allow the element to produce accurate predictions in the case of softening behavior, or in other words when the post-yield deformation path enters a segment with negative slope (softening). This formulation is also known as

Plastic Hinge and Plastic Zone Seismic Analysis of Frames, Fig. 3 Typical moment-rotation curve according to FEMA P695 (2009)





Plastic Hinge and Plastic Zone Seismic Analysis of Frames, Fig. 4 Modeling a portal frame using (a) displacement-based and (b) force-based elements

“force-based” or “flexibility-based.” Following the previous concept, Spacone et al. (1996) introduced a general mixed-type, fiber-based, beam–column element. At the integration sections, this element always maintains equilibrium of both forces and deformations and converges to a state that satisfies the constitutive laws within a specified tolerance.

Compared to lumped plasticity, the main shortcoming of distributed plasticity elements is that they require more computing resources. Numerical instabilities may be encountered if criteria that introduce abrupt loss of capacity are adopted, e.g., when trying to predict collapse. On the other hand, distributed plasticity elements do not require any special calibration and can be easily adopted for sections that consist of different materials since they use stress–strain relationships suitable for each fiber material. A discussion on distributed plasticity elements can be found in Neuenhofer and Filippou (1997), while comparisons between the force

and the displacement-based elements are presented in Neuenhofer and Filippou (1997) and Papaioannou et al. (2005). All studies converge that this formulation is accurate and robust and therefore suitable for a wide range of applications. A concise review on concentrated and distributed plasticity approaches can be found in Fragiadakis et al. (2015).

Section/Member Level

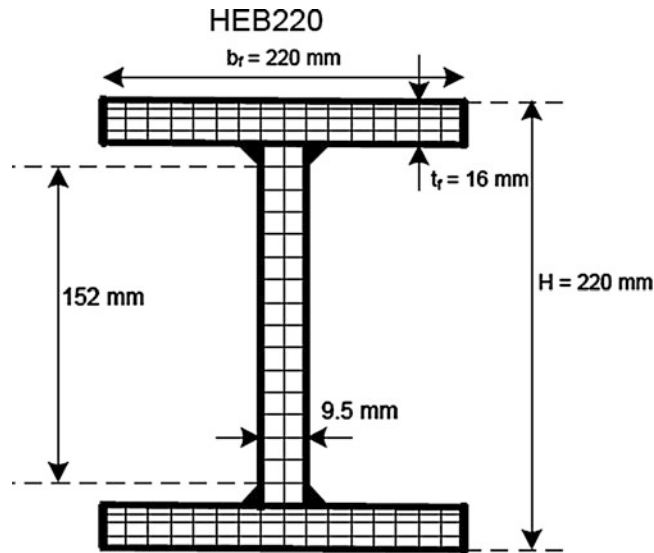
Fiber elements differ from lumped plasticity elements in the sense that they perform the integration over the fibers internally and therefore receive as input the parameters that define the stress–strain relationships of their fibers. On the other hand, lumped plasticity elements receive as input the moment–rotation relationships of the cross sections of interest, assuming that it is already known. In both cases, estimating the inelastic response requires integrating the stresses calculated at appropriately selected cross sections. Therefore, it is essential to be able to sufficiently capture the moment–rotation relationship of a cross section of a given geometry and reinforcement. Numerically, moment–rotation relationships are obtained using the fiber approach, i.e., discretizing the cross section to fibers that each follows its own material law.

Example

The simple example of the 1-story, 1-bay moment frame of Fig. 1 is modeled with wide-flange steel sections. The frame’s bay width and height are set to 4 m. The beam and columns of the steel frame are standard HEB 220 sections. For the distributed plasticity model, each section is divided into fibers where elastic perfectly plastic material with a yield stress $\sigma_y = 355$ MPa is assigned independently to each fiber. As shown in Fig. 5, the flanges and the web of the HEB 220 cross section are discretized into 64 fibers. For the concentrated plasticity model, the frame is modeled using a beam with hinge elements represented by rotational springs at the element end. The backbone curve describing the nonlinear behavior of the springs is calculated by integrating strains across the section (Fig. 5) and is presented in Fig. 6.

Plastic Hinge and Plastic Zone Seismic Analysis of Frames,

Fig. 5 Discretization of HEB 220 section to fibers



Plastic Hinge and Plastic Zone Seismic Analysis of Frames, Fig. 6 Backbone

curve of a HEB 220 steel section

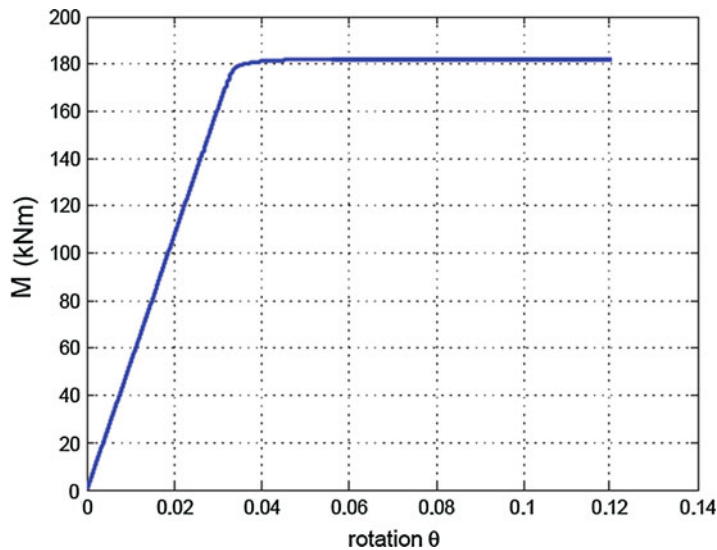


Figure 7a and b presents the pushover curves for both concentrated and distributed plasticity models and the corresponding incremental dynamic analyses (IDA), respectively. For the pushover analysis, the structure is pushed to a maximum 10 % roof drift. For the IDA case, the structural model is subjected to Canoga Park record from the 1994 Northridge earthquake scaled to multiple levels of intensity. The IDA curves of Fig. 7b plot an engineering demand parameter (EDP), in this case

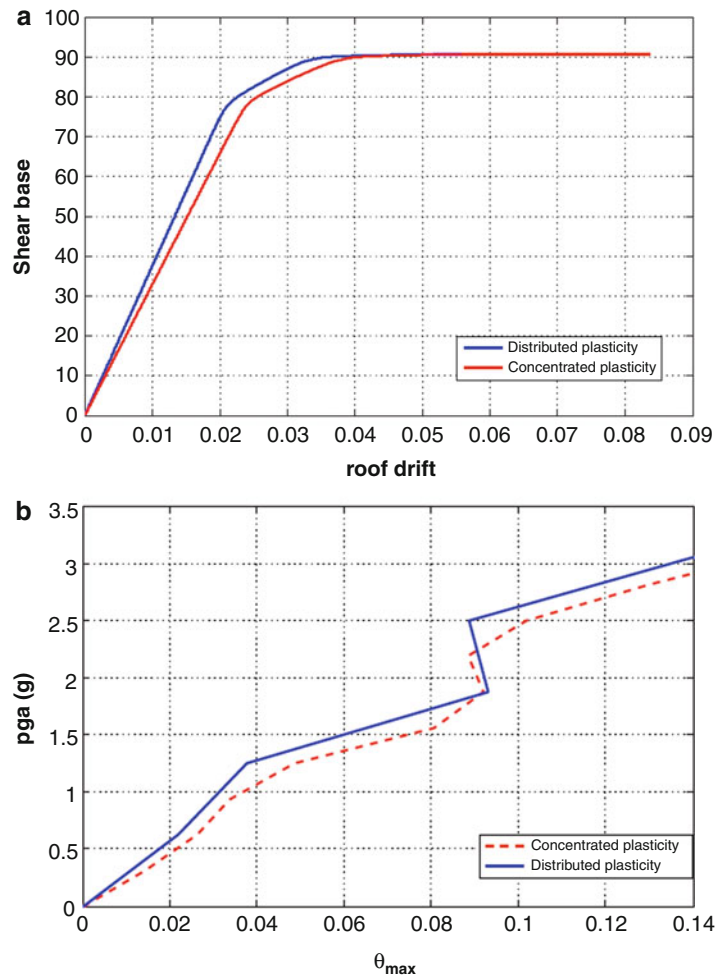
the maximum interstory drift $\theta_{\max}$, as a function of the earthquake intensity measure (IM) which is selected to be the maximum ground motion acceleration (PGA). From these figures the small difference on the structural response obtained by the two different models may be observed.

Summary

A critical review of the current state of the art of the computing practices adopted by the

Plastic Hinge and Plastic Zone Seismic Analysis of Frames,

Fig. 7 Comparison of (a) pushover and (b) IDA curves obtained with lumped and distributed plasticity models



earthquake engineering community is presented. The presentation extends from the finite element modeling of earthquake-resistant structures and the analysis procedures currently used to account for material nonlinearity. The two most commonly used approaches, namely, the plastic hinge and plastic zone approach, for estimating the structural capacity of seismically excited structures are discussed, and their relative advantages and disadvantages are presented together with the various aspects, developments, and implementations of these approaches over the past years. A comparison of the two approaches is presented in a simple example for both static pushover and incremental dynamic analysis approaches.

Cross-References

- [Assessment of Existing Structures Using Inelastic Static Analysis](#)
- [Incremental Dynamic Analysis](#)
- [Nonlinear Finite Element Analysis](#)

References

- Anagnostopoulos S (1981) Inelastic beams for seismic analysis of structures. *J Struct Eng* 107(7):1297–1311
- ASCE (2007) Seismic rehabilitation of existing buildings, ASCE/SEI Standard 41–06. American Society of Civil Engineers, Reston
- Assimaki D, Li W, Fragiadakis M (2012) Site effects in structural performance estimation. *Earthq Spectra* 28(3):859–88

- Ciampi V, Carlesimo L (1986) A nonlinear beam element for seismic analysis of structures. In: 8th European conference in earthquake engineering, Lisbon
- Clough RW, Johnston SB (1966) Effect of stiffness degradation on earthquake ductility requirements. In: Proceedings of Japan earthquake engineering symposium, p 232
- Dides MA, de la Llera JC (2005) A comparative study of concentrated plasticity models in dynamic analysis of building structures. *Earthq Eng Struct Dyn* 34(8):1005–1026
- FEMA P-695 (2009) Quantification of seismic performance factors, FEMA P-695 Report, prepared by the Applied Technology Council for the Federal Emergency Management Agency, Washington, DC
- Fragiadakis M, Papadrakakis M (2008) Modeling, analysis and reliability of seismically excited structures: computational issues. *Int J Comput Meth* 5(4):483–511
- Fragiadakis M, Vamvatsikos D, Karlaftis MG, Lagaros ND, Papadrakakis M (2015) Seismic assessment of structures and lifelines. *J Sound Vib* 334:29–56
- Giberson MF (1967) The response of nonlinear multi-storey structures subjected to earthquake excitation. EERL report. Earthquake Engineering Research Laboratory, California Institute of Technology, Pasadena
- Hellesland J, Scordelis A (1981) Analysis of RC bridge columns under imposed deformations. In: IABSE colloquium, Delft, Netherlands, pp 545–559
- Huang Z, Foutch D (2009) Effect of hysteresis type on drift limit for global collapse of moment frame structures under seismic loads. *J Earthq Eng* 13(7):939–964
- Ibarra LF, Medina RA, Krawinkler H (2005) Hysteretic models that incorporate strength and stiffness deterioration. *Earthq Eng Struct Dyn* 34(12):1489–1511
- Kaba SA, Mahin SA (1984) Refined modelling of reinforced concrete columns for seismic analysis. Report no. UCB/EERC-84/03, College of Engineering, University of California
- Neuenhofer A, Filippou FC (1997) Evaluation of nonlinear frame finite-element models. *ASCE J Struct Eng* 123:958–966
- NIST (2010) Applicability of nonlinear multiple-degree-of-freedom modeling for design. Report NIST GCR 10-917-9, prepared for the National Institute of Standards and Technology by the NEHRP Consultants Joint Venture, Gaithersburg
- Papadrakakis M, Papadopoulos V (1995) A computationally efficient method for the limit analysis of space frames. *J Comput Mech* 16(2):132–141
- Papaioannou I, Fragiadakis M, Papadrakakis M (2005) Inelastic analysis of framed structures using the fiber approach. In: Proceedings of the 5th international congress on computational mechanics (GRACM 05), Limassol, Cyprus
- Powell GH, Chen PF (1986) 3D beam-column element with generalized plastic hinges. *J Eng Mech* 112(7):627–641
- Scott MH, Fenves GL (2006) Plastic hinge integration methods for force-based beam-column elements. *J Struct Eng* 132(2):244–252
- Spacone E, Filippou FC, Taucer FF (1996) Fibre beam-column model for non linear analysis of R/C frames: part I formulation. *Earthq Eng Struct Dyn* 25(7):711–725
- Spiliopoulos KV, Lykidis GC (2006) An efficient three-dimensional solid finite element dynamic analysis of reinforced concrete structures. *Earthq Eng Struct Dyn* 35(2):137–157
- Vamvatsikos D, Fragiadakis M (2010) Incremental dynamic analysis for estimating seismic performance uncertainty and sensitivity. *Earthq Eng Struct Dyn* 39(2):141–163
- Wen YK (1976) Method for random vibration of hysteretic systems. *ASCE J Eng Mech Div* 102(2):249–263
- Zeris C, Mahin SA (1988) Analysis of reinforced concrete beam-columns under uniaxial excitation. *J Struct Eng* 114(4):804–820

Post-Earthquake Diagnosis of Partially Instrumented Building Structures

Eric M. Hernandez

Department of Civil and Environmental Engineering, College of Engineering and Mathematical Sciences, University of Vermont, Burlington, VT, USA

Synonyms

Damage detection; Instrumented buildings; Post-earthquake diagnosis; Structural dynamics; Wave propagation

Introduction

Post-earthquake diagnosis of building structures is essential in order to determine if a building can be safely re-occupied after a potentially damaging seismic event. The most traditional methodology is the use of visual inspections (ATC-20 2005). These inspections are made by qualified professionals and can take from a few days to possibly months. The effectiveness of the inspection will depend on a large part on the experience

of the inspector and accessibility to the structural damage. This last aspect was clearly manifested during the 1994 Northridge earthquake where hidden structural damage in buildings was not detected by qualified inspectors. Recently, researchers have begun to look at methods that rely on computer vision to perform the visual assessment in an automated fashion (German 2013). In principle this approach also relies on the damage being visible.

In the context of post-earthquake damage assessment, the term damage should be interpreted as reduction in stiffness and/or strength of any member in the structural system of the building. The loss of stiffness or strength can occur due to a single occurrence of a maximum relative displacement (drift) or due to short-cycle fatigue induced by alternating displacements that are characteristic of earthquake-induced ground motion. Depending on the type of material used for the building structure, visual symptoms of damage may or may not exist. On the other hand, visual symptoms do not necessarily imply that significant damage has occurred. For example, in reinforced concrete buildings, cracks may appear after a strong ground motion; however, this does not imply a reduction in strength. Due to the shortcomings with visual inspections, vibration-based post-earthquake assessment has been proposed.

Vibration-based post-earthquake damage assessment uses acceleration from the building response measured through accelerometers to detect, localize, and quantify the damage experienced by the building during the seismic event. Ideally, a building would be instrumented with six accelerometers per floor in order to capture the six possible independent motions of the floor (assuming rigid floor diaphragm). However, in most cases instrumentation can only take place at a limited number of locations, typically significantly less than the ideal number. The scientific challenge in vibration-based post-earthquake damage assessment is to develop algorithms that can use all available information (such as structural drawings, previous history of measured seismic events in the building, construction

information, etc.) together with limited acceleration measurements during the potentially damaging ground motion in order to perform the correct diagnosis.

Various algorithms have been proposed to perform post-earthquake diagnosis (Moaveni et al. 2011; Todorovska and Trifunac 2010; Bernal and Hernandez 2006; Lynch et al. 2006; Naeim et al. 2005; Revadigar and Mau 1999; Mau and Aruna 1994). Based on the relationship between the window(s) of time used to perform the diagnosis and the time of occurrence of the seismic event, existing methods can be broadly classified into “before-and-after” or “on-line” methods. As the name suggests, in the “before-and-after” strategy, the objective is to make a diagnosis based on measurements of the structural response prior to the potentially damaging event and measurements after the event, typically ambient vibrations. In the “on-line” strategy, the objective is to use the data during the seismic event to make the diagnosis.

Depending on the feature that is selected to characterize damage and on the definition of damage itself, existing methods can be broadly classified into:

- Spectral
- State estimation
- Wave propagation
- Energy methods
- Time series
- Model updating

As the name suggests, spectral methods search for changes in the spectral parameters of the structure (mode shapes and frequencies). Since damage is typically related to the loss of effective stiffness and modal parameters are directly related to the stiffness characteristics of the structure, it is intuitive to use these as a damage indicator. In this approach damage detection involves processing of acceleration records to determine changes in the frequency content of the measured acceleration using sliding windows Fourier analysis or time–frequency methods (Naeim et al. 2005; Todorovska and Trifunac 2010;

Moaveni et al. 2011; Asgariéh et al. 2012). If the effective frequency decreases during the strong motion window of time, then one would tend to infer that the stiffness of the system has decreased and structural damage has occurred.

State estimation methods operate by using partial measurements as feedback in order to estimate the complete dynamic response of the system and possibly track changes in the parameters of the model. Typically, some form of Bayesian filtering is used to perform the estimation (Chatzis et al. 2015). Damage features can be related to estimated functions of the state such as maximum drifts and/or damage indices (Erazo 2015; Hernandez and Erazo 2013). Alternatively, if changes in model parameters are used for damage assessment, then these need to be included in the state vector and estimated jointly with the traditional system state (Ching et al. 2006; Chatzi and Smyth 2009).

Another feature that has been proposed to detect damage in buildings is tracking seismic wave propagation through the building (Safak 1999). Wave propagation parameters include wave travel time between the floors and wave reflection and transmission coefficients of the floors. These have shown to be more sensitive to local damage than natural frequencies and provide yet another set of criteria for damage detection. However, the calculation of wave propagation parameters requires a dense array of sensor and recorders with a high rate of sampling. With the present sensor technology, the associated cost of the required level of instrumentation density may not be viable in all cases.

In energy methods the damage detection feature is the dissipated energy ratio which is a function of the energy dissipated by damping and the total energy dissipated. As mentioned previously and in contrast with existing methods previously described, dissipated energy is not an intrinsic property of linear systems and can be computed independent of whether the system is operating under linear or nonlinear range. In addition, the proposed damage feature has a very clear physical meaning and is by definition intimately related to earthquake-induced damage

of engineered structures (Uang and Bertero 1990; Sucuoglu and Erberik 2004; Teran-Gilmore and Jirsa 2007).

Time series methods have also been proposed. This family of methods relies on changes in the time history response or in the identified black-box input–output model coefficients. Lynch et al. (2006) developed a wireless sensor network system and used residuals generated from AR-ARX models as a damage-sensitive feature. The algorithm is based on the work of Sohn and Farrar (2001). Bernal and Hernandez (2006) developed a signal-processing algorithm based on the cumulative integral of residuals from partial ARX models. The partial ARX models are identified from non-damaging events and subsequently used to estimate the expected linear response of the building during the event of interest. The rate of accumulation of the cumulative integral of the residual between the partial ARX prediction and the measured response is used as a damage detection and classification feature.

The time series methods have the advantage (which can also be viewed as a disadvantage in some cases) of working with almost no information from a structural model of the system and relying completely on identified black-box models from measured data. This reduces the sensitivity to structural modeling errors, but the analysis is based entirely on numerical features not easily relatable to physical quantities pertinent to the problem. This makes time series methods less appealing since it is difficult to correlate the results with structural quantities of interest (such as forces, moments, drifts, etc.) related to structural damage.

Finally, model updating methods operate by identifying changes in the parameters of models formulated based on structural mechanics principles. Typically, these parameters are related to stiffness properties of the structural members of the model (beams, columns, walls, etc.). Model updating methods can operate within the “before-and-after” or “on-line” strategy. In the “before-and-after” strategy, the model parameters identified from ambient vibration before the ground motion or simply selected from experience are compared with the model parameters identified

from ambient vibrations after the ground motions (Simoen et al. 2015). If a significant change (or probability of change) is detected, then structural damage is deemed to have taken place. In the “on-line” strategy, as the name suggests, the model parameters are identified using the measured response during the ground motion (Ching et al. 2006; Chatzi and Smyth 2009). In this case, a more detailed evaluation can be conducted since in addition to stiffness parameters, hysteretic damage parameters can also be identified (which is not possible in the “before-and-after” strategy). Hysteretic behavior is known to be highly correlated to low-cycle fatigue damage typically experienced by buildings during strong earthquakes.

After this brief description of the various methodologies, a more detailed description of some of the methods is presented in the following sections.

Spectral Methods

Spectral methods operate under the assumption that the dynamic response of a building before and after the strong ground motion is governed by the following linear equation of motion:

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{C}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) = r\ddot{u}_g(t) \quad (1)$$

where $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are the mass, damping, and stiffness matrices of size $n \times n$. The response vector is $\mathbf{x}(t)$, $u_g(t)$ is the ground motion, and r is the ground motion influence vector. The corresponding undamped eigenvalue problem is

$$\mathbf{M}\phi_i\omega_i = \mathbf{K}\phi_i \quad (2)$$

where ω_i is the circular frequency of the i th mode and ϕ_i is the corresponding undamped mode shape. Spectral methods operate by interrogating changes in the mode shapes and frequencies before and after the ground motion, although more recently, data during the ground motion has been also used to identify the “changing” natural frequencies of the system as a sign of damage. The general assumption is that damage

induces reduction in stiffness, and reduction in stiffness is reflected in a reduction in the frequencies of the structure.

Challenges with the spectral methods include the following: (i) the intrinsic global nature of the modal properties makes it challenging (although not impossible) to point to the location of the damage; (ii) there are factors other than damage that produce similar effects on the damage-sensitive features, which are not easy to isolate (the effects of soil–structure interaction on the measured frequencies of vibration and environmental influences such as temperature); and (iii) only low-frequency modes can be reliably identified from vibration data in buildings, and these frequencies have low sensitivity to local damage.

State Estimation Methods

In state estimation methods, as the name suggests, the objective is to estimate the complete dynamic response (state) based on a model of the system and noise-contaminated partial measurements. For application in post-earthquake assessment, state estimation methods have been applied in two different avenues: (1) by using acceleration measurements to estimate the complete dynamic response, including displacements, velocities, interstory drifts, interstory shears, and overturning moments (the estimated quantities can be compared with structural capacity values of structural elements) (Mau and Aruna 1994; Hernandez and Bernal 2008), and (2) enlarging the state to include model parameters and identifying the model parameters (Ching et al. 2006). By identifying changes in model parameters, one can infer damage.

The nonlinear response during a strong ground motion can be modeled as

$$\mathbf{M}\ddot{\mathbf{x}}(t) + g_R(x, \dot{x}, \theta) = r\ddot{u}_g(t) \quad (3)$$

where g_R is a restoring force function, typically hysteretic. The restoring force function is defined by a set of parameter which we denote as θ .

By defining the state vector as $z^T = \{x^T \dot{x}^T \theta\}$, the equation of motion can be rewritten in first-order (state–space) form as

$$\dot{z}(t) = f(z(t)) + \mathbf{B}\ddot{u}_g \quad (4)$$

Measurements of the building response can be represented as function of the state as

$$y(t) = h(z(t)) + v(t) \quad (5)$$

where $v(t)$ is the measurement noise, typically small but not negligible during strong motions. The objective in state estimation methods, as the name suggests, is to estimate the complete state vector from noise-contaminated measurements $y(t)$. State estimation methods constitute an application of Bayes theorem:

$$p(z(t)|Y) = \frac{p(Y|z(t))p(z(t))}{p(Y)} \quad (6)$$

where $Y(t) = y([0 \ t])$. If the model is linear, then the application of Bayes theorem results in the Kalman filter (Grewal and Andrews 2001). If the model is nonlinear, several extensions of the Kalman filter have been proposed. Among these methods, the most popular ones are the extended Kalman filter, unscented Kalman filter, ensemble Kalman filter, and the particle filter (Sarkka 2013).

Wave Propagation Methods

This method uses data from acceleration sensors to detect changes in structural stiffness based on the analysis of travel times of seismic waves propagating through the structure (Todorovska and Trifunac 2007, 2010). The physical basis for this approach is based on the one-dimensional shear-wave propagation velocity:

$$V_s = \sqrt{\frac{\mu}{\rho}} \quad (7)$$

where μ is the shear stiffness and ρ is the density. Changes in the shear stiffness (due to damage) will result in a reduction in wave velocity and

consequently an increase in travel time between sensors. Wave propagation methods are more sensitive to local damage than the modal methods and should be able to point out to the location of damage with relatively small number of sensors. Additionally, the local changes in travel time should not be sensitive to the effects of soil–structure interaction, which is a major obstacle for the modal methods.

The spatial resolution of the wave methods is limited by the number of sensors. A minimum of two sensors (at the base and at the roof) are required to determine if the structure has been damaged, and additional sensors at the intermediate floors would help point out to the part of the structure that has been damaged. For example, one additional sensor between these two would help identify if the damage has been in the part of the structure above or beyond that sensor.

Energy Methods

Energy methods are based on identifying dissipated energy. Dissipated energy has a very clear physical meaning and is by definition intimately related to earthquake-induced damage of engineered structures (Uang and Bertero 1990; Sucuoglu and Erberik 2004; Teran-Gilmore and Jirsa 2007; Jehel et al. 2008).

To illustrate the concept, consider the governing equation of motion for an SDOF oscillator with viscous damping subject to a base motion time history $x_g(t)$:

$$m\ddot{y} + c\dot{x} + kx = 0 \quad (8)$$

where m is the mass, c is the viscous damping coefficient, and k is the stiffness coefficient. The absolute acceleration of the mass is $\ddot{y} = \ddot{x} + \ddot{x}_g$, and the relative displacement and velocity are x and $\dot{x}$, respectively. To obtain the energy balance relationship, we multiply both sides of the equation by $\frac{dx}{dt} dt$ and integrate between arbitrary time t_1 and t_2 :

$$m \int_{t_1}^{t_2} \ddot{y} \dot{x} dt + c \int_{t_1}^{t_2} \dot{x}^2 dt + k \int_{t_1}^{t_2} x \dot{x} dt = 0 \quad (9)$$

The third term in the left-hand side is the elastic strain energy stored between t_1 and t_2 :

$$\begin{aligned} \int_{t_1}^{t_2} kx \dot{x} dt &= \int_{t_1}^{t_2} kx \frac{dx}{dt} dt = \int_{x(t_1)}^{x(t_2)} kx dx \\ &= U = \frac{1}{2} kx^2 \Big|_{x(t_1)}^{x(t_2)} \end{aligned} \quad (10)$$

Therefore, in the limit, if the system starts from rest at t_1 and we let t_2 go to infinity until the system returns to its static equilibrium point, the strain energy integral U vanishes, and the dissipated energy (E_D) converges to the energy dissipated by viscous damping (E_{damp}), namely,

$$E_D = E_{\text{damp}} = c \int_0^\infty \dot{x}^2 dt = -m \int_0^\infty \ddot{y} \dot{x} dt \quad (11)$$

For a nonlinear system with hysteretic restoring force relationship, the integral in Eq. 2 is modified to incorporate the dissipated energy through hysteresis loops, resulting in

$$m \int_{t_1}^{t_2} \ddot{y} \dot{x} dt + c \int_{t_1}^{t_2} \dot{x}^2 dt + \int_{t_1}^{t_2} f_R \dot{x} dt = 0 \quad (12)$$

where f_R is the restoring force function. If the system starts from rest at $t_1 = 0$ and we let t_2 go to infinity until the system returns to one of its equilibrium points, the total energy dissipated by the system is given by

$$\begin{aligned} E_D &= c \int_0^\infty \dot{x}^2 dt + \int_0^\infty f_R \dot{x} dt = E_{\text{damp}} + E_h \\ &= -m \int_0^\infty \ddot{y} \dot{x} dt \end{aligned} \quad (13)$$

As can be seen, the expression for total dissipated energy, for linear (Eq. 4) and nonlinear (Eq. 6) hysteretic systems, is exactly the same. This invariance property makes the use of energy as a damage detection feature appealing. The expression for total dissipated energy by hysteresis is found from Eq. 6 as

$$E_h = \int_0^\infty f_R(x, \dot{x}) \dot{x} dt = -m \int_0^\infty \ddot{y} \dot{x} dt - c \int_0^\infty \dot{x}^2 dt \quad (14)$$

Based on the previous analysis, Hernandez and May (2013) proposed the following damage index (Z), defined for an SDOF as

$$Z = \frac{E_h}{E_D} = 1 - \frac{E_{\text{damp}}}{E_D} = 1 + \left(\frac{c}{m} \right) \frac{\int_0^{t_r} \dot{x}^2 dt}{\int_0^{t_r} \ddot{y} \dot{x} dt} \quad (15)$$

Other energy-based approaches that have been proposed rely on the identification of viscous damping (Satake et al. 2003). In these approaches, the premise is that effective viscous damping increases with damage. This is due to cracking, yielding, and other damage-related phenomena which results in higher levels of dissipated energy. The main drawback of this approach is that damping is difficult to identify accurately and its variability is high.

Time Series

Time series methods operate by identifying ARMA models of the form

$$\sum_{k=0}^l A_k y(p-k) = \sum_{k=0}^n B_k u(p-k) \quad (16)$$

where $A_k \in R^{m \times m}$ and $B_k \in R^{m \times p}$ are matrices, $y(h)$ is the measurement vector at time $t = h \Delta t$, and similarly $u(h)$ is the ground motion vector at time $t = h \Delta t$. To detect damage, time series methods search for changes in identified matrices A_k and B_k and(or) by examining the residuals between measurements during the ground motion and predictions provided by an ARMA model identified prior to the occurrence of the ground motion of interest. The residuals are defined as

$$\varepsilon(t) = y_m(t) - y(t) \quad (17)$$

where y_m is the measured acceleration and y is the model prediction. Bernal and Hernandez (2006) proposed two scalars, η and κ , where η measures the total extent of nonlinearity and κ captures the degree to which the structure recuperates the initial stiffness after the strong portion of the earthquake excitation is over. These metrics are computed as follows:

$$\gamma(t) = \frac{\int_0^t \varepsilon^2 dt}{\int_0^{t_{\max}} y_m^2 dt} \quad (18)$$

where

$$\eta = \gamma_{t=t_{\max}} \Big|_{\max \text{ over all channels}} \quad (19)$$

and

$$\kappa = \frac{S_e}{S_r} \Big|_{\max \text{ over all channels}} \quad (20)$$

where $t_{\max}$ = total duration of the earthquake record. It is implicit that Eq. 18 can be computed for every output channel and that the superscript t has been eliminated from the variables to simplify the notation. The numerator in Eq. 20 is the average slope of the curve given by Eq. 18 computed after it reaches 95 % of its final value, and the denominator is the slope of a line joining the 10 % to the 90 % values of the function in Eq. 18 for the earthquake used to identify the ARMA model. Various limits where proposed by Bernal and Hernandez (2006) on the constants γ and κ in order to separate undamaged buildings from potentially damaged ones. One drawback of this approach resides in the fact that residuals between building response and identified model predictions can occur due to many factors not related to damage, such as changes in building boundary conditions.

Model Updating

Model updating methods operate by identifying changes in the parameters (θ) of models formulated based on structural mechanics principles.

Typically, these parameters are related to stiffness properties of the structural members of the model (beams, columns, walls, etc.). Model updating methods can operate within the “before-and-after” or “on-line” strategy. In the “before-and-after” strategy, the model parameters identified from ambient vibration before the ground motion or simply selected from experience are compared with the model parameters identified from ambient vibrations after the ground motions (Simoen et al. 2015). If a significant change (or probability of change) is detected, then structural damage is deemed to have taken place. If a Bayesian approach is used, then the joint probability of the parameters given measurements is given by

$$p(\theta|Y_B) \propto p(Y_B|z\theta)p(\theta) \quad (21)$$

$$p(\theta|Y_A) \propto p(Y_A|z\theta)p(\theta) \quad (22)$$

where Y_B and Y_A are sets of ambient response measurements before and after the ground motion, respectively. More details regarding identification and model updating using ambient vibration can be found in Simoen et al. (2015). Damage is typically assessed by contrasting the two probability densities $p(\theta|Y_B)$ and $p(\theta|Y_A)$ or more simply by looking for shifts in the maximum likelihood value of these distributions (if they are unimodal).

In the “on-line” model updating strategy, as the name suggests, the model parameters are identified using the measured response during the ground motion (Ching et al. 2006; Chatzi and Smyth 2009). In this case, a more detailed evaluation can be conducted since in addition to stiffness parameters, hysteretic damage parameters can also be identified (which is not possible in the “before-and-after” strategy). Hysteretic behavior is known to be highly correlated to low-cycle fatigue damage typically experienced by buildings during strong earthquakes. Most of the work found in the literature uses shear building models and Bouc–Wen models (Bouc 1971) to represent the interstory restoring force function. The main drawback of model updating methods is their lack of robustness to model

errors (especially model class errors); this has been clearly illustrated in Ching et al. (2006). It is recommended that when using model updating methods, careful selection of the model class be exercised. In addition if the free parameter space is too large, uniqueness problems are to be expected; thus, the user needs to have good a priori knowledge of which elements are likely to exhibit damage and which ones not, thus excluding unnecessary parameters from the updating procedure.

Summary

Post-earthquake diagnosis of building structures is essential in order to determine if a building can be safely re-occupied after a potentially damaging seismic event. The most traditional methodology is the use of visual inspections. Depending on the type of material used for the building structure, visual symptoms of damage may or may not exist. Due to the shortcomings with visual inspections, vibration-based post-earthquake assessment has been proposed, developed, and implemented during the last few decades.

Vibration-based post-earthquake damage assessment uses acceleration from the building response measured through accelerometers to detect, localize, and quantify the damage experienced by the building during the seismic event. Various algorithms have been proposed to perform post-earthquake diagnosis. Existing methods for post-earthquake damage assessment can be broadly classified into spectral, state estimation, wave propagation, time series, energy methods, and model updating. This entry presents a summary of each method, their pros and cons, and range of applicability.

Cross-References

- [Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations](#)
- [Damage to Ancient Buildings from Earthquakes](#)

- [Damage to Buildings: Modeling](#)
- [Estimation of Potential Seismic Damage in Urban Areas](#)
- [Operational Modal Analysis in Civil Engineering: An Overview](#)
- [System and Damage Identification of Civil Structures](#)

References

- Asgarieh E, Moaveni B, Stavridis A (2012) Nonlinear structural identification of a three-story infilled frame using instantaneous modal parameters. In: Proceedings of the 30th international conference in modal analysis (IMAC-XXX), Jacksonville
- ATC-20 (2005) Field manual: post-earthquake safety evaluation of buildings, 2nd edn. Applied Technology Council, Redwood City, CA
- Bernal D, Hernandez E (2006) A data-driven methodology for assessing impact of earthquakes on the health of building structural systems. *Struct Des Tall Special Build* 15:21–34
- Bouc R (1971) Modèle mathématique d'hystérésis. *Acustica* 21:16–25
- Chatzi E, Smyth A (2009) The unscented Kalman filter and particle filter methods for nonlinear structural system identification with non-collocated heterogeneous sensing. *Struct Control Health Monit* 16(1):99–123
- Chatzis MN, Chatzi EN, Smyth AW (2015) An experimental validation of time domain system identification methods with fusion of heterogeneous data. *Earthq Eng Struct Dyn* 44:523–547
- Ching J, Beck J, Porter K, Shaikhutdinov R (2006) Bayesian state estimation method for nonlinear systems and its application to recorded seismic response. *ASCE J Eng Mech* 132(4):396–410
- Erazo K (2015) Non-linear state estimation with application to structural health monitoring. PhD dissertation, University of Vermont
- German SA (2013) Automated damage assessment of reinforced concrete columns for post-earthquake evaluations. PhD dissertation, Georgia Institute of Technology
- Grewal M, Andrews A (2001) Kalman filtering, theory and practice using MATLAB, 2nd edn. Wiley, New York
- Hernandez E, Bernal D (2008) State estimation in structural systems with model uncertainties. *J Eng Mech* 134(3):252–257
- Hernandez E, Erazo K (2013) Non-linear model-data fusion for post-earthquake damage assessment of structures. In: International workshop in structural health monitoring. Stanford University, Stanford, CA

- Hernandez EM, May G (2013) The Dissipated energy ratio as tool for earthquake induced damage detection and classification of instrumented structures. *ASCE J Eng Mech* 139(11):1521–1529
- Jehel P, Ibrahimbegovic A, Leger P, Davenne L (2008) On the computation of seismic energy dissipation in reinforced concrete frame elements. In: 8th world congress on computational mechanics, Venice
- Lynch JP, Wang Y, Lu KC, Hou TC, Loh CH (2006) Post-seismic damage assessment of steel structures instrumented with self-interrogating wireless sensors. In: Proceedings of the 8th national conference on earthquake engineering, San Francisco
- Mau TS, Aruna V (1994) Story-drift, shear and OTM Estimation from building seismic records. *ASCE J Struct Eng* 120(11):3366–3385
- Moaveni B, He X, Conte J, Restrepo J, Panagiotou M (2011) System identification study of a 7-story full-scale building slice tested on the UCSD-NEES shake table. *ASCE J Struct Eng* 137(6):705–717
- Naeim F, Hagie S, Alimoradi A, Miranda E (2005) Automated post-earthquake assessment and safety evaluation of instrumented buildings. JAMA report 2005-10639
- Revadigar S, Mau ST (1999) Automated multi-criterion building damage assessment from seismic data. *J Struct Eng ASCE* 125(2):211–217
- Safak E (1999) Wave-propagation formulation of seismic response of multistory buildings. *J Struct Eng ASCE* 125(4):426–437
- Sarkka S (2013) Bayesian filtering and smoothing. Cambridge University Press, Cambridge
- Satake N, Suda K, Arakawa T, Sasaki A, Tamura Y (2003) Damping evaluation using full-scale data of buildings in Japan. *J Struct Eng* 129(4):470–477
- Simoen E, DeRoeck G, Lombaert G (2015) Dealing with uncertainty in model updating for damage assessment: a review. *Mech Syst Signal Process* 56–57:123–149
- Sohn H, Farrar C (2001) Damage diagnosis using time-series analysis of vibrating signals. *J Smart Mater Struct* 10(3):446–451
- Sucuoglu H, Erberik A (2004) Energy-based hysteresis and damage models for deteriorating systems. *J Earthq Eng Struct Dyn* 33:69–88
- Teran-Gilmore A, Jirsa J (2007) Energy demands for seismic design against low-cycle fatigue. *J Earthq Eng Struct Dyn* 36(1):383–404
- Todorovska MI, Trifunac MD (2007) Damage detection in the Imperial County Services Building I: the data and time–frequency analysis. *Soil Dyn Earthq Eng* 27(6):564–76
- Todorovska M, Trifunac M (2010) Earthquake damage detection in the Imperial County Services Building II: analysis of novelties via wavelets. *J Struct Control Health Monit* 17(8):895–917
- Uang C, Bertero V (1990) Evaluation of seismic energy in structures. *J Earthq Eng Struct Dyn* 19(1):77–90

Principles of Broadband Seismometry

Nick Ackerley

Nanometrics, Inc, Kanata, Ottawa, ON, Canada

Synonyms

Weak-motion sensor; Very broadband seismometer

Introduction

There are many different types of instruments which can be used to detect ground motion. Broadband seismometers belong to a class of sensors called inertial sensors. In contrast, methods of sensing ground motion such as strainmeters and Global Positioning Systems (GPS) are not considered inertial sensors.

In this entry the principles of operation of broadband seismometers are discussed and contrasted with those of passive seismometers. The criteria for selecting a seismometer are discussed. Particular attention is paid to noise-generating mechanisms, including self-noise generated within the sensor, environmental sensitivities, and installation-related noise.

Applications

Broadband seismometers are very versatile instruments which have many applications in the field of earthquake engineering, including:

- Event detection and location (see ► [Seismic Event Detection](#) and ► [Earthquake Location](#))
- Earthquake magnitude and moment-tensor estimation (see ► [Earthquake Magnitude Estimation](#), ► [Earthquake Mechanism Description and Inversion](#) and ► [Long-Period Moment-Tensor Inversion: The Global CMT Project](#))
- Volcanic eruption early warning (see ► [Volcanic Eruptions, Real-Time Forecasting of](#))

and ► [Noise-Based Seismic Imaging and Monitoring of Volcanoes](#))

- Subsurface imaging
- Site response (see, e.g., ► [Site Response: Comparison Between Theory and Observation](#) and ► [Site Response for Seismic Hazard Assessment](#))
- Restitution of ground displacement (see ► [Selection of Ground Motions for Response History Analysis](#) and ► [Seismic Actions due to Near-Fault Ground Motion](#))

Classes of Inertial Sensors

In the broadest sense, a seismometer is any instrument which responds to ground motion, and a seismograph is any instrument which subsequently makes a recording of that ground motion. In practical usage however, a seismometer refers to a specific subclass of inertial sensors. One way to understand where the broadband seismometer is situated relative to the other classes of instruments, which measure ground motion, is to take a tour of the family tree of inertial sensors.

Inertial sensors consist of a frame and a “proof mass” suspended within the frame. Movement of the frame is sensed by sensing differential motion between the frame and the proof mass. The suspension usually constrains the proof mass to have one or more degrees of freedom, which can be either rotational or translational. If these degrees of freedom are purely rotational, the instrument is called a rotational seismometer or a gyroscope (see Lee et al. 2012).

Translational inertial sensors have many subclasses. One way of understanding their classification is to look primarily at the types of motion each is designed to measure: static or dynamic, strong or weak, horizontal or vertical, and long period or short period (these terms are explained in the paragraphs which follow).

Inertial sensors which are configured to detect static accelerations include gravimeters and tiltmeters, which are arranged to detect vertical and horizontal accelerations, respectively. Neither instrument is required to measure the full acceleration due to gravity. A tiltmeter can be leveled to produce zero output when it is installed; the mainspring of a gravimeter is

similarly adjusted to cancel a standard acceleration due to gravity. Thus the sensitivity of a gravimeter or a tiltmeter can be quite high because the range of accelerations to be measured is quite small. Finally, since traveling seismic waves are generally not of interest for gravimeters and tiltmeters, these instruments typically have a low-pass characteristic with an upper corner frequency which is relatively low, typically 1 Hz or lower.

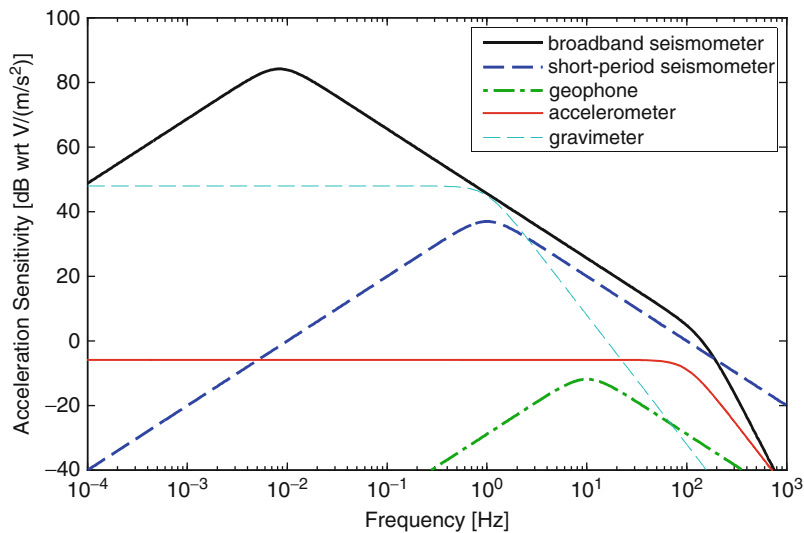
An ► [accelerometer](#) is another type of inertial sensor which is configured to detect accelerations down to zero frequency. Unlike a gravimeter or a tiltmeter, they are typically configured to read peak accelerations on the order of 1 g or larger without clipping and thus must have relatively low sensitivities. They can be used to detect static accelerations, for the purpose of integrating them to obtain position as in an inertial guidance system, or they can be used to detect strong seismic motions. Like a gravimeter they have a flat response to acceleration down to zero frequency, but the upper corner of their low-pass response will be at a much higher frequency.

A seismometer differs from gravimeters, tiltmeters, and accelerometers in that it is not required to respond to ground motion all the way down to zero frequency; this enables it to have a higher sensitivity and therefore measure much smaller motions than an accelerometer. They typically have transfer functions which are flat to velocity over at least a decade, but many hybrid responses, flat to acceleration over some significant band, are viable alternatives (Wielandt and Streckeisen 1982). “Strong-motion velocity meters” capable of detecting motions as large as those detectable on an accelerometer are also commercially available, but the term seismometer is generally reserved for an instrument which measures weak motion.

Traditionally the basic subgroupings of seismometers were long period or short period, depending on whether the lower corner period is above or below about 7 s period (see section “[Ground Motion Spectra](#)” below). Modern broadband seismometers span both frequency bands and have made long-period seismometers obsolete.

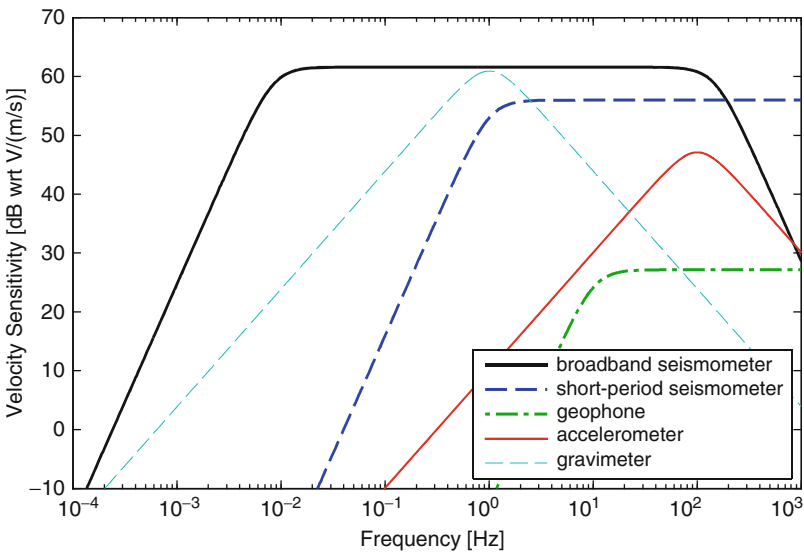
Principles of Broadband Seismometry,

Fig. 1 Response to acceleration of inertial sensors



Principles of Broadband Seismometry,

Fig. 2 Response to velocity of inertial sensors



Like a short-period seismometer, a geophone is usually a passive seismometer (see ► [Historical Seismometer](#) and ► [Passive Seismometers](#)), but one with a higher corner frequency and lower sensitivity.

The transfer functions of representative examples from the main classes of inertial sensors are plotted in Fig. 1 for ground motion in units of acceleration and in Fig. 2 for units of velocity. Seismometers have the highest sensitivity of all inertial seismometers, and broadband seismometers

have the highest sensitivity of all of these classes at long periods, but not at zero frequency.

The following sensors were taken as typical of the classes shown in Figs. 1 and 2:

- **Accelerometer:** a class A accelerometer (Working Group on Instrumentation, Siting, Installation, and Site Metadata 2008), e.g., Nanometrics Titan
- **Gravimeter:** based on the CG-5 (Scintrex Limited 2007)

- **Broadband seismometer:** based on the Trillium 120 (Nanometrics, Inc 2009)
- **Short-period seismometer:** based on the S-13 (Geotech Instruments, LLC 2001)
- **Geophone:** based on the SG-10 (Sercel - France 2012)

For more discussion of the importance of the sensitivity of a seismometer to its performance, see section “[Response](#).”

Force Feedback

The principles of operation, benefits, and drawbacks of an active inertial sensor are best understood in relation to the operation of a passive inertial sensor, since the latter is effectively a component in the former.

Modern short-period seismometers and geophones are passive inertial sensors which consist of a pendulum with a velocity transducer with sensitivity G in $\text{V}\cdot\text{s}/\text{m}$ mounted so as to measure the relative velocity $\dot{x}_b$ of a proof mass M and frame of the sensor, as shown in Fig. 3. The spring constant of the mainspring and suspension combined K and the viscous damping B determine the transfer function of the pendulum, required to relate the output voltage v_o to the input motion of the frame, $\dot{x}_i$.

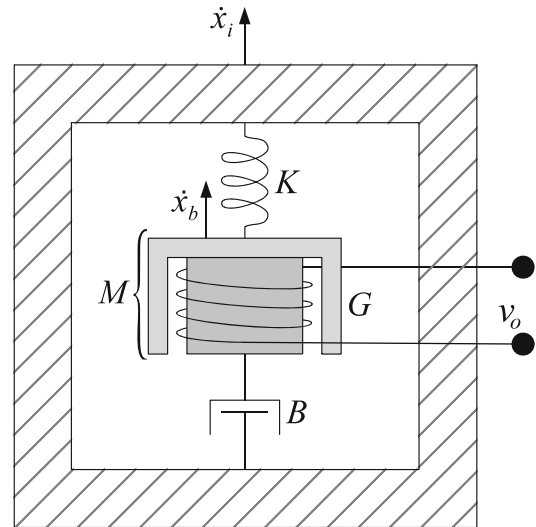
A block diagram of the whole passive seismometer system, from ground motion to the digitizer, is shown in Fig. 4.

The Laplace transform of the transfer function which converts frame velocity to an output voltage is (Aki and Richards 2002; Wielandt 2002)

$$\frac{V_o}{sX_i} = G \frac{s^2}{s^2 + \frac{B}{M}s + \frac{K}{M}} \left[\frac{\text{V}\cdot\text{s}}{\text{m}} \right]$$

Here the output voltage V_o and input displacement X_i are written in uppercase to emphasize that this transfer function operates in the frequency domain. It is a property of the Laplace transform that multiplication by the (complex) frequency s in the frequency domain is equivalent to differentiation in the time domain, so sX_i is the input *velocity*. Another property of the Laplace transform is that to evaluate the (complex) transfer function at a given frequency f in Hz or angular frequency $\omega = 2\pi f$ in rad/s, one makes the substitution $s = j\omega$.

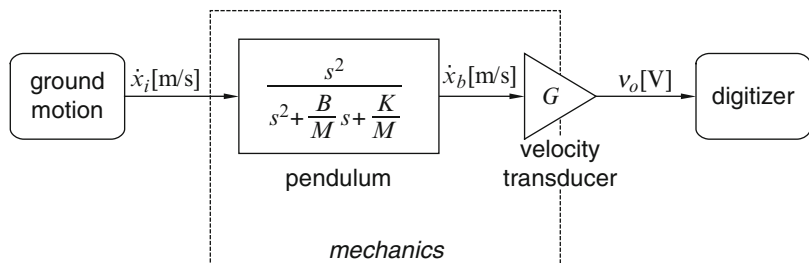
For high sensitivity in the passband, the generator constant of the velocity transducer must be large. Typically the velocity transducer is constructed like a voice coil in a speaker: many turns of copper passing through a narrow gap in a magnetic circuit energized by a strong magnet



Principles of Broadband Seismometry, Fig. 3 Mechanical schematic of a passive inertial sensor

Principles of Broadband Seismometry,

Fig. 4 System block diagram of a passive inertial sensor



with large diameter are required for a large generator constant. This is shown schematically in Fig. 3.

By comparison with the standard form of a second-order transfer function

$$F(s) = S \frac{s^2}{s^2 + 2\zeta_0 \omega_0 s + \omega_0^2}, \quad (1)$$

we can write expressions for the sensitivity S , the natural frequency f_o in Hz, the angular natural frequency $\omega_0 = 2\pi f_o$, and the (dimensionless) damping constant ζ_0 of the pendulum:

$$S = G$$

$$f_o = \frac{\omega_0}{2\pi} = \frac{1}{2\pi} \sqrt{\frac{K}{M}}$$

$$\zeta_0 = \frac{1}{2\omega_0} \frac{B}{M} = \frac{B}{2\sqrt{KM}}$$

Below the natural frequency, the sensitivity of passive seismometer drops off rapidly, as shown in Figs. 1 and 2. At the resonant frequency, the degree of “peaking” in the transfer function is controlled by the damping constant ζ_0 . Low damping corresponds to a strongly peaked response, something which is generally not desirable, because it means the output will be dominated by signal at that frequency, and because the system has a greater tendency to clip at that frequency.

For high sensitivity at low frequencies, the spring stiffness K [N/m] must be low and the mass M [kg] must be large. For maximum bandwidth while maintaining passband flatness, the viscous damping B [N/(m/s)] is generally set – often with the aid of an external damping resistor not modeled here – so that the dimensionless damping constant ζ_0 is approximately $1/\sqrt{2}$.

Passive seismometers have the obvious advantage of requiring no power to operate. Indeed the device is literally a generator, albeit one which produces very little power. However, modern digital seismology requires that some power be expended on digitization and subsequent

telemetry or storage, so the seismograph system as a whole always consumes some power.

The chief disadvantages of passive seismometers for recording ground motion at long periods are:

- (a) The suspension system and velocity transducer have nonlinearities as a function of the displacement of the proof mass from equilibrium. These nonlinearities can cause excursions in the output signals which cannot be removed by post-processing.
- (b) Large masses and compliant suspensions are required to achieve high sensitivity at long periods, making for an instrument, which is physically large, difficult to construct and susceptible to damage when being transported.
- (c) Conflicting damping requirements: As we shall see below, the damping must be as low as possible in order to minimize self-noise, but for a flat transfer function the damping must be set to a level much higher than the minimum. A transfer function which is strongly peaked will have a corresponding “notch” in its clip level, limiting its ability to record during large or nearby events.

A force-feedback or “active” seismometer solves these problems by taking a mechanical system like that of the passive seismometer, as shown in Fig. 3, and adding a displacement transducer T , as shown in Fig. 5. In an active seismometer, the voice coil G is used to produce a force on the proof mass by driving it with a feedback current i_f instead of using it to produce a voltage proportional to the velocity of the proof mass.

It is a remarkable fact that every generator is also a motor. The force transducer in the feedback path of an active seismometer has the same physical configuration as the velocity transducer at the output of a passive seismometer. Moreover the electromotive force generated per unit of velocity or “motor constant” of the former in V/(m/s) is identical to the force generated per unit of current or “generator constant” of the latter in N/A:

$$G = \frac{e}{\dot{x}} = \frac{f}{i} \left[\frac{\text{V}\cdot\text{s}}{\text{m}} = \frac{\text{N}}{\text{A}} \right]$$

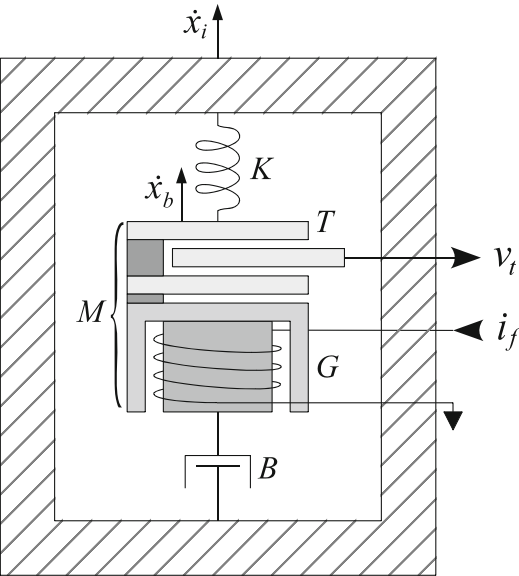
The displacement transducer shown in Fig. 5 is generally one of two types. A linear-variable differential transformer (LVDT) is effectively a transformer where the coupling to the secondary winding(s) is a function of the position of a movable magnetic core. A capacitive displacement transducer (CDT) is a capacitor where the

capacitance is varied by changing the spacing or overlapping area of the capacitor plates. In order to reduce nonlinearity and increase sensitivity, transducers of both types are normally implemented as part of a bridge circuit, requiring three windings in the case of an LVDT or three capacitor plates in the case of a CDT. In both cases the displacement transducers operate at a carrier frequency well above the passband of the instrument, and the outputs require demodulation back down to the passband before they are passed on to the compensator, output, and feedback stages.

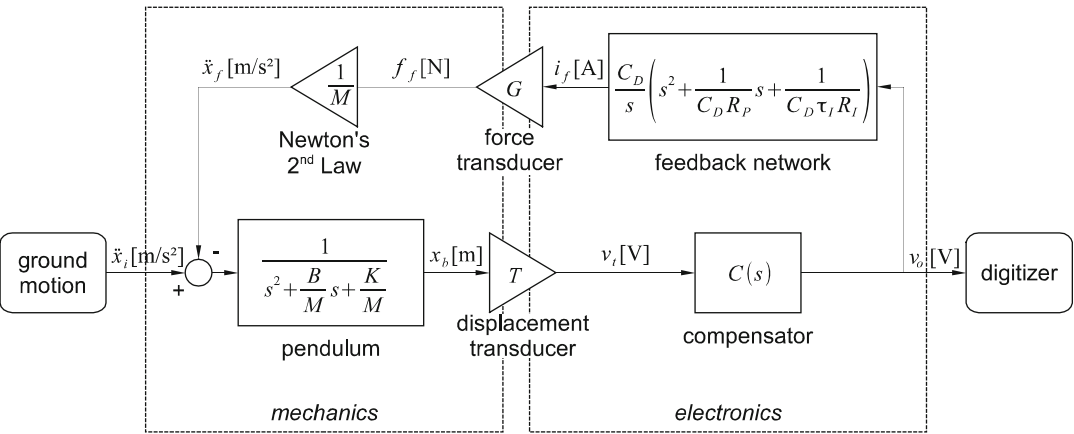
The feedback loop is closed using feedback electronics which take the displacement transducer output voltage and produce an appropriate feedback current as shown in the block diagram of Fig. 6. Two main blocks of interest in the feedback electronics are a compensator $C(s)$ in the forward path and the feedback network itself.

Discussions of the components, design, or limitations of the compensator are beyond the scope of this entry, but suffice it to say that the challenge is to simultaneously ensure high loop gain and stability and that the compensator design determines the shape of the roll-off of the closed-loop transfer function at high frequency.

A few of the key components in the feedback network are explicitly indicated in Fig. 6, namely, the “differential” feedback capacitor C_D , a “proportional” feedback resistor R_P , and an “integral” feedback resistor R_I . The integrator time constant



Principles of Broadband Seismometry, Fig. 5 Mechanical schematic of an active inertial sensor



Principles of Broadband Seismometry, Fig. 6 Block diagram of an active inertial sensor using force feedback

τ_I is another key parameter of the feedback circuit.

Feedback control systems have the property that as long as the loop gain is sufficiently high (and the system is stable), the transfer function of the system is the inverse of the feedback transfer function (Phillips and Harbor 1991). In the case of an inertial sensor, this means that the transfer function is determined by electronic components in the feedback network not the physical characteristics of the pendulum. Furthermore this same feedback holds the proof mass substantially at rest with respect to the frame of the sensor, greatly reducing the impact of nonlinearities in the suspension and transducers.

Thus, for high loop gain, a good approximation of the transfer function of the active seismometer of Fig. 6 is

$$\frac{V_o}{sX_i} = \frac{M}{GC_D} \frac{s^2}{s^2 + \frac{1}{C_D R_P} s + \frac{1}{C_D \tau_I R_I}} \left[\frac{\text{V}\cdot\text{s}}{\text{m}} \right]$$

By comparison with the standard form of a second-order transfer function (1), we can write expressions for the sensitivity, corner frequency, and damping of this force-feedback seismometer:

$$S = \frac{M}{GC_D}$$

$$f_o = \frac{\omega_0}{2\pi} = \frac{1}{2\pi} \sqrt{\frac{1}{C_D \tau_I R_I}}$$

$$\zeta_0 = \frac{1}{2C_D R_P \omega_0} = \frac{1}{2R_P} \sqrt{\frac{\tau_I R_I}{C_D}}$$

The properties which make an active force-feedback seismometer a good tool for broadband weak-motion seismology become apparent:

- (a) The effects of nonlinearities as a function of displacement of the proof mass from its rest position are greatly reduced, using a displacement transducer and high loop gain.
- (b) High sensitivity at long periods is achieved by selecting components in the feedback path. In order to achieve a particular natural

frequency, the electronic components required in a force-feedback seismometer are much smaller than the mechanical components required in a passive seismometer with the same response.

- (c) The damping of the pendulum can be made low without affecting the closed-loop response of the system, since the damping ζ_0 of the closed-loop system is determined entirely by the components in the feedback network.

The power required to operate a force-feedback seismometer must be considered its chief disadvantage relative to a passive seismometer. Although the power consumption can often be kept to a fraction of the rest of the recording system, the fact that power must be provided at all does increase the size and complexity of cables, connectors, and the digitizer itself. The versatility of the broadband seismometer relative to passive alternatives, however, can more than make up for this.

Ground Motion Spectra

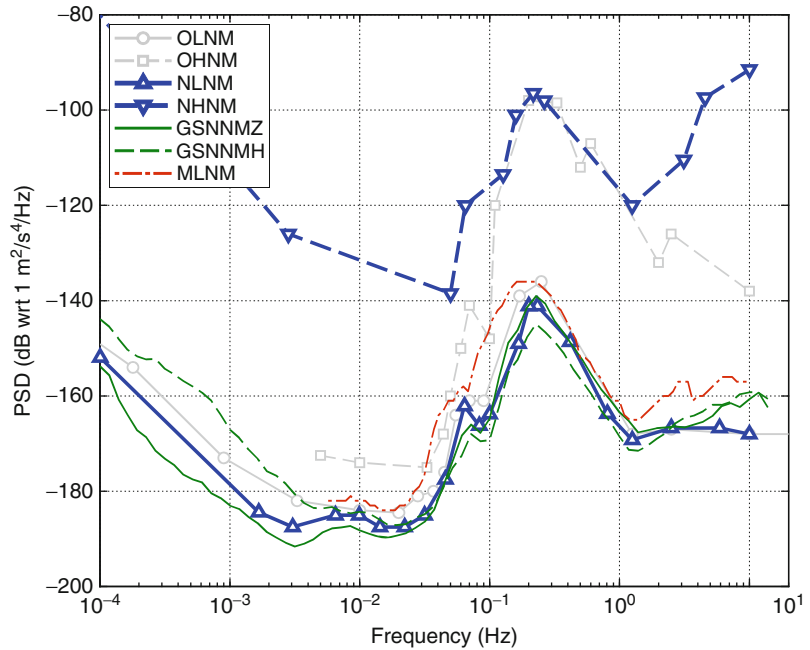
The sensing of ground motion due to earthquakes and other phenomena is fundamentally limited by the background motion of the earth. Efforts to define minimum, maximum, and typical ground motion models go back at least as far as the 1950s (Brune and Oliver 1959).

An important current standard consists of the new high- and low-noise models (NHNM and NHNM) (Peterson 1993) which in practice sets limits on the typical background motion at well-constructed vaults. It represented an improvement over the old noise models (OLNM and ONHM) because it included stations with the then-new STS-1 seismometer, the first high-quality very broadband sensor. These four models are plotted in Fig. 7 below.

In the band 0.05–1 Hz all models of ground motion are dominated by a peak called the microseismic peak. This peak is in fact made up of two peaks. The first peak, typically between 10 and 16 s period, corresponds to the natural period of waves generated by storm winds in mid-ocean. This peak is sometimes called the primary

Principles of Broadband Seismometry,

Fig. 7 Models of ground motion spectra



microseism. Because the mean pressure at the ocean bottom is proportional to the square of wave height, standing waves at mid-ocean produce a larger, double-frequency peak between 4 and 8 s period (Longuet-Higgins 1950). Ocean microseisms dominate recordings made by sensors with sufficiently low self-noise and divide ground motion into two important bands, commonly called short period (above the microseismic peak) and long period (below the microseismic peak).

Peterson dealt with the non-stationarity of the nonetheless random processes underlying the background motion of the earth by simply discarding time segments containing earthquakes or noise bursts. More recent models of ground motion spectra deal with the problem of non-stationarity in a more sophisticated way, using statistical approaches which evaluate the probability distribution of the power spectral density (PSD) as a function of frequency.

For example, studies of the Global Seismological Network (GSN) resulted in noise models of vertical (Z) and horizontal (H) ground motion (GSNNMZ and GSNNMH, respectively) (Berger et al. 2004), and a study of the continental United States has produced the mode low-noise

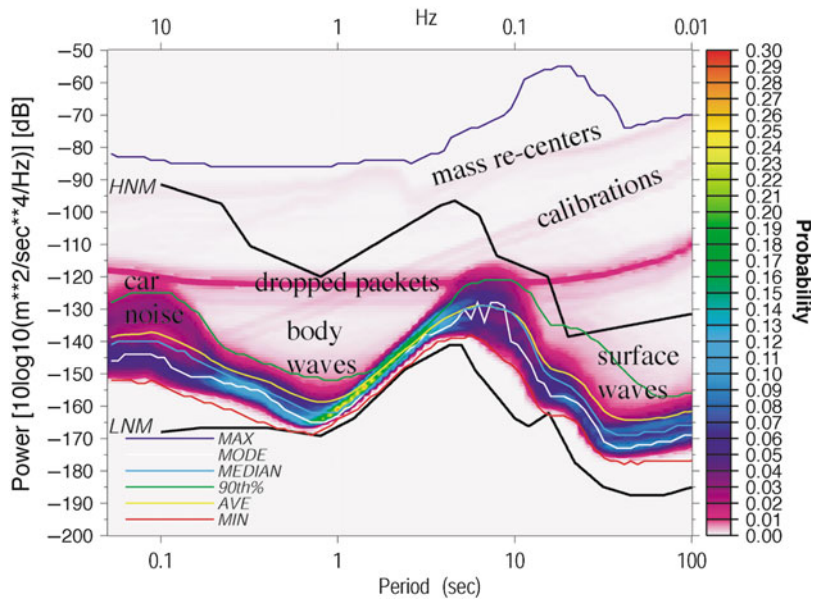
model (MLNM) (McNamara and Buland 2004): both studies employ the notion of the PSD probability density function (PDF) to produce estimates of ground motion spectra at quiet sites. Figure 8, reprinted from Figure 8 of McNamara and Buland (2004), shows how transient phenomena (e.g., calibration signal injected into the seismometer) are represented alongside stationary phenomena (e.g., the microseismic peak) using this approach.

The PSD PDF allows a more precise statistical definition of ground motion low-noise models. For example, the MLNM was constructed as the per-frequency minimum of the *modes* of the PDF of a set of broadband seismic stations distributed across the continental United States. Similarly, the GSNNM is the per-frequency minimum of the 1st (quietest) percentile of the PDF of the stations in the GSN. As shown in Fig. 7, the GSNNMZ is lower than the NLNM at very long periods (<0.01 Hz); this new lower limit is likely to still be limited by the performance of the STS-1.

Despite the advance in network performance monitoring that the PSD PDF represents, the NLNM remains the standard by which seismometer self-noise floors are measured, and many of

Principles of Broadband Seismometry,

Fig. 8 Typical acceleration power spectral density probability density function



the most common features of background ground motion are represented relative to the NLNM and NHNM.

Some strong generalizations can be made about background motion at long periods:

- Horizontal components are never quieter than vertical components at a given location. See section on “[Tilt](#).”
- Sites on alluvium are never quieter than nearby sites on hard rock. See section on “[Site Selection](#).”
- Surface sites are never quieter than nearby underground sites. See section “[Downhole and Ocean-Bottom](#).”

Indeed, the NHNM at long periods is made up of horizontal ground motion records at surface vaults on alluvium, while the NLNM at long periods is made up of vertical records on hard rock, mainly in subsurface vaults (Peterson 1993).

There are fewer features of ground motion at short periods (above the microseismic peak) which can be generalized. As at long periods, surface sites are never quieter at short periods than nearby underground sites, but horizontals and verticals tend to have similar levels. In both

bands “cultural noise” such as that due to traffic can be problematic, as can noise due to wind in trees or running water. Finally, there is always a minimum in the ground noise spectrum at or just above 1 Hz.

Although the NLNM was constructed from the records of land-based seismometers, it may also represent a hypothetical quietest site under the sea. Seismic noise surveys on the ocean bottom have not yet been undertaken on the scale that they have been on land. It is clear from the work which has been done so far, however, that the noise levels are generally significantly higher. Most of the ocean floor is covered with sediment which is more compliant than sediment on land, and seismic velocities are lower. Seafloor compliance means that excess site noise due to deformation and tilt can result from both variations in pressure due to waves at the surface and currents traveling along the bottom of the ocean (Webb and Crawford 2010). When precautions are taken to mitigate these effects, site noise can approach the levels observed on land.

In conclusion, although models based on larger or more regionally focused samples of sites exist, and although statistical methods have improved with time, Peterson’s NLNM remains the standard by which the noise performance of

broadband seismometers and sites are judged. It represents a hypothetical “quietest site on earth” so a good broadband seismometer should be able to resolve it. The NLNM is therefore commonly used as a reference curve on plots of broadband seismometer self-noise. In turn, at the very longest periods, it may be that the NLNM is limited by the available broadband sensor technology.

Ground Motion Unit Conversion

Ground motion can be expressed in the time domain units of displacement [m], velocity [m/s], or acceleration [m/s²]. In the frequency domain, a ground motion signal can furthermore be represented either as an amplitude spectrum [“per Hz,” e.g., m/s²/Hz] or as a PSD [“per $\sqrt{\text{Hz}}$,” e.g., m/s²/ $\sqrt{\text{Hz}}$]. In assessing the performance of a seismograph station, it is essential to be able to convert between various units fluently.

For example, comparison of earthquake spectra (typically a displacement amplitude spectrum or peak band-passed acceleration spectrum) to the clip level of a station (typically a maximum amplitude of acceleration or velocity) allows determination of the maximum earthquake magnitude which can be detected without clipping. Similarly, comparison of the same event spectra to seismograph self-noise or site noise (typically represented as an acceleration PSD) determines the minimum detectable magnitude. Finally, comparison of seismometer clip and self-noise levels as a function of frequency gives the clearest representation of seismometer dynamic range.

Acceleration-Velocity-Displacement

The simplest kind of conversion is integration and differentiation to obtain different time derivatives of ground motion. For example, to convert an acceleration to a velocity or from velocity to displacement at a given frequency f , one simply divides by the amplitude by the angular frequency $\omega = 2\pi f$. This follows straightforwardly from the equivalence of differentiation in the time domain to multiplication in the frequency domain by the Laplace frequency $s = j\omega$. Thus for sinusoidal amplitudes of acceleration $|a|$,

velocity $|v|$, and displacement $|d|$ at frequency f , the following relations hold:

$$|a| = 2\pi f |v| \quad (2)$$

$$|v| = 2\pi f |d| \quad (3)$$

These relations have wide applicability in converting amplitude spectra of finite energy signals of earthquakes between acceleration, velocity, and displacement and in converting PSDs of stationary between acceleration, velocity, and displacement.

Another type of conversion is needed for the comparison of finite energy signals and stationary signals. In particular, conversion from a PSD to the equivalent amplitude requires that a minimum of two factors be specified: a bandwidth and a crest factor.

Relative Bandwidth

Parseval’s theorem can be used to compute the root-mean-square amplitude a_{RMS} expected given the one-sided power spectral density as a function of frequency $a_{\text{PSD}}(f)$ and lower and upper frequency band limits $f_H > f_L$:

$$a_{\text{RMS}}^2 = \int_{f_L}^{f_H} a_{\text{PSD}}^2(f) df$$

This equation applies equally well to acceleration, velocity, or displacement, but note that if the RMS is in units, then the PSD is in units/ $\sqrt{\text{Hz}}$.

For narrowbands the spectrum can be assumed to be constant within the band so that

$$a_{\text{RMS}}^2 = a_{\text{PSD}}^2(f_H - f_L) = a_{\text{PSD}}^2 k_{\text{RBW}} f \quad (4)$$

$$a_{\text{RMS}} = a_{\text{PSD}} \sqrt{k_{\text{RBW}} f}$$

where the “relative bandwidth factor” is defined as

$$k_{\text{RBW}} \equiv \frac{f_H}{f} - \frac{f_L}{f}$$

Principles of Broadband Seismometry,
Table 1 Common choices of relative bandwidth

Bandwidth	b	n	k_{RBW}	$k_{\text{RBW}} _{\text{dB}}$
Octave	2	1	0.707	-1.5
1/6 decade	10	6	0.386	-4.1
1/2 octave	2	2	0.348	-4.6
1/3 octave	2	3	0.232	-6.4

It is common to define bands as a fraction $1/n$ of either an octave ($b, 2$) or a decade ($b, 10$), as follows:

$$\frac{f_H}{f_L} = b^{\frac{1}{n}}$$

and furthermore to dispose the band limits symmetrically (in a logarithmic scale) around the band center, so that

$$\frac{f_H}{f} = \frac{f}{f_L} = b^{\frac{1}{2n}}$$

so that the relative bandwidth factor k_{RBW} for various bandwidths can be evaluated using

$$k_{\text{RBW}} = b^{\frac{1}{2n}} - b^{-\frac{1}{2n}}$$

Some common choices of bandwidth are listed in Table 1. Two which appear commonly in seismometer specifications, 1/2 octave and 1/6 decade, are close enough that they are often considered interchangeable. The octave bandwidth is considered typical of a passive seismometer and has been used in some important studies of typical earthquake spectra (Clinton and Heaton 2002).

The choice of a relative bandwidth thus permits conversion of a power spectral density to an equivalent RMS amplitude.

Crest Factor

The second factor needed when comparing stationary and transient signals is called a “crest factor,” and is defined as the ratio of peak amplitude to RMS amplitude. The correct choice of crest factor depends on the nature of the signal in question. Broadband Gaussian noise will

surpass its standard deviation 32 % of the time and twice the standard deviation 5 % of the time, for example. These conditions would correspond to crest factors of 1 and 2, respectively.

The signals which make up background ground motion do often have Gaussian distributions (Peterson 1993), as do the self-noise of seismometers and digitizers. When a Gaussian signal is passed through a narrowband filter, the peak amplitudes of the signal which results have a Rayleigh distribution (Bormann 2002), corresponding to a crest factor:

$$k_{\text{crest}} = \sqrt{\frac{\pi}{2}} \cong 1.25$$

Given a crest factor, the equivalent amplitude can be estimated from a PSD using

$$a_{\text{peak}} = a_{\text{RMS}} k_{\text{crest}} \quad (5)$$

As with the equations accounting for relative bandwidth, this equation applies equally well to acceleration, velocity, or displacement.

Decibels

Ground motions are often expressed in decibels. By convention the reference amplitude for the decibel in this context is the corresponding metric unit. For example, an acceleration PSD is converted to dB using

$$a_{\text{PSD}}|_{\text{dB}} = 20 \log_{10} \left(\frac{a_{\text{PSD}}}{\frac{\text{m}}{\text{s}^2 \sqrt{\text{Hz}}}} \right)$$

while a velocity amplitude is converted using

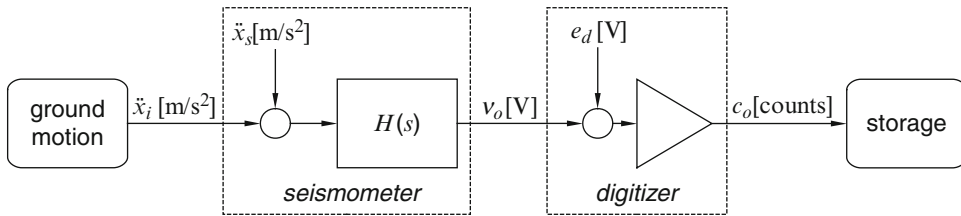
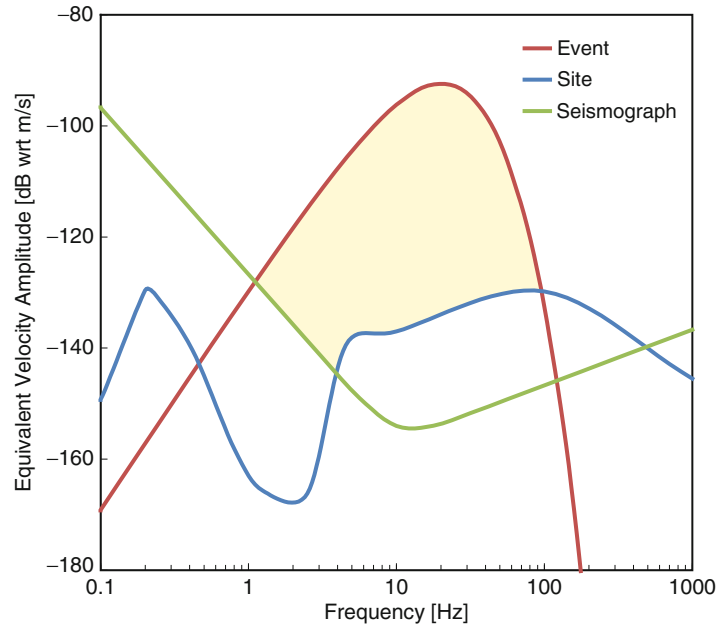
$$v_{\text{peak}}|_{\text{dB}} = 20 \log_{10} \left(\frac{v_{\text{peak}}}{\frac{\text{m}}{\text{s}}} \right)$$

Thus, the various unit conversions of Eqs. 2, 3, 4, and 5 become:

$$a|_{\text{dB}} = v|_{\text{dB}} + 20 \log_{10} \left(\frac{2\pi f}{\text{rad}} \right) \quad (6)$$

Principles of Broadband Seismometry,

Fig. 9 Representative station (noise) and small event (signal) spectra



Principles of Broadband Seismometry, Fig. 10 Components of a seismograph station

$$v|_{\text{dB}} = d|_{\text{dB}} + 20\log_{10}\left(\frac{2\pi f}{\text{rad}}\right) \quad (7)$$

$$a_{\text{RMS}}|_{\text{dB}} = a_{\text{PSD}}|_{\text{dB}} + 10\log_{10}(k_{\text{RBW}}) + 10\log_{10}\left(\frac{f}{\text{Hz}}\right) \quad (8)$$

$$a_{\text{peak}} = a_{\text{RMS}}|_{\text{dB}} + 20\log_{10}(k_{\text{crest}}) \quad (9)$$

These relations can be seen in action Fig. 13 below, where the band-passed peak velocity amplitudes of earthquakes are compared to the clip level and noise floor of a seismometer.

Components of Station Noise

The noise power at a seismic station is the sum of the noise powers of the seismograph and the background motion at the site:

$$\langle |\ddot{x}_{\text{station}}|^2 \rangle = \langle |\ddot{x}_{\text{seismograph}}|^2 \rangle + \langle |\ddot{x}_{\text{site}}|^2 \rangle$$

It is important that the station noise be low enough to not obscure the signals of interest when they arrive. This is shown in cartoon form in Fig. 9.

A modern seismograph station consists, in turn, of a seismometer and a digitizer, as shown schematically in Fig. 10. Both elements must be considered in determining the performance of the station as a whole.

Similarly the clip level of a seismograph station is a combination of the clip level of the seismometer and of the digitizer. A difference is that whereas noise powers are summed, the clip level of a system is the minimum of the clip level of all of the components, when referred to common units.

The importance of the sensitivity of a digitizer is now apparent. The sensitivity must be high enough to minimize the digitizer's contribution to the station noise, but at the same time it must be low enough that the clip level of the digitizer does not degrade the station clip level so much that events of interest will be clipped. Note that a system incorporating a seismometer which achieves a high sensitivity to ground motion using a high-gain output stage is equivalent to a system in which the digitizer subsystem includes a preamp with the same gain.

The self-noise of seismometers can be divided roughly into those that are stationary, in the sense that the probability distribution does not change with time, and those that are nonstationary.

Of the sources of stationary noise in a seismometer, there are two main subcategories, called "white" noise and "flicker" noise sources. White noise sources have the same amount of noise power per unit of (absolute) bandwidth in frequency; flicker noise sources have the same amount of noise power per fractional (relative) bandwidth, for example, per octave or per decade. The PSD of a white noise source is independent of frequency; the PSD of a flicker noise source varies as the inverse of frequency (Motchenbacher and Connelly 1993), i.e., proportional to $f^{-\alpha}$ where typically α is close to 1 but can be as high as 1.5.

In the process of referring the noise at various points in the feedback loop to the input of a seismometer, each source must be referred to the sensor input. This is similar to converting ground motion units from velocity to displacement and from displacement to velocity. Thus in principle various components of the noise floor of a seismometer can be proportional to f^n (or equivalently a slope of $10n$ dB/decade) where n is any positive or negative number, usually an integer. In practice, in a well-designed force-feedback seismometer, the equivalent acceleration PSD of the self-noise will be proportional to f^{-3} (−30 dB/decade) at low frequencies and proportional to f^4 (40 dB/decade) at high frequencies (Hutt and Ringler 2011).

For example, the displacement transducer in a feedback seismometer results in a white noise

proportional to displacement, but to determine the equivalent acceleration, a double integration is required, so the slope of the acceleration PSD is f^4 , and this tends to dominate the self-noise of the sensor at high frequencies. Similarly the digitizer white noise component, when converted from velocity to acceleration PSD, has a f^2 (20 dB/decade) characteristic and will tend to dominate the station noise near 10 Hz.

Two noise sources common to all seismometers, active and passive, are Johnson and Brownian noise. Johnson noise is a noise source due to thermal agitation of the electrons that make up the current flow in a resistor. For a resistance R [Ω] at temperature T [K], the expectation value of the voltage PSD across its terminals will be:

$$\langle |e_n|^2 \rangle = 4k_b TR$$

where $k_b = 1.38 \times 10^{-23}$ J/K is Boltzmann's constant. For low noise, resistors must have small values, but this generally results in greater power consumption.

Brownian noise is similar to resistor noise, both in terms of the statistical mechanics which explain it, and in the white noise spectrum which results. In particular, the Brownian motion of the boom of a pendulum results in an equivalent acceleration PSD of the frame of that pendulum equal to:

$$\langle |\ddot{x}_n|^2 \rangle = \frac{4k_b TB}{M^2}$$

Thus it is clear that for low noise, the proof mass must be large while the viscous damping must be small. All known voice-coil and capacitive transducers require narrow gaps for high sensitivity; this is a source of viscous damping and can only be avoided by pumping the pendulum enclosure down to a hard vacuum, something quite difficult in practice. The downside of large proof masses is increased size and fragility; many broadband seismometers required masses to be locked during transportation.

Nonstationary noise sources are also very important to the overall performance of a seismometer. A common example of nonstationary noise is what is sometimes referred to colloquially as a “pop.” The characteristic waveform is a randomly occurring disturbance having the shape of the impulse response of the seismometer. When such a signal is referred to the input of the seismometer, it turns out to correspond to a sudden step change in acceleration at the input to the seismometer, equivalent to a sudden step change in tilt of the instrument. Because “pops” occur at random times, they contribute a f^{-2} characteristic to the acceleration PSD noise spectrum.

“Pops” are often associated with temperature transients associated with initial installation and power-on. However, some sensors, which are poorly designed or have been insufficiently tested or damaged during handling, will exhibit an unusual susceptibility to pops after initial installation, and the excess “pop” noise will subsequently never die down to an immeasurable level. “Pop” noise is a low-frequency phenomenon which happens intermittently on a very long time scale. This is why, during testing of broadband seismometers, it is imperative to test sensors for a relatively long period of time, on the order of days or weeks, and include all data collected on a statistical basis, giving the median and maximum noise levels as much attention as the minimum noise level.

The PSD PDF (see section “[Ground Motion Spectra](#)”) is a particularly useful tool in performing this type of analysis. Figure 8 shows how the PSD PDF represents rare, transient events such as mass-centering operations are captured in the maximum of this distribution, while the typical behavior is captured in the mode. This allows the whole available record to be analyzed, without “window-shopping.” This is as it should be: single-occurrence “pops” can be as much of a problem in interpreting seismograms as stationary noise sources and should therefore be considered just as much a part of the overall self-noise performance. See ► [Seismometer Self-Noise and Measuring Methods](#).

Selection Criteria

Many factors need to be taken into consideration in choosing a broadband seismometer. Self-noise, clip level, and transfer function are the basic performance criteria, while power and size can crucially determine the cost of the overall installation. These factors and others are treated in this section.

Sensor Self-Noise

Broadband seismometers are required to have a self-noise less than the NLNM, a hypothetical “quietest site on Earth,” at least over some wide band. The self-noise cannot be measured deploying the sensor where there is no ground motion: the NLNM shows that there is no such place on Earth. Modern methods require at least three sensors be installed side by side on a rigid pier and a coherence-based analysis technique (Sleeman et al. 2006; Evans et al. 2010). Careful management of thermal, magnetic, and pressure shielding is required (see section on “[Installation Procedures](#)”), particularly to observe the performance of the best sensors at very long periods.

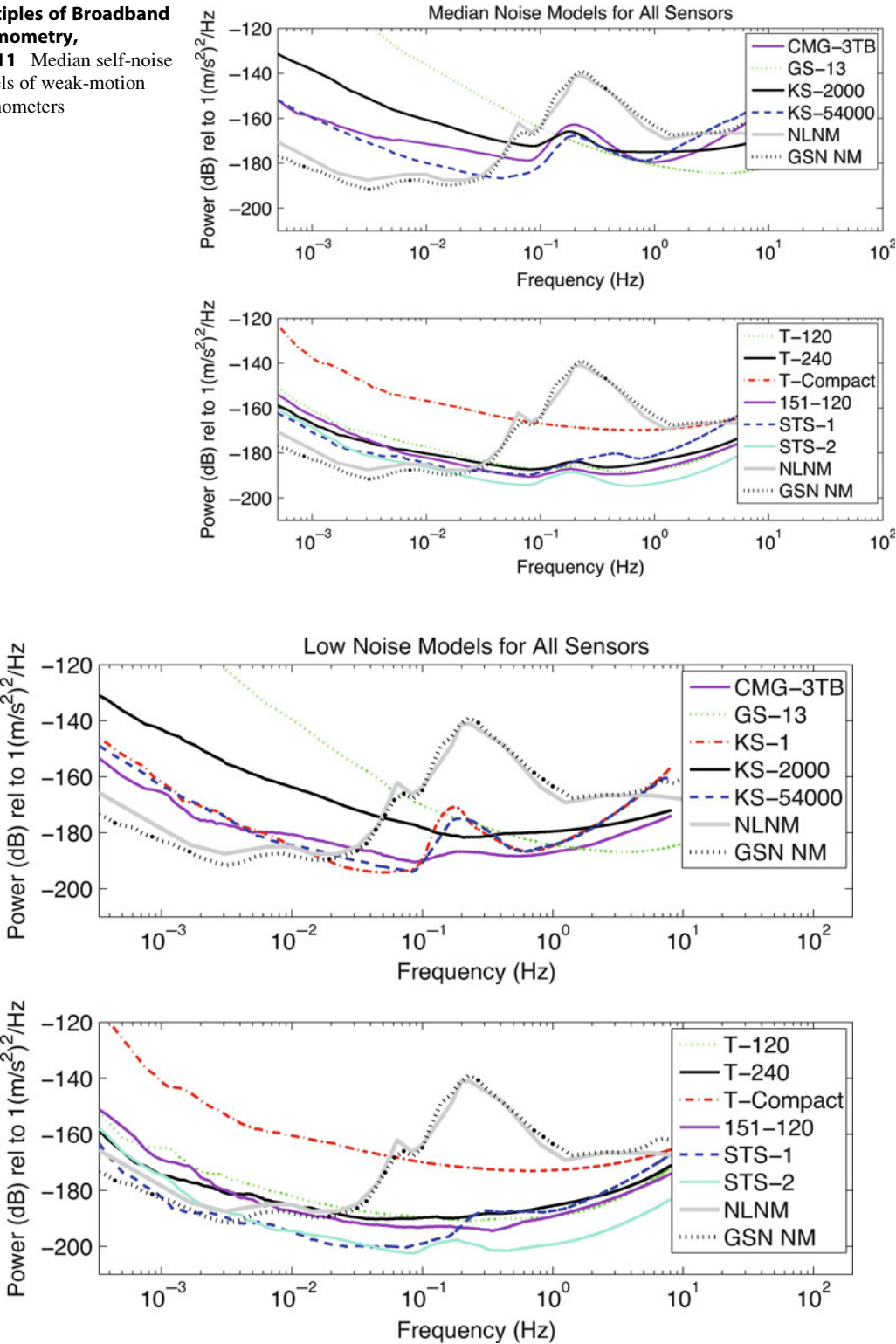
One comprehensive self-noise study (Ringler and Hutt 2010) includes all the important broadband seismometers as of its writing. Figures 11 and 12, reprinted from Figures 13 and 14 of that study, show the lowest 5th percentile and median noise of each sensor, respectively.

The abbreviations used in Figs. 11 and 12 are the same as those used in Table 2 below with the following exceptions: T-120, T-240, and T-Compact are the Nanometrics Trillium 120, Trillium 240, and Trillium Compact, respectively, and the 151–120 was the predecessor of the Geotech 151B-120. It is also worth noting that the STS-2 performance shown in these figures is that of a model with higher gain at 20,000 V·s/m and a correspondingly lower clip level, and that the CMG-3TB tested was a borehole variant of the standard CMG-3T.

The median self-noise was for some sensors close to the 5th percentile, while for others it was significantly higher. This can be taken as an indicator of the uniformity and/or the temporal stability of the sensors. The peak in the noise floors

Principles of Broadband Seismometry,

Fig. 11 Median self-noise models of weak-motion seismometers



Principles of Broadband Seismometry, Fig. 12 Low self-noise models of weak-motion seismometers

Principles of Broadband Seismometry, Table 2 Specifications of broadband seismometers

Manufacturer	Model	Sensitivity [V·s/m]	Clip [mm /s]	Period [s]	Upper corner [Hz]	Power [W]	Weight [kg]	Sensor volume [L]	Shield volume [L]
Geotech	KS-1	2,400	8	360	5	2.4	43	28	–
Streckeisen	STS-1	2,400	8	360	10	10.5	15	14	450
Nanometrics	Trillium 240	1,200	16	240	200	0.65	14	14	60
Nanometrics	Trillium 120	1,200	16	120	175	0.62	7.2	7.2	32
Guralp	CMG-3 T	1,500	13	120	50	0.75	14	8.4	–
Streckeisen	STS-2	1,500	13	120	50	0.8	13	11	72
Geotech	KS-2000	2,000	10	120	50	0.9	11	8.1	–
REF TEK	151B-120	2,000	10	120	50	1.1	12	12	–
Geodevice	BBVS-120	2,000	10	120	50	1.4	14	8.3	–
Nanometrics	Trillium Compact	750	26	120	100	0.16	1.2	0.8	7.8
Guralp	CMG-3ESP	2,000	10	60	50	0.6	9.5	4.7	–
Guralp	CMG-40 T	800	25	30	50	0.46	2.5	3.9	–

Volumes are for three components and are derived from footprint area and height alone

directly under the microseismic peak is not a feature of the sensor noise floor, but an artifact of imperfect misalignment correction.

Given the importance of the NLNM in setting the target performance of broadband seismometers, it is understandable that self-noise requirements are often specified in terms such as “below the NLNM from 30 s to 30 Hz.” Although the intention is clear, this is unfortunately a poor choice of wording. In some bands the NLNM is very steep – which is to say that the typical background acceleration PSD at a quiet site is also very steep – and so a great reduction in self-noise may not result in much improvement in the frequencies at which the self-noise crosses the NLNM. Conversely, where the slope of the NLNM is quite similar to that of the self-noise of a seismometer, a small improvement in self-noise will produce an exaggerated improvement in the NLNM-crossing frequency. Finally, care must be taken at high frequencies because the NLNM is simply not defined above 10 Hz (similarly the MLNM does not extend above 10 Hz and the GSNM does not extend above 13 Hz).

A better way to set a specification on seismometer self-noise is to pick two key frequencies near the edges of the band of interest and set limits in dB at those frequencies. For example,

to require the “median self-noise below –185 dB at 0.01 Hz and below –168 dB at 10 Hz” is equivalent to being “below the NLNM from 100 s to 10 Hz,” but for the purposes of comparing one seismometer to another, it is more instructive to compare their median performance in dB at specified frequencies than in the NLNM-crossing frequencies in Hz.

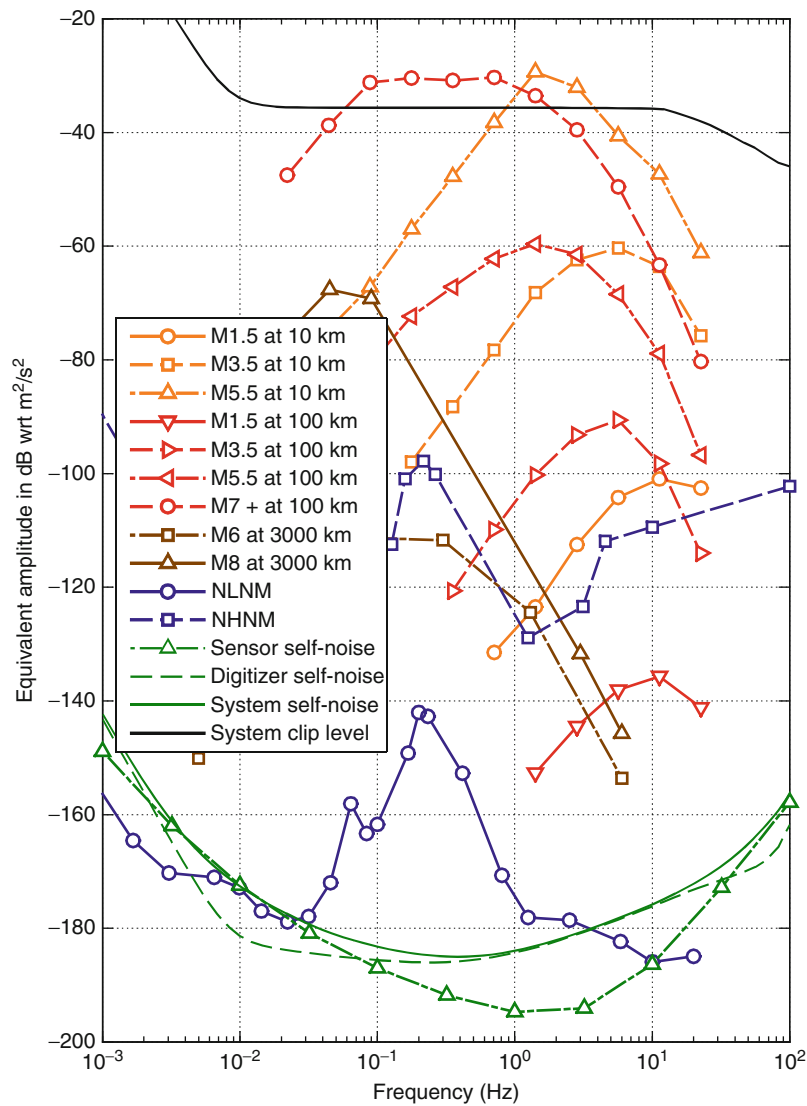
Clip Level

The largest measured ground motions resulting from earthquakes have approached peak accelerations of 40 m/s² and peak velocities of 3 m/s, although these limits tend to increase with the densification of networks and the passage of time (Strasser and Bommer 2009). No seismograph can measure both the largest and the smallest ground motions. The dynamic range of the current state of the art in analog-to-digital converters is insufficient to cover the >170 dB range required. Although weak-motion seismometers focus on measuring the smallest motions, they are still expected to measure relatively large motions without clipping, so that weak-motion studies are not “interrupted” by strong motion.

Whether the clip level of a particular seismometer is adequate depends on the expected levels of ground motion at the site of deployment.

Principles of Broadband Seismometry,

Fig. 13 Earthquake ground motions and seismograph clip levels and noise floors



The ground motion resulting from an earthquake varies with the magnitude but also with the distance from the event.

The ground motion resulting from earthquakes is a function of frequency. Figure 13 shows the typical ground motions for large and small earthquakes at local (10 km), regional (100 km) and teleseismic (3,000 km) distances (Clinton and Heaton 2002).

These event spectra are plotted for comparison against the noise floor and clip level of a representative broadband seismograph. It can be seen that the digitizer noise floor is limiting the

performance of the seismograph system for frequencies above the microseismic peak. The station operator could choose to increase the digitizer preamp gain or equivalently install a high-gain version of the seismometer. This would result in a reduction of the contribution of the digitizer to the system noise and therefore make it possible to see smaller events, but it would also reduce the clip level of the system, so that large events would be more likely to clip.

The clip level of a seismometer in the middle of its passband depends solely on the output clip level in volts and its sensitivity. Notwithstanding

a hypothetical seismometer which runs from very high voltage rails, a requirement for high clip level is low sensitivity, and vice versa. The sensitivity and clip levels of some representative broadband seismometers are given in Table 2, along with other critical performance criteria.

Broadband seismometers generally use displacement transducers to measure the relative displacement of the proof mass and the frame. The proof mass and its suspension together make a pendulum which has a response which is flat to acceleration and independent of its natural frequency above the natural frequency.

There is usually a critical frequency above which the clip level of a broadband seismometer becomes more or less flat to acceleration. For the seismometer depicted in Fig. 6, the acceleration clip level is 0.17 g and the velocity clip level is 16 mm/s, so this critical frequency is near 16 Hz. If this critical frequency is significantly higher than the peak frequency of the spectra of events of interest, it will not affect clipping behavior for real seismic signals.

Response

It is crucial to understand the transfer function of a broadband seismometer when converting a recording in counts or volts back to appropriate units of ground motion. It is a common mistake to think that the signals at frequencies below the lower -3 dB corner or above the upper -3 dB corner are not useful. In fact the only thing which determines whether or not useful signal is present is signal-to-noise ratio, as illustrated by Figs. 9 and 13. With careful application of the inverse response, useful estimates of ground motion can be obtained well outside the -3 dB band.

Seismometers are sometimes available with different options for lower or upper corner frequency and for the mid-band sensitivity. Note that the term “sensitivity” in the context of seismometers is interchangeable with “generator constant” and will have units of voltage per unit of velocity. When selecting a seismometer, it is important to have in mind the largest signal which is likely to be observed, given the seismicity of the nearest seismogenic zone. Seismic risk

maps are invaluable for determining the probability of recurrence of a given peak ground acceleration or velocity, which can then be directly compared to the configured seismograph system clip level. High sensitivity means the contribution of the digitizer to the station noise floor is reduced, but it also means the system clip level is reduced. Some users will prefer to use a low-gain seismometer and a digitizer with a built-in variable-gain preamp, so that they can “dial in” the correct station sensitivity after installation.

One feature of a seismometer not normally represented in its nominal transfer functions is the phenomenon of parasitic resonances. Well-characterized seismometers will have a specification for the lowest mechanical resonance; it is important to make sure that this frequency is above the range of frequencies of interest in a particular study.

Power

There are two advantages to low power. The first is that lower power means a physically smaller footprint and a less costly installation for stations which must be located far from main power systems. Such stations are common because the best seismic sites are generally located away from roads and cities and human activity in general, so-called sources of cultural noise. For a temporary deployment, lower power means fewer batteries are needed for a given length of time. For a permanent deployment at a remote site, lower power means less on-site power generation (e.g., fewer solar panels) is needed. A smaller footprint for power generation furthermore generally means less wind-induced seismic noise.

A second advantage to lower power relates specifically to performance at very long periods in vault-type installations. Power dissipation inside the sensor and digitizer means heat generation. This heat causes convection within the vault, and the resulting airflow tends to be turbulent and chaotic, heating and cooling various surfaces around the vault, in particular the floor, causing small but measurable tilts. Thus, for sensors which consume more power, it becomes more difficult to properly thermally shield well

enough to drive the resulting apparent horizontal accelerations down below the NLNM at very long periods.

The power consumption of some broadband seismometers is listed in Table 2.

Size and Weight

The physical size of a seismometer has a multiplying effect on the size and thus the cost of deployment of a seismic station. For vault installations, a larger sensor requires a larger volume to be reserved for thermal shielding. Since a good broadband vault must generally be built below ground level, a larger sensor means the minimum volume which must be dug out for the vault is larger.

For temporary deployments, the size and weight of a sensor can significantly affect ease of deployment. When it is a matter of driving to a remote location and hand-carrying the equipment even further away from the road, it can mean that significantly fewer stations can be set up per day, if the sensor is large and heavy.

Aside from the trouble it causes in a temporary deployment, a heavy sensor has an advantage over a light one, in that the associated thermal mass means better temperature stability.

The volume and weight of some broadband seismometers are listed in Table 2.

Enclosure, Leveling, and Topology

The choice of enclosure for broadband seismometers is an important one. Some common options are vault, borehole, posthole, and ocean bottom.

Enclosures designed for deployment in vaults need to be dust- and watertight, but are generally not designed for submersion to significant depths or durations (i.e., ingress protection ratings of IP66 or IP67 are common, but not IP68). There is no particular restriction on the overall diameter of a vault enclosure, but it should be designed for ease of leveling, orientation, and thermal isolation. For example, the connector should be oriented to allow cables to exit the enclosure horizontally near the surface of the pier. This makes it easy to strain-relieve the cable, minimizing the possibility of cable-induced noise, and

to place an insulating cover over top of it. The design of broadband seismometer vaults is described in “[Installation Procedures](#)” below.

Enclosures designed for deployment in cased boreholes generally need to have smaller diameters and a mechanism to lock the sensor in the hole. Drilling of boreholes is always more economical for smaller diameters than larger ones; a common casing diameter for broadband seismometers is 15 cm. The connector will generally exit at the top of the sensor and the whole assembly should be rated for continuous submersion to a significant depth (i.e., IP 68 to 100 m or more), since flooding of boreholes is common. Boreholes stray from verticality as they are dug deeper; a remote leveling range of up to $\pm 4^\circ$ is thus typically required. See ► [Downhole Seismometers](#) for a more detailed discussion.

Enclosures designed for deployment in shallow uncased holes called postholes do not need hole locks. Sensors are generally emplaced in backfilled soil or sand and are simply pulled out or dug out at the end of the deployment. As with borehole sensors, the connector should generally exit at the top and must be rated for submersion. There is less control of sensor leveling with deeper holes, and a remote leveling range of $\pm 10^\circ$ may be required. See ► [Downhole Seismometers](#) for a more detailed discussion.

Enclosures designed for ocean-bottom deployment have several requirements which other sensor types do not. Most of the ocean bottom is near 5 km depth, so in order to be deployable over most of the ocean bottom, a sensor would typically have a continuous submersion rating of 6 km. Most ocean-bottom deployments are done by releasing the sensor at the surface and without controlling exactly where it will come to rest on the ocean floor. The sensors are designed to level themselves, typically at a predetermined time after release, and since the exact resting place is not known in advance, a self-leveling range of $\pm 45^\circ$ or more is required. Prevention of corrosion and biofouling is additional crucial requirement for ocean-bottom enclosures. See ► [Ocean-Bottom Seismometer](#) for more information.

When an underground vault in bedrock is available, for example, in an inactive mine or in the basement of a building, then a vault-type enclosure is of course the best choice. When there is some significant overburden, so that to reach bedrock a borehole must be dug and cased, then a borehole-type enclosure is required. And of course ocean-bottom deployments require an ocean-bottom enclosure.

It is not uncommon however that a sensor must be deployed in a location where no preexisting vault or borehole is available. In such situations a posthole installation can give performance as good or better than a vault built according to best practices, at significantly less cost for the overall installation.

If leveling motors are included in the sensor, a remote leveling process is initiated via an external electrical signal or at a configurable time after power-on; otherwise manual leveling by adjustment of set screws is sometimes needed. All enclosure types except vault enclosures require remote leveling capability.

Sensor axis topology is a final consideration. Some studies may require only a single axis of seismic sensing, usually vertical. For triaxial seismic sensing, the sensor outputs should be horizontal (X, Y) and vertical (Z). However certain kinds of installation troubleshooting are easier to do if the internal sensing axes are not aligned to horizontal and vertical. See ► [Symmetric Triaxial Seismometers](#) for more information.

Environmental Sensitivities

Spurious signals due to environmental sensitivities are not normally considered part of the self-noise of a seismometer, but they can deleteriously affect the output signal in many of the same ways. Broadband seismometers are particularly sensitive to changing tilt, temperature, pressure, and magnetic fields.

To understand why, consider that in order to have self-noise just equal to the NLNM at 100 s period, you need to be able to discriminate ground motion from all other effects at a level of

$$a_{\text{PSD}} = 10^{\frac{-185}{20}} \frac{\text{m}}{\text{s}^2 \sqrt{\text{Hz}}} = 0.56 \frac{\text{nm}}{\text{s}^2 \sqrt{\text{Hz}}}$$

And since the self-noise of a seismometer is typically proportional to $1/f$ in this band, the noise in the decade around $f_{\text{PSD}} = 0.01$ Hz will be

$$a_{\text{NLNM}} = a_{\text{PSD}} \sqrt{f_{\text{PSD}} \ln(10)} \frac{\text{nm}}{\text{s}^2} = 0.08 \frac{\text{nm}}{\text{s}^2}$$

This tiny acceleration, measurable by very broadband seismometers (i.e., a weak-motion inertial sensor with a wide dynamic range over a very broadband), can be overwhelmed by spurious environmental sensitivities, as discussed below.

Tilt

Sensitivity to tilt is an inevitable consequence of inertial sensing, because gravitational equivalence principle tells us that gravity is indistinguishable from accelerations. An inertial sensor tilted from vertical by an angle θ measured in radians experiences an apparent horizontal acceleration of

$$\ddot{x} = g_0 \sin \theta$$

where $g_0 \cong 9.8 \text{ m/s}^2$ is the standard acceleration due to gravity near the surface of the Earth. For small tilt angles $\sin \theta \cong \theta$, so all inertial seismometers have the same tilt sensitivity α_T , that is, the same apparent horizontal acceleration in response to tilt:

$$\alpha_T \equiv \frac{\ddot{x}}{\theta} = g_0 \cong 9.8 \frac{\text{m/s}^2}{\text{rad}} \cong 0.17 \frac{\text{m/s}^2}{^\circ}$$

Some tilt and rotation is to be expected to accompany the translational motion of a traveling seismic wave, but locally generated non-seismic tilt can prevent critical observations from being made. It is because of their extreme sensitivity to tilt that all inertial translational sensors, including broadband seismometers, record higher levels of apparent horizontal motion than vertical motion at long periods.

In order to resolve the NLNM at 100 s, tilts in the decade band around that frequency would have to be kept smaller than

$$\Delta\theta = \frac{a_{\text{NLNM}}}{g_0} = 5 \times 10^{-10} \text{ }^\circ$$

This is an extremely small angle. It corresponds to lifting one side of a 10 m wide structure pier by just 1 Å, the order of magnitude of atomic radii.

Fortunately, locally generated “excess” tilt is not spontaneous but driven by some other environmental factor and can be greatly reduced with careful vault design. For example, it is common for such tilts to be driven by temperature or pressure sensitivity of the seismic vault or nearby subsurface geology. Tilts can also be driven by changes in insulation or water table, vehicular traffic or other cultural activity, or wind loading on nearby structures. Mitigating these sorts of effects is an overriding concern in designing a seismic vault, as described in the section “[Installation Procedures](#).”

Another, more subtle tilt-related effect is that a static tilt will increase off-axis coupling of horizontal motion into vertical. See ► [Symmetric Triaxial Seismometers](#) for more information. Other than this effect, the actual static tilt of a seismometer is generally not a problem, as long as it is within the operating range of the seismometer.

Most broadband seismometers have an integrator in the feedback circuit, and the operating range of this part of the circuit determines the tilt range of the seismometer. For the lowest possible noise at very long periods, the integrator output resistor must be large, and this restricts the tilt range of the seismometer. Thus very broadband seismometers are typically equipped with centering motors or leveling platforms to extend this tilt range. See ► [Downhole Seismometers](#) for more information.

Temperature

The operating temperature range of a seismometer is determined by the temperature coefficient of the mechanics and components in the force-feedback circuit. A vertical seismometer involves

balancing the acceleration due to the effect of gravity on the proof mass against forces supplied by a suspension. The temperature coefficient relates changes in deflection of the proof mass with temperature and so can be expressed in units of ppm (with respect to g_0 the acceleration due to gravity) per $^\circ\text{C}$. A temperature-compensated axis assembly is one in which changes in forces due to thermoelastic coefficients in the suspension cancel deflections due to coefficients of thermal expansion in the rest of the components (Wielandt 2002), such that a displacement transducer would register no movement of the proof mass.

Some broadband seismometers have very wide temperature ranges, encompassing the full range of possible deployment temperatures. Many of the broadband seismometers with the lowest self-noise, however, have operating temperature ranges of $\pm 10^\circ\text{C}$ or less. These seismometers are equipped with a re-centering mechanism which must be activated after the seismometer has been installed in a new vault, ideally after its temperature has stabilized. Just as horizontal sensitivity to tilt determines the tilt range, vertical sensitivity to temperature determines the temperature range.

Even for a sensor operating well within its temperature range, spurious horizontal or vertical output signals can result from tiny changes in temperature. The temperature sensitivity is typically a direct proportionality of equivalent input acceleration to change in temperature.

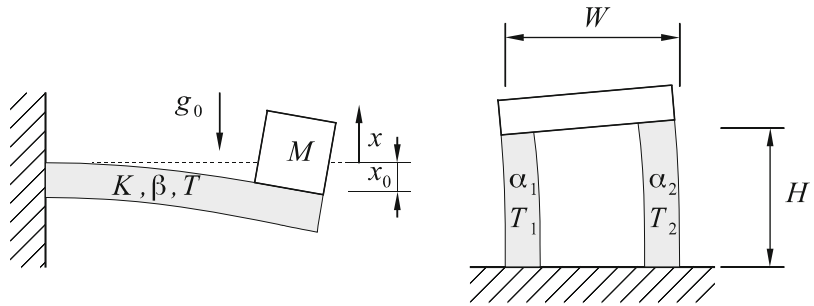
For a seismometer which is not temperature compensated, the temperature sensitivity is typically dominated by the thermoelastic coefficient of the mainspring, as shown in Fig. 14 (left). The cantilever balances the mass M against the force of gravity g_0 but as the temperature T changes the stiffness of the beam, represented as a spring constant K changes, and the apparent vertical acceleration $\ddot{x}$ changes.

Summing the forces on the mass M in Fig. 14 (left), we find

$$\sum F = -Kx + Mg_0 = M\ddot{x}$$

Principles of Broadband Seismometry,

Fig. 14 Schematic representation of vertical (left) and horizontal (right) temperature sensitivity



And the temperature sensitivity can be modeled as

$$K = K(\Delta T) = K_0(1 + \beta \Delta T)$$

So if the instrument is designed so that at $T = 0$, there is no apparent acceleration $\ddot{x} = 0$ and the deflection is static at x_0 :

$$x_0 = \frac{Mg_0}{K_0}$$

Now if we allow $T \neq 0$, the apparent acceleration is

$$\begin{aligned} \ddot{x} &= g_0 - \frac{x_0}{M} K_0(1 + \beta \Delta T) \\ &= g_0 - g_0(1 + \beta \Delta T) = g_0 \beta \Delta T \end{aligned}$$

Most copper alloys and steels have a thermoelastic coefficient on the order of $\beta = -300 \text{ ppm}/^\circ\text{C}$, the minus sign indicating that the mainspring relaxes with an increase in temperature.

This translates to a requirement for temperature stability of

$$\Delta T_z = \frac{a_{\text{NLNM}}}{|\beta|g_0} \cong 7 \times 10^{-9} \text{ } ^\circ\text{C}$$

A related problem is to measure ground motion on the order of the NLNM in the presence of temperature-generated tilts. This problem is significant both at the level of the seismometer and its subassemblies and at the level of the seismic vault and related superstructures. One way to model this effect is to visualize the sensor as a

platform with legs which either have different temperature coefficients or which are at different temperatures, as shown in Fig. 14 (right). For such a structure, small differentials produce small tilts:

$$\Delta a = \Delta \theta g_0 = (\alpha_1 T_1 - \alpha_2 T_2) \frac{H}{W} g_0$$

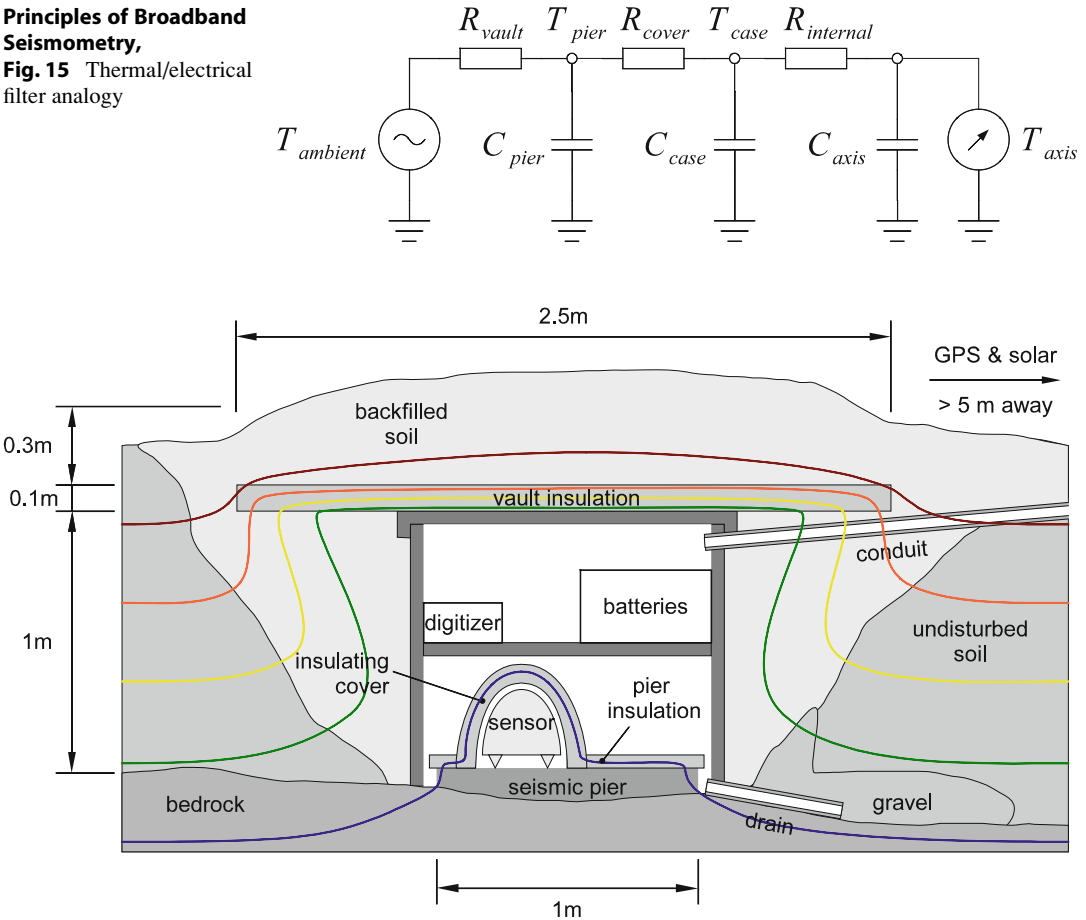
For an enclosure made out of a single material, all that matters is the difference in temperature across the structure. For an enclosure made out of steel or aluminum, the thermal coefficient of expansion is on the order of $\alpha = +20 \text{ ppm}/^\circ\text{C}$, with the positive sign indicating that the material expands with an increase in temperature. For the resulting equivalent horizontal acceleration to be less than the NLNM at a 100 s period, if the height is the same as the width, the temperature difference across the enclosure must be less than

$$\Delta T_x = \frac{a_{\text{NLNM}}}{\alpha g_0} \cong 1 \times 10^{-7} \text{ } ^\circ\text{C}$$

Obviously the actual dimensions of the structure can result in this effect being significantly amplified or attenuated, as can the geometry and relative stiffness of the members. Furthermore, it is important to note in both cases that static temperatures are not a problem because the seismometer does not respond to static acceleration. What matters is temperature variation with time, in the band of interest, in this case near 100 s period. Although this model is extremely simplistic, the point is that mechanisms for the conversion of changes in temperature into tilt abound, both inside a seismometer and outside.

Principles of Broadband Seismometry,

Fig. 15 Thermal/electrical filter analogy



Principles of Broadband Seismometry, Fig. 16 Idealized broadband vault

There is a subtle difference between the vertical temperature sensitivity and the horizontal temperature sensitivity of a seismometer. The vertical output of a seismometer is only sensitive to bulk temperature changes; the horizontal outputs are sensitive to differences in temperature across the surface of some part of the enclosure or differences in coefficient of thermal expansion combined with bulk changes in temperature.

The design of a thermal shield against bulk temperature changes is relatively straightforward and can be conceptualized by analogy with electronic circuits as designing a series of cascaded single-time-constant thermal low-pass filters. The design of bulk thermal isolation structures requires a series of concentric shells of thermally

resistive elements and thermally massive elements.

In Fig. 15 the first layer of thermal insulation R_{vault} shields the thermal mass C_{pier} of the pier, the sensor insulating cover R_{cover} shields the thermal mass C_{sensor} of the sensor, and the internal insulation of the sensor $R_{internal}$, if any, shields the thermal mass C_{pier} of the individual axes of the sensor. It is important when designing thermal shields to avoid accidentally including any thermal short circuits which decrease the effectiveness of the shielding. A thermal short circuit or thermal bridge is any path which crosses a thermal insulator and has high thermal conductivity. For example, if the electrical conduit in Fig. 16 was made of metal and therefore thermally conductive, it will act as a thermal short circuit and

degrade the performance of the vault insulation. This could be visualized in the electrical analogy of Fig. 15 as a low-value resistance in parallel with R_{vault} . A seismometer which includes some thermal insulation measures within its pressure vessel will require less external thermal shielding.

Bulk thermal isolation of the type described thus far primarily addresses the thermal sensitivity of the vertical output of a seismometer. Inhibiting thermally generated noise on the horizontal outputs of a seismometer is a different problem.

First, a vault must be free of drafts. At the same time, a fully sealed vault can make the vault respond to pressure with tilt, particularly if the vault is not installed on competent rock. The solution in some cases is to design the vault to have a single point at which it vents, so that the internal and external air pressures are equalized without generating drafts across the floor of the vault.

Second, air convection within the vault must be inhibited. This is done by reducing the power dissipated within the vault, by filling airspace within the vault with some material which inhibits airflow. The design of bulk thermal insulation is subtly different from the design of a shield intended to stop convective airflow. See the section on “[Installation Procedures](#)” below for more detail.

Pressure

A seismometer not contained within a pressure vessel will exhibit strong pressure sensitivity on the vertical output due to buoyancy of the proof mass. Consider a proof mass with a density $\rho_{\text{proof}} = 8 \text{ g/cm}^3$ at a temperature $T_{\text{air}} = 293 \text{ K}$ in dry air with a specific gas constant of $R_{\text{air}} = 287 \text{ J/kg} \cdot \text{K}$ in standard gravity g_0 . For such a seismometer, the vertical sensitivity to air pressure changes due to buoyancy is (Zürn and Wielandt 2007)

$$\alpha_B = \frac{g_0}{P_{\text{air}}} \frac{\rho_{\text{air}}}{\rho_{\text{proof}}} = \frac{g_0}{R_{\text{air}} T_{\text{air}} \rho_{\text{proof}}} = 15 \frac{\text{nm}}{\text{s}^2 \cdot \text{Pa}}$$

In order to be able to measure motions on the order of the NLNM at 100 s period, we need to keep variation in pressure under

$$\Delta P_B = \frac{a_{\text{NLNM}}}{g_0} R_{\text{air}} T_{\text{air}} \rho_{\text{proof}} = 0.006 \text{ mPa}$$

This requirement is stringent enough that if the pressure vessel of a seismometer is compromised, the vertical output will be dominated by this buoyancy effect.

A pressure vessel must be well designed in order to ensure that changes in atmospheric pressure do not produce equivalent horizontal or vertical outputs. Three critical specifications, then, are pressure attenuation of the pressure vessel and the pressure sensitivities of the vertical and horizontal outputs.

For the vertical channel of a seismometer in a pressure vessel, the limiting pressure effect is that due to atmospheric gravitation. Using the Bouguer plate model, in which the atmosphere above a station is modeled as a cylindrical plate having constant density, the gravitational pressure sensitivity due to atmospheric gravitation is (Zürn and Wielandt 2007)

$$\alpha_G = 2\pi \frac{G_0}{g_0} = 0.043 \frac{\text{nm}}{\text{s}^2 \cdot \text{Pa}}$$

Where the universal gravitation constant is $G_0 = 6.67 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2$. This, then, places a design constraint on a pressure vessel for a broadband seismometer: the pressure vessel will deform in response to changes in atmospheric pressure and result in a corresponding vertical acceleration, but the resulting pressure sensitivity should be less than α_G .

With this number in hand, we can reconsider the effect of buoyancy on the pressure vessel. An increase in atmospheric pressure will cause the volume of air inside the pressure vessel to become smaller; the rigidity of the vessel determines how much smaller. In order for the buoyancy effect to be much smaller than that due to the unavoidable effect of atmospheric gravitation, the pressure vessel must attenuate pressure changes by a factor much greater than

$$k_P = \frac{\alpha_B}{\alpha_G} = 340$$

For the horizontal channels of a seismometer, pressure sensitivity arises because atmospheric loading deforms the ground near the seismometer and produces measurable tilts. The level of sensitivity depends on geology and on depth, with shallow installations on unconsolidated sediment having the greatest sensitivity. At the Black Forest Observatory, the vault is 150–170 m below the surface in hard rock, and the measured admittance is typically (Zürn et al. 2007)

$$\alpha_T = 0.3 \frac{\text{nm}}{\text{s}^2 \cdot \text{Pa}}$$

Thus we can set a reasonable limit on the required horizontal pressure sensitivity of a broadband seismometer. When the pressure vessel deforms in response to pressure and causes the horizontal outputs to tilt and exhibit an apparent horizontal acceleration, the resulting pressure sensitivity should be less than α_T .

With both horizontal and vertical pressure sensitivities, a coherence analysis can be used to find a least-squares best-fit to the relative transfer function. This best-fit pressure sensitivity can then be used to correct seismic records for pressure, if a sufficiently sensitive microbarometer is colocated with the seismometer and recorded.

Magnetic

Temperature-compensated materials suitable for mainsprings tend to be magnetic, so inertial sensors tend to be susceptible to magnetic fields. The susceptibility takes the form of direct proportionality of equivalent output acceleration to magnetic field strength.

The magnetic sensitivity of a very broadband seismometer can vary between 0.05 and 1.4 m/s²/T (Forbriger et al. 2010). In order for the magnetic sensitivity of a seismometer to not interfere with measurement of ground motion down to the NLNM at 100 s during a magnetically quiet period, the vertical magnetic sensitivity must be less than

$$\alpha_M = 0.7 \frac{\text{nm}}{\text{s}^2 \cdot \text{T}}$$

For installations which must produce quiet records even during geomagnetic storms, a

magnetic shield can be used. Such shields are typically constructed of a high-permeability metal such as permalloy or mu metal.

Geomagnetic storms are not the only source of low-frequency magnetic fields. Hard drives and solar chargers are two examples of common equipment at seismograph stations which tend to generate interfering signals, and should therefore be located as far as away from the seismometer as is practicable.

As with pressure sensitivity, a best-fit magnetic sensitivity can then be used to correct seismic records for pressure, if a sufficiently sensitive magnetometer is colocated with the seismometer and recorded.

Site Selection

There is no substitute for a geological survey when it comes to site selection.

A site survey provides knowledge of the structures over which the seismometer will be installed. Where possible, seismometers should be installed on bedrock and as far away as possible from sources of cultural noise such as roads, dwellings, and tall structures.

The most important factor to consider in terms of geology is the composition of the uppermost stratum. For example, when the boundary of the uppermost layer is clearly defined as roughly horizontal, the S-wave velocity and thickness of that layer will determine the fundamental resonant frequency at that site. Lower velocities and larger drift thicknesses produce greater site amplification at lower frequencies.

Table 3, reprinted with permission from Trnkoczy (2002), grades site quality according to types of sediments or rocks and gives a sense of how the quality of a site relates to the S-wave velocity.

Low porosity is furthermore important as water seepage through the rock can cause tilts which overwhelm the seismic signal at long periods. Clay soils and, to a lesser extent, sand, are especially bad in this sense.

Principles of Broadband Seismometry,
Table 3 Classification of rock for site selection

Site quality		S-wave velocity (m/s)	
		Min	Max
1	Unconsolidated (alluvial) sediments (clays, sands, mud)	100	600
2	Consolidated clastic sediments (sandstone, marls); schist	500	2,100
3	Less compact carbonatic rocks (limestone, dolomite) Less compact metamorphic rocks; conglomerates, breccia, ophiolite	1,800	3,800
4	Compact metamorphic rocks and carbonatic rocks	2,100	3,800
5	Magmatic rocks (granites, basalts); marble, quartzite	2,500	4,000

Installation Procedures

The question of how to get the best performance out of a seismometer is a very involved topic. In this section the focus will be on broadband seismometers in vault enclosures, and in particular the edges of the usual band of interest, near 100 s period and 10 Hz frequency.

Underground Vaults

The STS-1 defined the limits of ground motion measurement at long periods after it was developed, around 1980. Each component of motion, vertical and horizontals, was detected by a mechanical assembly and controlled by a set of feedback electronics in a separate enclosure.

The advent of the STS-1 was accompanied by the development of systems used to maximize its performance (Holcomb and Hutt 1990). For vertical components, pressure and magnetic shielding was provided for vertical components using a glass bell jar and a permalloy shield, respectively.

Specially designed, so-called “warpless” baseplates prevented pressure-generated tilt noise from contaminating horizontal components. The electronics are housed in a separate enclosure which is not sealed and which can require regular replacement of desiccant to avoid anomalous response characteristics (Hutt and Ringler 2011).

The installation of an STS-1 is a delicate procedure; most of the innovations in the field of broadband seismometry since then have aimed at simplifying this procedure as well as reducing power and overall footprint.

To justify the performance of an STS-1, it is usually necessary to have an underground site in hard rock, something which is expensive to construct and unnecessary for earthquake engineering applications.

Shallow Broadband Vault

An idealized broadband vault design is shown in Fig. 16.

One practical procedure for constructing such a site is as follows. A hole is dug using a backhoe in which a large-diameter plastic tube is placed. A concrete slab is poured at the bottom to serve as a pier for the sensors to rest on. Thermal insulation is added around the sensors, and the digitizer is located in a separate compartment above the sensor. A cover is placed over the tube, and the earth which was dug out to make the hole is backfilled around the tube and tamped down up to the level of the lid. A layer of rigid foam insulation is placed across the lid before piling on the rest of the soil removed in digging the hole for the vault.

This same basic procedure can be tailored to the demands of temporary installations. The seismic vault designed for the “transportable array” of the USArray project (EarthScope 2013), for example, features most of the design elements shown in Fig. 16.

Thermal Insulation

Different thermal insulation components in a broadband vault serve different purposes. The sensor insulating cover serves as bulk insulation and as a breeze cover, and by restricting the airspace around the sensor, it stops convection around the sensor. An insulating layer laid on top of the seismic pier prevents convection-driven air currents from causing the pier to distort as they pass over its surface.

The thick layer of insulation over top of the vault serves to bring the vault closer in temperature to a deeper stratum of the ground. Otherwise,

a low thermal-resistance path from the vault to the surface would exist, and much of the benefit of burying a sensor in terms of thermal stability would be lost. Surface air temperature variation does not penetrate very deep into the ground; the effect of the insulation is to drive isotherms of temperature variation deeper into the ground, as shown approximately in Fig. 16. A rule of thumb for good-quality rigid Styrofoam insulation is that 2.5 cm of insulation provides the same thermal insulation as 30 cm of soil.

Pier Construction

The vault is drawn in Fig. 16 to accommodate a seismic pier which is significantly wider than a typical broadband seismometer plus its insulating cover. The reason for this is that some room must be left for the operator to stand beside the sensor and bend over it to orient the sensor to north, level it, and lock its feet. Vaults can be made significantly smaller if the seismometer is self-leveling, such as a ► [Downhole Seismometer](#), but of course the problem of sensor orientation still needs to be addressed.

The drain shown schematically in Fig. 16 will only be effective if the water table is at or below the depth of the seismic pier. Broadband vaults such as this one are prone to flooding; the surest remedy to this problem is to make use of a seismometer that is designed for submersion (e.g., ► [Downhole Seismometers](#)).

Because of the sensitivity of a broadband seismometer to tilt, the seismic pier should be physically decoupled from the vault wall. The soil at the surface will be constantly shifting due to wind and changes in water content or frost heave. Leaving a gap between the vault wall and the pier prevents such soil motion from being transmitted through the vault wall to the pier and producing measurable tilts.

The concrete for the pier should be made from 50 % Portland cement, 50 % sieved sand, and no aggregate. It should be vibrated to eliminate voids and allowed 24 h to harden before use. The pier must not be reinforced with steel; additional strength is not needed, and the different temperature coefficients would result in detectable tilts and cracking with temperature.

All classes of seismometer benefit from being sited on competent rock because levels of high-frequency (>1 Hz) noise of all kinds are lowest when the seismic wave velocities are highest. Broadband seismometers additionally benefit because hard rock sites are less susceptible to tilt, whether driven by pressure, cultural activity, or other phenomena, and horizontal site noise levels will be dominated by tilt at long periods (<0.1 Hz).

Broadband seismometers often must be sited, however, in areas where bedrock does not come near the surface. In these cases the vault design depicted in Fig. 16 is still useful; instead of pouring concrete, it may be convenient to place a stone block on tamped gravel and use that as a seismic pier.

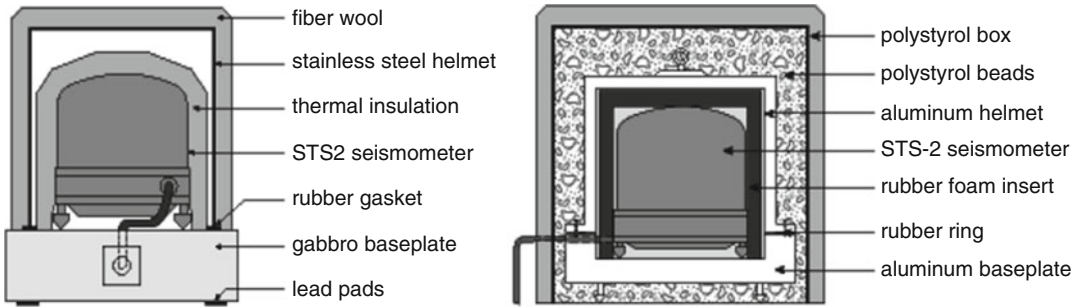
Thermal Shielding

Seismometer insulation is often implemented on a more or less ad hoc basis. Rigid Styrofoam insulation can be glued or taped together to make flat-faced shaped; large-diameter cardboard tubes lined with fiber wool can also be made.

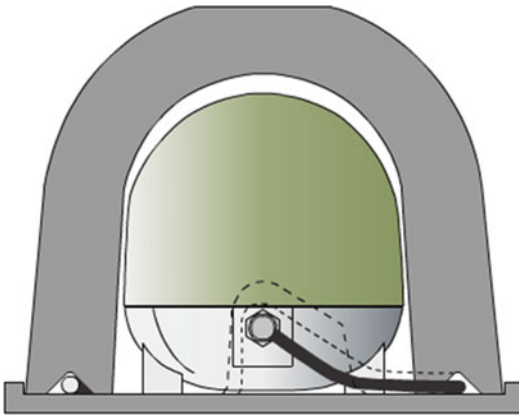
The “Stuttgart shielding” method used in the German Regional Seismic Network (GRSN) is a much more systematic approach, as shown in Fig. 17, reprinted with permission from Hanka (2002). It includes a thick gabbro baseplate, polished on one side, combined with a large stainless steel cooking pot, which provides pressure shielding to improve on the native pressure sensitivity of the STS-2. Both the outside of the pot and the outer surface of the sensor are wrapped with fiber wool to provide thermal insulation. Sometimes the whole assembly is further wrapped in a thermally reflective “space blanket” to provide additional thermal shielding.

A simplified setup was introduced for the GEOFON program. An aluminum enclosure including a “warpleless” baseplate provides the needed additional pressure shielding. Thermal shielding is provided by a foam rubber insert inside the aluminum enclosure and polystyrol beads outside the enclosure.

Some manufacturers provide thermal shields as accessories to seismometers. The thermal



Principles of Broadband Seismometry, Fig. 17 Thermal and pressure shielding for STS-2 (*left: GRSN, right: GEOFON*)



Principles of Broadband Seismometry, Fig. 18 Thermal shield for Trillium 240

shield for a Trillium 240 is shown in Fig. 18. This cover is made of rigid molded plastic filled with insulating foam. The cover is formfitting – without quite touching the sensor – so that it greatly restricts convective airflow around the sensor, without allowing forces to be transferred to the sensor body through the insulation. It includes a race for a turn of the sensor cable, which minimizes heat conduction through the cable. The shield includes a foam base gasket which raises the cover up off the pier and allows the cable to exit the shield, and provides a layer of insulation between the sensor and the surface of the seismic pier. Such shielding systems are rugged and easy to transport and provide excellent shielding performance and repeatability.

Other Installation Details

Many precautions which must be taken in installing a broadband seismometer should go without saying but must be stated anyway:

- Surface of pier must be clear of debris.
- Pier should be free of cracks, particularly beneath sensor.
- Insulating cover must be close to sensor but must not touch it.
- Adjustable feet must be firmly locked.
- Connectors must not touch insulating cover.
- Cable must be strain-relieved close to the sensor.
- Cable must not touch any other structure between connector and strain relief.

Cabling is an underrated source of excess long-period horizontal noise in vault-type installations. It is recommended to “strain-relieve” the cable to the pier, sometimes by placing a heavy weight on it close to the sensor. If this is not done, thermal expansion or other motion induced in the cable can be transferred to the seismometer, generating tilt. Sensitivity to cable-induced noise is particularly acute with stiff cables; cables designed for flexibility make the whole job much easier.

It is always important to follow the manufacturer’s instructions for sensor installation. Each sensor, for example, will provide different features for physical alignment of the sensor to north.

Downhole and Ocean Bottom

It is well known that site noise decreases with depth. This is the primary motivation for installations in deep, cased boreholes. Two newer techniques, which offer reduced installation costs and an even smaller surface footprint, are posthole and direct-burial installations (Nanometrics 2013). The design of such installations is however beyond the scope of this entry (see ► [Downhole Seismometers](#)).

The ocean bottom, in contrast, is an important environment for seismometer deployment not because of reduced levels of site noise, but because most of the earth is covered by ocean. It is similarly beyond the scope of this entry to treat the relevant installation techniques in detail, except to note some parallels with those used on land. Accurate leveling mitigates coupling of horizontal ground motion into the vertical output. Shallow burial can provide significant shielding from the effects of ocean-bottom currents. Colocated pressure sensors can be used to correct excess noise due to infragravity waves on the vertical. See ► [Ocean-Bottom Seismometers](#).

Summary

A seismometer is a kind of inertial sensor which can detect the smallest ground motions in some frequency band. Broadband seismometers are those which can detect motions which are as small as the background motion at a hypothetical quiet site (as represented by a model such as the NLNM) over a frequency band which extends both above and below the microseismic peak.

A broadband seismometer has much better performance than a passive seismometer of the same physical size. The use of displacement transducers and force feedback results in an instrument with better linearity, lower self-noise, and higher sensitivity at long periods.

A broadband seismometer must be installed in a carefully designed vault to ensure that spurious signals due to tilt driven by temperature and other environmental factors are minimized. A broadband seismometer capable of resolving the NLNM at 100 s should have a vertical

pressure sensitivity less than $\alpha_G = 0.04 \text{ nm/s}^2/\text{Pa}$, a horizontal pressure sensitivity less than $\alpha_T = 0.7 \text{ nm/s}^2/\text{Pa}$, and a magnetic sensitivity less than $\alpha_M = 0.7 \text{ nm/s}^2/\text{T}$.

The construction of an effective broadband seismic vault is described in some detail. Several different kinds of thermal insulation, serving different functions, are required for optimal performance at long periods.

References

- Aki K, Richards P (2002) Quantitative seismology, 2nd edn. University Science, Sausalito
- Berger J, Davis P, Ekström G (2004) Ambient Earth noise: A survey of the global seismographic network. *J Geophys Res* 209, B11307
- Bormann P (2002) Seismic signals and noise, chapter 4. In: Bormann P (ed) New manual of seismological observatory practice, vol 1. GeoForschungsZentrum, Potsdam
- Brune J, Oliver J (1959) The seismic noise of the Earth's surface. *Bull Seism Soc Am* 49:349–353
- Clinton F, Heaton T (2002) Potential advantages of a strong motion velocity meter over a strong motion accelerometer. *Seism Res Lett* 73(3):332–342
- EarthScope (2013) Transportable seismic network: imaging the Earth's interior. Retrieved 29 May 2014, from USArray: http://www.usarray.org/files/docs/pubs/TA_Host-a-Station_Imaging_0411-Final.pdf
- Evans JR, Followill F, Hutt CR, Kromer RP, Nigbor RL, Ringler AT, Steim JM, Wielandt E (2010) Method for calculating self-noise spectra and operating ranges for seismographic inertial sensors and recorders. *Seism Res Lett* 81(4):640–646
- Forbriger T, Widmer-Schmid R, Wielandt E, Hayman M, Ackerley N (2010) Magnetic field background variations can limit the resolution of seismic broad-band sensors. *Geophys J Int* 183(1):303–312
- Geotech Instruments, LLC (2001) Short-period seismometer model S-13 and GS-13. Retrieved 12 Nov 2013, from <http://www.geoinstr.com/ds-s13.pdf>
- Hanka W (2002) Parameters which influence the very long-period performance of a seismological station: examples from the GEOFON network, Section 7.4.4. In: Bormann P (ed) New manual of seismological observatory practice, vol 1. GeoForschungsZentrum, Potsdam, pp 64–74
- Holcomb LG, Hutt CR (1992) An evaluation of installation methods for STS-1 seismometers. Open-file report 92–302. US Geological Survey, Albuquerque
- Hutt CR, Ringler AT (2011) Some possible causes of and corrections for STS –1 response changes in the global seismographic network. *Seism Res Lett* 82(4):560–571

- Lee WHK, Evans JR, Huang B-S, Hutt CR, Lin C-J, Liu C-C, Nigbor RL (2012) Measuring rotational ground motions in seismological practice. In: Bormann P (ed) *New Manual of Seismological Observatory Practice 2* (NMSOP-2). Deutsches GeoForschungsZentrum GFZ, Potsdam, p. 1–27
- Longuet-Higgins MS (1950) A theory of the origin of microseisms. *Phil Trans R Soc A* 243(857):1–35
- McNamara DE, Buland RP (2004) Ambient noise levels in the continental United States. *Bull Seism Soc Am* 94(4):1517–1527
- Motchenbacher CD, Connelly JA (1993) *Low-noise electronic system design*. Wiley, New York
- Nanometrics Inc (2009) *Trillium 120P/PA seismometer user guide*. (15149R6). Nanometrics, Inc, Kanata
- Nanometrics Inc (2013) *Trillium posthole user guide*. (17217R5). Nanometrics, Inc, Kanata
- Peterson J (1993) Observations and modeling of seismic background noise. Open-file report 93–322. US Geological Survey, Albuquerque
- Phillips CL, Harbor RD (1991) *Feedback control systems*, 2nd edn. Prentice-Hall, Englewood Cliffs
- Ringler AT, Hutt CR (2010) Self-noise models of seismic instruments. *Seism Res Lett* 81(6):972–983
- Scintrex Limited (2007 June 21) *CG-5 Scintrex Autograv System operation manual*. Retrieved 7 May 2012, from Scintrex web site: <http://www.scintrexltd.com/documents/CG5.v2.manual.pdf>
- Sercel - France (2012) *Analog seismic sensors*. Retrieved 11 Dec 2013, from Sercel web site: <http://www.sercel.com/products/Lists/ProductSpecification/analog-seismic-sensors-specifications-Sercel-Seismometers.pdf>
- Sleeman R, van Wettum A, Trampert J (2006) Three-channel correlation analysis: a new technique to measure instrumental noise of digitizers and seismic sensors. *Bull Seism Soc Am* 96(1):258–271
- Strasser F, Bommer J (2009) Review: strong ground motions—have we seen the worst? *Bull Seism Soc Am* 99(5):2613–2637
- Trnkoczy A (2002) Factors affecting seismic site quality and site selection procedure, Section 7.1. In: Bormann P (ed) *New manual of seismological observatory practice*, vol 1. GeoForschungsZentrum, Potsdam, pp 1–14
- Webb S, Crawford W (2010) Shallow-water broadband OBS seismology. *Bull Seism Soc Am* 100(4):1770–1778
- Widmer-Schmid R, Kurrle D (2006) Evaluation of installation methods for Streckeisen STS-2 seismometers. Retrieved 25 Oct 2013, from <http://www.geophys.uni-stuttgart.de/~widmer/ge2.pdf>
- Wielandt E (2002) Seismic sensors and their calibration, chapter 5. In: Bormann P (ed) *New manual of seismological observatory practice*, vol 1. GeoForschungsZentrum, Potsdam
- Wielandt E, Streckeisen G (1982) The leaf-spring seismometer – design and performance. *Bull Seis Soc Am* 72(6):2349–2367
- Working Group on Instrumentation, Siting, Installation, and Site Metadata (2008) *Instrumentation guidelines*

- for the advanced national seismic system. Open file report 2008-1262. US Geological Survey, Reston
- Zürn W, Wielandt E (2007) On the minimum of vertical seismic noise near 3 mHz. *Geophys J Int* 168:647–658
- Zürn W, Exß J, Steffen H, Kroner C, Jahr T, Westerhaus M (2007) On reduction of long-period horizontal seismic noise using local barometric pressure. *Geophys J Int* 171(2):780–796

Probabilistic Seismic Hazard Models

Danielle Hutchings Mieler¹, Tatiana Goded² and Mark Stirling²

¹Earthquakes and Hazards Resilience Program, Association of Bay Area Governments, Oakland, CA, USA

²GNS Science, Lower Hutt, New Zealand

Synonyms

Deaggregation; Deterministic seismic hazard analysis; Ground motion prediction equations; Hazard curve; Hazard spectra; Probabilistic seismic hazard analysis, PSHA; Seismic hazard

Introduction

This entry provides an overview of probabilistic seismic hazard modeling, which has provided fundamental input to the engineering, planning, insurance sectors, and other fields for over 30 years. In essence probabilistic seismic hazard modeling uses the location, size, and occurrence rate of earthquakes to estimate the frequency or probability of damaging or potentially damaging earthquake motions that may occur at a site. By taking into account the frequency of earthquakes as well as the magnitude, the method captures the contribution to seismic hazard from all relevant earthquakes, from the frequent moderate earthquakes (magnitude $5 \leq M < 7$) to the infrequent large to great earthquakes ($M \geq 7$).

The following sections summarize the history and fundamental steps of probabilistic seismic hazard analysis (PSHA), on which probabilistic

seismic hazard modeling is based, providing example applications, discussing strengths and limitations, and describing current research and future needs.

Probabilistic Seismic Hazard Analysis: History, Method, and Outputs

History

Progression from Deterministic to Probabilistic Seismic Hazard Analysis

Prior to development of PSHA, seismic hazard models were primarily based on deterministic methods, in which a limited number of earthquake scenarios (usually just one or two) are selected as most relevant to the seismic hazard of the site. Based on the estimated magnitude of the scenario earthquake, distance from earthquake to site, and other considerations (e.g., slip type and site conditions), a ground motion is estimated for each scenario. While conceptually simple, deterministic methods do not consider the frequency of earthquakes. This can lead to overestimation of the hazard in the case where a source is associated with a very long recurrence interval. Conversely, ignoring sources that produce more frequent earthquakes of lesser size can lead to underestimation of the hazard. For example, a close moderate magnitude earthquake can produce higher ground motions than a more distant, large magnitude earthquake. Recognition of these deficiencies in deterministically based seismic hazard analysis led to the development of the methods of PSHA.

Development of the Probabilistic Hazard Analysis Framework

The framework for PSHA was first developed by Allin Cornell in the late 1960s (Cornell 1968). The motivation for developing PSHA came from the needs of engineers who must make fundamental trade-offs between the initial cost of designing and building a robust structure and the risk of economic loss due to natural disasters. It is important that engineers have the most accurate understanding of the hazard so they can

evaluate these trade-offs to achieve the optimal balance of cost, performance, and risk. Cornell's approach was to draw on the analogy of design winds or floods (e.g., 100-year flood) and provide engineers with a single ground motion value expressed in terms of return period. Cornell used the location, recurrence behavior, and predicted ground motion of earthquake sources to estimate the frequency or probability of exceeding a ground motion at a particular site or across a region. In this respect he saw PSHA as a way of providing a standardized method for the engineer or scientist to consider all possible earthquake scenarios and associated ground motions and their uncertainties to gain a clearer picture of the hazard for use in design.

Method

The basic four steps undertaken in a PSHA for a site are:

1. *Source definition.* Use geologic data and historical earthquake record to define the locations and dimensions of earthquake sources.
2. *Magnitude-frequency relationship.* Define the likely magnitudes of earthquakes that may be produced by each source and the associated frequencies of occurrence.
3. *Ground motion prediction.* Estimate the ground motion level that a source will produce at the site.
4. *Probability of exceedance calculation.* Quantify the frequency or probability at which a ground motion level is exceeded at the site, given all sources. Repeating this procedure for all sources and ground motion levels results in a hazard curve.

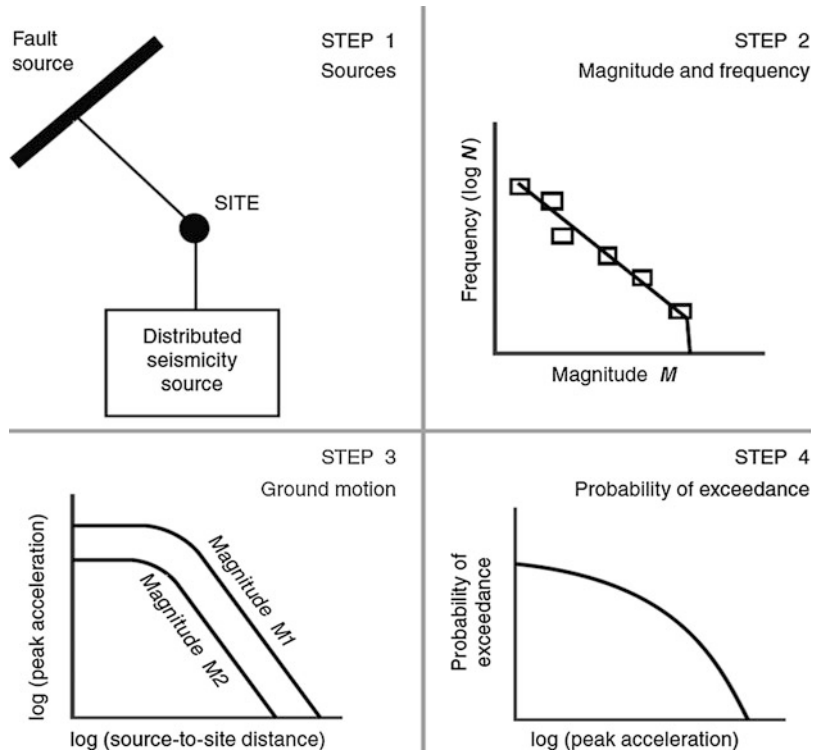
These steps are represented graphically in Fig. 1 and described in more detail in the following sections.

Source Definition

Seismicity catalogues, active fault data, and increasingly geodetic strain rates derived from global positioning system (GPS) data are used to define the locations and recurrence behavior of earthquake sources relevant to the site of

Probabilistic Seismic Hazard Models,

Fig. 1 The four steps of probabilistic seismic hazard analysis (After Cornell 1968)



interest. There are two seismic source classes: areal source zone, also referred to as a background or distributed seismicity source (Fig. 1) and fault source. Sources within 100–200 km from the site are usually considered in a PSHA, as sources beyond these distances generally produce ground motions that are too low to be of interest to engineering design.

Areal source zones are used to model the seismicity that is recorded in seismicity catalogues and, in particular, the seismicity that occurs away from known fault sources (e.g., Abrahamson 2011). The most important function of areal sources is therefore to represent sources that lack surface expression and have escaped detection by geologists. Areal source zones actually represent crustal volumes, as they are defined by a geographic area and an assumed thickness based on the maximum depth of seismicity (the seismogenic thickness). Fault sources are commonly modeled as planar features with length and depth that are based on surface fault traces and subsurface information.

Magnitude and Frequency

For fault sources, the dimensions and slip rate and/or paleoseismic (prehistoric) data defining the recurrence behavior of each source is used to develop a magnitude-frequency distribution of earthquakes. The distribution can range from a single magnitude and associated frequency (usually referred to as a characteristic earthquake; Schwartz and Coppersmith 1984) to a range of magnitudes. In the case of the characteristic earthquake, the magnitude is determined by the general equation

$$M_w = \log(A) + b, \quad (1)$$

where A is the fault area and b is a constant that is controlled by factors such as fault type, geology, regional tectonics, and stress drop. Various forms of the equation exist for different fault types and tectonic environments (e.g., Stirling et al. 2013), and there are also relationships that use the length of the fault to determine magnitude.

For areal sources, a range of earthquakes are usually considered for the source, utilizing the well-established Gutenberg-Richter relationship:

$$\text{Log}N(M) = a - bM, \quad (2)$$

where $N(M)$ is the activity rate, M is the magnitude, and a and b are empirical constants. The relationship was found to describe the seismicity of regions across the globe (Gutenberg and Richter 1944). The b -value is typically around 1, which means that there is roughly a tenfold decrease in the number of earthquakes between magnitude M and $M + 1$. The Gutenberg-Richter distributions are typically defined from seismicity catalogue data, and occasionally GPS and geologic data are used as well.

Ground Motion Model

Ground motion prediction equations (GMPEs), formerly called attenuation relations, are used to model the attenuation of ground motions with distance from the source. Various GMPEs have been developed for different tectonic regions (Abrahamson 2011). GMPEs typically account for magnitude, type of faulting, distance from the source, site response, and number of standard deviations. In PSHA the probability that the ground motion S_a exceeds the test value (z) is calculated for each magnitude (M) distance (r), and number of standard deviations (ε).

$$P(S_a > z | M, r, \varepsilon) \quad (3)$$

GMPEs have been developed for shallow crustal earthquakes in active tectonic regions (e.g., California), shallow crustal earthquakes in stable continental regions (e.g., eastern North America), and subduction zone earthquakes (e.g., Japan), rift environments, and others. The most recent suite of GMPEs has been produced by the PEER Next Generation Attenuation project (PEER-NGA; <http://peer.berkeley.edu/ngawest>). Over time, GMPEs have progressed from being developed for global application to being more focused on tectonic regimes and geographical regions (e.g., New Zealand; McVerry et al. 2006).

Probability of Exceedance Calculation

The final step of PSHA is to develop a hazard curve (Step 4 in Fig. 1), which gives the frequency or probability of exceedance for a suite of ground motion levels. The following equation is the fundamental equation of PSHA that produces the hazard curve:

$$v_i(S_a > z) = N_i(M_{\min}) \int_{r=0}^{\infty} \int_{M_{\min}}^{M_{\max}} \int_{\varepsilon_{\min}}^{\varepsilon_{\max}} f_{m_i}(M) f_{r_i}(r) f_{\varepsilon}(\varepsilon) P(S_a > z | M, r, \varepsilon) dr dM d\varepsilon \quad (4)$$

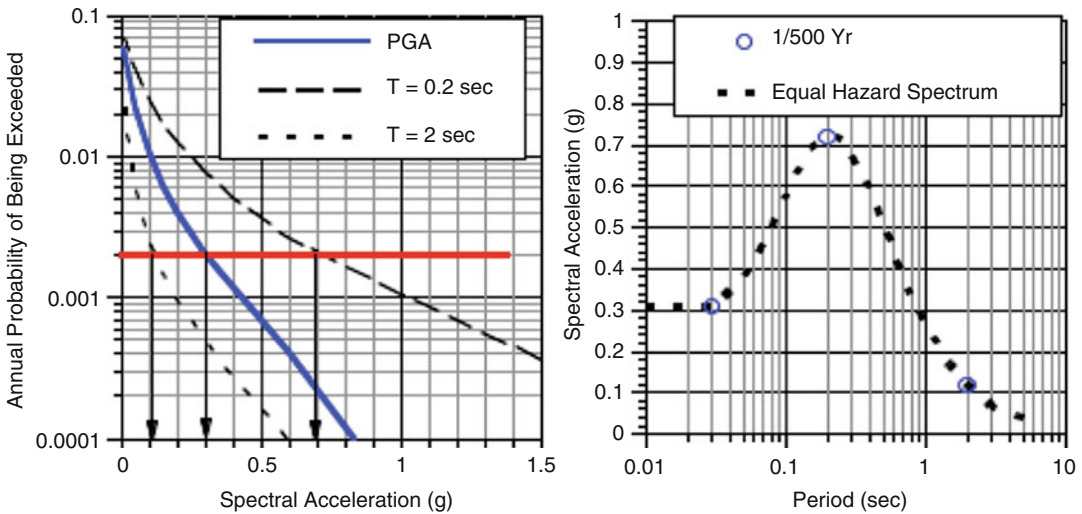
where $v_i(S_a > z)$ is the annual rate of events (v_i) on a single source that produce a ground motion parameter (S_a) exceeding a specified level (z) at the site of interest. The inverse of v is the return period in years. The hazard is therefore the integration over all possible magnitudes ($f_{m_i}(M)$), distances from the site to the source ($f_{r_i}(r)$), and standard deviations ($f_{\varepsilon}(\varepsilon)$). The ground motion for each individual scenario is calculated using a GMPE and the probability that the ground motion exceeds the test level is calculated (see “Ground Motion Model”).

PSHA is fundamentally a bookkeeping exercise. Instead of developing a small number of deterministic scenarios, a probabilistic seismic hazard model may develop many thousands of scenarios, each with relative contributions to the overall hazard at the site of interest. The rates of ground motions that exceed the specified level z are summed up over all sources to determine how often severe shaking occurs at a site, regardless of the source of the ground motion.

Outputs

Hazard Curve

The hazard curve (Step 4; Fig. 1) gives a suite of ground motion levels and their associated frequencies or return periods. A ground motion level is therefore read off the hazard curve at a user-specified annual frequency, and an important part of the probabilistic seismic hazard modeling process is selecting the appropriate



Probabilistic Seismic Hazard Models, Fig. 2 Procedure for developing uniform hazard spectra. In this example a return period of 475 years is used (Abrahamson 2011)

hazard level or return period. The return periods considered for engineering designs typically range from 475 years (often considered for ordinary buildings) to 2,500 years (special buildings, such as hospitals), which are, respectively, equivalent to 10 % and 2 % probability of exceedance in 50 years (e.g., Stirling et al. 2012). In contrast, nuclear facilities and major hydro-dam developments typically consider hazard estimates with 10,000-year return periods or longer. Hazard estimates for these three return periods typically show large quantifiable differences across regions like the USA, Europe, Japan, and New Zealand, reflecting the long-term tectonically driven differences in the expected future activity of earthquake sources across the regions.

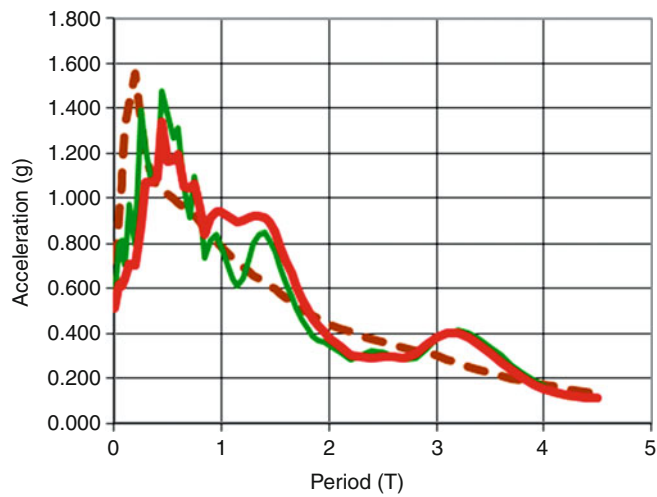
Uniform Hazard Spectra

Uniform hazard spectra, or equal hazard spectra (Fig. 2), can be rapidly developed from a probabilistic seismic hazard model to provide seismic design loadings for a range of return periods and spectral periods. The hazard curve (Step 4 in Fig. 1) is plotted for a suite of different spectral periods. At a chosen annual frequency or probability, the spectral acceleration (S_a) for each spectral period is measured from the hazard curve and plotted on a separate graph (Fig. 2). Spectral shapes differ for different sites due to local soil

conditions and the different mixes of earthquake magnitudes and distances surrounding the site of interest. The spectra therefore provide meaningful site-specific input to design loadings, including the selection of design earthquake scenarios and associated time histories (actual recordings of earthquakes used in engineering analysis). A response spectrum can also be plotted for a real event (Fig. 3) or a scenario event.

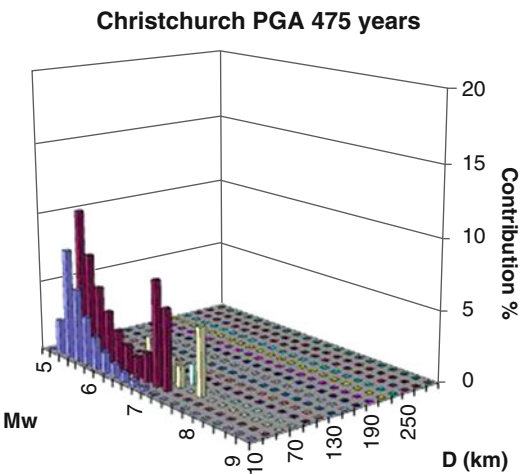
Deaggregation

Because the hazard curve is an ensemble of sources, magnitudes, and distances, it can be difficult to understand the relative contributions to the hazard at a site (Abrahamson 2011). The hazard curve can be broken down, or deaggregated, by magnitude and distance to identify the relative contribution of different earthquake scenarios to the hazard at a site. Similar scenarios are grouped together and the fractional contribution of different scenario groups to the hazard is computed and plotted on a deaggregation graph (Fig. 4). The results of the deaggregation will differ for different return periods and spectral periods. The deaggregation plots are often used to select realistic time histories for input to seismic loading analysis and to design scenario earthquakes for territorial authorities and others to plan for future earthquakes.



Probabilistic Seismic Hazard Models, Fig. 3 Examples of response spectra for Christchurch, New Zealand, for deep soil site conditions. The *solid lines* are spectral accelerations (SAs) recorded at selected strong motion stations in the city during the M6.2 2011 Christchurch

earthquake, and the *dashed line* is a response spectra derived from the New Zealand national seismic hazard model for a 10,000-year return period (Figure courtesy of Graeme McVerry, GNS Science)



Probabilistic Seismic Hazard Models, Fig. 4 Example of a deaggregation for the city of Christchurch derived from the New Zealand national seismic hazard model (Stirling et al. 2012). The deaggregation plot identifies two relevant classes of earthquakes that dominate the hazard of the city: earthquakes of M5-6.0 at distances of less than 10 km to the city and M6.0-7.5 at distances of 10–50 km. These classes of earthquakes encompass all of the major earthquakes of the Canterbury 2010–2012 earthquake sequence, despite the model being developed prior to initiation of the sequence

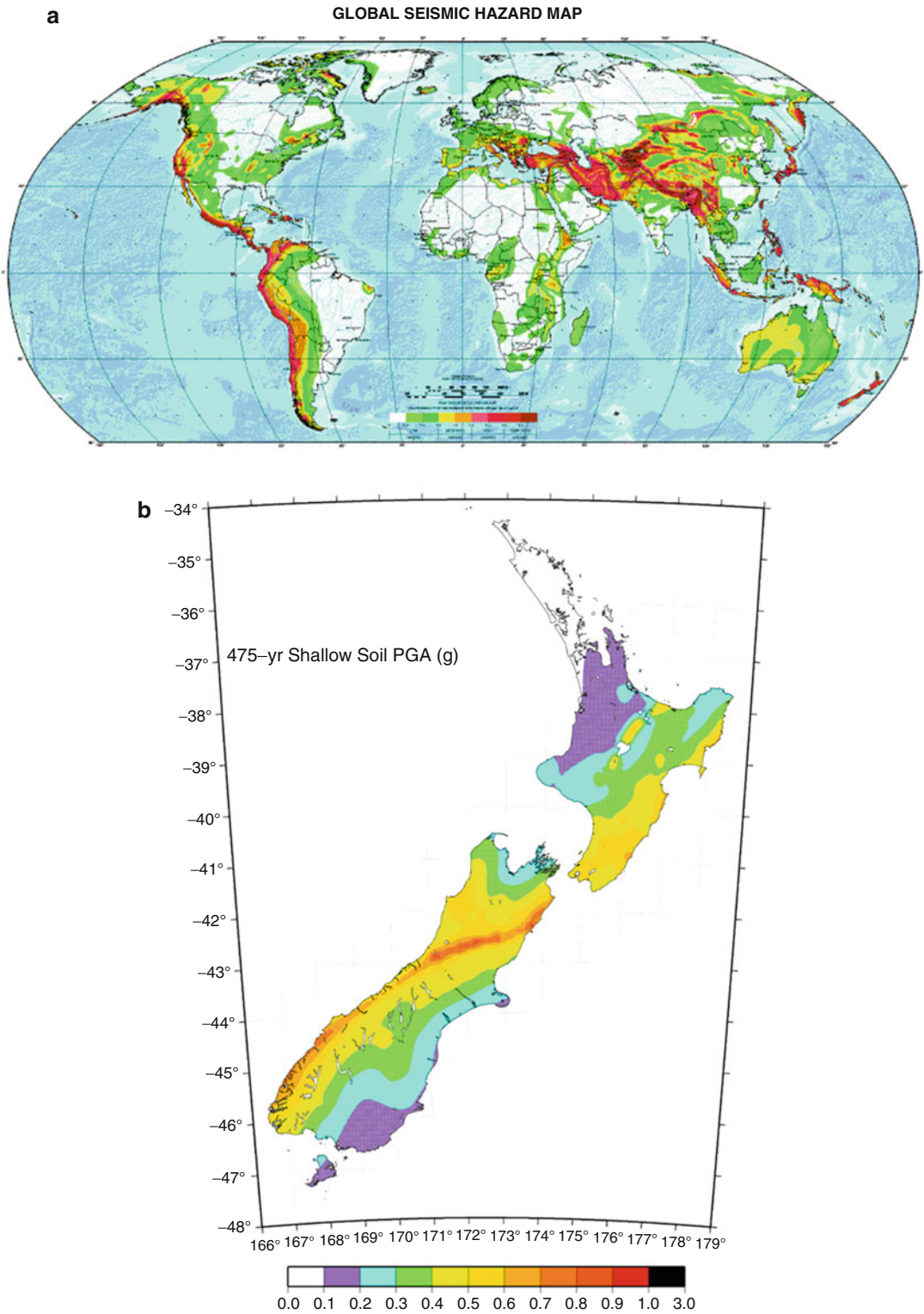
Example Applications

Regional, National, and Global Hazard Maps

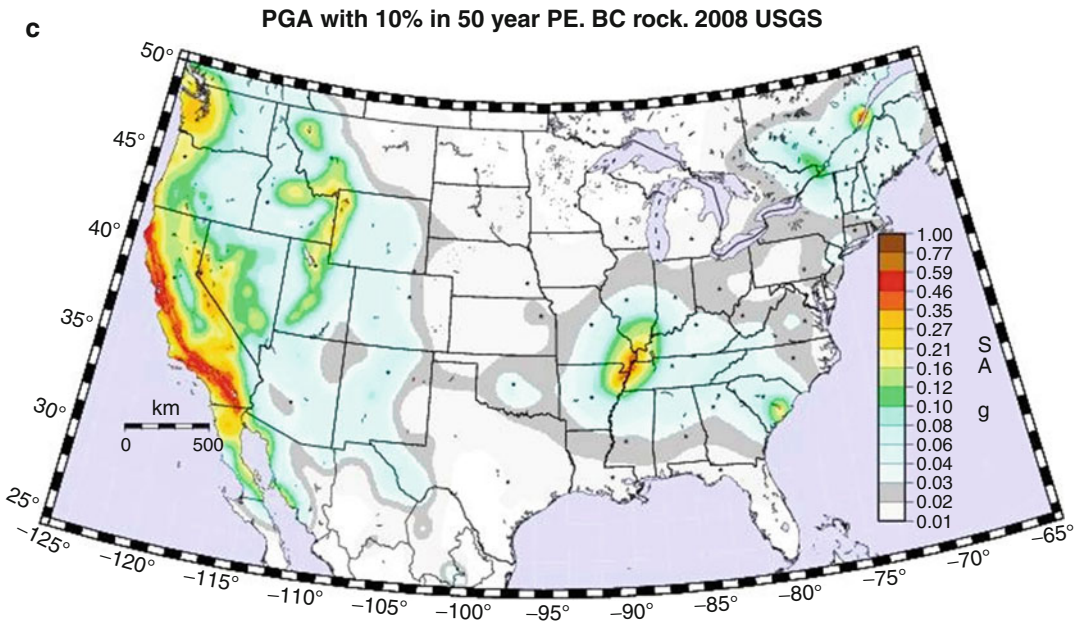
PSH maps of regions, countries, and the globe are routinely produced by the PSHA process (Fig. 5) at a grid of sites, and then the results are mapped for a given return period. The GSHAP (Global Seismic Hazard Analysis Program, the predecessor of the Global Earthquake Model GEM; globalquakemodel.org) map in Fig. 5a is an example of a global PSH map, which was developed to understand the global distribution of seismic hazard.

Two examples of national-scale PSH maps are from New Zealand (Fig. 5b) and the USA (Fig. 5c). These maps show high hazard along the main plate boundary areas and lower hazard away from the plate boundaries and provide very useful information for engineering and planning, including the development of design standards such as the New Zealand Loadings Standard NZS1170.5 (Standards New Zealand 2004).

At a regional scale, an example from the San Francisco Bay Area is shown in Fig. 6a, b. These



Probabilistic Seismic Hazard Models, Fig. 5 (continued)



Probabilistic Seismic Hazard Models, Fig. 5 Examples of global and national PSH maps: (a) Global seismic hazard analysis program (GSHAP) model (Giardini et al. 1999), (b) New Zealand national seismic hazard

model (Stirling et al. 2012), and (c) US national seismic hazard model (Petersen et al. 2008). Each map shows the peak ground accelerations (PGA) expected for a 475-year return period on soft rock sites

PSH maps were developed for use by the public and local governments to help drive hazard mitigation decision-making and policy with appropriate understanding of the likelihood of a significant earthquake event and the expected intensity. The PSHA map was deaggregated across a grid of the region to show the scenario with the highest contribution to the hazard in each location. The deaggregation map provides a guide for selecting the most appropriate earthquake scenarios for a given region or location. Scenario maps are available online for each of the major hazard sources identified in the deaggregation (quake.abag.ca.gov/earthquakes). A legend was developed that links MM (Modified Mercalli) intensity (the earthquake shaking intensity scale that is measured according to its effect on people, objects, and buildings) to expected nonstructural and structural damage of common dwelling types. In this way mitigation decisions by the public are risk informed and appropriate for the expected hazard (Brechwald and Mieler 2013).

Site-Specific Seismic Hazard in Wellington, New Zealand

New Zealand's capital city has long been a focus of site-specific PSHAs in New Zealand. The Wellington region is crossed by a number of major right lateral strike-slip faults and is underlain by the west-dipping subduction interface between the Pacific Plate and overriding Australian Plate (Hikurangi subduction zone) (Holden et al. 2013). In the short historic period of European settlement (ca. 160 years), the region has been shaken by large earthquakes, the largest being the M8.1-8.2 1855 Wairarapa earthquake. This earthquake also stands as the largest historical earthquake to have occurred in New Zealand since European colonization began in 1840. The earthquake was felt over a large part of the North Island and South Island of New Zealand and was severely damaging to settlements in the southern half of the North Island, particularly Wellington and Wanganui (Fig. 7).

Hazard curves for Wellington from the national seismic hazard model (Stirling

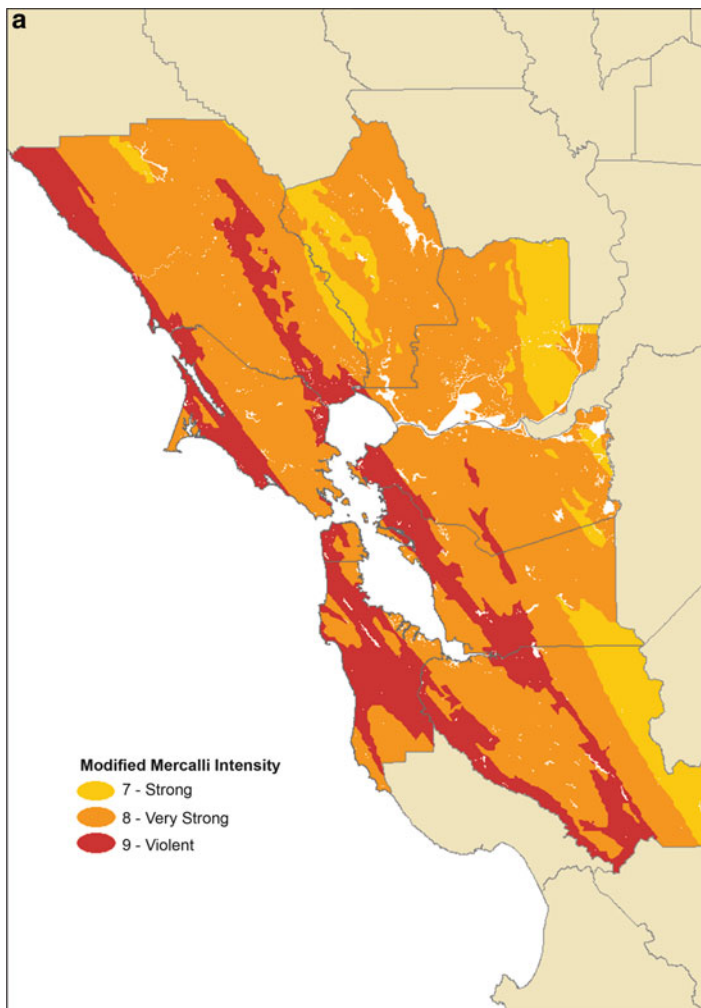
et al. 2012) are shown for several spectral periods in Fig. 8., The change in hazard as a function of return period is also illustrated by the two peak ground acceleration (PGA) hazard maps in Fig. 9a, b and by graphs of site-specific response spectra for Wellington city (Fig. 10). The highest overall spectrum is associated with the longest return period.

Wellington's 475-year PGA and S_a 1.0 s deaggregation are shown in Fig. 11a, b, respectively. The 475-year PGA hazard is dominantly controlled by fault sources. Peaks on the deaggregation plots show high contributions to overall hazard from the Wellington Fault

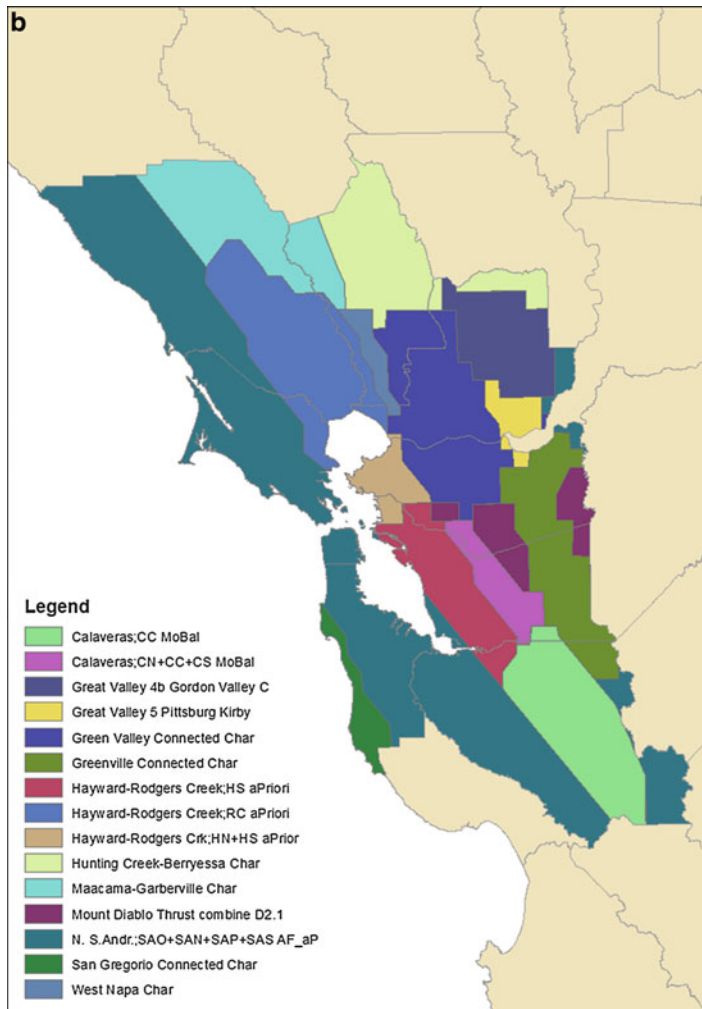
(M7.5 at less than 1 km; 20 % contribution), Ohariu Fault (M7.6 at 5 km; 20 % contribution), and Wairarapa Fault (M8.1 at 17 km; 13 % contribution). The 475-year S_a 1.0 s graph for Wellington shows an additional contribution to hazard from the local subduction zone (M8.1-9.0 at 23 km; 20 % contribution).

Limitations of Probabilistic Seismic Hazard Models

Recent, devastating earthquakes like the M9.0 2011 Tohoku, Japan, and M6.2 2011



Probabilistic Seismic Hazard Models, Fig. 6 (continued)

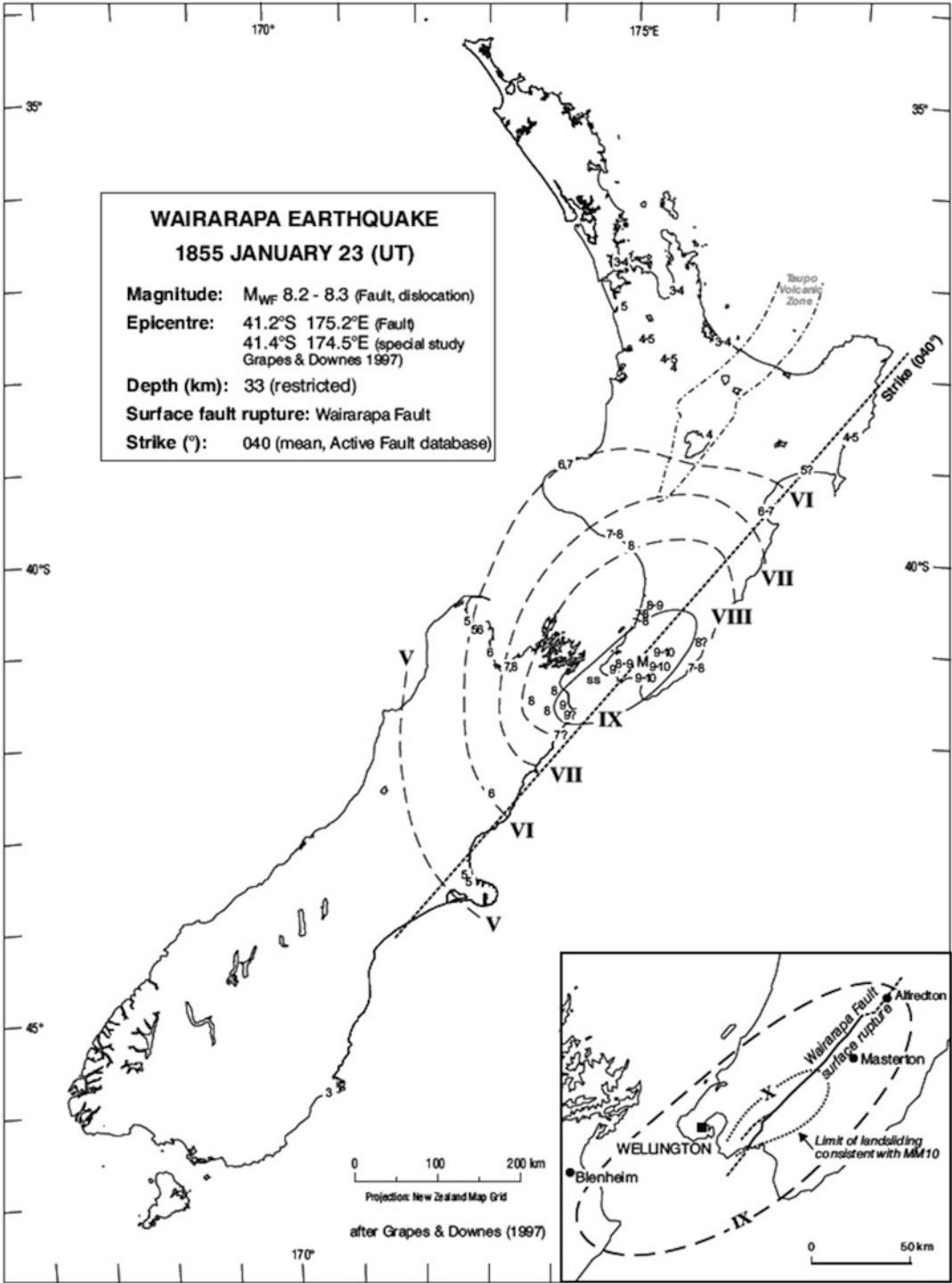


Probabilistic Seismic Hazard Models, Fig. 6 (a) PSH map for the San Francisco Bay Area, California, shown as an example of a regional-scale hazard map. MMI map obtained from 10 % probability in 50-year PGV values (equivalent to 475-year return period) (Brechtwald and

Mieler 2013). (b) Deaggregation of 10 % in 50-year hazard (a), showing fault scenario with most significant contribution to hazard by location (Brechtwald and Mieler 2013)

Christchurch earthquakes have resulted in considerable criticism of PSHA (e.g., Stein et al. 2011). The most frequent criticism is that PSHA did not provide any warning that these events were going to occur in 2011. While this is indeed the case, it is also correct to say that PSH models were never designed to provide short-term earthquake forecasts. The accelerated needs in Japan, New Zealand, and elsewhere to find short-term forecasting solutions are clearly beyond what standard PSHA can provide. Short-

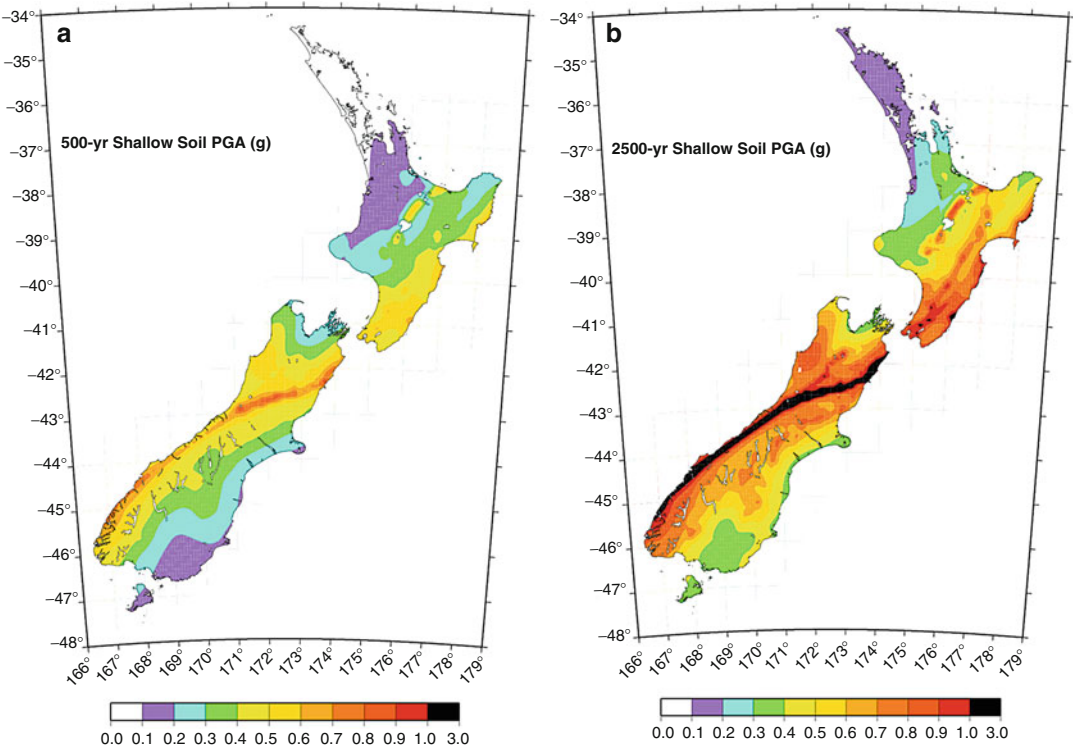
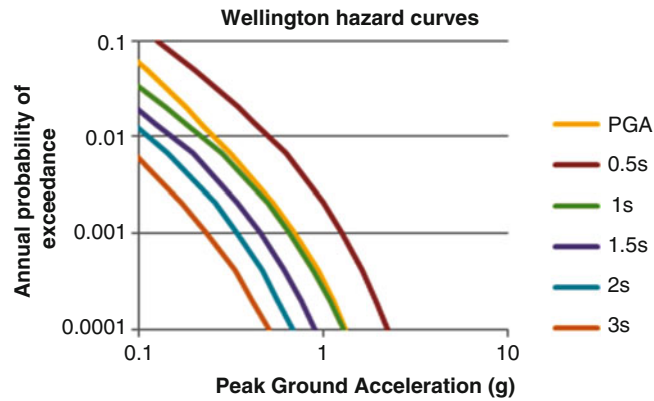
term forecasting requires construction of time-dependent or “time-varying” probability models. These models logically require two types of data: (1) detailed knowledge of the earthquake history and prehistory of well-studied faults, so the earthquake recurrence interval and elapsed time since the last earthquake faults can be determined, and (2) high-quality earthquake catalogues, which allow earthquake clustering behavior to be deciphered and modeled with time-varying rate or probability models (e.g., Rhoades et al. 2010).



Probabilistic Seismic Hazard Models, Fig. 7 Isoseismal map for the 1855 Wairarapa earthquake (maximum intensity MM9, possibly MM10; Downes and Dowrick 2009)

Probabilistic Seismic Hazard Models,

Fig. 8 Example of hazard curves as a result of a PSH analysis for the city of Wellington (data from the Stirling et al. 2012)



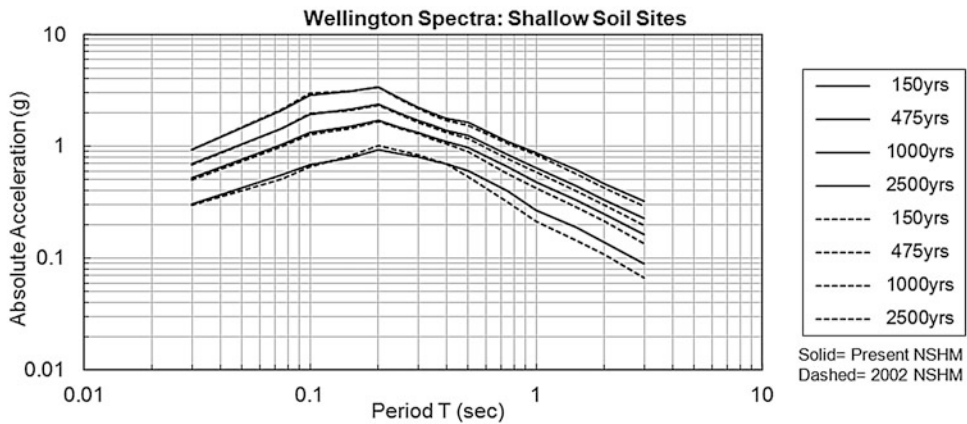
Probabilistic Seismic Hazard Models, Fig. 9 Seismic hazard maps for 475- and 2,500-year return periods (10 % and 2 % probability of exceedance in 50 years)

for class C (shallow soil) site conditions: (a) peak ground acceleration (PGA) for 475-year return period; (b) PGA for 2,500-year return period (Stirling et al. 2012)

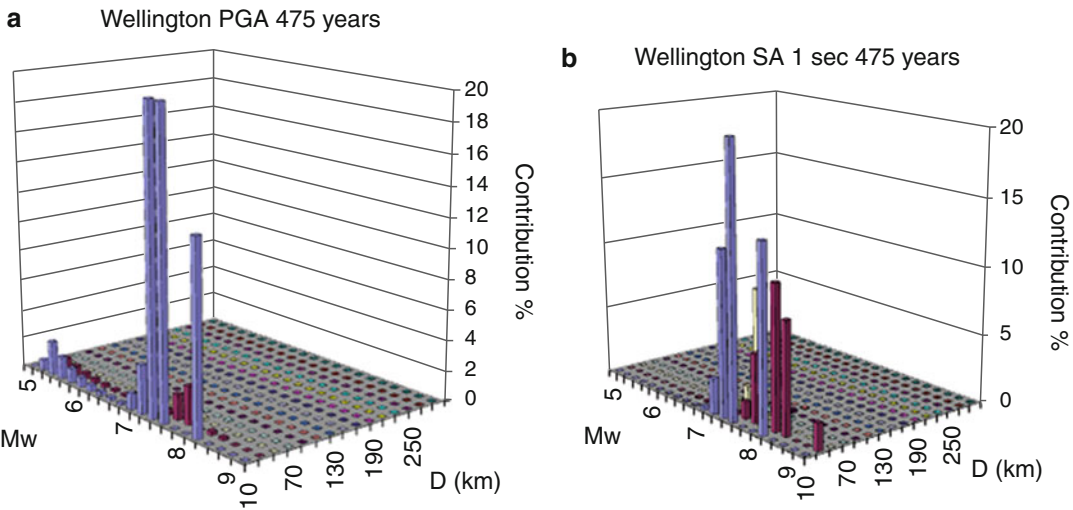
However, the resulting models generally differ greatly in terms of the resulting probabilities, and no one model is presently capable of providing a prospective short-term forecast of a large earthquake sequence that suddenly occurs in an area of low seismicity or seismic quiescence (as was the case for the Canterbury earthquake

sequence). The ability to provide actual short-term earthquake forecasts in areas of low seismicity still requires some significant advances in relevant scientific research and monitoring/detection.

Ground motion prediction (Step 3 of Fig. 1) is the source of large uncertainties for PSHA.



Probabilistic Seismic Hazard Models, Fig. 10 Response spectra for Wellington city for 150, 475, 1,000, and 2,500 years for class C (shallow soil) site conditions (Stirling et al. 2012). *Dashed lines* show the Stirling et al. (2002) spectra for comparison



Probabilistic Seismic Hazard Models, Fig. 11 Deaggregation graphs for the city of Wellington for (a) 475-year PGA for class C (shallow soil) site conditions and (b) 475-year SA 1 s to show the contribution of Hikurangi subduction interface sources at longer spectral periods (Stirling et al. 2012)

GMPEs are typically associated with large standard deviations (about 0.5 in natural log units of ground motion), which represents aleatory (random) uncertainty. Therefore, two earthquakes of the same magnitude and occurring the same distance away from a site can produce hugely different levels of shaking for unknown reason. These ground motion standard deviations do not appear to be reducing despite successive updates of ground motion prediction equations

(e.g., Watson-Lamprey 2013). In the Christchurch earthquake, PGAs of over 2 g were produced at some strong motion stations during the earthquake, and this was greatly in excess of what would normally be expected for an earthquake the size of the Christchurch earthquake (M6.2). Another issue associated with probabilistic seismic hazard models is that earthquakes often occur on sources that were previously unidentified. The causative fault of the main

shock of the Canterbury earthquake sequence (the M7.1 2010 Darfield earthquake) was unknown prior to the earthquake on 4 September 2010 due to the long recurrence interval and resulting lack of topographic expression of the fault in the relatively young Pleistocene outwash surface defining the Canterbury Plains. In the national seismic hazard model for New Zealand (Stirling et al. 2012), the earthquake was to some extent accounted for by the areal source model (i.e., consistent in terms of the long-term recurrence interval for Darfield-sized events), but these models do not inform where the earthquake sources and strongest shaking will occur and when the sources will produce the earthquakes. Again, this is an area of science that needs major advances in understanding and monitoring the changes that lead up to the occurrence of a large earthquake.

New Developments

Ground Motion Prediction

Many new GMPEs have been developed in the last decade, capturing the rich strong motion datasets produced by earthquakes in well-instrumented areas (e.g., Taiwan). The PEER-NGA project has involved some of the world's key GMPE developers producing a suite of GMPEs from the same quality-assured strong motion dataset. The models have incorporated more input parameters in an effort to improve ground motion prediction, particularly with respect to source geometry.

Improved Monitoring

Efforts to improve the recording of input data by seismic and GPS networks are of fundamental importance to PSHA. Seismic networks (e.g., Geonet in New Zealand, <http://geonet.org.nz>) are making large improvements to the detection threshold of earthquakes (the minimum magnitude for a complete record of earthquakes) and the ability to observe temporal and spatial changes in seismicity. GPS is being increasingly used to provide input to source models (e.g., distributed seismicity models and subduction

interface models). The generally short temporal coverage of GPS data is compensated for by a large spatial coverage, and as such it can provide a compliment to other source models. Satellite interferometry is another technique that is showing improvements in applicability and resolution over time, and these will allow greater ability to detect the coseismic deformation field (closely related to the source dimension) from earthquakes.

Detection and Characterization of Active Faults

Active fault datasets are the only PSHA input datasets that are able to extend the earthquake record back in time to prehistory. Great improvements in the ability to detect and characterize active faults for input to probabilistic seismic hazard models have been seen in the last 10 years. Fault mapping has improved significantly through accumulated experience and the availability of new tools (e.g., LIDAR). Greater ability to map the surface geometry of faults and distributions of displacement has led to improved characterization of fault sources in PSH models. The use of different disciplines and datasets together for fault characterization, particularly with respect to mapping fault ruptures in three dimensions, has yielded a great deal of understanding of rupture complexity and detail. Furthermore, increased age control on paleoearthquakes has made it possible to establish conditional probabilities of future ruptures and associated uncertainties.

Supercomputing to Consider All Possibilities

PSH models are increasingly drawing on diverse datasets and methods and utilizing high-end computing resources. The UCERF3 model (wgcep.org) incorporates hundreds of thousands of logic tree branches in its comprehensive source model, and access to supercomputers allows the complex model to be run through the four steps of PSHA (Fig. 1). Furthermore, physics-based seismic hazard modeling efforts such as CyberShake (SCEC.org) utilize supercomputers to run multiple realizations of earthquake scenarios from multiple sources, with shaking at the site computed

directly from source, path, and site effects for each earthquake. The millions of calculations required would not be possible without access to major computing resources and would not have been possible as recently as a decade ago. Plausible scenarios such as the linking of fault sources to produce extended ruptures and the range of uncertainties in magnitude-frequency statistics for the myriad sources are able to be handled without consideration of CPU demand. Already, exciting scientific results have emerged from the UCERF3 modeling efforts, such as the finding that seismicity on faults may not be well modeled by the Gutenberg-Richter relationship, as this produces a poor fit to the paleoseismic data in California (Ned Field pers. comm. 2013).

Future Needs

Correct Use of PSH Models

Many of the criticisms of PSHA in recent literature have been due to PSH models being used beyond their design capabilities. The models are not designed to be used as short-term forecasting tools, but more appropriately used to estimate seismic hazard for long return periods. The PSH models are also only as good as the input data and generalized methods of source parameterization, ground motion estimation, and probability estimation. However, as long as the associated uncertainties and limitations are fully expressed in the model commentary and appropriate cautionary advice is provided to the end users, the models will continue to provide valuable information. Use of deterministic models and appropriate parameterization of areal source models are two examples of solutions used to compensate for known or suspected deficiencies in PSH models.

Some recent efforts have focused on providing scientific forums to openly debate some of the criticisms leveled at PSHA and the associated input. The American Geophysical Union and Seismological Society of America have held "Earthquake Debates" sessions on several occasions over the last 5 years. Furthermore, the Powell Center for Analysis and Synthesis (<http://powellcenter.usgs.gov/>) has recently supported

a series of workshops that have brought together PSHA experts and critics from around the world, face to face, to address issues associated with maximum magnitude estimation, testability of PSHA, and development of global seismic source models (<http://www.nexus.globalquakemodel.org/powell-working-group/>). These meetings have been very productive, as people have been working together on common ground rather than talking past each other in the literature.

Earthquake Forecasting Without Earthquake Sequences

Clearly, a major advance in earthquake hazard estimation would be to achieve reliable short-term earthquake forecasts in areas of low seismicity, or in areas experiencing extended periods of seismic quiescence. To this end, future research needs to be focused on improving the ability to monitor and detect microseismicity and crustal deformation and on identifying reliable earthquake precursors. The Canterbury earthquake sequence has resulted in considerable advances being made in the modeling of short-term earthquake probabilities post-mainshock (i.e., aftershocks). These lessons are now being applied to the rebuilding of the city of Christchurch and will be applied to the rest of the country in the coming years. Clearly, these efforts need to be complimented with efforts to identify short-term precursors of future earthquakes in areas that are seismically quiet.

Reduction in Aleatory Uncertainty in Ground Motion Prediction

The aleatory uncertainty in ground motion estimation is very large and does not seem to have been reduced by the increasingly complex GMPEs available today. In other words there is still a very large range in the potential ground motions that could be produced at a single site due to earthquakes of the same magnitude, distance, and slip type. In contrast, the differences between GMPEs (epistemic uncertainties) do appear to have been reduced in recent years, at least within the NGA project. Clearly, effort needs to be focused on better understanding the source, path, and site effects that lead to the large

differences in ground motions observed in the strong motion databases.

Testability of Probabilistic Seismic Hazard Models

Finally, efforts need to be supported in the objective testing of PSH models, as to date PSH models have largely been developed in the absence of any form of verification. The Collaboratory for the Study of Earthquake Predictability (CSEP) has been developing testing strategies and methods for a wide variety of applications (SCEC.org), and collaborative work has also been focused on developing ground motion-based tests of the New Zealand and US national seismic hazard models. The Global Earthquake Model (GEM) Foundation (globalquakemodel.org) is including testing and evaluation as an integral part of the overall model development. The Yucca Mountain seismic hazard modeling project developed innovative approaches to consider all viable constraints on ground motions for long return periods for nuclear waste repository storage, prior to cancellation of the project in 2008 (Hanks et al. 2013). The need to verify hazard estimates for return periods of 10^4 – 10^6 years advanced the use of geomorphic criteria such as fragile geologic features (FGFs) to test the hazard estimates. The rationale is that these FGFs provide evidence for non-exceedance of ground motions for long return periods.

Complex PSHA on Normal Computers

If the future of PSHA is in the development of complex PSH models such as UCERF3, the reliance of these models on supercomputer resources will be a significant barrier to the widespread utility of these models and methods. Significant efforts in the future will therefore need to be focused on making these models usable on standard computers, or uptake will be extremely limited for everyday end-user PSHA applications.

Summary

This entry provides an overview of probabilistic seismic hazard (PSH) models which have provided fundamental input to the engineering,

planning, insurance sectors, and other fields for over 30 years. In essence PSH models use the location, size, and occurrence rate of earthquakes to estimate the frequency or probability of damaging or potentially damaging earthquake motions that may occur at a site. By taking into account the frequency of earthquakes as well as the magnitude, the models capture the contribution to seismic hazard from all relevant earthquakes, from the frequent moderate earthquakes (magnitude $5 \leq M < 7$) to the infrequent large to great earthquakes ($M \geq 7$). The entry summarizes the history and fundamental steps of PSHA, provides example applications, discusses strengths and limitations of PSHA, and describes current research and future needs.

Cross-References

- ▶ [Conditional Spectra](#)
- ▶ [Earthquake Recurrence](#)
- ▶ [Earthquake Recurrence Law and the Weibull Distribution](#)
- ▶ [Earthquake Response Spectra and Design Spectra](#)
- ▶ [Earthquake Return Period and Its Incorporation into Seismic Actions](#)
- ▶ [Engineering Characterization of Earthquake Ground Motions](#)
- ▶ [Physics-Based Ground-Motion Simulation](#)
- ▶ [Probability Seismic Hazard Mapping of Taiwan](#)
- ▶ [Review and Implications of Inputs for Seismic Hazard Analysis](#)
- ▶ [Seismic Actions Due to Near-Fault Ground Motion](#)
- ▶ [Seismic Risk Assessment, Cascading Effects](#)
- ▶ [Selection of Ground Motions for Response History Analysis](#)
- ▶ [Site Response for Seismic Hazard Assessment](#)
- ▶ [Spatial Variability of Ground Motion: Seismic Analysis](#)
- ▶ [Spectral Finite Element Approach for Structural Dynamics](#)
- ▶ [Time History Seismic Analysis](#)

References

- Abrahamson N (2011) CE 276: seismic hazard analysis and design ground motions [course notes]. Department of Civil Engineering, University of California, Berkeley
- Brechwald D, Mieler DH (2013) Sub-regional earthquake hazards and earthquake mapping update. Association of Bay Area Governments, Oakland. <http://quake.abag.ca.gov/projects/earthquake-mapping-update/>
- Cornell CA (1968) Engineering seismic risk analysis. *Bull Seismol Soc Am* 58(6):1583–1606
- Downes GL, Dowrick DJ (2009) Atlas of isoseismal maps of New Zealand earthquakes. GNS science monograph 25. Lower Hutt, New Zealand: GNS Science (Te Pū Ao)
- Giardini D, Grunthal G, Shedlock K, Zeng P (1999) The GSHAP global earthquake hazard map. *Annali Di Geofisica* 42:1225–1230
- Gutenberg, Richter (1944) Frequency of earthquakes in California. *Bull Seismol Soc Am* 34:185–188
- Hanks TC, Abrahamson NA, Baker JW, Boore DM, Board M, Brune JN, Cornell CA, Whitney JW (2013) Extreme ground motions and Yucca Mountain: U.S. Geological survey open-file report 2013–1245, 105p. <http://dx.doi.org/10.3133/ofr20131245>
- Holden C, Zhao J, Stirling M (2013) Ground motion modelling of a large subduction interface earthquake in Wellington, New Zealand. In: Proceedings of the New Zealand society of earthquake engineering, annual meeting 2013, Wellington. Paper 7, 8pp
- McVerry GH, Zhao JX, Abrahamson NA, Somerville PG (2006) New Zealand acceleration response spectrum attenuation relations for crustal and subduction zone earthquakes. *Bull N Z Soc Earthquake Eng* 39(1):1–58
- Petersen MD, Frankel AD, Harmsen SC, Mueller CS, Haller KM, Wheeler RL, Wesson RL, Zeng Y, Boyd OS, Perkins DM, Luco N, Field EH, Wills CJ, Rukstales KS (2008) Documentation for the 2008 update of the United States national seismic hazard maps: U.S. Geological survey open-file report 2008–1128, 61 pp
- Rhoades DA, Van Dissen RJ, Langridge RM, Little TA, Ninis D, Smith EGC, Robinson R (2010) Re-evaluation of the conditional probability of rupture of the Wellington–Hutt valley segment of the Wellington fault. *Bull N Z Natl Soc Earthquake Eng* 44:77–86
- Schwartz DP, Coppersmith KJ (1984) Fault behavior and characteristic earthquakes: examples from the Wasatch and San Andreas Fault Zones. *J Geophys Res Solid Earth* 89(B7):5681–5698
- Standards New Zealand (2004) Structural design actions–Part 5: earthquake actions – New Zealand, New Zealand Standard NZS 1170.5, Department Building and Housing, Wellington
- Stein S, Geller R, Liu M (2011) Bad assumptions or bad luck: why earthquake hazard maps need objective testing. *Seismol Res Lett* 82(5):623–626
- Stirling MW, McVerry GH, Berryman KR (2002) A new seismic hazard model for New Zealand. *Bull Seismol Soc Am* 92:1878–1903
- Stirling MW, McVerry GH, Gerstenberger M, Litchfield NJ, Van Dissen R, Berryman KR, Langridge RM, Nicol A, Smith WD, Villamor P, Wallace L, Clark K, Reyners M, Barnes P, Lamarche G, Nodder S, Pettinga J, Bradley B, Rhoades D, Jacobs K (2012) National seismic hazard model for New Zealand: 2010 update. *Bull Seismol Soc Am* 102(4):1514–1542
- Stirling MW, Gode T, Berryman K, Litchfield N (2013) Selection of earthquake scaling relationships for seismic-hazard analysis. *Bull Seismol Soc Am* 103(6):2993–3011
- Watson-Lamprey J (2013) Incorporating the effect of directivity in the intra-event standard deviation of the NGA West 2 ground motion prediction equations. In: Abstracts for annual meeting of the seismological society of America. Seismological Research Letters

Probability Density Evolution Method in Stochastic Dynamics

Jie Li and Jianbing Chen

School of Civil Engineering & State Key Laboratory for Disaster Reduction in Civil Engineering, Tongji University, Shanghai, China

Synonyms

Generalized density evolution equation; Global reliability; Nonlinear stochastic dynamics; PDEM; Stochastic harmonic function; Stochastic response

Introduction

The seismic ground motions are well recognized to be stochastic processes for over 60 years. Under such extreme loadings with large uncertainty, it is almost impossible for engineering structures subjected to earthquakes to avoid nonlinear behaviors during their service life (Roberts and Spanos 1990). Simultaneously, large uncertainties also exist in the models of structures, including the system mechanics parameters and such factors as non-structure effects, boundary conditions, geometric sizes, etc. For instance, the strength of concrete usually has a coefficient of variation (COV)

from 10 % to 23 %, whereas even the strength of steel, which is thought to be much more homogeneous, has a COV ranging from 7 % to 9 %. This may induce the fluctuation of static response of structures in the same order of magnitude of COV of the source uncertain parameters. However, the fluctuation in dynamic response may be enlarged greatly. In addition, the coupling of randomness in the system parameters and excitations will make the fluctuation of response much greater than that of response when the randomness is involved only in excitations (Chen and Li 2010). Therefore, to consider the randomness involved in both system parameters and excitations together with their coupling with the development of nonlinearity in structural behaviors is of paramount importance.

Engineering stochastic dynamics has been developed for over half a century, exhibited as two branches, i.e., the random vibration theory and stochastic structural analysis (stochastic finite element method). For linear structures, in both branches the probabilistic information of the second-order statistics could be well obtained (Ghanem and Spanos 1991; Li 1996). In random vibration when the excitations are white noise processes, the joint probability density function (PDF) is governed by the FPK equation, and the solution is well known as a joint Gaussian distribution. But in stochastic structural analysis, where the uncertainty of system parameters is dealt with, no analogous partial differential equation exists in the traditional theory (Li 1996; Ghanem and Spanos 1991). Moreover, in both random-parameter problems and random-excitation problems, huge difficulty exists in dealing with multi-degree-of-freedom (MDOF) nonlinear structures (Goller et al. 2013; Zhu 2006). The coupling of nonlinearity and randomness in MDOF systems is almost unbreakable. This is the common crucial difficulty in both branches.

In the past decade, a family of probability density evolution method (PDEM) was developed. In this method, the thought of physical stochastic systems was advocated (Li and Chen 2009). The principle of preservation of probability was adopted as a unified basis and revisited from the state space description and the random event

description (Li and Chen 2008). A decoupled generalized density evolution equation was derived and solved together with the embedded physical equations. By this the instantaneous PDF and reliability of MDOF nonlinear structures with randomness involved in both system parameters and external loadings could be captured (Li and Chen 2003, 2005; Chen and Li 2005; Li et al. 2012a; Goller et al. 2013). This entry will outline its theoretical basis and numerical algorithms and particularly put emphasis on earthquake engineering applications.

Basic Principles of the Probability Density Evolution Method

Without loss of generality, the equation of motion of an n -DOF structure subjected to seismic ground motion is

$$\mathbf{M}(\boldsymbol{\eta})\ddot{\mathbf{X}} + \mathbf{C}(\boldsymbol{\eta})\dot{\mathbf{X}} + \mathbf{f}(\boldsymbol{\eta}, \mathbf{X}) = -\mathbf{M}(\boldsymbol{\eta})\mathbf{I}a_R(\boldsymbol{\xi}, t) \quad (1)$$

where $\ddot{\mathbf{X}}, \dot{\mathbf{X}}, \mathbf{X}$ are the n -dimensional vectors of acceleration, velocity, and displacement relative to ground, respectively; $\mathbf{M}$ and $\mathbf{C}$ are the n by n mass and damping matrices, respectively; $\mathbf{f}$ is the linear or nonlinear restoring forces; $\mathbf{I}$ is the n -dimensional column vector with all components being 1, $a_R(\boldsymbol{\xi}, t)$ is the ground motion accelerogram, which could be specified by the models to be outlined in the later section; $\boldsymbol{\eta} = (\eta_1, \dots, \eta_{s_1})$ are the basic random parameters in the structural system properties; and $\boldsymbol{\xi} = (\xi_1, \dots, \xi_{s_2})$ are the basic random parameters in the excitation. For notational convenience, let $\boldsymbol{\Theta}(\varpi) = [\Theta_1(\varpi), \dots, \Theta_s(\varpi)] = (\boldsymbol{\eta}, \boldsymbol{\xi}) = (\eta_1, \dots, \eta_{s_1}, \xi_1, \dots, \xi_{s_2})$, where $s = s_1 + s_2$.

If the state vector $\mathbf{Y} = (\mathbf{X}^T, \dot{\mathbf{X}}^T)^T = (Y_1, \dots, Y_{2n})^T$ is introduced, Eq. 1 could be rewritten as a stochastic state equation:

$$\dot{\mathbf{Y}} = \mathbf{A}(\mathbf{Y}, \boldsymbol{\Theta}(\varpi), t) \quad (2)$$

where $\mathbf{A} = (A_1, \dots, A_{2n})^T = \left(\dot{\mathbf{X}}^T, [-\mathbf{M}^{-1}\mathbf{C}(\boldsymbol{\eta})\dot{\mathbf{X}} - \mathbf{M}^{-1}\mathbf{f}(\boldsymbol{\eta}, \mathbf{X}) - \mathbf{I}a_R(\boldsymbol{\xi}, t)]^T \right)^T$. The initial condition is given by $\mathbf{Y}(0) = \mathbf{Y}_0$.

Let us consider the PDF of $Y_\ell(t)$, the ℓ -th component of $\mathbf{Y}(t)$. Denote the PDF by $p_{Y_\ell}(y, t)$. To understand the evolution of $p_{Y_\ell}(y, t)$, consider the change of probability in an arbitrary interval $[y_L, y_R]$ during the time interval $[t, t + \Delta t]$:

$$\begin{aligned} \Delta P_D &= \int_{y_L}^{y_R} p_{Y_\ell}(y, t + \Delta t) dy - \int_{y_L}^{y_R} p_{Y_\ell}(y, t) dy \\ &= \left(\int_{y_L}^{y_R} \frac{\partial p_{Y_\ell}(y, t)}{\partial t} dy \right) \Delta t + o(\Delta t) \end{aligned} \quad (3)$$

This portion of change of probability is due to the probability transiting through the boundaries y_L and y_R . If during per unit time the probability passing a point z could be denoted by $J(z, t)$, then the change of probability in $[y_L, y_R]$ during $[t, t + \Delta t]$ is given by

$$\begin{aligned} \Delta P_B &= -J(y_R, t)\Delta t + J(y_L, t)\Delta t + o(\Delta t) \\ &= - \left(\int_{y_L}^{y_R} \frac{\partial J(y, t)}{\partial y} dy \right) \Delta t + o(\Delta t) \end{aligned} \quad (4)$$

According to the principle of preservation of probability, $\Delta P_D = \Delta P_B$. Substituting Eqs. 3 and 4 in it and considering the arbitrariness of $[y_L, y_R]$ yield

$$\frac{\partial p_{Y_\ell}(y, t)}{\partial t} = - \frac{\partial J(y, t)}{\partial y} \quad (5)$$

This is nothing but the continuity equation, in which $J(y, t)$ is the flux of probability, i.e., the probability passing a point during per unit time. According to this physical meaning, the flux of probability is $J(y, t) = \lim_{\Delta t \rightarrow 0} \frac{\Delta P_y}{\Delta t}$, where ΔP_y is the probability passing the point y during Δt , i.e., $\Delta P_y = p_{Y_\ell}(y, t)\Delta y + o(\Delta y)$, in which Δy is the displacement of the particle during Δt . Note that both $Y_\ell(t)$ and $\dot{Y}_\ell(t)$ depend on $\Theta(\omega)$. There is

$$\begin{aligned} \Delta P_y &= p_{Y_\ell}(y, t)\Delta y + o(\Delta y) \\ &= \int_{\Omega_\Theta} \left[\Delta Y_\ell(\boldsymbol{\theta}) p_{Y_\ell|\Theta}(y, t|\boldsymbol{\theta}) \right] p_\Theta(\boldsymbol{\theta}) d\boldsymbol{\theta} + o(\Delta y) \\ &= \left(\int_{\Omega_\Theta} [\dot{Y}_\ell(\boldsymbol{\theta}, t) p_{Y_\ell|\Theta}(y, \boldsymbol{\theta}, t)] d\boldsymbol{\theta} \right) \Delta t + o(\Delta t) \end{aligned} \quad (6)$$

where $p_{Y_\ell|\Theta}(y, t|\boldsymbol{\theta})$ is the conditional PDF and $p_{Y_\ell|\Theta}(y, \boldsymbol{\theta}, t)$ is the joint PDF of $(Y_\ell(t), \Theta)$. Clearly,

$$p_{Y_\ell}(y, t) = \int_{\Omega_\Theta} p_{Y_\ell|\Theta}(y, \boldsymbol{\theta}, t) d\boldsymbol{\theta} \quad (7)$$

According to Eq. 6,

$$J(y, t) = \lim_{\Delta t \rightarrow 0} \frac{\Delta P_y}{\Delta t} = \int_{\Omega_\Theta} [\dot{Y}_\ell(\boldsymbol{\theta}, t) p_{Y_\ell|\Theta}(y, \boldsymbol{\theta}, t)] d\boldsymbol{\theta} \quad (8)$$

Substituting Eqs. 7 and 8 in Eq. 5 yields

$$\begin{aligned} \frac{\partial}{\partial t} \int_{\Omega_\Theta} p_{Y_\ell|\Theta}(y, \boldsymbol{\theta}, t) d\boldsymbol{\theta} &= \\ - \frac{\partial}{\partial y} \int_{\Omega_\Theta} [\dot{Y}_\ell(\boldsymbol{\theta}, t) p_{Y_\ell|\Theta}(y, \boldsymbol{\theta}, t)] d\boldsymbol{\theta} \end{aligned} \quad (9)$$

which should hold for any arbitrary domain Ω_Θ , and therefore the integrand should be identical, i.e.,

$$\frac{\partial p_{Y_\ell|\Theta}(y, \boldsymbol{\theta}, t)}{\partial t} = -\dot{Y}_\ell(\boldsymbol{\theta}, t) \frac{\partial p_{Y_\ell|\Theta}(y, \boldsymbol{\theta}, t)}{\partial y} \quad (10)$$

This is the generalized density evolution equation (GDEE). More rigorous derivation could be found in Li and Chen (2008, 2009) and Chen and Li (2009).

Remark 1 The most important advantage of GDEE compared to the traditional equations, e.g., the FPK equation, is that the dimension is totally untied from the original dynamical system. Actually, although Eq. 10 is in one dimension, from the above heuristic deduction process, it is clear that if any arbitrary number of

components are of concern, a corresponding GDEE in appropriate dimensions exists.

Remark 2 In the PDEM there is no need for the stochastic process of concern to be Markovian. Actually, in most cases the process is not Markov. For instance, in engineering practice a complex structure may be modeled by the finite element method where nonlinear constitutive relationship of the material, say, the stochastic damage constitutive law for concrete, is embedded. In this case, usually quite a few of internal variables, say, the damage variables, are involved, and thus the response processes are not Markovian (Li et al. 2014).

Remark 3 In the GDEE, the randomness involved in the system parameters and external loadings is treated simultaneously in a unified way. Traditionally, in random vibration theory and stochastic structural analysis (stochastic finite element method), the methodologies are quite distinct. But as mentioned before in both branches, huge difficulty is encountered for nonlinear MDOF systems.

Stochastic Harmonic Function Representation of Seismic Ground Motions

Random Function Description of Stochastic Processes

Mathematically, a stochastic process $X(t)$ could be regarded as a family of random variables on a parametric set, say $t \in [0, T]$. For a continuous-parameter process, to characterize the probabilistic information of the stochastic process, the finite-dimensional distributions of PDFs, i.e., $p(x_1, t_1)$, $p(x_1, t_1; x_2, t_2)$, $p(x_1, t_1; x_2, t_2; \dots; x_n, t_n)$, $\dots$, should be specified. By doing so a stochastic process is regarded as a random function of time, but the dependence of X on t is specified not by an explicit expression of t but in an indirect way of specifying the complete cross-probabilistic information of X at all possible different time instants. This description is complete in mathematics. However, two deficiencies exist

for this description at least from the point of view of practical applications (Li et al. 2012b): (i) even if the finite-dimensional distributions are known, how the process X depends on t is still not clear in a sense of physics, but this might be very important for a practical physical problem, and (ii) to capture the high-dimensional distributions of general type other than joint normal distribution by observed information, huge data and prohibitively computational efforts are needed, which is usually impractical either due to lack of data or due to computational difficulty induced by the so-called curse of dimension.

A conceptually more accessible way to a stochastic process is the involvement of an abstract argument representing a sample point, and thus a stochastic process is denoted by $X(\varpi, t)$, where ϖ denotes a sample point in the sample space. By this it is very clear that X is a multi-argument function of ϖ and t . Because X is a function of ϖ , it is “stochastic” in nature. Simultaneously because X is also a function of t , it is a “process.” Thus $X(\varpi, t)$ is a stochastic process. For practical applications the deficiency of this description is that ϖ is a mathematically abstract point in the sample space and its relation to the physical entities is still not exposed. A further step could be made by introducing the embedded basic random variables in the physical problems under consideration, denoted by $\Theta(\varpi) = [\Theta_1(\varpi), \Theta_2(\varpi), \dots, \Theta_s(\varpi)]$ for convenience, and thus a stochastic process could be represented by $X(\varpi, t) = g(\Theta(\varpi), t)$, where $g(\cdot)$ is an explicit function of $\Theta(\varpi)$ and t . The form of $g(\cdot)$ could be determined by the embedded physical mechanism or by mathematical decomposition if phenomenological statistical models are involved, as shown in the following subsections.

Dynamic excitations encountered in engineering, e.g., earthquakes, wind, and waves, are originated with the embedded physical mechanism (Li et al. 2012b, Lin et al. 2012), although the knowledge on these physical phenomena is still at different levels, some even in very preliminary stage. One of the obstacles to understand these phenomena is the large degree of uncertainty involved and exposed as irregularity in the observed data. However, if the embedded

physical mechanism is extracted, then the problem will be understood and captured in a much clearer and easier way. For details, refer to Li et al. (2012b) and Wang and Li (2011).

Representations Based on Mathematical Decomposition

The first two moments, i.e., the mean and the correlation function, could usually capture the major characteristics of a stochastic process. Particularly, these two functions are complete for a Gaussian process. For a zero-mean Gaussian stationary process, the power spectral density function (PSD) is adequate to capture its probabilistic information. These functions essentially belong to phenomenological descriptions although in some cases the physical mechanism is involved when deriving the PSD, e.g., in the Kanai-Tajimi spectrum for ground motions (Tajimi 1960). However, due to its simplicity PSD models are widely employed in most engineering disciplines including earthquake engineering, ocean engineering, wind engineering, etc. Thus, how to represent a stochastic process in time domain by an explicit random function given its PSD is very important. A variety of methods including the Karhunen-Loève decomposition, the spectral representations and improvements were developed (Spanos et al. 2007; Shinozuka and Deodatis 1991; Grigoriu 2002).

The most widely used is the spectral representation method, by which a stochastic process is regarded as the sum of a series of harmonic functions with random phase, i.e.,

$$\hat{X}(t) = \sum_{j=1}^N A_j \cos(\omega_j t + \phi_j) \quad (11)$$

where A_j are deterministic amplitudes; ω_j are deterministic frequencies as inner points, say, uniformly spaced in the interval $[\omega_L, \omega_u]$ over which one-sided PSD is defined; and ϕ_j are independent random variables with identical uniform distribution over $[0, 2\pi]$. Clearly, $\mu_{\hat{X}} = E[\hat{X}(t)] = 0$, and $\sigma_{\hat{X}}^2(t) = E[\hat{X}^2(t)] = \sum_{j=1}^N \frac{A_j^2}{2}$. If the target one-sided

PSD is $G_X(\omega) = 2S_X(\omega)$ for $\omega \geq 0$ and otherwise $G_X(\omega) = 0$, where $S_X(\omega)$ is the double-sided PSD and symmetric to zero, then

$$\sigma_X^2(t) = \frac{1}{2\pi} \int_{\omega_L}^{\omega_u} G_X(\omega) d\omega \approx \sum_{j=1}^N \frac{1}{2\pi} G_X(\omega_j) \Delta\omega_j,$$

where $\Delta\omega_j$ is the length of the j -th frequency subinterval. Letting $\sigma_X^2(t) = \sigma_X^2(t)$ and comparing the terms one to one lead immediately to

$$A_j = \sqrt{\frac{1}{\pi} G_X(\omega_j) \Delta\omega_j} = \sqrt{\frac{2}{\pi} S_X(\omega_j) \Delta\omega_j}.$$

The properties of the representation in Eq. 11 were elaborately studied in Shinozuka and Deodatis (1991). It should be noted that the expression of the amplitude may vary in different literature due to the different position of the factor $\frac{1}{2\pi}$ in the Fourier transform. The spectral representation method is very simple and straightforward, but usually hundreds or even thousands of terms have to be retained. This leads to a large number of random variables, which induces difficulty in practice. To reduce the number of random variables is of great importance (Spanos et al. 2007).

A modification is to randomize the frequencies, and thus Eq. 11 is modified to

$$\tilde{X}(t) = \sum_{j=1}^N A(\tilde{\omega}_j) \cos(\tilde{\omega}_j t + \phi_j) \quad (12)$$

where $\tilde{\omega}_j$ are now random variables. For simplicity, assume the subintervals over which $\tilde{\omega}_j$ distribute are not overlapping, and construct a partition of $[\omega_L, \omega_u]$, i.e., the support of PDF of $\tilde{\omega}_j$ is $[\omega_{j-1}^{(p)}, \omega_j^{(p)}]$, $\omega_0^{(p)} = \omega_L$, $\omega_N^{(p)} = \omega_u$, $\omega_0^{(p)} < \omega_1^{(p)} < \dots < \omega_{N-1}^{(p)} < \omega_N^{(p)}$, and there are $\cup_{j=1}^N [\omega_{j-1}^{(p)}, \omega_j^{(p)}] = [\omega_L, \omega_u]$ and $[\omega_{j-1}^{(p)}, \omega_j^{(p)}] \cap [\omega_{k-1}^{(p)}, \omega_k^{(p)}] = \emptyset$. In this case, there is

$$\sigma_{\tilde{X}}^2(t) = E[\tilde{X}^2(t)] = \sum_{j=1}^N \frac{1}{2} E[A^2(\tilde{\omega}_j)] \quad (13)$$

and $\sigma_X^2(t) = \frac{1}{2\pi} \int_{\omega_L}^{\omega_u} G_X(\omega) d\omega = \sum_{j=1}^N \frac{1}{2\pi} \int_{\omega_{j-1}^{(p)}}^{\omega_j^{(p)}} G_X(\omega) d\omega$. Letting $\sigma_{\tilde{X}}^2(t) = \sigma_X^2(t)$ and

making the terms identical one to one lead to

$$E[A^2(\tilde{\omega}_j)] = \frac{1}{\pi} \int_{\omega_{j-1}^{(p)}}^{\omega_j^{(p)}} G_X(\omega) d\omega. \text{ If the PDF of } \tilde{\omega}_j \text{ is } p_{\tilde{\omega}_j}(\omega), \text{ then it follows that } \int_{\omega_{j-1}^{(p)}}^{\omega_j^{(p)}} A^2(\omega) p_{\tilde{\omega}_j}(\omega) d\omega = \frac{1}{\pi} \int_{\omega_{j-1}^{(p)}}^{\omega_j^{(p)}} G_X(\omega) d\omega, \text{ which leads to}$$

$$A(\tilde{\omega}_j) = \sqrt{\frac{G_X(\tilde{\omega}_j)}{\pi p_{\tilde{\omega}_j}(\tilde{\omega}_j)}} \quad (14)$$

The representation in Eq. 12 is called the stochastic harmonic function representation (SHF) and is of great flexibility by choosing different PDFs for the randomized frequencies (Chen et al. 2013). Particularly, if the PDF of $\tilde{\omega}_j$ takes the shape of PSD, i.e.,

$$p_{\tilde{\omega}_j}(\omega) = G_X(\omega) / \int_{\omega_{j-1}^{(p)}}^{\omega_j^{(p)}} G_X(\omega) d\omega, \text{ then Eq. 14}$$

$$\text{becomes } A(\tilde{\omega}_j) = \sqrt{\frac{1}{\pi} \int_{\omega_{j-1}^{(p)}}^{\omega_j^{(p)}} G_X(\omega) d\omega}. \text{ This is}$$

called the SHF of the first kind (SHF-I). If $\tilde{\omega}_j$ follows the uniform distribution over $[\omega_{j-1}^{(p)}, \omega_j^{(p)}]$, then from Eq. 14 there is $A(\tilde{\omega}_j) = \sqrt{\frac{1}{\pi} G_X(\tilde{\omega}_j) (\omega_j^{(p)} - \omega_{j-1}^{(p)})}$. This is called the SHF of the second kind (SHF-II).

What should be stressed is that the SHF representations are exact in the sense of reproducing the target PSD exactly. This is superior to the spectral representation method which is an approximate approach. Besides, studies show that although any arbitrary number of components could reproduce the target PSD, usually about 7–10 components are adequate considering the shape of the samples and the one-dimensional PDF of the generated process.

Numerical Algorithms

Although for some simple systems the analytical solution of GDEE is possible, for most practical problems, numerical algorithms are needed.

Procedures of Solution

The task is to obtain the instantaneous PDF $p_{Y_\ell}(y, t)$. For this purpose, the GDEE (Eq. 10) should be firstly solved under appropriate initial and boundary conditions which usually take

$$\left. \begin{aligned} p_{Y_\ell \Theta}(y, \theta, t) \Big|_{t=t_0} &= \delta(y - y_{\ell,0}) p_{\Theta}(\theta) \\ p_{Y_\ell \Theta}(y, \theta, t) \Big|_{y \rightarrow \pm\infty} &= 0 \end{aligned} \right\} \quad (15)$$

if the initial value $y_{\ell,0}$ is deterministic and independent of the random parameters. After Eq. 10 is solved Eq. 7 is employed to yield $p_{Y_\ell}(y, t)$. However, to solve the GDEE (Eq. 10), the coefficient $\dot{Y}_\ell(\theta, t)$ should be determined first. This comes from the embedded physical mechanism and is the solution of the physical equation (Eq. 1, or equivalently Eq. 2). Accordingly, the procedure of solution includes the following steps:

Step 1. Specify a representative point set $Q = \{\theta_q = (\theta_{1,q}, \dots, \theta_{s,q}) | \theta_q \in \Omega_\Theta, q = 1, 2, \dots, n_{pt}\}$ and the corresponding assigned probabilities $P_q, q = 1, 2, \dots, n_{pt}$, where n_{pt} is the total number of selected points and $0 < P_q < 1, \sum_{q=1}^{n_{pt}} P_q = 1$. This could be regarded as a first partial discretization of Eq. 10.

Step 2. For each specified $\{\Theta = \theta_q\}$, carry out deterministic analysis, say, the time integration method (Belytschko et al. 2000), to solve the physical equation (Eq. 1, or Eq. 2) to yield $\dot{Y}_\ell(\theta_q, t)$.

Step 3. Substitute $\dot{Y}_\ell(\theta_q, t)$ in the GDEE (Eq. 10) and solve it under the boundary and initial conditions (Eq. 15) by, say, the finite difference method with TVD scheme to yield the numerical result of $p_{Y_\ell \Theta}(y, \theta_q, t)$.

Step 4. Make a summation according to Eq. 7 to yield the numerical results of the instantaneous PDF $p_{Y_\ell}(y, t)$, i.e., $p_{Y_\ell}(y, t) = \sum_{q=1}^{n_{pt}} p_{Y_\ell \Theta}(y, \theta_q, t)$.

Optimal Selection of Representative Points

The point selection in Step 1 is of paramount importance to the efficiency and accuracy of PDEM. A consistent treatment is based on the partition of probability assigned space (Li et al. 2012a).

Let Ω_q denote the representative space of the point θ_q . Ω_q s construct a partition of Ω_Θ , i.e., $\bigcup_{q=1}^{n_{pt}} \Omega_q = \Omega_\Theta$ and $\Pr\{\omega \in \Omega_q \cap \Omega_k\} = 0$ for any $q \neq k$, where $\Pr\{\cdot\}$ denotes the probability of the event. The assigned probability of the point θ_q is then specified by $P_q = \int_{\Omega_q} p_\Theta(\theta) d\theta$, which clearly satisfies $0 < P_q < 1$ and $\sum_{q=1}^{n_{pt}} P_q = 1$. For problems with $2 \leq s \leq 4$, the method based on tangent spheres performs well, while for problems with $2 \leq s \leq 18$, the number theoretical method together with hyper-ball sieving could be employed (Li and Chen 2009). A versatile approach based on the generalized F-discrepancy (GF-discrepancy) was recently developed for general nonuniform, non-normal distributions.

Let the marginal cumulative distribution function (CDF) of the basic random variable Θ_j , $j = 1, 2, \dots, s$ be $F_j(\theta) = \int_{-\infty}^{\theta} p_{\Theta_j}(x) dx$. Denote the empirical marginal CDF of Θ_j by $\tilde{F}_j(\theta; Q) = \sum_{q=1}^{n_{pt}} P_q I\{\theta_{j,q} < \theta\}$, where $I\{\cdot\}$ is the indicator function of which the value is 1 when the bracketed event is true, and otherwise zero, $\theta_{j,q}$ is the j -th component of θ_q . By these a GF-discrepancy could be introduced:

$$D_{GF}(Q) = \max_{1 \leq j \leq s} \left\{ \sup_{-\infty \leq \theta \leq \infty} |\tilde{F}_j(\theta; Q) - F_j(\theta)| \right\} \quad (16)$$

It is demonstrated that the smaller the GF-discrepancy, the better the performance of the point set in the sense that the error of the involved high-dimensional integral is smaller (Chen and Zhang 2013). Therefore, the task is now to find some point set Q minimizing the GF-discrepancy, i.e.,

$$\begin{aligned} Q &= \arg[\min D_{GF}(Q)] \\ &= \arg \left[\min \left(\max_{1 \leq j \leq s} \left\{ \sup_{-\infty \leq \theta \leq \infty} |\tilde{F}_j(\theta; Q) - F_j(\theta)| \right\} \right) \right] \end{aligned} \quad (17)$$

Plenty of optimization methods could be adopted to complete this task and yield the optimal representative point set.

Finite Difference Method with TVD Scheme

In Step 3 a series of partial differential equations are solved. These equations are of the form

$$\frac{\partial p(y, t)}{\partial t} + a \frac{\partial p(y, t)}{\partial y} = 0 \quad (18)$$

for different values of a in the case of different θ_q . Quite a few difference schemes were developed. According to computational experiences the schemes with TVD (total variation diminishing) property are satisfactory (Li and Chen 2009). One of such schemes that is adopted is as follows:

$$\begin{aligned} p_j^{(k+1)} &= p_j^{(k)} - \frac{1}{2}(\lambda a - |\lambda a|) \Delta p_{j+\frac{1}{2}}^{(k)} - \frac{1}{2}(\lambda a + |\lambda a|) \Delta p_{j-\frac{1}{2}}^{(k)} \\ &\quad - \frac{1}{2}(|\lambda a| - |\lambda a|^2) (\psi_{j+\frac{1}{2}} \Delta p_{j+\frac{1}{2}}^{(k)} - \psi_{j-\frac{1}{2}} \Delta p_{j-\frac{1}{2}}^{(k)}) \end{aligned} \quad (19)$$

where $p_j^{(k)}$ denotes $p(x_j, t_k)$, $x_j = j\Delta x$, $j = 0, \pm 1, \pm 2, \dots$, $t_k = k\Delta t$, $k = 0, 1, 2, \dots$, and Δx and Δt are the space and time step, respectively; the mesh ratio $\lambda = \Delta t/\Delta x$; $\Delta p_{j+\frac{1}{2}}^{(k)} = p_{j+1}^{(k)} - p_j^{(k)}$; and $\Delta p_{j-\frac{1}{2}}^{(k)} = p_j^{(k)} - p_{j-1}^{(k)}$. The factors $\psi_{j+\frac{1}{2}}, \psi_{j-\frac{1}{2}}$ are called the flux limiters, which are related to the irregularity of the curve of $p(x, t)$ in terms of x . As $\psi_{j+\frac{1}{2}} = \psi_{j-\frac{1}{2}} \equiv 1$ the scheme in Eq. 19 reduces to the well-known Lax-Wendroff scheme, whereas as $\psi_{j+\frac{1}{2}} = \psi_{j-\frac{1}{2}} \equiv 0$ the scheme in Eq. 19 reduces to the one-sided scheme. To achieve the TVD property, $\psi_{j+\frac{1}{2}}, \psi_{j-\frac{1}{2}}$ should depend on the parameters characterizing the irregularity of $p(x, t)$, i.e.,

$$\begin{aligned} r_{j+\frac{1}{2}}^+ &= \frac{\Delta p_{j+\frac{3}{2}}^{(k)}}{\Delta p_{j+\frac{1}{2}}^{(k)}} = \frac{p_{j+2}^{(k)} - p_{j+1}^{(k)}}{p_{j+1}^{(k)} - p_j^{(k)}}, \\ r_{j+\frac{1}{2}}^- &= \frac{\Delta p_{j-\frac{1}{2}}^{(k)}}{\Delta p_{j+\frac{1}{2}}^{(k)}} = \frac{p_j^{(k)} - p_{j-1}^{(k)}}{p_{j+1}^{(k)} - p_j^{(k)}} \end{aligned} \quad (20)$$

The flux limiter $\psi_0(r) = \max(0, \min(2r, 1), \min(r, 2))$ could be recommended and then

$$\psi_{j+\frac{1}{2}}\left(r_{j+\frac{1}{2}}^+, r_{j+\frac{1}{2}}^-\right) = u(-a)\psi_0\left(r_{j+\frac{1}{2}}^+\right) + u(a)\psi_0\left(r_{j+\frac{1}{2}}^-\right) \quad (21)$$

where $u(\cdot)$ is Heaviside's function, of which the value is 1 if the argument is greater than 0 and is zero otherwise.

First-Passage Reliability and Global Reliability

First-Passage Reliability

The first-passage reliability is of great concern to engineering structures exposed to disastrous dynamic excitations. For instance, if the first-passage failure is defined in terms of $Y_\ell(t)$, the reliability is then

$$R(t) = \Pr\{Y_\ell(\Theta, \tau) \in \Omega_s, \text{ for } 0 \leq \tau \leq t\} \quad (22)$$

Note that the evolution of the probability density of $(Y_\ell(t), \Theta)$ is governed by Eq. 10. Eq. 22 means that once the probability outcrosses the boundary of the safety domain, then it will not return to the safety domain. This could be expressed mathematically as an absorbing boundary condition

$$p_{Y_\ell, \Theta}(y, \theta, t)|_{y \in \Omega_f} = 0 \quad (23)$$

where Ω_f is the failure domain. This of course is also equivalent to imposing the absorbing boundary condition $p_{Y_\ell}(y, t)|_{y \in \Omega_f} = 0$ for Eq. 5. Replacing the boundary condition in Eq. 15 by Eq. 23 and carrying out the procedures in the preceding section will yield the "remaining" PDF $\tilde{p}_{Y_\ell}(y, t)$, and thus the reliability is given by

$$R(t) = \int_{\Omega_s} \tilde{p}_{Y_\ell}(y, t) dy = \int_{-\infty}^{\infty} \tilde{p}_{Y_\ell}(y, t) dy \quad (24)$$

Global Reliability

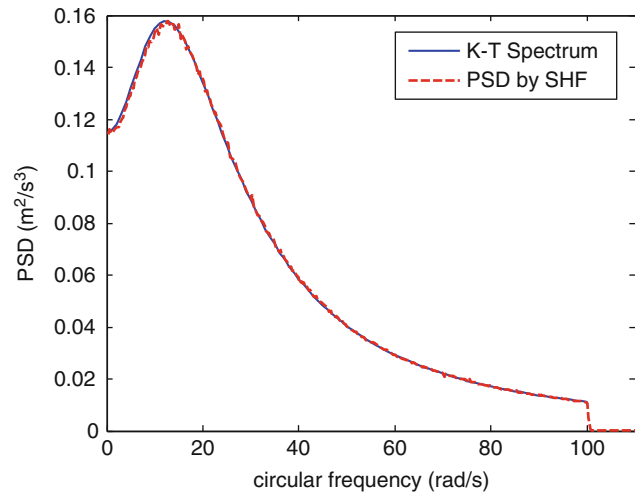
Traditionally, it is called the system reliability if more than one failure modes are involved.

However, due to the fact that the failure of a global structure is usually dependent on the whole process of development of nonlinearity in the system, the reliability should be captured in this whole process involving randomness. In this sense, it is more appropriate to call it the global reliability of a structure (Chen and Li 2007; Li and Chen 2009). For instance, if the failure of the global structure is determined by some combinations or even functional of $Y_\ell(\Theta, t)$ and $Y_k(\Theta, t)$, say, the Park-Ang model (Park and Ang 1985), then a corresponding equivalent process, denoted as $Z(t) = f[Y_k(\Theta, \tau), Y_\ell(\Theta, \tau); 0 \leq \tau \leq t]$ for convenience, as some kind of compound or even functional of $Y_\ell(\Theta, t)$ and $Y_k(\Theta, t)$, could be constructed such that the global reliability is given by $R_G(t) = \Pr\{Z(\Theta, \tau) \in \Omega_s, \text{ for } 0 \leq \tau \leq t\}$. Then the method similar to that described in the above subsection could be adopted, or the PDF of the equivalent extreme value $\max_{0 \leq \tau \leq t} Z(\Theta, \tau)$ could be obtained, and then the global reliability is yielded (Li et al. 2007, Li & Chen 2009).

Numerical Example of Applications

To illustrate the PDEM, a 10-story floor-shear structure with lumped masses on the floors subjected to stochastic ground motion is studied. To be simple, all the system parameters are taken as deterministic variables. The lumped masses from bottom to top are 1.5, 1.5, 1.5, 1.4, 1.4, 1.4, 1.3, 1.3, 1.3, and $0.7 (\times 10^5 \text{ kg})$, respectively; the inter-story stiffness from bottom to top are 4.25, 4.25, 4.25, 4.25, 4.0, 4.0, 4.0, 4.0, 3.5, and $3.5 (\times 10^{10} \text{ N/m})$. Rayleigh damping is adopted, i.e., the damping matrix $\mathbf{C} = a\mathbf{M} + b\mathbf{K}$, where $\mathbf{M}$ and $\mathbf{K}$ are the mass and stiffness matrix, and $a = 0.01$ and $b = 0.005$. The Bouc-Wen model is taken to characterize the hysteretic behavior of the restoring force (Ma et al. 2004; Li and Chen 2009), where the basic parameters take $A = 1$, $n = 1$, $q = 0$, $p = 600$, $d_\psi = 0$, $\lambda = 0.5$, $\psi = 0.2$, $\beta = 60$, $\gamma = 10$, $d_v = 200$, $d_\eta = 200$, and $\zeta = 0.95$. For the linear system $\alpha = 1$, and for the nonlinear system $\alpha = 0.01$.

**Probability Density
Evolution Method in
Stochastic Dynamics,
Fig. 1 Target and
reproduced PSD**



The Kanai-Tajimi spectrum is taken as the target spectrum of the stochastic ground motions, i.e.,

$$S(\omega) = \frac{1 + 4\zeta_0^2(\omega/\omega_0)^2}{(1 - (\omega/\omega_0))^2 + 4\zeta_0^2(\omega/\omega_0)^2} S_0 \quad (25)$$

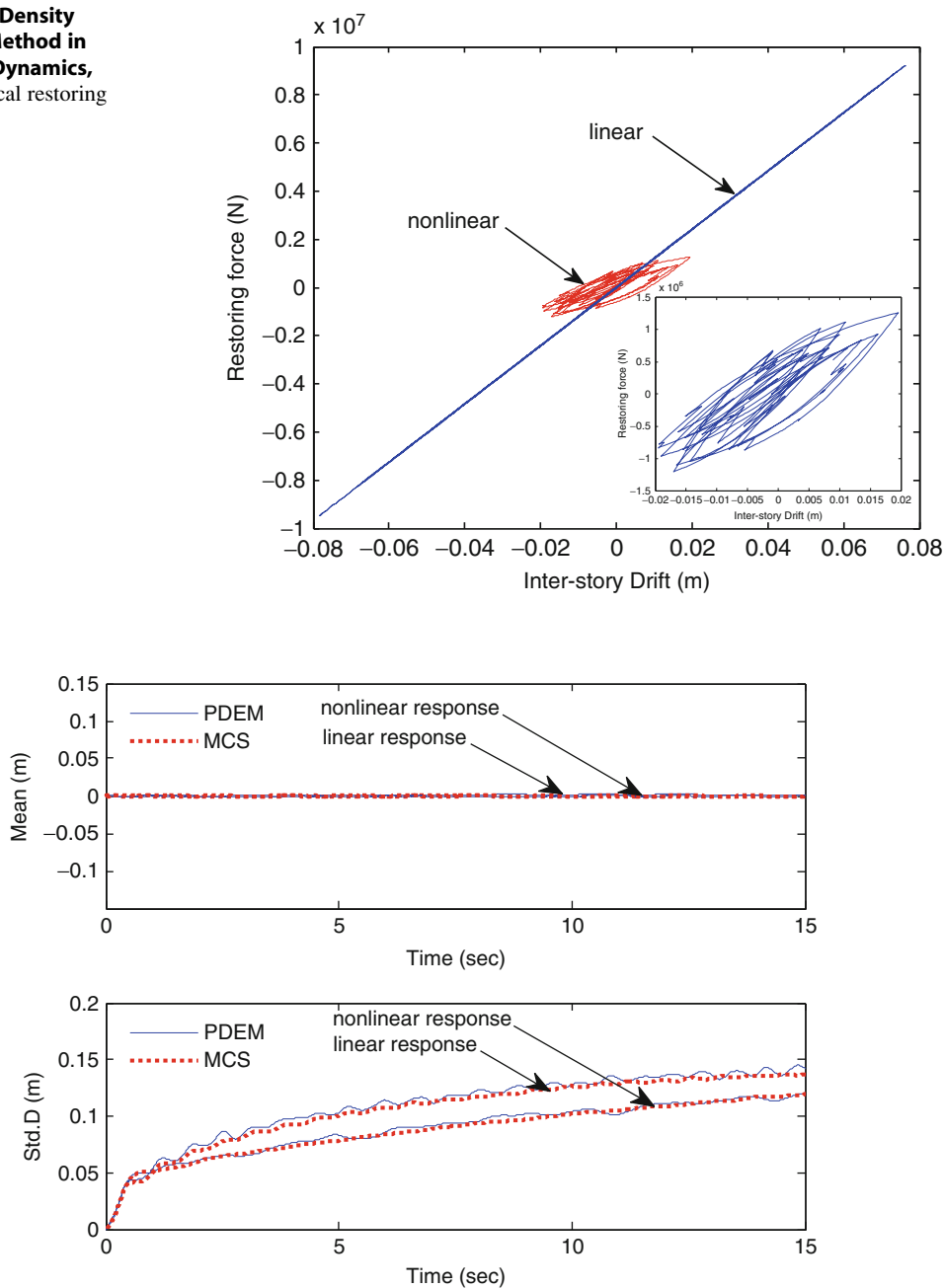
where $S(\omega)$ is the PSD, $\omega_0 = 16.9$, $\zeta_0 = 0.94$, and $\omega_u = 100$ rad/s. S_0 is such determined that the standard deviation of the ground motion acceleration is 1.0 m/s^2 .

In Fig. 1 shown is the target PSD in Eq. 25 (labeled as “K-T spectrum”) and the PSD reproduced by the SHF-II including 20 terms. It is seen that with only 20 SHF components, the PSD could be reproduced in high accuracy. This will also result with high accuracy in the second-order statistics of response of the linear system, but for the nonlinear structure, it seems that more terms should be retained. Figure 2 shows typical curves of restoring force vs. inter-story drift, demonstrating that in the nonlinear case strong hysteresis is involved. In Fig. 3 pictured are the mean and standard deviation of response (top displacement) of linear and nonlinear structures by PDEM and 20,000 times of Monte Carlo simulations (labeled as “MCS”), respectively. In this paper 800 representative time histories generated by SHF-II are adopted. Computational

experiences show that this number could still be reduced. Theoretically, the mean of linear and nonlinear structures should be zero due to the symmetry of distribution of the input ground acceleration process. This is clearly verified from the upper figure of Fig. 3, which shows that the mean by the PDEM is quite small in comparison to the standard deviation (the coordinate in the upper figure of Fig. 3 is in the order of magnitude of the standard deviation). Also it is seen that the standard deviation of the nonlinear response is smaller than that of the linear structure. Pictured in Fig. 4 are the PDF evolution surfaces during an identical period for linear and nonlinear systems. Note that if the input is Gaussian, the output should also be Gaussian for linear systems. Clear differences between the PDFs of linear and nonlinear responses could be observed from the figures.

It is noted that by PDEM the global reliability of the structure could also be obtained by imposing absorbing conditions. In Fig. 5 shown are the time-variant first-passage reliabilities of the nonlinear structure in ordinary and logarithmic coordinates for different threshold values by PDEM and Monte Carlo simulation (20,000 times). The thresholds 0.4, 0.3, 0.25, and 0.2 m correspond to the displacement angle of $1/75$, $1/100$, $1/120$, and

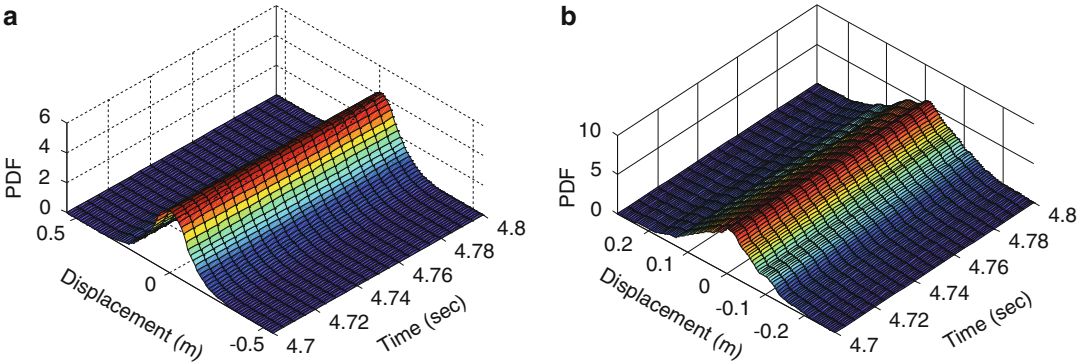
Probability Density Evolution Method in Stochastic Dynamics, Fig. 2 Typical restoring force



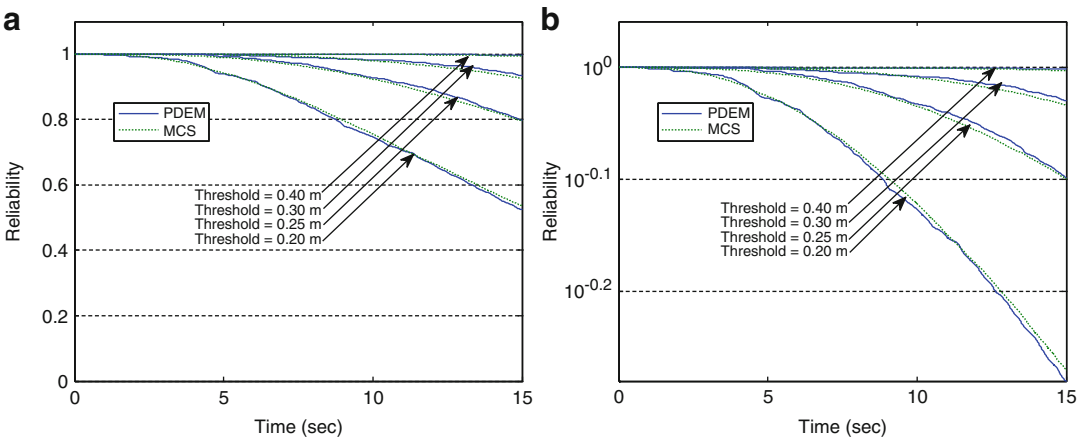
Probability Density Evolution Method in Stochastic Dynamics, Fig. 3 The mean and standard deviation

1/150, respectively. It is shown that the reliability decreases almost continuously from some time monotonically, but noticeably not in an exponential way as predicted by the Poisson assumption. Actually, from Fig. 5b it

is seen that the absolute value of slope of the logarithmic reliability is increasing, which means that the hazard rate of the structure exhibiting nonlinear behaviors is increasing as the time elapses.



Probability Density Evolution Method in Stochastic Dynamics, Fig. 4 PDF evolution surface. (a) Linear structure; (b) nonlinear structure



Probability Density Evolution Method in Stochastic Dynamics, Fig. 5 Reliability of the nonlinear structure. (a) Ordinary coordinate; (b) logarithmic coordinate

Summary

The probability density evolution method is outlined and illustrated in this entry. It is concluded that (1) the thought of physical stochastic systems provides a new perspective to stochastic dynamics and (2) the probability density evolution method shows its versatility in stochastic dynamics, particularly for MDOF nonlinear systems subjected to non-white noise excitations. However, improvements and extension of the physical stochastic models of dynamic excitations and more robust and efficient numerical algorithms are still needed.

Cross-References

- [Reliability Estimation and Analysis](#)
- [Stochastic Analysis of Nonlinear Systems](#)
- [Stochastic Ground Motion Simulation](#)
- [Structural Reliability Estimation for Seismic Loading](#)
- [Structural Seismic Reliability Analysis](#)

References

Belytschko T, Liu WK, Moran B (2000) Nonlinear finite elements for continua and structures. Wiley, Chichester

- Chen JB, Li J (2005) Dynamic response and reliability analysis of nonlinear stochastic structures. *Probab Eng Mech* 20(1):33–44
- Chen JB, Li J (2007) The extreme value distribution and dynamic reliability analysis of nonlinear structures with uncertain parameters. *Struct Saf* 29:77–93
- Chen JB, Li J (2009) A note on the principle of preservation of probability and probability density evolution equation. *Probab Eng Mech* 24(1):51–59
- Chen JB, Li J (2010) Stochastic seismic response analysis of structures exhibiting high nonlinearity. *Comput Struct* 88(7–8):395–412
- Chen JB, Zhang SH (2013) Improving point selection in cubature by a new discrepancy. *SIAM J Sci Comput* 35(5):A2121–A2149
- Chen JB, Sun WL, Li J, Xu J (2013) Stochastic harmonic function representation of stochastic processes. *J Appl Mech* 80(1):011001-1-11
- Ghanem R, Spanos PD (1991) *Stochastic finite elements: a spectral approach*. Springer, Berlin
- Goller B, Pradlwarter HJ, Schuëller GI (2013) Reliability assessment in structural dynamics. *J Sound Vib* 332:2488–2499
- Grigoriu M (2002) *Stochastic calculus*. Birkhäuser, Boston
- Li J, Chen JB, Fan WL (2007) The equivalent extreme-value event and evaluation of the structural system reliability. *Struct Saf* 29:112–131
- Li J (1996) *Stochastic structural systems: analysis and modeling*. Science Press, Beijing (in Chinese)
- Li J, Chen JB (2003) The probability density evolution method for analysis of dynamic nonlinear response of stochastic structures. *Acta Mech Sinica* 35(6):716–722
- Li J, Chen JB (2005) Dynamic response and reliability analysis of structures with uncertain parameters. *International Journal for Numerical Methods in Engineering* 62:289–315
- Li J, Chen JB (2008) The principle of preservation of probability and the generalized density evolution equation. *Struct Saf* 30:65–77
- Li J, Chen JB (2009) *Stochastic dynamics of structures*. Wiley, Singapore
- Li J, Chen JB, Sun W, Peng Y (2012a) Advances of probability density evolution method for nonlinear stochastic systems. *Probab Eng Mech* 28:132–142
- Li J, Yan Q, Chen JB (2012b) Stochastic modeling of engineering dynamic excitations for stochastic dynamics of structures. *Probab Eng Mech* 27(1):19–28
- Li J, Wu JY, Chen JB (2014) *Stochastic damage mechanics of concrete structures (in Chinese)*. Science Press, Beijing
- Lin N, Emanuel K, Oppenheimer M, Vanmarcke E (2012) Physically based assessment of hurricane surge threat under climate change. *Nat Climate Change* 2(6):462–467
- Ma F, Zhang H, Bockstedte A, Foliente GC, Paevere P (2004) Parameter analysis of the differential model of hysteresis. *J Appl Mech* 71:342–349
- Shinozuka M, Deodatis G (1991) Simulation of stochastic processes by spectral representation. *Appl Mech Rev* 44(4):191–204
- Spanos PD, Beer M, Red-Horse J (2007) Karhunen–Loève expansion of stochastic processes with a modified exponential covariance kernel. *J Eng Mech* 133(7):773–779
- Park YJ, Ang AH-S (1985) Mechanistic seismic damage model for reinforced concrete. *J Struct Eng* 111(4):722–739
- Roberts JB, Spanos PD (1990) *Random vibration and statistical linearization*. Wiley, Chichester
- Tajimi H (1960) A statistical method of determining the maximum response of a building structure during an earthquake. In: *Proceedings of second world conference on Earthquake engineering*, Tokyo, vol 11, pp 781–797
- Wang D, Li J (2011) Physical random function model of ground motions for engineering purposes. *Sci China Technol Sci* 54(1):175–182
- Zhu WQ (2006) Nonlinear stochastic dynamics and control in Hamiltonian formulation. *Appl Mech Rev* 59:230–248

Probability Seismic Hazard Mapping of Taiwan

Chin-Tung Cheng¹, Pao-Shan Hsieh¹,
Po-Shen Lin¹, Yin-Tung Yen¹ and
Chung-Han Chan^{2,3}

¹Disaster Prevention Technology Research Center, Sinotech Engineering Consultants, Inc., Taipei, Taiwan, ROC

²Department of Geosciences, National Taiwan University, Taipei, Taiwan, ROC

³Earth Observatory of Singapore, Nanyang Technological University, Singapore, Singapore

Synonyms

Hazard curve; Hazard map; Probabilistic seismic hazard assessment; Seismic hazard mitigation; Taiwan

Introduction

Studies on seismic hazard mitigation are important for seismologists, earthquake engineers, and

related scientists. Among studies, probabilistic seismic hazard assessments (PSHAs) provide the probability of exceedance for a specific ground motion level during a time interval (see entry “► [Site Response for Seismic Hazard Assessment](#)”). PSHA results can provide a key reference for the determination of hazard mitigation policies related to building codes and the site selection of public structures. Therefore, multidisciplinary scientists have attempted to build reliable systems for PSHAs.

Due to the plate boundary between the Eurasian and Philippine Sea Plates, Taiwan has a high earthquake activity (Fig. 1). In this region, devastating earthquakes lead to a loss of property and human life. Therefore, it is essential to develop a means of seismic hazard mitigation. One practical approach is to build a seismic hazard assessment system. Over the past few years, several studies have evaluated seismic hazards for Taiwan. For example, the Global Seismic Hazard Assessment Program (GSHAP, <http://www.seismo.ethz.ch/static/GSHAP/>) obtained a global probabilistic seismic hazard map that included Taiwan. However, this work employed earthquake catalogs obtained from global seismic networks rather than a detailed seismicity catalog from Taiwan. Another application was proposed by Campbell et al. (2002) who utilized seismic catalogs, active fault parameters, and ground motion prediction equations (GMPEs) for the world, the United States, and Taiwan. Cheng et al. (2007) evaluated seismic hazards for Taiwan and proposed a hazard map by integrating a catalog from a local network, active fault parameters, and seismogenic zones in Taiwan. Such studies are crucial for understanding seismic hazards in Taiwan. However, following these studies, many parameters and the database for seismic hazard assessments, such as understanding the tectonic setting, the distribution of active faults, GMPEs, and earthquake catalogs, have been revised and/or updated. By employing state-of-the-art parameters, an evaluation of seismic hazards can be more precise.

Seismic hazards for Taiwan are reevaluated through the PSHA approach as proposed by

Cornell (1968). According to this approach, parameters for seismogenic sources, which may result in seismic hazards, are required. Several seismogenic sources were characterized, including shallow regional sources, deep regional sources, crustal active fault sources, subduction intraslab sources, and subduction interface sources. The parameters for each source are discussed according to information from the tectonic setting, geology, geomorphology, geophysics, and earthquake catalog and present the results in the form of hazard maps and hazard curves. The results are compared with those proposed by previous studies and discuss their applicability for the future.

Methodologies and Seismic Activity Models

The PSHA Approach

The applied approach of PSHA was first developed by Cornell (1968) (see entry “Earthquakes and Tectonics: Probabilistic Seismic Hazard Assessment: An Overview”). According to the description of Kramer (1996), seismic hazards based on this approach can be assessed, as follows:

$$P[Y > y^*] = \iint P[Y > y^* | m, r] f_M(m) f_R(r) dm dr, \quad (1)$$

where $P[Y > y^*]$ is the probability, P , for a given ground motion parameter, Y , which exceeds a specific value, y^* . $P[Y > y^* | m, r]$ is the probability conditional on an earthquake with magnitude, m , imparted by a seismogenic source with the closest distance, r , between the site of interest and seismogenic source; and $f_M(m)$ and $f_R(r)$ are the probability density functions for magnitude and distance, respectively.

If there are N_S potential seismogenic sources near the site of interest, each of which has an average rate, ν_i . The total average exceedance rate, $\lambda_{y^*,0}$, for the region can be presented as follows:

$$\lambda_{y^0} = \sum_{i=1}^{N_S} v_i \iint P[Y > y^* | m, r] f_{M_i}(m) f_{R_i}(r) dm dr, \quad (2)$$

where $\sum_{n=1}^{N_S}$ is the summation of the contribution from N_S th seismogenic sources n .

PSHA uncertainties from different aspects were considered and properly treated. A logic tree approach was introduced to incorporate the uncertainties for seismogenic sources, the corresponding parameters of each source, and the GMPEs. The treatment of the weighting for each parameter is discussed in subsequent sections.

Seismic Activity Models

For the implementation of PSHA, the corresponding seismicity rate as a function of magnitude for each seismogenic source should be introduced (Eq. 2). Generally, there are two

models that present the relationships, the truncated exponential model and the characteristic earthquake model. In the following, both of these models are presented and discussed.

The Truncated Exponential Model

The truncated exponential model is based on Gutenberg-Richter's Law (G-R Law) (Gutenberg and Richter 1954), as follows:

$$\log(\dot{N}) = a - bM \quad (3)$$

where $\dot{N}$ is the annual rate for events with magnitudes larger than or equal to M and a and b are constants with values larger than 0. Following on G-R Law, the truncated exponential model represents the rate for a magnitude larger than the maximum magnitude, m_u , as 0. Thus, the cumulative annual rate, $\dot{N}(m)$, for a magnitude larger than or equal to m can be presented, as follows:

$$\dot{N}(m) = \dot{N}(m_0) \frac{\exp(-\beta(m - m_0)) - \exp(-\beta(m_u - m_0))}{1.0 - \exp(-\beta(m_u - m_0))} \quad \text{for } m_u \geq m \geq m_0 \quad (4)$$

where $\dot{N}(m_0)$ represents the cumulative annual rate for a magnitude of a minimum magnitude, m_0 , and m_0 and m_u represent the minimum and maximum magnitudes, respectively, of the seismogenic source. β can be represented as follows:

$$\beta = b \cdot \ln(10) \quad (5)$$

where b is the b -value in G-R Law (Eq. 3).

Wesnousky (1994) concluded that this model is suitable for regions with complex tectonic settings or multiple active faults. Thus, it was applied

for shallow regional sources, deep regional sources, and subduction intraslab sources.

The Characteristic Earthquake Model

The characteristic earthquake model was first proposed by Youngs and Coppersmith (1985). In addition to earthquake parameters, the model represents the seismicity rate by incorporating geological and geomorphological information. The cumulative annual rate for a magnitude larger or equal to m can be represented as follows:

$$\begin{aligned} \dot{N}(m) &= \dot{N}^e \frac{\exp(-\beta(m - m_0)) - \exp(-\beta(m_u - 1/2 - m_0))}{1.0 - \exp(-\beta(m_u - 1/2 - m_0))} + \dot{N}^c \quad \text{for } m_0 \leq m \\ &\leq m_u - \frac{1}{2}; \quad \dot{N}(m) = \dot{N}^c \frac{m_u - m}{1/2} \quad \text{for } m_u - \frac{1}{2} \leq m \leq m_u, \end{aligned} \quad (6)$$

where $\dot{N}^e$ and $\dot{N}^c$ represent the cumulative annual rates predicted by the truncated exponential and

the characteristic earthquake models, respectively. $\dot{N}^e$ can be presented, as follows:

$$\dot{N}^e = \frac{\mu A_f S \cdot (1 - \exp(-\beta(m_u - m_0 - 1/2)))}{\exp(-\beta(m_u - m_0 - 1/2)) \cdot M_0(m_u) \cdot \left[\frac{b \cdot 10^{-c/2}}{(c-b)} + \frac{b \exp(\beta)(1 - 10^{-c/2})}{c} \right]}, \quad (7)$$

where μ is the rigidity shear modulus, generally assumed to be 3×10^{10} Pascal (N/m²); A_f is the fault area; S is the slip rate; and c and d are constants. $\dot{N}^c$ can be represented as follows:

$$\dot{N}^c = \frac{1}{2} \dot{N}^e \frac{b \ln 10 \cdot \exp(-\beta(m_u - 3/2 - m_0))}{(1 - \exp(m_u - 1/2 - m_0))}. \quad (8)$$

In a comparison with the truncated exponential model, the characteristic earthquake model predicted lower rates for smaller magnitudes, whereas higher rates were predicted for larger magnitudes (Youngs and Coppersmith 1985).

To implement the characteristic earthquake model for PSHA, the cumulative annual rate should be in form of the rate $\lambda_n(m_i)$ (Eq. 2) between the magnitude bins of $m_i \pm dm/2$, which can be represented as follows:

$$\lambda_n(m_i) = N(m_i - dm/2) - N(m_i + dm/2), \quad (9)$$

where dm is the magnitude interval for the model and $N(m_i)$ is the rate for a magnitude larger or equal to m_i .

A few previous studies (Youngs and Coppersmith 1985; Wesnousky 1994) have suggested that the behavior of seismic activity along crustal active faults and subduction interfaces follows this model. Therefore, it was applied to the two seismogenic sources.

Seismogenic Tectonics in Taiwan

Tectonic Setting

Taiwan is located within the plate boundary between the Philippine Sea Plate and the Eurasian Plate (Fig. 1). Due to the interaction of the two plates, both subduction and collision take place in this region. Two subduction systems

surround this region. In the offshore of northeast Taiwan, the Philippine Sea Plate subducts to the north. As a result of back-arc spreading, the Okinawa Trough and the Ilan Plain formed in the northern section of the Ryukyu Volcanic Arc. In southern Taiwan, the Eurasian Plate subducts to the east. The Longitudinal Valley is the arc-continental collision boundary between the two plates. Collision began in the late Miocene in northern Taiwan. Due to lateral collision between plates, collision activity continues to migrate to the south. Currently, activity takes place in central and southwestern Taiwan. Northern Taiwan, in contrast, is a post-collision region with relatively low seismic activity.

The Earthquake Catalog

The utilized earthquake catalog was collected by the Central Weather Bureau (CWB) and provided earthquake parameters for events from 1900 to 2010 in Taiwan. Prior to 1973, a total of 15 stations equipped with Gray-Milne, Wiechert, or Omori seismographs were maintained. After 1973, the Taiwan Telemetric Seismic Network (TTSN) was established. The TTSN consists of 25 stations within the region of Taiwan. In the TTSN network, real-time signals are transmitted from field stations to a central station via leased telephone lines.

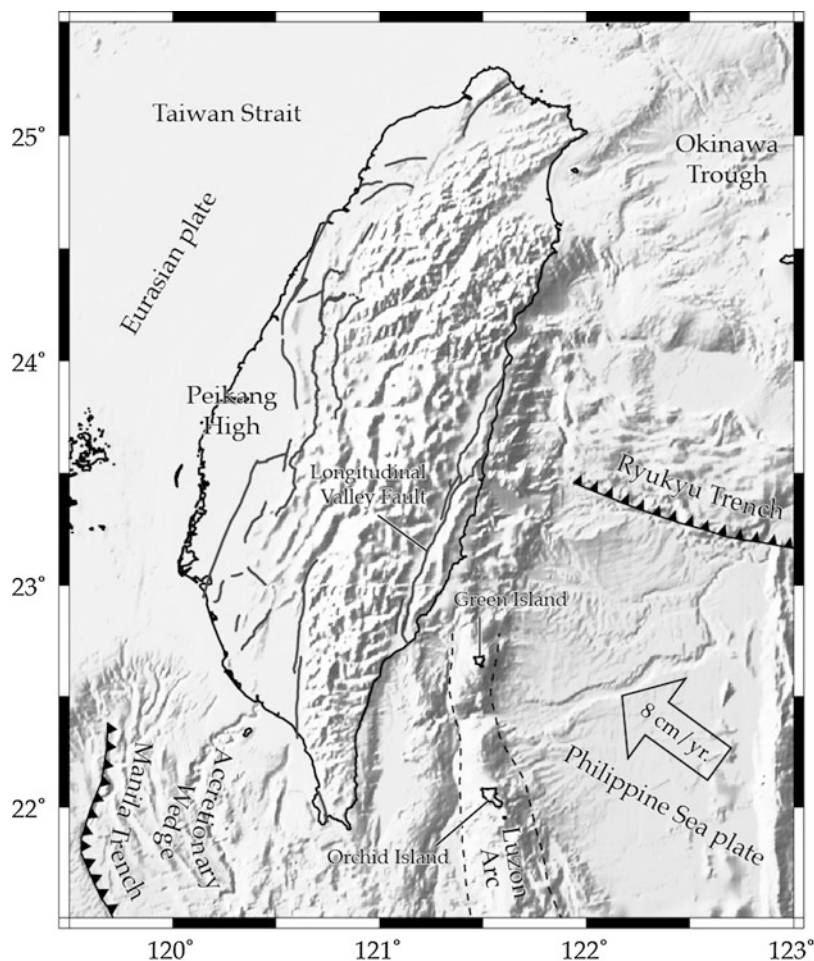
To assess seismic hazards, earthquake parameters should be analyzed using the following procedures: a magnitude harmonization from different scales, an evaluation of the magnitude of completeness, and a declustering process. In the following, each step of the procedure is described in detail.

Magnitude Harmonization

Since the magnitude scales for a catalog during different periods are generally different, it is critical to harmonize the magnitude scales during

Probability Seismic Hazard Mapping of Taiwan, Fig. 1

The tectonic setting in Taiwan and its vicinity (Modified from Cheng et al. 2007)

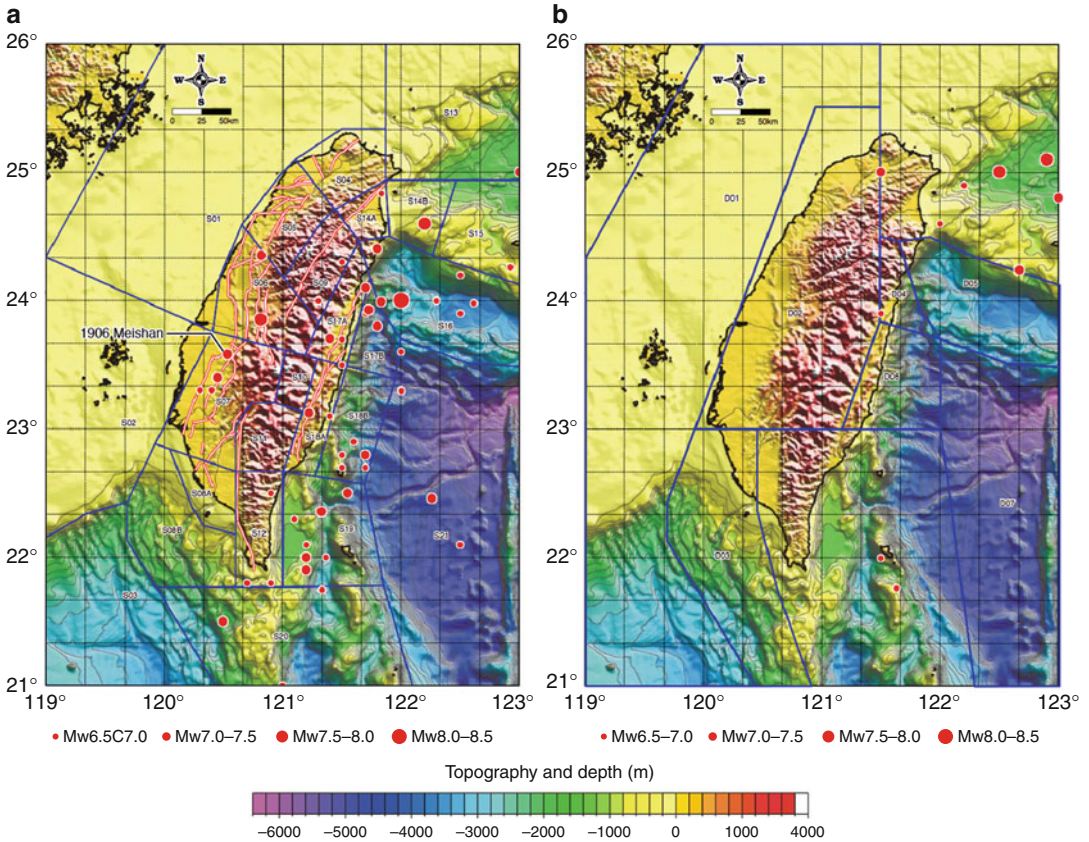


P

different periods. Previous studies (Hanks and Kanamori 1979 and references therein) have suggested a moment magnitude (M_W) for PSHA, since this scale is evaluated based on rupture dimensions and slip magnitudes. Additionally, the M_W scale is not affected by saturation at higher magnitudes. For example, the 1999 Chi-Chi, Taiwan, earthquake was determined to have a M_W of 7.6, whereas its corresponding M_L was 7.3. The discrepancy can be attributed to the saturation of M_L (Cheng et al. 2007). Therefore, the earthquake catalog obtained by Tsai et al. (2000) was considered. The magnitude scales of this catalog have been harmonized as M_W from 1900 to 1999. The magnitude scales of the catalog were harmonized according to the procedure of Tsai et al. (2000).

Magnitude of Completeness

To improve the reliability of the parameters for seismogenic sources, the catalog during the period when the network recorded all earthquakes with a certain magnitude threshold was considered. The threshold is known as the magnitude completeness, M_c . Thus, M_c for catalogs during different periods must be examined in advance. Chen et al. (2012) evaluated the M_c of the CWB catalog using the maximum curvature method. A higher M_c between 4.3 and 4.8 was obtained prior to 1973. Once the TTSN was established, M_c decreased to between 2.0 and 3.0. Based on the temporal distribution of M_c , $M \geq 6.0$ earthquakes after 1900 (Fig. 2) and $M \geq 2.0$ earthquakes after 1973 (Fig. 3) were considered.



Probability Seismic Hazard Mapping of Taiwan, Fig. 2 The distribution of earthquakes with $M \geq 6.0$ at the depth of (a) ≤ 35 km and (b) > 35 km since 1900. The distribution of shallow and deep regional sources is

illustrated by the *blue polygons* in Fig. 2a and b, respectively. In Fig. 2a, crustal active fault sources are presented as *red lines*

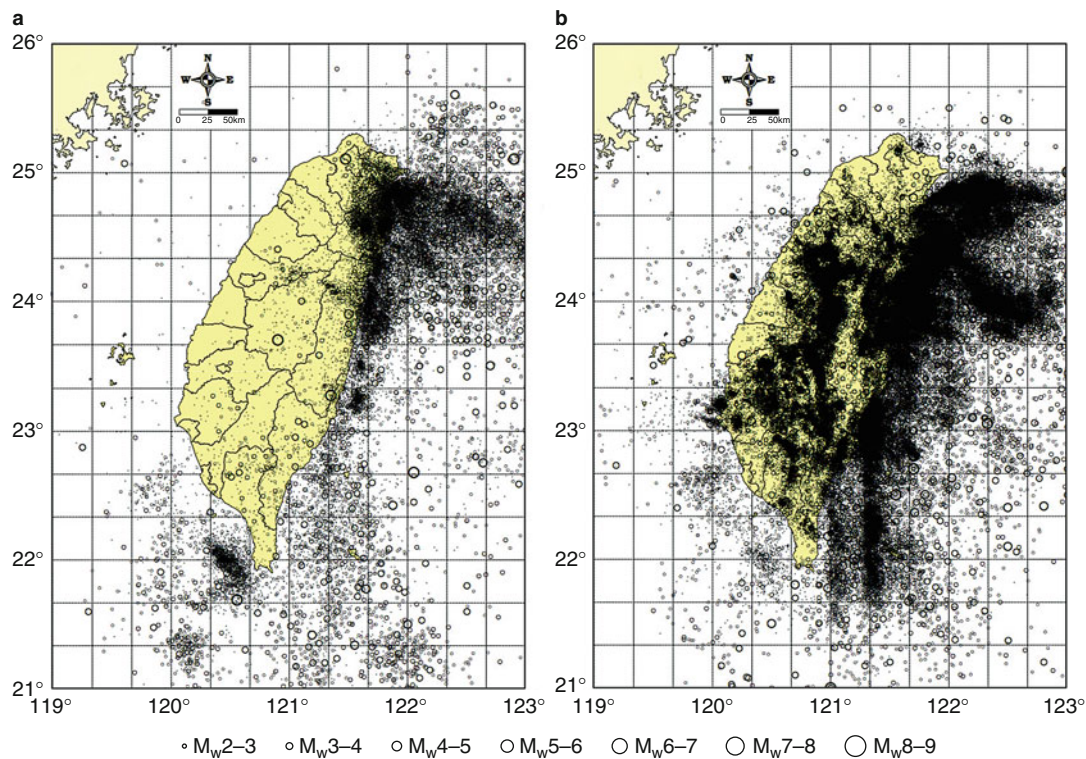
The Declustering Process

For application of the PSHA approach by Cornell (1968), it is assumed that the occurrence of earthquakes follows the Poisson procedure. In other words, earthquakes are independent of one another. However, in a catalog, earthquake sequences, which include foreshocks, mainshock, aftershocks, and swarms, can be observed. Therefore, it is critical to obtain a declustered catalog in respect to the PSHA (i.e., to remove foreshocks, aftershocks, and swarms from the catalog). The declustering approaches developed by Wyss (1979), Arabasz and Robinson (1976), Gardner and Knopoff (1974), and Uhrhammer (1986) were implemented. According to these approaches, earthquakes are considered dependent when their distance and

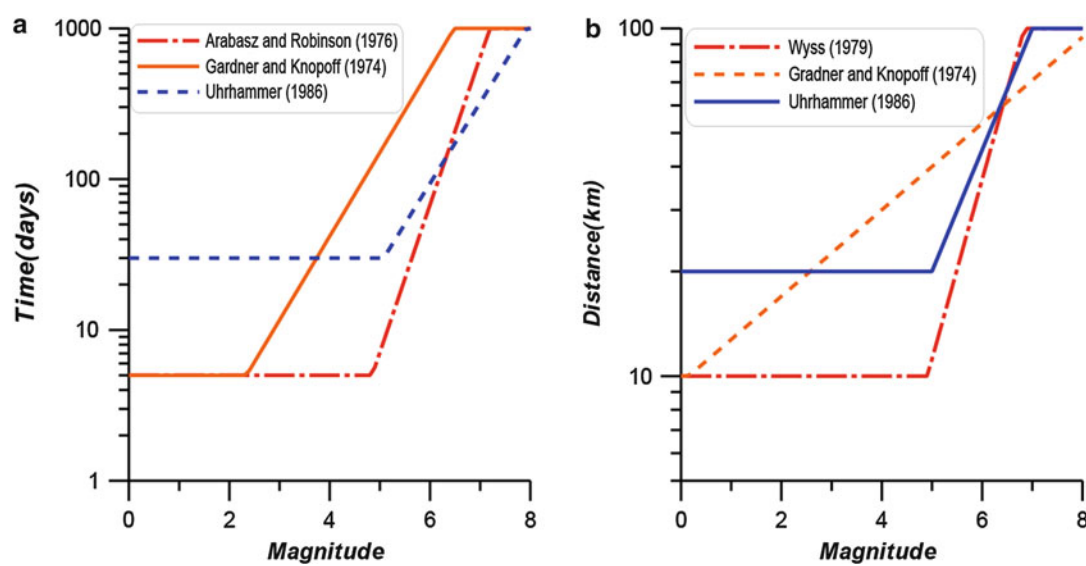
time are within the windows (Fig. 4). Earthquakes are regarded as foreshocks or aftershocks if they fulfill at least two of the four declustering approaches.

Focal Mechanisms

Based on the spatial distribution of the focal mechanisms, the seismogenic region for the PSHA can be distinguished. According to the focal mechanisms determined by Wu et al. (2010), the spatial distribution of the crustal stress state in Taiwan can be illustrated. In northern Taiwan, along the Central Range and in the Okinawa Trough, the stress states are normal favorable. In central Taiwan, southwestern Taiwan, and within the interfaces of subduction systems, the stress states are



Probability Seismic Hazard Mapping of Taiwan, Fig. 3 The distribution of earthquakes with $M \geq 2.0$ at a depth of (a) ≤ 35 km and (b) > 35 km since 1973. In the column of rake, N normal, RL right-lateral, LL left-lateral, T thrust



Probability Seismic Hazard Mapping of Taiwan, Fig. 4 The (a) distance and (b) time windows for each declustering approach. Earthquakes were considered dependent when their distance and time were within the windows

thrust favorable. In northwestern Taiwan, south Taiwan, and along the Longitudinal Valley, the stresses are strike-slip favorable. Additionally, the stress state within the two subduction systems can be comprehended using focal mechanisms determined by Wu et al. (2010).

The distribution of focal mechanisms can be associated with the tectonic setting, as mentioned above (in section “[Tectonic Setting](#)”). Representing the spatial distribution of stress states in each region would be of benefit for the PSHA in respect to the determination of seismogenic sources.

Seismogenic Sources

For application of the PSHA approach of Cornell (1968) (Eq. 2), seismogenic sources and corresponding parameters should be defined. Based on the understanding of each source, three types, Type I, Type II, and Type III (Kiureghian and Ang 1977), can be categorized. Type I is a source with a clear fault geometry, Type II is a source with a clear focal mechanism, and Type III is a source with a controversial fault geometry and mechanism. Type II sources were assumed as “regional sources.” By further considering the depth boundary of 35 km, a “shallow regional source” and a “deep regional source” were defined. Type I sources were treated as “crustal active fault sources” according to the distribution of active faults obtained by the Central Geological Survey (2010, http://fault.moeacgs.gov.tw/TaiwanFaults_2009/News/NewsView.aspx?id=3). Since two subduction systems exist in Taiwan (Fig. 1) and since the ground motion attenuation behaviors of intraslab and interface events are different, “subduction intraslab sources” and “subduction interface sources” were considered. In the following, each seismogenic source and the corresponding parameters are described in detail. For application of the logic tree in the PSHA, the weights for the parameters of each source are also required. In the following, the treatment of the weighting is described in detail.

Shallow Regional Sources

Using information on geomorphology, seismology, and geophysics, 28 shallow regional sources were defined in Taiwan and its vicinity (Fig. 5). The geometry and the corresponding parameters of each zone are outlined in the following sections.

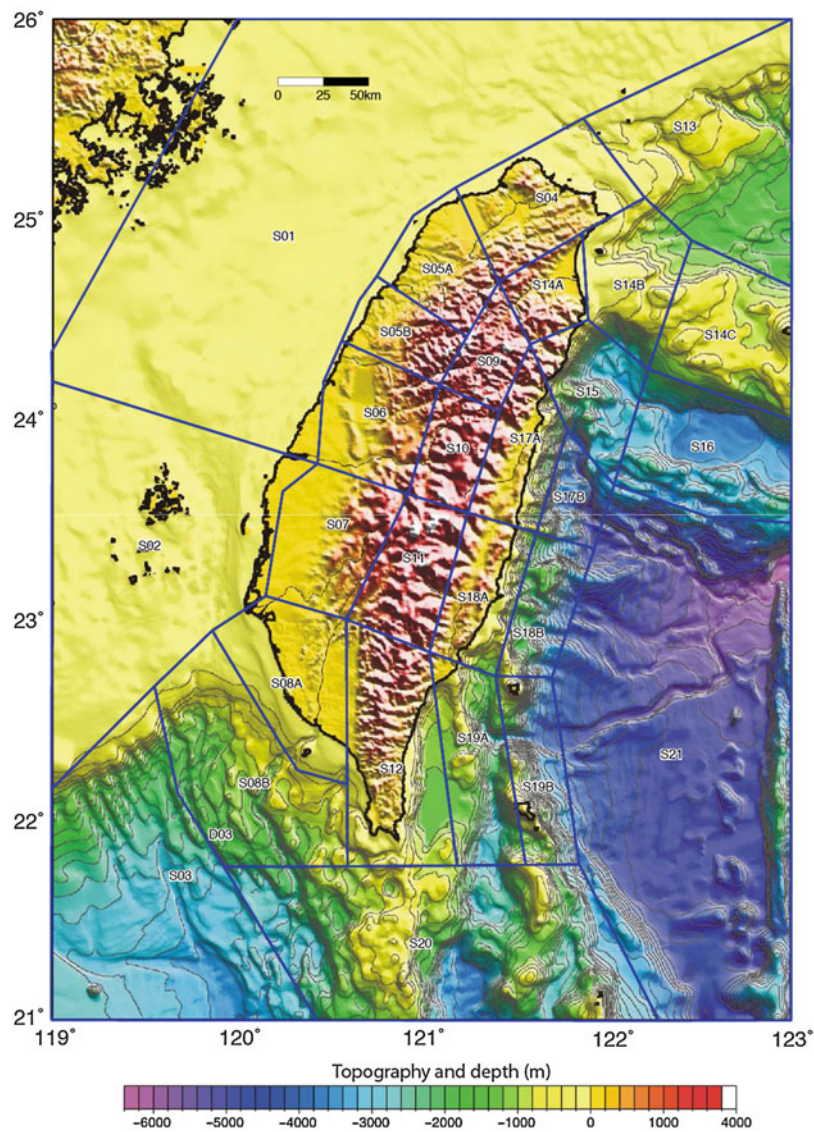
The Geometry of Each Source

S01, S02, and S03 are located in the stable Eurasian Continental Plate. In these sources, seismicity rates are relatively low in comparison to the region surrounding Taiwan Island. The boundary between S01 and S02 was determined based on the extended alignment of the structure in Taiwan. Additionally, earthquakes for the two sources present different focal mechanisms (Wu et al. 2010). The southern boundary of S02 is defined by the accretionary wedge of the southern subduction zone system. The eastern boundary of S03 is defined by the southern subduction zone system (Fig. 1).

At S04, the focal mechanisms suggest normal favorability, which is significantly different from that in its vicinity (Wu et al. 2010). S05A displays a transient mechanism from the normal mechanism in the northeast (S04) to the strike-slip and one located in the southwest (S05B). The eastern boundary of the two sources is defined based on different mechanisms from S09, where the dipping angles of earthquakes are close to vertical and mechanisms are normal favorable. S06 belongs to the frontal deformation region in the Western Foothills. Both the southern and northern boundaries are defined by fault segmentations and changes in the fault alignments (Fig. 2a). The western boundary is marked by the boarder of the Peikang High (Fig. 1). The eastern boundary is defined due to different mechanisms from S10, where the dipping angles are close to vertical and earthquakes are the normal favorable mechanism.

S07 is also located in the frontal deformation region within the Western Foothills. The eastern boundary is defined according to a significantly different seismicity rate from S11. S08A and S08B are located in the transition region between the frontal deformation region on the north and the subduction system on the south. In comparison with the thrust mechanisms in S07, earthquakes within S08A are strike-slip types with a

Probability Seismic Hazard Mapping of Taiwan, Fig. 5 The distribution of shallow regional sources



thrust component. The eastern boundary with S12 was determined according to heterogeneous deformation behaviors obtained from GPS observations (Hsu et al. 2009).

S09, S10, S11, and S12 are located within the transition region from the frontal deformation region in the west to the collision boundary between the Eurasian and Philippine Sea Plates in the east. The principal stress axis is vertical. The four sources are distinguished by strike orientations (from a NE-SW orientation in the north to a N-S orientation in the south) and seismicity rates. S13's source is located within the western

flank of the Okinawa Trough. Due to back-arc spreading, the mechanism of earthquakes in this source suggests normal favorability. S14A, S14B, and S14C result from back-arc spreading and subduction. In the three sources, the seismicity rates are high and the seismicity behaviors are complex. The southern boundary is defined according to the interface of the subduction system. The three sources are distinguished by their seismicity rates and mechanisms. S15 and S16 are located in the area where the Philippine Sea Plate subducts to the Eurasian Plate. In general, earthquakes have thrust mechanisms with low

dipping angles. In comparison, the mechanisms in S15 are complex due to the coexistence of plate collision and subduction. For the same reason, the seismicity rate in S15 is higher than that in S16.

S17A, S17B, S18A, and S18B reside along the eastern coastline and in the offshore region, which is in the collision zone between the two plates. The boundary between S17 and S18 is defined by heterogeneities in respect to active fault activity and deformation behavior according to GPS observations (Hsu et al. 2009). Additionally, seismicity rates in S17A and S18A to the west are higher than those in S17B and S18B to the east. Earthquakes in these sources are mainly thrusts with a strike-slip mechanism with high dipping angles.

S19A and S19B include Green Island and Orchid Island and their vicinity. Since 1900, eight earthquakes with $M \geq 6.0$ have taken place in these sources (Fig. 2a). The two sources are distinguished based on differences in the seismicity rate (Fig. 3a). S20 represents the southern offshore region of Taiwan. Additionally, it is above the subduction zone. S21 is located within the Philippine Sea Plate. Since it is located at a distance from the collision zone, the seismicity rate is relatively low (Fig. 1).

The Parameters of Each Source

In the above discussion, the distributions of shallow regional sources were determined. For application of the PSHA, parameters for each source were required. In the following, the acquirement of parameters including the seismicity rate models, maximum magnitudes, and their corresponding weights is described.

Due to insufficient information on fault geometry and the slip rate for regional sources, a truncated exponential model (Eq. 4) is applied for shallow regional sources. The model for each source was obtained through a regression of the declustered catalog using maximum likelihood estimates. For the regression, the magnitude interval for the model (dm in Eq. 7) was assumed to be 0.5. Therefore, the cumulative annual rate ($\dot{N}(m_0)$) and the b -value, as well as the corresponding deviations for each source, were obtained (Table 1).

Another key parameter for the PSHA was the maximum magnitude, m_u . For sources with active faults or subduction systems (i.e., S04, S05A, S05B, S06, S07, S08A, S11, S12, S15, S16, S17A, and S18A) (Fig. 2a), the corresponding m_u 's were generally obtained based on the maximum magnitude of active faults within the source (see section "Crustal Active Fault Sources"). Note that a maximum threshold of 6.5 in these sources was assumed since the occurrence of $M > 6.5$ earthquakes can be attributed to crustal active faults or subduction sources. However, m_u for some sources, S05A, S12, S15, and S16, was determined using other considerations. In S05A, since no earthquake with $M \geq 5.0$ has ever been recorded, a smaller m_u of 5.5 was assumed. In S12, S15, and S16, due to the existence of a subduction interface and other offshore active faults, a larger m_u of 7.0 was assumed. On the other hand, for sources without active faults (i.e., S01, S02, S03, S08B, S09, S10, S13, S14A, S14B, S14C, S17B, S18B, S19A, S19B, S20, and S21), the m_u 's were assumed based on the maximum observed earthquake plus 0.2 for each source (Table 1).

Deep Regional Sources

According to the distribution of seismicity and the tectonic setting, seven deep regional sources were defined in Taiwan and its vicinity (Fig. 6). Note that subduction interfaces were not included with respect to differences in ground motion attenuation behaviors (Lin and Lee 2008). The geometry and the corresponding parameters of each source are discussed in the following sections.

The Geometry of Each Source

Deep regional sources can be separated into two parts. To the east of longitude 121.5°, they are located within the subduction zone in northeast Taiwan. To the south of latitude 23°, they are regarded as another subduction zone in southern Taiwan. D01, D02, and D03 are located within the Eurasian Plate. The boundary between D02 and D03 is located along the latitude of 23°, which is regarded as the border of the continent to the north and the subduction zone to the south. The accretionary wedge of the southern

Probability Seismic Hazard Mapping of Taiwan, Table 1 The parameters and the corresponding standard deviations for the 28 shallow regional sources; the corresponding weights for the maximum magnitudes, m_u , are denoted in parentheses

NO	m_0	$N(m_0) (\pm\sigma_{N(m_0)})$	$b(\pm\sigma_b)$	m_u
S01	3.5	1.108(± 0.171)	0.865(± 0.114)	6.5(0.2) 6.6(0.6) 6.7(0.2)
S02	3.5	0.968(± 0.159)	0.772(± 0.114)	6.5(0.2) 6.6(0.6) 6.7(0.2)
S03	3.5	0.806(± 0.185)	0.852(± 0.162)	6.5(0.2) 6.6(0.6) 6.7(0.2)
S04	2.5	3.578(± 0.384)	0.783(± 0.066)	6.5(1.0)
S05A	2.5	4.498(± 0.349)	1.194(± 0.093)	5.5(1.0)
S05B	2.5	5.412(± 0.382)	0.811(± 0.053)	6.5(1.0)
S06	2.5	11.860(± 0.702)	0.816(± 0.041)	6.5(1.0)
S07	3	6.107(± 0.400)	0.643(± 0.034)	6.5(1.0)
S08A	2.5	9.990(± 0.649)	0.855(± 0.049)	6.5(1.0)
S08B	2.5	14.280(± 0.852)	1.028(± 0.051)	6.5(0.2) 6.6(0.6) 6.7(0.2)
S09	2.5	1.925(± 0.269)	0.0541(± 0.062)	6.7(0.2) 6.9(0.6) 7.1(0.2)
S10	3.5	0.935(± 0.157)	0.958(± 0.136)	6.5(0.2) 6.7(0.6) 6.9(0.2)
S11	3.5	1.543(± 0.201)	0.857(± 0.094)	6.5(1.0)
S12	2.5	23.510(± 0.966)	0.695(± 0.024)	7.0(1.0)
S13	3	5.185(± 0.497)	0.692(± 0.049)	6.3(0.2) 6.5(0.6) 6.7(0.2)
S14A	2.5	3.526(± 0.373)	0.676(± 0.058)	6.7(0.2) 6.9(0.6) 7.1(0.2)
S14B	2.5	1.523(± 0.198)	0.564(± 0.061)	7.4(0.2) 7.6(0.6) 7.8(0.2)
S14C	3.5	4.013(± 0.326)	0.720(± 0.048)	7.4(0.2) 7.6(0.6) 7.8(0.2)
S15	3.5	7.187(± 0.528)	0.629(± 0.037)	7.0(1.0)
S16	3.5	6.285(± 0.0491)	0.602(± 0.036)	7.0(1.0)
S17A	3.5	4.069(± 0.405)	0.713(± 0.055)	6.5(1.0)
S17B	3.5	1.795(± 0.268)	0.685(± 0.081)	7.3(0.2) 7.5(0.6) 7.7(0.2)
S18A	3.5	2.979(± 0.274)	0.616(± 0.048)	6.5(1.0)
S18B	3.5	2.524(± 0.253)	0.636(± 0.054)	7.3(0.2) 7.5(0.6) 7.7(0.2)
S19A	3.5	4.834(± 0.438)	0.678(± 0.049)	7.3(0.2) 7.5(0.6) 7.7(0.3)
S19B	3.5	3.017(± 0.353)	0.780(± 0.071)	7.3(0.2) 7.5(0.6) 7.7(0.3)
S20	3.5	4.866(± 0.456)	0.882(± 0.064)	7.1(0.2) 7.3(0.6) 7.5(0.2)
S21	3.5	9.581(± 0.498)	0.756(± 0.033)	7.1(0.2) 7.3(0.6) 7.5(0.2)

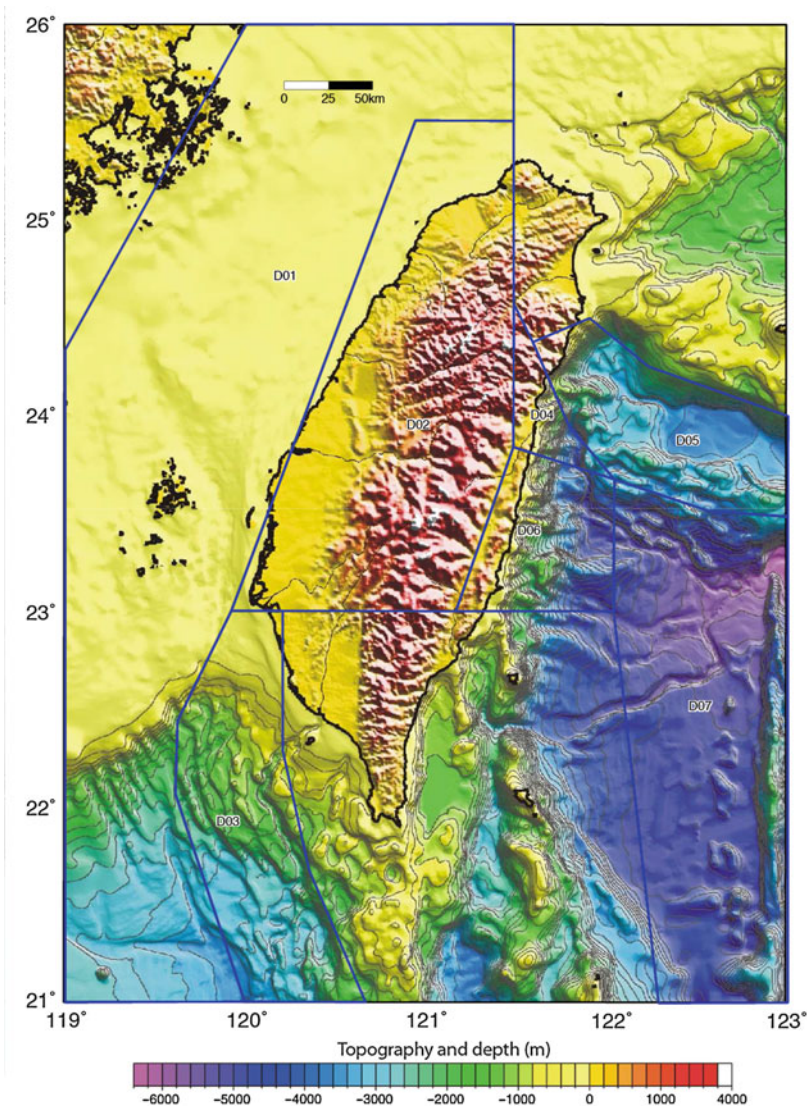
subduction zone system defines the western boundary of D03. The eastern boundary of D03 is illustrated by the distribution of the southern subduction zone system. To the east of Taiwan, four deep regional sources, including D04, D05, D06, and D07, were identified. Due to plate collision, higher seismicity rates were observed in D04 and D06 (Fig. 3b). In contrast, since it is a part of the deeper part of the Philippine Sea Plate, D07 has a lower seismicity rate. D05 is located within the region where the Philippine Sea Plate subducts to the Eurasian Plate.

The Parameters of Each Source

As a follow-up to the procedure for shallow regional sources, the acquirement of the

corresponding parameters for each deep regional source is described in the following. A truncated exponential model is used to present seismic activity. The cumulative annual rate and the b -value, as well as the corresponding deviations for each source, were obtained (Table 2). The maximum magnitude, m_u , for each source was assumed. Since D01, D02, and D03 are located within the stable Eurasian Continental Plate, a small m_u was assumed. For application of the logic tree, weights of 0.2, 0.6, and 0.2 were assumed for m_u 's of 6.5, 6.6, and 6.7, respectively. Due to the plate collision zone, higher m_u 's of 7.0, 7.2, and 7.4 were assumed for D04. Since earthquakes with $M \geq 6.7$ have never been recorded in D05, D06, and D07, m_u 's of 6.8, 7.0, and 7.2 were assumed.

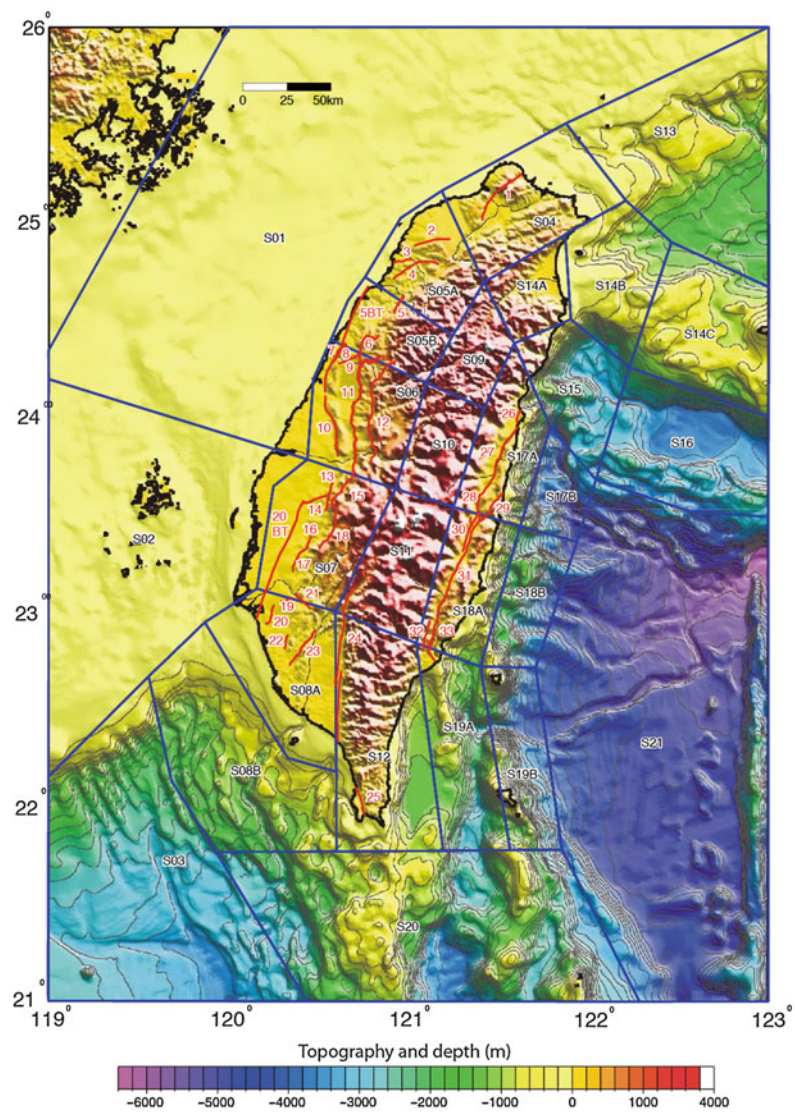
Probability Seismic Hazard Mapping of Taiwan, Fig. 6 The distribution of deep regional sources



Probability Seismic Hazard Mapping of Taiwan, Table 2 The parameters and the corresponding standard deviations for the seven deep regional sources; the corresponding weights of the maximum magnitudes, m_u , are denoted in parentheses

NO	m_0	$N(m_0) (\pm \sigma_{N(m_0)})$	$b(\pm \sigma_b)$	m_u
D01	3.5	$0.673(\pm 0.169)$	$0.837(\pm 0.162)$	6.5(0.2) 6.6(0.6) 6.7(0.2)
D02	3.0	$2.142(\pm 0.239)$	$0.799(\pm 0.078)$	6.5(0.2) 6.6(0.6) 6.7(0.2)
D03	3.0	$7.341(\pm 0.596)$	$0.750(\pm 0.055)$	6.5(0.2) 6.6(0.6) 6.7(0.2)
D04	3.5	$1.340(\pm 0.185)$	$0.668(\pm 0.077)$	7.0(0.2) 7.2(0.6) 7.4(0.2)
D05	3.5	$3.394(\pm 0.299)$	$0.669(\pm 0.052)$	6.8(0.2) 7.0(0.6) 7.2(0.2)
D06	3.5	$1.585(\pm 0.204)$	$0.665(\pm 0.072)$	6.8(0.2) 7.0(0.6) 7.2(0.2)
D07	3.5	$1.374(\pm 0.185)$	$0.554(\pm 0.070)$	6.8(0.2) 7.0(0.6) 7.2(0.2)

Probability Seismic Hazard Mapping of Taiwan, Fig. 7 The distribution of shallow regional (blue polygons) and crustal active fault sources (red lines). The corresponding shallow regional source and the *b*-values for each crustal active fault source are denoted in Table 4



Crustal Active Fault Sources

Thirty-three active faults obtained by the Central Geological Survey (2010, http://fault.moeacgs.gov.tw/TaiwanFaults_2009/News/NewsView.aspx?id=3) and three blind faults by Cheng et al. (2007) were considered as crustal active fault sources (Fig. 7). Crustal active fault sources are categorized as Type I sources by Kiureghian and Ang (1977). Sources of this type usually generate earthquakes with a characteristic magnitude repeatedly within a recurrence interval. Such seismic activity does not follow the truncated exponential model, but

fulfills the characteristic earthquake model (Wesnousky 1994). For application of this model, some parameters are required (Eqs. 6, 7, and 8). The fault parameters, including segmentation, length, depth, area, the rupture mechanism, dip angle, the possible magnitude of the slip, the slip rate, the recurrence interval, the last slip time, and the magnitude of a characteristic event, were obtained from various references and are listed in Table 3. For treatment of the *b*-value, a constant *b*-value is assumed in each shallow regional source. According to the distribution of shallow

Probability Seismic Hazard Mapping of Taiwan, Table 3 The parameters for crustal active fault sources; the corresponding weights of the slip rate and the maximum magnitude for each source are denoted in parentheses

ID	Length (km)	Depth (km)	Area (km ²)	Rake	Dip angle	Possible slip amount (m)	Slip rate (mm/year)	Recurrence interval (year)	Last occurrence time (year)	Characteristic magnitude(M _w)
1	61	0	15	1,009.6	N	>60°E	1.39 ~ 7.99	614 ~ 2,511	<11,000	7.20(0.5) 7.33(0.5)
	13	0	15	215.2	N	>60°E	0.23 ~ 1.30			
2	22	0	12	410.7	T	~40°S	0.46 ~ 3.20	117 ~ 527	<70,000	6.50(0.5) 6.64(0.5)
3	9	0	12	141	T-RL	~50°S	0.34 ~ 2.33	118 ~ 263	Late Pleistocene	5.98(0.5) 6.16(0.5)
4	28	0	12	672	T	~30°S	0.50 ~ 3.48	181 ~ 454	<300 year	6.64(0.5) 6.77(0.5)
5	12	0	12	158.8	T	>60°W	0.37 ~ 2.58	60 ~ 314	A.D.1935	6.15(0.5) 6.32(0.5)
5BT	37	2	15	962	T-RL	30°E	0.56 ~ 3.84	47 ~ 207		6.80(0.5) 6.91(0.5)
6	33	0	15	646.2	T	40~60°E	0.53 ~ 3.69	275 ~ 1,249	Holocene	6.73(0.5) 6.85(0.5)
7	8	2	15	147.1	T	40~50°E	0.46 ~ 3.20	74 ~ 362	Holocene	6.68(0.5) 6.80(0.5)
	22	2	15	404.4		40~50°E	0.32 ~ 2.24			
8	13	2	15	399.9	T	20~30°W	0.38 ~ 2.66	86 ~ 155	Holocene	6.19(0.5) 6.36(0.5)
9	14	0	15	210	RL	High angle	0.17 ~ 0.79	59 ~ 312	A.D.1935	6.23(0.5) 6.4(0.5)
10	36	2	15	936	T	18~45°E	0.55 ~ 3.81	33 ~ 119	Holocene	6.78(0.5) 6.9(0.5)
11	36	0	20	1,120.1	T	~40°E	0.55 ~ 3.81	112 ~ 280	A.D.1999	7.36(0.5) 7.43(0.5)
	48	0	20	1,493.5			0.61 ~ 4.22			
	14	0	20	435.6			0.39 ~ 2.73			
12	69	0	15	1,463.7	T	~45°E	0.69 ~ 4.79	670 ~ 2,651	A.D.1999	7.16(0.5) 7.24(0.5)
13	17	0	15	603.4	T	20~30°E	0.42 ~ 2.92	8 ~ 40	<18,540 BP	6.35(0.5) 6.6(0.5)
14	13	0	15	215.2	RL	>60°	0.15 ~ 0.72	22 ~ 117	A.D.1906	6.19(0.5) 6.36(0.5)
14BT	36	2	15	936	T	30°E	0.55 ~ 3.81	19 ~ 87		6.78(0.5) 6.90(0.5)
15	25	0	15	433	T-RL	>60°E	0.48 ~ 3.25	24 ~ 116	A.D.1999	6.57(0.5) 6.71(0.5)

16	7	0	15	154	T	~30°E	0.31 ~ 2.13	8.2(0.5) 9.3(0.5)	8 ~ 19	Late Pleistocene	5.83(0.5) 6.03(0.5)
17	17	0	15	373.9	T-LL	~30°E	0.42 ~ 2.92	8.2(0.5) 9.3(0.5)	21 ~ 40	<10,000	6.35(0.5) 6.50(0.5)
18	28	0	15	512.7	T	50-60°E	0.50 ~ 3.48	5(0.2) 10(0.6) 15(0.2)	26 ~ 121	<10,000	6.64(0.5) 6.77(0.5)
19	6	0	15	99.3	RL	>60	0.06 ~ 0.27	2.5(0.2) 5(0.6) 7.5(0.2)	12 ~ 72	A.D.1946	5.74(0.5) 5.95(0.5)
20	12	0	15	280	T	>35°W	0.37 ~ 2.58	4.0(0.5) 5.6(0.5)	23 ~ 58	Late Pleistocene	6.15(0.5) 6.32(0.5)
20BT	36	2	15	936	T	30°E	0.55 ~ 3.81	5(0.2) 10(0.6) 15(0.2)	23 ~ 104		6.78(0.5) 6.90(0.5)
21	10	0	15	165.5	LL	>60°N	0.11 ~ 3.81	1.3(0.2) 2.5(0.6) 3.8(0.2)	40 ~ 216	Mid-late Pleistocene	6.04(0.5) 6.22(0.5)
22	8	0	15	169.7	T	45°E	0.32 ~ 2.24	4.0(0.5) 5.6(0.5)	15 ~ 40	Mid-late Pleistocene	5.91(0.5) 6.10(0.5)
23	30	0	15	587.4	T	High angle	0.52 ~ 3.57	15(0.2) 3(0.6) 4.5(0.2)	85 ~ 385	7,189 BP	6.68(0.5) 6.80(0.5)
24	28	0	20	579.8	T-LL	70-80°E	0.50 ~ 3.48	2(0.2) 4(0.6) 6(0.2)	57 ~ 266	Late Pleistocene	6.64(0.5) 6.77(0.5)
	61			1,263			0.66 ~ 4.59	1.5(0.2) 3(0.6) 4.5(0.2)	26 ~ 122		7.09(0.5) 7.18(0.5)
25	39	0	20	830.1	T	70°E	0.57 ~ 3.92	4.2(0.5) 7.5(0.5)	61 ~ 159	Late Pleistocene	6.83(0.5) 6.94(0.5)
26	8	0	30	264.8	LL-T	60-70°E	0.32 ~ 2.24	10(0.2) 20(0.6) 30(0.2)	2 ~ 12	A.D.1951	5.91(0.5) 6.10(0.5)
27	30	0	30	1,174.9	LL-T	50°E	0.52 ~ 3.57	10(0.2) 20(0.6) 30(0.2)	6 ~ 29	Late Pleistocene	6.68(0.5) 6.80(0.5)
28	33	0	30	1,292.3	T-LL	40-50°E	0.53 ~ 3.69	12.5(0.5) 16.0(0.5)	13 ~ 25	A.D.1951	6.73(0.5) 6.85(0.5)
29	30	0	30	931.7	T	E	0.52 ~ 3.57	2(0.2) 4(0.6) 6(0.2)	40 ~ 183	Late Pleistocene	6.68(0.5) 6.80(0.5)
30	23	0	25	750.6	LL-T		0.47 ~ 3.25	8(0.2) 16(0.6) 24(0.2)	7 ~ 35	A.D.1951	6.52(0.5) 6.66(0.5)
31	67	0	30	2,217.8	T-LL	~65°E	0.69 ~ 4.74	26(0.5) 30(0.5)	17 ~ 26	A.D.2003	7.14(0.5) 7.23(0.5)
32	17	0	25	554.8	T	50°E	0.42 ~ 2.92	1.9(0.5) 3.0(0.5) 5.4(0.5) 5.5(0.5)	24 ~ 116	Late Pleistocene	6.35(0.5) 6.50(0.5)
33	20	0	25	551.7	T	65°E	0.45 ~ 3.09	8(0.2) 16(0.6) 24(0.2)	8 ~ 38	Late Pleistocene	6.44(0.5) 6.59(0.5)

Probability Seismic Hazard Mapping of Taiwan, Table 4 The corresponding shallow regional source and b -values for each crustal active fault source; the spatial distribution of the sources is presented in Fig. 7

Shallow regional sources	b -value	Active fault sources (ID)
S04	0.783	1
S05A	1.194	2, 3, 4,
S05B	0.811	5, 5BT, 6
S06	0.816	7, 8, 9, 10, 11, 12
S07	0.643	13, 14, 14BT, 15, 16, 17, 18, 20BT, 21
S08A	0.855	19, 20, 22, 23
S12	0.695	24, 25
S17A	0.713	26, 27, 28, 29
S18A	0.616	30, 31, 32, 33

regional sources and crustal active fault sources (Fig. 7), the corresponding b -value for each source was obtained. Table 4 summarizes the b -value for crustal active fault sources.

Subduction Interface Sources

Typical interface sources take place along plate boundaries within subduction zones. Two subduction systems are located within the vicinity of Taiwan (Fig. 1). Subduction interface sources exist in both systems (Fig. 8). In the northeast subduction system, an interface earthquake with M8.2 took place in 1920 (Tsai et al. 2000) and is the largest earthquake ever recorded in T01A. m_u 's of 8.0, 8.2, and 8.4 with corresponding weights of 0.2, 0.6, and 0.2 were assumed for this source. The interface in the southern subduction system was defined according to the profiles of earthquake distribution. Three interface sources, T02A, T02B, and T02C, were defined (Fig. 8). Segmentations were defined by the bending of the accretionary wedge. Parameters for subduction interface sources are summarized and presented in Table 5. The characteristic earthquake model provides the behaviors of the seismic activity for the sources.

Subduction Intraslab Sources

The geometries of intraslab sources for subduction systems are illustrated based on the profiles of the earthquake distribution. Ten and three

interface sources were defined for the northeastern and southern subduction zones, respectively (Fig. 8). Seismic activities of the sources were determined using the truncated exponential model. The corresponding m_u 's were inferred from the maximum magnitudes of the intraslab earthquakes surrounding Taiwan and the world. The parameters and the corresponding weights for subduction intraslab sources are summarized and presented in Table 6.

Ground Motion Prediction Equations (GMPEs)

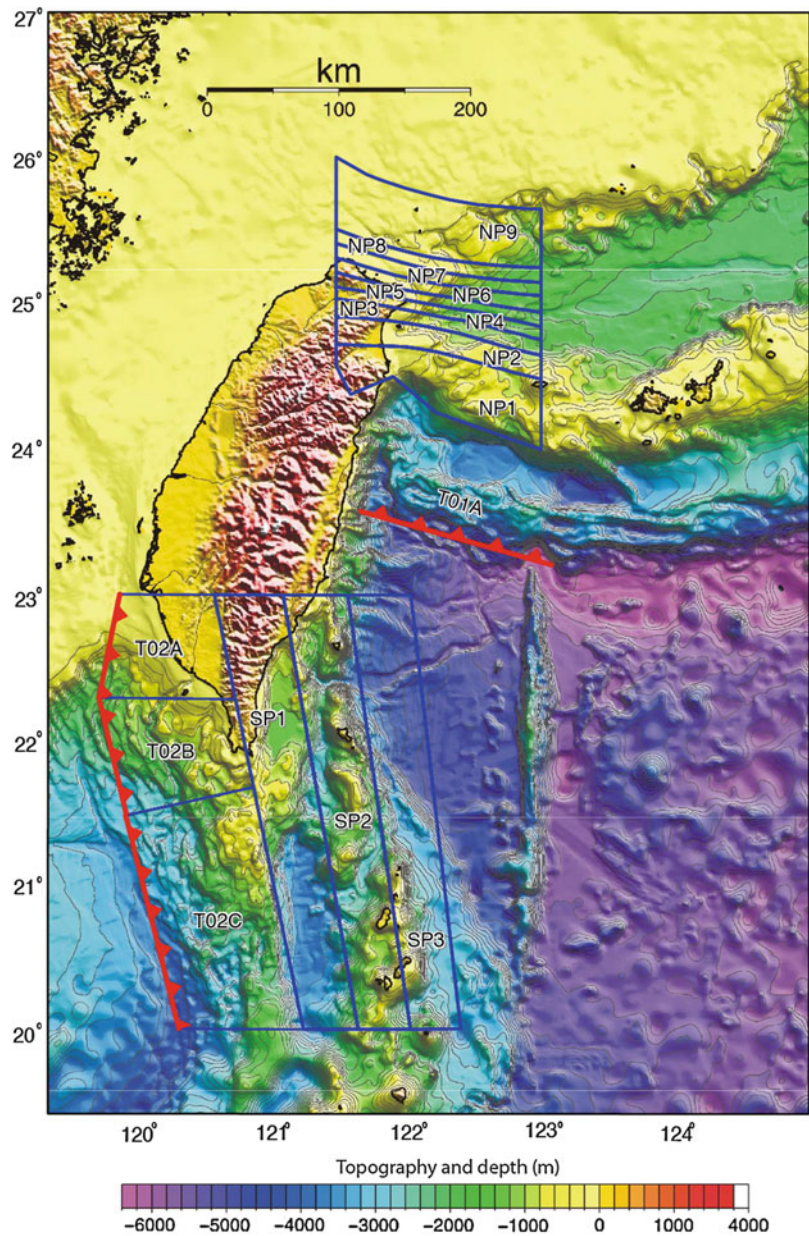
For the application of the PSHA, in addition to reliable seismogenic sources, another key factor is proper GMPEs (Encyclopedia of Earthquake Engineering: Ground Motion Prediction Equations). Lin (2009) pointed out that global GMPEs do not truly represent ground motion attenuation behaviors in Taiwan. Thus, only GMPEs obtained from the regression of strong ground motion observations in Taiwan were considered. Several studies have proposed GMPEs in order to model the attenuation behaviors for different types of earthquakes in Taiwan. For a more complete presence of attenuation behaviors, GMPEs with the form of an acceleration response spectrum (SA) are expected. In the form of SA, attenuation behaviors are presented as a response acceleration as a function of the response period. Lin (2009) considered 5,968 observations of 60 crustal earthquakes recorded by the Taiwan Strong Motion Instrumentation Program (TSMIP). GMPEs as a function of magnitude, m , are determined as follows:

$$\begin{aligned} \ln y = & C_1 + F_1 + C_3(8.5 - m)^2 \\ & + (C_4 + C_5(m - 6.3)) \ln \left(\sqrt{R^2 + \exp(H)^2} \right) \\ & + C_6 F_{NM} + C_7 F_{RV} + C_8 \ln(V_{S30}/1, 130.0), \end{aligned} \quad (10)$$

where y represents the response acceleration for PGA or SA in g ; R represents the shortest distance to the rupture surface in kilometers; F_{NM} is 1.0

Probability Seismic Hazard Mapping of Taiwan, Fig. 8

The distribution of the subduction interface and the intraslab sources illustrated by blue polygons. The surface alignments of the two trenches are presented in red



when the sources are normal mechanisms; F_{RV} is 1.0 when the sources are thrust mechanisms; V_{S30} represents the average shear-wave velocity from the ground surface down to a depth of 30 m; $F_1 = C_2(M_W - 6.3)$ for $M_W \leq 6.3$, whereas $F_1 = (-H \cdot C_5)(M_W - 6.3)$ for $M_W > 6.3$; and C_1 to C_8 and H are constants. The corresponding parameters for PGA and SA are presented in Table 7. Note that through these GMPEs, source

effects in the form of different focal mechanisms were considered and indicated that the largest amplitudes for thrust events and the smallest for normal ones were obtained based on the same magnitude and distance to the rupture surface. In order to present the reliability of these GMPEs, the corresponding standard deviations were analyzed and are presented as σ_{lny} in Table 8. These GMPEs were applied for crustal sources.

Probability Seismic Hazard Mapping of Taiwan, Table 5 The parameters for five of the subduction interface sources; the corresponding weights for the slip rate, dip angle, and maximum magnitudes, m_u , are denoted in parentheses

NO	m_0	Slip rate (mm/year)	Dip angle ($^{\circ}$)	Depth (km)	b -value	m_u
T01A	6.5	20(0.2)	18(0.3)	0–40	0.6	8.0(0.2)
		30(0.6)	20(0.4)			8.2(0.6)
		40(0.2)	22(0.3)			8.4(0.2)
T02A	6.5	5(0.2)	18(0.3)	0–35	0.7	7.6(0.2)
		10(0.6)	20(0.4)			7.8(0.6)
		15(0.2)	22(0.3)			8.0(0.2)
T02B	6.5	5(0.2)	18(0.3)	0–35	0.7	7.7(0.2)
		10(0.6)	20(0.4)			7.9(0.6)
		15(0.2)	22(0.3)			8.1(0.2)
T02C	6.5	5(0.2)	18(0.3)	0–35	0.7	7.9(0.2)
		10(0.6)	20(0.4)			8.1(0.6)
		15(0.2)	22(0.3)			8.3(0.2)

Probability Seismic Hazard Mapping of Taiwan, Table 6 The parameters and the corresponding standard deviations for 13 of the subduction intraslab sources; the corresponding weights of the maximum magnitudes, m_u , are denoted in parentheses

NO	m_o	$N(m_o) (\pm \sigma_{N(m_o)})$	$b(\pm \sigma_b)$	m_u
NP1	4.0	3.735(± 0.307)	0.908(± 0.063)	7.5(0.2) 7.7(0.6) 7.9(0.2)
NP2	4.0	2.335(± 0.247)	0.801(± 0.070)	7.5(0.2) 7.7(0.6) 7.9(0.2)
NP3	4.0	1.197(± 0.177)	0.866(± 0.104)	7.5(0.2) 7.7(0.6) 7.9(0.2)
NP4	4.0	0.580(± 0.119)	0.730(± 0.124)	7.6(0.2) 7.8(0.6) 8.0(0.2)
NP5	4.0	0.210(± 0.074)	0.938(± 0.297)	7.6(0.2) 7.8(0.6) 8.0(0.2)
NP6	4.0	0.263(± 0.083)	0.959(± 0.272)	7.6(0.2) 7.8(0.6) 8.0(0.2)
NP7	4.0	0.344(± 0.091)	0.730(± 0.173)	7.6(0.2) 7.8(0.6) 8.0(0.2)
NP8	4.0	0.374(± 0.098)	1.040(± 0.237)	7.6(0.2) 7.8(0.6) 8.0(0.2)
NP9	4.0	0.913(± 0.155)	0.913(± 0.138)	7.6(0.2) 7.8(0.6) 8.0(0.2)
SP1	4.0	1.041(± 0.165)	0.762(± 0.106)	7.5(0.2) 7.7(0.6) 7.9(0.2)
SP2	4.0	2.735(± 0.268)	0.831(± 0.068)	7.6(0.2) 7.8(0.6) 8.0(0.2)
SP3	4.0	1.660(± 0.209)	0.876(± 0.090)	7.6(0.2) 7.8(0.6) 8.0(0.2)

Many studies (e.g., Chan et al. 2013a) have indicated that ground motion attenuation behaviors for subduction earthquakes are different than those for crustal earthquakes. Since two subduction systems are located within the Taiwan region, as well as its vicinity (Fig. 1), the different GMPEs for both interfaces and intraslab events are considered to apply PSHAs. Lin and Lee (2008) proposed the first GMPEs for subduction events in Taiwan. In order to establish the GMPEs, they analyzed the strong motion records of subduction earthquakes obtained using TSMIP arrays and the Strong Motion Array in Taiwan Phase I

(SMARTI). The GMPEs of Lin and Lee (2008) are represented as follows:

$$\ln y = C_1 + C_2 m + C_3 (R + C_4 e^{C_5 m}) + C_6 H + C_7 Z_t, \quad (11)$$

where y represents the response acceleration for PGA or SA in g , R represents the hypocentral distance in kilometers, H represents the focal depth in kilometers, Z_t represents the subduction zone earthquake type ($Z_t = 0$ for interface earthquakes, whereas $Z_t = 1$ for intraslab earthquakes), and C_1 to C_7 are constants. The parameters and their corresponding standard

Probability Seismic Hazard Mapping of Taiwan, Table 7 The corresponding parameters and the standard deviations (σ_{Iny}) of the GMPEs for the PGA and each response period for crustal events (Eq. 8) as proposed by Lin (2009)

Period (s)	C ₁	C ₂	C ₃	C ₄	C ₅	H	C ₆	C ₇	C ₈	σ_{Iny}
PGA	1.0109	0.3822	0.0000	-1.1634	0.1722	1.5184	-0.1907	0.1322	-0.4741	0.627
0.01	1.0209	0.3822	-0.0003	-1.1633	0.1722	1.5184	-0.1922	0.1314	-0.4738	0.627
0.02	1.0416	0.3822	0.0017	-1.1668	0.1722	1.5184	-0.1942	0.1311	-0.4700	0.627
0.03	1.1961	0.3822	0.0038	-1.2028	0.1722	1.5184	-0.1990	0.1314	-0.4741	0.640
0.04	1.3834	0.3822	0.0087	-1.2499	0.1722	1.5184	-0.1959	0.1362	-0.4806	0.655
0.05	1.5612	0.3822	0.0153	-1.2957	0.1722	1.5184	-0.1922	0.1417	-0.4911	0.670
0.06	1.6907	0.3822	0.0210	-1.3218	0.1722	1.5184	-0.1984	0.1500	-0.4900	0.681
0.07	1.7673	0.3822	0.0261	-1.3336	0.1722	1.5184	-0.2011	0.1557	-0.4920	0.691
0.08	1.8689	0.3822	0.0273	-1.3440	0.1722	1.5184	-0.1947	0.1627	-0.4944	0.699
0.09	1.9430	0.3822	0.0276	-1.3435	0.1722	1.5184	-0.2011	0.1589	-0.4910	0.700
0.10	2.0218	0.3822	0.0254	-1.3409	0.1722	1.5184	-0.1817	0.1607	-0.4825	0.705
0.15	2.0521	0.3822	0.0100	-1.2578	0.1722	1.5184	-0.1851	0.1212	-0.4804	0.691
0.20	2.0333	0.3822	-0.0091	-1.1769	0.1722	1.5184	-0.2265	0.0999	-0.4350	0.676
0.25	1.9887	0.3822	-0.0293	-1.1153	0.1722	1.5184	-0.2355	0.0994	-0.4101	0.679
0.30	1.8827	0.3822	-0.0459	-1.0726	0.1722	1.5184	-0.2163	0.1036	-0.4361	0.686
0.35	1.7459	0.3822	-0.0600	-1.0307	0.1722	1.5184	-0.1949	0.1029	-0.4507	0.692
0.40	1.6821	0.3822	-0.0737	-1.0116	0.1722	1.5184	-0.1955	0.1099	-0.4734	0.695
0.45	1.6139	0.3822	-0.0861	-0.9939	0.1722	1.5184	-0.2011	0.1178	-0.4927	0.699
0.50	1.5288	0.3822	-0.0960	-0.9755	0.1722	1.5184	-0.2089	0.1142	-0.5035	0.699
0.60	1.3081	0.3822	-0.1133	-0.9407	0.1722	1.5184	-0.2212	0.1016	-0.5546	0.704
0.70	1.1383	0.3822	-0.1292	-0.9193	0.1722	1.5184	-0.1900	0.1036	-0.6037	0.710
0.80	1.0757	0.3822	-0.1442	-0.9167	0.1722	1.5184	-0.1865	0.1058	-0.6319	0.718
0.90	0.9935	0.3822	-0.1577	-0.9104	0.1722	1.5184	-0.1643	0.1165	-0.6577	0.723
1.00	0.8642	0.3822	-0.1687	-0.9001	0.1722	1.5184	-0.1505	0.1372	-0.6916	0.728
1.50	0.3150	0.3822	-0.2006	-0.8696	0.1722	1.5184	-0.0377	0.1572	-0.7582	0.738
2.00	-0.1760	0.3822	-0.2190	-0.8328	0.1722	1.5184	0.0780	0.1660	-0.7863	0.726
2.50	-0.4103	0.3822	-0.2319	-0.8415	0.1722	1.5184	0.0907	0.1648	-0.7939	0.709
3.00	-0.5019	0.3822	-0.2431	-0.8684	0.1722	1.5184	0.1195	0.1790	-0.7754	0.707
3.50	-0.7206	0.3822	-0.2479	-0.8689	0.1722	1.5184	0.1206	0.1629	-0.7673	0.708
4.00	-0.9383	0.3822	-0.2493	-0.8618	0.1722	1.5184	0.1267	0.1262	-0.7457	0.707
4.40	-1.0405	0.3822	-0.2559	-0.8472	0.1722	1.5184	0.1655	0.1486	-0.7042	0.717
5.00	-1.3694	0.3822	-0.2535	-0.8287	0.1722	1.5184	0.2208	0.1648	-0.6955	0.715

deviations for PGA and SA are presented in Table 8 and indicate the largest amplitude for intraslab events than interface ones when the magnitude and the hypocentral distance are the same. Since the GMPEs of Lin and Lee (2008) are among the only GMPEs from the regression of ground motion observations in Taiwan, they were applied to this study.

Seismic Hazards in Taiwan

The Hazard Maps

By considering all of the seismogenic sources, the corresponding parameters, and the GMPEs,

a PSHA was applied for Taiwan. The spatial distributions of seismic hazards were evaluated in the form of a seismic hazard map. The seismic hazards for 10 % (475-year return period) and 2 % (2,475-year return period) of the probability of exceedance for 50 years are presented (Fig. 9). The results suggest that seismic hazards in Taiwan are mainly contributed by active fault sources. The highest seismic hazard levels were determined from the Western Foothills to the Coastal Plain. A >1.0 g hazard for a 10 % probability of exceedance for 50 years was estimated near the Chiayi blind fault and the Tachienshan Fault (Fig. 9a). A high hazard was also evaluated along the eastern coastline and can be attributed

Probability Seismic Hazard Mapping of Taiwan, Table 8 The corresponding parameters and the standard deviations ($\sigma_{\ln y}$) of the GMPEs for the PGA and each response period for subduction events (Eq. 9) as proposed by Lin and Lee (2008)

Period	C1	C2	C3	C4	C5	C6	C7	$\sigma_{\ln y}$
PGA	-0.9000	1.0000	-1.9000	0.9918	0.5263	0.0040	0.3100	0.6277
0.010	-2.2000	1.0850	-1.7500	0.9918	0.5263	0.0040	0.3100	0.5800
0.020	-2.2900	1.0850	-1.7300	0.9918	0.5263	0.0040	0.3100	0.5730
0.030	-2.3400	1.0950	-1.7200	0.9918	0.5263	0.0040	0.3100	0.5774
0.040	-2.2150	1.0900	-1.7300	0.9918	0.5263	0.0040	0.3100	0.5808
0.050	-1.8950	1.0550	-1.7550	0.9918	0.5263	0.0040	0.3100	0.5937
0.060	-1.1100	1.0100	-1.8360	0.9918	0.5263	0.0040	0.3100	0.6123
0.090	-0.2100	0.9450	-1.8900	0.9918	0.5263	0.0040	0.3100	0.6481
0.100	-0.0500	0.9200	-1.8800	0.9918	0.5263	0.0040	0.3100	0.6535
0.120	0.0550	0.9350	-1.8950	0.9918	0.5263	0.0040	0.3100	0.6585
0.150	-0.0400	0.9550	-1.8800	0.9918	0.5263	0.0040	0.3100	0.6595
0.170	-0.3400	1.0200	-1.8850	0.9918	0.5263	0.0040	0.3100	0.6680
0.200	-0.8000	1.0450	-1.8200	0.9918	0.5263	0.0040	0.3100	0.6565
0.240	-1.5750	1.1200	-1.7550	0.9918	0.5263	0.0040	0.3100	0.6465
0.300	-3.0100	1.3150	-1.6950	0.9918	0.5263	0.0040	0.3100	0.6661
0.360	-3.6800	1.3800	-1.6600	0.9918	0.5263	0.0040	0.3100	0.6876
0.400	-4.2500	1.4150	-1.6000	0.9918	0.5263	0.0040	0.3100	0.7002
0.460	-4.7300	1.4300	-1.5450	0.9918	0.5263	0.0040	0.3100	0.7092
0.500	-5.2200	1.4550	-1.4900	0.9918	0.5263	0.0040	0.3100	0.7122
0.600	-5.7000	1.4700	-1.4450	0.9918	0.5263	0.0040	0.3100	0.7280
0.750	-6.4500	1.5000	-1.3800	0.9918	0.5263	0.0040	0.3100	0.7752
0.850	-7.2500	1.5650	-1.3250	0.9918	0.5263	0.0040	0.3100	0.7931
1.000	-8.1500	1.6050	-1.2350	0.9918	0.5263	0.0040	0.3100	0.8158
1.500	-10.3000	1.8000	-1.1650	0.9918	0.5263	0.0040	0.3100	0.8356
2.000	-11.6200	1.8600	-1.0700	0.9918	0.5263	0.0040	0.3100	0.8474
3.000	-12.6300	1.8900	-1.0600	0.9918	0.5263	0.0040	0.3100	0.8367
4.000	-13.4200	1.8700	-0.9900	0.9918	0.5263	0.0040	0.3100	0.7937
5.000	-13.7500	1.8350	-0.9750	0.9918	0.5263	0.0040	0.3100	0.7468

to crustal active fault sources with short recurrence intervals along the Longitudinal Valley (Fault ID 26–33 in Table 3). In contrast, in northern Taiwan, a low hazard level was obtained. Although the Sanchiao Fault is located in this region, the recurrence intervals of the fault are rather long (in between 614 and 2,511 years).

The Hazard Curves

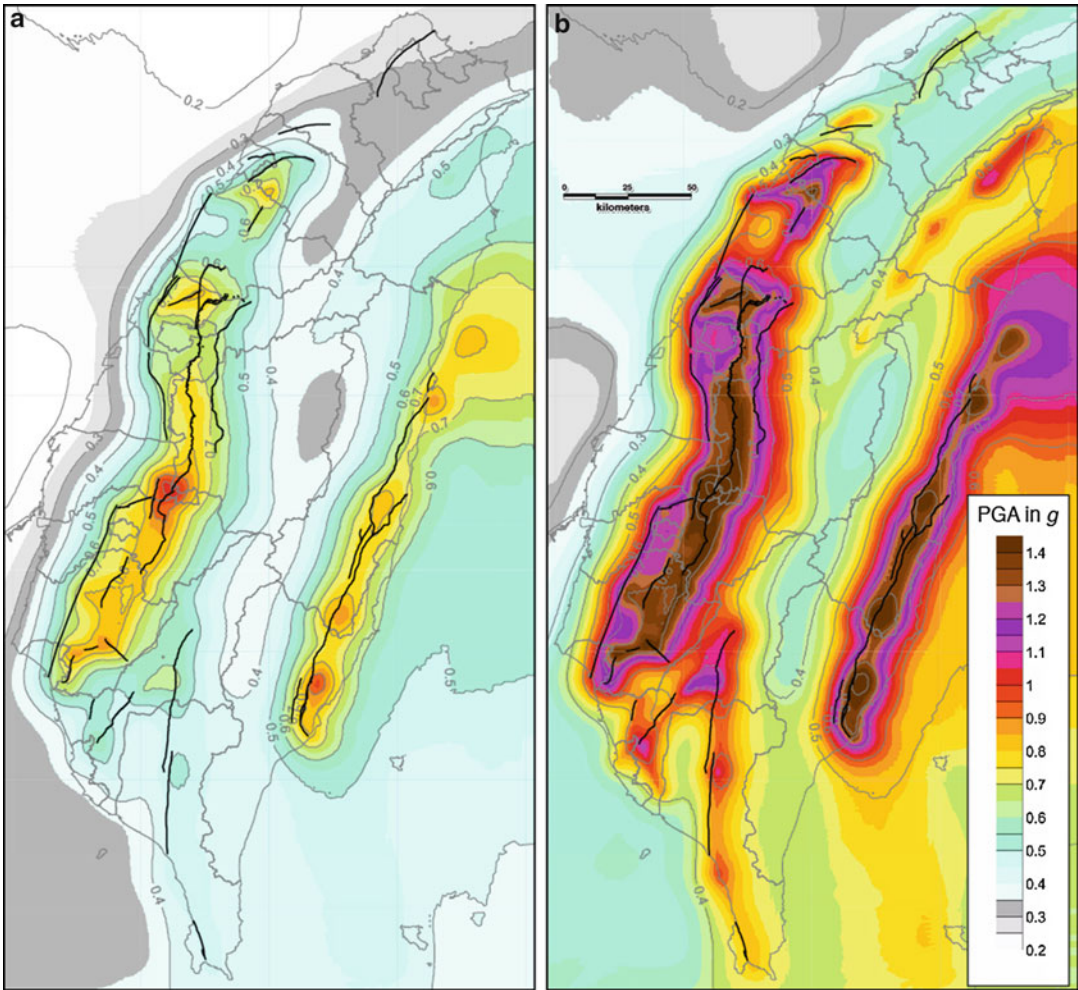
In order to represent the probability of annual exceedance as a function of ground motion level for different response periods, hazard curves are presented. The hazard curves were evaluated for the six municipalities, i.e., the Taipei, New Taipei, Taoyuan, Taichung, Tainan, and Kaohsiung

Cities, respectively (Fig. 10). The hazard curves for PGA and the response period of 0.3 and 1.0 s are obtained. Since some active faults are close to the Tainan and Taichung Cities (Fig. 9), higher hazards were determined. By contrast, the hazards in the Taipei, New Taipei, Taoyuan, and Kaohsiung cities are relatively low because of they are further away from crustal active fault sources with short recurrence intervals.

Discussion and Conclusions

A Comparison to Other PSHA Results

Probabilistic seismic hazards imparted by different seismogenic sources for the Taiwan region



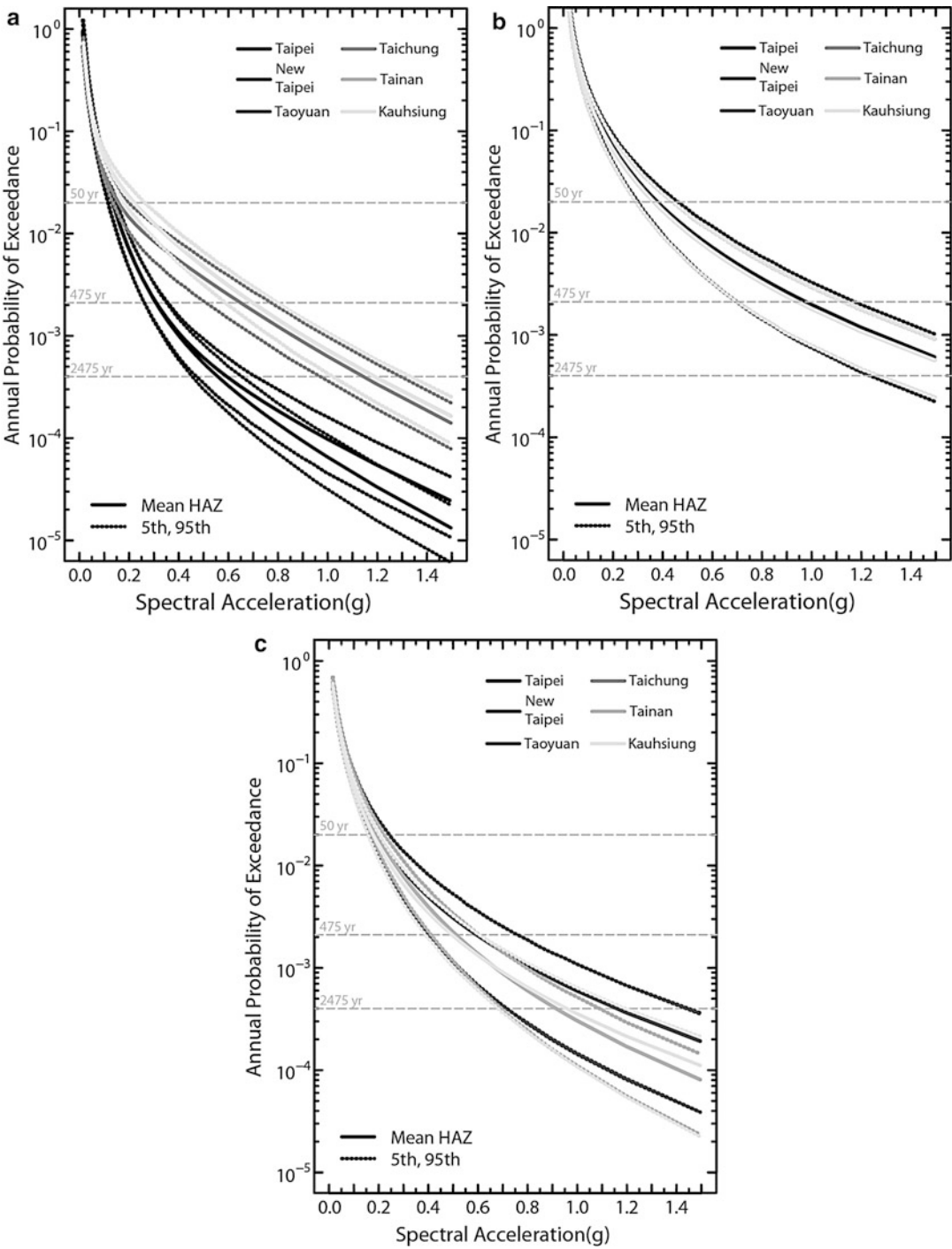
Probability Seismic Hazard Mapping of Taiwan, Fig. 9 The seismic hazard maps in (a) a 10 % (a 475-year return period) and (b) a 2 % (a 2,475-year return period) probability of exceedance for 50 years

were assessed. Higher hazards were determined along active faults in the Coastal Plain and the Longitudinal Valley. The pattern can be associated with the distribution of seismic activity in Taiwan. The pattern is similar to those obtained from previous studies (e.g., Cheng et al. 2007). In more detail, hazard peaks were evaluated near crustal active fault sources with short recurrence intervals. The results were similar to those obtained from other PSHA studies (e.g., Campbell et al. 2002; Cheng et al. 2007), which also considered the hazards imparted by active faults. The pattern can be attributed to the behaviors of GMPEs (Lin 2009), illustrating the significance

of larger ground shakings surrounding regions close to earthquake epicenters or rupture faults. The attenuation behavior is similar to observations of the 1999 Chi-Chi, Taiwan, earthquake. The shake map of this earthquake indicated significant high ground shakings in the form of PGA in the vicinity of the Chelungpu Fault, which ruptured during the coseismic period.

The Importance of the Short-Term Probabilistic Seismic Hazard Assessment

Recently, the concept of a traditional PSHA has been called into question. One of the standard procedures for traditional PSHAs is the



Probability Seismic Hazard Mapping of Taiwan, Fig. 10 The probability of annual exceedance as a function of hazard curves for PGA and the response period of

0.3 and 1.0 s for the six municipalities. The *solid* and *dashed lines* represent mean and two standard deviations of seismic hazards, respectively

construction of return periods for characteristic earthquakes that are independent of one another (see section “[The Declustering Process](#)” for details). Thus, an earthquake should be declustered prior to its incorporation for a PSHA. However, some recent cases have determined disadvantages for traditional PSAs. For example, the M_w 6.3, 21 February 2011, Christchurch earthquake can be regarded as an aftershock of the M_w 7.1 Darfield earthquake that took place on 4 September 2010. The Christchurch earthquake caused severe damage in downtown Christchurch due to a closer epicentral distance. Additionally, not only can aftershocks lead to devastating seismic hazards, but the next earthquake could expand hazards. An earthquake with a M_w 7.4 took place off the Pacific coast of Tohoku, Japan, on 9 March 2011. Due to its distance from urban regions, the resulting damage from this earthquake was negligible. Fifty-one hours after the earthquake (on March 11), a M_w 9.1 earthquake took place in the vicinity and resulted in disasters in Japan. Seismic hazards may also result from several events within an earthquake sequence. Based on spatial and temporal relationships, the 1904 Toulou and the 1906 Yanshuigang earthquakes can be regarded as a foreshock and an aftershock of the 1906 Meishan earthquake (Fig. 2), respectively. However, all three earthquakes caused severe disasters. The case indicates the importance of the seismic hazards imparted by each event in an earthquake sequence and raises the unsuitability of traditional PSAs (Chan et al. 2013b). Such instances indicate the importance of an earthquake sequence in respect to seismic hazard evaluations and suggest the reevaluation of seismic hazards immediately following large earthquakes.

Several studies have recognized disadvantages in traditional PSAs and have considered a time dependency for PSAs. For example, Chan and Wu (2012) and references therein estimated the temporal evolution of the seismicity rate and assessed seismicity hazards as a function of time. The feasibility of this approach has been tested using applications to several cases in Taiwan (e.g., Chan et al. 2013a, b).

The Applicability of PSHA Results and Seismogenic Parameters

An assessment of seismic hazards had been widely applied for each administrative region in Taiwan. In addition, hazard curves for the six municipalities were evaluated. According to the results, the probability of annual exceedance as a function of ground motion level for different response periods was presented. They would be valuable to lawmakers for determining building codes and to decision-makers for the site selection of public structures.

In addition, based on the PSHA results, further applications in respect to seismic hazard mitigation can be assessed (e.g., probabilistic seismic risk assessment). A risk assessment is presented as the probability that a given loss of property and human life exceeds a specific value during a time period. For its application, seismic hazards, exposure (distribution of population or construction), and a corresponding vulnerability are incorporated. The result could be of benefit to decision-makers for territorial planning.

The parameters implemented for the PSHA were acquired through the integration of state-of-the-art information regarding the tectonic setting, geology, geomorphology, earthquake catalog, geophysics, and the ground motion attenuation. The information could be used as a reference for an additional PSHA. For example, these parameters will be regarded as a branch in the logic tree approach for the Committee of Taiwan Earthquake Model (TEM, <http://tec.earth.sinica.edu.tw/TEM>). TEM is working on a seismic hazard map for the entire region of Taiwan. In the meantime, the work can also be incorporated within the Global Earthquake Model (see entry “The Global Earthquake Model (GEM)”), integrated by several research institutes, insurance, and reinsurance enterprises in order to evaluate seismic hazards around the world. Through collaboration with organizations on different scales and with different individuals, the GEM is expected to establish uniform and open standards for calculating and communicating earthquake hazards and risks worldwide.

Summary

A probabilistic seismic hazard assessment is implemented to the region of Taiwan. Seismogenic sources, which may result in seismic hazards, were divided into the following categories: shallow-crust regional sources, deep-crust regional sources, crustal active fault sources, subduction intraslab sources, and subduction interface sources. By further considering ground motion prediction equations for different types of sources and site conditions, hazard maps in Taiwan and hazard curves in the six municipalities were assessed.

Cross-References

► Site Response for Seismic Hazard Assessment

References

- Arabasz WJ, Robinson R (1976) Microseismicity and geologic structure in the northern south island, New Zealand. *J Geol Geophys* 19:569–601
- Campbell KW, Thenhaus PC, Barnard TP, Hampson DB (2002) Seismic hazard model for loss estimation and risk management in Taiwan. *Soil Dyn Earthq Eng* 22:743–754. doi:10.1016/S0267-7261(02)00095-7
- Chan CH, Wu YM (2012) The burst of seismicity after the 2010 M6.4 Jiashian earthquake and its implication of short-term seismic hazard in southern Taiwan. *J Asian Earth Sci* 51:231–239. doi:10.1016/j.jseae.2012.08.011
- Chan CH, Wu YM, Cheng CT, Lin PS, Wu YC (2013a) Time-dependent probabilistic seismic hazard assessment and its application to Hualien City, Taiwan. *Nat Hazards Earth Syst Sci* 13:1–16. doi:10.5194/nhess-13-1-2013
- Chan CH, Wu YM, Cheng CT, Lin PS, Wu YC (2013b) Scenario for a short-term probabilistic seismic hazards assessment (PSHA) in Chiayi, Taiwan. *Terr Atmos Ocean Sci* 24(4, Part II):671–683. doi:10.3319/TAO.2013.01.22.01(T)
- Chen C-H, Wang J-P, Yih-Min W, Chan C-H, Chang C-H (2012) A study of earthquake inter-occurrence times distribution models in Taiwan. *Nat Hazards*. doi:10.1007/s11069-012-0496-7
- Cheng CT, Chiou SJ, Lee CT, Tsai YB (2007) Study on probabilistic seismic hazard maps of Taiwan after Chi-Chi earthquake. *J Geo Eng* 2:19–28
- Cornell CA (1968) Engineering seismic risk analysis. *Bull Seismol Soc Am* 58(5):1583–1606
- Gardner JK, Knopoff L (1974) Is the sequence of earthquakes in southern California, with aftershocks removed, poissonian? *Bull Seismol Soc Am* 64:1363–1367
- Gutenberg B, Richter C (1954) Seismicity of the earth and associated phenomena, 2nd edn. Princeton University Press, Princeton, 310
- Hanks TC, Kanamori H (1979) A moment magnitude scale. *J Geophys Res* 84:2348–2350
- Hsu Y-J, Yu S-B, Simons M, Kuo L-C, Chen H-Y (2009) Interseismic crustal deformation in the Taiwan plate boundary zone revealed by GPS observations, seismicity, and earthquake focal mechanisms. *Tectonophysics* 479:4–18
- Kiureghian AD, Ang AH-S (1977) A fault-rupture model for seismic risk analysis. *Bull Seismol Soc Am* 67:1173–1194
- Kramer SL (1996) Geotechnical earthquake engineering. Prentice Hall, Upper Saddle River, 653 pp
- Lin PS (2009) Ground-motion attenuation relationship and path-effect study using Taiwan Data set. PhD dissertation, Institute of Geophysics, National Central University, Chung-Li, Taiwan (in Chinese)
- Lin PS, Lee CT (2008) Ground-motion attenuation relationships for subduction-zone earthquakes in north-eastern Taiwan. *Bull Seismol Soc Am* 98(1):220–240. doi:10.1785/0120060002
- Tsai YB, Wen KL, Chen KB, Kuo JY (2000) Summary of development of the Taiwan earthquake catalog and strong ground motion attenuation equations. Report for the National Project on the Disaster Prevention, 79 pp (in Chinese)
- Uhrhammer RA (1986) Characteristic of northern and central California seismicity. *Earthquake Notes* 57(1):21 (Abstract)
- Wesnousky SG (1994) The Gutenberg-Richter or characteristic earthquake distribution, which is it? *Bull Seismol Soc Am* 84:1940–1959
- Wu YM, Hsu YJ, Chang CH, Teng LS, Nakamura M (2010) Temporal and spatial variation of stress field in Taiwan from 1991 to 2007: insights from comprehensive first motion focal mechanism catalog. *Earth Planet Sci Lett* 298:306–316. doi:10.1016/j.epsl.2010.07.047
- Wyss M (1979) Estimating maximum expectable magnitude of earthquakes from fault dimensions. *Geology* 7:336–340
- Youngs RR, Coppersmith KJ (1985) Implications of fault slip rates and earthquake recurrence models to probabilistic seismic hazard estimates. *Bull Seismol Soc Am* 75:939–964

R

Radiocarbon Dating in Paleoseismology

Christopher Bronk Ramsey
Oxford Radiocarbon Accelerator Unit,
University of Oxford, Oxford, UK

Synonyms

Bayesian statistics: age modelling; Chronology;
Radiocarbon calibration

Introduction

Paleoseismology needs a chronological framework for past seismic events. Dating information can come from a variety of different sources, but primarily comes from stratigraphy combined with scientific dating techniques applied either to the seismic record or associated through major environmental signals (e.g., glacial-interglacial cycles).

For timescales spanning the last 50,000 years, radiocarbon dating is one of the most widely used dating techniques. It can be applied to organic material, biominerals (such as shell and coral), and in some cases to carbonate deposits and is therefore possible in a wide range of different environments. It is complementary to other techniques such as optically stimulated luminescence

dating (OSL), uranium-series dating, and argon-argon dating, relevant to different material types. Unlike those techniques, however, radiocarbon dating is a method that requires calibration against known-age samples dated by a different method. One implication of this is that the correct application of radiocarbon dating in seismology, as in other fields, often requires an understanding of the relevant parts of the carbon cycle. As with any dating technique, accurate interpretation of the data always relies on a careful examination of the relationship between dated samples and the events under investigation.

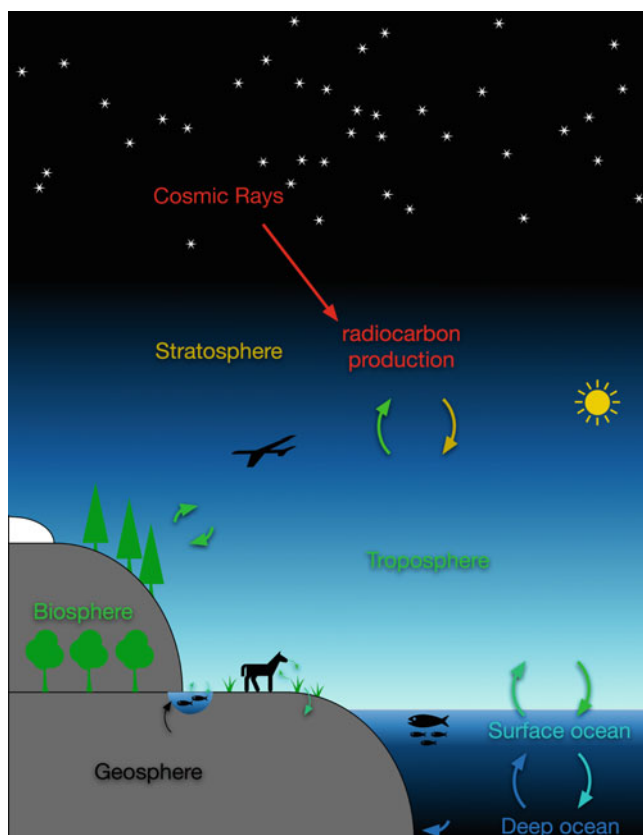
One tool that has become increasingly useful for the combination of information from stratigraphy and scientific dating techniques is Bayesian age modelling. These methods are most often applied with radiocarbon dating, but are also relevant to other dating techniques and are sufficiently flexible to be used with multiple methods. These Bayesian models are ideally suited to the generation of quantitative estimates of dates and intervals in paleoseismology.

Radiocarbon Dating

Radiocarbon dating (Libby et al. 1949) is a method that covers about the last 50,000 years. Primarily applicable to organic material, in principle, it can be applied to any material that draws its carbon directly, or indirectly, from the

Radiocarbon Dating in Paleoseismology,

Fig. 1 This schematic of the carbon cycle explains the distribution of radiocarbon through the different parts of the earth system. The ratios are highest in the stratosphere with a gradient down through to the geosphere. In practice, levels throughout the troposphere and (terrestrial) biosphere are very similar. The ratios are depleted by about 5 % in the surface oceans, and the geosphere has radiocarbon ratios below detection limits



atmosphere. The precision of the method is typically of the order of ± 100 years for the last few 1,000 years rising to the order of $\pm 1,000$ years at its limit. For optimum precision, samples should contain >1 mg of carbon, but samples smaller than this (by at least one order of magnitude) can be dated in some circumstances with reduced precision.

Radiocarbon is a cosmogenic isotope, produced in the upper atmosphere. It is incorporated into almost all living organisms, but once they die and stop taking up carbon from the atmosphere, the radiocarbon content decays away with a half-life of about 5,730 years.

Unlike some dating techniques (such as uranium-series dating on speleothems or corals), radiocarbon is not an absolute dating technique, but relies on the comparison of radiocarbon measurements on samples to those made on samples of known age (a process called calibration). In order to understand the limitations and

potential of the method, it is important to understand how the method works in some detail.

Radiocarbon in the Global Carbon Cycle

Radiocarbon is formed by the action of high-energy cosmic rays, largely from outside the solar system, on the upper atmosphere. The flux of these cosmic rays varies for a number of reasons, including changes in solar activity and in the earth's magnetic field.

Once produced the radiocarbon is quickly oxidized to $^{14}\text{CO}_2$. This radiocarbon-labelled carbon dioxide is mixed throughout the atmosphere through the atmospheric circulation. This mixing is very rapid within each hemisphere, but not so rapid across the equator. As part of the global carbon cycle (see Fig. 1), carbon dioxide exchanges between the surface oceans and the atmosphere and is also absorbed by photosynthesis into green plants and thereby into the biosphere.

The overall effect of the transport of carbon through the carbon cycle is that:

- The $^{14}\text{C}/^{12}\text{C}$ ratio is fairly uniform within each atmospheric hemisphere. There is a small-amplitude annual oscillation of the order of 0.4 % in the preindustrial era (Kromer et al. 2001) with longer-term fluctuations due both to changing production rates and changes in the carbon cycle itself (particularly ocean circulation and ventilation). The Southern Hemisphere is typically depleted in radiocarbon compared to the Northern Hemisphere by about 0.5 %.
- Living organisms that take their carbon directly from the atmosphere share the same radiocarbon content, though there is usually some isotopic fractionation between the atmosphere and biosphere, which can be corrected for by measuring the stable isotope ratio $^{13}\text{C}/^{12}\text{C}$. Reported radiocarbon dates are normally corrected for fractionation effects. Green plants usually form the lowest level in the food chain, and thus the radiocarbon throughout the biosphere is the same. However, plants and animals that live for significant periods might not be fully equilibrated with the atmosphere, and this has to be considered when interpreting radiocarbon dates.
- The surface oceans are depleted in radiocarbon compared to the atmosphere by ~5 %. The depletion in the deep oceans is much greater and highly dependent on local ocean circulation. Because of the complexities of the ocean system, the exact relationship between the surface oceans varies both spatially and temporally, making the dating of samples from the marine environment more complex.
- Carbon-containing geological formations made ultimately from living organisms (such as carbonates and coal) are sufficiently old that they do not contain any measurable radiocarbon.
- Rivers, lakes, streams, and groundwater can all contain a mixture of carbon from the atmosphere and from geological substrates. For this reason, samples from such systems will often be depleted in radiocarbon compared to the

atmosphere by amounts that have to be locally determined. This affects the radiocarbon dating of speleothems and organic deposits deriving their carbon from fully aquatic food systems.

Samples for Radiocarbon Dating

Information about the carbon cycle has a direct influence on the optimum choice of samples for dating seismic events. There are three main criteria for sample selection (assuming availability of suitable material), which are generally common to all applications of radiocarbon dating (Waterbolk 1971). These are: uncertainty in inbuilt radiocarbon age, association of the sample with the event, and state of preservation of the samples:

- Inbuilt radiocarbon age can be understood from the carbon cycle. Short-lived terrestrial plant remains, such as leaves, seeds, and small stems, normally have carbon directly from the atmosphere (via photosynthesis); these are ideal material for dating. Animal bones may have inbuilt age of the order of a few years to a few decades. Such offsets are often insignificant, and this makes these good materials for dating, but often unavailable in sites where seismic activity is clearly recorded. Wood from long-lived tree species is more problematic, as it can have inbuilt age of decades to centuries; in such cases, dendrochronological analysis might be appropriate, or outer rings can be selected for dating if available. Aquatic and marine environments are themselves depleted in radiocarbon and so have inbuilt radiocarbon age; it can be possible to correct for this, but in so doing, some precision is usually lost.
- Association between the sample and the seismic event of interest is clearly critical. In part, this comes down to stratigraphy, but also depends to some extent on the sample material itself. Fragile materials such as plant fragments are likely to be in their primary deposits because they are unlikely to survive significant reworking. Well-preserved plant macrofossils are therefore to be preferred over

charcoal fragments, which might well remain within sedimentary deposits during transport. In archeological contexts, articulated bones are another sample type where the nature of the sample indicates primary deposition. When dating raised shorelines, which can be useful for measuring tectonic uplift, marine species that are known to live close to the surface are best as they are less likely to date from previous uplift events (see, e.g., Shaw et al. 2010).

- For accurate dating samples also need to be well preserved. Most radiocarbon pretreatment methods work on the basis of recovering large molecules that remain in the sample from the living organism. These might be cellulose in plant remains, collagen in bone, or primary carbonates in the case of shell. Large organic molecules decay with time and with changes in water content of the sediments, and carbonates can be recrystallized over time. For these reasons, finding a well-preserved material can be a challenge.

There are often competing considerations in the choice of materials. For example, the optimum type of sample may not be available in close association with the evidence for the seismic event. In some instances, only poorly defined “organic carbon” from sediments is available for dating, which inevitably introduces considerable uncertainty into the dating, since there is no single link to a living organism, or even a coeval group of organisms.

Measurement of Radiocarbon

Radiocarbon dating has two main stages: pretreatment and measurement.

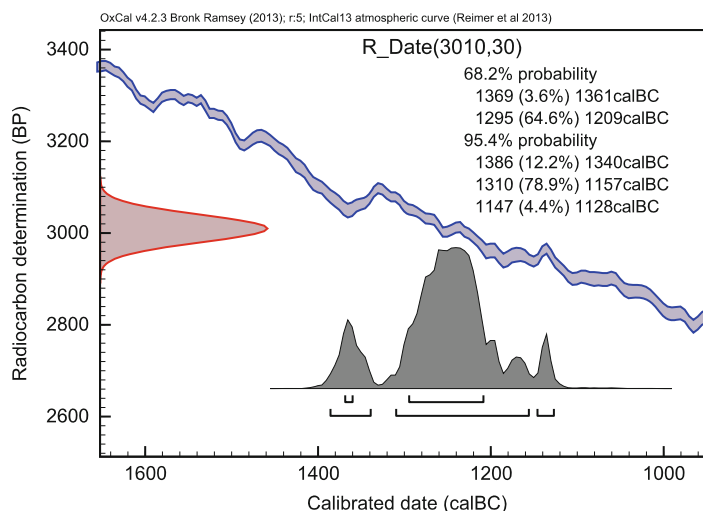
Pretreatment is the chemical extraction of material known or presumed to be from the original organism. It is necessary to remove contaminants that come from the environment, which might be of a very different age. In general, these methods work by purification of stable large molecules (cellulose, collagen, chitin), mineral complexes such as shell, or reduced carbon (in the form of charcoal). With the exception of

reduced carbon, these are carbon-containing materials, which would have been part of the original living organism and should therefore have a carbon content reflecting that of the prevailing environment. Reduced carbon is assumed to derive from such large molecular components and thus also provide material suitable for dating. Pretreatment usually involves using acid (for removal of deposited carbonates or dissolved carbon dioxide) and alkali (for the removal of humic acids). However, particularly for accurate dating of older samples, there are now much more elaborate pretreatment methods available. It is important to discuss the specific requirements with a radiocarbon laboratory with suitable experience.

The measurement process gives the isotope ratio of the sample. To correct for fractionation through the carbon cycle, it is necessary to measure both the $^{14}\text{C}/^{12}\text{C}$ and the $^{13}\text{C}/^{12}\text{C}$ ratios. The $^{13}\text{C}/^{12}\text{C}$ is normally measured using a stable isotope mass spectrometer, and the $^{14}\text{C}/^{12}\text{C}$ can either be measured by decay counting (gas or liquid scintillation counting, both of which need sample sizes of the order of 1 g) or by accelerator mass spectrometry (AMS, which requires a sample size of the order of 1 mg). In both cases, the radiocarbon date reported should contain the following information:

- A laboratory code, which identifies the laboratory and gives the information necessary to check any analytical details at a later stage.
- The nominal radiocarbon “age” calculated using the Libby half-life of 5568a with a 1σ standard uncertainty (or alternatively a ratio compared to the atmospheric ratio for 1950). This result should already be corrected for any isotopic fractionation. The units for radiocarbon dates are normally BP (sometimes written as ^{14}C BP) where the BP means “before present,” “present” being AD 1950.
- The $^{13}\text{C}/^{12}\text{C}$ ratio, expressed as $\delta^{13}\text{C}$ relative to the PDB standard.

These three components should always be reported as primary data with any study using



Radiocarbon Dating in Paleoseismology, Fig. 2 This shows a typical calibration plot for a single radiocarbon date. The radiocarbon measurement in this case is 3010 ± 30 BP (uncertainty at 1σ). The *red* distribution illustrates the uncertainty in this measurement on the radiocarbon date scale. The *blue* curve is the atmospheric calibration curve for this time period (with lines at $\pm 1\sigma$).

The *gray* probability distribution on the calibrated date axis shows the probability that the sample is any particular age. This can then be summarized into ranges either at the 68.2 % level (the most likely dates) or the 95.4 % level (ranges within which the date very likely lies). Note that the ranges are split and the distribution is multimodal; this is typical for radiocarbon calibrations

radiocarbon data. It is not sufficient to report the calibrated date (see next section) because calibration curves are updated periodically, and it is not possible to get back to the original measurement data from the calibrated date.

Calibration of Radiocarbon Dates

A radiocarbon “date,” as reported by a radiocarbon laboratory, is really just as an isotope ratio, and the conversion to a notional radiocarbon “date” is a convention that arose to put the measurement in a broad chronological context. The half-life used is that estimated by Libby and now known to be too low. The date is based on the assumption that the levels of radiocarbon in the atmosphere have remained constant over time, so to convert this into a real date, it is necessary to calibrate it against material of known age. To do this, it is necessary to use a calibration curve, which is a compilation of a very large number of radiocarbon measurements on material from records with independent chronologies. For the Holocene and Late Glacial

periods, these chronologies are based on dendrochronology; for earlier periods, the main control is uranium-series dating.

There are two main calibration curves, one for the atmosphere (currently IntCal13) and one for the oceans (Marine13) with details published in Reimer et al. (2013). These latest curves cover the full effective range of the radiocarbon dating technique (50,000 years).

When calibrating a radiocarbon date, it is first important to determine the reservoir from which the sample draws its carbon (Northern Hemisphere Atmosphere, Southern Hemisphere atmosphere, ocean, etc.). For the oceans, it is also necessary to know how different the local ocean radiocarbon levels are from the ocean average. A useful reference for this is the Marine Reservoir Database (<http://calib.qub.ac.uk/marine/>), which provides a compilation of preindustrial values. It should be noted, however, that changes in ocean circulation over time make these values inherently uncertain; this is especially true when considering pre-Holocene dates.

Calibration of a single radiocarbon date can be carried out using a number of different software packages – two common ones being Calib (<http://calib.qub.ac.uk/calib/>) and OxCal (<http://c14.arch.ox.ac.uk/OxCal.html>). Figure 2 shows an example calibration using OxCal and explains some of the main elements of the process.

Constraining Dates of Seismic Events

Using radiocarbon dates for the study of seismic events requires particular attention to the association between the samples that have been radiocarbon dated and the seismic event of interest. There are a number of different approaches that can be taken depending on the particular context.

The most straightforward approach is to try to find samples that relate very directly to the seismic event itself. Examples of this approach could be debris from a seismic event (e.g., destruction level in a human occupations site), or debris from a tsunami event, or marine material exposed by a tectonic uplift event (raised shoreline). In almost all cases, the material being dated will actually predate the seismic event, and efforts are made to ensure that the samples are not likely to predate the earthquake very significantly (see section on “[Samples for Radiocarbon Dating](#)”).

If there is more than one radiocarbon date associated with the seismic event and if all of the events can be assumed to come from within a year or so of the event (well within 10 years), then the different measurements can be combined before calibration. This might be the case, for example, in a tsunami or destruction event that has large quantities of short-lived plant material for dating. Before any such combination, the spread of measurements should be checked using a χ^2 test (Ward and Wilson 1978).

In most cases, however, it is very difficult to date a seismic event directly, and often all that can be done is to obtain dates that bracket the event in some way. If only one date is available, there is little that can be done other than calibrate it and discuss the implications

for the event in relation to this one constraint. If there are many dates, then it is possible to use a Bayesian statistical model to combine all of the available information and get quantitative estimates for the dates of the seismic events themselves.

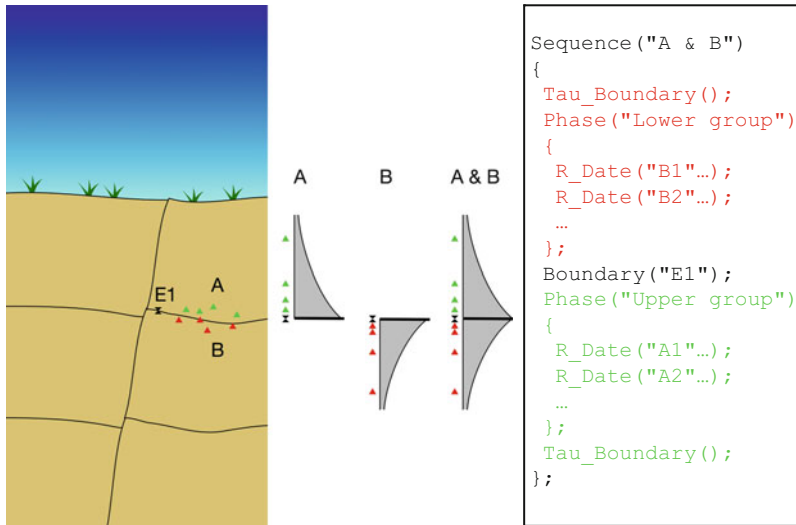
Bayesian Modelling of Chronologies

When dealing with radiocarbon dates, Bayesian statistics provides a useful framework for the integration of information from different sources and is widely used in archeology and environmental science. This is partly because of the non-normal nature of the uncertainties in calibrated radiocarbon measurements but also because such methods are very flexible and allow the inclusion of many different kinds of underlying model.

When dating seismic events, there are three main types of situation:

- Radiocarbon dates provide constraints for a single seismic event of interest.
- A sequence of seismic events is constrained by radiocarbon dates.
- Seismic events are recorded in a well-behaved sedimentary record, which can be dated by radiocarbon.

Each of these different possibilities requires slightly different approaches and so will be considered individually. There are a number of computer programs available for Bayesian analysis with stratigraphic constraint such as BCal (Buck et al. 1999) and OxCal (Bronk Ramsey 2009). For age-depth modelling, the same is also true with software such as OxCal (Bronk Ramsey 2008), BChron (Haslett and Parnell 2008), and Bacon (Blaauw and Christen 2011). Because OxCal uses a metalanguage, which makes it easy to specify models with text code, and because it can be used for both constrained modelling and full age-depth models, the examples given below are illustrated with that package. There is a primer on the use of OxCal for paleoseismology (Lienkaemper and Bronk Ramsey 2009).



Radiocarbon Dating in Paleoseismology, Fig. 3 The stratigraphy of deposit related to a seismic event (*E1*) is shown on the *left*. In the center are three different models that could be applied (*A*, if there is only overlying material; *B*, if there is only underlying material, older than the

event; *A* and *C*, where there is both). All models assume a concentration of dated samples close to the event. On the *right* is an outline of the OxCal code for the case where there is material below and above the event

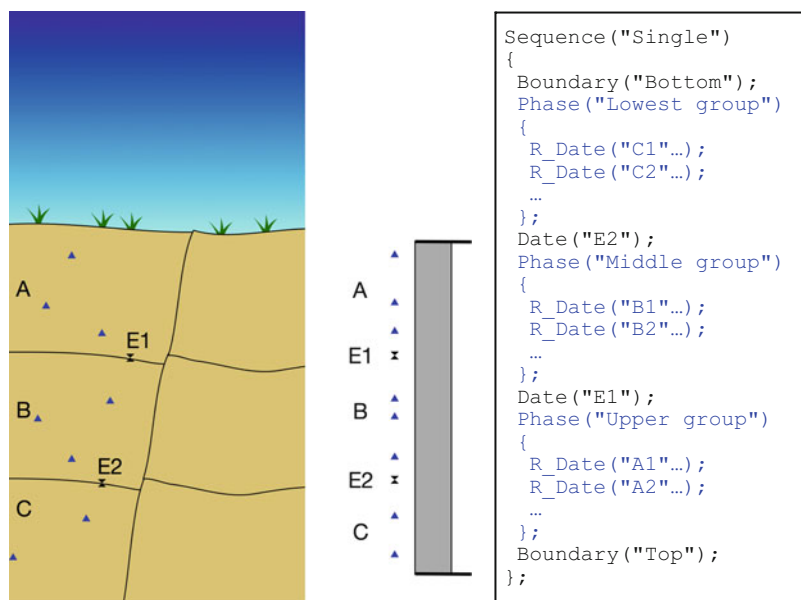
Constraints on a Single Seismic Event

The first situation to be considered is the case of a single seismic event. As discussed above, it is unusual to be able to date such an event directly. Quite often the dates associated with the event must all be older than the event because the samples are from material that was growing some time before. The model needs to take into account that most such material will be close in age to the event, but there is the possibility of occasional samples being substantially older. The most appropriate model for this scenario is one that treats the dated events as an exponential distribution through time (with an unknown time constant). Another less frequent situation is where there is only material overlying the seismic event to date. Ideally having both would constrain the event more precisely. Figure 3 shows such situations schematically. The program OxCal provides a metalanguage for the formulation of such models (Bronk Ramsey 2009). Note that in all cases the OxCal code definition follows the order of the events (from oldest to youngest) and therefore is reversed compared to the stratigraphic order.

Stratigraphic Sequences

The next situation is where there are several seismic events represented in the deposit. If the depositional processes are similar throughout, it is reasonable to assume that the dated samples and seismic events are all part of a sequence of events that have been sampled randomly. In this case, no assumptions are made about the stratigraphic sequence of samples between the seismic events. Figure 4 shows a schematic of such a situation and explains how such a model might be constructed in OxCal. Such a model will be realistic as long as the dates are randomly scattered through the sequence.

In practice, radiocarbon samples are often concentrated in particular parts of a sedimentary sequence, and this might well reflect underlying changes in the sedimentation processes. In such cases, the models should have “boundaries” which define the changes in sedimentation. In some cases, it may be that the seismic events correspond with significant changes in sedimentation type, and in this case the seismic events should be classed as boundaries. Such a situation is shown in Fig. 5, where there is a concentration



Radiocarbon Dating in Paleoseismology, Fig. 4 A sedimentary sequence with a scatter of radiocarbon dates in three groups separated by two seismic events, *E1* and *E2*, is illustrated on the *left*. The central panel shows a schematic for the model applied in this case, the underlying assumption being a random scatter of events throughout the sequence. The *right-hand* panel shows

the main elements of an OxCal model implementation; note that in this case, the radiocarbon dates are nested within “phases” implying that there is no information about the relative order within each group. This is similar to the example given in (Lienkaemper and Bronk Ramsey 2009)

of radiocarbon samples in an organic layer underlying each seismic event. The use of boundaries ensures that the concentration of radiocarbon dates in parts of the sequence does not itself bias the model as a whole.

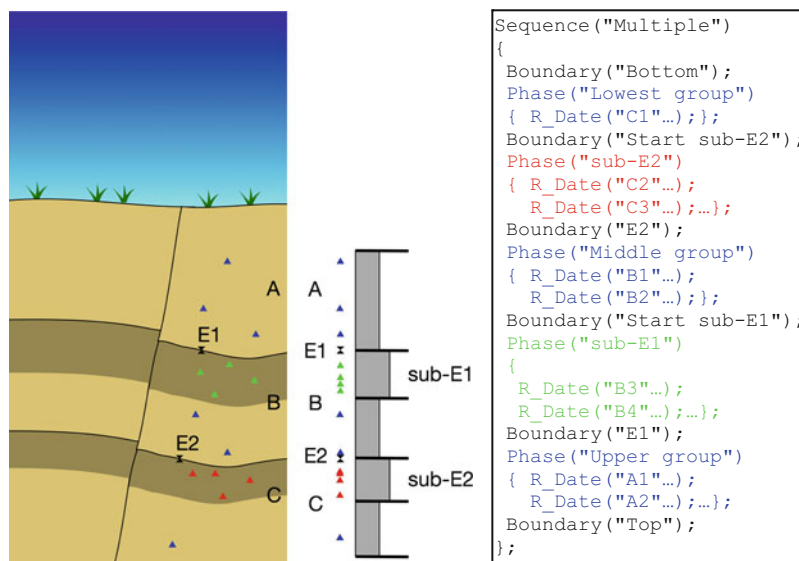
Deposition Models

Another type of Bayesian model that is useful for paleoseismology is an age-depth model for sedimentary sequences. Such models can be used when there are a series of radiocarbon dates in a single sediment column, with depth information. Most commonly, these are used for cores, but could also be applied to other types of section if a single depth scale can be defined. Like other Bayesian models, there needs to be a mathematical model for the deposition process. One such model is the Poisson process model implemented in OxCal. In this model, the assumption is that the deposition is random which allows for flexibility in an otherwise

approximately uniform deposition (Bronk Ramsey 2008). The degree of flexibility can be determined by the model, using the available data (Bronk Ramsey and Lee 2013). Figure 6 outlines the main aspects of such a model, in this case assuming that the sedimentary processes do not change significantly down the core. As with the sequences described above, boundaries can be introduced to split such a model into sections if there are major lithostratigraphic changes, which might imply an abrupt change in sedimentation rate.

Recurrence Information

So far, the discussion here has centered on retrieving quantified date information for seismic events. However, in many instances, it is equally important to get information about the interval between events, especially when estimating risk. Fortunately, the Bayesian models allow this information to be extracted and summarized.



Radiocarbon Dating in Paleoseismology, Fig. 5 The situation illustrated here is similar to that in Fig. 4, except that the dated samples are no longer randomly scattered and there is a clear stratigraphic boundary within the levels *B* and *C*. The central panel shows a schematic for the most appropriate model, which is divided up into levels within which the deposition of samples is more

likely to be random. The *top* and *bottom* of each of the organic layers that contain the concentration of radiocarbon samples are defined as boundaries in the model; in this case, this also coincides with the seismic events *E1* and *E2*. The *right-hand* panel shows, in outline, how such a model might be constructed in OxCal

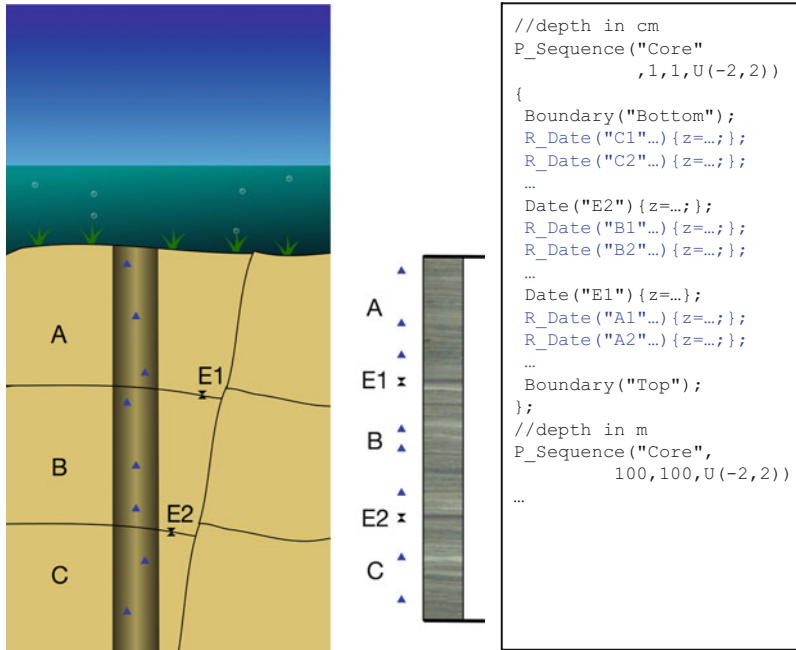
There are two main properties of interest: the mean recurrence interval and the recurrence error distribution (Lienkaemper and Bronk Ramsey 2009).

- The mean recurrence interval can be calculated if the difference in date between the first seismic event (e.g., E_n) and the final seismic event ($E1$) is known and the number of events (n). The mean recurrence interval is just given by $(E1 - E_n)/(n - 1)$. If extracted from a Bayesian model, this property will have a probability distribution which gives the uncertainty associated with this mean (but not the variability in the recurrence intervals).
- The variability in recurrence rates is also of interest. However, there is often considerable uncertainty in each of the successive recurrence intervals, and it is impossible to distinguish between this uncertainty and the actual variability in recurrence interval. The best

that can be done is to calculate each of the intervals (e.g., the interval between the last two events given by $I1_2 = E1 - E2$) and then to superimpose all of these to obtain a distribution using the sum function in OxCal (see details in Lienkaemper and Bronk Ramsey 2009 and Fig. 7). This summed distribution is a convolution of the original real recurrence interval distribution and the uncertainty in measuring the age differences between successive events.

Summary

Radiocarbon is a powerful dating technique, but its correct application requires a detailed consideration of the different parts of the method. In particular, the calibration of radiocarbon dates can only be carried out with a good understanding of the relationship of the dated samples to the global carbon cycle. Samples that derive their



Radiocarbon Dating in Paleoseismology, Fig. 6 Lake and marine cores provide a useful source of evidence for past seismic events. Where radiocarbon is used to date such cores, age-depth models can be used to integrate the information from the dated samples and the depths of the samples and events of interest. The central panel shows schematically the underlying assumption for a simple Poisson process model, where randomly varying

sedimentation continues over the whole sediment sequence. The *right-hand* pane shows the outline of the model definition within OxCal assuming the depth is given in cm, and depth output is required at cm intervals; if the depth is given in m, the first line of the code is different (see Bronk Ramsey and Lee 2013, for more details)

```
// Calculate intervals between seismic events
// and summed interval distribution (ID)
Sum("ID")
{
  I1_2=E1-E2;
  I2_3=E2-E3;
  I3_4=E3-E4;
};
// Calculate mean recurrence interval
RI=(E1-E4)/3;
```

Radiocarbon Dating in Paleoseismology, Fig. 7 Example of OxCal code (from Lienkaemper and Bronk Ramsey 2009) for calculating the mean recurrence

interval (RI) and the recurrence error distribution (ID) for a model where there are four seismic events $E1$, $E2$, $E3$, and $E4$

carbon fairly directly from the atmosphere are easiest to calibrate, whereas calibrating samples from marine environments requires more specific local environmental information.

Alongside developments in radiocarbon dating itself, comprehensive methods have been developed for the integration of chronological information into Bayesian age models. They are

most often applied to studies involving radiocarbon dating but can also be used with other dating techniques. Although these have been developed, initially for archeology and more recently for paleoenvironmental studies more generally, they are well suited to quantifying characteristics of seismic systems.

marine fauna in Greece: implications for the interpretation of recent earthquake and tectonic histories using lithophagid dates. *Earth Planet Sci Lett* 297(3–4): 395–404

Ward GK, Wilson SR (1978) Procedures for comparing and combining radiocarbon age-determinations – critique. *Archaeometry* 20:19–31

Waterbolk HT (1971) Working with radiocarbon dates. *Proc Prehist Soc* 37:15–33

Cross-References

- [Archeoseismology](#)
- [Earthquake Recurrence](#)
- [Probabilistic Seismic Hazard Models](#)

References

- Blaauw M, Christen AJ (2011) Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Anal* 6(3):457–474
- Bronk Ramsey C (2008) Deposition models for chronological records. *Quat Sci Rev* 27(1–2):42–60
- Bronk Ramsey C (2009) Bayesian analysis of radiocarbon dates. *Radiocarbon* 51(1):337–360
- Bronk Ramsey C, Lee S (2013) Recent and planned developments of the program OxCal. *Radiocarbon* 55(2–3):720–730
- Buck CE, Christen JA, James GN (1999) BCal: an on-line Bayesian radiocarbon calibration tool. *Internet Archaeol* 7
- Haslett J, Parnell A (2008) A simple monotone process with application to radiocarbon-dated depth chronologies. *J R Stat Soc Ser C (Appl Stat)* 57(4):399–418
- Kromer B, Manning SW, Kuniholm PI, Newton MW, Spurk M, Levin I (2001) Regional (CO₂)-C-14 offsets in the troposphere: magnitude, mechanisms, and consequences. *Science* 294(5551):2529–2532
- Libby WF, Anderson EC, Arnold JR (1949) Age determination by radiocarbon content: world-wide assay of natural radiocarbon. *Science* 109(2827):227–228
- Lienkaemper JJ, Bronk Ramsey C (2009) OxCal: versatile tool for developing paleoearthquake chronologies – a primer. *Seismol Res Lett* 80(3):431–434
- Reimer PJ, Bard E, Bayliss A, Beck JW, Blackwell PG, Bronk Ramsey C, Grootes PM, Guilderson TP, Hafliðason H, Hajdas I, Hatté C, Heaton TJ, Hoffmann DL, Hogg AG, Hughen KA, Kaiser KF, Kromer B, Manning SW, Niu M, Reimer RW, Richards DA, Scott EM, Southon JR, Staff RA, Turney CSM, van der Plicht J (2013) IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* 55(4):1869–1887
- Shaw B, Jackson JA, Higham TFG, England PC, Thomas AL (2010) Radiometric dates of uplifted

Random Process as Earthquake Motions

Luigi Carassale
Department of Civil, Chemical and
Environmental Engineering, University of
Genova, Genoa, Italy

Synonyms

Ground motion; Random process

Introduction

Earthquakes appear as an erratic phenomenon and are considered as unpredictable, at least from a deterministic point of view. They involve a randomness that is at least twofold involving the occurrence of earthquakes (when and where they appear) and the ground motion that they produce. An individual earthquake arrives in a random place on earth (even if some places are more likely to see a seismic event than others), appears at a random time and has variable intensity. On the other hand, the ground motion generated by the earthquake has random amplitude (e.g., peak acceleration), duration and spectral characteristics. Besides, the ground motion is variable in space due to a number of physical phenomena related to the propagation of the seismic waves. This intrinsic randomness suggests using probabilistic tools to represent and characterize earthquakes.

Of the two faces of the problem mentioned above, only the second one is considered here,

namely the stochastic characterization of the ground motion, given that an earthquake occurs at a specific site. The definition of stochastic models for this purpose have been of interest for a long time and several studies have been presented in the scientific literature mostly in the 1960s and 1970s of the last century.

Stochastic models for the representation of seismic ground motion are used to simulate artificial seismic acceleration time histories with specified characteristics, which may be employed in the framework of the Performance-Based Earthquake Engineering. The most common choice in this context is the use of natural acceleration records that are scaled and applied, as excitation, to a numerical model of the structure. This approach has several drawbacks including the scarce availability of natural recorded ground motions for the specific earthquake characteristics of interest (magnitude, distance, type of faulting, site conditions, etc.), the doubtful meaningfulness of the scaling procedure (e.g., Grigoriu 2011) and the questionable choice of the intensity measure used as reference for scaling.

Artificial ground motions can be obtained through Monte Carlo simulation from generative stochastic models that are parametrized with quantities that may be related to the site and earthquake characteristics. Simulated ground motions can be used as a surrogate of measured ones to calculate the dynamic response of a structure. Besides, the stochastic model can be directly used to predict the probabilistic response of the structure within the framework of the stochastic structural dynamics, whenever the structure is simple enough to enable this approach.

Seismic ground motions are randomly variable in time and space and are commonly modeled as random processes. As far as their temporal variability is concerned, the observation of recorded time histories of ground acceleration immediately reveals their nonstationary nature in terms of amplitude. A more careful analysis also reveals a significant variability of their harmonic content with the time. Figure 1 shows, as an example, a ground motion measured in Kobe, Japan, during the 1995 earthquake. The variation in time of the amplitude is

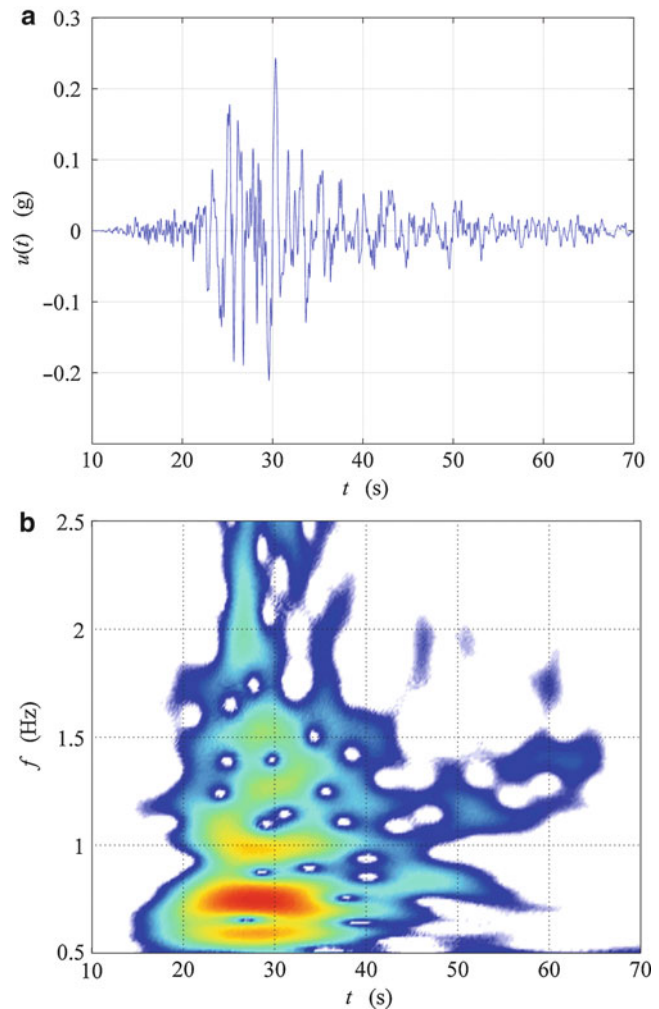
well evident observing the acceleration time history (Fig. 1a), while the variation of the harmonic content can be observed in the wavelet map (Fig. 1b).

Temporal nonstationarity can be easily modeled by multiplying a stationary process by a time-varying envelope function, while the modeling of spectral nonstationarity is definitely more complicated. Several researchers (e.g., Rezaeian and Der Kiureghian 2008) believe that, both these effects may be important, in particular when dealing with the analysis of nonlinear structures. In this case, indeed, the evolution of the dynamic characteristics of the structure due to degradation may match in an unfavorable way with the variation of the spectral content of the seismic input. As far as the spatial variability is concerned, several studies demonstrated that the earthquake ground motion measured in different points of a site may be significantly different due to a series of physical reasons. This difference increases as the mutual distance of the points increases and may have an important impact on structures that have large planimetric size such as bridges and pipelines.

A large number of stochastic models to represent the ground motion in a single point have been developed and described in several review papers (e.g., Liu 1970; Ahmadi 1979; Shinozuka and Deodatis 1988; Kozin 1988; Conte and Peng 1997). The existing stochastic models can be classified into four categories: (a) processes obtained by passing a Gaussian white noise through a filter, with subsequent modulation in time to achieve temporal nonstationarity; (b) processes obtained filtering an amplitude-modulated Gaussian white noise; (c) processes obtained through the amplitude modulation of a filtered Poisson process; (d) processes obtained by linear filtering of an amplitude modulated Poisson process; (e) processes realized as the output of time-variant linear filters driven by a Gaussian white noise. Models of type (a) are the simplest ones and are discussed in the following. Their main limitation involves their inability of reproducing time-variable frequency contents. Models of type (c) and (d) are rather flexible being able to produce

Random Process as Earthquake Motions,

Fig. 1 Time history (a) and wavelet map (b) of a ground acceleration (Kobe 1995)



amplitude and spectral nonstationarity, as well as non-Gaussian probability distributions. On the other hand, they require a quite difficult calibration process and it is difficult to relate model quantities with physical parameters obtained a priori from the characteristics of the site. The advantage of models type (e) is in their ability in modeling both temporal and spectral nonstationarity within the framework linear filtering (Rezaeian and Der Kiureghian 2008). An example of these models is briefly reported.

The models available to represent the spatial variability of the seismic motion are essentially expressed in terms of coherence function (e.g., Der Kiureghian 1996), therefore can be easily casted into the framework of models type (a).

Stochastic Models for the Seismic Ground Motion

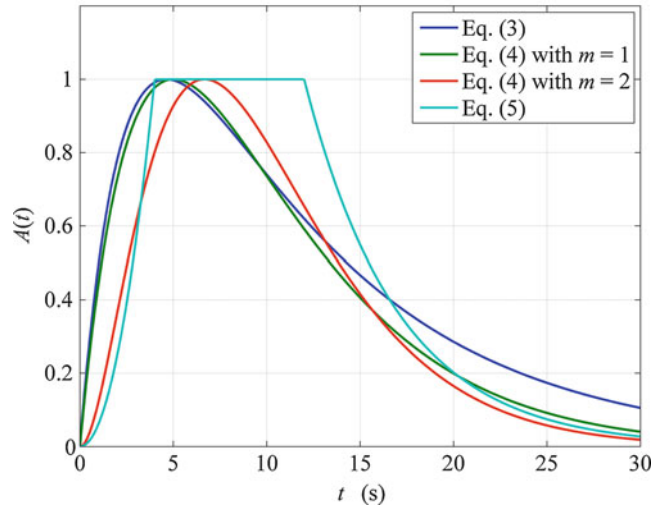
According to the models type (a) defined above, the ground acceleration $u(t)$ due to a seismic event is modeled as a zero-mean random process expressed in the form:

$$u(t) = A(t)y(t) \quad (1)$$

where $y(t)$ is a zero-mean stationary Gaussian random process (function of the time t) whose Power Spectral Density (PSD) function is $S_y(\omega)$, ω being the circular frequency; $A(t)$ is a slowly varying modulation function. Within the hypothesis that the variation of $A(t)$ is slow with respect

Random Process as Earthquake Motions,

Fig. 2 Time modulating functions. Equations 3, 4, and 5; $b_1 = 0.1$, $b_2 = 0.4$, $b = 0.2$, $m = 1$ or 2 , $t_1 = 4$ s, $t_2 = 12$ s



to the characteristic time scale of $u(t)$, the ground acceleration $u(t)$ can be represented through the Evolutionary Power Spectral Density (EPSD) function given as:

$$S_u(\omega, t) = |A(t)|^2 S_y(\omega) \quad (2)$$

A random process in the form of Eq. 1 is referred to as a uniformly modulated process; the amplitude of its PSD changes in time according to Eq. 2, but the harmonic distribution of its power is time-invariant.

A number of models for the modulating function $A(t)$ have been proposed, mostly in the 1960s and 1970s of the last century. Some of these models are reported here below (Shinozuka and Sato 1967; Iwan and Hou 1989; Jennings et al. 1969).

$$A(t) = A_0(e^{-b_1 t} - e^{-b_2 t}) \quad (3)$$

$$A(t) = A_0 t^m e^{-bt} \quad (4)$$

$$A(t) = \begin{cases} (t/t_1)^2 & t \leq t_1 \\ 1 & t_1 < t \leq t_2 \\ e^{-b(t-t_2)} & t > t_2 \end{cases} \quad (5)$$

where b, b_1, b_2, m, t_1, t_2 are positive constants and A_0 is usually calibrated to obtain the maximum value of $A(t)$ equal to unity. Figure 2 shows a

qualitative comparison of the four modulation functions obtained through Eqs. 3, 4, and 5.

As far as the random process $y(t)$ is concerned, a popular PSD model attributed to Kanai (1957) and Tajimi (1960) represents the ground acceleration as the absolute acceleration of a linear single-degree-of-freedom (sdof) system excited by a white noise. In this sense, the oscillator represents the terrain and the input white noise is the acceleration at the faulting point. Under this hypothesis the PSD of the process $y(t)$ is given as:

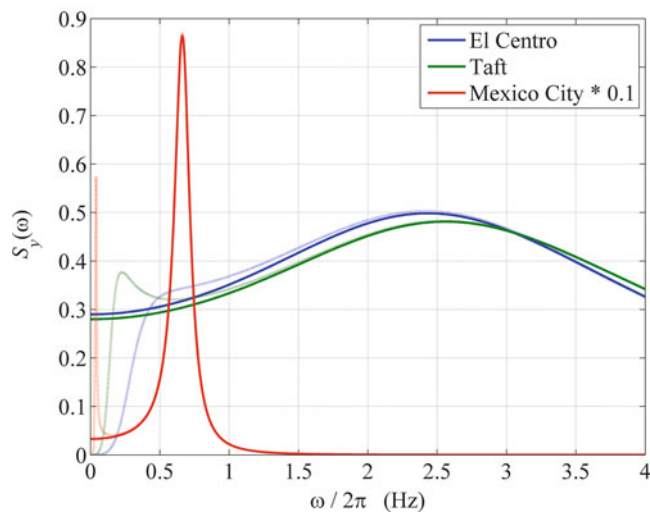
$$S_y(\omega) = S_{KT}(\omega) = \frac{(\omega_K^4 + 4\xi_K^2 \omega_K^2 \omega^2)}{(\omega_K^2 - \omega^2)^2 + 4\xi_K^2 \omega_K^2 \omega^2} S_0 \quad (6)$$

where the model parameters ω_K and ξ_K can be interpreted as natural frequency and damping ratio of the terrain; S_0 is a constant parameter scaling the variance of the random process and is related, in some sense, with the intensity of the earthquake. For the case of stiff terrains ω_K is about 4π – 5π and ξ_K is about 0.6–0.7.

A clear drawback of the Kanai–Tajimi (KT) model is its inconsistency in the low-frequency range due to its finite value for ω tending to zero. This implies that the PSD of displacement (and velocity) diverges giving rise to an unlimited amplitude of the ground displacement.

Random Process as Earthquake Motions,

Fig. 3 PSD of the parent stationary process $y(t)$. KT model (*solid line*) CP model (*dashed line*). Model parameters are given in Table 1. The ordinate for Mexico City earthquake is divided by 10



Random Process as Earthquake Motions, Table 1 Example of parameters for KT and CP models (after Yeh and Wen 1990)

Earthquake	ω_K (rad/s)	ζ_K	ω_P (rad/s)	ζ_P	S_0 (m ² s ⁻³)
El Centro	19.0	0.65	2.0	0.6	$2.9 \cdot 10^{-3}$
Taft	20.0	0.65	1.0	0.5	$2.8 \cdot 10^{-3}$
Mexico City	4.2	0.1	0.23	0.1	$3.3 \cdot 10^{-3}$

This problem can be easily circumvented by introducing a high-pass filter in series with the filter defining the KT model. A popular model obtained in this way is attributed to Clough and Penzien (1975) and is defined as:

$$S_y(\omega) = S_{CP}(\omega) = \frac{\omega^4}{(\omega_P^2 - \omega^2)^2 + 4\xi_P^2\omega_P^2\omega^2} S_{KT}(\omega) \tag{7}$$

where ω_P and ξ_P are further model parameters, which do not have any physical interpretation and are calibrated to achieve the desiderated high-pass characteristics of the filter. Obviously, other types of high-pass filter can be adopted for this purpose (e.g., Giaralis and Spanos 2009). Figure 3 shows the PSD of the process $y(t)$ obtained adopting the Clough-Penzien (CP) model, using the parameters reported in Table 1 estimated for three earthquakes (Yeh and Wen 1990).

As it has been mentioned, the models described above are unable to reproduce spectral nonstationarity. Limiting to the models based on

linear filtering of Gaussian processes, several approaches to produce spectral nonstationarity have been proposed. These usually adopt linear filters characterized by time-variable parameters that are calibrated to match some feature extracted from natural accelerograms. Several implementations are available, including discrete-time recursive models realized through several parameterizations. Here, a model based on the parametrization of the impulse-respose function IRF of the linear filter as proposed by Rezaeian and Der Kiureghian (2008) is briefly described since it can be easily seen as a generalization of the classical models presented above.

According to this model, the ground acceleration can be expressed as:

$$u(t) = A(t) \left[\frac{1}{\sigma(t)} \int_{-\infty}^t h(t - \tau; \boldsymbol{\theta}(\tau)) w(\tau) d\tau \right] \tag{8}$$

where $w(\tau)$ is a Gaussian white noise with intensity S_0 , $A(t)$ is an envelope function of the type of

Eqs. 3, 4, and 5, $\theta(\tau)$ a vector of time-variable model parameters defining the kernel h , and $\sigma(t)$ is the time-variable standard deviation of the filtered process, which can be obtained as:

$$\sigma(t) = 2\pi S_0 \int_{-\infty}^t h^2(t - \tau; \theta(\tau)) d\tau \quad (9)$$

Stochastic Models for Spatial Variability of the Ground Motion

The seismic ground motion is usually considered as fully correlated in space and it is represented by single-point functions, such as time histories, response spectra, PSDs, or evolutionary spectra. Single-supported structures are usually studied by expressing their equations of motion in terms of relative displacements between the structural elements (or lumped masses) and the ground. In this way, the inertia forces due to the seismic acceleration can be schematized as a system of apparent forces applied on the structural masses and proportional to the acceleration itself. This approach is justifiable for point-like facilities, such as compact buildings, where the lack of correlation of the ground motion may be ignored. However, for spatially extended structures, such as pipelines, or structures based on widely separated supports, such as bridges or viaducts, the partial correlation of the ground motion may be determinant for a reliable structural analysis.

Seismic accelerations recorded at different points of the ground differ because of the loss of coherence that gradually increases with distance and because of the phase shift due to the propagation of the seismic waves (Der Kiureghian 1996). A representation of the seismic acceleration that takes into account its partial correlation may be formulated in terms of multipoint statistics such as the cross-PSD. For multisupported structures, the displacement relative to the ground of the nodal points is not representative because of the variability of the seismic motion. The equations of motion can be conveniently written in terms of absolute displacements, separating the pseudostatic part, due to the static effect of the ground dragging, from the purely dynamic part of

the motion (e.g., Clough and Penzien 1975). The relative importance of these two contributions depends on the stiffness of the structure and on the characteristics of the soil determining the spatial correlation pattern of the ground motion. Many scientific papers investigated this problem concluding that fully correlated motions may produce higher or lower responses than partially correlated motions depending on the dynamic characteristics of structures, on the considered effects and on the cross-correlation between support motions (e.g., Zerva 1991; Loh and Ku 1995; Tubino et al. 2003).

According to Der Kiureghian (1996), four distinct phenomena give rise to the spatial variability of earthquake-induced ground motions: (1) loss of coherency of seismic waves due to scattering in the heterogeneous medium of the ground, as well as due to the differential superpositioning of waves arriving from an extended source, collectively denoted herein as the “incoherence” effect; (2) difference in the arrival times of waves at separate support points of the structure, denoted as the “wave-passage” effect; (3) gradual decay of wave amplitudes with distance due to geometric spreading and energy dissipation in the ground medium, denoted as the “attenuation” effect; and (4) spatially varying local soil profiles influencing the amplitude and frequency content of the ground motion under each foundation point, denoted as the “site-response” effect.

Empirical spectral models have been proposed and validated through array recordings of the ground motion along extended homogeneous areas (Bolt et al. 1982; Harichandran and Vanmarcke 1986; Abrahamson et al. 1991). These models assume implicitly that in each point of the terrain the acceleration can be represented through Eq. 1 as the amplitude modulation of a stationary random process. Accordingly, stochastic models to represent, jointly, the stationary parent process of the ground acceleration in pairs of points of the terrain have been developed. These models can be expressed in the frequency domain through the cross-PSD:

$$S_{y_j y_k}(\omega) = \sqrt{S_{y_j}(\omega) S_{y_k}(\omega)} \gamma_{jk}(\omega) \quad (10)$$

where S_{yj} and S_{yk} are the PSD of the stationary acceleration y_j and y_k in points j and k of the terrain, while γ_{jk} is the coherence function. According to the classification described above, the coherence function can be factorized as

$$\gamma_{jk}(\omega) = \gamma_{jk}^{(\text{inch})}(\omega) \gamma_{jk}^{(\text{wp})}(\omega) \gamma_{jk}^{(\text{att})}(\omega) \quad (11)$$

in which $\gamma^{(\text{inch})}$, $\gamma^{(\text{wp})}$ and $\gamma^{(\text{att})}$ are, respectively, the coherence functions representing the incoherence effect, the wave-passage effect and the attenuation effect encountered by the seismic wave. The site-response effect can be included into the PSD S_{yj} and S_{yk} .

The incoherence effect increases as the distance between the two points increases and is more significant for high-frequency harmonics. Der Kiureghian (1996), on the basis of statistical arguments, concluded that $\gamma^{(\text{inch})}$ can be expressed in the form

$$\gamma_{jk}^{(\text{inch})}(\omega) = \cos [\beta(d_{jk}, \omega)] \exp \left[-\frac{1}{2} \alpha^2(d_{jk}, \omega) \right] \quad (12)$$

where α and β are increasing functions of the distance d_{jk} between the two points, as well as of the frequency. These functions are often assumed as linear functions parametrized by the share-wave velocity.

The wave-passage effect is merely due to the different arrival time of the seismic waves in the two points. The coherence function corresponding to this effect has unitary modulus and can be written in the exponential as

$$\gamma_{jk}^{(\text{wp})}(\omega) = \exp \left[i\omega \frac{d_{jk}^L}{v_{\text{app}}} \right] \quad (13)$$

where d_{jk}^L is the distance between the two points projected along the wave propagation direction and v_{app} is the apparent wave propagation velocity.

The attenuation effect has negligible effects and it may be assumed $\gamma^{(\text{att})} = 1$.

Relationship Between PSD Models and Response Spectrum

For design purpose, the ground motion is usually assigned through acceleration, velocity, or displacement response spectra, while the stochastic representations discussed above are defined in terms of PSD of a parent stationary process and time-modulation functions. The relationship between these two representations is not obvious since it involves the expected maxima of the response of a sdof linear system subjected to the earthquake input, i.e., the response of the system:

$$\ddot{x}(t) + 2\xi_0\omega_0\dot{x}(t) + \omega_0^2x(t) = -u(t) \quad (14)$$

where $x(t)$ is the displacement of the oscillator with respect to the ground, the overdot denotes the time derivative, ω_0 is the natural frequency, and ξ_0 is the damping ratio. The oscillator is at rest when the earthquake arrives at $t = 0$.

Assuming that the input energy is distributed over a broad frequency band, and assuming that the structural damping ξ_0 is low, the response $x(t)$ is a narrow-band process which can be expressed through the Rice representation:

$$x(t) = a(t) \sin(\omega_0 t + \phi(t)) \quad (15)$$

where $a(t)$ and $\phi(t)$ are, respectively, the amplitude and phase modulating processes, whose time scale is assumed to be slow with respect to the carrier wave $\sin(\omega_0 t)$. This assumption reflects the circumstance that the frequency-response function of a linear system with small damping ratios has a very high peak in the neighborhood of $\omega = \omega_0$. Under these hypotheses, the amplitude modulation is a Markov-type random process governed by a first-order stochastic differential equation (Spanos and Solomos 1983) and its time-dependent variance is given by the expression:

$$\sigma_a^2(t) = \frac{\pi}{\omega_0^2} \exp(-2\xi_0\omega_0 t) \int_0^t \exp(2\xi_0\omega_0 \tau) S_u(\omega_0, \tau) d\tau \quad (16)$$

The standard deviation of the response amplitude can be formally related to the displacement response spectrum $S_d(\omega, \xi)$ through the equation:

$$S_d(\omega_0, \xi_0) = g \max_t [\sigma_a(t)] \quad (17)$$

where g is the peak factor (Vanmarcke 1976). Equation 17 establishes a relationship between the response spectrum and the EPSD of the ground motion. This relationship cannot be directly used as the peak factor g is unknown and, in general, depends on the stiffness and damping of the SDOF oscillator, on the frequency content and duration of the input process $u(t)$, and on the level of uncertainty furnished by the design spectrum as specified by codes. The peak factor is often estimated through formulations based on the extreme-value distribution, which, unfortunately, requires substituting the exact response with an equivalent stationary random process having limited duration. The brute-force alternative consisting of computing the peak factor by Monte Carlo simulation (Giaralis and Spanos 2009) seems to be more appropriate.

Closing Remarks

The seismic ground motion can be represented as a random process. Generative models can be formulated on the basis of several principles, including linear filtering and time modulation of a Gaussian white noise. According to this approach, a few classical models of ground acceleration have been described and the basic ideas behind a more recent generalization capable of spectral nonstationarity have been outlined.

When a structure or a facility is extended in space, the model seismic ground motion must reproduce the spatial correlation structure of the phenomenon. A class of spectral models able to represent multiple-point ground motion has been described.

While stochastic models represent the ground motion in terms of EPSD, usual engineering practice uses the concept of response spectrum. A formal link between the two representations is provided and discussed.

References

- Abrahamson NA, Schneider JF, Stepp JC (1991) Empirical spatial coherency functions for application to soil-structure interaction analysis. *Earthq Spectra* 7:1–28
- Ahmadi G (1979) Generation of artificial time-histories compatible with given response spectra – a review. *Solid Mech Arch* 4:207–239
- Bolt BA, Loh CH, Penzien J, Tsai YB, Yeh YT (1982) Earthquake strong motions recorded by a large near-source array of digital seismographs. *Earthq Eng Struct Dyn* 10:561–573
- Clough RW, Penzien J (1975) *Dynamics of structures*. Mc Graw-Hill, New York
- Conte JP, Peng BF (1997) Fully nonstationary analytical earthquake ground-motion model. *J Eng Mech ASCE* 123:15–24
- Der Kiureghian A (1996) A coherency model for spatially varying ground motions. *Earthq Eng Struct Dyn* 25:99–111
- Giaralis A, Spanos PD (2009) Wavelet-based response spectrum compatible synthesis of accelerograms – Eurocode application (EC8). *Soil Dyn Earthq Eng* 29:219–235
- Grigoriu M (2011) To scale or not to scale seismic ground-acceleration records. *J Eng Mech ASCE* 137:284–293
- Harichandran RS, Vanmarcke E (1986) Stochastic variation of earthquake ground motion in space and time. *J Eng Mech ASCE* 112:154–174
- Iwan WD, Hou ZK (1989) Explicit solution for the response of simple systems subjected to nonstationary random excitation. *Struct Saf* 6:77–86
- Jennings PC, Housner GW, Tsai C (1969) Simulated earthquake motions for design purpose. In: Rodrigo Flores A (ed) *Proceedings of the 4th world conference on earthquake engineering*. Santiago de Chile, vol A-1, pp 145–160
- Kanai K (1957) Semi-empirical formula for the seismic characteristics of the ground. *Bull Earthq Res Inst* 35:309–325
- Kozin F (1988) Autoregressive moving average models of earthquake records. *Probab Eng Mech* 3:58–63
- Liu SC (1970) Synthesis of stochastic representations of ground motions. *Bell Syst Tech J* 49:521–541
- Loh CH, Ku BD (1995) An efficient analysis of structural response for multiple-support seismic excitations. *Eng Struct* 17:15–26
- Pacific Earthquake Engineering Research Center (PEER), PEER-NGA database, <http://ngawest2.berkeley.edu/>
- Rezaeian S, Der Kiureghian A (2008) A stochastic ground motion model with separable temporal and spectral nonstationarities. *Earthq Eng Struct Dyn* 37:1565–1584
- Shinozuka M, Deodatis G (1988) Stochastic process models for earthquake ground motion. *Probab Eng Mech* 3:114–123
- Shinozuka M, Sato Y (1967) Simulation of nonstationary random processes. *J Eng Mech ASCE* 93:11–40

- Spanos PD, Solomos GP (1983) Markov approximation to transient vibration. *J Eng Mech Div ASCE* 109: 1134–1150
- Tajimi H (1960) A statistical method for determining the maximum response of a building structure during earthquake. In: Muto K (ed) *Proceedings of the second world conference on earthquake engineering*, Tokyo-Kyoto, vol 2. Gakujutsu Bunken Fukyu-Kai, Tokyo, pp 781–797
- Tubino F, Carassale L, Solari G (2003) Seismic response of multi-supported structures by proper orthogonal decomposition. *Earthq Eng Struct Dyn* 32:1639–1654
- Vanmarcke EH (1976) Structural response to earthquakes. In: Lomnitz C, Rosenblueth E (eds) *Seismic risk and engineering decisions*. Elsevier, Amsterdam
- Yeh CH, Wen YK (1990) Modeling of nonstationary ground motion and analysis of inelastic structural response. *Struct Saf* 8:281–298
- Zerva A (1991) Effect of spatial variability and propagation of seismic ground motions on the response of multiply supported structures. *Probab Eng Mech* 6:212–221

Reconstruction Following Earthquake Disasters

Suzanne Wilkinson¹,
Alice Yan Chang-Richards² and
James Olabode Bamidele Rotimi³

¹Department of Civil and Environmental Engineering, The University of Auckland, Auckland, New Zealand

²The University of Auckland, Auckland, New Zealand

³Auckland University of Technology, Auckland, New Zealand

Synonyms

Disaster reconstruction stages; Reconstruction expectations

Introduction

Disaster reconstruction has discernible stages which, once understood, can be used to help inform planning and reconstruction. Reconstruction is part of the recovery stage in a disaster and describes the physical regeneration of the built

environment, including housing, commercial buildings, and infrastructure. Disasters are commonly conceptualized in terms of reduction, readiness, response, and recovery stages. Similar to overall disasters, reconstruction passes through a number of stages, during which activities, such as funding, compliance, assessments, and rebuilding, are undertaken. Reconstruction is nonlinear, which means that housing reconstruction might go through reconstruction stages at different times to infrastructure reconstruction and in different locations within the same disaster. Housing, commercial properties, and infrastructure might be temporarily fixed and then permanently reinstated at a later date; however, the overall reconstruction stages are similar for each type of reconstruction. Drawing on previous studies, this article focuses on reconstruction following earthquakes and uses the earthquakes in Canterbury to illustrate the common stages of reconstruction.

The Disaster Reconstruction Process

Reconstruction can be considered in the wider context of an ongoing and fluctuating recovery. Governments have attempted to plan for recovery and to develop policies which incorporate increased resilience into recovery frameworks. An example of these policies can be found in New Zealand where the “Focus on Recovery” (MCDEM 2005), a framework encompassing the community and four environments, social, economic, natural, and built environment, was adopted into the overall civil defense and emergency management framework to provide multilevel and multiagency structures and processes for post-disaster recovery. The reconstruction of the built environment is seen as one of the key elements within recovery. According to Quarantelli (2008) reconstruction refers to the post impact rebuilding of the physical structures destroyed or damaged in a disaster. The relationship between reconstruction and community recovery has been summarized by Quarantelli (2008) as “The longer the reconstruction process, the slower the recovery of the community since recovery in other dimensions is also slowed.”

Reconstruction Following Earthquake Disasters,
Table 1 Wilkinson’s five stages of reconstruction mapped against Brunsdon and Smith’s (2004) reconstruction activities and other key activities

Reconstruction stage	Main activities
Chaos	Impact assessments
Realization	Proposals made
	Interim funding determined
	Governance and organizational structures setup
Mobilization	Reconstruction commences
Struggle	Reconstruction slows as problems emerge
	Problems with compliance, consents, shortages (labor, materials), quality, leadership, housing affordability, conflict
New normal	Establishing new ways of living

By understanding the stages of reconstruction, opportunities can be found for improving and increasing the speed of the overall recovery, including community recovery. The five stages of reconstruction, as discussed by Wilkinson (2013), are identified as chaos, realization, mobilization, struggle, and new normal. Other authors, such as Brunsdon and Smith (2004), listed key activities required for reconstruction as impact assessment, restoration proposal, funding, statutory compliance, and reconstruction. Brunsdon and Smith’s (2004) stages can be integrated into those of Wilkinson’s (2013) to create a fuller picture of the reconstruction process as seen in Table 1.

Reconstruction Stage 1: Chaos

In the initial stages of recovery, there tends to be significant chaos where the general situation is characterized by the question “what do we do?” The key reconstruction features found in this stage are the need for assessments to be done on buildings and communities. Brunsdon and Smith (2004) described the need for impact assessments. These assessment requirements are similar to the World Bank’s recommendations for assessing needs, identifying priorities, and

planning recovery, where one of the first activities usually undertaken is the assessment of damage and the impact of the damage on the community. The World Bank’s Rapid Damage and Loss Assessment (DaLa) methodology provides a way of undertaking the assessment activity, where DaLa incorporates physical damage, losses from that damage, and assessments of social, economic, and environmental impacts. At impact assessment, information is collated on the magnitude of the disaster event on individuals, communities, and the physical environment. The result of the impact assessment becomes the basis for future reconstruction works. Assessments require reviews and updating to take account of new information at later stages. Stakeholders in the reconstruction process are enlisted so that a comprehensive assessment report can be prepared. The success of assessment is greatly enhanced by information gathering, collation, and dissemination approaches, coupled with the level of interaction and planning arrangements that exist between the different disaster stakeholders (Rotimi 2010). Assessment processes are not always well established, so a certain amount of chaos is evident causing confusion and reassessment.

The Canterbury earthquakes can be seen to illustrate this initial chaos stage. Canterbury was significantly damaged by two major earthquakes and thousands of aftershocks. These earthquakes illustrate the five stages of reconstruction. In the first earthquake on the 4th of September 2010, there were no fatalities but widespread damage to housing, infrastructure, and public facilities throughout the city and surrounding areas. After the September earthquake, a large number of unreinforced masonry buildings in the central business district (CBD) were heavily damaged, and large areas of the CBD were cordoned off from the public for approximately 1 week. Engineers were quickly mobilized throughout the country for building damage assessment and safety evaluation. The earthquake produced rapid response and assessment, and structures were quickly put in place to deal with the situation. However, the violent magnitude 6.3 earthquake that devastated Christchurch on

22 February 2011 was the most severe of all the events in the Canterbury earthquake sequence causing the deaths of 185 people, and many buildings were severely damaged including further damage to infrastructure and widespread liquefaction leading to a more complex response, recovery, and reconstruction. The February earthquake also triggered land movement, the collapse of cliffs, and rock falls. As a result of the Canterbury earthquakes, more than 60 % of Christchurch's CBD buildings were severely damaged (CERA 2012). Another 60 % of the 5,000 businesses in the CBD and 50,000 employees were displaced. More than one third of central city businesses were unable to operate, with another third relocating to makeshift premises (DoL 2011). Over 150,000 homes (about three quarters of Christchurch's housing stock) sustained some damage from the earthquakes. The total number of individual buildings, land, and contents insurance claims received in the first year exceeded 600,000 (EQC 2011).

In terms of infrastructure damage, 1,021 km of roading needed rebuilding, which is about 52 % of Christchurch's urban sealed roads. The earthquakes also damaged 51 km of water supply mains and 58 km of the sewer system within the city (CERA 2012).

The initial building assessments for Canterbury earthquake demonstrate the chaos of the initial stages of reconstruction. In the need to act quickly, the assessment methods and protocols were not consistent, and the quality of assessment was variable (NZSEE 2011). The ways in which building assessments were conducted represents the chaos in the initial stages of reconstruction. There were different assessment techniques being used by different agencies, leading to different levels of quality and different information in the assessments. The level of training of the assessment teams varied, leading to the same buildings having different assessment outcomes. Reassessments were commonplace. Inaccurate and incomplete building assessments were used to make decisions, leading to fast-track demolition of buildings.

The chaos stage brings into focus the need to undertake a significant rethink of land use,

involving questions such as: Should we rebuild? Where? Questions about whether to rebuild and where caused uncertainty in the Canterbury earthquakes. Significant land-use reassessment was made to categorize land for rebuilding, but this again was initially characterized by inaccuracies and confusion. Since there was widespread water and waste water system failures, people had to use portable toilets and water tanks delivered water (Potangaroa et al. 2011). Failure to rapidly repair systems and poor quality assessments are typical in the chaos stage.

Reconstruction Stage 2: Realization

Following on from the chaos stage is the realization stage which is characterized by the common thought that "the disaster's impact is bigger than we thought." In this stage, the common elements are around establishing agencies; planning, especially land use and rezoning; new legislation; and the introduction of quickly produced new building codes (which are usually designed to improve past building practices and are aimed at improving resilience). Towards the end of the chaos stage, newly established recovery organizations emerge. For the community at large, there is often ongoing temporary accommodation or temporary displacement and no or minimal rebuilding occurring.

Brunsdon and Smith (2004) discussed the need for decisions on whether to repair, replace, or demolish affected properties and also the need to produce restoration proposals to give an outline of the anticipated reconstruction needs. Proposals outline a range of options for reconstruction, and an attempt to put costs to the reconstruction is often made.

The time period for complete reconstruction is relatively indefinite. It could last months, years, or decades after the disaster event. The sooner good organizational structures, staffed with qualified professionals, are in place, the better for overall recovery. Ensuring a smooth transition from response through relief and recovery is critical to the effectiveness of post-disaster recovery activity (Smart 2012). However, Davies (2006)

points out that "...many [disasters] concern the relentless pressure for rapid recovery from all quarters which is set against the normal demands for prudent planning, detailed consultation, reviews of safety requirements etc. There is also the demand for reform to be balanced with another pressure for realism or a return to pre-disaster norms." The realization stage of reconstruction brings this point into focus, as the debates of what to rebuild, where, and when become more frequent. The Christchurch earthquakes illustrate the structure and tensions of government agencies created to lead recovery and reconstruction. The Canterbury Earthquake Recovery Authority (CERA) was established as a key driver in determining how recovery policy and programs are designed and implemented. CERA as the primary agency for recovery has powers that were stipulated in a newly created Canterbury Earthquake Recovery Act to circumvent existing regulations. The Recovery Act also stipulated the role of the Minister for Canterbury Earthquake Recovery who was appointed by the Prime Minister to coordinate the recovery effort at the executive government level. This minister reports to the Cabinet Committee on Canterbury Earthquake Recovery, which was tasked to oversee and coordinate the government's response to support the recovery and reconstruction. Other government agencies and departments are coordinated through a Senior Officials Group, chaired by the Chief Executive of CERA. Elected members, commissioners, and leaders of the strategic partners are engaged through the Recovery Strategy Advisory Committee (RSAC). The Recovery Strategy, which was released in May 2012, is the overarching document to coordinate actions among CERA, other government agencies, local authorities, and strategic partners (CERA 2012). New Zealand followed Victoria and Queensland's lead in creating a new recovery agency and providing it with enabling legislation (Smart 2012). CERA is equivalent to the Victorian Bushfire Reconstruction and Recovery Authority (VBRRA) established following the bushfires in Victoria and the Queensland Reconstruction Authority (QRA) established following the floods in Queensland. Indeed, Queensland's

institutional response was specifically cited in a New Zealand Cabinet minute proposing the creation of CERA (New Zealand Cabinet Office 2011). Successful post-disaster recovery depends on strong leadership. In Christchurch, the recovery leadership has been undertaken by CERA at the national level. While the establishment of a dedicated response and recovery organization was gaining favor, one critique of this strategy is that the local authorities lost an opportunity to manage recovery and redevelopment of their own city. Tensions often arise at the realization stage, as the leadership and management of disaster recoveries are shuffled between local and central governments.

Another critical decision in relation to the rebuild of Christchurch City Centre was the creation of the Christchurch Central Development Unit (CCDU). The high degree of insurance penetration meant that many commercial building owners have the option of cashing out their buildings and moving their money elsewhere (Taylor et al. 2012). On the other hand, greater certainty of the future is needed to attract capital investment, retain employment, and boost economic development. Within this context, CCDU was established in April 2011 as a central government organization within CERA and tasked with laying out a blueprint for the CBD within 100 days, with the aim of creating confidence in the rebuild of the CBD.

In order to manage a large disaster rebuild, other new structures are often developed. Due to the scale of the February earthquakes, the Stronger Christchurch Infrastructure Rebuild Team (SCIRT) was created to rebuild the horizontal infrastructure (roads, bridges, water systems) and responsible for delivery of all asset assessments, project definition, concept and detailed design, and construction delivery. SCIRT adopted an innovative alliance delivery system for the reconstruction of the horizontal infrastructure. The SCIRT alliance was made up of eight partner organizations, consisting of three client organizations (Christchurch City Council (CCC), CERA, and New Zealand Transport Agency (NZTA)) and five main contractors. Each client plays a different role: CCC and NZTA act as asset

owner and founder, while CERA is mandated to coordinate the overall rebuild activity on behalf of the government. SCIRT is effectively a “virtual organization” which has a leadership team for governance and an Alliance Management Team (AMT) which looks after the Integrated Alliance Team (IAT) who are responsible for delivering the planning, design, and management functions to enable the delivery teams to do the work. The delivery teams are responsible for the construction.

The creation of the Christchurch Central Development Unit (CCDU) was seen as critical for the reconstruction as it created some certainty regarding the plan for new major public buildings and in encouraging investment into the new central city. The release of the Blueprint, the Christchurch Central Recovery Plan at end of July 2011, was regarded as a milestone for the city. The Plan identifies a number of major “anchor” projects that will contribute to recovery. Anchor projects will be progressed by the relevant organizations including CERA and other government agencies and the Christchurch City Council with involvement from the private and philanthropic sectors.

The alliance model adopted by SCIRT is a collaboration between a client, consultant, and contractor who mutually agree to undertake the work to target levels of quality, cost, and time. An additional rewards/sanctions mechanism is put in place to measure the performance of individual delivery contractors over time. Construction work is allocated to them based on their performance. The need to put in more collaborative structures for rebuilding, as demonstrated by SCIRT, has been recognized by Zuo et al. (2006), and the system adopted by SCIRT takes the alliance model concept and applies it in a unique way.

In addition to new structures, the realization stage of reconstruction brings with it decisions of a social, physical, environmental, and political nature which result in conflicts between implementing a speedy recovery and the needs for safety or quality (which reduces initial vulnerabilities) or speed versus wider community participation in decision making. Ingram

et al. (2006) conclude that relief and short-term response can be urgent and rapid; but longer-term recovery needs to be cautiously implemented. Long-term recovery programs have to be based upon comprehensive assessments of risk and vulnerabilities and balanced with overall recovery needs.

At the realization stage, information about the financial impacts of the reconstruction is clearer as reconstruction proposals and the apportioning of funding become key issues. Reconstruction funds may be raised privately, through insurance companies and from external donor agencies or charities. The outcome of funding and other statutory compliance applications may necessitate adjustments to initial restoration plans. Some other factors apart from economic considerations may impact on restoration programs. These may include structural integrity, safety, and functional/historical/cultural significance of the property to the owner. As with many disasters, in the wake of the Christchurch earthquakes, the first source of financial assistance for homeowners was from their own insurance companies. Loss of housing is covered by government-required earthquake insurance, provided by a national Earthquake Commission (EQC). New Zealand administers a disaster insurance scheme that is designed to cover residential property for many geological hazards and residential land for storms and floods. The scheme operates by applying a compulsory levy on insurance policies, the funds from which are pooled into the New Zealand Earthquake Commission (EQC) and used to fund property rebuilding and repair post-disaster.

Under the EQC, the New Zealand Government inherently assumes a degree of responsibility for the risks of events that are destructive but difficult to predict. The availability of insurance and the high level of coverage is a key aspect in financing Christchurch’s earthquake housing reconstruction. The government also shares in the cost of the city’s infrastructure repair. Funding for the infrastructure rebuild is provided through a combination of government subsidies from the New Zealand Transport

Agency (NZTA), CERA, and/or treasury, insurance companies, and council borrowings. The rebuild of the Christchurch central city is largely driven by the Christchurch Central Recovery Plan (CCRP), based around 17 anchor projects to boost business recovery in the CBD. In June 2013, a year after the launch of the CCRP, the central government and Christchurch City Council reached cost-sharing arrangements for the anchor projects in the CCRP and for the repair and replacement of the central city's essential horizontal infrastructure (CERA 2013). It is common at the realization stage to face the realities of the disaster in terms of the financial burden on the community and country, and different financial mechanisms are used to fund recovery.

Hasty reconstruction programs have longer-term impacts that may be difficult to undo (Ingram et al. 2006). Reconstruction decisions should therefore be a trade-off between idealistic goals and expediency. When systems are established, there is a desire to see some action, mainly in the form of rebuilding. Realization leads into the next stage with a need to increase activity; this is the mobilization stage.

Reconstruction Stage 3: Mobilization

Often marking the start of the mobilization stage is the 1 year anniversary of the disaster and the increasing mood to get on with the reconstruction. Mobilization is characterized by the common thought "we're getting on with it." The mobilization stage features high energy and increased decision making. Often public building repairs become evident, especially schools and community centers, as a desire to make physical statements about the rebuilding progress is made. In this stage, the first new buildings emerge, there is high activity in repairs, and an elevated concern about the wider resourcing problems is being encountered, leading to increasing costs and shortages of supply. However, there is much less uncertainty from population in this stage, governance and organization structures for reconstruction are in place, and

increasing activity leads to optimism. At this stage, how the recovery is funded will be a key driver. In Christchurch, because of the very high level coverage, insurance is driving recovery and problems with insurance have become prominent.

Mobilization also brings into focus the needs of housing repairs. Housing is a large component of the disaster recovery in Canterbury. About three quarters of the housing stock in Canterbury sustained some damage from the earthquakes and a mobilized workforce is required to undertake the repairs and rebuilding. Issues concerning insurance payouts have been cited most frequently as a major hurdle to advancing repair and rebuild.

For local residents, the wait for payment from the insurance, combined with the government decision on land zoning, creates additional stress in recovery.

Mobilization is characterized by increased decision making on properties, for instance, in Canterbury, following on from land rezoning, more than 7,800 property owners in the residential red zone were offered a buyout package by the government to leave their uninhabitable houses. The vast majority of red-zone home owners chose the buyout option, selling their property to the government. This residential red zone offer scheme has allowed property owners to move on from the most damaged areas promptly, and the government's assistance has progressed recovery faster than if the usual private insurance process had been left to manage the situation. Other housing decisions describing how the land is expected to perform in future earthquakes prescribed altered foundation. More than 10,000 homes were assessed as requiring substantial foundation work before the houses may be considered safe for living.

Tracking the reconstruction of Christchurch showed that by September 2013, 2½ years after the February 2011 earthquake, demolition of damaged buildings in the CBD was almost complete, and the need for a more tangible sense of momentum and progress through the visible proof of new buildings in the CBD was seen as immediate priorities (NZCID 2013).

Reconstruction Stage 4: Struggle

The increased optimism of mobilization is often followed by a struggle stage. The main mood of the struggle stage is one characterized by the feeling that “it’s really hard, it’s not going to plan.” What appears in this stage is the realization that there will be no fast recovery. Statutory application and documentation procedures have been known to slow down reconstruction programs (Burby et al. 2006). The entire process is worsened by the absence of skilled professionals and skills, such as processing officials, project managers, and essential construction skills. Thus, the period when statutory compliances/consents are pursued is usually characterized by disillusionment of affected individuals and the community. Delays and failures from the new structures and insufficient progress are features of this stage.

Recovery takes a very long time and is always beset with unanticipated problems which slow the recovery timeline. In the struggle stage, there is high activity, such as houses and public and commercial properties rebuilding and businesses returning. New business models emerge as prior models may not cope with the changed business environment. But, at the struggle stage, there are often costs escalating, materials hard to procure, skilled professionals in demand, reduced housing stock, and housing affordability problems. The struggle stage is very hard for the community, and more people start to reassess options, and there are population shifts from the region. In the pursuit of reconstruction objectives, it is usual for conflicts to occur between affected groups, government, and recovery providers. New groups that emerge after a disaster event may have difficulty working together with already established ones (Quarantelli 2008). Auf De Heide (1989) gives three reasons for conflict after a disasters:

- Scarcity of information and/or breakdown in communication among recovery stakeholders
- Challenges posed by the management of limited recovery resources
- Excessive response and recovery provisions by external aid agencies and outsiders

Conflict must be properly managed; otherwise they could have lasting effects on individuals and the community and affect the reconstruction. Picou et al.’s (2004) study suggests that the scale of litigations among stakeholders during reconstruction programs could become very extensive because of disaffection with the recovery process (and when recovery needs are unmet).

The reconstruction process involves the application for resource consents and building approvals. Consenting processes are usually painstaking for both the party(s) seeking approvals and the approving authorities. Approving authorities need to ensure that performance quality and safety provisions are not compromised. It is necessary to ensure that a considerable level of resilience is incorporated in all post-disaster development proposals.

Three years after the February 2011 earthquake, the pace of physical rebuild can be seen to represent the struggle stage, where reconstruction progress is gradual. In spite of the various stakeholders involved and different funding mechanisms and organizational structures applied in the rebuild sectors, other issues have emerged which have started to represent a struggle for the region’s reconstruction.

For instance, the Canterbury earthquakes exposed the lack of construction-related professionals for the rebuild. Demand for skills, such as demolition, repairs and construction, and associated professional services, has increased considerably in Canterbury since February 2011 (Chang-Richards et al. 2013). There have been difficulties in recruiting workers with desired skills. Resource pressures on rebuild projects remain primarily from human resources associated with structural, architectural, and land issues relating to rebuilding. Resource shortfalls have had an inflationary impact. The inflationary effects of increased construction professional fees and an increase in temporary house rentals for the inbound construction workforce have become major concerns.

Some engineering consultancies have reported ongoing issues sourcing people with high skill levels. Young engineers and mature project

management skills from Europe continue to be the largest inbound demographic group involved with the rebuild. Some construction organizations faced difficulties with immigration issues and housing their incoming workforces. At the same time, the inflationary impact which flowed through to higher property rents made attracting tradespeople from other parts of New Zealand more difficult.

According to official projections, there will be an additional 17,000 workers needed to meet demand at the height of the rebuilding activity in late 2014. Flowing on from the rebuild, another 15,000 workers will be needed in supporting roles and sectors, including such as administration, law, accounting, retail, accommodation, and services (MBIE 2013).

Given that the habitable housing stock has been greatly reduced in the earthquakes, Christchurch City has found it difficult to ensure the market provides enough affordable housing for displaced residents. Compounding this shortage was the need to house a large number of additional rebuild workers. A lack of temporary accommodation has been an ongoing concern which constrains labor supply.

The government created three temporary villages and developed about 40 permanent units to fill the gap of unmet housing needs for displaced residents. Housing New Zealand has initiated a progressive approach to rebuilding quake-damaged social housing properties in the eastern suburbs.

A survey undertaken by the Tenants Protection Association (2013) in June 2013 investigated the effects of rising rents on tenants. Of 365 surveyed tenants, 85 % reported to be negatively affected by rent increases. 60 % reported paying over 40 % of their annual income in rent. The average rent increase was \$43 dollars per week. Some areas such as “Inner North” and “North West” Christchurch recorded above-average rent increases of 39 % and 32 % respectively (MBIE 2013).

An estimated minimum of 10,000 homeowners and occupants will have to relocate to temporary accommodation while further repairs are carried out. As the rebuild proceeds,

construction-related inflation as a result of this rent inflation is likely to put extra pressure on the Canterbury labor market, community recovery, and regional economic development. The shortage of accommodation also appears to be a constraint on the rate at which the labor force is brought into the region.

One of the consequences of the sheer damage to housing is that there are thousands of buildings that need remedial work or rebuild for which a building consent will be required. This has led to the Christchurch City Council Building Consent Authority being overwhelmed by the volumes of building consents with criticism around the council’s inability to meet statutory timeframes for processing building consents.

Christchurch applicants had been expressing a significant level of concern over the time needed for consent. Given the projected rebuild activity for the coming years, the central government had to intervene to aid Christchurch City Council in changing its building consent functions with a more streamlined process around commercial and residential building consents.

There was also a lack of consistency of work procedures, difficulty in meeting requirements in terms of quality of work, and a lack of people to do the work. There was also a need for the construction industry to respond, plan, and train for the new/improved requirements of the rebuild.

The process from earthquake impact to longer-term reconstruction is complicated by a lack of prior capacity and experience in both local authorities and the insurance industry to deal with multiple events of large scale. The first 2 years’ reconstruction process was rife with uncertainties about land decisions, timeline of insurance payouts and housing repairs, as well as the eventual cost for the rebuild.

The rebuild has been slow because of the ongoing aftershocks, scale of the disaster, delays in resolving multi-event insurance claims, and skills shortages. Canterbury is still very much in the struggle stage of reconstruction, and this will continue for many years, as the new buildings, anchor projects, and housing are gradually rebuilt. The new normal stage follows the struggle stage.

Reconstruction Stage 5: New Normal

The new normal stage, which starts many years after the disaster, is characterized by the feeling that “this is how it is, there’s no going back.” There is always a new normal stage following a disaster. This stage finds landscapes now looking very different. Disaster recovery and reconstruction do not recreate the same environments seen prior to the disaster. Everything is different; the way people conduct business has changed; building work is still ongoing, but the architecture is different, and there are often delays; community configurations are different; and community and social facilities are relocated and often reconfigured. However, this new normal stage is a time where the community starts accepting the new normal environment. Where resilience has been introduced into the buildings and community, there is a sense of buildings being better, newer, and safer. The community starts living differently from prior to the disaster. For Canterbury, currently in the struggle stage of the recovery, there is hope for the future as the new normal stage comes closer.

Schneider (1992) suggests that stakeholder conflicts may also result from the emergence of new social norms after a disaster event which may not mesh well with traditional norms. From an individual or group perspective, stakeholders then struggle to reestablish or maintain their previously recognized roles, responsibilities, and boundaries. Chief among post-disaster management objectives is to enable a community to recover from the event while also future-proofing the community and its physical facilities against similar disaster events. New normal brings this debate into focus, where people feel that they are more likely to be prepared for any future disasters and that their facilities are better able to cope.

Summary

Development of the five stages of reconstruction (initial chaos, realization, mobilization, struggle, and new normal) through the lens of a recent

earthquake demonstrates the difficulty of disaster reconstruction. However, understanding the discernible stages can help inform the planning and reconstruction of subsequent disasters. The experience of the Christchurch rebuild has confirmed much of what is known about post-disaster recovery, including the importance of leadership, funding availability, participation of local communities, and timing of decisions. Reconstruction is a complex and dynamic process, full of uncertainties and stress, requiring a significant level of coordination and innovation. Reconstruction of housing, infrastructure, and commercial buildings requires an understanding of the stages, and how recovering cities move from one stage to another, and the elements requiring consideration on the reconstruction timeline. Although disaster reconstruction moves at different speeds, depending on the differing circumstances of the affected area, the reconstruction stages remain relatively static.

Cross-References

- ▶ [“Build Back Better” Principles for Reconstruction](#)
- ▶ [Economic Recovery Following Earthquakes Disasters](#)
- ▶ [Reconstruction in Indonesia Post-2004 Tsunami: Lessons Learnt](#)
- ▶ [Resilience to Earthquake Disasters](#)
- ▶ [Resourcing Issues Following Earthquake Disaster](#)

References

- Auf De Heide E (1989) Disaster response: principles and preparation and coordination. Mosby, St. Louis
- Brundson D, Smith S (2004) Summary notes from the infrastructure workshop. Paper presented at the NZ recovery symposium, Napier
- Burby RJ, Salversen D, Creed M (2006) Encouraging residential rehabilitation with building codes: New Jersey experience. *Journal of the American Planning Association*, 72, 183–186
- CERA (2012) Recovery strategy for greater Christchurch and economic recovery programme for greater Christchurch: a foundation for economic recovery and

- growth in greater Christchurch. Canterbury Earthquake Recovery Authority (CERA), Christchurch, Dec 2012
- CERA (2013) Christchurch central recovery plan – one year on. Greater Christchurch Recovery Update August 2013 (24):1–3; Third anniversary of 2010 earthquake brings major claim milestones. Greater Christchurch Recovery Update October 2013 (26)
- Chang-Richards Y, Wilkinson S et al (2013) Resourcing of the Canterbury rebuild: case studies of construction organisations. The University of Auckland, BRANZ, Resilient Organisations, Auckland. www.resorgs.org.nz/publications. Jan 2013
- Davies I (2006) Learning from disaster recovery: guidance for decision makers. UNDP, Geneva
- DoL (2011) A changing landscape: the impact of the earthquakes on Christchurch workplaces. Labour & Immigration Research Centre, Department of Labour, Wellington
- EQC Earthquake Commission (2011) Briefing to the incoming Minister. Dec 2011. www.eqc.govt.nz
- Ingram JC, Franco G, Rio CR-d, Khazai B (2006) Post-disaster recovery dilemmas: challenges in balancing short-term and long-term needs for vulnerability reduction. *Environ Sci Pol* 9(7–8):607–613
- MBIE (2013) Housing pressures in Christchurch: a summary of the evidence. Ministry of Business, Innovation and Employment (MBIE), Wellington. Available at <http://www.building.dbh.govt.nz/nz-housing-research>
- MCDEM (Ministry of Civil Defence & Emergency Management) (2005) Focus on recovery: a holistic framework for recovery in New Zealand. MCDEM, Ministry of Civil Defence & Emergency Management, Wellington, IS5/05
- New Zealand Cabinet Office (2011) Canterbury earthquake recovery: proposed governance arrangements. cabinet-paper-1-proposed-governance-arrangements-annex-2-march-2011
- NZCID (2013) Insights into Canterbury: findings of research on the Canterbury earthquakes recovery. New Zealand Council for Infrastructure Development (NZCID), Auckland
- NZSEE (2011) New Zealand Society of Earthquake Engineering (2011) Building safety evaluation following the Canterbury earthquakes. Report to the Royal Commission of inquiry into building failure, NZSEE, Wellington
- Picou JS, Marshall BK, Gill DA (2004) Disaster litigation and the corrosive community. *Social Forces* 82(4):1493–1522
- Potangaroa R, Wilkinson S, Zare M, Steinfort P (2011) The management of portable toilets in the eastern suburbs of Christchurch after the February 22, 2011 Earthquake. *Australasian Journal of Disaster and Trauma Studies*, 2 (Special Issue - A Focus on the Canterbury Earthquakes), 39–48
- Quarantelli DR (2008) A brief note on disaster restoration, reconstruction and recovery: a comparative note using postearthquake observations. Preliminary paper #359, Disaster Research Center, University of Delaware
- Rotimi JOB (2010) An examination of improvements required to legislative provisions for post disaster reconstruction in New Zealand. Doctoral thesis, University of Canterbury, Christchurch
- Schneider S (1992) Governmental response to disasters: the conflict between bureaucratic procedures and emergent norms. *Public Adm Rev* 52:135–145
- Smart J (2012) The role of post-disaster institutions in recovery and resilience: a comparative study of three recent disasters – Victorian Bushfires (2009), Queensland Floods (2010–11), and Canterbury Earthquake (2010–12). Working paper 12/02. Institute of Governance and Policy Studies, Victoria University, Wellington
- Taylor JE, Chang SE, et al (2012) Learning from Christchurch: technical decisions and societal consequences in post-earthquake recovery. Resilient Organisations research report 2012/08. Resilient Organisations, The University of British Columbia, The University of Canterbury, Christchurch
- Tenants Protection Association (Christchurch) (2013) TPA (CHCH) Rental Survey 2013: a study of increasing rents and housing conditions in the greater Christchurch area, June 2013. Tenants Protection Association Christchurch (TPA), Christchurch
- Wilkinson S (2013) Reconstruction stages, build, November Issue. BRANZ, New Zealand
- Zuo K, Wilkinson S, Le Masurier J, Shestakova Y (2006) Procurement systems for disaster reconstruction, 100th Anniversary earthquake conference. In: Proceedings of the 8th US national conference on earthquake engineering, San Francisco, pp 18–22

Reconstruction in Indonesia Post-2004 Tsunami: Lessons Learnt

Teddy Boen

PT Teddy Boen Konsultan, Jakarta, Indonesia

Synonyms

Build back better; Leadership; Post-disaster recovery

Introduction

On December 26, 2004, a powerful earthquake of magnitude 9.2 occurred in the Indian Ocean,

approximately 150 km off the west coast of Meulaboh, Sumatra Island, Indonesia. The resulting tsunami affected several countries around the Indian Ocean, with Indonesia suffering the greatest. In Aceh, the northern province of Sumatra, 131,000 were confirmed dead and 37,000 missing and approximately 500,000 internally displaced people (IDP) (Iwan, 2006). The tsunami caused a massive damage to housing and infrastructure, and the disaster significantly impacted the social fabric and government administration of the affected communities.

To ensure a speedy, effective rehabilitation and reconstruction of Aceh and Nias, on April 16, 2005, the government established the Rehabilitation and Reconstruction Agency (BRR, Badan Rehabilitasi dan Rekonstruksi NAD-Nias). BRR was an agency with the authority to plan, implement, control, and evaluate the process of rehabilitation and reconstruction. BRR's term of operation was 4 years and was based in Banda Aceh with a branch office in Nias and a representative office in Jakarta. BRR reported the progress of activities directly to the President of the Republic of Indonesia. The main mission was to restore livelihoods and strengthen communities in Aceh and Nias by designing, monitoring, and coordinating community-based reconstruction and development programs.

The December 26, 2004, tsunami that caused extensive damage in Aceh created an opportunity for mitigation, particularly with large sums of money available for rebuilding due to domestic appropriations and international relief. Opportunities arise to intervene in the course of earthquake. A large proportion of the fund was utilized for the reconstruction of approximately 127,400 houses, and unfortunately most of the houses built are not seismic resistant. The February 20, 2008, Simeulue earthquake demonstrated the weaknesses of the houses built; many were heavily damaged or collapsed, even while being constructed. Poorly built houses indicate that the mitigation opportunities to create an earthquake-resistant community were not implemented. The cause for ignoring mitigation opportunities could be lack of knowledge regarding disaster risk management and

the heavy dependency on external resources (financial and "experts") causing the loss of local control.

Having observed the rehabilitation and reconstruction of the houses in Aceh and Simeulue for almost 6 years, the author tries to summarize those shortcomings so that it can serve as lessons learned to other reconstruction projects all around the world, particularly in developing countries.

Below is a summary of reconstruction issues that were brought forward by the author and reconfirmed by others that resulted in poorly built houses in Aceh:

Needs Assessment and Site-Specific Information

A detailed and accurate disaster assessment is a very important factor for a successful planning and execution of the rehabilitation and reconstruction. Such assessment is expected to produce reliable data of beneficiaries of houses. The reliable data is vital prior to commencing the actual reconstruction (Boen 2008). However, in Aceh, there was limited awareness of the need for surveys and lack of expertise to specify or carry these out. The adequate surveys were not systematically completed (Ove Arup & Partners Ltd. 2010).

One of the major flaws was the poorly prepared beneficiary lists, which in turn caused many other problems; among others, the number of IDP and housing needs is constantly changing depending on the various sources. Beneficiary identification and verification was prone to corruption, for example, selling of family ID cards to outsiders by the head of the village, playing off one agency against another, or including relatives returned from other areas of Indonesia to receive multiple houses (IRIN Humanitarian News and Analysis 2009; Ove Arup & Partners Ltd. 2010). Besides that, many foreign "experts" introduced all sorts of house types which are not the prevailing practice, and many are culturally unacceptable. As the result of incorrect needs, to date many houses built during the reconstruction and rehabilitation period are still unoccupied or become wild occupations, and many already deteriorated.

The poor quality of houses was also a result of lack of assessment concerning the real need of materials, construction workers, and technical supervisors.

Directives From the Authority

Any successful reconstruction needs a capable authority with strong leadership that can provide clear directives and requirements concerning the type of buildings and the standards to be followed from the onset. Capable means professionals are familiar with construction of buildings as well as infrastructures and are knowledgeable about the disaster being dealt with.

The approval/permit system must be strictly enforced, and continuous qualified technical assistance and inspection need to be provided.

In Aceh, most nongovernmental organizations (NGOs) felt that BRR did not provide clear-cut guidance, particularly with regard to quality of the construction (Boen 2008; Ove Arup & Partners Ltd. 2010). In July 2005, BRR published the Building Code of the Province of Nanggroe Aceh Darussalam, which provided detailed technical requirements for houses. However, in this building code, there was no guidance on seismic resilient design, and this building code made no reference to national or international standards which were confusing for the donors/NGOs. Therefore, NGOs developed their own guidance. Unfortunately, the guidance in some cases provided incorrect or conflicting details. Several NGOs complied with the Building Code for Aceh assuming that it was sufficient and that local designers and contractors knew what they were doing without realizing that safe construction practices were not common practice. Multiple guidelines caused confusion as to what was deemed appropriate, rather than providing clarity as to which codes and standards should apply (Boen 2008; Ove Arup & Partners Ltd. 2010).

Relevant guidance for good practice of non-engineered buildings for new construction existed or had been developed for Indonesia, long before the reconstruction in Aceh and Nias,

since 1978 (Boen 1969, 1978; Arya et al. 1980; Ove Arup & Partners Ltd. 2010). However, lack of coordination, knowledge, and leadership within the shelter sector resulted in that these good references were not widely distributed and were frequently either not known about (Ove Arup & Partners Ltd. 2006, 2007). All those relevant materials were ignored, and instead, many foreign consultants made their own layout and adopted the confined masonry construction method, but leaving out the detailing for seismic resilience (Ove Arup & Partners Ltd. 2006, 2007; Boen 2008; Ove Arup & Partners Ltd. 2010).

From surveys in Aceh and Nias, almost all designers did not follow the existing guidelines and did not perform the necessary appropriate analysis and design for earthquake resistance (Ove Arup & Partners Ltd. 2006; Boen, 2008b, c; Ove Arup & Partners Ltd. 2010).

Differentiation Between Emergency Shelters and Transition and Permanent Houses

Reconstruction of houses after a disaster must be planned within the overall context of phases from emergency shelter to durable solutions. Differentiation between emergency shelters, transition houses, and permanent houses shall be made.

In Aceh, shelter policy was not developed within the overall context of the journey from emergency shelter to durable solutions, and over the first 6 months, there were considerable confusion and no clear policy as to the type of shelter assistance required (Ove Arup & Partners Ltd. 2010). In the early stage, many of the houses already built were of the transition type but built on permanent former lands belonging to recipients. Many of those “temporary/transition” houses become permanent, and the final reconstruction stage failed to materialize (Boen 2005; Ove Arup & Partners Ltd. 2007).

The need for “permanent” housing in part reflected the large amount of funding available and was articulated in terms of reconstruction rather than the recovery. This led to a focus on physical construction rather than how the process

of rebuilding can lead to economic activity and the role that shelter plays in meeting needs and allowing families to return home and carry out their livelihoods.

Influx of Local and Foreign “Experts”

In Aceh, although numerous contracting firms established themselves after the tsunami, there was no process of certification to guarantee their competency. Out of over 200 NGOs engaged in reconstruction, most had no previous experience in the trade of earthquake-resistant houses construction, and some of them had to terminate agreements mid-contract due to poor workmanship or faced expensive remedial works (Boen 2008; Ove Arup & Partners Ltd. 2010). The lack of immediately available site-specific information was coupled with the influx of many so-called experts (local and foreign) offering an endless number of earthquake-resistant building type “solutions” causing unsatisfactory results. Many NGOs ended in trying to “reinvent the wheel” by introducing house types defying the local culture. Only few NGOs constructed houses based on the prevailing culture in Aceh (Boen 2005; Ove Arup & Partners Ltd. 2007; Boen 2008).

Another major drawback was the fact that NGOs, world organizations, donors (countries), Red Cross, and Red Crescent did not have the necessary experience as well as expertise in post-disaster reconstruction of houses combined with lack of strong leadership from the BRR side in providing guidance as well as data for the post-disaster reconstruction. Some of them had experience in emergency response, and therefore, those “experts” lacked technical capacity and a clear understanding of the local building culture and the social order of the community as well as the ability to adapt disaster-resistant techniques to local styles and situations. Many of them had never previously worked in a post-disaster situation, particularly in reconstruction, and had not worked in Aceh before. They were often unaware of the specific objectives to reduce vulnerability and initially considered seismic resilience as an

optional rather than an essential requirement (Boen 2008; Ove Arup & Partners Ltd. 2010).

Even though Indonesia is highly vulnerable to earthquakes, seismic engineering does not form a core topic in undergraduate engineering degree programs leading to Bachelor of Engineering (Civil). During the reconstruction of Aceh, many of them had only recently graduated and had no practical experience or knowledge of seismic design and to specify and verify materials. Therefore, they failed to anticipate or spot problems and needed training and supervision. The knowledge of local engineering consultants could not be assumed to include seismic design particularly for earthquake-resistant design of houses (Boen 2008; Ove Arup & Partners Ltd. 2010). This was in spite of the fact that relevant guidance for good practice of non-engineered buildings for new construction existed or had been developed for Indonesia since 1978 (Boen 1969, 1978; Arya et al. 1980; Ove Arup & Partners Ltd. 2010).

Designs of Houses Are Culturally Inappropriate and Not Sustainable

The majority of the buildings, particularly houses that were destroyed by the tsunami in Banda Aceh City and villages in Lhoknga and Krueng Raya and Meulaboh City and villages along the west coast of Aceh, such as Calang, were non-engineered buildings consisting of two types: the first type was a one or two stories confined masonry buildings which were earthquake resistant if built with good quality materials and good workmanship and the second type was timber construction (Boen 2005).

However, as the economic condition is prospering, people tend to upgrade their timber houses into masonry because a measure of status is associated with the owners of such masonry houses. Therefore, for the reconstruction program, confined masonry houses shall be constructed as permanent houses because they are in accord with the local culture and will therefore be sustainable (Boen 2006a). In most cases, the quality and strength of masonry buildings

must be improved; however, it would be wrong if instead of enhancing the trade, they try to reinvent the wheel by introducing “alien” types of houses (Boen 2005).

Many NGOs introduced alien construction methodologies that will not be sustainable apart from being seismic resistant (Boen 2006a, 2008c). They did not realize that reemergence of alien construction methodologies was not preferred by the Indonesian people and culturally inappropriate. Some of the alien methodologies observed by the author in the second and fourth quarter of 2007 were precast construction, interlocking masonry, and light-steel construction (Boen 2008c; Ove Arup & Partners Ltd. 2010).

The trouble with these “alien” types of houses is that the buildings last as long as the fund is still available. As soon as the fund is stopped, the village people cannot employ the skills they have learned because they cannot afford the materials (Boen 2008b).

No Standard Construction Drawings nor Building Specification

Most agencies jumped to developing design solutions without attempting to comprehensively define quality or develop a brief or building specification. There was no shared understanding among communities, BRR, and the NGOs as to what quality comprised.

Many designs were developed without engineering input. Standard good practice such as the incorporation of ring beams, ties, and adequate laps between reinforcement was not shown on construction drawings, and specifications did not adequately cover material quality, testing, and workmanship. The facilitators and consultants hired were civil engineers and architects with no specific seismic experience and were unaware of the importance of ductile detailing. The design did not account for seismic loads, and the construction drawings prepared were inadequate, not highlighting the importance of ductile detailing (Boen 2008b, c; Ove Arup & Partners Ltd. 2010).

All NGOs were under pressure to build quickly to meet donor timescales and beneficiary expectations. Most of the time, this resulted in the use of substandard materials, especially poor-quality bricks. Brick quality was typically left unspecified, and the strength of bricks was not qualitatively tested on delivery. The quality of the bricks used in many housing constructions were in general below standard. The sizes were not uniform; it could be easily snapped in half or crushed underfoot, and some even “melted” if exposed to rain (Boen 2006a; Ove Arup & Partners Ltd. 2010).

The quality of the mortar for the masonry walls were also not well controlled, and the proportions of the mortar mixes were also not known and were left at the discretion of the foreman and the construction workers. In some places, the quality of mortar sand was good (the mud content is low), but in most cases the mud content was a bit high.

Main Target Is Numbers and Not Quality

Through the papers that the author published regarding the reconstruction of houses in Aceh, the author suggested to go beyond numbers for the reconstruction of houses and to concentrate on the technical as well as quality of the houses (Boen 2006a). However, until BRR was dismissed (April 2009), the main target was still numbers and not quality (Boen 2006a, b; Nazara and Resosudarmo 2007; Boen 2008c; Ove Arup & Partners Ltd. 2010). BRR still emphasized on providing houses rather than assistance to reconstruct. The focus was on physical construction rather than responding to the way that the process of rebuilding can lead to economic activity or the role that shelter plays in meeting needs and allowing families to return home and carry out their livelihoods (Ove Arup & Partners Ltd. 2010).

BRR claimed that the reconstruction of houses in Aceh and Nias was the fastest in the world (Multi Donor Fund for Aceh & Nias 2007). Some NGOs also claimed that reconstruction and rehabilitation of Aceh and Nias was

successful and could be instrumental for effective disaster response, both as immediate and longer-term strategies in future post-disaster responses in Indonesia and around the world (Multi Donor Fund for Aceh and Nias 2009). However, BRR and NGOs did not prioritize the quality of the built houses and missed the opportunity to “do it right the first time” with the aim to prevent vulnerabilities in future earthquakes. Most of the houses built so far were not earthquake resistant and need to be retrofitted (Boen 2005, 2006a; Ove Arup & Partners Ltd. 2006; Boen 2006a; Ove Arup & Partners Ltd. 2007; Boen 2008a).

Site Development and Infrastructure

The planning of site engineering design and infrastructure must run parallel to the planning of the site layout (Boen 2006a) and must be constructed during the site preparation stage. However, in Aceh, the engineering design still was left out (Boen 2008b; Ove Arup & Partners Ltd. 2010). The drainage, sanitary facilities, and water supply facilities were not constructed parallel with the construction of the housing complexes.

BRR negotiated agreements with the electricity and water boards to provide free connection to the houses built after the tsunami, but responsibility for notifying the electricity and water boards where connections were required was not clear. Many NGOs expected this to be done by BRR and they themselves tended only to make a recommendation to connect once the houses were completed. Consequently, supplies were seldom in place when the houses were first occupied. This resulted in many houses remaining unoccupied (Ove Arup & Partners Ltd. 2010).

Misinterpretation of Community-Based Reconstruction

Local communities will need to be actively involved in planning, decision-making, and implementation in most sectors if reconstruction is to be successful. At the onset of the reconstruction of houses in Aceh and Nias, one of

BRR’s missions was community-driven or community-based reconstruction. Community-based construction is a bottom-up model; beneficiaries are involved, jointly with the professionals, in planning and implementation (Boen 2008b; Boen 2008c). They are engaged in decisions about the project through discussions among themselves and professionals. Participation stimulates self-reliance, because people who participated in their own house building will be confident about problems and less dependent on outside agencies.

However, in Aceh, not all will contribute, and, in most communities, the collaboration with professionals will be largely left in the hands of representatives or local leaders. At the beginning of reconstruction of houses in Aceh, most of NGOs and other organizations provided the fund but left the purchasing of materials and hiring of construction workers to the beneficiaries. They forgot that with community-based mass housing, the quality control becomes a difficult issue. In the course of the reconstruction, due to many complications and misinterpretation of community-based construction, BRR, NGOs, and other donors moved to project type of construction by assigning a local contractor to do the job, and the beneficiaries were not involved in the construction of the houses.

Core Housing, Building Completion, and Further Extension

Core houses are frequently introduced but seldom really understood. Many architects working for NGOs as well as world organizations were suggesting that the 36 m² house should be considered as a core house that can be extended by the beneficiaries at a later date. Such opinion became very common in Aceh and Nias, while in fact, in earthquake-prone countries like Indonesia, it is not advisable to encourage beneficiaries to extend their houses unless the extension is predesigned and all the connections for the extension are already in place.

The extension and original core house must be united structurally to act as one integral unit when

shaken by earthquakes. The newly extended house must be reanalyzed because the new extended house will behave differently than the original core house. Or the extension house shall be structurally completely separated from the core house.

Summary

To prevent similar mistakes as in the reconstruction of Aceh, it is high time to produce a clear guideline for governments:

- Governments must develop and implement a solid strategy for the reconstruction or possible retrofit of the nation's buildings and infrastructure.
- Governments must appoint a professional to head the reconstruction with strong leadership qualification and not merely a political figure and must have track record in disaster mitigation.
- Government must allocate resources, tasks, and time to guarantee continuity in organizational structures and procedures across changing budget years, personnel assignments, and administrative regions.
- Governments should identify and utilize existing resources (including labor, expertise, materials, and funding) available within the country.
- Governments must organize themselves appropriately in order to screen intelligently resources and skills as they are offered and at the same time to resist/reject unneeded or unwanted supplies, personnel, experts, and advice.
- Governments should screen where support from other countries and organizations will be necessary.
- Donors and world organizations should avoid offering unneeded and unwanted supplies, personnel, advice, and particularly "experts" and avoid being arrogant by assuming that the disaster-stricken countries in the developing world cannot take care of themselves.

Cross-References

- ▶ [“Build Back Better” Principles for Reconstruction](#)
- ▶ [Earthquake Disaster Recovery: Leadership and Governance](#)
- ▶ [Interim Housing Provision Following Earthquake Disaster](#)
- ▶ [Learning from Earthquake Disasters](#)
- ▶ [Resilience to Earthquake Disasters](#)
- ▶ [Resourcing Issues following Earthquake Disaster](#)

References

- Arya A, Boen T, Grandori G, Moinfar A, Martemianov A, Grases J et al (1980) Guidelines for earthquake resistant non-engineered construction. The International Association for Earthquake Engineering, with financial support from UNESCO, Tokyo
- Boen T (1969) Dasar-Dasar Perencanaan Tahan Gempa. Direktorat Penelitian Bangunan, Departemen Pekerjaan Umum, Bandung
- Boen T (1978) Manual Bangunan Tahan Gempa – Rumah Tinggal (Detailer's manual for small buildings in seismic areas). Teddy Boen & Associates, Jakarta
- Boen T (2005) Reconstruction of houses in Aceh, seven months after the earthquake & Tsunami, Dec 26, 2004. In: Fourth symposium on new technologies for Urban Safety of Mega Cities in Asia (USMCA), Singapore
- Boen T (2006a) Building a safer Aceh, reconstruction of houses, one year after the Dec 26, 2004 Tsunami. In: 40th Anniversary of Trisakti University, “answering the challenges in today's civil engineering”, Jakarta
- Boen T (2006b) Reconstruction of houses in Aceh, twenty months after the Tsunami of Dec 26, 2004. In: Twelfth Japan Association Earthquake Engineering (JAEE) symposium, special session, Japan
- Boen T (2008a) Indonesian earthquake problem. In: International conference on earthquake engineering and disaster mitigation (ICEEDM), Jakarta
- Boen T (2008b) Lessons from the reconstruction of houses in Aceh after the December 26, 2004 Tsunami. In: Housing earthquake safety initiative (HESI) international symposium, Japan
- Boen T (2008c) Reconstruction of houses in Aceh, three years after the December 26, 2004 Tsunami. Paper presented at the international conference on Earthquake Engineering and Disaster Mitigation (ICEEDM), Jakarta
- IRIN Humanitarian News and Analysis (2009) Indonesia: Tsunami rebuilding a “Success” but challenges remain. IRIN

- Iwan WD (2006) Preface to the special issue on the great Sumatra earthquakes and Indian ocean tsunamis of 26 December 2004 and 28 March 2005, EERI Earthquake Spectra, Issue III, Volume 22, June 2006
- Multi Donor Fund for Aceh & Nias (2007) *Three Years after the Tsunami, Delivering Results, Supporting Transition, Progress Report IV, December 2007*
- Multi Donor Fund for Aceh, Nias (2009) Five years after the Tsunami: continuing the commitment to reconstruction, MDF Progress report, December 2009
- Nazara S, Resosudarmo B (2007) Aceh-Nias reconstruction & rehabilitation; progress and challenges at the end of 2006. Asian Development Bank (ADB)
- Ove Arup & Partners Ltd. (2006) Aceh & Nias Post Tsunami Reconstruction – Review of Aceh Housing Program, Ove Arup & Partners, April 2006
- Ove Arup & Partners Ltd. (2007) Tsunami appeal assurance mission, REP/121348/001, Disaster Emergency Committee, Ove Arup & Partners, 28 Nov 2007
- Ove Arup & Partners Ltd. (2010) Lessons from Aceh, key considerations in post-disaster reconstruction, Disaster Emergency Committee, Ove Arup & Partners, Nov 2009

common way of recording the signal nowadays is digitally. The signals from the sensor can be very small, in the order of microvolts, and cannot be used directly. The signals can also include undesired frequency content so before recoding the signal, it may have to be amplified and possibly filtered. In addition, the signals must be time stamped.

A recorder and a sensor is called a seismograph and consists in most cases of separate units. The reason for the separation is that the sensitive seismometer has to be placed at the most low noise site possible, with little disturbance from the recorder and people operating it. For field use, the ease of deployment makes it desirable to have one compact unit; in particular strong motion instruments – less sensitive – are often sold as complete units with sensor and recorder integrated.

The different steps needed in the process of recording seismic signals will be described.

Recording Seismic Signals

Gerardo Alguacil¹ and Jens Havskov²

¹Instituto Andaluz de Geofísica, University of Granada, Granada, Spain

²Department of Earth Science, University of Bergen, Bergen, Norway

Synonyms

Accelerometer; Digitizer; GPS timing; Recorder; Seismometer; Transmission of signals

Introduction

Seismic signals are detected by seismic sensors that can be seismometers, detecting ground velocity, or accelerometers, detecting ground acceleration. The sensors give out an electrical signal proportional to the ground motion. In order to use the signals for analyzing the ground motion, they must be recorded. A recording on paper is called a seismogram; however, the most

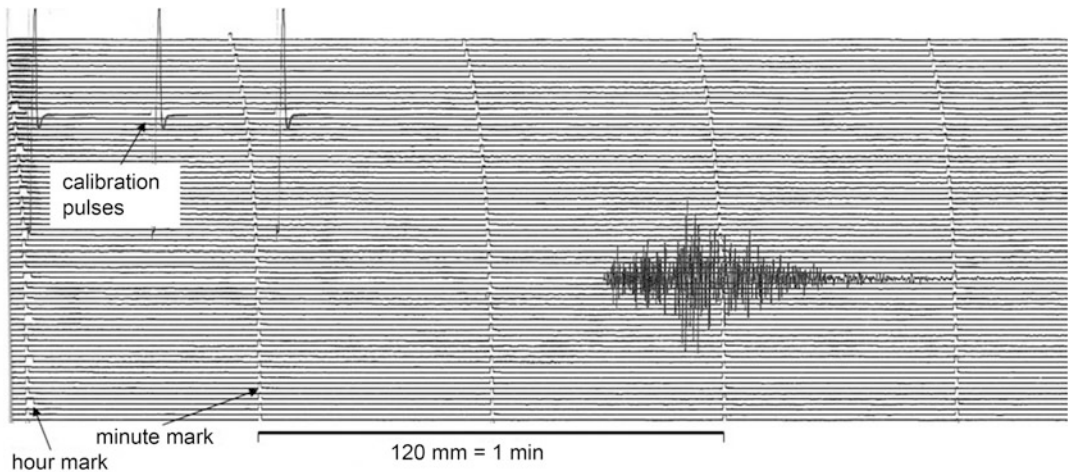
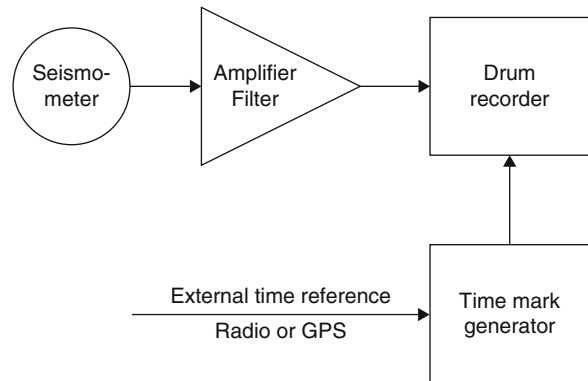
Analog Recording

Recording in analog form on paper continues to some degree despite the advance of digital technology (see Lee and Stewart 1981, for more details). Very few, if any, analog recorders are sold today, but there are probably hundreds still in operation. An analog recorder provides a simple, very reliable real-time backup recording. They are also popular for public relations. Thus, some observatories continue to have a few stations recording on analog recorders. A brief description of analog recorder will therefore be included here.

All analog paper recorders are based on a rotating drum with a pen which moves along an axis parallel to the rotating axis to provide a continuous trace for the whole recording period, usually 24 h. Old photographic recorders usually moved the drum itself along its rotating axis. For this reason, these recorders are often called helical drum recorder or helicorders. A block diagram of a complete system is represented in Fig. 1.

Recording Seismic Signals, Fig. 1

A complete analog recorder. The time mark generator is an accurate real-time clock which might be synchronized with an external time reference. GPS stands for Global Positioning System (Figure from Havskov and Alguacil 2010)



Recording Seismic Signals, Fig. 2 An example of an analog recording from a seismic recorder. The figure shows part of a seismogram. The time marks are 60 s apart

The time marks (Fig. 2) are usually generated at each minute, hour, and 24 h and have different lengths. The time mark generator can usually be synchronized with an external time reference.

The earlier recorders were recording on photographic paper, like the famous *World Wide Standard Seismographic Network* (WWSSN) system (for details, see Willmore 1979), while all recorders now record with either ink or a thermal pen on heat-sensitive paper. Figure 2 shows an example of a recording and Fig. 3 shows an example of a drum recorder with a seismogram.

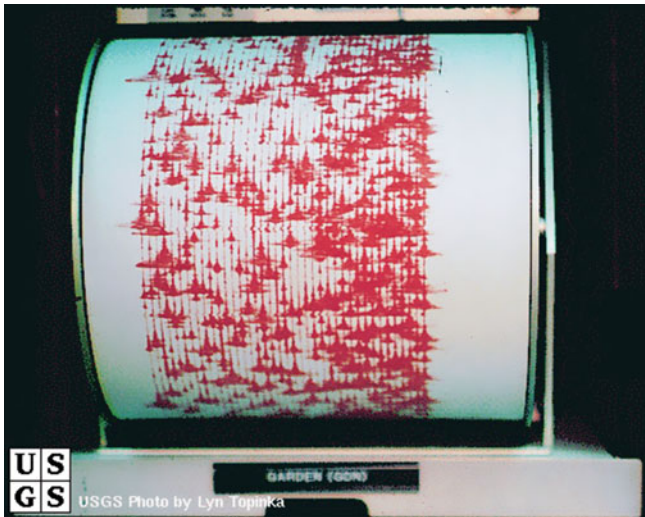
Apart from not recording digitally, the big drawback with the analog recording is its low dynamic range. The maximum amplitude is usually 100 mm and the minimum discernible

amplitude is about 0.5 mm, so the dynamic range is 200 or 46 dB. This can be compared to digital recorders, which achieve up to 150 dB dynamic range. This corresponds to an amplitude of 15 km on the paper recorder!

It should be noted that with digital recording it is of course possible to create a similar “paper” seismogram from the recorded signals and this is a quite common practice, although the seismogram rarely is printed but available on screen; see Fig. 4

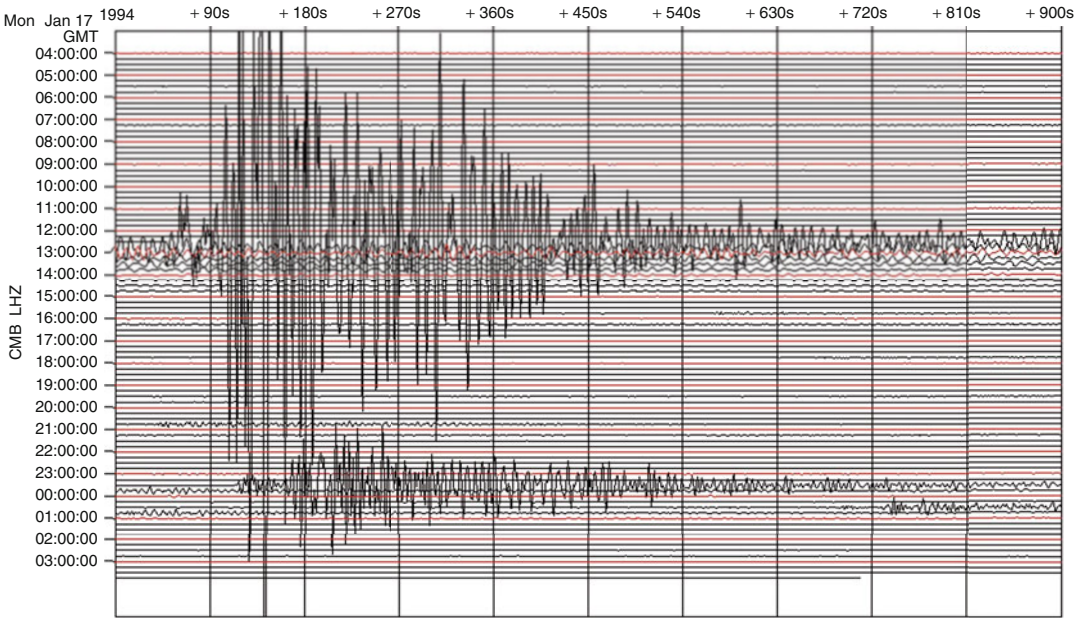
Digitization

A requirement for recording the data digitally is to convert the signal from analog to digital form.



Recording Seismic Signals, Fig. 3 An example of a drum recorder. The unit consists of a rotating drum with paper and a pen mounted on a pen motor (*above* plot, not seen) that translates from left to right as the drum rotates. The gain of the pen amplifier is variable to

adjust to different noise conditions. The seismogram shows the increase in earthquake activity during a dome-building eruption of Mount St. Helens (From vulcan.wr.usgs.gov/Images/Gif/Monitoring/Seismic/seis_garden.gif)

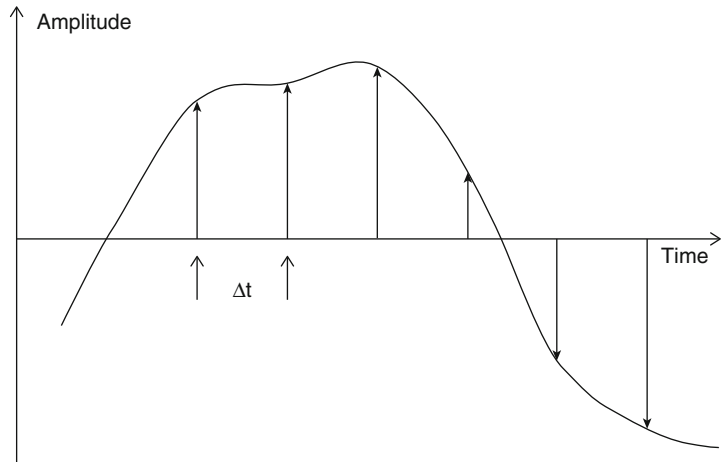


Recording Seismic Signals, Fig. 4 A 24-h seismogram produced from a digital recorded signal. This is a recording of the 17 January 1994 Northridge earthquake

(M6.7) near Los Angeles at 1230 GMT (Figure from topex.ucsd.edu/es10/lectures/lecture11/lecture11.html)

Recording Seismic Signals, Fig. 5

The analog-to-digital conversion process. The arrows show the location and values (amplitudes) of the samples, and the signal is thus approximated with a sequence of numbers available at time intervals Δt (Figure from Havskov and Alguacil 2010)



The process of converting a continuous analog signal to a series of numbers representing the signal at discrete intervals is called analog-to-digital conversion and is performed with analog-to-digital converters (ADC). Figure 5 shows a signal, where the amplitude is sampled at regular intervals Δt .

The process of analog-to-digital conversion involves two steps: first, the signal is sampled at discrete time intervals, and then each sample is evaluated in terms of a number which then corresponds to the amplitude at the time of sampling.

An ADC will give an output binary number when a specific voltage is input. In a computer, numbers are usually represented by a 2-byte word or 4-byte word (1 byte = 8 bits); hence there are 2^{16} or 2^{32} possible values. Since numbers are positive and negative, the corresponding ranges are $\pm 2^{15}$ and $\pm 2^{31}$, respectively (more exactly, $+2^{15}$ to $-2^{15} + 1$ or $+2^{31}$ to $-2^{31} + 1$, because the 0 value has to be included in the 2^n possible codes and the so-called two's complement code is used).

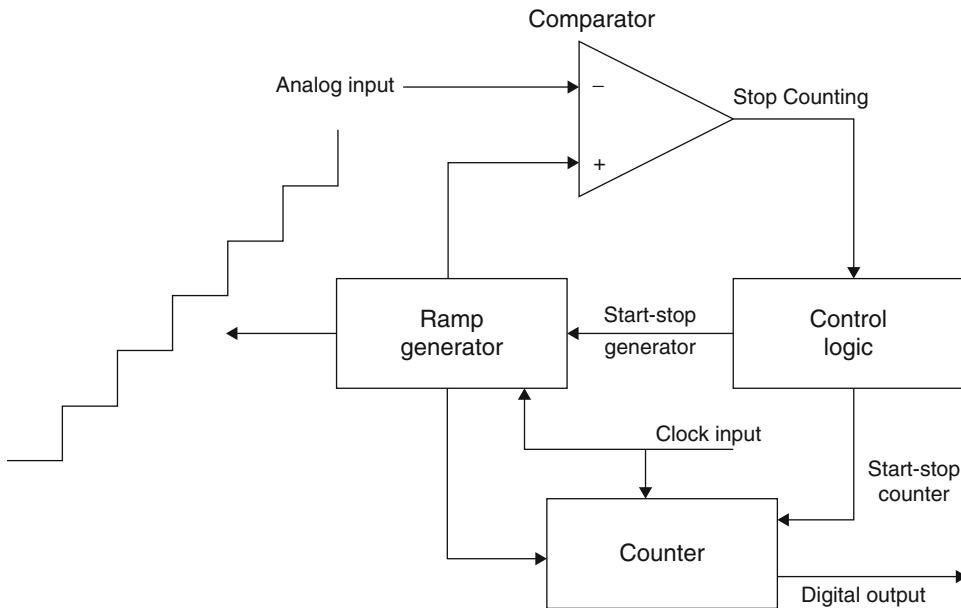
The best ADCs sold for seismology can resolve a step of $0.1 \mu\text{V}$ and has a range of $\pm 2^{24}$ although the most common maximum is $\pm 2^{23}$ (approximately $\pm 8 \times 10^6$). A converter resolving $\pm 2^{23}$ is called a 24-bit converter, and similarly a converter resolving $\pm 2^{15}$ is called a 16-bit converter.

The digitization process will introduce errors into the data and limit the information in them,

simply because a continuously varying signal is replaced by with a discrete set of points. Each sample point is converted from analog to digital form. Since only discrete amplitude values are available, there will be an error, the so-called quantization error. Also there is no information about the signal level between samples in time. This creates errors in both the amplitude and frequency content, and much of the efforts in improving the ADC process are related to minimizing these errors. Before getting into that discussion, it is illustrative to describe how a common ADC works in principle. For a mathematical theory of sampling, A/D conversion process, and other general aspects of seismic recorders, see Asch (2012).

A Typical ADC: The Ramp ADC

One of the simplest approaches of implementing an ADC is the ramp ADC. Figure 6 shows a simplified diagram. The control logic sends a signal to the ramp generator to start a conversion. The ramp generator then generates a ramp signal starting from level 0 (seen on left). The ramp signal enters the comparator, and once the ramp signal is larger than or equal to the input signal, the output from the comparator switches from zero to 1. At the same time as the ramp generator starts, the counter will start to count the number of levels on the ramp. When the comparator switches to level 1, the control logic will stop the counter and the number reached by the



Recording Seismic Signals, Fig. 6 Ramp ADC, see text (Figure from Havskov and Alguacil 2010)

counter is then a measure of the input voltage. After some time, the counter is reset and a new sample can be taken.

The ramp ADC is relatively slow and even slower if a high resolution is required. It also becomes slower as the number increase in size since the counter has to count longer for large amplitudes. An improvement of the ramp ADC is the successive approximation ADC, which is almost identical to the ramp ADC except that it has a more sophisticated control circuit. The converter does not test all levels, but first tests if the input level is below or above half the full scale, thus the possible range has been halved. It then tests if the input level is above or below the middle of this new range. The conversion time is much smaller than for the ramp ADC and constant. This design is the most popular of the classical type of ADCs. A typical 16-bit digitizer of this type may have a conversion time of 20 μ s, which is fast enough for multichannel seismic data acquisition.

Multichannel ADC

There is usually more than one channel to digitize. For three-component stations, there are 3, while for telemetric networks or small arrays,

there might be up to 100 channels. The simplest approach is to have one ADC for each channel. There are several ADC converters that have up to 64 channels with 16-bit resolution and sampling rates in the kHz range. These units only have one ADC. The ADC has, in the front, a so-called multiplexer which connects the ADC to the next analog channel as soon as a conversion is finished. The input signals are therefore not sampled at the same time, and there is a time shift, called skew, between the channels. If the ADC is fast, the skew might be very small, but in the worst case, the ADC just has time to take all the samples, and the skew is the sample interval divided by the number of channels. For many applications, like digitizing the signal from a network, skew has no importance, but in other applications where a correlation between the traces will be made like for arrays or three-component stations, the samples should be taken at the same time. The standard in seismic recorders now is to use one ADC per channel.

Some Basic ADC Properties

Resolution. The smallest step that can be detected which is the input voltage change corresponding to one count, i.e., a change of the *least significant*

bit (LSB, the rightmost bit). For a high dynamic range digitizer, this could be 0.1–1 μV . The number of bits is also often referred to as resolution. Most ADCs have an internal noise higher than one count: in this case, the *number of noise-free bits, rather than the total bit number, limits the effective resolution*. For instance, one count corresponds to 0.3 μV in a 24-bit ADC with a full scale of $\pm 2.5\text{ V}$, but it may have a noise of 2 μV peak to peak, and signals under this level cannot be resolved in practice.

Gain. The sensitivity expressed in counts/V. It can be derived from resolution. If, e.g., the resolution is 10 μV , the gain would be 1 count/ $(10^{-5}\text{ V}) = 10^5$ counts/V.

Sample rate. Number of samples acquired per second. For seismology, the usual rates are in the range 1–200 Hz, while for exploration seismology, sample rates can be more than 1,000 Hz. In general, the dynamic range of the ADC degrades with increasing sample rate, since its internal noise increases with bandwidth.

Dynamic range. Defined as the ratio between the largest and smallest value the ADC can give. For some digitizers, the lowest bits only contain noise, so the dynamic range is defined as the ratio between the largest input voltage and the noise level of the digitizer. This number can be substantially smaller than the theoretical largest dynamic range of a digitizer and may depend on the sampling frequency. So, to give one number for the dynamic range, a frequency bandwidth should ideally also be given. It is normally expressed in decibels (dB): every 20 dB means a factor 10, e.g., 120 dB is a dynamic range of $10^6:1$.

Dynamic range in terms of bits. The dynamic range can also be given as number of bits available in the output data sample. An n -bit converter then gives the numbers $0 - 2^n$ or in bipolar mode $\pm 2^{n-1}$. 12-, 16-, and 24-bit converters are the most used, with 24-bit converters dominating the market. 24-bit converters give out 3 bytes, while 25-bit converters use 4 bytes for storage.

Noise level. Number of counts out if the input is zero (subtracting DC offset). Ideally, an ADC should give out 0 counts if the input is zero. This is usually the case for low dynamic range digitizers 12–14 bits, but rarely the case for high

dynamic range digitizers. The noise level is most often given as an average in terms of RMS noise measured over many samples. A good 24-bit digitizer typically has an RMS noise level of 1–2 counts.

Cross talk. If several channels are available in the same digitizer, a signal recorded with one channel might be seen in another channel. Ideally, this should not happen, but it is always present (maybe at very low level) in practice. The specification is given in dB meaning how much lower the level is in the neighboring channel.

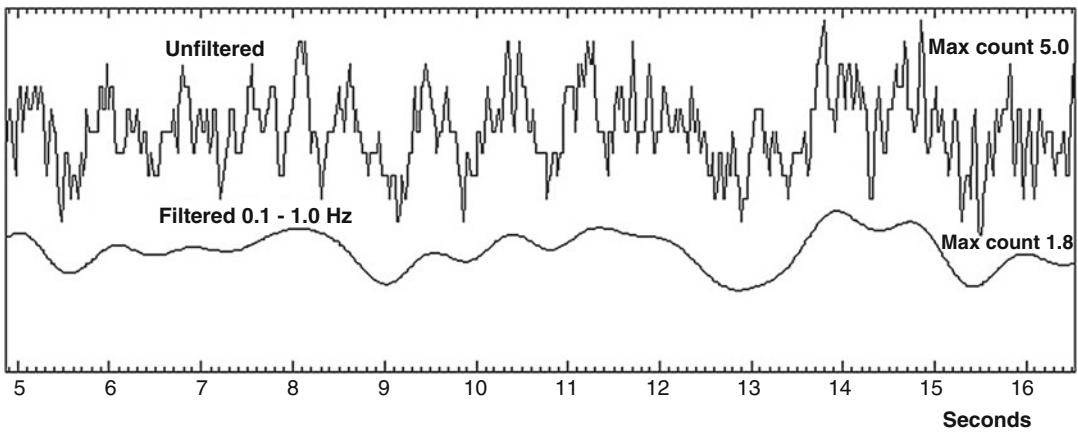
Nonlinearity. If the analog input is a linear ramp, the nonlinearity is the relative deviation of the converter output from the ideal value. It is expressed with relation to full scale (FS), e.g., 0.01 % of FS. For high dynamic range converters, it is important because a poor linearity may cause two different signals at the input to be intermodulated (the amplitude of one depends on the other) at the output. Usually, it is not a problem with modern 24-bit converters.

Digitizers for a Higher Dynamic Range

The Ramp digitizer has a practical limit of 16-bit dynamic range. This is not enough for most applications in seismology. Imagine a network recording local earthquakes. A magnitude two earthquake is recorded at 100-km distance with a maximum count value of 200, which is a lower limit if the signal should be recorded with a reasonable signal-to-noise ratio. What would be the largest earthquake at the same distance that could be recorded with a 16-bit converter before clipping? A 16-bit converter has a maximum output of 32,768 counts or 164 times larger. Assuming that magnitude increases with the logarithm of the amplitude, the maximum magnitude would be $2.0 + \log(164) = 4.2$. So a higher dynamic range is needed. In the following, some of the methods to get a higher dynamic range will be described.

Oversampling for Improvement of the Dynamic Range

The method of oversampling to improve the dynamic range of a digital signal consists of sampling the signal at a higher rate than desired,



Recording Seismic Signals, Fig. 7 Unfiltered and filtered record of seismic background noise in a residential area in western Norway on a hard rock site. The recording is made with a 4.5-Hz geophone and a 16-bit ADC at

a sample rate of 50 Hz. The filter is an eight-pole Butterworth filter with no phase shift (Figure from Havskov and Alguacil 2010)

low-pass filtering the digital signal, and resampling (decimating) it at a lower rate. Qualitatively what happens is then that the quantization errors of the individual samples in the oversampled trace are averaged over neighboring samples by the low-pass filter and the averaged samples therefore have more accuracy and consequently a higher dynamic range. In the frequency domain, the effect of oversampling is to spread the quantization noise over a wide frequency band, thus lowering its spectral density (e.g., Proakis and Manolakis 1992).

As an example, a digital recording of seismic background noise will be used; see Fig. 7. The top trace shows the unfiltered record which has a maximum amplitude of 5 counts. It is possible to see that there is a low-frequency signal superimposed on the cultural noise signal. After filtering, a smooth record of the microseismic noise with a typical period of 3 s is clearly seen. The microseismic noise is globally present and generated by ocean waves. Although the maximum amplitude is only 1.8 counts, it is clear that the resolution is much better than one count.

Noise in the signal, combined with oversampling, is therefore helping to increase the dynamic range. However, with a completely constant, noise-free signal, oversampling would not be able to increase dynamic range. Normally,

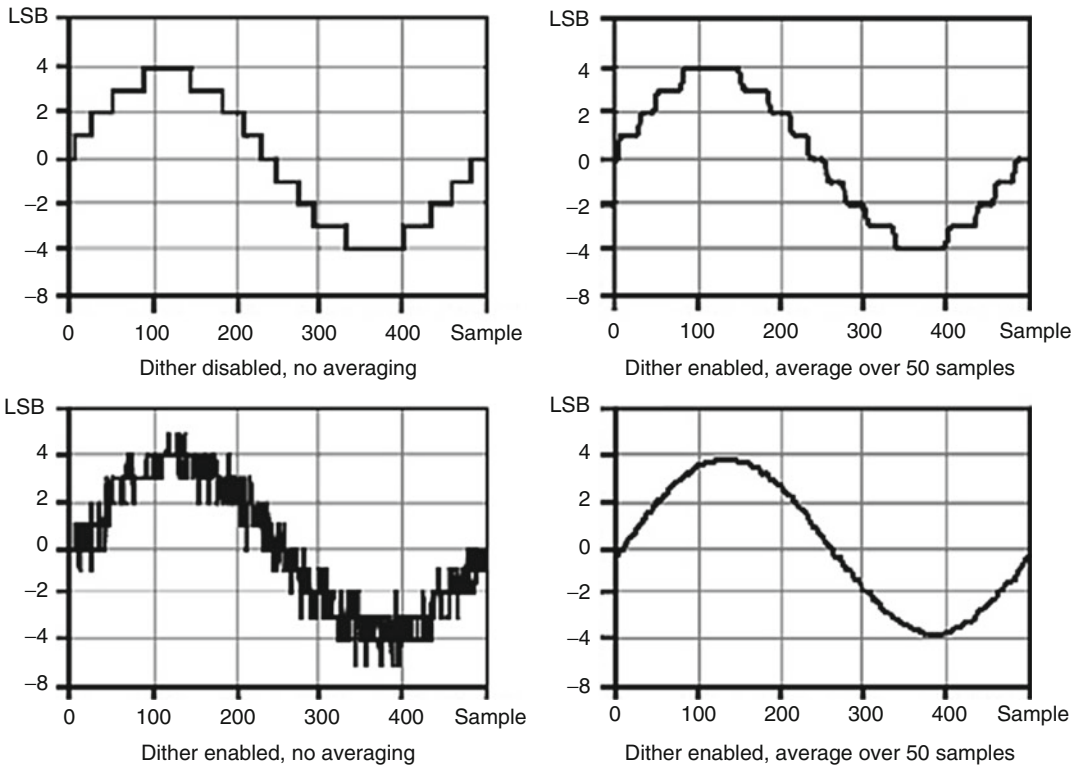
it is no problem to have noise in the signals, rather the contrary. However, in some designs, a Gaussian white noise of amplitude 0.5 LSB is added to the signal to get a higher dynamic range. This is also called dithering; see Fig. 8.

It is seen that the averaging of the signal without noise does not remove the quantization steps. It simply rounds them out a little.

This description of oversampling is very simplified. For more details, see, e.g., Scherbaum (2007) and Proakis and Manolakis (1992).

Sigma-Delta ADC (SDADC)

All ADCs will digitize the signal in steps, so, even with the highest resolution, there will be a quantization error. The idea behind the SDADC is to digitize with a low resolution but high sampling rate so as successive samples are highly correlated, get an estimate of the signal level, add the quantization error to the input signal, get a new estimate, etc. This process will continue forever, and the actual value of the input signal is obtained by averaging a large number of estimates. In this way, a higher resolution can be obtained than it is possible with the original ADC in much the same way as described with oversampling. Most SDADCs are based on a one-bit oversampling ADC that in reality is just a comparator that can determine if the level is



Recording Seismic Signals, Fig. 8 Effects of dithering and averaging on a sine wave input (Modified from www.ni.com/white-paper/3016/en/)

negative or positive. This ADC can be made very fast and accurate and is essentially linear.

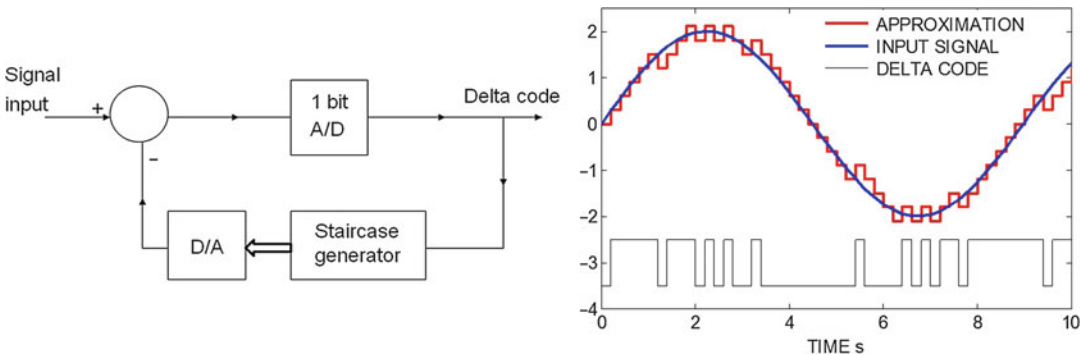
To understand how a sigma-delta works, it is useful to describe first what is called delta modulation, which was used to transmit voice signals in early times of digital telephony. In Fig. 9, we can see the principle. The signal is compared with a staircase signal generated in the following recursive form: if, at each clock pulse, the input signal is higher than the level of the staircase signal, a 1 is output from the A/D converter. This is now fed back to the staircase generator which increases its output with one step. If the level of the signal is smaller than the staircase signal level, a 0 is output, and the staircase signal is decreased by one step. If the signal is constant, a continuous stream of 1,0,1,0... is then the output. The step height needs to be high enough (or the sample rate high enough) to be able to follow the maximum signal slope, but the higher the step, the higher is the noise generated in the

flat zones of the signal. The output level of the D/A determines the maximum signal level that can be digitized.

The staircase function is proportional to the time integral of the output code. This output signal is itself a digital approximation of the time derivative of the input signal. Thus, integrating the delta code will decode it and give an approximation of the input signal. The output signal can also be said to approximate the level differences between one sample and the next.

This is used in the sigma-delta modulator (Fig. 10). In this case, the staircase generator is not needed, since the feedback integration and the input signal integration are performed by the unique integrator after the sum point, and simple 1-bit A/D and D/A do the job.

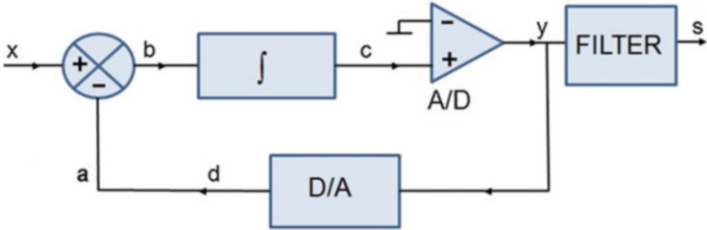
The device operates in discrete time. The digital converter, the integrator, and the digital-to-analog converter are all synchronized and controlled by a logic circuit (not shown).



Recording Seismic Signals, Fig. 9 Delta modulation. *Left*, simplified schematic of a delta modulator. *Right*, the input signal (blue) is compared with a staircase signal (red) that increases by a step when the input signal is

above its previous value and decreases otherwise. The *bottom* shows the output of the modulator which can be used to generate the staircase signal which is an approximation to the real signal

Recording Seismic Signals, Fig. 10 Simplified overview of the functional blocks of a sigma-delta ADC

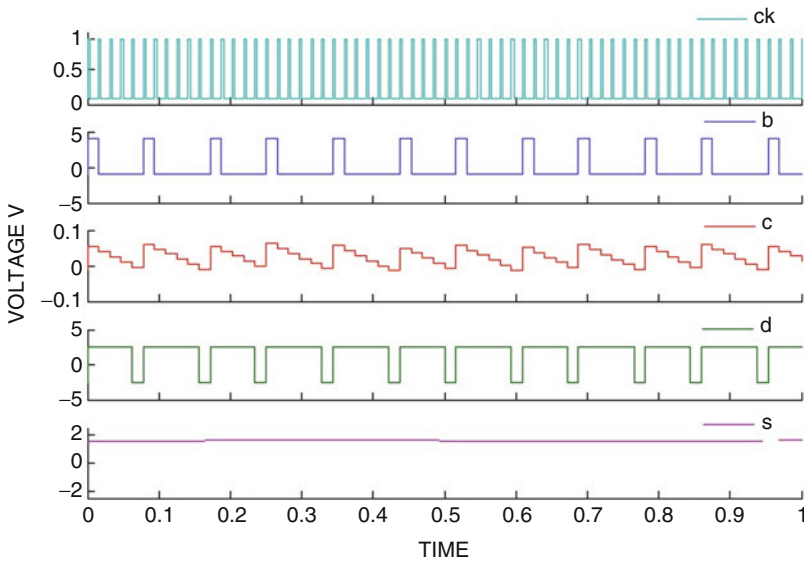


After a time step, the signal y from the A/D (in this example, a simple voltage comparator) is converted back to analog with the digital-to-analog converter (D/A). The output from the D/A is a signal of plus or minus the maximum input voltage (or reference voltage). This signal is subtracted from the present input signal x . The difference b is integrated. Since it is a discrete time circuit, this means that b multiplied with the time step is added to the previous value to give c . Due to the multiplication with the time step, the level of c is small. c is then feed into the one-bit A/D converter. Its output is 1 (high) for a nonnegative c signal and 0 for a negative c signal.

In Fig. 11, it is seen how it works for a constant input voltage of 1.5 V. Step 1: The integrated value is initially zero so 2.5 V is added to the signal b (Fig. 10). Step 2: The value at b is now $1.5 + 2.5$ which integrated becomes $(1.5 + 2.5)/64 = 0.06$. Step 3: A positive value at c (0.06) gives 1 at y , and 2.5 V is subtracted from the

signal at x so b is now $1.5 - 2.5$, and $(1.5 - 2.5)/64 = -0.016$ is added to the previous value of 0.06. This process repeats itself with the same negative steps until c becomes negative and a new positive pulse is generated. The average of the digital output at y is proportional with the analog input at x , and in this case, there are many 1's and few zeros so the output signal is clearly positive. If the input voltage is closer to the positive reference voltage, the negative steps will be smaller so there will be many more 1s compared to zeros. A zero input voltage would give an equal number of 0's and 1's to the average output signal is also zero (remember 0–1, represent the range -2.5 to 2.5 V). A negative input would give more 0's then 1's.

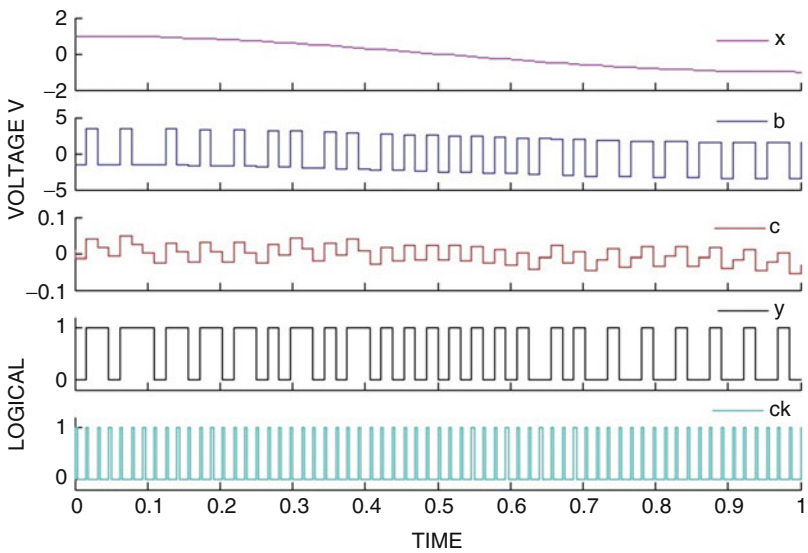
Figure 12 shows a signal that goes from positive to negative, and the proportion of 1s and 0s in the output code y varies accordingly. It can be shown theoretically (e.g., Proakis and Manolakis 1992; Aziz et al. 1996) that the output exactly represents the input if the sample rate is high enough.



Recording Seismic Signals, Fig. 11 Signals at several points of the simplified sigma-delta converter shown in Fig. 10 for an input voltage of 1.5 V. *ck* is the sampling clock. *b* is the difference between the actual input and the D/A output *d* for the previous step, *d* is the output of the digitizer converted to analog voltage. *c* is the integrator

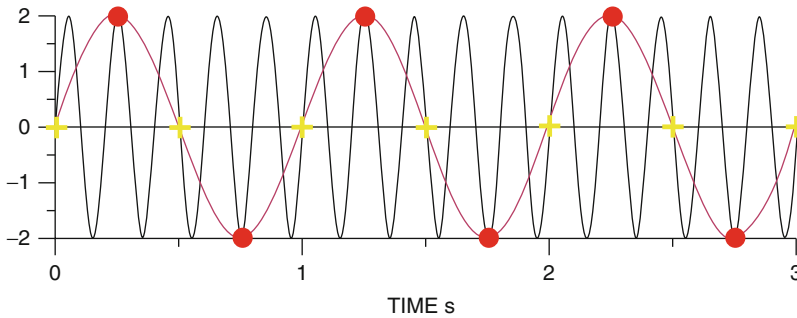
output. *s* is the filtered output. The sample rate is low in this example (64 samples per second, sps) in order to make the figure clear, and the step length is then 1/64 s. The maximum input voltage span (or reference voltage) is ± 2.5 V

Recording Seismic Signals, Fig. 12 Signals at several points of the simplified sigma-delta converter shown in Fig. 10 for a variable input voltage *x*. The signals' labels are the same than in Fig. 11. *y* is the output digital code



The output *y* of the sigma-delta modulator is low-pass filtered to *s* with a digital filter and decimated to a much lower output rate with high resolution. For instance, an internal sample rate of 3×10^5 sps and an output data rate of 100 sps may be typical values.

In practice, the integrator in high-resolution converters is multiple, typically 3–5 order. An *n* order SDAC has *n* difference-integrator circuits following each other. The effective dynamic range of a SDADC increases with the integrator order and the oversampling ratio.



Recording Seismic Signals, Fig. 13 An example of aliasing due to insufficient sample rate. A 5-Hz signal is digitized at a rate of 2 Hz. The digitization points are indicated with *red dots*. Depending on where the samples

are taken in time, the output signal can either be interpreted as a straight line (*yellow points in middle*) or a 1-Hz sine wave (*red*) (Figure from Havskov and Alguacil 2010)

Real SDADCs can be very complicated and there are many variations of the design compared to the description here. A typical SDADC uses several stages of digital filtering and decimation, so there are many ways to get the final signal. High-resolution SDADCs are inherently slow, with output data rates limited to about 1,000 (sps).

Aliasing

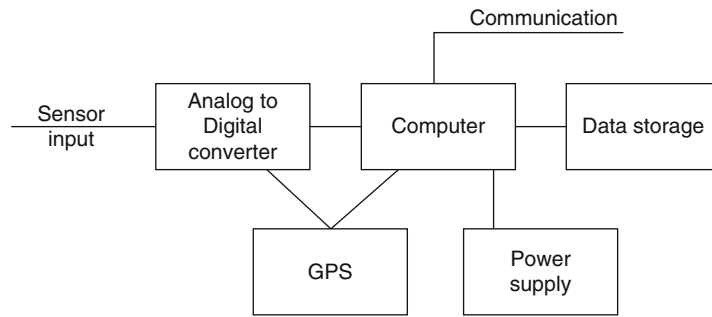
It is seen that there is a quantization error in amplitude due to the discrete resolution. Similarly, errors can also be introduced due to the discrete steps taken in time. Figure 13 shows an example of a 5-Hz signal digitized with a rate of 2 Hz or a sample interval of 0.5 s. The 2-Hz digitization rate is missing out on several oscillations that are simply not seen. If the samples happen to be taken on the top of the 2-Hz cycles, the digitized output signal will be interpreted as a 1.0-Hz sine wave. If the samples were taken a bit later, the output would be a constant level. From this example, it is clear that, in order to “see” a sine wave with frequency f , the sampling rate must be at least $2f$. In the above example, the sample interval must be at least 0.1 s or a sample rate of 10 Hz in order to “see” the 5-Hz signal. Thus, the general rule is that the signal must be sampled at a rate twice the frequency of the signal of interest. Or, in a given time series, it is only possible to recover signals at frequencies half the sampling rate. This frequency is called the Nyquist frequency.

The effect of sampling with a too low rate is not just that information of higher-frequency signals is lost, but more seriously that the high-frequency signals inject energy into the signals at frequencies lower than the Nyquist frequency. In the above example, the pure 5-Hz signal creates a nonexistent 2-Hz signal of the same amplitude as the 5-Hz signal so the digitized output would be completely distorted and have a wrong frequency content. This effect is called aliasing. The only way to avoid this problem is to make sure that the input signal does not contain energy above the Nyquist frequency. So an ADC must have a low-pass filter (an anti-alias filter) to remove unwanted high-frequency energy above the Nyquist frequency prior to the sampling. Alternatively, the sampling rate must be at least double the highest frequency with non-negligible content in the signal.

For traditional ADCs, the anti-alias filtering is done using an analog filter before digitization, while for ADCs with built-in oversampling and processing, most of the anti-alias filtering can be done digitally. Ideally, the anti-alias filter should remove all energy above the level of the LSB at the Nyquist frequency which is possible with high-order digital filtering.

Digital Recorder

The ADC has now been described as a component providing digital data readable by



Recording Seismic Signals, Fig. 14 Main units of a seismic recorder. There are no flow arrows between the units since all can have two-way communications. The GPS can be connected to the digitizer or the recorder. The

power supply may be common for all elements, or each may have its own regulator, but usually the power source is unique (e.g., a battery) (Figure from Havskov and Alguacil 2010)

a computer, in analogy with an amplifier providing an analog signal to be “received” by some analog storage medium. This data now has to be recorded by some computer. Currently, most digital recorders store all the data, the so-called continuous recording. One day of 24-bit (3 bytes) data of one channel, 100-Hz sampling rate, gives $3,600 \times 24 \times 100 \times 3 \approx 26$ Mb of data which is typical for seismological applications. For one month, a three-channel recorder would then record 2.4 Gb or 1.2 Gb if using a compressed recording format. In its simplest form, this can be done with a simple program dumping the data to disk on a general PC. For field use, this is usually not convenient due to power consumptions and that the equipment is not suitable for outdoor use.

The main elements of a digital seismic recorder are seen in Fig. 14.

The main tasks of the seismic recorder can be summarized as follows:

- Read data from one or several digitizers. The digitizers can be internal or external but usually internal.
- If digitized data do not have a time stamp, the recording computer must perform the time stamping of the data. This means that a time reference must be connected to the computer, usually a GPS receiver. If the recorder is connected to Internet, time



Recording Seismic Signals, Fig. 15 Centaur recorder from Nanometrics. The size is $20 \times 14 \times 8$ cm and the weight is 2.0 kg

- stamping could also be done with NTP (Network Time Protocol).
- Store data on a storage medium. Usually, there is a ring buffer, which means that after some time (hours to months depending on the size of the storage medium), the data is overwritten. For limited memory systems, there might not



Recording Seismic Signals, Fig. 16 Two examples of strong motion recorders with internal accelerometers. *Left:* model Etna from Kinometrics, with its lid removed. It is fixed to the floor with a central screw. On the back left

side the internal battery is located, and at the right side the sensors are seen. *Right:* model CMG-5TD from Güralp Systems anchored to the ground with its leveling base. The two cables are for power supply and GPS antenna

be a ring buffer, but just recording of declared events with some pre-event memory. Optionally, the data logger may simply stop recording when the storage media is full, instead of overwriting the oldest data.

- Check incoming data for seismic events (perform triggering) and store only real events. This might not be needed if the system has a large ring buffer but is often an option in order to get a quick idea of recorded events. Most present recording systems will record in continuous form.
- Provide communication with the outside world for change of parameters and download of data. This could be download of files with single events as well as a continuous stream of real-time data to be received at some data center.
- Act as a central recorder. In this case, the digitizer is in the field and it sends the data in real time to a central computer which then records the data.

Different manufacturers have constructed recorders using different types of computers,



Recording Seismic Signals, Fig. 17 The RefTek130S-01, a complete seismic station with six sensors, GPS, and a recorder. A short-period seismometer is seen to the right. The diameter is 22 cm and the height is 15 cm. The weight is 3.2 kg including battery and sensors (Figure from www.reftek.com)

ways of communications, and storage media, which has sometime made it difficult to interconnect different recorders and to use data from different recorders together. Fortunately, there is now more standardization and a typical up-to-date recorder using the most common standard can be as follows:

- The digitizer will have a resolution of 24 bits and be time stamped with a GPS.
- The recording computer will use a single-board Linux computer and record the data in a continuous form in MiniSeed format (the most used standard, see SEED 2012). The recorder will record the signal in flash memory, typical 64 Gb so a standard system will be able to record continuously for more than 4 years (3 channels, 100-Hz sample rate).
- Communications: Standard Ethernet with TCP/IP and serial lines (RS232) so the recorder can be connected to Internet and also accessed by modem or locally with a PC.
- Real-time transmission by Ethernet: The most common standard is to use SeedLink (www.seiscomp3.org). This standard can, by buffering, recover data from communication breakdown of hours to days so a central computer will not lose data in case of communication fault.
- Physical: The recorder will be waterproof, typically use 1 W of power, and have a weight of a few kg.

Seismic recorders for field use can be divided into three types:

A freestanding recorder with external sensors.

This is the most common type and can be used with all types of sensors. An example is shown in Fig. 15. This is a compact recorder with low-power consumption (1 W without Ethernet). It has a Linux operating system, a dynamic range of 140 dB, all required communication ports, high capacity fixed, and removable internal flash storage. It supports several communication

standards, can record MiniSeed, and communicates with SeedLink.

Strong motion recorders with built-in accelerometers.

This kind of recorder can be very simple with only registration of the events and no continuous recording, and it can be as sophisticated as a standard recorder. It is also very widespread and thousands have been built. Figure 16 shows two examples of recorders with built-in accelerometers, both with many units installed: Kinematics Etna, with only events recording capability, and Güralp CMG5-TD, with both events and continuous recording.

A seismic recorder with built-in seismometers.

This is not so common since, to protect the sensor from mechanical disturbances, it is often desirable to separate the seismometer from the recorder. On the other hand, it can be very convenient for field work. Figure 17 shows an example. This is a compact 6-component recorder with a short-period sensor, an accelerometer, and a battery built in. It only uses 0.4 W.

Summary

Recording of seismic signals starts with the process of digitizing the signal from the seismic sensor and doing an accurate time stamping of the data. The data is then sent to a recording computer which will store the data for days or years depending on recording capacity. The computer might also detect the events and/or send the data to a computer center in real time for central recording.

Cross-References

- [Earthquake Magnitude Estimation](#)
- [Ocean-Bottom Seismometer](#)
- [Passive Seismometers](#)

- ▶ Principles of Broadband Seismometry
- ▶ Seismic Event Detection
- ▶ Seismometer Self-Noise and Measuring Methods

References

- Asch G (2012) Seismic recording systems. In Bormann P (ed) New manual of seismological observatory practice (NMSOP-2). IASPEI, GFZ German Research Centre for Geosciences, Potsdam; <http://nmsop.gfz-potsdam.de>; doi:10.2312/GFZ.NMSOP-2
- Aziz PM, Sorensen HV, Van Der Spiegel J (1996) An overview of sigma-delta converters. *IEEE Signal Proc Mag* 13:61–84
- Havskov J, Alguacil G (2010) Instrumentation in earthquake seismology. Springer, Dordrecht, 358 pp
- Lee WHK, Stewart SW (1981) Principles and applications of microearthquake networks. Academic, New York, 293 pp
- Proakis JG, Manolakis DG (1992) Digital signal processing principles, algorithms and applications. Macmillan, New York, 969 pp
- Scherbaum F (2007) Of poles and zeros, fundamentals of digital seismology, revised 2nd edn. Springer, Dordrecht, 271 pp
- SEED (2012) SEED reference manual. Standard for the exchange of earthquake data, SEED format version 2.4. International Federation of Digital Seismograph Networks Incorporated Research Institutions for Seismology (IRIS), USGS. www.fdsn.org/seed_manual/SEEDManual_V2.4.pdf
- Wilmore PL (ed) (1979) Manual of seismological observatory practice, Report SE-20, World Data Center A for Solid Earth Geophysics, US Department of Commerce, NOAA

Regional Moment Tensor Review: An Example from the European–Mediterranean Region

Silvia Pondrelli and Simone Salimbeni
Istituto Nazionale di Geofisica e Vulcanologia,
Sezione di Bologna, Italy

Synonyms

Centroid moment tensor; Earthquake mechanisms and tectonics; Fault plane solution; Mechanisms of earthquakes; Seismic moment tensor inversion

Introduction

The seismic moment tensor is the complete mathematical representation of the movement on a fault during an earthquake, comprising of the couples of forces that produced it, the description of the fault geometry, and its size by means of the scalar seismic moment M_0 .

The computation of seismic moment tensor has become a widely diffused activity because of the relevance of this kind of data in seismotectonic and geodynamic studies and, in more recent times, because it allows obtaining rapid information about a seismic event immediately after its occurrence. This progress has been possible with the advent of modern standardized instruments since the early 1960s, above all of the very broadband seismographic stations that started to record in the late 1970s. Furthermore, time after time, the easier availability of digital data impressed a strong incentive to improve the procedures of source parameter computation.

Seismic moment tensor solutions are used to identify the activated faults during a seismic sequence, to understand their kinematics, and consequently to sketch the short-term possible evolving scenarios. Catalogs collecting all definitive seismic moment tensor solutions are one of the most relevant databases for geodynamic and seismotectonic studies at really different scales. At worldwide scale, the Global Strain Map (Kreemer et al. 2003, <http://gsrm.unavco.org/>, last access November 27, 2014) or the World Stress Map (Heidbach et al. 2008; <http://dc-app3-14.gfz-potsdam.de>, last access November 27, 2014) includes seismic moment tensor data to describe the worldwide strain and the stress maps. Effectively geodetic data measures the motion, borehole measurements give stress directions, and seismic moment tensors, in addition with geologic studies if available, give indications of the acting tectonic regimes.

At regional scale, merging seismic moment tensors with geology and geodesy data provides a picture of the present-day kinematics and tectonics also for a very complex region, e.g., Serpelloni et al. (2007) for the Western

Mediterranean. At local scale, the contribution of seismic moment tensors data is essential to identify the active faults and to study their kinematic behavior. Examples can be found for almost all well-known faults in the world, starting from the San Andreas and concerning the Mediterranean, the Marmara Sea in Turkey (Kinscher et al. 2013) or the Atlantic Ocean close to Portugal coasts (Stich et al. 2007) or along the Southern Apennines in Italy (Cucci et al. 2004).

The possibility to determine the seismic moment tensor of an earthquake is strongly related to the magnitude of the event and the threshold of the used computation method. For large global earthquakes, several research groups and agencies in the world produce routinely seismic moment tensors. Those having at present the longest experience and the greatest regularity are primarily the Global CMT (previously Harvard CMT) group at Lamont–Doherty Earth Observatory (LDEO) of the Columbia University (Dziewonski et al. 1981; see entry on “► Long-Period Moment-Tensor Inversion: The Global CMT Project”) and the United States Geological Survey (USGS, Sipkin 1994). These groups both started in the 1980s, employing different methods but having in common the use of teleseismic waveforms. More recently, also the source parameters produced by GEOSCOPE group at the *Institut de Physique du Globe de Paris* (Vallée et al. 2011) and those from GEOFON Earthquake Information Service (Cesca et al. 2010) are constantly available.

The Global Centroid Moment Tensor catalog (GCMT) produces centroid moment tensor (CMT) solutions for events that occurred since 1976 (Dziewonski et al. 1981; see entry on “► Long-Period Moment-Tensor Inversion: The Global CMT Project”). CMT solutions are routinely calculated for events located worldwide with a moment magnitude of M_w 5.0 and above. The final monthly catalog is published online about 4 months later (www.globalcmt.org). The traditional GCMT method uses long-period body and mantle waves (those with a period longer than 135 s); most of published solutions are calculated using an approach that, excluding minor modifications, is the same of its inception, in the 1980s.

In the last few years, only GCMT solutions are computed inverting also for intermediate period surface waves, variation that allows a lowering of the magnitude threshold down to 5.0 (see entry on “► Long-Period Moment-Tensor Inversion: The Global CMT Project”).

The United States Geological Survey (USGS) seismic moment tensor solutions have been determined routinely starting from 1993 (Sipkin 1994) for worldwide events with a magnitude greater than 5.5, using a technique different from the one employed for the computation of the Global CMTs. Namely, the “optimal filter design theory” has been applied to compute the source parameter solutions using only the long-period P-waveform vertical components, low-pass filtered at 20 s. The use of vertical components simply allows for shorter time to recover the necessary data, accelerating the source parameter computation. USGS moment tensors are available since 1980 because they have been computed backward in time. Presently, USGS seismic moment tensor solutions are regularly published in the *Physics of the Earth and Planetary Interiors* journal (PEPI) up to 2004 and at the same time are available on a dedicated web site (<http://earthquake.usgs.gov/earthquakes/eqarchives/sopar/>, last access November 27, 2014).

The list of agencies or research groups active in producing seismic moment tensors becomes longer when the magnitude threshold and the geographic scale changes from worldwide to regional. Moreover, the availability of these data is much more scattered, especially if we compare it among different areas. In fact, complete datasets with a long-time continuity are naturally available just in regions where low-to-moderate earthquakes are considered relevant and are recorded by a consistent number of seismographic stations, such as in Japan (Grid MT, Tsuruoka et al. 2009; NIED CMT, Kubo et al. 2002), in California (TMTS, <http://seismo.berkeley.edu/mt/>, last access November 27, 2014), or in the European–Mediterranean region (European–Mediterranean RCMT catalog or GFZ moment tensor solution dataset).

Here the efforts developed in the last decades in Europe and in the Mediterranean region for the

computation of regional seismic moment tensors for moderate-magnitude events are described. The earthquakes characterized by moderate energy release are those that more frequently affect this region. They are relevant given the high density of population and the concentration of human activities (e.g., tourism) where the seismic hazard is sometime incredibly high due to historical and social reasons. After a review of the applied techniques and the available catalogs for the Mediterranean region, a description of the European–Mediterranean RCMT catalog is added, followed by a review of typical tectonic trends of this region.

Seismic Moment Sensors for the European–Mediterranean Region

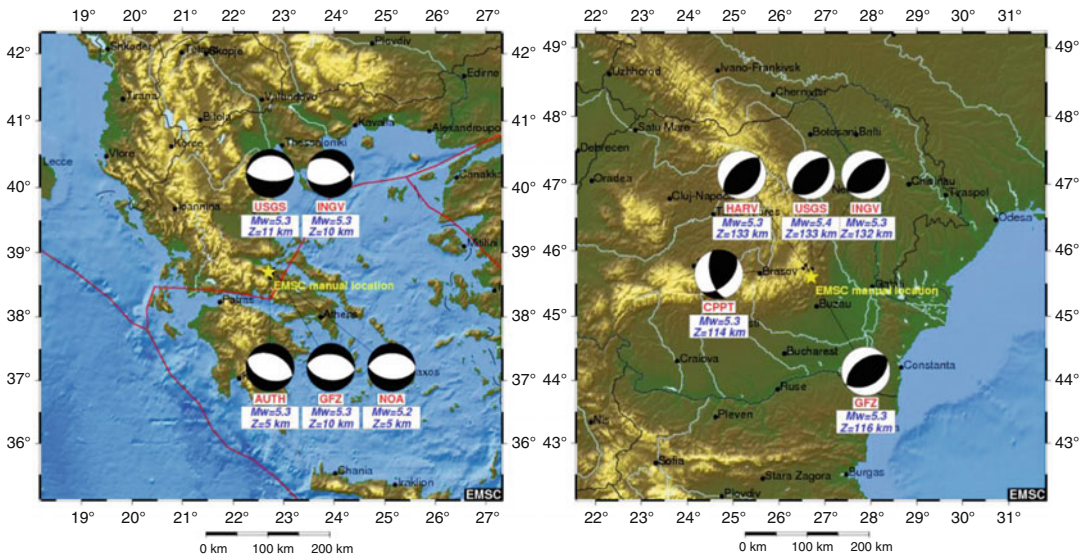
In the European–Mediterranean region, also a moderate-magnitude event may be particularly relevant due to the high seismic hazard of most of this region. The heterogeneity of the distribution of population since historical time, in some area particularly dense, together with a great variety of human activities, with tourism as one of the more relevant economy income, should be combined to a diffuse building vulnerability due to historical and social reasons.

Several agencies compute source parameters also for moderate-magnitude earthquakes for the whole region or for subregions. Figure 1 shows two maps downloaded from the European–Mediterranean Seismological Center (EMSC) web page (<http://www.emsc-csem.org/Earthquake>, last access November 27, 2014) showing all seismic moment tensors quickly computed by several research groups and published rapidly online after earthquake occurrences. Despite the moderate magnitude that is lower than 5.5 in both cases shown in the maps, at least five focal solutions are available. These results are obtained applying different methods and somehow different data; the added value is the quick check of their stability obtained by comparing them.

For moderate-magnitude earthquakes, the more constant contributors of quick solutions for the whole European–Mediterranean region are those

computed by the Quick RCMT project (Pondrelli et al. 2012 and references therein; <http://www.bo.ingv.it/RCMT>, last access November 27, 2014) and by the German Research Centre for Geoscience (Cesca et al. 2010; <http://geofon.gfz-potsdam.de/eqinfo/list.php?mode=mt>, last access November 27, 2014). Quick RCMTs are obtained with a semiautomatic process (Pondrelli et al. 2012) activated by the alerts sent by the EMSC after any events of interest, in this case any with a magnitude greater than 4.5 occurring in the geographic window between 10°W and 40°E of longitude and 25°N to 60°N of latitude. The magnitude threshold lowers to 4.0 for seismic events that occurred in Italy. Seismograms of body and surface waves mainly recorded at regional distance (within 15°–20° from the epicenter), collected from IRIS (<http://www.iris.edu>, last access November 27, 2014) and WebDC EIDA Data Center (<http://www.webdc.eu>, last access November 27, 2014), are used to invert for seismic moment tensor computation. The inversion process, better explained in the next paragraph, is an evolution of the Global CMT method (see entry on “► Long-Period Moment-Tensor Inversion: The Global CMT Project”). A visual inspection is done before the distribution online of inversion results, through several web sites, such as the <http://autorcmt.bo.ingv.it/quicks.html> (last access November 27, 2014) or the EMSC event web page (Fig. 1).

GFZ seismic moment tensor solutions are computed automatically using a spectral analysis technique that, at successive inversion steps, adopts the appropriate range of frequency, norm definition, and inversion method, to reach the best fitting source parameters (Cesca et al. 2010). Any solution is published online at <http://geofon.gfz-potsdam.de/eqinfo/> (last access November 27, 2014) and in the EMSC event web page. Since 2011, within this project, the seismic moment tensors for earthquakes with a magnitude threshold that changes with distance are constantly determined. Indeed, for European earthquakes, the GFZ dataset includes the source parameters for events with a magnitude greater than 4.5 and occasionally also for lower magnitude, down to 4.0.



Regional Moment Tensor Review: An Example from the European–Mediterranean Region, Fig. 1 A screenshot of two typical EMSC web pages, as examples of the several seismic moment tensor solutions computed for two events, September 16, 2013 M 5.3 in Greece and

October, 6, 2013 M 5.3 in Romania. Moment tensor solutions are quickly computed immediately after the earthquake occurrence and published routinely on the EMSC web page <http://www.emsc-csem.org/Earthquake>

Out of RCMT or GFZ MT, other contributors compute seismic moment tensors for earthquakes of the Mediterranean region. At the Swiss Seismological Service, in Zürich, the ETHZ moment tensor project is active. Before 2010, it guaranteed the automated computation of seismic moment tensors for large earthquakes worldwide and for large- to moderate-magnitude events in the European–Mediterranean region (http://www.seismo.ethz.ch/prod/tensors/mt_autom/index_IT, last access November 27, 2014; Bernardi et al. 2004). Computations were done using a consolidated technique based on the inversion of surface waves in the spectral domain (Giardini et al. 1993). Since 2010, the ETHZ moment tensor project produces automatically the seismic moment tensor for all events with M_l greater than 3.5 occurred in the Swiss region only. They adopted the Dreger and Helmberger (1993) method, adapted to local conditions and updated with local Green functions. This algorithm inverts the complete three-component broadband displacement waveforms to estimate a point-source solution.

This computation method is presently used also at the Instituto Geográfico Nacional seismic moment tensor project (IGN, <http://www.ign.es/ign>, last access November 27, 2014, Rueda and Mezcuá 2005). The source parameters for the seismicity having a magnitude greater than 3.0 of the Iberian Peninsula, Gibraltar Strait, and Alboran Sea, Northwestern Africa, in particular Canary Islands and Morocco, are determined using three-component seismograms recorded from the IGN seismographic network. The obtained solutions are then published quickly on the dedicated IGN web page and in the EMSC event web page.

Similarly, it occurs in the TDMT@INGV project for Italian seismicity. Since 2006, seismic moment tensors are rapidly computed also for smaller magnitude events, down to 3.5, using the Italian National Network seismic recordings and the Dreger procedures with the Green functions calibrated for the Italian peninsula (Scognamiglio et al. 2009; <http://cnt.rm.ingv.it/tdmt.html>, last access November 27, 2014).

In the Southeast Mediterranean areas, the Aristotle University of Thessaloniki (AUTH, Roumelioti et al. 2007) for the Greek seismicity and surrounding regions, and the Kandilli Observatory (KOERI) for Turkish seismic activity, have implemented dedicated procedure based on Dreger methods. The results of these analyses are regularly published on the EMSC event web site, such as the examples reported in Fig. 1. It is only at the National Observatory of Athens (NOA, Konstantinou et al. 2010) that source parameters are computed with a different technique, a linear time-domain moment tensor inversion method with a point-source approximation (Randall et al. 1995; Melis and Konstantinou 2006). Moment tensors are determined for all earthquakes, with a magnitude occasionally lower than 4.0, occurring in Greece and surrounding regions since 2005 (<http://bbnet.gein.noa.gr/HL/seismicity/moment-tensors>, last access November 27, 2014).

The amount of projects active at present to obtain rapidly good quality seismic moment tensors in the European–Mediterranean region enhances the importance given to source parameters in real-time seismic sequence scenarios. However, all datasets collected by each project, even if merged altogether, lack the geographic and temporal continuity that are requested to define a catalog to be used in seismic hazard or seismotectonic studies. The European–Mediterranean RCMT catalog only reached these features described in the following.

A Consolidate Example: The European–Mediterranean RCMT Catalog

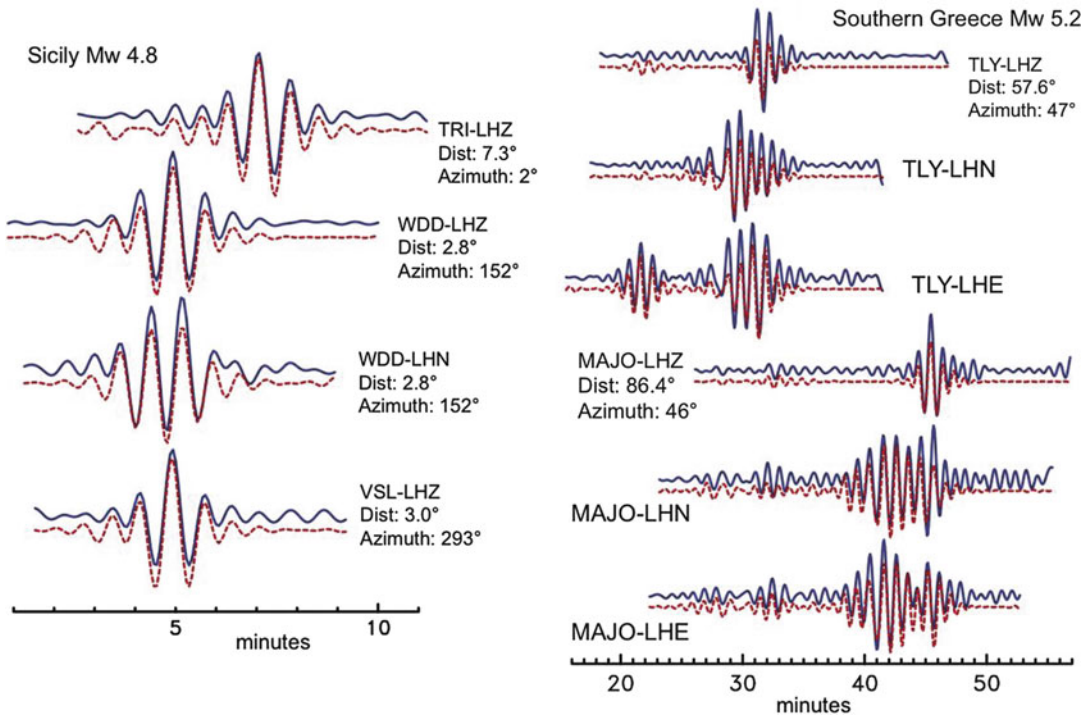
The RCMT project is the most relevant and most continuous initiative for seismic moment tensor computation on the European–Mediterranean scale. It starts with the 1997–1998 Central Italy seismic sequence. At that time, Arvidsson and Ekström (1998) have already developed the *regional* version of centroid moment tensor (CMT) computation technique, based on the modeling of surface waves of intermediate period (35–100 s) recorded at regional distance

(within 10° between station and epicenter). The method focuses on the determination of a centroid moment tensor also for moderate-magnitude events, e.g., a magnitude between 4.5 and 5.5, while standard CMT points to solve for earthquakes with a magnitude greater than 5.5.

When the RCMT project activated, each seismic moment tensor solution was determined using three-component seismograms recorded only by the very broadband stations of the regional network named MedNet. Seismograms were obtained through modem connection and computations were made using data from five to ten stations (Ekström et al. 1998). The first solutions in quasi-real time, namely the Quick RCMTs, have been available about 5 years later, in 2002, and regularly published in the MedNet web page, now not updated anymore. At the same time, the first version of the European–Mediterranean RCMT catalog became available (Pondrelli et al. 2011). RCMT solutions were obtained using three-component seismograms from all reachable very broadband stations, belonging to regional networks, e.g., MedNet or GEOFON (now GFZ) or worldwide networks, e.g., GSN IRIS/IDA network. Seismograms from stations closer than 90° were generally used and when the magnitude were lower than 5.0, only stations at distances up to 60° were included in the inversion.

In the firstly used RCMT algorithm, source parameters were recovered inverting for the complete three-component waveforms, including fundamental mode surface waves (Love and Rayleigh). This part of the seismic signal is, at regional distance, the one with greatest amplitude, but it is difficult to model because of its high sensitivity to crustal heterogeneities. On the other hand, modeling surface waves recorded at short distance allows the computation of moderate-magnitude events. This is the main advantage of RCMT algorithm and the reason of its extensive use in some regions, with respect to traditional CMT method.

In the RCMT computation process, the source excitation has been always computed in PREM (Dziewonski et al. 1981). Synthetic seismograms



Regional Moment Tensor Review: An Example from the European–Mediterranean Region, Fig. 2 On the left, a group of waveforms (real in blue and synthetics in dotted red) for stations located at regional distance (within 10°) for an event with Mw 4.8 occurred in Sicily in January 1998. On the right, a group of waveforms belonging to an earthquake Mw 5.2 occurred in Southern Greece

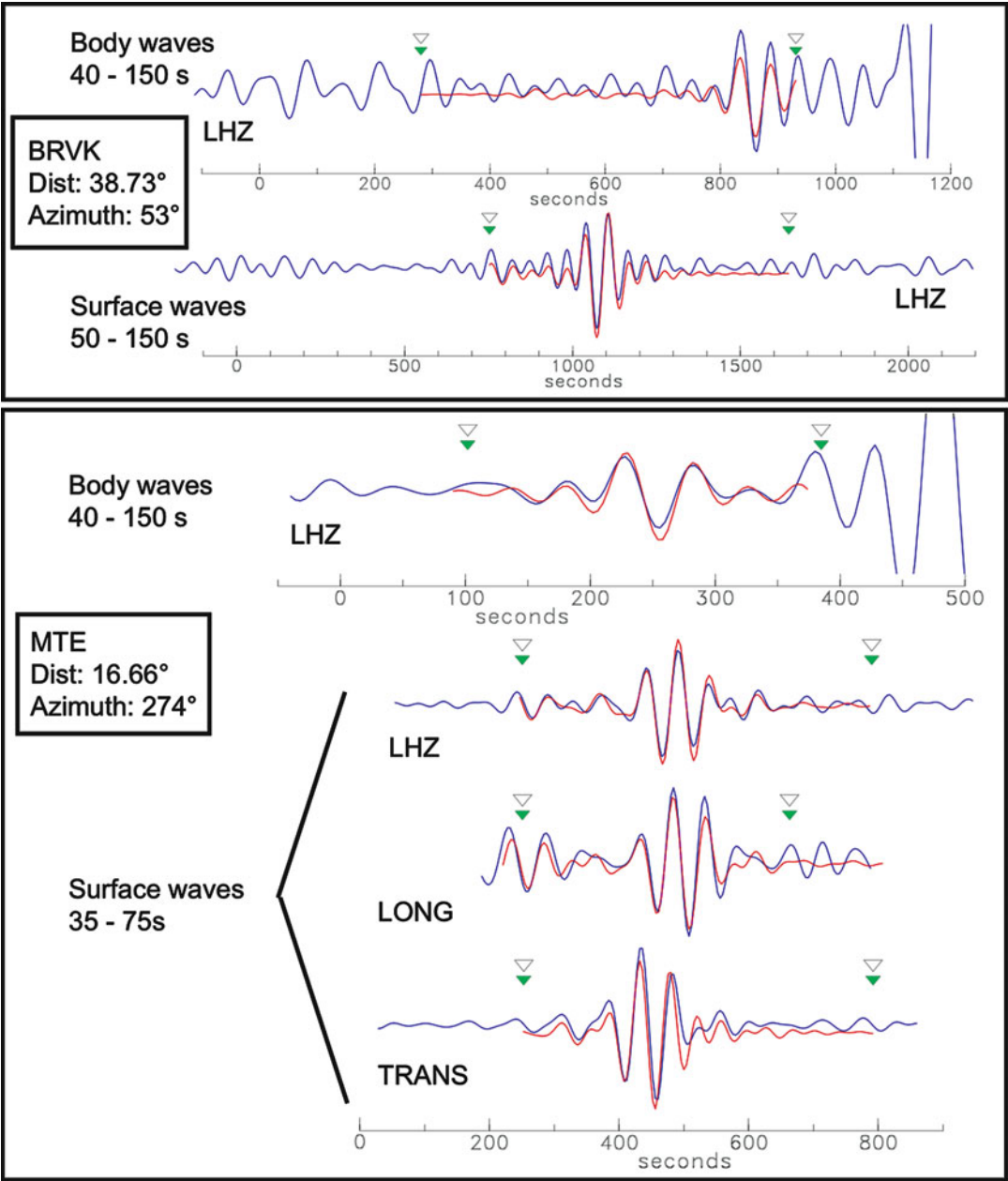
in October 1998, for stations located at teleseismic distance. For each waveform, the station name, the component (LHZ: vertical, LHN: N–S component, LHE: E–W component), the distance from the epicenter, and the azimuth, that is, the direction at which is located the station with respect to the epicenter, clockwise starting from the north, are reported

for fundamental mode surface waves have constantly determined by propagating the source pulse through the high-resolution global phase velocity models of Ekström et al. (1997). Normal mode summation was used only to model the overtone contribution, mostly evident when teleseismic distance waveforms were included in the inversion (Fig. 2). The observed and synthetic seismograms matched after a low-pass filtering with a cutoff down to 35 s and three to five iterations to minimize the misfit, estimating the moment tensor elements and a centroid location.

Since 2004, RCMTs are determined inverting also for long-period body waves, modeled by summation of the normal modes of the Earth corrected for large-scale three-dimensional mantle structure, in the same manner as in the Global

CMT Project (see entry on “► [Long-Period Moment-Tensor Inversion: The Global CMT Project](#)”).

An example of observed and synthetic seismograms for different distances is shown in Fig. 3. Body waves are included in the RCMT inversion when the distance between the epicenter and the station is large enough to separate body wave-train arrival from the surface wave one, maintaining at the same time a good signal-to-noise ratio. The use of the two different wave types has a stabilizing effect on the centroid location, in particular, and this is true for the earthquakes occurring near the border of our study region, e.g., out of Gibraltar Strait, since for these events the azimuthal station coverage is sometimes poor. Band-pass filtering is done between 40 and 150 s in velocity for body



Regional Moment Tensor Review: An Example from the European–Mediterranean Region, Fig. 3 Two examples of real (*blue*) and synthetic (*red*) waveforms for body and surface waves produced by the more recently used RCMT process computation. These data are for a Mw 5.2 earthquake, occurred in Southern Italy in

December 2014, recorded by two stations located at a different distance from the epicenter. In the lower panel, for surface waves, rotated horizontal components and the longitudinal and transverse components (LONG and TRANS) are used

waves, while for surface waves band-pass value changes with the magnitude, starting from 50 to 150 s for earthquakes with a magnitude of 5.5 to 35–75 s when the magnitude lowers to 4.5. Surface wave filtering is done always on the displacement signal.

At present, the RCMT computation procedure starts for all earthquakes occurring in the European–Mediterranean region with MI (Richter or local magnitude) greater than 4.5 or 4.0 for the Italian peninsula. The semiautomatic *pypaver* process (Pondrelli et al. 2012) produces a first version of Quick RCMT that is published, after a visual inspection, on the dedicated web pages (see previous paragraph for detail). After 3–4 months, all earthquakes, for which a trial to compute a QRCMT has been attempted, will be revised and all stable solutions are included in the European–Mediterranean RCMT catalog. It is searchable at <http://www.bo.ingv.it/RCMT/searchRCMT.html> (last access November 27, 2014).

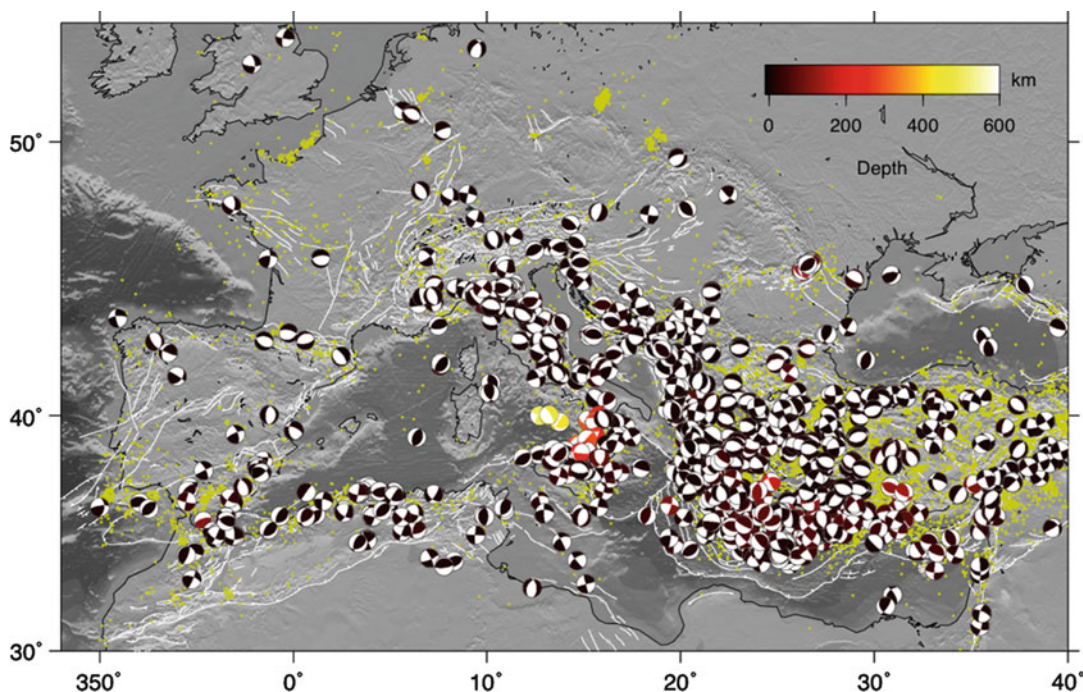
RCMTs are available in the European–Mediterranean RCMT catalog since 1997 up to the present, for events with magnitudes greater than 4.5, for a total of nearly 1700 seismic moment tensor solutions. The geographical space is limited to 25°–60° N in latitude and 10°W–40°E in longitude. For the Italian peninsula and Cyprus Island only, a backward extension of this catalog in time (up to 1976) and in magnitude (down to Mw 4.0) exists (Pondrelli et al. 2006; Imprescia et al. 2011; <http://www.bo.ingv.it/RCMT>).

The RCMT, definitive or quick, is considered stable when: (1) a minimum of seismograms of three stations azimuthally well distributed (with an angular distance of about 120° each other) is available, (2) the focal mechanism remains stable during five iterations needed to determine the centroid location, (3) the total root mean square of the misfit between seismograms and synthetics averaged for all station used is lower than 0.4, (4) the difference between initial and final coordinates is lower than 0.3°, and (5) the moment tensor should have a small non-double-couple component. This last point is quantified using thresholds arbitrarily used to define a non-double-couple moment tensor (see entry on

“► Long-Period Moment-Tensor Inversion: The Global CMT Project”).

In addition, a quality flag is given to all RCMT solutions. The A flag is given to all RCMT solutions that show full compliance to all criteria listed above, while the other letters (up to D) describe decreasing levels of quality. The B flag corresponds to moment tensors that have a difference between preliminary and final coordinates greater than 0.3° and lower 0.5°. If the variation is greater than 0.5°, the coordinates are kept fixed along the inversion and the flag becomes C. Coordinate variation is usually due to low quality of azimuthal distribution of stations when calculating RCMT for moderate-magnitude earthquakes. Flag D is given only when the obtained moment tensor has a high non-double component due to low signal-to-noise ratio. Changes in the difference between preliminary and final coordinates and large non-double-couple component can be considered indicators of complex seismic sources for large-magnitude events, but in the case of RCMT, computed mainly for moderate-magnitude earthquakes, they are commonly related to the low-quality, mainly noisy, seismographic data. The expedient of quality flags allowed being less severe in the quality criteria check when RCMT computation is applied to old or smaller earthquakes, as in the case of the Italian peninsula or for Cyprus Island.

The success ratio between the obtained RCMT solutions and the number of seismic events for which a computation trial is done is about 60 %. Several factors influence this ratio. Most of the earthquakes listed in the seismicity catalogs are tagged with a magnitude, in general MI (Richter or local magnitude), significantly larger than our resulting moment magnitude Mw; thus they are often too small to produce the necessary signal at long period. Another limiting factor is that moderate-magnitude events often are foreshocks or aftershocks of a seismic sequence. Consequently, waveforms include the signal of more than one event in the same time window, avoiding the possibility to separate single seismic source solution. Moreover, during a seismic sequence, seismograms are often strongly



Regional Moment Tensor Review: An Example from the European–Mediterranean Region, Fig. 4 Map of all the European–Mediterranean RCMT catalog solutions starting from January 1997 to December 2013. Focal mechanisms (*shaded areas*) change color with respect to the hypocentral depth of the earthquake (see legend at *top right*): black focal mechanisms represent shallower seismicity, while deeper events have a *red to yellow* beach

ball. Deeper events are mostly located close to the Gibraltar Strait, in the Southern Tyrrhenian Sea, and along the Hellenic Trench and South Turkey to Cyprus Island. In the background, the seismicity of the last 10 years from the European–Mediterranean Seismological Center (EMSC, <http://www.emsc-csem.org>) is mapped in *yellow* and major tectonic features are drawn in *white* (Barrier et al. 2005)

contaminated by the signal of the main events. Furthermore, a low quality of preliminary geographic location influences the final solidity of the RCMT. Earthquakes occurring in the marginal areas of the Mediterranean (out of the Gibraltar Strait or along part of the Northern Africa) often lack the sufficient recordings to reach as well a stable location and this weakens also definitive RCMT quality.

To ascertain the stability and quality of RCMT solutions, they have been compared to other seismic moment tensors computed with different conditions, finding a very good correspondence although station recordings and modeled waveforms were substantially different. Since the project started, RCMT solutions have been constantly compared with traditional CMT solutions and, successively, with other similar datasets. Several comparisons have been done

with ETH and GFZ data. A recent example is given in Cesca et al. 2010, where their seismic moment tensors are shown together with other catalog solutions, in particular RCMT (in the paper named MedNet, from the name of seismographic network data used at that time), Global CMT (still named Harvard CMT), ETH, and USGS seismic moment tensor solutions. In Fig. 4, the complete European–Mediterranean RCMT catalog is reported. Although surface wave signal is expected to be a little part of the seismogram in case of deep sources, it is sufficient to obtain stable RCMT solutions (Fig. 4).

The European–Mediterranean RCMT catalog considerably increases the number of earthquake moment tensors available in the European–Mediterranean region. For instance, the Global CMT catalog includes 424 events for the period 1997–2012 for this region, while adding RCMT

quadruples the total number of solutions (Fig. 5). This result is particularly significant for the Mediterranean region, where moderate-magnitude events have sometimes a greater impact with respect to other parts of the world, as a consequence of the particular density of population and human activities and the heterogeneity of seismic risk plan development, related to important social difficulties.

Focal Mechanisms and Tectonic Systems in the Mediterranean

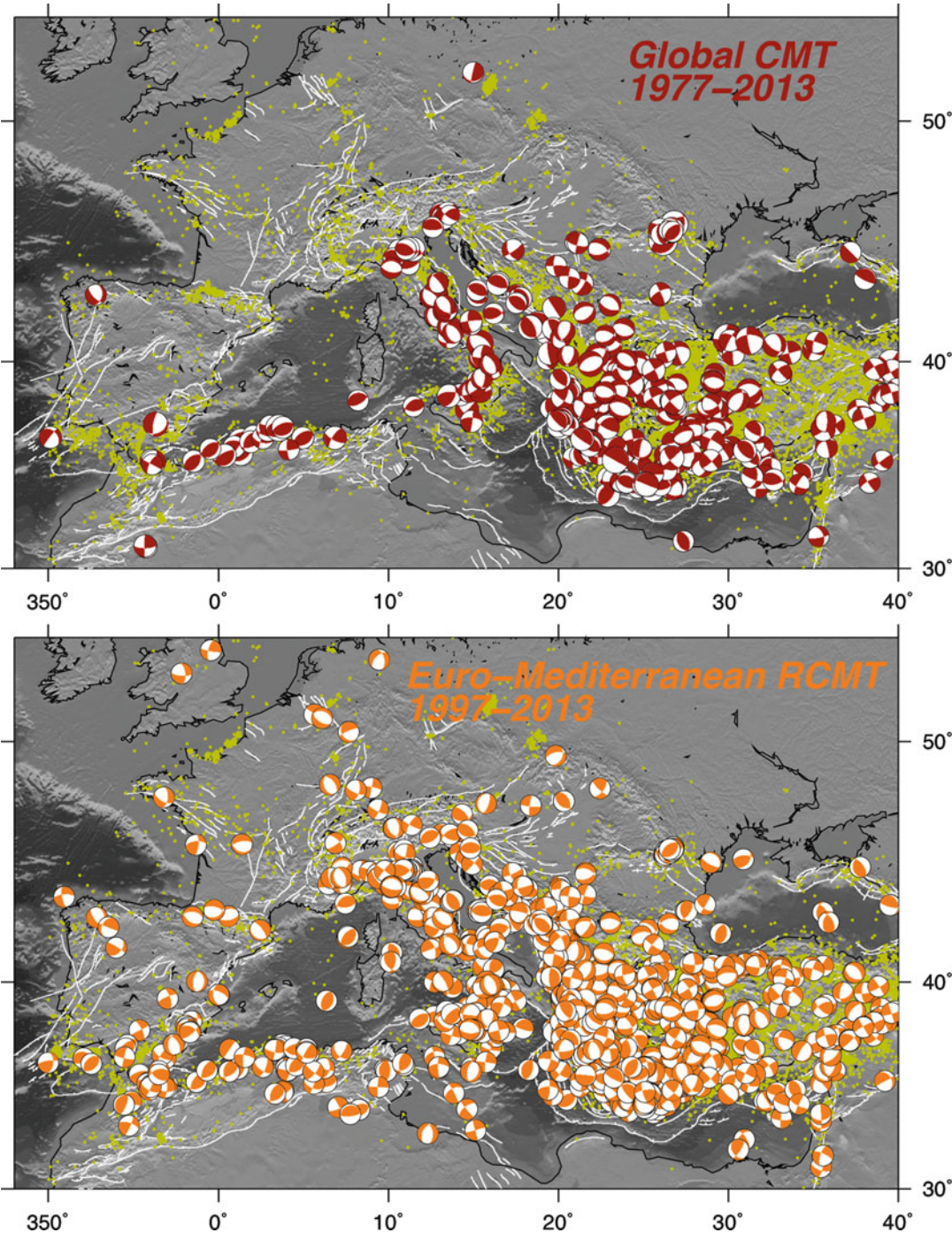
The Alpine–Mediterranean region is characterized by a really complex active tectonics, connected to the 100 My convergence between Africa and Eurasia plates, also involving several microplates (e.g., Adria and Anatolian ones). Earthquake focal mechanisms, representing the wide range of tectonic regimes acting here, contribute to the understanding of the active tectonics and yield fundamental information.

Seismicity in the Mediterranean area is rather diffuse and is characterized by mostly moderate energy release. Moderate-magnitude events are therefore particularly important because they are more widely spread than the relatively infrequent large earthquakes. This is evident comparing the two maps of Fig. 5, where 17 years of RCMT data give information for several regions (e.g., Central Europe, Pyrenees, Tunisia) not documented by 37 years of CMT data, as a consequence of the lowering of the magnitude threshold.

The most relevant geodynamic feature of the European–Mediterranean region is the tectonic boundary between Africa and Eurasia plates (yellow line in Fig. 6): starting from the Atlantic Ocean, it crosses from west to east of this area, entering the Mediterranean Sea through the Gibraltar Strait, following the coast of Northern Africa up to Sicily, then following the main mountain belts, such as the Apennines, the Alps, the Dinarides, and the Hellenides, and going further to the Hellenic Trench and the Cyprus Arc, moving toward east up to the Caucasus. The Africa and Eurasia plates converge for a long time, but in spite of generating solely a mountain

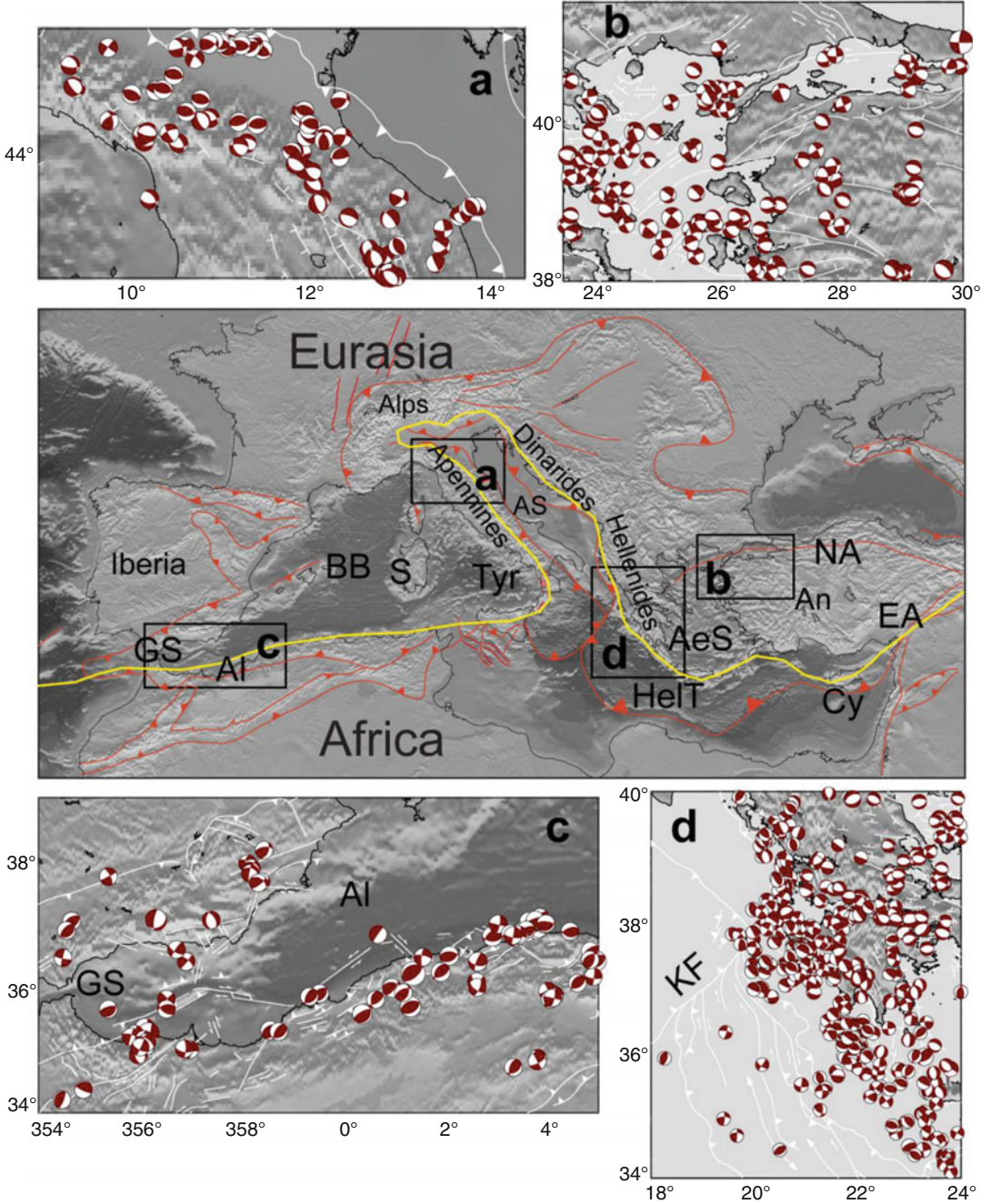
belt between them, a partitioning of the boundary gave life to several small geodynamic systems that altogether accommodated the plate convergence. Most of the basins that together compose the entire Mediterranean Sea, namely, the Alboran Sea, the Balearic Basin, the Tyrrhenian Sea, and the Aegean Sea (Fig. 6), opened as back-arc basins of different subducting systems active in different times. Only the Southern Tyrrhenian and the Hellenic subducting systems are still active as testified by deep earthquakes occurring down to 400 km of depth and more (Fig. 4). This amazing geologic history has been reconstructed using geological and geophysical data, starting from a realistic sketch of the present-day deformation, drawn with a major contribution of seismic moment tensor dataset (e.g., Vannucci et al. 2004).

In some parts of the Mediterranean, the complexity of geodynamic activity is shown by the coexistence of different focal mechanisms in short distances, few tens of km (Figs. 6 and 7). One example is the Northern Apennines, where normal, thrust, and transpressive (a mixture of strike-slip and compressive focal mechanisms) earthquakes occur together; another case is in the Northern Aegean Sea, where normal and strike-slip events represent the intersection between the transcurrent Northern Anatolian Fault motion occurring together with the extension of Aegean and Western Turkey region (Fig. 6a, b). The part of the Mediterranean where the plate collision exhibits a nearly constancy is Northwestern Africa, where most of earthquakes are thrust (Fig. 6c). However, also in this zone, strike-slip and few extensional events occur because even in this part of the boundary, the two main plates involved did not get in contact directly; they are separated by the Alboran Sea that is interpreted as a back-arc basin of a subduction system anciently active beneath Gibraltar Strait. Another part of the boundary where thrust earthquakes prevail is along the Hellenic Trench, where the African plate subducts under the Eurasian one (Fig. 6d). In fact, around it all types of focal mechanisms exist, related to the strike-slip faults that partition the trench (e.g., the Kefalonia Fault) or the



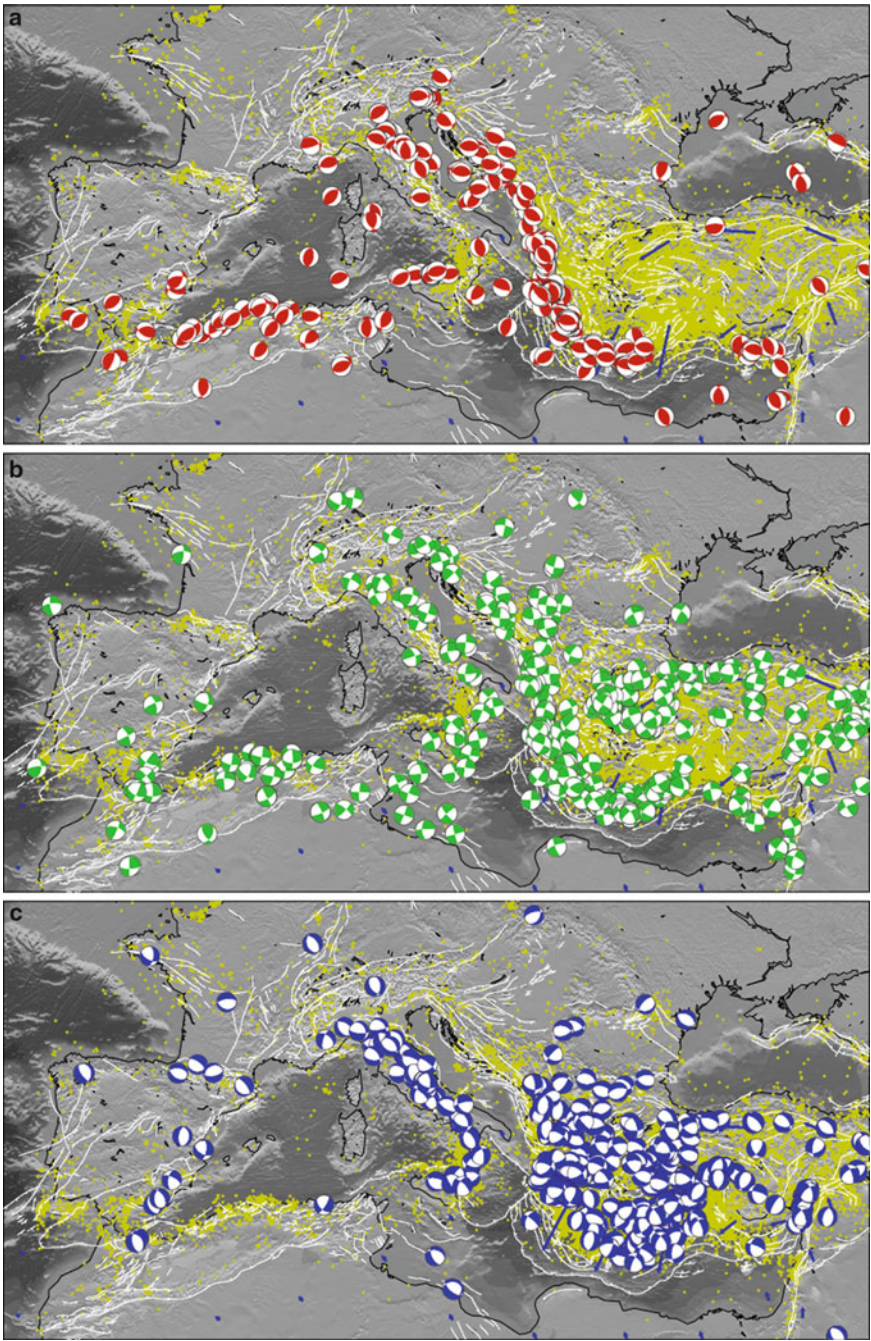
Regional Moment Tensor Review: An Example from the European–Mediterranean Region, Fig. 5 Map of all CMTs (*above*) and RCMTs (*below*) available in the European–Mediterranean region. CMT focal mechanisms are for earthquakes that occurred starting from 1977 to

2013 with a magnitude greater than 5.5, while RCMT solutions are for smaller seismic events, down to magnitude 4.0, occurred from 1997 to 2013. For seismicity and tectonic features in the background, see Fig. 4



Regional Moment Tensor Review: An Example from the European–Mediterranean Region, Fig. 6 Central map: main tectonic features of the Mediterranean region. Red lines are the principal active tectonic structures (Barrier et al. 2005). In yellow, a sketch of the plate boundary between Africa and Eurasia. GS, Gibraltar Strait; Al, Alboran Sea; BB, Balearic Basin; S, Sardinia; Tyr, Tyrrhenian Sea; AS, Adriatic Sea; HeT, Hellenic

Trench; AeS, Aegean Sea; An, Anatolian microplate; NA, Northern Anatolian Fault; EA, Eastern Anatolian Fault; Cy, Cyprus Island. In the four enlarged maps, all RCMTs and CMTs together for each zone are reported: (a) Northern Apennines; (b) transition between Northern Aegean and Marmara Sea following the Northern Anatolian Fault; (c) Alboran Sea and Northern Africa; (d) Northern Hellenic Trench (KF: Kefalonia Fault)



Regional Moment Tensor Review: An Example from the European–Mediterranean Region, Fig. 7 (a) Thrust (in red), (b) strike-slip (in green), and (c) extensional (in blue) seismic moment tensors for event with a

hypocentral depth shallower than 50 km mapped separately to better draw the different tectonic regimes active in the European–Mediterranean region. For seismicity and tectonic features in the background, see Fig. 4

extensional tectonics active in the front and back (e.g., in the Corinth Gulf).

In Fig. 7, in three separated maps, all thrust (in red), strike-slip (in green), and extensional (in blue) focal mechanisms are shown: these views help to study at larger scale the active tectonic styles. Thrust beach balls represent the compressive tectonic style and their distribution shows where the boundary between tectonic plates collides. This type of deformation occurs along a nearly continuous and narrow zone in Northern Africa, from Morocco to Northern Sicily, with two interruptions at the Gibraltar Strait (Fig. 6c) and south of Sardinia. Another continuous part of the tectonic boundary characterized by compressive features is along the eastern part of the Adriatic Sea, along Dinarides and Hellenides belts and partially around the Hellenic Trench. Other minor regions characterized by a compressive deformation are in the Northern Apennines and around Cyprus Island.

The distribution of strike-slip earthquakes is more diffuse (green in Fig. 7). In the Mediterranean, major strike-slip faults bound the Anatolian microplate (Fig. 6): at north the Northern Anatolian Fault can be clearly followed from the Marmara Sea to the Caucasus at east, while at southeast the Eastern Anatolian Fault runs from Cyprus Island to the Caucasus. Out of these two important tectonic structures, strike-slip focal mechanisms are however all along the Africa–Eurasia plate boundary. This is due to the fact that strike-slip earthquakes occur along the tectonic structures that dislocate different parts of the plate boundary and they accommodate its partitioning. The Africa–Eurasia plate boundary has a snake shape, and all of its curves are the concrete image of this partitioning, and in correspondence of most of them, strike-slip earthquakes prevail.

Normal focal mechanisms (blue in Fig. 7) characterize the areas where the lithosphere is affected by an extensional deformation, that is, where basins are opening or mountain belts are subjected to a collapse. In the Mediterranean, the two regions characterized by extensional tectonics are the Apennines and the whole Aegean Sea and surrounding regions (Fig. 6). Extensional

earthquakes typify also the Pyrenees and the Central Europe basins, such as the Rhine Graben.

A more careful description of earthquakes focal mechanisms in closer relation with tectonic structures of the European–Mediterranean region can be found in the *Atlas of Mediterranean Seismicity* (Vannucci et al. 2004).

Summary

The Mediterranean region is characterized by a complex tectonics and a widely distributed deformation, connected with the convergence between Africa and Eurasia plates. Smaller earthquake seismic moment tensors are extremely important to understand the geodynamics. The several methods used in this region to compute seismic moment tensors are described, with the main attention to RCMT computation process and to the European–Mediterranean RCMT catalog, the most relevant and most continuous initiative on regional scale.

Cross-References

- ▶ [Earthquake Mechanism Description and Inversion](#)
- ▶ [Earthquake Mechanisms and Tectonics](#)
- ▶ [Long-Period Moment-Tensor Inversion: The Global CMT Project](#)
- ▶ [Mechanisms of Earthquakes in Aegean](#)
- ▶ [Mechanisms of Earthquakes in Vrancea](#)
- ▶ [Moment Tensors: Decomposition and Visualization](#)
- ▶ [Non-Double-Couple Earthquakes](#)
- ▶ [Reliable Moment Tensor Inversion for Regional- to Local-Distance Earthquakes](#)

References

- Ardvisson R, Ekström G (1998) Global CMT analysis of moderate earthquakes $M_w > 4.5$ using intermediate period surface waves. *Bull Seismol Soc Am* 88:1003–1013
- Barrier E, Chamot-Rooke N, Giordano G, Morelli A, Brouillet J-F (2005) An innovative cartographic concept – the geodynamic map of the Mediterranean. *EPISODES* 28(3):193–196

- Bernardi F, Braunmiller J, Kradolfer U, Giardini D (2004) Automatic regional moment tensor inversion in the European–Mediterranean region. *Geophys J Int* 157:703–716
- Cesca S, Heimann S, Stammer K, Dahm T (2010) Automated procedure for point and kinematic source inversion at regional distances. *J Geophys Res* 115, B06304. doi:10.1029/2009JB006450
- Cucci L, Pondrelli S, Frepoli A, Mariucci MT, Moro M (2004) Local pattern of stress field and seismogenic sources in the Pergola–Melandro basin and the Agri valley (Southern Italy). *Geophys J Int* 156(3):575–583. doi:10.1111/j.1365-246X.2004.02161.x
- Dreger DS, Helmberger DV (1993) Determination of source parameters at regional distances with 3-component sparse network data. *J Geophys Res* 98:8107–8125
- Dziewonski AM, Chou T-A, Woodhouse JH (1981) Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *J Geophys Res* 86(B4):2825–2852
- Ekström G, Trömp J, Larson EWF (1997) Measurements and global models of surface wave propagation. *J Geophys Res* 102:8137–8158
- Ekström G, Morelli A, Boschi E, Dziewonski AM (1998) Moment tensor analysis of the Central Italy earthquake sequence of September–October 1997. *Geophys Res Lett* 25:1971–1974
- Giardini D, Boschi E, Palombo B (1993) Moment–tensor inversion from Mednet data: Regional earthquakes of the Mediterranean. *Geophys Res Lett* 20:273–276
- Heidbach O, Tingay M, Barth A, Reinecker J, Kurfes D, Müller B (2008) The world stress map database release 2008. doi:10.1594/GFZ.WSM.Rel2008
- Imprescia P, Pondrelli S, Vannucci G, Gresta S (2011) Regional centroid moment tensor solutions in Cyprus from 1977 to the present and seismotectonic implications. *J Seismol* 16(2):147–167. doi:10.1007/s10950-011-9254-7
- Kinscher J, Krüger F, Woith H, Luhr BG, Hintersberger E, Irmak TS, Baris S (2013) Seismotectonics of the Armutlu peninsula (Marmara Sea, NW Turkey) from geological field observation and regional moment tensor inversion. *Tectonophysics* 608:980–995. doi:10.1016/j.tecto.2013.07.016
- Konstantinou KI, Melis NS, Boukouras K (2010) Routine regional moment tensor inversion for earthquakes in the Greek region: the National Observatory of Athens (NOA) database (2001–2006). *Seismol Res Lett* 81:750–760. doi:10.1785/gssrl.81.5.750
- Kreemer C, Holt WE, Haines AJ (2003) An integrated global model of present-day plate motions and plate boundary deformation. *Geophys J Int* 154:8–34
- Kubo A, Fukuyama E, Kawai H, Nonomura K (2002) NIED seismic moment tensor catalogue for regional earthquakes around Japan: quality test and application. *Tectonophysics* 356(1):23–48 (26)
- Melis NS, Konstantinou KI (2006) Near real-time seismic monitoring in the Greek region: an example of the October 17, 2005 East Aegean Sea earthquake sequence. *Seismol Res Lett* 77(3):364–370
- Pondrelli S, Salimbeni S, Ekström G, Morelli A, Gasperini P, Vannucci G (2006) The Italian CMT dataset from 1977 to the present. *Phys Earth Planet Int* 159(3–4):286–303. doi:10.1016/j.pepi.2006.07.008
- Pondrelli S, Salimbeni S, Morelli A, Ekstrom G, Postpischl L, Vannucci G, Boschi E (2011) 1997–2008: 12 years of European-Mediterranean Regional Centroid Moment Tensor Catalog and its dissemination. *Phys Earth Planet Int* 185:74–81
- Pondrelli S, Salimbeni S, Perfetti P, Danecek P (2012) Quick RCMT solutions for the 2012 Emilia (Northern Italy) seismic sequence. *Ann Geophys* 55. doi:10.4401/ag-6146
- Randall G, Ammon C, Owens T (1995) Moment tensor estimation using regional seismograms from a Tibetan Plateau portable network deployment. *Geophys Res Lett* 22:1665–1668
- Roumelioti Z, Ganas A, Sokos E, Petrou P, Serpetsidaki A, Drakatos G (2007) Toward a joint catalogue of recent seismicity in western Greece: preliminary results. *Proceedings of the 11th international congress on geological society Greece, May 2007, Athens. Bulletin of the Geological Society of Greece, vol V XXXX, pp 1257–1267*
- Rueda J, Mezcuá J (2005) Near-real-time seismic moment-tensor determination in Spain. *Seismol Res Lett* 76:455–465
- Scognamiglio L, Tinti E, Michelini A (2009) Real-time determination of seismic moment tensor for the Italian region. *Bull Seismol Soc Am* 99:2223–2242
- Serpelloni E, Vannucci G, Pondrelli S, Argnani A, Casula G, Anzidei M, Baldi P, Gasperini P (2007) Kinematics of the Western Africa-Eurasia plate boundary from focal mechanisms and GPS data. *Geophys J Int* 00, 03367. doi:10.1111/j.1365-246X
- Sipkin SA (1994) Rapid determination of global moment tensor solutions. *Geophys Res Lett* 21(16):1667–1670. doi:10.1029/94GL01429
- Stich D, Mancilla F, Pondrelli S, Morales J (2007) Source analysis of the February 12th 2007, MW 6.0 Portugal earthquake: implications for the 1755 Lisbon earthquake. *Geophys Res Lett* 34:L12308. doi:10.1029/2007GL030012
- Tsuruoka H, Kawakatsu H, Urabe T (2009) GRiD MT (Grid-based Realtime Determination of Moment Tensors) monitoring the long-period seismic wavefield. *Phys Earth Planet Int* 175(Special issue: Earthquakes in subduction zones: a multidisciplinary approach):8–16
- Vallée M, Charléty J, Ferreira AMG, Delouis B, Vergoz J (2011) SCARDEC: a new technique for the rapid determination of seismic moment magnitude, focal mechanism and source time functions for large earthquakes using body-wave deconvolution. *Geophys J Int* 184:338–358
- Vannucci G, Pondrelli S, Argnani A, Morelli A, Gasperini P, Boschi E (2004) An atlas of Mediterranean seismicity. *Ann Geophys* 47:247–306

Reinforced Concrete Structures in Earthquake-Resistant Construction

Michael N. Fardis

Department of Civil Engineering, School of Engineering, University of Patras, Patras, Greece

Synonyms

Cast-in-situ buildings; Concrete frames; Concrete walls; Earthquake-resistant buildings; Flat-slab frames; Floor diaphragms; Foundation elements; Monolithic concrete buildings; Precast concrete systems; Wall-frame systems

Introduction

The *raison d'être* of most concrete structures is to create horizontal surfaces for use or occupancy (floors, a bridge deck, etc.) or, sometimes, for protection (a roof). Most of the mass, which generates the inertia forces in an earthquake, resides on these horizontal elements. Gravity loads travel from these elements to the ground via vertical elements, typically columns. Beams or girders often span between the columns to facilitate the collection of gravity loads from the horizontal surfaces and their transfer to the columns (Fig. 1). Concrete walls are often used in buildings (Fig. 2) to brace them laterally against second-order ($P-\Delta$) effects and to resist horizontal forces.

Walls can resist a horizontal earthquake very efficiently. And so does the combination of columns and floor beams – which exist anyway for the gravity loads – into frames. So, beam-column frames are the most common type of earthquake-resistant construction. Concrete walls are normally supplementary, because, unlike in masonry construction, it is not cost-effective to collect and transfer the floor gravity loads through concrete walls alone. So, the prime focus of this chapter is on concrete frames and walls, in their role as components of an earthquake-resistant system.

Inertia forces should find their way to the foundation via a smooth and continuous path in the structural system. From that point of view, cast in situ concrete is much better for earthquake-resistant construction than prefabricated elements – of concrete, steel, or timber – assembled on site: the connections between such elements create discontinuities in the flow of forces. So, cast in situ construction is the technique of choice for earthquake-resistant concrete structures, especially buildings; besides, monolithic construction is the quintessential use of concrete. For these reasons, section “[Frame, Wall, or Dual \(Frame-Wall\) Systems in Earthquake-Resistant RC Buildings](#),” below has cast in situ construction as the true focus, although precast systems are in principle covered as well. The special aspects of earthquake-resistant precast buildings are addressed in section “[Precast Concrete Systems](#)” supplementing section “[Frame, Wall, or Dual \(Frame-Wall\) Systems in Earthquake-Resistant RC Buildings](#),” in this respect. Section “[Floor diaphragms](#)” is devoted to concrete floors in their role as diaphragms, covering both fully cast in situ floors and those incorporating precast elements.

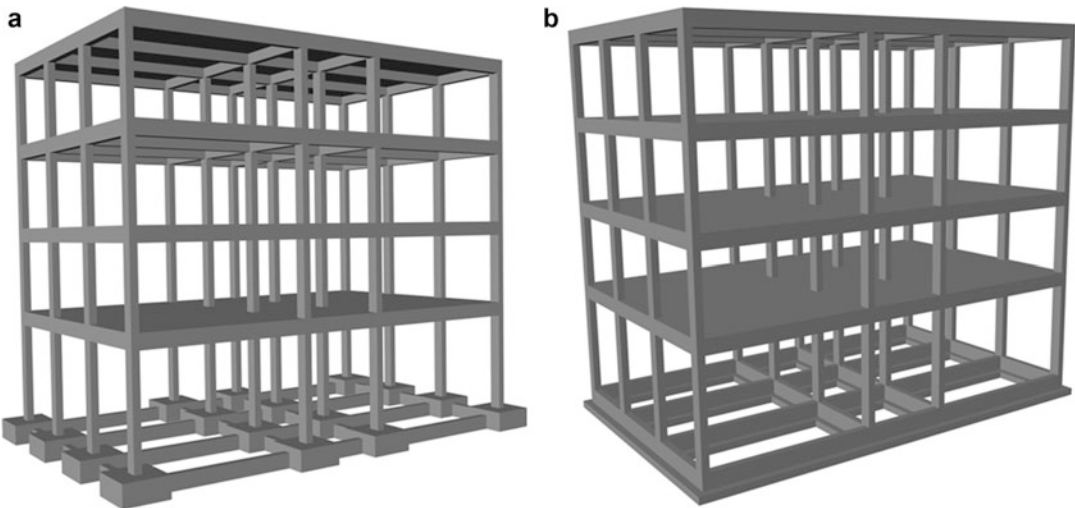
This chapter discusses the overall philosophy of earthquake-resistant reinforced concrete building design. It is intended for the experienced design professional and assumes familiarity with codes and frequently recurring problems in practice.

Frame, Wall, or Dual (Frame-Wall) Systems in Earthquake-Resistant RC Buildings

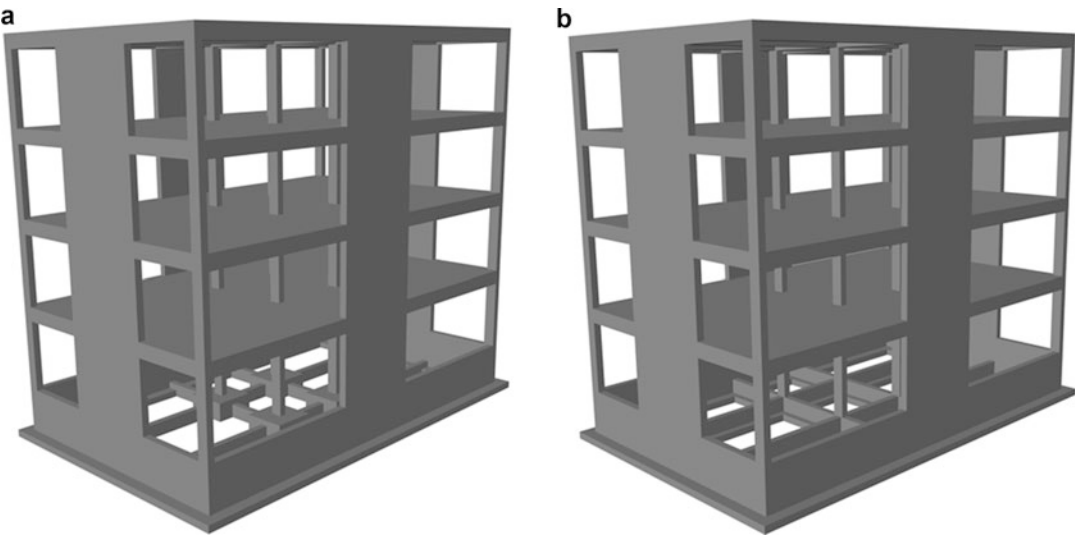
Frame Systems

Features of the Seismic Behavior of RC Frames

Walls resist the seismic overturning moments and the shears directly, through bending moments and shears, respectively, in the wall itself. By contrast, in a frame, the seismic overturning moment is resisted, not by the column moments but by axial forces in the columns (tensile at the windward side of the plan, compressive at the opposite – i.e., leeward – one, Fig. 3a).



Reinforced Concrete Structures in Earthquake-Resistant Construction, Fig. 1 Building with two-way frame system and foundation through (a) footings with two-way tie-beams and (b) two-way foundation beams



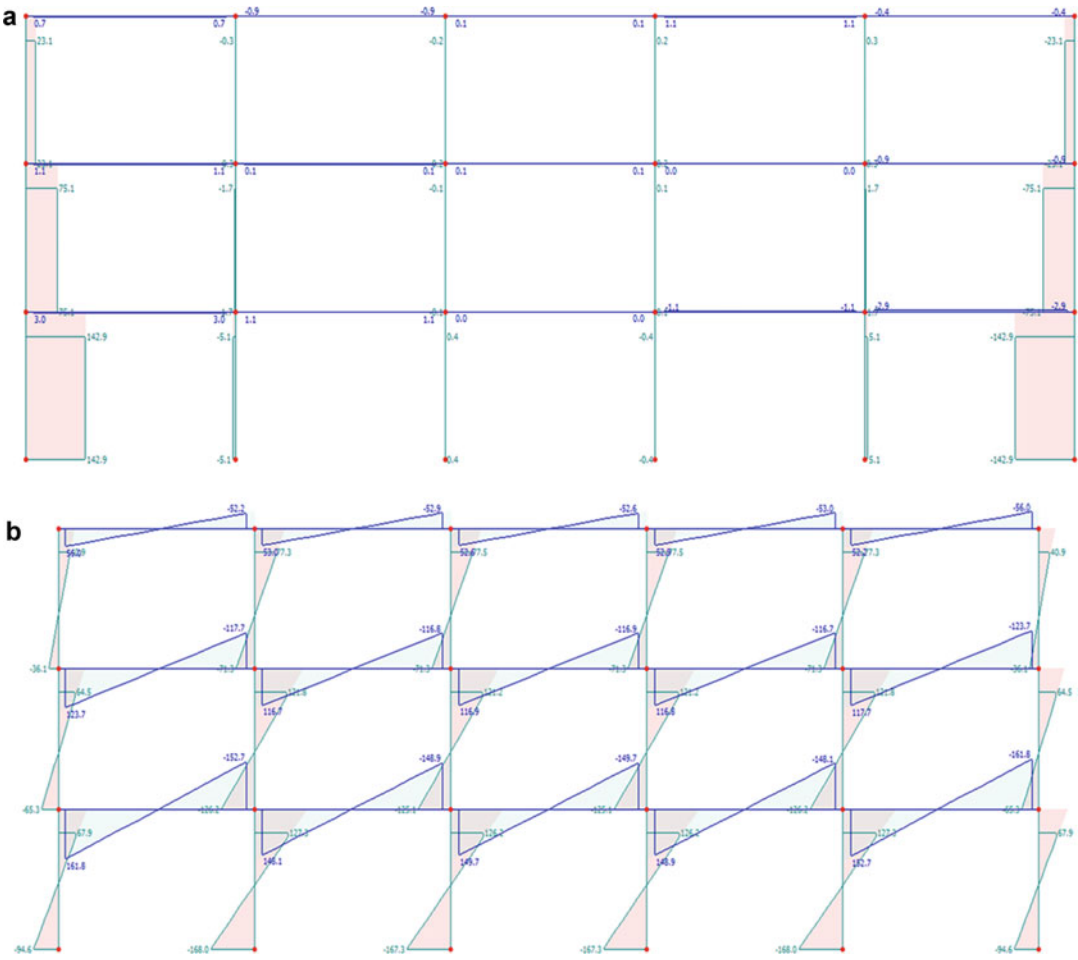
Reinforced Concrete Structures in Earthquake-Resistant Construction, Fig. 2 Frame-wall system with walls resisting lateral loads in both horizontal

directions; deep foundation beams for the walls; foundation of interior columns through (a) footings with two-way tie-beams and (b) two-way foundation beams

The column bending moments resist indirectly the seismic story shears: the algebraic difference between the bending moments at the top and the bottom of each column produces its contribution to the seismic shear of the storey (Fig. 3b). So, the seismic response of frame members is governed by flexure or strictly speaking by

normal action effects: bending moments and axial forces.

In a regular plane frame under lateral loading, the column inflection points are close to the story mid-height (Fig. 3b). Therefore, at that end of the column where the bending moment is larger, the shear span (moment-to-shear ratio) is



Reinforced Concrete Structures in Earthquake-Resistant Construction, Fig. 3 Typical seismic internal forces in a regular plane frame: (a) axial forces; (b) bending moments

normally between one-half and two-thirds of the column clear height. So, the most crucial components of a RC frame, its columns, normally have a shear span ratio (ratio of shear span to column depth) greater than 2.5. In beams, the most critical end is the one which, under earthquake loading, is under the largest hogging (negative) moment. The beam shear span ratio there is normally in the order of 3. For such values of shear span ratio, the inelastic behavior and failure of RC frame members are governed by flexure and are inherently ductile (provided that these members are designed not to fail in shear before their end sections yield in flexure).

Advantages and Disadvantages of RC Frames for Earthquake Resistance

The main advantage of RC frames for earthquake resistance is that, provided that they are capacity designed for plastic hinging in the beams and against preemptive shear failure of any member, their flexural behavior lends itself to the development of large global ductility and deformation capacity. Moreover, if there are several plane frames in the building, each one with several bays, the lateral-force-resisting system has high redundancy and offers multiple load paths. Therefore, with appropriate detailing of the end regions of their members, frames can be easily designed to resist strong earthquakes through

global ductility, rather than thanks to strength. For this reason, they are the structural system of choice in high seismicity regions, such as the Western USA and Japan. However, one can fully trust only fairly regular frames, having concentric beam-column connections. Strongly irregular frames and eccentric connections may have poor seismic performance. Besides, they are not sufficiently covered by present state of the art.

Although the seismic behavior of beams, columns, and frames made of reinforced concrete has been studied experimentally and analytically much more thoroughly than that of walls or wall systems, there are still significant uncertainties and gaps of knowledge about it. Examples are listed under disadvantage no. 4 below.

The advantages of RC frames for earthquake resistance may be summarized as follows:

1. The members of RC frames are inherently ductile.
2. Two-way frame systems, consisting of several plane frames in each horizontal direction with several bays each, have very high redundancy and multiple load paths.
3. Frames place few constraints on the architectural design, especially of the façade.
4. Provided that the frame has concentric connections and regular geometry, there is little uncertainty about its seismic response, because:
 - The seismic performance of frames and frame members is well known and understood, thanks to thorough experimental and analytical studies;
 - Frames are fairly easy to model and analyze for design purposes.
5. Certain features make frames attractive and cost-effective for earthquake resistance:
 - Beams and columns are needed in buildings anyway to support the gravity loads; so, why not use them for earthquake resistance as well?
 - Columns have strength and lateral stiffness against both horizontal components of the seismic action.
 - It is easy to design the foundation of smaller vertical elements (notably columns) than of larger ones (walls), with each foundation element transferring to the ground a small fraction of the seismic base shear.

Frames have also disadvantages for earthquake-resistant design:

1. Frames are inherently flexible and the size of their members is often governed by the inter-story drift limits under a moderate earthquake for which limitation of damage to structural and nonstructural elements is sought.
2. Column counterflexure within the same story may lead to soft-story mechanisms and a “pancake” type of collapse (Fig. 4).
3. Detailing of frames for ductility requires workmanship of high quality and strict supervision on site (especially for fixing the dense reinforcement and placing and compacting the concrete through the beam-column joints of two-way frames).
4. There are certain elements of uncertainty about the seismic response and performance of frames:
 - The effects of eccentric connections or strongly irregular layouts in 3D are not sufficiently known.
 - There is considerable uncertainty about the magnitude of the effective slab width in tension: it is pointed out that slab bars which are in this effective flange width and are parallel to the beam increase its flexural capacity for hogging moment and hence the beam capacity design shears, as well as the likelihood of plastic hinging in the columns.
 - The behavior of columns of two-way frames under the complex loading conditions (cyclic biaxial bending with varying axial force) to which they are subjected during real earthquakes is poorly known.
 - In two-way frames subjected to real earthquakes, columns that have been capacity-designed against plastic hinging may well form plastic hinges owing to the biaxial moment demands from beams connected to them in the two horizontal directions.



Reinforced Concrete Structures in Earthquake-Resistant Construction, Fig. 4 Soft-story mechanism due to plastic hinges in the columns of a frame (*left*) and pancake type of collapse (*right*)

Conceptual Seismic Design of RC Frames

The location of frames in plan, the span lengths, and often the depth of beams are normally controlled – sometimes even dictated – by architectural and functional considerations, as well as by design for gravity loads alone. However, the structural designer is normally left with considerable latitude for decisions and choices on the basis of seismic design considerations alone.

If a regular plane frame has:

- A constant span length in all bays,
- Beams and interior columns with constant cross section in each story, and
- An (effective) rigidity of the two exterior columns equal to 50 % of that of interior ones,

then the elastic seismic bending moments at the ends of all beams in the story will be the same, while those in the two exterior columns will be just half of the bending moments in the interior columns. In the elastic range, the seismic overturning moment will be resisted by axial forces in the exterior columns alone, approximately equal to the ratio of the seismic overturning moment at story mid-height to the distance between the axes of these columns. If members are dimensioned for the elastic seismic moments, all beam ends in a story will be subjected to about the same inelastic chord rotation demand; all columns, interior or exterior, will also develop about the same inelastic rotation demand at the story bottom; the same at column tops. This uniformity reduces uncertainty

about the distribution of seismic action effects among the members of the frame and within them.

Too short or too long bay lengths should be avoided:

- If the beam span is very long, the beam top reinforcement over the supports may be governed by factored gravity loads, rather than by seismic design. This penalizes the capacity design moments of columns at the joints and the capacity design shears of beams and columns. It produces flexural overstrength in beams with respect to the corresponding seismic demands and creates uncertainty about the inelastic response and the plastic mechanism. If plastic hinges do indeed form in a long beam, the large hogging moments at beam ends due to the concurrent gravity loads may prevent reversal of the inelastic flexural deformations there. Although the primary effect of non-reversal is positive, there is a collateral negative one: inelastic elongations accumulate in the reinforcement and the beam gradually grows longer, pushing out its supporting columns and possibly forcing exterior ones to separate from exterior beams at right angles to the elongating one(s). Last but not least, if the average clear beam span in the story is much longer than the clear storey height, then the cross section of the beam may have to be increased excessively for the story to meet the drift limits for purposes of damage control.

- Very short beam spans cause very high seismic shears in the beams, both from the seismic analysis and from the capacity design of the beam in shear. If the span is short in some bays but long in others, the high seismic shears in the short beams generate a large variation of the axial force in the adjoining columns upon reversal of the direction of the seismic action. This reversal will also cause an almost full reversal of the sign of shear at the ends of the short beam(s), because the concurrent gravity loads produce insignificant shear forces in short beams, especially as these beams usually support the secondary direction of floor slabs. A full reversal of high shears may exhaust the shear capacity in both diagonal directions or cause sliding shear failure along through-depth cracks at the end section(s) of the beam. Design against such shear effects may require diagonal reinforcement in the beam or shear reinforcement at $\pm 45^\circ$ to the beam axis. Last but not least, short beams have a low shear span ratio (often below 2.5) and, unless diagonally reinforced, low deformation capacity.

For the most common story heights and for ordinary gravity loads, the optimum beam span in earthquake-resistant buildings is between 4 and 5 m. Span lengths should be as uniform as possible within each frame.

RC Wall Systems

Definition of a Wall

Design codes define a concrete wall as a concrete vertical element with an elongated cross section. A limit of 4.0 for the aspect ratio (long-to-short dimension) of a rectangular cross section is conventionally adopted by most design codes to distinguish walls from columns. If a composite cross section consists of rectangular parts, one of which has an aspect ratio greater than 4, then this element is also classified as a wall. With this definition on the basis of the cross-sectional shape alone, a wall differs from a column in that:

- It resists lateral forces mainly in one direction, notably parallel to the long side of the section.

- It can be designed for such a unidirectional resistance by assigning flexural resistance to the two edges of the section (“flanges” or “tension and compression chords”) and shear resistance to the “web” between them, as in beams.

So for the purposes of flexural resistance and deformation capacity, one may concentrate the vertical reinforcement and provide concrete confinement only at the two edges of the section (Fig. 5). (If the cross section is not elongated, the vertical element is called upon to develop significant lateral-force resistance in both horizontal directions, then it is meaningless to distinguish between “flanges” on one hand, where vertical bars are concentrated and concrete is confined, and “web” on the other, where they are not).

The above definition of “walls” is consistent with concrete design codes and appropriate for dimensioning and detailing at the level of the cross section. It is not very meaningful though, in view of the intended role of “walls” in the structural system and of the correct practice in design, dimensioning, and detailing of walls, notably as entire elements and not just at the cross-sectional level. Seismic design often relies on walls for the prevention of a story mechanism in their long direction, without any verification that plastic hinges form in beams rather than in columns. Nevertheless, walls can enforce a beam-sway mechanism only if they act as vertical cantilevers (i.e., if their bending moment diagram does not change sign within at least the lower story(s) and they develop a plastic hinge only at the base (at the connection to the foundation)). As a matter of fact, unless the wall bending moment attains two values of large magnitude but of opposite sign within the full height of the wall (let alone within the same story), the wall cannot develop two plastic hinges in opposite bending (positive and negative) along its height and a story mechanism between them. Whether a “wall,” as defined above, will indeed act as a vertical cantilever and form a plastic hinge only at its base depends not so much on the aspect ratio of its section but primarily on how stiff and



Reinforced Concrete Structures in Earthquake-Resistant Construction, Fig. 5 Typical layout of reinforcement in wall section

strong the wall is relative to the beams it is connected to at story levels. For concrete walls to play their intended role, the length dimension of their cross section, l_w , should be large, not just relative to its thickness, b_w , but in absolute terms too (see Fig. 2 for examples). To this end, and for the beam sizes commonly found in buildings, a value of at least 1.5 m for low-rise buildings or 2 m for medium- or high-rise ones is recommended here for l_w .

A vertical element with a cross-sectional aspect ratio less than 4.0 (i.e., defined conventionally as a “column”) may work in an earthquake as a vertical cantilever and form a plastic hinge only at the base, if it is connected at story levels with very flexible beams or with no beams at all (as in flat-slab systems). However, normally the moment resistance at the base of such a vertical element is small, so that, given its long shear span (moment-to-shear ratio) there, it cannot contribute significantly to the base shear of the building. Moreover, its lateral stiffness is also low, so it is not effective in reducing inter-story drifts for damage limitation or P- Δ (second-order) effects. At the other extreme, vertical elements with cross sections sufficiently elongated to be classified as walls, but connected at story levels with very stiff and strong beams, may act as frame columns rather than as vertical cantilevers.

Optimal Length of Walls

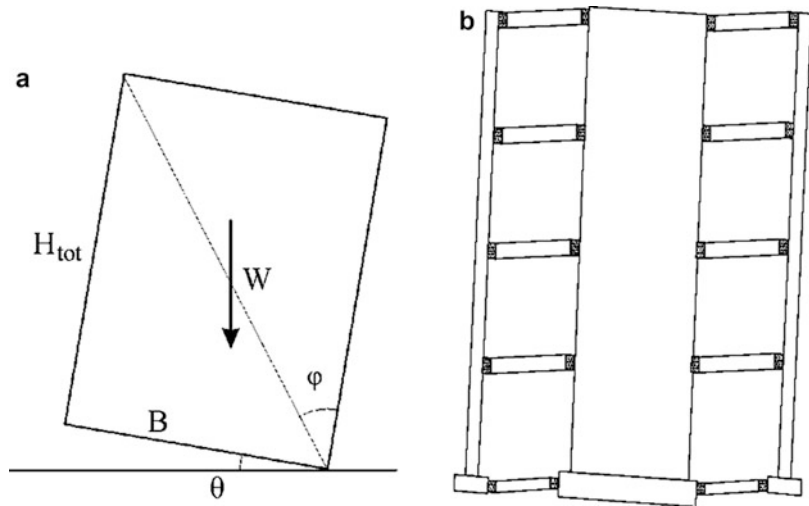
If the lateral-load-resisting system comprises only walls, the full seismic overturning moment at the base of the building is resisted directly by the (sum of) bending moments at the base of the walls, instead of indirectly by their axial forces. So at the base of a wall, the bending moment is large and the shear span (M/V ratio), L_s , long. If the beams are very flexible compared to the walls, each wall works as a vertical cantilever

subjected only to horizontal forces at story levels. Then L_s is about equal to $2/3$ of the total wall height, H_{tot} . For the usual beam sizes, L_s is about half of H_{tot} if the length l_w of the wall section is fairly large or about equal to 1.5 times the story height if l_w is short, near the limit of 4 times the width, b_w .

The shear strength of a concrete element is roughly proportional to its cross-sectional depth, h . Therefore, lumping the shear resistance into a few vertical elements with large values of h , instead of distributing it to many small-sized ones, does not save materials. On the other hand, the rigidity of RC members is roughly proportional to h^3 . So lumping the lateral stiffness into few vertical elements, rather than spreading it to many small ones, is very cost-effective. Using few walls with a large cross section is also cost-effective from the point of view of moment resistance and vertical reinforcement, because for a given base shear, V , and concrete volume per linear meter of building height, i.e., for a given wall cross-sectional area, the reduction of the shear span ratio, L_s/h , reduces the total vertical reinforcement ratio necessary to resist the wall base moment (Fardis 2009). As explained in the previous paragraph, for a wall with a fairly large value of $h = l_w$, the shear span, L_s , is roughly a known fraction of the building height. Therefore, the total vertical reinforcement ratio necessary to resist the wall base moment can be reduced by increasing h as much as reasonably feasible, i.e., by lumping the vertical elements into a few large walls. The optimum value of h is the one that gives values of L_s/h in the range between 2.5 and 3.0, below which the cyclic behavior of the wall and its ultimate deformation may be adversely affected by shear. For the typical value $L_s \approx 0.5H_{\text{tot}}$, a shear span ratio equal to $L_s/h = 3.0$ gives $l_w = h \approx H_{\text{tot}}/6$, i.e., $l_w \approx n_{\text{story}}/2$ for a typical story height of 3 m.

Reinforced Concrete Structures in Earthquake-Resistant Construction,

Fig. 6 Rocking of a large wall: (a) single rigid wall; (b) wall on footing connected to normal-size tie-beams



Foundation of Walls

The base section of a large and strong wall has a high moment resistance. It is difficult to transfer this large moment capacity to the ground through isolated footings. The maximum bending moment that an isolated footing can transfer to the ground is slightly less than the moment causing the footing to overturn, which is equal to $0.5NB$, where N is the vertical force and B is the dimension of the footing in the direction of bending. If the wall section is long, the parallel dimension of its footing is not much longer than the wall length $h = l_w$. So the maximum value of the nondimensional base moment in the wall, $\mu \equiv M/(bh^2f_c)$, that can be transferred by an isolated footing is $\mu \leq 0.5v$, where $v \equiv N/(bhf_c)$ is the nondimensional axial load at the base of the wall. Because walls have relatively low values of v (in the order of $v \approx 0.05$), the maximum value of μ that can be transferred to the ground through an isolated footing is also very low, in the order of the nondimensional moment at cracking of the wall base section (as the tensile strength of concrete is typically around 10 % of its compressive strength, f_c). Therefore, in order to develop its moment resistance at the base, a strong wall:

- Should be provided with a very large isolated footing, which is not cost-effective and introduces large uncertainty about the seismic response (it will uplift, following a nonlinear

relation between bending moment and uplift rotation which is hard to quantify, at least within everyday design practice), or

- Should be fixed at the top of a box-type foundation provided for the building as a whole (see section “[Box Type Foundation Systems](#)”), which fixes all walls at their base and maximizes their effectiveness

An isolated footing will uplift from the ground when the moment at its bottom exceeds the value corresponding to decompression at its edge. Then it starts rocking, with the value of the moment at its base approaching, but never reaching, the overturning moment of the footing, $0.5NB$, where B is the dimension of the footing in the vertical plane where rocking takes place (Fig. 6a). The wall and its footing will rock as a rigid body (Fig. 6b). As the seismic action is not static but dynamic, rocking is a very stable mode of response for the wall, provided that the concentrated force at the edge of the footing does not bring about bearing capacity failure of the foundation soil.

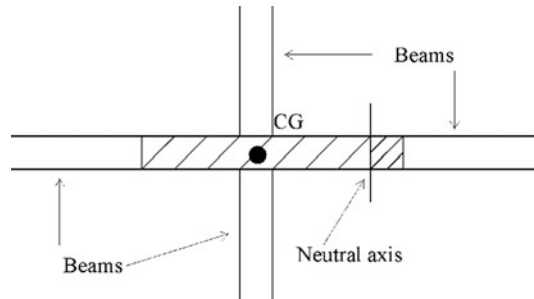
The relationship between the lateral force and the horizontal displacement during rocking under cyclic loading is nonlinear but nearly elastic, following an approximately bilinear envelope and recentering to approximately zero displacement for zero force. Therefore, rocking of a wall on an isolated footing may be considered as

a ductile mode of seismic response, almost as ductile as that of a fixed wall with a flexural hinge at the base. For this reason, some international guideline documents acknowledge for rocking a force reduction factor of the same order as that applying to ductile structural response. However, there is so much uncertainty about rocking – notably about its implications for the seismic action effects within the superstructure – that it cannot be reliably quantified and modeled in the context of earthquake-resistant design of wall systems.

Large RC Walls

Walls of large cross-sectional length which are fixed at the base exhibit certain features in their seismic response which resemble rocking of a wall with isolated footing. This is because, owing to the small magnitude of the wall axial force, the neutral axis of the cracked (or yielded) base section of the wall is very close to the edge of the section (Fig. 7) and far from its centroid (CG in Fig. 7). Therefore, flexure of the wall significantly lifts the centroid of the gross section and, with it, the tributary mass of the building supported on the wall. So part of the kinetic energy is temporarily and harmlessly converted to potential energy of these tributary masses, in lieu of damaging deformation energy of the wall itself. Besides, the end of any beams framing into the part of the wall section outside the compression zone is also raised (Fig. 7). So long as the other end of the beam is not lifted by the same amount, the beam shear force has a stabilizing effect on the wall, acting downward, increasing the wall axial compression to the benefit of its instantaneous strength, stiffness, and stability, while reducing the resultant moment on the section.

The above beneficial aspects of the wall behavior are due to the large horizontal dimension of the wall, combined with the no-tension feature of cracked concrete (similar to the interface between a footing and the ground) that causes the flexural rotation to take place about a pivot near the edge of the wall section. These phenomena are of purely geometric origin (due to coupling of the rotations with the vertical



Reinforced Concrete Structures in Earthquake-Resistant Construction, Fig. 7 Schematic plan view of wall section

displacement at the centroid of the wall section) and are neglected in ordinary geometrically linear analyses of the response, even when the analysis accounts for material nonlinearity; they are sources of good performance of structural systems consisting of large walls that are usually not taken into account.

Walls with a large horizontal dimension compared to their height cannot be designed effectively for energy dissipation through plastic hinging at the base, as they cannot be easily fixed there against rotation with respect to the rest of the structural system. Design of such walls for plastic hinging at the base is even more difficult, if the wall is monolithically connected with one or more transverse walls which are also large enough to be considered merely as “flange(s)” or “rib(s)” of the first wall. Walls with large horizontal dimensions will most likely develop little cracking and inelastic behavior under the design seismic action. Cracking will be mainly horizontal, at construction joints at floor levels. Flexural yielding, if it occurs, will also take place mainly there. Then the lateral deflections of large walls, acting as vertical cantilevers, will be produced by a combination of:

- A rotation of the foundation element of the wall with respect to the ground, most often with uplifting
- Similar rotations at sections of horizontal cracking and (possibly) flexural yielding at one or more floor levels, with the wall swaying as a stack of rigid blocks



Reinforced Concrete Structures in Earthquake-Resistant Construction, Fig. 8 Examples of systems of large walls

Owing to the rather low axial load level in large walls, these rotations will take place about a “neutral axis” very close to the compressed tip of the foundation element or to the compressed edge of the wall section at locations of cracking and (possibly) yielding. Such rotations induce significant uplift of the centroid of the sections, raising the floor masses tributary to the wall and the ends of beams framing into it to the benefit of the global response and stability of the system. Moreover, rigid-body rocking of the wall with its footing promotes radiation damping, which is particularly effective in reducing the high-frequency components of the input motion.

Eurocode 8 (CEN 2004) recognizes the ability of large walls to withstand large seismic demands through their geometry, rather than via strength and hysteretic dissipation due to vertical reinforcement. It defines a “large lightly reinforced wall” as a wall with a horizontal dimension, l_w , at least equal to 4.0 m or to two-thirds of its height, h_w (whichever is less) and gives it a special role. It also provides special design and detailing rules for such walls, allowing much less reinforcement than in “ductile walls,” provided that they belong in a lateral-load-resisting system consisting mainly of such walls (Fardis et al. 2005). Figure 8 shows examples of systems of large walls.

Walls with Non-rectangular Section or with Openings

Most of what has been said so far in section “**RC Wall Systems**,” as well as almost the entire body of knowledge about the cyclic behavior of concrete walls, is about walls with two-way-

symmetric rectangular or quasi-rectangular section (barbelled section, i.e., rectangular with each edge widened into a rectangular or square “column” or compact flange – with an aspect ratio of less than 4 – to enhance the moment resistance and prevent lateral instability of the compression zone). Such walls are modeled and dimensioned as prismatic elements with an axis passing through the centroid of their section. Lacking a better alternative, the same practice is applied when a rectangular wall runs into or crosses another wall at right angles to create a wall with a composite cross section of more than one rectangular parts – each part with an aspect ratio greater than 4 (L-, T-, C-, H-shaped walls, etc.). Such walls have high stiffness and strength in both horizontal directions. So they are subjected to biaxial bending and bidirectional shears during the earthquake. They are more cost-effective than the combination of their constituent parts as individual rectangular walls. However, present-day knowledge of their behavior under cyclic biaxial bending and shear is very limited, and the rules used for their dimensioning and detailing still lack a sound basis. Moreover, their detailing for ductility is complex and difficult to implement on site. For this reason, it is strongly recommended to make limited use of such walls in practical design. Designers choosing to use non-rectangular walls should opt for fairly simple geometries (e.g., one-way-symmetric C-section or two-way-symmetric H-section).

Large openings should be avoided in ductile walls, especially near the base, where a plastic hinge will form. If they are necessary for

functional reasons (e.g., for doors or windows), openings should be arranged at every floor following a very regular pattern, turning the wall into a coupled one, with the lintels between the openings serving and designed as coupling beams.

Advantages and Disadvantages of Walls for Earthquake Resistance

On the basis of what has been said so far in section “[RC Wall Systems](#),” the advantages of RC wall systems for earthquake-resistant design may be summarized as follows:

1. Walls are inherently stiff, so:
 - They are insensitive to the presence and any adverse effects (global or local) of masonry infills.
 - They prevent or limit damage in frequent or occasional earthquakes.
2. Walls offer excellent protection against collapse, as the lack of wall counterflexure within a story makes a soft-story mechanism physically impossible (Fig. 6b).
3. The seismic behavior and performance of individual walls is less sensitive than that of frames to lower quality design or poor workmanship on site.
4. Geometric effects and phenomena in large walls are favorable for the seismic response and performance.
5. All things considered, RC walls are more cost-effective for earthquake resistance than RC frames.
5. To avoid large eccentricities or low torsional stiffness of the stories, walls that contribute significantly to lateral stiffness and strength (e.g., those around service cores housing elevators, stairways, vertical piping, etc., close to the center in plan, or large perimeter walls) should be counterbalanced in plan by other elements with similar lateral stiffness and strength.
6. It is difficult to provide an effective foundation to a wall, especially with isolated footings.
7. There is large uncertainty about the seismic response of wall systems:
 - The cyclic behavior and seismic performance of RC walls and wall systems are less known than those of frames, because experimental research is more difficult to carry out and analytical models need to be more advanced and sophisticated.
 - The effects of rocking or of the rotations about the neutral axis of the wall cannot be accounted for reliably in practical design.
 - Walls are more complex to model, analyze, dimension and detail in practical design (especially those with a non-rectangular section).

Wall systems also have disadvantages for earthquake-resistant design:

1. They are inherently less ductile than RC beams or columns, more sensitive to shear effects, and harder to detail for ductility.
2. They offer limited redundancy and few alternative load paths.
3. They limit the freedom of the architectural layout, especially at the façade.
4. It is not cost-effective to use RC walls alone to support the gravity loads of the building; some beams and columns are needed anyway for that purpose.

Dual Systems of Frames and Walls

Both walls and frame systems, each have their advantages and disadvantages as lateral-load-resisting systems (see sections “[Advantages and Disadvantages of Walls for Earthquake Resistance](#)” and “[Advantages and Disadvantages of RC Frames for Earthquake Resistance](#)”, respectively). Although walls seem to have a better balance of advantages and disadvantages, one should keep in mind that there are almost always beams and columns in a building to carry gravity loads to the ground. It is a waste not to use them at all for earthquake resistance. So, it is cost-effective for earthquake resistance to combine within the same structural system frames and walls.

Many buildings combine frames and walls in their lateral-load-resisting system; sometimes the frames clearly dominate, sometimes the walls, while often none is clearly dominant; it is in such cases that the system is considered as

a dual (wall-frame) system. Eurocode 8 (CEN 2004) uses as a formal criterion to distinguish between systems the percentage of the seismic base shear taken, according to the linear analysis for the seismic action, by all the frames in the system or by all its walls:

- A “dual system” is one in which the percentage of the seismic base shear taken by the frames or the walls is between 35 % and 65 %; moreover, such a system is classified as a “wall-equivalent dual” or a “frame-equivalent dual,” if the percentage of the base shear taken by walls is between 50 % and 65 % or between 35 % and 50 %, respectively.
- If frames take more than 65 % of the seismic base shear, the lateral-load-resisting system is a “frame system.”
- If walls take more than 65 % of the seismic base shear, the lateral-load-resisting system is a “wall system.”

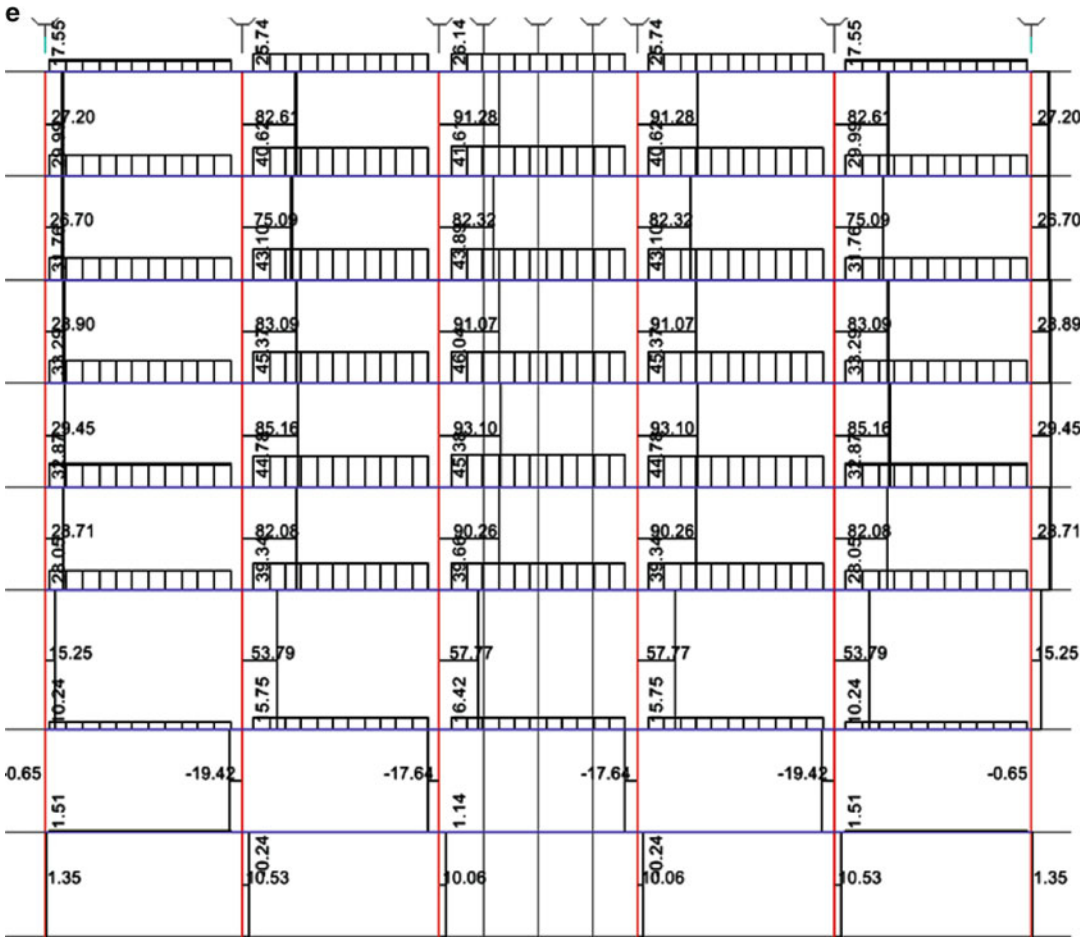
Dual systems combine the high strength and stiffness and insensitivity to soft-story effects of wall systems (advantages no. 1 and 2 of walls in section “[Advantages and Disadvantages of Walls for Earthquake Resistance](#)”) with the large ductility, deformation capacity, and redundancy of frames (advantages no. 1 and 2 of frames in section “[Advantages and Disadvantages of RC Frames for Earthquake Resistance](#)”). The walls offer protection from nonstructural damage in frequent, moderate earthquakes and help to meet the inter-story drift limits prescribed in seismic design codes. The frames may act as a second line of defense in very strong earthquakes, in case the deformation capacity of the inherently less ductile walls is exhausted and some walls lose a good part of their strength and stiffness. In view of this potential backup role of frames, US codes require the frames of dual systems to be designed for at least 25 % of the design seismic action, no matter the relative stiffness of the walls and the frames in the system. If this condition is met, US codes entitle dual systems of high ductility frames and walls to as high a value of the force reduction factor, R , as in frame systems.

Eurocode 8 (CEN 2004) assigns also the same value of the force reduction factor, q , to “frame systems” and to “dual systems,” i.e., to any system where frames take at least 35 % of the design base shear according to a linear seismic analysis.

Frames subjected to lateral loading have shear-beam-type lateral displacements, with the floors sliding horizontally with respect to each other. Inter-story drifts follow the height wise pattern of story seismic shears: they decrease from the base to the roof. By contrast, walls fixed at the base deflect laterally like vertical cantilevers, i.e., their inter-story drifts increase from the base to the roof. If frames and walls are combined in the same structural system, floor diaphragms impose on them roughly common floor displacements. As a result, the walls restrain the frames at lower floors, undertaking there the full inertia loads of the floor. Near the top of the building the frame is called upon not only to resist the full floor inertia loads but also to hold back the wall, which – if acting alone – would have developed a very large deflection at the top. In rough approximation, the walls of dual systems may be considered to be subjected to:

- The full inertia loads of all floors
- A concentrated force at the roof level, in the reverse direction with respect to the peak seismic response and the floor inertia loads

The magnitude of the concentrated force at the top exceeds that of the resultant inertia loads in the upper floors (i.e., of the story seismic shear there). Consequently, in the upper stories, the walls are often under reverse bending and shear with respect to the stories below (Fig. 9a–c). The frame may be considered to be subjected to just a concentrated force at the top, equal and opposite to the one applied there to the wall(s) and in the same sense as the floor inertia loads. Then the frame has roughly constant seismic shear in all stories and hence about the same bending moments in all of them (Fig. 9d, e). As a result, even when the cross-sectional dimensions of frame members are kept the same in all stories, their reinforcement requirements for the seismic



Reinforced Concrete Structures in Earthquake-Resistant Construction, Fig. 9 Seismic moments and shears in wall-frame system with box-type of basement: (a) schematic moment (M) and shear (V) diagrams in interior wall of seven-story building plus basement; (b)

example M-diagram in interior wall of six-story wall-frame building with two basement storeys; (c) V-diagram of wall in (b); (d) example M-diagram in interior frame of building in (b), (c); (e) V-diagram of frame in (d)

action do not increase from the top to the ground level. As a matter of fact, column reinforcement requirements may even decrease in lower stories, thanks to the favorable effect of the increased axial force on flexural strength. Therefore, in dual systems, column size should not decrease toward the roof.

Dual systems are geometrically more complex than frame or wall systems and have a more complicated seismic response than either one of them. Therefore, there is larger uncertainty about their seismic behavior and performance. This uncertainty is the main (if not the only) inherent drawback of dual systems against pure frame or wall

systems. Their conceptual design should aim at reducing this uncertainty. For instance, as the diaphragms of dual systems are called upon to impose common floor displacements to the two systems by transferring horizontal forces from the frame to the wall or vice versa, they should be thicker and stronger within their plane than what is required in pure frame systems. Another uncertainty arises from any rocking of the wall (s) at the base (Fig. 6). Such rocking shifts part of the story shears from the wall(s) to the frame. Rocking of wall footings with uplift is an intrinsically complex phenomenon, which cannot be reliably modeled in the framework of seismic

design practice. Underestimation of such rocking leads to unsafe design of the frames, while overestimation is unsafe for the wall(s). So, a prudent design would eliminate rocking in dual systems by providing full fixity of the walls at the foundation.

Note that in a system consisting only of walls, the distribution of seismic shear between them will be practically unaffected by the rotation of all the walls at the foundation level; the rotation will only increase the absolute magnitude of story drifts. The effect of the rotation of foundation elements is even smaller in pure frame systems: any rotation of the footing of a frame column has practically no consequences beyond the ground story; moreover, such a rotation will be much smaller than in a wall footing, because the higher axial load of the column resists uplift; more importantly, the smaller the cross section of a vertical element compared to the plan dimensions of its footing, the smaller its rotation. So, it is the design of systems that combine the two types of elements (walls and columns) that suffers from increased uncertainty, owing to the rotations of footings with respect to the ground.

Tall buildings often have a dual system comprising a strong wall near the center in plan (around a service core housing elevators, stairways, vertical piping, etc.) and stiff and strong perimeter frames. In such a system outrigger beams may be used to advantage, increasing the global lateral stiffness and strength of the system and mobilizing the perimeter frames against the seismic overturning moment.

Flat-Slab Frames

Beamless slabs (“flat slabs” called “flat plates” in North America, if supported on columns directly, without drop panels or column capitals) provide larger clear story height, unobstructed passage of services under the slab, and freedom for irregular layout of the column grid and potential modification of the layout of partitions. Moreover, if labor is expensive, they may be cost-effective for residential or office buildings. Most common are solid cast-in-place slabs, sometimes post-tensioned with bonded or unbonded tendons. Waffle slabs with drop panels around the columns

are also common. “Lift slabs” are all precast at ground level around the columns and lifted to their final position.

In flat-slab frames subjected to lateral loading, strips of the flat slab between the columns act and behave as beams. The effective width of such strips increases with increasing seismic demands but is quite uncertain. Regardless of this uncertainty, the stiffness and flexural capacity of the strips is relatively low compared to those of columns, conducive to a beam mechanism with column plastic hinging only at the base, as in strong-column-weak-beam designs. Owing to the flexibility of the flat slab, flat-slab frames may develop large second-order ($P-\Delta$) effects. There is also large uncertainty about the behavior of the region of the slab around the column under inelastic cyclic loading, and especially about its capacity to transfer to the column the floor gravity loads through vertical shear stresses, along with the slab moment due to the cyclic lateral loading. Experimental work on slab-column connections subjected to large amplitude cyclic deformations suggests that slab-column connections having a large safety margin against punching shear failure under the concurrent gravity loads can sustain without failing significant cyclic deformation demands (Fardis 2009).

According to currently accepted conventional wisdom, beamless frames of columns and flat slabs (“flat-slab frames”) are not considered suitable for earthquake resistance, owing to questions about their lateral displacement capacity. Indeed, buildings with “flat-slab frame” systems suffered heavy damage in the Northridge (1994) earthquake. However, in earthquakes that have inflicted heavy damage to urban centers in Europe, such buildings either performed fairly well, despite the lack of proper design and detailing for earthquake resistance, or were damaged in the vertical elements supporting the flat slab, but not at the connection. So, the statement that “flat-slab frames” are doomed in strong earthquakes is inconclusive. Nevertheless, owing to the gaps in knowledge mentioned above, modern seismic design codes do not have rules for the design and detailing of flat slabs as part of a ductile lateral-load-resisting system.

Precast Concrete Systems

Introduction

As pointed out in section “[Introduction](#),” the connections between precast elements are points of discontinuity in the flow of inertia forces from the masses to the ground: they may fail, either by failure of the joining element itself or due to damage of the concrete around the fastenings of this element. Depending on their connections, earthquake-resistant systems of precast elements:

1. Emulate monolithic construction, with a continuous and smooth force path, or
2. Are designed to safely accommodate the seismic displacements in connections which are weaker than the precast elements they connect (“jointed” systems)

Emulation of Monolithic Construction

Seismic design codes typically address only precast systems emulating monolithic construction, granting to them the same strength and ductility as those of geometrically similar cast-in-place structures.

The connections between the elements belong to one of the following types (fib [2003](#)):

1. Energy-dissipating ductile connections, based on yielding of those longitudinal bars of the elements which enter the connection and flexural plastic hinging at the end of the connected element as in cast-in-place construction; the connection and the connected elements follow the same code rules as in monolithic construction.
2. Overdesigned (“strong”) connections, capacity-designed to remain elastic while flexural yielding occurs at another location in the element, away from the connection.

“Jointed” Precast Systems

This type of construction is not covered in prescriptive seismic design codes. Some codes allow, though, experimental demonstration of “energy-dissipating connections” not complying with the prescriptive rules for precast or cast-in-place structures.

“Jointed” systems normally employ “dry connections” formed by welding or bolting reinforcing bars or steel embedments and dry packing or grouting. Inelastic seismic deformations do not take place in flexural plastic hinges in the connected elements, but at their interfaces, which continuously open and close during the earthquake.

A common example is tilt-up construction of precast walls, normally designed to remain essentially elastic in an earthquake.

Less common are innovative systems, which employ at the connections unbonded or partially bonded post-tensioning tendons alongside non-prestressed longitudinal bars and allow the whole structure to rock with respect to the ground, or its components to rock in a stable manner relative to each other as rigid bodies. Sometimes supplemental energy dissipation is added to reduce the displacement response. Such systems were proposed and tested in the end of the last century (Nakaki et al. [1999](#)). They comprise:

- Frames of precast concrete beams connected “dry” to precast columns via concentric unbonded post-tensioned strands, continuous through the joint
- Precast concrete walls, connected “dry” to the foundation via vertical concentric unbonded tendons

Such systems were applied during the first decade of the century in high seismicity areas: San Domingo (Stanton et al. [2003](#)), San Francisco (Englekirk [2002](#)), and Christchurch (where a building with this system survived unscathed the 2011 earthquake). However, the robustness of the system against progressive collapse under accidental extreme events should be demonstrated as well.

Floor Diaphragms

At any horizontal level where significant masses are concentrated, the vertical elements of the lateral-force-resisting system should be connected through floor diaphragms, transferring

the inertia forces from the masses to the lateral-load-resisting system and tying it together into an integral whole. The diaphragms should be designed to remain elastic while performing this in-plane action, alongside the out-of-plane flexure resisting the floor gravity loads and transferring them to the vertical elements.

Normally a cast-in-place solid slab qualifies as a diaphragm, if it is monolithically connected to the elements of the lateral-force-resisting system and has thickness and two-way reinforcement which meet the requirements of design codes concerning the flexural strength under factored gravity loads, the limits on deflection and crack width under service loads, and the minimum slab reinforcement.

In a floor incorporating precast elements, the in-plane shear transfer at the joints between these elements is questionable; it is strongly recommended to rely only on a cast-in-plane RC topping – for which Eurocode 8 (CEN 2004) requires at least 50 mm of thickness and the minimum slab reinforcement in both horizontal directions. The topping should be cast over a clean, very rough substrate; otherwise, shear connectors to the precast floor elements should be provided. If there is no topping meeting all these requirements, the joints between precast elements should be crossed by reinforcement capable of transferring the in-plane forces of the diaphragm.

Cases requiring special attention – and possibly verification of the diaphragm by calculation – include:

- Diaphragms with irregular geometry in plan, recesses and reentrances, irregular and large openings, etc.
- Irregular in-plan distribution of the masses and/or of the stiffness of the lateral-force-resisting system (e.g., with set-backs or offsets, vertical elements terminating below the top level, etc.)
- The roof of a basement with RC walls along only a part of the perimeter or of the ground floor area
- Cast-in-place waffle slabs with a thin, lightly reinforced top slab.

Shallow Foundation Systems

Introduction

In modern construction, the foundation system is practically always made of concrete, even when the superstructure is built of another material.

The term “shallow foundation” covers systems of isolated footings (pads) and tie-beams, foundation-beams, and rafts. The same types of foundations are often called “spread foundations.” Deep foundations, through piles, are not commonly used in buildings and are not treated here.

The role of the foundation is to transfer the gravity loads from the vertical members of the structure to the ground. So the natural choice for the foundation of a concrete column is to widen its base in order to adapt the section area, through which the vertical load passes, to the ground bearing capacity – lower than that of the concrete section. Each one of the resulting footings – normally concentric and square – is often connected to neighboring ones via horizontal tie-beams of rectangular section. The main role of tie-beams is to reduce the magnitude and impact of differential settlements of adjacent footings, due to large imbalances between their vertical loads and/or variations in the underlying soil. If a more interconnected foundation is essential, instead of placing a number of tie-beams and isolated footings in a row, a foundation beam is used: a deep beam with an inverted-T or L section, which transfers, through the underside of its bottom flange, vertical loads to the ground all along its length, not just around the column base. If the overall weight of the building and its contents is so large that the soil bearing capacity needs to be mobilized over most of its footprint area, it is normally cost-effective to combine all foundation elements in a raft, which acts as a single footing under the entire building, transferring vertical loads to the ground throughout its plan area.

Therefore, there are three types of shallow foundation systems for buildings, listed below

in ascending order of cost and effectiveness in transferring the gravity loads to the soil:

1. Isolated footings, with tie-beams (Figs. 1a and 2a) or without
2. (Two-way) Foundation beams (Figs. 1b and 2b)
3. Foundation rafts

Isolated Footings and Tie-Beams

If the building is subjected to significant seismic lateral loads, the foundation elements should be designed to transfer to the ground the large bending moments which develop at the base of each vertical element, as well as the action effects of the overall overturning moment due to the lateral loads, notably the uplifting of the “windward” side of the building and the additional vertical compression at the “leeward” side. In isolated footings, the vertical force that the footing must transfer to the ground acts at a large eccentricity (i.e., moment-to-vertical-force ratio) with respect to the center of the footing’s underside, especially when the vertical force due to the overall overturning moment is tensile. To accommodate this eccentricity, footings may have to be large in plan. To reduce the eccentricity, stiff tie-beams may connect every footing with the adjacent ones in both horizontal directions. The bending moments (in counterflexure) and the shears in tie-beams always have a sense of action that leads to a reduced moment reaction of the ground at the underside of the footing. So, for given moments and forces applied at the top of the footing by the vertical element, the larger the stiffness of the tie-beams, the smaller this moment reaction is at the base of the footing.

Tie-beams should be connected directly to the body of the footing in order to:

- Avoid creating a squat column between the soffit of the beam and the top of the footing
- Increase the effectiveness of tie-beams, by increasing their stiffness, EI/L_{cl} , through the reduction of their clear length, L_{cl} .

Tie-beams have another important role in design against lateral loads: to prevent differential horizontal slippage of footings, a role which may be played also by a horizontal slab between them (not integral to them).

Two-Way Systems of Foundation Beams

It is clear from Section “Isolated Footings and Tie-beams” that isolated footings may not be cost-effective for an earthquake-resistant building, even on competent ground. If the tie-beams have to be very deep and stiff to alleviate the eccentricity problem of footings, they may as well be merged with them into a foundation beam. That beam will work over its entire length as a single integral member, absorbing into its own bending moment diagram, the bending moments applied at its top flange by the columns. Thanks to its long overall length, it transfers efficiently to the ground the overall overturning moment applied to the beam by the columns it supports, with very little uplift of its windward end. So, especially for high seismicity, a two-way system of foundation beams, supporting all vertical elements of the building throughout the plan of the foundation, is the system of choice, especially for a tall building, regardless of the competence of the soil. It is also more cost-effective.

Perimeter frames are often stiffer than interior ones; sometimes they include walls. So they resist a larger share of the seismic action, while, owing to their smaller tributary area, they support a smaller fraction of the gravity loads. So their foundation may be more challenging than that for interior vertical elements. If the perimeter frames include walls, they may need deep foundation beams to prevent differential rocking of the walls at the base (Fig. 2). If there is a basement, the perimeter foundation beams can extend up to its top slab, at least along most of the length of each side of the perimeter. In that case, the building can be provided with the most effective foundation system for earthquake resistance, namely, the box system described in the next section.

Unlike two-way foundation beams, a foundation raft does not offer additional advantages for seismic design. Moreover, its analysis and design

are demanding even for gravity loads: the raft should be discretised with a fairly fine mesh of plate finite elements, each node being supported on the soil through a vertical Winkler spring. So, there is no special reason to choose a raft over a two-way system of foundation beams for purposes of earthquake resistance.

Box-Type Foundation Systems

The ideal shallow foundation for earthquake-resistant buildings is a box system, consisting of (Fardis et al. 2005):

1. Wall-like deep foundation beams along the entire perimeter of the foundation, possibly supplemented with interior ones across the full length of the foundation system. These deep beams are the main foundation elements transferring the seismic action effects to the ground. In buildings with a basement, the perimeter foundation beams may also serve as basement perimeter walls.
2. A concrete slab at the level of the top flange of the perimeter foundation beams (playing the role of the roof of the basement, if there is one), acting as a rigid diaphragm.
3. A foundation slab, or a grillage of tie-beams or foundation beams, at the lowest level of the perimeter foundation beams.

Such a system suits well buildings with basement(s), even those only partially embedded and/or with some openings along the perimeter. However, for the system to be fully effective, any openings between the top of the perimeter foundation beam and the soffit of the beam supporting the basement roof should be limited to a small fraction of the corresponding side of the perimeter.

If their perimeter foundation beam is deep enough to accommodate a basement story, the foundation system of the examples in Fig. 2 may be turned into a box type one by adding a slab at the top level of the foundation beams.

It is not essential to have a raft or a two-way system of foundation beams as in Fig. 2b at the lowest level of the deep foundation beams along the perimeter. Interior vertical elements may be

founded instead on individual footings, provided that the footings are connected to the bottom of the perimeter foundation beams through a diaphragm consisting of two-way tie-beams (as in Fig. 2a) or a slab (possibly serving also as the basement floor). As the transfer of the full seismic action from the ground to the superstructure (or vice versa) takes place through the perimeter beams, the foundation of interior elements and the abovementioned diaphragm may be at a level slightly above the bottom of the perimeter beams.

Owing to its large rigidity and strength, a box-type foundation works as a rigid body. Thus, it minimizes uncertainties about the distribution of seismic action effects over the interface between the ground and the foundation system and imposes the same rotation of all vertical elements at the level of their connection with this system, so that they may be considered as fixed there against rotation. Plastic hinges in walls and columns will develop just above the top of the box-type foundation. Fixity of interior vertical elements at the top of the foundation system is achieved through a couple of horizontal forces that develop at the levels of the top and bottom of this system. The main role of the diaphragm at the lowest level of the perimeter foundation beams is to provide the lower-level horizontal force for the fixity of vertical elements at the top of the foundation box.

The large rigidity and strength of a box-type foundation system protects the interior columns within the box and all its beams (including those at the roof of the basement) from high seismic moments and shears (see Fig. 9d, e) and from plastic hinging (except possibly at the column tops underneath the roof of the basement). However, interior walls develop large and, as a matter of fact, reverse seismic shears within the basement (see Fig. 9a, c), which makes their shear design challenging.

The seismic bending moment at the connection of interior vertical elements with their foundation element is small, and a footing (subjected in this case to essentially concentric compression) is sufficient for each one of them (see Fig. 9a, b, d).

The top slab of a box-type foundation transfers the seismic shears from the interior vertical elements to the perimeter wall-type foundation beams. It should have sufficient in-plane stiffness (i.e., thickness) and strength (two-way reinforcement) and should be free of large interior openings. The restraint of shrinkage of this slab by the stiff perimeter elements may induce through thickness cracking. Two-way reinforcement at both surfaces of the slab will reduce the consequences of such cracking on its in-plane stiffness. If the slab is cast shortly after the perimeter elements of the box to minimize their differential shrinkage, cracking due to the restraint of drying shrinkage is less likely. Another option is to lap-splice all slab reinforcement in each direction of the slab along a strip, all the way from one side of the perimeter to the opposite, and defer for a few days casting of the concrete in that strip and striking of the formwork of the entire slab. In this way, the early (and largest) part of the ultimate drying shrinkage of the parts of the slab cast first takes place without any restraint.

A box-type foundation offers additional advantages, not captured by conventional analysis of the seismic response (Fardis 2009):

- An integral rigid foundation for the whole building filters out differences in the seismic input between different points of the base, due to arrival there of the seismic waves with a time-lag and other minor differences. So it introduces in the superstructure a single seismic excitation, which at any given point in time is the average of the ground motion over the entire interface (horizontal and vertical) between the foundation system and the ground. This smoothing of the input motion removes high-frequency components. The larger the interface between the ground and an integral foundation (e.g., in buildings with a raft at the bottom or with deep embedment due to more than one basement stories), the more extensive the smoothing of the input motion.
- If the system of the foundation and the superstructure is stiff relative to the underlying soil stratum (more specifically, if its fundamental

period is shorter than that of the soil deposit) and, moreover, there is an integral and rigid foundation system for the entire building, some of the input energy, instead of being trapped within the building, is radiated back into the ground (“radiation damping”), reducing the seismic forces and deformations in the superstructure.

- In the extremely rare, but not inconceivable, case of surface faulting through the foundation (which happened at the Awaji island in Kobe, 1995, and at the Bolu viaduct in the Duzce, 1999, earthquake), an integral and rigid foundation may straddle the fault without collapse or severe damage in the superstructure due to the fault displacements – horizontal or vertical.

Summary

Earthquake-resistant concrete structures are typically cast in situ: monolithic construction suits best the nature of concrete as a material and helps ensure a smooth and continuous force path, which is essential for earthquake resistance. To serve this goal, precast concrete systems typically emulate monolithic connections; otherwise, they are designed to accommodate at the joints differential seismic displacements between the connected elements. Beam-column frames are very reliable for resistance to strong earthquake through energy dissipation and ductility; concrete walls are cost-effective for earthquake resistance and offer, through their stiffness, protection from damage under smaller or moderate earthquakes. “Dual” earthquake-resistant systems combine frames and walls and their advantages but are more challenging for a sound seismic design. Flat-slab frames are not trusted to resist the earthquake alone, because of their lateral flexibility and insufficient knowledge concerning the behavior of slab-column connections. Two-way foundation beams are much more efficient for earthquake-resistant construction than isolated footings, be it with two-way tie-beams; a box-type foundation is the best for that purpose but suits more buildings with basement(s).

Notation of Symbols

- B** Horizontal dimension of footing
 b_w Width of cross section
 h Depth of cross section
 H_{tot} Wall height
 l_w Wall length
 L_s Shear span (=M/V ratio)
 M Bending moment
 N Axial force in a vertical member
 V Shear force
 $\mu \equiv M/(bh^2f_c)$ Nondimensional bending moment
 $\nu = N/(bhf_c)$ Nondimensional axial force in a vertical member

Cross-References

- [Behavior Factor and Ductility](#)
- [Earthquake Response Spectra and Design Spectra](#)
- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- [Plastic Hinge and Plastic Zone Seismic Analysis of Frames](#)
- [Response Spectrum Analysis of Structures Subjected to Seismic Actions](#)
- [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)
- [Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling](#)
- [Seismic Strengthening Strategies for Existing \(Code-Deficient\) Ordinary Structures](#)
- [Soil-Structure Interaction](#)

References

- CEN (2004) European Standard EN1998-1:2004. In: Eurocode 8: design of structures for earthquake resistance. Part 1: general rules, seismic actions and rules for buildings. Comité Européen de Normalisation, Brussels
 Englekirk RE (2002) Design-construction of the Paramount – a 39-story precast prestressed concrete apartment building. PCI J 47(4):56–71
 Fardis MN (2009) Seismic design, assessment and retrofitting of concrete buildings (based on EN-Eurocode 8). Springer, Dordrecht

- Fardis MN et al (2005) Designers' guide to EN 1998-1 and EN 1998-5 Eurocode 8: design of structures for earthquake resistance. General rules, seismic actions, design rules for buildings, foundations and retaining structures. Thomas Telford Publishers, London
 fib (2003) Seismic design of precast concrete buildings, Bulletin 27. federation internationale du beton, Lausanne
 Nakaki S et al (1999) An overview of the PRESSS 5-story precast test building. PCI J 44(2):28–39
 Stanton J et al (2003) Design of the Cala building using PRESSS technology. In: Proceedings of the Fib symposium: concrete structures in seismic regions, Athens

Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach

N. Ganesh and Sayan Gupta
 Department of Applied Mechanics, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India

Synonyms

First passage; Karhunen Loeve expansions; Polynomial chaos; Random fatigue; Smolyak's quadrature; Sparse grid

Introduction

Failures in structural systems subjected to random dynamic loadings can be classified essentially into two broad categories. Failures that occur when the structure response exceeds safe threshold limits during its service life are due to overloading and are deemed to occur at the instant when the response crosses the threshold for the first time. These failures are known as first passage failures (Nigam 1983; Lin 1967). The alternative mode of failure in vibrating system occurs when the accumulated damage in the structure reaches a threshold limit due to the gradual degradation in the structure material on account of fatigue. It is obvious that for random loadings, both the first passage time and the

fatigue damage are random quantities and a reliability analysis of the structure against these modes of failure requires characterization of these quantities.

Reliability analysis of structures against random dynamic loadings is a problem in time-variant reliability analysis. Adopting a probabilistic approach provides a rationale framework within which the uncertainties in the loading and in the structure response can be adequately handled leading to quantification of the failure probability of the structure. In this approach, the random dynamic loads are typically modeled as stochastic processes. Consequently, the structure responses are also stochastic processes whose probabilistic characteristics are different from that of the loading. Here, the structure which behaves as a filter to the excitations significantly alters the probabilistic characteristics of the uncertainties associated with the loadings. Typically, the structure response can be viewed as a nonlinear transformation of the loading and are random processes in time. Time-variant reliability analysis of the structures requires characterization of the response processes.

When the structure response is a stochastic process, the time when the response exceeds the safe limits or when the total accumulated damage reaches a critical threshold are random quantities, quantification of which can be carried out using theories of probability and stochastic processes. The crux in estimating either the first passage failure probabilities or the random fatigue damage lies in evaluating the crossing statistics of the response using the well-known Rice's integral (Rice 1944). This requires estimating the joint probability density function of the response and its instantaneous time derivative. This, however, is not easy as even for Gaussian loadings the response of nonlinear structures is non-Gaussian whose marginal probability density function (pdf) itself is difficult to estimate.

First Passage Failures

Estimating the first passage failures in randomly vibrating structures requires estimating the probability of exceedance of the structure response for

the first time within the specified time duration. Let $Y(t)$ be the response of a randomly vibrating structural system; see Fig. 1. Let the dotted line be the safe threshold level the crossing of which constitutes a failure due to overloading. The time at which the first crossing of the response across the threshold level from below is defined as the first passage time, denoted by T_f . Clearly, for a random time history $Y(t)$, T_f is a random variable. If no failure is deemed to occur within the time period of interest, T , it implies that $Y(t)$ lies below the threshold α for all time t . Therefore, the mathematical statement for the failure probability against overloading can be expressed as

$$P_f = 1 - P[Y(t) \leq \alpha; \forall t \in (t_0, t_0 + T)], \quad (1)$$

where, t_0 is initial time, T is the duration, α is the safe threshold limit, P_f is failure probability, and $P[\cdot]$ is probability measure. Mathematical evaluation of P_f from Eq. 1 demands the knowledge of the infinite dimensional probability density function of $Y(t)$ which is an impossibility.

Alternatively, Eq. 1 can be written in the time invariant format by introducing a variable Y_m , as (Melchers 1999)

$$P_f = 1 - P[Y_m \leq \alpha] = 1 - P_{Y_m}(\alpha), \quad (2)$$

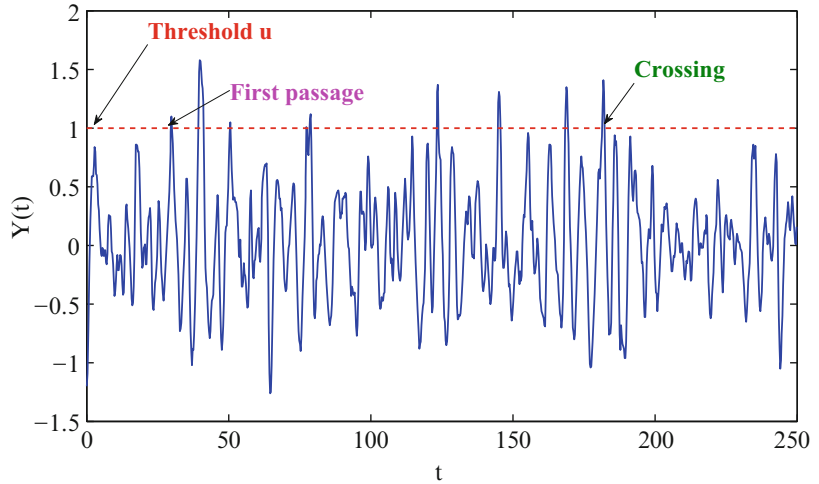
where, Y_m is the extreme value of the process $Y(t)$ given by

$$Y_m = \max_{t_0 \leq t \leq t_0 + T} \{Y(t)\}$$

and $P_{Y_m}(\alpha)$ is the probability distribution function (PDF) of Y_m and is referred to as the extreme value distribution (EVD) for the process $Y(t)$ in time duration T . If the threshold levels are sufficiently high, such that, the crossings can be assumed to be rare and independent, the crossings can be modeled as a Poisson counting process. Under these assumptions, it can be shown that the EVD can be approximated as (Cramer 1966),

$$P_{Y_m}(\alpha) = \exp \left[- \int_0^T \int_0^\infty \dot{y} p_{Y\dot{Y}}(\alpha, \dot{y}, t) \dot{y} d\dot{y} dt \right]. \quad (3)$$

Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach,
Fig. 1 Schematic diagram of first passage failure



Here, $p_{Y\dot{Y}}(\alpha, \dot{y}, t)$ is the joint transitional probability density function of the process and its instantaneous time derivative. It can be shown that for stationary processes $Y(t)$, Eq. 3 can be expressed in the simpler form

$$P_{Y_m}(\alpha) = \exp[-v_Y(\alpha)T], \quad (4)$$

where, $v_Y(\alpha)$ is the mean upcrossing intensity of the process $Y(t)$ across level α . The expression for estimating the mean crossing intensity $v_Y(\alpha)$ for a stationary process $Y(t)$ is available from Rice's formula (Rice 1944) and is given by

$$v_Y(\alpha) = \int_0^\infty \dot{y} p_{Y\dot{Y}}(\alpha, \dot{y}) d\dot{y}. \quad (5)$$

Here, $p_{Y\dot{Y}}(\cdot, \cdot)$ denotes the joint stationary pdf of the process $Y(t)$ and its instantaneous time derivative $\dot{Y}(t)$.

It must be noted here that the expression for the extreme value distribution in terms of the mean crossing intensity given in Eq. 4 is approximate whose quality depends on the accuracy of the assumptions of the level crossings being Poisson. Thus, as the levels become higher, the crossings become rarer and satisfy the assumptions of a Poisson counting process. Thus, at high levels α , Eq. 4 approaches the exact value. However, for lower levels of α , the quality of the approximation deteriorates as the assumptions

of a Poisson counting process are not met. Thus, the estimates of the failure probability based on the extreme value distributions are more accurate at the tails of the distribution. However, irrespective of the level α , the expression for the mean crossing intensity, given by Eq. 5, is always exact.

Random Fatigue Damage

For random response $Y(t)$ in a vibrating system, the accumulated linear fatigue damage due to $Y(t)$ is denoted by D_T and is given by

$$D_T = \sum_{j=1}^{Y(t)} \frac{1}{N(u, v)} \quad (6)$$

where, $N(u, v)$ denotes the strength of the structure material and is equal to the number of constant amplitude cycles, with amplitude range (u, v) that lead to fatigue. Usually,

$$N(u, v) = \frac{1}{f(u, v)} = \frac{1}{k(v - u)^a}, \quad (7)$$

where, k and $a \geq 1$ are material constants. Clearly, if $Y(t)$ is a random process, the stress amplitude ranges defined by (u, v) pair are random and obviously D_T is a random variable. For complete characterization of the random fatigue damage, one needs to characterize the pdf of D_T . This is, however, a difficult task. A simpler approach to

characterization of the random variable D_T would be to estimate its first moment, defined as $E[D_T]$ where, $E[\cdot]$ is the expectation operator.

Estimates of the mean fatigue damage can be carried out in the time domain using Monte Carlo simulations. This would imply simulation of a large ensemble of random time histories for $Y(t)$, evaluating D_T corresponding to each time history and taking their mean. This is however computationally too expensive, especially as fatigue damage is a gradual degradation mechanism and would require simulations of very long time histories of $Y(t)$. However, an alternative spectral approach for approximating the expected fatigue damage has been discussed in the literature. This is discussed next.

Assuming rainflow cycle counting, it can be shown that the expected fatigue damage is given by Rychlik (1993)

$$E[D_T] = T \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{12}(u, v) \mu(u, v) du dv, \quad (8)$$

where, $\mu(u, v)$ is known as the intensity of interval upcrossings, $E[\cdot]$ is the expectation operator, and $f_{12}(\cdot)$ is the second derivative of a function related to the damage accumulation rule. Estimating $\mu(u, v)$ is however a very difficult task. An easier option would be to estimate the bounds for intensity crossings. It can be shown that (Rychlik 1993)

$$\mu(u, v) \leq \min[\mu(u, u), \mu(v, v)] = \hat{\mu}(u, v), \quad (9)$$

where, $\mu(u, u) = \mu(u)$. The mean crossing rate $\mu(u)$ can be easily computed using the Rice's formula given in Eq. 5.

Difficulties in Estimating the Mean Crossing Intensity

The discussion in this section shows that the mean crossing intensity plays a crucial role in estimating the failure probability of vibrating structures, irrespective of the failure being defined in terms of first passage failures or random fatigue failures. The numerical evaluation of Rice's integral does not pose much challenge, even if the integral cannot be evaluated

analytically. However, the crux here lies in the knowledge of the joint pdf $p_{Y\dot{Y}}(\cdot, \cdot)$. For structural response which can be modeled as Gaussian, $Y(t)$ is Gaussian as well as its instantaneous time derivative. In such situations, $p_{Y\dot{Y}}(\cdot, \cdot)$ can be exactly obtained, which enables an analytical representation of $v_Y(x)$. However, it must be noted that $Y(t)$ is the structure response when subjected to random loadings, say $X(t)$. Thus, mathematically, one can write

$$Y(t) = g[X(t)], \quad (10)$$

where, the function $g[\cdot]$ behaves like a filter and depends on the structure behavior. Clearly, if $X(t)$ is Gaussian and $g[\cdot]$ is linear, $Y(t)$ is a Gaussian process. However, usually $g[\cdot]$ is nonlinear. In these situations, even if $X(t)$ is Gaussian, $Y(t)$ is non-Gaussian whose marginal pdf $p_Y(y, t)$ is itself difficult to estimate. Estimating the joint pdf $p_{Y\dot{Y}}(\cdot, \cdot)$ is an even harder task. This poses serious challenges in estimating the crossing statistics, and in turn, the failure probabilities against overloading or fatigue.

Studies exist in the literature which have focused on developing approximations for the joint pdf $p_{Y\dot{Y}}(\cdot, \cdot)$, especially when $X(t)$ is a Gaussian scalar or a vector process and $g[\cdot]$ is a quadratic transformation. In this entry, a methodology is presented by which one can derive analytical/numerical approximations for the crossing statistics of the response of nonlinear structures subjected to random vibrations, using a spectral approach. This involves transforming the problem into a mathematical subspace spanned by the basis vectors obtained from the projections of the spectral content of the input process. The basis functions are obtained from a Karhunen-Loève representation of the stochastic loading. Subsequently, the response is obtained as a polynomial chaos expansion in the same basis space. Finally, quantification of the probabilistic characteristics of the response are obtained.

The primary focus of this discussion is on the use of numerical algorithms when polynomial chaos expansion (PCE) based representation of the response is sought. After a brief introduction to polynomial chaos expansion, discussions focus

on the use of three methods that are used in representing the response using PCE. These include the stochastic Galerkin method, the stochastic collocation technique, and, finally, the sparse grid approach using Smolyak's algorithm. A comparative study is presented on the efficiency of these methods. Finally, the estimates of the failure probability are obtained against first passage failures and against random fatigue. The predictions obtained from PCE have been validated using full-scale Monte Carlo Simulations. The methodology has been demonstrated through a numerical example of obtaining the response of a vibrating 2-D airfoil subjected to gusty wind flow regime.

Karhunen-Loève Expansion

According to the Karhunen-Loève (K-L) theorem, a stochastic process on a bounded interval can be represented as an infinite linear combination of orthogonal functions, the coefficients of which constitute uncorrelated random variables. The basis functions in K-L expansions are obtained by eigendecomposition of the autocovariance function of the stochastic process and are shown to be its most optimal series representation. The deterministic basis functions, which are orthonormal, are the eigenfunctions of the autocovariance function and their magnitudes are the eigenvalues. The Karhunen-Loève expansion converges in the mean-square sense for any distribution of the stochastic process (Papoulis and Pillai 2002). A K-L representation of a zero-mean stochastic process $f(t, \theta)$ can be represented in the form

$$f(t, \theta) = \sum_{i=0}^{\infty} \xi_i(\theta) \sqrt{\lambda_i} \phi_i(t), \quad (11)$$

where, the coefficients λ_i and the functions $\phi_i(t)$ respectively are the eigenvalues and the eigenfunctions of the covariance function of the process, denoted by $R_{ff}(t, s)$, and are evaluated by solving the Fredholm integral equation of the second kind, given by

$$\int_0^T R_{ff}(t, s) \phi_i(s) ds = \lambda_i \phi_i(t). \quad (12)$$

Here, the parameter t indicates time and θ represents the random dimension. Note that the autocorrelation function $R_{ff}(t, s)$ need not be stationary. In Eq. 11, $\{\xi_i(\theta)\}$ is a vector consisting of uncorrelated random variables with zero mean and unit variance. Since $\phi_i(t)$ are eigenfunctions, it is obvious that

$$\int_0^T \phi_i(t) \phi_j(t) dt = \delta_{ij}, \quad (13)$$

where, δ_{ij} is the Kronecker-delta function. This implies that $\{\phi_i(t)\}$ constitute a set of orthonormal vectors and can be viewed as basis vectors.

Since Eq.11 is an infinite series, it becomes imperative to decide on the number of terms to be retained in the expansion of $f(t, \theta)$. It is seen that not all terms have a significant contribution in the expansion. If the series expansion is truncated after N terms, then

$$\tilde{f}(t, \theta) = \sum_{i=0}^N \xi_i(\theta) \sqrt{\lambda_i} \phi_i(t), \quad (14)$$

is the approximate representation for the process $f(t, \theta)$. The truncation error can be expressed as

$$e_N(t) = f(t, \theta) - \tilde{f}(t, \theta) = \sum_{j=N+1}^{\infty} \xi_j(\theta) \sqrt{\lambda_j} \phi_j(t) \quad (15)$$

and represents the remainder of the series after truncation. The mean square error due to truncation, for the entire time duration $[0, T]$ is given by

$$\begin{aligned} E \left[\left\{ \int_0^T e_N(t) dt \right\}^2 \right] &= \sum_{j=N+1}^{\infty} \sum_{k=N+1}^{\infty} E \left[\xi_j(\theta) \xi_k(\theta) \right] \\ &\quad \sqrt{\lambda_j} \sqrt{\lambda_k} \int_0^T \int_0^T \phi_j(t_1) \phi_k(t_2) dt_1 dt_2 \\ &= \sum_{j=N+1}^{\infty} E \left[\xi_j(\theta)^2 \right] \lambda_j \int_0^T \phi_j^2(t) dt \\ &= \sum_{j=N+1}^{\infty} \lambda_j. \end{aligned} \quad (16)$$

The simplifications in Eq. 16 follows from the property of the random variables $\xi(\theta)$ and Eq. 14. The result from Eq. 16 implies that the accuracy of the K-L approximation can be analyzed by studying the eigenvalues $\{\lambda_j\}$ obtained by solving Eq. 12. Typically, N is chosen such that the mean square error is less than a tolerance value ϵ_{tol} , usually taken to be of 10^{-3} or less.

For numerical calculations, the autocorrelation function $R_{ff}(t, s)$ is usually estimated from measurement data and are available only in discrete form as a matrix $\mathbf{R}$. It can be shown that the integral equation in Eq. 12 can be expressed in the matrix form as

$$\mathbf{R}\phi^T = \lambda\phi^T. \quad (17)$$

Thus, the integral equation in Eq. 12 is now expressed as an eigenvalue problem, where λ are the eigenvalues having the same meaning as in Eq. 12 and ϕ represents the eigenfunction in discrete form spanning the length $[0, T]$. It is obvious that the orthogonality conditions in Eq. 14 are satisfied.

The simulation of a random process using K-L expansion can now be summarized in the following algorithmic steps:

1. Generate the covariance matrix $\mathbf{R}$ from the autocovariance function of the random process. Let the dimension of $\mathbf{R}$ be $k \times k$
2. Solve the eigenvalue problem given by Eq. 17 and evaluate the k eigenvalues and eigenfunctions
3. Decide the number of terms N to be included in the series expansion based on the magnitude of eigenvalues, such that the mean square truncation error $\sum_{j=N+1}^k \lambda_j < \epsilon_{\text{tol}}$, where ϵ_{tol} is a specified tolerance value
4. Generate a set of standard normal random variables $\xi(\theta)$
5. Substitute the corresponding eigenvalues, eigenfunctions, and random variables in Eq. 15 to obtain realizations of the random process

It must be remarked here that there are alternative methods for simulating stationary Gaussian processes. Of these, the most commonly used

is the spectral representation proposed by (Shinozuka and Jan 1972). However, the advantage of using the K-L expansions for modeling the input excitations to nonlinear problems become evident when the response is obtained in terms of the basis vectors $\phi(t)$ that have been used for the K-L expansions. This is achieved by using polynomial chaos. This is discussed in the following section.

Polynomial Chaos Expansion

Since the KL expansion is obtained as a linear sum of Gaussian random variables, the series can be used to represent only Gaussian random processes. For spectral representation of non-Gaussian random processes, one uses the more general form known as the polynomial chaos expansion. Polynomial chaos expansion is a spectral representation of the random process in terms of orthonormal basis functions and deterministic coefficients. Based on the Cameron and Martin theorem (Cameron and Martin 1947), it can be shown that a zero mean, second order random process can be represented as

$$\begin{aligned} X(t, \theta) = & \hat{x}_0 \Gamma_0 + \sum_{i_1=1}^{\infty} \hat{x}_{i_1} \Gamma_1(\xi_{i_1}(\theta)) \\ & + \sum_{i_1=1}^{\infty} \sum_{i_2=2}^{\infty} \hat{x}_{i_1 i_2} \Gamma_2(\xi_{i_1}(\theta), \xi_{i_2}(\theta)) + \dots, \end{aligned} \quad (18)$$

where, $\{\Gamma_n(\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_n})\}$ are the orthogonal basis functions and $\{\hat{x}_0, \hat{x}_{i_1}, \hat{x}_{i_1 i_2}\}$ are the corresponding projections. Moreover, the orthogonal basis functions are assumed to be polynomial functions of standard Gaussian random variables $\xi = (\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_n})$. This is the reason for the terminology ‘‘polynomial chaos expansions.’’ Equation 18 is usually written in the more compact form

$$X(t, \theta) = \sum_{i=0}^{\infty} x_i(t) \Psi_i(\xi(\theta)), \quad (19)$$

where, Ψ_i denotes the basis functions and $x_i(t)$ are the deterministic projections along these basis functions. Since Ψ_i 's are the basis functions, they are mutually orthogonal and satisfy the following identity

$$\langle \Psi_i \Psi_j \rangle = \langle \Psi_i^2 \rangle \delta_{ij}, \quad (20)$$

where, δ_{ij} is the Kronecker delta function and $\langle \cdot \rangle$ is the expectation operator of the form

$$\langle x \rangle = \int_{\Omega} x w(x) dx. \quad (21)$$

Here, $w(x)$ is an appropriate weighting function and Ω is the integration domain. Truncating Eq. 19 to p terms, $X(t, \theta)$ can be approximated as

$$X(t, \theta) = \sum_{i=0}^p x_i(t) \Psi_i(\xi(\theta)). \quad (22)$$

The number of terms up to which the series is truncated, p , is known as the order of expansion. For n number of random variables and polynomial order n_p , p is given by the following:

$$p = \frac{(n + n_p)!}{n! n_p!} - 1. \quad (23)$$

The use of PCE to represent the solution of a nonlinear differential equation is discussed next.

Response Analysis Using Polynomial Chaos Expansion

Let us consider a differential equation of the form

$$\ddot{x} + f(x, \dot{x}) = g(t), \quad (24)$$

where, $g(t)$ is a Gaussian random process, expressed as a K-L expansion

$$g(t, \theta) = \sum_{i=0}^{\infty} \xi_i(\theta) \sqrt{\lambda_i} \phi_i(t). \quad (25)$$

Since the differential equation is nonlinear, it is clear that $x(t)$ is a non-Gaussian random process.

Now, let us assume that $x(t)$ can be represented as a PCE. Thus, Eq. 24 can now be written as

$$\begin{aligned} \sum_{i=0}^p \ddot{x}_i(t) \Psi_i(\xi) + f \left(\sum_{i=0}^p x_i(t) \Psi_i(\xi), \sum_{i=0}^p \dot{x}_i(t) \Psi_i(\xi) \right) \\ = \sum_{i=0}^M \xi_i \sqrt{\lambda_i} \phi_i(t), \end{aligned} \quad (26)$$

where, the basis functions $\Psi_i(\xi(\theta))$ are known but the coefficients $\{x_i\}$ are to be determined. Some of the techniques by which these deterministic coefficients can be determined are discussed in the subsequent sections.

Stochastic Galerkin Method

In Galerkin PC approach, the chaos expansion of the system response is substituted into the governing equations leading to a set of coupled equations in terms of the chaos coefficients

$$\begin{aligned} \langle \Psi_k^2 \rangle \ddot{x}_k(t) + \left\langle f \left(\sum_{i=0}^p x_i(t) \Psi_i, \sum_{i=0}^p \dot{x}_i(t) \Psi_i \right) \cdot \Psi_k \right\rangle \\ = \sum_{j=0}^M \langle \xi_j \cdot \Psi_k(\xi) \rangle \sqrt{\lambda_j} \phi_j(t), \end{aligned} \quad (27)$$

where, $k = 0, 1, 2, \dots, p$. These coupled deterministic equations typically have terms, such as, $E[\Psi_i(\xi) \Psi_j(\xi) \dots \Psi_k(\xi)]$ as coefficients. These coefficients represent multidimensional integrals of the form

$$E[\Psi_i(\xi) \Psi_j(\xi) \dots \Psi_k(\xi)] = \int \dots \int \psi_i \dots \psi_k p_{\xi}(\xi) d\xi \quad (28)$$

and need to be evaluated *a priori*.

For nonlinear dynamical systems, the form of these coupled deterministic equations are usually complicated and their solution can be tedious and time consuming. Depending on the nonlinearity, the number of terms in these expectations could be large. Evaluation of these terms imply multidimensional integration and need to be carried out prior to the solution of the coupled

deterministic equations. It must be noted that the application of the stochastic Galerkin method changes the form of the differential equations that need to be solved. Hence, the stochastic Galerkin method is referred to as an intrusive method. In order to circumvent these difficulties, several alternative methods are available in literature. Some of these are discussed in the subsequent sections.

Stochastic Collocation Method

The method of stochastic collocation is an efficient way that has been discussed in the literature (Ghanem and Spanos 1991; Desai and Sarkar 2010) for obtaining the PCE projections x_j . In contrast to the Galerkin method, the polynomial chaos expansions are not substituted in the governing equation and hence it is known as nonintrusive. Instead, here one uses the property that $\langle \Psi_i, \Psi_j \rangle = 0$ for $i \neq j$ for estimating the coefficients x_j . This involves multiplying both sides of Eq. 19 by Ψ_j and taking expectations. This leads to an estimate of $x_j(t)$ from the following formula

$$x_j(t) = \frac{\langle X(t, \xi(\theta)), \Psi_j \rangle}{\langle \Psi_j^2 \rangle}, \quad (29)$$

where, x_j is the PCE coefficient and the inner product in the numerator is given by the following multidimensional integral

$$\langle x(t, \xi(\theta)), \Psi_j \rangle = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} X(t, \xi(\theta)) \Psi_j w(\xi) d\xi. \quad (30)$$

The dimension of the integral is equal to the dimension of the vector ξ . More discussions on issues related to evaluation of integrals of the type shown in Eq. 30 are discussed in the next section.

Clearly, in the stochastic collocation technique, unlike in the Galerkin's method, one does not require transforming the original equations into any other form. Instead, the focus is on evaluating the multidimensional integrals. An inspection of Eq. 30 reveals that these integrals can be evaluated using suitable quadrature rules.

This implies selecting a set of collocation points along the domains and evaluating the structure response at these collocation points. As these methods do not alter the form of the governing equations, these methods are termed as nonintrusive methods. However, the price that one has to pay is that adopting collocation methods imply a solution of the equations of motion for various values of ξ corresponding to the collocation points. This increases the computational complexity of the problem, especially if the dimension of ξ is large. More discussions on these issues are presented in the next section.

Evaluating Multi-Dimensional Integrals

The successful implementation of PCE involves developing computationally efficient methods for the evaluation of multidimensional integrals such as Eq. 30. The multidimensional integrals are of the form

$$\langle x(t, \xi(\theta)), \Psi_j \rangle = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} X(t, \xi(\theta)) \Psi_j w(\xi) d\xi, \quad (31)$$

where, $w(\xi)$ is an appropriate weighting function. The dimension of the integral depends on the dimension of the vector ξ which again depends on the spectral content of the excitation process. Typically, in most engineering systems, the characterization of the input excitations require at least dimensions greater than 2. As in most cases closed form analytical expressions for these integrals are not possible to obtain, one has to resort to numerical methods. Therefore, one has to look for techniques to approximate these integrals with the expenditure of least possible computational effort. Various techniques for the same are available in the literature (Evans and Swartz 2000), a few of which are discussed here.

Let us consider the problem of evaluating a multidimensional integral of the form

$$\mathcal{I} = \int_{\Omega} f(\mathbf{x}) d\mathbf{x}, \quad (32)$$

where, the dimension of the integral is equal to the dimension of the vector $\mathbf{x}$ and $f(\cdot)$ is a function, which in general, can be treated to be nonlinear. It is assumed that analytical expressions for the integral in Eq. 32 are not available. Therefore, a suitable approximation for the above integral has to be developed.

Product Grid Methods

A tensor product of one-dimensional quadrature point set is used as the collocation point set in a product grid formula. Let us consider a one-dimensional quadrature rule

$$U^i(f) = \sum_{j=1}^{n_i} f(X_j^i) \cdot w_j^i \quad (33)$$

where, i denotes the dimension which in this case is $1X_j^i$ is the collocation point j in dimension i , w_j^i is the weight, and n_i is the number of collocation points in dimension i . In the multivariate case, a tensor product of such one-dimensional point set can be constructed as

$$(U^{i_1} \otimes \dots \otimes U^{i_d})f = \sum_{j_1=1}^{n_{i_1}} \dots \sum_{j_d=1}^{n_{i_d}} f(X_{j_1}^{i_1}, \dots, X_{j_d}^{i_d}) \cdot (w_{j_1}^{i_1} \otimes \dots \otimes w_{j_d}^{i_d}). \quad (34)$$

Equation 34 needs N_c collocation points. If $n_1 + \dots + n_d = n$, then $N_c = n^d$. So, for high dimensions ($d > 3$), the product grid method cannot be implemented efficiently. However, tensor product grids may be applied for problems involving lower dimensions. The collocation points and the corresponding weights can be computed using a variety of quadrature rules such as the Gauss-Legendre quadrature, Gauss-Hermite quadrature, Clenshaw-Curtis quadrature, and Gauss-Patterson quadrature.

Gauss-Legendre Quadrature

Gauss-Legendre quadrature, often simply referred to as Gauss quadrature is the most widely used method to numerically compute integrals over the domain $[-1, 1]$. The nodes are obtained

by evaluating the zeros of a Legendre polynomial of required order. If n is the order of the polynomial, then the Legendre polynomial $P_n(x)$ can be computed using the following recurrence relation:

$$\begin{aligned} (n+1)P_{n+1}(x) &= (2n+1)xP_n(x) - nP_{n-1}(x), \\ P_0(x) &= 1, P_1(x) = x. \end{aligned} \quad (35)$$

The nodes x_i obtained by computing the zeros of a n^{th} order polynomial are then used to compute the weights w_i as shown below:

$$w_i = \frac{2}{(1 - x_i^2) [P'_n(x_i)]^2}. \quad (36)$$

The nodes and weights thus computed are substituted in Eq. 33 to obtain the value of the integral. For the multidimensional case, Eq. 34 is used.

Gauss-Hermite Quadrature

When the integrand takes the form $f(x)e^{-x^2}$ and the domain of integration is $[-\infty, \infty]$, Gauss-Hermite quadrature is best suited. In this variant of Gauss quadrature, the nodes x_i are the roots of a n^{th} order Hermite polynomial and the weights are the Gaussian probability density function evaluated at these nodes. A recurrence relation for generating Hermite polynomials is given as

$$\begin{aligned} H_{n+1}(x) &= xH_n(x) - nH_{n-1}(x), \\ H_0(x) &= 1, H_1(x) = x. \end{aligned} \quad (37)$$

The above two Gauss quadrature rules are non-nested in nature. More discussions on nested quadrature rules will be discussed later.

Monte Carlo Simulations

As it was seen in the previous section, numerical evaluation of multidimensional integrals using quadrature rules is not computationally feasible for high dimensions. In such situations, Monte Carlo simulations(MCS) provide a method for

approximating the integral $\mathcal{I}$. This involves rewriting Eq. 32 as

$$\mathcal{I} = \int_{\Omega} \frac{f(\mathbf{x})}{p(\mathbf{x})} p(\mathbf{x}) \, d\mathbf{x}, \quad (38)$$

where, $p(\mathbf{x})$ can be assumed to be a probability density function (pdf), such that, $p(\mathbf{x}) \neq 0$ within the domain of integration, Ω . Equation 38 can be further rewritten as

$$\begin{aligned} \mathcal{I} &= \int_{-\infty}^{\infty} I[\mathbf{x} \in \Omega] \frac{f(\mathbf{x})}{p(\mathbf{x})} p(\mathbf{x}) \, d\mathbf{x}, \\ &= E \left[I[\mathbf{x} \in \Omega] \frac{f(\mathbf{x})}{p(\mathbf{x})} \right], \end{aligned} \quad (39)$$

where, $E[\cdot]$ is the expectation operator and $I[\cdot]$ is an indicator function that takes values of unity if $\mathbf{x}$ lies within the domain Ω and zero otherwise. Thus, an approximation for the expectation, or the mean, can be obtained as

$$\mathcal{I} = \sum_{j=1}^{N_m} I[\mathbf{x}_j \in \Omega] \frac{f(\mathbf{x}_j)}{p(\mathbf{x}_j)}, \quad (40)$$

where, N_m is the sample size. Here, $\mathbf{x}_j$ are simulated according to the pdf $p(\mathbf{x})$. The steps involved in MCS are:

1. Create a parametric model of the form $y = f(x_1, \dots, x_n)$
2. Generate samples of the parameters $x_{1_i}, \dots, x_{n_i}$, from a known probability distribution, $P(x)$
3. Evaluate the function f_i at each sample point
4. Compute the required statistical quantities like mean, variance, and higher order moments

The Monte Carlo simulation is easy to implement but suffers from a relatively lower convergence rate of $O(1/\sqrt{N_m})$.

Sparse Grid Collocation

The sparse grid method is one of the few methods to tackle the problem of the curse of dimensionality. This method was first introduced by

Smolyak and is based on constructing a multidimensional multilevel basis by truncation of the tensor product expansion of a one-dimensional basis. The number of function evaluations using a sparse grid is of $O(N(\log N)^{d-1})$ which is much lesser than that of a full tensor product which is of $O(N^d)$, where N is the number of collocation points and d is the dimension of the integral. Smolyak's algorithm finds application in a variety of high-dimensional problems like numerical integration (Gerstner and Griebel 1998; Barthelmann et al. 2000) and stochastic partial differential equations (Babuska et al. 2007; Xiu and Hesthaven 2005).

Smolyak's Algorithm

The Smolyak's quadrature rule enables creating a grid of collocation points in a multi-dimensional space with a minimal number of points. Let $f(x)$ be the function to be integrated over the d -dimensional domain Ω . Let the smooth function $f(x)$ be defined in $[0, 1]^d \rightarrow \mathfrak{R}$. For 1-dimensional case, i.e., when $d = 1$, the smooth function $f(x)$ can be approximated using the interpolation formula

$$\mathcal{U}^i(f) = \sum_{j=1}^{m_i} f(X_j^i) \cdot a_j^i, \quad (41)$$

where $\mathcal{H}^i = \{X_j^i\}$ denotes the set of nodal points defined in $[0, 1]$ and j takes values from 1 to m_i , i is an integer number and a_j^i are interpolating nodal basis functions. Note that $\mathcal{H}^i$ consists of m_i number of points.

It is next assumed that a sequence of formula corresponding to Eq. 41 can be generated for different values of i . In the multivariate case, i.e., when $d > 1$, the tensor product formula is given by

$$\begin{aligned} (\mathcal{U}^{i_1} \otimes \dots \otimes \mathcal{U}^{i_d})f &= \sum_{j_1=1}^{m_{i_1}} \dots \sum_{j_d=1}^{m_{i_d}} f(X_{j_1}^{i_1}, \dots, X_{j_d}^{i_d}) \\ &\cdot (a_{j_1}^{i_1} \otimes \dots \otimes a_{j_d}^{i_d}). \end{aligned} \quad (42)$$

This serves as the building block for the Smolyak's algorithm. A sparse interpolant $A_{q,d}$ is constructed

next using the Smolyak's algorithm using products of one-dimensional functions, given by

$$\mathcal{A}_{q,d}(f) = \sum_{q-d+1 \leq |\mathbf{i}| \leq q} (-1)^{q-|\mathbf{i}|} \cdot \binom{N-1}{q-|\mathbf{i}|} \cdot (\mathcal{U}^{i_1} \otimes \cdots \otimes \mathcal{U}^{i_d}), \quad (43)$$

where, $q \geq d$, $A_{d-1,d} = 0$, $\mathbf{i}$ denotes the multi-index $(i_1, \dots, i_d)$ and $|\mathbf{i}| = i_1 + \dots + i_d$. The index $i_k, k = 1, \dots, d$, denotes the level of interpolation along the k^{th} direction. Equation 43 implies that the interpolant is built up of one-dimensional functions of order i_k , such that the sum total $(|\mathbf{i}| = i_1 + \dots + i_k)$ across all dimensions is between $q - d + 1$ and q . Equation 43 can be rewritten in terms of incremental interpolants, Δ^i , given by

$$\Delta^i = \mathcal{U}^i - \mathcal{U}^{i-1} \quad (44)$$

with $\mathcal{U}^0 = 0$. Then, Smolyak's formula for computing a d -dimensional integral is

$$\mathcal{A}_{q,d}f = \sum_{i=1}^{q-1} (\Delta_i \otimes A_{q-1,d-1})f. \quad (45)$$

The interpolation, $\mathcal{A}_{q,d}f$, can be approximated by evaluating the function at specific nodes given by the sparse grid. A sparse grid $\mathcal{H}_{q,d}^i$ can be constructed using the points of the multivariate Smolyak's formula. If $\mathbf{X}^i$ is the set of points $\{X_1^i, \dots, X_{m_i}^i\}$ corresponding to $\mathcal{U}$, then

$$\mathcal{H}_{q,d} = \bigcup_{|\mathbf{i}| \leq q+d-1} \mathbf{X}_{i_1} \times \cdots \times \mathbf{X}_{i_d}, \quad (46)$$

where, $|\mathbf{i}| = i_1 + \dots + i_d$.

The construction of the interpolant using Smolyak's algorithm enables improving the level of approximation which uses all the previously generated interpolant functions. Thus, to extend the interpolation from level $i - 1$ to i , the function needs to be additionally evaluated at only the set of grid points $X_{\Delta}^i = X^i \setminus X^{i-1}$, where the bar indicates set difference. The set of grid

points at which the function needs to be evaluated for improving the interpolation from $(q - 1)$ order to q order is given by

$$\Delta \mathcal{H}_{q,d} = \bigcup_{|\mathbf{i}|=q} (X_{\Delta}^{i_1} \otimes \cdots \otimes X_{\Delta}^{i_d}), \quad (47)$$

where $i_1, \dots, i_d$ have the same meaning as before. The steps involved in generating a sparse grid and then utilizing it for computing a multidimensional integral are given below:

1. Choose a suitable one-dimensional quadrature with nodes X_i and weights a_i
2. Based on the number of dimensions d of the problem, decide the level of sparse grid q
3. Construct a sparse grid, $\mathcal{H}_{q,d}^i$, by evaluating Eq. 46
4. Define the difference quadrature formula $\Delta^i f$ as given in Eq. 44. Here f is the function which is to be integrated
5. Using Eq. 45, evaluate the integral of the function $f(\mathbf{x})$ over the domain Ω

The size of the sparse grid depends on the following factors:

- The level, q , of approximation adopted for the function
- The dimension of integration d
- The type of one-dimensional quadrature formula being used

Typically, the quadrature formulae that can be used are either of the non-nested type such as the Gauss-Legendre and Gauss-Hermite quadrature rules or nested type. A brief discussion on a couple of nested quadrature rules are presented next.

Nested Univariate Quadrature Rules

Some of the quadrature schemes have sets of nodal points which are nested. This means the set of nodal points at $(i + 1)^{th}$ level, X_{i+1} , includes all points in the set X_i . Symbolically this is written as $X_i \subset X_{i+1}$. Therefore, while moving to a higher level, only the points in the difference set, $X_{i+1} - X_i$, need to be considered for function

evaluation, thus saving a huge amount of computational effort (Barthelmann et al. 2000). The use of such nested quadrature schemes result in grids which are sparse and require fewer collocation points while evaluating multidimensional integrals. Specific forms of such quadrature schemes are discussed next.

Clenshaw-Curtis Quadrature

The Clenshaw-Curtis quadrature rule is widely used for generating sparse grids. It uses the roots of Chebyshev polynomials as the nodes. For any choice of $m_i > 1$, the nodes are given by

$$x_j^i = -\cos \frac{\pi \cdot (j-1)}{m_i-1}, \quad j = 1, \dots, m_i. \quad (48)$$

Also, when $m_i = 1$, $x_1^i = 0$. A nested set of points can be obtained by setting the following relation for m_i

$$m_i = 2^{i-1} + 1, \quad i > 1. \quad (49)$$

m_1 must be set to 1 to have the fewest number of points for a large d . Since m_1 is odd for any $i > 1$, the corresponding weights are given by

$$w_j^i = w_{m_i+1-j}^i = \frac{2}{m_i-1} \left(1 - \frac{\cos(\pi \cdot (j-1))}{m_i \cdot (m_i-2)} - 2 \sum_{k=1}^{\frac{(m_i-3)}{2}} \frac{1}{4k^2-1} \cdot \cos \frac{2\pi k \cdot (j-1)}{m_i-1} \right), \quad (50)$$

for $j = 2, \dots, m_i - 1$. The two missing weights are given by

$$w_1^i = w_{m_i}^i = \frac{1}{m_i \cdot (m_i-2)}, \quad (51)$$

where, m_i are given by Eq. 49.

Gauss-Patterson Quadrature

The Gauss-Patterson quadrature technique is an extension of the Gauss quadrature. In a regular Gauss quadrature, the set of nodes for a $(n+1)$ point quadrature is completely independent of the set of nodes for n point quadrature. This means that with an increase in n , the function evaluations already performed for a lower n value cannot be reused. Kronrod (Gerstner and Griebel 1998) proposed an optimal method by which all the nodes of lower order can be reused. This is carried out by augmenting a new set of nodes to the old set of nodes while moving from a lower level quadrature to one of a higher level. These new nodes are generally obtained as the roots of Stieltjes polynomials. Patterson (Gerstner and Griebel 1998) extended Kronrod's construction by developing an iterated scheme to find a nested set of nodes.

Product Grids

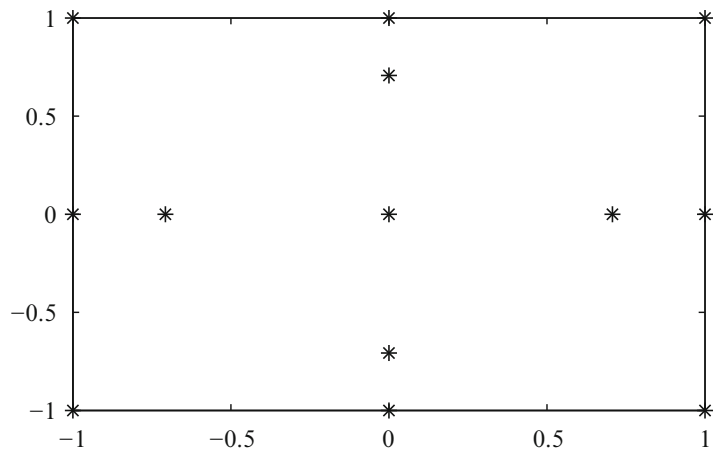
To illustrate the effect that the type of quadrature rule has on the sparse grid size, a series of two dimensional grids have been shown in Figs. 2, 3, 4, and 5. They are compared with the full tensor product grid. As can be observed, a considerable reduction in the number of nodal points between the full tensor grid and the sparse grid is seen in all the cases. For example, for a Clenshaw-Curtis grid with $d=2$ and $q=3$, the sparse grid requires 13 function evaluations while the full grid requires 25 function evaluations. The reduction in computations is therefore about 48 %. In the following section, the accuracy of this algorithm is examined with respect to a test function.

Accuracy of Sparse Grid Collocation

In this section, the accuracy of the sparse grid collocation technique for computing numerically multidimensional integrals is examined. A test function is examined whose exact numerical integral is available. For the one-dimensional quadrature rules used in the Smolyak's algorithm, only the nested quadrature rules, namely, the Clenshaw-Curtis(CC) rule and the Gauss-Patterson(GP) rule, have been used.

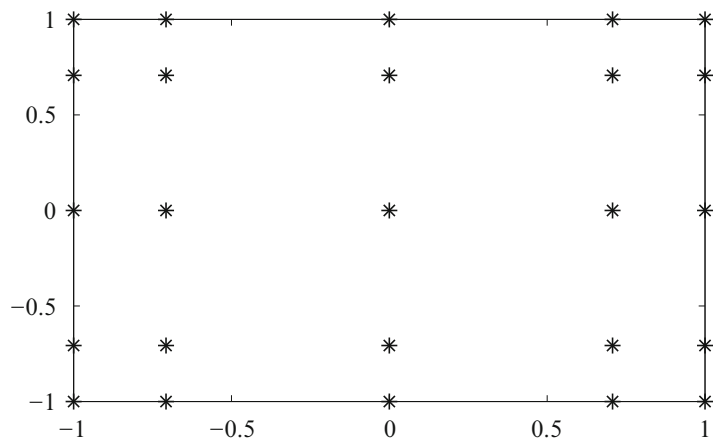
**Reliability Analysis of
Nonlinear Vibrating
Systems-Spectral
Approach,**

Fig. 2 Clenshaw-Curtis
sparse grid with
 $d = 2, q = 3$

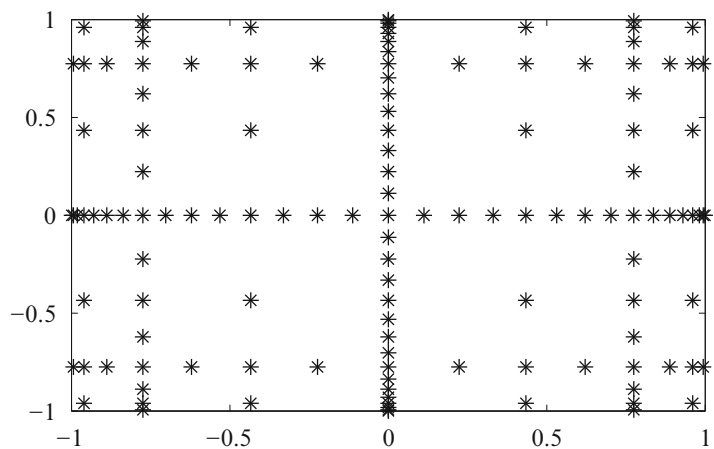


**Reliability Analysis of
Nonlinear Vibrating
Systems-Spectral
Approach,**

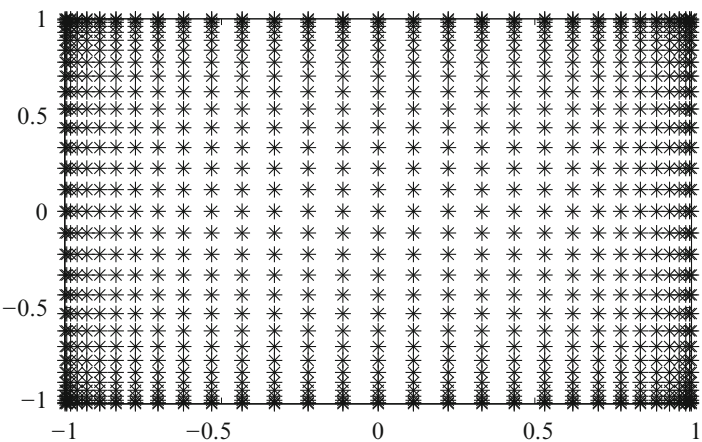
Fig. 3 Clenshaw-Curtis
full grid with $d = 2, q = 3$



**Reliability Analysis of
Nonlinear Vibrating
Systems-Spectral
Approach, Fig. 4** Gauss-
Patterson sparse grid with
 $d = 2, q = 5$



Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach, Fig. 5 Gauss-Patterson full grid with $d = 2, q = 5$



Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach, Table 1 Computational results of test function 1; CC Clenshaw-Curtis, GP Gauss-Patterson

Dimension	Level	Function calls CC	Function calls GP	Error CC	Error GP
2	1	5	17	-9.47×10^{-02}	-4.50×10^{-03}
2	5	65	129	-2.73×10^{-04}	-1.70×10^{-06}
2	7	321	769	$+5.47 \times 10^{-06}$	$+8.23 \times 10^{-09}$
4	3	41	49	-6.46×10^{-02}	-5.43×10^{-04}
4	5	401	769	$+1.60 \times 10^{-03}$	$+1.57 \times 10^{-05}$
4	7	2,929	7,937	-7.47×10^{-04}	$+2.84 \times 10^{-08}$
6	3	85	97	$+2.59 \times 10^{-01}$	-1.00×10^{-03}
6	5	1,457	2,561	-3.25×10^{-02}	$+2.85 \times 10^{-05}$
6	7	15,121	40,193	$+1.90 \times 10^{-03}$	$+3.51 \times 10^{-08}$

The numerical evaluation of the following integral is considered

$$\mathcal{I} = \int_{[0,1]^d} (1 + 1/d)^d \prod_{i=1}^d (x_i)^{(1/d)} \mathrm{d}\mathbf{x}, \quad (52)$$

where d denoted the dimension of the integral. Note that x is a d -dimensional vector. The exact value of the integral for any value of d is unity. A numerical estimate of the above integral is computed using both CC and GP quadrature rules. For a given dimension d , several levels of approximation q were considered and a comparison of the accuracy of the estimates are examined. The dimension d is varied from 2 to 6. Table 1 gives details of the number of function calls and the error incurred using the two methods. It is observed that for a given d , as level q is increased, the error in the approximation reduces. The error when GP quadrature rule is used is less than when CC is used.

However, as the level increases, the number of collocation points required in GP is usually more than twice than that when CC quadrature rule is used. For higher dimensions, GP appears to be more accurate than CC. For example, when $d = 6$, using GP with level 3 requires 97 function calls with accuracy of $O(10^{-3})$ while the same accuracy is obtained using CC with level 7 which requires 15121 function evaluations. Clearly, using GP with lower levels of approximation is better than using CC with higher levels of approximation. Thus, sparse grids based on GP rule will be used in all further studies carried out here.

Numerical Examples

The developed algorithm is used to obtain the response of a few randomly excited dynamical systems. A Duffing oscillator is considered whose

response has been obtained through the stochastic collocation and the sparse grid collocation methods. The Galerkin's method is avoided due to difficulties mentioned in section [Stochastic Galerkin Method](#). Next, a two-dimensional airfoil subjected to unsteady gusty wind flow regime is considered. The response is obtained through the sparse grid collocation method. Once the response of the vibrating systems is obtained in terms of a polynomial chaos expansion, approximations for the joint pdf of the response and its instantaneous time derivative are obtained using simulations. Subsequently, the crossing statistics are calculated using Rice's formula. Finally, estimates of the first passage failure probability and expected fatigue damage are calculated using the method discussed in sections [First Passage Failures](#) and [Random Fatigue Damage](#).

Duffing Oscillator

The problem of estimating the response of a randomly excited Duffing oscillator is considered. The governing equation of motion of a Duffing oscillator is given by

$$m\ddot{X} + c\dot{X} + kX + k_1X^3 = f(t, \theta), \quad (53)$$

where, m is the mass, c is damping, k is the linear stiffness, and k_1 is the nonlinear stiffness. When k_1 is small, the effect of the nonlinearity is not significant. Thus, in the numerical example, k_1 is assumed to be (a) 2 times k and (b) 10 times k . The numerical values considered are $m = 5$ kg, $c = 1.414$ Ns/m, $k = 1,000$ N/m, and $k_1 = 2 \times 10^3$ N/m³ and 1×10^4 N/m³.

Full Grid Stochastic Collocation

The excitation, $f(t, \theta)$, assumed to be a stationary Gaussian process, is approximated as a K-L expansion truncated up to M terms as shown below:

$$f(t, \theta) = f_0 + \sum_{i=1}^M \xi_i(\theta) \sqrt{\lambda_i} \phi_i(t). \quad (54)$$

Here, θ represents the stochastic dimension. The autocorrelation function of the excitation is given by

Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach, Table 2 Magnitude of eigenvalues for the K-L representation of the force on the linear single degree of freedom oscillator

Eigenvalue	Magnitude
λ_1	1920.10
λ_2	1213.80
λ_3	575.91
λ_4	210.59
λ_5	61.30
λ_6	14.68
λ_7	2.97
λ_8	0.52
λ_9	0.08
λ_{10}	0.01

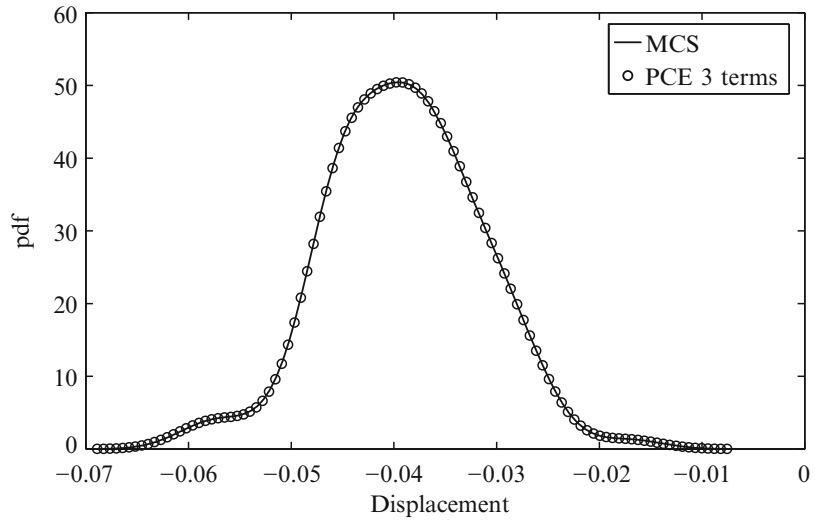
$$R_{ff}(\tau) = \sigma_f^2 e^{-c_0 \tau^2}, \quad (55)$$

where, $\sigma_f = 1$ and $c_0 = 1 \times 10^{-3}$. Table 2 lists the 10 highest eigenvalues of the spectral representation. First, the coefficients of the PCE representation are obtained using the method of stochastic collocation using a full tensor grid. As the number of collocation points becomes very large when a full tensor grid is used, the number of terms included in the K-L representation is limited to 2. A fifth order Gauss-Hermite quadrature is used. This results in 25 collocation points implying 25 deterministic runs of the governing equations of motion to estimate the PCE coefficients. A 3-term PCE representation is obtained for the response when $k_1 = 2 \times 10^3$ N/m³. These results are compared with those obtained from full-scale Monte Carlo simulations with 1×10^4 samples. Figures 6 and 7 respectively compare the pdf of the displacement and the velocity. While the pdf of the displacement appears to have a very close resemblance, a slight discrepancy is observed in the pdf of the velocity.

However, as the nonlinearity is increased, a 3-term PCE representation is found to be inadequate. When k_1 is increased to 1×10^4 N/m³, a 5-term PCE representation of the response is required. Comparisons between the mean of the displacement and the velocity are shown in Figs. 8 and 9, while the corresponding variance plots are shown in Figs. 10 and 11. It can be observed that while the mean have a good

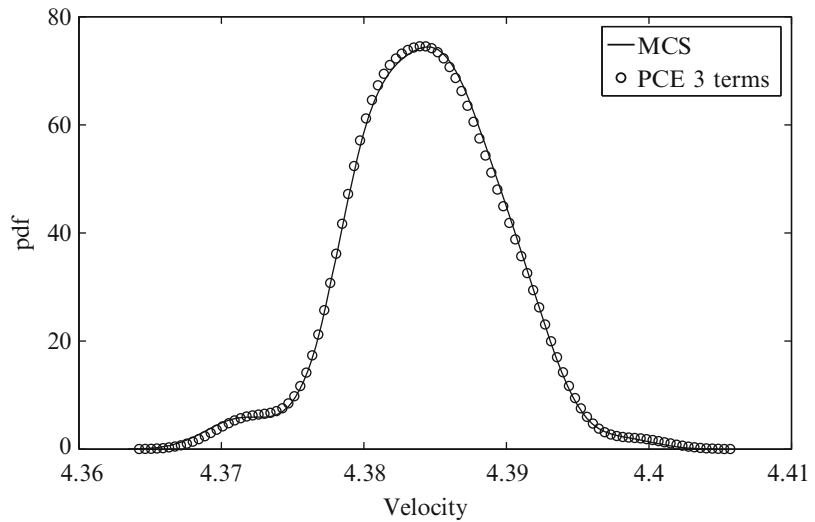
Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach,

Fig. 6 Marginal pdf $p_x(x)$; $k_1 = 2 \times 10^3 \text{ N/m}^3$



Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach,

Fig. 7 Marginal pdf $p_{\dot{x}}(\dot{x})$; $k_1 = 2 \times 10^3 \text{ M/m}^3$



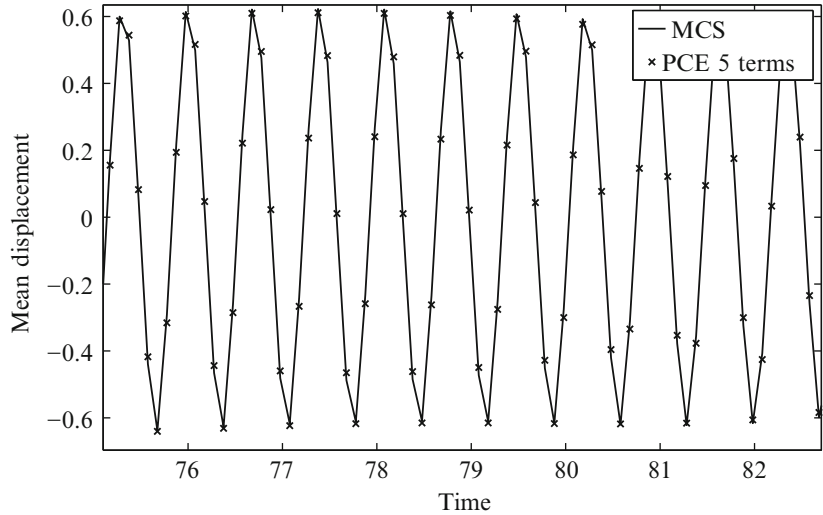
match, the variance plots reveal a slight mismatch. This implies that probably a 5-term PCE representation is not adequate and that higher order PCE representations are necessary. This becomes clear when one compares the marginal pdfs of the displacement and velocity of the Duffing oscillator shown in Figs. 11, 12, and 13. A significant mismatch is observed between the pdfs obtained from the 5-term PCE representation and Monte Carlo simulations. However, the CPU time required in the proposed method, which depends on the number of solutions of the governing equations is significantly smaller.

For a 5-term PCE representation, this required 25 solutions of the governing equation. In comparison, governing equation was solved 1×10^4 times when Monte Carlo simulations were used.

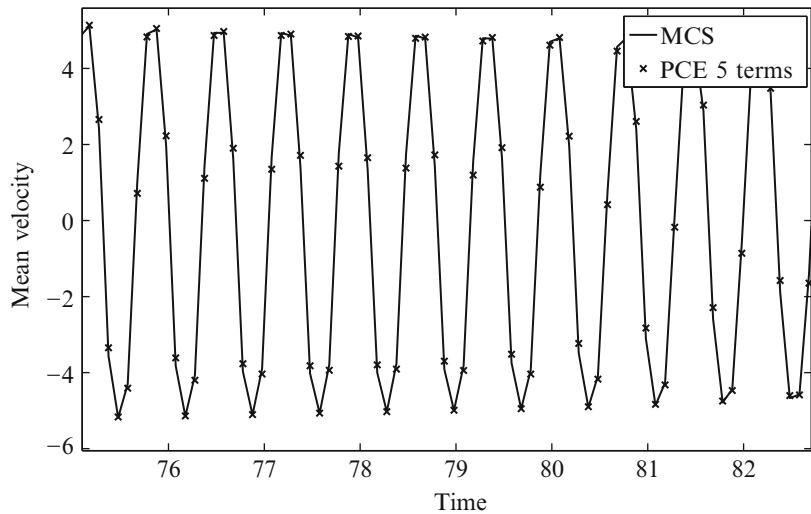
Sparse Grid Stochastic Collocation

Next, the sparse grid collocation method discussed in section [Sparse Grid Collocation](#) is used to obtain the PCE representation of the response. The use of sparse grid algorithm enables considering higher number of terms in the K-L representation of the excitation. Thus, now a 5-term K-L expansion is considered.

Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach, Fig. 8 Mean displacement; $k_1 = 1 \times 10^4 \text{ N/m}^3$



Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach, Fig. 9 Mean velocity; $k_1 = 1 \times 10^4 \text{ N/m}^3$

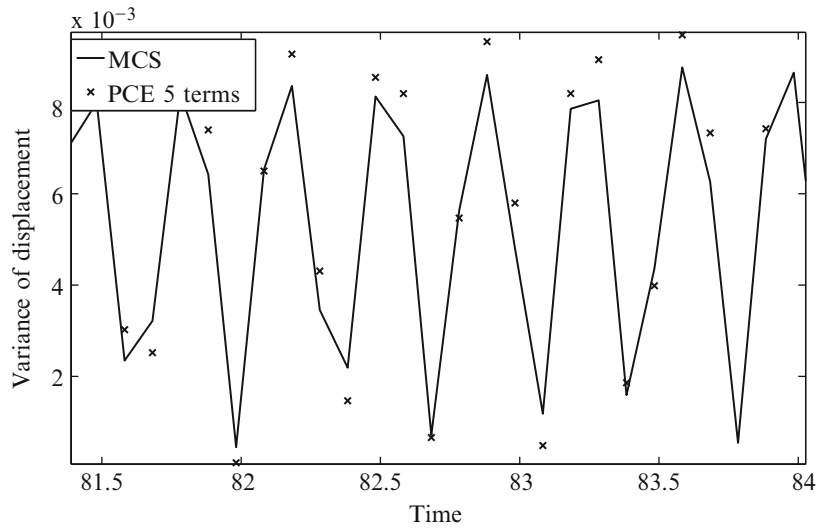


The sparse grid collocation method is now used to obtain the PCE coefficients of the response. A Gauss-Patterson based sparse grid collocation of level 6 is used for estimating the PCE coefficients. This involves a total 5,503 collocation points implying the need to solve an equivalent number of deterministic runs of the governing equation. This is still half the number of times the governing equation that needs to be solved using Monte Carlo simulations. The marginal pdfs of the displacement and velocity are shown in Figs. 14 and 15 respectively. A fairly good resemblance is observed. Figure 16 compares

the contour plots of the joint pdf $p_{X\dot{X}}(x, \dot{x})$ obtained using the proposed method and MCS. The mean crossing statistics are then computed using these joint pdf by numerically integrating Rice's formula. The mean crossing intensities calculated for various levels using the approximated joint pdf $p_{X\dot{X}}(x, \dot{x})$ obtained using the PCE-based method and MCS are shown in Fig. 17. Though the trends are observed to be very similar, the desired match is not obtained at the mean region. The first passage failure probabilities for various threshold levels are next calculated and are shown in Fig. 18. A very good

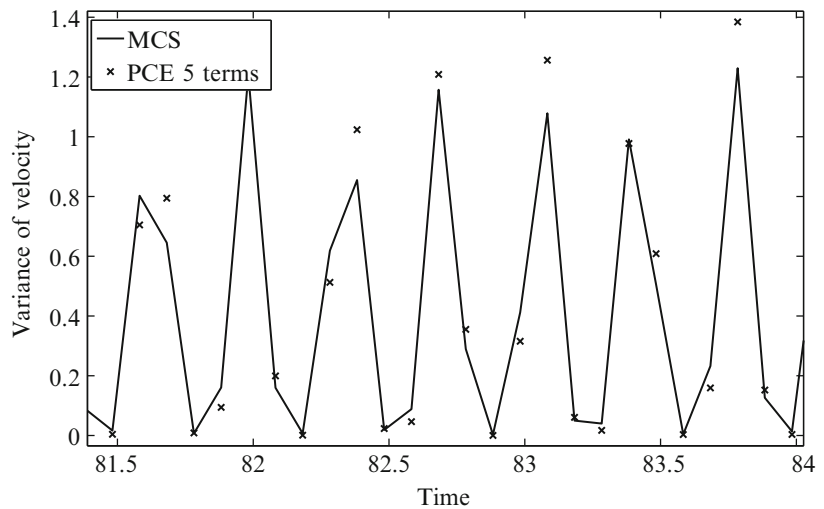
Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach,

Fig. 10 Variance of displacement; $k_1 = 1 \times 10^4$



Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach,

Fig. 11 Variance of velocity; $k_1 = 1 \times 10^4 \text{ N/m}^3$



match, when plotted in the log scales, is observed. This indicates that the order of error is negligible in comparison to the failure probability levels being estimated.

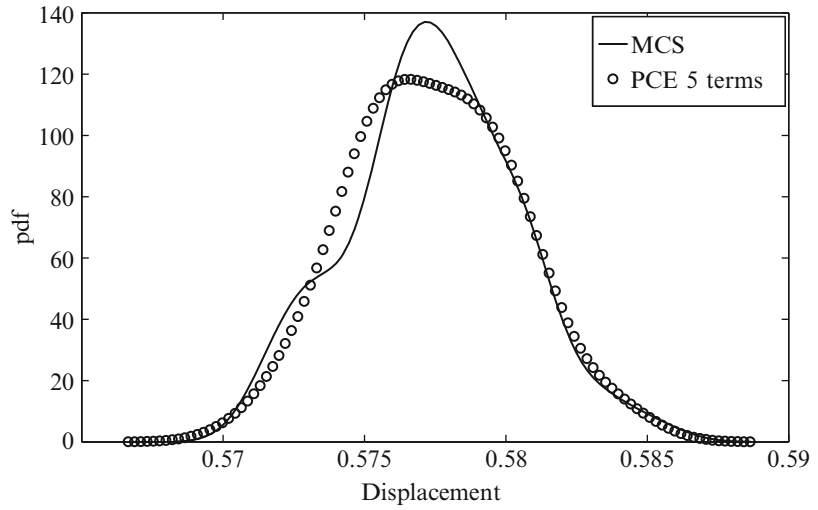
Next, estimates of the expected fatigue damage are obtained using the spectral method discussed in section [Random Fatigue Damage](#). These predictions are compared with those obtained from MCS using the procedure outlined in the previous section. For the sake of illustration, the numerical values for K and β are taken arbitrarily. The expected fatigue damage is

calculated for various values of β ranging from 2 to 8. These are shown in Fig. 19. It is observed that a very good match is observed with the proposed method and Monte Carlo simulations.

It is seen from the above results that the stochastic collocation method and the sparse grid method perform well in obtaining the response of randomly excited vibrating system. The sparse grid approach is next applied to a highly nonlinear fluid–structure interaction problem for response analysis and subsequently, reliability estimation.

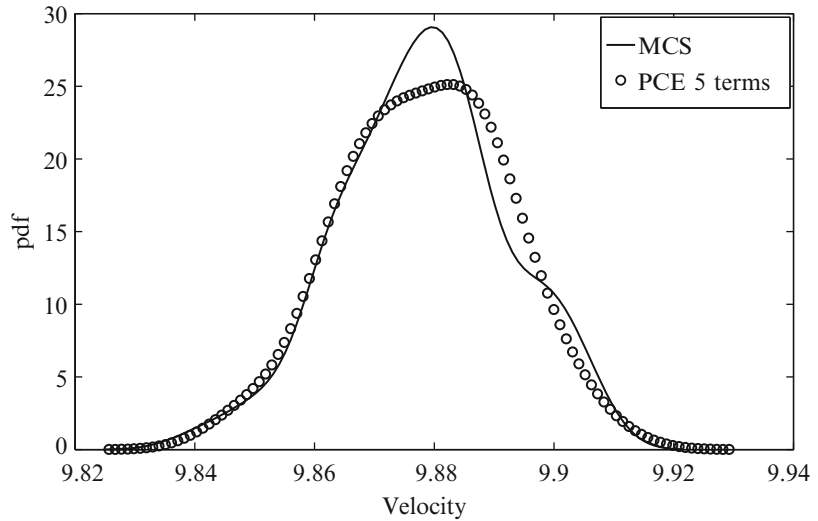
Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach,

Fig. 12 Marginal pdg
 $p_X(x); k_1 = 1 \times 10^4 \text{ N/m}^3$



Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach,

Fig. 13 Marginal pdf
 $p_X(\dot{x}); k_1 = 1 \times 10^4 \text{ N/m}^3$



2-D Airfoil Subjected to Unsteady Gusty Wind Flow Regime

Next, the oscillations of a two-dimensional airfoil in unsteady random flow are considered. The airfoil system is a nonlinear aeroelastic system with two degrees of freedom in the pitch and heave directions. The schematic diagram of the two-dimensional airfoil is shown in Fig. 20.

The governing equations of motion for the airfoil can be expressed as (Fung 1955)

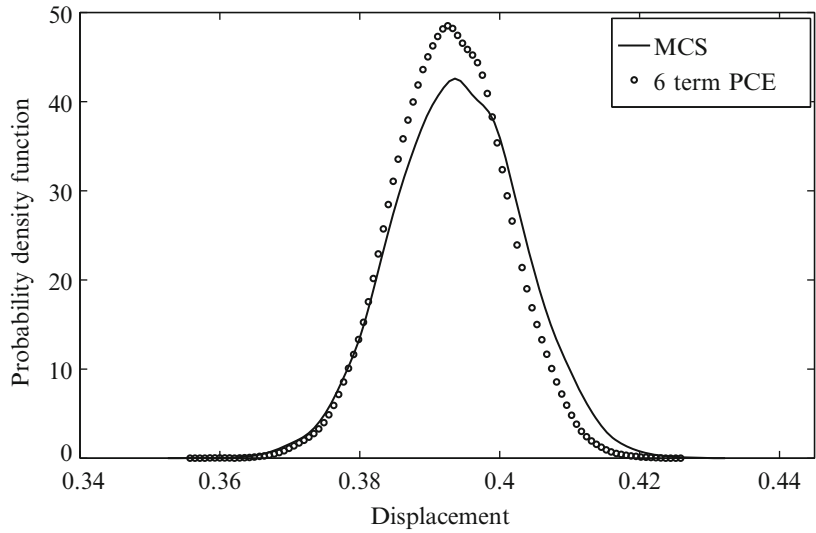
$$\epsilon'' + x_\alpha \alpha'' + 2\zeta_\epsilon \frac{\bar{\omega}}{U} \epsilon' + \left(\frac{\bar{\omega}}{U}\right)^2 (\epsilon + \beta_\epsilon \epsilon^3) = -\frac{1}{\pi\mu} C_L(\tau), \quad (56)$$

$$\frac{x_\alpha}{r_\alpha^2} \epsilon'' + \alpha'' + 2\frac{\zeta}{U} \alpha' + \frac{1}{U^2} (\alpha + \beta_\alpha \alpha^3) = \frac{2}{\pi\mu r_\alpha^2} C_M(\tau). \quad (57)$$

Here, $\epsilon = h/b$ is the nondimensional heave displacement, α is the pitch angle, m is the total mass per unit span, r_α is the radius of gyration about the elastic axis, ζ_ϵ and ζ_α are the damping ratios in plunge and pitch respectively, β_ϵ is the heaving stiffness co-efficient, β_α is the pitching stiffness, $a_h b$ denotes the distance of the elastic axis from the mid chord, and $x_\alpha b$ is the distance of the center of mass from the elastic axis; see Fig. 20

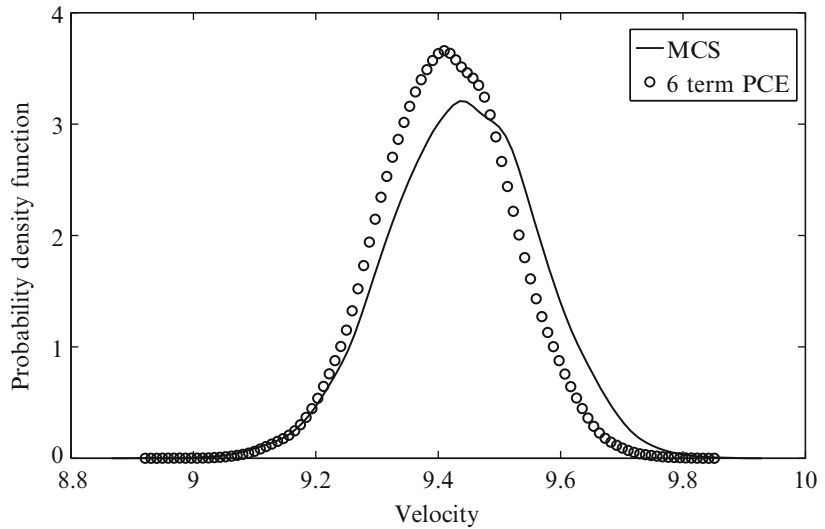
**Reliability Analysis of
Nonlinear Vibrating
Systems-Spectral
Approach,**

Fig. 14 Marginal pdf
 $p_X(x)$; 5 term K-L
expansion



**Reliability Analysis of
Nonlinear Vibrating
Systems-Spectral
Approach,**

Fig. 15 Marginal pdf
 $p_{\dot{X}}(\dot{x})$; 5 term K-L
expansion



R

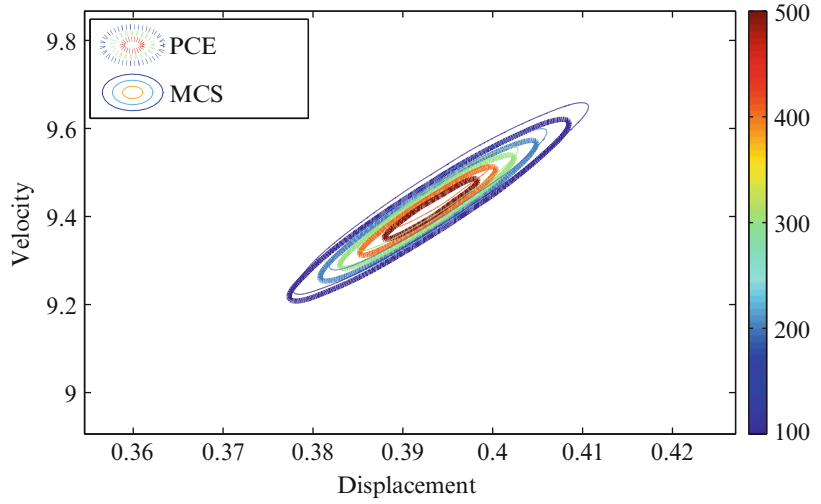
for a schematic. U is the nondimensional stream velocity given by $U = v/(b\omega_\alpha)$, $\overline{\omega} = (\omega_\epsilon/\omega_\alpha)$, where, ω_ϵ and ω_α are respectively the natural frequencies of the uncoupled plunging and pitching modes and $\tau = vt/b$ is the nondimensional time. The nonhomogeneous terms $C_L(\tau)$ and $C_M(\tau)$ represent the forcing terms and are usually represented as a set of coupled second order differential equations which are functions of α and ϵ (Alighanbari and Price 1996). Thus, these equations constitute a problem in fluid–structure interactions. The exponent $(\cdot)'$ represents differentiation with

respect to nondimensional time. Here, the coefficients and β_ϵ and β_α could be nonlinear functions of ϵ and α . If the flow is assumed to be incompressible and inviscid, the lift and pitching moment coefficients $C_L(\tau)$ and $C_M(\tau)$ are as given by Fung (1955) and are expressed as

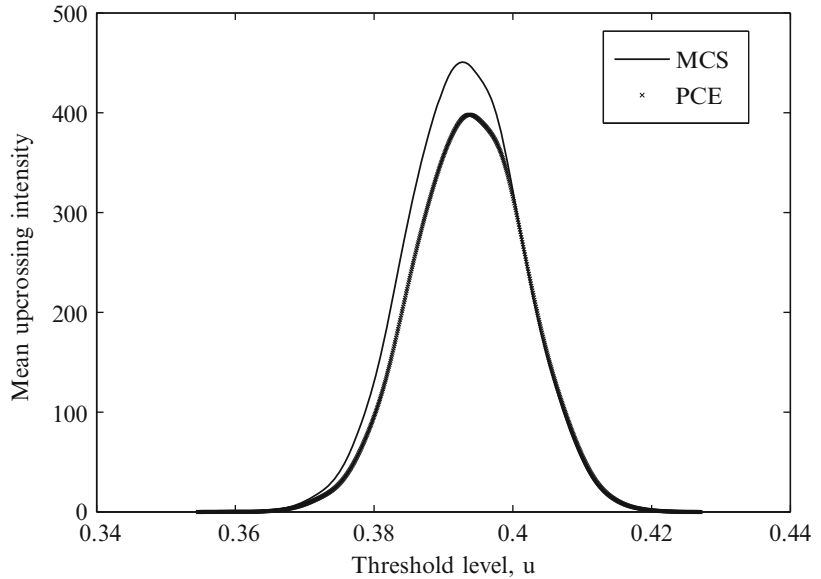
$$\begin{aligned}
 C_L(\tau) = & \pi(\epsilon'' - a_h \alpha'' + \alpha') \\
 & + 2\pi \left\{ \alpha(0) + \epsilon'(0) + \left[\frac{1}{2} - a_h \right] \alpha'(0) \right\} \phi(\tau) \\
 & + 2\pi \int_0^\tau \phi(\tau - \sigma) \left[\alpha'(\sigma) \epsilon''(\sigma) + \left[\frac{1}{2} - a_h \right] \alpha''(\sigma) \right] d\sigma,
 \end{aligned} \tag{58}$$

Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach,

Fig. 16 Contour plot of $p_{xx}(x, \dot{x})$



Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach, Fig. 17 Mean upcrossing intensity; Duffing oscillator



$$\begin{aligned}
 C_M(\tau) = & \pi \left[\frac{1}{2} + a_h \right] \times \left\{ \alpha(0) + \epsilon'(0) + \left[\frac{1}{2} - a_h \right] \alpha'(0) \right\} \phi(\tau) \\
 & + \pi \left[\frac{1}{2} + a_h \right] \times \int_0^T \phi(\tau - \sigma) \left\{ \alpha'(\sigma) + \epsilon''(\sigma) \right. \\
 & \left. + \left[\frac{1}{2} - a_h \right] \alpha''(\sigma) \right\} d\sigma \\
 & + \frac{\pi}{2} a_h (\epsilon'' - a_h \alpha'') - \left[\frac{1}{2} - a_h \right] \frac{\pi}{2} \alpha' - \frac{\pi}{16} \alpha''.
 \end{aligned}
 \quad (59)$$

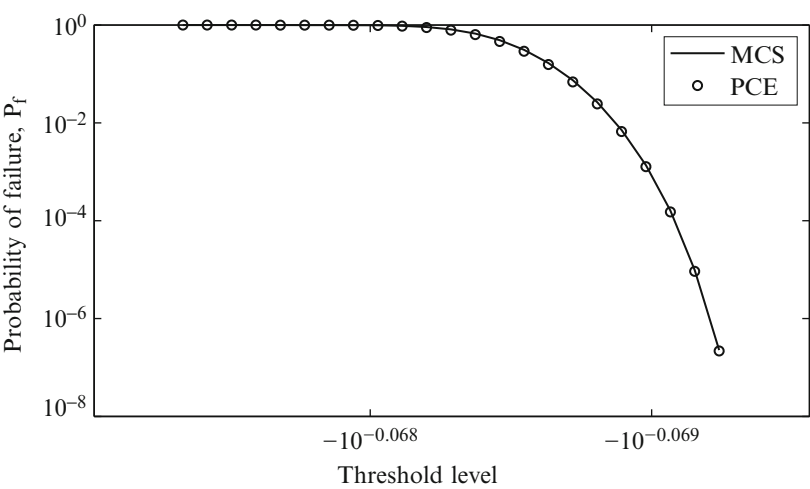
Here $\epsilon(0)$, $\epsilon'(0)$, $\alpha(0)$ and $\alpha'(0)$ respectively denote the initial conditions, $\phi(\tau)$ is the Wagner function expressed as

$$\phi(\tau) = 1 - \psi_1 e^{-\epsilon_1 \tau} - \psi_2 e^{-\epsilon_2 \tau} \quad (60)$$

and the constants $\psi_1 = 0.165$, $\psi_2 = 0.355$, $\epsilon_1 = 0.0455$, and $\epsilon_2 = 0.3$ are as reported in Lee et al. (1998).

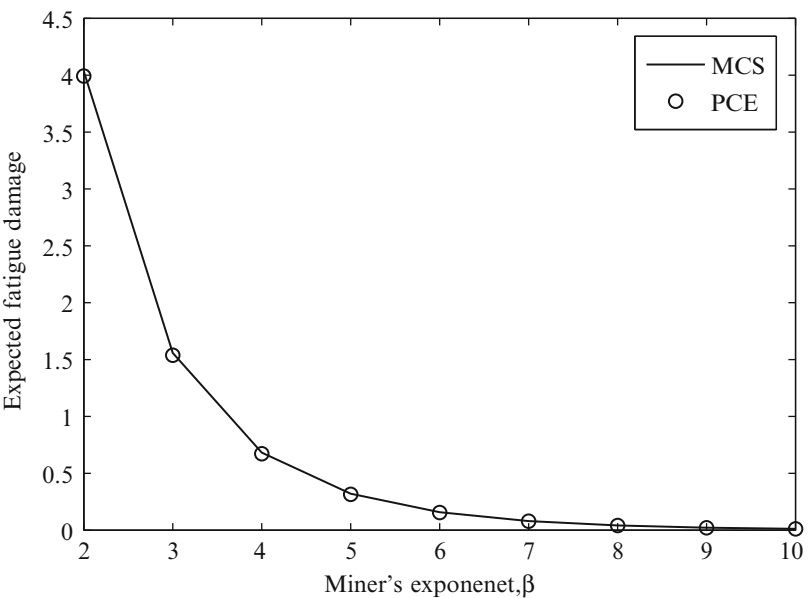
Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach,

Fig. 18 Probability of failure in log scale; Duffing oscillator



Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach,

Fig. 19 Expected fatigue damage; Duffing oscillator



R

In the absence of any gusts, the flow is constant with mean velocity U . The deterministic bifurcation plot for α is shown in Fig. 21. It is seen that bifurcation occurs at $U = 6.285$, after which the systems starts exhibiting limit cycle oscillations (LCO) (Alighanbari and Price 1996). The two branches after $U = 6.285$ represent the minimum and maximum of the LCO amplitude. The phase plot at $U = 16$ as shown in Fig. 22 depicts these limit cycle oscillations.

However, in real life situations, wind flow is not deterministic but has random fluctuations about a mean value. Hence, this randomness must be included in the model in order to gain better insights on the behavior of system. Therefore, the nondimensional free stream velocity U is assumed to be a stationary Gaussian random process with the autocorrelation function as given in Eq. 55 with $\sigma_f = 1$ and $c_0 = 1 \times 10^{-6}$. This implies a correlation length of 2.628×10^3 s. It must be

noted that wind velocity has very low frequency content with very large correlation lengths. The wind flow is represented by a KL expansion of the form

$$U = U_0 + \sum_{i=0}^N \xi_i(\theta) \sqrt{\lambda_i} \phi_i(t). \quad (61)$$

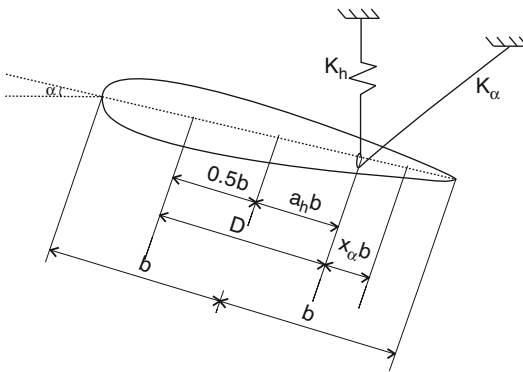
For a nondimensional time duration of 3,600 and $c_0 = 1e^{-6}$, it can be seen from Table 3 that only the first five eigenvalues λ are significant. Hence, a reasonable approximation for U is obtained with a five term K-L expansion. The mean nondimensional speed U_0 is taken as 16. The numerical values of the remaining parameters are taken to be the same as that in Lee

et al. (1998): $\mu = 100$, $\overline{\omega} = 0.2$, $a_h = -0.5$, $x_\alpha = 0.25$, $\zeta_\alpha = 0$, $\zeta_\epsilon = 0$, $r_\alpha = 0.5$.

It can be seen that the governing equations of motion, given by Eqs. 56–59 constitute a set of coupled second order integrodifferential equation and are hence highly nonlinear. These equations can be recast as a set of eight coupled first order nonlinear differential equations by introducing four auxiliary state variables; see Alighanbari and Price (1996). The system is solved using an adaptive 4th order Runge–Kutta algorithm and the results are presented in Figs. 23, 24, and 25.

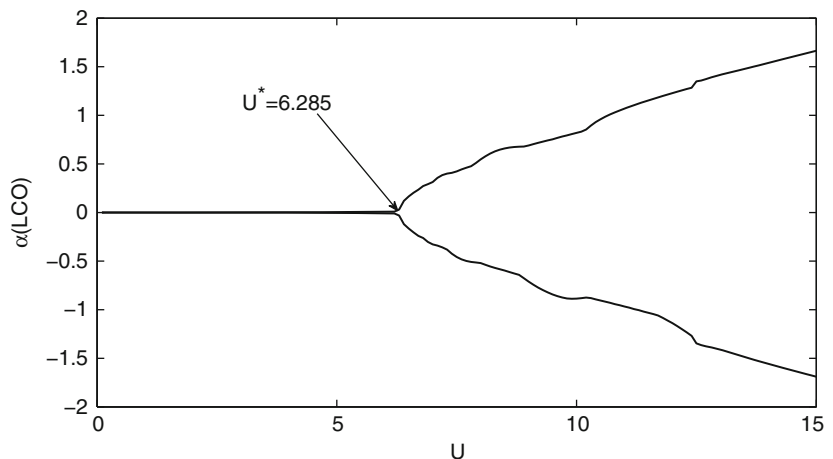
Twelve thousand realizations of the response are generated using both PCE and full-scale Monte-Carlo simulation. A Gauss-Patterson sparse grid of level 6 is used to obtain the PCE coefficients. The number of function evaluations in PCE and MCS are 5,503 and 12,000 respectively. The marginal pdfs of the pitching displacement and pitching velocity are shown in Figs. 23 and 24 respectively. A fairly good resemblance is observed between the predictions obtained by the proposed method and those obtained from MCS.

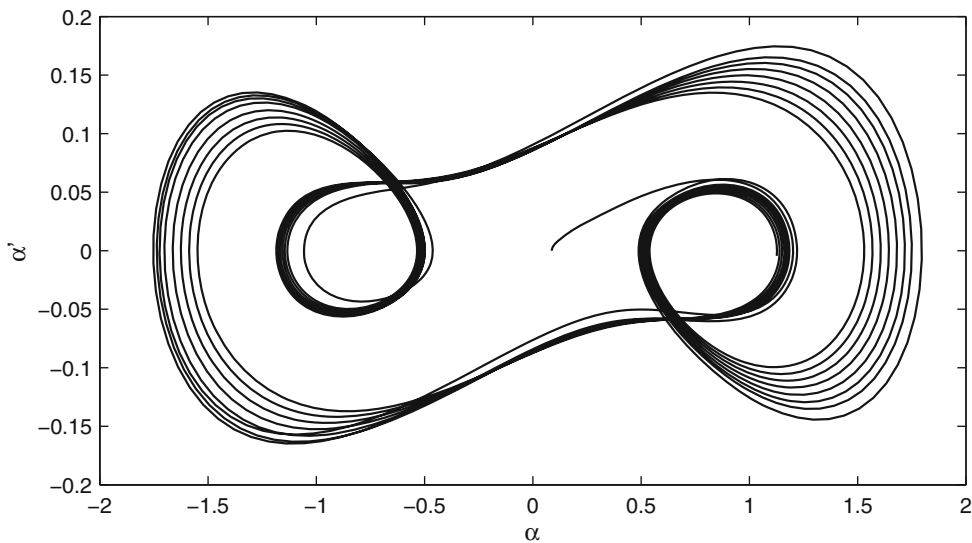
A PCE representation of the response enables generating samples of time histories of the response. Thus, next the joint pdf $p_{\alpha\dot{\alpha}}(\alpha, \dot{\alpha})$ are constructed and its contour plots are shown in Fig. 25. The corresponding pdf obtained from MCS are also plotted in the same figure for comparison. A very good agreement between the two is observed.



Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach, Fig. 20 Schematic diagram of a 2-D airfoil

Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach, Fig. 21 Deterministic bifurcation of α





Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach, Fig. 22 Phase plot at $U_0 = 16$

Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach, Table 3 Magnitude of eigenvalues for the K-L representation of the non-dimensional free stream velocity

Eigenvalue	Magnitude
λ_1	433.98
λ_2	298.66
λ_3	166.01
λ_4	70.42
λ_5	25.05
λ_6	7.46
λ_7	1.90
λ_8	0.42
λ_9	0.08
λ_{10}	0.02

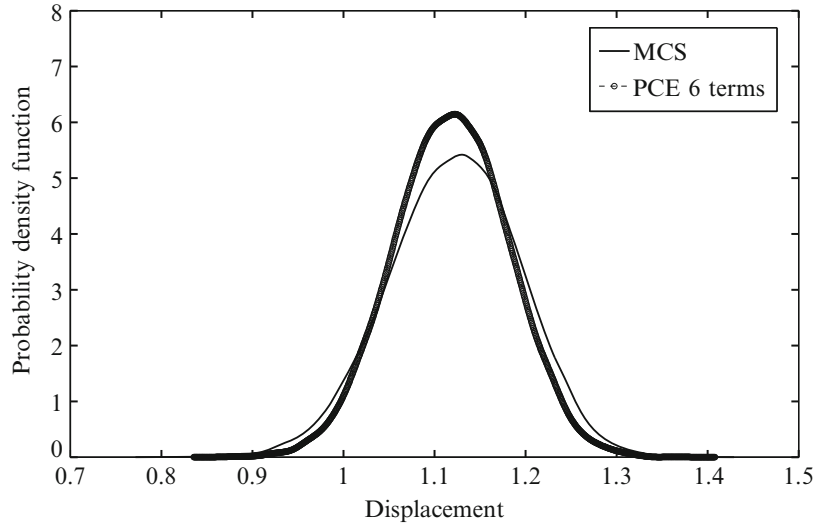
The first passage failure probabilities are next computed for various threshold levels. These are shown in Fig. 26. A good agreement between the predictions of the proposed method and MCS are observed. The expected fatigue damage obtained using the two methods for various values of f_t are shown in Fig. 27. It is seen that the proposed method leads to fairly close resemblance with MCS predictions. However, as f_t becomes larger, the proposed method slightly overestimates the expected fatigue damage and hence is more conservative from the perspective of reliability and residual life assessment.

Summary

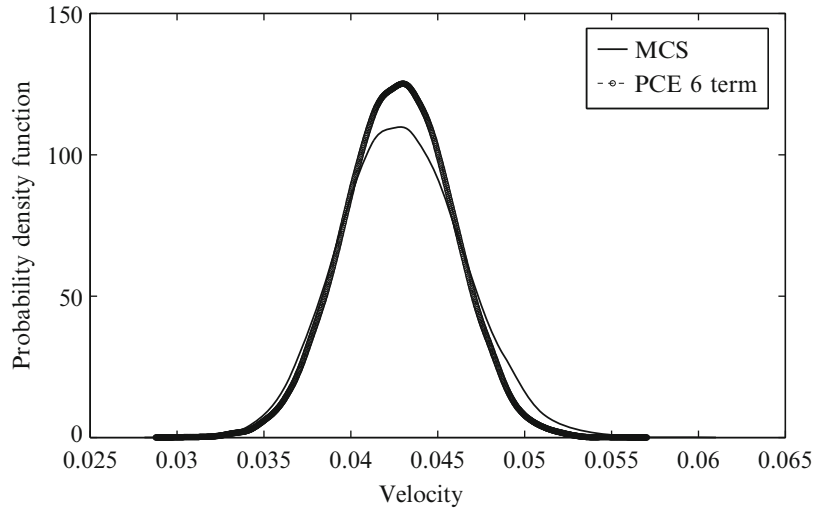
The focus of this entry has been on demonstrating the use of spectral-based methods for time-variant reliability estimation in nonlinear structural systems subjected to random loadings. The crux of this approach lies in approximating the joint pdf of the response and its instantaneous time derivative and subsequently using the Rice’s integral to numerically evaluate the crossing statistics. The loadings have been assumed to be stationary, Gaussian random processes and are represented in terms of the Karhunen-Lo  ve series expansion. Subsequently, polynomial chaos expansions have been used to represent the structure response. In using PCE, the main difficulty lies in estimating the PCE coefficients. In the literature, two general approaches for estimating the PCE coefficients have been discussed. The stochastic Galerkin method is an intrusive approach as the application of this method significantly alters the form of the governing equations of motion. Importantly, the use of this approach requires *a priori* evaluation of the multidimensional expectations.

The alternative and preferred numerical approach is to use the concept of stochastic collocation, which is a nonintrusive approach. This

Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach,
Fig. 23 Marginal pdf $p_x(\alpha)$



Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach,
Fig. 24 Marginal pdf $p_{\dot{x}}(\dot{\alpha})$



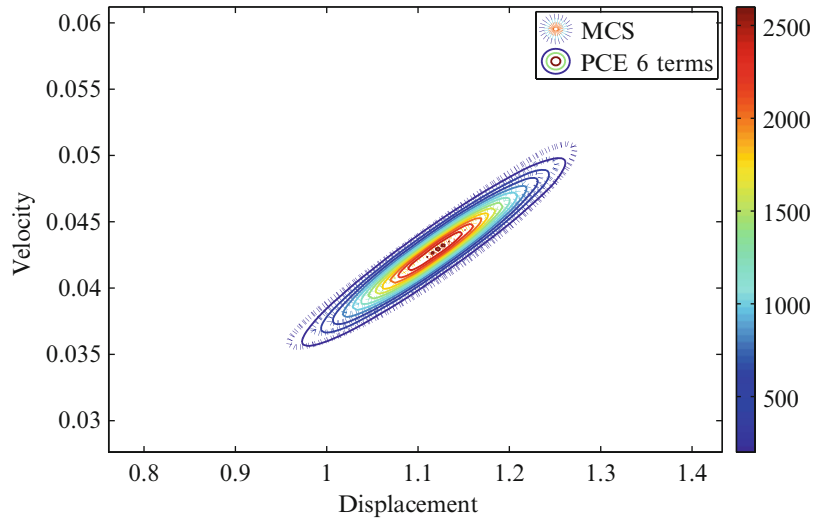
entry has focused on the use of the sparse grid collocation method based on Smolyak's algorithm as a computationally efficient tool for approximating the joint probability density function of the response and its instantaneous time derivative. For highly nonlinear problems such as the fluid-structure interaction problem considered in section 2-D Airfoil Subjected to Unsteady Gusty Wind Flow Regime, the proposed method provides an alternative method for obtaining the crossing statistics of the response. Usually, problems of this genre can be solved using Monte

Carlo simulations only. Though the proposed method requires significant computational effort, this is still less than 50 % of the computational efforts in comparison to Monte Carlo simulations.

The spectral-based method discussed in this entry is computationally efficient as long as the number of collocation points is less than the sample size used in Monte Carlo simulations. This in turn, depends on the number of terms retained in the PCE representation of the response. As the bases for the PCE representation of the response

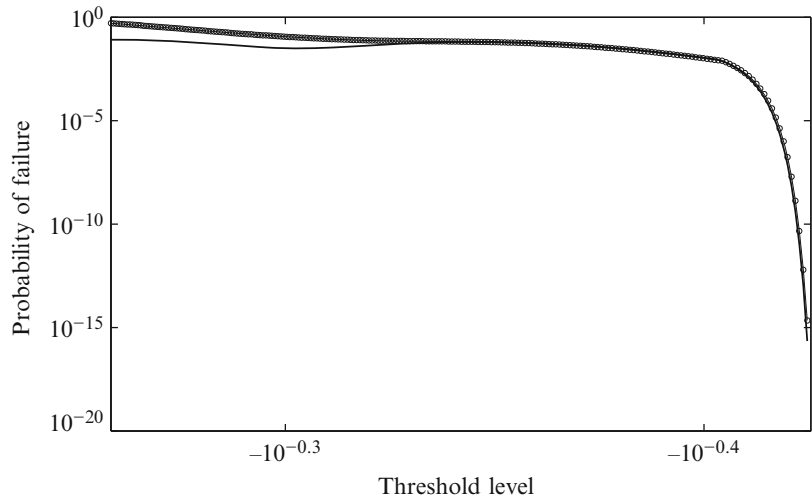
Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach,

Fig. 25 Contour plot of $p_{\alpha\dot{\alpha}}(\alpha, \dot{\alpha})$



Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach,

Fig. 26 Probability of failure 2-D airfoil



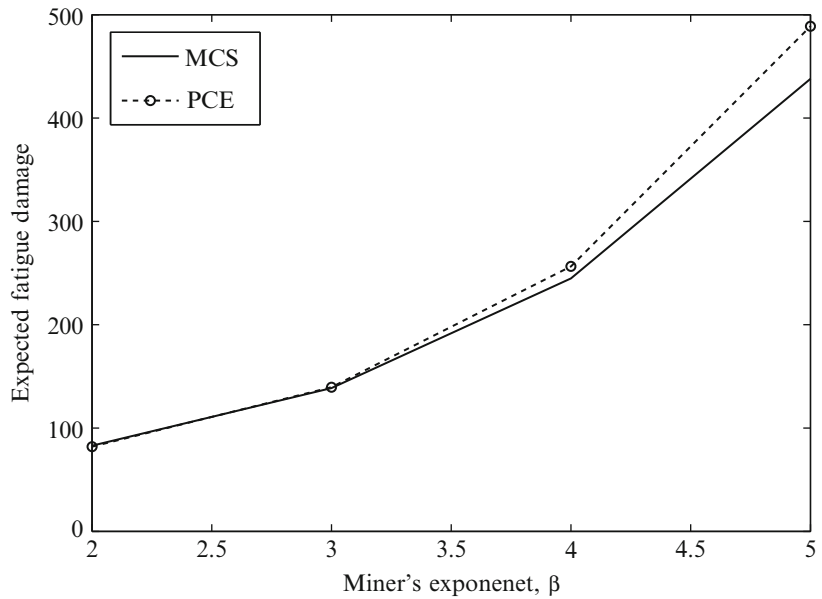
in turn depends on the basis dimension of the input loadings, this method is not suitable for problems where the excitations have a large spectral bandwidth. This is because the number of nonzero eigenvalues for the spectral content of the loading would be large. This, in turn, implies that the stochastic dimension would be large making the method computationally too intensive even with the application of the sparse grid-based algorithms. In such situations, MCS is more computationally efficient.

The excitations considered in this study have been limited to Gaussian excitations. This however

is not a restricting feature. For non-Gaussian excitations, an appropriate PCE representation can be obtained for the input as long as a transformation is available from an equivalent Gaussian process. This aspect of the study has not been addressed in this study.

The present study has considered only the randomness in the excitations, with the structure properties being considered to be deterministic. The studies can be extended to include uncertainties in the material properties as well. This however would increase the computational costs associated with the problem. An alternative

Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach,
Fig. 27 Expected fatigue damage; 2-D airfoil



approach could be to represent the system uncertainties in the spectral domain only and use hybrid approaches that make use of the spectral method discussed in this work along with Monte Carlo simulations. The proposed method can be used efficiently for spectral stochastic finite element method in problems involving characterization of uncertainties across multiple scales. Studies along these lines are currently being pursued in the group.

Cross-References

- [Probability Density Evolution Method in Stochastic Dynamics](#)
- [Reliability Estimation and Analysis](#)
- [Response Variability and Reliability of Structures](#)
- [Stochastic Analysis of Linear Systems](#)
- [Stochastic Analysis of Nonlinear Systems](#)

References

- Alighanbari H, Price SJ (1996) The post-hopf bifurcation response of an airfoil in incompressible, two-dimensional flow. *Nonlinear Dyn* 10(4):381–400
- Babuska I, Nobile F, Tempone R (2007) A stochastic collocation method for elliptic partial differential equations with random input data. *Soc Ind Appl Math* 45:1005–1034
- Barthelmann V, Novak E, Ritter K (2000) High dimensional polynomial interpolation on sparse grids. *Adv Comput Math* 12:273–288
- Cameron RH, Martin WT (1947) The orthogonal development of nonlinear functionals in series of Fourier-Hermite functionals. *Ann Math* 48:385–392
- Cramer H (1966) On the intersections between the trajectories of a normal stationary stochastic process and a high level. *Ark Math* 6:337–349
- Desai A, Sarkar S (2010) Analysis of a nonlinear aeroelastic system with parametric uncertainties using polynomial chaos expansion. *Math Probl Eng*, pages Article ID 379472. doi:10.1155/2010/379472
- Evans M, Swartz T (2000) Approximating integrals via Monte Carlo and deterministic methods. Oxford University Press, Oxford
- Fung YC (1955) An introduction to the theory of aeroelasticity. Wiley, New York
- Gerstner T, Griebel M (1998) Numerical integration using sparse grids. *Numer Algorithms* 18:209–232
- Ghanem R, Spanos PD (1991) Stochastic finite element: a spectral approach. Springer, Berlin
- Lee BHK, Jiang L, Wong YS (1998) Flutter of an airfoil with a cubic nonlinear restoring force. *AIAA* 98-1725:237–257
- Lin YK (1967) Probabilistic theory of structural dynamics. McGraw-Hill, New York
- Melchers RE (1999) Structural reliability analysis and prediction, 2nd edn. Wiley, Chichester
- Nigam NC (1983) Introduction to random vibrations. MIT Press, Cambridge
- Papoulis A, Pillai U (2002) Probability random variables and stochastic processes. Tata McGraw-Hill, New York

- Rice SO (1944) Mathematical analysis of random noise. *Bell Syst Tech J* 23:282–332
- Rychlik I (1993) On the narrow-band approximation for expected fatigue damage. *Probab Eng Mech* 8:1–4
- Shinozuka M, Jan CM (1972) Digital simulation of random processes and its applications. *J Sound Vib* 25:111–128
- Xiu D, Hesthaven JS (2005) Higher order collocation methods for differential equations with random inputs. *Soc Ind Appl Math* 127:1118–1139

Reliability Estimation and Analysis

Enrico Zio
European Foundation for New
Energy–Electricité de France, Ecole Centrale
Paris and Supélec, Paris, France
Energy Department, Politecnico di Milano,
Milan, Italy

Synonyms

Failure analysis; Reliability assessment; Reliability modeling

Introduction

Reliability is a fundamental attribute required to any structure, equipment, and system, because of the impact of failures on production efficiency, safety, and asset integrity. Its technical description requires two successive phases, one of analysis and one of estimation, which are at the basis of what is today called reliability engineering.

The first phase in the analysis of the reliability of a structure, equipment, or system is aimed at identifying the hazards associated to its operation and the mechanisms of failure to which it can be exposed during operation. The analysis is typically qualitative, based on expert judgment but driven by a systematic framework of procedures for organizing the expert knowledge. The output of the analysis is a list of the hazards and failure mechanisms, with technical information on the consequences they can provoke to the function provided by the structure, equipment, or system,

and indications on how to protect from them or mitigate them by design or maintenance. Completeness of this phase of the analysis is obviously fundamental, for the successive quantitative estimation.

The second and final phase aims at the estimation of the probability of failure of the structure, equipment, or system. A fundamental issue in this phase is the uncertainty in the failure occurrences, which must be given due account in the quantitative modeling.

In this entry, concepts and methods of reliability analysis and estimation are introduced and explained. Some of the material is adapted from Zio (2009a, b).

History of Reliability Engineering

The needs for reliability have significantly emerged in relation to mass production for the manufacturing of large quantities of goods from standardized parts (rifle production at the Springfield Armory, 1863, and Ford Model T car production, 1913) (Saleh and Marais 2006). A catalyst for this was the vacuum tube, specifically the triode invented by Lee de Forest in 1906, which at the onset of WWII initiated the electronic revolution, enabling a series of applications such as the radio, television, radar, and others. The vacuum tube is by many recognized as the active element that allowed the Allies to win the so-called wizard war. At the same time, it was also the main cause of equipment failure: tube replacements were required five times as often as all other equipment. After the war, this experience with the vacuum tubes prompted the US Department of Defense (DoD) to initiate a number of studies for looking into these failures.

A similar situation was experienced on the other side of the warfront by the Germans, where Chief Engineer Lusser, a program manager working in Peenemunde on the V1, prompted the systematic analysis of the relations between system failures and components faults.

These and other military-driven efforts eventually led to the rise of the new discipline of

reliability engineering in the 1950s, consolidated and synthesized for the first time in the Advisory Group on Reliability of Electronic Equipment (AGREE) report in 1957. The AGREE was jointly established in 1952 between the DoD and the American Electronics Industry, with the mission of (Coppola 1984):

1. Recommending measures that would result in more reliable equipment
2. Helping to implement reliability programs in government and civilian agencies
3. Disseminating a better education on reliability

Several projects, still military funded, developed in the 1950s from this first initiative (Coppola 1984; Knight 1991; Denson 1998). Failure data collection and root cause analyses were launched with the aim of achieving higher reliability in components and devices. These led to the specification of quantitative reliability requirements, marking the beginning of the contractual aspects of reliability. This inevitably brought the problem of being able to estimate and predict the reliability of a component before it is built and tested: this in turn led in 1956 to the publication of a major report on reliability prediction techniques entitled “Reliability Stress Analysis for Electronic Equipment” (TR-1100) by the Radio Corporation of America (RCA), a major manufacturer of vacuum tubes. The report presented a number of analytical models for estimating failure rates and can be considered the direct predecessor of the influential military standard MH-217 first published in 1961 and still used today to make equipment reliability predictions.

In the 1960s, the discipline of reliability engineering proceeded in two directions:

- Increased specialization in the discipline by sophistication of the techniques, e.g., redundancy modeling, Bayesian statistics, Markov chains, etc., and by the development of the concepts of reliability physics to identify and model the physical causes of failure and of structural reliability to analyze the integrity of buildings, bridges, and other constructions

- Shift of the attention from component reliability to system reliability and availability, to cope with the increased complexity of the engineered systems, like those developed as part of military and space programs like the Mercury, Gemini, and Apollo ones

Three broad areas characterized the development of reliability engineering in the 1970s:

- The potential of system-level reliability analysis (Barlow and Proschan 1975) motivated the rational analysis of the safety of complex systems such as the nuclear power plants (WASH-1400 1975).
- The increased reliance on software in many systems led to the growth of focus on software reliability, testing, and improvement (Moranda 1981).
- The lack of interest on reliability programs that managers often showed already at that time sparked the development of incentives to reward improvement in reliability on top of the usual production-based incentives.

In the following years, the last 40 years, the scientific and technical community has witnessed an impressive increase of developments and applications of reliability engineering, aimed at rationally coping with the challenges brought by the growing complexity of the systems and practically taking advantage of the computational power becoming available at reasonable costs. These developments and applications have been driven by a shift from the traditional industrial economy, valuing products, to the modern economy centered on service. This has led to an increased attention to service availability as a most important quality and to a consequent push in the development of techniques for its quantification, which requires considering a number of interrelated processes of component degradation, of failure and repair, and of diagnostics, prognostics, and maintenance.

Nowadays, reliability engineering is a well-established, scientific discipline which aims at providing an ensemble of formal methods to

investigate the uncertain boundaries between system operation and failure, by addressing the following questions (Cai 1996; Aven and Jensen 1999):

- Why failures occur, e.g., by using the concepts of reliability physics to discover causes and mechanisms of failure and to identify consequences
- How to develop reliable structures, equipment, and systems, e.g., by reliability-based design
- How to measure and test reliability in design, operation, and management
- How to maintain systems reliable, by fault diagnosis, prognosis, and maintenance

Reliability engineering addresses these questions by structured and formal methods of analysis, which entail the representation and modeling of the structure, equipment, or system based on the available knowledge and information, the quantification of the model based on the available data, and the representation, propagation, and quantification of the uncertainty in the behavior of the structure, equipment, or system as inferred from the knowledge, information, and data available, which is always incomplete, imprecise, and often scarce.

Methods of Reliability Analysis: Failure Mode and Effects Analysis (FMEA)

In this section, the qualitative analysis aimed at identifying the hazards and failure mechanisms associated to the operation of a system is exemplified by way of a very common method known as failure mode and effects analysis (FMEA). Actually in practice, a FMECA (failure mode, effects, and criticality analysis) is typically performed to arrive at also assigning a criticality class to each failure mode, for example, according to the following ranking:

- *Safe* = no relevant effects.
- *Marginal* = partially degraded system but no damage to humans.

- *Critical* = system damaged and damages also to humans; if no protective actions are undertaken, the accident could lead to loss of the system and serious consequences on the humans.
- *Catastrophic* = loss of the system and serious consequences on humans.

Such analysis is inductive, based on expert knowledge for identifying the failure modes and inducing their effects and eventually assigning their combined criticality. The analysis aims at identifying, in particular, those failure modes which could impair the function intended for the system, possibly leading to disabled operation or, even, initiating accidents with safety consequences.

The analysis is driven by a systematic procedure, whose basic tasks can be simplified as follows:

1. Decompose the system in functionally independent subsystems; for each subsystem identify the various operation modes (e.g., start-up, regime, shutdown, maintenance, etc.) and its configurations when operating in such modes (e.g., valves open or closed, pumps on or off, etc.).
2. For each subsystem in each of its operation modes, compile a table such as Table 1 below, listing all subsystem components, their failure modes, and their effects on the functionality of other neighboring components, on the subsystem, and eventually the whole system (Henley and Kumamoto 1992).

The analysis considers primarily the effects of single failures, except for the case of standby components for which the effects are considered conditioned on the failure of the main component.

For complex systems, a FMECA can be rather burdensome, in spite of the fact that several computer tools have been developed and made available on the market to aid organizing the implementation. Often, this analysis is used in support to reliability-centered maintenance programs.

Reliability Estimation and Analysis, Table 1 Typical FME(C)A table (Henley and Kumamoto 1992)

System											
Operation mode											
Component	Description	Failure mode	Effects on other components	Effects on subsystem	Effects on plant	Probability	Criticality	Detection methods	Protections and mitigation	Remarks	
		Failure modes relevant for the operational mode indicated	Effects of failure mode on adjacent components and surrounding environment	Effects on the functionality of the subsystem	Effects on the functionality and availability of the entire plant	Probability of failure occurrence (sometimes qualitative)	Criticality rank of the failure mode on the basis of its effects and probability (qualitative estimation of risk)	Methods of detection of the occurrence of the failure event	Protections and measures to avoid the failure occurrence	Remarks and suggestions on the need to consider the failure mode as accident initiator	

Methods of Reliability Estimation

In reliability engineering, the quantitative phase is focused on the time to failure T , which is a continuous random variable whose cumulative distribution function (cdf) $F_T(t)$ and probability density function (pdf) $f_T(t)$ are typically called the failure probability and density functions at time t . The complementary cumulative distribution function (ccdf) $R(t) = 1 - F_T(t) = P(T > t)$ is called *reliability* at time t and gives the probability that the structure, equipment, or system survives up to time t with no failures.

Another description of the failure behavior in time of a structure, equipment, or system, commonly used by reliability engineers and practitioners, is given by the probability that it fails within a time interval dt (mathematically infinitesimal) knowing that it has never failed before the lower bound, t , of the interval. This probability is expressed in terms of the product of the interval dt times a conditional probability density called *hazard function* or *failure rate* and usually indicated by the symbol $h_T(t)$:

$$\begin{aligned} h_T(t)dt &= P(t < T \leq t + dt | T > t) \\ &= \frac{P(t < T \leq t + dt)}{P(T > t)} = \frac{f_T(t)dt}{R(t)} \end{aligned} \quad (1)$$

The hazard function $h_T(t)$ gives the same information of the pdf and cdf to whom it is univocally related by Eq. 1 and its integration, i.e.,

$$F_T(t) = 1 - e^{-\int_0^t h_T(s)ds} \quad (2)$$

Figure 1 shows some common patterns of $h_T(t)$ in time, encountered in practice (Zeng 1997).

In the most general case, the hazard function follows the so-called “bathtub” curve (Fig. 1, case 1), which shows three distinct phases in the life of a component: the first phase corresponds to a failure rate decreasing with time, and it is characteristic of the *infant mortality* or *burn-in* period whereupon the more the structure, equipment, or system lives with no failures, the lower its probability of failure itself becomes (this period is

important for warranty analysis); the second phase, called *useful life*, corresponds to a failure rate independent of time: during this period, failures occur at random times with no influence on the usage time of the structure, equipment, or system; finally, the last phase sees an increase in the failure rate with time and corresponds to the development of aging processes.

Deviations from this general behavior (Fig. 1, cases 2–5) occur depending on the burn-in and maintenance procedures adopted by the particular industry.

The Weibull distribution is often used in reliability practice to describe the failure behavior in time of a structure, equipment, or system. The cdf of the failure time random variable is

$$F_T(t) = P(T \leq t) = 1 - e^{-\lambda t^\alpha} \quad (3)$$

and the corresponding pdf is

$$\begin{aligned} f_T(t) &= \lambda \alpha t^{\alpha-1} e^{-\lambda t^\alpha} & t \geq 0 \\ &= \lambda \alpha t^{\alpha-1} e^{-\lambda t^\alpha} & t < 0 \end{aligned} \quad (4)$$

The expected value and variance of the Weibull distribution are

$$\begin{aligned} E[T] &= \frac{1}{\lambda} \Gamma\left(\frac{1}{\alpha} + 1\right) \\ \text{Var}[T] &= \frac{1}{\lambda^2} \left(\Gamma\left(\frac{2}{\alpha} + 1\right) - \Gamma\left(\frac{1}{\alpha} + 1\right)^2 \right) \end{aligned} \quad (5)$$

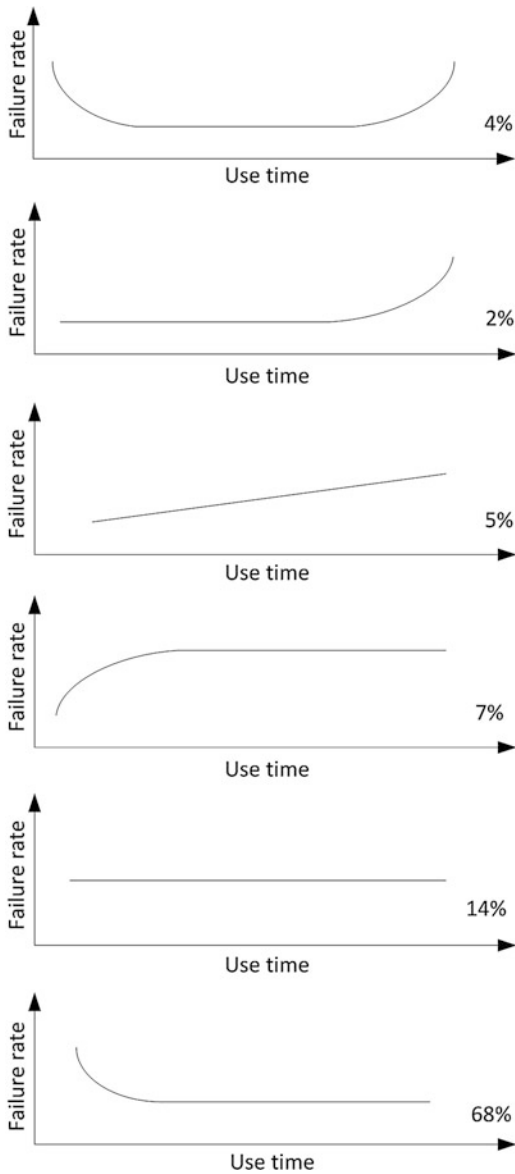
where the Gamma function $\Gamma(\cdot)$ is the generalization to non-integer numbers of the factorial and is defined as

$$\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} dx \quad k > 0 \quad (6)$$

which by integration by parts yields

$$\Gamma(k) = (k-1)\Gamma(k-1) \quad (7)$$

For $\alpha < 1$, the distribution can represent the burn-in period; for $\alpha > 1$, it can describe the aging period; for $\alpha = 1$, the distribution is called exponential, characterized by a constant failure rate,



Reliability Estimation and Analysis, Fig. 1 Patterns of time evolution of the hazard function (or failure rate) (Zeng 1997)

and describes the useful part of the life of the structure, equipment, or system. In this latter case, the expected value of the distribution is

$$E[T] = \frac{1}{\lambda} \quad (8)$$

and is called *mean time to failure* (MTTF).

The reliability of a series system made of N components, whose logic of operation is that all components must function for the system to function, is

$$R(t) = \prod_{i=1}^N R_i(t) \quad (9)$$

$R_i(t)$, $i = 1, 2, \dots, N$ are the individual components' reliabilities.

For exponential components of individual reliabilities $R_i(t) = e^{-\lambda_i t}$, $i = 1, 2, \dots, N$, the system reliability becomes $R(t) = e^{-\lambda t}$, where

$$\lambda = \sum_{i=1}^N \lambda_i = \text{system failure rate} \quad (10)$$

$$m = \frac{1}{\lambda} = \text{mean time to failure}$$

The series system is the only logic configuration in which components with constant failure rates induce a constant failure rate for the system. In all other configurations, the reliability of the system is not exponential. The system fails at $\min(t_1, t_2, \dots, t_N)$, where t_i is the failure time of the i th component in the system.

The reliability of a parallel system, in which all components are performing the same function so that anyone by itself can successfully carry out the system function, is

$$R(t) = 1 - \prod_{i=1}^N [1 - R_i(t)] \quad (11)$$

Since the system fails when all its components fail, the time to failure of the system is $\max(t_1, t_2, \dots, t_N)$.

More complex logics of system operation, including standby, multistate, etc., require more sophisticated modeling tools like Markov modeling (Zio 2009b) and Monte Carlo simulation (Zio 2013). Both of these methods are capable of accounting for the system behavior as described by its multiple states and the transitions among them. The system states are defined by the states of the units comprising the system. The units are not restricted to having only two possible states

but rather may have a number of different states such as functioning, in standby, degraded, partially failed, completely failed, under maintenance, etc.; the various failure modes of a unit may also be defined as states. The transitions between the states occur randomly in time, because of various mechanisms and activities such as failures, repairs, replacements, and switching operations, which are random in nature. Common cause failures may also be included as possible transitions occurring randomly in time.

Under specified conditions, the stochastic process of the system evolution may be described as a Markov process in which the system states and the possible transitions can be depicted with the aid of a state-space diagram, known as a Markov diagram, and be mathematically described by a probabilistic Markov system of equations. The Markov property states the following: given that a system is in state i at time t , the probability of reaching state j at time $t + v$ does not depend on the states visited by the system prior to t ($0 \leq u < t$): in other words, given the present state of the system, its future behavior is independent of the past.

However, in realistic conditions often the behavior of a system may not respect the Markov property. In these cases, the quantitative analysis may be effectively carried out by Monte Carlo simulation, which corresponds to performing a virtual experiment in which a large number of identical systems, each one behaving differently due to the stochastic character of the system behavior, are simulated during a given time and their failure occurrences are recorded. This, in principle, is the same procedure adopted in the physical (not virtual) reliability tests performed on individual units to estimate their failure rates, mean time to failure, or other parameter characteristics of their failure behavior; the difference is that for units, the tests are actually done physically in laboratory, at reasonable costs and within reasonable testing times (possibly by resorting to accelerated testing techniques, when necessary), whereas for systems this is obviously impracticable for the costs and times

involved in systems failures. Thus, instead of making physical tests on a system, the stochastic process of transition among its states is modeled by defining the probabilistic distribution governing the transition process, and a large number of realizations are generated by sampling from it the times and outcomes of the occurring transitions.

The Monte Carlo simulation of one single system life entails the repeated sampling from the probabilistic distribution of the time of occurrence of the next transition and of the new configuration reached by the system as outcome of the transition, starting from the current system configuration. Then, for the purpose of reliability estimation, a subset of the system configurations is identified as the set of fault states. Whenever the system enters one such configuration, its failure is recorded together with its time of occurrence. With reference to a given time t of interest, an estimate of the probability of system failure before such time, i.e., the system failure probability at time t , can be obtained by the frequency of system failures before t , computed by dividing the number of system life realizations which record a system failure before t by the total number of system life realizations simulated.

Summary

This entry contains some basic concepts, general knowledge, and classical methods for the analysis of the reliability of structures, equipment, and systems and its quantitative estimation. The presentation of the subject matter follows the two phases typically used in practice for the implementation of the analysis: the phase of qualitative analysis for hazards and failure modes identification and the phase of quantitative estimation of the reliability characteristics of interest. The concepts and practices of the first phase are exemplified via the common method of systematic analysis known as the failure mode, effects, and criticality analysis (FMECA). The mathematical, probabilistic concepts necessary

for the second phase of quantification are introduced for individual components and simple system configurations like the series and parallel logics. Also, a qualitative description is provided of advanced methods like Markov modeling and Monte Carlo simulation for the estimation of the reliability characteristics of complex systems.

Cross-References

- [Subset Simulation Method for Rare Event Estimation: An Introduction](#)
- [Uncertainty Theories: Overview](#)

References

- Aven T, Jensen U (1999) Stochastic models in reliability, vol 41. Springer, New York
- Barlow RE, Proschan F (1975) Statistical theory of reliability and life testing: probability models. Florida State University Tallahassee Holt, Rinehart and Winston
- Cai KY (1996) System failure engineering and fuzzy methodology an introductory overview. *Fuzzy Set Syst* 83(2):113–133
- Coppola A (1984) Reliability engineering of electronic equipment a historical perspective. *Reliab IEEE Trans* 33(1):29–35
- Denson W (1998) The history of reliability prediction. *Reliab IEEE Trans* 47(3):SP321–SP328
- Henley EJ, Kumamoto H (1992) Probabilistic risk assessment. IEEE Press, New York
- Knight CR (1991) Four decades of reliability progress. In: Reliability and maintainability symposium, 1991. Proceedings, annual. IEEE, pp 156–160
- Moranda PB (1981) An error detection model for application during software development. *Reliab IEEE Trans* 30(4):309–312
- Saleh JH, Marais K (2006) Highlights from the early (and pre-) history of reliability engineering. *Reliab Eng Syst Saf* 91(2):249–256
- US Nuclear Regulatory Commission (1975) Reactor safety study, WASH-1400 (NUREG-75/014)
- Zeng SW (1997) *Reliab Eng Syst Saf*. Elsevier 55:151–162
- Zio E (2009a) Computational methods for reliability and risk analysis. World Scientific Publishing
- Zio E (2009b) Reliability engineering: old problems and new challenges. *Reliab Eng Syst Saf* 94(2):125–141
- Zio E (2013) The Monte Carlo simulation method for system reliability and risk analysis. Springer, London

Reliability Estimation and Analysis for Dynamical Systems

V. S. Sundar and C. S. Manohar

Department of Civil Engineering, Indian Institute of Science, Bangalore, Karnataka, India

Synonyms

Seismic reliability; Seismic safety; Time variant reliability; Variance reduction

Introduction

The problem of time variant reliability of randomly parametered nonlinear dynamical systems is formulated. Various analytical and simulation-based strategies to tackle this problem are outlined. Questions on updating reliability models for instrumented structures are also briefly reviewed.

Engineering structures are designed for earthquake loads to display specific forms of inelastic behavior with certain modes of failures preferred over others. The analysis of earthquake response of structures involves transient dynamic analysis which allows for inelastic and geometric nonlinear structural behavior and takes into account uncertainties associated with earthquake-induced loads and structural dynamic properties in the inelastic transient regime. The earthquake-induced ground motions are typically modeled as a vector of non-stationary, Gaussian random processes, which allow for multicomponent and spatial variability characteristics. The uncertain structural properties are often modeled as a vector of mutually dependent non-Gaussian random variables and, in a more realistic framework, as a vector of non-Gaussian random fields evolving in space. When random field models are employed, the random fields are suitably discretized and represented as an equivalent set of random variables. Thus, the governing equation to be studied consists of a set of coupled nonlinear ordinary differential equations with random parameters and random external and (or) parametric excitations. Questions on

reliability measures associated with solution processes here pose significant challenge to the analyst. This entry aims to provide a selective overview of some of the methods and tools to tackle these questions.

Problem Statement

Semi-discretized equations of motion, typically resulting from the application of the finite element method to a structural mechanics problem, of the form

$$\begin{aligned} & \mathbf{M}(\boldsymbol{\theta}) \ddot{\mathbf{Y}}(t) + \mathbf{C}(\boldsymbol{\theta}) \dot{\mathbf{Y}}(t) + \mathbf{K}(\boldsymbol{\theta}) \mathbf{Y}(t) \\ & + \mathbf{Q}[\dot{\mathbf{Y}}(t), \mathbf{Y}(t)] + \bar{\mathbf{Q}}[\dot{\mathbf{Y}}(\tau), \mathbf{Y}(\tau), 0 \leq \tau \leq t] \\ & = \mathbf{F}(t) \mathbf{Y}(0) = \mathbf{Y}_0; \dot{\mathbf{Y}}(0) = \dot{\mathbf{Y}}_0 \end{aligned} \quad (1)$$

are considered. Here a dot represents differentiation with respect to time t ; $\mathbf{Y}(t)$ is a $\bar{d} \times 1$ vector of displacement vector; $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are, respectively, the mass, damping, and stiffness matrices; $\mathbf{Q}$ is a nonlinear function of instantaneous displacement and velocity vectors (arising from geometric nonlinear effects); $\bar{\mathbf{Q}}$ is a nonlinear function of system response time histories up to time t (arising from material nonlinear effects); $\mathbf{F}(t)$ is a vector of external forces; and $\boldsymbol{\theta}$ is a $p \times 1$ vector of system parameters. The forcing vector $\mathbf{F}(t)$ is taken to be a vector of mutually dependent random processes which are not necessarily white, Gaussian, or stationary. The parameter vector $\boldsymbol{\theta}$ encapsulates all the sources of uncertainties arising in specification of structural properties including geometric, mass, stress-strain relations, boundary conditions, and damping characteristics. In general, elements of $\boldsymbol{\theta}$ are modeled as a set of mutually dependent non-Gaussian random variables specified through a p dimensional joint probability density function (PDF) $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$. Let $h(t) \equiv h[\mathbf{Y}(t), \dot{\mathbf{Y}}(t), \ddot{\mathbf{Y}}(t), \boldsymbol{\theta}]$ be a response metric such that the probability measure $P_S = P\{h[\mathbf{Y}(t), \dot{\mathbf{Y}}(t), \ddot{\mathbf{Y}}(t), \boldsymbol{\theta}] \leq h^* \forall t \in [t_0, t_0 + T]\}$ represents the probability that response metric $h(t)$ stays below the permissible value h^* for all times during the interval

$[t_0, t_0 + T]$. Here P_S is the reliability and its complement $P_F = 1 - P_S$ is the probability of failure.

Remarks

1. The probability of failure here is defined with respect to a specified performance metric $h(t)$ which in general is a non-Gaussian random process.
2. Since the reliability here is defined with respect to permissible values of $h(t)$ over a given time duration $[t_0, t_0 + T]$, the problem of reliability analysis is classified as being time variant.
3. By defining $h_M(\boldsymbol{\theta}) = \max_{t \in [t_0, t_0 + T]} h[\mathbf{Y}(t), \dot{\mathbf{Y}}(t), \ddot{\mathbf{Y}}(t), \boldsymbol{\theta}]$, the maximum of the random process $h(t)$ over the duration $[t_0, t_0 + T]$, the probability of failure can be expressed as $P_F = P[h^* - h_M(\boldsymbol{\theta}) \leq 0] = \langle I[h^* - h_M(\boldsymbol{\theta}) \leq 0] \rangle$. Here $I[h^* - h_M(\boldsymbol{\theta}) \leq 0]$ is the indicator function such that $I[h^* - h_M(\boldsymbol{\theta}) \leq 0] = 1$ if $h^* - h_M(\boldsymbol{\theta}) \leq 0$ and zero otherwise, and $\langle \cdot \rangle$ is the mathematical expectation operator. Expressed in this form, the time variant reliability problem becomes time invariant in nature. This conceptual simplification, however, crucially depends upon the ability to characterize the PDF of $h_M(\boldsymbol{\theta})$ which, in itself, is a challenging problem.
4. The nature of the problem gets simplified if randomness is present only in specifying external excitations $\mathbf{F}(t)$ or only in specifying system parameters $\boldsymbol{\theta}$.
5. Problems of time variant system reliability, which involve more than one performance metric, can also be formulated, but their study is far more involved.

Exact solutions to the problem of determining P_F are not available for most cases of practical interest. One takes recourse to approximate analytical solutions or to Monte Carlo simulation-based methods. The analytical methods can be broadly classified into those based on level crossing statistics and those based on Markovian property of system responses. The simulation-based methods typically employ suitable strategies to control the sampling variance. A discussion of these topics forms the subject of the following sections.

Methods Based on Level Crossing Statistics

Introducing a counting process, where $N(h^*, 0, T)$ = the number of times the process $h(t)$ crosses level h^* during the interval $[0, T]$, and a random variable $T_f(h^*)$ = time taken by $h(t)$ to cross level h^* for the first time with $t \geq 0$, it can be shown that (Nigam 1983; Soong and Grigoriu 1993)

$$N(h^*, 0, T) = \int_0^T |\dot{h}(t)| \delta[h(t) - h^*] dt \quad (2)$$

It follows that $P[T_f(h^*) \geq t] = P[N(h^*, 0, t) = 0]$. Here it is assumed that at $t = 0$, the structure lies in the safe region. Furthermore, the random variable $h_M = \max_{t \in [0, T]} h(t)$ can be characterized using the relation $P[h_M \leq h^*] = P[T_f(h^*) \geq T]$. Thus the key to the determination of the reliability $1 - P_F = P[h_M \leq h^*]$ lies in the characterization of the counting process $N(h^*, 0, T)$. If h^* is taken to be large, such that, crossing of h^* by $h(t)$ can be considered as rare, $N(h^*, 0, T)$ can be approximated to be a Poisson random variable (see, Leadbetter et al. 1983, for the details of theoretical basis for this approximation) leading to

$$P[N(h^*, 0, T) = n] = \exp(-\lambda T) \frac{(\lambda T)^n}{n!}; \quad (3)$$

$$n = 0, 1, 2, \dots, \infty$$

Here the parameter λ is the average rate of crossing of level h^* given by

$$\lambda = \int_{-\infty}^{\infty} |\dot{h}| p_{h\dot{h}}(h^*, h; t) d\dot{h} \text{ which, for a stationary}$$

Gaussian random process, is given by $\lambda = \frac{\sigma_{\dot{h}}}{\pi \sigma_h} \exp\left\{-\frac{1}{2} \frac{h^{*2}}{\sigma_h^2}\right\}$ where σ_h and $\sigma_{\dot{h}}$ are, respectively, the steady-state variance of $h(t)$ and $\dot{h}(t)$. For more general class of models for $h(t)$, the evaluation of λ requires the knowledge of joint PDF of $h(t)$ and $\dot{h}(t)$. When $N(h^*, 0, T)$ is modeled as a Poisson random variable, the first passage time $T_f(h^*)$ gets distributed

exponentially and a Gumbel model for h_M can be further deduced from this. Several refinements to this basic idea are available in the literature (Nigam 1983; Soong and Grigoriu 1993): these involve development of one-step memory models for the counting process $N(h^*, 0, T)$ and models using envelope processes of $h(t)$.

Markov Vector Methods

When $F(t)$ arises as a vector of Gaussian white noise process or as a vector of filtered white noise processes, it becomes advantageous to interpret Eq. 1 as an Ito's stochastic differential equation (SDE) and represent it as

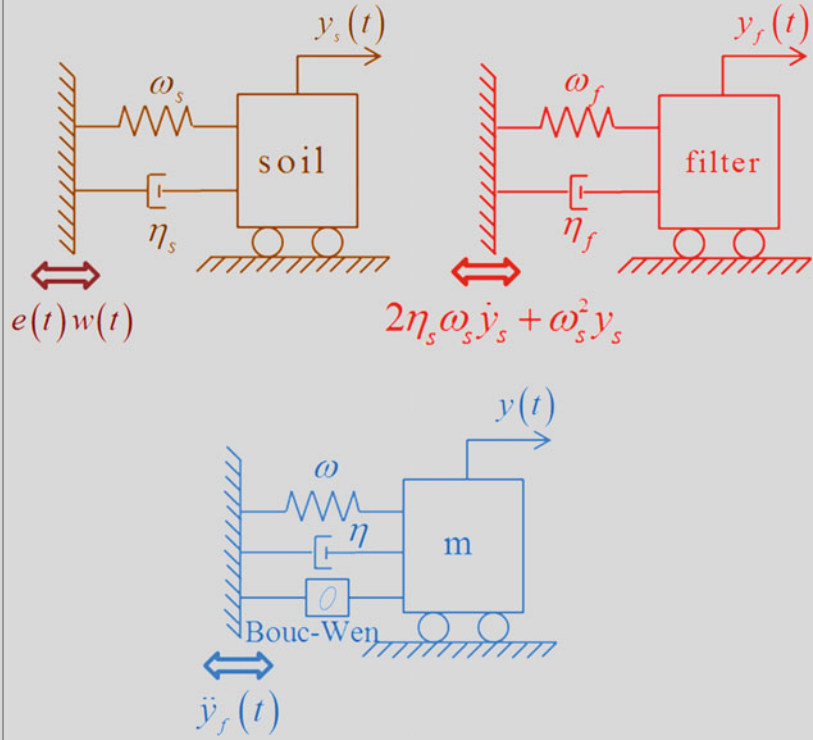
$$dX(t) = A[X(t), \Theta, t]dt + D[X(t), \Theta, t]dB(t)$$

$$X(0) = X_0 \quad (4)$$

Here $X(t)$ is the $d \times 1$ vector, A is the $d \times 1$ drift vector, D is the $d \times q$ diffusion matrix, and $dB(t)$ is a $q \times 1$ vector of increments of Brownian motion processes with $\langle dB(t) \rangle = 0$ and $\langle dB(t) dB^T(t + \tau) \rangle = \Sigma(t) \delta(\tau)$. The vector $X(t)$ not only includes displacement and velocity vectors [i.e., $Y(t)$ and $\dot{Y}(t)$] but also contains additional states arising in representation of hysteretic nonlinearities through internal variables and states appearing in representing applied excitations as filtered white noise processes. Box 1 illustrates the deduction of the governing equation in the above form for a single degree of freedom (SDOF) system with hysteretic nonlinearity and filtered white noise random excitation. Considering the vector of system parameters, Θ , to be deterministic, it can be shown that the response vector $X(t)$ possesses the Markovian property (Lin and Cai 1995). Such a process is completely described by the transition PDF $p(x; t|x_0; t_0)$ and the initial PDF $p(x_0; t_0)$. Associated with the governing SDE, one can derive equations governing the time evolution of the transitional PDF, $p(x, t|x_0, t_0)$. These equations can take three alternative forms, namely, the form of an integral equation

Box 1: Bouc–Wen Hysteretic SDOF System Under Kanai–Tajimi–Clough–Penzien Earthquake Excitation

$$\begin{aligned}
 & \ddot{y}_s + 2\eta_s \omega_s \dot{y}_s \\
 & + \omega_s^2 y_s = e(t)w(t)\ddot{y}_f \\
 & + 2\eta_f \omega_f \dot{y}_f + \omega_f^2 y_f \\
 & = 2\eta_s \omega_s \dot{y}_s + \omega_s^2 y_s \ddot{y} \\
 & + 2\eta \omega \dot{y} + \alpha y + (1 - \alpha) \bar{q} \\
 & = \ddot{y}_f \bar{q} = -\gamma |\dot{y}| \bar{q} |\bar{q}|^{v-1} \\
 & - \beta |\dot{y}| \bar{q}|^v + A \ddot{y}_s(0) = y_{s0}, \\
 & \dot{y}_s(0) = \dot{y}_{s0}, y_f(0) = y_{f0}, \\
 & \dot{y}_f(0) = \dot{y}_{f0}, y(0) = y_0, \\
 & \dot{y}(0) = \dot{y}_0
 \end{aligned}$$



Here $w(t)$ is modeled as a zero-mean white noise excitation with autocovariance $\langle w(t)w(t+\tau) \rangle = \sigma^2 \delta(\tau)$, $e(t) = A_0(e^{-\alpha_1 t} - e^{-\alpha_2 t})$ is a deterministic envelope function which imparts non-stationarity to the input at the bed rock level, and ω and η are the system parameters. The subscripts s and f denote, respectively, the parameters of the Kanai–Tajimi model and the Clough–Penzien high-pass filter. The parameters α , β , A , v and γ characterize the inelastic properties of the oscillator.

Representation as an Ito's SDE

$$\begin{aligned}
 x_1 &= x_s & dx_1 &= x_2 dt \\
 x_2 &= \dot{x}_s & dx_2 &= -2\eta_s \omega_s x_2 dt - \omega_s^2 x_1 dt - e(t)w(t)dt \\
 x_3 &= x_f & dx_3 &= x_4 dt \\
 x_4 &= \dot{x}_f & \Rightarrow dx_4 &= 2\eta_s \omega_s x_2 dt + \omega_s^2 x_1 dt - 2\eta_f \omega_f x_4 dt - \omega_f^2 x_3 dt \\
 x_5 &= x & dx_5 &= x_6 dt \\
 x_6 &= \dot{x} & dx_6 &= 2\eta_s \omega_s x_2 dt + \omega_s^2 x_1 dt - 2\eta_f \omega_f x_4 dt - \omega_f^2 x_3 dt - 2\eta \omega x_6 dt + \alpha x_5 dt + (1 - \alpha) x_7 dt \\
 x_7 &= \bar{q} & dx_7 &= -\gamma |x_6| x_7 |x_7|^{v-1} dt - \beta x_6 |x_7|^v dt + A x_6 dt
 \end{aligned}$$

(continued)

$$dX(t) = \mathbf{A}_1(\boldsymbol{\Theta})X(t) + \mathbf{A}[X(t), \boldsymbol{\Theta}, t]dt + \mathbf{D}(t)dB(t)$$

$$\text{with } \mathbf{X} = (x_s, \dot{x}_s, x_f, \dot{x}_f, x, \dot{x}, \bar{q})^t; \boldsymbol{\Theta} = (\omega_s, \eta_s, \omega_f, \eta_f, \omega, \eta)^t \sim p_{\boldsymbol{\Theta}}(\boldsymbol{\Theta})$$

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -\omega_s^2 & -2\eta_s\omega_s & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ \omega_s^2 & 2\eta_s\omega_s & -\omega_f^2 & -2\eta_f\omega_f & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ \omega_s^2 & 2\eta_s\omega_s & -\omega_f^2 & -2\eta_f\omega_f & \alpha & -2\eta\omega & (1-\alpha) \\ 0 & 0 & 0 & 0 & 0 & 0 & A \end{bmatrix}; \mathbf{A} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ -\gamma|x_6|x_7|x_7|^{v-1} - \beta x_6|x_7|^v \end{bmatrix}; \mathbf{D} = \begin{bmatrix} 0 \\ e(t) \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

(in which case it is called the Chapman–Kolmogorov–Smoluchowski equation), a diffusion type of partial differential equation with $\mathbf{x}$ and t as independent variables (in which case it is called the forward Kolmogorov equation or the Fokker–Planck–Kolmogorov equation), and a diffusion type of partial differential equation with $\mathbf{x}_0$ and t_0 as independent variables (in which case it is called the Kolmogorov backward equation). Furthermore, the governing equation for the survival function, $R(t|\mathbf{x}_0, t_0)$ = probability that the time for response originating at $\mathbf{x}_0$ at t_0 to cross a critical threshold is greater than t , also satisfies the operation appearing in the

Kolmogorov backward equation. These equations are listed below.

1. Chapman–Kolmogorov–Smoluchowski equation:

$$p(\mathbf{x}; t|\mathbf{x}_0; t_0) = \int p(\mathbf{x}; t|\bar{\mathbf{x}}; \tau) p(\bar{\mathbf{x}}; \tau|\mathbf{x}_0; t_0) d\bar{\mathbf{x}} \\ \forall t_0 < \tau < t$$

This is the consistency condition that the transition PDF needs to satisfy so as to ensure that $\mathbf{X}(t)$ is a Markov vector.

2. Fokker–Planck–Kolmogorov equation:

$$\frac{\partial}{\partial t} p(\mathbf{x}; t|\mathbf{x}_0; t_0) = - \sum_{i=1}^d \frac{\partial}{\partial x_i} \{p(\mathbf{x}; t|\mathbf{x}_0; t_0) A_i(\mathbf{x}, t)\} + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} \left\{ p(\mathbf{x}; t|\mathbf{x}_0; t_0) [\mathbf{D}(\mathbf{x}, t) \boldsymbol{\Sigma} \mathbf{D}^t(\mathbf{x}, t)]_{ij} \right\} \\ p(\mathbf{x}; t_0|\mathbf{x}_0, t_0) = \delta(\mathbf{x} - \mathbf{x}_0) = \prod_{i=1}^d (x_i - x_{0i}); \lim_{x_i \rightarrow \infty} p(\mathbf{x}; t|\mathbf{x}_0, t_0) \rightarrow 0, i = 1, 2, \dots, d$$

3. Kolmogorov backward equation:

$$\frac{\partial}{\partial t_0} p(\mathbf{x}; t|\mathbf{x}_0; t_0) = - \sum_{i=1}^d A_i(\mathbf{x}_0, t_0) \frac{\partial}{\partial x_{0i}} \{p(\mathbf{x}; t|\mathbf{x}_0; t_0)\} - \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d [\mathbf{D}(\mathbf{x}_0, t_0) \boldsymbol{\Sigma} \mathbf{D}^t(\mathbf{x}_0, t_0)]_{ij} \frac{\partial^2}{\partial x_{0i} \partial x_{0j}} [p(\mathbf{x}; t|\mathbf{x}_0; t_0)] \\ p(\mathbf{x}; t_0|\mathbf{x}_0, t_0) = \delta(\mathbf{x} - \mathbf{x}_0) = \prod_{i=1}^d (x_i - x_{0i}); \lim_{x_i \rightarrow \infty} p(\mathbf{x}; t|\mathbf{x}_0, t_0) \rightarrow 0, i = 1, 2, \dots, d$$

4. Kolmogorov backward equation for survival time:

Let $R(t|\mathbf{x}_0; t_0)$ denote the probability that the response of interest, $\mathbf{X}(t)$, starting from the safe

region Ω , takes time $\mathcal{T}$ to cross the boundary $\partial\Omega$ for the first time. Then, $R(t|\mathbf{x}_0; t_0) = P[\mathcal{T} > (t - t_0) | \mathbf{X}(t_0) = \mathbf{x}_0]$ and satisfies the Kolmogorov backward equation given by

$$\frac{\partial}{\partial t_0} R(t|\mathbf{x}_0; t_0) = -\sum_{i=1}^d A_i(\mathbf{x}_0, t_0) \frac{\partial}{\partial x_{0i}} [R(t|\mathbf{x}_0; t_0)] - \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d [\mathbf{D}(\mathbf{x}_0, t_0) \mathbf{\Sigma} \mathbf{D}'(\mathbf{x}_0, t_0)]_{ij} \frac{\partial^2}{\partial x_{0i} \partial x_{0j}} [R(t|\mathbf{x}_0; t_0)]$$

$$R(t_0|\mathbf{x}_0; t_0) = 1 \text{ if } \mathbf{x}_0 \in \partial\Omega; R(t|\mathbf{x}_0; t_0) = \begin{cases} 0, & \text{if } \mathbf{x}_0 \in \partial\Omega \\ 1, & \text{if } \mathbf{x}_0, \mathbf{x}(t) \in \Omega \end{cases}$$

5. Equation governing moments of function of $\mathbf{X}(t)$:

$$\frac{\partial}{\partial t} \langle \psi(\mathbf{X}, t) \rangle = \sum_{i=1}^d \left\langle A_i(\mathbf{X}, t) \frac{\partial}{\partial x_{0i}} \psi(\mathbf{X}; t) \right\rangle + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \left\langle [\mathbf{D}(\mathbf{X}, t) \mathbf{\Sigma} \mathbf{D}'(\mathbf{X}, t)]_{ij} \frac{\partial^2}{\partial x_{0i} \partial x_{0j}} \psi(\mathbf{X}, t) \right\rangle + \left\langle \frac{\partial}{\partial t} \psi(\mathbf{X}, t) \right\rangle$$

$$\langle \psi(\mathbf{X}, 0) \rangle = \psi_0$$

If one considers moments of the form $\langle X_1^{n_1}(t) X_2^{n_2}(t) \cdots X_m^{n_m}(t) \rangle$, it can be shown that, for linear systems, the moment equations form a closed set of equations and can be solved at least numerically. For nonlinear systems, however, the moment equations form an infinite hierarchy of equations which at no stage provide sufficient number of equations to solve for the moments. This limits the

applicability of the moment equations for such systems.

6. Equation governing moments of first passage time:

The PDF of the survival time $\mathcal{T}$ can be obtained as $p_{\mathcal{T}}(t) = \frac{d}{dt} [1 - R(t|\mathbf{x}_0, t_0)]$. Accordingly, the governing equations for moments of first passage time, $m_l = \langle \mathcal{T}^l \rangle$, $l = 1, 2, \dots$, are given by

$$(l+1)m_l - \sum_{i=1}^d A_i(\mathbf{x}_0, t_0) \frac{\partial m_{l+1}}{\partial x_{0i}} - \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d [\mathbf{D}(\mathbf{x}, t) \mathbf{\Sigma} \mathbf{D}'(\mathbf{x}, t)]_{ij} \frac{\partial^2 m_{l+1}}{\partial x_{0i} \partial x_{0j}}; l = 1, 2, \dots$$

$$m_1 = 1; m_{l+1} = \begin{cases} 0, & \text{if } \mathbf{x}_0 \in \partial\Omega \\ < \infty, & \text{if } \mathbf{x}_0 \in \Omega \end{cases}$$

Table 1 lists a few cases in which exact solutions for some of the above equations can be obtained. It may be noted that the determination of $p(\mathbf{x}, t|\mathbf{x}_0, t_0)$ can, in principle, facilitate the determination of the joint PDF $p_{h\dot{h}}(h^*, \dot{h}^*; t)$ and, hence, to the subsequent determination of the probability $P[h_M \leq h^*] = P[T_f(h^*) \geq T]$.

Monte Carlo Simulation Methods

The main idea here is to obtain, on a computer, random samples of $\boldsymbol{\Theta}$ and $d\mathbf{B}(t)$ compatible with the prescribed joint PDF and estimate the probability of failure as

$$\hat{P}_F = \frac{1}{N_1} \sum_{i=1}^{N_1} I[h^* - h_M^i(\boldsymbol{\Theta}) \leq 0] \quad (5)$$

Here $h_M^i(\boldsymbol{\Theta})$ is the maximum value of the response metric of interest corresponding to the i^{th} realizations of $\boldsymbol{\Theta}$ and $d\mathbf{B}(t)$, and N_1 is the sample size. This calculation is facilitated by discretizing Eq. 4 to obtain

$$\mathbf{x}_{k+1} = \boldsymbol{\phi}_k(\mathbf{x}_k, \boldsymbol{\theta}, \mathbf{v}_k); \mathbf{x}_0 = \mathbf{X}_0 \quad (6)$$

Here $\mathbf{x}_k$, $\boldsymbol{\phi}_k$ are $r \times 1$ vectors, $\mathbf{v}_k$ is $n_v \times 1$ vector of Gaussian random variables, and $k = 0, 1, 2, \dots, N$ represents the discretized

Reliability Estimation and Analysis for Dynamical Systems, Table 1 Few dynamical systems for which exact solutions can be obtained through the Markov vector approach

S. no.	System	Example	Quantities which are exactly determinable
1	Linear time invariant systems under external white noise or filtered white noise excitations (Nigam 1983)	$\ddot{u} + 2\eta\omega\dot{u} + \omega^2u = f(t)$ $\dot{f} + 2\eta_1\omega_1\dot{f} + \omega_1^2f = w(t)$ $u(0) = u_0, \dot{u}(0) = \dot{u}_0; f(0) = f_0, \dot{f}(0) = \dot{f}_0$ $\mathbf{x} = (u, \dot{u}, f, \dot{f})^t$	$p(\mathbf{x}; t)$ $p(\mathbf{x}; t \mathbf{x}_0; t_0)$ $\langle x_1^m(t) x_2^n(t) \cdots \rangle$ $\forall t$
2	Linear time invariant systems under parametric and (or) external white noise excitations (Soong 1973)	$\ddot{u} + 2\eta\omega\dot{u}[1 + w_1(t)] + \omega^2u[1 + w_2(t)] = w_3(t)$ $u(0) = u_0, \dot{u}(0) = \dot{u}_0$ $\mathbf{x} = (u, \dot{u})^t$	Time evolution of $\langle x_1^m(t) x_2^n(t) \rangle$ moments by solving a closed set of moment equations for all t
3	Nonlinear SDOF systems with nonlinear damping and (or) stiffness properties under external and white noise excitations (Caughey and Payne 1967; Lin and Cai 1995)	$\ddot{u} + 2\eta\omega\dot{u} + \omega^2u + f(u) = w(t)$ $u(0) = u_0, \dot{u}(0) = \dot{u}_0, \mathbf{x} = (u, \dot{u})^t$ $\ddot{u} + (\alpha + \beta u^2)\dot{u} + \omega^2[1 + w_1(t)]u = w_2(t)$ $u(0) = u_0, \dot{u}(0) = \dot{u}_0, \mathbf{x} = (u, \dot{u})^t$	$\lim_{t \rightarrow \infty} p(\mathbf{x}; t \mathbf{x}_0; t_0)$
4	A class of nonlinear multi-degree of freedom systems under combined parametric and external white noise excitation (Lin and Cai 1995)	$\ddot{u}_i + h_i(u, \dot{u}) = g_{ij}(u, \dot{u})w_j(t)$ $i = 1, 2, \dots, N_x; j = 1, 2, \dots, m;$ $\mathbf{x} = (u_1, \dot{u}_1, \dots, u_n, \dot{u}_n)^t$ $\ddot{u}_1 + \pi\mu(K_{11}\dot{u}_1 + 2\beta K_{12}\dot{u}_2)$ $\quad + \frac{\partial}{\partial u_1}H(u_1, u_2) = w_1(t)$ $\ddot{u}_2 + \pi\mu[2(1 - \beta)K_{12}\dot{u}_1 + K_{22}\dot{u}_2]$ $\quad + \frac{\partial}{\partial u_2}H(u_1, u_2) = w_2(t)$ $\langle w_i(t)w_j^T(t + \tau) \rangle = 2\pi K_{ij}\delta(\tau); i, j = 1, 2$ $\mathbf{x} = (u_1, \dot{u}_1, u_2, \dot{u}_2)^t$	$\lim_{t \rightarrow \infty} p(\mathbf{x}; t \mathbf{x}_0; t_0)$

time instants such that $t_k = k\Delta t$ and $N\Delta t = T$ with $\mathbf{x}_k$ serving as a discrete approximation to the response vector $\mathbf{X}(t)$ at $t = t_k$. Methods for discretization of SDEs are extensively described in the book by Kloeden and Platen (1992). Introducing a performance function $G(\boldsymbol{\theta}, \mathbf{v}) = h^* - \max_{0 \leq k \leq N} h[\boldsymbol{\theta}, \mathbf{x}_k]$, where $\mathbf{v} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N)$, the failure probability is obtained as $P_F = P[G(\boldsymbol{\theta}, \mathbf{v}) \leq 0]$. Accordingly, the estimator $\hat{P}_F$ is deduced to be given by

$$\hat{P}_F = \frac{1}{N_1} \sum_{i=1}^{N_1} I[G(\boldsymbol{\theta}^i, \mathbf{v}^i) \leq 0] \quad (7)$$

It can be shown that $\hat{P}_F$ is an unbiased and consistent estimator of the probability of failure with minimum variance for a specified N_1 . Furthermore, the sampling variance here is given

by $\text{Var}(\hat{P}_F) = P_F(1 - P_F)/N_1$. Accordingly, an increase in sample size N_1 is needed to reduce the sampling variance. The accuracy of the estimate obtained is quantified through its coefficient of variation defined as $\delta = \sqrt{(1 - P_F)/(N_1 P_F)}$. Thus, for instance, for $P_F = 10^{-5}$, and a target value of δ of 5 %, one needs about 4×10^7 samples to estimate P_F . The requirement on the number of samples increases to about 10^9 when δ is limited to 1 %. Given that the evaluation of $G(\boldsymbol{\theta}^i, \mathbf{v}^i)$, for each i , typically requires the execution of a finite element code, the evaluation of P_F using Eq. 7 is still not practically feasible. A way forward here is to devise an estimator for P_F in which the sampling variance can be controlled without taking recourse to increase in sample size. This class of methods is broadly termed as variance reduction methods and has been widely studied in the existing literature. The main idea here is to utilize known information

about the model to improve on the performance of the estimator. A few of the developments in this context, in the area of time variant structural reliability modeling, are outlined in the following sections.

Importance Sampling

To simplify the formulation, the basic random variables $(\boldsymbol{\theta}, \mathbf{v})$ are transformed into space of standard normal random variables denoted by $\mathbf{U} = (\mathbf{U}_{\boldsymbol{\theta}}, \mathbf{U}_{\mathbf{v}})$ where $\mathbf{U}_{\boldsymbol{\theta}}$ and $\mathbf{U}_{\mathbf{v}}$ are the vectors of standard normal random variables originating from $\boldsymbol{\theta}$ and $\mathbf{v}$, respectively, with the understanding that the transformations $\boldsymbol{\theta} = T_{\boldsymbol{\theta}}(\mathbf{U}_{\boldsymbol{\theta}})$ and $\mathbf{v} = T_{\mathbf{v}}(\mathbf{U}_{\mathbf{v}})$ exist (Melchers 1999). The performance function in terms of the standard normal variable vector $\mathbf{U}$ is denoted by $G(\mathbf{U}) = h^* - \max_{0 \leq k \leq N} \underline{h}(\mathbf{U}_{\boldsymbol{\theta}}, \mathbf{x}_k)$ where $\underline{h}(\mathbf{U}_{\boldsymbol{\theta}}, \mathbf{x}_k) = h[T_{\boldsymbol{\theta}}(\mathbf{U}_{\boldsymbol{\theta}}), \mathbf{x}_k]$, $k = 0, 1, \dots, N$. Instead of drawing samples from $p_{\mathbf{U}}(\mathbf{u})$, one could also select an importance sampling PDF $\pi_{\mathbf{U}}(\mathbf{u})$ and evaluate P_F using $P_F = \int_{-\infty}^{\infty} I[G(\mathbf{u}) \leq 0] \frac{p_{\mathbf{U}}(\mathbf{u})}{\pi_{\mathbf{U}}(\mathbf{u})} \pi_{\mathbf{U}}(\mathbf{u}) d\mathbf{u}$. The estimator $\hat{P}_F$ can now be modified to read as

$$\hat{P}_F^{IS} = \frac{1}{N_2} \sum_{i=1}^{N_2} \frac{p_{\mathbf{U}}(\mathbf{u}^i)}{\pi_{\mathbf{U}}(\mathbf{u}^i)} I[G(\mathbf{u}^i) \leq 0] \quad (8)$$

Here $\mathbf{u}^i$, $i = 1, 2, \dots, N_2$ are random samples drawn from $\pi_{\mathbf{U}}(\mathbf{u})$. It can be shown that the above estimator for P_F is unbiased and has the sampling variance given by

$$\text{Var}(\hat{P}_F) = \left\langle \left\{ I[G(\mathbf{U}) \leq 0] \frac{p_{\mathbf{U}}(\mathbf{U})}{\pi_{\mathbf{U}}(\mathbf{U})} - P_F \right\}^2 \right\rangle \quad (9)$$

The yet undetermined $\pi_{\mathbf{U}}(\mathbf{u})$ is now chosen such that the above sampling variance is minimized. It can be verified that for $\pi_{\mathbf{U}}(\mathbf{u}) = I[G(\mathbf{u}) \leq 0] p_{\mathbf{U}}(\mathbf{u}) / P_F$, the sampling variance becomes zero. This PDF is known as the ideal importance sampling PDF (ISPDF). As is evident, the determination of this ideal ISPDF

requires the prior knowledge of P_F , the very quantity that is deemed as unknown in the first place. This obviously indicates the limitation on the utility of ideal ISPDF to actually determine P_F . On the other hand, the significance of ideal ISPDF lies in the fact that it actually exists, and one could aim to construct suboptimal ISPDF which substantially reduces the sampling variance below that of the brute force Monte Carlo simulations. Thus, for example, the ISPDF could be obtained in standard normal space as a multivariate Gaussian PDF with mean placed at the design point (Schueller and Stix 1987) which has been found through a prior step based on the first-order reliability method (Melchers 1999).

Subset Simulation

The main idea here is to represent the failure region, Ψ , as a limit of decreasing sequence of failure events, Ψ_i , $i = 1, 2, \dots, M$, such that $\Psi_1 \supset \Psi_2 \supset \dots \supset \Psi_M = \Psi$ and $\Psi_k = \bigcap_{i=1}^k \Psi_i$, $k = 2, \dots, M$, so that the failure probability can be expressed as $P_F = P(\Psi_1) \prod_{i=1}^{M-1} P(\Psi_{i+1} | \Psi_i)$ (Au and Beck 2001). The simulation effort is focused on delineating the regions Ψ_i , $i = 1, 2, \dots, M$ and in estimating the probabilities $P(\Psi_1)$ and $P(\Psi_{i+1} | \Psi_i)$, $i = 1, 2, \dots, M - 1$. The effectiveness of the method is due to the fact that each of these probabilities is much larger than P_F and hence can be estimated reliably with fewer samples. Sampling strategies based on the application of Markov Chain Monte Carlo method have been adopted in this context. Here the conditional simulation is carried out using a modified version of the Metropolis Hastings algorithm.

Trajectory Splitting Methods

By considering the parameter vector, $\boldsymbol{\theta}$, to be a deterministic quantity, Eq. 4 can now be rewritten in a simpler form as

$$\begin{aligned} d\mathbf{X}(t) &= \mathbf{A}[\mathbf{X}(t), t] dt \\ &\quad + \mathbf{D}[\mathbf{X}(t), t] d\mathbf{B}(t); \mathbf{X}(0) \\ &= \mathbf{X}_0 \end{aligned} \quad (10)$$

The probability of failure is obtained as

$$P_F = P\left\{h^* - \max_{0 \leq t \leq T} h[X(t), t] \leq 0\right\} \\ = \left\langle I\left(h^* - \max_{0 \leq t \leq T} h[X(t), t] \leq 0\right) \right\rangle \quad (11)$$

The trajectory splitting methods are a class of methods that are based on driving response trajectories in state-space toward failure regions. The idea here is to associate with each trajectory a weight, and judge each of the trajectories by their proximity to the failure region.

For the purpose of illustration, consider the response of interest to be denoted by $Y(t)$. An approximation to the cumulative distribution function (CDF) of this response quantity, denoted by $P(y; t) = P[Y(t) \leq y]$, is written as

$$P(y; t) = \sum_{i=1}^{N_{dc}} I[Y^i(t) \leq y] W^i(t) \quad (12)$$

Here $Y^i(t)$, $i = 1, 2, \dots, N_{dc}$ are a set of sample realizations of the random process $Y(t)$, which is governed by an SDE of the form given in Eq. 10, and the weights $W^i(t)$, $i = 1, 2, \dots, N_{dc}$ have the following properties:

$$W^i(t) \geq 0, \quad \sum_{i=1}^{N_{dc}} W^i(t) = 1 \quad (13)$$

Furthermore, at any time instant t , the importance of each sample realization is quantified in terms of a heuristic energy term, $C^i(t) = [E_{PE}^i(t) + E_{KE}^i(t)] P_e(t) W^i(t)^{\bar{\beta}}$. Here $E_{PE}^i(t)$ is the potential energy, $E_{KE}^i(t)$ is the kinetic energy of the system, $P_e(t) = F(t)^T \dot{Y}(t)$ is the input power of the external excitation where $F(t)$ is the external force, $\dot{Y}(t)$ is the velocity of the response, and $\bar{\beta}$ is an importance parameter with $0.1 \leq \bar{\beta} \leq 0.5$.

In the double and clump method (Pradlwarter and Schueller 1999), the splitting of the most

favorable trajectories is accompanied by clumping of the least favored samples. Trajectories whose energy, $C(t)$, is estimated to be high are replicated, and the trajectories with lesser energy are clumped together. In every case, the interference with ensemble of samples is accompanied by suitable adjustment of associated weights so that the estimator under consideration remains unbiased.

In the Russian roulette and splitting method (Pradlwarter and Schueller 1999), the least favored trajectories are eliminated from further evolution, and those trajectories which evolve toward the failure region are split and multiplied thereby enhancing the number of samples moving toward the failure region. Here, every realization, $Y^i(t)$, $i = 1, 2, \dots, N_{dc}$, is allowed to take part in further simulations with a probability $p^i(t)$ and cease to exist with probability $1 - p^i(t)$. The expression for the CDF, given in Eq. 12, is modified by considering the additional survival probability terms as

$$P(y; t) = \sum_{i=1}^{N_{dc}} I[Y^i(t) \leq y] J^i(t) \bar{W}^i(t) \quad (14)$$

Here $J^i(t)$, $i = 1, 2, \dots, N_{dc}$ is a Bernoulli random variable such that $P[J^i(t) = 1] = p^i(t)$ and $P[J^i(t) = 0] = 1 - p^i(t)$, and $\bar{W}^i(t)$, $i = 1, 2, \dots, N_{dc}$ denote the weights after implementing the Russian roulette and are taken to be given by $W^i(t)/p^i(t)$ so that the estimator given in Eq. 14 remains unbiased. Also, the survival probability $p^i(t)$ is taken to be proportional to the quantity $1 - \{[C_m(t) - C^i(t)]/C_m(t)\}^2$ where $C_m(t) = \max_{1 \leq i \leq N_{dc}} C^i(t)$.

The selection of the importance descriptor, $C(t)$, based on the notion of distance between samples led to the development of a method called distance-controlled Monte Carlo simulation (Pradlwarter and Schueller 1999). Here the trajectories are further manipulated using the Russian roulette with splitting strategy so as to make the realizations to be uniformly distributed in space.

Some variants of the subset simulation method, which are based on splitting of trajectories, are as follows:

1. Subset simulation with splitting (Ching et al. 2005a): The trajectories here, after reaching the intermediate threshold level, are split into offsprings from the time instant they reach the intermediate threshold value, thus resulting in the new trajectories which always lie in the conditional failure region. This can be explained as follows:

Let t_m be the time instant when a trajectory $x(t)$ reaches the intermediate threshold level, α . This time history is represented as $x(t) = x^-(t)$ for $t \leq t_m$, and $x(t) = x^+(t)$ for $t > t_m$. The corresponding excitations are denoted by $u^-(t)$ and $u^+(t)$, respectively. According to the splitting algorithm, a response trajectory, $\tilde{x}(t)$, always lie in the intermediate failure region, provided the excitation is given as $\tilde{u}(t) = [u^-(t), \tilde{u}^+(t)]$. Here $\tilde{u}^+(t)$, $t > t_m$ is generated using Monte Carlo simulation according to the PDF $p[u^+(t)|u^-(t)] = p[u^+(t), u^-(t)]/p[u^-(t)]$.

2. The advantages of the original subset simulation (Au and Beck 2001) and subset simulation with splitting (Ching et al. 2005a) have been combined to result in a new methodology known as hybrid subset simulation (Ching et al. 2005b). Here the original subset simulation is made use of till the trajectories reach the intermediate threshold levels, and there onwards the splitting algorithms takes over.

The Girsanov Transformation-Based Variance Reduction Technique

In simulation studies on Markovian systems governed by Ito's SDEs of the form given in Eq. 10, an artificial control force can be introduced into the governing equation so as to achieve reduction in Monte Carlo sampling variance associated with desired response statistics (Kloeden and Platen 1992). The process of replacing a given SDE by a modified equation with artificial controls in this context is called the Girsanov transformation. The control force and

the estimators are formulated so that the estimate remains unbiased while at the same time striving to reduce the sampling variance. To illustrate this, consider the modified version of the SDE given in Eq. 10,

$$\begin{aligned} d\tilde{X}(t) &= A[\tilde{X}(t), t]dt + D[\tilde{X}(t), t]u(t)dt \\ &\quad + D[\tilde{X}(t), t]dB(t)d\Gamma(t) \\ &= -\Gamma(t)u'(t)dB(t)\tilde{X}(0) \\ &= X_0, \Gamma(0) = \Gamma_0 \end{aligned} \quad (15)$$

Here $u(t)$ is the additional control force, $\Gamma(t)$ is a scalar correction term introduced to account for the addition of controls, Γ_0 is the initial condition on $\Gamma(t)$ which can be taken to be deterministic, and $\tilde{X}(t)$ is the biased state corresponding to the modified excitation $D[\tilde{X}(t), t]u(t) + D[\tilde{X}(t), t]dB(t)$. It can be shown that (Grigoriu 2002; Macke and Bucher 2003)

$$\begin{aligned} &\left\langle I\left(h^* - \max_{0 \leq t \leq T} h[X(t)] \leq 0\right) \right\rangle \\ &= \left\langle \Gamma(T)I\left(h^* - \max_{0 \leq t \leq T} h[\tilde{X}(t)] \leq 0\right) \right\rangle / \Gamma_0 \end{aligned} \quad (16)$$

Here Γ_0 has been taken to be deterministic. An estimator for evaluating the expression on the right hand side can be obtained as

$$\tilde{P}_F = \frac{1}{N_2 \Gamma_0} \sum_{i=1}^{N_2} \Gamma(T) I\left\{h^* - \max_{0 \leq t \leq T} h[\tilde{X}^i(t), t] \leq 0\right\} \quad (17)$$

where $\tilde{X}^i(t)$ are random draws from Eq. 15. It follows that $\langle \tilde{P}_F \rangle = P_F$ and $\text{Var}(\tilde{P}_F)$ will be dependent on N_2 and the yet to be determined control vector $u(t)$. The control $u(t)$ is now selected such that $\text{Var}(\tilde{P}_F) \ll \text{Var}(\hat{P}_F)$. It can be shown that an ideal control $u^*(t)$ exists which yields $\text{Var}(\tilde{P}_F) = 0$, but its construction requires the knowledge of P_F , the very quantity being sought to be determined. The way forward would be to seek a suboptimal control $u(t)$ which helps to reduce the sampling variance

appreciably and not to seek the ideal situation of obtaining $\text{Var}(\hat{P}_F) = 0$. Here an associated dynamical system

$$\begin{aligned} dV(t) &= A[V(t), t]dt + D[V(t), t]u(t)dt; 0 \leq t \leq T \\ V(0) &= X_0 \end{aligned} \quad (18)$$

is considered. The control $u(t)$ is determined such that it minimizes the distance function given by

$$\beta(\tau^m) = \sqrt{\sum_{j=1}^q \int_0^{\tau^m} u_j^2(t)dt} \text{ subject to the constraint } h^* - h[V(\tau^m)] = 0 \text{ with } 0 < \tau^m \leq T.$$

For linear systems and linear performance functions, the suboptimal control force can be obtained as a function of impulse response functions reversed in time. Also from Eq. 15, it can be noted that the evaluation of the correction process, also known as the Radon–Nikodym derivative, is independent of the mathematical model for the structure under study. Thus, an acceptable choice for the control force can be made solely based on experimental techniques, and the estimator for the reliability can be deduced without taking further recourse to mathematical model for the structure under study. This permits the application of the Girsanov transformation-based variance reduction technique in the experimental study of time variant reliability of complex structural systems which are difficult to model mathematically (Sundar and Manohar 2014a).

Different variance reduction strategies can be combined into a single framework so as to leverage advantages associated with each one of them. Thus, for example, the subset simulation method and the Girsanov transformation-based method can be combined in such a way that subset simulations handle the uncertainties associated with parameters, and the Girsanov transformation takes care of the random excitations (Sundar and Manohar 2014b). Performance of few simulation-based variance reduction techniques with respect to few benchmark problems has been documented by Schueller and Pradlwarter (2007).

System Reliability

Structural systems typically consist of several components and are liable to fail in more than one way due to the failure of all or some of the components. Reliability analysis of such systems is carried out by defining a performance function corresponding to each failure element and identifying the possible modes that lead to the failure of the system as a whole. Unlike the performance function considered in the preceding sections, the performance indicators to vector-valued functions present a more realistic case whereby the failure of the system can be modeled by series systems (weakest link), parallel systems (exceedance of all the limit states), or more generally a combination of both. In the time-dependent reliability framework, it leads to the outcrossing problem of the vector-valued stochastic process whose elements are the performance functions.

To illustrate this, consider the performance metrics denoted by $h_{ij}[X(t), \Theta, t]$, $i = 1, 2, \dots, n_1$; $j = 1, 2, \dots, n_2$. The problem of system reliability is written as

$$P_S = 1 - P_F = 1 - P \left[\bigcup_{i=1}^{n_1} \bigcap_{j=1}^{n_2} (h_{ij}^* - h_{ij}^M \leq 0) \right] \quad (19)$$

where $h_{ij}^M = \max_{t_0 \leq t \leq T} h_{ij}[X(t), \Theta, t]$. Here the structural system consists of n_1 , the number of subsystems connected in series, and each subsystem consists of n_2 , the number of components connected in parallel. A Monte Carlo estimator for the problem on hand is given as

$$\hat{P}_F = \frac{1}{N_1} \sum_{i=1}^{N_1} I \left[\bigcup_{i=1}^{n_1} \bigcap_{j=1}^{n_2} (h_{ij}^* - h_{ij}^M \leq 0) \right] \quad (20)$$

Studies on joint distributions of level crossing counting processes, first passage times, and extremes for vector of Gaussian random processes have been carried by Gupta and Manohar (2005) and Song and Der Kiureghian (2006). The study by Gupta and Manohar (2005) employs multivariate Poisson counting process model

and deduce joint distribution of extreme values as multivariate distributions with Gumbel marginals. These studies represent efforts to analytically model reliability specified with respect to more than one performance functions.

Existing Structures

Once an engineering system comes into existence, it becomes possible to measure its performance under operating and (or) ambient loads. In such cases, it becomes relevant to ask how the time variant reliability models can be updated based on observed performance of the system. Analysis of reliability of existing structures has been the subject of several studies in the existing literature (see, e.g., Melchers 1999).

The first natural step in the study of instrumented structures consists of identifying the parameters of the postulated mathematical model. Beyond the step of system identification, there exist several questions related to mathematical model updating such as reliability and local/global sensitivity model updating. The problem of system identification can be carried out within the framework of the dynamic state estimation methods (Ristic et al. 2004), and this can further be extended to problems of updating time variant reliability models of nonlinear dynamical systems (Ching and Beck 2007; Radhika and Manohar 2010; Sundar and Manohar 2013). In the study by Sundar and Manohar (2013), the performance functions considered are taken to be nonlinear functions of system states which are not necessarily measured. The analysis consists of two ingredients: the first that involves estimation of system states and performance metric conditioned on measurements and the second that involves estimation of reliability against future random excitations. The template for this framework consists of the following components:

1. Mathematical model for the structure, typically based on finite element analysis
2. A set of noisy measurements on structural displacements, strains, applied loads, and

(or) reaction transferred to the supports under operating and (or) diagnostic loads

3. A mathematical model which relates measured quantities to the system states in the governing mathematical model for structure. Both the models for structural behavior and the measurements are taken to be imperfect, and this is accounted for by including appropriate random noise terms in the models.

To illustrate this, consider the dynamical system governed by the discretized equation of the form

$$\begin{aligned} \mathbf{x}_{k+1} &= \boldsymbol{\phi}_k(\mathbf{x}_k, \boldsymbol{\theta}, \mathbf{f}_k, \mathbf{w}_k), \\ k &= 0, 1, \dots, N, \mathbf{x}_0 = \mathbf{x}(0) \end{aligned} \quad (21)$$

Here $\boldsymbol{\theta}$ is a $p \times 1$ vector of unknown system parameters, $\mathbf{x}_k$ and $\mathbf{f}_k$ are the $r \times 1$ state and externally applied force vectors in the time discretized form, $\boldsymbol{\phi}_k$ is a nonlinear state transition vector, and $\mathbf{w}_k$ is the $n_w \times 1$ process noise which represents the error in arriving at mathematical model for the vibrating system, modeled as a sequence of zero-mean Gaussian random variables with known covariance, i.e., $\langle \mathbf{w}_k \mathbf{w}_j^t \rangle = \mathbf{Q}_k \delta_{kj}$ where $\mathbf{Q}_k$ is $n_w \times n_w$, and δ_{kj} denotes the Kronecker delta function. Clearly the response vector, $\mathbf{x}_k$ would possess Markovian property.

The structure is assumed to be instrumented with s number of sensors which measure time histories of structural response, $\mathbf{Z}_k$, $k = 1, 2, \dots, N$. Let the measured response be related to the system states, $\mathbf{x}_k$, through the equation

$$\mathbf{Z}_k = \mathbf{H}_k(\mathbf{x}_k, \boldsymbol{\theta}, \mathbf{v}_k), \quad k = 1, 2, \dots, N \quad (22)$$

Here $\mathbf{Z}_k$ is an $s \times 1$ vector of measurements from the s sensors; $\mathbf{v}_k$ is the $n_v \times 1$ measurement noise term, modeled as a sequence of zero-mean Gaussian random variables with $\langle \mathbf{v}_k \mathbf{v}_j^t \rangle = \mathbf{R}_k \delta_{kj}$ where $\mathbf{R}_k$ is $n_v \times n_v$ covariance matrix; and $\mathbf{H}_k$ is a nonlinear function that relates the measurements to system states through a pertinent mathematical model. The measurement noise arises inherently

due to the sensor noise and also takes into account the error associated in the mathematical model used to relate the system states to the measurements made. It may be noted that $\mathbf{Z}_k$, $k = 1, 2, \dots, N$ could include measurements on applied forces, $\mathbf{f}_k$, as well. The notations $\mathbf{x}_{0:k} = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k\}$, $\mathbf{f}_{1:k} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_k\}$, and $\mathbf{Z}_{1:k} = \{\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_k\}$ will be used to represent the state, force, and response measurement matrices, respectively. For the sake of simplicity, the time step used in time discretization of governing equation and the sampling step used in obtaining measurements are taken to coincide.

Equation 22 represents the additional information that has been gained by making measurements on structural behavior. If attention is limited to Eq. 21 alone, as would be done at the design stage of the structure before it had come into existence, one can obtain response descriptions such as the multivariate joint PDF, $p_{\mathbf{x}_{0:k}}(\mathbf{x}_{0:k})$, and measures of structural safety based on suitably defined performance functions, $g(\boldsymbol{\theta}, \mathbf{x}_k)$. In such evaluations, one would employ models for the PDF of the system parameters, $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$, and the external excitation, $p_{\mathbf{f}_{1:k}}(\mathbf{f}_{1:k})$, based on analysis of available data and engineering judgment.

Once the structure is constructed and measurements become available, the problem of updating reliability models can be posed as the problem of determination of reliability of the structural system conditioned on $\mathbf{Z}_{1:k}$, i.e., $P_S|\mathbf{Z}_{1:k} = 1 - P_F|\mathbf{Z}_{1:k}$, where $P_F|\mathbf{Z}_{1:k}$ is the posterior probability of failure. In future discussions, for the sake of simplicity, subscripts in denoting PDFs are omitted. For instance, $p_{\mathbf{x}_{0:k}}(\mathbf{x}_{0:k}|\mathbf{Z}_{1:k})$ would be written simply as $p(\mathbf{x}_{0:k}|\mathbf{Z}_{1:k})$.

Dynamic State Estimation

Dynamic state estimation technique provides a logical framework for assimilating the measurement given in Eq. 22 into the mathematical model expressed as Eq. 21. It is assumed here that the quantities $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ and $p(\mathbf{Z}_{k+1}|\mathbf{x}_k)$ can be determined from the knowledge of $\mathbf{w}_k$ and $\mathbf{v}_k$. The problem of dynamic state estimation is stated as determining the multivariate posterior PDF $p(\mathbf{x}_{0:k}|\mathbf{Z}_{1:k})$ and its marginal PDF $p(\mathbf{x}_k|\mathbf{Z}_{1:k})$. The basic idea here is to propagate the PDF of the state vector, $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$, from time step k to $k+1$ using the Markovian property of the vector $\mathbf{x}_k$ and Bayes' theorem to update the PDF of the states based on the available measurements. This can be expressed as a set of prediction and updating steps, given as follows:

$$\begin{aligned} \text{Prediction step : } p(\mathbf{x}_{k+1}|\mathbf{Z}_{1:k}) &= \int p(\mathbf{x}_{k+1}|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{Z}_{1:k})d\mathbf{x}_k \\ \text{Updating step : } p(\mathbf{x}_{k+1}|\mathbf{Z}_{1:k+1}) &= \frac{p(\mathbf{Z}_{k+1}|\mathbf{x}_{k+1})p(\mathbf{x}_{k+1}|\mathbf{Z}_{1:k})}{\int p(\mathbf{Z}_{k+1}|\mathbf{x}_{k+1})p(\mathbf{x}_{k+1}|\mathbf{Z}_{1:k})d\mathbf{x}_{k+1}}, \quad k = 0, 1, \dots, N \end{aligned} \quad (23)$$

Here $p(\mathbf{Z}_{k+1}|\mathbf{x}_{k+1})$ is called the likelihood function. The evolution of multidimensional

posterior PDF, $p(\mathbf{x}_{0:k+1}|\mathbf{Z}_{1:k+1})$, can be shown to be governed by the recursive relation

$$p(\mathbf{x}_{0:k+1}|\mathbf{Z}_{1:k+1}) = p(\mathbf{x}_{0:k}|\mathbf{Z}_{1:k}) \frac{p(\mathbf{Z}_{k+1}|\mathbf{x}_{k+1})p(\mathbf{x}_{k+1}|\mathbf{x}_k)}{p(\mathbf{Z}_{k+1}|\mathbf{Z}_{1:k})}, \quad k = 0, 1, \dots, N \quad (24)$$

For linear state-space models with Gaussian additive noise, the state estimation problem can be tackled exactly using Kalman filter, while for

more general class of nonlinear, non-Gaussian systems, particle filtering methods are employed. The second step involving reliability prediction

for both these classes of problems typically requires the application of Monte Carlo strategies.

Updating Reliability Models

Equations for the evolution of the updated PDF (i.e., after assimilating the measurements) similar to the ones given in section “[Markov Vector Methods](#)” can be derived. These are called the Kushner–Stratonovich equations. For most practical problems of interest, with nonlinear process and measurement equations, non-Gaussian noises, this equation remains theoretical in nature and suffers from the moment closure problem (Maybeck 1982). Thus updating reliability models is mostly carried out through the simulation-based methods.

The Kalman filter-based dynamic state estimation tools in combination with Monte Carlo simulation methods can be employed to estimate probability of failure in instrumented structures with performance functions encompassing unmeasured system states (Ching and Beck 2007). The variance reduction strategies developed in the context of reliability analysis when applied in conjunction with the dynamic state estimation techniques could be used to determine the updated probability of failure of the structural system. For example, the data-based extreme value analysis and the Girsanov transformation-based method can be used to determine the reliability of existing structures (Radhika and Manohar 2010; Sundar and Manohar 2013).

Summary

The problem of time variant reliability analysis of nonlinear dynamical systems can be studied to a limited extent through analytical methods and more comprehensively with simulation-based strategies. The analytical approximations are based on theory of outcrossing statistics and Markov vector approaches. Monte Carlo simulation-based methods invariably need to be reinforced with suitable variance reduction strategies so that the resulting tools become applicable to tackle

realistic problems. These tools are versatile to handle various complexities associated with treatment of structural nonlinearities, mutually dependent non-Gaussian random variables, spatially varying random fields, non-stationary random vector excitations, and estimation of probability of rare events. When structure comes into existence and its responses are measured, the analysis of reliability can be carried out by combining Bayesian updating/filtering tools with simulation-based strategies to assess structural reliability against future loads. This entry provides a snapshot of current state of art in this subject.

Cross-References

- ▶ [Analytic Fragility and Limit States \[P\(EDPI IM\)\]: Nonlinear Dynamic Procedures](#)
- ▶ [Nonlinear Dynamic Seismic Analysis](#)
- ▶ [Probability Density Evolution Method in Stochastic Dynamics](#)
- ▶ [Random Process as Earthquake Motions](#)
- ▶ [Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach](#)
- ▶ [Reliability Estimation and Analysis](#)
- ▶ [Response Variability and Reliability of Structures](#)
- ▶ [Seismic Reliability Assessment, Alternative Methods for](#)
- ▶ [Sensitivity of First-Excursion Probabilities for Nonlinear Stochastic Dynamical Systems](#)
- ▶ [Stochastic Analysis of Nonlinear Systems](#)
- ▶ [Structural Reliability Estimation for Seismic Loading](#)
- ▶ [Structural Seismic Reliability Analysis](#)
- ▶ [Subset Simulation Method for Rare Event Estimation: An Introduction](#)

References

- Au SK, Beck JL (2001) Estimation of small failure probabilities in high dimensions by subset simulation. *Probab Eng Mech* 16:263–277
- Caughey TK, Payne HJ (1967) On the response of a class of self-excited oscillators to stochastic excitation. *Int J Non Linear Mech* 2:125–151

- Ching J, Beck JL (2007) Real-time reliability estimation for serviceability limit states in structures with uncertain dynamic excitation and incomplete output data. *Probab Eng Mech* 22:50–62
- Ching J, Au SK, Beck JL (2005a) Reliability estimation for dynamical systems subject to stochastic excitation using subset simulation and splitting. *Comput Methods Appl Mech Eng* 194:1557–1579
- Ching J, Beck JL, Au SK (2005b) Hybrid subset simulation method for reliability estimation of dynamical systems subject to stochastic excitation. *Probab Eng Mech* 20:199–214
- Grigoriu M (2002) *Stochastic calculus: applications in science and engineering*. Birkhauser, Boston
- Gupta S, Manohar CS (2005) Multivariate extreme value distributions for random vibration applications. *ASCE J Eng Mech* 131(7):712–720
- Kloeden PE, Platen E (1992) *Numerical solution of stochastic differential equations*. Springer, Berlin
- Leadbetter MR, Lindgren G, Rootzen H (1983) *Extremes and related properties of random sequences and processes*. Springer, New York
- Lin YK, Cai GQ (1995) *Probabilistic structural dynamics: advanced theory and applications*. McGraw-Hill, Singapore
- Macke M, Bucher C (2003) Importance sampling for randomly excited dynamical systems. *J Sound Vib* 268:269–290
- Maybeck PS (1982) *Stochastic models, estimation, and control*, vol 2. Academic, New York
- Melchers RE (1999) *Structural reliability analysis and prediction*, 2nd edn. Wiley, Chichester
- Nigam NC (1983) *Introduction to random vibrations*. The MIT Press, Cambridge, MA
- Pradlwarter HJ, Schueller GI (1999) Assessment of low probability events of dynamical systems by controlled Monte Carlo simulation. *Probab Eng Mech* 14:213–227
- Radhika B, Manohar CS (2010) Reliability models for existing structures based on dynamic state estimation and data based extreme value analysis. *Probab Eng Mech* 25:393–405
- Ristic B, Arulampalam S, Gordon N (2004) *Beyond the Kalman filter: particle filters for tracking applications*. Artech House, Boston
- Schueller GI, Pradlwarter HJ (2007) Benchmark study on reliability estimation in higher dimensions of structural systems – an overview. *Struct Saf* 29:167–182
- Schueller GI, Stix R (1987) A critical appraisal of methods to determine failure probabilities. *Struct Saf* 4:239–309
- Song J, Der Kiureghian A (2006) Joint first-passage probability and reliability of systems under stochastic excitation. *ASCE J Eng Mech* 132(1):65–77
- Soong TT (1973) *Random differential equations in science and engineering*. Academic, New York
- Soong TT, Grigoriu M (1993) *Random vibration of mechanical and structural systems*. Prentice Hall, New Jersey
- Sundar VS, Manohar CS (2013) Time variant reliability model updating in instrumented dynamical systems based on Girsanov's transformation. *Int J Non Linear Mech* 52:32–40
- Sundar VS, Manohar CS (2014a) Random vibration testing with controlled samples. *Struct Control Health Monit*. doi:10.1002/stc.1646
- Sundar VS, Manohar CS (2014b) Estimation of time variant reliability of randomly parametered non-linear vibrating systems. *Struct Saf* 47:59–66

Reliable Moment Tensor Inversion for Regional- to Local-Distance Earthquakes

Giuliano F. Panza^{1,2,3,4}, Enrico Brandmayr^{1,2} and Andrea Magrin¹

¹Department of Mathematics and Geosciences, University of Trieste, Trieste, Italy

²The Abdus Salam International Centre for Theoretical Physics, SAND Group, Trieste, Italy

³Institute of Geophysics, China Earthquake Administration, Beijing, China

⁴International Seismic Safety Organization (ISSO), Arsita, Italy

Synonyms

Inversion; Moment tensor; Nonlinear; Shallow events; Short period

Introduction

The reliable determination of earthquakes source moment tensor and of its uncertainty is of key importance for both geodynamic investigation and seismic hazard assessment.

The knowledge of the focal mechanism (i.e., fault plane solution) in fact provides essential indications about the kind of rupture process affecting the fault and thus, besides the event's strength (i.e., seismic moment and magnitude), gives a first-order description of the local stress field acting on the fault. This information can be used not only to constrain geodynamic

interpretations relying on seismological and geophysical data modeling but also as an input in deterministic seismic hazard analysis (DSHA or NDSHA, Panza et al. 2012), providing the depth of the fault and its style of dominant displacement. The information provided by moment tensor inversion may enter as an input even in the probabilistic seismic hazard analysis (PSHA), providing fault geometry and dislocation mechanism that critically affect ground motion prediction equations (GMPEs), and, if incorrectly determined, may lead to unreliable results.

Many different methods use intermediate- to long-period waveforms, from regional to global distances, and are routinely applied to obtain “real-time” focal mechanism determinations for moderate to strong earthquakes. The routinely applied centroid moment tensor (CMT, Dziewonski et al. 1981) is one of the most popular methodologies for the determination of moderate to strong ($M > 5.5$) event’s focal mechanism, even if it has been proven that supplied focal depths can be severely in error (Chu et al. 2009). This method uses filtered body waves at period >45 s for moderate events and mantle waves filtered at period >135 s for stronger ones. CMT is not useful to determine focal mechanisms of low to moderate magnitude events, which are not resolved by teleseismic waves. Thus, in the last decades, inversion methods were extended to the use of regional to local seismograms and routinely used wherever the presence of a broadband seismic network is available. The regional centroid moment tensor (RCMT, Arvidsson and Ekström 1998), for example, provides focal mechanism determinations for events of $M > 4.5$ in the Euro-Mediterranean region, using regional surface wave records filtered at period >35 s (60 s for strongest events). RCMT is routinely used since 1997, i.e., since the broadband network coverage allows the use of the quite high number of stations needed by the inversion method. In fact a reanalysis by means of RCMT method (Pondrelli et al. 2006) of events that occurred in the period 1977–1997 in the Italian region shows that a low number of recording stations as well as their sparse azimuthal distribution badly constrains the inversion results.

Both CMT and RCMT methods suffer the instability of two elements of the moment tensor (namely, M_{zx} and M_{zy}) that are poorly resolved when waves whose wavelength significantly exceeds the hypocentral depth are used (Bukchin et al. 2010). To overcome this problem, with CMT and RCMT, usually the depth of shallow events is fixed during the inversion, either at 10 or 15 km of depth. Instability also affects moment tensor determination itself and gives rise to spurious large compensated linear vector dipole (CLVD) components and, in some cases, to pairs of best double-couple (DC) solutions equally well fitting the recorded waveforms (Henry et al. 2002).

In the following, the INPAR nonlinear inversion method (Šílený et al. 1992) is described that is suitable for the retrieval of earthquakes moment tensor and of its uncertainty for small to moderate earthquakes ($1.5 < M_w < 6.0$). INPAR that can handle periods as short as 0.2 s has been successfully applied in tectonic as well as in geothermal and volcanic environment, where it proved its capability to retrieve reliable non-DC components (e.g., Saraò et al. 2010).

With respect to the most inversion methods, INPAR (a) requires the use of a very limited number of full waveforms at the shortest wavelengths compatible with the source dimension and epicentral distance and (b) has the capacity to absorb, at least partially, spurious effects due to the epistemic uncertainty of the models describing the geological structure where the source is buried. On account of properties (a) and (b), the availability of average models and of a few well-recorded signals is sufficient for its application as shown here, for example, in Italy, Vrancea, Antarctica, and Egypt.

The instability of M_{zx} and M_{zy} can be successfully mitigated by the use of short-period waveforms and, therefore, largely improves the reliability of the moment tensor determination of shallow events. In addition, the flexibility of the method allows sorting out a spurious CLVD component from that arising from a mechanism changing in time and gives deeper insight in the rupture process of complex events.

Methodology

The INPAR method for the inversion of moment tensor adopts a point-source approximation. The retrieval of the six components of the moment tensor by waveform inversion is a nonlinear problem; anyway linearity can be preserved in the first step of the inversion by considering different time evolutions for each of the six components of the moment tensor, namely, the moment tensor rate functions (MTRFs, Panza and Saraò 2000). The k_{th} component of displacement at the surface is the convolution product of the MTRFs and (medium) Green's function spatial derivatives (hereafter Green's functions) and, using Einstein summation notation, can be written as

$$u_k(t) = \dot{M}_{ij}(t) * G_{ki,j}(t) \quad (1)$$

The moment rate functions are obtained by deconvolution of Green's functions from the data. The synthetic Green's functions are computed using the modal-summation technique (Florsch et al. 1991) at each grid point of a preassigned volume around estimated hypocentral coordinates and, by linear interpolation, in the intermediate points of this volume according to an a priori defined step. Thus, INPAR performs if necessary a dynamic relocation of the hypocenter, which is crucial, since its bad location may strongly affect the result of the inversion. The hypocentral location is searched until the difference (in terms of L_2 -norm) between synthetic and observed seismograms is minimized with respect to a preassigned threshold.

Considering each MTRF as independent functions in the first step of inversion leads to an over-parameterization of the problem which is advantageous for absorbing poor modeling of the structure (epistemic uncertainty, Kravanja et al. 1999).

In the second step, the source mechanism and time function are obtained after factorization of the MTRFs into a time constant moment tensor m_{ij} and a common source time function $f(t)$:

$$\dot{M}_{ij}(t) = m_{ij} \cdot f(t) \quad (2)$$

This means that the same time dependence for all moment tensor components, i.e., a rupture mechanism constant in time, is assumed, which is an acceptable approximation for weak to moderate earthquakes. Only the correlated part of the MTRFs is used in the factorization, thus absorbing the bias due to non-exact Green's functions (Panza and Saraò 2000); the problem is nonlinear and is solved iteratively by imposing constraints such as positivity of the source time function and the requirement of a mechanism consistent with clear readings of at least one first-arrival polarity.

The final solution and related uncertainties are obtained by means of a genetic algorithm (Šílený 1998) which finds out the point where the misfit function is a minimum and at the same time maps the model space in its vicinity with the aim of estimating the confidence regions of the model parameters.

In summary, the average mechanism and source time function obtained by the inversion are considered to be basically affected by three kinds of bias, generated, respectively, by: (1) the noise present in the data. (2) the horizontal mislocation of the hypocenter, and (3) the epistemic uncertainty in the structural models used to compute the synthetic Green's functions.

Examples of Applications and Discussion

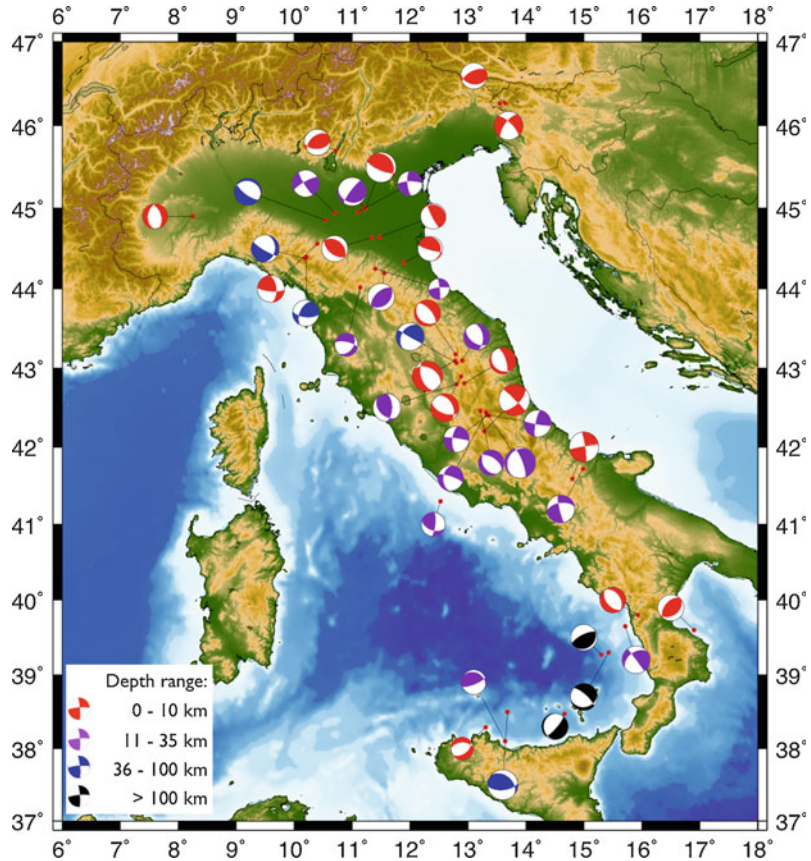
Application to Tectonic Events

The INPAR method has been applied for the retrieval of the moment tensor of tectonic events in different regions of the Earth, especially where the geographic features and logistic determine an uneven distribution of seismic stations (like in Italy, Antarctica, and Egypt).

In Italy INPAR has been routinely applied to determine the focal mechanisms of major, potentially damaging earthquakes ($M_w \geq 4.8$) since 1997. The comparison with the solution obtained using longer periods of data, performed by different authors (Guidarelli and Panza 2006; Brandmayr et al. 2013), points out differences in the hypocentral depth, especially in those cases in which depth is a priori fixed by CMT and RCMT.

Reliable Moment Tensor Inversion for Regional- to Local-Distance Earthquakes,

Fig. 1 Fault plane solution of the major events ($M_w \geq 4.8$) that affected Italy since 1997, determined by INPAR method. Different colors represent different source depth intervals



In some cases the relocation in depth is accompanied to a different fault plane solution, as a consequence of the improved resolvability of the M_{zx} and M_{zy} moment tensor components, allowed by the use in INPAR of much shorter periods than in CMT and RCMT.

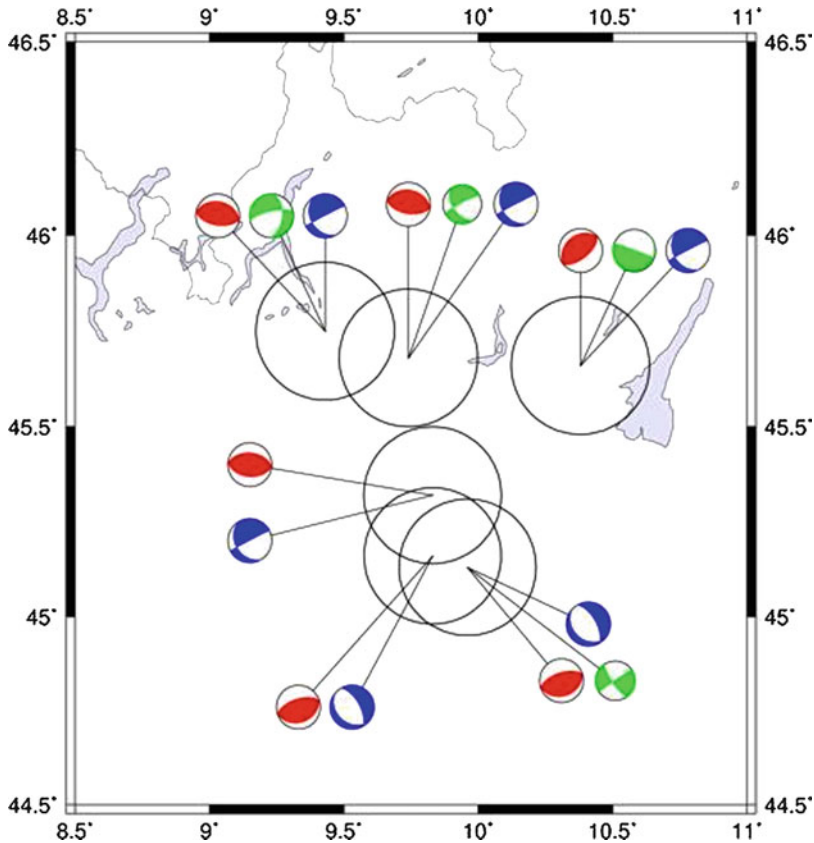
The synoptic view of all the events inverted by INPAR method (Fig. 1) is relevant for the understanding of the ongoing geodynamic process in the Italian region.

Most of the events occurred in central Apennines: two major seismic sequences hit Umbria and Marche regions (in 1997) and the city of L'Aquila (in 2009). Most of the hypocenters in this area are located in the crust (≤ 35 km of depth) and the prevalent fault mechanism is normal, in accordance with the ongoing extensional tectonics. Besides normal faulting, a relevant number of strike-slip events are present, likely representing zones of accommodation

of different extensional stress rates through the Apennines chain.

A more complex seismic pattern characterizes the northern part of the Apennines and the related Po Plain foredeep, where a major seismic sequence hit in 2012 (Brandmayr et al. 2013). Normal faulting prevails in the inner front, as shown by Pondrelli et al. (2006) in their review of the CMT and RCMT Italian dataset. Conversely, thrust to strike-slip faulting is dominant in the external front and on the buried faults below the Po Plain. Some lower crust earthquakes show a clear strike-slip mechanism, probably generated by buried faults located at the salient-recess interfaces.

Relevant intermediate (36–100 km) seismicity is present in the northernmost part of the Apennines, both on the inner and the external front. Northeastern Alps shows limited seismicity in the considered magnitude and time range, and the



Reliable Moment Tensor Inversion for Regional- to Local-Distance Earthquakes, Fig. 2 Focal mechanisms associated to the seismogenic nodes (circles) in the Po Valley from different sources: mean focal mechanism of seismogenic zone (Meletti and Valensise 2004) that contains the center of the node (blue), mean focal mechanism of events of EMMA database (Vannucci and

Gasparini 2004) that falls inside the node (green), focal mechanism of fault contained in Database of Individual Seismogenic Sources (DISS Working Group 2010) associated to the node (red). The reliability of the focal mechanism determination is crucial to define the average mechanism associated to each node and thus for realistic computation of hazard scenarios

observed shallow thrust to strike-slip mechanisms is fairly in agreement with the compressional tectonics of the area.

In southern Italy, major seismic activity is mostly concentrated along the Calabrian arc and the Tyrrhenian margin, with few events in the Ustica volcanic area. The only deep events (>100 km) in the Italian region are located in the inner front of the Calabrian arc, where the seismicity extends till 350 km of depth.

The reliable determination of the focal mechanisms also represents a key input for PSHA and NDSHA, to the latter in particular for the definition of the seismogenic nodes (Fig. 2).

Seismogenic nodes are earthquake-prone areas identified through morphostructural analysis (e.g., Gorshkov et al. 2002), and, in the framework of NDSHA, possible seismic sources are associated with them (Panza et al. 2012). Reliable focal mechanisms associated to nodes are crucial to compute the seismic hazard scenarios. Nevertheless, if a seismogenic node presents different associated mechanisms, parametric tests used in the NDSHA allow to compute different scenarios, in between which the worst can be taken as representative. Conversely, the way in which information about focal mechanism eventually enters in PSHA (i.e., as an average of different

mechanisms in the GMPE) badly affects the prediction scenario in case of unreliable moment tensor determinations.

Application of INPAR to tectonic events in Antarctica represents a suitable procedure in an environment such as the Scotia Sea region, where (1) the high level of seismic noise, typical of the oceanic environment, hampers the use of standard inversion techniques based on first-arrival polarities, (2) a low number of operating stations are present due to logistic, (3) the structural models are grossly known, and (4) the epistemic uncertainty is thus considerable.

Vuan et al. (2001) performed some feasibility tests, with particular attention to the distribution of the recording stations (i.e., their azimuthal coverage with respect to the source) for a synthetic event located at the South Sandwich Trench. The results show that a great improvement of the quality of the solution could be gained with the addition of only one ocean-bottom seismometer to the set of permanent seismic station deployed in 1997. The subsequent inversion of two real events, located by INPAR at depth of 90 and 15, respectively, is substantially in agreement with the results obtained by CMT. Guidarelli and Panza (2006) retrieved the moment tensor solutions of six events occurred in Bransfield Strait between 1997 and 1998, and Plasencia (2006) analyzed the seven major aftershocks of the Centenary Earthquake (2003/08/04, 04:37:19 GMT, 7.6 M_w), located along South Scotia Ridge.

The source mechanisms obtained are quite variable but consistent with the active tectonic processes and the complicated structure of the South Shetland Island region. The mechanisms around Elephant Island reflect the influence of the Elephant triple junction, with the intersection of the Shackleton Fracture Zone, the South Shetland Trench, and the South Scotia Ridge. In the forearc, north to the South Shetland Islands, the solutions obtained by Guidarelli and Panza (2006) seem to confirm the presence of active convergence along the South Shetland subduction zone, while Plasencia (2006) evidences the presence of both transpressive and transtensive

areas along the northern border of the South Orkney Microcontinent.

The inversion of local waveforms at high frequency (up to 5 Hz) by means of INPAR method has been applied both for shallow and deep weak to moderate tectonic events in Vrancea region and in eastern Carpathian bending zone (e.g., Ardeleanu et al. 2005).

The power of INPAR with respect to the traditional methods based on the inversion of polarities or amplitudes is obvious in the case of small earthquakes when few records are available and/or the noise distorts significantly or masks the first arrivals.

In these cases, the reliability of the solutions has been determined through a bootstrap-like procedure, consisting in inverting subsets of the available data combined with the rejection of obvious outliers, i.e., solutions that deviate strongly from the average. This procedure allowed limiting the distortion of source parameter due to the large epistemic uncertainty in the structural model of the region.

The estimated uncertainty of the resolved fault plane solutions for both shallow and intermediate depth events allowed reliable correlations with the stress field in the study area. The new data provided about the complex picture of the deformation field in the crustal and subcrustal domain of the Vrancea seismogenic region (a) indicate the transition from the compressive regime at intermediate depth to the extensional regime in the crust, characteristic over the Moesian platform, (b) point out the occurrence of events whose mechanism strongly deviates from the average mechanism of the area, and (c) show a large variability which could be caused by coseismic and postseismic relaxation associated with the Vrancea events generated in the subcrustal region.

Recently INPAR has been applied to a swarm of 15 small events ($1.7 < M < 3.7$) which occurred in August 2004 in the Abu Dabbab area (Egypt). The absence of a large seismic main shock, the periodically recorded swarms in the area, and the retrieved tensile earthquakes with a relatively high CLVD component are

considered evidence supporting igneous activity-related events (Ali et al. 2012).

Application to Volcanic and Geothermal Events

When dealing with seismic events in volcanic and geothermal environment, it is necessary to investigate full moment tensor, i.e., double-couple component (DC), compensated linear vector dipole component (CLVD), and isotropic component (V). In fact non-DC mechanisms in volcanic and geothermal environment can be indicators of local modifications of the stress fields induced by dikes injections as well as thermal cooling or high fluid pressure. It is thus crucial to sort out reliable non-DC components from the artificial ones that can arise from poor modeling of the structure, noise level, or inhomogeneous azimuthal station's coverage.

Panza and Saraò (2000) performed a systematic review of the possible sources of false non-DC components on low-magnitude events in volcanic and geothermal environment. They analyzed in particular events from Phlegraean Fields and Mt. Vesuvius, Larderello geothermal field (Tuscany), and Mt. Etna; spurious non-DC components can be identified with a confidence level between 95 % and 99 % through an appropriate error analysis like the one allowed by the use of a genetic algorithm (Šílený 1998), especially if the temporary evolution of the non-DC component is investigated as well. Reliable assessment of non-DC components in geothermal and volcanic environment is crucial for the geodynamic modeling of the magma or fluid sources, as shown by Saraò et al. (2010) who described, in the 2001 Mt Etna's eruption seismic swarm, an increase with increasing time of the non-DC component that is compatible with a complex stress regime naturally generated by vertical dike's propagation.

Summary

The reliable determination of earthquakes' moment tensor and of its uncertainty is of key importance for both geodynamic investigation

and seismic hazard assessment. Many different methods use intermediate to long period waveforms, from regional to global distances, and are routinely applied to obtain "real-time" focal mechanism determinations for moderate to strong earthquakes. A nonlinear inversion method is described that is suitable for the retrieval of earthquakes' moment tensor and of its uncertainty of small to moderate earthquakes ($1.5 < M_w < 6.0$), both in tectonic and volcanic environment, and that needs the use of a very limited number of full waveforms at the shortest compatible periods with the source dimension and epicentral distances.

The well-known instability of two elements of the moment tensor (M_{zx} and M_{zy}), which can give rise to very large spurious compensated linear vector dipole (CLVD) components and equally well-fitting pairs of solutions, is successfully mitigated by the use of short-period waveforms and, therefore, largely improves the reliability of the moment tensor determination of shallow events, which are affected by this instability when the relevant wavelength of the inverted signals significantly exceeds the source depth. In addition, the flexibility of the method allows to discriminate a spurious CLVD component from that arising from a mechanism changing in time and gives a deeper insight in the rupture process of complex events.

This method has been widely applied in different regions, whose seismic stations' spatial distribution often requires the use of few records like Italy, Vrancea, Antarctica, and Egypt, and some outstanding examples are given.

Cross-References

► [Probabilistic Seismic Hazard Models](#)

References

- Ali HA, El Hady S, Guidarelli M, Panza GF (2012) Source moment tensors of the earthquake swarm in Abu-Dabbab area, south-east Egypt. *Rend Fis Acc Lincei* 23:149–163. doi:10.1007/s12210-011-0158-9

- Ardeleanu L, Radulian M, Šílený J, Panza GF (2005) Source parameters of weak crustal earthquakes of the Vrancea region from short-period waveform inversion. *Pure Appl Geophys* 162:495–513. doi:10.1007/s00024-004-2618-y
- Arvidsson R, Ekström G (1998) Global CMT analysis of moderate earthquakes $M_w \geq 4.5$ using intermediate period surface waves. *Bull Seismol Soc Am* 88:1003–1013
- Brandmayr E, Romanelli F, Panza GF (2013) Stability of fault plane solutions for the major N-Italy seismic events in 2012. *Tectonophysics* 608:525–529. doi:10.1016/j.tecto.2013.08.034
- Bukchin B, Clévéde E, Mostinskiy A (2010) Uncertainty of moment tensor determination from surface wave analysis for shallow earthquakes. *J Seismol* 14(3):601–614. doi:10.1007/s10950-009-91858
- Chu R, Zhu L, Helmberger DV (2009) Determination of earthquake focal depths and source time functions in central Asia using teleseismic P waveforms. *Geophys Res Lett* 36:L17317. doi:10.1029/2009GL039494
- DISS Working Group (2010) Database of Individual Seismogenic Sources (DISS), Version 3.1.1: a compilation of potential sources for earthquakes larger than M 5.5 in Italy and surrounding areas. <http://diss.rm.ingv.it/diss/>. © INGV 2010 – Istituto Nazionale di Geofisica e Vulcanologia – All rights reserved; doi:10.6092/INGV.IT-DISS3.1.1
- Dziewonski AM, Chou TA, Woodhouse JH (1981) Determination of earthquake source parameters from waveform data for studies of global and regional seismicity. *J Geophys Res* 86:2825–2852
- Florsch N, Fäh D, Suhadolc P, Panza GF (1991) Complete synthetic seismograms for high-frequency multimode SH-waves. *Pure Appl Geophys* 136:529–560
- Gorshkov AI, Panza GF, Soloviev AA, Aoudia A (2002) Morphostructural zonation and preliminary recognition of seismogenic nodes around the Adria margin in peninsular Italy and Sicily. *J Seismol Earthq Eng* 4(1):1–24
- Guidarelli M, Panza GF (2006) Determination of the seismic moment tensor for local events in the South Shetland Islands and Bransfield Strait. *Geophys J Int* 167:684–692. doi:10.1111/j.1365-246X.2006.02953.x
- Guidarelli M, Panza GF (2007) INPAR, CMT and RCMT seismic moment solutions compared for the strongest damaging events ($M \geq 4.8$) occurred in the Italian region in the last decade. In: Proceedings of the national academy of sciences called XL, memoirs of physical and natural sciences, 124°, XXX, t. I, 81–98 Roma
- Henry C, Woodhouse JH, Das S (2002) Stability of earthquake moment tensor inversions: effect of the double-couple constraint. *Tectonophysics* 356:115–124
- Kravanja S, Panza GF, Šílený J (1999) Robust retrieval of seismic point source time function. *Geophys J Int* 136:385–394
- Meletti C, Valensise G (2004) Zonazione sismogenetica ZS9 – App.2 al Rapporto Conclusivo. In: Gruppo di Lavoro MPS (ed) Redazione della mappa di pericolosità sismica prevista dall'Ordinanza PCM 3274 del 20 marzo 2003. (Rapporto Conclusivo per il Dipartimento della Protezione Civile, INGV, Milano-Roma 2004) 65 pp, 5 enclosures
- Panza GF, Saraò A (2000) Monitoring volcanic and geothermal areas by full seismic moment tensor inversion: are non-double-couple components always artefacts of modelling? *Geophys J Int* 143:353–364
- Panza GF, Mura CL, Peresan A, Romanelli F, Vaccari F (2012) Seismic hazard scenarios as preventive tools for a disaster resilient society. *Adv Geophys* 53:93–165
- Plasencia M (2006) Lithospheric characteristics and seismic sources in the Scotia Arc through waveform inversion. PhD thesis, University of Trieste
- Pondrelli S, Salimbeni S, Ekström G, Morelli A, Gasperini P, Vannucci G (2006) The Italian CMT dataset from 1977 to the present. *Phys Earth Planet Int* 159(3–4):286–303. doi:10.1016/j.pepi.2006.07.008
- Saraò A, Cocina O, Privitera E, Panza GF (2010) The dynamics of the 2001 Etna eruption as seen by full moment tensor analysis. *Geophys J Int* 181:951–965. doi:10.1111/j.1365-246X.2010.04547.x
- Šílený J (1998) Earthquake source parameters and their confidence regions by a genetic algorithm with a “memory”. *Geophys J Int* 134:228–242
- Šílený J, Panza GF, Campus P (1992) Waveform inversion for point source moment tensor retrieval with optimization of hypocentral depth and structural model. *Geophys J Int* 109(259):274
- Vannucci G, Gasperini P (2004) The new release of the database of Earthquake Mechanisms of the Mediterranean Area (EMMA version 2). *Ann Geophys* 47:303–327
- Vuan A, Russi M, Costa G, Panza GF (2001) Moment tensor waveform inversion in the sub-Antarctic Scotia Sea region: feasibility tests and preliminary results. *Terra Antarctica* 8(2):55–62

Remote Sensing in Seismology: An Overview

Salvatore Stramondo
National Earthquake Center, Remote Sensing
Lab, Istituto Nazionale di Geofisica
e Vulcanologia, Rome, Italy

Synonyms

Coseismic deformation; Differential interferometry; Earthquake source; Modeling; Surface displacement; Synthetic aperture radar

Introduction

This entry is focused on the use of remote sensing in seismology for coseismic deformation detection, measurement, and modeling. Today the Earth sciences have a wide range of instruments and sensors that can provide scientists with a novel capability to study the physical processes driving earthquakes, volcanic eruptions, landslides, land subsidence, etc. In particular since the last 20 years, satellite missions devoted to Earth observation aims have been increased in number and improved in performances. Satellites allow systematic observation over large areas, over a long time period, and with frequent sampling. Such properties make satellites suitable to be used for studying the surface effects of earthquakes, which can be summarized in the effects on buildings and manufactures, and surface displacements.

This overview cannot be exhaustive of all the domains where remote sensing is applied in seismology. Indeed it is focused on the use of a specific techniques applied on satellite microwave data, named synthetic aperture radar (SAR), and on the application of a SAR processing technique known as InSAR (interferometric SAR) and its variation DInSAR (differential InSAR). It is worth noting how a comprehensive review about the use of remote sensing to cover the whole area of seismological studies would request to go deeper in details about more complex approaches for processing. The scope of the overview is simply to provide useful hints and references to who is interested to more specific studies.

In this overview concerning the use of remote sensing for seismological applications, some topics are beyond the aim of the entry, so as the technical details of InSAR and the use of optical satellite images for damage assessment, being the purpose of further contributions and will be debated elsewhere in the “Remote Sensing” section of the encyclopedia.

Satellite DInSAR

Among satellite sensors synthetic aperture radar has played a key role in the geosciences. The first

ever civilian SAR systems is the NASA Seasat (1978) satellite, the first satellite designed for remote sensing of the Earth’s oceans with a SAR sensor and used L-band (24-cm wavelength) radar. The aim of Seasat was to demonstrate the feasibility of global satellite monitoring of oceanographic phenomena and to help determine the requirements for an operational ocean remote sensing satellite system. Seasat was also used to provide the first demonstration of DInSAR to detect surface motion due to soil swelling (Fig. 1) in the Imperial Valley, California (Gabriel et al. 1989). The authors simply but efficiently summarize the essential properties of DInSAR, affirming: “can measure accurately extremely small changes in terrain over the large swaths associated with SAR imaging, especially since the sensor can work at night and through clouds or precipitations.”

Seasat was followed in 1991 by ERS-1 satellite launched by the European Space Agency (ESA), with a C-band (5.6-cm wavelength) SAR system, which was joined by ERS-2 a few years later (1995). In parallel, the National Space Development Agency of Japan (NASDA, now JAXA- Japan Aerospace Exploration Agency) developed an L-band SAR sensor that was launched into orbit onboard the JERS-1 satellite, in 1992. In 1995 the Canadian Space Agency (CSA) launched RADARSAT-1, an advanced Earth observation satellite (<http://www.asc-csa.gc.ca/eng/satellites/radarsat1/default.asp>), to monitor environmental change and to support resource sustainability. In 2002, ESA launched Envisat, an advanced polar-orbiting Earth observation satellite, which, among other instruments, operated an Advanced SAR (ASAR) system, a partially polarimetric instrument (capable of measuring different polarizations of the radar signal) with seven imaging swaths (I1–I7 modes), each with a different incidence angle, along with a wide-swath mode capable of producing 250-km-wide SAR images. In January 2006, JAXA launched the ALOS (Advanced Land Observing Satellite). ALOS carried optical instruments – PRISM and AVNIR – and the Phased-Array L-band Synthetic Aperture Radar (PALSAR). Starting in 2007, a new generation of

Remote Sensing in Seismology:

An Overview, Fig. 1 First demonstration of differential InSAR using Seasat data to detect vertical motions over a few days caused by soil swelling of irrigated fields in the Imperial Valley, California (From Gabriel et al. 1989). Differential interferogram (*top*) of the Imperial Valley and (*bottom*) double difference interferogram obtained by combining two interferograms as showed in the *top*. *Black areas* correspond to coherence loss

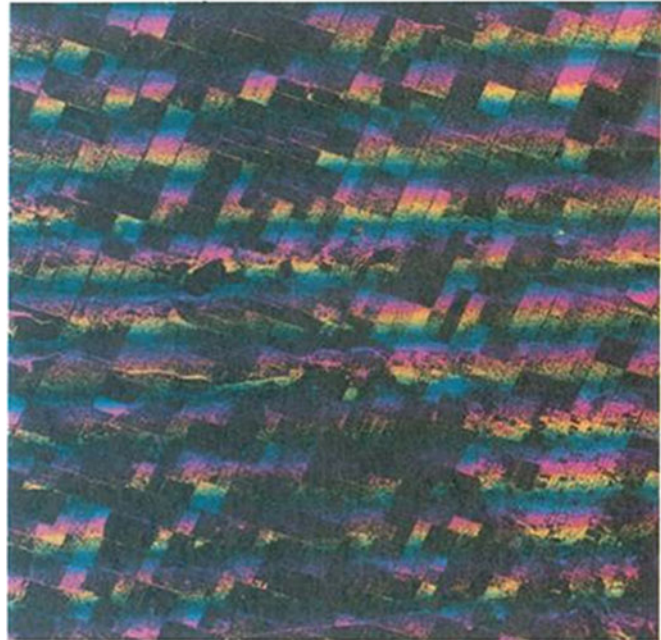


Plate 1a. Single-difference interferogram of Imperial Valley. Phases are presented as color and are shown. The fields with noisy phases had apparently been plowed between observations, resulting in a loss of phase coherence.

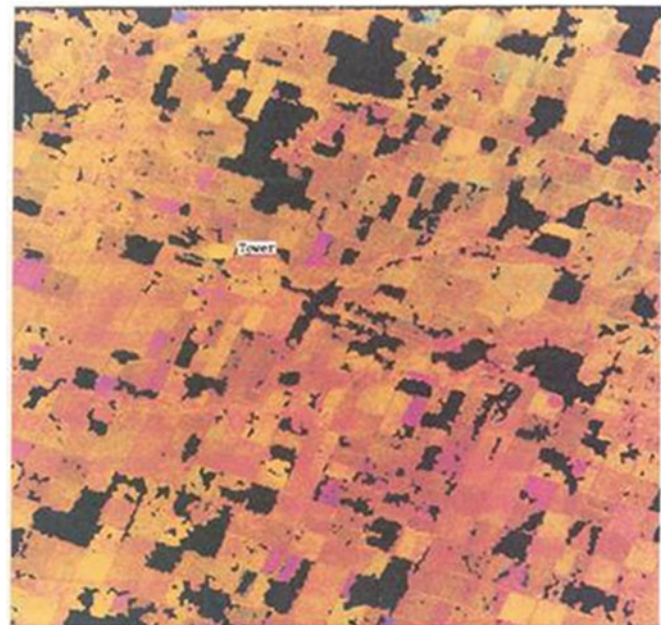


Plate 2. Double-difference interferogram of the Imperial Valley. Two interferograms like Plate 1 were differenced to generate this image, which shows changes that occurred between the observations. The dominant yellow color represents zero phase change, as evidenced by the very bright yellow spot on the left center, which is a radio tower. Black areas represent the loss of phase coherence, where the noisy phases of Plate 1 have been left out. The various colors, from blue to red to green, indicate small motions (2-3 cm) of the fields from swelling or shrinkage associated with watering.

high-resolution SAR systems was launched that revolutionized the panorama of available instruments. The German Aerospace Agency (DLR) launched TerraSAR-X, an X-Band system able

to operate in four modes (from High-Resolution Spotlight, 1-m resolution, to ScanSAR, 18 m). The Italian Space Agency (ASI) launched its first high-resolution X-Band SAR satellite,

COSMO-SkyMed-1, in 2007. COSMO-SkyMed's capabilities include very high resolution (up to 1 m), polarimetric modes and right-left lateral view options. With its RADARSAT-2 program, the CSA provides data continuity for RADARSAT-1 users. RADARSAT-2 has new capabilities including the high-resolution (3-m) Ultra Fine mode, full polarimetric capabilities, and right-left looking view.

It has been more than 20 years since the first successful applications of InSAR to measure large-scale surface deformation, characterizing the coseismic displacement field of the Landers earthquake (Massonnet et al. 1993). InSAR is a SAR data processing technique aimed at estimating any variation of the phase component of two or more SAR images acquired along the nominally same orbit (Rosen et al. 2000). The phase is related to the satellite-to-target distance, which is composed of a large number of integer wavelengths and the measured fractional phase component. The result of the application of InSAR is the so-called interferogram, that is, the pixel-to-pixel difference of the phase components of two SAR images covering the same area. Differential InSAR (DInSAR) is the technique applied to generate topographically corrected interferograms, where the topographic phase contribution measured in the interferogram is removed using a digital elevation model (DEM) (Rosen et al. 2000). The result of DInSAR application is an interferogram containing the displacement vectors along the satellite line of sight (LOS). This means that DInSAR measures the result of the projection of the North, East, and Up components of three-dimensional surface displacements into the LOS. In order to obtain a LOS displacement map, a "phase unwrapping" algorithm is applied to the result of DInSAR (Fornaro et al. 1996).

Coseismic Displacements

The capability of DInSAR to detect centimetric movements over large areas has been successfully used in seismology, where traditionally the main topic of scientists is to determine the source parameters (the fault parameters). An earthquake occurs

along a fault plane, that is, a discontinuity in the crust, when the stress applied exceeds a certain threshold. The size of an earthquake is measured by the seismic moment, while the moment magnitude (M_w) is the energy released (Kanamori 1977). M_w is measured over a scale of 10^0 (0–10, real numbers). From scientists' experience it is commonly shareable that surface deformation occurs in case of M_w over 5.0 and shallow hypocentral depth (up to 5 km). In recent years some reviews have been focused on defining such thresholds (Dawson and Tregoning 2007) and the trade-off magnitude-depth. DInSAR provides high spatial resolution of surface deformation and is able to provide reliable constraints to the fault location, to geometric parameters (depth, fault dimension, strike, dip), and to the definition of the average slip. DInSAR is also capable to provide an estimate of the geodetic moment, which now rarely differs from the seismic moment (Feigl 2002). Indeed the geodetic moment is the result of the estimation of released slip so as it derives from the surface displacements detected from DInSAR. Such displacement is the result of the coseismic deformation, plus secondary (but not negligible) contributions from post-seismic and aseismic deformations.

It is not far from truth saying that DInSAR has become a key tool in the Earth Sciences domain. Today it is a reliable input to modeling tools, either for simple or complex methods. Many models deals with the Okada formulation (Okada 1985) that implies the use of a uniform elastic half space. Improved data constraints and use of sophisticated modeling and inverse methods allow exploring variation of elastic properties in layers or more complex representation of faults and rheology (Masterlark and Hughes 2008). In recent years, DInSAR measurements combined with traditional seismological data have also been used for rupture models including constraints on rupture propagation and velocity (Fielding et al. 2013).

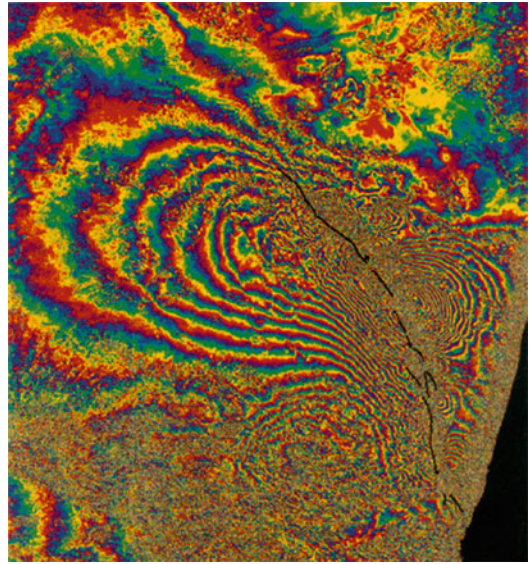
Case Studies

In the following sections, some examples of applications of DInSAR are showed. The aim is

to give the reader useful hints about the capabilities of DInSAR technique to measure earthquake-induced deformation. Today the earthquakes studied with the key contribution of DInSAR are more than 100. The case studies focused in the following are, for various reasons, among the most well known worldwide. Indeed among them is the first ever seismological application of DInSAR, the first ever in Europe to a seismic sequence with three mainshocks, one of the largest earthquakes in seismological instrumental history, etc.

Landers (1992)

The Mw (moment magnitude) 7.3 Landers earthquake of June 28, 1992, that ruptured over 85 km along a fault system is the first case study where DInSAR has been successfully applied to the measurement of the surface displacement field and allowed to provide a detailed image of the overall deformation field. Field observations and seismological data have shown right lateral slip up to 4 m and 6 m at about 10 km and 40 km north of the mainshock, for which the hypocentral depth is between 3 and 8 km. It is worth noting that the Landers earthquake was followed 3 h later by a Mw 6.2 (called the Big Bear) that, notwithstanding the magnitude, did not provoke any surface displacement. Coseismic “geodetic” displacements have been measured by GPS (Global Positioning System) (up to 3-m horizontal offset) (Hudnut et al. 1994) and pixel correlation from SPOT satellite (1-m displacement) (Michel and Avouac 2006). However the absolute novelty has been the analysis performed using a pair of SAR ERS-1 images (April 24–August 7, 1992). The result (Fig. 2) revealed a complex deformation surrounding a series of discontinuous strike-slip rupture segments in the Mojave Desert that stretched over about 70-km length. The maximum relative change in distance between the satellite and the ground (line of sight) produced by the earthquake was about 1 m across the fault, caused by around 6 m of slip on the fault. Deeper in detail the coseismic interferogram reveals no organized fringes in a band within 5–10 km of the fault trace, probably due to the high displacement gradient recorded in



Remote Sensing in Seismology: An Overview, Fig. 2 Coseismic interferogram for the Landers earthquake (From Massonnet et al. 1993). A color cycle represents a chance distance of 28 mm along the satellite LOS. *Black segments* depict the fault geometry as mapped in the field

this area exceeding a critical value detectable by DInSAR. Or, additionally, crustal block rotation may have reduced the coherence. A theoretical model based on Okada formulation has been performed by applying to each rectangular patch 2 km long a constant slip taken from the offset mapped on the field. The model extends from 0- up to 15-km depth, in order to match the seismic moment. It is interesting to see where the measured fringe pattern and the model differ. The main discrepancies are local complexities or short-wavelength features near the fault rupture, whereas the local concentration of strain is highly sensitive to the detailed geometry of the fault.

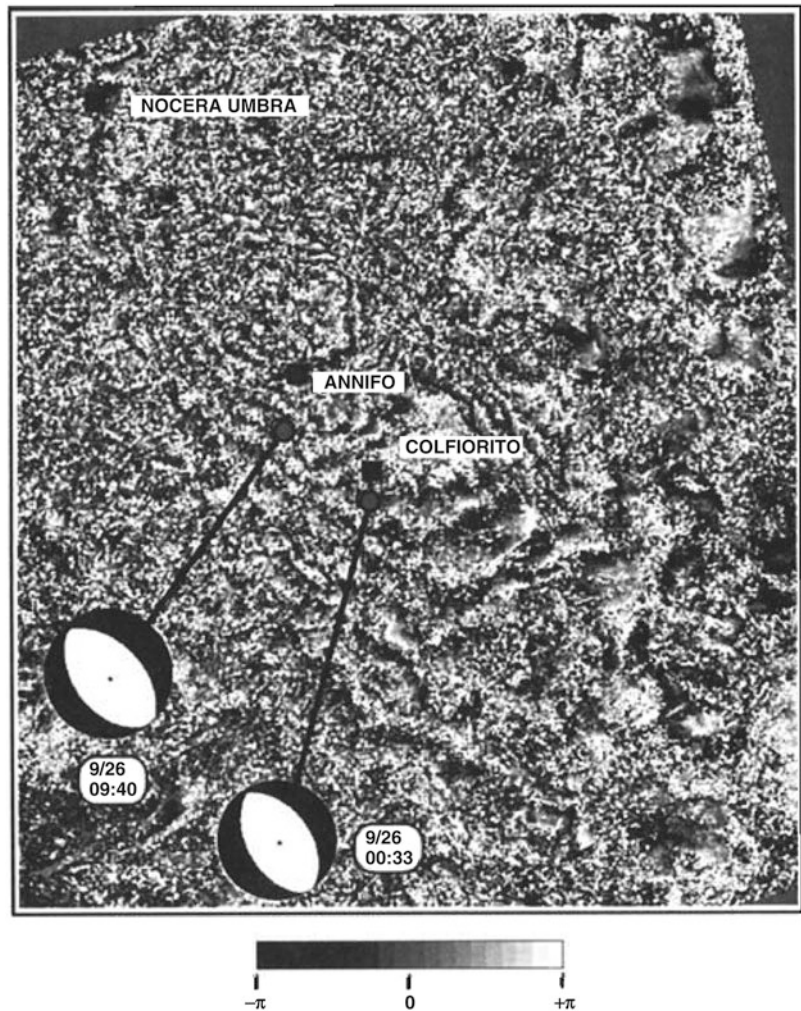
Umbria-Marche (1997)

In September 1997 a large area of Umbria-Marche Apennines (Central Italy) was struck by a strong seismic sequence. A foreshock occurred on September 3 (Mw 4.5), while two mainshocks took place on September 26 (00:33 GMT, Mw 5.7; 09:40 GMT, Mw 6.0). Some days later, on October 14 a Mw 5.6 occurred toward South, around the village of Sellano. This is one of the

Remote Sensing in Seismology:

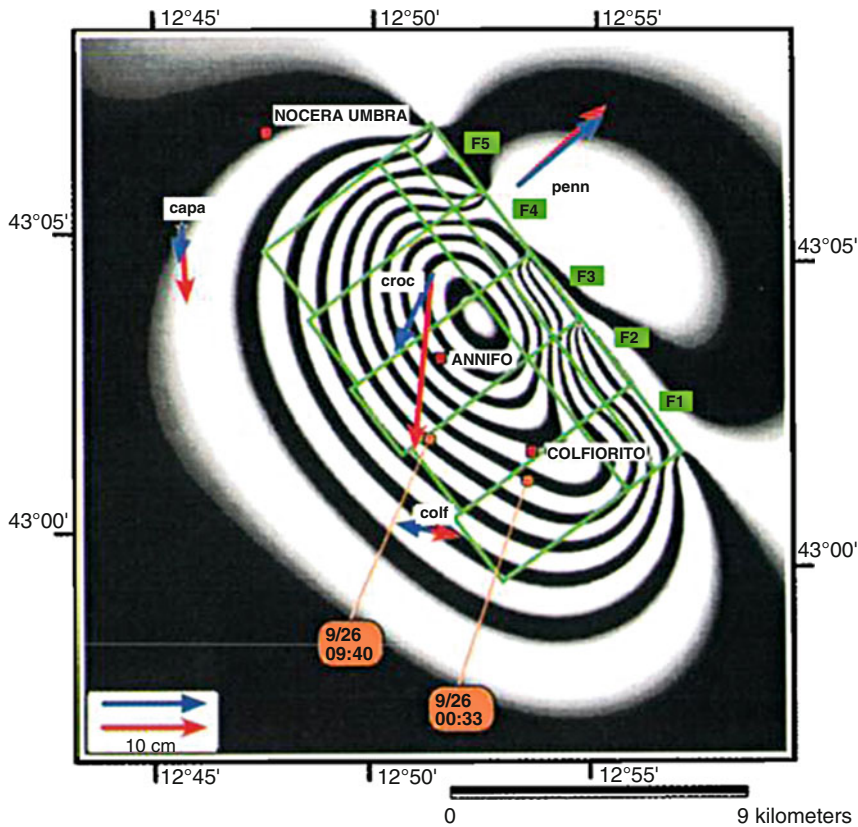
An Overview,

Fig. 3 Coseismic interferogram of the Umbria-Marche earthquake in September 1997. Fringes (black-white contours) correspond each to a LOS displacement of 28 mm. The two mainshocks are indicated by the source mechanisms (black-white beach balls), hour (GMT), and dates (From Stramondo et al. 1999)



earliest applications of DInSAR in seismology worldwide, the first in Europe (the second since ever after Grevena, 1995, Greece) concerning a complex sequence characterized by multiple mainshocks and thousands of minor events. Indeed 1,650 earthquakes with magnitude between 2.5 and 6 have been recorded in a period of 40 days in the area (Chiaraluce et al. 2003). This moderate-magnitude earthquake sequence caused a broad damage pattern and extensive concern among the population in the region. Due to the shallow depth of these earthquakes (less than 8 km), the ground shaking at the surface has been very large. The 1997 Umbria-Marche earthquake sequence in the

complex tectonic setting of the North Apennine offers the possibility to unravel the anatomy of a normal fault system and to investigate fault segmentation. The map of coseismic deformation has been obtained by a pair of ERS-2 images (September 7–October 12), so that it contains the surface movements due to both September 26 events. The interferogram (Fig. 3) shows some fringes (gray contours) that trace a bean-shaped pattern NW-SE oriented with a maximum LOS displacement of about 25 cm close to Annifo village. At the same time, one of the four GPS benchmarks in the epicentral area measured about 14-cm horizontal and about 24-cm vertical movements. Either GPS or DInSAR has been the



Remote Sensing in Seismology: An Overview, Fig. 4 Synthetic coseismic interferogram of the Umbria-Marche earthquake in September 1997. The fault plane has been divided in 15 patches at different

depths. Synthetic fringes (*black-white contours*) correspond each to a LOS displacement of 28 mm. Measured (*in red*) and modeled (*in blue*) GPS horizontal vectors are also shown (From Stramondo et al. 1999)

input data to an inversion modeling program that has provided an estimation of the fault parameters (Stramondo et al. 1999) and the distribution of the slip over the fault plane (Fig. 4).

The geophysicists and seismologists have benefitted from the additional information from DInSAR that has provided a picture of the overall displacements projected onto the LOS. Indeed differently from the results of ground surveys that give a sparse spatial sampling of the deformation field, DInSAR makes available to scientists a spatially continuous view of such field.

Bam (2003)

On December 26, 2003, southeastern Iran was shaken by a destructive earthquake (Mw 6.5). The event was located near the city of Bam and the Bavarat village (Fig. 5) and nucleated at

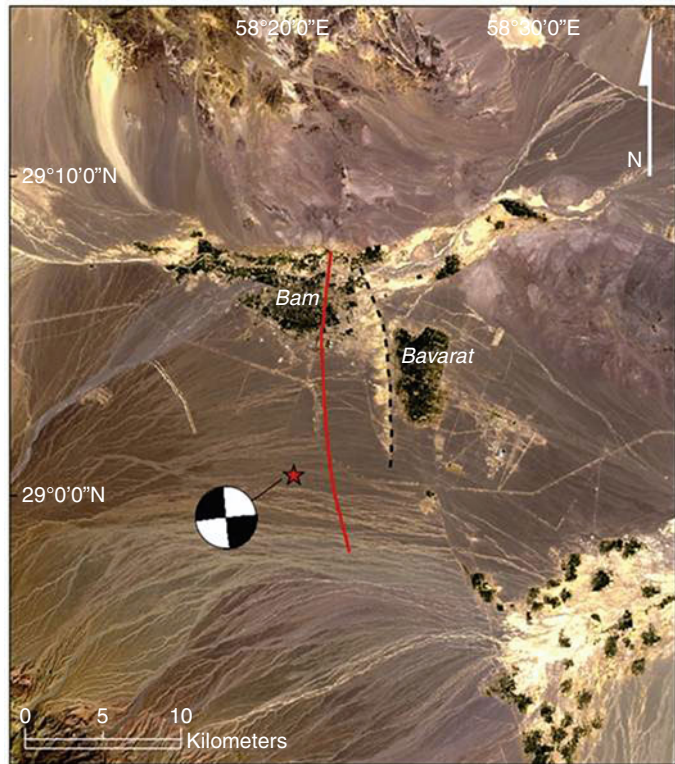
a depth below 10 km. Seismic focal mechanism suggests that slip occurred along a high-angle dipping fault plane, with a prevailing strike-slip sense of motion.

Before this earthquake occurred, the “Bam fault” was the major tectonic feature mapped in the epicentral area (Fig. 5), extending along the west side of Bavarat village (i.e., Fu et al. 2004). Initially, the December 2003 shock has been associated to this fault. Although the reports on field investigations describe the surface features produced by the event such as landslides, liquefactions, sinkholes, and fractures particularly concentrated between the city of Bam and Bavarat village (Zaré 2003), none of these surface features have been interpreted as evidences for surface dislocation directly associated to the Bam fault. In other words, such secondary effects

Remote Sensing in Seismology:

An Overview,

Fig. 5 Landsat 7 Enhanced Thematic Mapper Plus (ETM+) image of the Bam earthquake area (<http://www.parstimes.com/spaceimages/bam/landsat.html>). Red line is the fault inferred from InSAR data; this fault has been recognized as the earthquake source. Dotted black line is the Bam fault. The star indicates the epicenter location; focal mechanism (beach ball) is from USGS website (Modified from Stramondo et al. 2005)

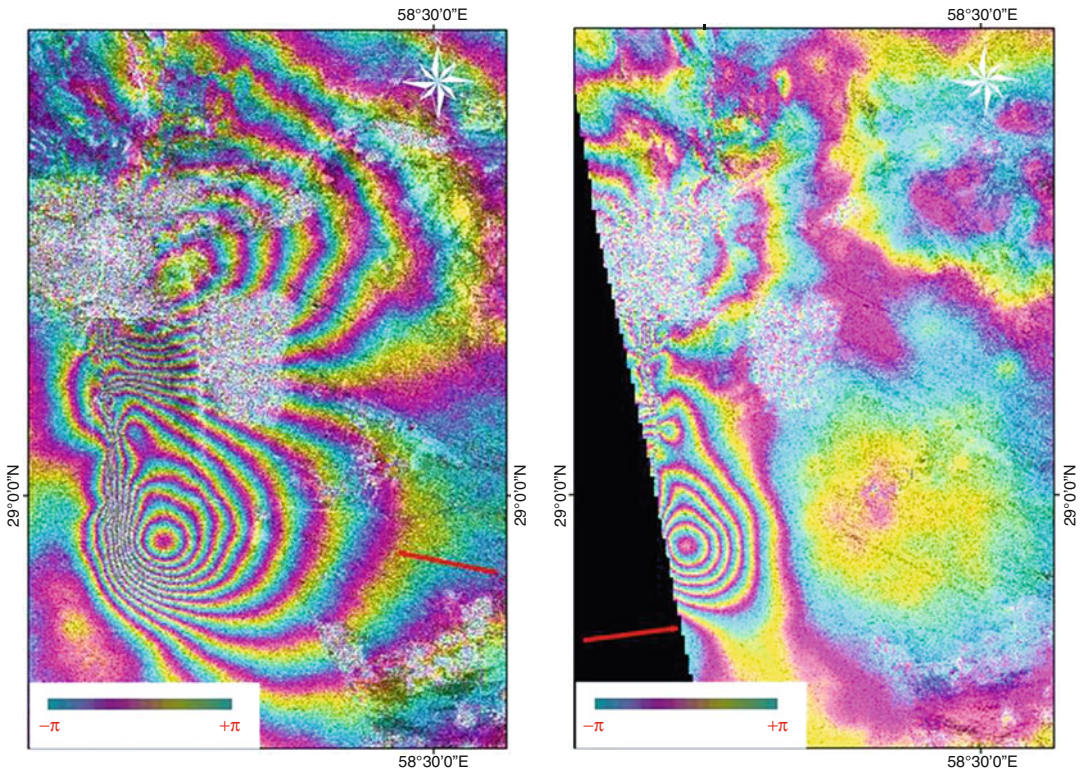


are not located around the fault responsible for the earthquake. In this case, DInSAR outcomes heavily contributed to infer location and geometry of the unknown fault that is the source of the earthquake. It is noteworthy how the hypocenter distribution of aftershocks recorded by a temporary seismic network installed some weeks after the seism is in good agreement with geometric parameters derived from DInSAR.

A dataset of Envisat ASAR data acquired from two different geometries (ascending and descending orbits) has been used for DInSAR analyses. In particular, about 4 km west of the Bam fault (black dashed line in Fig. 5), the fringes of displacement reach the maximum gradient (Fig. 6) along an averaging north–south direction suggesting location, length, and azimuth of the fault (red line in Fig. 5) (Stramondo et al. 2005a). The fault, named Arg-e-Bam, is almost 20 km long, and the northernmost portion (about 5 km) is exactly beneath the city of Bam.

The combination of the satellite LOS with the mechanism of the fault results in a positive

displacement (satellite-to-surface distance increase) measured in the descending interferogram of ~ 25 cm occurring within the southern lobe. Conversely, the lobe to the north has a maximum LOS movement equal to about -24.5 cm. This different sense of motion along the same fault side may be explained by the very low angle between the fault strike (fault alignment) and the satellite orbital direction and due to strike variations along the fault (see Fig. 7). The map of displacements has been used to simulate the distribution of slip over a fault plane, starting from the fault parameters constrained by DInSAR and seismological data. The synthetic interferogram obtained by an 18×12 km fault plane (see Stramondo et al. 2005b) fits well with the descending interferogram. The proposed model is a very simple representation of the reality, meaning that local complexities in the slip distribution and fault geometry have not been included (see Fig. 8). The aim, in the Bam case study and more generally in seismology, is first to provide a fast and draft view of the seismic



Remote Sensing in Seismology: An Overview, Fig. 6 (left) Coseismic interferogram from descending Envisat orbit (December 3, 2003–January 7, 2004); (right) coseismic interferogram from ascending Envisat orbit

(January 25, 2004–November 16, 2003). The *red lines* in both figures indicate the satellite LOS. Each color cycle corresponds to ~ 2.8 -cm displacement (From Stramondo et al. 2005)

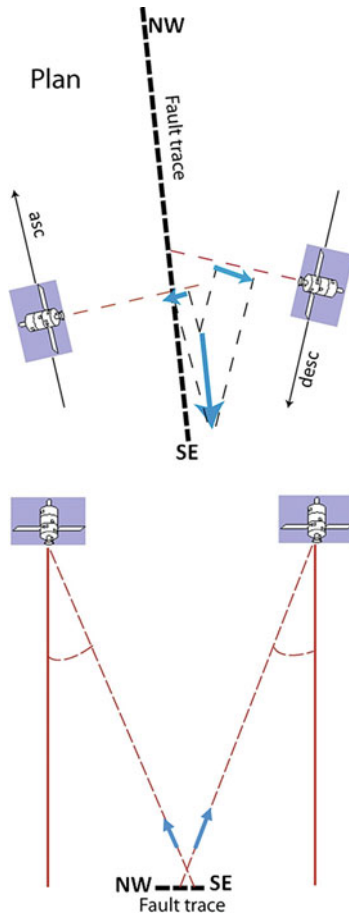
source. Later, if necessary and where available data allow it, deeper and more complex analyses can be performed.

L'Aquila (2009)

On April 6, 2009 (01:32GMT), a large portion of Abruzzi region and in particular the city of L'Aquila (Central Italy) were hit by a seismic sequence whose main event (M_w 6.3) occurred very close to the historical center (Walters et al. 2009). In L'Aquila and in the neighboring villages, the earthquake caused collapse or irreparable damage to over 15,000 buildings, killing 308 people and causing the relocation of over 65,000. The mainshock was followed by seven major aftershocks ($M_w > 5$), and over 6,000 smaller events occurred in the next few months in an area extended northwest–southeast for about 35 km.

L'Aquila 2009 is the first ever case study where X-, C-, and L-band SAR data have been available for the same earthquake (Stramondo et al. 2011b). SAR data coming from COSMO-SkyMed, Envisat, and ALOS satellites, allowed an unprecedented coverage of a coseismic displacement field in terms of different line of sights (LOS), ground resolutions, and frequency bands of the imaging sensors (Fig. 9). Notwithstanding this, in the early post-earthquake, attempts to determine the location of the seismic source were hampered by the lack of clear surface faulting. Indeed only 6 days after the seism, the first meaningful interferograms were obtained by COSMO-SkyMed and Envisat (descending orbit).

The COSMO-SkyMed interferogram shows a northwest–southeast concentric fringe pattern composed of 12 fringes, each corresponding to



Remote Sensing in Seismology: An Overview, Fig. 7 Planimetric and along-orbit (azimuth) oriented sketch of Envisat geometries along ascending and descending paths. Planimetric view (*top*) shows horizontal displacement vector and their projection in planimetry along each LOS (*in blue*); azimuth view (*bottom*) provides LOS projection of ascending and descending horizontal displacement vector. The fault trace is also showed

a range increment of 1.5 cm (going from the outer to the inner fringes). The maximum ground displacement measured along the satellite LOS (i.e., $\sim 36^\circ$ from nadir) reaches 18 cm at about 4 km SE from L'Aquila.

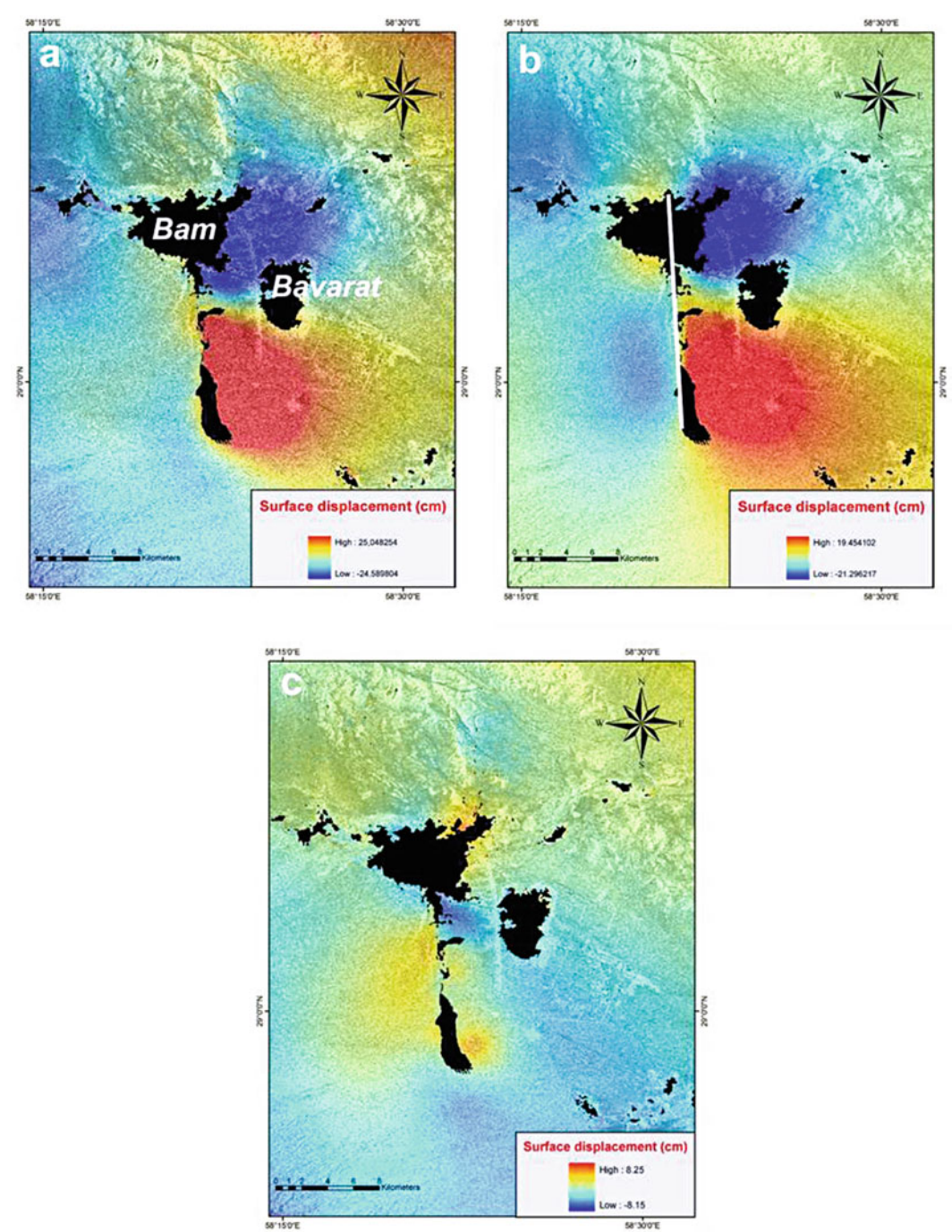
Concerning Envisat, two image pairs have been used, so that the coseismic displacements have been detected and measured from two different LOSs. Indeed a first interferogram (post-seismic image dated April 12) along descending orbit measured ten fringes (each 2.8 cm of fringe increment, total 28 cm), while in the Envisat

ascending interferogram (post-earthquake image dates April 15) nine fringes only (maximum displacement ~ 23 cm) probably due to signal loss. It is worth noting the different LOS incidence angles (angle from nadir, or from the vertical to the surface), $\sim 36^\circ$ versus $\sim 23^\circ$. Finally 2 weeks later (April 22), ALOS measured about 24–26-cm displacement (each fringe 11.5 cm of fringe increment) along $\sim 38^\circ$ from nadir. The surface movement is predominantly along a vertical axis, as expected, being originated from a southwest dipping normal faulting mechanism.

In order to explain the geometry and mechanism of the fault, a nonlinear inversion to constrain the fault geometries with uniform slip, followed by a linear inversion to retrieve the slip distribution on the fault plane, has been applied (Fig. 10). The input dataset to the modeling algorithm is composed of DInSAR data plus 30 GPS permanent stations that provided the horizontal vectors. The fault model from geodetic data well defines the position and geometry of the seismogenic fault, identified with the Paganica fault. The additional information provided from DInSAR to standard in situ data and ground surveys is absolutely relevant to the seismological analyses of the earthquake. The large number of SAR images spanning the mainshock recorded the coseismic surface displacement pattern in great detail. The most important results obtained from SAR data were the fast identification and characterization of the earthquake source, i.e., the Paganica fault, indeed previously overlooked as a possible source of a large earthquake. In addition, COSMO-SkyMed data, thanks to the use of a high-resolution (5-m) digital elevation model (DEM), detected local gravitational slope failures reactivated by the mainshock (Moro et al. 2011) where small (4–5-cm) deformations were measured. The L'Aquila sequence is today the only one where three SAR datasets from three different wavelengths have been used.

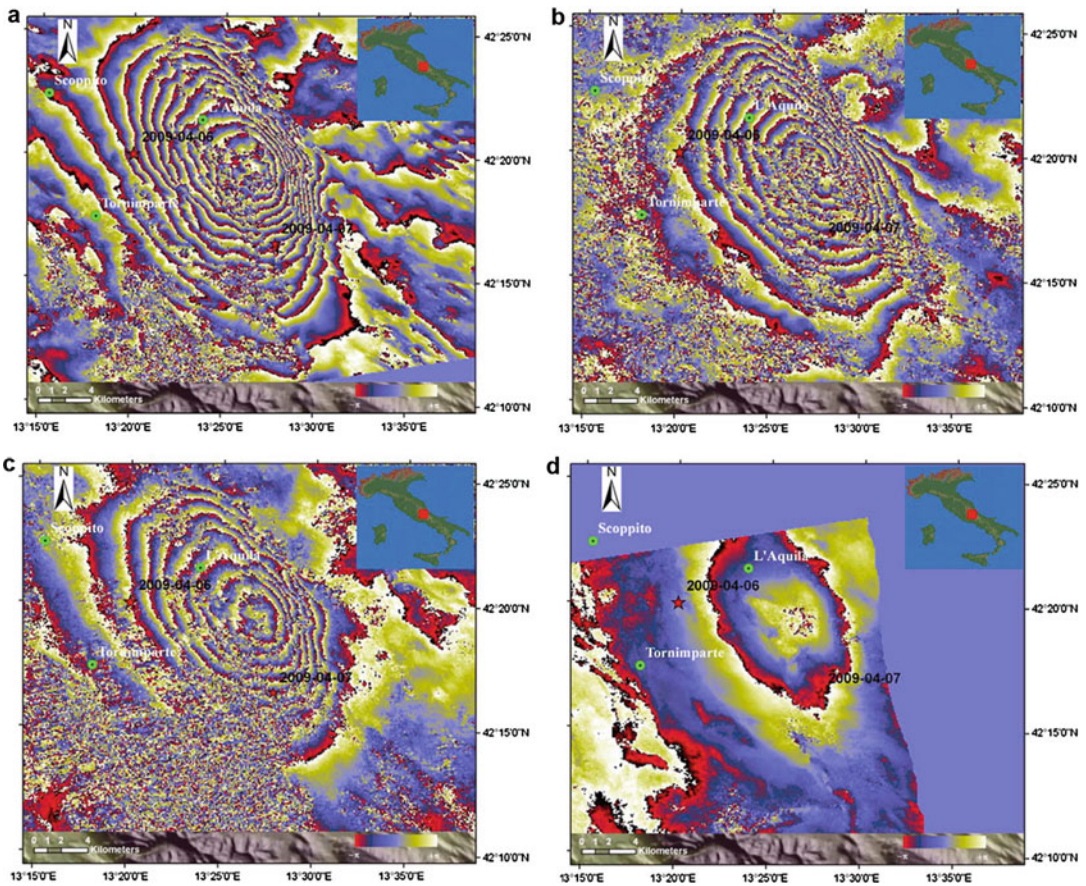
Tohoku-Oki (2011)

The disastrous Tohoku-Oki megathrust earthquake (M 9.0) occurred on March 11, 2011 (05:46:23 UTC), near the NE coast of Honshu island (Japan). The megathrust earthquake (M 9.0)



Remote Sensing in Seismology: An Overview, Fig. 8 (a) Unwrapped phase from the coseismic interferogram (Fig. 6, left). The city of Bam and Bavarat village

are in black, together with the near field of fault, (b) synthetic interferogram from (a). The white line is the surface trace of the modeled fault, (c) residual map

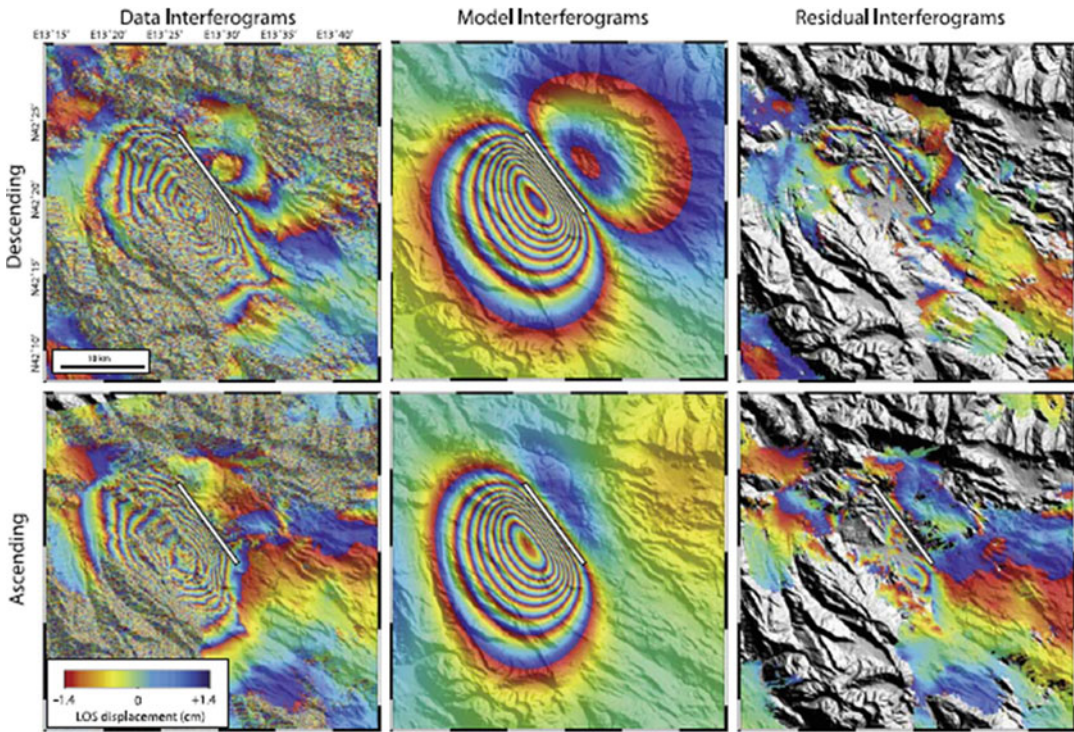


Remote Sensing in Seismology: An Overview, Fig. 9 Four differential interferograms obtained for the analysis of coseismic displacement of L'Aquila earthquake: (a) COSMO-SkyMed along ascending orbit; (b)

Envisat along descending orbit; (c) Envisat along ascending orbit; and (d) ALOS along ascending orbit. The study area is the red square in the top right inset (From Stramondo et al. 2011b)

originated from a thrust fault on the subduction zone plate boundary between the Pacific and North American plates (Liu and Yamazaki 2013), whose relative motion is 8–8.5 cm/year. The mainshock was preceded by a foreshock sequence lasting two days with a major event on March 9 (Mw 7.3), occurred along the same low-angle thrust plane and about 45 km away of the mainshock (Ide et al. 2011). Although the epicenter is offshore, located about 130 km E of Sendai at about 32-km depth, large coseismic deformation has been detected on land based on GEONET (GPS Earth Observation Network System). Indeed GPS station at the Oshika Peninsula registered up to 5 m toward ESE and 1 m

downward. In particular PRISM and AVNIR are two optical sensors that are not discussed in the manuscript (differently from PALSAR, for instance). A large tsunami estimated more than 10 m followed the seism and reached a few km inland causing serious damage to humans and buildings along the coastline (Watanabe et al. 2012). Maximum run-up heights up to 39.7 m were reported from the Sanriku region. The satellite Earth observation community has focused its attention on the investigation of the surface effects based on satellite data exploitation, in order to measure the surface displacement pattern, the surface changes due to damage, and the coastal changes due to inundation. The focal region from



Remote Sensing in Seismology: An Overview, Fig. 10 (left) Data, (middle) model, and (right) residual interferograms for Envisat (top) descending track 079 (dates: September 02, 2001–September, 04, 2012) and (bottom) ascending track 129 (dates: September 03, 2011–September 04, 2015), with the fault rupture modeled as a uniform dislocation in an elastic medium.

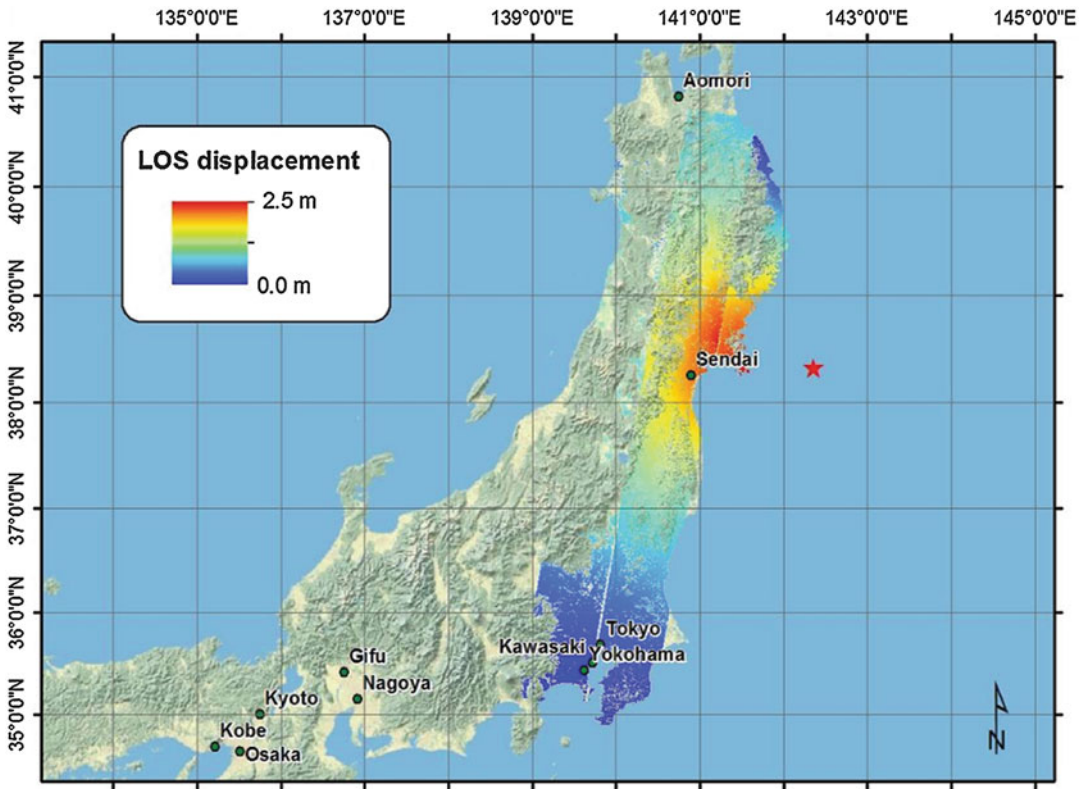
The white line in the interferograms is the up-dip surface projection of our model fault plane. All interferograms are overlain on SRTM topography illuminated from the NE. RMS misfit values for descending and ascending uniform slip models are 1.2 and 1.0 cm, respectively (From Walters et al. 2009)

seismological analysis indicated an area of about 500×200 km moved with a maximum displacement of about 24 m near the hypocenter. Soon after the earthquake, most of the space agencies made their satellite data of the epicentral region available free of charge via the Geohazards Supersites initiative (GEO's Tohoku-Oki Event Supersite Website <http://supersites.earthobservations.org/sendai.php>); JAXA (Japanese Aerospace Exploration Agency), ESA (European Space Agency), DLR (German Space Agency), NASA (National Aeronautics and Space Administration), and CNES (French Space Agency) provided a large number of SAR and optical images even though each agency has different data distribution policies (GEO's Tohoku-Oki Event Supersite Website). A huge dataset of SAR images has been made

available soon after the Tohoku-Oki earthquake thanks to the Group on Earth Observations (GEO) Geohazards Supersites initiative that decided to open the *Tohoku-Oki Event Supersite*.

The main rule to select a Supersite is that “access to spaceborne and in-situ geophysical data of selected sites” is provided to the scientific community. Indeed ERS-2, ASAR Envisat, ALOS PALSAR, TerraSAR-X, and COSMO-SkyMed data have covered the epicentral region. Although such large amount of acquisitions, Envisat and ALOS were the only two satellites able to fully cover the whole earthquake area.

The coseismic interferograms from the 2011 Tohoku-Oki earthquake show a great amount of surface deformation over all of northeastern Japan with a maximum line-of-sight



Remote Sensing in Seismology: An Overview,
Fig. 11 Map of coseismic displacements in Honshu island from Envisat ASAR data. Envisat images are all in IS6 mode, which means they are characterized by an

incidence angle of 40° at the center of each frame (or swath), along descending orbits. The red star indicates the epicenter and red arrows azimuth and range LOS vectors (From Stramondo 2013)

displacement of up to 3.7 m from the ascending PALSAR tracks and 2.4 m from the descending ASAR tracks (Fig. 11), respectively (Feng et al, 2011; Feng and Jónsson 2012). Such a large discrepancy is ascribable to the large horizontal motions, particularly in the eastern component (up to 5.03 m at GPS station 0550), and the large incidence angle. However, because the slip on the subduction interface was at a large distance from the onshore geodetic observations, the InSAR data contribute little new information in Japan, where ground-based GNSS observations are dense (Feng and Jónsson 2012). Nevertheless, the localized deformation due to M 6–7 aftershocks could be measured. In countries where dense GPS are not available, InSAR data alone can be used to invert for the slip distribution on the subduction interface

(Feng and Jónsson 2012; Pritchard et al. 2006; Kyriakopoulos et al. 2013).

New Zealand (2011)

The towns of Darfield (Canterbury) and Christchurch in New Zealand were hit by strong earthquakes within a time span of a few months, in September 2010 and February 2011. The Mw 7.1 Darfield earthquake took place on September 3, 2010, followed by a sequence of large aftershocks to the east, up to the occurrence of the Mw 6.3 Christchurch earthquake on February 21, 2011 (Fig. 12) (Beavan et al. 2010). The Darfield (Canterbury) earthquake occurred along a previously unrecognized east–west fault line, the strike-slip Greendale fault (Beavan et al. 2010; Stramondo et al. 2011b), that DInSAR allowed to detect and characterize.

The hypocenter of the Christchurch event was approximately 6 km southeast of Christchurch's city center, at a depth of 5–6 km, at the easternmost limit of the Darfield aftershocks.

One of the main topics in such case is identifying the possible cause-and-effect relationship, due to the stress transfer, between the September and the February main events. The research in the last decades demonstrated that over major active faults or fault systems, where seismologists registered the occurrence of an earthquake, the probability of occurrence of a second shock increases or decreases according to stress changes 1–4. Indeed, a mainshock perturbs the stress state in other sections of the same fault or in adjacent faults: this theory is known as Coulomb Stress Triggering.

The hypothesis is that once an earthquake occurs, the stress does not dissipate, but it propagates in the surrounding area, where it may increase the probability of occurrence of further earthquakes. In other words, the question is: do earthquakes interact with each other?

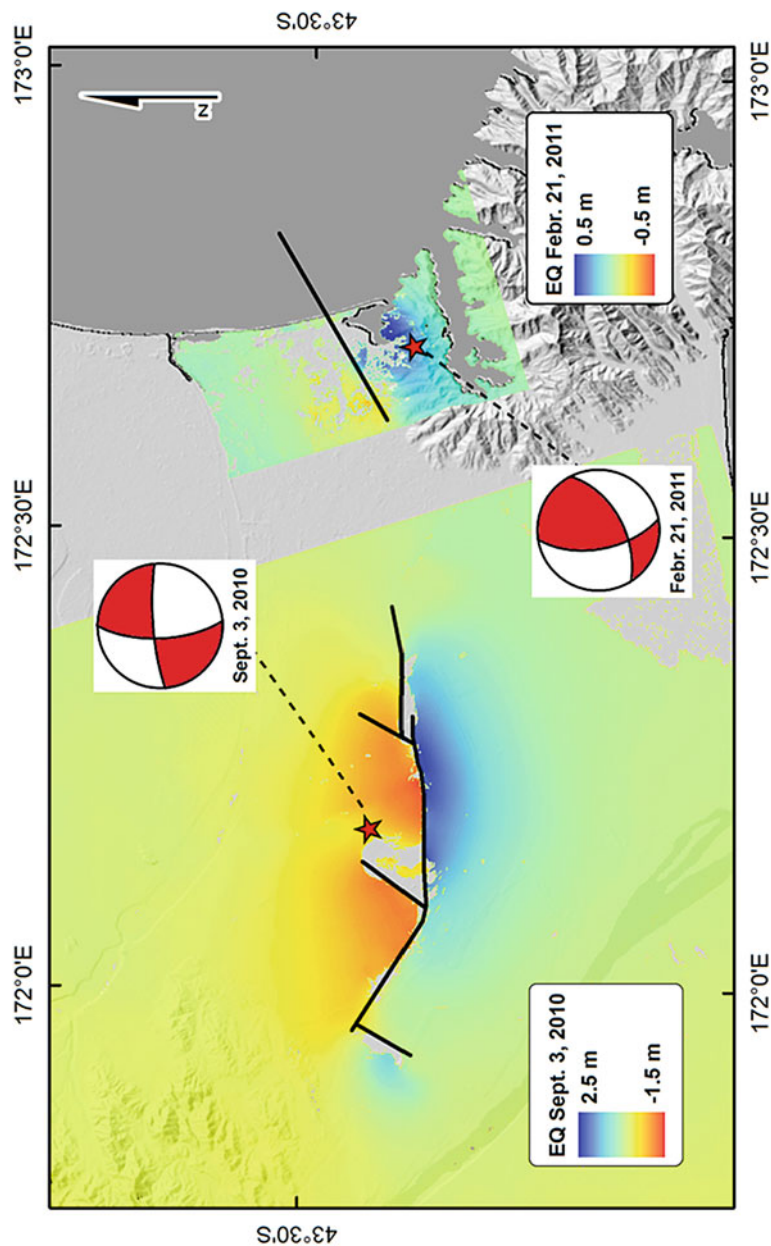
In both earthquakes DInSAR technique has been applied to ALOS PALSAR images. The Darfield earthquake led to an average displacement of 2.5 m, up to a maximum of 5 m. Stemming from DInSAR map of surface displacements, the seismogenic fault (Greendale fault) resulted a 44-km-long structure, with some minor lateral branches (see Fig. 12). Conversely, the Christchurch earthquake produced much minor deformation at the surface (average 50 cm), but the fault exactly crossed the city of Christchurch, so that strong damages and casualties were recorded.

In order to verify the hypothesis that an interaction between these earthquakes may have occurred and to understand the role of the first earthquake in promoting the rupture of the second event, the Coulomb failure function (CFF) has been evaluated onto the Christchurch fault plane. The CFF is obtained by computing the stress tensor corresponding to the elastic dislocation induced by the Darfield earthquake, projecting it onto the fault plane of the Christchurch earthquake, and evaluating the relative weights of the stress components (normal and shear), assuming an certain friction coefficient.

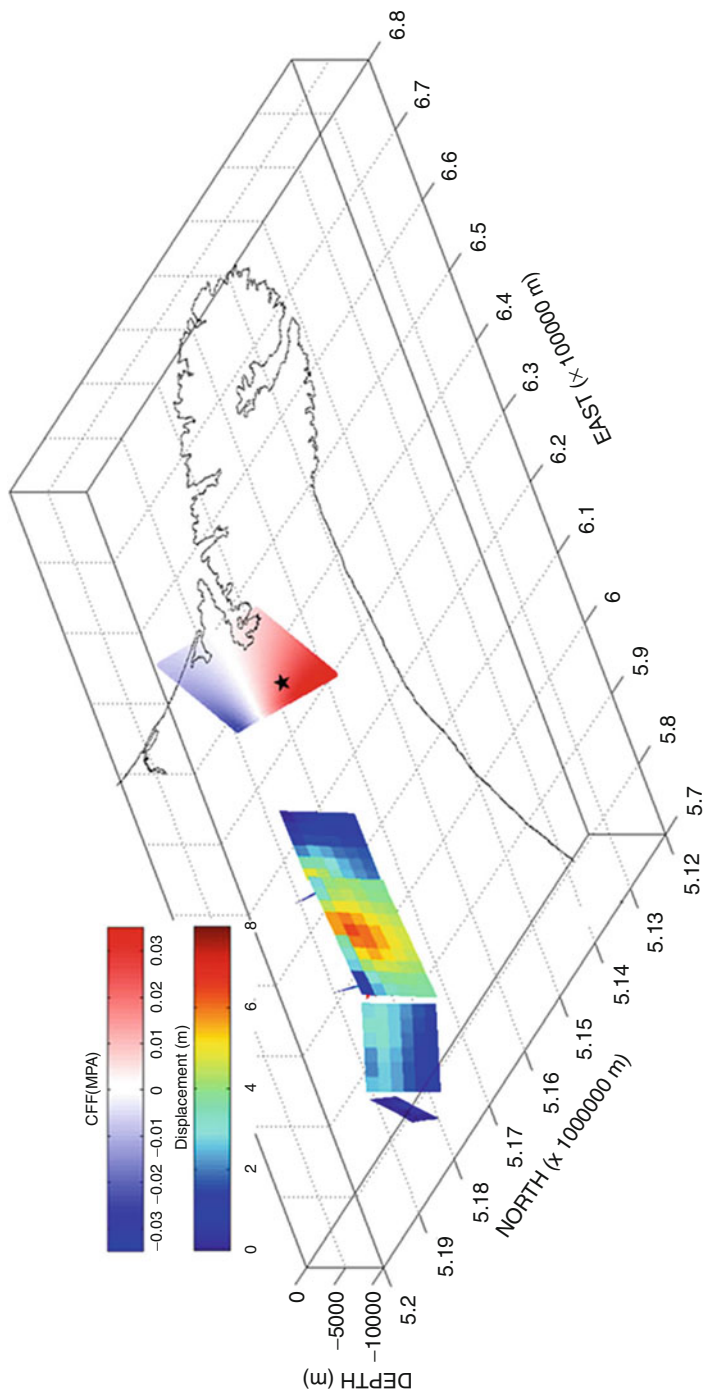
Positive or negative variations of the CFF indicate that the stress field is acting to promote or oppose the rupture, respectively. In our case study, the rupture of the 2010 Darfield earthquake loaded a large portion of the crust with stress values exceeding 1 bar. Taking into account the three-dimensional location of the Christchurch rupture plane (Fig. 13), it is clear that the shallower part of the fault (down to about 5-km depth) has actually been unloaded by the Darfield earthquake. The average CFF value on the loaded portion of the Christchurch fault is over 0.01 MPa, with peak values exceeding 0.03 MPa. These stress levels are definitely non-negligible, since a stress value of the order of 0.01 MPa is considered a threshold for effective triggering of seismic events. The stress perturbation due to the Darfield earthquake onto the Christchurch fault promoted the second earthquake, even though without a clear knowledge of the starting stress level of the second fault, it cannot be stated whether this latter was (or not) already likely to occur (Stramondo et al. 2011b).

Summary

In this entry the use of remote sensing in seismology, and in particular the investigation of surface displacements due to earthquakes, has been described. The first paragraph has been dedicated to provide a short overview of SAR satellite missions since the first ever used for soil applications up to the new generation of high-resolution SAR systems. The following paragraph has been concerning the introduction of the use of differential interferometric synthetic aperture radar (DInSAR) technique, a SAR data processing method that exploits the capability of satellite radar images to measure distance changes between satellite and ground surface with centimetric accuracy, for coseismic studies. The main part of the entry concerns the description of case studies relevant to the main topics, covering the time span from the first ever seismological application of DInSAR (Landers earthquake, 1993, see Massonnet et al. 1993).



Remote Sensing in Seismology: An Overview, Fig. 12 Unwrapped interferograms in LOS geometry. *Red stars* are the hypocenter location (from CMT) of the two events. Focal mechanisms are also shown (From Stramondo et al. 2011b)



Remote Sensing in Seismology: An Overview, Fig. 13 A 3D perspective view for the fault model of Greendale fault and Christchurch earthquake fault. In Greendale red and black stars indicate the hypocenter of the two earthquakes, respectively. Both fault the largest slip (about 6.5 m) is concentrated in the middle segment from 0- to 6-km depth. Coulomb stress change is estimated for the Christchurch fault plane. The panels are in UTM WGS84 coordinate system (From Stramondo et al. 2011b)

References

- Beavan J, Samsonov S, Motagh M, Wallace L, Ellis S, Palmer N (2010) The Darfield (Canterbury) earthquake: geodetic observations and preliminary source model. *Bull N Z Soc Earthq Eng* 43:228–235
- Chiaraluce L, Ellsworth WL, Chiarabba C, Cocco M (2003) Imaging the complexity of an active normal fault system: the 1997 Colfiorito (central Italy) case study. *J Geophys Res* 108(B6):2294. doi:10.1029/2002JB002166
- Dawson J, Tregoning P (2007) Uncertainty analysis of earthquake source parameters determined from InSAR: a simulation study. *J Geophys Res* 112: B09406. doi:10.1029/2007JB005209
- Feigl K (2002) Estimating earthquake source parameters from geodetic measurements. In: *International handbook of earthquake and engineering seismology*, vol 81A. Academic, Boston
- Feng G, Jónsson S (2012) Shortcomings of InSAR for studying megathrust earthquakes: the case of the Mw9.0 Tohoku-Oki earthquake. *Geophys Res Lett* 39(10):L10305
- Feng G, Ding X, Li Z, Mi J, Zhang L, Omura M (2011) Calibration of an InSAR-derived coseismic deformation map associated with the 2011 Mw-9.0 Tohoku-Oki earthquake. *IEEE Geosci Remote Sens Lett* 9(2):302–306
- Fielding EJ, Sladen A, Li Z, Avouac J-P, Bürgmann R, Ryder I (2013) Kinematic fault slip evolution source models of the 2008 M7.9 Wenchuan earthquake in China from SAR interferometry, GPS and teleseismic analysis and implications for Longmen Shan tectonics. *Geophys J Int* 194(2):1138–1166. doi:10.1093/gji/ggt155
- Fornaro G, Franceschetti G, Lanari R (1996) Interferometric SAR phase unwrapping using green's formulation. *IEEE Trans Geosci Remote Sens* 34:720–727
- Fu B, Ninomiya Y, Lei X, Toda S, Awata Y (2004) Mapping active fault associated with the 2003 Mw 6.6 Bam (SE Iran) earthquake with ASTER 3D images. *Remote Sens Environ* 92:153–157
- Gabriel AK, Goldstein RM, Zebker HA (1989) Mapping small elevation changes over large areas: differential radar interferometry. *J Geophys Res* 94(B7):9183–9191
- Hudnut KW, Bock Y, Cline M, Fang P, Feng Y, Freymueller J, Ge X, Gross WK, Jackson D, Kim M, King NE, Langbein J, Larsen SC, Lisowski M, Shen Z-K, Svarc J, Zhang J (1994) Co-seismic displacements of the 1992 Landers earthquake sequence. *Bull Seismol Soc Am* 84(3):625–645
- Ide S, Baltay A, Beroza GC (2011) Shallow dynamic overshoot and energetic deep rupture in the 2011 Mw 9.0 Tohoku-Oki Earthquake. *Science* 332(1426): 1426–1429. doi:10.1126/science.1207020
- Kanamori H (1977) The energy release in great earthquakes. *J Geophys Res* 82(20):2981–2876
- Kyriakopoulos C, Masterlark T, Stramondo S, Chini M, Bignami C (2013) Coseismic slip distribution for the M9 2013 Tohoku-Oki earthquake derived from 3D FEM modeling. *Journal of Geophysical Research*, doi:10.1002/jgrb.50265
- Liu W, Yamazaki F (2013) Detection of crustal movement from TerraSAR-X intensity images for the 2011 Tohoku, Japan earthquake. *IEEE Geosci Remote Sens Lett* 10(1):199–203
- Massonnet D, Rossi M, Carmona C, Adragna F, Peltzer G, Feigl K, Rabaute T (1993) The displacement field of the Landers earthquake mapped by radar interferometry. *Nature* 364:138–142
- Masterlark T, Hughes KLH (2008) Next generation of deformation models for the 2004 M9 Sumatra-Andaman earthquake. *Geophys Res Lett* 35, L19310. doi:10.1029/2008GL035198
- Michel R, Avouac J-P (2006) Coseismic surface deformation from air photos: the Kickapoo step over in the 1992 Landers rupture. *J Geophys Res* 111:B03408. doi:10.1029/2005JB003776
- Moro M, Chini M, Saroli M, Atzori S, Stramondo S, Salvi S (2011) Analysis of large, seismically induced, gravitational deformations imaged by high-resolution COSMO-SkyMed synthetic aperture radar. *Geology* 39:527–530. doi:10.1130/G31748.1
- Okada Y (1985) Surface deformation due to shear and tensile faults in a half-space. *Bull Seism Soc Am* 75: 1135–1154
- Pritchard ME, Ji C, Simons M (2006) Distribution of slip from 11 $M_w > 6$ earthquakes in the northern Chile subduction zone. *J Geophys Res*. doi:10.1029/2005JB004013
- Rosen Paul A, Hensley S, Joughin IR, Li FK, Madsen SN, Rodríguez E, Goldstein RM (2000) Synthetic aperture radar interferometry. *Proc IEEE* 88(3):333–382
- Stramondo S (2013) The Tohoku-Oki earthquake: a summary of scientific observations from remote sensing. *IEEE Geosci Remote Sens Lett* 10(4):895–897
- Stramondo S, Tesaro M, Briole P, Sansosti E, Salvi S, Lanari R, Anzidei M, Baldi P, Fornaro G, Avallone A, Buongiorno MF, Franceschetti G, Boschi E (1999) The September 26, 1997 Colfiorito, Italy, earthquakes: modeled coseismic surface displacement from SAR interferometry and GPS. *Geophys Res Lett* 26(7):883–886
- Stramondo S, Moro M, Tolomei C, Cinti FR, Doumaz F (2005a) InSAR surface displacement field and fault modelling for the 2003 Bam earthquake (southeastern Iran). *J Geodynamics* 40(2005):347–353
- Stramondo S, Moro M, Doumaz F, Cinti FR (2005b) The December 26th, 2003, Bam (Iran) earthquake: surface displacement from Envisat ASAR Interferometry, *International Journal of Remote Sensing*, vol. 26 (5):1027–1034
- Stramondo S, Chini M, Bignami C, Salvi S, Atzori S (2011a) X-, C-, and L-band DInSAR investigation of the April 6, 2009, Abruzzi earthquake. *IEEE Geosci Remote Sens Lett* 8(1):49–53
- Stramondo S, Kyriakopoulos C, Bignami C, Chini M, Melini D, Moro M, Picchiani M, Saroli M, Boschi E (2011b) Did the September 2010 (Darfield) earthquake

- trigger the February 2011 (Christchurch) event? Sci Rep 1. doi:10.1038/srep00098
- Walters RJ, Elliott JR, D'Agostino N, England PC, Hunstad I, Jackson JA, Parsons B, Phillips RJ, Roberts G (2009) The 2009 L'Aquila earthquake (central Italy): a source mechanism and implications for seismic hazard. *Geophys Res Lett* 36:L17312. doi:10.1029/2009GL039337
- Watanabe M, Motohka T, Miyagi Y, Yonezawa C, Shimada M (2012) Analysis of urban areas affected by the 2011 off the pacific coast of Tohoku earthquake and tsunami with L-Band SAR full-polarimetric mode. *IEEE Geosci Remote Sens Lett* 9(3):472–476
- Zaré M (2003) Seismological aspects of Bam (SE Iran) earthquake of 26 December 2003. Mw6.5: a preliminary reconnaissance report. Website: [http://www.mehdizare.com/lecture-present.htm#NEW%20Journal%20Papers%20\(updated%207%20April%20200](http://www.mehdizare.com/lecture-present.htm#NEW%20Journal%20Papers%20(updated%207%20April%20200)

Residual Strength of Liquefied Soils

Scott M. Olson

Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Synonyms

Critical state shear strength; Liquefied shear strength; Residual shear strength; Shear strength of liquefied soil; Steady state shear strength

Introduction

Liquefaction flow slides are one of the most catastrophic forms of ground failure. Flow failures often result in massive lateral and vertical movements of soil, occasionally for hundreds of meters laterally. While most commonly observed during earthquakes as a result of seismically induced liquefaction, flow failures also have occurred as a result of non-seismic (static) loading, displacements (i.e., shear strains) induced by global instability or creep, and dynamic loading. These failures are driven by static shear stresses that exceed the available shear resistance in the soil after the soil liquefies. The available shear

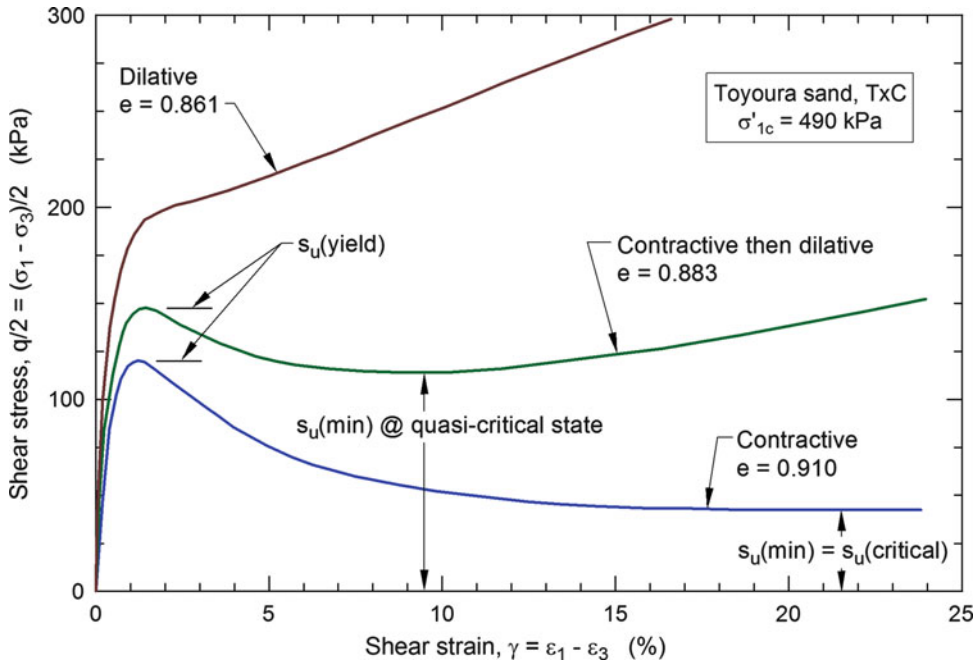
resistance after the soil liquefies, or residual shear strength of the liquefied soil, often is only a fraction of the drained shear strength of the soil. Despite the importance of the residual shear strength in analyzing and mitigating the potential for liquefaction flow failure, estimating this strength remains a difficult and controversial issue for geotechnical engineers. Difficulties largely stem from the inability to cost-effectively retrieve undisturbed samples of loose (i.e., contractive), saturated sandy soils that are most prone to liquefaction flow failure. As a result, empirical procedures constitute the state of practice for estimating the residual strength of liquefied soils, evaluating liquefaction flow failures, and designing remedial measures.

This entry reviews the basic definition of the residual shear strength of liquefied soil and the underlying mechanics that control its mobilization, describes failure mechanisms where the residual shear strength of liquefied soil is mobilized, presents a compilation of residual shear strength measurements from a large laboratory database, and illustrates back-analysis procedures used to interpret residual shear strengths from liquefaction case histories and the resulting empirical correlations that are widely used in practice to estimate site-specific values of the residual shear strength of liquefied soils.

Shear Strength of Liquefied Soil

Basic Behavior and Definitions

The residual shear strength of liquefied soil, or liquefied shear strength [$s_u(\text{liq})$], is defined as the shear strength mobilized at large displacement after liquefaction is triggered in a saturated, contractive, cohesionless (i.e., sandy) soil. In the laboratory, where drainage conditions are controlled, the term undrained (or constant volume) applies. However, some flow failures may experience porewater pressure redistribution or drainage (Stark and Mesri 1992; Fiegel and Kutler 1994; Kulasingam et al. 2004); therefore, the shear strength mobilized in the field may not be undrained. The term liquefied shear strength is used to describe the shear strength actually



Residual Strength of Liquefied Soils,
Fig. 1 Undrained stress–strain response in saturated sands (Data from Verdugo 1992). e , void ratio after consolidation; q , deviator stress; σ_1 , σ_3 , major and minor principal stresses; σ'_{1c} , major principal consolidation

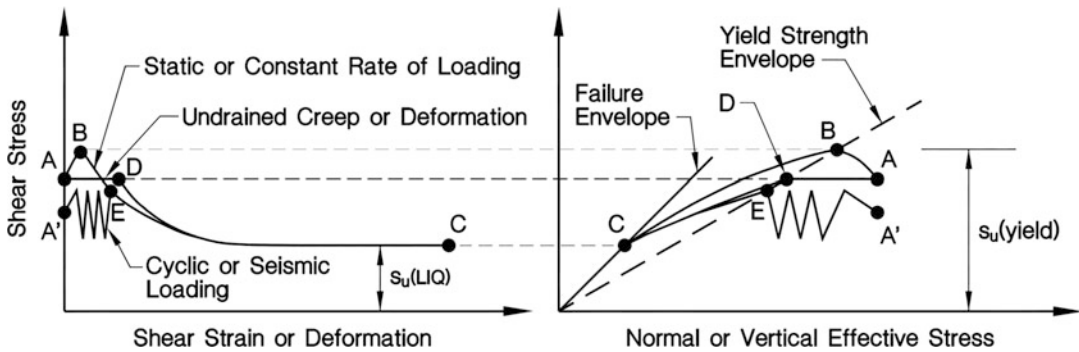
stress; $s_u(\min)$, minimum shear strength; $s_u(\text{critical})$, critical state shear strength; $s_u(\text{yield})$, yield shear strength; ϵ_1 , ϵ_3 , major and minor principal strains (After Olson and Mattson 2008; with permission from Canadian Science Publishing)

mobilized during flow liquefaction, including any potential effects of porewater pressure redistribution/drainage, soil mixing, hydroplaning, etc. Other terms, such as undrained steady state strength (s_{us} , Poulos 1981), undrained residual strength (s_{ur} , Seed and Harder 1990), and mobilized undrained critical strength ($s_u(\text{critical}, \text{mob})$, Stark and Mesri 1992), also have been used to describe the strength of liquefied soil mobilized in the field and in the laboratory. The term liquefied shear strength will be used here.

The liquefied shear strength represents just one aspect of undrained soil behavior. Figure 1 illustrates the response of three saturated Toyoura sand specimens monotonically loaded in undrained triaxial compression. As will be discussed in section “Critical State Framework,” the response of a sand depends on its state (consolidated void ratio and effective stress). Contractive response (or, more specifically, the tendency for contractive response during undrained loading) can be described as the sand

strain-hardens upon initial loading until reaching its yield (or peak) shear resistance; and upon reaching the yield shear resistance, the sand liquefies and strain-softens to a minimum shear resistance. In this case, the minimum shear resistance corresponds to the critical state shear strength, $s_u(\text{critical})$, as defined by Terzaghi et al. (1996) or the undrained steady state shear strength, s_{us} , as defined by Poulos (1981). Contractive undrained response occurs in very loose cohesionless soils and is of primary interest in liquefaction evaluation.

Contractive then dilative (tendency) response is commonly encountered in undrained laboratory testing of sandy soils. This response is similar to contractive undrained response, except that upon reaching a minimum shear resistance, the sand strain-hardens. Alarcon-Guzman et al. (1988) termed this minimum shear resistance the quasi-steady state shear strength based on steady state flow concepts (Poulos 1981); thus, the term quasi-critical state shear strength also



Residual Strength of Liquefied Soils, Fig. 2 Schematic undrained stress–strain and stress path response of saturated, contractive sandy soil (From Olson and Stark 2003; with permission from ASCE)

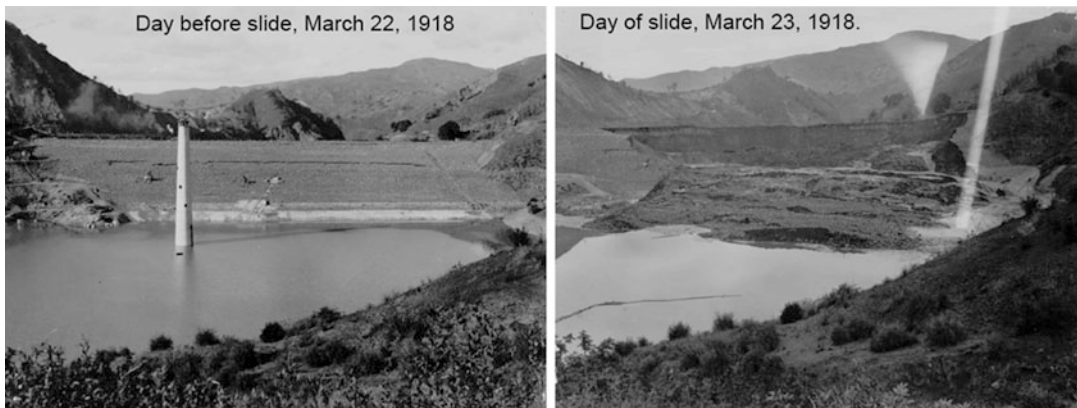
applies. Contractive then dilative undrained response occurs in loose to medium-dense sands and is relevant in liquefaction evaluation.

Dilative (tendency) response consists chiefly of strain-hardening behavior. The specimen reaches neither a yield nor a minimum shear resistance. Similar to the contractive then dilative undrained response, the sand often continues to strain-harden at large strains and may not reach its critical state within the displacement limits of conventional testing equipment. Dilative undrained response occurs in dense sands and generally indicates that the soil is not susceptible to liquefaction flow failure without significant void ratio redistribution, particle damage, or other changes to the soil.

A number of loading scenarios can trigger flow liquefaction where the liquefied shear strength is mobilized. These include (1) static loading, (2) deformation under a sustained shear stress, and (3) dynamic loading, such as seismic shaking. In each case, flow failures can only occur if the shear stress sustained under gravity loading, or static shear stress (τ_{static}), exceeds the liquefied shear strength. Figure 2 illustrates these scenarios, considering an element within a saturated, contractive sandy deposit either within or underlying an embankment during construction. Point A in Fig. 2 represents the prevailing stress and strain conditions in the element. Point A could have been reached by drained, partially drained, or completely undrained loading during embankment construction. During placement of the next fill lift, the element moves from Point

A to Point B, which is located on the yield strength envelope. This step assumes that the drainage boundaries and permeability of the element result in the fill lift causing a temporary undrained condition in the element. Point B represents the maximum shear resistance that the soil element can mobilize under undrained conditions. When the shear stress in this element induced by the embankment attempts to exceed Point B (the yield shear strength), the contractive soil structure yields and collapses, and liquefaction is triggered. The element then moves from Point B to Point C, the liquefied shear strength. Examples of this scenario include the Calaveras Dam and the Jamuna River bridge site failures (Olson 2001; Jefferies and Been 2006).

To illustrate deformation-induced failure, again consider a soil element with stress and strain conditions represented by Point A in Fig. 2. Point A could have been reached by drained, partially drained, or undrained loading, and the static shear stress carried by the element (Point A) is greater than its liquefied shear strength (Point C). In this case, the static shear stress in the soil is large enough to initiate shear strain, creep, or another deformation mechanism within the embankment and/or foundation. If the shear deformation is sufficiently large and element A is temporarily undrained, the element moves horizontally from Point A to Point D, which is located on the yield strength envelope. At Point D, liquefaction is triggered, and the element moves from Point D to Point C, the liquefied shear strength. Examples of this



Residual Strength of Liquefied Soils, Fig. 3 Upstream slope of Calaveras Dam on March 22 and March 23, 1918 (Photos: <http://damsafetyca.blogspot.com/2008/05/424-calaveras-dam.html>)

scenario include the Fort Peck Dam and the Nerlerk berm failures (Olson 2001; Jefferies and Been 2006).

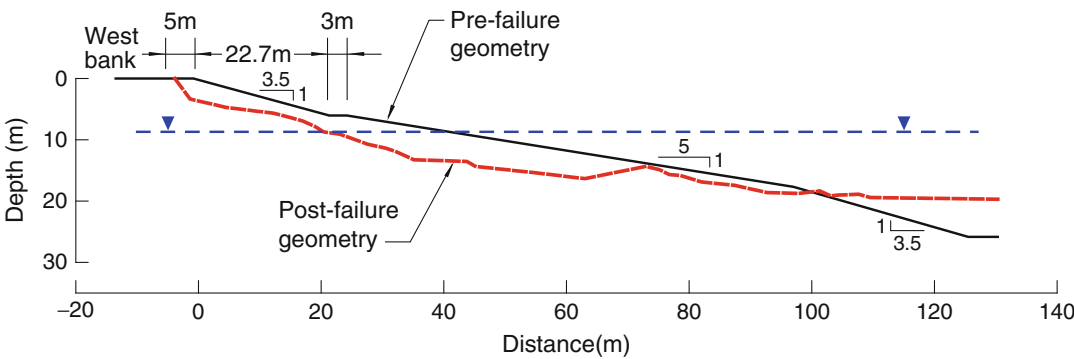
To illustrate dynamic loading-induced flow failure, consider a soil element with stress and strain conditions represented by Point A' in Fig. 2 (this point is not coincident with Point A for clarity only). Point A' could have been reached by drained, partially drained, or undrained loading, and the static shear stress carried by the element (Point A') is greater than its liquefied shear strength (Point C). The element is then subjected to a seismic or dynamic load. If the duration and intensity of the seismic/dynamic load is sufficient to generate porewater pressures that move the element from Point A' to Point E, liquefaction is triggered, and the element moves from Point E to Point C. Examples of this scenario are most numerous in the literature and include seismically induced failures such as Lower San Fernando Dam, Sheffield Dam, and the Kawagishi-cho apartment buildings and dynamic loading-induced failures such as Asele road embankment and Lake Ackerman highway embankment (Olson 2001; Jefferies and Been 2006).

Examples and Applications

As discussed in section “Basic Behavior and Definitions,” the liquefied shear strength can be mobilized in a variety of loading scenarios, including static loading-induced, deformation-induced, and

dynamic loading-induced failures. Mobilization of the liquefied shear strength requires large shear displacements and can result in liquefaction flow failure, loss of bearing capacity, and lateral soil flow around foundations. Figures 3 and 4 present examples of static loading-induced flow failures experienced at Calaveras Dam and the Jamuna River bridge site, respectively. The 1918 Calaveras Dam failure involved a massive flow failure of the hydraulic fill dam during construction. Rapid filling and a rising phreatic surface likely triggered liquefaction within the uncompacted sandy fills (Olson 2001; Jefferies and Been 2006). The 1995 failures at the Jamuna River bridge site were triggered by rapid cuts and large seepage pressures in the fine micaceous riverbed sands (Yoshimine et al. 1999; Jefferies and Been 2006; Ishihara 2008).

Figures 5 and 6 present examples of deformation-induced and dynamic loading-induced failures at Fort Peck Dam and Lake Ackerman highway embankment, respectively. Although failure occurred during construction, the liquefaction flow failure of Fort Peck Dam likely was triggered by global deformation, as the crest of the hydraulic fill dam had settled over 0.5 m in just a few hours prior to the upstream slope failure. This global movement was attributed to sliding along bentonitic shales in the foundation (Casagrande 1965). Liquefaction of the loose, end-dumped sandy fills at the Lake Ackerman highway embankment was triggered



Residual Strength of Liquefied Soils, Fig. 4 Pre- and post-failure geometries of December 3, 1995, flow slide at cross-section 1480 W3 at Jamuna River bridge site, Bangladesh (Modified from Ishihara 2008)



Residual Strength of Liquefied Soils, Fig. 5 Failed upstream slope of Fort Peck Dam on September 22, 1938 (Photo: www.fortpeckdam.com)

Residual Strength of Liquefied Soils,

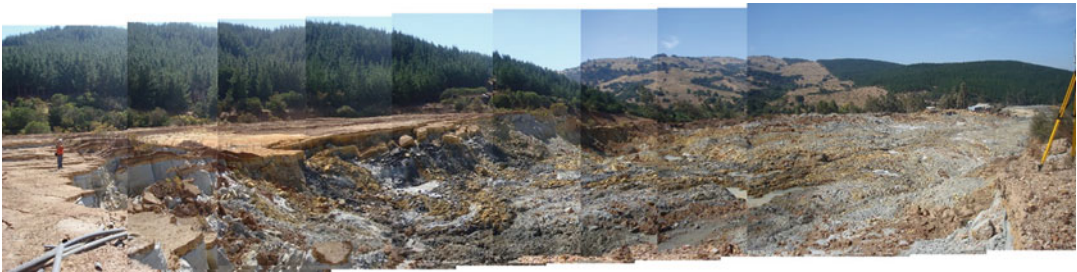
Fig. 6 Failed highway embankment on July 24, 1987, at Ackerman Lake, Michigan, during seismic exploration (Photo: Hryciw et al. 1990)





Residual Strength of Liquefied Soils, Fig. 7 Failed upstream slope of Lower San Fernando Dam after the February 9, 1971, San Fernando earthquake and during

draining of the reservoir (image at the right was taken by H.B. Seed) (Photos: <http://www.gf.uns.ac.rs/~wus/wus07/web4/liquefaction.html>)



Residual Strength of Liquefied Soils, Fig. 8 Failed slope of Las Palmas tailings dam after the February 27, 2010, Chile earthquake

during a 1990 seismic refraction study, resulting in a large flow slide into the lake (Hryciw et al. 1990).

Figures 7 and 8 present examples of seismic loading-induced failures of the Lower San Fernando Dam during the 1971 San Fernando earthquake and the Las Palmas tailings dam during the 2010 Chile earthquake. While the triggering mechanism among these failures differs, the pre-liquefaction static shear stress exceeded the liquefied shear strength in each case. As a result, once liquefaction was triggered, the failure mass began to accelerate downslope and displacements began to accumulate. In many of these cases, the shear strain within the liquefied material greatly exceeded 100 %.

While failures of embankments and dams are most commonly associated with the liquefied shear strength, the significant loss of shear resistance resulting from liquefaction can lead to bearing capacity failures of foundations. Figure 9

illustrates examples of bearing capacity failures resulting from liquefaction of the foundation soils during the 1964 Niigata, 1999 Dulce, 1999 Ismit, and 2010 Haiti earthquakes. While these bearing capacity failures do not involve displacements of the magnitude observed in many flow failures, the shear strains within the liquefied materials likely exceeded 50 %. Furthermore, in each of these cases, the static shear stresses resulting from the applied bearing pressures exceeded the liquefied shear strength in the saturated, contractive foundation soils.

Liquefaction-induced lateral spreading commonly occurs in mildly sloping ground and is the most common form of liquefaction-induced ground failure. Conventionally, lateral spreads are driven by combined static shear stresses and seismically induced shear stresses exceeding the available shear resistance. Because the failure is largely driven by seismic shear stresses, downslope displacements tend to accumulate in a



Residual Strength of Liquefied Soils, Fig. 9 Liquefaction-induced bearing capacity failures of buildings during several earthquakes: (a) 1964 Niigata (Photo: <http://www.ce.washington.edu/~liquefaction>), (b)

1999 Dulce (Photo: http://www.geerassociation.org/GEER_Post%20EQ%20Reports/Duzce_1999), (c) 1999 Ismit (Photo: <http://www.seas.ucla.edu/~wallace/earthquakes.htm>), and (d) 2010 Haiti

ratcheting manner, with displacements occurring when the seismic shear stresses are oriented downslope. Because the static shear stress is less than the liquefied shear strength, displacements cease when the ground motions end. However, in very loose soils (with very low liquefied shear strengths) or if porewater redistribution results in water layer formation, lateral soil flow can continue after ground motions end. In this case, failure (i.e., continued displacements) is driven by static shear stresses alone. While it is often difficult to differentiate these failure mechanisms, eyewitness reports are occasionally available to identify lateral spreads that continue after the end of shaking. Figure 10 illustrates the failure of the Showa Bridge during the 1964 Niigata earthquake which, according to eyewitness reports, failed as a result of post-shaking lateral flow.

Laboratory Characterization Using Critical State Concepts

Critical state soil mechanics was developed based on Casagrande's (1936) laboratory tests on loose (contractive) sands and the definition of the critical void ratio. The critical void ratio was defined as the prevailing global void ratio when shearing continues under constant shear resistance, constant effective confining stress, and constant volume. By the 1950s, the critical void ratio was redefined as a critical state. In sands, a critical state can occur at large shear displacement where the volume, shear stress, and effective confining stress remain constant (Taylor 1948; Roscoe et al. 1958). This state provides an intrinsic reference state for a sand that depends on its initial consolidated void ratio and consolidation stress, and all shear processes from this initial state eventually reach a critical



Residual Strength of Liquefied Soils, Fig. 10 Post-shaking failure of Showa Bridge during 1964 Niigata as a result of lateral soil flow (Photo: <http://www.ce.washington.edu/~liquefaction>)

state. Critical state soil mechanics is, therefore, a useful framework to understand soil behavior, interpret laboratory test results and field soil behavior, and establish failure criteria and post-failure behavior for constitutive models.

Critical State Framework

The locus of critical states for a given soil defines its critical state line. Figure 11 illustrates the critical state line for Batch 7 silty sand from Lower San Fernando Dam. In void ratio–log effective stress space, the critical state line can be approximated as

$$e_{cs} = e_{cs@1\text{kPa}} - \lambda \log \left(\frac{\sigma'}{1\text{ kPa}} \right) \quad (1)$$

where e_{cs} = critical state void ratio, $e_{cs@1\text{kPa}}$ = critical state void ratio at $\sigma' = 1\text{ kPa}$, λ = critical state line slope, and σ' = effective stress. Soil states after consolidation (i.e., consolidated void ratio, e_c , and consolidation stress, σ'_{vc}) can be compared with the critical state line to determine soil volumetric response to shear. Soil states that are above the critical state line tend to contract during shear, while soil states that are below the critical state line tend to dilate during shear. Been and Jefferies (1985) defined a state parameter, $\psi = e_c - e_{cs}$, to evaluate the potential

dilatancy of a soil. Positive values of state parameter correspond to contractive response, while negative values correspond to dilative response.

In void ratio–log effective vertical stress (σ'_v) space, the consolidated void ratio can be defined using the normal consolidation line or the critical state line as (see Fig. 12):

$$\begin{aligned} e_c &= e_{c@1} - C_c \log \left(\frac{\sigma'_{vc}}{1} \right) \\ &= e_{cs@1} - \lambda \log \left(\frac{\sigma'_{v,cs}}{1} \right) \end{aligned} \quad (2)$$

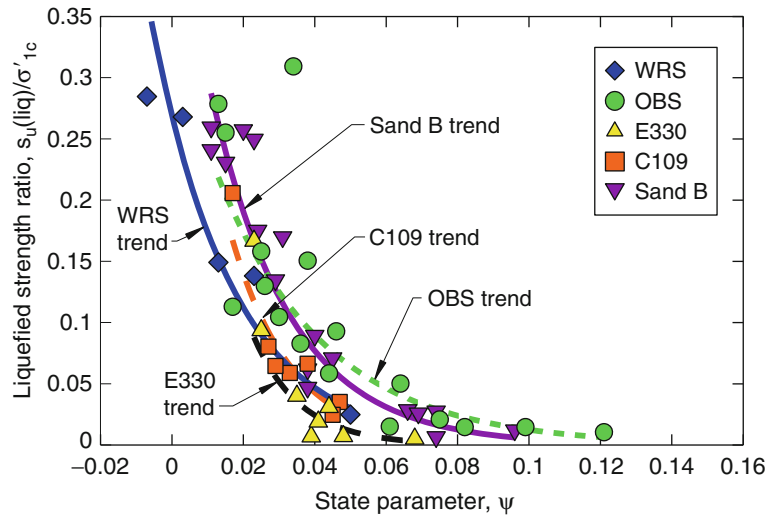
where $e_{c@1}$ = consolidated void ratio at $\sigma'_v = 1$ unit of stress, C_c = compression index, and σ'_{vc} = vertical consolidation stress. Considering the positions of e_c and e_{cs} along the CSL (see Fig. 12), the state parameter can be defined in terms of effective stresses, using effective vertical stresses here:

$$\begin{aligned} \psi &= e_c - e_{cs} \\ &= \left(e_{cs@1} - \lambda \log \sigma'_{v,cs} \right) \\ &\quad - \left(e_{cs@1} - \lambda \log \sigma'_{vc} \right) \\ &= \lambda \log \frac{\sigma'_{vc}}{\sigma'_{v,cs}} \end{aligned} \quad (3)$$

Residual Strength of Liquefied Soils,

Fig. 13 Relationship between state parameter and liquefied shear strength ratio for five sands in Olson (2001) database.

WRS = well-rounded sand (Data from Konrad 1990), OBS = Ottawa banding sand (Data from Dennis 1988), E330 = Erksak 330/0.7 sand (data from Been et al. 1991), C109 = Ottawa C109 sand (Data from Sasitharan et al. 1993, 1994), Sand B (Data from Castro 1969; Castro and Poulos 1977)



strength can be normalized by effective vertical stress to yield a critical state shear strength ratio:

$$\frac{s_u(\text{critical})}{\sigma'_{vc}} = 10^{-\psi/\lambda} \tan \phi'_{cs}. \quad (7)$$

Thus, the critical state shear strength ratio depends chiefly on the state parameter, the critical state line slope, and the critical state friction angle.

Influence of Soil Characteristics on Critical State (Liquefied) Shear Strength

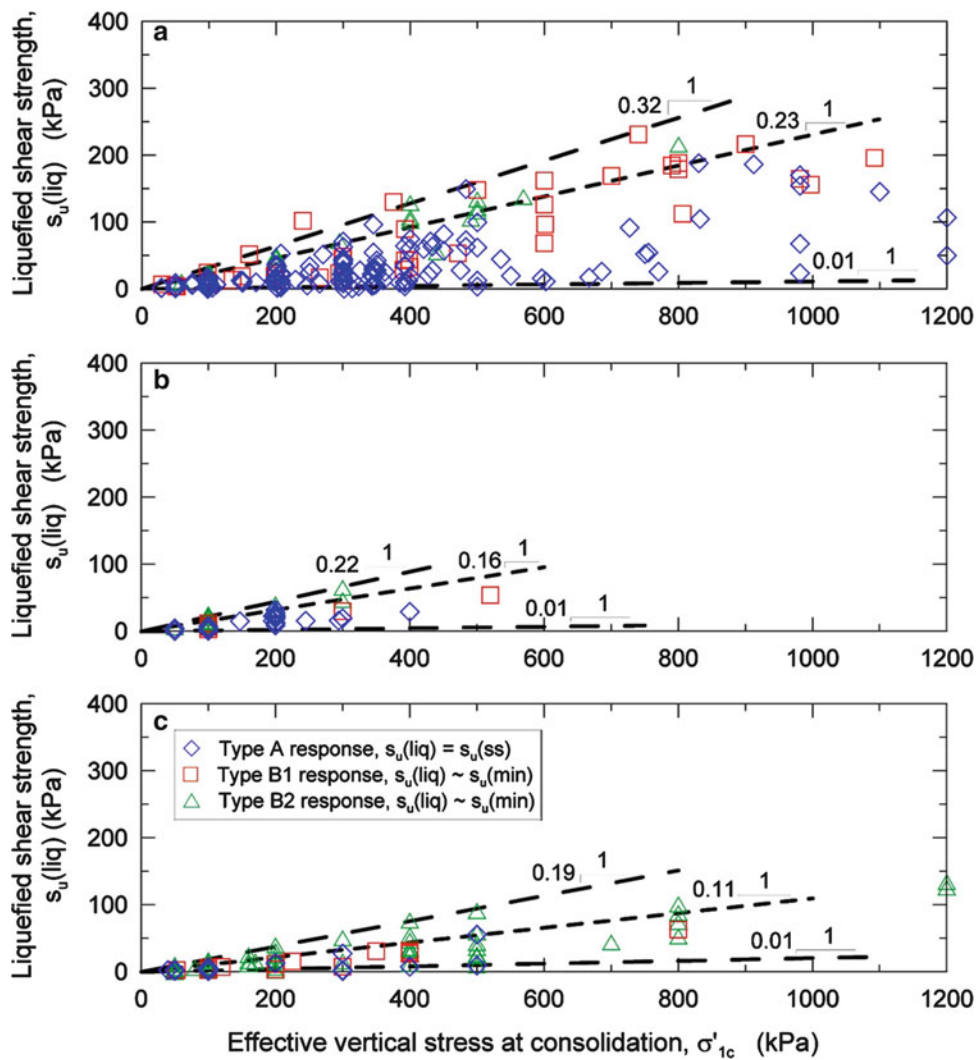
As shown in Eq. 7, the critical state shear strength depends on the state parameter, critical state line slope, and critical state friction angle. Using a large laboratory test database, Olson (2001) illustrated that $s_u(\text{critical})/\sigma'_{vc}$ is inversely related to ψ (see Fig. 13), confirming the relationship in Eq. 7. However, as also illustrated in Fig. 13, the relationship between $s_u(\text{critical})/\sigma'_{vc}$ and ψ is not unique.

The critical state line slope largely depends on gradation, mineralogy, and particle shape (Poulos 1981). All of these factors are related to compressibility. As such, the critical state line slope tends to increase as compressibility increases (Olson and Stark 2003). The critical state friction angle is chiefly a function of particle surface

roughness, which, in turn, largely depends on mineralogy and particle angularity (Sadrekarimi and Olson 2011). Sadrekarimi and Olson (2011) reported that while the critical state friction angle increases by a few degrees at effective stresses less than about 100 kPa, the friction angle remained nearly constant over a wide range of effective stresses (~100–800 kPa). Therefore, the key factors that influence the critical state shear strength ratio are the initial soil state (as quantified by state parameter, ψ) and the soil compressibility (as quantified by the critical state line slope, λ).

Laboratory-Measured Critical State (Liquefied) Shear Strength Ratios

As illustrated in Fig. 13, the critical state (liquefied) shear strength ratio varies uniquely with state parameter for individual sands. However, for positive values of state parameter (i.e., contractive and contractive then dilative responses), liquefied shear strength ratios vary over similar ranges for many sands with varying gradations and particle shapes. Olson and Mattson (2008) collected a database of 386 laboratory triaxial compression, direct simple shear, rotational shear, and triaxial extension test results to examine yield and liquefied strength ratio concepts.



Residual Strength of Liquefied Soils, Fig. 14 Summary of liquefied strength ratio data for (a) triaxial compression, (b) direct simple shear and rotational shear, and (c) triaxial extension tests. Type A = contractive response, Type B1 = contractive then dilative response where minimum shear resistance is maintained over range of shear strain greater than 3 %, Type

B2 = contractive then dilative response where minimum shear resistance is maintained over range of shear strain less than 3 %, $s_u(ss)$ = steady state shear strength, which is equivalent to critical state shear strength (Modified from Olson and Mattson 2008; with permission from Canadian Science Publishing)

Figure 14 summarizes the liquefied strength ratios measured in these tests. For contractive responses (Type A only), liquefied strength ratios range from 0.01 to 0.23 in undrained triaxial compression, 0.01 to 0.16 in undrained (or constant volume) direct simple shear and rotational shear, and 0.01 to 0.11 in undrained triaxial extension. As anticipated, the Type

B (contractive then dilative responses) tests correspond to specimens with lower state parameters, and therefore these specimens generally exhibit larger liquefied strength ratios than Type A (contractive only) specimens. For the Type A specimens, while the maximum values vary, the average liquefied shear strength ratios are similar for all modes of shear.

Field Characterization from Liquefaction Flow Failures

Critical state soil mechanics theory and laboratory undrained shear tests provide a valuable framework for understanding liquefaction behavior and interpreting flow failures in the field. However, it is difficult to retrieve undisturbed samples of sandy soils that are required for site-specific laboratory testing. While ground freezing has been used successfully to obtain nearly undisturbed samples, this process is too expensive for routine practice (and thus has rarely been used). Furthermore, as discussed in section “[Basic Behavior and Definitions](#),” many flow failures may experience porewater pressure redistribution or drainage; therefore, the shear strength mobilized in the field may not be undrained. As a result of these factors, liquefied shear strengths interpreted from field case histories of flow failures are commonly used to estimate site-specific liquefied shear strengths.

Back-Analysis from Liquefaction Flow Failures

Investigators typically have used three types of stability analyses (of increasing complexity) to back-calculate liquefied shear strengths and strength ratios from liquefaction flow failures. The analysis type depended on the detail and quality of information (e.g., pre- and post-failure geometries, phreatic surface location, sliding surface location, material that liquefied, eyewitness observations) available for each case history. For cases with minimal available information, an infinite slope-type analysis commonly is employed. For cases with sufficient information, a rigorous limit equilibrium analysis is employed. However, a limit equilibrium back-analysis of the pre-failure geometry (assuming a factor of safety of unity) results in an unconservative (too high) estimate of liquefied shear strength because, for the flow slide to occur, the liquefied shear strength had to be less than the back-calculated shear resistance. As a result, limit equilibrium analyses are commonly performed for the post-failure geometry. Figure 15 presents an example of a limit equilibrium analysis performed by

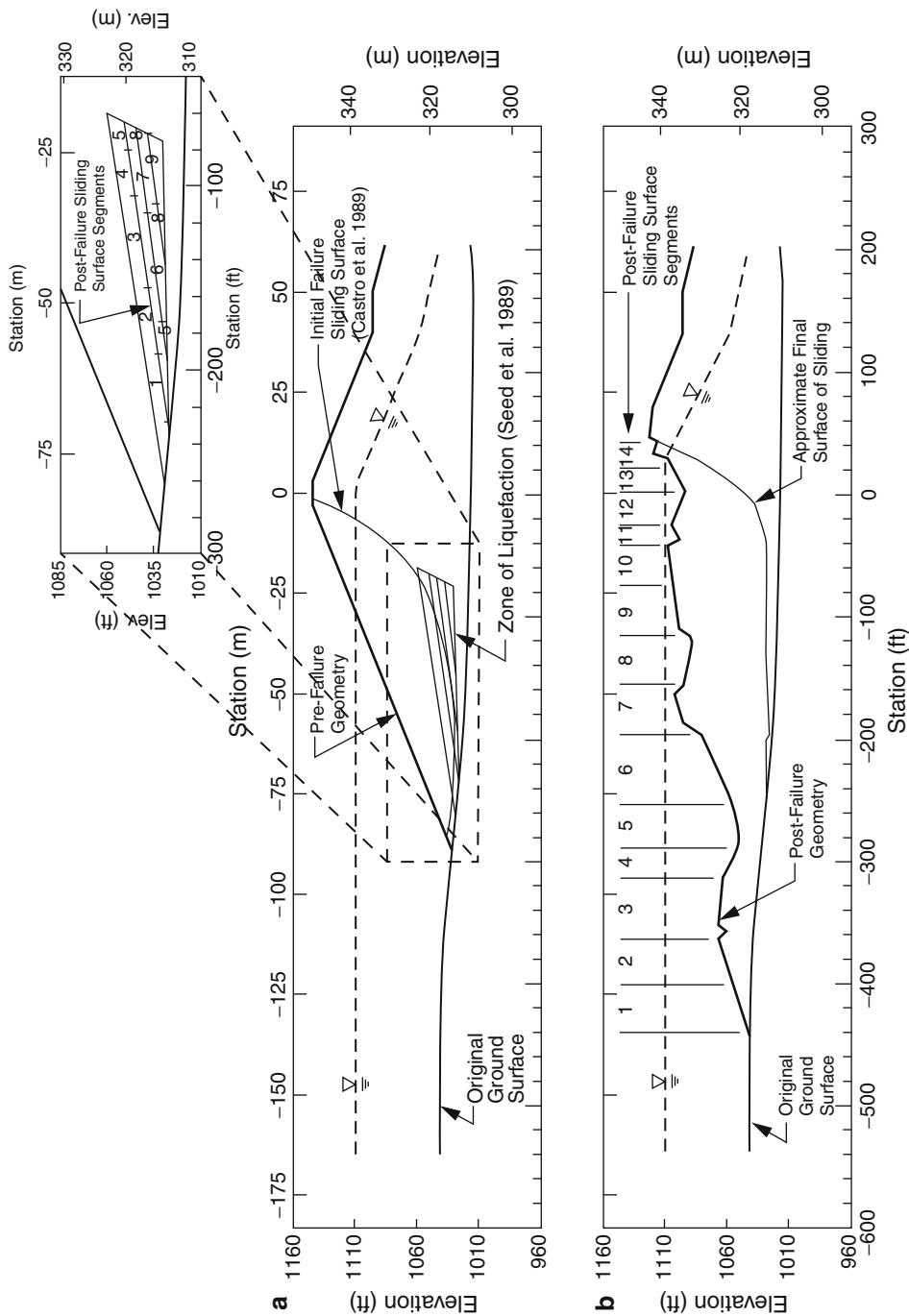
Olson (2001) to back-calculate the liquefied shear strength ratio for the 1971 liquefaction failure of Lower San Fernando Dam.

For cases with appropriate documentation and failure conditions, an additional kinetics analysis can be conducted. As noted above, liquefaction flow failures only occur if the static shear stress exceeds the liquefied shear strength. Because of this, as the failure mass starts to move, its velocity increases. When the failure mass deforms (i.e., moves downslope) sufficiently such that the static shear stress equals the liquefied shear strength, the mass has a finite velocity and will continue to deform and displace. Continued deformation decreases the static shear stress, and the mass decelerates. When the failure mass comes to rest, the static stress often is smaller than the liquefied shear strength, resulting in a factor of safety that is greater than unity. As a result, a limit equilibrium back-analysis of the post-failure geometry that assumes a factor of safety equal to unity results in a conservative (too low) estimate of liquefied shear strength. A kinetics analysis accounts for momentum in the back-analysis and generally is considered to yield the most reasonable back-calculated values of liquefied shear strength and strength ratio.

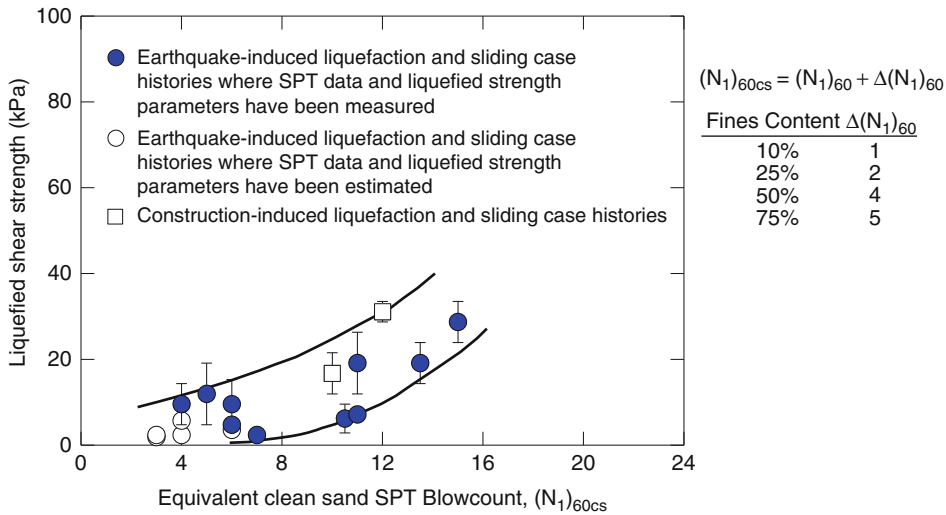
Correlations Based on Back-Analyzed Field Case Histories

Seed (1987) first proposed to back-analyze liquefaction failures to estimate liquefied shear strengths. However, as discussed by Olson and Stark (2002), there were a few limitations and inconsistencies in the back-analyses performed by Seed (1987). Seed and Harder (1990) expanded the work by Seed (1987), and their resulting correlation is still used in practice. Olson (2001) nearly doubled the number of back-analyzed case histories (from 17 to 33) used by Seed and Harder (1990) and performed kinetics analysis for 10 of the 33 case histories with sufficient documentation. This database has been widely used by numerous researchers (e.g., Idriss and Boulanger 2007; Robertson 2010) for evaluating liquefied shear strengths and strength ratios.

To estimate liquefied shear strengths or strength ratios for specific sites, Seed (1987)



Residual Strength of Liquefied Soils, Fig. 15 Example limit equilibrium analysis of Lower San Fernando Dam to back-calculate liquefied strength ratio. (a) Simplified pre-failure geometry used to determine pre-failure effective vertical stresses with initial failure sliding surface from Castro et al. (1989), and (b) simplified post-failure geometry and assumed final positions of liquefied soil segments (segments 1–9) (From Olsson and Stark 2002; with permission from Canadian Science Publishing)



Residual Strength of Liquefied Soils, Fig. 16 Seed and Harder (1990) relationship between liquefied shear strength and equivalent clean sand normalized SPT blow count (Modified from Seed and Harder 1990)

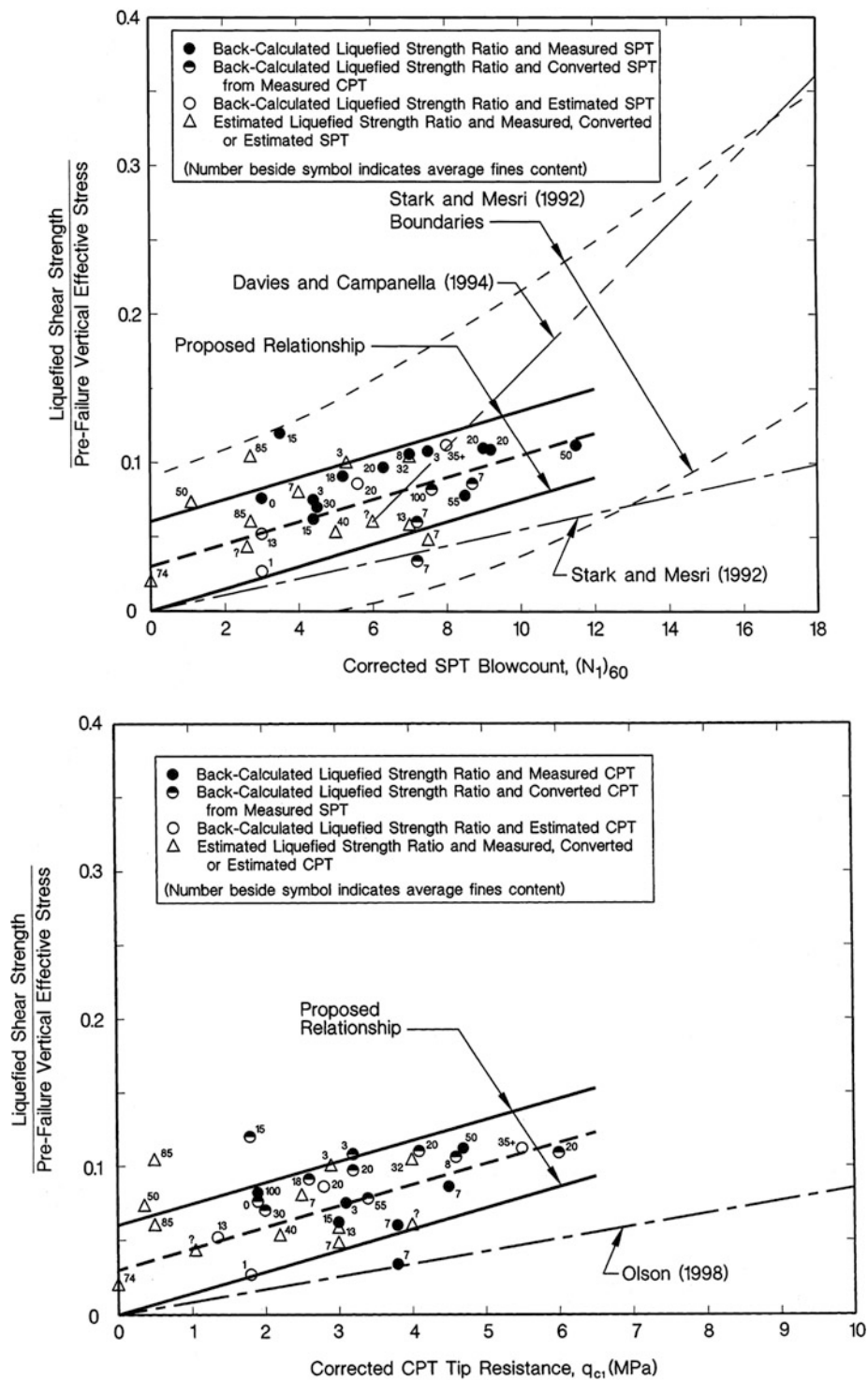
indexed the field case histories using standard penetration test (SPT) blow count. Furthermore, Seed (1987) proposed to adjust the overburden stress-normalized SPT blow count, $(N_1)_{60}$, for the presence of fines. This fines content-adjusted, overburden stress-normalized SPT blow count was termed the equivalent clean sand normalized blow count, $(N_1)_{60cs}$. Other researchers have used cone penetration test (CPT) tip resistance to index the field case histories. Correlations between liquefied shear strength (or strength ratio) and penetration resistance have been proposed because both liquefied shear strength and penetration resistance are functions of soil density and consolidation stress.

Figure 16 presents a correlation proposed by Seed and Harder (1990) between liquefied shear strength and $(N_1)_{60cs}$ based on the back-analysis of 17 liquefaction flow failures and liquefaction-induced lateral spreads. Seed and Harder (1990) updated the work from Seed (1987), adding a few case histories and considering the kinetics of failure for an unknown number of cases. This correlation remains widely used in practice for estimating the liquefied shear strength.

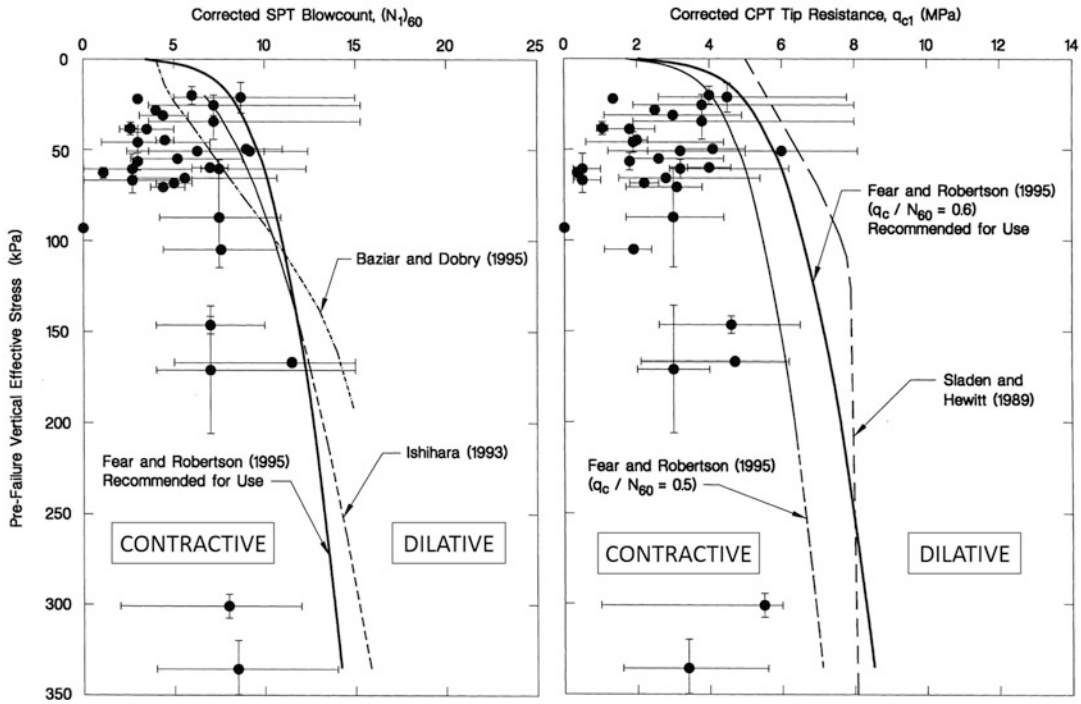
Figure 17 presents correlations proposed by Olson and Stark (2002) between liquefied shear strength ratio and $(N_1)_{60}$ and overburden stress-

normalized CPT tip resistance, q_{c1} . Olson and Stark (2002) did not include a fines content adjustment for penetration resistance, indicating that the liquefied shear strength ratios shown in Fig. 17 did not exhibit any trend with respect to fines content. They suggested that although soils with higher fines contents should exhibit lower values of penetration resistance (as a result of greater soil compressibility and smaller hydraulic conductivity), these soils are more likely to maintain an undrained condition during flow. The combination of these factors may, in effect, offset each other, resulting in no apparent difference in values of liquefied strength ratio for cases of clean sands and sands with higher fines contents.

Although the Olson and Stark (2002) correlations are linear, the relationships are not intended to be extrapolated indefinitely. These correlations were developed for contractive soils that mobilize a liquefied shear strength. As shown in Fig. 18, Olson and Stark (2003) proposed contractive–dilative boundaries for clean sands that limit the penetration resistance to which the liquefied shear strength ratio correlations can be extrapolated. For example, the CPT-based correlation in Fig. 17b is limited to $q_{c1} \leq 6.7$ MPa for $\sigma'_v = 100$ kPa, $q_{c1} \leq 7.7$ MPa for $\sigma'_v = 200$ kPa, and $q_{c1} \leq 8.4$ MPa for $\sigma'_v = 300$ kPa.



Residual Strength of Liquefied Soils, Fig. 17 Olson and Stark (2002) relationships between liquefied shear strength ratio and normalized SPT blow count and normalized CPT tip resistance (Modified from Olson and Stark 2002; with permission from Canadian Science Publishing)



Residual Strength of Liquefied Soils, Fig. 18 Olson and Stark (2003) relationships between liquefied shear strength ratio and normalized SPT blow count and

normalized CPT tip resistance (Modified from Olson and Stark 2003; with permission from ASCE)

Penetration resistance values higher than these contractive–dilative limits correspond to soils with a dilative undrained response where the undrained shear strength increases dramatically. For cases with penetration resistances that correspond to dilative conditions (in Fig. 18), Olson and Stark (2003) recommended that the mobilized shear strength be defined by the drained shear strength of the soil.

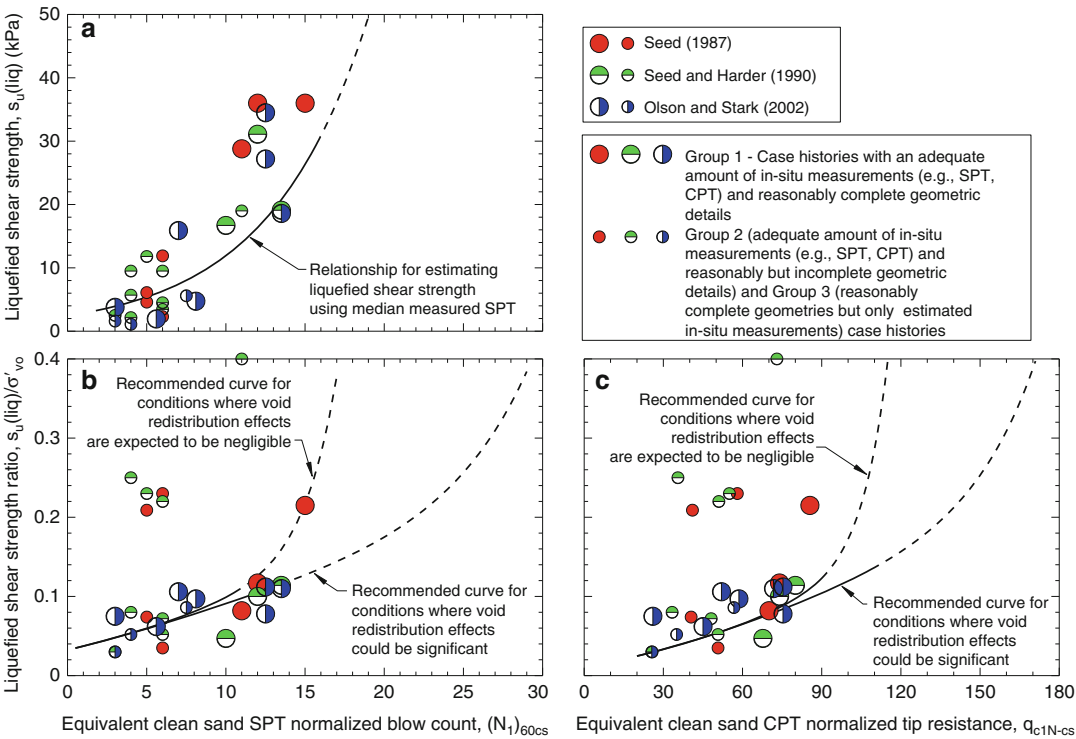
Idriss and Boulanger (2007) combined the back-analyses performed by Seed (1987), Seed and Harder (1990), and Olson and Stark (2002) and proposed correlations between liquefied shear strength and $(N_1)_{60-cs}$, liquefied shear strength ratio and $(N_1)_{60-cs}$, and liquefied shear strength ratio and q_{c1-cs} . Figure 19 presents these correlations. Idriss and Boulanger (2007) attempted to account for the potential for void ratio redistribution (i.e., water layer formation) by recommending conservative correlations for liquefied shear strength ratio, as shown in Fig. 19b and c. They also recommended that the

liquefied shear strength (or strength ratio) be limited to the drained shear strength of the soil.

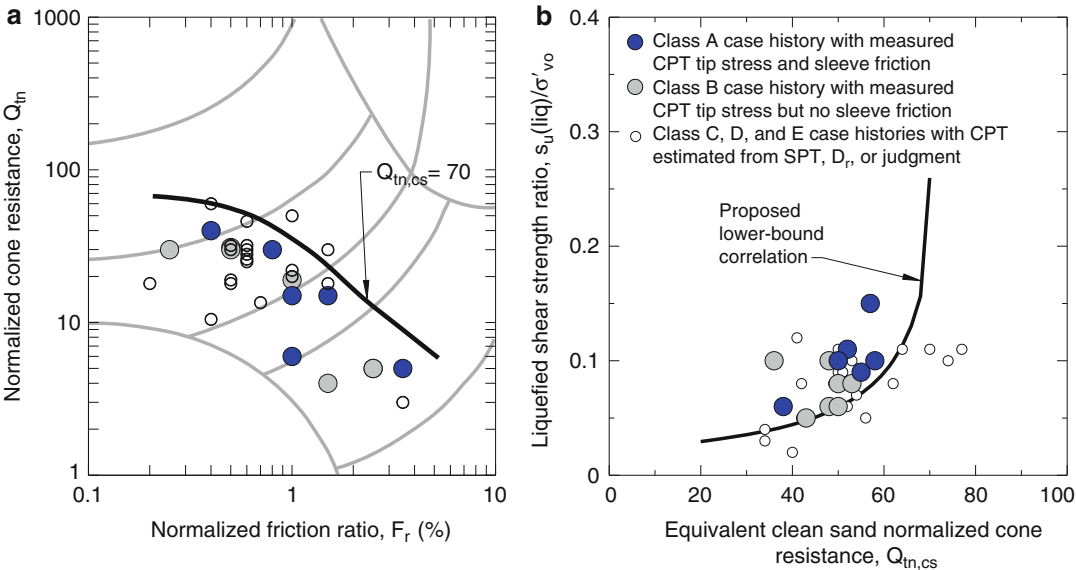
Robertson (2010) added three flow failure case histories to the Olson and Stark (2002) database and proposed a correlation that utilizes cone penetration test measurements (normalized net tip resistance, Q_{tn} , and normalized friction ratio, F_r) to define a contractive–dilative boundary and a correlation to estimate liquefied strength ratio using equivalent clean sand normalized net tip resistance, $Q_{tn,cs}$. These correlations are shown in Fig. 20. Robertson (2010) also recommended that the liquefied shear strength be limited to the drained shear strength of the soil.

Comparison to Laboratory Measurements

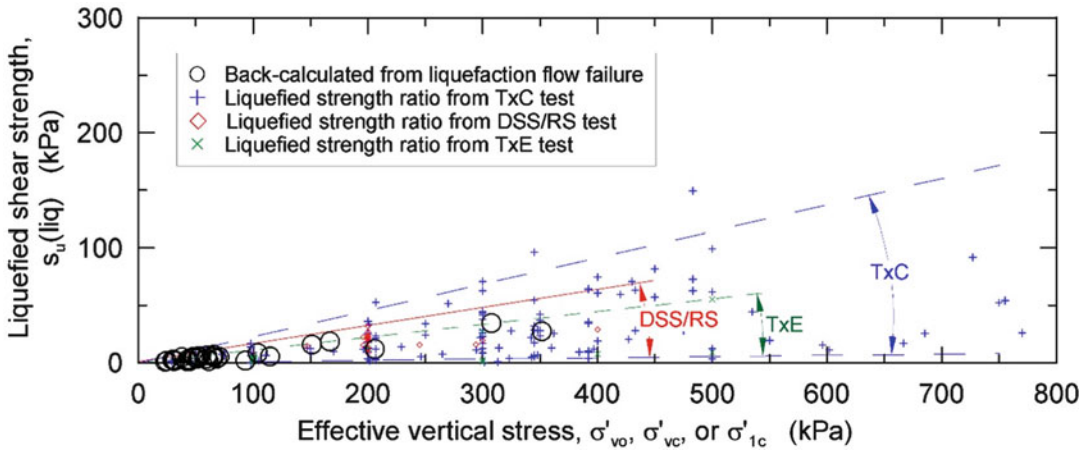
As noted above, the Olson and Stark (2002) back-analyses yielded liquefied shear strength ratios that ranged from approximately 0.05 to 0.12. Figure 21 compares the liquefied strength ratios back-calculated from field case histories with the values



Residual Strength of Liquefied Soils, Fig. 19 Idriss and Boulanger (2007) relationships between (a) liquefied shear strength and $(N_1)_{60cs}$, (b) liquefied shear strength ratio and $(N_1)_{60cs}$, and (c) liquefied shear strength ratio and q_{c1-cs} (Modified from Idriss and Boulanger 2007)



Residual Strength of Liquefied Soils, Fig. 20 Robertson (2010) relationships: (a) CPT-based contractive–dilative boundary and (b) liquefied shear strength ratio and $Q_{tn,cs}$ (Modified from Robertson 2010)



Residual Strength of Liquefied Soils, Fig. 21 Comparison of liquefied strength ratios for field and laboratory data. Flow failure case histories from Olson and Stark (2002). Laboratory data include only Type A (contractive) response where

$s_u(\text{liq}) = s_u(\text{critical})$. DSS direct simple shear, RS rotational shear, TxE triaxial extension, TxC triaxial compression (From Olson and Mattson 2008; with permission from Canadian Science Publishing)

obtained from the Olson and Mattson (2008) laboratory database. This comparison includes only the contractive undrained response (termed Type A) laboratory data because these tests reached unambiguous critical state strengths.

As illustrated in Fig. 21, the laboratory data envelope the case histories, and several case histories exhibit liquefied strength ratios larger than the upper bound for triaxial extension tests. This suggests that a triaxial extension mode of shear did not dominate liquefaction flow in these cases. Furthermore, the range of liquefied strength ratios mobilized in the field cases (0.05–0.12) is relatively small compared to the laboratory ranges, particularly the triaxial compression liquefied strength ratios (0.01–0.23). Importantly, the liquefied shear strengths mobilized in the field cases fall near the middle of the range for the direct simple/rotational modes of shear, supporting the use of effective vertical stress for liquefied shear strength normalization.

Summary

The residual strength of liquefied soil, or liquefied shear strength, is defined as the shear strength mobilized at large displacement after liquefaction

is triggered in a saturated, contractive, cohesionless (i.e., sandy) soil. While this definition is relatively straightforward, estimating the liquefied shear strength in practice remains a difficult task. This largely stems from the high cost and difficulty in retrieving high-quality, undisturbed samples of contractive, saturated sandy soils, as well as ongoing controversy regarding how the liquefied shear strength and critical state line should be determined.

As a result, the standard of practice for estimating the liquefied shear strength is using empirical correlations where liquefied shear strengths or strength ratios (liquefied shear strength divided by effective vertical stress) have been back-calculated from liquefaction failures. In these relationships, back-calculated liquefied shear strengths are correlated to measured or estimated standard or cone penetration test results. However, as a result of uncertainties related to a lack of well-documented case histories, as well as judgments related to the back-analyses and various interpretations of penetration resistance, a number of empirical estimates of liquefied shear strength and strength ratio are available in the literature. These correlations, while widely used in practice, require further research and validation.

Cross-References

- [Dynamic Soil Properties: In Situ Characterization Using Penetration Tests](#)
- [Geotechnical Earthquake Engineering: Damage Mechanism Observed](#)
- [Liquefaction: Countermeasures to Mitigate Risk](#)
- [Liquefaction: Performance of Building Foundation Systems](#)

References

- Alarcon-Guzman A, Leonards GA, Chameau JL (1988) Undrained monotonic and cyclic strength of sands. *J Geotech Eng ASCE* 114(10):1089–1109
- Baziar MH, Dobry R (1995) Residual strength and large-deformation potential of loose silty sands. *J Geotech Eng ASCE* 121(12):896–906
- Been K, Jefferies MG (1985) A state parameter for sands. *Geotechnique* 35(2):99–112
- Been K, Jefferies MG, Hachey J (1991) The critical state of sands. *Geophys J Roy Astron Soc* 41(3):365–381
- Casagrande A (1936) Characteristics of cohesionless soils affecting the stability of slopes and earth fills. *J Boston Soc Civ Eng* 23(1):13–32
- Casagrande A (1965) Second Terzaghi lecture: the role of “calculated risk” in earthwork and foundation engineering. *J Soil Mech Found Div ASCE* 91(SM4):1–40
- Castro G (1969) Liquefaction of sands. PhD thesis, Harvard University, Cambridge
- Castro G, Poulos SJ (1977) Factors affecting liquefaction and cyclic mobility. *J Geotech Eng Div ASCE* 103(GT6):501–516
- Castro G, Keler TO, Boynton SS (1989) Re-evaluation of the lower San Fernando Dam: report 1, the behavior of undrained contractive sand and its effect on seismic liquefaction flow failures of earth structures. US Army Corps’ of Engineers Contract Report GL-89-2, vols 1 and 2, US Army Corps of Engineers Waterways Experiment Station, Vicksburg
- Dennis ND (1988) Influence of specimen preparation techniques and testing procedures on undrained steady state shear strength. In: Donaghe RT, Chaney RC, Silver ML (eds) *Advanced triaxial testing of soil and rock*. ASTM STP 977. American Society for Testing and Materials, Philadelphia, pp 642–654
- Fiegel GF, Kutter BL (1994) Liquefaction induced lateral spreading of mildly sloping ground. *J Geotech Eng ASCE* 120(12):2236–2243
- Hryciw RD, Vitton S, Thomann TG (1990) Liquefaction and flow failure during seismic exploration. *J Geotech Eng ASCE* 116(12):1881–1899
- Idriss IM, Boulanger RW (2007) SPT- and CPT-based relationships for the residual shear strength of liquefied soils. In: Pitilakis KD (ed) *Proceeding of the 4th international conference on earthquake geotechnical engineering*, Springer, New York, pp 1–22
- Ishihara K (2008) Flow slides of underwater sand deposits in Jamuna River bed. In: *Geotechnical engineering for disaster mitigation and rehabilitation*. Science/ Springer, Beijing/Berlin/Heidelberg, pp 3–34
- Jefferies M, Been K (2006) Soil liquefaction: a critical state approach. Taylor & Francis, Oxon, UK, 479 p
- Jefferies MG, Been K, Hachey JE (1990) Influence of scale on the constitutive behavior of sand. In: *Proceeding of the Canadian geotechnical engineering conference*, Laval University, Quebec, vol 1, pp 263–273
- Konrad JM (1990) Minimum undrained strength of two sands. *J Geotech Eng ASCE* 116(6):932–947
- Kulasingam R, Malvick EJ, Boulanger RW, Kutter BL (2004) Strength loss and localization at silt interlayers in slopes of liquefied sand. *J Geotech Geoenviron Eng ASCE* 130(11):1192–1202
- Olson SM (2001) Liquefaction analysis of level and sloping ground using field case histories and penetration resistance. PhD thesis, University of Illinois at Urbana-Champaign, Urbana, 547 p
- Olson SM, Mattson BB (2008) Mode of shear effects on yield and liquefied strength ratios. *Can Geotech J* 45:574–587
- Olson SM, Stark TD (2002) Liquefied strength ratio from liquefaction flow failure case histories. *Can Geotech J* 39:629–647
- Olson SM, Stark TD (2003) Yield strength ratio and liquefaction analysis of slopes and embankments. *J Geotech Geoenviron Eng ASCE* 129(8):727–737
- Poulos SJ (1981) The steady state of deformation. *J Geotechn Eng Div ASCE* 107(GT5):553–562
- Robertson PK (2010) Evaluation of flow liquefaction and liquefied strength using the cone penetration test. *J Geotech Geoenviron Eng ASCE* 136(6):842–853
- Roscoe KH, Schofield AN, Wroth MA (1958) On the yielding of soils. *Geophys J Roy Astron Soc* 8(1):22–53
- Sadrekarimi A, Olson SM (2011) Critical state friction angle of sands. *Geophys J Roy Astron Soc* 61(9):771–783
- Sasitharan S, Robertson PK, Sego DC, Morgenstern NR (1993) Collapse behavior of sand. *Can Geotech J* 30:569–577
- Sasitharan S, Robertson PK, Sego DC, Morgenstern NR (1994) State-boundary surface for very loose sand and its practical implications. *Can Geotech J* 31:321–334
- Seed HB (1987) Design problems in soil liquefaction. *J Geotech Eng Div ASCE* 113(8):827–845
- Seed RB, Harder LF Jr (1990) SPT-based analysis of cyclic pore pressure generation and undrained residual strength. In: *Proceeding of the H. Bolton Seed memorial symposium*, vol 2. Bi-Tech Publishing, Vancouver, BC, Canada, pp 351–376
- Stark TD, Mesri G (1992) Undrained shear strength of liquefied sands for stability analysis. *J Geotech Eng ASCE* 118(11):1727–1747

- Taylor DW (1948) Fundamentals of soil mechanics. Wiley, New York
- Terzaghi K, Peck RB, Mesri G (1996) Soil mechanics in engineering practice, 3rd edn. Wiley, New York, 549 p
- Vasquez-Herrera A, Dobry R (1989) Re-evaluation of the Lower San Fernando Dam: report 3, the behavior of undrained contractive sand and its effect on seismic liquefaction flow failures of earth structures. US Army Corps' of Engineers contract report GL-89-2, US Army Corps of Engineers Waterways Experiment Station, Vicksburg
- Verdugo R (1992) Characterization of sandy soil behavior under large deformation. PhD thesis, Department of Civil Engineering, University of Tokyo, Tokyo
- Yoshimine M, Robertson PK, Wride CE (1999) Undrained shear strength of clean sands to trigger flow liquefaction. *Can Geotech J* 36(5):891–906

Resilience to Earthquake Disasters

John Vargo¹ and Erica Seville²

¹Resilient Organisations, University of Canterbury, Christchurch, New Zealand

²Resilient Organisations, Sheffield, New Zealand

Synonyms

Adaptive; Community resilience; Economic resilience; Individual resilience; Organizational resilience; Planning; Preparedness; Recover; Recovery; Response; Vulnerability

Introduction

Resilience to earthquake disasters should be a goal of any community living in a seismically active landscape. Resilience is the ability to not only survive but also to thrive in the face of adversity – in this case an earthquake. There are many aspects to achieving a resilient community. This entry will first explore the concept of resilience and the use of the term by different disciplines. It then goes on to look at the different systemic layers of resilience that are needed to achieve resilience to earthquake disasters, using examples from the Christchurch, New Zealand, 2010/2011 earthquakes to illustrate. The final

section of the entry looks at opportunities for improving resilience to earthquake disasters.

This entry serves as an introduction to the Disaster Recovery, Reconstruction, and Loss Modeling section of the Encyclopedia of Earthquake Engineering. Throughout the entry, readers are referred to other papers within the section where more information can be found on particular topics.

What is Resilience?

The recent popularity of the word resilience has led to diverse applications of the term – resilience seems at times to mean everything to everyone. Within the literature, there are many definitions of resilience, a selection of which are represented in Table 1. Why so many different definitions? The primary reason is that different disciplines use terms in ways that are pertinent to their field of study, both from a content and from a unit of analysis standpoint. In addition, some definitions and disciplines may take a process focus while others take an outcomes focus or speed of outcomes focus to viewing resilience. For example, ecology is concerned with the populations in the ecosystem and their ability to absorb change, while sociology and engineering are concerned with social and physical systems, respectively, and their ability to respond when disruption occurs.

In spite of the differences, however, there are many commonalities to the definitions of resilience provided in the literature. Virtually all of the definitions have certain characteristics in common:

- Resilience is a feature of a system (physical or social).
- The system is able to respond to change, disruption, or hazards through resistance, absorption, or adaptation or a combination of these.
- The system is able to continue to function at some level following the disruption: returning to its original state or potentially an adapted and improved state.

So although the language may vary by discipline, the fundamentals are largely the same.

Resilience to Earthquake Disasters, Table 1 Definitions of resilience from different disciplines (Kachali 2013)

Author(s)	Academic/research discipline	Definition
Holling	Ecology	A measure of the persistence of systems and of their ability to absorb change and disturbance and still maintain the same relationships between populations or state variables (Holling 1973)
Horne and Orr	Organizational and human resource development	Resilience is a fundamental quality of individuals, groups and organizations, and systems as a whole to respond productively to significant change that disrupts the expected pattern of events without engaging in an extended period of regressive behavior (Horne and Orr 1998)
Perrings	Environmental and resource economics	Resilience is a measure of the ability of a system to withstand stresses and shocks – its ability to persist in an uncertain world (Perrings 1998)
Comfort	Public and international affairs (public policy analysis)	The capacity to adapt existing resources and skills to new systems and operating conditions (Comfort et al. 1999)
Petak	Public administration	The system’s ability to make a smooth transition to a new stable state in response to the disturbance (Petak 2002)
Bruneau et al.	Earthquake engineering (community disaster resilience)	The ability of social units (e.g., organizations, communities) to mitigate hazards, contain the effects of disasters when they occur, and carry out recovery activities in ways that minimize social disruption and mitigate the effects of future earthquakes (Bruneau et al. 2003)
Tierney	Sociology	A property of physical and social systems that enables them to reduce the probability of disaster-induced loss of functionality, respond appropriately when damage and disruption occur, and recover in a timely manner (Tierney 2003)
Christopher and Peck	Logistics	The ability of a system to return to its original state or move to a new more desirable state after being disturbed (Christopher and Peck 2004)
Rose	Economics	The ability or capacity of a system to absorb or cushion against damage or loss. . . (more general definition) that incorporates dynamic considerations, including stability, is the ability of a system to recover from a severe shock (Rose 2004)
Walker et al.	Ecology	The capacity of a system to absorb disturbance and reorganize while undergoing change so as to still retain essentially the same function, structure, identity, and feedbacks (Walker et al. 2004)
Hollnagel et al.	Engineering	The ability of systems to anticipate and adapt to the potential for surprise and failure (Hollnagel et al. 2006)
Seville et al.	Business and engineering	An organization’s ability to survive, and potentially even thrive, in times of crisis (Seville et al. 2008)
UNISDR	Disaster reduction	The ability of a system, community or society exposed to hazards to resist, absorb, accommodate to and recover from the effects of a hazard in a timely and efficient manner, including through the preservation and restoration of its essential basic structures and functions (UNISDR 2009)

Resilience Required at Many Levels

For a community to both survive and be able to thrive following an earthquake, it is important for them to be resilient in all aspects vital to how that community functions. There are many different lenses by which a community’s resilience to

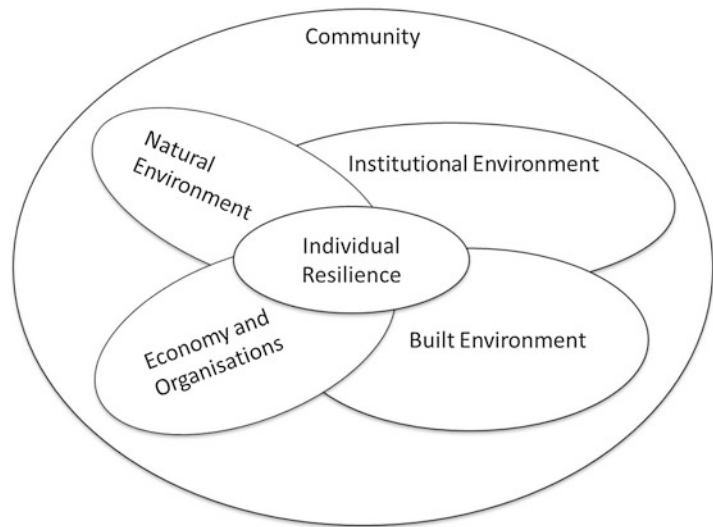
earthquakes could be considered; for the purposes of this entry, the framework in Fig. 1 is used.

Individual Resilience

How can an organization or community be resilient if the individuals that are in it are fragile?

Resilience to Earthquake Disasters,

Fig. 1 Different lenses that can be applied to a community's resilience



Clearly, the resilience of individuals is a key aspect to the resilience of communities and organizations. Individuals with high social capital, an optimistic outlook, good coping skills, sound physical health, and an ability to adapt will have increased levels of resilience. These characteristics can be supported and enabled by the community in which they live through mentoring, wise colleagues at work, hope inspiring leaders, and a community that encourages participation. The resulting community cohesion and a willingness to help each other under stressful circumstances have direct impacts on physical and mental health, stamina, and ability to recover in the aftermath of an earthquake.

Some individuals will have a greater capacity to “bounce back” than others. This capacity is determined by a range of characteristics that make some individuals more resilient and others more vulnerable. The entry “► [Community Recovery Following Earthquake Disasters](#)” highlights the fact that many communities, and the individuals within them, bring certain risk and protective factors to potential disaster situations. This combination of risks and resiliencies interacts with hazard and exposure variables to determine the extent to which each person and the communities of which they are part are affected. Within an affected community, different groups and individuals may experience the disaster in a multitude of ways.

Although most people will experience some psychosocial reaction to a disaster, the majority of people do recover from disaster events, with time and basic supports.

Bonanno et al. (2007) found a series of factors that support resilience in a range of settings, including age, sex, education, circle of friends and family, employment skills, location, and physical and mental health. All of these factors contribute in varying ways to individual resilience.

The Built Environment

The built environment provides shelter, workplace, community facilities, as well as essential infrastructure such as water, waste disposal, power, telecommunications, and transport. A modern community is dependent on this built environment to support the daily social, commercial, and cultural functioning of life. It is also this built environment that takes such a punishment when an earthquake strikes resulting in impassable roads, loss of electricity, “boil-water” notices, and collapsed buildings.

As highlighted in the “► [Damage to Buildings: Modeling](#)” entry, damage to the built environment is the root cause of many consequential impacts of earthquakes on a community. The potential for both structural and nonstructural damage to buildings and their contents leads to the potential for deaths and injuries, direct costs

for repair or replacement of the building and contents, as well as indirect losses that emerge as a result of the disruption caused by the damage.

For infrastructure systems, it is not just the damage to the individual components that is important for understanding community level impacts but also how the whole system consequently performs. As highlighted in the “► [Damage to Infrastructure: Modeling](#)” entry, it is essential also to understand the connections between these components and their spatial distribution relative to urban populations, since some components will be more critical than others in terms of their impact on system performance.

Brunsdon et al. (2013) bring another perspective, suggesting that resilience of the built environment be considered as comprising the following four elements:

- Robust assets capable of maintaining both low damage in more frequent events and life safety in major events
- Planning and preparation for failure, including key components and network routes having appropriate alternatives as well as having effective response arrangements in place
- Effective coordination of regulatory, design, and construction processes, to both ensure quality and to facilitate repair and reconstruction
- Realistic end-user expectations, with both owners and users understanding the risk and performance objectives and appropriate measures of backup arrangements

The key point from these elements is that resilience involves more than just the resilience of physical assets. Improving end-user knowledge of how the built environment is likely to perform in an earthquake, and encouraging users (particularly organizations with emergency response roles such as hospitals) to plan for a level of disruption in extreme events is also essential.

Economy and Organizations

The economy for a community consists of both the economic environment and the organizations

that populate that environment. The laws and regulations, the lack of corruption in government and commercial dealings, the taxation system, critical infrastructure, economic sector critical mass, and the competitive market all contribute to the economic environment. The entries “► [Economic Recovery Following Earthquakes Disasters](#)” and “► [Economic Impact of Seismic Events: Modeling](#)” together provide a detailed analysis of the ways in which earthquakes can affect the functioning of an economy, at both micro- and macro-levels.

Organizations provide employment, goods, and services and gathering places for the individuals within a community. The way in which a community’s organizations function within a disaster-disrupted environment contributes to both the state of economic resilience and the community’s overall resilience. In a disaster such as a large earthquake, organizations provide a rallying point where people can find encouragement from their work colleagues, a distraction from a badly damaged home, and solace from the loss of a loved one.

Organizational resilience can be conceptualized as (1) the adaptive capacity of an organization that is created by leadership and culture, (2) the internal and external relationships and networks that an organization draws on in both business as usual and when a crisis happens, and (3) the planning that is done to develop a clear direction that enables an organization to be change ready (Resilient Organisations 2013). So how does the economic environment support such organizational resilience? An example is given below of the Earthquake Support Subsidy (ESS).

The ESS was a central government initiative that was put in place shortly after the Christchurch earthquakes to provide a wage subsidy for businesses affected by the earthquakes that had fewer than 50 employees (more than 95 % of New Zealand businesses have fewer than 50 employees). The subsidy was provided for 6 weeks, with very low compliance costs to access the subsidy. The subsidy was given directly to the employer so that wages could be paid. The effect of the ESS was to provide time

for employers to get beyond the “fight or flight” stage of earthquake response and to do some forward planning for how they might recover. It provided assurance to employees that they would still be paid. Most importantly, in the critical first few weeks following the earthquakes, it maintained the employer-employee relationship. This initiative prevented the loss of jobs as well as a loss of employers. This in turn meant loss of population was minimized. This juncture of the economic environment and organizations was effective in maintaining economic activity, organizational coherence, and community resilience.

Institutional Resilience

Resilience is also needed within the institutional environment of a community. The institutional environment is the environment in which an individual, a community, or an organization functions and the supporting mechanisms that provide the boundaries for that functioning. The legal frameworks, building codes, political structures, insurance setting, and a range of other institutions and processes all provide this environment. The nature of these institutional arrangements can provide enhanced resilience, resources, and safety or add to the fragility of the wider community. Effective emergency services, a robust banking system, a proactive local council, an honest police force, and the availability of inexpensive earthquake insurance are all examples of institutional resilience.

An exemplar of this from the Christchurch earthquakes is the New Zealand Earthquake Commission (EQC). This quasi-government organization provides an earthquake insurance scheme for residential housing. The insurance is mandatory if the dwelling has fire insurance and as a consequence of the resulting high market penetration (95 % uptake) is very cost-effective. A consequence of this institutional environment is that following the Christchurch earthquakes, 95 % of homes were insured and will be repaired or rebuilt following this devastating event.

Natural Environment

The habitat that a community is situated in has an impact on its overall resilience. Availability of

clean water, fresh air, and uncontaminated land are a few of the things that we take for granted – but are essential to a healthy community which is able to respond effectively in the aftermath of a large earthquake. Following the 2010/2011 earthquake series in Christchurch, New Zealand, there were significant impacts on the natural environment. Substantial liquefaction, lateral spreading, and land subsidence are impacted on both the appearance and livability of some sections of the city. In areas deemed highly prone to liquefaction and lateral spreading, the land was “red zoned” with all houses within these areas slated for demolition, with no permitted re-habitation. In areas of significant land subsidence (in some areas in excess of a half meter), localized flooding has become an endemic problem now requiring judgments on habitability and flood protection.

Further impacts of earthquakes on the natural environment include issues of demolition debris, contaminated liquefaction, and other products of the devastation caused by seismic events. As highlighted in the entry “► [Waste Management Following Earthquake Disaster](#),” effective management of such waste disposal and recycling can either preserve or significantly undermine the natural environment. These impacts on the natural environment have flow on effects on the local population’s health, well-being, and resilience.

Community Resilience

A resilient community is one that can cope with, adapt to, and recover from the loss and disruption encountered through the experience of a disaster. As highlighted in the entry “► [Community Recovery Following Earthquake Disasters](#),” how effectively this is done is a function of how well people, communities, and societies can work together and use their resources to deal with the problems encountered. Resilience of individuals and the built environment, organizations, and the economy of institutions and the natural environment all contribute to the resilience of a community.

Paton and Johnston’s (2006) model of community resilience emphasizes the importance of

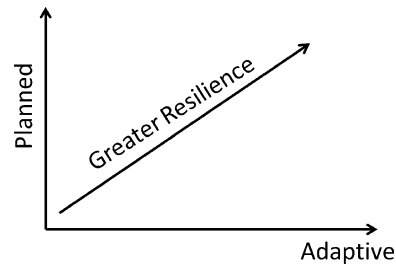
community development and participation. Aldrich (2012) asserts that social capital is a critical element of resilience and recovery for a community. Both of these point to the fact that it is the people and their interactions that are at the foundation of community resilience and the ability to recover from disasters. The built environment and economy, the organizations and institutions, and the natural environment are all there as supporting elements of this foundation, but in the end, as an old Maori (New Zealand) saying goes, “it is about the people, the people and the people.”

Improving Resilience

To be resilient to disasters, people, organizations, and communities need to be able to successfully develop both planned and adaptive resilience capabilities (Fig. 2).

Planning is needed to understand and, where possible, to mitigate the potential impacts of an earthquake on a community. Adaptive capabilities are needed to enable the community to respond creatively to the unexpected challenges that inevitably occur when disasters strike. Ideally, communities will hold both planned and adaptive capabilities in equal measure. Communities that fail to plan sufficiently for the risks they face can find themselves in a cycle of “reacting” to potential disasters. Communities that become too rigid in their thinking can restrict their ability to act and to change as circumstances require.

As discussed in the entry “► [Sustained Earthquake Preparedness: Functional, Social, and Cultural Issues](#),” the goal of planning activities should be to increase the likelihood of people and communities being in a position to be able to respond in planned and functional ways to the complex, challenging, emergent challenges and demands that earthquakes create, rather than having to react to them in ad hoc ways. But a community should also see that the benefit of planning comes from skills and capabilities developed through the planning process, rather than the main outcome being a “plan.” They



Resilience to Earthquake Disasters, Fig. 2 Both planned and adaptive capabilities are required for greater resilience

should consciously plan for how to develop their adaptive skills and capabilities – so that if there is a need to “react,” then they can adjust or develop new plans as required.

Planning should cover the 4 R’s of emergency management: reduction, readiness, response, and recovery (also known in some countries as prevention, preparedness, response, and recovery).

One of the first steps for a community in planning for earthquake disasters is to understand the risk to which they are exposed. This encyclopedia includes a number of entries relating to modeling, including ways a community can better understand the level of seismic hazard to which it is exposed using ► [Probabilistic Seismic Hazard Models](#) and how it can model potential impacts on community functions (e.g., “► [Damage to Buildings: Modeling](#),” “► [Damage to Infrastructure: Modeling](#),” and “► [Economic Impact of Seismic Events: Modeling](#)”).

Communities can also plan to be ready to both respond and recover from disaster. The most obvious focus for such planning is around ► [Emergency Response for Earthquake Disasters](#), but there are also other aspects, particularly related to recovery, that also benefit from prior planning. As discussed earlier, damage to the built environment in an earthquake is the root cause of many consequential impacts of earthquakes on a community. Planning for how to deliver an effective and efficient post-disaster reconstruction program is important. There are many different issues to consider in planning for ► [Reconstruction Following Earthquake Disasters](#). Of particular importance is ► [Earthquake](#)

Disaster Recovery: Leadership and Governance. There is also the potential need to change laws and regulations to facilitate a rapid and effective rebuild (► [Legislation Changes Following Earthquake Disasters](#)). In some circumstances, it may also be necessary to reevaluate ► [Land use planning following earthquake disasters](#) to understand if a community “should” rebuild in the same location. There is a need to consider how to deal with waste generated by the earthquake (► [Waste Management Following Earthquake Disaster](#)) and the resourcing implications that a reconstruction might have on the broader economy (► [Resourcing Issues Following Earthquake Disaster](#)).

Reconstruction takes time, in many cases decades, so it is important to also consider the transitional issues likely to emerge while the reconstruction and recovery takes place. For example, there is a need to plan for how to provide interim housing for people whose homes are damaged by the earthquake (► [Interim Housing Provision Following Earthquake Disaster](#)).

Resilience is also about finding the “silver lining” – seeking out the opportunities that always arise during a disaster, with the goal of emerging stronger and better than before. Throughout the literature, there is a concept emerging of the need to not just “bounce back” from disasters but to use disasters as an opportunity to “bounce forward” (Manyena et al. 2011), rebuilding communities to be better and more resilient than before (► [“Build Back Better” Principles for Reconstruction](#)). As earthquake disasters are relatively rare, it can often be difficult to build momentum for mitigation and preparedness activities. It is important therefore to leverage the disasters that do occur to focus collective attention and effort on improving resilience and ► [Learning from Earthquake Disasters](#).

Summary

This entry provides a high level overview of what it takes for a community to become “resilient” to earthquake disasters. Resilience is a term that is used within a number of disciplines,

all with slightly different definitions. At its core though, resilience refers to the ability of a system to be able to respond to change and to be able to continue to function at some level following the disruption. Resilience is not only about being able to survive a disaster but also requires being able to thrive in an ever changing environment. Communities situated in seismically active areas may have little ability to control the frequency or intensity of earthquakes they are exposed to. Communities can however make proactive efforts to reduce their vulnerabilities to earthquakes and to prepare themselves to both respond and recover from earthquakes when they do occur. Our communities need to become resilient to earthquake disasters.

Cross-References

- [“Build Back Better” Principles for Reconstruction](#)
- [Building Disaster Resiliency Through Disaster Risk Management Master Planning](#)
- [Community Recovery Following Earthquake Disasters](#)
- [Damage to Buildings: Modeling](#)
- [Damage to Infrastructure: Modeling](#)
- [Earthquake Disaster Recovery: Leadership and Governance](#)
- [Earthquake Risk Mitigation of Lifelines and Critical Facilities](#)
- [Earthquakes and Their Socio-economic Consequences](#)
- [Economic Impact of Seismic Events: Modeling](#)
- [Economic Recovery Following Earthquakes Disasters](#)
- [Emergency Response Coordination Within Earthquake Disasters](#)
- [Interim Housing Provision Following Earthquake Disaster](#)
- [Land Use Planning Following an Earthquake Disaster](#)
- [Learning from Earthquake Disasters](#)
- [Legislation Changes Following Earthquake Disasters](#)
- [Probabilistic Seismic Hazard Models](#)

- ▶ [Reconstruction Following Earthquake Disasters](#)
- ▶ [Reconstruction in Indonesia Post-2004 Tsunami: Lessons Learnt](#)
- ▶ [Resiliency of Water, Wastewater, and Inundation Protection Systems](#)
- ▶ [Resourcing Issues Following Earthquake Disaster](#)
- ▶ [Sustained Earthquake Preparedness: Functional, Social, and Cultural Issues](#)
- ▶ [Waste Management Following Earthquake Disaster](#)

References

- Aldrich DP (2012) Building resilience: social capital in post-disaster recovery. University of Chicago Press, Chicago
- Bonanno GA, Galea S, Bucciarelli A, Vlahov D (2007) What predicts psychological resilience after disaster? The role of demographics, resources, and life stress. *J Consult Clin Psychol* 75(5):671–682
- Bruneau M, Chang SE, Eguchi RT, Lee GC, O'Rourke TD, Reinhorn AM, von Winterfeldt D (2003) A framework to quantitatively assess and enhance the seismic resilience of communities. *Earthq Spectra* 19:733
- Brunsdon D, Vargo J, Seville E (2013) Resilient New Zealand, Aon Benfield hazards conference, 23–24 Sept 2013, Gold Coast Australia [2013–14]
- Christopher M, Peck H (2004) Building the resilient supply chain. *Int J Logist Manag* 15(2):1–14
- Comfort LK, Wisner B, Cutter S, Pulwarty R, Hewitt K, Oliver-Smith A, Krimgold F (1999) Reframing disaster policy: the global evolution of vulnerable communities. *Environ Hazards* 1(1):39–44
- Holling CS (1973) Resilience and stability of ecological systems. *Annu Rev Ecol Syst* 4:1–23
- Hollnagel E, Woods DD, Leveson NG (2006) Resilience engineering: concepts and precepts. Ashgate, Aldershot
- Horne JF, Orr JE (1998) Assessing behaviors that create resilient organizations. *Employ Relat Today* 24:29–40
- Kachali H (2013) Key element of sectoral recovery and resilience after the Canterbury earthquakes: a systems dynamics approach. PhD thesis, University of Canterbury. <http://www.resorgs.org.nz/Publications/research-theses.html>. Accessed 8 Dec 2013
- Manyena SB, O'Brien G, O'Keefe P, Rose J (2011) Disaster resilience: a bounce back or bounce forward ability? *Local Environ* 16(5):417–424
- Paton D, Johnston D (2006) Disaster resilience: an integrated approach. Charles C. Thomas, Springfield
- Perrings C (1998) Resilience in the dynamics of economy-environment systems. *Environ Resour Econ* 11(3):503–520
- Petak W (2002) Earthquake resilience through mitigation: a system approach. International Institute for Applied Systems Analysis, Laxenburg
- Resilient Organisations (2013) www.resorgs.org.nz. Accessed 22 Apr 2014
- Rose A (2004) Defining and measuring economic resilience to disasters. *Disaster Prev Manag* 13(4):307–314
- Seville EP, Brunsdon D, Dantas A, Le Masurier J, Wilkinson S, Vargo JJ (2008) Organisational resilience: researching the reality of New Zealand organisations. *J Bus Contin Emer Plan* 2(2):258–266
- Tierney KJ (2003) Conceptualizing and measuring organizational and community resilience: lessons from the emergency response following the 11 Sept 2001. Attack on the World Trade Center
- UNISDR (2009) <http://www.unisdr.org/eng/terminology/terminology-2009-eng.html>. Accessed 22 Apr 2014
- Walker B, Holling CS, Carpenter SR, Kinzig A (2004) Resilience, adaptability and transformability in social-ecological systems. *Ecol Soc* 9(2):5

Resiliency of Water, Wastewater, and Inundation Protection Systems

Craig A. Davis

Waterworks Engineer, Los Angeles Department of Water and Power, Los Angeles, CA, USA

Synonyms

Infrastructure service categories; Lifeline system earthquake resilience; Performance objectives; Post-earthquake service restoration

Introduction

Water, wastewater, and inundation protection systems are lifelines critically related to a communities' resilience to earthquakes. Herein, water, wastewater, and inundation protection systems respectively refer to potable water, sanitary sewer, and storm water control and flood protection systems (includes urban drainage, river dam and levee systems, sea wall barriers, etc.). Water, wastewater, and inundation protection systems in modern urban settings are critical for public health and human survival, and their ability to

recover and restore services following an earthquake is essential for community resilience.

Water and wastewater systems are generally well understood and do not require significant defining descriptions. However, inundation protection systems are not as well understood in the context described herein. Impacts from Hurricane Katrina in New Orleans, USA, identified how inundation protection systems serve as community lifelines and the need to understand their systemic performance for community protection. Events in Japan and New Zealand (both in 2011) have identified the critical need for post-earthquake performance of inundation protection systems. Some portions of inundation protection systems such as urban storm water drainage have great similarities with wastewater systems and are sometimes combined with wastewater sewer systems. Other water collection, storage, and conveyance systems are strongly related to water systems. However, the classic understanding of water, wastewater sewer, or storm water sewer systems does not include important aspects of the more general use of inundation protection systems such as water-retaining levees along major rivers and sea barriers protecting cities against atmospheric hazards (hurricane, typhoon, etc.). These are clearly critical lifeline systems and need to address the cascading effects of potential flood after earthquakes in order to properly address the inundation protection system along with the community resilience.

Water, wastewater, and inundation protection systems are large and complex geographically distributed systems traversing many geologic formations and hazards and can experience damage and resulting service losses when subjected to a widespread shock from an earthquake. A lifeline system's seismic resilience is dependent upon the amount of service losses sustained by an earthquake and the time required to return the services. The post-earthquake system performance is critical to the resilience of the communities served.

In order to assess a systems' serviceability and needed resilience following an earthquake and how those services impact community resilience, one must first understand the service types or

categories a system needs to provide after a potentially disastrous event (i.e., service customers have become dependent and reliant upon for personal and community-wide survival and sustainability).

The following section describes performance categories commonly expected of water, wastewater, and inundation protection systems. System services are shown as a subset of the performance categories. Each system performs functions to provide services to customers or portions of communities. For example, water systems have connections from the distribution network to homes or businesses to provide water to each customer. Wastewater systems have connections from homes or businesses to the collection network to remove sewage wastes for each customer. For these two systems each customer has a physical service connection, whereas inundation protection systems do not generally have customer connections as that described for water and wastewater systems. Instead inundation protection systems usually operate infrastructure to protect certain areas from being flooded. These areas may be farm lands, residential neighborhoods, business communities, entire cities, etc. The useful lands being protected are owned and utilized by people, businesses, governments, etc., who, in the context of this entry, are considered customers by the inundation protection system operators. Resilience is discussed in terms of the systems' ability to recover and return services over time and how the customers, and thus the communities, are impacted.

Water, Wastewater, and Inundation Protection System Performances and Services

Table 1 shows three primary system performance categories: (1) water, wastewater, and inundation protection services; (2) life safety; and (3) property protection. The service category is made up of at least five additional categories in Tables 2, 3, and 4, making a total of seven primary performance categories for each system. The total number of system performance categories is not

Resiliency of Water, Wastewater, and Inundation Protection Systems, Table 1 Primary water, wastewater, and inundation protection system performance categories

Performance category	Description
Water, wastewater, and inundation protection services	Provision of water, wastewater, and inundation protection services identified in Tables 2, 3, and 4, respectively
Life safety	Preventing injuries and casualties from direct or indirect damages to system facilities; includes safety matters related to response and restoration activities
Property protection	Preventing property damage as a result of damage to system components; also includes preventing system damages

Resiliency of Water, Wastewater, and Inundation Protection Systems, Table 2 Water service categories

Service categories	Description
Water delivery	The system is able to distribute water to customer service connections, but water delivered may not meet quality standards (requires water purification notice), pre-event volumes (requires water rationing), fire flow requirements (impacting firefighting capabilities), or pre-earthquake functionality (inhibiting system operations)
Quality	The water quality at service connections meets pre-earthquake standards. Potable water meets health standards (water purification notices removed), including minimum pressure requirements to ensure contaminants do not leach into the system
Quantity	Water flow to customer service connections meets pre-earthquake volumes (water rationing removed)
Fire protection	The system is able to provide pressure and flow of a suitable magnitude and duration to fight fires
Functionality	The system functions are performed at pre-earthquake reliability, including pressure (operational constraints resulting from the event have been removed/resolved)

Resiliency of Water, Wastewater, and Inundation Protection Systems, Table 3 Wastewater service categories

Service categories	Description
Wastewater collection/removal	The system is able to collect and remove wastewater at the customer service connections while preventing sewage flooding, but the system may not be able to treat collected wastewater to meet quality standards, properly dispose of wastewater, or meet pre-earthquake functionality (inhibiting system operations)
Quality	Wastewater is able to be treated to pre-earthquake volumes using intended processes and meet public health standards
Disposal	Entire wastewater volume is able to be properly disposed, protecting the environment, and meeting public health standards (including containment within pipe network)
Reclaimed source	Wastewater is able to be treated and used as an alternate source of water supply (note: this does not apply to all wastewater systems)
Functionality	The system functions are performed at pre-earthquake reliability (operational constraints resulting from the event have been removed/resolved)

limited to seven; however, any additional categories utilized to understand water, wastewater, or inundation protection system resilience are likely subordinate and needed to achieve the primary services presented in Tables 2, 3, and 4. Tables 2, 3, and 4 summarize service categories normally

provided by water, wastewater, and inundation protection systems using common network components and topology.

A system or portion of system meeting the description in Tables 2, 3, and 4 for each category is considered restored to the customer.

Resiliency of Water, Wastewater, and Inundation Protection Systems, Table 4 Inundation protection system service categories

Service categories	Description
Flood defense	The system is able to defend regions against flood hazard by collecting, storing, removing, and/or providing hydraulic containment and barriers to protect life and property from storm water and other forms of water runoff (e.g., snowmelt) as well as waves, surges, and other potential forms of inundation from water bodies; but some region(s) may be more vulnerable to flooding than prior to the earthquake and the system may not be able to treat water to meet quality standards, properly utilize or dispose of water, or meet pre-earthquake functionality (inhibiting system operations)
Quality	Water is able to be treated to pre-earthquake conditions using intended processes and meet public health standards (note: this does not apply to all inundation protection systems)
Disposal	Entire wastewater volume is able to be properly disposed, protecting the environment, and meeting public health standards (including containment within pipes, conduits, and conveyance lines)
Water supply	Storm/flood water is able to be treated and used as a water supply source (note: this does not apply to all inundation protection systems)
Functionality	The system functions are performed at pre-earthquake reliability, including the meeting or exceeding of pre-earthquake inundation protection levels (vulnerabilities and operational constraints resulting from the event have been removed/resolved)

The descriptions in Tables 2, 3, and 4, respectively, are intended to capture most aspects associated with modern water, wastewater, and inundation protection system services around the world; however, description modifications may be needed to suit certain regional practices. Some systems may not provide all the services listed in the above tables, but all systems have at least a primary category (e.g., water delivery, wastewater collection, flood defense) and functionality services. Local systems may need to customize Tables 2, 3, and 4 by modifying, removing, or adding services and related descriptions to ensure proper system resilience.

The ability of systems to provide the service categories in Tables 2, 3, and 4 can be assessed at any given time as a ratio of the number of customers with the service after an earthquake to the number of customers having the service before the earthquake. An assessment can be performed in relation to a hazard evaluation or an actual event for each service category in Tables 2, 3, or 4. As exemplified in Fig. 1a–c, the number of customer services restored over time can be monitored to track and evaluate overall system serviceability for each category.

The service restoration curves in Fig. 1a–c do not represent the performance of any specific system following any specific earthquake

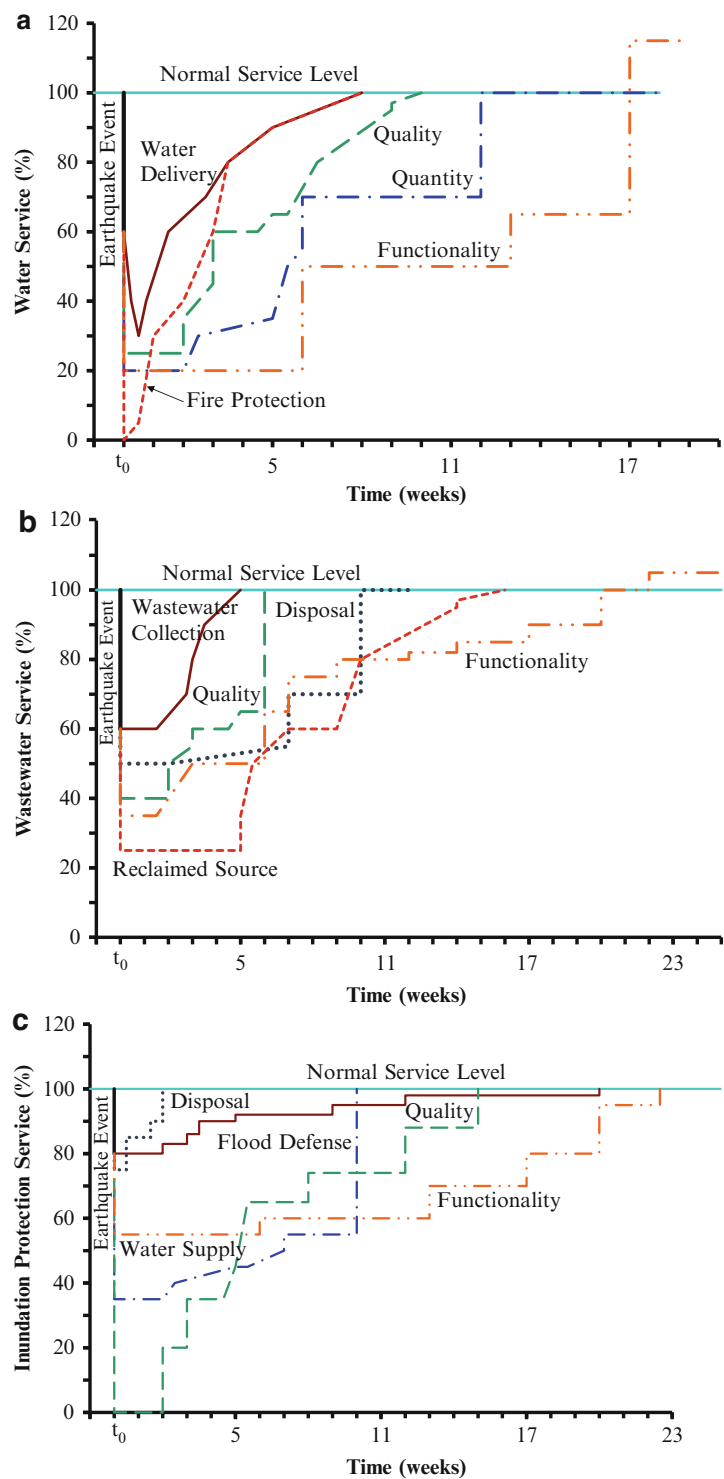
scenario; the curves are only intended to aid in understanding service restorations and system resilience. The units of weeks shown on the time axis are not intended to be typical. The restoration curves for each system hold specific characteristics that can be used as management tools for understanding how to develop and manage resilient systems. Davis (2011, 2014b) describes water system service characteristics and their interactions.

The flood defense service category in Table 4 has a related but important difference to the life safety and property protection performance categories in Table 1. The inundation protection systems serve to provide property protection against flooding and most coincidentally provide life safety protection against flooding; thus, the property protection and life safety performance categories are also service categories for the inundation protection systems. However, life safety and property protection issues go beyond the primary inundation protection purpose and include other aspects of system performance and operations, which are covered in the Table 1 performance categories (e.g., building collapse, unsafe work practices, and other examples provided in this entry), but not in the Table 4 flood defense service category.

The provision of services and the protection of life and property are arguably the most important

Resiliency of Water, Wastewater, and Inundation Protection Systems, Fig. 1

(a) Example water service restoration curves. (b) Example wastewater service restoration curves. (c) Example inundation protection service restoration curves



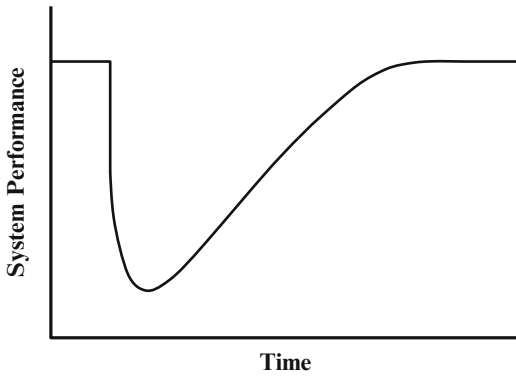
performances the water, wastewater, and inundation protection systems can achieve. Life safety can be threatened when system structures fail, dangerous chemicals (e.g., chlorine) are released into the atmosphere, large volumes of water are catastrophically released into populated areas (e.g., dam or levee failure), etc. Property damage can result from similar system damages as well as water from pipe breaks. Proper hazard mitigations to critical components within a system (e.g., strengthening dams, pipes, and other facilities, anchoring chemical containers, etc.) are important measures to prevent casualties and property damages. Although these mitigations are critical for life and property safety and aid in improving system resilience, they generally focus on specific components within the system and as a result do not always directly improve the ability to provide post-earthquake services. Services are those systemic aspects customers must rely upon, including safe and reliable: (A) water supply as described in Table 2; (B) wastewater collection, disposal, and associated items as described in Table 3; and (C) flood protection and associated items as described in Table 4.

The system serviceability categories can be used as direct measures of system resilience and related to community resilience; as a result, system resilience cannot be measured using a single performance category since no system can be defined with only one service category (there are at least two service categories for every system as described above). In contrast, the life safety and property protection categories are difficult to relate to system resilience as they cannot be used as a metric to track system performance over time, but often can be more directly related to community resilience. All performance categories are important to ensure community resilience, but the service categories are often the most important metrics for understanding water, wastewater, and inundation protection systemic resilience. This can be clarified through examples. If a large dam located within a city fails resulting in extreme devastation to a community, the community may be found to have low resilience to such a hazard (i.e., take an exceptionally long time to recover to pre-dam

failure conditions), but the water system may have high resilience on a systemic level and able to continue providing potable water services (all services shown in Table 2) to all customers (excepting any lost services from the flood) at all times because it has redundant supplies. This example shows the importance of the life safety and property protection performance categories and emphasizes the flood defense service category of inundation protection systems. Now, presume the water system has mitigated the seismic risks to the above described dam so that it does not fail during a great earthquake striking the city, but thousands of other water system components, which do not pose a life safety or property protection concern, are severely damaged removing the ability to supply water to the majority of customers for a very long time, leading to public health concerns. In this second example water supply remains behind the dam, but the system fails and is unable to deliver water. The water system is less resilient to the earthquake than the dam failure hazard, even though the dam failure may be more catastrophic to the community. Certainly the water system and the community are both in worse condition and less resilient if the dam were to fail during the earthquake. From these examples the relative community resilience between the dam failure and earthquake hazards cannot be judged. The main point of these examples is not to weigh the services versus life safety or property protection performance categories, but instead to show how all performance categories are important, and addressing only a few performance categories does not necessarily improve system or community resilience to all important hazards potentially impacting a community.

System Services and Seismic Resilience

McDaniels et al. (2008) illustrates system resilience, shown in Fig. 2, as a measure of system performance over time, describing system performance by the number of customers served. Bruneau et al. (2003) illustrates system resilience similarly describing the vertical axis as the



Resiliency of Water, Wastewater, and Inundation Protection Systems, Fig. 2 Infrastructure system resilience

quality of infrastructure of a community. From this it is easy to see how the water, wastewater, and inundation protection system service restoration curves in Fig. 1 can be linked to resilience. Figure 1 also illustrates how system resilience is more complex than shown by a single curve in Fig. 2. Understanding this complexity is important to: (1) developing, maintaining, and operating a resilient infrastructure system and (2) creating a resilient community.

As previously noted, water, wastewater, and inundation protection system resilience cannot be characterized by any single service category. These system resiliencies are dictated by multiple service categories and how they interact with the regional community. Community dependence on water, sewage disposal, and flood protection for survival does not allow these system's resiliencies to be defined independent of the community resilience. This is compounded by the dependency relationships with other lifeline infrastructure (e.g., power to run pump stations). Thus, system resilience cannot be measured only by the service-time lost, but also by how it helps to improve overall community resilience; this is clarified using some examples below.

Water System Examples

If an urban water system were able to provide nearly full water delivery and quality service throughout the distribution area but limited fire protection service following a major earthquake,

and several ignitions grow into large fires, then this could result in serious community fire damage, especially if the fires merged into a major conflagration. This situation would not be a characteristic of a resilient community, but if only one of the service restorations was reviewed (e.g., water delivery, as is commonly done following an earthquake), the water company could potentially claim that it was highly resilient to the earthquake. As another example, water systems capable of providing post-earthquake services to other lifelines and emergency operations such as hospitals, emergency operation centers, evacuation centers, etc., in a manner that does not significantly disrupt their critical operations help increase the community resilience. This is a characteristic of a resilient water system even if the services to residents and other businesses are temporarily disrupted.

Wastewater System Examples

A wastewater system transmission pipe network located in a large urban area is severely damaged in an earthquake and cannot be used to convey sewage from the customer service connections to the treatment plant. However, most customers can dispose sewage from their buildings into the wastewater collection network directly through their service connection lines, as long as the transmission system can remove the waste and keep it from backing up and flooding the streets and homes. To continue providing collection services, the wastewater system operators begin to dispose of waste directly into local rivers and streams passing through the city. This allows the wastewater collection service to continue but results in severe environmental contamination and potential health hazards. This example shows how one service may be provided at nearly 100 % (sewage collection) and if this were the only metric tracked (commonly done after an earthquake), the wastewater system could claim it is resilient from the perspective of providing collection services. However, the transmission subsystem is dictating overall system performance and shows poor resilience relative to the severe and potentially unacceptable impacts on the community and overall environment.

Inundation Protection System Examples

A local community is protected from river flooding by a levee system located along the river banks. The local rainwater runoff within the community is pumped from the drainage pipe network to a local treatment plant then discharged into the river. An earthquake causes severe damage to the levee system, the pump station, and treatment plant. The drainage pipe network has also been significantly damaged but maintains ability to collect and flow water. In many locations the levees have settled below the annual mean river elevation. The earthquake struck at the end of the dry season so fortunately the water level is lower than the mean and no catastrophic water release occurred. However, a relatively significant storm arrived days after the earthquake, before all levees and pump stations could be repaired, and the levee system was overtopped in one location resulting in severe flooding of homes and businesses for many weeks. This inundation protection system may appear to have performed resiliently immediately after the earthquake and even retained water in all but one location throughout its entire length, however since the system functionality was not adequately restored in time to protect against flooding severe impacts to the system and the community result, revealing the low level seismic resilience within the system due to the large number of customers impacted. However, had the system operators known of the potential vulnerabilities and had emergency materials, supplies, contracts, equipment, and other necessary resources on hand to respond in a manner to protect against the overtopping, then the inundation protection system would be shown as resilient to the earthquake strike.

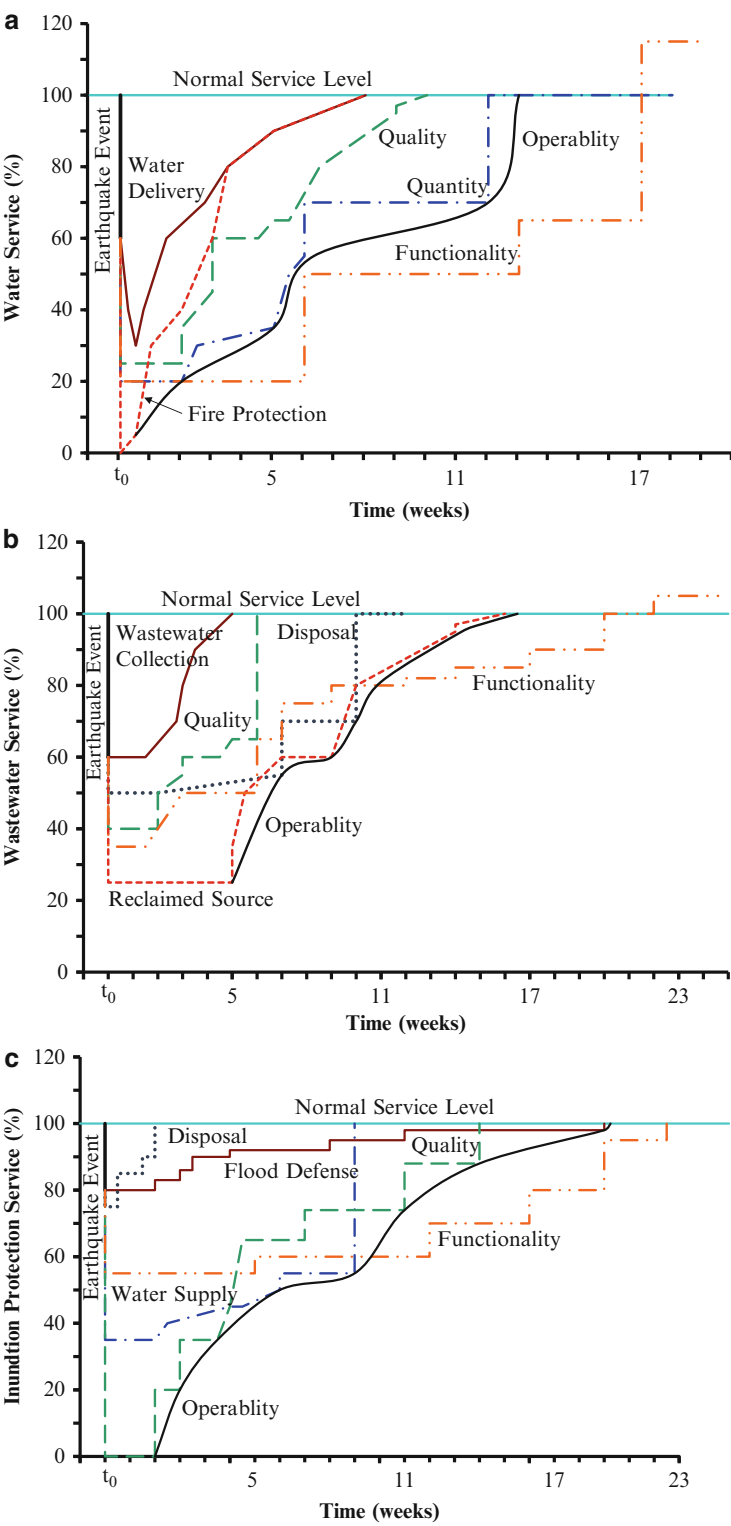
Functionality Versus Operability

Functionality services describe the ability of a system to reliably perform. A highly functional system can achieve its basic purpose (i.e., for water systems provide water delivery, quality, quantity, and fire protection services) prior to completing all infrastructure repairs (e.g., Davis

et al. 2012). Operability describes the cumulative restoration of all services except functionality.

Except for functionality services, the distinction between the other services for each system in Tables 2, 3, and 4 is relatively easy to understand (i.e., for water systems the distinction between water delivery, quantity, quality, and fire protection services is relatively easy to understand). The difference between functionality services and operability for each system is not so inherent and is described in more detail here. Damage imposes constraints that do not allow the system to function with its pre-event performance and reliability in advance of completing all necessary repairs, even when the other services are completely recovered. For example, in a water system after the water delivery, quality, quantity, and fire protection services reach 100 % restoration the system may be fully operational and able to completely service customers (herein, at this state the system is termed operational and has operability) prior to the system being fully functional. The state of complete operability is mostly from the customers' perspective of having all services restored, but the system may not operate or function as it did prior to the extreme event. Full functionality services are not recovered until the constraints imposed by the event (e.g., component damages) have been removed or resolved by completing repairs, possible new construction, and/or operational modifications. A complete functionality recovery returns all services to their pre-event performance, reliability, and redundancy level, at which time the system can operate or function as it did prior to the event. Thus, the functionality services are the last to be fully restored and may be completed after or along with some or all of the other services. In the inundation protection system example provided in the previous section, if the levees were adequately repaired in time to resist all upcoming river flows the system could be considered operable, but not fully functional until all repairs were completed and returned to at least pre-earthquake conditions. Figure 3 shows bounding curves identifying operability using the same serviceability

Resiliency of Water, Wastewater, and Inundation Protection Systems, Fig. 3 Example service restorations showing bounding operability curves (a) water services; (b) wastewater services; (c) inundation protection services



examples of Fig. 1. Operability may be controlled by one or more service categories.

This view of functionality is consistent with that normally considered for resilience modeling (e.g., Bruneau et al. 2003; Cimellaro et al. 2010), with the exception presented herein distinguishing the important difference between the system operability and functionality. For example, in the context of this entry, Cimellaro et al. (2010) define recovery time as the period necessary to restore system functionality to a desired level that can operate or function the same, close to, or better than the original one (which is not achieved until all significant system components are restored to pre-event conditions, i.e., when all significant operations restraints are removed). Bruneau et al. (2003) identifies restoration as the time when the infrastructure is completely repaired. Understanding the difference between system functionality and operability is important for distinguishing the difference in system resilience and the systems' support to community resilience. Operability recovery provides a direct measure of a systems' support to community resilience, and functionality recovery is a measure of water, wastewater, or inundation protection system resilience. Operability recovery is critical for community resilience, and functionality recovery is critical for ensuring the community recovery is complete and sustainable. The difference in operability and functionality restoration and recovery times lies mostly in system redundancy and resourcefulness, which are dimensions of resilience (Bruneau et al. 2003).

As previously described, operability is achieved through a combination of service restorations, but cannot by itself be described as a service. The time it takes to obtain operability status throughout a system can only be viewed as the temporal sum to incrementally restore all the services other than functionality (e.g., wastewater collection, quality, disposal, and reclaimed source for wastewater systems) and therefore serves as a descriptive milestone delineating when customers resume receipt of their accustomed services. Thus, it is important to understand how to assess each service category and its respective characteristics.

Service Equilibrium Shift

Figures 1 and 3 assume the normal pre-earthquake service level (i.e., the horizontal line at 100 %) remains stable during and following the earthquake event. The previous descriptions only indicated the functional services may have a final restoration that may not match the pre-event level for water and wastewater systems (see Fig. 1a, b). However, it is possible for all service categories to have a short-term or long-term stability that does not match the pre-event service levels. This involves a service equilibrium shift (SES) in the net loss or gain in serviceability (a gain is shown for functionality in Fig. 1a, b). In many cases, any potential SES is a function of the system's interaction with the greater community's resilience. A net loss in services may result either from the inability of a damaged system to support the pre-event number of services or the inability of the community as a whole to sustain the number of people and industries, regardless of the systems' ability to support the services. The March 11, 2011, Great East earthquake and tsunami disaster in the Tohoku region of Japan provides numerous examples of this problem. The destroyed communities were not resilient to the tsunami hazard resulting in a permanent decrease in the total number of water, wastewater, and inundation protection services. At some point after an event causing a permanent SES, the curves will need to be re-normalized by the post-event number of service connections to account for the permanent loss in demand.

Figure 4 shows an example water system quantity restoration curve having a temporary 40 % reduction and a permanent 20 % reduction. This may result from an earthquake severely impacting water supply such as ground water wells, aqueducts, or other sources. In this example a 40 % or greater conservation is needed between times t_0 and t_1 , time of earthquake strike and time when temporary SES is removed, respectively. At time t_1 an additional 20 % of water supply is restored, but the community must now adjust to a permanent 20 % loss in water supplies. Restoration curves similar to

Resiliency of Water, Wastewater, and Inundation Protection Systems, Fig. 4 Example quantity service equilibrium shift (SES)

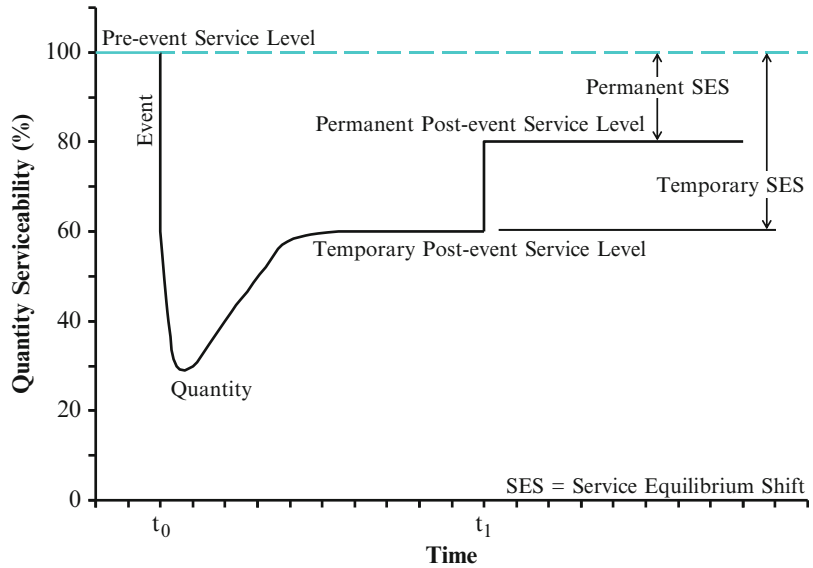


Fig. 4 may be created for any of the service categories presented in Tables 2, 3, and 4 and may even show an increase in services as indicated for functionality in Fig. 1a, b. The SES characteristic shown in Fig. 4 as a permanent loss is not inherent to all systems, but may result from certain extreme events.

The examples in Fig. 1 show all categories having a common trend of initial service declination then an increase in service restorations until completed. In real post-earthquake situations, the ability to maintain a continuous upward trend, once initiated, is highly dependent upon how the restoration activities are managed. Managers in charge of restoration should make decisions to target a continued upward trend in restoration after controls have been established to stop the downward trend. It is possible to have a drop in services within the upward trend. An actual case where this has occurred is presented in the Case Study section later in this entry. Repeated serviceability drops (even just one) can have significant ripple effects and potentially serious impacts to the community. The inundation protection system example in the [System Services and Seismic Resilience](#) section has a severe reduction in flood defense services a few days after the earthquake.

Quantifying Post-earthquake System Services

The service restoration calculation methodology is as follows: (1) identify the area (s) where services are not being met, (2) count the number of services (or people, businesses, etc.) in each area, and (3) calculate the ratio for number of post-earthquake services to pre-earthquake services for each category in Tables 2, 3, and 4. The water, wastewater, and inundation protection service restoration curves are plots of this quantification over time. For post-event restoration estimates, the calculation for all services except functionality is relatively independent of system layout and operations, whereas an assessment for an earthquake scenario may require an understanding of system layout and some hydraulic analysis. Functionality services, however, cannot be estimated in any case without a full understanding of systemic capabilities. For this reason the remainder of this section describes quantification of post-earthquake functionality.

Functionality characterizes the ability of a system to operate in a damaged state and provide services in a manner customers are used to receiving. Damage imposes operational

constraints, and the system cannot function at its pre-earthquake performance and reliability in advance of removing these constraints even when all the other system services (i.e., Tables 2, 3, and 4) are completely recovered. Tables 2, 3, and 4 identify performance and reliability as critical service aspects captured by functionality.

Functionality services S_f can be estimated from

$$S_f = \frac{m}{N} = \frac{1}{N} \sum_i^{\eta} m_i \frac{F_P F_R}{1 + R} \Big|_{\text{equiv}} = \frac{1}{N} \sum_i^{\eta} m_i E_i \quad (1)$$

where m is the weighted average number of services providing functionality, m_i is the number of services in service area A_i , η is the total number of service areas to be evaluated, F_P is the performance factor applied to each line or node to quantify its operational or hydraulic performance capability, F_R is the reliability factor applied to each line or node to quantify its seismic or structural reliability, and R is the redundancy factor. The number of services m_i may be evaluated in alternative forms using total service connections, population, land area, etc. $E_i = F_P F_R / (1 + R) \Big|_{\text{equiv}}$ is the equivalent performance-reliability/redundancy (PRR) calculated for all complete flow paths to or from A_i .

Equation 1 evaluates the water, wastewater, and inundation protection systems as networks of lines and nodes. Lines are conduits carrying water (e.g., channels, pipes, tunnels, etc.). Nodes are points of line initiation, intersection, or ending and may represent different types of water facilities (e.g., reservoir, pump station, treatment plant, etc.) or the connection of one or more lines. The term component is used herein to describe lines and nodes in a general sense.

$F_P \geq 0$ and is defined as the ability of a component to operate or provide flow during and following an earthquake in relation to the components' pre-earthquake performance. For components operating after an earthquake but not with the pre-earthquake performance, then

$0 < F_P < 1$. $F_P = 0$ for each component having a complete loss in operational ability because damage, either to that component or another upstream component, prevents the ability to operate. $F_P = 1$ for operable components having the ability perform at pre-earthquake levels, even if they are not utilized (i.e., water is available and component can perform at full capacity, but component may remain isolated from utilization). From this F_P for each component within each subsystem can be quantified as the ratio of pre- to post-earthquake performance capability. Thus, F_P is quantified by a variety of metrics including flow rate for conduits; volumes for supply and storage; flow rate and pressure for pumping; turbidity, taste, bacteria, etc., for water quality; and so on. Additionally, flow and volume change relative to customer demand/protection and therefore F_P can vary depending on time of day and year. As a result, S_f may change depending on the metric used to evaluate F_P (e.g., for flow, average daily, peak daily, average annual, or other demand values may be used). Therefore, it is important to clearly define and report the metrics used in the evaluations.

$F_R \geq 0$ and is defined as the probability that a component will function during and following an earthquake in relation to the components' pre-earthquake reliability. For components functioning after an earthquake but not with the pre-earthquake capability, then $0 < F_R < 1$ (components are damaged and weakened but operable to some degree). $F_R = 1$ for operable components having the ability to function at pre-earthquake reliability levels (essentially undamaged). $F_R = 0$ for each component having a complete loss in reliability (i.e., damage has removed all ability to operate); in such cases $F_P = F_R = 0$. F_P is dependent upon F_R when damage removes total operability; however, F_P and F_R are not necessarily correlated when $F_R \neq 0$. For example, a pipe can be damaged with $0 < F_R < 1$ while $F_P = 1$; this means the pipe is still fully operable but with less reliability than its pre-earthquake state. An undamaged storage tank with $F_R = 1$ may not be operable ($F_P = 0$)

because it cannot receive water to store due to other system damages.

Each redundancy is assigned a coefficient R_j :

$$R_j = \frac{\sum Q_j}{Q_j} - 1; \quad Q_j \leq \bar{Q}_i; \quad \sum R_j \geq 0 \quad (2)$$

where Q_j is the flow capacity through route or path j and $\bar{Q}_i$ is the total pre-earthquake flow demand to or from A_i or a point p_i .

The equivalent PRR for route k made up of j components in series is

$$E_k = \prod \frac{F_{P_j} F_{R_j}}{1 + R_k} \quad (3)$$

The equivalent PRR for l routes in parallel is calculated from

$$E_k = \sum_r \frac{\prod_j F_{P_{jl}} F_{R_{jl}}}{1 + R_l} \quad (4)$$

where Σ and Π are the operators for summation and product, respectively. $F_{P_{jl}}$ and $F_{R_{jl}}$ are the factors for the j th component on route l . Equations 3 and 4 are used to assemble the equivalent PRR for all paths to a service area. Davis (2014a) provides an example on how to apply the above relationships.

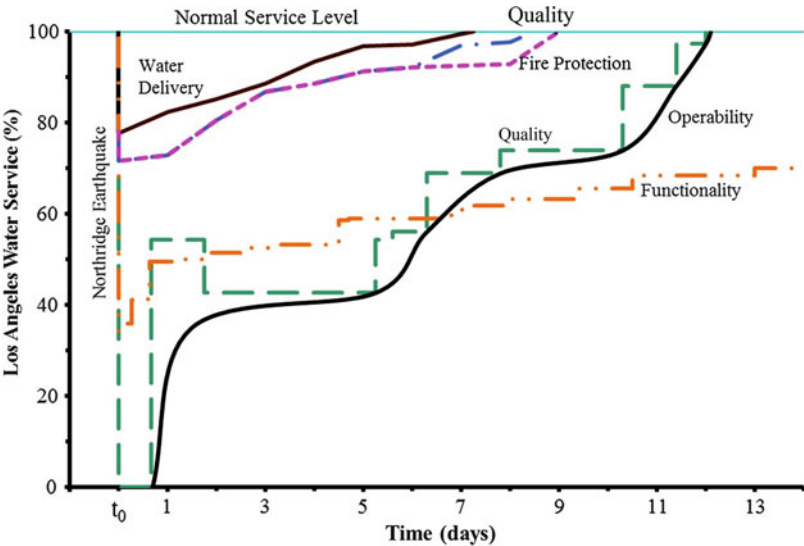
Case Study of Los Angeles Water System Restoration

To further illustrate service restorations, results of an actual water system earthquake performance and post-event service restorations are presented for Los Angeles following the 1994 Northridge earthquake. This case identifies the importance for documenting post-earthquake restorations for all service categories. Case studies quantifying system performance capabilities improve our currently limited knowledge on the complex interactions between (a) the service categories and (b) service restoration and community resilience.

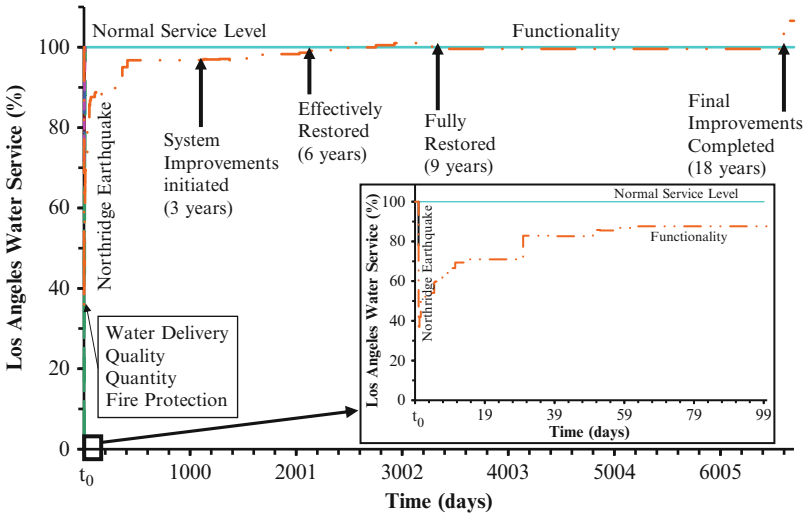
On January 17, 1994, a moment magnitude 6.7 earthquake struck the northern area of Los Angeles and caused significant damages to the Los Angeles Department of Water and Power (LADWP) infrastructure. In summary, there were 14 repairs made to the raw water supply conduits, more than 60 repairs to treated water transmission pipes, 1,013 repairs to distribution pipe, over 200 service connection repairs, 7 reservoirs were damaged, half the treatment plant was temporarily removed from service, and some other incidental damages. Total water system repair costs reached \$41 million. Davis et al. (2012) describes the damage, water outage areas, post-earthquake system performance, and service restorations, which are summarized here.

Figure 5 shows the water service restoration curves for the LADWP following the 1994 Northridge earthquake. As seen in Fig. 5 the water delivery service dropped to about 78 % shortly after the earthquake due to water leaking from broken pipes. The LADWP's ability to contain the impacted area and initiate restorations rapidly allowed the water delivery services to increase soon after the earthquake. The quantity and fire protection services dropped to a low of about 72 % on January 17, 1994. The quality service dropped immediately to zero because a water purification notice was issued across the entire city within 3 h after the earthquake. As shown in Fig. 5, the water delivery service was restored to 100 % at about 7 days, quantity and fire services at about 8.5–9 days, and quality service at 12 days after the earthquake. The rapid increase in quality service within 1 day after the earthquake resulted from recognition that much of the system was not damaged and water quality was maintained in the undamaged areas. The drop in quality service for about 3 days, around 2 days after the earthquake, resulted from the realization that a portion of the pipe network where the water purification notice had previously been removed suffered greater damage than initially recognized; upon recognition the notice was reinstated. The remainder of quality restoration was primarily due to

Resiliency of Water, Wastewater, and Inundation Protection Systems, Fig. 5 Los Angeles water system service restorations following the 1994 Northridge earthquake



Resiliency of Water, Wastewater, and Inundation Protection Systems, Fig. 6 Los Angeles water system functionality service restorations following the 1994 Northridge earthquake



disinfecting the broken pipe network after sufficient repairs were completed to restore flows and pressures. Figure 5 shows the operability state was completely governed by water quality restorations.

The Los Angeles water system has a high level of supply and transmission redundancy which was utilized to provide continued services through the non-damaged portions of the system. Figure 6 shows the entire functionality service restoration, which is calculated using the

methodology described in the previous section for the supply, treatment, transmission, and distribution subsystems. The entire functionality restoration can be viewed at different scales for the first 15 days in Fig. 5, the entire duration in Fig. 6, and the first 100 days are blown up in the inset of Fig. 6. The functionality services initially dropped to about 34 % and began to improve soon thereafter as restorations were undertaken to the supply and transmission subsystems. As seen in Fig. 5, functionality service

does not initially track with operability or any of the other service categories; this is because damage to the supply reduced reliability to a much greater area than the main damage zone. The functionality service rapidly increased to about 60 % once critical repairs to major supply and transmission lines were completed a few days after the earthquake, followed by a relatively linear increase to 70 % for about the next 2 weeks. At 30 days after the earthquake functionality services reached 82 %, after which, as shown in Fig. 6, there were long periods between relatively small incremental service improvements until the functionality returned to normal at about 9 years after the earthquake. Several improvements were made to the system as a direct result of knowledge gained and repairs made following the 1994 earthquake; these were initiated at about 6 years and increased reliability above the pre-earthquake levels after about 9 years. All improvements were completed after 18 years resulting in a positive SES, thereby increasing the post-earthquake functionality to 105 %. The water delivery, quantity, fire protection, quality, and functionality curves are calculated from recorded data and show how the service categories in Table 2 are applied to characterize post-earthquake system performance in an actual operating system severely impacted by an earthquake.

Summary

Descriptions of normally provided water, wastewater, and inundation protection service performance categories have been presented in relation to their importance to post-earthquake system restorations. A clear distinction is made herein between functionality service restoration and system operability (restoration of the other service categories for each system) and their respective relations to (a) water, wastewater, and inundation protection system resilience and (b) community resilience. Methodologies useful for estimating and tracing the service restorations over time have been presented. A case study on service restorations for the Los Angeles

Water System applied these methodologies and showed how the service categories are applicable to actual systems that experience earthquake damage. The relationship between water, wastewater, and inundation protection system service restorations and seismic resilience was illustrated; system resilience cannot be fully described independent of community resilience.

Cross-References

- ▶ [Resilience to Earthquake Disasters](#)
- ▶ [Resourcing Issues Following Earthquake Disaster](#)
- ▶ [Seismic Design of Dams](#)
- ▶ [Seismic Design of Pipelines](#)

References

- Bruneau M, Chang SE, Eguchi RT, Lee GC, O'Rourke TD, Reinhorn AM, Shinozuka M, Tierney K, Wallace WA, Winterfeldt D (2003) A framework to quantitatively assess and enhance the seismic resilience of communities. *EERI Spectra* 19(4):733–752
- Cimellaro GP, Reinhorn AM, Bruneau M (2010) Framework for analytical quantification of disaster resilience. *Eng Struct* 32:3639–3649, Elsevier
- Davis CA (2011) Water system services and relation to seismic performance. In: *Proceedings of 7th Japan-US-Taiwan workshop on water system seismic practices*, JWWA/WRF, Niigata
- Davis CA (2014a) Quantifying post-earthquake potable water system services. In: *6th China-Japan-US trilateral symposium lifeline earthquake engineering*, Chengdu, China, May 28 June 1, 2013, TCLEE monograph no 38, ASCE, Reston, pp 19–26
- Davis CA (2014b) Water service categories, post-earthquake interaction, and restoration strategies. *EERI, Earthquake Spectra* 30(4), doi:10.1193/022912EQS058M
- Davis CA, O'Rourke TD, Adams ML, Rho MA (2012) Case study: Los Angeles water services restoration following the 1994 Northridge Earthquake. In: *15th world conference on earthquake engineering*, Lisbon, 24–28 Sept 2012, paper no 0364
- McDaniels T, Chang S, Cole D, Mikawoz J, Longstaff H (2008) Fostering resilience to extreme events within infrastructure systems: characterizing decision contexts for mitigation and adaption. *Glob Environ Change* 18:310–318, Elsevier

Resourcing Issues Following Earthquake Disaster

Alice Yan Chang-Richards¹,
Suzanne Wilkinson², Erica Seville³ and
David Brunsdon⁴

¹The University of Auckland, Auckland,
New Zealand

²Department of Civil and Environmental
Engineering, The University of Auckland,
Auckland, New Zealand

³Resilient Organisations, Sheffield, New Zealand

⁴Kestrel Group Ltd., Wellington, New Zealand

Synonyms

Demand surge; Inflation; Migrant labor; Post-disaster reconstruction; Resourcing

Introduction

Resourcing for post-disaster reconstruction is a complex and dynamic process that goes beyond the provision of building materials, labor, and plant. The interplay between economic recovery following earthquake disasters and time compression often poses resourcing issues, including a lack of desired skills and qualifications, the competition for limited resources, demand surge, and the management of migrant labor. Resourcing post-disaster reconstruction projects tends to be considered a procurement responsibility for construction professionals rather than an integral issue in disaster recovery. However, the outcomes largely depend on the way and extent to which the stakeholders leverage their influence and value into resourcing. Reconstruction following a major disaster poses a particularly challenging problem of planning recovery programs in a way that it can respond to both short-term and long-term resource availability.

The United Nations International Strategy for Disaster Reduction UNISDR defines a disaster as, “A serious disruption of the functioning of a community or a society causing widespread

human, material, economic, or environmental losses which exceed the ability of the affected community or society to cope using its own resources.” Disasters often lead to a reduction of production capacity but also to an increase in the demand for the reconstruction sector and goods. Thus, in theory, the reconstruction acts as a stimulus on the economy.

It is usually the inadequacy of resources and capacities to rebuild that makes the implementation of reconstruction programs particularly difficult. Following a major disaster, most local production facilities and supply systems in manufacturing industries are likely to be damaged, and the construction market tends to be in disorder and contested. Disruption of transportation and energy supply systems can exacerbate supply problems. In addition, preexisting problems within the construction industry can exacerbate the difficulty in procuring building expertise and materials. Resourcing issues thus arise more frequently following a large disaster due to the restructuring of economic activities as a consequence either of the disaster itself or of the reconstruction.

Characteristics of a Post-disaster Resource Market

In the immediate aftermath of a disaster, the assessment of resource requirements tends to be undertaken in parallel with damage assessment. In order to combat the shortage of resources for the rebuild, it becomes imperative for public agencies to develop quantitative models for projecting demand of various resources over the recovery timeframe. Based on an extensive mapping exercise, decisions need to be made quickly on what resources are needed, where the shortages are, what the sources are, and how to ensure their availability in a timely manner. Sometimes national- or local-level policies and plans may exist to guide the development of materials and workforce capacity to meet the reconstruction task, but all these are subject to change as more precise damage information becomes available.

Post-disaster reconstruction is affected by the interplay between time compression and natural

disaster economics. Time compression requires an unusual pace of capital expenditures and reconstruction. Under the pressure of limited time, the need to replace lost homes, infrastructure, and commercial properties generates a significant demand surge for materials and construction workers. In the short term, issues such as material and skill shortages for reconstruction invariably alter the landscape of resource demand and supply. It is therefore common to find environmentally related resource consents are fast tracked for the quarry industry to increase their production capacity. It is also common to find the training and education programs are consolidated for those who are eager to enter the building industry, catching the reconstruction boom.

Unfortunately, the resource shortages and their effects on reconstruction tend to be considered to be a short-term issue. The fact that resource demand is very evenly distributed across different stages and different sectors of reconstruction tends to be overlooked by policy makers. In some cases, the repercussions of a lack of resources for reconstruction can be felt over a long period. For example, 3 years after the 2004 Indian Ocean tsunami, post-disaster housing reconstruction in Indonesia faced:

- A large unused inventory of imported construction products such as windows, doors, and roofs as a result of changes to design
- Suspended progress on house rebuilding due to overstretched budgets (Fig. 1, photo: Chang-Richards)
- Many newly built houses that had to be demolished as a result of the poor quality of construction materials and workmanship
- A multitude of tsunami-affected beneficiaries who were still living in temporary shelters awaiting permanent homes promised by donors

There are several factors which, through their interaction and combined effects, inflate resource demand in post-disaster reconstruction. These include revised building codes and standards, regulatory requirements, construction innovation, environmental concerns, altered housing culture, and budgetary constraints. Research in



Resourcing Issues Following Earthquake Disaster, Fig. 1 Suspended progress on construction of a post-tsunami house in Aceh, Indonesia

this area demonstrates that a dynamic resourcing perspective and approach is a key requirement for dealing with these uncertainties and complexities inherent in the reconstruction process.

Resourcing Approaches for Post-disaster Reconstruction

Resourcing broadly encompasses a wide range of activities that have a bearing on resource management for post-disaster reconstruction projects, embracing pre-event resource planning and preparedness, resource procurement, resource delivery, and the development of resource alternatives. Conventional measures have been employed in past reconstruction practice to address resourcing problems, such as new investment in production and importing resources from outside affected areas. These measures tend to be ad hoc and seem to be unable to perform well to alleviate resource shortages in the long run. In addition, the absence of pre-event planning and preparedness, the inadequacy of efficient and flexible institutional arrangements, and the lack of proactive engagement of the construction industry in disaster management all contribute to undermining resourcing performance in a post-disaster environment.

The empirical research conducted by Chang et al. (2010) recognized that, in order to arrive at

a resilient and sustainable built environment after a disaster, resourcing efforts should be made around four components:

- A resourcing facilitator: legislation and policy
- A resourcing implementer: the construction industry
- A resourcing platform: the construction market
- Resourcing access: the transportation system

The dynamic resourcing process for post-disaster reconstruction reveals that resource availability depends on how relevant stakeholders address resource constraints including resource cost, quality, quantity, environmental concerns, and cultural acceptance.

The type of resourcing approach can be defined in terms of the way and extent to which the stakeholders leverage their influence and value into resourcing activities. Four main resourcing approaches have been widely applied in past disaster reconstruction practice.

1. *Government-driven resourcing approach*: post-disaster reconstruction resource availability is mainly driven by governmental agencies and other authorities.
2. *Donor-driven resourcing approach*: donors play a dominant role in resourcing efforts for a post-disaster reconstruction project.
3. *Market-driven resourcing approach*: the instruments, forces, and rules in the construction market have a major influence on resource availability for post-disaster reconstruction.
4. *Owner-driven approach*: house owners are responsible for rebuilding their own houses through self-maintenance with limited external financial, technical, and material assistance.

In 1974, Cyclone Tracy destroyed the city of Darwin in North Australia. The post-Darwin cyclone reconstruction adopted a government-oriented resourcing approach. The Australia government initiated restrictions on building projects in order to control resource prices. However, post-cyclone inflation was about 75 %, and this

impact even extended to Townsville in the state of Queensland (Walker and Minor 1979).

In contrast, the 1989 Newcastle earthquake in Australia was different because a more market-driven resourcing approach was allowed and the government only controlled building standards. The post-earthquake inflation in Newcastle was about 20 % (Shephard et al. 1997). This inflation was largely due to the differing view of clients' engineers and insurers on what remedial work was required (Walker 1995).

The 1994 Northridge earthquake in the United States shows that resource availability can be further enhanced by government commitment to improve its environment and to help facilitate the procurement and development of resources. Wu and Lindell (2004) highlighted that having a pre-impact recovery plan in Los Angeles, with resource availability as one of its targets, facilitated housing reconstruction and allowed local officials to manage the reconstruction process more effectively. Other cases, such as recent events in New Zealand and Japan, show that by investing in construction apprenticeship training and sectoral programs, the government enhanced the market players' capacity to cope with large-scale disaster reconstruction demand.

Donor-driven resourcing occurs when nongovernment organizations (NGOs) or other designated organizations are tasked with housing rebuilding projects for beneficiaries, often in underdeveloped countries where the indigenous resources and capability are unlikely to cope with large-scale disasters.

In comparison with donor-driven resourcing, the owner-driven/community-driven approach is empowering and participatory and thus was popular in post-tsunami reconstruction in Indonesia among NGOs, such as the United Nations High Commissioner for Refugees (UNHCR), the United Nations Children's Fund (UNICEF), and World Vision, which consider community redevelopment and participation as being among their main objectives.

The UNDP, in conjunction with UN-HABITAT, designed the Aceh Nias Settlements Support Program (ANSSP). This program promoted self-construction with funding support

from aid agencies in forms of cash grant or transfer. Aid agencies also provided technical assistance with regard to material selection and procurement in some affected rural areas in Aceh. UN-HABITAT also introduced “People’s Process” in Aceh to allow a fully empowering community-organized reconstruction to make the best use of local resources and capabilities.

The four models work holistically and are essentially transferable under a specific reconstruction circumstance. For example, the government may control the supply of cement, which is purchased by a contractor or an NGO for construction work, with or alongside a community constructing its own houses. Hence, there is a continual shifting between these models in post-disaster resourcing for reconstruction. It is when this balance is distorted that resourcing issues appear. For example, there were cases in which local governments restricted the supply of cement in Sri Lanka and the supply of timber in Simeulue, Indonesia, after the 2004 Indian Ocean tsunami and the supply of roofing in Pakistan after the 2005 Kashmir earthquake. International NGOs imported these resources from outside the affected areas where local contractors and communities were unable to cope with resource shortages, and longer-term resourcing problems appeared.

The four types of resourcing approaches can overlap or combine in the same recovery operation. Government-driven resourcing does not necessarily imply that the government gets involved in actual resource procurement for housing reconstruction projects but that the authorities retain full control over the resource provision and supply process. The success of market-oriented and donor-led resourcing in post-disaster housing reconstruction also requires continual government participation and lies largely in the way and extent to which the authorities provide facilitation.

Resourcing Issues Following a Major Disaster

Resourcing for post-disaster reconstruction is often construed as a procurement responsibility

for construction professionals rather than an integral issue in disaster recovery. This approach underestimates the complexity of the real economic challenge that disasters pose to communities. In reality, reconstruction following a major disaster must confront important resourcing issues such as a lack of desired skills and qualifications, the competition for limited resources, demand surge, and the management of migrant labor.

The tendency in post-disaster situations is often to revise building codes and standards in order to enhance the resilience of structures to future disasters. However, this regulatory change induces the changes to the requirement of materials, skills, tools, and techniques in new construction. New materials and skills may not be available in the time of reconstruction as it requires a considerable time for manufacturers to undertake the research and development, to test and release these new materials onto the market. The additional costs of the new construction requirements are often not well reflected in the regulatory impact statement due to uncertainties and the real-time nature of disaster events.

Therefore, governments tend to set their recovery policy on the assumption that the prices of building materials will rise rapidly when the rebuild ramps up in earnest but will gradually fall again as the peak of reconstruction passes the spikes in construction demand. Agencies are inclined to respond to the anticipated resource shortages during the reconstruction boom time or simply leave it to the market. Research in this area has shown that when the resourcing issue is not handled well, it will lead to a secondary economic disaster manifested by inflationary chaos, “Dutch disease” or demand surge. “Dutch disease” is named after the experience in the Netherlands of deindustrialization in the wake of large inflows of export revenues from North Sea Oil in the late 1970s. It is known as the negative impact on the non-booming sectors due to the simultaneous and competing demand for inputs between booming sectors and non-booming sectors. These adverse market responses can worsen local economies, causing time and cost effects on disaster reconstruction projects.

More resource challenges may appear in countries where the local natural resources are already severely deprived prior to a disaster. “Business as usual” projects, capital works, and development projects may also play a part, absorbing wider market space from recovery projects. The impacts of resource shortages on recovery are more profound when time is tight, and the government is under political pressures. In post-tsunami reconstruction in Indonesia, for instance, some house owners or aid agencies turned to available inferior resources or sought to import materials from outside the region with lengthy lead times. As could be expected, without appropriate site supervision and quality control in place, a variety of construction defects and failures associated with poor building materials and workmanship occurred (Fig. 2, Photo: Chang-Richards).

Implications of Demand Surge

Demand surge is a prominent feature associated with large-scale natural disasters, one which entails higher repair or rebuild costs resulting from higher labor wages and material prices following a large- versus small-scale disaster. Increased repair and rebuild costs after large-scale disasters have fallen into a broad range between 20 % and 250 %. For example, after the 1994 Northridge earthquake (US), insurers observed a 20 % increase in the costs to settle claims (Kuzak and Larsen 2005, p. 113). The Cyclone Larry (Australia) saw an increase of 50 % in reconstruction costs (Australian Securities and Investments Commission 2007). Most needed materials brick, aggregate, and cement had costs increase by 127 %, 125 %, and 30 % during the first 6 months of China’s Wenchuan earthquake reconstruction (Chang et al. 2012). Increases of 200–250 % in the costs of construction materials were observed in tsunami-affected Aceh of Indonesia 2 years after the tsunami (Steinberg 2007).

A number of project actors are often confronted with the problem of escalated costs for indemnifying properties exposed to disasters.



Resourcing Issues Following Earthquake Disaster, Fig. 2 Poor building materials and workmanship in a newly rebuilt house in Aceh, Indonesia

These actors might include agencies of national and regional governments, local government, international agencies, nongovernmental organizations, financial institutions, insurers, reinsurers, and the affected populations. In the last few decades, anticipating demand surge in the event of a natural disaster has been of a primary interest to many insurance and reinsurance companies. Common demand surge models include products from AIR Worldwide, EQECAT, and RMS (Risk Management Solutions).

Although the circumstances contributing to increased reconstruction costs are disaster-specific, there are common explanations for demand surge across events. Research in this area suggests the following factors: the total amount of repair work; the costs of reconstruction materials, labor, and equipment; reconstruction timing; construction contractor fees; general economic conditions; insurance claims handling; and decisions of an insurance company (Olsen and Porter 2013). Among these factors, the increased cost of reconstruction labor and materials is the most common explanation for demand surge.

A more obvious implication of demand surge is likely to be a marked increase in salaries as organizations compete for a decreasing cohort of able workers. Simultaneously, those experienced professionals are likely to be poached by competitor organizations who may offer better fringe benefits and remuneration packages. Spiraling employment linked to tight labor market

conditions ultimately impacts upon the cost of construction projects. The reinforcing effect of interplay between competition and demand surge may persist in the longer term even after the demand surge has passed as evidenced in Aceh of Indonesia following the 2004 Indian Ocean tsunami.

As a result of demand surge, many post-disaster housing projects may suffer from funding shortfalls, further strains on local government expenditure, revised scope of projects, and even project failure. This can be an insurmountable problem and, at worst, result in a huge waste of resources when projects are left half constructed. Some aid agencies in Indonesia, for example, found themselves having to compromise their housing number commitments to local government and beneficiaries, while others experienced significant reconstruction delays and project failures.

The amplification of demand surge impacts comes when there are other developments elsewhere that may pose increased and competing demand and place additional pressure on both human and material resources. For instance, current and future investment in Queensland, Australia's mineral and energy projects are anticipated to have a substantial impact on the demand for labor resources, while the reconstruction and recovery of communities from 2010 to 2011 Queensland floods and cyclones is underway. Whether for reputational advantage, market efficiency, or political reasons, minimizing the impacts of demand surge on the economy has moved from being a concern of local economists to a major component of post-disaster reconstruction strategy, labor market, and economic recovery programs.

Management of Migrant Labor

Those resourcing methods which are based on neoclassical economics and deal mostly with the larger economy tend to consider resource availability as a consequential result of market processes. An organizational perspective is often missing which explains both internal resourcing

dynamics and the linkages between construction organizations and the wider recovery environment. As part of a business initiative, construction organizations may try to upskill workers or import skills from elsewhere outside the disaster zone. The difficulty here is that even when people are trained with construction and engineering skills, only rarely can they be readily competent for the job. Organizations often face a dilemma where they are committed to using local resources but still need to look for expertise from outside to fill vacancies, especially those high-value jobs.

At an organizational level, resourcing skills for reconstruction projects and programs come from a variety of sources. Research in this area demonstrates that sourcing construction professionals from outside for their rebuild projects often have positive results from organizations' perspectives in terms of their confidence in work quality, their economic position, and business performance. In order to attract and retain those nonlocal workers, organizations may have to offer appealing remuneration packages, which may vary from organization to organization. For example, assistance with relocation by an organization in Christchurch, New Zealand, may include the facilitation of immigration, providing temporary accommodation, schooling arrangement for young children, and social networking and employment assistance for the worker's spouse (see Table 1).

In some situations, the capacity of the local construction sector is insufficient for large-scale reconstruction and a large number of workers from outside are needed. There are often challenges associated with managing migrant labor within a local context such as culture, housing, lifestyle, and work ethos. For instance, it was difficult for migrant workers from Java to adapt to a completely different lifestyle and livelihood customs in Aceh after the Indian Ocean tsunami. There was a high rate of worker turnover between projects and organizations. Organizations found it difficult to secure these skills unless constantly increasing labor wages. Influx of labor may also pose a significant challenge to an already strained housing market. A housing crisis may result when

Resourcing Issues Following Earthquake Disaster, Table 1 Demographic features of migrant workers for rebuild in Christchurch

	Engineers	Building control professionals
Origin	Earthquake-prone countries – the USA, Italy, Spain, Chile	Europe, Australia, the USA
Age cohort	25–35	30–50
Work experience	New graduates or more than 2 years experience	At least 5–10 years plus experience

Source: Chang-Richards et al. 2012

there is a lack of reasonable alternative housing available for disaster victims, and/or construction workers compete with locals for limited housing stock and putting those lower-income people at risk of displacement.

In addition, a lack of accommodation for construction workers may become a major constraint to the rebuilding process. The construction sector may experience difficulties sourcing temporary accommodation for out-of-town workers in suitable and affordable solutions. Demand for housing from construction workers is likely to compound the shortage of houses available to residents displaced by the earthquakes. Competing demand for temporary accommodation is likely to contribute to post-disaster demand surge and thus inflation region wide. For example, following the 22 February 2010 earthquake in Christchurch, New Zealand, the rental market failed to meet the needs of inbound construction workers seeking affordable accommodation. Some areas such as “Inner North” and “North West” recorded above-average rent increases of 39 % and 32 %, respectively (MBIE 2013).

In general, the decision to develop site accommodation for workers is based on factors such as impacts of workers’ accommodation on communities, quality of housing stock, financial viability of investment, scale of demand, and accessibility to land and infrastructure. In low-income nations, local government and construction organizations play a large role in providing on-site temporary accommodation for workers. This may be

assisted by nongovernment organizations. However, gaining consents from local authorities for developing such an accommodation project on public or private land might be a particularly daunting task in high-income countries. Local government needs to revise the housing and planning policies to allow either temporary or semi-permanent structures to be built. It also requires the revision of zoning and building code controls to facilitate and encourage the development of different housing models.

Summary

Resourcing for post-disaster reconstruction is a complex and dynamic process in which resource demand is unevenly distributed across different stages and different sectors of rebuild. Varying the resourcing approach involves leverage of stakeholders’ influence and market forces to a varied degree. This implies facilitating organizational responsibility over sourcing and managing resources to meet reconstruction tasks. It also implies facilitating collective responsibility among project actors over resource needs to mitigate the resource risk for the rebuild.

Resourcing bears on a variety of issues inherent in the post-disaster reconstruction context. This process requires a systemic and dynamic support on recovery planning systems, industry capacity, regulatory framework, and economic issues at both project and organizational levels. Research has shown that following a large-scale disaster, resource risks will exist but can be managed. By linking resourcing with broader plans for sustainable and equitable post-disaster reconstruction, more fundamental actions both pre- and post-disaster can be taken to reduce socially and physically produced resource shortages.

Cross-References

- [Earthquakes and Their Socio-economic Consequences](#)
- [Economic Recovery Following Earthquakes Disasters](#)

- [Interim Housing Provision Following Earthquake Disaster](#)
- [Legislation Changes Following Earthquake Disasters](#)

References

- Australian Securities and Investments Commission (2007) Making home insurance better, report 89. Australian Securities and Investments Commission, Canberra
- Chang Y, Wilkinson S, Seville E, Potangaroa R (2010) Resourcing for a resilient post-disaster reconstruction environment. *Int J Disaster Res Built Environ* 1(1):65–83
- Chang Y, Wilkinson S, Seville E, Potangaroa R (2012) Changes in resource need for post-disaster reconstruction: a longitudinal study in China. *Build Res Inform* 40(3):327–336
- Chang-Richards Y, Seville E, Wilkinson S, Brunsdon D (2012) Challenges and opportunities for construction organisations, APEC project on SME Resilience: Resilient Organisations Research Report 2012/10. Available at: http://www.resorgs.org.nz/images/stories/pdfs/Organisationsfacingcrisis/apec_report5_reconstruction_final.pdf
- Kuzak D, Larsen T (2005) Use of catastrophe models in insurance rate making. In: Grossi P, Kunreuther H (eds) *Catastrophe modeling: a new approach to managing risk*. Springer, New York
- MBIE (2013) Housing pressures in Christchurch: a summary of the evidence. Ministry of Business, Innovation and Employment (MBIE), Wellington. Available at: <http://www.building.dbh.govt.nz/nz-building-research>
- Olsen AH, Porter KA (2013) Storm surge to demand surge: exploratory study of hurricanes, labor wages, and material prices. *Nat Hazard Rev (ASCE)* 14(4):247–257
- Shephard RB, Smith EGC, Spurr DD (1997) Earthquake insurance loss assessment for regions of Australia. *Bull N Z Natl Soc Earthq Eng* 30(1):32–39
- Steinberg F (2007) Housing reconstruction and rehabilitation in Aceh and Nias, Indonesia – rebuilding lives. *Habitat Int* 31(1):150–166
- Walker G (1995) Discussion on physical reconstruction. Paper presented at the Wellington after the quake: the challenges of rebuilding cities, Wellington
- Walker GR, Minor JE (1979) Cyclone Tracy in retrospect: a review of its impact on the Australian community. Paper presented at the fifth international conference on wind engineering, Fort Collins
- Wu JY, Lindell MK (2004) Housing reconstruction after two major earthquakes: the 1994 northridge earthquake in the United States and the 1999 Chi-chi earthquake in Taiwan. *Disasters* 28(1):63–81

Response Spectrum Analysis of Structures Subjected to Seismic Actions

Michalis Fragiadakis

School of Civil Engineering, Laboratory for Earthquake Engineering, National Technical University of Athens (N.T.U.A.), Athens, Greece

Synonyms

Lateral force method; Modal analysis; Modal response spectrum analysis; Modal superposition; Response spectrum; Seismic performance assessment

Introduction

Response spectrum analysis (RSA) is a method widely used for the design of buildings. Conceptually the method is a simplification of modal analysis, i.e., response history (or time history) analysis (RHA) using modal decomposition, that benefits from the properties of the response spectrum concept. The purpose of the method is to provide quick estimates of the peak response without the need to carry out response history analysis. This is very important because response spectrum analysis (RSA) is based on a series of quick and simple calculations, while time history analysis requires the solution of the differential equation of motion over time. Despite its approximate nature, the method is very useful since it allows the use of response spectrum, a very convenient way to describe seismic hazard.

RSA is very appealing to practicing engineers because seismic loading is defined by means of a response spectrum. All design codes worldwide define seismic input (or hazard) by means of a code-compliant, typically smooth, response spectrum that can be easily adjusted according to the site seismic hazard. Such spectra are able to implicitly take into consideration the fact that structures are designed to resist seismic actions

by deforming inelastically. Two variations of RSA are offered in almost every seismic design code. Following the notation of Eurocode 8 (EC8 2004), these variations are the “lateral force method” and the “modal response spectrum analysis.” This entry discusses the underlying concepts of both methods in a comprehensive manner. The interested reader is also advised to consult the classic textbooks of Chopra (2000) and Clough and Penzien (1975).

Modal Analysis

Before proceeding to the discussion of response spectrum analysis, it is essential to have a good understanding of modal analysis, also known as modal superposition method. Modal analysis aims at transforming a fully coupled problem of N degrees of freedom to N uncoupled single-degree-of-freedom (SDOF) problems that can be solved individually, and, in the end, the individual solutions can be superimposed to obtain the solution of the initially coupled problem. The advantage is that simpler methods, including closed-form methods, can be used to solve the single-degree-of-freedom problem. From a computing standpoint, the number of operations required is substantially reduced when modal decomposition is preferred (Bathe 1996). A second reason is that only a small number of modes need to be taken into consideration, while the rest can be neglected with a minimum loss of accuracy. This considerably speeds up the whole process, while the number of modes required depends on the type of loading and the structural properties (i.e., stiffness, damping).

The equation of motion of a system with N degrees of freedom is written as

$$\mathbf{m}\mathbf{u} + \mathbf{c}\dot{\mathbf{u}} + \mathbf{k}\mathbf{u} = \mathbf{q}(t) \quad (1)$$

where $\mathbf{m}$, $\mathbf{c}$, and $\mathbf{k}$ are the mass, damping, and stiffness matrices, respectively, while $\ddot{\mathbf{u}}$, $\dot{\mathbf{u}}$, and $\mathbf{u}$ denote the vectors of acceleration, velocity, and displacement and are always functions of time. $\mathbf{q}(t)$ is the vector of applied loads, again defined as function of time. If the dynamic load is

a ground motion history (acceleration time history), then $\mathbf{q}(t)$ is obtained as

$$\mathbf{q}(t) = -\mathbf{m}\boldsymbol{\varsigma}\ddot{u}_g(t) \quad (2)$$

where $\boldsymbol{\varsigma}$ is a vector of order N whose entries are equal to 1 for translational degrees of freedom in the direction of the ground motion and zero otherwise. It is also reminded that the number of modes is N , equal to the number of the degrees of freedom of the system.

By definition, the mode shape vector $\boldsymbol{\Phi}_n$ describes the shape of the n th mode, and the vector of displacements is equal to the sum:

$$\mathbf{u}(t) = \sum_{n=1}^N \boldsymbol{\Phi}_n y_n(t) \quad (3)$$

The coupled equation of motion of Eq. 1 can be uncoupled after substituting $\mathbf{u}(t)$ using Eq. 3 and left-multiplying it with $\boldsymbol{\Phi}^T$. This is achieved with the aid of the orthogonality condition, where the products $\boldsymbol{\Phi}_m^T \mathbf{m} \boldsymbol{\Phi}_n$, $\boldsymbol{\Phi}_m^T \mathbf{c} \boldsymbol{\Phi}_n$ and $\boldsymbol{\Phi}_m^T \mathbf{k} \boldsymbol{\Phi}_n$ are equal to zero if $n \neq m$. The following generalized quantities can be defined when $n = m$:

$$\begin{aligned} M_n &= \boldsymbol{\Phi}_n^T \mathbf{m} \boldsymbol{\Phi}_n \\ C_n &= \boldsymbol{\Phi}_n^T \mathbf{c} \boldsymbol{\Phi}_n \\ K_n &= \boldsymbol{\Phi}_n^T \mathbf{k} \boldsymbol{\Phi}_n \end{aligned} \quad (4)$$

It is also assumed that

$$Q_n(t) = \boldsymbol{\Phi}_n^T \mathbf{q}(t) = -\boldsymbol{\Phi}_n^T \mathbf{m} \boldsymbol{\varsigma} \ddot{u}_g(t). \quad (5)$$

The coupled equation of motion (Eq. 1) is equivalent to the following set of equations:

$$\begin{aligned} M_1 \ddot{y}_1 + C_1 \dot{y}_1 + K_1 y_1 &= Q_1(t) \\ M_2 \ddot{y}_2 + C_2 \dot{y}_2 + K_2 y_2 &= Q_2(t) \\ \dots &\dots \\ M_n \ddot{y}_n + C_n \dot{y}_n + K_n y_n &= Q_n(t) \end{aligned} \quad (6)$$

If we divide by M_n and assume $C_n = 2\zeta_n \omega_n$ and $K_n = M_n \omega_n^2$, the equations are further simplified to

$$\ddot{y}_n + 2\zeta_n \omega_n \dot{y}_n + \omega_n^2 y_n = -\frac{L_n}{M_n} \ddot{u}_g(t) \quad (7)$$

where

$$L_n = \Phi_n^T \mathbf{m} = \sum_{j=1}^N m_{nj} \Phi_{nj} \quad (8)$$

For brevity, only the last equation of Eq. 7 is shown. Equations 6 or 7 are not coupled and therefore can be solved separately to obtain y_n . Equation 7 can be seen as the equation of motion of a SDOF with frequency ω_n subjected to a strong ground motion whose amplitude $\ddot{u}_g(t)$ has been multiplied with L_n/M_n . Otherwise, the solution of Eq. 7 is the solution of a SDOF with frequency ω_n subjected $\ddot{u}_g(t)$ and then multiplied with L_n/M_n .

The ratio L_n/M_n is the participation factor of mode n , which always sums to one and may take both positive and negative values. This quantity depends on how the modes have been normalized and is not equal to the contribution of a mode to a response quantity.

The vector of displacements $\mathbf{u}$ is back-calculated using Eq. 3, and every other response quantity can be subsequently determined once $\mathbf{u}$ is known. For example, the vector of elastic forces will be

$$\begin{aligned} \mathbf{f}(t) &= \mathbf{k}\mathbf{u}(t) = \dots \\ &\dots = \mathbf{k}\Phi_1 y_1(t) + \mathbf{k}\Phi_2 y_2(t) + \dots + \mathbf{k}\Phi_n y_n(t). \end{aligned} \quad (9)$$

For damped systems, the above decoupling is feasible only if the damping matrix $\mathbf{c}$ is orthogonal, otherwise Eq. 6 does not apply. Usually $\mathbf{c}$ is derived as a combination of $\mathbf{k}$ and $\mathbf{m}$, i.e., $\mathbf{c} = \alpha_0 \mathbf{k} + \alpha_1 \mathbf{m}$, where α_0, α_1 are constants. This form of damping is known as Rayleigh damping. Another approach would be to form the damping matrix so that every mode is damped with $C_n = 2\zeta_n \omega_n$. This is achieved with the aid of the formula (Chopra 2000):

$$\mathbf{c} = \mathbf{m} \left(\sum_{n=1}^N \frac{2\zeta_n \omega_n}{M_n} \Phi_n^T \Phi_n \right) \mathbf{m} \quad (10)$$

A fundamental property of modal analysis is that the response can be accurately captured if a, relatively, small number of modes are

considered. Therefore, the question is how many modes should be included in modal analysis. The answer depends on the applied load and on the response parameter examined. For example, more modes are required to accurately capture the base shear than the roof displacement. Since the exact solution is not a priori known, one has to identify the response parameters that are likely to be sensitive to high modes and then decide the number of modes to include. In lieu of the above, some building codes suggest that the number of modes considered should be chosen so that the effective modal masses for the modes taken into account amounts to at least the 90 % of the total mass of the structure. Due to the relationship of the effective modal mass and the base shear, this rule implies that the error in the base shear estimate should be less than 10 %. Moreover, according to Eurocode 8 (EC8 2004), the analysis should include all modes with effective modal masses greater than 5 % of the total.

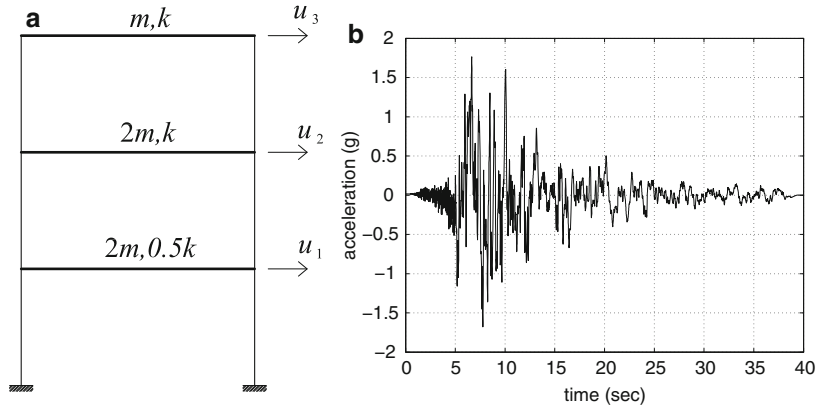
Example 1: Modal Analysis of a Plane Frame Determine the response of the three-storey plane frame of Fig. 1a when subjected to the Imperial Valley (1979) ground motion record (Fig. 1b). Assume $k = 10,000$ kN/m, $m = 20$ t, and 5 % damping of the critical for every mode.

The mass and the stiffness matrices are obtained as follows:

$$\begin{aligned} \mathbf{k} &= \begin{bmatrix} 1.5k & -k & 0 \\ -k & 2k & -k \\ 0 & -k & k \end{bmatrix} \\ &= \begin{bmatrix} 15,000 & -10,000 & 0 \\ -10,000 & 20,000 & -10,000 \\ 0 & -10,000 & 10,000 \end{bmatrix} \text{ kN/m} \\ \mathbf{m} &= \begin{bmatrix} 2m & 0 & 0 \\ 0 & 2m & 0 \\ 0 & 0 & m \end{bmatrix} = \begin{bmatrix} 40 & 0 & 0 \\ 0 & 40 & 0 \\ 0 & 0 & 20 \end{bmatrix} t. \end{aligned}$$

The eigenvalue problem, $[\mathbf{k} - \mathbf{m}^2] = 0$, is first solved to obtain the eigenmodes (or just “modes”) and the eigenperiods (or just “periods”) of the frame. The eigenperiods are $T = [0.98, 0.31, 0.21]$ s and the mode vector is

Response Spectrum Analysis of Structures Subjected to Seismic Actions, Fig. 1 (a) Plane frame geometry, (b) the Imperial Valley record (15 Oct. 1979)



$$\Phi = [\Phi_1 \Phi_2 \Phi_3] = \begin{bmatrix} 0.69 & -0.95 & 0.38 \\ 0.92 & 0.16 & -0.83 \\ 1.00 & 1.00 & 1.00 \end{bmatrix}$$

The first column gives the first mode and every row corresponds to the modal displacement of the respective storey. The system of Eq. 7 now becomes

$$\begin{aligned} \ddot{y}_1 + 0.64\dot{y}_1 + 40.8y_1 &= -\frac{84.22}{72.62}\ddot{u}_g(t) \\ \ddot{y}_2 + 2.04\dot{y}_2 + 418.2y_2 &= \frac{11.32}{56.90}\ddot{u}_g(t) \\ \ddot{y}_3 + 3.03\dot{y}_3 + 916.0y_3 &= -\frac{2.09}{53.60}\ddot{u}_g(t) \end{aligned}$$

The modal displacements $y_n(t)$ are obtained solving the above equations either using Duhamel's integral or, preferably, numerically using the Newmark's method (Bathe 1996; Chopra 2000). The modal response histories of $y_n(t)$ are shown in Fig. 2. According to the three plots, the modal displacement of the first mode is considerably larger than that of modes two and three (note the different scale of the vertical axes). Moreover, the response histories of higher modes have a more rich frequency content compared to that of lower modes. Figure 3 shows the modal displacements of the third storey, obtained using Eq. 3. The first three plots show the displacement response history of every mode and the fourth plot shows the actual/total displacement of the third storey. Regarding the amplitude and the frequency content, the observations of Fig. 2 still hold.

Moreover, the maxima of every response history occur at different time instants. The figure also implies that we can have a good approximation of the final solution using the first mode only and discarding modes two and three with a minor loss of accuracy.

$$\mathbf{c} = \begin{bmatrix} 71.5068 & -28.9232 & -5.0115 \\ & 75.9453 & -26.4174 \\ \text{sym} & & 40.4784 \end{bmatrix}$$

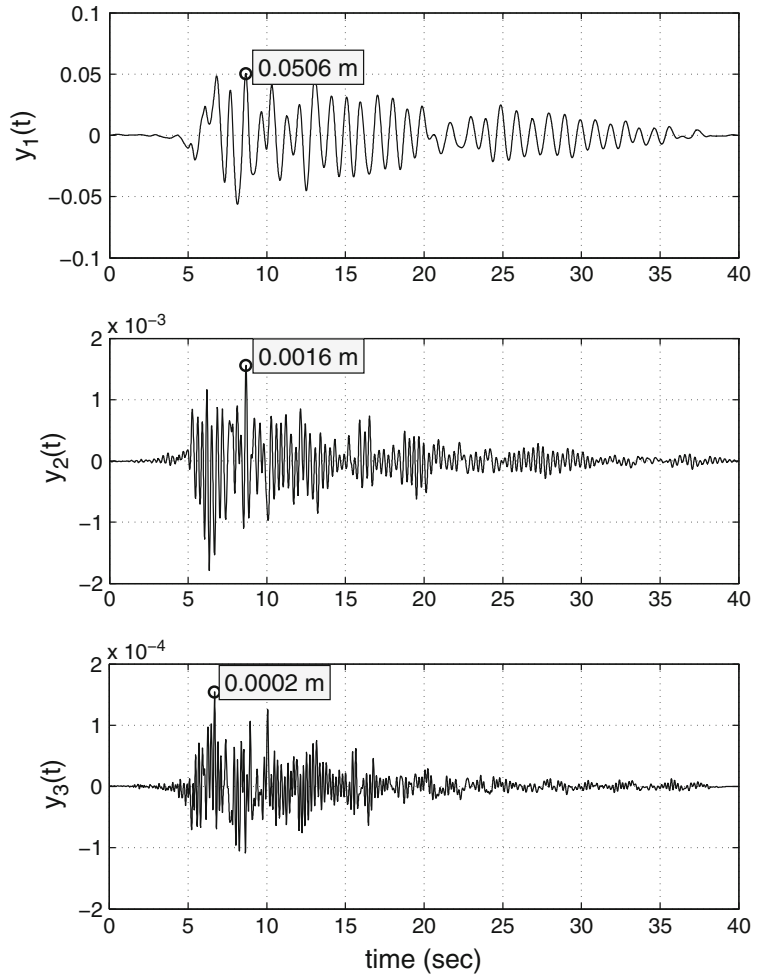
The final solution of the last plot in the bottom of Fig. 3 has been verified also by solving the equation of motion (Eq. 1) with the aid of Newmark's method. For this purpose, we use Eq. 10 to obtain the damping matrix $\mathbf{c}$ shown above. Direct integration algorithms are usually simpler and their additional computational effort is easily handled by modern computers. Modal analysis may be preferred because it (i) helps decompose the problem and allows drawing qualitative conclusions, (ii) is faster when only the most significant modes are considered, and (iii) is the basis of response spectrum modal analysis.

Response Spectrum

In seismic design problems, in most occasions, it is sufficient to know only the peak values of the response, defined as the maximum of the absolute value of a response quantity $r(t)$ that varies in time (Chopra 2005):

$$r_0 = \max|r(t)| \quad (11)$$

**Response Spectrum
Analysis of Structures
Subjected to Seismic
Actions, Fig. 2** Response
histories of modal
displacements $y_n(t)$ (Eq. 3)



where r_0 is the absolute value of $r(t)$. The response spectrum of a quantity is the plot of the maximum (or peak) value of a response quantity (e.g., displacement, velocity, acceleration) against the full range of interest of the natural vibration period values T_n .

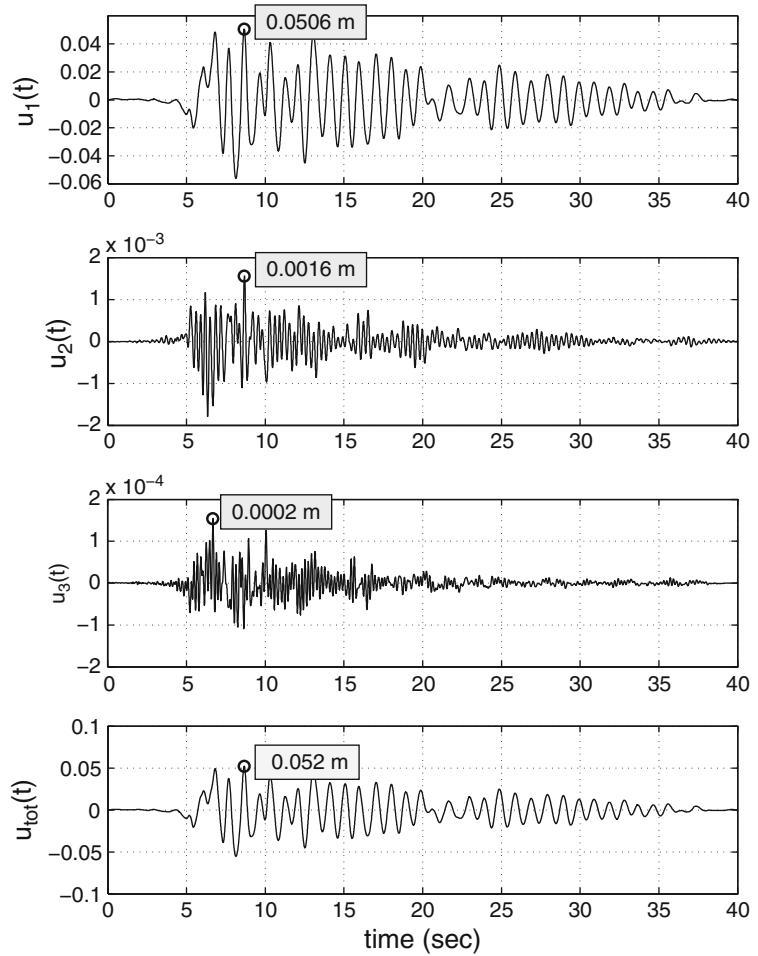
The word spectrum (plural spectra or spectrums) was first used in science within the field of optics to describe the rainbow of colors in visible light when separated using a prism. In other words, “spectrum” implies a broad range of conditions or behaviors grouped and studied together. If a given period value T_n refers to a different SDOF structure, a response spectrum of a quantity by definition provides *the maximum value of the response quantity of interest under a given ground motion for the whole range of*

structures of interest. Therefore, such plots give the maximum response for the whole range (or spectrum) of SDOF structures.

Properties of Response Spectra

Figure 4 shows the response spectra of three recorded ground motion records. Some very interesting observations regarding the properties of the response spectra can be made. All three spectra intersect the vertical axis at an acceleration value equal to the peak ground acceleration, i.e., the maximum acceleration value of the recorded ground motion. This point corresponds to a very rigid SDOF system with period equal to zero. Moreover, the Manjil (1990) and the Assisi-Stallone (1997) ground motions are clearly stronger compared to the Parkfield (1966) record,

Response Spectrum Analysis of Structures Subjected to Seismic Actions, Fig. 3 Response history of the modal displacements of the third storey. The plot in the bottom shows the actual/total response history of the third storey



while the Manjil record has a richer frequency content compared to the Assisi-Stallone record. This stems from the fact that the Manjil record maintains high acceleration values for a wide range of periods, while the Assisi-Stallone spectrum decays quickly. A rich frequency content is also observed in the Parkfield record.

Apart from the observations of the above paragraph, the most important property of a response spectrum is that it provides the maximum acceleration of a single-degree-of-freedom (SDOF) system without the need of performing a dynamic analysis, since all necessary calculations have been already performed in order to draw the spectrum. If $S_a(T_n, \xi)$ is the $\xi\%$ -damped spectral acceleration (ordinate of the acceleration spectrum), the peak deformation of a SDOF with period T_n and damping ξ will be

$$u_{\max} = \omega_n^2 S_a(T_n, \xi) = \left(\frac{T_n}{2\pi}\right)^2 S_a(T_n, \xi) \quad (12)$$

Similarly the peak value of the equivalent static force $f_{s,\max}$ will be

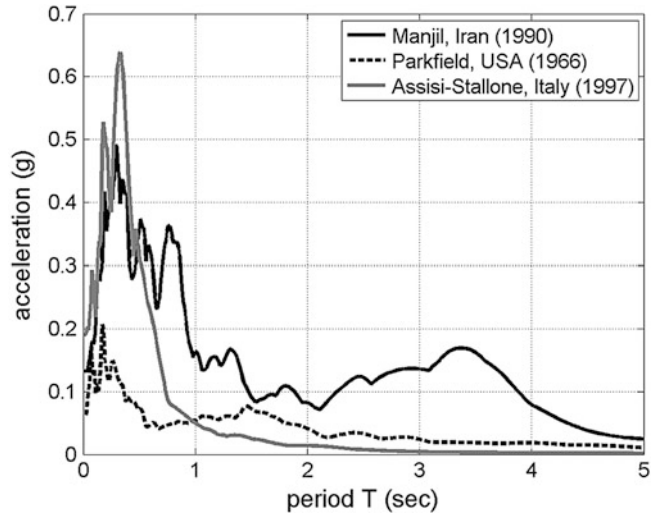
$$\begin{aligned} f_{s,\max} &= k u_{\max} = m S_a(T_n, \xi) \\ &= \frac{S_a(T_n, \xi)}{g} W \end{aligned} \quad (13)$$

where k , m , and W are the stiffness, mass, and weight of the SDOF, respectively.

Integrating the acceleration response history would give the velocity, while the integral of the velocity history provides the displacement response history. This means that acceleration, velocity, and displacement are proportional

Response Spectrum Analysis of Structures Subjected to Seismic Actions,

Fig. 4 Acceleration response spectra for three strong ground motion records



(by a factor ω), except for a phase shift. The phase shift does not influence the maximum response values and therefore the following approximations can be made:

$$\begin{aligned} S_d(T_n, \xi) &= |u_{\max}| \\ S_v(T_n, \xi) &= |\dot{u}_{\max}| \approx \omega S_d(T_n, \xi) = PSV \\ S_d(T_n, \xi) &= |\ddot{u}_{\max}| \approx \omega^2 S_d(T_n, \xi) = \omega PSV = PSA \end{aligned} \quad (14)$$

where PSV and PSA are the pseudospectral velocity and the pseudospectral acceleration, respectively. They are not the exact maxima of velocity and acceleration, but they are very close to the actual values and are usually used in engineering practice.

Design or Code-Compliant Spectra

The response spectra of Fig. 4 refer to the response of all possible SDOF systems of specific ground motion records. On the other hand, a design spectrum recognizes that ground motions at a site may originate from different seismic sources. Therefore, a design spectrum is representative of an ensemble of ground motions and, contrary to the single-record spectra of Fig. 4, its shape is always smooth. Design spectra have been derived from statistical analysis from an ensemble of response spectra and correspond either to a mean or to a mean plus one standard deviation spectrum. They are also known as

“uniform hazard spectra,” since they refer to the same probability of exceedance for the whole range of periods.

A procedure for constructing a design spectrum from peak values of ground acceleration, velocity, and displacement has been presented by Newmark and Hall (1982). The procedure uses the amplification factors of Table 1 and is applied in a logarithmic tripartite log paper where the horizontal axis denotes frequency (Hz). The steps required are quite simple (Fig. 5): (i) first, draw the ground parameters (peak acceleration, velocity, and displacement) each parallel to the corresponding axis. (ii) Then draw lines parallel to them at distances obtained from Table 1. For example, for damping 5 %, the amplification factor of the median acceleration is $3.21 - 0.68 \ln(5) = 2.11$ and 1.65 and 1.39 for the velocity and the displacement spectrum, respectively. (iii) Draw a line connecting the response spectrum at 8 Hz with the ground motion at 33 Hz so that the response of very stiff structures is equal to that of the ground. (iv) Finally, draw another line connecting the response spectrum at 1/10 Hz with the ground motion curve at 1/33 Hz so that the peak response of very flexible systems coincides with the peak ground deformation. Other approaches for creating both elastic and inelastic spectra can be found in the literature.

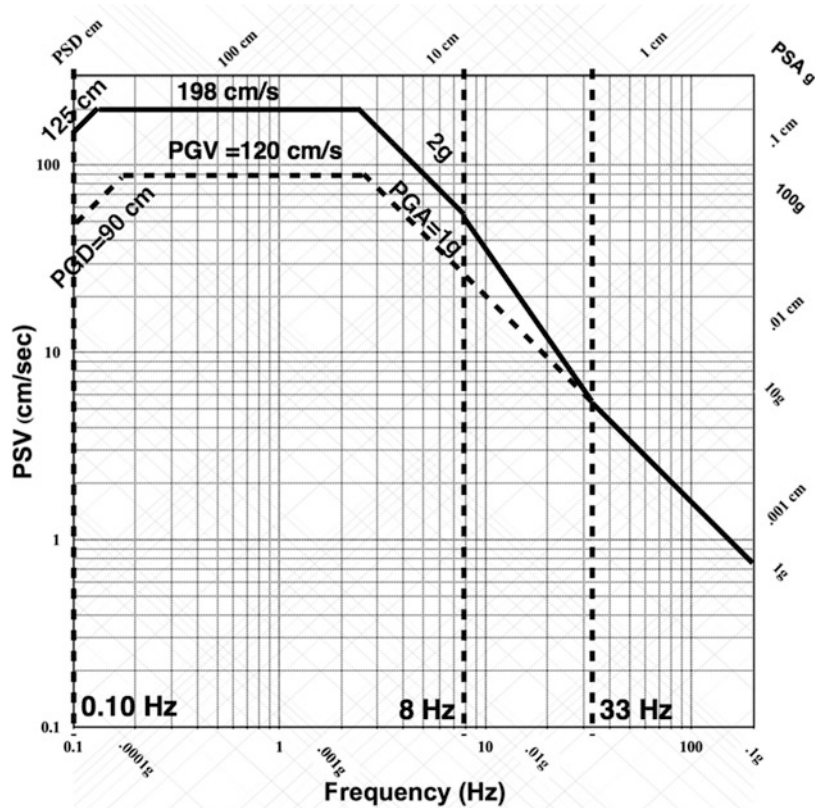
Design codes give smooth elastic or design spectra. The distinction between elastic and design

Response Spectrum Analysis of Structures Subjected to Seismic Actions, Table 1 Spectral amplification factors for 84-percentile confidence (Newmark and Hall 1982)

Damping (%)	Acceleration	Velocity	Displacement
50 percentile (median)	3.21–0.68 ξ	2.31–0.41 ξ	1.82–0.27 ξ
84 percentile	4.38–1.04 ξ	3.38–0.67 ξ	2.73–0.45 ξ

**Response Spectrum
Analysis of Structures
Subjected to Seismic
Actions,**

Fig. 5 Construction of design spectra according to Newmark and Hall (1982)



spectrum is that the design spectrum is derived from the elastic by dividing the ordinates of the latter by a response modification factor R (according to the terminology used in the USA) or a behavior factor called q in the European practice. Some codes, such as Eurocode 8 (European Committee for Standardization 2004), anchor the spectrum to the peak ground acceleration, but

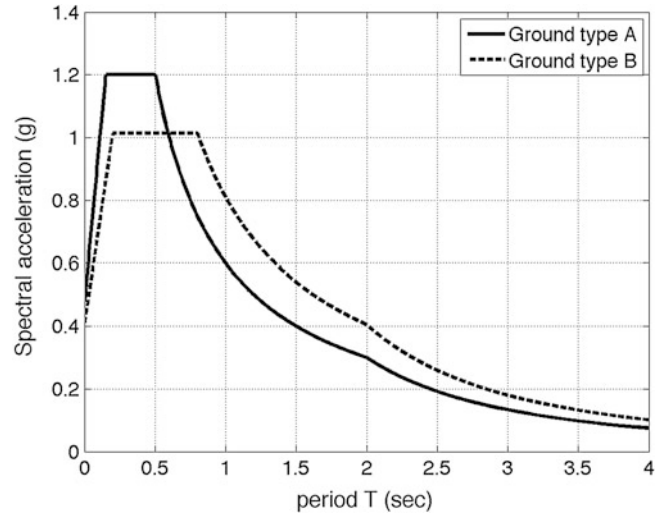
this is not always the case. In seismic codes that do not specify explicitly the PGA value, it can be inferred either from the acceleration value at period equal to zero or by dividing the flat segment of the spectrum (plateau) with the acceleration amplification factors of Table 1.

The Eurocode 8 elastic spectrum is given by the expression below and is plotted in Fig. 6:

$$S_a(T, \xi) = \begin{cases} \alpha_g S [1 + (\eta 2.5 - 1)(T/T_B)] & , \text{ if } 0 \leq T \leq T_B \\ \alpha_g S \eta 2.5 & , \text{ if } T_B \leq T \leq T_C \\ \alpha_g S \eta 2.5 (T_C/T) & , \text{ if } T_C \leq T \leq T_D \\ \alpha_g S \eta 2.5 (T_C T_D/T^2) & , \text{ if } T_D \leq T \leq 4\text{sec} \end{cases} \quad (15)$$

Response Spectrum Analysis of Structures Subjected to Seismic Actions, Fig. 6

The elastic design spectrum of European Committee for Standardisation (2004)



where S is the soil factor, and T_B , T_C , and T_D are period values that change the shape of spectrum according to the soil conditions. Eurocode 8 recommends the above spectrum for earthquakes of magnitude above 5.5, while another expression is recommended for smaller earthquakes. η is the damping correction factor equal to 1 when $\xi = 5\%$ (typical for reinforced concrete structures) and obtained as

$$\eta = \sqrt{10/(5 + \xi)} \geq 0.55 \quad (16)$$

α_g is the design acceleration which is equal to the importance factor γ_I times the design peak ground acceleration (PGA). Figure 6 shows two response spectra obtained with the formula of EC8 (Eq. 15). The spectrum with the solid line refers to ground type A ($S = 1.2$, $T_B = 0.15$, $T_C = 0.5$, $T_D = 2.0$) and will be used for the design of a reinforced concrete structure ($\xi = 5\%$). The importance factor (γ_I) is considered equal to 1.0 and the design PGA is 0.40 g. The spectrum with the dashed line refers to ground type B ($S = 1.35$, $T_B = 0.2$, $T_C = 0.8$, $T_D = 2.0$), PGA equal to 0.30 g, and $\gamma_I = 1.0$. Note that both spectra intercept the vertical axis at a value equal to $a_g S$, while the horizontal plateau (for $\xi = 5\%$) is at $a_g S 2.5$. 2.5 is the

acceleration spectral amplification factor approximately equal to that given by the expression of Table 1. Both PGAs adopted for the site under consideration have a 10 % probability of being exceeded in 50 years (return period $T_m = 475$ years), and therefore the seismic hazard of both response spectra is uniform and equal to a 10 % probability of being exceeded in 50 years.

Response Spectrum Modal Analysis

Peak Modal Responses

Modal analysis provides the entire response history for a given ground motion record. For design purposes, its application requires a design ground motion record that is representative of the seismic hazard at the site. The use of such records is rare and requires special skills to select them. Hence, in engineering practice, the seismic hazard is defined preferably with the aid of regional response spectra as discussed the “[Properties of Response Spectra](#)” section.

Moreover, for design purposes, we usually use the maximum value of a response parameter and not the entire response history. Since every mode can be treated as an independent SDOF system, the maximum response values of a mode can be

easily obtained from the corresponding response spectrum (Eqs. 12 and 13).

If $S_d(T_n, \xi)$, $S_v(T_n, \xi)$, and $S_a(T_n, \xi)$ denote the spectral displacement, velocity and acceleration, respectively, the maximum modal displacements are obtained from a response spectrum as

$$y_{n,\max} = \frac{L_n}{M_n} S_d(T_n, \xi) = \frac{L_n}{M_n} \frac{T_n^2}{4\pi^2} S_a(T_n, \xi) \quad (17)$$

The maximum displacement and the equivalent lateral force of the j th storey will be

$$\begin{aligned} u_{n,\max}^j &= \frac{L_n}{M_n} S_d(T_n, \xi) \Phi_{nj} \\ &= \frac{L_n}{M_n} \frac{T_n^2}{4\pi^2} S_a(T_n, \xi) \Phi_{nj} \end{aligned} \quad (18)$$

$$s_{n,\max}^j = \frac{L_n}{M_n} m_j S_a(T_n, \xi) \Phi_{nj} \quad (19)$$

Alternatively, the maximum values of the member forces due to the n th mode can be obtained by static analysis by loading the structure with the maximum equivalent lateral forces of Eq. 19. The calculation would be

$$\mathbf{f}_n = \left(\frac{L_n}{M_n} \mathbf{m} \Phi_n \right) S_a(T_n, \xi) \quad (20)$$

and the modal displacements are obtained as $\mathbf{s}_{n,\max} = \mathbf{k}^{-1} \mathbf{f}_n$.

Modal Combination Rules

RSA provides only the maximum response parameter values of every mode. Once the maxima of every mode are known, we need to calculate the exact maximum value of a response quantity. This calculation is not straightforward. If the maximum value of every mode had occurred exactly at the same time instant, then combining Eqs. 3 and 17, we would get

$$\mathbf{u}_o = \sum_{n=1}^N \left| \frac{L_n}{M_n} \frac{T_n^2}{4\pi^2} \Phi_n S_a(T_n, \xi) \right| \geq \max(\mathbf{u}(t)) \quad (21)$$

The equation above provides a conservative estimation of $\mathbf{u}_{\max}$, since, as also shown in Fig. 2, the maxima of every mode do not occur concurrently. This upper bound is usually too conservative and rarely used for design. In practice, modal combination rules, as discussed in the next section, are preferred.

The Square Root of the Sum of Squares (SRSS) Rule

The most common rule for modal combination is the *Square Root of Sum of Squares* (SRSS) rule. According to this rule, the peak response of every mode is squared and then the squares are summed. The estimation of the maximum response quantity of interest is the square of the sum. If we denote r as an arbitrary response quantity (e.g., displacement, stress resultants) and r_n is its peak modal response, their SRSS combination is written as

$$r_o \approx \sqrt{r_1^2 + r_2^2 + \cdots + r_n^2} \quad (22)$$

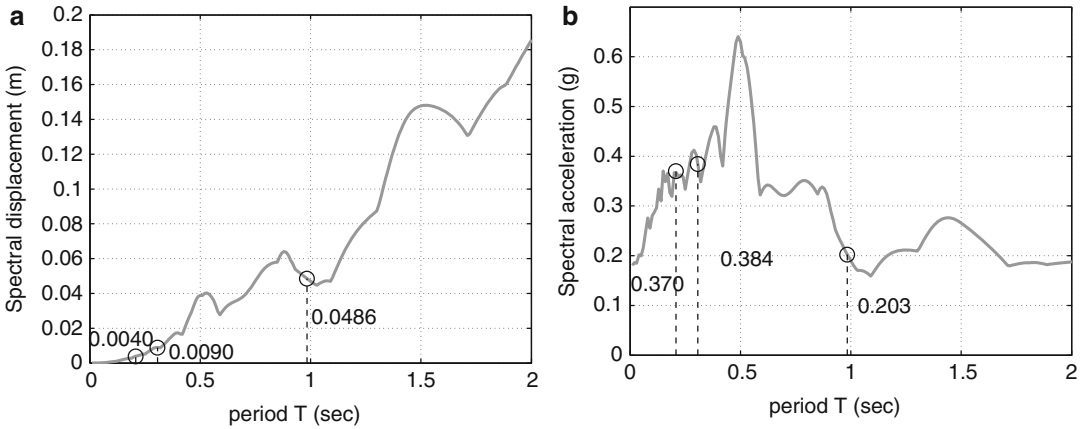
In general, this combination rule gives good response estimates, especially for two-dimensional problems. The major limitation is that in order to produce satisfying estimates, the modes should be well separated, i.e., the eigenfrequencies should not have close values. If this condition is not met, the CQC method should be used instead. A criterion to determine if two modes are well separated is

$$\frac{1}{\beta_{nm}} \geq 1 + 0.1 \sqrt{\xi_n \xi_m} \quad (23)$$

where $\beta_{nm} = \omega_m/\omega_n = T_n/T_m$ and ξ_n, ξ_m is the damping ratio of modes n and m , respectively.

The Complete Quadratic Combination (CQC) Rule
The complete quadratic combination (CQC) rule overcomes the limitations of the SRSS rule and should be adopted if the condition of Eq. 23 is not satisfied. The rule is expressed as follows:

$$r_o \approx \sqrt{\sum_{n=1}^N \sum_{m=1}^N \epsilon_{nm} r_n r_m} \quad (24)$$



Response Spectrum Analysis of Structures Subjected to Seismic Actions, Fig. 7 (a) Displacement response spectrum, (b) acceleration response spectrum

where ϵ_{nm} is a correlation coefficient that takes values in the 0,1 range and is equal to 1 when

$n = m$. If β_{nm} is that of Eq. 23, the correlation term is calculated as

$$\epsilon_{nm} = \frac{8\sqrt{\xi_n \xi_m}(\xi_n + \beta_{nm}\xi_m)\beta_{nm}^{3/2}}{(1 - \beta_{nm}^2)^2 + 4\xi_n \xi_m \beta_{nm}(1 + \beta_{nm}^2) + 4(\xi_n^2 + \xi_m^2)\beta_{nm}^2} \quad (25)$$

If the same modal damping is used for modes n and m ($\xi_n = \xi_m = \xi$), the equation reduces to

$$\epsilon_{nm} = \frac{8\xi^2(1 + \beta_{nm})\beta_{nm}^{3/2}}{(1 - \beta_{nm}^2)^2 + 4\xi^2\beta_{nm}(1 + \beta_{nm}^2)} \quad (26)$$

Note that according to Eq. 26 when $\beta_{nm} = 1$, i.e., for two modes with equal frequencies, their correlation coefficient ϵ_{nm} is equal to 1. When the modes are well separated, the $n \neq m$ terms of ϵ_{nm} become very small and the CQC rule gives estimates very close to those of the SRSS rule. ϵ_{nm} can take both positive and negative values and the CQC estimate may be above or below that of SRSS. Further discussion on the CQC rule can be found in Der Kiureghian (1981) and Wilson et al. (1981).

Both the SRSS and the CQC method should be applied directly on the response quantity of

interest according to its modal value. Obtaining the value of a response quantity by combining the peak values of other quantities must be avoided as it may result to gross errors.

Example 2: Response Spectrum Modal Analysis of a Plane Frame For the building of Fig. 1a, estimate the storey displacements using response spectrum analysis. The structure is again subjected to the Imperial Valley ground motion record of Fig. 1b.

The spectral displacements and the spectral accelerations of the Imperial Valley (1979) ground motion record are shown in Fig. 7. We show both displacement and acceleration spectra for comparison (although the acceleration spectrum is always adopted in practice). The modal displacements $y_{n,\max}$ are calculated with either of the expressions of Eq. 17. Thus, for the first mode, we obtain:

$$\begin{aligned}
 y_{1,\max} &= \frac{L_n}{M_n} S_d(T_n, \xi) \\
 &= \frac{84.22}{72.62} \times 0.0486 = 0.056m, \text{ or} \\
 y_{1,\max} &= \frac{L_n}{M_n} \frac{T_n^2}{4\pi^2} S_a(T_n, \xi) \\
 &= \frac{84.22}{72.62} \frac{0.98^2}{4\pi^2} \times 0.203 \times 9.81 = 0.056m
 \end{aligned}$$

Similarly, we obtain $y_{2,\max} = 0.0018m$ and $y_{3,\max} = 0.0002m$ for the second and the third mode, respectively. Both methods yield practically the same value; the minor differences are due to round-off errors. These values should coincide with the maxima of Fig. 2. The differences observed are again due to round-off errors. Note that we have dropped the negative signs using the absolute values of all response quantities.

The modal displacements of the first mode are (Eq. 3)

$$\begin{aligned}
 \mathbf{u}_{1,\max} &= \Phi_1 y_{1,\max} = \begin{bmatrix} 0.69 \\ 0.92 \\ 1.0000 \end{bmatrix} \times 0.056 \\
 &= \begin{bmatrix} 0.0388 \\ 0.0519 \\ 0.0565 \end{bmatrix} m.
 \end{aligned}$$

The $\mathbf{u}_{n,\max}$ values are summarized in Table 2.

If the maxima of $\mathbf{u}_n(t)$ occurred at the same time instant, then we could apply Eq. 21 to get the maximum storey displacements. Although this is

not correct, it can be used to give a conservative upper bound. Therefore

$$\begin{aligned}
 \mathbf{u}_{\max} &< \left| \sum_{n=1}^N \Phi_n y_{n,\max} \right| \\
 &= |\Phi_1 y_{1,\max}| + |\Phi_2 y_{2,\max}| + |\Phi_3 y_{3,\max}| \\
 &= \begin{bmatrix} 0.69 \\ 0.92 \\ 1.0 \end{bmatrix} \times 0.0565 + \begin{bmatrix} 0.95 \\ 0.16 \\ 1.0 \end{bmatrix} \times 0.0018 \\
 &\quad + \begin{bmatrix} 0.38 \\ 0.83 \\ 1.0 \end{bmatrix} \times 0.0002 = \begin{bmatrix} 0.0403 \\ 0.0519 \\ 0.0580 \end{bmatrix} m.
 \end{aligned} \tag{27}$$

The above values clearly overestimate the actual values obtained with response history analysis which are shown in the last row of Table 2. For a more accurate and less conservative estimation of the maximum displacements, we use the SRSS or the CQC rule. A sample calculation of the top storey displacement using the SRSS rule will be

$$\mathbf{u}_{3,o} = \sqrt{0.0565^2 + 0.0018^2 + 0.0002^2} = 0.0565m.$$

For storeys 1 and 2, the maximum displacements are $\mathbf{u}_{1,o} = 0.0389m$ and $\mathbf{u}_{2,o} = 0.0519m$ (Table 2).

The CQC method requires first to calculate β_{nm} and ϵ_{nm} . The values of β_{nm} are shown in Table 3, and Table 4 shows ϵ_{nm} . A sample calculation of β_{12} would give $\beta_{12} = T_2/T_1 = 0.31/0.984 = 0.3123$. ϵ_{12} is obtained as

$$\epsilon_{12} = \frac{8 \times 0.05^2 \times (1 + 0.3123) \times 0.3123^{3/2}}{(1 - 0.3123^2)^2 + 4 \times 0.05^2 \times 0.3123 \times (1 + 0.3123)^2} = 0.005587 \tag{28}$$

Moreover, when $N = 3$, Eq. 24 becomes

$$\begin{aligned}
 r_o^2 &= r_1^2 + r_1^2 + r_2^2 + r_3^2 + (\epsilon_{12} + \epsilon_{21})r_1r_2 + \\
 &\quad + (\epsilon_{23} + \epsilon_{32})r_2r_3 + (\epsilon_{13} + \epsilon_{31})r_1r_3.
 \end{aligned}$$

With the aid of Table 4, the maximum displacement of the third storey becomes

$$\begin{aligned}
 u_{3,o}^2 &= 0.0565^2 + 0.0018^2 + 0.0002^2 \\
 &\quad + 2 \times 0.0056 \times 0.0565 \times 0.0018 \\
 &\quad + 2 \times 0.0592 \times 0.0018 \times 0.0002 \\
 &\quad + 2 \times 0.0026 \times 0.0565 \times 0.0002 \\
 &= 0.0032 + 1.4 \cdot 10^{-6} + 4.2 \cdot 10^{-8} + 5.9 \cdot 10^{-8} \\
 &= 0.0032 u_{3,o} = \sqrt{0.032} = 0.0565
 \end{aligned} \tag{29}$$

Response Spectrum Analysis of Structures Subjected to Seismic Actions, Table 2 Response spectrum modal displacements

	Storey 1	Storey 3	Storey 3
$u_{1,max}$	0.0388	0.0519	0.0565
$u_{2,max}$	0.0017	0.0003	0.0018
$u_{3,max}$	0.0001	0.0001	0.0002
Sum (Eq. 21)	0.0403	0.0519	0.0580
SRSS (Eq. 22)	0.0389	0.0519	0.0565
CQC (Eq. 24)	0.0389	0.0519	0.0565
RHA	0.0333	0.0468	0.0520

Response Spectrum Analysis of Structures Subjected to Seismic Actions, Table 3 β_{nm} values used for the CQC method

Mode	1	2	3
1	1.0000	0.3123	0.2110
2	3.2021	1.0000	0.6757
3	4.7396	1.4799	1.0000

Note that for the problem studied, the correlation terms are very small (less than 10^{-6}), and thus the final prediction is similar to that of the SRSS method. Also, cautiousness is required due to the fact that in the CQC calculation, the r terms maintain their sign (here, everything is positive). This means that the CQC estimates could be greater or smaller than that of SRSS.

Another approach to obtain the modal displacements is through applying the equivalent lateral forces of Eq. 20. For the first mode, the calculation will be

$$\begin{aligned} \mathbf{f}_1 &= \left(\frac{L_1}{M_1} \mathbf{m} \Phi_1 \right) S_a(T_1, \xi) \\ &= \frac{84.22}{72.62} \begin{bmatrix} 40 & 0 & 0 \\ 0 & 40 & 0 \\ 0 & 0 & 20 \end{bmatrix} \begin{bmatrix} 0.6870 \\ 0.9184 \\ 1.0000 \end{bmatrix} 0.202g \\ &= \begin{bmatrix} 63.3358 \\ 84.6704 \\ 46.0954 \end{bmatrix} \text{ kN.} \end{aligned}$$

The equivalent lateral forces are applied on the structure to the yield the modal displacements:

$$\mathbf{u}_1 = \mathbf{k}^{-1} \mathbf{f}_1 = \mathbf{k}^{-1} \begin{bmatrix} 63.3358 \\ 84.6704 \\ 46.0954 \end{bmatrix} = \begin{bmatrix} 0.0388 \\ 0.0519 \\ 0.0565 \end{bmatrix} m$$

Response Spectrum Analysis of Structures Subjected to Seismic Actions, Table 4 ϵ_{nm} values used for the CQC method

Mode	1	2	3
1	1.0000	0.0056	0.0026
2	0.0056	1.0000	0.0592
3	0.0056	1.0000	0.0592

The reader should appreciate that the base shear of the first mode, V_{s1} , is equal to the sum of the entries of $\mathbf{f}_1$, which yields $V_{s1} = 194.10 \text{ kN}$.

Multistorey Buildings with Arbitrary Plan Configuration

Response spectrum analysis has been so far discussed for plane frames. The extension to three-dimensional buildings, although straightforward, requires some further attention. When three-dimensional multistorey buildings are studied, we assume that each storey behaves as a rigid diaphragm with three inplane degrees of freedom, two translational, and one rotational. Therefore, the total number of degrees of freedom is $3N$, where N is the number of storeys. Moreover, the direction of seismic loading is arbitrary but can always be analyzed to two components parallel to the x and y axes and denoted $\ddot{u}_{g,x}$, $\ddot{u}_{g,y}$, respectively. The equation of motion (Eq. 1) is now written:

$$\tilde{\mathbf{m}} \begin{bmatrix} \ddot{\mathbf{u}}_x \\ \ddot{\mathbf{u}}_y \\ \ddot{\mathbf{u}}_\theta \end{bmatrix} + \tilde{\mathbf{c}} \begin{bmatrix} \dot{\mathbf{u}}_x \\ \dot{\mathbf{u}}_y \\ \dot{\mathbf{u}}_\theta \end{bmatrix} + \tilde{\mathbf{k}} \begin{bmatrix} \mathbf{u}_x \\ \mathbf{u}_y \\ \mathbf{u}_\theta \end{bmatrix} = -\tilde{\mathbf{m}} \begin{bmatrix} \ddot{u}_{g,x} \\ \ddot{u}_{g,y} \\ 0 \end{bmatrix} \quad (30)$$

where the mass matrix is

$$\tilde{\mathbf{m}} = \begin{bmatrix} \mathbf{m}_x & 0 & 0 \\ 0 & \mathbf{m}_y & 0 \\ 0 & 0 & \mathbf{I}_O \end{bmatrix} \quad (31)$$

and $\tilde{\mathbf{c}}$ and $\tilde{\mathbf{k}}$ are the corresponding damping and stiffness matrices, respectively. $\mathbf{I}_O$ is a diagonal matrix of order N whose entries are the moments of inertia of every storey about the vertical axis that passes through the center of mass. The modal

vector consists of three subvectors, $\Phi_n = \langle \phi_{xn} \phi_{yn} \phi_{\theta n} \rangle$, and the participation factor L_n^h/M_n is now calculated from the quantities:

$$\begin{aligned} L_{xn}^h &= \phi_{xn}^T \mathbf{m}, \text{ for } \ddot{u}_{gx} \\ L_{yn}^h &= \phi_{yn}^T \mathbf{m}, \text{ for } \ddot{u}_{gy} \end{aligned} \quad (32)$$

and

$$M_n = \Phi_n^T \tilde{\mathbf{m}} \Phi_n \quad (33)$$

K_n and C_n are defined as before using Eq. 4. Equation 32 depends on the component of ground motion considered ($\ddot{u}_{g,x}$ or $\ddot{u}_{g,y}$). For every mode, we solve again Eq. 7 to obtain the modal displacements y_n and the lateral displacements and the torsional rotation of the diaphragm:

$$\mathbf{u}_n = \langle \mathbf{u}_{xn} \mathbf{u}_{yn} \mathbf{u}_{\theta n} \rangle = \langle \phi_{xn} \phi_{yn} \phi_{\theta n} \rangle y_n(t) \quad (34)$$

The modal displacements can be obtained from the design response spectrum using Eq. 17 or performing static analysis of the building applying the equivalent static forces associated with the n^{th} mode peak response (Eq. 20). In this case, assuming $\ddot{u}_g = \ddot{u}_{gx}$, the lateral forces $\mathbf{f}_n$ will be

$$\mathbf{f}_n = \frac{L_{nx}}{M_n} \mathbf{m} \Phi_n S_a(T_n, \xi) \quad (35)$$

and

$$\mathbf{u}_n = \tilde{\mathbf{k}}^{-1} \mathbf{f}_n. \quad (36)$$

Response Spectrum Modal Analysis in Seismic Design Codes

Various design codes and guidelines, including European and US codes, suggest the use of both linear and nonlinear methods of analysis. Nonlinear methods are beyond the scope of this entry, while linear methods are based on response spectrum modal analysis and on a simplified method that considers only the first mode of response. According to the terminology of

Eurocode 8, linear elastic analysis can be performed either with (a) the “lateral force method of analysis” for buildings meeting specific conditions or (b) with the “modal response spectrum analysis,” applicable to all types of buildings and being the reference method for obtaining seismic response estimates.

According to the “lateral force method of analysis,” a lateral load pattern that follows the first mode is applied (e.g., Eq. 35). In Eurocode 8, the method is applicable only if the fundamental period of vibration is less than 2s or $4T_C$, where T_C is the corner period of the design spectrum. A second condition that also has to be satisfied is that the structure should meet the criteria for regularity in plan. Different criteria may be found in other codes. The “modal response spectrum analysis” is recommended for all other cases, with the only exception of buildings with seismic isolation provided by highly nonlinear devices. Practically every design code in the world uses these two methods for linear elastic analysis and recommends criteria to determine if their applicable.

Summary

Response spectrum modal analysis has been presented, discussing that it is a simplified version of modal analysis and appropriate for structural design. The method allows the use of smooth design spectra for the assessment and the design of structures. All concepts discussed are presented in a numerical example, while the adaptation of the method by modern design codes has been conceptually explained.

Cross-References

- [Assessment of Existing Structures Using Response History Analysis](#)
- [Modal Analysis](#)
- [Nonlinear Dynamic Seismic Analysis](#)
- [Time History Seismic Analysis](#)

References

- Bathe KJ (1996) *Finite element procedures*. Prentice Hall, New York
- Chopra AK (2000) *Dynamics of structures: theory and applications to earthquake engineering*, 2nd edn. Prentice Hall, Englewood Cliffs
- Chopra AK (2005) *Earthquake dynamics of structures a primer*, 2nd edn. Earthquake Engineering Research Institute, Oakland
- Clough RW, Penzien J (1975) *Dynamics of structures*. McGraw-Hill, New York, 634 pp
- Der Kiureghian A (1981) A response spectrum method for random vibration analysis of MDF systems. *Earthq Eng Struct Dyn* 9:419–435
- European Committee for Standardisation (2004) Eurocode 8 (EC8) *Design of structures for earthquake resistance – part 1: general rules for buildings*. Brussels
- Newmark NM, Hall WJ (1982) *Earthquake spectra and design*. Earthquake Engineering Research Institute, Berkeley
- Wilson EL, Der Kiureghian A, Bayo EP (1981) Replacement for the SRSS method in seismic analysis. *Earthq Eng Struct Dyn* 9(2):187–192

Response Variability and Reliability of Structures

George Stefanou

Institute of Structural Analysis & Antiseismic Research, School of Civil Engineering, National Technical University of Athens, Athens, Greece
Institute of Structural Analysis & Dynamics of Structures, Department of Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

Synonyms

Failure probability; Fragility; Response statistics; Stochastic dynamics

Introduction

Computational methods, such as the finite element method, are nowadays necessary for the analysis and design of large-scale engineering

systems. The considerable influence of inherent uncertainties on system behavior has also led the scientific community to recognize the importance of a stochastic approach to engineering problems. Engineering experience has shown that uncertainties are involved not only in the loading but also in the material and geometric properties of engineering systems. The rational treatment of these uncertainties cannot be addressed rigorously in the framework of the traditional deterministic approach. Stochastic methods do provide this possibility at the expense of increasing the complexity of the system model and, consequently, of the required computational effort for the solution of the problem. Therefore, the exploitation of the available computational resources and the development of enhanced solution algorithms are of paramount importance in the application of stochastic methods to real-world problems and to their further dissemination to the engineering community.

The problems of response variability and reliability of structures with stochastic properties under dynamic loading are currently the subject of extensive research in the fields of computational stochastic dynamics and earthquake engineering. Both problems deal with the computation of the statistical characteristics of the response (statistical moments, probability of failure), and their solution is time consuming, particularly in the case of large-scale realistic structures. An important practical application of structural response variability and reliability is the estimation of seismic fragility curves used to assess the vulnerability of structures due to earthquakes. Fragility curves provide the probability of exceeding a prescribed level of damage for a wide range of ground motion intensity, usually expressed in terms of peak ground or spectral accelerations.

The existing analytical methods for the evaluation of structural response variability and reliability can only be used in few special cases and are not applicable to realistic engineering problems. Therefore, this chapter will be focused on approximate methods and simulation. The first class of methods is mostly based on the

approximation of the underlying Fokker–Planck–Kolmogorov equation or of the limit state function, while the second comprises the direct Monte Carlo simulation and its variants. In sections “[Response Variability of Stochastic Systems](#)” and “[Reliability of Stochastic Systems](#),” distinction will be made between linear and nonlinear systems as the available techniques depend on structural characteristics. Nonlinear systems are significantly more involved, and their treatment requires the development of specialized approaches. A numerical example involving the computation of the response variability and reliability of a steel frame will be provided in the last section.

Response Variability of Stochastic Systems

Linear Systems

The case of stochastic linear systems will be examined first. The equation of motion of these systems in the time domain t is the following:

$$\mathbf{M}(\boldsymbol{\theta})\ddot{\mathbf{u}}(t, \boldsymbol{\theta}) + \mathbf{C}(\boldsymbol{\theta})\dot{\mathbf{u}}(t, \boldsymbol{\theta}) + \mathbf{K}(\boldsymbol{\theta})\mathbf{u}(t, \boldsymbol{\theta}) = \mathbf{F}(t, \boldsymbol{\theta}) \quad (1)$$

where $\mathbf{M}$, $\mathbf{C}$, $\mathbf{K}$ are the stochastic mass, damping, and stiffness matrices, respectively, and $\mathbf{u}$, $\dot{\mathbf{u}}$, $\ddot{\mathbf{u}}$ denote the stochastic displacement, velocity, and acceleration vectors of the structural system (parameter $\boldsymbol{\theta}$ denotes the randomness of a quantity).

Direct Monte Carlo simulation (MCS) is the simplest method for computing the dynamic response variability of stochastic structural systems. In this method, $NSIM$ samples of the stochastic system matrices are generated using a random number generator, and the equation of motion (Eq. 1) is solved $NSIM$ times, leading to a population (sample) of the response vector. Based on this population, the response variability of the system is calculated using simple relationships of statistics. For example, if u_i is the displacement at the i th d.o.f. of the discretized structure, then the unbiased estimates of the

mean value and variance of the corresponding sample are

$$E(u_i) = \frac{1}{NSIM} \sum_{j=1}^{NSIM} u_i(j) \quad (2)$$

$$\sigma^2(u_i) = \frac{1}{NSIM - 1} \left[\sum_{j=1}^{NSIM} u_i^2(j) - NSIM \cdot E^2(u_i) \right] \quad (3)$$

It is obvious that the accuracy of the estimation depends on the number of samples, and, in particular, the estimate of standard deviation σ is inversely proportional to $\sqrt{NSIM}$. A small number of samples, e.g., $NSIM \approx 50$, permits only a rough approximation of the mean value and variance of the response. With a larger sample size, e.g., $NSIM \approx 500$, it is possible to estimate the cumulative distribution function (CDF) of the response. A number of samples as large as 10^4 is usually needed in dynamical problems in order to compute the probabilistic characteristics of the response with sufficient accuracy. This leads to an excessive computational cost especially in the combined case of large-scale systems and of considerable stochastic dimension. However, the development of robust and efficient solution algorithms in conjunction with the increasing availability of powerful computers and the suitability of the method to parallel processing with ideal efficiency alleviates this limitation to a large extent (Johnson et al. 2003; Stefanou 2009).

The direct MCS described above is the basic version of the method and is often used in the literature as an exact (reference) approach for validating the results of other methods (Adhikari 2011). Several variants of this approach (e.g., importance sampling, subset simulation, line sampling) have been developed in the last twenty years especially for the efficient solution of reliability problems where the calculation of small failure probabilities requires a very large number of samples. These variants will be examined in section “[Reliability of Stochastic Systems](#).”

Recently, an exact nonstatistical method has been proposed for the dynamic analysis of finite

element (FE)-discretized uncertain linear structures in the frequency domain (Falsone and Ferro 2007). This procedure is based on the properties of the structural deformation modes and, in particular, on the number of principal deformation modes of the FE type used for the structural discretization. The method leads to an exact relationship between the response and the random variables representing the structural uncertainties and thus allows obtaining an optimum level of accuracy through a reduced computational effort.

An approximate method for the response variability calculation of dynamical systems with uncertain stiffness and damping ratio can be found in Papadimitriou et al. (1995). This approach is based on complex mode analysis where the variability of each mode is analyzed separately and can efficiently treat a variety of probability distributions assumed for the system parameters. A probability density evolution method (PDEM) has also been developed for the dynamic response analysis of linear stochastic structures (Li and Chen 2004). In this method, a probability density evolution equation (PDEE) is derived according to the principle of preservation of probability. With the state equation expression, the PDEE is further reduced to a one-dimensional partial differential equation from which the instantaneous probability density function (PDF) of the response and its evolution are obtained. Finally, variability response functions have been recently proposed as an alternative to direct MCS for the accurate and efficient computation of the dynamic response of linear structural systems with uncertain Young modulus (Papadopoulos and Kokkinos 2012).

Nonlinear Systems

The general form of the equation of motion for a multi-degree-of-freedom nonlinear system is the following:

$$\mathbf{M}(\boldsymbol{\theta})\ddot{\mathbf{u}}(t, \boldsymbol{\theta}) + \mathbf{f}_D(\mathbf{u}, \dot{\mathbf{u}}, t, \boldsymbol{\theta}) + \mathbf{f}_S(\mathbf{u}, \dot{\mathbf{u}}, t, \boldsymbol{\theta}) = \mathbf{F}(t, \boldsymbol{\theta}) \quad (4)$$

where $\mathbf{u}$, $\dot{\mathbf{u}}$, $\ddot{\mathbf{u}}$ denote the stochastic displacement, velocity, and acceleration vectors of the structural

system; $\mathbf{M}$ is the stochastic mass matrix; and $\mathbf{f}_D(\mathbf{u}, \dot{\mathbf{u}}, t, \boldsymbol{\theta})$, $\mathbf{f}_S(\mathbf{u}, \dot{\mathbf{u}}, t, \boldsymbol{\theta})$ are the damping and restoring force terms, which are usually nonlinear functions of the displacement and velocity of the system ($\boldsymbol{\theta}$ denotes the randomness of a quantity, as in Eq. 1).

In contrast to the linear case, the stochastic analysis of dynamic systems with nonlinear mechanisms involved either in the excitation process or in the mechanical properties poses a major challenge in the field of stochastic dynamics (Manolis and Koliopoulos 2001). This can be explained by the fact that most of the methods developed for the analysis of linear systems are inefficient or inappropriate for the nonlinear case. For example, the analysis of uncertain nonlinear systems is generally not feasible using frequency domain analysis techniques (Iwan and Huang 1996). Major research efforts in this area concentrate on the development of new and the adaption of existing methods to assess the stochastic response. In the past, analytical procedures dominated the developments in nonlinear stochastic dynamics, but in recent years, the emphasis has been placed on numerical methods, sometimes combined with analytical procedures. It is worth noting that the majority of earthquake engineering applications fall within this class of stochastic problems (Desceliers et al. 2004).

Most of the analytical solutions for nonlinear systems are based on the Fokker–Planck–Kolmogorov (FPK) diffusion equation:

$$\frac{\partial p}{\partial t} + \sum_{i=1}^N \frac{\partial}{\partial x_i} (a_i p) - \frac{1}{2!} \sum_{i=1}^N \sum_{j=1}^N \frac{\partial^2}{\partial x_i \partial x_j} (b_{ij} p) + \dots = 0 \quad (5)$$

where p represents the transition probability between states and a_i , b_{ij} represent the drift and diffusion terms depending in general on the state vector $\mathbf{x}$. These require an excitation process with broadband characteristics in order to be modeled by a white noise approximation, leading to Markovian properties. Exact solutions are available for a limited class of problems. Other solution methods, e.g., the FE method or the path integral

method, require extremely large computational efforts for systems of order higher than two. Alternative types of approximate methods, such as the stochastic averaging and statistical linearization techniques, are currently limited to systems with very few degrees of freedom (Schuëller 2006; Kougiumtzoglou and Spanos 2013).

The existing numerical methods for response statistics calculation are mostly based on simulation or on the perturbation approach (Schuëller and Pradlwarter 1999; Muscolino et al. 2003; Liu et al. 1986). Studies can also be found on the statistical equivalent linearization (EQL) method for the response variability and reliability estimation of discrete nonlinear systems (Schuëller and Pradlwarter 1999). It is well known that the main drawback of the perturbation approach is the significant loss of accuracy when the level of uncertainty of the system properties is high. On the other hand, the computational effort required by statistical approaches such as direct MCS for the analysis of large-scale structures is considerable thus making essential the use of efficient solution strategies and parallel processing. In addition, the validity of EQL is questionable in some cases, and the method may even produce misleading results (Bernard 1998). Alternatively, an extension of the PDEM has been developed for the dynamic response analysis of nonlinear stochastic structures (Li and Chen 2006).

Reliability of Stochastic Systems

The reliability analysis of structures under dynamic loading consists in the computation of the probability that a specific response quantity will not exceed a critical threshold at various time instants. The characteristics of the loading and material properties can be described in most cases with sufficient accuracy by utilizing the theory of stochastic processes and fields along with experimental data information for estimating the appropriate parameters. While for wind loading, and also in some cases for wave loading, linear structural models suffice, nonlinear models are definitely required for earthquake loading as explained in subsection “Nonlinear Systems.”

A recent comprehensive review on reliability assessment in structural dynamics can be found in Goller et al. (2013).

Linear Systems

The direct MCS examined in section “Response Variability of Stochastic Systems” becomes inefficient for the solution of reliability problems where a large number of low-probability realizations in the failure domain must be produced. In order to alleviate this problem without deteriorating the accuracy of the solution, numerous variants of this approach have been developed. An important class of improved MCS is variance reduction techniques where the generation of samples of the basic random variables is controlled in an efficient way.

The most prominent representative of this class of methods is importance sampling (IS), in which the generation of samples is controlled by a sampling distribution concentrated in the “important” (low-probability) region of the failure domain. The main challenge in the application of IS to physical problems is the determination of the sampling distribution, which depends on the specific system at hand and on the failure domain (Schuëller 2006). The optimal choice of the sampling distribution (for which the variance of the estimator of the probability of failure p_F vanishes) is practically infeasible since an a priori knowledge of p_F is required for this purpose. Thus, several techniques based on kernel density estimators or design points have been proposed in order to produce a sampling distribution characterized by a reduced variance of the estimator of p_F :

$$\hat{p}_F = \frac{1}{N} \sum_{i=1}^N \frac{1_F(\boldsymbol{\theta}^{(i)}) h(\boldsymbol{\theta}^{(i)})}{f(\boldsymbol{\theta}^{(i)})} \quad (6)$$

In this equation, N is the number of samples, 1_F denotes the indicator function of the failure domain, h is the joint probability distribution of the basic random variables, and the samples $\{\boldsymbol{\theta}^{(i)}\}_{i=1}^N$ are generated according to the sampling distribution f .

IS is efficient for the reliability assessment of static linear and nonlinear systems characterized by a small number of basic random variables. However, for the dynamic reliability analysis of large nonlinear systems in high stochastic dimensions, the computational effort needed to construct a suitable sampling distribution may exceed the effort required by the direct MCS (Schuëller 2006).

Apart from sampling-based algorithms, approximate methods have also been developed for the evaluation of the reliability of linear dynamical systems. Most of these methods are based on the solution of the FPK diffusion equation by means of the FE method or the path integral approach with the goal of determining the response PDF which is used for the estimation of the associated failure probability. Although these diffusion process-based methods provide accurate results, their application to large-scale structural systems is limited. Therefore, some recent research efforts focus on the enhancement of the PDEM (see section “Response Variability of Stochastic Systems”) for large-scale structural reliability estimation. The first-order reliability method (FORM) is also frequently applied in the field of dynamics. The basic idea of this approach is the approximation of the limit state function by a hyperplane through the so-called design point, i.e., the point on the limit state function with the minimum distance to the origin. However, due to the underlying assumption of a single failure region with a linear performance function, the accuracy of FORM in the reliability assessment of dynamical systems is often not satisfactory (Goller et al. 2013). Finally, an efficient IS technique is combined in Jensen and Valdebenito (2007) with approximate representations of performance functions for the reliability analysis of linear dynamical systems.

Nonlinear Systems

The reliability analysis of nonlinear systems is defined similarly to the linear case as the computation of first excursion probabilities over specified thresholds. While in linear problems these thresholds are specified in terms of

displacements, the reliability thresholds of nonlinear responses might be more appropriately specified in terms of damage caused by excessive strains. Since most types of materials are characterized by a nonlinear stress–strain relationship and deteriorating strength, nonlinear analysis is the correct choice to assess the structural reliability under severe loading conditions, such as earthquakes.

Direct MCS is the most general method for the reliability analysis of nonlinear systems, where the accuracy and efficiency are independent from the number of random quantities and the structural model type, and depends only on the number of generated samples (see section “Response Variability of Stochastic Systems”). However, a very large number of samples is required for the accurate computation of small failure probabilities (in the range of $10^{-7} < p_F < 10^{-4}$) thus making direct MCS inefficient in practice. In order to overcome the inefficiency of direct MCS in calculating small failure probabilities, a novel approach called subset simulation (SS) has been proposed by Au and Beck (2001). SS is a powerful tool, simple to implement, and capable of solving a broad range of reliability problems, e.g., Au and Beck (2003). The basic idea of SS is to express the failure probability p_F as a product of larger conditional probabilities by introducing a decreasing sequence of intermediate failure events (subsets): $\{F_i\}_{i=1}^m$ such that $F_m = F$ and $F_1 \supset F_2 \dots \supset F_m = F$:

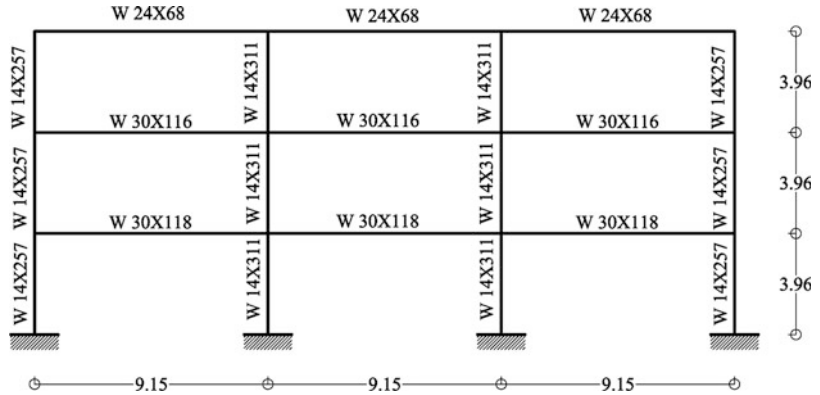
$$p_F = P(F_m) = P(F_1) \prod_{i=1}^{m-1} P(F_{i+1}/F_i) \quad (7)$$

With a proper choice of the intermediate events, the conditional failure probabilities can be made sufficiently large. Therefore, the original problem of computing a small failure probability is reduced to calculating a sequence of larger conditional probabilities, which can be efficiently estimated by means of direct MCS with a small number of samples.

Another recently developed technique which permits the efficient treatment of high-dimensional reliability problems is line sampling (LS) (Koutsourelakis et al. 2004). This technique

Response Variability and Reliability of Structures,

Fig. 1 A three-story steel frame (With kind permission from Springer Science + Business Media: (Stefanou and Fragiadakis 2011), © Springer Science + Business Media B.V. 2011)



takes advantage of an implicitly available performance function (data points on the limit state surface) obtained directly from FE analyses. As already mentioned in the case of IS, the optimal choice of the sampling distribution is practically infeasible. However, something quite close to optimal sampling can be achieved by using LS and computing an important direction α which points toward the failure domain nearest to the origin. Neither the vector α is required to point exactly to the design point, nor are any assumptions made with respect to the shape of the limit state surface. In contrast to IS, where an inappropriate choice of the sampling distribution leads to worse estimates compared to direct MCS, LS performs at least as well as direct MCS even in the worst possible case where the direction α is selected orthogonal to the optimal direction (Koutsourelakis et al. 2004). In comparison to IS, LS requires far less performance evaluations (FE analyses) to obtain a similar accuracy. The advantages of LS become more pronounced in high stochastic dimensions as it is shown in Schuëller et al. (2004) where a comparison between different approaches for reliability estimation is presented.

The PDEM can also be efficiently applied to compute the probability density of the response of dynamically excited nonlinear structures (Li et al. 2012). It is finally worth noting that the theory of non-Gaussian translation processes has been applied directly to the reliability analysis of dynamic systems under limited information. This method delivers accurate results for the case of linear and nonlinear dynamic systems

assuming stationary output but can be easily extended to a special class of nonstationary, non-ergodic output (Field and Grigoriu 2009).

A Numerical Example

In this section, the response variability and reliability of a three-story steel moment-resisting frame shown in Fig. 1 are computed using direct MCS (Stefanou and Fragiadakis 2011). The frame has been designed for a Los Angeles site, following the 1997 NEHRP (National Earthquake Hazard Reduction Program) provisions in the framework of the SAC/FEMA program (SAC 2000). The dynamic response of the building is dominated by the fundamental mode which has a period value equal to $T_1 = 1.02$ s when the mean value of the modulus of elasticity is used. Response history analyses are performed using a force-based, beam-column fiber element implemented on a general purpose FE program (Taylor 2000). Geometric nonlinearities are not considered in the analysis. Rayleigh damping is used to obtain a damping ratio of 2 % for the first and the fourth mode. The material law is considered to be bilinear with pure kinematic hardening (with unloading and reloading branches parallel to the elastic stiffness). The applied gravity loading is 32.22 kN/m for the first two stories and 28.76 kN/m for the top story. These values are also used to obtain the nodal masses resulting to a lumped mass matrix.

Three sets of five strong ground motion records are used as input to the nonlinear dynamic procedure (Table 1). The three sets correspond to

Response Variability and Reliability of Structures, Table 1 Details of the ground motion records

ID (level)	Earthquake	Station	ϕ^{oa}	Mw ^b	R ^c	Soil ^d	PGA ^e
1 (1/1)	Imperial Valley, 1979	Plaster City	045	6.5	31.7	C,D	0.042
2 (2/1)	Imperial Valley, 1979	Plaster City	135	6.5	31.7	C,D	0.057
3 (3/1)	Imperial Valley, 1979	Westmoreland Fire Station	180	6.5	15.1	C,D	0.11
4 (4/1)	Imperial Valley, 1979	Westmoreland Fire Station	090	6.5	15.1	C,D	0.074
5 (5/1)	Imperial Valley, 1979	Compuertas	285	6.5	32.6	C,D	0.147
6 (1/2)	Northridge, 1994	LA, Baldwin Hills	090	6.7	31.3	B,B	0.239
7 (2/2)	Imperial Valley, 1979	Plaster City	090	6.5	31.7	C,D	0.057
8 (3/2)	Loma Prieta, 1989	Sunnyvale Colton Ave	270	6.9	28.8	C,D	0.207
9 (4/2)	Superstition Hills, 1987	Wildlife Liquefaction Array	090	6.7	24.4	C,D	0.18
10 (5/2)	Loma Prieta, 1989	Sunnyvale Colton Ave	360	6.9	28.8	C,D	0.209
11 (1/3)	Superstition Hills, 1987	Wildlife Liquefaction Array	360	6.7	24.4	C,D	0.2
12 (2/3)	Northridge, 1994	LA, Hollywood Storage FF	360	6.7	25.5	C,D	0.358
13 (3/3)	Loma Prieta, 1989	Hollister South & Pine	000	6.9	28.8	-,D	0.371
14 (4/3)	Loma Prieta, 1989	WAHO	000	6.9	16.9	-,D	0.370
15 (5/3)	Loma Prieta, 1989	WAHO	090	6.9	16.9	-,D	0.638

^aComponent^bMoment magnitude^cClosest distance to fault rupture^dUSGS, Geomatrix soil class^eIn units of g

three levels of increasing hazard: low, medium, and high. The chosen records differ in terms of amplitude, frequency content, duration, etc., and therefore, this variability is expected to be transferred to the statistics of the analysis, producing significant record-to-record variability. The 15 natural records are used as input to the analysis in order to compute the mean response quantities, the dispersion, and the reliability of the frame for the three intensity levels.

The spatial variability in Young modulus and yield stress of the frame is described by two uncorrelated 1D homogeneous non-Gaussian translation stochastic fields with zero mean and coefficient of variation (COV) equal to 0.10. A slightly skewed shifted lognormal distribution defined in the range $[-1, +\infty]$ is assumed for the two stochastic fields. E and σ_y are simultaneously varying in all the cases examined. The spectral density function of the underlying Gaussian field is given by

$$S_{gg}(\kappa) = \frac{\sigma_g^2 b}{2\sqrt{\pi}} \exp\left(-\frac{b^2 \kappa^2}{4}\right) \quad (8)$$

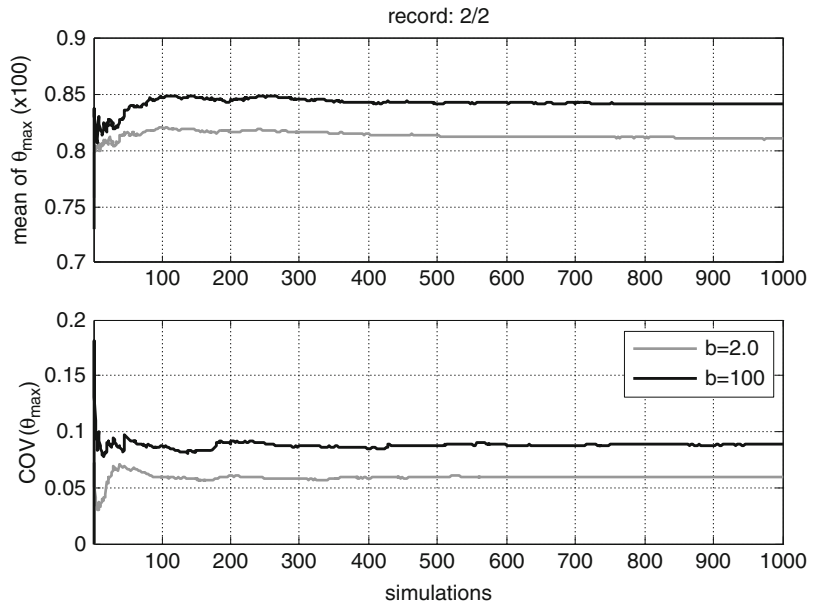
where σ_g denotes the standard deviation of the stochastic field and b is a parameter that

influences the shape of the spectrum and is proportional to the correlation length of the stochastic field.

The representative response quantity whose statistics are monitored is the maximum interstory drift $\theta_{\max}$. This parameter is a well-known engineering demand parameter that captures the seismic demand and its distributions along the height of the structure. The response statistics are calculated using 1000 Monte Carlo simulations. This number of simulations represents sufficiently well the prescribed first two moments of the response, while as shown in Fig. 2, statistical convergence is practically achieved after 400 simulations. The same trend was observed for all ground motions considered. The sensitivity of $\theta_{\max}$ with respect to the scale of correlation of the stochastic fields, quantified with the aid of the correlation length parameter b , is examined for the ground motions of the three sets. For this purpose, several sets of sample functions of E and σ_y are generated, each for a different value of parameter b . Six representative values of b varying from weak to strong correlation are considered ($b = 0.2, 1.0, 2.0, 10, 20$ and 100).

Response Variability and Reliability of Structures,

Fig. 2 Statistical convergence of mean and COV of maximum interstory drift $\theta_{\max}$ – lognormal distribution of E , σ_y with COV = 0.1 (With kind permission from Springer Science + Business Media: (Stefanou and Fragiadakis 2011), © Springer Science + Business Media B.V. 2011)



Based on Fig. 3, an important observation can be made regarding the variability of $\theta_{\max}$. In contrast to the static case where the displacement variability shows always the same trend, starting from small values for small correlation lengths corresponding to white noise stochastic fields up to large values for large correlation lengths, the COV of $\theta_{\max}$ varies significantly not only with the correlation length b but also in different ways among the records of the same intensity level (Fig. 3b). In some cases, the effect of b becomes negligible, and then the record-to-record variability is predominant (e.g., records 1/1, 5/1, and 3/3). In addition, a large magnification of uncertainty is observed in some cases, which is more pronounced for records 1/1, 4/1, 4/2, and 5/2, where the response COV tends to values that are 1.4–1.8 times greater than the corresponding input COV ($=0.1$). In contrast, the mean value of drift, although presenting an important record-to-record variability, is practically not affected by the correlation length parameter b (see Fig. 3a).

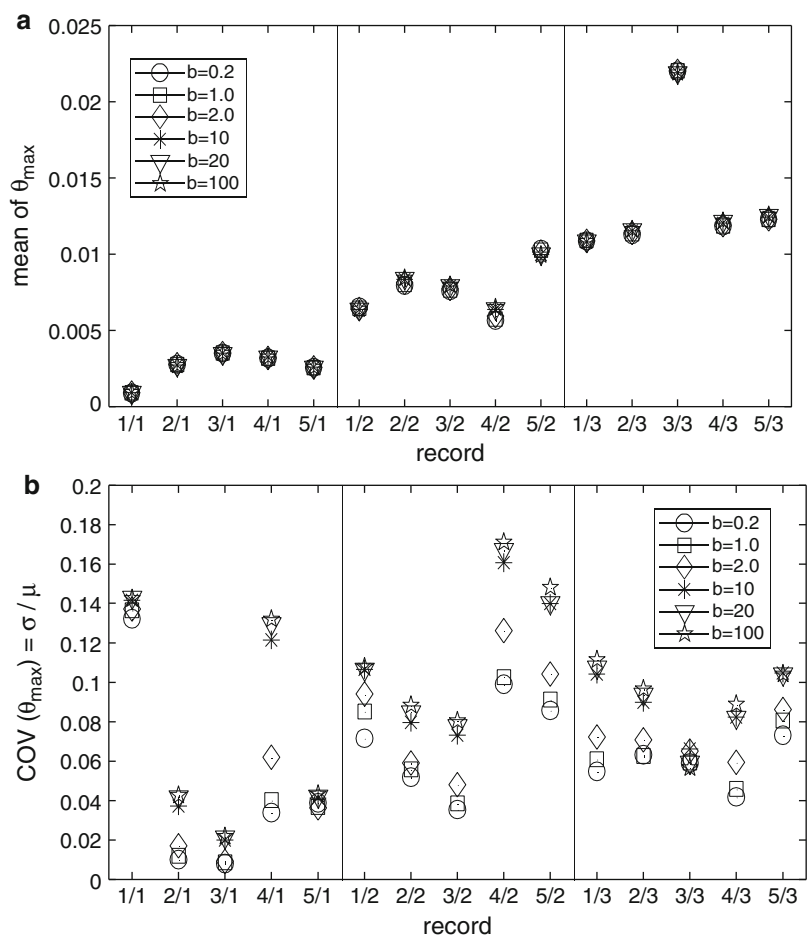
The skewness of $\theta_{\max}$ has been also calculated and is depicted in Fig. 4 where a significant influence of the correlation length parameter b and an important record-to-record variability can be observed. The values of skewness obtained in

each record and seismic intensity are substantially different from the skewness of the lognormal distribution describing the material properties (which is equal to 0.30) thus underlying the strong nonlinearity of the problem that causes a significant change in the statistical moments (and distribution) of the response.

The reliability of the frame can finally be calculated. Fig. 5 shows the CDF of $\theta_{\max}$ of record 4/2 for two different values of $b = 1, 100$. If the reliability of the frame is defined as the maximum interstory drift not exceeding a threshold, e.g., 6.5×10^{-3} , the reliability can be obtained from Fig. 5 for both cases of b as 0.915 and 0.555, respectively. It is worth noting that the reliability is substantially smaller in the second case. Fig. 6 shows the PDF of $\theta_{\max}$ of record 4/2 for the same two values of b . Simultaneously shown are the normal and lognormal distributions with a mean and standard deviation identical to those of the computed PDF and the extreme value distribution with the same mean as that of the computed PDF. It can be observed that these widely adopted probability distributions are quite different from the real PDF of the response, which clearly has a bimodal form especially in the case of small correlation length.

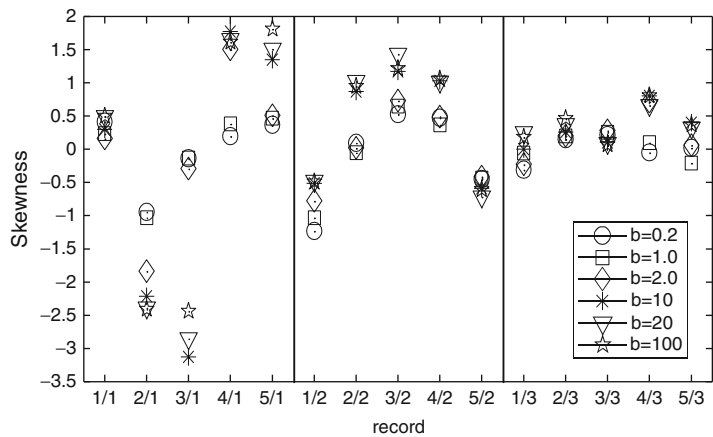
Response Variability and Reliability of Structures,

Fig. 3 (a) Mean and (b) $COV(\theta_{max})$ for different values of correlation length parameter b and the 15 natural ground motions of Table 1 (With kind permission from Springer Science + Business Media: (Stefanou and Fragiadakis 2011), © Springer Science + Business Media B.V. 2011)



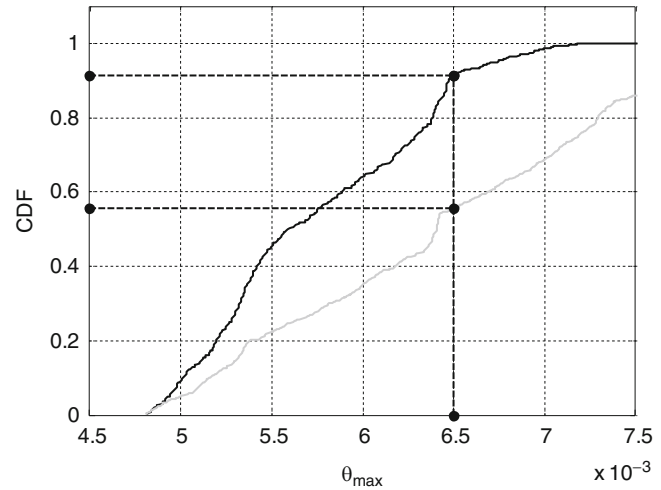
Response Variability and Reliability of Structures,

Fig. 4 Skewness of θ_{max} for different values of correlation length parameter b and the 15 natural ground motions of Table 1 (With kind permission from Springer Science + Business Media: (Stefanou and Fragiadakis 2011), © Springer Science + Business Media B.V. 2011)



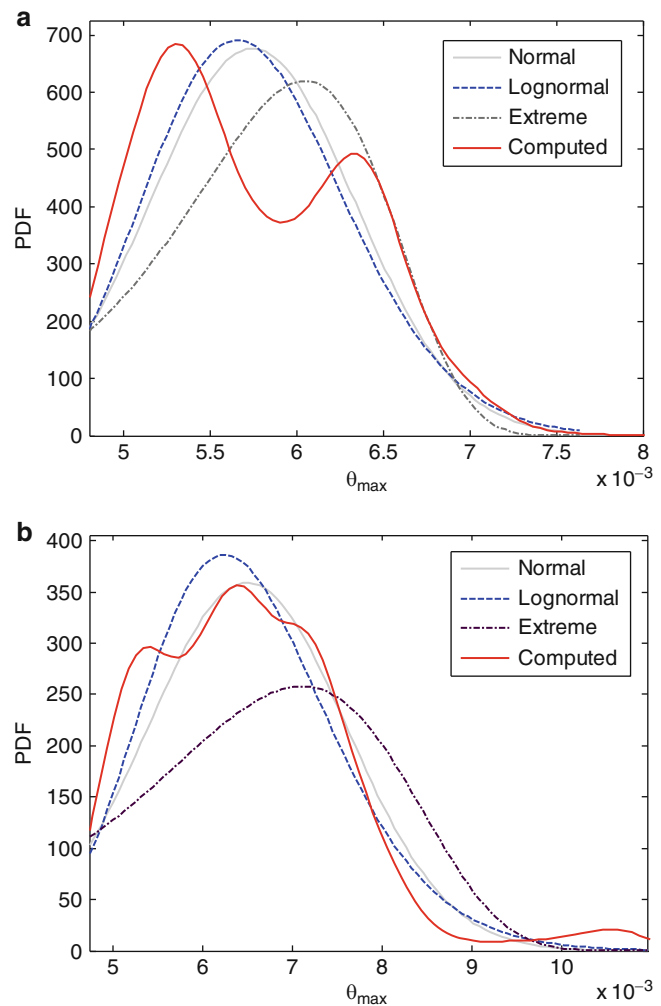
Response Variability and Reliability of Structures,

Fig. 5 CDF of $\theta_{\max}$ of record 4/2 for correlation length parameter $b = 1$ (black line) and $b = 100$ (gray line) (With kind permission from Springer Science + Business Media: (Stefanou and Fragiadakis 2011), © Springer Science + Business Media B.V. 2011)



Response Variability and Reliability of Structures,

Fig. 6 PDF of $\theta_{\max}$ of record 4/2 for correlation length parameter (a) $b = 1$ and (b) $b = 100$ (With kind permission from Springer Science + Business Media: (Stefanou and Fragiadakis 2011), © Springer Science + Business Media B.V. 2011)



Summary

The problems of response variability and reliability of structures with stochastic properties under dynamic loading are currently the subject of extensive research in the fields of computational stochastic dynamics and earthquake engineering. Both problems deal with the computation of the statistical characteristics of the response and have important practical applications, such as the estimation of seismic fragility curves which are used to assess the vulnerability of structures due to earthquakes. As the existing analytical methods for the evaluation of structural response variability and reliability can only be used in few special cases, this chapter is mainly focused on approximate methods and simulation. A numerical example involving a steel frame is also provided to illustrate the presented theoretical concepts.

Cross-References

- [Probability Density Evolution Method in Stochastic Dynamics](#)
- [Reliability Estimation and Analysis](#)
- [Stochastic Analysis of Linear Systems](#)
- [Stochastic Finite Elements](#)
- [Subset Simulation Method for Rare Event Estimation: An Introduction](#)

References

- Adhikari S (2011) A reduced spectral function approach for the stochastic finite element analysis. *Comput Methods Appl Mech Eng* 200:1804–1821
- Au SK, Beck JL (2001) Estimation of small failure probabilities in high dimensions by subset simulation. *Probab Eng Mech* 16:263–277
- Au SK, Beck JL (2003) Subset simulation and its application to seismic risk based on dynamic analysis. *J Eng Mech (ASCE)* 129:901–917
- Bernard P (1998) Stochastic linearization: what is available and what is not. *Comput & Struct* 67:9–18
- Desceliers C, Soize C, Cambier S (2004) Non parametric-parametric model for random uncertainties in non-linear structural dynamics: application to earthquake engineering. *Earthq Eng & Struct Dyn* 33:315–327
- Falsone G, Ferro G (2007) An exact solution for the static and dynamic analysis of FE discretized uncertain structures. *Comput Methods Appl Mech Eng* 196:2390–2400
- Field RV Jr, Grigoriu M (2009) Reliability of dynamic systems under limited information. *Probab Eng Mech* 24:16–26
- Goller B, Pradlwarter HJ, Schuëller GI (2013) Reliability assessment in structural dynamics. *J Sound Vib* 332:2488–2499
- Iwan WD, Huang CT (1996) On the dynamic response of nonlinear systems with parameter uncertainties. *Internat J Non-Linear Mech* 31:631–645
- Jensen HA, Valdebenito MA (2007) Reliability analysis of linear dynamical systems using approximate representations of performance functions. *Struct Saf* 29:222–237
- Johnson EA, Proppe C, Spencer BF Jr, Bergman LA, Székely GS, Schuëller GI (2003) Parallel processing in computational stochastic dynamics. *Probab Eng Mech* 18:37–60
- Kougioumtzoglou IA, Spanos PD (2013) Response and first-passage statistics of nonlinear oscillators via a numerical path integral approach. *J Eng Mech* 139:1207–1217
- Koutsourelakis PS, Pradlwarter HJ, Schuëller GI (2004) Reliability of structures in high dimensions, part I: algorithms and applications. *Probab Eng Mech* 19:409–417
- Li J, Chen JB (2004) Probability density evolution method for dynamic response analysis of structures with uncertain parameters. *Comput Mech* 34:400–409
- Li J, Chen JB (2006) The probability density evolution method for dynamic response analysis of nonlinear stochastic structures. *Int J Numer Methods Eng* 65:882–903
- Li J, Chen JB, Sun W, Peng YB (2012) Advances of probability density evolution method for nonlinear stochastic systems. *Probab Eng Mech* 28:132–142
- Liu WK, Belytschko T, Mani A (1986) Probability finite elements for nonlinear structural dynamics. *Comput Methods Appl Mech Eng* 56:61–81
- Manolis GD, Koliopoulos PK (2001) Stochastic structural dynamics in earthquake engineering. WIT Press, Boston
- Muscolino G, Ricciardi G, Cacciola P (2003) Monte Carlo simulation in the stochastic analysis of nonlinear systems under external stationary Poisson white noise input. *Int J Non-Linear Mech* 38:1269–1283
- Papadimitriou C, Katafygiotis LS, Beck JL (1995) Approximate analysis of response variability of uncertain linear systems. *Probab Eng Mech* 10:251–264
- Papadopoulos V, Kokkinos O (2012) Variability response functions for stochastic systems under dynamic excitations. *Probab Eng Mech* 28:176–184
- SAC (2000) State of the art report on system performance of steel moment frames subjected to earthquake ground shaking. FEMA-355C. Federal Emergency Management Agency, Washington, DC
- Schuëller GI (2006) Developments in stochastic structural mechanics. *Arch Appl Mech* 75:755–773

- Schuëller GI, Pradlwarter HJ (1999) On the stochastic response of nonlinear FE models. *Arch Appl Mech* 69:765–784
- Schuëller GI, Pradlwarter HJ, Koutsourelakis PS (2004) A critical appraisal of reliability estimation procedures for high dimensions. *Probab Eng Mech* 19:463–474
- Stefanou G (2009) The stochastic finite element method: past, present and future. *Comput Methods Appl Mech Eng* 198:1031–1051
- Stefanou G, Fragiadakis M (2011) Nonlinear dynamic response variability and reliability of frames with stochastic non-Gaussian parameters. In: Papadrakakis M et al (eds) *Computational methods in stochastic dynamics*. Series: “Computational Methods in Applied Sciences”, vol 22. Springer, New York, pp 171–185
- Taylor RL (2000) FEAP: a finite element analysis program. User manual, version 7.3. Department of Civil and Environmental Engineering, University of California at Berkeley, Berkeley. <http://www.ce.berkeley.edu/~rlt/feap/>

Response-Spectrum-Compatible Ground Motion Processes

Pierfrancesco Cacciola and Laura D’Amico
School of Environment and Technology,
University of Brighton, Brighton, UK

Synonyms

Artificial accelerograms; Earthquake modelling; Gaussian process; Nonstationary earthquakes; Simulation; Stochastic; Spectrum compatible

Introduction

Several problems of engineering interest require the use of ground motion time histories to analyze and design relevant structures. Recorded and artificial time histories can be both used for this purpose. The use of recorded time histories presents certainly some benefits but they might provide misleading information if used without appropriate consideration. As also it is rarely possible to collect a sufficient number of records in a given location, the use of artificial time histories presents a valuable strategy for the seismic

analysis and design of structures. The use of artificial earthquakes is allowed by international seismic codes (see e.g., Eurocode 8 2010). On the other hand, seismic codes do not provide the method to follow in order to generate artificial accelerograms, but provide only general conditions to be fulfilled in the professional practice. Specifically, seismic codes allow practitioners to employ artificial accelerograms in the seismic assessment and design of structures if the mean response spectrum calculated from each individual artificial time history matches the target response spectrum within a prefixed tolerance over a fixed range of frequencies. The conditions imposed by the seismic codes lead clearly to an inverse problem whose solution is not single. Various approaches can be used to address this problem, and in the last four decades, a plethora of methods has been proposed. Ground motion arising from seismic waves is a phenomenon that generally cannot be described in a deterministic fashion, being in its own nature randomly variable in space and time. The seismic waves resulting from the occurrence of an earthquake are affected by several factors (i.e., source patterns, path, site effects, etc.), as a consequence only a probabilistic approach provides a rigorous representation of the resulting action. The definition of which hypothesis and methods address more realistically the description of the ground motion phenomenon is still an open discussion in the scientific community.

Due to the random nature of the seismic action, the ground motion acceleration $\ddot{u}_g(t)$ recorded at a given location can be seen as a sample of a zero-mean stochastic process. In this regard a number of stochastic models have been proposed in literature and mainly categorized as Gaussian or non-Gaussian models (see Shinozuka and Deodatis 1988). Due to their relative simplicity, the Gaussian models are the most used. It is well known that a zero-mean Gaussian stochastic process is fully defined by the knowledge of the correlation function $R(t_1, t_2)$ or by the generally known power-spectral density function (PSD) $S(\omega, t)$. The latter is commonly preferred due to its physical interpretation related to the energy of the earthquake process

that is highly relevant for design applications. Moreover the one-sided (called also unilateral) power-spectral density function $G(\omega, t)$ is extensively used in literature, and according to the theory of the evolutionary spectra (Priestley 1981), it can be written as follows:

$$\begin{aligned} G(\omega, t) &= 0 & \omega < 0, \\ G(\omega, t) &= 2S(\omega, t) = |a(\omega, t)|^2 G(\omega) & \omega \geq 0; \end{aligned} \quad (1)$$

where $a(\omega, t)$ is the frequency dependent modulating function while $G(\omega)$ is the one-sided power-spectral density function of the stationary counterpart of $\ddot{u}_g(t)$. The processes defined by the evolutionary power-spectral density (EPSD) function in Eq. 1 enable to feature the ground motion acceleration by retaining both amplitude and frequency variation with time, generally these processes are known as fully nonstationary (or non-separable). In the case in which only the amplitude of the process varies with respect to time (i.e., $a(\omega, t) = a(t)$), the process is generally known as quasi-stationary (or separable or also uniformly modulated). Accordingly, Eq. 1 for quasi-stationary processes modifies as follows:

$$\begin{aligned} G(\omega, t) &= 0, & \omega < 0; \\ G(\omega, t) &= 2S(\omega, t) = |a(t)|^2 G(\omega), & \omega \geq 0. \end{aligned} \quad (2)$$

It is observed that the adoption of the time-amplitude nonstationarity through the modulating functions implies an empirically based decision-making process. Several authors proposed modulating functions generally calibrated on different arrays of accelerograms. Among the proposed modulating functions, probably the most diffused are that one proposed by Shinozuka and Sato (1967):

$$a(t) = \delta[\exp(-\beta_1 t) - \exp(-\beta_2 t)]; \beta_1, \beta_2 > 0, \beta_1 < \beta_2 \quad (3)$$

with

$$\delta = \beta_1/(\beta_2 - \beta_1) \exp[\beta_2/(\beta_2 - \beta_1) \ln(\beta_2/\beta_1)] \quad (4)$$

and the following one proposed by Iwan and Hou (1989):

$$a(t) = \delta t^m \exp(-\beta_3 t); \quad m, \beta_3 > 0 \quad (5)$$

$\delta = (\beta_3/m)^m \exp(m)$ that extended the analytical formula proposed by Hsu and Bernard (1978) defined for $m = 1$. Furthermore Jennings et al. (1969) proposed a piecewise modulating function given by the following formula:

$$a(t) = \begin{cases} \left(\frac{t}{t_1}\right)^2 & t < t_1 \\ 1 & t_1 \leq t \leq t_2 \\ \exp[-\beta_4(t - t_2)] & t > t_2 \end{cases} \quad (6)$$

extensively used in the generation of artificial earthquakes because it possesses a stationary time window of duration $T_s = t_2 - t_1$.

Finally, in the particular case in which the modulating function presented in Eq. 2 figures as $a(t) = 1$, the process is called stationary. Note that in this case the power-spectral density function depends only on the circular frequency and not on the time.

Once the power-spectral density function $G(\omega, t)$ is defined, the r th sample of ground acceleration process can be generated via the superposition of N harmonics as follows (e.g., Shinozuka and Deodatis 1988):

$$\ddot{u}_g^{(r)}(t) = \sum_{i=1}^N \sqrt{2G(i\Delta\omega, t)\Delta\omega} \cos(i\Delta\omega t + \varphi_i^{(r)}) \quad (7)$$

where $\Delta\omega$ is the size interval of the discretized frequency domain, $G(i\Delta\omega, t)$ is the power-spectral density function, and $\varphi_i^{(r)}$ are independent random phases uniformly distributed over the interval $[0, 2\pi)$.

The mean of the simulated response spectra obtained from the artificial accelerograms generated by Eq. 7 has to be compared to the response spectra defined by the seismic codes. If the recommended criteria are satisfied, the time histories can be used for design purposes (i.e., these are spectrum compatible).

Clearly, the proper definition of the power-spectral density function is the crucial step to address the stochastic modeling of the seismic action. In the context of defining spectrum-compatible ground motion processes, the issue consists in determining which is the inverse relationship between the power-spectral density function and the target response spectrum. It has to be emphasized that the evaluation of the response-spectrum-compatible power-spectral density function possesses the twofold aspects: (i) allowing the generation of artificial accelerograms through Eq. 7 and (ii) allowing the direct stochastic analysis.

The following sections are devoted to the review of the most relevant models of ground motion acceleration processes aimed to satisfy the code requirements. Namely, the stationary, quasi-stationary, and fully nonstationary as well as spatially variable stochastic spectrum-compatible ground motion will be discussed.

Stationary/Quasi-stationary Models

The most dated models of ground motion process rely on the simplistic hypothesis that only the energy of the ground motion varies with respect to time. According to this hypothesis the quasi-stationary models can reliably represent the ground motion process. Due to the separable feature of the power-spectral density function, the study of the quasi-stationary processes is generally addressed modeling independently the stationary counterpart from the modulating function (see e.g., Eq. 2). This approach is clearly approximated, but it has been shown of being able to provide accurate results. In this framework several procedures have been proposed in literature in order to determine spectrum-compatible power-spectral density functions. Herein the models proposed by Vanmarcke and Gasparini (1977), Kaul (1978), Sundararajan (1980), Pfaffinger (1983), Preumont (1980); Der Kiureghian and Neuenhofer (1992), Park (1995), and Cacciola et al. (2004) are addressed. The list is not fully comprehensive and the readers can refer to the review papers from Ahmadi (1979) and Cacciola (2011) for further references.

Problem Position

In order to address the problem of simulating artificial accelerograms compatible to a given response spectrum, firstly it is necessary to determine the generally known response-spectrum-compatible power-spectral density function. Under the hypothesis of zero-mean Gaussian stationary ground motion process, Vanmarcke and Gasparini (1977) pointed out the fundamental relationship between the target response spectrum and the power-spectral density function of the ground motion through the generally known “first passage problem,” that is,

$$RSA(\omega_0, \zeta_0) = \omega_0^2 \eta_U(T_s; p) \sigma_U \quad (8)$$

where $RSA(\omega_0, \zeta_0)$ is the pseudo-acceleration response spectrum for a given damping ratio ζ_0 and natural circular frequency ω_0 , η_U is the dimensionless peak factor, T_s is the time observing window, p is the not-exceeding probability, and σ_U is the standard deviation of the displacement response, which corresponds to the square root of the zero-order stationary spectral moment $\lambda_{0,U}$. The latter is in turn a function of the power-spectral density function of the ground motion through the following general formula:

$$\lambda_{i,U}(t) = \int_0^\infty \omega^i |H(\omega)|^2 G(\omega) d\omega \quad i = 0, 1, 2 \quad (9)$$

where $|H(\omega)|^2 = \left((\omega_0^2 - \omega^2)^2 + 4\zeta_0^2 \omega_0^2 \omega^2 \right)^{-1}$ is the energy transfer function of the single degree of freedom system of natural frequency ω_0 and damping ratio ζ_0 and $G(\omega)$ is the stationary response-spectrum-compatible power-spectral density function to be determined.

In the last four decades, Eq. 8 has been used by several authors to define the spectrum-compatible power-spectral density function. The methods proposed in literature mainly differentiate one from another for the hypothesis adopted to define the peak factor and for the approximations involved in the evaluation of the response spectral moments.

State of the Art for the Available Solutions

The authors who firstly introduced the implicit relationship for the evaluation of the response-spectrum-compatible power-spectral density function were Vanmarcke and Gasparini (1977); the formula proposed is nowadays still used by researchers and practitioners. To reach this objective Vanmarcke and Gasparini (1977) proposed an approximate expression of the standard deviation of the response displacement:

$$\sigma_U \simeq \frac{1}{\omega_0^2} \left\{ \left[G(\omega_0) \omega_0 \left(\frac{\pi}{4\zeta_0} - 1 \right) \right] + \int_0^{\omega_0} G(\omega) d\omega \right\}^{1/2} \quad (10)$$

Equation 10 is determined under the hypothesis of small damping ratio ζ_0 and approximating the energy transfer function with its value $|H(0)|^2 = 1/\omega_0^4$ in the range of frequencies between 0 and ω_0 . Furthermore they adopted the peak factor expression provided under the hypothesis of independent outcrossing of given barriers (see, e.g., Vanmarcke 1976), through its upper bound value (relatively to damping higher than 10 %), that is,

$$\eta_{U, \text{upp}} = \sqrt{2 \ln \left(\frac{2N_U}{-\ln p} \right)} \quad (11)$$

and for a lower bound valid for undamped systems

$$\eta_{U, \text{low}} = 1 + 0.25 \ln \left(\frac{N_U}{-\ln p} \right) \quad (12)$$

In Eqs. 11 and 12, p is the value of the exceeding probability chosen equal to 0.5 determining the median value of the peak factor; it is noted that the assumption of $p = 0.5$ introduces a further approximation that the median coincides with the mean value. N_U is the number of outcrossings of the given barrier over the fixed time window T_s assumed equal to the strong motion phase of the ground motion process, given by

$$N_U = v_U T_s \quad (13)$$

where v_U is the mean outcrossing rate of the zero level, that is,

$$v_U = \frac{1}{2\pi} \sqrt{\frac{\lambda_{2,U}}{\lambda_{0,U}}} \quad (14)$$

in which $\lambda_{0,U}$ and $\lambda_{2,U}$ are, respectively, the zeroth and the second-order spectral moments of the response process as defined in Eq. 9. Furthermore, the dependence of the peak factor from the power-spectral density function of the response is overcome by approximating the response spectral moments with their solution for the white noise input leading to the formula

$$N_U \simeq \frac{\omega_0}{2\pi} T_s \quad (15)$$

Finally the stationary spectrum-compatible unilateral power-spectral density function can be expressed by using Eqs. 8, 10, and 11 as follows:

$$G(\omega_0) = \frac{1}{\omega_0 \left(\frac{\pi}{4\zeta_s} - 1 \right)} \left[\frac{RSA^2(\omega_0, \zeta_0)}{\eta_U^2} - \int_0^{\omega_0} G(\omega) d\omega \right] \quad (16)$$

in which also the damping ratio ζ_0 has been substituted with the fictitious damping ζ_s to take into account of the transient part of the response process

$$\zeta_s = (1 - e^{2\zeta_0 \omega_0 t})^{-1} \zeta_0 \quad (17)$$

which tends to the natural damping of the oscillator once the steady-state response has developed.

Furthermore, the two authors developed the computer software SIMQKE (Gasparini and Vanmarcke 1976) providing a practical tool to simulate time histories from a spectrum-compatible power-spectral density function which is implemented as

$$G(\omega_i) = \frac{4\zeta_s}{\omega_i \pi} \left(\frac{RSA^2(\omega_i, \zeta_0)}{\eta_U^2} - \Delta\omega \sum_{k=1}^{i-1} G(\omega_k) \right) \quad (18)$$

in which a more accurate expression of the peak factor is adopted under the hypothesis of barrier outcrossings in clumps, that is,

$$\eta_U(\omega_0, \zeta_0) = \sqrt{2 \ln \left\{ \frac{\omega_0 T_s}{\pi(-\ln p)} \left[1 - \exp \left[-\delta_U \sqrt{\pi \ln \left(\frac{\omega_0 T_s}{\pi(-\ln p)} \right)} \right] \right] \right\}} \quad (19)$$

where δ_U is the spread factor measuring the narrowness of the power-spectral density function of the response process defined in the range between 0 and 1 and approximated by the following expression for white noise input process:

$$\delta_U = \sqrt{1 - \frac{\lambda_{1,U}^2}{\lambda_{0,U} \lambda_{2,U}}} \simeq \sqrt{\frac{4\zeta_0}{\pi}} \quad (20)$$

After selecting a time-modulating function, the software employs the proposed power-spectral density function for simulating synthetic accelerograms via superposition of random harmonic functions through Eq. 7.

Another contribution has been provided by Kaul (1978); in order to provide a simplified expression of the power-spectral density function compatible with a given response spectrum, the author adopted the peak factor in Eq. 11 in conjunction with the hypothesis that the response spectral moments are determined under the hypothesis of white noise input process. Therefore, N_U is determined according to Eq. 15 and the zero-order response spectral moments is given as follows:

$$\lambda_{0,U} \simeq \frac{\pi}{4\zeta_0 \omega_0^3} G(\omega_0) \quad (21)$$

After some simple algebra the following stationary power-spectral density function is determined:

$$G(\omega_0) = \frac{4\zeta_e}{\pi \omega_0} \frac{RSA^2(\omega_0, \zeta_0)}{\eta_U^2} \quad (22)$$

where the damping of the oscillator has been adjusted as suggested by Rosenblueth and Elorduy (1969) as follows:

$$\zeta_e = \zeta_0 + \frac{2}{\omega_0 t_f} \quad (23)$$

in which t_f is the time duration of the stochastic process. As Kaul assumed the process as stationary t_f is coincident with T_s .

Equation 23 has been used by Der Kiureghian and Neuenhofer (1992) in order to investigate a response spectrum methodology for multi-support seismic excitation, in which a constant value of the peak factor was adopted. Currently their stationary power-spectral density function formula is adopted in Eurocode 8 part 2 (2010), recommended for the seismic analysis of bridges, that is,

$$G(\omega_0) = 2\omega_0^2 \left(\frac{2\zeta_0 \omega_0}{\pi} + \frac{4}{\pi T_s} \right) \left[\frac{RSD^2(\omega_0, \zeta_0)}{2.5^2} \right] \quad \omega \geq 0 \quad (24)$$

where $RSD(\omega_0, \zeta_0) = RSA(\omega_0, \zeta_0)/\omega_0^2$ is the displacement response spectrum.

Sundararajan (1980) introduced an iterative method for defining the spectrum-compatible power-spectral density function, differently from the direct approaches followed by Vanmarcke and Gasparini (1977), Kaul (1978), and Der Kiureghian and Neuenhofer (1992). The iterative procedure established by Sundararajan (1980) starts from a trial solution of the first passage problem given in Eq. 8 and variance of the response process given by Eq. 21:

$$G(\omega_0) = \frac{4\zeta_0 \omega_0}{\pi} \frac{RSV^2(\omega_0, \zeta_0)}{\eta_U^2} \quad (25)$$

where $RSV(\omega_0, \zeta_0) = RSA(\omega_0, \zeta_0)/\omega_0$ is the velocity response spectrum and the peak factor is assumed independent of the input according to the hypothesis of small damping. Both Davenport (1964) and Amin and Gungor (1971) formulae

have been considered suitable to this purpose. However, in the numerical analysis conducted by the authors, a constant value $\eta_U = 3$ has been selected.

After determining the first trial solution of the power-spectral density function approximate as piecewise of linear functions, the author expressed the variance of the displacement response accordingly as

$$\sigma_U = \sum_{k=0}^N \int_{\omega_k}^{\omega_{k+1}} \frac{G(\omega)}{(\omega_0^2 - \omega^2)^2 + 4\zeta_0^2 \omega_0^2 \omega^2} d\omega \quad k = 1, \dots, N \quad (26)$$

in which the integrals have been determined in closed form, that is,

$$\int_{\omega_k}^{\omega_{k+1}} \frac{G(\omega)}{(\omega_0^2 - \omega^2)^2 + 4\zeta_0^2 \omega_0^2 \omega^2} d\omega = \frac{G_k}{(\omega_{k+1} - \omega_k)} [\omega_{k+1} J(\omega_k, \omega_{k+1}) - K(\omega_k, \omega_{k+1})] + \frac{G_k}{(\omega_{k+1} - \omega_k)} [\omega_k J(\omega_k, \omega_{k+1}) - K(\omega_k, \omega_{k+1})] \quad (27)$$

where $J(\omega_k, \omega_{k+1})$ and $K(\omega_k, \omega_{k+1})$ are the solutions of the integrals given by the following equations

$$J(\omega_k, \omega_{k+1}) = \int_{\omega_k}^{\omega_{k+1}} \frac{1}{(\omega_0^2 - \omega^2)^2 + 4\zeta_0^2 \omega_0^2 \omega^2} d\omega \quad (28)$$

$$K(\omega_k, \omega_{k+1}) = \int_{\omega_k}^{\omega_{k+1}} \frac{\omega_0}{(\omega_0^2 - \omega^2)^2 + 4\zeta_0^2 \omega_0^2 \omega^2} d\omega \quad (29)$$

and determined analytically. The iterative procedure proceeds by comparison of the target response spectrum and the approximate response spectrum evaluated through the semi-analytical expression of the variance of the response displacement. The values G_k are then updated scaling their values by the ratio of the square of the target response spectrum and the approximate response spectrum.

In the same year Preumont (1980) coupled direct and iterative solutions in order to solve the inverse problem of spectrum-compatible ground motion processes obtained from given response spectra; following the approach used

by Kaul and Sundararajan, the suggested solution is given as

$$G(\omega_0) = \frac{4\zeta_e}{\pi\omega_0} \frac{RSA^2(\omega_0, \zeta_0)}{\eta_U^2} \quad (30)$$

in which the semiempirical peak factor given in Eqs. 19 and 20 but considering the adjusted spread factor $\delta_U^{1,2}$ (Vanmarke 1972) is employed. The fictitious damping ζ_e given by Eq. 17 is also used.

The approximate solution given in Eq. 30 provides accurate results for low and intermediate frequencies. Therefore the author considered the solution proposed by Vanmarcke and Gasparini (1977) given by Eq. 10 with the adjusted spread factor as further improvement of Eq. 19. As the peak factor evaluation depends on the whole power-spectral density function, which is intrinsically related to the spectral characteristics of the unknown response process (see Eqs. 9, 13, 14, and 20), the author developed an algorithm which enables to update the spectral characteristics of the response process every frequency step of definition of the frequency domain. The iterative scheme mainly differs from the procedure proposed by Sundararajan (1980) as not only the zeroth-order spectral moment is updated at each step but also the peak factor. To this aim the

first- and second-order spectral moments are also updated in the iterative scheme, using a stepwise representation of the power-spectral density function as

$$G(\omega_k) = \frac{1}{2} [G(\omega_k) + G(\omega_{k+1})] \quad k = 1, \dots, N \quad (31)$$

The spectral moments of the response are analytically determined according to the following formula

$$\lambda_{i,U} = \sum_{k=1}^N \int_{\omega_k}^{\omega_{k+1}} \frac{\omega^i G(\omega)}{(\omega_0^2 - \omega^2)^2 + 4\zeta_0^2 \omega_0^2 \omega^2} d\omega \quad k = 1, \dots, N \quad (32)$$

and approximating the first term for $i = 0, 1, 2$, as

$$\lambda_{0,U} = \int_{\omega_k}^{\omega_{k+1}} |H(\omega)|^2 G(\omega) d\omega \simeq \frac{\pi G(\omega_0)}{4\zeta_0 \omega_0^3} \varphi(\omega_k, \omega_{k+1}, \omega_0, \zeta_0) \quad (33)$$

$$\lambda_{1,U} = \int_{\omega_k}^{\omega_{k+1}} |H(\omega)|^2 G(\omega) d\omega \simeq \frac{\pi G(\omega_0)}{4\zeta_0 \omega_0^2 \sqrt{1 - \zeta_0}} \theta(\omega_k, \omega_{k+1}, \omega_0, \zeta_0) \quad (34)$$

$$\lambda_{2,U} = \int_{\omega_k}^{\omega_{k+1}} |H(\omega)|^2 G(\omega) d\omega \simeq \frac{\pi G(\omega_0)}{4\zeta_0 \omega_0} \psi(\omega_k, \omega_{k+1}, \omega_0, \zeta_0) \quad (35)$$

where $\varphi(\omega_k, \omega_{k+1}, \omega_0, \zeta_0)$, $\theta(\omega_k, \omega_{k+1}, \omega_0, \zeta_0)$, and $\psi(\omega_k, \omega_{k+1}, \omega_0, \zeta_0)$ are determined in closed form.

In the scenario of iterative methods to define the power-spectral density function from a known response spectrum, Pfaffinger (1983) proposed a procedure in which a discretization by piecewise polynomials is numerically implemented. The free parameters are determined iteratively by a least squares fit. The peak factor used the

expression provided by Davenport (1964), assuming the statistical independence of outcrossing of a given barrier, that is,

$$\eta_U = \sqrt{2 \ln(2N_U)} + \frac{\gamma}{\sqrt{2 \ln(2N_U)}} \quad (36)$$

where $\gamma = 0.5772$ is the Euler constant and N_U is given by Eq. 15.

The unilateral spectrum-compatible power-spectral density function is discretized in the interval $\omega_k < \omega < \omega_{k+1}$ and obtained by summation of the piecewise polynomials as follows:

$$G(\omega) = \sum_{k=1}^N p_k P_k(\omega) \quad k = 1, \dots, N \quad (37)$$

where p_i are either free parameters or have to be assigned to satisfy specified conditions and $P_i(\omega)$ are the interpolation polynomials.

According to the approximation given Eq. 37, the standard deviation and mean outcrossing rate of the zero level are obtained from the following equations, respectively:

$$\sigma_U = \left(\sum_{i=1}^N p_i I_{0,i}(\omega_0, \zeta_0) \right)^{1/2} \quad (38)$$

and

$$v_U = \frac{1}{2\pi} \left[\frac{\sum_{i=1}^N p_i I_{2,i}(\omega_0, \zeta_0)}{\sum_{i=1}^N p_i I_{0,i}(\omega_0, \zeta_0)} \right]^{1/2} \quad (39)$$

where $I_{2j,i}$ is given by

$$I_{2j,i}(\omega_0, \zeta_0) = \int_{\omega_k}^{\omega_{k+1}} \frac{\omega^{2j} P_k(\omega)}{(\omega_0^2 - \omega^2)^2 + 4\zeta_0^2 \omega_0^2 \omega^2} d\omega \quad j = 0, 1 \quad (40)$$

That has been derived analytically. The parameters p_i and the polynomial coefficients are finally determined through a least squares fitting.

Based on the iterative model proposed by Pfaffinger (1983), Park (1995) defined a method

to derive the equivalent power-spectral density function from a given target response spectrum.

The approach to the modeling consists of assuming the bilateral power-spectral density function as the summation of Dirac's delta functions with unknown amplitudes, that is,

$$S(\omega) = \sum_{j=1}^m S(\omega_j) \Delta\omega_j \delta(\omega_j) = \sum_{j=1}^m p_j \delta(\omega_j) \quad (41)$$

where $S(\omega)$ is the unknown two-sided power-spectral density function, $\delta(\omega_j)$ are Dirac's delta functions for each component, $\Delta\omega_j$ is the incremental frequency step, and p_j are the discretized power components.

Therefore by superimposing a series of Dirac's delta functions which may represent a narrowband process, the result will be a wide-band process as the real earthquake frequency content characterization.

The discretized target response spectrum is obtained as a superposition of the components of the power and of the response spectrum in itself as follows:

$$RSA^2(\omega_k, \zeta_0) = \sum_{j=1}^m p_j RSA_{k,j}^2(\omega_k, \omega_j, \zeta_0) \quad \text{for } k = 1, 2, \dots, m \quad (42)$$

where $RSA(\omega_k, \zeta_0)$ is the discretized target response spectrum and $RSA_{k,j}(\omega_k, \omega_j, \zeta_0)$ is the

peak acceleration response of a single degree of freedom system with natural frequency ω_k , damping ratio ζ_0 , and forced by the narrowband $\delta(\omega_j)$ Dirac's delta functions. $RSA_{k,j}(\omega_k, \omega_j, \zeta_0)$ is determined using the peak factor approximation by Davenport (1964) with N_U given by Eq. 15 and equivalent duration T_e defined as

$$T_e = \frac{\int_0^{t_f} a(t) dt}{\max[a(t)]} \quad (43)$$

where $a(t)$ is a selected time-modulating function and t_f is the total duration of the earthquake.

The unknown coefficients p_j defining the power-spectral density function are evaluated by minimizing the following penalty function:

$$\sum_{k=1}^n \left[RSA^2(\omega_k, \zeta_0) - \sum_{j=1}^m p_j RSA_{k,j}^2(\omega_k, \omega_j, \zeta_0) \right]^2 \quad p_j \geq 0; \quad (44)$$

where n represents the number of frequencies discretizing the target response spectrum.

Following the direct approach proposed by Vanmarcke and Gasparini (1977), Cacciola et al. (2004) developed a handy and accurate recursive expression for determining the spectrum-compatible power-spectral density function as follows:

$$G(\omega_i) = 0, \quad \forall 0 \leq \omega \leq \omega_x$$

$$G(\omega_i) = \frac{4\zeta_0}{\omega_i \pi - 4\zeta_0 \omega_{i-1}} \left(\frac{RSA(\omega_i, \zeta_0)^2}{\eta_U^2(\omega_i, \zeta_0)} - \Delta\omega \sum_{k=1}^{i-1} G(\omega_k) \right), \quad \forall \omega > \omega_x \quad (45)$$

where $\omega_x \cong 1 \text{ rad/s}$ is the lowest bound of the existence domain of then peak factor and η_U is the peak factor evaluated according to Vanmarcke

(1972) for white noise input (and $p = 0.5$) here rewritten for clarity's sake

$$\eta_U(\omega_0, \zeta_0) = \sqrt{2 \ln \left\{ \frac{\omega_0 T_s}{\pi(-\ln p)} \left[1 - \exp \left[-\delta_U^{1.2} \sqrt{\pi \ln \left(\frac{\omega_0 T_s}{\pi(-\ln p)} \right)} \right] \right] \right\}} \quad (46)$$

and spread factor (Vanmarke 1976)

$$\delta_U = \left[1 - \frac{1}{1 - \zeta_0^2} \left(1 - \frac{2}{\pi} \arctan \frac{\zeta_0}{\sqrt{1 - \zeta_0^2}} \right)^2 \right]^{\frac{1}{2}} \quad (47)$$

Recently Di Paola and Navarra (2009) proposed a closed form expression for the spectrum compatible power-spectral density function which parameters were defined through a best fitting of the model proposed by Cacciola et al. (2004).

Numerical Application

The direct models described are herein applied to simulate artificial quasi-stationary accelerograms compatible with a given response spectrum. Specifically, Eq. 7 in conjunction with Eq. 2 is used to generate the ground motion time histories.

Jennings et al. (1969) time-modulating function presented in Eq. 6 is used for this purpose. In this application the values of the parameters t_1 , t_2 , and β_4 are evaluated by imposing that the energy of the stochastic ground motion reaches the values of the 5 % and 95 %, respectively, in t_1 and t_2 , namely, by the use of the generalized Husid function extended for stochastic processes as proposed by Cacciola and Deodatis (2011)

$$H(t) = \frac{\int_0^t \int_0^\infty a^2(\tau) G(\omega) d\omega d\tau}{\int_0^\infty \int_0^\infty a^2(\tau) G(\omega) d\omega d\tau} \quad (48)$$

and imposing that $H(t_1) = 0.05$ and $H(t_2) = 0.95$, Eq. 47 leads to the following analytical values of Jennings's parameters as function of the duration of the stationary part T_s , that is,

$$\beta_4 = \frac{9}{T_s}; t_1 = \frac{2.5}{\beta_4}; t_2 = \frac{11.5}{\beta_4} \quad (49)$$

The stationary power-spectral density function in Eq. 48 is determined according to the direct procedures proposed by Gasparini and Vanmarke

(1976), Vanmarcke and Gasparini (1977), Kaul (1978), Preumont (1980), Der Kiureghian and Neuenhofer (1992), and Cacciola et al. (2004).

The Eurocode 8 response spectrum is used for illustrative purpose; specifically for a 5 % damping ratio, the response spectrum is given by the following equations:

$$\begin{aligned} RSA(T_0) &= a_g S \left[1 + \frac{T_0}{T_B} (1.5) \right] & 0 \leq T_0 \leq T_B \\ RSA(T_0) &= 2.5 a_g S & T_B \leq T_0 \leq T_C \\ RSA(T_0) &= 2.5 a_g S \left[\frac{T_C}{T_0} \right] & T_C \leq T_0 \leq T_D \\ RSA(T_0) &= 2.5 a_g S \left[\frac{T_C T_D}{T_0^2} \right] & T_D \leq T_0 \leq 4s \end{aligned} \quad (50)$$

where a_g is the design ground acceleration, S is the soil factor, T_0 is the natural period, and T_B , T_C , and T_D are the periods specified relatively to the soil type.

According to the Eurocode 8 provisions, the mean response spectrum of the simulated time histories has to match the target response spectrum within the prescribed tolerance along the prefixed range of periods according to the following condition:

$$\max \left\{ \frac{RSA(T_0) - \overline{RSA}(T_0)}{RSA(T_0)} \times 100 \right\} \leq 10\% \quad (51)$$

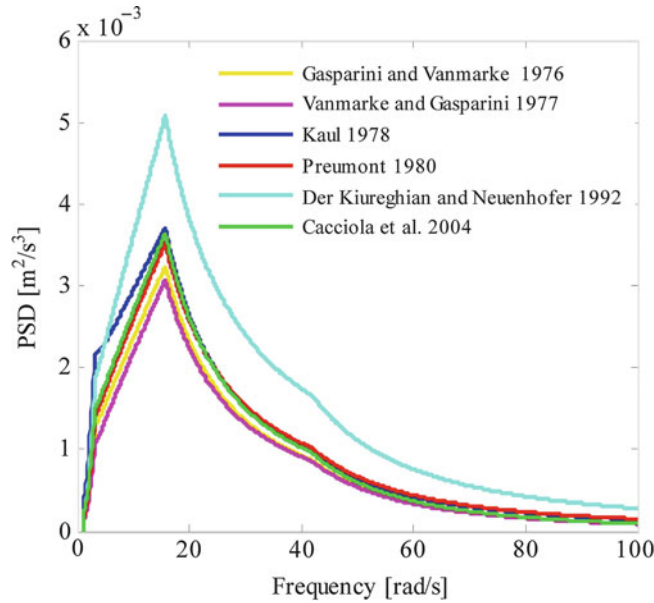
where $\overline{RSA}(T_0)$ is the mean response spectrum of at least three simulated accelerograms and $RSA(T_0)$ is the target one: moreover it has to be verified that

$$\overline{RSA}(0) > a_g S \quad (52)$$

The parameters chosen in this numerical application correspond to the ground Type A, namely, $S = 1$, $T_B = 0.1$ s, $T_C = 0.4$ s, $T_D = 2.0$ s; the maximum ground acceleration set is $a_g = 1$ m/s². Furthermore according to the Eurocode 8, the duration of the stationary part has to be taken at least of 10 s. Therefore the strong motion of Jennings's modulating function is set equal to $T_s = 10$ s

Response-Spectrum-Compatible Ground Motion Processes,

Fig. 1 Comparison between the power-spectral density functions



Response-Spectrum-Compatible Ground Motion Processes,

Fig. 2 Comparison between different peak factor's definitions adopted in the compared models

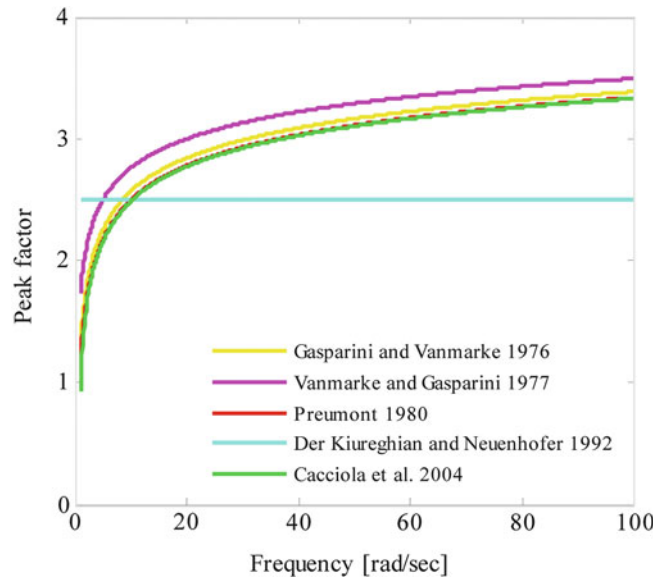


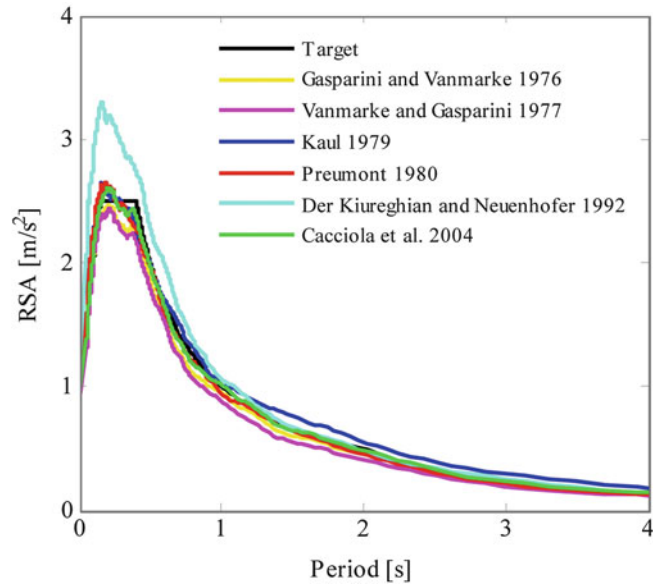
Figure 1 shows the power-spectral density functions according to the selected models which were numerically evaluated assuming the frequency step of 0.1 rad/s. Slight differences are noticeable between all the models apart from Der Kiureghian and Neuenhofer (1992) and Kaul (1978); all the other power-spectral density functions tend to similar values outside a certain range

of frequencies, in which the amplitude of the functions tends to negligibility.

Figure 2 elucidates the influence of assuming a constant value for the peak factor. The closed form expression given in the Eurocode 8 part 2, in which the peak factor is considered equal to 2.5, maintains considerable higher energy content compared with the others. From Fig. 2

Response-Spectrum-Compatible Ground Motion Processes,

Fig. 3 Comparison of the mean ensemble of the simulated response spectra



it can be observed that the constant value 2.5 is quite below the other peak factors in the range between 10 and 100 rad/s. Therefore the power-spectral density function is clearly overestimated by the function proposed in Eurocode 8 for $T_s = 10$ s.

Note that the peak factor relative to the model of Kaul (1978) is identical to the peak factor presented for Vanmarcke and Gasparini (1977); hence in Fig. 2, its variation with the frequency is not mentioned.

Successively 100 samples of accelerograms have been generated via superposition of harmonics through Eq. 7 using a time sampling step of 0.01 s and of 0.1 rad/s for the frequency domain. The time histories were generated in order to evaluate the mean of the ensemble of the simulated response spectra as shown in Fig. 3.

From Fig. 3 it is observed that the model proposed by Eurocode 8 leads to a mean simulated response spectrum quite above the recommended one; however it is spectrum compatible as it fulfils the conditions Eqs. 50 and 51.

No iteration has been adopted in order to improve the matching between the prescribed response spectrum and the mean-simulated stochastic response spectral accelerations, which

generally can be pursued by the following iterative scheme:

$$\begin{aligned} G_0^{(1)}(\omega_0) &= G_0(\omega_0); \quad G_0^{(j)}(\omega_0) \\ &= G_0^{(j-1)}(\omega_0) \left[\frac{RSA(\omega_0, \zeta_0)}{RSA^{(j-1)}(\omega_0, \zeta_0)} \right]^2 \end{aligned} \quad (53)$$

where $RSA^{(j-1)}$ is the approximate pseudo-acceleration spectrum determined at the j th iteration and RSA is the target response spectrum.

Nonstationary Models

In the previous sections the most common procedures to determine stationary/quasi-stationary power-spectral density functions compatible with a given response spectrum have been presented. It should be noted at the outset that the ground motion time histories generated from quasi-stationary stochastic processes have energy content only variable with time. Although this approach is convenient and accurate for the seismic analysis of traditional linear behaving structures, it does not lead to a comprehensive description of the seismic phenomenon.

It must be highlighted that the variation of the frequency content of stochastic ground motions has been recognized of primary importance in the seismic assessment of relevant engineering structures (see e.g., Yeh and Wen 1990; Wang et al. 2002); therefore the influence of the amplitude and the frequency time variation should be equally hypothesized to feature more realistically the structural response to earthquake ground motions.

Problem Position

Assuming the ground acceleration process as zero-mean Gaussian nonstationary non-separable

process, the evaluation of spectrum-compatible power-spectral density functions could be undertaken by the “first passage problem” pursued for the nonstationary case (Corotis et al. 1972; Cacciola 2011), that is,

$$\exp \left[- \int_0^{t_f} \alpha_U(RSA(\omega_0, \zeta_0), t) dt \right] = p \quad (54)$$

where t_f is the time interval in which the ground motion possesses not negligible energy and α_U is the hazard function defined as

$$\alpha_U(RSA(\omega_0, \zeta_0), t) = \frac{1}{\pi} \sqrt{\frac{\lambda_{2,U}(t)}{\lambda_{0,U}(t)}} \frac{1 - \exp \left(- \sqrt{\frac{\pi}{2}} \frac{RSA(\omega_0, \zeta_0)/\omega_0^2}{\sqrt{\lambda_{0,U}(t)}} \delta_U(t) \right)}{\exp \left(\frac{(RSA(\omega_0, \zeta_0)/\omega_0^2)^2}{2\lambda_{0,U}(t)} \right) - 1} \quad (55)$$

where the time-varying spread factor is given by

$$\delta_U(t) = \sqrt{1 - \frac{\lambda_{1,U}^2(t)}{\lambda_{0,U}(t)\lambda_{2,U}(t)}} \quad (56)$$

where $\lambda_{i,U}(t)$ ($i = 0, 1, 2$) represent the nonstationary spectral moments of the response (see e.g., Mikealov et al. 1999).

Ideally by Eqs. 54, 55, and 56 and the definition of the nonstationary spectral moments of the response, the evolutionary power-spectral density function pertinent to the codes prescriptions could be derivable. However this approach not yet attempted, to the best knowledge of the authors, it could be computationally burdensome, the convergence is not assured, and moreover it may lead to physically unacceptable results; therefore alternative approaches are usually preferred.

In the framework of Gaussian spectrum compatible nonstationary ground motion models, very few procedures have been proposed in literature. Herein the ground motion models developed by Spanos and Vargas Loli (1985),

Preumont (1985), and Cacciola (2010) are described; moreover the explicit formulae among those proposed are compared in a numerical application.

State of the Art for Available Solutions

Spanos and Vargas Loli (1985) for the first time developed the relationship between the evolutionary power-spectral density function and a known response spectrum. The authors defined an approximate analytical expression whose parameters were derived from an optimization procedure.

In order to determine the link between the power-spectral density function and the target response spectrum, the authors derived the following relationship:

$$RSV(\omega_0, \zeta_0) = r\sigma_{U_{\max}}^2(\omega_0, \zeta_0) \quad (57)$$

where $RSV(\omega_0, \zeta_0)$ is the velocity target response spectrum, $\sigma_{U_{\max}}$ is the maximum of the variance of the response velocity, and r is a scaling factor which value is taken equal to $\sqrt{\pi/2}$.

The variance of the velocity response process is approximated by the authors by

$$\begin{aligned}\sigma_{\dot{U}}^2(t) &\approx \omega_0^2 \sigma_U^2(t) \\ &= \pi \exp(-2\zeta_0 \omega_0 t) \int_0^t \exp(2\zeta_0 \omega_0 t) S(\omega, t) dt\end{aligned}\quad (58)$$

where $S(\omega, t)$ is the bilateral power-spectral density function to be determined. In order to determine the power-spectral density function from Eq. 56, the authors assumed the evolutionary power spectrum as a linear combination of m separable functions through the following mathematical model:

$$S(\omega, t) = \sum_{k=1}^m C_k t^2 \exp(-\alpha_k t) S_k(\omega) \quad (59)$$

$S_k(\omega)$ is the Kanai-Tajimi power-spectral density function

$$S_k(\omega) = \frac{1 + 4\zeta_k^2 \left(\frac{\omega_0}{\omega_k}\right)^2}{\left\{1 - \left(\frac{\omega_0}{\omega_k}\right)^2\right\}^2 + 4\zeta_k^2 \left(\frac{\omega_0}{\omega_k}\right)^2} \quad (60)$$

where ω_k and ζ_k , which are respectively the natural frequency and the damping ratio of the single degree of freedom system representing the soil, are parameters to be determined along with C_k and α_k in Eq. 58.

By substituting Eq. 58 into Eq. 57, the maximum of $\sigma_{\dot{U}}^2(t)$ is given by

$$\sigma_{\dot{U}_{\max}}^2 = \frac{\pi}{2\zeta_0 \omega_0} \sum_{k=1}^m C_k t^{*2} \exp(-\alpha_k t^*) S_k(\omega_0) \quad (61)$$

Therefore by means of Eqs. 60 and 56, the relationship between the velocity target response spectrum and the modeled evolutionary power-spectral density function is expressed by the equation

$$RSV^2(\omega_0, \zeta_0) = r^2 \frac{\pi}{2\zeta_0 \omega_0} \sum_{k=1}^m C_k t^{*2} \exp(-\alpha_k t^*) S_k(\omega_0) \quad (62)$$

The differences between the target spectrum on the left-hand side and the right-hand side of the above formula are minimized through the Levenberg-Marquardt procedure by best fitting of the constant parameters $t_1^*, t_2^*, \dots, t_n^*, C_k, \alpha_k, \zeta_k, \omega_k$.

After defining the evolutionary power-spectral density function samples of ground motion process are generated by Eq. 7. An iterative procedure is successively required in order to ensure the matching of the simulated and target response spectra (Giaralis and Spanos 2009).

In the same year, the relevance of the nonstationarity of the frequency content in the modeling of the ground motion has been highlighted by Preumont (1985); the author defined an empirical model of the non-separable power-spectral density function. The model assumes that the high-frequency components are magnified in the early part of the time history. Specifically, the following evolutionary spectrum is assumed

$$G(\omega, t) = t^2 e^{-\alpha(\omega)t} G(\omega) \quad (63)$$

with

$$\alpha(\omega) = \alpha_0 + \alpha_1 \omega + \alpha_2 \omega^2 \quad (64)$$

where α_0 , α_1 , and α_2 are adjustable parameters calibrated to impose the equation between the energy of the separable and non-separable process.

The matching of the mean simulated and target response spectra is ensured by equating at each frequency the energy of the separable spectrum-compatible process and that one of the non-separable, that is,

$$G^S(\omega) \int_0^\infty a(t)^2 dt = G(\omega) \int_0^\infty t^2 e^{-\alpha(\omega)t} dt = G(\omega) \frac{2}{\alpha(\omega)^3} \quad (65)$$

where $G^S(\omega)$ is the stationary response-spectrum-compatible power-spectral density function determined according to the procedures described in the previous sections and $a(t)$ is the time-modulating function (see e.g., Eq. 6)

Therefore through Eq. 64 $G(\omega)$ in Eq. 62 is readily determined as follows:

$$G(\omega) = \frac{\int_0^\infty a^2(t) dt}{\int_0^\infty t^2 e^{-\alpha(\omega)t} dt} G^S(\omega) = \frac{\alpha(\omega)^3}{2} G^S(\omega) \int_0^\infty a^2(t) dt \quad (66)$$

Once $a(t)$ is defined, the evolutionary power-spectral density function is pursued by means of Eqs. 62 and 63 in which the coefficients α_0 , α_1 , and α_2 are suitably determined.

An alternative approach has been proposed by Cacciola (2010); the author's contribution allows a straightforward evaluation of a non-separable power-spectral density function compatible with a target response spectrum. In the model proposed by Cacciola (2010), it is assumed that the nonstationary spectrum-compatible evolutionary ground motion process is given by the superposition of two independent contributions: the first one is a fully nonstationary known counterpart which accounts for the time variability of both intensity and frequency content; the second one is a corrective term represented by a quasi-stationary zero-mean Gaussian process that adjusts the nonstationary signal in order to make it spectrum compatible. Therefore the ground motion can be split in two contributions:

$$\ddot{u}_g(t) = \ddot{u}_g^R(t) + \ddot{u}_g^C(t) \quad (67)$$

Taking into account of the statistical independence of the two contributions, the evolutionary spectrum-compatible power-spectral density function can be expressed as

$$G(\omega, t) = G^R(\omega, t) + G^C(\omega, t) \quad (68)$$

where $G^R(\omega, t)$ is the joint time-frequency distribution of the recorded accelerogram and $G^C(\omega, t)$

is the separable power-spectral density function representing of the corrective term given by

$$G^C(\omega, t) = a^2(t)G(\omega) \quad (69)$$

Specifically $G^R(\omega, t)$ is attained by the fully nonstationary model of Conte and Peng (1997), which hypothesizes the stochastic ground motion as a series of zero-mean independent uniformly modulated Gaussian processes as

$$G(\omega, t) = \sum_{j=1}^N |a_j(t)|^2 G_j(\omega) \quad (70)$$

where

$$a_j(t) = \alpha_j(t - \vartheta_j)^{\beta_j} \exp(-\gamma_j(t - \vartheta_j)) U(t - \vartheta_j) \\ j = 1, \dots, N \quad (71)$$

where $U(t - \vartheta_j)$ is the unit step function and $G_j(\omega)$ a stationary counterpart obtained as

$$G_j(\omega) = \frac{v_j}{\pi} \left[\frac{1}{v_j^2 + (\omega + \eta_j)^2} + \frac{1}{v_j^2 + (\omega - \eta_j)^2} \right] \\ j = 1, \dots, N \quad (72)$$

in which the parameters N , α_j , β_j , γ_j , ϑ_j , η_j , v_j are determined by a best fit procedure in order to minimize the differences between the analytical model and the joint time-frequency distribution of a real earthquake.

In order to define the corrective term $G^C(\omega, t)$ according to the quasi-stationary model described in the previous section, $a(t)$ is the modulating function given in Eq. 6 whose parameters are calibrated by the knowledge of the recorded accelerograms and $G(\omega)$ is determined modifying Eq. 45 as follows:

$$G^C(\omega_i) = \begin{cases} 0, & 0 \leq \omega \leq \omega_x \\ \frac{4\zeta_0}{\omega_i \pi - 4\zeta_0 \omega_{i-1}} \left(\frac{RSA(\omega_i, \zeta_0)^2 - RSA^R(\omega_i, \zeta_0)^2}{\eta_U^2(\omega_i, \zeta_0)} - \Delta \omega \sum_{k=1}^{i-1} G^C(\omega_k) \right), & \omega > \omega_x \end{cases} \quad (73)$$

where η_U is the peak factor defined in Eq. 19 and RSA^R is the pseudo-acceleration response spectrum for the recorded ground motions $\ddot{u}_g^R(t)$. Equation 72 is defined for $RSA(\omega_i, \zeta_0) > RSA^R(\omega_i, \zeta_0)$, as a consequence a preliminary scaling procedure might be required.

It has to be emphasized that due to the mathematical structure of the model proposed by Conte and Peng (1997), Eq. 67 can be rewritten in the form

$$\begin{aligned} G(\omega, t) &= G^R(\omega, t) + G^C(\omega, t) \\ &= \sum_{j=1}^N |a_j(t)|^2 G_j(\omega) + a_0^2(t) G(\omega) \\ &= \sum_{j=0}^N |a_j(t)|^2 G_j(\omega) \end{aligned} \quad (74)$$

which allows a compact representation of the evolutionary spectrum-compatible power-spectral density function. It is noted that in the formulation proposed by Cacciola (2010), the nonstationary frequency content is not modeled through empirical analytical formulas, but it is selected from real records.

Numerical Application

In this section the ground motion stochastic processes modeled according to Preumont (1985) and Cacciola (2010) are presented through a numerical example. The models have been selected as those holding an explicit formulation of the evolutionary power-spectral density function among those herein presented.

The spectrum-compatibility of the two procedures is illustrated considering a given target response spectrum as prescribed by Eurocode 8 and expressed in Eq. 50. The response spectrum for Type B soil and Type 1 seismicity is selected; the parameters corresponding to the recommended spectrum are, namely, $S = 1.2$, $T_B = 0.15$ s, $T_C = 0.5$ s, $T_D = 2.0$ s and the design ground acceleration was set equal to 0.35 g.

To implement the model proposed by Preumont (1985), the values relatively to the time-modulating function seen in Eqs. 62 and 63

were set respectively equal to $\alpha_0 = 0.3$, $\alpha_1 = 0.01$, and $\alpha_2 = 0$. Also, the time-modulating function is selected according to the author preferences; hence Jennings et al.'s (1969) envelope function was taken for a strong motion of $T_s = 13$ s, $t_1 = 2$ s, $t_2 = 15$ s, and total duration $t_f = 20$ s.

In order to apply the procedure proposed by Cacciola (2010), firstly the evolutionary power spectrum as developed by Conte and Peng (1997) is obtained specifically for El Centro 1940 earthquake, north-south component SOOE (N-S) of the Imperial Valley.

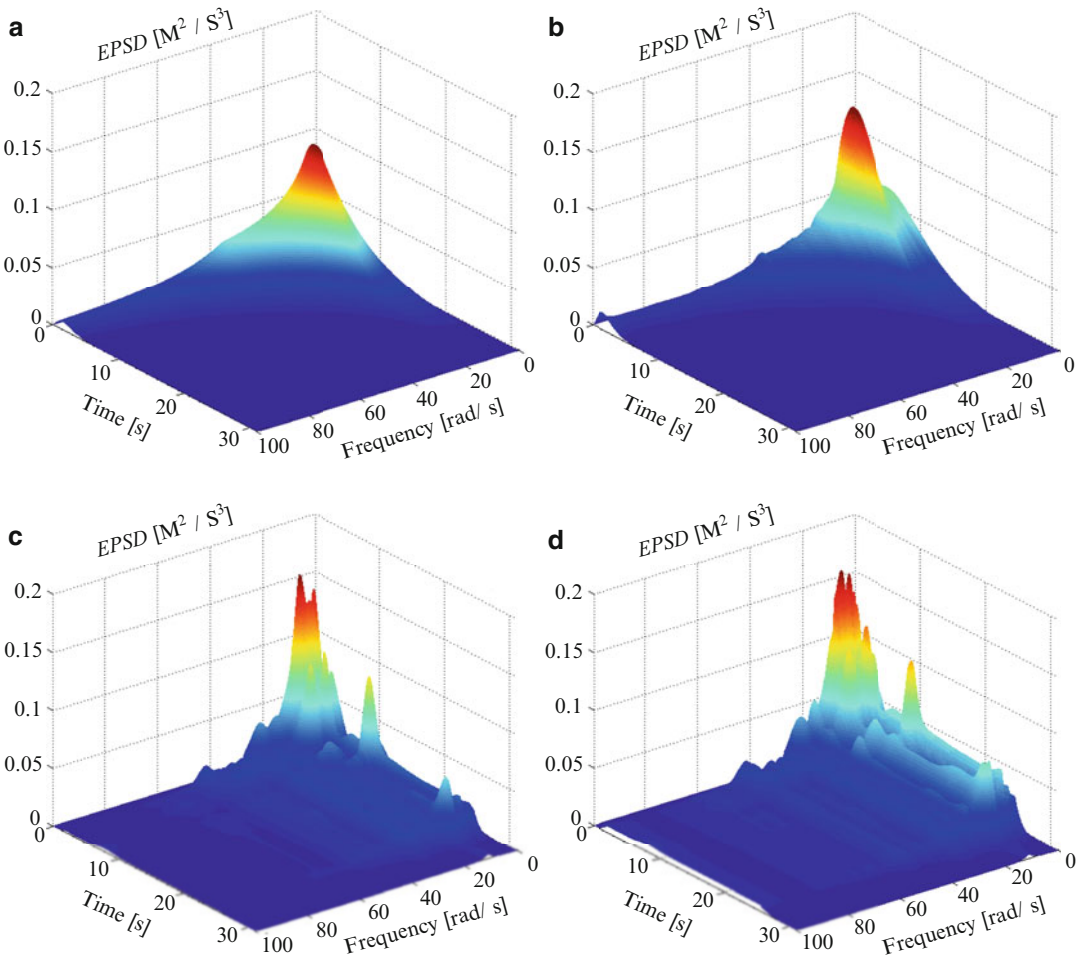
Jennings et al.'s (1969) model is chosen also for this model as in Eq. 6. The parameters for the time-modulating function are defined according to the real strong motion quantified as $T_s = 23.86$ s, namely, these are $t_1 = 1.65$ s, $t_2 = 25.51$ s total duration $t_f = 31$ s and parameter $\beta_4 = 3/(t_f - t_2)$. Finally to evaluate the simulated response spectra a number of 100 time histories were generated via Monte Carlo simulation method by Eq. 7.

The evolutionary power-spectral density functions proposed by the authors are displayed in Fig. 4a, b; in general, due to the nonstationarity and the approximations involved in the models, the iterative scheme as seen in Eq. 52 is required. Therefore Fig. 4c, d shows the iterated power-spectral density functions; markedly it should be observed that although the different joint time-frequency distribution both the models can be used according to the seismic provisions for simulating ground motion accelerograms and evidently tending to reach the spectrum-compatibility criteria as shown in Fig. 5.

Figure 5 displays that both the models reached the required spectrum-compatibility after the contained number of 5 iterations over the recommended range of period.

Nonstationary Vector Processes

Ground motion arising from seismic waves is a phenomenon that by its nature varies with time and in space. Even if, in earthquake engineering



Response-Spectrum-Compatible Ground Motion Processes, Fig. 4 Evolutionary power-spectral density functions according to (a) Preumont (1985); (b) Preumont

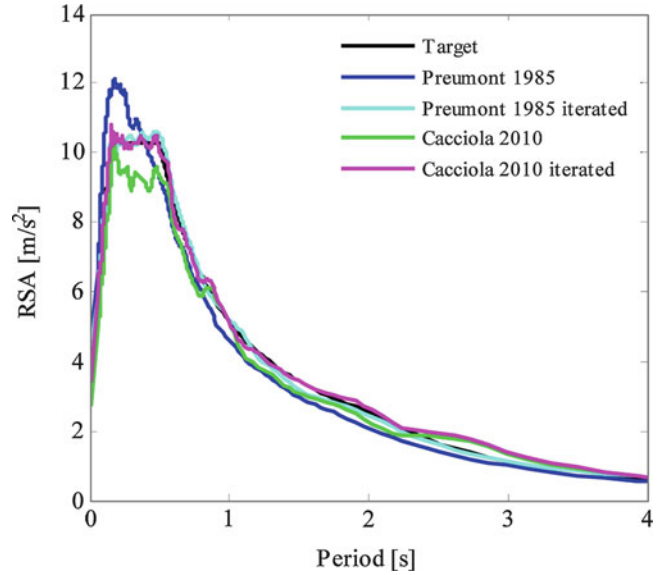
(1985) after 5 iterations; (c) Cacciola (2010); (d) Cacciola (2010) after 5 iterations

practice, the attention is generally focused on the time variability feature and its effects on the structural response; on the other hand, it is well known that the spatial variability of earthquake ground motion can influence significantly the response of structures, especially if they are long and/or rigid such as bridges and nuclear power plants. Zerva (2009) presents an overview of the engineering applications in which ground motion spatial variability has been taken into account and for an in-depth discussion of the state of the art along with the challenges

involved in modeling the spatial variability of earthquake ground motion. The spatial variability of the earthquake ground motion can be modeled assuming the hypothesis of a zero-mean Gaussian multivariate vector process. Following the spectral-representation method proposed by Shinozuka and Deodatis (1988), assuming the 1D- m V (one-dimensional, m -variate), the nonstationary ground motion stochastic vector process with components $f_j^0(t)$, ($j = 1, \dots, m$) is defined by the cross-correlation matrix given by

Response-Spectrum-Compatible Ground Motion Processes,

Fig. 5 Comparison of the ensemble of the simulated response spectra before and after 5 iterations



$$\mathbf{R}_f^0(t, t + \tau) = \begin{bmatrix} R_{11}^0(t, t + \tau) & R_{12}^0(t, t + \tau) & \cdots & R_{1m}^0(t, t + \tau) \\ R_{21}^0(t, t + \tau) & R_{22}^0(t, t + \tau) & \cdots & R_{2m}^0(t, t + \tau) \\ \vdots & \vdots & \ddots & \vdots \\ R_{m1}^0(t, t + \tau) & R_{m2}^0(t, t + \tau) & \cdots & R_{mm}^0(t, t + \tau) \end{bmatrix} \quad (75)$$

$$\mathbf{S}_f^0(\omega, t) = \begin{bmatrix} S_{11}^0(\omega, t) & S_{12}^0(\omega, t) & \cdots & S_{1m}^0(\omega, t) \\ S_{21}^0(\omega, t) & S_{22}^0(\omega, t) & \cdots & S_{2m}^0(\omega, t) \\ \vdots & \vdots & \ddots & \vdots \\ S_{m1}^0(\omega, t) & S_{m2}^0(\omega, t) & \cdots & S_{mm}^0(\omega, t) \end{bmatrix} \quad (76)$$

Or alternatively by the corresponding cross-spectral density matrix given by

whose elements are expressed in the following form:

$$\begin{aligned} S_{jj}^0(\omega, t) &= |a_j(\omega, t)|^2 S_j(\omega), & j &= 1, 2, \dots, m \\ S_{jk}^0(\omega, t) &= a_j(\omega, t) a_k(\omega, t) \sqrt{S_j(\omega) S_k(\omega)} \gamma_{jk}(\omega), & j, k &= 1, 2, \dots, m; \quad j \neq k \end{aligned} \quad (77)$$

where $a_j(\omega, t)$ and $S_j(\omega)$ ($j = 1, 2, \dots, m$) are the (non-separable) modulating function and the (stationary) power-spectral density function of component $f_j^0(t)$, ($j = 1, 2, \dots, m$), respectively, and $\gamma_{jk}(\omega)$ ($j, k = 1, 2, \dots, m; j \neq k$) is the complex coherence function between $f_j^0(t)$ and $f_k^0(t)$. The elements of the cross-correlation matrix are related to the corresponding elements of the cross-spectral density matrix through the transformations:

$$R_{jj}^0(t, t + \tau) = \int_{-\infty}^{\infty} a_j(\omega, t) a_j(\omega, t + \tau) S_j(\omega) e^{i\omega\tau} d\omega; \quad j = 1, 2, \dots, m \quad (78)$$

$$\begin{aligned} R_{jk}^0(t, t + \tau) &= \int_{-\infty}^{\infty} a_j(\omega, t) a_k(\omega, t + \tau) \\ &\quad \sqrt{S_j(\omega) S_k(\omega)} \gamma_{jk}(\omega) e^{i\omega\tau} d\omega; \\ &\quad j, k = 1, 2, \dots, m; \quad j \neq k \end{aligned} \quad (79)$$

For the special case of uniformly modulated nonstationary stochastic vector process, the modulating functions $a_j(\omega, t)$ ($j = 1, 2, \dots, m$) are independent of the frequency ω , that is,

$$a_j(\omega, t) = a_j(t), \quad j = 1, 2, \dots, m \quad (80)$$

In this case, Eqs. 77 and 78 reduce to

$$R_{jj}^0(t, t + \tau) = a_j(t)a_j(t + \tau) \int_{-\infty}^{\infty} S_j(\omega)e^{i\omega\tau}d\omega; \quad j = 1, 2, \dots, m \quad (81)$$

$$R_{jk}^0(t, t + \tau) = a_j(t)a_k(t + \tau) \int_{-\infty}^{\infty} \sqrt{S_j(\omega)S_k(\omega)}\gamma_{jk}(\omega)e^{i\omega\tau}d\omega; \quad j, k = 1, 2, \dots, m; j \neq k \quad (82)$$

After determining the power-spectral density matrix samples of 1D- m -variate ground motion process can be determinate (Deodatis 1996). Specifically, Cholesky's decomposition is applied to decompose the power spectra density matrix as follows:

$$\mathbf{S}_f^0(\omega, t) = \mathbf{H}(\omega, t)\mathbf{H}^{T*}(\omega, t) \quad (83)$$

where $\mathbf{H}(\omega, t)$ is a lower triangular matrix and the superscript T denotes the transpose of a matrix. $\mathbf{H}(\omega, t)$ is written as

$$\mathbf{H}(\omega, t) = \begin{bmatrix} H_{11}(\omega, t) & 0 & \cdots & 0 \\ H_{21}(\omega, t) & H_{22}(\omega, t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ H_{m1}(\omega, t) & H_{m2}(\omega, t) & \cdots & H_{mm}(\omega, t) \end{bmatrix} \quad (84)$$

The diagonal elements of $\mathbf{H}(\omega, t)$ are real and nonnegative functions of ω satisfying

$$H_{jj}(\omega, t) = H_{jj}(-\omega, t), \quad j = 1, 2, \dots, m; \quad \forall t \quad (85)$$

while the off-diagonal elements are generally complex functions of ω . Once the cross-spectral density matrix $\mathbf{S}_f^0(\omega, t)$ is decomposed according to Eq. 82 and 83, the nonstationary ground motion vector process $f_j^0(t)$, $j = 1, 2, \dots, m$ can be simulated by the following series as $N \rightarrow \infty$

$$f_j(t) = 2 \sum_{r=1}^m \sum_{s=1}^N |H_{jr}(\omega_s, t)| \sqrt{\Delta\omega} \cos [\omega_s t - \vartheta_{jr}(\omega_s, t) + \phi_{rs}], \quad j = 1, 2, \dots, m \quad (86)$$

Or explicitly

$$f_1(t) = 2 \sum_{s=1}^N |H_{11}(\omega_s, t)| \sqrt{\Delta\omega} \cos [\omega_s t - \vartheta_{11}(\omega_s, t) + \phi_{1s}] \quad (87)$$

$$f_2(t) = 2 \sum_{s=1}^N |H_{21}(\omega_s, t)| \sqrt{\Delta\omega} \cos [\omega_s t - \vartheta_{21}(\omega_s, t) + \phi_{1s}] + 2 \sum_{s=1}^N |H_{22}(\omega_s, t)| \sqrt{\Delta\omega} \cos [\omega_s t - \vartheta_{22}(\omega_s, t) + \phi_{2s}] \quad (88)$$

where

$$\vartheta_{jk}(\omega, t) = \tan^{-1} \left(\frac{\text{Im}[H_{jk}(\omega, t)]}{\text{Re}[H_{jk}(\omega, t)]} \right) \quad (89)$$

with $\text{Im}[\cdot]$ and $\text{Re}[\cdot]$ denoting the imaginary and real part of a complex number, respectively. Furthermore, the ϕ_{rs} ($r = 1, 2, \dots, m$; $s = 1, 2, \dots, N$) are m sequences of N independent random phase angles distributed uniformly over the interval $[0, 2\pi)$.

Clearly, the proper definition of the power-spectral density matrix is the crucial step to address for the stochastic modeling of the

seismic action. In this regard the challenge is to simulate artificial accelerograms, that are compatible with response spectra defined in different locations and correlated through the coherence function $\gamma_{jk}(\omega)$ introduced in Eq. 76. The definition of the coherence function is still an open issue in the scientific community. The readers might refer to the monograph of Zerva (2009) for a detailed description of coherence functions.

Based on the data provided from the main Strong Motion Array Taiwan (SMART-1), Hao et al. (1989) developed a method for generating stationary and quasi-stationary multi-support inputs for any given set of location which are compatible with target response spectra. To this aim a coherence function is developed for pairs of stations, that is,

$$\begin{aligned} \gamma_{ij}(d_{ij}^L, d_{ij}^T, \omega) = & \exp(-b_1 d_{ij}^L - b_2 d_{ij}^T) \\ & \exp\left[-\left(-a_1 d_{ij}^{L1/2} + a_2 d_{ij}^{T1/2}\right)\omega^2\right] \\ & \exp\left(i\omega \frac{d_{ij}^L}{v_{app}}\right) \end{aligned} \quad (90)$$

where d_{ij}^L, d_{ij}^T are the station pairs distances with same projected separation in the longitudinal and transverse direction of the preferential wave propagation, ω is the circular frequency, v_{app} is the apparent wave velocity, the parameters a_1, a_2, b_1 , and b_2 were evaluated by least squares fitting using SMART-1 data.

The averaged power-spectral density function $S_k(\omega)$ for all the stations was calculated by best fitting the Kanai-Tajimi model. Finally, the modulating function amplitude variation was evaluated assuming the following model:

$$a(t) = \delta t \exp(-\beta_5 t^2) \quad (91)$$

whose parameters are determined through best-fitting procedures. Once defined the elements of the power-spectral density matrix, the simulation of quasi-stationary vector process is pursued through a specific algorithm. The matching with the target response spectrum is guaranteed through an iterative procedure, scaling the

Fourier transform of the generated accelerograms through the ratio of the target and simulated response spectra and then determining the spectrum-compatible ground motion acceleration through the inverse Fourier transform.

Deodatis (1996) proposed a general iterative procedure to determine quasi-stationary response-spectrum-compatible ground motion vector process. The procedure requires the initial selection of the modulating functions and coherence functions representative of a given site. The procedure starts assuming the stationary power-spectral density function constant for every location:

$$S_j(\omega, t) = a^2(t) S_j(\omega) \quad (92)$$

where $a(t)$ is the time-modulating function and $S_j(\omega)$ the stationary power-spectral density function.

The problem solution starts assuming a constant value $S_j(\omega)$ at each frequency.

The ground motion accelerograms at each point on the free field are then simulated according to the spectral-representation method described in Eqs. 76, 77, 78, 79, and 80 and 83, 84, 85, 86, 87, 88, and 89. The matching is guaranteed by updating individually each power-spectral density functions $S_j(\omega)$ according to the iterative scheme described in Eq. 52.

By using a wavelet-based approach, a method to define fully nonstationary spectrum-compatible vector processes has been introduced by Sarkar and Gupta (2005). The proposed scheme assumes that a suitable accelerogram conforming to the local source and site conditions is available. The procedure is developed assuming that each point at the free field possess the same response spectrum. The accelerogram is first modified to make it spectrum compatible to the target response spectrum through a deterministic iterative procedure. The power-spectral density function from the response-spectrum-compatible accelerogram at a given time instant is determined as

$$S_X(\omega_j) = \frac{K}{a_j} E \left[|W_{\psi} X(a_j, b_j)|^2 \right] \left| \hat{\psi}_{a_j, b_j}(\omega_j) \right|^2 \quad (93)$$

where $W_{\psi}X(a_j, b_j)$ is the wavelet coefficient for the response-spectrum-compatible accelerogram at scale parameter a_j and shift parameter $b_j = (i - 1)\Delta b$ in which Δb is the sampling time interval of the accelerogram, K is a scaling coefficient, and $\hat{\psi}_{a_j, b_j}(\omega_j)$ is the Fourier transform of $\psi_{a_j, b_j}(t_j)$ defined from the mother wavelet.

The off-diagonal terms in Eq. 76 are determined imposing a coherence function. The simulation of m -variate response-spectrum-compatible accelerograms is then performed extending the spectral-representation method described in Eqs. 76, 77, 78, 79, and 80 and 83, 84, 85, 86, 87, 88, and 89 to a wavelet simulation-based procedure.

Recently, the procedure developed by Deodatis (1996) for quasi-stationary response-spectrum-compatible vector process has been extended by Cacciola and Deodatis (2011) to the fully nonstationary case. According to the model introduced by Cacciola (2010), the authors proposed a response-spectrum-compatible ground motion model given by the superposition of two contributions:

$$f_j^{SC}(t) = f_j^L(t) + f_j^C(t) \quad j = 1, 2, \dots, m \quad (94)$$

where $f_j^L(t)$, $j = 1, 2, \dots, m$ represents the j th component of a fully nonstationary vector process reflecting the “local” geological and seismological conditions of a given site and $f_j^C(t)$, $j = 1, 2, \dots, m$ represents the j th

component of a quasi-stationary corrective term, whose power-spectral density matrix has to be determined.

The terms of the evolutionary spectrum-compatible power-spectral density matrix are then given by the following equation:

$$S_{jj}^{SC}(\omega, t) = S_{jj}^L(\omega, t) + S_{jj}^C(\omega, t) \quad j = 1, 2, \dots, m \quad (95)$$

and the off-diagonal terms

$$S_{jk}^{SC}(\omega, t) = \sqrt{S_{jj}^{SC}(\omega, t)S_{kk}^{SC}(\omega, t)}\gamma_{jk}(\omega) \quad j, k = 1, 2, \dots, m; j \neq k \quad (96)$$

The corrective term $f_j^C(t)$, $j = 1, 2, \dots, m$ in Eq. 93 is assumed quasi-stationary with evolutionary power-spectral density defined by

$$S_{jj}^C(\omega, t) = a_j^2(t)S_j^C(\omega) \quad j = 1, 2, \dots, m \quad (97)$$

where $a_j(t)$ is the j th modulating function preserving the amplitude time variability of the process $f_j^L(t)$ and $S_j^C(\omega)$ is the stationary power-spectral density of the corrective process. The latter is then determined modifying the recursive formula proposed by Cacciola et al. (2004) for the unilateral power-spectral density function $G_j^C(\omega)$ related to the bilateral by $G_j^C(\omega) = 2S_j^C(\omega)$, as follows:

$$G_j^C(\omega_i) = \begin{cases} 0, & \omega > \omega_x \\ \frac{4\zeta_0}{\omega_i\pi - 4\zeta_0\omega_{i-1}} \left(\frac{RSA^{(j)}(\omega_i, \zeta_0)^2 - RSA^{(f_j^L)}(\omega_i, \zeta_0)^2}{\eta_U(\omega_i, \zeta_0)^2} - \Delta\omega \sum_{j=1}^{m-1} G_j^C(\omega_k) \right) & \omega > \omega_x \end{cases} \quad (98)$$

where $RSA^{(j)}(\omega_i, \zeta_0)$ is the target response spectrum and $RSA^{(f_j^L)}(\omega_i, \zeta_0)^2$ is the pseudo-acceleration response spectrum relatively to the “local” stochastic process and the peak factor η_U has been defined in Eq. 46

The spectrum compatibility is achieved by correcting the $G_j^C(\omega)$ of Eq. 52 iteratively once simulated response spectra are obtained through the generation technique described in Eqs. 76, 77, 78, 79, and 80 and 83, 84, 85, 86, 87, 88, and 89.

Summary

In this entry the simulation of response-spectrum-compatible ground motion processes has been addressed. Gaussian stationary, quasi-stationary, and fully nonstationary processes have been described. The assumption of stationarity of the ground motion process provided a robust relationship between the target response spectrum and the power-spectral density function of the input. This relationship determined in the early 1970s is still used by researchers and practitioners as a vehicle to develop more advanced ground motion models. Although the assumption of stationary ground motion process is physically unacceptable, the stationary model and the related power-spectral density function provide a useful tool for determining the stochastic response of structures, through handy analytical formulas. Note that the current modal combination rule (i.e., CQC) implemented in the commercial software for the seismic analysis of structures is determined using the hypothesis of stationary Gaussian ground motion process. Moreover, stationary models can be used as a vehicle for predesign structures and/or control devices in a probabilistic sense. The quasi-stationary ground motion processes have been then introduced in literature by modulating the energy of the ground motion through empirically based analytical functions. The quasi-stationary model has the advantage to be suitable for the simulation of artificial earthquakes preserving the handy analytical formulation of the stationary process. When the nonstationarity is involved, it has to be emphasized that a certain level of subjectivity is introduced to the modeling even for the simplest quasi-stationary case. Therefore, inputs from seismological and geological data are necessary to elaborate a reliable nonstationary model. Moreover, as the nonstationary frequency content of real earthquakes has been found to be relevant for the seismic assessment of structures, more advanced response-spectrum-compatible models have been proposed and discussed in

this entry. Various approaches were proposed due to the challenge involved in the relationship between the target response spectrum and the evolutionary power-spectral density function of the ground motion process. It has to be emphasized that different models are determined to match a given response spectrum (for a given damping ratio), but they might lead to completely different results for structures with damping ratios different from that used in the calibration of the response-spectrum-compatible model and also for nonlinear behaving structures. This can be considered as a pitfall in the definition of the spectrum-compatibility criteria. Finally response-spectrum-compatible ground motion models taking into account the spatial variability of the ground motion through the generally known coherence function have been also addressed. The spatial variability is of particular relevance for bridge analysis and in general for long and/or rigid structures. Numerical applications completed the description of the response-spectrum-compatible ground motion processes.

In conclusion the definition of a good ground motion model is still an open issue in the scientific community; current trends show major interest between analytical modeling and data gathered from record's databases, trying to evaluate the relevant features of the real earthquakes.

Cross-References

- [Building Codes and Standards](#)
- [Random Process as Earthquake Motions](#)
- [Stochastic Analysis of Linear Systems](#)

References

- Ahmadi G (1979) Generation of artificial time-histories compatible with given response spectra – a review. *SM Arch* 4(3):207–239
- Amin M, Gungor I (1971) Random vibration in seismic analysis: an evaluation. ASCE National meeting on structural engineering, Baltimore, pp 19–23

- Cacciola P (2010) A stochastic approach for generating spectrum compatible fully non-stationary earthquakes. *Comput Struct* 88(15–16):889–901
- Cacciola P (2011) Stochastic ground motion modelling for the seismic analysis of structures: a review. *Comput Technol Rev* 4:65–91
- Cacciola P, Deodatis G (2011) A method for generating fully non-stationary and spectrum-compatible ground motion vector processes. *Soil Dyn Earthq Eng* 31(3):351–360
- Cacciola P, Colajanni P, Muscolino G (2004) Combination of modal responses consistent with seismic input representation. *J Struct Eng ASCE* 130(1):47–55
- Conte JP, Peng BF (1997) Fully non-stationary analytical earthquake ground-motion model. *J Eng Mech* 123(1):15–24
- Corotis RB, Vanmarke EH, Cornell CA (1972) First passage of non-stationary random processes. *J Eng Mech Div* 98:401–414
- Davenport AG (1964) Note on the distribution of the largest value of a random function with application to gust loading. *Proc Inst Civil Eng* 28:187–196
- Deodatis G (1996) Non-stationary stochastic vector processes: seismic ground motion applications. *Probab Eng Mech* 11:149–168
- Der Kiureghian A, Neuenhofer A (1992) A response spectrum method for multiple-support seismic excitation. Report no UCB/EERC-91/08. Earthquake Engineering Research Center, University of California, Berkeley
- Di Paola M, Navarra G (2009) Stochastic seismic analysis of MDOF structures with nonlinear viscous dampers. *Struct Control Health Monit* 16:303–318
- Eurocode 8 (2010) Design of structures for earthquake resistance, part 1: general rules, seismic actions and rules for buildings; part 2: bridges: Comité Européen de Normalisation, Brussels
- Gasparini DA, Vanmarke EH (1976) Simulated earthquake motions compatible with prescribed response spectra. Evaluation of seismic safety of buildings, report number 2
- Giaralis A, Spanos PD (2009) Wavelets based response spectrum compatible synthesis of accelerograms-Eurocode application (EC8). *Soil Dyn Earthq Eng* 29:219–235
- Hao H, Oliveira CS, Penzien J (1989) Multiple-station ground motion processing and simulation based on SMART-1 array data. *Nucl Eng Des* 111:293–310
- Hsu TI, Bernard MC (1978) A random process for earthquake simulation. *Earthq Eng Struct Dyn* 6:347–362
- Iwan WD, Hou ZK (1989) Explicit solutions for the response of simple systems subjected to non stationary random excitation. *Struct Saf* 6:77–86
- Jennings PC, Housner GW, Tsai C (1969) Simulated earthquake motions for design purpose. In: Proceedings of the 4th world conference earthquake engineering, Santiago, A-1, pp 145–160
- Kaul MJ (1978) Stochastic characterization of earthquakes through their response spectrum. *Earthq Eng Struct Dyn* 6:497–509
- Mikealov G, Sarkani S, Lutes LD (1999) Spectral characteristics of non-stationary random processes- a critical review. *Struct Saf* 21:223–244
- Park YJ (1995) New conversion method from Response Spectrum to PSD functions. *J Eng Mech* 121(12):1391–1392
- Pfaffinger DD (1983) Calculation of power spectra from response spectra. *J Eng Mech* 109(1):357–372
- Preumont A (1980) A method for the generation of artificial earthquake accelerograms. *Nucl Eng Des* 59:357–368
- Preumont A (1985) The generation of non-separable artificial earthquake accelerograms for the design of nuclear power plants. *Nucl Eng Des* 88:59–67
- Priestley MB (1981) Spectral analysis and time series. Academic Press, London
- Rosenblueth E, Elorduy J (1969) Response of linear systems to certain transient disturbances. In: Proceedings 4th world conference earthquake engineering, Santiago, A-1, pp 185–196
- Sarkar K, Gupta VK (2005) Wavelet based simulation of spatially correlated and spectrum compatible accelerograms. In: Proceedings of the national symposium on structural dynamics, random vibration and earthquake engineering, Bangalore, pp 69–78
- Shinozuka M, Deodatis G (1988) Stochastic process models of earthquake ground motion. *J Probab Eng Mech* 3(3):114–123
- Shinozuka M, Sato Y (1967) Simulation of nonstationary random process. *J Eng Mech ASCE* 93:11–40
- Spanos PD, Vargas Loli LM (1985) A statistical approach to generation of design spectrum compatible earthquake time histories. *Soil Dyn Earthq Eng* 4(1):2–8
- Sundararajan C (1980) An iterative method for generation of seismic power spectral density functions. *Nucl Eng Des* 61:13–23
- Vanmarcke EH, Gasparini DA (1977) Simulated earthquake ground motions. In: Proceedings of the 4th international conference on Smirt, K1/9, San Francisco
- Vanmarcke EH (1972) Properties of spectral moments with applications to random vibration. *J Eng Mech Div ASCE* 98:425–446
- Vanmarcke EH (1976) Structural response to earthquake. In: Lomnitz C, Rosenblueth E (eds) Seismic risk and engineering decisions. Elsevier, New York
- Wang J, Fan L, Qian S, Zhou J (2002) Simulations of non-stationary frequency content and its importance to seismic assessment of structures. *Earthq Eng Struct Dyn* 31:993–1005
- Yeh CH, Wen YK (1990) Modelling of non-stationary ground motion and analysis of inelastic structural response. *Struct Saf* 8(1–4):281–298
- Zerva A (2009) Spatial variation of seismic ground motions: modelling and engineering applications. CRC Press/Taylor & Francis Group, Boca Raton

Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used

Ioanna Papayianni

Laboratory of Building Materials, Department of Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

Synonyms

Compatibility issues; Earthquake; Fiber reinforced polymers; Grouts; Masonry of heritage structure; Mortars; Preservation; Protection; Retrofitting; Selection

Introduction

Masonry heritage structures (MHS) constitute a countless number of different types of constructions dated from prehistory to present in which many “value contexts” (Lipe 1984) such as historic archeological, aesthetic, symbolic, social, cultural, scientific/technological, and economic are included making them a real treasury of human civilization. According to Burman (Burman 2001), their survival is essential to the spiritual, emotional, and economic well-being of humans.

The preservation of these structures concerns both life expectancy and protection from collapse occurred from earthquakes in seismic regions or from other natural or anthropogenic disasters. The former is closely related to the conservation/consolidation from decay phenomena due to the ageing effects of the environment impact on the buildings. The collapse is mostly attributed to the inherent inadequacy of historical masonry structural systems to bear horizontal loads.

The great diversity in the typology of MHS due to the various components, techniques of construction, morphology, type of reinforcement, and functionality makes the study of it, in terms of time and cost, difficult.

Moreover, internationally accepted Charters and Documents (Venice Chapter 1964) (Nara

Document on Authenticity 1994) recommend low invasive interventions for achieving a balance between keeping authentic characteristics and taking strengthening measures.

The peculiarity of the problems with MHS as well as some unforeseen failures after retrofitting in the past years led to the adoption of a step-by-step approach to strengthening, called observational method (Lowrenco 2006). By this way, a better compromise between traditional and innovative materials and techniques is achieved.

The society awareness about heritage structure preservation has globally increased internationally because of the higher recognition of the “values” associated with civilization of human genius and economy related to cultural tourism. The value of a monumental structure or area is increased after its preservation. Therefore, strategic policies of preserving built heritage have been promoted including: development of preventing seismic strengthening measures and establishing regulating frames and management systems. Technological advances in this field have fueled the market with many innovative materials and techniques or even new concepts of confronting seismic risk.

The Ancient Masonry Structures

The ancient masonry structures are composed of load bearing horizontal and vertical masonry elements which often are inadequately connected to resist seismic actions. Besides, the flooring systems and roofing forms (arches, vaults, domes) do not always provide enough diaphragm stiffening. These features, as well as the type of masonry morphology (one, two, or three leaf rubble masonry), predetermine their respond to loading (monotonic or cyclic) and the type of failure.

The choice of an appropriate and reliable analytical model for the study and assessment of a masonry structure seismic capacity prerequisites a thorough knowledge of its characteristics and behavior, as well as its pathology and degree of degradation.

This knowledge acquisition is achieved by surveying the MHS and applying on site

nondestructive measurements and other diagnostic tools of analysis at laboratory.

Masonry is a composite building material consisting of three discrete phases: units, joint mortar, and unit-mortar interface. The last is the weakest phase governing the behavior of masonry to horizontal loads.

As units may serve stone pieces shaped or irregularly cut, as well as mud bricks/earth blocks and fired bricks, or a combination of them. Successive courses of bricks are often found in stone masonry of city walls. It is also common, in large size piers or tower walls, a rubble material, or even mortar to have been used as infill.

The joint mortars are mainly lime-based mixtures of a binding system and aggregates with the following characteristics (Papayianni 2004):

- Apparent specific density 1.4–1.7
- High porosity 20–40 %
- Low strength (1–5 MPa)
- Low elastic modulus of elasticity (2–6 GPa)
- High deformability

They contain aggregates of small grains max size 4 or 8 mm not excluding greater sizes even pebbles of 16–30 mm, particularly in the cases of castle walls and churches with thick mortar joints of byzantine architecture. In earth block masonries of vernacular architecture, mud mortars are often used for jointing.

The strength capacity of the masonry depends on the strength of its components and their volume proportion in the masonry mass according to Eurocode 6, 1995. In some relevant equations found in literature (Tassios 1986), mortars' joint thickness is also taken into account for the estimation of masonry strength. Furthermore, the mortar strength is primarily influenced by the type of binding system.

Apart from the effect of masonry components on its strength, the type and morphology of the masonry (one or multiple leaf system) plays an important role to failure mechanism. For example, in three leaf masonry walls, separation of leaves is often observed (Binda et al. 2007). Moreover, if the unit-mortar interface has been weakened, shear or tensile type failure occurs easily.

In Table 1 some characteristics of often found components in ancient masonry are given for comparison with the modern materials used in construction (Papayianni and Tsolaki 1995). The structure of an ancient mortar and brick is shown in Fig. 1.

For the understanding of the behavior of an ancient masonry, the inside stress and stress-strain characteristics are measured on site by nondestructive test methods (NDT). Among the most widely used NDT are those of single and double flat jacks, sonic measurements, boroscopy, and active thermography.

Moreover, long-term monitoring of deformations by establishing a proper system of gauges is essential for the estimation of the existing stresses. In addition, the determination of mechanical, chemical, and physical characteristics of the materials by applying test methods at laboratory allows the evaluation of residual strength and elasticity of masonry, as well as the diagnosis of the decay and the degree of degradation. This holistic process of studying heritage structures is necessary because apart from cracks, a “softening” of the masonry mass may occur due to ageing or chemical attack (sulfate attack, efflorescence) that leads to significant strength decrease of both components of the masonry. For example, the strength of mud bricks is strongly influenced by their moisture content. Relevant to the effect of degradation to the strength of mortars and brick is shown in Table 2 (Papayianni and Tsolaki 1995).

Then, an analytical model for the masonry structure is adopted to study the respond of it to different scenarios of loading. The kind of applied loads, the values for strength and elasticity of masonry components, the degree of stiffness, and the safety factors have to be decided for a realistic estimation of the degree of retrofitting and selection of remedial measures. This is a job of great responsibility since there are not adequate regulative frames and every monumental structure is a particular case study (Penelis 1996).

Compatibility Issues

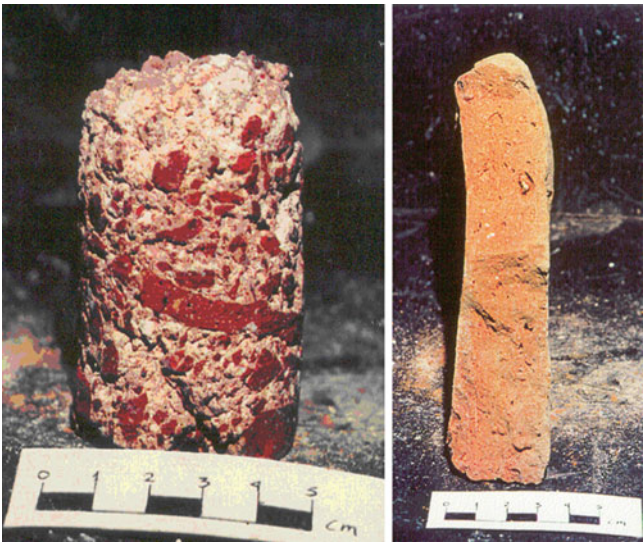
From experimental and analytical research works (Thomasen 2003), it is clear that the seismic

Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used, Table 1 Characteristics of components of ancient masonries

Monument	Historical period	Type of mortar	Binder	Lime content %	Ap. specific density	Strength in compression (MPa)	Brick plates (cm)	Absorption (%)	Strength in compression (MPa)
Galerius palace	Roman	Structural	Lime + pozzolan + soil	30	1.7	3.0–4.0	30 × 40 × 6	15.4	8.0–10.0
Acheiropoietos church	Middle Byzantine, seventh century	Structural	Lime + pozzolan + brick dust	35	1.55	3.0–3.5	30 × 40 × 5	17.5	5.0–9.0
Hagia Aikaterini Church	Late Byzantine thirteenth century	Structural	Lime + pozzolan + soil	20	1.65	2.0–2.5	30 × 30 × 4	14.5	10.0–12.0
Bezesteni	Ottoman, fifteenth century	Structural	Lime	40	1.70	1.0–1.5	30 × 30 × 4	20.0	10.0–14.0

Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used,

Fig. 1 Ancient mortar core and brick taken from Hagia Sophia in Thessaloniki (seventh century AD)



Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used, Table 2 Influence of the grade of deterioration on the bricks and mortars

Monument	Material	Absorption %	Salt content	Cracking	Strength in compression MPa
St. Panteleimon	Brick 1	14	Low	No	17.1
Fourteenth century AD	Brick 2	28	High	Yes	5.1
Hagia Aikaterini	Mortar 1	18	Medium	A few	2.5
Thirteenth century AD	Mortar 2	20	Medium	A few	1.9
Acheiropoietos	Mortar 1	20	Medium	No	3.3
Seventh century AD	Mortar 2	35	Medium	Yes	1.5

capacity of masonry is inherently low and anti-seismic protection demands explicitly strengthening. Moreover, MHS materials suffer very often from decay, and consolidation of the masonry mass must precede before any strengthening.

The dilemma is how much strengthening is appropriate for the historic structures, which do not conform to modern seismic design codes, addressing concrete constructions. Internationally accepted documents (Venice Charter 1964, ICOMOS, (Burras Charter 1981, 1988), Nara Document on Authenticity 1994), require respect to the “values” of monumental character structures, including architectural integrity, authenticity of the materials, and their morphology. What is recommended in retrofitting is to use compatible to authentic materials and traditional skills for their appliance.

There are many poor examples of interventions on monuments with concrete, which was supposed to cooperate as repair material with the old masonry. The destructive consequences from the use of incompatible to authentic matrix repair materials are mostly related to different physico-chemical characteristics of cement-based or polymeric matrix repair materials, which block the moisture movement, because of very low porosity and different porosity properties in comparison to old mortars. In addition, the aforementioned repair materials differ in elasticity from the existing old ones and consequently in deformability under the action of mechanical or hydrothermal loads.

However, for MHS in seismic regions, which are to be inhabited the value “safety” is of first priority. Therefore, strengthening methods are selected to be technical and cost-effective, as well as less invasive (Tassios and Mamillan 1985).

Homogenization and consolidation of the old masonry mass by inserting new repair materials is often applied with or without any reinforcement technique. It is worth mentioning how compatibility aspects could be interpreted in technical terms concerning their design (Papayianni 2004) (Van Balen et al. 2005). The most important characteristics or criteria by which compatibility is driven are:

- Surface features (color, texture, roughness)
- Composition (type of binder, granulometry of aggregate)
- Strength level (compressive, tensile)
- Elasticity (modulus of elasticity, deformability)
- Porosity (porosity, pore size distribution, capillarity)
- Coefficient of thermal dilation

These criteria have been applied in designing the repair mortars of a great number of interventions in Greece (Papayianni 2004), from 1990s to present, and proved very successful and long lasting. It must be pointed out that compatible repair material does not mean imitation of the authentic one. What is pursued is that having understood the functionality of the masonry, the repair material should not unsettle the long-term balance between local environment and structure and change its behavior. Based on experience from practice, it seems that by adopting mixed-type binding systems and suitable admixtures, the properties of repair materials may be adjusted to meet compatibility criteria.

Materials and Techniques for Retrofitting and Strengthening Masonry Heritage Structures

Introduction

The increasing awareness of the society about safeguarding heritage buildings promotes strategies of preventing their seismic retrofit. Cracks and other damages must firstly be repaired for reintegrating the continuity of the masonry corpus and then proceed to strengthening

alternatives. By this way, the resistance to earthquake and durability of MHS are globally improved. Of course, a thorough comprehension of the behavior of MHS must precede.

A solution of another concept to the problem is to reduce the seismic impact in the heritage structure by using a base isolation system (BIS) with supplementary viscous dampers at the base and the top of the building (Saito 2006). A successful example is the case of Los Angeles City Hall in which the BIS allows to remain stationary while the ground moves up to 500 mm in horizontal direction (Nabih 1996).

The methods of interventions are practically referred to materials and techniques. They are divided into irreversible methods, in which new materials are embodied in the masonry corpus and cannot be removed, and reversible ones, in which new materials are able to be activated on the surface or near the surface of the masonry undertaking local loads and can be replaced.

Irreversible Interventions

Irreversible interventions aim at consolidating the mass of the masonry, in particular when it suffers from softening due to intensive deterioration. The most common are:

- Grout injection
- Flowable mortar infills
- Deep repointing
- Reconstruction of masonry missing parts
- Punching steel bars (stainless steel) to saw cracks.

Of course, these interventions contribute to strengthening, but often additional techniques are applied to increase diaphragm stiffening and achieve adequate connection of horizontal and vertical structural elements of old masonry, aiming at increasing its seismic capacity. In the past, jacketing of piers with concrete, adding concrete beams as “chainage” or masonry confinement, has been widely used. They were characterized as invasive interventions altering the function and character of the old structure. Similarly, ferrocement surface coating or casting shotcrete over a steel mesh has been used for strengthening.

In irreversible interventions a major compromise between principles of restoration and safety and durability issues for the constructions is continuously under question. As declared in Nara Document (1994) “keeping the authenticity of heritage structures, materials and architecture is of great importance since all the values of a monumental or listed construction are transferred to future generation with materials.” However, as mentioned in the previous entry, in the case of inhabited historical buildings in urban areas which are to be preserved but also to be safe, authenticity issues are not of first priority.

Grouts and mortars are inserted into the masonry mass using different techniques to fill cracks, void in brick-mortar interfaces, and reduce the intrinsic heterogeneity and anisotropy of the masonry. After longtime and extensive research works and experience from field applications, some recommendations about designing grouts and mortars are found in literature (Tomazevic 1992). The conflicting aspects encountered in designing grout and mortar for repair of MHS make it an attractive research topic of paramount importance for the preservation of the heritage structures and their durability.

In most cases a tailored to the specific monumental structure grout/mortar mixture is proposed, taking into account the characteristics and pathology of the structure. Ready mixed grouts or mortars are not usually used since they do not often comply with specific requirements imposed by compatibility. Responsible for the design and application technique is the engineer who will select the binding system. This may be a combination of lime and pozzolan, with small quantities of cement, reactive silica, and admixtures, for rheology improvement, strength development, and volume stability. Commercial hydraulic lime is often used as an alternative binder. A series of tests are recommended for ensuring their quality (Toumbakari 2002).

Grouts

The technique of the consolidation with injection of grouts was at first well known in geotechnical and concrete engineering. Later, it was

extended for the homogenization of the mass of heritage structure. From the 1970s onward, grout injections have been used in many historical buildings and monumental structures. Grouted anchors are often used when tierods are inserted into masonry to improve tensile strength capacity. In 1990–2000, a number of research works and PhD thesis (Miltiadou 1990) (Valluzzi 2000) (Toumbakari 2002) have contributed much to the evolution of materials and techniques improving the quality and performance of the grouts.

The grouts are slurries of hydraulic nature binding system with or without inert fines and admixtures. The water/binder ratio is usually around 1.0. To be efficient, grouts must present in fresh state adequate fluidity and consistence, penetrability, and volume stability. In addition, after hardening they must develop good bonding within old masonry microstructure and mechanical strength.

To keep compatibility principles in designing grouts for MHS implies the selection of inorganic and relatively low strength potential binders, such as lime-based binding systems, which are the constituents of the most joint mortars of ancient masonries. An exception is the earth block masonries, in which the soil is the main binder for blocks and mortars and grouts should be soil-based mixtures.

In the first grouting intervention on MHS, portland cement was used as binder, but gradually the cement was diluted with lime or replaced by hydraulic lime or ternary binding system of the lime-pozzolan-cement plus some additives. Based on experimental work some researchers suggested (Toumbakari 2002) that the grouts injection results in significant strengthening if its binding system contains at least 30 % of mass portland cement.

An example of application with mixed-type binding system constitutes the consolidation of masonry element of the Byzantine Church Acheiropoietos (dated from the seventh century AD) by grouting in the 1990s with the following mixture. The choice of the binding system was based on the experimental study of materials and trial mixtures.

Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used, Table 3 Characteristics of lime-based grouts

Parts by weight					W/B	Fresh state properties			Compressive strength (MPa)	
Hydrated lime	Hydraulic lime	Natural pozzolan	White cement	Brick dust		Fluidity (sec)	Penetrability (sec)	Vol. stability (%)	28d	90d
1					1.06	10.8	3.15	1.5	–	0.69
	1				0.61	9.33	3.41	0.7	1.76	–
1		1			1.10	9.8	4.8	1.2	0.89	4.45
1		0.8	0.2		1.00	9.34	3.1	0.8	1.73	4.79
1		0.7	0.3		0.97	9.70	2.10	0.6	2.51	5.39
1		0.6		0.4	0.93	9.59	2.10	1.5	0.82	3.33
1		0.6	0.2	0.4	0.90	10.20	1.95	1.0	1.72	4.75

Composition of grout used in Acheiropoietos:

Hydrated lime: 40 %

White cement: 30 %

Ground pozzolan: 15 %

Brick dust: 15 %

Epoxy resin grouts which are widely used in filling cracks in concrete members are inappropriate in MHS grouting since the different nature of synthetic materials in terms of water and vapor impermeability makes moisture to be entrapped inside the MHS that is detrimental for old masonry. The ideal grout composition for MHS should present chemical and physical compatibility with the matrix of the MHS mortars like the transfused blood to human body.

The use of additives and admixtures in small quantities results in modifying grouts to develop the desirable properties. It should be pointed out that lime-based mixtures keep their fluidity for longer time than those of cement-based ones. This is a significant benefit for working in the field, especially in the summer period of Mediterranean climate.

Some lime-based grout mixtures with their characteristics in fresh and hardened state are shown in Table 3 (Papayianni and Pacht 2012).

In literature, there are recommendations (Papayianni 2004) about the selection of raw materials and methods of testing grout

performance and checking the filling effect after hardening. Some of them are mentioned below:

- The hydrated lime powder must be reactive and its reactivity could be checked by measuring $\text{Ca}(\text{OH})_2$ content by DTA-TG. Apart from lime hydroxide, hydraulic lime is also used in grouts because it hardens itself in humid environment and it is more convenient in the field work. However, there are differences between alternatives, concerning strength development and microstructure.
- Natural pozzolans are considerably activated if milled and contribute to earlier strength development, as well as to long-term strength.
- Sand or gravel of selected max size must present an even granulometry and be free of organic and soluble salts. Limestone dust or brick powder is also often added as fines.
- The salt content of all constituents of the grout mixture should be as low as possible. For example, efflorescence of sulfates is often observed due to sulfate content of brick powder. Even superplasticizers used to arrange fluidity must be free of sulfates.
- In case of cement addition it should be of low alkali and sulfate content.

Fluidity is commonly checked by Marsh cone (ASTM C939-87), while penetrability or injectability is tested by using the sand-column test (NORM NFP 18–891, 1986, EN 1771).

For volume stability the test method according to DIN 4227 Teil 5 standards is used. Furthermore, other more specific tests may be used, such as water retentivity and drying shrinkage, according to ASTM C1506 and ASTM C474, respectively.

Cracks, voids, and other discontinuities inside the mass are detected by nondestructive methods, such those of ultrasonic sonometer, radar, and IR thermography before and after the grout hardening (Tomazevic 1992). A draft assessment of achieved filling is made by this way.

Flowable Mortars

In case of cracks of high opening (2–10 cm), it is preferable to use a flowable mortar instead of grout, because the sand content of greater grain max size (0–8 mm) contributes to volume stability reducing shrinkage cracking. If necessary, stainless steel rods are often inserted transversally to the crack in order to saw the separate parts of masonry. The binders of flowable mixtures may consist of hydrated lime powder, pozzolan, and fines, such as limestone powder added by superplasticizers and viscosity modifiers. The composition of the flowable mortar must be based on the analysis of the existing old mortar. It has also to be tested for adequate flowability, even one hour after making the mixture, following the known test methods for fluidity and robustness, in relevant European guidelines 2005, such as flow table and L-box.

A type of flowable mortar has been applied in filling cracks of a part of wall masonry in Galerius ensemble, by gravity or low pressure. The composition and mechanical characteristics are given in Table 4 and Fig. 2.

Deep Repointing Mortar

Retrofitting of masonry joints with mortar is a common intervention work of restoration projects. Before designing the mortar’s composition, a systematic analysis of existing mortars is made to meet requirements for compatibility and a decision is made about strength demand and durability. The selected traditional type binding system, such as lime-pozzolan, may also include a small percentage of brick bust or cement and admixtures to meet the compatibility and

Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used, Table 4 Composition and strength of a flowable mortar applied in parts of Galerius ensemble

Raw materials	Parts by mass
Hydrated lime powder	1
Ground pozzolan	1
Limestone filler	0.7
Sand (0–4 mm)	6.5
Water/binder ratio	0.61
Superplasticizer (% by mass of binder)	2.4
Retarder (% by mass of binder)	0.3
Porosity	20.8 %
Flexural strength (specimens: 4 × 4 × 16 cm)	2.65 MPa
Compressive strength (specimens: 4 × 4 × 16 cm)	6.5 MPa

durability requirement. Apart from binders, coarse aggregates are often used especially in thick mortar joints. Some examples of these mortars used in interventions are shown in Table 5.

Reversible Interventions

In general, they are considered low invasive and effective in terms of expectancy and resistance to environmental conditions. In most of this type of interventions, the materials are industrially manufactured, standardized, and commercially known with brand names. They are usually applied with specific technique. The problems presented in practice are rather related to technique than to materials.

An analytical assessment of how the MHS responds to earthquake is necessary, since the positions and the degree of retrofitting must be known before any type of reversible intervention.

A long list of this type of interventions could be mentioned since there are many technological advances in this field. Some of them are:

- Pre-stressed (stainless) steel cables offered many solutions in the last decades of previous century (Fig. 3; Ignatakis and Stylianidis 1989).
- Stainless steel bars or strips into bed joints repointing with lime-based repair mortars.

Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used,

Fig. 2 Galerius palace ensemble, third century AD. Cracks filled with flowable lime-based mortar



Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used, Table 5 Proposal of deep repointing of Byzantine monuments

Monument: Saint Nicholaos Church, Aiges, Sparti (Post Byzantine period)

Analysis of mortar samples

Color: grayish white

Stratigraphy: One layer

Stereoscopic Observation: Compact paste with micropores, cracks with re-crystallized secondary material (photo) with concentration of calcite (photo). High content in aggregates, max size up to 8–10 mm, of mixed-type siliceous and calcitic origin

Compressive strength: 14.1 kg/cm²

Porosity: 19.15 %

Specific gravity: 1.74 g/cm³

Chemical composition %:

Na ₂ O	K ₂ O	CaO	MgO	Fe ₂ O ₃	Al ₂ O ₃	SiO ₂	L.I.
0.36	0.33	31.5	6.18	1.04	3.29	27.7	30.0
Content in anions:			Cl ⁻	NO ₃ ⁻	SO ₄ ²⁻		
			0.17	0.12	0.19		

Insoluble residue: 48.8 % by mass

Gradation curve: Max size of grains: 8 mm

Proposal of repair mortar

Constituents	Parts by weight		
Hydrated lime	1		
Pozzolan ground (milaiki gaia) (10 % residue on 45 µm sieve)	1		
River sand (0–1 mm)	0.2		
Local limestone crushed (2–4 mm)	1.6		
Coarse limestone (4–16 mm)	0.2		
Superplasticizer (free of sulfates) 1 % by mass of the binder			
Water required for 15 ± 1 cm expansion on the flow table			
The trial mixes showed that the achieved 28-days compressive strength ranges from 19 to 25 kg/cm ² and the porosity from 17 % to 20 %.			



Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used, Fig. 3 St Panteleimon Church, Thessaloniki. Substitution of wooden tierods

with stainless steel ones (Photograph courtesy of Prof. Ch. Ignatakis and Prof. K. Stylianidis)

Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used, Fig. 4 St Panteleimon Church, Thessaloniki. Metallic braces for confinements of small part of pier above the column capital (Photograph courtesy of Prof. Ch. Ignatakis and Prof. K. Stylianidis)



R

Bars or strips could also be made of carbon fiber reinforced polymer (CFRP) covered with lime-based mortars (Modena and Valluzzi 2003).

- Metallic braces anchored on the masonry for confinement of the out of plane failure deformations (Fig. 4; Ignatakis and Stylianidis 1989).
- Bonded to surface fiber reinforcement with polymer or cementitious matrices (FRP or FRG).

There is much progress in the materials and techniques of anchoring for FRP/FRG textiles, resulting in higher performance and cost-

effectiveness, but still these types of interventions require special and expensive work. There are different types of advanced fiber reinforced sheets or grids of polymeric (FRP) or cement-based (FRG) matrix bonded on the surface of masonry elements. The use of this type of retrofit has impressively increased and continuously evolved. For example, tie bars are placed on the under strengthening areas of masonry with Shape Memory Alloy (SMA) devices by which recover from strain and relief of stress are achieved (Desroches and Smith 2004).

Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used, Table 6 Characteristics of fibers of FRP(s)

Type of fibers	Glass	Carbon	Aramid	Steel
Tensile strength (MPa)	1,300–3,400	2,000–5,600	2,500–3,620	1,500–3,500
Modulus of elasticity (GPa)	22–62	150–325	48–76	185–210
Ultimate strain (mm/mm)	0.03–0.05	0.01–0.015	0.02–0.036	0.04
Coefficient of thermal dilation (10^{-6} m/m/°K)	5.5	0.0	–0.5	6.5
Density (g/cm ³)	2.5–2.6	1.7	1.4	7.9
Melting point (°C)	1,100	310	420	1,300

Source: Based on ACI Manual of Concrete Practice ACI 544 1R-96 (2002)



Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used, Fig. 5 Commercial steel and carbon fiber polymer sheets (Photograph courtesy of G. Mitolidis)

Bonded to Surface- Reinforcement Techniques

These low invasive alternatives have started to be used in construction from the 1990s. They consist of a fibric system and a matrix system bonded in different forms, such as pultruded bars, sheet, or textiles and grids. The most known fibers are steel, carbon, aramid, glass and alkali-resisting glass (AR glass), basalt and biocomposites (flax, hemp), or recycled biopolymers based on lignin, cellulose, and pectin. Some characteristics of fibers are found in the literature in Table 6 (Meier 1995). Carbon fiber reinforced polymer sheets (CFRP) and steel fiber reinforced polymer sheets (SFRP) are the most widely used products in reinforcing constructions (Fig. 5). They

undertake tensile stresses developed in parallel to fiber direction. The matrix in which fibers are embodied unidirectionally or bidirectionally is polymer, such as epoxy resins, polyester, and vinyl ester. In cement-based textiles or grids, a polymer-modified cement mortar with 1 % by volume long chopped AR glass fiber is used. Fibers could be categorized in three classes (Casadei and Agneloni 2009; Mitolidis 2009):

- Of high modulus of elasticity, such as different qualities of steel and carbon
- Of medium modulus of elasticity such as aramid, basalt, and glass
- Of low modulus of elasticity such as Flax and Hemp

Indicative values for the characteristics of the epoxy resin matrices are:

- Density 1.1–1.7
- Modulus of elasticity 0.5–20GPa
- Flexural strength 9–30 MPa
- Ultimate strain 0.5–5 %

The main type of failure of surface bonded CFRP(s) concerns the development of considerable shear stresses on the surface leading to detachment. That is why the anchorage length must be estimated for each type of matrices and be based on reliable measurements of bonding.

The procedure of applying FRP includes:

- Check the bonding capacity of masonry substrate.
- Leveling the surface to succeed good bonding that is essential for transferring stresses in particular in the case of not strong anchorage.
- Apply FRP sheets on dry surface with good quality of adhesive material.
- Ensure sufficient anchorage. Metallic haunch connections are preferable for low invasive FRP.
- Use of proper dye for protection from UV radiation.

More details are available in literature (ACI 440 2R-08, 2008).

When retrofitting masonry walls or strengthening arches and domes, the values of practical interest in relation to FRP are:

- Ultimate tensile strength (i.e., 5–6 MPa)
- Ultimate elongation (i.e., 1.62 %)
- Transition zone tensile stress (i.e., 2.1 MPa)
- Tensile modulus of elasticity (i.e., 0.37GPa)

As advantages of FRP are considered the following:

- The low weight or high strength to weight ratio
- The relatively easy installation and high productivity of the work
- The minimal dimensional change of the masonry elements

- The resistance to corrosion, chemical attack, and environmental impact

However, many disadvantages have also been noted mainly because of the polymer nature of the matrix, such as:

- Low resistance to fire
- Degradation of polymeric matrix by UV
- Application needs special equipment and specialized staff as well as dry surfaces and low temperatures
- The incompatibility of resins to masonry structure has an impact to the bond of FRP on masonry surface. Moreover, FRP may create secondary side effects that make the benefits of reversibility questionable. This occurs because of the lack of vapor permeability that leads to moisture accumulation inside the masonry mass which favors deterioration phenomena, particularly in the case of large masonry area rapping with FRP. Furthermore, FRP or AR fiber glass textiles exhibit brittle failure mechanism when their bonding is based on adhesion and friction. In fiber grids with cement-based matrix, these effects are mitigated.

Some outstanding examples of MHS strengthening with FRP mentioned in literature, such as the strengthening of the San Vitale Church in Parma, Italy, and the repair of the monastery of St Andreas, Mount Athos, 2011 (Figs. 6 and 7).

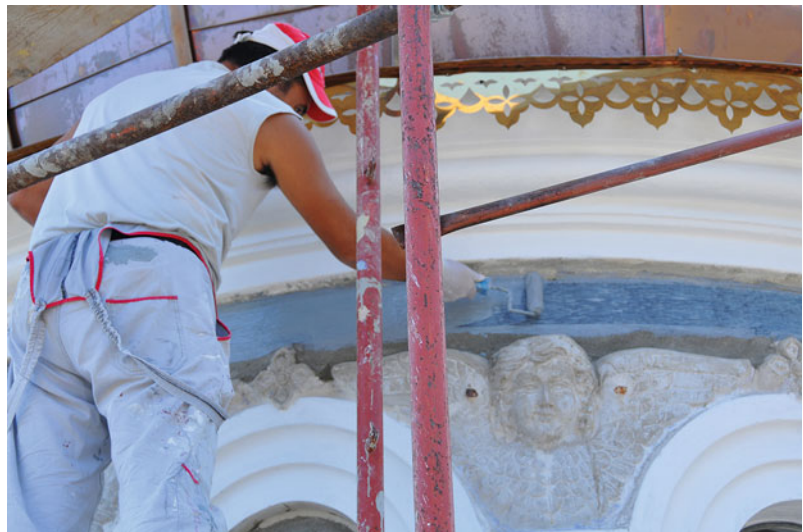
An interesting evolution in the field of reversible external intervention is SMA devices, which are systems of advanced technology applied in MHS for defending earthquake. The main materials of SMA are alloys of NiTi, which recover from large strain through heat action. They are considered superelastic and high damping systems so as by the removal of stress a recovery from strain occurs. The high cost, the difficulty in handling the material because of its hardness, and the welding deficiency are mentioned as disadvantages, as well as the dependency of the SMA properties from the temperature and their sensitivity to compositional changes of alloys. (MANSIDE Consortium 1998).

An example of SMA application, mentioned in literature, is the first application of these



Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used, Fig. 6 The damages of the dome of the monastery at Andreas in Mount Athos before repair intervention (Photograph courtesy of G. Mitolidis)

Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used, Fig. 7 The use of CFRP for strengthening the base of the dome of St Andreas in Mount Athos, 2011 (Photograph courtesy of G. Mitolidis)



devices in S. Giorgio Church in Trigano, Italy, by Prof. Indirli (2001) (Indirli et al. 2001) and reinforcing of St. Francisco of Assisi Church by Professors Groci and Gastellano (Groci 2001).

Summary

Masonry Heritage Structures (MHS) have inherent weakness in carrying horizontal loads, and their preservation and protection in seismic areas require retrofitting and strengthening interventions. However, principles of philosophy of conservation about compatibility and keeping authenticity should be followed in parallel to safety and durability issues valid in any type of construction inhabited, according to

modern codes and regulations for anti-seismic design of structures.

The nature of building materials of old masonries which are porous and of low strength capacity is different from modern compact cement-based building materials (reinforced with steel bars) and of low porosity, high bearing capacity in compression, and bending and tensile loads. Similarly, materials of organic polymer-based matrix are impermeable and hardly cooperating with inorganic matrix materials. These contradictions introduce particular difficulties in retrofitting and strengthening heritage masonries enhanced by the fact of the great variety of masonry types, in terms of their components, way of constructing, functionalism in the structural ensemble, and various loading conditions.

After earthquakes of the last three decades of the twentieth century in Italy and Greece, countries with a tremendous heritage building asset, the MHS “de facto” separation in historical masonry buildings of urban areas and monumental buildings was followed for emergency repairs. This temporary separation allowed more freedom in compromising safety and compatibility issues in the case of historical buildings of old urban centers, while compatibility and authenticity issues prevailed in the case of monumental buildings. This option was also suggested by distinguished scientists (Tassios and Mamillan 1985), because it helps decision makers about the degree of strengthening.

Addressing properly the aforementioned issues in lifesaving interventions on MHS pre-requisites the consideration of a holistic approach of analysis of existing structures. This involves using measurements on site and at laboratory, as well as assessment of MHS behavior by adequate models. In this manner, a thorough understanding of the structure can be obtained before deciding about repair materials and techniques.

In the past, many failures occurred concerning mainly side effects due to invasive retrofitting interventions on the MHS ensembles. The current evolution in this field focuses on the development of compatible grouts, flowable mortars, and repointing mortars for irreversible type consolidation of the mass of the masonry corpus. Besides, steel and other fiber reinforcement bonded on the surface of masonry with materials of polymeric- or cement-based matrix are used as reversible interventions.

Advances in the field of irreversible interventions could be considered the use of upgraded lime-based grouts and mortars, i.e., with addition of new generation admixtures prolonging fluidity and increasing volume stability, or by using nano-modified slurries for the enhancement of bonding of interfaces between old and new materials.

Furthermore, new types of reinforcement are expected, which would be light and resistant to environmental influences, providing a high deformability capacity with strong and yet less harmful anchorage, and meet aesthetic and compatibility requirements are some expectations in reversible strengthening of HMS.

Moreover, the quality of materials and intervention works is of primary importance for the longevity and economy of interventions. However, the system of regulations, recommendations, proper test methods, and specific standards has not been fully developed, in spite of continuous efforts of the last decade on national and European level. There is still a gap between knowledge and practice, but undoubtedly the continuous education and joint research efforts, as well as the awareness of the society about safeguarding cultural patrimony, have resulted in significant improvement of consolidation and strengthening works.

Cross-References

- ▶ [Ancient Monuments Under Seismic Actions: Modeling and Analysis](#)
- ▶ [Damage to Ancient Buildings from Earthquakes](#)
- ▶ [Masonry Components](#)
- ▶ [Seismic Behavior of Ancient Monuments: from Collapse Observation to Permanent Monitoring](#)
- ▶ [Seismic Strengthening Strategies for Heritage Structures](#)
- ▶ [Strengthening Techniques: Masonry and Heritage Structures](#)

References

- Binda L, Gardani G, Cantini L, Tirabisi C (2007) On site and laboratory detection of the quality of masonry in historic buildings. In: Arun G (ed) Proceeding of the international symposium studies on historical heritage, Antalya Turkey, pp 667–682
- Burman PATI (2001) What is cultural heritage? In: Baer NS, Snickars F (eds) Rational decision-making in the presentation of cultural property. Dahlem University Press, Berlin
- Casadei P, Agneloni E, (2009) Advanced composite materials for strengthening, blast protection and seismic protection. EC-COST ACTION C26 urban habit constructions under catastrophic events. <http://www.tecinn.com>
- Desroches R, Smith B (2004) Shape memory alloys in seismic resistant design and retrofit: a critical review of their potential and limitations. *J Earthq Eng* 3(8):415–429

- Groci G (2001) Strengthening of the Basilica of St. Francis of Assisi after the September 1997 Earthquake. *Struct Eng Int* 11:207–210
- Ignatakis C, Stylianidis K (1989) Criteria for the substitution of wooden tierods in Byzantine Churches—St. Panteleimon Church of Thessaloniki. In: Brebbia CA (ed) *Proceedings of the 1st international conference on structural studies, repairs and maintenance of historical buildings*, Florence, pp 385–393
- Indirli M, Castellano M, Clemente P, Martelli A (2001) Demo application of SMA devices. The rehabilitation of the St. Giorgio Church bell tower. In: *Proceedings of the SPIE, Smart Systems for Bridges, Structures and Highway*, New port Beach, California (US) March, 2001
- Lipe WD (1984) Value and meaning in cultural resources. In: Cleve H (ed) *Approaches to the archaeological heritage: a comparative study of world cultural resource management system*. Cambridge University Press, Cambridge
- Lowrenco PB (2006) Structural restoration of monuments: recommendations and advances in research and practice. In: *Proceedings of the 1st international conference restoration of heritage masonry structures*, Minho Portugal, KP05 01–16
- MANSIDE Consortium (1998) Guidelines for the passive seismic protection of structures using shape alloy based devices. Report of the Brite/Euram, Bruxelles project BE 95–2168
- Meier U (1995) Strengthening of structures using carbon fibre/epoxy composites. *Constr Build Mater* 6(9):341–351. Elsevier Science, London
- Miltiadiou A (1990) Contribution a l' etude des coulis hydrauliques pour la reparation et le renforcement des structures et des monuments historique en masonnerie. PhD thesis, ENPC, Paris
- Mitolidis G (2009) Experimental investigation of concrete beams behaviour with carbon or steel fibre reinforced polymers. PhD thesis, AUTH, Thessaloniki
- Modena C, Valluzzi MR (2003) Repair techniques for creep and long- term damage of massive structures. In: Brebbia CA (ed) *Proceedings of the conference structural studies, repairs and maintenance of heritage architecture VII*. WIT Press, pp 141–150
- Nabih Y (1996) Application on hybrid damping system with base isolation for the seismic retrofit of Los Angeles City Hall. www.cctk.ac.in/nicee/wcee/article/n_2090.pdf
- Papayianni I (2004) Design of compatible repair materials for the restoration of monuments. *Int J Restor* 10(6):623–636
- Papayianni I, Pacht V (2012) High performance lime based grouts for repair of historic masonries. In: Jasienko J (ed) *Proceedings of the SAHC 2012 conference structural analysis of historical constructions*, pp 1897–1904
- Papayianni I, Tsolaki A (1995) Microstructural and mechanical strength of ancient mortars. In: Papayianni I, Astrinidou P (eds) *Proceedings of the NATO SFS workshop on materials for restoration*, Thessaloniki Greece, pp 35–48
- Penelis G (1996) Techniques and materials for structural restoration. In: *Proceedings of the 14th WCEE*, Acapulco
- Saito T (2006) Basic concept of base isolation system for buildings. Japan Society of Seismic Isolation. <http://www.cibw114.net/report/Japan02.pdf>
- Tassios TP (1986) *Mechanics of masonry*. Edition of National Technical University of Athens, Athens, pp 6–13 (in greek and italian)
- Tassios TP, Mamillan M (1985) Structural investigation of ancient monuments. *Valutazione Strutturale dei Monumenti antichi*. Ed. Kappa, Rome
- Thomasen SE (2003) Performance- based seismic design criteria for historic buildings. In: Brebbia CA (ed) *Structural study, repairs and maintenance of heritage architecture VII*. WIT Press, Southampton, pp 507–512
- Tomazevic M (1992) Laboratory and in sites test of the efficacy of grouting and tying of stone masonry. In: Binda L (ed) *Proceedings of the international work CNR-ENDT effectiveness of injection techniques of retrofitting of stone and brick masonry walls in seismic areas*, Milan, pp 95–117
- Toumbakari E (2002) Lime-pozzolan-cement grouts and their structural effects on composite masonry walls. PhD thesis, K.U. Leuven
- Valluzzi MR (2000) Mechanical behaviour of masonry strengthened by lime-based materials and techniques. PhD thesis, University of Padua
- Van Balen K, Papayianni I, Van Hess R, Binda L, Waldum A (2005) Introduction to requirements for functions and properties of repair mortars. *Mater Struct/Materiaux et Constr* 38(282):781–785

Retrofitting and Strengthening Measures: Liability and Quality Assurance

George Gr. Penelis

Civil Engineering Department, Aristotle

University of Thessaloniki, Thessaloniki, Greece

Synonyms

Emergency measures; Intervention; Propping; Quality control; Quality management; Quality plan; Rehabilitation; Repair; Restoration; Risk; Shoring; Supervision

Introduction

Buildings or civil engineering works in seismic regions constitute very large-scale property investments which affect the lives of a great number of people.

The original construction and any later retrofitting and strengthening for seismic protection require **the coordination** of many people, including investors, engineers, foremen, technicians, and laborers, in a large, organized effort. The number of people participating in the construction of the works or of the interventions and the number of subsequent users are sources of risk for people and goods due to failures during construction or during the life-time of the works. For the above reasons, very early in human history, organized societies incorporated into their **penal** and **civil** codes articles against the violation of rules of construction, providing for **penalties** and **liabilities** and aiming in this way to reduce the incidence and effects of failures on goods and people. It is interesting to note the reference to this issue in the Code of Hammurabi (1772 BCE) (King 1915) (Fig. 1).

At the same time, rules for building procedures have been established, accompanied by quality control systems. All these measures constituting what we call today **quality assurance** have on one hand aimed at reducing probabilities of failure and on the other at creating a framework of checkpoints for the objective determination and dissemination of liabilities to the personnel involved in the construction of buildings and public works.

For a clear approach to issues of quality assurance and liability in regard to retrofitting and strengthening of buildings or civil engineering works, it is necessary to clarify first the various actions of **assessment**, **design**, and **construction** in retrofitting where mistakes may occur and therefore where liabilities and penalties may arise.

Taking all the above into account, the following structure will be adopted in the following sections:

- Steps of procedure for seismic retrofitting and strengthening of buildings and civil works
- Quality assurance mechanisms applied at each step of the retrofitting procedure
- Liabilities emerging in case of defects or failures

Retrofitting and Strengthening Procedures

Seismic Assessment

General

The seismic assessment of existing buildings has as its main objective making decisions about the need for retrofitting each individual building. In this respect seismic assessment always precedes retrofitting activities. Keeping in mind that the retrofitting cost (structural–nonstructural) is very high, ranging around a mean value of 10–12 % of the original cost of construction, and that the number of buildings or civil engineering works involved in a mitigation program is usually very large, it is apparent that an evaluation procedure must be established **for screening** the buildings or works under consideration in successive steps of accuracy, until the number of buildings or other works (e.g., bridges) chosen for eventual intervention has been radically reduced.

National programs for seismic risk mitigation through retrofitting must be classified as either **post-earthquake** or **pre-earthquake**. The first category is triggered by a strong earthquake and the damage caused, while the second category is triggered by “**political decisions**” for seismic risk reduction for a special category of works such as schools, hotels, hospitals, etc. It is apparent that the first category is the most usual, as it is obligatory for a government to activate such a program after a disastrous earthquake. No matter which one of these two categories is followed, the flowchart of the whole procedure through rehabilitation is given in Fig. 2 (Penelis and Penelis 2014).



229. If a builder build a house for some one, and does not construct it properly, and the house which he built fall in and kill its owner, then that builder shall be put to death.

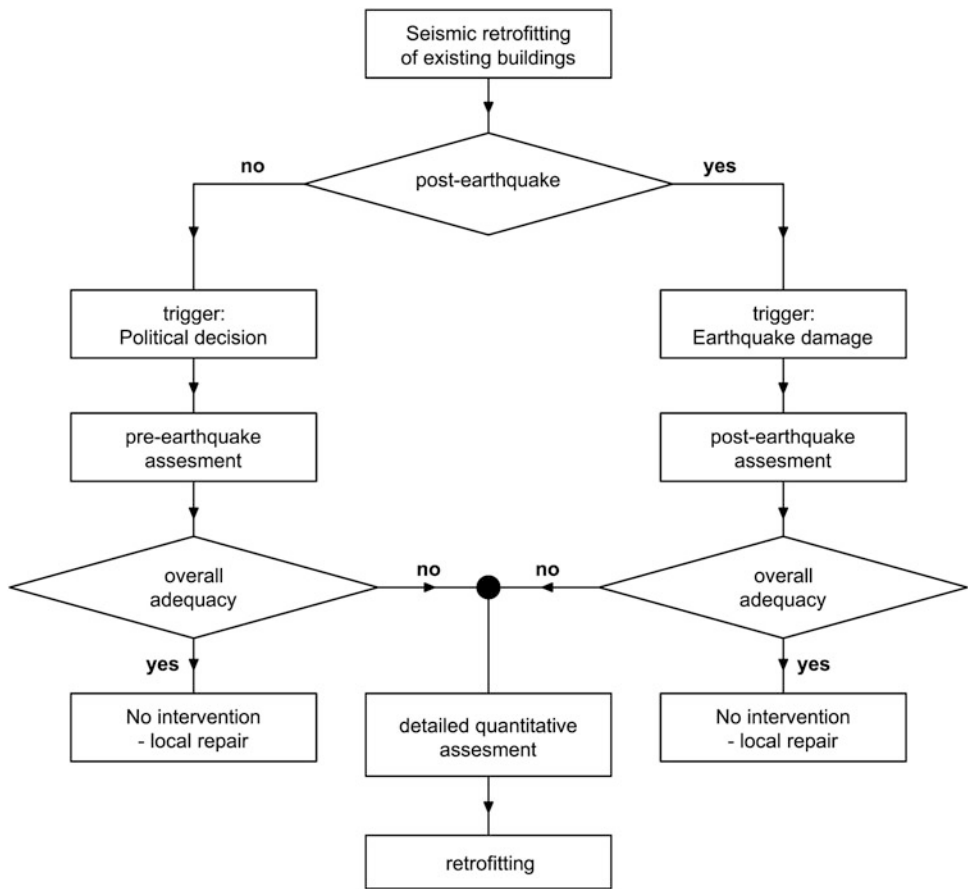
230. If it kill the son of the owner the son of that builder shall be put to death.

231. If it kill a slave of the owner, then he shall pay slave for slave to the owner of the house.

232. If it ruin goods, he shall make compensation for all that has been ruined, and inasmuch as he did not construct properly this house which he built and it fell, he shall re-erect the house from his own means.

233. If a builder build a house for some one, even though he has not yet completed it; if then the walls seem toppling, the builder must make the walls solid from his own means.

Retrofitting and Strengthening Measures: Liability and Quality Assurance, Fig. 1 The Code of Hammurabi: reference to building defects liabilities and penalties (King 1915)



Retrofitting and Strengthening Measures: Liability and Quality Assurance, Fig. 2 Flowchart for pre- and post-earthquake assessment and retrofitting (Adapted from Penelis and Penelis (2014), C.R.C. Press)

Post-earthquake Assessment

General Post-earthquake assessment activities must be classified into two main levels:

- Emergency post-earthquake damage inspection
- Post-earthquake seismic evaluation

The first category of inspections is performed by the State. During this procedure, damage characterization for each independent building and designation of its damage level are carried out. At the same time, emergency propping activities and demolitions are executed for direct risk reduction due to aftershocks.

The second category is performed by the owners of each individual building: During this

procedure the residual resistance of every individual damaged structure is assessed and the degree of intervention is decided, that is, **repair** or **strengthening** of the structure.

Emergency Post-earthquake Damage Inspection A strong earthquake puts on trial not only citizens but the State itself. The authorities have to face chaotic situations due to lack of information, the extent of the affected region, and the multiple requests for assistance and for inspections of damaged buildings. The foregoing remarks aim at describing the environment in which the structural engineer is called upon to do an emergency post-earthquake damage inspection.

This should be the prevailing element for the organization of the earlier operation. Indeed,

since damage evaluation sometimes involves thousands of buildings that have to be assessed in a short period of time by hundreds of engineers, a special procedure has to be followed, completely different from that used for the detailed evaluation of a single building before retrofitting and even from that used in pre-earthquake assessment programs.

Therefore, after a strong earthquake, only a **damage-oriented evaluation procedure** may be implemented. In this case structural engineers come face to face with buildings or public works struck by the earthquake, without being able to use the scientific tools that they possess, namely, **in situ measurements, laboratory tests, and the analysis**. They are compelled by circumstances to restrict themselves to qualitative evaluations and make decisions based solely on **visual observation**, using at the same time their **good engineering judgment**.

In any case the engineer can make one of the following decisions after this first contact with the damaged structure:

- (a) Allow use of the building without restrictions, designating it as **“green.”**
- (b) Classify it as temporarily unusable with limited access due to local damage, designating it as **“yellow.”**
- (c) Classify it as out of use because of extensive damage, designating it as **“red.”**
- (d) Classify it as near collapse and activate proping measures together with restrictions on approaching the area around the building, or even demolitions.

The assessment procedure is depicted in the flowchart in Fig. 3 (Penelis and Kappos 1997). From the above diagram it can be seen that the whole procedure is articulated at three levels related to the seriousness of damage and is therefore aimed at reducing the risk of assessment errors.

The procedure described above, being to a great degree subjective, requires special concern for the formation of a proper quality assurance system. At the same time this subjectivity should be taken seriously into account for

liability issues. All these will be examined in detail in the sections that follow.

However, it should be noted that, as may be observed in the case of Thessaloniki (earthquake of June 20, 1978), the results of emergency inspection should be characterized as rather reliable, since the repair or strengthening that followed, which was based on a detailed post-earthquake assessment, involved (Penelis et al. 1987):

- 11.5 % of the number of “green” buildings
- 45.2 % of the number of “yellow” buildings
- 81.9 % of the number of “red” buildings

Post-earthquake Assessment The main criterion for a post-earthquake assessment and the decision on the degree of retrofitting is the extent of damage and its seriousness. This approach radically decreases the number of interventions and, therefore, the retrofitting cost of an affected region and the time required for the whole operation.

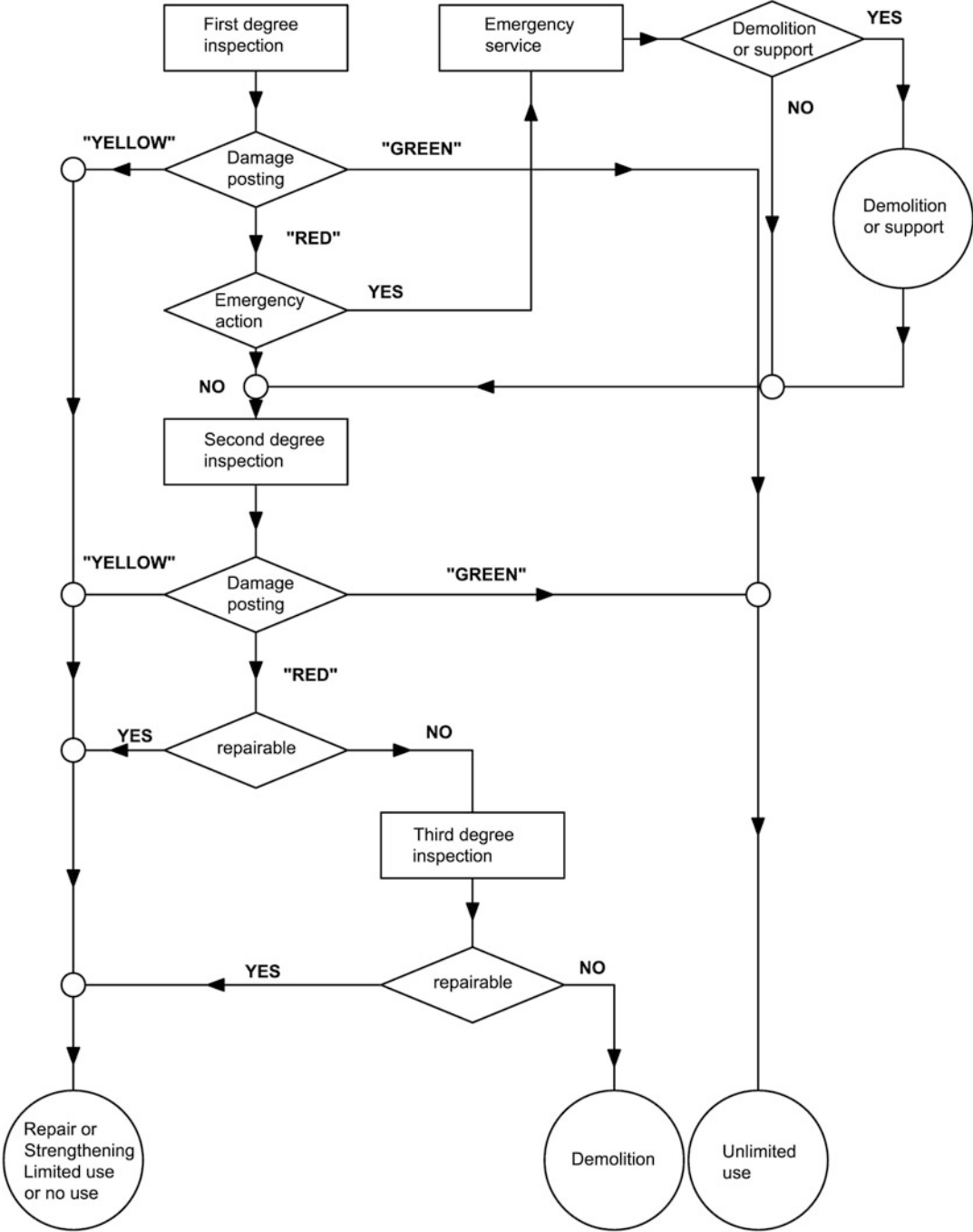
Based on the above, the approach to the intervention procedure may be stipulated as follows:

- In structures with light damage of local nature, intervention should be limited to **repair** (Fig. 4).
- In structures with extensive or heavy damage, global-type intervention should extend to **strengthening** of the structure.

It is obvious that only buildings or other types of structures posted as “yellow” or “red” are evaluated. The evaluation is a damage-oriented quantitative procedure, and it is based mainly on the ratio of the residual seismic resistance V_D of the damaged structure to its capacity V_c before damage (Fig. 4).

The result of this assessment is a decision either to repair (local damage) or strengthen (global-type damage).

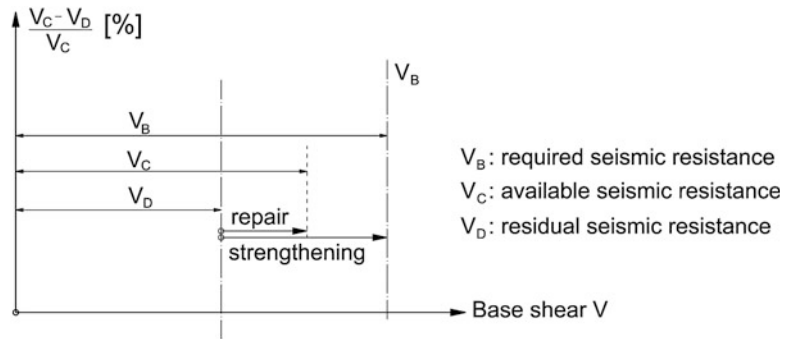
Strengthening is carried out for buildings following the procedure specified by FEMA 356, 2002/ASCE SEI 41–06, or EC8-3/2005 (see next paragraph, tier 3).



Retrofitting and Strengthening Measures: Liability and Quality Assurance, Fig. 3 Flowchart of emergency inspection procedure in Thessaloniki 1978 (From Penelis and Kappos (1997) with permission of C.R.C. Press)

Retrofitting and Strengthening Measures: Liability and Quality Assurance, Fig. 4

Schematic presentation of the difference between “repair” and “strengthening” (From Penelis and Kappos (1997), with permission of C.R.C. Press)



Pre-earthquake Assessment

Keeping in mind that the majority of the buildings in use do not comply with modern seismic codes and therefore present various levels of vulnerability, it has been realized, particularly in the USA since the early 1980s, that a drastic reduction of seismic risk could be achieved only by an “active” intervention in the existing building stock. Thus, a framework of procedures for seismic assessment of buildings has been developed in various countries, with that in the USA being the soundest (FEMA 310/1998/ASCE 31–02).

In most cases this procedure is articulated in three successive steps of higher detail and accuracy from the first step to the third.

The FEMA 310/1998/ASCE 31–02 procedure is presented in the flowchart in Fig. 5.

The above standard provides for a **three-tiered procedure** aiming at the localization of buildings that might be hazardous after a strong earthquake and for the **life safety** of buildings intended to be used for **immediate occupancy**. These **three tiers** are the following:

- Screening phase (Tier 1)
- Evaluation phase (Tier 2)
- Detailed evaluation phase (Tier 3)

Screening phase, Tier 1, allows a rapid seismic evaluation of the building. The level of analysis required at this step is minimal.

Tier 2 includes a complete but simplified linear analysis of the building or any similar structure.

Tier 3 includes a detailed seismic evaluation and must be assigned to the detailed retrofitting

procedure (FEMA 356, 2002/ASCE SEI 41–06, or EC8-3/2005).

Detailed Seismic Assessment and Retrofitting

General

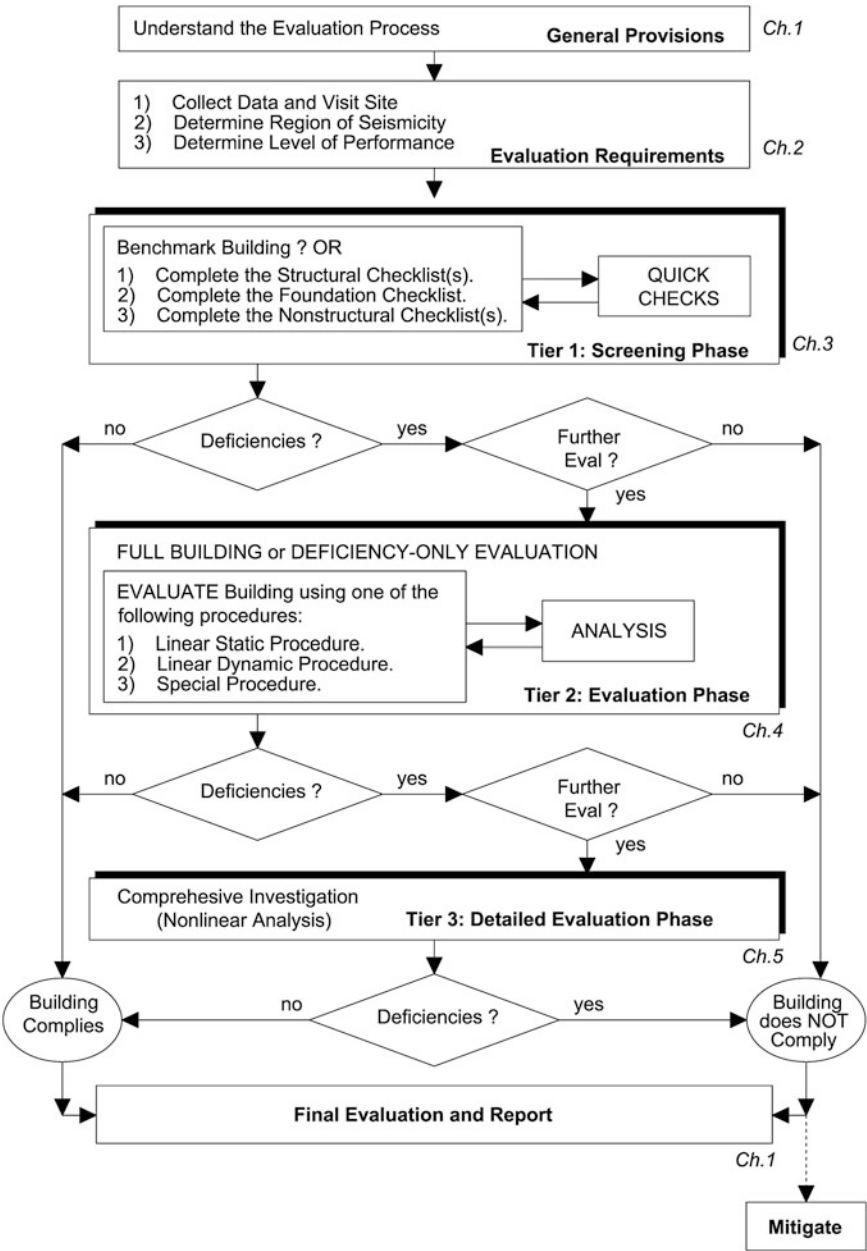
As clarified in section “[Seismic Assessment](#),” detailed seismic assessment and retrofitting of buildings or similar works is activated after a seismic evaluation procedure has been completed and structures designated for detailed investigation and retrofitting have been identified.

Whether or not the designation of these buildings is the result of a pre-earthquake or a post-earthquake procedure, the method for this detailed assessment and retrofitting is the same, and for the time being, this is specified by FEMA 356, 2002/ASCE SEI 41–06 in the USA and EC8-3/2005 in the EU. It should be noted that any retrofitting activity aims at raising the seismic performance of an existing structure to a level similar to that of new structures designed and constructed in accordance with contemporary seismic codes.

Detailed Assessment

The task of detailed assessment is the quantitative determination of the existing global safety factor of the structure and at the same time the existing local safety factors in all critical sections of the structural elements.

In this context, on one hand the need for eventual retrofitting is verified, and on the other these results constitute a very good guide for the retrofitting scheme that will be adopted for the strengthening of the existing structure.



Retrofitting and Strengthening Measures: Liability and Quality Assurance, Fig. 5 Evaluation process according to FEMA-310/1998/ASCE31-02 (From FEMA-310/1998/ASCE 31-02 Handbook for the Seismic Evaluation of Buildings, Federal Emergency Management Agency, Washington D.C. with permission)

Bearing in mind that the quantification of safety issues requires the existence of reliable information on the structure, the foundation ground, and the seismic actions that will be used as an input in well-defined analytical structural procedures, it is evident that a detailed assessment procedure specifies:

- Requirements for the necessary information of the structure itself, the foundation ground, and

the seismic actions, based on site investigations, laboratory tests, investigation of existing drawings of past conditions, etc.

- Specified methods of analysis
- Specifications of the limit state that should be examined
- Global and local verifications of seismic resistance

Design of Seismic Retrofitting

The design procedure for seismic retrofitting includes the following steps:

- Conceptual design
- Analysis
- Safety verifications
- Drawings
- Technical reports

The scope of this entry does not allow detailed reference to the above steps. However, brief reference should be made to some important issues:

- Conceptual design is of particular importance, because it formulates the type and configuration of the retrofit scheme.
- Analysis and safety verifications need special consideration, as the prevailing procedure is the **displacement-based** design, in contrast to the **force-based design** which is the prevailing procedure for new seismic-resistant structures.
- Drawings should include three groups, namely:
 - Demolition drawings
 - Temporary propping drawings
 - Retrofitting drawings
- Technical reports should be similar to those for new structures.

Execution

The execution of retrofit works must be carried out by a qualified contractor.

During retrofitting activities, due to the demolition works on the structure, temporary propping works are required. This stage of the works is the riskiest, as unexpected collapses and casualties may occur.

In retrofitting activities special materials and techniques are used that are very rarely applicable to new structures. Therefore, a detailed study of their characteristics is required, as well as careful supervision during their application.

The form and the extent of interventions cannot be completely foreseen at design stage. The contractor and his technical staff are often compelled to improvise in order to adjust materials and techniques to the needs of unexpected existing conditions.

Concluding Remarks

The above brief presentation has been considered necessary for treatment of the quality assurance and liability issues for retrofitting and strengthening which follow.

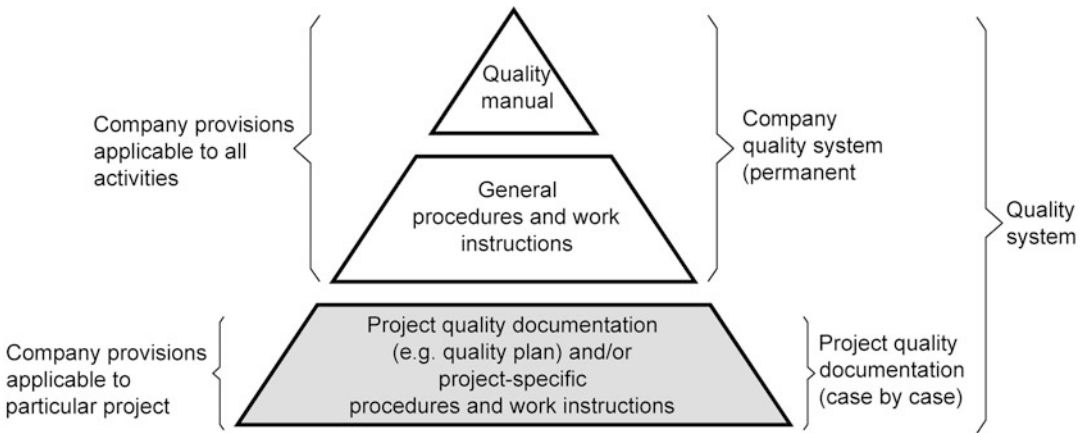
In fact, although seismic assessment, retrofitting, materials, and techniques are presented by other authors in detail in other chapters of this encyclopedia, it was not possible for us to develop quality assurance and liability issues for activities that have not been incorporated, even summarily, in a properly integrated structure, in compliance with the steps that follow.

Quality Assurance

General

For the clarification of various issues that will be presented below, some definitions of quality management must be given in advance:

- **Quality** of a work (e.g., retrofitting) or of an action (e.g., assessment) is its conformity to performance requirements specified by the code.
- **Quality assurance** (ISO 8204 text, CEB, Bulletin 241 [1998](#)) expresses all the planned and systematic activities implemented within a quality system and demonstrated as needed, in order to provide adequate confidence that a work (e.g., retrofitting) or action (e.g., assessment) will fulfill performance requirements for quality.



Retrofitting and Strengthening Measures: Liability and Quality Assurance, Fig. 6 Conceptual form of quality system (Adapted from CEB Bulletin 241 (1998). Quality Management guidelines, Thomas Telford, London)

- **Quality plan** is a document setting out the specific quality practices, resources, and sequence of activities relevant to a particular work or project.
- **Quality control** in the construction industry is a set of documented procedures used to monitor principal activities and processes during realization of a project in order to reduce nonconformities with performance requirements to a minimum. It is therefore applied during all stages of the project, including design, work preparation, execution, and handing over.

Over the last 20 years, an effort has been made to reform the quality assurance system. In the past, primary importance was given to the control of the quality of the final product and of services at the interfaces during the construction process. Current quality assurance focuses more on process control during production. In this respect, quality assurance means quality management of processes and is directed as much toward the internal organization of each participant and to the interfaces between them. The purpose of this change is the **reduction of failure cost** due to deficiencies in the project.

Contrary to most other industries, the construction industry usually operates at temporary locations. This requires the development of

procedures for each particular project. In this context the quality system in the construction industry includes two parts (see Fig. 6) (CEB, Bulletin 241 1998):

- The company quality system (permanent)
- The project quality plan and its documentation (case by case)

In other words, the result of a project quality plan is adequate project quality documentation. This includes procedures and instructions developed mainly for the given project. It is apparent that in the following subsections, only reference to the project quality plan and its documentation will be made, since the company quality system is beyond the scope of this entry. At the same time, an effort will be made to adapt the general rules in effect mainly for new structures to seismic retrofitting and strengthening of existing buildings and public works.

Quality Assurance of Assessment Procedures

According to the presentation of section “[Seismic Assessment](#),” assessment procedures may be classified in the following categories:

- Post-earthquake emergency damage inspections
- Post-earthquake assessment
- Pre-earthquake assessment

The above activities have nothing to do with building production procedures as they are services. However, the quality of assessments influences decisions in regard to retrofitting and therefore the performance of buildings or similar works in future earthquakes.

The detailed assessment of existing structures must be examined together with retrofitting activities, as it constitutes a part of the retrofitting design.

Quality Assurance of Emergency Post-earthquake Inspections

The quality plan of the project (emergency post-earthquake inspections) must refer to the following issues:

- Determination of the affected areas in order to estimate the scale of the operation
- Qualifications of the personnel who will undertake the inspection and evaluation activities at various levels and departments, that is:
 - Personnel for first-level inspection of conventional buildings
 - Personnel for inspection of buildings of special interest (e.g., hospitals, schools, etc.)
 - Personnel for second-level inspections
 - Personnel for third-level inspections
 - Personnel for emergency proppings or demolitions (design–execution)

The qualifications and number of the members of inspection teams, in the case of the earthquake of June 20, 1978, in Thessaloniki and all others in Greece since then are depicted in Fig. 7 (Penelis and Kappos 1997).

- Personnel resources. It is very important for the whole operation to ensure in advance the personnel that will be used for the inspections. Public administration, public corporations, and the private sector supply the usual personnel resources (see Figs. 7 and 8) (Penelis and Kappos 1997; Penelis and Penelis 2014).
- Selecting in advance the building for the operational headquarters. This should be a building designed for immediate occupancy (e.g., an existing communications center).
- Specifications of the activities of the headquarters should be clarified in the quality

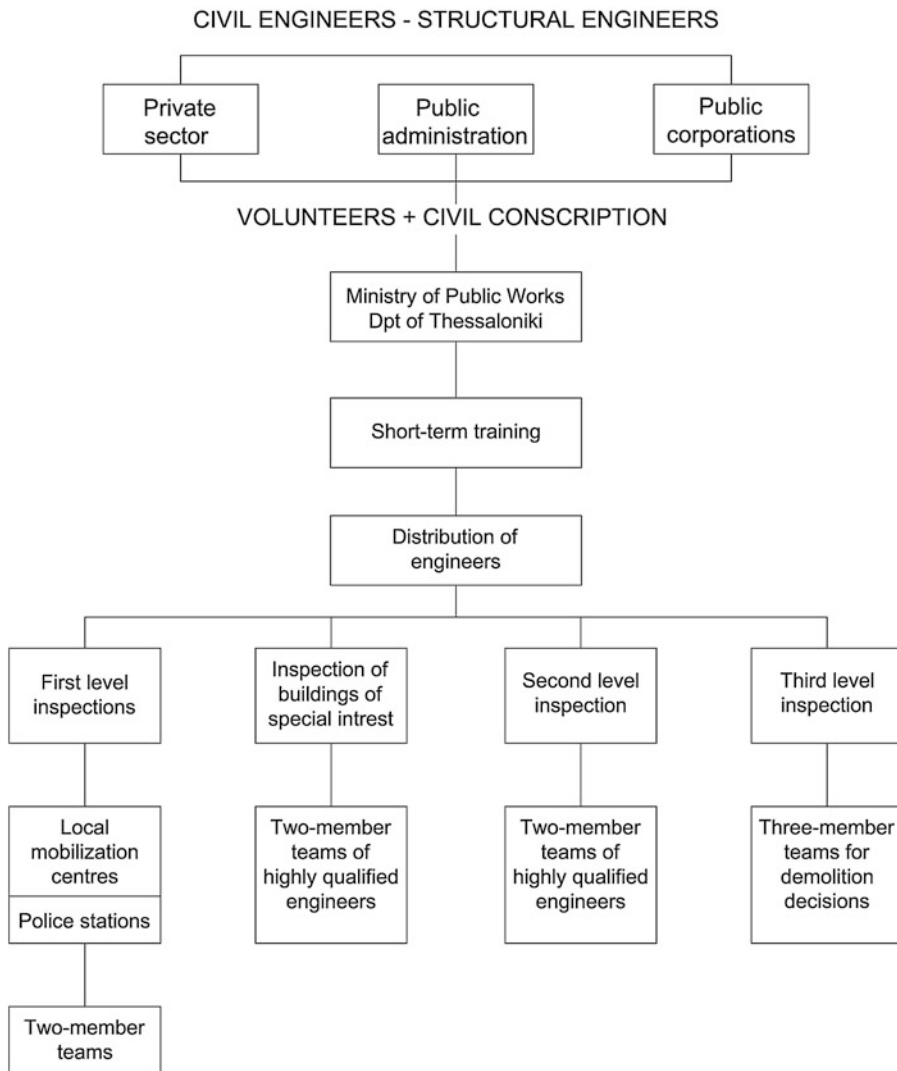
plan of emergency inspection and the relevant documentation, that is:

- Coordination with other services charged with the disaster relief
- Mobilization of the personnel for inspections and distribution of these personnel to local inspection centers
- Coordination of the activities of local inspection centers
- Determination of the position of the local inspection centers in a city or a district
- Distribution of inspection personnel depending on qualifications into first-level teams, second-level teams, committees for checking buildings of special interest, etc.
- Central management of all materials and equipment needed for the inspection teams such as inspection forms, measuring tapes, hammers, chisels, flashlights, batteries, etc.
- Organizing brief training courses on damage assessment for inspection personnel
- Provision of summary manuals prepared in advance on seismic pathology, the method of emergency inspections, and the filling in of the inspection forms that the teams are obliged to prepare after each inspection
- Listing of buildings or other works of vital importance for the post-earthquake operation of the area which must be inspected first by highly qualified structural engineers

The responsibility for the preparation of the quality plan and the relevant quality documentation belongs to the “Governmental Emergency Management Agency,” which should prepare, in cooperation with local authorities, quality plans for all districts of high seismicity of a country.

The key to the success of the above plan of operations is the personnel:

- They must all be at least civil engineers with ascending qualifications in structural engineering and long experience in higher-level inspections and decision-making.
- They must be teams of two persons so that they can cross-check their evaluations.



Retrofitting and Strengthening Measures: Liability and Quality Assurance, Fig. 7 Organizational chart for the inspection service in Thessaloniki after the earthquake

of 20 June 1978 (From Penelis and Kappos (1997) with permission of C.R.C. Press)

R

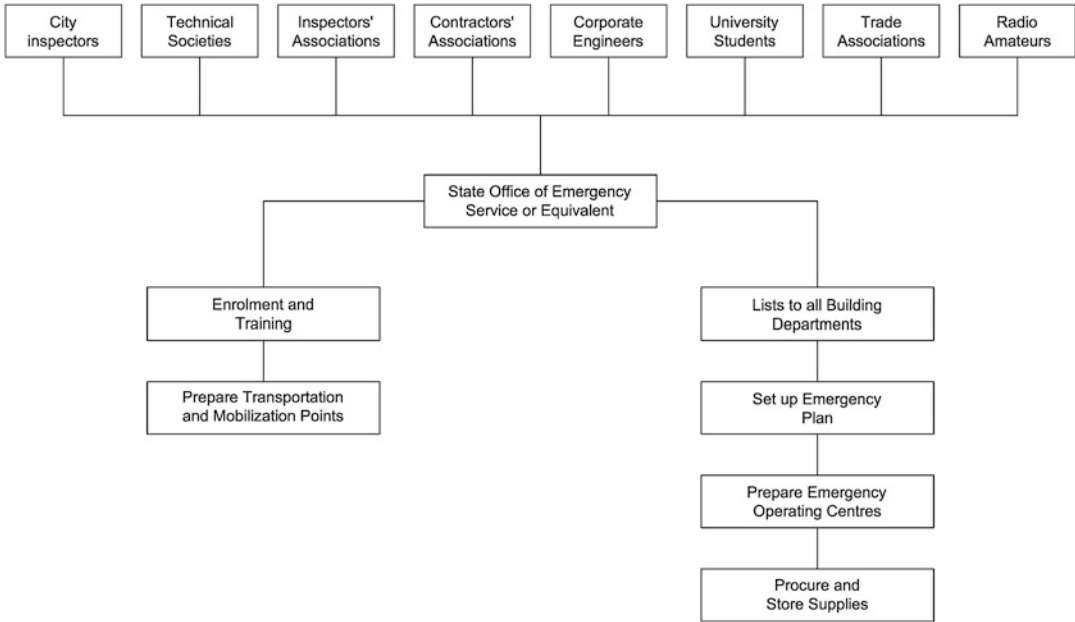
Quality Assurance of Post-earthquake and Pre-earthquake Assessment

The quality plan and the corresponding documentation must refer to the following issues:

– References and applicable documents

Documents to be used as a source of reference in the procedure shall be listed. The following may be mentioned by way of example:

- Contract documents
- Project specifications
- Standards and guidelines for the assessment procedure (e.g., FEMA 310/1998/ ASCE 31–02 or relevant ministerial decrees at national level)
- Other documents relevant to the procedure (e.g., existing files with drawings and calculations)



Retrofitting and Strengthening Measures: Liability and Quality Assurance, Fig. 8 Organizational chart for the pre-earthquake preparation period according to

ATC3-06 (USA) (From Penelis and Kappos (1997) with permission of C.R.C. Press)

– **Preparation and handling**

- The most important aspects of the evaluation must be condensed in the quality plan, identifying:
- Collection of data and visit to the building or public work
- Determination of the seismicity of the region
- Determination of the level of performance
- The number of quality-related activities (e.g., the three tiers of FEMA 310/ASCE 31-02 standard and the number of the checklists of each tier)
- The checklists that must be used for each activity (e.g., structural checklists, foundation checklists, and nonstructural checklists)

All the above are depicted in the flowchart in Fig. 5 (FEMA 310/1998/ASCE 31-02).

- **Simplified analysis and design verifications at all critical regions of the structural system.**
- **Identifying damages for any reason.**

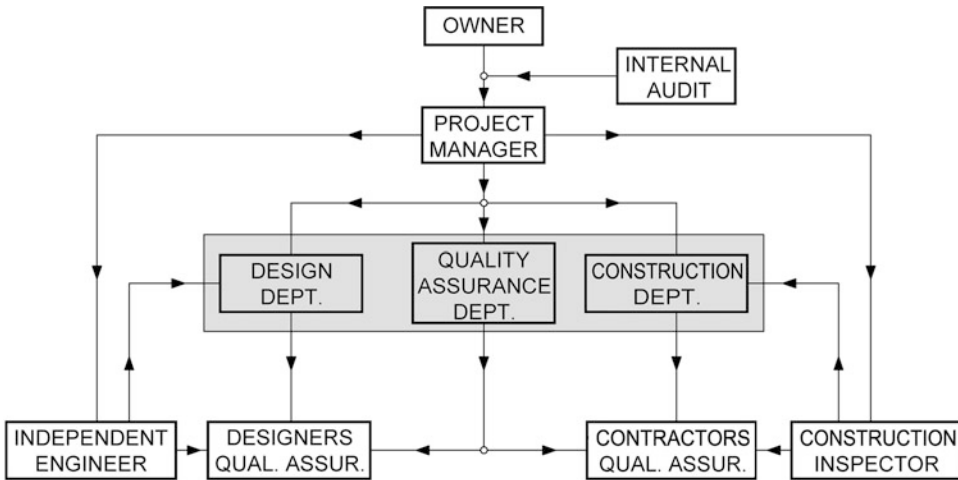
– **Detailing deficiencies.**

- **Quality control.** This may be carried out by an **independent engineer** or by the **internal audit** of the agency responsible for the evaluation.

Quality Assurance of Detailed Assessment and Retrofitting

General

For a successful structural intervention, coordinated measures are needed in order to ensure quality of design and construction. The detailed assessment is included in the design procedure. In this respect two separate quality plans must be elaborated, the first for the design and the second for the construction. The first is elaborated by the company that is responsible for the design, the second by the contractor. Both project quality plans must be approved by the project manager and then are added to each company's quality system (see Fig. 6) and to the contract with the owner of the building (Fig. 9).



Retrofitting and Strengthening Measures: Liability and Quality Assurance, Fig. 9 Flowchart of quality management procedure of retrofitting activities

Design Quality Plan

Though the design quality plan should be tailored to the particular project, such a plan should include the main phases of the project and address at least the following issues:

– References and applicable documents

Documents to be used as a source of reference in the procedure shall be listed. The following are mentioned by way of example:

- Contract documents
- Project specifications
- Standards and guidelines for the detailed assessment and retrofit design (e.g., FEMA-356/2000/ASCE SEI 41–06 or EC 8–3/2005, National Annexes, etc.)
- Other documents relevant to the procedure (e.g., existing files with drawings and calculations)

– The most important aspects of the detailed assessment and retrofitting design must be condensed in the quality plan identifying such elements as:

- The level of technical data generated by the field investigation, in relation to the **knowledge levels** that will be introduced in the analysis and design and particularly with respect to:

- Geometry
- Details of the structural system
- Materials of the structural system
- The analytical methods to be used to identify and quantify the building response and vulnerabilities
- The structural performance levels or limit states for which the structure shall be assessed and redesigned
- The principles of conceptual design to be followed
- Drawings and documents that should be included in the design study, with particular reference to:
 - Drawings of demolitions
 - Drawings of temporary propping
 - General drawing of retrofit
 - Detail drawings
- and to the following reports:
 - Report of detailed assessment
 - Report of redesign of the rehabilitated structure
 - Technical descriptions of the intervention scheme
 - Technical specifications for materials and works
 - Quantity estimates
 - Bill of quantities
 - Cost estimates

- Testing and inspection plan at construction stage
 - The main phases of the project for which the contractor shall elaborate a method statement and shall submit it to the project manager for approval
- **Quality control of design**
- Seismic retrofitting (repair or strengthening) requires an appropriate scheme of counterchecking of design documents. When the design has been completed, an additional check should be performed by an **independent engineer** providing an independent second opinion (ATC 40–1996). This may be a governmental or private agency responsible for verifying the criteria and checking the calculations, the drawings, and the technical reports, certifying that they conform to the criteria and regulations of the building codes and relevant guidelines. No matter who is financially responsible for quality control (the owner or the designer), this special requirement for a third-party control is incorporated into the quality plan and its documentation.
- **Design review**
- The design files (drawings, technical reports, and calculations) together with the checking of the “independent engineer” are **reviewed** and approved by the project manager of the work (see Fig. 9) before releasing the tender documents for construction bids.
- **Building permit**
- The obligation of the “designer” to get the building permit for the retrofit from the relevant governmental or municipal agency by preparing all necessary documents must be referred to in the quality plan of the project.

Construction Quality Plan

Construction of seismic retrofitting works must be checked for quality and general compliance with the intent of the drawings, technical documents, and specifications of the retrofitting design, via the construction quality plan of the project. As was previously mentioned, this document will be prepared by the contractor and will be checked and approved by the project manager (Fig. 9).

This document should at a minimum include the following:

1. Required contractor quality control procedure including elements such as the following:
 - Materials testing
 - Submission to the project manager of the required quality certificates for industrial materials (e.g., steel reinforcement, FRP laminates, etc.)
 - Temporary propping procedures
 - Demolition procedures
 - Reinforcement detailing
 2. Required “designer’s” services including steps such as the following:
 - Review of required contractor submittals
 - Monitoring of required inspection reports and test results
 - Consultation with the contractor on the intent of the construction documents

The “designer” is either the original designer of the retrofitting design or an associate of the project manager of the work with the proper qualifications.
 3. The main phases of the project for which the contractor shall elaborate and submit method statements to the project manager for approval
- The designer shall be responsible for performing periodic structural observation of the retrofitting works at significant stages of construction like temporary propping, demolitions, reinforcements, details, etc. This procedure should be in addition to any special **inspection and testing**.

The project manager of the work shall engage on behalf of the owner the services of a special inspector to observe construction and prepare **quality records and nonconformity and corrective action reports**.

The above inspector shall be responsible for verifying the special test requirements as described in the construction quality plan.

The standard FEMA-356/2000/IASCE SEI 41–06 introduces a strong involvement of the Code Official, that is, the governmental or municipal agency for permits for the quality control of design and construction of seismic retrofitting.

Liability

General

When, despite the existing quality assurance system, deficiencies occur, then liability issues arise. Some legal terms will be briefly clarified below since they will be used in the discussion that follows.

Liability is one of the most significant terms in the field of law since it describes the condition of being actually or potentially subject to a legal obligation. Liability may be **criminal** or **civil**.

The basic assumption in **criminal liability** is that there is both a “mental element” and a “physical element” to the offense. It should be noted that various offenses in relation to, for example, traffic law or environmental law have been so structured that the “mental element” is in fact not required for a conviction. In this category are also classified offenses in violation of building law. These offenses are not characterized as **crimes** but rather as **misdemeanors** and are basically related to “**negligence**.”

Civil liability gives a person rights to obtain redress from another person, e.g., the ability to sue for damages or for personnel injury. For there to be an award of damages, the injured party has to have suffered an actual loss, be it personal injury, damage to property, or financial loss. The burden of proof is “the balance of probability,” which is much lower than for criminal matters.

Professional indemnity insurance is the kind of insurance that protects the insured from claims by dissatisfied clients in disputes over errors. The most usual claims involve:

- Negligence
- Intellectual property
- Loss of documents/data
- Dishonesty
- Breach of duty
- Defective products

In this respect **professional indemnity insurance** can cover mainly **civil liabilities** owing to **professional negligence**.

Liability for Buildings and Civil Engineering Works in Criminal Law

Most European countries have incorporated into their Penal Law articles against the violation of the rules of building.

For example, Greek Penal Law stipulates the following:

“Anybody who violates by probable intention or negligence the commonly accepted technical rules during the design, management, or execution of buildings or public works or a demolition, thus putting at risk the life or health of persons, is punishable for up to 2 years in prison.”

Similar stipulations are also met with in the penal codes of other European Countries. According to the above stipulation, the person responsible for the violation of the law may be:

- The designer
- The manager of the building or civil engineering work
- The contractor who executes the work or the owner if he executes the works using employees

Bearing in mind that during design and construction of large-scale projects, each of the above “responsibilities” corresponds to a large number of persons, the law gives the judge the discretion to distribute liabilities based on the evidence of witnesses and existing documents.

According to the law, liability arises in the case of **violation of the commonly accepted technical rules**. These rules comprise presidential decrees related to building activities, codes of practice, standards, technical specifications, etc. A basic parameter for the judgment is the elaboration of **expert evidence**, which the court solicits from an independent expert.

A basic parameter for a conviction is the objective justification of the risk of life or health and the evidence of causality between violation of building rules and damage. In practice it is very difficult for the judge to justify a conviction without any injury or death in the work. It should be noted that in the case of death or injury, the article of criminal liability in building works is combined cumulatively with the

relevant article of Penal Law for death or injury **owed to criminal negligence**.

Liability for Buildings and Civil Engineering Works in Civil Law

Liability for building works is included in the **tortious liability** provisions of the civil code of each country. Similarly, in the EU the Product Liability Directive 85/374/EEC of 25 July 1985 has established for EU countries legislation regarding **strict product liability** for damage arising from defective products. This liability is in addition to any existing rights that consumers enjoy under domestic law.

Strict liability applied to product liability suits makes a manufacturer or seller responsible for all product defects and particularly **design defects, manufacturing defects, and defects in marketing**. Strict liability wrongs do not depend on the degree of carefulness of the defendant. This means that the defendant is liable when it is shown that the product is defective and the product has caused harm to the consumer.

The term “**product**” in the above directive refers to **movable** personal property. In the corresponding US law, it refers to **tangible** personal property, and in this respect the relevant law in the USA may also extend to the products of the building industry. The above directive must be related to Directive 89/106/EEC, which refers to the technical characteristics particularly of the **construction products** which are incorporated into buildings or public works. In fact, the Council of the EU, recognizing the significance of building materials and of building procedures for the protection of persons and property, has established Directive 89/106/EEC of 21 December 1988. This is a **performance-oriented** directive, since technical characteristics of the “**construction products**” are influenced by the performance-based six essential requirements for the works, namely:

- Mechanical resistance and stability
- Safety in case of fire
- Hygiene, health, and the environment
- Safety in use

- Protection against noise
- Energy economy and heat retention

European codes, standards, and technical approvals are the consequence of the above directive, and in this respect also of [EC8-3/2005](#), which applies to the detailed seismic assessment and retrofitting of buildings.

In other words, particularly for “**construction products**” and building or civil works, the EU has established a very strict system for quality assurance (codes, standards, etc.). Thus, in case of defects in the works, a causality between defects and code violation can be easily found and liabilities may be determined and justified.

Liability for Retrofitting and Strengthening of Buildings or Public Works

It is apparent that this type of liability lies within the bounds described above of criminal and civil liability for construction products and buildings or civil engineering works.

However, a distinction should be made here between seismic assessment (section “[Seismic Assessment](#)”) and detailed seismic assessment and retrofitting (section “[Detailed Seismic Assessment and Retrofitting](#)”), since these two main activities differ drastically.

Liability for Seismic Assessment

Emergency seismic inspections and assessment procedures are conducted under the pressure of time and cost as they involve thousands of buildings and public works. In particular, emergency inspections are conducted in emergency situations. Therefore, the engineers are not able to use their main scientific tools, that is, on-site measurements, laboratory tests, and analysis. In this respect it is the author’s opinion that the assessment personnel should not be held personally **liable** for any damage that may occur to persons and property as a result of any act or omission in engineers carrying out their duties during emergency inspections. Otherwise, the personnel becomes very conservative and the damage findings of the assessment appear to be much more extensive than the reality. The case of the Kozani earthquake in Greece (1996) is

characteristic, where at the first-level inspection 90 % of the schools of the region were found to be “red” (extended structural damage), while at a second-level inspection these were limited to only 10 % “red” and 15 % “yellow.” For the time being, existing legislation in EU countries does not make any reference to this issue, leaving the application of the tortious liability provisions to the discretion of the judge.

It is important to note here that in US Presidential Executive Order 12941/1-12-94 (1994) for “**Seismic Safety of Existing Federally Owned or Leased Buildings**,” where the assessment procedure for the above building is specified, it is explicitly stated that: “Nothing in this order is intended to create any right to administrative or judicial review, or any other right, benefit or trust responsibility, substantive or procedural, enforceable at law by any party against the U.S., its agencies or instrumentalities, its officers or employees or any person.” In other words, this paragraph of the above Executive Order releases anybody involved in the assessment procedure from any liability, civil or criminal.

Liability for Detailed Seismic Assessment and Retrofitting

As presented in section “[Detailed Seismic Assessment and Retrofitting](#),” this activity includes:

- Detailed assessment
- Design
- Execution of work

In this respect it is an activity similar to that for a new building or civil engineering work. Therefore, the legislation for liability applying to new buildings or civil engineering works also applies to seismic retrofitting and strengthening.

Summary

This chapter deals with the steps of procedure for seismic retrofitting and strengthening of buildings and civil works. The quality assurance

mechanisms applied at each step of the retrofitting procedure are also outlined in a detailed fashion. Finally the liabilities emerging in case of defects or failures are outlined and described so that the practicing engineers as well as the relevant authorities can evaluate their significance, during crisis design making.

Cross-References

- ▶ [Assessment and Strengthening of Partitions in Buildings](#)
- ▶ [Earthquake Risk Mitigation of Lifelines and Critical Facilities](#)
- ▶ [Seismic Vulnerability Assessment: Lifelines](#)
- ▶ [Seismic Vulnerability Assessment: Masonry Structures](#)
- ▶ [Seismic Vulnerability Assessment: Reinforced Concrete Structures](#)
- ▶ [Strengthened Structural Members and Structures: Analytical Assessment](#)
- ▶ [Strengthening Techniques: Bridges](#)
- ▶ [Strengthening Techniques: Code-Deficient R/C Buildings](#)
- ▶ [Strengthening Techniques: Code-Deficient Steel Buildings](#)
- ▶ [Strengthening Techniques: Masonry and Heritage Structures](#)

References

- ATC 40–1996 Seismic evaluation and retrofit of concrete buildings. Applied Technology Council, Redwood City
- CEB, Bulletin 241 (1998) Quality management guidelines. Thomas Telford, London
- DIRECTIVE 85/374/EEC, Liability for defective products 25 July 1985, European Economic Community, Brussels
- DIRECTIVE 89/106/EEC, Construction products 21 December 1988, European Economic Community, Brussels
- EC8-3/2005 (Chap. 17)
- EXECUTIVE ORDER 12 941 (1994) Seismic safety of existing federally owned or leased buildings. Executive order of December 1, 1994. The White House, Washington, DC
- FEMA 310/1998/ASCE 31–02 (Chap. 17)
- FEMA-356/2000/ASCE SEI 41–06 (Chap. 17)

- King LW (1915) *The Code of Hammurabi* (translation) Liling Galdman Law Library, 127 Wall Street, New Haven, 2008
- Penelis GG, Kappos AJ (1997) *Earthquake-resistant concrete structures*. SPON E&FN (Chapman & Hall), London
- Penelis GrG et al (1987) *A statistical evaluation of damage to buildings of Thessaloniki in the 1978 earthquake*, technical report of the lab of R/C structures. Aristotle University of Thessaloniki (in Greek)
- Penelis GrG, Penelis GGr (2014) *Concrete buildings in seismic regions*, CRC Press (Taylor and Francis Group), London. ISBN-10: 0415537630, ISBN-13: 978-0415537636

Retrofitting and Strengthening of Contemporary Structures: Materials Used

Theodoros Rousakis

Department of Civil Engineering, School of Engineering, Democritus University of Thrace, Xanthi, Greece

Synonyms

Concrete; Fiber-reinforced polymers; Injective resins; Materials; Retrofit; Seismic; Steel; Strengthening

Introduction

Contemporary Structures, Seismic Retrofit, and Strengthening

A variety of materials have emerged into the retrofit and strengthening field. Interdisciplinary approaches in material's science, in structural retrofit, and in earthquake engineering have better addressed unidentified gaps in knowledge. Growing interlinks among industry, researchers, and practitioners favor introduction of advanced materials per different kinds of structure, application, and retrofit technique. Modern retrofit design combines different materials utilizing their unique attributes (physical properties, mechanical as well as time- and loading type-

dependent behavior and durability-related properties *fib* Model Code 2010, *fib* 14, ACI440-08) toward optimized performance of the structure as a whole.

This chapter presents the materials used for the seismic retrofit of structures covered by modern retrofit design codes and recommendations such as reinforced concrete, steel, and composite structures as well as masonry structures (EC8-3 2005; FEMA 547 2006; Fardis 2009; CNR-DT 200/2004). It discusses materials for both repair and strengthening of contemporary structures with special focus on reinforced concrete buildings and bridges (Priestley et al. 1996; Fardis 2009) as well as on advanced materials (high-performance cement-based materials, shape-memory alloys, fiber-reinforced polymers, textile-reinforced mortars, etc.).

Jackets made of concrete and reinforcing steel as well as repair mortars have been widely used in the retrofit of structures for many decades. In the 1960s strengthening of concrete structures included the emerging technique of structural steel plates bonded to the surface of the tension zone with adhesives and bolts. Further, the use of fiber-reinforced polymers (FRPs) has been established in construction in the late 1980s, and they have been increasingly used ever since (TR 55 2012). Specifically, FRPs were firstly used in Europe and in Japan, while their use was adopted in the United States and Canada later on (Bank 2006). FRP systems and applications focused on seismic retrofit and strengthening were developed by the Japanese, while United States followed in the early 1990s. Today a wider variety of retrofit and strengthening material exist including injective resins, polymeric mortars, fiber-reinforced concrete, nonmetallic reinforcements, and even nanoenriched materials, and their availability has broadened remarkably the options of practitioners. Application performance-based approaches and standards for the design, application, and quality control of materials used have raised retrofit efficiency, reliability, and safety. Earthquake-resistant repair and strengthening requires advanced performance of added materials as they usually need to resist extreme stresses, together with the

elements of the existing structure, to ensure collapse prevention and avoid human loss.

In what follows common materials used in retrofit and strengthening of structures are presented. They are classified according to the nature of their basic constituent.

Retrofit and Strengthening Materials

Concrete and Cement-Based Materials

Cast-in-place ordinary concrete, fiber-reinforced concrete, shotcrete or gunite, self-consolidated concrete, and advanced mortars (polymeric or fiber enriched) comprise a class of materials widely used for the purpose of retrofit and strengthening of existing structures. Most of these materials serve as the matrix for metallic reinforcements to form external retrofitting or strengthening jackets or for the repair of damaged concrete members. Advanced mortars may be used as matrix for nonmetallic reinforcement (mainly textiles) in jacketing. Advanced fiber-reinforced concretes or mortars (with steel, polymeric, carbon, glass, natural, or other discontinuous fibers) may be used as sole strengthening systems. Repair patches, mortars, and cement grouts are used in damaged concrete and masonry members. Special materials of the category may be used as protective finishing. All materials of this class should ensure adequate bond with existing old concrete surface in bond-critical applications. The added concretes and mortars may contain appropriate fibers (steel, polypropylene, PVA, etc.), constituents, or additives to minimize the effects of shrinkage or creep deformations and prevent the deterioration of the interface bond.

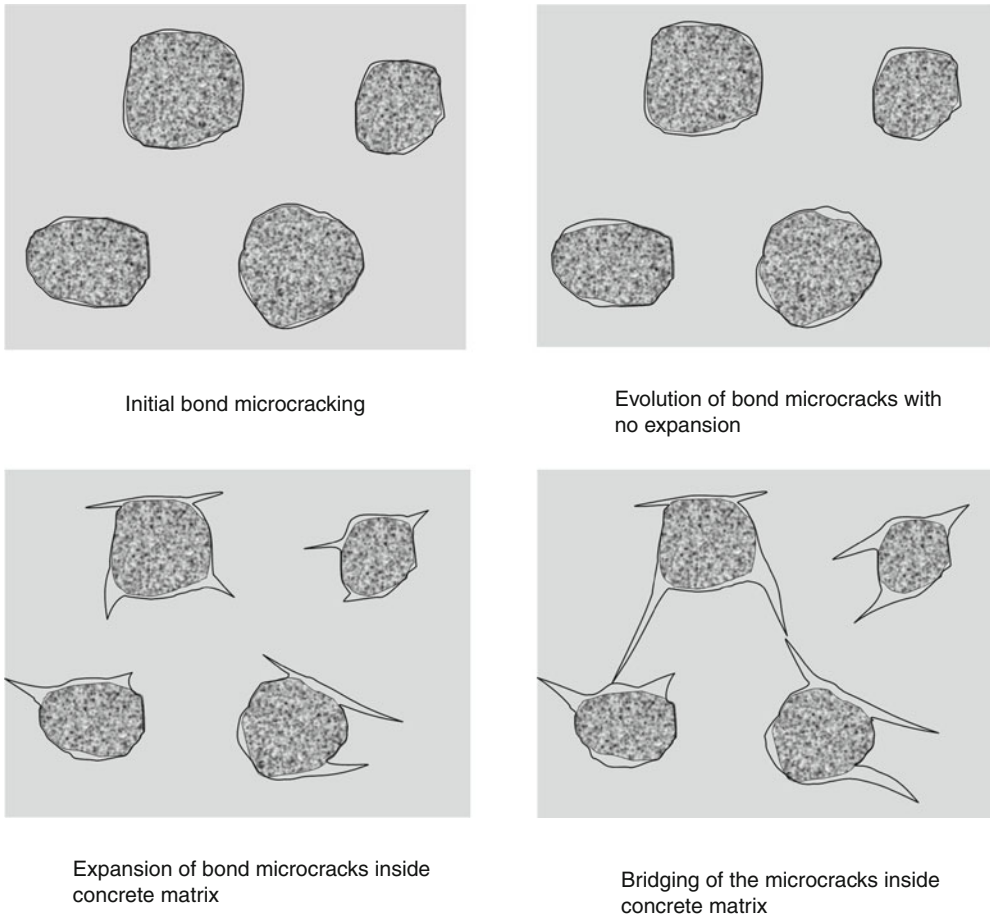
Many advanced cement-based materials are already under consideration for the next generation codes (*fib* Model Code 2010), while most recent and emerging developments in the field (Fardis 2012) are expected to follow.

Concrete Under Earthquake Excitations

Concrete is widely used in retrofit and strengthening of existing structures. Recent developments in cement-based materials are aiming at

offering performance-based solutions per application. The mechanical behavior of concrete is usually considered in a macroscopic level. In this regard, the effect of concrete composition and of its constituents is neglected and is taken as homogeneous continuum. However, concrete is a composite heterogeneous material, as it consists of gravels bonded together with a mixture that consists of cement, sand, and water. A microstructural level approach may reveal the effect of the size, quantity, and properties of the constituents of the concrete on the mechanical behavior of the material at the macroscopic level. The mechanical performance of the concrete depends on the mechanical performance of the gravels, the cement paste, their bond, and their ratio. Specifically, it should be taken into account that extensive bond microcracks exist at the interface between gravels and mortar even before the loading of concrete. Many of the microcracks are caused by segregation, shrinkage, or thermal expansion in the mortar. Further, the cement paste has a high porosity of about 30 % that corresponds to presence of water and/or air. During concrete loading, microcracks are developed (Fig. 1) because of the differences in stiffness between aggregates and the mortar. Consequently, the aggregates/mortar interface becomes the weakest link in the composite material. This is the reason for the low tensile strength of concrete. Moreover, microcracks explain the nonlinear behavior of concrete during loading. Accurate modeling of the behavior of concrete in microstructural level should take into account the abovementioned issues. The effects of partial plasticization and of the fracture of concrete on the inelastic behavior of the material should be considered as well.

In what follows, the concrete is considered from a macroscopic point of view. As already mentioned, ordinary concrete used in jackets possesses a low tensile strength. Thus, from a macroscopic point of view, cracking normal to tensile stresses (even low stresses) is expected. Under inelastic uniaxial compressive loading, concrete elements exhibit a gradual degradation of their elastic modulus due to accumulation of internal damage. This degradation is fairly small

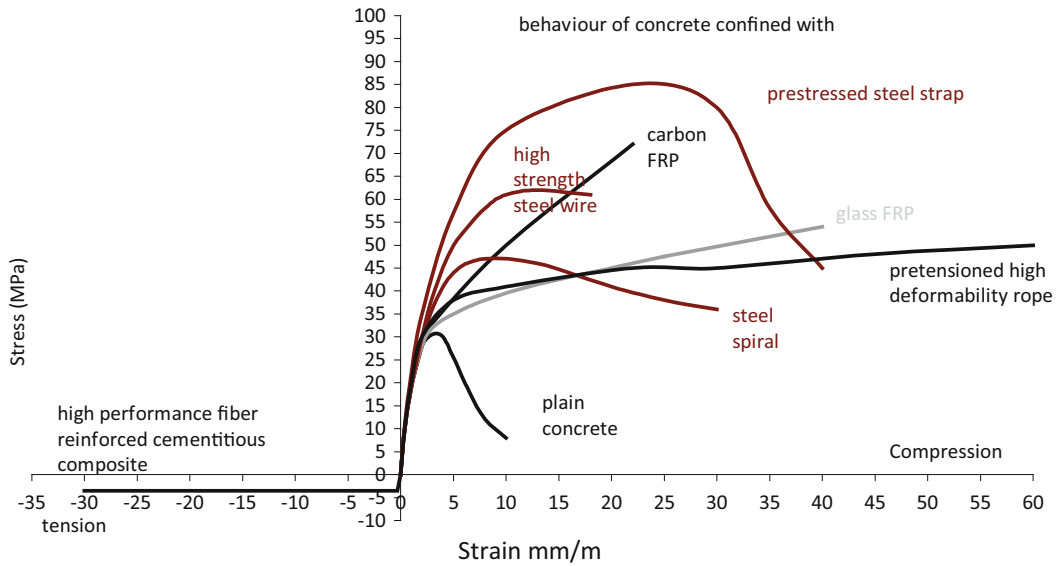


Retrofitting and Strengthening of Contemporary Structures: Materials Used, Fig. 1 Development of microcracks inside concrete during loading

in the pre-peak response but becomes significant in post-peak loading. Moreover, in seismic redesign of concrete structures, low-cycle fatigue effects on concrete under compression are not expected to be important. This is because the duration of strong ground shaking and the predominant periods of common buildings suggest normally a low number of high compressive stress cycles, not more than 10 (Fardis 2009). Thus, the monotonic compressive stress-strain behavior of existing or added concrete may serve as the envelope curve of its cyclic behavior under earthquakes in redesign.

The use of higher-quality concrete in added jackets than the existing ones magnifies the inherent brittleness of the post-peak behavior of the material under uniaxial compression. In concrete

members, transverse reinforcing materials are used to confine and thus restrain adequately lateral expansion of concrete (steel or FRPs, etc.). The concrete is then under triaxial compression, and a significant increase of its strain ductility (and secondly of its strength) takes place (see Fig. 2). The effects from the confinement of concrete are discussed in the sections “[Steel and Metal-Based Materials](#)” and “[Nonmetallic Reinforcements and Resins](#)” of the main confining materials. In general, the monotonic behavior of confined concrete may also serve as the envelope curve of its cyclic behavior. Finally, concrete-to-concrete interface bond may deteriorate under cyclic loading and should be taken carefully into account in the design of bond-critical interfaces.



Retrofitting and Strengthening of Contemporary Structures: Materials Used, Fig. 2 Typical stress–strain behavior of cement-based materials and different concretes plain or confined

Advances in Cement-Based Materials

Recently, a variety of fiber-reinforced cementitious products have been developed and applied in retrofit and strengthening of existing structures. Their main characteristic is the enhancement of the tensile characteristics with respect to ordinary concrete. Ultrahigh-strength fiber-reinforced concrete (UHPFRC) shows an increased tensile strength capacity without pseudo strain hardening. On the other hand, fiber-reinforced cementitious composites (FRCCs) exhibit pseudo strain-hardening tensile characteristics and are divided in two main sub-categories: strain-hardening cementitious composites (SHCC) and high-performance fiber-reinforced cement composites (HPRFRC). SHCC exhibit high tensile deformation at failure, while HPRFRC lie in between SHCC and UHPFRC exhibiting both pseudo strain hardening and increased tensile strength and strain.

The tensile stress of SHCC is relatively low compared to that of UHPFRC, while UHPFRC is superior to SHCC in terms of controlling crack damage. Therefore, SHCC is mainly used for the areas of tension in a flexural member. The SHCC mixture exhibits excellent performance in crack-damage mitigation and energy dissipation.

Thus, it is suitable for seismic-resistant structures and sections subjected to high stress concentrations. Further, HPRFRC exhibits several desirable characteristics as well, such as strain-hardening behavior, formation of multiple fine cracks, increased energy dissipation, and crack-damage tolerance. HPRFRC can also be used to retrofit reinforced concrete structures and to enhance their strength, ductility, and seismic performance. The unique tensile characteristics of HPRFRC can provide an alternative retrofit solution for extensively damaged concrete members in their critical region or in short members. High-ductility fiber-reinforced cementitious composite under tension had replaced the existing concrete core of a column under seismic loading. The overall lateral load-carrying and deformation capacities were improved when compared to the conventional reinforced concrete column specimen. The retrofitted reinforced concrete columns presented better controlled concentrated local damage in the zone of the plastic hinge. Flexural and shear cracks of the concrete, spalling of the cover concrete, buckling of the longitudinal reinforcing bars, and compressive crushing of the concrete were better controlled in the columns with HPRFRC (Cho et al. 2012).

Steel and Metal-Based Materials

This class of materials comprises reinforcing steel, prestressing steel, and structural steel (EC8-3 2005, *fib* Model Code 2010, FEMA 547 2006). The characteristic ductile strain behavior of steel after yielding enables the concrete section to exhibit increased curvature ductility in order to meet the performance-based criteria of modern design (required section curvature ductility or member chord rotation ductility that corresponds to the target structure displacement ductility). Reinforcing steel in the form of ribbed bars and of appropriate class (Eurocodes) or welded fabric is used in concrete jacketing or as near-surface mounted (NSM) reinforcement in primary concrete members (i.e., NSM stainless steel bars, Bournas and Triantafyllou 2009). Modern earthquake-resistant retrofit design requires the use of ribbed bars with characteristic yield strength between 400 and 600 MPa and hardening behavior within specified limits. Also, characteristic strain at maximum stress should be higher than 5 % or 7.5 % strain (see Fig. 3). For high-ductility class, the upper characteristic value of the actual yield strength should not exceed the nominal value by more than 25 % (Eurocodes). Reinforcing steel is also used as transverse reinforcement (stirrups) in retrofit and strengthening. Furthermore, extensively buckled or fractured bars after a severe earthquake may be replaced in damaged members. Steel dowels and hangers may be necessary to ensure force transfer between old and added concrete (see Fig. 4). In this case, the issue of steel weldability needs to be considered. Moreover, steel stud shear connectors may be used for the same reason in cases of reinforced concrete jacketing of steel members.

Prestressing steel (see typical stress–strain diagram in Fig. 3) is used in external tendons (posttensioning) in the form of strands or wires or bundle of monostrands (inside corrugated metal or plastic ducts or inside smooth steel or plastic pipes) in concrete structures.

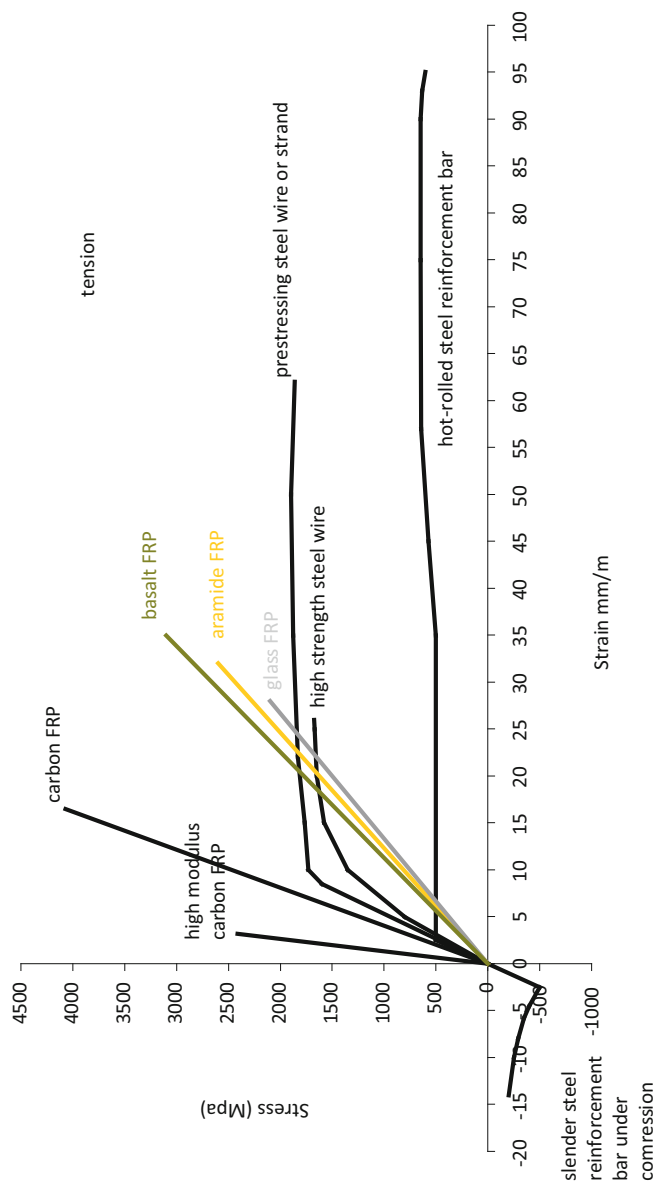
Structural steel corner angles, plates, and rods may be used as external steel jacketing of concrete members or in retrofit of masonry. Steel jacketing with two half shells welded together

to form a tube with a gap from the existing column, grouted with mortar, is another alternative mainly for bridge piers (Priestley et al. 1996). Structural steel (or aluminum) components may be used in energy dissipative zones of braced frames. Additional steel components may be used in concrete or in steel or in composite steel–concrete buildings. In every case, the characteristics of structural steel should conform to the criteria included in the relevant building codes of practice for steel structures.

Steel Under Earthquake Excitations

Reinforcing steel in external reinforced concrete jackets for the retrofit and strengthening of existing reinforced concrete or steel members comes mainly in the form of bars. Usually, the stress–strain behavior under tension–compression is crucial to assess the longitudinal bars' contribution. The stress–strain behavior under tension is of interest for the transverse reinforcement (stirrups). Steel dowels and stud shear connectors may involve shear (or tensile) behavior of steel- or bond-related issues. External steel plates may be considered under multiaxial loading (i.e., in cases of shear strengthening in order to estimate their yield stress). Bond-related issues may concern bar–concrete interfaces, bar–resin–concrete interfaces (NSM applications), or plate–resin–concrete interfaces under cyclic loading.

As far as the longitudinal bars are concerned, the modeling of their stress–strain behavior requires simple one-dimension elements. Yet, the reinforcing steel exhibits varying stress–strain behavior under cyclic tension–compression with respect to the type of the concrete member, the position inside the section of the member, and the type of the connection between existing member and added jacket (chemical bond, steel dowels, or steel hangers). Also, existing reinforced concrete members usually have inadequate anti-buckling detailing of bars. In existing concrete members subjected to seismic loading, bars normally yield in tension before they do so in compression. Buckling usually takes place during σ - ε branch of unloading from tension to compression that exhibits the



Retrofitting and Strengthening of Contemporary Structures: Materials Used, Fig. 3 Typical stress-strain curves of steel reinforcements and cured FRP sheets



Retrofitting and Strengthening of Contemporary Structures: Materials Used, Fig. 4 Details of jacket with dowels and hangers (Adapted from Tsakiris et al. 2012)

Bauschinger effect (gradual deviation of the tangent modulus of elasticity from linearity – before yield stress – to zero modulus value, after the first yield plateau) or the hardening branch that follows it (Fardis 2009). Buckling of bars depends on the lateral restraint by the transverse steel and the concrete cover, the flexural crack opening, the interaction with dilating concrete core, etc. Notice typical stress–strain diagram of slender bar under compression with stirrup spacing to bar diameter ratio 12 in Fig. 3. Concrete jackets may prevent premature buckling of old bars in the existing columns, but considerations remain for the bars of the jackets.

Typically, beams or columns with low axial compression exhibit predominantly tensile yielding. Therefore, the monotonic curve provides a reasonable envelope to the cyclic response in the tension range, while compressive behavior deviates. In columns with high compression stress levels and high confinement ratios, reinforcing bars may be subject to strain reversals of almost equal magnitude. In that case the stress level for a given strain may substantially exceed the stress indicated by the monotonic behavior (Priestley et al. 1996).

Concrete confinement by steel increases remarkably stress and especially strain of concrete at failure. Notice in Fig. 2 typical stress–strain curves for concrete confined externally with steel spiral. Further, it favors the force transfer between existing column and concrete jackets. Adequate compressive strain ductility of concrete ensures the success of the performance-based redesign in the member level. In general, the monotonic behavior of confined concrete may also serve as the envelope curve of its cyclic behavior with the stipulation that the issues raised for the longitudinal bars' buckling as well as force transfer issues through interfaces and strain gradient in sections subjected to cyclic bending are accounted for.

Recent Advances in Metal-Based Materials

The use of shape-memory alloy (SMA) reinforcements is very promising in retrofit and strengthening of existing structures. SMAs have more than one crystal structure. This is called polymorphism. The prevailing crystal structure or phase in polycrystalline metals depends on both temperature and external stress. They are a class of metallic alloys that can “remember” their initial geometry during transformations (forward and reverse) between two main phases at their atomic level (austenite and martensite).

SMAs include copper–zinc–aluminum, copper–aluminum–nickel, and nickel–titanium (NiTi) alloys. NiTi alloys exhibit better mechanical properties than copper-based SMAs which were first developed in the early 1960s. SMAs have been widely used in retrofit and strengthening projects in the field of damping, active vibration control, and prestressing or posttensioning of structures with fibers and tendons, while recently, iron-based SMAs were developed that can reduce the cost.

Among several recent advances, NiTi SMA wires have been pretensioned before application and then wrapped around a concrete column as external confining reinforcement while being at the martensite phase. Then, by raising the temperature of the wire, transformation to the austenite phase takes place. In this manner, the wire recovers its original shape, and stresses are

imposed by the wire to the concrete core. The effects of active confinement through SMA were investigated analytically on a RC bridge column retrofitted with SMA spirals. The behavior of the column was studied under cyclic loading and earthquake excitations. Active confinement improved the seismic behavior of the columns in terms of strength, effective stiffness, and residual drifts. Early increase in concrete strength associated with using active confinement delayed the damage experienced by both concrete and steel (Andrawes et al. 2010).

Active confinement, and its subsequent beneficial effects, has also been achieved by means of prestressed metal strips. Recently, posttensioned metal straps were used to retrofit a full-scale reinforced concrete structure tested in shake table (Garcia et al. 2014). Also, unbonded high-strength steel wire was used to passively confine concrete cylinders. Typical curves for prestressed straps or high-strength wire spirals are illustrated in Fig. 2.

Nonmetallic Reinforcements and Resins

Nonmetallic reinforcements consist of continuous organic or inorganic fibers, embedded in a matrix (*fib* Model Code 2010). Fiber-reinforced polymers (FRPs) are nonmetallic reinforcing fibers embedded in a polymeric matrix. Typical fibers have relatively high strength and high modulus of elasticity. The polymer transfers the load among the fibers and protects them against abrasion or aggressive environments. The polymeric matrix has relatively low strength and low modulus of elasticity. The most common resin systems used in FRPs for structural applications are the thermosetting epoxy and vinylester polymers. The glass transition temperature, T_g , of the impregnating polymers (typically far lower than the corresponding resistance at elevated temperatures of the fibers) and of the bonding adhesives is of great importance. T_g of the polymer controls the performance of the FRP system as a whole, at elevated temperature, especially in bond-critical applications. After T_g temperature the polymer falls into a plastic state, it softens, and the mechanical properties degrade (especially the flexural and bond strength of the FRP are affected).

Principal fibers are made of carbon, glass, and aramid (*fib* 14 2001). Recently, basalt (inorganic fibers) and PBO (polyphenylene benzobisoxazole) FRPs are used in retrofit and strengthening applications (Hollaway and Teng 2008; *fib* Model Code 2010) as well as PEN and PET FRPs. Other interesting emerging thermoplastic fibers are ultrahigh-molecular-weight (UHMW) polyethylene fibers and polyvinyl alcohol (PVA) fibers for FRP bars and FRP sheets (Bank 2006). Most of them present linear elastic behavior up to failure (see typical curves of different cured FRPs in Fig. 3), while some present divergence from linearity (bilinear behavior for PET, PEN). Different carbon fibers exhibit modulus of elasticity ranging between 215 and 800 GPa, tensile strength between 1,800 and 6,000 MPa, and ultimate strain between 0.20 % and 2.30 %. Aramid fibers have modulus of elasticity ranging between 70 GPa and 130 GPa, tensile strength between 2,750 and 4,100 MPa, and ultimate strain between 2.40 and 5.00 %. Common glass fibers (high electrical resistance) in retrofit and strengthening are subdivided in E-glass, S-glass (stronger and stiffer than E-glass), and alkali-resistant (AR) glass. Glass fibers have modulus of elasticity ranging between 70 and 90 GPa, tensile strength between 1,900 and 4,800 MPa, and ultimate strain between 3.00 % and 5.70 % (*fib* 14 2001; Hollaway and Teng 2008). The mechanical properties of basalt fibers range between 89 and 95 GPa for the modulus of elasticity, between 3,000 and 4,900 MPa for the tensile strength and between 3.00 % and 5.00 % for the ultimate strain. The mechanical properties of different reinforcing fibers are gathered in Table 1. The mechanical properties of the FRP (fibers and polymer) depend on the fiber volume fraction, the orientation of the fiber, the method of manufacture, the temperature and duration of the cure cycle, and the age of the polymer composite (Hollaway and Teng 2008). The fiber volume fraction in most FRP bars is estimated between 50 % and 60 % and for FRP strips between 60 % and 70 % (Bank 2006). Carbon FRP systems present high resistance to alkalinity/acidity exposure and to creep rupture and fatigue. Glass FRP systems present thermal expansion coefficient similar to concrete. Glass and aramid FRPs are excellent insulators and

Retrofitting and Strengthening of Contemporary Structures: Materials Used, Table 1 Typical mechanical properties of reinforcing fibers

		Modulus of elasticity (GPa)	Tensile strength (MPa)	Ultimate strain
Fibers	Carbon fiber	215–300	3,500–6,000	0.015–0.023
	High modulus carbon	370	3,500–4,410	0.0095–0.012
	Ultrahigh modulus carbon	570–800	1,800–2,300	0.002–0.0166
	Glass	70–90	1,900–4,800	0.03–0.057
	Aramid	70–130	2,750–4,100	0.024–0.05
	Basalt	89–95	3,000–4,900	0.03–0.05
	PBO	270	5,200–5,400	0.01–0.02
	PEN	22–27	790–1,030	0.045–0.05
	PET	6.7–18	740–920	0.07–0.138
	Vinylon	16	735	0.046
	Polypropylene	2	400	0.20
	UHMW polyethylene	175	2,400	0.019

present high impact tolerance. Basalt fibers are derived from volcanic deposits in a single-melt process. They offer better thermal stability, heat and sound insulation properties, vibration resistance, as well as durability than glass fibers.

FRPs in retrofit and strengthening are met in the form of pre-cured bars and laminates as near-surface mounted (NSM) reinforcement. FRP tendons are used in external posttensioning. They are often applied without duct. FRP strips, laminates, profiles, sheets or fabrics, and filaments are used as externally bonded reinforcements. FRP sheets or fabrics are applied by wet layup or pre-impregnated (prepreg). FRP prestressing systems have been developed in the form of wires, strands, bars, or plates. Various different nonmetallic reinforcing elements are shown in Fig. 5. Design recommendations and codes include the use of FRPs for static and some of them for seismic actions (*fib* Model Code 2010; *fib* 14 2001; JSCE 2001; ACI440-08 2008; CSA S6-06 2006; *fib* 24 2003; CNR-DT 200/2004 2004; TR55 2012 among else). Furthermore, FRP grids and fabrics may be embedded in shotcrete or mortar jackets.

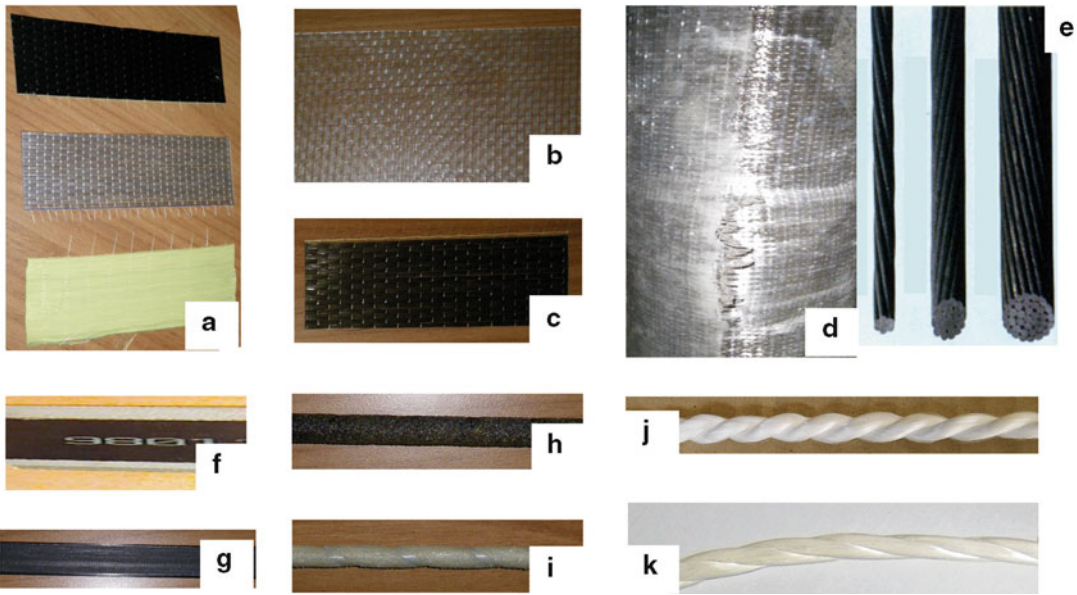
The aforementioned FRP products concern mainly reinforced concrete structures. Yet, carbon FRP plates (with or without prestress) and wraps have been widely used for the retrofit and strengthening of steel structures (increase of member and connection structural capacity,

fatigue crack control, relieve of permanent stresses, corrosion protection). Furthermore, carbon, glass, and aramid FRP plates, grids, and sheets or glass FRP bars (NSM or internal) have been widely used for the retrofit and strengthening of masonry structures (walls, piers, arches, and vaults), bonded with epoxy or cement-based materials. Finally, carbon and glass FRPs have been used in timber structures since the 1960s (Hollaway and Teng 2008).

Different types of resins (mainly epoxy based) are used widely in concrete repair as injected resins for capillary cracks or cracks with higher width or as protective coatings.

Nonmetallic Materials Under Earthquake Excitations

The early systematic research by Priestley et al. (1996) and Seible et al. (1997) in FRP-retrofitted reinforced concrete columns has served as a solid basis for further seismic retrofit developments. Also, it revealed some of the key features of the use of FRPs over conventional techniques. As already discussed, FRPs possess high tensile strength as well as tensile modulus of elasticity and tensile strain at failure that varies among different materials. FRP jacketing has been proven to be a very advantageous alternative to reinforced concrete jackets or steel jackets or steel caging for the following cases of local seismic strengthening of existing concrete members (within certain



Retrofitting and Strengthening of Contemporary Structures: Materials Used, Fig. 5 Different reinforcing fibers and FRPs: (a) unidirectional carbon, glass, and aramid fiber sheet, (b) transparent bidirectional glass FRP sheet after curing, (c) basalt fiber sheet, (d)

fractured glass FRP confining jacket, (e) carbon FRP tendons, (f) carbon FRP laminate, (g) carbon FRP strip for near-surface mounting (NSM), (h) carbon FRP bars, (i) glass FRP bars, (j) polypropylene fiber rope, (k) nylon rope

case-dependent limits): (a) enhancement of the compressive axial strain ductility of concrete through confinement (transverse fiber reinforcement), (b) delay of the premature buckling of the longitudinal steel bars through confinement, (c) upgrade of inadequate lap splices of longitudinal steel bars through confinement, and (d) enhancement of the shear capacity with externally bonded FRPs (transverse fiber reinforcement). In all the above cases, FRPs provide adequate lateral restriction against evolution of undesirable concrete cracking through their unique elastic tensile behavior up to failure. That is, if the flexural capacity of the member (mainly column or wall) is adequate in redesign, then FRPs can prevent brittle shear failure of concrete and loss of bond between steel lap-spliced bars and concrete, in order to develop their full tensile yield stress. They can enhance slender bars in order to develop adequate compressive stress. Having suppressed the abovementioned premature failures and by enabling increased compressive concrete strains, the member can meet the section curvature ductility and the corresponding member deformation

requirements of redesign. FRP shear strengthening of beams or of external beam–column joints or repair of damaged ones with injective resins is also an alternative – within certain limits in any case. In all the above cases, the FRP reinforcement is exposed primarily to variable cyclic tensile deformation in the direction of its fibers, through the polymer–concrete interface. Furthermore, FRP is under compression through its bond with concrete interface. Compression is mainly sustained by the polymer matrix that should maintain its integrity to favor the performance of the tensioned fibers. Seismic flexural strengthening requires the FRP jacket to go through the beam–column connection. Besides, fibers along the axis of the member and under compression require special detailing. Generally, their contribution is neglected, while local buckling should be prevented. Thus, most of the guidelines do not refer to seismic FRP flexural strengthening (EN1998-3) or prevent flexural FRPs when under compression because of cyclic flexure or accidental actions if no measures for local buckling are taken. CNR-DT 200/2004 also includes seismic

FRP flexural strengthening with provisions on anchorage of FRP and proper confinement to avoid fiber delamination and concrete spalling under cyclic loads. In most cases, the use of FRP concerns local seismic retrofit and strengthening as transverse reinforcement. The increase of existing members' sections and of their elastic stiffness because of the use of FRP jackets in these cases is negligible. Thus, FRPs can be applied locally without affecting the distribution of the developed internal forces within the existing three-dimensional structure and can provide structure with significant ductility. Usually, seismic strengthening of existing structures requires also horizontal displacement control through added stiffness or strength enhancement. Therefore FRPs are used in combination with other materials and intervention techniques at a global level (reinforced concrete jackets, new reinforced concrete walls, etc.). Furthermore, heavily damaged reinforced concrete walls (under lateral load reversals) can be retrofitted with properly designed and anchored FRPs to achieve enhanced flexural and shear capacity (Antoniades et al. 2007). Finally, FRP jackets have been proven an effective measure of repair for corrosion-damaged columns (Pantazopoulou et al. 2001).

In general, the discussion for steel-confined concrete and the behavior of bars under earthquake excitations is valid for FRP-confined reinforced concrete as well. However, while ordinary steel transverse reinforcement can only provide a reduced lateral restraint to the dilating concrete after yielding, concrete confinement by FRP increases remarkably stress and especially strain of concrete at failure. This is because FRPs exhibit a linear elastic tensile behavior up to their failure. Therefore, contrary to steel confinement, FRPs exert an ever-increasing lateral stress up to their failure and may restrict, alter, or even prevent concrete dilation up to failure. In general, the monotonic behavior of FRP-confined concrete may also serve as the envelope curve of its cyclic behavior (Rousakis et al. 2008 among others), having in mind mainly the issues raised for the longitudinal bars' buckling as well as force transfer issues through laps and strain gradient in sections subjected to cyclic bending.

Advantages of Nonmetallic Reinforcements over Conventional Retrofit and Strengthening

The main advantages of FRP reinforcements that favor their increasing use over conventional materials in common seismic retrofitting are (e.g., *fib* 14 2001; Hollaway and Teng 2008; Fardis 2009 among others):

- (a) High strength-to-weight ratio. This leads to structural jacket with a negligible added thickness to the original section. It does not add significant additional dead load to the structure and thus does not increase load bearing demands. It enables for selective local (or global) interventions without adding uncertainties or altering the distribution of internal forces in the structure. It maximizes the net floor plan area when applied to vertical members.
- (b) Continuous and flexible structural fibers impregnated and bonded with polymers in situ. In combination with (a) it provides an easy to handle reinforcement that can follow the form and shape of the existing members. It can be used in multiple layers, in different directions to satisfy design needs and form a composite reinforcement after resin curing. Polymer-bonded jacket allows for direct and full strain compatibility all over the jacket surface and increases the effectiveness in bond-critical applications. It allows for easy laps and self-anchoring. Continuous fibers lead to the minimum material waste during installation:
- (c) Compatibility with other retrofitting techniques in cases of combined intervention.
- (d) FRP is a low-maintenance material and can be used to halt corrosion in existing members.
- (e) Immunity to corrosion of the FRP itself.
- (f) All above advantages may lead to minimized disruption of use during installation, minimized labor, and other costs.

Furthermore, FRPs may be unique solution in special cases of demanding interventions such as fully reversible ones or when transparency of substrate should be maximized. Finally, special design requirements may oblige for the

use of specific FRP systems because of their unique physical properties.

Recent Advances in Nonmetallic Reinforcements and Resins

Numerous novel nonmetallic materials have been employed in seismic retrofit and strengthening of concrete members combining different technologies and desirable properties and mechanisms (Fardis 2012). Advanced polymer-modified cement-based mortars have been used as matrix for textiles (TRM, textile-reinforced mortars) offering significant structural capacity upgrade and resistance at elevated temperature (Fardis 2009 among others). Recently, TRMs have been used as confining reinforcement in columns together with stainless steel or FRP near-surface mounted (NSM) flexural strengthening. Significant flexural strengthening under seismic loading was achieved by NSM or externally bonded longitudinal FRPs, provided proper confinement with FRP or TRM jackets that may prevent local buckling of compressed longitudinal FRP (Bournas and Triantafillou 2009 among others).

Another issue of concern is the fire performance of FRPs. Upgraded fire resistance of NSM FRP flexural strengthening applications in beams loaded up to their service load could be achieved. The strengthened beams had proper insulation systems and detailing to improve the performance of the polymer adhesive or expansive mortar was used as bond adhesive (Palmieri et al. 2013). Other applications utilize heat-resistant polymer matrices to enhance the performance of FRP reinforcements.

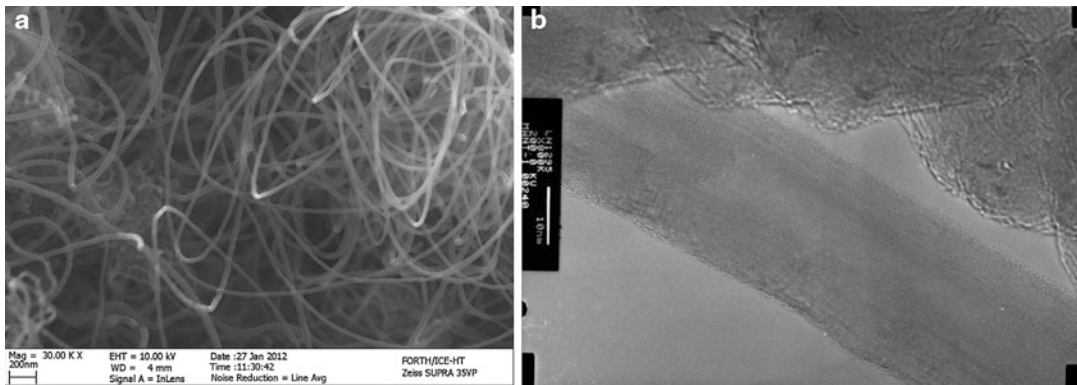
Active confinement with FRPs (filament, straps, or composite shells) is a viable alternative to metal ones as presented in numerous applications. Beneficial effects of unbonded prestressed FRPs on shear strengthening of beams have been identified also in Lees et al. (2002).

Recently, high-deformability materials (PEN, PET, vinylon, polypropylene, etc.) were explored as transverse reinforcement in seismic strengthening of columns. They were used as reinforcing fibers of FRPs to upgrade the axial strain ductility of concrete columns under axial or seismic loading (Anggawidjaja et al. 2006; Dai et al. 2011).

Other applications concerned external confinement of columns with high-deformability continuous fiber ropes or tapes that were unbonded and non-impregnated (“green” perspective). The tapes and ropes were used alone or in combination with FRPs as passive (Rousakis et al. 2014) or active confinement (Rousakis and Tourtouras 2014). Vinylon and polypropylene ropes are presented in Fig. 5j, k. Furthermore, multiwall carbon nanotubes (MWCNTs in Fig. 6) have been used to enrich epoxy resins used for glass fiber sheet impregnation or as injected resins in concrete crack repair (Rousakis et al. 2014). The use of CNTs in resins raises not only their mechanical properties but is expected to upgrade their temperature resistance as well.

Summary

The chapter presents the main materials used in retrofit and strengthening of contemporary structures. The materials that are already incorporated in design recommendations and design codes are the core of the presentation. Recent advances are also discussed in order to identify the main trends, urgent needs, and potential future emerging materials. The materials of the chapter mainly concern the retrofit and strengthening of reinforced concrete structures such as buildings and bridges, while references to existing structures made of other materials are available. The materials are classified according to the technology of their basic constituent. The three main material classes reviewed in this entry include concrete and cement-based materials, steel and metal-based material, and nonmetallic reinforcements. The chapter presents the basic properties of the materials and their performance in retrofit and strengthening applications with respect to the different techniques and scopes of use in design. The discussion is focused on mechanical behavior, interaction with other existing materials, and open issues from a materials’ point of view to ease the understanding of chapters to follow. The variety of different materials available reveals a trend toward performance-based solution per specific application (more advantageous than others) in accordance with the



Retrofitting and Strengthening of Contemporary Structures: Materials Used, Fig. 6 Scanning electron microscopy image of the MWCNTs (a). Transmission

electron microscopy image of the nanotube (b) (Adapted from Rousakis et al. 2014)

specific design requirements. The basic principles of efficiency, safety, and reliability of retrofit are better served through more specialized materials, as earthquake-resistant structures may have to overcome potential overloads, in order to avoid collapse and human loss.

Cross-References

- ▶ [Assessment of Existing Structures Using Inelastic Static Analysis](#)
- ▶ [Assessment of Existing Structures Using Response History Analysis](#)
- ▶ [Nonlinear Dynamic Seismic Analysis](#)
- ▶ [Nonlinear Finite Element Analysis](#)
- ▶ [Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used](#)
- ▶ [Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions](#)
- ▶ [Seismic Collapse Assessment](#)
- ▶ [Strengthening Techniques: Bridges](#)
- ▶ [Strengthening Techniques: Code-Deficient R/C Buildings](#)

References

- ACI Committee 440.2R-08 (2008) Guide for the design and construction of externally bonded FRP systems for strengthening concrete structures. American Concrete Institute, Farmington
- Andrawes B, Shin M, Wierschem N (2010) Active confinement of reinforced concrete bridge columns using shape memory alloys. *ASCE J Bridge Eng* 15(1):81–89
- Anggawidjaja D, Ueda T, Dai J, Nakai H (2006) Deformation capacity of RC piers by new fiber-reinforced polymer with large fracture strain. *Cem Concr Compos* 28(10):914–927
- Antoniades KK, Salonikios TN, Kappos AJ (2007) Evaluation of hysteretic response and strength of repaired R/C walls strengthened with FRPs. *Eng Struct* 29(9):2158–2171
- Bank LC (2006) Composites for construction: structural design with FRP materials. Wiley, Hoboken, p 551. ISBN 0-471-68126-1
- Bournas DA, Triantafillou TC (2009) Flexural strengthening of reinforced concrete columns with near-surface-mounted FRP or stainless steel. *ACI Struct J* 106(4):495–505
- Canadian Standard Association (CSA) (2006) Canadian highway bridge design code, CAN/CSA-S6-06. Canadian Standards Association, Rexdale, p 733
- Cho C, Kim Y, Feo L, Hui D (2012) Cyclic responses of reinforced concrete composite columns strengthened in the plastic hinge region by HPFRC mortar. *Compos Struct* 94(7):2246–2253, ISSN 0263–8223
- CNR-DT 200/2004 (2004) Guide for the design and construction of externally bonded FRP systems for strengthening existing structures. National Research Council, Advisory committee on technical recommendations for construction, Rome
- Dai J, Bai Y, Teng JG (2011) Behavior and modeling of concrete confined with FRP composites of large deformability. *J Compos Constr* 15(6):963–973
- Eurocode 8 (EC 8) (2005) Design of structures for earthquake resistance – part 3: assessment and retrofitting of buildings. EN 1998–3: 2005, 2005
- Fardis MN (2009) Seismic design assessment and retrofitting of concrete buildings. Springer, Dordrecht

- Heidelberg London New York <http://link.springer.com/book/10.1007%2F978-1-4020-9842-0>
- Fardis MN (2012) Innovative materials and techniques in concrete construction. ACES workshop. Springer, Dordrecht Heidelberg London New York <http://www.springer.com/engineering/civil+engineering/book/978-94-007-1996-5>
- FEMA 547 (2006) Techniques for the seismic rehabilitation of existing buildings, 547/2006 Edition. FEMA, Washington, DC
- fib* (2003) Seismic assessment and retrofit of RC buildings. *fib* Bulletin 24. Federation Internationale du Beton, Lausanne
- fib* Bulletin 14 (2001) Externally bonded FRP reinforcement for RC structures. *fib* Bulletin 14, Lausanne, (convener of TG 9.3 Triantafyllou A)
- fib* Model Code 2010 (2013) *fib* model code for concrete structures 2010. Wiley-VCH Verlag GmbH & Co. KGaA Germany <http://www.ernst-und-sohn.de/en/fib-model-code-for-concrete-structures-2010>
- Garcia R, Hajirasouliha I, Guadagnini M, Helal Y, Jemaa Y, Pilakoutas K, Mongabure P, Chrysostomou C, Kyriakides N, Ilki A, Budescu M, Taranu N, Ciupala MA, Torres L, Saiidi M (2014) Full-scale shaking table tests on a substandard RC building repaired and strengthened with post-tensioned metal straps. *J Earthq Eng* 18(2):187–213
- Hollaway LC, Teng JG (2008) Strengthening and rehabilitation of civil infrastructures using fibre-reinforced polymer (FRP) composites, vol 25, Woodhead Publishing Limited, Great Abington, Cambridge, England www.woodheadpublishing.com
- JSCE (2001) Recommendation for upgrading of concrete structures with use of continuous fiber sheets, vol 41, Concrete engineering series. Japan Society of Civil Engineers, Tokyo
- Lees JM, Winistoerfer AU, Meier U (2002) External prestressed carbon fibre reinforced polymer straps for shear enhancement of concrete. *J Compos Constr* ASCE 6(4):249–256
- Palmieri A, Matthys S, Taerwe L (2013) Fire endurance and residual strength of insulated concrete beams strengthened with near-surface mounted reinforcement. *J Compos Constr* 17(4):454–462
- Pantazopoulou SJ, Bonacci JF, Sheikh S, Thomas MDA, Hearn N (2001) Repair of corrosion-damaged columns with FRP wraps. *J Compos Constr* 5(1):3–11
- Priestley MJN, Seible F, Calvi GM (1996) Seismic design and retrofit of bridges. Wiley, New York
- Rousakis TC, Tourtouras IS (2014) RC columns of square section – passive and active confinement with composite ropes. *J Compos B Eng* 58:573–581
- Rousakis TC, Karabinis AI, Kioulos PD, Tepfers R (2008) Analytical modelling of plastic behaviour of uniformly FRP confined concrete members. *Elsevier J Compos B Eng* 39(7–8):1104–1113
- Rousakis TC, Kouravelou KB, Karachalios TK (2014) Effects of carbon nanotube enrichment of epoxy resins on hybrid FRP – FR confinement of concrete. *J Compos B Eng* 57:210–218
- Seible F, Priestley NM, Hegemier GA, Innamorato D (1997) Seismic retrofit of RC columns with continuous carbon fiber jackets. *ASCE J Compos Constr* 1(2):52–62
- The Concrete Society TR55 (2012) Design guidance for strengthening concrete structures using fibre composite materials, 3rd edn, The Concrete Society, Blackwater, Camberley, Surrey
- Tsakiris SA, Rousakis TC, Karabinis AI (2012) Confinement effects and strain transfer in reinforced concrete jackets of different detailing for the strengthening of old-type concrete columns. In: 15th world conference on earthquake engineering, 24–28 Sept 2012, Lisbon

Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions

T. P. Tassios

Department of Structural Engineering –
Reinforced Concrete Laboratory, National
Technical University of Athens (N.T.U.A.),
Zografou, Athens, Greece

Introduction

Preamble

An engineering structure may be partly interpreted as an **organism**, since it comprises various mutually supported members; besides, these members undergo changes in time, due to internal (e.g., corrosion) and/or external causes (e.g., differential settlement). Thus, the way structures are designed, constructed, and maintained could profitably be inspired by scientific branches of medicine applicable to living organisms, such as anatomy, physiology, obstetrics, pediatrics, pathology, surgery, gerontology, etc.

Traditionally, however, structural engineering used to limit its interests only up to a level equivalent to the “birth” of an organism, as if it were supposed to be “invulnerable” (to aging and to environmental effects) and, therefore, “immortal.” Nevertheless, some empirical rules were

developed in time, regarding retrofitting and strengthening of structures damaged due to a variety of causes. Such rules, however, were not always reliably resulting in economical and safe solutions. The need to rationalize these methods of structural intervention became more than apparent during the last decades of the twentieth century.

Thus, structural engineering moved further along the analogy of medical branches to develop new sectors contributing to a better understanding (i) of pathological causes as well as (ii) of the mechanical behavior of damaged regions of structural members.

Historical Background

Still, long before the above discussed relatively recent rationalization of the field, empirical repair of structures has been taken place:

- Some emblematic examples from ancient Greek history show that the need for structural repair and strengthening after various pathological events was deemed essential.

The most emblematic example is perhaps the rehabilitation of the great temple of Zeus at Olympia, after an **earthquake** that damaged the temple during the second century BCE. Professor Dinsmoor (1941) has convincingly proved that at least eight columns located at the NE and the SW corners of the temple were damaged: “A disastrous earthquake left the columns standing, though with their drums jutting outward and inward. The only method of overcoming this irregularity is to dismantle the columns as low down as the root of disturbance and then to rebuild them” (p. 415). And it is very interesting to observe today the construction traces of this huge intervention, although all columns are lying on the ground. Previous milder earthquakes have produced on the same monument dangerous openings of vertical joints between stone elements; their repair was made by means of “hook clamps, placed horizontally across the vertical joint” (p. 405).

Another historical structural intervention is the rehabilitation of the temple of Erechtheion

on the Acropolis of Athens. After a disastrous **fire** (first century BCE), thermal fracturing of several building elements (e.g., a heavy lintel over the north doorway) required their taken down in pieces, and, upon appropriate propping of the overlaying part, their replacement by new ones slid in position. To this aim, grooves were carved on the underlying surfaces, and small diameter iron spheres were interposed to facilitate horizontal sliding of the new lintel into its final desired position (Korres 1997; Caskey et al. 1927):

- After the Second World War, repair and strengthening of **bridges** was intensified. Nevertheless, only traditional materials were used, such as concrete and steel, whereas redesign was based on rather simple calculations.

New materials (such as composite materials, thin steel plates, shotcrete etc.) and new intervention methods (such as seismic isolation, buckling restrained braced systems, etc.) were developed after the 1970s.

The most recent development in the field seems to be the recognition of the importance of the force transfer mechanisms along the old-to-new material interface.

However, it must be admitted that extensive and well-organized efforts in several countries toward preseismic strengthening of existing public buildings were not always materialized: economic and social costs (including occasional shortsightedness) decelerate the process.

Principles of Structural Intervention

In view of the complexity of an actual or a potential seismic damage, structural interventions risk to be ill-conceived (both in terms of safety and economy), unless a **scientific approach** is followed to the extent that available knowledge and data do allow for. To this end, some fundamental rules of broad applicability (“principles”) should be adopted, regarding the three well-known stages, i.e., of assessment, redesign, and construction (repair or strengthening). Thanks to these principles, it is hoped that the entire process of structural intervention will be rationalized, resulting in more safe and economical repair

and strengthening of existing structures, as well as in avoidance of gross errors.

In what follows, several basic principles, to be elaborated later in the text, are briefly introduced:

- **Knowledge first**, action next: The entire bearing system and its possible environmental and man-made damages of an existing structure must be known in detail, before any intervention. The designer should also have a clear knowledge of the pathological causes (and their mechanisms) of existing or potential future damages. Besides, apart from the knowledge of the available resistance to static loads (i.e., strength and stiffness) of structural elements, their ductility capacity should also be known in the case of a seismic intervention.
 - **Consistency to the level of reliability by which action effects and structural capacity are determined**: Poor knowledge of structural properties is not compatible with sophisticated design methods. Similarly, current advanced analytical tools (for the determination of action effects) can only be used when the properties of structural members (e.g., member dimensions) are known within an acceptable level of confidence.
 - The error of “monolithism”: Forces along an “old-to-new” material interface can be mobilized only when a relative displacement occurs along this **interface**.
 - The different **loading histories** of existing and strengthened structural members should be taken into account. In doing so, upon a damaging earthquake, an unloading of the existing element will be considered, whereas added materials for repair or strengthening will remain free of stress until a new earthquake happens.
 - Redesign of existing structures should be carried out with a differentiated target **reliability level**. Consequently, different partial safety factors are usually applied, both for actions and for added materials.
 - Any seismic upgrading of an existing structure is welcome instead of doing nothing, waiting for better future opportunities. This social requirement should however be satisfied in a more or less systematic way, by selecting an appropriate and feasible “**redesign target**” each time.
 - A **conceptual** redesign should precede any quantification of structural intervention measures. Similarly, **maintenance** against environmental effects seems to be the first step of any seismic upgrading.
 - Several intervention strategies are offered, regarding the extent of the intervention (“to the members or to the system”), as well as the question “resistance or ductility upgrading.” An optimization is sought in every case: the multiple costs should be minimized, including the disproportionately high social costs for the relocation of tenants, interruption of professional activities, etc.
 - Repairs or strengthening should contribute toward a uniform distribution of the **margins of safety** along the height of the building under structural retrofitting. Thus, the probability is increased of numerous building elements to dissipate seismic energy, without a disproportionately high ductility to be locally required.
 - Numerous and rapidly developing intervention methods (some of them being rather sophisticated) should be thoroughly known to designers; the traditional schism “**designer/constructor**” does not seem to be promising.
 - Upgrading of monuments against earthquakes necessitates the application of additional principles, because of the possible adverse effects of structural measures against the so-called monumental values, i.e., aesthetics, historicity, etc.
 - The quality of the end product of any technology may equally be jeopardized by deficiencies in conception, in design, in construction, and/or in maintenance. In the field of aseismic structural intervention of existing structures, construction phase is much more subtle and sophisticated than in the case of new structures. Thus, field inspection in this case should be meticulous and well organized; this need should always be recognized.
- Among others, the above “principles” proved to be influential in code development and in

structural engineering practice during the past few decades. Therefore, it is deemed useful to be further elaborated in the remainder of this article.

It is maintained that despite their seemingly theoretical nature, the contents of this article may be of considerable practical importance.

“First Understand, then Withstand”

Preamble

- (a) In the case of **strengthening** of a structure, full knowledge of the existing weaknesses of the structure is needed, before undertaking any intervention. Such knowledge is only feasible if adequate and reliable data is available and properly understood, such as:
- A thorough documentation of the existing structure
 - An appropriate post-elastic analytical model, able (be it in a rough way) to reproduce analytically the observed pathology of the structure
 - Strength, stiffness, and ductility of all structural elements under imposed cyclic actions

A reliable assessment of the structure will then be feasible, and a valid prediction of its behavior under a targeted “design” seismic action can follow.

In the case of simple **repair** of a damaged structure, although certain structural weaknesses may have already been revealed, mapping the complete picture of its structural condition should be pursued. First, an interpretation of the damage morphology is needed in order to understand the failure mechanisms behind it. Second, an estimate of the severity of the seismic actions the structure was exposed to should be reached. Finally, the designer may be obliged to decide if a strengthening is further needed, up to a targeted (design) level of seismic actions, higher than the one that caused the actual damage.

Consequently, for the assessment of structures to be repaired, the data mentioned in the

case of structures to be redesigned (see point (b), here below) need also to be available and well understood:

- (b) After the assessment of its seismic bearing capacity, a structure may be needed to be strengthened. The relevant **redesign** of an existing structure is a distinct process requiring knowledge of further data that should be well understood, such as:
- Appropriate redesign seismic actions
 - Properties of added materials
 - Mechanical behavior of the interface of “existing-to-added” materials
 - Sufficient knowledge of the various available methods of interventions
- (c) In conclusion, a considerably better **understanding** before deciding, before designing, and before constructing is needed in the case of assessment and redesign of a seismic structural intervention, as compared to the case of a new structure. In this regard, certain comments are further included in the following subsections underpinning the general principle “first understand, then withstand.”

“Do Not Trust Magicians”

Several professions may be exercised in the society based on sound experience, without necessarily a deep knowledge of the related phenomena; this may also be the case with empirical design and construction of simple buildings. However, it is unlikely that a safe assessment or a safe and economical seismic strengthening is feasible without a deep theoretical knowledge of several modern branches of structural engineering and substantial subsequent experience. This, one may say, is a matter of professional ethics – but it constitutes a principle of practical importance.

Documenting Is (Part of) Designing

Existing structures, especially buildings, can be viewed as black boxes: a series of important information is frequently missing regarding basic structural data. Thus, documentation constitutes a fundamental part of assessment and redesign, and it should be organized as follows

(see, e.g., Greek Code for Structural Interventions, 2013):

- (a) Data needed for the assessment and the redesign of an existing structure should be systematically collected **separately** for each category, such as structural system and its in-time modification, ground and foundation conditions, environmental conditions affecting durability, geometric data regarding cross sections of bearing and secondary materials, detailing, previous and future uses of the structure and respective loadings, in situ strengths of constitutive materials, previous alterations or damages and their possible previous repair, potential identification of design errors or of construction gross errors, etc.
- (b) The different levels of reliability of data (LRD) should be distinguished, each of them being described in terms of the source and the completeness of available information. Different LRD should be considered for **each** of the basic categories of data needed; e.g., in the case of an existing reinforced concrete (RC) structure, data regarding the geometry of cross sections may be very satisfactory, as opposed to data regarding quantity and quality of reinforcing bars, which is usually less reliable. If this is the case, one cannot consider an “average” LRD; instead, each category of data should be handled in calculations with its appropriate qualification (e.g., EC8-3, Greek Annex)
- (c) Depending on its LRD, **each** category of data should be used in assessment and redesign with possibly different values of partial safety factors (e.g., reduced loading safety factors for dead loads and material safety factors are commonly taken since their LRD level is usually high).

Similarly, the level of sophistication of the adopted methods of analysis should be related to certain minimal reliability levels of input data, in order to avoid the use of inconsistently accurate tools compared to the available data.

Advanced Models of Analysis and Resistance Determination

For the majority of existing seismically vulnerable structures, it is wished that, after a “reasonable” structural intervention and under a critical earthquake, only repairable damage may occur or (at least) collapse will be avoided. Thus, it is expected that several regions of the structure will be found in a post-yield stage. This means that the most appropriate analysis should be **nonlinear**. The use of such analytical models is much more needed in the case of intervention than in the case of new structures. And to this end, typical hysteretic behavior of structural elements should be known.

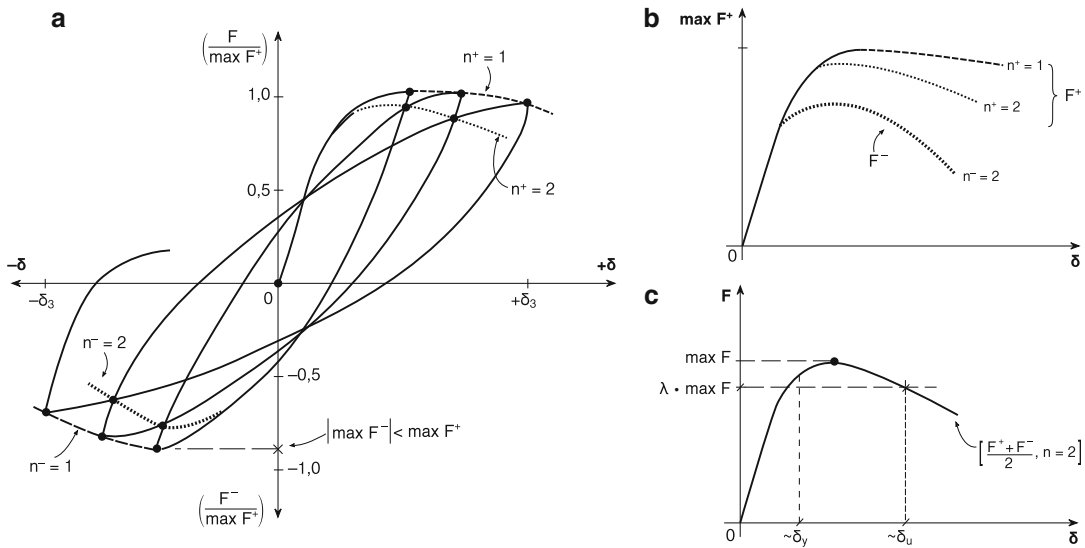
On the other hand, since (legitimately) several regions of structural members will be in a post-yield stage, their force response behavior should be taken into account in the design, under the following conditions:

- Under a targeted imposed post-yield deformation (in other words, under a certain ductility demand)
- After an expected equivalent number, n , of fully reversed deformations imposed by the earthquake (e.g., $n = 2$), this number being a characteristic of the seismic action itself

In Fig. 1 the hysteretic behavior of a structural element is schematically shown. A force response degradation is obviously taking place under cyclically imposed deformations $\pm \delta$. The most appropriate constitutive law “F, δ ” under seismic actions seems to be the one corresponding to $n = 2$; the so-called envelop (corresponding to $n^+ = 1$) has not much physical sense under cyclic seismic conditions. Real force response for $n = 2$ is considerably lower than for $n = 1$; moreover, available ductilities for $n = 2$ (see Fig. 1b) are very much smaller than for $n^+ = 1$. Thus, seismic design based on “envelops” ($n^+ = 1$) does not seem to be realistic.

In this respect, a couple of other observations need also be made:

- (i) Force response under “negative” imposed deformations in the schematic example of



Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions, Fig. 1 Schematic hysteretic behavior (a) and simplified skeleton curves (b), depending on the number of full

reversals “n” and the direction of the imposed deformations (+δ or −δ). Average values of force response and indicative ductility are shown in (c)

Fig. 1 is considerably lower (see in Fig. 1b the curves $n^+ = 2$ and $n^- = 2$). Thus, in most cases average values are used.

- (ii) In the cases where specific values of available ductility factors are used, it is customary to define as “conventional failure deformation” “ δ_u ” the value corresponding to a force response level equal to $\lambda \cdot \max F$ (see Fig. 1c), where the value “ λ ” (< 1) to be selected depends mainly on the acceptable degradation of the element under consideration. Besides, it is expected that, thanks to a **redistribution** of action effects in neighboring structural elements, part of the strength loss $(1 - \lambda) \cdot \max F$ of the damaged element may be undertaken by some neighboring elements. Thus, finally, appropriate “ λ ” values depend on the targeted performance and available ductility of the elements, as well as on the redundancy of the structure. However, only conventional λ values are normally used (ranging from 0.85 for RC and steel structures to as low as 0.70 for masonry structures), without requesting explicit verifications justifying these values in every specific design case.

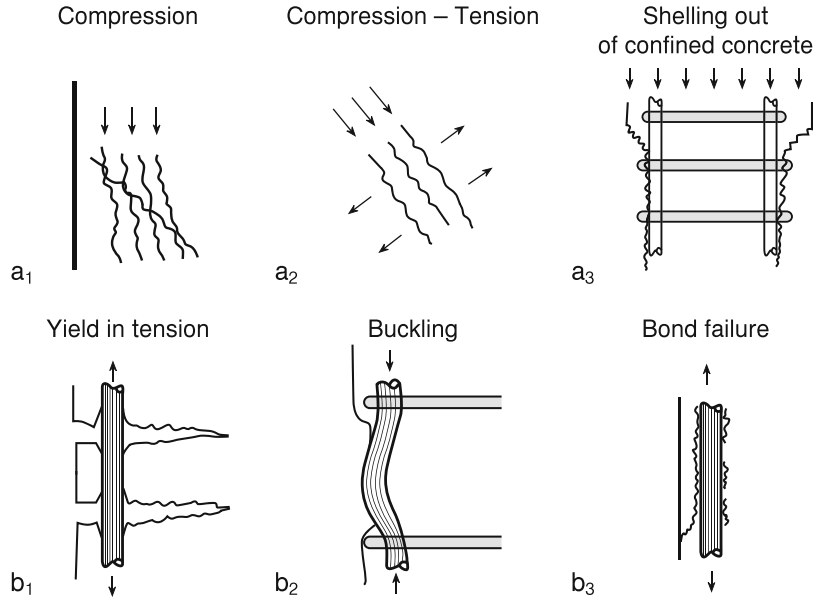
Causes and Mechanisms of Damage

Damages that appear on the structure after an earthquake constitute a sort of “confession” of the weaknesses of the structure. Correct “reading” of this confession is of paramount importance for the seismic assessment of the structure. To this aim, it is useful to identify elementary failure mechanisms (and corresponding damage morphology) attributed to local forces, such as those presented in Fig. 2 for the case of RC structures. Based on such so to say “vocabulary” of elementary damages, more complex damage patterns due to a combination of effects may be analyzed and better understood. However, **differential diagnosis** may also be needed in order to avoid misinterpretation of causes of damage with similar morphology. Some typical examples of such a case may be (i) a concrete crack along a steel bar due to corrosion, misinterpreted as a local steel-concrete bond failure, and (ii) a diagonal crack in a wall, due to differential settlement, misinterpreted as a seismic damage.

In conclusion, a series of conditions and situations (such as those indicatively shown in this section) should be well understood by the

Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions,

Fig. 2 Some indicative elementary failure mechanisms of concrete (“a”) and reinforcing bars (“b”) in the case of RC structural elements may formulate a more complicated “vocabulary,” for a sound interpretation of damage mechanisms



designer prior to any assessment or seismic redesign. Thus, in principle, the level of knowledge needed in the field of structural interventions is considerably higher and more complicated than in designing new structures.

Consistent Level of Reliability

Advancements in engineering are not uniform in all fields; e.g., thanks to the rapid developments in numerical and computational methods, available tools for the analysis of structures are significantly richer and more sophisticated, when compared to available models for resistance determination of critical regions. As a consequence, it may occasionally happen to verify the basic safety inequality

$$E > R \quad (1)$$

in which an action effect “E” is determined by means of, say, a dynamic finite element method, whereas the corresponding capacity (or resistance R) was found by means of a poor empirical method. Such a **difference of reliability** between the two sides of the inequality undermines the desirable rationality of the safety verification sought.

In the case of structural interventions, such an undesirable situation may happen frequently. Sophisticated analytical methods are equally applicable in this case, but the capacity (in terms of forces or deformations) of composite cross sections is not always easily modeled. In fact, capacity is commonly roughly estimated by means of oversimplified empirical methods.

It is therefore needed to formulate the following principle: “In safety verifications, methods of Analysis and methods of capacity/resistance determination should be of comparable reliability”. Regulatory documents should offer more specific rules on this matter.

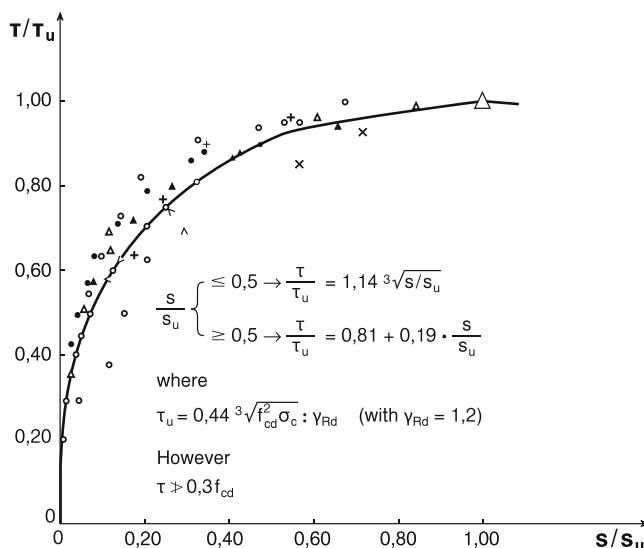
A similar situation of inconsistent level of accuracy in determining action effects and structural capacity may appear in the case of structural interventions in an **inadequately documented** structure, i.e., when:

- Its geometry is not completely known.
- The strengths of its materials were not directly measured.
- Previous damages and repairs were not identified.

If this is the case, the attempted seismic retrofitting may have doubtful results, despite the apparent “precision” of the calculations used

Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions,

Fig. 3 Concrete-to-concrete friction mobilized along a rough interface (after cracking), as a function of the sliding “ s ” normalized by the ultimate sliding “ s_u ” (~ 2 mm for normal concrete) that corresponds to the maximum friction resistance “ τ_u ” under a total normal stress “ σ_c ”



in its redesign. An ensuing principle should be formulated: “Poor documentation is not compatible with sophisticated design methods”.

The Error of “Monolithism”

Some regulatory documents require that, after a seismic structural intervention, “existing and added materials should behave monolithically.” Although (in a qualitative way) this requirement may be understood as a concern for an adequate “collaboration” between two or more different materials, it has however to be reminded that the transfer of shear or tensile forces from the existing (old) to the added (new) material necessitates a certain **local relative displacement** at their **interface**; otherwise, the “reaction” of the added material cannot be mobilized (such local displacements being a “sliding” for shear or a “crack opening” for tensile forces). Consequently, the corresponding constitutive law of structural behavior along the interface should be known, as indicatively shown in Fig. 3 for concrete-to-concrete monotonic friction (Tassios and Vintzileou 1987) or in Fig. 4 for friction resistance under cyclic sliding. By means of such laws, the mobilized forces are calculated as a function of the allowed displacements at the interface, depending on the targeted performance level after the intervention. Thus, the requirement

of “monolithic behavior” should be reformulated as “permissible interface displacement,” resulting in corresponding mobilized resistances of the composite structural element after the intervention.

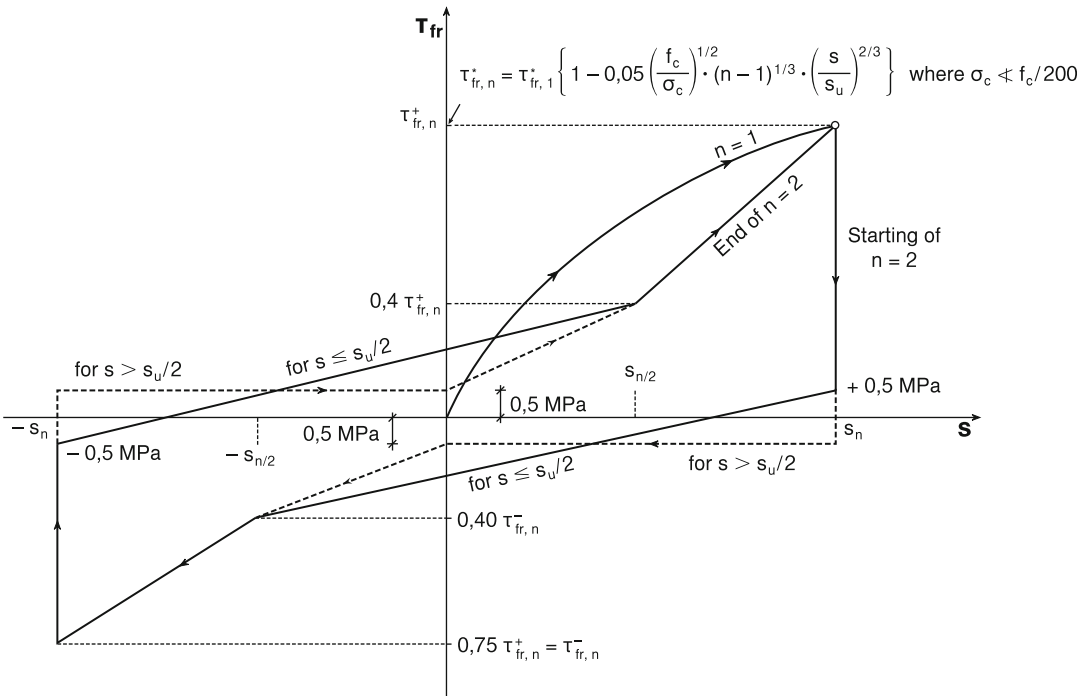
In this connection, however, the **sensitivity** of the interfaces versus environmental actions should be specifically taken into account: moisture penetration, steel corrosion, and possible aging of composite materials should be appropriately considered.

The Principle of the Loading History Differences

It is important to keep in mind that added materials for repair or strengthening are submitted to quite **different loading histories** than those applied to the materials of existing structural elements. This fact should be correctly taken into account during the pertinent redesign.

Introductory Example

As an **introductory example**, consider the load paths followed in the case of an RC column, before and after a critical earthquake, vis-a-vis those followed during a subsequent intensive earthquake acting on the same column strengthened by means of an RC jacket. Let the existing and the added materials be indicatively

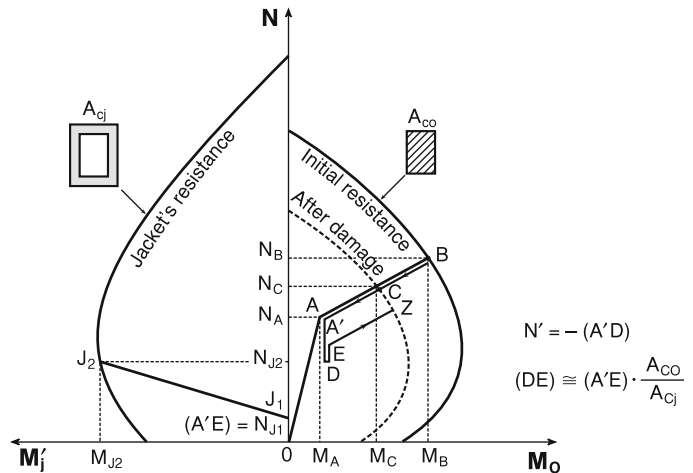


Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions, Fig. 4 A formalistic constitutive rule of friction stress response along

a rough concrete-to-concrete interface under cyclically imposed sliding $\pm S_n$, for “n” cycles

Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions,

Fig. 5 Force paths: (i) of a RC column before (B) and after (C) seismic damage, during its unloading (A'), at its propping (D) and its seismic reloading (Z); and (ii) of a RC jacket after props' removal (J1) and during its seismic reloading (J2)



considered without any connection along their interface. In Fig. 5, the initially critical “M, N” resistance curve of the existing cross section of the considered RC column is shown. The column (of cross section A_{co}) follows the load path OA

under normal loading and the path AB under seismic loading. At point “B,” damage takes place, and the column is instantaneously unloaded up to a point “C” on the residual resistance curve after damage. After the earthquake,

the representative point moves back to point A ($\equiv A'$), if redistribution of action effects is not considered. At this stage, the structural intervention may be initiated: (i) Propping is installed, introducing a small axial load upwards $N' = -(AD)$. (ii) An appropriate RC jacket is constructed (with a cross section A_{cj}). (iii) After its hardening, props are removed, and their force is transferred to the column (DE) and to the jacket (OJ_1).

In case an earthquake occurs upon redesign, the column will follow the EZ loading path, while the jacket will follow (a different) J_1J_2 loading path, up to their respective critical curves. The total seismic demands taken by the composite element read as:

$$\begin{aligned} \text{Seismically induced axial load } (\Delta N)_E &= N_{EZ} + N_{J_1J_2} \\ \text{Seismic flexural moment } M_E &\succ M_Z + M_{J_2} \end{aligned} \quad (2)$$

Such differences in the loading histories of existent and new material as illustrative by the above example should be appropriately considered in dimensioning the composite element. Ideally, they should also take into account the interaction along the interface between existing and added materials.

New RC Walls

The significance of the subject of differential loading histories in the field of structural interventions is also shown in the important case of **new RC walls**, a posteriori constructed or instilled in a multi-bay RC frame building. The considerable axial force acting on the columns embedded within the new wall does **not** act through the entire wall. Notably, the axial forces carried by the original columns are required to avoid large rocking response of the wall and the corresponding drastic reductions of its effective stiffness.

Nevertheless, during such an almost inevitable rocking of “added” walls, the pole of rotation being located eccentrically, the aforementioned encapsulated columns may be slightly lifted up, so that part of their axial loads to be carried by the entire new wall.

In conclusion, the application of the principle of differential loading necessitates a complete knowledge of the stages of the intervention and of its technical details.

Modified Reliability Level

- (a) Theoretically, in the design of new structures, the selection of an appropriate “target reliability level” depends (i) on the cost of losses in case of failure and (ii) on the cost of safety measures. Normally, the acceptable failure probability P_f corresponds to the minimum of a “generalized” cost function accounting for the design, construction, and maintenance capitalized costs, as well as the cost of all types of losses due to failure, multiplied by P_f . In the specific case of existing structures (provided that their use remains the same in future), the costs of safety measures to upgrade their performance are disproportionately high, compared to the costs of equal safety improvement in the case of a new structure. Consequently, the acceptable failure probability to be chosen in redesigning existing structures is expected to be relatively **higher for the same time of exposure** (see point (b) in section “[Better Do ‘Something than Nothing’](#)”).

Such probabilistic considerations are mainly suitable for the **assessment** of the seismic performance of existing structures, particularly in the case of doubts about design or construction errors, occurrence of exceptional incidents, indications of expiration of the designed service life of the structure, and/or a planned change of its use (see RILEM 2001):

- (b) For ordinary structures (normal occupancy buildings and even normal bridges), the verification with respect to a particular limit state (performance level) is made by means of the **partial safety factor format**. But in the case of existing structures, the representative values of the variables and the values of the partial factors need to be appropriately

modified, in order to meet the reliability requirements discussed in the previous point (a) (see also fib Model Code 2010, Sect. 3.3.3.1).

(c) Redesign of permanent loads.

Characteristic values may be reduced provided that dependable estimation of existing dead loads is available.

Partial factors are usually the same as per new structures, although the inherently lower uncertainty could possibly allow for some reduction of these factors.

Redesign of variable actions:

Characteristic values and partial factors need not be modified, **except** for the following cases:

- If the category of use of the structure is not changed but the frequency of its loading is possibly increased because of the intervention itself.
- Seismic actions (considered as variable actions in regions of moderate or high seismicity) should be appropriately selected for redesign, in accordance with the “something-than-nothing” principle of section “[Better Do ‘Something than Nothing’](#)”. In all cases, however, $\gamma_E = 1$.

(d) Design values of material strength:

(d.1) In the case of force-based redesign (linear methods of analysis in several or nonlinear analysis in case of brittle building elements):

- Existing materials.
 - No additional strength measurements are required. Appropriate “default values” are reduced by partial safety factors used in designing new structures.
 - Appropriate additional strength measurements (mainly in situ) should be carried out at building regions of relatively uniform strength of materials (e.g., tops of columns): (i) resistances of critical **cross sections** are calculated with representative strengths in such building regions (local mean value minus

one standard deviation) and reduced as follows. Conversion factor of concrete strength is not needed; thus, “ γ_i ” may be practically taken equal to 1.30. Values of “ γ_s ” for steel are usually not reduced, unless a strict control of the exact position and diameter of bars is carried out, in which case γ_s may be taken equal to 1.05. (ii) Resistances of large critical volumes of existing RC, appropriately strengthened, may be calculated with mean measured or with default strength values.

- In both cases, for lower levels of reliability of data (see point (c) in section “[Documenting Is \(Part of\) Designing](#)”), further higher partial factors should be used, applicable only for the specifically deficient category of data (e.g., insufficiently known anchorage lengths).

• Added materials.

Normal characteristic strength values and partial safety factors are used. However, in case of added very thin concrete layers or in case of any material added in areas of the structure difficult to inspect, γ_M values may be increased.

(d.2) In the case of deformation-based redesign (nonlinear analysis for quasi-ductile elements or linear analysis with verification carried out by means of the method of local ductility factors “m”):

- Measured or “default” values of strengths of existing materials may be used with their **mean** values for quasi-ductile elements and with partial safety factors equal to unity (unless in case of lower level of reliability of data, point (c) in section “[Documenting Is \(Part of\) Designing](#),” when higher factors should be used).

Added materials' strengths in these cases may be reduced with appropriately lower safety factors.

- The case of point (d.1) (ii) above is also valid:

(d.3) Strength of materials in large critical volumes may be used with their **mean** values. Moreover, independent of the volume of critical regions, whenever in a resistance model a factored strength may result in higher resistance values, the corresponding " γ_M " should be taken equal to unity. This is, for instance, the case with the calculation of the mechanical volumetric ratio " ω_w " of confining materials, in which concrete strength appears in the denominator and, if factored, results in a fictitious increase of " ω_w " (by 30 % or 50 %) that is against safety.

(e) Model uncertainty factors.

Design of structural interventions against seismic actions needs to use model uncertainty factors much more frequently than in any other structural engineering field, for various reasons, including:

- In case of previous structural damages, a rather unknown degree of redistribution of action effects has taken place; hence the occasional need for γ_{Ed} values to increase calculated action effects.
- The number of variables entering the models of redesign is larger than in any other case, resulting in higher uncertainty. Besides, force transfer mechanisms along interfaces between "added-to-new" materials show disproportionately higher scattering. It is therefore customary to use γ_{Rd} values to reduce calculated resistance values.
- Feedback information to validate the redesign models used is commonly insufficient. Consequently, relatively higher γ_{Rd} values are used.

It is noted that γ_{Rd} values used in calculating capacity design quantities do not entirely belong to this category of factors.

Better Do "Something than Nothing" (Selection of Redesign Targets)

- (a) The social interest for the seismic safety of existing buildings is better served if the following principle is applied: instead of maximalistic targets of structural intervention, necessitating unavailable funds and very long time of execution, it is better to select a less ambitious redesign target, easier to be implemented (in terms of both investment and social costs).

In fact, the relatively high burden put on the present generation which should supposedly undertake the full strengthening of existing structures in favor of future generations hinders in many cases the implementation of any strengthening. Thus, the risk of social losses due to future earthquakes remains high for rather long periods, even in relatively wealthy seismic regions. That is why it is reasonable to encourage owners to proceed to any economically "feasible" intervention whenever this is possible for them and the local community – instead of the ambition to apply on an existing structure the seismic design rules required for new structures.

- (b) In doing so, however, it is advisable to follow as much as possible a more systematic scheme of redesign target selection, as it is roughly described below.

For the redesigned structure, a lower future performance level may be adopted (see Table 1), provided that this is compatible with the social function of the structure.

On the other hand, as it was discussed in point (a) of section "[Modified Reliability Level](#)," it is reasonable to select a much higher acceptable probability of exceedance P_f of seismic action than those used in designing new structures. If, for instance, a P_f value as high as 50 % is chosen (in a time reference of 50 years for buildings), the seismic ground acceleration for redesign may be reduced approx. by 40 % of the value applicable to a 10 % probability of exceedance.

The combination of a redesign performance level and of an acceptable probability

Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions, Table 1 Redesign performance levels

	A	B	C
Future performance	Uninterrupted function	Life and property protection	Non-collapse
Acceptable damage	Insignificant	Repairable local plastic hinges	Beyond repair

of exceedance constitutes a “**redesign target**.” Lower levels of such targets are easier to be implemented, provided that they contribute to the gradual increase of seismic safety of the existing building stock:

- (c) By definition, an existing structure has consumed part of its **life period expectancy** L_t . Consequently, the duration of its remaining life is reduced, and (for a given probability of exceedance P_f) the value of the seismic action to be used in the redesign of the structural intervention could be reduced, in accordance with the following well-known equations:

$$\begin{aligned} \text{Mean recurring period } T_m &= L_t : [-\ln(1 - P_f)] \\ \text{Ground acceleration } a_g &= \exp(k_1 \log T_m + k_2) \end{aligned} \quad (3)$$

where k_1 , and k_2 are appropriate site-specific constants.

Nevertheless, shorter remaining life periods can rarely be reliably taken into account for the following reasons:

- Conventional L_t values (e.g., 50 years for buildings) are not confirmed in practice; normally, buildings are much longer used, while the “technical” life of bridges may be shorter than anticipated.
- Environmental influences may significantly modify lifetime periods, producing additional uncertainty in any numerical consideration of this issue.

Conceptual Redesign

- (a) Prior to any decision about the kind and the extent of a structural intervention in an

existing building, some basic qualitative rules of conceptual nature should first be considered. Happily enough, these rules are very similar to those applied in the conceptual design of new buildings. A brief list of such basic rules follows below (see also Fardis 2009, Ch. 2). Qualitative decisions and subsequent quantitative redesign of structural interventions regarding seismic upgrading should consider the possibility to **improve** the following characteristics of the building:

- Clarity of the lateral load resisting system, including its simplicity and uniformity
- Contribution to the symmetry and the regularity in plan
- Possible increase of torsional stiffness about a vertical axis
- Possible regularity in elevation of mass and lateral stiffness distribution
- Adequate redundancy
- Reestablishment of the continuity of vertical and lateral force paths, without stress or deformation concentrations
- Contribution to the effectiveness of floor diaphragms at all levels
- Possible reduction of total mass
- Avoidance of possible adverse effects of infills

- (b) Similarly, the first step of any structural intervention should be the evaluation and upgrading of the possible **environmental** influences that directly reduce both the available strength and ductility of materials. As an example, it is reminded here that corrosion of steel bars reduces the available diameter of a bar, it may produce longitudinal cracking to concrete reducing the sufficiency of anchorages (i.e., threatening the possibility of the bar to reach its yielding stress), and finally it may reduce the ductility capacity of steel

itself. Consequently, **maintenance** seems to be the first step of any seismic upgrading (see, e.g., Schiessl 1992).

Selection of Intervention Strategies

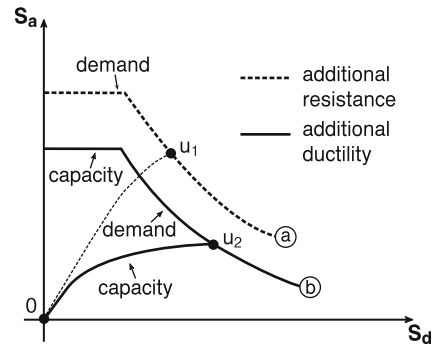
Two categories of such strategies are encountered in the practice of structural intervention. Optimal solutions are sought in both cases.

“Spread-Soft” or “Local-Strong” Intervention?

- (a) Critical regions of a rather large number of building elements are repaired and/or strengthened: this apparently soft method, however, increases disproportionately the social costs of the intervention; tenants should normally be relocated to other buildings, whereas previous professional activities in the building under structural interventions are canceled.
- (b) Instead, additional (external or inserted) walls made of RC or appropriate steel bracings are connected with the existing vulnerable frame in few selected bays and appropriately founded. This alternative intervention reduces considerably the disturbance of the inhabitants but necessitates additional design and construction efforts – especially in connection with problems related to the rotation (or rocking) of the foundation of such new walls.

Additional Resistance (Strength and Stiffness) or Additional Ductility?

Theoretically, for a given performance level, it is possible to satisfy safety requirements by means of upgrading (i) resistance (i.e., strength and stiffness) and/or (ii) ductility. In the first case, in terms of a “pushover” analysis, seismic demand is also increased, whereas the critical displacement (ending the capacity curve) is reduced (see point u_1 in Fig. 6). In the second case, seismic demand is reduced, and the critical capacity displacement is increased (point u_2 in Fig. 6). These two equivalent results however may significantly differ in terms of costs; ductility upgrading may occasionally be more expensive. An **optimum solution** should be sought.

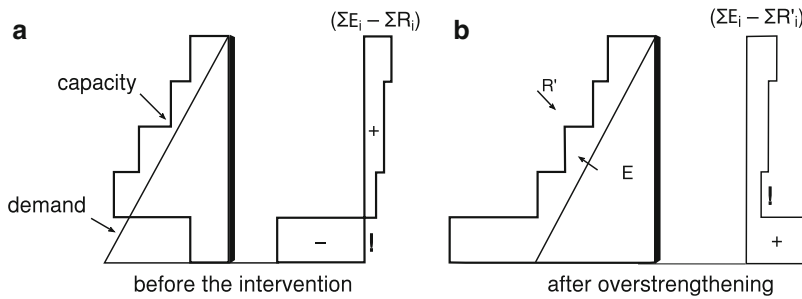


Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions, Fig. 6 In a pseudo-acceleration vs. pseudo-displacement plot, two intervention strategies are depicted, against a given seismic elastic spectrum: (a) upgrading of resistance (u_1) and (b) upgrading of ductility of the system (u_2). Optimization is sought

In-Elevation Uniformity of Safety Margins

Any structural intervention should contribute as much as possible toward a uniform distribution of the final margins of safety (in terms of resistance against lateral loading) along the height of buildings – provided however that a correct provision of seismic action effects (ΣE_i) on each floor is available and a realistic calculation of the resistances of building elements (ΣR_i) is possible. A quasi-uniform distribution of the safety margin ($\Sigma E_i - \Sigma R_i$) offers the possibility that all floors yield almost simultaneously. The latter ensures uniform dissipation of seismic energy along the entire height of the structure. Thus, the formation of “soft” stories is avoided, and the ductility demand at each floor is reduced.

A typical deviation from this principle is related with the ground floor of some multistory buildings in Southern Europe and Southern America: brickwork infills are frequently missing in ground floors (thus creating a soft floor); besides, a strong strengthening of the building elements of this floor relocates “softness” to the above floors, i.e., induces increased probability for this floor to suffer from locally concentrated damage (Fig. 7).



Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions, Fig. 7 (a) Initial soft ground floor ("pilotis"): seismic damage will be

concentrated there. (b) Overstrengthened ground floor relocates "softness" to the above floors

"Designer and Constructor"

In the field of repair and strengthening of structures, a plethora of alternative innovative materials and new construction methods exist. Therefore, it would be inconceivable for a designer to ignore them. Consequently, in this field (more than in designing new structures), the usual dichotomy **designer/constructor** should be blunted. For example, developments in the use of flat jacks to facilitate (and quantitatively check) propping of building elements under intervention should be well known to any designer.

Besides, **interdisciplinarity** in this field is more desirable. For instance, a structural engineer may commit a considerable mistake if they neglect the soil nonlinearity, due to which the rotation center of the foundation of an added wall may lay in a quite eccentric position (see section "[Introductory Example](#)").

Aseismic Upgrading of Monuments

(a) Introduction

- An inadequate structural intervention in a monument violates the value of human life and the survival of the monument.
- An excessive structural intervention violates a series of "monumental values," such as:
 - Aesthetic form
 - Historicity

Further, such an excessive structural intervention may violate the desirable reversibility of restoration measures.

- "Social values" may also be at risk, including the need for economy and for possible new uses of the monument under consideration.
- An optimization should be sought, with the difficulty that these antithetic values are of different nature, and they may not be amenable to a common unit (i.e., money).

(b) Selection of redesign seismic action

In view of this complex situation of "values at risk," it would be wrong to use seismic actions prescribed by codes of practice for the design of new buildings or bridges in seismic redesigning of monuments. Theoretically (for a given conventional reference time), a new probability of exceedance should be sought, taking into account that any additional upgrading of safety would entail "costs" in terms of:

- Money
- Modification of aesthetic form
- Devaluation of historicity
- Reduction of reversibility
- Possible difficulties in new uses

It is expected that such an original increase of "costs" would occasionally result in a considerable increase of the acceptable probability of exceedance of the value of seismic actions that will be used in the redesign of the resistance of a monument.

In more practical terms, however, an interdisciplinary group of experts offers justified

opinions on two or three alternative solutions based on lower and relatively higher seismic loads, depending, i.e., on their consequences on the aforementioned monumental values (see Tassios 2011).

The Retrofitting “Construction Principle”

- (a) The originality and the subtlety of structural interventions are recognized.

In the case of added **structural materials**, the following difficulties are usually encountered, inter alia:

- Precision regarding the forces induced by propping
- Cleanness of interfaces
- Geometrical precision in added thin layers
- Sensitivity of operations such as gluing, welding and the like
- Originality of methods like grouting and injections
- Health protection of personnel working in damaged buildings or handling chemicals in several cases

In the case of installation of dampers, base isolators or active protection devices, construction operations become closer to mechanical engineering.

- (b) Because of these rather delicate conditions, structural upgrading construction-stages need additional and better organised Inspection. Safety is a multidimensional vector with several components; and it may be jeopardised by a deficient Inspection, especially in the case of structural intervention.

Consequently, good seismic upgrading-design files should be accompanied by an appropriately detailed set of technical Specifications and with explicit suggestions for additional Inspection wherever needed.

Summary

This article reviewed the basic principles of structural interventions for the seismic

assessment, repair, and strengthening of existing structures. After a brief historical background, the need for such principles is first described and subsequently several basic views toward a rational foundation of assessment and redesign are presented.

To this end, a basic motto “First understand, then withstand” is established underlining the importance of complete documentation, systematic knowledge of pathological mechanisms, long experience on available traditional and modern methods of retrofitting, as well as familiarity with the use of advanced redesign models. The significance of the displacements along the interface between existing and added materials, as well as the principle of the differential loadings of these two categories of materials, has been subsequently explained.

The problem of the target reliability level in redesigning structural interventions is then discussed in detail, together with the modified values of safety factors. Further, the selection of “redesign targets” (possibly new performance levels and new probabilities of exceedance of the seismic action) is discussed, together with possibly modified remaining life periods of the existing structure.

Finally, alternative intervention strategies are presented, together with the principle of the final uniformity of safety margins in the elevation of the retrofitted building. The entry closes with some considerations regarding the execution of a structural intervention and the role of the designer in this respect.

References

- Caskey D, Fowler HN, Paton JM, Stevens GP (1927) *The erechteium*. Harvard Univ Press, Cambridge Mass
- Dinsmoor WB (1941) An archaeological earthquake at Olympia. *Am J Archaeol* XLV(3). 45:399–427
- Fardis M (2009) *Seismic design, assessment and retrofitting of concrete buildings*. Springer, Dordrecht
- fib Model Code (2010) Ernst u. Sohn
- Korres M (1997) Restoration and reconstruction work on monuments in Antiquity. In: *Proceedings of the La Reintegrazione nel restauro dell’Antico*, Gangemi
- RILEM (2001) Probabilistic assessment of existing structures – JCSS report. RILEM, Paris

- Schiessl P (1992) Repair strategies for concrete structures damaged by steel corrosion. In: Proceedings of the international RILEM/CSIRO/ACRA conference, Melbourne
- Tassios T (2011) Problemi assiologici nel restauro strutturale dei monumenti. Liguori, Napoli
- Tassios T, Vintzileou E (1987) Concrete-to-concrete friction. ASCE Struct Eng 113:832–849

Review and Implications of Inputs for Seismic Hazard Analysis

Lalliana Mualchin

Headquarters: Studio legale Avv. Wania Della Vigna, International Seismic Safety Organization (ISSO), Arsita (TE), Italy

Synonyms

Critique on seismic hazard analysis; Limitations on seismic hazard analysis; Preference of DSHA over PSHA; Problems with probabilistic seismic hazard analysis

Introduction

In earthquake-prone regions of the world, it is a common practice to assess anticipated hazard severity for land-use planning, emergency management, and structural design load considerations for public safety applications. Generally, strong ground motion hazards would impact a large area, and displacement hazards could affect a relatively small area localized along the fault trace of surface-faulting earthquakes. This section will *review* inputs for obtaining a generic ground motion as the most pervasive earthquake hazard and that without site-specific considerations. Note that soft soil sites are generally more hazardous than stiff soil or rock sites, and such site-specific conditions should be characterized realistically in practice for ground response analysis.

Implications of the inputs for seismic hazard assessed results will be discussed. The hazard

assessment is accomplished by performing seismic hazard analysis (SHA) procedures using either a deterministic or a probabilistic approach (Reiter 1990). The deterministic results provide hazards anticipated from the *largest* potential events and the probabilistic from prescribed *recurrent* event(s). The ground motion hazard is obtained from the mean or a certain level above the mean as applied to the ground motion prediction equations (GMPEs) in both the deterministic and probabilistic approaches according to policy or guidelines.

Background

SHA and its results have been well known and used, but the inputs are not well understood by the public, users, and even some analysts, especially those using PSHA software. Analysts and users should understand first the inputs, the approaches, and the limitations in the results before the intended applications.

The deterministic seismic hazard analysis (DSHA) and its enhanced neo-DSHA (NDSHA) define hazards anticipated from a single magnitude of the *largest* potential earthquake or maximum credible earthquake (MCE) scenario centered on each of the considered seismogenic faults (e.g., Mualchin and Jones 1992; Panza et al. 2012), independent of the earthquake recurrence times or the ground motion return periods. The justification is that (1) such a single *largest* earthquake encompasses and thus automatically considers impacts from all other possible smaller events, (2) earthquakes can occur at any time regardless of its short or long recurrence estimated times or low or high estimated slip rates of the faults, and (3) ground motion for MCE would exceed any physically realistic ground motion return periods. In addition, NDSHA provides at each site a set of synthetic seismograms generated for all the earthquake scenarios. The usual criticisms on DSHA are its lack of magnitude recurrence times or ground motion return periods and lack of quantifying uncertainty in the results, but these are truly more of misunderstanding in the practice of DSHA.

The probabilistic seismic hazard analysis (PSHA) determines hazards anticipated from a number of earthquakes on each of the same seismogenic sources as used in DSHA whose magnitudes, however, are defined according to their recurrence times (e.g., Cornell 1968). The justification is that (1) hazards from more or less active faults are “equitably” accounted for through the recurrences or its equivalents, (2) integrating or combining impacts from all significant magnitudes and not just one single MCE event is comprehensive, and (3) uncertainty has been accounted for in each step of the analysis, in the GMPEs, and in the final results. The usual criticism is its lack of transparency in the results with respect to the inputs (due to an elaborate analysis with emphasis on uncertainty) which has created enormous problems. Revealed problems on several aspects and results of PSHA that remain unresolved (e.g., Klügel 2005, 2007; Kossobokov and Nekrasova 2012; Nekrasova et al. 2013; Wang 2011; Wang and Cobb 2012; Wyss et al. 2012) began to accumulate and put the future of the probabilistic approach uncertain.

In general, DSHA/NDSHA provides more conservative ground motion values when compared with PSHA’s short to moderate ground motion return periods. However, PSHA can produce extremely large ground motion values for very long return periods, much greater than would be produced by MCE magnitude events. Such PSHA’s extreme ground motions are controversial and physically not supportable even though it can be easily calculated theoretically. Both DSHA and PSHA approaches have advantages and disadvantages as well as pros and cons, but a detailed discussion is outside the scope of this section.

For a coherent framework, DSHA and PSHA inputs will be simplified into two, namely, (1) earthquake magnitudes and (2) strong ground motion attenuation relationships, or GMPEs appropriate for the region. These two will be understood as *derived* inputs because they are obtained from analysis and interpretation of (a) seismologic and geologic *data* of instrumentally recorded, historic, and paleo-earthquakes; (b) strong ground motion recorded *data*; and (c) earthquake source or fault *data*. These data,

subjected to analysis and interpretation, are the *fundamental* inputs. Note that GMPEs are not required in NDSHA (e.g., Panza et al. 2013) which has an advantage over PSHA.

Moment magnitude scale (Hanks and Kanamori 1979) is used in modern practice to avoid saturation of large-magnitude earthquakes in the Richter scale. The moment magnitude scale takes advantage of the seismic moment derived for an equivalent double-couple earthquake source and which measures the strength (i.e., magnitude) of an earthquake without any saturation (Aki and Richards 2002). Hereafter, magnitude means moment magnitude.

Inputs for Seismic Hazard Analysis

Magnitude for DSHA

Only MCE magnitude for each of the considered seismogenic faults is derived as the required input. The magnitude is generally estimated from basic inputs of the length and area of fault by applying it to the empirical relationships between fault lengths or areas and magnitudes established from observations (Wyss 1979; Bonilla et al. 1984; Wells and Coppersmith 1994). The relationships are also established according to the style of dominant fault displacements, namely, strike-slip, normal, and reverse/thrust faults. Fault displacement can also be used for magnitude estimates but such data may not be readily available for all the faults.

Fault length is obtained by measuring it, usually from geologic maps based on field investigations. The width of the fault is estimated from the dip and depth of the fault, noting that fault width (W) is given by:

$$W = h / \cos(\delta),$$

where h is the depth of the fault and δ is the angle of the dip.

The dip can be obtained from geologic mapping or focal mechanism of earthquakes on the fault system. The focal depth distribution of earthquakes in the area can provide the dip and depth of the fault. Fault area is calculated from

fault length and width. Though actual fault surface would be quite complex, this representation is an approximation as a plane.

The length, width, and area of a causative fault can be approximated from the aftershock distribution zone of earthquake, preferably in conjunction with a fault map. This seismological method is possible only *after* the occurrence of an earthquake and provides data for developing empirical relationships between observed fault parameters and magnitudes.

By applying fault length or area to the appropriate empirical relationships, magnitude is obtained in a straightforward manner. Typical examples from Wells and Coppersmith (1994), without the standard deviations of the regression equations, are:

$$M = a + b * \log_{10}(\text{SRL}),$$

where SRL is the surface fault rupture length (km) and $a, b = 5.16, 1.12$ for strike-slip fault; $5.00, 1.22$ for reverse fault; $4.86, 1.32$ for normal fault; and $5.08, 1.16$ for all faults.

$$M = a + b * \log_{10}(\text{RA}),$$

where RA is the fault rupture area (sq-km) and $a, b = 3.98, 1.02$ for strike-slip fault; $4.33, 0.90$ for reverse fault; $3.93, 1.02$ for normal fault; and $4.07, 0.98$ for all faults.

For example, for a strike-slip fault, a change in fault length by twice or half will change the magnitude by about 0.3 magnitude unit, and an increase or decrease in fault length by 10 or 20 km by about 0.1 magnitude unit. For well-mapped faults, errors in lengths by 10–20 km are most unlikely and would be much less, and the uncertainty in MCE magnitude estimate would be negligible for all practical purposes. Therefore, uncertainty in MCE magnitude is not an issue in DSHA. An enhanced exploitation of geologic information is allowed by NDSHA (Panza et al. 2013).

Magnitudes for PSHA

PSHA used a range of magnitudes, from a minimum value, typically around magnitude

4, to the maximum or upper bound magnitudes for faults or areal sources. The upper bounds are comparable to MCE magnitudes. Considerable efforts have been made to estimate the upper bound magnitudes by using logic-tree analysis to quantify the uncertainty. However, such analysis has not demonstrated better estimates of MCE magnitudes. The weights or inputs at every step in the logic-tree analysis are not real data but experts' opinions with undefined uncertainty.

Again, magnitudes for seismogenic faults and areal zones are defined according to their *recurrence times* in the frequency magnitude relationships based on historic or paleo-seismic (magnitudes and occurrence times) regional data. A homogeneous earthquake catalog with “completeness” of earthquakes is the prerequisite for PSHA but not usually achieved. From such data, the most common frequency magnitude relationship or the “Gutenberg-Richter” equation for the source region has been derived, as proposed by Richter (1958):

$$\log_{10}N = a - bM,$$

where N is number of earthquakes of magnitude M or greater per year and a and b are constants usually estimated with regression analysis.

The recurrence time (in years) is the reciprocal of N, or $1/N$. Even though there are variations, the general form of this long-established equation has been the rule over the years. Availability of reliable earthquake data is required and applicability of the equation to specific faults is an open question that remains unresolved.

It should be noted that there is no formal time-series analysis operated on magnitudes and origin times to obtain recurrence times. Magnitudes in the observed time period are lumped together for calculating N arithmetically. There are no criteria for the spatial extent and data availability for developing the equation. Implications of these points are that recurrent magnitudes may not be as assumed and have undefined uncertainty.

Recurrence can also be obtained from slip rates on a seismogenic fault from paleo-seismic “displacement” data (Wallace 1970; McCalpin 1996).

Slip rate is derived from the total slip (i.e., displacement) along a fault over a long period of time even though the slip (centimeters to meters) during an earthquake takes only seconds. Mean recurrence interval (RI) is calculated from the equation:

$$RI = D/(S - C),$$

where D is displacement during a single earthquake, S is coseismic slip rate, and C is creep slip rate which is assumed zero for most faults. RI is equivalent to $1/N$, where N is given in the Gutenberg-Richter equation.

Note that displacements on a fault surface during an earthquake are truly multivalued and the meaning of D is not yet defined. Therefore, the value of slip rates has undefined uncertainty.

In principle, displacement can be correlated to magnitude by using empirical relationships between magnitude and displacement (Wells and Coppersmith 1994) given by:

$$M = a + b * \log_{10}(MD),$$

where MD is maximum displacement (m) and a , $b = 6.81, 0.78$ for strike-slip fault; $6.52, 0.44$ for reverse fault; $6.61, 0.71$ for normal fault; and $6.69, 0.74$ for all faults.

$$M = a + b * \log_{10}(AD),$$

where AD is average displacement (m) and a , $b = 7.04, 0.89$ for strike-slip fault; $6.64, 0.13$ for reverse fault; $6.78, 0.65$ for normal fault; and $6.93, 0.82$ for all faults.

In general, displacement or slip on many faults is not readily available, and therefore, slip rate may not be applicable in a large majority of seismogenic faults. Note that the significance of a specific displacement observed on a fault is undefined in characterizing the associated earthquake.

Ground Motion Prediction Equations (GMPEs) for DSHA and PSHA

The most commonly used ground motion parameter in SHA is peak ground acceleration (PGA) even though it is not well correlated with induced

structural damage by earthquakes. Nevertheless, it is used as a scaling factor for developing seismic design load spectra and simulating synthetic seismograms. Obviously, structural damage is best correlated with intensity such as Modified Mercalli which has been neglected in favor of using quantitative instrumental recordings. Reexamination of intensity scale is in order for complementing or supplementing current practices because of difficulty in interpreting instrumental records for structural damage and performance.

GMPEs show PGA as a function of “distance” from earthquake source for various magnitudes, starting with high PGAs near the source and gradually decreasing with increasing distances from the source (e.g., Schnabel and Seed 1973; Sadigh et al. 1997; Joyner and Boore 1981; Campbell 1981). Common distances used are epicentral distance, nearest distance to fault surface, and nearest distance to surface projection of fault. For proper comparisons of various GMPEs, the adopted distances should be related correctly.

The so-called Next Generation Attenuation (NGA) GMPEs have generated a new series of publications (Abrahamson and Silva 2008; Boore and Atkinson 2008; Campbell and Bozorgnia 2008; Chiou and Youngs 2008; Idriss 2008). The main idea of NGA was to have a uniform data for GMPE researchers so that they produce better or comparable results. It remains to be seen if the expectation is true. It should be noted that uniform data preparation is not necessarily a neutral activity and serious bias can occur in the process for various reasons, including intention to influence on the results for “special interest” applications. Moreover, independence of researchers is most valuable and working under such a condition may not produce best objective results.

In general, variations from the general trend of attenuation curves or GMPEs are, to a first approximation, due to variabilities in (1) subsurface physical configurations and conditions along the wave propagating path from source to site and (2) preferential focusing of seismic energy in the direction of fault rupture propagation.

The (aleatory) variability and (epistemic) uncertainty quantified as the standard deviation

“sigma” of $\log_{10}$ (PGA) in the GMPEs have remained stable within 0.15 and 0.35 over the last 40 years (Stasser et al. 2009). In other words, GMPEs have been developed as best as can be in spite of certain defects. GMPEs have been used in confidence over the years with prudent professional judgments.

Ground Motion Prediction for NDSHA

NDSHA produces realistic time series describing possible ground motions and does not require appealing to any variant of GMPEs. In the coming years, the reliability of NDSHA will be tested by real events to become more and more mature.

Basic Inputs for DSHA and PSHA

These are listed without discussion:

Historic seismicity from earthquake catalog, including completeness and homogeneity
Strong motion recorded raw and analyzed data and interpreted results
Paleo-seismicity from investigations and interpretation
Detailed fault information: location, geologic age, three-dimensional geometry including continuity or segmentation, style of movement, associated site and regional seismicity, assumed extent of rupture
Information about the structural model for the studied area (for NDSHA)

Implications of Inputs on Hazard Results

SHA results will be as good as the inputs. Accurate inputs will produce reliable results. Questionable inputs will produce doubtful results. Implications on the quality of results for DSHA/NDSHA and PSHA when the same inputs are used will be discussed. The nature of inputs and the implied hazard results will be examined for DSHA/NDSHA and PSHA.

Magnitudes

1. For DSHA/NDSHA, MCE magnitudes are based on fault input that are based on mapping and earthquake analysis. Fault lengths

are well established and its width and area are reliably derived. The nature of the empirical relationships is such that the MCE magnitude estimates are robust given the unlikely and conservative uncertainty in fault length. With such inputs, DSHA/NDSHA results are stable and expected to be reliable.

2. For PSHA, recurrent magnitudes will vary according to the quality and quantity of historic and paleo-seismic earthquake data. Regions with well-recorded history such as in Europe and China are expected to have better recurrent magnitudes. Well-studied faults such as the San Andreas Fault in California are expected to provide recurrent magnitudes for the largest events but not necessarily for the lower magnitudes. Faults with inadequate or no earthquake data will not have reliable recurrent magnitudes. Because in real earthquakes, slips are multivalued, and therefore, slip rates based on specific slip measurements have undefined uncertainty. The recurrence equation itself has no formal time-series analysis but just lumped together seismicity data within the observed time periods and therefore contains undefined uncertainty. Translating frequency magnitude relationships of a region to specific fault (s) has undefined uncertainty. A recent research reported that the concept of recurrence time itself can be misleading or even meaningless and the quantitative determination of such time remains only a dream (Bizzari and Crupi 2013).

Based on the above points and several involved undefined uncertainties, recurrent magnitudes are not reliable and so are the PSHA results.

Ground Motion Prediction

3. Both DSHA and PSHA used GMPEs. The stable nature of GMPEs over the last 40 years as noted above demonstrated the maturity and reliability of GMPE. Improvement may be expected from better understanding of earthquake source physics and unlikely from more statistical analysis such as

separating into aleatory variability and epistemic uncertainty. GMPEs as the derived inputs for DSHA and PSHA are stable and well established. GMPEs have been used in confidence over the years with prudent professional judgments.

4. The obvious limitations of GMPEs are due to neglecting (a) variation in subsurface conditions along the wave propagating path from source to site, (b) site-specific effects, and (c) preferential focusing of seismic energy in the direction of fault rupture propagation.
5. NDSHA computes realistic time series of possible ground motions without any variant of GMPEs. This is a distinct advantage to be tested by real events in the coming years and become mature for wide applications.

Using Same Inputs by DSHA/NDSHA and PSHA

6. For given faults and a given GMPE, the quality of results for DSHA/NDSHA depends on the accuracy of estimated MCE magnitude based on fault parameters and (a) GMPE for DSHA or (b) the computation for NDSHA. However, the quality of results for PSHA depends on additional parameters: earthquake recurrence based on seismicity or slip rate data; recurrence has undefined uncertainty.

Because of more inputs required by PSHA, which are associated with undefined uncertainty, the quality of the results will not be as good and transparent as that for DSHA/NDSHA. It will be difficult to evaluate the quality of results in PSHA as it does not produce transparent results like DSHA/NDSHA.

Comments on Inputs for DSHA/NDSHA and PSHA

7. DSHA/NDSHA requires fault geometry and geographic location of faults which can be readily obtained from tectonic maps and earthquake locations from seismicity maps. In a nutshell, input MCEs for sources are estimated as stated above and ground motion hazards can be obtained by using input GMPE in DSHA and by direct computation in NDSHA.

8. PSHA requires comprehensive homogeneous earthquake catalog and displacement data of faults. Earthquake catalogs for many areas are not always comprehensive, homogeneous, or complete. Recurrence equations have to be derived from such data but the applicability of such equations on individual faults remains problematic. Fault displacements provide paleo-earthquakes to extend earthquake history and can provide slip rate for deriving recurrence intervals. However, displacements are available only for a few faults. Moreover, the significance of specific displacements is undefined, being just one of the multivalued displacements during an earthquake centered on a fault. Furthermore, recurrence times are not based on time-series analysis and thus have undefined uncertainty. The concept of recurrence may be even meaningless and the quantitative determination remains only a dream.
9. PSHA inputs are more difficult and involved with several undefined uncertainties. PSHA's distinctiveness is to incorporate differences in seismic activity of faults. Unfortunately, this has not been achieved so far in a realistic manner.
10. In contrast, DSHA/NDSHA inputs are easier to obtain, and the results are stable and therefore expected to be reliable. Obviously, the implication is that DSHA/NDSHA seems more reliable and much more economic than PSHA.

Summary

The inputs for DSHA/NDSHA are simple, limited, and easy to obtain. The hazard results are stable and expected to be reliable. The absence of recurrence is an advantage because short or long recurrent earthquakes can occur at any time. NDSHA is advantageous without the need to use GMPEs. The procedures for DSHA should be more formalized as in PSHA. The simplicity and strength of DSHA should not be confused with the complexity and sophistication of PSHA.

The inputs for PSHA are not easily obtained and quite involved. Derived quantities such as

recurrence and slip rate are associated with undefined uncertainties. Therefore, the results are expected to be doubtful. A simpler PSHA procedure than the currently advanced may be more realistic.

The concept of recurrence as the foundation for PSHA can be misleading or even meaningless and the quantitative determination of such recurrence time remains unrealistic, and the value of such results questionable. In contrast, the concept of MCE and its magnitude determination are well established.

PSHA is much more expensive due to data requirements and complex analysis. There is a diminishing return in spending more money for PSHA. In contrast, DSHA is economic and the results are stable and expected to be reliable.

Cross-References

- [Earthquake Magnitude Estimation](#)
- [Earthquake Recurrence](#)
- [Earthquake Return Period and Its Incorporation into Seismic Actions](#)
- [Paleoseismology: Integration with Seismic Hazard](#)
- [Probabilistic Seismic Hazard Models](#)
- [Site Response for Seismic Hazard Assessment](#)

References

- Abrahamson NA, Silva WJ (2008) Summary of the Abrahamson and Silva NGA ground motion relations. *Earthq Spectra* 24:67–97
- Aki K, Richards P (2002) *Quantitative seismology: second edition*. University Science, Sausalito
- Bizzari A, Crupi P (2013) Linking the recurrence time of earthquake to source parameters: a dream of a real possibility? *Pure Appl Geophys*. doi:10.1007/s00024-013-0743-1
- Bonilla MG, Mark RK, Lienkaemper JJ (1984) Statistical relations among earthquake magnitude, surface rupture length, and surface fault displacement. *Bull Seismol Soc Am* 74:2379–2411
- Boore DM, Atkinson GM (2008) Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5 %-damped PSA at spectral periods between 0.01 s and 10.0 s. *Earthq Spectra* 24:99–138
- Campbell KW (1981) Near-source attenuation of peak horizontal acceleration. *Bull Seismol Soc Am* 71:2039–2070
- Campbell KW, Bozorgnia Y (2008) NGA ground motion model for the geometric mean horizontal component of PGA, PGV, PGD and 5 % damped linear elastic response spectra for periods ranging from 0.01 to 10 s. *Earthq Spectra* 24:139–171
- Chiou B, Youngs RR (2008) Chiou-Youngs NGA ground motion relations for the geometric mean horizontal component of peak and spectral ground motion parameters. *Earthq Spectra* 24:173–215
- Cornell CA (1968) Engineering seismic risk analysis. *Bull Seismol Soc Am* 58:1583–1606
- Hanks TC, Kanamori H (1979) A moment magnitude scale. *J Geophys Res* 84:2348–2350
- Idriss IM (2008) An NGA empirical model for estimating the horizontal spectral values generated by shallow crustal earthquakes. *Earthq Spectra* 24:217–242
- Joyner WB, Boore DM (1981) Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 imperial valley, California, earthquake. *Bull Seismol Soc Am* 71:2011–2038
- Klügel J-U (2005) Problems in the application of the SSHAC probability method for assessing earthquake hazards at Swiss nuclear power plants. *Eng Geol* 78:285–307
- Klügel J-U (2007) Error inflation in probabilistic seismic hazard analysis. *Eng Geol* 90:186–192
- Kossobokov V, Nekrasova A (2012) Global seismic hazard assessment program maps are erroneous. *Seism Instrum* 48:162–170
- McCalpin JP (ed) (1996) *Paleoseismology*. Academic, San Diego
- Mualchin L, Jones AL (1992) Peak acceleration from maximum credible earthquakes in California. California Division of Mines & Geology Open File Report 92–01 and Maps
- Nekrasova A, Kossobokov V, Peresan A, Magrin A (2013) The comparison of the NDSHA, PSHA seismic hazard maps and real seismicity for the Italian territory. *Nat Hazards* 68. doi:10.1007/s11069-013-0832-6
- Panza GF, La Mura C, Peresan A, Romanelli F, Vaccari F (2012) Seismic hazard scenarios as preventive tools for a disaster resilient society. *Adv Geophys* 53:93–165
- Panza GF, Peresan A, La Mura C (2013) Seismic hazard and strong ground motion: an operational neo-deterministic approach from national to local scale. *Geophysics and geochemistry*. In: UNESCO-EOLSS Joint Committee (ed) *Encyclopedia of life support systems (EOLSS)*. Developed under the auspices of the UNESCO. Eolss Publishers, Oxford
- Reiter L (1990) *Earthquake hazard analysis: issues and insights*. Columbia University Press, New York
- Richter CF (1958) *Elementary seismology*. W. H. Freeman, San Francisco
- Sadigh K, Chang C-Y, Egan JA, Makdisi F, Youngs RR (1997) Attenuation relationships for shallow crustal earthquakes based on California strong motion data. *Seismol Res Lett* 68:180–189

- Schnabel PB, Seed HB (1973) Acceleration in rock for earthquakes in the western United States. *Bull Seismol Soc Am* 63:501–516
- Stasser FO, Abrahamson NA, Bommer JJ (2009) Sigma: issues, insights, and challenges. *Seismol Res Lett* 80:40–56
- Wallace RE (1970) Earthquake recurrence intervals on the San Andreas fault. *Geol Soc Am Bull* 81:2875–2890
- Wang Z (2011) Seismic hazard assessment: issues and alternatives. *Pure Appl Geophys* 168:11–25
- Wang Z, Cobb JC (2012) A critique of probabilistic versus deterministic seismic hazard analysis with special reference to the New Madrid Seismic Zone. In: Cox RT, Tuttle MP, Boyd OS, Locat J (eds) *Recent advances in North American paleoseismology and neotectonics east of the Rockies*. Geological Society of America special paper, vol 493, pp 259–275
- Wells DL, Coppersmith K (1994) New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bull Seismol Soc Am* 84:974–1002
- Wyss M (1979) Estimating maximum expectable magnitude of earthquakes from fault dimensions. *Geology* 7:336–340
- Wyss M, Nekrasova A, Kossobokov V (2012) Errors in expected human losses due to incorrect seismic hazard estimates. *Nat Hazards* 62:927–935

Robust Control of Building Structures Under Uncertain Conditions

Izuru Takewaki¹ and Kohei Fujita²

¹Department of Architecture and Architectural Engineering, Graduate School of Engineering, Kyoto University, Kyoto, Japan

²RIKEN Advanced Institute for Computational Science, Kyoto, Japan

Synonyms

Base-isolated building; Building structure; Earthquake response; Interval analysis; Passive damper; Robust control; Structural control; Uncertain parameter

Introduction

Design can be regarded as an action with decision making for finding an acceptable outcome

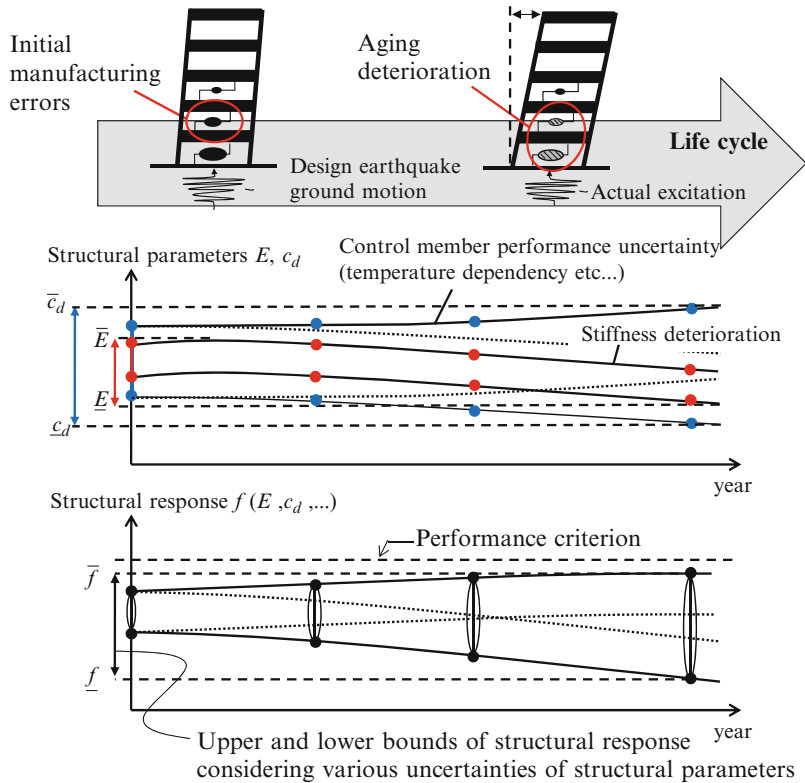
satisfying (usually optimally) a set of specified requirements called design constraints. In general, the design model (design object), the objective function, and the design constraints include various uncertainties. For example, mechanical properties of structural members have variability due to initial manufacturing errors, material deterioration, and temperature dependence (see Fig. 1). Young's moduli and strengths of concrete exhibiting large variability and strengths of steel indicate not a small uncertainty. In addition, it is well known that passive dampers used in structural control and base-isolation systems have a large degree of variability compared to ordinary structural members. This characteristic is reflected in the structural design of passive-controlled structures and base-isolated structures in Japan. From another viewpoint, it is often pointed out that there remains a large uncertainty in the modeling and analysis of structural systems, e.g., shear strengths of reinforced concrete members, reinforced concrete shear walls, steel plate shear walls, steel braces, and super high-rise buildings (Minami et al. 2013). Some of these phenomena possess inherent uncertainties which may not be possible to overcome by the enhancement and upgrade of modeling.

Regarding the disturbances for design objects, design input ground motions have a larger variability compared to the structural member variability. This may result from the insufficient research in the modeling of ground motions together with physical unclearness of ground media (Takewaki et al. 1991, 2011, 2012, 2013; Conte et al. 1992; Takewaki 2006, 2008, 2013; Ariga et al. 2006; Celebi et al. 2012).

Based on these observations, it is recognized well that the reliable design of building structures with greater robustness and reliability can be achieved by taking into account the input disturbance uncertainties and structural parameter uncertainties. In order to get a reliable design with greater robustness, the introduction and definition of robustness are essential. An example measure of the robustness is explained here and its application to the robust control of building structures with uncertain input disturbance and structural parameters is presented. Since

Robust Control of Building Structures Under Uncertain Conditions,

Fig. 1 Variation of structural response under uncertain structural parameters in a passively controlled building



a method for quantifying the response variability due to input disturbance and structural parameter uncertainties is needed in the robust control of such building structures, an enhanced and efficient methodology is explained for evaluating the robustness of an uncertain building structure with passive dampers and an uncertain base-isolated building structure.

A number of studies on uncertainty analysis have been accumulated so far. Most studies are aimed at investigating the upper bound of the structural responses considering the uncertainties of structural parameters (e.g., see Ben-Haim and Elishakoff 1990; Ben-Haim 2001, 2006; Takewaki and Ben-Haim 2005; Takewaki 2006, 2013; Takewaki et al. 2012; Kanno and Takewaki 2006; Elishakoff and Ohsaki 2010). It is well understood and accepted that the interval analysis is a representative of the reliable uncertainty analysis methods. It seems that the concept of interval analysis was introduced by Moore (1966). Alefeld and Herzberger (1983) have then conducted the pioneering work. They treated

linear interval equations, nonlinear interval equations, and interval eigenvalue analysis by using interval arithmetic. Since their innovative achievements, various interesting interval analysis techniques based on the interval arithmetic algorithm have been proposed by many researchers (e.g., Dong and Shah 1987; Koyluoglu and Elishakoff 1998; Mullen and Muhanna 1999; Qiu 2003).

More recently, some sophisticated interval analyses using Taylor series expansion have been proposed by Chen et al. (2003, 2009), Chen and Wu (2004), and Fujita and Takewaki (2011a, b, 2012a, b). In the early stage of the interval analysis using Taylor series expansion, the first-order Taylor series expansion was introduced and investigated for the basic problems of static response and eigenvalue. Chen et al. (2009) developed a matrix perturbation method using the second-order Taylor series expansion and obtained an approximation of the bounds of the objective function without interval arithmetic. They pointed out that the computational load can be reduced

from the number of calculation 2^N (N : number of interval parameters) to $2N$ by neglecting the non-diagonal elements of the Hessian matrix of the objective function with respect to interval parameters. Furthermore, Fujita and Takewaki (2011a) have proposed the so-called updated reference-point (URP) method. In this method, the critical uncertain structural parameters can be obtained by the approximation of second-order Taylor series expansion, and the upper bound of the structural responses can be evaluated by reanalyzing the structural response using critical uncertain structural parameters. Their method has been extended to nonlinear inelastic problems (Fujita and Takewaki 2012a).

In the structural design procedure, the robustness of building structures should be taken into account and incorporated under various uncertainties of structural parameters and inputs (e.g., see Takewaki et al. 2012). Ben-Haim (2001) has proposed an index, called the robustness function, for measuring the degree of the robustness based on the info-gap decision theory. The concept of the info-gap models and the robustness functions will be explained in the following section. In the info-gap model, the uncertainty of structural parameters is assumed to be given by a non-probabilistic model, e.g., an interval model used in the interval analysis. According to the definition of the robustness function, it can be regarded as a quantitative index of the robustness of the building structure (Takewaki and Ben-Haim 2008).

Here an efficient evaluation method using the robustness function with respect to the constraint on seismic performance is presented and explained by taking advantage of the sophisticated uncertainty analysis method called the URP method. A planar shear building model with passive viscous dampers is used for the robustness analysis (see Fig. 1). By comparing the robustness functions for various damper distributions, a preferable damper distribution is investigated to enhance the robustness of the building model under various uncertainties of structural parameters.

An application to base-isolated buildings with hysteretic responses is also explained.

Robustness Function for Seismic Performance

In the structural design of buildings in earthquake-prone countries, the constraints on dynamic responses for earthquake loadings are major concerns and should be taken into account in an appropriate manner. In these design constraints, the dynamic responses such as maximum horizontal displacement and member stress evaluated by a reliable time-history response analysis are required especially in important structures to check the satisfaction of the performance criteria. The necessity of using the time-history response analysis depends on the seismic-resistant design codes of the countries. Even if all the design constraints are satisfied at the initial construction stage, some responses to external loadings during service life may violate such constraints due to various factors resulting from randomness, material deterioration, temperature dependence, etc. To overcome and remedy such difficulty, the introduction of the robustness function which represents the degree of robustness of the objective building structure may be one of the effective solutions for the design of more robust building structures under uncertainties (Takewaki et al. 2012).

In this section, a definition of the robustness of building structures for the seismic performance is introduced based on the info-gap model (Ben-Haim 2001). According to the info-gap model, the uncertainty of structural parameters is defined as a non-probabilistic model.

As for convex models, several useful methods can be used (Ben-Haim and Elishakoff 1990; Ben-Haim et al. 1996; Pantelides and Tzan 1996; Tzan and Pantelides 1996; Baratta et al. 1998). A convex model is defined mathematically as a set of functions. Each function is a realization of an uncertain event. Convex models for ground motion modeling depend on the level of prior information available. Examples are a local energy-bound convex model, an integral energy-bound convex model, an envelope-bound convex model, a Fourier-envelope convex model, and a response-spectrum-envelope convex model (Ben-Haim et al. 1996). One of the merits of the convex

models is the capability of prediction of the maximum or extreme response of structures to unknown inputs. The details of convex models can be found in appropriate references (Ben-Haim and Elishakoff 1990; Ben-Haim et al. 1996).

In the following, uncertain structural parameters are assumed to be described by an interval model. The interval parameter $\mathbf{X}^I$ is defined by

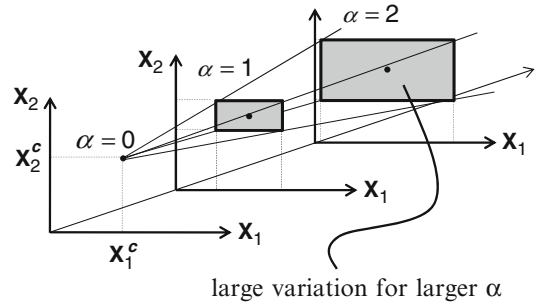
$$\mathbf{X}^I = \{X_i^I | [X_i^c - \Delta X_i, X_i^c + \Delta X_i], i = 1, \dots, N\} \quad (1)$$

In Eq. 1, $()^I$ and $[a, b]$ represent the definition of an interval parameter where a and b are the lower and upper bounds of the interval parameter, respectively. Furthermore, $()^c$, $\Delta()$ and N denote the nominal value of an interval parameter, half the varied range of the interval parameter, and the number of interval parameters, respectively. When the uncertainty of structural parameters is described by the interval vector, it means that the feasible domain of interval parameters is constrained into an N -dimensional rectangle.

In the info-gap model (Ben-Haim 2001), the level (or degree) of uncertainty is defined by a single uncertain parameter α . Based on the definition of an uncertain parameter α in the info-gap model, the feasible domain of the interval parameter $\mathbf{X}^I$ can be represented by an uncertainty set $\mathbf{X}^I(\alpha) \in \mathbf{R}$ which is described by

$$\mathbf{X}^I(\alpha) = \{X_i^I | [X_i^c - \alpha \Delta X_i, X_i^c + \alpha \Delta X_i], i = 1, \dots, N\} \quad (2)$$

In Eq. 2, ΔX_i is a prescribed value of half the varied range of the interval parameters. It should be noted that the level (or degree) of uncertainty can be described by the combination of ΔX_i and α . Therefore, it can be mentioned that the uncertainty level of the uncertainty set $\mathbf{X}^I(\alpha)$ varies according to the variation of uncertain parameter α . Figure 2 shows the variation of a two-dimensional interval model with an uncertain parameter α . When $\alpha = 0$, the uncertainty set $\mathbf{X}^I(0)$ corresponds to a nominal vector of structural parameters.



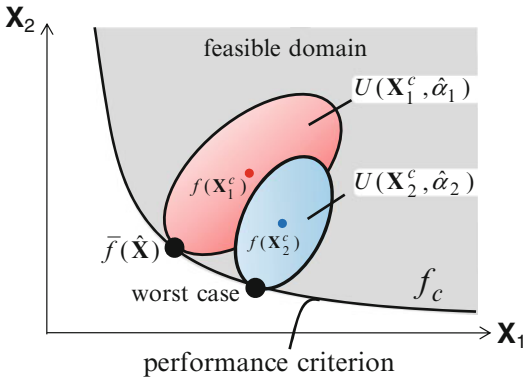
Robust Control of Building Structures Under Uncertain Conditions, Fig. 2 Variation of uncertainty set of interval model

The robustness function $\hat{\alpha}$ for the design constraint of the seismic performance can be defined as

$$\hat{\alpha}(\mathbf{X}^c, f_c) = \max\{\alpha | f \leq f_c, f \in U(\mathbf{X}^c, \alpha)\} \quad (3)$$

where f , f_c , and $U(\mathbf{X}^c, \alpha)$ denote the objective function, the performance criterion value, and the set of the possible structural responses in the domain of the uncertainty set $\mathbf{X}^I(\alpha)$. In Eq. 3, the robustness function $\hat{\alpha}$ is the maximum value of the uncertain parameter α which satisfies the performance criterion $f \leq f_c$. If the nominal value $f(\mathbf{X}^c)$ of the objective function corresponding to the nominal design $\mathbf{X}^c$ violates f_c or just coincides with f_c without considering a safety factor, the robustness function $\hat{\alpha}$ is regarded as zero, which means that no variability due to the uncertainty of structural parameters can be allowed. If the relation $\hat{\alpha}_1(\mathbf{X}_1^c, f_c) > \hat{\alpha}_2(\mathbf{X}_2^c, f_c)$ holds, a design more robust than $\mathbf{X}_2^c$ can be achieved by $\mathbf{X}_1^c$.

Figure 3 illustrates the relationship between the robustness function and the feasible domain of structural design to satisfy the performance criterion $f \leq f_c$ for two-dimensional interval parameters. The robustness function $\hat{\alpha}$ is derived as the worst case of the objective function, i.e., the upper bound of the objective function $\bar{f}$ in $U(\mathbf{X}^c, \hat{\alpha})$. However, when the number of the combinations of uncertain parameters is extremely large, it may be hard to evaluate the worst case of the objective function reliably. For this reason, an efficient uncertainty analysis method is desired which can evaluate the upper bound of



Robust Control of Building Structures Under Uncertain Conditions, Fig. 3 Robustness function for performance criterion

the objective function considering the uncertainty of the structural parameters accurately and reliably.

As a simple example, let us consider a shear building model as a vibration model, as shown in Fig. 4, with viscous dampers in addition to masses and story stiffnesses. It is well recognized in the field of structural control and health monitoring that viscous damping coefficients c_i of dampers in a vibration model are quite uncertain compared to floor masses and structural stiffnesses. This is because most of such dampers possess high temperature and frequency dependencies and it is often difficult to restrain those properties into an acceptable range. By using a specific method for describing such uncertainty, the uncertain viscous damping coefficient of a damper can be expressed in terms of the nominal value $\tilde{c}_i$ and the unknown uncertainty level (band) α as shown in Fig. 5a (Takewaki and Ben-Haim 2005).

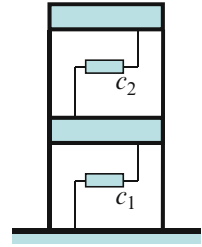
$$C(\alpha, \tilde{\mathbf{c}}) = \left\{ \mathbf{c} : \left| \frac{c_i - \tilde{c}_i}{\tilde{c}_i} \right| \leq \alpha, \quad i = 1, \dots, N \right\}, \alpha \geq 0 \quad (4a)$$

The inequality in Eq. 4a can be rewritten as

$$(1 - \alpha)\tilde{c}_i \leq c_i \leq (1 + \alpha)\tilde{c}_i \quad (4b)$$

This description is the same one used in the interval analysis (Moore 1966; Mullen et al. 1999; Koyluoglu and Elishakoff 1998).

Robust Control of Building Structures Under Uncertain Conditions, Fig. 4 Shear building model with uncertain viscous dampers



Another definition of uncertainty may be expressed by

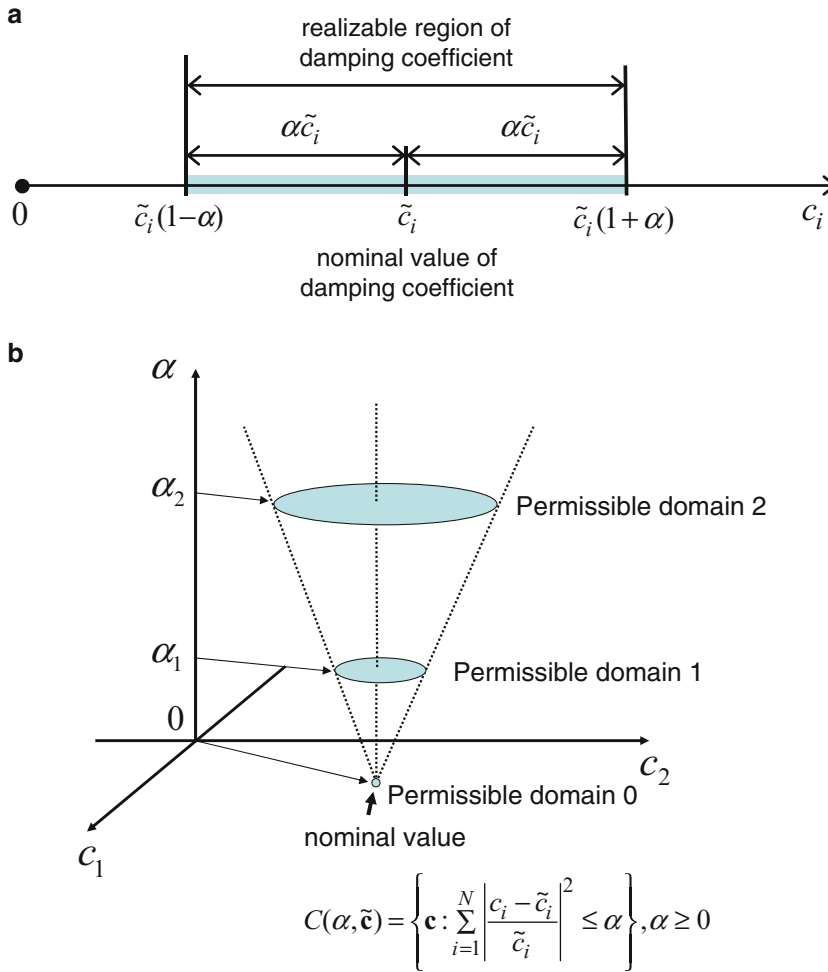
$$C(\alpha, \tilde{\mathbf{c}}) = \left\{ \mathbf{c} : \sum_{i=1}^N \left| \frac{c_i - \tilde{c}_i}{\tilde{c}_i} \right|^2 \leq \alpha \right\}, \alpha \geq 0 \quad (4c)$$

The implication of Eq. 4c can be found in Fig. 5b. It seems that Eq. 4c is too mathematical oriented and does not express the vagueness of a real problem. However, unless a matter is expressed in a mathematical form, an uncertainty problem cannot be formulated. Furthermore, a simple model shown in Fig. 4 has been used to enable the drawing of a three-dimensional figure in Fig. 5b.

Figure 6 shows illustrative representation of the concept of info-gap robustness function. When the uncertainty level is α_2 , the response domain is just encircled by the permissible (or feasible) domain. In this case, the info-gap robustness function is given by α_2 .

Efficient Uncertainty Analysis Based on Interval Analysis

In this section, the URP (updated reference-point) method proposed originally for stochastic input is explained (Fujita and Takewaki 2011a). This method can be used as an efficient uncertainty analysis to obtain the robustness function $\hat{\alpha}$ explained in the previous section. Since the URP method takes full advantage of an approximation of first- and second-order Taylor series expansion in the interval analysis, the formulation of Taylor series expansion in the interval analysis and the achievement of second-order Taylor series expansion proposed by Chen et al. (2009) are explained briefly.



Robust Control of Building Structures Under Uncertain Conditions, Fig. 5 Description of uncertainty with info-gap model

R

Interval Analysis Using Taylor Series Expansion

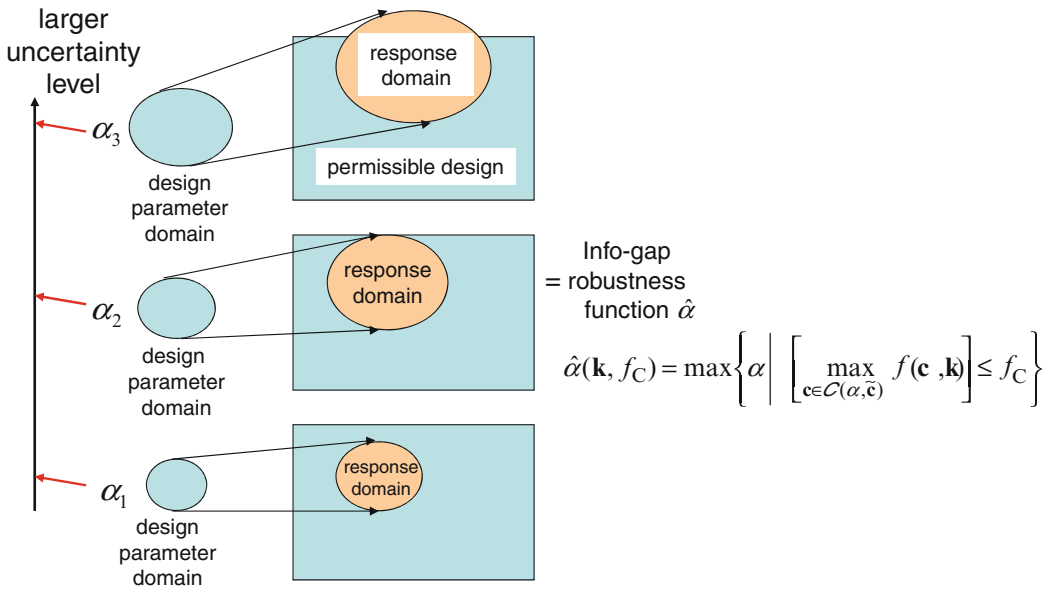
An approximate objective function f^* using Taylor series expansion up to second order around the nominal model $\mathbf{X}^c$ can be expressed as

$$f^*(\mathbf{X}) = f(\mathbf{X}^c) + \sum_{i=1}^N f_{,X_i}(X_i - X_i^c) + \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N f_{,X_i X_j}(X_i - X_i^c)(X_j - X_j^c) \quad (5)$$

In Eq. 5, the notations $(\cdot)_{X_i}$ and $(\cdot)_{X_i X_j}$ denote first-order differentiation $\partial f(\mathbf{X}) / \partial X_i|_{X_i=X_i^c}$ and second-

order differentiation $\partial^2 f(\mathbf{X}) / \partial X_i \partial X_j|_{X_i=X_i^c, X_j=X_j^c}$ of the objective function at the nominal value. Therefore, $f_{,X_i}$ and $f_{,X_i X_j}$ correspond to the gradient of f and to the Hessian of f for the nominal model, respectively.

In order to evaluate the upper bound of f^* approximately, a basic theorem of “inclusion monotonic” in the interval analysis is often assumed. The theorem of “inclusion monotonic” is assumed in some of the previous studies on interval analysis, e.g., Chen et al. (2009). If the natural interval extension f^I of f is inclusion monotonic, the objective function f satisfies



Robust Control of Building Structures Under Uncertain Conditions, Fig. 6 Illustrative representation of concept of info-gap robustness function

$$\begin{aligned} & \{f(\mathbf{X}) : x_i \in X_i^I, i = 1, 2, \dots, N\} \\ & \subseteq f(X_1^I, X_2^I, \dots, X_N^I) \end{aligned} \quad (6)$$

The right-hand side of Eq. 6 denotes the interval (range) of the function f determined by the endpoint combinations. In other words, as long as the arguments of f are constrained between intervals (the lower and upper bounds a and b in $[a, b]$ are also called “intervals”), the variation of f for any interval parameter value within the intervals should be included in the range of values for the intervals. Based on the theorem of “inclusion monotonic,” the upper bound of f can be derived by iterative calculations with all endpoint combinations, i.e., the upper and lower bounds of interval parameters. However, when the number N of interval parameters is extremely large, this primitive approach needs much computational time and load caused by a large number of combination of interval parameters. Although the computational effort can be reduced effectively by the approximation of Taylor series expansion in Eq. 5, the number of iterative calculations with all endpoint combinations is the same with the

interval analysis method based on the theorem of “inclusion monotonic,” e.g., Dong and Shah (1987).

The approximation of Taylor series expansion enables the avoidance of iterative response analyses such as time-history analysis for evaluating the objective function. However, the computation of full elements of the Hessian matrix requires a huge computational load when N is large, especially for numerical sensitivity analysis, i.e., the finite difference analysis using gradient vectors. A simpler approach has therefore been proposed by Chen et al. (2009) where the non-diagonal elements of the Hessian matrix are neglected.

As a simple approximation, an objective function f^{**} using second-order Taylor series expansion with only diagonal elements can be rewritten as

$$\begin{aligned} f^{**}(\mathbf{X}) &= f(\mathbf{X}^c) \\ &+ \sum_{i=1}^N \left\{ f_{,x_i}(X_i - X_i^c) + \frac{1}{2} f_{,x_i x_i}(X_i - X_i^c)^2 \right\} \end{aligned} \quad (7)$$

From Eq. 7, the increment of the objective function can be evaluated by using first- and second-order Taylor series expansion as the sum of the increments of the objective function in each one-dimensional domain. The perturbation $\Delta f_i(\mathbf{X})$ of the objective function by the variation of the i -th interval parameter X_i can be described as

$$\Delta f_i(X_1^c, \dots, X_i, \dots, X_N^c) = f_{,X_i}(X_i - X_i^c) + \frac{1}{2} f_{,X_i X_i}(X_i - X_i^c)^2 \quad (8)$$

The validity of ignoring cross terms can be found in the reference (Fujita and Takewaki 2011). In Eq. 8, the interval extension of the one-dimensional perturbation can be derived as

$$\begin{aligned} \Delta f_i^I(X_1^c, \dots, X_i, \dots, X_N^c) \\ = \left[\min[\Delta f_i(X_1^c, \dots, \underline{X}_i, \dots, X_N^c), \Delta f_i(X_1^c, \dots, \bar{X}_i, \dots, X_N^c)], \right. \\ \left. \max[\Delta f_i(X_1^c, \dots, \underline{X}_i, \dots, X_N^c), \Delta f_i(X_1^c, \dots, \bar{X}_i, \dots, X_N^c)] \right] \end{aligned} \quad (9)$$

Equation 9 implies that the upper bound $\overline{\Delta f_i}$ of Eq. 7 can be derived by the comparison of Δf_i with structural parameter sets $\mathbf{X} = \{X_1^c, \dots, \underline{X}_i, \dots, X_N^c\}$ and $\mathbf{X} = \{X_1^c, \dots, \bar{X}_i, \dots, X_N^c\}$. Since Eq. 8 is a function of $\mathbf{X}$, it is natural to define an upper bound of that function. Finally, substituting $\overline{\Delta f_i}$ ($i = 1, \dots, N$) into Eq. 7, the interval extension of the approximate objective function f^{**} can be obtained as

$$\begin{aligned} f(\mathbf{X}^I) \approx \left[f(\mathbf{X}^c) + \left[\sum_{i=1}^N \underline{\Delta f_i}(X_i^I) \right], \right. \\ \left. f(\mathbf{X}^c) + \left[\sum_{i=1}^N \overline{\Delta f_i}(X_i^I) \right] \right] \end{aligned} \quad (10)$$

It is interesting to note that the number of calculations in Eq. 10 is reduced to $2N$ from 2^N in Eq. 7. However, it should also be mentioned that, because of the approximation by Taylor series expansion, the deterioration of accuracy cannot be avoided when the level of uncertainties of interval parameters is large.

Search Algorithm for Critical Combination of Interval Parameters

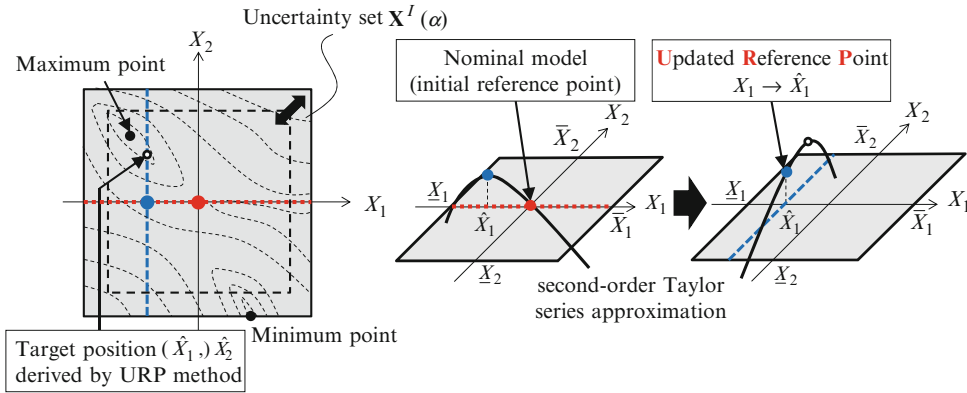
The approximation based on Taylor series expansion can reduce computational load dramatically in the interval analysis. However, it should be noted that the result by such approximation may (probably certainly) include errors. Furthermore, although some of the interval analysis methods

are based on “inclusion monotonic,” it is not necessarily appropriate to assume the monotonic variation of the objective function for dynamic responses. This is because a resonance (specific phenomenon in dynamic problems) could produce a non-monotonic response property. When the objective function has a non-monotonic property in $U(\mathbf{X}^c, \alpha)$, the extreme value of the objective function may occur not on the bound of interval parameters but in an inner feasible domain of interval parameters. Even in such a case, the robustness evaluation can be achieved by reanalyzing the structural response via a reliable response analysis for the estimated critical combination of interval parameters (worst case). In this section, an efficient search algorithm for the critical combination of interval parameters is explained which makes the approximate objective function maximum by using first- and second-order Taylor series expansion.

Consider Eq. 8 again. When the perturbation (or variation) $X_i - X_i^c$ of the structural parameter in Eq. 8 is denoted by ΔX_i , Eq. 8 can be transformed into

$$\Delta f_i(\Delta X_i) = \frac{1}{2} f_{,X_i X_i} \left(\Delta X_i + \frac{f_{,X_i}}{f_{,X_i X_i}} \right)^2 - \frac{f_{,X_i}^2}{2 f_{,X_i X_i}} \quad (11)$$

This transformation is just a simple mathematical transformation. From Eq. 11, it can be seen



Robust Control of Building Structures Under Uncertain Conditions, Fig. 7 Conceptual diagram of the URP method

that the increment of the objective function with respect to X_i is parabolic in the one-dimensional domain. By using Eq. 11, the target position can be found of the uncertain structural parameter X_1 which maximizes the objective function based on

the second-order Taylor series expansion. For instance, when $f_{,X_iX_i} < 0$ is satisfied, the target position $\hat{X}_i$ of the i -th interval parameter X_i which maximizes Eq. 11 can be derived explicitly as

$$\hat{X}_i = \begin{cases} X_i^c - f_{,X_i}/f_{,X_iX_i} & (|f_{,X_i}/f_{,X_iX_i}| \leq \Delta X_i) \\ X_i^c + \Delta X_i & (-f_{,X_i}/f_{,X_iX_i} \geq 0, |f_{,X_i}/f_{,X_iX_i}| > \Delta X_i) \\ X_i^c - \Delta X_i & (-f_{,X_i}/f_{,X_iX_i} < 0, |f_{,X_i}/f_{,X_iX_i}| > \Delta X_i) \end{cases} \quad (12)$$

The first case of Eq. 12 indicates that the critical value occurs in the inner domain. On the other hand, the second and third cases mean that the critical value occurs at the boundaries.

The target positions of the other interval parameters can be obtained successively in a similar way. The feature of the methodology explained here is that a possibility is considered of occurrence of the extreme value in an inner range of interval parameters and that only the first- and second-order sensitivities of the objective function are needed.

Equation 12 indicates that the target position of X_i can be derived by first- and second-order sensitivities $f_{,X_i}$, $f_{,X_iX_i}$ of the objective function with respect to X_i . For evaluating these sensitivities, it is needed to define a reference point. From the general point of view, it may be natural that the first- and second-order sensitivities with diagonal elements only are calculated at the reference

point of the nominal model. However, it is difficult to consider the influence of the interaction between the interval parameters in this approach. Fujita and Takewaki (2011a) developed the *updated reference-point method* (URP method) where the different computational procedure for the evaluation of first- and second-order sensitivities is applied. A detailed flow of the computation procedure of the URP method is explained below. The conceptual diagram of the URP method for two-dimensional interval parameters is shown in Fig. 7.

Step 1 Calculate the first-order sensitivities $f_{,X_i}$ ($i = 1, \dots, N$) of the objective function for the nominal model.

Step 2 Sort the absolute values $|f_{,X_i}|$ ($i = 1, \dots, N$) of the first-order sensitivities in descending order to give the priority to the parameter with the largest sensitivity. Sort also

the interval parameters as $\mathbf{X}_A = \{X_{A1}, \dots, X_{AN}\}$ corresponding to this.

Step 3 Calculate the second-order sensitivity of the objective function with respect to the interval parameter X_{Ak} . The second-order sensitivity $\partial^2 f / \partial X_{Ak}^2$ can be derived as a scalar value. When $k \geq 2$ holds, i.e., the reference point of the objective function has been updated, the first-order sensitivity of the objective function with respect to the interval parameter X_{Ak} should be calculated again.

Step 4 Derive the target position $\hat{X}_{Ak}$ of the interval parameter where the approximate objective function $f^{**}(\mathbf{X})$ is maximized. The problem in this step can be stated as

$$\begin{aligned} &\text{Find } \hat{X}_{Ak} \\ &\text{so as to maximize or minimize } f^{**}(\mathbf{X}) \\ &\text{subject to } X_{Ak} \in (X_{Ak})^I \\ &\text{where } X_{Al} \text{ is the current value } (l \neq k) \end{aligned} \quad (13)$$

Step 5 Update the set of interval parameters from current one X_{Ak} to $\hat{X}_{Ak}$.

Step 6 Update the corresponding system structural matrices such as $\mathbf{C}$ and $\mathbf{K}$ at the new reference point.

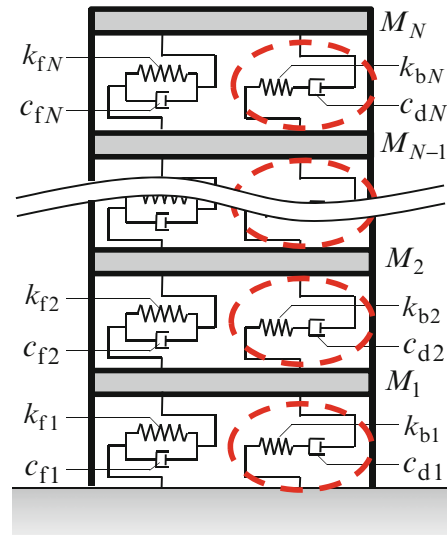
Step 7 Update k to $k + 1$. Repeat Step 2 through Step 6 until k becomes N .

Application to Building Structure with Passive Dampers

Any structural properties or responses, such as eigenvalue, static, and dynamic responses, can be employed as the objective function in the URP methods. From the viewpoint of seismic structural design, the applicability of the URP method to building structures with passive dampers is investigated as an example where the objective function is defined as maximum interstory drift for a set of recorded ground motions.

Structural Model with Passive Dampers and Selection of Uncertain Parameters

Consider an N -story planar shear building model, as shown in Fig. 8, with viscous dampers and their



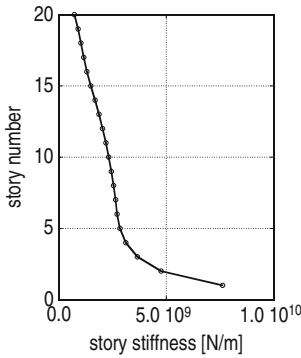
Robust Control of Building Structures Under Uncertain Conditions, Fig. 8 Structural model with passive damper

supporting members. Let M_i , k_{fi} , c_{fi} , c_{di} , and k_{bi} ($i = 1, \dots, N$) denote the floor mass, the story stiffness of the frame, the structural damping coefficient, the damping coefficient of the passive damper, and the supporting member stiffness of the damper in the i -th story, respectively. It has been shown that the supporting member stiffness of dampers plays an important role in the optimal distribution of dampers (Fujita et al. 2010a). The frame stiffness distribution in the nominal model is shown in Fig. 9 ($N = 20$). The properties of structural parameter of the nominal model are shown in Table 1.

The equations of motion of the building with viscous passive dampers subjected to the horizontal ground motion can be expressed in time domain by

$$\begin{aligned} \mathbf{M}\ddot{\mathbf{u}}(t) \\ + (\mathbf{C} + \mathbf{C}_D)\dot{\mathbf{u}}(t) + (\mathbf{K} + \mathbf{K}_b)\mathbf{u}(t) = -\mathbf{M}\ddot{\mathbf{r}}_g(t) \end{aligned} \quad (14)$$

where $\mathbf{M}$, $\mathbf{C}$, $\mathbf{C}_D$, $\mathbf{K}$, and $\mathbf{K}_b$ are the system mass, structural damping, damper damping, structural stiffness, and supporting member



Robust Control of Building Structures Under Uncertain Conditions, Fig. 9 Story stiffness distribution

Robust Control of Building Structures Under Uncertain Conditions, Table 1 Structural parameters of main frame

	20-story building
Floor mass [kg]	$1,024 \times 10^3$
Total damper capacity [Ns/mm]	$6,000 \times 10^8$
Supporting member stiffness [N/mm]	Ratio 1.0 to frame story stiffness
Structural damping ratio (stiffness-proportional damping)	0.02
Fundamental natural circular frequency with damper [rad/s] ^a	3.927

^aComplex eigenvalue analysis without dampers

stiffness matrices, respectively. Furthermore, $\mathbf{r} = \{1, \dots, 1\}^T$ is the influence coefficient vector. Assume that a viscous damper connected in series with its supporting member is treated as a detailed model in which a small lumped mass is allocated between the components of the dashpot and the spring. Then the components of the $\mathbf{K}$, $\mathbf{K}_b$, and $\mathbf{C}_D$ can be given by a linear combination of structural parameters k_{fi} , c_{di} , and k_{bi} ($i = 1, \dots, N$). The Newmark- β method ($\beta = 1/4$) is used to evaluate the maximum interstory drift.

The structural parameters $\mathbf{c}_d = \{c_{di}\}$, $\mathbf{k}_b = \{k_{bi}\}$, and $\mathbf{k}_f = \{k_{fi}\}$ are employed as interval parameters. The interval parameters of these uncertain

structural parameters are described with an uncertain parameter α by

$$\begin{aligned} \mathbf{c}'_d &= [\mathbf{c}_d^c - \alpha \Delta \mathbf{c}_d, \mathbf{c}_d^c + \alpha \Delta \mathbf{c}_d] \\ \mathbf{k}'_b &= [\mathbf{k}_b^c - \alpha \Delta \mathbf{k}_b, \mathbf{k}_b^c + \alpha \Delta \mathbf{k}_b] \\ \mathbf{k}'_f &= [\mathbf{k}_f^c - \alpha \Delta \mathbf{k}_f, \mathbf{k}_f^c + \alpha \Delta \mathbf{k}_f] \end{aligned} \quad (15a, b, c)$$

The uncertainty levels of interval parameters $\mathbf{X}' = \{\mathbf{c}'_d, \mathbf{k}'_b, \mathbf{k}'_f\}$ are given by $\boldsymbol{\varepsilon} = \{\varepsilon_i\}$ ($i = 1, \dots, 3N$) defined by

$$\varepsilon_i = \begin{cases} \Delta c_{di}/c_{di}^c = 0.3 & (i = 1, \dots, N) \\ \Delta k_{bi-N}/k_{bi-N}^c = 0.3 & (i = N+1, \dots, 2N) \\ \Delta k_{fi-2N}/k_{fi-2N}^c = 0.1 & (i = 2N+1, \dots, 3N) \end{cases} \quad (16)$$

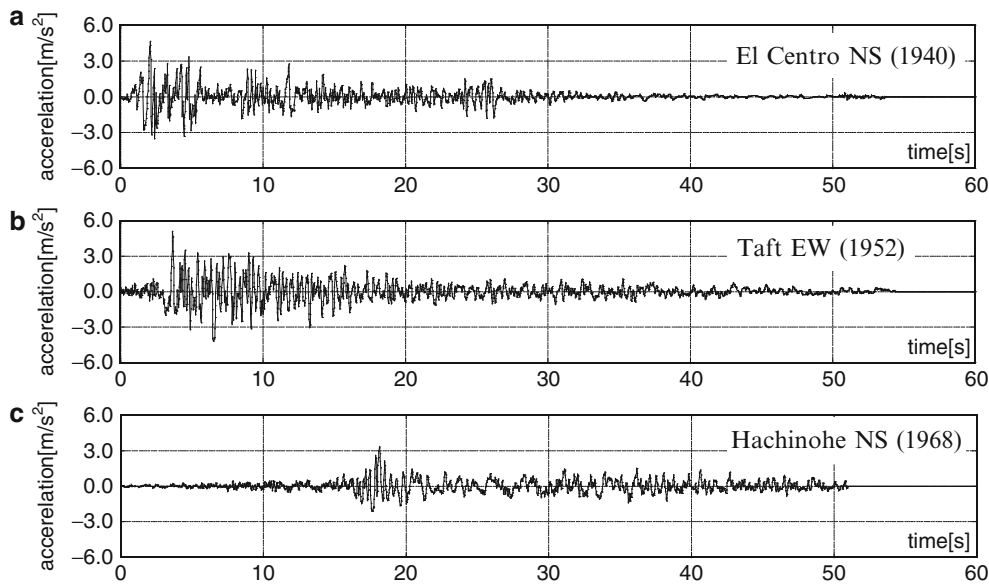
Equation 16 assumes that the degrees of uncertainties of interval parameters are constant for all stories on the same structural properties in the following numerical examples.

Recorded Ground Motions

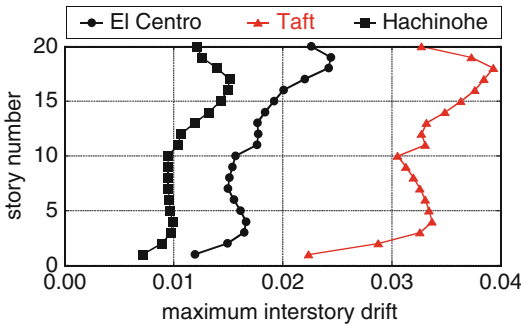
El Centro NS 1940, Taft EW 1952, and Hachinohe NS 1968 are used as representative recorded ground motions, whose maximum ground velocities are normalized as 0.5 [m/s]. These earthquake ground motions are often used in the structural design (level 2 of large earthquake ground motion) of high-rise and base-isolated buildings in Japan. Figure 10a–c shows the normalized recorded ground motions.

Robustness Function for Various Damper Distributions

In this section, the robustness functions defined by Eq. 3 for various damper distributions are evaluated by the URP method. Figure 11 shows the comparison of maximum interstory drifts of the building structure without dampers for three recorded ground motions. It may be observed from Fig. 11 that the interstory drift is maximized at upper stories above the 15th floor. For enhancing the seismic performance of the building structure, let us consider three different



Robust Control of Building Structures Under Uncertain Conditions, Fig. 10 Normalized recorded ground motions (a) El Centro NS (1940), (b) Taft EW (1952), (c) Hachinohe NS (1968)



Robust Control of Building Structures Under Uncertain Conditions, Fig. 11 Maximum interstory drift without dampers

damper distributions: (a) uniform distribution, (b) added only from 11th to 20th story, and (c) optimum distribution. In these various damper distributions, the total quantity of damping coefficients of the viscous dampers is given by a constant value as shown in Table 1. Figure 12 shows the comparison of the maximum interstory drifts with three different damper distributions shown in Fig. 13.

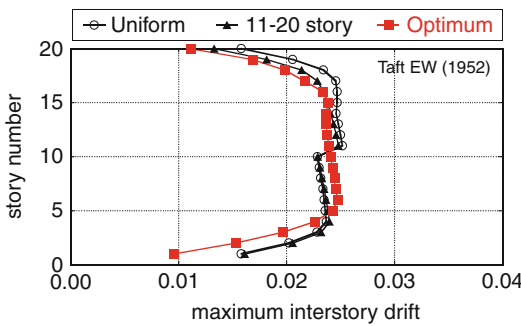
The optimum damper distribution has been derived by the optimization algorithm developed in the reference (Fujita et al. 2010b) to minimize the maximum amplitude of interstory-drift transfer function.

Figure 14a–c shows the robustness functions with respect to the constraint on the maximum interstory drift for the three representative recorded ground motions. These figures can be obtained by using the method in section “Efficient Uncertainty Analysis Based on Interval Analysis” and evaluating the maximum interstory drift for various values of $\hat{\alpha}$. The maximum interstory drift of the nominal model without uncertainty is shown in Fig. 14a–c at the uncertainty parameter $\hat{\alpha} = 0$. By comparing these nominal values of the objective function for various damper distributions, it is understood that the most preferable response reduction can be obtained by the damper distribution from 11th to 20th story in El Centro NS (1940) and Hachinohe NS (1968). On the other hand, the most preferable response reduction can be obtained by the optimal damper distribution in

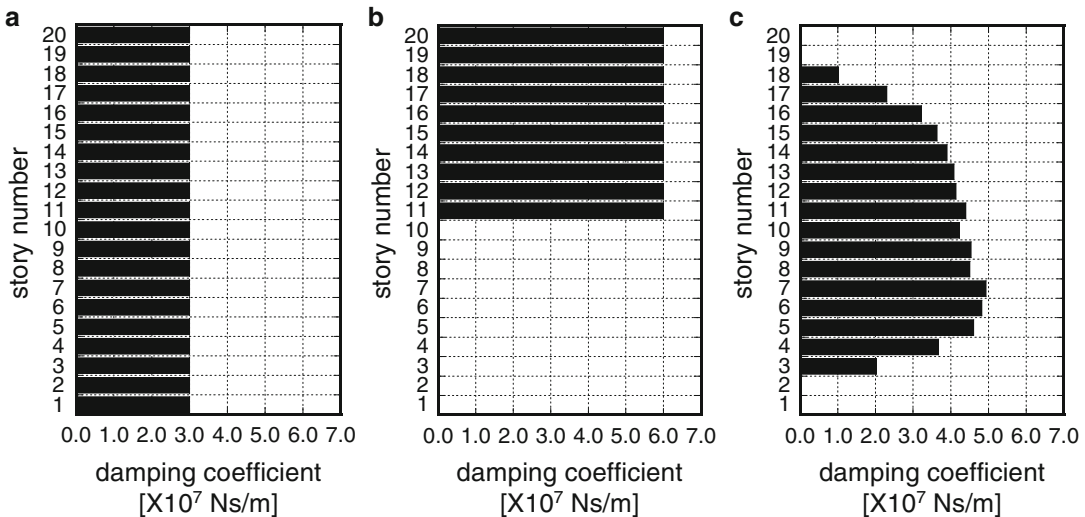
Taft EW (1952). Since the optimal damper distribution is aimed at suppressing the maximum amplitude of interstory-drift transfer function at the fundamental natural circular frequency, the drastic response reduction can be achieved for the excitation, such as Taft EW (1952), whose predominant frequency is resonant to the fundamental natural circular frequency of the objective building structure. It can be observed from Fig. 14b that if the performance criterion with respect to the maximum interstory drift is given by 0.03 [m], the robustness function $\hat{\alpha}$ is nearly

0.8 for the optimal damper distribution, while nearly 0.4 for the 11th–20th story distribution. From this comparison, it can be concluded that a large robustness can be obtained by the optimal damper distribution.

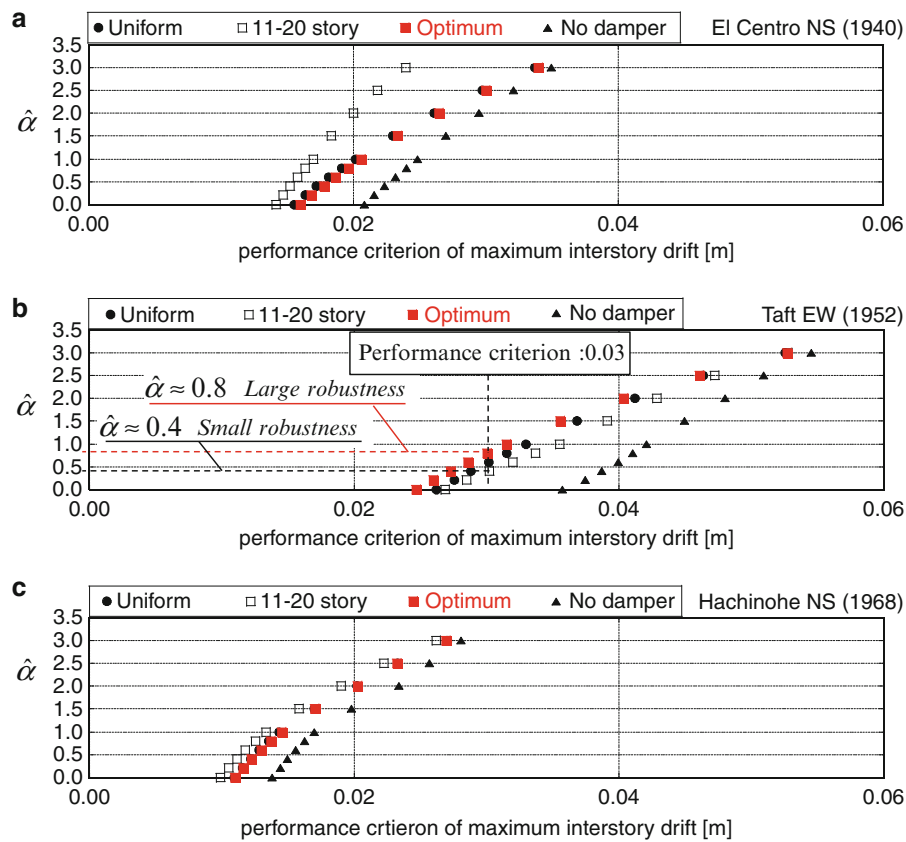
From the viewpoint of seismic structural design, the total quantity of added dampers may be a principal design parameter. In the general design procedure, the total quantity of dampers is determined based on the requirement whether the dynamic responses satisfy the constraints with a certain safety factor ρ . However, it is often ambiguous whether the values of these safety factors are appropriate or not. On the other hand, such total damper quantity can be derived in a more logical manner by using the “robustness function” for the structural uncertainty. Figure 15a shows the conceptual diagram of the redesign approach for determining the total damper quantity for a robust building structure, which can be derived by varying the robustness function with respect to the total damper quantity. The thick curve can be drawn by evaluating the dynamic response for various total quantities of dampers (nominal parameters). In Fig. 15a, when the performance criterion f_c and the robustness function $\hat{\alpha}$ are given (0, 0.5, 1.0), an appropriate



Robust Control of Building Structures Under Uncertain Conditions, Fig. 12 Comparison of maximum interstory drift with various damper distributions



Robust Control of Building Structures Under Uncertain Conditions, Fig. 13 Various damper distributions: (a) uniform, (b) 11th–20th story, (c) optimum



Robust Control of Building Structures Under Uncertain Conditions, Fig. 14 Comparison of robustness functions for various damper distributions. (a) El Centro NS (1940), (b) Taft EW (1952), (c) Hachinohe NS (1968)

total damper quantity can be found (Design 1, 2, 3). If the value of $\hat{\alpha}$ is assumed to be large, the total damper quantity will increase and a large robustness can be achieved. Figure 15b illustrates the relation of total quantities of passive dampers for the robustness function representation and the safety factor representation (Design 4). It can be observed that the robustness function of Design 4 with a safety factor ρ is between 0.5 and 1.0. A clear understanding may be possible of the meaning of the safety factor in terms of the structural robustness.

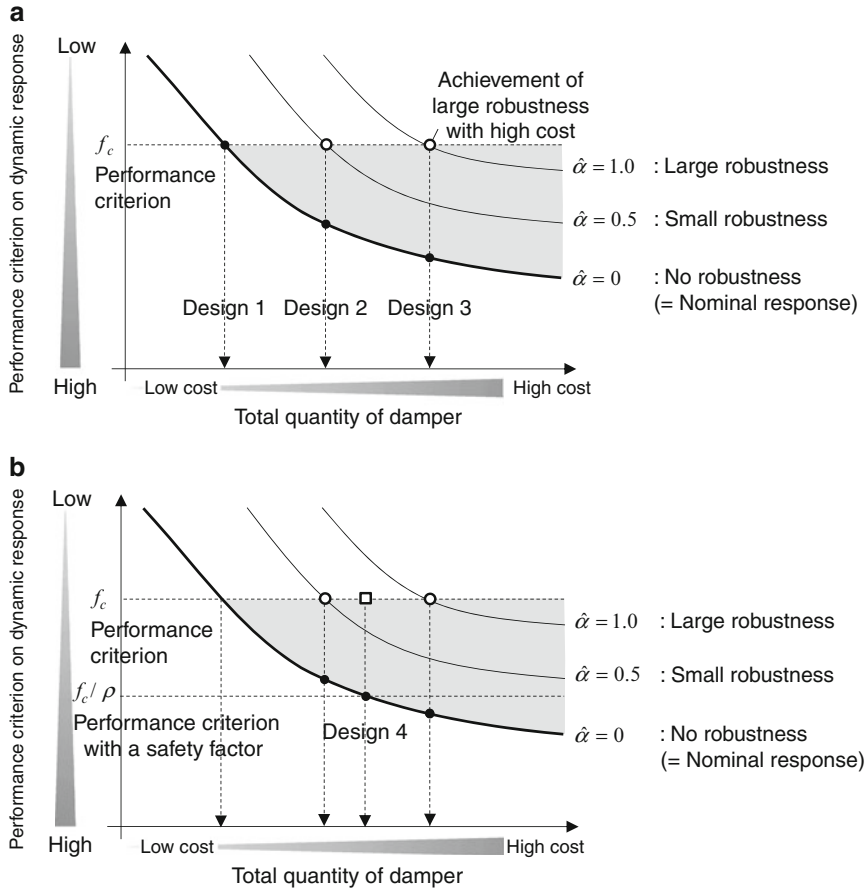
Figure 16 shows the degree of robustness of building structures with different damper distributions in terms of different robustness functions. If the maximum interstory drift is specified as 0.03 [m], the building structure with the optimum

damper distribution exhibits the maximum robustness. The building structure with a uniform damper distribution is the second and that with the 11–20th story distribution is the third.

Application to Base-Isolated Building with Hysteretic Response

Estimation of the Variation of the Objective Function by Second-Order Taylor Series Expansion

If the critical combination of interval parameters maximizing the objective function can be predicted, it becomes possible to evaluate the robustness accurately. Therefore, the Taylor

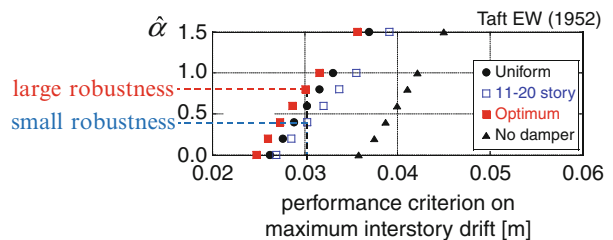


Robust Control of Building Structures Under Uncertain Conditions, Fig. 15 Redesign of total quantity of passive dampers for robust structures, (a) total quantity of dampers for no robustness, small robustness and large

robustness for a given performance criterion, (b) total quantity of dampers for the robustness function representation and the safety factor representation

Robust Control of Building Structures Under Uncertain Conditions,

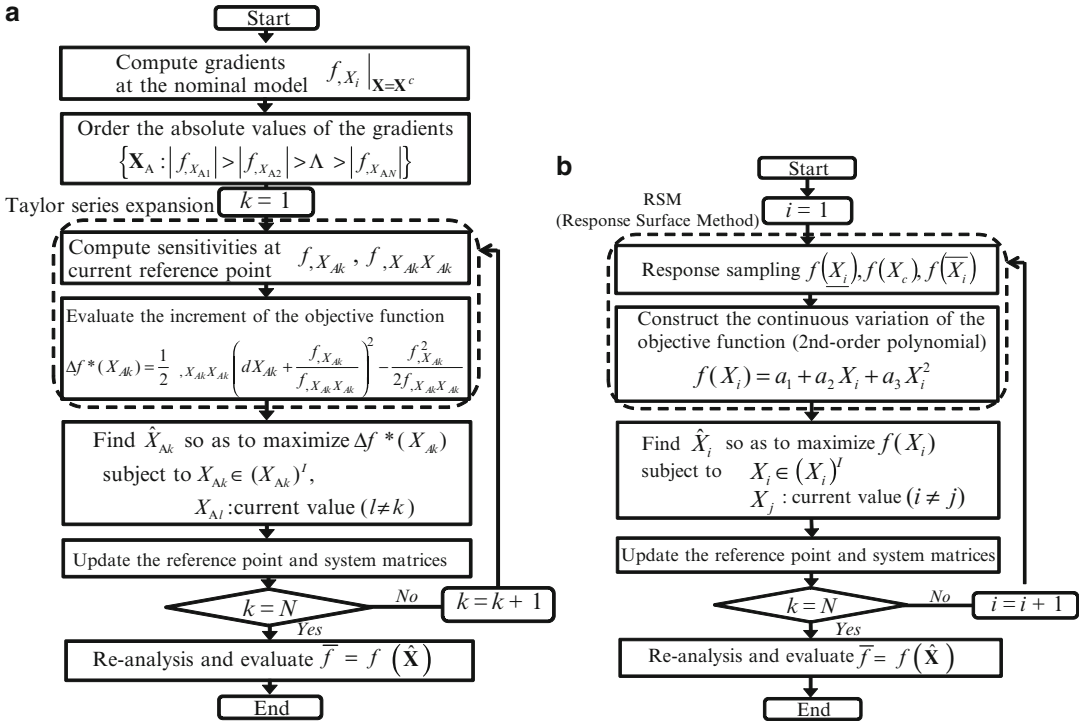
Fig. 16 Degree of robustness in terms of different robustness function



series approximation with diagonal elements only is used to predict this critical combination. Consider the variation $\Delta f_i(dX_i)$ for X_i and let us regard $dX_i = X_i - X_i^c$ as a variable satisfying $-\Delta \bar{X}_i \leq dX_i \leq \Delta \bar{X}_i$. Then the variation $\Delta f_i(dX_i)$ can be rewritten as

$$\Delta f_i(dX_i) = \frac{1}{2} f_{,X_i X_i} (dX_i + f_{,X_i} / f_{,X_i X_i})^2 - f_{,X_i}^2 / 2 f_{,X_i X_i} \quad (17)$$

This is the same as Eq. 11. The function defined by Eq. 17 is a quadratic function and



Robust Control of Building Structures Under Uncertain Conditions, Fig. 17 Flowchart of URP method. (a) Second-order Taylor series approximation, (b) response surface method (RSM)

the value dX_i maximizing or minimizing Δf can be obtained explicitly. Consider the case satisfying $f_{,X_iX_i} < 0$. If the notation $\Delta \bar{X}_i = \Delta \underline{X}_i = \Delta X_i$ is

introduced, the interval parameter maximizing Δf_i can be obtained as follows:

$$\hat{X}_i = \begin{cases} X_i^c - f_{,X_i}/f_{,X_iX_i} & (|f_{,X_i}/f_{,X_iX_i}| \leq \Delta X_i) \\ X_i^c + \Delta X_i & (-f_{,X_i}/f_{,X_iX_i} \geq 0, |f_{,X_i}/f_{,X_iX_i}| > \Delta X_i) \\ X_i^c - \Delta X_i & (-f_{,X_i}/f_{,X_iX_i} < 0, |f_{,X_i}/f_{,X_iX_i}| > \Delta X_i) \end{cases} \quad (18)$$

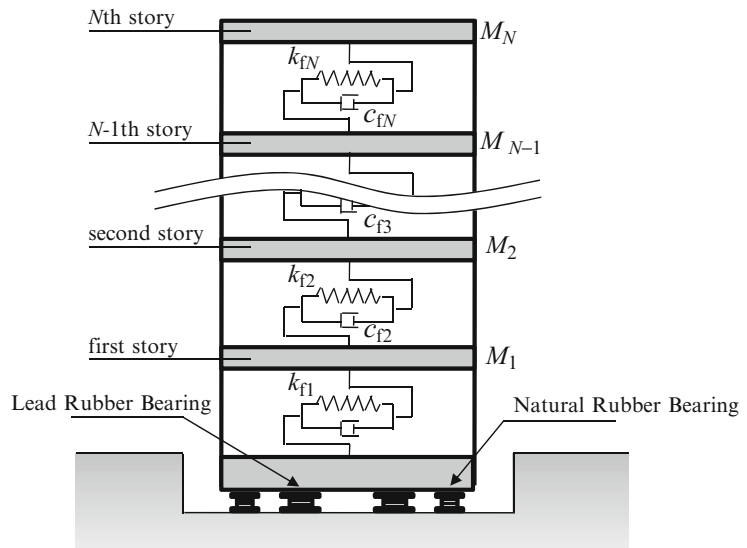
This is the same as Eq. 12. The evaluation of first- and second-order sensitivities $f_{,X_i}$ and $f_{,X_iX_i}$ is made at the nominal model (reference point). It should be noted that the correlation among interval parameters is not taken into account. Here a method called the URP (updated reference-point) method (Fujita and Takewaki 2011a) is used. This method changes the reference point step by step. The flowchart for finding the upper bound of the objective function is shown in Fig. 17a.

Estimation of the Variation of the Objective Function by Response Surface Approach

In the method using Taylor series expansion as explained in previous sections, the accuracy of the robustness evaluation depends directly on the reliability of the numerical sensitivity analysis. For this reason, when the evaluation of numerical sensitivities has some difficulties resulting from the elastic-plastic structural property of isolators, another URP method should be introduced where the variation of the objective function is

Robust Control of Building Structures Under Uncertain Conditions,

Fig. 18 Base-isolated structural model (Fujita and Takewaki 2012a)



estimated using a kind of response surface methods (RSM). RSM is well known as a statistical response evaluation technique which can provide the continuous variation of the objective function by several response samplings. Here three response values $\bar{f}$ evaluated at the nominal, lower and upper ends of an interval parameter X_i are used, and the following second-order polynomial model is introduced:

$$f(X_i) = a_1 + a_2X_i + a_3X_i^2 \quad (19)$$

In Eq. 19 $a_i (i = 1, 2, 3)$ are unknown variables derived from reference responses (some sampling points). By using this approximation of the objective function, the URP method can be applied to the uncertainty analysis as shown in section “[Estimation of the Variation of the Objective Function by 2nd-Order Taylor Series Expansion](#).” Figure 17b shows the flowchart of the URP method using RSM.

Application of URP Method to Base-Isolated Building

Application of the URP method to a base-isolated building model is shown in this section. The validity of the URP method using second-order Taylor series approximation (section “[Estimation of the Variation of the Objective Function by 2nd-Order](#)”

[Taylor Series Expansion](#)”) and RSM (section “[Estimation of the Variation of the Objective Function by Response Surface Approach](#)”) is demonstrated through the comparison with the other method (Monte Carlo Simulation (MCS)). Although the number of interval parameters shown later is large in this simulation, the response analyses are performed for randomly selected combination of endpoints of interval parameters.

Modeling of Base-Isolated Buildings and Uncertainty of Structural Parameters

Consider an N -story base-isolated shear building model as shown in Fig. 18. A simple model is used here to present an essential feature of the robustness evaluation method. Since the use of natural rubber bearings (NRB) only as isolator elements may lead to a higher level of acceleration at the top story, both NRB and lead rubber bearings (LRB) are used here. Furthermore, an additional damping in the base-isolation story is to be given by oil dampers.

It is necessary to take into account the response variability (robustness) of the building under earthquake ground motions due to the variability of mechanical properties of isolators, dampers, and superstructures. Table 2 shows an example of variability ratios (variable band) of isolators of NRB and LRB.

Robust Control of Building Structures Under Uncertain Conditions, Table 2 Variability ratio of structural properties of isolators (NRB and LRB) (Fujita and Takewaki 2012a)

Isolator	Structural parameter	Bound	Manufacturing error (%)	Temperature dependency (%)	Aging deterioration (%)
NRB (natural rubber bearing)	Elastic shear modulus	Upper	+10	+6	+10
		Lower	−10	−5	0
LRB (lead rubber bearing)	Plastic stiffness	Upper	+10	+6	+11
		Lower	−10	−5	0
	Hysteretic strength	Upper	+10	+23	0
		Lower	−10	−21	0

Interval variables in this example are the damping coefficient c_0 and elastic shear stiffness k_0 of NRB, the plastic shear stiffness K_d and elastic shear stiffness K_e of LRB, and the damping coefficient vector $\mathbf{c}_f$ and the horizontal stiffness vector $\mathbf{k}_f$ of the superstructure. The set of interval variables is denoted by

$$\mathbf{X}^I = \{c_0^I, k_0^I, K_d^I, K_e^I, \mathbf{c}_f^I, \mathbf{k}_f^I\} \quad (20)$$

For members with the same properties, it may be appropriate to assign the same variability in each group of members in a building. It should be noted that the present formulation is applicable to such realistic case.

Let us define the ratios of the upper and lower variations of interval variables to the corresponding nominal values.

$$\bar{\alpha} = \left\{ \frac{\Delta \bar{c}_0}{c_0^c}, \frac{\Delta \bar{k}_0}{k_0^c}, \frac{\Delta \bar{K}_d}{K_d^c}, \frac{\Delta \bar{K}_e}{K_e^c}, \frac{\Delta \bar{c}_{f_1}}{c_{f_1}^c}, \dots, \frac{\Delta \bar{k}_{f_1}}{k_{f_1}^c}, \dots \right\} \quad (21a)$$

$$\alpha = \left\{ \frac{\Delta c_0}{c_0^c}, \frac{\Delta k_0}{k_0^c}, \frac{\Delta K_d}{K_d^c}, \frac{\Delta K_e}{K_e^c}, \frac{\Delta c_{f_1}}{c_{f_1}^c}, \dots, \frac{\Delta k_{f_1}}{k_{f_1}^c}, \dots \right\} \quad (21b)$$

The structural property of a 20-story base-isolated building ($N = 20$) is summarized in the following. The floor mass per story is $1,024 \times 10^3$ (kg). This corresponds to $32 \text{ m} \times 32 \text{ m}$ floor plan ($1,000 \text{ kg/m}^2$). The floor mass per story is denoted by m , and the lowest mode of vibration of the super building with fixed base-isolation story is given by a straight line. Then the story stiffness can be expressed as

$$k_{f_i} = \frac{1}{2} \{N(N+1) - i(i-1)\} m \omega_1^2 \quad (i = 1, \dots, N) \quad (22)$$

where ω_1 is the fundamental natural circular frequency of the super building with fixed base-isolation story and is given by 3.93 [rad/s] (natural period = 1.6 [s]). The super building is assumed to behave elastically. The damping coefficients of the super building are given by stiffness-proportional damping with a lowest-mode damping ratio $h = 0.02$.

In the design of the base isolator, the diameter, thickness, and shear modulus have to be determined. The required conditions are that (1) the stress under dead load is within an allowable value and (2) the deformation capacity and the fundamental natural period are appropriate. When only NRB isolators are used, the horizontal stiffness k_0 of the base-isolation story is often determined by specifying the natural circular frequency of the model with the rigid super building. On the other hand, since LRB isolators have the elastic–plastic characteristic, it may be difficult to derive k_0 for the target natural frequency of the base-isolated building. Table 3 shows the property of NRB and LRB isolators of the example model. The fundamental natural circular frequency of this base-isolated building is 1.16 [rad/s], i.e., natural period = 5.40 [s]. The additional damping coefficient of the dampers in the base-isolation story is given by specifying the damping ratio of the model with a rigid super building as 0.1 .

Input Ground Motions

Since it is well known that base-isolated buildings are vulnerable for long-period and

Robust Control of Building Structures Under Uncertain Conditions, Table 3 Parameters of isolators of base-isolated building (Fujita and Takewaki 2012a)

	NRB	LRB
Diameter of rubber [mm]	1100	
Total rubber thickness [mm]	252	
Shear modulus of rubber [N/mm ²]	0.390	0.385
Diameter of lead plug [mm]	200	
Shear modulus of lead plug [N/mm ²]	0.588	
Total number of isolators	8	15

long-duration ground motions (Ariga et al. 2006; Takewaki 2008; Takewaki et al. 2012), six long-period ground motions are used in this example. Tomakomai EW motion is famous in Japan as the first one which brings attention to the long-period ground motion in the design of super high-rise buildings and base-isolated buildings (Ariga et al. 2006). The other ones are the ground motions recorded during the 2011 Tohoku earthquake and its aftershocks. These ground motions are scaled to the same maximum ground velocity = 0.5 [m/s]. This scaling is used in the structural design of high-rise buildings in Japan. Figure 19a–e, f presents the time histories of acceleration and velocity of those ground motions. The velocity response spectra are shown in Fig. 20.

Interval Analysis for Drift of Base-Isolation Story

The uncertainty analysis is shown here for the maximum drift of base-isolation story. The level of variability in Eq. 20 is given by

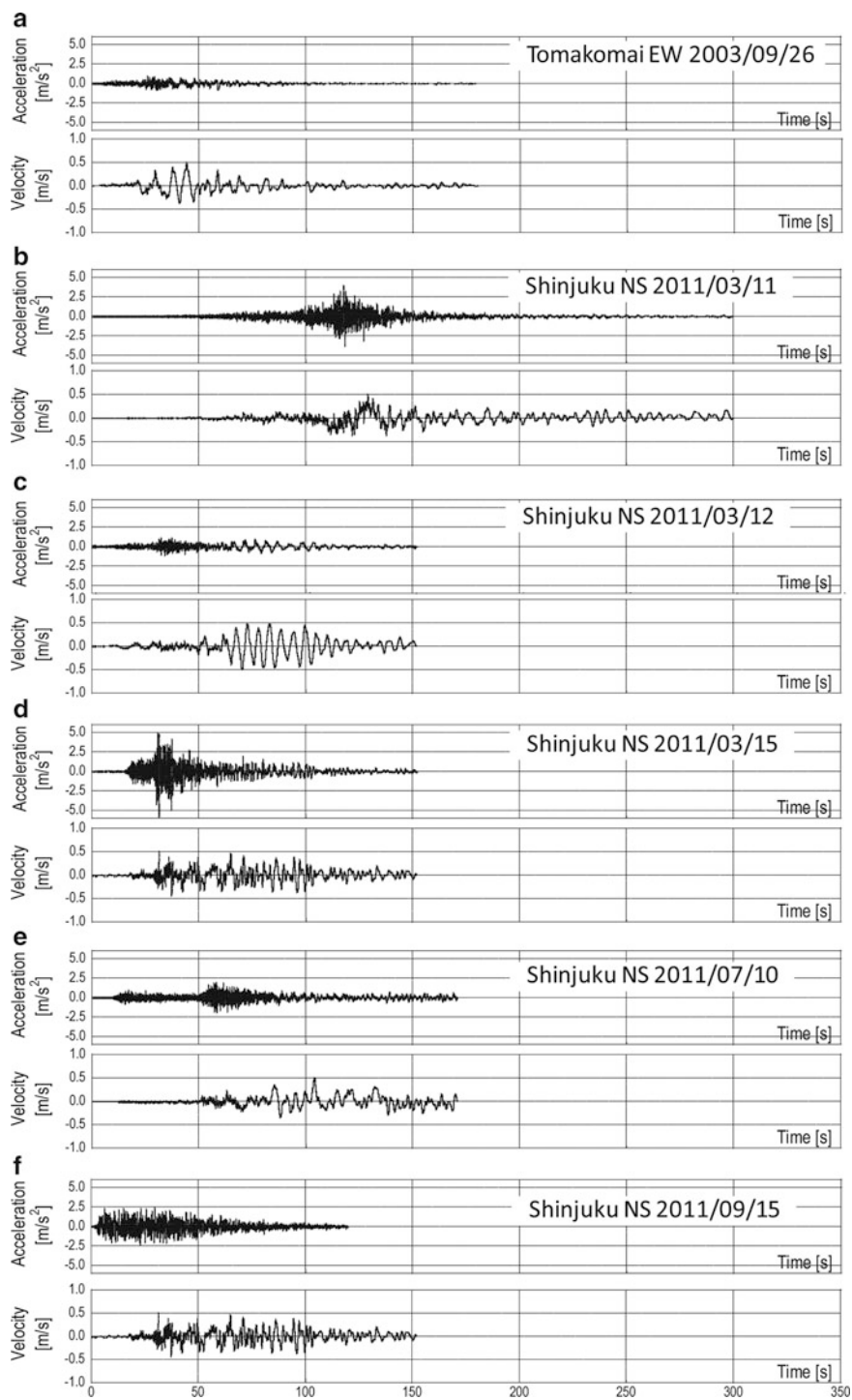
$$\bar{\alpha} = \begin{cases} 0.3 & (i = 1) \\ 0.26 & (i = 2) \\ 0.27 & (i = 3) \\ 0.26 & (i = 4) \\ 0.1 & (i = 5, \dots, N + 4) \\ 0.05 & (i = N + 5, \dots, 2N + 4) \end{cases}, \quad (23)$$

$$\underline{\alpha} = \begin{cases} 0.3 & (i = 1) \\ 0.15 & (i = 2) \\ 0.15 & (i = 3) \\ 0.3 & (i = 4) \\ 0.1 & (i = 5, \dots, N + 4) \\ 0.05 & (i = N + 5, \dots, 2N + 4) \end{cases}$$

where i denotes the interval variable number. The case of $i = 1$ corresponds to the damping coefficient of oil dampers in the base-isolation story. The case of $i = 2$ indicates the elastic shear stiffness k_0 of the NRB. In addition, the case of $i = 3$ gives plastic stiffness K_d of LRB, and the case of $i = 4$ corresponds to elastic shear stiffness K_e of LRB. Here intervals of k_0 and K_d are treated to be unsymmetrical in increasing and decreasing directions. This is due to the aging effect (hardening) of isolators. Furthermore, the case of $i = 5, \dots, N + 4$ provides the damping coefficients $\mathbf{c}_f$ of the super building, and the case of $i = N + 5, \dots, 2N + 4$ corresponds to the story stiffness $\mathbf{k}_f$ of the super building.

Figure 21 shows the upper bounds of the maximum drift of the base-isolation story compared for various methods (URP methods with second-order Taylor series approximation/with RSM and the Monte Carlo Simulation (MCS)). As explained before, the difference between the URP method with second-order Taylor series approximation and that with RSM is how to estimate the variation of the objective function. In the former one, the numerical sensitivities, i.e., the gradient vector and the Hessian matrix, of the objective function are needed. On the other hand, in the latter one, a kind of RSM is applied where appropriate response samplings are made and the gradient vector and the Hessian matrix are evaluated from the constructed approximate function.

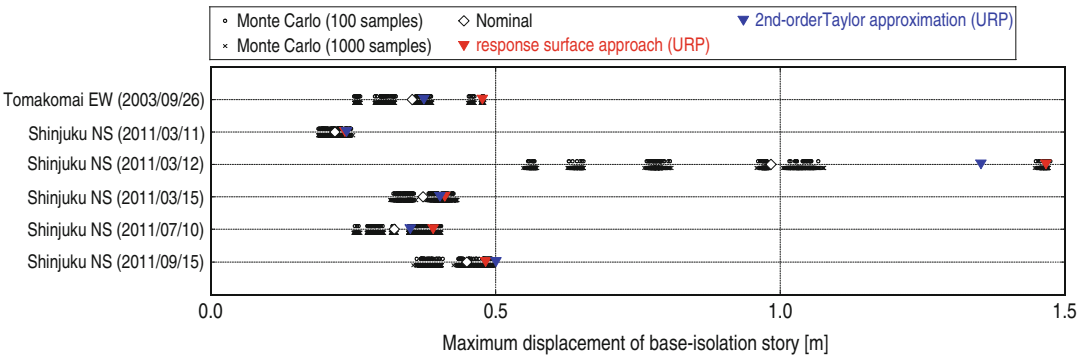
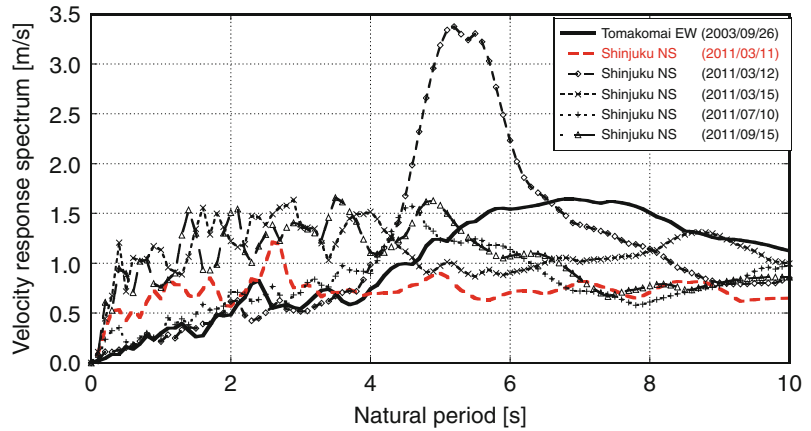
Figure 22 illustrates the comparison of the hysteresis loop and time history of drift in the base-isolation story for models with nominal parameters and critical parameters under Shinjuku NS (2011/03/12). This ground motion is a result of the aftershock for March 11, 2011,



Robust Control of Building Structures Under Uncertain Conditions, Fig. 19 Recorded ground motions (Max. Vel. is 0.5 m/s) (Fujita and Takewaki 2012a)

Robust Control of Building Structures Under Uncertain Conditions,

Fig. 20 Velocity response spectrum ($h = 0.05$) (Fujita and Takewaki 2012a)



Robust Control of Building Structures Under Uncertain Conditions, Fig. 21 Comparison of maximum interstory drift of base-isolation story derived by URP method with MCS (Fujita and Takewaki 2012a)

Tohoku (Japan) earthquake. It can be observed that the critical parameters certainly give a larger response than the nominal parameters.

Figure 23 presents the upper bounds of the maximum top-story floor acceleration by the interval analysis with various methods (URP methods with second-order Taylor series approximation/with RSM and the Monte Carlo simulation). It can be observed from Fig. 23 that the level of variability of the maximum value into the increasing side derived by the URP method with RSM is larger than that of second-order Taylor series approximation. In other words, while the URP method with RSM provides a definite upper bound for the Monte Carlo simulation, the URP method with second-order Taylor series approximation does not necessarily assure the upper bound. This may result from the difficulty

in the sensitivity analysis in the process of the second-order Taylor series approximation for elastic-plastic responses.

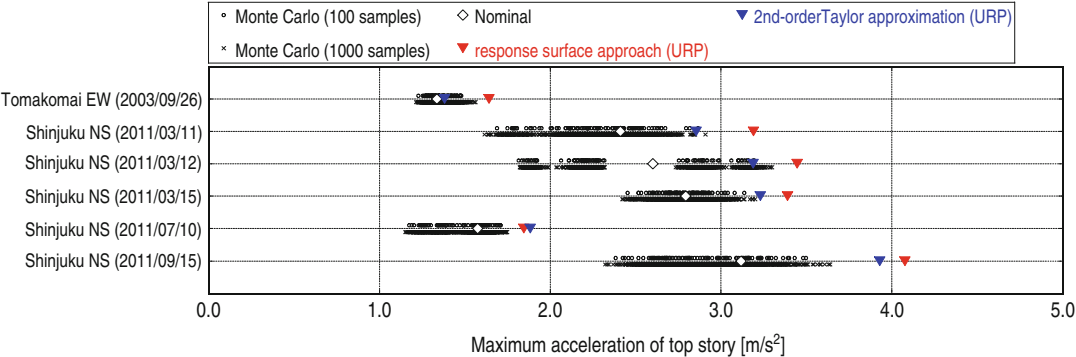
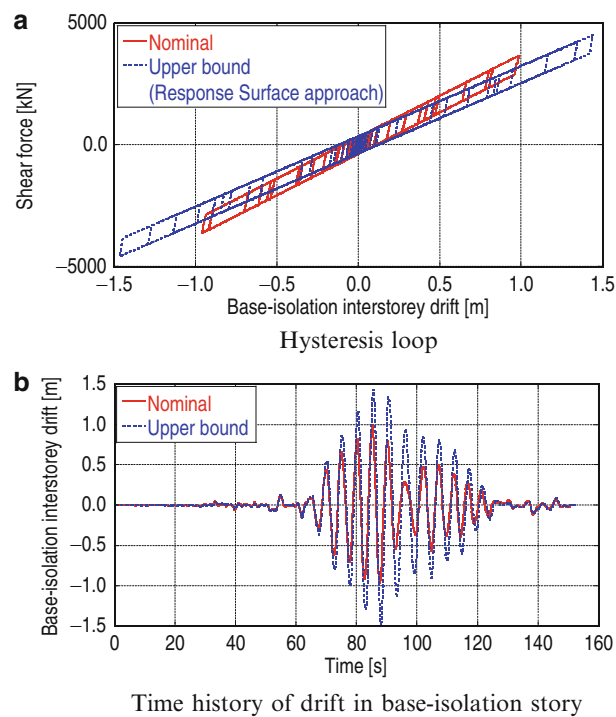
Summary

The design of building structures with greater robustness and reliability can be achieved by taking into account the input disturbance uncertainties and structural parameter uncertainties (Takewaki et al. 2012). In order to get a reliable design with greater robustness, the introduction and definition of robustness are essential.

A robustness function for the constraint on dynamic response of building structures with passive dampers subjected to ground motions has been defined based on the info-gap decision

Robust Control of Building Structures Under Uncertain Conditions,

Fig. 22 Comparison of hysteresis loop and time history of drift in base-isolation story for models with nominal parameters and critical parameters under Shinjuku NS (2011/03/12) (Fujita and Takewaki 2012a)



Robust Control of Building Structures Under Uncertain Conditions, Fig. 23 Comparison of maximum top-story acceleration derived by URP method with MCS (Fujita and Takewaki 2012a)

theory (Ben-Haim 2001). For evaluating the reliable robustness, an efficient uncertainty analysis methodology for the robustness evaluation of a damped structure has been explained which is aimed at finding the upper bound of dynamic response under uncertainties of structural parameters.

In the explained uncertainty analysis method, a model of uncertain parameters has been defined by using an interval model. Although the basic

theorem of “inclusion monotonic” is assumed in some of the interval analyses proposed so far, the critical combination of interval parameters in a feasible domain, not only on the bounds but also in an inner domain of interval parameters, has been derived explicitly by using the explained method. By evaluating the extreme value of the objective function via the approximation of second-order Taylor series expansion, the upper bound of the objective function can be

obtained straightforwardly for the predicted structural parameter set. To evaluate the upper bound of the objective function more accurately within a reasonable task, the URP (updated reference-point) method has been introduced where the reference point to calculate first- and second-order sensitivities has been updated according to the variation of uncertain structural parameters.

Numerical examples using the robustness function has been presented for a 20-story planar shear building including passive viscous dampers with supporting members by applying the URP method. A detailed comparison of the robustness of the structures where the additional dampers have been distributed (a) uniformly, (b) from 11th to 20th story, and (c) optimally has been conducted for representative recorded ground motions. The optimum damper distribution has been derived by the optimization algorithm developed to minimize the maximum amplitude of interstory-drift transfer function. By comparing the robustness functions among the various damper distributions, the large robustness can be obtained in the design of optimal damper distribution especially for the excitation whose predominant frequency is resonant to the fundamental natural frequency of the building structure.

An approximate contour plot of the robustness function with respect to a varied total damper quantity has been constructed. By comparing the contour plots of the robustness function for various damper distributions, a total damper quantity can be derived which satisfies the performance criterion under the uncertainty of the structural parameters. It has also been shown that it is possible to clarify the relation of total quantities of passive dampers for the robustness function representation and the safety factor representation. It is expected that this leads to a more robust damper design.

Application of the URP method to the base-isolated building model with hysteretic responses has been shown. The validity of the URP method using second-order Taylor series approximation and RSM (response surface method) is demonstrated through the comparison with the other method (Monte Carlo Simulation (MCS)).

Cross-References

- ▶ [Lead-Rubber Bearings with Emphasis on Their Implementation to Structural Design](#)
- ▶ [Nonlinear Dynamic Seismic Analysis](#)
- ▶ [Passive Control Techniques for Retrofitting of Existing Structures](#)
- ▶ [Robust Design Optimization for Earthquake Loads](#)
- ▶ [Seismic Strengthening Strategies for Existing \(Code-Deficient\) Ordinary Structures](#)
- ▶ [Structural Optimization Under Random Dynamic Seismic Excitation](#)

References

- Alefeld G, Herzberger J (1983) Introduction to interval computations. Academic, New York
- Ariga T, Kanno Y, Takewaki I (2006) Resonant behavior of base-isolated high-rise buildings under long-period ground motions. *Struct Des Tall Special Build* 15(3):325–338
- Baratta A, Elishakoff I, Zuccaro G, Shinozuka M (1998) A generalization of the Drenick-Shinozuka model for bounds on the seismic response of a single-degree-of-freedom system. *Earthq Eng Struct Dyn* 27: 423–437
- Ben-Haim Y (2001) Information-gap decision theory: decisions under severe uncertainty. Academic, London
- Ben-Haim Y (2006) Information-gap decision theory: decisions under severe uncertainty. Academic, London
- Ben-Haim Y, Elishakoff I (1990) Convex models of uncertainty in applied mechanics. Elsevier, New York
- Ben-Haim Y, Chen G, Soong TT (1996) Maximum structural response using convex models. *J Eng Mech ASCE* 122(4):325–333
- Çelebi M, Okawa I, Kashima T, Koyama S, Iiba M (2012) Response of a tall building far from the epicenter of the 11 March 2011 M 9.0 Great East Japan earthquake and aftershocks. *J Struct Design Tall Special Build* 23 (6):427–441
- Chen SH, Wu J (2004) Interval optimization of dynamic response for structures with interval parameters. *Comp Struct* 82:1–11
- Chen SH, Lian H, Yang X (2003) Interval eigenvalue analysis for structures with interval parameters. *Finite Element Anal Design* 39:419–431
- Chen SH, Ma L, Meng GW, Guo R (2009) An efficient method for evaluating the natural frequency of structures with uncertain-but-bounded parameters. *Comp Struct* 87:582–590
- Conte JP, Pister KS, Mahin SA (1992) Nonstationary ARMA modeling of seismic motions. *Soil Dyn Earthq Eng* 11(7):411–426

- Dong W, Shah H (1987) Vertex method for computing functions of fuzzy variables. *Fuzzy Set Syst* 24(1):65–78
- Elishakoff I, Ohsaki M (2010) Optimization and anti-optimization of structures under uncertainty. Imperial College Press, London
- Fujita K, Takewaki I (2011a) An efficient methodology for robustness evaluation by advanced interval analysis using updated second-order Taylor series expansion. *Eng Struct* 33(12):3299–3310
- Fujita K, Takewaki I (2011b) Sustainable building design under uncertain structural- parameter environment in seismic-prone countries. *Sustain Cities Soc* 1(3):142–151
- Fujita K, Takewaki I (2012a) Robustness evaluation on earthquake response of base-isolated buildings with uncertain structural properties under long-period ground motions. *Architectonica J* 1(1):46–59
- Fujita K, Takewaki I (2012b) Robust passive damper design for building structures under uncertain structural parameter environments. *Earthq Struct* 3(6):805–820
- Fujita K, Moustafa A, Takewaki I (2010a) Optimal placement of viscoelastic dampers and supporting members under variable critical excitations. *Earthq Struct* 1(1):43–67
- Fujita K, Yamamoto K, Takewaki I (2010b) An evolutionary algorithm for optimal damper placement to minimize interstorey-drift transfer function. *Earthq Struct* 1(3):289–306
- Kanno Y, Takewaki I (2006) Sequential semidefinite program for maximum robustness design of structures under load uncertainties. *J Optim Theory Appl* 130(2):265–287
- Koyluoglu HU, Elishakoff I (1998) A comparison of stochastic and interval finite elements applied to shear frames with uncertain stiffness properties. *Comp Struct* 67:91–98
- Minami Y, Yoshitomi S, Takewaki I (2013) System identification of super high-rise buildings using limited vibration data during the 2011 Tohoku (Japan) earthquake. *Struct Control Health Monitor* 20(11):1317–1338
- Moore RE (1966) Interval analysis. Prentice-Hall, Englewood Cliffs
- Mullen RL, Muhanna RL (1999) Bounds of structural response for all possible loading combinations. *J Struct Eng ASCE* 125:98–106
- Pantelides CP, Tzan SR (1996) Convex model for seismic design of structures: I analysis. *Earthq Eng Struct Dyn* 25:927–944
- Qiu ZP (2003) Comparison of static response of structures using convex models and interval analysis method. *Int J Numer Methods Eng* 56:1735–1753
- Takewaki I (2006) Critical excitation methods in earthquake engineering. Elsevier, London
- Takewaki I (2008) Robustness of base-isolated high-rise buildings under code-specified ground motions. *Struct Design Tall Special Build* 17(2):257–271
- Takewaki I (2013) Critical excitation methods in earthquake engineering, 2nd edn. Elsevier, London
- Takewaki I, Ben-Haim Y (2005) Info-gap robust design with load and model uncertainties. *J Sound Vibrat* 288(3):551–570
- Takewaki I, Ben-Haim Y (2008) Info-gap robust design of passively controlled structures with load and model uncertainties, chapter 19. In: Tsompanakis Y, Lagaros N, Papadrakakis M (eds) Structural design optimization considering uncertainties. Taylor & Francis, London, pp 531–548
- Takewaki I, Conte JP, Mahin SA, Pister KS (1991) A unified earthquake-resistant design method for steel frames using ARMA models. *Earthq Eng Struct Dyn* 20(5):483–501
- Takewaki I, Murakami S, Fujita K, Yoshitomi S, Tsuji M (2011) The 2011 off the Pacific coast of Tohoku earthquake and response of high-rise buildings under long-period ground motions. *Soil Dyn Earthq Eng* 31(11):1511–1528
- Takewaki I, Moustafa A, Fujita K (2012) Improving the earthquake resilience of buildings: The worst case approach. Springer, London
- Takewaki I, Fujita K, Yoshitomi S (2013) Uncertainties in long-period ground motion and its impact on building structural design: Case study of the 2011 Tohoku (Japan) earthquake. *Eng Struct* 49:119–134
- Tzan SR, Pantelides CP (1996) Convex models for impulsive response of structures. *J Eng Mech ASCE* 122(6):521–529

Robust Design Optimization for Earthquake Loads

Wolfgang Graf and Marco Götz
Institute for Structural Analysis, Technische
Universität Dresden, Dresden, Germany

Synonyms

Earthquake; Fuzzy analysis; Fuzzy probability-based randomness; Robustness; Uncertain design; Uncertain process; Uncertainty

Introduction

The design of structures, with the help of numerical procedures, is one of the most challenging engineering tasks. The structure has to withstand

all dead loads, live loads, and unpredictable events as earthquakes. Earthquake analysis itself is, due to the high lack of information, sophisticated. The numerical simulation of these structures needs advanced methods to capture the nonlinear dynamic behavior. The recent advancements in the field of finite element analysis (FEA) enable the possibility to simulate the deterministic structural behavior.

But a realistic simulation of structures needs to consider uncertainty for actions and resistances. Hence, the realistic consideration of earthquake loads needs to describe the uncertainty of, e.g., intensity and duration of the earthquake. Therefore, uncertainty models are necessary, and this contribution is focused on the parameterized description of uncertainty by using uncertain variables; see section “[Uncertainty Modeling](#).” A well-established model to describe uncertainty is a random variable, which can model the uncertainty characteristic variability.

But the unfulfillable conditions to apply random variables in an engineering application need models to describe epistemic uncertainty, as expert knowledge and small sample sizes; therefore, the uncertainty model fuzziness is used. The combination of both models, fuzzy probability-based randomness, combines the advantages and is a sophisticated uncertainty model. The necessary numerical algorithms are formulated in section “[Uncertain Structural Analysis](#).”

The consideration of uncertain variables in processes, as earthquake analysis, needs the formulation of algorithms to gather the additional time dependency. This extension of the uncertain analysis to uncertain earthquake analysis is shown in section “[Earthquake Analysis Under Consideration of Uncertain Data](#).” A specific aspect is the numerical efficiency of the approaches; therefore, discussion about the possibility to decrease the computational costs is included.

There are several definitions of the robustness of structures. This contribution uses a robustness measure, based on the comparison of the uncertainty of the input quantities and the output

quantities; see section “[Robustness](#).” The measure applies information reducing measures to quantify the amount of uncertainty of the quantities. This computation allows a relative statement of the robustness of the structure (for the given loads).

The main idea of this contribution is to design robust structures under consideration of earthquake loads. The synthesis of uncertainty (to quantify the robustness of structures) and design of structures under consideration of time-dependent responses (earthquake-induced processes) are described in section “[Numerical Design of Processes with Uncertain Parameters](#).” The design of structures can be done by simply observing several variants or by using optimization technologies. Another possibility, especially in an early stage of design, is the solution of inverse problems. The focus of this contribution is on parameterized optimization, as design technology, with the objective to find robust structures. This is known as robust design optimization (RDO). The consideration of uncertain design variables is solved by a split into deterministic design and uncertain a priori variables. This section shows the necessary mathematical formulations for optimization under consideration of uncertain variables (design and other variables can be uncertain) and a possibility for numerical implementation, by using the numerical analysis concepts for uncertain data, shown in section “[Uncertain Structural Analysis](#).”

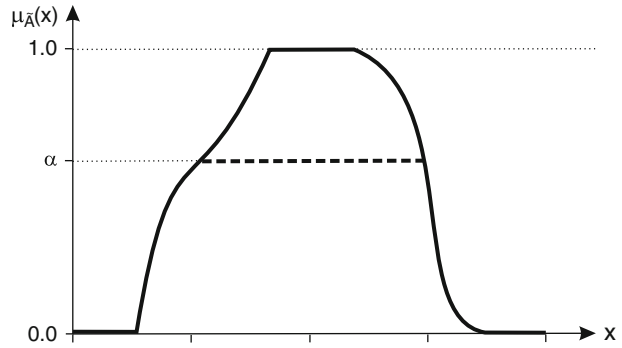
This contribution is focused on numerical procedures for robust design, including uncertainty modeling and computation, for structures under earthquake loads. The example shows the applicability of the approaches; see section “[Example](#).”

Uncertainty Modeling

Uncertainty models can be categorized in two main fields. On the one hand, there are basic uncertainty models, and on the other hand, there are generalized (polymorphic) uncertainty models (Möller and Beer 2004; Beer et al. 2013).

Robust Design Optimization for Earthquake Loads,

Fig. 1 Visualization of a normalized fuzzy set and an α -level



With generalized uncertainty models, more than one uncertainty characteristic can be considered. This contribution will present fuzziness and randomness as basic models and furthermore fuzzy probability-based randomness as a polymorphic model (see Pannier et al. 2013).

Fuzziness

Fuzzy Set

The idea of fuzzy sets is the gradual weighting of a crisp subset $A \in \mathbb{R}$. The crisp subset is defined by the characteristic function

$$\xi_A : \mathbb{R} \rightarrow \{0; 1\}, x \mapsto \begin{cases} 1, & x \in A \\ 0, & x \notin A. \end{cases} \quad (1)$$

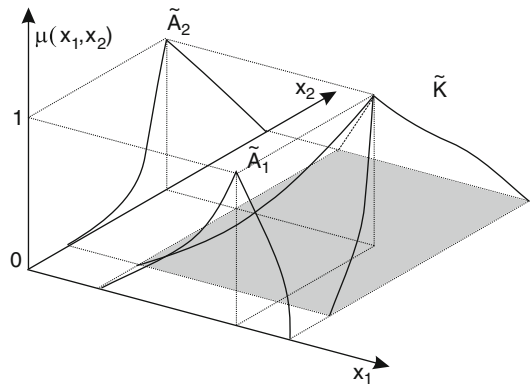
The gradual weighting can be found by the mapping $\mu : A \rightarrow [0, 1]$, i.e., the set of ordered pairs

$$\{(x, \mu(x)) | x \in X \subseteq \mathbb{R}, \mu(x) \geq 0\}. \quad (2)$$

This mapping is the membership function (or the fuzzy set of A). If the membership function μ satisfies $\sup_{a \in A} \mu(A) = 1$, then μ is called a normalized fuzzy set, e.g., see Fig. 1. The membership function can be interpreted as the possibility of occurrence. The uncertainty model fuzzy variable can be used, if reliable information is spare or if only subjective information, as expert knowledge, is available.

Cartesian Product

The membership function of a multidimensional fuzzy set $\mathcal{F}(A)$ is defined by the Cartesian product



Robust Design Optimization for Earthquake Loads,
Fig. 2 Cartesian product of two fuzzy variables

$$\tilde{K} = (\tilde{A}_1 \times \cdots \times \tilde{A}_n), \quad (3)$$

yielding to the membership function

$$\begin{aligned} \mu_{\tilde{K}} : \mathbb{R}^n &\rightarrow [0; 1] \\ &: ((x_1, \dots, x_2) \mapsto \min(\mu_{\tilde{A}_1}(x_1); \dots; \mu_{\tilde{A}_n}(x_1))). \end{aligned} \quad (4)$$

A visualization for a two-dimensional fuzzy set is shown in Fig. 2. In general, the set of all fuzzy sets of A is denoted as $\mathcal{F}(A)$.

α -Level Discretization

In view of a numerical treatment, it is necessary to discretize the fuzzy set $\tilde{A}$ into α -level cuts. Due to the usage of normalized fuzzy sets, the discretization needs to be applied for $\alpha \in (0, 1]$. The α -level cut

$$\tilde{A}_\alpha = \{x \in \mathbb{R} \mid \mu_{\tilde{A}}(x) \geq \alpha\} \quad (5)$$

is a crisp subset $\tilde{A}_\alpha \subseteq \mathbb{R}$. Summarizing, each fuzzy set $\tilde{A}$ is uniquely determined by the family of α -level cuts, that is,

$$\tilde{A} = (\tilde{A}_\alpha)_{\alpha \in (0;1]}. \quad (6)$$

The points of view of this contribution are convex fuzzy variables. Convexity is given, if all α -level cuts are convex, that means for $x_1, x_2 \in \mathbb{R}$ and $\lambda \in [0; 1]$, it holds

$$\mu_{\tilde{A}}(\lambda \cdot x_2 + (1 - \lambda) \cdot x_1) \geq \min(\mu_{\tilde{A}}(x_1), \mu_{\tilde{A}}(x_2)). \quad (7)$$

An example of a convex fuzzy set is the triangular-shaped fuzzy variable $\langle q, r, s \rangle$, with $q, r, s \in \mathbb{R}$ holding

$$\begin{aligned} \langle q, r, s \rangle : \mathbb{R} &\rightarrow [0, 1] \\ : x &\mapsto \begin{cases} \frac{1}{r-q} \cdot (x-q) & \text{for } q < x \leq r, \\ \frac{1}{s-r} \cdot (s-x) & \text{for } r < x \leq s, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (8)$$

Randomness

The probability space (Ω, Σ, P) is composed of a set of elementary events Ω , a σ -algebra Σ and a probability measure P (Pannier et al. 2013). The sample space Ω contains all possible elementary events ω . For continuous random variables, e.g., $\Omega \in \mathbb{R}$, the Borel σ -algebra, denoted by $\mathcal{B}(\mathbb{R})$, contains all possible intervals $I \subseteq \mathbb{R}$. The probability measure P assigns for each event $s \in \Sigma$ a real value in $[0,1]$, representing the probability of s . The probability measure is the mapping

$$P : \Sigma \rightarrow [0, 1], \quad (9)$$

holding the axioms of Kolmogorov:

$$(I) \quad \forall A \in \Sigma : 0 \leq P(A) \leq 1 \quad (10)$$

$$(II) \quad P\phi = 1 - P(\Omega) = 0 \quad (11)$$

$$(III) \quad P\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} P(A_i), \quad (12)$$

with $(A_i)_{i \in \mathbb{N}} \in \Sigma^{\mathbb{N}}$, $A_i \cap A_j (i \neq j, i, j \in \mathbb{N})$.

For a real-valued observation space, the random variable X is $X : \Omega \rightarrow \mathbb{R}$ that satisfies the condition

$$\forall I \in \Sigma : X^{-1}(I) := \{\omega \in \Omega; X(\omega) \in I\} \in \Sigma. \quad (13)$$

The probability measure P_X , related to the random variable X , is defined as

$$\begin{aligned} P_X : \mathcal{B}(\mathbb{R}) &\rightarrow [0; 1] : I \mapsto P_X(I) = P_X(X^{-1}(I)) \\ &= P(\{\omega \in \Omega | X(\omega) \in I\}). \end{aligned} \quad (14)$$

The distribution of a random variable X can be described by a cumulative distribution function (CDF) F_X (see Fig. 3) and the related probability density function (PDF) f_X , respectively, yielding the probability for an interval I ,

$$P_X(I) = \int_I f_X(t) dt = \int_{x_l}^{x_r} f_X(t) dt \text{ with } I = [x_l; x_r]. \quad (15)$$

The CDF F_X contains no information of the underlying probability space, there is just information about the distribution. Therefore, the CDF is analyzed for intervals $I = [-\infty; x]$, such that

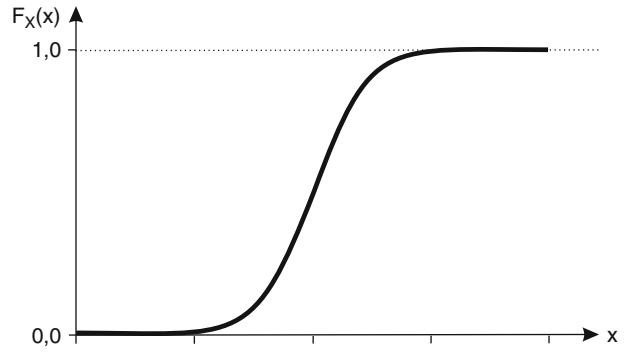
$$F_X(x) = \int_{-\infty}^x f_X(t) dt. \quad (16)$$

For a continuous f_X , the probability of an interval $I = [x_l; x_r]$ is

$$F_X(I) = F_X(x_r) - F_X(x_l). \quad (17)$$

The uncertainty model randomness can be used to model large data sets showing the uncertainty characteristic variability. Furthermore, the

**Robust Design
Optimization for
Earthquake Loads,**
Fig. 3 Visualization of a
random variable



samples need to be independent and identically distributed. Due to small sample sizes and the mostly unfulfilled i.i.d. paradigm, the applicability of random variables for engineering tasks is limited.

Fuzzy Probability-Based Randomness (fp-r)

Definition

A type of a polymorphic uncertainty model is the fuzzy probability-based randomness, taking variability and incompleteness into account. For fuzzy probability-based random variables, the probability measure P of the random number, Eq. 9, is defined as an evaluated set of probability functions. This means that every event is represented by a fuzzy value and not by a real number. The fuzzy probability space is the triple $(\Omega, \Sigma, \hat{P})$. Ω and Σ are the same as in the random number definition. The fuzzy probability $\hat{P}$ is a family of mappings

$$\hat{P} = (\hat{P}_\alpha)_{\alpha \in (0,1]}, \quad (18)$$

where $\hat{P}_\alpha$ assigns to each $A \in Z$ an interval $[\hat{P}_{\alpha,1}(A), \hat{P}_{\alpha,r}(A)]$, such that

$$0 \leq \hat{P}_{\alpha,1}(A) \leq \hat{P}_{\alpha,r}(A) \leq 1 \quad (19)$$

holds. The relating measurable mapping $X : \Omega \rightarrow \mathbb{R}$ is introduced as a fuzzy probability-based random variable. To describe a cumulative distribution function, it is necessary to define the set

$$(F_X)_\alpha := \{G : \mathbb{R} \rightarrow [0;1] \text{ cdf} \mid \forall x \in \mathbb{R} : \hat{P}_{\alpha,l}(X^{-1}((-\infty; x])) \leq G(x) \leq \hat{P}_{\alpha,r}(X^{-1}((-\infty; x]))\}. \quad (20)$$

The fuzzy cumulative distribution function is the family of the sets

$$F_X = ((F_X)_\alpha)_{\alpha \in (0;1]}. \quad (21)$$

For the family F_X , a fuzzy set $\hat{s} = (\hat{s}_\alpha)_{\alpha \in (0,1]}$ can be defined, such that for each $\alpha \in (0,1]$ and for the parameters $s \in \hat{s}_\alpha$ a unique cumulative distribution function F_s can be found. The fp-r variable F_X is formulated by a bunch parameter representation

$$F_X = (\{F_s \mid s \in \hat{s}_\alpha\})_{\alpha \in (0,1]}, \quad (22)$$

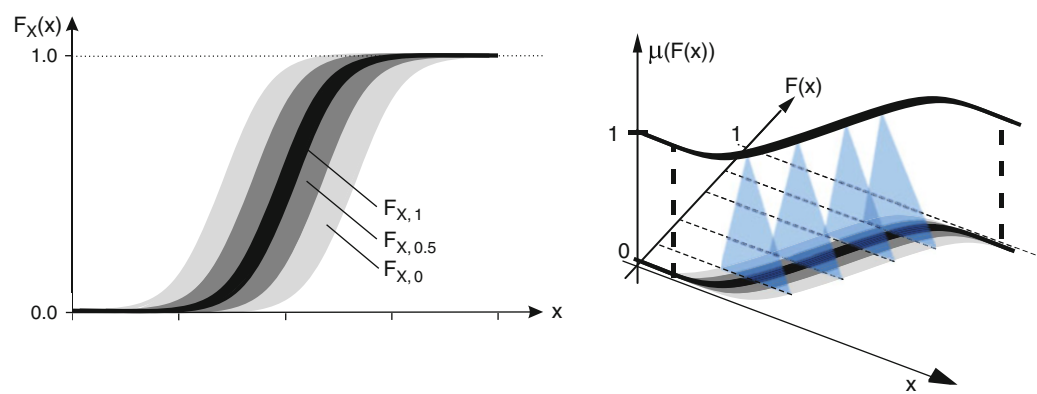
(see also Möller and Beer 2004). In Fig. 4 a visualization of a fuzzy cumulative distribution function and the modeled uncertainty characteristics are shown.

Exemplary Normal Distributed fp-r Variable

An example for a fp-r variable is the normal distributed variable X^{fp-r} , with fuzzy triangular-shaped (Eq. 8) bunch parameters,

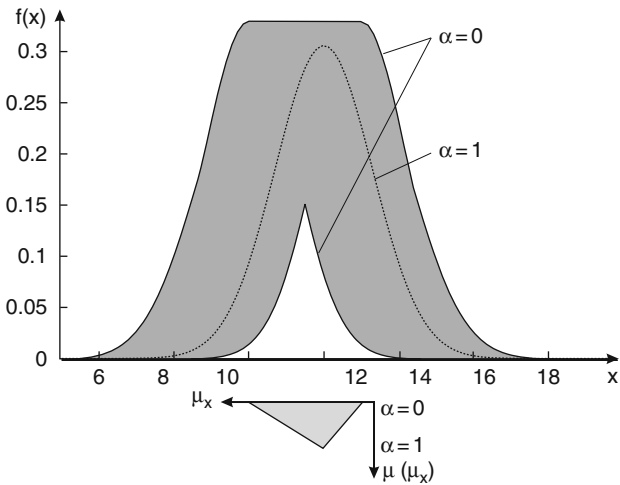
$$X^{fp-r} \sim NV(\mu_x^f = \langle 10, 12, 13 \rangle \sigma_x^f = \langle 1.2, 1.3, 1.4 \rangle). \quad (23)$$

In Fig. 5 all possible probability density functions are shown as grayed area.



Robust Design Optimization for Earthquake Loads, Fig. 4 Visualization of different uncertainty characteristics within a fp-r variable

Robust Design Optimization for Earthquake Loads, Fig. 5 Example of the fuzzy probability-based random density function



Uncertain Structural Analysis

Fuzzy Analysis

Mathematical Formulation

Fuzzy analysis (FA) is the mapping of fuzzy input variables to a fuzzy output variable

$$f^{FA} : \mathcal{F}(A) \rightarrow \mathcal{F}(B) \quad (24)$$

with

$$A \subseteq \mathbb{R}^n \text{ and } B \subseteq \mathbb{R}. \quad (25)$$

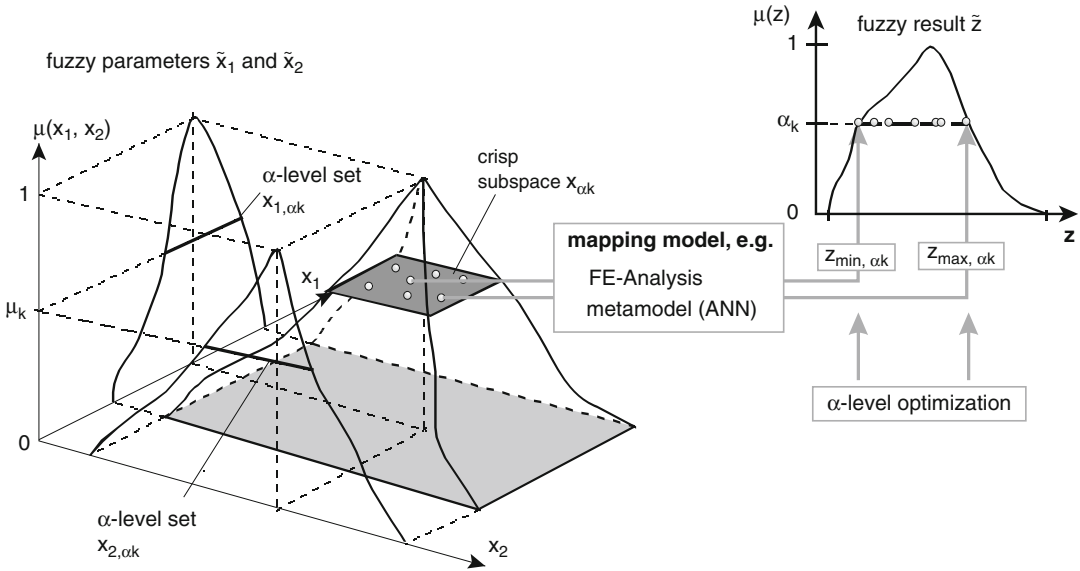
A possibility to apply the mapping f^{FA} is the application of the extension principle (Zadeh 1975; Yager 1986):

on the basis of a deterministic analysis

$$f : X \subset \mathbb{R}^n \rightarrow \mathbb{R} : \underline{x} \mapsto z. \quad (27)$$

The computation of $\hat{x}(a)$ needs to consider Eq. 4. The extension principle is accompanied by a high numerical effort, which is not feasible for practical applications. Therefore, an approximate numerical algorithm, called α -level optimization, was adopted in Möller and Beer (2004).

$$\tilde{\underline{x}} \mapsto \left(z \mapsto \sup_{\underline{a} \in f^{-1}(\{z\})} \tilde{\underline{x}}(\underline{a}) \right), \quad (26)$$



Robust Design Optimization for Earthquake Loads, Fig. 6 Sketch of fuzzy analysis based on α -level optimization

α -Level Optimization

The aim of the α -level optimization is the computation of the membership function of the fuzzy result value $\tilde{z}$. Therefore, the min-max operator of the extension principle (Eq. 26) is replaced by optimization tasks. The necessary discretization of the fuzzy input quantities into α -level is defined in Eq. 5. The main idea of this approach is that the necessary information is the minimum and maximum value for the response only, which can be reached at each α -level $\alpha_k \in (0,1]$. The membership values of the result are

$$\alpha_k = \mu(z_{\min, \alpha_k}) = \mu(z_{\max, \alpha_k}). \quad (28)$$

The elements $z | \mu(z_{\min, \alpha_k}) < z < \mu(z_{\max, \alpha_k})$ have no relevance. The task is the computation of the minimum and maximum result value at each α -level; it can be formulated by two optimization tasks

$$\begin{aligned} z_{\min, \alpha_k} &= f(x_1, \dots, x_n) \\ &\rightarrow \min | (x_1, \dots, x_n) \in X_{\alpha_k} \end{aligned} \quad (29)$$

$$\begin{aligned} z_{\max, \alpha_k} &= f(x_1, \dots, x_n) \\ &\rightarrow \max | (x_1, \dots, x_n) \in X_{\alpha_k} \end{aligned} \quad (30)$$

X_{α_k} is the crisp subset (at α -level α_k) of the Cartesian product of the input quantities and

represents the constraints of the optimization problem. In Fig. 6 the procedure is visualized for two input quantities ($n = 2$) and one output quantity. An efficient optimization algorithm is the modified evolutionary optimization algorithm, introduced in Möller et al. (2000).

Stochastic Analysis

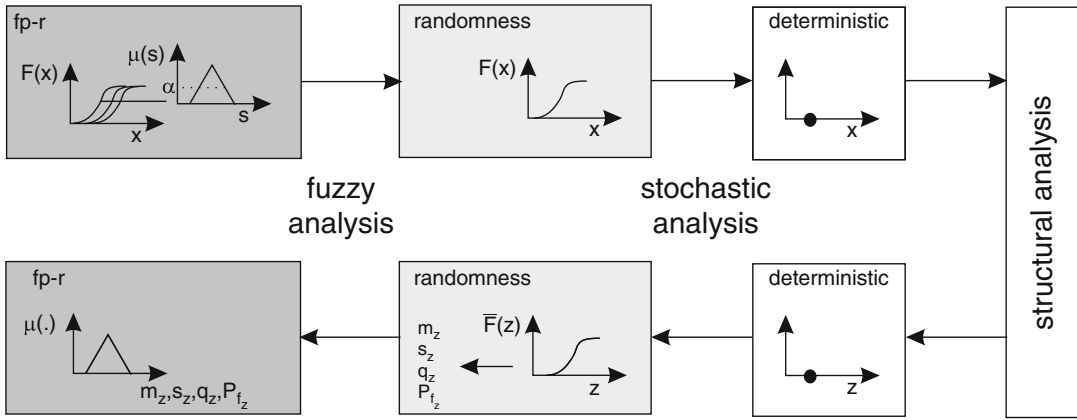
An established method to perform an analysis with stochastic quantities is the well-known Monte Carlo simulation (MCS) method. MCS enables a simple method to compute high-dimensional integrals or statistical moments. An engineering-relevant example is the computation of the probability of failure:

$$P_f = \int_{-\infty}^{\infty} I(g(\underline{x})) \cdot f(\underline{x}) d\underline{x}. \quad (31)$$

The input quantities $\underline{x} \in \mathbb{R}^n$ are characterized by the PDFs $f(\underline{x})$.

The indicator function $I(g(\underline{x}))$ is of binary manner and defines failure or not, in dependency of the limit state functions: $\underline{g}(\underline{x})$

$$I(g(\underline{x})) = \begin{cases} 1 & \text{if } g(\underline{x}) \leq 0 \text{ (failure)} \\ 0 & \text{if } g(\underline{x}) > 0 \text{ (no failure)} \end{cases}. \quad (32)$$



Robust Design Optimization for Earthquake Loads, Fig. 7 Data workflow in uncertain analysis with fp-r quantities

The idea of the MCS is the approximate solution of the integral by analyzing different variants/combinations of input quantities, called samples. Yielding the approximate probability of failure

$$\hat{P}_f = \frac{1}{N} \sum_{i=1}^N I(g(x_i)). \quad (33)$$

The sample size is N and the samples x are considered according to the inverse transformation of random quantities:

$$x_i = F^{-1}(u_i). \quad (34)$$

F^{-1} is the inverse function of the CDF (Eq. 16), and $u_i \in [0,1]$ are arbitrary generated, union-distributed random numbers.

Other information of interest, as empirical mean value m_z and empirical variance s_z , can be computed:

$$m_z = \frac{1}{N} \sum_{i=1}^N z_i \quad (35)$$

$$s_z = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (z_i - m_z)^2}. \quad (36)$$

In this method, samples according their PDFs are generated and simulations are computed; the

resulting set of samples yields to an empirical CDF, which is analyzed for P_f , m_z , s_z , or quantile values. This step can be interpreted as information reduction.

Uncertain Analysis with Fuzzy Probability-Based Random Quantities

The application of polymorphic uncertainty models needs specific sequential numerical algorithms for considering fuzzy probability-based random variables. To apply the approach, the fp-r variables need to be defined by the bunch parameters $\underline{s}$ (Eq. 22). The basic uncertainty models and the other polymorphic uncertainty models are included as special cases.

In general, the algorithm can be separated into fuzzy analysis and stochastic analysis. The fuzzy analysis handles the fuzzy properties of the considered uncertainty model; see section “[Fuzzy Analysis](#).” The stochastic analysis takes the uncertainty characteristic variability into account. It is evaluated by standard or advanced MCS; see section “[Stochastic Analysis](#).” In Fig. 7 the sequential reduction of uncertainty is shown. The application of the fuzzy analysis evaluates the incompleteness characteristic by α -level optimization. The variability of the resulting random variable is evaluated by MCS. The deterministic structural analysis (e.g., FEA) computes for the input values $\underline{x}$ and the output quantities $\underline{z}$. The stochastic analysis results in empirical

distribution functions. These distribution functions need to be evaluated and reduced to a deterministic value, e.g., probability of failure P_f , empirical mean values m_z , or empirical variances s_z . By means of the fuzzy analysis, the membership function of these values can be computed. Summarizing, the sequential reduction of uncertainty allows the usage of basic numerical methods for the analysis considering fp-r and lower-level uncertain quantities. The result of an uncertain analysis with fp-r input quantities is a fuzzy quantity for empirical moments of the stochastic analysis.

Earthquake Analysis Under Consideration of Uncertain Data

Algorithmic Formulation

Under consideration of uncertain data, a deterministic function $f : \mathbb{R}^n \rightarrow \mathbb{R} : \underline{x} \mapsto f(\underline{x})$ becomes

$$f^u : \mathcal{D}(Q, R) \rightarrow \mathcal{D}(V, W) : \underline{p}^u \mapsto f^u(\underline{p}^u). \quad (37)$$

The set of uncertain input parameters is abbreviated by $\mathcal{D}(Q, R)$, neglecting further properties. For example, $\mathcal{D}(\mathbb{R}, [0, 1])$ is a set of fuzzy numbers. The variable $\underline{p}^u \in \mathcal{D}(Q, R)$ contains all uncertain parameters. The uncertainty of parameters can be observed for actions and resistance. This means that, e.g., dead loads, earthquake accelerations, or material parameters cannot be described by a deterministic value.

If the uncertain function depends on parameters $\underline{t} = \{\tau, \underline{\theta}, \underline{\varphi}\}$ as time $\tau \in \mathbb{R}^+$, spatial coordinates $\underline{\theta} \in \mathbb{R}^3$, and other parameters $\underline{\varphi} \in \mathbb{R}^{n_{op}}$ with $\underline{t} \in \mathbb{R}^{n_p}$, it yields

$$\begin{aligned} \hat{f}^u : \mathbb{R}^{n_p} \times \mathcal{D}(Q, R) &\rightarrow \mathbb{R}^{n_p} \times \mathcal{D}(V, W) \\ : (\underline{t}, \underline{p}^u) &\mapsto \hat{f}^u(\underline{t}, \underline{p}^u), \end{aligned} \quad (38)$$

with the total number of additional parameters n_p . For a numerical implementation, a discretization of the dependent parameters is necessary (see Möller et al. 2009). For time dependency

only, the uncertain function is called uncertain process and Eq. 38 results in

$$\begin{aligned} \hat{f}_\tau^u : \mathbb{R} \times \mathcal{D}(Q, R) &\rightarrow \mathbb{R} \times \mathcal{D}(V, W) \\ : (\tau, \underline{p}^u) &\mapsto \hat{f}^u(\tau, \underline{p}^u). \end{aligned} \quad (39)$$

For a nonlinear earthquake analysis (time-history analysis), the general equation of motion in the incremental form

$$\underline{M}\Delta\ddot{\underline{v}}(\tau) + \underline{D}\Delta\dot{\underline{v}}(\tau) + \underline{K}_T\Delta\underline{v}(\tau) = \Delta\underline{R}(\tau) \quad (40)$$

is to be solved. The displacement, velocity, and acceleration are described by $\underline{v}(\tau)$, $\dot{\underline{v}}(\tau)$, and $\ddot{\underline{v}}(\tau)$, respectively. The matrix $\underline{M}$ contains the nodal masses and the matrix $\underline{D}$ the nodal damping. The tangential stiffness of the current time step is the content of the matrix $\underline{K}_T$. The external nodal loads are composed in the load vector $\underline{r}(\tau)$. For earthquake loads, that means extrinsic accelerations of the foundation parts of a structure, the displacements can be separated into rigid body displacements $\underline{v}_{\text{rig}}$ and an according relative displacement $\underline{v}_{\text{rel}}$,

$$\underline{v}(\tau) = \underline{v}_{\text{rig}}(\tau) + \underline{v}_{\text{rel}}(\tau). \quad (41)$$

Yielding the modified equation of motion

$$\begin{aligned} \underline{M}\Delta(\ddot{\underline{v}}_{\text{rig}}(\tau) + \ddot{\underline{v}}_{\text{rel}}(\tau)) + \underline{D}\Delta(\dot{\underline{v}}_{\text{rig}}(\tau) + \dot{\underline{v}}_{\text{rel}}(\tau)) \\ + \underline{K}_T\Delta(\underline{v}_{\text{rig}}(\tau) + \underline{v}_{\text{rel}}(\tau)) = \Delta\underline{R}(\tau). \end{aligned} \quad (42)$$

Under consideration of the independence of internal forces and rigid body displacements and the independence of material damping and rigid body velocities, it yields

$$\begin{aligned} \underline{M}\Delta(\ddot{\underline{v}}_{\text{rig}}(\tau) + \ddot{\underline{v}}_{\text{rel}}(\tau)) + \underline{D}\Delta(\dot{\underline{v}}_{\text{rel}}(\tau)) \\ + \underline{K}_T\Delta(\underline{v}_{\text{rel}}(\tau)) = \Delta\underline{R}(\tau). \end{aligned} \quad (43)$$

If the earthquake acceleration is the only load to be considered ($\Delta\underline{R}(\tau)$), the equation of motion, which is to be solved, can be simplified as

$$\begin{aligned} \underline{M}\Delta(\ddot{\underline{v}}_{\text{rel}}(\tau)) + \underline{D}\Delta(\dot{\underline{v}}_{\text{rel}}(\tau)) + \underline{K}_T\Delta(\underline{v}_{\text{rel}}(\tau)) \\ = \underline{M}\Delta(\ddot{\underline{v}}_{\text{rig}}(\tau)) \end{aligned} \quad (44)$$

The inertial displacements of the accelerated soil $\underline{v}_{\text{soil}}(\tau)$ can be considered with

$$\underline{v}_{\text{rig}}(\tau) = \underline{Z} \cdot \underline{v}_{\text{soil}}(\tau), \quad (45)$$

by defining an allocation vector $\underline{Z}$. The solution gives the time-dependent structural response. For a numerical applicability, the necessary discretization into time steps $\tau_i = \tau_{i-1} + \Delta\tau i \in \{1, 2, \dots, n_t\}$ yields

$$\begin{aligned} f_{\tau=\tau_i}^u : \mathcal{D}(Q, R) &\rightarrow \mathcal{D}(V, W) \\ : \underline{p}^u &\mapsto \hat{f}_{\tau}^u(\tau_i, \underline{p}^u). \end{aligned} \quad (46)$$

This discretization is realized by implicit and explicit time integration methods in the framework of FEA.

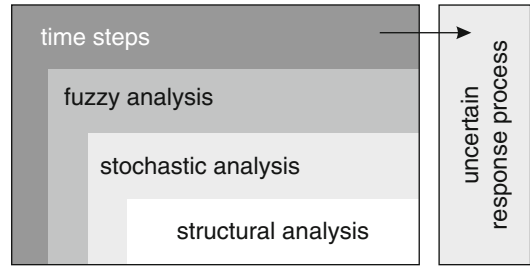
It has to be remarked that the formulations shown in the following, for the consideration of time dependency, are only valid if the uncertain structural parameters are not time dependent.

Numerical Variants for the Analysis of Uncertain Dynamical Processes

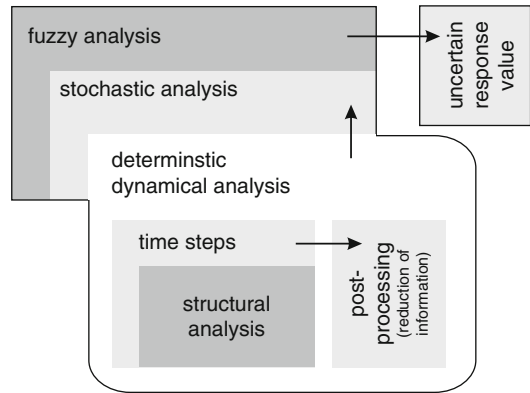
The consideration of polymorphic uncertain variables, e.g., fp-r variables, in a numerical procedure, can be accomplished by a sequential reduction of the level of uncertainty; see section “Uncertain Analysis with Fuzzy Probability-Based Random Quantities.”

The time dependency of additional parameters can be considered by two different approaches. On the one hand, there is variant I, which needs to compute the uncertain analysis for each time step $\tau_i \in \tau$. The result is an uncertain process of the response variable. This concept is highly time-consuming. The structure of the algorithm for fp-r variables is shown in Fig. 8.

On the other hand, in variant II the dynamical process is computed in the inner sequence. This concept is less time-consuming, due to the necessary reduction of information for the deterministic dynamical process. This means that the response process has to be reduced to a deterministic value, e.g., average or maximal value. The result is one uncertain response value, according to the selected information reducing measure. The algorithm is sketched in Fig. 9.



Robust Design Optimization for Earthquake Loads, Fig. 8 Variant I for uncertain dynamical processes



Robust Design Optimization for Earthquake Loads, Fig. 9 Variant II for uncertain dynamical processes

Increasing Numerical Efficiency

The applicability of polymorphic uncertainty models for earthquake analysis is, due to the enormous numerical effort, limited. The consideration of polymorphic uncertainty models makes it necessary to reduce the computational effort. There are three essential possibilities:

- Reduction of calculation time for the deterministic analysis or reduction of necessary number of evaluations of the deterministic solutions
- Replacement of the deterministic solution with efficient metamodels
- Parallel evaluation of deterministic simulations

The reduction of the necessary number of evaluations of the deterministic analysis is possible for polymorphic uncertainty models, due to

the sequential evaluation of fuzzy and stochastic properties; see section “[Uncertain Structural Analysis](#),” for these two sequences. The stochastic solution can be applied by standard MCS or by more efficient technologies as subset sampling. The solution for fuzzy quantities can be improved by applying efficient optimization technologies for the α -level optimization or a specific evaluation of needed points of the membership function.

Both have in common that the numerical effort increases with growing numbers of uncertain parameters. This means that reducing the number of uncertain quantities improves the numerical efficiency. This necessary categorization of uncertain quantities into relevant and nonrelevant is possible by analyzing the sensitivity for each input quantity according to the output quantity of interest (see, e.g., Saltelli et al. 2008). Metamodels are versatile, e.g., they are applicable for pattern classification, function approximation, or computing sensitivity measures. The possibility to approximate functional data, a mapping of input quantities to output quantities

$$f^* : \mathbb{R}^n \supset H^n \rightarrow \mathbb{R}^m \quad (47)$$

can be done by several types of metamodels (see Søndergaard 2003). The definition of the region of interest $H^n = [a_1; b_1] \times \cdots \times [a_n; b_n] | a_i, b_i \in \mathbb{R}, a_i \leq b_i, i \in \{1, \dots, n\}$ is necessary to avoid extrapolating. The metamodel f^* can be found on the basis of a set of support points $\mathcal{N} = \{(\underline{x}, \underline{z}) | \underline{x} \in H^n, \underline{z} \in H^m\}$, which provide the relations between input and output quantities of the original function $f : \mathbb{R}^n \rightarrow \mathbb{R}^m, \underline{x} \mapsto \underline{z}$ pointwise. There are well-established metamodels to approximate functional data, as artificial neural networks (ANN) or radial basis function networks (RBFN).

Robustness

Robustness Measure

Several interpretations of robustness are known and introduced. Robustness can be the resistance against extraordinary events, the limited variations of structural response and redundancy,

prevention of progressive failure in the case of local failure, and the ratio of direct risks to indirect risks. A main contribution in the field of designing robust structures is developed by Beer and Liebscher (2008). The interpretation of robustness in this contribution is the sophisticated ratio of uncertainty of input quantities $\underline{x}^u$ to the uncertainty of result quantities $\underline{z}^u$. The robustness measure R was introduced in Graf et al. (2010) and Sickert et al. (2010):

$$R = \frac{\sum_{i=1}^l k_i \cdot P_i(\mathcal{U}(x_i^u)) + \sum_{k=l+1}^n k_k \cdot P_k(\mathcal{U}(x_k^u))}{\sum_{i=1}^l k_i \cdot \left(\sum_{j=1}^m k_{i,j} \cdot P_{i,j}(\mathcal{U}(z_{i,j}^u)) \right)}. \quad (48)$$

The robustness measure for a structure is analyzed considering the uncertain quantities with n uncertain input parameters x_w^u . In contrast, l load and failure scenarios, represented by the uncertain load input variables, are considered separately. There are m uncertain result variables z_w^u , which have to be taken into account. All uncertain variables $y^u \in \{x_w^u, z_w^u\}$ have in common that the measures describing the uncertainty

$$\mathcal{U} : \mathcal{D}(Q, R) \rightarrow \mathbb{R} : y^u \mapsto \mathcal{U}(y^u) \quad (49)$$

have to be found; selected measures are described in section “[Information Reducing Measures for Uncertain Quantities](#).” The measures need to be selected for the related uncertainty model and a deliberated selection has to provide the comparability of the measures. The penalty functions $P_w(\cdot)$ and the weighting factors k_w are necessary to enable the possibility of a problem-dependent adaption of the robustness measure. The proposed robustness measure is relative and does not represent an absolute value about the robustness of structures. The main benefit is the independence from the used uncertainty model.

Information Reducing Measures for Uncertain Quantities

Information reducing measures $\mathcal{M}$ map uncertain quantities to deterministic values:

$$\mathcal{M} : \mathcal{D}(Q, R) \rightarrow \mathbb{R}. \quad (50)$$

In general, the measures for uncertain quantities can be classified into representative values $\mathfrak{M}$ and values to quantify the amount of uncertainty $\mathfrak{U}$. For fuzzy quantities, the reduction of information is called defuzzification. This contribution shows information reducing measures for random and fuzzy variables. To avoid confusion, random variables $Y \in \mathcal{D}(\Omega, \mathbb{R})$ are defined by the CDF $V_y: \mathbb{R} \rightarrow [0; 1]$ and the PDF $v_y: \mathbb{R} \rightarrow [0; \infty)$. A fuzzy variable $\tilde{y} \in \mathcal{F}(\mathbb{R})$ is given by the membership function $\mu_{\tilde{y}}: \mathbb{R} \rightarrow [0; 1]$ and by the family of α -level cuts, according to Eq. 6.

Representative Measures

For random variables, the representative measures

- Mean value

$$\mathfrak{M}_E^r(Y) = \int_{-\infty}^{\infty} y \cdot v_Y(y) dy \quad (51)$$

- Quantile value

$$\mathfrak{M}_Q^r(Y) = V_Y^{-1}(q) | q \in (0, 1) \quad (52)$$

can be found. For fuzzy variables, the representative measures mean value, level rank method, and extremal values can be found. The mean value, also centroid method, is equal to the centroid of the area of the membership function. The level rank method represents the mean of the center point of each α -level (see Rommelfanger 1988). The extremal values can be used for “worst-case” observations.

- Mean value

$$\mathfrak{M}_E^f(\tilde{y}) = \frac{1}{\mathfrak{U}_A^f(\tilde{y})} \int_{-\infty}^{\infty} y \cdot \mu_{\tilde{y}}(y) dy \quad (53)$$

- Level rank method

$$\mathfrak{M}_L^f(\tilde{y}) = \frac{1}{i} \sum_{k=1}^i \frac{\tilde{y}_{\alpha_{k,l}} + \tilde{y}_{\alpha_{k,r}}}{2} \quad (54)$$

- Minimum

$$\mathfrak{M}_{\text{Min}}^f(\tilde{y}) = \tilde{y}_{\alpha_{0,l}} \quad (55)$$

- Maximum

$$\mathfrak{M}_{\text{Max}}^f(\tilde{y}) = \tilde{y}_{\alpha_{0,r}} \quad (56)$$

Uncertainty Measures

The measures describing the uncertainty for random quantities are

- Variance

$$\mathfrak{U}_V^r(Y) = \int_{-\infty}^{\infty} v_y(y) \cdot (y - \mathfrak{M}_E^r(Y))^2 dy \quad (57)$$

- Entropy

$$\mathfrak{U}_E^r(Y) = \int_{-\infty}^{\infty} v_y(y) \cdot \ln(v_y(y)) dy \quad (58)$$

The uncertainty measures for fuzzy variables are

- Area

$$\mathfrak{U}_A^f(\tilde{y}) = \int_{-\infty}^{\infty} \mu_{\tilde{y}}(y) dy \quad (59)$$

- Variance

$$\mathfrak{U}_V^f(\tilde{y}) = \frac{1}{\mathfrak{U}_A^f(\tilde{y})} \int_{-\infty}^{\infty} \mu_{\tilde{y}}(y) \cdot (y - \mathfrak{M}_E^f(\tilde{y}))^2 dy \quad (60)$$

- Entropy

$$\mathfrak{U}_E^f(\tilde{y}) = -\frac{1}{\ln(2)} \int_{-\infty}^{\infty} \mu_{\tilde{y}}(y) \cdot \ln(\mu_{\tilde{y}}(y)) + (1 - \mu_{\tilde{y}}(y)) \cdot \ln(1 - \mu_{\tilde{y}}(y)) dy \quad (61)$$

Representative and uncertainty measures of fp-r variables can be found in Pannier (2011).

Numerical Design of Processes with Uncertain Parameters

The numerical design of a process can be accomplished by different procedures. The easiest way is to analyze different variants; other procedures are optimization or the solution of the inverse problem. The aim of the design can be described by an objective function for, e.g., efficiency, profitability, reliability, robustness, or sustainability. These goals can be formulated as various constraints or objective functions.

The method, proposed in this contribution, is an optimization approach for parameterized optimization problems. The idea is to find a combination of various uncertain parameters yielding the best uncertain response, whereas the best design is related to the highest robustness.

Uncertain Optimization Task and Split of Uncertain Variables

The optimization objective function under consideration of uncertain quantities is written as

$$\hat{f}_Z^u : \mathcal{D}(Q, R) \rightarrow \mathcal{D}(V, W). \quad (62)$$

The uncertain input quantities $\mathcal{D}(Q, R)$ can be classified into uncertain design variables $\underline{x}_d^u \in \mathcal{D}(Q_x, R_x)$ and uncertain a priori parameters $\underline{p}^u \in \mathcal{D}(Q_p, R_p)$, holding

$$\mathcal{D}(Q_x, R_x) \times \mathcal{D}(Q_p, R_p) = \mathcal{D}(Q, R). \quad (63)$$

The value of a priori parameter is independent of the current design. But, an optimization algorithm for the direct consideration of polymorphic uncertain design variables does not exist. This means that the application of uncertain design variables is not possible, directly. A solution is the application of an affine transformation $\mathcal{T}$ to the uncertain design variables. This transformation splits them to deterministic design variables $\underline{d} \in \mathbb{R}^{n_x}$ (suitable for optimization algorithms) and further (constant) uncertain a priori variables, containing the set $\underline{p}^u$. The affine transformation has to be defined for each uncertainty model.

Random Variables

$$\begin{aligned} \mathcal{T}^r : \mathbb{R}^{n_x} &\rightarrow \mathcal{D}(\Omega, \mathbb{R}^{n_x}) \\ : \underline{d} &\mapsto (\underline{\omega} \mapsto c(\underline{d}) \cdot U(\underline{\omega}) + \underline{d}) \\ &= \underline{X}_d | U \in \mathcal{D}(\Omega, \mathbb{R}^{n_x}) \end{aligned} \quad (64)$$

Fuzzy Variables

$$\begin{aligned} \mathcal{T}^f : \mathbb{R}^{n_x} &\rightarrow \mathcal{F}(\mathbb{R}^{n_x}) \\ : \underline{d} &\mapsto (c(\underline{d}) \cdot \underline{u} + \underline{d} \mapsto \tilde{u}(\underline{u})) \\ &= \tilde{\underline{X}}_d \quad | \tilde{\underline{X}}_d \in \mathcal{F}(\mathbb{R}^{n_x}) = \mathcal{D}(\mathbb{R}^{n_x}, [0, 1]) \end{aligned} \quad (65)$$

Fuzzy Probability-Based Random Variables

$$\mathcal{T}^{fp-r} = \mathcal{T}^r \text{ (due to not involved probability measure)} \quad (66)$$

The application of the affine transformation yields to the objective function (Eq. 62)

$$\begin{aligned} f_Z^u : \mathbb{R}^{n_x} \times \mathcal{D}(Q_p, R_p) &\rightarrow \mathcal{D}(V, W) \\ : (\underline{d}, \underline{p}^u) &\mapsto \hat{f}_Z^u(\mathcal{T}(\underline{d}), \underline{p}^u). \end{aligned} \quad (67)$$

The “uncertain” optimization task can be formulated as

$$\begin{aligned} &\text{determine } L \subseteq X_d^+ \text{ such that } \forall \underline{d} \in X_d^+ \text{ and } \underline{d}_{min} \in L \text{ it holds :} \\ &f_Z^u(\underline{d}_{min}, \underline{p}^u) \text{ is “lower or equal than” } f_Z^u(\underline{d}, \underline{p}^u). \end{aligned}$$

The objective is to find

$$\underset{\underline{d} \in X_d^+}{\text{minimum}} f_Z^u(\underline{d}, \underline{p}^u), \quad (68)$$

under consideration of the uncertain permissible range

$$\begin{aligned} X_d^+ &= \{\underline{d} \in \mathbb{R}^{n_x} | \forall j \in \{1; \dots; a_g\} : \\ g_j^u(\underline{d}, \underline{p}^u) \text{ “lower than” } 0 \\ \wedge \forall k \in \{1; \dots; a_h\} : h_k^u(\underline{d}, \underline{p}^u) \text{ “equal” } 0\} \end{aligned}, \quad (69)$$

with the uncertain constraints

$$\begin{aligned} g_j^u : \mathcal{D}(Q, R) &\rightarrow \mathcal{D}(V, W) \text{ and } h_k^u \\ : \mathcal{D}(Q, R) &\rightarrow \mathcal{D}(V, W). \end{aligned} \quad (70)$$

Due to the missing of a general rule for comparing uncertain quantities, relational operators as $<$, $>$, and $=$, used in the introduced design task and the related constraints and the usually used minimum operator “min” for a deterministic case in Eq. 68, are not applicable. This means the solution of the design task with uncertain quantities can be found for surrogate problems only. The surrogate problem can be formulated as passive or active approach (see Pannier 2011; Götz et al. 2013). Both approaches mainly differ in the order of evaluation of uncertainty task and solving the optimization problem. This contribution shows an algorithm for computing a robust design by an application of the active approach.

Numerical Optimization Algorithm for Consideration of Uncertain Quantities

The active approach is also denoted as *here-and-now* strategy (Tintner 1960). The generalized formulation for polymorphic uncertain quantities is

$$\begin{aligned} &\text{determine } L \subseteq X_d^+ \text{ such that} \\ &\forall \underline{d} \in x_d^+ \text{ and } \underline{d}_{min} \in L \text{ it holds :} \\ &\mathcal{M}\left(f_z^u(\underline{d}_{min}, p^u)\right) \leq \mathcal{M}\left(f_z^u(\underline{d}, p^u)\right). \end{aligned}$$

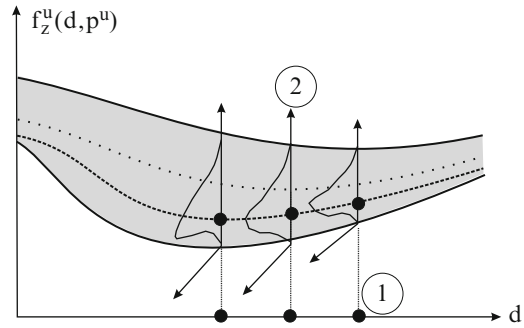
Selected deterministic design quantities d (Fig. 10, ①) were transformed to uncertain design quantities x_d^u and the uncertain result quantities (Fig. 10, ②) can be computed. The application of information reducing measures

$$\mathcal{M} : \mathcal{D}(V, W) \rightarrow \mathbb{R} \quad (71)$$

allows the reduction to deterministic values and further optimization processing; see section “Information Reducing Measures for Uncertain Quantities.” The numerical algorithm, for the design of an uncertain process type variant II (Fig. 9), is shown in Fig. 11.

There are seven major steps:

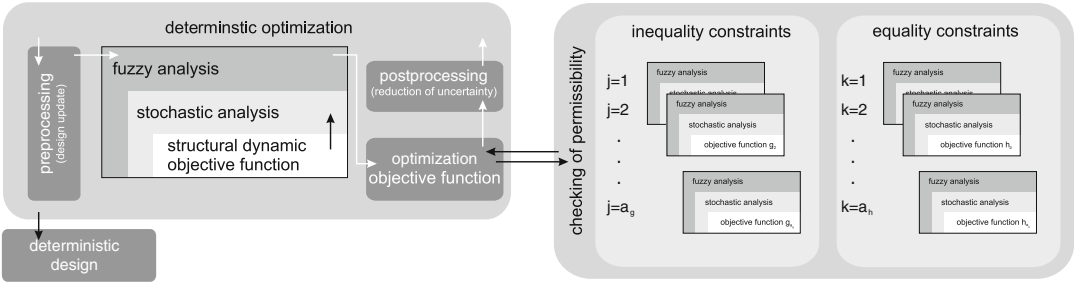
1. The preprocessing provides the objective function, with deterministic input variables, according to Eq. 67. The actual deterministic design variables and the uncertain a



Robust Design Optimization for Earthquake Loads, Fig. 10 Optimization procedure for uncertain objective functions

priori parameter are combined by applying the affine transformation.

2. The uncertainty analysis for processes, with the sequential reduction of uncertainty, computes the uncertain result; see section “Numerical Variants for the Analysis of Uncertain Dynamical Processes.” For variant II it results in an uncertain response, representing the main characteristics of the uncertain process.
3. The structural dynamic objective function is the core of this design algorithm; it provides the deterministic relations between deterministic inputs and deterministic (time-dependent) outputs. This part could be applied by an FEA, but numerical efforts should be reduced by the application of metamodels; see section “Increasing Numerical Efficiency.”
4. The optimization objective function checks constraints (maybe further uncertain analysis is necessary), Eq. 70. For the application of variant I, it is necessary to reduce the uncertain response process to a representative uncertain value.
5. The postprocessing has the main focus of this contribution; it computes the robustness of the current design (see section “Robustness Measure”), taking the uncertainty of the current design and the uncertainty of the response variables into account. It provides a deterministic value, representing the quality of the current design. For other objectives of optimization, only the application of information reducing measures is necessary.



Robust Design Optimization for Earthquake Loads, Fig. 11 Algorithm for the design of processes under consideration of fp-r variables

- 6. The deterministic optimization algorithm updates the deterministic design variables.
- 7. The result is a deterministic design yielding to a robust structure, under consideration of dynamical structural behavior and uncertainty of data and information.

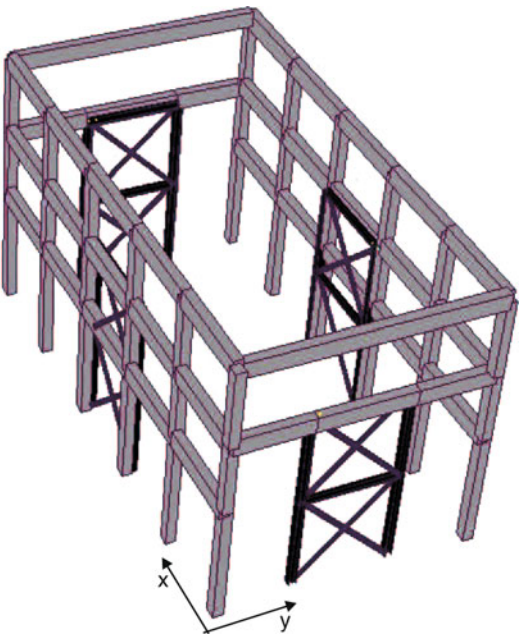
This approach is the basis for reliability-based optimization (RBO) methods (see, e.g., Enevoldsen and Sørensen 1994).

Example

Model Description

In this example, an existing RC structure, Fig. 12, should be strengthened for earthquake loading (Sickert et al. 2011). Previous studies (e.g., Steinigen et al. 2012) found that the RC frame has two independent mode shapes for longitudinal and lateral directions. The point of view in this example is the behavior in lateral direction. The RC frame should be coupled with a steel frame, and the objective of the optimization is to find optimal parameters of the connecting device, yielding a robust design, under consideration of an uncertain action process, with respect to the introduced variant II.

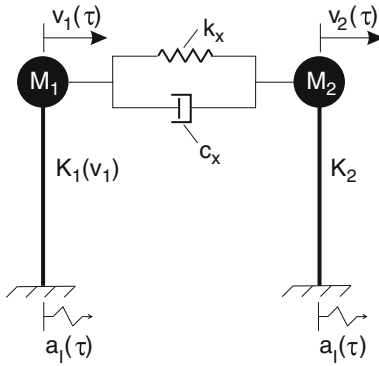
In Table 1 the uncertain input parameters of the structure are given. Due to the missing of any information about loading and material, the input parameters are modeled as triangular-shaped fuzzy variables (Eq. 8). The lateral structure is modeled as nonlinear 2DOF system; see Fig. 13. With the help of a fuzzy pushover analysis,



Robust Design Optimization for Earthquake Loads, Fig. 12 RC frame structure

Robust Design Optimization for Earthquake Loads, Table 1 Input data of the uncertain process for the 2DOF system

Fuzzy parameter	Value
$\tilde{\gamma}$ [mm/kN]	$\langle -3.3, 0.0, 6.0 \rangle \cdot 10^{-3}$
$\tilde{d}(F)$ [mm]	$d(F) + \tilde{\gamma} \cdot F$
$\tilde{K}_2$ [kN/m]	$\langle 50, 52.5, 55 \rangle \cdot 10^3$
$\tilde{M}_2$ [t]	$\langle 1.2, 1.5, 1.8 \rangle$
$\tilde{a}_I(\tau)$ [m/s ²]	$\frac{a_I(\tau)}{2.5} \cdot \langle 2, 2.25, 2.5 \rangle$
$\tilde{c}_{x,p}$ [kNs/m]	$\langle 0.9, 1.0, 1.1 \rangle$
$\tilde{k}_{x,p}$ [kN/m]	$\langle 0.0, 100.0, 150.0 \rangle$



Robust Design Optimization for Earthquake Loads,
Fig. 13 2DOF simulation model of the strengthened RC structure in lateral direction

Fig. 14, the uncertain nonlinear behavior of the RC frame could be included in the simulation model. This nonlinear force-displacement kernel curve is scaled by the parameter $\tilde{\gamma}$. The stiffness and mass of the steel frame are modeled by $\tilde{K}_2$ and $\tilde{M}_2$, respectively.

The load cases of the structure are ground motions; three acceleration-time dependencies of measured earthquakes are regarded: the acceleration-time dependencies $a_I(\tau)$ measured during the earthquakes in El Centro 1940, Calitri 1980, and Taiwan 1999, for instance (see Fig. 15). The acceleration-time dependencies are scaled to a peak ground acceleration (PGA) value of 0.25 g, applied to the 2DOF system. For a realistic modeling of earthquake loads, it is necessary to consider the uncertainty of the measured accelerations; therefore, the scaling with a fuzzy value yields $\tilde{a}_I(t)$.

Uncertain Analysis

Due to the application of the variant II to analyze uncertain processes, a realization of the uncertain parameters is determined in the first sequence. The solution of the dynamic structural analysis, in this case the solution of the equation of motion, Eq. 40, is found in the inner sequence. The time-displacement dependencies are computed for the 2DOF system. Realizations of the time-displacement functions during the Taiwan earthquake of both parts of the systems can be seen in

Fig. 16. The maximum displacements of the RC structure, during each earthquake acceleration, $\underline{v}_{MD,i} = \max v_1(\tau) \mid i \in \{1, 2, 3\}$, is the result of the postprocessing procedure. The result of the uncertain analysis is the fuzzy maximum displacements $\tilde{v}_{MD,i} \mid i \in \{1, 2, 3\}$.

Robust Design of Connection Device

Design Objectives

The design parameters are the viscosity c_x and stiffness k_x of the damping device:

$$\tilde{c}_x = c_{x,d} \cdot \tilde{c}_{x,p} \quad \text{and} \quad (72)$$

$$\tilde{k}_x = k_{x,d} \cdot \tilde{k}_{x,p}. \quad (73)$$

The split of uncertain a priori and deterministic design part is applied. The uncertain a priori parameters $\tilde{c}_{x,p}$ and $\tilde{k}_{x,p}$ are shown in Table 1. The design ranges of the parameters are

$$c_{x,d} \in [0, 10000] \text{ kN/m and} \quad (74)$$

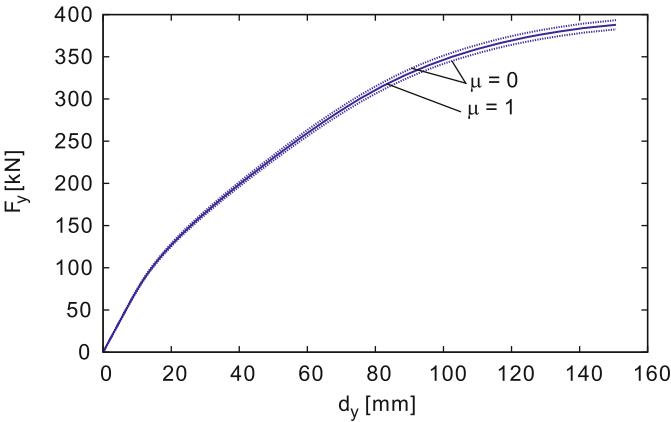
$$k_{x,d} \in [0, 10000] \text{ kNs/m.} \quad (75)$$

The general optimization task is to “compute optimal connection device parameters for a robust structure.” Therefore, the robustness measure, Eq. 48, is evaluated for the three load cases. The uncertainty is assessed by applying the information reducing measure entropy for fuzzy quantities (Eq. 61). The weighting factors and penalty functions of the robustness measure are here not considered. For the design d , it yields

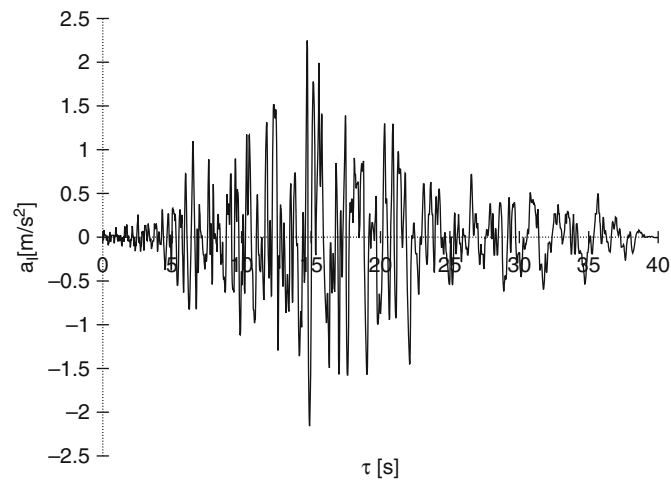
$$R_d = \frac{\sum_{i=1}^{l=3} \left(\mathcal{U}_E^f(\tilde{c}_{x,d,i}) + \mathcal{U}_E^f(\tilde{k}_{x,d,i}) \right)}{\sum_{i=1}^3 \mathcal{U}_E^f(\tilde{v}_{MD,d,i})} \quad (76)$$

$$= \frac{3 \cdot (0.1453346 \cdot c_{x,d} + 109.001)}{\sum_{i=1}^{l=3} \mathcal{U}_E^f(\tilde{v}_{MD,d,i})}. \quad (77)$$

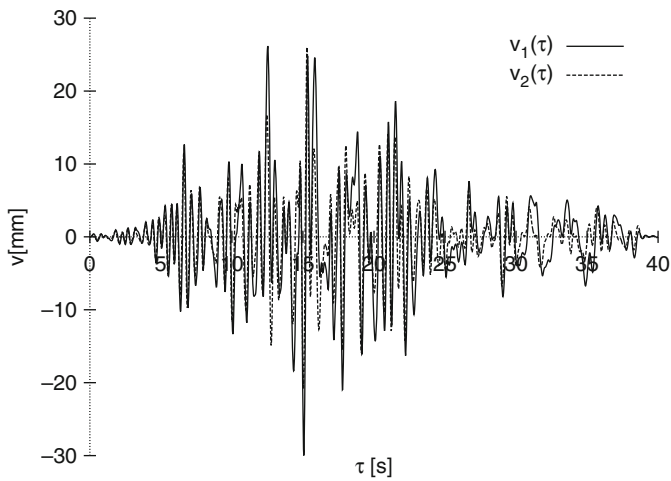
**Robust Design
Optimization for
Earthquake Loads,**
Fig. 14 Fuzzy force-
displacement dependency
of the RC frame structure



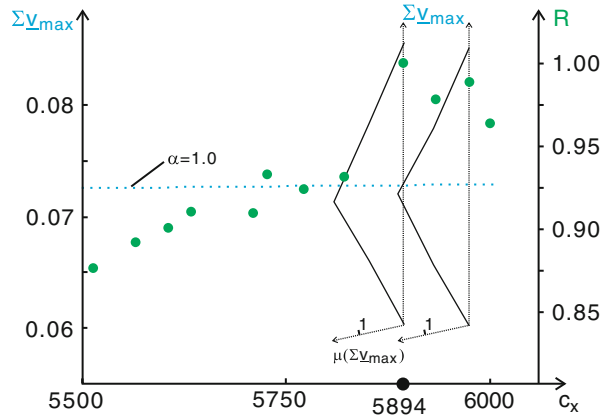
**Robust Design
Optimization for
Earthquake Loads,**
Fig. 15 Considered
earthquake acceleration,
Taiwan 1999



**Robust Design
Optimization for
Earthquake Loads,**
Fig. 16 Time-
displacement dependency
for 2DOF system for
Taiwan earthquake 1999



Robust Design Optimization for Earthquake Loads, Fig. 17 Result of the uncertain design of processes



Results

In Fig. 17 the result of the robust design analysis is shown. In the shown part of the design space $c_{x,d} \in [5,500, 6,000]$ kN/m and $k_{x,d} = 0.0$ kNs/m, the summed result of the earthquake analysis $\sum v_{MD}$ is set on the left axis and the related robustness R_d is set on the right axis. The blue dotted line shows the summed displacement at $\alpha = 1$. The fuzzy result $\sum v_{MD}$ is shown for the two examples. The robustness for the observed designs is shown with green dots (scaled to $\max(R) = 1.0$). The optimal robust design is found for the parameters $c_{x,d} = 5,894$ kN/m and $k_{x,d} = 0.0$ kNs/m taking uncertainty into account. This means,

$$\begin{aligned} \tilde{c}_x &= \langle 5304.6, 5894.0, 6483.4 \rangle \text{ kN/m and} \\ \hat{k}_x &= \langle 0.0, 100.0, 150.0 \rangle \text{ kN/m} \end{aligned} \quad (78)$$

were considered.

Summary

In this contribution, parameterized robust design of structures under earthquake loads is shown. The necessary capturing of uncertainty is done by using fuzzy, random, and fuzzy probability-based random variables. The numerical algorithms are shown in detail, with the focus of

reusing existing uncertainty analysis methods and available optimization approaches. Robustness of a structure is defined as the ratio of the amount of uncertain input quantities to the amount of uncertain output quantities. The example shows the applicability of robust design optimization for earthquake loads.

Cross-References

- [Analytic Fragility and Limit States \[P\(EDPI IM\)\]: Nonlinear Dynamic Procedures](#)
- [Bayesian Statistics: Applications to Earthquake Engineering](#)
- [Nonlinear Dynamic Seismic Analysis](#)
- [Stochastic Analysis of Nonlinear Systems](#)
- [Structural Optimization Under Random Dynamic Seismic Excitation](#)
- [Structural Reliability Estimation for Seismic Loading](#)
- [Uncertainty Theories: Overview](#)

References

- Beer M, Liebscher M (2008) Designing robust structures a nonlinear simulation based approach. *Comput Struct* 86:1102–1122
- Beer M, Ferson S, Kreinovich V (2013) Imprecise probabilities in engineering analyses. *Mech Syst Signal Process* 37:4–29
- Enevoldsen I, Sørensen J (1994) Reliability-based optimization in structural engineering. *Struct Saf* 15(3): 169–196

- Götz M, Graf W, Serafinska A, Kaliske M (2013) Surrogate models in early design stages. In: Deodatis G, Ellingwood B, Frangopol D (eds) *Safety, reliability, risk and life-cycle performance of structures and infrastructures* (11th ICOSSAR). Columbia University/CRC Press, New York/Boca Raton, p 383 (book of abstracts, full paper in print)
- Graf W, Sickert JU, Pannier S, Kaliske M (2010) Robust design with uncertain data and response surface approximation. In: Beer M, Muhanna RL, Mullen RL (eds) *4th international workshop on reliable engineering computing (REC 2010)*, vol 4. National University of Singapore. Research Publishing Services, pp 554–573
- Möller B, Beer M (2004) *Fuzzy randomness – uncertainty in civil engineering and computational mechanics*. Springer, Berlin
- Möller B, Graf W, Beer M (2000) Fuzzy structural analysis using α -level optimization. *Comput Mech* 26:547–565
- Möller B, Graf W, Sickert J-U, Steinigen F (2009) Fuzzy random processes and their application to dynamic analysis of structures. *Math Comput Model Dyn Syst* 15:515–534
- Pannier S (2011) *Effizienter numerischer Entwurf von Strukturen und Prozessen bei Unschärfe*. Dissertation, TU Dresden, Veröff. Institut für Statik und Dynamik der Tragwerke, H. 23
- Pannier S, Waurick M, Graf W, Kaliske M (2013) Solutions to problems with imprecise data – an engineering perspective to generalized uncertainty models. *Mech Syst Signal Process* 37(Special issue: imprecise probabilities):105–120
- Rommelfanger H (1988) *Fuzzy Decision Support Systeme – Entscheiden bei Unschärfe*. Springer, Heidelberg
- Saltelli A, Chan K, Scott EM (2008) *Sensitivity analysis*. Wiley, New York
- Sickert JU, Mandara A, Kaliske M, Graf W (2010) Robust design of seismic up-grading of r.c. structures with innovative bracing systems. In: Mazzolani FM (ed) *Urban habitat constructions under catastrophic events – proceedings of the COST C26 action final conference*. CRC Press, Leiden, pp 595–600
- Sickert JU, Pannier S, Graf W, Kaliske M (2011) Robustness based design of structures under consideration of generalized uncertainty models. In: Nishijima K - (ed) *International conference on applications of statistics and probability in civil engineering (ICASP)*, Zurich. Taylor & Francis Group, London, pp 533–540
- Søndergaard J (2003) *Optimization using surrogate models – by the space mapping technique*. PhD thesis, Informatics and Mathematical Modelling, Technical University of Denmark, Lyngby
- Steinigen F, Sickert J-U, Graf W, Kaliske M (2012) Chapter 7: Fuzzy and fuzzy stochastic methods for the numerical analysis of reinforced concrete structures under dynamical loading. In: Papadrakakis M, Stefanou G, Papadopoulos V (eds) *Computational methods in stochastic dynamics*, vol 2. Springer, Dordrecht, pp 113–130
- Tintner G (1960) A note on stochastic linear programming. *Econometrica* 28(2):490–495
- Yager RR (1986) A characterization of the extension principle. *Fuzzy Set Syst* 18(3):205–217
- Zadeh L (1975) The concept of a linguistic variable and its application to approximate reasoning. *Inform Sci* 8(3):199–249

Rockfall Seismicity Accompanying Dome-Building Eruptions

Nick Varley

Facultad de Ciencias, Centre of Exchange and Research in Volcanology, Universidad de Colima, Colima, Mexico

Synonyms

Dome collapse; Earthquakes; Effusive eruption; Magma emplacement; Magma extrusion

Introduction

Dome-building eruptions are very common at volcanoes and can present a considerable threat to the nearby population. In this entry they are discussed with reference to the seismicity that is generated by the dome growth process, which often provides an insight into the associated processes. Small earthquakes frequently occur during dome emplacement, and modelling their generation has helped in the understanding of the processes occurring during magma extrusion (e.g., Iverson et al. 2006; Varley et al. 2010). The second type of seismogenic phenomena are dome collapses, which result from the emplacement of magma at the surface and the inherent instability of the resulting dome structure. In this case studying the seismic signals can provide estimates of the characteristics of the resulting rockfall or pyroclastic flow.

Analysis of the seismicity that accompanies dome growth is improving our understanding of

their construction and can aid in the mitigation of associated risks. This entry discusses the formation of lava domes and the associated risk to populations close to the volcano. The seismicity generated by magma ascent is discussed: how rockfalls are generated from domes and what are the characteristics of the resulting seismic signals. A comparison is made with the characteristics of the seismicity produced by other mass movements down slopes, with the utility of these signals highlighted.

Volcanic Domes

Magma erupts in a variety of fashions, mainly determined by its physical properties, the amount of gas contained, and its ability or inability to escape the magma during decompression. Explosive activity varies from relatively weak eruptions, called Strombolian, that form scoria cones, the most common continental volcanic feature, to large dangerous Plinian eruptions, characteristic of tall stratovolcanoes associated with subduction of plates along many boundaries across the globe. Usually the most significant threat comes from the possibility of a large explosive eruption, which may produce pyroclastic flows that can reach many tens of km from the volcano and destroy everything in their path. However, the formation of a dome also has the potential for large-scale destruction.

Viscous lava will accumulate upon extrusion in an eruption forming a lava dome rather than flowing away as is the case for more fluid low-silica magma such as basalt. Stratovolcanoes are usually constructed from intermediate magma, and by nature, the magma erupts either with high volatile contents and fast ascent producing explosive eruptions or more slowly and with low gas contents resulting in dome construction. Domes can be isolated structures, e.g., San Pedro, Nayarit, Mexico (Fig. 1), or in groups such as Lassen Peak in California, or as part of chains such as the Mono Lake-Inyo Craters in California, USA, where four of 13 domes have erupted in the last 600 years (Hildreth 2004). As in the case of Lassen Peak, domes can reach several

hundreds of meters in height. However, the majority of domes represent a distinct episode in the complex story of magma ascent and emplacement and explosions with pyroclastic flows and tephra fall that together construct a stratovolcano.

Domes grow to fill or partially fill summit craters or lower on the flanks for eruptions associated with lateral dykes. Examples of recent summit dome-forming eruptions are those at Volcán de Colima, Mexico (Lavallée et al. 2012; Fig. 2), or Redoubt, USA (Coombs et al. 2013). However, the eruptive history of the volcano might result in growth at a different location on the volcano's flanks, for example, the long-lived dome-building eruption at Santa Maria, Guatemala. The dome complex Santiaguito started growing in 1922 and, after several shifts in the focus of extrusion, continues to produce magma today within the scar left by the 1902 eruption, located below the summit (Rose 1972; Fig. 3). This dome complex is an example of where cyclicity can be observed with variation in the effusion rate over distinct periods of time (Barmin et al. 2002). The period of such cycles can vary from years to just days, such as observed at Soufrière Hills Volcano, Montserrat (Loughlin et al. 2010; Fig. 4). Any foresight regarding the effusion rate for a growing dome would be invaluable for the mitigation of the risks associated with collapse events.

Risks Associated with Dome Growth

Volcanic domes can present a major risk due to the likelihood of collapse. Some of the principal recent volcanic disasters have been associated with this phenomenon, the most notable being the eruption of Mt. Pelee, Martinique, in 1902, during which the city of was wiped out by a pyroclastic flow, killing around 28,000 people. These flows develop when the dome is pressurized; a partial collapse will produce decompression which initiates fragmentation of the magma. The resulting mixture of ash and hot gas, accompanied by large blocks, will descend the volcanic flank at high speed due to gravity.

**Rockfall Seismicity
Accompanying Dome-**

Building Eruptions,
Fig. 1 San Pedro dome,
Nayarit, Mexico. This
dome was emplaced within
a caldera



**Rockfall Seismicity
Accompanying Dome-**

Building Eruptions,
Fig. 2 Volcán de Colima
summit dome, Mexico
(Photo taken on
3 November 2010 during
the 2007–2011 eruption.
The active lobe on the
right-hand side of the dome
can be observed, which was
the source of many
rockfalls each day down the
western flank (Fig. 7))



R

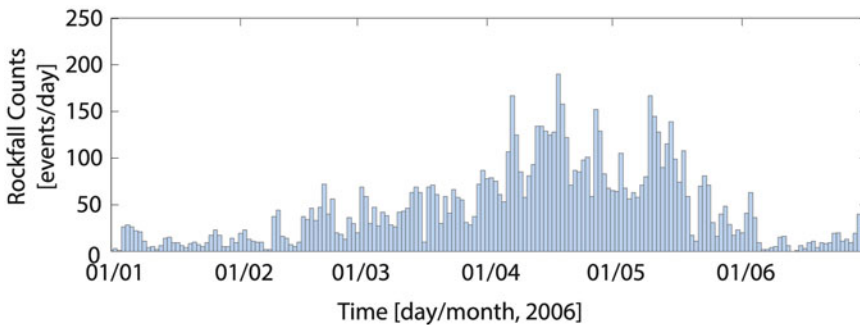
The gas gives the flow mobility, and their runout distance can be many km, much greater than a rockfall event. Pressurization within the dome is determined by the exsolution of volatiles, which in turn is partly controlled by the extrusion rate. At Volcán de Colima, pyroclastic flows were rare and only very small during the very slow dome growth that occurred between 2007 and 2011, whereas the faster effusion rate which was witnessed in 2004 resulted in frequent

pyroclastic flows, the largest reaching 6.1 km from the volcano. As a comparison, rockfalls at this volcano do not reach more than 2 km of distance.

Volcanic domes grow at different rates from extremely slow like the case of Volcán de Colima from 2007 until 2011, which grew at a rate of $0.02 \text{ m}^3 \text{ s}^{-1}$ (Varley et al. 2010) to a fast rate such as the case of Chaitén, Chile, which initially grew at $66 \text{ m}^3 \text{ s}^{-1}$ (Pallister et al. 2013).

Rockfall Seismicity Accompanying Dome-Building Eruptions,

Fig. 3 Santiaguito dome, Guatemala (Photo taken 13 January 2007 showing the erupting Caliente dome to the left with the ash cloud from an explosion)



Rockfall Seismicity Accompanying Dome-Building Eruptions, Fig. 4 Soufrière Hills Volcano, Montserrat. Seismic counts from rockfalls and long-period events (due

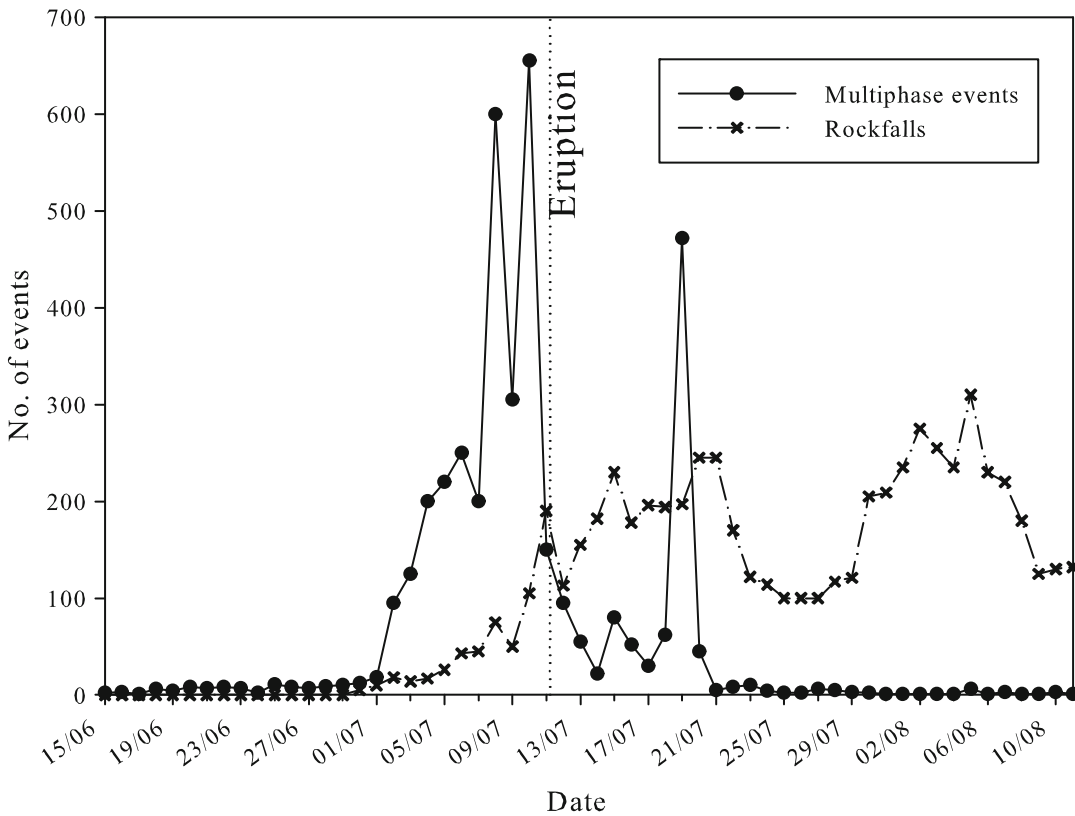
to both rockfalls and internal earthquakes); 1 January to 1 July 2006 (Figure provided by Silvio De Angelis)

A faster growth rate implies greater pressurization within the dome, which in turn increases the potential for far-reaching pyroclastic flow and hence far greater risks. Once a dome growing within the summit crater of a stratovolcano has reached the rim, it will begin to overflow in the form of rockfalls and possibly form a lava flow.

Merapi in Indonesia deserves a special mention with its long history of frequent pyroclastic flow generation associated with the growth of a dome on the upper flanks. In 1994 a flow killed over 60 people, and it was thought to have been generated by pure gravitational collapse of the

growing dome, whereas events of 1997 and 1998 were more fluid due to gas pressurization with the lava dome (Voight et al. 2000). The 1994 collapse was estimated at $2.5 \times 10^6 \text{ m}^3$ with the various pyroclastic flows occurring over a period of about 7 h, reaching 6.5 km from the volcano. The event did not coincide with periods of high extrusion rates which had declined over 3 months prior to the collapse (Voight et al. 2000).

A seismic event classification scheme specific to Merapi included “multiphase” events which were detected during rapid dome growth. They are somewhat similar to hybrid events in other



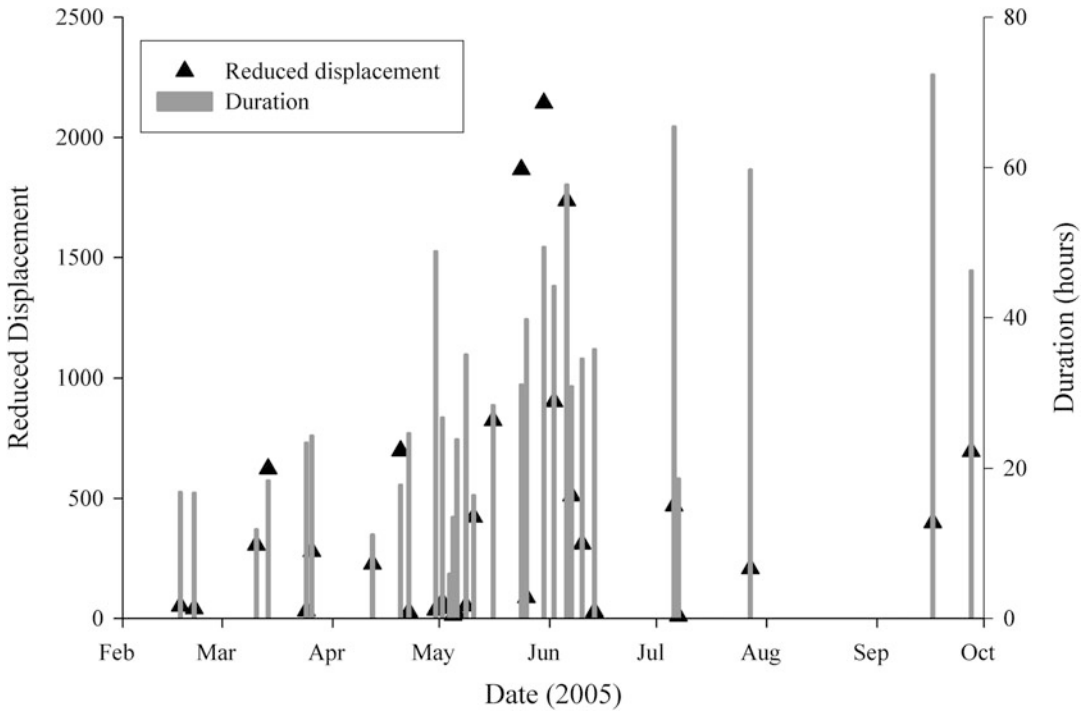
Rockfall Seismicity Accompanying Dome-Building Eruptions, Fig. 5 Seismicity recorded at Merapi, Indonesia, on 15 June–12 August 1998 from rockfall events and multiphase earthquakes, which are associated with

dome growth at Merapi. The indicated eruption on 11 July was a dome collapse (Adapted from Voight et al. 2000)

classification schemes for volcano seismicity. These events occur at a rate of up to several 100 per day, are shallow, have an emergent waveform, and are characterized by frequencies of 3–4 Hz; however, their generation mechanism remains poorly understood. Seismicity recorded from a broadband seismometer was used to analyse dome collapse events during the 1994 eruption, with the assumption that the volume was proportional to the seismic magnitude for a similar source area and descent path (Brodscholl et al. 2000). Figure 5 shows the increase in multiphase-type seismicity associated with a dome collapse in 1998, with 33 events occurring per week in mid-June to 2029 in the 6 days ending 12 July (Voight et al. 2000). The increase in seismicity was accompanied by an increase in

the number of rockfalls (as well as tilt); the correlation indicating that the parameters were related to increasing effusion rates. These pyroclastic flows did not result in casualties since clear precursors were detected allowing evacuations to take place, even though the volume of material was larger than in 1994 at $8.8 \times 10^6 \text{ m}^3$.

The eruption at Soufrière Hills Volcano, Montserrat, has produced many dome collapses which have dominated the hazards presented to the local population. Decompression from the large December 1997 dome collapse produced a large explosion with highly mobile pyroclastic density currents (Woods et al. 2002). These flows have been at times complex with multiple pulses (Loughlin et al. 2002). The role of gas-rich magma batches has been considered as a



Rockfall Seismicity Accompanying Dome-Building Eruptions, Fig. 6 Swarms of LP events at Volcán de Colima related to magma ascent and dome formation in 2005. A poor correlation between the duration of the

seismic swarm and the magnitude of the seismicity associated with the explosion that subsequently occurred destroying any domes forming within the crater

contributing factor in the destabilizing process prior to major collapse events such as in 2006 (De Angelis et al. 2007).

Magma Ascent and Emplacement

Ignoring external factors, such as the interaction with water, it is the magma ascent rate, combined with its viscosity and volatile contents, which determines whether an eruption is explosive or effusive. The more efficiently the magma can release its volatile cargo, the more likely the eruption will result in lava domes or flows, rather than undergo fragmentation with the rapid expansion of gases and the emission of ash in an eruption column.

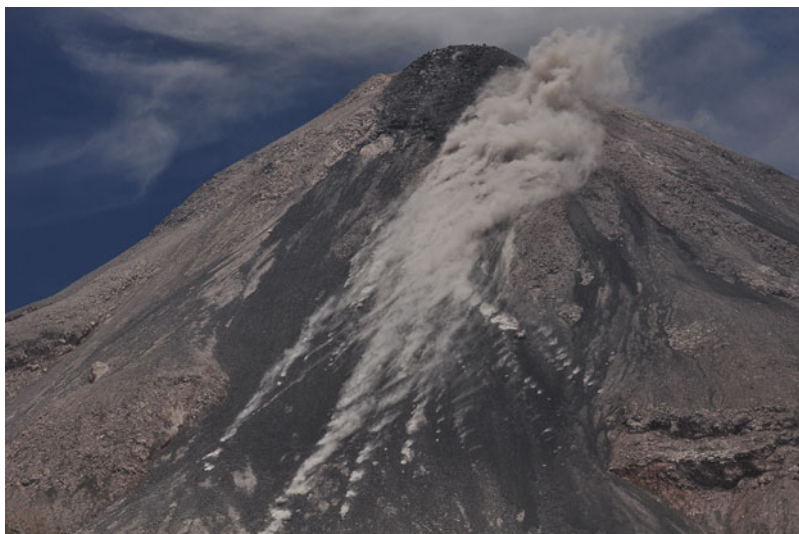
A particular type of seismicity is often recorded during slow magma ascent. As already discussed, in the case of Merapi, so-called multiphase events can be used to qualitatively

estimate the magma ascent rate (Voight et al. 2000). Many other volcanoes have produced low-frequency seismicity during effusive eruptions, often in form of swarms, where the events are concentrated in time and/or multiplets, when they are also concentrated in space.

Modelling magma ascent in the conduit has highlighted the most important factors, with interactions along the conduit wall dominating, whether it is gas loss through the walls themselves or the transformation from ductile to brittle behaviour due to increased shear; they are critical in determining the final eruptive style. At Volcán de Colima, small-magnitude long-period (LP) events were recorded in swarms lasting from 6 hours to 3 days, prior to moderately large Vulcanian explosions that produced pyroclastic flows reaching up to 5.4 km from the volcano (Varley et al. 2010). As shown in Fig. 6 there was little relation between the duration of the swarm and the size of the associated

Rockfall Seismicity Accompanying Dome-Building Eruptions,

Fig. 7 Small rockfall descending the western flank of Volcán de Colima during period of dome growth (30 April 2013). A cloud of dust has been generated with the impacts of individual rocks visible on the slope



explosion as recorded by the seismicity resulting from the event. The majority of swarms continued for a considerable time after the explosion, with an increase in the magnitude of the LP events. It was suggested that the magnitude was related to the strain rate and thus the ascent velocity of the magma in the conduit, with brittle deformation generating the events (Goto 1999; Neuberg et al. 2006; Varley et al. 2010). There was clear evidence for magma ascent since small domes were also observed at Volcán de Colima emplaced during the height of the explosive activity.

Generation of Rockfalls

Rockfalls can be described as mass movements of material down a slope by gravity, whereby the detachment occurred along discontinuities such as fractures, joints, or bedding planes and the motion is one of free fall, bouncing, and rolling. Bourrier et al. (2013) recognize that a more precise definition is required and propose two end-members being fragmental rockfall (where there is a greater interaction with the slope than between individual particles) and rock mass fall (particles interact with each other and travel as a deforming mass). Clearly for dome-forming volcanic eruptions, the fragmental rockfall would be

more usual. The magnitude of a rockfall can vary from very small, perhaps just a few rocks, to an event producing a column large enough to be mistaken for one generated by an explosion (as was the case at Mt. St. Helens; Moran et al. 2008). Although typically the term “rockfall” is used to describe all collapse events associated with growing lava domes, which lack a dominant gas phase to generate a pyroclastic density current, the term “rockslide” might be applicable in some cases, where initial movement is one of translation rather than free fall.

Rockfalls associated with volcanic activity are most often associated with growing domes or the front of lava flows, the former being of a larger magnitude. Figure 7 shows a relatively small rockfall from the slowly growing dome at Volcán de Colima. The effusion of silicic lava produces blocks which accumulate with low stability. The morphology of these structures is dependent on various factors including: composition, lava temperature, volatile contents, crystal contents, and effusion rate. A steeper front will produce more frequent and greater-magnitude rockfalls. When there is gas pressurization within the dome, a rockfall can easily transform into a pyroclastic density current which has greater mobility due to the high proportion of hot gases in its composition. Small events might have a volume of $<10 \text{ m}^3$, whereas a large dome collapse, such as

the events at Soufrière Hills Volcano, can be up to $>5 \times 10^7 \text{ m}^3$ (Woods et al. 2002).

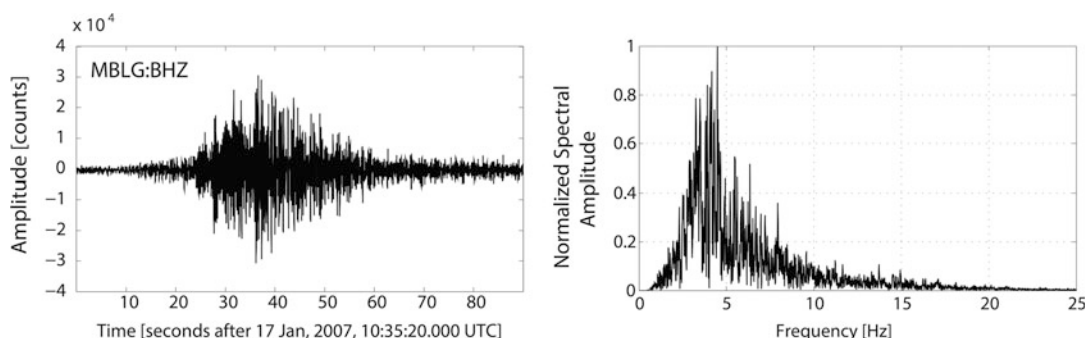
The relationship between effusion rate and rockfall production is not necessarily simple (Loughlin et al. 2010). Physical factors will influence the relationship, e.g., the presence of multiple sources for effusion (e.g., Lavallée et al. 2012) or variations in the morphology as a function of time. Determining what controls the effusion rate at a volcano involves the consideration of a complexity of variables. Modelling has produced some insight into the processes, for example, taking a simple geometric model and simulating an increase in pressure within the magma chamber have been shown to influence the effusion rate (de' Michieli Vitturi et al. 2010). Increases in effusion rate can have a critical influence on the risks as was witnessed at Soufrière Hills, where an increase has resulted in large increases in eruptive activity, either as Vulcanian explosions or potentially dangerous pyroclastic flows descending the flanks.

Estimating the effusion rate of a volcano can prove to be a challenge. If the magma extrusion results in purely dome growth, an estimation of the volume variation over time is sufficient. This can be carried out using photogrammetry, with recent software advances allowing the use of standard cameras to produce digital elevation models to an accuracy of around 5 % (Diefenbach et al. 2012; James and Varley 2012). Maintaining an acceptable accuracy becomes difficult once the dome extends to lava flows or when the loss of material through rockfalls becomes significant. Although estimating the area of a lava flow presents no problems, the height usually remains largely unknown, and estimates typically assume a fairly arbitrary and constant thickness, whereas in reality this will vary as a function of the gradient. It is difficult to calculate the volume lost through rockfalls, whereby estimates can only be made in the case of particularly favourable conditions (Mueller et al. 2013).

Rockfalls can also represent precursors for a larger-scale slope failure such as debris flow or debris avalanche. Where there are steep slopes, resulting in free fall of detached material, a linear

relationship is observed between magnitude and frequency in log-log space (Dussauge-Peisser et al. 2002; Rosser et al. 2007), similar to the Gutenberg-Richter relationship for seismicity. This power law for the volume distribution of rockfalls allows the use of statistics to predict the occurrence rates of events of different volumes based on historic records. For the long-term hazards in mountainous regions, this has important implications; however, for dome-derived rockfalls, episodes are usually relatively short in duration lasting just months or in a few cases a few years, making it difficult to compile a reliable volume distribution during the eruption. With frequently active volcanoes, however, if the dome growth has similar characteristics to previously documented episodes, it remains possible to assume repetition of a distribution and hence produce a probabilistic assessment of the hazards. Dussauge-Peisser et al. (2002) published a summary of the *b* values obtained in theirs and other studies when correlating rockfall volume with occurrence (nonvolcanic cases). The volumes range from $0.01\text{--}10 \text{ m}^3$ to $10^7\text{--}10^{10} \text{ m}^3$ and the study periods from 2 months to 10,000 years, with the *b* exponent ranging from 0.40 to 1, but with the majority of values in the range 0.4–0.5. For volcanoes similar time-frequency variations have been observed for pyroclastic flows from Soufrière Hills Volcano, Montserrat (Jolly et al. 2002).

The strain and seismic energy release associated with slope failure should accelerate as a power law (Voight 1988). By examining the increase in seismic energy release over time, it is possible to predict the failure associated with explosive eruptions (Arámbula-Mendoza et al. 2011; Budi-Santoso et al. 2013). For slope failure there is less evidence; however, for Soufrière Hills Volcano, Hammer and Neuberg (2009) showed that swarms of LP events detected prior to dome collapses showed an increase in magnitude that followed the material failure law, thus verifying this method as a potential tool to predict these major collapse events. However, Bell et al. (2011) argue that the failure forecast method is unreliable and that the non-Gaussian distribution of errors needs to be



Rockfall Seismicity Accompanying Dome-Building Eruptions, Fig. 8 Typical seismic waveform from a rockfall with spectrum from Soufrière Hills Volcano, Montserrat in 2007 (figure prepared by Silvio De Angelis)

considered. They suggest the use of a generalized linear model, which is a generalization of least-squares linear regression.

Generation and Characteristics of Seismicity

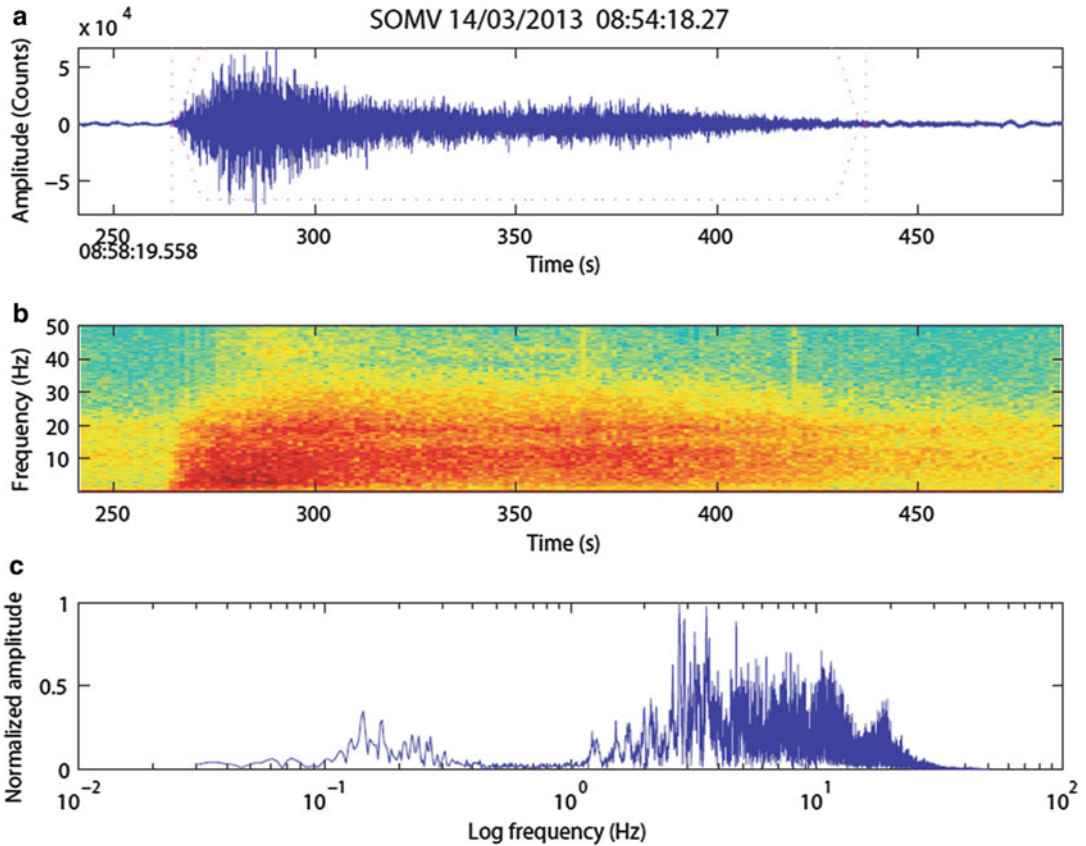
Due to the vastly superior body of work, initially seismicity from nonvolcanic mass movements is considered here. The differences between these distinct types of events and those derived from volcanic domes are minimal. During the transport of material down steep slopes, part of the energy transferred becomes seismic energy producing characteristic waves. The proportion of potential energy transformed into seismic energy is small, with estimates varying from 0.25 (Vilajosana et al. 2008) to between 10^{-6} and 10^{-3} (Deparis et al. 2008; Hibert et al. 2011).

A typical seismic signal generated from a mass movement event has an elongated symmetric gradually increasing then decreasing waveform; the initial steady acceleration in the liberation of energy results in emergent seismic signals (e.g., Pankow et al. 2014). The position of the peak within the waveform should be related to the time when the largest volume is in motion on the slope, producing the largest number of impacts with the ground (Dammeier et al. 2011). There might be a wide variety of block size, shape, and initial location on the slope, producing diverse durations of individual

transport times and extending the overall event duration. Whereas the generation of tectonic seismic events is modelled by double-couple or moment tensors, seismicity generated by landslides or rockfalls has been modelled as single force (Fukao 1995).

For rockfalls the spectral content is wide (1–20 Hz) but will depend upon the source-station distance, with a larger distance implying larger low-frequency content due to the attenuation being frequency dependent (Suriñach et al. 2005). Path and source effects will influence the signal recorded, particularly since the source is moving at a high velocity, so its direction in relationship to the seismometer location will have a large effect. The majority of energy tends to be concentrated below ~ 3 –4 Hz and the higher frequencies might be generated by the collision between blocks (Dammeier et al. 2011). Figures 8 and 9 show typical rockfall waveform and spectrum from Soufrière Hills Volcano in 2007 and Volcán de Colima in 2013.

Examination of the seismicity generated by landslides has determined differing characteristics for different wavelengths, with low-frequency components being detected prior to high frequency with their peak also occurring earlier (Yamada et al. 2012). The topography can significantly influence the seismic signal, as can the rheology of the flow (Favreau et al. 2010). Due to the emergent waveform, the uncertainty of the first arrival makes it more difficult to estimate the location of the event or calculate the



Rockfall Seismicity Accompanying Dome-Building Eruptions, Fig. 9 Typical seismic waveform from a rockfall with spectrum and spectrogram from Volcán de

Colima during small dome collapses in 2013 (figure prepared by Raúl Arámbula)

magnitude of the seismicity using standard seismic methods. The local M_L and coda or duration magnitude m_d can be calculated using an assumed location. Large values for the difference between these calculated magnitudes $M_L - m_d$ are a characteristic of seismicity from surface events. However, successful source location has been achieved, for example, in the case of Piton de la Fournaise volcano (Hibert et al. 2011). It is possible to record the signal from large dome collapses at quite a distance. Long-period seismic signals were generated by the 1997 collapse at Soufrière Hills Volcano and recorded at stations 450 and 1,261 km away. Modelling permitted the description of the avalanche event as two horizontal and one vertical force, the transition between each force occurring over a period of 70 s (Zhao et al. 2013).

Utility of Rockfall Seismicity

As with any application of seismicity, the availability of information is a function of the network configuration and response of each seismometer. With a large network including closely located stations, analyses of the seismicity of mass movements can provide a wealth of information, including the source of the rockfall. For volcanoes, this can be important since the risk presented by an event descending one flank is likely to be distinct to another. In an artificially generated rockfall, it was shown that a suitable configured network of broadband stations could be used to take advantage of the linear polarization of the P-waves of a rockfall to obtain a location and size estimate for rockfalls (Vilajosana et al. 2008). Recently an automated

system has been developed at Piton de la Fournaise Volcano, Reunión, to not only locate events but to estimate their volume (Hibert et al. 2014). Using the technique of fuzzy sets, a success rate of 92 % has been achieved.

The application of rockfall seismicity to mitigate the associated risk can be divided into several areas:

Monitoring of Slope Conditions

The application of seismic networks for monitoring slope conditions at open-cast mines has generated abundant data in the event of slope failure. An enormous $65 \times 10^6 \text{ m}^3$ landslide occurred at the Bingham Canyon copper mine in the USA in 2013 (Pankow et al. 2014). In this case, seismometers that detected the event were located between 6 and 400 km. The landslide, classified as a rock avalanche, was actually two closely spaced events. The seismicity was dominated by long-period events with differences between the two episodes: only the first one produced a high-amplitude peak near the end of the coda and there was a large difference when the maximum amplitude occurred.

Retrospective Analysis of Mass Movement Events

The simplest application of seismicity is as a means of estimating the duration and velocity of avalanches. Incorporating spectral analysis extends the possibility, with long-period seismicity in particular having provided important insights into the physical characteristics of flows, e.g., erosive processes of the 40–60 Mm^3 Mount Steller, Alaska, rock-ice avalanche were qualitatively described (Moretti et al. 2012).

The volume of a moving mass can be estimated by quantifying the seismic energy. Yamada et al. (2012) formulated an empirical relationship which correlated approximately the square of the volume with the energy parameter. It is important to note that in general a specific relationship derived for one location will be explicit for that location, though attempts have been made to formulate more widely applicable correlations between different characteristics. Just taking five basic parameters of the seismicity

(duration, peak magnitude, velocity envelope area, rise time, and average ground velocity) Dammeier et al. (2011) obtained relationships with five fundamental rockslide event parameters: volume runout distance, drop height, and potential energy, for a series of different events. Particularly using the first three parameters, they suggested their results could be adapted to allow event characterization from seismic data. Obviously for risk mitigation, the volume and runout distance of a collapse event represent the most important parameters.

With more complex analysis, further dynamic information can be obtained about mass movement events. Suriñach et al. (2005) observed an increase in time of the higher frequencies of the seismicity associated with landslides. When the direction of the movement is toward the seismometer, a decrease in the attenuation of higher frequencies with increasing distance from the source would result in this pattern.

Seismicity associated with avalanches has been used to calibrate a model to describe its motion (Schneider et al. 2010). One of the events studied was on an Alaskan volcano, Iliamna, which typifies the more simple topography when comparing volcanoes with mountains. This results in a simpler seismic signal from any mass movement down its flanks. This type of model produces an estimate of the inundation zone of the avalanche. In the case of an erupting lava dome, the volatiles being released from the magma upon decompression will make a pyroclastic density current more likely than a less mobile avalanche.

Precursors of Eruptions

The quantification of rockfall frequency can make a useful contribution to volcano monitoring, with significant increases possibly representing precursors to enhanced activity. At Augustine, USA, rockfall-generated seismicity was recorded from 1997 to 2009 and revealed a typical duration of 30 s for rockfall events, with their frequency $>4 \text{ Hz}$ (DeRoin and McNutt 2012). The number of annual events increased dramatically from an average of 28 per year up to 340 in 2005, which was precursory to an

eruption which started in early 2006. Subsequently there were rockfalls during the growth of a lava dome and its frequent collapse. Instability of the dome meant the continuation of rockfalls after effusion had ceased. Also in Alaska at Iliamna, for some ice-rock avalanches seismic precursors were recorded, demonstrating the value of seismic monitoring of hazardous collapse events (Huggel et al. 2007).

Studies of Eruption Processes

Detailed examination of the characteristics of the seismicity associated with rockfalls has been used to understand the mechanism of the destabilization process and collapses (De Angelis et al. 2007; West 2013). The amalgamation of the fields of rock mechanics and volcano seismology is resulting in a far more detailed understanding of the generation process of rockfalls or pyroclastic flows. Studies of time series of rockfalls can provide important information regarding the evolution of the eruptive process. Their associated seismicity provides a convenient method to identify their occurrence in time. Calder et al. (2005) looked at the repose period between rockfalls generated by the growing of Soufrière Hills Volcano, Montserrat. They determined that a log-logistic probabilistic density function fitted the data, which reflects the failure mechanism. Within the seismic classification scheme at Soufrière Hills Volcano, two distinct rockfall classes are defined: normal rockfalls and long-period rockfall signals (Luckett et al. 2002). The latter are thought to be produced by the superposition of two phenomena, violent degassing on the dome surface triggering a nearby rockfall. For pyroclastic flows descending the flanks, the seismicity was used to estimate the location of the flow from its onset thus proving values for their extension and speed (Jolly et al. 2002).

At Augustine, Alaska, rockfalls were observed to follow a seasonal pattern, with rainfall having some influence (DeRoin and McNutt 2012), an important information for risk mitigation.

A special case of volcanic rockfall, not associated with dome growth, is the events arising from instabilities in the walls of pit craters which can occur at basaltic volcanoes. Very

long-period events have been identified as being generated from rockfalls within the pit crater at Kilauea (Chouet and Dawson 2013). The idea behind their generation mechanism is oscillation of the conduit and inversion of the waveforms provided the volume of the rockfalls ($200\text{--}4,500\text{ m}^3$) as well as the dimensions of the magma column, dyke system, magma flow rate, and its viscosity.

For volcanoes, a direct correlation between rockfall frequency and effusion rate has often been assumed but studies quantifying the two parameters are virtually nonexistent. To quantify the volume of rockfalls using seismicity, first the relationship between the two needs to be established. It was observed that the seismic magnitude correlated with the rockfall volume at Mt. St. Helens on the same slope (Norris 1994). Mueller et al. (2013) estimated the effusion rate of Volcán de Colima by observing directly the volume of material lost by the dome in various rockfalls. The associated seismicity was calibrated by comparing both the duration and the integrated energy with the estimated rockfall volume. At this time the overall dome was not increasing in volume, but fresh magma produced an exogenic lobe located on the steep western side with emplaced material all lost in rockfalls.

One study carried out at Piton de la Fournaise, Réunion, analyzed hundreds of events and established a relationship between rockfall volume and the seismic signal (Hibert et al. 2011). These rockfalls occurred from the steep crater walls of the basaltic volcano during periods of inflation and deflation rather than from lava emplacement.

Summary

Growing lava domes can be hazardous with the possibility of large collapses impacting the surrounding area. The characteristic seismicity produced during rockfalls can be used to determine important parameters associated with the eruption, such as the calculation of the rate of effusion, or to understand the processes associated with lava emplacement. Studies of the seismicity

therefore represent an important component of risk mitigation. In this entry results from studies of nonvolcanic rockfalls and other mass movements are presented and compared to the seismicity measured for their volcanic counterparts. The factors controlling the emplacement of domes are briefly summarized, along with how the seismicity is generated and then utilized in volcano monitoring.

Cross-References

- [Noise-Based Seismic Imaging and Monitoring of Volcanoes](#)
- [Seismic Monitoring of Volcanoes](#)
- [Very-Long-Period Seismicity at Active Volcanoes: Source Mechanisms](#)
- [Volcanic Eruptions, Real-Time Forecasting of](#)
- [Volcano-Tectonic Seismicity of Soufriere Hills Volcano, Montserrat](#)

References

- Arámbula-Mendoza R, Lesage P, Valdés-González C, Varley NR, Reyes-Dávila G, Navarro C (2011) Seismic activity that accompanied the effusive and explosive eruptions during the 2004–2005 period at Volcán de Colima, Mexico. *J Volcanol Geotherm Res* 205(1–2):30–46
- Barmin A, Melnik O, Sparks RSJ (2002) Periodic behavior in lava dome eruptions. *Earth Planet Sci Lett* 199(1–2):173–184
- Bell AF, Naylor M, Heap MJ, Main IG (2011) Forecasting volcanic eruptions and other material failure phenomena: an evaluation of the failure forecast method. *Geophys Res Lett* 38(15): L15304
- Bourrier F, Dorren L, Hungr O (2013) The use of ballistic trajectory and granular flow models in predicting rockfall propagation. *Earth Surf Process Landf* 38(4):435–440
- Brodscholl A, Kirbani SB, Voight B (2000) Sequential dome-collapse nuees ardentes analyzed from broadband seismic data, Merapi Volcano, Indonesia. *J Volcanol Geotherm Res* 100(1–4):363–369
- Budi-Santoso A, Lesage P, Dwiyono S, Sumarti S, Subandriyo S, Suroso S, Jousset P, Metaxian J-P (2013) Analysis of the seismic activity associated with the 2010 eruption of Merapi Volcano, Java. *J Volcanol Geotherm Res* 261:153–170
- Calder ES, Cortés JA, Palma JL, Luckett R (2005) Probabilistic analysis of rockfall frequencies during an andesite lava dome eruption: the Soufrière Hills Volcano, Montserrat. *Geophys Res Lett* 32:L16309. doi:10.1029/2005GL023594
- Chouet B, Dawson P (2013) Very long period conduit oscillations induced by rockfalls at Kilauea Volcano, Hawaii. *J Geophys Res Solid Earth* 118(10):5352–5371
- Coombs ML, Sisson TW, Bleick HA, Henton SM, Nye CJ, Payne AL, Cameron CE, Larsen JF, Wallace KL, Bull KF (2013) Andesites of the 2009 eruption of Redoubt Volcano, Alaska. *J Volcanol Geotherm Res* 259:349–372
- Dammeier F, Moore JR, Haslinger F, Loew S (2011) Characterization of alpine rockslides using statistical analysis of seismic signals. *J Geophys Res Earth Surf* 116(F4):F04024
- De Angelis S, Bass V, Hards V, Ryan G (2007) Seismic characterization of pyroclastic flow activity at Soufrière Hills Volcano, Montserrat, 8 January 2007. *Nat Hazard Earth Syst Sci* 7:467–472
- de' Michieli Vitturi M, Clarke AB, Neri A, Voight B (2010) Transient effects of magma ascent dynamics along a geometrically variable dome-feeding conduit. *Earth Planet Sci Lett* 295(3–4):541–553
- Deparis J, Jongmans D, Cotton F, Baillet L, Thouvenot F, Hantz D (2008) Analysis of Rock-Fall and Rock-Fall Avalanche Seismograms in the French Alps. *Bulletin of the Seismological Society of America*, 98(4):1781–1796
- DeRoin N, McNutt SR (2012) Rockfalls at Augustine Volcano, Alaska: the influence of eruption precursors and seasonal factors on occurrence patterns 1997–2009. *J Volcanol Geotherm Res* 211–212:61–75
- Diefenbach A, Crider J, Schilling S, Dzurisin D (2012) Rapid, low-cost photogrammetry to monitor volcanic eruptions: an example from Mount St. Helens, Washington, USA. *Bull Volcanol* 74(2):579–587
- Dussauge-Peisser C, Helmstetter A, Grasso JR, Hantz D, Desvarreux P, Jeannin M, Giraud A (2002) Probabilistic approach to rock fall hazard assessment: potential of historical data analysis. *Nat Hazard Earth Syst Sci* 2(1/2):15–26
- Favreau P, Mangeney A, Lucas A, Crosta G, Bouchut F (2010) Numerical modeling of landslides. *Geophys Res Lett* 37(15):L15305
- Fukao Y (1995) Single-force representation of earthquakes due to landslides or the collapse of caverns. *Geophys J Int* 122(1):243–248
- Goto A (1999) A new model for volcanic earthquake at Unzen Volcano: Melt rupture model. *Geophys Res Lett* 26(16):2541–2544
- Hammer C, Neuberg JW (2009) On the dynamical behaviour of low-frequency earthquake swarms prior to a dome collapse of Soufrière Hill volcano, Montserrat. *Geophys. Res. Letters* 36: L06305, doi:06310.01029/02008GL036837
- Hibert C, Mangeney A, Grandjean G, Shapiro NM (2011) Slope instabilities in Dolomieu crater, Réunion Island: from seismic signals to rockfall characteristics. *J Geophys Res* 116(F4):F04032

- Hibert C, Mangeney A, Grandjean G, Baillard C, Rivet D, Shapiro NM, Satriano C, Maggi A, Boissier P, Ferrazzini V, Crawford W (2014) Automated identification, location, and volume estimation of rockfalls at Piton de la Fournaise volcano. *J Geophys Res Earth Surf* 119. doi:10.1002/2013jf002970
- Hildreth W (2004) Volcanological perspectives on Long Valley, Mammoth Mountain, and Mono Craters: several contiguous but discrete systems. *J Volcanol Geotherm Res* 136(3–4):169–198
- Huggel C, Caplan-Auerbach J, Waythomas CF, Wessels RL (2007) Monitoring and modeling ice-rock avalanches from ice-capped volcanoes: a case study of frequent large avalanches on Iliamna Volcano, Alaska. *J Volcanol Geotherm Res* 168(1–4):114–136
- Iverson RM, Dzurisin D, Gardner CA, Gerlach TM, LaHusen RG, Lisowski M, Major JJ, Malone SD, Messerich JA, Moran SC, Pallister JS, Qamar AI, Schilling SP, Vallance JW (2006) Dynamics of seismogenic volcanic extrusion at Mount St Helens in 2004–05. *Nature* 444:439–443
- James MR, Varley N (2012) Identification of structural controls in an active lava dome with high resolution DEMs: Volcán de Colima, Mexico. *Geophys Res Lett* 39(22):L22303
- Jolly AD, Thompson G, Norton GE (2002) Locating pyroclastic flows on Soufrière Hills Volcano, Montserrat, West Indies, using amplitude signals from high dynamic range instruments. *J Volcanol Geotherm Res* 118(3–4):299–317
- Lavallée Y, Varley N, Alatorre-Ibargüengoitia M, Hess KU, Kueppers U, Mueller S, Richard D, Scheu B, Spieler O, Dingwell D (2012) Magmatic architecture of dome-building eruptions at Volcán de Colima, Mexico. *Bull Volcanol* 74(1):249–260
- Loughlin SC, Calder ES, Clarke AB (2002) Pyroclastic flows and surges generated by the 25 June 1997 dome collapse, Soufrière Hills Volcano, Montserrat. In: Druitt TH, Kokelaar BP (eds) *The Eruption of Soufrière Hills Volcano, Montserrat, from 1995 to 1999*. *GSL Memoirs*, London, pp 231–262
- Loughlin SC, Luckett R, Ryan G, Christopher T, Hards V, De Angelis S, Jones L, Strutt M (2010) An overview of lava dome evolution, dome collapse and cyclicity at Soufrière Hills Volcano, Montserrat, 2005–2007. *Geophys Res Lett* 37:L00E16
- Luckett R, Baptie B, Neuberg J (2002) The relationship between degassing and rockfall signals at Soufrière Hills Volcano, Montserrat. In: T.H. Druitt and B.P. Kokelaar (Editors), *The Eruption of Soufrière Hills Volcano, Montserrat, from 1995 to 1998*. *Geological Society Memoirs* 21, London, pp. 595–602
- Moran SC, Matoza RS, Garcés MA, Hedlin MAH, Bowers D, Scott WE, Sherrod DR, Vallance JW (2008) Seismic and acoustic recordings of an unusually large rockfall at Mount St. Helens, Washington. *Geophys Res Lett* 35(19):L19302
- Moretti L, Mangeney A, Capdeville Y, Stutzmann E, Huggel C, Schneider D, Bouchut F (2012) Numerical modeling of the Mount Steller landslide flow history and of the generated long period seismic waves. *Geophys Res Lett* 39(16):L16402
- Mueller SB, Varley NR, Kueppers U, Lesage P, Reyes Davila GA, Dingwell DB (2013) Quantification of magma ascent rate through rockfall monitoring at the growing/collapsing lava dome of Volcán de Colima, Mexico. *Solid Earth* 4(2):201–213
- Neuberg JW, Tuffen H, Collier L, Green D, Powell T, Dingwell D (2006) The trigger mechanism of low-frequency earthquakes on Montserrat. *J Volcanol Geotherm Res* 153(1–2):37–50
- Norris RD (1994) Seismicity of rockfalls and avalanches at three Cascade Range volcanoes: implications for seismic detection of hazardous mass movements. *Bull Seismol Soc Am* 84(6):1925–1939
- Pallister JS, Diefenbach AK, Burton WC, Muñoz J, Griswold JP, Lara LE, Lowenstein JB, Valenzuela CE (2013) The Chaitén rhyolite lava dome: Eruption sequence, lava dome volumes, rapid effusion rates and source of the rhyolite magma. *Andean Geol* 40(2): 277–294
- Pankow KL, Moore JR, Hale JM, Koper KD, Kubacki T, Whidden KM, McCarter MK (2014) Massive landslide at Utah copper mine generates wealth of geophysical data. *GSA Today* 24(1):4–9
- Rose WI (1972) Santiaguito volcanic dome, Guatemala. *GSA Bull* 83:1413–1434
- Rosser N, Lim M, Petley D, Dunning S, Allison R (2007) Patterns of precursory rockfall prior to slope failure. *J Geophys Res Earth Surf* 112(F4):F04014
- Schneider D, Bartelt P, Caplan-Auerbach J, Christen M, Huggel C, McArdell BW (2010) Insights into rock-ice avalanche dynamics by combined analysis of seismic recordings and a numerical avalanche model. *J Geophys Res Earth Surf* 115(F4):F04026
- Suriñach E, Vilajosana I, Khazaradze G, Biescas B, Furdada G, Vilaplana JM (2005) Seismic detection and characterization of landslides and other mass movements. *Nat Hazard Earth Syst Sci* 5(6): 791–798
- Varley N, Arámbula-Mendoza R, Reyes-Dávila G, Stevenson J, Harwood R (2010) Long-period seismicity during magma movement at Volcán de Colima. *Bull Volcanol* 72(9):1093–1107
- Vilajosana I, Suriñach E, Abellán A, Khazaradze G, García D, Llosa J (2008) Rockfall induced seismic signals: case study in Montserrat, Catalonia. *Nat Hazard Earth Syst Sci* 8(4):805–812
- Voight B (1988) A method for prediction of volcanic eruptions. *Nature* 332:125–130
- Voight B, Young KD, Hidayat D, Subandrio, Purbawinata MA, Ratdomopurbo A, Suharna, Panut, Sayudi DS, LaHusen R, Marso J, Murray TL, Dejean M, Iguchi M, Ishihara K (2000) Deformation and seismic precursors to dome-collapse and fountain-collapse nuées ardentes at Merapi Volcano, Java, Indonesia, 1994–1998. *J Volcanol Geotherm Res* 100(1–4): 261–287

- West ME (2013) Recent eruptions at Bezymianny volcano – a seismological comparison. *J Volcanol Geotherm Res* 263:42–57
- Woods AW, Sparks RSJ, Ritchie LJ, Batey J, Gladstone C, Bursik M (2002) The explosive decompression of a pressurized lava dome: the 26 December (Boxing Day) 1997 collapse and explosion of Soufrière Hills Volcano, Montserrat. In: Druitt TH, Kokelaar BP (eds) *The Eruption of Soufrière Hills Volcano, Montserrat, from 1995 to 1999*. *GSL Memoirs* 21. Geological Society, London, pp 457–466
- Yamada M, Matsushi Y, Chigira M, Mori J (2012) Seismic recordings of landslides caused by Typhoon Talas (2011), Japan. *Geophys Res Lett* 39(13):L13301
- Zhao J, Mangeney A, Stutzmann E, Capdeville Y, Moretti L, Calder ES, Smith PJ, Cole P, Le Friant A (2013) Model space exploration for determining landslide source history from long period seismic data. *Geophys Res Abstr* 15:EGU2013–EGU2479

Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding

Panayiotis C. Polycarpou and

Petros Komodromos

Department of Civil and Environmental Engineering, University of Cyprus, Nicosia, Cyprus

Synonyms

Adjacent buildings; Collision; Elastomeric pads; Impact; Neighboring buildings; Rubber bumpers; Seismic gap; Seismic isolation; Standing distance

Introduction

In densely resided areas and city centers, neighboring buildings are usually constructed very close to each other without adequate clearance between them. Therefore, during strong earthquakes, structural pounding may occur between adjacent buildings, due to deformations of their stories “► [Learning from Earthquake Disasters](#).” Consequences of such pounding occurrences, ranging from light local damage to severe structural

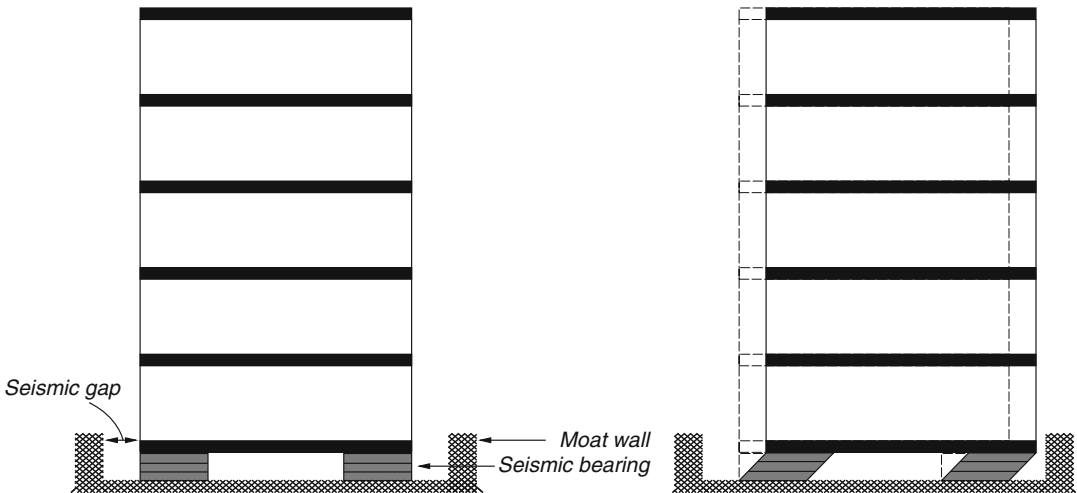
damage or even collapse, have been observed and reported in past strong earthquakes “► [Post-Earthquake Diagnosis of Partially Instrumented Building Structures](#)” (Anagnostopoulos 1995; Bertero 1987; EERI 1990, 2000). In case of structural pounding, both floor accelerations and interstory deflections may be significantly amplified, threatening the functionality and the contents of the building (Papadrakakis and Mouzakis 1995; Anagnostopoulos 1988; Komodromos et al. 2007). The photograph in Fig. 1 shows a pounding incidence between two adjacent buildings, as reported from an EERI/PEER reconnaissance team after the L’Aquila Earthquake, which hit Central Italy in April 2009 (EERI 2009). During that seismic event, the roof of a two-story building hit an adjacent four-story structure causing considerable damage to the columns of the latter at that level. However, the third and the fourth stories of the building did not experience any significant damage. The limitation of the damage of the four-story building at the level of impact indicates the destructive effect of structural pounding.

Furthermore, pounding may also take place in cases of seismically isolated buildings, which exhibit quite different dynamic characteristics from conventionally fixed-support buildings. In particular, pounding of a seismically isolated building occurs primarily as a consequence of the large relative displacements at the isolation level (Fig. 2), due to the excess flexibility that is provided there through the seismic isolators rather than due to the deformation of the superstructure. This fact, in combination with a limited width of the seismic gap that is provided around the building, results in impacts between the structure and the surrounding moat wall during very strong, near-source “► [Seismic Sources from Landslides and Glaciers](#)” seismic excitations (Komodromos et al. 2007).

At the pounding floors, short-period impulses of high amplitude are observed in the acceleration response, while their amplitude is affected by the impact stiffness. The presence of high spikes in the acceleration response due to structural pounding is a very critical issue, especially for buildings that may house sensitive equipment.

Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 1

Damage of a four-story conventional building due to pounding with its adjacent two-story building, during the L'Aquila earthquake in Italy, in April 2009 (Source: <http://www.eqclearinghouse.org/italy-090406/>)



Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 2 (a) Configuration of a seismically isolated building; (b) Mode of deformation during an earthquake

Therefore, it is very vital to consider potential impact mitigation measures that could be employed in practice.

Certain mitigation measures have already been proposed by other researchers who investigated this problem in buildings and bridge decks, in an effort to alleviate the damaging effects of

structural pounding (Warnotte et al. 2007). One of the proposed measures is the incorporation of layers of soft material, such as elastomeric material, on certain locations, where impact is likely to occur, in order to act as a shock absorber. The idea is based on similar measures that are usually taken in ports or harbors, where soft bumpers are

Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 3 Rubber shock absorber attached within the seismic gap of a seismically isolated building in Cyprus



attached at the wharfs in order to avoid damage on the docking ships due to collisions with the docks. Figure 3 shows a case of applying rubber bumpers in the seismic gap of a seismically isolated building located in Nicosia, Cyprus. The incorporated rubber shock absorbers can act as restrainers to prevent larger horizontal displacements than the seismic isolators can accommodate and avoid sudden impact pulses due to potential structural pounding. However, there is a question about the effectiveness of this technique, since the introduction of such material, with a certain thickness, reduces accordingly the width of the available seismic gap between the colliding structures. It is widely known that the width of the seismic gap affects significantly the dynamic response during pounding and, specifically, as the width of the seismic gap increases the detrimental effects of pounding are generally alleviated.

In order to assess the effectiveness of such an impact mitigation measure, proper numerical simulations and parametric studies are needed, considering various types and configurations of structures under different dynamic excitations. Nevertheless, the behavior of rubber shock absorbers under impact loading must be sufficiently well modeled. Undoubtedly, the most precise modeling can be achieved with the development of a detailed finite element model using special elements and material laws to

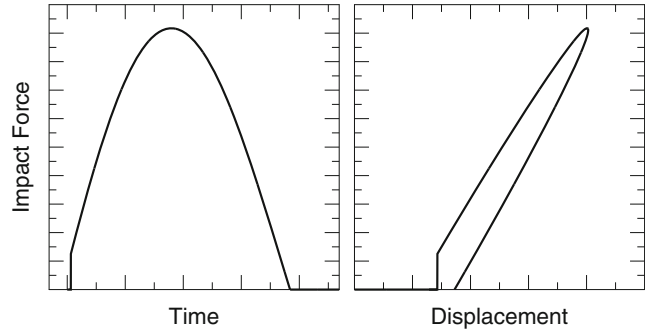
properly represent the nonlinearity of the problem. However, such simulations are computationally very demanding and cannot be effectively employed for a parametric investigation of the problem. Therefore, an impact model should be used in numerical simulations of simple multidegree of freedom dynamic systems, which will be able to provide, with sufficient accuracy, the pounding forces considering the usage of rubber shock absorbers at impact locations (Polycarpou and Komodromos 2009, 2013). In the following paragraphs, an efficient methodology for simulating the behavior of rubber bumpers of certain thickness and material properties is described.

Impact Modeling

Usually, in numerically simulated dynamic systems, such as multistory buildings under earthquake excitations, structural impact is considered using force-based methods, also known as “penalty” methods. These methods allow relatively small interpenetration between the colliding structures, which can be justified by the local deformability at the point of impact. The interpenetration depth is used together with an impact stiffness coefficient, which represents an impact spring, to calculate the impact forces that act on the colliding structures and push them apart.

Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 4

The modified linear viscoelastic impact model, used for simulating concrete-to-concrete impacts



Based on the mathematical relation between the impact force and the interpenetration depth, the impact models can be classified as linear and non linear models. Moreover, some models assume that an impact dashpot acts in parallel to the contact spring in order to take into account the energy that is dissipated during an impact (Polycarpou and Komodromos 2010).

Concrete-to-Concrete Impacts

In the proposed methodology, impacts between concrete surfaces are simulated assuming a linear impact spring and an impact dashpot exerting, in parallel, impact forces to the colliding structures whenever their separation distances are exceeded. In particular, when a contact is detected, the impact force is estimated at each time step using the following formulas (Komodromos et al. 2007):

$$F_{imp}(t + \Delta t) = \begin{cases} k_{imp} \cdot \delta(t) + c_{imp} \cdot \dot{\delta}(t) & \text{when } F_{imp}(t) > 0 \\ 0 & \text{when } F_{imp}(t) \leq 0 \end{cases} \quad (1)$$

where $\delta(t)$ is the interpenetration depth, $\dot{\delta}(t)$ is the relative velocity between the colliding bodies, k_{imp} is the impact spring's stiffness, and c_{imp} is the impact damping coefficient. The latter is computed according to the following formulas, provided by Anagnostopoulos (1988):

$$c_{imp} = 2 \cdot \xi_{imp} \sqrt{k_{imp} \cdot \frac{m_1 \cdot m_2}{m_1 + m_2}} \quad (2)$$

$$\xi_{imp} = -\frac{\ln(COR)}{\sqrt{\pi^2 + (\ln(COR))^2}} \quad (3)$$

In the aforesaid formulas, m_1 , m_2 are the masses of the two bodies and COR is the coefficient of restitution, which depends on the plasticity of the impact and is defined as the ratio of relative velocities after and before impact. Its value rates between 0, which corresponds to totally plastic impact (without rebound), and 1, which represents

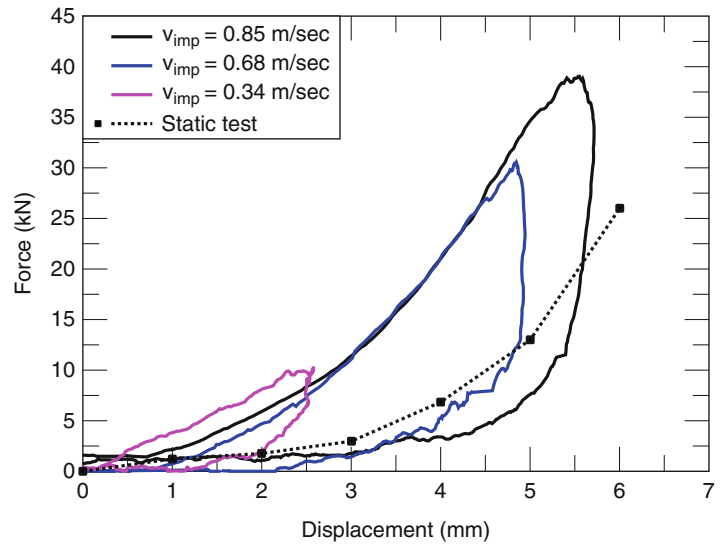
the case of elastic impact. An estimated value of the coefficient of restitution has to be provided before the analysis, taking into account the geometry and material characteristics at the vicinity of the impact. For example, the value of the COR, for the case of structural pounding between concrete structures, is commonly set to a value between 0.5 and 0.7. Actually, the above impact model (Fig. 4) is a small variation of the classical linear viscoelastic impact model that had been initially proposed by Anagnostopoulos (1988), in which the tensile forces that arise at the end of the restitution period are omitted and a small plastic deformation is introduced, which increases the available clearance between the adjacent structures.

Modeling of Rubber Bumpers – Proposed Impact Model

A significant part of this numerical problem has to do with the simulation of the behavior of

Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding,

Fig. 5 Experimental results, involving both static and impact tests of rubber shock absorbers, (Obtained from Kajita et al. (2006))



rubber bumpers under impact loading. The usage of linear impact models for simulating the response of rubber during impact loading does not seem to be the most suitable, considering the stress–strain curves obtained from experiments (Kajita et al. 2001, 2006; Kawashima et al. 2002; Shim et al. 2004). In particular, static and dynamic compressive tests of rubber reveal an exponential relationship between the compressive load and the corresponding displacement, as shown in the plots of Fig. 5. Therefore, it would be more appropriate to simulate the incorporation of rubber bumpers by using a nonlinear impact model. Furthermore, since a rubber shock absorber has a finite thickness, there is a possibility to reach its ultimate compressive strain during severe impacts, whereas the impact stiffness should represent the material behind the rubber (e.g., concrete) and not the rubber bumper, after the ultimate strain of the rubber is exceeded.

Subsequently, a new impact model is proposed in the following paragraphs, taking into account the observations from relevant impact

tests as well as the special characteristics of rubber bumpers. Specifically, the proposed nonlinear impact model assumes that the impact force is increasing exponentially with the interpenetration depth during the approach phase of impact, which is very close to the profile of the stress–strain curves obtained from the corresponding experiments. In order to take into account the dissipation of kinetic energy during impact, a different path of the force–displacement curve is followed during the restitution phase, forming a hysteresis loop (Fig. 6). In particular, the impact force for a certain indentation value during the detaching phase is reduced compared to the corresponding impact force for the same indentation during the approach phase. Actually, this reduction, which depends on certain parameters, such as the material characteristics and the impact velocity, determines the area of the hysteresis loop and, therefore, the amount of dissipated energy during impact. The impact force during the approaching phase is provided by the formula

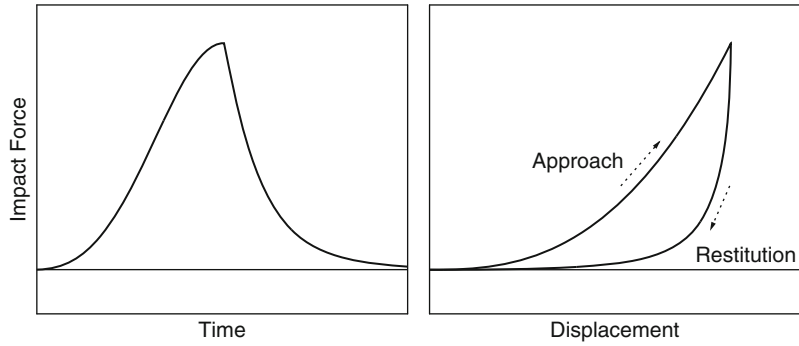
$$F_{imp}^A = \left\{ \begin{array}{ll} k_{imp} \cdot \delta^n & \text{for } \delta < \delta_u \\ k_{imp} \cdot \delta_u^n + k_{imp_PY} \cdot (\delta - \delta_u) & \text{for } \delta > \delta_u \end{array} \right\} \text{ when } \dot{\delta} > 0 \quad (4)$$

Equation. 4 takes into account the case of exceeding the ultimate compressive strain of the

material, during the approach phase. Specifically, it is assumed that, after a certain indentation, δ_u ,

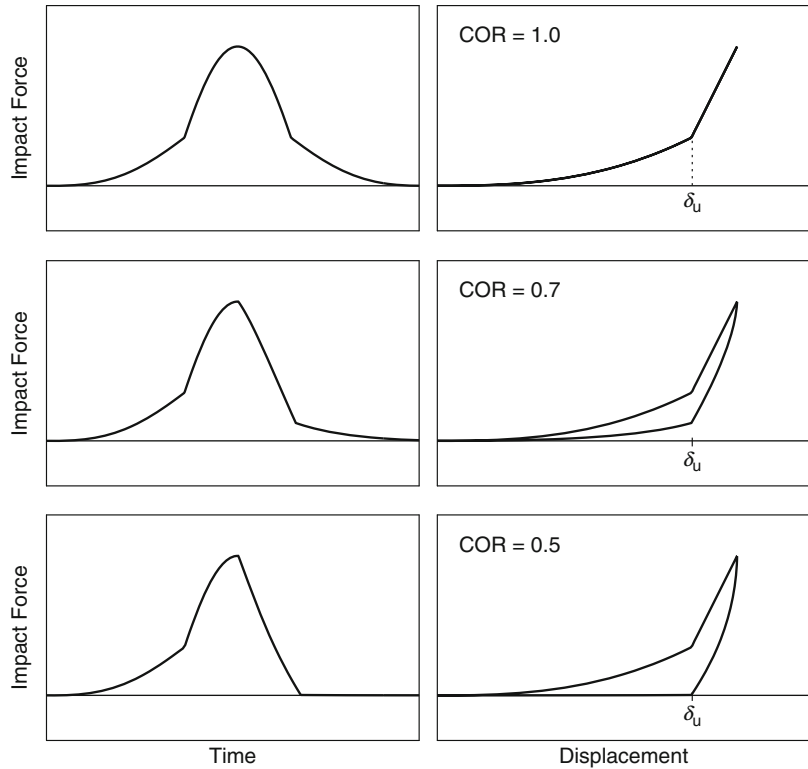
Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 6

Force-displacement diagram of the non-linear impact model with hysteretic damping



Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 7

Impact force in terms of time and displacement in the case of exceeding the ultimate compressive strain capacity of the rubber bumper



which corresponds to the ultimate compressive capacity of the rubber bumper, the exponential trend becomes linear with a linear postyield stiffness, k_{imp_PY} (Fig. 7).

During the restitution phase, the impact force is described by the following expression:

$$F_{imp}^R = k_{imp} \cdot \delta^n \cdot (1 + C_{imp} \cdot \dot{\delta}) \quad \text{for } \dot{\delta} < 0 \quad (5)$$

In Eq. 4, k_{imp} is the impact stiffness, δ is the indentation, and n is the impact exponent ($n > 1$). The dot over δ denotes derivation with respect to time, so $\dot{\delta}$ represents the approaching relative velocity of the colliding bodies. The impact stiffness is given by the following expression:

$$k_{imp} = \alpha \cdot k_{st} = \alpha \cdot \frac{A \cdot K_r}{d^n} \quad (6)$$

where k_{st} is the bumper's static stiffness and $\alpha > 1$ is a multiplier that expresses the strain rate dependency of the material stiffness, and for simplicity, it can be assumed that it ranges between the values of 2 and 2.5 for common earthquake-induced structural pounding velocities, based on relevant experimental results (Kajita et al. 2006). A is the contact area of the bumper, d is the bumper's thickness, and K_r expresses the material stiffness. The unknown parameters that have to be determined in Eq. 6 are the material stiffness K_r and the exponent n . The values of both parameters depend on the material characteristics, and therefore, their evaluation can be done experimentally. In particular, a static test curve of a rubber specimen (see, e.g., Fig. 5a) can be approximated with an exponential curve of the form

$$f(x) = c \cdot x^b \quad (7)$$

In this way, c can represent k_{st} , while b can represent the exponent n . Then, the material stiffness can be calculated by substituting these values in Eq. 6 and solving for K_r . After obtaining the material properties K_r and n , the impact stiffness of any rubber bumper with the same material and any dimensions can be calculated using Eq. 6.

The damping term C_{imp} in Eq. 5 is given by the formula (Polycarpou et al. 2013)

$$C_{imp} = \frac{(1 - COR^2)}{2 \cdot v_{imp}} \cdot \frac{\ln^3(COR)}{COR \cdot (2 + \ln^2(COR) - 2 \cdot \ln(COR)) - 2} \quad (8)$$

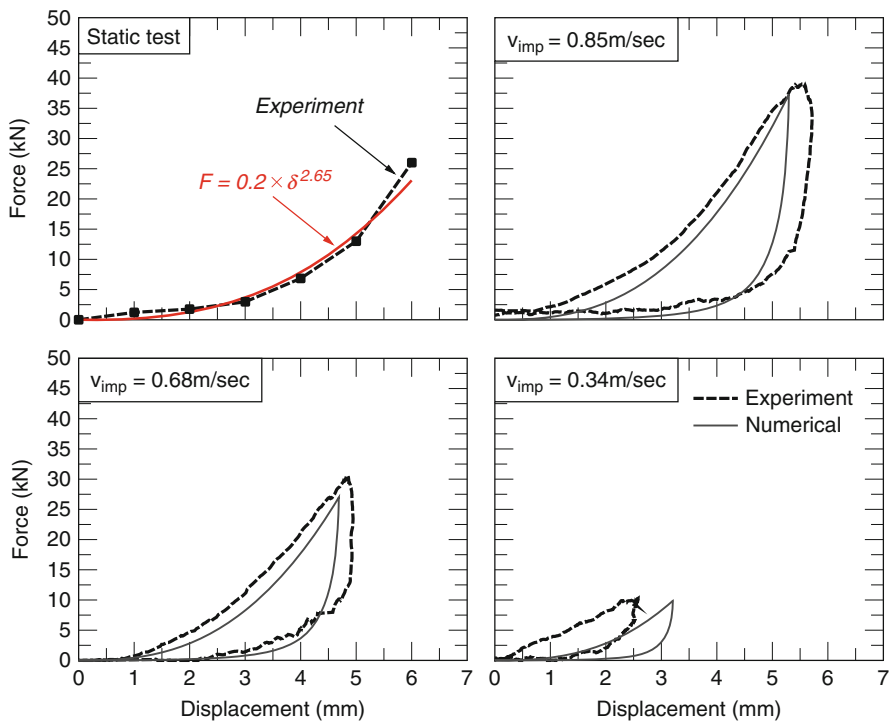
where v_{imp} is the impact velocity, which is the relative velocity of the two bodies just before impact.

The force-time and force-displacement diagrams of the proposed nonlinear model for simulating the response of rubber bumpers under impact loading are shown in Fig. 6. Figure 7 demonstrates the same diagrams in the case of exceeding the ultimate compressive capacity of the bumper for three different values of the coefficient of restitution.

Impact Model's Validation and Parameters' Determination Based on Experimental Data

In order to validate the accuracy of the proposed nonlinear hysteretic impact model, the load-displacement curves obtained from the collision tests are compared with the corresponding results from numerical analyses, simulating the impact of two free bodies and considering the proposed impact model. The estimation of the impact parameters is based on the static load-displacement curve of the rubber that was used in the experiments. In particular, the static test curve is approximated with an exponent $n = 2.65$ and a static stiffness $k_{st} = 0.2 \text{ kN/mm}^{2.65}$. Consequently, considering the dimensions of the shock absorber that was used in the experiments ($40 \times 40 \times 10 \text{ mm}$) and substituting to Eq. 6, the material's stiffness K_r is found to be equal to $55,835 \text{ kN/m}^2$. The strain rate multiplier α is taken to be equal to 2.25, and thus, the impact stiffness for the dynamic response is calculated through Eq. 6 to be equal to $0.45 \text{ kN/mm}^{2.65}$. The energy loss during similar impact tests was found to range around 40–50 % of the initial kinetic energy when using the rubber shock absorbers (Kajita et al. 2001). Accordingly, the coefficient of restitution is assumed to be equal to 0.5 for the simulations. Nevertheless, the value of the coefficient of restitution, in the proposed impact model, does not affect the value of the maximum impact force but only the trend of the restitution phase, determining the hysteretic energy loss.

The plots in Fig. 8 present the experimental results from Kajita et al. (2006) in comparison with the corresponding numerical results, obtained from the performed simulations considering the proposed nonlinear hysteretic impact model for the simulation of the rubber bumpers. In particular, the force-displacement diagrams are presented for the static test and three different impact velocities, considering an impact shock absorber of 10 mm thick attached between two colliding masses of 300 kg each. It is observed that the trends obtained from numerical analysis, using the proposed impact model, correlate very well with the corresponding experimental data. In particular, there is a good approximation of the maximum impact force as well as the shape and



Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 8 Comparison between experimental and numerical force-displacement

diagrams of a 10 mm thick rubber shock-absorber for various impact velocities

size of the hysteresis loop of the proposed impact model with the corresponding experimental results.

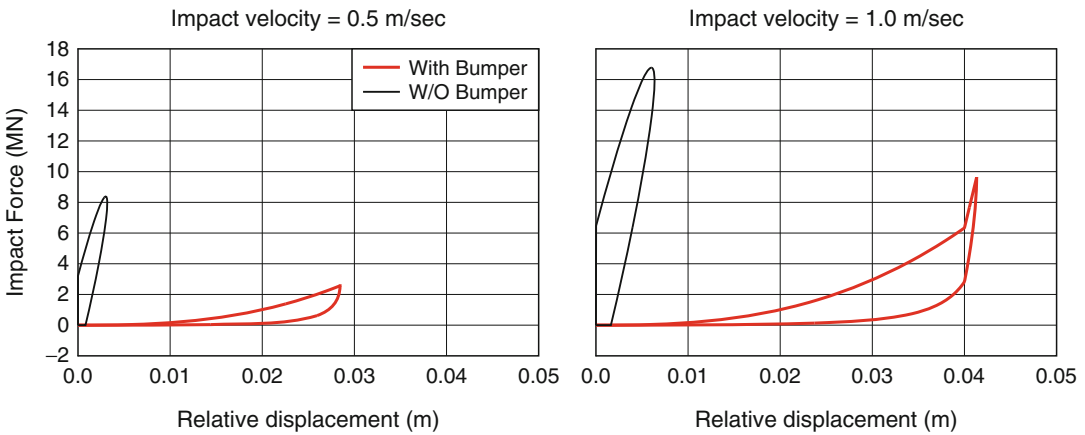
Numerical Example Considering Two Rigid Bodies

In order to examine the effect of using a rubber shock absorber on the computed responses after impact, a simple numerical example of two free rigid bodies of equal masses that collide with a constant relative velocity has been performed. Two different circumstances have been considered regarding the area of contact. In the first case, concrete-to-concrete impact was considered and the modified linear viscoelastic impact model was used (Eq. 1). In the second case, a rubber bumper 5 cm thick was assumed to be incorporated at the area of contact, which is simulated using the nonlinear impact model with hysteretic

Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Table 1 Impact parameters for the cases without and with rubber bumper

Property	No bumper	With bumper
Impact model	Linear	Non-linear
Exponent (n)	1.0	2.65
Impact stiffness (k_{imp})	2,500 kN/mm	0.36 kN/mm ^{2.65}
Coefficient of Restitution (COR)	0.6	0.5
Bumper thickness (d)	—	5 cm
Bumper's max strain (δ_u/d)	—	0.8
Post-yield impact stiffness (k_{imp_PY})	—	2,500 kN/mm

damping (Eqs. 4 and 5). The impact parameters that were used in both cases are provided in Table 1. The masses of the two colliding rigid structures are assumed to be 320 t each, while two different values of impact velocity were used, specifically 0.5 and 1.0 m/s.



Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 9 Impact force – displacement diagram for the cases of two

impacting rigid bodies, with and without the incorporation of a rubber bumper and for two different values of the impact velocity

Figure 9 demonstrates the load–displacement diagrams for the two cases of the impact velocity and for both cases of with and without the use of the rubber bumper. The results show that the indentation, which represents the local deformation at the vicinity of an impact, is much larger in the case of having the rubber shock absorber due to the reduced impact stiffness. Furthermore, in the case of the relatively high impact velocity of 1.0 m/s, the deformation exceeds the maximum compressive capacity of the 5 cm thick rubber bumper, and the impact force begins to rise rapidly, since the postyield linear impact stiffness is used.

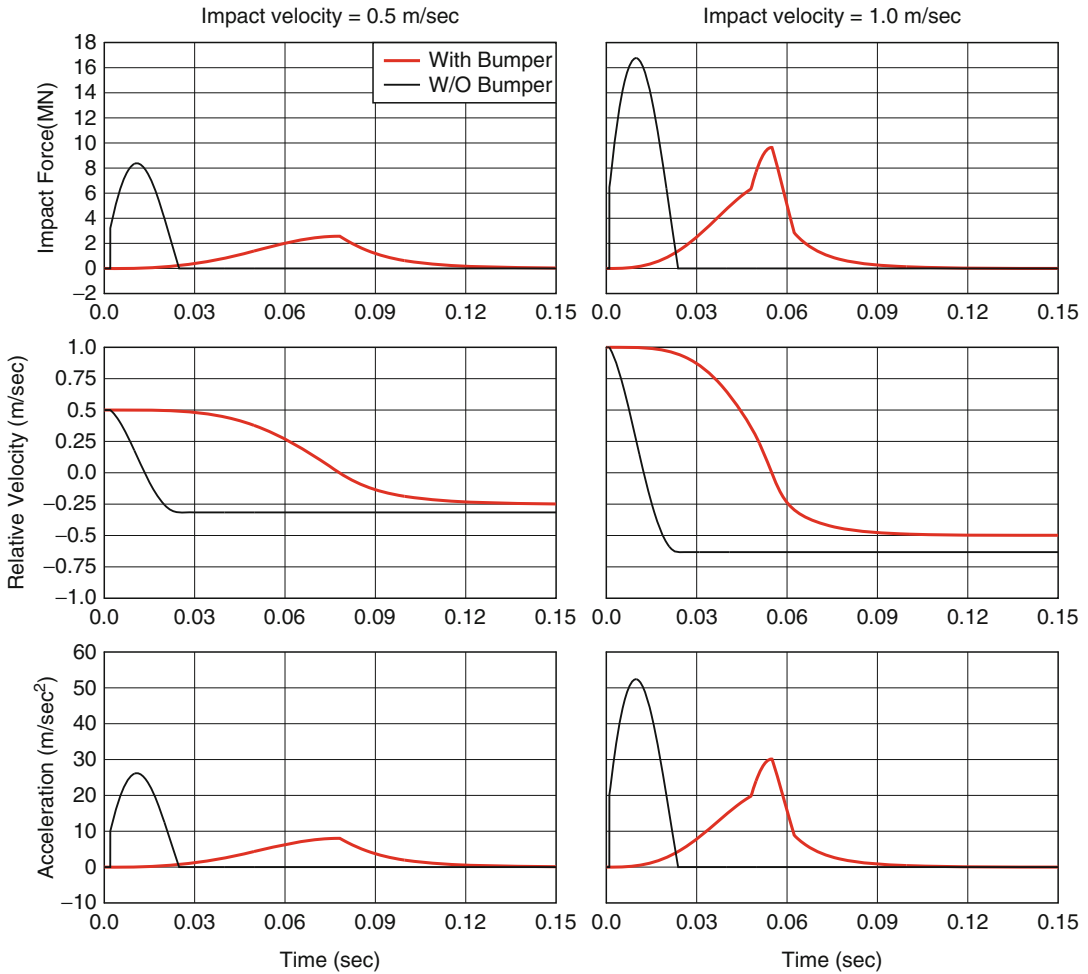
The plots in Fig. 10 show the impact force, relative velocity, and acceleration time histories for the same cases. It is evident that the usage of the rubber bumper elongates the duration of impact and reduces both the maximum impact force and the maximum acceleration. The ratio between the relative velocity after and before impact is equal to the coefficient of restitution used in the corresponding impact model, which verifies the correctness of the two impact models used in the simulations.

Although the use of rubber bumpers seems to be an effective impact mitigation measure in the case of two colliding free bodies, its effectiveness has to be assessed in the case of earthquake-induced pounding of adjacent buildings, since

the incorporation of the bumpers reduces the available seismic gap, increasing the possibility of pounding. For this purpose, two practical examples are presented in the following paragraphs in order to assess the effectiveness of using rubber shock absorbers as an impact mitigation measure for both cases of seismically isolated and conventionally fixed-support buildings. For the numerical simulations, a specialized software application has been developed using modern object-oriented programming in order to efficiently perform dynamic analyses of buildings in two dimensions “► [Time History Seismic Analysis](#)” modeling the consideration of potential poundings with adjacent structures.

Application Example Using a Seismically Isolated Building

A four-story seismically isolated building is considered in this example, assuming shear beam behavior for the superstructure and bilinear behavior for the base isolation system “Elastomeric Bearings and their Implementation into Structural Design and Analysis” (Fig. 11). The initial seismic gap around the building is considered to be equal to 25 cm. The same building is considered under a second configuration, where rubber shock absorbers of 5 cm thick are



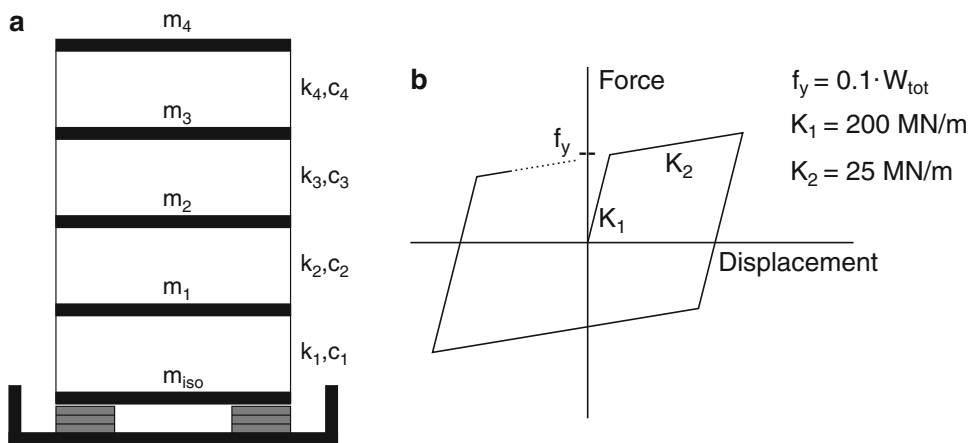
Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 10 Impact force, relative velocity and acceleration time-histories for the

case of two impacting rigid bodies, with and without the incorporation of a rubber bumper and for two different cases of the impact velocity

attached around the building at the isolation level, with a clearance of 20 cm.

For the case without the use of bumpers, where the building hits against the surrounding moat wall (concrete-to-concrete impact), the linear viscoelastic impact model is used with an impact stiffness of 2,500 kN/mm and a coefficient of restitution equal to 0.7. In the case of incorporating the rubber shock absorbers, the proposed nonlinear hysteretic impact model is used with an exponent $n = 2.65$, $k_{imp} = 0.36$ kN/mm^{2.65} and a $COR = 0.45$. The impact stiffness of the

bumpers was calculated using Eq. 6, considering four pieces of rubber with an area of 15 cm × 15 cm each, attached at each side of the seismically isolated building (Fig. 12). The Sylmar Converter Station record “► [Recording Seismic Signals](#)” ($PGA = 0.897g$) of the Northridge 1994 earthquake ($M_w = 6.7$) is used as the ground excitation for the simulations. For the seismically isolated building under consideration, the maximum induced unconstrained displacement at the isolation level due to this earthquake record is 31.67 cm.



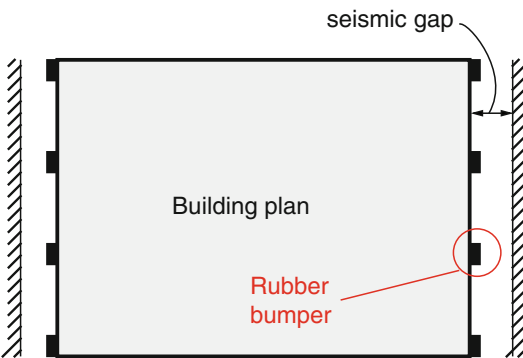
c

Parameter	Value
Story stiffness (k_i)	600 MN/m
Story mass (m_i)	320 ton
Top story mass (m_n)	250 ton
Superstructure's damping ratio (ξ_{sup})	2%
Mass at isolation level (m_{iso})	320 ton
Isolation's viscous damping ratio (ξ_{iso})	5%

Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 11 (a) Model of a seismically isolated building; (b) Bilinear model of the

isolation system behavior; (c) Structural properties of the considered seismically isolated building

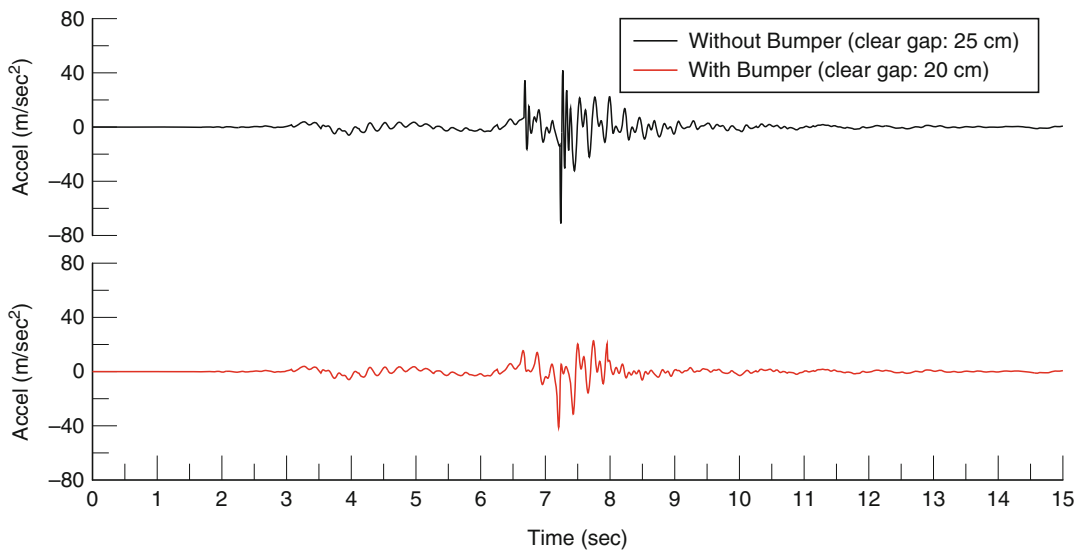
R



Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 12 Locations of rubber shock absorbers in a plan view of the building

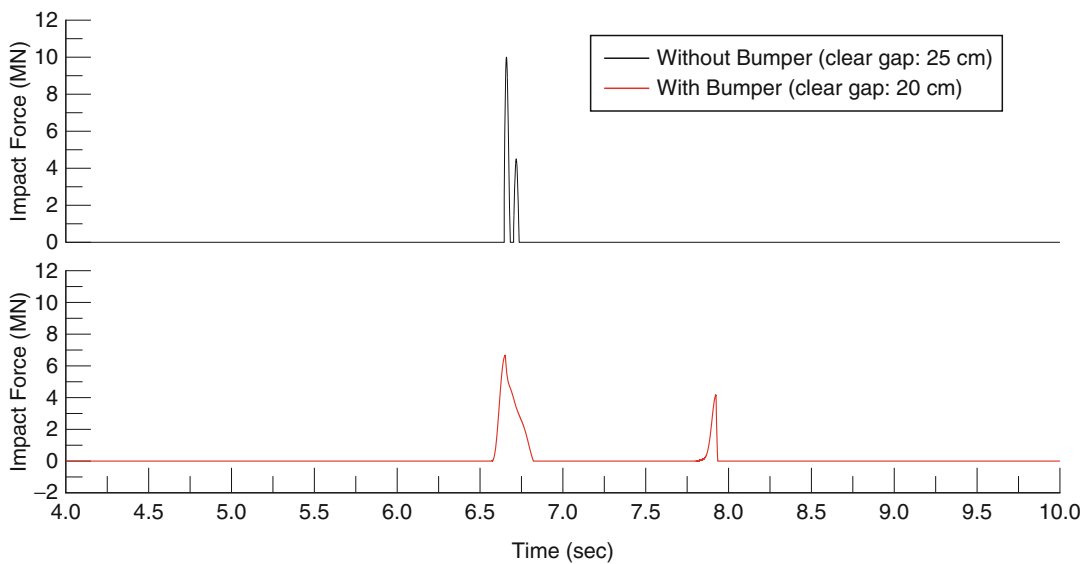
Figure 13 presents the acceleration time histories at the base of the seismically isolated building, where poundings occur, for both cases, without and with bumpers. Although in the case of having bumpers the available clearance is reduced from 25 to 20 cm, it is observed that the maximum acceleration response is lower than the corresponding peak acceleration without bumpers. In particular, the high spikes in the acceleration response seem to be eliminated due to the usage of the rubber shock absorbers.

This can be explained by observing the impact force time history as presented in Fig. 14.



Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 13 Differences on the acceleration time-history at the isolation level due to

the attachment of 5 cm wide rubber shock absorbers the impact locations

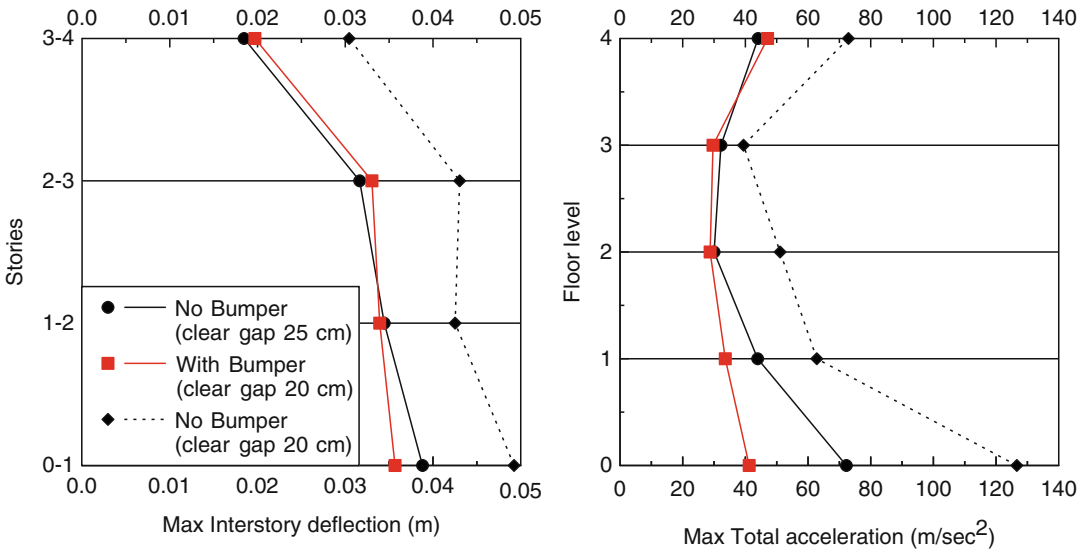


Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 14 Differences on the impact force time-history at the isolation level due to

the attachment of 5 cm wide rubber shock absorbers at the impact locations

Specifically, the plot indicates that the impact forces in the case of adding bumpers not only become smaller but also the duration of the impact is longer, smoothening the acceleration response at the corresponding floor level.

Figure 15 displays the maximum responses at all floors of the seismically isolated building for the two configurations examined, i.e., with and without rubber bumpers, as well as the case of having a seismic gap of 20 cm without using bumpers.



Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 15 Differences on the maximum responses of the seismically isolated

building due to the attachment of 5 cm wide rubber shock absorbers at the impact locations

While for the case of reducing the seismic gap size to 20 cm without adding any bumpers, the response is substantially amplified, no significant increase of the maximum interstory deflections or the maximum floor acceleration at the upper floors were observed due to the decreased gap size when rubber bumpers are used. On the contrary, the maximum interstory deflection at the first story, which is the largest among all stories, slightly decreases after the incorporation of the rubber. Nevertheless, it has to be mentioned that these observations concern only the specific earthquake excitation and the specific seismically isolated building. There is a need for further investigation, performing numerous simulations considering different characteristics of the structures and different earthquake records, where the effectiveness of such impact mitigation measures will be more generally assessed.

Application Example and Parametric Analyses Considering Two Conventional Fixed-Supported Buildings

Next, another example is presented to examine also the case of using rubber shock absorbers as

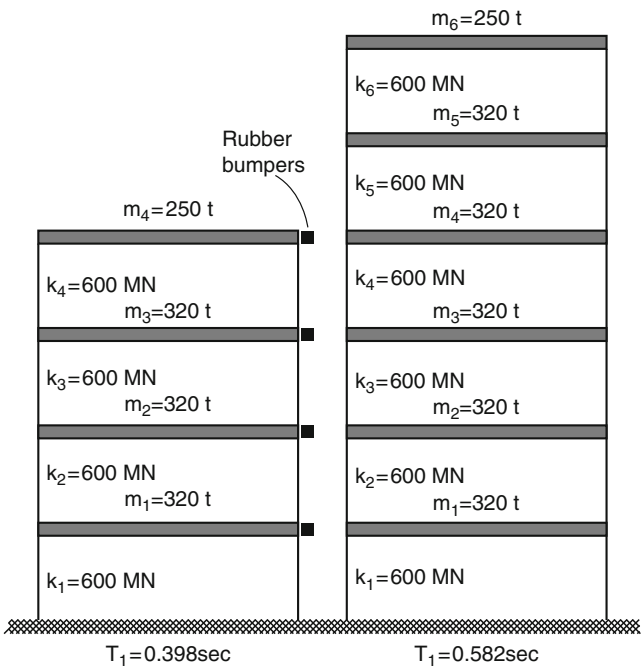
an impact mitigation measure for cases of narrow seismic gap sizes between adjacent multistory conventionally fixed-support buildings. Again, the simulated buildings are modeled in 2D as multi-degree of freedom (MDOF) systems, with shear beam behavior and the masses lumped at the floor levels, assuming linear elastic behavior during earthquake excitations.

A four-story and a six-story fixed-support buildings are considered in series for the performed simulations, as shown in Fig. 16. Each floor has a lumped mass of 320 t, except the top floor where a mass of 250 t is considered. Each story has a horizontal stiffness of 600 MN, while a constant viscous damping ratio of 5 % has been considered for both buildings. The floors of the neighboring buildings are assumed to be at the same levels.

For the particular structural system, the performed analysis examined whether the incorporation of rubber bumpers at the locations of potential impacts, which reduces the available seismic gap width, would be beneficial for the colliding buildings or not. For the performed simulations, 5 cm thick bumpers were assumed to be installed at all floor levels, as shown in Fig. 16. Therefore, by applying the rubber shock

Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 16

The two multistory buildings considered in the simulations and their structural properties



absorbers at the side of one of the two buildings, the existing seismic gap size reduces by 5 cm, when compared with the case without the bumper. Consequently, the results obtained from the simulations considering the use of rubber bumpers are compared with the corresponding results from the case without bumpers but with a clearance that is 5 cm wider. In the performed parametric analyses, the available seismic gap was varied in the range of 5–25 cm, which corresponds to a clearance width of 0–20 cm in the case of incorporating rubber bumpers.

In order to investigate the effect of the earthquake characteristics, three different seismic records (Table 2) from relatively strong and widely known earthquakes were selected and used as ground excitations.

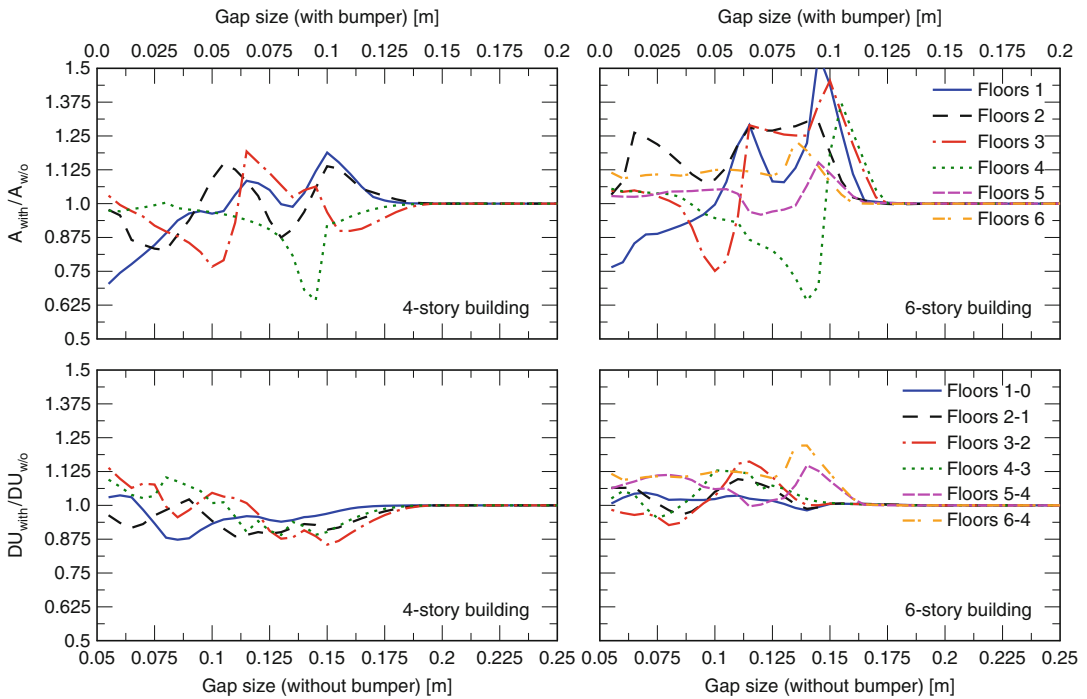
Plots in Fig. 17 demonstrate the effect of using rubber bumpers on the computed response of the four-story and the six-story buildings, in terms of the size of the seismic gap for the Kobe earthquake record. In particular, the plots present the amplification of the peak floor accelerations and peak interstory deflections due to the incorporation of rubber shock absorbers with a thickness of 5 cm, between the two buildings. The

Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Table 2 Earthquake records that were used in the simulations

Earthquake	Mw	Recording station	P.G.A. (g)
Kobe, Japan 1995	6.9	0 KJMA	0.821
Northridge, USA 1994	6.7	74 Sylmar – Converter	0.897
San Fernando, USA 1971	6.6	Pacoima Dam, S16	1.170

amplification of the response is defined as the ratio of the response obtained after the incorporation of rubber bumpers, which unavoidably reduce the available clearance, to the corresponding response, without the usage of bumpers. Therefore, the usage of rubber bumpers has beneficial effects on the corresponding response quantity when the amplification ratio value is smaller than 1.0.

The results indicate that the size of the seismic gap affects the effectiveness of the rubber bumpers in a different manner on each floor and for each building. For example, the peak floor acceleration at the fourth floor of the four-story



Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 17 Amplification of the peak floor accelerations and interstory deflections of

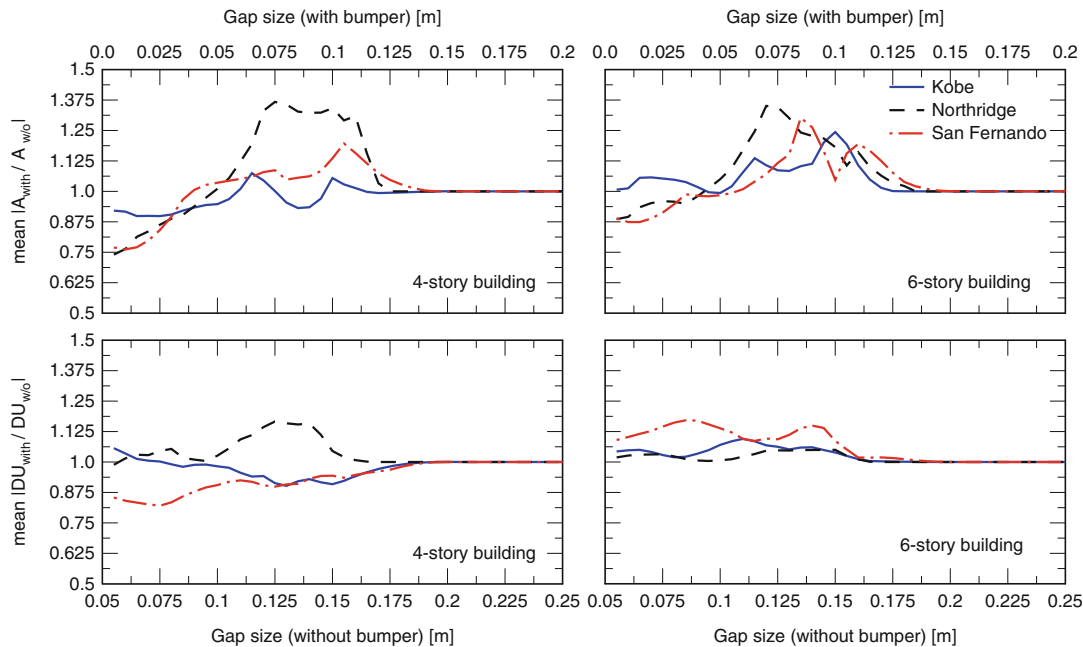
the 4-story and the 6-story buildings, due to the usage of rubber shock-absorbers, in terms of the width of the seismic gap, considering the Kobe earthquake record

building is reduced after the incorporation of the bumper almost for all seismic gap sizes, while at the same time the peak acceleration at the second floor of the six-story building amplifies up to 30 %. Moreover, the maximum interstory deflections are not affected in the same way with the floor accelerations, since the latter may be amplified after the use of the rubber bumper, while the former are reduced for a certain gap size. Nevertheless, peak floor accelerations seem to be more sensitive to the usage of bumpers than interstory deflections, since the variations of the curves in the plots of Fig. 17 are more pronounced in the former case.

In order to be able to provide the computed results from all three earthquake records in the same plots, the mean peak responses among all floors of the buildings are computed and plotted in Fig. 18. Specifically, these plots demonstrate, in a more general form, the effect of applying rubber bumpers of 5 cm thick inside the available

gap on the overall seismic response of the two buildings. It is observed that the characteristics of the earthquake excitation affect the effectiveness of this kind of an impact mitigation measure, in combination with the size of the available clearance. It can be also observed that, under the considered circumstances, the incorporation of such a shock absorber amplifies, in most of the times, the response, especially in the case of the six-story building. However, there are some cases of relatively narrow gap sizes in which the usage of rubber bumpers seems to be beneficial.

In the previously presented simulations, it has been assumed that, after the attachment of rubber bumpers on the side of the seismically isolated building, the reduction of the available clearance from the surrounding moat wall equals the corresponding thickness of the bumpers. However, the rubber bumpers could be attached in small cavities on the buildings' walls, taking full advantage of the compressible width of the



Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 18 Mean values of the peak responses among all floors of the 4-story and

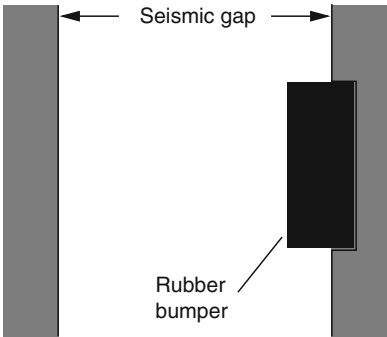
the 6-story buildings, due to the usage of rubber shock-absorbers, in terms of the width of the seismic gap

rubber, as shown in Fig. 19, without unnecessarily decreasing further the width of the seismic gap. For example, if the thickness of a rubber bumper is 5 cm and its maximum compressive strain equals 0.8, then the compressible width δ_u of the bumper is 4 cm. Therefore, if the particular 5 cm thick shock absorber is attached in a cavity that is 1 cm deep, its effective width of 4 cm can be fully utilized, without unnecessarily decreasing further the width of the available seismic gap by 1 cm.

The above technique seems to be quite efficient since the corresponding amplification ratios due to the incorporation of the bumpers, shown in Fig. 20, are substantially reduced in relation to those of Fig. 18.

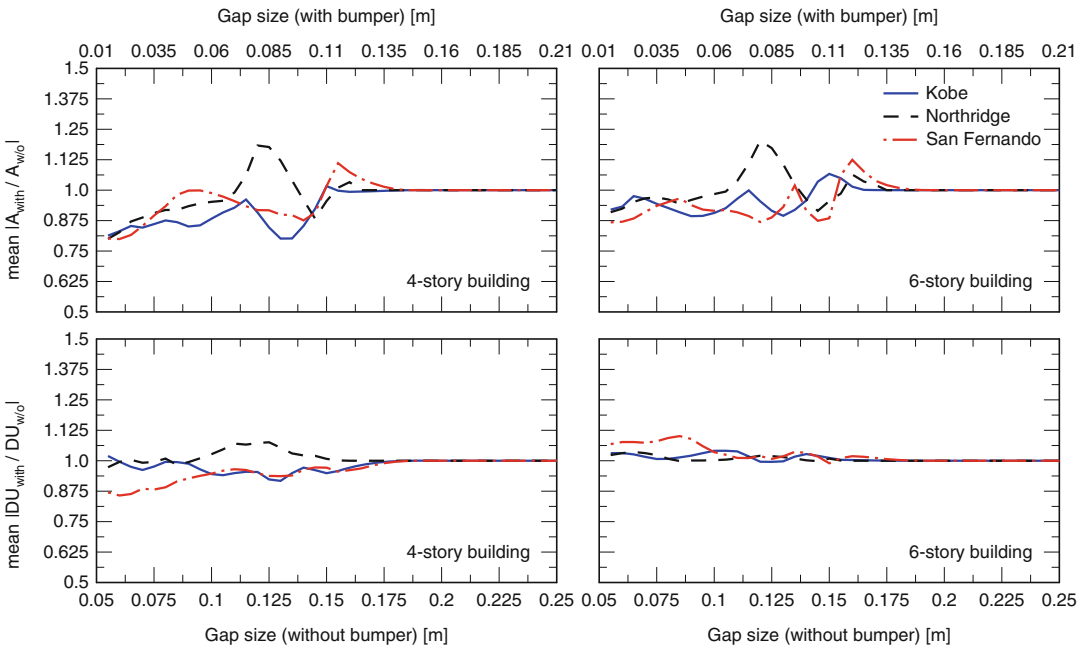
Summary

The current study investigates numerically the possibility of incorporating layers of rubber or other elastomeric material between neighboring



Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 19 Attachment of a rubber shock-absorber in a cavity on the building's wall

structures with relatively narrow seismic gaps in order to act as collision bumpers and mitigate the detrimental effects of potential earthquake-induced pounding. The efficiency of this potential impact mitigation measure is parametrically examined, considering both cases of conventionally fixed-support and seismically isolated buildings, which are subjected to various earthquake



Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding, Fig. 20 Mean values of the amplification of peak responses among all floors of the 4-story and the 6-story buildings, due to the usage of

rubber shock-absorbers with 1 cm cavity (full advantage of bumper effective thickness, i.e. 4 cm), in terms of the width of the seismic gap

excitations. The results indicate that only under certain circumstances the incorporation of rubber bumpers in an existing seismic gap can reduce the amplifications of the peak responses of the structures due to pounding. Specifically, the effectiveness of the bumpers depends on the existing gap size in combination with the earthquake characteristics and the structural properties (e.g., number of stories) of the adjacent buildings. The attachment of the bumper in cavities on the building's wall, taking full advantage of the whole compressible width of the rubber, improves their efficiency.

Cross-References

- Learning from Earthquake Disasters
- Post-Earthquake Diagnosis of Partially Instrumented Building Structures
- Recording Seismic Signals
- Time History Seismic Analysis

References

- Anagnostopoulos SA (1988) Pounding of buildings in series during earthquakes. *Earthq Eng Struct Dyn* 16:443–456
- Anagnostopoulos SA (1995) Earthquake induced poundings: state of the art. In: Duma (ed) 10th European conference on earthquake engineering, Balkema, Rotterdam
- Bertero VV (1987) Observations on Structural Pounding. In: *Proceedings international conference on Mexico earthquakes*, ASCE, 264–278
- Earthquake Engineering Research Institute (EERI) (1990) Loma prieta earthquake reconnaissance report, Rep. No. 90–01. In: Benuska L (ed) EERI, Oakland
- Earthquake Engineering Research Institute (EERI) (2000) Kocaeli, Turkey, earthquake of August 17, 1999 reconnaissance report, Publ. No. 00–03. In: Youd TL, Bardet J-P, Bray JD (ed) EERI, Oakland
- Earthquake Engineering Research Institute (EERI) (2009) L'Aquila, Italy earthquake clearinghouse – Observations from-EERI/PEER team, <http://www.eqclearinghouse.org/italy-090406/>
- Kajita Y, Kitahara T, Nishimoto Y, Otsuka H (2006) Estimation of maximum impact force on natural rubber during collision of two steel bars. In: *First European conference on earthquake engineering and seismology (1st ECEES)*, Geneva, 3–8 Sept

- Kajita Y, Nishimoto Y, Ishikawa N, Watanabe E (2001) Energy absorption capacity of the laminated fiber reinforced rubber installed at girder ends. In: High performance materials in bridges, international conference on high performance materials in bridges, vol 122. doi:10.1061/40691(2003)17, Kona, p 17
- Kawashima K, Shoji G, Koshitoge M, Shimanoe S (2002) Design of an earthquake-resistant expansion joint with unseating prevention system. In: FIB congress, E-282 (CD-ROM), Osaka
- Komodromos P, Polycarpou PC, Papaloizou L, Phocas MC (2007) Response of seismically isolated buildings considering poundings. *Earthq Eng Struct Dyn* 36:1605–1622
- Papadrakakis M, Mouzakis H (1995) Earthquake simulator testing of pounding between adjacent buildings. *Earthq Eng Struct Dyn* 24:811–834
- Polycarpou PC, Komodromos P (2009) Simulating the use of rubber shock absorbers for mitigating poundings of seismically isolated buildings during strong earthquakes. In: 2nd international conference on computational methods in structural dynamics and earthquake engineering (COMPDYN 2009), Rhodes, 22–24 June
- Polycarpou PC, Komodromos P (2010) On the numerical simulation of impacts for the investigation of earthquake-induced pounding of buildings. In: The tenth international conference on computational structures technology (CST2010), Valencia, 14–17 Sept
- Polycarpou PC, Komodromos P, Polycarpou A (2013) A nonlinear impact model for simulating the use of rubber shock absorbers for mitigating the effects of structural pounding during earthquakes. *Earthq Eng Struct Dyn* 42:81–100
- Shim VPW, Yang LM, Lim CT, Law PH (2004) A visco-hyperelastic constitutive model to characterize both tensile and compressive behavior of rubber. *J Appl Polym Sci* 92:523–531
- Warnotte V, Stoica D, Majewski S, Voiculescu M (2007) State of the art in the pounding mitigation techniques. *Intersections/Intersectii* 4(3):102–117, ISSN: 1582–3024

Safety Assessment of Piled Buildings in Liquefiable Soils: Mathematical Tools

Subhamoy Bhattacharya
University of Surrey, Guildford, UK

Synonyms

Case study of pile damage during the 1995 Kobe earthquake; Geotechnical analysis of pile foundation in liquefiable soils; Numerical Modeling; Pile foundations in liquefiable soils; Reliability of pile foundations; Seismic Analysis of Pile Foundations; Soil-Structure Interaction

Introduction

Piles are routinely used as foundations to support short-to-medium span bridges and buildings typically over four stories and other structures. Collapse and/or severe damage of pile-supported structures is still observed in liquefiable soils after most major earthquakes such as the 1995 Kobe earthquake (Japan), the 1999 Kocaeli earthquake (Turkey), and the 2001 Bhuj earthquake (India). The failures not only occurred in laterally spreading (sloping) ground but were also observed in level ground where no lateral spreading would be anticipated. A good discussion on the failure modes can be found in Bhattacharya and Madabhushi (2008).

The failures were often accompanied by settlement and tilting of the superstructure, rendering it either useless or very expensive to rehabilitate after the earthquake. Following the 1995 Kobe earthquake, investigations have been carried out to find the failure pattern of the piles. Piles were excavated or extracted from the subsoil, borehole cameras were used to take photographs, and pile integrity tests were carried out. These studies hinted at the location of the cracks and damage patterns for the piles. Of particular interest is the formation of plastic hinges in the piles. This indicates that the stresses in the pile during and after liquefaction exceeded the yield stress of the material of the pile, despite large factors of safety which were employed in the design. Hinges were found to have occurred at various depths along the pile: at the pile head, at the middle of the liquefiable layer, and toward the interface of liquefiable/non-liquefiable layer. In this context it must be mentioned that piles are currently designed with adequate factor of safety for geotechnical load-carrying capacity (maximum allowable load in the pile) and against bending failure. Bending moment can occur in a pile due to lateral loads arising from: (a) inertia load acting at the pile head and (b) kinematic loads from the ground. Further discussion on the dynamics of the problem can be found in Adhikari and Bhattacharya (2008).

Liquefaction of soil around the pile can affect a pile-supported structure in the following ways:

1. A pile will be laterally unsupported in the zone of liquefaction, i.e., no lateral restraint from the soil to the pile in the liquefied zone. As a result, a pile may be vulnerable to buckling instability if the axial load is high enough and the unsupported length of pile is sufficient.
2. The pile will lose its shaft resistance in the zone of liquefaction, i.e., no liquefied soil-pile friction, and as a result the axial load on the pile in the liquefiable zone will increase.
3. The time period of the structure will change due to liquefaction as the foundation becomes flexible. While calculating the period of a building under non-liquefied condition, it is assumed that foundation is rigid, and as a result only the dimensions of the buildings are adequate to obtain the period. However, as soon as a soil liquefies, the time period may increase.
4. The damping of a pile-supported structure also increases during and after liquefaction.

The above mechanisms were verified experimentally by carefully designed model tests and can be found in Bhattacharya (2003), Bhattacharya et al. (2004), Bhattacharya et al. (2005), and Lombardi and Bhattacharya (2014). Bhattacharya (2003) showed through dynamic centrifuge tests that axial load alone can cause a pile to fail if the surrounding soil liquefies in an earthquake and the mechanism being buckling instability.

Lombardi and Bhattacharya (2014) showed through high-quality shaking table tests that the time period of pile-supported structures will increase owing to liquefaction. In the experiments, the soil was liquefied progressively through a broadband white noise signal with increasing magnitude. Furthermore, they also showed that the overall damping ratio of the structures may increase in excess of 20 %. These have important design implications.

All current design methods, such as JRA (1996), NEHRP (2000), IS 1893 (2001), and Eurocode 8 (CEN 2004), focused on bending strength of the pile to avoid bending failure due to lateral loads (combination of inertia and lateral spreading). In contrast to these conventional

design codes, which advocate bending mechanism as the main design consideration, recent research showed that an axially loaded pile can be laterally unsupported in liquefied soils and is susceptible to buckling failure.

Buckling instability under the interaction of axial and lateral loads can be a more critical design consideration because of its sudden nature and sensitivity to imperfection; see, for example, Dash et al. (2010). More recently, Bhattacharya et al. (2009) included the effects of dynamics on the combined axial and lateral loads on a pile foundation. Essentially, piles in liquefied soils may be better regarded as columns carrying lateral loads rather than laterally loaded beams.

In design, beam bending and column buckling are approached differently. Bending is a stable mechanism as long as the pile remains elastic and secondary failure (e.g., local buckling) is not a possibility. This failure mode depends on the bending strength (e.g., yield moment capacity and plastic moment capacity) of the member under consideration. In contrast, buckling is an unstable mechanism and it occurs suddenly and drastically when the elastic critical load is reached. It is the most destructive mode of failure and depends on the geometrical properties of the member, i.e., slenderness ratio, rather than the member strength. For example, steel pipe piles with identical length and diameter but having different yield strengths (e.g., 200 MPa, 500 MPa, and 1,000 MPa) will buckle at similar axial loads but can resist different amounts of bending. In other words, bending failure may be avoided by increasing the yield strength of the material, but it may not suffice to avoid buckling. To prevent buckling failure, there should be a minimum pile diameter depending on the depth of liquefiable soils. Therefore, designing against bending would not automatically satisfy buckling requirements. It is envisaged that there are plenty of existing pile-supported structures that may need retrofitting.

This entry therefore describes a probabilistic and a deterministic method to assess the reliability of pile foundation for a scenario earthquake. Before the methodology is described, a discussion is presented on the main loading on pile foundations during earthquakes in liquefiable soils.

Methodology

Different Stages of Loading on a Pile-Supported Structure During Earthquakes

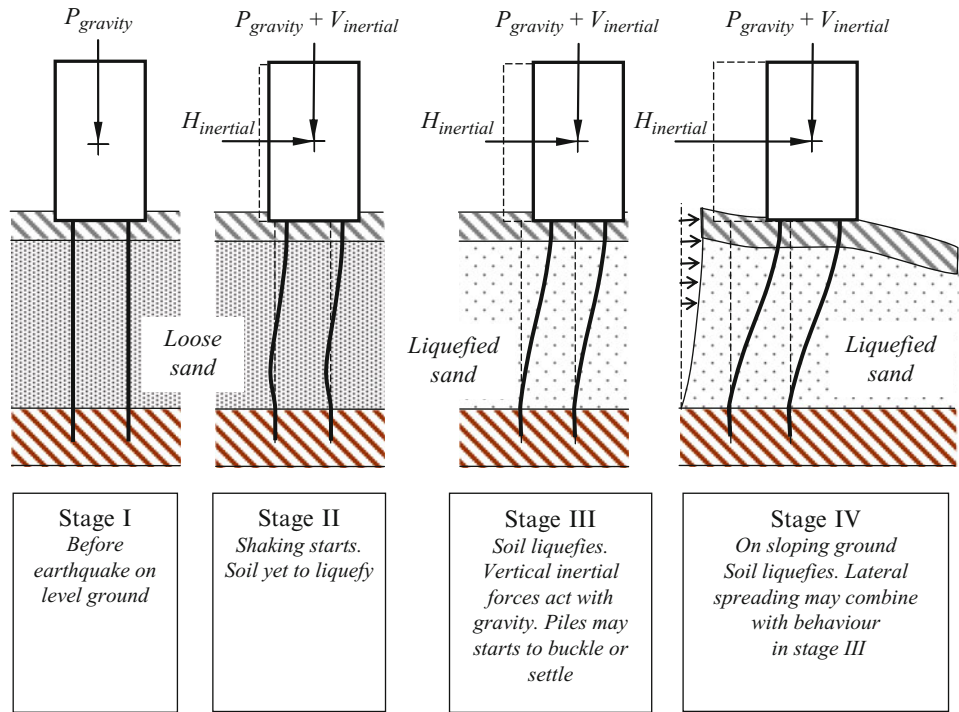
Figure 1 shows the different stages of loading of a pile-supported structure during a seismic liquefaction-induced event. $P_{gravity}$ (Stage I) represents the axial load on the piles in normal condition. This can be estimated based on static equilibrium. This axial compressive load may increase/decrease further by $V_{inertial}$ due to inertial effect of the superstructure and kinematic effects of the soil flow past the foundation. $H_{inertial}$ is the inertial lateral loads due to the oscillation of superstructure (Stage II and III). Ground movement causes kinematic loads on the pile foundations. This load can be of two types: transient (during shaking, due to the dynamic effects of the soil mass) and residual (after the shaking ceased due to soil flow, often known as “lateral spreading”) (Stage IV). The various forms of feasible failure mechanisms of pile foundations are shear failure, bending failure, buckling instability, and dynamic failure.

Formulation

Bhattacharya (2006) discusses the deterministic approach to determine the factor of safety of pile foundations against the buckling instability failure. Bhattacharya and Goda (2013) developed a probabilistic procedure for determining the occurrence of a buckling failure of existing piled foundations due to a scenario earthquake.

The methodology is based on assessing two length parameters:

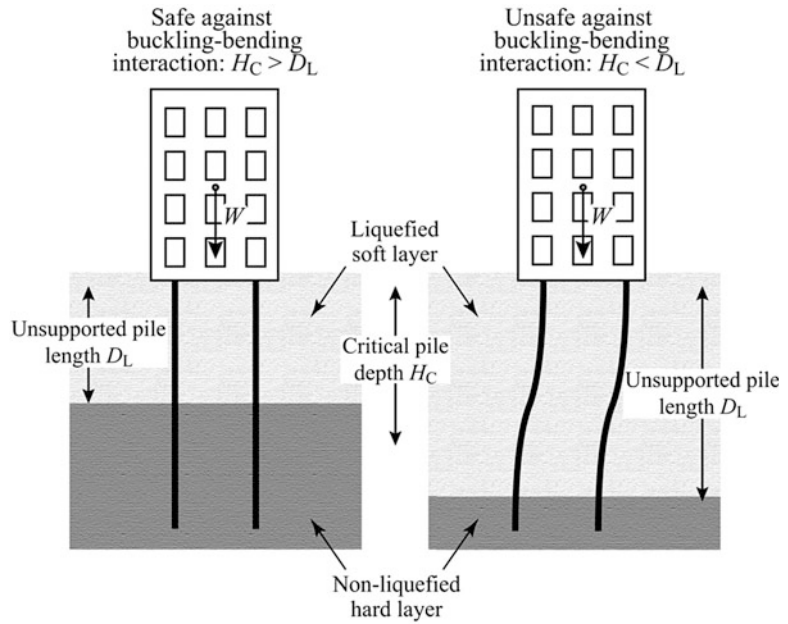
1. Critical length of a pile denoted (H_C). Essentially this is the unsupported length of the pile that can sustain without collapse due to combined axial and lateral loading. This can be estimated based on manipulation of the Euler’s buckling load equation considering the correct boundary condition of the pile. Specifically H_C depends on the type and dimension of superstructure (bridge or building), bending stiffness, axial load acting on the pile, dynamic characteristics of



Safety Assessment of Piled Buildings in Liquefiable Soils: Mathematical Tools, Fig. 1 Different stages of loading and failure mechanism of pile during earthquake

Safety Assessment of Piled Buildings in Liquefiable Soils: Mathematical Tools,

Fig. 2 Concept of critical length (H_C) and unsupported length (D_L)



superstructure, and boundary conditions of the pile at the top and bottom of the liquefiable layer.

- Length of unsupported pile due to liquefaction (D_L): this can be obtained based on the depth of liquefaction due to the design earthquake.

Figure 2 shows the concept of the two lengths. Comparing H_C with D_L , potential failure of the underground pile due to buckling is predicted when $H_C < D_L$ (see Figs. 1 and 2). In reliability terms, H_C is the capacity variable, while D_L is the demand term, and thus the failure criterion can be regarded as the limit state function:

$$g = (H_C - D_L) \quad (1)$$

Probabilistic and Deterministic Approach

In a deterministic approach, we obtain a single value of the two parameters and can obtain the factor of safety against failure. However, the advantages of assessing using probabilistic approach are (i) the outcome is expressed as a (estimated) likelihood of failure by taking into account various sources of uncertainty involved in the assessment; (ii) sensitivity analysis can be conducted to identify key factors that affect the outcome (the results are useful to improve the

methodology and tool); and (iii) the probabilistic framework provides a straightforward way to integrate geotechnical assessment tools into existing probabilistic seismic hazard methods as well as decision-support tools for implementing earthquake risk mitigation measures (Fig. 3).

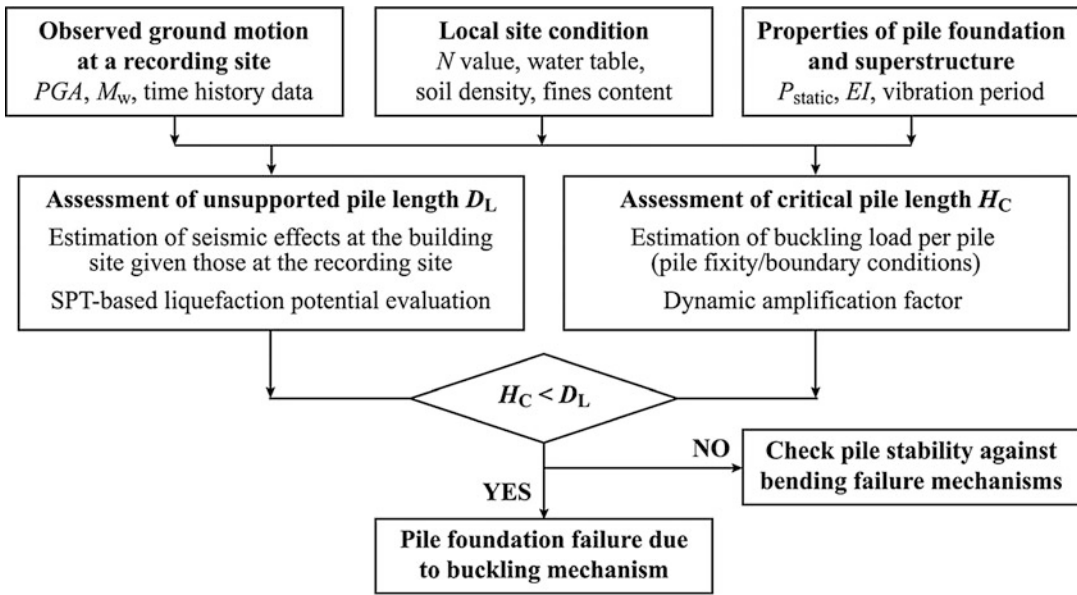
Step-by-Step Methodology

The steps are:

- Estimation of peak ground acceleration (PGA) at building/bridge site:** The assessment of D_L requires (i) estimation of ground motion parameters (typically, peak ground acceleration (PGA)) at a building site due to a specified scenario and (ii) assessment of liquefaction initiation potential at different depths along the pile length. Equation 2 is a typical equation to obtain PGA.

$$\text{Log } PGA_{\text{bldg}} = f(M_w, R_{\text{bldg}}, V_{S30, \text{bldg}}) + \varepsilon_{\text{bldg}} \quad (2)$$

where PGA_{bldg} is PGA value at the building site, M_w is the moment magnitude, R is the distance measured (typically closest distance to fault rupture plane), V_{S30} is the average shear-wave velocity in the uppermost 30 m in (m/s), and $\varepsilon_{\text{bldg}}$ is the random error that is



Safety Assessment of Piled Buildings in Liquefiable Soils: Mathematical Tools, Fig. 3 Schematic diagram showing the methodology

modeled as a normal variable with zero mean and logarithmic standard deviation.

(b) **Assessment of the depth of liquefiable soil:**

The next step is to conduct probabilistic liquefaction potential evaluation of a soil column (along a pile) at the building site to estimate D_L for a given seismic excitation level. Such assessment is often conducted by using simplified stress-based methods of Seed and Idriss (1971) based on standard penetration test (SPT) data, cone penetration test (CPT) data, and shear-wave velocity (V_S) data. Recently, probabilistic procedures for liquefaction initiation have been developed by considering different in situ measures for describing soil strength. Using a sophisticated Bayesian regression analysis and well-screened case

studies, Cetin et al. (2002, 2004) developed a statistical model for calculating the probability of liquefaction initiation based on SPT data, while Moss et al. (2006) developed a counterpart using CPT data. Using the first-order reliability method, Juang et al. (2005) developed a similar model based on V_S data. The significance of these developed models is that key uncertainties associated with the input data/parameters and the adopted models themselves are taken into account; the developed models can produce unbiased potential of liquefaction initiation and are useful to conduct probabilistic liquefaction hazard analysis (Goda et al. 2011). The probability of liquefaction initiation P_L at a depth of interest can be estimated as

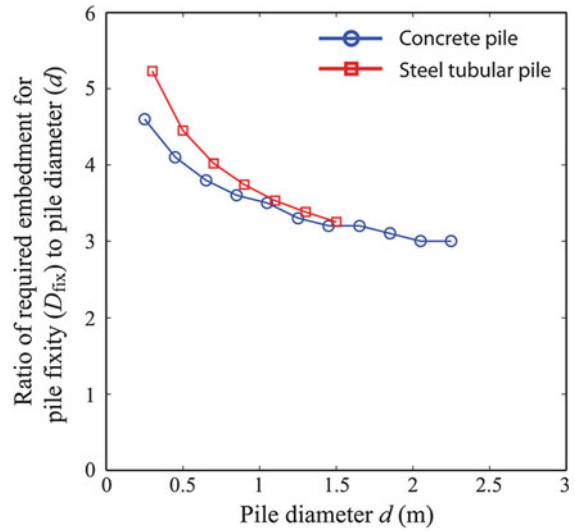
$$P_L = \Phi \left(-\frac{N_{1,60}(1 + 0.004FC) - 13.32\ln(CSR_{eq}) - 29.53\ln(M_w) - 3.70\ln(\sigma'_v/P_a) + 0.05FC + 16.85}{2.7} \right) \quad (3)$$

where Φ is the standard normal function, $N_{1,60}$ is the corrected SPT counts (but not

adjusted for fines content), FC is the fines content (in percentage), CSR_{eq} is the cyclic

**Safety Assessment of
Piled Buildings in
Liquefiable Soils:
Mathematical Tools,**

Fig. 4 Depth required for fixity a pile



stress ratio (but not adjusted for the moment magnitude), and σ'_v is the vertical effective stress.

- (c) **Unsupported length of the pile (D_L):** The next step is to determine the unsupported pile length (D_L) based on the liquefaction profile. D_L is equal to the thickness of liquefied soil layers plus some additional length necessary for fixity at the bottom of the liquefied soils. Typical calculations show that the fixity depth is about three to five times the diameter of the pile (see Fig. 4); see Bhattacharya and Goda (2013). If a relatively thin non-liquefied layer is overlain and underlain by thick liquefied layers, lateral restraint of the pile at the non-liquefied layer might not be expected, i.e., the pile is unsupported. In such a case, the thin non-liquefied layer can be ignored in determining D_L , and the unsupported length needs to be extended until a thick non-liquefied layer is encountered. However, if there is a complex soil profile with alternating liquefiable and non-liquefiable soil layers, more detailed analysis is required.
- (d) **Assessment of the critical pile length H_C :** This section describes the mathematical background behind the rationale of obtaining the critical pile depth H_C . Before the onset of shaking, the static axial load P_{static} acts on

each pile beneath a building, assuming that each pile is equally loaded during static condition and neglecting any eccentricity of loading. During an earthquake, inertial action of the superstructure imposes the dynamic axial load on the piles, which will increase the axial load on some piles. These piles with increased axial loads may be vulnerable to buckling. An estimate of the maximum axial compressive load acting on a pile can be given by

$$P_{dynamic} = (1 + \alpha)P_{static} \quad (4)$$

where α is termed as the dynamic axial load factor and is a function of type of superstructure, height of the center of mass of the superstructure, and characteristics of the earthquake shaking (e.g., frequency content and amplitude).

For buckling instability analysis, each pile needs to be evaluated with respect to its end conditions, i.e., fixed, pinned, or free. Each pile in a group of identical piles will have the same buckling resistance as a single pile. If a group of piles is fixed in a stiff pile cap and embedded sufficiently at the tip, as in Fig. 2, the pile group will buckle in side sway.

The elastic critical load of a single pile, P_{cr} , can be estimated as

Safety Assessment of Piled Buildings in Liquefiable Soils: Mathematical Tools, Table 1 Values of K in Eq. 7

Boundary condition of the pile at the top and bottom of the liquefied layer		Effective length	K	Examples
Top	Bottom			
Fixed	Fixed [sufficient embedment at the dense layer]	$L_{\text{eff}} = 0.5H_c$	0.5	Pile groups with raked piles
Free to translate but restrained against rotation – sway frame	Pinned [insufficient embedment at the dense layer]	$L_{\text{eff}} = 2H_c$	2	See, for example, NFCH (Niigata Family Court House) building in Bhattacharya (2003)
Free to translate but restrained against rotation – sway frame	Fixed [sufficient embedment at the dense layer]	$L_{\text{eff}} = H_c$	1	Most cases fall under such category; see, for example, Fig. 2
Fixed in direction but free to rotate	Fixed [sufficient embedment at the dense layer]	$L_{\text{eff}} = 0.7H_c$	0.7	Pile groups with raked piles. Improper pile-pile cap connection
Fixed in direction but free to rotate	Pinned [less embedment at the dense layer]	$L_{\text{eff}} = H_c$	1	Pile groups with raked piles. Improper pile-pile cap connection
Free, i.e., unrestrained against rotation and displacement	Fixed [sufficient embedment at the dense layer]	$L_{\text{eff}} = 2H_c$	2	Piles in a row such as the Showa Bridge piles

$$P_{cr} = \frac{\pi^2 EI}{(L_{\text{eff}})^2} \quad (4)$$

where L_{eff} is the effective length, i.e., Euler's equivalent buckling length of a strut pinned at both ends, and EI is the bending stiffness of the pile. The effective length of the pile (L_{eff}) can be found in Table 1 or from any structural mechanics textbook or codes of practice. The unsupported length of the pile D_L is equal to the thickness of liquefiable soil plus some additional length necessary for fixity at the bottom of the liquefiable soil.

The applicability of the elastic critical load, as in Eq. 4, to pile buckling failure is an important factor. Experiments show that the actual failure load of a slender column is much lower than that predicted by Eq. 4. Rankine (1866) recognized that the actual failure involves an interaction between elastic and plastic modes of failure. Lateral loads and inevitable geometrical imperfection lead to creation of bending moments in addition to axial loads. Bending moments have to be accompanied by stress resultants that diminish the cross-sectional area available for carrying the axial load; thus the actual failure load is likely to be less than the elastic critical load,

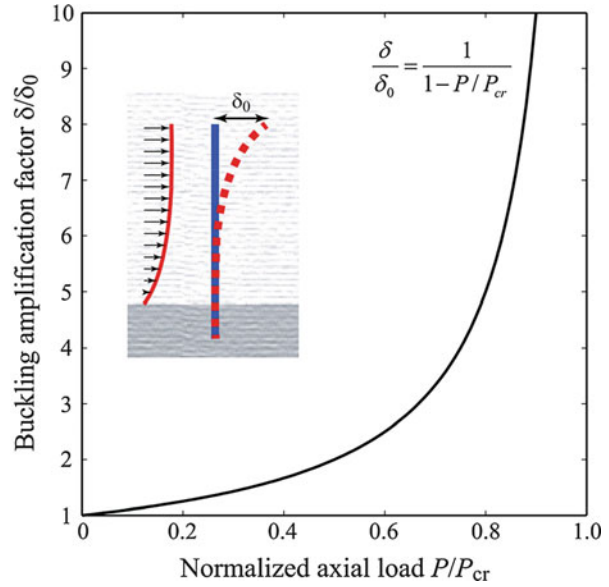
i.e., $P_{\text{failure}} < P_{cr}$. Equally, the growth of plastic bending zones reduces the effective elastic modulus of the section, thereby resulting in the decreased critical load for buckling (i.e., capacity). Furthermore, these processes feed each other, and as elastic critical loads are approached, all bending effects are magnified. Stability analysis of elastic columns showed that if lateral loads in the absence of axial load would create a maximum lateral displacement δ_0 in the critical mode shape of buckling, then the displacement δ under the same lateral loads but with the concurrent axial load P is given by

$$\frac{\delta}{\delta_0} = \frac{1}{1 - P/P_{cr}} \quad (5)$$

The term δ/δ_0 is the buckling amplification factor (i.e., amplification of lateral displacement due to the presence of the axial load). Figure 5 presents a graph of the buckling amplification factor plotted against the normalized axial load P/P_{cr} , where P denotes the applied axial load. It can be observed from Fig. 5 and Eq. 6 that if the applied load is 50 % of P_{cr} , the amplification of lateral deflection due to lateral loads is about two times. At these large deflections, secondary moments

Safety Assessment of Piled Buildings in Liquefiable Soils: Mathematical Tools,

Fig. 5 Buckling amplification factor versus normalized axial load



will be generated due to P - Δ moment, leading to more deflection. It is therefore important to remain in the linear regime and not in anyway near the asymptotic region, where the buckling amplification factor increases dramatically (e.g., $P/P_{cr} > 0.6$).

Moreover, it would also be unwise to use a factor of safety less than three against the Euler load of a pile, i.e., ($P/P_{cr} = 0.33$). Such consideration is consistent with general design practice where structural engineers use a factor of safety of at least three against linear elastic buckling to take into account the eccentricity of load, deterioration of elastic stiffness due to plastic yielding, and unavoidable imperfection. The actual failure load $P_{failure}$ is therefore some factor ϕ ($\phi < 1$) times the theoretical Euler's buckling load given by Eq. 5.

$$P_{failure} = \phi P_{cr} \quad (6)$$

Based on the above discussion, it may be inferred that buckling instability is initiated at around 0.35, i.e., $\phi = 0.35$. It is noted that in reality, this factor depends on the axial load, imperfection of piles, and residual stress in the pile due to driving. The selection of ϕ is one of the significant sources of uncertainty in determining the critical pile depth H_C .

Determination of "Critical Depth" H_C

In the limit state condition of failure, $P_{dynamic} = P_{failure}$. For the type of structure shown in Fig. 2, $L_{eff} = H_C$ in Eq. 4. In order to generalize the boundary condition of the pile (i.e., pile head fixity condition with pile cap/superstructure and the fixity at the interface between liquefiable and non-liquefiable layers at deeper depths), one may write Eq. 7:

$$L_{eff} = K \times H_C \quad (7)$$

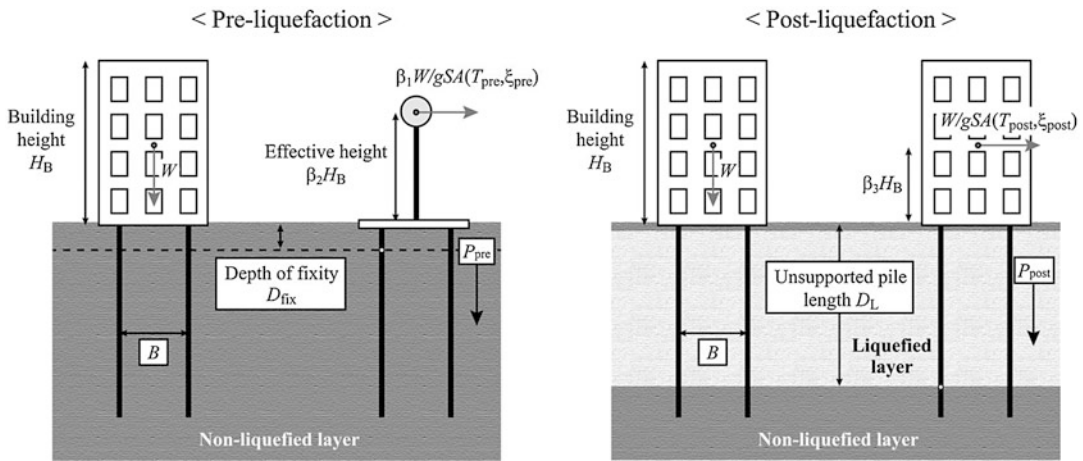
where K is the column effective length factor (e.g., $K = 1$ for free lateral translation but restrained against rotation-sway motion). Values of K for other boundary conditions of the pile are given in Table 1.

With the abovementioned assumptions, Eq. 6 can be rewritten as

$$P_{dynamic} = \phi P_{cr} = \frac{\phi \pi^2 EI}{K^2 H_C^2} \quad (8)$$

Rearranging Eq. 8 gives the estimate of the critical depth H_C for a pile as follows:

$$H_C = \sqrt{\frac{\phi \pi^2 EI}{K^2 P_{dynamic}}} = \sqrt{\frac{\phi \pi^2 EI}{K^2 (1 + \alpha) P_{static}}} \quad (9)$$



Safety Assessment of Piled Buildings in Liquefiable Soils: Mathematical Tools, Fig. 6 Loading condition of pile foundation in pre-liquefaction and post-liquefaction stages

Therefore, the assessment of H_C is based on the calculation of the critical buckling load for a pile foundation surrounded by liquefied soils (from which the pile cannot receive sufficient lateral support). Key parameters in Eq. 9 are ϕ , K , and α . Reasonable values of ϕ and K can be selected based on engineering judgment.

Based on the procedure described above, both D_L and H_C can be assessed probabilistically. The occurrence of pile foundation failure due to buckling mechanism is indicated if $H_C < D_L$. By sampling D_L and H_C many times, probabilistic assessment of liquefaction-induced pile foundation failure due to buckling mechanism can be carried out.

Formulation for Determination of Dynamic Load Amplification Factor (α)

One of the critical factors is the assessment of α . This can be facilitated by examining axial forces acting on a pile in pre-liquefaction and post-liquefaction situations as shown in Fig. 6. In a pre-liquefaction stage, the plane of fixity of a building with pile foundation surrounded by non-liquefied soils can be estimated by following a similar procedure as shown in Fig. 4, i.e., depth from ground surface to the plane of fixity is denoted by D_{fix} in Fig. 6 and is typically a few meters. By assuming that the natural vibration period of the building before liquefaction is T_{pre} ,

the maximum axial force acting on a pile due to inertia can be calculated as

$$P_{pre} = \frac{\beta_1 (W/g) \times SA(T_{pre}, \xi_{pre}) \times (D_{fix} + \beta_2 H_B)}{(N_p/2) \times B} \quad (10)$$

where W is the total weight of a building (note: $W = P_{static} N_p$), $SA(T_{pre}, \xi_{pre})$ (g) is the spectral acceleration at T_{pre} with damping ratio ξ_{pre} (typically, 2–5 %), β_1 is the coefficient to account for modal mass for the fundamental vibration mode (typically, 0.8–0.9), β_2 is the coefficient to account for the effective height where the inertia due to the modal mass acts (typically, 0.65–0.75), H_B is the height of building, N_p is the number of piles (assuming that an equal number of piles are positioned in two rows), and B (m) is the foundation width between the two rows of piles (note: the foundation width B is along the direction where axial force is induced by overturning moment due to lateral inertia). T_{pre} can be estimated by using an empirical equation, such as $T_{pre} = 0.09 H_B / B^{0.5}$; see Anderson et al. (1952) also adopted in IS 1893.

In a post-liquefaction situation, the building is supported by piles that have relatively long unsupported lengths of D_L (as evaluated based on the liquefaction initiation analysis). In this

case, the natural vibration period in a post-liquefaction stage, T_{post} , can be calculated as

$$T_{\text{post}} = 2\pi \sqrt{\frac{W/g}{N_p \times 12EI/D_L^3}} \quad (11)$$

where $12EI/(D_L)^3$ is the lateral stiffness of each pile. Then, the maximum axial force acting on a pile is given by

$$P_{\text{post}} = \frac{(W/g) \times SA(T_{\text{post}}, \xi_{\text{post}}) \times (D_L + \beta_3 H_B)}{(N_p/2) \times B} \quad (12)$$

where $SA(T_{\text{post}}, \xi_{\text{post}})$ (g) is the spectral acceleration at T_{post} with damping ratio ξ_{post} (typically, 10–30 % representing the damping of liquefied soil) and β_3 is the coefficient to account for the effective height where the inertia acts in a post-liquefaction condition (typically, 0.5). The underlying assumption of Eqs. 11 and 12 is that the building is considered as a rigid mass and the pile provides the primary lateral stiffness to the building. It must be mentioned that this approach is very simple, and many uncertainties, such as effects of vertical inertia, timing of max inertia and timing of loss of lateral support due to liquefaction, and potential effects due to lateral spreading, are not taken into consideration.

By defining $\alpha = \max(P_{\text{pre}}/P_{\text{static}}, P_{\text{post}}/P_{\text{static}})$, α is given by:

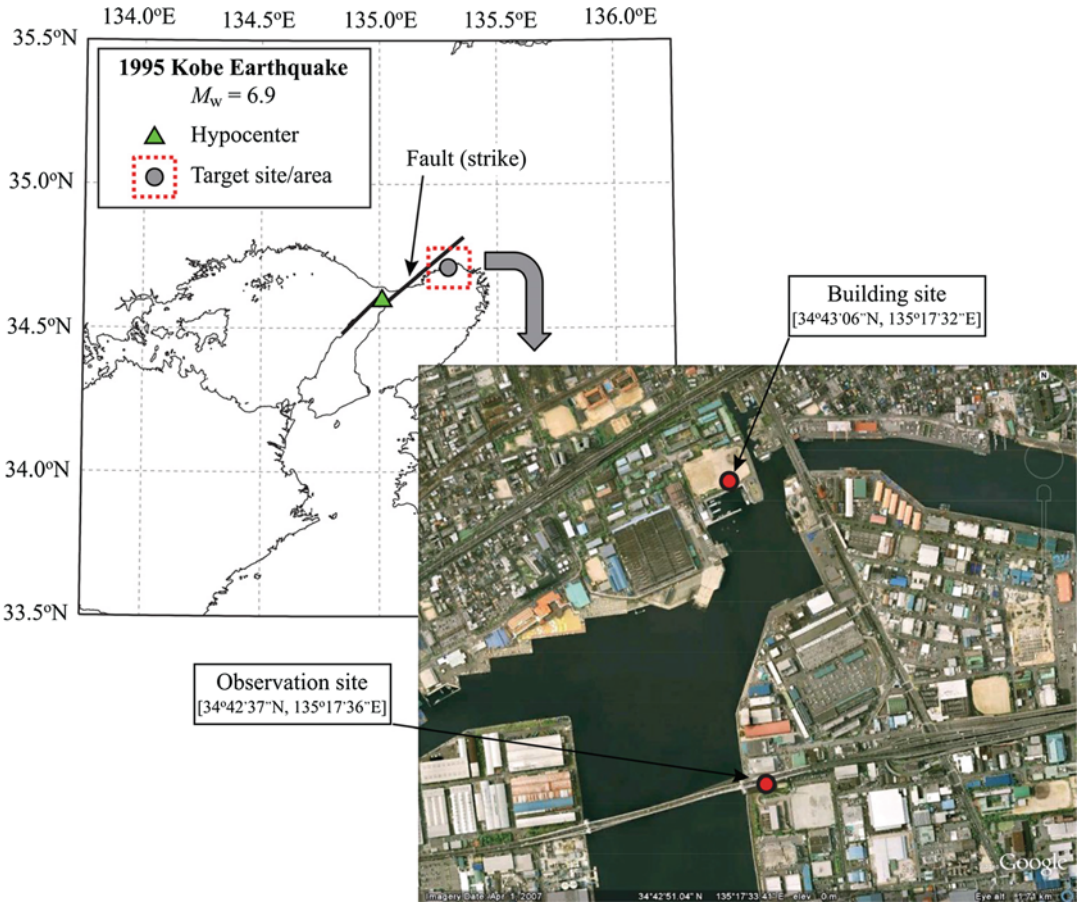
$$\alpha = \frac{2}{gB} \max[\beta_1 SA(T_{\text{pre}}, \xi_{\text{pre}})(D_{\text{fix}} + \beta_2 H_B), SA(T_{\text{post}}, \xi_{\text{post}})(D_L + \beta_3 H_B)] \quad (13)$$

The additional information needed for evaluating α is the estimated spectral acceleration values at vibration periods T_{pre} and T_{post} with damping ratios ξ_{pre} and ξ_{post} (note: T_{pre} and T_{post} are random variables; in particular, T_{post} is significantly affected by liquefaction initiation analysis).

A case study: collapse of a 5-story RCC building in Kobe during the 1995 Kobe earthquake:

Several buildings with pile foundation had to be demolished due to severe liquefaction damage during the 1995 M_w 6.9 Kobe earthquake. Because of the dramatic consequences, several detailed post-earthquake investigations were conducted to examine the cause and failure mechanism of these cases (Tokimatsu et al. 1997). For illustration, one of such case studies (Uzuoka et al. 2002; Bhattacharya 2006) is focused upon. A 5-story reinforced concrete frame building (total height $H_B = 14.5$ m) was located at 6 m from the quay wall on a reclaimed fill in the Higashinada area of the Kobe City; the distance from the rupture plane to the building site was about 5 km (Fig. 7). The Kobe earthquake caused a lateral displacement of 2 m to the quay wall toward the sea, and the building was tilted by 3° due to lateral spreading. The schematics of the post-earthquake investigation of the building and pile foundation are shown in Fig. 8. At the building site, significant lateral spreading was observed (about 1.0–1.5 deformation/movement of the ground toward the sea; Tokimatsu et al. 1997).

The building was supported on 38 hollow pre-stressed concrete piles (there were two pile rows separated by 7.5 m and 19 piles were aligned in each row); the pile length was 20 m with exterior and interior diameters of 0.4 and 0.24 m, respectively. Figure 9 shows variations of soil profiles and SPT N counts with depth at the building site. The site has fill/sand layers with relatively low N counts (e.g., 2–9 m and 12–16 m), which are susceptible to liquefaction (i.e., saturated sand layers with low strength); the average shear-wave velocities in the uppermost 12 and 30 m are estimated to be about 147 and 216 m/s, respectively. The water table level was about 2 m below ground surface. The post-earthquake investigation by Tokimatsu et al. (1997) indicated that soil layers shallower than 9 m were liquefied (based on a simplified stress method). Moreover, the borehole logging data shown in Fig. 8 suggest that the sandy silt



Safety Assessment of Piled Buildings in Liquefiable Soils: Mathematical Tools, Fig. 7 Location of the building site and observation site

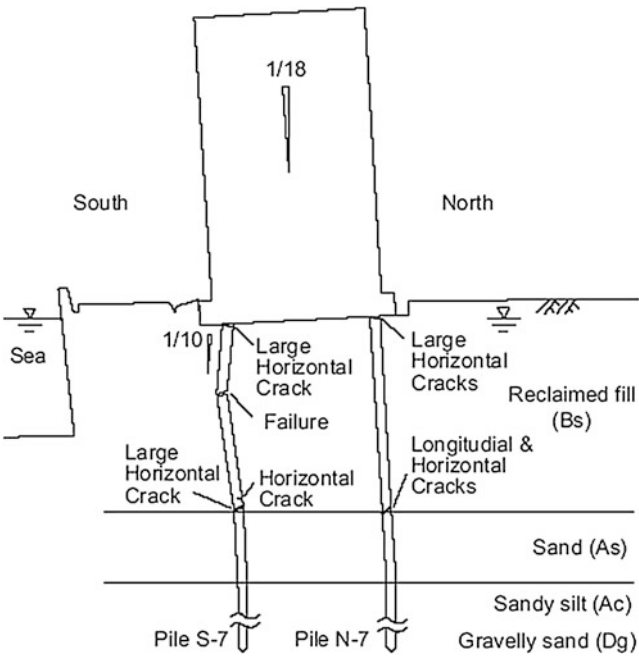
layer between 12 m and 16 m, having low strength, may be disturbed by strong ground motion; thus this layer may not offer much fixity to the pile.

The nearby ground motion recording was obtained at the Higashi Kobe Bridge, which is away about 0.9 km from the building site; distance to the rupture plane for the observation and building sites is about 5.2 and 4.4 km, respectively. The site condition at the observation site is similar to that at the building site (typically, NEHRP site class D or E). The recorded acceleration time histories at the observation site are shown in Fig. 9 (Public Works Research Institute 1995).

The 5 % damped response spectra of the two horizontal components and their geometric mean are presented in Fig. 9 and are compared with the median GMPE (Ground Motion Prediction Equations) by Zhao et al. (2006); this relation is used as a representative regional model throughout this study to estimate ground motion parameters at the building site. The comparison of the calculated response spectra with the Zhao et al. relation indicates that the observed response spectra have less spectral content at vibration periods less than 1.0 s, while they contain rich spectral content at vibration periods greater than 1.0 s. The responses at short vibration periods are likely

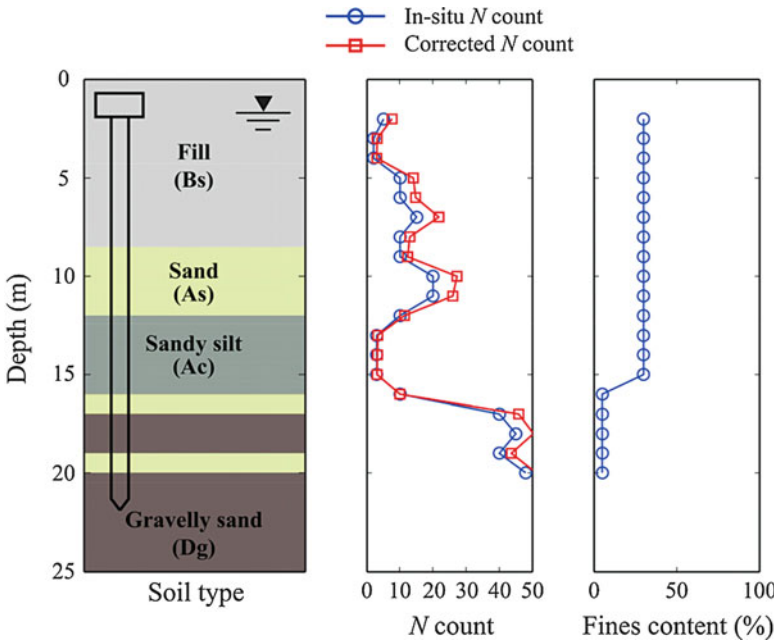
Safety Assessment of Piled Buildings in Liquefiable Soils: Mathematical Tools,

Fig. 8 Post-earthquake investigation of a case study (Tokimatsu et al. 1997)



Safety Assessment of Piled Buildings in Liquefiable Soils: Mathematical Tools,

Fig. 9 Boring log of the soil at the building site



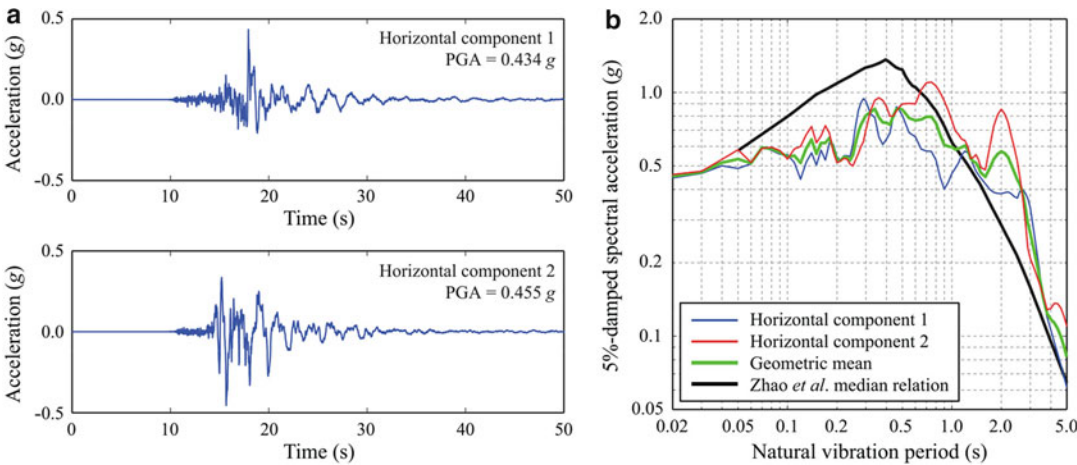
to be affected by nonlinear site amplification (i.e., de-amplification), and at such a site, liquefaction-induced ground failure may be expected. Further to note, geographical positions of the fault plane,

hypocenter, and the observation site (Fig. 6) are of typical “near-fault motions” due to forward directivity (Mavroeidis and Papageorgiou 2003); this can be corroborated by large response spectra

Safety Assessment of Piled Buildings in Liquefiable Soils: Mathematical Tools, Table 2 Summary of probabilistic information of input parameters

Parameter	Mean	Coefficient of variation	[Lower and upper limits]	Distribution type
Moment magnitude M_w	6.9	0.1 ^a	[6.6, 7.2]	Normal
Peak ground acceleration PGA (g)	Equation 2		–	Lognormal
Water table level (m)	2.0	–	[1.0, 3.0]	Uniform
FC (%)	Fig. 9	0.1	–	Lognormal
N count	Fig. 9	0.15	–	Lognormal
Vertical total stress σ_v (Pa)	– ^{b,c}	0.1	–	Lognormal
Vertical effective stress σ'_v (Pa)	– ^{b,c}	0.15	–	Lognormal
Pre-liquefaction period T_{pre} (s)	0.5	0.1	–	Lognormal
EI of a pile (MNm^2)	32.35	0.1	[24.26, 48.53]	Lognormal
Static axial force per pile P_{static} (kN)	412	0.1	[309, 618]	Lognormal

^aThis is the standard deviation
^bIt depends on the water table
^cDry and wet soil densities are set to 1.76 and 1.92 g/cm³



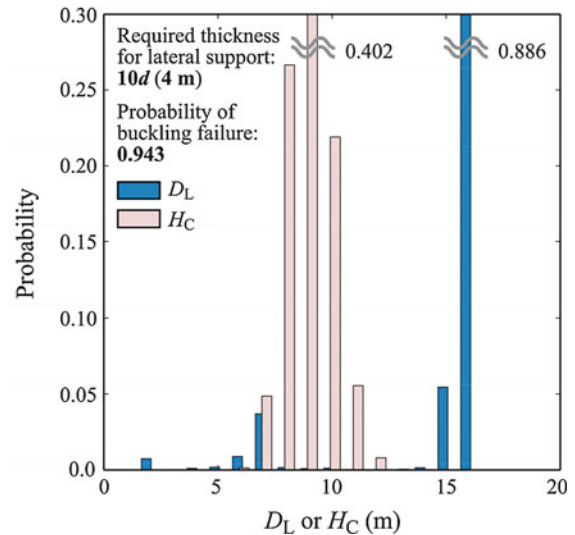
Safety Assessment of Piled Buildings in Liquefiable Soils: Mathematical Tools, Fig. 10 Ground motion time history (a) and 5 % damped response spectra (b) at the Higashi Kobe Bridge observation site

values at long periods (Fig. 9b) and by inspecting velocity time histories of the acceleration data where large velocity pulses are clearly visible. Table 2 provides a summary of the input parameters used in the analysis, and Fig. 11 shows the probability density function

for D_L and H_C which shows the high likelihood that the building collapsed due to buckling of the piles (probability of buckling failure is 0.943). Further details of this methodology can be found in Bhattacharya and Goda (2013) (Fig. 10).

Safety Assessment of Piled Buildings in Liquefiable Soils: Mathematical Tools,

Fig. 11 Comparison of probability mass functions for D_L and H_C



Summary

A probabilistic-based method to reassess such safety has been formulated in this entry. This method can easily be coded in a program, such as MATLAB or Fortran. This method checks the stability of the foundation against buckling instability at full liquefaction, i.e., when the soil surrounding the pile is at its lowest possible stiffness. Two parameters, namely, “critical depth of the pile H_C ” and the “unsupported length of the pile due to liquefaction D_L ,” are estimated. Critical depth is a function of axial load acting on the pile (P), flexural stiffness of the pile (EI), and the boundary condition of the pile above and below the liquefiable soil. On the other hand, D_L mainly depends on the earthquake characteristics, soil profile, and ground conditions. A case study is considered to illustrate an application of the methodology.

Cross-References

- [Damage to Buildings: Modeling](#)
- [Geotechnical Earthquake Engineering: Damage Mechanism Observed](#)
- [Seismic Collapse Assessment](#)

- [Seismic Reliability Assessment, Alternative methods for](#)
- [Seismic Vulnerability Assessment: Reinforced Concrete Structures](#)

References

- Adhikari S, Bhattacharya S (2008) Dynamic instability of pile-supported structures in liquefiable soils during earthquakes. *Shock and Vibration* 16(6):665–685
- Anderson AW, Blume JA, Degenkolb HJ, Hammill HB, Knapik EM, Marchand HL, Powers HC, Rinne JE, Sedgwick GA, Sjöberg HO (1952) Lateral forces of earthquake and wind. *Trans ASCE* 117:716–80
- Bhattacharya S (2003) Pile instability during earthquake liquefaction. Ph.D Thesis, University of Cambridge, U.K.
- Bhattacharya S (2006) Safety assessment of existing piled foundations in liquefiable soils against buckling instability. *ISCT J Earthquake Technol* 43:133–147
- Bhattacharya S, Goda K (2013) Probabilistic buckling analysis of axially loaded piles in liquefiable soils. *Soil Dyn Earthquake Eng* 45:13–24
- Bhattacharya S, Madabhushi SPG (2008) A critical review of methods for pile design in seismically liquefiable soils. *Bull Earthquake Eng* 6:407–446
- Bhattacharya S, Madabhushi SPG, Bolton MD (2004) An alternative mechanism of pile failure in liquefiable deposits during earthquakes. *Geotechnique* 54:203–213
- Bhattacharya S, Madabhushi SPG, Bolton MD (2005) Reply to the two discussions on the paper “An alternative mechanism of pile failure in liquefiable deposits during earthquakes”. *Geotechnique* 55(3):259–263

- Bhattacharya S, Adhikari S, Alexander NA (2009) A simplified method for unified buckling and free vibration analysis of pile-supported structures in seismically liquefiable soils. *Soil Dyn Earthquake Eng* 29:1220–1235
- CEN (2004) Eurocode 8, design of structures for earthquake resistance – Part 1: general rules, seismic actions and rules for buildings, EN 1998-1:2004. Comite Europeen de Normalisation, Brussels, Belgium
- Cetin KO, Der Kiureghian A, Seed RB (2002) Probabilistic model for the initiation of seismic soil liquefaction. *Struct Safety* 24:67–82
- Cetin KO, Seed RB, Der Kiureghian A, Tokimatsu K, Harder LF Jr, Kayen RE, Moss RES (2004) Standard Penetration Test-based probabilistic and deterministic assessment of seismic soil liquefaction potential. *J Geotech Geoenviron Eng* 130:1314–1340
- Dash SR, Bhattacharya S, Blakeborough A (2010) Bending-buckling interaction as a failure mechanism of piles in liquefiable soils. *Soil Dyn Earthquake Eng* 30:32–39
- Goda K, Atkinson GM, Hunter JA, Crow C, Motazedian D (2011) Probabilistic liquefaction hazard analysis for four Canadian cities. *Bull Seismol Soc Am* 101:190–201
- IS 1893 (2001) Indian Standard for Seismic Design
- JRA (1996) Japanese Road Association, Specification for Highway Bridges, Part V, Seismic Design
- Juang CH, Yang SH, Yuan H (2005) Model uncertainty of shear wave velocity-based method for liquefaction potential evaluation. *J Geotech Geoenviron Eng* 131:1274–1282
- Lombardi D, Bhattacharya S (2014) Modal analysis of pile-supported structures during seismic liquefaction. *Earthquake Eng Struct Dyn* 43(3):119–138. doi:10.1002/eqe.2336
- Mavroidis GP, Papageorgiou AS (2003) A mathematical representation of near-fault ground motions. *Bull Seismol Soc Am* 93:1099–1131
- Moss RES, Seed RB, Kayen RE, Stewart JP, Der Kiureghian A, Cetin KO (2006) CPT-based probabilistic and deterministic assessment of in situ seismic soil liquefaction potential. *J Geotech Geoenviron Eng* 132:1032–1051
- NEHRP (2000) National Earthquake Hazards Reduction Program (NEHRP, 2000): commentary for Federal Emergency Management Agency (FEMA, USA 369) on seismic regulations for new buildings and other structures
- Public Works Research Institute (1995) Strong-motion acceleration records from Public Works in Japan (No. 21). Ministry of Construction, Tsukuba, Japan
- Rankine WJM (1866) Useful rules and tables. London
- Seed HB, Idriss IM (1971) Simplified procedure for evaluating soil liquefaction potential. *J Soil Mech Found Div* 97:1249–1273
- Tokimatsu K, Ohoka H, Shamoto Y, Asaka Y (1997) Failure and deformation modes of piles due to liquefaction-induced lateral spreading in 1995 Hyogoken-Nambu earthquake. *J Struct Constr Eng, AIJ* 495:95–100
- Uzuoka R, Sento N, Yashima A, Zhang F (2002) 3-dimensional effective stress analysis of a damaged group-pile foundation adjacent to a quay wall. *J Jpn Assoc Earthq Eng* 2:1–14
- Zhao JX, Zhang J, Asano A, Ohno Y, Oouchi T, Takahashi T, Ogawa H, Irikura K, Thio HK, Somerville PG, Fukushima Y, Fukushima Y (2006) Attenuation relations of strong ground motion in Japan using site classification based on predominant period. *Bull Seismol Soc Am* 96:898–913

SAR Images, Interpretation of

Antonio Iodice and Daniele Riccio

Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione, Università degli Studi di Napoli Federico II, Napoli, Italy

Introduction

A synthetic aperture radar (SAR) is a remote sensing microwave imaging system which consists of a radar sensor mounted on a moving platform, such as an airplane or a satellite. As the platform flies along an approximately straight line, the radar emits microwave pulses at fixed rate (pulse repetition frequency (PRF)) and receives corresponding returns (echoes) backscattered by the illuminated scene. A SAR is able to distinguish points at different distances from the line of flight based on different delays of their returns; in addition, points at different positions along a direction parallel to the line of flight are distinguished by forming a very long (and therefore very directive, with a very narrow beam) synthetic array. This is obtained by properly combining pulses received by the sensor at different positions along the line of flight, so that the synthetic array length is equal to the length of the portion of line of flight such that a given ground point remains within the real antenna beamwidth. In this way, a two-dimensional (2D) image is obtained, which is the projection

of the scene onto the plane containing the look direction (range direction) and the line of flight (azimuth direction). This is at variance with optical images, which are the projection of the scene onto the plane perpendicular to the look direction.

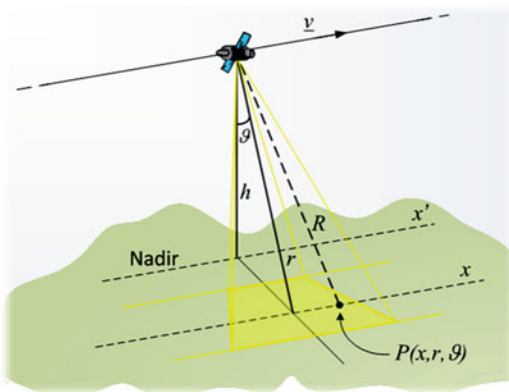
A SAR is an active sensor (i.e., it illuminates the scene), so that it can work during both the day and the night; in addition, it relies on microwaves, which can propagate through clouds, so that it can image the Earth's surface even in the presence of a cloud cover.

An important parameter characterizing an imaging sensor is its resolution, which is the minimum distance between two points such that they can be distinguished by the sensor. For a SAR sensor, the resolution along the range direction is $c/(2B)$, where c is the speed of light and B is the pulse bandwidth, whereas the resolution along the azimuth dimension is $\lambda r/(2X)$, where λ is the electromagnetic wavelength, r is the sensor-to-ground distance, and X is the synthetic array length. In its usual acquisition mode ("strip map" mode), the SAR radar antenna is constantly pointed along a direction perpendicular to the line of flight, forming a significant angle ("look angle") with the nadir direction (i.e., SAR is a "side looking" sensor); see Fig. 1. In this case, the synthetic antenna length X is $\lambda r/L$, where L is the real antenna effective azimuth length, so that the azimuth resolution turns out to be $L/2$. Note that the SAR resolution is independent of the

sensor-to-scene distance, at variance with the one of optical sensors. In order to obtain a higher (better) resolution, although with a smaller illuminated scene, a "spotlight" acquisition mode can be used, in which the SAR antenna beam is steered during the flight to constantly illuminate a given spot on the ground. In this way, a longer synthetic array can be obtained, this implying a better azimuth resolution. The resolution of modern spaceborne SAR systems spans from about 10 m to fractions of meter.

A SAR image provides information on the imaged scene which is in some sense complementary with respect to that provided by an optical image. In fact, while the intensity of a pixel in an optical image mainly depends on the chemical properties of the surface of the imaged objects, the intensity of a pixel in a SAR image depends on electromagnetic properties (permittivity and conductivity) of imaged objects and on their roughness at wavelength (i.e., centimetric) scale: smooth surfaces (calm water, concrete or asphalt surfaces, etc.) appear as dark areas on the image, whereas surfaces with increasing roughness appear as increasingly bright areas.

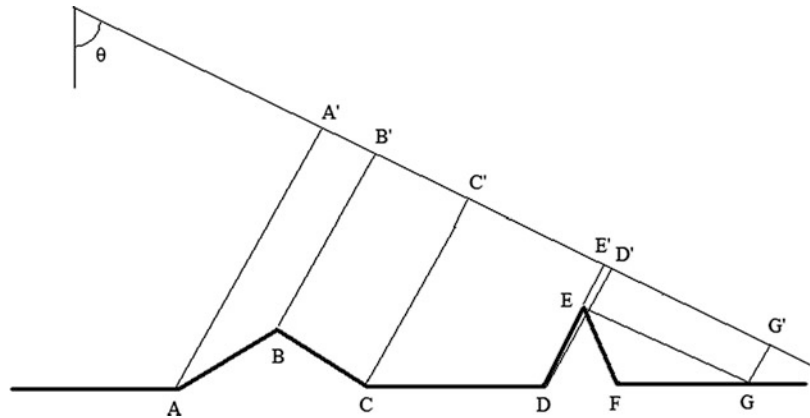
Finally, another important peculiarity of SAR sensors with respect to optical ones is their "coherent" nature: in fact, they are able to emit a coherent radiation and to measure not only the intensity of the received signal but also its phase. This allows using interferometric techniques (interferometric SAR (InSAR) and differential interferometric SAR (DInSAR)) to obtain terrain topography and to monitor small terrain movements and tomographic techniques (SAR Tomography) for the three-dimensional reconstruction of imaged objects. In addition, using two orthogonally polarized transmitting and/or receiving antennas, information can be extracted by observing how the polarization of the backscattered wave is modified with respect to the transmitted one (SAR Polarimetry). However, the coherent nature of SAR also causes the appearance of the "speckle" noise, which gives a "salt-and-pepper" look to SAR images: a macroscopically homogeneous area appears to be composed of pixels of randomly varying intensity. Speckle noise can be reduced, at the expense of



SAR Images, Interpretation of, Fig. 1 Geometry of SAR acquisition

SAR Images, Interpretation of,

Fig. 2 Geometric distortions. Foreshortening-compression ($A'B'$), foreshortening-dilation ($B'C'$), layover ($E'D'$), shadow ($D'-G'$)



geometric resolution, by averaging over adjacent pixels (multi-look image), or by proper filtering.

By summarizing, SAR sensors have significant advantages with respect to optical ones (day-and-night, all-weather capabilities, coherent nature), but SAR images are more difficult to be visually interpreted than optical ones, due to both geometrical (i.e., image projection plane including the look direction) and radiometric (i.e., involved dependence of image intensity on terrain electromagnetic and roughness properties, speckle) issues. Fundamentals on SAR systems and applications can be found, e.g., in Ulaby et al. (1986), Elachi (1988), and Curlander and McDonough (1991). ▶ [InSAR](#) and ▶ [SAR Tomography](#) are the subjects of other entries of this encyclopedia, whereas SAR Polarimetry is analyzed in detail in Lee and Pottier (2009).

Foreshortening, Layover, Shadow, and Geocoding

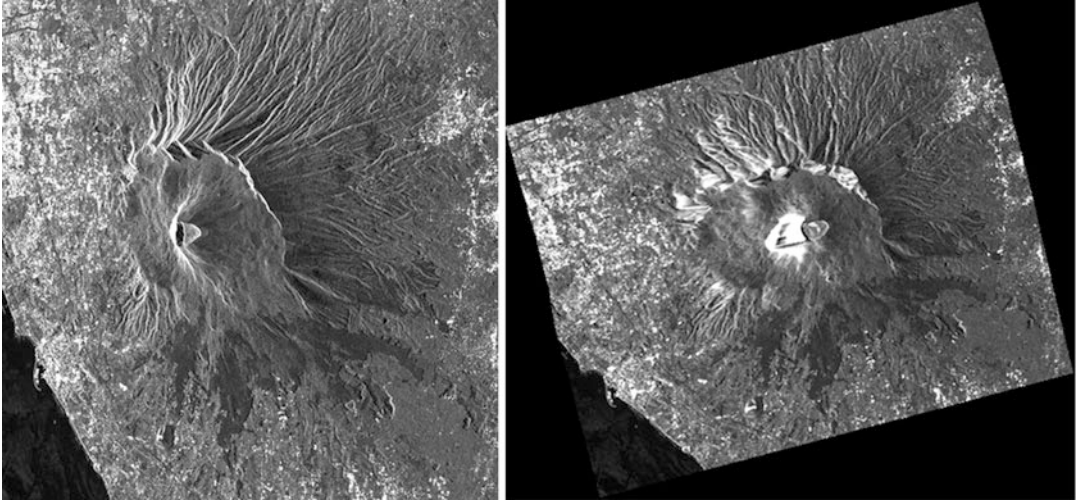
Due to the fact that a SAR image is the projection of the scene onto a plane including the look direction, geometric distortions on SAR images are very different from those experienced in optical images.

If terrain slope is smaller than the look angle θ , the resolution cell on the ground is compressed with respect to the horizontal terrain case if the surface is tilted toward the sensor; otherwise it is dilated (see Fig. 2.). This effect is termed “foreshortening.”

If the surface is tilted toward the sensor and its slope is larger than the look angle, then there is an “inversion” of the SAR geometry: the positions along the range direction of peaks and bases of hills or mountains are exchanged, and the sides of hills or mountains are “folded” onto the valleys in front of them (so that a single pixel corresponds to both an area on the hill’s side and an area on the valley, and a very bright area on the image appears) (see Figs. 2 and 3). This effect is called “layover.”

Finally, if the surface is tilted away from the sensor and its slope is larger than 90° minus the look angle, then a portion of the surface is not illuminated (see Fig. 2), and no return is present in that portion of the SAR image: a “shadow” appears.

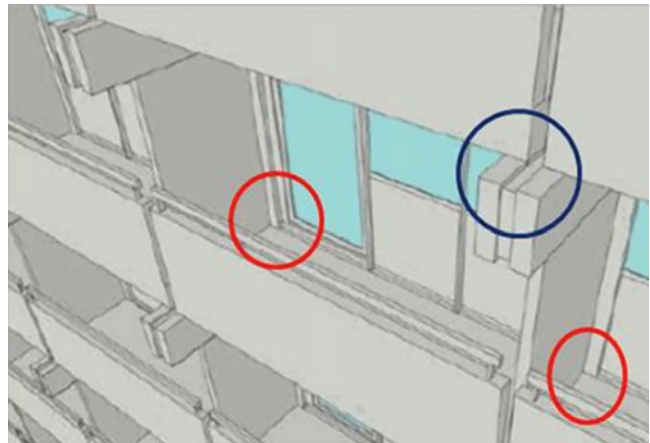
If SAR platform trajectory and terrain topography are known, abovementioned geometric distortions can be mitigated by a post-processing step, called “geocoding,” that allows representing the SAR images in a standard cartographic map projection. In this way, the image can be, for instance, easily integrated in a GIS. Note however that, although foreshortening can be corrected by geocoding, layover and shadow effects imply a loss of information that cannot be recovered by this post-processing step. In addition, geocoding implies an interpolation process that may alter the image information content: therefore, in some applications it may be preferable to extract the physical parameter of interest directly from the image in SAR native geometry and then to geocode the obtained final



SAR Images, Interpretation of, Fig. 3 COSMO/SkyMed SAR image of the area of Mt. Vesuvius, Italy (*left*). Near range is on the left. Geocoded version of the

same image (*right*). The very bright area near the crater corresponds to a layover area

SAR Images, Interpretation of, Fig. 4 Elements on a building facade forming dihedral and trihedral structures

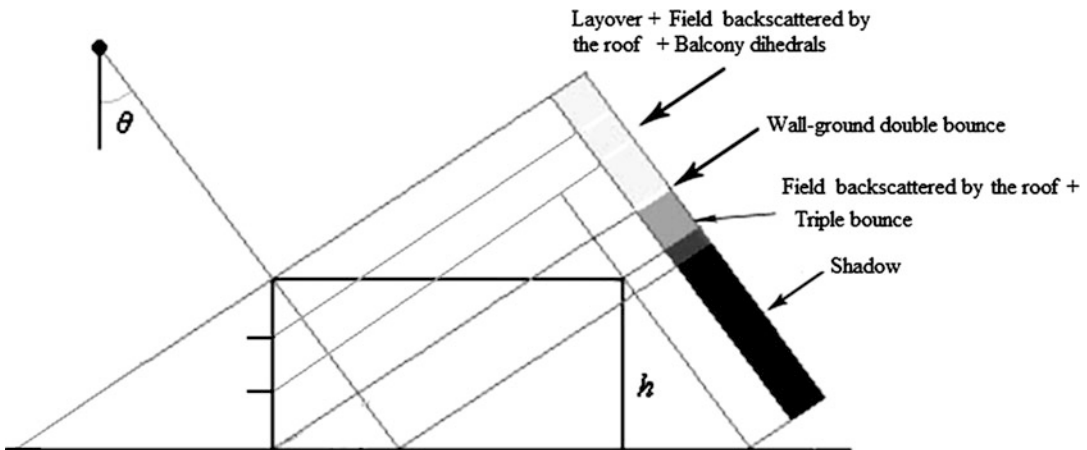


map (Guida et al. 2008; Di Martino et al. 2012). An example of SAR image before and after geocoding is shown in Fig. 3.

Interpreting SAR Images of Buildings

Geometric distortions described in the previous section are particularly severe in SAR images of urban areas, which are of prominent interest if, for instance, a fast post-event earthquake damage assessment is needed. With the launch of COSMO/SkyMed and TerraSAR-X missions,

very-high-resolution (VHR) SAR images of urban areas have become routinely available. In particular, in the spotlight acquisition mode, COSMO/SkyMed SAR sensors are able to obtain a resolution even better than 1 m. Accordingly, in principle a lot of information on objects present in the urban scenario can be extracted from such images; however, due to the above-cited severe geometric distortions and due to the involved interaction between incident electromagnetic wave and imaged scene, direct interpretation of VHR SAR images is not straightforward. It is easy to realize (see Figs. 4 and 5) that such



SAR Images, Interpretation of, Fig. 5 SAR image formation for a single building

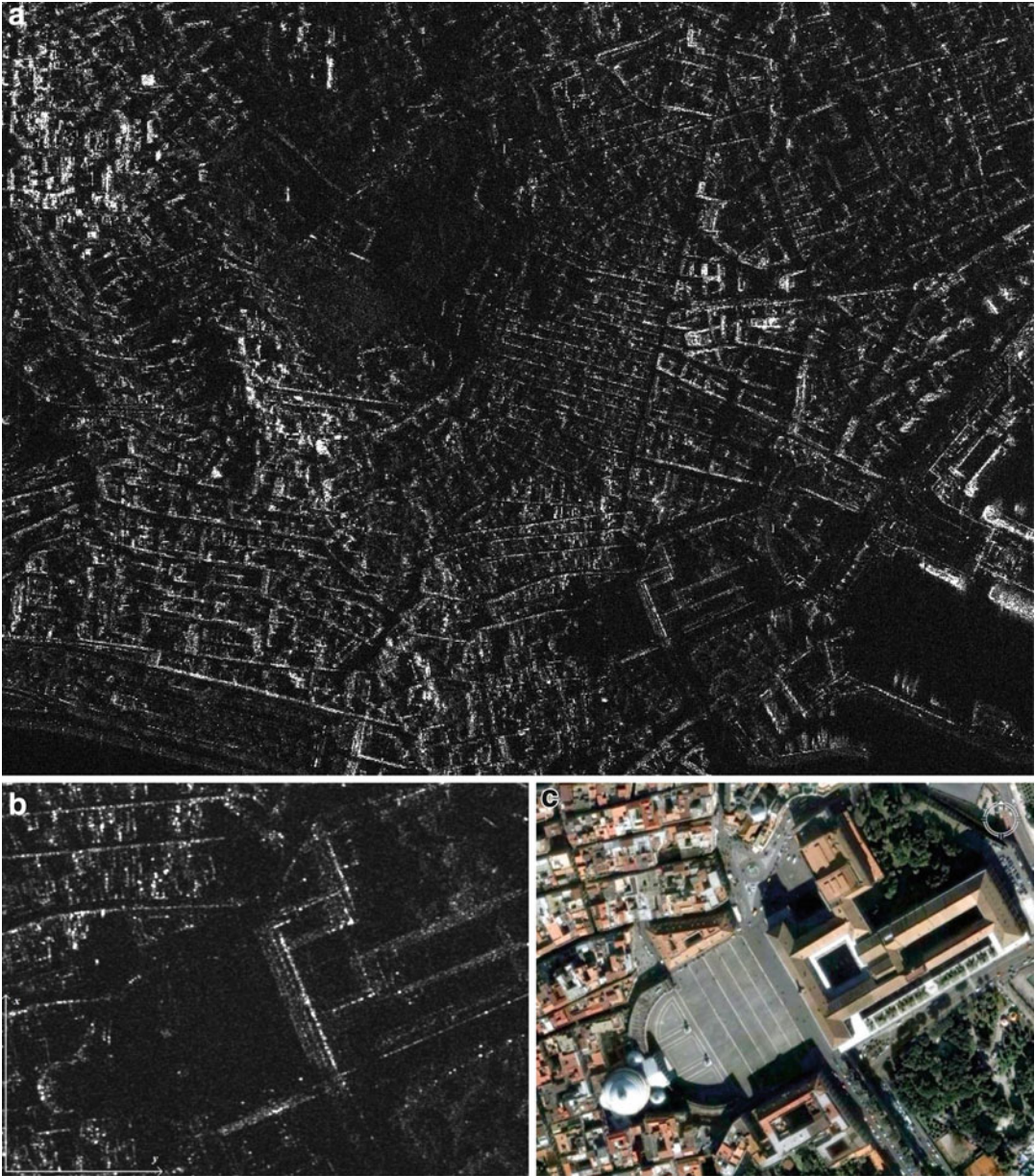
images are dominated by the combination of single scattering from terrain and buildings and multiple scattering from dihedral and trihedral structures. For moderate- and low-resolution systems, they may be simultaneously present in a single resolution cell. Nonetheless, for very-high-resolution SAR systems, the resolution cell is so small that dihedral and trihedral returns dominate with respect to the single scattering background. Accordingly, a realistic description of VHR SAR amplitude images can be obtained by considering sparse brilliant points, or lines, over a dark background: the positions of these brilliant points and lines can be considered randomly distributed, unless they belong to a building facade, in which case an ordered, periodic spatial distribution is expected.

In fact, if we consider a single building, as illustrated in Fig. 5, by moving from near to far range, we first find a layover area in which each pixel is the superposition of two or three contributions: one from the ground, one from the vertical wall, and possibly one from the roof. Ground and roof are usually so smooth that corresponding contributions are weak, whereas often, on the vertical walls, dihedral or trihedral structures of balconies and windows are present (see Fig. 4) whose contributions to the backscattered signal are significant. In fact, with regard to dihedral structures, it is easy to realize that, if the structure is aligned with the SAR line of flight, all double-

bounce paths have the same length, corresponding to twice the path length from the sensor to the internal edge of the dihedral, and hence they reach the sensor simultaneously. Accordingly, a very bright line appears on the image. Similarly, if we consider a trihedral structure, for a wide range of structure orientations, all triple-bounce paths have the same length, corresponding to twice the path length from the sensor to the internal corner of the trihedral, and hence they reach the sensor simultaneously. Accordingly, a very bright point appears on the image. Since balconies and windows usually are spatially distributed in an ordered way, they form periodic patterns of brilliant lines and/or points within the layover area of the SAR image of the building.

By continuing to move from near to far range, the end of the layover area is marked by a bright line representing the double-bounce return from the large dihedral structure formed by the vertical wall and the ground. This line is particularly evident if the wall is aligned with the sensor line of flight. Then, a dark area is present, including the very weak triple-bounce return (wall-ground-wall and/or ground-wall-ground) and, possibly, the weak return from the roof. And, finally, we find a very dark shadow area.

An example of VHR SAR image of an urban area is reported in Fig. 6a. This is a 1-m resolution TerraSAR-X image of Naples, Italy. Near range



SAR Images, Interpretation of, Fig. 6 TerraSAR-X image of Naples, Italy (a). Near range is on the left. Excerpt of the previous image showing the area of Piazza del Plebiscito (b). An optical image of the same area (c)

is on the left. In Fig. 6b an excerpt of the previous image is reported, and in Fig. 6c an optical image of the same area is shown. The square at the center of this area is Piazza del Plebiscito, and the building at its right side is the Palazzo Reale (Royal Palace). In agreement with our previous discussion, in correspondence of the façade of

this building, by moving from left to right (i.e., from near to far range), we can distinguish three bright lines, followed by a very bright line and by a dark area. The three bright lines correspond to an architectural structure at the roof edge and to two lines of balconies, whereas the very bright line corresponds to the wall-ground

double-bounce return. Finally, the dark area corresponds to building roof and shadow.

Model-Based Interpretation of SAR Images

Above presented description allows a qualitative interpretation of SAR images of natural and urban areas. Quantitative information on the geometry (i.e., distances between objects, building size, floors' height, etc.) can be also obtained from such an analysis if look angle and pixel spacing are known. However, in order to obtain quantitative relations between image intensity and scene properties, a deeper analysis is needed. First of all, the scene must be described in terms of parameters of interest: for instance, for a natural scenario, soil moisture and composition, terrain roughness, vegetation biomass, etc. Then, a direct electromagnetic scattering model must be used to express the backscattered field in terms of such scene parameters. Inversion of this model allows retrieving the scene parameters of interest from SAR images of the scene.

This field is currently the subject of intense research activity, and the description of scattering models and retrieval algorithms goes beyond the scope of the present work. However, a review of scattering models can be found in some textbooks (e.g., Ulaby et al. 1986; Tsang et al. 2000; Franceschetti and Riccio 2007) and scientific papers (e.g., Fung et al. 1992; Franceschetti et al. 2002). Examples of retrieval algorithms can also be found in the scientific literature, both for natural scenes (Iodice et al. 2011; Di Martino et al. 2012) and urban areas (Guida et al. 2008, 2010).

RGB Compositions

Visual interpretation of SAR images can be made easier by combining different SAR images of the same area to create a false color image.

For instance, if a SAR polarimetric system is employed, a combination of different polarimetric channels can be used. Available channels are

HH (i.e., transmit a horizontally polarized electromagnetic field and measure the horizontally polarized component of the received field), VV (i.e., transmit a vertically polarized electromagnetic field and measure the vertically polarized component of the received field), and HV or VH (i.e., transmit a horizontally polarized electromagnetic field and measure the vertically polarized component of the received field, or vice versa). For a wide range of scattering surfaces (reciprocal scatterers), HV and VH channel returns are equal, so that only one of the two can be actually used. A very useful combination consists of HH + VV, HH-VV, and HV channels, and it is called a "Pauli decomposition" (Lee and Pottier 2009). In fact, it turns out that the sum of HH and VV returns is dominated by single scattering from rough surfaces (soil surfaces, sea surfaces), the difference of HH and VV returns is dominated by double scattering (terrain-building walls, or ground-tree trunks), and HV return is dominated by volumetric scattering or extremely rough surface scattering (vegetation). Accordingly, an RGB color image can be obtained by loading the HH + VV signal onto the blue channel, HH-VV onto the red channel, and HV onto the green one. Accordingly, blue areas on the image will correspond to bare or little vegetated soils, or sea; red areas to built-up areas or trees with little foliage; and green areas to very vegetated soils or forests. Intermediate colors will correspond to pixels containing combinations of the previous targets. An example of false color SAR image obtained by using the Pauli decomposition is reported in Fig. 7.

Another possibility is to load, onto the three different color channels, SAR images of the same area acquired at different times. This allows to easily identify areas subjected to changes between two different acquisitions. For instance, in Fig. 8 a false color SAR image of an area in a semiarid region (Tougou basin, Burkina Faso) is shown. An image acquired during the dry season is loaded onto the blue channel, another image acquired in the wet season is loaded onto the green channel, and the interferometric coherence between the two acquisitions (see ► [InSAR and](#)



SAR Images, Interpretation of, Fig. 7 False color image obtained by performing a Pauli decomposition on the polarimetric SAR image of an agricultural area



SAR Images, Interpretation of, Fig. 8 False color SAR image of the Tougou basin area (Burkina Faso)

A-InSAR: Theory is loaded onto the red channel. Accordingly, different colors can be interpreted as follows (Amitrano et al. 2015):

Black areas: permanent basin water (dark on both images, low coherence)

Blue areas: wet season basin water (dark on wet season image, intermediate on dry season image, low coherence)

Green areas: wet season vegetation (bright on wet season image, intermediate on dry season image, low coherence)

Red or white areas: man-made objects, village (intermediate or bright on both acquisitions, high coherence)

Blue-green intermediate color, high intensity (cyan): trees (bright on both acquisitions, low coherence)

Blue-green intermediate color, intermediate intensity (Prussian blue or dark green): bare soils (intermediate on both acquisitions, intermediate coherence)

Note that the one described above is just an example: different RGB compositions can be employed, according to the considered application and to the scene characteristics that the user is interested to highlight.

Summary

SAR sensors have significant advantages with respect to optical ones: day-and-night, all-weather capabilities, and the possibility of measuring terrain topography and monitoring small terrain movements, due to its coherent nature. In addition, SAR images of virtually any area of the Earth surface are today routinely available, due to the different SAR satellite missions currently in orbit. Therefore, their use has a huge potential impact on a number of applications, among which the fast post-event earthquake damage assessment.

However, visual interpretation of SAR images by a human operator requires that the latter is properly trained to get used to the peculiar characteristics of SAR images: geometric distortions, involved dependence of image intensity on terrain electromagnetic and roughness properties, and speckle noise. Visual interpretation can be made easier by properly combining different images to form a color image (RGB composition).

Finally, automatic quantitative interpretation of SAR images requires the availability or the development of electromagnetic scattering models and of corresponding retrieval algorithms. This field is currently the subject of an intensive research activity.

References

Amitrano D, Di Martino G, Iodice A, Riccio D, Ruello G (2015) A new framework for SAR data RGB multitemporal representation: rationale and products. *IEEE Trans Geosci Remote Sens* 53:117–133

- Curlander JC, McDonough RN (1991) *Synthetic aperture radar: systems and signal processing*. Wiley, New York
- Di Martino G, Riccio D, Zinno I (2012) SAR imaging of fractal surfaces. *IEEE Trans Geosci Remote Sens* 50:630–644
- Elachi C (1988) *Spaceborne radar remote sensing: applications and techniques*. IEEE Press, New York
- Franceschetti G, Riccio D (2007) *Scattering, natural surfaces and fractals*. Academic, Burlington
- Franceschetti G, Iodice A, Riccio D (2002) A canonical problem in electromagnetic backscattering from buildings. *IEEE Trans Geosci Remote Sens* 40:1787–1801
- Fung AK, Li Z, Chen KS (1992) Backscattering from a randomly rough dielectric surface. *IEEE Trans Geosci Remote Sens* 30:356–369
- Guida R, Iodice A, Riccio D, Stilla U (2008) Model-based interpretation of high-resolution SAR images of buildings. *IEEE J Sel Topics Appl Earth Observ* 1:107–119
- Guida R, Iodice A, Riccio D (2010) Height retrieval of isolated buildings from single high resolution SAR images. *IEEE Trans Geosci Remote Sens* 48:2967–2979
- Iodice A, Natale A, Riccio D (2011) Retrieval of soil surface parameters via a polarimetric two-scale model. *IEEE Trans Geosci Remote Sens* 49:2531–2547
- Lee JS, Pottier E (2009) *Polarimetric radar imaging: from basics to applications*. CRC Press, Boca Raton
- Tsang L, Kong J, Ding K (2000) *Scattering of electromagnetic waves – theory and applications*. Wiley, New York
- Ulaby FT, Moore RK, Fung AK (1986) *Microwave remote sensing, active and passive*. Artech House, Norwood

SAR Tomography for 3D Reconstruction and Monitoring

D. Reale and G. Fornaro
Institute for Electromagnetic Sensing of the Environment, National Research Council,
Naples, Italy

Synonyms

Cosmo-Skymed; Differential interferometry; DInSAR; Displacement measurement; Infrastructures monitoring; Multidimensional SAR imaging; SAR tomography; Synthetic-aperture radar, SAR; TerraSAR-X

Introduction

Synthetic-aperture radar (SAR) is one of the most important Earth remote sensing sensors whose applications have grown dramatically in the recent years. It provides images at microwaves with resolution comparable to that of optical systems, but with the crucial advantage of all-time, day/night, and all-weather, imaging capability. Similarly to classical surveillance radars, SAR measures the distance (range) from sensor to target: resolutions of the order of meters are achieved through the pulse compression of large bandwidth (frequency-modulated) signals. Very high resolution in the along-track direction is achieved as well through the coherent combination of target echoes received over the illumination interval, thus implementing a virtual (synthetic) array of antennas (i.e., a very large antenna) by exploiting the movement of a very small antenna mounted on board airplanes or satellites. The latter feature turns SAR systems to be imaging radars.

SAR data are nowadays used in many areas of environmental risks monitoring situations such as flooding, glaciers, land cover, and forest monitoring (Curlander and McDonough 1991; Moreira et al. 2013). Among all, one of the primary applications of SAR is the 3D reconstruction and monitoring of the Earth surface displacements through the use of interferometric techniques.

SAR interferometry (InSAR) exploits the coherent properties of the sensor, i.e., the capability to accurately control not only the envelope but also the phase of the transmitted radiation. SAR images are in fact complex data characterized by an amplitude measurement (envelope), related to the backscattering properties of the scene, as well as a phase signal related to both phase of the backscattering coefficient and to the distance of the target from the sensor to an accuracy of the order of the wavelength (centimeters at microwaves). Similarly to the human vision system, acquiring images with a slight angular diversity allows SAR to be sensitive to the 3D scene properties, i.e., to estimate the topography of the observed scene. Topography is not accessible in a single SAR acquisition because the

imaging process returns only a 2D projection of the 3D reality. The Shuttle Radar Topography Mission (SRTM) in the last decade has represented the first case of an extensive use of InSAR for the generation of a worldwide (except for the poles) digital elevation model (DEM), that is, a digital topography map of the Earth surface (Van Zyl 2001). SRTM DEM (90 m spacing DTED-1 standard) has been extensively used in many applications. Nowadays, the TerraSAR-X/TanDEM-X mission (launched in 2007 and 2011) is providing a “refreshing” with higher resolution (12 m spacing DTED 3 standard) of the Earth DEM on a global scale by exploiting the simultaneous acquisitions of two twin SAR sensors flying in a close formation (Krieger et al. 2007). On the other hand, differential interferometry (DInSAR) takes advantage of the very high precision of radar systems in measuring phase to estimate differential displacements of the area imaged at different time instants with an accuracy of the order of a fraction of the used wavelength. DInSAR is today routinely used to estimate displacements induced by large earthquakes as well as to monitor volcanic activities producing ground movements, subsidence caused by water and/or oil extraction, mining activities, and also slow moving landslides (Massonnet et al. 1993; Carnec et al. 1995; Fornaro and Franceschetti 1999; Crosetto et al. 2005).

The availability of long-term data archives of former C-Band ESA ERS1/2 and ENVISAT SAR satellites has pushed the development of multipass interferometric processing techniques which coherently process large dataset of tens of images. In this way intrinsic limitations of classical, single pair, DInSAR, as the presence of atmospheric phase contribution and unwrapping procedures needs, affecting the unambiguous estimation of the useful deformation signal, are overcome. These techniques are mainly categorized based on the assumption on the scattering on the ground. From one hand, the approaches referred to as Persistent Scatterer Interferometry (PSI) prioritize the spatial resolution and uses all available baselines for accurate monitoring of “strong” (i.e., persistent) scatterers, typically located on anthropic structures, exhibiting

a temporal response stability over the whole observation period (Ferretti et al. 2000, 2001). On the other hand, multipass DInSAR techniques as the Small Baseline Subset (SBAS) limits the processing to interferograms characterized by short temporal separation as well as reduced angular diversity and makes use of a spatial multilooking to enhance the signal quality at the expense of a spatial resolution loss. SBAS techniques, known also as DInSAR stacking, are devoted to the monitoring of wide areas including rural zones (Berardino et al. 2002; Ferretti et al. 2011).

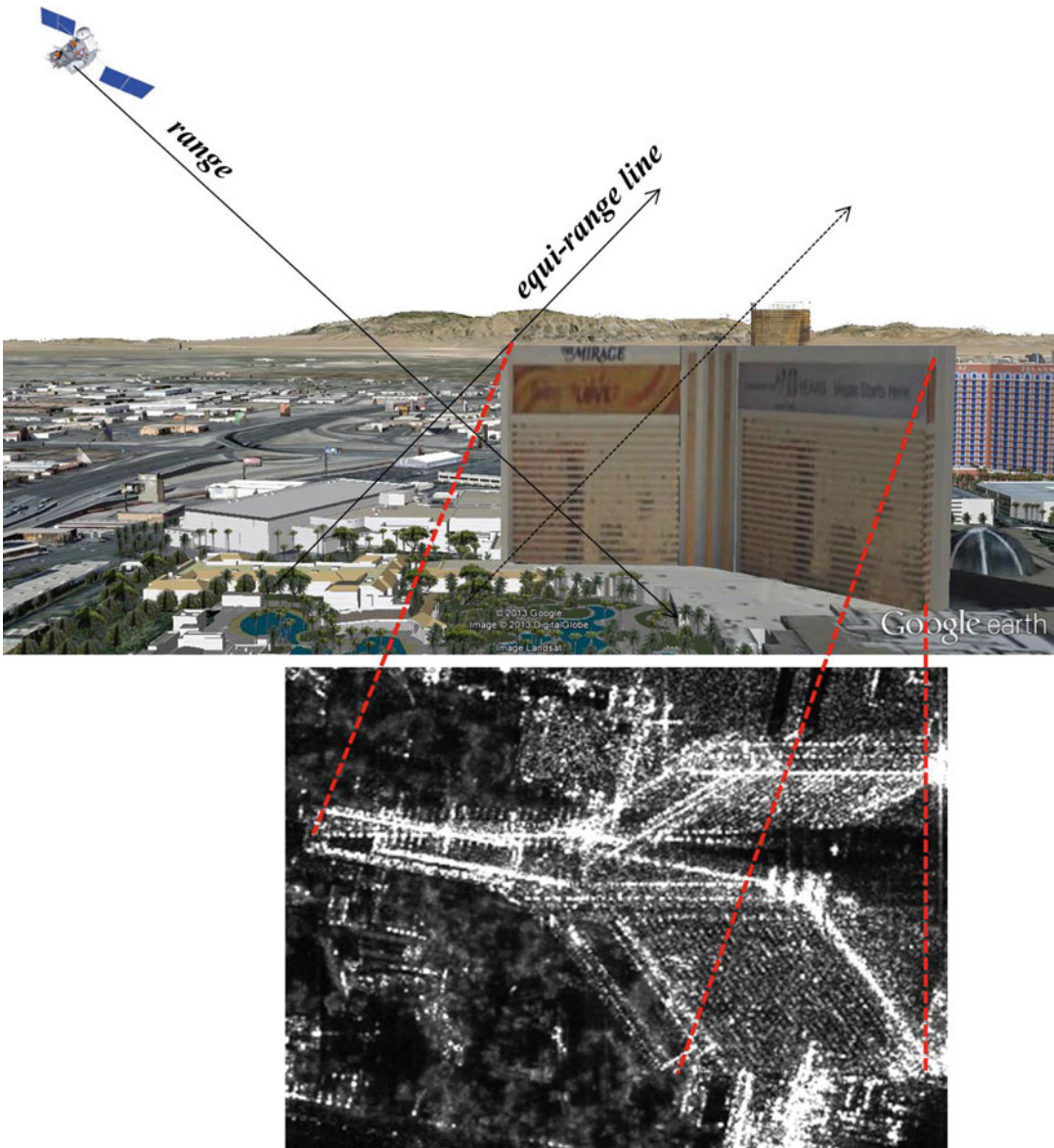
PSI and SBAS approaches have been used to investigate several aspects of risks for monitoring of coseismic and post-seismic deformation corresponding to several major earthquakes, volcanoes deformation, landslides, etc., as well as for the monitoring of buildings and infrastructures (Cascini et al. 2007; Arangio et al. 2013).

A recent advance for the technological viewpoint in the application to building reconstruction and monitoring is provided by SAR tomography which extends the SAR imaging concept to the third dimension of height: SAR tomography is also known as multidimensional SAR imaging technique due to the capability of full 3D imaging plus time monitoring. As for PSI, SAR tomography exploits full resolution data with angular diversity but for the use of the whole complex measured data then introducing the virtual antenna array along the height: the larger height antenna extent allows reducing again, as for the azimuth, the antenna beamwidth and reconstructing with finer resolution the backscattering along height (3D imaging) (Reigber and Moreira 2000; Gini et al. 2002; Fornaro et al. 2005). This latter allows, from one side, estimating height parameters of scatterers with better accuracy but, above all, detecting the presence of possible multiple scattering mechanisms which may interfere within the same radar spatial resolution cell. This interference, known as the layover effect, is a direct consequence of the imaging principle of radar systems which discriminates scatterers in distance: in the typical SAR side-looking geometry, in presence of steep topography, as for buildings, walls, and in

general vertical surfaces, it happens that scatterers located at different heights may be sensed by the radar at the same distance and therefore their returns imaged in the same pixel. A pictorial explanation of the layover induced on a building, as well as its effect on a very high (1 m) resolution TerraSAR-X amplitude image, is provided in Fig. 1: backscattering returns from targets located on the left part of the roof are imaged at first, and then the contributions from the facade and finally the base of the building are imaged in far range. Consequently, the building is tilted toward the sensor in the resulting SAR image: notice that because of the very high resolution, returns from the facades are spread over a large number of pixels in which contribution from ground is also expected. The vertical synthetic aperture exploited by SAR tomography allows improving resolution and tightening responses of the different interfering scatterers and then giving a chance to detect and localize separately each scattering mechanism (Fornaro and Serafino 2006). This capability has a major importance in the processing of urban areas, where the presence of buildings causes very frequent occurrence of layover between building facades, surrounding structures and ground. It is worth to note that this feature is own of the tomographic approach: none of interferometric approaches, including the PSI, can counteract this interference because of the leading assumption of only one scattering mechanism per pixel and of the phase-only signal model which does not cope with an imaging viewpoint.

As a direct extension of DInSAR, differential tomography (4D imaging) has been also proposed: it extends the 3D imaging to measure also the deformation parameters of scatterer in the focused 3D space (Lombardini 2005; Fornaro et al. 2009a; Zhu and Bamler 2010a). Along the same lines of PSI, also time series explaining the temporal evolution of deformation can be extracted, even separately for each interfering scatterers possibly exhibiting different deformation behaviors (Fornaro et al. 2009).

SAR technology is also evolved to specifically accomplish requirements and improve monitoring performances of multipass techniques.



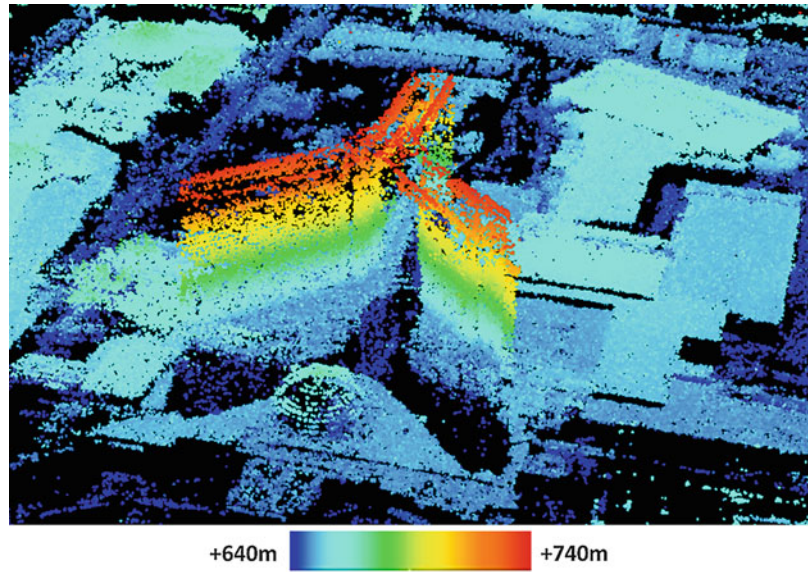
SAR Tomography for 3D Reconstruction and Monitoring, Fig. 1 Pictorial illustration of the layover distortion on SAR images induced by the side-looking imaging principle of radar and evidence on a 1 m spatial resolution

TerraSAR-X image: closest targets located on the roof are imaged in near range respect to those on the base of the building. Optical image courtesy of Google

Recent X-Band sensors as the TerraSAR-X and the Cosmo-Skymed constellation are providing images with resolution never achieved in the past, in the meter/submeter regime with reduced revisiting times. As far as the latter aspect is concerned, the Cosmo-Skymed mission is worldwide the largest constellation for civilian

application: it is composed by four small-size satellites that ensure an average revisiting time of 4 days at the maximum acquisition rate, which dramatically impacts in unexpected emergency situations when immediate imaging of damaged areas is required. High resolution allows capturing much more information from the scene,

SAR Tomography for 3D Reconstruction and Monitoring, Fig. 2 3D visualization of scatterers detected by the SAR tomography for the building imaged in the amplitude image of Fig. 1 (Reale et al. 2011a). Colors are set according to the estimated height



notably increasing the density of monitored scatterers (Gernhardt et al. 2010). At the same time, the layover becomes a major issue in the imaging of urban areas: the finer is the resolution, the larger will be the number of pixels affected by the layover induced by the buildings facades. Multidimensional imaging have demonstrated to be an effective tool in resolving the distributed layover on very high-resolution data, allowing fully separating contribution from ground and facade for the accurate 3D reconstruction of buildings. An example of the capability of multidimensional imaging for 3D building reconstruction and layover solution is provided in Fig. 2 which represents the 3D point cloud of the scatterers detected over the building imaged in Fig. 1 by the SAR tomography on a dataset of 25 TerraSAR-X very high-resolution spotlight images (Reale et al. 2011a). The precise reconstruction proves the effectiveness of this advanced processing which takes benefit of the detailed imagery provided by 1 m resolution acquisition capability.

This chapter is intended to introduce the principal concepts of the multidimensional imaging approach. A brief introduction on classical interferometric approach is firstly provided to introduce the concepts of angular and temporal

diversity in SAR imagery and their relationship with physical parameters of interest as topography and surface movements. SAR tomography is introduced as an extension which turns the interferometric processing into an imaging problem, allowing both improving performances in estimation of parameters as well as allowing extracting further information from the data with respect to interferometry as the separation and detection of interfering scatterers in layover areas. Finally, examples of the application of both multipass DInSAR (small scale), and tomographic (full resolution) processing carried out on recent Cosmo-Skymed data are provided to point out the potentiality of the joint use of these technique for a complete monitoring of risk situations at different scales, from regional up to the level of the single infrastructure, spatial scales.

The SAR Interferometry Background

SAR imaging allows discrimination of targets along the azimuth and range directions. Assuming SAR sensors to fly locally rectilinear trajectories (airborne) or orbits (spaceborne), azimuth x (directed along the sensor velocity vector) and range r (distance orthogonal from the flight track)

represent two coordinates of the natural radar cylindrical reference system with the axis coincident with the flight track. By using large bandwidths, reaching the order of hundreds of megahertz, modern sensors distinguish targets in range with a resolution degree that reaches the meter/submeter scale. In addition, a high azimuth resolution capability of the final 2D images is achieved by synthesizing antennas in the order of kilometers through the exploitation of the intrinsic motion of the platform along its orbit (Curlander and McDonough 1991).

The height information, which is not accessible in a single SAR image, can be estimated through the interferometric concept. As in any electromagnetic coherent system, the phase information is related to the traveled path and therefore to the distance of the scene from the imaging radar. Classical SAR interferometry exploits the phase difference of, at least, two images acquired with an angular diversity induced by a slight orbital offset (spatial baseline), to retrieve the 3D localization of ground scatterers. Conversely, temporal separation (temporal baseline) is exploited by the differential SAR interferometry to measure possible displacements along the radar's line-of-sight (LOS) occurring at each acquisition epoch. The winning aspect of interferometry is the very high precision of radar in the estimation of the phase values, whose accuracy is of the order of a fraction of the wavelength: this allows estimating movements with sub-centimetric accuracy, using C- and recently X-Band radars from the space. Classical DInSAR has been extensively applied to measure large deformation caused mainly from earthquakes or volcanic activities. However, the presence of additional phase disturbing contributions does not allow pushing accuracy to a millimeter accuracy through the exploitation of just few images. As stated above, the interferometric phase difference $\Delta\varphi$ of a radar system working with a wavelength λ is related to the difference δr in the traveled path of the signal forming the two images. This phase difference is composed of multiple contributions associated with different source (Fornaro and Franceschetti 1999):

$$\Delta\varphi = \frac{4\pi}{\lambda} \delta r = \frac{4\pi}{\lambda} (\delta r_z + \delta r_d) + \Delta\varphi_a + \Delta\varphi_o + \Delta\varphi_n \quad (1)$$

The first term $\delta r_z = (b/r)s$, with s being the slant height (orthogonal to the azimuth/range imaging plane corresponding to the master image) ($s = z/\sin(\vartheta)$, ϑ is the look angle), accounts for the distance variation induced by the presence of topography on the ground which plays a role in the presence of an imaging parallax measured by spatial baseline b . Conversely, δr_d is the distance variation associated with possible deformation signal (LOS component) measured among the two acquisition epochs. Subsequent terms play a disturbance role. The term $\Delta\varphi_a$ is due to the propagation delay variation between the two acquisitions induced by changes in atmosphere. Slowing produces a time delay which is mapped into a range variation. The atmospheric propagation delay (APD) exhibits a spatial correlation over hundreds of meters: it is typically on the same level of deformations which are also often spatially correlated; $\Delta\varphi_o$ and $\Delta\varphi_n$ are associated with orbital inaccuracies and noise. In applications devoted to the estimation of the scene topography, simultaneous acquisitions are preferable because both deformation and atmospheric contributions are absent: this was the case of the Shuttle Radar Topography Mission (SRTM) which employed a dual-antenna system to reconstruct the digital elevation model (DEM) of the Earth and recently of the TerraSAR-X/TanDEM-X mission which employs a pair of twin satellites flying in close formation to produce a better resolved, 12 m spatial resolution, DEM. Differently, when deformation is of interest, as in differential interferometry, repeated passes over the exactly same orbit should be required to avoid the impact of topographic phase contribution. Indeed, this requirement is problematic to enforce, and then DInSAR interferograms are produced by subtracting from the original interferogram an estimation of δr_z evaluated from an external DEM (typically the SRTM DEM): such operation, referred to as zero baseline steering (ZBS), aims at eliminating or at least mitigating the fringes corresponding to

the topography. The fringe pattern retrieved by the external DEM is usually referred to as “synthetic interferogram.” Estimation of deformation is however affected by the presence of atmospheric contribution, and then classical two-pass DInSAR configuration is usually applied to estimate predominant deformation caused, f.i., by large earthquakes.

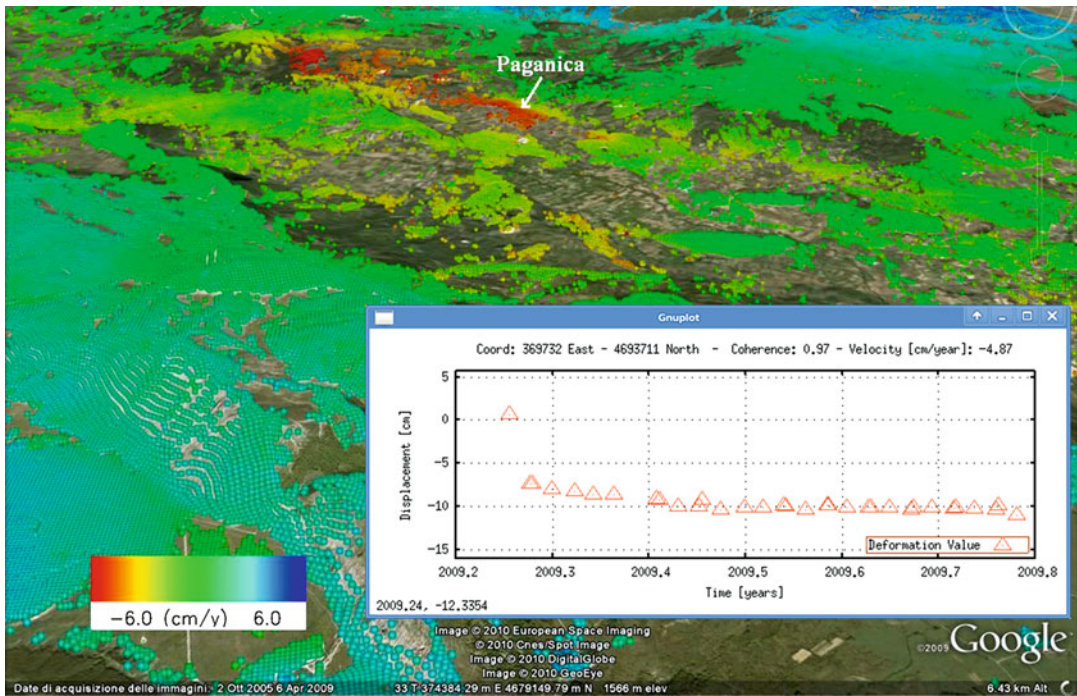
The assumption leading to the differential phase in Eq. 1 is that the scene backscattering involved in the complex conjugate interferometric products are the same in both the images. Changes lead to the presence of noise contribution in the interferometric phase $\Delta\varphi$ which is referred to as decorrelation (Bamler and Hartl 1998; Fornaro and Franceschetti 1999). The coherence measures the degree of decorrelation: it measures the modulus of the cross-correlation index between the two images, which is a measure of the linear predictability of the two random variables corresponding to the master and slave SAR image values in the given pixel. The coherence is a product of several decorrelation effects, the most important being associated with the change of the imaging geometry due to the spatial orbit offset, the temporal separation between the acquisitions and the thermal noise. Spatial decorrelation is caused by the change of the imaging angle which determines a change of the backscattering coefficient, because of the presence in each resolution cell of a large number of scattering sources. The temporal decorrelation is due to the change of the scene backscattering properties over the time. This decorrelation source is of main importance in repeat pass interferometry, especially with systems operating at higher frequencies (f.i., X-Band). It is critical over the sea and in vegetated areas where the growth and in general the change of vegetation lead to strong variation of the backscattering coefficient. The last term, the thermal decorrelation, is due to the presence of thermal noise in the receiving apparatus. It is particularly evident in areas characterized by very low scene backscattering. Other decorrelation sources are associated with variations of the imaging aspect angle (Doppler centroid decorrelation) and to processing artifacts.

It is worth pointing out that what is really measured in interferometry is only a restricted version, corresponding to the modulus 2π of $\Delta\varphi$: one of the most critical steps of the interferometric processing is therefore represented by the unwrapping procedure which is required to retrieve the exact, unrestricted, absolute differential phase from whom quantity of interest can be estimated. Decorrelation effects, as well as large phase discontinuities, f.i., associated to steep height variations, affect the reliability of phase unwrapping algorithms (Ghiglia and Pritt 1998).

Multipass SAR Interferometry

Satellites regularly repeat orbits over the time. As a consequence, stacks of multipass acquisitions, characterized by angular and temporal diversity, are available in remote sensing archives for most of the Earth surface. In order to achieve a higher accuracy in the estimation of the deformation, modern advanced DInSAR processing algorithms jointly process all the images in stacks of multitemporal acquisitions: this processing allows in fact to provide a discrimination between the atmospheric and deformation as well as to cancel possible residual topography components.

Among all, interferometric techniques can be distinguished in two main categories which are characterized by complementary assumptions about the ground scattering: the multipass DInSAR stacking techniques and the Persistent Scatterers Interferometry (PSI). The class of the DInSAR stacking methods is a direct extension of the classical two-pass DInSAR technique. It assumes the scattering to be spatially distributed over the resolution cell and is based on the exploitation of both only small (temporal and spatial) baseline interferograms (hard baseline thresholding) and of a spatial multilook in order to limit the effects of decorrelation and reduce the phase noise in the interferograms, as for the Small Baseline Subset technique (Berardino et al. 2002). It is tailored to the monitoring of wide areas and scattering mechanisms that exhibit decorrelation including rural areas with



SAR Tomography for 3D Reconstruction and Monitoring, Fig. 3 Post-seismic deformation velocity map and co/post-seismic time series of a point in the Paganica

area close to L'Aquila obtained by processing a dataset of 33 Cosmo-SkyMed images acquired between April 4 and October 13, 2009 (Reale et al 2011b)

slow temporal correlation losses. This technique relies on the inversion of the linear system relating interferometric (differential) phase values after phase unwrapping to the phase values on each acquisition. The separation of displacement and atmospheric contribution is carried out by exploiting their statistical characterization in terms of spatial and temporal correlation. DInSAR stacking methods analyze interferograms typically generated by pairing acquisitions characterized by small spatial baselines and temporal separation, thus limiting the decorrelation. They play a favorable role in the design of a two-scale processing: besides effectively counteracting decorrelation and phase noise, the spatial multilooking allows also to under-sample the interferograms in the image spatial (azimuth and range) coordinates, thus reducing the amount of data and consequently the computational effort for the analysis of large areas. Therefore, this class of algorithms is particularly suitable for the analysis of large areas at lower resolution

(small scale). Additionally, through the analysis of multiple interferograms, the atmospheric phase contribution can be estimated and compensated from the data. This latter, coupled with the removal of the background deformation signal occurring on a small scale, are used to phase calibrate the full resolution data for processing at large scale, i.e., at the level of single building and infrastructure.

A typical product of the processing through DInSAR stacking techniques is reported in Fig. 3: it represents the deformation mean velocity map, superimposed to a Google Earth image, corresponding to the slope of the deformation time series for each pixel selected by looking at a quality index measuring the temporal consistency of measurements after phase unwrapping. The processing has been carried out through the application of the Enhanced Spatial Differences (ESD) technique which extends the classical SBAS approach by exploiting a model for the phase differences between adjacent pixels to

counteract the effect of atmospheric contribution, thus supporting the phase unwrapping step and increasing the coverage and accuracy of the retrieved deformation measurements (Fornaro et al. 2009b). The result is relevant to the dataset of Cosmo-Skymed acquisitions over the area of L'Aquila, in the central part of Italy, struck by the 2009 earthquake. During the emergency acquisition plan, the Cosmo-Skymed constellation, at that time composed by three of the four satellites nowadays operatively, intensified the acquisition to its highest possible rate (almost 1 acquisition every 5 days on average) to acquire in only 6 months datasets on ascending and descending orbits on different beams including a sufficient number of images for multipass differential interferometric processing. The result presented in Fig. 3 is relevant to the processing of 33 H-image ascending acquisitions (beam 09 corresponding to an incidence angle of about 40°) taken between April 4 and October 13 2009. Colors in the map move from blue, associated to movements toward the radar los (uplift), to red, associated to movements away from the sensor (subsidence). The velocity map (evaluated only on the post-seismic dates from April 12 to October 13) overlaid to a Google Earth image shows an area of subsidence with a rate exceeding the limit of 6 cm/year affecting the city of Paganica, in the eastern part of L'Aquila. In the same image, the plot of the time series of a point in the subsidence area shows the jump associated to the main shock (April 6) and the exponential decay of the subsidence with the aftershocks. Measurements were shown to be in full agreement with traditional leveling (Reale et al. 2011b) and GPS measurements (D'Agostino et al. 2012).

The second class of multipass techniques, the Persistent Scatterer Interferometry, as for SAR tomography, works at the highest spatial resolution to determine the deformation of single dominant scatterers typically associated with man-made structures (dihedral and trihedral of walls edges, poles, gratings, etc.) (Ferretti et al. 2000, 2001). In this case, to achieve also high accuracy in the estimation of the localization of scatterers, no limitations on the spatial baseline are introduced. Similarly to DInSAR stacking

and differently from SAR tomography, PSI uses only the phase information and assumes the presence of a single persistent scatterer per resolution cell retaining correlation over the time (persistent scatterers). The use of the model however presumes the compensation of phase contributions such as the atmospheric phase delay. Such a compensation can be carried out either by analyzing the phase on persistent scatterers (PS) candidates, which are strong scatterers where the phase is less affected by noise or by using the coarse resolution product of previous stacking techniques. In the latter case, a good practice is also to subtract the low-resolution, spatially correlated, deformation so to obtain also a zero deformation steering to carry out high-resolution analysis on residual phase signals; the following model is assumed for the vector $\boldsymbol{\varphi}$ collecting the compensated phase values in the N available acquisitions at full resolution

$$\boldsymbol{\varphi} = \frac{4\pi}{\lambda} \frac{\mathbf{b}}{r} s + \frac{4\pi}{\lambda} \mathbf{d}(s, \mathbf{t}) + \boldsymbol{\varphi}_n \quad (2)$$

where $\mathbf{b}$ is the N -dimensional vector of spatial baselines and $\mathbf{d}(s, \mathbf{t})$ is the vector collecting the displacements measured at the acquisition instants collected in the vector $\mathbf{t}$. Following the compensation for low-resolution components, deformations are divided into the linear, described by the average (mean) velocity v of the pixel corresponding to the slope of the time series wrt to the epochs, and nonlinear $\mathbf{d}_{nl}$ addenda, i.e., $\mathbf{d}(s, \mathbf{t}) = v\mathbf{t} + \mathbf{d}_{nl}$.

PSI assumes nonlinear terms having a small amplitude and carries out, for each pixel, a measure of the correlation of the measured signal with the model in Eq. 2, through the maximization of the scalar product in the following, which returns also an estimation of the (s, v) parameters:

$$C = \max_{(s, v)} \frac{1}{N} \left| \left(e^{j\boldsymbol{\varphi}} \right)^H e^{j\frac{4\pi}{\lambda} \left(\frac{\mathbf{b}}{r} s + v\mathbf{t} \right)} \right| \quad (3)$$

Only pixels for which C is above a fixed threshold are labeled as persistent scatterers and therefore the algorithm provides the temporal series as a product.

In the recent literature, the SqueeSAR approach has been proposed as an extension of the PSI technique to handle the impact of target decorrelation of distributed scatterers (Ferretti et al. 2011). SqueeSAR performs a phase filtering of the interferograms by exploiting the correlation matrix estimated from the data. SqueeSAR extends PSI to partially correlated (i.e., decorrelating) scatterers: the algorithm looks for a persistent scatterer (PS) mechanism equivalent to the real distributed scatterer, i.e., it assumes the response of the equivalent scatterer characterized only by phase variations with constant amplitude. Similarly to DInSAR stacking, SqueeSAR is tailored to the analysis of rural areas, however it does not perform a hard threshold on the baselines but rather it uses in a weighted way all the interferograms.

SAR Tomography System Model

SAR tomography is a step forward, respect to PSI: both are designed to monitor, at the full available spatial resolution, the deformation affecting ground scatterers. The key difference is however the different assumptions of the nature of the scattering which reflects into the interpretation of the received signal. As PSI processes only the phase information interferograms and therefore assumes the presence of a single scattering mechanism, SAR tomography removes this latter hypothesis and considers the complex value of each image pixel measured at the generic n th acquisition as the superposition of multiple elementary backscattering contributions distributed along the elevation s (Reigber and Moreira 2000; Gini et al. 2002; Fornaro et al. 2005). Differential tomography exploits the multitemporal characteristics to allow tomography to also monitoring deformation of scatterers. Particularly, a Fourier expansion of the deformation term $d(s, t)$ is introduced and, assuming the atmospheric phase delay caused by the propagation in atmosphere being compensated through a preliminary multipass DInSAR processing (f.i., SBAS), the measured signal at the generic antenna is therefore modeled as (Lombardini

2005; Fornaro et al. 2009a; Zhu and Bamler 2010a):

$$g_n = \iint_{I_s I_v} \gamma(s, v) e^{j \frac{4\pi b_n s}{\lambda r}} e^{j \frac{4\pi v t_n}{\lambda}} ds dv + w_n \quad (4)$$

which shows that, but for the noise term w_n , a 2D Fourier transform (FT) relationship stands between the data g_n and the backscattering distribution γ in the elevation/velocity (s, v) domain with respect to the Fourier conjugate variables $\xi_n = -2b_n/(\lambda r)$, associated with the elevation s , and $\eta_n = -2t_n/\lambda$, associated with the velocity v . In particular, as a consequence of the deformation term Fourier expansion, $\gamma(s, v)$ plays in v the role of the spectrum of the motion-related signal for elevation s . For linear deformation, the spectral velocity coincides with the deformation rate, i.e., $d(s, t_n) = vt_n$, and then v is usually referred to as deformation mean velocity, whereas for more complex motion, v identifies the (velocity) harmonic involved in the motion.

The tomographic problem consists of the estimation, in each image pixel, of the scene backscattering distribution $\gamma(s, v)$ starting from the N samples g_n and involves, in the most general case, the inversion of Eq. 4, that is, a spectral analysis of the data. This analysis moves the interferometric processing toward an imaging problem approach which extends the classical azimuth compression concepts widely known in the SAR 2D image focusing also for the third (elevation) dimension. Large antenna spans are coherently processed to achieve narrow responses and improve height resolution to the order of meters, allowing separating backscattering from source which are located at different heights. The scattering sources in the pixel can be spatially concentrated (compact scatterers), as for the layover in urban area where scatterers are typically located on the roofs and facades of building and interfere with those at lower heights, f.i., on the ground, or can be distributed along the elevation as for application in forest scenario where separation of ground level from canopy is of interest (Reigber and Moreira 2000; Cloude 2006; Tebaldini 2010).

The 4D model in Eq. 4 represents the most general imaging model which can be

particularized under specific conditions: in case of simultaneous acquisitions implying the absence of temporal diversity, as well as in case of absence of deformation, the signal model reduces to a 1D Fourier transform, and then the backscattering profile $\gamma(s)$ is of interest. This introduces to the 3D imaging framework. On the other hand, whereas uniform motion mostly applies for classical risk situations associated to slow, long-term deformation phenomena, more complex behavior can be also taken into account: the analysis of latest X-Band SAR data points out a higher sensitivity to small changes as those caused by the thermal dilation of materials (Reale et al. 2011a; Zhu and Bamler 2011). Moreover, since revisiting times are reduced with respect to the former generation of SAR sensors, the time requested to collect a sufficient number of images for reliable application of tomographic processing reduces to the order of 1 year, whereas typical C-Band ENVISAT and ERS acquisitions spanned a temporal interval of observation spanning several years. The reduced observation times may imply possible correlation with linear deformation behavior and impair the estimation of the deformation mean velocity (Reale et al. 2013). Even so, the sensitivity to thermal dilation can be exploited by SAR tomography as well. By extending the deformation model to account also for a second contribution linearly related to the average temperature T_n at the acquisition instants, i.e., $d(s, t_n) = vt_n + kT_n$, the order of the tomographic imaging can be extended (5D imaging) to estimate also a coefficient k which measures the expansion along the line-of-sight for each degree of temperature change (Zhu and Bamler 2011; Reale et al. 2013). In application to the monitoring of strategic infrastructures, this strategy allows estimating the stress induced by the temperature changes over the different segments of the structures (Fornaro et al. 2013).

SAR Tomography Imaging Algorithms

Multidimensional SAR imaging algorithms proposed in the literature typically work on

a discretized version of the model in Eq. 4: letting $\boldsymbol{\gamma} = [\gamma(s_0, v_0), \dots, \gamma(s_{M-1}, v_{M-1})]^T$ be the vector that collects the $M = M_s \times M_v$ samples of $\gamma(s, v)$ at the discrete points, hereafter called bins, (s_m, v_m) , with $m = 0, \dots, M-1$, belonging to the $M_s \times M_v$ elevation/velocity discretization grid (T defines the transposition operator), and $\mathbf{g} = [g_0, \dots, g_{N-1}]^T$ and $\mathbf{w} = [w_0, \dots, w_{N-1}]^T$ the vectors collecting, for each pixel, the measured complex data and the noise contribution at each acquisition, respectively, the FT operator in Eq. 4 can be rewritten in the discrete case as

$$\mathbf{g} = \mathbf{A}\boldsymbol{\gamma} + \mathbf{w} \quad (5)$$

where $\mathbf{A} = [\mathbf{a}_0, \dots, \mathbf{a}_{M-1}]^T$ is the $N \times M$ system matrix collecting the steering vectors associated with each discretization bin synthetically defined as $\mathbf{a}_m = \mathbf{a}(s_m, v_m)$, whose generic element is $[\mathbf{a}_m]_n = \exp[-j2\pi(\xi_n s_m + \eta_n v_m)]/\sqrt{N}$.

Several techniques can be used to implement the imaging, that is, the inversion of Eq. 5 that leads to the estimation of the backscattering distribution in the elevation/velocity plane (s, v) . Each is characterized by a different trade-off between simplicity, computational efficiency, sidelobes reductions, and super-resolution capability.

The beamforming (BF) technique represents the classical method to perform the inversion of Eq. 5: it makes use of the conjugate operator $\mathbf{A}^H$, with H being the Hermitian (conjugate transpose) operator, to profile the backscattering along the elevation bins (Fornaro et al. 2009a):

$$\hat{\boldsymbol{\gamma}} = \mathbf{A}^H \mathbf{g}. \quad (6)$$

Once the backscattering in the (s, v) plane has been estimated, scatterers are selected by looking for strong peaks in $\hat{\boldsymbol{\gamma}}$. In this context, a tool for the effective selection of reliable scatterers is required: since in real data the useful information is corrupted by noise, a detection stage is required to control the false alarm rate, defined as the probability to declare the presence of a scatterer, whereas the scatterer is not really present on the ground. With reference to the case of a single

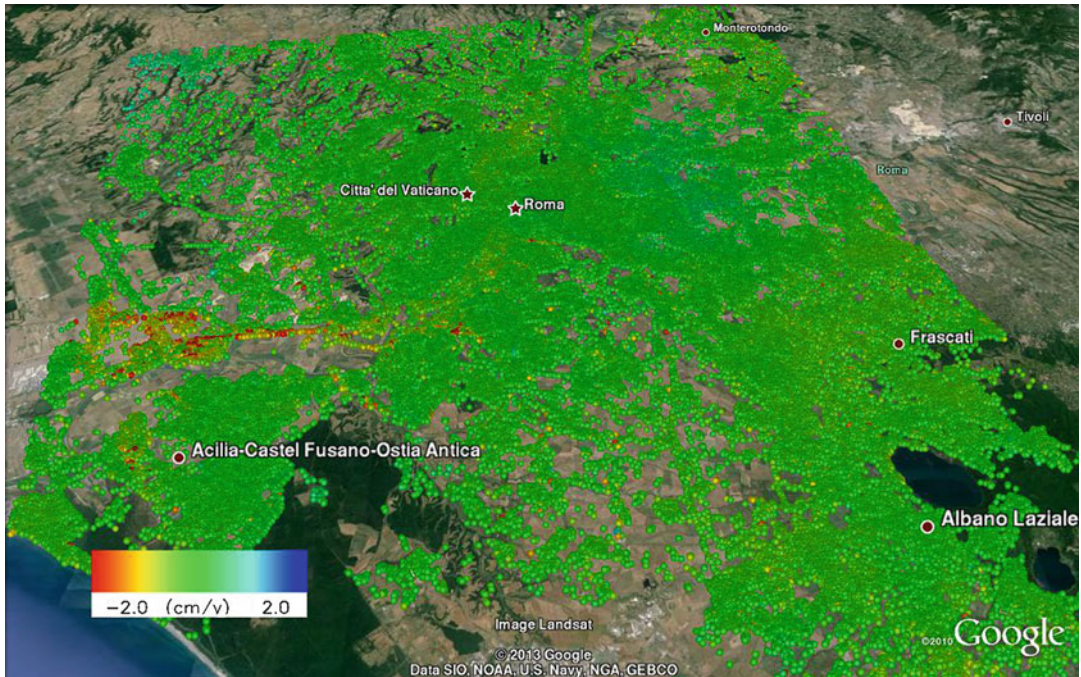
scatterer pixel, a test statistic based on the generalized likelihood ratio test has been proposed in the literature: it provides at the same time the maximum likelihood estimation of the (s, v) pair and declares the presence (hypothesis $\mathcal{H}_1$) or absence (hypothesis $\mathcal{H}_0$) of the scatterer for a given probability of false alarm by exploiting a test statistic which is strictly related to the BF reconstruction and is expressed as (De Maio et al. 2009):

$$\max_{(s,v)} \frac{|g^H a(s,v)|}{\|g\| \|a(s,v)\|} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\geq}} T \quad (7)$$

where T is the detection threshold, belonging to the $[0, 1]$ interval and set according to the desired level of false alarm. It is worth noting that for single scatterers, the test statistic represents the highest peak of the normalized BF reconstruction and the argument of the maximization is just the

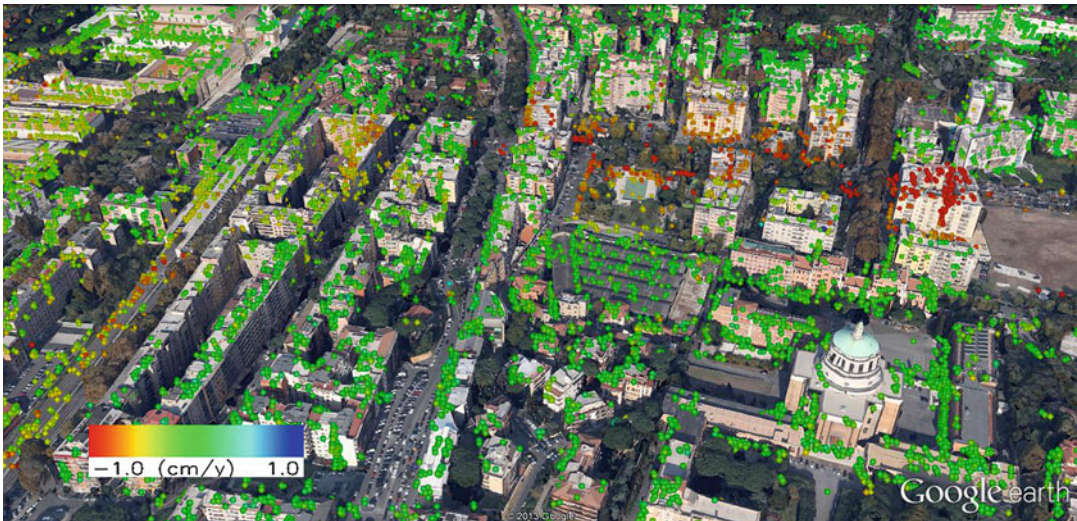
ML estimation of the (s, v) parameters. The GLRT in Eq. 7 allows also demonstrating the effectiveness of an imaging approach, as in SAR tomography, with respect to the classical interferometric processing. The PSI technique, in fact, exploits a similar test but for the use of only the phase information of each element of the vector g (Ferretti et al. 2000, 2001): the exploitation of the whole complex, amplitude plus phase, data as in Eq. 7 provides a significant increase, with respect to PSI, of the detection probability for a fixed false alarm rate.

Typical products achieved by the sequence of low-resolution DInSAR stacking techniques for the derivation of coarse-scale deformation and calibration of data for the subsequent full resolution processing is reported in Figs. 4 and 5. The processing involved a dataset of 29 Cosmo-SkyMed images acquired over the city of Rome, Italy, from April 2011 to October 2012 in H-image mode which provides images with 3 m



SAR Tomography for 3D Reconstruction and Monitoring, Fig. 4 Small-scale deformation mean velocity map estimated by the ESD technique on a dataset of 29 Cosmo-SkyMed images over the city of Rome, Italy.

Colors are set on the estimated velocity and correspond to movements toward the sensor (uplift) in blue and away (subsidence) in red. Optical image courtesy of Google



SAR Tomography for 3D Reconstruction and Monitoring, Fig. 5 3D visualization of the scatterers, represented as *dots*, detected by the GLRT after 4D beamforming reconstruction. *Colors* are set according to

the estimated deformation mean velocity. The area is relevant to the Grotta Perfetta area in Rome, close to the Tevere River. Optical image courtesy of Google

of spatial resolution. According to the SBAS approach, a total number of 80 interferograms has been generated by imposing maximum spatial and temporal baseline spans of 1,000 m and 150 days, respectively. A multilook has also been carried out through the use of a 16×16 pixels moving average filter followed by a subsampling of the same factor. In Fig. 4 it is shown the deformation mean velocity map, superimposed to a Google Earth image, corresponding to the slope of the deformation time series for each pixel selected by looking at a quality index measuring the temporal consistency of measurements after phase unwrapping. As for the L'Aquila dataset, the low-resolution processing has been carried out through the ESD technique. Colors in the map move again from blue (uplift) to red (subsidence). The results highlight the presence of a distributed deformation pattern in the western part of the city which largely affects also the Rome airport. The large number of detected scatterers can be appreciated, although visualization constraints of the Google Earth environment impose a reduction of the total number of points, in this case of a factor 5. The dynamic range also

hides some deformation signals occurring in the central part of the city, which is of interest for the application of full resolution tomographic analysis provided in a following section.

Figure 5 provides a close view, in a Google Earth framework, of the results of the application of 4D imaging and subsequent GLRT detection on the Cosmo-Skymed dataset over Rome introduced before. It allows attesting the capabilities of imaging approaches in the single building monitoring. The area, in the Rome city center, is affected by a severe subsidence induced by the consolidation of the alluvial sediments of the Tevere river. Each dot corresponds to a detected scatterer after geocoding from the original radar geometry to the natural geographic coordinates system: the correct estimation of heights allows effectively positioning scatterers, as they overlap the 3D models provided by Google Earth. Furthermore, the colors correspond to deformation velocity rate estimated for each scatterer: several buildings in the area suffer for large subsidence rates which cause extensive structural damages affecting their stability (Arangio et al. 2013).

The detection problem can be straightforwardly extended to the case of multiple scatterers per resolution cell, which may occur in layover areas. In this framework, a detector which tests the presence of possibly a maximum of two scatterers based on a sequential use of the GLRT discussed above has been also presented (Pauciuolo et al. 2012). With reference to the separation of interfering scatterers within the same image pixel, in application to a real acquisition scenario, the acquisition geometry poses limitation on the imaging capabilities of BF: the baseline distribution defines the unambiguous elevation range that is the maximum extension allowed in elevation direction to avoid aliasing phenomena. Given Δb_n be the spatial separation between the successive antennas, the unambiguous elevation interval will be $\Delta_s = \lambda r / (2\Delta b)$, where Δb is the average value of Δb_n in case of nonuniform baseline distribution. Moreover, the final elevation resolution of BF cannot exceed the Rayleigh resolution given by $\Delta_s = \lambda r / (2B)$ with B being the total baseline span. In the same way, unambiguous velocity interval and velocity resolution can be also defined as $\Delta_v = \lambda / (2\Delta t)$ and $\Delta_v = \lambda / (2T)$, where Δt and T are the average temporal separation and the total temporal span, respectively. The elevation resolution plays a key role in the capability of distinguishing possible multiple scattering contributions within the same pixel: it implies the minimum separation of scatterers as they can be distinguished as separate scatterers. Latest satellite SAR missions as the TerraSAR-X are characterized by a narrow radius of the orbital tube, leading to small baseline excursions that result in final height resolution of some tens of meters. In application to urban areas, such a resolution can be very restrictive for the solution of the interference due to the layover. Finally, in all the practical applications, baselines are far from being uniformly distributed, thus BF gives poor reconstruction performances in terms of sidelobes and leakage in the (s, v) point spread function. Alternative strategies can be adopted to improve both the resolution performances and also the quality of the reconstruction of $\hat{\gamma}$ in the

presence of highly uneven baseline distributions and at the same time achieve some super-resolution capabilities, i.e., the possibility to push the height resolution below the inherent Rayleigh limit.

The use of the singular value decomposition (SVD) of the operator $\mathbf{A}$ in Eq. 5 allows regularizing the inversion by restricting the solution space and benefiting of the inclusion of very limited a priori information on the expected scene elevation extent. The regularization, obtained through the so-called Truncated SVD, allows avoiding noise amplification and inversion instabilities and hence generally provides a better sidelobe reduction and as well as slight super-resolution than plain BF (Fornaro et al. 2009a).

Compressed sensing (CS) is a recent technique used in linear inversion problems for signal recovery that takes benefit of the hypothesis that the signal to be reconstructed have (in some basis) a sparse representation, i.e., a small number of nonzero entries. Under certain assumptions of the measurement matrix, the signal can be reconstructed from a small number of measurements. SAR tomography in urban areas is a favorable application scenario for CS due to the fact that for typical operative frequencies, the scattering occurs only on some scattering centers associated with ground, facades, and roofs of ground structures (Zhu and Bamler 2010b; Budillon et al. 2011).

Summary

Available high-resolution synthetic-aperture radar imaging sensors are capable of providing, in a systematic regular basis, images of single buildings and ground targets with very high spatial details. Persistent Scatterers Interferometry techniques have been already shown to largely benefit of the use of very high-resolution data to monitor buildings. Despite the spatial resolution increase, the steepness of the topography corresponding to vertical structured targets

generates critical distortion effect, the most critical ones are shadow and layover. These are major impairing sources in the analysis of SAR images corresponding to urban areas. SAR tomography represents a powerful technique which allows the implementation of a radar scanner from the space with large 2D antennas to scan the details of building and overcoming problems of layover to generate dense point-cloud measurements of buildings. Multidimensional SAR imaging, based on the concept of SAR tomography, is a tool that represents the most advanced method in the 3D reconstruction and monitoring of buildings. PSI and SAR tomography with very high-resolution sensors provide a unique tool for application of spaceborne microwave radar imaging to urban areas, which is expected in the near future to play a key role in vulnerability and damage assessment of buildings and infrastructures.

Cross-References

- [Building Monitoring Using a Ground-Based Radar](#)
- [InSAR and A-InSAR: Theory](#)
- [SAR Images, Interpretation of](#)
- [Urban Change Monitoring: Multi-temporal SAR Images](#)

References

- Arangio S, Calò F, Di Mauro M, Bonano M, Marsella M, Manunta M (2013) An application of the SBAS-DInSAR technique for the assessment of structural damage in the city of Rome. *Struct Infrastruct Eng*. doi: 10.1080/15732479.2013.833949
- Bamler R, Hartl P (1998) Synthetic aperture radar interferometry. *Inverse Probl* 14:1–54
- Berardino P, Fornaro G, Lanari R, Sansosti E (2002) A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms. *IEEE Trans Geosci Remote Sens* 40:2375–2383
- Budillon A, Evangelista A, Schirinzi G (2011) Three-dimensional SAR focusing from multipass signals using compressive sampling. *IEEE Trans Geosci Remote Sens* 49:488–499
- Carnec C, King C, Massonnet D (1995) Measurement of land subsidence by means of differential SAR interferometry to sites of small extent. *Geophys Res Lett* 23:3579–3582
- Cascini L, Ferlisi S, Peduto D, Fornaro G, Manunta M (2007) Analysis of DInSAR Data via geotechnical criteria. *Rivista Italiana di Geotecnica* 4:50–67
- Cloude SR (2006) Polarization coherence tomography. *Radio Sci* 41:RS4017. doi:10.1029/2005RS003436
- Crosetto M, Crippa B, Biescas E (2005) Early detection and in-depth analysis of deformation phenomena by radar interferometry. *Eng Geol* 79:81–91
- Curlander JC, McDonough R (1991) *Synthetic aperture radar – system and signal processing*. Wiley, New York
- D’Agostino N, Cheloni D, Fornaro G, Giuliani R, Reale D (2012) Space-time distribution of afterslip following the 2009 L’Aquila earthquake. *J Geophys Res* 117: B02402. doi:10.1029/2011JB008523
- De Maio A, Fornaro G, Pauciuolo A (2009) Detection of single scatterers in multi-dimensional SAR imaging. *IEEE Trans Geosci Remote Sens* 47:2284–2297
- Ferretti A, Prati C, Rocca F (2000) Nonlinear subsidence rate estimation using permanent scatterers in differential SAR interferometry. *IEEE Trans Geosci Remote Sens* 38:2202–2212
- Ferretti A, Prati C, Rocca F (2001) Permanent scatterers in SAR interferometry. *IEEE Trans Geosci Remote Sens* 39:8–20
- Ferretti A, Fumagalli A, Novali F, Prati C, Rocca F, Rucci A (2011) A new algorithm for processing interferometric data-stacks: *squeeSAR*. *IEEE Trans Geosci Remote Sens* 49:3460–3470
- Fornaro G, Franceschetti G (1999) SAR interferometry, Chapter IV. In: Franceschetti G, Lanari R (eds) *Synthetic aperture radar processing*. CRC Press, Boca Raton
- Fornaro G, Serafino F (2006) Imaging of single and double scatterers in urban areas via SAR tomography. *IEEE Trans Geosci Remote Sens* 44:3497–3505
- Fornaro G, Serafino F, Lombardini F (2005) 3D multipass SAR focusing: experiments with long-term spaceborne data. *IEEE Trans Geosci Remote Sens* 43:702–712
- Fornaro G, Reale D, Serafino F (2009a) Four-dimensional SAR imaging for height estimation and monitoring of single and double scatterers. *IEEE Trans Geosci Remote Sens* 47:224–237
- Fornaro G, Pauciuolo A, Serafino F (2009b) Deformation monitoring over large areas with multipass differential SAR interferometry: a new approach based on the use of spatial differences. *Int J Remote Sens* 30:1455–1478
- Fornaro G, Reale D, Verde S (2013) Bridge thermal dilation monitoring with millimeter sensitivity via multidimensional SAR imaging. *IEEE Geosci Remote Sens Lett* 10:677–681

- Gernhardt S, Adam N, Eineder M, Bamler R (2010) Potential of very high resolution SAR for persistent scatterer interferometry in urban areas. *Ann GIS* 16:103–111
- Ghiglia DC, Pritt MD (1998) Two-dimensional phase unwrapping: theory, algorithms, and software. Wiley-Interscience, New York
- Gini F, Lombardini F, Montanari M (2002) Layover solution in multibaseline SAR interferometry. *IEEE Trans Aerosp Electron Syst* 38:1344–1356
- Krieger G, Moreira A, Fiedler H, Hajnsek I, Werner M, Younis M, Zink M (2007) TanDEM-X: a satellite formation for high-resolution SAR interferometry. *IEEE Trans Geosci Remote Sens* 45:3317–3341
- Lombardini F (2005) Differential tomography: a new framework for SAR interferometry. *IEEE Trans Geosci Remote Sens* 43:37–44
- Massonnet D, Rossi M, Carmona C, Adragna F, Peltzer G, Fiegl K, Rabaute T (1993) The displacement field of the Landers earthquake mapped by radar interferometry. *Nature* 364:138–142
- Moreira A, Prats-Iraola P, Younis M, Krieger G, Hajnsek I, Papathanassiou KP (2013) A tutorial on synthetic aperture radar. *IEEE Geosci Remote Sens Mag* 1:6–43
- Pauciullo A, Reale D, De Maio A, Fornaro G (2012) Detection of double scatterers in SAR tomography. *IEEE Trans Geosci Remote Sens* 50:3567–3586
- Reale D, Fornaro G, Pauciullo A, Zhu X, Bamler R (2011a) Tomographic imaging and monitoring of buildings with very high resolution SAR data. *IEEE Geosci Remote Sens Lett* 8:661–665
- Reale D, Nitti DO, Peduto D, Nutricato R, Bovenga F, Fornaro G (2011b) Post-seismic deformation monitoring with the COSMO/SKYMED constellation. *IEEE Geosci Remote Sens Lett* 8(4):696–700
- Reale D, Fornaro G, Pauciullo A (2013) Extension of 4-D SAR imaging to the monitoring of thermally dilating scatterers. *IEEE Trans Geosci Remote Sens* 51:5296–5306
- Reiger A, Moreira A (2000) First demonstration of airborne SAR tomography using multibaseline L-band data. *IEEE Trans Geosci Remote Sens* 38:2142–2152
- Tebaldini S (2010) Single and multipolarimetric SAR tomography of forested areas: a parametric approach. *IEEE Trans Geosci Remote Sens* 48(5):2375–2387
- Van Zyl J (2001) The Shuttle Radar Topography Mission (SRTM): a breakthrough in remote sensing of topography. *Acta Astronaut* 48:559–565
- Zhu X, Bamler R (2010a) Very high resolution spaceborne SAR tomography in urban environment. *IEEE Trans Geosci Remote Sens* 48:4296–4308
- Zhu X, Bamler R (2010b) Tomographic SAR inversion by L1 norm regularization – the compressive sensing approach. *IEEE Trans Geosci Remote Sens* 48:3839–3846
- Zhu X, Bamler R (2011) Let's do the time warp: multicomponent nonlinear motion estimation in differential SAR tomography. *IEEE Geosci Remote Sens Lett* 8:735–739

School Seismic Safety and Risk Mitigation

Marla Petal¹, Ben Wisner^{2,3}, Ilan Kelman^{4,17}, David Alexander², Omar-Dario Cardona⁵, Djillali Benouar⁶, Sanjaya Bhatia⁷, Jitendra Kumar Bothara^{8,9}, Amod Mani Dixit⁸, Rebekah Green¹⁰, Ram Chandra Kandel¹¹, Tracy Monk¹², Bishnu Pandey¹³, Janise Rodgers¹⁴, Zeynep Türkmen Sanduvaç¹⁵ and Rajib Shaw¹⁶

¹Risk RED (Risk Reduction Education for Disasters), Los Angeles, CA, USA

²University College London, London, England

³Oberlin College, Oberlin, OH, USA

⁴Institute for Risk and Disaster Reduction and Institute for Global Health, University College London, London, England

⁵Universidad Nacional de Colombia, Manizales, Colombia

⁶University of Bab Ezzouar, Algiers, Algeria

⁷UNISDR Recovery Platform, Kobe, Japan

⁸NSET, Kathmandu, Nepal

⁹Miyamoto Impact, Christchurch, New Zealand

¹⁰University of Western Washington, Bellingham, WA, USA

¹¹Toronto, ON, Canada

¹²Families for School Seismic Safety, Vancouver, Canada

¹³University of British Columbia, Vancouver, Canada

¹⁴GeoHazards International, Menlo Park, USA

¹⁵Risk RED, Istanbul, Turkey

¹⁶University of Kyoto, Kyoto, Japan

¹⁷Norwegian Institute of International Affairs, Oslo, Norway

Synonyms

Comprehensive school safety; Critical infrastructure; Safe school construction; Safe school facilities; School construction

Introduction

Access to education is a basic human right. It is enshrined in Convention on the Rights of the

Child (1990), the World Declaration on Education for All (in 1990), and the World Education Forum (WEF 2000). It is one of the Millennium Development Goals for the decade starting in 2005 and continues to be part of the “post-2015” development agenda. Education is strongly associated with poverty reduction, and there are strong global and national drives to implement it. The Global Partnership for Education has 29 national partners, supporting the implementation of universal, free, quality basic education in 57 partner developing countries. In GPE’s 2012–2015 Strategic Plan, the first of its four strategic goals is “All children have access to a safe, adequately equipped space to receive an education with a skilled teacher.” However, none of its monitoring indicators mention safety. Since 2004, the GPE has contributed to build, rehabilitate, and equip close to 53,000 classrooms (GPE 2014). However, up until at least 2013, there was no systematic due diligence with respect to disaster-resilient construction. In the rush to fulfill the right to education, are children being put at risk?

This entry assesses seismic threats to schools and reviews incidents of children and teachers killed by structural failure of school buildings as well as structural damage to schools and near misses. It reviews progress, good practices, and lessons learned based on these threats. The entry goes on to overview school vulnerability, global experiences in seismic-resistant school construction, and retrofit. A summary of progress in school seismic safety along with a recommended systematic all-hazards approach to comprehensive school safety set the stage to make the case for continued advocacy for school seismic safety.

The Threat

According to estimates made by the Center for International Earth Science Information Network at Columbia University in 2006, there are more than 100 million school-aged children exposed to significant seismic risk throughout the world (New York Times 2008). In 2004, 10 of the 16 contributors to this article initiated recording

of case studies on progress and struggles for school seismic safety which are updated in the entry “► [School Seismic Safety: Case Studies](#).” They set out the magnitude of concern and main arguments for advocacy in an unpublished article. At the time, the authors posited the gruesome estimate of “4,800 school children killed by earthquake-related school collapse or severe damage over the next decade. . . It might be reasonable and prudent to plan to avoid a loss of student life in earthquakes of somewhere between 2,000 and 5,000 in a 10 year period.” At the time it was written, this estimate seemed to the authors to be somewhat alarmist.

The following year, shortly after the unprecedented destruction caused by the Indian Ocean earthquake and tsunami, 168 countries agreed to the 2005–2015 Hyogo Framework for action. Over the course of this 10 year period, this dire predication has been exceeded fourfold as the result of only two major earthquakes during school hours: the 8 October 2005 Kashmir earthquake which killed more than 18,000 students, in addition to staff, in schools, and the 12 May 2008 Sichuan earthquake which killed more than 5,300 students, in addition to staff, in their schools (UNISDR 2008).

In the powerful earthquake and massive tsunami on 11 March 2011 in northern Japan, schools themselves were by and large structurally sound and resisted earthquake damage, but tsunami-retaining walls were breached as the tsunami was larger than expected and land subsidence had not been factored in. Disaster drills and practice of “tendenko” (automatic tsunami evacuation), by many school children, saved many lives. Some schools provided vertical evacuation, and many survived at evacuation and shelter centers. But instances of confusion occurred and many school pupils and teachers also died. Today, the students now displaced by the resulting nuclear disaster recognize this neglected threat as the most catastrophic of all. The international community is virtually silent on this threat.

In common with other infrastructure, school buildings are subject to damage and collapse in earthquakes. Many of these have resulted in children killed while being educated (Table 1).

Pictorial evidence of historic earthquake damage to schools is available in the NISEE, Earthquake Engineering Online Archive.

There have also been many cases when an earthquake destroyed school buildings when they were not in session, and thus deaths and injuries were narrowly avoided (Table 2). However the severe impact on continuity of education and the potential magnitude of loss of life in these events further highlight the

importance of ensuring the seismic safety of schools.

Making the Case for School Seismic Safety

Many public buildings and different sorts of critical infrastructure are threatened by earthquakes. The case can be made for giving priority to

School Seismic Safety and Risk Mitigation, Table 1 Children killed by structural failure of school buildings

Date/local time (Source)	Location/magnitude ^a	Consequences/schools	Consequences/children
12 Jan 2010 16:53 (CNN 2014)	Port-au-Prince, Haiti M 7.0	MoE estimates 4,992 schools affected (23 % of the nation's schools)	Deaths and injuries unknown. Many children with disabling injuries. Some schools were holding their third shifts. Est. 1.3 m children and youth affected
12 May 2008 14:28 (COGGS 2008)	Wenchuan, China M.7.9	175 schools (7,000 classrooms) in Sichuan and Shaanxi provinces were destroyed	>5,300 school children died in dozens of schools In the Beichuan Middle school, 1,300 of 2,999 students and teachers died
6 Mar 2007 11:00 (COGGS 2008)	Western Sumatra M 6.4	The wall of a primary school collapsed. Fire followed. Up to 329 schools affected by several earthquakes (2005–2010)	4 primary school children died
8 Oct 2005 St. 08:50 (UNISDR 2008)	Kashmir, Pakistan, and India M 7.6	More than 10,000 schools collapsed 80 % of Maheshehra's 2,749 66 % of Batagram's 678, and 37 % of Abbottabad's 1,829 public schools were destroyed or seriously damaged	>18,000 school children died >50,000 school children were seriously injured
1 May 2003 03:20 (Rodgers 2012)	Bingöl, Turkey M 6.4	4 school buildings collapsed. Only the dormitory was occupied	84 students killed and 114 survived in the dormitory
24 February 2003 10:03 (COGGS 2008)	Bachu, Xinjiang, China M 6.4	900 classrooms collapsed	Students were outside in physical education at the time of the earthquake. At least 20 students killed in one middle school collapse
31 October 2002 11:40 (COGGS 2008)	San Giuliano di Puglia, Molise, Italy M 5.9	San Giuliano infant school collapsed	26 children and 3 adults killed. 35 children rescued alive from the building but some reports suggest that one child died later
26 January 2001 Friday 08:16 Republic day holiday (COGGS 2008)	Gujarat, India M 7.6	1,884 school buildings collapsed. 5,950 classrooms destroyed. 36,584 unfit for instruction	971 school children and 31 teachers were killed in school activities. 1,051 students and 95 teachers seriously injured 32 children died at Swaminarayan School

(continued)

School Seismic Safety and Risk Mitigation, Table 1 (continued)

Date/local time (Source)	Location/magnitude ^a	Consequences/schools	Consequences/children
13 February 08:22 and 13 January 2001 (COGGS 2008)	El Salvador M 6.6	85 schools damaged beyond repair. In aftershock 22 preschoolers and their teacher were killed	50 % of fatalities were children
9 July 1997 15:24 (COGGS 2008)	Cariaco, Venezuela M 7.0	Two out of five school buildings collapsed. Four reinforced concrete buildings had serious structural defects	46 students killed
10 May 1997 12:57 (COGGS 2008)	Ardekul, Iran M 7.3	Elementary school collapsed	110 young girls were killed
1992 (COGGS 2008)	Erzincan, Turkey M 6.9	6-story medical school collapsed	62 students were killed.
7 December 1988 11:41 (COGGS 2008)	Spitak, Armenia M 6.8	380 children and youth institutions destroyed. 105 of 131 in Spitak and Leninakin destroyed	Likely thousands of school children killed. At least 400 children died in the collapse of a Dzhrashen elementary school
27 July 1976 03:42 (COGGS 2008)	Tangshan, China M 7.8	Most school buildings destroyed	2,000 students killed in the dormitory of the College Mining Institute
13 April 1949 11:58 (COGGS 2008)	Olympia, Washington, USA M 7.1	10 schools destroyed, 30 damage. Spring break	2 children in school were killed
31 October 1935 (COGGS 2008)	Helena, Montana, USA M 6.2	Newly built secondary school wing collapsed	2 students killed. Classes not in session, could have been much worse
10 March 1933 Long Beach (COGGS 2008)	Long Beach, California, USA M 6.4	70 schools destroyed. 120 with major damage. Classes held in tents for 2 years. First legislation for safe school construction	2 children died in gymnasium collapse. Spring break, classes not in session, could have been much worse

^aMagnitudes from USGS or Wikipedia

schools from three perspectives: Duty bearers have moral and legal obligations to fulfill children's rights to both safety and survival and educational continuity. In more affluent countries, the cost benefits of investments in public safety, the importance of safeguarding development investments, and preventing educational disruption are undisputed. And, the uses of school buildings as multipurpose community centers and disaster shelters, even when children are not harmed, have cascading social and economic

consequences beyond the replacement cost of school buildings themselves. In most cases public discussion and debate on these issues tend to mix these ethical and pragmatic arguments.

Human Rights Argument

The human rights argument suggests that no society should tolerate a choice between the safety of children's lives and their education. The right to

School Seismic Safety and Risk Mitigation, Table 2 School structural damage from earthquakes

Date (Source)	Location/magnitude ^a	Structural and educational impacts
2 July 2013 14:37 (Pandey 2013)	Aceh, Indonesia M 6.2	966 schools affected
4 April 2010 15:40 (Rodgers 2012)	California, USA, and Baja, Mexico M 7.2	Significant structural damage to several schools in Mexico. Significant nonstructural damage to several schools in the USA. Cost of repair almost 20 % of annual budget for one school district. School was on spring break. Nonstructural damage would have caused injuries and blocked egress. In California hazardous asbestos from collapsed walkways and mercury from light fixtures closed schools for extended periods
30 September 2009 17:16 (Rodgers 2012)	Padang, West Sumatra, Indonesia M 7.6	2,164 severely damaged, 1,447 moderately damaged, 1,137 lightly damaged School was recently dismissed for the day. Temporary school buildings of timber frame and corrugated steel
2 September 2009 14:55 (Pandey 2013)	West Java, Indonesia M 7.0	716 schools affected
21 September 2009 14:53 (Rodgers 2012)	Mongar, Bhutan M 6.1	91 schools affected: 6 destroyed, 17 required major repair, 44 required partial repair, 24 required minor repair (cost 12.9 m USD) plus damage to boarding schools, water, and sanitation School was dismissed early for holiday. Temporary learning facilities inadequate for weather
6 April 2009 03:32 (Rodgers 2012)	L'Aquila, Italy M 6.3	78 schools had extended closures and 12 partial closures
12 September 2007 18:10 (Pandey 2013)	Bengkulu, Indonesia M 8.5	240 schools affected (2005–2010)
15 August 2007 23:40 (Rodgers 2012)	Pisco, Peru M 8.0	116 schools were severely damaged. 478 classrooms were needed to restore school activities
27 May 2006 05:53 (Rodgers 2012)	Yogyakarta, Indonesia M 6.3	Yogyakarta: 2155 educational facilities damaged or destroyed; Central Java: 752 damaged or destroyed. Damage and losses estimated at 1.7 trillion Indonesian Rupiah
26 December 2004 early am (UNISDR 2008)	Indonesia, Sri Lanka, Maldives, Thailand M 9.1–9.3	School earthquake and tsunami damage combined: Indonesia – 750 destroyed 2,135 damaged. Sri Lanka – 51 destroyed, 100 damaged. Maldives – 44 destroyed or damaged. Thailand – 30 destroyed
26 December 2003 5:26 (COGGS 2008)	Bam, Iran M 6.6	67 of 131 schools collapsed. The remaining 64 were heavily damaged and unusable 33,000 students were affected
22 September 2003 12:45 (UNISDR 2008)	Puerto Plata, Dominican Republic M 6.4	50 public schools damaged, 140 classrooms impacted 18,000 students were without classrooms
21 May 2003 19:48 (COGGS 2008; OECD 2004)	Boumerdes, Algeria M 6.8	130 schools damaged beyond repair. 753 schools extensively damaged or destroyed The earthquake occurred out of normal school hours, so children were not at school. Cost of school rehabilitation \$79 million+
24 February 2003 (COGGS 2008)	Xinjiang, China	Dozens of schools collapsed The earthquake struck 27 minutes before thousands of children would have been in classrooms
25 April 2002 22:41 (Rodgers 2012)	Tbilisi, Georgia M 4.5	Approximately \$8 million US in school damage. No collapses; 1 school with very heavy damage; 35 with substantial damage; 68 with moderate damage; 98 with negligible or slight damage

(continued)

School Seismic Safety and Risk Mitigation, Table 2 (continued)

Date (Source)	Location/magnitude ^a	Structural and educational impacts
21 September 1999 1:47 (COGGS 2008)	Chi-Chi, Taiwan M 7.7	51 schools collapsed. 786 schools nationwide were damaged. 22 % of schools and 71 % of post-secondary institutes damaged The earthquake happened in the middle of the night, so no one was in the building. Cost of repair and reconstruction \$1.3 billion
June 23 2001 15:33 (COGGS 2008)	Arequipa, Peru	98 school buildings seriously damaged School not in session on Saturday
17 August 1999 3:02 (COGGS 2008)	Kocaeli, Turkey M 7.6	43 schools were damaged beyond repair. 381 minor to moderate damage In Istanbul 60 km away 35 schools were unsafe and demolished School was not in session but was suspended for 4 months. In Istanbul 131 schools were closed temporarily, for inspection
25 January 1999 13:19 (OECD 2004)	Pereira and Armenia, Colombia M 6.2	74 % of schools in Pereira and Armenia were damaged School was not in session
20 August 1998 (UNISDR 2008)	Udayapur, Eastern Nepal M 6.6	1,200 schools heavily damaged or destroyed. 6,000 affected
17 July 1998 00:19	Papua, New Guinea M 7.0	Schools destroyed
9 July 1998 5:19 (COGSS 2008)	Faial, Azores, Portugal M 6.2	Schools damaged School was not in session
20 May 1998	Afghanistan/Tajikistan M 6.6	Unknown
12 November 1996 15:33 (COGSS 2008)	Nazca, Peru	93 school buildings seriously damaged
1996 (OECD 2004)	Temouchent, Algeria M 5.6	6 schools destroyed, 17 moderate damage, 36 light damage
17 January 1995 5:46 (COGSS 2008)	Hanshin-Awaji, Japan M 6.9	54 buildings school damaged beyond repair. Extensive earthquake and fire damage to 4,500 educational buildings. ¥94 billion. School was not in session
1994 (OECD 2004)	Beni Chourgrane, Aleeria M 5.6	4 schools destroyed, 16 moderate damage, 30 light damage
17 January 1994 04:31 (FEMA 2011)	Northridge, California, USA M 6.7	24 of 127 affected schools suffered significant structural damage. Suspended lighting and ceiling systems were damaged in 1,500 buildings 2, 204 schools were used as shelters. Had this occurred during the school day, significant injuries and lack of safe egress for thousands would have resulted. The Los Angeles Unified School District, amongst others, embarked on projects for nonstructural mitigation, now the responsibility of school maintenance personnel
25 March 1993 (COGSS 2008)	Scott Mills, Oregon, USA	Part of masonry school building collapsed Spring break, school was not in session
17 October 1989 17:04 (EERI 1990)	Loma Prieta, California, USA M 6.9	7 schools in three districts and one headquarters sustained severe damage. 1,544 were schools surveyed. Total value of damage \$81 million USD
10 October 1989 (OECD 2004)	El Asnam, Algeria M 7.3	70–85 schools suffered extensive damage or collapsed The earthquake occurred out of normal school hours, so children were not at school
20 August 1988 4:39 (COGSS 2008)	Bihar, India, and Udaypur Nepal M 6.6	950 school buildings were damaged in Bihar 6,000 schools destroyed in Nepal School was not in session

(continued)

School Seismic Safety and Risk Mitigation, Table 2 (continued)

Date (Source)	Location/magnitude ^a	Structural and educational impacts
8 November 1988 (UNISDR 2008)	Yunan, China	1,300 schools destroyed in earthquake
19 September 1985 :17 (COGSS 2008)	Mexico City, Mexico M 8.0	137 school buildings collapsed, 1,687 school buildings were damaged Schools were not yet open
2 May 1983 23:42 (COGSS 2008)	Coalinga, California, USA	Extensive nonstructural damage noted
10 October 1980 13:25 (OECD 2004)	El Asnam, Algeria M 7.3	70 schools totally destroyed, 25 moderately damaged School was not in session
9 February 1971 06:01 (State of California 2009)	Sylmar, California, USA M 6.6	Only 4 of 1,544 buildings surveyed suffered severe damage. Nearly all damage was nonstructural School was not in session
31 May 1970 4:23 (COGSS 2008)	Chimbote, Peru M 7.9	6,730 classrooms collapsed and hundreds seriously damaged
27 March 1964 (COGSS 2008)	Alaska, USA	Primary school destroyed by an earthquake-induced landslide. Half of Anchorage's schools were significantly damaged The earthquake struck on a holiday, Good Friday, so schools were closed
1963 (COGSS 2008)	Skopje, Macedonia	44 schools (57 % of urban stock) were damaged 50,000 students affected. Sunday, school not in session
21 July 1952 4:52 (COGSS 2008)	Kern County, California, USA M 7.3	20 schools damaged or destroyed (most built before 1933). Significant nonstructural damage also noted
4 March 1952 (USGS 2003)	Sapporo, Japan	400 schools collapsed in Sapporo
10 March 1933 (COGSS 2008)	Long Beach, California, USA	70 schools collapsed The earthquake hit early in the evening after children had left for the day which saved their lives. Five students were killed in a gymnasium
3 February 1931 (Dowrick and Rhoades 2004)	North Island, New Zealand	Several schools were severely damaged The earthquake happened at mid-morning during school playtime when the children were outdoors enjoying the summer weather. Some students were killed, but the death toll could have been several hundreds
17 June 1929 10:17	Murchison, New Zealand	College tower and dormitory roofs collapsed School was not in session
18 April 1906 05:12	San Francisco, California USA	28 schools burned in fire. 41 schools damaged or destroyed Classes were not in session

^aMagnitudes from USGS

Bibliographic references on many structural impacts on schools are available on the internet (Rodgers 2012)

life and the right to education are both recognized human rights, and both should be met. This argument takes on additional salience in view of the current international effort to increase school enrollment and attendance by girls, disabled children, and children of the very poor and marginalized groups in society.

Around the world, at least 100 million children of school age do not attend school representing about 14 % of the world's children (UNESCO 2004). Providing facilities to educate them requires construction of schools and rapid expansion of building programs. The Education for All campaign originally

hoped to enroll 24 million of these children in a decade. Millennium Development Goals (MDGs) specifically aim to “[e]nsure that, by 2015, children everywhere, boys and girls alike, will be able to complete a full course of primary schooling.”

In 2004 it was estimated that more than 7,500 new schools were needed within the next 3 years solely in Afghanistan, a country with a significant seismic hazard. It would be ironic and tragic if in the course of achieving one MDG, another is undermined.

Another MDG target is to reduce the under-five mortality rate by two thirds, between 1990 and 2015. On the one hand, the international community is seeking to save the lives of under-fives, only to put them at risk a few years later when they go to school.

Educational authorities charged with the construction and maintenance of schools are also the ones tasked with many other functions: They develop curricula, hire teachers, and choose educational resources such as textbooks and computers. School safety issues have to find a place in the capital, maintenance, and operation budgets of school buildings and school operation.

Retrofitting schools for seismic safety can be perceived to compete for funds with the rest of the educational process. The question facing decision makers can actually appear to be: “What is more important: an up-to-date textbook and good laboratory facilities now or a building that can withstand an extreme event which might or might not occur with the next few decades?”

Under most circumstances, young people do not lobby for their own rights to health and safety. Children cannot refuse to go to school because a building is unsafe. By law, they must attend school, though teachers, parents, and others may advocate on their behalf. Faculty and support staff in schools should also be concerned for their occupational safety and theoretically be natural advocates of safe school facilities. Yet there are no examples mentioned to date of teachers unions becoming involved in the issue of school disaster vulnerability.

Cost-Effectiveness Arguments

There are two forms that cost-effectiveness arguments may take. One asserts that the authority responsible for education incurs greater cost in the long run to repair and replace schools damaged by earthquakes than the cost of enforcing building codes and making sure that every new school is a safe school (or even of retrofitting older or poorly built schools). In some cases, replacement of unsafe schools is more cost-effective than repair (e.g., see entry “► [School Seismic Safety: Case Studies](#),” for examples from Algeria, Colombia, and Turkey). Notable studies of the benefits and costs of retrofitting schools in the USA, Italy, Mexico, and Peru have been published in the decade between 2004 and 2014.

A more ambitious and difficult case to make concerns the relative cost-effectiveness of investments in school seismic safety when compared to investing that money in other kinds of public health, safety, and welfare. In cases where child and infant mortality is high, longevity is shorter, basic vaccinations are not universal, or safe domestic water and sanitation facilities are inadequate, then the relative ranking of school safety as a cost-effective public health intervention may be low. Competition for public health funds could occur in trying to decide between clean water and vaccinations for everyone versus school seismic safety. In more affluent countries, the cost-effectiveness of saving lives in a future disaster usually has a high place among prioritized goals.

Of course, the physical safety of children, both in schools and out in the world at large, goes well beyond school seismic safety. HIV/AIDS, malnutrition, sexual violence, malaria, labor practices, and forced military service are day-to-day threats to the physical safety of many of the world’s children. The small potential for an earthquake over the next century might appear to pale beside other concerns which daily kill many more children.

However, in places where school seismic safety is a prominent issue – such as Tehran, Vancouver, Kathmandu, Bogotá, and Wellington – a significant earthquake has a high

probability of happening during the lifetime of schools currently standing and, therefore, for the gradually changing cohort of children during the life of the building. If earthquakes happen with equal probability around the clock, then approximately a 6–23 % chance exists of schoolchildren being in the school during a damaging earthquake. Cost-benefit studies of seismic construction estimate that it would add about 5 % to the cost of building a school in the USA, and in other countries the highest estimates are about 15 %, making the expectation that “every new school be a safe school” a realistic expectation.

When a population at risk is predominantly children, depending on the country, each death represents 40–70 years of lost life and productivity, and each injury represents 40–70 years of potentially expensive medical care, such as for brain or spinal injuries. Fix schools and several generations of children are protected. Health economics and medical ethics agree that the greatest social benefit comes from investment in the health and capacities of children. Aside from saving lives, the cost of education interrupted, and the serious potential for drop out adds another cost factor that seismic safety could help avoid.

Argument from the School’s Multiple Functions

The symbolic, cultural, economic, and political significance of schools as a community hub gives them an importance beyond merely being the site for educating children. Schools often play roles as central places for meetings and group activities, including literacy classes, religious services, political activities, and marriage ceremonies, particularly in rural areas where the school might be the only location big enough to hold such an event. Schools may also provide essential nutrition programs and serve as makeshift hospitals or vaccination centers even in normal times.

Where schools are the safest buildings in a community, they often serve as temporary shelter from storms and floods. They may be staging areas for first aid or rescue operation or other disaster response functions and even provide

temporary housing, while still fulfilling their role as an education facility.

Thus schools have a value in the social fabric of a community, providing adult education, promoting public health, building and maintaining sustainable livelihoods, and protecting people. The monetary value of those social gains defies estimation but clearly adds value and further justifies investment in safe school construction and maintenance.

We know from many disasters the important role that schools play in anchoring and speeding community recovery. Rapid school re-opening has tangible benefits in terms of children who are safe, supervised, and progressing towards their educational goals. Intangible benefits of schools functioning normally following a disaster include the psychosocial support in the face of loss and change. The importance of operational continuity of schools is linked to community recovery.

To take another example, retrofitting can spread a message far beyond the school. When children see their school being seismically retrofitted, they may have and may be designed to have ripple effects on safer residential construction. However, this is by no means automatic, and just how to maximize school construction or retrofit experience into a wider learning opportunity is a promising line of pursuit. Schools certainly serve as community hubs for propagating the seismic safety messages. School seismic safety can not only protect a community’s children but also educate communities to protect themselves.

Progress, Good Practices, and Lessons Learned

Assessing School Safety from Disasters, A Global Baseline Report (UNISDR 2012) found several consistent threats to safe school facilities:

- **Failure to assure every new school is a safe school:** Neither donors, governments, nor NGO associations have unequivocally committed to providing evidence or assurances or submitted to monitoring to assure that every new school is a safe school. Many small-scale

donors are particularly unaccountable and are not reached by the same accountability mechanisms and efforts of UN agencies and major international non-governmental agencies.

- **Multi-hazard awareness is often lacking:** In the construction of school facilities, there are many examples of fulfilling resilience to one hazard, while failing to mitigate against others – sometimes resulting in schools being dangerous in spite of good intentions or lying unused.
- **Impact of construction on education and family life not well understood:** School remodeling, retrofit, and replacement all have an impact on existing school programs and families. Planning these projects to minimize adverse impacts continues to be a concern.
- **Opportunity for construction and retrofit as an educational experience is untapped:** School construction and retrofit provide ideal opportunities for students and communities to learn the many principles of disaster-resilient construction to be applied throughout their communities. This opportunity is typically wasted as school sites are hidden from view and the experience is not used as a learning opportunity.
- **Lifeline infrastructure failures threaten school attendance:** Vulnerabilities in roads, bridges, and transportation systems must be prioritized when school attendance is threatened.
- **Failure to prioritize school re-opening jeopardizes community recovery:** Schools play a critical role in disaster recovery and community resilience where adults cannot return to work (UNISDR 2012).

The same study found consensus around the following core commitments required for safe school facilities: (1) Every new school must be a safe school. (2) Legacy schools should be prioritized for replacement and retrofit. (3) Lifeline infrastructure and nonstructural safety should be assessed locally and measures taken to mitigate [dangers]. (4) School furnishings and equipment should be designed and installed to minimize

potential harm they might cause to school occupants.

The expert review process that was part of the *Guidelines for Safer School Construction* (INEE 2010) yielded identification of a rich set of enabling factors associated with successful and sustained programs for school structural safety that all school safety advocates need to consider awareness, community ownership, partnership and dialogue, quality assurance, appropriate technology, integrated education, cultivating innovation, encouraging leadership, and continuous assessment and evaluation.

Overview of School Building Vulnerability

Rodgers (2012) reviewed earthquake damage assessment reports through 2009, for 32 earthquakes globally and aggregated findings from 31 school building vulnerability assessments. Table 3 shows the most commonly cited sources of vulnerability from both sources.

The general lack of agreement between vulnerability assessment and damage data likely reflects fragmentary and typically inadequate efforts to collect school damage data following past earthquakes, as well as a tendency for vulnerability assessments to identify common characteristics (such as plan irregularities) that rarely lead to the severe damage noted in post-earthquake damage surveys and reconnaissance reports. More complete earthquake damage data would provide the best indicator of the vulnerability-creating characteristics more likely to cause severe damage, because many vulnerability assessments do not differentiate the severity of damage expected from observed deficiencies.

The sources of and characteristics of structural vulnerability can be summarized in terms of: configuration (large windows with partial height walls below create captive columns or narrow piers, large windows on one side, weak or soft stories, large rooms, buildings one bay wide often with irregular plans), building type (vulnerable forms of vernacular and engineered construction,

School Seismic Safety and Risk Mitigation, Table 3 Characteristics found in damage and vulnerability assessments

Characteristics	Cited in 25 % or more		Cited in 15–24 %	
	Damage assessments	Vulnerability assessments	Damage assessments	Vulnerability assessments
Captive columns due to partial height masonry infill walls under windows	✓	✓		
Non-ductile reinforced concrete frame construction	✓	✓		
Generally poor construction quality	✓			✓
Poor-quality engineered materials	✓			
Soft or weak story		✓		
General plan irregularity		✓		
Exterior falling hazards		✓		
Maintenance deferred or lacking		✓		
Inadequate doors, windows, halls/corridors, or stairs		✓		
Vulnerable masonry construction		✓	✓	
Lack of seismic design understanding by engineers			✓	
Interior architectural and contents hazards			✓	✓
Windows reducing solid wall area in masonry construction				✓
Torsion				✓
General vertical irregularities				✓

Rodgers (2012), pp. 4–5

safer traditional construction forms and practices abandoned, standard building plans with seismic deficiencies, heavy roofs), location (sites susceptible to ground failure, sites that amplify ground motions), construction practices (poor quality, unskilled or low-skilled local labor, reducing quality to save money or time), materials (poor-quality engineered materials, weak local materials), lack of construction inspection, lack of maintenance, subsequent modifications, falling hazards, and inadequate exit pathways (Rodgers 2012).

Underlying drivers create an environment conducive to the vulnerability-creating characteristics cited above. Published literature identifies the following: unregulated community-based construction, scarcity of resources, inadequate building codes or zoning, lack of code enforcement, corruption of enforcement mechanisms, unskilled or unaware building professionals, lack of accountability, lack of awareness, failure to prioritize school safety, and urgent need for large numbers of new schools (Rodgers 2012).

Overview of Global Experiences in Seismic-Resistant School Construction

Some of the major policy and programmatic endeavors to assure seismic resilient construction of schools, worldwide, as of 2013, have involved important steps such as providing risk maps for safe school site selection, construction guidelines, standards, and oversight and commitments to safe school construction in the context of both post-disaster reconstruction and new school construction to meet Millennium Development Goals.

The **provision of risk maps** for safe school site selection requires both national and subnational coordination and often several different agencies reporting on the full spectrum of geophysical and hydrometeorological risks and taking into account nuclear, biological, and chemical hazards. In Peru, a pool of trained consultants based in universities around the country are now available to advise Regional Education Offices on safe school site selection. They draw

from existing risk maps for 115 towns (UNISDR 2008).

In the area of **construction guidance and standards**, California's Field Act in 1933 stands as the starting point of the movement. The Act required 15 % higher performance standards for new school construction and introduced stringent supervision. Legacy school construction was raised as a policy issue as early as 1938 (Garrison Act) but was not enforced until 1968. The oversight system involves structural plans prepared by engineers and approved by the Division of the State Architect. There is recurring on-site inspection and a final verification process.

The more common approach is the development of technical guidance for planning, design, construction, and local ongoing maintenance. There are numerous variations on this theme. For example, in the Philippines, in 2007, the Department of Education adopted the Principal-Led School Building Program approach where principals or school heads take charge of the implementation of management of the repair and/or construction. Assessment, design, and inspection functions are provided by Department of Education engineers who assist the principal during the procurement process. The Parent, Teacher, and Community Association and other community stakeholders are responsible for auditing procurements (INEE 2010). Interestingly, in Panama, it was the development and implementation of the maintenance guidance tool that paved the way for new school construction standards (UNISDR 2012).

There have been several examples of post-disaster commitment to "building back better," emerging from a general consensus following the 2005 Indian Ocean earthquake and tsunami, on the need use humanitarian assistance and reconstruction financing more responsibly. However, in the area of school seismic safety, these good intentions have only translated vaguely to measurable improvements in safe school construction. In Pakistan, 4 years after a devastating earthquake there, the National Education Policy 2009 section 5.5 addressed Education in Emergencies with several policy actions including requirements for school construction according

to international standards (UNISDR 2012). Following the devastating 2010 earthquake in Haiti, many donors stated that the schools that they are supporting seismic, hurricane, and flood-resilient school reconstruction, though there is no program that monitors progress in this regard.

In Indonesia in 2009, the Center for Disaster Mitigation, Institute of Technology Bandung (CDM – ITB), and Save the Children International published a handbook of typical school design and a manual on retrofitting of existing vulnerable school buildings for the Aceh and West Sumatra Earthquake Response programs. The guidelines take into account lessons learned in safe school construction, weaknesses in oversight of local government construction, and the need to incorporate design of dual-purpose multi-hazard shelters. In 2014 they were considered ready for an update.

In the Philippines, following devastating typhoons in 2006, 99 disaster-resilient schools and 26 day-care centers were constructed with the support of the Department of Education engineers, school principals, and community members. The new buildings, with water and sanitation facilities, can also serve as evacuation centers with flexibility to accommodate large numbers of people for emergency shelter (Global Education Cluster 2011).

Following the 1999 Kocaeli earthquake, 820 of 1,651 schools that were 60 km away in Istanbul, were found to have sustained some damage. Thirty-five schools were replaced, 59 schools were strengthened, and 59 were repaired (COGSS 2008).

Clear warrants and commitments from donors IGOs or INGOs when it comes to safe school construction are still clearly much needed.

There have also been too few and/or too quiet **commitments to safe school construction** in the context of the Millennium Development Goals, in spite of the fact that the Global Partnership for Education states as its first strategic goal the provision of a quality basic education in a *safe* environment. The most important and notable has been in Uttar Pradesh, India, where 23.5 million children attend school in this moderate to severe seismic risk zone; 21,000 new school buildings

(30/day) were to be built in a 2-year period. In 2006–2007 the Elementary Education Department proposed to integrate earthquake-resilient design into all new school buildings. One primary school, two upper primary, and three additional classroom designs were prepared with detailed construction manuals. Disaster-resilient measures added 8 % to the construction costs. To cope with massive scale of the project, a cascading approach prepared 4 master trainers for each of 70 districts. These individuals trained 1,100 Junior Engineers and Education Officers. Ten thousand masons were also trained. In Uttar Pradesh every new school is now a safe school (UNISDR 2008).

Overview of Global Experiences in School Seismic Retrofit

In Sichuan, China, Prior to the 2008 Sichuan earthquake, school principal Ye Zhiping pestered local authorities until they consented to retrofit the buildings of Sangzao Middle School to improve their safety. He also initiated regular evacuation drills. The result of his efforts was that during the devastating earthquake, this school provided life safety for all of its students and staff.

The United Nations Center for Regional Development in Kobe began promoting school earthquake safety initiatives in 1999, in the process of resilience-building following the Hanshin-Awaji earthquake. A multi-country school seismic retrofit initiative (2005–2008) sought to make schools safer through self-help, cooperation, and education. The project engaged local communities, governments, and resource institutions in demonstration vulnerability assessments and school retrofit projects in four to six schools each, in Fiji, India, Indonesia, and Uzbekistan. In 2006, the state of Uttar Pradesh, in India, undertook large-scale disaster-resilient construction of new schools (Bhatia 2008). GeoHazards International also conducted small-scale screening and retrofit demonstration projects in vulnerable schools in Delhi, India (Rodgers 2012), and helped Bhutan's Ministry

of Education develop the process and tools for a nationwide school vulnerability assessment program, which is currently underway.

More than a dozen countries have developed approaches, conducted significant vulnerability assessments, and/or made commitments made to school retrofit since 2000. Several of these were inspired by unacceptable levels of damage experienced in recent large earthquakes. Many are instructive or inspiring, in terms of their scope, methods, and limitations. Looking at these regionally allows an overview of both limited scope and adequacy.

Middle East and North Africa: In Algeria, vulnerability assessment was done on 526 buildings in 190 schools across 9 municipalities in Algiers, using simple survey forms (Rodgers 2012). In Syria, UNDP is supporting earthquake school safety program incorporated into 5-year plan and institutions for disaster risk reduction are being consolidated (UNISDR 2012). The Arab League is currently considering a regional approach to disaster risk reduction, which will hopefully include a comprehensive approach to school safety.

North America: In British Columbia, Canada, Vancouver school buildings were surveyed in 1990 (Rodgers 2012). Responding to advocacy efforts of the local "Families for School Seismic Safety," in 2004, the provincial government committed \$1.5 billion Canadian to ensure that BC Schools meet acceptable seismic life safety standards by 2019.

In the USA, there has been detailed assessment of 26 school buildings in Kodiak, Alaska, with recommendation for retrofit of four. In California, a desk assessment of 9,659 pre-1978 school buildings found 7,537 potentially vulnerable buildings. Twenty thousand uncertified projects have been mapped (California Watch 2011). The state of Oregon conducted collapse risk assessment of 2,185 K-12 school buildings using FEMA 154 rapid visual screening (RVS) and produced structural engineering reports for more than 300 buildings. South Carolina has completed a prioritization exercise on all public schools; six have been retrofitted. In Tennessee 49 buildings in 202 schools have been screened using ATC-21

plus local methods, and in Utah, RVS was used on a sample of 128 of 1,085 schools in the state (Rodgers 2012).

Latin America and the Caribbean: The Organization of American States began a commitment to school safety in 1992. A coordinated regional action plan was developed to benefit Costa Rica, El Salvador, Guatemala, Honduras, Nicaragua, and Panama. Development assistance donors and local organizations contributed to strategies and capacity to carry out retrofitting of educational facilities. School infrastructure experts from each country received training.

In Bogotá, Colombia, in 1997, seismic microzonation studies paved the way for seismic-resistant building codes in 1998. In 2000 the Directorate of Prevention and Attention of Emergencies in Bogotá found 434 of 710 schools vulnerable to earthquake damage, 3 in flood areas, and 20 in landslide-prone areas. Two hundred and one were prioritized for retrofit or replacement. Between 2004 and 2008, an investment of \$460 m USD in school replacement, retrofit, and risk management promotion provided structural reinforcement of 172 schools, “nonstructural” risk reduction in 326 schools, and the construction of 50 new mega-schools, compliant with earthquake-resistance requirements. Three hundred thousand children are safer as a result (see entry “► [School Seismic Safety: Case Studies](#)” for case study of Colombia). In Ecuador, initial screening of 340 high-occupancy school buildings, and modified RVS of 60 most vulnerable, detailed analysis for 20, and retrofit designs for 15 have taken place (Rodgers 2012). In Lima, Peru, 28 schools in Barranco and 80 schools in Chorrillos were evaluated using ATC 21 RVS and EMS_98 estimation of damage potential (Rodgers 2012). A retrofit solution was developed to mitigate potentially devastating structural defect of “short columns.” And in Venezuela, 50-year-old schools were identified as needing retrofitting in moderate and above seismic zones, whereas 20–30 year-old “box” schools only require retrofit in higher-risk zones. Practical retrofitting techniques were developed. As of 2007, 28,000 schools were being surveyed in a national program. Twelve

schools were selected for pilot retrofits (Rodgers 2012).

Europe and Central Asia: In Europe, discussion has been robust in Italy and Portugal, innovations have been led by UNICEF and partners in Central Asia, and World Bank financing has supported Turkey to make significant progress in seismic safety (see entry “► [School Seismic Safety: Case Studies](#)” for case study on Turkey).

In Yerevan, Armenia, full assessments have been conducted by teams of dozens of people, mobilized from as many as seven different government agencies, over several days. Every year 40 of Yerevan’s 200 schools are slated for special maintenance, upgrading, and retrofitting. It has been noted that a 2-person expert team spending 2 h per conducting a rapid assessment would require 6 FTE years to assess Armenia’s 1,500 schools. In Kyrgyzstan, a national school safety assessment of over 3,000 learning facilities with support from USAID found that more than 80 % were vulnerable to earthquake damage. Public access to this information is made possible through an online portal (UNICEF 2011).

In Uzbekistan 1,000 school buildings were assessed, revealing that 51 % require demolition and replacement, 26 % require capital repair and reinforcement, and 27 % are life safe and require no intervention (Khakimov et al. 2007). Eleven design institutes participated in building code revision for school building construction. Typical designs were created for new schools of different sizes. A database of typical construction and technical decisions seismic reinforcement were developed. UNCRD provided financial and technical support for demonstration projects on reinforced concrete frame, masonry, and frame panel buildings. The incremental cost of seismic reinforcement was shown to be between 3 % and 14 % depending on intensity zone, type of construction, number of floors, capacity, and ground conditions (Khakimov et al. 2007).

In Italy, a substantial contribution comes in the form of an overall risk management framework developed for retrofit prioritization (Grant et al. 2007). Some schools have now been assessed in Emilia-Romagna (Rodgers 2012). Portugal has demonstrated an important

innovation by incorporating school vulnerability assessment and retrofit into its ongoing modernization program. At least 330 public school buildings have been assessed and retrofits designed (Rodgers 2012; UNISDR 2012).

The Istanbul Seismic Risk Mitigation and Emergency Preparedness Project (ISMEP) (with loans from World Bank and EIB) allowed for retrofitting of 250 schools and reconstruction of 36 schools in 2007–2008 with 600 more undergoing assessment and feasibility studies. In 2009 the remaining 450 schools were slated for retrofitting.

South Asia: Bhutan has begun a nationwide vulnerability assessment of school buildings. The first phase, covering 5 of Bhutan's 20 districts, began in 2013, with funding from UNICEF.

In India there are several examples of large-scale seismic vulnerability assessments: In Gujarat a modified RVS was conducted for 153 schools following the 2001 earthquake (Rodgers 2012). In Shimla, SEEDS India took a stepwise approach: Step one was low-cost mass scale RVS of school buildings. From these, a smaller number were selected for simplified vulnerability assessment using limited engineering analysis. The highest-risk buildings were identified for detailed vulnerability analysis (SEEDS 2006). Retrofitting designs were drawn up for 20 schools and implementation carried out in ten schools. Guidelines were developed for retrofit and training of local masons and engineers and delivery of skill training. "Nonstructural mitigation plans" were carried out in 20 schools. An awareness campaign reached out to all 750 schools, including nearly 100,000 students, 7,500 teachers and local builders, engineers, and officials (SEEDS 2006). The Government of India's National School Safety Program plans to seismically retrofit more than 40 schools throughout the country as demonstration projects. The National Center for Peoples' Action in Disaster Preparedness (NCPDP), GeoHazards International, and others also carried out school assessment and retrofit programs.

Nepal has also made some strides in both vulnerability assessment and retrofit planning. There are an estimated six million children and 140,000 teachers at risk of death and injury in schools. In the Kathmandu Valley, 643 schools (1,100 buildings) have been inventoried and 378 (695 buildings) surveyed for vulnerability. Seventy-five percent are expected to be damaged beyond repair, in a scenario earthquake. A school day earthquake would kill 29,000 children and teachers and injure 43,000 (Dixit et al. 2013). The MoE has planned to retrofit 900 schools in the Kathmandu Valley over 5 years (Dixit et al. 2012). In Lamjung and Nawalparasi, vulnerability screening has covered 745 and 636 buildings, respectively, some with detailed assessments (Rodgers 2012).

In Pakistan, in 2008, the Aga Khan Planning and Building Services, Habitat Risk Management Program in Northern Pakistan, used retrofitting of four schools to demonstrate structural and nonstructural seismic retrofitting, to train builders and to train female village youth in mapping, land-use planning, and disaster management (INEE 2010).

Southeast Asia: There has been relatively sparse activity when it comes to seismic safety of schools in Southeast Asia. It may be that the frequency of cyclones and flooding and even the threat of tsunami precede thoughts of earthquake risks. It may also be that the rapid pace of development and the increasing numbers of new children being brought into school have led to natural prioritization of safe new construction rather than retrofit. In the Philippines, local authorities are responsible for school construction. However, assessment, design, and inspection functions are provided by Department of Education engineers who assist the principal during the procurement process. The Parent, Teacher, and Community Association and other community stakeholders are responsible for auditing procurements. Earthquake, typhoon, flood, and even volcanic ashfall resilience must often be factored in (INEE 2010).

East Asia: School seismic safety has been on the agenda in Japan for many years, but it is only since 2005 that 125,000 public school buildings nationwide have been assessed by the Ministry of Education (MEXT) (Rodgers 2012). Sixty-two percent of these were constructed before 1981, when the current anti-seismic code enforcement began. About 25 % of schools are considered safe, but 48,000 older school buildings were found needing assessment or retrofitting. 10,000 of these were found to be at high risk of collapse in expected earthquakes. The Ministry of Education raised subsidies for vulnerable school buildings from 50 % to 67 % in 2008 when 229 billion JPY was allocated to meet the new goal of retrofitting of all highest-risk school buildings within 4 years.

Oceania: School seismic safety is also on the agenda in New Zealand, where a walk through survey of 21,000 buildings at 2,361 public schools in 1998 triggered a follow-up investigation in 2000 (Rodgers 2012). A World Bank GFDRR project demonstrated school retrofit in six schools in two districts (2008–2009).

Summary

In the course of the past decade, an approach to all hazards and all aspects of school safety has emerged in both the literature and practice of global advocacy. The Global Alliance for Disaster Risk Reduction and Resilience in the Education Sector (led by UNESCO, UNICEF, UNISDR, IFRC, INEE, Save the Children, Plan International, World Vision) use the shared Comprehensive School Safety framework. The framework takes a multi-hazard approach and addresses the many different factors needed to address safe school facilities, school disaster management, and disaster reduction education. While seismic vulnerability (and related secondary hazards) to school buildings are naturally of concern to earthquake engineers and many others, it is important to fit this into an

all-hazard and comprehensive approach so that the solutions to seismic safety do not ignore coexisting vulnerabilities to cyclones, floods, and volcanic eruption nor conflated with the broader approach that also addresses disaster management and education (Global Alliance for DRRR in the Education Sector 2014).

Overall, the threat of earthquake damage to school buildings has not been sufficiently well appreciated. School safety issues have not featured in the major global campaign for increased school attendance (“Education for All” and the Millennium Development Goals). The full extent of the risk to school buildings and to students remains to be fully defined.

A global effort at mapping schools (by density of occupancy and quality of construction) in relation to seismic and other hazards has been proposed by the World Bank Global Facility for Disaster Risk Reduction and Recovery, to begin in 2014. The full impacts of earthquakes on the education sector cannot end with calculating the value of structural and nonstructural damage. The impacts on *children’s education* are almost entirely unmeasured. Research is needed to understand how educational outcomes such as enrollment, attendance, and achievement are impacted by earthquakes.

There are strong arguments that support giving school seismic safety increased priority and a higher profile. An initial step in raising the visibility of this issue was the adoption of school safety as one of the focal points for advocacy in preparation for the Hyogo Framework for Action 2005–2015 adopted at the World Conference on Disaster Reduction held in Kobe, Japan, in January 2005. The development of the Comprehensive School Safety framework in 2013 has begun to articulate how school facilities safety can be understood within the wider context that includes school disaster management as well as risk reduction and resilience education. As a post-2015 agenda for both development and disaster risk reduction are currently under consideration, it

continues to be extremely important to raise the profile of school safety. In preparation for this, child-centered organizations have formed a Global Alliance for Disaster Risk Reduction and Resilience in the Education Sector.

Based in part on the case studies (see entry “► [School Seismic Safety: Case Studies](#)”), it seems evident that low-cost, accessible technology and design exists with which to build new schools and to retrofit existing ones. A community-based approach holds great promise involving many stakeholders, including local buildings, masons, contractors, etc. Promising demonstration and large-scale projects in Nepal, India, Turkey (see entry “► [School Seismic Safety: Case Studies](#)”), Central Asia, and the Caribbean islands all provide strong experience to build upon for case studies.

Case studies also make clear that child rights advocates, parents, and seismic safety experts together, lobbying for school seismic safety, can be extremely effective in achieving policy change, as case studies of British Columbia and Bogotá (see entry “► [School Seismic Safety: Case Studies](#)”) show.

School seismic safety has been the subject of both research and policy since the 1933 Long Beach earthquake that spurred California’s landmark *Field Act*, requiring that school construction meet seismic safety standards. As both seismic risk assessment and building codes have progressed, so too have expectations for selection of performance standards. However, globally, the application of these standards and codes falls short in several major respects: community-built schools are frequently constructed using high-tech materials intended for engineered construction, without the corresponding understanding, training, or supervision; where building codes exist they are not known, understood, or consistently applied; and safe site selection is frequently skipped, and site-specific hazards are not factored in. Privately built schools are often exempt from the same standards of construction as public schools. The need for programs and people that bridge the available engineering knowledge with scalable on-the-ground national programs is significant.

In 2009, *Guidance notes on safer school construction* (INEE 2010) was published to synthesize and kick-start systematic guidance. An important global resource for documents to guide safe school construction was initiated by UNESCO IPRED, immediately after the Haiti earthquake (UNESCO IPRED 2010). This resource database endeavors to compile both building codes and the now numerous documents produced by NGOs or at the national level with standard designs for safe school construction, and in some cases with construction guidance.

The past decade has seen several relevant scientific papers suggesting methods for vulnerability screening (e.g., in Italy, Grant et al. 2007), and detailing approaches to seismic retrofit. The challenge is whether or not the guidance and the science are put into practice. The written record does not suggest that these approaches are yet systematic, are supported with training, are monitored, or are applied to both public and private schools. Community-built and un-engineered construction has been addressed in far fewer publications and has not specifically addressed school construction.

There have been a small number of significant programmatic efforts to support seismic safety. UNICEF’s regional office for Central and Eastern Europe and the Commonwealth of Independent States, with support from the World Bank and DIPECHO, has partnered with national governments in Central Asia and the South Caucasus to address school safety. Part of that work has included developing a broad regional framework for assessing and ranking school facilities based upon exposure and vulnerability to earthquakes and other natural hazards. Drawing upon INEE’s *Guidance notes on safer school construction*, UNICEF elaborated a list of 17 simple indicators that local experts could use as part of a rapid visual assessment of school facilities in order to identify schools at risk of heavy damage in seismic events.

In 2012, engineers in Kyrgyzstan localized this framework and carried out a national school safety assessment of over 3,000 learning facilities with USAID funding. They reported to the national government that over 80 % of learning

facilities were vulnerable to damage in seismic events and provided public access to the assessment through an online portal. Similar national assessment strategies are being piloted in Kazakhstan, Tajikistan, Armenia, and Azerbaijan.

Similarly, UNCRD (UN Centre for Regional Development) showcased community-based comprehensive school earthquake safety in selected countries of Asia Pacific. Under the program “Reducing Vulnerability of School Children to Earthquakes,” school communities carried out seismic retrofitting of their school buildings with expert guidance from Bandung Institute of Technology (ITB) in Indonesia. The retrofitting works in public schools were used for community awareness on earthquake safety through community visits in the school premises during construction time. Pilot school assessment and retrofitting in Fiji led to the National Disaster Management Office (NDMO) adopting school safety program under regular government that also developed seismic retrofit guidelines and mason’s training manual. Tashkent city government (Hokimiyat) in Uzbekistan appraised neighborhood associations on schools retrofitting programs and used school constructions for training of engineers on seismic safety.

The United Nations International Strategy for Disaster Reduction (UNISDR) launched the 2006–2007 biennial awareness campaign “Disaster Reduction Begins in Schools.” This was followed up in 2010 with the *Resilient Cities Global Campaign for One Million Safe Schools and Hospitals Campaign*. The 10-point checklist that 1,643 Mayors have signed on for, includes assessing and upgrading the safety of schools. These successes deserve praise but should not induce complacency. There is a long way to go with respect to school seismic safety.

Initial programs and guidance for safe school facilities have been provided by OECD (2004), UNCRD (2008), INEE/World Bank GFDRR/UNISDR (2010), and several other programs, with modest support of donors and lenders. These approaches experiences are now ripe for implementation at scale. These include regional hazard mapping and revision (where necessary), the potential for crowd-sourced

mapping of local hazards; enforcement of seismic building codes by national, provincial, and local governments; training of engineers and significant capacity-building efforts to train local masons and other builders; and invention of more innovative models for funding reinforcement of schools.

It is important, however, not to fetishize the safety of school *buildings* and to take care not to separate the safety of the community of users and educational continuity planning, which is not limited to the buildings themselves. Neither should the focus be solely on fatality prevention. There is much similar work to be done to prevent disability and injury especially by securing the contents of the buildings and to assure educational continuity. All-school, participatory school disaster management planning, local risk assessment and risk reduction, mastery of emergency response skills, and regular drills to practice and improve readiness are important. A culture of safety is necessarily multifaceted, and activism in one area encourages changes in consciousness, expectations, and demands.

The enthusiasm for making education accessible to all does not absolve duty-bearers from assuring that school is safe from infrequent but high-impact hazards such as earthquake and various secondary hazards. It would be an ironic and tragic result if the achievement of one Millennium Goal (increased school attendance) is marred by increased death and injury of young people, thus setting back the achievement of another Millennium Development Goal (reduction of child mortality).

Cross-References

- ▶ [“Build Back Better” Principles for Reconstruction](#)
- ▶ [Building Codes and Standards](#)
- ▶ [Earthquake Protection of Essential Facilities](#)
- ▶ [Earthquake Risk Mitigation of Lifelines and Critical Facilities](#)
- ▶ [School Seismic Safety: Case Studies](#)

References

- Bhatia S (2008) Earthquake-resistant school buildings in India. *Urb Plan Des* 161(4):147–149
- California Watch (2011) K-12/public safety/spotlight/on shaky ground. 22 Dec 2011. Accessed 3 Feb 2012
- CNN (2014) Haiti earthquake fast facts. <http://edition.cnn.com/2013/12/12/world/haiti-earthquake-fast-facts/index.html>. Accessed 6 Jun 2014
- Coalition for Global School Safety (2008) School disaster reduction: think globally, act locally. 'Mother' slide presentation
- Dixit A et al (2012) Institutionalization of school earthquake safety program in Nepal. In: Proceedings of the 15th WCEE. http://www.iitk.ac.in/nicee/wcee/article/WCEE2012_5260.pdf. Accessed 16 Feb 2014
- Dixit et. al. (2013) Public School Earthquake Safety Program in Nepal, Geomatics, Natural Hazards and Risk, DOI: 10.1080/19475705.2013.806363
- Dowrick D, Rhoades DA (2004) Risk of casualties in New Zealand earthquakes. *Bull N Z Soc Earthq Eng* [http://www.nzsee.org.nz/db/Bulletin/Archive/38\(2\)0053.pdf](http://www.nzsee.org.nz/db/Bulletin/Archive/38(2)0053.pdf). Accessed 5 Mar 2015
- EERI (1990) Earthquake Engineering Research Institute, Oakland. Loma Prieta Earthquake 17 Oct 1989. Preliminary reconnaissance report. https://www.eeri.org/wp-content/uploads/USA_CA_LomaPrieta_PrelimReport.pdf. Accessed 16 Feb 2014
- FEMA (2011) Reducing the risks of non-structural safety – a practical guide, 4th edn. Washington, DC. <http://www.fema.gov/media-library/assets/documents/21405?id=4626>. Accessed 19 Feb 2014
- Global Alliance for Disaster Risk Reduction and Resilience in the Education Sector (2014) Comprehensive school safety. <http://preventionweb.net/go/31059>. Accessed 15 Jan 2014
- Global Education Cluster (2011) Disaster risk reduction in education in emergencies: a guidance note for education clusters and sector coordination groups. <http://preventionweb.net/go/20366>. Accessed 16 Feb 2014
- Global Partnership for Education (2014) Results and data. <http://www.globalpartnership.org/results/>. Accessed 16 Feb 2014
- Grant DN, Bommer JJ, Pinho R, Calvi GM, Goretti A, Meroni F (2007) A prioritization scheme for seismic intervention in school buildings in Italy. *Earthq Spectra* 23(2):291–314
- INEE/World Bank GFDRR/UNISDR (2010) Guidance notes on safer school construction. INEE (Inter-agency network for education in emergencies), New York. <http://www.preventionweb.net/go/10478>
- Khakimov S, Nurtaev B, Tursunov K (2007) School earthquake safety program in Uzbekistan, slide presentation. Tashkent
- New York Times (2008) School-age children in earthquake zones, 27 May 2008. <http://www.nytimes.com/imagepages/2008/05/27/science/27school.graphic1.ready.html>. Accessed 16 Feb 2014
- OECD (2004) Keeping schools safe in earthquakes. Organisation for Economic Co-operation and Development, Paris. http://www.oecd.org/document/36/0,2340,en_2649_201185_33630308_1_1_1_1,00.html. Accessed 19 Feb 2014
- Pandey B (2013) Political economy study on safe schools, World Bank, Global Facility for Disaster Reduction and Recovery
- Rodgers JE (2012) Why schools are vulnerable to earthquakes. In: Proceedings, 15th world conference on earthquake engineering, Lisbon, 24–28 Sept 2012
- SEEDS (2006) Earthquake Safety in Mountain Regions. School Earthquake Safety Initiative. Shimla, Delhi
- State of California, Seismic Safety Commission (2009) The field act and its relative effectiveness in reducing earthquake damage in California's public schools. <http://www.seismic.ca.gov/pub/CSSC%2009-02%20Field%20Act%20Report.pdf>. Accessed 16 Feb 2014
- UNCRD (2008) Reducing school vulnerability to earthquakes. United Nations Center for Regional Development, Hyogo, Japan. <http://www.hyogo.uncrd.or.jp/publication/pdf/Report/2008SESI/UNCRD%20SESI%20publication%202008.pdf>. Accessed 16 Feb 2014
- UNESCO (2004) 'Education for all week 19–25 April 2004: the extent of the problem', information from UNESCO's (United Nations Educational, Scientific and Cultural Organization) http://portal.unesco.org/education/en/ev.php-URL_ID=28702&URL_DO=DO_TOPIC&URL_SECTION=201.html. Accessed 25 Jan 2014
- UNESCO IPRED (2010) Resource/database, disaster preparedness and mitigation. <http://www.unesco.org/new/en/natural-sciences/special-themes/disaster-preparedness-and-mitigation/natural-hazards/earthquakes/ipred/unesco-ipred-database/>. Accessed 25 Jan 2014
- UNICEF (2011) Toward safer schools – methodology for nationwide benchmarking of school safety http://www.unicef.org/ceecis/towards_safer_school_FINAL.pdf. Accessed 5 Mar 2015
- UNISDR (2008) Disaster prevention for schools – guidance for education sector decision-makers. Geneva. <http://www.preventionweb.net/go/7344>. Accessed 25 Jan 2014
- UNISDR (2012) Assessing school safety from disasters, a global baseline report. <http://preventionweb.net/go/35274>. Accessed 15 Jan 2014
- USGS (2003) 'Magnitude 8.3 – Hokkaido, Japan region 25 Sep 2003 19:50:06 UTC'. <http://earthquake.usgs.gov/recenteqsww/Quakes/uszdap.htm>. Accessed 25 Jan 2014
- USGS (2015) Historic World Earthquakes. <http://earthquake.usgs.gov/earthquakes/world/historical.php>. Accessed 5 Mar 2015
- WEF (2000) UNESCO World Education Forum. http://www.unesco.org/education/efa/wef_2000/. Accessed 5 Mar 2015

School Seismic Safety: Case Studies

David Alexander¹, Sanjaya Bhatia², Djillali Benouar³, Jitendra Kumar Bothara^{4,5}, Omar-Dario Cardona⁶, Amod Mani Dixit⁴, Rebekah Green⁷, Manu Gupta⁸, Ram Chandra Kandel⁹, Ilan Kelman^{10,17}, Tracy Monk¹¹, Bishnu Pandey¹², Marla Petal¹³, Zeynep Türkmen Sanduvaç¹⁴, Rajib Shaw¹⁵ and Ben Wisner^{1,16}

¹University College London, London, England

²UNISDR Recovery Platform, Kobe, Japan

³University of Bab Ezzouar, Algiers, Algeria

⁴NSET, Kathmandu, Nepal

⁵Miyamoto Impact, Christchurch, New Zealand

⁶Universidad Nacional de Colombia, Manizales, Colombia

⁷University of Western Washington, Bellingham, WA, USA

⁸SEEDS of India, Delhi, India

⁹Toronto, ON, Canada

¹⁰Institute for Risk and Disaster Reduction and Institute for Global Health, University College London, London, England

¹¹Families for School Seismic Safety, Vancouver, Canada

¹²University of British Columbia, Vancouver, Canada

¹³Risk RED (Risk Reduction Education for Disasters), Los Angeles, CA, USA

¹⁴Risk RED, Istanbul, Turkey

¹⁵University of Kyoto, Kyoto, Japan

¹⁶Oberlin College, Oberlin, OH, USA

¹⁷Norwegian Institute of International Affairs, Oslo, Norway

Synonyms

Education sector; Examples; Policies; Policy; Progress; Schools; Seismic safety

Introduction

The case studies in this entry accompany the previous entry on School Seismic Safety and

Mitigation. They tell the stories of a variety of national efforts to improve school seismic safety. The contributors to the case studies are engineers and parents, social workers, and international development specialists. They examine policy, advocacy, vulnerability, and solutions. They contain observations about stepwise progress, motivation, political will, technical approaches, innovations, moderate successes, and long roads ahead. There are more stories to be added. The intention is to provide school seismic safety advocates with both elements of inspiration and way points on a road map with many options to consider.

The case studies and their contributors are:

Algeria – Djilali Benouar

Canada – Tracy Monk

China – Sanjaya Bhatia

Colombia – Omar Dario Cardona

India (Delhi, Shimla) – Manu Gupta

India (Uttar Pradesh) – Sanjaya Bhatia

Italy – David Alexander

Japan – Rajib Shaw

Nepal – Amod Mani Dixit, Jitendra Kumar Botendra, Ram Chandra Kandel, and Bishnu Pandey

Turkey – Marla Petal and Zeynep Türkmen Sanduvaç

USA (California) – Marla Petal and Rebekah Green

It is important to note that school *seismic* safety should not be isolated from school safety from disasters and climate change impacts in general. It cannot be considered responsible to approach the rights of children from our narrow professional silos. It is incumbent upon all advocates for children to see the matter from their point of view, that is, from the perspective of all of the physical threats to their right to survival and safety and from all of the threats to their right to education and development.

Similarly, key stakeholders in the seismic safety of schools (engineers, architects, builders) must make the effort to think not primarily about the *structures*, but rather about the people who

use the structures. The users must also have safe access and egress. For the users, infill walls that fall out of plane and unsecured nonstructural building elements or building contents are a far greater threat than their “minor damage” designation suggests. If the building is going to be one of the strongest structures in the community, then it may also need to be planned to be as a cyclone shelter or to double as a shelter for people displaced after an earthquake. In that case, water and supply storage and extra sanitation facilities have to be considered. And the job is not complete when the key is handed over. Buildings that do not come with a user’s manual and a school maintenance calendar stand little chance of being safely maintained. Thus, the concern for school seismic safety does not begin and end with the structures themselves, but must take a holistic approach and, of course, include the user community.

Algeria

Ninety percent of Algeria’s population of 30 million is concentrated in a band about 60 km wide and 1,200 km long along the Mediterranean coast on the African and Eurasian tectonic plate boundary. This region has repeatedly experienced moderate-to-strong earthquakes. During the twentieth century, earthquakes claimed at least 10,000 lives, injured about 27,000, and made about half a million homeless. In addition to building collapse due to earthquakes, considerable damage from liquefaction and landslides was observed.

School buildings have also suffered considerable damage in earthquakes, varying according to the period during which they were built: (1) those degraded through aging and lack of maintenance, built during the colonization era (1830–1962), account for about 30 % of the school building stock; (2) those built after independence, during the 1970s, with rapidly growing population and democratization of educational opportunity (when primary school became free and compulsory), when school construction accounted for the largest single item in Algeria’s budget, were designed and built without taking into account

seismic risk; and (3) those built with technical supervision after 1983 and the introduction of Algeria’s seismic building code in 1981.

Schools in Algeria are all state owned and were built by the government. The government adopted one typical structure for all schools that could be duplicated easily across the country. The standard architectural design of schools involves two fundamental elements: the classroom and the circulatory corridors. Classrooms are 7×4 m and circulatory corridors are 2.5 m wide. These are far from those of an ideal seismically resistant structure as recommended by Algeria’s own seismic codes. Other standard design elements also unfortunately reduce the resilience of these school structures.

Numerous reports show the deficiencies in design, construction techniques, and materials (poor quality of concrete) with respect to particular earthquakes at El Asnam (1980), Chenoua-Tipaza (1989), Beni Chougrane-Mascara (1994), Ain Temouchent (1999), and Boumerdes-Algiers (2003) (Bendimerad 2004). The following typical damage to school buildings was recorded in recent earthquakes:

- Rupture of staircases
- Destruction of joints
- Destruction of short columns
- Damage in Masonry
- “Pancake” collapse due to weak columns, overly strong beams, and heavy roofs composed of reinforced concrete slabs

Such damage causes enormous financial loss to the government. For instance, after the Boumerdes-Algiers earthquake disaster of 2003, 100 primary schools had to be rebuilt completely for the sum of US\$4.28 million and 253 rehabilitated for \$10.65 million. In addition, 12 junior high schools were completely rebuilt for the sum of \$10.28 million and 111 rehabilitated for \$20.85. Also, 10 high schools were rebuilt for the sum of \$21.42 million, and 58 were rehabilitated for \$12 million.

So far these destructive earthquakes have occurred after school hours or on weekends, and thus, no loss of life or injuries have been recorded

at schools in Algeria. This good luck may have made the government and the civil society alike less aware of the high vulnerability of the schools and partially explains why there has so far been no implementation of a Ministerial instruction dating from 1989 that required application of “technical expertise and the eventual reinforcement of all public buildings and in particular schools and universities.” Instead, the introduction of new materials such as reinforced concrete in the absence of proper seismic-resistant design, building codes, and enforceable regulations has increased the risk to structures and their occupants. Relatively minor reinforcements could reduce the potential for damage to these structures.

Canada

British Columbia’s (BC) west coast is Canada’s region of highest seismic hazard. Two-thirds of the province’s 3.9 million people live within the zone of highest risk. The region has experienced ten moderate-to-large earthquakes since 1870. In recent millennia, an earthquake on the order of the largest magnitude experienced globally in the last 100 years has struck approximately every 500 years. In addition to potential building collapse induced by ground shaking, significant damage from liquefaction, tsunamis, and landslides are expected.

Older BC schools were built from some of the most seismically vulnerable materials – in the early 1900s, unreinforced masonry and then, in the mid-1900s, non-ductile concrete frame. Thus, in many communities, the school might be one of the buildings at highest risk for earthquake damage. A formal risk assessment of all BC school buildings was initiated in June 2004 with the full report due in October 2004. Initial estimates from the provincial government suggest that 800 of BC’s schools might need some form of seismic upgrading.

In Vancouver, BC’s largest city with a core population of about 560,000 and a metropolitan population of about two million, a 1989 rapid seismic risk assessment found that 30 % of the city’s school buildings were at high risk of

experiencing structural damage in an earthquake, and 15 % were at moderate risk (Taylor 1989). Between 1990 and 2004, 11 schools have been upgraded, so if the current pace continues, this work would be completed in 2064.

While the seismic hazard facing schools in greater Vancouver is similar to that in Seattle, Washington, school safety has not been a priority on the Canadian side of the border. Indeed, differences in seismic hazard mapping techniques used by Canadian and US geoscientists may actually underestimate the risk to Vancouver schools.

The current National Building Code of Canada ranks buildings according to their priority as critical infrastructure. The higher the number assigned, the higher the priority. Average houses are assigned an Importance Factor of 1.0, schools are designated 1.3, and hospitals, police stations, and prisons are assigned 1.5. Schools – unless they are designated as post-disaster shelters – get a lower priority than hospitals, police stations, and prisons. Vancouver City Council is funding the seismic upgrading of community centers so that they could be used as post-disaster receiving centers. Due to the differences in funding sources, some community centers are being upgraded, while nearby schools, which might be less seismically resistant, are sometimes not.

In general in BC there is high awareness of earthquake risk. For example, the City of Vancouver has seismically upgraded its water supply system and bridges, and the electric utility, BC Hydro, is systematically upgrading its buildings and infrastructure, including dams. Even some provincially run liquor outlets have been seismically upgraded. With seismic safety so clearly on the policy agenda in many sectors in British Columbia, why have public schools received so little attention?

The answer is that in BC funding for public school seismic upgrades has been part of the capital budget of the provincial Ministry of Education. Local school boards oversee this work and must proactively request provincial funding for projects that they deem to be high priority. Advocacy work by one of the authors on behalf of Families for School Seismic Safety British

Columbia (FSSS) identified and tackled concerns with this funding system.

First, there was no district-to-district standardization of approach. Each local school district was individually approaching the issue of seismic safety for only their schools. FSSS pressed the government to unify the approach taken by centralizing expertise. The earthquake engineering community, through its professional association, the Association of Professional Engineers and Geoscientists of BC, is now advising the government on standardized, peer-reviewed methods for assessing and addressing seismic risk to schools.

Second, local authorities were concerned that informing parents of the seismic risks to local schools could cause panic and could be politically damaging. FSSS's and others' work has ensured that Vancouver parents are now well informed about the issue and are actively involved in trying to solve it. This process did not cause panic. Instead, bringing parents into the consortium has yielded an active and effective lobbying group.

Finally, there is the problem of setting priorities. The primary concern of school boards – quite rightly – is the day-to-day education of children. Seismic safety of a school building does not lead to improved education, so school boards have sometimes had difficulty making the issue a high priority. FSSS is trying to help public officials see school seismic safety as an infrastructure, public health, and human rights issue and to obtain new funding from outside of the Ministry of Education, that is, from provincial and national, authorize with mandates in those areas. The aim is that this work be seen as an infrastructure project for children and not perceived as competing for funding with their day-to-day educational needs.

Ultimately, the two basic human rights of children, to an education and to physical safety, should not be competing for the same funds. The expert community is now driving the initiative and the government appears to be listening. Many positive steps have been taken in BC, but there is much work left to be done.

China

Following the 2008 Sichuan earthquake in which at least 15,000 children lost their lives in schools, the Ministry of Education and the Ministry of Construction and National Development and Reform Commission (NDRC) jointly released the Design Instructions for School Planning and Construction after the Sichuan earthquake. These standards require that school sites are assessed before the schools are built in accordance with national regulation, performance objectives are determined by the country-level government, schools are built or retrofitted to meet performance objectives, and schools' furnishings and equipment are designed and installed to minimize potential harm they might cause to school occupants. The quality-monitoring bureau leads monitoring on the safety of equipment installation.

In Sichuan in the spirit of "building back better," the investigation, design, construction, supervision, inspection, and acceptance of school construction are conducted in line with relevant national construction standards. In Sichuan, steps were also taken to make sure that there are mechanisms to ensure that schools' maintenance is financed and executed. From 2000 to 2005 the first and second session of school renovation and maintenance was conducted. After 2006 a long-term mechanism for school building maintenance was put into place. MOE and UNICEF collaborated and prepared and revised the National Guidelines for Safe School Construction and Management and also collaborated to develop construction standards for kindergartens and preschools.

In 2009 the Ministry of Education (MoE) initiated a 3-year national "School Construction Safety Programme" to upgrade the safety of primary and middle school buildings all over the country with the aim of making schools the safest places in China. The program has several key elements:

- To screen and assess the quality of all school buildings across the country, to understand the qualities of buildings resistant to local disaster

risks, and to input the data and information of the assessment into a database

- To understand disaster risks in the regions where the schools are located, such as determining whether local disaster risks come from floods, landslides, earthquakes, or rainstorms threatening the safety of the school buildings
- To determine whether to repair, strengthen, or reconstruct school buildings that have not reached official standards based on the intensity of the identified disaster risks in the region where the schools are located
- To allocate funds and to start the construction work to upgrade the primary and middle school buildings that are at risk (International Recovery Platform 2010)

Colombia

The capital city of Colombia, Bogotá, is the most important political, administrative, economic, and cultural center of the country and has one million school children. Bogotá's population was estimated to be around 7.6 million in 2013. As a result of social investment over the past decade, more than 12 % live below the poverty line.

Among the most common hazard events affecting Bogotá are earthquakes and landslides. Although there has not been a severe earthquake in Bogotá since 1917, there is certainly the potential for one. Also, elsewhere in Colombia, 74 % of the schools in the cities of Pereira and Armenia suffered damage in the 1999 earthquake. Fortunately this occurred during the lunch hour, when no children were in the school buildings.

Several risk identification methods were put in place in the city prior to 2004. These include compilation of records of hazard events, generation of hazard maps, studies of physical and social vulnerability, and studies of environmental degradation. One of the means of reducing risk from earthquakes and landslides in Bogotá is the assessment of seismic risk of bridges, hospitals, and schools. This has become a core part of the city's economic and social development plan. Of these assessment programs, the best-known

is the Department of Education's effort to identify school seismic risk and to reinforce schools.

Much of the educational infrastructure in Bogota is more than 50 years old and does not meet minimal standards of safety. The Department of Education commissioned a systematic review of schools that ran from 1997 to 2003 (Secretaría de Educación del Distrito Capital de Santafé de Bogota 2000). This study covered approximately 2,800 buildings at 706 schools serving roughly 54 % of the student population in Bogotá. The other 46 % of the student population attends private schools and was not covered in this review. The review found that 434 of the schools presented high risk to students. Some 772 buildings at these schools fell into this category (16 %). The study also found that 60 schools had buildings in immediate and urgent need of reinforcement.

During the next city administration, from 2004 to 2008, after a detailed technical explanation on the need of schools' retrofitting, the city mayor decided to implement a retrofitting program of the 200 most vulnerable schools. A special risk analysis was performed on each building that was identified and prioritized. From this analysis, the structural reinforcement requirements were defined according to the seismic building code updated in 1998, with new special provisions for schools. The comprehensive improvement program had to be adjusted. Taking into account other technical, urban, economic, and environmental issues, many schools were not retrofitted: 67 schools were demolished and full restitution was made for a total of 107 new schools. This additional program was called the "50 Macro-schools plan" whose goal was to provide an educational infrastructure of maximum specifications and supplement the retrofitting and integrated improvement program. At the end, due to the costs involved in reducing vulnerability for 434 vulnerable schools, 201 schools were considered in critical condition according to risk studies that were conducted to prioritize and rank the schools. The reinforcement of these buildings had a cost about US\$ 200 million and the total program including the new schools was about

US\$ 430 million. Additionally, this program was based on the implementation of a teaching strategy to incorporate risk management into the culture. Both structural and nonstructural objectives were implemented to obtain a comfortable and safe school environment and a high-quality education service.

Assessing and reducing the risk to schools in Bogotá took place in a more general planning and management context. For example, zones of high risk of landslide, where no mitigation works are possible, are declared to be protected land. Human occupation is restricted in these areas as well as those considered at high risk to floods. In 2003 it was estimated that some 185,000 people lived in informal settlements in a total of 34,230 informal housing units. In Bogotá there are 173 illegal settlements that account for 14 % of the total land area. The city administration has developed a massive legalization program since 1995, thus reducing the number of informal settlements from 1,451 to its current number, an eightfold reduction in less than 10 years.

Nevertheless, as much as 60 % of the population of the city lives in informally constructed dwellings. While most of these are located in legal settlements, they still represent a challenge to seismic safety. The year 2000 land-use master plan for Bogotá contains hazard and risk maps that determined land use, details of special treatment for high-risk areas, and arrangements for issuance of building permits, as well as protection plans for utilities and services. The city also relies on community-based networks to control illegal land occupation and has developed a large-scale relocation program for families living in high-risk conditions.

The city of Bogotá has disaster risk reduction at the center of its planning process, and in this context school seismic safety ranks very high. Having diagnosed the scale and urgency of the problem in Bogotá, steps were being taken to reinforce the most hazardous school buildings. The challenges the city still faces include completing the program, making the retrofitting of the second priority group of vulnerable schools, extending its school safety program to private schools that cover nearly half the school-aged

population, and accelerating the rate of school reinforcement to improve the safety of children and teachers. Bogota has had different governance problems during the last two administration periods. Unfortunately, two mayors have been removed or suspended by the national general attorney due to corruption and inefficiency. Due to these situations and perhaps changing priorities in risk management, the school safety retrofitting program was not continued.

India: Delhi

NGO partners SEEDS and GeoHazards International (GHI), working with the Government of Delhi, demonstrated earthquake nonstructural risk reduction in a public school. The school welfare committee comprised of faculty, staff, and local community members learned to identify the nonstructural building elements and building contents that could fall, slide, or collide during a likely Delhi earthquake, as well as fire and evacuation hazards. They were exposed to simple low-cost techniques for reducing these risks (moving some items, fastening others) and came up with innovative solutions of their own. The logic of regular fire and earthquake drills became readily apparent to these new stakeholders. A handbook for schools on nonstructural risk reduction developed by the NGO partners and published by the Government of Delhi provides a new resource for generalizing these lessons.

India: Shimla

A small-scale demonstration project for school retrofit was carried out by SEEDS of India and GeoHazards International. Structural assessment of school buildings was carried out using a filtering method: The first step was low-cost mass scale Rapid Visual Assessment Survey of school buildings for potential seismic hazards. Based on these surveys a smaller number were selected for Simplified Vulnerability Assessment using limited engineering analysis. The highest-risk

buildings were identified for detailed vulnerability analysis. Retrofitting designs were drawn up for 20 schools and implementation of retrofit carried out in 8 schools. Guidelines were developed for retrofit and training of local masons and engineers and delivery of skill training. “Nonstructural mitigation plans” were carried out in 20 schools. An awareness campaign was designed to reach all 750 schools in the region including nearly 100,000 students, 7,500 teachers, and local builders, engineers, and officials (SEEDS 2006).

India: Uttar Pradesh

There are 23.5 million children attending school in this moderate-to-severe seismic risk zone. As part of the Education for All campaign, the State Government of Uttar Pradesh constructed 82,000 additional elementary school classrooms and 7,000 buildings in 2006–2007. To ensure seismic resilience of the buildings, UNDP provided earthquake engineers who examined the blueprints for the schools and modified the design to integrate seismic resilience. The marginal cost increase of 8 % to assure seismic safety was funded by the government.

For effective implementation of the new modified designs, training and orientation programs were initiated by the government, supported by the local UNDP office, building the capacity of 40 architects, over 200 engineers, and over 10,000 masons. To ensure transparent monitoring, the designs were widely circulated to the local communities where the schools were constructed, so they could monitor the quality, along with support of departmental engineers. To complete the safety of the schools, school-level safety committees were established, school emergency plans developed, and mock drills became a part of the school program.

In 2006–2007 the Elementary Education Department proposed to integrate earthquake-resilient design into all new school buildings. To prepare for this, one design of primary school buildings and two upper primary and three additional classroom designs were prepared with

detailed construction manuals. The disaster-resilient measures added 8 % to the construction costs. To cope with the massive scale of the project, a cascading approach prepared 4 master trainers for each of the 70 districts. These individuals in turn conducted trainings for 1,100 fellow Junior Engineers and Education Officers. Ten thousand masons were also trained. This program ensures that *every new school will be a safe school*. The problem of preexisting stock of 125,000 unsafe school buildings in need of retrofit remains to be tackled (Bhatia 2008).

Italy

Of the 8,102 municipalities in Italy, all are regarded as “seismically active.” Up until 2003, 2,965 of them (representing 40 % of the land surface and 45 % of the population) were considered “highly seismic” category. As a result, new construction in this category must observe stringent anti-seismic building codes. Subsequently, a more sophisticated classification was introduced, based on a 50-year recurrence interval and local estimates of peak ground acceleration (PGA). This had the effect of increasing the areas classified as highly seismic. Whereas in previous classifications, some municipalities were effectively regarded as aseismic, that is no longer the case and all 8,102 are now considered to be in a seismic zone to a greater or lesser extent. Italy bases its seismic classification on historical records and calculated return periods. Where these are a poor reflection of seismic hazard, it can underestimate the earthquake threat. Hence, severe damage occurred in northern Emilia and southern Lombardy in the earthquakes of May 2012, an area that had not had a major seismic disaster since 1574.

Several hundred municipalities are faced with the highest seismic risk in the Strait of Messina (including eastern Sicily) and southern Calabria (the Aspromonte) areas. The Apennine Mountains, which form the “backbone” of the Italian peninsula, are, in a tectonic sense, divided into blocks, which means that seismicity varies significantly from one locality to another. However, the vulnerability of buildings, including schools,

is almost universally high: modest local taxation revenues inhibiting retrofitting and maintenance. The largest seismic event of the past century remains that in Avezzano, in 1915.

An event that for Italians most encapsulates the seismic risk to schools was that of 31 October 2002 at 11:40 a.m. where in the Apennine town of San Giuliano di Puglia (population 1,195), the infants' school collapsed onto a class of small children (Augenti et al. 2004). Twenty-six small children and three teachers were crushed to death, and 35 children were rescued and lived. The building had been constructed with regional development funds in the early 1960s and had had its roof renewed a year before the earthquake. Evidently, the roof, of reinforced concrete with a ring beam, was too rigid and too heavy for the underlying structure, a concrete frame building with hollow-brick infill. There were signs that the quality of the cement was poor and the reinforcing steel was not used as it should have been. Moreover, despite mounting evidence that the Molisan Apennines are significantly affected by periodic earthquake activity, the revision of local building codes to take account of the new data on seismic risk cannot be applied as easily to existing structures as it can to ones that are about to be built (Augenti et al. 2004, p. S258).

The school at San Giuliano di Puglia succumbed because it contravened simple, well-known laws of dynamic response in structures affected by seismic acceleration. Inertial forces applied to a heavy roof sitting upon a weak frame structure amount to a recipe for tragedy.

Consider the schools of the Lunigiana, a surprisingly remote mountainous area of northwestern Tuscany. The Lunigiana has a sparse and dispersed population. Children attend elementary schools in the villages and secondary schools in major population centers, which they reach by bus or car. Many of the school buildings were constructed in the 1950s and 1960s to cater for the postwar population boom, and, in a rural area of relative economic stagnation, they have neither been built to be fully anti-seismic nor retrofitted. Indeed, in the minor population centers, they are decidedly dilapidated, nor do their staffs seem to have much interest in repeatedly

practicing evacuation drills. Yet the area awaits a magnitude 6 earthquake, which it is predicted may kill up to 120 people and injure more than a thousand. How many of them will be school children? This situation is typical of the seismic risk that affects highland Italy.

Elsewhere in Italy, much more progress has been made. The civil protection departments of several regional governments have introduced comprehensive *Scuola sicura* (safe schools) programs, notably in the northern regions of Lombardy, Piedmont, and Emilia-Romagna and in the autonomous Region of Sicily. The programs involve a combination of structural measures and school disaster management efforts, such as evacuation drills and lessons in civil protection. In many of the major cities, fire brigades and volunteer civil protection services are heavily involved in the programs, with public-private partnerships supporting attractive safety literature for school children.

Despite these developments, as in other seismic countries so in Italy, the building stock of schools continues to age and the civil protection educators must fight against the indifference of teachers, principals, and administrators. In many respects, mass mortality in Italian schools during recent seismic events has been avoided mostly by the lack of major earthquakes during school hours, a situation that will not prevail forever. Given an overwhelming need to upgrade the seismic performance of schools in Italy, the response of the national government has been to rank the buildings in terms of the deficit between design requirements (a function of the rules that prevailed at the time they were built) and the latest assessment of peak ground acceleration (PGA). Priority funding is given to those schools that have the greatest "PGA deficit" (Grant et al. 2007). However, the problem of unsafe school stock is simply too expensive to solve in the short to medium term.

Japan

In general, Japan is understood to be a leader in evaluating seismic risk and in implementing building codes for seismic-resilient construction.

Two publications available in English are MEXT's school seismic retrofit handbook (MEXT 2008a) and school nonstructural reference book (MEXT 2008b). In addition to high seismic performance standards for schools, following the 1995 Kobe earthquake, Japan also began providing guidance for mitigation of hazards due to building nonstructural elements and contents in schools. Nonetheless, the East Japan earthquake and tsunami [of 11 March 2011, with magnitude 9.0 earthquake off coast of East Japan], destroyed 6,284 in the affected region with different damage levels. Most of these were affected by tsunami waves, rather than earthquake. This was due to the location of the school buildings [proximity to coastal areas], the layout and structure of the buildings, and the subsidence of local tsunami retention walls. At the immediate aftermath, some schools [with higher stories] were used as temporary evacuation sites, and later people were rescued by helicopters. In some schools, located in higher ground, people took shelter, which lasted till 6 months in some cases, which caused serious disruption of school education.

The disaster pointed out several dimensions of role of schools and disaster education: (1) Schools can play an important public infrastructure of the community; however, the structural safety of the building needs to be linked to operational [including supplies of emergency kit] and locational issues. (2) School-community linkage is an important element, and local communities played an important role along with school teachers for the management of the evacuation sites. (3) While it is unavoidable to use the school as shelter, the continuity of education in emergency is a crucial issue. (4) School-based community recovery emerged as an effective concept, where the reconstruction of school building was linked to enhanced community cohesion (Shaw and Takeuchi 2012).

Nepal

A risk estimate for Kathmandu Valley, the economic, political, and technological hub of Nepal,

based on a scenario earthquake similar to the 1934 Bihar-Nepal earthquake, suggests that more than six million children and 140,000 teachers are at risk in schools (Bothara et al. 2002). A survey of 900 public schools in greater Kathmandu Valley estimated that more than 75 % of school buildings would suffer severe damage beyond repair (estimated at US\$7 million), and other 25 % would suffer repairable damage. In the absence of intervention, an estimated 29,000 children could be killed in their schools. With intervention 24,000 of these could be saved and buildings protected (NSET 2000).

A more recent assessment of school buildings in other parts of the country shows that more than 9,000 school buildings, more than 10 % of the total in Nepal, would suffer partial to complete collapse, resulting in very high casualties.

Most Nepalese school buildings are community built, by local craftsmen who have no formal training and are often illiterate. Technically trained people are not part of this process, unless it is funded by the government. Construction is characterized by the high degree of informality. The local availability of the construction materials, such as fired or unfired bricks, stone in mud mortar, and timber, controls the construction process. The use of modern materials such as cement, concrete, and steel bars is limited by affordability and accessibility and is confined to urban areas and areas accessible by transport.

In Nepal, there is no mandatory policy to control school design and construction. While some schools are supported by international donor agencies and/or the government requires design/drawing details, many are directly constructed by the communities without standard design criteria or technical supervision. Likewise, site-specific hazards are also not considered during the design and construction. Some design details are available, but they may not be entirely suitable for specific sites/locations. At most local levels, people lack the capacity to understand and implement the earthquake-safe construction method.

Low budgets for most school construction and lack of awareness and knowledge on the part of

graduate engineers of traditional and informal construction methodology result in most school buildings lacking earthquake resilience.

The National Society for Earthquake Technology in Nepal (NSET), a national NGO, conducted a program to strengthen existing school buildings and promote structural as well as nonstructural components of the school buildings for seismic safety, leveraging the decentralized, traditional, and informal approach to construction (Bothara et al. 2004). This program involved craftsman training, technology development and transfer, and community awareness raising. Many local masons became master masons. On-site master masons worked in residence, supported by visiting engineers with far-reaching effects. Shake-table demonstrations of typical versus seismic-resilient construction impressed communities with the effectiveness and feasibility of seismic-resistant measures. By raising awareness in schools, the entire community is reached because lessons trickle down to parents, relatives, and friends.

The approach developed took into account sociocultural and economic issues, with outreach to all stakeholders – school staff, students, local community, local clubs, and the local and central government. They have all been involved in the process so that they become aware of the risk and support the solution. School building construction was taken as an opportunity to train masons and to transfer simple but effective technology to others in the community, including house owners.

Following this approach, NSET retrofitted more than 40 schools, mostly of unreinforced masonry buildings. The program was found successful in transferring technology to local craftsmen who were quite keen to learn about the complete process and to adopt the technology. These masons became the propagators of the safety message in the vicinity of these schools and the replication of earthquake-resilient construction. The long-term sustainability of these impacts has yet to be assessed, but NSET's experience shows that seismic retrofitting and earthquake-resistant new construction are both affordable and technically viable.

Turkey

Turkey has more than eight million children attending schools in 64 provinces in the first- and second-degree seismic risk zones. The 1999 Kocaeli (moment magnitude Mw 7.4) and Duzce (Mw 7.2) earthquakes with approximately 20,000 fatalities raised awareness of the school safety question, and the 2002 (Mw 6.0) Afyon-Sultandagi and 2003 Bingöl (Mw 6.4) earthquakes kept awareness high.

During the 1999 earthquake in Kocaeli, 43 schools were damaged beyond repair, and 381 sustained minor-to-moderate damage. School was suspended for 4 months causing major disruption to the lives of families and children. In Istanbul, 60 km away, there was damage at 820 of the 1,651 schools. Damage at 131 of these sites necessitated at least temporary school closure. Thirteen were immediately demolished, and another 22 were later slated for demolition when retrofitting proved too costly. Fifty-nine schools were strengthened and 59 repaired.

In the Bingöl earthquake of 2003, out of 29 schools in the affected area, 4 school buildings collapsed completely, 10 were heavily damaged, 12 slightly or moderately damaged, and 3 undamaged.

Public schools in the Kocaeli earthquake fared better than residential buildings and private schools. Had children been at school during the Kocaeli earthquake, far fewer would have lost their lives. The fatality rate in residential buildings in the Kocaeli earthquake was 1.5 % in heavily damaged buildings and 16.5 % in totally collapsed buildings (Petal 2009). Similar damage in higher occupancy buildings of the same type would cause higher fatality rates. In the single example of the school dormitory in the Bingöl earthquake where 84 children died, the fatality rate was 44 %. Average risks are theoretical and don't occur. Instead, the reality is that either the school is not occupied and no one dies or it is occupied and the fatality rates are high, and the tragedy is wholly unacceptable.

There is much that is right with school construction in Turkey. As a result of an assigned

importance factor of 1.5, public schools are designed to withstand a 50 % increase in earthquake design loads. Schools have regular symmetrical structural designs, and those that are only one or two stories have fared well, for the most part meeting standards for life safety, if not continuous occupancy. The lethality of school buildings is almost entirely attributable to shoddy construction and is particularly lethal in taller buildings that may also have design defects.

For decades all public construction was under the authority of the Ministry of Public Works and Settlement. Earthquake building codes on the books since the 1930s were updated most recently in 1976 and 1998, yet the existence of these laws has not guaranteed the safety of construction. The reasons are numerous.

Beyond an undergraduate or graduate degree, there have been no independent or nonacademic professional qualifications, proficiency standards, continuing education requirements, or licensure for architects or engineers nor any qualifications for building contractors. There are also no guidelines for reliable and systematic building inspection during construction. Penalties for noncompliance with building codes are beset with bureaucratic and social impediments and often are simply not applied. Legal liability in some future event with low-frequency occurrence can hardly be a deterrent with so many to share blame. Public construction has also suffered from a standard (though not legally required) preference for the lowest bid in public tenders. The civil service employment system also lacks proficiency standards and qualifications for professional staff; so at the local level there is a wide variety in the capacity for project supervision and control. Wage and salary levels are low, and there has been opportunity for both favoritism and corruption. There are no ombudsman or advocacy services to support consumer whistle-blowers.

Istanbul provides a dramatic example of three overlapping tasks:

- Immediate response to damages caused by the 1999 earthquake

- Implementation of a comprehensive retrofitting and replacement for seismic risk mitigation
- Follow-through on an ambitious program of school expansion and construction initiated to respond to the acute shortage of class space occasioned by three additional years of compulsory education enacted in 1998

After the 1999 earthquake, the responsibility for school construction was shifted to the Ministry of Education's Division of Investments and Facilities (DIF). In turn, DIF appointed consultants from the private sector to oversee the new facility design and construction. DIF also developed standard designs for the new facilities, and new school construction was financed by a combination of government funds and charitable contributions raised by not-for-profit foundations. New construction and procurement laws also went into effect; however, the cumulative impact of these changes and pressures is not yet known (Gülkan 2004).

In Istanbul of all projects the highest priority is given to regional boarding schools, then to schools in the 12 highest-risk districts, and to those in proximity to the Marmara seacoast. The overall mitigation and retrofit effort targets more than 1,800 buildings that constitutes the 80 % of stock predating the 1998 Building Code. This ambitious program is budgeted for US\$320 million (Yüzügüllü et al. 2004).

An additional problem in Turkey is that awareness of nonstructural hazards remains low. Classroom doors often open inward and shelving and laboratory equipment remains unfastened. However, concern that children advised to "drop, cover, and hold" might be injured by flimsy wooden desks led to production and distribution of 80,000 steel desks to more than 500 schools in the most vulnerable areas.

Schools are a well-distributed means of public education, and children can play a leading role in the dissemination of public safety messages. Thus, Professor Isikara, former head of KOERI, a major earthquake research institute in Istanbul, toured the country visiting schools, becoming known as "Grandpa Quake," and produced the

first children's books and popular educational and rap music cartoons for earthquake awareness. Both the Istanbul Governor's Office and KOERI's newly established Istanbul Community Impact Project (ICIP) produced handouts distributed to all school children. Schools also received books and CDs. At the national level an introduction to natural hazards was integrated into the primary school curriculum in Environmental Studies in 2002. Annual school-wide earthquake drills and preparedness and remembrance activities were initiated on 11 November 2001, to coincide with the Duzce earthquake anniversary.

Between 2001 and 2003, a cascading model of training and instruction called "ABCD Basic Disaster Awareness" was implemented by KOERI's Istanbul Community Impact Project. A curriculum was developed to address specific assessment and planning activities, physical risk reduction, and response preparedness measures to be taken prior to a disaster. This was a significant reorientation from previous "awareness" programs that began with what to do "during the shaking." A single full day of instructor training for 3,600 teachers was provided in collaboration with the Ministry of Education Provincial Directorates and outside donors. These teachers in turn communicated with 121,000 school personnel and through them with 1.68 million school children as well as with 700,000 parents. The project established an Internet-based monitoring system to monitor dissemination. Based on the success of this project, almost a decade later an even more ambitious scale-up was attempted.

In 2010 the Ministry of National Education is committed to taking the program nationwide with technical support from Risk RED, in a follow-up project supported by the American Red Cross. A distance-learning self-study curriculum was developed consisting of 1 course (10 lessons) in household disaster preparedness and 1 course (9 lessons) in school disaster management, with the goal of reaching 25,000 school-based instructors (Petal and Türkmen 2012). During the first year of deployment of the courseware in 2012, more than 79,000 MoNE employees completed one or more lessons. More than 65,600 users completed the School Disaster and Emergency

Management Course and more than 50,000 passed the final test. Almost 50,000 users have completed the Individual and Household Disaster Preparedness Course and more than 40,000 users have passed the final test. There were a total of 114,700 course completions, 92,800 final tests passed, and more than one million lessons were successfully completed by users. Within the first year, active users completed an average of almost 15 lessons each (Petal and Türkmen 2012).

In 2005, a loan to implement the Istanbul Seismic Risk Mitigation and Emergency Preparedness (ISMEP) was funded by the World Bank and European Investment Bank, to help prepare Istanbul for a probable earthquake in the Marmara Region. The project had broad aims to enhance the institutional and technical capacity of emergency management-related institutions, raise public awareness, assess priority public buildings for retrofit or reconstruction, and support building code enforcement (www.ipkb.gov.tr). "Component B" addressed seismic risk mitigation for priority public buildings. This included a feasibility study for retrofit of 1,128 education buildings on 796 school sites. Of these, 506 were strengthened and 148 were reconstructed. At the time of completion of the project, further 34 were slated for retrofitting and 27 for reconstruction.

USA: California

In 2008, seismic safety advocates in California launched the first Great Southern California ShakeOut. The now annual event has grown to include five western states and several central US states and has been conducted in cities in five other countries. In its first year four million children and adults participated through 207 school districts plus almost 750 individual schools (Risk RED 2009). In 2013 there were more than 9.6 million participants in the Great California ShakeOut drill. Globally there were almost 25 million registered participants in similarly inspired Great ShakeOut Drills. Of these almost 75 % were school-based participants (Earthquake Country Alliance 2013). This regular public

awareness event has heightened interest and concern in safe school facilities. With 3.6 million children enrolled in 262 public school districts in seven counties in Southern California, a major earthquake in the region could cause an unprecedented catastrophe for schools, children, and teachers.

School seismic safety has been a policy and a community concern in California since the 1933 Long Beach earthquake, and school emergency planning has been required statewide since 1984. With 75 years of public policy leadership to support school safety, new school construction standards are higher than those for regular buildings and come close to assuring life safety. An advice regarding non-structural mitigation measures (fastening furnishings, etc.) has been in place for 20 years and requirements for such mitigation have been in place for 10 years.

The 1933 Field Act implemented immediately after the Long Beach earthquake that year required that schools be built to 15 % higher performance standards than normal construction. In 1938 the Garrison Act required examination and improvements to pre-1933 construction, but went unenforced until 1968. The Uniform Building Code enacted in 1976 is now the current standard for safe school construction. The Field Act has been hailed as a high point in school seismic safety and California schools are considered the safest in the United States. Some school facility managers feel that its requirements are too stringent and too costly, and many seismic safety advocates feel that it does not go far enough. The Act requires that structural plans be prepared by licensed structural engineers and approved by an independent state agency (the Division of the State Architect (DSA)). Schools have continuous on-site inspection (rather than periodic), by a DSA-approved project inspector. Project architect and engineers must perform construction observation and administration, and a final verified report must be filed by the project architect, engineers, inspectors, testing labs, and the contractor (State of California, DSA 2007, 2009).

In 2007 the California Seismic Safety Commission found that: (1) The cost of compliance

with the Field Act is incremental and minimal. (2) Timeliness, consistency, accuracy, and communication are being improved by the Division of the State Architect. (3) The exemplary performance of school buildings is directly attributable to the stringent seismic design provisions, plan review, field inspection, and testing required by the act and which go beyond the standard building codes. (4) All public schools should be covered. A 2009 study reported that in the four major earthquakes since the Field Act, there have been no public school collapses. The construction to Uniform Building Code, the special enforcement and quality control provisions, an oversight by the Office of the State Architect, and the 2003 publication guiding mitigation of nonstructural hazards are all judged to be successful. Nonetheless, the work is not yet complete.

In 1999 Assembly Bill 300 required desk assessment of 9,659 pre-1978 school buildings. The final report based on woefully incomplete records was released in 2002 and found 7,537 potentially vulnerable buildings requiring detailed seismic evaluation. The cost of retrofit was estimated at \$4.5 billion (State of California 2002). Due to fear of planning and financial implications, details were not released to the public. In 2005 an investigative reporting series by California Watch finally achieved this (Risk RED 2009). In 2011, 20,000 uncertified projects were released on an interactive map (California Watch 2011).

A review of school seismic safety in California identifies four remaining areas of concern:

- There are still some 7,537 school buildings in California constructed before 1978 that are of questionable safety.
- Portable classrooms, which may account for one-third of all classrooms in California, may be particularly hazardous if not properly supported and fastened.
- Private schools are not currently required to meet these same construction standards as public schools.
- Nonstructural mitigation measures continue to require consistent application to protect children and adults from both injury and death.

- Each school district and private school is strongly recommended to conduct its due diligence and report on these issues transparently to parents, staff, and students, so that collective action can be taken to address these serious vulnerabilities. Neither fear nor California's persistent financial crisis in the education sector should be acceptable excuses for inaction (Risk RED 2009).

Summary

Most of these case studies have focused on the primary importance of safe school facilities, through both standards for new school construction and strategies for school vulnerability assessment and planning for retrofit and replacement. While sound earthquake engineering expertise is fundamental to advocacy, communications, planning, and execution of these efforts, it is also important to retain the perspective of the primary beneficiaries: children, teachers, and school communities. This necessitates going beyond the obvious: site selection, design, and construction. As some of these case studies indicate, it is also important to consider nonstructural mitigation, ongoing maintenance, safe access to school, school function as temporary emergency shelters, and even structural awareness education and the use of construction as an educational opportunity for children and communities. By taking this wider (and multi-hazard) view, the focus on safe school facilities overlaps with both ongoing school disaster management and with risk reduction education.

Cross-References

► School Seismic Safety and Risk Mitigation

References

- Augenti N, Cosenza E, Dolce M, Manfredi G, Masi A, Samela L (2004) Performance of school buildings during the 2002 Molise, Italy, earthquake. *Earthq Spectra* 20(S1):S257–S270
- Bendimerad F (2004) Earthquake vulnerability of school buildings in Algeria. In: Group meeting on earthquake safety of schools. OECD, Paris, pp 35–44
- Bhatia S (2008) Earthquake-resistant school buildings in India. *Urban Plan Des* 161(4):147–149
- Bothara JK, Guragain R, Dixit A (2002) Protection of educational buildings against earthquake. National Society for Earthquake Technology, Nepal
- Bothara JK, Pandey B, Guragain R (2004) Seismic retrofitting of low strength unreinforced masonry non-engineered school buildings. *Bull N Z Soc Earthq Eng* 37(1):13–22
- California Watch (2011) K-12/public safety/spotlight/on shaky ground. 22 Dec 2011. Accessed 3 Feb 2012
- Earthquake Country Alliance (2013) Great shake out earthquake drills. <http://www.shakeout.org>. Accessed 30 March 2015
- Grant DN, Bommer JJ, Pinho R, Calvi GM, Goretta A, Meroni F (2007) A prioritization scheme for seismic intervention in school buildings in Italy. *Earthq Spectra* 23(2):291–314
- Gülkan P (2004) Seismic safety of school buildings in Turkey: obstacles impeding the achievable? In: OECD (ed) Keeping schools safe in earthquakes. Organization for Economic Cooperation and Development, Paris, pp 64–87
- International Recovery Platform (2010) Case reports – China. United Nations International Strategy for Disaster Reduction. http://www.recoveryplatform.org/outfile.php?id=548&href=http://www.recoveryplatform.org/assets/recoverycases_reports/china/CHINA_300910.pdf. Accessed 16 Feb 2014
- MEXT (2008a) School seismic retrofit handbook <http://www.nier.go.jp/shisetsu/pdf/e-taishinjirei.pdf>. Accessed 30 March 2015
- MEXT (2008b) School non-structural reference book <http://www.nier.go.jp/shisetsu/pdf/e-jirei.pdf>. Accessed 30 March 2015
- NSET (2000) Seismic vulnerability of the school buildings of Kathmandu valley and methods for reducing it. Kathmandu Valley Earthquake Risk Management Project implemented by National Society for Earthquake Technology-Nepal, Kathmandu
- Petal M (2009) Evidence-based public education for disaster prevention: the causes of deaths and injuries in the 1999 Kocaeli earthquake, VDM. Doctoral dissertation, UCLA, 2004
- Petal M, Türkmen Z (2012) Dreams for Turkey: a case study of scale and reach of distance learning self-study for individual and household preparedness and school disaster management. <http://www.riskred.org/activities/dreamsforturkey.pdf>. Accessed 30 March 2015
- Risk RED (2009) School disaster readiness: lessons from the First Great Southern California ShakeOut. <http://www.riskred.org/schools/shakeout2009.pdf>. Accessed 31 March 2015
- Secretaría de Educación del Distrito Capital de Santafé de Bogota (2000) Análisis de vulnerabilidad sísmica de las edificaciones de la Secretaría de Educación del

- Distrito y diseños de rehabilitación de algunas de ellas. Proyectos y Diseños, P&D, Ltda, Bogotá
- SEEDS of India (2006) Earthquake safety in mountain regions. School Earthquake Safety Initiative, Shimla/Delhi
- Shaw R, Takeuchi Y (2012) East Japan earthquake and tsunami: evacuation, communication, education and voluntarism, Research Publisher
- State of California Department of General Services (2002) Seismic safety inventory of public schools (AB300 final report)
- State of California, Seismic Safety Commission (2007) The field act and public school construction: a 2007 perspective. Accessed 30 March 2015
- State of California, Seismic Safety Commission (2009) The field act and its relative effectiveness in reducing earthquake damage in California's public schools. Available online at <http://www.seismic.ca.gov/pub/CSSC%2009-02%20Field%20Act%20Report.pdf>. Accessed 16 Feb 2014
- Taylor G (1989) Seismic assessment of Vancouver school buildings. TBG (Transit Bridge Group) Engineering Firm, Vancouver
- Yüzügülü Ö, Barbarosoglu G, Erdik M (2004) Seismic risk mitigation practices of school buildings in Istanbul. In: OECD (ed) Keeping schools safe in earthquakes. Organization for Economic Cooperation and Development, Paris, pp 176–183

Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis

Nicola Nisticò and Alessandro Proia
Dipartimento di Ingegneria Strutturale e
Geotecnica, Università La Sapienza, Rome, Italy

Synonyms

Nonstructural elements; Secondary
Nonstructural Elements; Secondary Structural
Elements.

Introduction

Seismic evaluation of a construction has to include, with a given level of accuracy, the interaction phenomena between (1) soil,

(2) foundation, (3) structural part of the construction to be intended as the part in elevation, and (4) nonstructural part to be intended as a part of the construction with minor or no task to the structural capacity.

From a conceptual point of view, it could be easy to assert the following principle:

the seismic evaluation of a construction has to be performed based on 1) the definition of Structural Resisting System (SRS) 2) a proper model implementation of the SRS 3) a proper analysis of the SRS.

Given that the previous principle could resolve the problem, the SRS verification remains one of the goals of the analysis process: it can be pursued with a probabilistic approach (ATC 58 2012) (1) defining the required performance (2) based on the predictable loss (direct and indirect) consequent to a given seismic event.

A general approach, in which the strategy can be framed, is the performance-based approach that (1) defines a given number of performance levels (PLs), (2) chooses a seismic level for each PL, and (3) requires a given performance for each PL. Examples of PLs are operational, immediate occupancy, life safety, and collapse prevention.

The performance-based approach delegates the applicator of it the definition or selection of the most appropriate tools to be applied for the (1) identification of the resisting system (RS), (2) structural analysis (modeling included), (3) capacity definition, and (4) verification.

The first step (RS identification) is not an easy task: the RS includes the soil and the construction that on the other hand can be split in structural and nonstructural elements which are not supposed to have any role in the global resistance of the construction with regard to neither of the so-called vertical loads or of the horizontal loads such as those that schematize wind and seismic actions.

The nonstructural elements, having no role in the seismic capacity, are generally considered as additional weight to be included in the mass evaluation, neglecting the structural interaction between them and the structural resisting system.

For many of the nonstructural elements, the absence of interaction can be considered realistic so that they can be considered as attached elements (from which the name attachment derives) that having a proper structure (mass, stiffness, structural capacity) have to be verified with regard to the seismic action transferred to them (from the resisting system).

The attachments (as previously introduced) are objects with their morphology so that they can be schematized with either continuous or discrete models opportunely connected to the structure they are attached to. Examples of attachments are (1) furnitures, (2) technical systems, and (3) art objects of a museum.

If the structure-attachment interaction is negligible as well as the soil-structure interaction, a cascade procedure can be adopted evaluating the seismic action (in terms of time histories or response spectrum) (1) at the base of the structure in elevation and then (2) transferring it at the points to which the nonstructural element is attached.

The cascade procedure, not considering the primary (SRS) and secondary (attachments) systems as a whole entity (PS system), cannot be adopted when the two systems are tuned; that means their periods are similar and the attachment could be acting as a tuned mass damper (TMD) for the primary system.

The seismic analysis of the attachments can be performed by means of different strategies (Villaverde 1997; Chen and Soong 1988) among which linear and nonlinear analyses can be included: the PS system can be analyzed as a global system with an evident computational effort.

The need of efficient and accurate methods to analyze the PS systems inspired methodologies (Igusa and Der Kiureghian 1985a, b) based on (1) modal synthesis, (2) perturbation theory, and (3) random vibrations.

The decoupling of the secondary system (from the P system) allows the evaluation of the seismic action in terms of Floor Response Spectra whose approach is similar to the approach that governs the decoupling between soil and structures: a Response Spectra is defined and applied at the

base of the structure, including in it the effect of the propagation of the seismic action in the soil. Similarly, a spectra (FRS) is defined at the base of the attachment, including in it, with a cascade procedure, the effect of the propagation of the action at the soil (at first) and, subsequently, at the elevation structure.

The definition of a Response Spectra at the base of the attachment solves the problem since the attachment can be analyzed with traditional methodologies that are, for example, Seismic Modal Analyses or Time Histories Analyses based on acceleration histories compatible with the given Floor Response Spectrum.

Usually the effect of the propagation of the seismic event (from the soil to the attachment) is performed considering a linear behavior of the primary structure: this is supposed a realistic assumption for new conceived structure when operational and immediate occupancy PLs are considered.

The linear structural behavior of the principal system could be considered as nonrealistic in some cases, where the system's nonlinearity could produce effects (on the attachment) more severe than those evaluable with a linear behavior assumption (Chaudhri and Villaverde 2008).

The attachments, as discussed so far, are secondary elements that do not give any contribution to seismic resistance of the primary system and, in these terms, can be classified as secondary nonstructural elements (NSEs) to be distinguished from the secondary structural elements (SEs) that have a negligible role in supporting the seismic action but can have a specific role in transferring the vertical load to the foundation system.

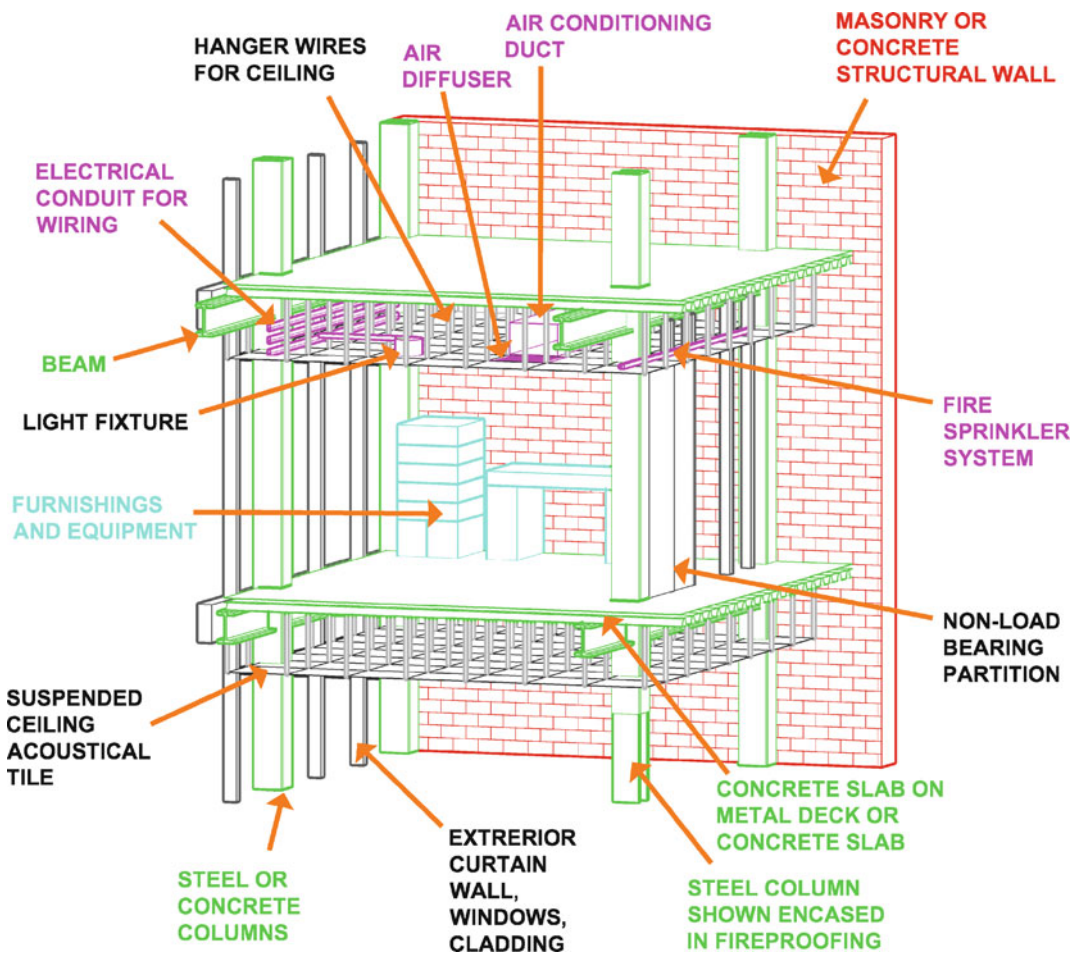
Secondary elements, either attachment (secondary NSEs) or structural elements (SEs), are both systems having their structure, opportunely linked to the primary structure: they have to be adequately modeled in order to be analyzed with the strategies common to the seismic branch, such as (1) static analyses (linear and nonlinear), (2) modal response spectrum analyses, and (3) time domain analyses either linear or nonlinear.

Morphological and Phenomenological Aspects Versus Modeling and Analysis

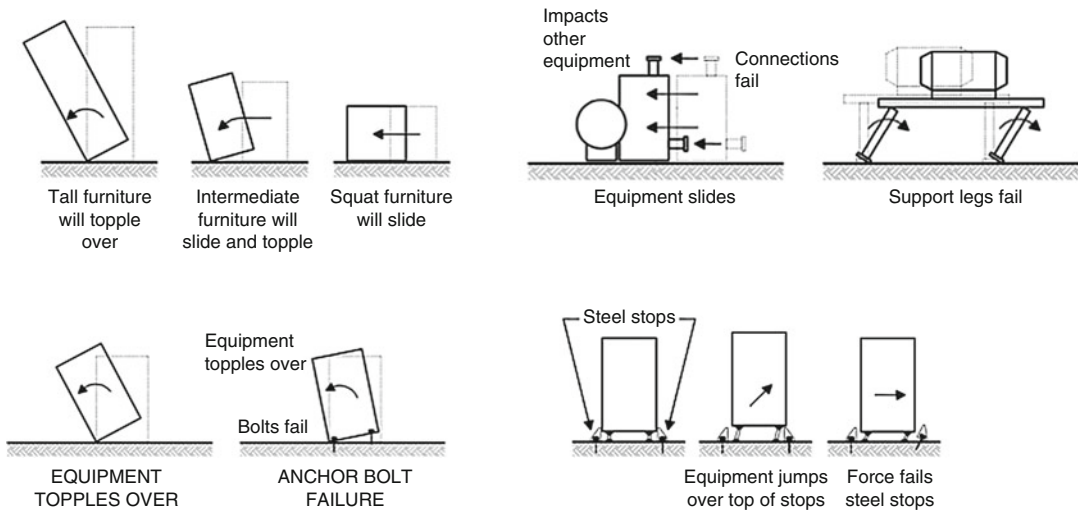
Depending on the structural resisting system typology and the construction usage, the NSE typology can be wide (see Fig. 1, for the building case) and their taxonomy can be found in Taghavi and Miranda (2003) where a comprehensive data-base of nonstructural components is presented covering different aspects such as, among others, cost information from which is deducible that the structural cost of a building could be not relevant with respect to the global cost: the office buildings structural costs, even if relevant, are only 18 % of the total cost of construction that can be split in

cost of (1) structural elements, (2) secondary structural elements, and (3) nonstructural elements such as the contents are. The cost of the nonstructural elements can be estimated to reach the 70 % of total construction costs if the hotel buildings are concerned, while it is lower in office buildings (62 %) and hospitals (48 %) where contents (such as medical equipment) can be estimated to be 44 % of the total cost.

Economic loss due to seismic nonstructural damage can be relevant: during the 1994 Northridge earthquake, the nonstructural damage was (Kircher 2003) about 50 % of the global building damage which was estimated to be \$18.5 billion.



Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Fig. 1 Typology of building nonstructural elements. Reproduced from FEMA 74 (2005)



Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Fig. 2 Principal rigid body mechanism

Most of the NSEs have limited seismic performance because they are not properly attached to the primary structures, so that, depending on their slenderness, they can (see Fig. 2) (a) topple over, (b) slide and topple, and (c) slide. The loss of capacity of the NSEs or their connections can cause damage to other equipment (see Fig. 2) and injury to people, so careful attention has to be paid to the design of the connection (see Fig. 2).

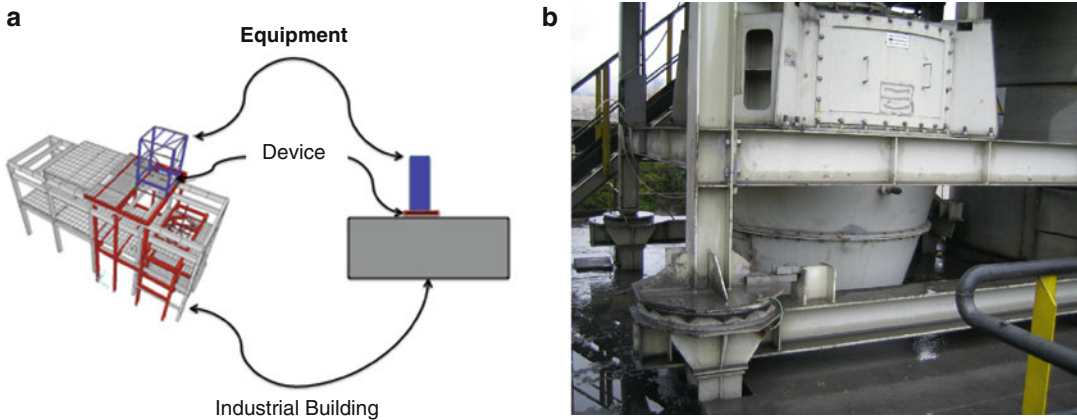
Unlike the old conceived NSEs, the new-generation elements can have a good seismic performance, thanks to the wide range of connections that can be adopted to link the NSE to the structure. Depending on the case at hand, the design can include (1) seismic joints opportunely designed to accommodate seismic displacements, (2) seismic isolators to reduce the acceleration level, and (3) dissipative device to reduce the level of acceleration, velocity, and displacement.

So the new-generation NSEs cannot be conceived without an adequate strategy for the connection design, an example of which is reported in Fig. 3 where the case of a machinery (for cement production) mounted on a steel structure attached on a reinforced concrete building has been reported: dissipative devices have been introduced at the base (of the steel structure) to

reduce the seismic action on both the machinery and building.

Referring (see Fig. 3) to the previously introduced example (where the equipment can be considered as an attachment of the primary structure), the following indications can be given for the seismic analysis:

1. If the NSE is rigidly connected to the structure and its mass (M_{NS}) is not negligible with respect to the building mass (M_S), a global analysis of the PS system is required. In case of modal spectra time history (TH) analysis, some approximation in the damping definition is needed due to the different damping of the attachment (steel structure) with respect to the reinforced concrete structure. More appropriate step-by-step TH analyses can consider the element damping, properly modeling it by means of dashpots when *nonclassical* damping is present.
2. If M_{NS} is negligible with respect to M_S , but the NSE period (T_{NS}) is close to structural period (T_S), the so-called *tuning* happens and a global analysis of the PS system is required, not excluding positive effect.
3. If M_{NS} is negligible with respect to M_S , and T_{NS} is not close to T_S , the decoupling could be considered and the strategy for the analyses



Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Fig. 3 (a) Cement industrial building. Machinery mounted upon an *rc* structure. (b) Cement Industrial Building Applications: Italy

could be oriented to the definition of the seismic action at the level of the connection, considering the structural system as a stand-alone element subjected to a seismic action at the base of it.

The following strategy can be considered: (a) time history analyses applied either to the whole system or to the stand-alone attachment, considering the acceleration histories recorded at the level of the attachment connection and (b) spectrum-based analyses defining an acceleration spectrum consistent with the time histories recorded at the points where the attachment is connected to the structures: the generated spectrum is usually named Floor Response Spectrum (FRS) even if (as the case reported in Fig. 4) the FRS has been evaluated where the attachment is linked and not at the floor level.

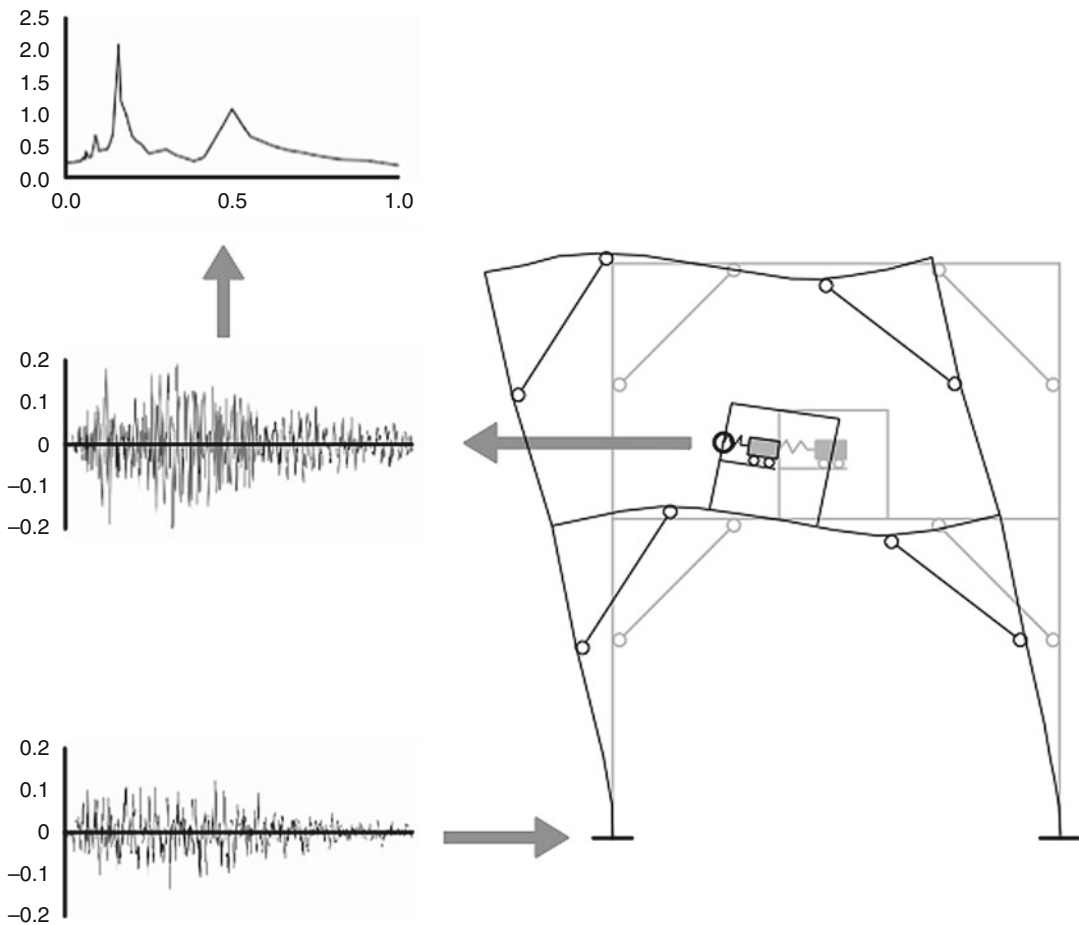
The FRS definition is an important task for the NSE analysis, being the referred analysis tool to be adopted, due to its recognized simplicity in conjunction with good level of reliability (in those cases where the decoupling can be adopted). It is possible to affirm that the usually adopted acceleration spectrum *is* to structural analysis *as* the FRS *is* to the analysis of nonstructural elements, so that the FRS is a period-dependent function that can be evaluated for different value of the attachment damping

given a specific soil and structure characterized by their own periods and dampings.

Differently from the secondary nonstructural elements, the secondary structural elements (SE) require different strategies of modeling and analysis. It is useful to introduce them as reported in CEN (2008): some structural elements (i.e., beams and columns) can be designed as seismic secondary elements, neglecting their contribution to the global seismic resistance so that their stiffness and strength can be neglected. As further specified in CEN (2008), the SE and the joint (that link them with structure) have to be designed considering (1) the vertical gravitational load, (2) displacement consequent to the seismic action, and (3) second-order effects that include the flexural moments evaluated considering the deformed element shape (P-Delta effect).

Clearly the previous definition of the SE supposes that they have a negligible influence in the global structural behavior.

Starting from the classification of the secondary elements in NSEs (attachments) and SEs, the following can be asserted: (1) given a construction, it includes a principal (P) and a secondary (S) structure; (2) if the S structure has a negligible influence on the P structure, the whole construction (PS) structural behavior can be decoupled; (3) the S structure can be classified in structural (SEs) and nonstructural (NSEs) elements; (4) the SEs have to be designed for vertical gravitational



Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Fig. 4 FRS generation

load (transferred from the PS to them, including the self-weight loads) considering the seismic-induced displacement (P-Delta effect included); (5) the NSEs have to be verified with regard to self-weight loads and seismic action transmitted by the P structure; (6) the secondary structure elements and their supports (links) have to be verified in order to avoid that their partial or total failure can induce injury to people or important objects; and (7) if the interaction between P and S systems is not negligible, a global PS analysis is required.

In the following, some of the principal characteristic regarding secondary structural and nonstructural elements will be described.

Secondary Structural Elements

Typical case of SE is internal and external building partitioning system (Glass Systems included): their seismic contribution is usually neglected in the seismic analysis (1) accepting (for severe earthquake) their damage and (2) imposing that out-of-plane collapse (see Fig. 5) has to be prevented.

The infilled partitioning system can have a role in transferring the vertical load even if they can have a minor contribution in seismic global capacity. If they have no role in the vertical load as well as in the seismic P structure capacity, their classification as NSE (attachments) is reasonable.

Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis,

Fig. 5 Molise (Italy): Seismic event occurred in 2002 (October 31 ($M = 5.4$) and November 1 ($M = 5.3$)). Example of damaged infilled frame: in-plane and out-of-plane mechanisms



If no flexible joints are considered (in between the P and S systems they represent), the absence of collaboration with the P system is not judicable by means of qualitative considerations, but it can require a structural analysis of the PS system including them as structural elements. In this case, the designer can follow some suggestions such as those included in CEN (2008) that consider a structural system as SE if its global stiffness is lower than 15 % of the P system stiffness. The models concerning the infilled frames systems are well known.

The damage level in the partitioning system is usually controlled imposing a threshold to the interstory drift (see Table 1) as function of the performance level and construction usage (ASCE 2002).

Attachments

Typical cases of attachments are parapets, windows, partitioning systems, antennas, electrical power systems, and furnitures. Depending on their components, they can be sensitive to the seismic acceleration or deformation (see Table 2).

Modeling and Analysis

NSEs are elements characterized by their mass and stiffness, and independently of the seismic action they are subjected to, they can be modeled and analyzed based on FEM strategies considering either their linear or nonlinear behavior.

In general the NSE is a system composed by subsystems with a structural complexity (see Fig. 6) that can require 3D complex models to be calibrated by means of experimental tests (Fig. 6a) including identification strategies: dynamic tests can be carried out by means of shaking tables (Fig. 6c).

The experimental tests in support of modeling and analysis implementation are especially required either when the importance of NSE usage is considered strategic or when the cost of it justifies the experimental activity. In some cases, a qualification procedure can be required, generally ruled by international standard (Gilani et al. 1999; IEEE 2005).

Modeling

Modeling has to take into account all the components that give stiffness and strength

Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Table 1 Drift control: usually adopted values as function of usage and performance level

Performance level	Damage description and downtime/loss	Drift control
Immediate occupancy: usually required for construction which usage is considered strategic	Negligible structural damage; essential systems operational; minor overall damage. Downtime/loss: 24 h	0.3 % (stiff joint), 0.6 % (deformable joints)
Life safety	Probable structural and nonstructural damage; no collapse; minimal falling hazards; adequate emergency egress. Downtime/loss: possible total loss	0.5 % (stiff joint) 1.0 % (deformable joints)
Collapse prevention	Several structural and nonstructural damage; incipient collapse; probable falling hazards; possible restricted access. Downtime/loss: probable total loss	Not required

contributions, including the connection elements that, if needed, have to be modeled as nonlinear elements.

In many cases, such as the bushing sketched in Fig. 6a, an accurate modeling requires informations about all the subcomponents (coil springs, valves) in terms of mass stiffness and strength. The needed informations are not usually known and the element investigation has to be supported by means of experimental tests devoted either to global information acquisition (frequencies, modal shapes) or to evaluation of the level of

performance given a defined seismic action. Experimental tests can include shaking table tests or static tests: this aspect is strictly linked to the qualification process (IEEE 2005).

Seismic Action Modeling and Structural Analyses

Seismic action can be simulated according to the usually adopted strategies that, for the case at hand, include (1) time histories (usually in terms of acceleration) and (2) response spectrum finalized either to modal analyses or to static linear or nonlinear pushover analyses.

Seismic level will depend on the referred performance level that (see Table 1) identifies the required performance associable to a seismic event with a given return period, to be defined based on cost-benefit analysis.

General rules valid for secondary elements are the following:

1. Mass and stiffness uncertainties have to be considered together with spatial distribution of seismic effect in case of extended SE systems.
2. Seismic effects on SE have to take into account, in general, both horizontal and vertical components to be evaluated based on a structural model of the principal system.
3. If the SE behavior can be decoupled from the principal system, the datum method for the evaluation of the peak acceleration at the SE is based on the Floor Response Spectrum (FRS) that given an SE element, with a defined structural period and damping, attached to a given part of a structure, having its mechanical properties, subjected to a given seismic event (*E*), allows to define the peak acceleration to which the element will be subjected when the seismic event (*E*) is transferred at the base of NS element.

Based on the knowledge of the FRS, one of the following methods can be adopted: (a) static equivalent forces (including nonlinear pushover analysis), (b) modal analysis, and (c) time history (linear or nonlinear) analyses based on accelerograms compatible with the FRS.

Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Table 2 NSE classification (ATC/BSSC 1997) and element sensitivity with regard to acceleration and deformation

Component	Sensitivity		Component	Sensitivity	
	A	D		A	D
A. Architectural			B. Mechanical equipment		
1. Exterior skin			1. Mechanical equipment		
Adhered veneer	S	P	Boilers and furnaces	P	
Anchored veneer	S	P	General mfg. and process machinery	P	
Glass blocks	S	P	HVAC equipment, vibration isolated	P	
Prefabricated panels	S	P	HVAC equipment. Nonvibration isolated	P	
Glazing systems	S	P	HVAC equipment, mounted in-line with ductwork	P	
2. Partitions			2. Storage vessels and water heaters		
Heavy	S	P	Structurally supported vessels (category 1)	P	
Light	S	P	Flat bottom vessels (category 2)	P	
3. Interior veneers			3. Pressure piping	P	S
Stone, including marble	S	P	4. Fire suppression piping	P	S
Ceramic tile	S	P	5. Fluid piping, not fire suppression		
4. Ceilings			Hazardous materials	P	S
(a) Directly allied to structure	P		Nonhazardous materials	P	S
(b) Dropped, furred, gypsum board	P		6. Ductwork		
(c) Suspended lath and plaster	S	P			
(d) Suspended integrated ceiling	S	P			
5. Parapets and appendages	P				
6. Canopies and marquees	P				
7. Chimneys and stacks	P				
8. Stairs	P	S			

A acceleration sensitive; D deformation sensitivity; P primary response; S secondary response

Floor Response Spectra-Based Evaluation

Floor Response Spectra are functions that define the response spectrum of a given response parameter (e.g., acceleration, velocity, displacement) as a function of period and damping of a given structure (attachment) localized at a given point of the construction.

The generally adopted technique for the FRS definition consists in (1) analyzing the P structure (to which the S structure is attached) in the time domain, considering n time histories (e.g., acceleration TH), (2) evaluating (for each TH), at a given point of the structure, the TH of the acceleration and the related response spectrum for a given damping value, and (3) defining one representative spectrum (based on the n available FRSs) having a given overcoming probability (usually a 50 % probability is considered): for

low values of n (e.g., minor than 7), an envelope spectrum has to be considered.

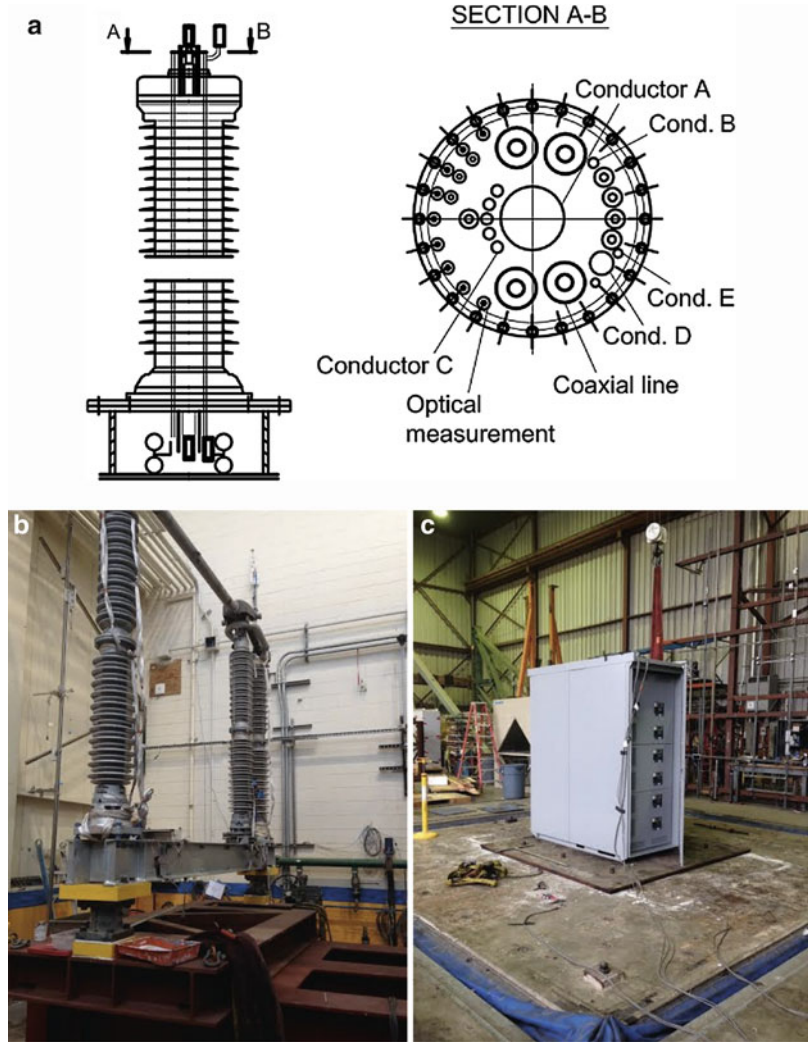
Usually the location of the attachment is not known in advance, so that the previous procedure can be applied considering p points obtaining p Response Spectra. For those points that are located at the same level (*floor*) of the P structure, a single spectrum can be evaluated (enveloping the Response Spectra), naming it Floor Response Spectrum.

If the goal is the evaluation of conservative FRSs, for each floor, a set of points has to be opportunely selected so that both translational and rotational effects are captured: they usually include the floor centroids and one or more corners for each floor.

It is worth mentioning that having defined the P system structural model, it is possible to

Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis,

Fig. 6 (a) Morphology of a bushing. (b) Bushing experimental test carried out at UC Berkeley (CA). (c) Cabinet experimental test carried out at UC Berkeley (CA)



evaluate a transferring function H_p that (1) knowing the Fourier Transform (F_i , input FT) of a given accelerogram (2) allows the definition of the Fourier Transform of the TH acceleration at given point (F_o , output FT) so that (3) the inverse Fourier Transform of F_o gives the TH at the considered point that is the required information for the Response Spectrum evaluation.

Alternatively, if the Power Spectrum G_i of a given earthquake or of a family of earthquakes is known as well as the previous defined Input Transferring Function (F_i), the output Power Spectrum (G_o) is evaluable according to Eq. 1.

$$G_o(w) = G_i(w) |H_p(w)|^2 \quad (1)$$

Knowing $G_o(w)$ and the transferring function (H_{SDOF}) of a single-degree-of-freedom system (SDOF), the SDOF Spectra Power Density is evaluable (see Eq. 2) and the related Response Spectrum is the required FRS.

$$G_{SDOF}(w) = G_o(w) |H_{SDOF}(w)|^2 \quad (2)$$

The previously presented approaches are not usually adopted for conventional structures as they are time consuming, so that predictive expressions are proposed in literature or enforced in

international recommendations: given the peak ground acceleration, the floor peak acceleration is evaluated by multiplying the PGA by an analytical function, named S_a in the following.

The usually proposed functions (S_a) are based on (1) simplified expressions for the evaluation of the floor acceleration and (2) simplified shape functions representative of the required FRS.

It is worth mentioning that given a structure, knowing of it (1) the prevalent modal shape (Φ), (2) the prevalent period (T_s), and (3) the modal participation factor (Γ) if a seismic action is considered, the absolute structural acceleration ($\ddot{u}_i$) associated to the single modal coordinate (Φ_i) can be evaluated according to Eq. 3 where $R_S(T_s)$ is the value of the normalized acceleration spectrum, for a given value of the structural damping (ζ_s) at the prevalent period of the structure.

$$\ddot{u}_i = \text{PGA} \cdot \Gamma \cdot \{1 + [R_S(T_s) - 1] \cdot \Phi_i\} \quad (3)$$

Assuming a given analytical function (R_{FRS}), being it dependent on the period (T_{NS}) and the damping (ζ_{NS}) of the nonstructural element, the required FRS, associable to the Φ_i modal coordinate, is equal to

$$\text{FRS}_i = \text{PGA} \cdot \Gamma \cdot \{1 + [R_S(T_s) - 1] \cdot \Phi_i\} \cdot R_{FRS} \cdot (T_{NS}) \quad (4)$$

Usually, Eq. 4 is simplified adopting (1) a constant value for $R_S(T_s)$ evaluated at the plateau of the acceleration spectrum (assumed in the range

of 2–3), (2) a simplified expression for the evaluation of modal displacement Φ_i of a given floor supposed to be equal to z/H where z is the level of the considered floor and H is the total construction height, and (3) a value of Γ between 1 and 1.5.

Based on the previous assumptions, Eq. 4 can be simplified as follows, having assumed $R_S = 3$, $\Gamma = 1$:

$$\text{FRS}_i = \text{PGA} \cdot \Gamma \cdot [1 + 2 \cdot (z/H)] \cdot R_{FRS} \cdot (T_{NS}) \quad (5)$$

In the following sections, some literature expressions (CEN 2008; FEMA 369 2001; AFPS 2007; KTA 2012) will be given, expressing them in terms of the normalized FRS (S_a) that corresponds to FRS_i evaluated for $\text{PGA} = 1$.

It has to be specified that in order to show the trend of the S_a functions, they will be plotted, contextualizing it to the simple structure described in the Case Study section.

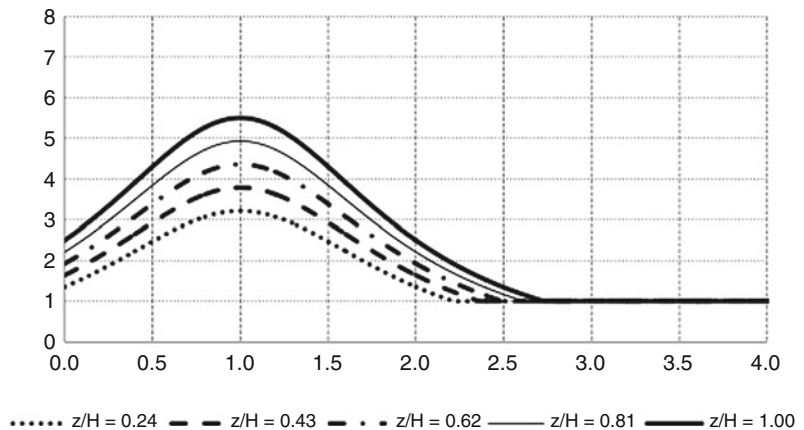
CEN (2008)

The following Eq. 6, plotted in Fig. 7, is proposed, supposing $\zeta_s = \zeta_{NS} = 5\%$.

$$S_a = 3 \cdot \frac{1 + \frac{z}{H}}{1 + \left(1 - \frac{T_{NS}}{T_s}\right)^2} - 0.5 \geq 1 \quad (6)$$

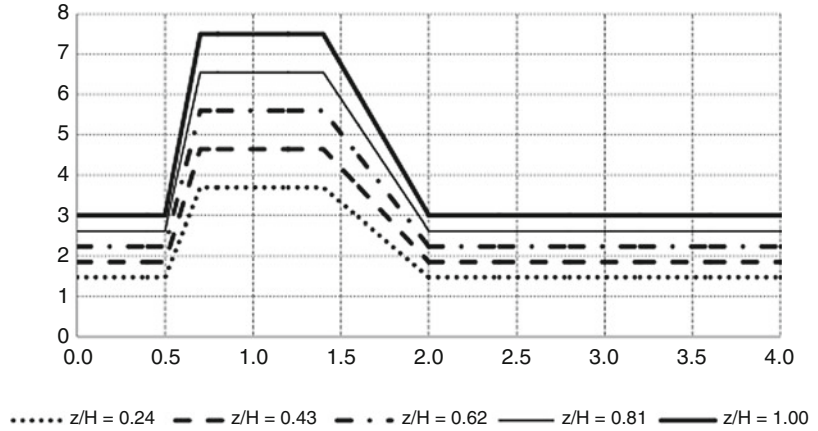
Equation 6 can be evaluated assuming $T_{NS} = 0$, obtaining (1) the value of the expression

Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis,
Fig. 7 Amplification factor S_a for different values of z/H



Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis,

Fig. 8 Amplification factor (S_a) shape versus T_{NS}/T_N



adopted for the evaluation of the normalized floor acceleration (Eq. 7a) and (2) the value of the adopted expression for the evaluation of R_{FRS} (Eq. 7b). So that Eq. 6 can be rearranged as reported in Eq. 7c.

It is possible to recognize that (1) a value of 2.5 has been assumed for the evaluation of R_s and (2) the expression adopted for the evaluation of R_{FRS} is supposed to be dependent on the normalized floor height.

$$\frac{\ddot{u}_i}{PGA} = \left\{ 1 + [R_s(T_s) - 1] \cdot \frac{z}{H} \right\} = 1 + 1.5 \cdot \frac{z}{H} \quad (7a)$$

$$R_{FRS} = \frac{1 + \frac{z}{H}}{1 + 1.5 \cdot \frac{z}{H}} \cdot \frac{3}{1 + \left(1 - \frac{T_{NS}}{T_s}\right)^2} - \frac{0.5}{1 + 1.5 \cdot \frac{z}{H}} \quad (7b)$$

$$S_a = \left(1 + 1.5 \cdot \frac{z}{H} \right) \cdot \left\{ \frac{1 + \frac{z}{H}}{1 + 1.5 \cdot \frac{z}{H}} \cdot \frac{3}{1 + \left(1 - \frac{T_{NS}}{T_s}\right)^2} - \frac{0.5}{1 + 1.5 \cdot \frac{z}{H}} \right\} \geq 1 \quad (7c)$$

FEMA 369 (2001)

The following Eq. 8, plotted in Fig. 8, is proposed supposing $\zeta_s = \zeta_{NS} = 5\%$, where the values of R_{FRS} are reported in Table 3:

$$S_a = \left(1 + 2 \cdot \frac{z}{H} \right) \cdot R_{FRS} \quad (8)$$

Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Table 3 Values of R_{FRS} as function of T_{NS}/T_N

T_{NS}/T_N	R_{FRS}
$T_{NS}/T_N < 0.5$ and $T_{NS}/T_N > 2.0$	1.0
$0.7 \leq T_{NS}/T_N < 1.4$	2.5
$0.5 \leq T_{NS}/T_N < 0.7$	$(7.5 \times T_{NS}/T_N) - 2.75$
$1.4 \leq T_{NS}/T_N < 2.0$	$6 - (2.5 \times T_{NS}/T_N)$

It is possible to recognize that (1) a value of 3.0 has been assumed for the evaluation of R_s and (2) the expression adopted for the evaluation of R_{FRS} is supposed to be independent on the normalized floor height.

AFPS (2007)

The following expression is proposed:

$$S_a = \sqrt{1 + \Gamma_s^2 \cdot R_s^2} \cdot \left(\frac{z}{H} \right) 2\alpha \cdot R_{FRS} \quad (9)$$

where (1) α is a parameter to be calibrated in order to minimize the difference between the effective modal displacements (Φ_i) and the proposed simplified expression (z/H), (2) R_s is the value of the normalized structural acceleration evaluated at the fundamental structural period (T_s) for a considered value of the structural damping, (3) the participation factor (Γ_s), evaluable according to Eq. 10a, assumes a

maximum value of 1.6, if $\alpha = 1.5$ is imposed, and (4) R_{FRS} values are reported in Table 3 as a function of parameter A (see Eq. 10b) that takes into account the damping (ζ_{NS}) of the nonstructural element.

$$\Gamma_S = \frac{2\alpha + 1}{\alpha + 1} \tag{10a}$$

$$A = \frac{35}{2 + \zeta_{NS}} \tag{10b}$$

It is worth mentioning that Eq. 9 derives from Eq. 4, with the following assumptions: (1) a unitary participation factor is considered for the ground acceleration, while the principal mode participation factor is considered according to Eq. 10a. (2) The spectral acceleration $R_s(T_s)$ is considered for the evaluation of the floor relative acceleration instead of the spectral relative normalized acceleration ($R_s(T_s) - 1$). (3) The ground acceleration and the relative structural acceleration are combined through the SRSS (Square Root of the Sum of the Squares) combination rule (Table 4 and Fig. 9).

KTA (2012)

The proposed expression does not give any information to evaluate the floor acceleration (a_g), but it only defines the amplification shape (R_{FRS}) reported in Fig. 10 (left), where the f -axis is the component frequency axis (in logarithmic scale) and f_1 , f_n , and f_{limit} , respectively, are (1) lowest decisive eigenfrequency of the principal system

at the lower limit value in the variation range of the system parameters, however, not lower than the rightmost corner frequency of the highest plateau of the associated response spectrum; (2) highest decisive eigenfrequency of the principal system for the upper limit value in the variation range of the component parameters, however, not lower than the rightmost corner frequency of the highest plateau of the associated response spectrum; and (3) upper limit frequency of the associated response spectrum.

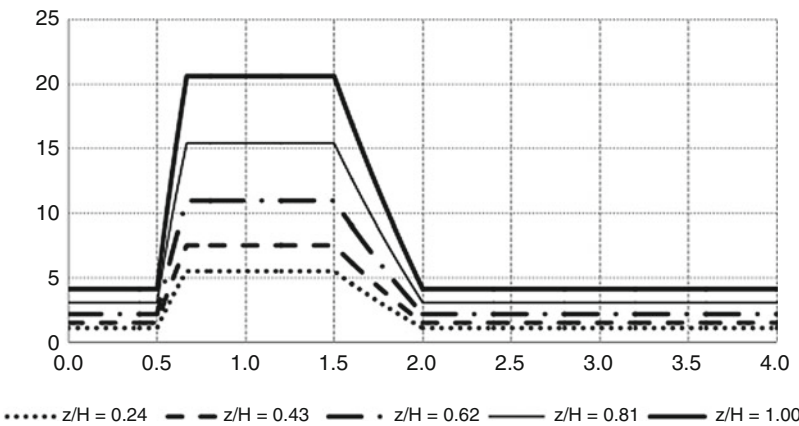
The maximum value of the amplification factor is reported in Fig. 10 (right), where D_1 and D_2 are respectively the damping ratios (in percentage of critical damping) of structural and nonstructural elements whose suggested values are reported in Table 5.

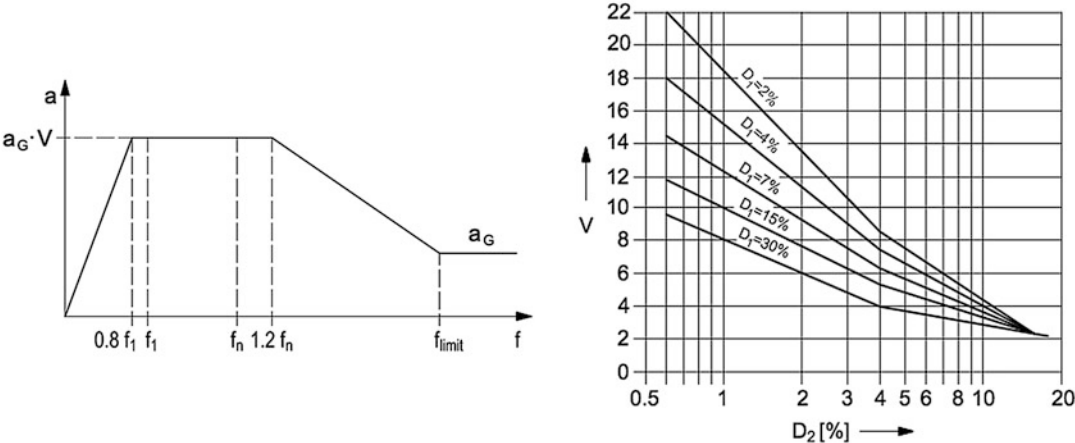
In order to compare the obtainable FRS with those previously discussed, the amplification factor (S_a) is plotted in Fig. 11, having assumed (1) a 5 % damping for both structural and nonstructural

Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Table 4 Values of R_{FRS} as function of T_{NS}/T_N

T_{NS}/T_N	R_{FRS}
$T_{NS}/T_N < 0.5$ and $T_{NS}/T_N > 2.0$	1.0
$2/3 \leq T_{NS}/T_N < 3/2$	A
$0.5 \leq T_{NS}/T_N < 1.5$	$A - (A - 1) \times \frac{(\log_3^3 T_{NS})}{(\log_4^3)}$
$2 < T_{NS}/T_N$	$K_T = A - (A - 1) \times \frac{(\log_3^2 T_{NS})}{(\log_5^2)}$

Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Fig. 9 Amplification factor (S_a) shape versus T_{NS}/T_N ($\alpha = 1.5$, $\Gamma_S = 1.6$, $R_S = 2.5$)





Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Fig. 10 Amplification factor shape (*left*) and maximum amplification factor (*right*)

Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Table 5 Suggested damping values (in percent of critical damping). Column A: to be adopted for verifying the load-carrying capacity and integrity and for determining the spectra. Column B: in the case of mechanically active components for which the functional capability is verified by a deformation analysis

Components	Damping ratios	
	A	B
Pipes	4	2
Steel with welded connections and welded components (e.g., vessels, valves, pumps, motors, ventilators) ^a	4	2
Steel with SL or SLP bolt connections (<i>SL</i> – structural bolt connection with a borehole tolerance ≤ 2 mm; <i>SLP</i> – fitted bolt connection with a borehole tolerance ≤ 0.3 mm)	7	4
Steel with SLV(P) or GV(P) bolt connections (<i>SLV(P)</i> – preloaded fitted bolt connection; <i>GV(P)</i> – fitted friction-grip bolt connection)	4	2
Cable support structures	10 ^b	7
Fluid media	0.5	0.5

^aIf, on account of the design, deformations are possible only in small regions of the structure (low structural damping), the values as listed shall be halved (special cases)

^bIn well-substantiated cases, the damping ratio may be increased up to 15 %

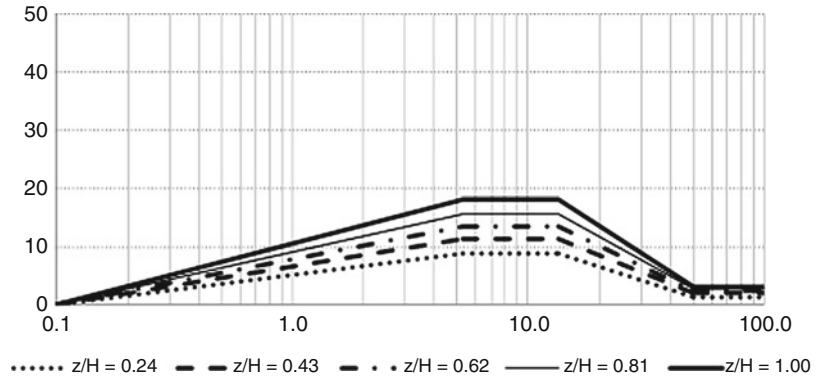
elements, (2) the FEMA expression (see Eq. 11) for the evaluation of the normalized floor acceleration, (3) $f_1 = 6.66$ Hz that is the rightmost corner frequency of the acceleration plateau of CEN (2008) type 1 Spectrum (A soil), and (4) $f_n = 11.1$ Hz that is the highest decisive eigenfrequency of the principal system described in the Case Study section.

$$a_g = 1 + 2 \cdot \frac{z}{H} \tag{11}$$

Verification

As stated in KTA (2012), the verification process has to regard (1) the load-carrying capacity in terms of strength, stability, and secure positioning (e.g., their protection against falling over, dropping down, impermissible slipping); (2) the integrity, that is, the capability of a component above and beyond its load-carrying capacity to meet the respective requirements regarding leak tightness and deformation restrictions; and (3) the functional capability, that is, the capacity of a system or component above and beyond its load carrying capacity to fulfill the designated tasks by way of its respective mechanical or electrical function.

Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis,
Fig. 11 Normalized floor acceleration based on Eq. 11



Depending on the importance of the element to verify and the material (conventional or nonconventional material), the verification process could include experimental tests either for the validation of the numerical model or for the qualification of the element itself. The verification procedure can include a) analysis, b) physical experiments, and c) analogies and plausibility considerations.

Based on the introduced classification that distinguishes secondary element in structural and nonstructural, the following criteria can be defined:

1. Secondary structural elements have to be verified with regard to the vertical loads transmitted from the P structure, opportunely combined with the other action considered to be contemporary to the seismic action. The connections have to be verified with regard to the seismic-induced action, including second-order effects such as those induced by the axial load in the deformed configuration (P- Δ effects).
2. Secondary nonstructural elements have to be verified with regard to the self-weight loads opportunely combined with the other actions considered to be contemporary to the seismic action.
3. For both types of elements (S and NS elements), the action supposed to act contemporary with the seismic action can be consequent to different events such as those pertaining to collisions, explosions, and fires.

The verifications have to consider potential damage induced to other elements, which loss of capacity could induce either human or economic loss.

The verification is performed checking that the element capacity will be greater than the demand, defined in terms of different mechanical properties (stresses, forces, displacement) depending on the adopted materials.

In order to define the design forces, the considered floor response acceleration spectra can be reduced to take into account the nonstructural element ductility. If the FRSs have been numerically evaluated, they have to be modified (see Fig. 12) to take into account the structural stiffness uncertainties: (1) an adequate plateau has to be imposed in correspondence of the structural period, (2) the linear envelope has to be properly introduced, and (3) the ductility of the nonstructural element can be considered, properly reducing the FRS (see Fig. 12b, c).

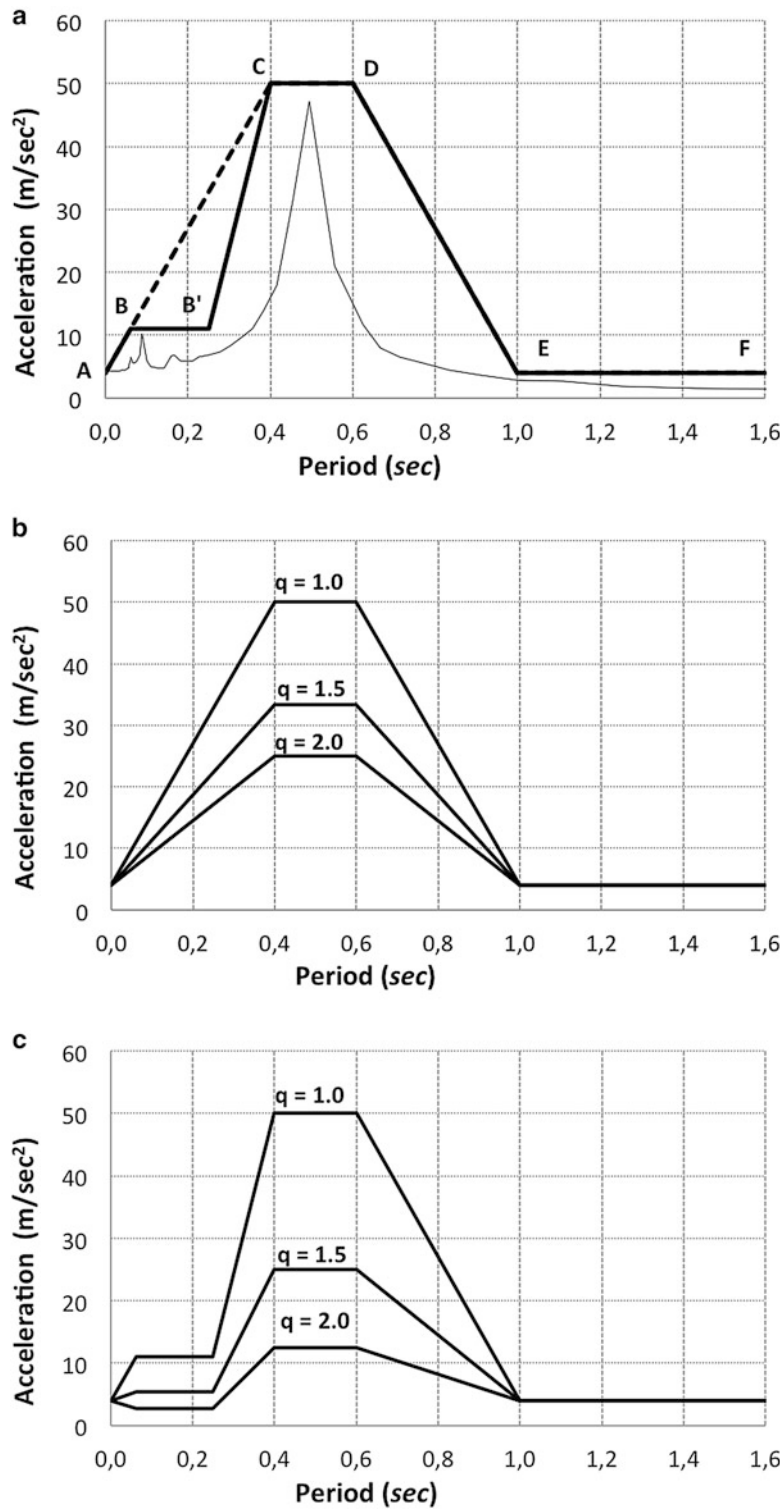
Case Study

The previously described procedures to determine FRSs will be applied to a steel frame system hosting a set of equipments, whose characteristics and localization are reported in Table 6 and Fig. 13 (left), reproduced from KTA (2012).

The maximum acceleration of each equipment can be evaluated by means of (1) time histories considering the interaction between the principal structure and the equipment or (2) FRSs

Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis,

Fig. 12 (a) Design spectra definition. Shape modification: alternative solution. (b) Design spectra definition. Design spectra based on ABCDEF spectrum. (c) Design spectra definition. Design spectra based on ABB'CDEF spectrum



Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Table 6 Equipment mass (*Ton*) and period (*sec*)

Equipment		Mass	Floor	Period
1		20	1	0.051
2		10	1	0.093
3,4,5		10	2	0.070
6		10	3	0.034
7		20	3	0.060
8		30	4	0.033
9,10		15	5	0.036

evaluated based on the previously described *cas-cade* procedure or on the predictive expressions already presented.

Time History-Based Evaluation of Equipment Accelerations

A detailed model of the PS system could include the secondary system modeled as reported in Fig. 13 (right): PS principal modal shapes are those reported in Fig. 14a. Alternatively equipments can be modeled by means of mass lumped at pertinent position of the floor as reported in Fig. 13 (center): PS principal modal shapes are those reported in Fig. 14b.

Both models have been analyzed by means of time histories, carried out (1) generating 7 accelerograms compatible with the acceleration spectra ($PGA = 0.15\text{ g}$) suggested in CEN (2008) for B soil and low-magnitude events ($M < 5.5$) (see Fig. 15), (2) considering a constant damping value of 2 % for the structural model and for the

equipments, (3) performing a dynamic modal TH analysis evaluating, for each accelerogram, the maximum absolute value of a given quantity (acceleration), and (4) averaging the maximum values obtained (at step 3) for each analysis.

If the detailed model with interaction (WI) is considered, the evaluated quantities are the mass accelerations of the single equipment.

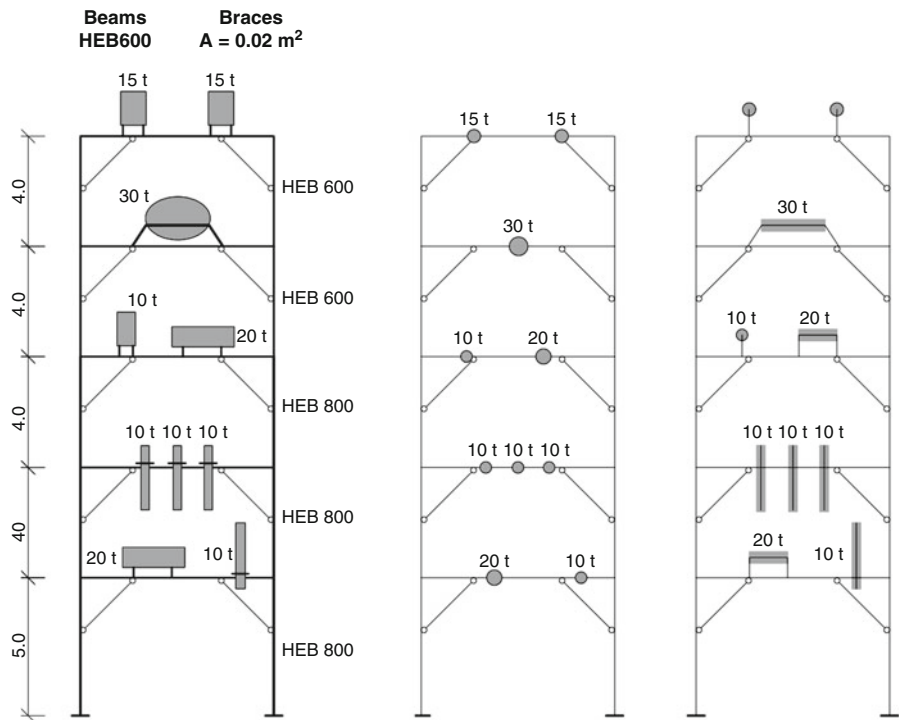
Regarding to the lumped mass system, (1) for each considered accelerogram, an FRS has been generated (Fig. 16a): the FRS is relative to the acceleration of the top left floor corner (no sensible variations in FRSs have been observed if other floor points are considered); (2) having generated (for each floor) seven FRSs, the averaged FRS has been evaluated (Fig. 16b); and (3) for each equipment, depending on the floor it is attached to and its period (T_{NS}), the acceleration has been evaluated through the resulting FRS.

The results of the performed evaluation are reported in Table 7 where the acceleration of each floor and the acceleration of each equipment are reported, calculated with or without interaction.

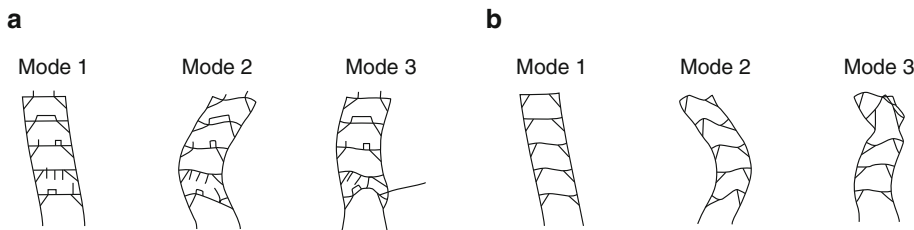
Comparing the maximum acceleration evaluated considering the PS system with those evaluated through the cascade procedure, a significant difference (if the Equipment 2 is concerned) between the two approaches can be noticed: the difference is aspecktable since the equipment period (0.093 s) is close to the period of the third modal shape (0.090 s) so that (see Fig. 16c) in a very small period range (in between 0.09 and 0.1 s), the acceleration ranges between 7.0 and 4.5 m/s – the already-mentioned *tuning* effect causes the equipment acceleration reduction if the complete PS system is analyzed in order to include the P-S interaction.

Analytical FRS-Based Evaluation of Equipment Accelerations

It has been already outlined that CEN proposal (CEN 2008) and FEM proposal (FEMA 369 2001) are based on a fixed value (5 %) of structural and equipment damping. So that only the proposal reported in AFPS (2007) and KTA (2012) will be considered in the following.

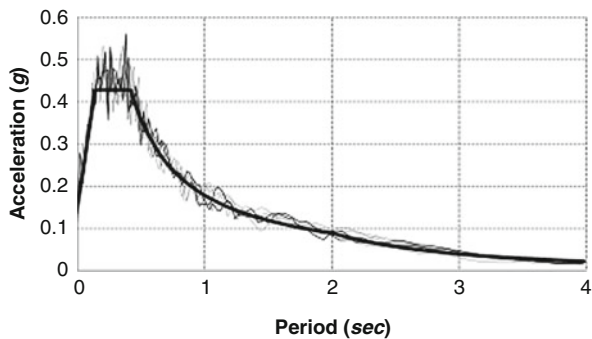


Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Fig. 13 Case study: geometry (left) and models (center, right)



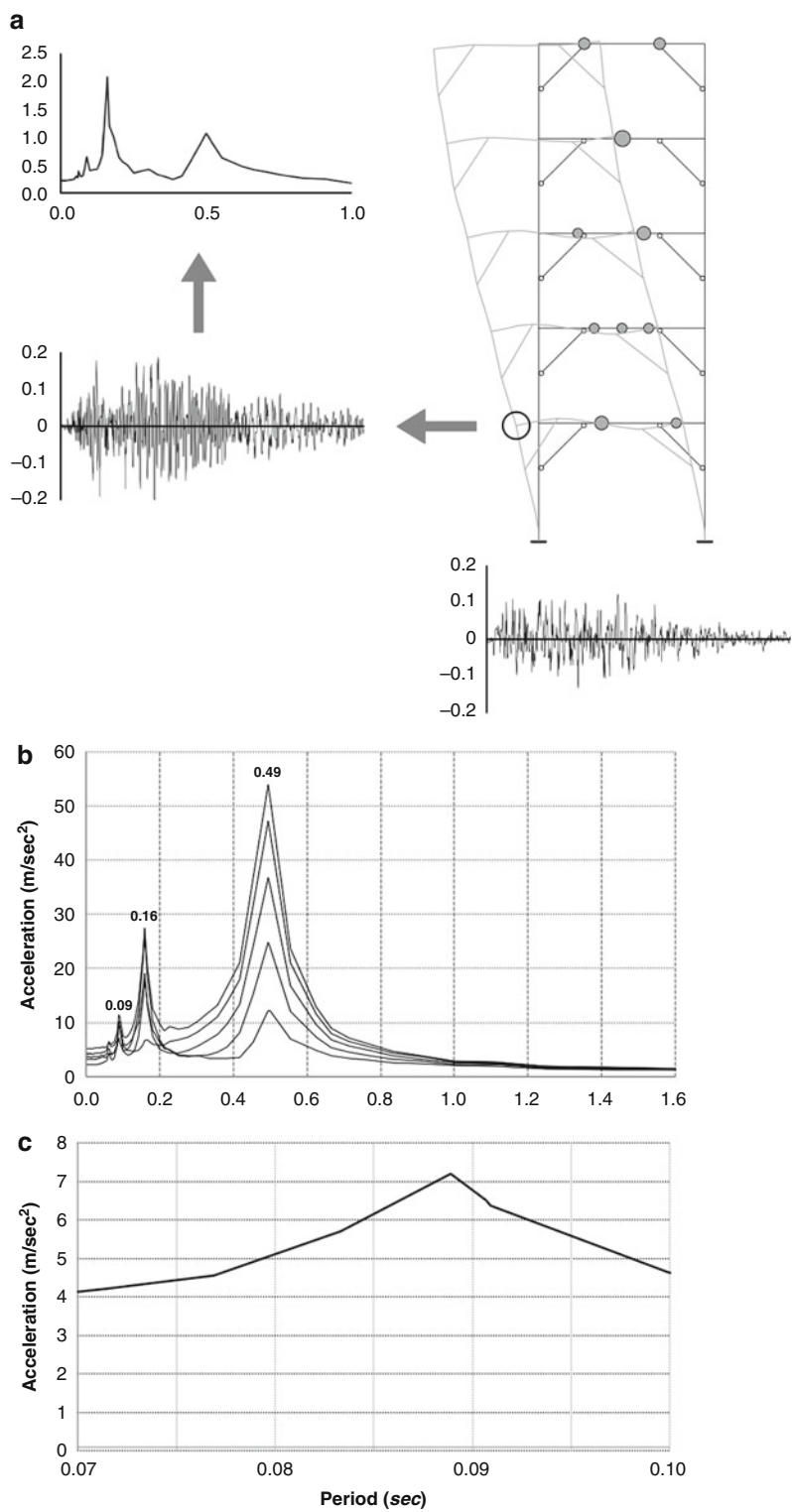
Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Fig. 14 (a) Modal shape: direct modeling of the equipment. Periods (sec): 0.48, 0.158, 0.121. (b) Modal shape: equipment modeled as lumped masses. Periods (sec): 0.49, 0.159, 0.09

Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Fig. 15 Target acceleration spectrum (PGA = 0.15 g) and spectra of the generated accelerograms



Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis,

Fig. 16 (a) Floor Response Spectra generations: scheme. (b) Generated Floor Response Spectra. (c) Generated Floor Response Spectra: 1° Floor; Period range 0.07–0.1 s



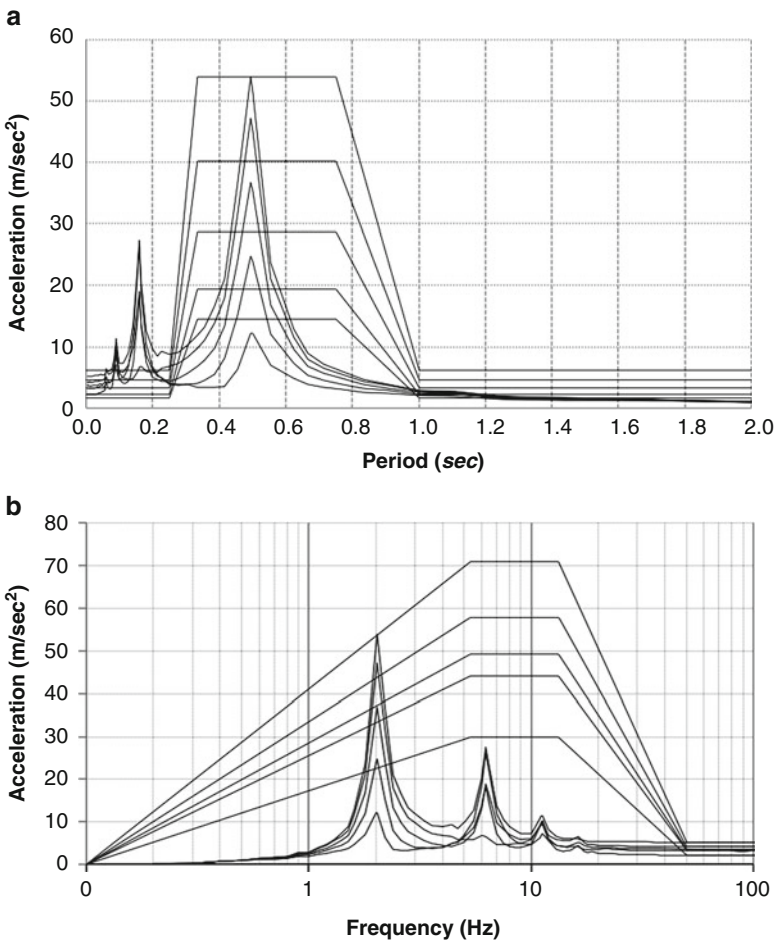
Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Table 7 Floor acceleration (FA, m/sec^2) and equipment acceleration evaluated without (F_w/o) and with (F_w) interaction

N_E	N_F	Floor acceleration	FRS W/o	FRS W
1	1°	2.21	2.7	2.1
2	1°	2.21	6.5	3.2
3	2°	3.26	4.20	4.0
4	2°	3.26	4.20	4.0
5	2°	3.26	4.20	4.0
6	3°	3.65	3.70	3.8
7	3°	3.65	5.10	3.8
8	4°	4.28	4.30	4.3
9	5°	5.25	5.3	5.3
10	5°	5.25	5.3	3.3

It has to be specified that (1) concerning the KTA (2012) proposal, the floor acceleration evaluated by means of TH analyses (see Table 7) has been considered and (2) concerning the AFPS (2007) proposal, the FRSs have been evaluated by means of Eq. 9, evaluating the spectral acceleration corresponding to the first structural period (0.49 s) and a participation factor (Γ_S) equal to 1.6 that correspond to $\alpha = 1.5$ (see Eq. 10a).

It is clear (see Fig. 17a, b) that considering equipment periods close to the lowest structural periods, the KTA proposed expressions are more conservative while the AFPS expressions are less conservative (Tables 8 and 9).

Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Fig. 17 (a) Generated FRSs. Acceleration versus period: (1) numerical simulation and (2) AFPS predictive equations evaluated based on Eq. 9 ($\alpha = 1.5$, $\Gamma_S = 1.6$, $R_S = 2.4$). (b) Generated FRSs. Acceleration versus frequency: numerical simulation and KTA predictive equations



Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Table 8 Normalized floor displacement

Floor	MD	z/H
1°	0.21	0.24
2°	0.45	0.42
3°	0.68	0.62
4°	0.87	0.81
5°	1.00	1.00

Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis, Table 9 Equipment (Eq) acceleration

N _E	N _F	AFPS	KTA
1	1°	2.2	22.0
2	1°	2.2	34.0
3,4,5	2°	3.3	42.0
6	3°	3.7	22.0
7	3°	3.7	42.0
8	4°	4.3	25.0
9,10	5°	5.3	35.0

Summary

The chapter deals with the methodologies focused on the seismic analyses of the so-called *secondary* (sometimes *attachments*) elements that are part of a construction whose seismic resistance is delegated to a *primary* resistant structure.

Although secondary elements can be decontextualized from the primary resistant structures, they will be subjected to seismic action as well and, having their own structures, need to be modeled and analyzed by means of methods included in the general methodologies proper of seismic branch.

Among the methodologies usually adopted for secondary element analyses, the Floor Response Spectra (FRS)-based analyses become popular due to their recognized simplicity.

FRSs provide acceleration (consequently velocity and displacement) to which the

secondary element (with a given period and damping) will be subjected to when *attached* (from which the alternative name *attachments* derives) to a given part of the structures such as a building floor (from which the name Floor Response Spectra derives).

Given that FRS generation could require onerous numerical analyses, simplified expressions are proposed in literature and discussed in the following together with the general methodologies tailored to secondary element modeling and seismic analyses.

Cross-References

- Behavior Factor and Ductility
- Building Codes and Standards
- Classically and Nonclassically Damped Multi-degree of Freedom (MDOF) Structural Systems, Dynamic Response Characterization of
- Code-Based Design: Seismic Isolation of Buildings
- Earthquake Risk Mitigation of Lifelines and Critical Facilities
- Equivalent Static Analysis of Structures Subjected to Seismic Actions
- European Structural Design Codes: Seismic Actions
- Modal Analysis
- Nonlinear Dynamic Seismic Analysis
- Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers
- Response Spectrum Analysis of Structures Subjected to Seismic Actions
- Seismic Fragility Analysis
- Soil-Structure Interaction
- Spatial Variability of Ground Motion: Seismic Analysis
- Tuned Mass Dampers for Passive Control of Structures Under Earthquake Excitations

References

- AFPS (2007) Guide methodologique pour la conception, l'installation et le diagnostic des equipment en zone sismique. Association Francaise de Genie Parasismique, Paris
- ASCE 7-02 (2002) Minimum design loads for buildings and other structures. American Society of Civil Engineers, Reston
- ATC 58 (2012) Seismic performance assessment of buildings. Applied Technology Council, Redwood City
- ATC/BSSC (1997) NEHRP commentary on the guidelines for the seismic rehabilitation of buildings. Report no FEMA 274, prepared by the Applied Technology Council for the Building Seismic Safety Council, published by the Federal Emergency Management Agency, Washington, DC
- CEN (2008) Eurocode 8: design of structures for earthquake resistance, part 1: general rules, seismic actions and rules for buildings. Comité Européen de Normalisation, Brussels
- Chaudhri SR, Villaverde R (2008) Effect of building nonlinearity on seismic response of nonstructural components: a parametric study. *J Struct Eng* 134(4):661–670
- Chen Y, Soong TT (1988) State-of-the-art review seismic response of secondary systems. *Eng Struct* 10:218–228
- FEMA 369 (2001) NEHRP recommended provisions for seismic regulations for new buildings and other structures: part 2-commentary. Building Seismic Safety Council, National Institute of Building Sciences, Washington, DC
- FEMA 74 (2005) Earthquake hazard mitigation for nonstructural elements. Field Manual, Emergency Management Agency, Washington, DC
- Gilani AS, Whittaker AS, Fenves GL, Fujisaki E (1999) Seismic evaluation of 55° kV porcelain transformer. Pacific Earthquake Engineering Research Center, Berkeley
- (IEEE) (2005) Recommended practice for seismic design of substations, Institute of Electrical and Electronics Engineers, New York
- Igusa T, Der Kiureghian A (1985a) Dynamic characterization of two degree-of-freedom equipment-structure systems. *J Eng Mech ASCE* 111(1):1–19
- Igusa T, Der Kiureghian A (1985b) Dynamic response of multiply supported secondary systems. *J Eng Mech ASCE* 111(1):20–41
- Kircher CA (2003) It makes dollars and sense to improve nonstructural system performance. In: Proceedings of ATC 29–2 seminar on seismic d, performance, and retrofit of nonstructural components in critical facilities, Newport Beach, 23–24 Oct 2003
- KTA (2012) Safety standards of the nuclear safety standards commission: design of nuclear power plants against seismic events; Part 4: components. KTA-Geschaefsstelle c/o BfS, Willy-Brandt-Str. 5, 38226 Salzgitter
- Taghavi S, Miranda E (2003) Response assessment of nonstructural building elements. Pacific Earthquake Engineering Research Center, PEER 2003/05 Sept
- Villaverde R (1997) Seismic design of secondary structures: state of the art. *J Struct Eng* 123(8):1011–1019

Seismic Accelerometers

Gerardo Alguacil¹ and Jens Havskov²

¹Instituto Andaluz de Geofísica, University of Granada, Granada, Spain

²Department of Earth Science, University of Bergen, Bergen, Norway

Synonyms

Broadband seismometers; Force-balance accelerometers; MEMS accelerometers; Passive seismometers; Seismic recorders; Servo-accelerometers; Strong ground motion

Introduction

Earthquake ground motion ranges in amplitude from a few nanometers (e.g., for a distant earthquake or a local microearthquake) to several meters close to the fault causing a big quake. It is very difficult to record that wide range of signals with a single type of instrument. Seismometers are designed to be very sensitive for detecting weak signals of ground motion and have a response proportional to ground velocity in a frequency band typically from 0.01 to 100 Hz. On the other hand, strong ground motion instruments are in general less sensitive and may manage on scale large ground motion amplitudes. This strong ground motion may reach peak accelerations above 2 g (g is gravity acceleration) and peak velocities higher than 3 m/s (Anderson 2010). The motion experienced at specific points in a building or structure may be even higher. Seismic accelerometers were developed for recording on scale vibrations up to such range of amplitude either on ground or in structures.

Due to instrumental and practical reasons, the preferred motion measure for strong ground motion is acceleration, since the inertial force acting on a structure due to a seismic action is proportional to ground acceleration. Nevertheless, some strong-motion velocimeters are commercially available, but their use is much less widespread.

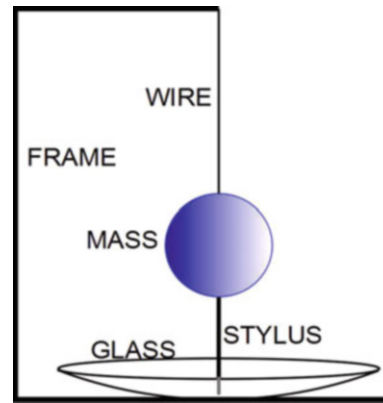
Weak motion from local small earthquakes or moderate to strong earthquakes at regional or large distances are studied fundamentally on seismograms, proportional to ground velocity in a wide frequency band. However, even for strong motion, ground velocity and displacement are also of interest: Ground velocity is directly related to important parameters such as shake energy or soil liquefaction potential and is linked to damage on intermediate period structures and buried pipelines (e.g., Akkar and Bommer 2007). Ground displacement is significant for large structures, where differential displacements may cause damage, and from a seismological view, it is proportional to seismic moment, a source parameter defining the earthquake size. Therefore, it is desirable that velocity and displacement may be estimated from acceleration records.

Historical Overview

This section partly follows the review article by Trifunac (2009). See also “► [Historical Seismometer](#).”

Earlier attempts to record strong ground motion were made with the so-called seismoscopes (Fig. 1), most of which were limited to record the horizontal ground motion on a (curved) surface without a time reference. These instruments consisted of a pendulum, either normal or inverted, whose mass motion was drawn on a curved surface (usually a smoked glass) by a stylus.

As a result of the program of strong-motion earthquake investigation of the US Coast and Geodetic Survey and cooperative institutions (McComb and Ruge 1937), the first instrument designed to record the strong ground motion versus time was developed in the decade of 1930.



Seismic Accelerometers, Fig. 1 Simplified schematics of a seismoscope. A mass is suspended from the frame by a wire, and a stylus writes the relative motion on a spherical cap surface

This instrument was built with a mass suspended on torsion wires, based on the Wood-Anderson seismometer principle (known because it was used by Richter to define his magnitude scale). Damping was achieved – like for the Wood-Anderson seismometer – using a conductive copper mass in which the field of a permanent magnet induced parasitic currents. A more robust pivoted suspension and a spring were later incorporated. Recording was made on photographic paper on a drum, with a “starter” system built with an independent undamped pendulum closing an electric contact when a strong motion occurred. A clock mechanism interrupted the light beam to produce time marks on the record. The first strong ground motion accelerograms were obtained on three stations in 1933 from the Long Beach earthquake (McComb and Ruge 1937).

With analog records and the computational means available at that epoch, it was very difficult to estimate displacement from acceleration. Therefore, some seismometers with unity amplification and photographic recording were used to obtain displacement records (see “► [Passive Seismometers](#)”).

A remarkable effort was made to test these first instruments with a shaking table to assess the independent motion of the components, zero shifts (instability of the mass rest position),

parasitic vibrations (instrument vibration modes different from fundamental), the ability to calculate velocity and displacement from acceleration records, and other features (McComb and Ruge 1937, Ruge and McComb 1943).

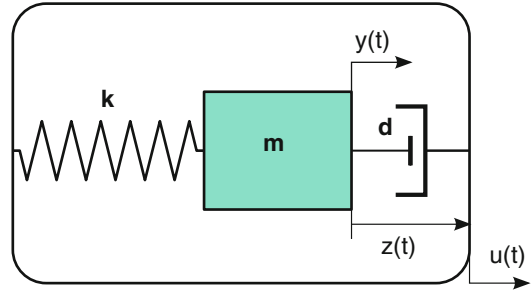
It was not until early 1960 that accelerographs become commercially available. Around that time the digital computation also made it possible to calculate the velocity and displacement histories from the corresponding accelerograms. In this decade, some seismic accelerometers were still of torsion wire or inverted pendulum types, but the first servo-accelerometers using the concept of force balance appeared (Reasenberg 1969; Eisenberg and McEvilly 1971). This concept (see later) is to apply a feedback force on the mass to keep it at rest relative to the ground. So the applied force is the mass times the ground acceleration. The main advantages are improved linearity and dynamic range (i.e., the relation between the maximal non-clipped signal amplitude and the minimal resolvable amplitude; see later). Recording was analog on photographic paper, but some prototypes used digital recording (Reasenberg 1969). The first digitizers used, however, had a too low resolution to match the dynamic range of servo-accelerometers.

Following the period of photographic recording, some models adopted analog recording on magnetic tape using frequency modulation, but digital recording was soon the norm due to its better characteristics. Presently all high-performance seismic accelerometers use force-balance sensors and high-resolution digital recording.

Basic Principles of Accelerometers

Seismometers and accelerometers are both based on the inertia principle.

In Fig. 2, a schematic view of the well-known passive damped oscillator is drawn. A mass m is fixed to one end of a spring of elastic constant k . The other end of the spring is fixed to a frame anchored to the ground. The motion is mechanically limited to the horizontal axis shown. When the ground moves a displacement



Seismic Accelerometers, Fig. 2 A schematic single degree of freedom oscillator. A spring makes a force proportional and opposed to the relative displacement of the mass $z(t)$, and a dashpot damps this motion with a force proportional to the relative velocity. The ground (and the frame) displacement is $u(t)$

$u(t)$ with respect to some inertial frame, the mass displacement is $y(t)$ in this inertial frame, and $z(t) = y(t) - u(t)$ is the mass displacement relative to the instrument casing or the ground. A dashpot represents a damping device (usually magnetic) that acts on the mass with a force proportional to its relative velocity. This damping avoids the oscillation of the system with its free period when it is excited. The two real forces acting on the mass are the spring force and the damping force, so the dynamics equation is (Havskov and Alguacil 2010)

$$-k \cdot z - d \cdot \dot{z} = m \cdot y(t) = m[\ddot{u}(t) + \ddot{z}(t)] \quad (1)$$

It is useful to write the equation as a function of the two coefficients experimentally measurable. The free or natural angular frequency ω_0 can be defined as

$$\omega_0 = \sqrt{\frac{k}{m}} = 2\pi f_0 = \frac{2\pi}{T_0} \quad (2)$$

where T_0 is the suspension-free period and its inverse f_0 is the natural frequency. The damping coefficient or damping fraction h is

$$h = \frac{d}{2\omega_0 m} \quad (3)$$

Rearranging Eq. 1 then gives

$$\ddot{z}(t) + 2h\omega_0 \cdot \dot{z}(t) + \omega_0^2 \cdot z = -\ddot{u}(t) \quad (4)$$

Both ω_0 and h can be measured, e.g., from the system transient response or by exciting the system with steady-state harmonics of frequencies in a band including f_0 . For low frequencies, the mass velocity and acceleration terms in Eq. 4 are small, so mass displacement z is then proportional to the ground acceleration $\ddot{u}$.

The system response may be characterized in the Laplace transform domain. Let $Z(s)$ and $U(s)$ be the Laplace transforms of $z(t)$ and $u(t)$. By assuming null velocities and displacements at initial time, the Laplace transform of Eq. 4 becomes

$$s^2 Z(s) + 2h\omega_0 s Z(s) + \omega_0^2 Z(s) = -s^2 U(s) \quad (5)$$

The transfer function between the mass relative motion Z and the ground acceleration $s^2 U$ is then

$$T_a(s) \equiv \frac{Z(s)}{s^2 U(s)} = \frac{-1}{s^2 + 2h\omega_0 s + \omega_0^2} \quad (6)$$

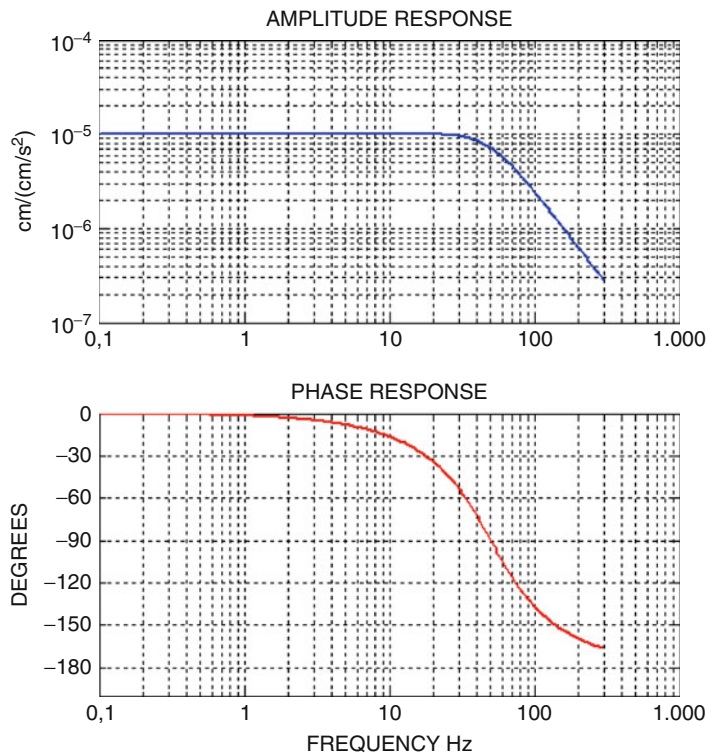
The explicit frequency response function is obtained by substitution of s by $i\omega$ in Eq. 6, that is,

$$H_a(\omega) = \frac{-1}{-\omega^2 + 2h\omega_0 \cdot i\omega + \omega_0^2} \quad (7)$$

For a sinusoidal motion of arbitrary frequency $f = \omega/2\pi$, the modulus of this complex function represents the amplitude relation between the mass relative motion $z(t)$ and the ground acceleration $\ddot{u}(t)$, and its phase is the relative phase between them. Both are plotted in Fig. 3 as functions of frequency. The amplitude response is flat for acceleration up to the corner frequency f_0 and decays at higher frequencies as f^{-2} . The frequencies of interest are usually in the flat zone, so the instrumental correction for amplitude is simply a constant factor. Observe, nevertheless, that the phase response deviates from flatness even for frequencies well under the natural frequency. This response function is formally like the response of a second-order low-pass filter with cutoff frequency $f_0 = \omega_0/2\pi$.

Seismic Accelerometers,

Fig. 3 *Top*: amplitude response of a mechanical accelerometer for ground acceleration. *Bottom*: phase response. The free period is $T_0 = 0.02$ s ($f_0 = 50$ Hz) and $h = 0.70$



Displacement Transducer

The mass motion has to be measured and recorded by some device. In early accelerometers, a light beam was reflected in a mirror which rotated with the mass motion. Presently, almost all seismic accelerometers use a capacitive transducer, which gives a voltage output proportional to the mass displacement. This type of transducer is very sensitive – it can resolve displacements of the order of pm (10^{-9} mm).

Two types of capacitive transducers may be used: variable gap or variable area. The most used arrangement, Fig. 4, is a pair of capacitors with a common central moving plate and two fixed plates or vice versa (variable gap). When the central plate moves with the mass relative to the fixed plates, one of the capacities increases and the other decreases. The same happens with the variable area type. An identical sinusoidal or square signal with frequency of several kHz is fed to each capacitor with opposite sign. The signal amplitude at the common point of both capacitors is proportional to the capacitance

difference, and the phase depends on the sign of this difference. This may be demodulated with a phase-sensitive demodulator (PSD) circuit. While this is the most used technique, other approaches are possible.

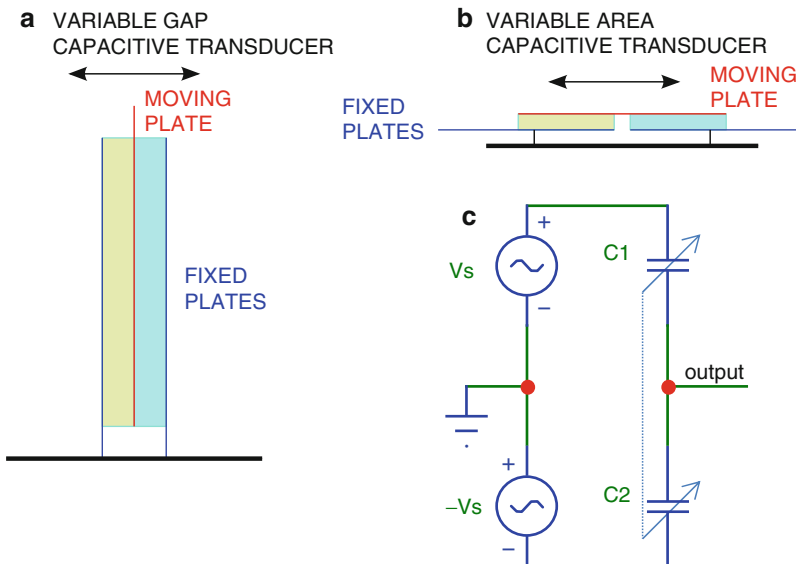
The output of the detector is proportional to the capacitance difference. Let A be the overlapping area between plates and d the distance (gap) between them. If d is much smaller than the plate dimensions, the capacitance of a parallel plane capacitor is

$$C = \frac{\varepsilon \cdot A}{d} \quad (8)$$

where ε is the dielectric permittivity (for air $\varepsilon = 8.85 \cdot 10^{-3}$ pF/mm). In the variable gap transducer, for a displacement x of the central plate, it may be easily shown that the capacitance difference is

$$C_1 - C_2 = \varepsilon \cdot A \cdot \frac{2x}{d^2 - x^2} \quad (9)$$

which is not linear with x .

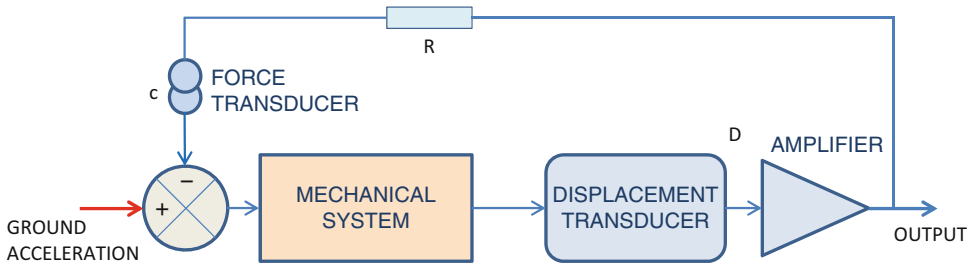
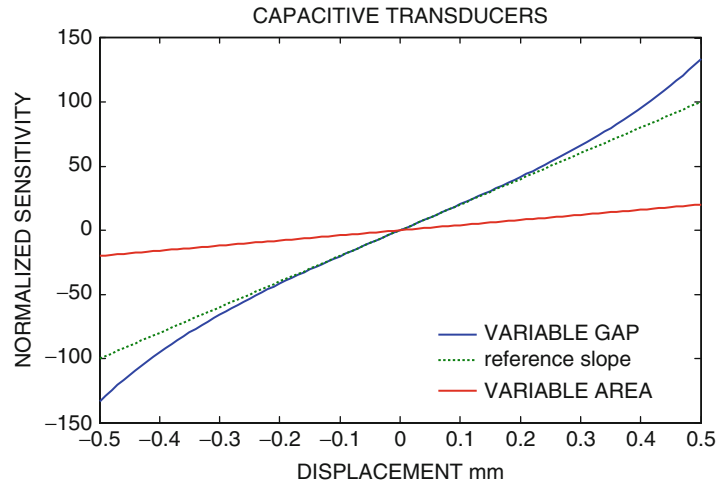


Seismic Accelerometers, Fig. 4 Schematics of variable capacitance displacement transducers. (a) Variable gap. As the central plate (red) moves right, the gap with the left fixed plate (blue) increases (capacitance of left capacitor, yellow, decreases), and the gap with the right plate (blue) decreases (capacitance of the right capacitor, light blue,

increases). (b) Variable area. The effective area of each capacitor (same colors as in a) varies in opposite sense when the central plate (red) moves horizontally. The gap is not at scale, since in practice it is very narrow in relation to plate dimensions. (c) Circuit schematics, see text

Seismic Accelerometers,

Fig. 5 Normalized sensitivity of capacitive transducers of variable gap (blue) and variable area (red). A straight line is plotted for reference (dotted) with the slope of the variable gap at the origin



Seismic Accelerometers, Fig. 6 Block schematic of an FBA

In a variable area transducer with plate overlap dimensions $a \times b$, the capacitance difference is linear with the displacement x :

$$C_1 - C_2 = \frac{\epsilon}{b} \cdot 2b \cdot x \quad (10)$$

In Fig. 5, these functions – excluding ϵ – are plotted for an overlapping area of $A = 100 \text{ mm}^2$ and gap distance of $d = 1 \text{ mm}$ and 0.5 mm for variable gap and variable area, respectively. Two features are clear: variable gap type is more sensitive but nonlinear, and variable area type is less sensitive but linear.

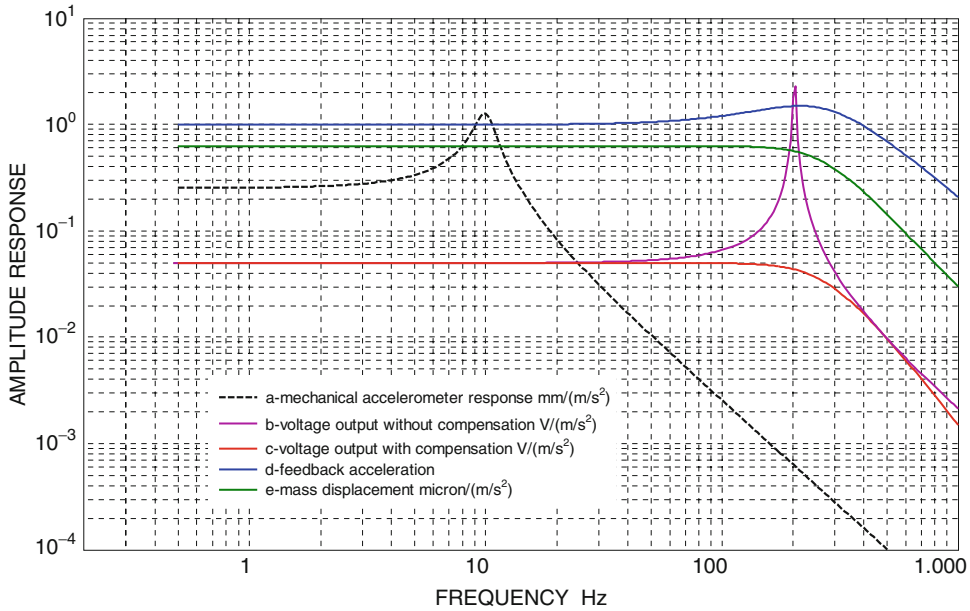
Variable gap transducers are often used in feedback systems, like force-balance accelerometers, since the mass displacement is held very small – within the linear zone – by the servo control.

Variable area transducers are used in some micro-electro-mechanical (MEM) accelerometers (see later) without feedback.

The Force-Balance Accelerometer (FBA)

Most present seismic accelerometers are of the force-balance type. The principle is to apply a feedback force on the inertial mass opposed to its motion, in such a way that this motion is reduced to a minimum. A simplified block schematic is shown in Fig. 6.

The ground acceleration produces an inertial force on the mass, like in the basic passive accelerometer of Fig. 1. The relative mass motion z is sensed with a displacement transducer, whose output is amplified, with a total factor D (volt/meter, V/m). The amplifier output v_o is applied



Seismic Accelerometers, Fig. 7 Several responses relative to ground acceleration. See text. Note the different units for each curve

through a resistor R to a force transducer (normally a coil-magnet system) with constant c (newton/ampere, N/A). The feedback force $-c \cdot v_o/R = -c \cdot A \cdot z/R$ is then applied to the mass. Equation 1 has now to be modified to

$$-k \cdot z - d \cdot \dot{z} - \frac{Dc}{R} \cdot z = m[\ddot{u}(t) + \ddot{z}(t)] \quad (11)$$

It is clear that the effect of feedback is to increase the effective stiffness of the system from k to $k + D \cdot c/R$. Arranging terms of Eq. 11 to get the form of Eq. 4, it is found that with feedback, a new effective “free” angular frequency is ω_f

$$\omega_f = \sqrt{\frac{k + D \cdot c/R}{m}} = \sqrt{\omega_0^2 + \frac{D \cdot c}{R \cdot m}} \quad (12)$$

which is higher than the “open-loop” free angular frequency ω_0 .

On the other hand, the new damping coefficient with feedback h_f is decreased to

$$h_f = h \frac{\omega_0}{\omega_f} \quad (13)$$

This is not desirable, since a low damping will produce a ringing transient response or even an unstable system. The frequency response (Fig. 7) has the same form as Eq. 7, but in this case it will show a resonance peak.

Actually, the phase-sensitive demodulator (PSD) associated with the displacement transducer will often include a low-pass filter at frequencies above the seismic band, thus introducing an additional phase delay in this band.

Some additional circuitry is then required to improve the transient response and prevent a possible self-oscillation of the servo system. Several techniques for this are possible. A simple one consists of including a second feedback loop to add a small force proportional to the time derivative of the mass displacement, i.e., the mass velocity. This allows the feedback force to control the mass motion “in advance,” and the servo loop is stabilized.

Figure 7 shows the responses relative to ground acceleration of several systems and output points of the systems. An under-damped mechanical accelerometer (without feedback

force) is drawn in curve *a*, in this example with a free oscillation frequency of 10 Hz. Curve *b* corresponds to an FBA built with this system by adding a displacement transducer and a feedback force proportional to the mass displacement, which increases the resonance frequency. The damping is still too low and a strong resonance peak at 200 Hz appears. Curve *c* shows the voltage output response for the same FBA once stabilized by the addition of a small feedback force proportional to the mass velocity. Curve *d* shows the feedback acceleration response (i.e., the feedback force per unit mass), which is nearly equal to the ground acceleration (so the amplitude response is 1) up to the frequencies around the resonance, where the feedback force needs to be a little higher to stabilize the motion. Curve *e* plots the mass motion response measured in microns (μm) relative to ground acceleration in m/s^2 . It is approximately 0.6 in the useful band, so a ground acceleration of 1 g will produce a mass motion of $9.8 \text{ m} \cdot \text{s}^{-2} \cdot 0.6 \mu\text{m}/(\text{m} \cdot \text{s}^{-2}) \approx 6 \mu\text{m}$, a very small displacement, for which the capacitive transducer will be quite linear.

It is also illustrative to view the transient response of the FBA (Fig. 8). The input is a simulated ground acceleration pulse of 1 m/s^2 and with duration of 50 ms. It can be seen that the output voltage has only a small overshoot, due to the compensation circuit in the feedback. So, there is an unavoidable small distortion affecting the higher frequencies, like any instrument with a

limited bandwidth, but this will not affect the band of interest. The feedback acceleration has a little overshoot to counteract the resonance and keep the mass as stationary as possible. The mass displacement is also shown (with sign changed) and has an amplitude about $0.6 \mu\text{m}$, as predicted by the frequency response in Fig. 7.

The FBA principle is the basis not only of servo-accelerometers but also of [broadband seismometers](#) (BB). These instruments have a flat frequency response for ground velocity in a band from ~ 0.01 to several tens of Hz. The feedback loop keeps the mass almost at rest relative to the ground, but the signal is integrated in the loop, and the output is taken from a point where it is proportional to ground velocity in a wide band. For more details, see the entry “[Principles of Broadband Seismometry](#).”

Characteristic Parameters of Accelerometers

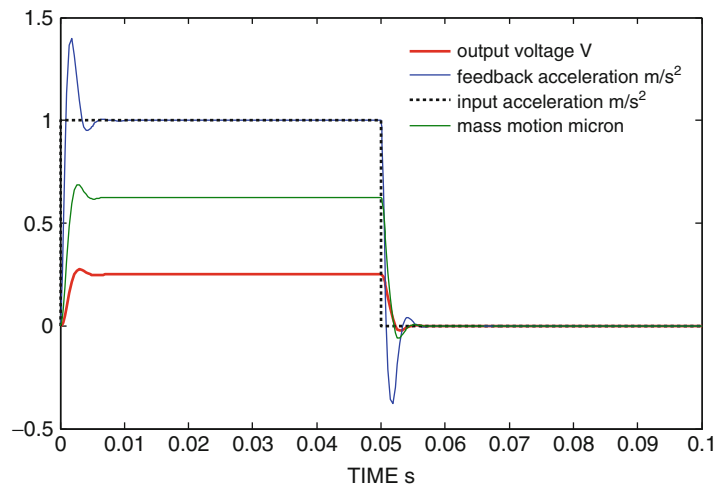
The performance of a seismic accelerometer may be characterized by several parameters.

Sensitivity: The relation between the voltage output and the input acceleration in V/m/s^2 or V/g (gravity normal acceleration = 9.807 m/s^2). Typical values are 0.5–5 V/g .

Input Full Scale (FS): The maximum ground acceleration before the instrument gets saturated. A large FS prevents saturation with strong ground

Seismic Accelerometers,

Fig. 8 Transient response of the example FBA (Fig. 7) for a pulse in acceleration input of $1 \text{ m} \cdot \text{s}^{-2}$



motion but at a cost of sensitivity. In general, for accelerometers to be installed in zones where large ground motion is expected, FS should be 2–4 g, but in low-moderate seismicity areas, 0.5–1 g would be suitable. Nevertheless, it should be kept in mind that some of the largest peak ground motions ever recorded (up to 4.36 g for Iwate-Miyagi 2008 earthquake; Yamada et al. 2009) have been produced by moderate-size earthquakes (Anderson 2010). Some accelerometers allow user setting its FS.

Resolution: It is the minimum acceleration amplitude that it may measure. In early accelerometer digital recorders, it was limited by the digitizer (A/D converter) resolution. In modern instruments, the digitizer is usually 24 bit, and the resolution is related to the self-noise level of both the accelerometer and the digitizer. So it is usually given as self-noise level, or it may be obtained from the dynamic range. Good-quality accelerometers have noise levels under $1 \mu\text{m/s}^2$ rms (root mean square).

Dynamic Range: The relation (expressed in decibels, dB) between the maximum acceleration amplitude (FS) and the resolution. A ratio between two amplitudes a_1 and a_2 expressed in dB is $20 \cdot \log(a_1/a_2)$; for energy or power ratios, a factor 10 instead of 20 is used. The best accelerometers now have a dynamic range up to more than 150 dB, which means a ratio $10^{150/20} > 30 \cdot 10^6$, but most instruments achieve 120–140 dB.

MEMS Accelerometers

The same principles described so far are valid regardless the size of the accelerometer, but accelerometer sensors of micrometric size have specific characteristics, so a brief description follows.

Originally, the aim of reducing the size of accelerometer sensors was due to the manufacturing of integrated accelerometers for industry and navigation (airborne) applications. The space exploration requirements, among other applications, led to the adaptation of some of the

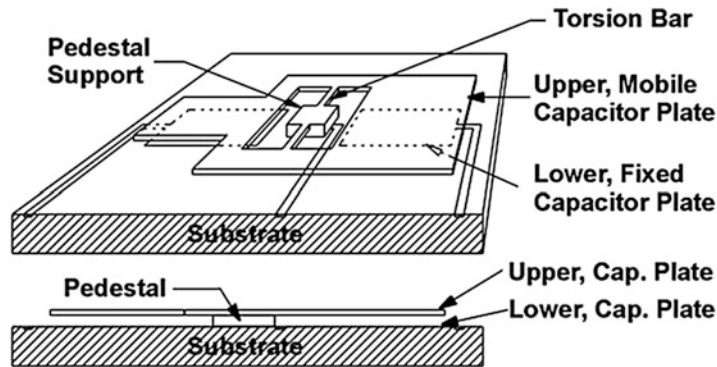
techniques of manufacturing integrated circuits to the production of very lightweight and low-power-sensitive accelerometers.

In recent years, the industry has found a large number of consumer applications for this kind of accelerometers. This has led to the development of micro-sized, light and low-power sensors, suitable for mass production and thus inexpensive. These are the *micro-electro-mechanical system* (MEMS) devices. MEMS accelerometers are used for automotive industry (e.g., for air-bag triggering), inertial navigation, medical applications, game consoles, cellular phones, tablets and laptops, etc. Most of these devices have a poor performance for seismic recording, but a few commercial models are suitable for low-resolution strong ground motion recording and earthquake early warning (EEW) systems or for triggering shutoff procedures on critical facilities (trains, gas valves, lifts, critical machinery) in case of a strong earthquake (e.g., Yanada et al. 2002). Furthermore, some seismic instrumentation companies have developed MEMS accelerometers for seismic-grade recorders with comparable performance of conventional sensors. The idea is to build the mass-spring system within the multilayer structure of an integrated circuit (Fig. 9), using the same techniques as for electronic components and circuits.

The internal noise of a spring-mass system is due to the thermal-mechanical coupling by means of the damping dissipative forces. For instance, the air molecules interchange kinetic energy with the mass, which then “dances” randomly with a Brownian motion.

It may be shown (e.g., Aki and Richards 2002; Havskov and Alguacil 2010) that this noise is proportional (Eq. 14) to the mechanical damping coefficient h (not included the feedback effect) and inversely proportional to the suspended mass. The power spectral density (PSD) $\ddot{Z}_n^2$ of the mass acceleration noise (the PSD integrated between two frequencies gives the noise variance in this band) is

$$\ddot{Z}_n^2 = \frac{8k_B T}{m} h \omega_0 \quad (14)$$



Seismic Accelerometers, Fig. 9 Principal elements of a MEMS (micro-electro-mechanical system) accelerometer with capacitive transducer. The mass is the *upper* mobile capacitor plate which can rotate around the torsion bars. The displacement, proportional to acceleration, is sensed

with the variance in the capacitance. For high-sensitive applications, a feedback circuit is added which controls a restoring electrostatic force, thus making an FBA. The size of the sensor above is about 2 mm (Figure from www.silicondesigns.com/tech.html)

where k_B is the Boltzmann constant ($1.38 \cdot 10^{-23}$ J/K), T is Kelvin temperature, m the suspended mass, h the mechanical damping coefficient, and ω_0 the free angular frequency. This equation explains why a good part of the effort in improving MEMS accelerometer noise is focused (Walmsley et al. 2009) in increasing the proof mass and reducing damping.

Thus, for a micro mass, the open-loop damping should be kept as low as possible for an acceptable noise level at ambient temperature. High-purity silicon spring elements are almost perfectly elastic, and the spring is operating in vacuum to avoid air damping, so the mechanical damping can be held quite low.

Several strategies have been tested in prototypes for sensing the mass motion. Most commercial devices use some kind of variable capacitance. Other sensing designs are based on tunnel effect, optical diffraction and interferometry, piezo-resistivity, piezoelectricity, resonance, or electrochemical phenomena. Very small capacitive transducers are not practical in general, since the circuitry parasitic capacitances make the relative changes too small. Nevertheless, these parasitic capacitances may be reduced to a minimum in MEMS devices, since the associated circuit is integrated within the same chip (see, for instance, Li et al. 2001).

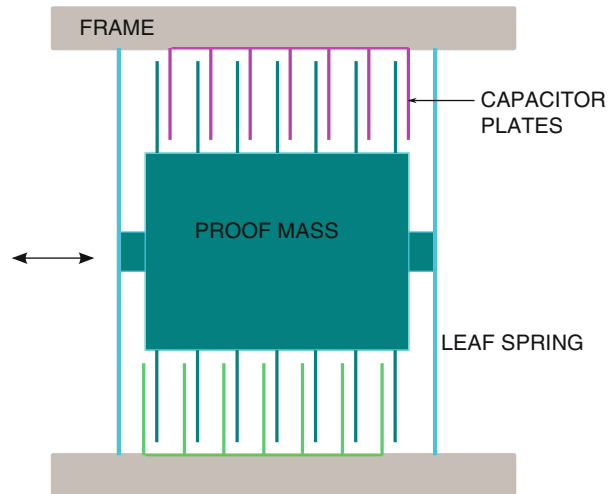
The two capacitive techniques described above are used: variable gap between plates and variable area. The first is quite nonlinear and usually requires electrostatic feedback to linearize the transducer response. Recently, electro-magnetic feedback has also been proposed for MEMS accelerometers (Dwyer 2011). On the other hand, the variable area technique is more linear, and some devices using it operate in open-loop mode (e.g., Homeijer et al. 2011). Figure 10 shows a possible arrangement for a MEMS accelerometer.

MEMS sensors are now made with electro-chemical sensors (Deng et al. 2013) with a noise PSD of $0.1 (\mu\text{m/s}^2)/\sqrt{\text{Hz}}$ or $10 \text{ ng}/\sqrt{\text{Hz}}$ at 1 Hz, so these sensors could be used at medium noisy sites.

All of these devices have two distinct parts within the same package: the mechanical sensor and the associated electronics, typically implemented by an ASIC (application-specific integrated circuit). Some commercial models have a sigma-delta modulator (a device yielding a sequence of logical pulses whose average value is proportional to its analog voltage input) included (see “► [Recording Seismic Signals](#)”), and so its output is a pulse-density digital signal, able to be directly interfaced to a microcontroller or computer. The integrated MEMS has to be

Seismic Accelerometers,

Fig. 10 A possible arrangement of the mass, spring, and capacitor plates in a MEMS accelerometer. All elements are in a layer a few microns thick



mounted in a printed circuit board (PCB) with some more electronics, at least a power supply unit, and optionally assembled in a casing with connectors, able to be fixed to another element. This assembly has to be rigid enough not to introduce parasitic resonances in the seismic band. Including electronics and housing, the total weight of these sensors might be around 0.1–0.5 kg. The sensor chip itself typically weighs less than 1 g.

The (US) Working Group on Instrumentation, Siting, Installation, and Site Metadata of the Advanced National Seismic System (ANSS 2008) defined four classes of strong-motion stations – A, B, C, and D – in terms of performance. Class A has the highest performance, with resolution better than $7\ \mu\text{g}$ and broadband dynamic range $> 111\ \text{dB}$ ($\geq 20\text{bit}$). Class B has a resolution between 7 and $107\ \mu\text{g}$ and a dynamic range 87–111 dB (or 16–20 bit). Class C resolution is $107\text{--}1,709\ \mu\text{g}$ and dynamic range 63–87 dB (12–16 bit). Class D has poorer performance than C. Currently (2014) there are no commercial MEMS accelerometers that fulfill the class A specifications, which are only met by classical macroscopic FBA devices. One of the reasons is the difficulty of achieving low noise levels with a so small suspended mass.

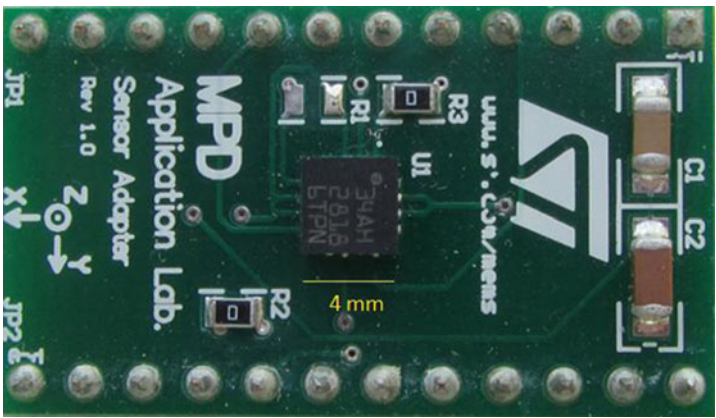
Several manufacturers use MEMSs for accelerograph recorders with class B performance, e.g., GeoSIG GMS-18 (www.geosig.com/productfile.html?productid=10319).

And some MEMS manufacturers offer class B accelerometers, e.g., Silicon Designs 1221 (www.silicondesigns.com/pdfs/1221.pdf) has typical noise PSD of $5\ \mu\text{g}/\sqrt{\text{Hz}}$ – an rms noise of $50\ \mu\text{g}$ over 100 Hz bandwidth. Recent designs (Homeijer et al. 2011) report noise levels under $100\ \text{ng}/\sqrt{\text{Hz}}$ or even $10\ \text{ng}/\sqrt{\text{Hz}}$ above 1 Hz (Milligan et al. 2011) with new capacitance transducers operating without feedback. These performances make it even suitable for use as exploration sensors, substituting geophones.

Class C devices are cheaper and mostly used in consumer products. Those models with stable response and enough bandwidth may be suitable to be applied in seismic strong ground motion monitoring, for instance, instrumental intensity estimation (shake maps), structural response, earthquake early warning, and shut off of critical facilities.

As an example of this class of sensors, Fig. 11 shows a MEMS triaxial accelerometer LIS 344 ALH (www.st.com/web/en/catalog/sense_power/FM89/SC444) from STMicroelectronics mounted on a small printed circuit board, which is sold as an evaluation board, since hand soldering the chip for prototyping is quite a difficult task. This chip accelerometer has a user-selectable full-scale $\pm 2\ \text{g}$ or $\pm 6\ \text{g}$ and comes with factory-trimmed sensitivity and offset. Its bandwidth may be selected by an external capacitor.

Seismic Accelerometers,
Fig. 11 A MEMS triaxial accelerometer LIS 344 ALH from STMicroelectronics (*center*) mounted in a small printed circuit with the minimal external components



Seismic Accelerometers,
Fig. 12 *Upper*, accelerogram obtained with a class B accelerograph from an earthquake of moment magnitude $m_w = 3.8$ at an epicentral distance of 16 km. *Lower*, the simulated record with the MEMS sensor of Fig. 11 by adding the recorded self-noise

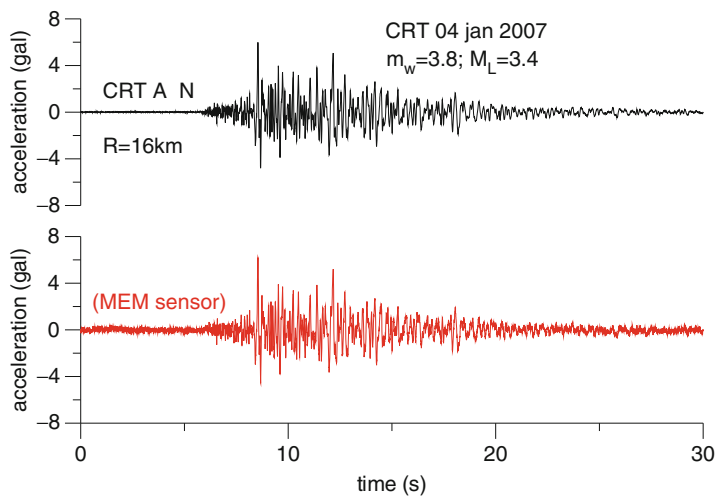


Figure 12 plots a small local earthquake recorded by a standard accelerometer and the simulated record with this sensor, using its real noise. The MEMS sensor shows a higher noise, but a useful signal is still available.

A comparative test of these sensors (Evans et al. 2014) shows that some of them could be very useful for low-cost seismic networks. Some models have performances suitable for strong ground motion recording with acceptable SNR to be used as class C seismic accelerometers.

Accelerometer Examples

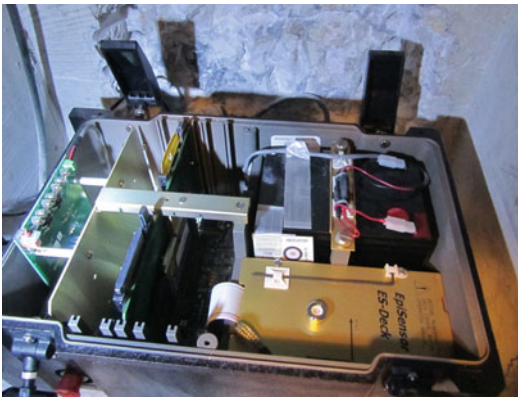
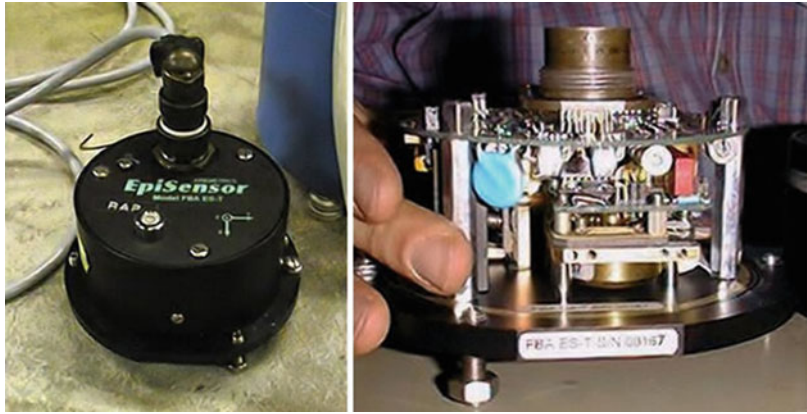
Many seismic accelerometers commercially available are sold assembled with the corresponding

recorder in a case, which is called an accelerograph. Some of them may be acquired separately, with an analog voltage output for each axis. For earthquake recording, the standard is a triaxial instrument. Practically all the recorders have GPS synchronization of internal clock. Some models have the possibility of Internet timing, for sites where GPS signal cannot be received.

Most accelerograph manufacturers offer several models with different performance and price. The models shown below are just examples and may not be representative of the best instrument offered by each company.

Figure 13 shows one of the accelerometers with highest dynamic range, the EpiSensor from Kinemetrics (www.kinemetrics.com/p-87-EpiSensor-ES-T.aspx), a triaxial FBA with

Seismic Accelerometers, Fig. 13 The EpiSensor accelerometer from Kinemetrics. *Left*, surface sensor package. *Right*, the case is removed to show the components



Seismic Accelerometers, Fig. 14 The Etna accelerograph from Kinemetrics with its lid removed. It has an internal triaxial accelerometer (EpiSensor). The hermetic case is made of fiberglass



Seismic Accelerometers, Fig. 15 CMG 5TD from Guralp Systems. This model can include a recorder and an Ethernet interface. The photo shows a unit installed in an underground gallery with the mounting base anchored to the concrete ground

full-scale selectable up to ± 4 g. Also available is a uniaxial and with different packages, including borehole. It may be mounted inside a compact accelerograph package like Etna model (Fig. 14). This robust accelerograph model has been in the market for many years, so, in spite of the good accelerometer inside, its performance as recorder (dynamic range, communication interface, data storage) does not fully match the sensor specifications, and it may be considered a class B accelerograph. The same manufacturer offers new recorder models with higher performance.

Guralp Systems (www.guralp.com/products/instruments/cm-g-5) CMG5 (Fig. 15) is an FBA

with high dynamic range that can be supplied with built-in digitizer, recorder, and communication facilities. It is a class A accelerograph. A borehole version is also available.

Figure 16 shows Nanometrics Titan (www.nanometrics.ca/products/titansma). It is one of the newest class A accelerographs, with internal FBA and high dynamic range in a very compact



Seismic Accelerometers, Fig. 16 Nanometrics Titan accelerograph. Its size is approximately 18 x 12 x 10 cm



Seismic Accelerometers, Fig. 18 The REFTEK 148-1 “QuakeRock” accelerograph, with MEMS sensor (Photo from REFTEK (www.reftek.com/products/motion-recorders-148-01.htm))



Seismic Accelerometers, Fig. 17 FBA with MEMS technology AC-43 from GeoSIG. It weighs 2 kg and its size is 19 x 11 x 9 cm (Figure from GeoSIG (www.geosig.com/AC-4x-id10357.html))

package. It has a web interface and can store data in a removable SD memory card.

Finally, two accelerometers with MEMS technology will be shown. In Fig. 17, the GeoSIG AC-43 (www.geosig.com/AC-4x-id10357.html) is a triaxial FBA with class B performance when mounted inside or connected to a recorder, such as the GSR-18, from the same company. The model 148-01 “QuakeRock” from REFTEK (Fig. 18)

is a class C accelerograph with only event recording (not continuous) and limited dynamic range but may operate unattended for 2 years with two “D” size batteries. The internal clock is free running (no GPS).

Accelerograph Installation

Two kinds of accelerometer or accelerograph installations are possible: (a) free-field installation and (b) structural-monitoring installation. The purpose of the first is to record the ground motion unaffected by man-made structures. The acceleration records obtained from this kind of stations may be considered the base-level excitation of any building or structure in the zone, if the building itself does not interact with the ground at the station. The second type is an accelerograph in a building or engineering structure at different levels and positions to study and monitor the structure vibrations in response to ground seismic motions.

The main consideration is that the installation setup should not affect the accelerograph records.

For free-field stations, it is usual to build a small concrete pier on which the accelerometer (or accelerograph if the sensor is inside it) must be firmly anchored so a strong motion cannot move the instrument relative to ground! Most commercial instruments include a suitable base with anchoring holes or a similar system. Other materials for pier, like a table made with steel bars, are not suitable since it may resonate with very low damping in the seismic band of interest.

Free-field accelerographs should not be close to tall buildings that may modify the ground motion in their proximity. Actually some accelerograph stations are installed at the basement of buildings and are considered as free-field but may not truly fulfill this condition, due to the soil-structure interaction.

If the free-field installation is to be done on soil, the pier should have a suitable foundation to assure that it is well coupled to the ground, but not being so heavy as to modify the local soil dynamic response substantially. Free-field accelerographs must be protected with a cover or small hut from the weather. Provisions have to be made for power supply, GPS antenna, and communication, usually Ethernet via cable, satellite link, or cellular modem. And a fence around the installation would protect it from animals, human-made noises, and eventually vandalism. Accelerographs are not very sensitive to weak motions, but modern high-resolution instruments are capable of detecting human activity such as traffic or machinery working at short distance, so this kind of noise should be avoided as far as possible.

Structural-monitoring installations are usually done under cover, and additional weather protection is not required, but the instrument may have to be protected from human activity or other disturbances. Normally a pier is not required, since the sensor may be anchored directly to a structural element. GPS reception for time synchronism may be a problem if the station is far from the open sky (e.g., in a dam gallery), but there exist technical solutions: e.g., a GPS receiver may be outside, and the signal is “repeated”; the accelerograph may be

synchronized via Ethernet or the accelerometer is installed on the site, and the recorder is separately installed near the open sky so the GPS antenna can be placed outside.

Summary

Seismic accelerometers sense the ground or structure seismic vibrations and, together with a suitable recorder, are called accelerographs. Most modern seismic accelerometers are of force-balance type (FBA), a servo system in which a feedback force is applied to the suspended inertial mass to keep its motion as small as possible. This improves the instrument linearity and dynamic range. Usually the mass motion is measured by a sensitive capacitive transducer.

MEMS accelerometers are integrated micromachined electromechanical devices widely used in industry that presently do not match the classical FBA performance, but are useful for some seismic applications.

Examples of commercial seismic accelerometers and accelerographs and some brief guidelines for the installation of free-field and structure-monitoring accelerographs are given.

Cross-References

- ▶ [Passive Seismometers](#)
- ▶ [Principles of Broadband Seismometry](#)
- ▶ [Recording Seismic Signals](#)
- ▶ [Seismic Network and Data Quality](#)

References

- Aki K, Richards PG (2002) Quantitative seismology. University Science Books, Sausalito CA, 700 pp
- Akkar S, Bommer JJ (2007) Empirical prediction equations for peak ground velocity derived from strong-motion records from Europe and the Middle East. *Bull Seismol Soc Am* 97:511–530. doi:10.1785/0120060141
- Anderson JG (2010) Source and site characteristics of earthquakes that have caused exceptional ground accelerations and velocities. *Bull Seismol Soc Am* 100:1–36

- ANSS Working Group on Instrumentation, Siting, Installation, and Site Metadata of the Advanced National Seismic System Technical Integration Committee (2008) Instrumentation guidelines for the advanced national seismic system, U.S. Geological Survey. Open-File Report 2008–1262, 41 pp
- Deng T, Chen D, Wang J, Chen J, He W (2013) A MEMS based electrochemical vibration sensor for seismic motion monitoring. *J Microelectromech Syst* 99:920–923
- Dwyer PD (2011) MEMS accelerometer. US Patent 8–065.915 B2, 29 Nov 2011
- Eisenberg A, McEvilly TV (1971) Comparison of some widely-used strong-motion earthquake accelerometers. *Bull Seismol Soc Am* 61:383–397
- Evans JR, Allen RM, Chung AI, Cochran ES, Guye R, Hellweg M, Lawrence JF (2014) Performance of several low-cost accelerometers. *Seismological Research Letters* 85:147–158
- Havskov J, Alguacil G (2010) Instrumentation in earthquake seismology. Springer, Dordrecht, 358 pp
- Homeijer B, Lazaroff D, Milligan D, Alley R, Wu J, Szepesi M, Bicknell B (2011) Hewlett Packard's seismic grade MEMS accelerometer. In: *Proceedings of the MEMS*, Jan 2011, Cancún, Mexico, pp 585–588
- Li B, Lu D, Wang W (2001) Micromachined accelerometer with area-changed capacitance. *Mechatronics* 11:811–819
- McComb HE, Ruge AC (1937) Tests of earthquake accelerometers on a shaking table. *Bull Seismol Soc Am* 27:325–329
- Milligan DJ, Homeijer BD, Walmsley RG (2011) An ultra-low noise MEMS accelerometer for seismic imaging. In: *Sensors 2011 IEEE*, October 2011, Limerick, DOI 10.1109/ICSENS.2011.6127185, pp 1281–1284
- Reasenberg P (1969) Non-linearity test of an accelerometer system designed for the seismic near-field measurement. *Bull Seism Soc Am* 59:1213–1225
- Ruge AC, McComb HE (1943) Tests of earthquake accelerometers on a shaking table. *Bull Seismol Soc Am* 33:2–12
- Trifunac MD (2009) 75th anniversary of strong motion observation—a historical review. *Soil Dyn Earthq Eng* 29:591–606
- Walmsley RG, Kiyama LK, Milligan DM, Alley RL, Erickson DL, Hartwell, PG (2009) Micro-G silicon accelerometer using surface electrodes. In: *IEEE sensors 2009 conference*, October 2009, Christchurch, DOI 10.1109/ICSENS.2009.5398201, pp 971–974
- Yamada M, Mori J, Heaton T (2009) The slapdown phase in high-acceleration records of large earthquakes. *Seismol Res Lett* 80:559–564. doi:10.1785/gssrl.80.4.559
- Yanada T, Furukawa H, Ichida S, Takubo K, Shimizu Y, Koganemaru K, Suzuki T (2002) Earthquake sensor. *Sens Update* 11:159–216. DOI:10.1002/seup.200211101

Seismic Actions Due to Near-Fault Ground Motion

George P. Mavroeidis

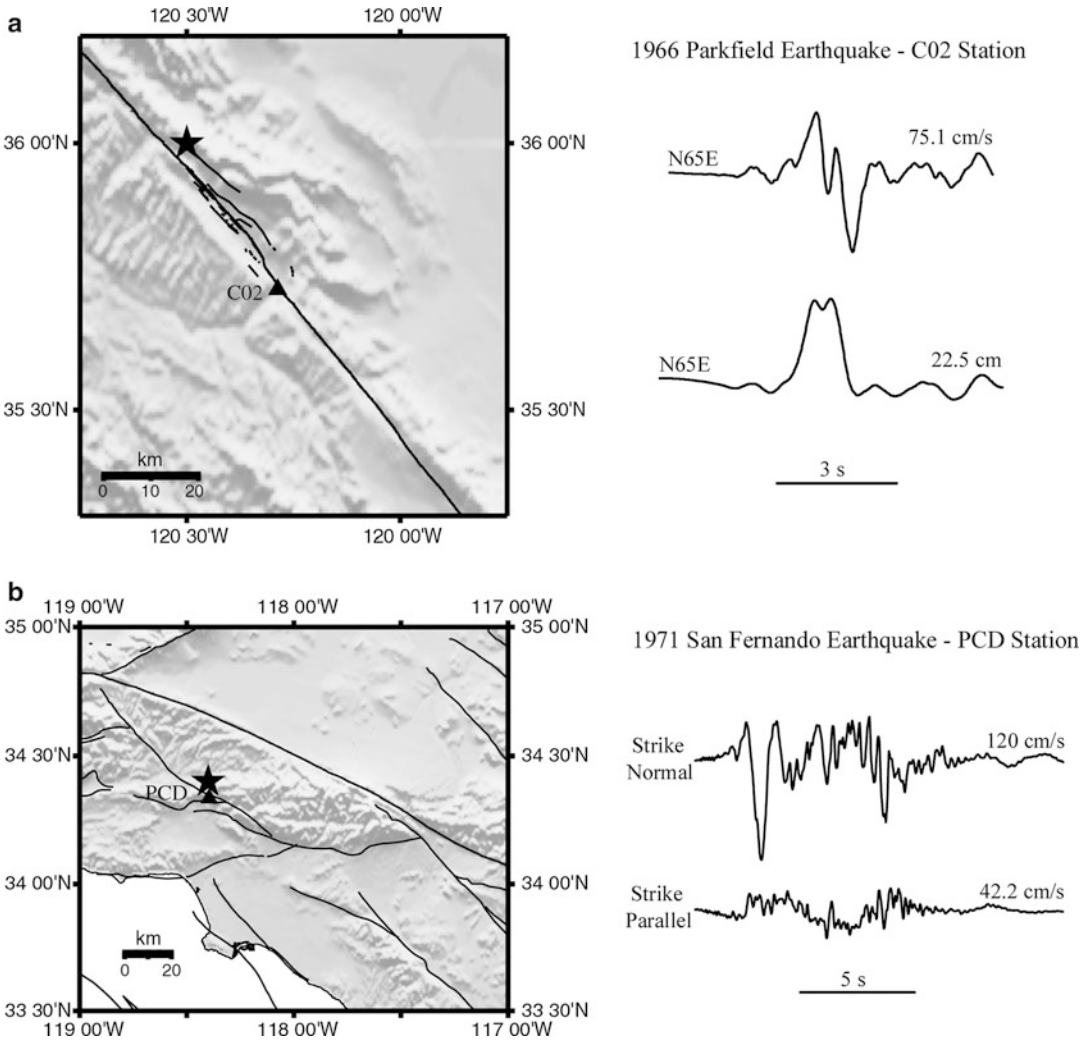
Department of Civil and Environmental Engineering and Earth Sciences, University of Notre Dame, Notre Dame, IN, USA

Synonyms

Damping coefficient; Forward directivity; Near field; Near source; Permanent translation (fling); Response spectrum; Seismic ground excitation; Strength reduction factor; Time history

Introduction

Near-fault seismic ground motions are frequently characterized by intense velocity and displacement pulses of relatively long duration that clearly distinguish them from typical far-field ground motion records. This observation, along with its engineering significance, was first made with respect to the C02 record (Fig. 1a) generated by the 1966 Parkfield earthquake at a distance of only 80 m from the fault break (Housner and Trifunac 1967). The damage that the Olive View Hospital sustained during the 1971 San Fernando earthquake was also attributed to the effect of near-fault ground motions on flexible structures (Bertero et al. 1978). That was perhaps the first time that earthquake engineers linked the structural damage caused by an earthquake to the impulsive character of near-fault ground motions (Fig. 1b). However, it was not until the 1994 Northridge and the 1995 Kobe earthquakes that the majority of engineers recognized the destructive potential of near-fault ground motions and started considering methods to incorporate near-source effects into engineering design. Code provisions have historically been developed based on recorded ground motions not sufficiently close to the causative fault. Thus, the effect of near-fault pulse-like ground motions on the dynamic



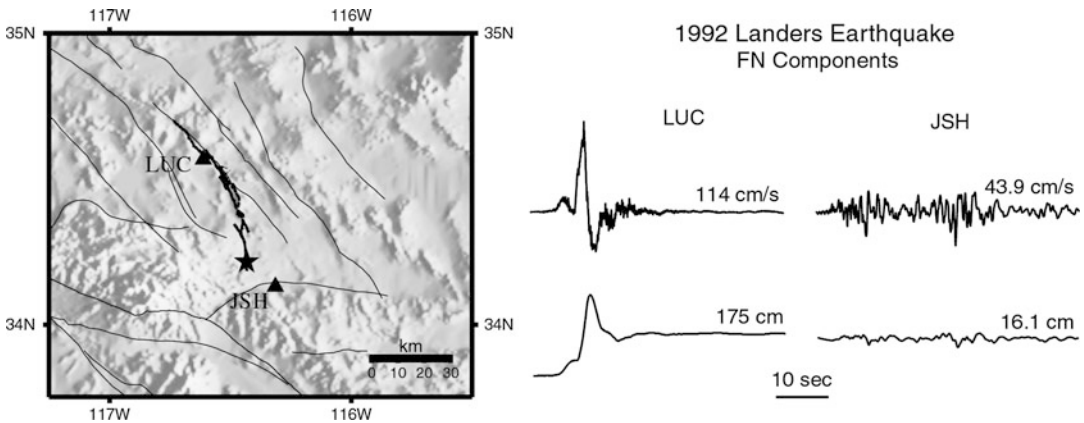
Seismic Actions Due to Near-Fault Ground Motion, Fig. 1 Characteristic examples of near-fault pulse-like ground motion records: (a) Station No. 2 (C02) record from the 1966 Parkfield, California, earthquake and (b)

Pacoima Dam (PCD) record from the 1971 San Fernando, California, earthquake (Reprinted from Mavroeidis and Papageorgiou (2003). Copyright © 2003 Seismological Society of America)

response of engineering structures has received much attention over the past two decades.

This entry focuses on the description of seismic actions due to near-fault ground motions with particular emphasis on the following topics: (1) characteristics of near-fault ground motions, (2) effect of fault rupture parameters on near-fault seismic excitations, (3) synthesis of near-fault ground motion time histories for earthquake engineering applications, and (4) derivation of response spectra, strength reduction factors and

damping coefficients for engineering analysis and design in the near-fault region. The material presented in this entry is primarily based on previous articles published by the author and Professor Apostolos S. Papageorgiou and is presented in a manner that provides established knowledge in the disciplines of engineering seismology and earthquake engineering to technically inclined and informed readers. It should be emphasized that this entry does not intend to be a specialized research article advancing the current state of



Seismic Actions Due to Near-Fault Ground Motion, Fig. 2 Characteristic examples of forward and backward directivity from the 1992 Landers, California, earthquake

(Reprinted from Mavroeidis and Papageorgiou (2002). Copyright © 2002 Earthquake Engineering Research Institute)

knowledge or a review article summarizing the vast amount of archived research literature on the subject.

Characteristics of Near-Fault Ground Motions

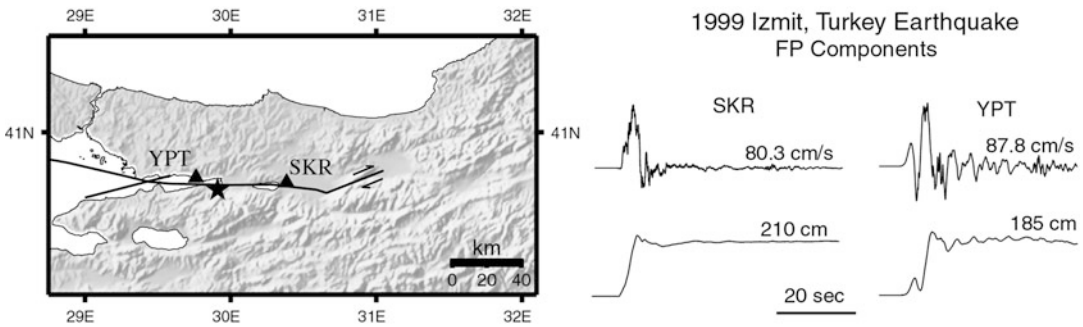
Not all ground motion time histories recorded at stations in the vicinity of a fault exhibit intense velocity pulses. The existence of pulse-like ground motions in near-fault records primarily depends on the relative position of the station that recorded the motion with respect to the direction of propagation of rupture on the causative fault plane and on the magnitude and direction of slip on that segment of the fault that is located in the vicinity of the station. Whenever these ground motion pulses do occur, they are typically caused by the forward directivity and/or permanent translation (fling) effects.

Forward directivity occurs when the fault rupture propagates toward a site with a rupture velocity that is approximately equal to the shear wave velocity. In this case, most of the energy arrives coherently in a single, intense, relatively long-period pulse at the beginning of the record representing the cumulative effect of almost all the seismic radiation from the fault. Forward directivity pulses are polarized in the fault-normal direction for both strike-slip and dip-slip

faults. Figure 2 illustrates a characteristic example of forward directivity from the 1992 Landers earthquake. The fault rupture propagated to the north along the indicated strike-slip fault. The fault-normal velocity and displacement time histories recorded at Lucerne Valley (LUC) station (which is located in the forward direction with respect to the propagation of rupture) are characterized by intense pulse-like motions. In contrast, the ground motion recorded at Joshua Tree (JSH) station (which is located in the backward direction with respect to the propagation of rupture) is relatively weak.

Permanent translation (fling) is a consequence of permanent fault displacement due to an earthquake; it appears in the form of step displacement and one-sided velocity pulse in the strike-parallel direction for strike-slip faults or in the strike-normal direction for dip-slip faults. In the latter case, directivity and permanent translation effects “build up” in the same direction. Figure 3 illustrates characteristic examples of permanent translation (fling) from the 1999 Izmit earthquake. The fault-parallel velocity and displacement time histories recorded at Yarimca (YPT) and Sakarya (SKR) stations are affected by the permanent displacement along the right-lateral strike-slip North Anatolian Fault.

Even though emphasis has traditionally been given to the investigation of forward directivity and permanent translation (fling) effects, other



Seismic Actions Due to Near-Fault Ground Motion, Fig. 3 Characteristic examples of permanent translation (fling) from the 1999 Izmit, Turkey, earthquake

(Reprinted from Mavroeidis and Papageorgiou (2002). Copyright © 2002 Earthquake Engineering Research Institute)

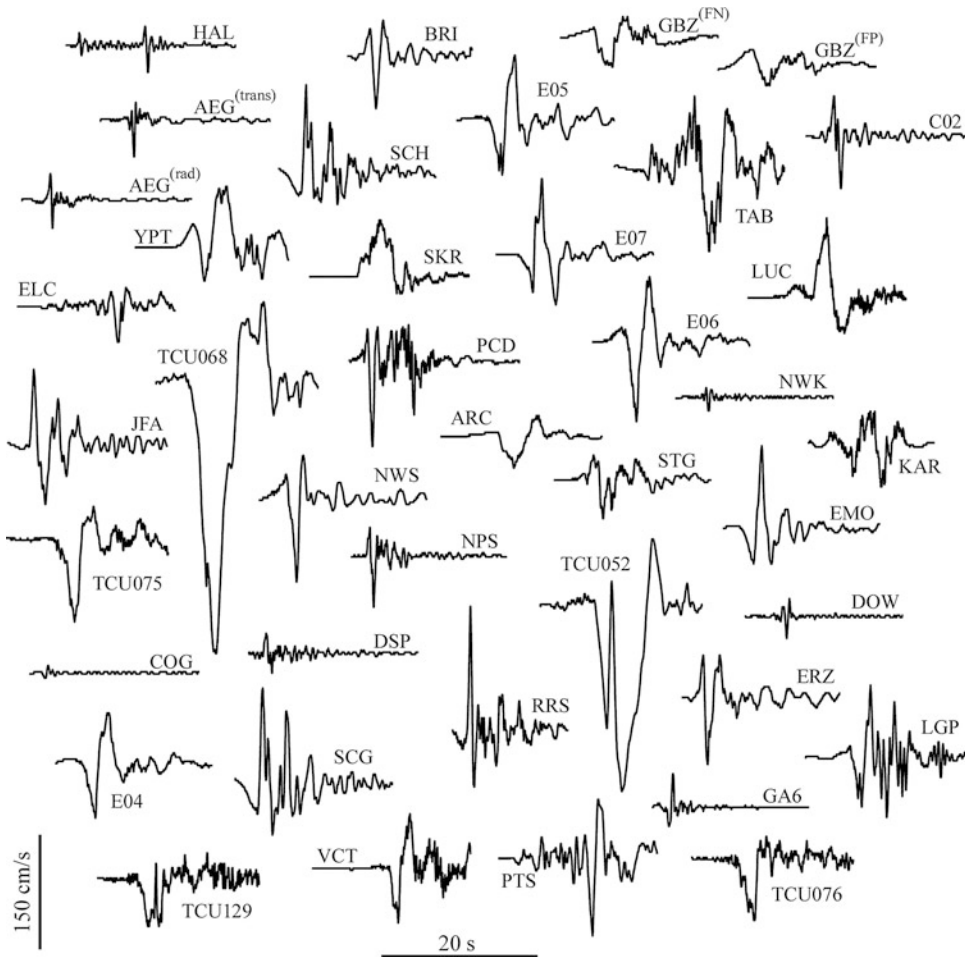
conditions may also give rise to near-fault pulse-like motions. A comprehensive review of the factors that influence near-fault ground motions, along with a detailed list of references on the subject, has been presented by Mavroeidis and Papageorgiou (2002, 2003). Figure 4 illustrates a large number of actual near-fault ground motion records with “distinct” velocity pulses. These records are part of the near-fault ground motion database compiled by Mavroeidis and Papageorgiou (2003). It is evident that the pulse duration (or period), the pulse amplitude, as well as the number and phase of half cycles are the key parameters that define the waveform characteristics of the near-fault velocity pulses.

Effect of Fault Rupture Parameters on Near-Fault Ground Motions

The effect of fault rupture characteristics on near-fault ground excitations has been investigated by Mavroeidis and Papageorgiou (2010) using a kinematic modeling approach. In order to associate fault rupture characteristics (such as slip, rupture velocity, and state of stress) with near-fault ground motions, four well-documented seismic events (1979 Imperial Valley, 1985 Michoacan, 1989 Loma Prieta, and 1999 Izmit) were considered along with the concept of isochrones. An isochrone is the locus of all those points on the fault plane, the radiation of which arrives at a certain observer at a specified time.

Isochrones are frequently used in seismology to provide intuitive insight into factors that strongly influence the generation of strong ground motions. By plotting the *S*-wave isochrones on the fault plane of the investigated seismic events, the long-period velocity pulses of the near-fault ground motions can be directly associated with specific regions and characteristics of the fault rupture.

The results indicated that the seismic energy radiated from the high-isochrone-velocity region of the fault arrives at the receiver within a time interval that coincides with the time window of the long-period ground motion pulse recorded at the site. Furthermore, the near-fault ground motion pulses are strongly correlated with large slip on the fault plane locally driven by high stress drop. In addition, the local rupture velocity seems to be inversely correlated to the spatial distribution of the strength excess over the fault plane confirming findings of previous studies (e.g., Bouchon 1997). As an example, Fig. 5 illustrates time histories of near-fault ground motions and *S*-wave isochrones for selected stations of the 1979 Imperial Valley earthquake. These stations are located close to the ruptured fault where the effect of forward directivity was pronounced. The spatial distribution of the static slip offset and rupture time inferred by Archuleta (1984) and the spatial distribution of the static stress drop and strength excess calculated using the methodology proposed by Bouchon (1997) are also illustrated.



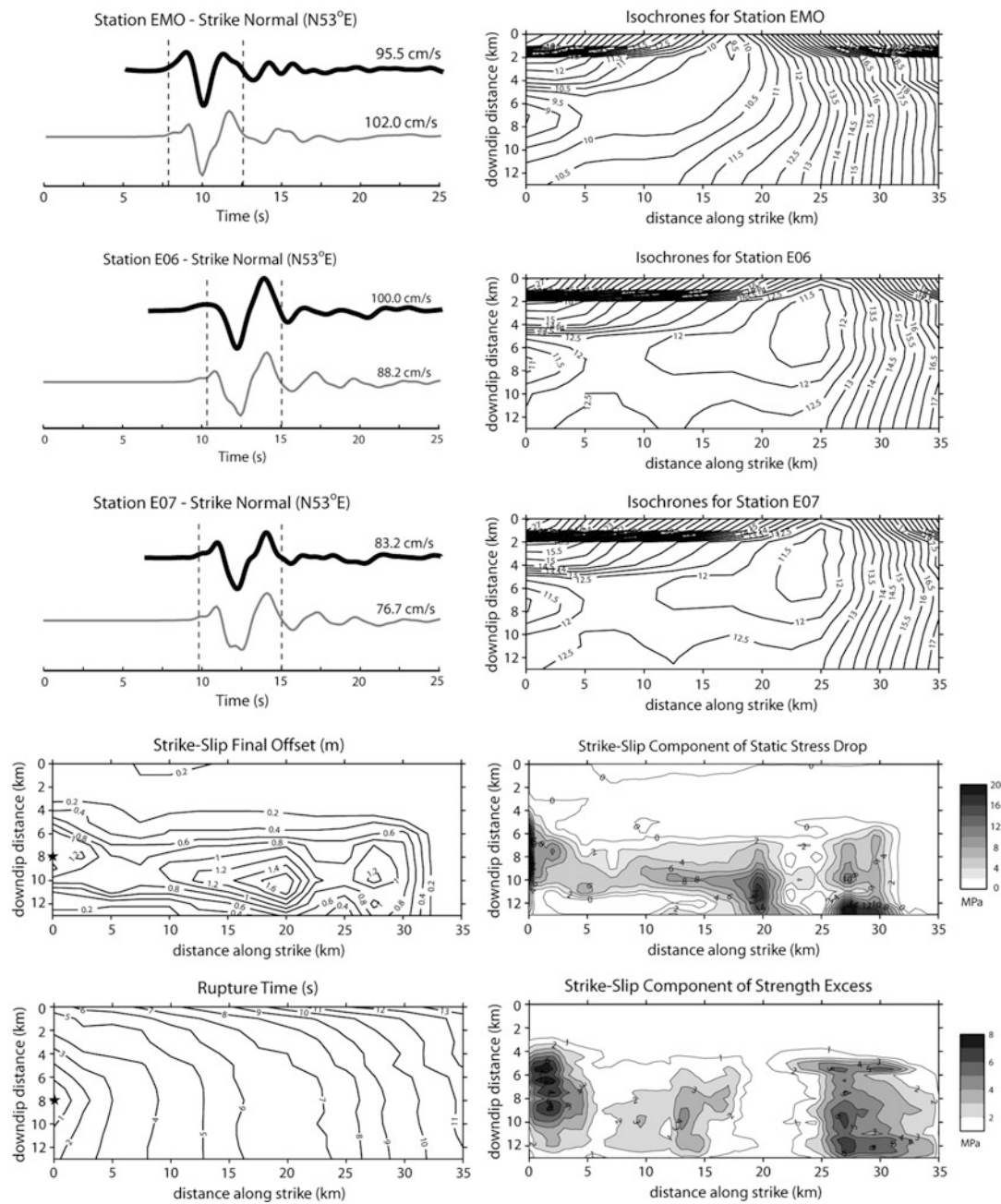
Seismic Actions Due to Near-Fault Ground Motion, Fig. 4 Strong motion records with “distinct” velocity pulses (Reprinted from Mavroeidis and Papageorgiou (2003). Copyright © 2003 Seismological Society of America)

Finally, it was found that for various events, the area of the fault that contributes to the formation of the near-fault pulse encompasses more than one patch of significant moment release (subevent) (e.g., 1979 Imperial Valley, 1989 Loma Prieta). This observation explains why a dislocation model with average properties (i.e., slip, rise time, etc.) reproduces successfully near-fault ground motions for strike-slip faults and for dip-slip faults with intermediate-to-large earthquake magnitudes. However, for very large earthquakes, such as megathrust events on subduction zones (e.g., 1985 Michoacan), the fault region that contributes to the pulse formation

encompasses individual subevents, and, consequently, crack-like slip functions (rather than dislocation models) may be more appropriate for the simulation of the near-fault ground motions. The interested reader may find a detailed discussion in Mavroeidis and Papageorgiou (2010).

Time Histories of Near-Fault Ground Motions

The advent of performance-based earthquake engineering, the growth of computer processing power, and the associated increased availability



Seismic Actions Due to Near-Fault Ground Motion, Fig. 5 Recorded (*black trace*) and synthetic (*gray trace*) near-fault ground motion time histories and *S*-wave isochrones for selected stations of the 1979 Imperial Valley, California, earthquake. Tomographic images of the static

slip offset, rupture time, static stress drop, and strength excess are also illustrated (Reprinted from Mavroeidis and Papageorgiou (2010). Copyright © 2010 Seismological Society of America)

of structural analysis software have made possible the performance of sophisticated nonlinear structural analysis on a routine basis. However, the overall seismic performance-based assessment of a given structure hinges on the use of realistic earthquake ground motions that reflect the seismic hazard at the site of the structure, as well as the local site conditions (Halldorsson et al. 2011).

In general, earthquake engineers have the following options for selecting ground motion input when performing nonlinear structural analysis in the near-fault region: (1) use actual records of near-fault ground motion, (2) generate synthetic records of near-fault ground motion using physical models of the seismic source, and (3) generate synthetic records of near-fault ground motion using phenomenological models.

Recorded Near-Fault Ground Motions

The gradually increasing number of recorded near-fault ground motions has recently enabled strong motion seismologists to compile these records in publicly available ground motion databases (e.g., Pacific Earthquake Engineering Research Center Ground Motion Database, Center for Engineering Strong Motion Data, European Strong Motion Database, among others). Even though the number of near-fault records is still limited, they have served as an invaluable resource to earthquake engineers. Researchers have also proposed methodologies for identifying and extracting pulse-like motions from actual near-fault records using wavelet analysis (e.g., Baker 2007; Vassiliou and Makris 2011).

However, the selection of strong motion records for nonlinear structural analysis is not always a straightforward process. For instance, the available records may not reflect the appropriate earthquake magnitude, source mechanism, site conditions, or source-site configuration. While this could be a problem for far-field sites, it is even a greater challenge for sites in the immediate vicinity of the fault. For the above reasons, it is of paramount importance to earthquake engineers to have the ability to generate suites of realistic broadband ground motion time

histories, both in the far-field and near-fault regions (Halldorsson et al. 2011).

Synthetic Near-Fault Ground Motions Using Physical Models of the Seismic Source

Strong motion seismologists have utilized various schemes of deterministic and stochastic simulation techniques to generate broadband ground motion time histories at specific locations in the vicinity of the fault (see, e.g., Papageorgiou 1997 and references provided therein). These simulation methods are based on source mechanics principles and wave propagation theory. Site effects are also frequently taken into account. Regardless of the degree of sophistication of the various ground motion simulation methods, the Earth crustal structure and the seismic source should sufficiently be characterized and quantified.

For regions of intense seismic activity, the crustal structure is frequently defined in terms of one-dimensional velocity models. Detailed three-dimensional crustal models have also become available for specific regions in the benefit of three-dimensional wave propagation codes that may effectively take into account basin effects and complex fault geometries at the cost of increased computational demands. The characterization of the seismic source is a more complicated issue. For kinematic descriptions of the earthquake source, source parameters such as slip, rise time, rupture velocity, and slip function should properly be quantified and a priori defined. On the other hand, for dynamic descriptions of the earthquake source, the source parameters may vary as long as the elastodynamics equation with a prescribed fracture criterion on a predetermined fault plane is satisfied. The selected initial conditions and failure criterion determine the time and space evolution of the fault rupture in a dynamic source model.

Once the seismic source and Earth crustal model have been adequately described, near-fault ground motion simulations in the low-frequency range (e.g., below 1 Hz) can be performed using deterministic modeling techniques [e.g., discrete wavenumber method (DWN), finite difference method (FDM), finite element method (FEM), boundary element

method (BEM), spectral element method (SEM), or hybrids of them] that involve calculations of synthetic Green's functions. In order to generate broadband synthetic ground motions, the low-frequency waveforms should be combined with high-frequency ground motions (e.g., above 1 Hz) simulated using: (1) the empirical or semiempirical Green's function method or (2) a stochastic modeling technique utilizing a source model that provides an unambiguous way to distribute the seismic moment of the simulated event on the fault plane. This matter is of great importance for near-fault ground motion simulations due to the proximity of the point of observation to the source. It should be mentioned that high-frequency ground motion simulations can be carried out using synthetic Green's functions as well, excluding site effects and small-scale heterogeneities.

Synthetic Near-Fault Ground Motions Using Phenomenological Models

Ground motion simulation techniques based on kinematic or dynamic source models are not always appealing to earthquake engineers because specialized seismological knowledge and, quite frequently, demanding computational resources are required. Therefore, in practice, earthquake engineers utilize actual near-fault records of past earthquakes to investigate the dynamic response of engineering structures to pulse-like seismic excitations and rely on strong motion seismologists only for generating site-specific near-fault ground motions for the design of special structures.

To overcome this deficiency, earthquake engineers have introduced idealized waveforms, intending to represent typical ground motion pulses observed in the near-fault region, in an effort to investigate the dynamic response of engineering structures to near-fault ground motions (e.g., Makris 1997; Sasani and Bertero 2000; Alavi and Krawinkler 2001; Mavroeidis and Papageorgiou 2003, among others). These idealized waveforms should successfully capture the impulsive character of the near-fault records both qualitatively and quantitatively. In addition, their input parameters should have a clear physical

interpretation and scale, to the extent possible, with physical parameters of the faulting process.

Mathematical Representation of Near-Fault Ground Motion Pulses Proposed by Mavroeidis and Papageorgiou (2003)

The mathematical formulation for the representation of the near-fault ground velocity pulses proposed by Mavroeidis and Papageorgiou (2003) is the product of a harmonic oscillation with a bell-shaped function. That is:

$$v(t) = \begin{cases} A \frac{1}{2} \left[1 + \cos \left(\frac{2\pi f_p}{\gamma} (t - t_0) \right) \right] \cos [2\pi f_p (t - t_0) + v], \\ t_0 - \frac{\gamma}{2f_p} \leq t \leq t_0 + \frac{\gamma}{2f_p} \text{ with } \gamma > 0, \text{ otherwise} \end{cases} \quad (1)$$

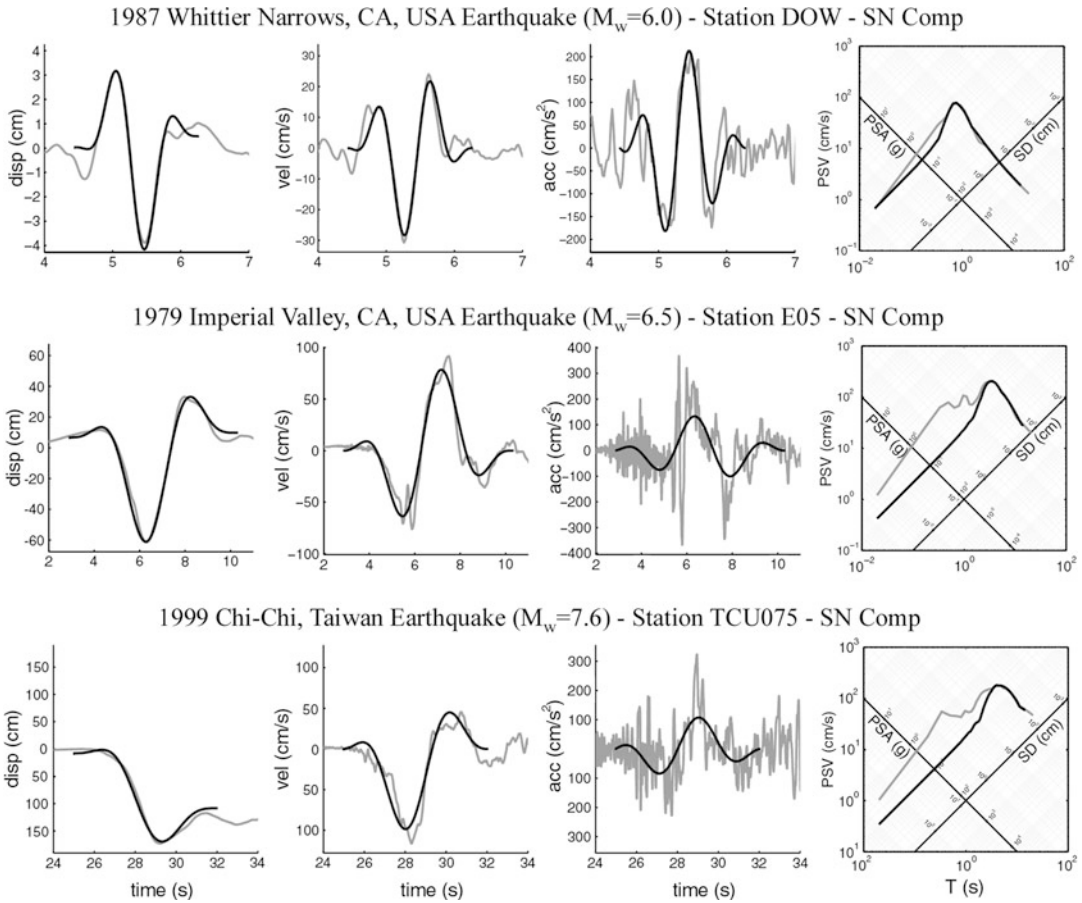
where A controls the amplitude of the signal, f_p is the frequency of the amplitude-modulated harmonic (or the prevailing frequency of the signal), v is the phase of the amplitude-modulated harmonic, γ is a parameter that defines the oscillatory character of the signal, and t_0 specifies the epoch of the envelope's peak. The pulse period (T_p) is defined as the inverse of the prevailing frequency (f_p) of the signal, thus providing an "objective" assessment of this important parameter. That is:

$$T_p = \frac{1}{f_p} \quad (2)$$

The model input parameters have a clear physical meaning as they coincide with the key features that determine the waveform characteristics of the near-fault pulses (i.e., amplitude, duration, phase and number of half cycles).

The mathematical model proposed by Mavroeidis and Papageorgiou (2003) was calibrated using a large number of actual near-fault records. It successfully replicated a large set of displacement, velocity, and, in many cases, acceleration time histories, as well as the corresponding elastic response spectra. A sample of the quality of fitting of the synthetic waveforms to actual near-fault records is illustrated in Fig. 6.

The scaling characteristics of the model input parameters were also investigated through



Seismic Actions Due to Near-Fault Ground Motion, Fig. 6 Sample of synthetic waveforms (*black trace*) fitted to actual near-fault records (*gray trace*). Ground motion time histories (displacement, velocity, and acceleration),

as well as the 5% damped elastic response spectra are illustrated (Reprinted from Mavroeidis et al. (2004). Copyright © 2004 John Wiley & Sons, Inc.)

regression analysis, and simple empirical relationships were proposed. By performing least-squares analysis (Fig. 7a), the following relationship was obtained between the pulse period and the earthquake magnitude:

$$\log T_P = -2.2 + 0.4 M_w \quad (3a)$$

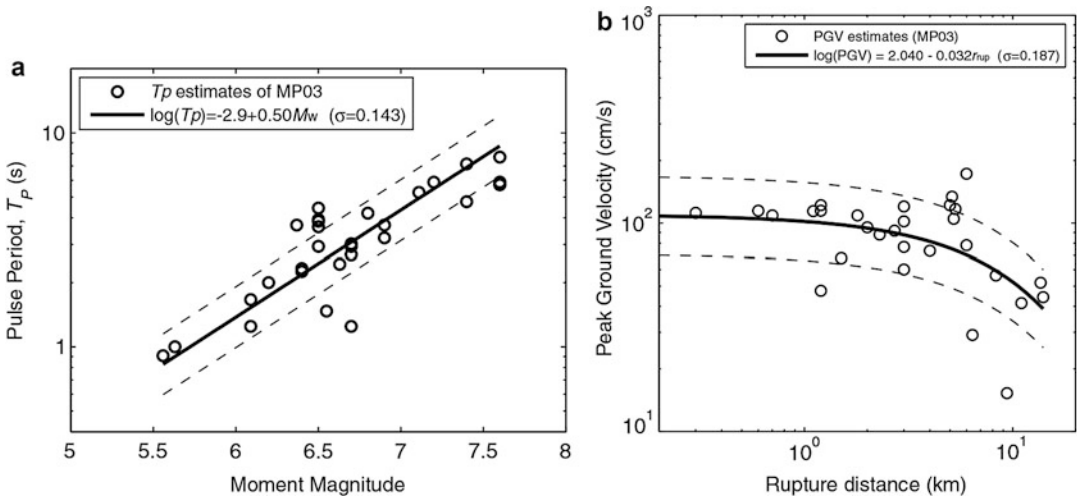
If the empirical relationship is required to satisfy the self-similarity condition, the following equation can be obtained:

$$\log T_P = -2.9 + 0.5 M_w \quad (3b)$$

Equation (3) was derived by Mavroeidis and Papageorgiou (2003) using near-fault ground

motion records affected by forward directivity. Similar scaling equations have been proposed by other investigators (e.g., Somerville 2003; Bray and Rodriguez-Marek 2004; Baker 2007, among others). However, it should be pointed out that the definition of the pulse period in these studies is not the same as the definition provided by Mavroeidis and Papageorgiou (2003), and therefore caution should be utilized when the mathematical model of Eq. 1 is used in conjunction with scaling laws for the pulse period proposed by other investigators.

Mavroeidis et al. (2004) also derived an equation that relates the pulse period (T_P) with the rise time (τ) (i.e., the time it takes for a representative point on the fault plane to reach its final displacement). The rise time is an important physical



Seismic Actions Due to Near-Fault Ground Motion, Fig. 7 (a) Scaling of the pulse period with earthquake magnitude, according to Mavroeidis and Papageorgiou (2003), and (b) attenuation of peak ground velocity with

rupture distance, according to Halldorsson et al. (2011) (Reprinted from Halldorsson et al. (2011). Copyright © 2011 American Society of Civil Engineers)

parameter of the fault rupture process that greatly affects strong ground motion characteristics. In fact, the rise time (and therefore the pulse period) is related to one of the characteristic corner frequencies of the source spectrum (i.e., the spectrum of seismic waves radiated by the earthquake source before these are modified by propagation path and site effects).

On the other hand, the peak value of the near-fault velocity records appears to be a fairly stable parameter. A value of 100 cm/s effectively represents peak ground velocities within a few kilometers from the causative fault regardless of the earthquake magnitude (Mavroeidis and Papageorgiou 2003). This observation is in good agreement with the typical slip velocity value of 100 cm/s frequently considered by seismologists. As indicated by Mavroeidis et al. (2004), there are solid physical reasons that explain the stability of the velocity amplitude close to the fault. More recently, Halldorsson et al. (2011) proposed the following attenuation relationship for peak ground velocity (PGV) with rupture distance (R) (Fig. 7b):

$$\log PGV = 2.040 - 0.032 R \quad (4)$$

Once PGV has been determined using Eq. 4, parameter A that controls the amplitude of the

synthetic velocity pulse can be defined by considering that $A \sim (0.85-1.00)$ PGV. Other investigators have also proposed attenuation relationships for PGV (e.g., Bray and Rodriguez-Marek 2004, among others).

Finally, parameter γ varies from a value slightly larger than 1 up to a maximum value of 3, while the phase angle (ν) varies from 0° to 360° . Halldorsson et al. (2011) have provided the probability density functions of γ and ν , assuming that these two parameters are normally distributed.

Simplified Methodology for Generating Broadband Near-Fault Ground Motions Proposed by Mavroeidis and Papageorgiou (2003)

As explained by Mavroeidis and Papageorgiou (2003), the mathematical expression of Eq. 1 replicates accurately the intermediate-to-long-period (“coherent”) features of the near-fault ground motions. The high-frequency components that are “incoherent” cannot be simulated using simplified mathematical models (see, e.g., acceleration time histories and the short-period range of response spectra in Fig. 6).

A simplified methodology for generating realistic, broadband, near-fault ground motions that are adequate for engineering analysis and design was proposed by Mavroeidis and Papageorgiou (2003).

Based on this technique, the coherent (long-period) ground motion component is simulated using the mathematical model of Eq. 1, while the incoherent (high-frequency) seismic radiation is synthesized using the specific barrier model (Papageorgiou and Aki 1983a, b) in the context of the stochastic modeling approach. The specific barrier model is a physical model of the seismic source that applies both to the “near-fault” and “far-field” regions, allowing for consistent ground motion simulations over the entire frequency range and for all distances of engineering interest. The specific barrier model has been calibrated to shallow crustal earthquakes of three different tectonic regions: interplate, intra-plate, and extensional regimes (Halldorsson and Papageorgiou 2005).

This simplified methodology has been applied to hypothetical and actual earthquakes (e.g., Mavroeidis and Papageorgiou 2003; Halldorsson et al. 2011) and is demonstrated in this entry through the case study of the 1971 San Fernando earthquake (Mavroeidis 2004). According to Heaton (1982), the 1971 San Fernando earthquake with M_W 6.6 may have been a double seismic event that occurred on two subparallel thrust faults, the Sierra Madre and San Fernando Faults, as indicated in Fig. 8a. The slip distribution on the causative faults, inferred by inversion of teleseismic, strong motion, and geodetic data, is illustrated in Fig. 8b. The damage that the Olive View Hospital sustained during the earthquake has been attributed to the destructive potential of near-fault ground motions on flexible structures (Bertero et al. 1978). No strong motion instruments were installed in the immediate vicinity of the hospital building. However, there are indications that the ground motion that the Olive View Hospital sustained was equivalent or greater than the ground motion recorded at the nearby Pacoima Dam (PCD) station.

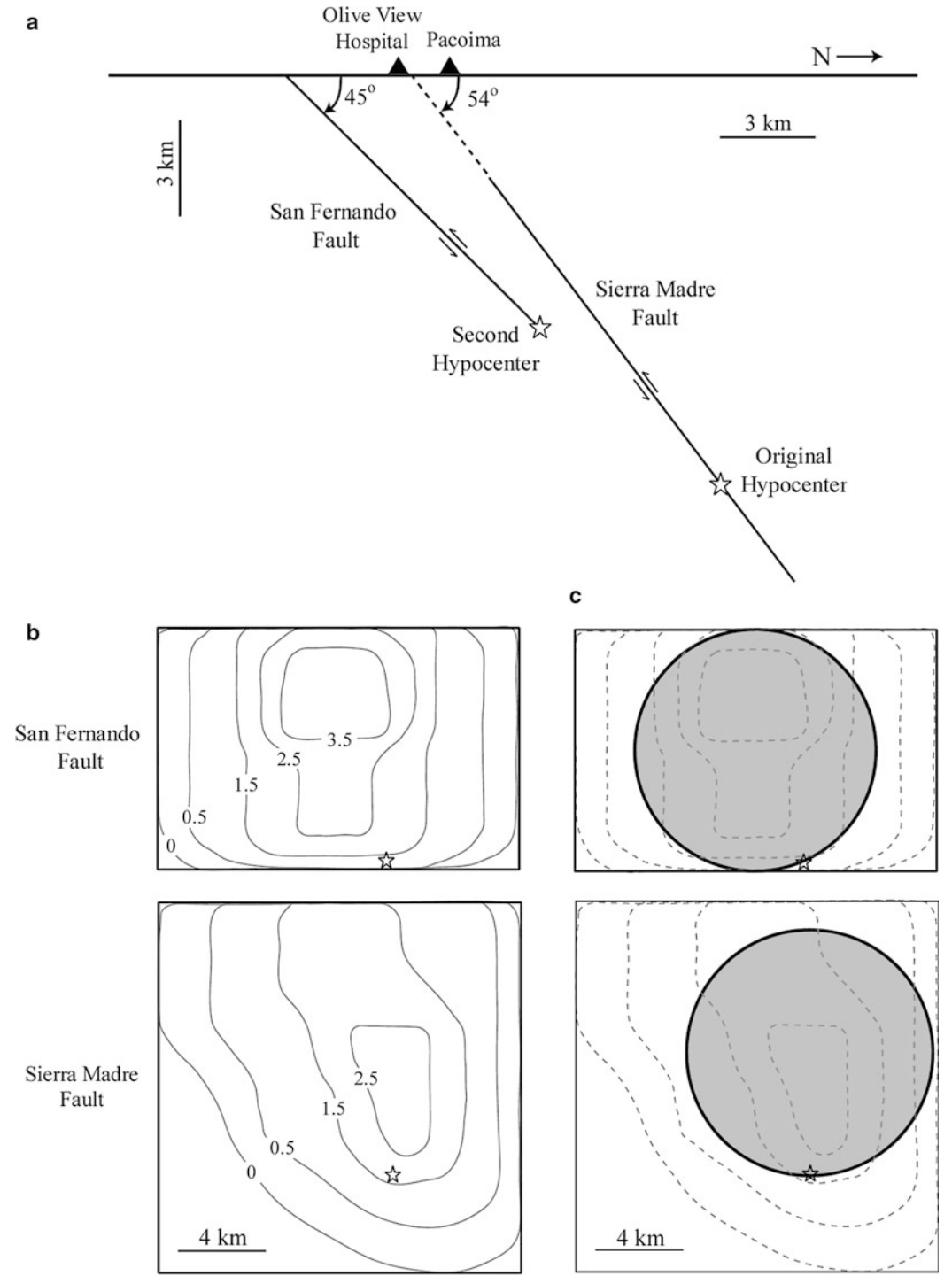
Broadband synthetic time histories are generated at the location of the Olive View Hospital for the fault-station geometry of Fig. 8a. The mathematical model of Eq. 1 is first employed to generate the coherent component of the ground motion at the Olive View Hospital. The values for the input parameters A , f_p , γ , and ν are those inferred by fitting the mathematical model of

Eq. 1 to the ground motion time histories and response spectra of the PCD record (see Mavroeidis and Papageorgiou 2003). For the synthesis of the incoherent seismic radiation at the location of the Olive View Hospital, the specific barrier model of Papageorgiou and Aki (1983a) is utilized. The selected parameters for the specific barrier model are consistent with the values inferred by Papageorgiou and Aki (1983b) for the 1971 San Fernando earthquake. The model consists of two subevents (Fig. 8c), in agreement with the two distinct slip patches of similar size inferred for this event by Heaton (1982) (Fig. 8b). The site characterization at the Olive View Hospital is assumed to be NEHRP site class D consistent with available information.

Figure 9a illustrates the synthetic ground motions (strike-normal component) at the Olive View Hospital. The top and middle panels display the incoherent and coherent ground motion components, while the bottom panels show the superposition of the above two components. For comparison purposes, the ground motion recorded at the nearby PCD station due to the 1971 San Fernando earthquake is also shown in Fig. 9b. The overall agreement between the synthetic ground motions at the Olive View Hospital and the recorded ground motions at PCD station is very good. It is evident that acceleration amplitudes larger than those recorded at the PCD station characterize the synthetic accelerogram at the Olive View Hospital. On the other hand, the corresponding velocity and displacement time histories are very similar. These differences in acceleration amplitudes may be attributed to the different site conditions at the locations of the Olive View Hospital (NEHRP site class D) and Pacoima Dam (NEHRP site class B; rock good enough to serve as the foundation of a concrete dam).

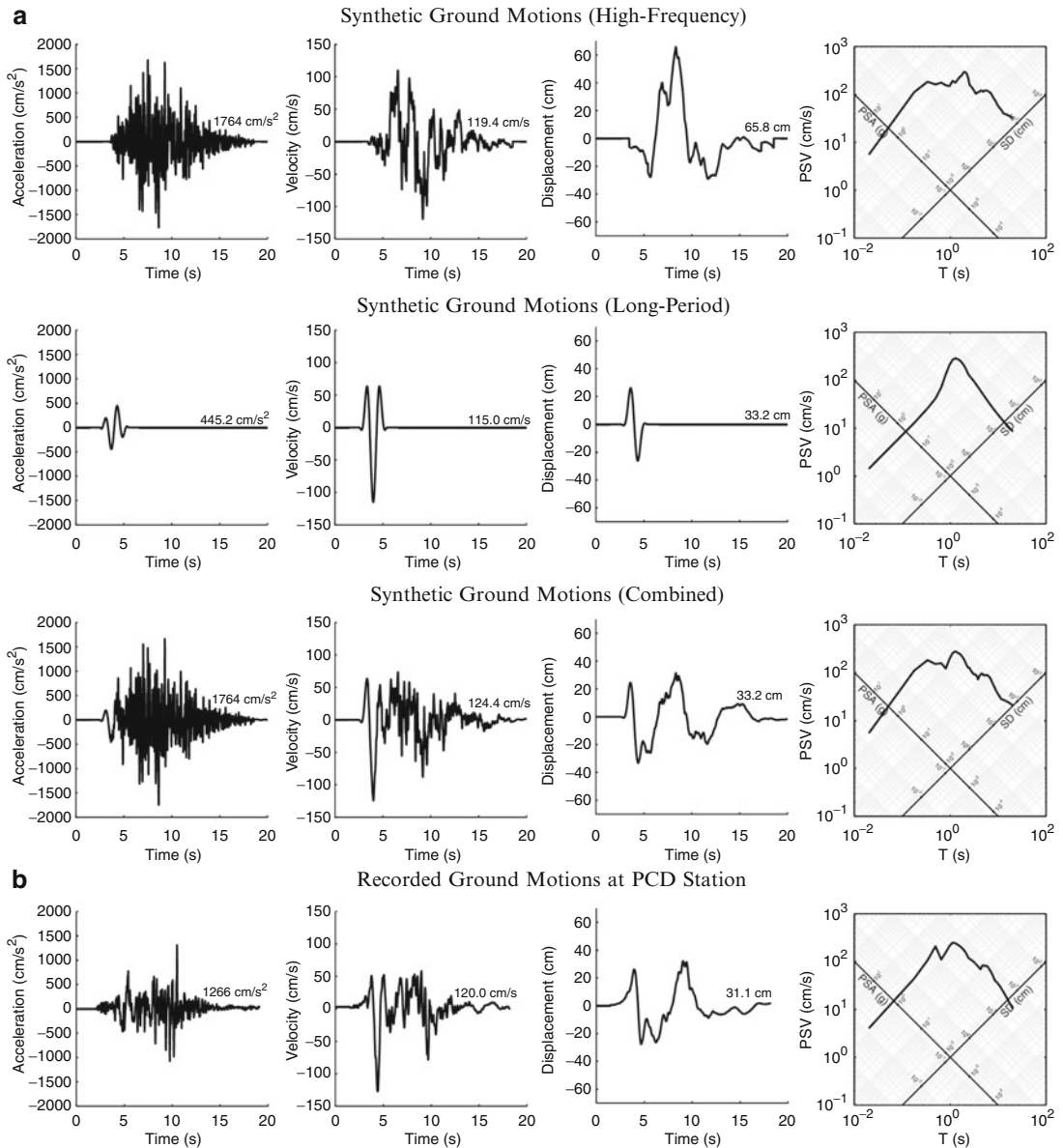
Response Spectra, Strength Reduction Factors, and Damping Coefficients for Near-Fault Ground Motions

In this section, the primary characteristics of near-fault ground motion response spectra are discussed, and recommendations are made for



Seismic Actions Due to Near-Fault Ground Motion, Fig. 8 (a) Cross-sectional view of the causative faults of the 1971 San Fernando, California, earthquake (Heaton 1982), (b) slip distribution in meters for the two

subparallel thrust faults of Fig. 8a (Heaton 1982), and (c) subevents of the specific barrier model represented by two circular cracks (ω^2 -model) (Reprinted from Mavroeidis (2004). Copyright © 2004 G. P. Mavroeidis)



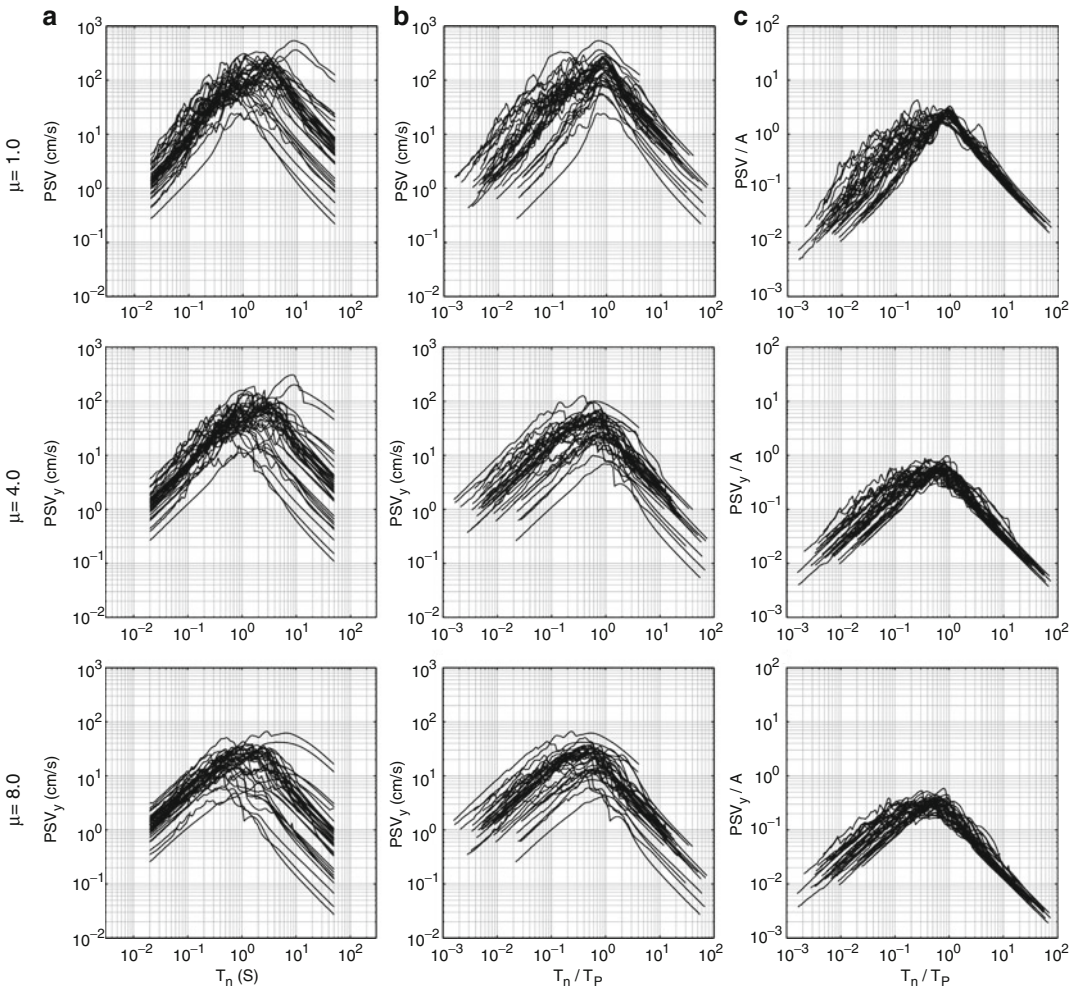
Seismic Actions Due to Near-Fault Ground Motion, Fig. 9 (a) Synthesis of near-fault ground motions at the location of the Olive View Hospital for the fault-station geometry illustrated in Fig. 8a; the 5% damped elastic

response spectra are also shown. (b) Actual ground motions recorded at the nearby Pacoima Dam (PCD) station (Reprinted from Mavroeidis (2004). Copyright © 2004 G. P. Mavroeidis)

design spectra, strength reduction factors and damping coefficients for analysis, and design in the near-fault region. The interested reader may find additional information in Mavroeidis et al. (2004) and Hubbard and Mavroeidis (2011).

Response Spectra

Figure 10a displays the 5% damped equal-ductility pseudo-velocity response spectra of elastic-perfectly plastic single-degree-of-freedom (SDOF) systems subjected to a large



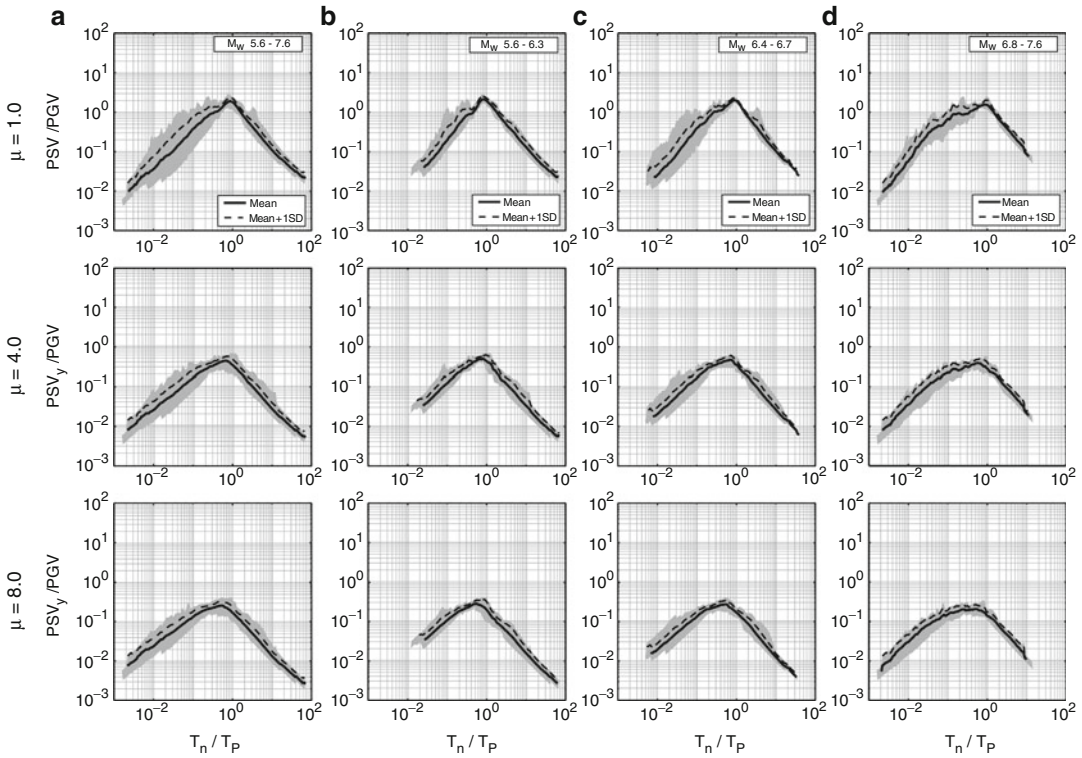
Seismic Actions Due to Near-Fault Ground Motion, Fig. 10 Standard and normalized 5% damped equal-ductility ($\mu = 1.0, 4.0, 8.0$) pseudo-velocity response spectra of elastic-perfectly plastic SDOF systems

subjected to actual near-fault ground motion records: (a) PSV versus T_n , (b) PSV versus T_n/T_P , and (c) PSV/A versus T_n/T_P (Reprinted from Mavroeidis et al. (2004). Copyright © 2004 John Wiley & Sons, Inc.)

number of actual near-fault ground motion records (Mavroeidis et al. 2004). Inspection of this figure reveals that peak spectral amplitudes of near-fault records vary significantly, especially for smaller values of the ductility factor (μ). Furthermore, the periods that correspond to peak spectral amplitudes are characterized by significant dispersion.

Figure 10b illustrates the equal-ductility pseudo-velocity response spectra of Fig. 10a with the period axis normalized with respect to the corresponding T_P values estimated by

Mavroeidis and Papageorgiou (2003). This abscissa normalization yields response spectra characterized by peak spectral amplitudes that lie within a very narrow range of the normalized period (i.e., $T_{\text{peak}}/T_P \approx 0.7\text{--}1.0$ for elastic spectra). If the ordinates of the equal-ductility pseudo-velocity response spectra are further normalized with respect to A , the normalized response spectra of Fig. 10c are obtained; these spectra are characterized by small dispersion of normalized peak spectral amplitudes. In addition, they exhibit smaller dispersion in the normalized



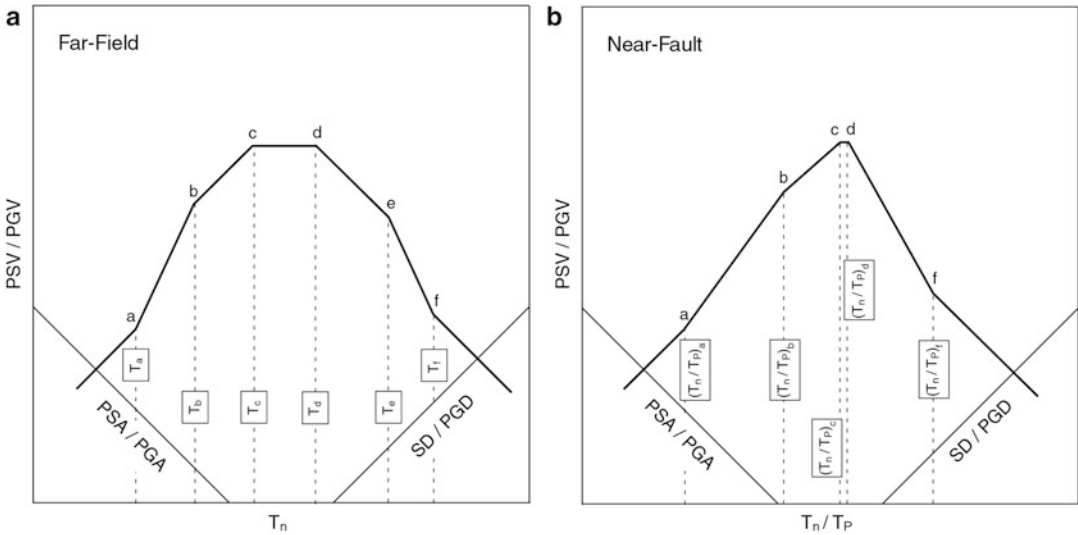
Seismic Actions Due to Near-Fault Ground Motion, Fig. 11 Normalized 5% damped equal-ductility pseudo-velocity response spectra of elastic-perfectly plastic SDOF systems: (a) all earthquakes (M_W 5.6–7.6), (b) moderate earthquakes (M_W 5.6–6.3), (c) moderate-to-large earthquakes (M_W 6.4–6.7), and (d) large earthquakes

(M_W 6.8–7.6). The solid and dashed lines represent the mean and mean-plus-one-standard-deviation pseudo-velocity response spectra, respectively. The gray region represents the range of variation of the spectral amplitudes (Reprinted from Mavroeidis et al. (2004). Copyright © 2004 John Wiley & Sons, Inc.)

long-period range and larger dispersion in the normalized high-frequency range (controlled by the coherent and incoherent ground motion components, respectively).

Thus, it may be concluded that parameters A and T_P can be used to normalize the response spectra of actual near-fault records. This development facilitates the systematic investigation of the response spectrum characteristics of the SDOF system subjected to near-fault ground motions. It should be noted that the normalization of the ordinates of the equal-ductility pseudo-velocity response spectra of Fig. 10a with respect to PGV values yields normalized response spectra very similar to those illustrated in Fig. 10c. This is anticipated because parameter A effectively approximates PGV.

In order to investigate the effect of the earthquake magnitude on the normalized response spectra of Fig. 11a, the seismic events have been grouped into three categories: moderate (M_W 5.6–6.3), moderate-to-large (M_W 6.4–6.7), and large (M_W 6.8–7.6) earthquakes. The normalized equal-ductility pseudo-velocity response spectra of these three categories are displayed in Fig. 11b, c, d, respectively. It becomes evident that, for smaller values of T_n/T_P , the normalized spectral amplitudes increase with earthquake magnitude. However, for larger T_n/T_P values, the normalized response spectra appear to exhibit a uniform behavior regardless of the variation in earthquake magnitude. As a consequence, the normalized response spectra of large earthquakes exhibit flatter shapes around their peaks than the



Seismic Actions Due to Near-Fault Ground Motion, Fig. 12 Schematic illustrations of idealized response spectra in four-way logarithmic plots for far-field and

normalized response spectra of moderate earthquakes.

The ensemble of the normalized elastic response spectra illustrated in the first panel of Fig. 11 can be utilized to derive normalized elastic design spectra for moderate, moderate-to-large, and large earthquakes, as well as for the entire set of seismic events considered by Mavroeidis et al. (2004). The solid and dashed lines in the top panel of Fig. 11 represent the mean and mean-plus-one-standard-deviation 5% damped normalized elastic response spectra. These average elastic response spectra can be used to derive normalized elastic design spectra for two different nonexceedance probability levels.

Figure 12a displays a sketch of the standard idealized elastic design spectrum derived from far-field ground motion records. This standard design spectrum has been used for many decades since it was first introduced in engineering practice. The acceleration-, velocity-, and displacement-sensitive regions of this design spectrum can readily be identified in Fig. 12a. On the other hand, the normalized response spectra of the near-fault ground motion records (see Fig. 11) can be approximated by the sequence of linear segments displayed in Fig. 12b. The values of the characteristic normalized periods

near-fault ground motion records (Reprinted from Mavroeidis et al. (2004). Copyright © 2004 John Wiley & Sons, Inc.)

$[(T_n/T_P)_a, (T_n/T_P)_b, (T_n/T_P)_c, (T_n/T_P)_d, \text{ and } (T_n/T_P)_f]$ are provided in Mavroeidis et al. (2004).

Strength Reduction Factors

The earliest and perhaps the simplest recommendation of a procedure to construct inelastic spectra from elastic spectra using ductility-dependent strength reduction factors (R_y) is based on the work of Veletsos and Newmark (1960). These results were further developed by Newmark and Hall (1982) based on a suite of far-field ground motion records. Mavroeidis et al. (2004) checked the validity of the reduction factors proposed by Newmark and Hall (1982) for the response spectra of near-fault ground motion records by normalizing the period intervals of the R_y design equations as follows:

$$R_y = \begin{cases} 1, & \frac{T_n}{T_P} < \left(\frac{T_n}{T_P}\right)_a \\ \sqrt{2\mu-1}, & \left(\frac{T_n}{T_P}\right)_b < \frac{T_n}{T_P} < \left(\frac{T_n}{T_P}\right)_{c'} \\ \mu, & \frac{T_n}{T_P} > \left(\frac{T_n}{T_P}\right)_c \end{cases} \quad (5)$$

where $\left(\frac{T_n}{T_P}\right)_{c'} = \frac{\sqrt{2\mu-1}}{\mu} \left(\frac{T_n}{T_P}\right)_c$. The characteristic values of $(T_n/T_P)_a$, $(T_n/T_P)_b$, and $(T_n/T_P)_c$ are

associated with the normalized elastic design spectrum for near-fault ground motions (see Fig. 12b) proposed by Mavroeidis et al. (2004).

Figure 13 compares the computed values of R_y from the mean elastic and inelastic 5% damped normalized response spectra of Fig. 11 with the R_y values obtained from Eq. 5. The agreement between the two sets of curves is very good over the entire period range, for all specified ductility factors, and for all earthquake magnitude categories. Figure 13 demonstrates that the Veletsos-Newmark-Hall design equations can be used for near-fault ground motions as well, provided that normalized response spectra are utilized and appropriate values of $(T_n/T_P)_a$, $(T_n/T_P)_b$, and $(T_n/T_P)_c$ are selected.

Damping Coefficients

Damping coefficients are frequently used in earthquake engineering as a simple way to adjust the pseudo-acceleration or displacement response spectra associated with a viscous damping ratio of 5% to the higher values of viscous damping needed for design of structures equipped with base isolation and/or supplemental energy dissipation devices. Damping coefficients are also frequently used for predicting the maximum displacement demands of an inelastic structure from the maximum displacement demands of its equivalent linear system characterized by a longer natural period and a higher viscous damping ratio.

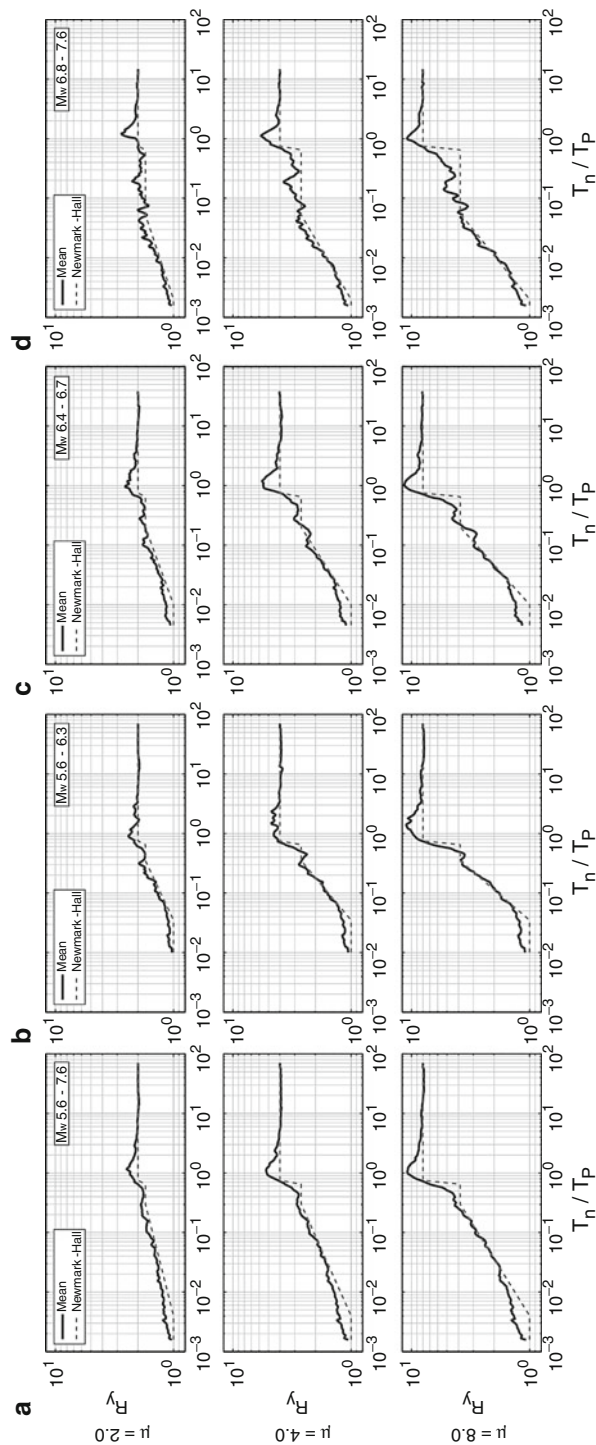
Damping coefficients (B) are defined as $B(T, \beta) = PSA(T, \beta = 5\%) / PSA(T, \beta)$, where T is the elastic period of vibration of the structure, β is the viscous damping ratio, and PSA are the ordinates of the pseudo-acceleration response spectra for particular values of T and β . Damping coefficients (B) are also known as “damping adjustment factors.” The reciprocal of B is often used in the literature and referred to as “damping correction factor,” “damping reduction factor,” “spectral scaling factor,” or “damping modification factor.”

Hubbard and Mavroeidis (2011) calculated damping coefficients for the SDOF system subjected to near-fault pulse-like ground motions for a large range of periods and damping levels.

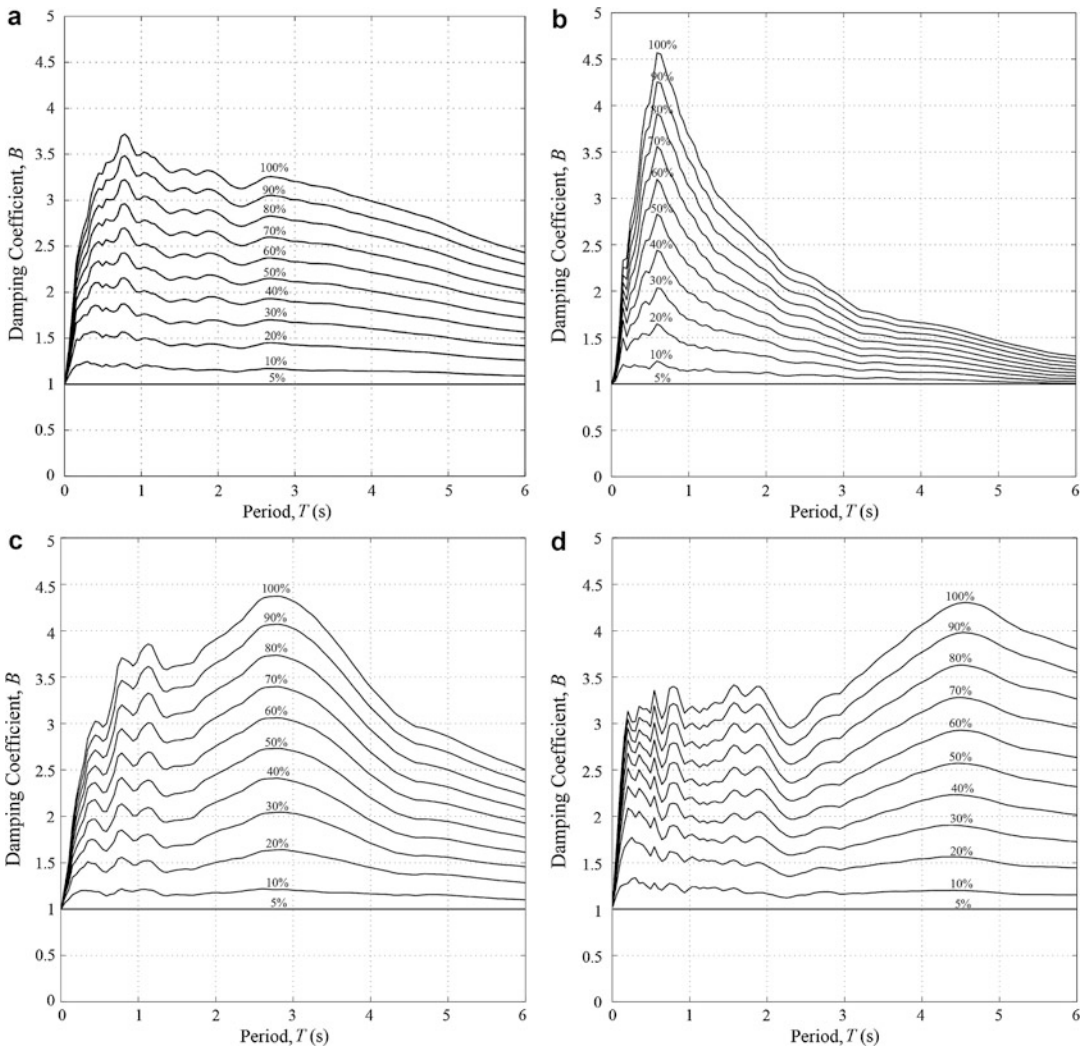
The results indicated that damping coefficients proposed in design codes and previous studies, based primarily on far-field ground motion records, tend to not be conservative for near-fault pulse-like seismic excitations. Figure 14a illustrates the relationships between damping coefficient and period that were established for viscous damping ratios in the range of 5–100%. These damping coefficient curves were generated using the definition of B factors and the median pseudo-acceleration response spectra for various levels of damping.

In order to investigate the effect of earthquake magnitude on damping coefficients using near-fault records, the seismic events were again grouped into three categories labeled as moderate (M_W 5.6–6.3), moderate-to-large (M_W 6.4–6.7), and large (M_W 6.8–7.6) earthquakes. The variation of damping coefficient with period for these three categories is displayed in Figs. 14b, c, d for viscous damping ratios in the range of 5–100%. While the B factors for all three earthquake magnitude categories attain approximately the same peak values for a given damping ratio, the period range over which these peak values occur clearly depends on earthquake magnitude. In addition, the damping coefficient curves of Fig. 14a derived from the entire ground motion ensemble smooth out the effect of earthquake magnitude and therefore do not capture the particular features of the damping coefficient plots illustrated in Figs. 14b, c, d.

Figure 15 indicates that the normalization of the period axis of the B plots with respect to T_P yields damping coefficient curves that show a much stronger resemblance to each other. More specifically, the normalized damping coefficient curves for all groups of records attain comparable peak values for a given damping ratio. These maximum values tend to be slightly closer than the peaks observed on the non-normalized damping coefficient curves displayed in Fig. 14. In addition, the normalized periods over which these peak values occur coincide at a value slightly lower than 1.0 on the normalized period axis, a statistic that had previously varied greatly by earthquake magnitude. The B curves illustrated in Fig. 15 may allow



Seismic Actions Due to Near-Fault Ground Motion, Fig. 13 Comparison of mean values of the strength reduction factor with the Veletsos-Newmark-Hall design equations for: (a) all earthquakes (M_w 5.6–7.6), (b) moderate earthquakes (M_w 5.6–6.3), (c) moderate-to-large earthquakes (M_w 6.4–6.7), and (d) large earthquakes (M_w 6.8–7.6); $\mu = 2.0$, 4.0, 8.0, and $\zeta = 5\%$ (Reprinted from Mavroeidis et al. (2004). Copyright © 2004 John Wiley & Sons, Inc.)



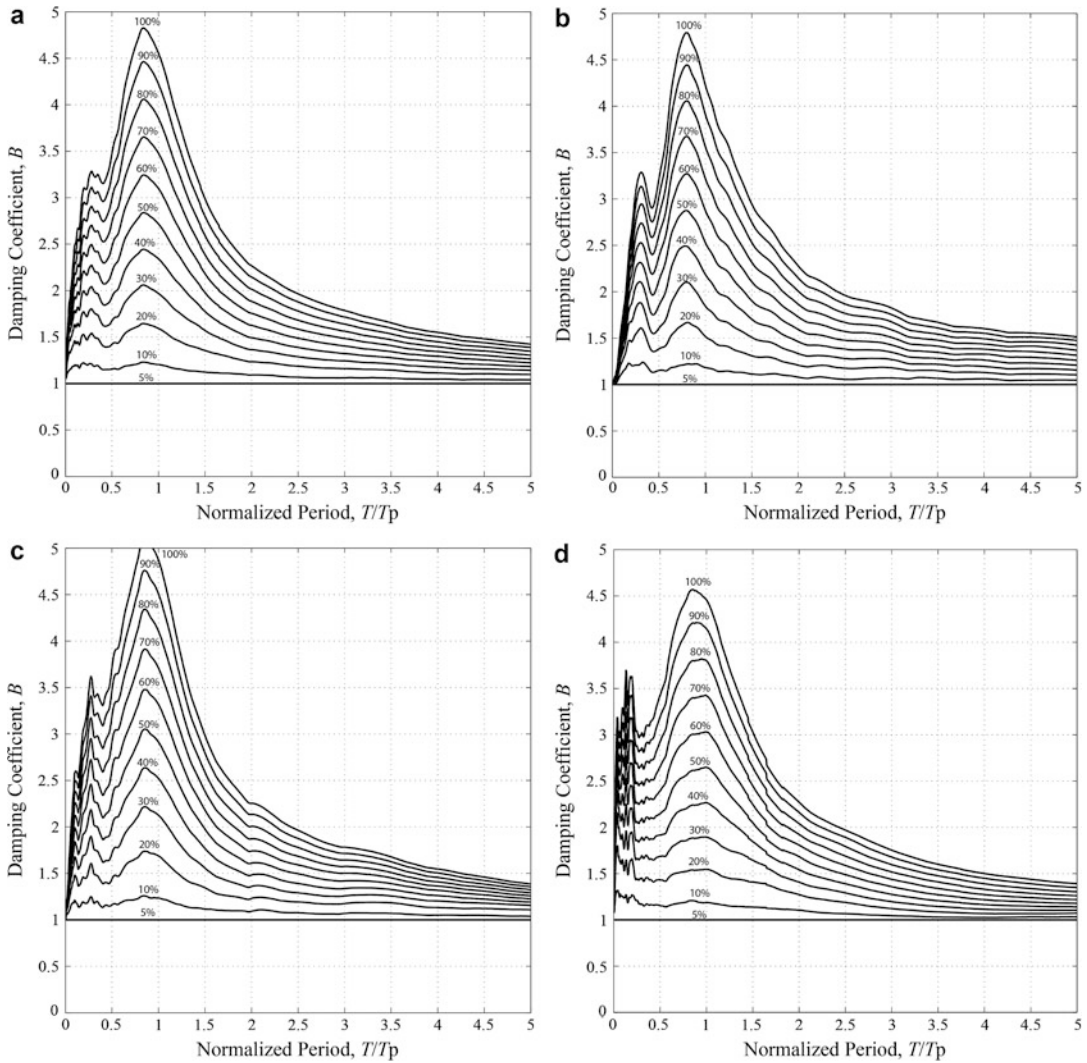
Seismic Actions Due to Near-Fault Ground Motion, Fig. 14 Calculated damping coefficients for near-fault ground motion records: (a) all earthquakes (M_W 5.6–7.6), (b) moderate earthquakes (M_W 5.6–6.3), (c)

moderate-to-large earthquakes (M_W 6.4–6.7), and (d) large earthquakes (M_W 6.8–7.6) (Reprinted from Hubbard and Mavroeidis (2011). Copyright © 2011 Elsevier B.V.)

for a single set of empirical equations to represent near-fault damping coefficients that are normalized by T_p .

An empirical equation was developed by Hubbard and Mavroeidis (2011) to fit the main behavior of damping coefficients observed in the set containing all records (Fig. 15a). In order to effectively model these B curves, two equations were needed to describe different ranges of the

damping ratio. Equation 6 is designed to be representative of the damping coefficients at normalized periods greater than ~ 0.83 , a normalized period where the peak in value seems to occur. Below this characteristic normalized period, the damping coefficients linearly reduce to one at a normalized period of zero. Damping ratios are represented as fractions instead of whole numbers within the context of Eq. 6:



Seismic Actions Due to Near-Fault Ground Motion, Fig. 15 Damping coefficients for near-fault ground motion records with the period axis normalized with T_p : (a) all earthquakes (M_w 5.6–7.6), (b) moderate

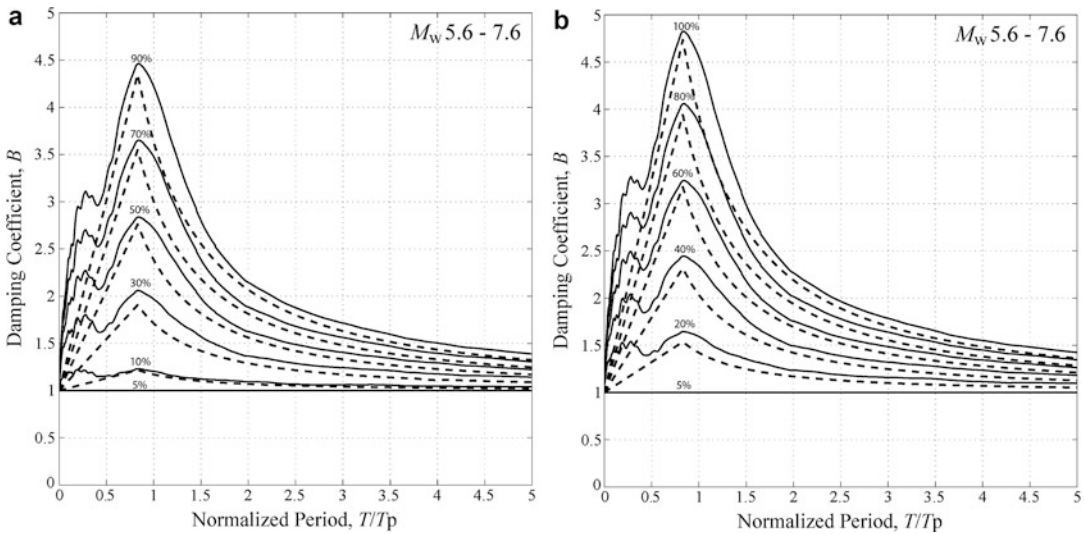
earthquakes (M_w 5.6–6.3), (c) moderate-to-large earthquakes (M_w 6.4–6.7), and (d) large earthquakes (M_w 6.8–7.6) (Reprinted from Hubbard and Mavroeidis (2011). Copyright © 2011 Elsevier B.V.)

$$B = 3.4 \frac{\beta^{1.3}}{(T/T_p)^{1.3}} + 1 \quad \text{for } 0.10 \leq \beta \leq 0.50 \quad (6a)$$

$$B = 2 \frac{(\beta + 0.3)^{1.5}}{(T/T_p)^{1.3}} + 1 \quad \text{for } 0.50 < \beta \leq 1.00 \quad (6b)$$

The empirical expression that was developed through this method is shown in Fig. 16 as a direct

comparison to the computed damping coefficients for the set containing all records. For the considered ranges of damping ratio and normalized period, Eq. 6 provides a model that is conservative without exception. It also does an adequate job of capturing the shape of the curves, ensuring that there are no damping coefficients that are greatly overconservative. In addition, the model remains conservative and captures the behavior of the damping coefficient plots of the different earthquake magnitude groupings



Seismic Actions Due to Near-Fault Ground Motion, Fig. 16 A comparison between the proposed method and the calculated damping coefficients with the period axis

of the dataset (Figs. 15b, c, d) as shown by Hubbard and Mavroeidis (2011).

Summary

This entry focuses on the description of seismic actions due to near-fault ground motions. Particular emphasis was given on synthesizing broadband near-fault ground motion time histories for earthquake engineering applications, using a simple mathematical model for the representation of the coherent ground motion component and a physical model of the seismic source for the description of the incoherent seismic radiation. In addition, recommendations on design spectra, strength reduction factors, and damping coefficients were made for engineering analysis and design in the near-fault region. This included the normalization of the period axis with respect to the period of the ground velocity pulses. The pulse period is controlled by the rise time on the fault plane and scales directly with earthquake magnitude.

Acknowledgments The material presented in this entry is primarily based on previous articles published by the author and Professor Apostolos S. Papageorgiou.

normalized with T_p (Reprinted from Hubbard and Mavroeidis (2011). Copyright © 2011 Elsevier B.V.)

The author is grateful to Professor Papageorgiou for providing him with the opportunity to work on this research, as well as for his guidance and advice.

This material is based upon work supported by the National Science Foundation (NSF) under Grant Nos. EEC-9701471, CMS-0077504, and CMMI-1032504; by the Federal Highway Administration (FHWA) under Contract DTFH61-98-C-00094; and by the US Geological Survey (USGS) under Award No. 04HQGR0029. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the funding agencies.

Cross-References

- ▶ [Earthquake Response Spectra and Design Spectra](#)
- ▶ [Engineering Characterization of Earthquake Ground Motions](#)
- ▶ [Physics-Based Ground-Motion Simulation](#)
- ▶ [Response Spectrum Analysis of Structures Subjected to Seismic Actions](#)
- ▶ [Response-Spectrum-Compatible Ground Motion Processes](#)
- ▶ [Selection of Ground Motions for Response History Analysis](#)
- ▶ [Stochastic Ground Motion Simulation](#)
- ▶ [Time History Seismic Analysis](#)

References

- Alavi P, Krawinkler H (2001) Effects of near-fault ground motion on building structures (Report). CUREE-Kajima Joint Research Program, Richmond
- Archuleta RJ (1984) A faulting model for the 1979 Imperial Valley earthquake. *J Geophys Res* 89:4559–4585
- Baker JW (2007) Quantitative classification of near-fault ground motions using wavelet analysis. *Bull Seismol Soc Am* 97:1486–1501
- Bertero VV, Mahin SA, Herrera RA (1978) Aseismic design implications of near fault San Fernando earthquake records. *Earthq Eng Struct Dyn* 6:31–42
- Bouchon M (1997) The state of stress on some faults of the San Andreas system as inferred from near-field strong motion data. *J Geophys Res* 102:11731–11744
- Bray JD, Rodriguez-Marek A (2004) Characterization of forward-directivity ground motions in the near-fault region. *Soil Dyn Earthq Eng* 24:815–828
- Halldorsson B, Papageorgiou AS (2005) Calibration of the specific barrier model to earthquakes of different tectonic regions. *Bull Seismol Soc Am* 95:1276–1300
- Halldorsson B, Mavroeidis GP, Papageorgiou AS (2011) Near-fault and far-field strong ground motion simulation for earthquake engineering applications using the specific barrier model. *J Struct Eng ASCE* 137:433–444
- Heaton TH (1982) The 1971 San Fernando earthquake: a double event? *Bull Seismol Soc Am* 72:2037–2062
- Housner GW, Trifunac MD (1967) Analysis of accelerograms – Parkfield earthquake. *Bull Seismol Soc Am* 57:1193–1220
- Hubbard DT, Mavroeidis GP (2011) Damping coefficients for near-fault ground motion response spectra. *Soil Dyn Earthq Eng* 31:401–417
- Makris N (1997) Rigidity-plasticity-viscosity: can electrorheological dampers protect base-isolated structures from near-source ground motions? *Earthq Eng Struct Dyn* 26:571–591
- Mavroeidis GP (2004) Modeling and simulation of near-fault strong ground motions for earthquake engineering applications. PhD dissertation, Department of Civil, Structural and Environmental Engineering, State University of New York at Buffalo, Buffalo
- Mavroeidis GP, Papageorgiou AS (2002) Near-source strong ground motion: characteristics and design issues. In: Proceedings of the seventh U.S. national conference on earthquake engineering, Boston
- Mavroeidis GP, Papageorgiou AS (2003) A mathematical representation of near-fault ground motions. *Bull Seismol Soc Am* 93:1099–1131
- Mavroeidis GP, Papageorgiou AS (2010) Effect of fault rupture characteristics on near-fault strong ground motions. *Bull Seismol Soc Am* 100:37–58
- Mavroeidis GP, Dong G, Papageorgiou AS (2004) Near-fault ground motions, and the response of elastic and inelastic single-degree-of-freedom (SDOF) systems. *Earthq Eng Struct Dyn* 33:1023–1049
- Newmark NM, Hall WJ (1982) Earthquake spectra and design. Earthquake Engineering Research Institute, Berkeley
- Papageorgiou AS (1997) Engineering seismology. In: Beskos DE, Anagnostopoulos SA (eds) Computer analysis and design of earthquake resistant structures. Computational Mechanics Publications, Southampton, pp 153–200
- Papageorgiou AS, Aki K (1983a) A specific barrier model for the quantitative description of inhomogeneous faulting and the prediction of strong ground motion. I. Description of the model. *Bull Seismol Soc Am* 73:693–722
- Papageorgiou AS, Aki K (1983b) A specific barrier model for the quantitative description of inhomogeneous faulting and the prediction of strong ground motion. II. Application of the model. *Bull Seismol Soc Am* 73:953–978
- Sasani M, Bertero VV (2000) Importance of severe pulse-type ground motions in performance-based engineering: historical and critical review. In: Proceedings of the twelfth world conference on earthquake engineering, Auckland
- Somerville P (2003) Magnitude scaling of the near fault rupture directivity pulse. *Phys Earth Planet In* 137:201–212
- Vassiliou MF, Makris N (2011) Estimating time scales and length scales in pulse-like earthquake acceleration records with wavelet analysis. *Bull Seismol Soc Am* 101:596–618
- Veletsos AS, Newmark NM (1960) Effect of inelastic behavior on the response of simple systems to earthquake motions. In: Proceedings of the second world conference on earthquake engineering, Tokyo

Seismic Analysis of Concrete Bridges: Numerical Modeling

Andreas Kappos
Department of Civil Engineering, City
University London, London, UK

Synonyms

Concrete bridges; Modal analysis; Nonlinear analysis; Seismic loading

Introduction

Bridges are deceptively simple systems, since they are typically single-storey structures

wherein the horizontal members (the deck) can often be modeled either as a continuous beam or as a series of simply supported beams. In fact, the continuous beam might be a valid approximation in seismic analysis of bridges if a “spine” model is adopted; in such a model the bridge deck is simulated using 3D beam elements with 6 degrees of freedom (DOFs) at each node, located at the centroid of the cross section. It is worth noting that in gravity load analysis the geometric complexity of the deck is usually represented in the computer model to greater detail, as compared to that used for estimating the seismic response. On the other hand, bridges present peculiarities that are not commonly encountered (or are far less important) in buildings, such as the modeling of bearings, shear keys, and expansion joints, as well as the modeling of soil-structure interaction at all bridge supports, including those at the abutments which can play a major role in some cases.

In the sections that follow, modeling of the various bridge components is first addressed (section “[Modeling of Bridge Components](#)”), followed by an overview of methods currently used for seismic analysis of bridges (section “[Bridge Analysis Methods](#)”), which is a topic that is also addressed in other articles of the encyclopedia; section “[Bridge Analysis Methods](#)” also includes the presentation of a case study that illustrates the modeling and analysis procedures described in the previous sections, while both sections “[Modeling of Bridge Components](#)” and “[Bridge Analysis Methods](#)” include specific modeling examples and selected results. Finally some concluding remarks are provided in section “[Summary and Concluding Remarks](#).” Methods of seismic analysis are presented in various parts of the encyclopedia (e.g., article by Vayas and Iliopoulos (2014) focusing on modeling of steel and composite bridges). Hence, the focus herein is on analysis of concrete bridges, wherein some specific issues arise as discussed in the following; of course, several of the described models and techniques are also applicable to bridges made of other materials.

Modeling concrete bridges for seismic design purposes should always take into account the

intended plastic mechanism, which, contrary to buildings, involves primarily yielding in the vertical members, i.e., the piers, while seismic energy dissipation can also take place in the bearings; in a small number of (important) bridges, supplementary damping devices (such fluid or friction dampers) are provided and have to be accounted for in the analysis. In practical design applications bridges are analyzed in the elastic range, and any inelasticity (material nonlinearity) effects, wherever entering of some members in the inelastic range under the design earthquake is allowed, are accounted for by simply reducing the design response spectrum by a “behavior” (or force reduction) factor. However, special types of bridges have been analyzed using advanced inelastic analysis tools (typically as a verification of an initial design based on the results of equivalent elastic analysis), and in general, inelastic analysis methods are gaining ground in recent years. Therefore, an attempt is made herein to present some basic concepts and models suitable for inelastic analysis, with emphasis on those that are better suited for practical application.

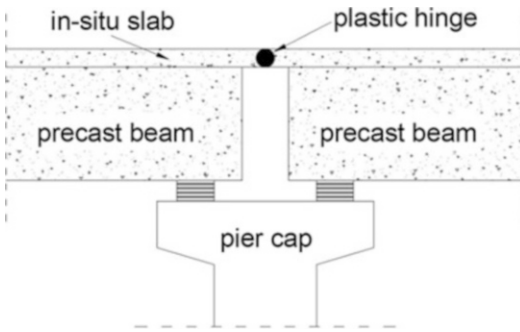
Modeling of Bridge Components

Deck

In concrete bridges the deck can have various forms, i.e., solid slab, voided slab, beams (usually precast I-beams) with cast in situ top slab, box girder (single-cell or multicell), and other, less common, ones. The material used is typically prestressed concrete, except for some short span bridges with slab-type deck, where ordinary reinforced concrete can be used.

Very important, both for the seismic behavior of the bridge and for its modeling, is the type of pier to deck connection. There are three basic options in this respect:

- **Monolithic connection:** Very common in slab bridges and box-girder bridges, especially when the cantilever method of construction is used in the latter.



Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 1 Pier region of a bridge with deck consisting of precast posttensioned beams and cast in situ slab continuous over the piers

- **Bearing connection:** The deck rests on the piers through two or more bearings, which in the case of modern concrete bridges are typically elastomeric bearings; this is the usual type of connection in the case of beam-type decks and also in box-girder bridges when the incremental launching method of construction is used.
- **Mixed type of connection:** Some piers (typically the taller ones) are monolithically connected to the deck, whereas others (typically the squat ones) have bearing connections. This is a fairly new solution, which has advantages when bridge configurations with substantially unequal pier heights have to be used (e.g., in ravine bridges).

In the case of bearing connections (very common in older concrete bridges), it is clear that the deck does not carry moments due to lateral (seismic) loading nor forms part of the energy dissipation mechanism of the bridge. An exception is the case of continuity slabs, i.e., the parts of a top slab in beam-type decks that continue over the piers (whereas beams terminate on each side of the pier), hence providing a continuous deck (Fig. 1). For these regions modern codes like Eurocode 8-2 (CEN 2005a) allow formation of plastic hinges (in bending about the transverse axis, see Fig. 1). When the focus of the analysis is on the response to the design earthquake and inelastic behavior is allowed and anticipated,

actual hinges can be introduced beforehand at these locations.

In the usual case that the deck remains elastic under the design seismic action, it can be modeled with elastic elements of any appropriate type: beam-column elements forming frames or grillages or even shell elements. For seismic analysis the recommended approach is the use of 3D beam columns (spine model), which is the simplest option and generally an adequate one for the purposes of this analysis. Relevant guidelines like those of ACI 341 (2014) and FHWA (2006) recommend four to five elements per span, but usually more elements are used in practice, since these are elastic members and do not noticeably affect the computational demands. Besides, prestressed concrete members like beams and box girders often have cross sections that vary along the span (thicker webs of beams and box girders toward the piers, where shear forces are maximum); hence, the 3D beam elements should be arranged in such a way that they properly reproduce this gradual change in geometry. Moreover, since masses required for the dynamic analysis of the bridge are typically lumped at the nodes of the model (even when uniformly distributed masses are automatically calculated from the geometry of the elements), use of a sufficient number of elements leads to a more accurate representation of the mass distribution in the bridge and hence of its dynamic characteristics.

In the case of *monolithic* connections, deck-to-pier joints carry bending moments due to seismic loading, and in principle, parts of the deck may become inelastic. However, in most cases the strength of the deck that is governed by the substantial gravity loading on the bridge plus traffic loads (very high in railway bridges) is clearly higher than that of the pier that is typically governed by seismic moments in medium to high seismicity areas; hence, the deck remains in the elastic range and the previous comments apply. In this case, accurate modeling of box girders requires accounting for shear lag (nonuniform distribution of the longitudinal stress across the flange width due to the shear deformations within the flange); the FHWA (2006) manual specifies that the flexural stiffness of the superstructure taken

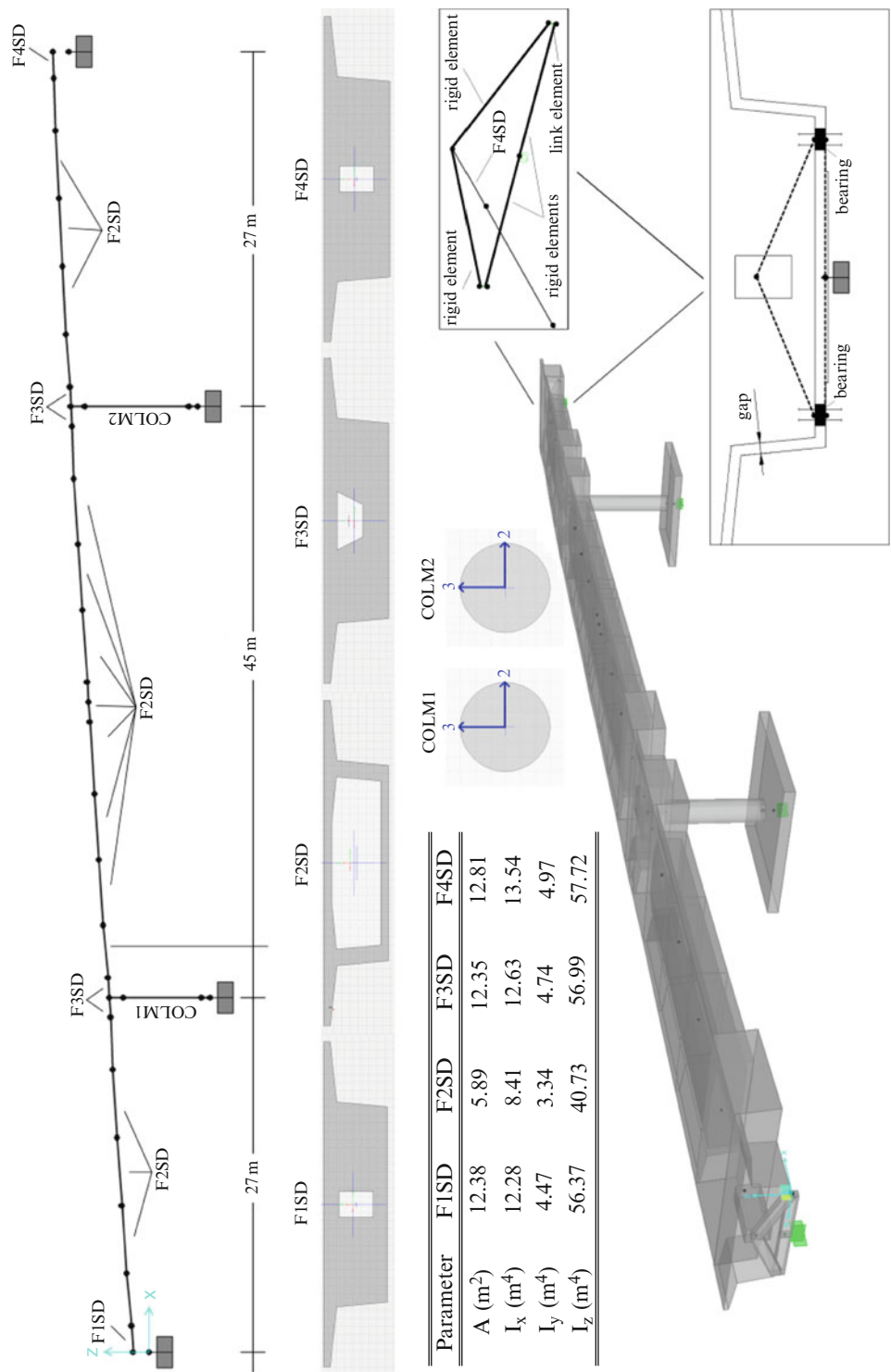
about a transverse axis should be reduced near piers when there is moment transfer between the superstructure (deck) and the pier, without providing values for this reduction. According to the ATC-32 report (ATC 1996), stiffness in these regions is based on an effective width that should be no greater than the width of the column plus twice the cap beam depth. If this width is practically equal to the entire width of the superstructure, no reduction in stiffness due to shear lag is required in the model.

An example is shown in Fig. 2 depicting the spine model of an overpass bridge (Kappos et al. 2013a) with monolithic pier to deck connections and free sliding connections at the abutments. It is seen that, depending on the length of the span, up to 11 3D beams have been used. The thickening of the webs of the box-girder section toward the end of the spans is properly modeled in the software used, SAP 2000 (CSI 2011). Figure 2 also conveys an idea of what the model would look like in case shell elements were used. As shown in a study by Kappos et al. (2002), it is possible to achieve a good match of the dynamic characteristics of a bridge modeled using shell elements, using a simple spine model; the match was particularly good (differences in significant natural periods between 4 % and 11 %) when reduced flexural rigidity was specified for the 3D beam elements close to the monolithic connection. That study also confirmed that it is not necessary to use too many elements for the deck; in fact a model involving 488 3D beam elements predicted almost identical dynamic characteristics (periods and mode shapes) as another one using only 77 elements (8 per span).

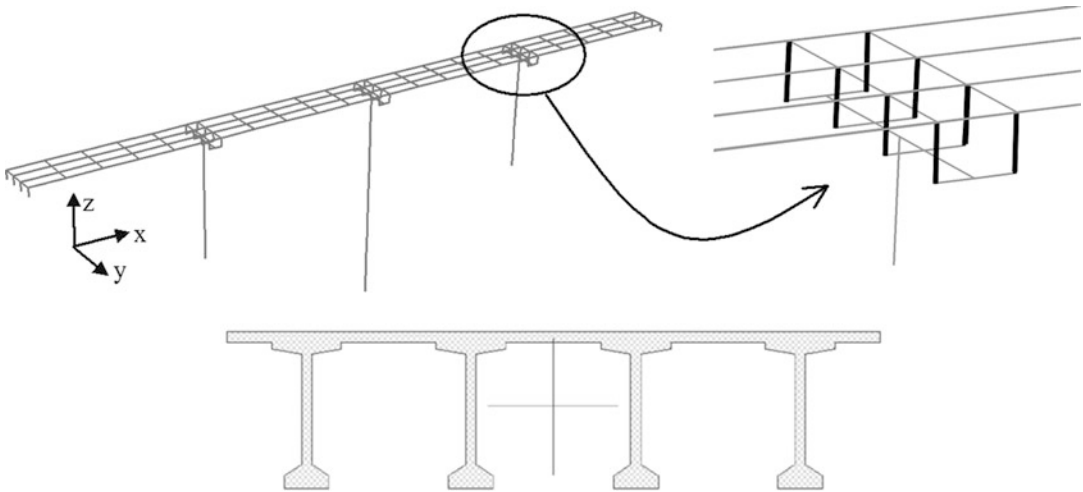
The value of *flexural rigidity* (EI) to be assigned to the elements used for modeling the deck depends on the material used. In the usual case of prestressed concrete, it is allowable to assume negligible cracking and use the value for the gross cross section (EI_g). In ordinary (non-prestressed) reinforced concrete decks, EI should account for cracking effects. Recommended values in Caltrans (2013) are 50–75 % of EI_g ; the lower bound represents lightly reinforced sections and the upper bound represents heavily reinforced sections.

The results of seismic analysis in the transverse direction of a bridge with box-girder deck will be influenced by the assumption made regarding the *torsional* stiffness of the deck (which is substantial, as opposed to that of “open-type” orthotropic decks like beams with top slab). It is recommended to assume 20 % of the uncracked value, based on the ratios ($10 \div 30$ %) of cracked-to-uncracked torsional stiffness estimated by Katsaras et al. (2009).

In between the aforementioned simple (spine) and complex (shell) modeling approaches for the deck is the *grillage* model, i.e., a horizontal planar system of longitudinal and transverse 3D beams, as shown in the example of Fig. 3, referring to a ravine bridge with a top slab on posttensioned I-beams supported through laminated elastomeric bearings, studied by Ntotsios et al. (2009). The longitudinal members of the grillage have the properties of the I-beams and the tributary part of the slab, the transverse elements above the piers have the properties of the actual transverse beams in the actual bridge, while the intermediate four transverse elements of the grillage represent the coupling of the longitudinal beams in the transverse direction due to the presence of the deck. Grillage models are particularly suited for slabs or orthotropic decks like that of Fig. 3 and provide a good balance between accuracy and practicability. They are not very easy to set up nor offer particular advantages over the spine model in the case of box-girder sections. An important issue here is the modeling of the torsional stiffness of the deck which cannot be estimated from the properties of the individual elements of the grillage but rather should be derived for the entire box girder and then distributed among the longitudinal members. In case the grillage model is used for solid slab decks, it is recommended to add diagonal braces to account for the interaction between longitudinal and transverse deck action due to Poisson effects, which cannot be neglected (as in the case of orthotropic decks with primary girders, see Vayas and Iliopoulos (2014) for orthotropic composite decks). More details on the input parameters required for deck modeling with grillages (flexural and torsional rigidities, shear areas)



Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 2 Finite element modeling of an overpass bridge with a box-girder superstructure monolithically connected to the piers (Kappos et al. 2013)



Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 3 Grillage model and detail of the connection between the deck and the pier (*above*) for a ravine bridge with beam with top slab deck (*below*)

can be found in Chapter 2, of the recent book by Kappos et al. (2012). In that chapter a brief discussion of some advanced topics like the effects of skewness and curvature in plan and/or in elevation on the seismic response, and the verification of deck deformation demands when this is necessary for seismic assessment, can also be found.

Piers and Their Foundations

Piers commonly used in concrete bridges are of the following types:

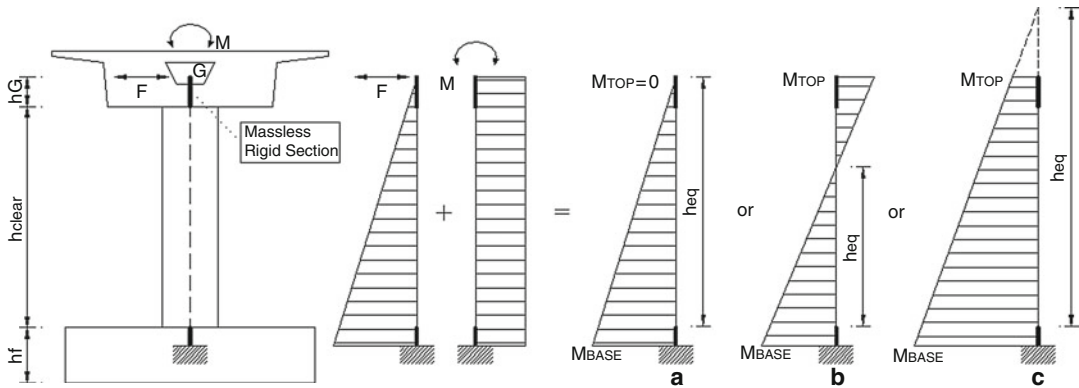
- Single columns with solid circular or (less frequently) rectangular section
- Single columns with hollow circular or rectangular section
- Multicolumn bents (frame-type piers, running in the transverse direction of the bridge)
- Wall-type piers (usually of large dimensions, especially in the transverse direction)
- Braced piers (usually V shaped)

It is noted that hollow rectangular piers are typically of large dimensions (could reach 7 m or more in the transverse direction), and their structural behavior is closer to that of walls, rather than of hollow circular columns.

Geometric Considerations

Single-column piers, whether solid or hollow, are generally modeled as “sticks” using beam-column elements; four to five elements are usually enough, unless complex geometries (e.g., flared columns) are involved. As shown in Fig. 4, the top element of the column is connected via a rigid link to the centroid G of the deck section (which is the location of the corresponding horizontal element); in most software packages there is no need for specifying a different element, but rather the top element of the column extends up to G , and its end portion is specified as a rigid offset. In the common case of columns monolithically connected to the box girder of the deck, there is no need for additional elements in the column, but when the box girder is supported on bearings (which will be the case at the abutments), it is necessary to introduce transverse elements extending from the first to the last bearing and link these to the end node of the deck model using rigid links (see bottom right of Fig. 2).

As shown in Fig. 4, the moment distribution along the height of the pier is influenced by the rotational restraint at the top, which in turn depends on the torsional rigidity of the deck. This is substantial in the case of box girders



Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 4 Pier modeling and transverse response accounting for the torsional stiffness of the deck

(as in Fig. 4) and low in the case of precast beams with top slab and other similar, “open-type,” sections.

Multicolumn bents are naturally modeled as 2D frames, again with beam-column elements, each of which has the properties of the corresponding member (column or cap beam). It is worth noting that although columns usually have circular sections, cap beams are rectangular. Rigid offsets at the element ends properly capture the effect of the finite size of the beam-column joints that are quite massive members in bridges. In case the cap beam is monolithically cast with the deck, the torsional resistance of the top of the bent is substantially higher than that of the cap beam alone; Aviram et al. (2008) recommend multiplying the torsional resistance of the cap beam by 10^2 . Clearly this does not apply in the case that the deck is bearing-supported on the cap beam. *V-shaped* piers are modeled in a similar way, but of course, vertical elements are inclined rather than upright.

There are several options available for modeling *wall-type* piers and the aspect ratio is an important parameter in this case. The simpler model is clearly the stick one, previously mentioned for the case of single-column piers. Decks are usually bearing-supported on wall-type piers (which are the preferred solution in seismic isolation designs wherein seismic energy dissipation takes place in the bearings and, whenever present, the dampers); hence, it is essential in this case to

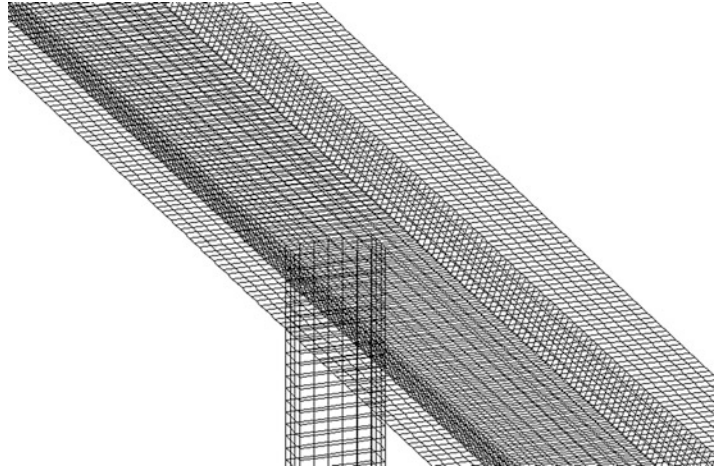
add the aforementioned horizontal rigid element at the top of the stick model. A more refined model could consist of a vertical grillage (see notes on grillage models in section “Deck”). The most refined model, feasible only in elastic analysis, is the use of shell elements; an example (not necessarily a recommended one, unless for research purposes) is shown in Fig. 5; a mesh of 18,885 shells was set up for the 3-span bridge, keeping the aspect ratio of the shells rather low at 1.2, since these elements perform best when their shape is close to square. As noted in section “Deck,” the main normal modes of that bridge were captured with reasonable accuracy using a simple spine model with only 77 beam-column elements (Kappos et al. 2002).

Stiffness Considerations

Even in elastic analysis of reinforced concrete (R/C) piers for seismic loading, it is essential to account for the effect of cracking, to make sure that displacements are not underpredicted. Practically all existing codes adopt approximate values of the pier stiffness, corresponding to yield conditions, and this stiffness is assumed as known when design seismic actions (e.g., modal forces) are estimated. These approximate values are either very rough estimates, like the $0.5 EI_g$ (50 % of uncracked section rigidity) adopted by both Eurocode 8-1 (CEN 2004) and AASHTO (2010), or slightly more sophisticated ones taking into account the level of axial loading on the pier

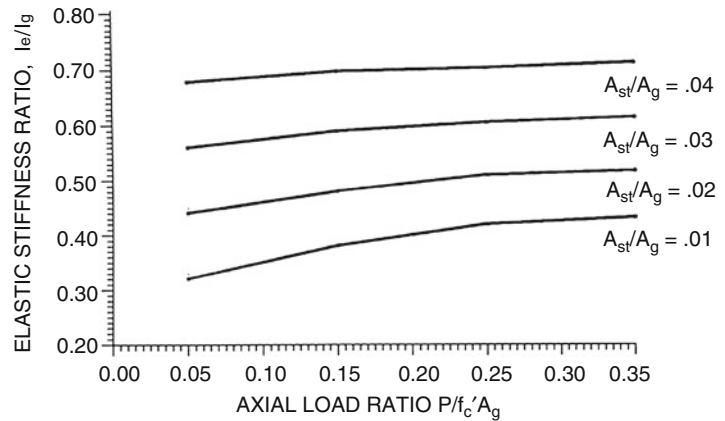
Seismic Analysis of Concrete Bridges: Numerical Modeling,

Fig. 5 Finite element (*shell*) mesh in the pier to deck connection area of a bridge with box-girder superstructure monolithically connected to hollow rectangular piers (Kappos et al. 2002)



Seismic Analysis of Concrete Bridges: Numerical Modeling,

Fig. 6 Effective stiffness of cracked reinforced concrete circular sections (AASHTO 2010)



(which, in general, is not significantly affected by seismic actions) and/or the reinforcement ratio.

Eurocode 8-2 for Seismic Design of Bridges (CEN 2005a) in its (informative) Annex C suggests the following relationship for the effective moment of inertia of R/C ductile columns:

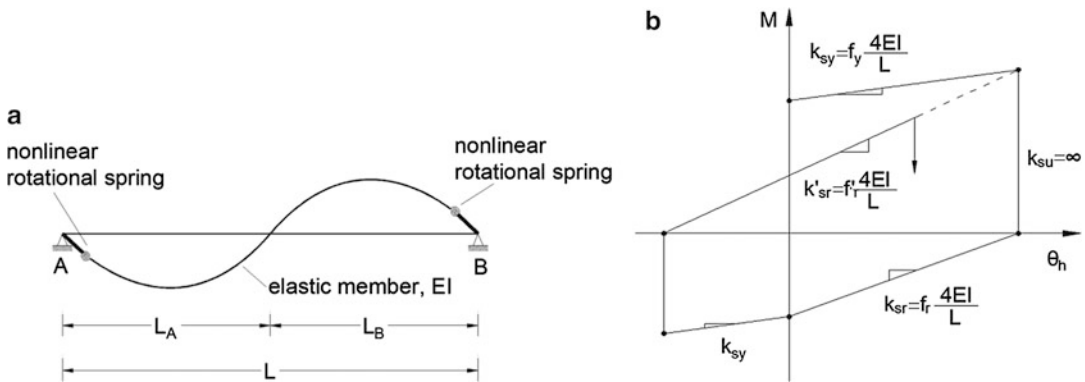
$$I_{\text{eff}} = 0.08 I_g + I_{\text{cr}} \quad (1)$$

where the cracked section inertia can be calculated as the secant value at yield (M_y is the yield moment and ϕ_y the yield curvature, E_c the concrete modulus)

$$I_{\text{cr}} = M_y / (E_c \cdot \phi_y) \quad (2)$$

Obviously, I_{cr} can only be estimated from Eq. 2 when the pier has been designed, so that both strength and yield curvature can be calculated; hence, use of these relationships is feasible only when iterative elastic analyses, or inelastic analysis, are used.

The Caltrans Seismic Design Criteria (2013) adopt the same concept as EC8-2 (secant value at yield), the only exception being that the $0.08 I_g$ term (accounting for tension stiffening effects) is not included in Eq. 1. As an alternative, the Caltrans Criteria allow the calculation of effective stiffness as a function of the axial load ratio and the pier reinforcement ratio from graphs provided by Priestley et al. (1996); diagrams like that of Fig. 6 can be directly implemented for carrying



Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 7 Lumped plasticity model (a) basic configuration; (b) hysteresis law for springs

out elastic analysis, assuming a reasonable reinforcement ratio (e.g., $A_{st}/A_g = 0.01$, which is the usual minimum reinforcement ratio), while, in principle, analysis should be repeated if the resulting reinforcement is substantially different.

Nonlinear Models for Piers

As mentioned in section “[Introduction](#),” the most common plastic mechanism on which seismic design is based is that involving inelastic response of the piers. Therefore, inelastic (material nonlinear) modeling of piers is important, not only for research but also for practical assessment purposes (which includes assessing an existing, probably substandard, bridge as well as a newly designed bridge that is important enough to warrant this additional design effort). Space limitations do not allow a detailed treatment of this important issue, and only a brief overview of the main available models will be provided herein; more detailed information on nonlinear modeling of bridge piers and several case studies can be found i.a. in the recent book by Kappos et al. (2012).

Nonlinear models for piers can be classified into three categories:

- Lumped plasticity models
- Distributed plasticity models
- Continuum models

Several subcategories can be defined for each of the above, as briefly discussed in the following.

Lumped plasticity models, also known as “point-hinge” models, are based on the simplifying assumption that all inelastic behavior takes place at the plastic hinge points that are typically located at the member ends. This concept can be materialized in different ways, the most efficient one consisting in inserting two nonlinear rotational springs at the element ends, as shown in Fig. 7a; more accurately, the springs are inserted at the ends of the rigid offsets located at the element ends to model the finite width of joints (e.g., between the cap beam and the column, in a multicolumn bent). All post-yield flexural deformation takes place in these springs, whereas the remainder of the beam-column element remains elastic throughout; it is emphasized that the flexural rigidity EI of the “elastic” element should account for cracking effects, as discussed in section “[Stiffness Considerations](#).” The (local) stiffness matrix for this lumped plasticity element, relating chord rotations at the ends to the corresponding bending moments, can be readily set up assuming a series connection between the springs and the beam, hence adding the flexibility matrices of each component, i.e.,

$$[F] = [F_e] + [F_s] = \begin{bmatrix} \frac{L}{3EI} + \frac{1}{K_{Si}} & -\frac{L}{6EI} \\ -\frac{L}{6EI} & \frac{L}{3EI} + \frac{1}{K_{Sj}} \end{bmatrix} \quad (3)$$

where K_{si} and K_{sj} are the stiffnesses of the springs at ends i and j , which are assumed for simplicity to be uncoupled (no off-diagonal terms K_{sij}) and can be different, e.g., when one end has yielded while the other is still in the pre-yield range. The main advantage of this simple model is that K_s values can be defined on the basis of any constitutive law, whether simple or complex. Figure 7b shows a typical moment versus rotation hysteresis law (with stiffness degradation) that can be applied to either spring; note that prior to exceeding the yield moment the springs are rigid ($K_s = \infty$), hence all elastic deformation takes place in the elastic member. The 2×2 flexural stiffness matrix relating end moments to chord rotations can be easily derived by inverting $[F]$ and then transforming to the 6×6 matrix including the rigid body modes and the axial deformations (axial stiffness EA is usually assumed to remain unaffected by flexural yielding).

Bridge piers, in particular those of the wall type, develop significant shear deformations subsequent to shear cracking, which may occur before or after flexural yielding. In the lumped plasticity context, shear deformations can be treated either in a simplistic way by using the stiffness matrix of a Timoshenko beam (involving GA' terms in addition to EI ones) and a rough reduction factor for GA' , or, more rigorously, by introducing additional springs at the ends representing the relationship between shear force and shear deformation; the issue of inelastic shear is discussed later on in relation to another model. Another important source of deformation in R/C piers, especially those with substandard detailing with respect to earthquake requirements, is *bond slip*, which can give rise to substantial local rotations at the member ends ("fixed end rotations"). This effect can either be modeled indirectly by decreasing the stiffness of the $M - \theta$ law (Fig. 7b) at the member ends, something that requires proper calibration, or be modeled directly by introducing additional rotational springs to the model of Fig. 7a.

Despite their crudeness, lumped plasticity models (first develop in the 1960s) remain quite popular due to their simplicity, as well as the fact

that they can relatively easily account for the effects of shear, as well as of bond slip, and are easier to calibrate against experimental results than other, more sophisticated (and complex), models.

Distributed (or spread) plasticity models, still of the beam-column type, drop the assumption of point hinges and directly account for the spread of inelasticity along the bridge member, hence leading, in principle, to more accurate results. There are several different approaches in this respect, i.e., inelastic response can be monitored at several predetermined sections of the element and the stiffness matrix be synthesized on the basis of the tangent stiffness of each such section, or variable length plastification zones be defined, typically at the member ends, assuming the rest of the element is quasi-elastic (as in lumped plasticity models). The latter option retains some of the simplicity of point-hinge models while being more rational and, in principle, accurate and will be described in the following.

A recent spread plasticity model (Mergos and Kappos 2012) accounting for inelastic response in all mechanisms (flexure, shear, bond slip) is shown in Fig. 8. The length of the plastified zones at the ends (respective rigidities EI_A and EI_B) is defined on the basis of the moment diagram of the element and the corresponding yield moments, e.g., the left zone has a length $\alpha_A \cdot L$, where

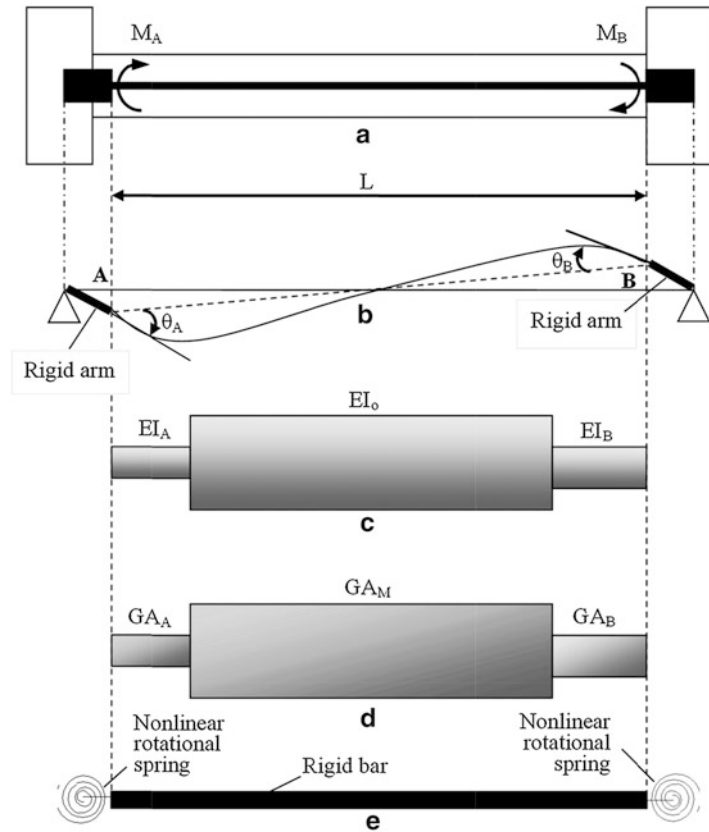
$$\alpha_A = \frac{M_A - M_{yA}}{M_A - M_B} \leq 1 \quad (4)$$

The flexural stiffness matrix can be set up using the principle of virtual work for the case of members with variable cross section. The current rigidities are calculated from the moment versus curvature relationship at each member end. Models with more than three parts have been proposed but are not deemed appropriate for practical application.

A similar procedure can be followed in the case of the shear sub-element (Fig. 8d), which represents the hysteretic shear behavior of the R/C member prior and subsequent to shear cracking, flexural yielding, and yielding of the shear reinforcement. In this case, the current shear

Seismic Analysis of Concrete Bridges: Numerical Modeling,

Fig. 8 Distributed plasticity model: (a) geometry of R/C member; (b) beam-column finite element with rigid arms; (c) flexural sub-element; (d) shear sub-element; (e) anchorage slip sub-element

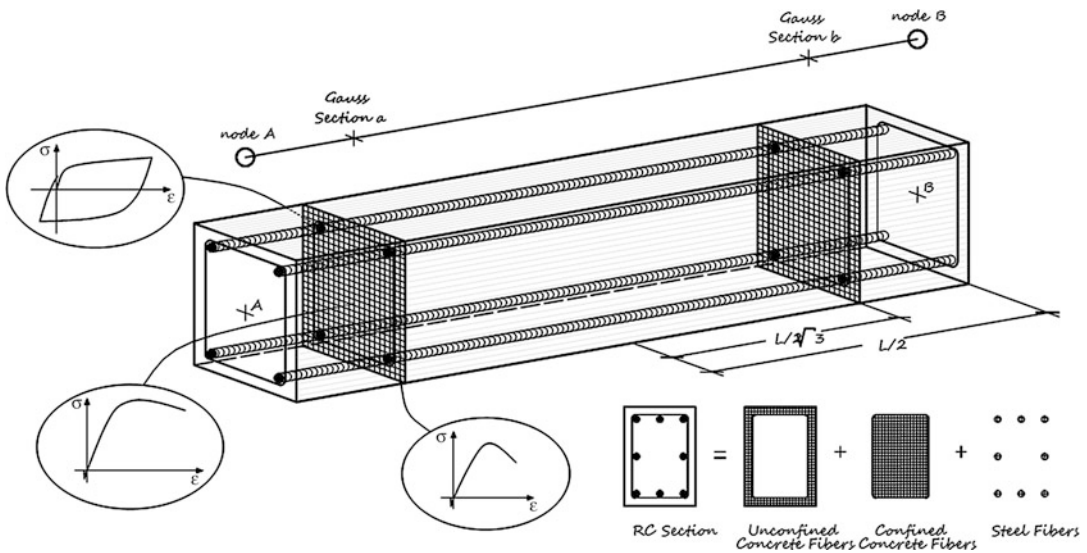


rigidities (GA_A , GA_B) are calculated from the $V - \gamma$ (shear versus shear deformation) curves at each end, details of which are given in Mergos and Kappos (2012). After determining the distribution of GA along the R/C member at each step of the analysis and by applying the principle of virtual work, the coefficients of the flexibility matrix of the shear sub-element are given by the following equation:

$$f_{ij}^{sh} = \frac{a_{As}}{GA_A \cdot L} + \frac{1 - a_{As} - a_{Bs}}{GA_M \cdot L} + \frac{a_{Bs}}{GA_B \cdot L} \quad (i, j = A, B) \quad (5)$$

Finally, rotations due to slip at the end anchorages are captured with the simple slip sub-element of Fig. 8e that consists of a rigid bar with two uncoupled nonlinear springs at the ends. The $M - \theta_{slip}$ skeleton curve is derived assuming uniform bond stress along different segments of the anchored reinforcement bar (details in Mergos and Kappos 2012).

In the previously presented models, the element stiffness matrix in the post-yield range is set up on the basis of EI values that are estimated from predefined constitutive laws relating bending moment to either end rotation or end curvature; when axial (EA) and shear (GA) rigidities are not taken as constant, their values are estimated from similar predefined laws (e.g., $V - \gamma$). Another option is the *fiber model*, wherein the stiffness parameters are not estimated from predefined laws, but rather from moment – curvature analysis (and, far less often, shear force versus deformation analysis) of a number of “monitoring” or “control” sections, which are divided in a number of “fibers” (in the general, biaxial, case, these are squares or rectangles rather than horizontal fibers), using Bernoulli’s principle and the stress-strain constitutive laws of the pertinent materials, i.e., confined and unconfined concrete (for the core and the cover, respectively) and steel bars, as shown



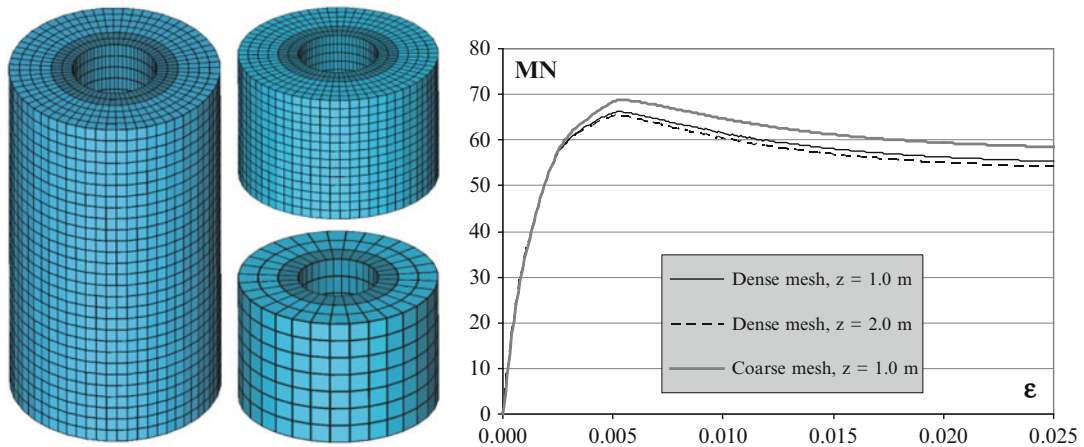
Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 9 Discretization of R/C member using the fiber approach

in Fig. 9 (Kappos et al. 2012). The model has been implemented in point-hinge models but has found its main use in distributed plasticity models, wherein four (Fig. 9) or more sections are “monitored” and the element stiffness matrix is set up assuming linear variation of stiffness (or flexibility) between monitoring sections; the latter are typically taken as the end sections and the Gauss points used in the integration required for deriving the element stiffness matrix (e.g., using the Gauss-Lobatto quadrature scheme).

A rigorous application of the fiber model involves a number of difficulties, and different approaches have been put forward, some of them based on the stiffness approach (and involving displacement shape functions) and others on the flexibility approach (involving force shape functions, which do not change in the inelastic range); hybrid procedures have also been used. The flexibility approach is numerically more advantageous but computationally more demanding. Details of all these procedures fall beyond the scope of this entry and can be found in the literature (e.g., Fardis 1991; Kappos et al. 2012). However, it has to be emphasized here that although the fiber approach is more rigorous than the “phenomenological” approaches based on predefined force – deformation relationships,

it is not necessarily more accurate than the latter, except in the case of R/C members with negligible effect of shear and bond-slip deformations, which is not very common, even in well-designed bridge members. For instance, a pier usually has such an aspect ratio that shear deformations cannot be ignored. There are versions of the fiber model wherein shear deformations are included (Kappos et al. 2012), but the computational cost involved is particularly high. Finally, because most of the available fiber model-based software packages (like SeismoStruct) include constitutive laws assuming that concrete is initially uncracked, they overestimate the actual stiffness of real bridges (and other structures) that are cracked prior to being subjected to seismic loading (due to shrinkage, traffic and ambient vibrations, and possible previous small earthquakes).

Nonlinear *continuum models* are typically used for research purposes, whose aim is to study in detail the response of critical regions in piers. Material nonlinearity is taken into account using either a “standard” plasticity model or a combination of plasticity and damage models, the latter being able to affect the elastic component of the deformation as (seismic) damage propagates. Geometric nonlinearity is less critical in R/C piers, except for very tall ones. The type of



Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 10 Modeling of hollow circular piers with solid elements using different heights and mesh densities

finite elements used can be quite “heavy,” i.e., shell and 3D solid (“brick”) elements have been used for concrete piers. Figure 5 shows an example of using shell elements for both the deck and the piers, but nonlinear behavior was not taken into account in that model.

Figure 10 is taken from a study (Papanikolaou and Kappos 2009) focusing on the effect of confinement on the strength and ductility of solid and hollow piers; solid elements were used for concrete, whose behavior was governed by a sophisticated plasticity model accounting for confinement effects, while line elements embedded to the solid elements were used for the reinforcement (transverse reinforcement consisted of spiral or hoop circular reinforcement, with or without transverse links). It is seen that the axial load versus axial deformation curves resulting from the “coarse” mesh are not substantially different from those from the dense mesh; it is worth pointing out that an upper limit of about 6,000 solid elements was found, beyond which the computational cost and volume of results were excessive, which is a good indicator of the type of models that can be analyzed in such a context. Further examples of applications of continuum models to concrete pier components can be found in Kappos et al. (2012).

Soil-Structure Interaction Effects

A key aspect in proper modeling of piers (whether inelastic or equivalent elastic) is

capturing the effect of foundation compliance, i.e., of the fact that pier foundations, especially the shallow but also the deep ones, like piles, do not provide full fixity to the base of the pier, but displace (horizontally), rotate, and even settle (vertically), as the bridge is subjected to seismic loading. Large-size foundations have also the effect of modifying the seismic input to the structure (“kinematic” interaction or “wave scattering” effect), but this issue is not further addressed herein. The interaction of the foundation ground with the bridge substructure (the piers and their foundations) does not only modify the fixity conditions but also increases the damping (“radiation” damping at the ground-foundation interface); in simplified analysis this can be safely ignored, but there are ways to explicitly include it, such as the addition of dash-pot elements at the base of the piers.

A commonly adopted practical approach for calculating the pseudo-static interaction between the bridge foundation and the soil is the *Winkler spring* model, wherein the soil reaction to the foundation movement is represented by independent (linear or nonlinear) unidirectional translational spring elements. In the case of surface foundations, the vertical springs are distributed below the surface of the footing, while in pile foundations horizontal springs are distributed along the pile shaft. Although approximate, Winkler formulations are widely used not only

because their predictions are in good agreement with results from more rigorous solutions but also because the variation of soil properties with depth can be relatively easily incorporated. Moreover, they are efficient in terms of computational time required, thus allowing for easier numerical handling of the structural inelastic response, where this is deemed necessary.

In the case of *surface foundations*, a simple system of three translational (two horizontal, x and y , one vertical, z) and two (less often three) rotational springs can be used at the base of the footing; the spring constants for the x , y , and z springs can be estimated from relationships (ASCE 2007) derived from the solution of the problem of a rigid plate resting on the surface of a homogeneous half-space:

$$\begin{aligned} K_x &= \frac{GB}{2-v} \left[3.4 \left(\frac{L}{B} \right)^{0.65} + 1.2 \right] \\ K_y &= \frac{GB}{2-v} \left[3.4 \left(\frac{L}{B} \right)^{0.65} + 0.4 \frac{L}{B} + 0.8 \right] \\ K_z &= \frac{GB}{2-v} \left[1.55 \left(\frac{L}{B} \right)^{0.75} + 0.8 \right] \end{aligned} \quad (6)$$

where G is the shear modulus of the ground, v the Poisson ratio (0.35 for unsaturated soils and 0.5 for saturated soils), L the larger dimension of the (rectangular) footing, and the smaller one; similar relationships are given in ASCE (2007) and FHWA (2006) for the rotational springs. These relationships are particularly convenient to use since they only include very fundamental properties of the soil, which can always be estimated (e.g., the initial modulus G_0 can be estimated from the shear wave velocity and the specific weight of the soil). However, an upper and lower bound approach to defining stiffness and (in nonlinear models) capacity is recommended because of the uncertainties in the soil properties and the static loads on the foundations of existing bridges. The large-strain effective shear modulus, G , can be roughly estimated on the basis of the anticipated peak ground acceleration (PGA); for regions of low-to-moderate seismicity, a value of $G = 0.5 G_0$ is recommended, while

for regions of moderate-to-high seismicity, $G = 0.25 G_0$ is suggested (FHWA 2006).

In *pile foundations* the mechanical parameters for the springs are frequently obtained from experimental results (leading to P-y curves for lateral and N-z curves for axial loading) as well as from very simplified models. A commonly used P-y curve is the lateral soil resistance versus deflection relationship proposed by the American Petroleum Institute (API):

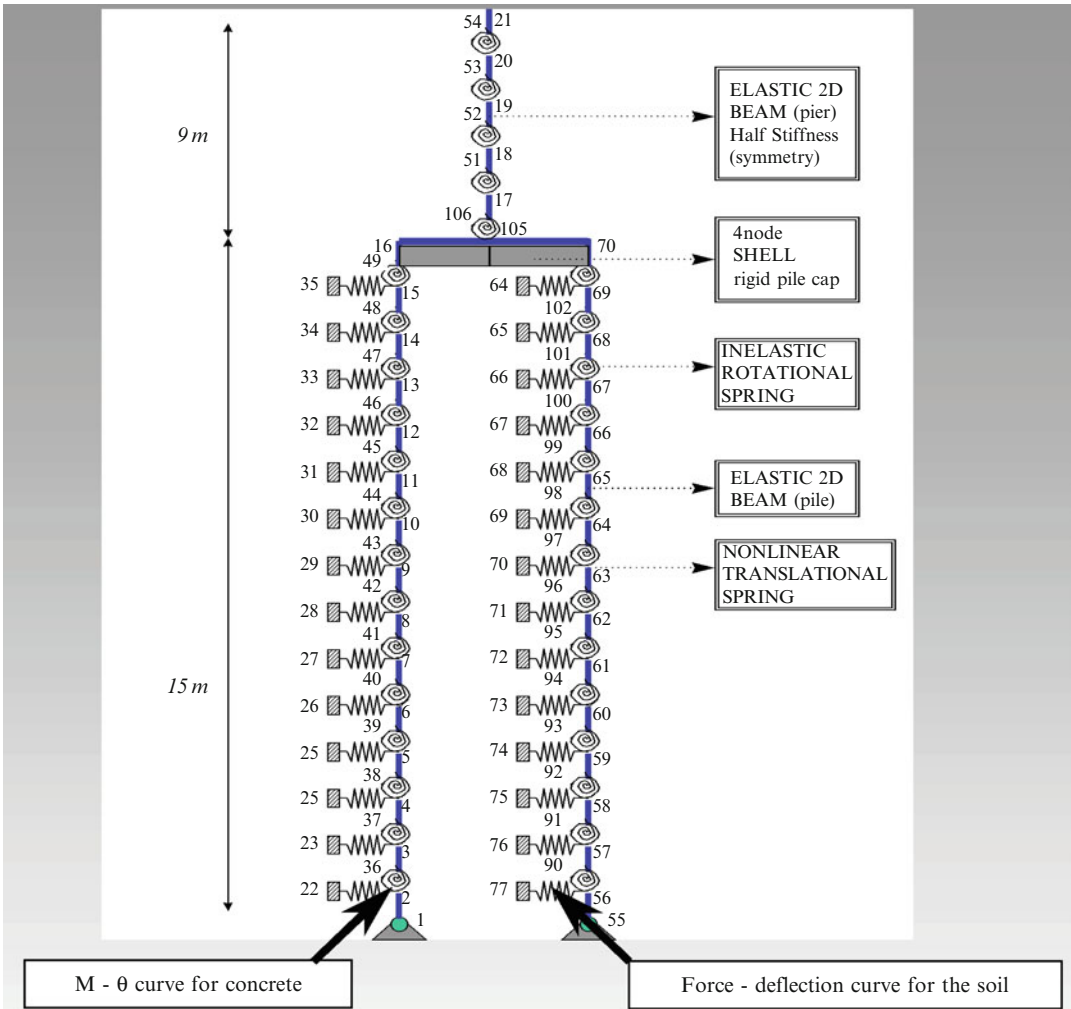
$$P = 0.9p_u \tanh \left[\frac{kH}{0.9p_u} y \right] \quad (7)$$

where p_u is the ultimate bearing capacity at depth H , y is the lateral deflection, and k is the initial modulus of subgrade reaction. The tip of the pile can either be modeled using a vertical spring (with a carefully selected axial stiffness) or assumed to be vertically fixed but free to rotate.

Figure 11 shows an example (Kappos and Sextos 2001) of modeling a pile group using Winkler springs with initial properties calculated from Eq. 7. This is a fully inelastic model wherein both the piles and the pier (modeled using the lumped plasticity approach, but with several elements for each pile) can yield; in practical applications equivalent linear properties (based on secant stiffness at the estimated maximum displacement) are often used, especially for the soil. The limitations of the P-y approach are discussed in detail in Kappos et al. (2012), where an overview of more advanced soil-bridge interaction (kinematic and inertial) models is also provided. It will only be mentioned here that despite the abundance of models and software, nonlinear analysis wherein both the bridge and the foundation ground are modeled with nonlinear models is not only cumbersome but often leads to convergence problems; hence, explicit treatment of nonlinearity should be confined where the main interest lies, typically in the piers and perhaps in the piles (cf. Fig. 11).

Abutments and Backfills

There are two common types of bridge abutments:

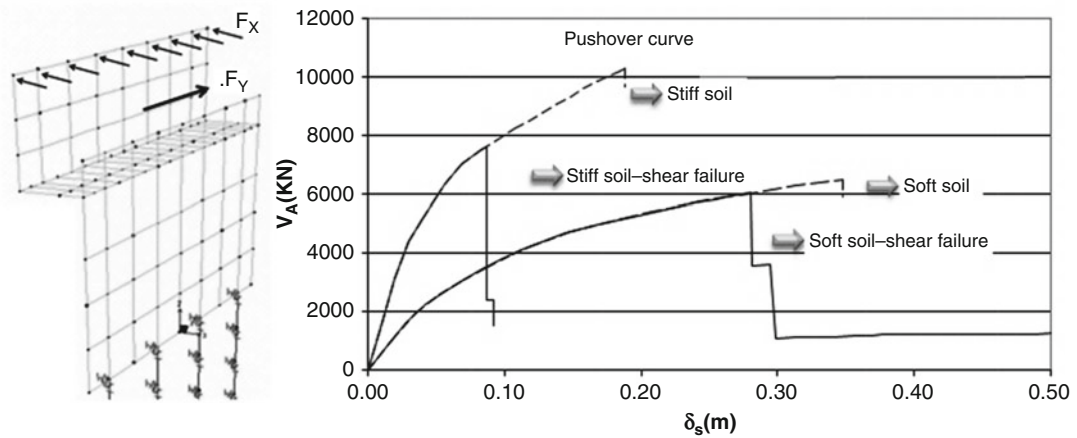


Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 11 Modeling of soil-pile foundation-pier system using Winkler springs

- Seat type (the deck is bearing-supported on the horizontal seat of the abutment; see Fig. 2 bottom right)
- Integral or diaphragm type (abutment monolithically connected to the end of the deck)

The modeling of the abutment system, which also includes the wing walls and the foundation (footing or piles) can vary from very simple to very complex, depending on the situation. Some common cases are briefly discussed in the following.

Integral- or *diaphragm-type* abutments are always included in the model of the bridge, as a continuation of the deck model; in addition to the body of the abutment, they also include the foundation, which typically consists of piles (often relatively flexible ones). It is essential that the flexibility of the abutment foundation be modeled; otherwise the displacements of the integral bridge are seriously underestimated; besides problems with earthquake analysis, failure to capture the flexible end supports of the bridge also results in unrealistic stresses from



Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 12 Modeling of abutment system and resulting pushover curves

temperature variations and shrinkage of concrete. In addition to the foundation, the backfill and the embankment have to be accounted for in the model, especially in short bridges with stiff deck; a concept usually adopted in these cases is the “effective” (or “critical”) length of the embankment, whose properties are calculated and then introduced as springs (discussed later in this section) at the end of the elements representing the integral abutment.

Seat-type abutments can be modeled in a simple way, as shown in Fig. 2, i.e., by just including in the model the bearings (as springs, see section “[Bearings](#)” for property definition) of the abutment seat; so long as the end connection of the bridge is dominated by the properties of the bearings (in modern concrete bridges these are either elastomeric or pot bearings that can slide in all directions, see section “[Bearings](#)”), this simple model is quite adequate. If pot bearings are used, even a simple sliding connection (roller support) can be defined in the model; this ignores friction forces at the pot bearings, which are small if monolithic connections or fixed bearings are used elsewhere in the bridge. However, when a longitudinal joint closes, or a shear key blocks the transverse movement of the deck, or the gap between the deck and the abutment stem wall closes (see Fig. 2 bottom right), the abutment-backfill system is activated and significant forces

can develop at the bridge ends; further movement (in either direction) can be captured by the model only if the flexibility of the system, which includes both concrete members and the backfill soil, is modeled.

In abutments the important aspects of soil-structure interaction (section “[Soil-Structure Interaction Effects](#)”) are modeled in practice-oriented applications through a system of linear or nonlinear *springs* at the ends of the bridge. The properties of the springs can best be defined by an analysis of the abutment-backfill system, preferably accounting for nonlinear effects directly, or at least by proper selection of reduced properties (e.g., estimates of G of the ground consistent with the expected deformations). Figure 12 shows the modeling of an abutment system (Kappos and Sextos 2009); the abutment wall is modeled with 2D shell elements, while the piles with frame elements supported on (depth dependent) nonlinear horizontal springs. In the vertical direction, friction springs were used along the piles and an appropriate vertical stiffness was introduced with the use of a (compression only) spring at the tip of the piles; a simpler model was also analyzed wherein infinite vertical stiffness was assumed (tip displacements restrained). On the right of Fig. 12 are shown the *pushover curves* (i.e., seismic force versus monitoring point displacement) derived for the transverse direction

(force F_y), wherein the behavior of the abutment system is dominated by the nonlinear response of the piles. For investigation purposes, the analysis was performed both for soft soil conditions (which was the actual case in the specific abutment) and for the case of a significantly stiffer supporting soil. These pushover curves reveal the sensitivity of the abutment response to the soil conditions, as well as the importance of accounting for all possible failure modes in the analysis (shear failure of the piles limits the deformation capacity, i.e., the ductility, of the entire system).

Equivalent linear or nonlinear springs based on the stiffness defined by curves like those in Fig. 12 can be used for modeling abutments to which the deck forces are directly transferred (due to the previously mentioned reasons). A full-range model, covering all stages of the response, should include two (nonlinear) springs in series, one representing the stiffness of the bearings and one the stiffness of the abutment system; a gap element (with the width of the longitudinal joint of the bridge) should be added if analysis is carried out in a single run; otherwise separate analyses with open and closed joint should be carried out (a usual practice in the USA).

In lieu of carrying out a proper analysis of the abutment, simplified procedures can be used for estimating a reasonable stiffness (and, in nonlinear analysis, strength) for the abutment. Arguably the most popular procedure is that prescribed in Caltrans (2013), wherein the *longitudinal* stiffness can be calculated from the initial embankment fill stiffness $K_i \approx 28.7 \text{ kN/mm}$ (m width of the wall), and this has to be adjusted proportionally to the backwall (or diaphragm) height (h):

$$K_{\text{abut}} = K_i \times w \times (h/1.7) \quad (8)$$

where w is the projected width of the backwall or diaphragm, for seat and diaphragm abutments, respectively, and $(h/1.7)$ is a proportionality factor based on the 1.7 m height of the diaphragm abutment specimen tested at UC Davis (the actual relationship is not linear, but so far there is no sufficient data to develop a more sophisticated

one). The aforementioned K_i applies to well-compacted backfills as required by Caltrans (2013); otherwise it should be reduced by 50 %. The ultimate abutment load was assumed to be limited by a maximum static soil passive pressure of 240 kPa; the latter is multiplied by the corresponding surface, e.g., the product of the backwall width and height in seat-type abutments as well as the proportionality factor $(h/1.7)$. The stiffness value for Eq. 8 applies when the elastic response of the bridge is dominated by the abutments; when this is not the case, Caltrans prescribes reductions depending on the ratio of the longitudinal displacement demand at the abutment (from elastic analysis) to the effective longitudinal abutment displacement at idealized yield (ratio of strength to stiffness).

In the *transverse* direction, a nominal abutment stiffness equal to 50 % of the elastic transverse stiffness of the adjacent bent can be used; this nominal stiffness has no direct correlation or relevance to the actual residual stiffness (if any) provided by the failed shear key but is meant to suppress unrealistic response modes associated with a completely released end condition.

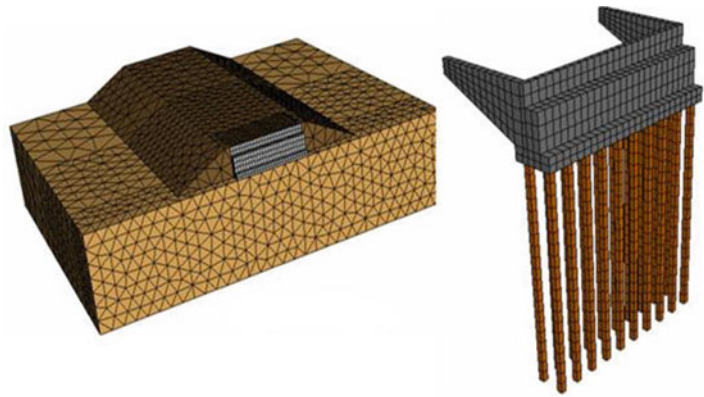
Clearly, a full model for the abutment-backfill system should also include the backfill soil, as well as part of the embankment that is activated during the seismic excitation of the bridge; as mentioned previously, the latter is important in the case of short bridges (like two-span overpasses) wherein the embankment plays a key role in the seismic response of the bridge. In particular, consideration of the abutment-soil system participating mass has a critical effect on the mode shapes and consequently the dynamic response of the bridge. The critical length L_c of the embankment to be considered in the analysis can be estimated from the relationship (Zhang and Makris 2002):

$$L_c \approx 0.7 \sqrt{SB_c H} \quad (9)$$

where S is the slope of the embankment, H its height, and B_c its width at the crest. It should be borne in mind that L_c actually changes with the level of the seismic action, but this is difficult to capture in practical analysis.

Seismic Analysis of Concrete Bridges:

Numerical Modeling,
Fig. 13 Modeling of abutment and backfill system using 3D finite elements



An example of a “heavy” finite element model of the entire system is shown in Fig. 13 from Kappos et al. (2012) wherein further examples and details can be found. Results of analysis of a number of typical abutment and backfills using sophisticated models such as that of Fig. 13, which considered the soil (backfill, embankment, and foundation) as the nonlinear material mechanism, have shown that both the stiffness and the strength estimated according to the pre-2013 Caltrans provisions (adopting $K_i = 11.5$ kN/mm/m, i.e., 60 % lower than the new value) underestimated the values found from the 3D FE models.

So long as the soil behind the abutment has been analyzed and its (macroscopic) stiffness reduced to a spring constant, the entire bridge can be modeled by combining in series the aforementioned translational springs (one for the backfill-embankment system and one for the abutment and its foundation) in each direction of the bridge. More details on modeling abutments and backfills using a system of nonlinear springs can be found in Aviram et al. (2008).

Bearings, Joints, and Shear Keys

Bridge furnishings include a number of components, i.e.,

- Bearings
- Joints
- Parapets – rails
- Waterproofing system

Among these, critical components of the bridge, particularly in an earthquake resistance context, are the bearings and the joints. Modeling of these critical components for seismic analysis is discussed in the remainder of this section, which also covers some other components, normally located close to, or even within, the bearings, i.e., shear keys and damping devices.

Bearings

Bearings are mechanical systems which:

- Transmit loads from the superstructure (deck) to the substructure (piers, abutments)
- Accommodate relative displacements between them

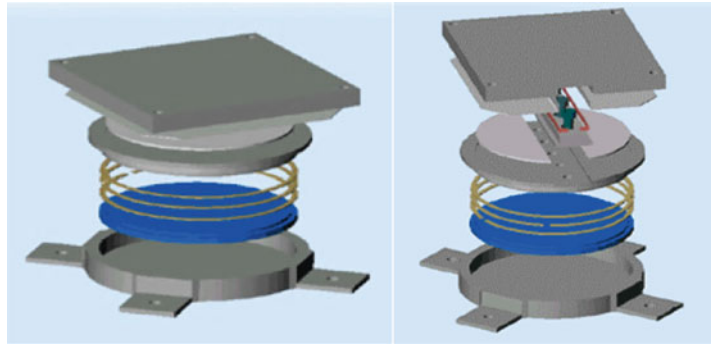
In the past, steel bearings of the pin, roller, rocker, or sliding type have been used, and they are still used in some bridges, in particular steel ones. In modern concrete bridge construction, bearings typically belong to one of the following categories:

- Pot bearings
- Elastomeric (common) bearings
- Elastomeric (special) bearings

Pot bearings (Fig. 14) allow sliding and rotation and consist of a shallow steel cylinder (or “pot”) on a vertical axis with a neoprene disk which is slightly thinner than the cylinder and fitted tightly inside. A steel piston fits inside the cylinder and bears on the neoprene, while flat

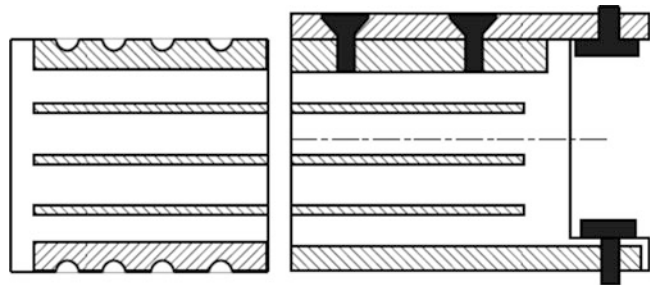
Seismic Analysis of Concrete Bridges: Numerical Modeling,

Fig. 14 Pot bearings (www.agom.it): free-sliding (*left*) and transversely guided sliding



Seismic Analysis of Concrete Bridges: Numerical Modeling,

Fig. 15 Laminated bearing with outer steel plates: profiled (*left*) or allowing fixing (EN1337, CEN 2005b)



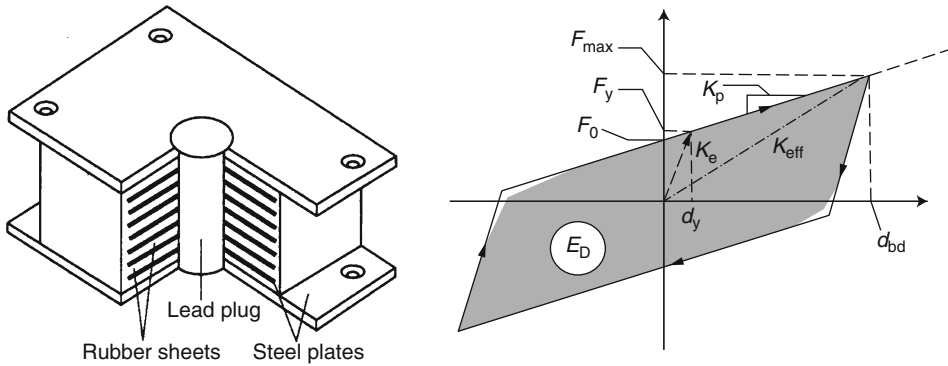
brass rings are used to seal the rubber between the piston and the pot; the rubber behaves like a viscous fluid, flowing as rotation occurs. Sliding can either take place in any direction or be guided (through a groove and sliding bar system; see Fig. 14-right) in a specific direction. Use of pot bearings is very common, especially at the seats of the abutments.

In general, pot bearings do not have to be explicitly modeled; it suffices to release the corresponding degrees of freedom at the support of the deck.

Common elastomeric bearings are made of elastomer, i.e., either natural or synthetic rubber (e.g., neoprene), which is flexible in shear (low GA) but very stiff against volumetric change. To avoid bulging (lateral expansion that adversely affects the properties of the elastomer) different types of reinforcement are used in the elastomer (fiberglass, cotton, steel). The most common and efficient (and also the most expensive) type of reinforced elastomeric bearings is that reinforced with thin steel plates as shown in Fig. 15; these are constructed by vulcanizing elastomer to these steel plates.

The design of elastomeric bearings is carried out (in European and some other countries) according to the European Standard EN1337 (CEN 2005b). This standard prescribes maximum strains due to vertical load, rotations and horizontal actions, such as loads or displacements, and minimum thickness of the internal and external steel plates. It also prescribes a number of ultimate limit state verifications (limitation of distortion and rotation, tension in the steel plates, bearing stability (buckling), and slip prevention). The procedure for designing a bridge so that the seismic action is resisted entirely by elastomeric bearings on all supports (“seismic isolation”) is prescribed in Chap. 7 of Eurocode 8-3 (CEN 2005a).

Proper modeling of elastomeric bearings is essential in the framework of seismic design. In all cases at least the horizontal shear stiffness (K_h) should be captured, but in more refined models the flexural (K_b) and the axial (K_v) stiffness of the bearings are also introduced in the model. These three stiffness values can be calculated from the following relationships:



Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 16 Lead rubber bearing (rectangular section) and corresponding hysteresis law

$$K_h = GA/t_r \quad (10a)$$

$$K_b = 0.329 E_c I / t_r \quad (10b)$$

$$K_v = E_c A / t_r \quad (10c)$$

where t_r is the thickness of the elastomer (not the total height of the bearing) and the other symbols have their usual meaning. The shear modulus G can be taken as 0.9 MPa for static loading and 1.8 MPa for dynamic loading (rubber is a visco-elastic material, i.e., its resistance increases with the loading rate).

Special elastomeric bearings are those that provide the high energy dissipation needed to resist strong earthquakes. The most common types used today are:

- High damping elastomeric bearings
- Lead-rubber bearings (laminated bearings with lead core)

Bearings of the first category can provide fairly high values of effective damping ratio ($\xi_{\text{eff}} \approx 10\text{--}25\%$). The main concern about them is durability, as rubber properties are known to deteriorate with time; of course, companies manufacturing them claim design lives appropriate for normal uses. Lead-rubber bearings (LRB) are laminated bearings with a cylindrical lead core, as shown in Fig. 16; they are either circular or rectangular. The selection of lead as the core

material is due to its high hysteretic energy dissipation (“fat” elastoplastic hysteresis loops under cyclic loading; see Fig. 16 right). In LRBs the equivalent viscous damping ratio can reach quite high values (in the range of 30 %). In addition to damping, the lead core also provides resistance to service lateral loads.

Modeling of special elastomeric bearings is similar to that of common bearings as far as stiffness is concerned (relationships 10), but the additional damping has to be properly introduced in the model. In all types of elastic analysis, a viscous damping ratio higher than that of reinforced concrete (5 %) is introduced, and the corresponding spectra are used to derive the seismic actions; in Eurocode 8 the ordinates of the elastic spectrum for $\xi \neq 5\%$ are estimated by multiplying the reference spectrum by

$$\eta = \sqrt{10/(5+\xi)} \geq 0.40 \quad (11)$$

The value of effective damping ratio to be used can be estimated from (CEN 2005a)

$$\xi_{\text{eff}} = \frac{1}{2\pi} \left[\frac{\sum E_{D,i}}{K_{\text{eff}} d_{\text{cd}}^2} \right] \quad (12)$$

where $\sum E_{D,i}$ is the sum of dissipated energies of all special bearings i in a full deformation cycle at the design displacement d_{cd} and $K_{\text{eff}} = \sum K_{\text{eff},i}$, i.e., the sum of the composite stiffnesses of the

isolator unit and the corresponding substructure (pier) i . For the specific case of LRBs, the dissipated energy $E_{D,i}$ is calculated from the pertinent elastoplastic hysteresis loop (Fig. 16 right).

An interesting type of special bearing, used exclusively for seismic isolation, is the *friction pendulum* shown in Fig. 17, wherein the sliding surface of the bearing is concave; hence, the restoring force is provided by the horizontal component of the structure itself. Sliding on the concave surface is resisted by friction of the contact material which is PTFE (polytetrafluoroethylene, most common commercial name Teflon); the friction coefficient is high initially (hence, no swaying of the superstructure takes place under normal loading conditions) but substantially higher under high velocities induced by earthquake.

The *articulation* of the bridge, i.e., the arrangement of the different types of bearings, is a critical aspect of the design of the bridge, in particular the seismic one. Figure 18 shows an

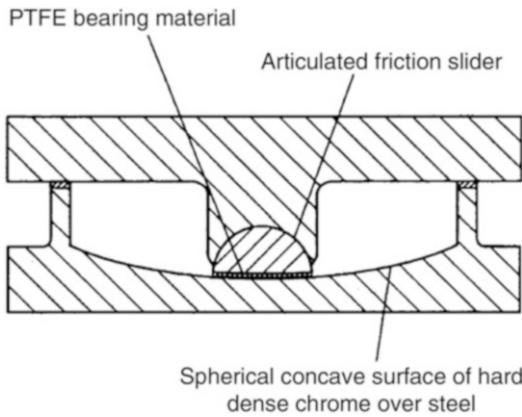
example of an actual railway bridge wherein a seismic isolation approach was adopted, involving a combination of lead rubber bearings and viscous fluid dampers; it is seen that the LRBs located toward the end of the bridge are movable horizontally, while free sliding pot bearings are used at the abutments.

Supplemental Damping Devices

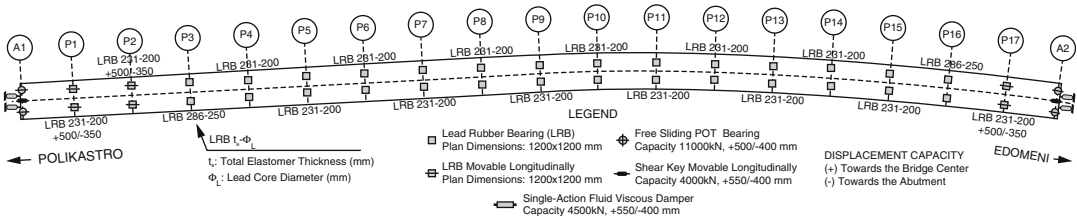
On several occasions involving large bridges, the amount of damping that can be provided by realistic arrangements of special bearings is not sufficient for limiting displacements to the required limits (recall that long-span bridges are long-period structures and when isolation is used their fundamental period can become very long, more than 3 s). In such cases a more efficient solution can be the use of special damping devices (separate from the bearings) that will supplement the energy dissipation provided by the bearings (high damping or LRB); alternatively, common elastomeric bearings or friction pendulum bearings can be used in combination with the damping devices. The most commonly used devices are

- Viscous fluid dampers
- Steel yielding devices

Viscous fluid dampers are based on the concept (long used in the automotive industry) of forcing through a piston a viscous fluid (usually silicon oil) through an orifice. Another, more recent, alternative are shear panels containing high-viscosity fluids. The constitutive law of such dampers is not restricted to the well-known linear dependence on velocity (through C , the



Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 17 Friction pendulum bearing



Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 18 Arrangement of bearings in a seismically isolated railway bridge (Kappos et al. 2014)

damping coefficient) but is often nonlinear, of the form

$$F_D = C \cdot |\dot{u}|^\alpha \cdot \text{sgn}(\dot{u}) \quad (13)$$

where F_D is the damping force and α is an exponent between 0 and 1; $\alpha = 1$ corresponds to standard viscous damping, but in practical applications lower values are used since they lead to higher F_D at lower velocities (e.g., in the bridge shown in Fig. 18, $\alpha = 0.15$ was selected for the viscous dampers, which had $C = 5,440 \text{ kN}\cdot\text{s/m}$).

Modeling of viscous fluid dampers can be made using two approaches: in the simpler approach, appropriate for practice-oriented elastic analysis, the dampers are substituted by an effective value which is the sum of the basic damping ξ_0 (typically 5 %) and the contribution of the fluid dampers (having exponent α and damping coefficients C_j). Considering the fundamental mode of the bridge (modal displacements ϕ_i at each mass m_i) and calculating the energy dissipated by the nonlinear dampers in a cycle of sinusoidal motion, it can be shown (Hwang 2002) that

$$\xi_{\text{eff}} = \xi_0 + \frac{\sum_j \lambda C_j \phi_{ij}^{1+\alpha} \cos^{1+\alpha} \theta_j}{2\pi A^{1-\alpha} \omega^{2-\alpha} \sum_i m_i \phi_i^2} \quad (14)$$

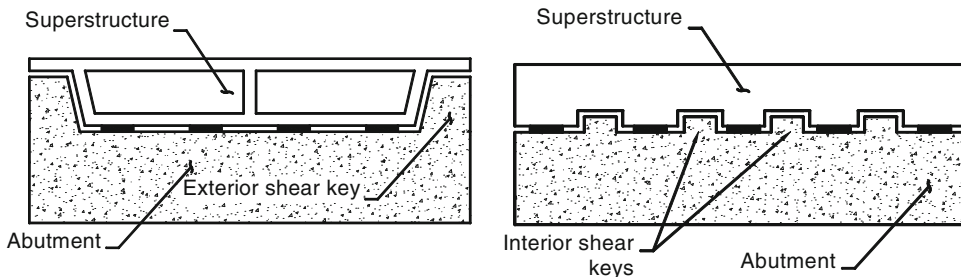
where ω is the circular frequency, θ_j is the angle of inclination of the damping device j , λ is a function of the exponent α (tabulated values of λ are given in FEMA 1997), and $u_i = A\phi_i$ (A is the amplitude) are the actual displacements of the masses m_i .

In a more involved, nonlinear, analysis, the viscous dampers can be directly introduced in the model at the particular locations of the bridge wherein they are installed; such elements are available in some software packages but are rarely used in practical design.

Steel yielding devices provide an almost elastoplastic hysteresis loop, and it is possible to get a great variety of damping ratio values by properly selecting the yield displacement and post-yield stiffness ratio (hardening) of the isolation system. They can be modeled either by expressing the hysteretic energy dissipation as an equivalent damping (by equating the area of the hysteresis loop to that of the ellipse representing viscous damping energy) or by directly including yielding elements at the pertinent positions of the bridge (such elements are available in most programs).

Shear Keys

Shear keys serve the purpose of preventing the displacement of the bridge deck in a certain direction and can be located at several positions in a bridge; a typical one is at the abutments for blocking the transverse movement of the deck. They can be either external as shown in Fig. 19 left (i.e., forming part of the seat) or as interior short cantilevers interlocking with corresponding grooves in the deck (Fig. 19 right); exterior shear keys are preferable because they are easy to inspect and repair. Shear keys might directly bear on the surrounding part of the deck, immediately blocking its movement, or be located at a selected distance, forming a local joint (see section “Joints” for modeling of joints).



Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 19 Exterior and interior shear keys in bridges (Kappos et al. 2012)

Modeling of shear keys should be consistent with their design “philosophy,” i.e., whether the shear key is meant to remain essentially elastic for the entire range of seismic response of the bridge considered in design, or a “sacrificial” element whose capacity should be limited with a view to protecting other more critical and/or more difficult to repair components of the bridge; a popular design concept in California is that the capacity of the shear keys should not exceed or be close to the shear capacity of the piles (on which the abutment is supported). Eurocode 8-2 (CEN 2005a) vaguely specifies that the design actions for the seismic links (one type of which is shear keys) should be derived as capacity design effects, with the horizontal resistance of the bearings assumed equal to zero, without explaining how these effects should be derived (clearly one possibility could be to relate the strength of shear keys to that of piles, as mentioned previously).

The strength of shear keys can be derived from sophisticated models such as strut and tie ones or simplified ones, usually based on the shear friction concept. Bozorgzadeh et al. (2006) proposed the following relationship for the nominal capacity of a shear key:

$$V_n = \frac{\mu_f \cdot \cos \alpha + \sin \alpha}{1 - \mu_f \cdot \tan \beta} \cdot A_{vf} \cdot f_{su} \quad (15)$$

where μ_f is a kinematic friction coefficient, β is the angle of inclined face of the shear key (Fig. 19 left), α is the angle of kinking of the vertical bars with respect to the vertical axis (recommended value from test results 37°), and f_{su} is the (ultimate) tensile strength of the vertical reinforcement that has an area A_{vf} . Such models strongly depend on the friction coefficient that varies substantially depending on the detailing of the shear key; for sacrificial keys with smooth finishing of the concrete interface, $\mu_f = 0.36$ is recommended (Bozorgzadeh et al. 2006), but for properly detailed joints of adequate roughness, much higher values apply (up to 1.4 for keys monolithically cast with the abutment seat).

Having established a proper value for the shear key strength, the key can be modeled simply as a rigid-plastic spring (or “link”) with a

displacement capacity of around 100 mm or using more sophisticated multilinear constitutive laws with ascending and descending branches (see more details in Kappos et al. 2012). In linear, “code-type,” analysis, one usual option in the USA is to carry out two analyses, one with and one without the shear keys (essentially one assuming displacement is blocked in the pertinent direction and one with the deck allowed to displace freely).

Joints

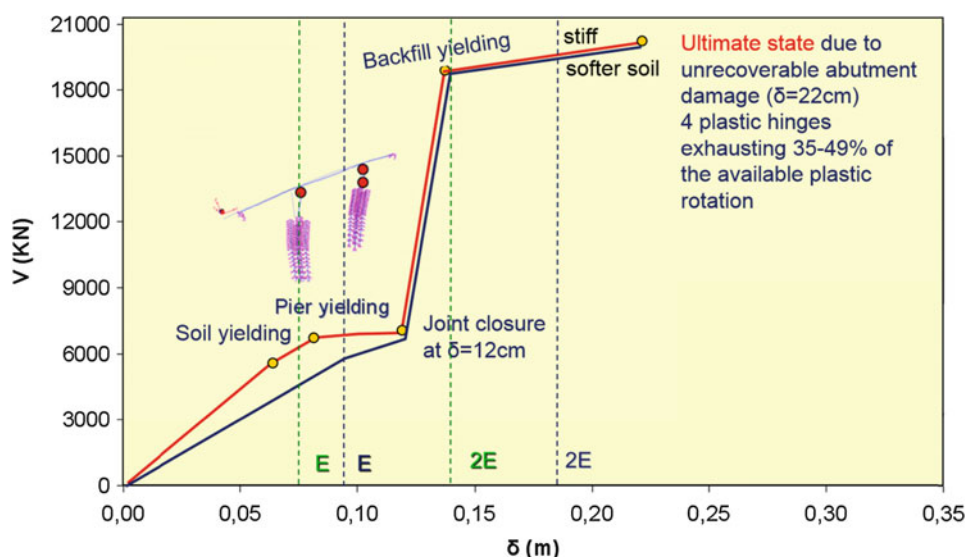
Joints (also called *expansion joints* which is inappropriate when they are also designed for seismic actions) are required to accommodate (with negligible resistance and noise) movements of the deck due to:

- Thermal expansion/contraction
- Shrinkage and creep of concrete
- Earthquake-induced horizontal movement

There are several types of joints (compression seal, strip seal, finger plate, sliding plate, modular), each of them appropriate for up to a certain design movement of the bridge. In bridges designed for high seismic actions, the joint gap might result as quite substantial (over 200 mm). It is noted that current seismic codes like Eurocode 8-2 require the joint gap to satisfy

$$d_{ED} = d_E + d_G + \psi_2 d_T \quad (16)$$

where d_G is the long-term horizontal displacement due to permanent and quasi-permanent actions (posttensioning, shrinkage, and creep), d_T the displacement due to thermal actions ($\psi_2 = 0.5$ for road bridges), and d_E is the seismic displacement calculated as $\pm \eta \mu_d d_{Ee}$ where d_{Ee} is the displacement derived from the analysis for the seismic loading combination (with the design spectrum reduced by the behavior factor q), μ_d is the design ductility ($\mu_d = q$ in the common case that the fundamental period of the bridge $T \geq T_o = 1.25T_c$ where T_c is the corner period of the design spectrum) and η is the damping correction factor for the design spectrum ($\eta = 1$ for $\xi = 5\%$).



Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 20 Pushover curves for the longitudinal direction of an overpass bridge (Kappos and Sextos 2009); “E” denotes the design earthquake level

When large movements ($d_{Ed} > 100$ mm) have to be accommodated, *modular* joints are used, wherein sealing elements and rail elements are coupled; in this case intermediate elements (rails), supporting elements, and linkage elements (e.g., folding trellis) causing equal gap widths are needed, and the total movement accommodated by the joint is the sum of the distances of the rails.

It is common practice in Europe to provide substantial joint gaps (as well as ample seat lengths) and make sure that the joints remain open during the design earthquake, without carrying out specific verifications for the case that the gap closes. In California, “dual” analysis is the recommended (by Caltrans) practice, wherein the bridge is analyzed assuming either free movement or full restraint at the “compression end” of the bridge (movement is always free at the “tension end” where the deck moves away from the abutment) and taking the most unfavorable response quantities from either set of analyses; clearly this is conservative and might result in increased costs. Another interesting difference in US and European practice is that *restrainers* (either cables or bolt linkages, the bolts passing through holes in the deck) are often used at the

joints of US (and New Zealand) bridges, with a view to preventing *unseating* during earthquakes stronger than the design one. In Europe this is seen as an option mainly in *retrofitting* of existing bridges with insufficient seating lengths.

Modeling of joints in seismic analysis is essential, since the bridge boundary conditions are drastically different when joints are open or closed. For instance, in the longitudinal direction of the bridge, there are at least two joints at the ends (over the seat-type abutments) that are essential for accommodating “non-seismic” displacements (d_G and d_T); when either of these joints closes during the earthquake, substantial forces are transferred to the abutment-backfill system (see section “[Abutments and Backfills](#)” and Fig. 20 later in this section), and the horizontal forces developed in the bridge-abutment-backfill system can be much higher than those developed when longitudinal movement was free.

Proper modeling of joints requires *nonlinear* analysis using the gap (or other special link) elements currently available in most software packages. This element is often combined in series with the spring (or link) elements modeling the bearings and/or shear keys at the abutment.

It is noted that although the gap element has an elastic behavior subsequent to gap closure, the analysis of the bridge is nonlinear, as the boundary conditions change during the analysis. The importance of capturing the effect of joint closure on the seismic behavior of a bridge can be seen in Fig. 20, where pushover curves are given for the longitudinal direction of a typical overpass (whose abutment system is shown in Fig. 12). Recall that two different soil conditions were studied; this has a noticeable effect on the initial stiffness of the bridge, but as soon as the longitudinal gap (120 mm in this case) closes, a drastic increase in both stiffness and strength is noticed, as the abutment-backfill system is now activated. The final failure of the bridge is estimated to take place during this second stage of the response (at a displacement of almost twice the gap length) and is attributed to unrecoverable damage to the soil behind the abutment (50 % loss in strength), while the piers are still well within their rotational capacity (35–49 %). A different failure mechanism (exceedance of available ductility of piers) would have been predicted had the end support been simulated as longitudinal restraint (as per the Caltrans simplified approach).

Bridge Analysis Methods

Methods of analysis can be classified as:

- Elastic (equivalent) static
- Elastic dynamic (response spectrum)
- Inelastic static (pushover)
- Inelastic dynamic (response history)

The basics of all these methods are presented in a series of articles in the encyclopedia and will not be repeated herein. Instead, some aspects of analysis (other than member modeling that was presented in section “[Modeling of Bridge Components](#)”) specific to concrete bridges will be briefly described, and the current trends in nonlinear static (pushover) analysis of bridges (not covered elsewhere here) will also be

presented. These will be followed by a case study involving application of different analysis method to an actual concrete bridge.

Code-Prescribed Analysis of Concrete Bridges

Among the important documents (codes-guidelines) for the *analysis* of concrete bridges are those regularly published by Caltrans (2013, latest version) and the ACI (2014, latest version) in the USA and Eurocode 8-2 (CEN 2005a) in Europe. Concrete bridges are the sole type covered in the ACI Report 341.2, while the Caltrans Criteria and Eurocode 8-2 also cover bridges made of other materials (steel and composite). For *retrofitting* of bridges a comprehensive document is the FHWA (2006) Manual. Some key aspects of these documents specific to the analysis of concrete bridges are discussed in the following. Other important sections relevant to the design of concrete bridges are those prescribing the procedures for resistance verification of reinforced concrete sections and the detailing of R/C members (piers, abutments, and retaining walls), which fall beyond the scope of this entry that focuses on modeling for structural analysis. Of course, calculation of R/C member strength and ductility (especially for piers) is essential for nonlinear analysis of concrete bridges; aspects of this issue are covered in the case study presented later in section “[Comparative Case Study](#).”

- All types of analysis are permitted for concrete bridges, the *equivalent static procedure* being subject to a number of limitations regarding the effect of higher modes; this type of analysis is usually suitable for the longitudinal direction of straight bridges, which is dominated by a single mode (that is often the fundamental mode of the bridge). Three versions of the method (Rigid Deck Model, Flexible Deck Model, Individual Pier Model) are prescribed by Eurocode 8-2 (see basic aspects of these methods in Vayas-Iliopoulos (2014)). The reference method in practical design is the (elastic) *dynamic response spectrum analysis*, while *nonlinear* methods are only used in practice for the verification of the design of

some important bridges that have initially been designed using response spectrum analysis. Nonlinear methods are much more common in the case of *assessment* of existing bridges. The FHWA (2006) Retrofit Manual provides sufficient guidance (especially with respect to estimating strength and deformation capacity) for the application of both the push-over and response-history analysis methods, which are covered more briefly in Eurocode 8-2; notably, assessment of existing bridges is not currently covered by the Eurocode package (it is one of the issues that will be added at the next stage of development).

- Application of elastic methods to bridges is the same as for other structures, but there are two aspects specific to concrete bridges that have to be properly addressed:
 - In most concrete bridges both prestressed and ordinary (non-prestressed) concrete are used, the former for the deck, the latter for the piers and abutments. The *damping* ratio ξ is different for these two materials (primarily due to the different degrees of cracking in each), i.e., 5 % for reinforced concrete and 2 % for prestressed concrete. In EC8-2 this is accounted for by considering the response spectrum for an equivalent damping ratio:

$$\xi_{\text{ef}} = \frac{\sum \xi_i E_{\text{di}}}{\sum E_{\text{di}}} \quad (17)$$

where E_{di} is the deformation energy induced in member i by the seismic action. This quantity is not a standard output of common structural analysis programs, and in order to avoid ad hoc spreadsheet calculations, designers often prefer to simply use an average value of 3.5 %. It is worth noting here that E_{di} is *not* the energy *dissipation* through yielding mechanisms, but rather refers to the pre-yield state; hence, it is not appropriate to consider a spectrum for $\xi = 5$ % on the basis that only reinforced concrete members yield and dissipate energy (prestressed concrete decks remain quasi-elastic as already mentioned in section “[Deck](#)”).

- The most important factor in reducing the elastic response spectrum to the design one (which is, in fact, an inelastic spectrum except when elastic response is foreseen for the design seismic action) is the force reduction factor, called *behavior factor* (q) in Eurocode 8 and *response modification factor* (R) in the US codes. Values for q for concrete bridges depend on whether the bridge is designed as ductile or “limited ductile.” For *ductile* concrete bridges with *vertical* piers (working in prevailing bending) $q = 3.5\lambda(\alpha_s)$, where $\lambda(\alpha_s) = 1.0$ for $(\alpha_s) \geq 3$ ($\alpha_s = L_s/h$ is the shear span ratio of the pier, where L_s is the distance from the plastic hinge to the point of zero moment and h is the depth of the cross section in the direction of flexure of the plastic hinge) and $\lambda(\alpha_s) = \sqrt[3]{(\alpha_s/3)}$ for $3 > (\alpha_s) \geq 1$. For ductile concrete bridges with piers consisting of *inclined* struts (e.g., V shaped), $q = 2.1\lambda(\alpha_s)$. For *limited ductile* concrete bridges, the corresponding values are 1.5 (vertical piers) and 1.2 (inclined struts). For abutments rigidly connected to the deck (integral bridges) $q = 1.5$, except in “locked-in” structures, i.e., bridge structures whose mass essentially follows the horizontal seismic motion of the ground (hence, they do not experience significant amplification of the horizontal ground acceleration), in which case $q = 1$. These values, as well as similar ones specified by the American Code (AASHTO 2010), are in many cases conservative. In a recent study evaluating the actual force reduction factors for existing bridges in Europe, Kappos et al. (2013) found that in all bridges studied the available q values were higher than those used for design in both the longitudinal and transverse directions. In fact, in many cases the code-specified values (in particular those of AASHTO for single-column bents) seem to significantly underestimate the actual energy dissipation capacity of concrete bridges. Seen from another perspective,

this is a clear indication that modern bridges possess adequate margins of safety and are able to withstand seismic actions that are often substantially higher than those used for their design. This high performance is due to their ductility, as well as their overstrength; previous studies that have ignored the latter led to unrealistically low estimation of q -factor values.

Nonlinear Static Analysis of Concrete Bridges

Interesting and useful work has been carried out in the last decade on nonlinear (inelastic) static, also known as pushover, analysis of bridges; nevertheless this is clearly less than that for buildings. A recent book presenting all available methods for pushover analysis of bridges is that by Kappos et al. (2012), which also includes a substantial number of case studies involving the comparative application of several methods. Due to space limitations, only one approach will be presented herein which, in the writer's opinion, combines sufficient accuracy with relatively limited effort and the possibility to be applied using available (commercial) software tools, with very limited need for additional spreadsheet calculations; in fact, software for "single-run" application of the method is currently at an advanced level of development. The method is usually referred to as *(multi-)modal pushover analysis (MPA)*; it was presented in a comprehensive form for buildings by Chopra and Goel (2002) and was extended to bridges by Paraskeva et al. (2006). The key idea is to perform multiple pushover analyses of the structure, one for each significant mode, and combine statistically the resulting displacements and rotations. The steps involved in the latest version of the method (Paraskeva and Kappos 2010), which includes a number of improvements, are summarized in the following.

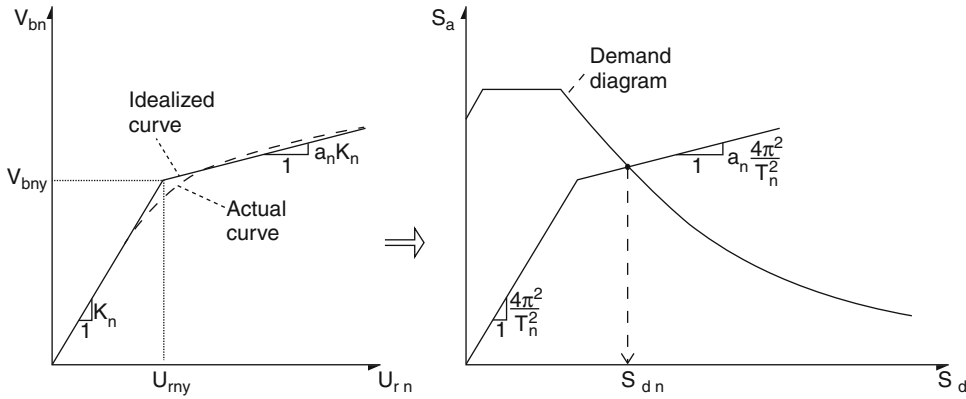
Step 1: Compute the natural periods, T_n , and mode shapes, ϕ_n , for linearly elastic vibration of the structure.

Step 2: Carry out separate pushover analyses for force distribution $\mathbf{s}_n^* = \mathbf{m} \cdot \phi_n$ for each

significant mode of the bridge and construct the pushover curve (base shear V_{bn} vs. displacement u_{cn} of the "control" or "monitoring" point) for each mode; $\mathbf{s}_n^*$ are loading patterns; hence, the relative significance of each mode is not accounted for at this stage; this will be done at Step 5, through the target displacement for each modal pushover analysis. Gravity loads are applied before each MPA and $P-\Delta$ effects are included, if significant (e.g., in bridges with tall piers).

Step 3: The pushover curve must be idealized as a bilinear curve so that a yield point and ductility factor can be defined and subsequently used to appropriately reduce the elastic response spectra representing the seismic action considered for assessment. This idealization can be done in a number of ways, some more involved than others; it is suggested to do this once using the full pushover curve (i.e., analysis up to "failure" of the structure, defined by a drop in peak strength of about 20 %) and the equal energy absorption rule (equal areas under the actual and the bilinear curve). It is noted that the remaining steps of the methodology can be applied even if a different method for producing a bilinear curve is used.

Step 4: Several procedures are available (FEMA 1997; Chopra and Goel 2002; CEN 2004, all referring to buildings) for defining the earthquake displacement demand associated with each of the pushover curves derived in Step 3. Paraskeva et al. (2006) adopted the capacity and demand spectra procedure based on inelastic demand spectra (Fajfar 1999); hence, Step 4 consists in converting the idealized $V_{bn} - u_{cn}$ pushover curve of the multi-degree-of freedom (MDOF) system to a "capacity diagram" (Fig. 21). The base shear forces and the corresponding displacements in each pushover curve are converted to spectral acceleration (S_a) and spectral displacements (S_d), respectively, of an equivalent single-degree-of-freedom (SDOF) system, using the relationships (Chopra and Goel 2002):



Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 21 Idealized pushover curve of the n th mode of the MDOF system and corresponding capacity curve for the n th mode of the equivalent inelastic SDOF system

$$S_a = \frac{V_{bn}}{M_n^*} \quad (18a)$$

$$S_d = \frac{u_{cn}}{\Gamma_n \cdot \phi_{cn}} \quad (18b)$$

wherein ϕ_{cn} is the value of ϕ_n at the control (or “monitoring”) point, $M_n^* = L_n \cdot \Gamma_n$ is the effective modal mass, with $L_n = \Phi_n^T \mathbf{m} \cdot \mathbf{1}$, $\Gamma_n = L_n/M_n$ is a mass participation factor, and $M_n = \Phi_n^T \mathbf{m} \cdot \Phi_n$ is the generalized mass, for the n th natural mode. For *inelastic* behavior, estimation of the displacement demand at the monitoring point is made with the aid of inelastic spectra.

Step 5: Since the displacement demand calculated in Step 4 (for each mode) refers to SDOF systems with periods equal to those of the corresponding modes, the next step is to correlate these displacements to those of the actual bridge. Hence, Step 5 consists in converting the displacement demand of the n th mode inelastic SDOF system to the peak displacement of the monitoring point, u_{cn} of the bridge, using Eq. 18b. The selection of this point is a critical issue for MPA of bridges, and as discussed by Paraskeva et al. (2006), several choices of monitoring point are acceptable as long as the derived pushover curve has a reasonable shape, but they do not

lead to equally good results as far as the final response quantities are concerned. For practical purposes, a good selection is the deck point above the most critical support (pier or abutment) of the bridge.

Step 6: In this step, a correction is made of the displacement of the monitoring point of the bridge, which was calculated at the previous step. The correction is necessary only for cases that significant inelasticity develops in the structure. If the structure remains elastic or close to the yield point, the MPA procedure suggested by Paraskeva et al. (2006) is used to estimate seismic demands for the bridge. The response displacements of the structure are evaluated by extracting from the database of the individual pushover analyses the values of the desired responses at which the displacement at the control point is equal to u_{cn} (see Eq. 18b). These displacements are then applied to derive a new vector Φ_n' , which is the deformed shape (affected by inelastic effects) of the bridge subjected to the given modal load pattern. The target displacement at the monitoring point for each pushover analysis is calculated again with the use of Φ_n' , solving Eq. 18b for u_{cn}' , and recalculated Γ_n using Φ_n' .

Step 7: The response quantities of interest (displacements, plastic hinge rotations, forces in the piers) are evaluated by extracting from

the database of the individual pushover analyses the values of the desired responses r_n , due to the combined effects of gravity and lateral loads for the analysis step at which the displacement at the control point is equal to u_{cn} (or u_{cn}').

Step 8: Steps 3–7 are repeated for as many modes as required for sufficient accuracy; there is little merit in adding modes whose participation factor is very low (say less than 1 %), and application of the method to a number of bridges shows that it is not necessary to assure that the considered modes contribute to 90 % of the total mass.

Step 9: The total value for any desired response quantity (and each level of earthquake intensity considered) can be determined by combining the peak “modal” responses r_{no} using an appropriate modal combination rule, e.g., SRSS or CQC. This simple procedure is used for both displacements and plastic hinge rotations, which are the main quantities commonly used for seismic assessment of bridges. If member forces (e.g., pier shears) have to be determined accurately, a more involved procedure of combining modal responses should be used, consisting in correcting the bending moments at member ends (whenever yield values were exceeded) on the basis of the relevant moment versus rotation ($M - \theta$) diagram and the value of the calculated plastic hinge rotation; this procedure blends well with the capabilities of currently available software.

Comparative Case Study

The overpass shown in Fig. 2 (some aspects of its modeling were discussed in section “[Modeling of Bridge Components](#)”) has three spans and total length equal to 100 m, typical in modern motorway construction. Piers have a cylindrical cross section, while the pier heights are 8 m and 10 m. The deck is monolithically connected to the piers, while it rests on its two abutments through elastomeric bearings; movement in both the longitudinal and transverse directions is initially allowed at the abutments, but transverse displacements are restrained whenever the 150 mm gap shown

at the bottom of Fig. 2 is closed. The Greek Seismic Code design spectrum (similar to that of EC8) scaled to a PGA of 0.16 g was used for seismic design. The design spectrum corresponded to ground category “B” (close to ground “C” in the final version of EC8 (CEN 2004)). The bridge was designed as a ductile structure (plastic hinges expected in the piers) for a behavior factor $q = 2.4$.

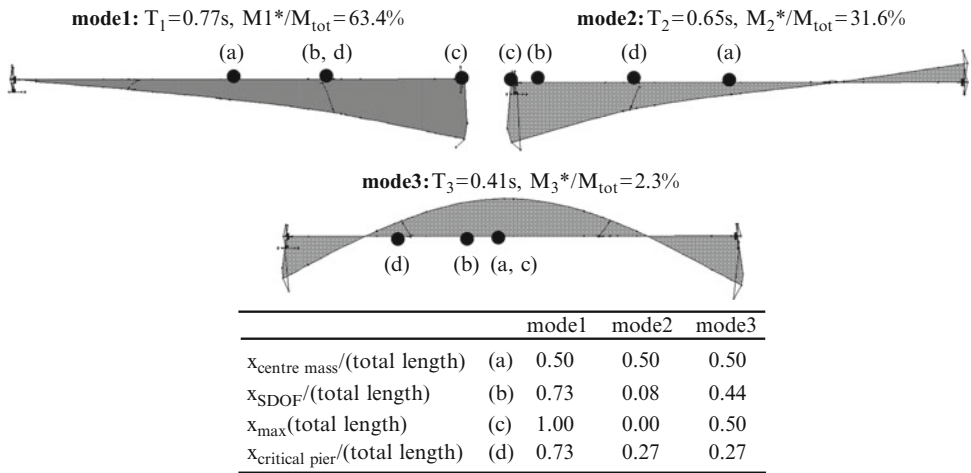
The bridge was analyzed applying a number of popular assessment procedures, i.e.:

- Modal analysis
- “Standard” pushover analysis (SPA) (first mode loading)
- Pushover analysis for a “uniform” loading pattern (as required by Eurocode 8 (CEN 2004) and by the ASCE Standard 41–06 (ASCE/SEI 2007))
- Modal pushover analysis (MPA) as proposed in Paraskeva et al. (2006)
- Improved modal pushover analysis as proposed by Paraskeva and Kappos (2010)
- Nonlinear response-history analysis (NRHA), for artificial records closely matching the demand spectrum (see Paraskeva et al. 2006)

All inelastic analyses were carried out using the SAP2000 software package (CSI 2011). Plastic hinging in the piers had to be modeled slightly differently in the NRHA and the pushover analysis, due to limitations of the software used. More specifically, nonlinear rotational spring elements were used in the finite element models used in NRHA, while the built-in beam hinge feature of SAP2000 was implemented in the models set up for pushover analysis. In both cases, though, the same moment versus rotation ($M - \theta$) relationship was used (i.e., bilinear with 2–6 % hardening, depending on the calculated ultimate moment), with input parameters defined from fiber analysis performed for each pier section, utilizing the in-house developed computer program RCCOLA-90.

Nonlinear Static Analysis

The dynamic characteristics of the bridge were determined using standard modal (eigenvalue)



Seismic Analysis of Concrete Bridges: Numerical Modeling, Fig. 22 Modal force distribution, location of the equivalent SDOF systems, and modal parameters for the main transverse modes of the overpass bridge

analysis. Figure 22 illustrates the first three transverse mode shapes of the overpass bridge, together with the corresponding participation factors and mass ratios, as well as the locations of alternative monitoring points for each mode. Consideration of the modes shown in Fig. 22 assures that more than 90 % of the total mass in the transverse direction is considered. For MPA, applying the modal load pattern of the n th mode in the transverse direction of the bridge, the corresponding pushover curve was constructed and then idealized as a bilinear curve (Fig. 21). As noted under Step 4 of the MPA procedure (see section “Nonlinear Static Analysis of Concrete Bridges”), the inelastic demand spectra method was used for defining the displacement demand for a given earthquake intensity.

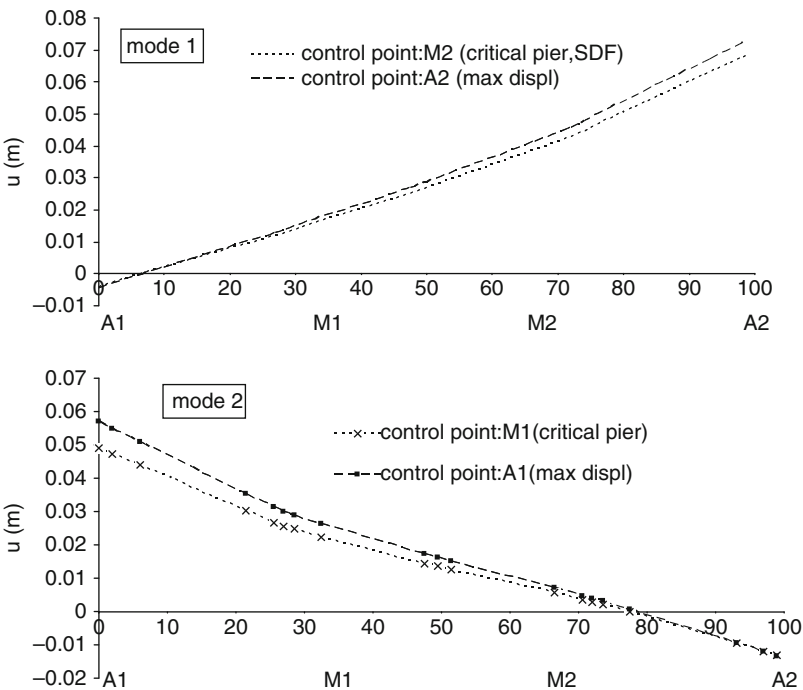
Figure 23 illustrates the deck displacements of the bridge derived using pushover analysis for each mode independently, as well as the MPA procedure initially proposed in Paraskeva et al. (2006). If the structure remains elastic for the given earthquake intensity, both spectral displacement S_d and the product $\Gamma_n \cdot \Phi_n$ will be independent of the selection of the control (monitoring) point; this means that deck displacements are independent of the location of the monitoring point. On the contrary, it was found that deck displacements derived with respect to different control points, for *inelastic* behavior of the

structure, are not identical but rather the estimated deformed shape of the bridge depends on the monitoring point selected for drawing the pushover curve for each mode.

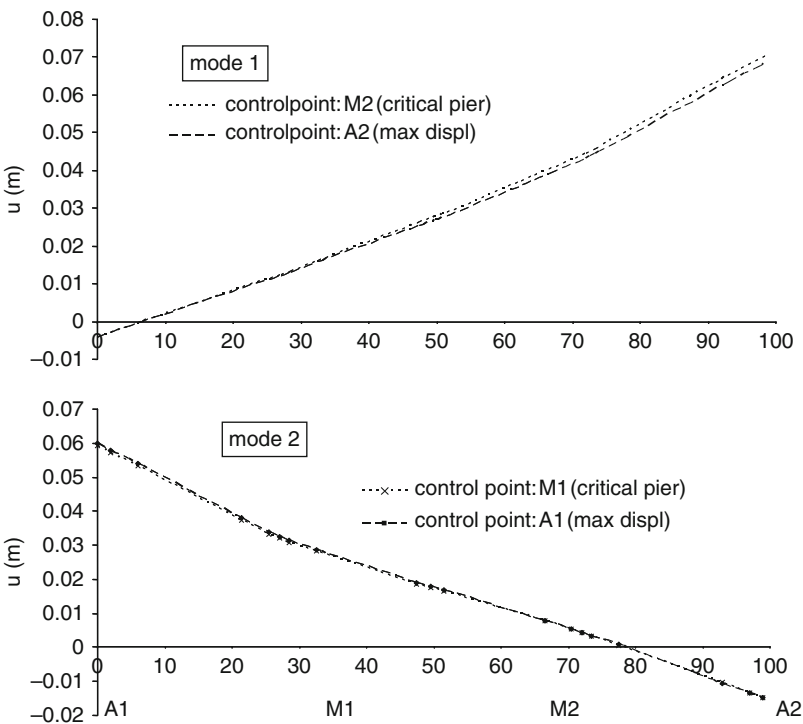
For inelastic behavior, it appears that the estimated values of u_{cn} are different not only because of the deviation of the elastic mode shape Φ_n from the actual deformed shape of the structure but also due to the fact that the spectral displacement S_d is dependent on the selection of monitoring point if the structure exhibits inelastic behavior (due to the bilinearization of the capacity curve). An improved target displacement of the monitoring point is calculated (from Eq. 18b) using Φ'_n , the actual deformed shape of the structure (see Fig. 23), while the spectral displacement remains the same. The response quantities of interest are evaluated by extracting from the “database” the values of the desired responses, r_n , for the analysis step at which the displacement at the control point is equal to u_{cn}' (the improved estimate of u_{cn} derived on the basis of Φ'_n).

Figure 24 illustrates the deck displacements of the overpass bridge, calculated from MPA using u_{cn}' as target displacement for each mode. It is noted that, due to the approximations involved in the capacity and demand spectra procedure, deck displacements derived with respect to different control points are not the same, but differences are significantly reduced and results are deemed

Seismic Analysis of Concrete Bridges: Numerical Modeling,
Fig. 23 “Modal” deck displacements derived with respect to different control points – *inelastic* behavior of the overpass bridge ($A_g = 0.16\text{ g}$)

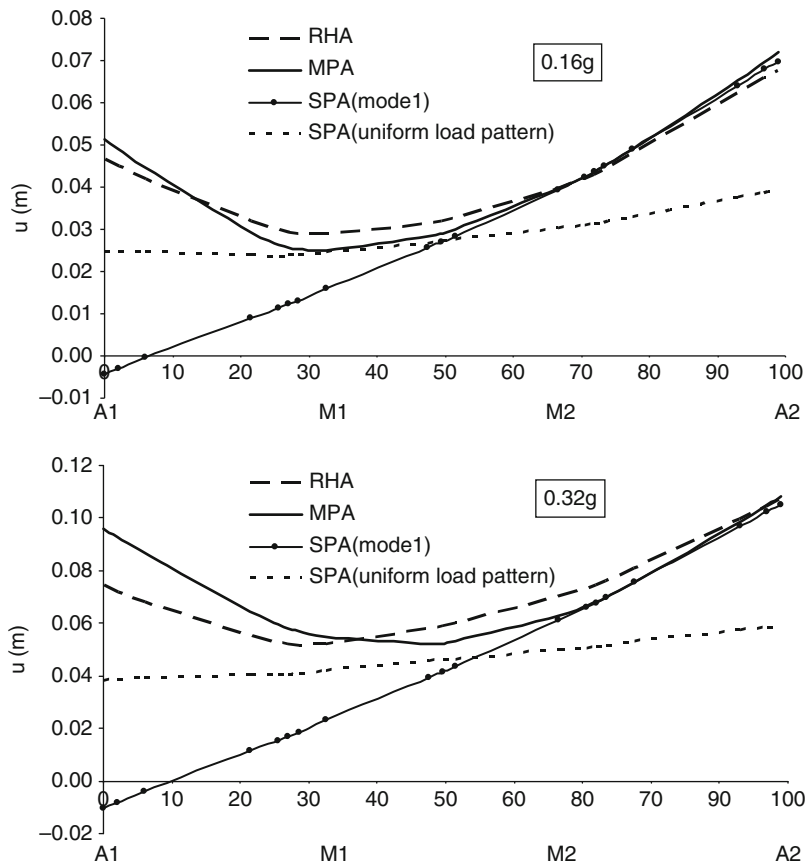


Seismic Analysis of Concrete Bridges: Numerical Modeling,
Fig. 24 “Modal” deck displacements derived with respect to different control points using u_m' as target displacement according to the *improved* MPA procedure – overpass bridge ($A_g = 0.16\text{ g}$)



Seismic Analysis of Concrete Bridges: Numerical Modeling,

Fig. 25 Response to the design earthquake ($A_g = 0.16$ g) and to twice the design earthquake ($A_g = 0.32$ g) calculated from SPA, MPA, and NRHA: deck displacements of the overpass bridge



acceptable for all practical purposes. Paraskeva and Kappos (2010) also studied other bridges with different configurations and noted that the differences between deck displacements derived with respect to different control points, as well as the improvement in the prediction of deck displacements using the procedure described in section “Nonlinear Static Analysis of Concrete Bridges,” are more significant in bridges longer than the overpass of Fig. 2, especially if the increased length is combined with significant curvature in plan, which amplifies the complexity of its dynamic behavior and results in more significant contribution of the higher modes.

Reliability of Static Analysis Procedures

Results of the standard and modal pushover approaches were evaluated by comparing them with those from nonlinear response-history

analysis, the latter considered as the most rigorous procedure to estimate seismic demand. To this effect, a series of NRHAs was performed using five artificial records compatible with the design elastic spectrum. The Newmark $\gamma = 1/2$, $\beta = 1/4$ integration method was used, with time step $\Delta t = 0.0025$ s and a total of 10,000 steps (25 s of input). A uniform damping value of 5 % was assumed for all modes of vibration, while hysteretic damping was accounted for through the elastoplastic behavior of the structural members.

The displacements determined by the SPA and MPA procedures were checked against those from NRHA for increasing levels of earthquake excitation, as shown in Fig. 25. It is noted that the deck displacements shown in the figures as the NRHA case are the average of the peak displacements recorded in the structure during the five response-history analyses. In this study the

displacement demand was estimated independently in static and dynamic (response-history) inelastic analysis, whereas in some previous studies comparisons of displacement profiles were made assuming the same maximum displacement in both cases; the choice adopted here is deemed as more relevant for practical applications, as it permits an evaluation of all aspects of the proposed procedure (including the uncertainty in estimating displacement demand in pushover analysis).

As shown in Fig. 25, the MPA procedure predicts well (i.e., matches closely the values from the NRHA approach) the maximum transverse displacement. On the other hand, the SPA procedure underestimates the displacements of the deck at the location of the abutment A1 and the first pier of the bridge, compared to the more refined NRHA approach. This is not surprising if one notes the differences between the first two mode shapes in the transverse direction (Fig. 22), which are strongly affected by torsion (they contribute more than 90 % of the torsional response, as well as over 90 % of the transverse response of the bridge) due to the unrestrained transverse displacement at the abutments (until the 150 mm gap closes), combined with the different stiffness of the two piers caused by their different height. What is essentially achieved by the MPA is the combination of these first two modes (the 3rd transverse mode is not important in this particular bridge), each of which dominates the response in the region of the corresponding abutment. In the case of applying ground motions with twice the design earthquake intensity (also shown in Fig. 25), where the structure enters deeper into the inelastic range and higher mode contributions become more significant (without substantial alteration of the mode shapes), it is noted that the displacement profile derived by the MPA method tends to match that obtained by the NRHA, whereas SPA predictions remain poor. Note that, regardless of earthquake intensity, the uniform loading pattern (also shown in Fig. 25) fails to capture the increased displacements toward the abutments; nevertheless its overall prediction of the displacement profile could be

deemed better than that resulting from using a single modal load pattern.

Additional case studies reported by Paraskeva and Kappos (2010) confirm that SPA predicts well (i.e., matches closely the values from the NRHA approach) the maximum transverse displacement, when applied to bridges of regular configuration, where the higher mode contribution is not significant. In such cases the improvement of the displacements derived by the MPA procedure is not significant even for high levels of earthquake excitation.

Summary and Concluding Remarks

It is clear that today the bridge engineer has at his/her disposal a set of powerful analysis tools that can be used for the seismic design or assessment of any bridge type. The potential of these tools, when properly utilized, was revealed by their success in predicting the response of bridges tested under high levels of earthquake actions that caused substantial amounts of inelasticity; an example was presented in section “[Comparative Case Study](#).” The information on bridge modeling presented in section “[Modeling of Bridge Components](#)” offers to researchers and designers the necessary information regarding the available models for the various parts of the bridge (deck, bearings and shear keys, isolation and energy dissipation devices, piers, foundation members), as well as tools for modeling the dynamic interaction between piers, foundation, and soil, as well as the abutment-embankment-superstructure system. It also provides information on important parameters that help ensuring that inelastic analysis of bridge earthquake response is conducted properly. It has to be emphasized in this respect that the power and versatility of the analysis tools also makes the results particularly sensitive to improper application.

Special emphasis was given to modeling of piers (section “[Piers and Their Foundations](#)”), as these members are both the ones wherein energy dissipation through plastic hinging is intended to

occur (unless a seismic isolation system is used) and those whose inelastic response is relatively easier to model in inelastic analysis (compared, for instance, to the abutment-backfill system or some foundation types or, indeed, some types of joints). Having said this, it is also clear from the material presented in this entry that proper modeling of the other components of the bridge, even those that are typically assumed to remain elastic during the seismic excitation (such as prestressed concrete decks), is also important, since, through their stiffness characteristics, they affect the dynamic characteristics of the bridge and the way seismic actions are transferred to the dissipating zones. Of great importance is also the modeling of the various connections in the bridge system, i.e., those between piers and deck, abutments and deck, and, in the common case (especially in the transverse direction) that the movement of the deck is restrained at the location of the abutment, the proper modeling of the response of the abutment-backfill system. As noted in section “[Abutments and Backfills](#),” in a practical context and when the main objective of the analysis is the response of the bridge itself (rather than that of the surrounding ground), the recommended solution is to carry out an independent analysis of the abutment-embankment system, determine its resistance curves (in all relevant directions), and use them to describe the nonlinear response of the equivalent springs to which the bridge model will be connected. If such an analysis cannot be afforded, the properties of these springs can be defined on the basis of simplified guidelines from the literature. For pier-foundation-soil interaction, the existing literature is more mature and often it is not necessary to carry out separate analysis of the system to derive the nonlinear properties of the soil-foundation dynamic impedance to be introduced in the bridge model, especially when surface foundations are used; in these cases information from the literature can be used to account approximately for the interaction with the surrounding ground.

Regarding the feasibility and reliability of different methods used for the analysis of concrete

bridges, these depend primarily on the configuration of the bridge analyzed. As a rule, the longitudinal direction of the bridge is the easier one to analyze, and even simple, equivalent static elastic methods can lead to a reasonable design. In most other cases dynamic analysis is required to properly capture the higher mode effects that are important, especially in the areas close to the abutments. When the expected plastic mechanism does not involve more-or-less uniform yielding in the energy dissipation zones of the bridge (this is the case of irregular pier configurations), inelastic analysis is strongly recommended for verifying the design initially carried out using standard modal (response spectrum) analysis. Nonlinear analysis is clearly the preferred choice in the case of assessing existing bridges not properly detailed for seismic performance; for such bridges the simplest choice is standard pushover analysis, but whenever more than one mode affects the response (this is very often the case in the transverse direction of the bridge) more sophisticated tools like the multimodal pushover method presented in sections “[Nonlinear Static Analysis of Concrete Bridges](#)” and “[Comparative Case Study](#)” have to be used. The use of such analytical tools is expected to increase when the software required for applying them in a single run becomes widely available.

Cross-References

- [Buildings and Bridges Equipped with Passive Dampers Under Seismic Actions: Modeling and Analysis](#)
- [Seismic Analysis of Steel and Composite Bridges: Numerical Modeling](#)

References

- AASHTO (American Association of State Highway and Transportation Officials) (2010) LRFD bridge design specifications, 5th edn. American Association of State Highway and Transportation Officials, Washington, DC

- ACI (American Concrete Institute) Committee 341 (2014) Seismic analysis and design of concrete bridge systems (ACI341.2R-14). American Concrete Institute, Detroit
- ASCE (American Society of Civil Engineers) (2007) Seismic rehabilitation of existing buildings – ASCE standard 41-06. American Society of Civil Engineers, Reston
- ATC (Applied Technology Council, USA) (1996) Improved seismic design criteria for California bridges: provisional recommendations. ATC report no ATC-32. Applied Technology Council, Redwood City
- Aviram A, Mackie KR, Stojadinović B (2008) Guidelines for nonlinear analysis of bridge structures in California. PEER report 2008-03. University of California, Berkeley
- Bozorgzadeh A, Megally S, Restrepo JI, Ashford SA (2006) Capacity evaluation of exterior sacrificial shear keys of bridge abutments. *J Bridge Eng ASCE* 11(5):555–565
- Caltrans (California Department of Transportation) (2013) Seismic design criteria version 1.7. Caltrans Division of Engineering Services, Sacramento, California
- CEN (2005a) Eurocode 8: design of structures for earthquake resistance – part 2: bridges (EN 1998-2). CEN, Brussels
- CEN (2005b) Structural bearings – part 3: elastomeric bearings (EN 1337-3). CEN, Brussels
- CEN (Comité Européen de Normalisation) (2004) Eurocode 8: design of structures for earthquake resistance – part 1: general rules, seismic actions and rules for buildings (EN1998-1-1). CEN, Brussels; Corrigenda: July 2009 and Jan 2011
- Chopra AK, Goel RK (2002) A modal pushover analysis procedure for estimating seismic demands for buildings. *Earthq Eng Struct Dyn* 31(3):561–582
- Computers & Structures Inc. (CSI) (2011) SAP2000 – version 15.0.1: linear and non linear static and dynamic analysis and design of three-dimensional structures. Computers & Structures Inc. (CSI), Berkeley
- Fajfar P (1999) Capacity spectrum method based on inelastic demand spectra. *Earthq Eng Struct Dyn* 28(9):979–993
- Fardis MN (1991) Member-type models for the non-linear seismic response analysis of reinforced concrete structures. In: Donea J, Jones PM (eds) *Experimental and numerical methods in earthquake engineering*. CEC, JRC, Ispra
- FEMA (1997) NEHRP guidelines for the seismic rehabilitation of buildings, FEMA-273. FEMA, Washington, DC
- FHWA (Federal Highway Administration, USA) (2006) Seismic retrofitting manual for highway bridges part 1 – bridges (FHWA- HRT-06-032). Turner-Fairbank Highway Research Center, McLean
- Hwang JS (2002) Seismic design of structures with viscous dampers. In: *International training program for seismic design of building structures*. NCREE, Taiwan
- Kappos AJ, Sextos AG (2001) Effect of foundation type and compliance on the lateral load response of R/C bridges. *J Bridge Eng ASCE* 6(2):120–130
- Kappos AJ, Sextos AG (2009) Seismic assessment of bridges accounting for nonlinear material and soil response, and varying boundary conditions. In: *Coupled site and soil-structure interaction effects with application to seismic risk mitigation*, NATO science for peace and security, series-C. Springer, Dordrecht, pp 195–208
- Kappos AJ, Manolis GD, Moschonas IF (2002) Seismic assessment and design of R/C bridges with irregular configuration, including SSI effects. *Eng Struct* 24(10):1337–1348
- Kappos AJ, Saiidi M, Aydinoglu N, Isakovic T (2012) Seismic design and assessment of bridges: inelastic methods of analysis and case studies. Springer, Dordrecht
- Kappos AJ, Gkatzogias KI, Gidas I (2013a) Extension of direct displacement-based design methodology for bridges to account for higher mode effects. *Earthq Eng Struct Dyn* 42(4):581–602
- Kappos AJ, Paraskeva TS, Moschonas IF (2013b) Response modification factors for concrete bridges in Europe. *J Bridge Eng ASCE* 18(12):1328–1335
- Kappos AJ, Papadopoulos I, Tokatlidis A (2014) Design of a seismically isolated railway viaduct over axios river in Northern Greece. In: *Railways 2014: the second international conference on railway technology*, Ajaccio, Corsica, 8–11 Apr, Civil-Comp Press, paper no 80
- Katsaras CP, Panagiotakos TB, Koliass B (2009) Effect of torsional stiffness of prestressed concrete box girders and uplift of abutment bearings on seismic performance of bridges. *Bull Earthq Eng* 7:363–375
- Mergos PE, Kappos AJ (2012) A gradual spread inelasticity model for R/C beam-columns, accounting for flexure, shear and anchorage slip. *Eng Struct* 44:94–106
- Notsios E, Karakostas C, Lekidis V, Panetos P, Nikolaou I, Papadimitriou C, Salonikios T (2009) Structural identification of Egnatia Odos bridges based on ambient and earthquake induced vibrations. *Bull Earthq Eng* 7(2):485–501
- Papanikolaou VK, Kappos AJ (2009) Numerical study of confinement effectiveness in solid and hollow reinforced concrete bridge piers: Methodology. *Comput Struct* 87(21–22):1427–1439
- Paraskeva TS, Kappos AJ (2010) Further development of a multimodal pushover analysis procedure for seismic assessment of bridges. *Earthq Eng Struct Dyn* 39(2):211–222
- Paraskeva TS, Kappos AJ, Sextos AG (2006) Extension of modal pushover analysis to seismic assessment of bridges. *Earthq Eng Struct Dyn* 35(10):1269–1293
- Priestley MJN, Seible F, Calvi GM (1996) *Seismic design and retrofit of bridges*. Wiley, New York
- Vayas I, Iliopoulos A (2014) *Numerical modeling guidelines for seismic analysis of steel and composite*

bridges. In: Encyclopedia of earthquake engineering. Springer, Berlin

Zhang J, Makris N (2002) Kinematic response functions and dynamic stiffnesses of bridge embankments. *Earthq Eng Struct Dyn* 31(11):1933–1966

Seismic Analysis of Masonry Buildings: Numerical Modeling

Christos A. Zeris and Elizabeth Vintzileou
Department of Structural Engineering, Faculty of
Civil Engineering, National Technical University
of Athens, Athens, Greece

Synonyms

Adobe; Confined masonry; Finite elements; Load-bearing masonry; Masonry building reinforced concrete; Reinforced masonry; Seismic analysis; Unreinforced masonry

Introduction

Masonry building construction encompasses the large inventory of structures built worldwide solely or partially of individually laid masonry units bonded or not together with some type of mortar, possibly with the incorporation of reinforcement; depending on the constituent materials, stacking, and bonding techniques, different technologies are adopted for building the masonry elements that comprise the entire building; thus, materials and techniques adopted worldwide vary with local customs, socioeconomic conditions, and available technology. A comprehensive continuously updated description of the different materials and technologies of construction around the world is given in the World Housing Encyclopedia (EERI/IAEE).

Because of their widespread use, masonry structures house not only the population but also important social and economic operations such as hospitals and schools, high congregation areas, business, small to medium industry, and civil

administration (UNIDO 1983); furthermore, being historically one of the earliest forms of construction, they also house the people's cultural heritage, most often being such by themselves (Figs. 1, 2, 3, and 4).

Due to their geographic spread, masonry structures are exposed to different levels of seismic hazard; in fact, a significant portion of cultural heritage structures are located in earthquake prone-areas of Europe, Asia, and South America. Recent devastating earthquakes (e.g., in New Zealand, Italy, Chile, India, Pakistan, and elsewhere) have shown that existing masonry structures are quite vulnerable to seismic actions, as also recorded in many seismic damage reconnaissance reports published after major earthquakes (e.g., among others, Hughes et al. 1990; Rossetto et al. 2009; DesRoches and Comerio 2011).

As a consequence of this fact, social and economic requirements for human safety and operability as well as the need for preservation of cultural heritage require more and more often that these structures be analyzed for seismic actions, in order to be designed (new construction), or their seismic vulnerability can be assessed and evaluated (existing construction) for the purpose of repair, rehabilitation, and/or strengthening to current seismic standards. The seismic performance assessment of existing masonry buildings also follows performance-based design (PBD) and analysis procedures, similar to other types of structures (e.g., concrete and steel buildings): hence, in the design of new construction, normative regulations, practices, and experience of good seismic performance have been encompassed (e.g., EC6 2005 for gravity load design and EC8 2004 for seismic design). On the other hand, codes and guidelines are being drafted for the assessment and retrofit of existing masonry buildings (e.g., FEMA-356 2000) or for cultural heritage structures (Moro 2007; ICOMOS; ISCARSAH).

Compared to the more recently evolved types of construction like steel and reinforced concrete, however, masonry buildings have certain inherent idiosyncrasies, which their modeling for seismic analysis should account for:



Seismic Analysis of Masonry Buildings: Numerical Modeling, Fig. 1 The stone masonry castle of Mycenae, Peloponnese, Greece, 2nd millennium BC (Source <http://en.wikipedia.org/wiki/Mycenae>)



Seismic Analysis of Masonry Buildings: Numerical Modeling, Fig. 3 The Alcantara bridge (104 AD), Spain (Source: http://en.wikipedia.org/wiki/Roman_bridge)



Seismic Analysis of Masonry Buildings: Numerical Modeling, Fig. 2 The Byzantine church of Hagia Sophia (360 AD), Istanbul, Turkey



Seismic Analysis of Masonry Buildings: Numerical Modeling, Fig. 4 The Aaos river bridge (eighteenth century), Hepirus, Greece

- (i) Masonry buildings are both non-engineered, namely, structures built with traditional means and experience, and engineered, namely, structures designed and constructed following engineering principles and, more recently, code regulations (e.g., EC6 2005; EC8 2004). In fact, earlier engineered masonry buildings have been designed for gravity loads only; only some types of recent masonry construction, following the evolution of seismic regulations, have also been designed for earthquake (IAEE 2008;

ASCE 2013; EC6 2005). In addition to the possible lack of proper design, they are also characterized by the usual problems in the load-bearing system similar to other structures, such as: irregularities in plan and/or elevation, improper foundation conditions, a history of (possibly undocumented) modifications in plan and elevation, and the decay of the material properties under environmental exposure.

- (ii) Depending on the prevailing socioeconomic conditions, masonry buildings are and have



Seismic Analysis of Masonry Buildings: Numerical Modeling, Fig. 5 Residential masonry buildings in Greece, built in the nineteenth and early twentieth century, reflecting different levels of socioeconomic status

been constructed using a variety of locally available technologies and materials (Fig. 5). Often, currently acceptable levels of quality control and supervision were not enforced. New masonry buildings in developed countries may use units manufactured industrially and with quality control; this is not the case for existing buildings in all countries or new masonry construction in underdeveloped countries (see typical examples offered in the World Housing Encyclopedia, [EERI/IAEE](#)).

- (iii) Masonry, unless suitably reinforced or confined, as in the case of new construction, cracks and fails in a brittle manner. Seismic analysis methods for existing or historical unreinforced masonry buildings need to take this fact into account.
- (iv) Masonry buildings suffer from structural inadequacies in the load-bearing system inherent in this type of construction, such as: the presence of diaphragms which are

poorly connected to the masonry for lateral load transfer, the relatively high flexibility of the floor diaphragms, and the way the masonry wall elements are constructed and tied together through the thickness and/or at their intersections. Seismic analysis modeling techniques should not overlook these particularities, if the analysis results are to be reliable.

Depending therefore on the problem at hand, reliable modeling for seismic analysis of masonry structures will have to account of these building characteristics, in order to reliably predict the damages expected and to identify the methods and extent of intervention required for strengthening these structures. Following a brief review of masonry building characteristics, the methods and limitations of different methodologies adopted for the seismic analysis of masonry construction are subsequently considered.

Modeling Techniques for Seismic Analysis of Masonry Buildings: Classification and Definitions

Masonry Materials Used

According to the type of masonry units employed, masonry is classified as stone masonry (Fig. 5), industrially manufactured clay brick (solid, hollow), site-produced mud brick or plinth (also known as adobe masonry), or industrially manufactured hollow or solid concrete masonry units (CMU) in fully, partially grouted, or ungrouted construction. Other industrially manufactured types of block are also used worldwide such as lightweight concrete, cinder, fly ash, or autoclaved concrete, among others. Depending on the type of construction, reinforcement is also used to enhance the bearing capacity of masonry elements and to provide ductility to the structure. Another construction system (covered by modern codes, e.g., EC6 2005) is that of confined masonry: horizontal and vertical reinforced concrete elements are provided, at distances depending on the dimensions of the building and on the seismicity of the region. The ties, constructed during the construction of the masonry, function as linear tensioned members.

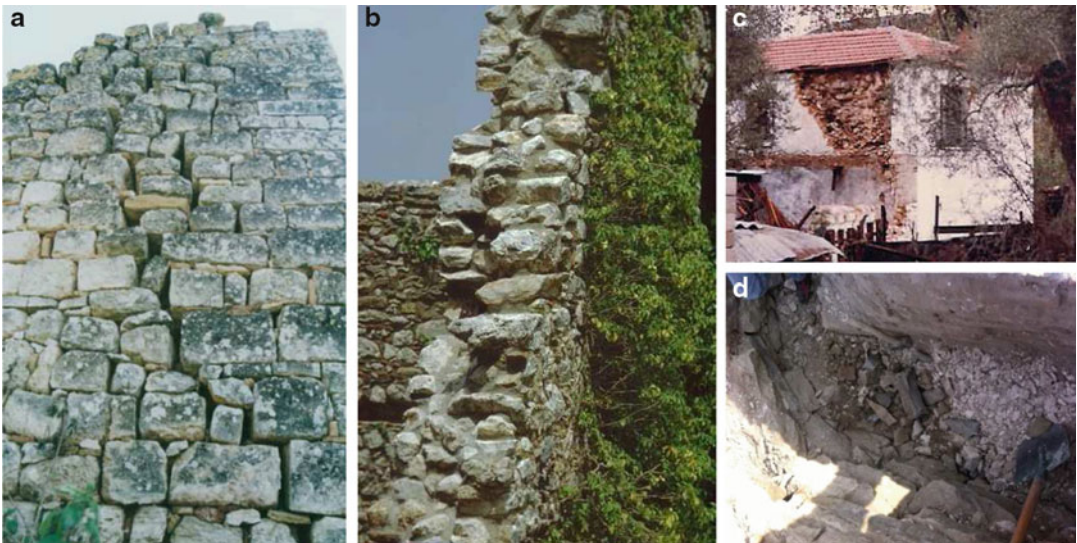
They assist masonry in taking shear and bending (in and out of plane) while they also contribute to the confinement of plain masonry, thus enhancing the ductility of the system.

In cultural heritage buildings in Europe and elsewhere, most frequent construction types of stone masonry are double-leaf masonry, made of two leaves, either independent or connected between them with sporadic header stones (Fig. 6), and three-leaf masonry, made of two independent leaves, with the space between them being filled with a more or less loose material of poor mechanical properties. Other types of masonry, like cavity masonry or timber reinforced one, are also quite frequent in historic structures.

Finally, the material that bonds the masonry units together may be none (also called dry construction), mud, or different types of mortar such as lime, lime–pozzolan, lime mortar reinforced with animal hair, or, in case of modern masonry construction, cement or lime–cement.

Load-Bearing Function of the Masonry

The type of load path of the vertical and transverse loads down to the foundation defines two basic types of masonry:



Seismic Analysis of Masonry Buildings: Numerical Modeling, Fig. 6 Unreinforced masonry construction technologies. (a) Dry masonry construction (Tiryns,

Argos, -Peloponnese, thirteenth to fourteenth century BC). (b) Double-leaf masonry wall. (c) Three-leaf rubble stone masonry. (d) Cavity wall (Van, Turkey)

- (i) **Load-bearing (LB) masonry**, bearing all the gravity and lateral loads from floors and roof to the foundation. Thus, the masonry elements themselves are responsible to provide overall lateral stability and to transfer the seismic inertia loads under in-plane and out-of-plane actions, down to the foundation, through combined flexural, axial, shear, or rigid body rocking mechanisms. Seismic analysis of this type of masonry involves adequate modeling of the entire load-bearing structural system response at the individual element level, including the diaphragms and the foundation, accounting for both in-plane and out-of-plane response, as discussed in the next section. As a special case of LB masonry construction, masonry arches are constructed with keystone elements at the apex, in order to provide the path of the line of thrust to the foundation.
- (ii) **Non-load-bearing (NLB) masonry**, in which the bearing function is provided by a structural skeleton made of other structural materials such as steel and timber or reinforced concrete (RC). Such is the case of the widely adopted masonry panel infilled RC building construction, a form of confined masonry, whose modeling is covered in section “[Modeling of NLBM Infilled Frame Buildings for Seismic Analysis.](#)”

If anything else, NLBM infill panels have to support with adequate resistance capacity and in a stable manner their own inertia forces as these are materialized at the building elevation relative to ground in which they are supported. Furthermore, through the deformation compatibility between the infill and the confining frame structure, they are forced to resist seismic load through a friction contact separation mechanism with the frame under lateral load response. Therefore, unless specifically isolated by adequate details from the bearing frame (over the entire lateral inelastic deformation expected), NLBM panels are seismically bearing, and consequently, even NLBM elements need to be included in seismic analysis of the entire building in which they are constructed.

Because of this contact mechanism, practical seismic analysis of this type of masonry construction involves macromodeling using axial load-bearing struts, as briefly discussed in section “[Modeling of NLBM Infilled Frame Buildings for Seismic Analysis.](#)”

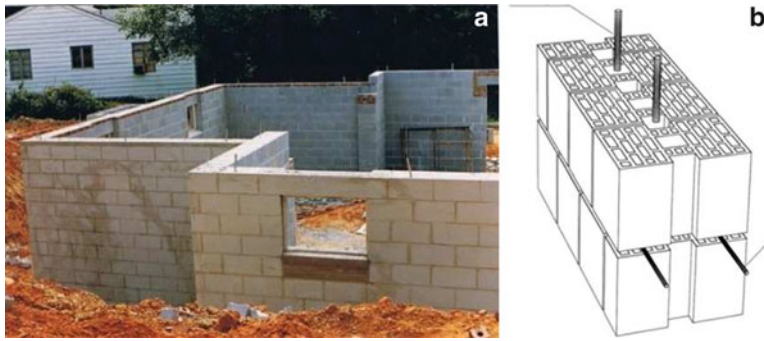
Masonry Wall Construction Types and Resisting Mechanisms

Masonry walls, whether LBM or NLBM, range from single-leaf or multiple-leaf walls (Fig. 6) with interior cavities among the leaves, which are either void (e.g., the common double-leaf walls with space for sliding window frames) or are filled with rubble (typical for old masonry buildings, cultural heritage buildings included) or concrete (in modern masonry construction of reinforced cavity walls according to EC6 2005). Multiple-leaf walls may be tied to each other in modern construction (e.g., modern veneer brick masonry walls) or, as often the case in existing structures, they are untied.

The laying of the masonry units varies according to the local techniques, the material, and the production form of the unit. Industrial units are laid in courses (use of header courses or random laying construction and keystones at the intersections) and the provision of collar joints filled with mortar (and possibly steel) and/or the inclusion of bond beam elements. The vertical elements are traditionally built on a stone or brick or concrete footings, on which they rest or are tied to with reinforcement. Often, interior walls are in fact lighter construction partitions of brick or wood not tied to the load-bearing system, or the wall intersections were poorly connected.

Irrespective of material and load-bearing type, masonry is classified according to its resisting mechanisms as unreinforced masonry, reinforced masonry (prestressing possibly included), and confined masonry, depending on whether reinforcement or additional confining elements within the masonry are used. The use of each type in new constructions in seismic regions depends on the seismicity of the region, on the number of stories, etc. (see, e.g., EC8 2004).

- (i) **Unreinforced masonry (URM)**, namely, masonry without any or very small amounts



Seismic Analysis of Masonry Buildings: Numerical Modeling, Fig. 7 Reinforced masonry construction technology and detail. (a) Modern CMU reinforced masonry residential building (Source: [http://en.](http://en.wikipedia.org/wiki/Masonry_structure)

[wikipedia.org/wiki/Masonry_structure](http://en.wikipedia.org/wiki/Masonry_structure)). (b) Reinforced masonry construction detail using clay bricks with improved thermal insulation properties

of reinforcement included and without any additional confining members within the masonry body. This form of masonry construction is the most frequently encountered type of construction in existing or historical buildings and is associated with the largest amount of modeling problems, due to the variety of techniques and materials adopted and the variability of their properties. Seismic analysis methods for URM buildings need to account for the mechanical characteristics of the material, obtained from in situ evaluation using flat jacks or through testing (e.g., Clough et al. 1979; Magenes et al. 2008; Vintzileou and Miltiadou-Fezans 2008; Ruiz-Garcia and Negrete 2009; and the TCCMaR test series, Kingsley 1994); furthermore, modeling should account for the brittle nature of its response. Only elastic methods of analysis are therefore meaningful in their seismic performance assessment, whereas a low behavior factor (~ 1.50) is applicable for evaluation of the seismic forces.

- (ii) **Reinforced masonry (RM)** is masonry with horizontal reinforcement in mortar bed joints and vertical reinforcement positioned in a cavity or in holes of the vertically perforated masonry units or in grooves between adjacent blocks (Fig. 7a, b).

In modern construction, reinforcement comprises steel reinforcing in the form of bars, trusses or cut wire mesh, dovetails,

and other special shaped proprietary ties, placed vertically and horizontally. Other reinforcing materials include prestressing strands (prestressed masonry) and polymers, used for external strengthening existing masonry. In addition to steel reinforcement used in modern construction, timber elements have traditionally been used in historical masonry structures and are still used in new buildings constructed in seismic regions in the developing world, with very good seismic performance characteristics (also called timber-laced buildings, Figs. 8 and 9).

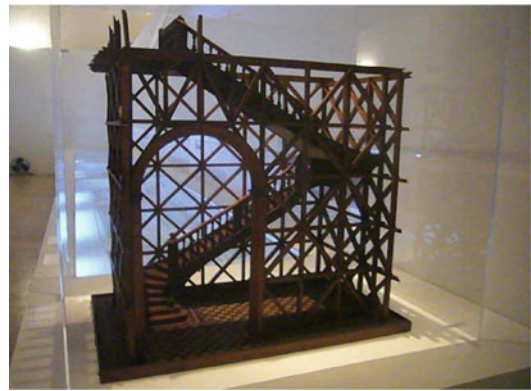
Generally, reinforcement serves to connect the leaves of the masonry wall, to allow for uniform distribution of vertical loads at floors and roof levels, to act as shear and flexural reinforcement for in-plane and out-of-plane seismic load transfer, and also to provide ductility, when reinforcement yielding precedes brittle failure of the masonry; therefore, its beneficial role should be included in seismic analysis; furthermore, for new construction, when ductility is enforced by design, both elastic and inelastic methods of analysis are meaningful in seismic performance assessment, with a suitable response reduction coefficient applied in the expected seismic loads; in this context, behavior factors up to 3.0, comparable to RC wall construction, can be used in modern RM building design (EC6 2005).



Seismic Analysis of Masonry Buildings: Numerical Modeling, Fig. 8 Traditional timber-laced stone masonry buildings in earthquake-prone areas. (a) Antalya, Turkey. (b) Kastoria, Greece



Seismic Analysis of Masonry Buildings: Numerical Modeling, Fig. 9 Bhatar construction: traditional unreinforced load-bearing timber-laced masonry structure in Pakistan in modern construction (Source: <http://www.holcimfoundation.org/Projects/advocacy-of-traditional-earthquake-resistant-construction-north>)



Seismic Analysis of Masonry Buildings: Numerical Modeling, Fig. 10 Model of the Gaiola pombalina masonry confinement with timber, developed in Lisbon after the 1755 earthquake ([http://en.wikipedia.org/wiki/Gaiola_\(construction\)](http://en.wikipedia.org/wiki/Gaiola_(construction)))

- (iii) **Confined masonry (CM)** is masonry which encompasses during construction horizontal and/or vertical confining RC or RM elements, monolithically bonded into the masonry structure (e.g., RC lintel beams and pilasters). In fact, modern clay brick walls in seismic regions make use of such horizontal and vertical RC elements. In the case of RC confining elements, similar modeling conventions for seismic analysis are adopted as for RM buildings, using however higher seismic forces; in this case, a behavior factor of 2.0 is adopted in EC8 (2004).

A special form of URM and LBM construction originating in older (historic) buildings (a well-known case being the Pombalinas, constructed during the rebuilding of Lisbon following the devastation of the city on 1755) and still adopted in less developed countries (e.g., the *dhajji* in Pakistan) makes use of timber reinforcing elements in the form of tension–compression braces; they are built within the masonry walls at the floor levels or between floors, in order to confine the masonry infilling the voids (Fig. 10), thereby enhancing its in-plane shear resistance and providing damping under seismic response. This type of LBM construction is also behaving as CM and partially RM, due to the load-carrying mechanism provided by the timber elements and

the in-plane stabilization and stiffening role provided by the masonry infill, confined in turn by the timber structure; seismic modeling of such systems follows the macromodel approach (see section “[FE Modeling of LBM Structures for Seismic Analysis](#)”), whereby the confining timber elements are included in the model together with the masonry and are verified accordingly.

Seismic Analysis Methods of Masonry Buildings

Depending on the problem at hand, both linear and nonlinear analysis methods, as also employed for seismic analysis of RC and steel buildings, are being applied to LBM buildings according to the limit state objective of the verification and the ductility capacity of the masonry elements (if any).

Linear Elastic Analyses

For serviceability limit state (cracking, service load deformation) verifications, linear analysis methods are adequate, up to the onset of cracking of the masonry. The use of linear elastic analysis for the ultimate limit state (strength) design verification of URM buildings under seismic load combinations provides meaningful verification results only under seismic load levels obtained using a response reduction coefficient equal to 1.0; for the design of new RM buildings, it is possible to use linear elastic methods with higher behavior factors (up to 3.0), due to the inherent ductility capacity of the masonry. It should be noted, however, that in case of historic buildings, lower seismic actions than those prescribed by modern codes are frequently adopted. Actually, the application of the requirements of current codes may lead to invasive interventions that are against the internationally accepted charters for the preservation of the built cultural heritage (e.g., [ICOMOS](#); [ISCARSAH](#)).

Linear elastic analysis methods can also be adopted for ultimate limit state verifications of URM buildings, provided the structure does not have excessive irregularities ([EC6 2005](#)) and lacks significant torsional effects, thereby

exhibiting clearly separated orthogonal modes of vibration. Linear elastic modeling neglects cracking of the masonry elements and its influence in the kinematics and the redistribution of forces. Even if cracking is neglected, however, linear elastic models provide useful information about the structure and the model adopted since: (i) they give an indication of the areas of increased tensile or compressive stress in the masonry, which potentially need to be strengthened or rehabilitated; (ii) they can be easily compared with a visual or in situ measurement of cracking and stresses in the masonry, thus allowing to verify the reliability of the model and the reasons for the existing condition of the structure; and (iii) possible interacting factors of overstress can be established (e.g., earthquake following a long-term preexisting foundation settlement).

Static and Modal Elastic Analyses

Linear elastic analysis includes both equivalent static and modal analysis:

- (i) Static analysis can be used in buildings in which higher modes are not dominant in the response, the building is orthogonal in plan, and it does not exhibit major irregularities: two different lateral load distributions with height should be considered, representative of different modes of lateral deformation, namely, (a) triangular distribution and (b) uniform distribution. In the case of micromodels (see section “[FE Modeling of LBM Structures for Seismic Analysis](#)”), possible refinements on the linear elastic model can be considered through local modification of the finite element (FE) stiffness characteristics in order to account for cracking of the masonry, whether such is predicted from an elastic analysis or it is obtained from field inspection of the condition of the structure.
- (ii) Modal analyses, when used, should include all the modes with a modal mass that is greater than 5 % of the total mass of the building and enough modes whose modal masses sum to at least 75 % of the total building mass.

For the evaluation of the modal response characteristics – deformations and internal forces – using macromodels (see section “[FE Modeling of LBM Structures for Seismic Analysis](#)”), the square root of the sum of squares combination rule of modal quantities can be used if the modes differ with each other by as much as 90 %, or, better, using the complete quadratic combination rule, giving accurate maxima for closely coupled mode combinations. For seismic modal analyses using micromodels (e.g., shell finite elements), the peak response characteristics (deformations and internal stresses) $E_{\max}$ should be evaluated following:

$$E_{\max} = \max_t \left\{ \sum_{j=1}^n E_j(t) \right\} \quad (1)$$

where $E_j(t)$ is the time history of the corresponding parameter due to response in the j th eigenmode.

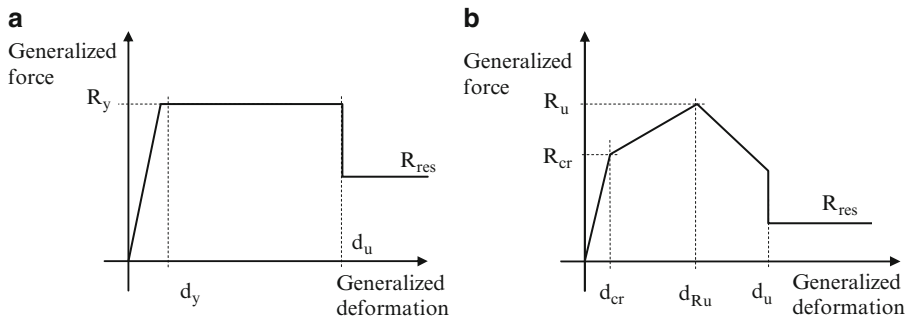
Inelastic Analyses

In currently accepted methodologies for the assessment of existing LBM buildings (both RM or URM) in the context of PBD, nonlinear static analysis methods and corresponding modeling conventions are employed with different levels of modeling detail, depending on whether cracking, post-ultimate, and cyclic hysteretic characteristics are included in the model (Fig. 11). The purpose of these methodologies is to evaluate the inelastic lateral load deformation

of the building, taking into account the actual inelastic characteristics of the elements, brittle or ductile; evaluation of this capacity curve yields the expected target deformation demands, under different seismic excitation levels (performance levels), at which point the onset, distribution, and extent of structural damages are obtained and compared to code damage levels (FEMA-356 2000), either in terms of element resistance (brittle element response) or in terms of inelastic deformations (ductile element response). For the evaluation of these generalized force–deformation characteristics, different failure mechanisms – accounting also for out-of-plane effects – can be considered, and the weakest governing mechanism should be adopted as governing the failure response.

Modeling of LBM Buildings for Seismic Analysis

The primary load-bearing elements of LBM construction are the vertical load-supporting elements, namely, the floor and roof structure, as well as the perimeter and interior bearing masonry walls. All these elements carry the vertical loads (including self-weight) and the lateral forces to the foundation. Secondary elements, not part of the lateral resisting system (such as light partitions), are not included in the model as earthquake resisting elements. Only their self-weight alone is taken into account in analysis.



Seismic Analysis of Masonry Buildings: Numerical Modeling, Fig. 11 Inelastic analysis constitutive models of masonry building elements for in-plane only action (bending or shear). (a) Simplified. (b) Refined

Floor and Roof Elements and Diaphragmatic Action

Floor structures and the roof structure transfer the vertical loads and in-plane inertia seismic loads to the rest of the bearing elements. While in modern masonry buildings floor elements are typically stiff RC diaphragms, in typical existing or historical masonry construction, light wooden floors on wood or steel purlins have been used traditionally. Brick or stone masonry vaults are also typical in the lower story of several historic structural systems; other variants include arched brick constructions between joists or two-way Zoellner diaphragms made of brick infilled voids. Purlins usually span one way and are simply resting or encased in the bearing walls at each end. A peculiarity of masonry buildings is that masses are not concentrated at floor levels; they are distributed over the height of the building.

Unlike modern construction concrete slabs which provide diaphragmatic action that distributes the inertia loads in plan and tie all the vertical masonry elements at the floor level, existing masonry structures were constructed with flexible diaphragms which deform in plane and operate differently during the earthquake (Fig. 12a). Therefore, the presence or absence of diaphragmatic action and the way the diaphragm is tied to the vertical elements are two important aspects to consider in seismic modeling, since the diaphragm stiffness will affect both the dynamic characteristics of the masonry building and the transfer of forces among the stiff vertical wall elements.

Along the same context, the function of the diaphragm (one-way flexible or two-way stiff action) shall also define the distribution of the floor plan masses to the walls: one-way joist diaphragms will only distribute inertia reaction loads across the walls at which their wooden joists are infamed, making the usual uniform mass distribution assumption in the building model, namely, a lumped rotational/translational mass at the center of mass, incorrect.

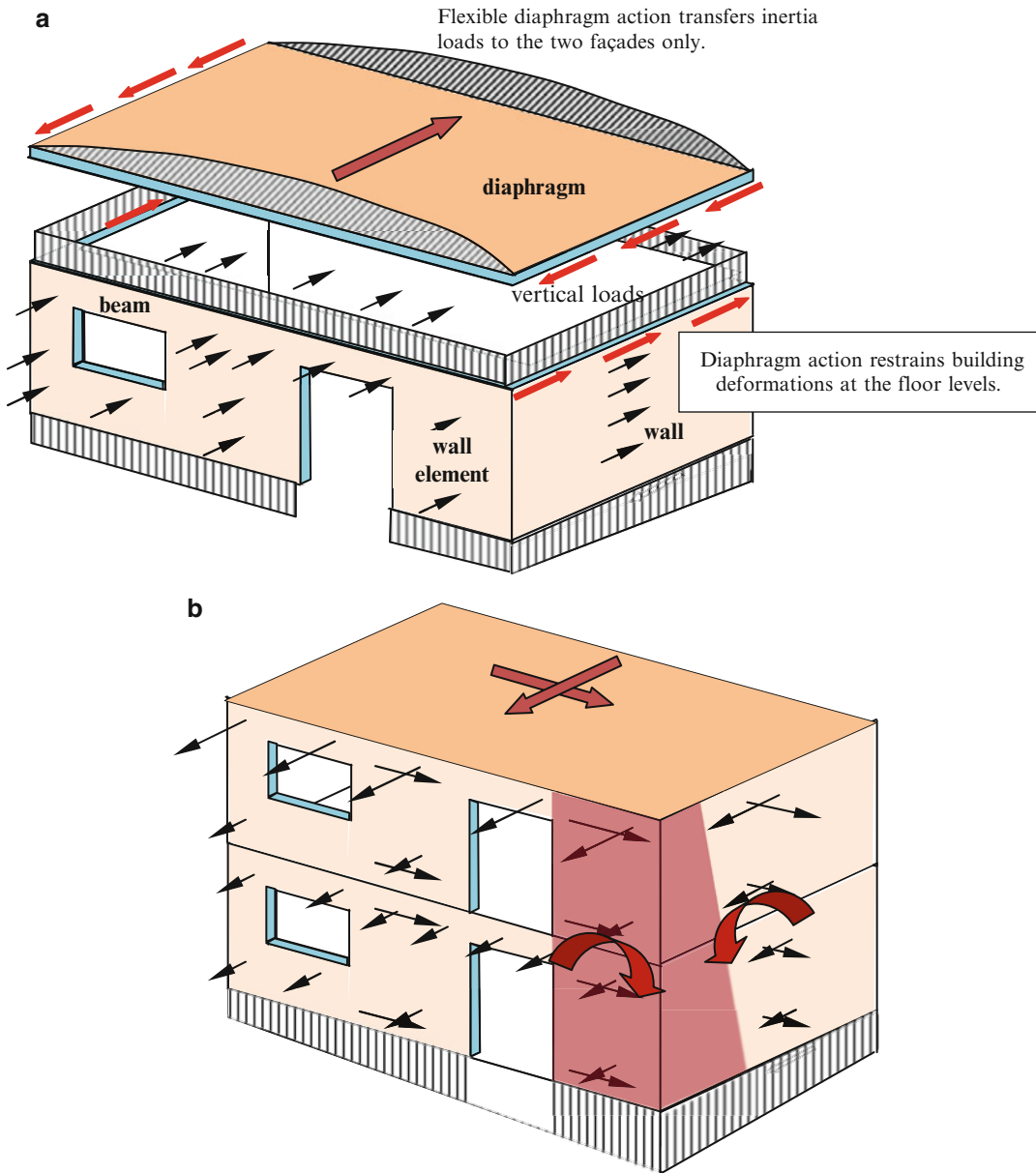
In addition to load distribution, failure of the diaphragm is also important to model: apart from failure of arch-supported diaphragms that tend to be sensitive to relative in-plane deformations of

the supporting walls, failure or collapse of flexible diaphragms takes the form of failure (usually pullout) of the diaphragm-to-wall connection, due to inertia force transfer or excessive out-of-plane deformation of the wall and loss of support of the timber elements to the wall. This type of behavior depends mostly on the vertical support system of the diaphragm and the detailing of the connections, which all need to be part of the building model (see, e.g., Vintzileou et al. 2007, for a description of the system used in the traditional masonry structures in Lefkada, Greece).

Seismic Load-Bearing (Primary) Vertical Elements

The primary lateral load-bearing elements of LBM buildings are the perimeter and interior bearing walls, which are typically perforated with openings, forming wall elements, spandrel beams (also arched lintel beams), and pilasters (Fig. 12b). The in-plane lateral load transfer of the wall elements depends on their aspect ratio (the height to width ratio): walls with relatively longer width compared to the element height (height to width ratio less than 2–3) tend after initial cracking to transfer the lateral force to the lower level directly through an inclined strut (including also the vertical load), whose horizontal component equals the lateral load; more slender walls or multiple wall elements (wallets) created in a wall with openings may opt for a more flexure-dominated behavior (similar considerations apply also to the horizontal spandrel beams between openings), while failure of these elements is brittle.

Since the walls are the primary lateral load resisting elements, their distribution in plan and their stiffness (namely, geometric size and percentage of opening area) determine the eccentricity between the center of application of the floor inertia forces (the center of mass for stiff diaphragms) and the center of rigidity of the building in plan. Consequently, irregularities that may arise in plan and also in elevation due to the wall distribution and geometry will influence the distribution of the lateral loads among the different resisting elements, something which is crucial to model in seismic analysis. It should



Seismic Analysis of Masonry Buildings: Numerical Modeling, Fig. 12 Good practices in the seismic analysis of masonry buildings. (a) Contributions of the seismic forces to the masonry façade in a typical URM building

through diaphragmatic action (where this exists). (b) Contributions of the three-dimensional analysis to the in-plane and out-of-plane actions on the facades and to the corner element forces

further be overemphasized that these structural characteristics are dependent on the intensity of the earthquake, since masonry (like concrete) cracks in tension or softens under extreme compression stresses, and therefore the relative stiffness of the bearing elements changes beyond

what has been assumed in an elastic analysis, redistributing forces as well as increasing the lateral deformation demands during the an earthquake. The analysis sophistication therefore and the modeling capabilities should reflect the level of response prediction.

In-plane response failure of the walls or other structural elements (where present) will take place due to inadequate resistance or excessive interstory drift. Masonry element failure includes wall pier, pilaster, or spandrel beam failure under in-plane actions. Depending on the aspect ratio of these elements and the existence of diaphragmatic action, failure of the wall elements affects individual masonry elements and will jeopardize the entire vertical load path to the lower floors, the building dynamic characteristics, as well as the redistribution of forces in and out of plane. Element failure takes the following forms:

- (i) Flexure-dominated failure including cracking and/or rocking of the wall, compressive toe failure of the wall
- (ii) Shear failure in plan, including sliding at mortar bed joints, diagonal cracking or diagonal crushing of the wall between the cracks
- (iii) Entire wall movement
- (iv) Apart from these in-plane response failure modes, walls may also fail under concurrent in-plane and out-of-plane action. In this case failure is closely associated with the existence or not of adequate diaphragmatic action and a suitable tensile diaphragm connection to the façade walls that will mobilize the entire building façade from the footing line (below grade) to the roof (Fig. 12a). Special cases of out-of-plane buckling failure under in-plane compressive action are also included in this combined failure mode, primarily for slender wall elements though, e.g., unsupported multiple-leaf walls.
- (v) Failure under combined biaxial effects (combined axial load) characterizes the corner walls and pilasters, due to the three-dimensional nature of the building response. Bidirectional rocking of the building (Fig. 12b) induces axial loads that are considerably higher than uniaxial predictions, together with biaxial bending and shear forces. In this case, a three-dimensional model of the building is needed to predict such overstress in the corners, with the results of plane analyses only being unconservative.

Secondary Elements

Such elements are typically the interior masonry (or other) column elements that support the diaphragm or narrow, slender elements on the façade that function as columns. The failure of these elements does not alter significantly the response of the building, and therefore, they are ignored in the seismic analysis model. However, their ability to bear vertical loads under the seismic deformation of the building should still be verified.

Foundation

The foundation of the masonry walls is not necessarily responding as rigid and non-deforming, as typically assumed in seismic modeling. Foundations may settle in the long term under the weight of the walls, inducing tensile cracks in the building that change the distribution of forces and the way the vertical elements respond (rocking rather than flexing). Furthermore, due to the lack of reinforcement, contact type of connection develops between the URM footing, the masonry wall, and the soil, with possible uplift and rocking under tensile or rocking response. These mechanisms should be captured in the seismic model, if the vertical and lateral loads are such as to allow for this kind of motion. Consequently, if preliminary analyses indicate this to be the case, the soil resistance to the footing stresses should be modeled using an elastic but tensionless type of behavior (e.g., a Winkler model with uplift), in order to obtain the proper footing flexibility as is the case in situ. An in situ geotechnical study and adequate knowledge of the foundation shall provide the soil constitutive characteristics and the foundation conditions (e.g., the possible existence of a well or a septic tank near the footing).

Principle Modeling Requirements for Seismic Analysis of LBM Buildings

In summary, the building seismic model used should identify the basic force transfer mechanisms of the structural system, irrespective of the method of analysis used:

- (a) *Plane or three-dimensional model.* Plane frames are often used as an approximation for modeling the building. It is important to

note that the proper idealization of seismic behavior is through modeling of the three-dimensional response. Two-dimensional models are unable to monitor the three-dimensional response of the entire structure; consequently their use will not predict the spatial response effects, namely, (i) the simultaneous action of seismic axial forces and biaxial bending effects in the corner piers, (ii) the corner element vertical deformation compatibility under concurrent actions in the two orthogonal directions, (iii) the influence of out-of-plane bending in the resistance of in-plane actions of the vertical elements, and (iv) the possible in plan torsional effects which will enter into the response in the case of a relatively rigid diaphragm and eccentric distributions of the mass (e.g., an opening) or the stiffness (e.g., asymmetric facade opening distributions and/or interior masonry walls).

- (b.1) *Modeling of the diaphragm.* The presence or lack of a diaphragm in the model should be in accordance with the function of the diaphragm in the structural system at hand. The in-plane rigidity of the diaphragm is an important consideration to account for in the model, particularly because of the fact that masonry buildings tend to be very stiff and the relative diaphragm-to-wall element in-plane stiffness will define the distribution of inertia loads from the floor to the vertical seismic load resisting elements.
- (b.2) *Modeling of the diaphragm connection with the walls.* Furthermore, as far as the modeling of the diaphragm is concerned, it is important to consider in the model whether its connections with the vertical masonry elements justify the use of a deformation fully compatible FE nodal connection, transferring load. The same is true for the roof structure, which is normally simply supported on the masonry walls, and sliding of the roof trusses on the walls is often possible under differential lateral seismic movements across walls.
- (c) *Modeling of the mass.* The conventional frame analysis assumption of a lumped mass

idealization is not justified in masonry modeling, and, instead, the model should incorporate distributed masses; due to the large masonry element size (compared to the normal operational loads of the building) and often the lack of a heavy concrete diaphragm, loads are not distributed according to the lateral stiffness of the elements. For this reason, three-dimensional modal dynamic analysis should at best be adopted, reflecting more accurately the system deformation and the load path of the inertia forces from the diaphragm to the foundation. Similarly, out-of-plane effects need also to be taken into account in a “distributed with height” sense for assessing the in- and out-of-plane interaction of forces to the walls, the dynamic connection forces at the wall intersection, and the seismic deformations and forces at the roof-to-wall connections.

- (d) *Modeling of the building foundation.* Masonry building response and past deformation history are affected by the deforming foundation at the base of the building. Consequently, full foundation fixity is often an unrealistic assumption, given that the stresses under the masonry wall may be relatively high. Evidence of cracking around openings or unsymmetrical distribution of cracks in an existing masonry building is often the effect of differential settlements due to variable ground conditions in plan, such as the presence of an abandoned well or a septic tank at one end, improper ground preparation at the time of construction, or a partial plan base-ment. For modeling the foundation, an acceptable practical modeling approach makes use of tensionless elastic springs, providing vertical, lateral, and bending restraint in the embedded footings.

FE Modeling of LBM Structures for Seismic Analysis

From the analysis of the complexity of the response, the failure modes, and the fact that different technologies and materials comprise

masonry construction, practical modeling of masonry buildings for seismic analysis relies on the use of FE models and follows two different techniques (Fig. 13):

- (i) The less refined “global” (phenomenological) macromodels, suitable for the analysis of entire LBM buildings. In this case, the walls, piers, and spandrel beams of the masonry structure are modeled using one-dimensional line elements with nonlinear characteristics. These macromodel elements (Fig. 13d) are characterized by equivalent axial load, bending, and shear interaction response characteristics. Such models have been used both for elastic and inelastic three-dimensional seismic analysis for design, assessment of seismic vulnerability, and fragility studies of masonry buildings under monotonic or cyclic loading.
- (ii) The refined “local” FE micromodels where the entire masonry building or a plane portion of the building, such as a masonry façade under planar response, is modeled using two- or three-dimensional FE approximations, with associated material and loading description (Fig. 13c). In terms of geometric representation, thick shell or across the length and through thickness brick FE approximations are adopted, possibly coupled with beam FE for the reinforcement, if any. Even more refined micromodel approximations have been adopted in research studies, in which the actual masonry unit and mortar have been separately modeled using brick and plate elements, respectively.

In terms of material approximation, phenomena such as cracking and compression nonlinearity or the presence of steel are smeared within the element integration area through equivalent stiffness and resistance modifications at the FE integration point. Models of this type have been used in the parametric investigation of conventional or historical masonry construction and for the validation of test results. Complexity and computational cost and resources are the

primary issue in this case as well as in certain aspects, the actual ability to model the material behavior under cyclic loading conditions.

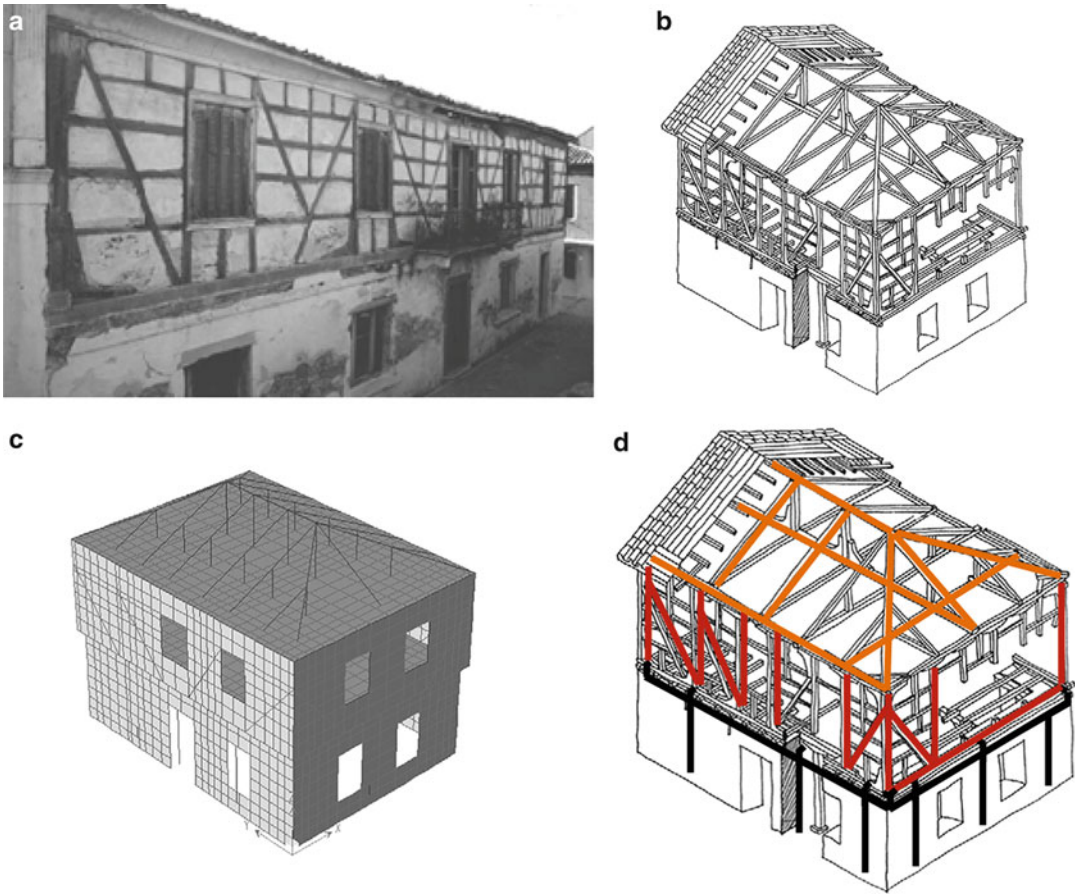
Macromodels for Entire Building Analysis

Since the use of micromodels is too expensive and complicated for entire building analysis under seismic excitations and due to the fact that available software capabilities are often limited in handling such FE micromodels, equivalent FE macromodels are being used for practical seismic analysis of LB or NLB masonry construction.

For load-bearing masonry, the usual modeling conventions adopted in conventional frame building analysis are also adopted for masonry buildings as well: roof and diaphragm elements are modeled using line FE, or plain diaphragmatic action is enforced (if it exists). Similarly, wall piers, columns (pilasters), and spandrels are modeled at their centerline using equivalent or actual property line FE with linear or nonlinear characteristics, while the element joint regions are modeled using infinitely stiff elements and/or rigid zone transformation models depending on the analysis software conventions.

For inelastic analysis, phenomenological axial, shear, and flexural constitutive relations should be specified, possibly with cracking and post-failure modeling capabilities and, if possible, interaction of axial/flexural and shear stresses. Where this is not possible and given that the variation of axial loads is not high for low-rise buildings, uncoupled values may be assumed based on initial state vertical load levels.

Masonry building macromodels for inelastic analysis evolved from: (i) the simplified weak spandrel strong pier model, whereby spandrels crack early and are neglected (therefore the piers are considered to act as uncoupled cantilevers joint by hinged rigid link beams at the floor levels) (Fig. 14b); (ii) the strong spandrel weak pier model (the shear frame analogy), whereby the piers crack first and are therefore assumed to deform with their inelastic characteristics, the spandrels remaining relatively rigid (Fig. 14c); and (iii) the equivalent frame model, where the masonry structure is modeled as an assembly of



Seismic Analysis of Masonry Buildings: Numerical Modeling, Fig. 13 Modeling example of a historical masonry building in Lefkada, Greece, built of timber confined plinth masonry walls at the upper floor on

a stone masonry ground floor. (a) Photo. (b) Bearing structure. (c) Micromodel. (d) Macromodel (Vintzileou et al. 2007). Note that several load-bearing elements in the model are omitted for clarity

vertical pier and horizontal spandrel line elements interconnected by rigid joint regions (Fig. 14d): the geometry of the joints is obtained from the geometry of openings and an equivalent pier height, which is defined by the extent of cracking observed in the vertical elements following an earthquake or, if uncracked, assuming a crack inclination at about 30° that extends from the opening toward the joint (Lagomarsino et al. 2013).

For the constitutive modeling of inelastic FE macromodels under static seismic-type load, the multilinear (simplified or more refined) shear force–interstory drift or bending moment–pier rotation diagrams of Fig. 11 have been proposed

by several investigators for modeling both the in-plane shear and the flexural response of masonry wall piers, incorporating the different failure mechanisms of these elements (Fig. 15).

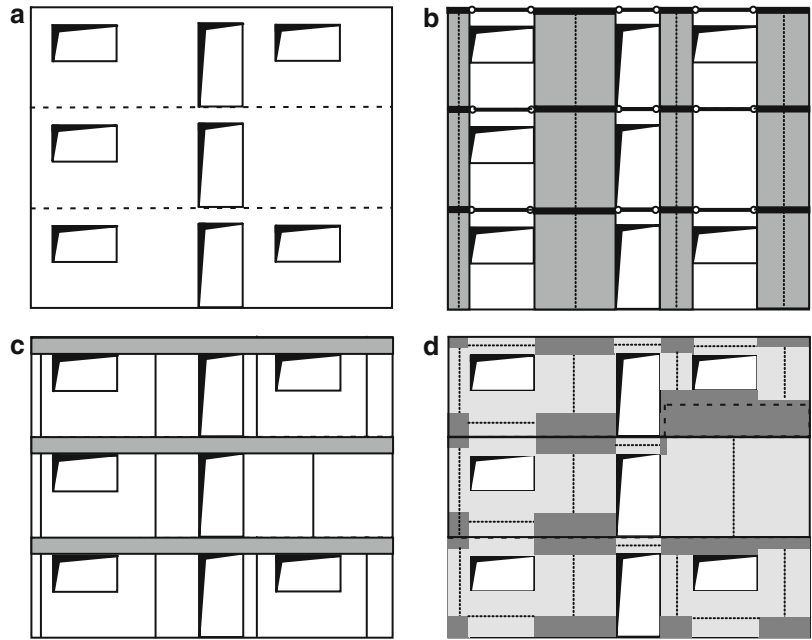
Magenes and Calvi (1997) proposed that the peak shear resistance R_u (Fig. 15) of a rocking masonry wall pier bearing an axial load P is given by:

$$R_u = \frac{D^2 t}{H_0} \frac{p}{2} \left(1 - \frac{p}{k f_u} \right) \quad (2)$$

where D and t are the length and thickness of the pier; H_0 is the effective height equal to the shear span, namely, the height to zero moment, taken as

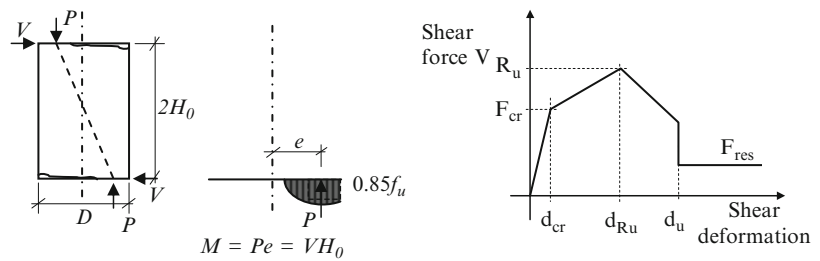
Seismic Analysis of Masonry Buildings: Numerical Modeling,

Fig. 14 Equivalent frame model for load-bearing masonry construction. (a) The masonry facade to be modeled. (b) Weak spandrel and strong cantilever pier model. (c) The weak pier and rigid spandrel shear frame model. (d) The equivalent frame model using spandrels, piers, and rigid joint regions (Lagomarsino et al. 2013)



Seismic Analysis of Masonry Buildings: Numerical Modeling,

Fig. 15 Model of the wall pier and lateral load deformation constitutive relation under axial force



being equal to the pier height, for cantilever piers, and half the pier height, for piers in contraflexure; p is the average vertical load pressure, equal to the axial load P divided by the wall area ($p = P/Dt$); f_u is the compressive strength of the masonry; and k is a coefficient that depends on the stress distribution at the toe of the wall ($k = 0.85$ for an equivalent rectangular stress block; see Fig. 15).

For shear failure mechanisms of brick masonry walls, when mortar bed and head stone failure is involved, they proposed to adopt a Mohr–Coulomb friction model for the wall, namely:

$$R_u = Dt\tau_u = Dt(c + \mu p) = Dt\left(c + \mu \frac{P}{Dt}\right) \quad (3)$$

where, in addition to the parameters defined above, τ_u is the average masonry shear strength; μ is the Coulomb friction coefficient; and c the cohesion of the wall (namely, the average frictional resistance at $P = 0$). Both of the latter parameters are global material constants for the wall and are obtained from testing of masonry elements. Following micromodel analysis at failure and comparison with test results, their model accounted for two types of shear wall failure, namely:

- (i) Failure at the cracked wall ends, with the peak shear strength τ_u being resisted in the compression area of the cracked section D' by t , where D' (the compressed portion of

the wall length D) is obtained by equilibrium of normal force and moment (Eq. 4a):

$$D' = \beta D = \left(1.5 - \frac{3V}{P} \frac{H_0}{D}\right) D \quad (4a)$$

- (ii) Failure at the mid-height of the wall, with the maximum shear strength τ_u being resisted by the entire wall thickness (area D by t) decreasing in an inverse linear manner with the wall shear span ratio H_0/D :

$$R_u = Dt \frac{\tau_u}{1 + \frac{H_0}{D}} = Dt \left(\frac{c + \mu p}{1 + \frac{H_0}{D}} \right) \quad (4b)$$

The shear strength of the wall in diagonal tensile failure is therefore given by the minimum strength of the two resisting mechanisms – failure modes:

$$R_u = Dt \tau_u, \quad \tau_u = \min \left\{ \left(\frac{1.5\bar{c} + \bar{\mu}}{1 + \frac{3\bar{c}H_0}{pD}} \right), \left(\frac{\bar{c} + \bar{\mu}p}{1 + \frac{H_0}{D}} \right) \right\} \quad (5)$$

It is further noted in the above expressions that the coefficients of friction and cohesion μ and c may be modified so as to obtain effective values $\bar{\mu}$ and $\bar{c}$, respectively, corrected for the geometry of the masonry unit, in accordance with the fact that the expressions above sometimes overestimated the experimental value of strength, due to the influence of the masonry headjoints.

- (iii) In addition to shear failure by Coulomb friction at the joints, shear failure due to in-plane cracking of the bricks was also experimentally observed for weak brick and strong mortar, in the presence of high axial stresses. It was proposed that the shear strength be estimated in this case in terms of the tensile strength of the bricks f_{bt} , following Eq. 6:

$$R_u = Dt \tau_b = Dt \frac{f_{bt}}{2.3 \left(1 + \frac{H_0}{D}\right)} \sqrt{1 + \frac{p}{f_{bt}}} \quad (6)$$

and the minimum of the values (Eqs. 5 and 6) used for resistance. In terms of deformation, the wall pier deformation corresponding to the ultimate strength (d_{Ru} , Fig. 15) was found to be close to 0.5 % of the wall height H in most test results they performed or evaluated (Magenes and Calvi 1997).

Detailed Micromodels for Seismic Analysis

In addition to the macromodels above, micromodels have been proposed and are employed for equivalent static linear or nonlinear seismic analysis of masonry buildings and (primarily) historical structures; furthermore, micromodels have also been used to calibrate macromodel topology and the masonry wall failure and constitutive response. Micromodels idealize masonry in detail using:

- (i) Two-dimensional thick plate and shell elements that account for both in-plane and out-of-plane stiffness and resistance characteristics (for out-of-plane bending effects of the walls). The use of shell element elastic models is quite common in the seismic analysis of masonry structures, since they do not require excessive computational resources and provide the basic load path and demand concentrations within the building, accounting for complex geometries and multitude of materials, such as timber, masonry, etc. (Fig. 13c). These are therefore suitable to use in entire building seismic analysis models.
- (ii) Three-dimensional (brick) FE models of the entire masonry structure, taking into account in the model both complex geometric idealizations of the structure and the foundation and the material complexities associated with the presence and interaction of several different materials with complex constitutive characteristics, such as: stress-strain

nonlinearity due to cracking or crushing; tri-axial capacity interaction and volumetric dilatancy under loading for the mortar, concrete, stone, or brick; yielding or pullout phenomena of the reinforcement where it exists; interface failure between brick, mortar, or steel (where it exists); and inelasticity of timber.

Depending also on the capabilities of the software adopted for the seismic evaluation, micromodels adopted for modeling of inelastic seismic response of LBM buildings include:

- The smeared representation models. These model nonlinearity as spatially averaged, by considering the distributed cracking of masonry in the vicinity of the FE integration points and/or the average constitutive response of masonry over the entire FE integration volume (area) using suitable two- or three-dimensional inelastic constitutive behavior and a strength interaction surface. Plane stress or three-dimensional yield surface characteristics have been proposed for CMU, brick masonry, or stone, possible candidates being, among others, the models proposed by Gambarotta and Lagomarsino (1997) and Stavridis and Shing (2010).
- The discrete representation models. These model nonlinearity discretely through detailed FE modeling of all different material regions involved, namely, individual modeling of the mortar as a brick, shell, or zero-length contact FE, the masonry units and the concrete (where present) as a brick or plain stress shell element (primarily for industrially manufactured units of constant geometry), and the steel or timber reinforcement (where present) as truss or beam elements (for CM or RM buildings).
- As a special case of these are the discrete crack representation models, which further monitor crack formation using fracture energy criteria and the evolution of cracking within the masonry element through mesh redefinition (average material representation) and/or through or along predefined mortar beds (where these exist in physical and model

space), using suitable contact friction elements.

- Models based on the discrete element idealization that fall into this category have been promoted for modeling primarily historical monuments; in this case, the brick units (or stone building blocks) are modeled as individual deformable or undeformable volume elements, and their interface is described with Coulomb frictional contact-separation characteristics.

Modeling of NLBM Infilled Frame Buildings for Seismic Analysis

Due to the abundant use of CM infilled panels in steel and RC frame structures, the seismic analysis modeling of entire frame buildings with NLBM infills is also briefly examined herein for completeness. Only the modeling of the infills is considered herein, since the modeling of the entire frame is beyond the scope of this text. It should be noted, however, that in the design of RC frame buildings, the infills are (and have been) neglected in the structural model assuming that these contribute only to the inertia mass. Only recently modern seismic codes (EC8 2004) provide structural forming and detailing guidelines for taking into account possible adverse response effects due to the presence of the infill; furthermore, PBD methodologies for the assessment and retrofit of existing RC frames require that these be fully accounted for in the seismic model, in their as-built configurations and properties (FEMA-356 2000).

Infill panels provide a large increase in the lateral stiffness of the confining frame with a disproportionate increase in its mass; consequently, masonry infilled frames exhibit short fundamental periods compared to the bare frame structure and therefore attract higher inertia forces at shorter drifts. Properly engineered infill panels, constructed of good quality modules and mortar and adequately wedged into the panel, without excessive openings and a regular in plan and in height configuration, will provide the bare frame structure with considerable overstrength and stiffness enhancement.

Seismic Analysis of Masonry Buildings: Numerical Modeling,

Fig. 16 Damages to infilled RC frames following the 1985 Mexico City earthquake. (a) Damage of the confining frame elements. (b) Out-of-plane failure of the infills. Note that the panels were confined by RC lintels and pilasters tied to the RC frame

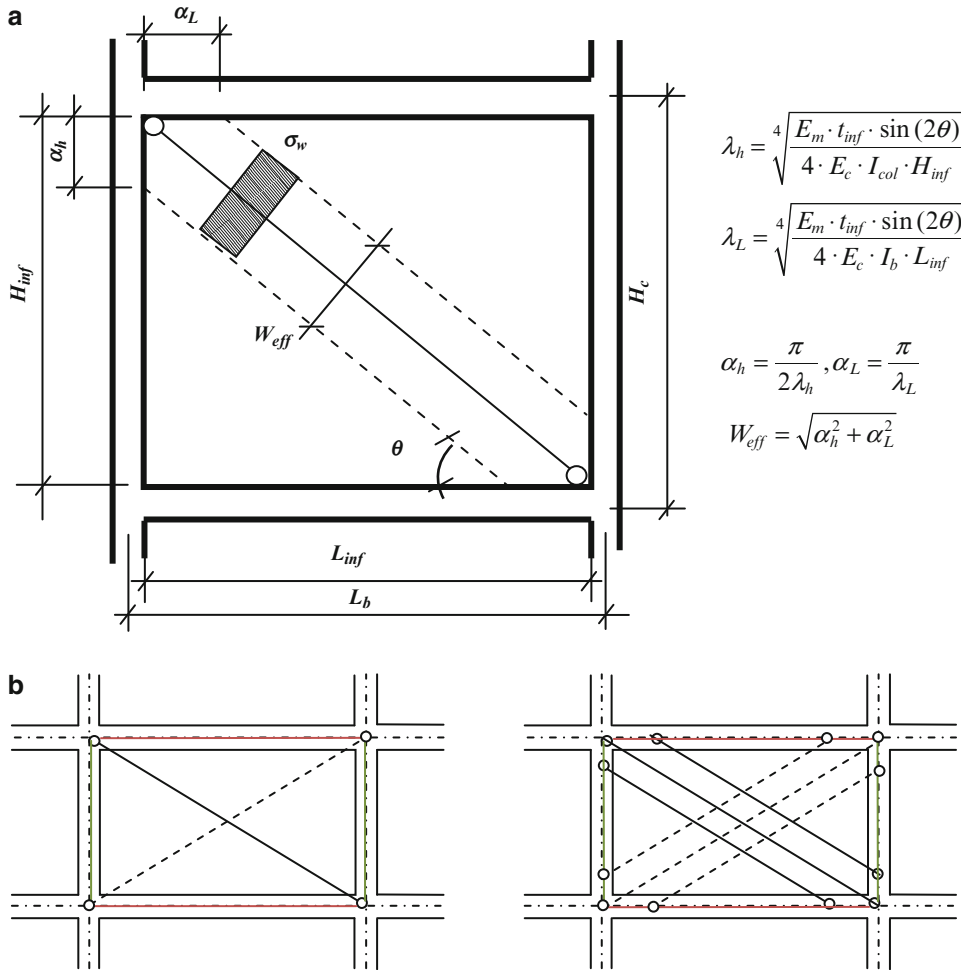


Masonry infills, like other masonry elements already discussed, tend to fail in plane or out of plane (Fig. 16) following similar mechanisms as other masonry elements (diagonal thrust or tension, corner compression, sliding at joints through brick or mortar, out-of-plane collapse, etc.). Such a full or partial failure of the infills will lead to local failures of the confining elements due to the formation of unintentional short column effects or due to a shear failure of the top of the column or the beam–column joint. For these reasons, in the PBD assessment approach of existing structures, the infills need to be accounted for in seismic analysis.

In modeling infilled frame buildings in practical seismic analysis of entire structures, macromodels are used for the panels, as shown in Fig. 17. Micromodels have also been adopted in the literature, however primarily for research applications and in order to calibrate the macromodel parameters. The use of macromodels stems from the observation that, due to the partial contact separation of the panel from the surrounding frame elements under lateral response of the infilled frame, the function of the infill can be modeled as an equivalent axial load-bearing diagonal strut element resisting

compression only in the direction of the lateral force. The strut has a thickness t_{inf} equal to the panel thickness and an equivalent width W_{eff} based on the frame to panel contact length; proposed strut widths adopt contact lengths that account for the relative stiffness characteristics of the confining frame and the infill panel in terms of $\lambda_{h,L}$ (Stafford Smith and Carter 1969), where $\lambda_{h,L}$ expresses the relative stiffness of the infill and the concrete elements assuming elastic contact (Fig. 17a):

where E_m and E_c are the Young's moduli of masonry and concrete; I_{col} is the uncracked moment of inertia of the confining column elements; θ is the geometric inclination of the infill strut (Fig. 17a); L_{inf} and H_{inf} are the clear length and height of the infill; and H_c is the centerline distance between the beams. For seismic applications cross strut configurations are used with compression only properties (Fig. 17b). In order to model the effect of short column formation in the infilled frame response, multiple strut configurations as shown in Fig. 17b have also been proposed (Crisafulli et al. 2000). Their use is recommended in the case of asymmetric infill configurations within a floor, such as infilled frame bays next to open bays or end bays of the infilled frame.



Seismic Analysis of Masonry Buildings: Numerical Modeling, Fig. 17 Equivalent strut macromodel of infilled RC frames and effective strut width equation. (a)

Model proposed by Stafford Smith and Carter (1969). (b) Single and multiple strut macromodels of infilled RC frames for static lateral load and cyclic dynamic analysis

Summary

The seismic response and forms of failure for different types of masonry construction have been presented and discussed. It is commonly believed that masonry buildings, being the most common form of past construction in all seismic affected areas, as well as a financially viable alternative for new low- to mid-rise construction, they need to be evaluated for seismic loading, due to the seismic vulnerability of such structures as it has been proven in numerous seismic events, past and recent.

Several modeling conventions have been proposed so far and are still under rigorous research investigations, for the seismic analysis of masonry buildings. Overall, these range, depending on the capabilities of the FE code at hand and also on the reliability of the structural information available, from simple strut-and-tie models (the equivalent frame analysis models, which are also referred to as macromodels) to the refined micromodels, using plane or three-dimensional FE analysis tools, both linear and nonlinear material-wise and geometry-wise.

Not all methods are suitable for all cases, and often analysis “overkill” for a problem that is quite complex to model will give a false sense of security: due to the complexity of the problem, involving different masonry materials (brick, mortar, stone, rubble, and also RC or steel and timber), different types of construction (LB or NLB, CM), and different structural topologies (low or mid rise, with or without diaphragms, irregular in plan or elevation, with openings, with flexible foundation), equally complex methods should be used, justifying the accuracy of the input information, namely, the material properties for all materials involved in the construction and an adequate knowledge of their interaction.

For entire structural models, macromodels or area micromodels with averaged properties over the FE region are adequate to capture global quantities (forces and deformations) and seismic performance. For detailed damage prediction (cracking, crushing, region disintegration under overload), three-dimensional FE models are adopted, using either deformable elements with smeared or discrete cracking representation or discrete FE models with contact friction interfaces, where crack spreading is not feasible (e.g., URM construction and historic building analyses). These models, although they provide realistic damage predictions compared to observation and test, are too detailed and complex to be applied to ordinary construction, and they need to be substantiated by adequate testing in order to establish the material behavior as input to the FE model.

Cross-References

- [Ancient Monuments Under Seismic Actions: Modeling and Analysis](#)
- [Assessment of Existing Structures Using Inelastic Static Analysis](#)
- [Equivalent Static Analysis of Structures Subjected to Seismic Actions](#)
- [Masonry Modeling](#)
- [Masonry Structures: Overview](#)
- [Nonlinear Finite Element Analysis](#)
- [Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings](#)

- [Seismic Vulnerability Assessment: Masonry Structures](#)
- [Seismic Vulnerability Assessment: Reinforced Concrete Structures](#)
- [Strengthening Techniques: Masonry and Heritage structures](#)

References

- ASCE (2013) Building code requirements and specification for masonry structures. Standards ASCE/SEI 5-13, 6-13. American Society of Civil Engineers, Reston
- Clough R, Mayes R, Gülkan P (1979) Shaking table study of single-story masonry houses, vol 3, Summary, conclusions and recommendations, UCB/EERC-79/25. Earthquake Engineering Research Center, University of California, Berkeley
- Crisafulli F, Carr A, Park R (2000) Analytical modeling of infilled frame structures; a general review. *Bull N Z Soc Earthq Eng* 33(1):30–47
- DesRoches R, Comerio M (2011) The 2010 Haiti Earthquake. *Earthquake Spectra*, vol 27, S1. Earthquake Engineering Research Institute, Oakland
- EC6 (2005) Eurocode No. 6, Design of masonry structures – Part 1-1: General rules for reinforced and unreinforced masonry structures, EN-1996-1. European Committee for Standardization, Brussels
- EC8 (2004) Eurocode No. 8, Design of structures for earthquake resistance – Part 1: General rules, seismic actions and rules for buildings, EN-1998-1. European Committee for Standardization, Brussels
- EERI/IAEE. World Housing Encyclopedia. <http://www.world-housing.net/>
- FEMA-356 (2000) Prestandard and commentary for the seismic rehabilitation of buildings. FEMA, Washington, DC
- Gambarotta L, Lagomarsino S (1997) Damage models for the seismic response of brick masonry shear walls. Part II: The continuum model and its applications. *Earthq Eng Struct Dyn* 26:441–462
- Hughes R, Nichols R, Paul T, Riding V (1990) The Augusta, Sicily earthquake of 13 December 1990, a field report. EEFIT, Institution of Structural Engineers, London
- IAEE (2008) Regulations for seismic design: a world list. <http://www.iaee.or.jp/worldlist.html>
- ICOMOS. International Council on Monuments and Sites. <http://www.icomos.org/en/>
- ISCARSAH. The International Scientific Committee on the Analysis and Restoration of Structures of Architectural Heritage. <http://iscarsah.icomos.org/>
- Kingsley G (1994) The U.S.-TCCMAR full-scale five-story masonry research building test, Part 2, Report SSRP-94/02. University of California, San Diego
- Lagomarsino S, Penna A, Galasco A, Cattari S (2013) TREMURI program: an equivalent frame model for

the nonlinear seismic analysis of masonry buildings. *Eng Struct* 56:1787–1799

- Magenes G, Calvi GM (1997) In-plane seismic response of brick masonry walls. *Earthq Eng Struct Dyn* 26(11):1091–1112
- Magenes G, Morandi P, Penna A (2008) Test results on the behaviour of masonry under static cyclic in plane lateral loads. ESECMaSE project, report RS-01/08. Department of Structural Mechanics, University of Pavia, Pavia
- Moro L (coord. editor) (2007) Guidelines for evaluation and mitigation of seismic risk to cultural heritage. Gangemi Editore, Rome
- Rossetto T, Peiris N, Alarcon J, So E, Sargeant S, Libberton C, Verrucci E, Del Re D, Free M (2009) The L'Aquila (Italy) Earthquake of 6th April 2009, a field report. EEFIT, Institution of Structural Engineers, London
- Ruiz-Garcia J, Negrete M (2009) Drift-based fragility assessment of confined masonry walls in seismic zones. *Eng Struct* 31:170–181
- Stafford Smith B, Carter C (1969) A method of analysis for infill frames. *Proc Inst Civil Eng* 44:31–48
- Stavridis A, Shing PB (2010) Finite element modelling of nonlinear behaviour of masonry-infilled RC frames. *J Struct Eng* 136(3):285–296
- UNIDO (1983) Repair and strengthening of reinforced concrete, stone and brick masonry buildings. Building construction under seismic conditions in the Balkan region, vol 5. UNDP/UNIDO Project RER/79/015. United Nations Industrial Development Organization, Vienna
- Vintzileou E, Miltiadou-Fezans A (2008) Mechanical properties of three-leaf stone masonry grouted with ternary or hydraulic lime based grouts. *Eng Struct* 30(8):2265–2276
- Vintzileou E, Zagkotsis A, Repapis C, Zeris C (2007) Seismic behaviour of the historical structural system of the Island of Lefkada, Greece. *Construct Build Mater* 21(1):225–236

Seismic Analysis of Steel and Composite Bridges: Numerical Modeling

Ioannis Vayas¹ and Aristidis Iliopoulos²

¹Traffic Engineering Laboratory, National Technical University of Athens (N.T.U.A.), Athens, Greece

²Peikko Greece SA, Marousi, Athens, Greece

Synonyms

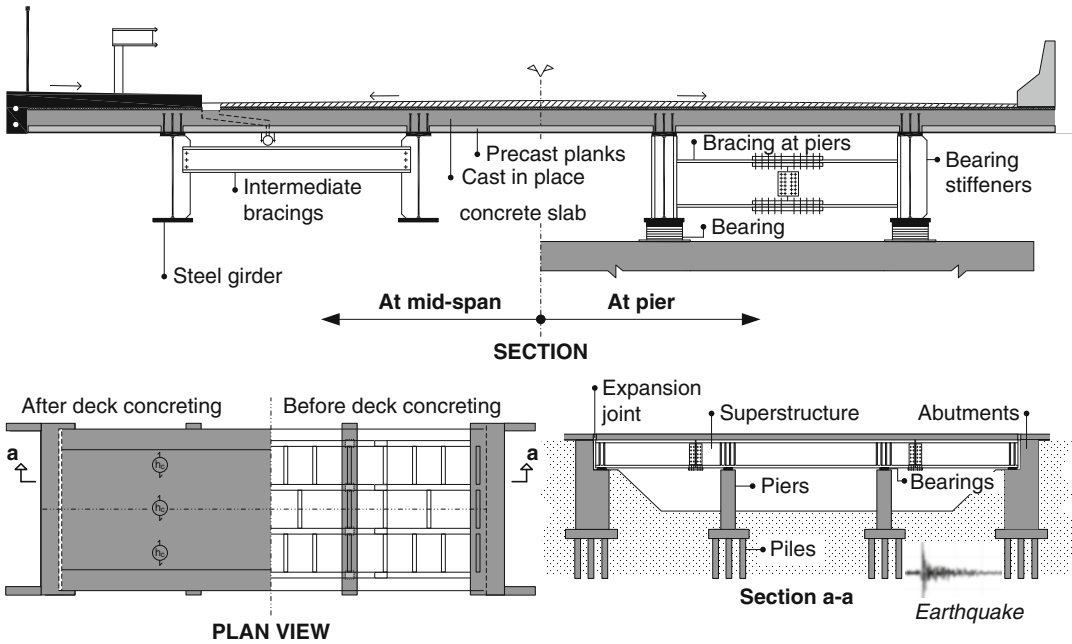
Bearings; Bridge modeling; Composite bridge; Seismic design; Seismic isolation

Introduction

Reinforced concrete slabs rigidly connected with steel girders have been used to form the basic superstructure of large numbers of deck bridges for many decades. This is due to the fact that the composite construction method offers the bridge engineers a great variety of solutions for different types of problems. A typical composite cross section of a highway bridge is shown in Fig. 1. A series of parallel steel girders are rigidly connected with a reinforced concrete slab through shear connectors. The shear connectors installed are mostly welded studs allowing use of the deck as part of the top flange (*deck plate girders*). The longitudinal bending of the composite T-girders, at sagging bending areas, results in tension in steel and compression in concrete. The simultaneous operation of both of these materials generates the composite action which is the most important feature for the formation of stiff and high-strength cross sections. At hogging moment areas, concrete is considered to be fully cracked, and only the slab reinforcement, but not concrete, contributes to bending resistance.

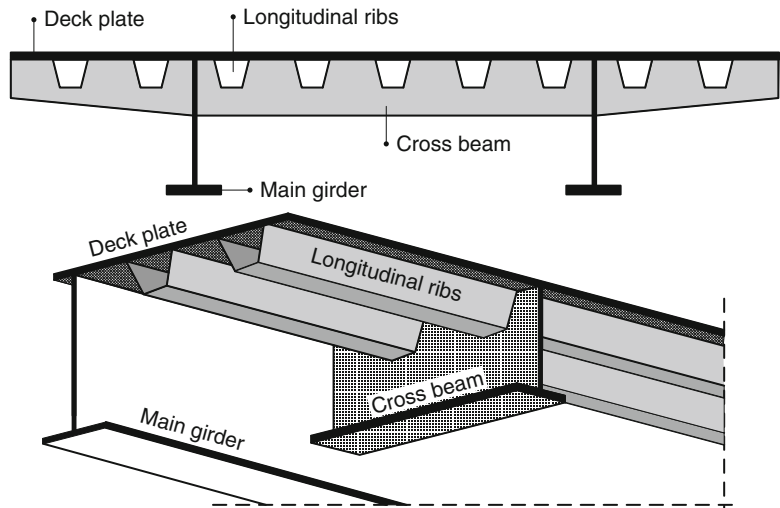
In pure steel bridges the reinforced concrete slab is replaced by an orthotropic steel deck. This is composed of the steel deck plate; the longitudinal, mostly trapezoidal, stiffeners; and the crossbeams (Fig. 2). Orthotropic are lighter than concrete decks. However, they require high fabrication costs due to intensive welding operations and are susceptible to fatigue. Therefore, pure steel bridges (see EN 1993-2 2006) are nowadays mostly limited to cases where it is essential to limit the deck weight, e.g., for very large spans, movable bridges, etc. The present article refers mainly to composite bridges that constitute the vast majority in modern steel bridge construction.

Modeling for analysis is required in order to determine internal forces and moments, deformations, and vibrations of bridge decks including bearings, piers, abutments, piles, etc. In addition models should include the foundation when soil-structure interaction is accounted for. A bridge analysis model should be based on the following criteria:



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 1 Layout of a typical composite deck bridge

Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 2 Steel bridge with orthotropic deck



- It should reflect the structural response in terms of deformation, strength, and local and global stability.
- It should include as many as possible structural elements and parts of the bridge deck (cross frames, stiffeners, etc.) and their possibly eccentric connections.
- It should also include bearings and piers individually, piles, etc.
- It should cover all construction stages and loading cases.

- Loads should be easily introduced.
- It should allow the performance of dynamic analysis and include the most important modes of vibration.
- The resulting output should be such that it enables easily the execution of the code-prescribed verifications.
- It should be supported by commercial analysis and design software.

Models for Seismic Analysis

There exist several possible models for bridge seismic analysis that could be employed depending on the bridge configuration, the bearing types, the connection between superstructure (bridge deck) and piers or abutments, the type of foundation with possible soil-structure interaction, etc.

Figure 3 shows possible modeling levels for seismic analysis of steel and composite bridges, starting from the simple to the comprehensive ones. Simple models, like the first three in Fig. 3, can be used for preliminary seismic analysis. They may be employed for the derivation of seismic forces and displacements on bearings, piers, foundations, or the soil. Comprehensive models, like the last three of Fig. 3, are mostly used at the main design phase since they also cover analysis for construction stages and service conditions where other loads due to traffic, wind, temperature, time-dependent concrete effects, etc. must be taken into account.

In the following the various analysis models for steel and composite steel-concrete bridges are presented. Models for superstructures (decks) are linear since decks are expected to remain elastic in the seismic situation. Any inelastic activity should be restricted to bearings, piers, piles, soil, etc. For bearings and piers, nonlinear characteristics are provided. For other elements (piles, abutments, soil), reference is made to other chapters.

Rigid Deck Model

In this model the superstructure is represented as a single mass, M_{dir} , acting on a spring of stiffness

K_{dir} . M_{dir} includes the entire mass of the deck without the mass of the piers. The global spring stiffness K_{dir} represents the combined stiffness of bearings and piers and the foundation of soil and is calculated from

$$\frac{1}{K_{dir}} = \frac{1}{\sum K_{bearings}} + \frac{1}{\sum K_{piers}} + \frac{1}{\sum K_{foundation}} \quad \text{where } dir = \text{direction X or Y} \quad (1)$$

The fundamental period for this single mass oscillator is calculated from (see Fig. 3a)

$$T_{dir} = 2 \cdot \pi \cdot \sqrt{\frac{M_{dir}}{K_{dir}}}, \quad M_{dir} = \sum m \quad (2)$$

The seismic forces acting on the entire deck $F_{EA, dir, tot}$ are determined from the relevant response spectrum. The seismic forces of bearings at the top of one pier i may be determined from

$$F_{bearing, i} = \frac{\sum (K_{bearing, i} + K_{pier, i} + K_{foundation, i})}{K_{dir}} \cdot F_{AE, dir, tot} \quad (3)$$

This model may be applied for the longitudinal direction of straight bridges with continuous decks, when the mass of the piers is less than 20 % of the tributary mass of the deck. The model may be also applied for the transverse direction provided that all conditions (a) to (c) referred below apply (see EN 1998-2 2005):

- (a) $L/b \leq 4.0$

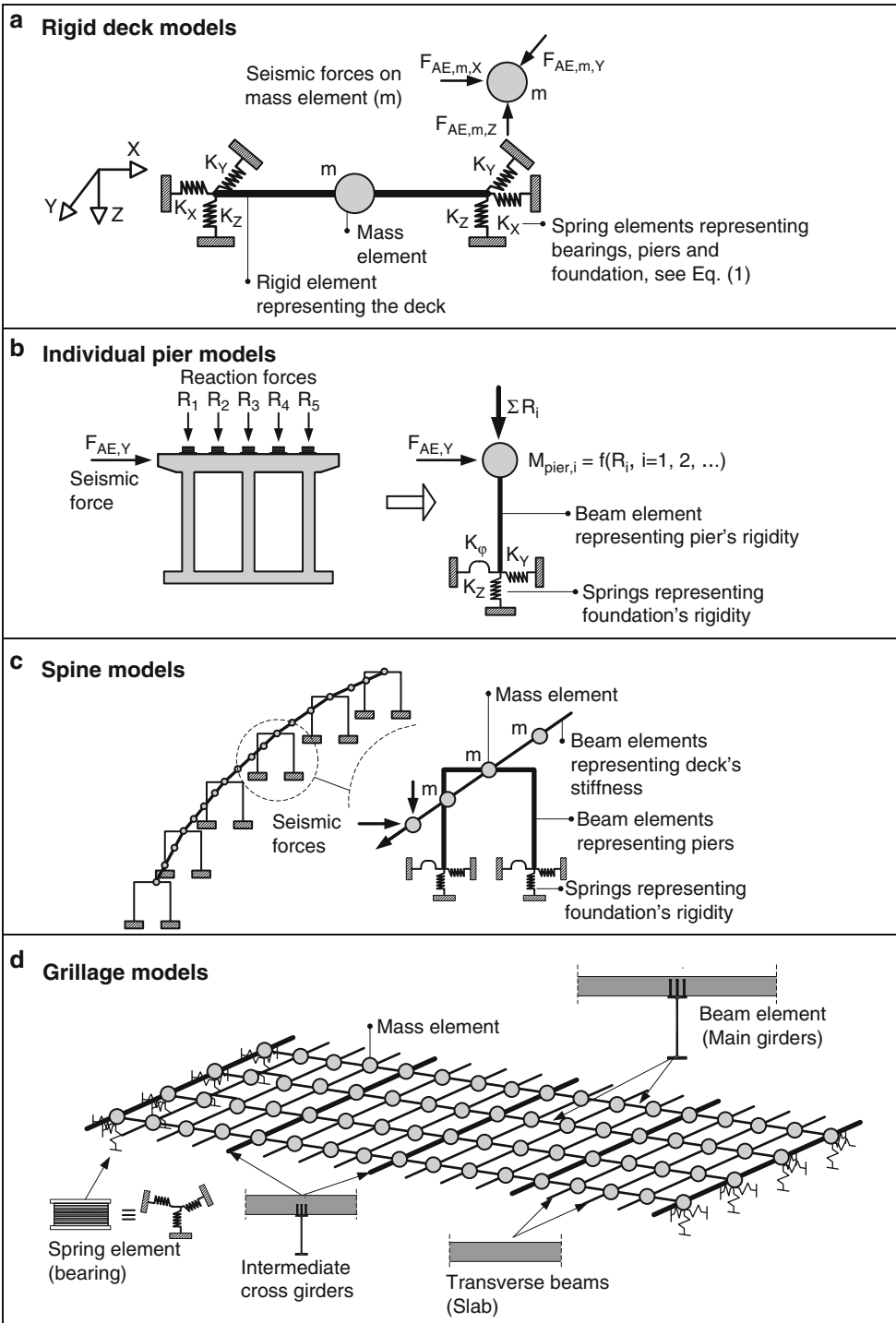
where

L is the total bridge length and

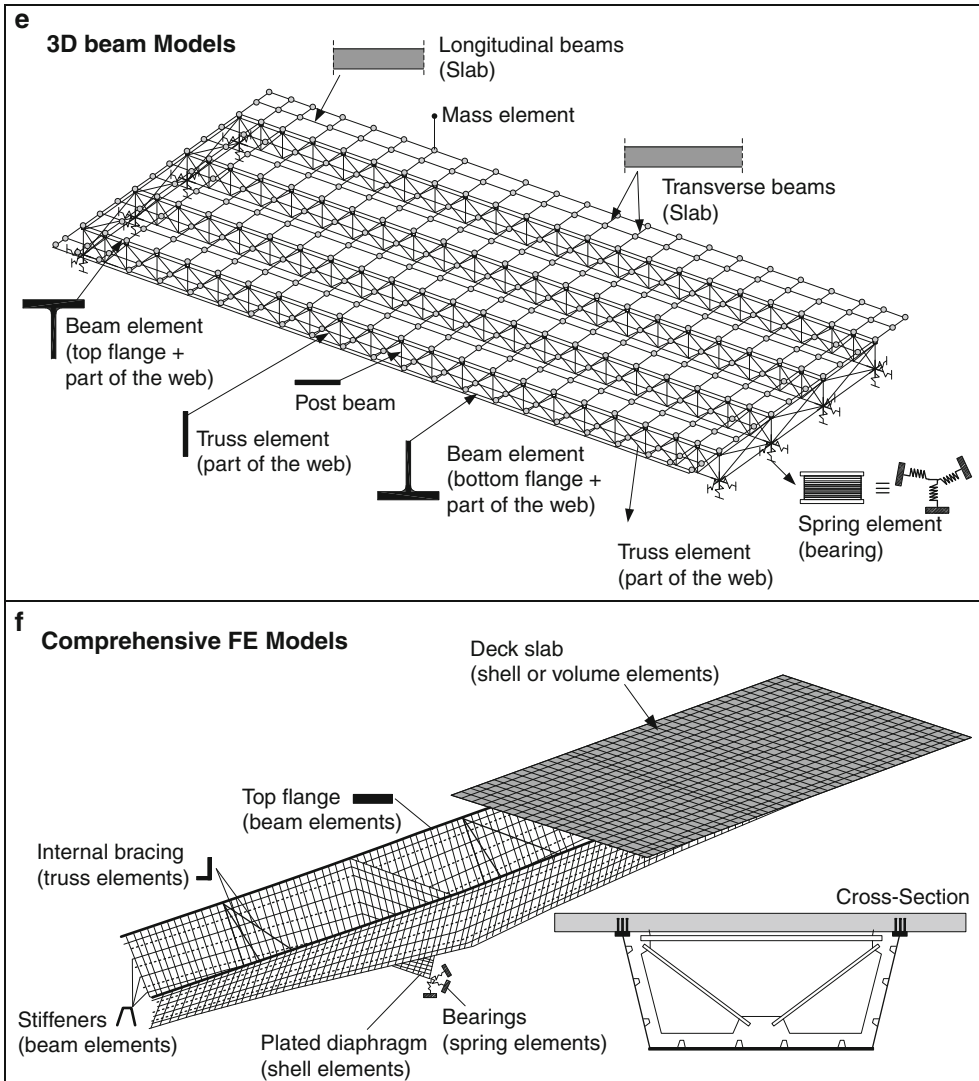
b is the width of the deck.

- (b) $\Delta d/d_a \leq 0.2$

where Δd and d_a are, respectively, the maximum difference and the average value of the displacements in the transverse direction of all pier tops under $F_{AE, Y}$.



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 3 (continued)



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 3 Analysis models for seismic design of steel and composite bridges. (a) Rigid deck

models. (b) Individual pier models. (c) Spine models. (d) Grillage models. (e) 3D beam models. (f) Comprehensive models

- (c) The theoretical eccentricity between the mass center of the deck and the stiffness center of the supporting members does not exceed 5 % of the deck's length.

The application of the rigid deck model may be extended to cases where bearings with damping properties are employed. The bearing stiffness K_{bearing} is expressing the secant

stiffness of the bearing device and is determined iteratively to correspond with the bearing's displacements. In addition higher damping values are achieved, so that a reduction factor must be employed to the resulting seismic forces. More detailed information is given in sections “[High-Damping Reinforced Elastomeric Bearings](#)” and “[Lead Rubber Bearings \(LRB\)](#).”

Individual Pier Model

This model may be used for seismic analysis of bridges in the transverse direction. Each pier and the associated part of the superstructure is considered separately and represented as a single mass oscillator. The mass of the oscillator is $M_{\text{pier}, i}$, and includes the mass of the deck between half distances of piers, while $K_{\text{pier}, i}$ represents the pier stiffness.

The fundamental period for pier i is calculated from

$$T_{\text{pier}, i} = 2 \cdot \pi \cdot \sqrt{\frac{M_{\text{pier}, i}}{K_{\text{pier}, i}}} \quad (4)$$

Based on the fundamental period, the seismic forces acting on top of the pier $F_{\text{AE}, Y}$ are determined from the relevant response spectrum. The individual pier model is not appropriate for curved or skew bridges, bridges with varying spans, varying pier lengths, etc. It may be used for long bridges where each pier is able to act independent of the rest of the bridge. These requirements are met when following conditions apply:

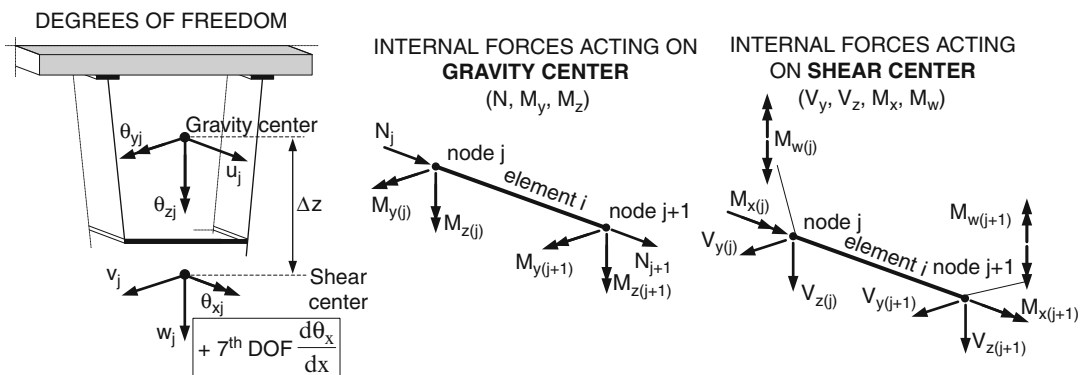
- The seismic action in transverse direction is mainly resisted by piers.
- There is no significant interaction between adjacent piers.
- $0.90 \leq \frac{T_{\text{pier}, i}}{T_{\text{pier}, i+1}} \leq 1.10$.

Spine Models

Spine models are appropriate for performing multimodal dynamic analysis on bridges and let the seismic forces resisted by each pier determined in a “natural” way in accordance with their relative stiffness. They may be employed for “normal” bridges, normal meaning more or less straight, low skew, narrow deck, and cross section with limited distortion, i.e., with rigid closely spaced transverse frames or crossbeams.

In spine models the bridge deck is represented by beam elements that are positioned at the centroid of the cross section and have six (6) degrees of freedom (DOFs) at end nodes (Fig. 4). The degrees of freedom are the translations (u, v, w) along the principal axes coordinate system (x, y, z) and the corresponding rotations ($\theta_x, \theta_y, \theta_z$). The resulting internal forces (N, M_y, M_z) act at the gravity center, while the shear forces and torsion moments (V_y, V_z, M_x) at the shear center. Cross-sectional warping may be taken into account by introduction of an additional seventh DOF per node as independent variable ($\theta'_x = d\theta_x/dx$), which results in the bimoment M_w as additional internal moment (see Kindmann and Kraus 2011).

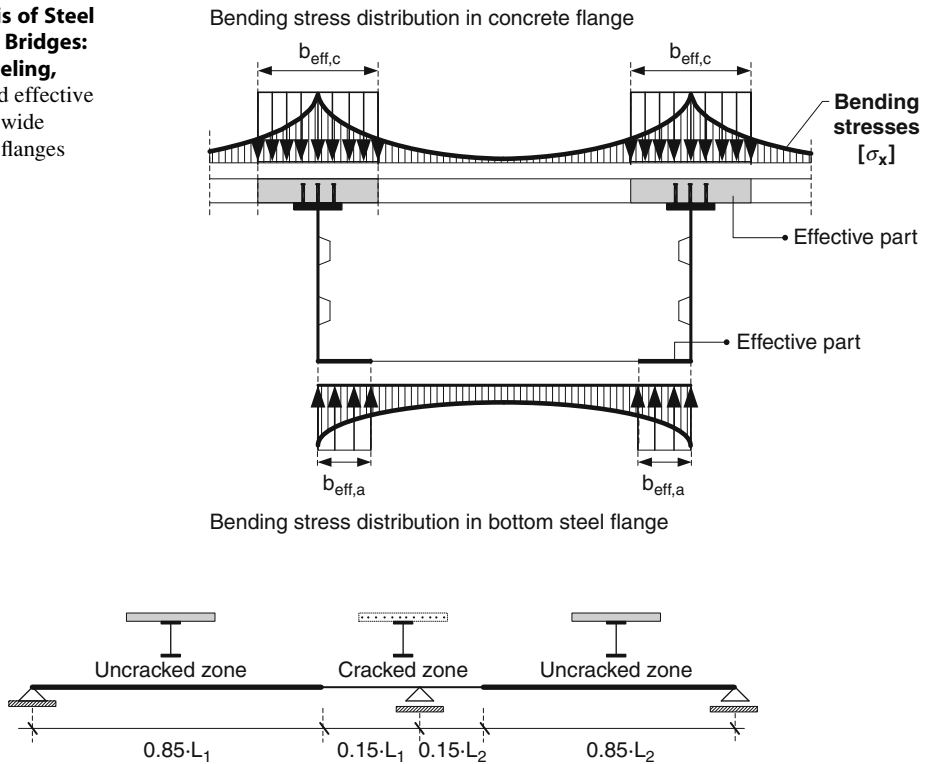
The initial beam cross section is the gross cross section. However, for wide flanges, either of concrete or steel, effective widths must be introduced to consider shear lag effects (Fig. 5). Although effective widths vary along the length of the bridge, they are smaller at internal supports



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 4 Representation of the bridge deck by 7 DOF beam elements

Seismic Analysis of Steel and Composite Bridges: Numerical Modeling,

Fig. 5 Gross and effective cross section for wide concrete or steel flanges



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 6 Determination of the cracked regions for composite bridges

than at spans, fixed values, equal to those at spans, are considered for global analysis. It should be mentioned that effective widths are calculated differently for concrete flanges than for steel flanges (see EN 1993-2 2006 and EN 1994-2 2005).

For composite bridges the flexural stiffness, denoted as $E \cdot I_1$, is calculated for the uncracked section in regions where concrete is in compression. For seismic analysis, the short-term modulus of elasticity of concrete is considered. At hogging moment areas, concrete is in tension for beam-type bridges. Cracking of concrete in those regions is considered by introducing the flexural stiffness $E \cdot I_2$ of the “cracked” section in which the contribution of the concrete slab is neglected. For continuous bridges the cracked region may be considered to be adjacent to the internal supports in a length equal to 15 % of the corresponding span length (Fig. 6). The 15 % rule constitutes a rough

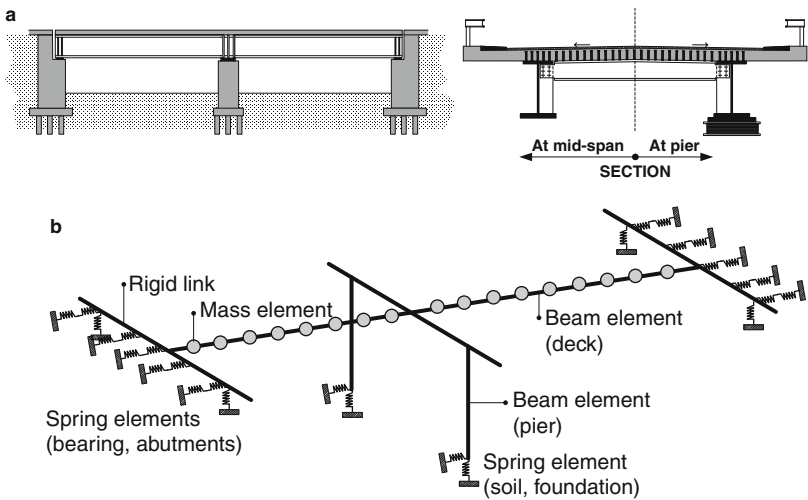
approximation of the true behavior (EN 1994-2 2005), however, for the purpose of seismic analysis is considered as adequately precise.

As an alternative to the fully cracked section, the *tension stiffening effect* of concrete may be taken into account in the cross-sectional properties by increasing the area of steel reinforcement by the factor $1/[1 - (0.5 \cdot f_{ctm})/(\rho_{s, \text{tot}} \cdot f_{sk})]$ where $\rho_{s, \text{tot}}$ is the total reinforcement ratio, f_{ctm} the mean tensile strength of concrete, and f_{sk} the characteristic yield strength of the reinforcing steel (see Vayas and Iliopoulos 2013).

The torsional stiffness $G \cdot I_t$ for composite box girders may be calculated from mechanics, where the shear modulus of concrete is taken into account by introducing 0.2 or 0 as the Poisson’s ratio for uncracked and cracked regions correspondingly.

Although the superstructure is represented by a single beam in spine models, bearings and piers

Seismic Analysis of Steel and Composite Bridges: Numerical Modeling,
Fig. 7 Two-span continuous bridge. (a) Physical model and (b) numerical spine model



appear individually in the model. Figure 7a shows for a two-span continuous composite girder bridge that is supported by two pile bent abutments and one two-column bent. The cross section is a composite section consisting of two I-girders and a concrete deck. The girders rest individually on bearings. Figure 7b shows the spine model for this bridge. The cross section of the superstructure is represented by a single beam element, while each bearing by two horizontal springs. The beam and the springs are coupled by a rigid link that represents the crossbeam. Below the crossbeam are the two piers that are represented by beam elements that rest on translational and rotation springs representing the pile foundation and the soil. Similar conditions apply to the abutments, where each bearing is represented by two horizontal springs, coupled in series with horizontal abutment/foundation springs.

Grillage Models

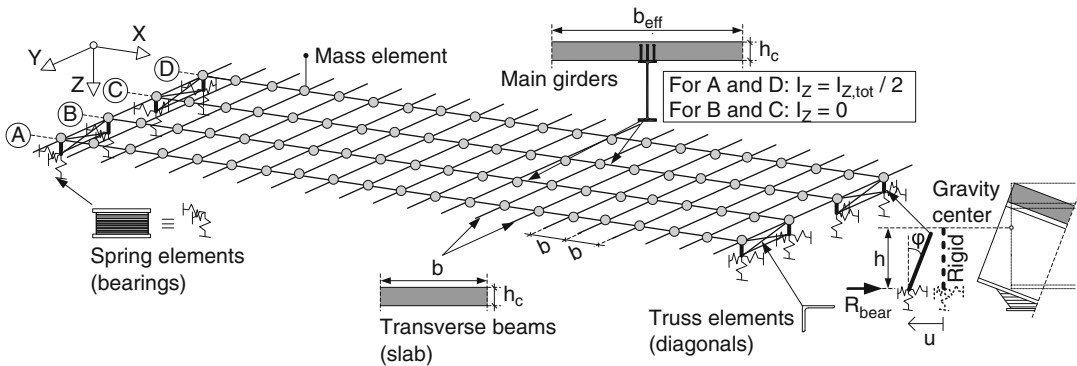
General

The most popular computer-aided modeling method for the analysis of composite bridges is the simulation by means of a plane grillage system. This is due to the fact that this system is easy to apply and comprehend as well as that it has been proved to be sufficiently accurate for a wide

variety of bridge decks. In this model, the structure is idealized by means of a series of longitudinal and transverse beam elements rigidly interconnected at nodes. Each element is given an equivalent bending and torsion inertia to represent the relevant portion of the deck.

Plate Girder Bridges

Figure 8 illustrates a grillage representation of a simply supported composite bridge with four main girders that may similarly be applied to continuous systems or different number of main girders. Longitudinal grillage members are arranged to represent the main girders with the inertia properties of the composite section (steel section with a part of the slab corresponding to the effective width). Transverse members represent the deck slab with thickness h_c equal to the thickness of the slab and width b equal to the distance between transverse beams; it is convenient to select b equal to the distance of the axle loads. A non-cracked flexural rigidity for the slab elements is usually applied. The torsional rigidity of the transverse slab elements can be set to zero. The total in-plane second moment of area of the slab is equally shared to the two extreme main girders (A and D), while the intermediate girders (B and C) are given $I_z = 0$. This is because wind loads mainly act on the edge girders of the bridge. In case of intermediate cross girders whose stiffness may influence the transverse distribution of



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 8 Grillage model of a plate girder bridge

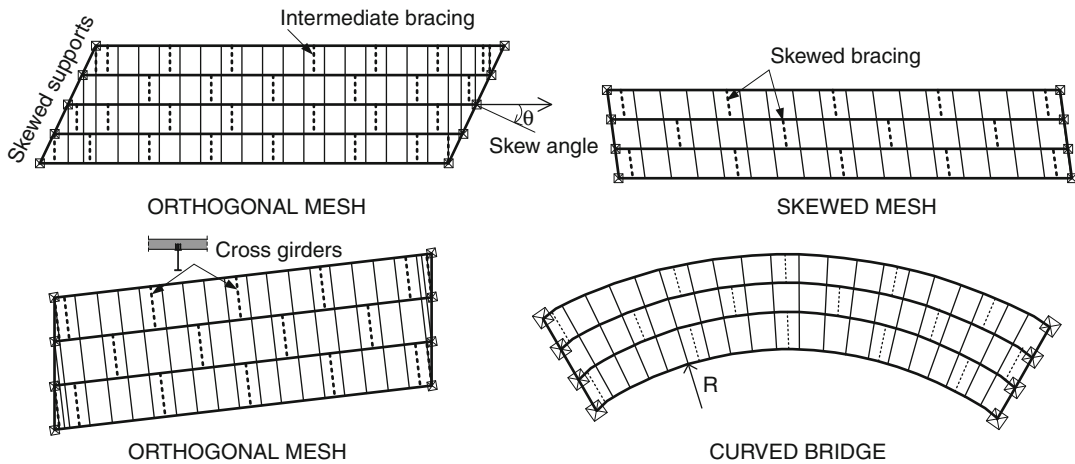
the vertical loads, these girders are taken into account by appropriate modification of the transverse member properties. In case of long distances between adjacent transverse beams, X-bracing concrete elements may need to be set in the deck's plane in order to simulate the diaphragmatic behavior of the deck slab.

Reinforced elastomeric bearings, usually implemented in bridges, are represented by three axial springs of equivalent stiffness corresponding to the relevant stiffness properties in horizontal and vertical directions; the calculation of the bearings' stiffness is presented in section "[Reinforced Elastomeric Bearings](#)." The axes of the main beams coincide with the center of gravity of their cross sections. However, the bearings are positioned beneath the lower flange. Accordingly, rotations of the main girders result in horizontal deformations u of the bearings and additional support reactions R_{bear} . The support nodes are therefore put at a lower level from the grillage members and are connected to the longitudinal beams by rigid vertical bars whose height h is equal to the distance between the center of gravity of the main composite beams and the bottom flange; for better accuracy, the shear center of the cross section should be used which is assumed to be the "real" center of rotation. In case of intermediate cross girders whose stiffness may influence the transverse distribution of the vertical loads, these girders are taken into account with beam elements of an appropriate stiffness.

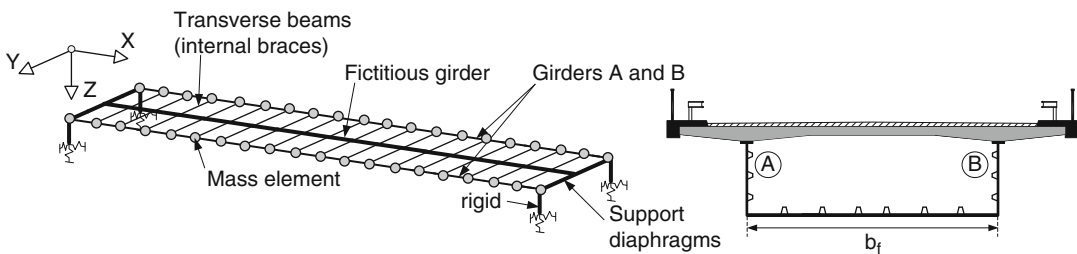
At piers truss elements are used for the representation of the cross braces. Due to the height h of the rigid elements, the geometry of the bracings in model may not follow the exact geometry of the bracings in real structure. A height adjustment for the rigid elements may then be necessary. This should be done only for the purpose of estimating the forces of the bracing members due to horizontal loadings, i.e., wind or earthquake. It has to be stated that in most bridges the gravity center of the composite cross sections is located near the top flange. For such cases a height adjustment has little influence on final results.

Skew Bridges

In skew bridges the support abutments or piers are placed at angles other than 90° from the longitudinal centerlines of the girders (Fig. 9). The skew angle is usually defined as the angle between the longitudinal axis of the bridge and a line square to the supports. The presence of skew affects the geometry and the behavior of the structure. Special phenomena, like twisting and out-of-plane rotation of the main girders during concreting, uplifting forces at bearings, and fatigue problems due to out-of-plane web distortion, make the analysis and design of skewed bridges intricate. The transverse elements representing the slab are usually oriented perpendicular to the main girders (orthogonal mesh); this is the most usual grillage model used by the designers. Alternatively, the transverse members



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 9 Grillage models for skew and curved plate girder bridges



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 10 Grillage model of a simply supported box girder bridge

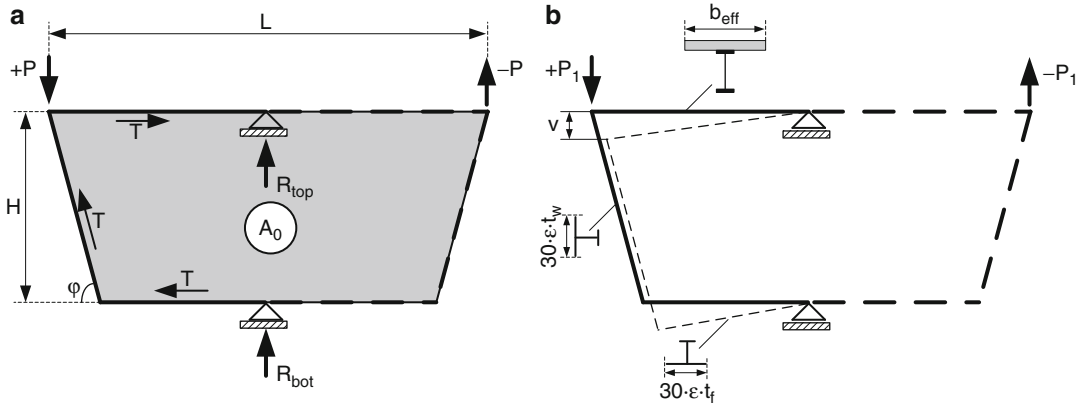
can be placed parallel to the line of supports (skewed mesh). Generally the skewed mesh is convenient for low skew angles ($\theta < 20^\circ$) or when the intermediate bracing is not arranged square to the main girders.

Curved Bridges

Curved decks pose no particular problem for grillage modeling (Fig. 9). A curved bridge deck can be represented by a grillage of curved members or of straight members. Some computer programs support curved members but others do not. Generally, a grillage of straight beams with a very fine mesh is for small values of curvature sufficiently accurate. For highly curved bridges, 3D – or FE – models should be used (Adamakos et al. 2011).

Box Girder Bridges

A grillage model can be implemented also for single-box girder bridges (Fig. 10). The box girder is divided in two opened composite cross sections in which the shear lag effect in the deck slab and the lower flange is considered through the effective widths. The grillage is thus composed of two main composite girders A and B transversely connected with beams representing the internal braces or diaphragms, not the slab. The torsional rigidity of the composite box girder is represented in the model by a fictitious girder located between the main composite girders. The central girder comprises also the whole bending ($I_{Z, tot}$) – and shear stiffness of the deck slab ($A_Y \approx$ slab area). The flanges for girders A and B extend over their effective



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 11 (a) Shear flow of the box section due to *St. Venant* torsion. (b) Deformation of cross frames

widths. Care should be given that these widths are different for concrete and for steel flanges and also different for analysis and for design.

Longitudinal stiffeners in the bottom flange are considered by “smearing” them over the flange width. The bottom flange is then considered having a total thickness of $t_{tot} = t + t_{add}$, where t is the thickness of the panel and $t_{add} = \sum A_{stiffeners} / b_f$. The same procedure is followed for bridges with orthotropic steel deck, where the top flange panel is taken into account with increased thickness due to “smearing” of the longitudinal stiffeners. However, the increase in thickness does not apply to the web panel due to the fact that existing longitudinal stiffeners are usually placed only to enhance its buckling resistance and are not necessarily continuous.

Transverse beams in the global model represent the flexibility of the cross frames or of cross braces. Since a single beam shall represent an entire frame, the definition of its properties requires some preliminary analysis. The beams are considered as rigid in bending but flexible in shear deformations. Accordingly they are assigned an infinite in-plane second moment of area and a shear area A_s . Cross frames or cross braces, and accordingly the transverse beam, resist part of the torsion while they are not strained due to global bending. A global torsion moment M_t is partly resisted by the *St. Venant* action and partly by the cross frames or cross

braces. This torsion is split into antisymmetric loading $\pm P = M_t / L$ (see Fig. 11). The *St. Venant* shear flow is given by

$$T = \frac{M_t}{2 \cdot A_0} = \frac{P \cdot L}{2 \cdot A_0} \quad (5)$$

where A_0 is the shaded area in Fig. 11a.

Therefore, the forces resisted by the cross elements are equal to

$$P_1 = P - T \cdot H_{inclined} \cdot \sin \varphi = P - T \cdot H \quad (6)$$

where $P = M_t / L$. It may be seen that the cross elements are resisting part of the global torsion. If the webs are not inclined ($\varphi = 90^\circ$), the force resisted by the cross elements is equal to:

$$\begin{aligned} P_1 &= P - \frac{P \cdot L}{2 \cdot H \cdot L} \cdot H \\ &= \frac{P}{2}, \quad \text{i.e., 50 \% of the acting forces} \end{aligned} \quad (7)$$

Subsequently the vertical deformation v of the cross frame due to antisymmetric loading $\pm P_1$ as determined before is numerically calculated (Fig. 12). Due to symmetry, only half of the cross section is analyzed and hinges are placed in the middle of the flanges. The two flanges and the web are represented by beam elements with cross

sections composed of the transverse stiffeners or transverse girders and an associated flange width. The effective width for the web and the bottom flange panels is equal to $30 \cdot \varepsilon \cdot t$, $\varepsilon = \sqrt{235 \cdot f_y}$, where f_y = yield stress in MPa and t = thickness of the web or bottom flange.

Orthotropic steel decks are represented in a similar way. For composite bridges with concrete decks, the cross section of the relevant beam element is composed of the transverse girder and the effective width of the concrete slab.

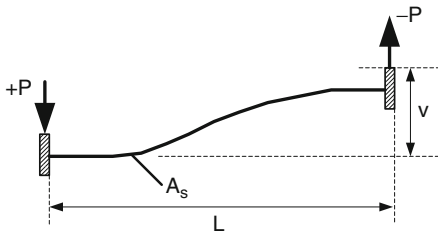
The deformation of a transverse beam is equal to $v = P \cdot L / (A_s \cdot G)$ where G is the shear modulus (see Fig. 12). By setting this deformation equal to the deformation of the cross frame calculated previously, the shear area of this beam is defined as equal to

$$A_s = \frac{P \cdot L}{v \cdot G} \quad (8)$$

where v is the deformation of the system of Fig. 11 due to the load P_1 given before.

Half-Through Bridges

Half-through bridges may be also represented by plane grillage models (Fig. 13). Main and cross



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 12 Deformation of the transverse beam of the global model

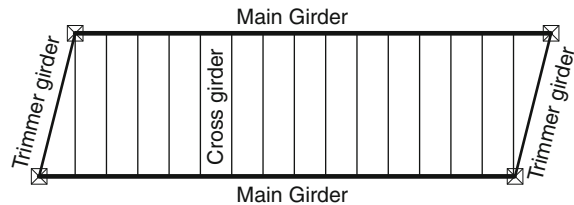
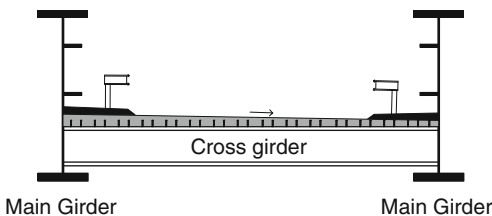
girders are represented by beam elements, the former with steel sections, the latter with composite section including the effective slab width. Lateral torsional buckling phenomena of the top compression flange cannot be captured by this grillage model. 3D models are then recommended.

3D Models

General

The structural representation of bridge decks with truss girders or I-shaped plate girders may be done by means of 3D models as proposed from Vayas et al. (2010, 2011) and Vayas and Iliopoulos (2013). Truss girders are represented by their chord and bracing members, while plate girders are transformed to equivalent trusses. Such models have been proven to be advantageous for modeling orthogonal, skewed, and curved bridges. Unlike grillage models, they are able to consider:

- Eccentricities among the structural elements of a bridge and therefore additional internal forces and possible load distributions
- The transversal variation in the level of the neutral axis
- Torsion and distortional warping effects
- The dispersed structural behavior of the deck slab, in which bending takes place in two directions
- Buckling phenomena of the steel girders during erection stages
- Diaphragms, bracing systems, and stiffeners – possible overload or fatigue effects are taken into account



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 13 Grillage model of a half-through bridge

Girder Representation

Steel and composite I cross sections are modeled by a “hybrid” truss as shown in Fig. 14. For the steel girder, the flanges of the truss are beam elements with a cross section composed of the flange and a part of the web of the steel girder. Comparative analyses showed that one third of the web height may be associated to the flange. Therefore, the flanges of the truss are T-sections consisting of the flange of the steel girder and one third of the web and are positioned at the center of gravity of the T-section. The webs are represented by diagonal truss elements with width equal to one third of the web height and thickness equal to the web thickness. It has been also shown that the cross-sectional area $A_d = h_w \cdot t_w / 3$ for the diagonals adequately corresponds to the shear stiffness of the web. The post-beams are located at a spacing $s \approx 5\%$ of the span of the bridge. This distance is generally acceptable for small and medium span bridges because the angle between the diagonals and the flange elements usually remains between 35° and 45° . Post-beams represent both the in-plane and out-of-plane stiffness of the web.

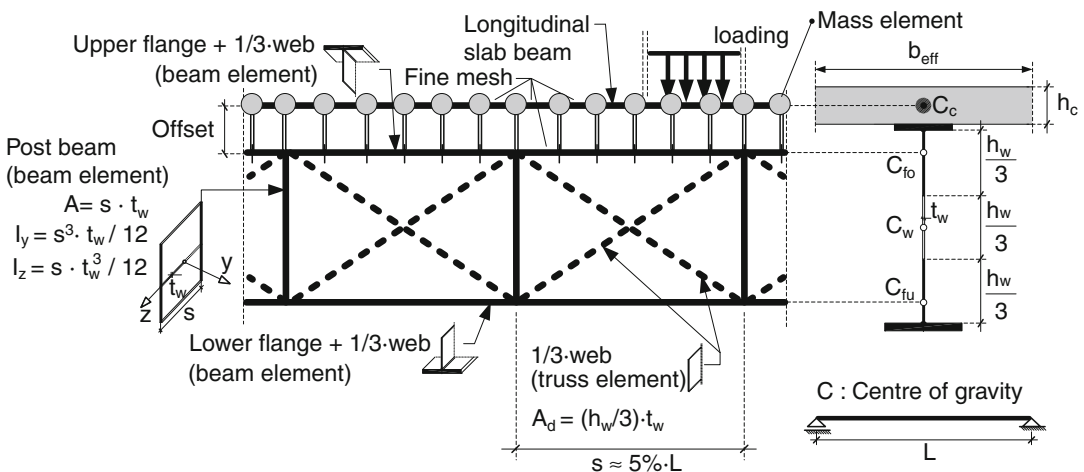
For a composite section the same procedure is followed, with the concrete slab represented by another beam element connected with the upper flange of the truss through the appropriate offset: offset = distance between the centroids C_c

and C_{fo} . The nodes of the elements that represent the slab are the same nodes of those representing the upper flange of the truss. It is recommended that a fine mesh is used for the beam elements of the concrete slab and the top flange of the steel girder so that a full shear connection is achieved. Without a fine mesh the beam elements of the slab may deflect differently than those of the top flange.

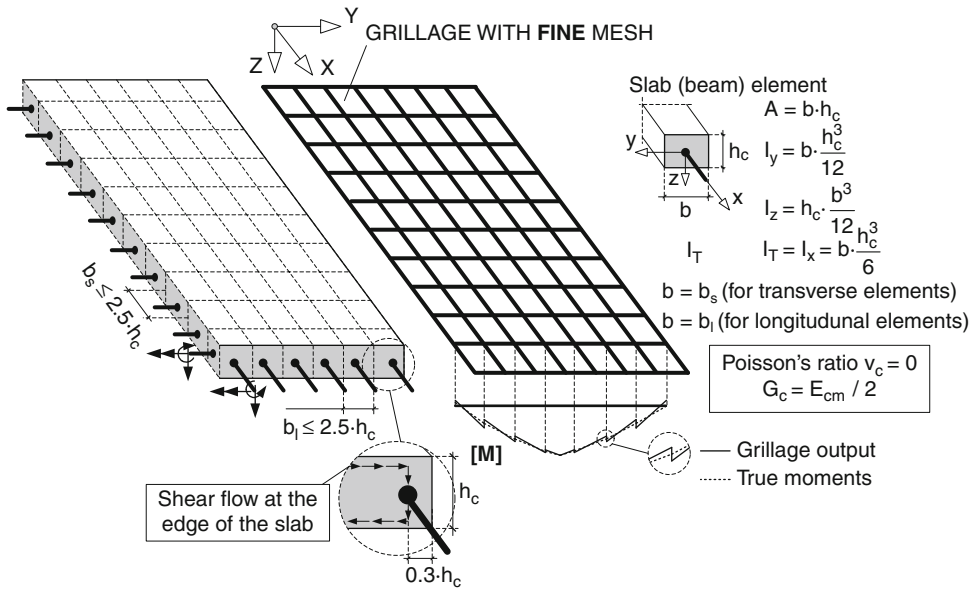
Slab Representation

Slabs are structurally continuous in both directions X and Y, and they resist applied loads by shear forces, bending moments and torques which are coupled with each other. For this reason it was previously mentioned that the transverse slab elements of the grillage models should not be used for the final design of the slab.

A grillage model which considers the dispersed bending and torsion stiffness of a solid slab is illustrated in Fig. 15. The grillage mesh should be sufficiently fine so that the grillage deflects in a smooth surface in a similar way as a real slab. A smooth deflected surface is equivalent to the requirement that the twist $\partial^2 w / \partial x \partial y$ is the same in orthogonal directions and that $m_{yx} = m_{xy}$. The spacing of the beams shouldn't be less than 2.5 times the slab depth. Transverse beams should have spacing similar to that of the longitudinal beams. It is also recommended that



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 14 Truss idealization for a steel-concrete composite girder



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 15 Grillage model for an isotropic solid slab

the row of longitudinal beams at each edge of the grillage should be located in a distance of $0.3 \cdot h_c$ from the edge of the slab, where h_c is the slab depth (see Hambly 1990). This is where the resultant of the shear flows is located. The width of the edge member for the calculation of I_T should be therefore reduced to $b - 0.3 \cdot h_c$.

The 3D Model Implementation

The grillage model for the slab's representation in Fig. 15 can be combined with the truss model which is shown in Fig. 14. Figure 16 illustrates a 3D model which is recommended for the structural analysis both of simple and continuous composite bridges. Attention must be paid so that the grillage has its longitudinal members coincident with the center lines of the steel sections. At sagging moment areas, longitudinal slab elements are used with their uncracked properties. At hogging moment areas, concrete is considered as fully cracked and the total reinforcement is positioned due to simplicity at center of the slab. Transverse slab elements can be considered with their uncracked properties.

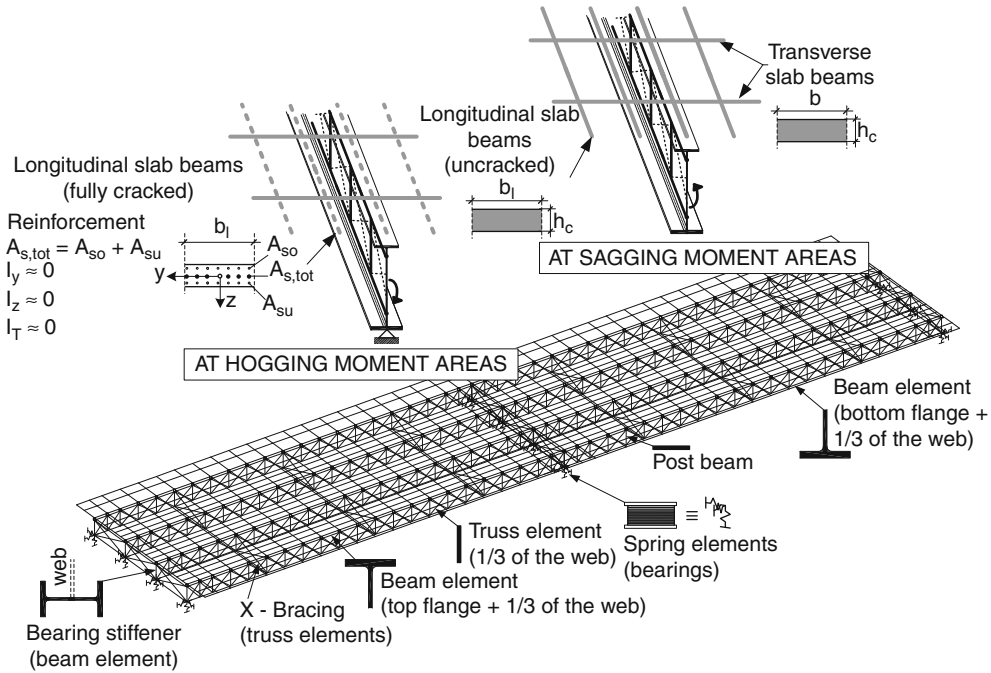
One can see that the model can be set up in a detailed way by taking into account all the

necessary structural elements, i.e., cross-bracings, bearings, etc. Imperfections, precambering, and girders with variable cross sections can also be implemented in the model. Therefore, structural phenomena which may be difficult or impossible to investigate with plane grillages are included in the outputs of the 3D model, e.g., arch effects in integral bridges with longitudinally variable cross sections.

Modeling of Bearings

General

Bridge decks rest usually on bearings that are important elements for seismic analysis and design. Most of the bearings, especially modern bearings, provide seismic isolation to the bridge, reducing the forces on piers, foundations, and the soil. Seismic isolation is provided due to the low stiffness of such bearings that results in a shift to longer fundamental periods and lower spectral values (see Chopra 1995). An appropriate representation of bearings in the structural model is essential for seismic bridge design and will be presented in the following. Bearings may deform



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 16 3D model of a two-span continuous bridge

during seismic action in the nonlinear range. Accordingly linear and nonlinear characteristics are provided.

T_e is the nominal thickness of the elastomer layers.

dir = global direction X or Y.

Reinforced Elastomeric Bearings

Reinforced elastomeric bearings (Fig. 17) consist of uniformly spaced layers of elastomer (natural or synthetic rubber) and reinforcing steel plates that obtain their bonding through the process of vulcanization. The equivalent viscous damping ratio for common elastomers ξ is less than 6 %.

Reinforced elastomeric bearings are introduced as linear springs in global analysis. The spring stiffness in each unrestrained horizontal direction may be obtained from

$$K_{dir} = \frac{A \cdot G_b}{T_e} \quad (9)$$

where:

$A = a \cdot b$ or $\pi \cdot D^2/4$ is the plan area of the bearings.

The shear modulus G_b considered has an increased value compared to $G = 0.9$ MPa of elastomers to account for the speed of loading in the seismic situation so that (see EN 1998-2 2005):

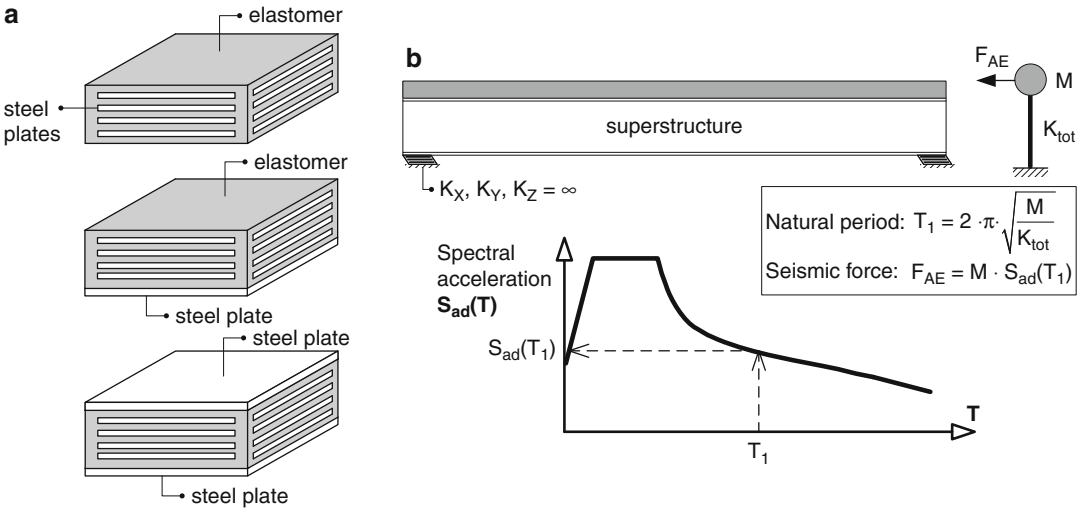
$$G_b = 1.1 \cdot G \quad (10)$$

Beyond this, upper and lower values of the shear modulus are introduced in the seismic combination with recommended values as follows:

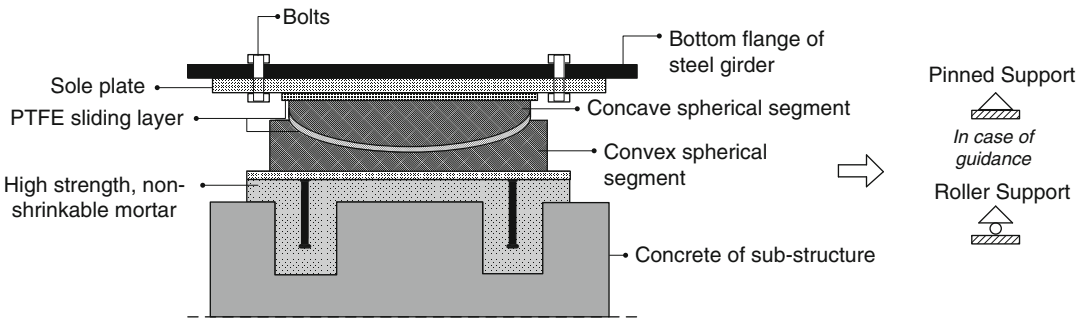
$$G_{b,max} = 1.5 \cdot G_b = 1.65 \cdot G \quad (11a)$$

$$G_{b,min} = 1.0 \cdot G_b = 1.1 \cdot G \quad (11b)$$

Upper values are supposed to result in maximum forces, minimum values maximum displacements. However, this might not be always the case, depending on the resulting periods and



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 17 (a) Composition of reinforced elastomeric bearings. (b) Single DOF seismic analysis of a simply supported composite bridge with a rigid deck model



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 18 Layout and representation of spherical bearings

the spectral values. In vertical direction the bearing is practically incompressible so that the vertical displacement is considered to be restrained without use of springs.

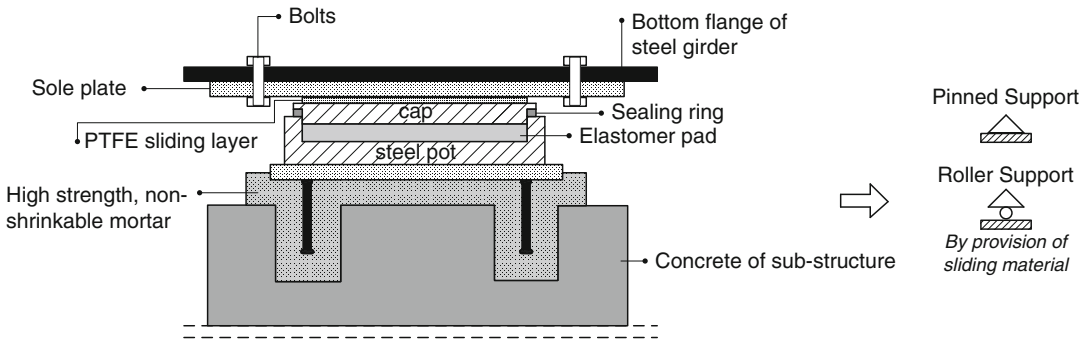
High-Damping Reinforced Elastomeric Bearings

These are bearings in which the common elastomer is substituted by high-damping elastomer. Their equivalent viscous damping ratio ξ reaches values between 10 % and 20 %, while common elastomeric bearings have damping ratios ξ below 6 %. These bearings are modeled in global analysis as linear springs like common elastomeric bearings. The reduction in seismic forces

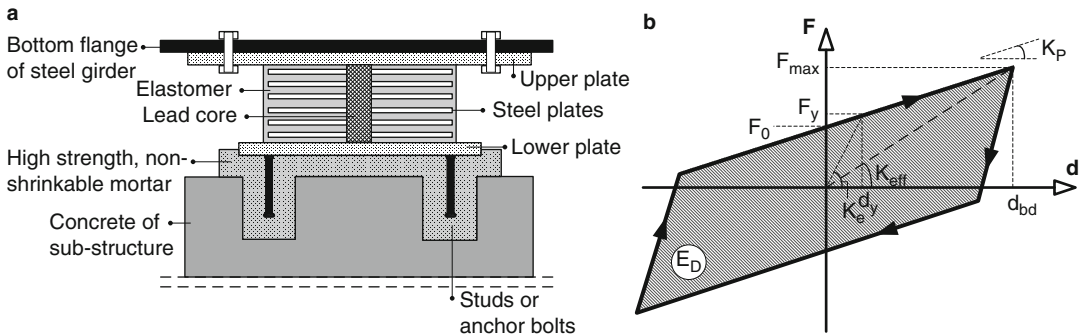
is taken into account by application of the reduction factor $\eta = \sqrt{10/(5 + \xi)} \geq 0.55$ on the forces determined from a response spectrum with 5 % damping.

Spherical Bearings

Spherical bearings consist of a sole plate that transfers loads from the superstructure, a concave spherical segment and a convex spherical segment that provides a mating surface for the concave segment and transfers load to the substructure (Fig. 18). PTFE (Teflon) sliding layers are provided between the three parts to allow horizontal displacements and rotation in all directions. Spherical bearings provide



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 19 Layout and representation of pot bearings



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 20 (a) Layout of lead rubber bearings. (b) Force-displacement behavior of lead rubber bearings

therefore only vertical support. They are accordingly represented by hinges allowing also horizontal displacements. Spherical bearings may be also guided to restrain displacements in one or both horizontal directions. In this case displacements in the relevant horizontal direction are restrained.

Pot Bearings

Pot bearings are based on the incompressibility of natural rubber when placed in a closed steel pot where natural rubber behaves like a fluid (Fig. 19). Pot bearings are able to transfer high compression forces in a small surface and allow rotations around all axes. They provide horizontal and vertical support while rotations are free. Provision of a sliding material allows horizontal displacements in which case the bearings provide only vertical support.

Lead Rubber Bearings (LRB)

These are common reinforced elastomeric bearings with low-damping elastomer and a cylindrical lead core that may reach damping values up to 40 % (Fig. 20a).

During cyclic loading the lead core is yielding and strain hardening so that the hysteretic response of the bearings, as illustrated in Fig. 20b, is typical for a yielding and strain hardening material. In analysis the bearings may be accordingly represented by a bilinear horizontal springs with elastic stiffness K_e for displacements up to the yield displacement and post-elastic tangent stiffness K_p for larger displacements.

The elastic stiffness may be determined from

$$K_e = K_R + K_L \quad (12)$$

where:

K_R is the shear stiffness of the elastomers.

K_L is the shear stiffness of the lead core.

The yield force is determined from

$$F_y = F_{Ly} \cdot \left(1 + \frac{K_R}{K_L}\right) \quad (13)$$

where F_{Ly} is the yield force of the lead core.

The force at zero displacement is equal to

$$F_0 = F_y - K_p \cdot d_y \quad (14)$$

The post-elastic stiffness is attributed to the lead core only and is calculated from

$$K_p = \frac{F_{\max} - F_y}{d_{bd} - d_y} = K_R \quad (15)$$

However, the above bilinear representation is appropriate for nonlinear analyses. For linear analysis the spring properties are linearized by introduction of the secant stiffness. This effective stiffness is written as

$$K_{\text{eff}} = \frac{F_{\max}}{d_{bd}} \quad (16)$$

The analysis must be iterated in this case by introduction of an initial value for K_{eff} and update it consecutively as a function of the resulting maximum displacement of the previous step.

Figure 20b indicates that lead rubber bearings dissipate energy when displaced beyond the yield displacement. The energy dissipation per cycle equals to the shaded area of Fig. 20b and may be determined from

$$E_D = 4 \cdot (F_y \cdot d_{bd} - F_{\max} \cdot d_y) \quad (17)$$

The energy dissipation may be transformed to increased damping with an effective damping ratio ξ_{eff} determined from

$$\xi_{\text{eff}} = \frac{E_D}{F_{\max} \cdot d_{bd}} = \frac{E_D}{2 \cdot \pi \cdot K_{\text{eff}} \cdot d_{bd}^2} \quad (18)$$

where all symbols are indicated in Fig. 20b. For linear analysis the value of ξ_{eff} is updated in each iteration step in order to determine the reduction factor for increased damping on the resulting seismic forces. When the rigid deck model is employed, all bearings are represented by a single spring with an equivalent stiffness and damping ratio determined from

$$K_{\text{eff}, \text{tot}} = \sum K_{\text{eff}, i} \quad (19)$$

$$E_{D, \text{tot}} = \sum E_{D, i} \quad (20)$$

where the summation refers to all bearings i .

The equivalent damping ratio for all bearings is calculated from

$$\begin{aligned} \xi_{\text{eff}, \text{tot}} &= \frac{E_{D, \text{tot}}}{2 \cdot \pi \cdot K_{\text{eff}} \cdot d_{bd}^2} \\ &= \frac{1}{2 \cdot \pi \cdot d_{bd}^2} \cdot \frac{E_{D, \text{tot}}}{\sum K_{\text{eff}, i}} \end{aligned} \quad (21)$$

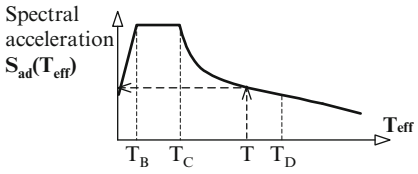
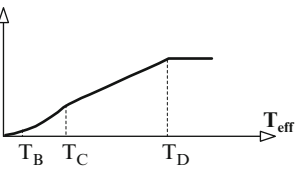
The steps of the iterative procedure for the rigid deck model are as follows:

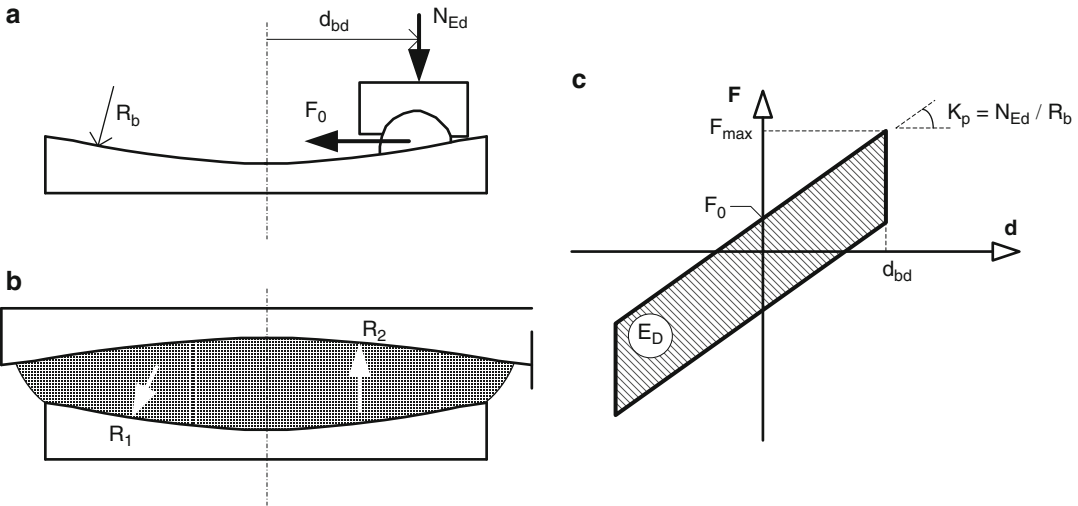
1. Selection of the bearings
2. Initial assumption for the displacement d_{bd}
3. Calculation of $K_{\text{eff}, \text{tot}}$, $\xi_{\text{eff}, \text{tot}}$ and $E_{D, \text{tot}}$ as a function of d_{bd}
4. Determination of the fundamental period of the equivalent system $T_{\text{eff}} = 2 \cdot \pi \cdot \sqrt{M_{\text{tot}}/K_{\text{eff}, \text{tot}}}$ where M_{tot} is the total mass
5. Determination of the seismic forces and the new displacement from Table 1
6. Repetition of steps 2–5 until converge of the displacements is achieved

Friction Pendulum Bearings (FPS)

Friction pendulum bearings constitute a variation of the spherical bearings, where the plane sliding top surface is replaced by a curved one (Fig. 21a, b). They exhibit, unlike spherical bearings, a re-centering capability and are not subjected to rotations. The force at zero displacement is provided by the friction and is given by

Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Table 1 Seismic forces and design displacements (EN 1998-2 2005)

Fundamental period	Seismic forces	Design displacement
$T_C \leq T_{\text{eff, tot}} \leq T_D$	$F_{AE} = \beta \cdot \frac{T_C}{T_{\text{eff, tot}}} \cdot n \cdot A_g \cdot W$	$d_{bd} = \frac{T_{\text{eff, tot}}}{T_C} \cdot d_C$
$T_D < T_{\text{eff, tot}}$	$F_{AE} = \beta \cdot \frac{T_C \cdot T_D}{T_{\text{eff, tot}}^2} \cdot n \cdot A_g \cdot W$	$d_{bd} = \frac{T_D}{T_C} \cdot d_C$
Spectra	Acceleration spectrum	Displacement spectrum
		
Notation	A_g peak ground acceleration, W total weight, and $\eta = \sqrt{10/(5 + \xi_{\text{eff, tot}})} \geq 0.55$	

**Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 21** Layout of (a) single- and (b) double-friction pendulum bearings. (c) Force-displacement behavior of friction pendulum bearings

$$F_0 = \mu \cdot N \quad (22)$$

The kinematics of the bearings leads to following relations:

$$\sin \varphi = d/R \quad (23a)$$

$$\tan \varphi = F/N \quad (23b)$$

good approximation when $d/R \leq 0.25$ and considering the developing friction between surfaces, the following relation may be derived for the restoring force:

$$F = \frac{d \cdot N}{R} + \mu \cdot N \cdot \text{sign}(\dot{d}) \quad (24)$$

where:

Equating the above equations for small angles $\sin \varphi \approx \tan \varphi$ that may be considered as valid with N is the vertical force.

d is the horizontal displacement.

R is the radius of the spherical surface.

μ is the friction coefficient.

$\text{sign}(\dot{d})$ is the sign of the velocity.

It may be seen that the restoring force increases linearly with the displacement and the support reaction, i.e., with the mass. For two spherical surfaces, similar relations apply, where $R = R_1 + R_2$ (Fig. 21b).

Figure 21c displays the force-displacement characteristic of friction pendulum bearings that indicates a symmetrical response. The effective linear spring stiffness is proportional to the compression force and is determined from

$$K_{\text{eff}} = \frac{F}{d} = N \cdot \left(\frac{1}{R} + \frac{\mu}{d} \right) \quad (25)$$

The energy dissipation in a cycle is given by

$$E_D = 4 \cdot \mu \cdot N \cdot d \quad (26)$$

and the equivalent effective damping ratio by

$$\xi_{\text{eff}} = \frac{E_D}{2 \cdot \pi \cdot F \cdot d} = \frac{2 \cdot \mu}{\pi \cdot (d/R + \mu)} \quad (27)$$

When the rigid deck model is employed, the equivalent stiffness of all bearings is determined from

$$K_{\text{eff, tot}} = \sum K_{\text{eff, i}} = W_{\text{tot}} \cdot \left(\frac{1}{R} + \frac{\mu}{d} \right) \quad (28)$$

where the summation refers to all bearings i and W_{tot} is the total weight of the bridge in the seismic situation.

The individual spring stiffness for each bearing i is then

$$K_{\text{eff, i}} = \frac{N_i}{W_{\text{tot}}} \cdot K_{\text{eff, tot}} \quad (29)$$

where

N_i is the compression force of bearing i .

If bearings with identical radius R are selected, the damping ratios for each bearing

and for the overall system are equal to $\xi_{\text{eff, tot}} = \xi_{\text{eff, i}}$.

The steps of the iterative procedure for the rigid deck model are as follows:

1. Selection of the bearings, specifically the friction coefficient μ and the radius R
2. Initial assumption for the displacement d
3. Calculation of $K_{\text{eff, tot}}$ and $\xi_{\text{eff, tot}}$ as a function of d
4. Determination of the fundamental period of the equivalent system
 $T_{\text{eff}} = 2 \cdot \pi \cdot \sqrt{M_{\text{tot}}/K_{\text{eff, tot}}}$, where M_{tot} is the total mass
5. Determination of the seismic forces and the new displacement from Table 1
6. Repetition of steps 2–5 until converge of the displacements is achieved

The equivalent spring stiffness and energy dissipation may then be calculated for each individual bearing.

Fixed Bearings

Until the 1950s, steel bearings were used that consisted of four types: *pins*, *rollers*, *rockers*, and *metal sliding bearings*. Pins are fixed bearings allowing rotations. Rollers and rockers allow translation and rotation, while sliding bearings utilize one plane metal plate sliding against another, with PTFE as intermediate lubricant material, to accommodate translations. These older bearing types are modeled as hinges (see Fig. 22). Fixed bearings may also be represented by springs, their flexibility resulting from the flexibility of the bearing bars and the elongation of the anchor bolts.

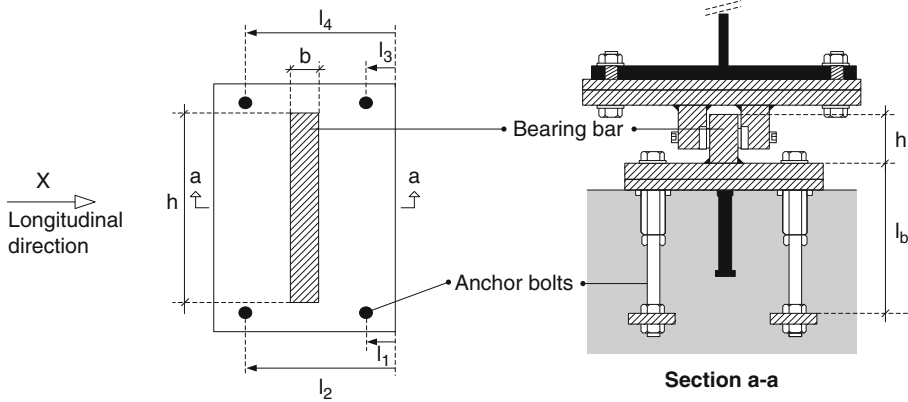
The stiffness in longitudinal direction of fixed bearings may be determined from

$$\frac{1}{K} = \frac{h^3}{3 \cdot E \cdot I} + \frac{h \cdot l_b}{E \cdot A \cdot \sum_1^4 l_i} \quad (30)$$

where:

E is the modulus of elasticity for steel.

I is the second moment of area of the bearing bar ($I = b^3 \cdot h/12$).



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 22 Fixed bearing

h is the height of the bearing bar.

A is the area of the anchor bolts.

l_b is the length of the anchor bolts.

l_i are the distances of the anchor bolts from the tip of the base plate.

The stiffness in transverse direction is given by a similar equation, neglecting the first term due to the high stiffness of the bearing bar in this direction and changing the direction of the lengths l_i . If a bearing set is going to be represented by one translational and one rotational spring, the relevant spring stiffness may be expressed by

$$K_{hor} = \sum_1^n K_i \quad (31a)$$

and

$$K_{rot} = \sum_1^n K_i \cdot l_{bi}^2 \quad (31b)$$

where:

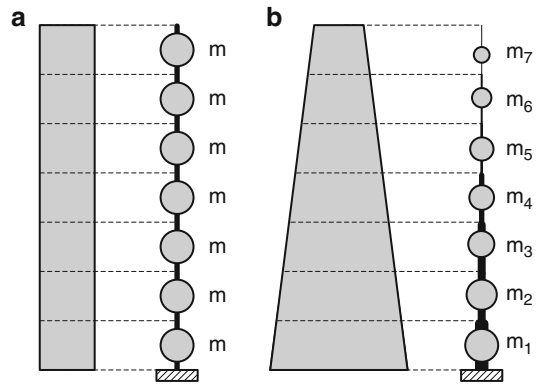
K_i is the stiffness of each bearing as determined before.

n is the number of bearings in the set.

l_{bi} is the distance of bearing i to the centerline of the bridge deck.

Modeling of Piers

Piers are generally represented by means of 6 DOF beam elements. The pier is usually

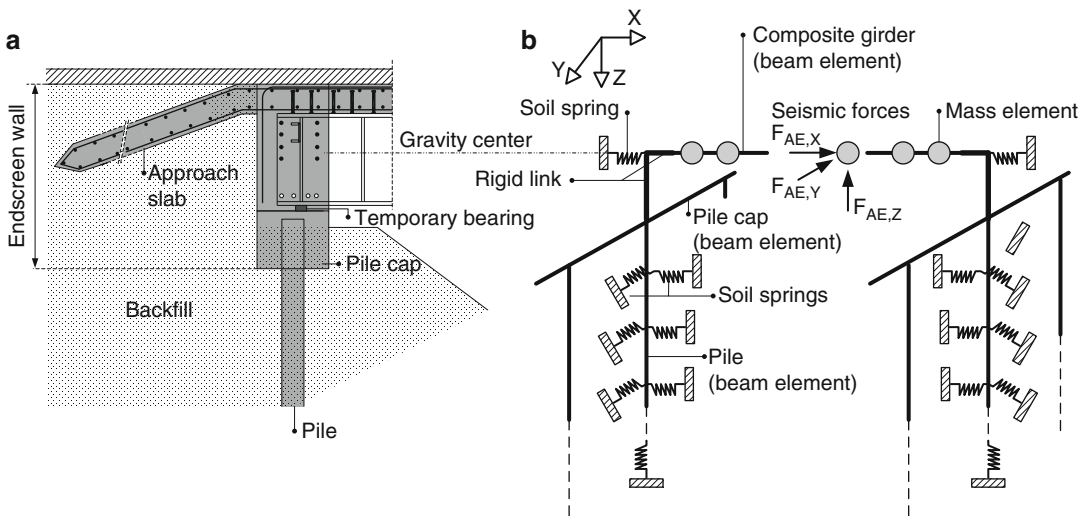
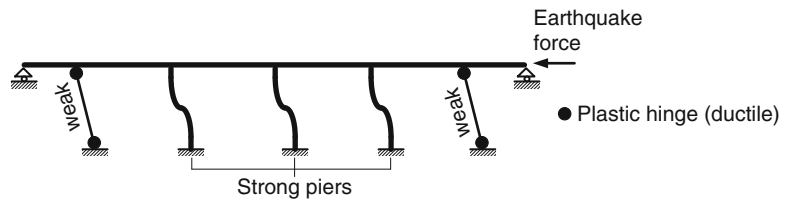


Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 23 Representation of piers through discrete beam mass models – pier with (a) constant and (b) variable dimensions

subdivided in more elements, especially if its cross section varies along the height. The cross section of each element corresponds to the average pier section within its length (Fig. 23).

In frame bridges the superstructure runs continuously and is rigidly connected to the piers so that bearings and expansion joints are avoided. Such bridges are allowed to behave nonlinear in the seismic situation. Unlike in moment-resisting frame buildings where plastic hinges are developing in the beams, plastic hinges are expected to develop in bridges in the piers so that the superstructure behaves elastic (Fig. 24).

Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 24 Plastic hinges in frame bridges



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 25 (a) End connection of superstructure to abutments in integral bridges. (b) Numerical grillage model of the end region of integral bridges

Accordingly for nonlinear analyses, the nonlinear properties of the piers shall be modeled. Two types of models may be used:

- (a) **Plastic hinge models:** The nonlinear behavior of the pier elements is located in rotational springs at the ends of the elastic behavior part of the element. The characterization of a plastic hinge requires a moment-curvature diagram to be defined. This is obtained from the cross-sectional response to monotonic loading that is derived from a finite element analysis.
- (b) **Plastic zone models:** This accounts for spread-of-plasticity effects in sections and along the beam-column element. Beam-column finite fiber elements may be used that are able to better characterize reinforced concrete elements. Consequently, higher accuracy in the structural damage estimate is attained, even for the case of high inelasticity levels.

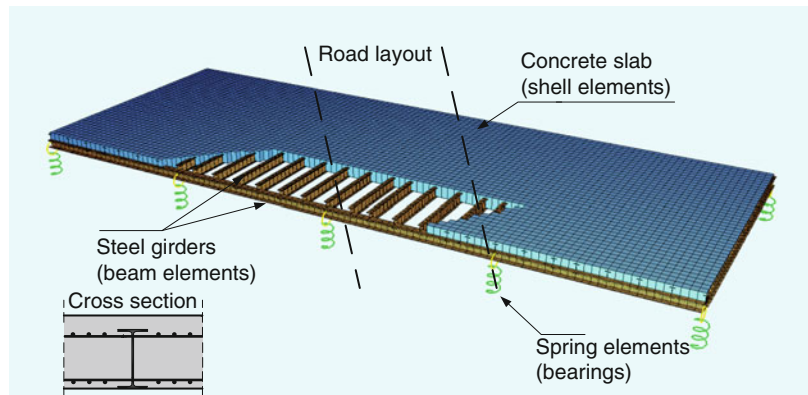
A structural model that includes nonlinearity in a distributed fashion, using finite fiber elements, is able to characterize in higher detail the reinforced concrete elements and thought to capture more accurately response effects on such elements. Geometrical and material properties are the only required ones as input.

Modeling of Integral Bridges

Integral bridges are those where the steel girders are encased in the abutment walls so that bearings and expansion joints are avoided (Fig. 25a). Steel girders are connected to the abutment through reinforcement that passes through their web and by shear studs in the flanges or the web. Movements of the bridge due to thermal actions or time-dependent deformations of concrete are accommodated by the flexibility of piles that

Seismic Analysis of Steel and Composite Bridges: Numerical Modeling,

Fig. 26 Finite element model of a filler deck bridge



support the abutments. To increase the support flexibility, steel piles are oriented with their flanges parallel to the girders so that bending occurs around their weak axis. Integral bridges are allowed for straight and skewed bridges but not for curved bridges. Settlements of the backfill are accommodated by an approach slab. Integral bridges exhibit a good seismic performance due to increased redundancy, smaller displacements, and larger damping due to nonlinear soil-pile-structure interaction.

A grillage model for the end region of integral bridges is illustrated in Fig. 25b. Steel girders and piles are represented by 6 DOF beam elements. Beam elements representing the girders are assigned the properties of the cracked composite cross section due to the fact that negative moments develop in abutment so that the concrete slab is in tension. This section includes the section of the steel girder and the reinforcement within the effective width. Due to the small distance between piles, the pile cap is supposed to be rigid so that piles are connected with horizontal rigid elements and not with beam elements. The connection of the embedded girder and the abutment is supposed to be rigid. Accordingly the beam elements of the girders are connected by vertical rigid elements with the pile cap. Finally, horizontal rigid elements are connecting the beam elements of the girders. The vertical pile support is supposed to be hinged. Springs in two horizontal directions represent the soil that supports the piles, with stiffness varying in depth.

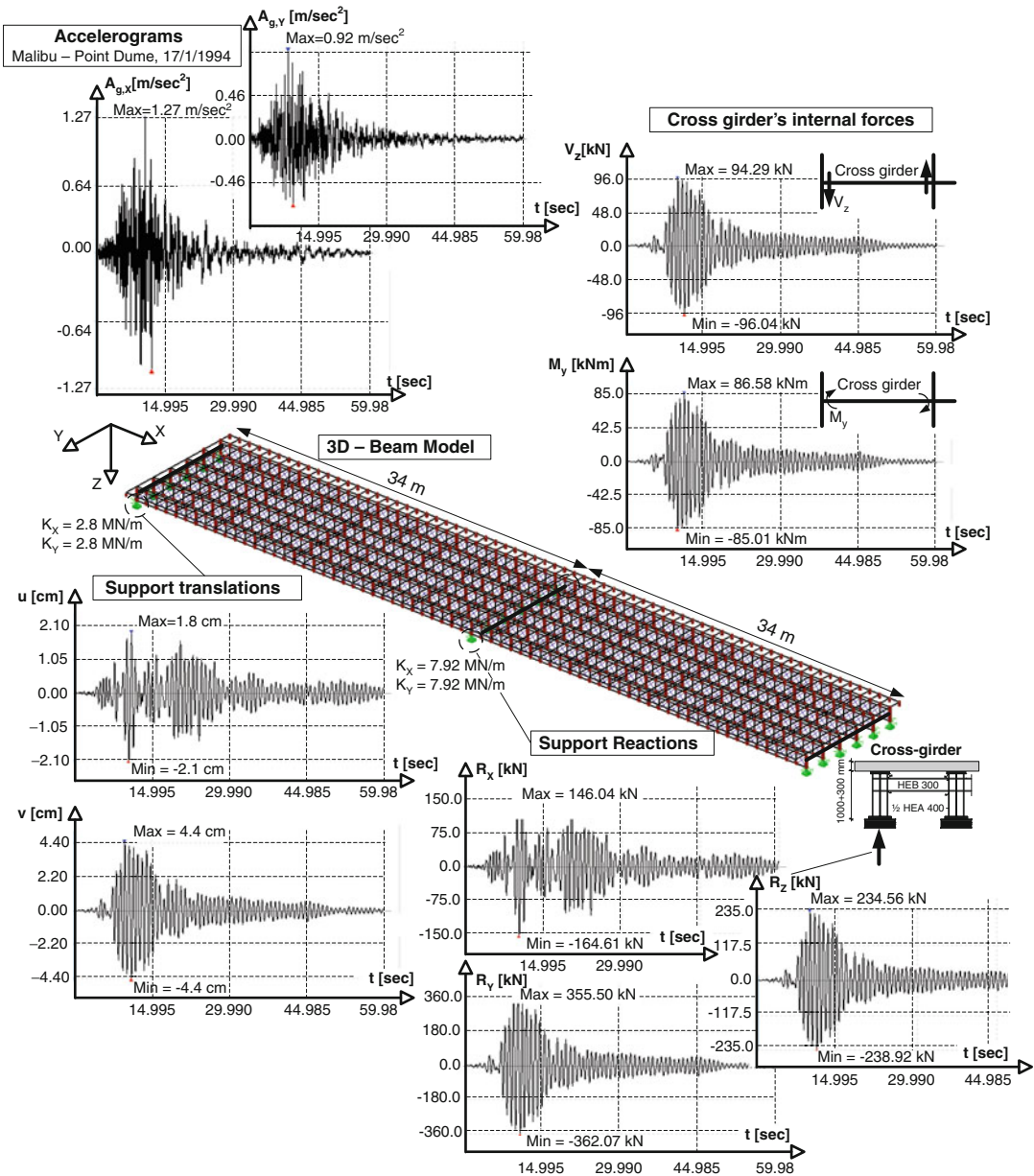
Horizontal springs in the longitudinal direction represent the soil-passive pressure behind the abutment.

When nonlinear methods of analysis are employed, the soil springs and the beam elements of the piles are assigned nonlinear properties. In an extension of the model, the connection between girders and abutment is considered flexible so that compression-only springs are introduced at the contact points between girders and concrete material.

Modeling of Filler Deck Bridges

Filler beam where small distanced steel beams are encased in concrete are small span bridges and are used both as simply supported (max. span ≈ 15 m) and continuous systems (max. span ≈ 30 m). Shear connection is ensured by transverse reinforcement that passes through holes at the webs of the steel beams. Figure 26 shows such a bridge that is composed of closely spaced longitudinal IPE girders resting on two transverse steel H-shaped (HEA section) girders that are supported by reinforced elastomeric bearings. The bridge is orthogonal but the road is skew.

The longitudinal elements are represented by beam elements their section being composed from the steel girder and the associated concrete embedment. The transverse H girders are also represented by beam elements with an HEA cross section. Shell finite elements represent the



Seismic Analysis of Steel and Composite Bridges: Numerical Modeling, Fig. 27 Time history analysis of 2×34 m long span plate girder bridge with a 3D beam model

concrete slab that serve for an easy application of traffic loads, distribute vertical and horizontal loading between longitudinal beams, and ensure diaphragm action. In order to correctly represent the stiffness of the deck, the concrete of the longitudinal beam elements is assigned a very

low modulus of elasticity. However, concrete strength is introduced with its design value in order to allow correct design of the composite action. Deflections are calculated as average values between uncracked and cracked analysis, the latter being simulated by considering the

thickness of the concrete shell elements not as the total slab thickness but as the thickness of the compression zone. Bearings are represented by linear springs. The above-described model is a combination of a grillage and a FEM shell element model.

Application: Time History Analysis with a 3D Beam Model

A time history analysis of a two-span composite bridge with six main girders is indicatively demonstrated in Fig. 27. The bridge has been modeled according to the recommendations of the section “3D Models.” The superstructure rests on low-damping elastomeric bearings whose stiffness has been estimated through Eqs. 8, 9, 10, and 11a and 11b. For the analysis a damping factor equal to 2 % for the three first natural modes has been implemented, value taken from the fabricator of the bearings. The mass of the superstructure includes all the permanent loads and 20 % of the traffic loads. One can see that with the 3D beam model, the time variation of the vertical reaction forces R_z can be calculated; such a calculation with a grillage model would not be feasible. Moreover, the internal forces of the cross girders are calculated. Not only the grillage model but a detailed FE model as well would not give the possibility of such a calculation since shell or volume elements offer stresses and not forces or bending moments as final results. With the 3D beam model, the frame action of the cross girders (HEB 300) and the transverse stiffeners (1/2 HEA 400) could be adequately investigated.

Summary

Different types of models are used for the design and seismic analysis of steel and steel-concrete composite bridges. In this entry 1D, grillage, spine, 3D beam, and finite element models are presented. The advantages and drawbacks of each model are debated, and modeling

recommendations for plate, filler deck, and box girder bridges are provided. The implementation of structural bearings in the models is also discussed, and guidance on the calculation of the bearings’ stiffness and the seismic forces is given. The entry ends with an example of a time history analysis of a continuous composite beam and the calculation of reaction forces and displacements.

Cross-References

- [Buildings and Bridges Equipped with Passive Dampers Under Seismic Actions: Modeling and Analysis](#)
- [Seismic Analysis of Concrete Bridges: Numerical Modeling](#)

References

- Adamakos T, Vayas I, Petridis S, Iliopoulos A (2011) Modeling of curved composite I-girder bridges using spatial systems of beam elements. *J Construct Steel Res* 67(3):462–470
- Chopra AK (1995) *Dynamics of structures*. Prentice Hall, Upper Saddle River
- EN 1993-2 (2006) CEN European Committee for Standardization. *Design of steel structures. Part 2: Steel bridges*
- EN 1994-2 (2005) CEN European Committee for Standardization. *Design of composite steel and concrete structures. Part 2: General rules and rules for bridges*
- EN 1998-2 (2005) CEN European Committee for Standardization. *Design of structures for earthquake resistance. Part 2: Bridges*.
- Hambly EC (1990) *Bridge deck behaviour*, 2nd edn. E. F. & N. Spon, London
- Kindmann R, Kraus M (2011) *Steel structures – design using FEM*. Ernst & Sohn, Berlin
- Vayas I, Iliopoulos A (2013) *Design of steel concrete composite bridges to Eurocodes*. CPC Press, Taylor and Francis Group, Boca Raton
- Vayas I, Iliopoulos A, Adamakos T (2010) Spatial systems for modelling steel-concrete composite bridges – comparison of grillage systems and FE models. *Steel Construct Des Res* 3:100–11
- Vayas I, Adamakos T, Iliopoulos A (2011) Three dimensional modeling for steel-concrete composite bridges using systems of bar elements – modeling of skewed bridges. *Int J Steel Struct* 11(2):157–169

Seismic Analysis of Steel Buildings: Numerical Modeling

Charis J. Gantes

Institute of Steel Structures, School of Civil Engineering, National Technical University of Athens, Zografou Campus, Athens, Greece

Synonyms

Numerical simulation of steel buildings; Structural design of steel buildings in seismic regions

Introduction

The aim of this contribution is to present the state of the art in numerical modeling of steel buildings, aiming at their structural analysis and design. To that effect, the successive stages of this process will be presented, including the conceptual design, numerical modeling, application of support conditions and loads, structural analysis, and design checks. The two latter issues, structural analysis and design checks, will be discussed briefly, to the extent that they interact with the modeling process, as they are also covered in detail in other parts of this encyclopedia. The interaction with other engineering disciplines throughout the design process will be demonstrated. For the purpose of this presentation, a prototype building will be used, which is described in the next section.

Description of Prototype Building

A simple single-story laboratory building, part of an industrial project, will be used as prototype building for this presentation. The building's plan view is 28.00 m by 15.00 m (Fig. 1), and there is a small loft in part of the building (Fig. 2). The roof is double pitched with minimum required clearance equal to 4.50 m (Fig. 3). The locations of exterior doors are fixed, imposing restrictions on

the layout of structural elements, as will be described in the following section. The preliminary architectural drawings include proposed column locations, indicated by their I-sections in the plan views of Figs. 1 and 2.

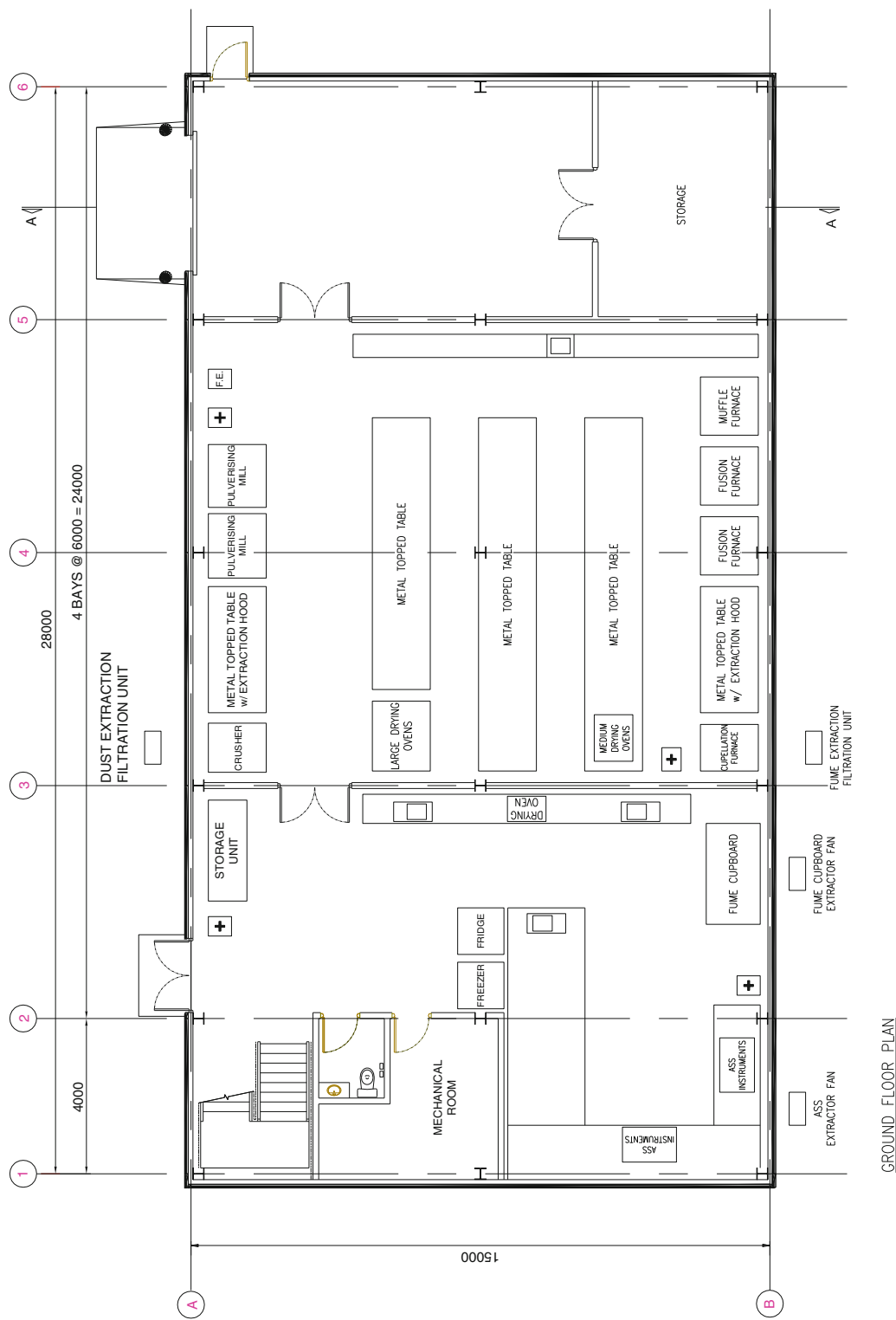
Conceptual Design

The initial stage of the structural design process consists of the so-called conceptual design, in which the different structural elements that are needed to support the structure are selected and are geometrically defined, taking into account the architectural requirements. In parallel, all restrictions imposed by the building's use, as well as its location and associated environmental actions have to be satisfied.

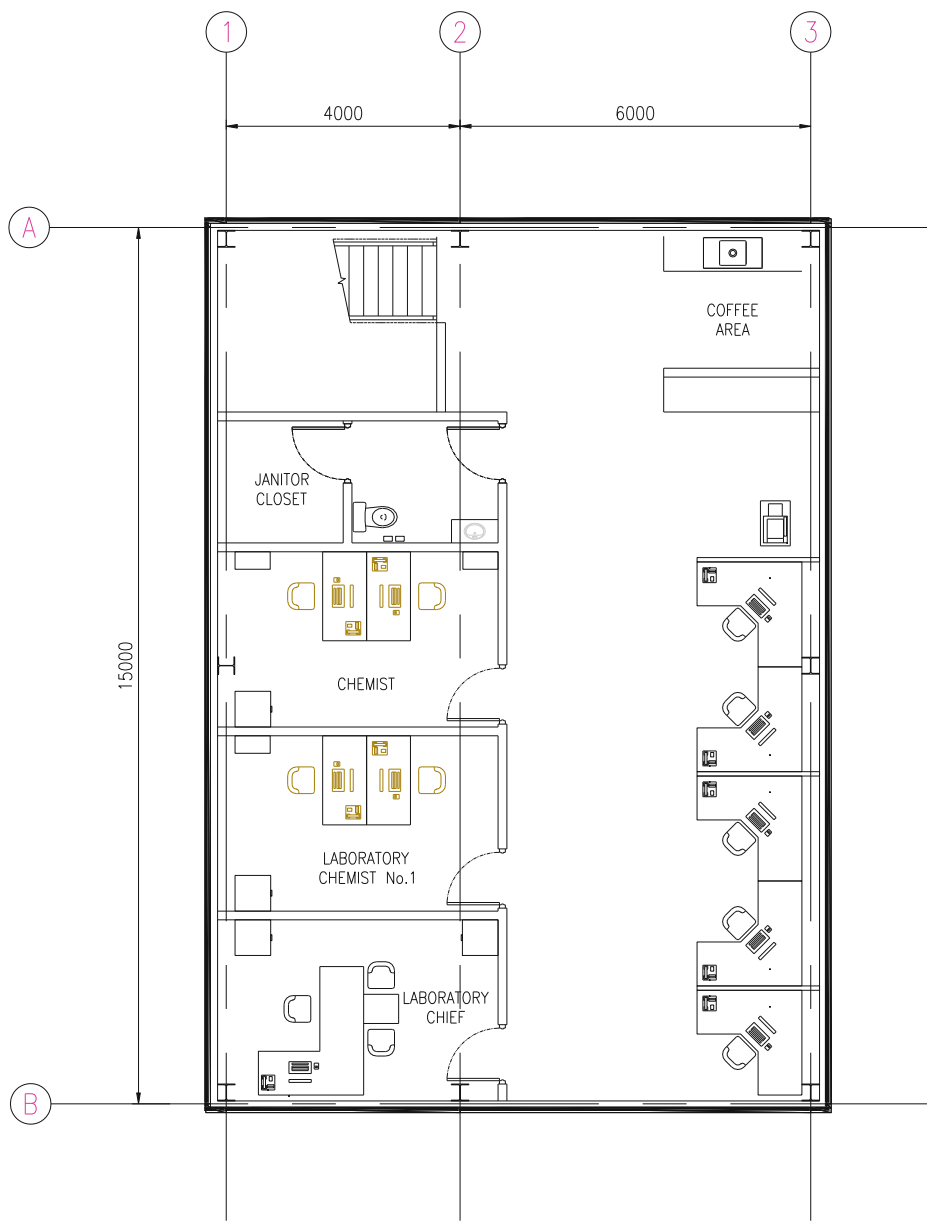
To that effect, close cooperation and interaction of the structural engineer with other involved engineering disciplines are necessary. Such disciplines include, at a minimum:

- The architectural engineering team, to determine the overall geometry; positions of openings; types of materials for floors, cladding, etc.; and associated dead loads and, in general, to ensure that structural choices do not interfere with the building's desired function and aesthetics
- The mechanical/electrical/HVAC engineering team, to determine loads that are associated with these functions of the building and to prevent conflicts with required shafts and other openings for passage of lifelines
- The geotechnical engineering team, to obtain data about the soil and its mechanical properties and to determine the type of the building's foundation

In conceptual design, all basic decisions, regarding material properties, structural system, loads, and codes to be applied, are taken. All structural engineering experts agree that this is the most important design stage and that these decisions affect significantly the project's design and construction schedule and



Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 1 Architectural plan view of ground floor of prototype building (out of scale)



FIRST FLOOR PLAN

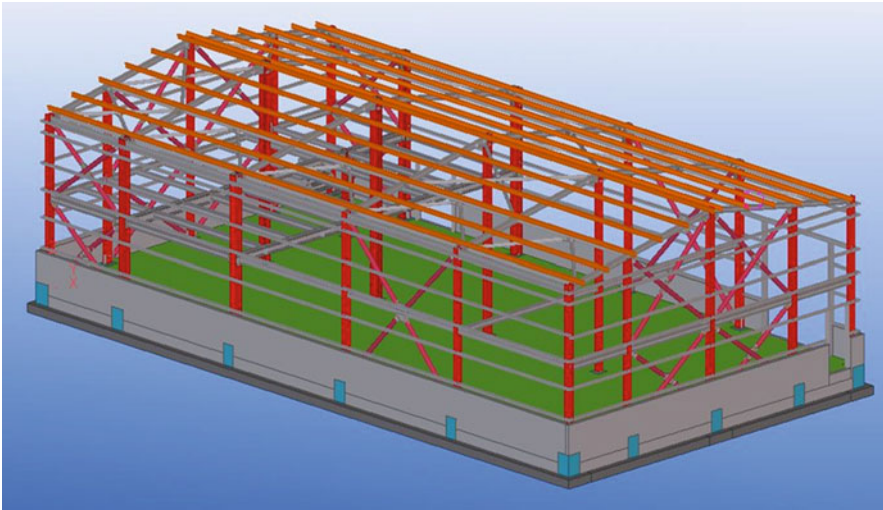
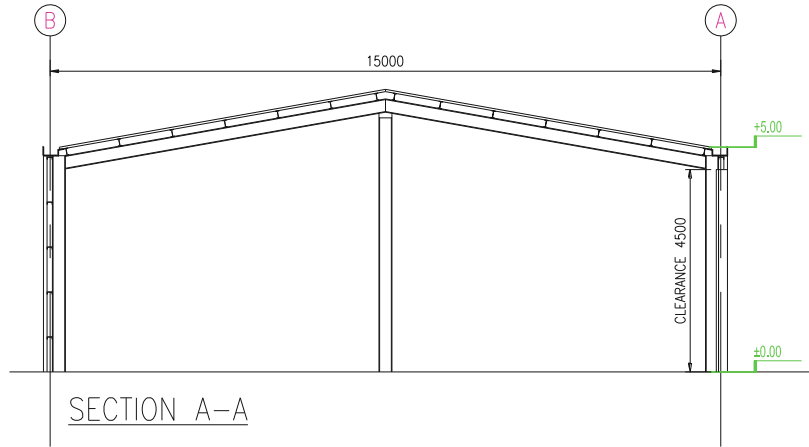
Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 2 Architectural plan view of loft of prototype building (out of scale)

cost, sometimes even its feasibility. This is a process mostly dictated by the available experience of all participating design teams but can nowadays be assisted by modern software

tools, both for preliminary design of the different building’s facilities but also for communication between involved teams (building information modeling, BIM). However, a

Seismic Analysis of Steel Buildings: Numerical Modeling,

Fig. 3 Transverse section A-A of prototype building (out of scale)



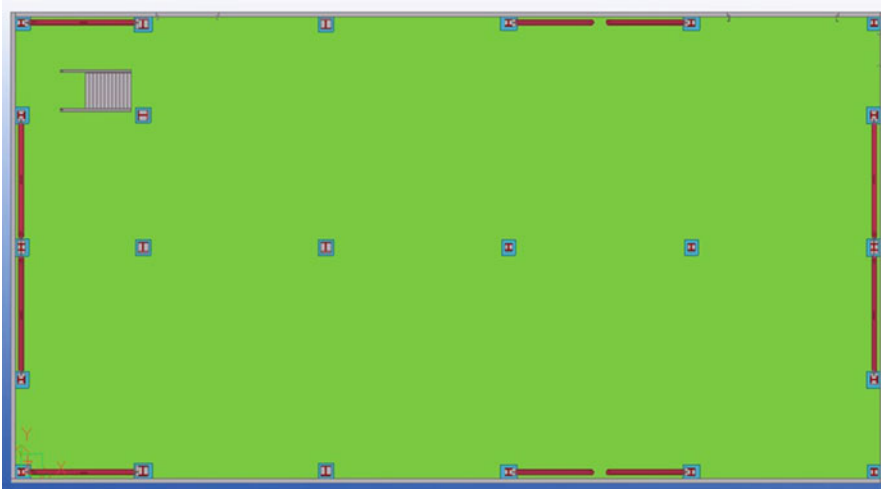
Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 4 3D view of prototype building's 3D model

more detailed treatment of this design stage is beyond the scope of this presentation.

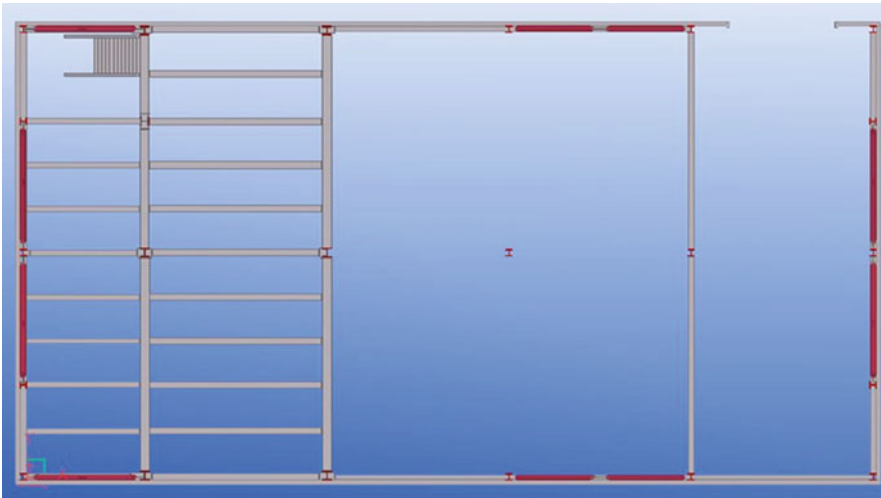
Several conceptual design choices for the prototype building described in section “[Description of Prototype Building](#)” are illustrated in Figs. 4, 5, 6, 7, 8, 9, 10, and 11. The figures have been obtained from a 3D model of the building's structure, made with a specialized modeling software, which enables integration of architectural and structural engineering and other solutions and production planning systems, working for both detailing and producing steel structures of any type.

The building's structural system consists of:

- Two-span main frames spaced at 6.00 m (with the exception of the first span, which is equal to 4.00 m) in accordance with the architectural proposal. Columns and girders have I-section, oriented so that the strong axis acts for in-plane loading. It is noted that a 6.00 m distance between main frames is optimum for avoiding loss of material in purlins and longitudinal beams, taking into account that steel members are commonly fabricated in 12.00 m lengths. Haunches are provided between



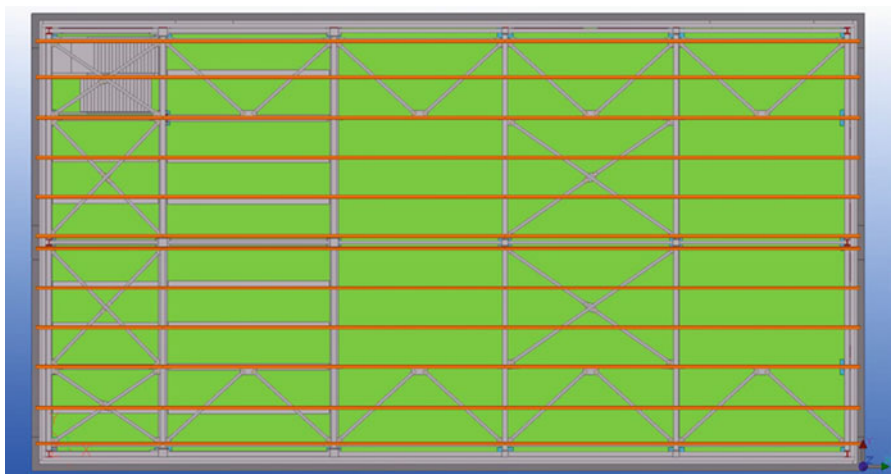
Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 5 Plan view of prototype building's 3D model above foundation level



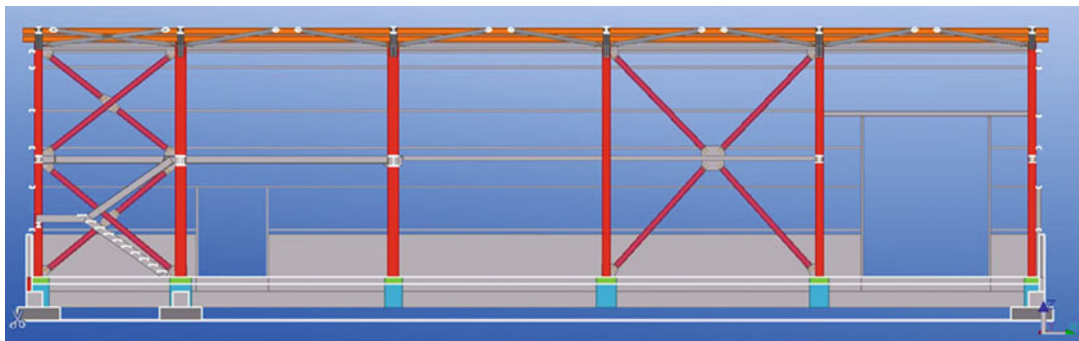
Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 6 Plan view of prototype building's 3D model above loft level

columns and girders at internal frames, in order to achieve reduction of girder cross section and to accommodate the required number of bolts for a moment connection. External frames are equipped with vertical X-bracing, so that truss action dominates rather than frame action. As a result, smaller column and girder sections are possible, and there is no need for haunches between columns and girders.

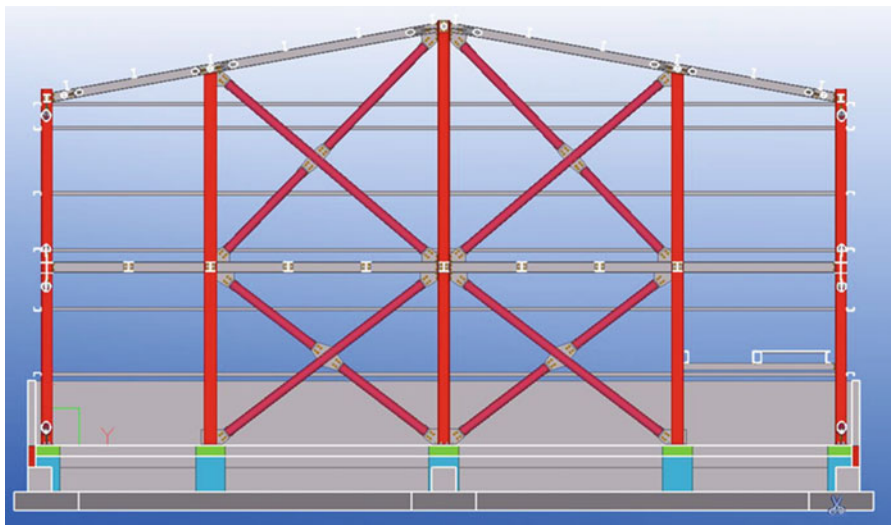
- Longitudinal beams connecting the main frames at the top and at mid-height of columns. On one side of the building, the mid-height head beam is eliminated in the last span, to accommodate a required door at that location. A beam over the door is used instead. It is noted that this is not an optimum structural configuration, but it is an acceptable compromise to respect functional requirements.



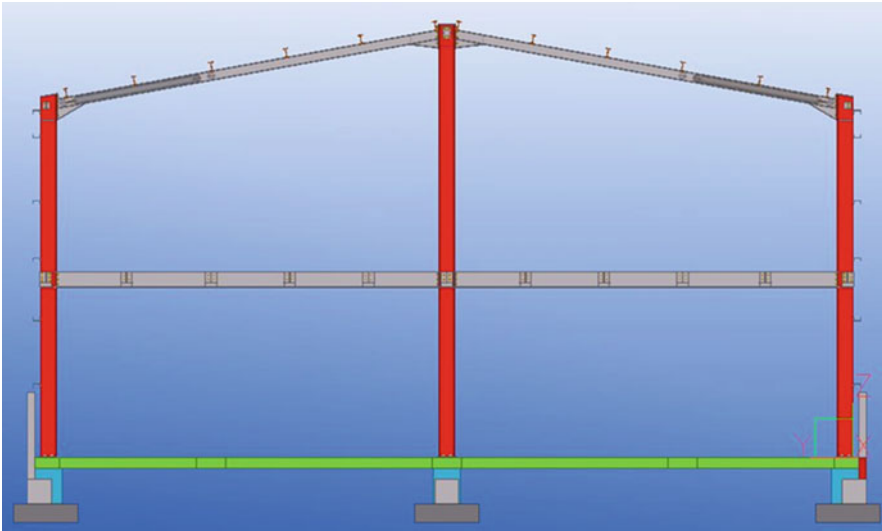
Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 7 Roof plan view of prototype building's 3D model



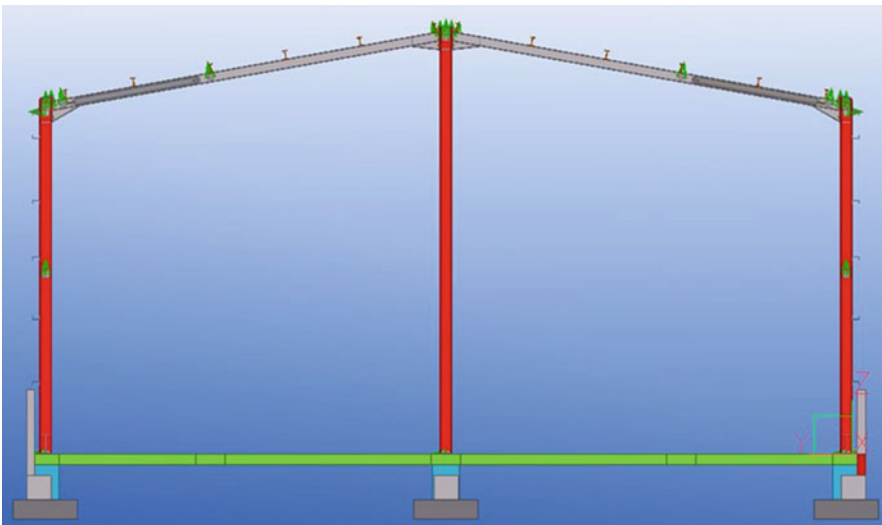
Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 8 Side view of prototype building's 3D model



Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 9 Front view of prototype building's 3D model



Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 10 Transverse section of prototype building's 3D model at spans with loft



Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 11 Transverse section of prototype building's 3D model at spans without loft

- Vertical X-bracing at two spans to provide lateral stability in the longitudinal direction. The type of bracing depends on panel geometry, aiming at inclination that is as close as possible to 45° .
- Horizontal roof bracing at the same spans where vertical bracing is provided, as well as

at the roof's peripheral panels, to ensure sufficient diaphragm action.

- Roof and side purlins to support cladding.

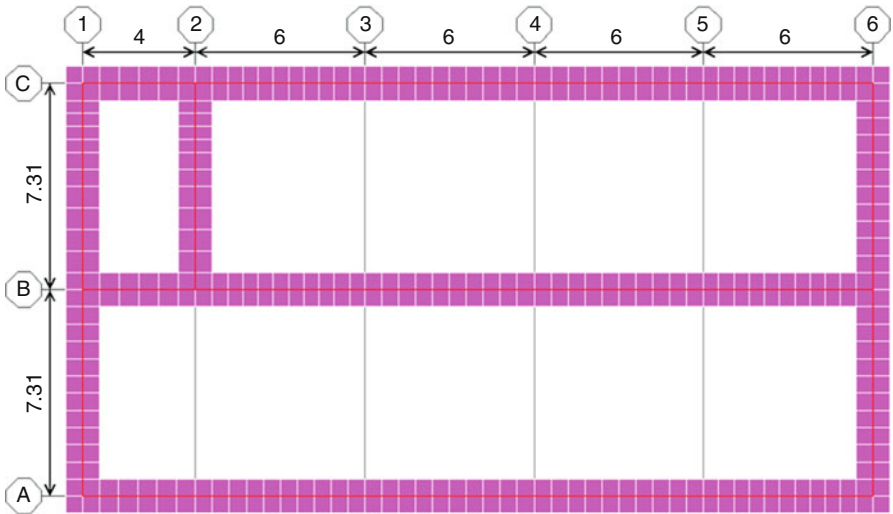
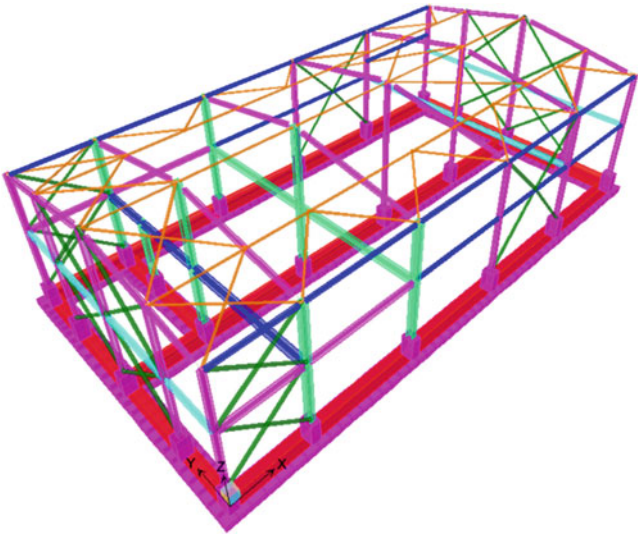
A peripheral reinforced concrete wall is provided, to ensure water tightness and to protect the steel columns from eventual vehicle collisions.

Taking also into account the available geotechnical information for this site, foundation beams are provided at the periphery of the building, along the axis of intermediate columns and below the auxiliary columns supporting the loft. A steel staircase is provided to connect the ground floor with the loft.

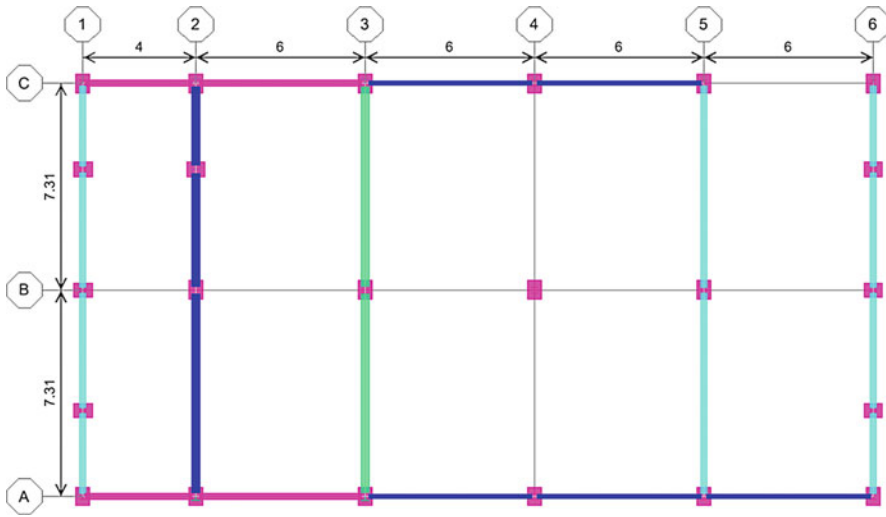
Numerical Modeling

For the simulation, analysis and design of superstructure and foundation use is made of an appropriate finite element software. Views of the finite element model are shown in Figs. 12, 13, 14, and 15.

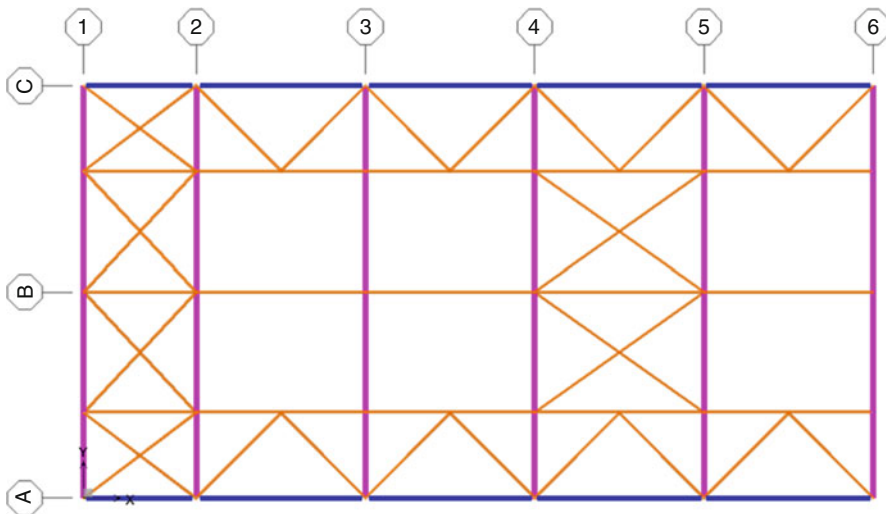
Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 12 3D view of prototype building's finite element model



Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 13 Foundation plan view from prototype building's finite element model



Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 14 Loft's plan view from prototype building's finite element model



Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 15 Roof's plan view from prototype building's finite element model

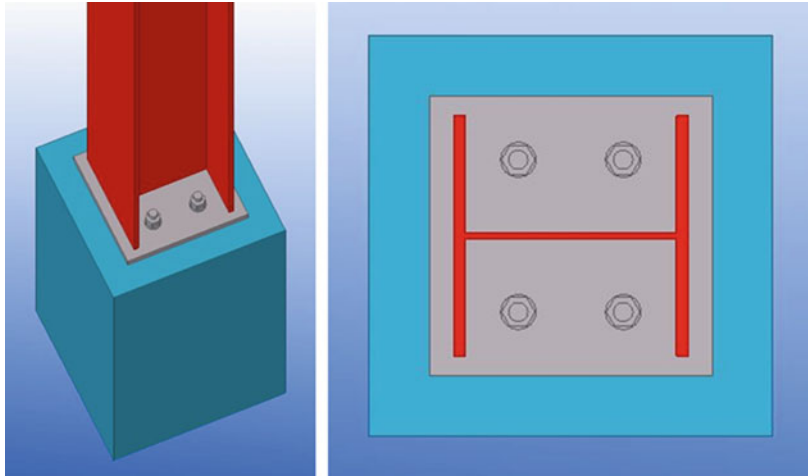
Some recommendations of good practice when setting up the finite element model are the following:

- It is recommended not to include roof and side purlins and other secondary members in the numerical model. The type of connection between purlins and main frame, which is a simple support so that both purlins and frame

columns and girders are continuous but do not transmit moments to each other, would require rigid links for correct modeling, which is possible in most structural programs, but would render the model unnecessarily complex. Only those members that have an active role in the building's spatial response to either vertical or horizontal actions need to be modeled. Other members can be modeled, analyzed, and

Seismic Analysis of Steel Buildings: Numerical Modeling,

Fig. 16 Column base connections of prototype building, modeled as hinges



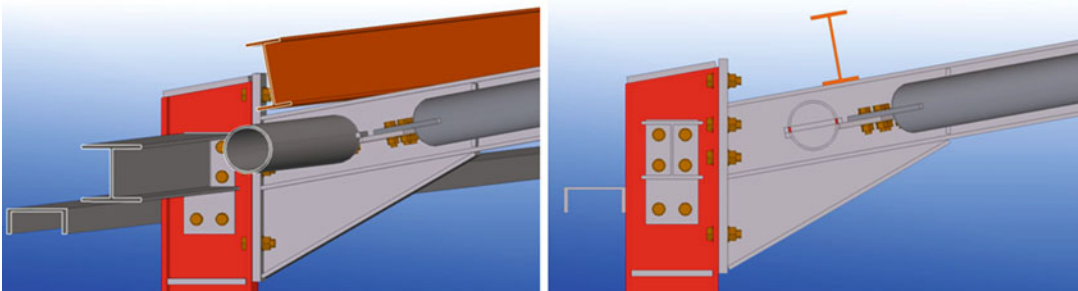
dimensioned independently, and the loads they transfer to the frames can be applied directly upon the frame columns and girders in the 3D finite element model.

- Haunches can be modeled by means of uniform beam section over the entire length of the haunch, usually the one at haunch mid-length. This provides a satisfactory representation of increased girder stiffness in the area of the haunch and enables safe prediction of strength for the design phase.
- Particular attention is needed for the appropriate modeling of supports and connections, by releasing the appropriate member degrees of freedom depending on the connection detail. Even though real connections between steel members are neither perfect hinges nor perfect moment connections, it is common practice to model all connections as either hinges or moment connections, at least initially. Modern codes provide tools for eventually classifying connections as semirigid (e.g., European Committee for Standardisation 2003c), in which case either the connections should be stiffened to become rigid or a rotational spring should be inserted in the model to represent the partially restricted connection between the members in question. The decision about the type of joint to be included in the model depends mainly on the lever arm between bolts in the connection. Connections of I-beams where all bolts are between the

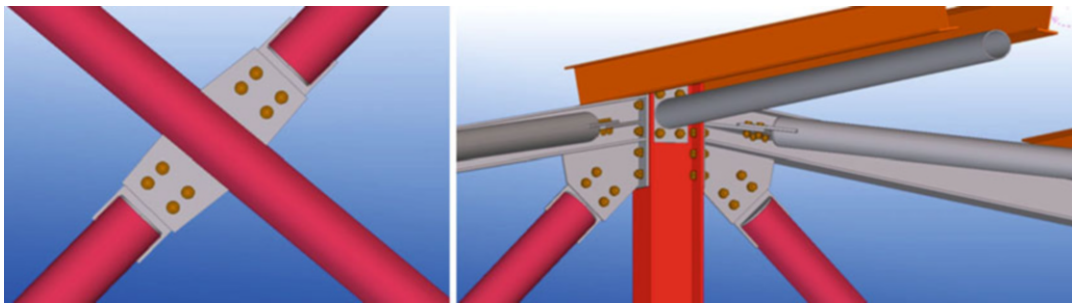
flanges are commonly modeled as hinged, as is the case with the supports of all columns in the prototype building (Fig. 16). In case bolts are provided outside the flanges, rigid moment connections are considered in the model, as is the case with the column to girder connections in the prototype building (Fig. 17). Welded connections are usually modeled as rigid, while connections between or at the end of bracing members (Fig. 18) are modeled as hinged.

For moment connections, no intervention is required in the model, as this is the default option for frame structures. Hinged connections are modeled by means of releases of appropriate rotational degrees of freedom at the members' ends. In Figs. 19 and 20, moment releases at the ends of roof and vertical bracing as well as secondary beams are denoted by black dots.

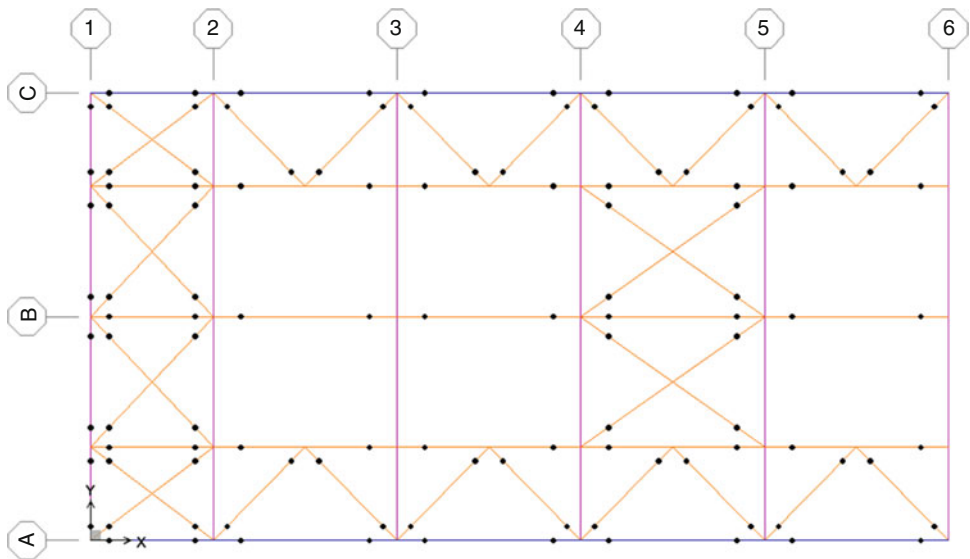
- If there is a diaphragm in the structure, as, for example, is the case with the composite slab of the loft in the prototype building, it is recommended to avoid the more complex option of modeling it with shell elements and instead create a "master" node that will be connected with all nodes of the diaphragm (Fig. 21). In this case, in order to account for uncertainties in the location of masses and in the spatial variation of the seismic motion, the calculated center of mass at each floor shall be considered as being displaced from its



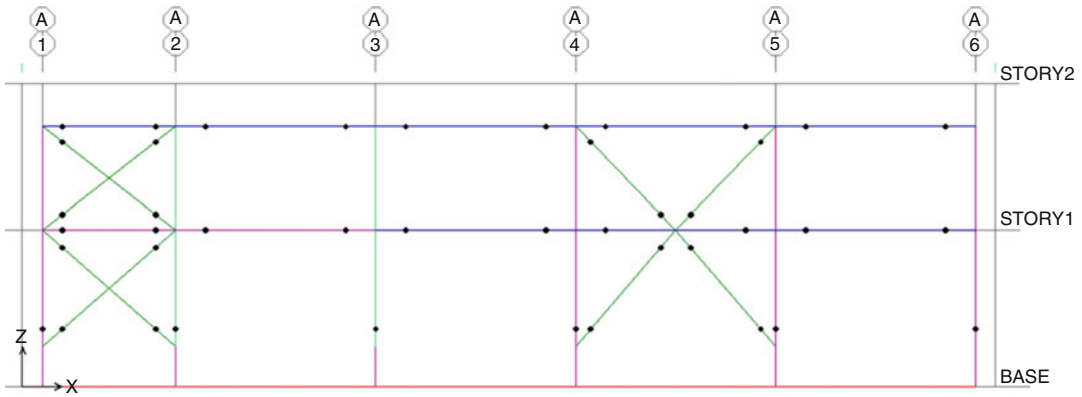
Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 17 Column to girder connections of prototype building, modeled as rigid moment connections



Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 18 Connections between or at the end of bracing members, modeled as hinges



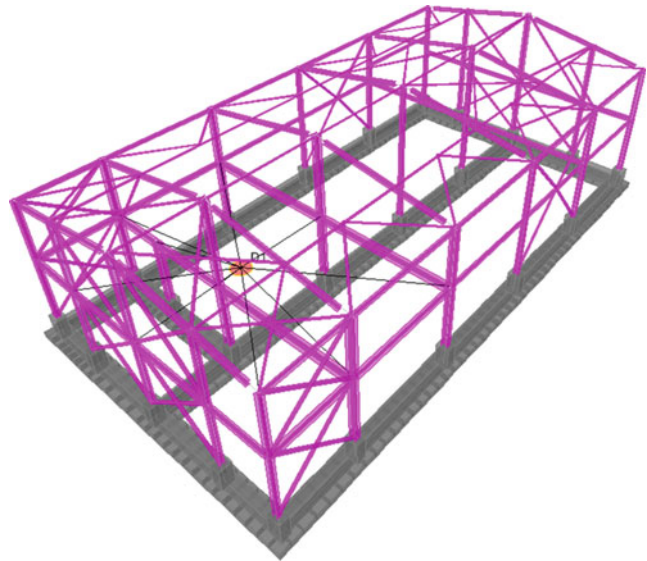
Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 19 Roof's plan view from numerical model illustrating with dots the joint releases on roof's bracing and on secondary beams



Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 20 Building's side view from numerical model illustrating with dots the joint releases on column

bases, on vertical bracing, and on secondary beams (reinforced concrete pedestals on which steel columns are supported are also shown)

Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 21 Diaphragm definition of loft slab



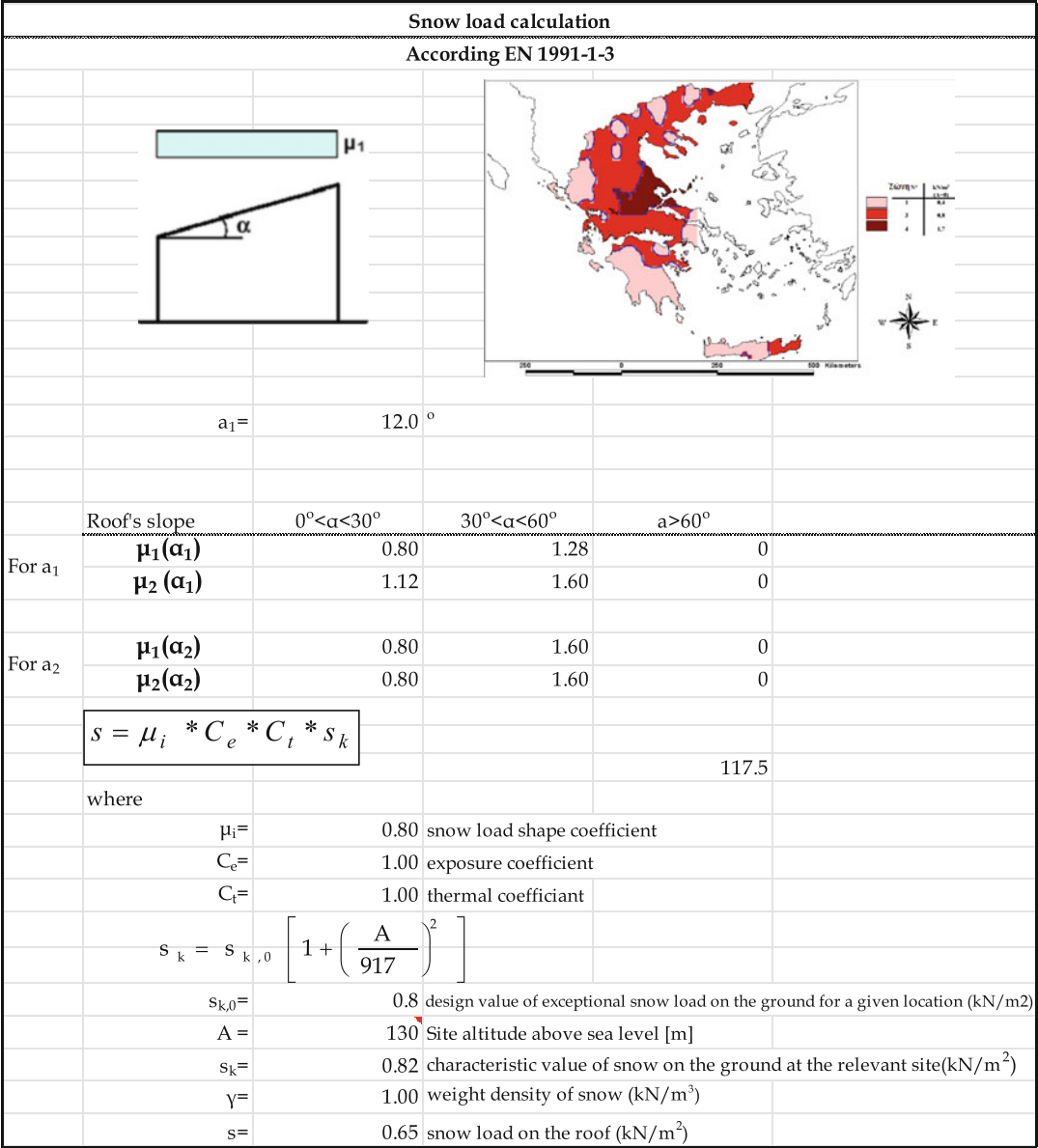
nominal location in each direction by an accidental eccentricity, commonly taken equal to $e = 0.05 \cdot L$, where L is the corresponding span length.

Loads

The next phase of the design process consists of setting up appropriate loading situations including basic loads and load combinations. Basic loads commonly include permanent loads, live loads depending on the building's intended use,

and environmental loads such as those due to snow, wind, and temperature, as well as seismic actions in case the building is located in a seismic region. Characteristic nominal values of live, snow, and wind loads and seismic actions are prescribed by pertinent codes (e.g., European Committee for Standardisation 2001b, 2003a, b, d, 2004a).

For the prototype building presented here, Excel spreadsheets for the calculation of snow, wind, and seismic loads are illustrated in Figs. 22, 23, 24, and 25. Issues that may require special attention include snow accumulation due

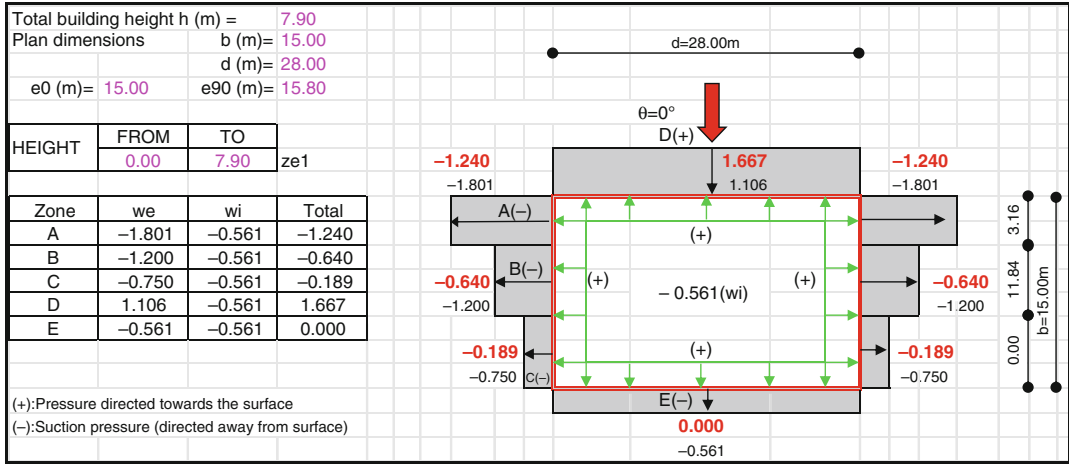


Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 22 Excel spreadsheet for snow load calculation for the prototype building according to Eurocode 1 – Part 1–3

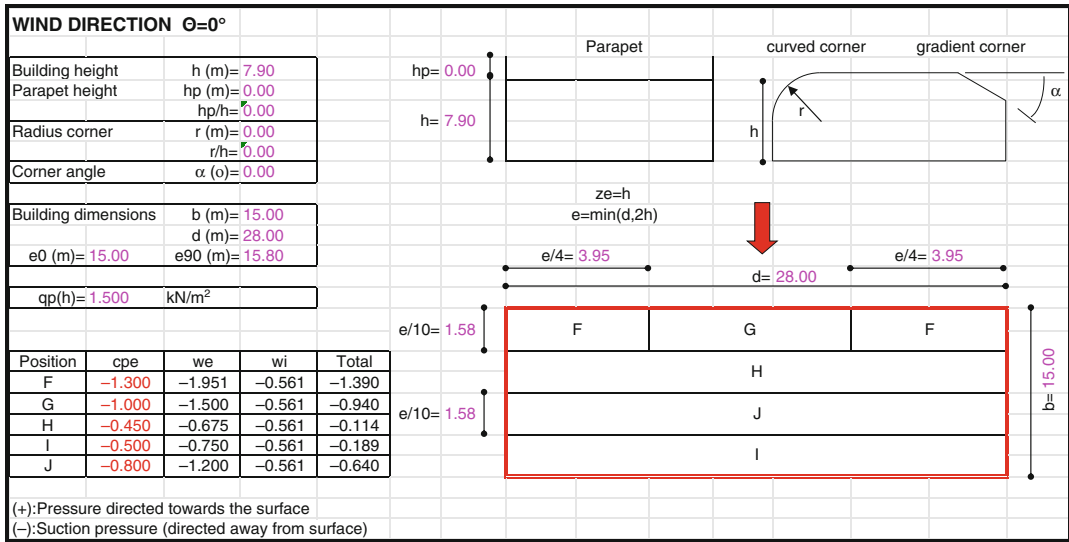
to roof geometry, increased snow density in regions of low temperatures, or unusual building shapes not covered by the code guidelines for wind loads. Choices that differentiate the design spectrum of steel structures from those of structures made of different materials are the values of damping, commonly taken equal to 2 % for welded and 4 % for bolted steel structures, and

the values of behavior factor, for which recommendations are provided in the pertinent seismic codes, depending also on the type of structural system.

In case part of the structure is buried (e.g., a basement), it is recommended to define accordingly the level above which the structure vibrates independently of the ground.



Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 23 Excel spreadsheet for wind load calculation on the side walls of the prototype building according to Eurocode 1 – Part 1–4



Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 24 Excel spreadsheet for wind load calculation on the roof of the prototype building according to Eurocode 1 – Part 1–4

Following load calculation, basic loads are then applied on the model, as shown schematically in Figs. 26, 27, 28, and 29.

The next step is to describe pertinent load combinations in accordance with the relevant codes (e.g., European Committee for Standardisation 2001a), including combinations in the ultimate (ULS), serviceability (SLS), and seismic limit states. Common load combinations

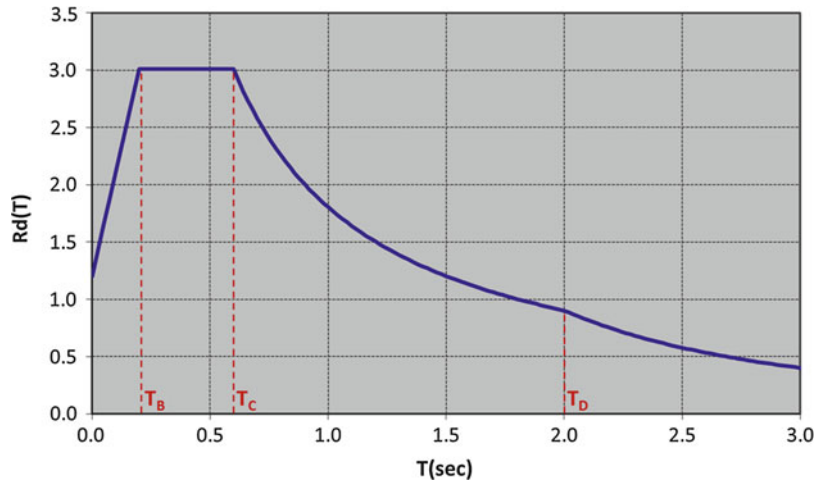
for ordinary steel buildings are provided in Table 1.

In Table 1:

- Live load corresponds to all relevant types of live load, such as service load or load due to snow.
- Wind load corresponds to four basic individual load cases for each direction, including

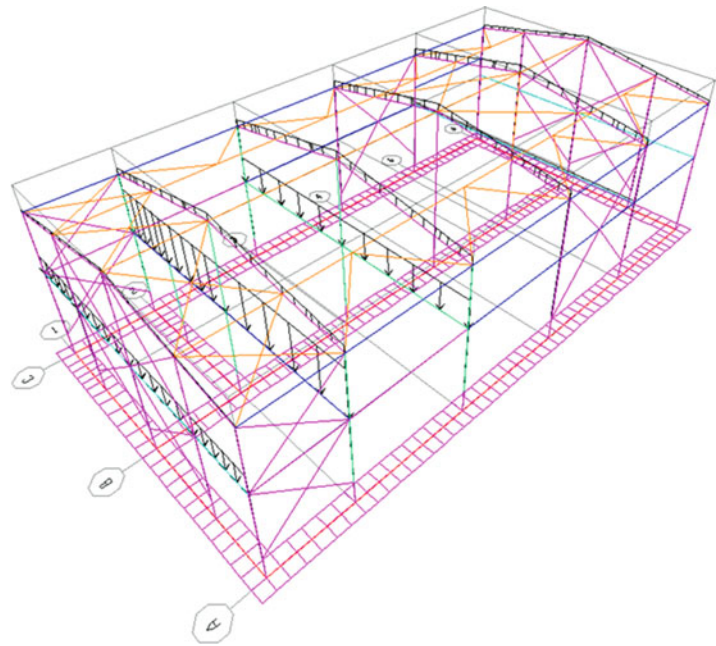
Seismic Analysis of Steel Buildings: Numerical Modeling,

Fig. 25 Horizontal component of elastic response spectrum of the prototype building according to Eurocode 8



Seismic Analysis of Steel Buildings: Numerical Modeling,

Fig. 26 Application of dead loads on the finite element model of the prototype building



S

uplift or downward pressure on the roof and positive or negative pressures on the walls.

- Earthquake load corresponds to three basic individual load cases, one for each global direction.
- Temperature load corresponds to two basic individual loads (increase, decrease).

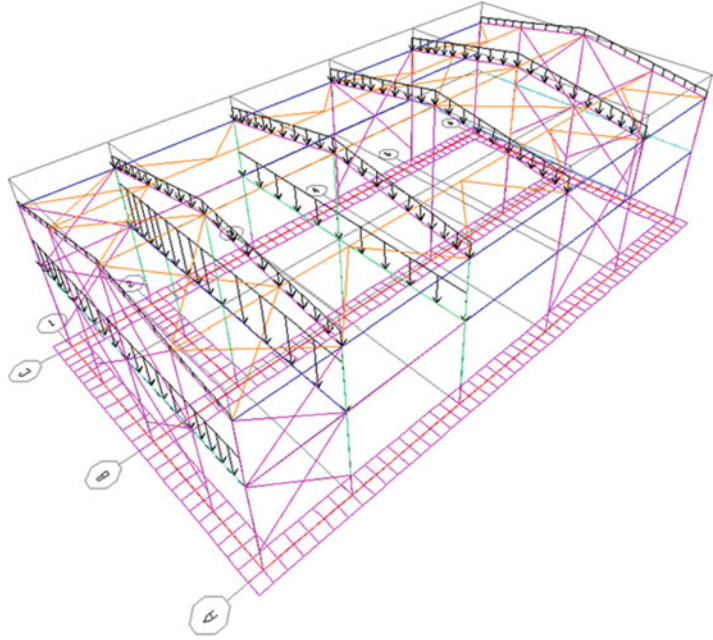
Structural Analysis

A decision that has to be taken next concerns the type of structural analysis to be performed.

A variety of linear and nonlinear (geometrically and/or material) and static or dynamic analysis algorithms are commonly available in commercial structural software (e.g., Gantes and Fragkopoulos 2010). However, in the vast majority of common steel buildings, linear static analysis is used in design practice for all loads and load combinations, with the exception of seismic actions, for which frequently both equivalent static as well as response spectrum dynamic analysis are carried out. The principle of

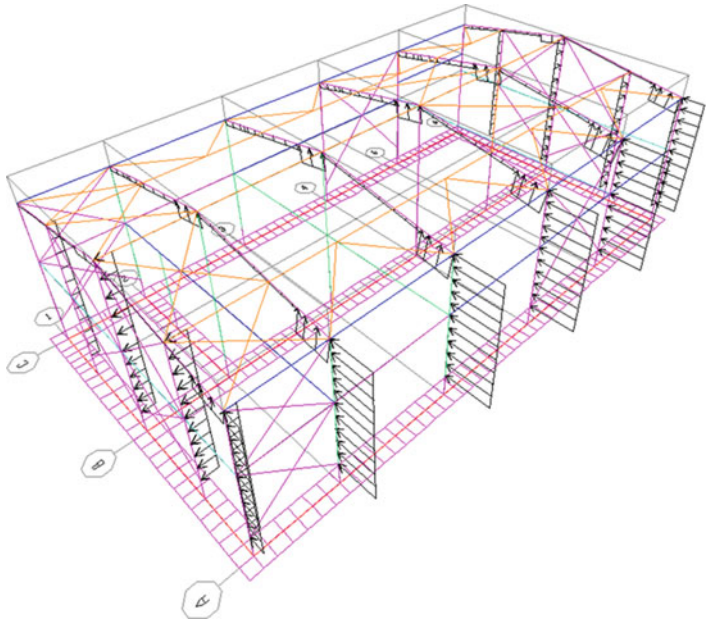
Seismic Analysis of Steel Buildings: Numerical Modeling,

Fig. 27 Application of live loads on the finite element model of the prototype building



Seismic Analysis of Steel Buildings: Numerical Modeling,

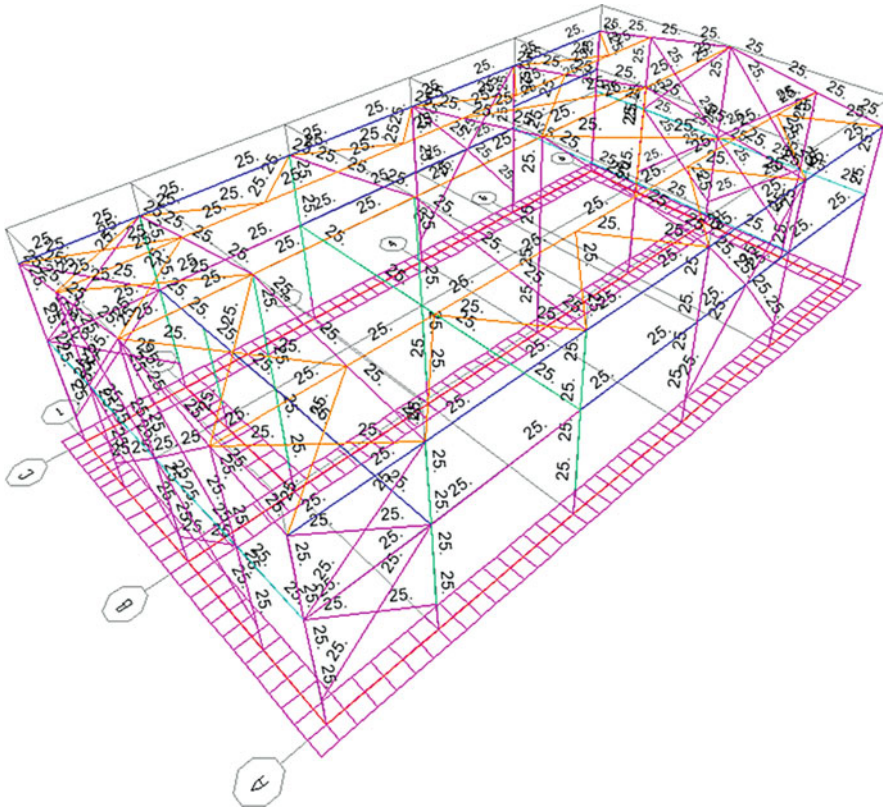
Fig. 28 Application of wind loads on the finite element model of the prototype building



superposition is routinely employed to obtain action effects for load combinations.

In the equivalent static method, first the base shear is calculated, which is then distributed over the height of the building as specified by the pertinent seismic code. In the response spectrum analysis, for a given direction of acceleration, the

maximum displacements, forces, and stresses are computed throughout the structure for each vibration mode. These modal values for a given response quantity (displacements, forces, or stresses) are then combined appropriately to produce a single, positive result for the given direction of acceleration, using, for example, the SRSS



Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 29 Application of temperature increase on the finite element model of the prototype building

or the CQC method. It is noted that most modern codes contain a requirement of modal participating mass ratios at least equal to 90 %. In case this is not satisfied, the response spectrum analysis must be repeated with higher number of participating modes.

The analysis results include mode shapes and natural periods of vibration as well as deformed shapes and internal force diagrams for all individual loads and load combinations. A qualitative evaluation of analysis results is always highly recommended in order to detect possible modeling errors. For that purpose, it is a good practice to start with mode shapes and natural periods of vibration, proceed with deformed shapes for individual load cases, continue with internal force diagrams for individual load cases, and finally, conclude with internal force diagrams for load combinations and envelopes, which are also

used for design. Indicative results for the prototype building are presented in the following figures, including mode shapes (Figs. 30 and 31), seismic deformed shapes (Figs. 32 and 33), and internal force diagrams for individual load cases (Fig. 34) and load combination envelopes (Figs. 35, 36, 37, and 38).

Structural Design Checks

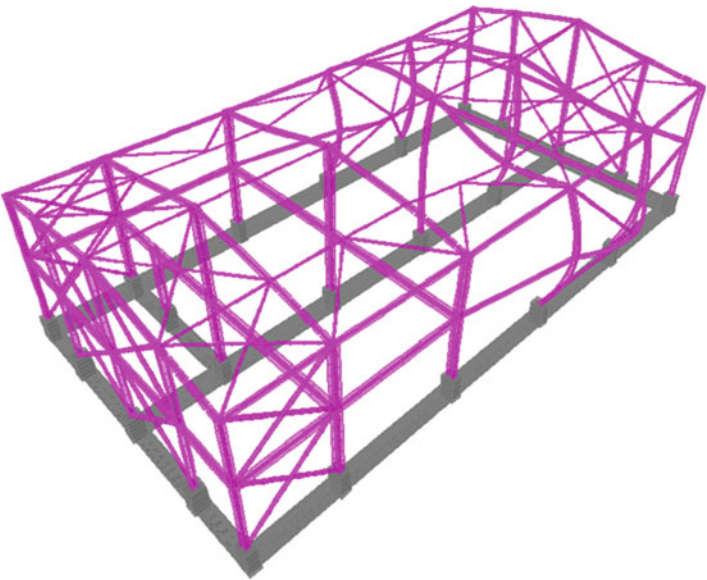
Following the qualitative evaluation of analysis results on the basis of mode shapes and corresponding vibration periods, deformed shapes and internal force diagrams, and provided that confidence is gained for the correctness of these results, design checks are performed, consisting of general checks, member checks, and connection checks. Foundation checks are

Seismic Analysis of Steel Buildings: Numerical Modeling, Table 1 Common load combinations for ordinary steel buildings in the ultimate (ULS), serviceability (SLS), and seismic limit states according to the Eurocodes

	Load combination	Load factors $\gamma_{unfavourable}; \gamma_{favourable}$				
		Dead	Live	Wind	Temperature	Earthquake
ULS	COMB1	1.35	1.50			
	COMB2	1.35	1.50	0.90		
	COMB3	1.35	1.50		0.90	
	COMB4	1.35	1.50	0.90	0.90	
	COMB5	1.35;1.00		1.50		
	COMB6	1.35;1.00	0.90	1.50		
	COMB7	1.35;1.00		1.50	0.90	
	COMB8	1.35;1.00	0.90	1.50	0.90	
	COMB9	1.35			1.50	
	COMB10	1.35	0.90		1.50	
	COMB11	1.35		0.90	1.50	
	COMB12	1.35	0.90	0.90	1.50	
SLS	COMB13	1.00	1.00			
	COMB14	1.00	1.00	0.60		
	COMB15	1.00	1.00		0.60	
	COMB16	1.00	1.00	0.60	0.60	
	COMB17	1.00		1.00		
	COMB18	1.00	0.60	1.00		
	COMB19	1.00		1.00	0.60	
	COMB20	1.00	0.60	1.00	0.60	
	COMB21	1.00			1.00	
	COMB22	1.00	0.60		1.00	
	COMB23	1.00		0.60	1.00	
	COMB24	1.00	0.60	0.60	1.00	
SEISMIC	COMB25	1.00	0.30			1.00

Seismic Analysis of Steel Buildings: Numerical Modeling,

Fig. 30 Dominant vibration mode in transverse direction of prototype building



Seismic Analysis of Steel Buildings: Numerical Modeling,

Fig. 31 Dominant vibration mode in longitudinal direction of prototype building

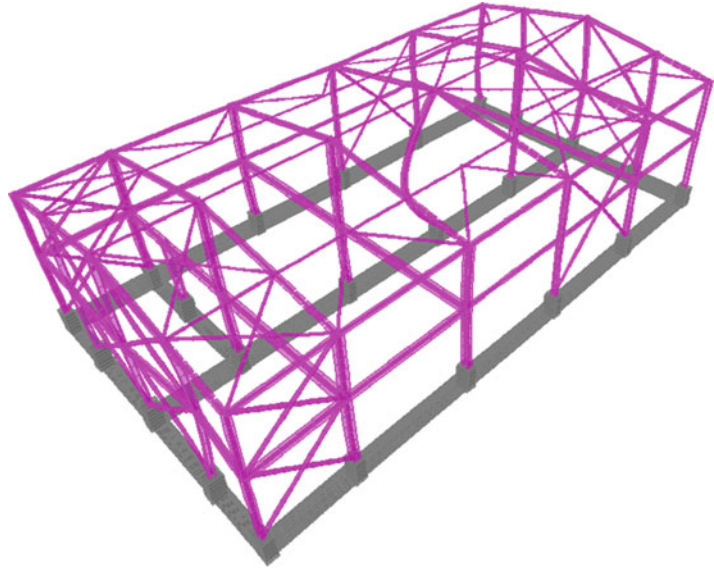
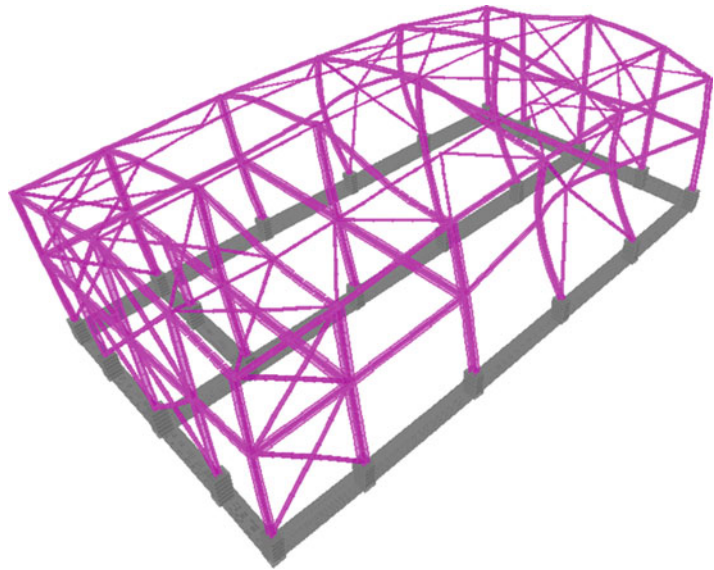
**Seismic Analysis of Steel Buildings: Numerical Modeling,**

Fig. 32 Deformed shape of prototype building from equivalent static method – transverse direction



S

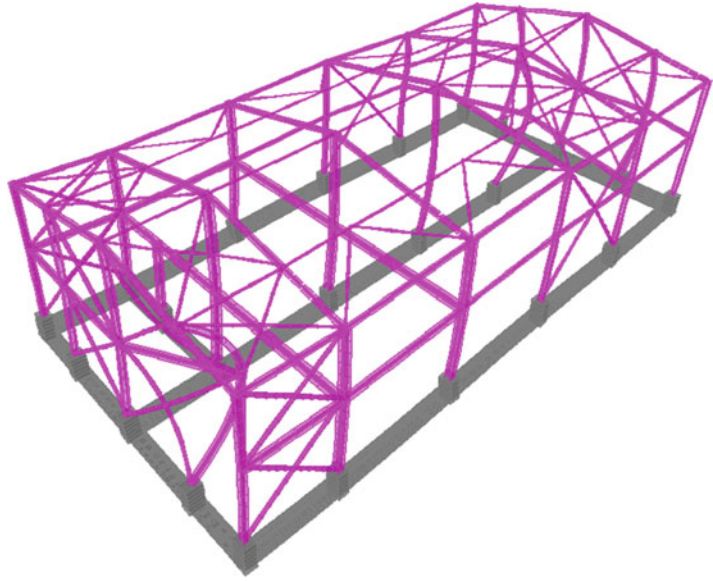
also part of this process, which are presented in other parts of this encyclopedia. This is an iterative process, and frequently, there is a need to return to the model, modify it, and run a new set of analyses and checks, as schematically illustrated in Fig. 39.

General checks (e.g., European Committee for Standardisation 2004b) consist mainly of confirming that overall structure deformations

are acceptable for all load combinations. This includes interstory drifts and overall building drift for all load combinations with predominantly horizontal components, such as wind and seismic combinations. Depending on the use of the building, general checks may also include restrictions about the vertical vibration frequencies, associated with a sensation of unease of the users, as is the case in stadium grandstands. In case such checks

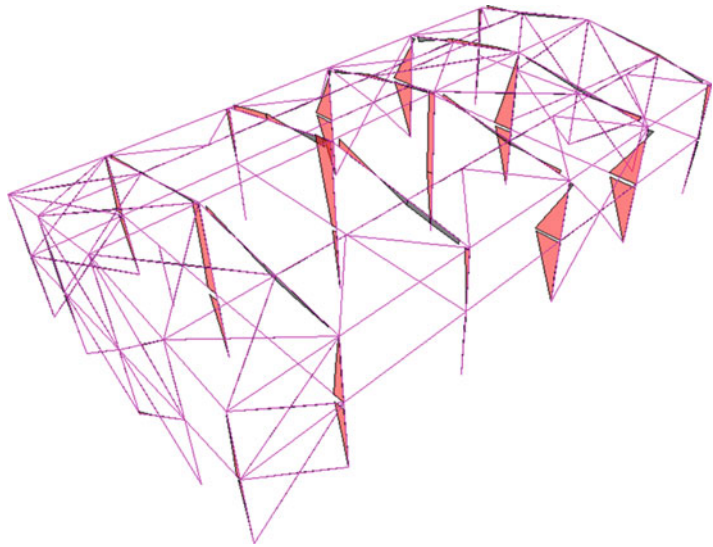
Seismic Analysis of Steel Buildings: Numerical Modeling,

Fig. 33 Deformed shape of prototype building from equivalent static method – longitudinal direction



Seismic Analysis of Steel Buildings: Numerical Modeling,

Fig. 34 Moment M_3 diagrams of prototype building from equivalent static method – seismic action in transverse direction



are violated by far, a modification of the structural system may be the only solution, while smaller violations may be addressed by increasing member cross sections.

Member checks (e.g., European Committee for Standardisation [2004b](#)) consist of comparison between actions and resistances in the ultimate limit state and comparison of

maximum deflections to allowable upper bounds in the serviceability limit state. As flexural and lateral-torsional buckling are in most cases critical for steel members, it is highly recommended that the engineer reviews and modifies as needed the buckling lengths initially proposed by the software for the corresponding checks. It is also

Seismic Analysis of Steel Buildings: Numerical Modeling,

Fig. 35 Envelope of moment M_3 diagrams of prototype building for ULS combinations

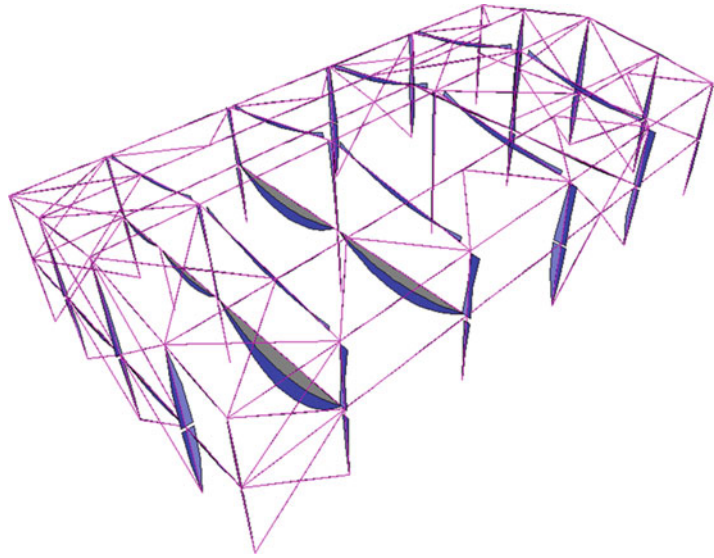
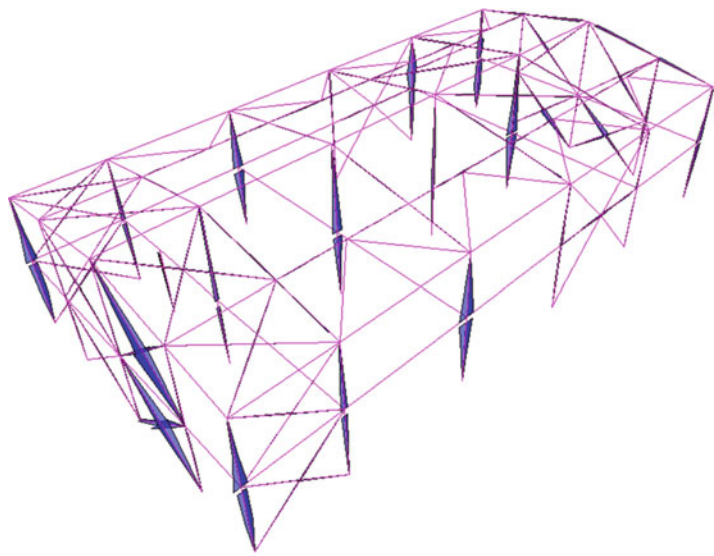
**Seismic Analysis of Steel Buildings: Numerical Modeling,**

Fig. 36 Envelope of moment M_2 diagrams of prototype building for ULS combinations



S

noted that most seismic design codes require capacity design checks, leading, for example, to the necessity of larger bracing sections to satisfy minimum slenderness requirements or larger column sections, to adhere to the weak beam-strong column design approach.

Connection checks (e.g., European Committee for Standardisation 2003c) include strength checks, to ensure safe transfer of internal actions between members, as well as stiffness checks, so that the behavior of actual connections (Fig. 40) is in accordance with the hinged/semirigid/rigid assumptions adopted in

Seismic Analysis of Steel Buildings: Numerical Modeling,

Fig. 37 Envelope of shear V_2 diagrams of prototype building for ULS combinations

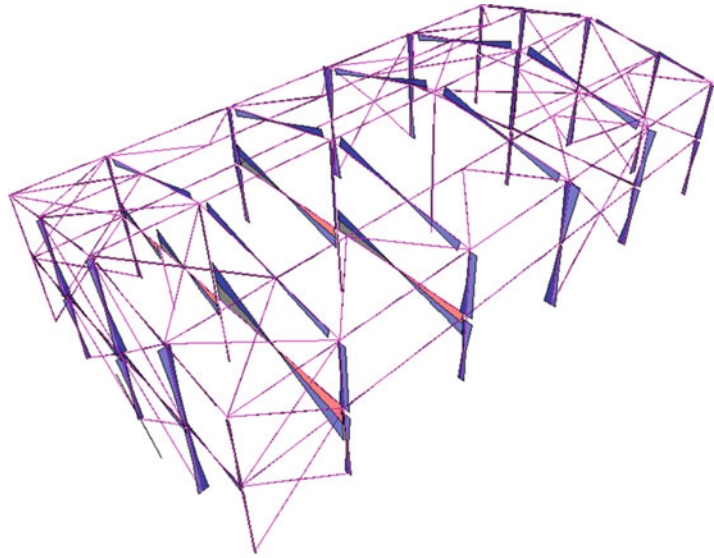
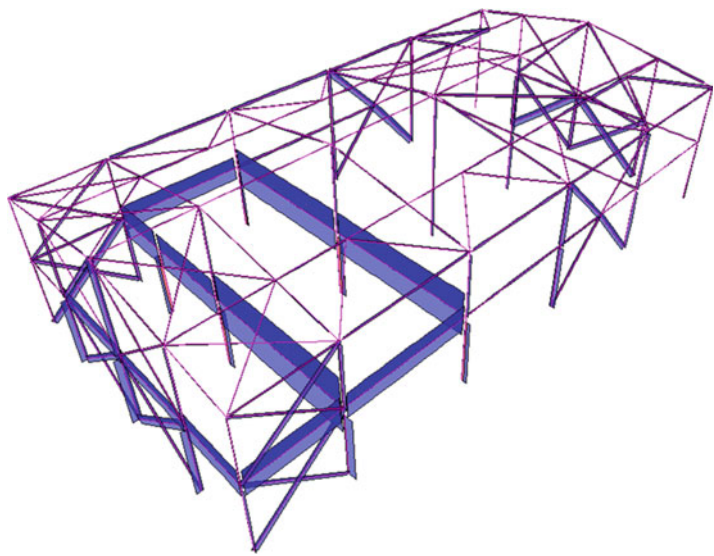
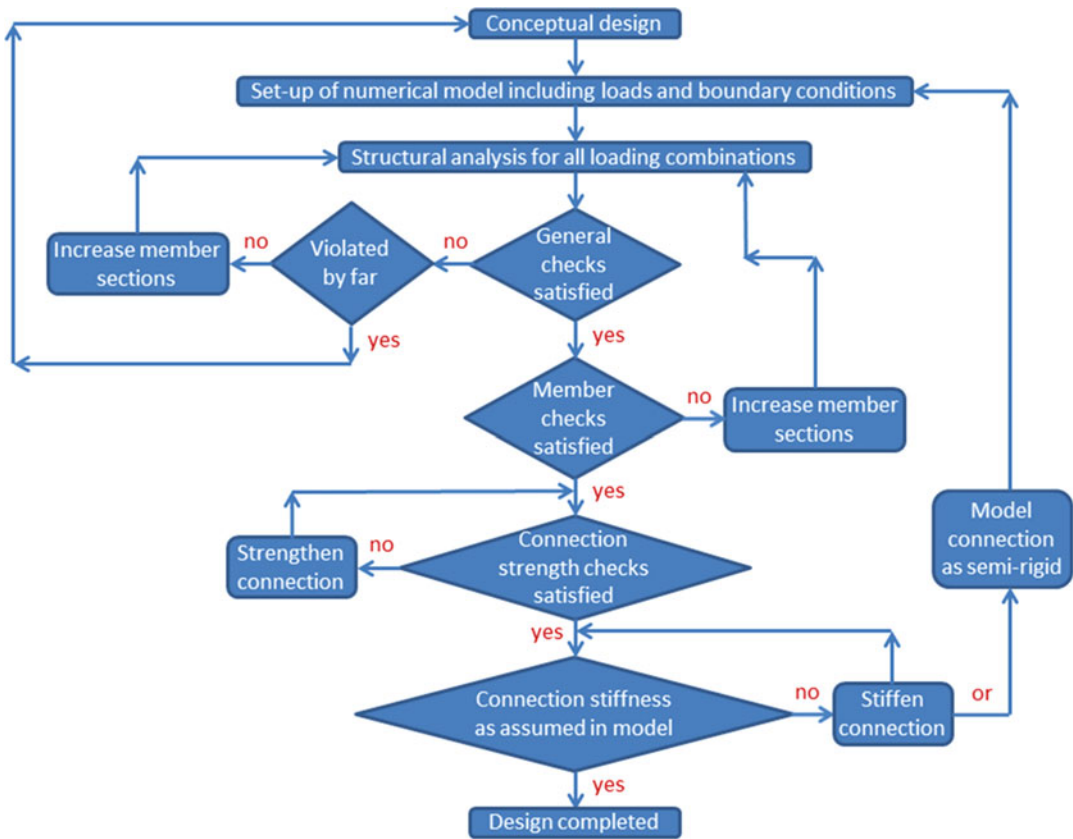
**Seismic Analysis of Steel Buildings: Numerical Modeling,**

Fig. 38 Envelope of axial force N diagrams of prototype building for ULS combinations

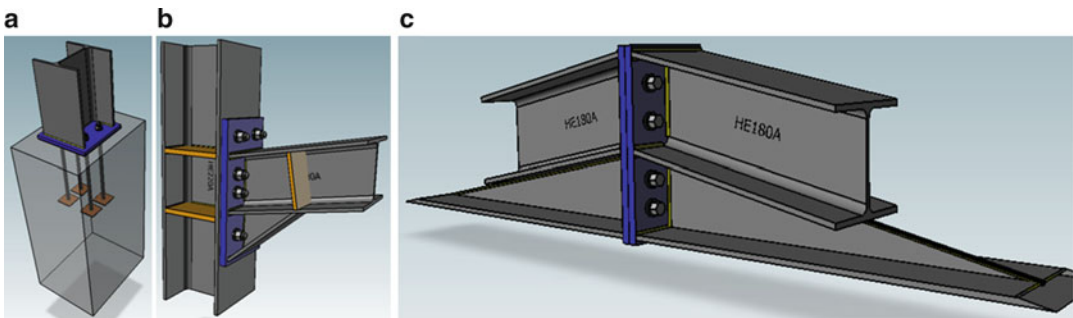


the numerical simulation. As there is so far limited experience with the behavior of semi-rigid connections under cyclic loading, it is recommended to avoid their use in seismic regions. Instead, it is proposed to stiffen the connection so that it can be classified as rigid. Strength calculation and stiffness classification of steel connections are commonly performed by means of dedicated software.

Foundation checks depend on the type of foundation (mat foundation, foundation beams, spread footings, pile foundation, etc.) and consist of general stability checks (overturning, sliding), soil bearing capacity checks, comparison of absolute and differential settlements to allowable values, and calculation of reinforcement for the reinforced concrete elements.



Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 39 Flowchart of structural design process for steel buildings



Seismic Analysis of Steel Buildings: Numerical Modeling, Fig. 40 (a) Column base, (b) girder to column, (c) frame ridge connections of prototype building

Summary

The state of the art in numerical modeling of steel buildings has been presented, from the point of view of the practicing structural engineer designing such structures in seismic regions.

Conceptual design, numerical modeling, structural analysis, and design checks have been discussed, but emphasis has been directed toward modeling, as the other design phases are covered in detail in other parts of this encyclopedia.

Cross-References

- [Assessment of Existing Structures Using Response History Analysis](#)
- [Behavior Factor and Ductility](#)
- [Buckling-Restrained Braces and Their Implementation in Structural Design of Steel Buildings](#)
- [Earthquake Response Spectra and Design Spectra](#)
- [Earthquake Return Period and Its Incorporation into Seismic Actions](#)
- [Equivalent Static Analysis of Structures Subjected to Seismic Actions](#)
- [European Structural Design Codes: Seismic Actions](#)
- [Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis](#)
- [Modal Analysis](#)
- [Plastic Hinge and Plastic Zone Seismic Analysis of Frames](#)
- [Response Spectrum Analysis of Structures Subjected to Seismic Actions](#)
- [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)
- [Seismic Analysis of Steel and Composite Bridges: Numerical Modeling](#)
- [Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling](#)
- [Soil-Structure Interaction](#)
- [Steel Structures](#)
- [Strengthening Techniques: Code-Deficient Steel Buildings](#)
- [Structural Design Codes of Australia and New Zealand: Seismic Actions](#)
- [Time History Seismic Analysis](#)

References

- European Committee for Standardisation (2001a) Eurocode – basis of structural design
- European Committee for Standardisation (2001b) Eurocode 1: actions on structures – part 1–1: general actions – densities, self-weight, imposed loads for buildings
- European Committee for Standardisation (2003a) Eurocode 1: actions on structures – part 1–3: general actions – snow loads

- European Committee for Standardisation (2003b) Eurocode 1: actions on structures – part 1–5: general actions – thermal actions
- European Committee for Standardisation (2003c) Eurocode 3: design of steel structures, part 1–8: design of joints
- European Committee for Standardisation (2003d) Eurocode 8: design of structures for earthquake resistance – part 1: general rules, seismic actions and rules for buildings
- European Committee for Standardisation (2004a) Eurocode 1: actions on structures – part 1–4: wind actions
- European Committee for Standardisation (2004b) Eurocode 3: design of steel structures, part 1–1: general rules and rules for buildings
- Gantes CJ, Fragkopoulos KA (2010) Strategy for numerical verification of steel structures at the ultimate limit state. *Struct Infrastruct Eng* 6(1–2):225–255

Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling

Konstantinos Daniel Tsavdaridis

School of Civil Engineering, University of Leeds, Leeds, UK

Synonyms

Concrete-encased steel sections; Connections; Constitutive models; Cyclic behavior; Damping; Frames; Hysteretic rules; Nonlinear; Numerical modeling; Panel zones; Seismic analysis; Steel–concrete composite; T-stub components

Introduction

Steel–Concrete Composite (SCC) Systems

Composite construction includes a wide range of structural systems, e.g., framed structures employing all steel–concrete composite (SCC) members and components (e.g., composite beam-to-columns and connections) and sub-assemblages of steel and/or reinforced concrete (RC) elements. Such components and elements are employed to optimize the resistance and deformation capacity (Uchida and Tohki 1997). SCC structures have been used

extensively in recent years because of benefits in combining the two construction materials. SCC structures are also known for their excellent earthquake performance owing to their high strength, high ductility, and large energy absorption. Their good structural damping properties arising from the friction between the steel–concrete interfaces make them an even more attractive alternative for seismic resistance.

Consequent effects of combining the two materials are the enhanced lateral strength and stiffness of the frame, with apparent effects of the alteration of the structural natural period of vibration and the complex local behavior of beam-to-column connections. Furthermore, SCC beams subjected to lateral loading show complex behavior due to several factors, including the slip between the concrete slab and the steel beam, the variation of longitudinal stress across the width of the slab, and the overall configuration of the numerous different types of models, while the steel and concrete parts can be subjected to different actions in every case. For the above reasons, the calculation of the seismic response of composite structures is not a straightforward task due to the interaction of local and global effects and hence the unexpected failure modes that might incur. Consequently, it is very important for the analysis of such structures to account for the local interactions (e.g., interface behavior between steel and concrete) as well as the local behavior of structural systems (e.g., beam-to-column and base-to-column response). All these factors make the analysis of SCC structures and their individual components an intriguing but challenging task.

Although experimental procedures can be performed in order to enhance the understanding of the behavior of SCC structures under earthquake loading, they are typically expensive and time-consuming and do not cover a broad range of SCC structures and elements. As a result, numerical modeling procedures have been developed and tested in order to facilitate the analysis of such structures.

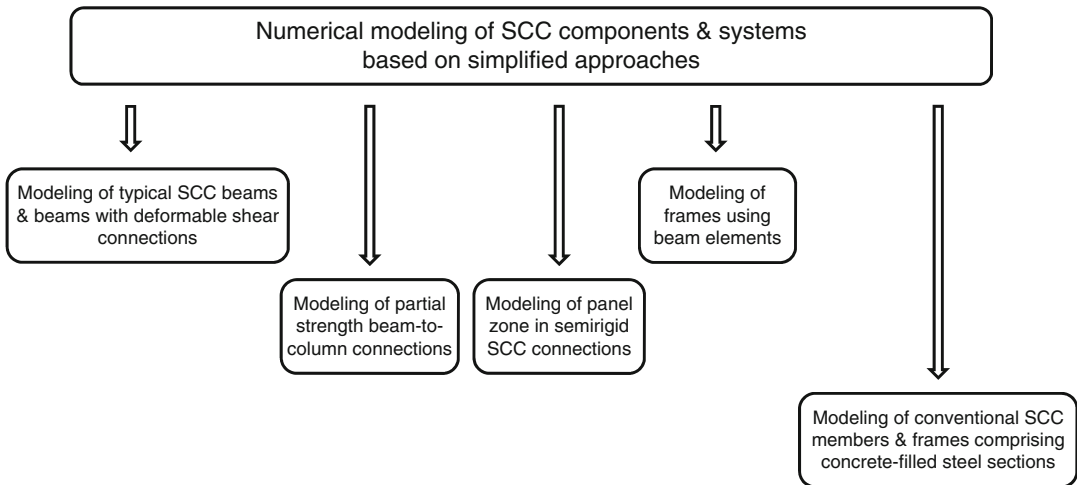
Most finite element (FE) packages (e.g., ANSYS, ABAQUS, ADINA, DIANA, LS-DYNA, MIDAS FEA, etc.) rely on the use of

constitutive models which emphasize on the description of post-peak material characteristics such as strain hardening and softening, tension stiffening, shear-retention ability, etc. (Cotsovos and Kotsovos 2011). The derivation of such constitutive models has been based on a variety of theories and their combination. However, the application of FE packages in practical structural analysis has shown that the constitutive relationships are case-study dependent, since the solutions obtained are realistic only for particular problems. Therefore, the applicability of packages to a different set of problems requires modifications of the constitutive relationships. This is entirely dependent on the interpretation of the observed material behavior as well as the use of the experimental data to validate the constitutive relationships.

To this end, the aim of the present chapter is to provide an indication of the concepts which are widely used for modeling the steel–concrete composite behavior and to develop numerical guidelines for the nonlinear analysis of such structures and their components, considering the seismic actions under earthquake events. The numerical analyses presented herein model the behavior of SCC structures/components using macro-models (i.e., the use of line elements and spring connections) rather than micro-models (continuum FE models) due to their simplicity and accuracy in nonlinear analysis. Different aspects of modeling including the geometry, material nonlinearity (through the constitutive laws adopted), hysteretic behavior, and geometrical nonlinearity as well as other parameters important for seismic analysis are also presented in this chapter.

Chapter Synthesis

The modeling of SCC elements and frames is based on three approaches. The first one is the simplified modeling approach presented in this chapter, engaging the use of springs and line elements for the elementary simulation of the behavior of each component to the entire frame assembly. The scope is to initiate numerical guidelines based on the simplified approach and to present modeling examples of SCC beam cross sections, flooring systems, fully composite



Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Chart 1 Structure of the Chapter

members, beam-to-column connections, as well as holistic approaches modeling a frame. The breakdown of this chapter is given in Chart 1.

Based on the bottom-up approach (i.e., combining different structural components starting from the most fundamental of a system and giving rise to grander systems), the engineer will gain a decent understanding on the parameters to be considered during the computational modeling procedure. Modeling of these structural components will enable the computation of their response to different load histories and moreover will enable the engineer to carry out the state determination of a member from a frame assembly.

Requirements for Collapse Analysis of Composite Structures

Numerical modeling procedures should aim to address a number of issues regarding the local, intermediate, and global level of SCC structural design. On the *local* level, aspects such as the cyclic behavior of the steel and the concrete members (including the softening and hardening of the material), the local buckling of steel flanges, the load carrying capacity, the curvature ductility of the components, as well as the effects of

confinement should be carefully studied. On the *intermediate* level, the ductility of the member in terms of rotation/displacement should be established. Additionally, second-order effects ($P-\Delta$) on forces and deformations should be taken into account through the provisions for large displacement analysis. Modeling the beam-to-column connection is also essential when the fully rigid assumption is not suitable. On the *global* level, the overall ductility and strength of the structure should be established through force–displacement relations. The progressive yielding and the hinge formation at the structural frame should also be established through moment–rotation relationships. The complete list of requirements for the collapse analysis and the modeling of SCC structures subjected to earthquake actions is presented below:

- Stress–strain relationships for the steel material including strain hardening and softening
- Provision for the effects of local buckling in the steel section
- Stress–strain relationships for the concrete material including cycling loading regimes and the effect of the confinement on the peak stress and corresponding strain

- Explicit representation of the slip boundary conditions of the shear connection both at local and global levels
- Provision for second-order effects on forces and deformations
- Effective beam-to-column connection models, including panel distortion
- Iterative and advanced dynamic analysis techniques for analyzing the structural response near collapse state

Modeling of Steel–Concrete Composite (SCC) Beams

A variety of different models have been developed by researchers in order to capture the behavior of SCC beams, based on either concentrated or distributed plasticity. In concentrated plasticity models, all the inelasticity is concentrated at the ends of the member; therefore, it deals with material nonlinearity in an approximate but efficient manner. On the contrary, distributed plasticity models simulate the inelastic behavior along the length of the member. This approach is more accurate but at the same time is more computationally demanding. Most of the formulations for both approaches are rather complex and not amenable to generic and routine application in structural engineering design.

The present subchapter presents a simplified (new) modeling approach based on the work of Zhao et al. (2012) for the nonlinear analysis of SCC beams and composite frames with deformable shear connections (based on the distributed plasticity approach) using line elements to simulate the structural beam and column members, layered fiber section to simulate the reinforced concrete slabs, and nonlinear spring elements for the simulation of the interface between the structural steel beams and the reinforced concrete slab. Vertical interactions between the slab and steel beams are not expected to be significant, therefore are not accounted into the analysis. The geometry of the model, along with a simple set of details, is outlined below. The assembled model is shown in Fig. 1.

Model Geometry of a Typical SCC Beam

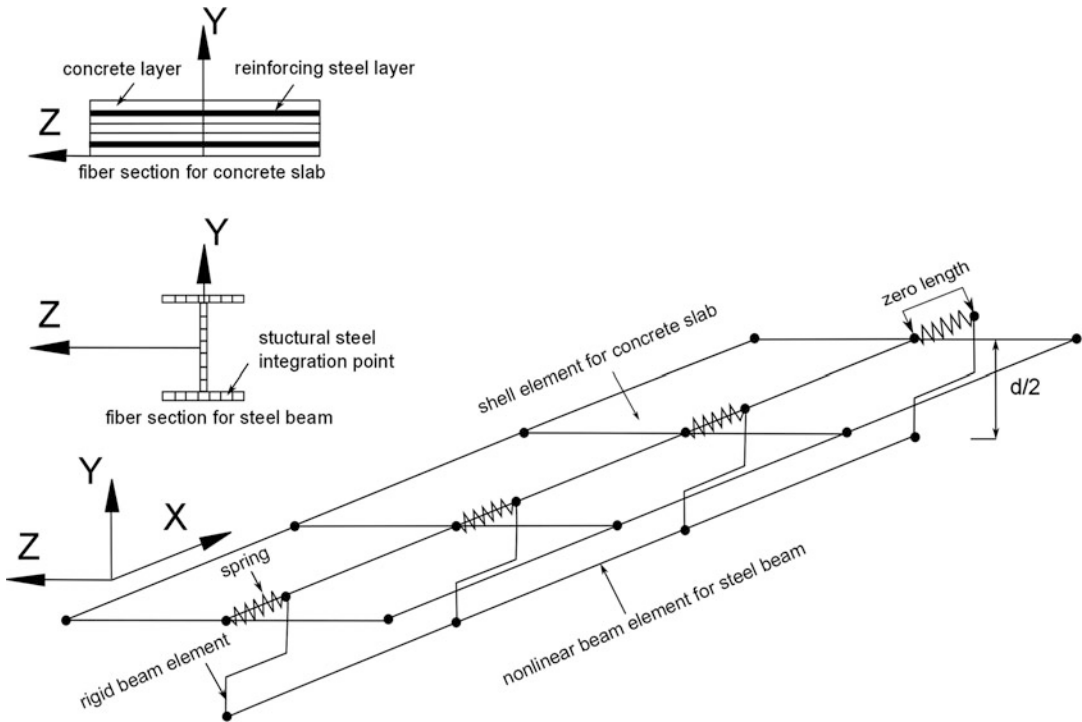
To model the geometry of the macromodel for a typical SCC beam, the following assemblies should be utilized:

1. Four-node-layered shell elements representing the concrete slab. Reinforcement layers comprising steel material properties should be used to simulate the steel reinforcement located at the top and bottom of the concrete slab.
2. Two-node fiber beam-to-column elements for modeling the steel beam. The reference surface of the slab will be located at the centroid of the steel beam cross section.
3. Dummy nodes at the same locations as the beam-to-column element nodes simulating the connection between the nodes of the steel beam and the shell elements.
4. Rigid beam elements connecting the dummy nodes and the corresponding ones of the shell elements, located on the same x- and z-coordinates.
5. Discrete spring elements with only translation in the z-direction connecting both the dummy nodes and the beam-to-column element nodes in order to control the interface shear-slip surface along the length of the beam.

Model Geometry of Beam with Deformable Shear Connection

Modeling of two-dimensional beams with deformable shear connection is based on the Newmark et al. (1951) model, in which (i) the Euler–Bernoulli beam theory applies to both components of the SCC beam and (ii) the deformable shear connection is represented by an interface model with distributed bond allowing interlayer slip as well as enforcing contact between the steel and concrete components.

A local coordinate system should be established to enhance the understanding of kinematics of Newmark’s model. With reference to Fig. 2, Z axis is parallel to the beam axis and the vertical plane YZ is the plane of geometrical and



Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Fig. 1 Assembled macromodel representing a typical steel-concrete composite beam

material symmetry of the cross section. Loads are also assumed symmetric with respect to the YZ plane. The displacement field u of a material point of the beam is given by:

$$u(y, z) = v(z)j + \left[w_a + (y_a - y)v'(z) \right]k \quad \text{on } A_a (a = 1, 2) \quad (1)$$

where w_a is the axial displacement of the reference point of domain A_a , the ordinate of which is y_a ($a = 1$: concrete slab, $a = 2$: steel beam); v is the vertical displacement of the cross section; and j and k denote the unit vectors along the Y and Z axes, respectively. It is observed that the transverse displacements and rotations of the slab and of the steel beam are equal due to the enforced contact between the two components. The only nonzero strain components are the axial strain ϵ_{za} and the interface slip δ :

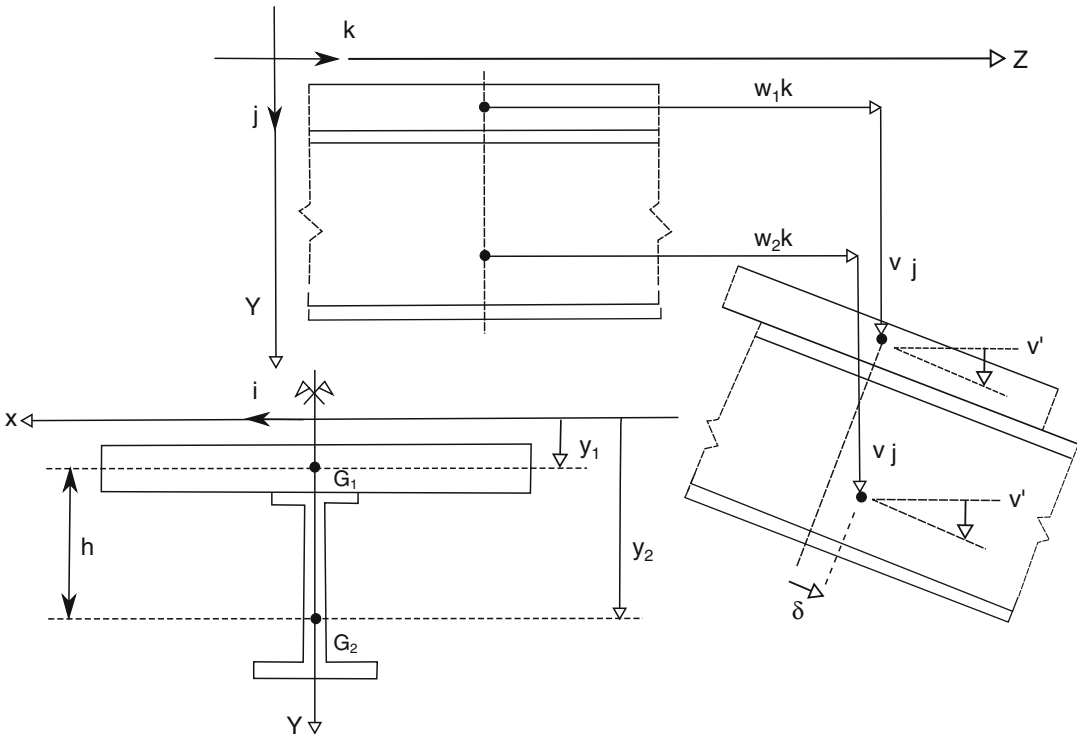
$$\epsilon_{za}(y, z) = w'_a(z) + (y_a - y)v''(z) \quad \text{on } A_a (a = 1, 2) \quad (2)$$

$$\delta(z) = w_2(z) - w_1(z) + hv'(z) \quad (3)$$

where $h = y_2 - y_1$ is the distance between the reference points (G_1 and G_2 in Fig. 2) of the two components. At the locations of the longitudinal reinforcement, Eq. 2 also provides the strain in the reinforcement, due to the assumption of perfect bond between the steel and the concrete.

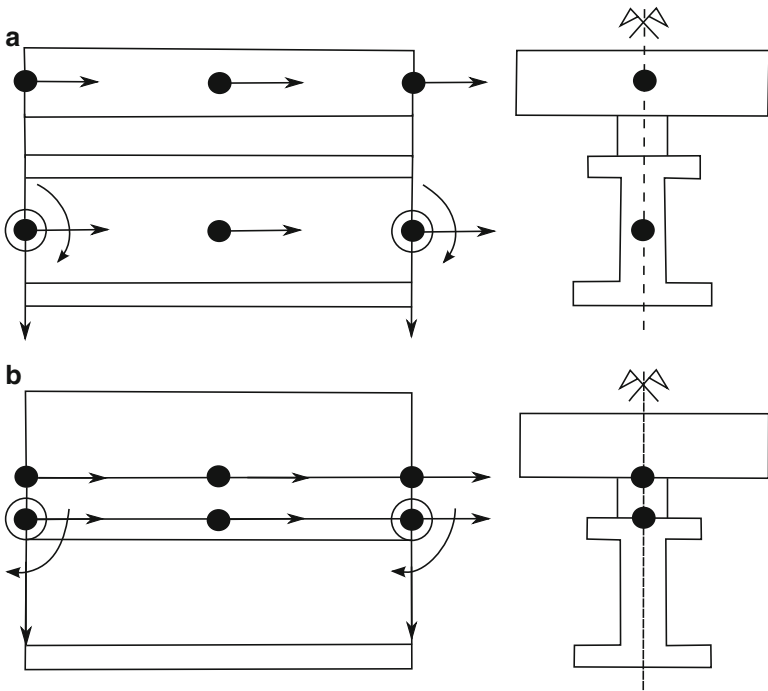
FE Formulations

A simple two-dimensional 10 degree-of-freedom (DOF) SCC frame element with deformable shear connection is presented herein, similarly to Zona et al. (2008). With reference to Fig. 3, 8 of the 10DOFs are external (4 DOFs per end node) allowing for the axial displacement, the transverse displacement, and the rotation of the steel beam and 1DOF for the axial displacement



Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Fig. 2 Kinematics of two-dimensional composite beam model and reference system

Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Fig. 3 10DOF SCC beam element used (a) references defined at the beam and slab centroids and (b) references defined at the slab–beam interface



of the concrete slab. The remaining internal 2DOFs allow for axial displacement of the steel beam and the concrete slab (Fig. 3b).

Modeling of Inertia and Damping Properties

Modeling the inertia properties of the frame elements can be achieved using lumped masses at the DOFs of the external nodes. Consequently, the inertia properties of the FE model are independent of the type of finite elements employed (i.e., the structure's mass matrix can be obtained using force, displacement, or mixed-based formulation frame elements).

Even though the friction between steel beams and concrete slabs in SCC frames may be a strong source of structural damping, quantitative information about this energy dissipating mechanism usually referred to as structural damping is limited owing to the partial availability of experimental dynamic data. Consequently, the well-known and widely used Rayleigh damping model can be used by the practicing engineer. In this model, the damping matrix can be obtained using the classical Rayleigh damping relationship (Eq. 4), where the damping matrix is proportional to the mass matrix and the initial stiffness matrix:

$$[C] = \mu[M] + \lambda[K] \quad (4)$$

where

μ = mass proportional Rayleigh damping coefficient

λ = stiffness proportional Rayleigh damping coefficient

M = system structural mass matrix

K = system structural stiffness matrix

Note: The proposed model presented in the above sections considers only rigid beam-to-column connections. Nevertheless, semirigid connections can be considered in the same numerical procedure by introducing special joint elements with prescribed constitutive behavior.

Constitutive Stress–Strain Relationships

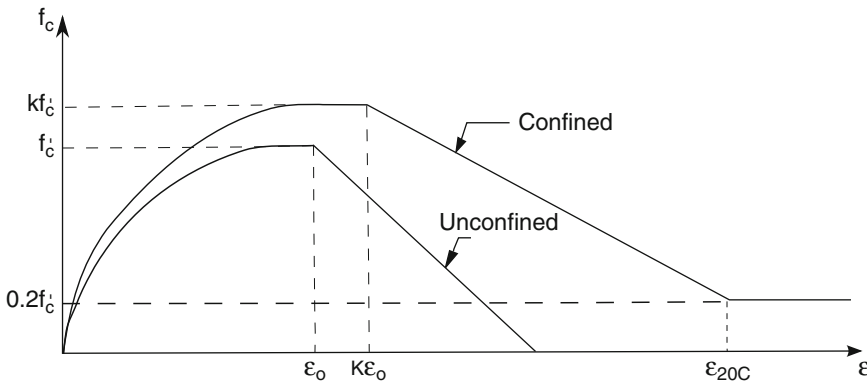
For modeling purposes, the material properties (such as the Young's modulus, Poisson's ratio,

elastic and plastic strength, and strain hardening) can be obtained from the uniaxial stress–strain curves derived from coupon tests and then applied to the corresponding fibers across the composite cross section. In order to accurately simulate the behavior of SCC beams under earthquake conditions, robust material models capable of simulating the material nonlinearity as well as other damaging effects under dynamic or cyclic loading (i.e., softening/hardening) need to be employed. Several models have been developed to achieve the aforementioned scope, some of which are presented in the following sections.

Constitutive Law for Concrete Parts (Based on the Kent–Park Model)

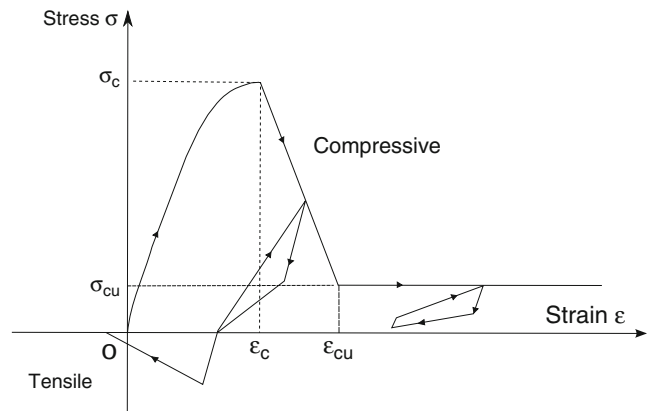
The proposed constitutive law modeling the concrete in monotonic compression for the cases of confined and unconfined concrete is the Kent–Park model as described in Park and Paulay (1975). As it is shown in Fig. 4, the material follows a parabolic stress–strain curve up to a maximum stress equal to the cylinder's strength, after which it decays linearly with strain until the residual strength is reached. In tension, the model assumes a linear stress–strain behavior up until the tensile limit of the material is reached, and then the stiffness and strength decays with increasing strain (Fig. 5).

The cyclic behavior of the concrete can be described by the Blakeley–Park model also presented in Park and Paulay (1975). The stress–strain response lies within the Kent–Park envelope; however, the effect of concrete confinement is not taken into account. The model assumes that unloading and reloading takes place along a line without energy dissipation or stiffness deterioration for strains smaller or equal to the strain corresponding to peak stress ($\epsilon < \epsilon_c$). Beyond this point, the stiffness deterioration is taken into account through the introduction of reduction factors, given by Blakeley and Park. Along the first unloading branch, the stress is reduced approximately 50 % without any reduction in strain. The reloading branch with slope equal to $f_c E$ extends back to the envelope (Fig. 5).



Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Fig. 4 Monotonic stress–strain Kent and Park model for unconfined and confined concrete

Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Fig. 5 Cyclic stress–strain Blakeley–Park model for concrete



Constitutive Law for Concrete (Based on the Popovics–Saenz Law)

The constitutive law for concrete is a uniaxial cyclic law with monotonic envelope defined by the Popovics–Saenz law (Balan et al. 1997). Linear unloading–reloading branches with progressively degrading stiffness characterize the cyclic behavior of the material. The response of concrete under cyclic loading is shown in Fig. 6. According to the same figure, after each unloading–reloading, the monotonic envelope is reached again when the absolute value of the largest compressive strain attained so far is surpassed. The tensile behavior of concrete is characterized by the same loading–unloading–reloading rules with the same initial stiffness and appropriate values for the other parameters.

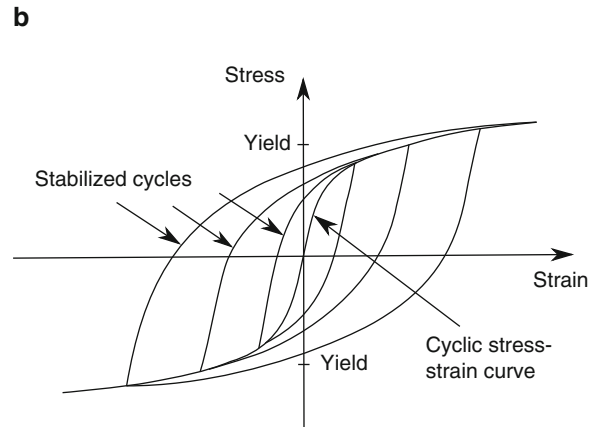
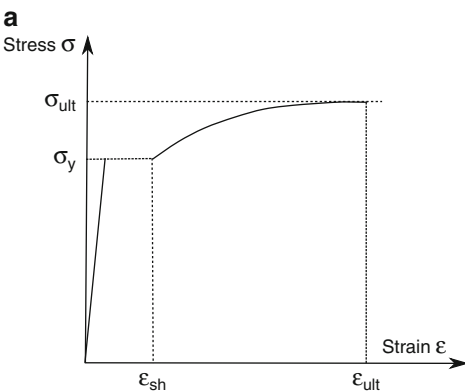
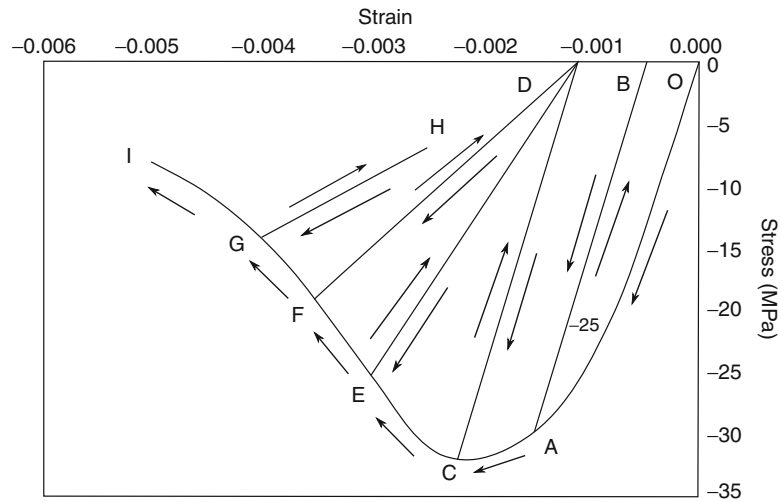
Constitutive Law for Steel

Figure 7a, b describes the elastoplastic response of the steel under monotonic and cyclic loading respectively. For monotonic loading, the characteristic yield plateau in the stress–strain model followed by a region of increased strength owing to strain hardening of the material. The unloading from the yielded condition is elastic; thereafter, the Bauschinger effect can be represented by a Ramberg–Osgood relationship (Eq. 5) until the yield stress is reached. This model uses a single nonlinear equation to characterize the observed curvilinear response of steel subjected to monotonic loading:

$$\varepsilon = \frac{\sigma}{E} + K \left(\frac{\sigma}{E} \right)^\eta \quad (5)$$

Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling,

Fig. 6 Hysteretic concrete material model under compression



Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Fig. 7 (a) Stress–strain model under monotonic loading. (b) Steady-state cyclic response of mild steel

where

$\frac{\sigma}{E}$ = equal to the elastic part of the strain
 $K\left(\frac{\sigma}{E}\right)^\eta$ = accounts for the plastic part of the strain
 K and η = parameters that describe the hardening behavior of the material

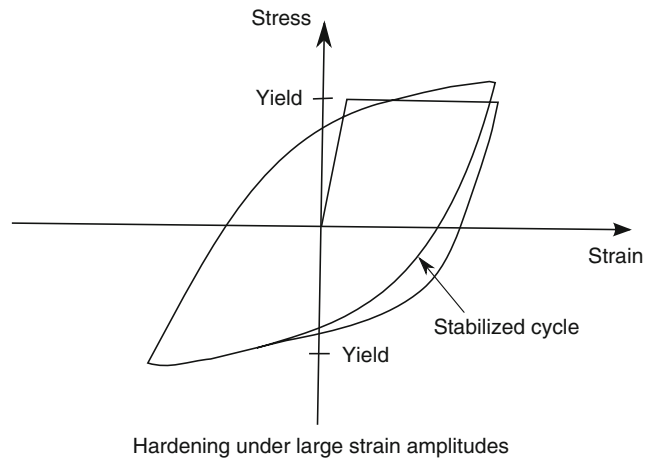
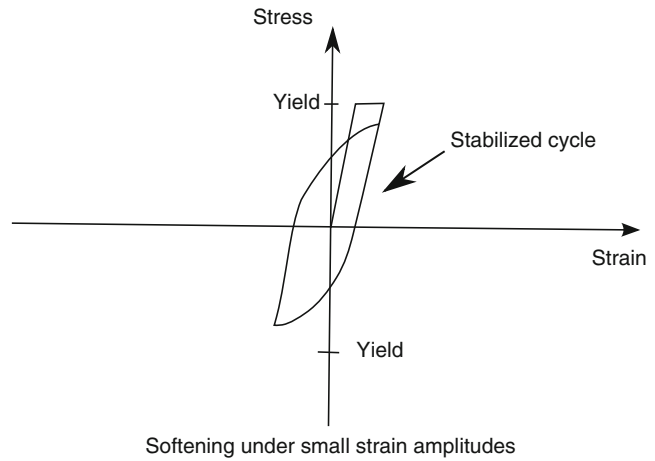
When the steel material is subjected to constant strain amplitude under cyclic loading, it exhibits a response that converges to a stabilized saturation loop which depends only on the cycling amplitude (Fig. 7b). As it is shown in Fig. 8, the response of the steel material under

constant strain amplitude cycles is described by strain hardening for large amplitudes and strain softening for small amplitudes. For the accurate simulation of the steel material response under an arbitrary load, the constitutive model needs to account for all the monotonic response, the steady-state cyclic behavior, as well as the transient behavior involving softening and hardening. This can be achieved using an efficient simplistic computationally bilinear model.

Bilinear Stress–Strain Steel Model In this bilinear model, the elastic range remains

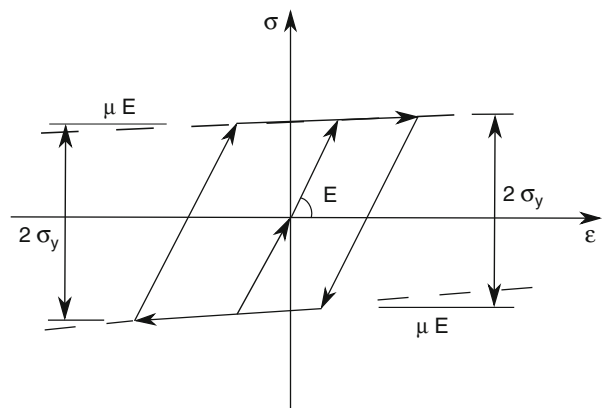
Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling,

Fig. 8 Cyclic response of steel under constant strain amplitude cycles



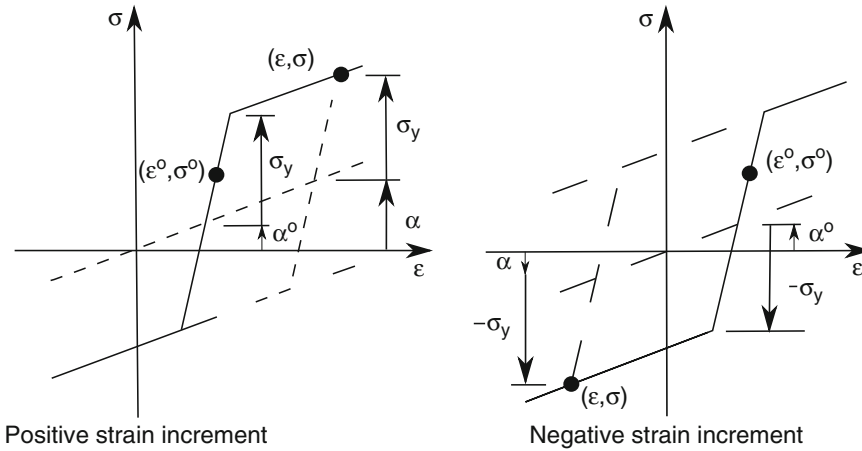
Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling,

Fig. 9 Loading and unloading paths of bilinear kinematic model



constant throughout the various loading stages. The kinematic hardening rule for the yield surface is assumed to be a linear function of the increment of plastic strain

(Fig. 9). The calculation of the current stress state is expressed mathematically using Eqs. 6, 7, and 8, and it is presented graphically in Fig. 10:



Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Fig. 10 Stress determination with the bilinear kinematic model

$$\left. \begin{aligned} \alpha &= \alpha^0 \\ \sigma &= \sigma^0 + E(\varepsilon - \varepsilon^0) \end{aligned} \right\} \quad (6)$$

If $\left(\frac{\alpha - \sigma_y - \sigma^0}{E} \right) \leq (\varepsilon - \varepsilon^0) \leq \left(\frac{\alpha^0 + \sigma_y - \sigma^0}{E} \right)$

$$\left. \begin{aligned} \alpha &= \alpha^0 + \mu E \left(\varepsilon - \varepsilon^0 - \frac{\alpha^0 + \sigma_y - \sigma^0}{E} \right) \\ \sigma &= \alpha + \sigma_y \end{aligned} \right\} \quad (7)$$

If $(\varepsilon - \varepsilon^0) > \left(\frac{\alpha^0 + \sigma_y - \sigma^0}{E} \right)$

$$\left. \begin{aligned} \alpha &= \alpha^0 + \mu E \left(\varepsilon - \varepsilon^0 - \frac{\alpha^0 - \sigma_y - \sigma^0}{E} \right) \\ \sigma &= \alpha - \sigma_y \end{aligned} \right\} \quad (8)$$

If $(\varepsilon - \varepsilon^0) < \left(\frac{\alpha^0 - \sigma_y - \sigma^0}{E} \right)$

where according to Figs. 9 and 10:

E = Young's modulus

μ = strain hardening parameter

σ_y = initial yield surface

ε = current strain

σ = current stress

α = current center of elastic range

Subscript "0" denotes values at the start of an increment

Constitutive Law for Steel (Based on the Menegotto–Pinto Model)

The constitutive law describing the behavior of the steel material is the uniaxial Menegotto–Pinto model (1973). This computationally efficient nonlinear law is capable to model both kinematic and isotropic hardening as well as the Bauschinger effect, allowing for accurate simulation and reproduction of experimental results. The response of the steel material is defined by the following nonlinear equation:

$$\sigma = b \varepsilon + \frac{(1 - b) \varepsilon}{(1 + \varepsilon^R)^{\frac{1}{R}}} \quad (9)$$

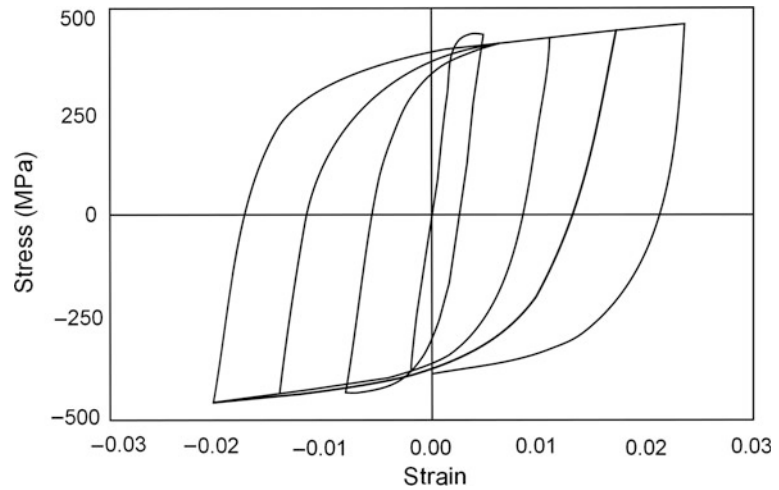
where the effective strain and stress (ε , σ) are a function of the unload–reload interval, b is the ratio of the initial to final tangent stiffness, and R defines the shape of the unloading–reloading curves. Figure 11 presents a typical stress–strain response based on the Menegotto–Pinto model. The model assumes a symmetric response for loading in compression and tension.

Interaction of Material Surfaces: Evaluation of Spring Properties

The degree of composite action and interaction between the steel beam and the concrete slab is a fundamental mechanism that needs to be

Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling,

Fig. 11 Menegotto–Pinto material constitutive model for structural steel; typical cyclic stress–strain response



considered by the engineer during the modeling procedure of SCC beams and structural systems owing to the implications on the serviceability and ultimate limit states, the energy dissipation under cyclic loading, and the local stress distribution.

Two different modeling approaches can be used for the description of partial bond in SCC structural systems. The concentrated bond approach is based on the use of concentrated springs for the modeling of the connection. The springs are attached at the location of each connector, modeling either the action of the shear stud connectors between the steel and the concrete slab or the friction in concrete-filled hollow sections and partially encased steel sections. The second approach is based on the distributed bond model, which assumes a continuous bond stress and bond slip along the contact surface. For both approaches, the uplift is typically neglected; therefore, it is considered that the concrete slab and the steel beam have the same vertical displacement and curvature.

Shear–slip relationships are widely available providing information regarding the behavior of the connectors. Figure 12a presents a simplified bilinear shear–slip relationship based on a widely used shear–slip model proposed by Ollgaard et al. (1971). The Ollgaard model is described by the following exponential function (Eq. 10) representing an experimentally observed large reduction of stiffness with increasing slip:

$$\frac{N_v}{N_{vu}} = (1 - e^{-ns})^m \quad (10)$$

where

N_{vu} = connection (ultimate) strength

N_v = shear load

s = slip between two components of the composite beam (interface slip)

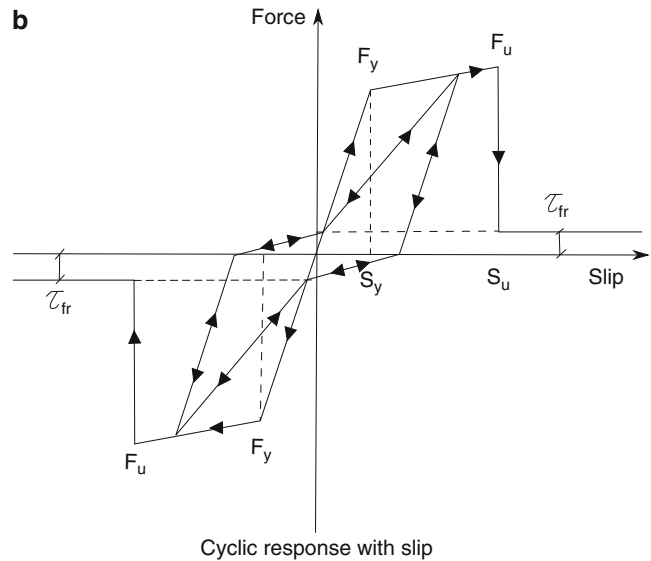
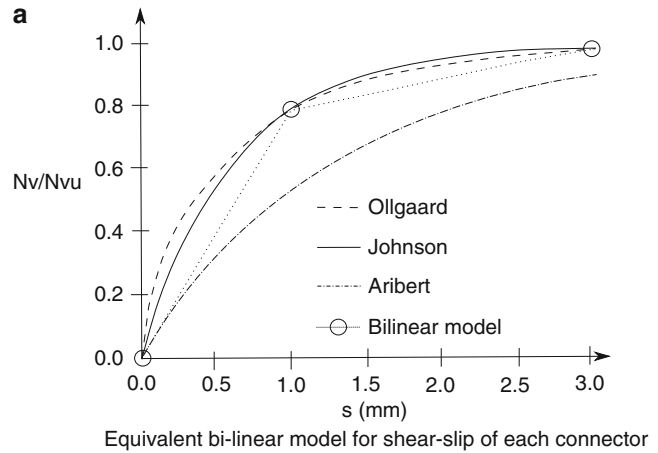
n, m = empirical parameters defining the shape of the curve calibrated from experimental data

In Fig. 12b, the monotonic envelope is presented by the definition of an ultimate slip, S_{ult} . When ultimate slip, the shear force–slip behavior follows zero stiffness with constant shear force $N_v = \pm \tau_{fr}$, where τ_{fr} is the residual shear force.

Modeling of Steel–Concrete Composite Beam-To-Column Partial-Strength Semirigid Connections

Compared to traditional bare steel structures, SCC frames can achieve more effective beam-to-column connections through the contribution of the concrete slab in resisting bending moments under gravitational and lateral loads. Additionally, these structures comprising partial-strength partially restrained beam-to-column joints designed in such a way to exhibit ductile seismic

Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Fig. 12 Shear–slip relationships

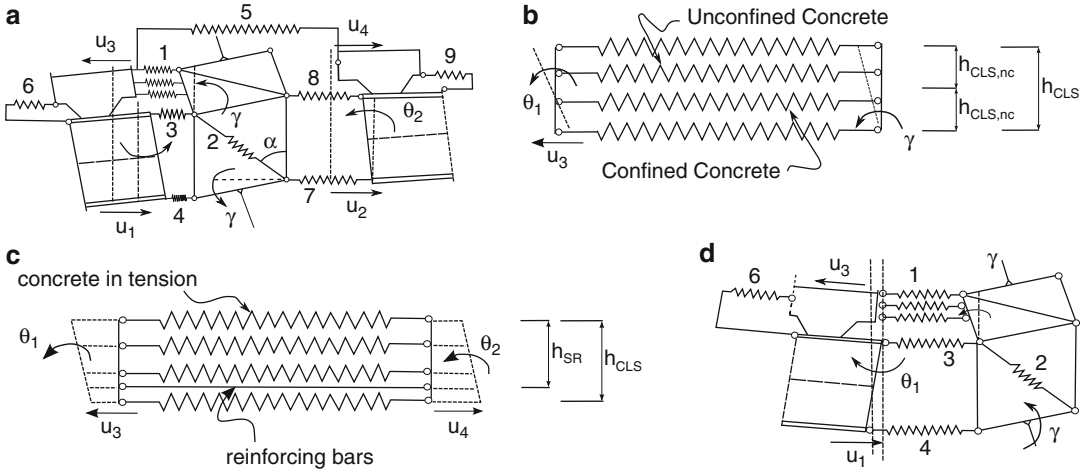


response through the plastic deformation of their components can achieve the formation of a desirable beam hinging global frame mechanism, with large hysteretic energy dissipation capacity and reduced force demand on the columns.

This section of the chapter presents a simplified approach based on the work done by Braconi et al. (2007) for the nonlinear analysis of a partial-strength beam-to-column connection using a component model. The behavior of partial-strength beam-to-column connections under the application of seismic load is described using

nonlinear spring elements as shown in Fig. 13. With reference to the same figure, the elements comprising the model should account for the response of the:

1. Concrete in compression
2. Column web panel in shear
3. Upper T-stub in compression (+Ve moment)
4. Lower T-stub in tension (+Ve moment)
5. Concrete slab in tension
6. The shear studs (+Ve moment)
7. Upper T-stub in tension (–Ve moment)
8. Upper T-stub in compression (–Ve moment)
9. Shear studs (–Ve moment)



Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Fig. 13 Component model: (a) overall joint model, (b) kinematics of the

concrete slab in compression (component 1), (c) kinematics of the concrete slab in tension (component 5), and (d) overall exterior joint model

Kinematics

Equilibrium must be maintained between the force acting in the nine components and the internal and external forces. The response of the assembled model is therefore defined by a set of eight equations related to translational equilibrium between components in the same beam-to-column connection, equilibrium between the shear studs and the beam steel profile, rotational equilibrium between the internal forces in the beam-to-column connection and the bending moment of the beam framing in it, as well as rotational and translation equilibrium acting on the column web panel. The set of eight equations can be then solved using a numerical procedure (i.e., Newton–Raphson) considering the storey deformation (drift, δ) as such an external action in the format of imposed deformations.

On the basis of the small displacement theory, the local kinematics could be described using a total of 7DOFs: the horizontal displacement of the bottom surface of the slab on both sides of the column, u_1 and u_2 ; the horizontal displacement of the top surface of the slab on both sides of the column, u_3 and u_4 ; the relative rotation of the two beams with respect to the column faces, θ_1 and θ_2 ; and the column panel zone shear distortion, γ . The deformations, δ , of the nine components

comprising the model are linked to the seven degrees of freedom through the following equations:

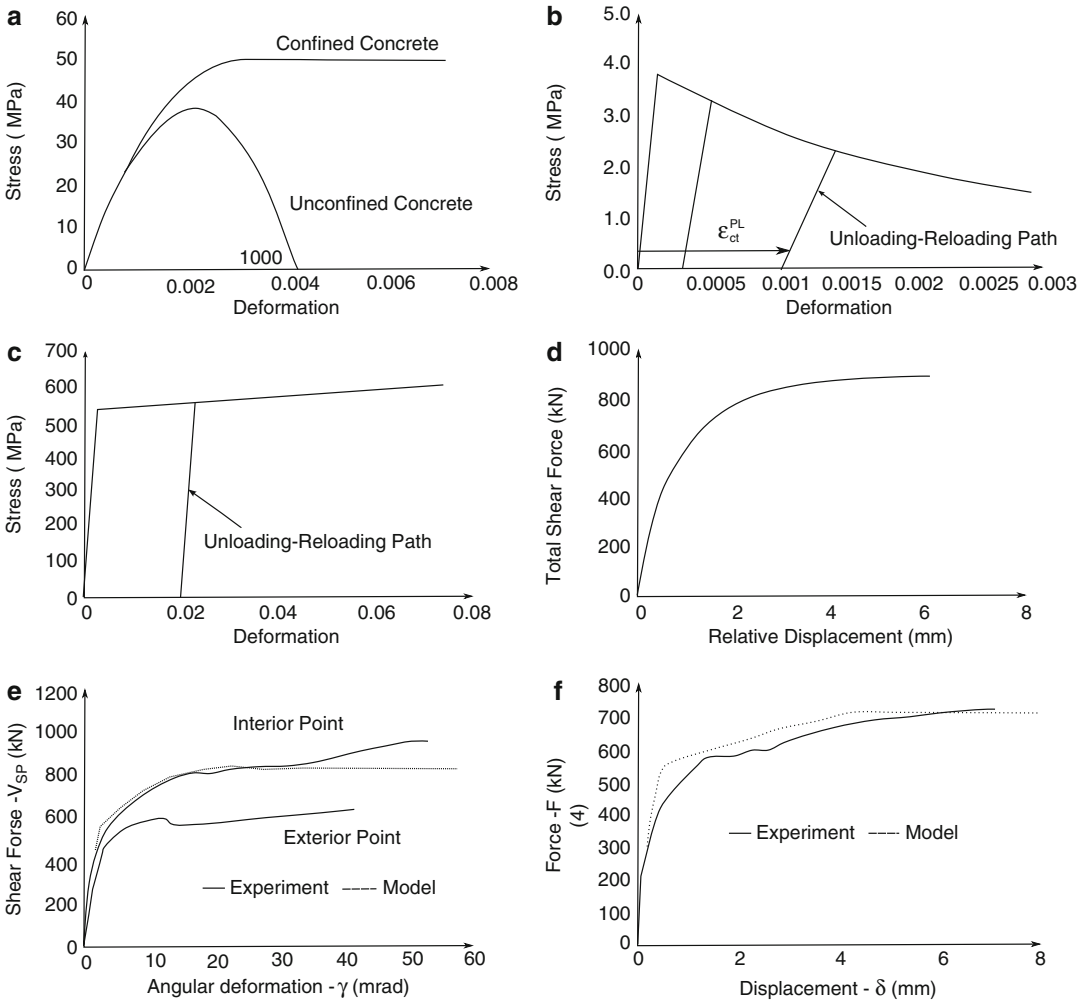
$$\delta_{1,i}^{nc} = -u_3 - \frac{\Delta_{nc}}{2} (2i - 1) \tan(\theta_1) + \left[h_{cs} - \frac{\Delta_{nc}}{2} (2i - 1) \right] \tan(\gamma) \quad (11)$$

$$\delta_{1,i}^c = -u_3 - \left[\frac{\Delta_c}{2} (2i - 1) + h_{cls,nc} \right] \tan(\theta_1) + \left[h_{cls} - \frac{\Delta_c}{2} (2i - 1) \right] \tan(\gamma) \quad (12)$$

$$\delta_3 = -u_1 - \left(h_{cs} + \frac{t_{bf}}{2} \right) \tan(\theta_1) \quad (13)$$

$$\delta_4 = -u_1 - \left(h_{cs} + h_b - \frac{t_{bf}}{2} \right) \tan(\theta_1) \quad (14)$$

$$\delta_{5,i} = u_4 - \left[h_{cs} - \frac{\Delta_{ct}}{2} (2i - 1) \right] \tan(\theta_2) - u_1 - \left[\frac{\Delta_{ct}}{2} (2i - 1) \right] \tan(\theta_1) \quad (15)$$



Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Fig. 15 Component constitutive relationships: (a) concrete in compression,

(b) concrete in tension, (c) reinforcing steel, (d) shear stud, (e) panel zone in shear, and (f) lower T-stub in tension

Concrete slab in compression can be modeled using stress–strain relationships proposed in EC2 or any of the stress–strain relationships presented in sections “[Constitutive Law for Concrete Parts \(Based on the Kent–Park Model\)](#)” and “[Constitutive Law for Concrete \(Based on the Popovics–Saenz Law\)](#)” for both unconfined (concrete above reinforcement level) and confined (concrete below reinforcement level) conditions. A typical example of the stress–strain relationship that can be used in analysis is presented in Fig. 15a for both concrete

conditions. In tension (Fig. 15b), the behavior of concrete can be represented using a linear response until cracking, followed by a softening branch where the tensile resistance reduces exponentially as proposed by Stevens et al. (1991). Linear unloading–reloading branches can be adopted intersecting the deformation axis at a residual plastic deformation ϵ_{ct}^{PL} given by Eq. 22 and also shown in Fig. 15b:

$$\epsilon_{ct}^{PL} = 146 (\epsilon_{ct}^{max})^2 + 0.523 (\epsilon_{ct}^{max}) \quad (22)$$

Steel Reinforcement

Bilinear stress–strain relationships as shown on Fig. 15c can be adopted for representing the behavior of steel reinforcement.

Shear Connectors

The slip between the concrete slab and the beam owing to the flexibility of the shear stud connectors can be modeled using force–deformation relationships similar to those presented earlier in section “Interaction of Material Surfaces: Evaluation of Spring Properties” (“simulation of composite action”). The recommended force–deformation relationship for this particular model is based on the model proposed by Aribert and Lachal (2000) (and is tailored to the guidelines of EC4 and EC8 for the calculation of the ultimate shear stud resistance), and it is presented in Fig. 15d and Eq. 23.

$$F = Q_u \left(1 - e^{C_1(u_3 - u_4)} \right)^{C_2} \quad (23)$$

where

Q_u = ultimate shear stud resistance calculated according to EC4 and EC8

C_1, C_2 = coefficients suggested by Aribert and Al Bitar (1989) depending on the height of studs and type of steel profile

Panel Zone in Shear

The behavior of the panel zone plays a significant role in determining the overall stiffness and capacity of the frame. In terms of seismic design, the panel zone can have a significant influence on the distribution of plasticity and energy dissipation on the overall performance of the structure. A multi-linear shear force to shear deformation response is retained for the web panel zone. The following equations describe this multi-linear relationship through the gradients of elastic, post-elastic, and strain hardening stiffness:

$$K_{el, wp} = G_S \frac{A_{vc}}{z} \quad (23)$$

$$K_{t, wp} = 1.04 G_S \frac{b_{cf} t_{cf}^2}{z} \quad (24)$$

$$K_{s, wp} = \frac{G_S}{a_H} \frac{(h_c - t_{cf})t_{cw}}{z} \quad (25)$$

where

G_S = shear modulus of steel

A_{vc} = shear area of the column section

z = centerline vertical distance between the column stiffeners

b_{cf} = column flange width

t_{cf} = column flange thickness

h_c = column depth

t_{cw} = column web thickness

a_H = hardening coefficient related to the thickness of the column web panel and the column flanges

Using the elastic and post-elastic stiffness, shear forces in the panel zone can be obtained. The elastic stiffness, $K_{el, wp}$, is obtained from EC3 and is applicable until the shear force, V_{wp} , reaches the yield limit, $V_{el, w}$, which is also specified in EC3 (Eq. 26). The shear forces correspond to the post-elastic branches, described by Krawinkler’s model (Krawinkler 1978), in which the shear force is obtained through Eqs. 27 and 28. The three different branches of the shear force–deformation relationship are presented in Fig. 15e (dashed line):

$$V_{el, wp} = \frac{0.9 f_{y, cw} A_{vc}}{\sqrt{3}} \quad (26)$$

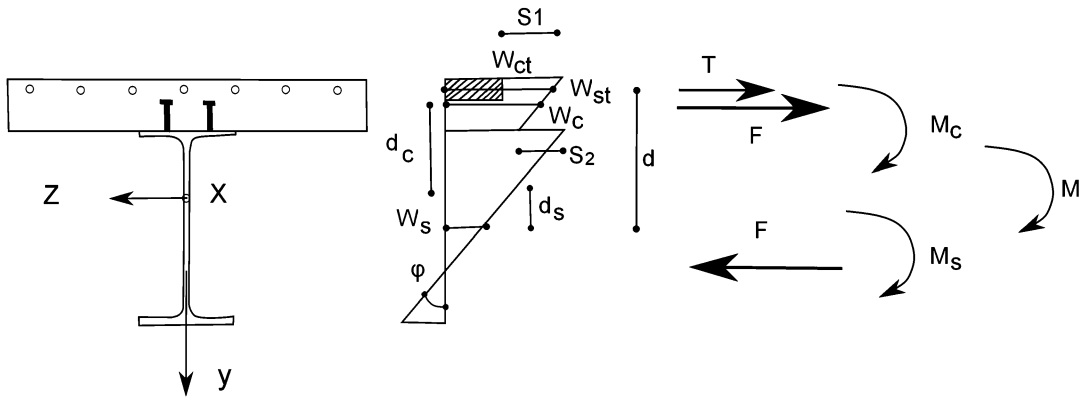
$$V_{t, wp} = \frac{0.9 f_{y, cw} A_{vc}}{\sqrt{3}} + 1.04 G_S \frac{b_{cf} t_{cf}^2}{z} 3 \gamma_y \quad (27)$$

$$V_{s, wp} = \frac{f_{u, cw} A_{vc}}{\sqrt{3}} + \frac{f_{u, cf} b_{cf} t_{cf}^2}{4z} \quad (28)$$

where

γ_y = panel distortion at yield = $V_{el, wp}/K_{el, wp}$

$f_{u, cw}, f_{u, cf}$ = ultimate tensile stress for column web and flanges



Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Fig. 16 Cross-sectional model under negative bending

T-Stubs Components

The geometry of the equivalent T-stub components can be determined using the effective width concept presented in EC3 for stiffened columns and end-plates. The force–deformation relationship of T-stub elements required for component modeling can be obtained from the model proposed by Piluso et al. (2001) (dashed line) (Fig. 15f).

Simplifications and Assumptions

1. The difference in column web shear stiffness between the concrete-encased and non-encased segments of the columns should be taken into account in modeling through the assumption of infinite stiffness of the upper encased portion of the column and a flexible diagonal spring in the bare steel column web over the beam depth.
2. Bolt pretension effects need to be considered in modeling through modification of the stiffness of the equivalent T-stub springs in the elastic range.
3. There is an interaction between the connections on both column sides due to the continuity of the slab and the slab longitudinal steel reinforcement between the two beams. Hence, including the anchorage steel bars, the negative moment capacity of one side is dependent on the concrete capacity of the opposite side while essentially transferring the tensile forces acting from the reinforcing steel to the

column. The modeling engineer can use more advanced modeling procedures (i.e., as proposed by Fabbrocino et al. 2002) when continuous SCC beams are considered for the assessment of the behavior of the connection capacity, in terms of global quantities such as the rotations and deflections as well as local quantities such as the slip and the curvature of members, the interaction forces, and the rebar strain.

4. Column web buckling and beam flange buckling under compression are not considered in such modeling procedure.

Fabbrocino et al. 2002

1. Modeling the cross section of composite beams is achieved through a modification of the well-known Newmark's kinematic model, as shown in Fig. 16. This approach requires the definition of the slab effective width depending on the type of loading (hogging or sagging) and on the connection detailing at the beam end. A linear strain pattern is then applied to each component of the cross section. Under the assumption that the curvature and the rotation for each of the components are the same (e.g., for both the concrete slab and the steel profile), the uplift is neglected. Using analytical procedures, the tensile stresses developed in the concrete slab; the slip between the different components of the cross section; the interaction force, F ; the

global bending moments in the steel profile, M_s ; and the concrete slab, M_c , as well as the moment–curvature relationship of the cross section can be obtained.

2. Modeling the continuous composite beam is based on a combination of the main behavioral aspects on the different regions of the beam (i.e., Newmark's model is used for sagging moments, whereas its modified version is used when cracked zones of the beam are considered). The moment–curvature relationship in each section of the beam can be then defined through an iterative process. Once the generalized moment–curvature relationship is established, rotations and displacements can be obtained by integration of the curvature distribution. The numerical procedure for the solution of a simple structural system of a beam characterized by geometrical and mechanical symmetry is based on the compatibility method; therefore, the support bending moment is the main unknown and the beam is statically determined. The reader is referred to Fabbrocino et al. (2002) for a step-by-step guidance for the solution of the composite section and beam.

Modeling of the Panel Zone in Semirigid Steel–Concrete Composite (SCC) Connections

When moment-resisting frames are subjected to horizontal loading such as earthquake excitation, unbalanced moments occur at the beam-to-column connections resulting to shear deformations in the panel zones of the columns. Therefore, the behavior of the panel zone plays a significant role in determining the overall stiffness and capacity of the frame. In addition to this, in terms of seismic design, the panel zone can have a significant influence on the distribution of plasticity and energy dissipation mechanisms and in turn significant effects on the overall performance of the structure.

This section deals with the modeling of the panel zone region within the beam-to-column joints of SCC moment-resisting frames. This

particular model proposed by Castro et al. (2005) is based on a realistic stress distribution at the edges of the panel, aiming to account for the location of the neutral axis. This methodology enables assessment of the shear stress distribution through the panel depth representing the distribution of the plasticity in the vicinity of that region. Both shear and bending deformations are considered in the elastic and post-elastic stages. The additional resistance of the panel zone owing to the contribution of the column flanges is also taken into account by considering the column depth and flange thickness (Fig. 17).

Procedures and Details

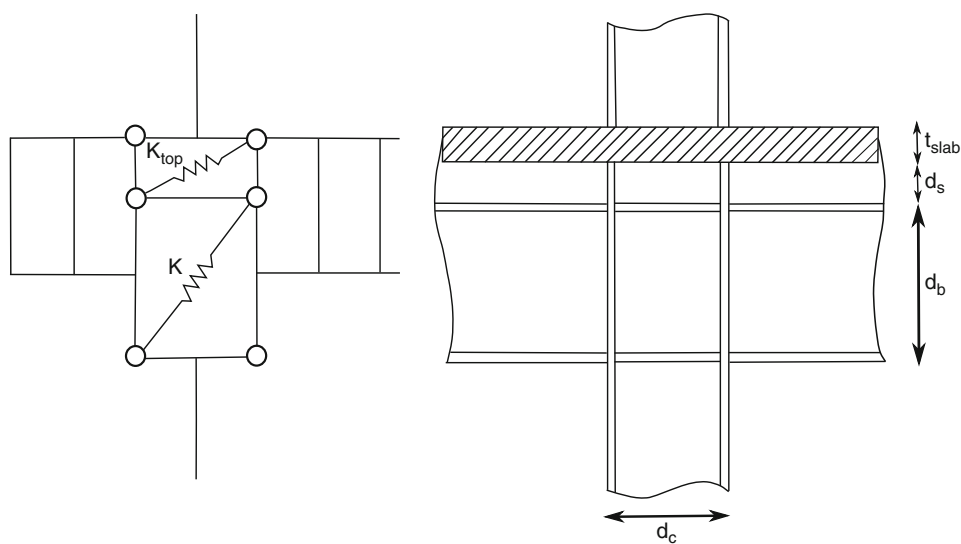
With reference to Fig. 18, the physical dimensions of the panel zone, d_c and d_b , are taken into account. The part of the column being in contact with the slab is modeled through an assemblage of links on top of the panel. This modeling approach is essentially a determination of the spring properties of both the “panel zone” and the “top panel” which are derived analytically allowing for implementation in frame analysis software. The procedure establishes an analogy between the analytical (consisting of the actual connection) and the corresponding numerical model for frame analysis, as shown in Fig. 18.

The location of neutral axis is calculated considering a linear stress distribution, based on the assumption that the SCC beam behaves elastically until the panel yields.

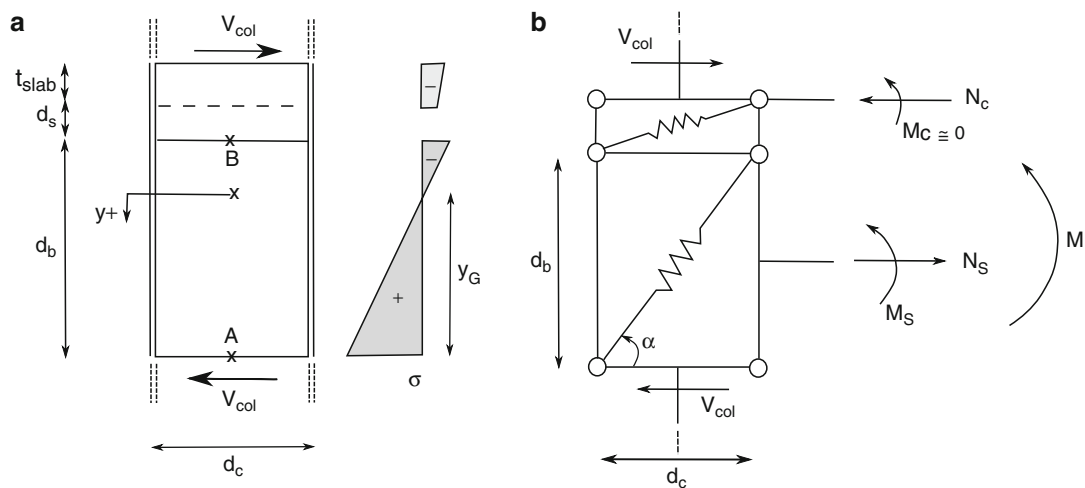
An assessment of the effective width of the slab in the vicinity of the connection is also required. For positive moment (sagging), it is assumed that the slab is limited to the contact width of the column flange width, b_c . For negative moment (hogging), the slab is not considered, under the assumption that the reinforcement is not anchored to the column.

Notion of Calculations

The main aspect of the proposed model is the determination of the spring properties for both the panel zone and the top panel. For a given moment carried by the SCC beam, a corresponding equivalent shear is applied to the panel in the numerical model. The stiffness



Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Fig. 17 Numerical and analytical representation of joint models



Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Fig. 18 Analytical and numerical representation of joint models

that should be used in the numerical model can be then determined from the distortion caused to the analytical model owing to the application of moment. Furthermore, the difference between load level corresponding to first and full yielding of the panel can be derived given that the shear stress distribution through the panel is known. The procedure needed to be followed by the engineer for derivation of these important parameters

for both elastic and post-elastic ranges is outlined below.

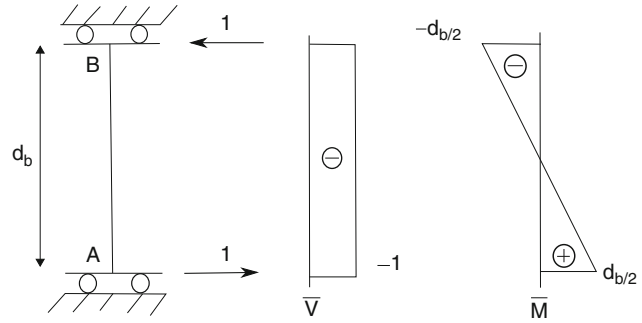
This section serves as a numerical guideline; the reader is referred to J. M Castro et al. (2005) to exploit the complete set of equations.

Elastic Range

1. The neutral axis location of the steel beam and in turn its second moment of area can be

Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling,

Fig. 19 The virtual system used to find panel deformations



calculated under the assumption that the steel beam behaves in an elastic manner up to yielding of the panel zone.

- The ratio of axial force (N_s) and bending moment (M_s) carried by the beam can be then obtained using analytical procedures.
- The total moment (M) acting to the connection can be then calculated in the numerical model using Eq. 29. The moment (M_c) developed in the slab is considered to have insignificant influence and therefore it is ignored:

$$M = M_s + N_s \left[\frac{d_b}{2} + d_s + \frac{t_{slab}}{2} \right] \quad (29)$$

- The equivalent moment carried by the panel can be then obtained from Eq. 30:

$$V_{eq} = \frac{M_s}{d_b} + \frac{N_s}{2} - V_{col} \quad (30)$$

where $V_{col} = \frac{M}{h_s}$ and h_s is the storey height.

- From the analytical model, knowing the normal stress distribution of the composite section, the shear force and bending moment distributions can be obtained.
- Using the principle of virtual work (Fig. 19) and the calculated shear force and bending moment distributions, the relative horizontal displacement can be obtained. Having the application of opposite unit forces in the virtual system, the internal virtual forces (V_{int} , M_{int}) can also be obtained.
- Using the calculated equivalent shear force (V_{eq}) and the relative horizontal displacements ($|\Delta_{shear}|$, $|\Delta_{bending}|$), the elastic stiffness

to be used in the numerical model is calculated by

$$K_{el} = \frac{V_{eq}}{|\Delta_{shear}| + |\Delta_{bending}|} \quad (31)$$

- Using the calculated elastic stiffness, the relative drift of the panel ($\Delta_{y,el}$) at the onset of yielding can be obtained analytically.
- Finally, using the calculated elastic stiffness and the relative drift of the panel zone, the elastic stiffness and the relative drift of the spring can be obtained using the Eqs. 32 and 33:

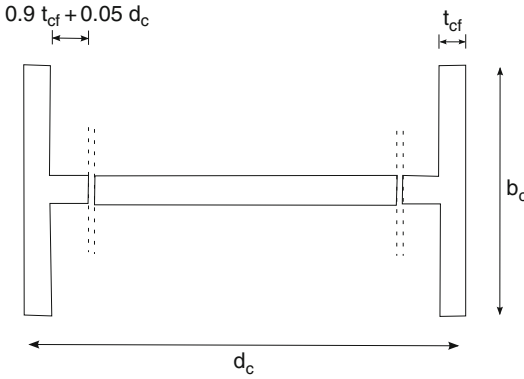
$$K_{el, spring} = \frac{K_{el}}{\cos^2 a} \quad (32)$$

$$\Delta_{el-y, spring} = \frac{\Delta_{y,el}}{\cos a} \quad (33)$$

where a is the angle of the spring as indicated in Fig. 18.

Post-Elastic Range

- Beyond the yielding of the panel, the shear stiffness of the column web is assumed to drop to the strain hardening stiffness of the material. Consequently, the post-elastic stiffness of the panel is provided by the strain hardening of the column web by the flanges and a portion of the column web delimiting the panel zone, as shown in Fig. 20. Following the same assumption for the beam remaining largely elastic, and by following similar procedures to those described for the elastic



Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Fig. 20 The cross-sectional definition for the post-elastic range

range, the post-elastic stiffness for the panel can be obtained from Eq. 34:

$$K_{pl} = \left[\mu \frac{V_{eq}}{\Delta_{shear}} \right] + \left[\frac{V_{eq}}{\Delta_{bending}} \frac{2I_t}{I_{col}} \right] \quad (34)$$

where

μ = strain hardening parameter

I_t = second moment of area of the T-section obtained analytically

I_{col} = second moment of area of the column obtained analytically

2. The relative deformation, $\Delta_{y,pl}$, of the panel zone in the post-elastic range can be obtained using the principle of virtual work and can be readily derived as:

$$\Delta_{y,pl} = \Delta_{y,el} + \frac{\bar{f}_y d_b^2}{6E d_{CG}} \quad (35)$$

where

d_{CG} = distance from the centroid of the T-section to the external fiber of the column flange

$\bar{f}_y$ = reserve stress in the same fiber after shear yielding of the panel zone

3. Having known the calculated post-elastic stiffness and relative drift of the panel zone, the post-elastic stiffness and relative drift of the spring can be obtained using the Eqs. 36 and 37:

$$K_{pl, spring} = \frac{K_{pl}}{\cos^2 a} \quad (36)$$

$$\Delta_{pl-y, spring} = \frac{\Delta_{y, pl}}{\cos a} \quad (37)$$

4. Finally, the stiffness provided by the panel zone owing to strain hardening in shear (the first term of Eq. 34) is given by:

$$K_{S, Hardening} = \mu \frac{V_{eq}}{\Delta_{shear}} \quad (38)$$

Therefore, the strain hardening stiffness of the diagonal spring can be determined from

$$K_{S, H-Spring} = \frac{K_{S, Hardening}}{\cos^2 a} \quad (39)$$

Moment–Rotation Relationship of Panel Zone

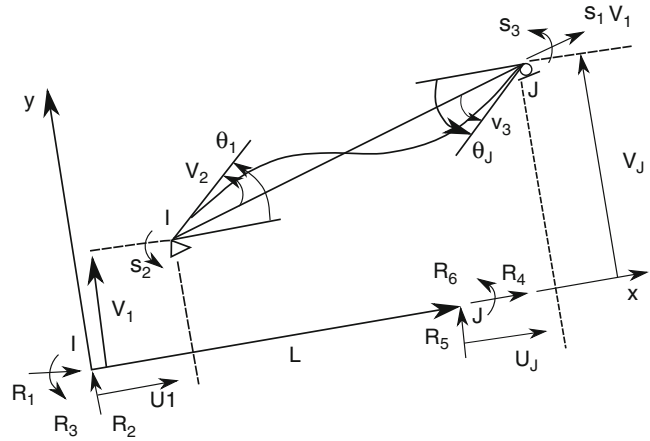
The rotational springs in the panel zone are modeled using trilinear moment–rotation relationships while using the stiffness derived from the above expressions for the elastic and post-elastic range (later in Fig. 22). According to that figure, the panel zone is expected to behave in an asymmetric manner in tension and compression owing to the presence of concrete slab which influences the stiffness, yield moments, and cyclic behavior of the panel zone. Under negative moment, the concrete slab contribution is ignored owing to cracking of the concrete. Yield points (M_{py}^+ , θ_{py}^+) and (M_{py}^- , θ_{py}^-) are controlled by the steel yielding, and the ultimate points (M_{pu}^+ , θ_{pu}^+) and (M_{pu}^- , θ_{pu}^-) are controlled by the ultimate strength of the concrete. Under cycling loading, unloading occurs in a straight line with the same slope as the initial stiffness K_e . The reloading is directed towards the previous peak, thereby considering some strength and stiffness degradation.

Modeling of Frames Using Beam Elements

This section presents an analytical approach for the assessment of SCC frames under earthquake excitation using two-dimensional SCC beam

Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling,

Fig. 21 Element relative forces and deformations in local coordinate system



elements (Kim and Engelhardt 2005). In order to model the behavior of SCC beams under earthquake excitation, factors such as the beam-to-column connection details, the local crushing of concrete, the loading pattern on beam, as well as the bond behavior between the reinforcing steel and the concrete need to be considered into the modeling. It is widely accepted that three-dimensional FE models can accurately predict the behavior of SCC beams at the expense of time and computational efficiency. On the other hand, these simpler but reasonably accurate two-dimensional SCC beam models can provide an alternative tool for frame response assessment.

Beam Elements

The beam elements are described as a one-component series hinge-type model combining analytical formulations calibrated against to experimental data and to other data from “sophisticated” described models, while intended to represent the clear span of beams in moment frames (i.e., the length of the beams between column flanges). The two-dimensional SCC beam elements are described by a linear elastic beam with a nonlinear zero-length hinge at each end; the resulting element is referred as “complete element.” Each of the hinges is described using nonlinear rigid-plastic moment–rotation relationships in order to simulate the real structural

behavior which was observed from experiments. Each complete element is characterized by two external and two internal nodes. The internal nodes are located between the connection of the linear elastic beam element and the hinges, while the external nodes connect the entire structure. Each of the external nodes has 3DOFs, 2-translations, and 1-rotation in the local coordinate system as presented in Fig. 21.

In the local coordinate system, the element can be considered as a simply supported beam given that the rigid body motions are removed. Based on equilibrium, using the values of relative forces (s_1 , s_2 , s_3), all the components of local nodal forces (R_1 to R_6) can be calculated. The transformation of forces is defined using the following relationship:

$$R = A s \quad (40)$$

where

A = force transformation matrix (this is well known and can be found in literature, i.e., Przemieniecki (1968))

From the geometry, the transformation from the local displacements, r , to the relative deformations (v_1 , v_2 , v_3) is performed by:

$$v = A^T r \quad (41)$$

where

$$\mathbf{r}^T = \{U_I, V_I, \theta_I, U_J, V_J, \theta_J\} \quad (42)$$

Element Stiffness

The initial stiffness of the aforementioned “complete element” is that of the linear elastic beam. As gradual yielding occurs at the hinges owing to increased moments at the element ends, the stiffness of the “complete element” reduces accordingly. In order to obtain the reduced stiffness at any load step after yielding, the instantaneous tangent flexibility of the nonlinear rigid-plastic force–deformation relationship for a hinge is combined with the flexibility of the elastic beam element.

A flexibility matrix, $\mathbf{f}$, is first formed for the elastic element including the effects of elastic shear deformation through the following relationship:

$$dq = \mathbf{f} ds \quad (43)$$

where

$dq = (dq_1, dq_2, dq_3) =$ elastic deformation increment at the internal nodes

$ds =$ action increment in which
 $ds^T = \{dF, dM^I, dM^J\} = \{ds_1, ds_2, ds_3\}$

For hinges at nodes I and J, the incremental action–deformation relationship is expressed by

$$dw_p = \begin{Bmatrix} 0 \\ d\theta_p^I \\ d\theta_p^J \end{Bmatrix} = \begin{Bmatrix} dv_1 - dq_1 \\ dv_2 - dq_2 \\ dv_3 - dq_3 \end{Bmatrix} = \mathbf{f}_p ds \quad (44)$$

where

$dw_p =$ vector of plastic hinge deformations at nodes I and J

$d\theta_p^I, d\theta_p^J =$ incremental plastic rotation at nodes I and J

$\mathbf{f}_p =$ hinge or plastic flexibility matrix in which nonzero terms are the second and third elements in the diagonal

Using Eqs. 43 and 44, the action–deformation relationship can be obtained for the “complete

element” expressed in terms of degrees of freedom, $\mathbf{v}$, as follows:

$$dv = ds + dw_p = \mathbf{F}_t ds \quad (45)$$

The hinge flexibility coefficients, $\mathbf{f}_p$, can be simply added to the appropriate coefficients of the elastic element flexibility matrix, $\mathbf{f}$, in order to obtain the tangent flexibility matrix, $\mathbf{F}_t$, for the “complete element,” as shown in Eq. 46. Once the 3×3 tangent flexibility matrix is obtained, it is then inverted to obtain the 3×3 tangent stiffness matrix, $\mathbf{K}_t$:

$$\mathbf{F}_t = \begin{vmatrix} \frac{L}{EA} & 0 & 0 \\ 0 & \frac{L}{EA^+} \mathbf{F}_{ii} + \frac{1}{GA_s L} + \mathbf{f}_p^i & -\frac{L}{EA^+} \mathbf{F}_{ij} + \frac{1}{GA_s L} \\ 0 & -\frac{L}{EA^+} \mathbf{F}_{ij} + \frac{1}{GA_s L} & \frac{L}{EA^+} \mathbf{F}_{jj} + \frac{1}{GA_s L} + \mathbf{f}_p^j \end{vmatrix} \quad (46)$$

where

$EA^+ =$ flexural rigidity of the composite beam

$EA =$ axial rigidity of the composite beam

$GA_s =$ effective shear rigidity of the composite beam

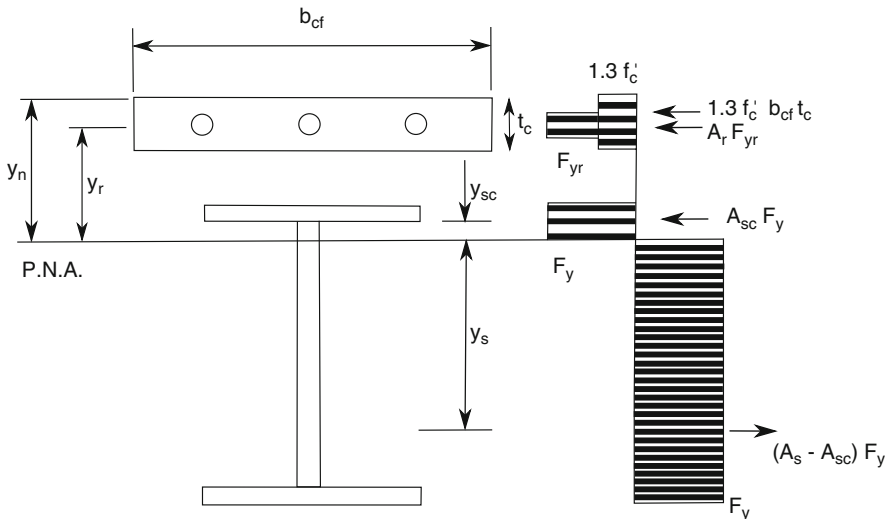
$\mathbf{F}_{ii} = \mathbf{F}_{ij} = 1/3$ and $\mathbf{F}_{jj} = 1/6$ for a uniform member

$\mathbf{f}_p^i, \mathbf{f}_p^j =$ flexibility of hinges at nodes I and J

Hysteretic Rules

Due to the cross-sectional asymmetry of the SCC beam, the response will be different for positive and negative moments. As a result, a hinge must discern the load paths to model the hysteretic behavior of SCC beam for an arbitrary cyclic loading. Apart of the cross-sectional asymmetry, the hysteretic rules employed for the complete element need to take into account factors such as the strength deterioration and the stiffness degradation.

In this part of the chapter, the hysteretic rules are determined from the modification of Lee’s model (1987) in order to better fit the curves and the nonlinear behavior of experimental test specimens. This model employs a specified



Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling, Fig. 23 Plastic stress distribution for composite beam

The engineer is then required to calculate the effective width of the concrete slab on each side of the beam centerline for computing the positive elastic stiffness. This will enable the calculation of the second moment of inertia of the transformed SCC section. The effects of slip between the concrete and the steel on the positive elastic stiffness are taken into account through the use of a fraction of the second moment of inertia of the transformed SCC section. The calculation of the moment of inertia for the negative elastic stiffness needs to be performed, taking into account the steel beam section and the reinforcing steel bars within the effective slab width. The varying moment of inertia of SCC and cracked sections along the length of the beam is assumed to be equivalent to a uniform moment of inertia of a cracked section with reinforcing bars within the effective width.

Calculation of Moment at the Yielding and Ultimate Point Using a plastic stress distribution for SCC beams (Fig. 23), the ultimate moment (M_{max}) at the connection can be obtained. The positive yield moment (M_y^+) is assumed to be a fraction of the calculated ultimate moment. The negative yield moment (M_y^-)

is the plastic moment of both the steel beam section and the reinforcing steel bars within the effective width.

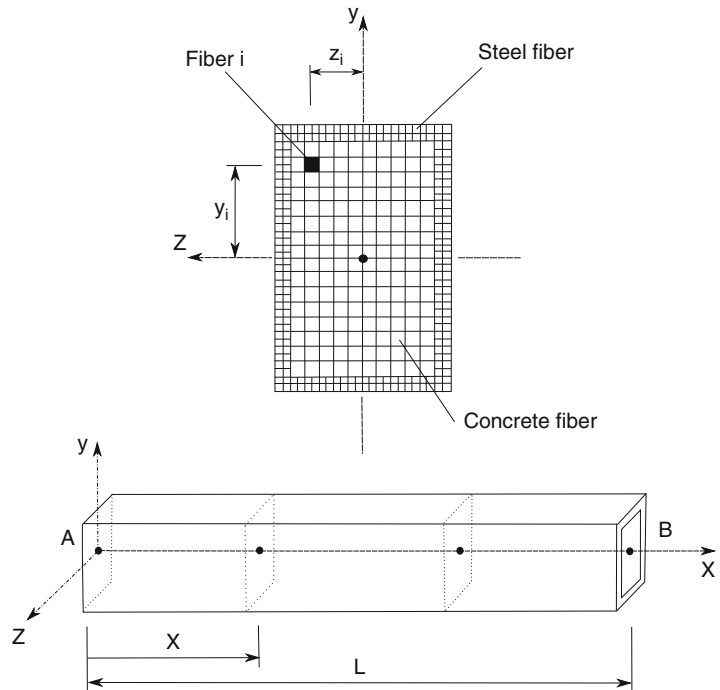
The contribution of the concrete slab to the ultimate moment at the connection is determined using the column width and a concrete compressive bearing stress of $1.3f'_c$, due to concrete confinement near the face of the column.

Modeling of SCC Frames with Concrete-Filled Steel Columns

The advantages of concrete-filled steel (CFS) structures in terms of high strength, high ductility, and large energy absorption led to their extensive use in high-rise structures in earthquake-prone regions.

This section presents a numerical procedure for the nonlinear inelastic analysis of CFS frames based on a fiber beam-to-column element. The nonlinear response of SCC frames is captured through the inelasticity of materials or due to changes in the frame geometry. Global geometric nonlinearities ($P-\delta$ effects) are taken into account by the use of stability functions derived from the exact stability solution of a beam-to-column element subjected to axial forces and bending

Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling,
Fig. 24 Fiber hinge concept



moments. The spread of plasticity over the cross section and along the member length is captured by tracing the uniaxial stress–strain relationships of each fiber on the cross sections located at the selected integration points along the member length. The nonlinear equilibrium equations can be then solved using an incremental iterative scheme, based on the generalized displacement control method.

Fiber Beam-to-Column Element and Material Nonlinearity

The gradual plastification of a composite cross section can be described using the concept of fiber section model, similarly to the modeling of the concrete slabs in section “[Introduction](#).” The fiber model is presented in Fig. 24. The concept behind this model is rather simple; the cross-sectional area of the SCC element is subdivided into fibers represented by their area, A_i , and coordinate location (y_i , z_i – with origin the centroid of the section). Different material properties (e.g., concrete confined and unconfined, steel, reinforced steel) can be assigned to each of the fibers. Based on the relevant constitutive material

models, the fiber strains are used to calculate the fiber stresses, which are in turn integrated over the cross-sectional area to obtain stress resultants (i.e., forces and moments).

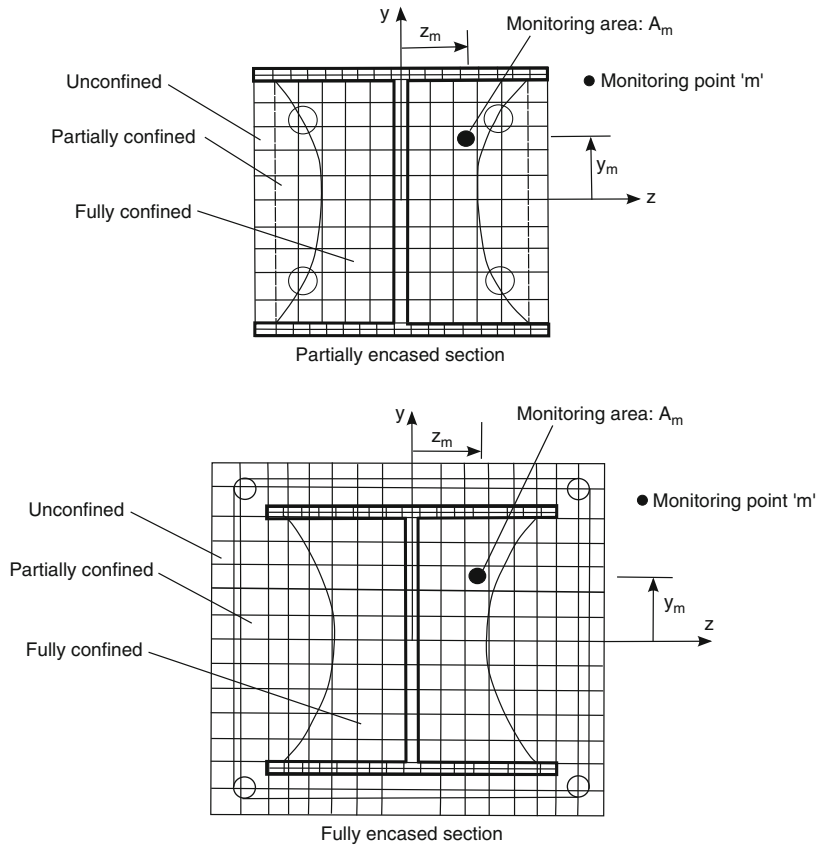
Using the fiber model, a number of assumptions have been made as follows:

1. Sections remain plane after bending.
2. Due to the latter assumption, cracking is considered to be smeared and normal to the member axis.
3. Torsional and shear effects are ignored.
4. Multi-axial stress states (due to the confinement effects) can be included in the model by increasing the concrete strength and modifying its post-peak response.
5. Local buckling effects or initial stress arising from thermal effects or erection loads are typically not included.

Confinement of Concrete-Encased Steel Sections

In order to utilize the material constitutive models described in the previous sections, the concrete confinement zones need to be identified (Fig. 25).

Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling,
Fig. 25 Confinement zones and distribution of monitoring points of SCC sections



Several methods for the identification of the confined zones have been suggested for partially and fully encased sections, and confinement factors, k , have been developed for use in numerical modeling. According to EC8, the confinement factor, k , is given by:

$$k = 1.0 + 5.0 \alpha_c f_l / f_{co} \quad \text{for } f_l / f_{co} < 0.05 \quad (47)$$

$$k = 1.125 + 2.5 \alpha_c f_l / f_{co} \quad \text{for } f_l / f_{co} \geq 0.05 \quad (48)$$

where

α_c = confinement effectiveness coefficient (area of concrete/area of confined concrete)

f_l = lateral confinement pressure from transverse reinforcement

f_{co} = unconfined concrete compressive strength

Local Flange Buckling

The flange buckling is a phenomenon that largely depends on the width-to-thickness ratio, the boundary (i.e., restraint) conditions, and the material properties of the components comprising the section. The effect of the local flange buckling is the reduction of the ultimate strength of the section and/or the diminishing of its rotational capacity in the inelastic range. Additionally, the ductility of encased SCC members is adversely affected by local buckling, and this needs to be considered when estimating the rotational capacity.

To account for the local buckling of bare and encased steel sections, simple methods have been developed and can be readily utilized in a frame analysis software. One of the most popular methods has been developed by Ballio et al. (1987). In this approach, the cross section is divided into a finite number, i , of strips with each strip having an assigned area. If the

compressive strain in any strip exceeds the critical strain, ϵ_{cr} , the area of the strip reduces to zero for the subsequent load cases. In order to extend the applicability of this method beyond the elastic limit into the inelastic range, the elastic critical stress is divided by the yield strain, as it is represented in the following relationship:

$$\frac{\epsilon_{cr}}{\epsilon_y} = \frac{k \pi^2 E}{12 \sigma_y (1 - \nu^2) \left(\frac{\lambda_i}{t}\right)^2} \quad (49)$$

where

λ_i = distance between the centroid of the strip and the plate connection

t = thickness of the plate

k = confinement factor

E = Modulus of elasticity of steel

ν = Poisson's ratio

Geometric Nonlinear P- δ Effect

Geometric nonlinearities can be classified in two categories. The first category is related to the global geometric nonlinearities, usually referred to as P- δ effects. The second category is related to local geometric nonlinearities (i.e., local buckling), which are generally neglected in frame analysis (while they are carefully considered in advanced finite element analyses with discretized models). The global geometric nonlinearities can be incorporated in the models following basic procedures used in nonlinear frame analysis. One of these procedures employs the updated Lagrangian formulation in order to account for geometric nonlinearities such as large displacements and rotations.

In most of the analyses of multi-storey structures subjected to earthquake excitation, the effects of the combination of gravitational forces and lateral displacement are ignored. Such effects are often referred as second-order effects. The reason behind overlooking the second-order effects can be explained by the fact that traditionally, in low-rising reinforced concrete structures

(i.e., structures with low natural period and small lateral displacement response) subjected to earthquake excitation, the second-order effects are insignificant and therefore neglected. As steel structures become taller nowadays, the P-Delta effects are amplified due to the corresponding increase of lateral displacement.

The effect of axial force acting through the relative transverse displacement of the member ends known as P- δ effect can be taken into account in the modeling by using the geometric stiffness matrix, $[K_g]$, as:

$$[K_g] = \begin{bmatrix} [K_s] & -[K_s]^T \\ -[K_s]^T & [K_s] \end{bmatrix} \quad (50)$$

where

$$[K_s] = \begin{bmatrix} 0 & \alpha & -b & 0 & 0 & 0 \\ \alpha & c & 0 & 0 & 0 & 0 \\ -b & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (51)$$

and

$$\alpha = \frac{M_{zA} + M_{zB}}{L^2} \quad b = \frac{M_{yA} + M_{yB}}{L^2} \quad c = \frac{P}{L} \quad (52)$$

M_{zA} , M_{zB} , M_{yA} , M_{yB} = end moments with respect to z and y axes, respectively

P = axial force

L = length of the element

The tangent stiffness matrix of a beam-to-column element is then obtained by the following relationship:

$$[K]_{12 \times 12} = [T]_{6 \times 12}^T \times [K_e]_{6 \times 6} [T]_{6 \times 12} + [K_g]_{12 \times 12} \quad (53)$$

where

$$[T]_{6 \times 12} = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1/L & 0 & 1 & 0 & 0 & 0 & 1/L & 0 & 0 & 0 \\ -b & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/L & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1/L & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1/L & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \end{bmatrix} \quad (54)$$

and

$$\begin{aligned} [K_e]_{6 \times 6} &= \text{element stiffness matrix given by } \{\Delta F\} \\ &= [K_e] \{\Delta d\} \end{aligned}$$

The reader is referred to Thai and Kim (2011) for details on how to derive the element stiffness matrix.

Constitutive Models

Any of the constitutive models presented in the previous sections for steel, concrete, and steel reinforcement can be applied for modeling the material characteristics of the concrete-encased steel sections of the SCC frames.

Summary

The validity of the above case studies has been verified by comparing the numerical predictions with experimental data obtained from a wide range of structural systems subjected to static and hysteretic loadings. Full details of these comparative studies are presented in the literature. However, such constitutive laws and models are usually dependent on parameters which are evaluated through the particular use of the experimental data, and it is in the designers' discretion to choose and interpret these when data are used for specific purposes. In this chapter, it has been attempted to generalize the constitutive models for a number of applications. On the other hand, the lack of generality and objectivity that characterizes most FE packages can only be balanced through the use of material models which are compatible with valid experimental information. In fact, the work presented in

this chapter is considered as a step towards these directions.

The modeling of SCC members in this chapter (i) serves primarily the computation of the response of such members when they are subjected to seismic actions and (ii) acts as a vehicle for carrying out the state determination of the section (or integration point) to a frame element and ultimately to the whole frame assembly. The outcome of the former application is typically the moment–curvature response under a constant axial load. The latter application typically returns section forces that correspond to given section deformations (in uniaxial bending axial strain, and curvature).

In earthquake engineering, the stiffness of the column members is one of the most important parameters of the entire structural systems since it governs the lateral resistance of the frame. The natural period of vibration of the frame decreases with increasing stiffness but also increases with increasing mass. Therefore, the members comprising the frame need to be accurately simulated in order to derive the stiffness and mass matrices to be accounted for the frame analysis. According to the typical response spectra, the acceleration response of a SCC frame reduces with increased natural period. This implies that a composite structure will have to resist lower base shear; therefore, the earthquake effects will be less significant. The displacement response of the structure also increases proportionally with increasing natural period. For the case of increased lateral displacements, second-order effects ($P-\delta$ effects) could be developed which will determine the design and amplify the demand on the structure. The engineer is required to reduce the influence of the second-

order effects by controlling the lateral displacement of the frame providing ductility in the beams, columns, and connections. As a result, the accurate analysis of each individual component plays an important role in the aseismic design. In the holistic frame assembly, global geometric nonlinearities can be incorporated in the models following basic procedures used in the nonlinear frame analysis (i.e., modification of the stiffness matrix).

Using nonlinear static analysis, the engineer can obtain information on the global ductility and strength of the structure through force–displacement relations. At each point on the force–displacement curve, the engineer can check the member behavior and see whether the limit states are fulfilled. Weak areas and progressive hinge formation on the structural frame are revealed during the analysis.

Cross-References

- [Equivalent Static Analysis of Structures Subjected to Seismic Actions](#)
- [Nonlinear Analysis and Collapse Simulation Using Serial Computation](#)
- [Nonlinear Dynamic Seismic Analysis](#)
- [Nonlinear Finite Element Analysis](#)
- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- [Response Spectrum Analysis of Structures Subjected to Seismic Actions](#)
- [Seismic Analysis of Steel and Composite Bridges: Numerical Modeling](#)
- [Seismic Analysis of Steel Buildings: Numerical Modeling](#)
- [Steel Structures](#)
- [Seismic Vulnerability Assessment: Reinforced Concrete Structures](#)

References

Aribert JM, Al Bitar A (1989) Optimisation du dimensionnement en connection partielle de poutres

de planchers mixtes réalisés avec un bac de tôle mince nervurée (Optimization of the dimensioning in the case of partial connection of beams of composite floors carried out with a thin ribbed sheet vat). *Construction Métallique* 4 (1989) 12, pp 4–33

Aribert J-M, Lachal A (2000) Cyclic behaviour of the shear connection component in composite joints. In: *Behaviour of steel structures in seismic area. Proceedings of STESSA 2000 conference*, Montreal, Canada. Balkema, Rotterdam, pp 105–112

Balan TA, Filippou FC, Popov EP (1997) Constitutive model for 3D cyclic analysis of concrete structures. *J Eng Mech* 123(2):143–153

Ballio G, Calado L, Iori I, Mirabella Roberti G (1987) I problemi delle grandi costruzioni in zona sismica. Associazione Italiana Cemento Armato e Precompresso, Roma, pp 31–44

Braconi A, Salvatore W, Tremblay R, Bursi OS (2007) Behaviour and modeling of partial-strength beam-to-column composite joints for seismic applications. *Earthq Eng Struct Dyn* 36:142–161

Castro JM, Elghazouli AY, Izzuddin BA (2005) Modelling of the panel zone in steel and composite moment frames. *Eng Struct* 27(1):129–144

Cotsovos DM, Kotsovos MD (2011) Constitutive modelling of concrete behaviour. In: Papadrakakis M, Fragiadakis M, Lagaros ND (eds) *Computational methods in earthquake engineering*, Computational methods in applied sciences. Springer, Netherlands, pp 147–175

Fabbrocino G, Manfredi G, Cosenza E (2002) Modelling of continuous steel-concrete composite beams: computational aspects. *Comput Struct* 80(27–30): 2241–2251

Kim DK, Engelhardt MD (2005) Composite beam element for nonlinear seismic analysis of steel frames. *J Struct Eng* 131:715–724

Krawinkler H (1978) Shear in beam-column joints in seismic design of steel frames. *Eng J* 15(3):82–91

Lee SJ (1987) Seismic behavior of steel building structures with composite slabs. PhD thesis, Dept. of Civil Engineering, Lehigh Univ., Bethlehem

Menegotto M, Pinto PE (1973) Method of analysis for cyclically loaded R.C. plane elements including changes in geometry and non-elastic behaviour of elements under combined normal force and bending. In: IABSE, Report of the working commissions, Band 13, symposium on resistance and ultimate deformability of structures acted on by well defined repeated loads, Lisbon

Newmark NM, Siess CP, Viest IM (1951) Tests and analyses of composite beams with incomplete interaction. *Proc Soc Exp Stress Anal* 9(1):75–92

Ollgaard JG, Slutter RG, Fisher JW (1971) Shear strength of stud connectors in lightweight and normal-weight concrete. *AISC Eng J* 8(2):55–64

Park R, Paulay T (1975) Reinforced concrete structures. Wiley, New York

- Piluso V, Faella C, Rizzano G (2001) Ultimate behavior of bolted T-stubs I: theoretical model. *J Struct Eng* 127(6):686–693
- Przemieniecki JS (1968) *Theory of matrix structural analysis*. McGraw-Hill, New York
- Stevens NJ, Uzumeri SM, Collins MP, Will GT (1991) Constitutive model for reinforced concrete finite element analysis. *ACI Struct J* 88(1):49–59
- Thai HT, Kim SE (2011) Nonlinear inelastic analysis of concrete-filled steel tubular frames. *J Construct Steel Res* 67:1797–1805
- Uchida N, Tohki H (1997) Design of high-rise building using round tubular steel composite columns. In: *IABSE international conference on composite construction conventional and innovative*, Innsbruck, 16–18 Sept 1997
- Zhao H, Yuan Y, Ye Y (2012) Simplified nonlinear simulation of steel-concrete composite beams. *J Construct Steel Res* 71:83–91
- Zona A, Barbato M, Conte JP (2008) Nonlinear seismic response analysis of steel-concrete composite frames. *ASCE J Struct Eng* 123(6):986

Seismic Analysis of Wind Energy Converters

Giuseppe Failla

Dipartimento di Ingegneria Civile, dell'Energia, dell'Ambiente e dei Materiali, (DICEAM), University of Reggio Calabria, Reggio Calabria, Italy

Synonyms

Aerodynamic damping; Response spectrum; Seismic risk; Soil structure interaction; Wind energy converter

Introduction

Wind power is one of the fastest-growing renewable energy segments on a percentage basis. In 2013, over 35 GW of new wind capacity was installed all over the world, bringing the total wind capacity to 318 GW at the end of 2013 (GWEC 2014). The total installed wind capacity is expected to reach 365 GW by the end of 2014,



Seismic Analysis of Wind Energy Converters, Fig. 1 Horizontal axis wind turbine (HAWT) in Egel, Germany (photo by Hadhuey)

enough to provide about 4 % of the global electricity demand (GWEC 2014).

While many different design solutions have been considered in the early stages, for commercial use the modern wind industry has now stabilized on horizontal axis wind turbines (HAWT). A typical example is shown in Fig. 1: a land-based tower with a nacelle mounted on the top, containing the generator, a gearbox, and the rotor. Typically, three-bladed upwind rotors are used.

With the continuous increase of wind power production, the search for optimal design is facing new and challenging tasks. The design of land-based HAWTs has been traditionally driven by high wind speed conditions. However, following the introduction of new technologies such as variable pitch and active control in larger, lighter, and cost-effective HAWTs, in some cases the design-driving considerations have been changed, with fatigue and turbulence being considered in addition to high wind speed conditions.

For these lighter HAWTs, especially when installed in seismically active areas, a question has soon arisen as to whether seismic loads shall be considered among design loads. On the other hand, the need to investigate the potential importance of seismic loads has been corroborated by the damage that occurred to land-based HAWTs, following the 1986 North Palm Springs Earthquake, USA, and the 2011 Kashima City Earthquake, Japan. Post-earthquake surveys in the wind farms nearest the epicenter of North Palm Springs Earthquake documented that 48 out of 65 HAWTs were damaged, generally due to buckling in the walls of the supporting tower (photographs are available in the report by Swan and Hadjian (1988)). Earthquake-induced failure may occur also at the foundation level, as for the case of the footing of a HAWT in the Kashima wind farm (photographs are available in the paper by Umar and Ishihara (2012)). In this context, the seismic assessment of HAWTs has drawn an increasing attention in the last years, and as a result, seismic loading has been progressively included in International Standards (ISs) and Certification Guidelines (CGs) (DNV/Risø 2002; GL 2010; IEC 2005; AWEA 2011).

The key points in the seismic assessment of HAWTs can be briefly summarized as:

- Selection of the load combinations
- Use of a specific analysis method
- Definition of the structural model

On these points, sufficient information is generally available in existing ISs and CGs. However, because a certain flexibility is allowed, especially in the definition of the structural model and the selection of an appropriate analysis method, it is important that engineers be aware of the potential options available and how they may affect the reliability of the results. In an attempt to respond to these needs, this entry will provide, first, a preliminary introduction to the relevant issues involved in the seismic assessment of HAWTs. Hence, detailed prescriptions of existing ISs and CGs will be reported and, finally, examples of the possible options for the implementation of the seismic assessment will be

presented. Throughout the entry, land-based HAWTs will be referred to.

Seismic Assessment of HAWTs

Load Combinations

The selection of appropriate load combinations for seismic assessment is a relevant issue addressed by ISs and CGs. In general, they are recommended based on the observations that follow.

At sites with a significant seismic hazard, there is a reasonable likelihood that an earthquake occurs while the HAWT is in an operational state, i.e., while the rotor is spinning; in this case, the HAWT is subjected to simultaneous earthquake loads and operational wind loads. It shall be considered, also, the possibility that the earthquake triggers a shutdown and that, as a result, the HAWT is subjected to simultaneous earthquake loads and emergency stop loads. Another possible scenario is that the earthquake strikes when the turbine is parked, i.e., not operating due to wind speeds exceeding the cutoff wind speed of the turbine; specifically, blades may be locked against motion (fixed pitch turbines) or feathered such that no sufficient torque is generated for the rotor to spin (active pitch turbines). In recognition of these observations, the load combinations generally suggested by ISs and CGs for the seismic assessment of HAWTs are:

- Earthquake loads and operational wind loads
- Earthquake loads and emergency stop loads
- Earthquake loads and wind loads in a parked state

Both earthquake loads and wind loads are stochastic processes. The wind process is generally treated as a stationary process. Samples can be generated from well-established power spectral densities (PSDs) in the literature (e.g., Von Karman PSD or Kaimal PSD, see Manwell et al. (2010)), with parameters to be set depending on site conditions. Wind acts on the blades of the rotor and along the tower.

Obviously, wind loading on the blades varies significantly depending on whether the rotor is spinning or not; to generate wind loading on a spinning rotor, concepts of classical aerodynamics are used, for instance, those of Blade Element Momentum (BEM) theory and subsequent modifications (Manwell et al. 2010). The earthquake process is inherently nonstationary. Spectrum-compatible samples may be synthesized from site-dependent response spectra or site-specific historical records may be used, according to the prescriptions of the adopted ISs and CGs.

Structural Analysis Method

The computation of the HAWT response to the different load combinations is a crucial step of the seismic assessment. In general, two approaches can be pursued:

- A fully coupled time-domain simulation computing the response to simultaneously acting wind loading and seismic loading
- A decoupled analysis where the responses to wind loading and seismic loading are computed separately and then superposed

A fully coupled time-domain simulation is the most desirable approach. The reason is that it allows the actual wind loads on the blades to be evaluated correctly, taking into account that the oscillations of the tower top, induced by the earthquake ground motion, affect the rotor aerodynamics (in particular, the relative wind speed at the blades, depending on which lift and drag forces are calculated). However, for the implementation of fully coupled time-domain simulations, dedicated software packages are required, capable of solving the nonlinear motion equations of the structural system under simultaneous wind and seismic excitations.

When performing a decoupled analysis, instead, the responses to wind loading and seismic loading are built separately. This means that wind loads are evaluated as no earthquake ground motion is acting at the tower base. Correspondingly, the response to the earthquake ground motion is computed as no wind loading is acting on the rotor. It is evident that this approach is

approximate since, as explained earlier, the actual wind loads on the blades depend on the oscillations of the tower top, to which contributes also the earthquake ground motion at the tower base. Nevertheless, numerical comparisons with benchmark results obtained by fully coupled time-domain simulations have shown that decoupled analyses can yield accurate results, provided that the separate response to earthquake loading is computed using an appropriate level of damping. In particular, it has been found that:

- A percentage equal to 5 % of critical damping is appropriate when the separate response to earthquake loading is to be combined with the response to operational wind loading, i.e., for the load combination = earthquake loads + operational wind loads (Witcher 2005; Prowell and Veers 2009; Prowell 2011).
- In contrast, a percentage varying between 0.5 % and 2 % of critical damping is appropriate when the separate response to earthquake loading is to be combined with the response to wind loading in a parked state, i.e., for the load combination = earthquake loads + wind loads in a parked state (Prowell and Veers 2009; Prowell 2011; Stamatopoulos 2013).

Such a variability of the damping level, to be considered in a decoupled analysis, can be explained as follows:

- The low damping ratios (0.5–2 %), which are required for the load combination = earthquake loads + wind loads in a parked state, are motivated by the fact that when the turbine is parked, the only damping source is the structural damping of the tower that, as customary in steel structures, is generally low.
- The higher damping ratio (5 %), which is required for the load combination = earthquake loads + operational wind loads, reflects the fact that the earthquake loading significantly affects the aerodynamics of a spinning rotor. In particular, a motion of the tower top due to the earthquake loading, against or in the wind direction, causes, respectively, an increase or a decrease of the instantaneous

thrust force, with respect to that computed assuming no earthquake ground motion at the tower base (such increase or decrease of the instantaneous thrust force mirrors an increased or decreased relative wind speed at the blades). Since, in both situations, this alternation of the thrust force is oriented opposite to the tower top motion, it can be understood that its effects are to be modeled by introducing additional damping with respect to the structural damping, when computing the separate response of the HAWT to earthquake loading. The same observation holds true when a decoupled analysis is performed to compute the response of HAWTs to combined wind loading and wave loading (Kuhn 2001).

The difference between damping for the parked state and operational state is generally referred to, in the literature, as *aerodynamic damping*, to mean that its source is essentially the rotor aerodynamics.

Decoupled analyses may be performed in time and frequency domain. Especially frequency-domain formulations have been awarded a considerable attention, because in this case the separate response to earthquake loading can be built by coded response spectra, a concept most engineers are familiar with. However, in light of the earlier observations on the appropriate level of damping, it is evident that particular care shall be taken when following this approach. In fact, the typical 5 % damped response spectra for building structures (ICC 2012) will be suitable only for the load combination = earthquake loads + operational wind loads, while 0.5–2 % damped response spectra shall be used for the load combination = earthquake loads + wind loads in a parked state. It is evident that selecting the correct damping ratio is crucial: if the response to earthquake loading was obtained from 5 % damped response spectra, nonconservative results would be certainly obtained when the turbine is parked. Some ISs and CGs take account of these issues (IEC 2005; AWEA 2011), but no explicit indications on the damping ratio to be adopted are given in the others (DNV/Risø 2002; GL 2010).

Structural Model

A fundamental assumption of existing ISs and CGs, with regard to the structural model, is material linearity. This assumption is essentially justified by the fact that the primary intent is to ensure power production for the design life of the HAWT, usually 20 years, and that nonlinear deformation (damage) to the turbine would interrupt reliable operation. Material linearity means low operational stresses, and this provides some safety margins against failure (Bazeos et al. 2002). Therefore material linearity will be, in general, a prerequisite of ISs and CGs also when assessing the response to seismic excitations.

Starting from the assumption of material linearity, in general, two types of structural modeling are feasible:

- Simplified models which model the tower and consider the rotor-nacelle assembly (RNA) as a lumped mass at the tower top
- Full system models which describe the whole turbine, including the nacelle and rotor with a certain level of detail

Simplified models are appealing since the complexities involved in modeling the rotor are avoided. Full system models include the rotor blades and, in general, turbine components such as power transmission inside the nacelle and pitch and speed control devices, with a different degree of accuracy depending on the specific modeling adopted, for instance, a finite element (FE) or a rigid multi-body modeling.

Simplified or full system models can be used depending on the selected structural analysis method. In particular:

- Fully coupled time-domain simulations involve only full system models as they require modeling the rotor aerodynamics, with the earthquake ground motion simultaneously acting at the tower base.
- Decoupled analyses may be implemented using either a full system model or a simplified model. If a simplified model is adopted, seismic loads are built considering the mass of

RNA lumped at the tower top, while wind loads are obtained by a dedicated software package, capable of modeling the rotor aerodynamics, with no earthquake ground motion at the tower base, since the analysis is decoupled.

International Standards and Certification Guidelines

Guidance for seismic loading on HAWTs can be found in the following ISs and CGs:

- DNV/Risø: Guidelines for design of wind turbines (DNV/Risø 2002). Released by Det Norske Veritas (DNV) and Risø National Laboratory
- GL 2010: Guideline for the certification of wind turbines (GL 2010). Released by Germanischer Lloyd (GL)
- IEC 61400-1: Wind turbine generator systems. Part 1: Safety requirements (IEC 2005). Released by International Electrotechnical Commission (IEC)
- ASCE/AWEA RP2011: Recommended practice for compliance of large land-based wind turbine support structures (ASCE/AWEA 2011). Released by American Society of Civil Engineers (ASCE) and American Wind Energy Association (AWEA)

DNV/Risø Guidelines

DNV/Risø Guidelines are meant to provide a basic introduction to the most relevant subjects in wind turbine engineering (DNV/Risø 2002). Consistently with this general purpose, quite general suggestions are given to deal with seismic loading.

It is prescribed that earthquake effects should be considered for HAWTs located in areas that are considered seismically active based on previous records of earthquake activity (Section 3.2.8). For those areas known to be seismically active but with no sufficient information available for a detailed characterization of seismicity, an evaluation of the regional and local geology is recommended to determine the

location of the HAWT relative to the alignment of faults, the epicentral and focal distances, the source mechanism for energy release, and the source-to-site attenuation characteristics. In this case, the evaluation should aim to estimate both the design earthquake and the maximum expectable earthquake, taking into account also the potential influence of local soil conditions on the ground motion.

No specific recommendations are given on the earthquake-wind load combinations to be considered. However, since it is prescribed that in seismically active areas the HAWT should be designed so as to withstand earthquake loads, it is implicit that the three, typical load combinations described earlier (i.e., earthquake loads and operational wind loads, earthquake loads and emergency stop loads, earthquake loads occurring in a parked state) shall be referred to.

As for what concerns the method of analysis, DNV/Risø provides explicit suggestions only for the response spectrum method, as used in a decoupled analysis. In particular, the use of a single degree-of-freedom (SDOF) system with a lumped mass on top of a vertical rod is suggested, with the rod length equal to the tower height and the lumped mass including the mass of the rotor-nacelle assembly (RNA) and $\frac{1}{4}$ of the mass of the tower. It is prescribed that the fundamental period of the SDOF system is used in conjunction with a design acceleration response spectrum to determine the loads set up by the ground motion, by analogy with the simplified procedures used in building codes. Analyses shall be performed for horizontal and vertical earthquake-induced accelerations. However, no explicit recommendations are given on the criterion to translate the resulting spectral response acceleration into design seismic loads, as well as on the damping ratio to be used. Since in the absence of specific guidance on this matter, a most intuitive choice of engineers could be using the typical procedures of the International Building Code (ICC 2012), it has to be remarked that the 5 % damping ratio, embedded in the standard design response spectrum, is appropriate only for seismic loading acting during an operational state, but overestimates considerably the

actual damping in a parked state. This aspect should be well kept in mind when referring to DNV/Risø for seismic assessment of HAWTs.

Regarding the structural model, attention is drawn to the need of including the actual stiffness of the structural component of the foundation and an appropriate model of the supporting or surrounding soil, the latter through a proper soil structure interaction (SSI) modeling (Section 8.4). Although, for this purpose, nonlinear and frequency-dependent models are recommended in principle, appropriate linearized models are allowed, depending on the expected strain level in the soil (typically, it may be up to 10^{-1} for earthquake loading and considerably larger than for other loading conditions). The linearized models consist of translational and rotational springs for circular footings and piles.

GL Guidelines

GL 2010 guidelines aim to set a number of requirements for the certification of wind turbines (GL 2010). For this reason, they are quite prescriptive and provide detailed information on some particular aspects of seismic risk.

In agreement with DNV/Risø, GL 2010 prescribes that seismic loading shall be taken into account in seismically active areas (Section 4.2.4.2.3). Earthquake loading is included in a group of design load cases (Table 4.3.2) classified as load cases accounting for “extended” design situations, including special applications and site conditions. These design load cases are not mandatory for certification purposes, but may be chosen for the verification of the HAWT to complement the applicability in specific design situations. The response to seismic loading is to be assessed both in the operational state and the parked state (Table 4.3.2) under normal wind loading. For the operational state it is also suggested to consider the activation of the emergency shutdown triggered by the earthquake. The safety factor for all the loads to be combined with seismic loading is equal to 1.0 (Section 4.3.5.4). A return period of 475 years is prescribed as the earthquake design level. To model the seismic loading, recommendations of the local building

code should be applied or, in the absence of locally applicable regulations, those of either Eurocode 8 (2004) or American Petroleum Institute (API 2000).

Regarding the method of analysis, GL 2010 specifies that fully coupled or decoupled analyses are possible, with at least three modes in both cases. Time-domain simulations shall be carried out considering at least six simulations per load case. As with DNV/Risø, no guidance is provided on the damping ratio to be adopted when using the design response spectrum in a decoupled analysis. Again, because of the lack of guidance on this matter, it shall be kept in mind that the 5 % damping ratio is appropriate only in the operational state and that lower damping ratios shall be considered in the parked state.

GL 2010 gives no particular prescriptions on the structural model to be adopted. However, because at least three modes have to be included in the vibration response, the use of a multi-degree-of-freedom (MDOF) structural model is implicitly suggested. In general, a linear elastic behavior shall be assumed. A ductile response can be considered only when the support structure has a sufficient static redundancy, such as a lattice tower. However, if ductile behavior is assumed, the structure shall be mandatorily inspected after occurrence of an earthquake.

IEC Standards

IEC 61400-1 Standards aim to specify essential design requirements to ensure structural integrity of wind turbines (IEC 2005). They have the status of national standards in all European countries whose national electrotechnical committees are CENELEC members (CENELEC = European Committee for Electrotechnical Standardization).

IEC 61400-1 recommends that, in seismically active areas, the integrity of the HAWT is demonstrated for the specific site conditions (Section 11.6), while no seismic assessment is required for sites already excluded by the local building code, due to weak seismic actions. The seismic loading shall be combined with other significant, frequently occurring operational

loads. In particular, IEC 61400-1 prescribes that the seismic loading shall be superposed with operational loads, to be selected as the higher of:

- (a) Loads during normal power production, by averaging over the lifetime
- (b) Loads during emergency shutdown, for a wind speed selected so that the loads prior to the shutdown are equal to those obtained with (a)

No explicit reference is made, however, to the loaded case of an earthquake loading striking in a parked state.

The safety factor for all load components to be combined with seismic loading shall be set equal to 1.0. The ground acceleration shall be evaluated for a 475-year recurrence period based on ground acceleration and response spectrum requirements as defined in local building codes. If a local building code is not available or does not provide ground acceleration and response spectrum, an appropriate evaluation of these parameters shall be carried out.

Regarding the method of analysis, fully coupled or decoupled analyses are possible (11.6). In time-domain analyses, sufficient simulations shall be undertaken to ensure that the operational load is statistically representative. It is prescribed that the number of tower modes used in either of the above methods shall be selected in accordance with a recognized building code. In the absence of a locally applicable building code, consecutive modes with a total modal mass of 85 % of the total mass shall be used.

IEC 61400-1 gives no particular indications on the structural model for seismic analysis. In agreement with GL 2010, however, it is implicit that the structure shall be modeled as a MDOF system, since the use of consecutive modes with a total modal mass equal to at least 85 % of the total mass is recommended. In general, the response should be linearly elastic, while a ductile response with energy dissipation is allowed only for specific structures, in particular for lattice structures with bolted joints.

Annex C of IEC 61400-1 presents a simplified, conservative method for the calculation of seismic loads. This procedure is meant to be used when the most significant seismic loads can reasonably be predicted on the tower, and shall not be used if it is likely that the earthquake ground motion may cause significant loading on the rotor blades or the structural components of the foundation. The principal simplifications in Annex C are ignoring the modes higher than the first tower bending mode and the assumption that the whole structure is subjected to the same acceleration. Upon evaluating or estimating the site and soil conditions required by the local building code, or adopting conservative assumptions while detailed site data are not available, the simplified method can be applied as follows:

- The acceleration at the first tower bending natural frequency is set using a normalized design response spectrum and a seismic hazard-zoning factor. For this, a 1 % damping ratio is assumed.
- Earthquake-induced shear and bending moments at the tower base are calculated by applying, at the tower top, a force equal to the total mass of the RNA + $\frac{1}{2}$ the mass of the tower times the design acceleration response.
- The corresponding base shear and bending moments are added to the characteristic loads calculated for an emergency stop at rated wind speed, i.e., the speed at which the limit of the generator output is reached.
- The results are compared with those obtained against the design loads or the design resistance for the HAWT. If the tower can sustain the resulting combined loading, no further investigation is needed. Otherwise, a thorough investigation shall be carried out on a MDOF structural model.

With regard to such a simplified method, described in Annex C, it shall be pointed out that ignoring the second tower mode is a significant nonconservative simplification (e.g., see Zhao and Maisser (2006) on the role of the second tower mode in the seismic response of HAWTs).

This is somehow compensated for by incorporating $\frac{1}{2}$ of the tower mass with the tower head mass and prescribing superposition with the characteristic loads calculated for an emergency stop at rated wind speed, which represent quite conservative aerodynamic loads.

ASCE/AWEA Recommended Practice

The general purpose of ASCE/AWEA RP2011 is to clearly identify specific US national recommendations for wind turbine design, which are compatible with IEC 61400-1 but may provide proper recommendations for those cases in which US practice and IEC 61400-1 differ. As for what concerns seismic assessment, ASCE/AWEA RP2011 makes a quite comprehensive effort to harmonize some relevant specific prescriptions of certification agencies with the traditional perspectives of US standards ASCE/SEI 7-05, which sets the minimum design loads for buildings and structures in general (ASCE 2006). For the level of detailed information provided, ASCE/AWEA RP2011 can be considered a very useful and comprehensive reference tool for the seismic assessment of HAWTs.

ASCE/AWEA RP2011 points out that although standard HAWT classes shall be generally designed for normal wind conditions, extreme wind, conditions and other environmental conditions including temperature and air density, specific prescriptions on the criteria for the design of HAWTs subjected to earthquake ground motions are necessary, in recognition of the fact that earthquake events are common in many US jurisdictions. According to ASCE/AWEA RP2011, it is of critical importance to recognize that seismic loads plus operational loads may in some cases govern tower and foundation design. For these reasons, load combinations involving earthquake occurring in an operational state and earthquake triggering emergency stop loads, an earthquake occurring in a parked state should be considered (Section 5.4.4). Seismic ground motion values should be determined based on the acceleration response spectrum or site-specific ground motion procedures as prescribed by ASCE/SEI 7-05 (see Section 11.4 and Chapter 21 in ASCE/SEI 7-05).

Unlike the alternative ISs and CGs, ASCE/AWEA RP2011 provides quite detailed prescriptions on a “best practice” load combination including seismic loads plus operational loads:

$$U = (1.2 + 0.2S_{DS})D + 0.75(\rho Q_E + 1.0M) \quad (1)$$

$$U = (0.9 - 0.2S_{DS})D + 0.75(\rho Q_E + 1.0M) \quad (2)$$

where:

U = factored load effect

D = dead load

M = operational loading equal to the greater of (1) loads during normal power production at the rated wind speed or (2) characteristic loads calculated for an emergency stop at rated wind speed

Q_E = effect of horizontal seismic (earthquake-induced) forces

S_{DS} = design spectral response acceleration parameter at short periods

$\rho = 1.0$, redundancy factor (for nonbuilding structures not similar to buildings $\rho = 1.0$, according to Chapter 12.3.4.1 of ASCE/SEI 7-05)

ASCE/AWEA RP2011 suggests Equations 15.4-1 and 15.4-2 of ASCE/SEI 7-05 for the seismic response coefficient C_s (nonbuilding structures), if Equation 12.8-1 of ASCE/SEI 7-05 is used to compute the seismic base shear. Specifically, in Eq. 15.4-2 of ASCE/SEI 7-05 for C_s a response modification factor $R = 1.5$ is recommended by ASCE/AWEA RP2011. The use of $R = 1.5$ does not necessarily imply that a ductile response or material overstrength is expected but accounts for a certain conservatism in the seismic response coefficient C_s prescribed for nonbuilding structures.

In Eqs. 1 and 2, the use of a load factor of 0.75 on both seismic loads and operational loads is supported by results of time-domain analyses on HAWTs ranging from 65 to 5 MW, subjected to 100 earthquake ground motion records, for varying orientation of wind and earthquake loads (Prowell 2011). It is observed that when the seismic hazard at a particular site is dominated by

known faults, consideration of site-specific prevailing wind direction and maximum earthquake component direction may be appropriate. In this case no load factor may be applicable, if wind and wave propagation directions are expected to coincide.

ASCE/AWEA RP2011 recommends that for load combinations not including operational loads, the spectral response acceleration parameter should be based on a 1 % damping ratio, which reflects the low inherent damping of typical steel support structures for HAWTs. The multiplicative spectral adjustment factor, B , to adjust spectral response acceleration, S_a , from 5 % (standard IBC value for determining S_a) to 1 % damped values is equal to 1.40 (Table 5-6). For load combinations that include operational loads, the spectral response acceleration parameter should be based on 5 % damped values. ASCE/AWEA RP2011 points out that this increase in damping is based on the aerodynamic damping inherent to an operating HAWT as verified by experimental and numerical results showing that a damping level of 1 % produces overly conservative results (Prowell 2011).

Regarding the method of analysis, according to ASCE/AWEA RP2011 a fully coupled time-domain analysis and decoupled analyses based on equivalent lateral force method or modal response spectrum method are acceptable, as permitted by the local building code. For the specific implementation of each method of analysis, the local building code or ASCE/SEI 7-05 is referred to. In particular, if the equivalent lateral force procedure is used, the vertical distribution of seismic forces should be calculated based on the procedure given in ASCE/SEI 7-05, Chapter 12.8.3, with some modifications: the seismic forces corresponding to the seismic weight of the RNA should be located at the turbine's center of gravity, and those corresponding to the seismic weight of the tower structure (including ladders, platforms, railings, etc.) should be distributed to nodes distributed along the tower height. No further prescriptions are given on a specific structural model to be adopted, when implementing a fully coupled or a decoupled analysis.

As for what concerns decoupled analyses, ASCE/AWEA RP2011 suggests that in those cases when only the peak seismic loads and peak operational loads are available, the proposed combination method for seismic loads and operational loads may be overly conservative, especially in recognition of the fact that the respective peak loads do not occur at the same instant of time and in the same loading direction. Therefore, to reduce potential design conservatism and obtain a more accurate prediction of the response, fully coupled time-domain analyses are suggested, considering earthquake ground acceleration in combination with operational or emergency stop loads. Seismic analysis should comply with the requirements of ASCE/SEI 7-05, Chapter 16, concerning, for instance, the minimum number of simulated earthquake ground motions. It is recommended that time-domain analyses be conducted with analysis software capable of simulating the structural response and global turbine dynamics, including the aerodynamic interaction.

ASCE/AWEA RP2011 gives some interesting points of view on the applicability of typical prescriptions of building codes to HAWTs design. For instance, according to ASCE/AWEA RP2011, enhanced performance objectives may be established to meet specific owner requirements and to improve expected behavior during and after an earthquake. For this purpose the use of a performance factor, similar to an importance factor of 1.5 for essential facilities, is suggested. This performance factor shall be agreed with the wind turbine manufacturer to establish acceleration thresholds for turbine components that will ensure operational performance. Also, according to ASCE/AWEA RP2011, no specific drifts or displacement need to be defined. This is motivated essentially by the fact that thorough analysis and design considerations of the ultimate and fatigue limit states implicitly limit the displacements of the tower. In addition, the wind turbine controls monitor and limit the possible tower top accelerations, to prevent exceeding the design loading.

Another distinctive feature of the ASCE/AWEA RP2011 is the emphasis put on

consideration of seismic forces in the foundation design, for areas with historical earthquake activity. Evaluation of earthquake effects should be performed in accordance with the requirements of the local building code or IEC 61400-1. In any case, geotechnical evaluation of earthquake effects should include ground shaking, liquefaction, slope instability, surface fault rupture, seismically induced settlement/cyclic densification, lateral spreading, cyclic mobility, and soil strength loss. In areas susceptible to earthquake effects, appropriate mitigation should be provided for foundations. For projects located near active faults, the characteristics of the fault including type, seismic setting, subsurface conditions, ground motion attenuation, and maximum earthquake magnitude should be considered. At any rate, HAWTs should be located with adequate setbacks from fault zones. Where relatively loose unsaturated cohesionless soils are present at the project site, the effect of ground shaking from a design level earthquake should be taken into account. Also, potential settlement due to cyclic densification of the site soils should be evaluated.

Implementation of Seismic Assessment

Despite the prescriptions given by existing ISs and CGs (DNV/Risø 2002; GL 2010; IEC 2005; ASCE/AWEA 2011), engineers dealing with the seismic assessment of HAWTs may face a few issues, which are only partially addressed by ISs and CGs. This section is meant to provide some insights into these aspects, illustrating the most relevant studies for this purpose. Although investigations are not fully accomplished, it is worth mentioning these studies, as they may provide engineers with very useful data for a correct seismic assessment of HAWTs.

As mentioned earlier, an important step is the construction of the structural model. Simplified models, which avoid the complexities involved in modeling the rotor, are frequently used, especially for a preliminary design. In contrast, for a comprehensive investigation of all factors relevant to seismic risk, full system models are generally considered. They allow prediction of

component loads instead of only tower loads, which cannot be estimated in a simple tower-based model. In this regard, it is remarked that higher modes involving the rotor dynamics may play an important role, as they may fall in the region of maximum spectral response acceleration (Prowell et al. 2010).

Regardless of the structural model adopted, another important issue in the seismic assessment of HAWTs is SSI modeling. Modeling the base as fixed, with no consideration of the SSI, could be justified in the case of overdesigned foundations and stiff soil conditions. However, HAWTs may be installed on relatively soft soils or loose soils containing alluvial deposits, and under these circumstances the SSI modeling, particularly for dynamic loads, could become a major concern in the design of the foundation and, consequently, of the entire support structure of the HAWT (Bazeos et al. 2002; Zhao and Maisser 2006). A proper SSI modeling may play an important role also in consideration of the ground motion amplification effects on soft and loose soils.

In the following, simple and full system models will be described briefly as used in the recent literature, along with relevant information on SSI modeling.

Simple Models

One of the first studies on the seismic response of HAWTs has been carried out by Bazeos and coworkers (Bazeos et al. 2002). They have investigated a 38 m high HAWT resting on a concrete block, located in a site with 0.12 g peak ground acceleration and semi-rock soil conditions. Seismic analyses have been conducted on two different models, a FE model of the tower with shell elements and a simplified FE model with lumped masses along the tower height and 3D beam elements approximately mapping mechanical and geometrical properties of the tower. In both models, a top mass has been added to model the RNA; SSI has been modeled by a set of discrete springs and dashpots and, adding to the mass of the concrete foundation block (modeled as rigid), a virtual soil mass moving in phase (Mulliken and Karabalis 1998). Parked conditions only have been considered, with no aerodynamic loads

along the tower. Seismic analysis has been carried out in the time domain using ground motions compatible with the elastic response spectrum, as prescribed by the Greek Aseismic Code with 0.5 % damping. For the relatively low ground acceleration under consideration (0.12 g), low stress levels have been found due to seismic excitations, with respect to the stress levels due to wind in either operational or emergency states. Results obtained by the time-domain simulations on the two FE models have been validated by a response spectrum analysis on a SDOF system with a mass set equal to the total mass of the system and a stiffness computed from the first natural period of the FE model with shell elements. Interestingly, Bazeos and coworkers (2002) have showed that higher tower modes can be significantly affected by SSI modeling. This result is important in consideration of the fact that the natural frequencies of the higher tower modes may fall within the region of maximum spectral acceleration (Haenler et al. 2006).

Umar and Ishihara (2012) have focused on the construction of a response spectrum for HAWTs under seismic excitations only, i.e., in a parked state. The need for a specific response spectrum is motivated by the observation that the support structures for HAWTs exhibit, unlike buildings, long period, heavy top, and different mass distribution along the height. Besides this general observation, by carrying out numerical simulations using a database of strong earthquake ground motions, they have shown that the very low damping levels in parked conditions determine excessive fluctuations in the response spectrum, and such uncertainty cannot be captured by existing damping correction factors in Eurocode 8 (2004) and Japanese Building Standard Law (BSL 2004). Umar and Ishihara (2012) have modeled the HAWT as a MDOF system with a lumped mass at the top, and a sway-rocking model to take into account SSI effects. They have proposed a modified correction factor for the damping ratio of the BSL response spectrum used in Japan (BSL 2004), depending on the natural period and the targeted reliability. They have shown that the maximum seismic loads, as obtained by a complete quadratic combination of

five modal responses obtained by the specified design spectrum, match very well the corresponding values obtained with time series analyses. Results have been provided for HAWTs with different size.

Stamatopoulos (2013) has addressed the response of HAWTs to near-fault ground motions. He has investigated a 53.95 m tall turbine resting on a circular footing. A FE model of the tower and the circular footing has been used, with a lumped mass at the tower top modeling the RNA; SSI has been modeled by uncoupled nonlinear springs distributed below the footing. The response to near-fault ground motion has been investigated by three methods: a response spectrum method based on the elastic acceleration spectrum provided by the Greek Aseismic Code, suitably increased by 25 % to account for proximity to a seismic fault; a response spectrum method involving an elastic local acceleration spectrum built based on actual records for the project site; and a time history analysis using synthesized ground motions compatible with the elastic local acceleration spectrum. The two response spectrum analyses have been carried out on a FE model with the tower grounded by a linearly elastic rotational spring, with stiffness computed as the ratio of the bottom bending moment to the bottom rotation. An iterative procedure has been implemented, since the bottom bending moment and rotation depend on the seismic loading computed from the spectral acceleration; the latter depends on the first natural period, which depends in turn on the stiffness of the grounded rotational spring. Stamatopoulos (2013) has shown that the acceleration spectrum provided by the Greek Aseismic Code, although appropriately increased to account for proximity of seismic fault, significantly underestimates the shear and bending moment demand at the tower base by 55 %.

Nuta and coworkers (2011) have proposed a methodology to assess the probability of failure of a HAWT under seismic excitations in a parked state. Despite material linearity is generally prescribed for design, in fact, nonlinear behavior has to be taken into account when examining the potential failure mechanisms induced by seismic

excitations. For a HAWT with 1.65 MW rated power and 80 m hub height, they have built log-normally distributed fragility curves to estimate the probability of reaching a defined damage state. Parameters of the lognormal distributions have been obtained from nonlinear incremental dynamic analyses, assuming the magnification factor with respect to the design earthquake as intensity measure, and peak displacement, peak rotation, residual displacement, and peak stress as damage measures. For the specific sites of this study, two in Canada and one in the USA, no significant probability of failure has been found for the 1.65 MW HAWT under consideration (Nuta et al. 2011).

Full Models

In the last few years, a significant effort has been spent on developing advanced tools that may allow a full system modeling of HAWTs. Due to computational complexity involved in a FE modeling of all components of a HAWT, models with a limited number of degrees of freedom have been built, and, in general, a modal approach, a multi-body approach, or a combination of the two has been used. Many high quality full-modeling software packages are now available for the wind industry, such as GH BLADED (Bossanyi 2000) and FAST (Jonkman and Buhl 2005), developed at the United States National Renewable Energy Laboratory (NREL). A few comments on these packages are in order, especially with regard to the options available for seismic analysis.

GH BLADED uses a multi-body dynamics approach in conjunction with a modal representation of the flexible components like tower and blades (Bossanyi 2000). A fully coupled time-domain simulation is feasible, with wind and seismic loadings simultaneously generated. Two methods are available for simulating seismic loading. The first method allows recorded acceleration time histories to be used, while the second method uses an iterative procedure to synthesize acceleration time histories providing an elastic response spectrum that closely matches a specified design response spectrum. SSI can be also modeled. GH BLADED has been validated by GL for calculating operational loads associated

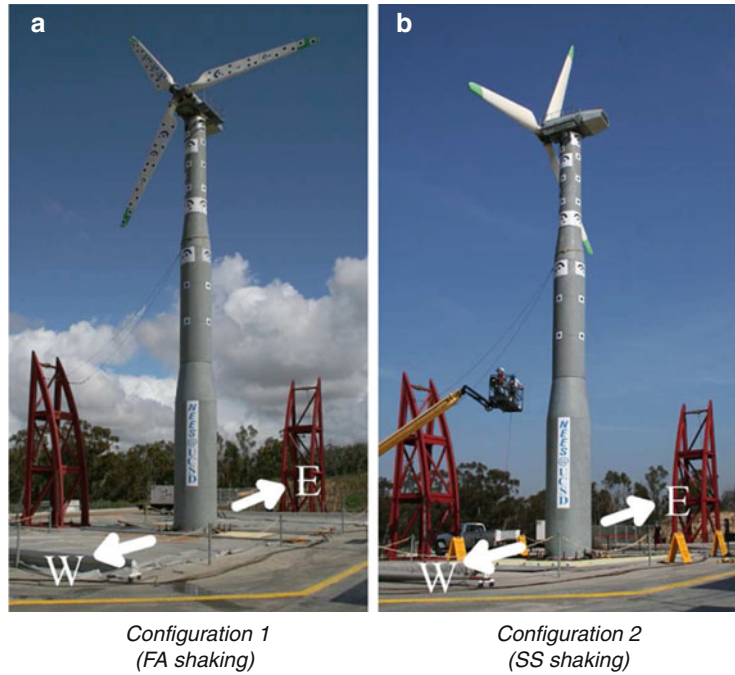
with typical load cases. A combined multi-body dynamics and modal formulation is adopted also by FAST (Jonkman and Buhl 2005), with flexible components modeled based on user-provided mode shapes. A fully coupled time-domain simulation can be implemented, with seismic loading generated as a user-defined loading imposed at the tower base. Like GH BLADED, FAST has been validated by GL.

Using BLADED, Witcher (2005) has compared the results from a response spectrum method and time-domain simulations as applied to a 60 m tall 2 MW turbine subjected to earthquake ground motion in both operational and parked cases. He has found that the elastic design spectrum with 5 % damping ratio yields a maximum bending moment at the tower base in a very good agreement with that computed by time-domain simulations, thus inferring that aerodynamic damping experienced by an operating turbine is quite close to 5 %. However, the results of the response spectrum method and time-domain analyses were very different in the parked case, with the first significantly underestimating the maximum tower base bending moment. This result has confirmed that in the parked case no aerodynamic damping is generated and that using the response spectrum method with a 5 % damping ratio does lead to nonconservative results. Although, in the specific case under examination, the bending moment demand due to earthquake loading in the parked case was lower than that due to earthquake loading in the operational case, Witcher (2005) has drawn the attention to the fact that, in some cases, the driving load can be that corresponding to earthquake loading in the parked case and has recommended further investigations on this issue.

Using FAST, one of the most comprehensive and fruitful studies on the seismic assessment of HAWTs has been carried out by Prowell and coworkers (2013). They have run a multiyear research program including extensive numerical simulations with FAST on HAWTs featuring different sizes and rated power (65 kW; 900 kW; 1.5 MW and 5 MW) and experimental tests on a HAWT (65 kW rated power, 22.6 m hub height, and a 16 m rotor diameter) mounted on the

Seismic Analysis of Wind Energy Converters,

Fig. 2 HAWT on the outdoor shake table at the University of San Diego, CA. The *arrows* indicate the direction of shaking (From Prowell et al. 2013)



outdoor shake table at the University of California, San Diego (Fig. 2). Prowell and coworkers have collected a considerable amount of data, which has also served as a basis for ASCE/AWEA RP2011 prescriptions. Using a set of 99 ground motions with different magnitude and source-to-recording distance, they have run numerical simulations showing that the considered earthquakes may produce, in the 5 MW HAWT, a bending moment demand at the tower base well above that from extreme wind events. This result has been found for parked, operational, and emergency shutdown simulations and confirmed that seismic loads may be design driving for large turbines in regions of high seismic hazard.

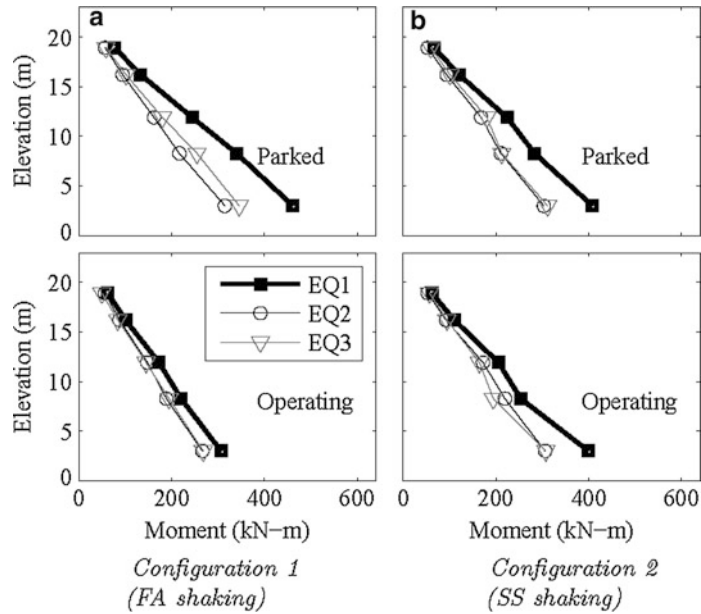
As for what concerns the experimental tests on the 65 kW HAWT, only a few key results are reported here, but interested readers can find detailed information on both numerical and experimental findings in the study by Prowell and Veers (2009), Prowell et al. (2010, 2013), Prowell (2011), and references therein. Experimental tests were carried out in operational and parked states. In each state, shaking has been imparted in two directions, one parallel (FA =

fore-aft) and another perpendicular (SS = side-to-side) to the rotation axis of the rotor. Structural response characteristics have been recorded for motions imparted in both configurations and both operational states (Prowell et al. 2013).

The results have shown that for shaking imparted in the SS direction, no appreciable differences are encountered between operational and parked states, in terms of bending moment envelopes (Fig. 3b). In contrast, for shaking imparted in the FA direction, the bending moment demand at the tower base in the operational state was reduced by approximately 15–33 % from that while parked (Fig. 3a). This reduction of demand confirmed that in the operational state, aerodynamic damping has to be accounted for when performing a decoupled analysis, with separately generated wind loading and seismic loading (Prowell et al. 2013). However, due to the influence of many factors such as wind speed, earthquake magnitude, wind and earthquake relative directions, and SSI modeling, multiple simulations including likely distributions of wind speed and earthquake shaking have been recommended for an accurate quantification of the aerodynamic damping (Prowell et al. 2013).

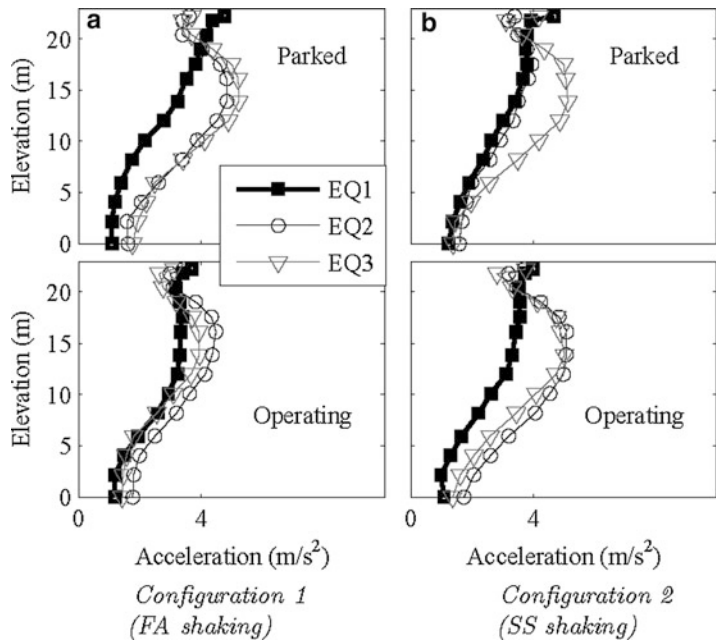
Seismic Analysis of Wind Energy Converters,

Fig. 3 Experimental bending moment envelope for three earthquake ground motions (From Prowell et al. 2013)



Seismic Analysis of Wind Energy Converters,

Fig. 4 Experimental acceleration envelope for three earthquake ground motions (From Prowell et al. 2013)



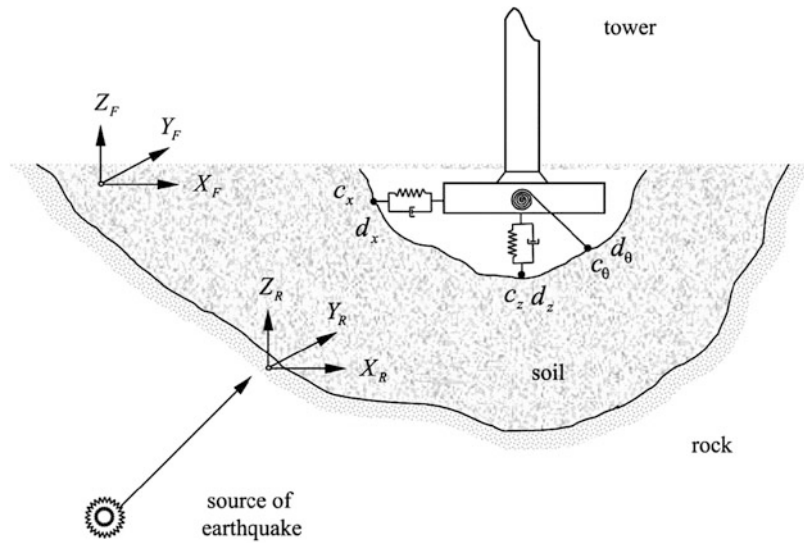
The same simulations should serve for a probabilistic description of the influence of seismic excitations on extreme loads.

A further important result obtained by Prowell and coworkers (2013) concerns the relative contributions of various tower modes. The maximum

absolute acceleration envelopes in the FA and SS direction have shown that, in addition to the first mode, the second mode contributes significantly, as indicated by the high acceleration values at two thirds of the tower height (Fig. 4). This result has confirmed what predicted by other

Seismic Analysis of Wind Energy Converters,

Fig. 5 SSI modeling for seismic assessment of HAWTs (From Zhao and Maisser 2006)



researchers (Haenler et al. 2006) on the importance of the second tower mode for estimation of seismic loads on large turbines. However, on the basis of the response PSDs, it has been found that amplification of energy imparted near the frequency of the second modes was not significantly influenced by operational state, thus implying that aerodynamic damping shall preferably be accounted for only in the first mode response (Prowell et al. 2013).

Besides the software packages used by the wind industry, there exist also a few full models of HAWTs, which have not been translated in software packages for the wind industry yet, but have provided very interesting results on the seismic response of HAWTs.

For instance, Zhao and coworkers (Zhao and Maisser 2006; Zhao et al. 2007) have developed a hybrid multi-body system (MBS) for full modeling of HAWTs. By this approach, the elastic tower is discretized into a series of rigid bodies coupled elastically by constraint joints and springs. The wind rotor is treated as a rigid disk; nacelle and machine carrier are coupled by an in-plane joint and treated as a rigid ensemble connected to the tower top through a revolute joint. The governing equations are derived using Lagrange's equations. This approach, though more mathematically rigorous, does not require external calculation of component mode shapes.

The MBS has been used by Zhao and Maisser (2006) to assess the seismic response of a 65 m high HAWT. SSI has been modeled approximately by a frequency-independent discrete parameter model, as a 3D set of uncoupled spring-damper devices, including translations and rotations (Fig. 5).

Seismic analysis has been carried out in operational conditions with a three-component weak real earthquake record. Numerical results have showed that while the top displacement is dominated by the wind thrust, force and bending moment in the longitudinal direction at the tower base are affected considerably by earthquake loads. In addition, in the lateral direction (where there are no wind loads), the force and bending moment are essentially decided by the earthquake loads and amplify several times.

Another interesting result concerns the lateral reaction force at the main bearing that is found to be significantly increased with respect to the case of no earthquake loading. This increase of the lateral reaction force is attributed to the gyroscopic forces that arise because of the change of the wind rotor direction, due to the earthquake loading. A further result of the study by Zhao and Maisser (2006) is that SSI modeling affects significantly the higher tower modes, especially the second lateral bending mode. In this respect, this result confirms the results obtained by Bazeos

et al. (2002) using a simplified model of the HAWT and shall be taken into account in consideration of the fact that especially the higher tower modes may fall within the region of maximum spectral acceleration (Haenler et al. 2006).

Using a multi-body system with flexible parts (tower, blades) described by a variable number of modes and including SSI modeling, Haenler and coworkers (2006) have investigated the seismic response of a HAWT with an 80 m rotor diameter and 60 m hub height, operating under a 13 m/s wind speed and subjected to an earthquake ground motion with a 0.3 g peak ground acceleration (PGA). They have showed that the full system model predicts modes at frequencies in the region of maximum spectral response acceleration for typical design response spectra. An important contribution of this study regards the relative increase in higher mode response. It has been found that, for normal wind loading, 80 % of the tower energy is associated with the first mode, while, during the considered earthquake, the energy in the first mode is reduced to 54 % percent only of the tower energy, thus concluding that higher tower modes are more important for earthquake loading than for typical wind loading.

Another full model of HAWT has been proposed by Ishihara and Sarwar (2008). Starting from the observation that, unlike wind loads, seismic waves may excite a wide range of frequencies including those of higher modes, they have developed a nonlinear FEM code (CAST) for a full FE modeling of the HAWT and its components. Beam elements with a linear material have been used to model tower and blades. Analyses performed by the FE code on HAWTs, in a parked state, have been used in conjunction with the Japanese BSL response spectrum (BSL 2004) to derive design formula for the prediction of seismic loads on two turbines. In particular, following a semi-theoretical codified method provided by the Japan Society of Civil Engineering (JSCE 2007), a profile of seismic loads (shear and moment demand) acting on a HAWT in parked conditions was estimated from a base shear built as the sum of a shear force due the first mode, obtained by using the BSL acceleration response spectrum, and the shear force

contributions due to higher modes, obtained on the basis of the FE analysis. Results have been provided for two HAWTs, one with 400 kW rated power, 36 m hub height, and 31 m rotor diameter and the other one with 2 MW rated power, 67 m hub height, and 80 m rotor diameter. By comparison with the results obtained from time-domain simulations, it has been shown that the BSL 5 % damped response spectrum provides nonconservative seismic load profiles for both the 400 kW HAWT and the 2 MW one, thus confirming that a 5 % damping ratio is not appropriate for HAWTs in parked conditions, which experience much lower damping level. It is worth remarking that the time-domain simulations carried out by Ishihara and Sarwar (2008) highlighted that contribution of higher modes may become significant for large HAWTs (2 MW) under earthquake excitations, confirming similar results obtained by other researchers (Haenler et al. 2006).

Summary

The importance of seismic loading in the design of HAWTs is recognized in existing ISs and CGs, and in seismically active areas attention shall be paid to the possibility that design is driven by load combinations involving seismic excitations.

Existing methods for seismic assessment of HAWTs are fully coupled time-domain simulations, computing the response to simultaneously acting wind loading and seismic loading, and decoupled analyses, where the responses to wind loading and seismic loading are built separately and then superposed. Fully coupled time-domain simulations, although complex and time consuming, remain the indispensable benchmark tool for an accurate seismic assessment of HAWTs. However, decoupled analyses are also important, especially in the early stages of design.

For the implementation of fully coupled time-domain simulations, a full system model must be employed. Simplified models, where the rotor is not modeled, are allowed to build separate seismic loads in decoupled analyses.

Despite the prescriptions of existing ISs and CGs, a few important aspects are actively being investigated, such as the influence of SSI modeling, the importance of higher modes in the seismic response, the potential failure mechanisms under earthquake loading, and the estimation of aerodynamic damping for decoupled analysis. A full understanding of these aspects is desirable for a correct seismic assessment of HAWTs.

Cross-References

- [Earthquake Response Spectra and Design Spectra](#)
- [European Structural Design Codes: Seismic Actions](#)
- [Modal Analysis](#)
- [Response Spectrum Analysis of Structures Subjected to Seismic Actions](#)
- [Soil-Structure Interaction](#)
- [Stochastic Ground Motion Simulation](#)
- [Time History Seismic Analysis](#)

References

- API (2000) Recommended practice for planning, designing and constructing fixed offshore platforms, API RP-2A. American Petroleum Institute, Washington, DC
- ASCE (2006) Minimum design loads for buildings and other structures, ASCE/SEI 7-05. American Society of Civil Engineers, Reston
- ASCE/AWEA (2011) Recommended practice for compliance of large land-based wind turbine support structures, ASCE/AWEA RP2011. American Society of Civil Engineers/American Wind Energy Association, Reston/Washington, DC
- Bazeos N, Hatzigeorgiou GD, Hondros ID, Karamaneas H, Karabalis DL, Beskos DE (2002) Static, seismic and stability analyses of a prototype wind turbine steel tower. *Eng Struct* 24:1015–1025
- Bossanyi EA (2000) Bladed for windows user manual. Garrad Hassan and Partners, Bristol
- BSL (2004) The building standard law of Japan. The Building Centre of Japan, Tokyo (in English and Japanese)
- DNV/Risø (2002) Guidelines for design of wind turbines. Det Norske Veritas/Wind Energy Department, Risø National Laboratory, Copenhagen/Roskilde
- Eurocode 8 (2004) Design of structures for earthquake resistance. Part 1: general rules, seismic actions and rules for building. EN 1998-1. European Committee for Standardization, Brussels
- GL (2010) Guideline for the certification of wind turbines. Germanischer Lloyd, Hamburg
- GWEC (2014) Global wind report: Annual market update 2013. Global Wind Energy Council, Brussels
- Haenler M, Ritschel U, Warnke I (2006) Systematic modelling of wind turbine dynamics and earthquake loads on wind turbines. In: Proceedings of the European Wind Energy Conference & Exhibition (EWEC), 27 Feb–2 Mar 2006. Athens
- ICC (2012) International building code. International Code Council, Country Club Hills
- IEC (2005) Wind turbine generator systems. Part 1: safety requirements, 3rd edn, IEC 61400-1. International Electrotechnical Commission, Geneva
- Ishihara T, Sarwar MW (2008) Numerical and theoretical study on seismic response of wind turbines. In: Proceedings of the European Wind Energy Conference & Exhibition (EWEC), 31 Mar–3 Apr 2008. Brussels
- Jonkman JM, Buhl ML (2005) FAST user's guide, NREL/EL-500-38230. NREL, Golden
- JSCE (2007) Guidelines for design of wind turbine support structures and foundations, Japan Society of Civil Engineers, Tokyo (in Japanese)
- Kuhn M (2001) Dynamics and design optimisation of offshore wind energy conversion systems. Report no. 2001.002, Delft University Wind Energy Research Institute (DUWIND), Delft
- Manwell JF, McGowan JG, Rogers AL (2010) Wind energy explained: theory, design and application, 2nd edn. Wiley, Chichester
- Mulliken JS, Karabalis DL (1998) Discrete models for through-soil coupling of foundations and structures. *Earthq Eng Struct Dyn* 27:687–710
- Nuta E, Christopoulos C, Parker JA (2011) Methodology for seismic risk assessment for tubular steel wind turbine towers: application to Canadian seismic environment. *Can J Civil Eng* 38:293–304
- Prowell I (2011) An experimental and numerical study of wind turbine seismic behavior. PhD dissertation, University of California, San Diego
- Prowell I, Veers P (2009) Assessment of wind turbine seismic risk: existing literature and simple study of tower moment demand, SAND2009-1100 report. Sandia National Laboratories, Albuquerque
- Prowell I, Elgamal A, Uang C, Jonkman J (2010) Estimation of seismic load demand for a wind turbine in the time domain. In: Proceedings of the European Wind Energy Conference (EWEC), 20–23 Apr 2010, Warsaw
- Prowell I, Elgamal A, Uang C, Luco JE, Romanowicz H, Duggan E (2013) Shake table testing and numerical simulation of a utility-scale wind turbine including operational effects. *Wind Energy*. doi:10.1002/we.1615
- Stamatopoulos GN (2013) Response of a wind turbine subjected to near-fault excitation and comparison

- with the Greek Aseismic Code provisions. *Soil Dyn Earthq Eng* 46:77–84
- Swan S, Hadjian AH (1988) The 1986 North Palm Springs earthquake: effects on power facilities, NP-5607 research project 2848. Electric Power Research Institute (EPRI), Palo Alto
- Umar AB, Ishihara T (2012) Seismic load evaluation of wind turbine support structures considering low structural damping and soil structure interaction. In: Proceedings of the European Wind Energy Association Conference (EWEA), 16–19 Apr 2012. Copenhagen
- Witcher D (2005) Seismic analysis of wind turbines in the time domain. *Wind Energy* 8(1):81–91
- Zhao X, Maissner P (2006) Seismic response analysis of wind turbine towers including soil-structure interaction. *Proc Inst Mech Eng K J Multi-Body Dyn* 220(1):53–61
- Zhao X, Maissner P, Wu J (2007) A new multibody modeling methodology for wind turbine structures using a cardanic joint beam element. *Renew Energy* 32:532–546

Seismic Anisotropy in Volcanic Regions

Jessica H. Johnson

School of Earth Sciences, University of Bristol
Wills Memorial Building, Bristol, UK

School of Environmental Sciences, University of
East Anglia, Norwich, UK

Synonyms

Seismic anisotropy/shear wave splitting at (active) volcanoes; Shear wave splitting in volcanic regions; Using seismic anisotropy/shear wave splitting to monitor (active) volcanoes; Using seismic anisotropy/shear wave splitting to track/monitor/detect volcanic/magmatic activity

Introduction

Modern geophysical techniques enable changes to be observed at some volcanoes before magmatic eruptions: detection of seismicity (see “► [Seismic Monitoring of Volcanoes](#)”) from magma pushing through cold country rock is one of the most common and successful monitoring techniques and can lead to short-term

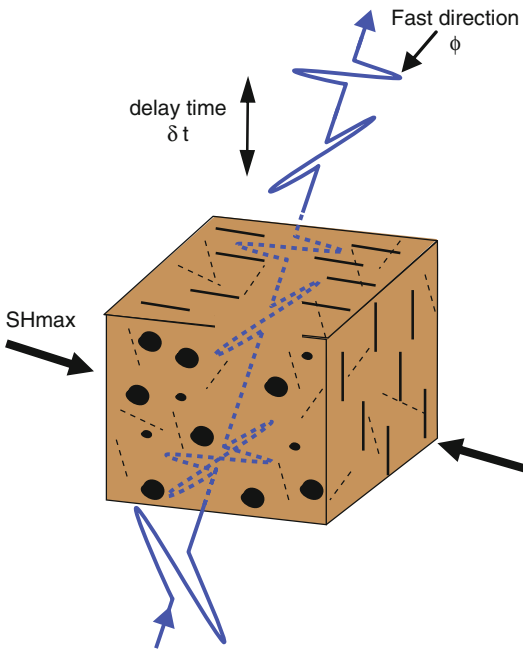
forecasting (see “► [Volcanic Eruptions, Real-Time Forecasting of](#)”). Another example of geophysical precursors to eruptions is surface deformation from inflation or deflation of a volcano due to magma movement. However, some volcanoes do not display these clues, and there remains a need for techniques that are sensitive to other physical attributes that might change in conjunction with the eruption process. Any overpressured magma storage reservoir, be it a system of dikes, sills, conduits, spherical chamber, or a combination of these, will exert a stress on the surrounding country rock that may or may not be manifest as observable strain. Detecting and understanding this stress may be a key to predicting if and when a volcano will erupt.

Definition

Seismic anisotropy is the variation of seismic wave speed with direction. It is an indicator of geometric ordering in a material, where features smaller than the seismic wavelength (e.g., crystals, cracks, pores, layers, or inclusions) have a dominant alignment. This alignment leads to a directional variation of elastic wave speed. Seismic anisotropy can be measured using many seismological techniques but is most frequently observed using shear wave splitting, which requires clear shear wave energy recorded on three-component seismometers. Measuring the effects of anisotropy in seismic data can provide important information about processes in the Earth such as stress conditions, material flow, and the structure of the subsurface and deep Earth.

Historical Background

Hess (1964) first made a significant observation of large-scale anisotropy when seismic refraction measurements in oceans showed that the P-wave velocity of the upper mantle (Pn) was consistently higher for profiles recorded perpendicular to an oceanic spreading center (i.e., parallel to the direction of spreading or plate movement) than for profiles recorded parallel to the spreading center. These measurements were attributed to the alignment of olivine crystals in the mantle lithosphere because of flow during the formation



Seismic Anisotropy in Volcanic Regions, Fig. 1 Shear wave splitting in an anisotropic medium. Anisotropy is caused by preferentially aligned cracks due to a maximum horizontal compressive stress (S_{Hmax}). A vertically propagating shear wave that is arbitrarily polarized gets split into a fast wave with polarization (ϕ) parallel to crack alignment and a slow wave, which is polarized at 90° to ϕ . The waves are separated with delay time δt

of the oceanic plate at the ridge. Since the 1970s, improvements in computing power and memory and in seismic field observation have led to a greater understanding of the seismic anisotropy of the Earth at all levels and scales (Savage 1999). Measurements of seismic anisotropy have been used to detect fabric and stress in the Earth's crust, flow in the upper mantle, topography of the core–mantle boundary, and differential rotation of the inner core.

The measurement of seismic anisotropy has been found to be a proxy for determining the direction of maximum horizontal compressive stress (S_{Hmax}) in the crust; applied stress can cause microcracks to preferentially open parallel to the maximum compressive stress, creating an anisotropic medium with the fast direction parallel to S_{Hmax} (Fig. 1). The mechanism of aligned microcracks is thought to be the only one that allows seismic anisotropy to vary on observable

timescales, and temporal changes are traditionally interpreted as stemming from variations in the stress field due to large earthquakes or magmatic intrusions. There is mounting evidence, however, that the dominant mechanism for seismic anisotropy can switch between a static condition, such as aligned fractures in fault zones, and a dynamic process, such as stress causing aligned microcracks to dilate. In areas where there are strong changes in S_{Hmax} direction and magnitude on observable timescales, such as at active volcanoes, seismic anisotropy analysis has proven a valuable tool when combined with ground deformation or other seismological observations for interpretation of volcanic processes such as magma migration (e.g., Gerst and Savage 2004; Bianco and Zaccarelli 2009; Unglert et al. 2011).

Seismic Anisotropy

Shear Wave Splitting

Shear wave splitting occurs when a shear wave travels through a seismically anisotropic medium, i.e., one in which seismic waves travel faster in one direction or with one polarization than another (Fig. 1). In the Earth's upper crust, anisotropy is most likely to be caused either by stress conditions preferentially aligning microcracks parallel to the maximum compressive stress or by pervasive structural features. For a near-vertical propagation direction, the shear wave with the displacement in the plane of the open cracks will travel faster than that crossing the plane of the cracks, and so a fast shear wave with orientation ϕ and a slow shear wave orthogonal to ϕ , separated by a delay time δt , will be observed (Babuška and Cara 1991). Crack-induced anisotropy has in some studies been considered a direct indicator of present-day stress, with ϕ providing information about the orientation of S_{Hmax} and δt giving information about the strength of anisotropy and the amount of time that the wave spent traversing the anisotropic medium. Studies that combine independent stress estimation methods with shear wave splitting results have found that strong geological fabric or aligned structures rather than the maximum

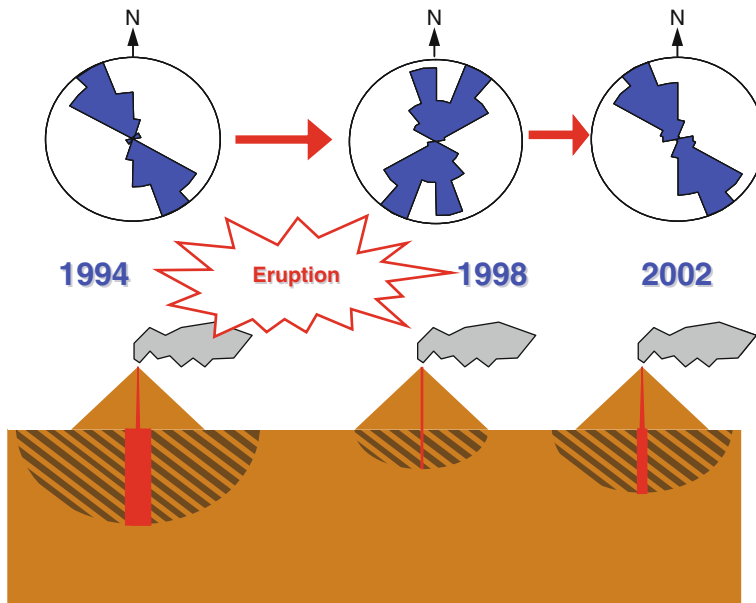
stress can govern the observed ϕ in some situations. Both stress-induced and structure-related anisotropies have been observed in the same regions and, in some cases, have been detected at the same station (e.g., Zinke and Zoback 2000; Johnson et al. 2011).

The subject of temporally varying anisotropy is a hotly debated topic (e.g., Crampin and Peacock 2008). Temporal variations in shear wave splitting can take the form of a rotation in the fast direction of anisotropy, an increase or decrease in the delay time, or a combination of both. The changes are thought to stem from perturbations of the elastic properties of the crust due to crack opening either by local concentration of shear stress or by a change in pore-fluid pressure. Monitoring these changes can therefore provide crucial information about the state of stress and pore content around active volcanoes. However, spatial variations of anisotropy masquerading as temporal variations are one of the main caveats associated with the interpretation (Johnson et al. 2011). This can occur when the location of the source earthquakes changes or migrates throughout the duration of the observation period leading to the seismic waves experiencing different anisotropic conditions due to heterogeneity in the anisotropic structure. Temporal variations in anisotropy associated with major magmatic eruptions, such as the 1995/1996 eruptions of Mount Ruapehu, have been subject to doubts about the possibility of shear wave splitting variations being due to changing source earthquake location. Therefore, the methods must be fully understood before changes in anisotropy with time can truly be used as an eruption forecasting tool. To do this, one must first explore the spatial variations in anisotropy at volcanoes before examining the temporal changes.

Mt. Ruapehu volcano in the North Island of New Zealand has been the subject of some of the most comprehensive shear wave splitting studies of any volcano on Earth. Therefore, to illustrate the spatial versus temporal variation issue, anisotropy studies at Mt. Ruapehu will be explored. Temporal variation of shear wave splitting at Mt. Ruapehu was investigated by Miller and Savage (2001) and Gerst and Savage (2004).

Miller and Savage (2001) measured shear wave splitting from earthquakes in 1994 and 1998 and observed a change in the dominant ϕ spanning the time of the last major magmatic eruption in 1995/1996. That study was extended by Gerst and Savage (2004), who used the same techniques and an additional deployment of three-component seismometers in 2002 to observe further changes in ϕ (Fig. 2). The results of both studies were interpreted as being caused by a dike-shaped magma reservoir, or system of dikes, trending NE–SW. According to this model, the magma reservoir was pressurized before the eruption, producing a local stress field different from the regional stress field. This interpretation is favored above one in which dikes are intruded and solidify, causing a new structural anisotropy, because of the lack of detectable deformation and seismicity associated with dike intrusion. The model suggests that following the eruption the reservoir was less full and correspondingly less pressurized, meaning that the local stress returned to that of the surrounding region. The Gerst and Savage (2004) study suggested that the later changes in ϕ were due to repressurizing of the reservoir in response to an increase of magma in the system because ϕ from deep earthquakes displayed the regional trend, while ϕ from shallow earthquakes was oriented to the pre-eruption direction. The return of anisotropy to the pre-eruptive state also supports the interpretation of stress-controlled anisotropy as the intrusion of dikes would be a permanent change.

Johnson et al. (2011) investigated the spatial variations in anisotropy in more detail using a dense seismometer deployment in 2008, to compare future changes in anisotropy and to identify the regions and causes of past changes in anisotropy with more confidence. Johnson et al. (2011) divided the mapped anisotropy into regions in which the fast polarizations agreed with stress estimates from focal mechanism inversions, suggesting stress-induced anisotropy, and those in which the fast polarizations were aligned with structural features such as faults or metamorphic fabric, suggesting structural anisotropy. Using this benchmark of anisotropy,



Seismic Anisotropy in Volcanic Regions, Fig. 2 Schematic stress and anisotropy model after Gerst and Savage (2004). In 1994, a pressurized dike system created a local stress field. In 1998, after the eruption, when the dike system was depressurized, stress directions partially returned to the regional trend. In 2002,

the dike system refilled, and the stress field in the anomalous region was dominated by the dike again. The alignment of cracks was not as strong as in 1994, so the anisotropy in the anomalous region was not strong enough to affect fast directions from deep events

Johnson and Savage (2012) examined temporal changes in shear wave splitting from 1994 to 2010. They observed a region of strong anisotropy centered on Mt. Ruapehu in 1995, the time of a major magmatic eruption, agreeing with Miller and Savage (2001) and Gerst and Savage (2004), which was interpreted to be due to an increase in fluid-filled fractures during the eruption. They also observed strong anisotropy and a change in fast direction ($\sim 80^\circ$) at Mt. Tongariro in 2008, which was initially interpreted to be due to a change in the geothermal system but was later the location of a small eruption in 2012.

Measuring Shear Wave Splitting

Many methods have been developed to measure shear wave splitting in the Earth's crust (e.g., Crampin and Gao 2006). The goal is to identify the orientation of the fast split shear wave and the delay time between the fast and slow split shear waves. These parameters can then be used to infer rock properties such as crack distribution and

geometry, pore content, or stress regime. Ideally, the procedure will be able to accurately process large quantities of three-component data in an efficient, unbiased, and objective way, without operator intervention, while providing quantitative evaluation of the uncertainties for each measurement of shear wave splitting.

A manual method for analyzing shear wave splitting is that of direct inspection of the 3D particle motion projected onto a horizontal plane (hodograms). This method relies heavily on observer judgment and therefore can produce biased results and is inefficient for large datasets. Other methods employ partially automated programs, where the user must pick a window around the shear wave arrival and/or evaluate the results. There are several approaches commonly used to semiautomatically determine shear wave splitting parameters: the covariance matrix technique searches for nonzero eigenvalues to "unsplit" the shear wave; the cross-correlation technique searches for the rotation

and time delay that yields the highest cross-correlation between orthogonal components; the aspect ratio technique searches for the rotation and time delay that yields the most linear particle motion with the maximum aspect ratio; and the vector amplitude technique uses the maximum amplitude in a time window to identify the split shear waves. Any degree of user interaction may introduce some subjectivity and is usually time-consuming, although many practices of automated quality control will result in loss of significant amounts of data. Completely automated shear wave splitting analysis has proven elusive, as the step that is difficult to automate is the picking of the shear wave arrival. Several methods boast full automation other than this problematic step. In general, since phase arrivals are often picked for previous analyses such as hypocenter location, these automated methods are preferable to the ones with a lot of user interaction; more data can be evaluated, reducing the effect of loss of data through the quality control steps.

Interpreting Shear Wave Splitting: Some Assumptions

Most shear wave splitting observations are interpreted under some assumptions, which are rarely completely true in the Earth:

1. **That the medium possesses hexagonal symmetry with a horizontal axis:** The majority of anisotropic rocks in the Earth have, or can be approximated to have, hexagonal symmetry. This is because the most common symmetries have patterns that do not differ significantly for horizontal fast axis alignment and near-vertical incidence angles. The simplest models used to explain variations in two orthogonal directions are hexagonally symmetric models. Therefore, shear wave splitting is usually interpreted in terms of transverse anisotropy with a horizontal symmetry axis.
2. **That the anisotropic medium is in a single, homogeneous layer:** When a shear wave passes through multiple anisotropic layers, the observed splitting parameters depend strongly on the thickness and strength of anisotropy of the layers, on the relative fast directions, and on the wavelength of the wave. If the shear wave has been sufficiently split in the first layer that the fast and slow waves are separated, then when it enters the second layer, which has a fast direction 20–70° different, both of the quasi-shear waves will be split again. In this case both waves will now have the fast and slow directions of the second layer. However, when the splitting from the first layer is weak so that the two quasi-shear waves are not more than one wavelength apart, both waves are still resplit, but the result is a complex waveform that is difficult to interpret but can still be meaningful. In general, shear wave splitting fast direction is representative of the last layer that the wave passed through and the delay time can be approximated as accumulating along the path.
3. **That the anisotropic medium is localized beneath the receiver:** As seen in the previous point, in general, the anisotropy parameters are representative of the last anisotropic medium that the wave travelled through. Therefore, the assumption that the medium is local to the station is usually justified. However, in regions with heterogeneous anisotropic structure, data at a single station will not be consistent and the assumption will not be appropriate.

Further discussions of these points can be found in Silver and Savage (1994), Savage (1999), and Johnson et al. (2011).

Other Estimates of Seismic Anisotropy

Shear wave splitting analysis is becoming a very popular method for determining seismic anisotropy in the crust. This is in part due to the abundance of data and methods available and also to the relative insensitivity to the source–receiver geometry (other than deep enough earthquakes for the rays to arrive within the shear wave window) and independence from the need for dense networks of seismometers (see “► [Seismic Network and Data Quality](#)”). Another benefit of

shear wave splitting analysis is that, even though the results can be averaged over multiple measurements, they represent a snapshot of anisotropy. There are, however, other methods for determining seismic anisotropy. Most of the other methods involve calculating anisotropic velocities using tomographic techniques (see “► [Seismic Tomography of Volcanoes](#)”) with body or surface waves from active or passive sources, or using ambient seismic noise (see “► [Noise-Based Seismic Imaging and Monitoring of Volcanoes](#)” and “► [Seismic Noise](#)”).

Tomographic methods (see “► [Seismic Tomography of Volcanoes](#)”) require larger amounts of data with relatively even coverage of sources and sensors, which (usually) take longer to acquire, thereby rendering the methods less practical for time-lapse investigations than shear wave splitting investigations. Tomographic methods that are used to calculate seismic anisotropy are different to techniques that conduct tomographic inversions on shear wave splitting data. Anisotropy parameters are derived from the inversions in the former case, while the latter is an inversion of the anisotropy parameters.

Inversions of body wave arrival times for three-dimensional velocity structures are common practice at volcanoes using teleseismic waves, local earthquake sources, and active seismic sources. These inversions can also account for 3D V_p azimuthal anisotropy, which is parameterized with a percent anisotropy and an orientation of the fast axis. Seismic anisotropy can be detected using inversions of Love and Rayleigh surface waves from large earthquakes in the same manner.

Surface waves constructed from cross-correlations of ambient seismic noise can be inverted for 3D seismic velocity structure. Seismic anisotropy from ambient noise tomography can be calculated. These calculations are different from the time-lapse studies that detect temporal variations in isotropic seismic velocities using ambient noise interferometry (see “► [Tracking Changes in Volcanic Systems with Seismic Interferometry](#)”).

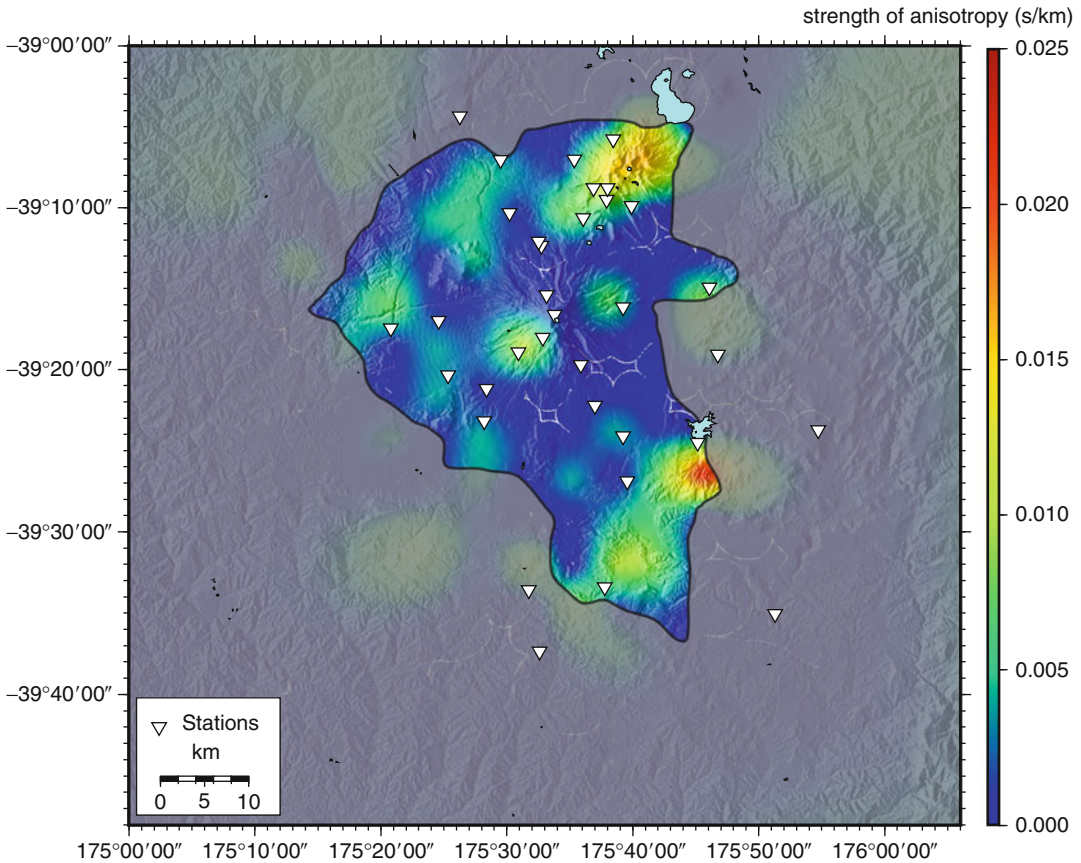
Variations in Shear Wave Splitting

Spatial Variation of Shear Wave Splitting

Most shear wave splitting results are plotted as rose diagrams (circular histograms) of fast direction at the station at which the measurements were made (e.g., Fig. 2). This implicitly assumes that the anisotropy is localized beneath the station. In many regions this may be appropriate; however, when there is lateral heterogeneity, the rose diagrams will become scattered or multimodal. Johnson et al. (2011) detected distinct splitting parameters for measurements using earthquakes from several different clusters in the region of Mt. Ruapehu. This backazimuthal dependence implies that the ϕ value obtained from shear wave splitting analysis is highly dependent upon the path that the ray takes, which has also been found in other regions. Furthermore, this suggests that the anisotropy changes over shorter distances in the crust than is often observed for mantle anisotropy and that averaging ϕ over the whole region may not be appropriate. If the causes of the different regions of anisotropy are known, it is easier to identify and map the differences using shear wave splitting analysis. If the crustal stresses or fabrics are more complex or unknown, then it is more difficult to map and interpret heterogeneous anisotropy, although a denser array of seismic stations and a broad range of backazimuths increase the likelihood of identifying the source of heterogeneity.

For time-lapse studies, it is important to mitigate the chance of spatial variations of shear wave splitting being erroneously interpreted as temporal variations. One way to do this would be to map spatial variations in detail at a time when there are no hypothesized temporal changes. This becomes difficult, but essential, when the anisotropy is very heterogeneous and the seismograph array is less dense than the spatial changes. This section outlines some examples of shear wave splitting tomography studies, which attempt to solve the problem of spatial variation.

Shear wave splitting tomography is difficult because of the nonlinear effect of multiple layers



Seismic Anisotropy in Volcanic Regions, Fig. 3 Delay time tomography from the inversion of shear wave splitting data at Mount Ruapehu Volcano, New Zealand. Warm colors indicate strong anisotropy,

shaded area shows the limit of statistical significance calculated from the model variance matrix, white inverted triangles are seismic stations (After Johnson et al. (2011))

of anisotropy on waveforms. This is different to travel time tomography, in which the travel time of a wave is often linearly related to the velocity structure of the media it has passed through. Because of this difficulty, many techniques treat ϕ independently from δt . Three-dimensional tomographic inversions can be carried out on the delay time data to characterize fracture density distribution. This tomography uses only δt from local earthquakes to investigate anisotropy strength in the crust. In this way, regions of high anisotropy can be identified, but information regarding fast directions is not accounted for.

Audoine et al. (2004) presented a simple method of 2D spatial averaging to examine heterogeneous anisotropy in the crust. A grid was

constructed with nodes regularly spaced between each earthquake and station. This grid was then treated as a new dataset, and ϕ for each node within a polygon, or within a box of a regular lattice, was averaged. This created average ϕ values at regular intervals that could be denser than the station spacing, hereby identifying spatial trends in fast directions, but not accounting for delay times. Johnson et al. (2011) adapted a combination of the above methods to map the heterogeneous seismic anisotropy field around Mt. Ruapehu in New Zealand (Fig. 3). To constrain the locations of high anisotropy, they conducted a two-dimensional tomographic inversion on the delay time estimates. They then used a spatial averaging technique similar to that of

Audoine et al. (2004), but with the fast polarizations weighted according to the strength of anisotropy calculated from the tomography. The method uses a quad-tree gridding system to enable higher resolution where the data permit and couples the two shear wave splitting parameters, even though they are not used in a joint inversion.

Abt and Fischer (2008) carried out full 3D shear wave splitting tomography for the mantle. The method parameterized the mantle as a 3D block model of crystallographic orientations. Nonlinear properties of shear wave splitting were accounted for by applying the inversion iteratively and recalculating partial derivatives after each iteration. Using this method, Abt and Fischer (2008) modeled an idealized subduction zone with uniform stations and sources. When applied to real data, they found that the geometry of stations and observed seismicity in the Nicaragua–Costa Rica subduction zone yielded partial to good resolution. This method also has the potential to be applied to crustal studies such as active volcanoes.

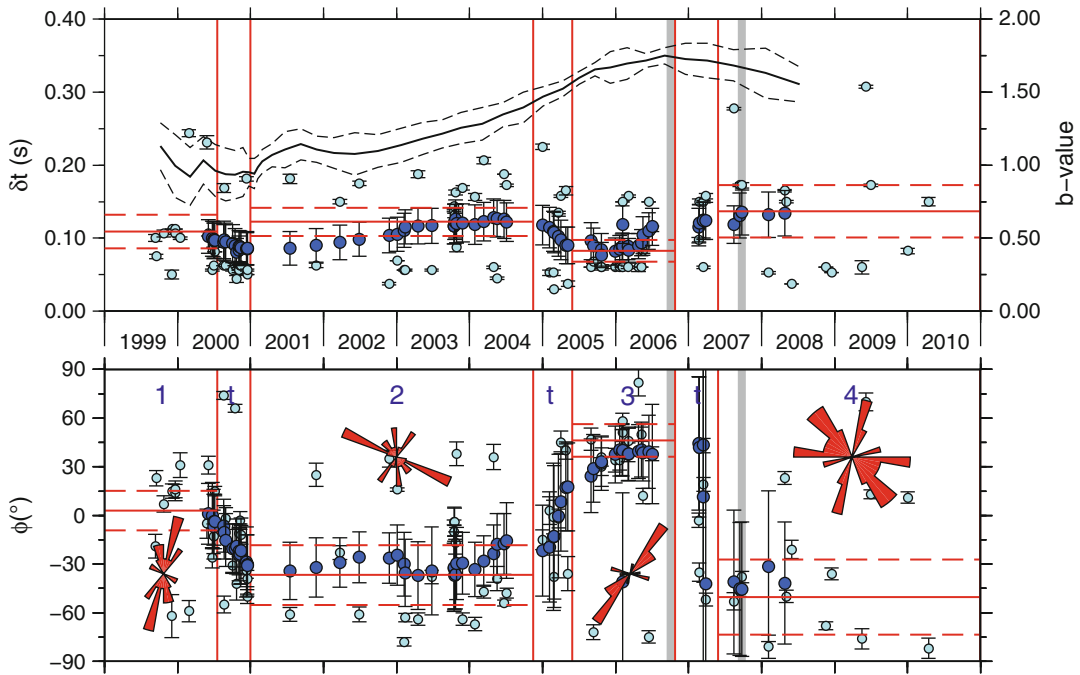
Temporal Variation of Shear Wave Splitting in the Crust

The temporal variation of shear wave splitting and its interpretation is highly controversial. The main point of dispute is whether the accumulation of stress before significant events such as eruptions or large earthquakes allows the time and magnitude of impending events to be stress forecast through shear wave splitting monitoring. The effects have been reported with hindsight before at least 15 earthquakes. Crampin and Peacock (2008) present a summary of observations of temporal variations in shear wave splitting attributed to stress-aligned fluid-saturated microcracks. However, alternative interpretations of these observations of temporal changes suggest there may be observer bias in data selection, unsound statistical analyses, misinterpretation of spatial variation, and lack of correlation with other stress-determining factors/correlation with structural evidence.

Clear evidence has been obtained that crustal shear wave splitting can vary over short distances and can be caused by structural features in the

crust, which would not change with changing stress. However, it is also clear that anisotropy due to stress-aligned fluid-saturated microcracks can change with time, as has been demonstrated in industry when small changes associated with injection and removal of fluids from reservoirs were examined. The use of similar earthquakes (i.e., those that have the same source mechanism and location) for shear wave splitting analysis helps to eliminate some of the discrepancies with interpretation. Johnson et al. (2010) used shear wave splitting analysis, a multiplet of 25 similar earthquakes and double-difference relocation to examine temporal variations in seismic properties prior to and accompanying magmatic activity associated with the 2008 eruption of Okmok Volcano. They found a general change in ϕ but could not rule out dependence on backazimuth and no significant change using the multiplet. Using earthquakes originating from the subducted slab in order to reduce the effect from changing paths, several modes of ϕ were identified, relating to the anisotropy of the mantle wedge, regional stress direction, and local stress induced by the pressurization and depressurization of the magma reservoir. These modes were found to have different prominence at different times throughout the eruptive cycle.

Once again returning to the example of Mt. Ruapehu, Miller and Savage (2001) and Gerst and Savage (2004) observed general changes in seismic anisotropy throughout the eruptive cycle. Johnson et al. (2011), however, showed that seismic anisotropy around Mt. Ruapehu is heterogeneous and that averaging the whole region is not appropriate. Therefore, Keats et al. (2011) used seismicity generated from a consistently active area of seismicity about 20 km to the west of Mt. Ruapehu. Shear wave splitting results revealed a decrease in delay time in the 2006–2007 eruption period and a significant variation in the fast shear wave polarization in the same time period (Fig. 4). These changes were attributed to an increase in pore-fluid pressure in the region due to fluid movement, and it was suggested that this fluid movement may be associated with the eruptions in 2006 and 2007.



Seismic Anisotropy in Volcanic Regions, Fig. 4 Moving average plot of fast polarization (ϕ) and delay time (δt) using earthquakes within the Erua swarm at station FWVZ at Mount Ruapehu, New Zealand. Individual measurements for ϕ and δt are displayed in light blue and 10-point moving averages are displayed in dark blue. The error bars indicate 95 % confidence intervals. The four time periods, marked by the numbers 1–4, and three transition zones, marked by a t, are indicated with vertical red lines and the mean for each period are shown

by the red horizontal bars with 90 % confidence interval (dashed red lines). The times of the two phreatomagmatic eruptions that occurred are also marked with grey bars. Rose diagrams of ϕ are displayed in their respective time periods. The b-value for the Erua swarm catalogue is also plotted against time in black at the top using a window of 40 events and an eight-event overlap. Dashed black lines indicate 95 % confidence interval (After Keats et al. (2011))

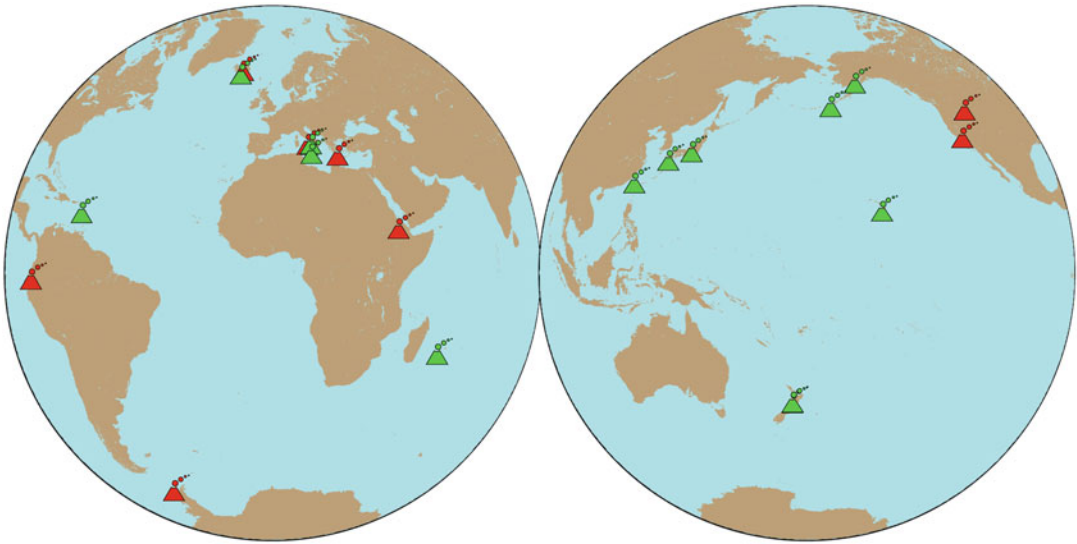
Temporal variation of shear wave splitting parameters as a stress indicator has been employed at several other volcanoes. While most studies concentrate on either stress-induced anisotropy or structurally dominated anisotropy, several studies have found that the dominant mechanism can change between the two (Keats et al. 2011; Johnson and Poland 2013) and that more than one mechanism can be dominant over short distances, potentially creating an apparent temporal change that is actually an artifact of changing earthquake location (Zinke and Zoback 2000).

Even though we have seen that shear wave splitting analysis can be used as an indicator of stress and of fluid saturation in the crust, surprisingly few studies have been conducted on shear

wave splitting around volcanoes (Fig. 5). This is due, in part, to the generally noisy waveforms and complicated interpretation of such observations when taking into account heterogeneity and complex stress regimes. It should be noted that there is significant literature about shear wave splitting in the mantle beneath active volcanoes because these regions are invariably of interest tectonically, but this entry focuses on crustal studies only.

Relating Shear Wave Splitting to Other Observations

Shear wave splitting analysis has been proven to be a useful indicator of stress and of fluid



Seismic Anisotropy in Volcanic Regions, Fig. 5 Map showing volcanoes at which shear wave splitting investigations have been carried out. *Green volcano symbol*

indicates that temporal variations were observed, *red* indicates that no temporal variations were observed

saturation in the crust but can also be caused by structures such as macroscopic fractures or mineral alignment. At volcanoes, the generally noisy waveforms, complicated heterogeneity, and complex stress regimes can make interpretation of shear wave splitting difficult. For this reason, shear wave splitting investigations are often coupled with other stress or strain indicators and structural geology data to minimize the ambiguity in the interpretation.

Volcanic Seismicity, Fault Plane Solutions, and b-Value

Studies of shear wave splitting in areas of developed industry may use direct indicators of the stress field in the crust, such as borehole break-outs, for comparison with shear wave splitting results. However, few active volcanoes have these clues, so other stress monitoring techniques must be employed.

The occurrence of volcano-tectonic (VT) (see “► [Volcano-Tectonic Seismicity of Soufriere Hills Volcano, Montserrat](#)”) earthquakes is an immediate indicator of differential stress in the crust around a volcano. Changes in the stress field may trigger more VTs, the detection of which is one of the most successful tools of seismic

volcano monitoring. For this reason, many shear wave splitting studies have been designed to investigate temporal changes in anisotropy during a volcanic crisis (e.g., Savage et al. 2010; Johnson et al. 2010; Roman and Gardine 2013). Bianco and Zaccarelli (2009) calculated a background seismic anisotropy at Mount Vesuvius of 4 %, but at times of seismic crisis, the average anisotropy was 8 % and the fast directions tended to flip by 90°.

As noted above, however, heterogeneity in the anisotropic media must be carefully considered, especially when there is a migration of seismicity associated with magma movement. These migrating earthquakes are used as sources of shear wave energy, but temporal variations observed in the anisotropy measurements could be due to spatial variations.

VT earthquakes contain additional information about the state of stress in the crust. Fault plane solutions (FPS) provide information about the orientation of the fault plane that slipped, as well as the slip vector for each earthquake. From the FPSs, the pressure and tension axes can be identified, which give information about the stress regime. The FPS P-axis azimuth, which represents the orientation of the principal

compressional axis of the moment tensor, may in reality differ significantly from the orientation of maximum compressive stress, depending on the orientation of preexisting planes of weakness (faults). This is sometimes countered by calculating large numbers of FPS P-axis azimuths to highlight the major trends that most likely represent the actual orientations of S_{Hmax} (e.g., Roman and Cashman 2006; Roman and Gardine 2013). Another step can be taken by inverting for the full stress tensor using the FPS results (e.g., Johnson et al. 2011). The true direction of S_{Hmax} can then be compared with shear wave splitting parameters to jointly interpret the results.

The orientations of FPS of VT earthquakes have been observed to display systematic changes related to episodes of magmatic activity at several volcanoes (Roman and Cashman 2006). The pressure axes of FPS have been observed to rotate orthogonal to the dominant regional stress orientation in some cases, indicating that these earthquakes may result from dike inflation in the direction of minimum compressive stress (Roman and Cashman 2006). Roman and Gardine (2013) investigated shear wave splitting and earthquake fault plane solutions at Redoubt volcano around the time of the March 2009 eruption. They found that an approximately $\sim 90^\circ$ change in the polarization of shear wave splitting fast polarization accompanied the earliest signs of seismic unrest in 2008 and continued through the eruption before diminishing in 2009. A similar change in the orientation of fault plane solutions occurred 18–48 h prior to the eruption onset on March 23, 2009, but almost 2 months after a strong increase in the rate of shallow VT earthquakes. The shear wave splitting and FPS results suggested a protracted period of slow magma ascent followed by a short period of rapidly increasing magma pressurization beneath the volcano. Roman and Gardine (2013) found that after the eruption, both shear wave splitting fast polarization and fault plane solution pressure axes had a direction more consistent with the regional stress than before the eruption.

The b-value of a cluster of earthquakes is often used to describe the size distribution. The b-value

comes from the Gutenberg–Richter law and is a frequency magnitude (see “► [Frequency-Magnitude Distribution of Seismicity in Volcanic Regions](#)”) relation. For crustal earthquakes, b-values are typically ~ 1 for tectonic earthquakes, though they tend to be higher in volcanic regions (see “► [Frequency-Magnitude Distribution of Seismicity in Volcanic Regions](#)”). The b-value has been related to physical properties such as stress, material homogeneity, and pore pressure and can therefore be useful in comparison with seismic anisotropy measurements to determine changes in physical properties around volcanoes. Keats et al. (2011) used seismicity from a discrete cluster of earthquakes near Mt. Ruapehu in New Zealand to compare shear wave splitting and b-values over 12 years (Fig. 4). The joint analysis allowed them to attribute temporal changes to an increase in pore-fluid pressure due to fluid movement associated with eruptive activity.

Attenuation, V_p/V_s Ratio, and Noise Cross-Correlations

Perturbations of the elastic properties of the crust around volcanoes are often linked to crack opening either by local concentration of shear stress or by an increase of pore-fluid pressure. Shear wave splitting analysis is one way to monitor these properties, but there are other techniques that can provide additional information about the state of stress or pore content. Earthquake coda (the part of a high-frequency seismogram following the P and S waves) is generated by random scattering processes in the crust. The coda decay parameter (or Q_c^{-1}) is often used to parameterize characteristics of the medium, and temporal changes in Q_c^{-1} have been linked to crack opening (Del Pezzo et al. 2004).

Increased gas content in pores and cracks has the effect of lowering P-wave speed due to the higher fluid compressibility, but not significantly affecting the shear modulus and hence S-wave speed. The ratio between P-wave velocity (V_p) and S-wave velocity (V_s), V_p/V_s , is therefore useful for characterizing pore-fluid content. Gas-enriched pore space has been reported to affect V_p/V_s above magmatic intrusions and has

been combined with shear wave splitting studies to characterize pore content at volcanoes (Unglert et al. 2011; Johnson and Poland 2013).

The perturbations of the elastic properties of the crust around volcanoes caused by changes in stress or pore fluid can be monitored continuously using cross-correlations of ambient seismic noise (see “► Noise-Based Seismic Imaging and Monitoring of Volcanoes” and “► Seismic Noise”). Monitoring of Q_c^{-1} , V_p/V_s ratio, and isotropic velocity from seismic noise cross-correlations therefore naturally complements shear wave splitting investigations as the mechanisms of change are so similar (Del Pezzo et al. 2004).

Deformation

Geodetic techniques that determine changes in strain, such as GPS and InSAR, can provide some of the stress tensor elements and can therefore be useful to compare with shear wave splitting investigations. Studies such as those by Savage et al. (2010) and Unglert et al. (2011) use strain data to determine that the anisotropy is due to stress-oriented microcracks in the upper crust. At Mount Asama, Savage et al. (2010) correlated GPS baseline length measurements with shear wave splitting measurements to analyze stress changes accompanying the eruption in 2004. They found that the best model from the GPS analysis of a vertical dike and conduit also fit the shear wave splitting measurements, as did the temporal variations. From this, a crack aspect ratio of 2.6×10^{-5} was calculated and a differential horizontal stress of 6 MPa at 3 km depth was inferred. Unglert et al. (2011) performed shear wave splitting analyses on local earthquakes around Aso Volcano between 2001 and 2008 and compared the results to strain from GPS measurements in the area. They observed, using clusters with relatively stable epicenters, that two stations showed a significant change in ϕ in 2004–2005. Models from seismic tomography and receiver functions were found to fit both the anisotropy and strain measurements. Other studies have used the observations of no correlation between geodetic measurements and shear wave splitting results to infer that the mechanism of anisotropy is structurally

controlled. An anticorrelation between geodetic measurements and shear wave splitting results was observed at Kīlauea Volcano, implying that some volcanic process was affecting both measurements but that the mechanism was different to the traditional interpretation of a pressurizing magma reservoir. Johnson and Poland (2013) used these observations, combined with V_p/V_s ratio to interpret these changes as being due to increased gas flux.

Geology

Numerous investigations have found that in some regions, the fast direction of anisotropy is oblique to the direction of maximum horizontal compressive stress. In these cases, other clues for the cause of anisotropy are sought and the anisotropy is frequently found to align preferentially with macroscopic structures at the surface. Macroscopic fractures are often aligned because they are caused by faulting in a regional stress field. These macroscopic structures also cause anisotropy and shear wave splitting that often aligns with S_{Hmax} . Exceptions arise when strike-slip faults initially align at 45° to the direction of S_{Hmax} , faults are a product of a paleostress, or the rock has been deformed since the faulting. Velocity anisotropy is also strongly dependent on rock fabrics and metamorphic rocks with distinct foliations can have anisotropies of up to 20 % even in the absence of cracks (Babuška and Cara 1991, and references therein). This is mainly due to the preferred alignment of intrinsically anisotropic minerals such as biotite and hornblende. Geological textures can be identified through analysis of in situ rocks or oriented drill cores. Shear wave splitting that is caused by structural features is unlikely to change over time periods of investigation. However, due to changing conditions in stress or pore content, the dominant mechanism of anisotropy can change, rendering a temporal variation in shear wave splitting parameters. Johnson and Poland (2013) investigated shear wave splitting changes at Kīlauea Volcano associated with the onset of the summit eruption in 2008. They found that the orientation of fast shear waves at Kīlauea was usually controlled by structure, but in 2008 showed changes

with increased SO₂ emissions preceding the start of the summit eruption.

Tectonic Versus Magmatic Stress-Controlled Anisotropy

Studies that have recorded rotations in the fast direction of anisotropy due to a localized perturbation in stress often relate the background stress regime to the regional tectonics of the area. Rotations of ϕ by 90° are commonly observed, but several mechanisms for the rotation have been proposed. The 90° rotations should be treated carefully because cycle skipping, when the match of the fast and slow waveforms has a factor of T/2 ambiguity where T is the dominant period, can lead to an error of 90° in the recorded fast direction and a false delay time. When the rotations are real, Crampin et al. (2002) suggest that they are “flips” caused by wave propagation through cracks containing fluids at high pore-fluid pressures. However, a 90° rotation would also be expected due to a dike intrusion (Gerst and Savage 2004). A dike will exert pressure on the surrounding rock, generating a local stress field that is superimposed on the regional stress field. The stresses of such an elongated structure are mainly oriented perpendicular to the strike axis. When the pressure in the dike system is high enough, the generated stress field locally reorients the principal stresses as well as the local crack alignment. The pattern of stress perturbation around an intrusion can be complex, however, and the interaction with the regional, or tectonic, stress has been the subject of several studies (e.g., Roman and Cashman 2006; Vargas-Bracamontes and Neuberg 2012). Vargas-Bracamontes and Neuberg (2012) found that in the presence of a dominant regional stress field, the stress perturbation from an intrusion will be negligible. As the pressure from the intrusion increases, both the regional and the local stress field will coexist. This phenomenon was observed at Okmok Volcano, Alaska, when the shear wave splitting results displayed a mode corresponding with the regional stress direction and one corresponding with the pressurization of the magma reservoir (Johnson et al. 2010). As the pressure is progressively increased,

Vargas-Bracamontes and Neuberg (2012) found that the stress patterns gradually approach those corresponding to the absence of a regional stress field. Therefore, in cases with extremely high magma pressures, such as before large explosions at Soufriere Hills Volcano, Montserrat, the regional stress field may be omitted in numerical models.

The Future of Seismic Anisotropy in Volcanic Regions

Using seismic anisotropy to measure stress at active volcanoes holds enormous potential as a monitoring and eruption forecasting tool. Changes in seismic anisotropy associated with volcanic activity have already been detected; however, interpretation of seismic anisotropy observations is usually qualitative, and researchers struggle to use the results to quantify the magnitude of stress variations and the cause. Eruption forecasting is increasingly evolving from empirical pattern recognition to forecasting based on models of the underlying dynamics. For shear wave splitting monitoring to be beneficial, physical models must be developed that can explain not only changes in general trends but also the scatter in the data and anomalous observations. Central to this are the links between stress changes and fracture or crack compliance in the country rock and the role that fluids in cracks play, particularly the effects of hydrothermal circulation, pore pressure, and gas flux. Another limitation is that the majority of current research seeks to interpret only ϕ or only δt observations, but the two parameters are inherently linked. Models used to explain shear wave splitting observations must predict both parameters.

To quantify the response of seismic anisotropy to pre-, co-, and post-eruption subsurface magma movement, shear wave splitting data should first be used to map the areas affected by changing stress and identify the mechanism of seismic anisotropy in areas that do not have stress-controlled anisotropy. Methods have recently been developed for the inversion of

geomechanical parameters such as crack size, geometry, density, and content (e.g., Wuestefeld et al. 2012). Monitoring of these properties over time will not only indicate the occurrence of changes but will elucidate the nature of the subsurface changes.

Increasingly realistic numerical and experimental models of the fluid dynamics and elastodynamics are becoming possible. Integration of multidisciplinary data within these models will enable greater understanding of the underlying mechanisms and may be used to calculate the shear wave splitting and other important parameters for different volcanic scenarios. Ultimately, it may be possible to use information contained in seismic anisotropy to monitor subsurface magma movement and forecast changes in a volcano's behavior by establishing the characteristic stress field response for a given volcano, or through a deeper understanding of the complex relationships between seismic anisotropy, local crustal stresses, and the physical mechanisms of magma migration.

Summary

Seismic anisotropy is the variation of seismic wave speed with direction. Shear wave splitting occurs when a shear wave travels through a seismically anisotropic medium. Temporal variations in shear wave splitting can take the form of a rotation in the fast direction of anisotropy, an increase or decrease in the delay time, or a combination of both. The measurement of seismic anisotropy has been found to be a proxy for determining the direction of maximum horizontal compressive stress in the crust; applied stress can cause microcracks to preferentially open parallel to the maximum compressive stress, creating an anisotropic medium with the fast direction parallel to the maximum horizontal compressive stress. The mechanism of aligned microcracks is thought to be the only one that allows seismic anisotropy to vary on observable timescales, and temporal changes are traditionally interpreted as stemming from variations in the stress field due to large earthquakes or magmatic intrusions.

The changes are thought to stem from perturbations of the elastic properties of the crust due to crack opening either by local concentration of shear stress or by a change in pore-fluid pressure. There is mounting evidence, however, that the dominant mechanism for seismic anisotropy can switch between a static condition, such as aligned fractures in fault zones, and a dynamic process, such as compressive stress causing aligned microcracks to dilate. In areas where there are strong changes in maximum compressive stress direction and magnitude on observable timescales, such as at active volcanoes, seismic anisotropy analysis has proven a valuable tool when combined with ground deformation or other seismological observations for interpretation of volcanic processes such as magma migration. By inverting shear wave splitting data for geomechanical parameters and integrating the results with numerical and experimental models, it may be possible to monitor subsurface magma movement.

References

- Abt DL, Fischer KM (2008) Resolving three-dimensional anisotropic structure with shear wave splitting tomography. *Geophys J Int* 173(3):859–886. doi:10.1111/j.1365-246X.2008.03757.x
- Audoine E, Savage MK, Gledhill K (2004) Anisotropic structure under a back arc spreading region, the Taupo Volcanic Zone, New Zealand. *J Geophys Res* 109(B11):305. doi:10.1029/2003JB002932
- Babuška V, Cara M (1991) Seismic anisotropy in the Earth, vol 10. Kluwer Academic, Dordrecht
- Bianco F, Zaccarelli L (2009) A reappraisal of shear wave splitting parameters from Italian active volcanic areas through a semiautomatic algorithm. *J Seismol* 13(2):253–266. doi:10.1007/s10950-008-9125-z
- Crampin S, Gao Y (2006) A review of techniques for measuring shear-wave splitting above small earthquakes. *Phys Earth Planet In* 159(1–2):1–14. doi:10.1016/j.pepi.2006.06.002
- Crampin S, Peacock S (2008) A review of the current understanding of seismic shear-wave splitting in the Earth's crust and common fallacies in interpretation. *Wave Motion* 45(6):675–722. doi:10.1016/j.wavemoti.2008.01.003
- Crampin S, Volti T, Chastin S, Gudmundsson A, Stefansson R (2002) Indication of high pore-fluid pressures in a seismically-active fault zone. *Geophys J Int* 151(2): F1–F5. doi:10.1046/j.1365-246X.2002.01830.x

- Del Pezzo E, Bianco F, Petrosino S, Saccorotti G (2004) Changes in the coda decay rate and shear-wave splitting parameters associated with seismic swarms at Mt. Vesuvius, Italy. *Bull Seismol Soc Am* 94(2):439–452. doi:10.1785/0120030141
- Gerst A, Savage MK (2004) Seismic anisotropy beneath Ruapehu Volcano: a possible eruption forecasting tool. *Science* 306(5701):1543–1547. doi:10.1126/science.1103445
- Hess HH (1964) Seismic anisotropy of the uppermost mantle under oceans. *Nature* 203(494):629. doi:10.1038/203629a0
- Johnson JH, Poland MP (2013) Seismic detection of increased degassing before Kilauea's 2008 summit explosion. *Nat Commun* 4:1668. doi:10.1038/ncomms2703
- Johnson JH, Savage MK (2012) Tracking volcanic and geothermal activity in the Tongariro Volcanic Centre, New Zealand, with shear wave splitting tomography. *J Volcanol Geotherm Res* 223–224:1–10. doi:10.1016/j.jvolgeores.2012.01.017
- Johnson JH, Prejean S, Savage MK, Townend J (2010) Anisotropy, repeating earthquakes, and seismicity associated with the 2008 eruption of Okmok volcano, Alaska. *J Geophys Res* 115, B00B04, doi:10.1029/2009JB006991. (Figure 1)
- Johnson JH, Savage MK, Townend J (2011) Distinguishing between stress-controlled and structural shear wave anisotropy at Mount Ruapehu volcano, New Zealand. *J Geophys Res-Sol Ea* 116:B12. doi:10.1029/2011JB008308
- Keats BS, Johnson JH, Savage MK (2011) The Erua earthquake cluster and seismic anisotropy in the Ruapehu region, New Zealand. *Geophys Res Lett* 38(L16):315. doi:10.1029/2011GL049014
- Miller V, Savage M (2001) Changes in seismic anisotropy after volcanic eruptions: evidence from Mount Ruapehu. *Science* 293(5538):2231–2233. doi:10.1126/science.1063463
- Roman DC, Cashman KV (2006) The origin of volcano-tectonic earthquake swarms. *Geology* 34(6):457. doi:10.1130/G22269.1
- Roman DC, Gardine MD (2013) Seismological evidence for long-term and rapidly accelerating magma pressurization preceding the 2009 eruption of Redoubt Volcano, Alaska. *Earth Planet Sci Lett* 371–372:226–234. doi:10.1016/j.epsl.2013.03.040
- Savage MK (1999) Seismic anisotropy and mantle deformation: what have we learned from shear wave splitting? *Rev Geophys* 37(1):65–106. doi:10.1029/98RG02075
- Savage MK, Ohminato T, Aoki Y, Tsuji H, Greve SM (2010) Stress magnitude and its temporal variation at Mt. Asama Volcano, Japan, from seismic anisotropy and GPS. *Earth Planet Sci Lett* 290(3–4):403–414. doi:10.1016/j.epsl.2009.12.037
- Silver PG, Savage MK (1994) The interpretation of shear-wave splitting parameters in the presence of two anisotropic layers. *Geophys J Int* 119(3):949–963. doi:10.1111/j.1365-246X.1994.tb04027.x
- Unglert K, Savage MK, Fournier N, Ohkura T, Abe Y (2011) Shear wave splitting, vP/vS, and GPS during a time of enhanced activity at Aso caldera, Kyushu. *J Geophys Res-Sol Ea* 116(B11):203. doi:10.1029/2011JB008520
- Vargas-Bracamontes D, Neuberger J (2012) Interaction between regional and magma-induced stresses and their impact on volcano-tectonic seismicity. *J Volcanol Geotherm Res* 243–244:91–96. doi:10.1016/j.jvolgeores.2012.06.025
- Wuestefeld A, Verdon JP, Kendall J-M, Rutledge J, Clarke H, Wookey J (2012) Inferring rock fracture evolution during reservoir stimulation from seismic anisotropy. *Geophysics* 76(6):WC157–WC166. doi:10.1190/GEO2011-0057.1
- Zinke JC, Zoback MD (2000) Structure-related and stress-induced shear-wave velocity anisotropy: observations from microearthquakes near the Calaveras Fault in central California. *Bull Seismol Soc Am* 90(5):1305–1312. doi:10.1785/0119990099

Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring

Vincenzo Gattulli and Francesco Potenza
DICEAA - Dipartimento di Ingegneria Civile,
Edile-Architettura, Ambientale, CERFIS -
Centro di Ricerca e Formazione in Ingegneria
Sismica, University of L'Aquila, L'Aquila, Italy

Synonyms

Masonry failure; Monumental structures; Seismic damage; Seismic vulnerability; Structural analysis; Structural health monitoring

Introduction

The earthquake occurred on April 6, 2009, in Italy at L'Aquila has been a catastrophic event for both the city and the University of L'Aquila (Ceci et al. 2010). Since then, numerous scientific activities have accompanied both the immediate recovery and the long reconstruction program in different fields of earthquake science and engineering. In particular, the development of new

systems and technologies for both understanding and enhancing the structural behavior of significant historical palaces and churches, densely populating the L'Aquila city center, has attracted the attention of specialists and researchers coming from a broadband spectrum of scientific disciplines. This rich multidisciplinary approach has permitted the development of a new paradigm in the preparatory works necessary for planning the retrofitting and the reconstruction of ancient buildings. Therefore, the impact of new technologies in the area of observation, survey, testing, modeling, restoration, retrofitting, and monitoring of historic constructions and monuments, diffusely used at L'Aquila, merits to be discussed and briefly reported, here, as starting point for future novel findings and process optimization.

In this respect, it appears reasonable to mention the geophysical and geotechnical research efforts conducted to characterize various aspects of what has been observed and measured during foreshock, mainshock, and aftershock in the L'Aquila valley. Starting from a historical seismological study regarding the earthquakes that took place in the area of L'Aquila (central Italy) from the ancient Roman period to the late Middle Ages, the persistence and magnitude of earthquakes demonstrated to have a strong bearing on the economy and culture of the communities (Guidoboni et al. 2012). Furthermore, understanding the rupture slip distribution for the 2009 L'Aquila mainshock has allowed for the development of a complex high-resolution 3D FE model incorporating surface topography and rheological heterogeneities, deduced from real tomography (Volpe et al. 2012). The 3D approach provides a more concentrated and localized slip distribution on the rupture plane, evidencing a single area of high-slip release SE of the hypocenter. Moving from the underground level to the ground level, scientific studies have tried to correlate the distribution and the severity of the damage with the geological setting of the area, taking into account the characteristics of the building stock through time (Tertulliani et al. 2012). Strong-motion records and ambient noise measurements taken soon after the mainshock and during the entire aftershock

sequence showed variability in ground motion amplification throughout the city. General conclusions highlight that the building stock of the city suffered different levels of damage that can be partially explained by the combination of building vulnerability and surface geology. In particular, the observation of damages in RC buildings highlighted that the few collapse cases ($<1\%$ of the whole RC buildings stock) are mainly the consequence of structural deficiency, inadequate structural layout, and maybe site effects (Decanini et al. 2012). The majority of RC construction survived the earthquake with minor damages or with damages to nonstructural elements. The extensive observed damage to infill walls that can be associated to a significant amount of energy dissipation prevented the development of plastic hinges in RC columns, thus allowing the maintenance of structural integrity. Concerning the seismic demand, the study highlighted the noticeable difference between the elastic and inelastic demand, indicating that the displacement demand is moderate, associable to a non-extraordinary event, even if, in several cases, it overcomes the threshold of serviceability limit, producing economic losses and potential injuries and deaths (Ceci et al. 2013).

Much more complex than the damage scenario affecting the RC buildings is the case of technical observations made on which occurred specifically in historic constructions and monuments at L'Aquila (D'Ayala and Paganoni 2011; Da Porto et al. 2012; Indirli et al. 2013) and in the small historical centers of the valley (D'Ayala and Paganoni 2011; Carocci 2012; Indirli et al. 2013). These studies show that within the historic city center of L'Aquila, the number of total collapses observed was minor, while the proportions of partial collapses of the upper stories were perhaps greater than would have been expected given the level of shaking (D'Ayala and Paganoni 2011). In a minority of cases, total collapse of the facade was observed, and these were usually associated to substantial alteration of the original structure. Although the number of undamaged masonry buildings is very modest, the majority suffered either minor damage or repairable structural damage.

This situation allowed on one hand to identify and recognize collapse mechanisms and on the other hand to better correlate ultimate capacity to strong-motion characteristics and hence lends itself to a more detailed in situ survey and analytical assessment. In the specific case of churches, a statistical analysis has defined the percentages of possible mechanisms related to local construction practices and architectural features of the region. For each possible mechanism, the relative percentage of activation and calculated average damage allowed some conclusions to be drawn on the seismic vulnerability of the various structural and nonstructural elements making up a church. These conclusions may be regarded as widely valid and, if further corroborated by other post-earthquake surveys, may be used to define hierarchies of interventions, also on non-damaged structures (Da Porto et al. 2012). The study showed that it would be useful to introduce weights to damage mechanisms when calculating damage indices evaluated through survey form compilation, which is still a powerful tool in the management of emergency and post-emergency phases. Furthermore, it has been pointed out that the recorded high-peak ground accelerations reached high values, both in the horizontal and vertical directions. The simultaneous combination of these acceleration components, has produced a reduction of the masonry shear capacity, especially in the upper floors. Consequently, the widespread vulnerability of unreinforced masonry buildings resulted in extensive damage, collapse, and victims in a medium-size city. The underestimation of the earthquake actions is a considerable drawback that can be overcome only by integrating the study of earthquake scenarios with more accurate methodologies. Moreover, with regard to the structural vulnerability, most of the unreinforced masonry structures were built (and structurally modified with subsequent interventions) before the adoption of the 2002 updated requirements; thus, the percentage of unreinforced masonry buildings heavily damaged or collapsed in the epicenter area is relatively high (approximately 45 %). The first cause of widespread overall disruption originates from the very poor quality

of the disaggregated masonry, made by arbitrary materials, unable to resist horizontal forces. Several damage/collapse types have been encountered, classified, and associated with the simplified mechanisms foreseen by the Italian MEDEA procedure (Indirli et al. 2013).

The needs of simplified performance-based assessment procedures in order to support the interpretation of observed damage is evidenced in all recent earthquakes, such as those occurred in L'Aquila (2009), Christchurch (2010–2011), and Emilia Romagna (2012), which have caused not only a significant death toll and huge economic losses but also heavy damage to the worldwide cultural heritage (Parisi and Augenti 2013). Earthquake damage to monumental constructions has been discussed and critical issues affecting the seismic response of historic masonry structures are identified such as masonry quality, connections among structural elements, diaphragm flexibility, out-of-plane resistance of masonry walls, structural irregularities, wrong retrofit interventions, and earthquake ground motion characteristics.

Moreover, since 2009, the reconstruction idea of the city of L'Aquila and its surrounding was a main goal, primarily of the citizens and, with them, of the Italian Government. However, during every catastrophic earthquake, Italian citizens and institutions seem to face a series of typical problems due to bureaucratic burdens, corruption, excess legislation, political clientelism, poor performance, and weak public control, and they are the combined effects of cultural, institutional, and political factors at the governance level (Özerdem and Rufini 2013).

Notwithstanding several difficulties, numerous applied research activities have been conducted in the area of new and sophisticated methods for the assessment of masonry complex structures, which need retrofitting interventions. Most of the works conducted at L'Aquila in this area are characterized by a series of correlated activities ranging from historic search, survey, and analysis of the structure with regard to proposals for repair and seismic strengthening. In particular, several problems are generally encountered in the various levels of survey and

damage interpretation and those of modeling and analysis of the structural behavior of complex and clustered masonry buildings (Da Porto et al. 2013). The in-depth comparative studies between the surveyed damages and the results of structural assessment through different approaches have produced a great enhancement in the understanding of the complex behavior of masonry structures, evidencing that use of the “Natural Laboratory of Earth” is the best approach to evaluate the seismic behavior and the performance of structural systems as well as the failures occurring in reality and is a non-substitutable step for the advancement of knowledge in the field of seismic engineering (Brandonisio et al. 2013). For example, the observation made on large historical palace, such as the case of Palazzo Centi, has shown that, during the 2009 earthquake, differently from other historical masonry buildings, it responded “reasonably” well in the main structural parts also due to recent retrofit interventions carried out in 2003. These mechanical interventions proved to be the reason of the good seismic behavior of the building and that “saved” it from more serious damages demonstrating the effectiveness of the traditional techniques in designing and retrofitting masonry buildings (Lucibello et al. 2013).

The structural assessment and retrofitting of important churches in L’Aquila, some of them having a monumental value, has been elaborately studied. Results concerning the seismic behavior of masonry churches have shown that the dynamic excitation due to the seismic ground motion activates many vibration modes of the building structure, though all of them are characterized by small participation factors. Therefore, churches are not behaving through a superposition of global modes but more as dynamic interaction of localized modes. Consequently, in many examined case studies, the ratio between the total base shear and the church total weight ranged between 20 % and 30 %, evidencing a reduction with respect to the plateau value of the spectral acceleration provided by the Italian code. Therefore, appropriate choices of the force reduction factor should be adopted for these monumental buildings different from the case of

traditional residential buildings characterized by shear-type behavior. Further, the activation of many local modes also calls for retrofit interventions, which should “tie up” the building, thus avoiding the local failure modes that are often observed (Brandonisio et al. 2013).

Detailed 3D nonlinear numerical analyses have been performed on several palaces and churches and also on the macroelements composing them such as towers and facades. Linear and nonlinear structural analyses conducted by means of a full 3D finite element model have been used to partially explain the collapse of the transept area of the Basilica of S. Maria di Collemaggio. The transept structural system formed by arches, pendentives, and barrel vaults transferring the vertical loads onto the multilobed pillars has been demonstrated to be vulnerable with respect to both transversal and longitudinal actions. In particular, even in the case of longitudinal actions, the nonlinear static analysis evidences the presence of high tensile stresses in the arches connecting the pillars to the apse walls. This mechanism, observed in the numerical simulations and amplified by the strong vertical acceleration component registered close to the site, together with the minimal resistance of the material inside the core of the collapsed pillars, is the most probable explanation of the implosion of the transept structures in the church (Gattulli et al. 2013).

The bell tower of the San Pietro di Coppito Church in L’Aquila suddenly collapsed during the devastating earthquake, which occurred in April 2009. A series of analyses were performed to provide an ex post evaluation of the causes at the base of the collapse and give operative design information for the reconstruction (Milani et al. 2012). For this aim nonlinear static and limit analysis has been performed evidencing the role played by the actual geometry and by the masonry mechanical characteristics of the tower through reconstructing the failure mechanisms by both modeling approaches. More recently, in a similar modeling approach, an attempt has been made to take into account the effects of fiber reinforcement polymers positioned on a large facade of Palazzo Camponeschi,

as example of engineering analyses on the effectiveness of specific retrofitting intervention with new material on ancient masonry structure in order to augment their seismic capacity (Gattulli et al. 2014).

Finally, the dynamic behavior of masonry structures under three-dimensional shaking is not completely understood, and it needs specific investigations. In this sense, the direct observations and measuring of such behavior through the use of novel structural permanent monitoring systems may help to reach such a goal.

Before the 2009 L'Aquila earthquake, very few structures were equipped by permanent structural monitoring systems managed by the Department of Civil Protection (DPC). However, the response of the Pizzoli Town Hall during the mainshock has been recorded and analyzed by DPC, giving special insights on the potentiality of these systems for immediate evaluation of the damages occurring during an earthquake. Successively, the large amount of installed, temporary or permanent, different types of devices (accelerometers, smart wireless devices, displacement and velocity transducers, inclinometers, etc.) reach a number of around three hundred (300) or more, evidencing a large impact of this technology in the post-earthquake emergency phase, especially during the earthquake swarms (Gattulli 2013). In particular several monitoring systems have been installed in the emergency phase, to understand the occurred behavior in damaged building (Cimellaro et al. 2012; Foti et al. 2013), or during the construction of temporary scaffolding, in order to verify the efficacy of the added structural system especially in the case of monumental building (Russo 2013). Because of this scope, in many cases, the permanent monitoring has worked ranging from a limited number of hours up to several months. In other cases, the monitoring system is permanently installed on the structure for years; it can be used also to determine the change that will occur in the structural behavior during the reconstruction phase (Federici et al. 2012; Gattulli 2013).

In several cases, the structural monitoring system uses only accelerometers, starting from very few measurements to a larger number of devices

with different characteristics and sensitivities. Instead, more complex monitoring systems are used in monumental churches and buildings where accelerometers are accompanied by crackmeters, inclinometers, temperature measurement devices, etc. (Casarin et al. 2012; Russo 2013; Gattulli et al. 2014).

Within this broad range of activities, in the present chapter are summarized specific findings of the researchers of the University of L'Aquila constantly involved to define sustainable reconstruction interventions, in the framework of broad multidisciplinary collaboration, open to the use of innovative technologies and accompanied by valuable training of young scholars.

Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring

The historic center of L'Aquila, which was widely extended before the earthquake, was inhabited by about 7,000 residents and 8,000 students, constituting a large and widespread university campus. In this situation, dense of monuments of great value, characterized by a particular urban fabric, having an orthogonal system due to the Angioina division into lots even if modified by an accelerated stratification, made by previous earthquake, are placed more than 800 businesses, several professional offices, and headquarters of the administrations of various public institutions. This complicated system is developed for about 10 km, including historical centers, consolidated tissues, parts of the city under construction, and the natural park of Gran Sasso and Sirente-Velino. The earthquake has seriously compromised the system and this area corresponds, quite closely, to the so-called seismic crater in which the most important damage occurred. In the revitalization of this large zone affected by the earthquake, a central role is played by the recovery of the L'Aquila historical center accompanied by the similar ones in the neighboring small villages. Hence, the characterization of this system, using a large scale, starting from an urban point of view up to geomorphological and geotechnical vision, is an element of strong relevance also to understand what happens

to the smaller scale of the individual monument, integrating the specific observations with an overview of the problem. Indeed, the numerous campaigns for the characterization of the seismic behavior of the L'Aquila's valley made by a microzoning mesh of the historical center indicated a number of site effects which are very important for better understanding the L'Aquila earthquake and its effects. Recent studies, characterizing the underground of the historical center, point out a sudden transition between the L'Aquila breccia and silty clay loam through a highly detailed reconstruction.

The problem of the reconstruction of historical buildings and monuments, especially within the walls of the historic center, brings to investigate the main open issues concerning the interventions that ensure a real improvement. The structural behavior of the monuments, comprising a masonry structure, still today open to a full understatement, depends on factors such as the mechanical properties of the different materials composing the masonry walls, the features of the horizontal structure and their connection with the vertical ones, and the used construction methods. Therefore, it is important to realize an optimal use of the wealth in the technical and scientific knowledge, stored over the years, to evaluate properly the physical consistency and safety of historical buildings. The answer, of course, cannot be unique but should result in a critical and competent attitude. Some ideas may be received from the reading of a book dealing with the technical problem of preservation. For example, in a classical book on this issue is reported "... In this type of intervention is of particular importance the respect of the typological, structural and functional quality of the system, avoiding those transformations that alter the character." The matter is still very complicated to be solved. Indeed, the properties of fragility, heterogeneity, and anisotropy of masonry and their different typological varieties make it difficult to well describe unitarily their mechanical behavior. Recent technical codes take into account these uncertainties, allowing to use simplified numerical model, implemented "ad hoc," to evaluate the consistency of masonry structures, avoiding the use of

numerical checks that not always ensure a complete reliability. Moreover, the physical nature of masonry structures, product of improvements obtained following successive attempts, and empirical rules lead to the need to examine each case as a special case, making prohibitive to have a general characterization also due to the historic analyses which produces a series of unique cases. However, the entire reconstruction process must also take into account other problems such as energy efficiency, architectural restoration, and refunctionalization. All these actions have to make considering permanent monitoring of the monument so as to better schedule maintenance and following the target of long-term operation of the object.

A series of research groups of the University of L'Aquila have conducted numerous activities related to earthquake engineering immediately after the earthquake to support the city and the entire community strongly hit by the catastrophic event. Here, we attempt to report this huge effort motivated by the willingness to recover as soon as possible to the original state. This entails a path that goes from collapse observation to permanent monitoring, thus characterizing a broadband research activity and picking, according to the authors' opinion, the most significant results. Therefore, the presentation is subdivided in several specific fields.

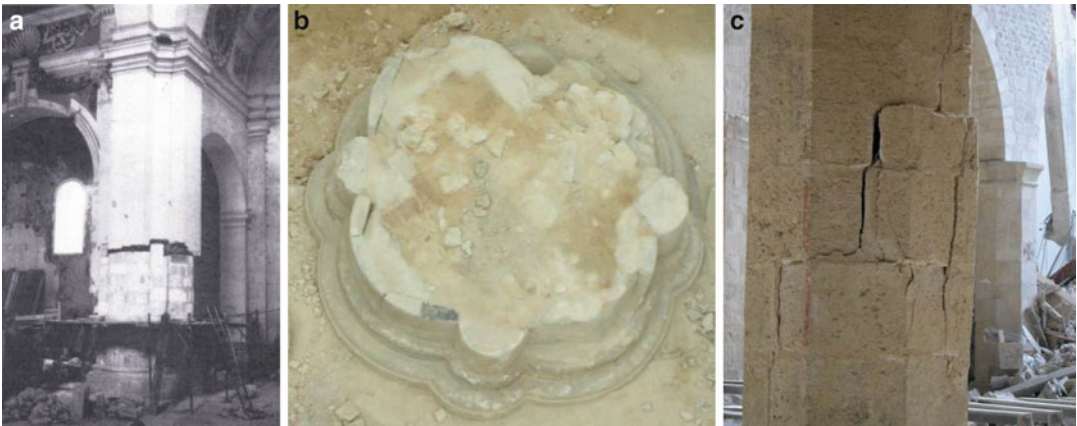
Collapse and Damage Observation and Survey

A series of partial or complete collapses have occurred in historical buildings at the center of L'Aquila, some of them having monumental value. Among them, one of the most studied is the collapse of the transept of the Basilica di S. Maria di Collemaggio (Gattulli et al. 2013).

The major damage suffered by the Basilica in 2009, indeed, was probably induced by a mechanism of implosion of all the structures that compose the transept: the great multilobed pillars, the triumphal arch and the wall above, the barrel vaults, the dome, and the roofing structures (Fig. 1). In particular, the failure occurred almost simultaneously in the two large pillars and the arches connecting the nave walls with the chancel, closed at the end by the apse, and it was



Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring,
Fig. 1 Basilica of S. Maria di Collemaggio transept collapse: (a) aerial view and (b) internal view



Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring,
Fig. 2 Pillars and columns of S. Maria di Collemaggio:

(a) shaping of the pillars in a multilobed medieval form, (b) material pillar core evidenced by the collapse, and (c) column damages

followed by the overtopping dome collapse. It should be pointed out that the transept is realized through a structural system in which the arch and vault thrusts, induced by permanent vertical loads, play an important stabilizing role. During the simultaneous vertical and horizontal actions of the 2009 earthquake, the equilibrium state in the masonry element was significantly altered. In particular, due to the proximity to the epicenter, the vertical and horizontal acceleration components had similar amplitudes.

Moretti's restoration intervention in the early 1970s was intended to remove the Baroque decorative features added to the Basilica and to unearth the "powerful original structure" of the thirteenth-

century plan, but only the works on the nave had been completed, leaving the transept with its original Baroque decoration. At the same time, he shaped the pillars in a multilobed form, removing material along the height starting from the original multilobed base (Fig. 2). In performing the restoration, the surface of the pillars was realized using blocks of freestone placed around the masonry core of the Baroque pillars having only an aesthetic function with no resistance increase. Small stones placed randomly, and probably coming from previous earthquake debris, were used for the core material.

Moreover, a subhorizontal fracture positioned at the height of around 2 m is present in the north

external wall, hosting the Holy Door, apparently created at the interface between the lower masonry part subjected to grout injection performed during the restoration works conducted in 1999 and the upper masonry part preserving the original workmanship.

Besides that, several of the threaded rods of the bracing system under the central nave roof were broken, almost all in the vicinity of the large crack on the north side wall of the church. In the north aisle, slippage of bolts designed to anchor the wooden trusses was noticed. Both of these cases of damage show the presence of transverse deformations of the longitudinal walls mainly located in the north, which may occur even after the transept collapse.

The damage pattern which affects the columns of the nave walls was developed in a complex manner; there were deep cracks due to heavy compression, initially appearing only in the central columns of the nave (Fig. 2c) and then progressively emerging, with the aftershocks, in almost all the others as well. The progression of damage was very similar to that found during the intervention in the last restoration of the 1970s, which consisted of inserting some missing fragments and grout with cement mortar in the opened joints. The only column that initially exhibited minor damage was the first one towards the transept, which had been completely rebuilt, although simply with brick facing of limited thickness and with an inner core done in small stones immersed in cement mortar.

In all the other columns, large cracks were visible. These columns, in distinction from the first ones, were made using stones shaped like truncated pyramids; these stones were juxtaposed, in an almost dry manner, in order to form the octagonal section. In most cases, cracks occurred in the stone blocks, despite the apparently high resistance of the material. The area of the apse was also characterized by severe damage. This area was affected by clear cracking in the mortar joints of the hewn stone blocks of the main apse, where there were also expelled blocks forming a wedge shape above the mullioned window. Observation of the interior reveals cracks in the apse vault, which is one of the most loaded

elements, a partial detachment, and the presence of permanent relative displacements, which occurred in one of the ribs. The lateral chapels suffered more serious damage. However, the study of the mechanical behavior of the monument cannot avoid to consider the low quality of the masonry, even present in peculiar elements as arches and vaults, exposed clearly in evidence by the collapse (Fig. 3a) and the peculiar realization of the masonry walls at L'Aquila that along the time due to the occurrence of several earthquakes is the product of different craftsmanship (Fig. 3b).

Laser Scanning Measurements

Besides the evident collapses easily visible, the relief of the damage, caused by the earthquake on April 6, 2009, has produced continuous and stimulating activities in damage observation and survey, producing a positive and critical approach about different aspects of the problem. Indeed, in particular regarding this specific area, the catastrophic event was an opportunity to apply the most advanced available technologies in the field. For the first time, the laser scanner technique was extensively used for the surveying of a great portion of L'Aquila territory and in particular for the historical buildings and monuments positioned within the old L'Aquila walls. The survey of the portion of the city hit by an earthquake has been considered, in several cases, a useful support for other structural analysis and conservation and restoration studies. Laser scanning has been used just after the earthquake in L'Aquila and during several summer schools that took place every year since the earthquake, to survey the facades of numerous buildings. The point cloud has been treated and filtered, in order to quantify and verify exactly the deformations due to the earthquake, compared with the original shape and geometry of the building. Laser scanning surveys allow the acquisition of the shape complexity of the building, with its irregularity, not only in some essential lines, as it is in topography. The whole information package is more complex than what is obtained with traditional techniques (topography and photogrammetry), but it is extremely useful for survey and measurements



Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring, Fig. 3 Masonry features at L'Aquila: (a) masonry quality

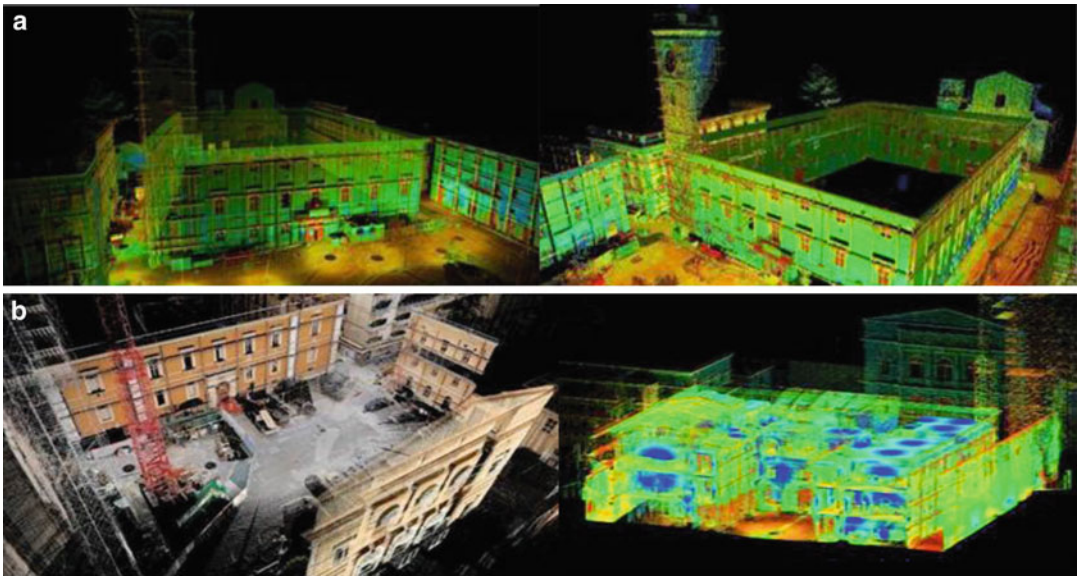
evidenced by the collapse and (b) masonry walls realized with material belonging to different periods and with different craftsmanship

of structural deformations and cracking. Moreover, the data acquisition system permits to acquire information on difficult geometry from a certain distance, avoiding difficulties related to traditional methods of measurement. Moreover, the University of L'Aquila placed important resources in the field, through several labs (CERFIS, CSE, etc.), renewing the technical equipment and using specialist teams, which operate for the geometric and damage relief. Data processing, made by Leica instrumentation (total station HDS2600 and software Leica Cyclone 7.2) and by a scanner velocity max of 1,016,727 point/s and flow rate in the range of ambiguity of 79 m, permits a return of three-dimensional model (see Fig. 4) from which plants, prospects, and sections can be obtained. The opportunity of having this great volume of data has stimulated to develop interactions between different disciplines, and so they constitute a new wealth of information. In particular the detailed reconstruction of crack path in the main facade of historical buildings (like Camponeschi Palace) allowed a verification of the numerical models to reproduce the observed behavior under

the earthquake and to predict the retrofitting improvements (Gattulli et al. 2014).

Applications of Infrared Thermography for the Investigation of Historic Structures

Infrared thermography is used as a nondestructive technique to identify the most significant inhomogeneity of materials and properties in the surface and structures that are difficult to appreciate by simple visual inspection but are instead detectable by the relief of anomalies associated to thermal phenomena. A portable instrumentation, which consists in thermal imagers, enables the observation of the materials constituting the main and secondary walls (e.g., infill, detachments, fills, and so on) and naturally the discontinuities. The infrared thermography is a noninvasive technique that uses infrared energy emitted spontaneously by any body or object: electromagnetic radiation whose wavelength depends on the temperature of the same body (Wien's law). Because the temperature of the bodies in ambient condition have a wavelength that belongs to the range of infrared spectrum ($\lambda = 0.3 \div 0.78 \mu\text{m}$), the technique is called



Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring, Fig. 4 Reconstruction of tridimensional geometrical models of the structures damaged by the earthquake

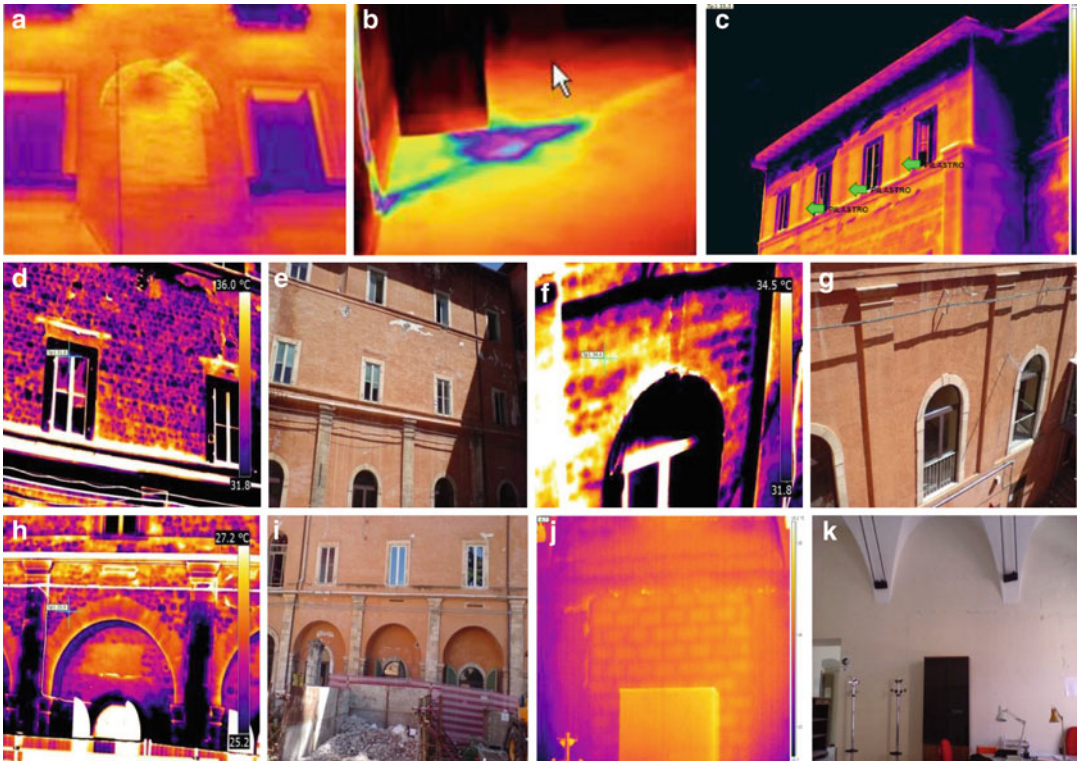
using laser scanner techniques: (a) Margherita Palace, headquarter of L'Aquila Municipality, and (b) De Amicis school building

infrared thermography. The value of radiant energy issued change when the objective of the camera passes from a point at a higher temperature to one with a lower temperature, and so also the wavelength of the radiation assumes two different values (Boltzmann's law). Where there is a thermal gradient, the scanner captures the energy variation that is linked with the body temperature through Stefan's law, and so it can determinate the value of emissivity's coefficient. Using a dedicated scanner optical system that investigates point by point all the recorded fields, the camera reproduces an electrical signal representative of the image's thermal information and proportional to its intensity, which will be shown in a kine-scope. In particular for each range of temperature variation, there will be a corresponding range of radiant intensity variation to which the instrument assigns a particular shade of color. The main characteristics of the instrumentation used are the following: minimum focusing distance, 0.3 m; temperature range, from -20°C to $+120^{\circ}\text{C}$; field of view (FOV), $24^{\circ} \times 18^{\circ}$; thermal sensitivity (NETD mK), $<40\text{ mK @ }+30^{\circ}\text{C}$ ($+86^{\circ}\text{F}$); accuracy, $\pm 1^{\circ}\text{C}$ o $\pm 1\%$ of readings;

resolution IR, 640×480 pixel; and spectral field, from 7.5 to 13 m. An extensive campaign of infrared thermography has been conducted to characterize the masonry of Camponeschi Palace, where the Letter and Philosophy Faculty is located. Figure 5 shows a series of findings. Moreover, this technique can be used without any interruption of the current activities and works. The fast data processing allows for the recognition of peculiar masonry zones with doors or windows under the plaster that are partially or completely closed, as well as with the presence of chimneys and underground utilities, masonry wall textures, warping of the floor, nonhomogeneity of the materials, structural damages, delamination of plaster as well as malfunctions of electrical and plumbing systems, infiltration of moisture, leaks, and thermal bridges.

Ultrasonic Testing

In ultrasonic testing (UT), very short ultrasonic pulse waves with center frequencies ranging from 0.1 to 15 MHz and occasionally up to 50 MHz are transmitted into materials to detect internal flaws



Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring, Fig. 5 Infrared thermography results obtained in Camponeschi Palace at L'Aquila: (a) evidence of opening closed with arched lintel, (b) failure of pipes, (c)

identification of concrete elements, (d–g) reading of textures walls under plaster, (h, i) reading of weaving arched lintels, (l, m) presence of openings in the internal masonry walls transversal to the perimeter walls

or to characterize materials. An ultrasound transducer connected to a diagnostic machine is passed over the object being inspected.

There are two methods of receiving the ultrasound waveform: reflection and attenuation. In reflection (or pulse-echo) mode, the transducer produces and receives pulsed waves when the “sound” is reflected back to the device. Reflected ultrasound comes from an interface, such as the back wall of the object or from an imperfection within the object. The diagnostic machine displays these results in the form of a signal with amplitudes representing the intensity of the reflection and the distance and registering the arrival time of the reflection. In attenuation (or through-transmission) mode, a transmitter sends ultrasound through one surface, and a separate receiver detects the transmitted signal on another surface after traveling through the

medium. Imperfections or other conditions in the space between the transmitter and receiver reduce the amount of sound transmitted, thus revealing their presence.

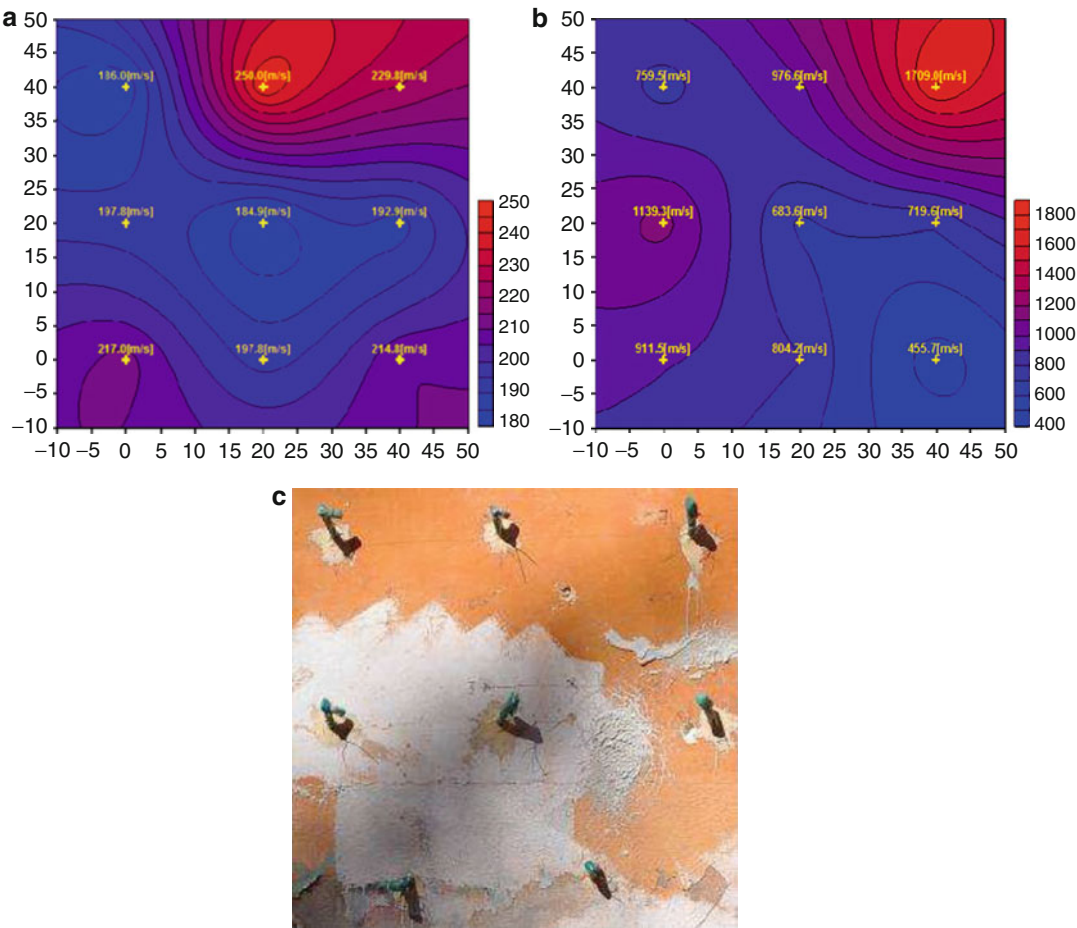
Ultrasonic testing, used on masonry, permits to obtain information about the homogeneity and composition of the masonry wall of the analyzed building. This technique has the benefit to be a nondestructive test and can be done in a very rapid and simple way in wide region of structural parts, but on the contrary, this large volume of data needs a specific interpretation. For this reason, it is important to combine the sonic tests with destructive ones, performed in a limited and significant structural zone, to have a well-specific and useful reference which defines the right relation between the speed of sound and the mechanical characteristics of the investigated masonry. Therefore, it is useful to highlight the main

technical properties of the instrumentation used in a lot of campaign tests at L'Aquila by the University:

- Acquisition, measuring range from 100 mv to 20 v
- Time bases from 20 ns to 81.9 s
- Sample resolution 8 bit
- Samples for event 8,192 for contact measurements, 640 for log, bandwidth 50 mhz, filter for ultrasounds: central frequency 50 kHz, measurement channel 1
- Probes through contact and sonic test with hammer: resonant frequency 53 kHz, diameter

48 mm, maximum frequency for pulse emission: 1 per second, minimum measurement pitch: 10 mm

In Fig. 3 images obtained from sonic tests are reported. From the comparison of the graphs reported in Fig. 3a, b, the distribution of pulse wave speed in the same portion of masonry before and after fluid mortar injection grouting is highlighted. In the first case the maximum speed is 250 m/s, while in the second case it grows up to 1,709 m/s. The different velocity transmission indicates clearly the increase of stiffness due to the grouting (Fig. 6).



Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring, Fig. 6 Sonic tests on a wall of Camponeschi Palace:

distribution of the velocity of propagation (a) before and (b) after the fluid mortar injection grouting and (c) locations of injection pipes

Georadar (Ground Penetrating Radar: GPR) Tests

The GPR methodology that employs electromagnetic waves permits the boundary definition of mechanical characteristics of the investigated body by the identification of the interface between layers that have different resistivity and dielectric constant. Structural degradation is classified (four classes of QRI “Quality Radar Index”) by a qualitative analysis of the power radar signal behavior with the depth by both radar anomalies in the section (presence/absence, point 0 and 1) and energy recovery, intense and localized (also called “pick”) in the graphs (point 1 for the pick up to 6 dB, 2 point up to 12 dB, and 3 point over 13 dB). The total QRI is the sum of the two scores described above. The method is calibrated using correlation between radar data and a direct destructive test made on hundred scans acquired on masonry tester. This method permits a direct or relative classification depending on whether there are direct data on each structure or not. The relative classification is divided in four classes: QRI 4, class of low quality; QRI 3, class of medium-low quality; QRI 2, class of medium quality; and QRI 1, class of medium-high quality. In Table 1 the main results of the georadar tests are reported and the obtained classification, in particular, is highlighted in the last column.

In Fig. 7a a grid in the scanned surface of a portion of a masonry wall of the Camponeschi Palace is presented while the obtained results through the georadar test are described in

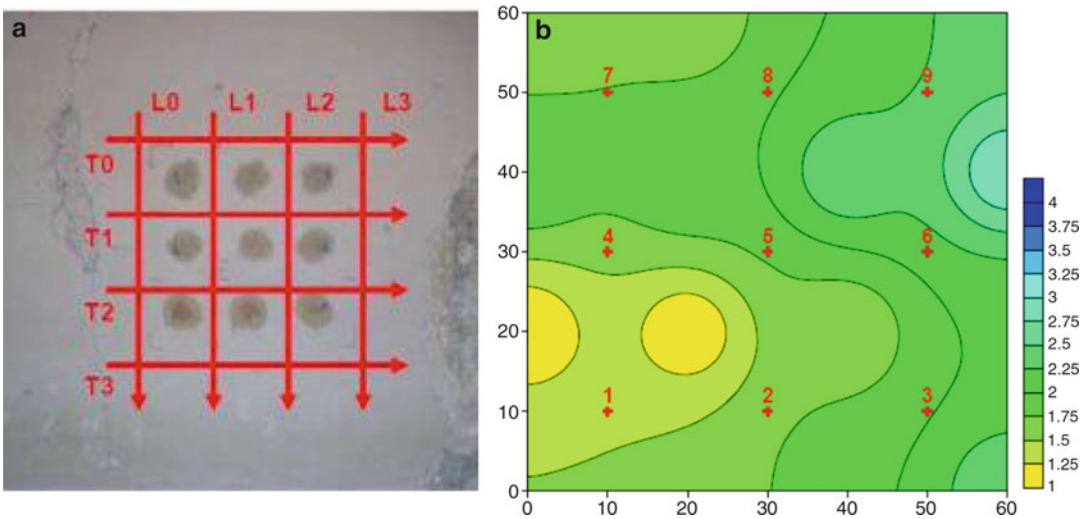
Fig. 7b in terms of QRI distribution, evidencing a relevant variability of QRI ranging from 1 to 3.25.

Removal of Plaster and Endoscopic Tests

Among minimal invasive tests, the removal of plaster rendered a zone of inhomogeneity, eventually pointed out by infrared thermography, visibly permitting to check the true nature of the walls and of the composing elements. Successively, endoscopic tests have been usually made in several historical masonry walls at L’Aquila. Endoscopy is simply an extension of the essential visual survey into areas inaccessible to the naked eye. The equipment ranges from relatively simple borescopes, consisting of a light source, a small-diameter rigid tube with built-in optics, and an eyepiece, to complex controllable systems with numerous specialized attachments. By drilling a hole (normally less than 12 mm) and inserting the tube, it is possible to inspect voids under floors or behind paneling, for example. Any hidden problems such as fungal growth can, in theory, be identified. The more sophisticated and expensive equipment is fully flexible and can be steered by wires built into the casing. Systems are available down to 6 mm in diameter, and more specialized systems may reach less than 2 mm. It is possible to attach still or video cameras to the eyepiece to record the findings. The theory is fairly simple, but in practice it can be very difficult to retain a sense of scale of the observed image and keep track of the location and orientation of the tip. The focal range, depth of field, and strength of

Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring, Table 1 Main information deduced by GPR tests

Scansions		GPR graphics			GPR sections	
		Max	Max height of the picks equalized	Point	Presence/absence	
Codex	Length	db				
LEA10000	0.6	5	5.0	1	0	1
LEA10001	0.6	2	2.0	1	0	1
LEA10002	0.6	5	5.0	1	1	2
LEA10003	0.6	7	7.0	2	1	3
TEA10000	0.6	9	9.0	2	0	2
TEA10001	0.6	8	8.0	2	1	3
TEA10002	0.6	5	5.0	1	0	1
TEA10003	0.6	7	7.0	2	0	2



Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring, Fig. 7 Georadar tests on masonry of Camponeschi Palace: (a) grid scans and (b) QRI distribution

light are greatly reduced in the smaller diameter systems. It is not unknown for insulation lagging to be misidentified as dry-rot. However, reliable reconstruction of the stratification in the main masonry walls has been obtained, to know the construction technique and to assess qualitatively the physical conditions. The employed instrumentation has permitted to register short movies of what is visible inside the cavities, producing a large quantity of data in digital form to be processed.

Flat-Jack Tests

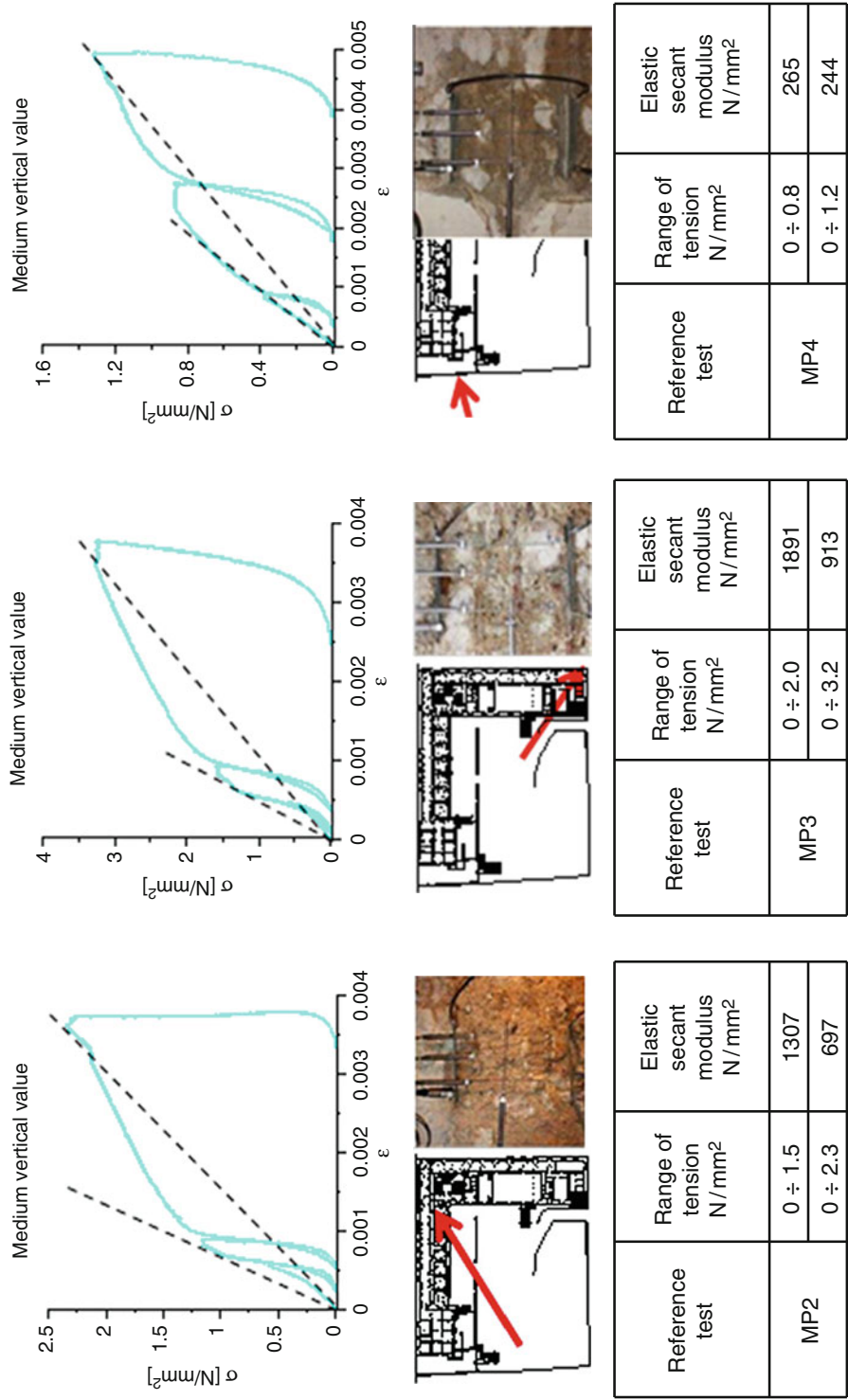
Several tests of single and double flat jacks were carried in different historic buildings in the center of L'Aquila including Camponeschi Palace, De Amicis school, the headquarter of the Carispaq Bank and Margherita Palace, and others. Moreover, diagonal tests have also been performed to characterize the shear strength in selected masonry panel individuated in existing walls in order to characterize the mechanical properties of the main typology of historical masonry walls at L'Aquila. A probabilistic characterization of the available data on the mechanical features of the historical masonry at L'Aquila is going on in order to have the possibility to investigate the structural assessment of historical buildings,

taking into account explicitly the inherent uncertainties of the problem. A more reliable description of the seismic behavior of the masonry mechanical characteristic will realize a well-balanced design of retrofitting interventions.

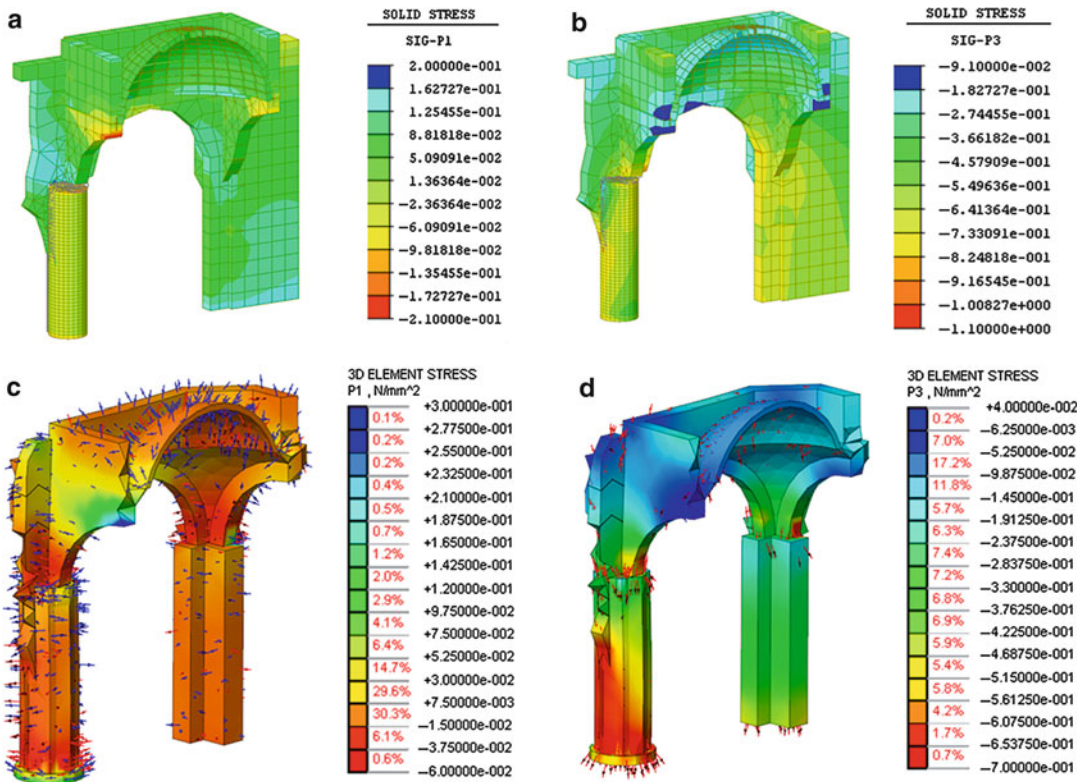
Figure 8 summarizes the results obtained through double flat-jack tests realized at different locations of the Camponeschi Palace. The three performed tests evidence the variability in both elastic modulus and measured strength through the adopted testing procedure. Indeed, in the analyzed case, the average strength value is 2.47 N/mm^2 , while the variance is 0.736 N/mm^2 , evidencing the large scatter in the material properties belonging to the same historical building.

Computational Mechanics and Masonry Structural Modeling

Structural analysis and numerical modeling of the seismic behavior of masonry structures is still an open research topic, especially in Italy and in particular after the L'Aquila earthquake. Relevant issues pertain to the inherent uncertainties related to the evaluation of the seismic adequacy of complex masonry buildings through reliable numerical models. Starting from the limit analysis in which the simplified assumption of describing the incipient collapse behavior through chains



Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring, Fig. 8 Results of double flat-jack tests on Camponeschi Palace



Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring, Fig. 9 Principal stresses due to static loads in the

transept: model A, (a) maximum P1 and (b) minimum P3; model B, (c) maximum P1 and (d) minimum P3

of rigid masonry blocks, a series of different modeling approaches have been experimented within the activities of the computational Mechanics Lab of the University of L'Aquila directed by the first author. On this respect, the experience gained at L'Aquila benefits from the large amount of data available for possible comparison with the numerical simulations. Therefore, potentiality and drawbacks of several modeling assumptions have been evidenced along the path, which have shown to be particularly useful also to enrich the tools for retrofitting design. The structural modeling of the Basilica of S. Maria di Collemaggio constitutes a relevant case study in this field. Figure 9 reports, synthetically, the results of two stress analyses induced by live loads conducted by means of two different finite element models of the entire Basilica: in the first case (model A), four-node quadrilateral isoparametric plate-shell elements have been

used to model the main walls in which a prevalent bidimensional behavior has been recognized (Fig. 9a, b) (see also Gattulli et al. 2013); in a second case (model B), a more refined model has been considered in which all the masonry elements have been modeled using eight-node isoparametric solid elements (Fig. 9c, d). It is evident as the stress flux is more clearly defined in the arches and in the pillars by the more refined model, even if the level of stress amplitude is comparable in the two cases. At the same time, the analysis evidences that in the pillars is present an important flexural deformation induced by the thrust differences in the longitudinal direction between the last arch belonging to the nave walls and the arch of the transept on which is laying the dome and the vault. Even if the results are obtained under the hypothesis of modeling the masonry material as an elastic and homogeneous solid, they evidence the role played by the

three-dimensional geometry and the consequent internal action distribution. Figure 9c evidences as the maximum tension flux in the column is prevalently radial, while the minimum one in Fig. 9d is vertical.

A different activity has been carried out on the Camponeschi Palace which represents a good example of how damage observation and experimental survey, performed using the techniques described above, can be a valuable aid in supporting the modeling choices for both damage interpretation and simulating the effect of retrofitting interventions.

The Camponeschi Palace is an important historical building located in the historical city center of L'Aquila. The palace was built near the end of the sixteenth century by the Jesuit order that established and hosted in the structure an institution of higher education. Several modifications to the original configuration and structural interventions have been made during the years. The palace sets upon a rather sloping site and it is characterized by an L-shaped plan with two arms of equal length delimiting an internal courtyard. The building was hosting the Faculty of Philosophy and Letters of University of L'Aquila at the time of the 2009 earthquake.

Due to the relevant structural damage, which occurred, the Department of Civil Protection classified the palace structurally unsafe. The crack pattern of the facade due to the 2009 L'Aquila mainshock and successive earthquake swarms was detected through automated laser-based surveys. The crack pattern distribution of the main facade is shown in Fig. 10a. The major damage was concentrated on the first story. On the left side of the facade, subvertical and diagonal cracks developed at the end sections of the spandrels above and below the openings.

Single and double flat-jack tests were carried out on site in order to evaluate the compressive strength and the deformability properties of the masonry for the entire building.

A numerical model was built specifically to simulate the nonlinear static behavior of the facade and of a retrofitting system using fiber reinforcement polymers for the masonry panels. Eight-node isoparametric plane stress elements

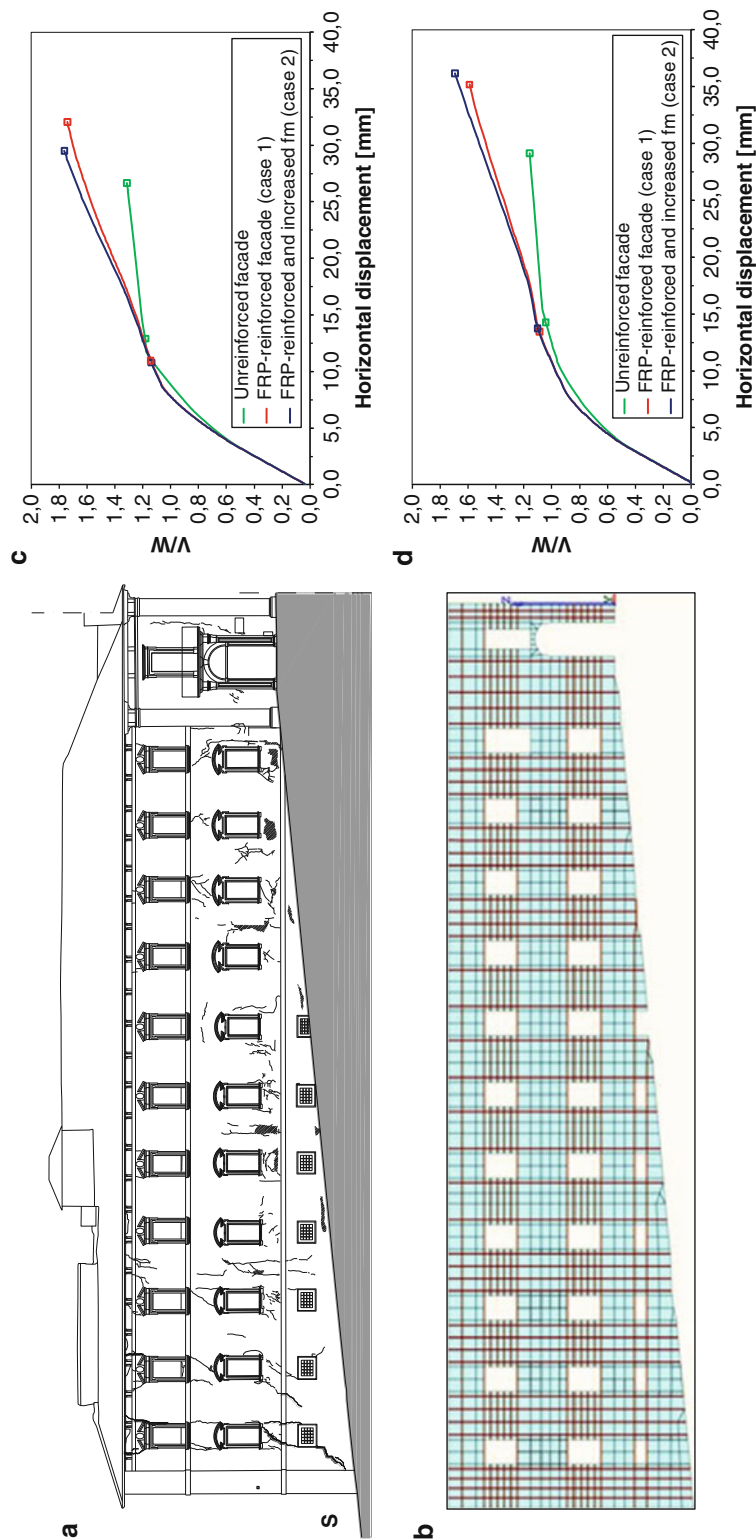
with a 3×3 Gauss integration scheme were used. Figure 10b shows the adopted mesh. The incremental/iterative solution procedure employed is a modified Newton–Raphson approach. A constitutive law, comprising a parabolic hardening rule and a parabolic exponential softening branch, modeled the compression behavior of the masonry; a linear hardening branch followed by a linear softening branch characterized the tension behavior. The FRP reinforcement was modeled using truss elements directly connected to the nodes of the mesh of the panels without using interface elements. Thus, perfect adhesion between the nodes of trusses and the corresponding nodes of the mesh was considered. The strips were considered not able to carry compression loads. The tensile behavior of the FRP was represented with a stress–strain relationship, characterized by a linear elastic behavior in tension, followed by exponential degradation.

Nonlinear incremental static (pushover) analyses were carried out on unreinforced and reinforced facade, separately for the positive (Fig. 10c) and negative x direction (the horizontal one) (Fig. 10d) and for both the unreinforced and FRP-reinforced masonry (Gattulli et al. 2014).

In general, a correct and reliable implementation of the model allows performing in-depth analyses of the expected structural behavior under seismic loads enabling robust design choices.

Structural Monitoring for Historic Structures

Structural health monitoring is an emerging tool for reliable structural assessment. Efficient monitoring programs may help in the characterization of the progressive decay of structural performance, both in the short and in the long term, providing useful information for optimized maintenance and safety. Generally, electronic equipment deployed with the purpose of structural health monitoring is implemented via wired systems. In this case, the development of the system can be limited by a series of factors such as the high cost of measurement equipment, difficulty in installation due to large instruments size, and need of a wired communication and power infrastructure. In recent years, great attention has been



Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring, Fig. 10 Camponeschi Palace: (a) damage pattern on the facade (b) mesh of the FEM model with the FRP layout; numerical pushover curves for horizontal loads (c) +x direction and (d) -x direction

paid to the development of wireless sensor networks that are able to overcome several problems linked to the development of the entire layout but evidencing other issues related to the new technology. One of which is connected to the collection and transmission of a great quantity of data. In fact relevant information can be obtained by the registrations of the vibration data induced by ambient sources, wind, earthquakes with low intensity, operation of machines, and traffic. The sensors are placed in specific points of the structure and appropriately selected and the choice of which is based, for example, on numerical modeling. Using the acquired data, it is possible to identify dynamical models, which predict the observed behavior on the basis of error minimization. Differences in the identified systems produced by data acquired in different periods separated by a long interval of time may permit to identify relevant changes in the behavior of the systems which can be linked to degradation of the structural performance or incurred damages. For example, in largely damped, like masonry, structures, it is convenient to use high sampling times for acquisition as opposed to frame structures. Although the interesting structural modes can be placed within an interval 0–10 Hz, the structural response may be highly damped and attenuated with a reasonable amplitude level contained in few seconds. The acquired time histories under seismic action will be composed of a few samples that make difficult the reconstruction of the frequency spectrum. Some important benefits are that smart sensors include microprocessors that are able to run some basic algorithmic tasks and then store some condensed information to be exchanged between the other sensors (nodes) that comprise the wireless network (WN).

The long-term activity at the University of L'Aquila aims to develop innovative automated techniques, which permit to automatically update the appropriately selected models, which in turn may address questions regarding the seismic safety of the monitored structures. In particular, following a parametric identification approach, using the distributed available measurements from sensors embedded in the

reinforcements, the mechanical parameters characterizing the local behavior of the reinforcement either in the reinforcing fibers and in their support may be opportunely monitored. Subsequently, through information data fusion made available by arrays of sensitized reinforcements distributed on a masonry macroelement, mechanical parameters influencing the global behavior may be identified. Possible developments concern the use of measurements at different levels of excitation amplitudes, which are available in seismic areas. They may be used to define a methodology for identifying nonlinear behavior, distinguishing between scenarios where damage occurs in the fiber, in the support, or in the masonry structure.

The implementation of a full SHM system in historical buildings is still an open problem especially in aspects relating to the correct fusion of vibration measurements, integrated with other type of measurements. For example, in concrete structures it has proven useful to follow the trend of humidity and temperature in order to evaluate deterioration and updated material performances. In this manner early assessment and warning of incipient problems may be used to drive quick interventions or the planning and scheduling of the maintenance programs. To the contrary, in monumental structures often numerous cracks are present which can be the objective of permanent static monitoring. Indeed, following the slow opening and closing of the crack may be used as an important correlation measurement for validating information provided for the identification of other parameters. The opening and closing may depend on the cyclic trend of the average temperature corresponding to a certain time interval (day, month, season). Similarly to this trend, it may possible to observe a reduction or increase of stiffness or a change of configuration that may induce also a cyclic trend in the identification of natural frequencies.

Of course, this set of actions permits both a rationalization of the cost and an improvement of the service-life predictions, potentially also for monumental structures, especially if they have been recently restored and retrofitted with a consequent important investment.

In general there are no particular rules for the design of a fully operating SHM system, but each case is unique and it requires particular care especially with respect to the goals to be achieved. The observation and the investigation on the damage reported by the structure after a seismic event can lead to a customized system able to provide specific information for continuously assessing structural condition. Correlations among obtained results should be sought for and only after analyzing all available information; the responsible team, formed by professionals with different skills (structural, geotechnical, architectural, electronic, and informatics), on the basis of their experience and sensibility, can formulate a final diagnosis and define the actions to be performed.

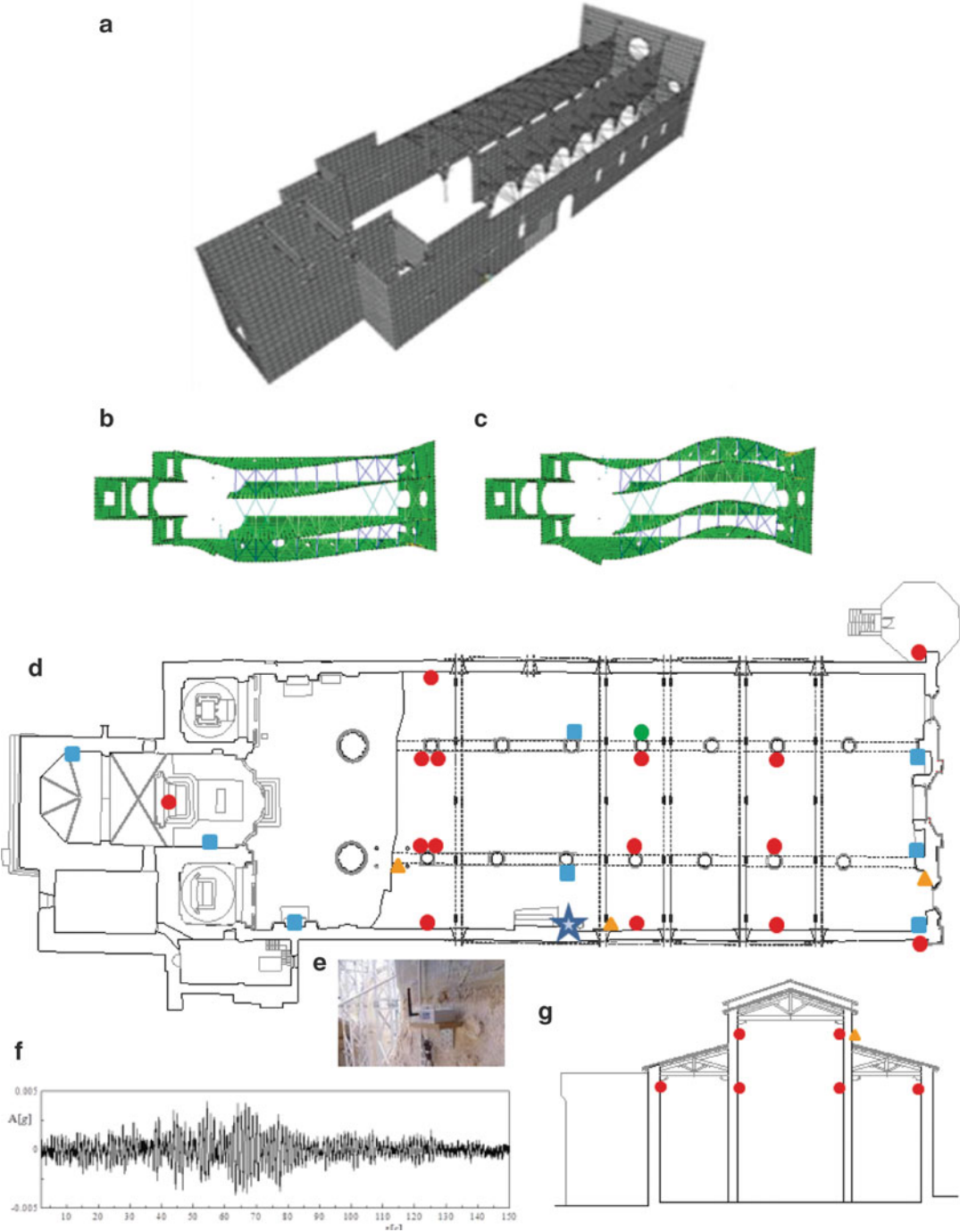
Within such a context, in the first phase of the SHM system deployment, at the Basilica di Santa Maria di Collemaggio, different actions have been performed including design, implementation, installation, and test of a network monitoring accelerations. The main goal was an accurate measurement of the basic structural response, both to environmental actions and to low-amplitude seismic events, which occasionally continue to occur at the site. A wireless sensor network was used as a platform combining a MEMS tri-axial accelerometer and also sensors to measure humidity, temperature, and luminosity (Gattulli et al. 2014). The proper positioning of the sensors has been designed through the study of different numerical models accounting for the interaction between the main wall element and the temporary scaffolding (Fig. 11a). In particular, the dynamic behavior can be observed looking at the principal modal characteristics (frequency, modal shape, and percentage of participating mass) of the linear dynamic response (Fig. 11b, c). The analyses suggest, for example, that, by placing the sensors along the internal longitudinal facade, it is possible to survey the transversal displacements especially during a seismic event. Until today, after almost 3 years of permanent monitoring, the system is still able to capture relevant data during the occurrence of several seismic events coming from far and near fields. Generally, the registered maximum

response amplitude in the acquired time history (one of this shown in Fig. 11f) determines the possibility to get information regarding the main modal features.

Nowadays, in order to derive the processed information, from experimental data, different modal identification procedures can be used which are generally classified based on whether they operate in the time domain or in the frequency domain. However, the amount of contained information is relatively independent of the analysis domain.

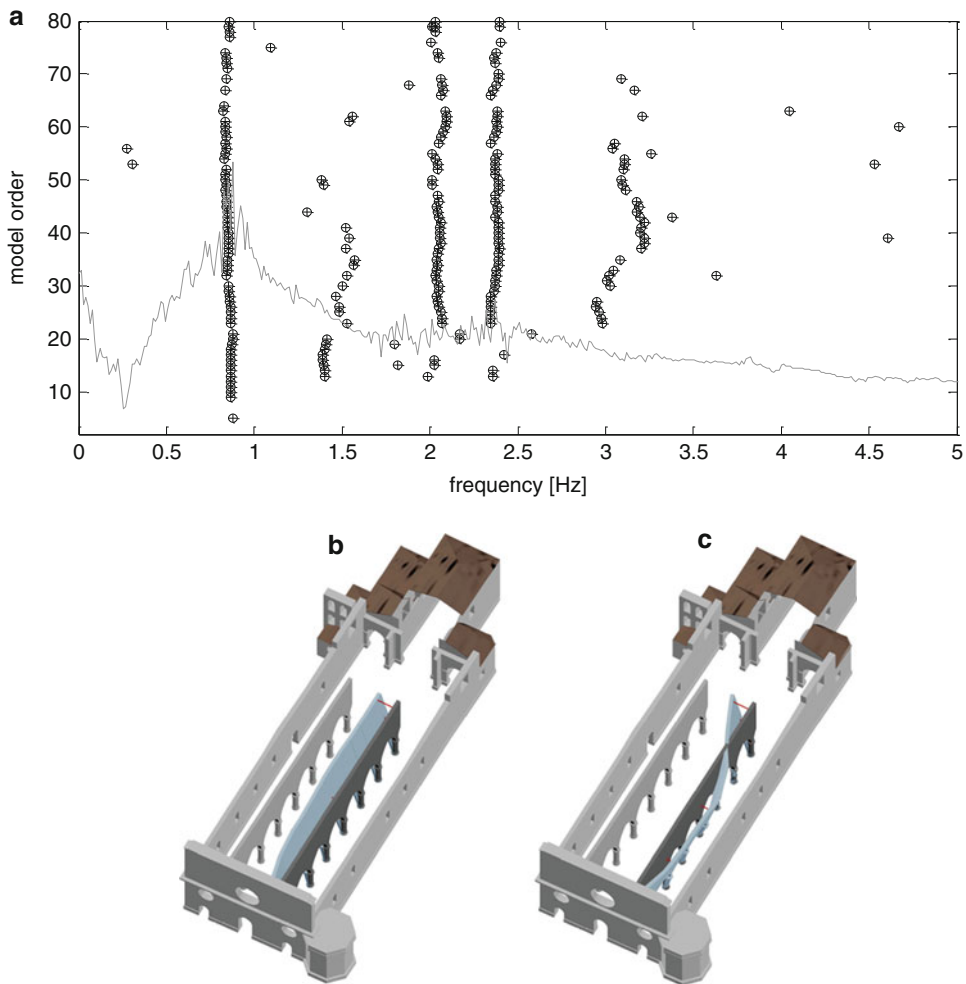
Among several procedures probably, the two most popular algorithms are the Enhanced Frequency Domain Decomposition (EFDD) and the Stochastic Subspace Identification (SSI) (Peeters and De Roeck 2001), and the first operates in the frequency domain and the second in the time domain.

The two procedures assume that the system input (loads) is white noise that can be thought as a representative model of environmental noise. However, even if the adoption of the white noise model in the case of a seismic event may be questionable, the straightforward applications of the abovementioned procedures to response data acquired during small earthquake still allow to derive valuable information especially if special care is taken, such as in the case of the studied system at the Basilica (Gattulli et al. 2015a, b). Indeed, here, using the SSI procedure, for example, the evaluation of the identified stable poles may include dominant components of the input, producing possible errors or misinterpretation of the dynamic behavior. However, the direct comparison with results obtained by numerical models driven by engineering judgment permits to extract realistic structural modal data. Moreover, the robustness of this data can be pursued through the use of several different registrations generated by earthquakes with different characteristics dependent on the source location. Following the reliable identification of the modal characteristics, the second step is the updating of the numerical model parameters in order to minimize the error between measured and numerical modal features. In the error index formulation, natural frequencies and/or modal shapes



Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring, Fig. 11 Structural monitoring at the Basilica of S. Maria di Collemaggio: (a) finite element model accounting for the interaction between the main wall

element and the temporary scaffolding; (b) first (1.09 Hz) and (c) second (1.97 Hz) mode shape; layout of the multifunction wireless sensor network: (d) plant, (e) sensor, (g) section, and (f) acquired registration during a seismic event



Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring, Fig. 12 Identified modal model by structural monitoring:

(a) stabilization diagram of the SSI procedure; (b) first (0.90 Hz) and (c) second (1.45 Hz) identified mode shape

may be considered. In the first case, a simple difference in percentage is generally considered, while the Modal Assurance Criteria (MAC) is used to compare measured and numerical modal shapes. The MAC criterion indicator permits to evaluate the differences between similar modes but also the cross-orthogonality between different numerical and experimental modal shapes. As an example, Fig. 12a illustrates the stabilization diagram obtained from the time histories acquired by the wireless sensor network using SSI during a seismic event. The ordinates include the model order, which simply refer to the size of

the Hankel matrix considered in the identification process, while the abscissa is reporting the frequency; the circles with a cross inside indicate stable modes, while the others, colored in gray, are unstable. At least one stable mode is present at different frequencies, but in the case of two specific frequencies, stable modes seem to be present at almost all different model orders considered in the analysis. These stable poles are present at 0.90 Hz and in the range between 1.45 and 2 Hz. In the latter case, a comparison with numerical models permits possible interpretation of the modal identification. The first two

identified modes are illustrated in Fig. 12b, c illustrating the deformed shape of the portion of the Basilica in which the response used in the identification procedure has been registered. The second phase of the monitoring system development involves the deployment of crackmeter and inclinometer sensors at the Basilica. Figure 11d summarizes the employed instrumentation, where the red circles indicate the wireless sensor network relative to the accelerometer; the blue squares and the orange triangles correspond to the crackmeters and the inclinometers, respectively. Moreover, the green circles refer to an accelerometer placed at the base of a pillar necessary for acquiring the acceleration of the seismic input, while the star indicates the positioning of the node gateway that collects the data of each sensor and sends them to the server. The data fusion of the two sets of information is currently under investigation.

Summary

The earthquake, which occurred on April 6, 2009, has been a catastrophic event for both the city and the University of L'Aquila. Nevertheless, the disaster has transformed itself into a tremendous opportunity to revitalize the area, providing the potential for the national and international scientific community to test the effectiveness of new systems and technologies and consequently to establish main directives for the monitoring of monumental structures. The applicability of new technologies in the areas of observation, surveying, testing, modeling, restoration, retrofitting, and monitoring of ancient monuments, of which the city center is densely populated, is extensively documented herein, through an organized outline and a series of case studies.

Cross-References

- [Ambient Vibration Testing of Cultural Heritage Structures](#)
- [Damage to Ancient Buildings from Earthquakes](#)

- [Masonry Modeling](#)
- [Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used](#)
- [Seismic Vulnerability Assessment: Masonry Structures](#)
- [Stochastic Structural Identification from Vibrational and Environmental Data](#)

References

- Brandonisio G, Lucibello G, Mele E, Luca AD (2013) Damage and performance evaluation of masonry churches in the 2009 L'Aquila earthquake. *Eng Fail Anal* 34:693–714
- Carocci CF (2012) Small centres damaged by 2009 L'Aquila earthquake: on site analyses of historical masonry aggregates. *Bull Earthq Eng* 10(1):45–71
- Casarin F, Benetta MD, Modena C, da Porto F, Valluzzi MR, Cantini L, Tedeschi C, Condoleo P (2012) Structural monitoring and investigation campaign in the church of St. Giuseppe dei minimi in L'Aquila, after the 6th of April 2009 earthquake. RILEM book series, Springer, Berlin, vol 6. pp 1187–1193
- Ceci AM, Contento A, Fanale L, Galeota D, Gattulli V, Lepidi M, Potenza F (2010) Structural performance of the historic and modern buildings of the University of L'Aquila during the seismic events of April 2009. *Eng Struct* 32(7):1899–1924
- Ceci AM, Gattulli V, Potenza F (2013) Serviceability and damage scenario in RC irregular structures: post-earthquake observations and modelling predictions. *J Perform Constr Facil* 27(1):98–115
- Cimellaro GP, Piantà S, De Stefano A (2012) Output-only modal identification of ancient L'Aquila city hall and civic tower. *J Struct Eng* 138(4):481–491
- D'Ayala DF, Paganoni S (2011) Assessment and analysis of damage in L'Aquila historic city centre after 6th April 2009. *Bull Earthq Eng* 9(1):81–104
- Da Porto F, Silva B, Costa C, Modena C (2012) Macro-scale analysis of damage to churches after earthquake in Abruzzo (Italy) on April 6, 2009. *J Earthq Eng* 16(6):739–758
- Da Porto F, Munari M, Prota A, Modena C (2013) Analysis and repair of clustered buildings: case study of a block in the historic city centre of L'Aquila (Central Italy). *Construct Build Mater* 38:1221–1237
- Decanini LD, Liberatore L, Mollaioli F (2012) Damage potential of the 2009 L'Aquila, Italy, earthquake. *J Earthq Tsunami* 6(3):art. no 1250032, 32 pp
- Federici F, Graziosi F, Faccio M, Gattulli V, Lepidi M, Potenza F (2012) An integrated approach to the design of wireless sensor networks for structural health monitoring. *Int J Distrib Sens Netw Art. No.* 594842, 16 pp
- Foti D, Gattulli V, Potenza F (2013) Output-only identification and model updating by dynamic testing in

- unfavorable conditions of a seismically damaged building. *Comput Aided Civ Infrastruct Eng* 29 (9):659–675
- Gattulli V (2013) Advanced applications in the field of structural control and health monitoring after the 2009 L'Aquila earthquake. In: Sebastiano D'Amico (ed) *Engineering seismology, geotechnical and structural earthquake engineering*. InTech, Croatia, pp 208–235. ISBN: 978-953-51-1038-5, doi:10.5772/55438
- Gattulli V, Antonacci E, Vestroni F (2013) Field observations and failure analysis of the Basilica S. Maria di Collemaggio after the 2009 L'Aquila earthquake. *Eng Fail Anal* 34:715–734
- Gattulli V, Lampis G, Marcari G, Paolone A (2014) Simple modeling of FRP reinforcement: from masonry panels to a historic palace facade. *Eng Struct* 75:604–618
- Gattulli V, Potenza F, Federici F, Graziosi F, Colarieti A, Faccio M (2015a) Distributed structural monitoring for a smart city in a seismic area. *Key Eng Mater* 628:123–135
- Gattulli V, Potenza F, Lepidi M, Federici F, Graziosi F, Colarieti A (2015b) Seismic structural monitoring by wireless sensor networks of the Basilica S. Maria di Collemaggio after the transept collapse. *Earthq Eng Struct Dyn* (in preparation)
- Guidoboni E, Comastri A, Mariotti D, Ciuccarelli C, Bianchi MG (2012) Ancient and medieval earthquakes in the area of L'Aquila (northwestern Abruzzo, central Italy), A.D. 1–1500: a critical revision of the historical and archaeological data. *Bull Seismol Soc Am* 102(4):1600–1617
- Indirli M, Kouris LAS, Formisano A, Borg RP, Mazzolani FM (2013) Seismic damage assessment of unreinforced masonry structures after the Abruzzo 2009 earthquake: the case study of the historical centers of L'Aquila and Castelvecchio Subequo. *Int J Archit Herit* 7(5):536–578
- Lucibello G, Brandonisio G, Mele E, Luca AD (2013) Seismic damage and performance of Palazzo Centi after L'Aquila earthquake: a paradigmatic case study of effectiveness of mechanical steel ties. *Eng Fail Anal* 34:407–430
- Milani G, Russo S, Pizzolato M, Tralli A (2012) Seismic behavior of the San Pietro di Coppito Church bell tower in L'Aquila, Italy. *Open Civ Eng J* 6(1):131–147
- Özderdem A, Rufini G (2013) L'Aquila's reconstruction challenges: has Italy learned from its previous earthquake disasters? *Disasters* 37(1):119–143
- Parisi F, Augenti N (2013) Earthquake damages to cultural heritage constructions and simplified assessment of artworks. *Eng Fail Anal* 34:735–760
- Peeters B, De Roeck G (2001) Stochastic system identification for operational modal analysis: a review. *J Dyn Sys Meas Control* 123(4):659–667
- Russo S (2013) On the monitoring of historic Anime Sante church damaged by earthquake in L'Aquila. *Struct Control Health Monit* 20(9):1226–1239
- Tertulliani A, Leschiutta I, Bordoni P, Milana G (2012) Damage distribution in L'Aquila city (central Italy) during the 6 April 2009 earthquake. *Bull Seismol Soc Am* 102(4):1543–1553
- Volpe M, Piersanti A, Melini D (2012) Complex 3-D finite element modelling of the 2009 April 6 L'Aquila earthquake by inverse analysis of static deformation. *Geophys J Int* 188(3):1339–1358

Seismic Collapse Assessment

Christoph Adam¹ and Luis F. Ibarra²

¹Department of Engineering Science, Unit of Applied Mechanics, University of Innsbruck, Innsbruck, Austria

²Department of Civil and Environmental Engineering, University of Utah, Salt Lake City, USA

Synonyms

Collapse criterion; Collapse variability; Component deterioration; Nonlinear dynamic analysis; P-delta effect; Structural modeling strategies

Introduction

In earthquake engineering, structural collapse is defined as the local or global failure of a system that occurs due to the loss of vertical load-carrying capacity in the presence of seismic events. The two primary modes of global collapse are sidesway and vertical collapse. Sidesway collapse is the global failure of the system caused by a reduction of the lateral load-bearing capacity due to large horizontal displacements, whereas vertical collapse is caused by a direct loss of the gravity load-bearing capacity in one or several structural components (Krawinkler et al. 2009). Vertical collapse is a type of progressive collapse, which is the total or disproportionate failure of the system triggered by an initial local failure that spreads out from element to element.

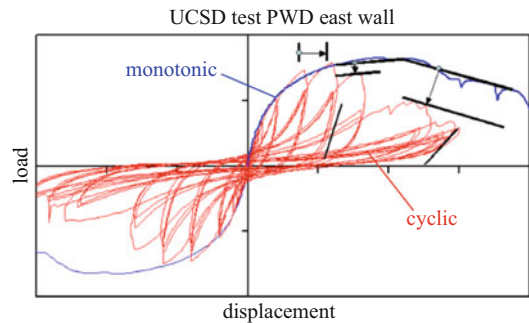
Commonly, in ductile frame structures, sidesway collapse is the predominant mode of

collapse during catastrophic earthquake events. These systems are affected by component material deterioration and/or P-delta effects. Material deterioration reduces the component's strength under large inelastic displacements. Moreover, the strength and stiffness of the structural components deteriorates in each load cycle, reducing the load-bearing capacity. The importance of material deterioration decreases for long-period structures, where the P-delta effects (also referred to as second-order effects) may lead to a negative post-yield stiffness that in combination with large inelastic displacements accelerates the onset of collapse.

In most non-ductile frames, however, columns lose the capacity to sustain gravity loads due to shear and subsequent axial-load failures, and as a result, the collapse mechanism is controlled by the loss in vertical load-bearing capacity. As outlined in Baradaran Shoraka (2013), this type of collapse of older concrete buildings may be precipitated by axial-load failure of columns, punching shear failure of slab-column connections, failure of slab-diaphragm connections, or axial-load failure beam/column joints.

Prediction of earthquake-induced collapse requires the identification of all possible modes of collapse, appropriate structural modeling of building components under cyclic loading, and suitable representation of earthquake excitation including the consideration of epistemic and aleatory uncertainties. Krawinkler et al. (2009), Haselton et al. (2009), and Baradaran Shoraka (2013) provide insights into the phenomenon of dynamic collapse of earthquake-excited structures. For basic concepts in prediction of collapse and its implementation to performance-based earthquake engineering, the reader is referred to Zareian et al. (2010). Villaverde (2007) presents a comprehensive literature that provides a path through which seismic collapse assessment has evolved in the last decades.

The required accuracy and the available resources determine the selection of a seismic collapse assessment procedure. In a first step, the governing seismic modes of collapse such as vertical collapse or sidesway collapse are identified. Then a structural model is created that

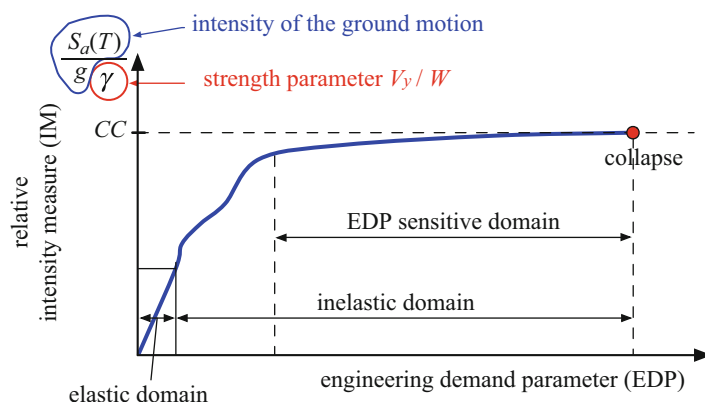


Seismic Collapse Assessment, Fig. 1 Force-displacement relations from a monotonic and a cyclic test (Modified from Gatto and Uang 2002; With permission from Prof. Uang, UC San Diego)

captures these failure modes, and modeling uncertainties are quantified. Evaluation of seismic collapse should be based on structural analyses that incorporate strength and stiffness material deterioration, as well as geometric nonlinearities (P-delta effects). Collapse of single-degree-of-freedom (SDOF) systems was first studied by including only P-delta effects in seismic response, which may cause a negative inelastic tangent stiffness that eventually leads to system collapse (Jennings and Husid 1968; Bernal 1987; MacRae 1994). The development of hysteretic models that include strength and stiffness deterioration (Kunnath et al. 1990; Sivaselvan and Reinhorn 2000; Song and Pincheira 2000; Ibarra and Krawinkler 2005) improved the assessment of collapse capacity. As observed in Fig. 1, collapse assessment requires models that can predict strength reduction under large displacements, as well as strength and stiffness deterioration due to cyclic loading.

This article focuses on sidesway collapse of frames, which has been traditionally estimated by means of non-degrading systems, in which judgmental limits are assigned to response quantities such as roof drift, story drift, ductility, and hysteretic energy. These parameters are usually referred to as engineering demand parameters (EDPs) (Moehle and Deierlein 2004). The development of degrading models permits tracking the collapse limit state, but EDPs become very sensitive when the system is close to collapse, and

Seismic Collapse Assessment, Fig. 2 IDA curve for a single ground motion up to collapse: elastic, inelastic, and EDP-sensitive domain; collapse (With permission from Adam 2014)



small perturbations in input parameters lead to large variations in the response. For this reason, global collapse methodologies are commonly based on an intensity measure (IM) instead of an EDP. A relative IM is the ratio of ground motion intensity (demand) to a structure strength parameter (capacity). For a given structure and ground motion, collapse evaluation consists of a series of response history analysis starting with a relative intensity that produces an elastic response of the system. Then the relative intensity is increased to cover the full range of interest or until collapse takes place. This series of response history analysis is called incremental dynamic analysis (IDA) (Vamvatsikos and Cornell 2002). As an example, Fig. 2 shows an IDA curve that results from increasing the IM up to collapse. Collapse is imminent when a very small increment in relative intensity causes a very large increment in the EDP of the system, a condition that indicates dynamic instability. In Fig. 2 dynamic instability occurs when the slope of the IDA curve approaches zero.

The global seismic collapse capacity of a system is defined as the largest earthquake intensity at which the structure still maintains dynamic stability (Krawinkler et al. 2009), denoted as CC in Fig. 2. This definition is associated with a specific structure under a given ground motion. Dynamic analysis requires the specification of system properties and input ground motions, but variations in these quantities may produce large dispersion in the response. Therefore, collapse

needs to be evaluated in a probabilistic framework that includes uncertainties in the frequency content of the ground motions (e.g., record-to-record (RTR) variability) and in the input parameters of the system (e.g., epistemic uncertainty). Then, nonlinear response history analyses usually include a set of ground motion records that represents the RTR variability at the site of the structure. The effect of uncertainty in nonlinear deterioration parameters on the variance of collapse capacity can be quantified by means of approximate methods (e.g., FOSM method) or carrying out Monte Carlo simulations (Ibarra and Krawinkler 2011; Ugurhan et al. 2014).

The variance in collapse capacity can be used to generate collapse fragility functions that provide the probability of collapse given that an event with a certain IM occurred. Shome and Cornell (1999) summarize the arguments used to consider the collapse fragility as a log-normally distributed function. For simplified procedures, only the median, 16th, and 84th percentiles of the individual collapse capacities are required to find an approximation of the collapse fragility. A final product of collapse assessment is the mean annual frequency of collapse, which is obtained by combining fragility curves with the hazard information at the site.

In this contribution some of the issues involved in seismic sidesway collapse assessment of moment-resisting frames are addressed in more detail, including nonlinear dynamic analysis, structural modeling, strength and stiffness

material deterioration, and P-delta effect. The probabilistic framework used to evaluate collapse and the effect of epistemic uncertainties and aleatory variability on the collapse capacity variance are discussed, although ground motion selection is not part of this article. The last sections deal with the development of simplified methods to evaluate collapse capacity.

Nonlinear Dynamic Analysis

In the last two decades, significant progress has been achieved in the nonlinear dynamic analysis of structures. It is now established as the most reliable tool to predict realistically the seismic collapse capacity of a building, and several provisions such as FEMA P-695 (2009), ATC-58 (2012), and ATC-78 (2013) propose collapse assessment methodologies based on nonlinear dynamic analysis.

Incremental Dynamic Analysis

If sidesway collapse is the governing failure mechanism, the IDA procedure can be used to determine the largest earthquake intensity at which the structure still maintains stability. In this approach nonlinear dynamic response history analyses are performed repeatedly, increasing in each subsequent run the ground motion intensity. The IDA curve of the structure for the considered earthquake record is obtained by plotting an appropriate record intensity measure against its corresponding EDP (Vamvatsikos and Cornell 2002). The analysis can be stopped when the EDP satisfies a certain failure criterion that may correspond to structural sidesway collapse. In this case, the corresponding intensity of the ground motion is referred to as collapse capacity of the building for this specific ground motion record.

IDA is commonly performed for a representative set of ground motions to account for RTR variability. Examples of such sets are the LMSR ground motions (Medina and Krawinkler 2003) and FEMA P-695 sets for far-field and near-field ground motions (FEMA P-695 2009). In this sense, the ground motions are used as standardized excitation protocols, and the effort of

selecting ground motions is avoided (Baker 2013). These records need to be scaled to the same IM, and since there is not a unique definition, several IMs have been proposed: (a) elastic ground motion-based scalar IMs such as peak ground acceleration (PGA), peak ground velocity (PGV), and peak ground displacement (PGD); (b) elastic and inelastic spectral-based IMs such as spectral acceleration and spectral displacement at the fundamental period of the structure, as well as spectral values related to higher mode effects or period elongation; and (c) vector valued IMs (e.g., Baker and Cornell 2005). Currently, the most widely used IM is the 5 % damped spectral pseudo-acceleration $S_a(T_1, \zeta = 0.05)$ at the fundamental period T_1 of the structure. For example, if $S_a(T_1, \zeta = 0.05)$ is normalized by the product of the acceleration of gravity g and the base shear coefficient γ , the relative seismic collapse capacity of the building subjected to the ground motion “ i ” reads as

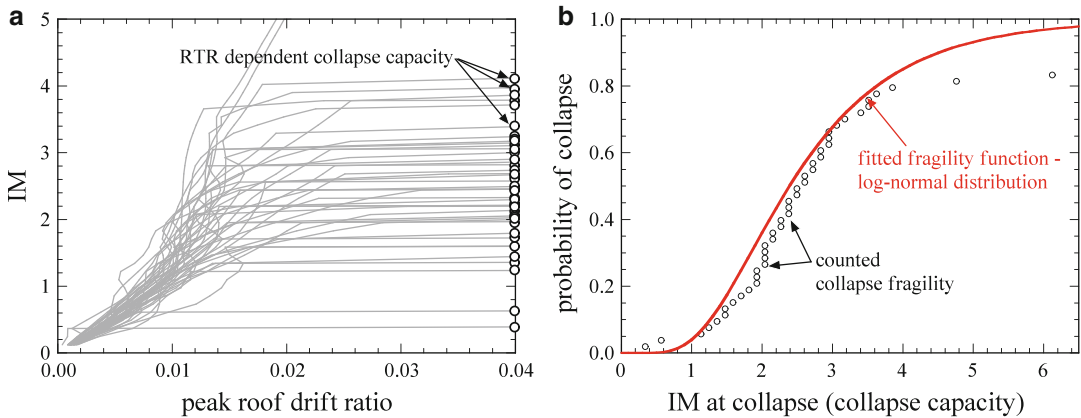
$$CC_i = \frac{S_{a,i}}{g\gamma} \Big|_{\text{collapse}} \quad (1)$$

The base shear coefficient γ is defined as the base shear V_y at onset of yielding divided by the total weight W .

The collapse fragility function is obtained by counting the fraction of collapse capacities CC_i ($i = 1, 2, \dots, n$) that correspond to the n ground motions of the considered set. When all ground motions lead to collapse, the collapse capacity distribution is usually log-normal distributed (Ibarra and Krawinkler 2005). As an example, Fig. 3 shows the IDA curves of a structure subjected to the 44 ground motions of the FEMA P-695 far-field ground motion suite and the corresponding collapse fragility function.

Multiple Stripe Analysis

The IDA approach cannot consider certain site-specific ground motion conditions. As outlined by Baker (2013), some of the ground motion properties relevant for collapse assessment are the spectral shape of the elastic response spectrum, the duration, and the effects of near-fault ground motions and velocity pulses. The



Seismic Collapse Assessment, Fig. 3 (a) IDA curves of a structure for all records of a ground motion set. (b) Counted collapse fragility and corresponding fitted

fragility function according to a log-normal distribution (With permission from Adam 2014)

expected spectral shape of a ground motion does not only vary with location but it also depends strongly on the ground motion amplitude (Baker and Cornell 2006). For instance, at different seismic intensity levels, the spectral shape of the expected earthquake is dissimilar. These differences in the spectral shapes can be ascribed to the soil condition, magnitude, distance, and parameter ε , which quantifies the difference between the selected ground motion's spectral acceleration at a specific period and the median of the ground motion spectral acceleration obtained from attenuation relationships (Baker 2013). Because the same set of records are often used for different locations, however, IDA yields the same risk of sidesway collapse for two equal buildings located in different locations with similar seismic hazard, although the anticipated ground motion spectral shapes are different.

To address some of the above issues, Jalayer and Cornell (2009) developed a multiple stripe analysis (MSA) that accounts for specific site conditions and for ground motion properties that change with increasing magnitude. In this approach, the ground motions are reselected at each IM level according to the changing spectral shape. For each record the peak response of the EDP derived from nonlinear response history analysis and the number of events that leads to sidesway collapse is plotted as a function of the corresponding IM, thus leading to "stripes."

The counted fraction of observed individual collapses at each IM level leads then to the collapse fragility curve.

Collapse Criteria

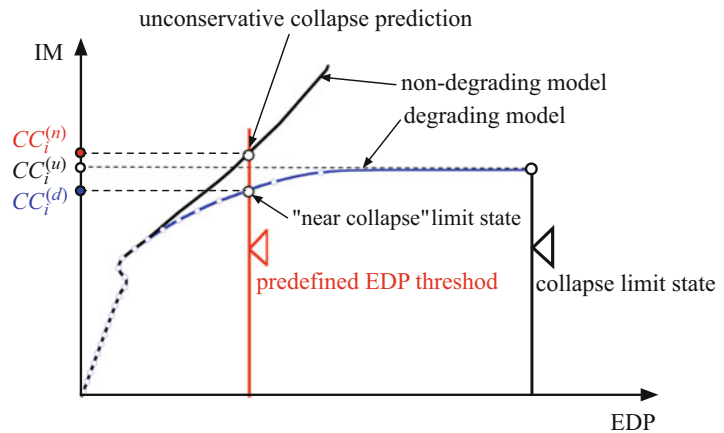
The numerical sidesway collapse prediction depends on the model capabilities. Historically, the first nonlinear numerical models did not include deteriorating characteristics, and damage indexes were developed to predict collapse based on maximum displacements and/or dissipated energy. The most widespread cumulative damage model was the Park–Ang model (Park and Ang 1985), which was developed specifically for reinforced concrete components. The damage measure (DM) consisted of a linear combination of displacement and energy demands:

$$DM = \frac{\delta_{\max}}{\delta_{\text{ult}}} + \frac{\beta}{F_y \delta_{\text{ult}}} \int dHE \quad (2)$$

where $\delta_{\max}$ is the maximum displacement of the system, δ_{ult} denotes the monotonic displacement capacity of the system, β represents the structural performance parameter, and HE is the hysteretic energy. In concept, the parameters δ_{ult} and β of this cumulative damage model are equivalent to the deteriorating parameters of recent hysteretic models that include cumulative deterioration. However, damage index models were based on non-deteriorating models, leading to unrealistic

Seismic Collapse

Assessment, Fig. 4 Two IDA curves for the same building and ground motion record based on two different structural models. Definition of the collapse capacity: $CC_i^{(u)}$ collapse capacity associated with infinite deformation. $CC_i^{(d)}$, $CC_i^{(n)}$: collapse capacity predictions based on a predefined EDP limit (With permission from Adam 2014)



displacement–force relationships for large inelastic excursions.

In recent practical applications, when non-degrading models are used to predict collapse, this limit state is associated with the deformation of a single structural component that exceeds a predefined EDP threshold, often represented by a plastic rotation demand (Haselton et al. 2009). Eurocode 8 (CEN 2005) provisions, for instance, are based on capacity thresholds for the deformation of RC members at yielding and failure developed by Fardis and collaborators (e.g., Panagiotakos and Fardis 2001; Fardis and Biskinis 2003). As a further example, ASCE/SEI 41-13 (2014) procedure compares the component demands of an existing building with component acceptance criteria that define collapse prevention. In such an approach, the component that fails first governs the state of the entire structure. That is to say, component capacity thresholds do not capture the global behavior and the capability of the structure to redistribute forces after failure of one component.

A structure may exhibit several modes of seismic-induced collapse such as sidesway collapse or vertical collapse, depending on the history of spread of inelastic deformations, load distribution, etc. If the corresponding degrading structural model captures only a specific type of collapse such as sidesway collapse, the other collapse mechanisms that might occur are referred to non-simulated modes of collapse. That is, they only can be predicted by an artificial

criterion such as the exceedance of a predefined EDP threshold. In particular non-ductile collapse modes such as column shear failure and subsequent vertical collapse are frequently considered as non-simulated collapse modes (e.g., Liel and Deierlein 2013).

The IDA curves of Fig. 4 illustrate sidesway collapse prediction for degrading and non-degrading structural models. The intensity $CC_i^{(u)}$ is the ultimate collapse capacity associated with infinite EDP deformations for the degrading model, and it corresponds to the “exact” collapse capacity. However, if a predefined deformation threshold is used to define collapse, the use of the degrading model always results in a conservative near-collapse prediction. That is, the intensity at this EDP threshold, $CC_i^{(d)}$, is smaller than $CC_i^{(u)}$. On the other hand, the use of a non-degrading model may underestimate or overestimate the actual collapse capacity, depending on the preestablished EDP threshold. In Fig. 4 the corresponding intensity denoted as $CC_i^{(n)}$ results in a nonconservative prediction because the predefined EDP threshold is rather large.

Baradaran Shoraka (2013) developed a robust collapse assessment framework to predict vertical collapse, such as column shear failure and subsequent axial failure in non-ductile reinforced concrete frames. In this approach, progression of damage is tracked throughout the numerical analysis by comparing stepwise floor level gravity demands with corresponding capacities. Explicit modeling of progressive collapse involves

element (commonly column) removal until the structure attains a state that cannot resist the vertical load demand. Baradaran Shoraka et al. (2014) also compared the results of explicit gravity load collapse modeling and those obtained from nonlinear acceptance collapse prevention criteria specified in ASCE/SEI 41-13 (2014) for retrofit of non-ductile frames.

Structural Modeling Strategies for Assessing Sidesway Collapse

Structural Modeling on the System Level

The structural model may be two-dimensional or three-dimensional, but when the building exhibits a coupled bending-torsional response, a three-dimensional system better predicts the building seismic behavior. Such a system requires sophisticated modeling of the nonlinear constitutive behavior at the component level that is not well understood. Thus, three-dimensional models are avoided for collapse analysis as long as it is feasible, and a more generic two-dimensional structural model is utilized.

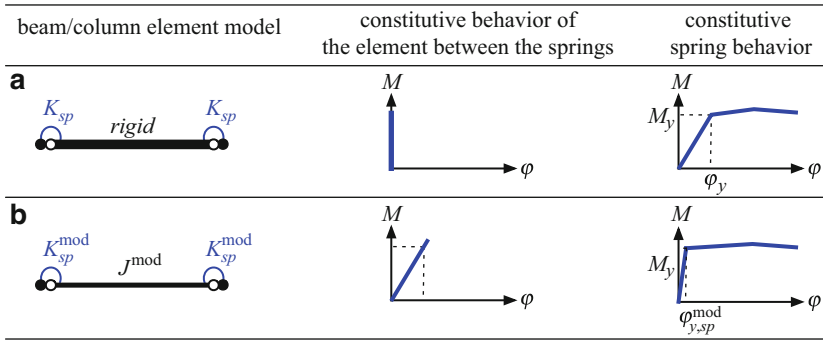
As outlined by Krawinkler et al. (2009), there are two practical modeling strategies for frame structure evaluation: fiber models and lumped plasticity models. In a fiber model, an axial stress–strain relation is separately assigned to each fiber of the cross section. For example, different constitutive laws can be used in reinforced concrete sections to represent confined and unconfined concrete and the longitudinal steel bars. In this manner, the spread of inelastic deformations within the structural component is captured. In many situations this approach can accurately predict earlier limit states such as yielding and crack initiation, among others. However, several problems have been observed in fiber models. Since these models usually are based on axial stress–strain relations, shear deformations cannot be modeled appropriately. According to Krawinkler et al. (2009), further phenomena that are difficult and often impossible to capture are, for example, bond slip, rebar fraction and rebar buckling, fracture of concrete, and local and lateral torsional buckling in steel. Since

these phenomena may govern the deterioration process, strain “artificial” limits need to be specified in fiber models (Krawinkler et al. 2009).

In a lumped plasticity model, locations within the structure that might undergo inelastic deformations in a severe seismic event are identified up front. Inelastic nonlinear zero-length springs are assigned to these locations, while the rest of the model components remain elastic. These simple phenomenological component models are calibrated with experimental outcomes to represent appropriate cyclic response behavior.

For instance, in a moment-resisting frame, rotational springs are placed at the end of beams and columns. Frames are designed according to the strong column-weak beam philosophy, in which beams and bottom end of base columns are intended to exhibit inelastic deformations at their ends when subjected to strong ground motions. However, many columns may undergo inelastic deformations at the story level when approaching the collapse limit state. Thus, nonlinear spring elements need to be placed at these locations to accurately predict the seismic collapse capacity.

A concentrated plasticity model includes beam/column elements of different degree of sophistication. In the simplest case, all elastic, inelastic, and deteriorating properties of the element are assigned to rotational springs at both ends of the element, whereas the element itself is considered to be rigid; see Fig. 5a. This model may capture the development of plastic hinges. As a disadvantage the spring stiffness must be defined a priori, which is a constant regardless of the potential changes in the moment gradient of the element. To avoid this drawback, Ibarra and Krawinkler (2005) suggest to keep the beam/column element elastic and to add rotational springs at the ends, whose rotational stiffness is several times stiffer than the rotational stiffness of the elastic beam/column element. In this model, an elastic beam/column element is connected to two springs at the ends, and thus, the properties of the parameters can be calibrated according to the component “real” behavior (Fig. 5b). Jäger and Adam (2012) showed that the latter element can describe column and beam behavior sufficiently



Seismic Collapse Assessment, Fig. 5 Different beam/column elements and their constitutive behavior to monotonic loading. (a) Rigid element with two nonlinear

rotational springs. (b) Modified elastic element with two nonlinear rotational springs (Modified from Jäger and Adam 2012)

accurate for collapse assessment. The rigid element with two nonlinear springs at its ends should be used for beams only because when it is used in columns, the elastic axial-bending loading interaction cannot be taken into account. A shortcoming of lumped plasticity models is that the nonlinear column spring strength remains constant during the response history analysis, independently of the moment - axial force interaction.

Figure 6a shows a concentrated plasticity model for a multistory three-bay steel moment-resisting frame, whose plan configuration is shown in Fig. 6b. At both ends of each beam and column, a nonlinear rotational spring is arranged with an offset from the node to account for the member's cross section. In contrast, generic frame models used for research purposes are often “centerline models,” i.e., the lumped plasticity elements are located at the nodes without any offset from the nodal point.

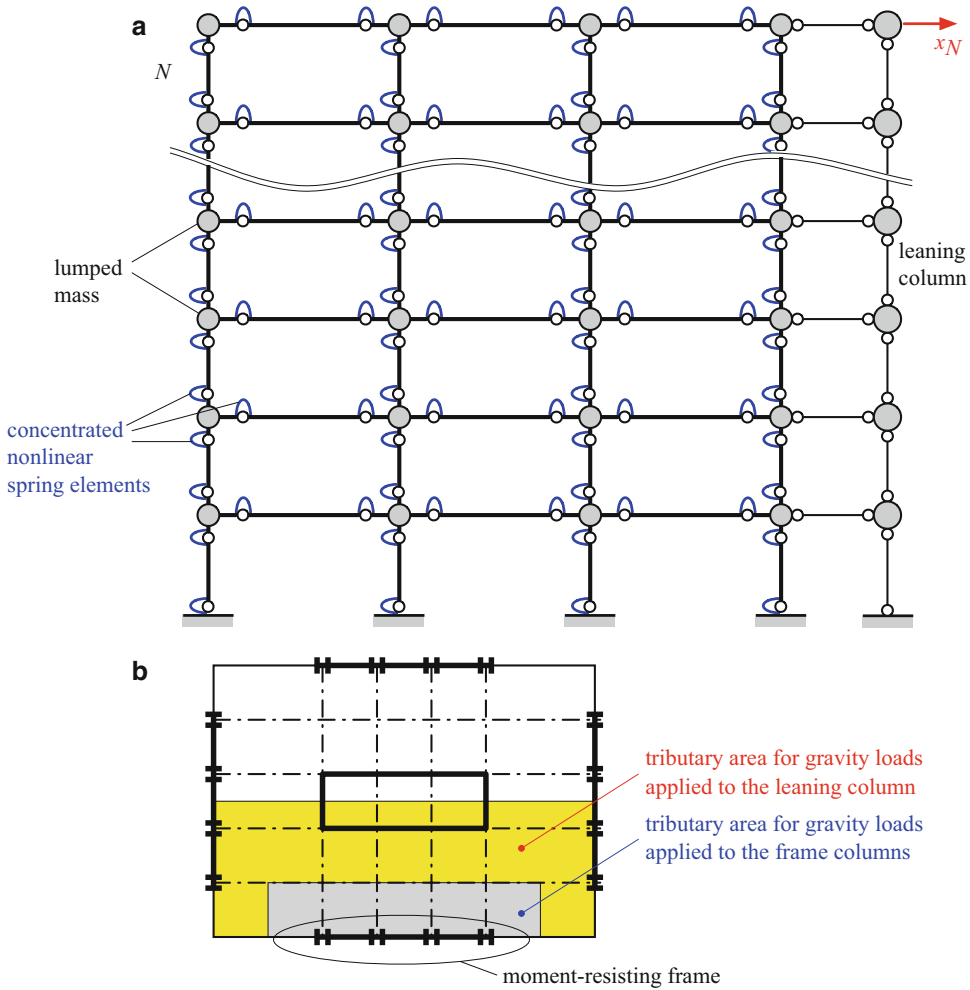
The structural model of Fig. 6a includes a leaning column with zero flexural stiffness to account for the P-delta effect resulting from gravity loads that are not directly applied to the frame (Haselton et al. 2009). This leaning column is loaded with a vertical load at each story level representing half of the total system gravity load that is not directly tributary to the columns of the frame (NIST GCR 10-917-8 2010), as shown in Fig. 6b.

Another issue that needs to be properly addressed is modeling of viscous damping for

nonlinear dynamic analysis. Several studies have revealed that inappropriate consideration of viscous damping leads to spurious damping forces that distort the nonlinear dynamic response prediction. Particularly affected are lumped plasticity models with elements that exhibit large initial stiffness (Charney 2008). According to the ATC-72-1 report (ATC-72-1 2010), modal damping is the preferred choice for modeling of damping in inelastic structures, but in many software packages, this is not possible. In such cases a combination of mass and stiffness proportional damping (e.g., Rayleigh damping) should be used. Modeling guidelines for damping in inelastic structures are provided in ATC-72-1 (2010).

Structural Modeling of Cyclic Component Behavior

Sidesway collapse assessment based on lumped plasticity structural models requires hysteretic models that include all the important modes of deterioration observed in experimental tests. As illustrated in Fig. 7, monotonic tests on structural components show that strength of the backbone curve is “capped” and is followed by a negative tangent stiffness. In addition, the cyclic response indicates that strength and stiffness deteriorate with the number and amplitude of cycles. As a result, the strength predicted by the monotonic backbone curve is usually not reached during cyclic loading. Several hysteretic models have been developed to represent this behavior (e.g., Song and Pincheira 2000). However, few models



Seismic Collapse Assessment, Fig. 6 Example of structural modeling of a frame building. (a) Overall lumped plasticity structural model of the moment-

resisting steel frame. (b) Rectangular plane configuration of the frame building (Modified from NIST GCR 10-917-8 [2010](#))

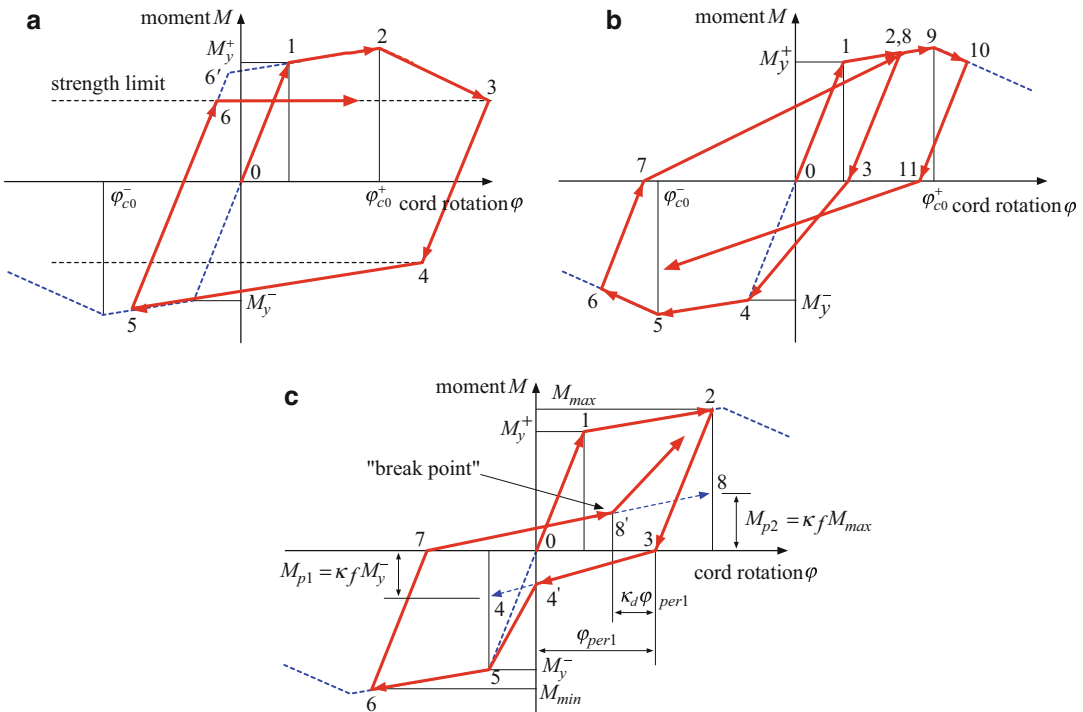
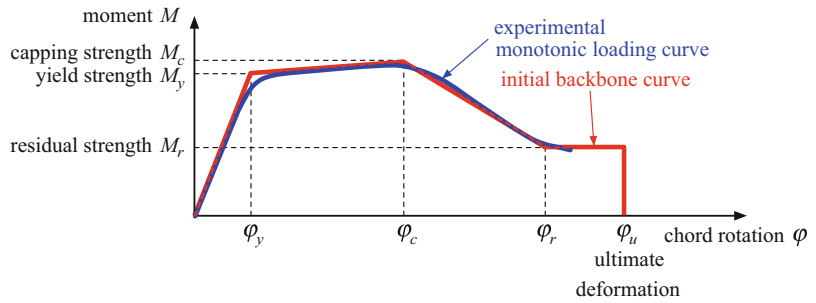
integrate strength deterioration in the backbone curve, as well as cyclic deterioration of strength and stiffness. This section describes exemplarily the deteriorating models developed by Ibarra et al. (2005) for bilinear, peak-oriented, and pinched hysteretic systems.

Figure 8 presents bilinear, peak-oriented, and pinching deteriorating models based on a backbone curve that includes a post-capping negative slope, and sometimes a residual strength, as shown in Fig. 7. The deteriorating models preserve most of the rules of the original non-deteriorating models. For instance, the

bilinear model (Fig. 8a) is based on the standard hysteretic bilinear rules with kinematic strain hardening, but a “strength limit” is introduced when the backbone curve includes a branch with negative slope. The limit for strength is the smallest strength reported on the post-capping branch in previous excursions. If this condition were not established, the strength in the loading path could increase in later stages of deterioration. Note that the original peak-oriented and pinching models exhibit stiffness deterioration in the reloading path. However, in this article the term “deterioration” is used for hysteretic

Seismic Collapse Assessment,

Fig. 7 Initial backbone curve for the modified Ibarra–Krawinkler model (Modified from Krawinkler et al. 2009)



Seismic Collapse Assessment, Fig. 8 Basic deteriorating hysteretic models: (a) bilinear, (b) peak oriented, and (c) pinching (Modified from Ibarra and Krawinkler 2005)

models that possess a post-capping stiffness branch in the backbone curve. These models may also be subjected to cyclic deterioration.

The peak-oriented model (Fig. 8b) keeps the basic hysteretic rules proposed by Clough and Johnston (1966), but the backbone curve is modified to include strength capping and residual strength. The negative post-capping stiffness does not modify any basic rules of the model. The pinching model is similar to the peak-

oriented one, but the reloading consists of two parts. First the reloading path is directed toward a “break point,” which is a function of the maximum permanent deformation and the maximum load experienced in the direction of loading. The break point is defined by the parameter κ_f that modifies the maximum “pinched” strength (points 4 and 8 of Fig. 8c) and the parameter κ_d that defines the displacement of the break point (points 4' and 8').

Component Deterioration Based on Hysteretic Energy Dissipation

Consideration of cyclic component deterioration is important for collapse prediction, especially for short-period highly ductile systems. The Ibarra–Krawinkler hysteretic model includes four cyclic deterioration modes based on energy dissipation that are activated once the yielding point is surpassed. As observed in Fig. 9, basic strength and post-capping strength deterioration effects translate the strain-hardening and post-capping branch toward the origin, unloading stiffness deterioration decreases the unloading stiffness, and reloading (accelerated) stiffness deterioration increases the target maximum displacement.

The amount of deterioration depends on the parameter $\beta_{m,i}$, which is based on the hysteretic energy dissipated when the component is subjected to cyclic loading (Ibarra and Krawinkler 2005),

$$\beta_{m,i} = \left(\frac{E_i}{E_{t,m} - \sum_{j=1}^i E_j} \right)^{c_m} \quad (3)$$

where E_i is the hysteretic energy dissipated in excursion i , $\sum E_j$ represents the hysteretic energy dissipated in all previous positive and

negative excursions, and $E_{t,m}$ denotes the hysteretic energy dissipation capacity for the cyclic deterioration mode “ m .” The original hysteretic rules defined $E_{t,m}$ as a function of twice the elastic strain energy, $E_{t,m} = \gamma_m M_y \varphi_y$ (Ibarra and Krawinkler 2005). The modified Ibarra–Krawinkler model defines $E_{t,m}$ as a function of the plastic chord rotation φ_p , $E_{t,m} = \lambda_m M_y \varphi_p$, or $E_{t,m} = \Lambda_m \varphi_p$ (Lignos and Krawinkler 2012). The parameters γ_m , λ_m , and Λ_m are calibrated from experimental tests, and reasonable results are obtained if all cyclic deterioration parameters have the same value. Exponent c_m is the rate of deterioration parameter. $c_m = 1$ implies a constant rate of deterioration.

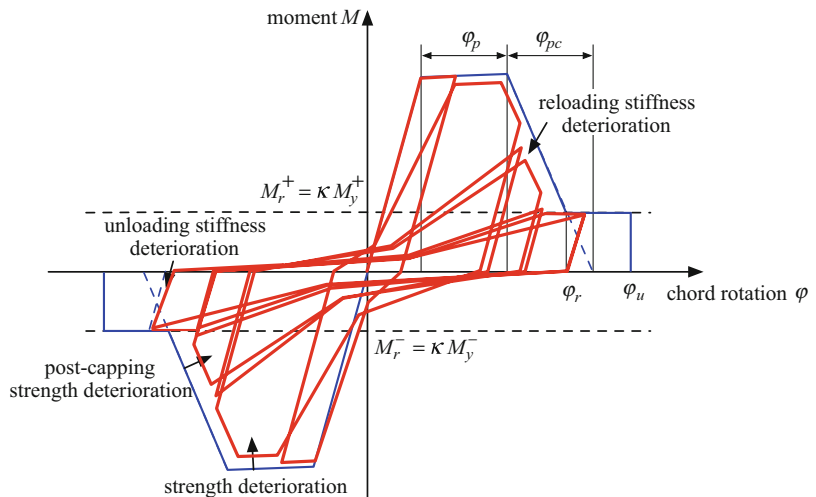
The parameter $\beta_{m,i}$ modifies the strength and stiffness properties of the previous cycle. For basic strength deterioration, for instance, the strain-hardening branch moves toward the origin by an amount equivalent to reducing the yield strength as follows (for this mode index $m = s$):

$$M_{y,i}^+ = (1 - \beta_{s,i}) M_{y,i-1}^+ \quad (4)$$

where $M_{y,i}^+$ is the deteriorated yield strength after excursion i and $M_{y,i-1}^+$ is the deteriorated yield strength before excursion i . Except for the unloading stiffness mode, a positive and a negative value is defined for each deterioration parameter because the algorithm deteriorates the strength independently in both directions.

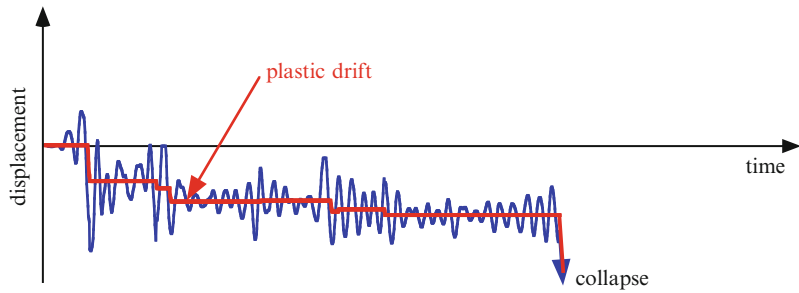
Seismic Collapse

Assessment, Fig. 9 Four modes of component deterioration in a pinching component (Modified from Lignos and Krawinkler 2013; With permission from American Society of Civil Engineers)



Seismic Collapse Assessment,

Fig. 10 Inelastic structural response history with ratcheting effect due to P-delta (With permission from Adam 2014)



That is to say, $M_{y,i}^-$ is updated after every positive inelastic excursion, and $M_{y,i}^+$ is updated after every negative inelastic excursion. For example, $\beta_{s,i}$ is calculated each time the inelastic path crosses the horizontal axis. These parameters are calibrated from experiments. For instance, Lignos and Krawinkler (2012, 2013) conducted deterioration modeling of steel components based on an experimental database.

Global Seismic Sidesway Collapse and the P-Delta Effect

In a vibrating structure, gravity loads acting through large horizontal deflections induce additional bending moments, which in turn increase the lateral deflections. In structural analysis, this destabilizing effect of gravity loads (i.e., global P-delta effect) is usually interpreted as a reduction of the lateral structural stiffness by the so-called geometric stiffness. Generally, for a real building within its elastic range of deformation, this stiffness decrease is of minor importance because its magnitude is small compared to the first-order elastic stiffness. During severe seismic excitations, however, inelastic deformations combined with gravity may cause a structure to approach a state of dynamic instability if the post-yield tangent stiffness becomes negative. In such a condition, the displacement response tends to amplify in a single direction due to the ratcheting effect, as shown in Fig. 10. Very flexible structures are especially affected by seismic-induced sidesway collapse in the P-delta mode, and in many of these cases, cyclic component material deterioration can be disregarded.

However, this consideration not only depends on the structure characteristics but also on the ground motion applied to the structure. For example, if a strong-motion earthquake causes structural collapse after a few load cycles, component degradation is of minor significance compared to a situation where a long-duration earthquake exhibits many cycles before the structure fails.

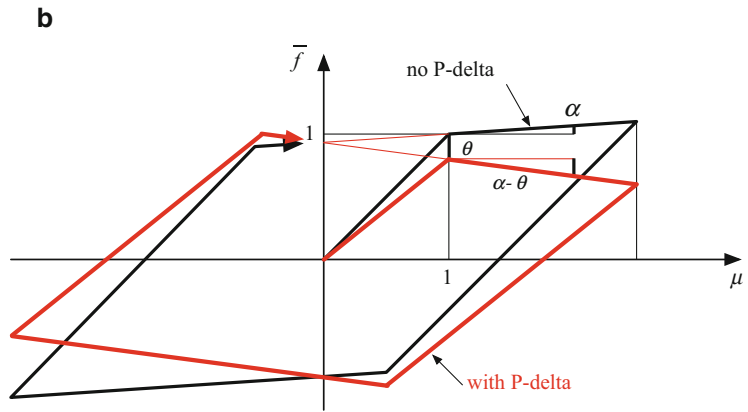
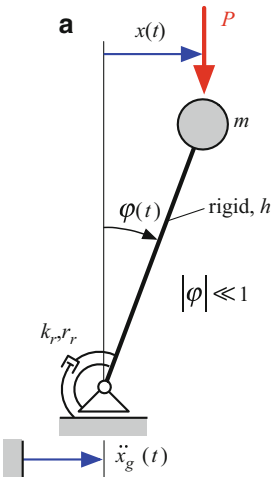
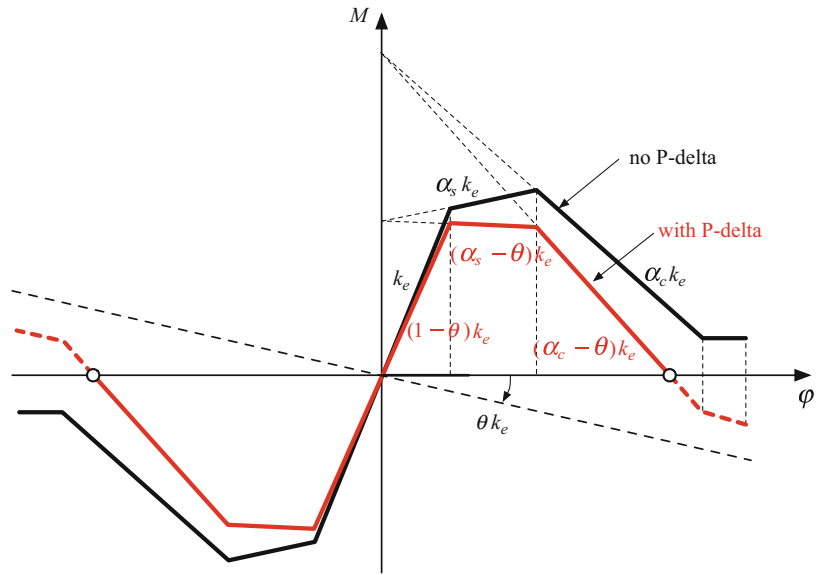
In an inelastic SDOF system, the gravity load generates a uniform shear deformation of its hysteretic force–displacement relationship, as shown exemplarily in Fig. 11, for the backbone curve of an inelastic SDOF system. Characteristic displacements (such as the yield displacement) of this relationship remain unchanged, whereas the characteristic forces (such as the strength) are reduced. As a result, the slope of the elastic and post-elastic branch rotates. The magnitude of this reduction can be expressed by means of the stability coefficient θ (MacRae 1994). The parameter θ is a function of the gravity load, geometry, and stiffness. For instance, for the bilinear SDOF system shown in Fig. 12a, the stability coefficient reads as

$$\theta = \frac{Ph}{k_r} \quad (5)$$

where P is the gravity load, h the length of the rod, and k_r the initial stiffness of the elastoplastic rotational spring at the base. In the example of Fig. 12, the post-yield stiffness is negative, because the stability coefficient θ is larger than the hardening ratio α . In a nonmaterial degrading SDOF system, a negative post-tangential stiffness ratio, i.e., $\theta - \alpha > 0$, is a necessary condition for potential structural collapse under severe earthquake excitations.

Seismic Collapse Assessment,

Fig. 11 P-delta effect on the backbone curve of an inelastic SDOF system (Modified from Ibarra and Krawinkler 2005)

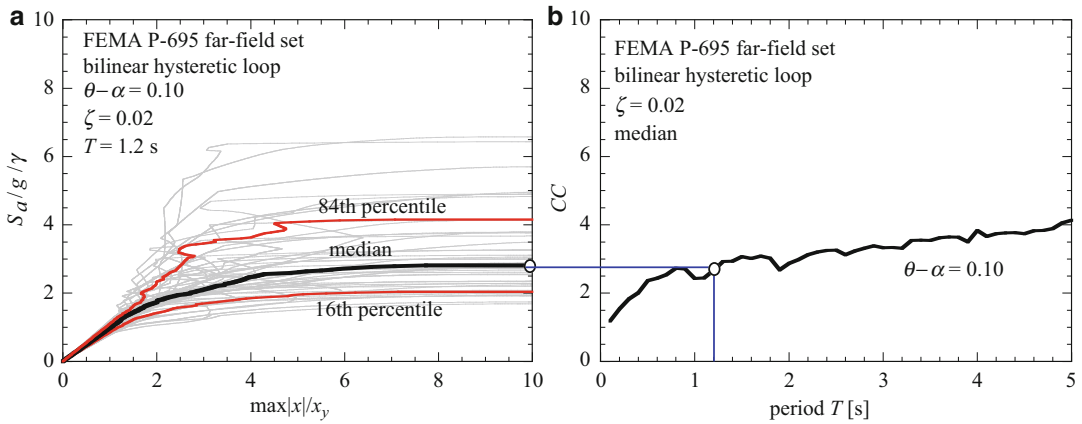


Seismic Collapse Assessment, Fig. 12 (a) Model of an SDOF system vulnerable to P-delta. (b) Normalized bilinear cyclic structural behavior of an SDOF system with and without P-delta (With permission from Adam and Jäger 2012a)

Adam and Jäger (2012a) showed that the collapse capacity of a P-delta vulnerable inelastic SDOF system with non-deteriorating hysteretic behavior and bilinear backbone curve is a function of the negative slope of the post-tangential stiffness ratio $\theta - \alpha$, the elastic structural period of vibration T , the viscous damping coefficient ζ , and the shape of the hysteretic loop such as bilinear, peak oriented, or pinching. Based on this observation, Adam and Jäger (2012a) and Jäger

and Adam (2013) conducted a rigorous parametric IDA study to reveal the impact of characteristic parameters (i.e., period T , negative post-yield stiffness ratio $\theta - \alpha$, damping ζ , and hysteretic loop) on the collapse of those systems, presenting the outcomes in the form of “collapse capacity spectra.”

In a collapse capacity spectrum, the P-delta vulnerable SDOF system collapse capacity is represented as a function of the initial period of

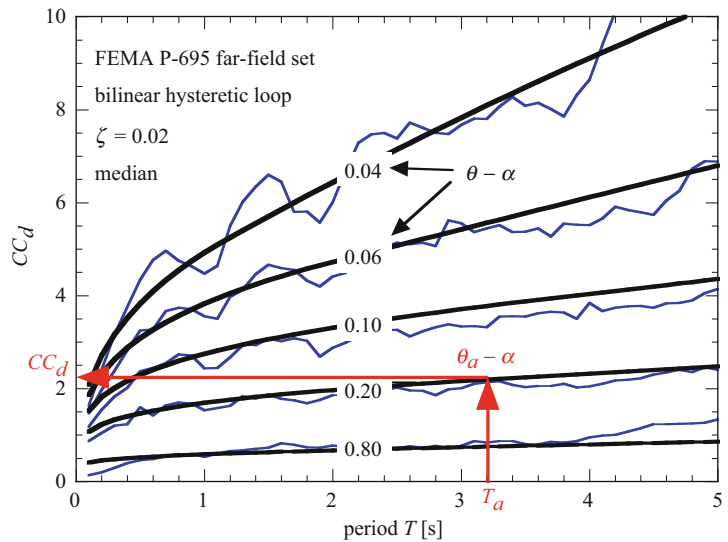


Seismic Collapse Assessment, Fig. 13 (a) IDA curves of an SDOF system with parameters defined in the figure for all records of the FEMA P-695 far-field ground motion

set and statistically evaluated results. (b) Corresponding median “instability collapse” capacity spectrum (With permission from Adam 2014)

Seismic Collapse Assessment,

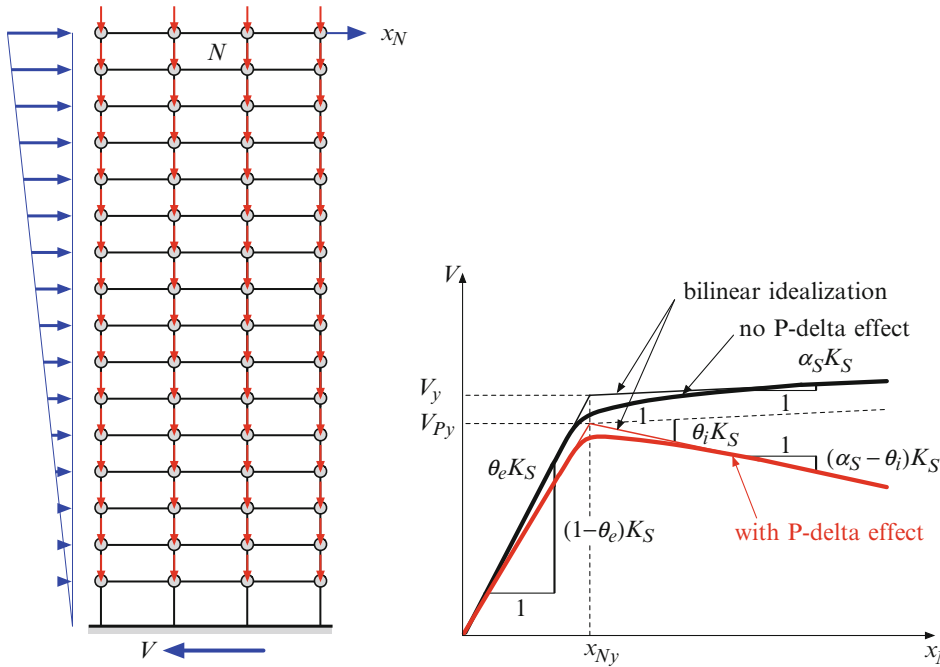
Fig. 14 Median collapse capacity spectra (blue lines) for a set of negative post-yield stiffness ratios and corresponding smooth collapse capacity spectra (With permission from Adam 2014)



vibration T for a fixed set of characteristic parameters. For example, Fig. 13a shows the individual IDA curves of an SDOF system for the 44 ground motions of FEMA P-695 far-field set (in Adam and Jäger (2012a) referred to as ATC63-FF set) up to “instability collapse,” as well as median, 16th, and 84th percentile curves. Collapse capacity spectra are obtained by repeating this procedure for different vibrational periods within a certain interval. As before, collapse capacity spectra can be computed for a specific suite of earthquake records, leading to median, 16th, and

84th percentile collapse capacity spectra (Fig. 13b). Based on a collapse capacity spectrum database, Adam and Jäger (2012a) and Jäger and Adam (2013) formulated analytical expressions of median, 16th, and 84th percentile collapse capacity spectra with a “smooth” shape via nonlinear regression analyses, as shown in Fig. 14.

In multistory frames, gravity loads may substantially impair the complete structure or only a subset of stories (Gupta and Krawinkler 2000). The local P-delta effect may induce collapse



Seismic Collapse Assessment, Fig. 15 Global pushover curves disregarding P-delta (*black graphs*) and considering P-delta (*red graphs*) (With permission from Adam 2014)

of a structural element, without necessarily affecting the stability of the complete structure. An indicator of the severity of the local P-delta effect is the story stability coefficient proposed in the Eurocode 8 (CEN 2005). However, a consistent relationship between the local P-delta effect and the global P-delta effect, which characterizes the overall impact of gravity loads on the structure, cannot be established due to dynamic interaction between adjacent stories in a multistory frame structure (Gupta and Krawinkler 2000).

Strong evidence for the vulnerability of a regular frame structure to P-delta-induced global seismic collapse is obtained from a first-mode pushover analysis (Adam and Jäger 2012b). For instance, the red curve of Fig. 15 illustrates the global pushover curve of a very flexible structure that exhibits a negative post-yield stiffness due to effect of gravity loads. When severe seismic excitation drives the structure into its inelastic branch, a state of dynamic instability may be approached, and the global collapse capacity is attained at a rapid rate. The pushover curve plotted in black

disregards gravity loads in the nonlinear static analysis.

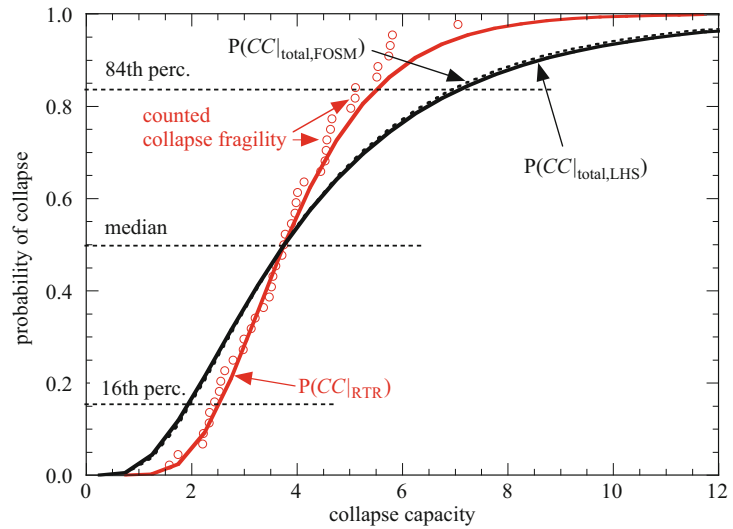
Global pushover curves of multi-degree-of-freedom (MDOF) structures do not exhibit a uniform stability coefficient in the entire range of elastic and inelastic deformations, unlike SDOF systems. As observed in Fig. 15, two stability coefficients can be identified in a bilinear approximation of pushover curves with and without P-delta effect, that is, a stability coefficient in the elastic range of deformation (θ_e) and a stability coefficient in the post-yield range of deformation (θ_i). According to Medina and Krawinkler (2003), θ_i can be much larger than θ_e : $\theta_i > (>) \theta_e$.

Variance of Collapse

Earlier studies on quantification of epistemic uncertainties on the seismic response concluded that the contribution of epistemic uncertainties to the variance of the system's seismic performance is in general much smaller than that from aleatory uncertainties, and thus it can be omitted

Seismic Collapse Assessment,

Fig. 16 Collapse fragility curves considering the RTR uncertainties only (red line, red markers) and combined RTR and epistemic uncertainties (black lines)



(e.g., Esteva and Ruiz 1989). In contrast, recent studies indicate that the epistemic uncertainty caused by degrading nonlinear system parameters is very relevant when assessing the total variance of the collapse limit state (Ibarra and Krawinkler 2005; Zareian et al. 2010). For example, Haselton et al. (2011) evaluated the total variance of collapse capacity of a four-story reinforced concrete special moment frame, including RTR variability and modeling uncertainties. He estimated the logarithmic standard deviation associated with modeling uncertainties as 0.45, a value comparable to the standard deviation of the logarithm of RTR variability, which varies in general between 0.25 and 0.45 (Liel et al. 2009; Adam and Jäger 2012a). As an example, Fig. 16 shows the collapse fragility of a P-delta vulnerable SDOF structure as shown in Fig. 12a, considering the RTR uncertainty only, and the total variability considering RTR and epistemic uncertainties.

The effect of RTR variability on the variance of collapse is directly obtained by performing IDAs for a set of ground motions on a deterministic system. The variance of collapse capacity of nonmaterial deteriorating systems is also affected by uncertainty in parameters, such as the yield moment and post-yield hardening ratio, but their effect on the variance of collapse capacity is small compared to that originated by RTR variability.

For deteriorating models, however, epistemic uncertainty can be relevant for collapse capacity evaluation because of the large uncertainty on degrading nonlinear parameters, such as plastic rotation and post-capping stiffness (Fig. 17) (Ibarra and Krawinkler 2011; Uguirhan et al. 2014).

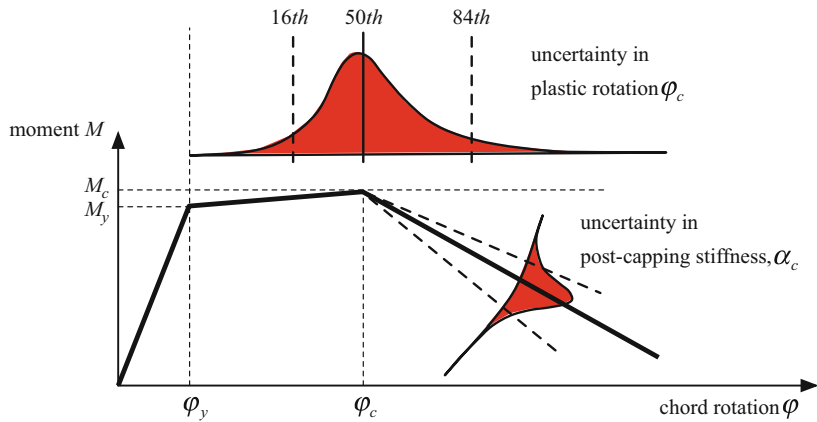
The first-order second-moment (FOSM) method (e.g., Melchers 1999) can be used to compute the additional variance of collapse capacity due to uncertainty in the system parameters. The total variance of the collapse capacity, $\sigma_{\ln CC(TOT)}^2$, based on FOSM is (Ibarra and Krawinkler 2011):

$$\sigma_{\ln CC(TOT)}^2 \cong \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\partial g(\bar{X})}{\partial x_i} \frac{\partial g(\bar{X})}{\partial x_j} \right)_{\bar{X}=\bar{\mu}_x} \rho_{x_i, x_j} \sigma_{x_i} \sigma_{x_j} + \sigma_{\ln CC(RTR)}^2 \quad (6)$$

The first term on the right hand side of Eq. 6 represents the contribution to the variance of collapse capacity due to uncertainty of the system parameters, whereas $\sigma_{\ln CC(RTR)}^2$ is the contribution to the variance of collapse capacity due to RTR variability, and $\bar{X}$ represents the set of system random parameters. The function $g(\bar{X})$ is the collapse capacity as a function of the variation on system parameters, and the variance is computed from the gradient of $g(\bar{X})$, which is

Seismic Collapse Assessment,

Fig. 17 Uncertain parameters of the backbone curve (Modified from Ibarra and Krawinkler 2005)



linearized using a Taylor series expansion about the mean $\bar{X}$. The derivative $\partial g(\bar{X})/\partial x_i$ can be approximated by evaluating the performance function at two system parameter values. The derivative is estimated as the slope of the straight line that connects the pairs $[x_i, g(x_i)]$. The standard deviation of the system parameters, σ_{x_i} , and the correlation coefficients, ρ_{x_i, x_j} , need to be estimated in advance.

Ibarra and Krawinkler (2005, 2011) used the FOSM method to evaluate the effect of uncertainty in system parameters on the collapse capacity of SDOF systems. They found that uncertainty in the displacement at the peak (cap) strength and the post-capping stiffness contributes significantly to the variance of collapse capacity. The uncertainty in deterioration parameters on the variance of collapse capacity may be comparable to that caused by RTR variability.

Although the FOSM method is relatively simple to implement, it cannot predict the shift in the median caused by system uncertainties. Vamvatsikos and Fragiadakis (2010) concluded that this shortcoming should not be considered important for practical applications because for more cases the shift in the median is less than 10 %. Liel et al. (2009) found shifts in the median closer to 20 % and proposed a simplified method, termed ASOSM (approximate second order second moment), that uses FOSM to predict the increase in fragility's logarithmic standard deviation and the shift in the median of the limit state fragility. The total variance can also be approximated using the point estimate method, in which

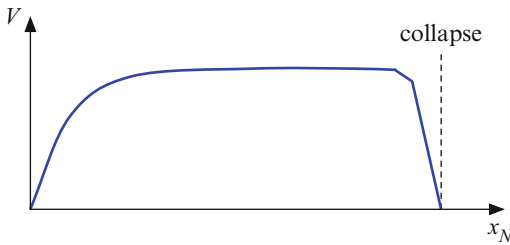
the first moments of a function are calculated in terms of the first moments of the random variables (Rosenblueth 1981). The total variance in collapse capacity can be more accurately predicted using Monte Carlo simulation, but this method is computationally expensive. An alternative is to use the Latin hypercube sampling (LHS) method to reduce Monte Carlo realizations. LHS is a stratified sampling that allows efficient estimation of the quantity of interest by reducing the variance of Monte Carlo simulations (Vamvatsikos and Fragiadakis 2010). In the example of Fig. 16, the FOSM method and the LHS method deliver almost the same prediction of the collapse fragility function.

Simplified Methods

IDA and MSA procedures require the numerical solution of the governing equations of motions of the structural model together with constitutive equations in each time step for each response history analysis. As a consequence, it is time-consuming and computationally expensive. Thus, there is a need for simple to apply, but yet sufficient accurate methods to predict the global collapse of MDOF structures under seismic excitations (Lignos and Krawinkler 2012).

Methods Based on Nonlinear Static Analysis

One of the simplest approaches is to conduct a nonlinear static analysis to estimate the seismic collapse capacity. During this so-called pushover



Seismic Collapse Assessment, Fig. 18 Global pushover curve up to global collapse (With permission from Adam 2014)

analysis, gravity loads are applied to the structural model that considers explicitly the inelastic constitutive behavior of its elements. Subsequently the model is subjected to lateral forces with a predefined invariant load pattern that is amplified incrementally in a displacement-controlled procedure. A first-mode load pattern may be used. Alternatively, the load pattern recommended in ASCE 7-10 (2010) can be utilized. As a result the global pushover curve of the structure is obtained, at which the base shear is plotted against the displacement of the control node (usually the roof displacement). The global pushover curve represents the global capacity of the building against horizontal loads, and it is assumed that it reflects the global or the local mechanism involved when the structure approaches dynamic instability. In the perspective of a “near-collapse” limit state, collapse is assumed if one of the selected EDPs, such as story drifts, plastic hinge rotations, etc., exceeds a certain threshold. When the base shear drops to zero (Fig. 18), global “instability collapse” is assumed to be attained.

Pushover analysis was originally developed for first-mode dominated structures and later refined to account also for higher mode effects (e.g., Chopra and Goel 2002). The basic assumption of these procedures is that the nonlinear static response can be related to the nonlinear dynamic response. This is, however, not the case for many structures because the dynamic response is strongly path dependent, and effects, such as cyclic deterioration, damping, and duration of an earthquake, among others, cannot be captured. Nonetheless, nonlinear static analysis

may be reasonable applied as an ingredient for the collapse assessment of regularly shaped low-rise buildings whose seismic response is dominated by the first mode. Even if the pushover procedure is not appropriate for a full collapse assessment, it can be used to identify weaknesses in the structural design, to debug a structural model used for collapse assessment, and to obtain a better understanding of the strength and deformation demands of the structure (Deierlein et al. 2010). In contrast, Villaverde (2007) states that it is doubtful that nonlinear static methods can be used reliably to predict the seismic collapse capacity of structures and to estimate their margin of safety against a global collapse. These methods, however, are commonly utilized in practical procedures. For example, Shafei et al. (2011) proposed a simplified methodology for collapse assessment of degrading moment-resisting frames and shear walls utilizing closed-form equations, given that the global pushover curve is provided. These equations correlate median and dispersion of collapse fragility curves and were generated through multivariate regression analysis from a comprehensive database of collapse fragilities and pushover curves.

Methods Based on Equivalent Single-Degree-of-Freedom Systems

Several studies use equivalent single-degree-of-freedom (ESDOF) systems to predict the global seismic collapse capacity (e.g., Bernal 1998; Fajfar 1999; Adam et al. 2004). Application of ESDOF systems implies that the first mode dominates the dynamic response and the collapse mode. A global collapse capacity assessment of low- to medium-rise buildings by means of an ESDOF system is straightforward when the story drifts remain rather uniformly distributed over the height, regardless of the extent of inelastic deformation. Adam and Jäger (2012b) showed that the application of these simplified systems yields reasonable collapse capacity predictions for P-delta-sensitive regular high-rise buildings, because P-delta-induced collapse is primarily controlled by the first mode. However, if a partial mechanism develops, the global collapse capacity is greatly affected by the change of the

deflected shape, and it will be amplified in those stories in which the drift becomes large (Bernal 1998).

Capacity Spectrum Method (N2 Method)

In the last decade, the capacity spectrum methodology became a popular tool for assessing the seismic performance of regular first-mode dominated structures. It represents a compromise between nonlinear dynamic analyses with complex modeling strategies and simplified linear static analysis methods. In Eurocode 8 (CEN 2005), a version of this method, i.e., the N2 method (Fajfar 1999), is implemented.

In this approach a global lateral load–displacement relation, which is the outcome of a pushover analysis, represents the seismic capacity of the structure and is referred to as capacity curve. The base shear V and the roof displacement x_N of a bilinear idealization of this curve are transformed into the domain of the corresponding ESDOF system according to the following Eurocode 8 equations (CEN 2005) (see Fig. 19):

$$\begin{aligned} f^* &= V/\Gamma, & D &= x_N/\Gamma, \\ \Gamma &= L^*/m^*, & L^* &= \sum_{i=1}^N m_i \phi_i, & m^* &= \sum_{i=1}^N m_i \phi_i^2 \end{aligned} \quad (7)$$

Here, f^* is the equivalent spring force, and D the displacement of the ESDOF system. The equation includes the story masses m_i , $i = 1, \dots, N$, of the ductile N -story structure, and the N components ϕ_i of the shape vector ϕ (with $\phi_N = 1$), which prescribes the vertical distribution

of the displacements of the structure in its ESDOF approximation (Fajfar 2002). Note that ϕ should be affine to the horizontal load pattern of the pushover analysis.

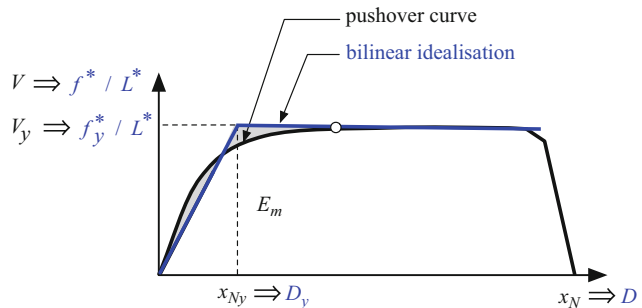
Dividing f^* by the equivalent mass L^* renders its ordinate in the “acceleration” dimension; see Fig. 19. This transformation permits the comparison of the capacity curve with the seismic demand represented by the response spectrum in the ADSR format of the actual site. In this format for an inelastic SDOF system with a target ductility μ , the spectral acceleration S_{av} at yield is plotted against the corresponding peak spectral displacement S_d (Fajfar 1999), as shown in Fig. 20a. Note that the ductility defines the ratio of the maximum imposed (inelastic) deformation to the deformation at onset of yield.

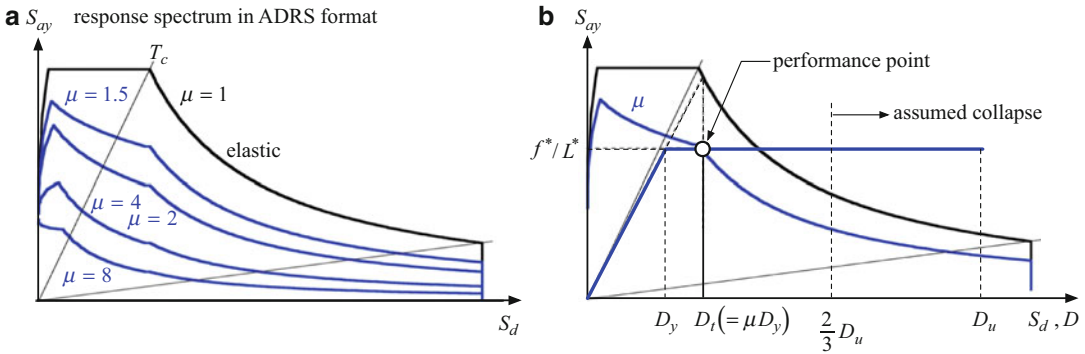
In the subsequent step, the intersection point between a bilinear approximation of the capacity curve and the response spectrum is obtained. If this point is in the inelastic branch of deformation, the elastic spectrum must be reduced such that the ductility of the capacity curve at this point and of the corresponding inelastic spectrum coincide, Fig. 20b. For periods larger than the corner period T_c , this intersection point, which is referred to as performance point, is found by application of the equal displacement rule.

According to Eurocode 8 (CEN 2005), structural stability under seismic excitation cannot be verified if the displacement demand at the performance point is larger than two-thirds of the ultimate deformation capacity of the structure or if no performance point can be found. Consequently, collapse is indicated, as presented in Fig. 20b.

Seismic Collapse Assessment,

Fig. 19 Capacity curve
With permission from
Adam (2014; Modified
from CEN 2005)





Seismic Collapse Assessment, Fig. 20 (a) Elastic and corresponding constant ductility response spectra in the ADRS format. (b) Performance point With permission from Adam (2014; Modified from Fajfar 1999)

Collapse Capacity Spectrum Method

The collapse capacity spectrum method is another example of a simplified ESDOF system-based method. This method predicts the global collapse capacity of regular P-delta-sensitive moment-resisting frame structures, where cyclic component deterioration can be omitted (Adam and Jäger 2012b). In a first step, the vulnerability of the structure to P-delta-induced collapse is evaluated based on global first-mode pushover curve of the structure with applied gravity loads. When this curve exhibits a negative post-yield stiffness, as shown in Fig. 15, the building may collapse under severe earthquake excitation in the P-delta mode. In this situation the structure is transformed into an ESDOF system.

Collapse capacity spectra (Adam and Jäger 2012a) are the further ingredient of this method. However, application of collapse capacity spectra is not straightforward for an MDOF structure, because the backbone curve of the ESDOF system is derived from the global first-mode pushover curves with and without considering P-delta. As discussed before, bilinear idealization of these curves does not exhibit a uniform stability coefficient as in a real SDOF system, but an elastic θ_e and an inelastic stability coefficient θ_i that is always larger, in some cases even much larger than θ_e : $\theta_i > (>) \theta_e$. Assigning a specific auxiliary backbone curve (Ibarra and Krawinkler 2005) to the ESDOF system solves this problem because its rotation by the uniform auxiliary stability coefficient θ_a ,

$$\theta_a = \frac{\theta_i - \theta_e \alpha_s}{1 - \theta_e + \theta_i - \alpha_s} \quad (8)$$

yields the backbone curve considering P-delta in analogy to a real SDOF system; see Fig. 21. The parameter α_s is the strain-hardening coefficient of the global pushover curve without P-delta. The discussed backbone curves of the ESDOF system are depicted in Fig. 21: The bilinear curve with the largest strength represents the auxiliary backbone curve, and the graph with the smallest strength is the backbone curve with P-delta. Subsequently the negative post-yield stiffness ratio $\theta_a - \alpha_s$ of the auxiliary ESDOF system is evaluated. Since the initial stiffness of the auxiliary backbone curve is larger than the original one, the period of the ESDOF system with assigned auxiliary backbone curve is (Adam et al. 2004)

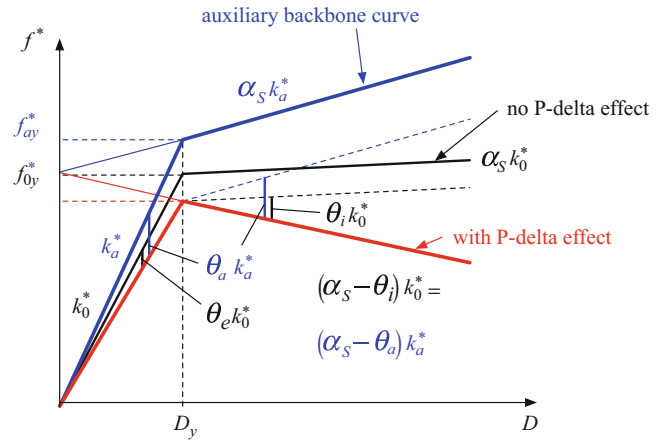
$$T_a = 2\pi \sqrt{\frac{1 - \alpha_s}{1 - \theta_e + \theta_i - \alpha_s}} \sqrt{\frac{x_{Ny}}{V_y}} \sqrt{\sum_{i=1}^N \phi_i m_i} \quad (9)$$

In Eq. 9 x_{Ny} and V_y denote the roof displacement and the base shear, respectively, at the onset of yield of the global pushover curve without P-delta; see Fig. 15.

Now the appropriate collapse capacity spectrum with respect to the underlying ground motion set, viscous damping ζ , the hysteretic loop, and the negative post-yield stiffness $\theta_a - \alpha_s$ is consulted to obtain the median (CC_d), 16th

Seismic Collapse Assessment,

Fig. 21 Backbone curves with and without P-delta and auxiliary backbone curve (Modified from Adam and Jäger 2012b)



(CC_d^{p16}), and 84th percentile (CC_d^{p84}) collapse capacities of a SDOF system at the period T_a of the auxiliary ESDOF system. For instance, in Fig. 14 CC_d is read at a period $T_a = 3.2$ s for $\theta_a - \alpha_S = 0.20$. These collapse capacity quantities are transformed into the domain of the ESDOF system:

$$CC_{\text{ESDOF}} = \frac{CC_d}{\lambda_{\text{MDOF}}},$$

$$\lambda_{\text{MDOF}} = \left(\sum_{i=1}^N \phi_i m_i \right)^2 / \left(\sum_{i=1}^N m_i \sum_{i=1}^N \phi_i^2 m_i \right) \quad (10)$$

λ_{MDOF} is the transformation coefficient that relates the actual structure with the corresponding SDOF system. For details it is referred to Fajfar (2002). CC_{ESDOF} is an estimate of the median collapse capacity of the MDOF system, $CC_{\text{MDOF}} \approx CC_{\text{ESDOF}}$. The same applies to the percentiles: $CC_{\text{MDOF}}^{p16} \approx CC_{\text{ESDOF}}^{p16}$ and $CC_{\text{MDOF}}^{p84} \approx CC_{\text{ESDOF}}^{p84}$. Since the collapse capacities follow generally a log-normal distribution, an approximation of the collapse fragility function is derived from

$$\ln N(m, \sigma^2), \quad m = \ln(CC_{\text{ESDOF}}),$$

$$\sigma = \ln \sqrt{CC_{\text{ESDOF}}^{p16} CC_{\text{ESDOF}}^{p84}} \quad (11)$$

Application of this method for assessing the seismic collapse fragility of a series of generic

and real moment-resisting frame structures showed that the collapse capacity can be estimated quick, but yet accurate, without conducting time-consuming dynamic analyses (Adam and Jäger 2012b). The collapse capacity spectrum method is particularly useful in engineering practice because the structure can be evaluated with respect to its seismic collapse capacity in the initial design process without a detailed dynamic analysis.

Summary

This article summarizes procedures currently used for assessing global collapse of structures induced by strong-motion earthquakes. Through intensive research in the last two decades, substantial progress was achieved in this field, and the outcomes of these studies are compiled in numerous publications. Thus, in the present contribution, only some of the issues involved in seismic collapse assessment are elaborated in detail. The selection of the presented material is naturally biased by research and experience of the authors.

Seismic collapse assessment is a branch of earthquake engineering that combines multidisciplinary fields of research such as seismology, structural dynamics, materials science, applied mathematics, and computational mechanics. It includes adequate prediction of the seismic hazard, ground motion selection, identification of

possible modes of collapse, modeling of cyclic component deterioration, appropriate consideration of hysteretic and viscous damping, quantification of modeling and parameter uncertainties, and nonlinear dynamic analyses based on stable numerical algorithms.

For accurate assessment of seismic structural collapse, however, there are still challenges that need to be addressed. For example, understanding of nonlinear cyclic component behavior in a condition of oblique bending requires experimental tests and advances in constitutive modeling. Another issue is the quantification of the contribution of nonstructural elements on the seismic collapse capacity. On the other hand, from engineering practice, there is a strong demand on simplified assessment strategies that are simple to apply and at the same time reasonably accurate.

Cross-References

- [Analytic Fragility and Limit States \[P\(EDP|IM\)\]: Nonlinear Dynamic Procedures](#)
- [Analytic Fragility and Limit States \[P\(EDP|IM\)\]: Nonlinear Static Procedures](#)
- [Conditional Spectra](#)
- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- [Seismic Risk Assessment, Cascading Effects](#)
- [Site Response for Seismic Hazard Assessment](#)

References

- Adam C (2014) Structural dynamics and earthquake engineering (in German). Internal report. University of Innsbruck
- Adam C, Jäger C (2012a) Seismic collapse capacity of basic inelastic structures vulnerable to the P-delta effect. *Earthq Eng Struct Dyn* 41:775–793
- Adam C, Jäger C (2012b) Simplified collapse capacity assessment of earthquake excited regular frame structures vulnerable to P-delta. *Eng Struct* 44:159–173
- Adam C, Ibarra LF, Krawinkler H (2004) Evaluation of P-delta effects in non-deteriorating MDOF structures from equivalent SDOF systems. In: *Proceedings of the 13th world conference on earthquake engineering (13 WCEE)*, Vancouver, 1–6 Aug 2004, digital paper, 15 pp
- ASCE/SEI 7-10 (2010) Minimum design loads for buildings and other structures. American Society of Civil Engineers, Reston
- ASCE/SEI 41-13 (2014) Seismic evaluation and retrofit of existing buildings. American Society of Civil Engineers, Reston
- ATC-58 (2012) Seismic performance assessments of buildings, 100% draft. Applied Technology Council, Redwood City
- ATC-72-1 (2010) Modeling and acceptance criteria for seismic design and analysis of tall buildings. Applied Technology Council, Redwood City
- ATC-78 (2013) Identification and mitigation of seismically hazardous older concrete buildings: interim methodology evaluation. Applied Technology Council, Redwood City
- Baker JW (2013) Trade-offs in ground motion selection techniques for collapse assessment of structures. In: Adam C, Heuer R, Lenhardt W, Schranz C (eds) *Proceedings of the Vienna Congress on recent advances in earthquake engineering and structural dynamics 2013 (VEESD 2013)*, Vienna, 28–30 Aug 2013, digital paper, paper no. 123, 10 pp
- Baker JW, Cornell CA (2005) A vector-valued ground motion intensity measure consisting of spectral acceleration and epsilon. *Earthq Eng Struct Dyn* 34:1193–1217
- Baker JW, Cornell CA (2006) Spectral shape, epsilon and record selection. *Earthq Eng Struct Dyn* 35:1077–1095
- Baradaran Shoraka MB (2013) Collapse assessment of concrete buildings: an application to non-ductile reinforced concrete moment frames. PhD thesis, The University of British Columbia, Vancouver
- Baradaran Shoraka MB, Elwood KJ, Yang TY, Liel AB (2014) Collapse assessment of non-ductile, retrofitted and ductile reinforced concrete frames. In: Elwood KJ, Kim InSung, Dragovich J (eds) *Seismic assessment of existing reinforced concrete buildings: new developments*, ACI special publication ACI SP 297, SP297-8, 19 pp., American Concrete Institute, Farmington Hills, MI
- Bernal D (1987) Amplification factors for inelastic dynamic P-Δ effects in earthquake analysis. *Earthq Eng Struct Dyn* 15:635–651
- Bernal D (1998) Instability of buildings during seismic response. *Eng Struct* 20:496–502
- CEN (2005) Eurocode 8: design of structures for earthquake resistance. Comité Européen de Normalisation (CEN), Brussels
- Charney FA (2008) Unintended consequences of modeling damping in structures. *J Struct Eng* 134:581–592

- Chopra AK, Goel RK (2002) A modal pushover analysis procedure for estimating seismic demands for buildings. *Earthq Eng Struct Dyn* 31:561–582
- Clough RW, Johnston SB (1966) Effects of stiffness degradation on earthquake ductility requirements. In: Proceedings of the 2nd Japan earthquake engineering symposium, Tokyo, October 1966, pp 227–232
- Deierlein GG, Reinhorn AM, Willford MR (2010) Nonlinear structural analysis for seismic design. A guide for practicing engineers. NEHRP Seismic Design technical brief no. 4, produced by the NEHRP Consultants Joint Venture, a partnership of the Applied Technology Council and the Consortium of Universities for Research in Earthquake Engineering, for the National Institute of Standards and Technology, Gaithersburg, NIST GCR 10-917-5
- Esteva L, Ruiz S (1989) Seismic failure rates of multistory frames. *J Struct Eng* 115:268–284
- Fajfar P (1999) Capacity spectrum method based on inelastic demand spectra. *Earthq Eng Struct Dyn* 28:979–993
- Fajfar P (2002) Structural analysis in earthquake engineering – a breakthrough of simplified non-linear methods. In: Proceedings of the 12th European conference on earthquake engineering, London, 9–13 September 2002, CD-ROM paper, Paper ref. 843. Elsevier, p 20
- Fardis MN, Biskinis DE (2003) Deformation capacity of RC members, as controlled by flexure or shear. In: Toshimi K, Hitoshi S (eds) Proceedings of the international symposium on performance-based engineering for earthquake resistant reinforced structures, a volume honoring Prof. Shunsuke Otani, Tokyo, 8–9 September 2003, pp 511–530
- FEMA P-695 (2009) Quantification of building seismic performance factors. Federal Emergency Management Agency, Washington, DC
- Gatto KS, Uang CM (2002) Effects of loading protocol and rate of loading on woodframe shearwall response. In: Seventh U.S. national conference on earthquake engineering (7 NCEE), Boston, MA, 21–25 July 2002
- Gupta A, Krawinkler H (2000) Dynamic P-delta effects for flexible inelastic steel structures. *J Struct Eng* 126:145–154
- Haselton CB, Liel AB, Deierlein GG (2009) Simulating structural collapse due to earthquakes: model idealization, model calibration, and numerical simulation algorithms. In: Papadrakakis M, Lagaros ND, Fragiadakis M (eds) Proceedings of the 2nd international conference on computational methods in structural dynamics and earthquake engineering (COMPDYN 2009), Rhodes, 22–24 June 2009, CD-ROM paper, paper no. CD497
- Haselton CB, Liel AB, Deierlein GG, Dean BS, Chou JH (2011) Seismic collapse safety of reinforced concrete buildings. I: assessment of ductile moment frames. *J Struct Eng* 137:481–491
- Ibarra LF, Krawinkler H (2005) Global collapse of frame structures under seismic excitations. Report No. PEER 2005/06, Pacific Earthquake Engineering Research Center, University of California, Berkeley
- Ibarra LF, Krawinkler H (2011) Variance of collapse capacity of SDOF systems under earthquake excitations. *Earthq Eng Struct Dyn* 40:1299–1314
- Ibarra LF, Medina RA, Krawinkler H (2005) Hysteretic models that incorporate strength and stiffness deterioration. *Earthq Eng Struct Dyn* 34:1489–1511
- Jäger C, Adam C (2012) Seismic collapse assessment of non-deteriorating frames with irregular structural properties vulnerable to P-delta. In: Proceedings of the 15th world conference on earthquake engineering (15 WCEE), Lisbon, 24–28 Sept 2012, digital paper, paper no. 0245, 10 pp
- Jäger C, Adam C (2013) Influence of collapse definition and near-field effects on collapse capacity spectra. *J Earthq Eng* 17:859–878
- Jalayer F, Cornell CA (2009) Alternative non-linear demand estimation methods for probability-based seismic assessments. *Earthq Eng Struct Dyn* 38:951–972
- Jennings PC, Husid R (1968) Collapse of yielding structures during earthquakes. *J Eng Mech Div Proc ASCE* 94:1045–1065
- Krawinkler H, Zareian F, Lignos DG, Ibarra LF (2009) Prediction of collapse of structures under earthquake excitations. In: Papadrakakis M, Lagaros ND, Fragiadakis M (eds) Proceedings of the 2nd international conference on computational methods in structural dynamics and earthquake engineering (COMPDYN 2009), Rhodes, 22–24 June 2009, CD-ROM paper, paper no. CD449
- Kunnath SK, Reinhorn AM, Park YJ (1990) Analytical modeling of inelastic response of R/C structures. *J Struct Eng ASCE* 116:996–1027
- Liel AB, Deierlein GG (2013) Cost-benefit evaluation of seismic mitigation alternatives for older reinforced concrete frame buildings. *Earthq Spectra* 29:1392–1411
- Liel AB, Haselton CB, Deierlein GG, Baker JW (2009) Incorporating modeling uncertainties in the assessment of seismic collapse risk of buildings. *Struct Saf* 31:197–211
- Lignos DG, Krawinkler H (2012) Sidesway collapse of deteriorating structural systems under seismic excitations. Report no. TR 177, The John A. Blume Earthquake Engineering Research Center, Department of Civil and Environmental Engineering, Stanford University, Stanford
- Lignos DG, Krawinkler H (2013) Development and utilization of structural component databases for performance-based earthquake engineering. *J Struct Eng ASCE* 139:1382–1394
- MacRae GA (1994) P-Δ effects on single-degree-of-freedom structures in earthquakes. *Earthq Spectra* 10:539–568

- Medina RA, Krawinkler H (2003) Seismic demands for nondeteriorating frame structures and their dependence on ground motions. Report no. 144. The John A. Blume Earthquake Engineering Research Center, Department of Civil and Environmental Engineering, Stanford University, Stanford
- Melchers RE (1999) Structural reliability analysis and prediction, 2nd edn. Wiley, Chichester
- Moehle J, Deierlein GG (2004) A framework methodology for performance-based earthquake engineering. In: Proceedings of the 13th world conference on earthquake engineering (13 WCEE), Vancouver, 1–6 Aug 2004, digital paper
- NIST GCR 10-917-8 (2010) (ATC-76-1). Evaluation of the FEMA P-695 methodology for quantification of building seismic performance factors. National Institute of Standards and Technology, Gaithersburg
- Panagiotakos TB, Fardis MN (2001) Deformations of reinforced concrete members at yielding and ultimate. *ACI Struct J* 98:135–148
- Park Y, Ang A (1985) Mechanistic seismic damage model for reinforced concrete. *J Struct Eng ASCE* 111:722–739
- Rosenblueth E (1981) Point estimates for probability. *Appl Math Model* 5:329–335
- Shafei B, Zareian F, Lignos DG (2011) A simplified method for collapse capacity assessment of moment-resisting frame and shear wall structural systems. *Eng Struct* 33:1107–1116
- Shome N, Cornell CA (1999) Probabilistic seismic demand analysis of nonlinear structures, reliability of marine structures program report no. RMS-35, Department of Civil and Environmental Engineering, Stanford University
- Sivaselvan MV, Reinhorn AM (2000) Hysteretic models for deteriorating inelastic structures. *J Eng Mech* 126:633–640
- Song J, Pincheira J (2000) Spectral displacement demands of stiffness and strength degrading systems. *Earthq Spectra* 16:817–851
- Ugurhan B, Baker JW, Deierlein GG (2014) Uncertainty estimation in seismic collapse assessment of modern reinforced concrete moment frame buildings. In: Proceedings of the tenth U.S. national conference on earthquake engineering (10 NCEE), Anchorage
- Vamvatsikos D, Cornell CA (2002) Incremental dynamic analysis. *Earthq Eng Struct Dyn* 31:491–514
- Vamvatsikos D, Fragiadakis M (2010) Incremental dynamic analysis for estimating seismic performance sensitivity and uncertainty. *Earthq Eng Struct Dyn* 39:141–163
- Villaverde R (2007) Methods to assess the seismic collapse capacity of building structures: state of the art. *J Struct Eng* 133:57–66
- Zareian F, Krawinkler H, Ibarra L, Lignos D (2010) Basic concepts and performance measures in prediction of collapse of buildings under earthquake ground motions. *Struct Des Tall Spec Build* 19:167–181

Seismic Design of Dams

Guoxi Wu

Wutec Geotechnical International, New Westminster, Metro Vancouver, BC, Canada

Synonyms

CSR; Design seismic parameters; Earthfill dams; Factor of safety; Granular and plastic soils; Limit equilibrium; Newmark sliding block; Nonlinear finite element analysis; Performance-based design; PWP model; Risk analysis; Soil liquefaction; SPT; Time history

Introduction

Design of new dams or safety evaluation of existing dams for seismic loads is standard practice and routinely required. In a broad term, dams can be classified into three types – concrete dam, rock-fill dam, and earthfill dam – and seismic design can be at a dam safety level and/or at a serviceability level.

This entry provides contents and discusses methods for design and analysis of earthfill dams at a dam safety level where the ultimate limit state is applied for the highest level of design earthquake loads. For dam safety, dam failure is the primary concern; for serviceability consideration, the dam should remain functional and any damages should be easily repairable under this level of earthquake loads.

Dam Performance in Past Earthquakes

Damages to dams under earthquakes can result from ground shaking, soil liquefaction, ground cracking, ground displacements (lateral spreading and settlement), and in extreme cases surface rupture along an earthquake-active fault. Examples of recent big and devastating earthquakes include the 2008 Sichuan earthquake (a crustal

earthquake with a magnitude of M8.0 occurred in China) and the 2010 Chile earthquake (a subduction earthquake with M8.8 occurred off the coast of central Chile).

Historical data show that dams around the world have performed well and satisfactory and that the probability of a dam failure under strong ground motion shaking is low. Nearly all well-built and well-compacted embankment dams can withstand moderate earthquake shaking with peak ground accelerations (PGA) greater than 0.2 g. Dams constructed of clay soils on clay, or on rock, or on overburden foundations resistant to liquefaction have withstood (with no apparent damages) extremely strong shaking with PGAs from 0.35 to 0.8 g.

Soil liquefaction, either in the dam fills or in the foundation, is the most damaging factor that affects the performance of dams under earthquakes. Dams built of sandy soils, especially hydraulic or semi-hydraulic fills, or built on foundations of loose (low density) sandy soils are highly susceptible to earthquake damage or vulnerable to failures due to the potential for soil liquefaction. A famous case history of such is the near failure of the Lower San Fernando Dam in California, USA, in a 1971 earthquake (Seed 1979).

Design Seismic Loads Based on a Risk Analysis

Public tolerance to seismic risk for the consequence of dam failure ultimately determines the adequacy of an existing dam or the criteria for the design of a new dam. For instance, at a specific location with the same tectonic setting or geological condition and thus the same earthquake hazard condition, a nuclear reactor facility may be designed using a PGA of 1.0 g for a very low probability earthquake event, a hydroelectric dam using a PGA of 0.7 g for a low probability earthquake event, a residential building using a PGA of 0.5 g for a median probability of happening, and a temporary bridge for construction traffic using a PGA of 0.3 g for a relatively high probability of

occurring. This is mainly because the societal risk tolerance on a relative scale increases in order from nuclear radiation leak to flood from a dam breach, to collapse of a residential building, and to loss of a temporary structure.

The first factor contributing to seismic risk, such as from an earthquake event with a PGA of 0.7 g, is the occurrence rate of such earthquake which is often measured by annual exceedance probability (AEP). The other factor is the consequence as a result of an earthquake, which is ultimately measured in terms of loss of life and sometimes economic loss. At a conceptual level, seismic risk can be expressed as the product of seismic hazard probability and consequence, and it represents the probabilistic expectation of the consequence.

Ideally, seismic design of a dam should be based on a risk analysis, including calculation of the actual probability of a dam failure and its consequence. The adequacy of the dam would then be judged from the seismic risk (such as annual probability of single or multiple fatalities) that is acceptable to the society for the loss of life involved in a dam failure.

The risk-based approach to dam safety evaluation should balance the public risk and the limited societal resources available to manage the particular risk. As shown in Fig. 1, it is considered generally acceptable that the maximum level of societal risk to fatality is less than 10^{-3} per annum for loss of one life and the risk is less than 10^{-5} per annum when more than 100 lives would be lost in the event of a dam failure. The principle that the risks should be as low as reasonably practicable (ALARP) is generally followed in practice, and it is thus reasonable to use an annual probability of 10^{-5} and 10^{-6} for 100 and more fatalities.

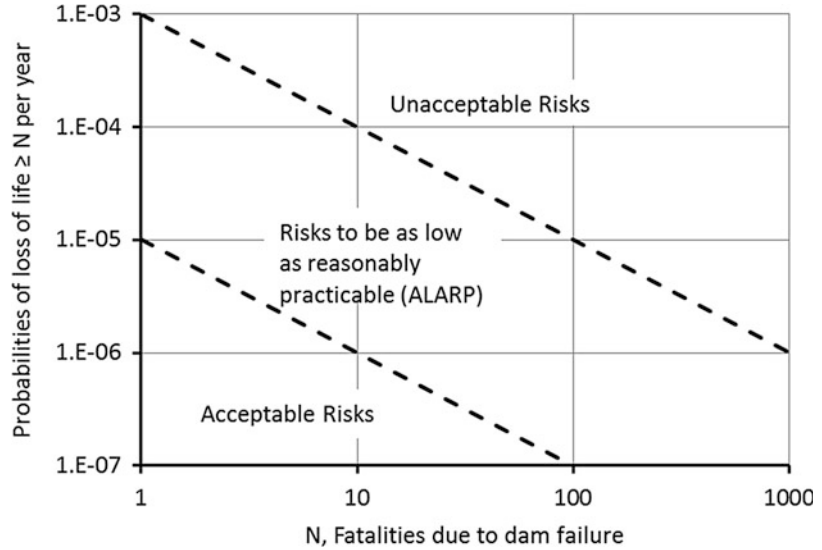
In mathematical terms, seismic risk to loss of life (fatality) is calculated using the following two equations:

$$P_{\text{Fatality}} = P_{\text{Failure}} \times P_{\text{Fatality/Failure}}$$

and

$$P_{\text{Failure}} = P_{\text{Earthquake}} \times P_{\text{Failure/Earthquake}}$$

Seismic Design of Dams,
Fig. 1 An example
 relationship to manage
 societal risk for dam safety



where:

P_{Fatality} = unconditional probability of fatality from earthquake, i.e., seismic risk.

$P_{\text{Earthquake}}$ = unconditional probability for an earthquake to occur, such as the one with a PGA of 0.7 g.

$P_{\text{Failure/Earthquake}}$ = conditional probability of a dam failure in event of the earthquake.

$P_{\text{Fatality/Failure}}$ = conditional probability of loss of life in the event of a dam failure.

The probabilities caused by all dam failure initiating events (failure modes) need to be aggregated or compounded in order to obtain the total probability to a dam failure. Failure modes (such as soil liquefaction) for earthquake loading should be identified in advance in order to perform a risk calculation. The risk calculation for a particular failure mode (P_{Failure}) is carried out using an event tree approach where significant events are sequenced in levels. Construction of an event tree for a dam requires special knowledge in geotechnical earthquake engineering and a comprehensive understanding of the dam.

Using “soil liquefaction” as an example failure mode, an event tree for dam failure could consist of earthquake acceleration (Level 1), contributing earthquake magnitude (Level 2), liquefaction analysis method (Level 3), soil

liquefaction capacity (Level 4), dam crest settlement magnitude (Level 5), reservoir water level (Level 6), and dam damage level (Level 7). For illustration purpose, at Level 1 for earthquake acceleration, it may have four scenarios with a PGA of 1.0, 0.7, 0.5, and 0.3 g, respectively, and each PGA scenario has its probability of occurrence (AEP). At Level 3 assuming two models for liquefaction analysis, the weighting factors (e.g., 0.55 and 0.45) would be assigned to each model to make a total weighting of 1.

In many cases where a dam consequence class is available after completion of a life safety model and analysis, it can be conservatively assumed that the conditional probability ($P_{\text{Fatality/Failure}}$) equals to 1.

Although in recent years the risk-based seismic dam safety evaluation is increasingly used in dam safety management, it is not widely used in engineering design or seismic safety evaluation due to the limited ability to perform such a complex risk-based analysis. As described below, the standard-based or traditional approach is more commonly used in seismic design.

Standard-Based Method and Design Earthquake AEP

The standard-based method is a semi-probabilistic method that defines the seismic

hazard using the probabilistic approach, but does not explicitly calculate the probability of a dam failure and its consequence. Without any quantifiable risk calculation, this is an empirical or experience-based approach to risk evaluation and management. The consequence, i.e., conditional fatalities in the event of a dam failure, is indirectly evaluated by using a dam consequence class scheme (FEMA 1998; CDA 2007). In some countries the highest consequence class is “Extreme” when potential (expected) fatalities are 100 and more.

With this approach, a seismic hazard level (AEP) is selected according to the consequence class of a dam so that the selected AEP conforms to the societal acceptable risk level. For a dam under an “Extreme” class, it would be reasonable to target a dam failure probability of 10^{-5} to 10^{-6} per annum, and an earthquake with an AEP of 10^{-5} or higher could be adequate, taking into account the satisfactory post-earthquake performance of dams around the world, i.e., the low conditional probability of a dam failure under strong shaking. For background information, the nuclear industry uses the 84th percentile spectra at AEP of 10^{-4} or approximately median-mean spectra at a probability of 10^{-5} per annum. With less stringent safety criteria than the nuclear industry, dam safety design can adopt a design earthquake with an AEP of 1/10,000 (i.e., 10^{-4}) based on the mean spectra. The mean is the expected value given the epistemic uncertainties, and the mean hazard value typically such as in Canada varies from 65th to 75th percentiles in the hazard distribution.

The design earthquake AEP for an “Extreme” class dam can vary from one region to another or from one country to another depending on the risk tolerance ability, and it is normally jointly selected by the dam owner and the government regulatory agency. In some developing countries with lower societal reliability than the developed countries, an earthquake AEP of 1/1,000 is used for dam safety design. On the other hand, even in developed countries such as Canada, an earthquake AEP of 1/2,475 is used for building safety design as a result to balance public risk and economic cost.

Once a seismic hazard level is selected, conventional analyses such as limit equilibrium or finite element analyses are conducted using seismic loads corresponding to the selected AEP. The results (such as stresses, ground displacements, or stability in a dam or its foundation) and consequences are evaluated deterministically using standards, specifications, and design codes. The potential for a dam failure is evaluated by comparing deterministically the resulting stresses and displacements with ultimate stability and established failure criteria. The evaluations of results are primarily done by empirical evidence, past experience, and engineering judgment.

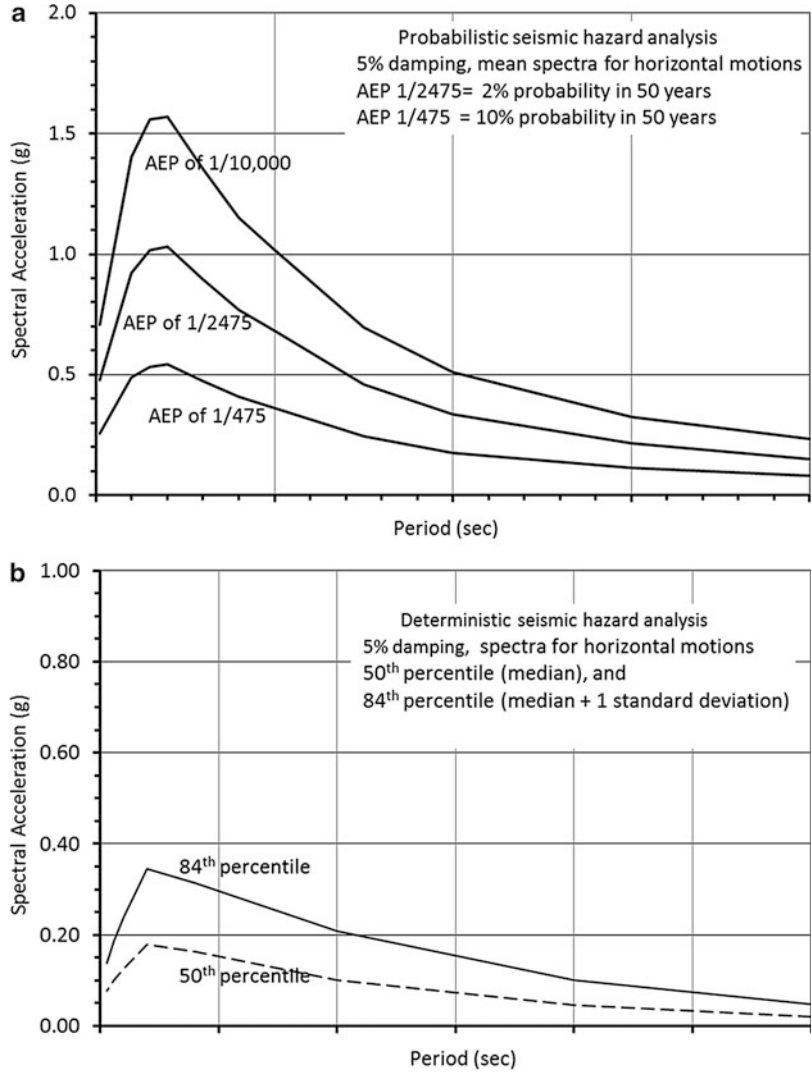
Seismic Hazard Evaluations

A deterministic seismic hazard analysis (DSHA) is normally conducted for each known earthquake source to determine site ground motion parameters such as PGA, response spectra at 50th and 84th percentile, magnitude of earthquake, and duration of strong shaking. This would include assessment of potential seismic hazards from earthquake activities along local or regional known faults for the life of a dam.

A probabilistic seismic hazard analysis (PSHA) is usually conducted to determine design seismic parameters at various AEP levels, in addition to DSHA where known active faults or subduction zones are identified. A PSHA normally consists of **identification** of earthquake sources by developing a source model and the earthquake occurrence rate for each source zone with the understanding of local geology and regional past earthquake history, **application** of ground motion prediction equations (Earthquake Spectra 2008) that are appropriate to the region (e.g., hardness and shear wave velocity of the bedrock) and the types of seismic sources, and **determination** of earthquake response spectra at various levels of probability by the integration of hazard contributions over all earthquake magnitudes and distances for all seismic source zones.

The response spectra, either from DSHA or from PSHA, are normally defined for a series of discrete natural periods or frequencies (such as

Seismic Design of Dams,
Fig. 2 Example results of
seismic hazard analyses (a)
probabilistic method and
(b) deterministic approach



0.1, 0.2, 0.5, 1, 2, 5, 10, and 33 Hz; the frequency for PGAs can range from 33 to 100 Hz) that are used in the hazard calculation, and they are always associated with an uncertainty level (e.g., median, mean, or 84th percentile).

A seismic hazard analysis determines the potential intensity of seismic loads that could hit a damsite; and it always proceeds to, but does not always relate to, the design of a new dam or performance evaluation of an old dam for earthquake loading. However, seismic design of dams uses results of a seismic hazard analysis. Figure 2 shows

typical response spectra from DSHA and PSHA at the most commonly used damping ratio of 5 %.

Design Seismic Parameters

Seismic parameters for design of dams consist mainly of peak ground accelerations (PGAs), site-specific response spectra for horizontal and vertical accelerations, magnitude and site-source distance for the design earthquake, duration of strong shaking, ground motion time histories as

required for dynamic analyses, and fault displacements in rare situations when a dam is (or to be) on an active fault.

Design Response Spectra and PGAs

Response spectra and PGAs from DSHA can be used directly in seismic design, and they are conventionally computed at an 84th percentile value in the hazard distribution, i.e., one standard deviation (using $\varepsilon = 1$ in ground motion prediction equations) above the predicted median value. In some cases for seismic faults with low rate of activities, the 50th percentile values (the median) are used in design.

When seismic hazard is evaluated using a probabilistic approach, response spectra and PGAs from PSHA are used for seismic safety design of dams, using the results for the selected AEP (e.g., mean spectra with AEP of 1/10,000).

In current practice, the vertical ground motions do not seem to have much effect on the performance of earthfill dams and thus are not normally included in design analysis.

Design Earthquake Magnitude and Site-Source Distance

The PSHA results represent at a specific AEP level a composite of hazard contributions from earthquakes of all magnitudes and distances. The response spectra from a PSHA are also called the uniform (or equal) hazard response spectra (UHRS).

Deaggregation of the composite seismic hazard is performed to identify relative contributions of individual earthquakes or scenario earthquakes with various magnitudes and distances. For seismic design of a dam, representative earthquake magnitudes (M) and site-source distances (D) are obtained by deaggregation of the uniform hazard at natural periods that are significant and critical to the dam seismic response.

The representative M and D for the design earthquake are then used in seismic analysis where magnitude and duration of the earthquake are needed such as in the soil liquefaction analysis, or for selection of ground motion records needed in a time-history analysis.

Input Ground Motion Time Histories

The current trend in seismic design of dams is to conduct linear or nonlinear time-history analysis to obtain dynamic response of dam to earthquake loads. Time-history analysis of dam requires input ground motion time histories (acceleration, velocity, and displacement).

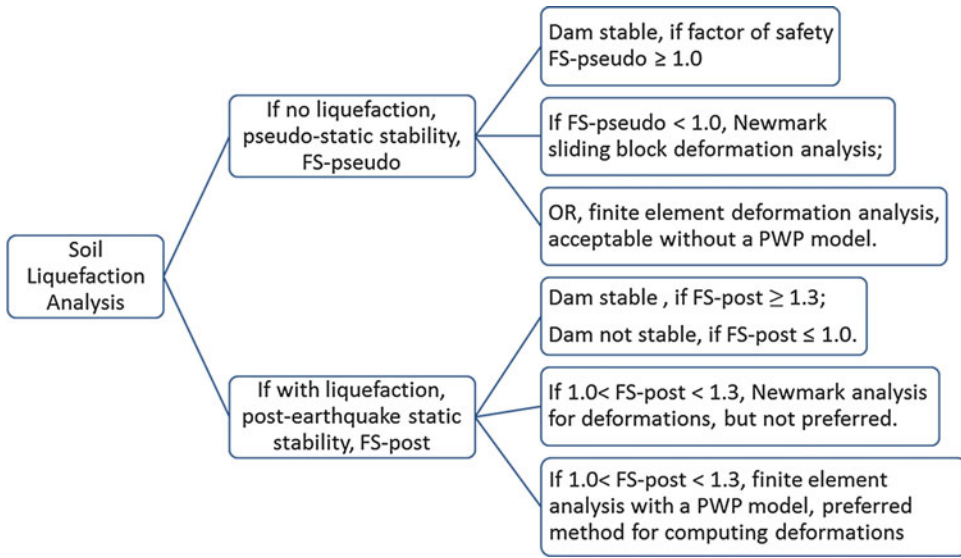
For dynamic analysis of earthfill dams, multiple ground motion records (five to eight commonly adopted) from past earthquake are selected, and each record is then linearly scaled to fit the design response spectra (such as a UHRS) over the range of natural periods that are appropriate for the dam. Using this method, an individual ground motion record is multiplied by a single scale factor to increase or decrease the magnitude of the motion and its spectrum, without modifying the shape of its spectrum or its frequency contents.

The selected ground motion records should be consistent with the seismic parameters and representative to the design earthquake in terms of magnitude, site-source distance, duration of strong shaking, tectonic setting and source mechanism, and consistency of site conditions between the recording station and the dam.

In some cases when uniform scaling of recorded ground motions is unable to meet the requirements, the design earthquake ground motions would be obtained by modifying a recorded ground motion in the time or frequency domain.

General Approach to Seismic Design and Analysis

For dams under static conditions, limit equilibrium analyses for slope stability are normally conducted for long-term operating reservoir conditions (steady-state seepage) and for short-term reservoir drawdown conditions. In general, for satisfactory stability factor of safety (FS) would be minimum 1.5 for long-term conditions ($FS > 1.5$) and 1.3 for short-term conditions ($FS > 1.3$).



Seismic Design of Dams, Fig. 3 Types of seismic analyses based on liquefaction susceptibility (PWP = excess pore water pressures from earthquake loads)

For dams under seismic loading, two primary modes of dam failure must be addressed, the overtopping failure caused by excessive settlement of dam crest and the internal erosion and piping failure caused by cracks in dams, damage of filter layer upstream, or drains downstream of the core zones.

Some of the common design measures to mitigate these failures include the following: remove or improve problematic foundation soils by ground treatment and adequately compact the dam fills, use wide core zones of plastic soils that are resistant to erosion, use well-graded wide filter zones upstream of the core, construct chimney drains downstream of the core to lessen soil saturation and reduce downstream seepage, and use dam crest and downstream slope details to provide protection of the dam in the event of an overtopping.

Seismic analyses appropriate to a dam and its foundation are conducted using design seismic parameters to provide adequate information on expected stresses and ground displacements for evaluation of expected performance of the dam. Seismic analyses (see Fig. 3) can include a soil liquefaction evaluation, a pseudo-static stability analysis, a Newmark sliding block deformation

analysis, a post-earthquake static stability analysis, and a finite element dynamic analysis for computing permanent ground deformations.

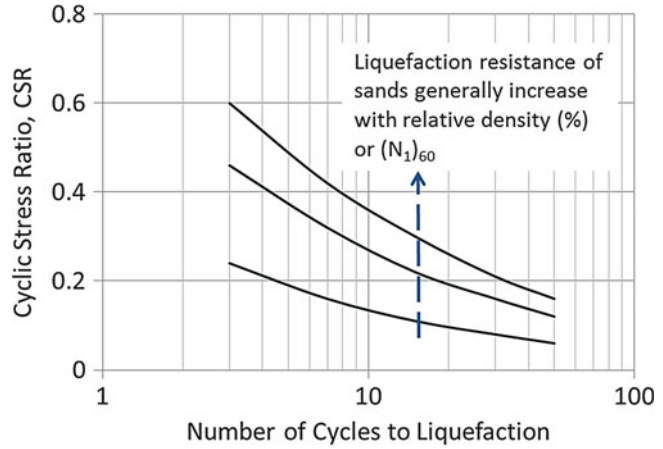
Evaluation of Soil Liquefaction

Liquefaction of Sands Under Cyclic Loads

Sand liquefaction is a fluidization process of saturated sand mass subject to cycles of shear stress. Under shaking, it can be easily observed that loose sands in a dry container will experience volume contraction and settle to a more compact state. In a saturated condition, immediate volume change of the sand would not occur because water in the pore does not drain quickly enough under the rapid earthquake loading. Instead, the potential for volume change translates into a quick increase in excess pore water pressure (PWP) in the sand mass. Liquefaction occurs when the PWP exceeds a threshold value that the pore water effectively suspends sand particles. Sand boiling to ground surface is a surficial expression of liquefaction of sands in the ground.

Liquefaction strength, or capacity, or resistance (more commonly used) of sands can be measured by laboratory testing using

Seismic Design of Dams,
Fig. 4 Liquefaction
resistance trend curves for
clean sands (<5 % fines)



reconstituted samples prepared to a target relative density. It is found (see Fig. 4) that liquefaction resistance generally increases with relative density as expected, and under the same relative density, it decreases with more cycles of shear stress used in the testing.

Because liquefaction is triggered by multiple cycles of shear stresses, liquefaction resistance of soils is represented using both the cyclic shear stress level (τ_{cyc}) and the number of cycles for such uniform (constant amplitude) stresses to trigger a liquefaction failure. Cyclic stress ratio (CSR) refers to the ratio of the cyclic shear stress to the initial vertical normal stress (σ'_{v0}) at which the sample is under cyclic shearing, i.e., $CSR = \tau_{cyc}/\sigma'_{v0}$. For earthquake loading, the number of cycles is related to earthquake magnitude and duration. In engineering analysis, nonuniform cycles of shear stress from a M7.5 earthquake are artificially set equal to, in terms of net effect on soil, a shear stress level with 15 uniform cycles. Thus, liquefaction resistance of soil is typically characterized by CSR at 15 cycles or CSR_{15} as shown in Fig. 5.

Liquefaction Resistance of Sandy Soils

Except with ground-frozen sampling technique that is practically not applicable, liquefaction resistance of in situ sands is considered not measurable from laboratory testing.

The most widely adopted test for field evaluation of relative density for sands is the standard penetration test (SPT) with split-barrel soil

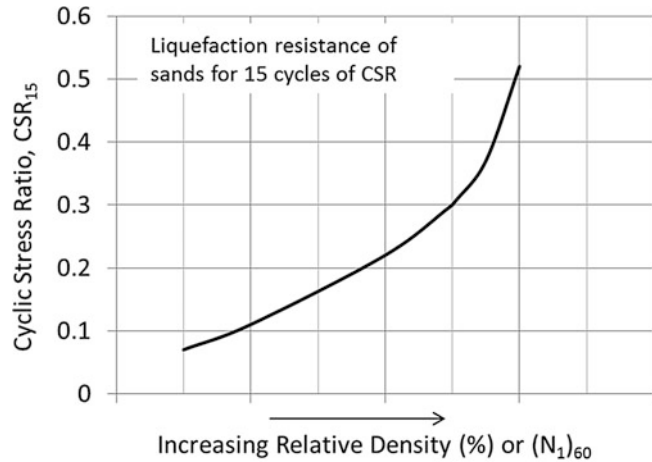
sampling. A key result of SPT is SPT N value that is the number of blow counts for a penetration depth of 305 mm (1 ft.), and $(N_1)_{60}$ is defined as the N value corrected for the soil under an in situ effective vertical normal stress of about 100 kPa (1 atmospheric pressure) and to the standard SPT hammer energy level (i.e., at 60 % of the theoretic maximum hammer energy). The general understanding is that the measured SPT N value is corrected downward (i.e., reduced) when the test is performed for soil at a depth with a confining stress higher than 100 kPa (e.g., at 300 kPa).

Current technology or methodology for evaluating liquefaction resistance of sandy soils was developed empirically by indirect studies of performance case histories of dams in historical earthquakes, where in some cases soil liquefaction was assessed to have occurred and in other cases liquefaction was believed not triggered.

Field soil liquefaction resistance of sands (particle size from 0.075 to 4.75 mm) has been derived from field performance of soils under various levels of historical earthquakes, and CSR_{15} from case history studies is plotted with in situ measured $(N_1)_{60}$ (Seed and Harder 1990) as an equivalence to the relative density of sand. Similar to the curve shown in Fig. 5, sands with $(N_1)_{60} > 30$ are considered to be dense and generally not liquefiable; sands with $(N_1)_{60} < 10$ are generally loose and highly susceptible to liquefaction.

Seismic Design of Dams,

Fig. 5 An example liquefaction resistance curve for clean sands



Unless completely confined with no drainage path, gravels (particle size from 4.75 to 75 mm) are not so much vulnerable to soil liquefaction due to its high permeability that allows fast dissipation of excess pore water pressures. However, sandy soils containing some amount of gravels are liquefiable, and their liquefaction resistance can be tested in the field using a large penetration hammer such as the Becker Hammer Penetration Test (Youd et al. 2001).

At a given relative density or $(N_1)_{60}$, sandy soils containing significant amount of fines (particle size < 0.075 mm) are known to have higher liquefaction resistance than clean sands with no fines (Youd et al. 2001). Nonplastic silts, containing 100 % fines but with very low plasticity (such as plasticity index $PI < 5$ %), are considered to behave under cyclic loading in a similar manner as sands. The liquefaction resistance of nonplastic silts is evaluated as sands with upward correction on fines content.

Liquefaction Resistance of Plastic Soils

A more descriptive terminology for liquefaction failure of plastic soils (silts or clayey silts), containing 100 % fines and with relatively high PI (such as $PI > 10$ %), is “cyclic strain softening” which emphasizes more on the structural breakdown of the material under cyclic loading than on the buildup of excess pore water pressure. Evaluation of liquefaction resistance for plastic soils is an area that requires more research. In the

1980s and 1990s of the last century, soil index parameters (water content, liquid limit, PI , and fines content) were used as basis for liquefaction assessment, but the current trend method is to use results of laboratory cyclic tests (Finn and Wu 2013), especially for evaluation of more critical structures such as for dam safety.

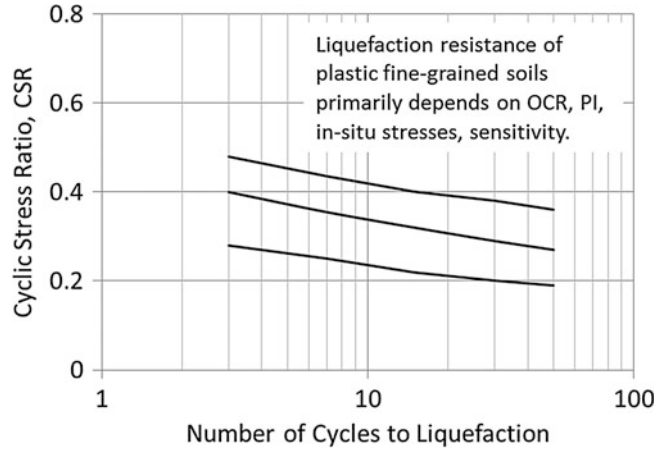
Liquefaction resistance of in situ plastic soils can be measured directly by laboratory testing on in situ Shelby samples. Soil samples in a cyclic direct simple shear test are normally first consolidated to its pre-consolidation pressure (the highest historic pressure experienced by the soil or σ'_p) and then sheared by cycles of τ_{cyc} under σ'_{v0} and initial shear stress (τ_{st}) similar to the in situ stresses of the sample.

As compared to sands where the relative density or $(N_1)_{60}$ is the single parameter, there are more parameters and factors affecting and contributing to the liquefaction resistance for plastic soils (see Fig. 6). The curves for plastic soils are generally flatter than curves for sands; that is, the resistance decreases less with increasing number of cycles. The two key factors are the over-consolidation ratio ($OCR = \sigma'_{v0}/\sigma'_p$) and PI . Liquefaction resistance of plastic soils generally increases with increasing OCR and PI but decreases with increasing τ_{st} .

As a general observation, plastic soils having $CSR_{15} > 0.4$ (for 15 cycles) are considered to be hard, insensitive to earthquake loading, and unlikely to experience much strength loss. Plastic

Seismic Design of Dams,

Fig. 6 Liquefaction resistance trend curves for plastic soils, generally flatter than sands



soils having CSR_{15} near or less than 0.15 may be sensitive to shaking and can experience significant loss of strengths in a strong earthquake.

Earthquake-Induced CSR

For liquefaction analysis, seismic load in a soil element is represented by time history of shear stresses (τ_{dyn}) imposed by earthquake shaking, and they can be calculated from a site response analysis using the input ground motions at the damsite and dynamic properties of dam fills and foundation soils.

Soil dynamic properties for each of the soil zones in the dam and its foundation primarily consist of low-strain shear modulus (G_{max}) and the damping characteristics (Seed et al. 1986). The low-strain shear modulus (G_{max}) is commonly computed from shear wave velocity (V_s) that can be measured by seismic downhole or crosshole surveys. Shear modulus would reduce from G_{max} as shear strain increases; reduction of soil shear modulus (stiffness) with increasing strains is more significant for gravels than for sands and more for sands than for clays. Soil damping ratio would be less than 5 % at low strain, but it increases with increasing strain. The maximum damping ratio is about 25 % for gravels and sands and in the order of 20 % for clays and silts.

A seismic site response analysis (► [Site Response: 1-D Time Domain Analyses](#)) is commonly performed in a total stress analysis, without including PWP effect, using a 1D soil column

analysis for a low dam or using a 2D finite element dynamic analysis to model the geometric effect for a high dam. A site response analysis, either 1D or 2D model, can be conducted using the equivalent linear method (Idriss et al. 1974) or a true nonlinear method (Finn et al. 1986) for the simulation of shear modulus degradation (decrease) and damping ratio increase with the increase of shear strain. A true nonlinear approach is considered more appropriate when the ground shaking level is high, such as with a PGA of 0.4 g or higher.

Factor of Safety Against Liquefaction

Upon completion of a site response analysis, cyclic stress ratio ($CSR = \tau_{dyn}/\sigma'_{v0}$) in each soil zone is calculated by converting the nonuniform cycles of shear stresses from the earthquake to equivalent cycles of uniform shear stresses (Wu 2001) and then normalized to the in situ vertical normal stress (σ'_{v0}).

The liquefaction resistance (CSR_{15}) of soils is determined, for sands using $(N_1)_{60}$ from SPT and for plastic soils using liquefaction resistance curves from laboratory cyclic shear tests.

A factor of safety against soil liquefaction for each soil zone is calculated to be the ratio of liquefaction resistance of soil and seismic shear stress from earthquake, i.e., $FS_{LIQUEFACTION} = CSR_{15}/CSR$. A computed $FS_{LIQUEFACTION}$ near or less than 1.0 indicates triggering of soil liquefaction, and $FS_{LIQUEFACTION} > 1.5$ means not liquefiable.

Pseudo-Static Limit Equilibrium Stability Analysis

A pseudo-static analysis is a limit equilibrium method which includes additional seismic inertia forces in a conventional static slope stability analysis. Seismic coefficients (k_h for horizontal and k_v for vertical) are often used in such analysis, and they are normally taken as the peak ground accelerations (PGAs) as a fraction of the gravity acceleration. The seismic inertia forces are represented by $k_h W$ and $k_v W$ in horizontal and vertical direction, respectively, where W is the weight of a sliding block. Due to the nature of seismic shaking, shear strengths of soils for rapid loading conditions are conventionally used for characterization of soil resistance to earthquake loadings.

In a scenario that soils in a dam and its foundation would not develop significant PWP from shaking, a pseudo-static factor of safety greater than 1 (FS-pseudo >1) is a very strong indication that there would be little or no damage to the dam from an earthquake.

However, a pseudo-static factor of safety less than 1 (FS-pseudo <1) does not necessarily represent dam instability or unsatisfactory performance; instead due to the transient nature of earthquake motions, the dam would undergo some deformations for the short time interval when the ground acceleration exceeds the yield acceleration. In such case seismic ground deformations are computed and used for performance assessment.

It is not possible to predict failure of a dam or estimate seismic ground deformations by a pseudo-static analysis, and therefore other types of more comprehensive analysis are required to provide a more reliable basis for seismic design of dams.

Newmark Sliding Block Seismic Deformation Analysis

Newmark (1965) pointed out that for seismic loading permanent ground deformations of a dam, in addition to the pseudo-static factor of

safety, should be considered to be tolerable or not for dam performance assessment. The Newmark deformation method assumes that deformation of a dam is modeled by a rigid block sliding on an assumed failure surface under the design ground motions at the damsite.

Sliding deformation is assumed to occur whenever the acceleration of a sliding block exceeds the yield acceleration, which is the horizontal acceleration that results in a factor of safety of 1 in a pseudo-static slope stability analysis, and the sliding stops when the acceleration falls below the yield acceleration. Mathematically, the Newmark sliding block displacements are computed by double integration of acceleration time history on the sliding block for the portion of accelerations that exceeds the yield acceleration (Fig. 7).

Although it has been widely used at a time when there is lack of direct methods for computing seismic ground deformations, the Newmark sliding block model is generally recognized to be not a very good representation of how embankment dams deform under earthquake, especially for dams with low factors of safety. Due to its simplicity, the method may provide a range of expected ground deformations, but it would neither give distribution of ground displacements in the dam nor provide direct calculation of dam crest settlement.

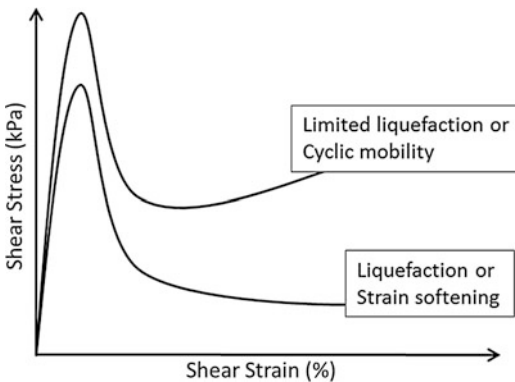
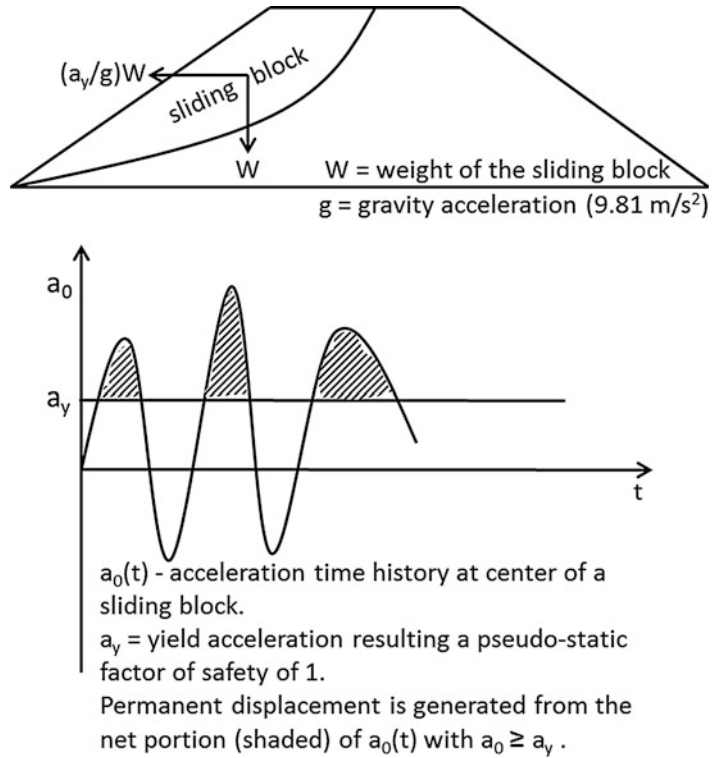
Post-Earthquake Static Stability Analysis

A direct consequence of soil liquefaction is a great reduction of shear strength, in particular for loose and contractive sands (Fig. 8), for which the shear strength may become a small fraction of its static and drained value due to the buildup of excess pore water pressures and fluidization. In design practice, shear strength of liquefied soil is also described as residual strength and sometime as post-liquefaction strength.

As a frictional material and under static loading, loose sand can have a friction angle possibly ranging from 30° to 35° ; however, the post-liquefaction strength of the same loose sand

Seismic Design of Dams,

Fig. 7 Newmark sliding block deformation calculation diagram



Seismic Design of Dams, Fig. 8 Shear stress – strain response indicating great strength loss after liquefaction and some strength loss with cyclic mobility

may decrease to an equivalent friction angle of 5° or perhaps less (Seed and Harder 1990), indicating the degree of shear strength reduction due to liquefaction.

Residual strength of plastic soils after initial strain softening could decrease to 50–80 % of its pre-earthquake static value, and after sufficient

shaking disturbance, it can eventually decrease to the remolded strength value. Depending on sensitivity, remolded strength of silts and clayey silts may range from about 25 % of its peak static strength for low-sensitivity soils to only 10 % for high-sensitivity soils.

A static limit equilibrium analysis, without seismic loads but using the post-earthquake soil strength, is conducted in order to assess the post-earthquake static stability of the dam. The analysis would use residual strength for liquefied soils and reduced strength in soils that do not liquefy but with the buildup of excess pore water pressures from the design earthquake shaking.

A factor of safety less than 1 from a post-earthquake static stability analysis (FS-post < 1) is a strong indication of dam instability and a potential failure during or immediately after the design earthquake. As such, the design for a new dam would be improved and ground strengthening for an existing dam would be required, in order to increase the post-earthquake dam stability and meet the dam safety requirements.

If $FS_{\text{post}} > 1$ is calculated, ground deformations from the design earthquake are often computed for dam performance evaluation, preferably using a nonlinear finite element dynamic analysis. The Newmark sliding block deformation analysis should be avoided for computing deformations of dams involving soil liquefaction.

Seismic Deformations by Finite Element Dynamic Analysis

A more rigorous and direct method for computing earthquake-induced ground deformations, with and without soil liquefaction, is a nonlinear finite element dynamic analysis (Finn et al. 1994) that overcomes the uncertainties associated with the approximation embedded in the pseudo-static stability calculation and in the Newmark deformation analysis. With PWP calculations built into effective stress soil models, the finite element method has the capability to simulate the coupled effects from dynamic site response, generation of excess pore water pressures, soil liquefaction, and post-liquefaction behavior, and this is a trend method for quantitative evaluation of earthquake-induced ground deformations.

A well-known analysis procedure for the finite element approach is the one developed by Finn et al. (1986) using soil data from centrifuge tests and then verified using soil data and field measured deformations of the Upper San Fernando Dam in California, USA, in a 1971 earthquake (Wu 2001), and recently extended to include a dilative silt model (Finn and Wu 2013). This analysis procedure is highly regarded by practicing engineers and referenced for seismic design of dams by the US Federal Emergency Management Agency (FEMA 2005).

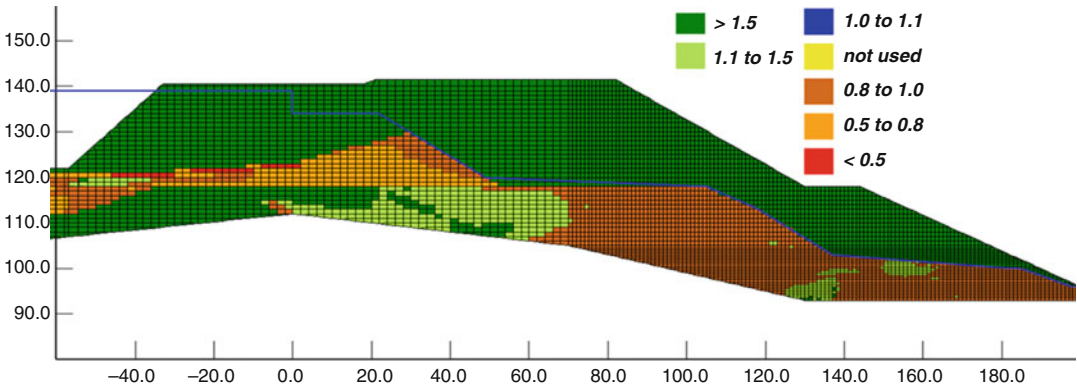
The Finn-Wu finite element procedure is simple and well suitable for practical engineering analysis. This analysis method adopts Mohr-Coulomb criteria for modeling shear strength of soils and uses a nonlinear hyperbolic model to simulate hysteresis response of soils under cyclic

loads. Excess pore water pressures (PWP) caused by earthquake loads can also be computed using effective stress models with calculation of factor of safety against liquefaction for each soil element. In addition, the procedure provides two options for applying input ground motions: (1) ground accelerations applied at the model's rigid base and (2) outcropping ground velocities applied at the model's viscous and elastic base. The elastic base option is often used when the model base stiffness is not rigid relatively to the model body that prevents incident seismic waves from effectively reflecting back to the model body.

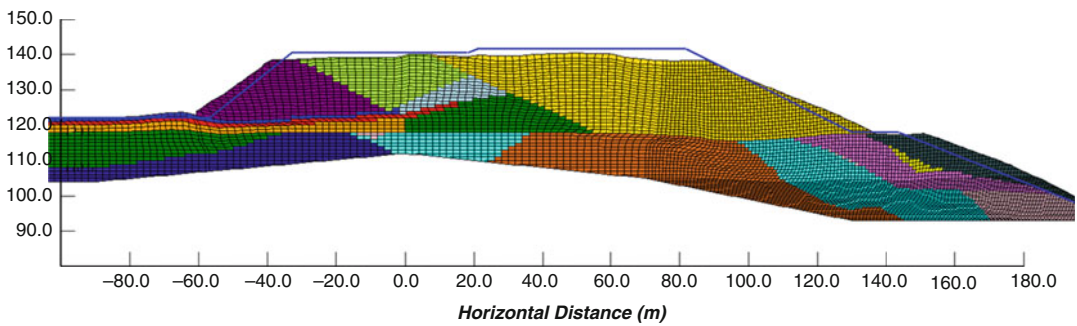
As an example of Finn-Wu finite element procedure, factors of safety against liquefaction computed from a dynamic analysis using a subduction earthquake input motion are shown in Fig. 9 (Finn and Wu 2013). The results showed that almost the entire saturated zone (below the water table shown in blue) under the downstream slope of the dam would undergo large cyclic strains with $FS_{\text{LIQUEFACTION}} < 1$. A computed deformed mesh of the dam, with soil material zones shown in colors, immediately after the earthquake using the same subduction ground motion is shown in Fig. 10, indicating a deep-seated ground deformation pattern and significant lateral movement and settlement of the downstream dam crest. The original dam surface is outlined in blue in the figure.

Seismic Performance Evaluation

In some cases, expected performance of dam can be assessed to either clearly safe or clearly unsafe, based on results of simple limit equilibrium stability analyses. In majority cases, however, dam performance is assessed from estimated ground deformations and strains. Ultimately, the expected ground deformations and the associated ground strains, either from a Newmark analysis or a nonlinear finite element analysis, should not exceed what the dam can safely withstand without catastrophic release of the reservoir water due to overtopping or internal



Seismic Design of Dams, Fig. 9 Computed $FS_{LIQUEFACTION}$ shown in colors for an example finite element model of an earthfill dam (undeformed, partial model, dimension in m)



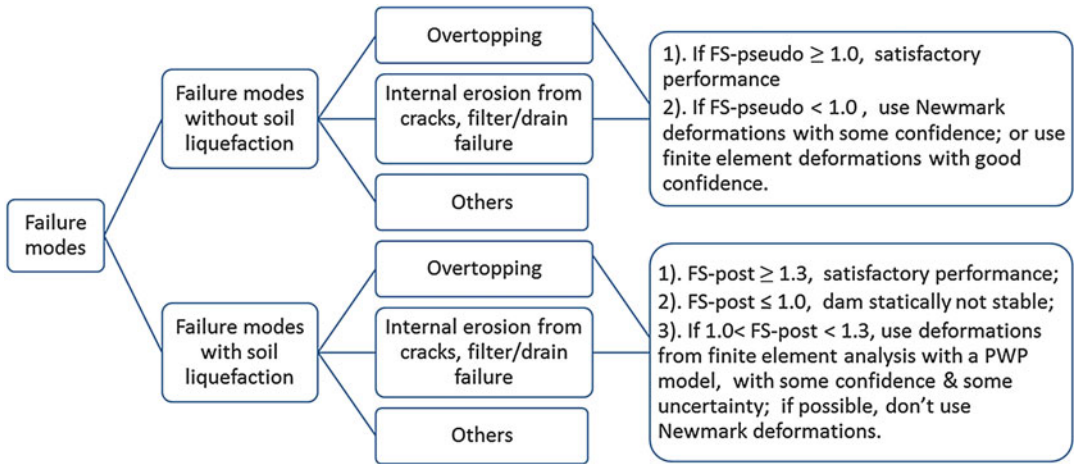
Seismic Design of Dams, Fig. 10 Deformed dam cross section after a subduction earthquake (displacement scale factor 1.0, elevation for vertical axis in m)

erosion. In the process, uncertainties associated with the calculated factor of safety or with the computed ground deformation need to be understood to achieve the confidence on the expected performance. In some cases, the consequences of misjudging uncertainty levels would also be taken into considerations.

Two key factors contributing to uncertainties in predicting the expected performance of dam under earthquake are related to soil liquefaction. The first is “Will liquefaction be triggered?”, and the second is “What is the residual strength after soil liquefaction?” Ideally, a risk analysis would be beneficial to quantify the uncertainties. However, state of practice is mostly based on a deterministic approach for dam performance assessment, such as using analysis steps outlined in Fig. 11.

Overtopping of a dam in earthquake can be caused either by slope instability (failure) or by excess ground deformations (settlement) at dam crest. If post-earthquake stability analyses indicate a factor of safety well above 1 such as $FS_{post} > 1.3$, historical dam performance experience in earthquakes would indicate that the dam will have limited or small deformations and will perform satisfactorily. When liquefaction is not involved, factor of safety greater than 1 with seismic force in a pseudo-static analysis ($FS_{pseudo} > 1$) would also indicate that the dam will perform well in a design earthquake.

On the other hand, confidence in dam safety decreases and probability of a dam failure due to overtopping or internal erosion increases when a post-earthquake factor of safety (FS_{post}) near or less than 1.0 is calculated using residual strengths



Seismic Design of Dams, Fig. 11 Evaluation of dam performance for critical failure modes under earthquake

for liquefied soils. A FS-post less than 1.0 would indicate slope instability under post-earthquake static conditions, and a dam slope failure can occur immediately after the earthquake.

In general, when a wedge or circular sliding surface has a low post-earthquake factor of safety, the deformations along the slip surface would be large or sometime excessive. In these cases, ground deformations calculated preferably from a finite element dynamic analysis are used to assess the potential of a dam failure by overtopping, or internal erosion, or loss of soils in piping.

Engineering judgment is carefully applied in assessing the level of uncertainties in design parameters and analytical methodology and thus the confidence level in the use of calculated deformations. When soil liquefaction is not an issue, deformations estimated from a Newmark sliding block analysis would prove to be adequate for many cases. For dams with liquefaction issues, seismic design or evaluation would use deformations computed from a finite element dynamic analysis, which has adequate soil models for simulation of soil liquefaction and post-liquefaction behavior and also can compute large-strain ground deformations. Even using a finite element approach, computed deformations for dams without involving soil liquefaction would be more reliable (with less uncertainties) than those involving liquefaction.

Summary

This entry describes basic principles and methodology used in seismic analysis and design for earthfill dams.

The risk-based approach for dam safety evaluation is introduced to indicate that seismic design criteria are governed by societal tolerance of seismic risk which consists of both seismic hazard and its consequence. In engineering practice, a full risk analysis is not commonly performed due to its complexity; instead, consequence of a dam failure in an earthquake is often represented using a dam consequence class scheme. A design level of seismic hazards is then selected from the consequence class for a dam. Seismic hazard of a damsite can be evaluated on a probabilistic approach. Once design seismic parameters are determined using the selected seismic hazard level, seismic analysis and design are conducted using a deterministic approach where seismic demands are compared with the ultimate capacity of soils.

State-of-practice methodology for evaluation of soil liquefaction is described, which include measurement of liquefaction resistance of soils by laboratory tests, assessment of liquefaction resistance from in situ tests in the field, soil parameters that impact the liquefaction resistance, and typical values of liquefaction resistance. Liquefaction potential of a soil element is

evaluated by comparing the seismic shear stress from earthquake with the liquefaction resistance of the soil.

Seismic analyses discussed in this entry include pseudo-static stability analysis, Newmark sliding block deformation analysis, post-earthquake static stability analysis, and finite element dynamic analysis for computing permanent ground deformations.

It is pointed out that soil liquefaction is the key factor contributing to uncertainties in predicting the expected performance of dam under earthquake, which are based on results of limit equilibrium stability analyses and in majority cases from estimated ground deformations. Engineering judgment should be applied in assessing the level of uncertainties in design parameters and analytical methodology and thus the confidence level in the use of calculated factors of safety or ground deformations.

Cross-References

- [Probabilistic Seismic Hazard Models](#)
- [Site Response: 1-D Time Domain Analyses](#)

References

- Canadian Dam Association (CDA) (2007) Dam safety guidelines. CDA
- Earthquake Spectra (2008) Special issue on the next generation attenuation project. *Earthq Spectra Spec Issue* 24(1):639–644
- Federal Emergency Management Agency (FEMA) (1998) Federal guidelines for dam safety: hazard potential classification systems for dams, FEMA 333, Oct 1998. Reprinted April 2004
- Federal Emergency Management Agency (FEMA) (2005) Federal guidelines for dam safety: earthquake analyses and design of dams, FEMA 65, May 2005
- Finn WDL, Wu G (2013) Dynamic analyses of an earthfill dam on over-consolidated silt with cyclic strain softening, Keynote lecture. In: 7th international conference on case histories in geotechnical engineering, Chicago, Apr 2013
- Finn WDL, Yogendrakumar M, Yoshida N, Yoshida H (1986) TARA-3: a computer program to compute the response of 2-D embankments and soil structure interaction systems to seismic loadings, Department of Civil Engineering, University of British Columbia, Vancouver
- Finn WDL, Ledbetter RH, Marcuson WF (1994a) Seismic deformations in embankments and slopes. In: Balasubramaniam AS et al (eds) *Proceedings symposium on developments in geotechnical engineering from Harvard to New Delhi 1936–1994*, Bangkok. Balkema Publication. ISBN 90 5410 522 4, pp 233–264
- Finn WDL, Ledbetter RH, Wu G (1994b) Liquefaction in silty soils: design and analysis. *Ground failures under seismic conditions*. ASCE Geotech Spec Publ 44:51–76
- Idriss IM, Seed HB, Serff N (1974) Seismic response by variable damping finite elements. *J Geotech Eng ASCE* 100(1):1–13
- Newmark NM (1965) Effects of earthquakes on dams and embankments. *Geotechnique* 15(2):139–160
- Seed HB (1979) 19th Rankine lecture: considerations in the earthquake resistant design of earth and rockfill dams. *Geotechnique* 29(3):215–263
- Seed RB, Harder LF (1990) SPT-based analysis of cyclic pore pressure generation and undrained residual strength. In: *Proceedings of H. Bolton seed memorial symposium*, vol 2. BiTech Publishers, Vancouver, pp 351–376
- Seed HB, Wong RT, Idriss IM, Tokimatsu K (1986) Moduli and damping factors for dynamic analyses of cohesionless soils. *J Geotech Eng ASCE* 112(11):1016–1032
- Wu G (2001) Earthquake induced deformation analyses of the Upper San Fernando dam under the 1971 San Fernando earthquake. *Can Geotech J* 38:1–15
- Youd TL et al (2001) Liquefaction resistance of soils: summary report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of liquefaction resistance of soils. *J Geotech Geo-Environ Eng ASCE* 127(10):817–833 (21 authors)

Seismic Design of Earth-Retaining Structures

Luigi Callisto

Department of Structural and Geotechnical Engineering, Sapienza Università di Roma, Rome, Italy

Introduction

Earth-retaining structures provide support to excavations or earth fills: the earth pressure is either transmitted to the soil located below the foundations of the structure, as in gravity

retaining walls, or balanced by a combination of horizontal passive forces in the ground and the reaction forces of additional structural elements, as in the case of embedded retaining structures. Seismic forces have a detrimental effect on these structures, producing an increase in the earth pressure and a decrease in the available soil resistance. Commonly, the seismic design of an earth-retaining structure is carried out using a pseudo-static approach, in which the seismic forces are derived from a uniform acceleration field and are applied statically to the soil-structure system. This approach has the advantage of being relatively simple to implement, but the rationale for the choice of the constant acceleration values needs careful consideration. Therefore, the following section provides a simple conceptual model explaining how the pseudo-static acceleration values can be derived from the estimated maximum acceleration and from the acceptable seismic performance of the structure. The remaining sections provide the essential formulas for the evaluation of the seismic earth pressure and deal with specific issues relative to the seismic design of gravity and embedded retaining structures.

Conceptual Model for the Seismic Behavior of Retaining Structures

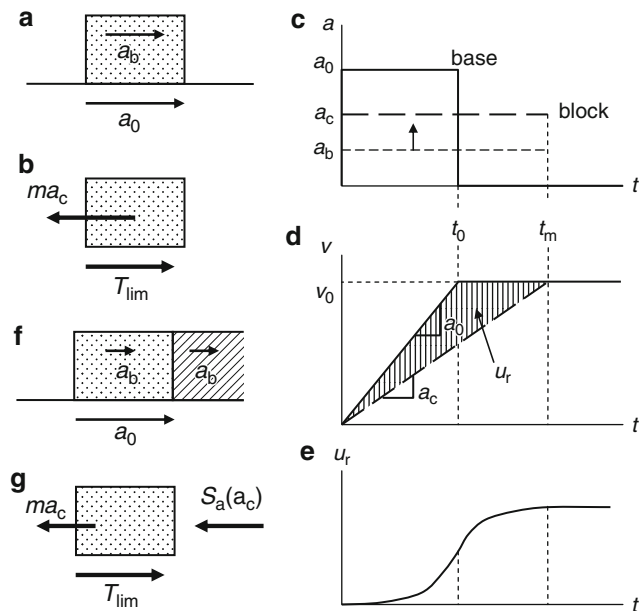
During an earthquake, the structure and the retained soil are accelerated and are therefore subjected to the corresponding inertial forces that, combining with the preexisting gravitational forces, produce additional loading on the retaining structure. It is important to note that the seismic inertial forces are different from the gravitational forces, as they vary cyclically in amplitude and direction, and have a transient nature. This has important consequences on the seismic design of a retaining structure.

Displacing Retaining Structures

An effective illustration of the seismic behavior of a retaining structure is provided by the simple scheme of a rigid block with a mass m resting on a horizontal base (Fig. 1a). The base is subjected to a simple dynamic excitation, consisting of a horizontal acceleration a_0 that is kept constant over a time interval t_0 , and is then removed (Newmark 1965). If a_b is the block's acceleration, then the corresponding inertial force is ma_b that must be balanced by the resultant force T applied by the

Seismic Design of Earth-Retaining Structures,

Fig. 1 Schematic representation of a rigid block subjected to an acceleration pulse at the base



base to the bottom surface of the block (Fig. 1b). If the capacity at the block-base contact is T_{lim} , then the acceleration of the block cannot be larger than its critical value $a_c = T_{\text{lim}}/m$.

Now suppose that a_0 is larger than a_c (Fig. 1c): since the acceleration of the block cannot go beyond a_c , a relative acceleration occurs between base and block and the corresponding velocity plots tend to diverge, as shown in Fig. 1d. The shaded area between the two velocity plots represents the relative displacement u_r between base and block; if the horizontal acceleration at the base of the block were kept constant at a_0 for a long period of time, the system would be loaded permanently by the inertial forces, and the relative displacement would increase rapidly with an upward concavity. This would be by definition a collapse mechanism: the system capacity is attained under constant forces, and the displacements accelerate.

In the present case, since a_0 is maintained only for the time interval t_0 , it can be seen that as t gets larger than t_0 , the sign of the relative acceleration changes, and the relative velocity decreases. Therefore, the relative displacements increase with a downward concavity and eventually stop when the relative velocity equals zero (Fig. 1e): at the end of this specific event, since the system's capacity has been attained only transiently, the system has not collapsed, but rather suffered a damage in the form of a permanent relative displacement. Whether this damage is acceptable or not depends on the specific function of the system.

This schematic illustration shows that because of the transient nature of the seismic forces, the attainment of the full strength of the system during an earthquake does not imply a collapse, but rather a damage deriving from the temporary activation of a plastic mechanism. In principle, it would be possible to design the structure in such a way that its resistance be never attained during the seismic event ($a_c = T_{\text{lim}}/m > a_0$), but this approach to the design is not particularly desirable, especially for severe, low probability events, because it may be uneconomical and also because the seismic behavior would not benefit

from the dissipation of energy taking place during the activation of the plastic mechanism.

In addition to the inertial forces related to its own mass, a retaining structure is loaded by the earth thrust that in turn increases because the masses of the retained soil are accelerated. Therefore, a scheme more appropriate to a retaining structure is that of Fig. 1f, where the block is loaded additionally by $S_a(a_b)$, which is the resultant force of the total earth pressures corresponding to the acceleration a_b . (Expressions for $S_a(a_b)$ are given in the next section.) Hence in this case a given acceleration produces larger inertial forces. The critical acceleration a_c is derived from the equilibrium equation (Fig. 1g):

$$m a_c + S_a(a_c) = T_{\text{lim}} \quad (1)$$

and is smaller than that of the previous scheme of Fig. 1a: the conceptual model is essentially the same, but from Fig. 1d and e, it is evident that a smaller critical acceleration corresponds to larger relative velocities and hence to larger displacements.

The schematic model of Fig. 1 evidences an important principle: since the block's acceleration cannot exceed a_c , the internal forces acting within the block cannot go beyond those evaluated for $a = a_c$. Since a_c is related to the seismic resistance of the system, it follows that stronger systems are called to resist larger internal forces.

With reference to the seismic design of a retaining structure, one can depict two different cases:

1. The critical acceleration of the system a_c is larger than the maximum acceleration at the base a_0 ; in this case, the maximum acceleration of the block is a_0 , there are no relative displacements, and the maximum internal forces are evaluated with $a = a_0$.
2. The critical acceleration of the system a_c is smaller than the maximum acceleration at the base a_0 ; in this second case, the maximum acceleration of the block is a_c ; the

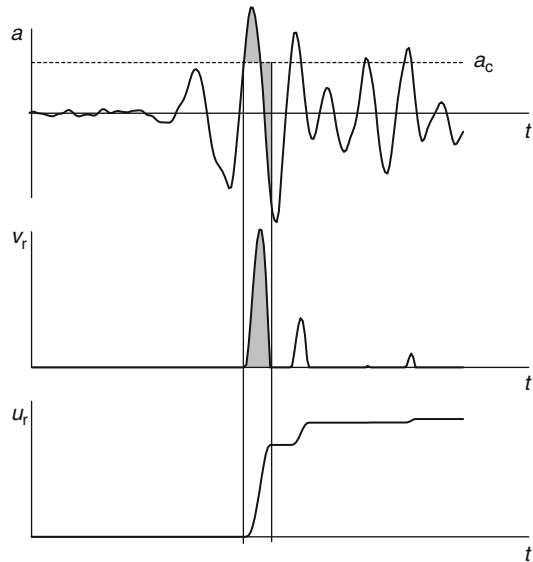
displacements increase as a_c decreases; the internal forces are evaluated with $a = a_c$ and increase with a_c .

This conceptual model can be used to illustrate a performance-based approach to the design of earth-retaining structures. A given limit state corresponds to a certain design seismic action that must be found accounting for the local site response. The required seismic performance of a retaining structure can be expressed by its permanent displacement and can be chosen by the designer for compatibility with the limit state under consideration. The design consists on one hand in endowing the system with a critical acceleration sufficiently large to produce displacements that do not exceed the specified limit and on the other hand in protecting the structural elements from a premature yielding by making sure that their structural capacity is sufficient to carry the maximum internal forces occurring during the development of the displacements.

Evaluation of Seismic Displacements

The conceptual model of Fig. 1 can be used directly to evaluate, for a given acceleration time history at the base, the seismic displacements of a retaining structure. This is accomplished as shown in Fig. 2: whenever the base acceleration $a(t)$ exceeds a_c , the time history of the relative acceleration $a(t) - a_c$ is integrated to provide the relative velocity v_r . The integration is halted when v_r equals zero and is restarted when $a(t)$ becomes again larger than a_c . It is recognized that a retaining wall can move only downhill; therefore, the positive and the negative portions of the acceleration time histories are considered one at a time. The displacements are obtained from the integration of the time history of the relative velocity.

The direct evaluation of the displacements depicted in Fig. 2 is very simple, but can be performed only if the seismic action is described as a set of accelerograms. This implies the evaluation of the outcrop seismic motion that must be inserted into a detailed site response calculation, which in turn requires an accurate geotechnical



Seismic Design of Earth-Retaining Structures, Fig. 2 Illustration of the Newmark (1965) integration scheme

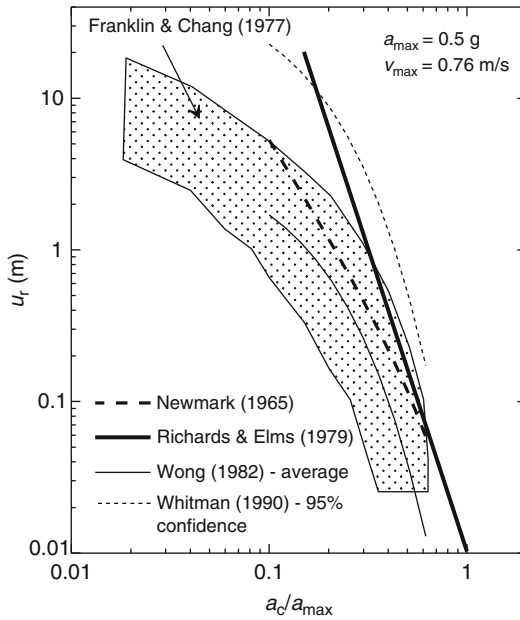
characterization of the soil deposits, starting from the bedrock.

For simple earth-retaining structures, the outcrop seismic action is more commonly expressed as an elastic response spectrum, and the local site response is considered through simple amplification coefficients, usually based on the classification of the subsoil according to the shear wave velocity of the topmost, say, 30 m. In this case, acceleration time histories are not available: a direct calculation of the displacements is not possible, and it is necessary to use a simplified approach.

The final displacement induced by the acceleration time history of Fig. 1 can be evaluated as

$$u_r = \frac{v_0^2}{2a_c} \left(1 - \frac{a_c}{a_0} \right) \quad (2)$$

where $v_0 = a_0 t_0$ is the maximum acceleration at the base. Newmark (1965) argued that a given acceleration time history can be approximated as $a_{\max}/a_c$ cycles equivalent to the simple impulse of Fig. 1, where $a_{\max}$ is the maximum recorded acceleration. This assumption results in the following expression for the final displacement:



Seismic Design of Earth-Retaining Structures,
Fig. 3 Seismic displacements plotted as a function of
 the critical to maximum acceleration ratio

$$u_r = \frac{v_{\max}^2}{2a_c} \left(1 - \frac{a_c}{a_{\max}}\right) \cdot \left(\frac{a_{\max}}{a_c}\right) \quad (3)$$

where $v_{\max}$ is the maximum recorded acceleration.

Franklin and Chang (1977) selected a number of accelerograms scaled to $a_{\max} = 0.5 \text{ g}$ and $v_{\max} = 0.76 \text{ m/s}$; they produced a parametric integration of these accelerograms for different values of the ratio $a_c/a_{\max}$. Richards and Elms (1979), based on the same acceleration time histories, produced the following relationship for the prediction of the seismic displacements:

$$u_r = 0.087 \frac{v_{\max}^2}{a_{\max}} \cdot \left(\frac{a_{\max}}{a_c}\right)^4 \quad (4)$$

The displacements predicted by the above relationships are plotted in Fig. 3 as a function of the ratio $a_c/a_{\max}$, together with the results by Franklin and Chang (1977) and with the curves proposed by Wong (1982) for the average values and by Whitman (1990) for the values associated to a 95 % confidence. In principle, the use of the

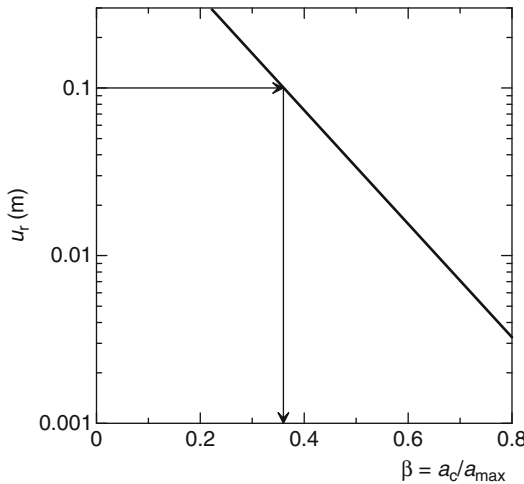
previous relationships or of the diagram of Fig. 3 is quite immediate; the critical acceleration expresses the resistance of the retaining structure to the seismic action and can be evaluated using the earth pressure formulas given in the following; $a_{\max}$ is the maximum expected acceleration for the limit state under consideration; the maximum velocity $v_{\max}$ can be estimated, for instance, from the constant-velocity portion of the elastic spectra provided by the construction codes (e.g., EN 1998-5 2003).

However, Fig. 3 shows that for a given ratio $a_c/a_{\max}$, the available relationships result in a very large uncertainty for the displacement. The reason for this substantial scatter is that the maximum recorded acceleration alone is not sufficient to represent the characteristics of an accelerogram. While in principle it is possible to produce more accurate expressions that relate the displacements to additional properties of the seismic motion (frequency content, Arias intensity, duration, etc.), often the designer that wishes to adopt a simplified procedure is provided only with an estimate of $a_{\max}$. An alternative way to obtain a more accurate prediction of the seismic displacement is to use the structure of the above predictive equations for the development of local relationships based on regional seismic records.

Equivalent Seismic Action on Retaining Structures

Consider a relationship between the displacement u_r and the ratio $a_c/a_{\max}$ that can be deemed appropriate to the local seismicity (Fig. 4). If u_0 is the tolerable displacement for the limit state under consideration and β is the corresponding value of the ratio $a_c/a_{\max}$, then the requisite that $u \leq u_r$ implies that the critical acceleration a_c must be larger than $\beta a_{\max}$ (with $\beta < 1$). This means that a check that the desired performance is met can also be done by controlling that under the conventional seismic forces associated to the acceleration $\beta a_{\max}$, the resistance of the system is not exceeded.

In other words, the performance of the retaining structure that in principle should be evaluated in terms of displacements can also be checked in terms of equivalent forces. This is



Seismic Design of Earth-Retaining Structures, Fig. 4 Relationship between seismic displacements and the ratio $a_c/a_{\max}$ developed by Rampello et al. (2010) for stiff soils based on the Italian seismicity

commonly expressed in terms of seismic coefficients: seismic forces deriving from a horizontal seismic coefficient,

$$k_h = \beta a_{\max}/g, \quad (5)$$

generally used in conjunction with a vertical seismic coefficient expressed as a fraction of k_h , must be such that the system capacity is not exceeded.

Note that no global or partial safety coefficient should be used in this check, as this is not an assessment of the safety with respect to a collapse mechanism, but rather an indirect evaluation of the seismic performance of the system.

Non-displacing Retaining Structures

In some circumstances, a retaining structure cannot undergo permanent deformation resulting from the activation of a plastic mechanism. This may be due either to the design requirements for the specific limit state implying a negligible damage to the structure or to the retaining structure being significantly constrained by additional structural elements, like props, rakers, etc.

In the absence of displacements, the coefficient β of Eq. 5 is equal to 1, and therefore the seismic earth pressure can be evaluated using the formulas given in the following section, using a horizontal seismic coefficient corresponding to the maximum acceleration. However, there are cases in which a different approach is necessary, and the soil should be regarded as an elastic material. These are dealt with at the end of the next section.

Evaluation of Earth Pressure

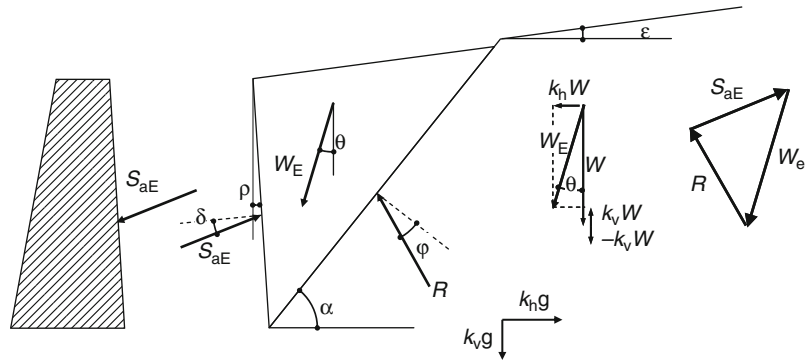
Earth Pressure in Active and Passive Limit States

When a plastic mechanism is mobilized, forces and/or stresses transmitted by the soil in contact with the retaining structure may be evaluated by assuming that the soil is in limit conditions, that is, its shear strength is completely mobilized. If the retaining structure displaces away from the soil, the soil is in an active limit condition; conversely, if the soil is contrasting the displacement of the retaining structure, it is in a passive limit condition.

For a purely frictional strength criterion, the resultant force applied by the soil to the retaining structure may be found using the Mononobe-Okabe approach (Okabe 1924; Mononobe and Matsuo 1929), which is an extension to seismic conditions of the Coulomb's limit equilibrium method. In active limit conditions, the total earth thrust S_{aE} is provided by a volume of soil that slides downward along a planar surface with an inclination α to the horizontal; this mechanism is shown in Fig. 5, where φ is the angle of shearing resistance of the soil and δ is the angle of friction at the soil-structure contact. The earth thrust is found by maximizing with respect to α the force that ensures the translational equilibrium for the soil wedge of Fig. 5, subjected to the bulk force W_E and to the resultant force R applied by the in situ soil onto the sliding plane. Since the shear strength of the soil is fully mobilized along the sliding surface, R is inclined by φ with respect to the normal to the

Seismic Design of Earth-Retaining Structures,

Fig. 5 Illustration of the Mononobe-Okabe approach for the evaluation of the active seismic force on a retaining structure



sliding plane and S_{aE} has an inclination δ with respect to the normal to the retaining structure. The seismic forces are included in the analysis by considering the horizontal and vertical seismic coefficients k_h and k_v in the evaluation of W_E .

The value of S_{aE} , maximized with respect to α , is given by the expression

$$S_{aE} = \frac{1}{2} \gamma \cdot (1 - k_v) \cdot H^2 \cdot K_{aE} \quad (6)$$

where γ is the soil unit weight, H is the height of the retaining structure, and K_{aE} is the coefficient of active thrust, evaluated with the following equation:

$$K_{aE} = \frac{\cos^2(\varphi - \beta - \theta)}{\cos \theta \cos^2 \beta \cos(\delta + \beta + \theta) \left[1 + \sqrt{\frac{\sin(\varphi + \delta) \sin(\varphi - \varepsilon - \theta)}{\cos(\delta + \beta + \theta) \cos(\varepsilon - \beta)}} \right]^2} \quad (7)$$

in which the angles β and ε have the meaning depicted in Fig. 5 while the angle θ defines the direction of W_E :

$$\theta = \arctan \frac{k_h}{1 - k_v} \quad (8)$$

The inclination α_{cr} of the sliding surface that maximizes S_{aE} can be evaluated with the following expression:

$$\alpha_{cr} = (\varphi - \theta) + \arctan \left[\frac{\{\tan a (\tan a + \cot b) [1 + \tan(\delta + \beta + \theta) \cot b]\}^{\frac{1}{2}} - \tan a}{1 + \tan(\delta + \beta + \theta) (\tan a + \cot b)} \right] \quad (9)$$

with

$$\begin{aligned} a &= \varphi - \varepsilon - \theta \\ b &= \varphi - \beta - \theta \end{aligned} \quad (10)$$

The active pressure S_{aE} given by Eq. 6 is a function of the horizontal acceleration, through

the dependency of K_{aE} on k_h (Eqs. 7 and 8); therefore, it can be regarded as the function $S_a(a_b)$ introduced in the previous section and used accordingly.

Equations 6, 7, and 8 imply quite obviously that S_{aE} increases with the horizontal seismic coefficient k_h . Conversely, it can be seen from

Eqs. 9 and 10 that the angle α_{cr} decreases with k_h , implying that the dimensions of the sliding wedge increase with the earthquake intensity. In other words, for larger accelerations, the volume of soil that interacts with the retaining structure is larger, and this finding has several implications, for instance, in the selection of the relevant strength parameters or for the design of a drainage system that needs to be effective in controlling the pore pressure within the entire volume of soil interacting with a retaining wall.

Since the Mononobe-Okabe method considers only translational equilibrium, it provides no direct information on the point of application of S_{aE} .

For passive limit conditions, the Mononobe-Okabe approach is not recommended, because

the assumption of a planar sliding surface is not conservative, especially when the soil-wall frictional coefficient δ approaches φ .

A more general approach to evaluate the seismic pressure in active and passive limit conditions is to use the lower-bound theorem of limit analysis. A convenient expression for vertical walls was derived by Lancellotta (2007) for passive limit conditions and extended by Rampello et al. (2011) to active limit conditions. The effective normal stress acting on the retaining structure in active or passive limit conditions $\sigma'_{a,p}$ is expressed as the product of a suitable thrust coefficient $K_{a,p}$ and the notional vertical effective stress σ'_v ; the thrust coefficient has the expression

$$K_{a,p} = \left[\frac{\cos \delta}{\cos(\varepsilon - \theta) \pm \sqrt{\sin^2 \varphi - \sin^2(\varepsilon - \theta)}} \cdot \left(\cos \delta \mp \sqrt{\sin^2 \varphi - \sin^2 \delta} \right) \right] \frac{\cos \varepsilon}{\cos \theta} (1 - k_v) \cdot e^{\mp 2\psi \tan \varphi} \quad (11)$$

with

$$\psi = 0.5 \left\{ \arcsin \left(\frac{\sin \delta}{\sin \varphi} \right) \mp \arcsin \left[\frac{\sin(\varepsilon - \theta)}{\sin \varphi} \right] \mp \delta + (\varepsilon - \theta) + 2\theta \right\} \quad (12)$$

In the above expressions, the upper and lower operators apply to active and passive conditions, respectively. Equations 11 and 12 imply that K_a increases and K_p decreases with θ , that is, with the seismic acceleration.

Providing normal stresses that are proportional to the vertical effective stresses, this lower-bound solution yields the entire distribution of soil-wall contact stresses that can be used to study the equilibrium of the retaining structure for the selected limit state.

Effect of Pore Water Pressure

The presence of pore water pressures produces a change in the seismic contact stresses. This effect

is better evaluated using the above lower-bound solution. The total contact stress $\sigma_{a,p}$ is equal to the sum of the effective contact stress $\sigma'_{a,p}$ and the pore pressure u :

$$\sigma_{a,p} = K_{a,p} \sigma'_v + u = K_{a,p} (\sigma_v - u) + u \quad (13)$$

The change in total contact stresses $\sigma_{a,p}$ is not caused only by the introduction of u in the above expression but also by the coefficients of earth pressure in seismic condition depending on the pore water pressure.

For most practical purposes, it can be assumed that during the seismic motion, the pore water vibrates in phase with the soil skeleton (Matsuzawa et al. 1984 suggest that this is the case for permeability coefficients smaller than 10^{-5} m/s). The evaluation of the seismic pressure should consider that the inertial forces in the soil are proportional to the total mass of the soil skeleton and the pore water while the shear strength is proportional to the effective stresses. Callisto and Aversa (2008) showed that this

effect can be taken into account by modifying the expression of θ in Eq. 8, which becomes

$$\theta' = \arctan\left(\frac{\sigma_v}{\sigma'_v} \frac{k_h}{1 - k_v}\right) \quad (14)$$

where σ_v is the total vertical stress. This expression leads to larger values of K_a and to smaller values of K_p , showing that pore water pressures have a detrimental effect on the coefficients of earth pressure. Expression (14) implies that K_a and K_p may vary with depth even in a homogeneous soil. For the particular, and quite uncommon, case of a hydrostatic pore water pressure distribution with the total hydraulic head at the ground surface, expression (14) becomes independent of depth:

$$\theta' = \arctan\left(\frac{\gamma}{\gamma'} \frac{k_h}{1 - k_v}\right) \quad (15)$$

where γ is the unit weight and γ' is the submerged weight of the soil. Use of expression (15) is explicitly suggested by Eurocode 8 part 5 (EN 1998-5 2003).

Effect of Cohesion

The relationships listed above were developed for a purely frictional material. For some of these solutions, the possibility exists to account rigorously for the effect of cohesion. However, small values of the cohesion c' can be incorporated in the analysis with sufficient accuracy using the Rankine expression:

$$\sigma'_{a,p} = \mp 2c' \sqrt{K_{a,p}} + K_{a,p} \sigma'_v \quad (16)$$

Drainage Conditions

For a saturated soil, it can be assumed that the seismic event occurs in undrained conditions, even for coarse-grained materials. The excess pore water pressures Δu generated during the earthquake can be easily incorporated into Eq. 13, to provide (for a purely friction material)

$$\sigma_{a,p} = K_{a,p} [\sigma_v - (u + \Delta u)] + (u + \Delta u) \quad (17)$$

If a wall retains an excavation in a fine-grained soil, in the long-term, the removal of soil produces a significant over-consolidation and the volumetric response of the soil becomes dilatant. In this case, the seismic event is likely to produce a decrease of the pore water pressures ($\Delta u < 0$), and this beneficial effect may be neglected. However, for relatively loose coarse-grained soils, the earthquake may produce a significant increase Δu of the pore water pressure. The evaluation of Δu for coarse-grained material is beyond the scope of this entry; the reader may refer to the sections devoted to liquefaction and to Seed and Booker (1977).

An excavation in fine-grained material can be taken to occur in approximately undrained conditions, generating negative excess pore water pressures. Traditionally, the static design of an excavation in these conditions is carried out expressing the shear strength of the soil in terms of total stresses, using the Tresca criterion and a limiting shear stress equal to the undrained shear strength S_u . Yet for the analysis of the seismic conditions, a total stress analysis is seldom appropriate, because the time interval corresponding to a significant dissipation of the excess pore water pressure may be considered to be much smaller than the return period of any severe seismic event. Therefore, the seismic design is controlled by the seismic events that, having a large return period (e.g., 475 years for a 10 % probability over a life span of 50 years), may be deemed to produce effects on a retaining structure that is already in drained conditions.

Evaluation of Earth Pressure for Non-displacing Retaining Structures

Although the case of zero displacement can be addressed simply by inserting $\beta = 1$ into Eq. 5 and therefore evaluating the earth pressure using the maximum predicted acceleration, there are cases in which active limit conditions may not develop at the rear of a retaining structure, because of the very small deformation allowed during the excavation or because the wall has been backfilled after its construction. In these situations, if the deformation of the retaining structure is impeded during the seismic event, a

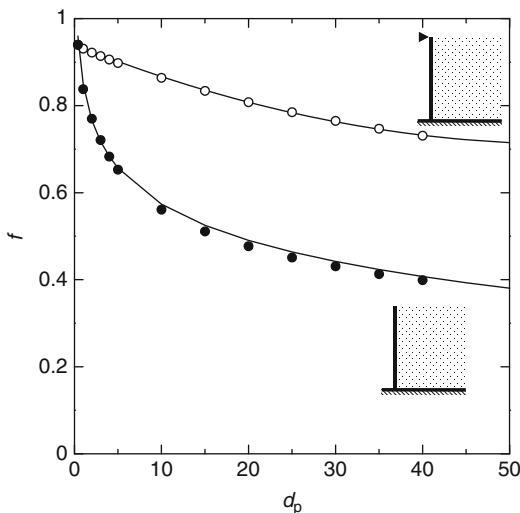
different approach is in order to evaluate the earth pressure, based on the hypothesis that the soil does not attain its strength and can be modeled as a linearly elastic material. The increase in the horizontal stress produced by the earthquake can then be expressed as

$$\Delta\sigma_{hE} = \gamma H \frac{a_{\max}}{g} \cdot f \quad (18)$$

where H is the height of the retaining structure. For an infinitely rigid wall, the coefficient f may be taken equal to 1 (Wood 1973). For a flexible wall, the coefficient f is a function of the soil-wall relative stiffness d_p , expressed as (Younan and Veletsos 2000)

$$d_p = \frac{G_s H^3}{(EI)_w} \quad (19)$$

In the above relationship, G_s is the shear modulus of the backfill, and the quantity $(EI)_w$ is the bending stiffness of the retaining structure (per out-of-plane length). Figure 6 shows the coefficient f plotted as a function of the relative stiffness d_p as obtained by Younan and Veletsos (2000) for two different restraining conditions. This graph,



Seismic Design of Earth-Retaining Structures,
Fig. 6 Coefficient f of Eq. 18 plotted as a function of the relative stiffness d_p

together with Eqs. 18 and 19, may be used to evaluate the design earth pressures.

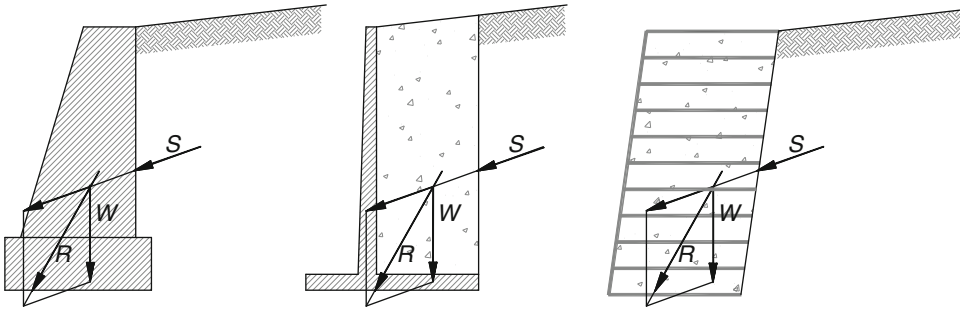
Gravity Retaining Walls

In static conditions, gravity retaining walls transfer the sub-horizontal earth thrust S to the foundation soil, thanks to their considerable self-weight. Figure 7 shows that the weight of the wall W combines with S to yield a resultant force R that (i) goes through the base of the wall and (ii) has a small inclination to the vertical. As shown in Fig. 7, different types of gravity retaining walls have a similar global behavior, the only difference being the relative amount of soil contributing to the total weight of the structure.

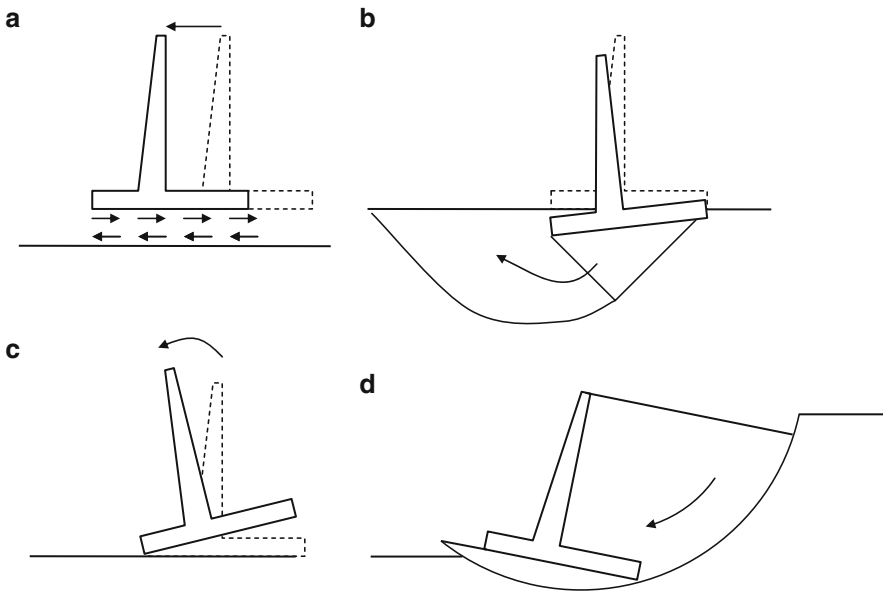
Since these structures are endowed with a significant mass, seismic accelerations produce large inertial forces that must be taken into account, considering both the horizontal and vertical acceleration components.

Construction codes provide the values of the seismic coefficient that should be used for the design, essentially in the form of the reduction coefficient β of Eq. 5, depending on the allowable displacement at the end of the earthquake. Using the appropriate seismic coefficient, the design of a gravity retaining wall is carried out studying global and local mechanisms. The global mechanisms are essentially the same for the different types of retaining structures depicted in Fig. 7: (a) sliding along the base, (b) bearing capacity of the foundation, (c) overturning, and (d) overall slip-circle mechanism (Fig. 8). Local mechanisms are relative to the attainment of the structural strength or to the mobilization of the reinforcement strength in a reinforced earth structure.

The details of each safety check are given by the construction codes and therefore are not dealt with here. However, a general appreciation of the seismic design of a retaining wall is provided by the example of Fig. 9, relative to a cantilever retaining wall with a height $H = 3.5$ m subjected to a seismic event characterized by a maximum acceleration $a_{\max} = 0.25$ g. Restricting the



Seismic Design of Earth-Retaining Structures, Fig. 7 Transfer of earth pressure to the foundation soils for gravity retaining walls

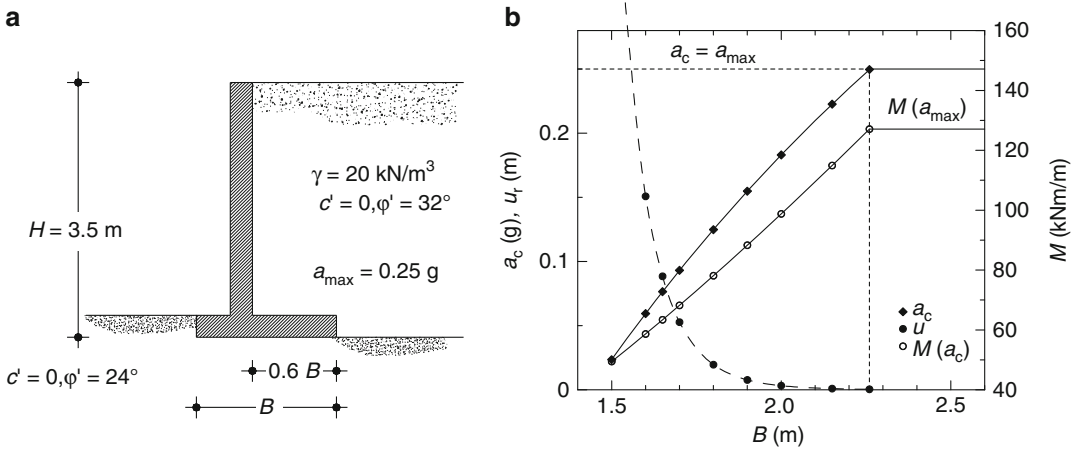


Seismic Design of Earth-Retaining Structures, Fig. 8 Global limit states for a gravity retaining wall

attention to the sliding mechanism only, Fig. 9 shows the values of the critical acceleration a_c plotted as a function of the width B of the wall base and of the corresponding displacements evaluated using the relationship of Fig. 4. It is evident that wider walls have a larger critical acceleration and therefore undergo smaller displacements. The figure also shows the maximum bending moment $M(a_c)$ in the wall stem, evaluated when the wall is critically accelerated. It is evident that since the bending moments increase with a_c , they must show an increasing trend with B . Hence, if a retaining wall is designed

with a large base to have a particularly good seismic performance, it must be endowed with a correspondingly large structural strength to resist the internal forces associated with its large critical acceleration. Following this line of thought, it is evident that a wall with a critical acceleration equal to $a_{\max}$ will suffer negligible displacement and will be subjected to internal forces that cannot be larger than those evaluated with an acceleration equal to $a_{\max}$.

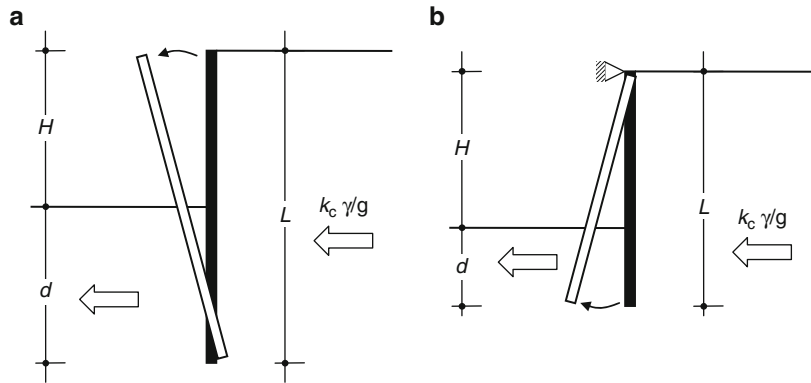
It should be noted that any “safe” provision, such as the use of strength or resistance factors, or an underestimation of the soil strength, leads to



Seismic Design of Earth-Retaining Structures, Fig. 9 (a) Example of a retaining wall; (b) critical acceleration, displacement, and maximum bending moment plotted as a function of base width

Seismic Design of Earth-Retaining Structures, Fig. 10

Reference schemes of a cantilevered (a) and a singly propped (b) embedded retaining wall, showing the direction of the inertial forces activating the plastic mechanisms



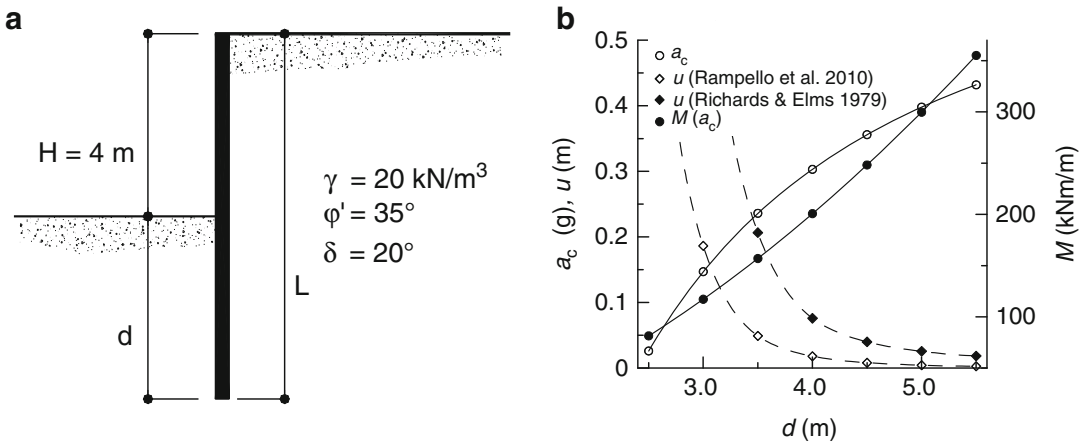
design acceleration smaller than the critical one. Figure 9 shows clearly that a larger critical acceleration implies on one hand smaller displacements and on the other hand larger internal forces. For the structural design of the wall, it is then recommendable to evaluate always the internal forces corresponding to the actual critical acceleration of the wall.

Embedded Retaining Walls

The seismic behavior of embedded retaining walls is different from that of gravity retaining walls, in that the earth thrust is resisted by the passive strength of the soil in front of the wall and by additional constraints (anchors, props); therefore, it is not necessary that the structure

possesses a large mass and the corresponding inertial forces are small; the effect of vertical acceleration can usually be neglected.

On the other hand, cantilevered or singly propped walls (Fig. 10) bear some similarities with gravity retaining walls, because if the strength of the soil interacting with the wall is attained during the seismic event, a plastic mechanism may develop, and the earth pressure relationships may be used in a limit equilibrium computation to evaluate the critical acceleration that activates the mechanisms. Figure 11 shows an example in which a cantilevered retaining wall is subjected to a maximum acceleration $a_{\max} = 0.5$ g. For this wall, the critical acceleration can be obtained studying the rotational mechanism of Fig. 10 using, for instance, the Lancellotta (2007) closed form expressions for



Seismic Design of Earth-Retaining Structures, Fig. 11 (a) Example of an embedded retaining wall; (b) critical acceleration, displacement, and maximum bending moment plotted as a function of embedded length

the active and passive pressures. The results are very similar to those developed for the example retaining wall of Fig. 9, showing that increasing the embedded length of the wall the critical acceleration increases, and therefore the permanent displacement decreases but the internal forces in the wall become larger. For these wall types, it is recommended that the internal forces be evaluated considering the critical value of the acceleration, that is, considering the activation of the plastic mechanisms depicted in Fig. 10.

When the structure has several additional constraints, the attainment of the soil strength is not sufficient for the activation of a plastic mechanism: if the design requires that the capacity of the wall itself and of the constraints cannot be reached, then these should be classified as a non-displacing wall and the earth pressure should be evaluated accordingly, as explained in a preceding section. In this perspective, no consideration of plastic mechanisms is needed, and the objective of the design is essentially the evaluation of the maximum internal forces in the structural members that should derive necessarily from a study of the soil-structure interaction.

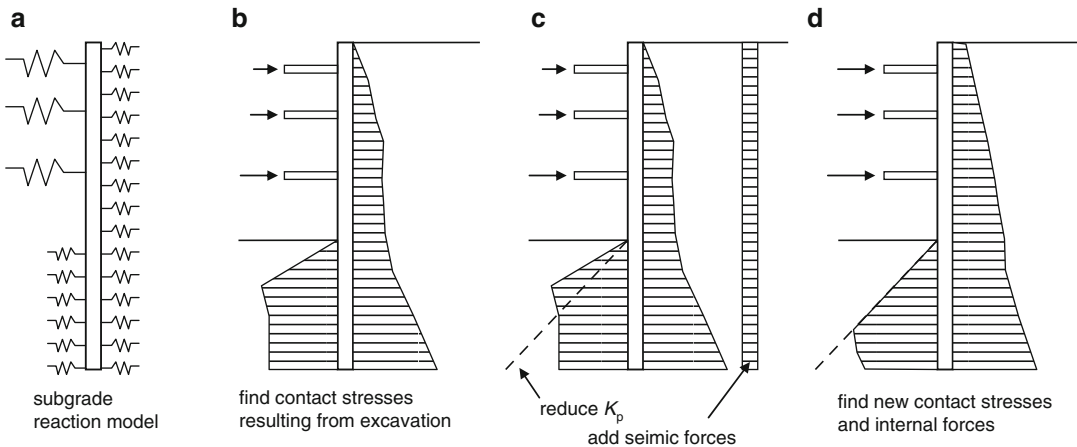
However, the analysis of the soil-structure interaction under seismic conditions is not straightforward and can be carried out at different levels of complexity. A very basic representation of the soil-structure interaction is represented by the subgrade reaction methods that constitute a

standard method for the static design of multi-constrained retaining wall (Fig. 12a). While there is not a general consensus on the use of this method in seismic conditions, a simple strategy for taking into account the inertial force consists of the following steps:

- (i) Carry out a static analysis to model the construction sequence (Fig. 12b).
- (ii) Evaluate the additional forces due to the earthquake and the reduction of the coefficient of passive resistance, as a function of the maximum acceleration estimated for the site.
- (iii) Apply these forces as external loads on the subgrade reaction model, and reduce the coefficient of passive resistance (Fig. 12c).
- (iv) Run the program for equilibrium and compatibility (Fig. 12d).

Summary

The current trend in the design of earthquake-resistant structures is based on the evaluation of the seismic performance of the system. For a retaining structure, the seismic performance is expressed by its cumulative displacement at the end of the earthquake. Although it would be desirable to evaluate directly the performance of the structure, it has been shown that a design



Seismic Design of Earth-Retaining Structures, Fig. 12 Use of a subgrade reaction model for the seismic analysis of a multi-propped retaining wall

requirement limiting the seismic displacements can be transformed into an equivalent requirement on the seismic action that needs to be considered in a conventional design process, which is commonly based on forces. Specifically, the conventional seismic coefficient used in a pseudo-static calculation can be made to decrease as the allowable displacement decreases. Specific cases of non-displacing walls have been evidenced that may call for a different approach, in which the internal forces are evaluated assuming that the soil does not mobilize its strength during the earthquake.

References

- Callisto L, Aversa S (2008) Dimensionamento di opere di sostegno soggette ad azioni sismiche. In: *Opere geotecniche in condizioni sismiche*, MIR 2008. Pàtron, Bologna, pp 273–308
- EN 1998-5 (2003) Eurocode 8: design of structures for earthquake resistance – part 5: foundations, retaining structures and geotechnical aspects. CEN European Committee for Standardization, Brussels
- Franklin AG, Chang PK (1977) Earthquake resistance of earth and rockfill dams. Report 5, Permanent displacement of earth embankments by Newmark sliding block analysis. MP S-71-17, Soils and Pavements Laboratory, US Army Waterways Experiment Station, Vicksburg
- Lancellotta R (2007) Lower-bound approach for seismic passive earth resistance. *Géotechnique* 57(3):319–321
- Matsuzawa H, Ishibashi I, Kawamura M (1984) Dynamic soil and water pressures of submerged soils. *J Geotech Eng* 111(10):1161–1176
- Mononobe N, Matsuo H (1929) On the determination of earth pressure during earthquakes. In: *Proceedings of the world engineering congress*, Tokio, vol 9, pp 177–185
- Newmark NM (1965) Effects of earthquakes on dams and embankments. Fifth Rankine lecture. *Géotechnique* 15(2):139–193
- Okabe S (1924) General theory on earth pressure and seismic stability of retaining wall and dam. *J Jpn Civ Eng Soc* 10(5):1277–1323
- Rampello S, Callisto L, Fargnoli P (2010) Evaluation of slope performance under earthquake loading conditions. *Ital Geotech J* 44(4):29–41
- Rampello S, Callisto L, Masini L (2011) Spinta delle terre sulle strutture di sostegno. *Atti delle XXIII Conferenze Geotecniche di Torino*
- Richards R, Elms DG (1979) Seismic behavior of gravity retaining walls. *J Geotech Eng Div ASCE* 105(GT4):449–464
- Seed HB, Booker JR (1977) Stabilization of potentially liquefiable sand deposits using gravel drains. *J Geotech Eng* 103(7):757–768
- Whitman RV (1990) Seismic design and behavior of gravity retaining walls. In: *Proceedings of the design and performance of earth retaining structures*, Ithaca, ASCE GSP 25, pp 817–842
- Wong CP (1982) Seismic analysis and improved seismic design procedure for gravity retaining walls. MSc thesis, Department of Civil Engineering, M.I.T., Cambridge, MA
- Wood JH (1973) Earthquake-induced soil pressures on structures. PhD thesis, California Institute of Technology, Pasadena
- Younan AH, Veletsos AS (2000) Dynamic response of flexible retaining walls. *Earthq Eng Struct Dyn* 29:1815–1844

Seismic Design of Pipelines

Hany El Naggar

Department of Civil and Resource Engineering,
Dalhousie University, Halifax, NS, Canada

Synonyms

Earthquakes; Lifelines; Liquefied soil; PGD;
Pipeline breakings; Seismic-induced settlement

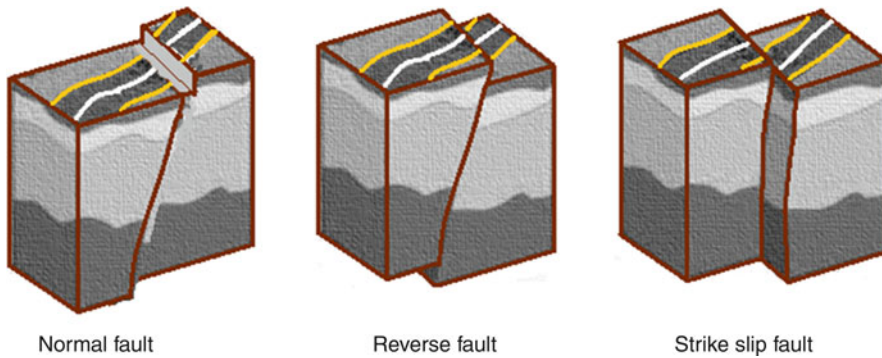
Introduction

Pipelines are a cost-effective means for the transportation of water supply and sewers, in addition to commercial fluids such as oil and gas. The designation of pipeline systems as “lifelines” indicates that they should be designed to function and operate at all times for public safety and well-being and also for economic reasons. Engineers must consider the different types of loads that are going to be imposed on the pipeline, the environment that the pipeline will travel in, and the type of material that the pipeline is going to convey. Consequently, pipelines must be designed for different loads such as stresses due to pressure generated by the flow (internal pressure), external pressure generated by the weight of earth and by live loads for buried pipelines, or external pressure generated by fluid if the pipe is submerged underwater. Also, seismic loads (earthquake loads) in medium and high seismicity zones are detrimental and should be considered. In general, pipeline design includes several general steps: (1) load determination, (2) critical performance evaluation (determining the critical stress and/or deformation), (3) comparison of the critical performance with the limiting criteria established by adopted codes and standards, and finally (4) selection of the pipe.

Over the past century, several catastrophic earthquakes caused severe damage to buried pipelines. In the 1906 San Francisco earthquake, one of the main reasons that caused the extensive damage was due to the failure of several water

pipelines which hindered firefighting efforts. Several years later, in the 1971 San Fernando earthquake, about 80 % of the reported destruction were in welded steel buried pipelines. Following the 1989 Loma Prieta earthquake, the East Bay Municipal Utilities District identified over 120 water pipeline breaks, and the San Jose Water Company reported another 155 pipe breaks. However, most of the serious damage reported by Pacific Gas and Electric Company (PG&E) occurred to natural gas mains and service lines. During the Northridge earthquake (1994) in California, some natural gas pipelines were severely damaged and the leak of containment fluid caused a large explosion in the Northridge town resulted in several deaths (Lau et al. 1995; O'Rourke and Palmer 1996). In the 1995 Hyogo-Ken Nanbu earthquake (in Japan), the natural gas leakage from buried pipelines resulted in numerous number of fires (531 cases reported) which started primarily due to gas release and electricity sparks affecting areas of over one square kilometer totally burnt (Scawthorn and Yanev 1995). More recently, the 1999 Chi-Chi earthquake (in Taiwan) also caused severe damage to natural gas distribution systems. More than 100,000 customers were affected after the earthquake, and the estimated economic loss of five major natural gas companies was approximately US\$ 25 million (Chen et al. 2000). In the 1999 Izmit earthquake (in Turkey), Tupras refinery suffered serious fire damage as the pipeline conveying water from neighbor lake was damaged, so the refinery was dependent on internal water reservoirs which were insufficient. In 2001 two earthquakes occurred (1 month apart) in San Salvador; several pipelines in rural areas experienced severe destruction caused by huge landslides. Till now, in several very recent earthquakes occurred in Chile (2010), New Zealand (Darfield, 2010), Japan (Tohoku, 2011), and Italy (earthquake of Emilia, 2012), serious damages to pipelines have been witnessed.

Accordingly, pipelines should be designed to function and operate during and following design earthquakes for life safety and economic reasons. As mentioned above several earthquakes in the



Seismic Design of Pipelines, Fig. 1 Main types of faults

last few decades resulted in too many pipeline breakings, and that extensive pipe breakage has the potential to lead to great economic harm to our urban communities.

The main focus of this chapter will be centered on the seismic analysis and design of pipelines.

Permanent Ground Deformation and Seismic Wave Propagation Hazards

In seismic events, buried pipelines can be damaged mainly by either the permanent ground deformations (PGD) or by the transient seismic wave propagation. PGD movements include faulting, landslides, lateral spreading due to liquefaction, and seismic settlement. Even though PGD risks are usually restricted to small regions within the pipeline, the chances of them causing severe damage is substantial since they impose large deformations. On the other hand, the seismic wave propagation risks typically affect the whole pipeline, but with lower damage rates (as the total deformations are general less and not permanent). For example, only 5 % of the affected area experienced lateral spreading during the 1906 San Francisco earthquake; approximately 52 % of all pipeline breaks occurred within one city block of the lateral spreading (O'Rourke et al. 1985).

Permanent Ground Deformation (PGD)

This section describes in details the four different forms of permanent ground deformation and

presents methods to calculate the amount of PGD as well as the extent of the PGD zone. Equations to quantify the amount of PGD are provided. Also, useful observations and notes to determine the extent of the PGD zone are presented and discussed.

Fault

Stresses in the earth's crust push the two sides of the fault. Eventually enough stress builds up and the rocks slip suddenly releasing energy in waves that travel through the rock to cause earthquake. Accordingly, earthquakes occur on faults. A fault is a thin zone of crushed rock separating blocks of the earth's crust. When an earthquake occurs on one of these faults, the rock on one side of the fault slips with respect to the other. The fault surface can be vertical, horizontal, or oblique to the surface of the earth. If the earthquake magnitude is large enough, the offset along the fault will propagate all the way to the earth's surface causing surface rupture (fault offset). Figure 1 below shows the main types of faults. In normal and reverse faults, the major ground displacement is vertical with a minor horizontal displacement. These ground displacements pose axial tension/compression and bending stresses in the pipeline depending on the direction of movement. On the other hand, in strike slip fault the main deformation (the offset) occurs in the horizontal plane, which poses axial tension/compression and bending stresses in the pipeline depending on the intersection angle of the pipeline and the fault.

The following empirical equations by Wells and Coppersmith (1994) are the most recognized equations to estimate the average fault displacement relative to the size of the considered earthquake:

$$\text{For the normal fault} \rightarrow \log \delta_f = -4.45 + 0.63 M \quad (1)$$

$$\text{For the reverse fault} \rightarrow \log \delta_f = -0.74 + 0.08 M \quad (2)$$

$$\text{For the strike slip fault} \rightarrow \log \delta_f = -6.32 + 0.90 M \quad (3)$$

where δ_f is the average fault displacement in meters and M is the moment magnitude of the earthquake.

Landslides

Seismically induced landslide involves a wide range of downslope mass ground movements, which can occur in offshore, coastal, and onshore environments.

Offshore Landslides In offshore landslides, the response of the pipeline is mainly governed by the orientation of the pipeline to the direction of ground movement. The imposed displacements for undersea slides are so large that pipeline response is likely controlled by the maximum force available at the soil-pipe interface.

Onshore Landslides There are several types of onshore landslides based on soil movements, geometry of the slide, and the types of material involved. The main types are rock falls, rock topples, slides, and lateral spreads. Rock fall and rock topple can cause direct damage to above-ground pipelines by the impact of falling rock fragments. In an earth slide the earth moves relatively as a block; they typically develop along natural slopes and embankments.

Based on Newmark's Block model for landslides (Newmark 1965), the critical acceleration at which the slide will be triggered α_c can then be determined from

$$a_c = g (\text{FOS} - 1) \sin \alpha \quad (4)$$

where g is the acceleration due to gravity, FOS is the factor of safety, and α is the angle of the slope.

Then the displacement of the block can be calculated by double integrating the ground acceleration. Jibson and Keefer (1993) proposed the following equation to estimate the Newmark displacement, δ_s , in centimeters as

$$\text{Log } \delta_s = 1.460 \log I_a - 6.642 a_c + 1.546 \quad (5)$$

where α_c is the critical acceleration in g s and I_a is the Arias Intensity in m/s .

Arias Intensity can be calculated using the following relationship developed by Wilson and Keefer (1983):

$$\text{Log } I_a = M - 2 \log R - 4.1 \quad (6)$$

where M is the earthquake magnitude and R is source distance in kilometers.

Lateral Spreading Due to Liquefaction

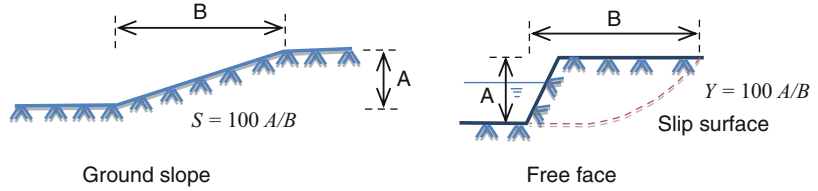
Seismic shaking may trigger the liquefaction of saturated loose cohesionless soils. The liquefaction process increases the pore water pressure in the ground to a level where the effective stress approaches zero at which point the soil loses entirely its shear strength, which in turn results in the lateral movement of the soil (lateral spreading). In past major earthquakes, large number of pipelines suffered massive damage caused by liquefaction-induced ground failures (Zhang et al. 2004).

Youd et al. (2002) proposed two empirical equations to approximately estimate the PGD due to liquefaction:

For lateral spreads down gentle ground slopes (GS),

$$\begin{aligned} \text{Log } \delta_L = & -16.213 + 1.532 M - 1.406 \log R^* \\ & - 0.012 R + 0.338 \log S + 0.54 \log T_{15} \\ & + 3.413 \log (100 - F_{15}) \\ & - 0.795 \log (D50_{15} - 0.1) \end{aligned} \quad (7)$$

Seismic Design of Pipelines, Fig. 2 Sketch showing the ground slope and the free face lateral spreads



For lateral spreads at a free face (FF),

$$\begin{aligned} \log \delta_L = & -16.710 + 1.532M - 1.406 \log R^* \\ & - 0.012R + 0.592 \log Y + 0.54 \log T_{15} \\ & + 3.413 \log(100 - F_{15}) \\ & - 0.795 \log(D50_{15} - 0.1) \end{aligned} \quad (8)$$

where δ_L is the PGD in meters, M is the earthquake magnitude, S is the ground slope %, γ is the free face ratio (in % see Fig. 2), T_{15} is the thickness in meters of the saturated cohesionless soil layer with a corrected standard penetration test (SPT value of less than 15), F_{15} is the percentage of average fines contents in T_{15} (in %), $D50_{15}$ is the mean grain size in mm in T_{15} , and R^* is an adjusted distance parameter in kilometers given by

$$R^* = R + 10 e^{(0.89M - 5.64)} \quad (9)$$

where R is the horizontal distance in kilometers from the site of interest to the nearest bound of the seismic energy source (do not use less than 0.5 km).

Seismic Settlement

Seismic-induced settlement may be caused by densification of cohesionless soils, consolidation of cohesive soils, or consolidation of liquefied soil. In this section we will only discuss liquefaction-induced ground settlement as it can cause larger settlement and hence higher potential for damage to buried pipelines.

Takada and Tanabe (1988) proposed the two following empirical equations to calculate liquefaction-induced settlement at embankments and leveled sites:

For embankments

$$\delta_{GS} = 0.11 H_1 H_2 \frac{a_{\max}}{N} + 20 \quad (10)$$

For leveled sites

$$\delta_{GS} = 0.30 H_1 \frac{a_{\max}}{N} + 2 \quad (11)$$

where δ_{GS} is the liquefaction-induced settlement in centimeters, H_1 is the thickness of saturated cohesionless soil layer in meters, H_2 is the height of embankment in meters, N is the SPT N -value in the cohesionless layer, and $a_{\max}$ is the ground acceleration in gals.

Seismic Wave Propagation

For the seismic analysis and design of buried pipelines, the effect of the seismic wave propagation on the pipeline is usually characterized by the induced ground strain and curvature. Newmark (1967) developed a straightforward method to estimate the ground strain. The general form of a traveling wave in Newmark's method is given by

$$U = f \left(\frac{t}{T} + \frac{x}{\lambda} \right) \quad (12)$$

where U is the function of the separation distance between the two points, x , and the speed of the seismic wave; T is the period of the repeating motion; and λ is the wavelength.

For particle motion parallel to the direction of propagation (R-waves), ϵ_g , the ground strain along the direction of propagation can be calculated as

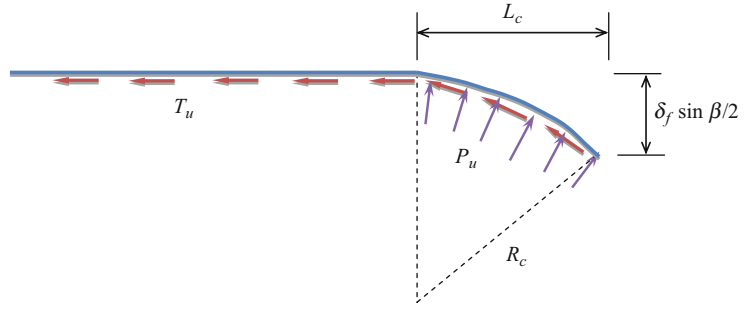
$$\epsilon_g = \frac{V_{\max}}{C_R} \quad (13)$$

where $V_{\max}$ is the maximum horizontal ground velocity in the direction of wave propagation and C_R is the propagation velocity of the R-wave.

For particle motion perpendicular to the direction of propagation (S-waves), C_g , the curvature can be calculated as

Seismic Design of Pipelines,

Fig. 3 Kennedy et al. (1977) model for one side of the fault



$$C_g = \frac{A_{\max}}{C_s^2} \quad (14)$$

where $A_{\max}$ is the maximum ground acceleration perpendicular to the direction of wave propagation and C_s is the propagation velocity of the S-wave.

Wave propagation with particle motion parallel to the pipeline (R-wave) would induce large axial strain in the pipeline. On the other hand, S-waves propagating parallel to the pipeline with particle motion perpendicular to the pipeline direction would induce only bending strains.

For S-waves traveling in a horizontal plane and at an angle with respect to the pipeline, the S-wave particle motion (perpendicular to its direction of propagation) would have one component parallel to the pipeline inducing axial strain, ε_g , and another component perpendicular to the pipeline inducing bending strain, ε_b , which can be calculated as

$$\varepsilon_g = \frac{V_{\max}}{2 C_s} \quad (15)$$

and

$$\varepsilon_b = \frac{\pi D V_{\max}}{\lambda C_s} \quad (16)$$

where ε_b is the upper bound bending strain (maximum), $V_{\max}$ is the peak ground velocity, C_s is the apparent propagation velocity of the S-wave, λ is the wavelength, and D is the pipe diameter.

The bending strains induced in a pipeline due to traveling waves (S-waves, L-waves, or the vertical component of R-waves) are generally

small compared to the axial strains induced in the ground by the traveling waves.

Pipelines Response to Faulting

This section presents the response of continuous pipelines subject to fault offsets. PGD due to faulting can be resolved into two components: longitudinal PGD (parallel to the pipeline) and transverse PGD (perpendicular to the pipeline) axis. In the case of “normal fault” type (see Fig. 1), the pipeline will be subjected to bending axial tensile force, caused by the transverse and longitudinal components, respectively. In this case, tensile rupture would be the most probable failure mechanism. In the “reverse fault”-type case, the pipeline will be subjected to bending axial compressive force, caused again by the transverse and longitudinal components, respectively. In this case, buckling would be the most likely failure mechanism. In the last case of “strike slip fault,” the pipeline can be subjected to either tension or compression depending on the intersection angle between the pipeline and the fault and pipe and the relative movement at the fault.

Kennedy et al. (1977) proposed a simplified method to analyze the tensile and bending behavior of pipelines due to fault movements. Figure 3 shows the Kennedy et al. model for one side of the fault.

According to Kennedy et al. (1977), the total strain in the pipe (bending + tensile) is given by

$$\varepsilon = \varepsilon_a + \varepsilon_b = \frac{\Delta L}{L} + \frac{D}{2R_c} \quad (17)$$

where ΔL is the total elongation of the pipeline, L is the total length of the pipeline, D is the pipe diameter, and R_c is the radius of curvature of the curved portion. R_c can be estimated by

$$R_c = \frac{\sigma \pi D t}{P_u} \quad (18)$$

where σ is the axial stress at fault crossing and P_u is the peak lateral pipeline-soil interaction force per unit length which can be calculated using the 1984 ASCE guideline using the following relations for sand and clay, respectively:

For sand

$$P_u = \gamma H N_{qh} D \quad (19)$$

For clay

$$P_u = c_u N_{ch} D \quad (20)$$

where γ is the unit weight of the soil, H is the embedment depth of the pipeline, and N_{qh} and N_{ch} are the horizontal bearing capacity factors for sand and clay, respectively. Figure 4 presents

horizontal bearing capacity factors for sand and clay after Hansen (1961).

The total elongation of the pipeline, ΔL , can be estimated using the following equation:

$$\Delta L = \delta_f \cos \beta + \frac{(\delta_f \sin \beta)^2}{3L_c} \quad (21)$$

where δ_f is the average fault displacement, β is the fault angle, and L_c is the horizontal projection length of the laterally deformed pipeline (see Fig. 3).

L_c can be approximately calculated using the following simplified equation:

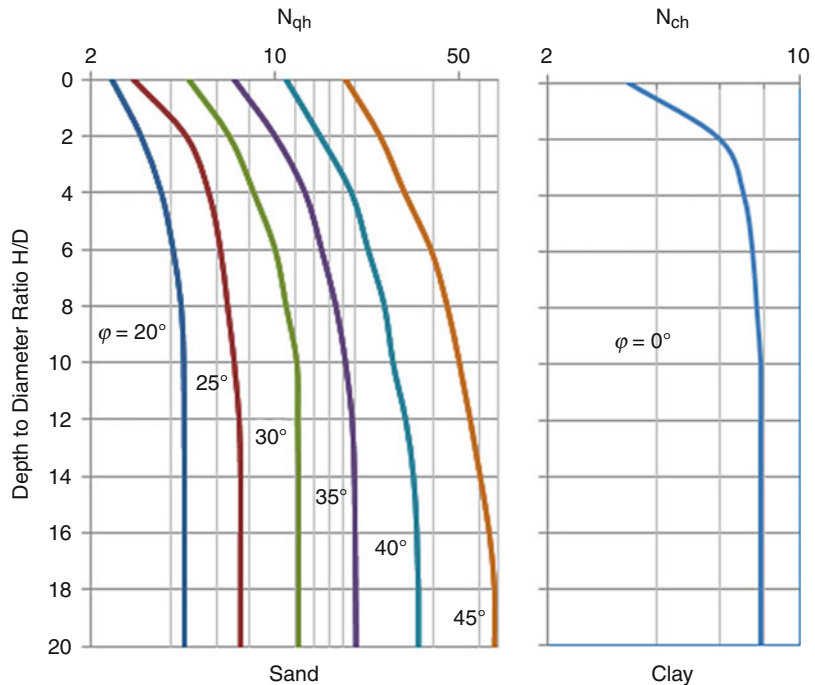
$$L_c = \sqrt{R_c \delta_f \sin \beta} \quad (22)$$

Pipelines Response to Longitudinal PGD

This section presents the response of continuous pipelines subject to longitudinal permanent ground deformations, PGD, where the soil movement is in the same direction as the pipeline.

Seismic Design of Pipelines,

Fig. 4 Horizontal bearing capacity factors for sand and clay (Reproduced after Hansen 1961)



O'Rourke et al. (1995) proposed an inelastic pipeline model to analyze the response of pipelines to longitudinal PGD. Figure 5 shows the considered model. To develop this model O'Rourke et al. utilized an idealized block pattern in which a mass of soil having length L moves down a slight incline. The soil displacement on either side of the PGD zone is zero, while the soil displacement within the zone is a constant value δ .

In O'Rourke et al. model, a block of soil between Points B and D moves to the right pulling the pipe laterally with it with the soil forces acting on the pipeline within the PGD zone to the right. On the other hand, the soil between Points A and B and the soil between Points D and E resist the pipeline movement and these soil restraint forces are directed to the left. The combined pipeline-soil interaction forces result in a region of pipe axial tension between points A and B and a region of axial compression between points D and E. The conditions outlined in Fig. 5 correspond to the case where the PGD, δ , is comparatively large and the length of the PGD zone, L , is comparatively short. In that case, the maximum pipe displacement is less than the ground displacement and the pipe strain is controlled by L .

Figure 6 presents the other possibility of the O'Rourke et al. model where the length of the PGD zone is relatively large while the amount of PGD is comparatively short. Also in this case, there is still axial pipe tension between points A and B and axial compression between points D and E; however, the zone is long enough that the pipe displacement matches that of the ground between Points C and D where the axial force and strain in the pipe are zero.

As it can be seen from Figs. 5 and 6, the axial force in the pipeline in the segment AB is linearly proportional to the distance from Point A. Accordingly, the pipelines strain and displacement can be evaluated using the following Ramberg-Osgood model relations:

$$\varepsilon(x) = \frac{\beta_p x}{E} \left[1 + \frac{n}{1+r} \left(\frac{\beta_p x}{\sigma_y} \right)^r \right] \quad (22)$$

$$\delta(x) = \frac{\beta_p x^2}{E} \left[1 + \frac{2}{2+r} \frac{n}{1+r} \left(\frac{\beta_p x}{\sigma_y} \right)^r \right] \quad (23)$$

where n and r are Ramberg-Osgood parameters (given in Table 1 below), E is the modulus of elasticity of steel, σ_y is the effective yield stress, and β_p is the pipe burial parameter, defined as the friction force per unit length t_u divided by the pipe cross-sectional area A .

The pipe burial parameter β_p can be obtained from the following:

For sand

$$\beta_p = \frac{\tan \varphi \gamma H}{t} \quad (24)$$

For clay

$$\beta_p = \frac{\alpha c_u}{t} \quad (25)$$

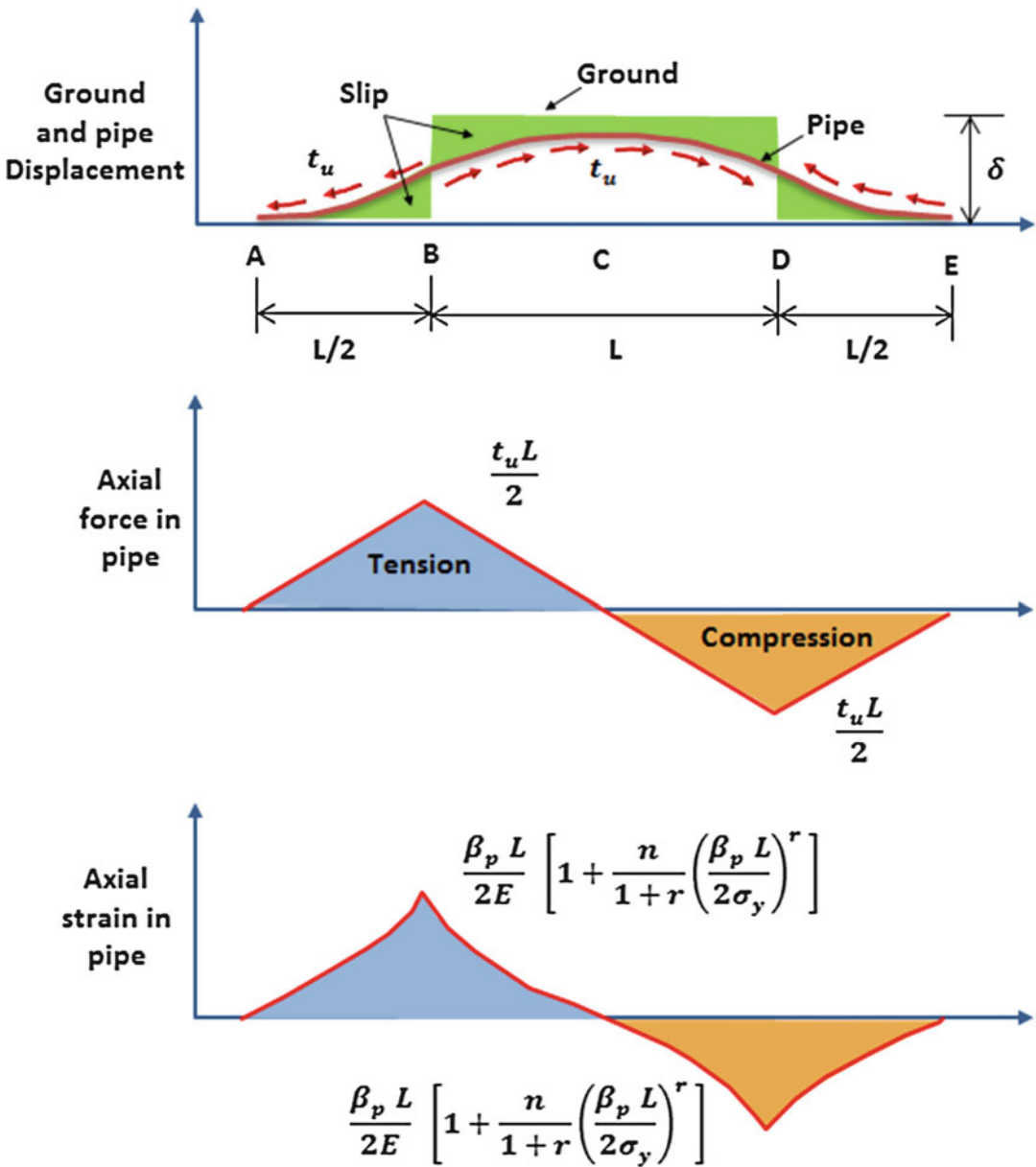
where φ is the angle of shear resistance, γ is the effective unit weight of the soil, H is the embedment depth of the pipeline, α is the adhesion factor for clay, c_u is the undrained cohesion of the clay (see Fig. 7 below), and t is the pipe wall thickness.

Wrinkling of the Pipe Wall in Compression

Substituting a critical local buckling strain into Eq. 22, one can obtain the critical length of PGD zone L_{cr} . This can then be used to calculate the critical ground movement δ_{cr} from Eq. 23. The critical strain in compression may be taken as $0.175 t/R$.

Pipeline Response to Transverse PGD

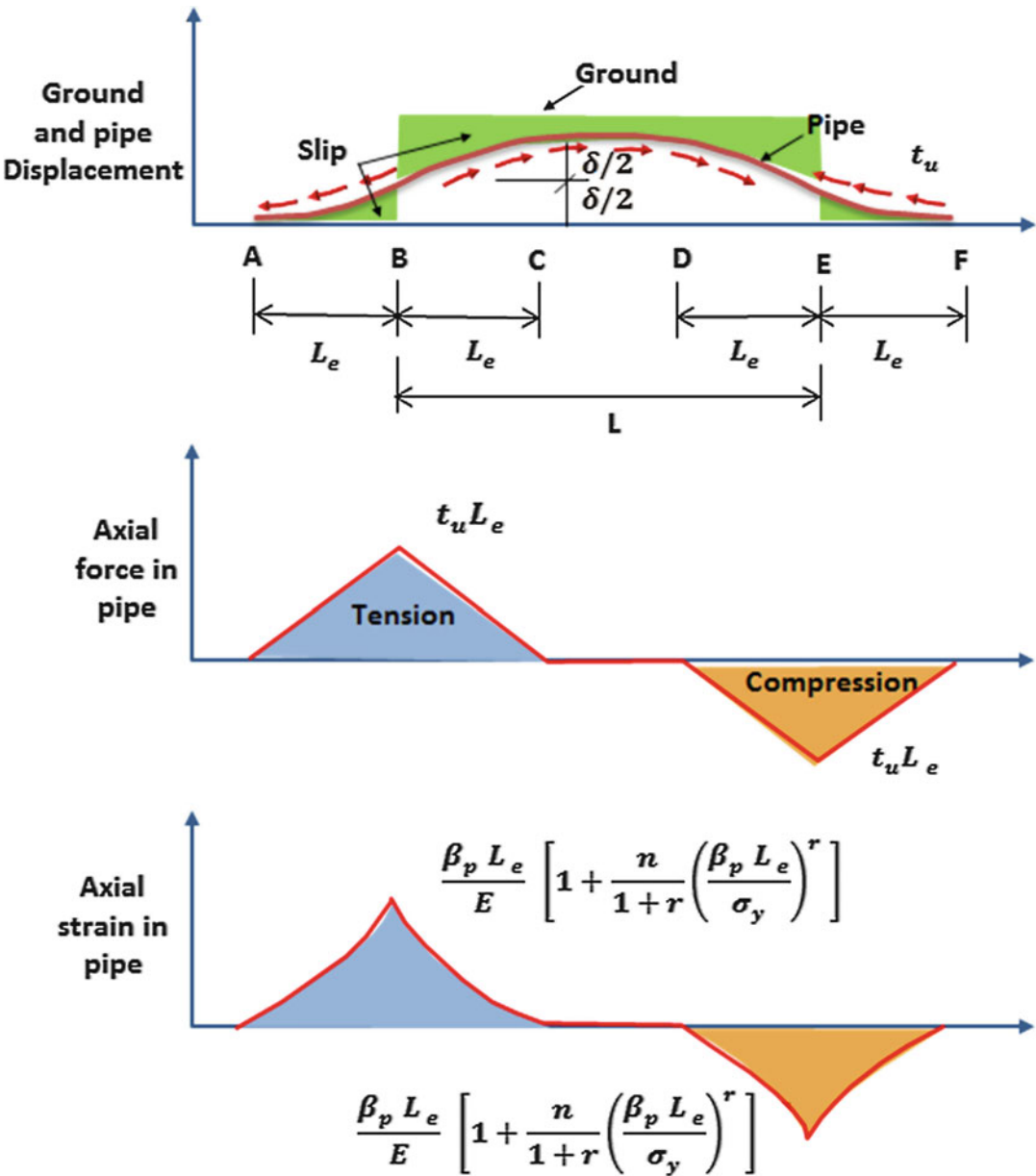
This section presents the response of continuous pipelines subject to transverse permanent ground deformations, PGD, where the soil movement is perpendicular to the pipeline. When subjected to transverse PGD, the pipeline will stretch and bend as it attempts to conform to the transverse ground movement profile. In this case, the failure



Seismic Design of Pipelines, Fig. 5 O'Rourke et al. model to analyze the response of pipelines to longitudinal PGD

mechanism of the pipeline will be governed by the relative magnitude of the axial tensile strain and the bending strain. If the tensile strain is relatively small, the pipe wall may buckle in compression due to excessive bending. Instead, if tensile strain is relatively large, the pipe may rupture in tension due to the combined effects of the tensile and bending stresses.

In general, the response to transverse PGD is a function of the magnitude of PGD, the width of the PGD zone, and the pattern of ground deformation. Two types of transverse ground deformation patterns are discussed here: the spatially distributed transverse PGD pattern and the abrupt transverse PGD pattern. Figure 8 shows sketch of the considered patterns.



Seismic Design of Pipelines, Fig. 6 O'Rourke et al. model to analyze the response of pipelines to longitudinal PGD (PGD zone is relatively large, while the amount of PGD is comparatively short)

Spatially Distributed Transverse PGD

O'Rourke (1989) proposed a simple model to analyze the response of pipelines to spatially distributed transverse PGD. In this model O'Rourke considered two types of response (wide and narrow width PGD zones) as shown in Fig. 9 below. In the wide width PGD zone, the

pipeline is relatively flexible and its lateral displacement is assumed to closely conform to the soil outline. Accordingly, the pipeline strain is expected to be mainly due to the ground curvature (i.e., displacement controlled). On the other hand, for the narrow width PGD case, the pipeline is relatively stiff and the pipeline lateral

displacement is significantly smaller than that of the soil. Hence, the pipeline strain is anticipated to be due to loading at the pipeline-soil interface (i.e., load controlled).

The maximum bending strain, ε_b , in the pipeline is given by the following:

– For the wide width PGD zone

$$\varepsilon_b = \pm \frac{\pi^2 \delta D}{W^2} \quad (26)$$

– For the narrow width PGD zone

$$\varepsilon_b = \pm \frac{p_u W^2}{3\pi E t W^2} \quad (27)$$

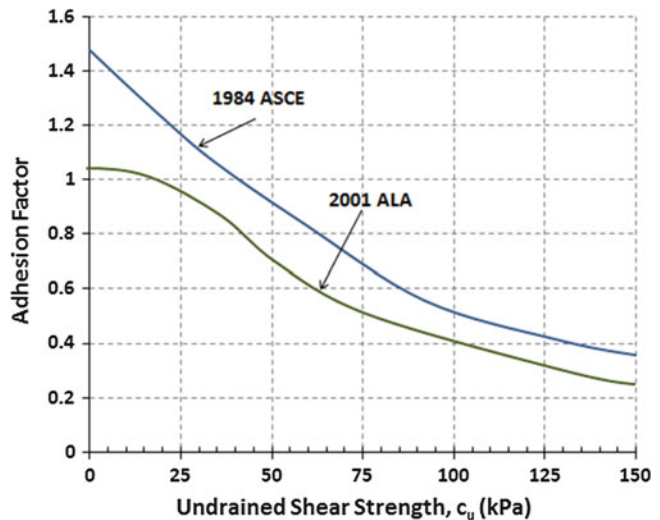
And the average axial tensile strain, ε_a , in the pipeline is estimated by the following:

Seismic Design of Pipelines, Table 1 Yield stress and Ramberg-Osgood parameters for mild steel and X-grade steel

	Grade B	X-42	X-52	X-60	X-70	X-80
Yield stress (MPa)	241	290	359	414	483	552
n	10	15	9	10	5.5	16
r	100	32	10	12	16.6	16

Seismic Design of Pipelines,

Fig. 7 Adhesion factors
(Reproduced after Honegger and Nyman 2004)



– For the wide width PGD zone

$$\varepsilon_a = \left(\frac{\pi}{2}\right)^2 \left(\frac{\delta}{W}\right)^2 \quad (28)$$

– For the narrow width PGD zone

The axial tension in this case is small and neglected.

where δ is the magnitude of the PGD, D is the pipe diameter, W is the length of the PGD zone, P_u is the maximum lateral force per unit length at the pipeline-soil interface, E is the elastic modulus of the pipeline material, and t is the pipe wall's thickness.

Abrupt Transverse PGD

Parker et al. (2008) proposed a simple model to analyze the response of pipelines to transverse PGD. Figure 10 presents the geometric and force details of the model. In this model the width of the abrupt transverse PGD is $2W_1$. Within this width the pipeline is subject to a lateral force per unit length P_{u1} . This lateral load is resisted by soil resistance forces P_{u2} over a distance W_2 on each side of the abrupt transverse PGD zone (see Fig. 10). Therefore, from horizontal equilibrium in the direction of pipeline, we get

$$P_{u1}W_1 = P_{u2}W_2 \quad (29)$$

The tensile force in the pipeline is assumed to be a constant value T_o within the PGD zone. Beyond the margins the pipeline axial tension decreases linearly at Points C and E to zero at Points A and G (see Fig. 10).

According to Parker et al. (2008), the total elongation due to the pipeline deformation

(Point B to Point F) can be estimated using the following equation:

$$\Delta L = \frac{1}{3} P_{u1}^2 W_1^3 (1 + P_{u1}/P_{u2}) / T_o^2 \quad (30)$$

Thus, the axial pipe strain is calculated as

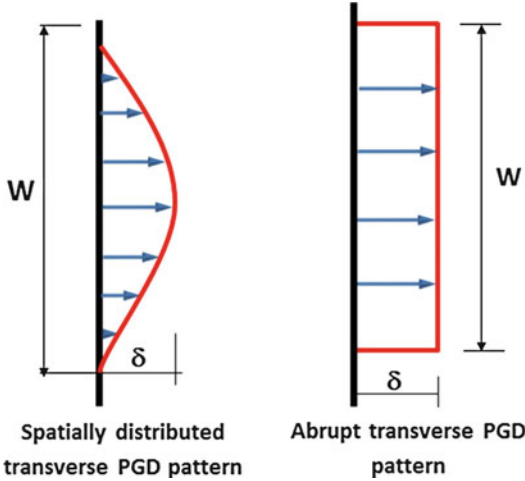
$$\varepsilon_a = \frac{\Delta L}{L} \quad (31)$$

and the bending strain in the pipeline can be evaluated as

$$\varepsilon_b = \frac{D P_{u1}}{2 T_o} \quad (32)$$

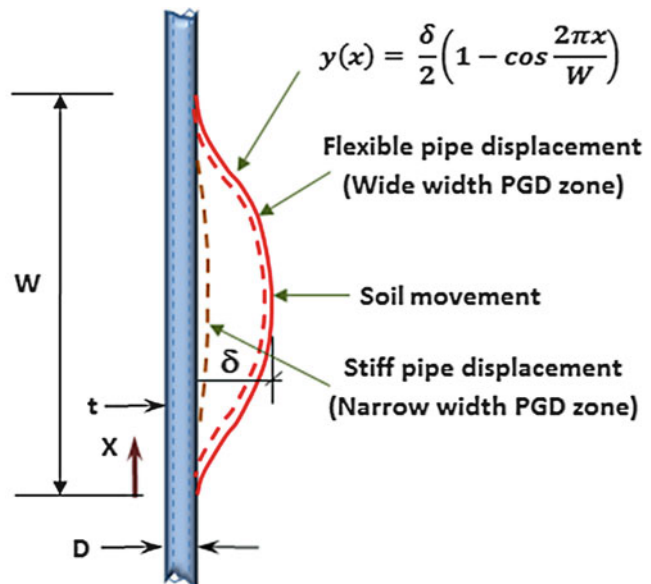
Pipelines in Liquefied Soil

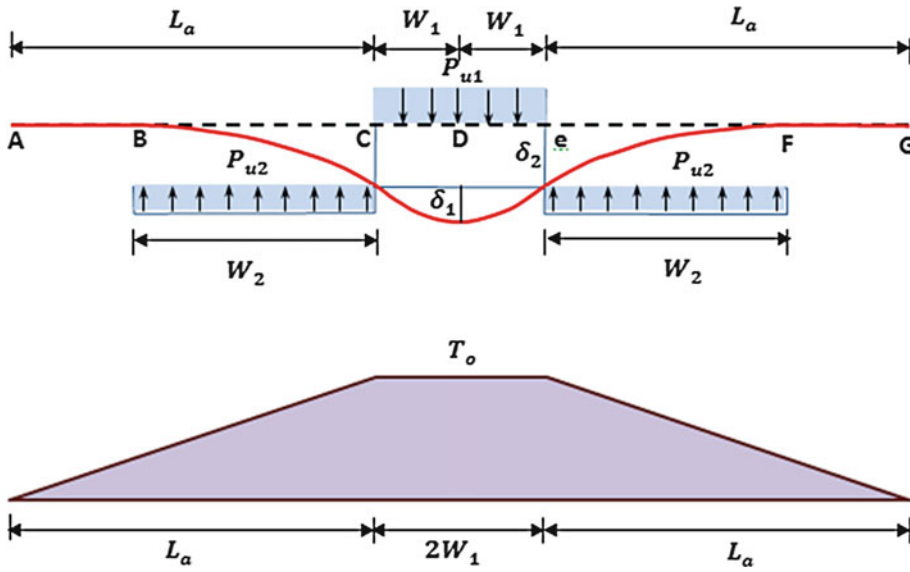
It is common practice that pipelines are buried at depths of 2 m or less from the ground surface. Thus, the top of the susceptible liquefiable soil layer is commonly located below the bottom of the pipeline. However, in some cases when the pipeline is buried at a river bed in saturated sand, for example, the soil surrounding the pipeline may liquefy during strong seismic shaking event. In this case,



Seismic Design of Pipelines, Fig. 8 Sketch of the considered patterns

Seismic Design of Pipelines,
Fig. 9 O'Rourke (1989)
model to analyze the
response of pipelines to
spatially distributed
transverse PGD





Seismic Design of Pipelines, Fig. 10 Geometric and force details of the Parker et al. (2008) model

the pipeline will probably deform laterally following the flow of the liquefied soil downward a mild slope or move upward due to buoyancy, particularly when something restrains the pipeline at one point or a compressive load acts on the pipeline.

Horizontal Movement

When a pipeline is surrounded by liquefied soil, the pipeline may move laterally due to the flow of liquefied soil downslope. The response of a buried pipe surrounded by liquefied soil subject to spatially distributed transverse PGD can be analyzed using the O'Rourke (1989) method presented earlier.

Vertical Movement

When a pipeline is surrounded by liquefied soil, the pipeline may uplift due to the buoyancy and moves upward. Hou et al. (1990) proposed analytical method to analyze the response of pipelines subjected to vertical movements. According to Hou et al. (1990), the uplifting force per unit length, P_{uplift} , acting on the pipeline within the liquefied zone is given by

$$P_{\text{uplift}} = \frac{\pi D^2}{4} (\gamma_{\text{soil}} - \gamma_{\text{contents}}) - \pi D t \gamma_{\text{pipe}} \quad (33)$$

where D is the pipe diameter, γ_{soil} is the unit weight of the liquefied soil, γ_{contents} is the unit weight of the contents inside the pipe (water, oil, gas, etc.), γ_{pipe} is the unit weight of the pipe material, and t is the pipe wall's thickness.

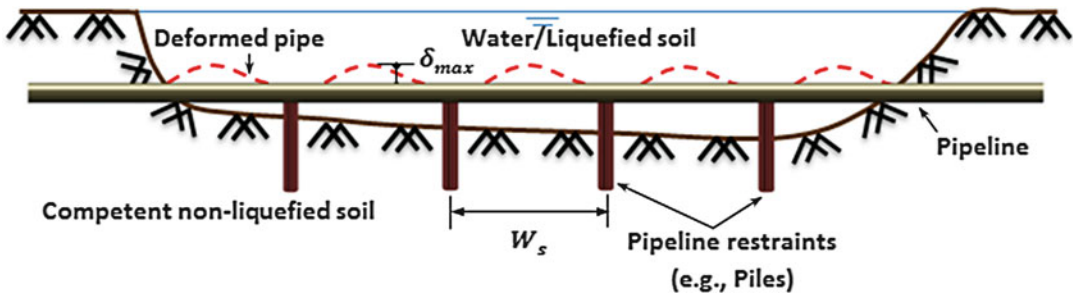
The maximum uplift displacement and/or the spacing for pipeline restraints is given by (see Fig. 11 below)

$$A \delta_{\text{max}}^3 + 16 I \delta_{\text{max}} - \frac{16 P_{\text{uplift}} W_s^4}{E \pi^5} \quad (34)$$

and the maximum strain in the pipeline is then given by

$$\epsilon_{\text{max}} = \pm \frac{\pi^2 \delta_{\text{max}} D}{W_s^2} + \frac{\pi^2 \delta_{\text{max}}^2}{4 W_s^2} \quad (35)$$

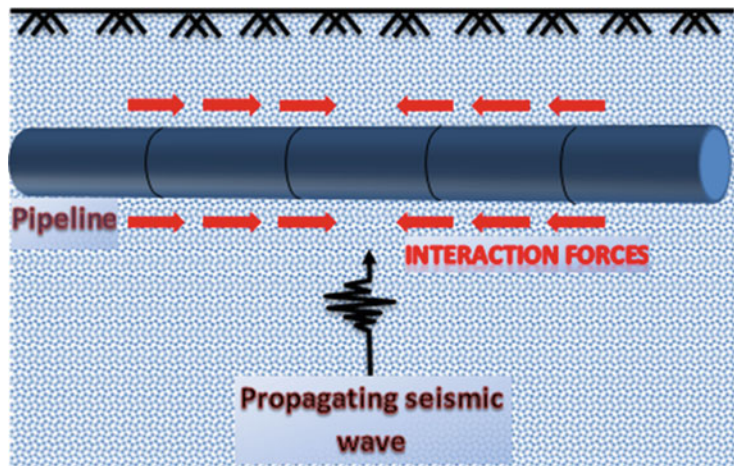
where A is the cross-section area, I is the moment of inertia, W_s is the spacing of the restraints, and E is the elastic modulus of the pipe's material.



Seismic Design of Pipelines, Fig. 11 Profile of pipeline crossing liquefied zone

Seismic Design of Pipelines,

Fig. 12 External forces on the buried pipeline by the adjacent soil during seismic shaking



Seismic Design Guidelines and Pipeline-Soil Interaction

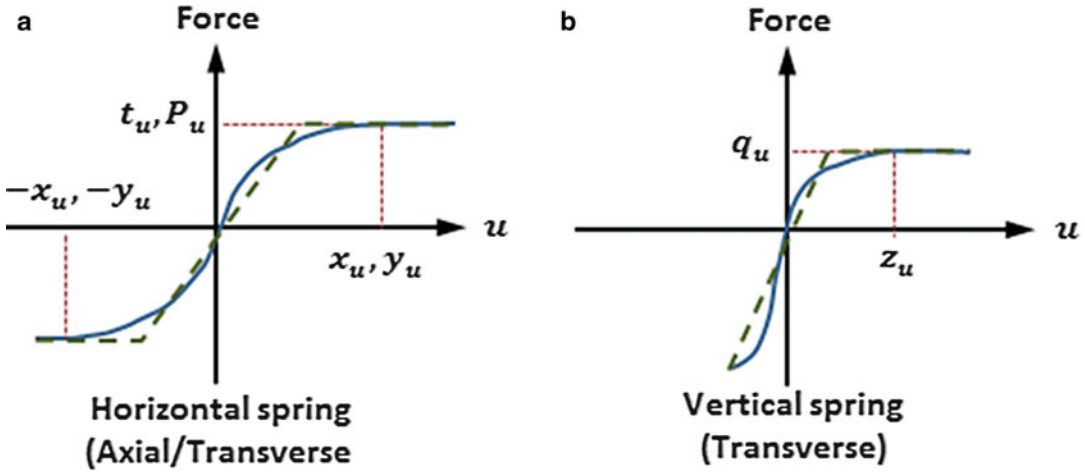
The pipeline-soil interaction effect exerts external forces on the buried pipeline by the adjacent soil when seismic motions are provided. The type of soil around the pipeline plays a significant role in its seismic behavior. In general, the soil displacement will produce friction like forces at the pipeline-soil interface (see Fig. 12 below). The overall seismic performance of buried pipeline is strongly related to the force-deformation relationship at the pipeline-soil interface (the p-y curves). For cohesionless soils, the probability of liquefaction becomes higher in loose materials. In cohesive soils, softer soils will undergo greater differential settlement due to consolidation and higher amplification effects and accordingly greater interaction forces.

Pipelines in Competent Non-liquefied Soil

The 1984 ASCE Guidelines suggest for the purpose of analysis idealized elastoplastic models for the force-deformation relationship at the pipeline-soil interface (see Fig. 13 below).

Longitudinal Movement

In this case the relative movement is parallel to the pipeline which results in axial forces at the pipeline-soil interface. The 1984 ASCE Guidelines provide relations for both cohesionless and cohesive soils. For cohesionless materials, the longitudinal resistance is due to the friction in the longitudinal direction at the pipeline-soil interface. The normal pressure which leads to the axial friction is the overburden and the lateral soil pressures. In the 1984 ASCE Guidelines, the normal pressure is taken as the average of the vertical and at rest lateral soil pressures acting



Seismic Design of Pipelines, Fig. 13 Idealized elastoplastic models for the force-deformation relationship at the pipeline-soil interface (Reproduced after ASCE 1984)

on the pipeline. (The 1984 ASCE Guidelines assumes $X_u \cong 2.5 \sim 5.0$ mm).

$$t_u = \pi D \gamma H \left(\frac{1 + k_o}{2} \right) \tan k\varphi \quad (36)$$

For cohesive materials, the longitudinal resistance is proportional to the adhesion at the pipeline-soil interface:

$$t_u = \pi D \alpha c_u \quad (37)$$

where D is the pipe diameter, γ is the effective unit weight of the soil, H is the depth of the pipeline, φ is the angle of shear resistance, k_o is the coefficient of lateral soil pressure at rest, k is a friction factor, α is the adhesion factor (given in Fig. 7), and c_u is the undrained shear strength of the soil.

Transverse-Horizontal Movement

In this case the relative movement is perpendicular to the pipeline which results in transverse-horizontal forces at the pipeline-soil interface. The 1984 ASCE Guidelines provide relations for both cohesionless and cohesive soils. For cohesionless materials, the maximum soil resistance in horizontal transverse direction may be calculated using the following equation:

$$P_u = \gamma H D N_{qh} \quad (38)$$

and the maximum elastic relative displacement in horizontal transverse direction is

$$y_u = \begin{cases} (0.07 \sim 0.10)(H + D/2) & \text{for loose sand} \\ (0.03 \sim 0.05)(H + D/2) & \text{for medium sand} \\ (0.02 \sim 0.03)(H + D/2) & \text{for dense sand} \end{cases} \quad (39)$$

For cohesive materials, the maximum soil resistance in horizontal transverse direction may be evaluated using the following equation:

$$P_u = c_u N_{ch} D \quad (40)$$

The maximum elastic relative displacement in horizontal transverse direction is

$$y_u = (0.03 \sim 0.05)(H + D/2) \quad (41)$$

where γ is the unit weight of the soil, H is the embedment depth of the pipeline, and N_{qh} and N_{ch} are the horizontal bearing capacity factors for sand and clay, respectively (Fig. 4).

Pipelines in Liquefied Soil

The response of continuous pipelines buried in liquefied soil layer is very sensitive to the

stiffness of the soil (Suzuki et al. 1988; Miyajima and Kitaura 1989). Based on several experimental results, it is recommended that the stiffness of liquefied soil ranges from 1/100 to 3/100 of that for non-liquefied soil (Yoshida and Uematsu 1978; Matsumoto et al. 1987; Yasuda et al. 1987; Tanabe 1988). Accordingly, the reduced stiffness will be used at the pipeline-soil interface. Analysis is then performed using the same procedures as in the competent non-liquefied soil case (on the conservative side).

Summary

This chapter focused on the seismic analysis and design of pipelines. In seismic events, buried pipelines can be damaged mainly by either the permanent ground deformations (PGD) or by the transient seismic wave propagation. Hence, different analysis methods for pipelines subjected to permanent ground deformation (PGD) or transient seismic wave propagation hazards were presented and discussed in details. In addition, the response of continuous pipelines subject to fault offsets was discussed. Also, several methods were introduced to predict the response of pipelines to either longitudinal or transverse PGD.

Earthquake shaking may trigger the liquefaction of saturated loose cohesionless soils. The liquefaction process increases the pore water pressure in the ground to a level where the effective stress approaches zero at which point the soil loses entirely its shear strength. A separate section in this chapter presented and discussed methods of analyzing and designing pipelines buried in liquefiable soils.

The overall seismic performance of buried pipelines is strongly related to the pipeline-soil interaction. Consequently, this chapter dedicated a section to provide guidelines for the seismic design of pipelines considering the pipeline-soil interaction effects as it plays a significant role in its seismic behavior.

Cross-References

- [Earthquake Magnitude Estimation](#)
- [Earthquake Mechanisms and Tectonics](#)
- [Earthquake Response Spectra and Design Spectra](#)
- [Liquefaction: Performance of Building Foundation Systems](#)
- [Seismic Vulnerability Assessment: Lifelines](#)
- [Soil-Structure Interaction](#)

References

- American Society of Civil Engineers (ASCE) (1984) Guidelines for the seismic design of oil and gas pipeline systems. Committee on Gas and Liquid Fuel Lifelines, ASCE
- Chen WW, Shih BJ, Wu CW, Chen YC (2000) Natural gas pipeline system damages in the Ci Ci earthquake (The City of Nantou). In: Proceedings of the 6th international conference on seismic zonation, Palm Springs, California.
- Hansen JB (1961) The ultimate resistance of rigid piles against transversal forces. Bulletin 12, Danish Geotechnical Institute, Copenhagen
- Honegger, D.G. and Nyman, D. (2004). PRCI guidelines for the seismic design and assessment of natural gas and liquid hydrocarbon pipelines. Pipeline Research Council International, Published by Technical Toolboxes, Houston, Texas
- Hou Z, Cai J, Liu X (1990) Response calculation of oil pipeline subjected to permanent ground movement induced by soil liquefaction. In: Proceedings of the China-Japan symposium on lifeline earthquake engineering, Beijing, pp 107–114
- Jibson RW, Keefer DK (1993) Analysis of the seismic origin of landslides: examples from the new Madrid seismic zone. *Geol Soc Am Bull* 105:521–536
- Kennedy RP, Chow AW, Williamson RA (1977) Fault movement effects on buried oil pipeline. *J Transp Eng Div ASCE* 103(TE5):617–633
- Lau DL, Tang A, Pierre J-R (1995) Performance of lifelines during the 1994 Northridge earthquake. *Can J Civ Eng* 22:438–451
- Matsumoto H, Sasaki Y, Kondo M (1987) Coefficient of subgrade reaction on pile in liquefied ground. In: Proceedings of the second national conference on soil mechanics and foundation engineering, Tokyo, Japan, pp 827–828 (in Japanese)
- Miyajima M, Kitaura M (1989) Effects of liquefaction-induced ground movement on pipeline. In: Proceedings of the Second U.S.-Japan workshop on liquefaction, large ground deformation and their effects on lifelines, Buffalo, Technical report NCEER-89-0032,

- Multidisciplinary Center for Earthquake Engineering Research, Buffalo, pp 386–400
- Newmark NM (1965) Effects of earthquakes on dams and embankments. *Geotechnique* 15(2):139–160
- Newmark NM (1967) Problems in wave propagation in soil and rocks. In: *Proceedings of the international symposium on wave propagation and dynamic properties of earth materials*, University of New Mexico Press, pp 7–26
- O'Rourke MJ (1989) Approximate analysis procedures for permanent ground deformation effects on buried pipelines. In: *Proceedings of the second U.S.-Japan workshop on liquefaction, large ground deformation and their effects on lifelines*, Buffalo. Technical report NCEER-89-0032, Multidisciplinary Center for Earthquake Engineering Research, Buffalo, pp 336–347
- O'Rourke TD, Palmer MC (1996) Earthquake performance of gas transmission pipelines. *Earthq Spectra* 12(3):493–527
- O'Rourke TD, Grigoriu MD, Khater MM (1985) Seismic response of buried pipes. *Pressure vessel and piping technology – a decade of progress*, ASME, pp 281–323
- O'Rourke MJ, Liu XJ, Flores-Berrones R (1995) Steel pipe wrinkling due to longitudinal permanent ground deformation. *J Transp Eng* 121(5):443–451
- Parker E, Moore R, Evans T, Usher N (2008) Evaluation of landslide impact on deepwater submarine pipelines. OTC 19459, *Offshore Technology Conference*, Houston
- Scawthorn C, Yanev PI (1995) Preliminary report 17 January 1995, Hyogoken Nambu, Japanese earthquake. *Eng Struct* 17(3):14657
- Suzuki H (1988) Damage to buried pipes caused by large ground displacement. In: *Proceedings of the first Japan-U.S. workshop on liquefaction, large ground deformation and their effects on lifeline facilities*, Tokyo, pp 127–132
- Takada S, Tanabe K (1988) Estimation of earthquake induced settlements for lifeline engineering. In: *Proceedings of the ninth world conference earthquake engineering*, vol VII, pp 109–114
- Tanabe K (1988) Fundamental study on seismic assessment and design of buried pipelines subjected to ground failure during earthquake. Doctoral dissertation, Kobe University (in Japanese)
- Wells DL, Coppersmith KJ (1994) New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement. *Bull Seismol Soc Am* 84(4):974–1002
- Wilson RC, Keefer DK (1983) Dynamic analysis of a slope failure from the 6 August 1979 Coyote Lake, California earthquake. *Bull Seismol Soc Am* 73(3):863–877
- Yasuda S, Saito K, Suzuki N (1987) Soil spring constant on pipe in liquefied ground. *Proceedings of the nineteenth JSCE conference on earthquake engineering*, Tokyo, Japan, pp 189–192 (in Japanese)
- Yoshida T, Uematsu M (1978) Dynamic behavior of a pile in liquefaction sand. *Proceedings of the fifth Japan earthquake engineering symposium*, Tokyo, Japan, pp 657–663 (in Japanese)
- Youd TL, Hansen C, Bartlett S (2002) Revised multilinear regression equations for prediction of lateral spread displacement. *J Geotech Geoenviron Eng ASCE* 128(12):1007–1017
- Zhang G, Robertson PK, Brachman RWI (2004) Estimating liquefaction-induced lateral displacements using the standard penetration test or cone penetration test. *J Geotech Geoenviron Eng ASCE* 130:861–871

Seismic Design of Tunnels

Youssef M. A. Hashash¹ and
Maria Ines Romero-Arduez²

¹Department of Civil and Environmental Engineering, 2230c Newmark Civil Engineering Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL, USA

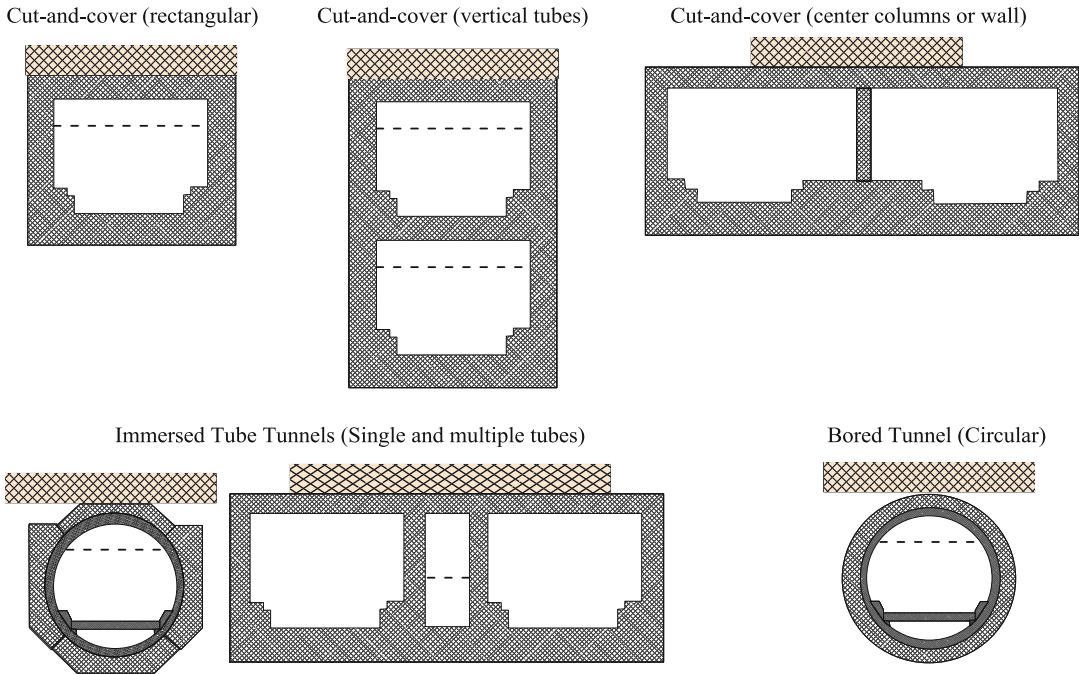
²Department of Civil and Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, USA

Synonyms

Cut-and-cover structures; Earthquakes; Free-field deformations; Immersed-tube tunnels; Nonlinear response; Racking; Soil-structure interaction; Tunnels

Introduction

The complete enclosure in soil or rock makes their seismic behavior different than that of aboveground structures or superstructures. Underground structure seismic response is constrained by seismic response of the surrounding soil and cannot experience free vibrations as is the case for aboveground structures. This section focuses on the seismic analysis and design of large linear underground structures commonly used for metro structures, highway tunnels, and large water/sewage transportation ducts in urban areas and can be grouped into three broad



Seismic Design of Tunnels, Fig. 1 Typical cross sections of underground structures

categories: (1) bored or mined tunnels, (2) cut-and-cover tunnels, and (3) immersed tunnels (Fig. 1).

The section starts with a selected review of performance of underground structures during seismic events. This is followed by presentation of a performance-based framework for design and analysis of underground structures considering both permanent and transient deformations. A number of seismic design issues are then discussed including vertical ground shaking and response, interaction of temporary and permanent structures, impact of superstructure and adjacent structures, transitions and tunnel joints, seismic retrofit of existing facilities, design considerations for structural support members, and precast tunnel lining.

Performance of Underground Facilities During Seismic Events

Based on several studies that documented earthquake damage to underground facilities in the past and the behavior of underground facilities

in recent large magnitude earthquakes (e.g., Tohoku, Japan, 2011; Maule, Chile, 2010), underground structures suffer appreciably less damage than surface structures. However, damage or failure of a limited section of an underground structure can be disruptive to post-earthquake recovery or operation in densely populated urban areas as such damage can interrupt the function of an entire system whether it is part of a mass transit or vehicular transportation network or large water or sewage transportation tunnels.

Damage is related to a number of parameters including ground motion intensity, ground conditions, and structural support system. Shallow tunnels tend to be more vulnerable to earthquake shaking than deep tunnels, and those constructed in soil can undergo more deformation than those constructed in competent rock. Circular bored tunnels are less susceptible to earthquake damage than cut-and-cover tunnels. Shaking damage can be reduced by stabilizing the ground around the tunnel and by improving the contact between the lining and the surrounding soil using grouting. Stiffening the lining without stabilizing the

Seismic Design of Tunnels, Fig. 2 Street view of Dakai subway station collapse (Iida et al. 1996)



surrounding poor ground may only result in excessive seismic forces in the lining. Damage at tunnel portals may be caused by slope instability (Hashash et al. 2001).

Underground Structures in the United States

The Bay Area Rapid Transit system (BART) in San Francisco, California, sustained the 1989 Loma Prieta earthquake without damage and was operational after the earthquake. It consists of underground stations and tunnels embedded in soft bay mud deposits connected to Oakland via the transbay-immersed-tube tunnel. It was one of the first underground facilities designed with seismic considerations: special seismic joints were designed to accommodate differential movements. Limited displacements were measured at that joint.

During the same earthquake, the Alameda Tubes, a pair of immersed-tube tunnels that connect Alameda Island to Oakland in California, experienced structural cracking on the ventilation buildings and limited water leakage due to liquefaction of loose deposits above the tubes.

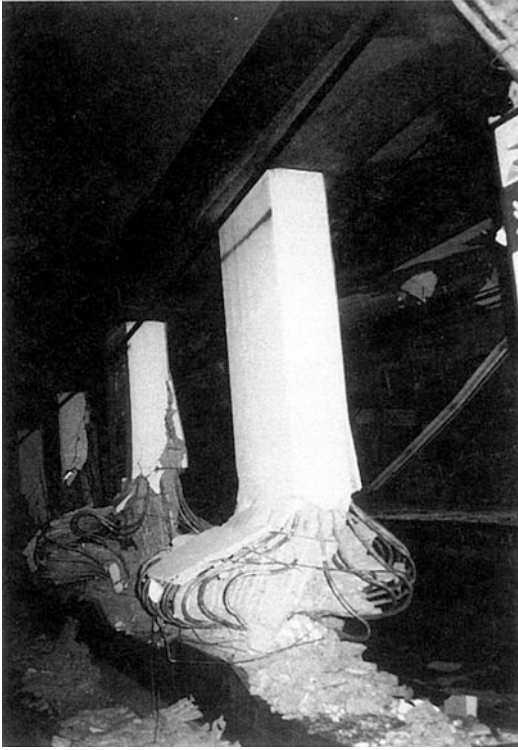
The 1994 Northridge earthquake caused no damage to concrete lining of bored tunnels of the metro system in Los Angeles, California (Hashash et al. 2001).

Underground Structures in Japan

The Dakai subway collapse (Fig. 2) during the 1995 Hyogoken-Nambu earthquake in Kobe, Japan, was the first collapse of an urban underground structure due to earthquakes shaking. The collapse experienced by the center concrete columns (Fig. 3) was due to lack of shear reinforcement, leading to a collapse of the ceiling slab and settlement of the soil cover. The 1962 station design did not include specific seismic provisions. However, in the 2011 Tohoku earthquake, the underground subways in Sendai experienced strong shaking with no reports of damage. Only a water distribution plant in Kashima City and a wastewater treatment plant in Itako City were observed to have been damaged by liquefaction. The damages include uplift of buried tanks, offsets in underground tunnels, damage to support utilities, and damage to major trunk lines on and off the site (Ashford et al. 2011).

Underground Structures in Taiwan

The 1999 Chi-Chi earthquake did not cause visible signs of damage in several highway tunnels located in central Taiwan. The main damage occurred at tunnel portals because of slope instability. No damage was reported in the Taipei subway, located 100 km from the ruptured fault



Seismic Design of Tunnels, Fig. 3 Dakai subway station collapse (Iida et al. 1996)

zone (Fig. 4). However, some tunnels in mountainous areas were severely damaged due to slope failure (Fig. 5).

Underground Structures in Turkey

The August 17, 1999, Kocaeli earthquake had minimal impact on the Bolu twin tunnels, a 1.5-billion-dollar project under construction at that time. It had an excavated section of 15 m tall by 16 m wide and crossed several minor faults parallel to the North Anatolian Fault. After the earthquake, continuous monitoring showed no movement due to the earthquake. The November 12, 1999, earthquake caused collapse of both tunnels 300 m from its eastern portal, in a clay gauge material in the unfinished section (Hashash et al. 2001).

Underground Structures in Chile

Cut-and-cover highway box structures with three lanes of traffic in one direction and approximately 1 km long constructed in relatively stiff

gravels, the Santiago Metro tunnels and underground stations, and highway tunnels on the southbound of Route 5 appeared to be undamaged by the 2010 Maule earthquake in Chile, as shown in Fig. 6. In the northbound of Route 5, “La Calavera” tunnel near Calera has a rock block dislodged, but the tunnel was old and had problems before the earthquake (Elnashai et al. 2010).

In summary, well-engineered underground structures performed well even under strong shaking in recent earthquakes in different parts of the world. However, underground structures are vulnerable to permanent ground displacements such as liquefaction, slope stability, and fault displacement. There is also vulnerability to transient ground motions when insufficient structural detailing is provided, the underground structure is constructed in loose ground, masonry lining is used, or near field effects are present.

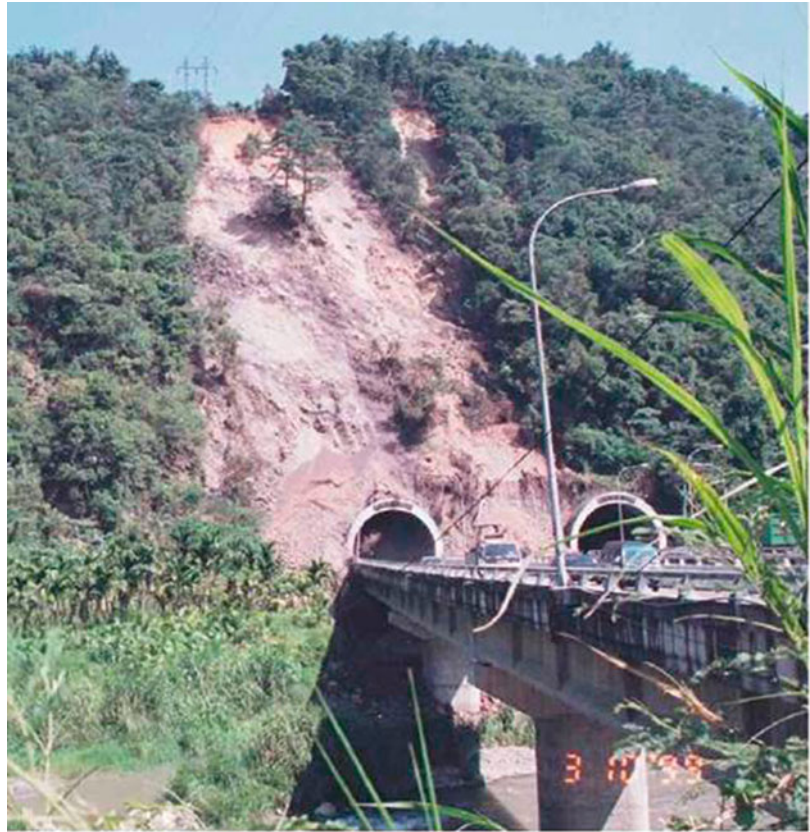
Performance-Based Seismic Evaluation Framework

Underground structures under earthquake effects can undergo permanent deformations and/or transient deformations. Factors influencing these effects include shape, dimensions, and depth of the structure, properties of the surrounding soil or rock, properties of the structure, and severity of ground shaking (Hashash et al. 2001). Table 1 summarizes a performance-based framework for seismic design and analysis of underground structures. The framework consists of three main steps: definition of seismic environment, evaluation of ground response to shaking, and assessment of structure response due to seismic shaking.

Step 1: Definition of Seismic Environment

Seismic analysis of underground structures starts with site-specific definition of its seismic environment. A detailed field and laboratory investigation program is necessary; the field investigation program should include definition of the site stratigraphy and direct measurements of shear wave velocity profiles and cone

Seismic Design of Tunnels, Fig. 4 Slope failure at tunnel portal, Chi-Chi earthquake, central Taiwan (Hashash et al. 2001)



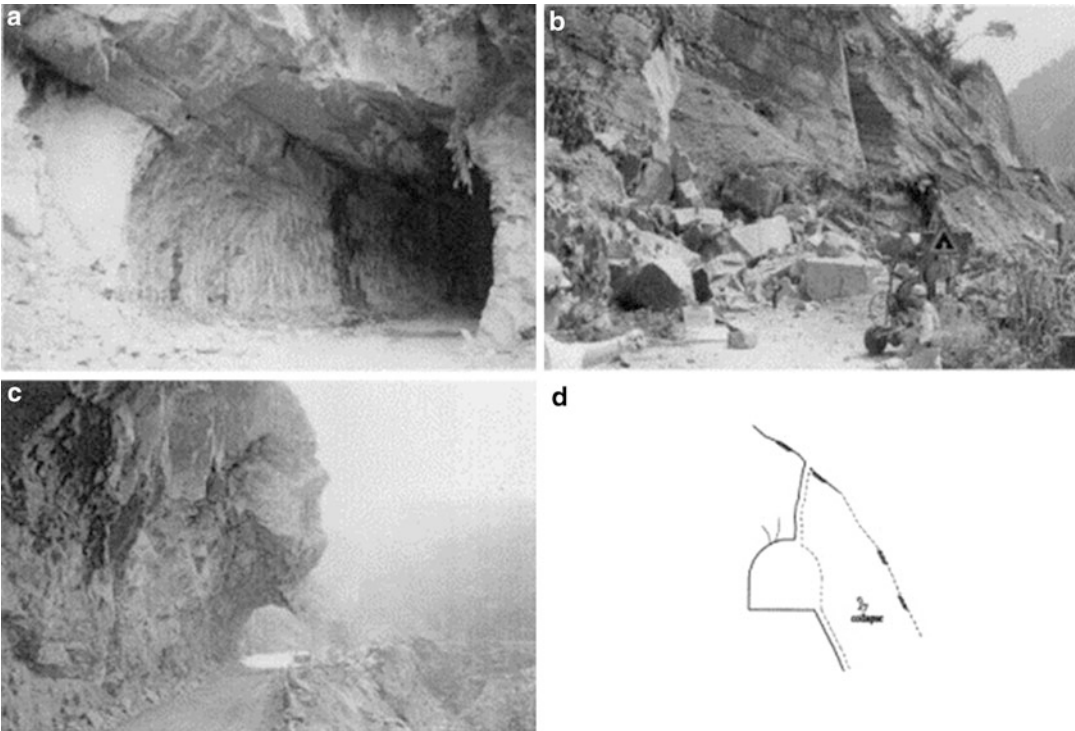
penetration resistance of soft soils as well as assessment of potential geo-hazards including slope instability, fault displacement, and lateral spreading. Appropriate static and cyclic laboratory tests for major soil units are also required.

Site-specific probabilistic and/or deterministic seismic hazard analyses, as well as hazard analyses using conditional (mean) spectra, are needed to define seismic hazard levels for permanent condition of the structure (operational and maximum levels). Increasingly seismic hazard, using a shorter return period, is being considered for temporary conditions during construction.

Seismic performance criteria selection is a crucial aspect in the design of an underground structure. Performance objectives include explicit target performance of the structure and system performance. Performance objectives are not purely technical requirements and should include owner and user requirements, policy considerations, and life-cycle costs. It is an iterative process

based on analysis findings to answer the question of what is feasible and at what cost as illustrated in Fig. 7.

Often a two-level criterion is adopted: operating design earthquake (ODE) and maximum design earthquake (MDE). Those are defined using response spectra developed in the seismic hazard analysis. A suite of three component motions is needed for each of the design earthquake levels for site response analysis and soil-structure interaction modeling. It is preferred to use recorded motions instead of synthetic motions to spectrally match the target spectra. Ground motion spatial incoherence must be taken into account for long structures including (1) wave passage, (2) extended source effects, (3) ray-path effects, and (4) local site effects. One-dimensional equivalent linear and nonlinear site response analyses are conducted to assess how the ground motion is affected by the soil column. One-dimensional site response

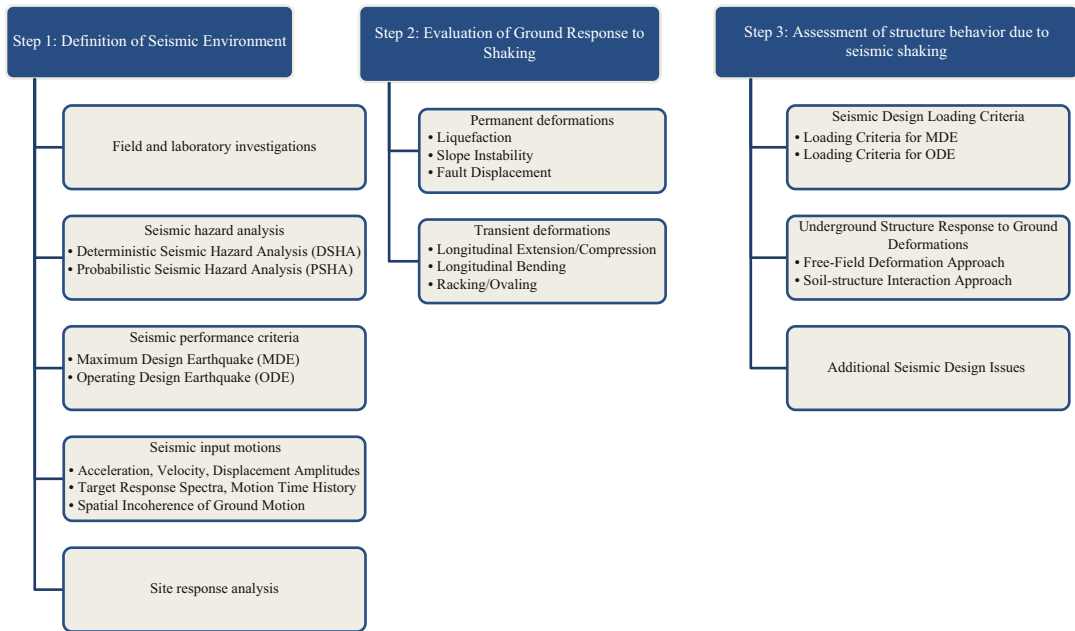


Seismic Design of Tunnels, Fig. 5 Chi-Shue tunnel before and after Chi-Chi earthquake (Wang et al. 2001)



Seismic Design of Tunnels, Fig. 6 Left: Highway box structures in Santiago, Right: Highway tunnel in Route 5 South

Seismic Design of Tunnels, Table 1 Performance-based framework for design and analysis of underground structures



analyses are used in the analysis of underground structures to:

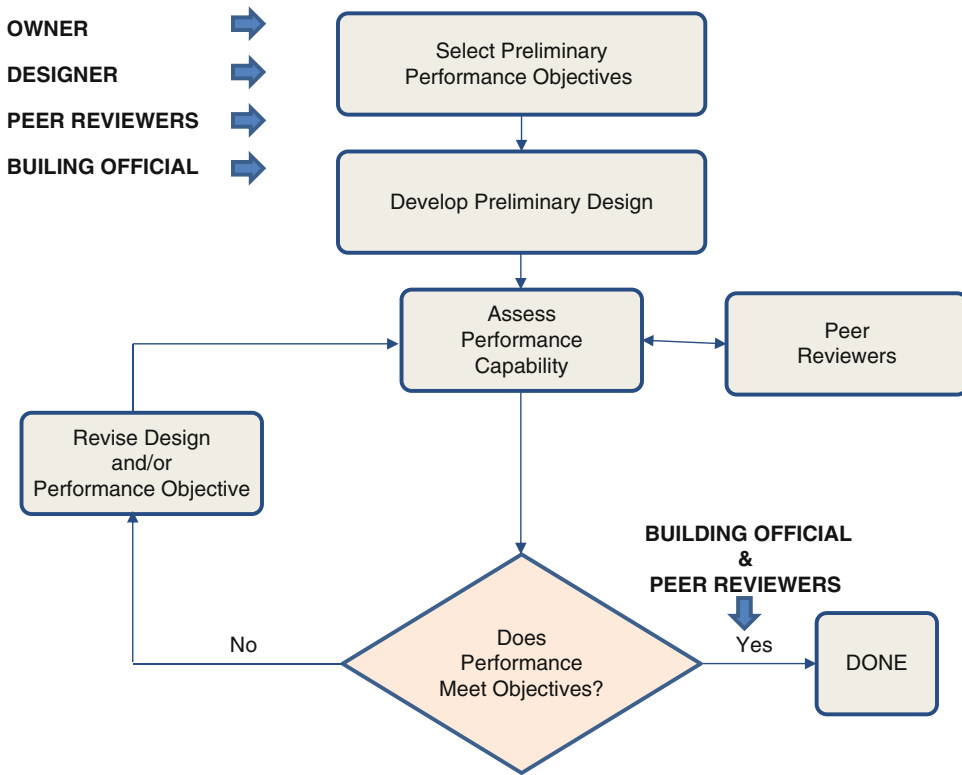
- (a) Obtain free-field racking deformations (differential sideways movements between the top and bottom elevations of rectangular structures) along the underground structure height which can be used in pseudo-static soil-structure interaction.
- (b) Obtain input motions for dynamic soil for dynamic soil-structure interaction analysis.
- (c) Obtain strain-compatible soil properties for use in pseudo-static and dynamic soil-structure interaction.
- (d) Assess potential liquefaction and ground failure.

Step 2: Evaluation of Ground Response to Shaking

Evaluation of ground response to shaking can be divided into permanent deformations or ground failure and transient deformations or ground shaking.

Permanent deformations or ground failure includes liquefaction, slope instability, and fault displacement. Liquefaction, prevalent in loose sand and fill deposits, can result in generation of sand boils, loss of shear strength, lateral spreading, and slope failure. Tunnels in liquefiable deposits can experience increased lateral pressures, loss of lateral passive resistance, flotation or sinking, lateral displacements if lateral spreading happens, permanent settlement, and compression/tension failure after soil consolidation. A landslide intercepting a tunnel can result in concentrated shearing displacements and collapse of a cross section. The potential for these failures is greatest when a pre-existing landslide intersects the tunnel, in shallower parts of tunnel, and at tunnel portals. An underground structure may have to pass across an active fault zone; in these situations the tunnel must tolerate the expected displacements. The design for permanent deformations is discussed in section “[Design for Permanent Deformations.](#)”

Transient deformations can be quite complex due to interaction of seismic waves with surficial



Seismic Design of Tunnels, Fig. 7 Performance-based design procedure after Hamburger and Hooper (2011)

deposits. Underground structures undergo three primary modes of deformation during seismic shaking: compression-extension, longitudinal bending, and ovaling/racking. The design for transient deformations is discussed in section “[Performance Evaluation Under Transient Ground Deformations](#).”

Step 3: Assessment of Structure Behavior Due to Seismic Shaking

The evaluation of structure behavior will be primarily a deformation controlled soil-structure interaction problem. Pseudo-static and dynamic soil-structure interaction approaches have been used in the evaluation of structure response and are discussed in section “[Performance Evaluation Under Transient Ground Deformations](#).”

Section “[Additional Seismic Performance Issues](#)” discusses additional seismic design issues such as vertical ground shaking and response;

interaction of temporary and permanent structures; permanent changes in state of stress of soil; impact of superstructure and adjacent structures; tunnel joints such as portals, stations, and tunnel segment; seismic retrofit of existing facilities; design considerations for structural support members; precast tunnel lining; and seismic design of buried reservoirs.

Design for Permanent Deformations

Designing underground structures for permanent deformations may not be viable, but ground stabilization techniques can help prevent large deformations. Some solutions include ground improvement, drainage, soil reinforcement, grouting, earth retaining systems, or even removing problematic soils or relocating the tunnel alignment.

Underground Structures Crossing Active Faults

In the Century Area Tunneling Safety and Fault Investigations TAP report for the Los Angeles Metro (Cording et al. 2011), some recommendations are provided for consideration when an underground tunnel segment crosses an active fault:

- The segments must be designed to accommodate fault displacement without collapse and with the capability of being repaired.
- The alignments should be selected so that the tunnel crosses at a relatively sharp angle to the fault zone to minimize the length of tunnel that must accommodate fault displacements.
- Methods employed to allow fault displacement on a tunnel lining include excavating to a larger section at the crossing to facilitate realignment, providing the tunnel with a strong but flexible lining like ductile steel segments or articulated joints, and placing crushable backpacking material around the structural lining.
- As there is no precedent of placing metro underground stations on active faults, designing for it will be extremely difficult and cost-prohibitive. If possible, it is advisable to avoid this scenario.

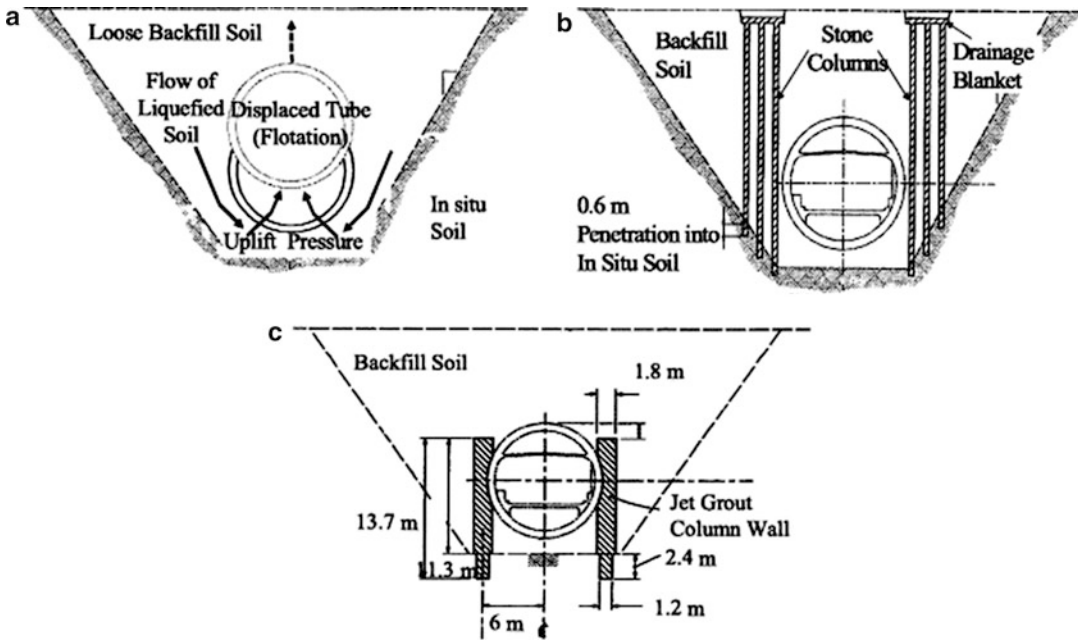
The design approach for tunnels crossing active faults will change depending on the displacement magnitude and the tunnel width over which the displacement is distributed:

- (a) If fault displacements are small and/or distributed over a relatively wide zone, providing articulation of the tunnel lining through ductile joints is a possible solution. The closer the joint spacing, the better the tunnel performance will be; this is more viable in soft soils where displacements can be effectively redistributed over the tunnel lining. The tunnel can then deform in an S-shape through the fault zone without rupture. It is always necessary to keep the tunnel watertight when using joints.
- (b) If large displacements are concentrated in a narrow zone, retrofit will consist of enlarging the tunnel section across and beyond the displacement zone. The length over which the enlargement is made is a function of fault displacement and permissible curvature of the road or track; the longer the enlarged tunnel, the smaller the post-earthquake curvature (Power et al. 1996). This solution has been implemented in the San Francisco BART system and Los Angeles Metro rapid transit tunnel system. Concrete-encased steel ribs provide sufficient ductility to accommodate distortions with little strength degradation. Under axial displacements, even though compression is more damaging to the tunnel lining than extension, both will result in unacceptable water inflow. A solution for water tightness is flexible couplings (Wang 1993), used for the Southwest Ocean Outfall in San Francisco. Cellular concrete may also be used within the enlarged tunnel, because it has a low yield strength that can minimize the loads on the tunnel liner while also providing adequate resistance for normal soil pressures and other seismic loads.

Estimating fault displacement is a key issue to design tunnels crossing active faults. One option to estimate fault displacement is using empirical relationships that express expected displacements in terms of some source parameter. Deterministic and probabilistic fault displacement hazard analyses can be used to assess fault displacement hazard where a displacement attenuation function is used in a probabilistic seismic hazard analysis (Coppersmith and Youngs 2000; Youngs et al. 2003).

Flotation in Liquefiable Deposits

Liquefaction evaluation is discussed elsewhere. If liquefaction is limited to soil layers above the underground structure, then it is unlikely to influence the racking deformations of the structure. However, if the structure is partially or entirely embedded in liquefiable soil, additional evaluations are required. Underground structures may experience flotation in liquefiable deposits.



Seismic Design of Tunnels, Fig. 8 Isolation principle, use of cutoff walls to prevent tunnel uplift (Schmidt and Hashash 1999)

As shown in Fig. 8a, when the tunnel experiences uplift due to flotation, the liquefied soil moves underneath the displaced tunnel and lifts it further up (Schmidt and Hashash 1999). Uplift can be prevented through isolation using cutoff walls, such as sheet pile walls; stone columns (Fig. 8b); or jet grout columns (Fig. 8c). Sheet piles with drainage capability can also reduce excess pore water pressure. The rise in excess pore water pressure is prevented at the bottom of the tunnel and in the soil underneath with these barrier walls. With longer barrier walls and a wider structure, uplift is more difficult. After the liquefaction potential is mitigated, flexible joints can be used to allow for differential displacements at tunnel connection joints.

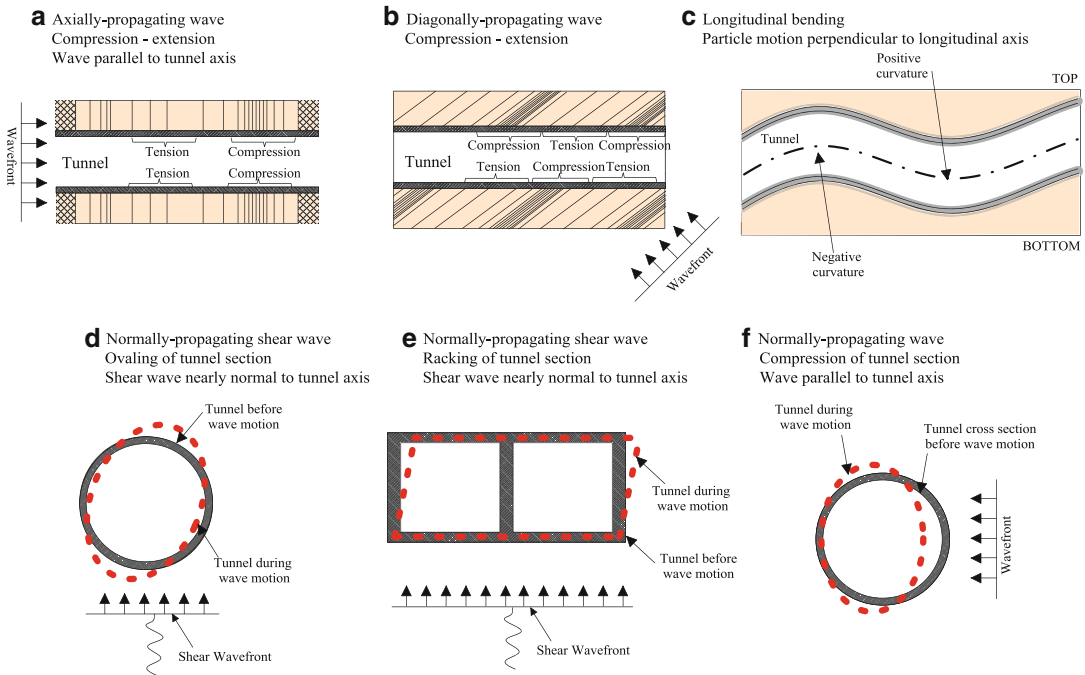
Slope Instability and Lateral Spreading

Stabilizing the soil or changing the alignment is often the most practical way to mitigate slope instability and lateral spreading. It is challenging to design an underground structure to resist or take these movements unless the hazard is localized and the movement is small (Power et al. 1996).

Performance Evaluation Under Transient Ground Deformations

If permanent deformations are not expected, then the underground structure must be designed for transient ground deformations. In this case, the underground structure response is controlled by the ground deformation and the peak ground velocity.

The focus of underground structure seismic design is on free-field deformations of the ground and their interaction with the structure, since the inertia of the surrounding soil is large relative to the inertia of the structure. Figure 9 shows response of underground structures to seismic motions: axial compression and extension, longitudinal bending, and ovaling/racking. Axial deformations are due to seismic waves producing motions parallel to the tunnel axis, bending is due to seismic waves producing particle motion perpendicular to the longitudinal axis, and ovaling/racking is due to shear waves propagating normally to the tunnel axis. Design considerations for axial and bending deformations are generally in the direction along



Seismic Design of Tunnels, Fig. 9 Deformation modes of tunnels due to seismic waves

the tunnel axis and in the transverse direction for ovaling/racking.

There are four main approaches to tackle the design for transient deformations: free-field deformation methods, pseudo-static soil-structure interaction analyses, dynamic soil-structure interaction finite-element analyses, and dynamic earth pressure methods. The merits and drawbacks of these methods are summarized in Table 2.

Free-field deformation methods assume that the underground structure deformations are identical to those of the surrounding ground. They do not take into account soil-underground structure interaction and are most appropriate when the structure (racking) stiffness is equivalent to that of the surrounding ground.

Pseudo-static soil-structure interaction models account for the kinematic interaction between the soil and the underground structure neglecting inertial interaction. They are often used for practical design purposes when the structure is not too complex (NCHRP 611, Anderson et al. (2008)).

Nowadays, the ease of access to high performance computers makes it possible to perform

dynamic soil-structure interaction analyses within a reasonable amount of time. These types of analyses allow problems with complicated tunnel geometry and ground conditions to be solved efficiently. However, the selection of parameters for a complex problem requires expertise; therefore, it is important to always verify the computer model solution with simpler pseudo-static or closed-form solutions.

The presence of a rectangular frame structure in the ground will induce dynamic earth pressures acting upon the structure. Complex shear and normal stress distributions along the exterior surfaces of the structures are expected, but quantifying those distributions require rigorous dynamic soil-structure interaction, since they heavily depend on how the interface is modeled.

In the past, the Mononobe-Okabe method was used to calculate the seismic-induced dynamic earth pressures on underground structures. The method assumes the earthquake load is caused by inertial forces of the surrounding soil and calculates the load using soil properties and a determined seismic coefficient. This method is not applicable in the case of underground structures,

Seismic Design of Tunnels, Table 2 Comparison of seismic design approaches

Approaches	Advantages	Disadvantages	Applicability
Free-field deformation methods	1. Comparatively easy to formulate, many 1D wave propagation programs available	1. Nonconservative for tunnel structure more flexible than ground 2. Conservative for tunnel structure stiffer than ground 3. Overly conservative for tunnel structures significantly stiffer than ground 4. Less precision with highly variable ground conditions	For tunnel structures with equal stiffness to ground
Pseudo-static soil-structure interaction methods	1. Good approximation of soil-structure interaction 2. Comparatively easy to formulate 3. Reasonable accuracy in determining structure response 4. Computationally efficient 5. Sensitivity analysis can be easily performed	1. Ignores inertial effects 2. Less precision with highly variable ground 3. Shear displacement not transmitted uniformly to shallow box structures	Most conditions except for variable soil profile, shallow structures
Dynamic soil-structure interaction finite-element analysis	1. Best representation of soil-structure system 2. Best accuracy in determining structure response 3. Capable of solving problems with complicated tunnel geometry and ground conditions (significant variations in soil stiffness)	1. Computationally demanding 2. Uncertainty of design seismic input parameters may be several times the uncertainty of the analysis	All conditions
Dynamic earth pressure methods	1. Serve as additional safety measures against seismic loading	1. Lack of rigorous theoretical basis 2. Resulting in excessive deformations for tunnels with significant burial 3. Use limited to certain types of ground properties	None

since they will move with the ground and will not form an active wedge.

When designing underground structures for transient deformations, sufficient ductility is needed to absorb imposed deformations without losing the capacity to carry static loads. Care should be exercised in not increasing the stiffness of the structure as this tends to attract additional loads thus increasing the demand on the structure.

Free-Field Deformation Approach

Free-field deformations are the deformations caused by seismic waves on a given soil profile in the absence of structures or excavations. The interaction between the soil and the underground structure is neglected, but provides a first-order

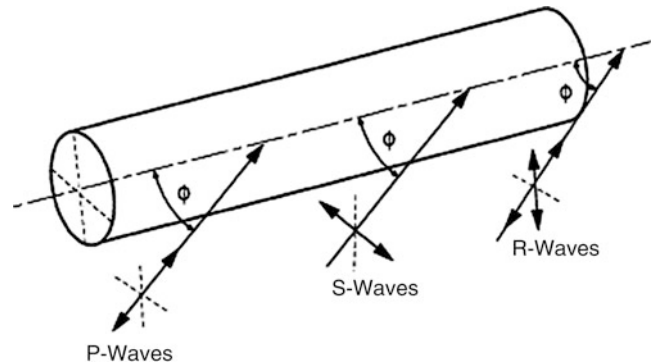
estimate of the underground structure deformation. Imposing the free-field deformations directly on the underground structure can underestimate or overestimate the structure deformations.

Closed-Form Elastic Solutions

Initial estimates of strains and deformations in a tunnel can be developed using simplified closed-form solutions. There are many assumptions within the formulation of these methods:

- (a) The seismic wave field is considered as a plane of wave with the same amplitude at all locations along the tunnel, differing only in their arrival time. Wave scattering and 3D wave propagation are neglected. The results

Seismic Design of Tunnels, Fig. 10 Seismic waves causing longitudinal axial and bending strains (Power et al. 1996)



of these analyses should be interpreted with care (Power et al. 1996).

- (b) Harmonic wave propagating at a given angle of incidence in a homogeneous, isotropic, elastic medium. The critical incidence angle resulting in the maximum strain is typically used (Newmark 1967). However, the strain order of magnitude estimated by this method is useful as initial design tool and design verification method.
- (c) St John and Zahrah (1987) developed free-field solutions for axial and curvature strains due to compression, shear, and Rayleigh waves. Figure 10 shows the seismic waves causing the strains. Treating the tunnel as an elastic beam allows the calculation of combined axial and curvature deformations.

The strain bending component is relatively small compared to axial strains, but if the tunnel radius increases, the curvature contribution increases. Tunnel cracks may open and then close in the lining due to the cyclic nature of the axial strains. As long as the cracks are small, are uniformly distributed, and do not affect the performance of the tunnel, even unreinforced concrete linings are considered adequate. It is important to emphasize that the p- and s-wave velocities used are those of the deep rock. The range for s-wave is between 2 and 4 km/s and p-wave between 4 and 8 km/s (Power et al. 1996).

Ovaling and Racking Deformation

Ovaling deformations, developed by waves acting perpendicular to the circular tunnel lining, are

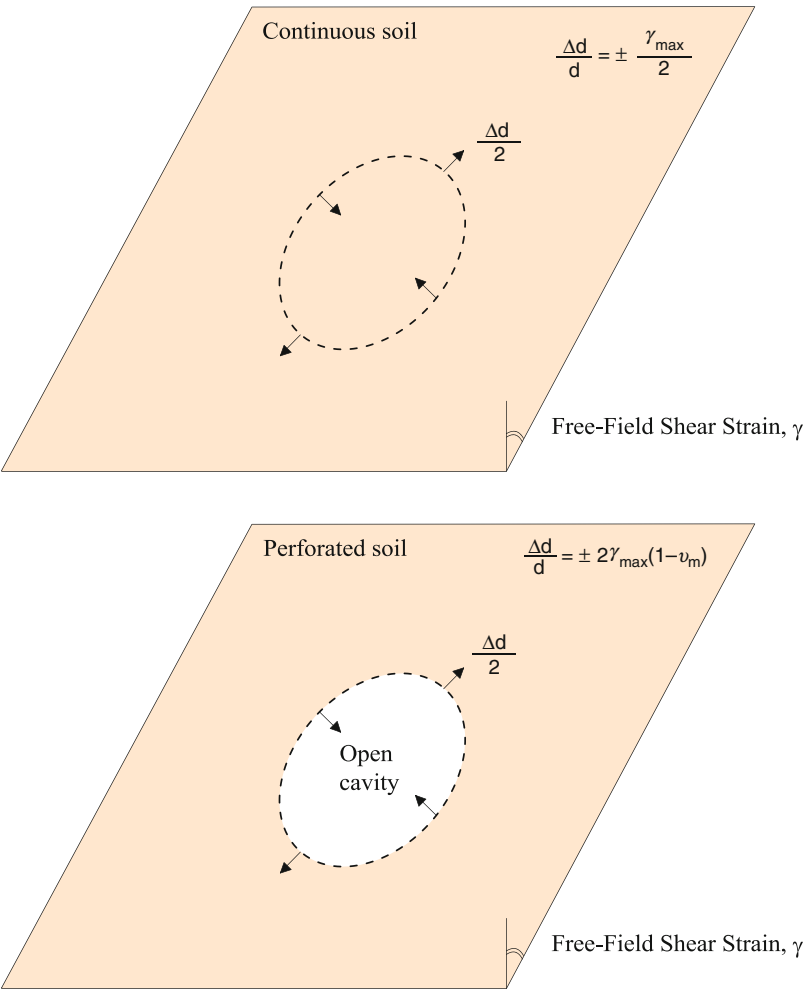
caused predominately by vertically propagating shear waves (Wang 1993). Ground shear distortions can be defined assuming a non-perforated ground or a perforated ground. As shown in Fig. 11, both cases ignore the tunnel lining (soil-structure interaction), where the maximum diametric strain is in terms of the maximum free-field shear strain ($\gamma_{\max}$) and the Poisson ratio (ν_m). The first can be used to approximate the behavior of a tunnel lining whose stiffness is equal to the medium it replaces. The second can be used to approximate the behavior of a tunnel lining whose stiffness can be neglected in comparison with the stiffness of the medium.

A rectangular box structure will undergo transverse racking deformations (Fig. 12) when subjected to earthquake shear distortions. Racking deformations are defined as the differential sideways movements between the top and bottom elevations of rectangular structures. If an initial calculation is needed, it can be calculated based on St John and Zahrah (1987) equations.

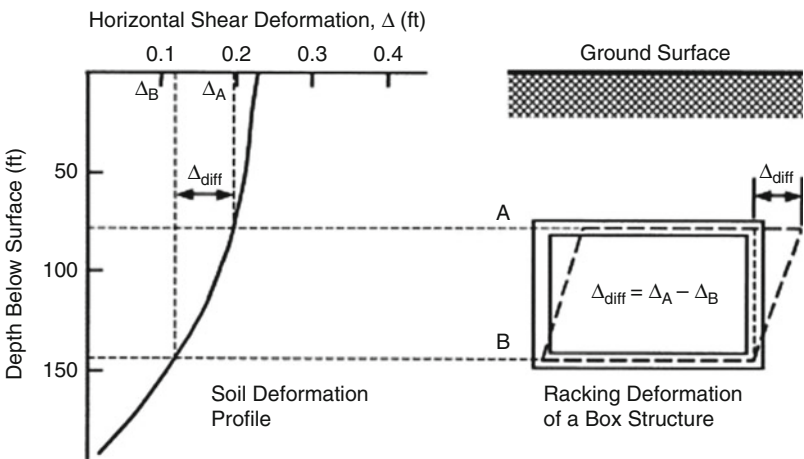
Numerical Analysis

Many computer programs are available to estimate free-field shear distortions: SHAKE (Schnabel et al. 1972), FLUSH (Lysmer et al. 1975), D-MOD (Matasovic 1993), and DEEPSOIL (Hashash et al. 2011; Hashash and Park 2001), among others. One-dimensional site response analyses can be used to characterize the change in the propagating ground motions on variable soil profiles, but these analyses only take into consideration vertically propagating shear waves. However, these are the waves that

Seismic Design of Tunnels, Fig. 11 Free-field shear distortions of perforated and non-perforated ground



Seismic Design of Tunnels, Fig. 12 Typical free-field racking deformation imposed on a rectangular frame (Wang 1993)



carry most of the seismic energy. The analyses can be performed in equivalent linear frequency domain or nonlinear time domain. The resulting free-field shear distortion can be expressed in the form of shear strain or shear deformation profile with depth.

Applicability of Free-Field Deformation Approach
The free-field deformation approach is a simple and effective design tool when earthquake-induced ground motions are small. However, in structures located within soft soil profiles, the method gives overly conservative designs, because free-field ground distortions in these soils are large. It also neglects the difference in stiffness between the lining and the surrounding soil. The presence of an underground structure modifies the free-field deformations; methods to model this interaction will be described in the following sections.

Pseudo-static Soil-Structure Interaction

In pseudo-static soil-structure interaction analyses, the soil and structure inertia due to seismic shaking is neglected. The problem is simplified to that of a structure in a soil medium subjected to simple shear on horizontal and vertical planes.

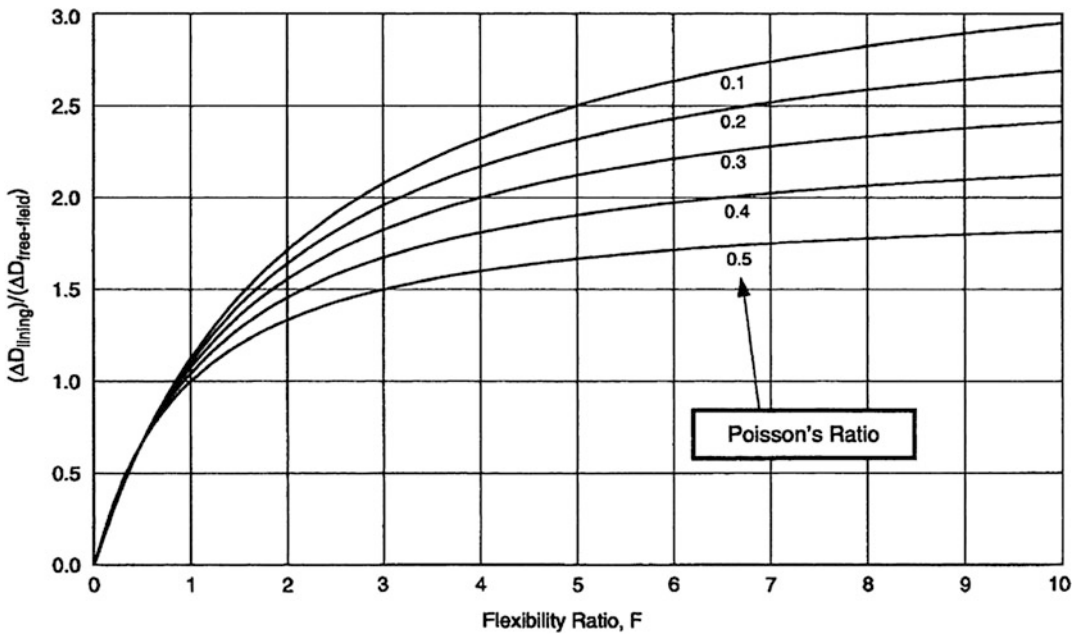
The beam-on-elastic foundation approach is used to model soil-structure interaction effects. Both the lining and the medium are assumed to be linear elastic. Wang (1993) presents a summary of closed-form elastic solutions for axial force and moment developed for circular tunnels due to seismic waves propagating along and perpendicular to the tunnel axis. Adding stiffness and strength to the structure may attract more forces, so a better solution would be to add ductility. These solutions are dependent on the estimates of appropriate spring coefficients compatible with anticipated displacements and wave lengths. They are often limited to idealized seismic wave forms.

Most pseudo-static SSI analyses focus on the interaction of vertically propagating shear waves with the transverse section of a tunnel. These analysis approaches are described next.

Transverse: Ovaling Deformations of Circular Tunnels

Peck et al. (1972) proposed closed-form solutions in terms of thrust, bending moments, and displacement under external loading. The lining response was a function of structure compressibility and flexibility ratios, in situ overburden pressure, and at-rest earth coefficient. To adapt to seismic loading, the free-field shear stress replaces the in situ overburden pressure and earth coefficient. The stiffness of the tunnel relative to the ground is measured by the compressibility (C) and flexibility (F) ratios. Those are the extensional stiffness and flexural stiffness of the medium relative to the lining.

Under this framework, Wang (1993) presented solutions for the diametric strain, the maximum thrust, and the bending moment under full-slip conditions, meaning normal force but no tangential shear force are present between the lining and the medium. For most cases the interface condition is between full slip and no slip. Slip interface can only happen in tunnels in soft soils or cases of severe seismic loading and full-slip assumption may lead to underestimation of the maximum thrust. As shown in Wang (1993) and NCHRP 611 (Anderson et al. 2008), for ground Poisson's ratio less than 0.5, thrusts decrease with decreasing compressibility ratio, but for Poisson's ratio of 0.5, the thrust response is independent of compressibility. The normalized lining distortion can be a plotted as function of flexibility ratio, as shown in Fig. 13. When $F < 1.0$, the lining is considered stiffer than the ground and deforms less than the ground. When $F > 1.0$, the lining is expected to deform more than the free field with an upper limit equal to the perforated ground case as described in Table 3. Penzien (2000) provides an analytical procedure to evaluate racking deformations of rectangular and circular tunnels. His solutions for ovaling deformations in terms of thrust and moment are very close to those of Wang (1993) for full-slip condition. However, the value of thrust for no-slip condition is much smaller in Penzien (2000) than in Wang (1993), differing in one order of magnitude.



Seismic Design of Tunnels, Fig. 13 Normalized lining deflection vs. flexibility ratio, full-slip interface, circular lining (Wang 1993)

Seismic Design of Tunnels, Table 3 Explanation of flexibility ratio

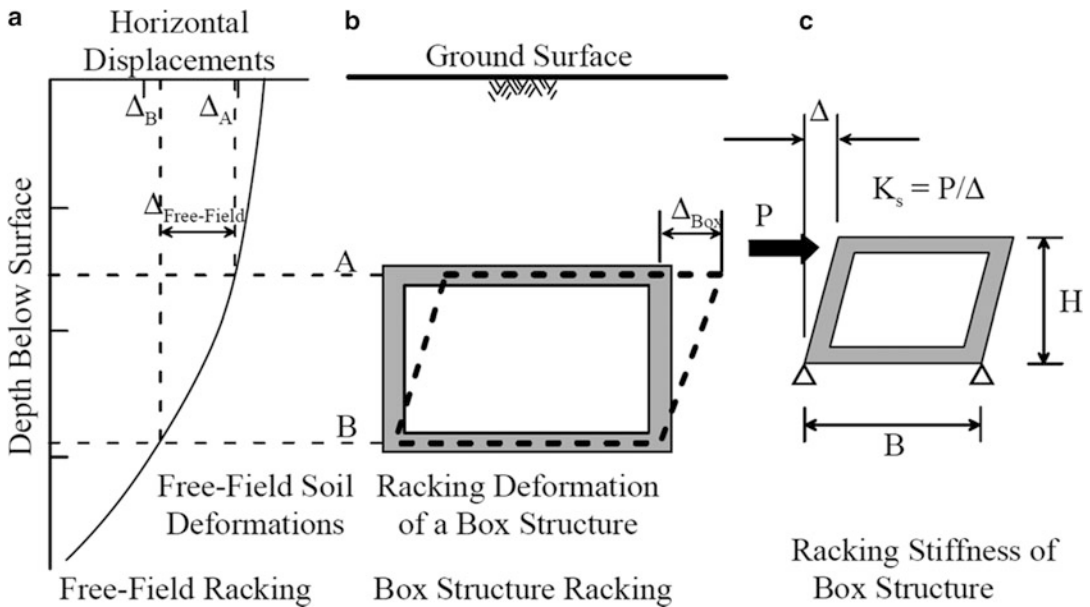
Flexibility ratio F	Meaning
$F \rightarrow 0$	The structure is rigid, so it will not rack regardless of the distortion of the ground
$F < 1$	The structure is stiff relative to the medium and will therefore deform less. Usually soft soil, and racking deformations are large
$F = 1$	The structure and medium have equal stiffness, so the structure will undergo approximately free-field distortions
$F > 1$	The structure racking stiffness is smaller than that of the soil. Usually stiff soil, and racking deformations are small
$F \rightarrow \infty$	The structure has no stiffness, so it will undergo deformations identical to the perforated ground

Hashash et al. (2005) compared the two analytical solutions to finite-element method numerical analyses to validate which of the solutions provide the correct solution to this problem.

The results from the numerical analyses agree with Wang (1993) solutions, highlighting the limitations of the other analytical solution. Sedarat et al. (2009) show that interface condition between the tunnel lining and the surrounding soil has an important impact on the computed thrust in the lining but limited impact on computed lining deformation.

Transverse: Racking Deformations of Rectangular Tunnels

Box-shaped cut-and-cover tunnels, common for transportation tunnels, have seismic characteristics different from circular tunnels because the walls and slabs of the box-shaped tunnels are stiffer. They are also often placed at shallower depths compared to circular tunnels. Therefore, it is important to carefully consider the soil-structure interaction due to increased stiffness and the increased seismic ground deformations at shallow depths (Hashash et al. 2001). Numerical analyses are often employed to compute the response of the tunnel structure to deformation of the surrounding soil.



Seismic Design of Tunnels, Fig. 14 Racking ratio, free-field and structure racking (Hashash et al. 2010)

Wang (1993) and Anderson et al. (2008) employed such techniques to develop relationships between racking ratio and flexibility (Fig. 14).

Huo et al. (2006) present an analytical solution for deep rectangular structures with a far-field shear stress. Complex variable theory and conformal mapping were used to develop the solution of structures in homogeneous, isotropic, elastic medium.

The relative stiffness between soil and structure, structure geometry, input earthquake motions, and tunnel embedment depth are factors that contribute to the soil-structure interaction effect. The most important of those is the flexibility ratio (F) that refers to the shear stiffness of the soil relative to the structure that replaces it (Wang 1993):

$$F = \frac{G_m \cdot B}{P \cdot H}$$

where G_m is the shear modulus of soil or rock medium, B is the width of the structure, P is the force required to cause a unit racking deflection of a rectangular frame structure, and H is the height of the tunnel as illustrated in Fig. 14. For a rectangular frame with an arbitrary configuration, the flexibility ratio can be determined by

performing a simple frame analysis; even for simple one-barrel frames, no computer analysis is needed.

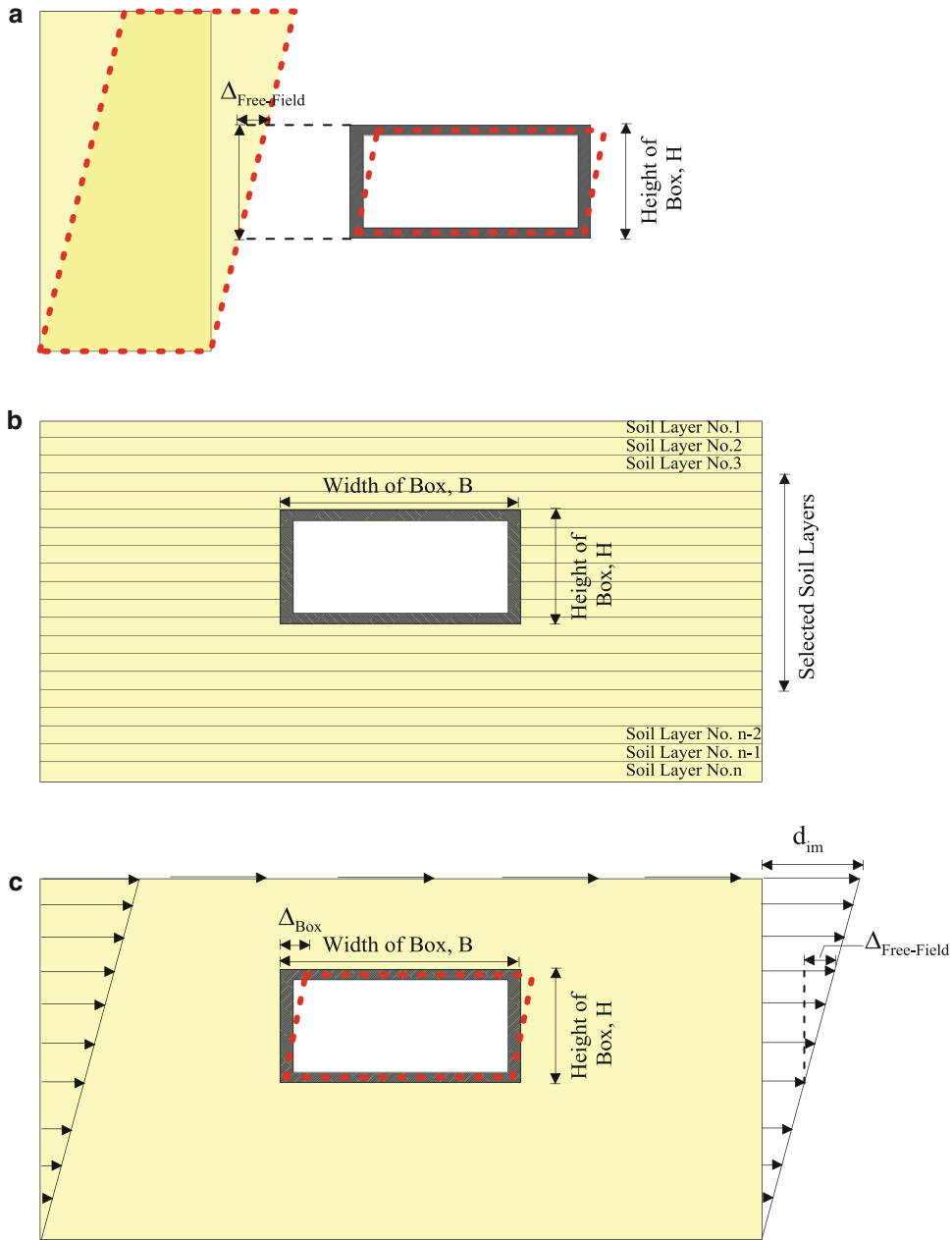
The racking ratio (R) is the ratio structure racking distortion to the free-field ground distortion:

$$R = \frac{\Delta_{\text{structure}}}{\Delta_{\text{free-field}}}$$

NCHRP 611 (Anderson et al. 2008) developed a relationship between the racking ratio (R) of rectangular conduits and the flexibility ratio (F), whereby

$$R = \frac{2F}{1 + F}$$

Hashash et al. (2010) describes the procedure to be used for performing 2D pseudo-static racking analysis of rectangular tunnels (Fig. 15). In this type of analysis, the soil is assumed to be massless and the section deforms in pure shear. In a first step, Fig. 15a, the free-field racking deformation time history is computed from 1D site response analyses using hazard-compatible ground motion time histories.



Seismic Design of Tunnels, Fig. 15 2D pseudo-static numerical analysis (Hashash et al. 2010)

To perform the 2D numerical analysis, the second step (Fig. 15b) is to define the elastic properties of a uniform soil medium as the average strain-compatible elastic properties of the selected soil layers (Anderson et al. 2008). As recommended in Hashash et al. (2010), layers

1–3 m above and below the structure should be included. Shear modulus values can be selected using the strain-compatible shear wave velocities from site response analysis in the selected layers, from which the average shear modulus over the selected layers can be calculated. In this step, the

structural member properties are needed: E (stiffness), I (moment of inertia), and A (cross-sectional area).

Finally, in a 2D numerical analysis, the lateral displacement (d_{im}) time histories obtained from 1D site response analysis are applied at the left, right and top boundaries of the model to impose the free-field racking calculated in the first step on the model, as shown in Fig. 15c. With the numerical analysis, the soil medium will transmit shearing deformations to the box structure and the box racking deformation time history can be obtained. This is used to obtain the racking ratio (R).

Two-dimensional pseudo-static numerical analyses can be a very useful tool, but they have some limitations. The ground surface shear displacements for shallow box structures cannot be transmitted uniformly. The model can be artificially extended to address this problem. Racking deformations are assumed to vary uniformly over the height of the box structure. The response of individual layers is not represented. This becomes a problem when layers with very different stiffness are part of the soil profile.

Dynamic Soil-Structure Interaction

The complex soil-structure interaction of underground structures during seismic loading can be simulated using numerical analysis tools which include lumped mass/stiffness methods and finite-element/difference methods.

Lumped mass/stiffness methods are useful to analyze the 3D behavior of a tunnel lining in a simplified manner. Many parameters for the springs that represent the structure stiffness and the soil stiffness must be defined to have a realistic model.

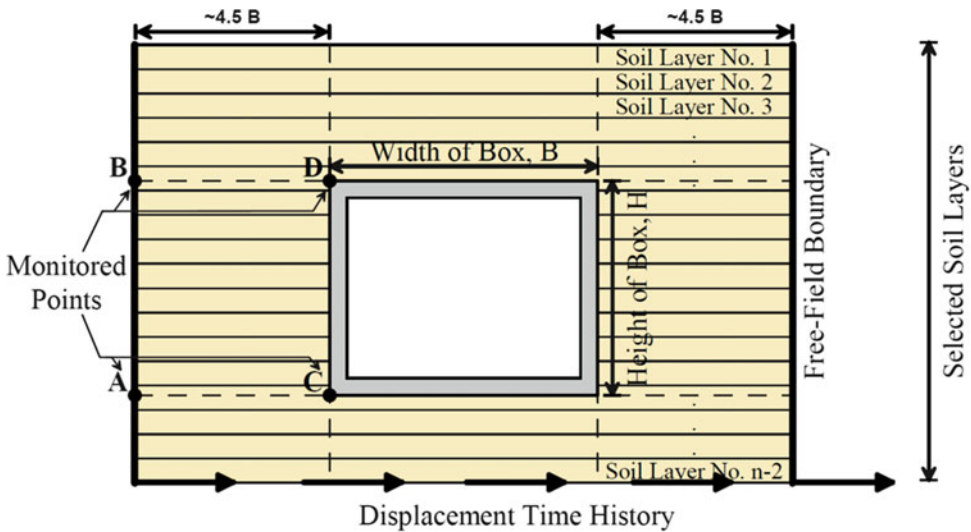
In finite-element/difference models, the tunnel structure is discretized and the soil surrounding the tunnel can be either discretized or represented by springs. 2D and 3D models can also be used to analyze the inelastic sections of the tunnel cross section. Discrete element models are useful when weak planes in the soil profile exist, since in this method the soil/rock mass is represented by an

assemblage of distinct blocks (rigid or deformable) with a prescribed constitutive relationship.

One of the advantages of these dynamic soil-structure interaction analyses is that the soil layers are modeled to reflect the idealized site stratigraphy; each soil layer can be either modeled as a linear elastic material with strain-compatible shear modulus and damping values or characterized via soil constitutive models that represent soil nonlinearity and hysteretic response at small strains. However, the use of the nonlinear constitutive models requires careful selection of input parameters and thus more advanced testing to define those input parameters. The nonlinear behavior and the frequency content of the free-field environment contribute to the structural racking behavior.

Hashash et al. (2010) provides a simplified 2D dynamic soil-structure interaction procedure that makes computational effort manageable for design purposes of transverse response of rectangular tunnels. The first step is to perform a 1D site response analysis to obtain the acceleration and displacement time history throughout the soil profile and then obtain the strain-compatible shear wave velocities and damping ratios for the 2D model layers.

The numerical analysis involves applying the displacement time history at the model base, then propagating the ground motion through the soil to simulate the soil-box interaction. The displacement time histories must be obtained at four monitored points, as shown in Fig. 16, to calculate the relative box displacement $\Delta_{box} = \max[abs(\delta_{h,C} - \delta_{h,D})]$ and the free-field relative displacement $\Delta_{ff} = \max[abs(\delta_{h,A} - \delta_{h,B})]$ to then calculate the racking ratio $R = \frac{\Delta_{box}}{\Delta_{ff}}$. The complete soil profile does not need to be included; a limited thickness of soil that captures the characteristics of wave propagation is sufficient. In the dynamic soil-structure analysis, the soil profile represented in the analysis reflects the actual soil profile, as there is no need to use an average soil layer needed in pseudo-static analyses.



Seismic Design of Tunnels, Fig. 16 2D dynamic soil-structure interaction (Hashash et al. 2010)

Comparison of Pseudo-static and Dynamic Racking Soil-Structure Interaction Analyses

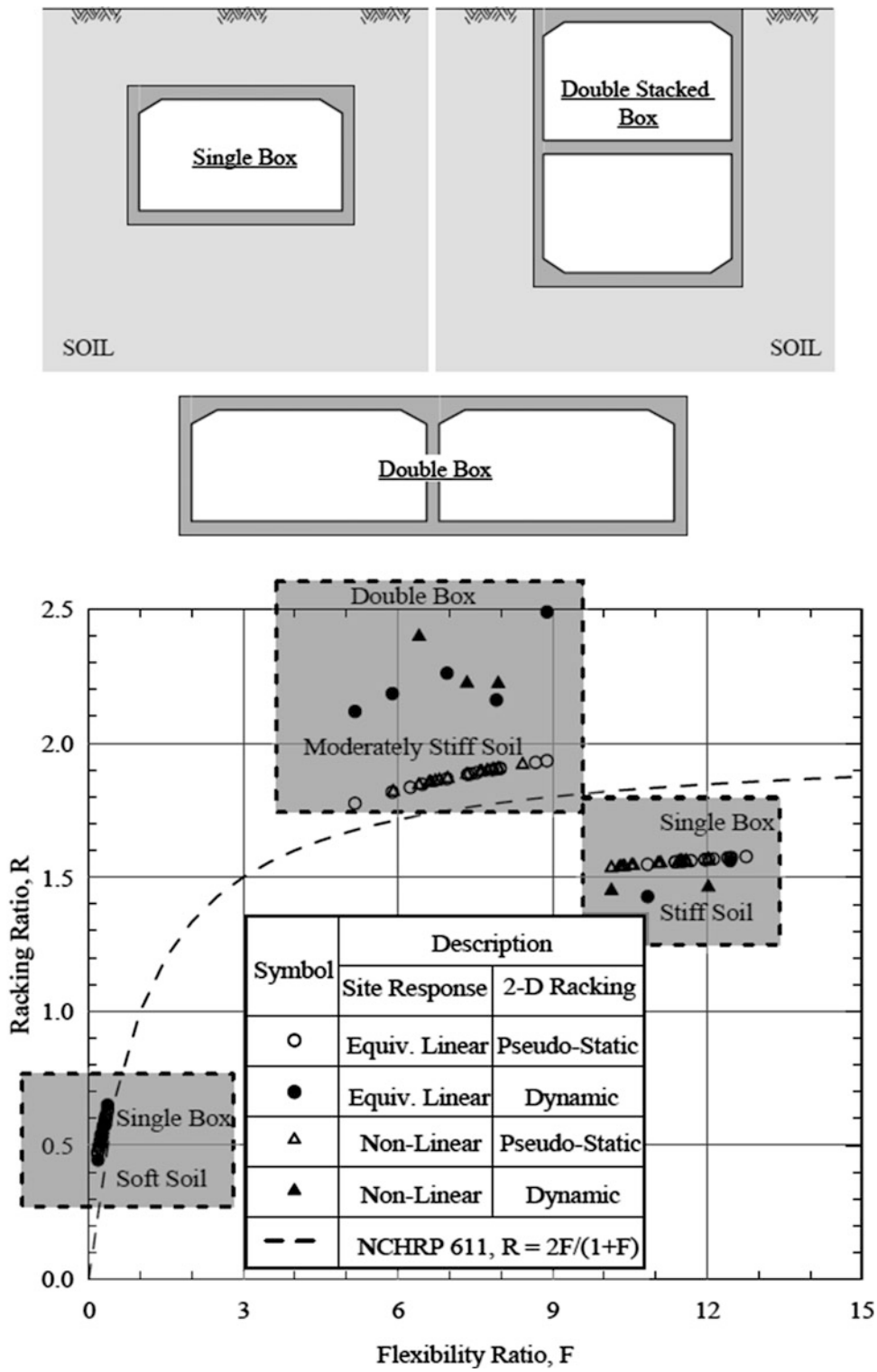
Hashash et al. (2010) performed a series of pseudo-static and dynamic soil-structure interaction analyses of single- and double-box structures in stiff and soft soil profiles using equivalent linear and nonlinear site response analysis with 14 ground motion time histories. The results are shown in Fig. 17. The results of the study found that for $F < 1$ the dynamic analysis and pseudo-static analysis appear to be quite similar and slightly above NCHRP 611 (Anderson et al. 2008). For $4 < F < 9$, the dynamic results show more scatter and higher racking ratios compared to pseudo-static results, both above NCHRP611 (Anderson et al. 2008). Dynamic analyses for $10 < F < 13$ show slightly lower racking ratios than pseudo-static analysis, both below NCHRP611 (Anderson et al. 2008). Some of the main conclusions from Hashash et al. (2010) are: (a) numerical approaches provide results and trends that are consistent with results obtained from simplified closed-form solutions, (b) there is a need to account for variability in the input ground motions and site response analysis methods as they affect the flexibility ratio (F), and (c) dynamic analyses must be

performed to verify and supplement the results of pseudo-static soil-structure interaction.

Performance Evaluation for Immersed-Tube Tunnels

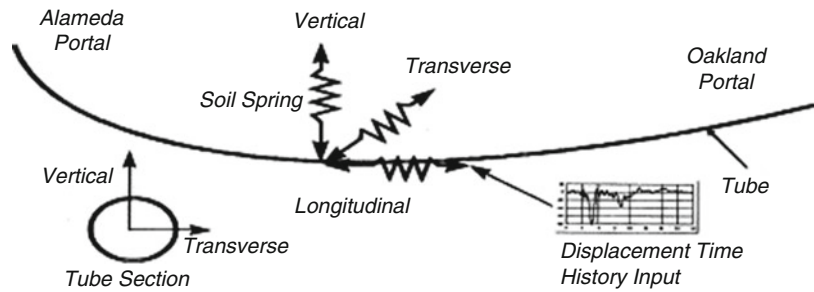
Pseudo-static longitudinal 3D models can be used to analyze axial and bending deformations in immersed-tube tunnels. In a lumped mass analysis approach, the tunnel lining is divided into individual segments with different masses and stiffnesses. The masses are then connected by springs that represent the axial, shear, and bending stiffness of the tunnel as shown in Fig. 18 (Hashash et al. 1998). Free-field displacement time histories that consider effects of wave passage/phase shift and incoherence are calculated at selected locations along the tunnel's length. The computed free-field displacement time histories are then applied at the ends of springs to represent soil-tunnel interaction in a quasi-static analysis. If a dynamic analysis is needed, appropriate damping factors need to be incorporated into the structure as well as springs to represent the soil.

Recent work from Anastasopoulos et al. (2007) focuses on nonlinear response to



Seismic Design of Tunnels, Fig. 17 Results from comparison of 2D pseudo-static and dynamic numerical analyses (Hashash et al. 2010)

Seismic Design of Tunnels, Fig. 18 3D model for global response of immersed-tube tunnel (Hashash et al. 1998)



strong seismic shaking of deep immersed tunnels (≈ 70 m). The free-field acceleration time histories are computed at the base of the tunnel through 1D wave propagation analysis using equivalent linear and nonlinear analyses. The computed free-field acceleration time histories are then imposed on the supports of the tunnel in the form of excitation. The tunnel is modeled as a multi-segment beam connected to the ground through calibrated springs, dashpots, and sliders. Wave passage effects are taken into account using Eurocode 8 (EC8 2002); however, the geometric incoherence was not considered because it did not make a difference when added to wave passage. The soil is assumed to be uniform along the tunnel.

A finite-element analysis is used to perform a nonlinear dynamic transient analysis of the tunnel. Tunnel segments are modeled using beam elements that take into account shear rigidity. The joints are modeled with nonlinear hyperelastic elements. The bored tunnels at the end of both segments are incorporated in the analysis as beams on viscoelastic foundation. Influence of segment length and joint properties was then investigated parametrically.

The results from Anastasopoulos et al. (2007) show that seismic response of immersed tunnel correlates better with PGV than PGA consistent with prior studies. There are some key conclusions applicable to immersed tunneling projects. First, a properly designed immersed tunnel can resist near-fault soil-amplified excitation with a PGA as large as 0.6 g and PGV as large as 80 m/s and containing long period pulses. Also, the net tension and excessive compression between segments can be avoided by suitable design of joint

gaskets and relatively small segment lengths. However, it is important to note that this study did not examine the heterogeneous nature of the soil conditions and ground motion incoherency and did not investigate the time-dependent stress relaxation on the rubber gasket or the effect of tectonic displacements from fault rupturing.

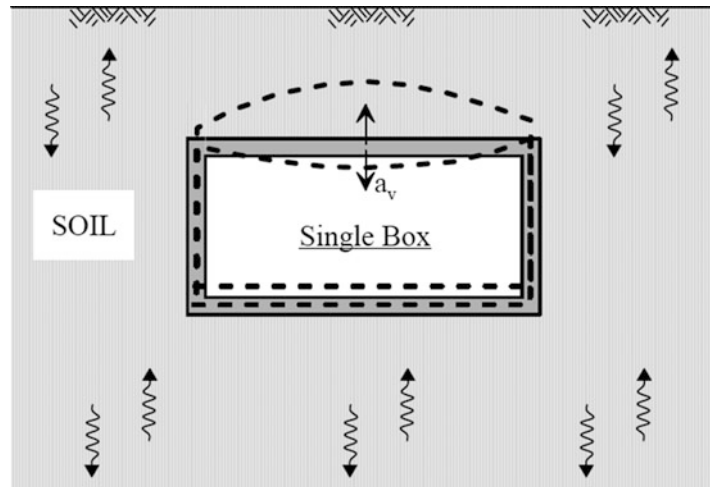
Additional Seismic Performance Issues

A number of additional items must be considered when evaluating the seismic performance of box structures. The detailed treatment of all these items is beyond the scope of this document, but a brief overview of some of them is presented.

Response to Vertical Ground Shaking

Significant vertical ground motions often due to near-fault effects can impose vertical loads on the roof of box structures. Figure 19 shows a schematic of the vertical loading on relatively shallow box structure (Hashash et al. 2010). Two types of analyses can be performed to assess the vertical loads. The vertical acceleration near the ground surface can be estimated as part of the seismic hazard analysis and then be used to compute pseudo-static inertial load of the soil on the tunnel roof. Alternatively, 2D dynamic soil-structure interaction analysis representing both the underground structure and the soil can be performed to compute the vertical inertial loading on the roof. There remain significant uncertainties in selecting appropriate soil properties for propagating vertical ground motion in such a model.

Seismic Design of Tunnels, Fig. 19 Vertical ground shaking effects (Hashash et al. 2010)



Interaction of Temporary and Permanent Structures

Braced excavations are used when space is needed to construct a shallow underground structure. Temporary excavations in highly seismic urban areas are also being seismically designed though for a lower level of seismic shaking than a permanent structure. Dynamic soil-structure interaction analyses are preferable to analyze racking in temporary structures, because it allows the dynamic load increments and levels of deformation on the temporary system to be estimated without significant simplifications.

It is customary to neglect the contribution of the temporary shoring wall on the permanent tunnel box. This is based on the assumption that the presence of the temporary walls will enhance the performance of the system and neglecting it is prudent. The effect of the presence of shoring walls on the seismic response of permanent box structures was studied in Hashash et al. (2010). For many projects only a few feet of the top of shoring is system is cut off after completion of the box construction while the rest of the wall remains in place. The study analyzed a single box surrounded by soft clay using both 2D pseudo-static and dynamic methodologies, as discussed earlier. Three cases were considered in Fig. 20a.

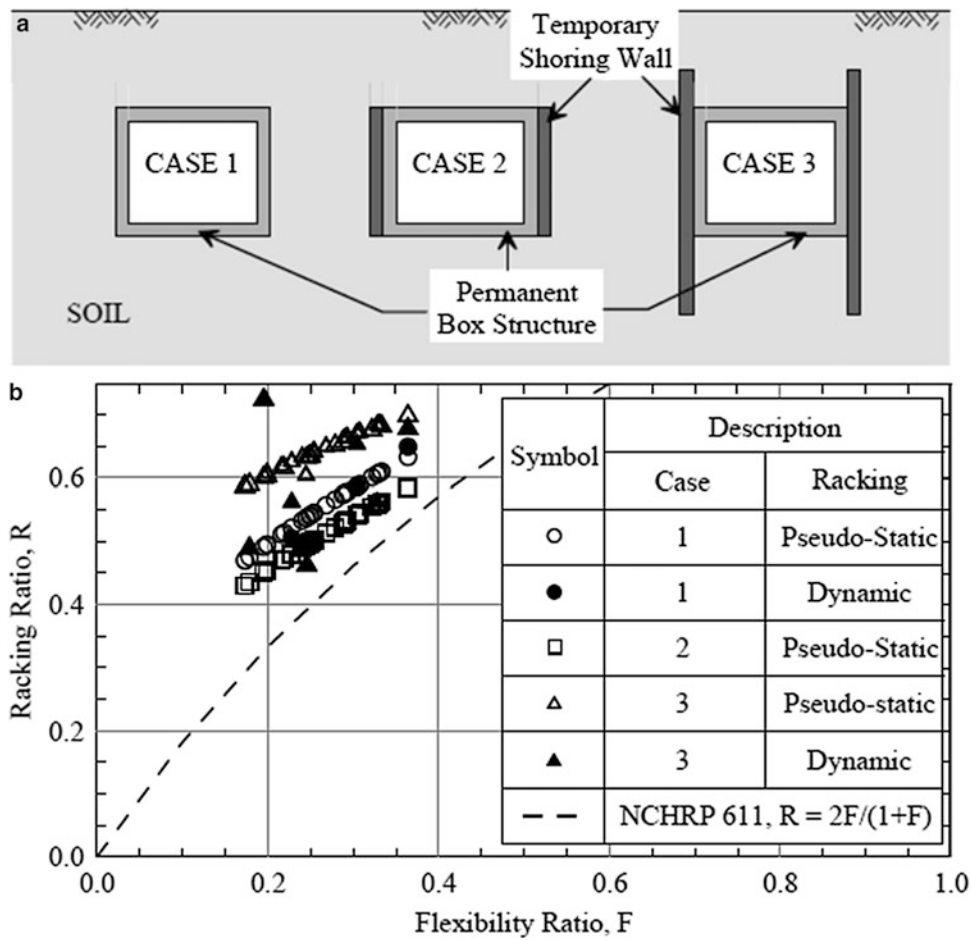
Figure 20b shows that modeling the shoring wall over the height of the structure only slightly

reduces the racking ratio of the tunnel structure, and hence neglecting the presence of the temporary walls is acceptable. An increase between 15 % and 20 % in the racking ratios is estimated when the temporary shoring wall is modeled above and below the permanent structure. The shoring walls act as extended wings, transferring soil loads to the structure from above and below the structure and therefore creating more racking deformations (Hashash et al. 2010). Therefore, neglecting the contribution of the shoring walls in this case will lead to underestimating of the racking deformation and is thus not advisable.

Impact of Superstructure and Adjacent Structures

Shallow underground structures for public transportation are a key component of sustainable cities. In dense urban environments, underground structures are built in close proximity to high-rise building foundations. Tall buildings have the potential to change the ground motions in the foundation soil and therefore transmit significant forces and base shear to adjacent underground structures. It is therefore important to evaluate the impact of transmitted forces from the superstructure to the soil to the underground structure under earthquake loading.

If aboveground structures are built over existing underground structures, the interaction between the structures must be evaluated.



Seismic Design of Tunnels, Fig. 20 Effect of shoring walls on permanent structures (Hashash et al. 2010)

The global system performance can be evaluated by means of a numerical analysis of seismic soil-box-structure interaction.

A related item is that the placement of wide box structures in deep excavations in soft soils may lead to significant changes in the soil properties beneath the structure. The influence of those changes on the long-term dynamic response of the structure should be taken into account (Hashash et al. 2010).

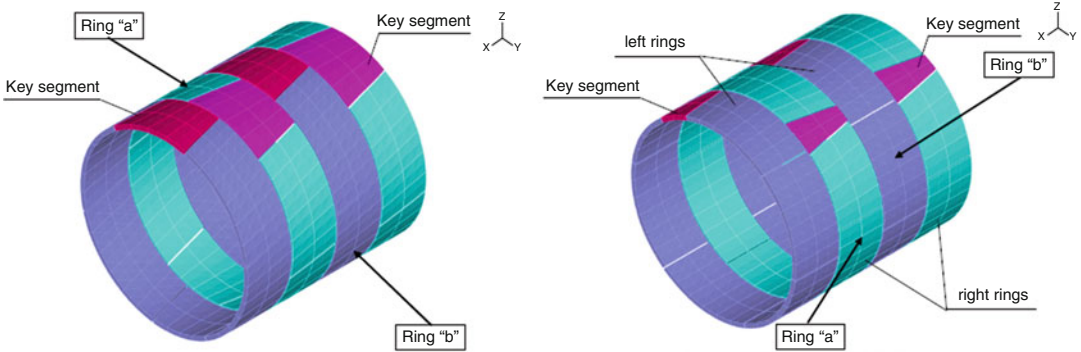
Tunnel Joints: Portals, Stations, and Tunnel Segments

Stiffness differences in the tunnel structure may generate differential movements and stress concentration. Some examples include tunnel-building or tunnel-station connections, tunnel

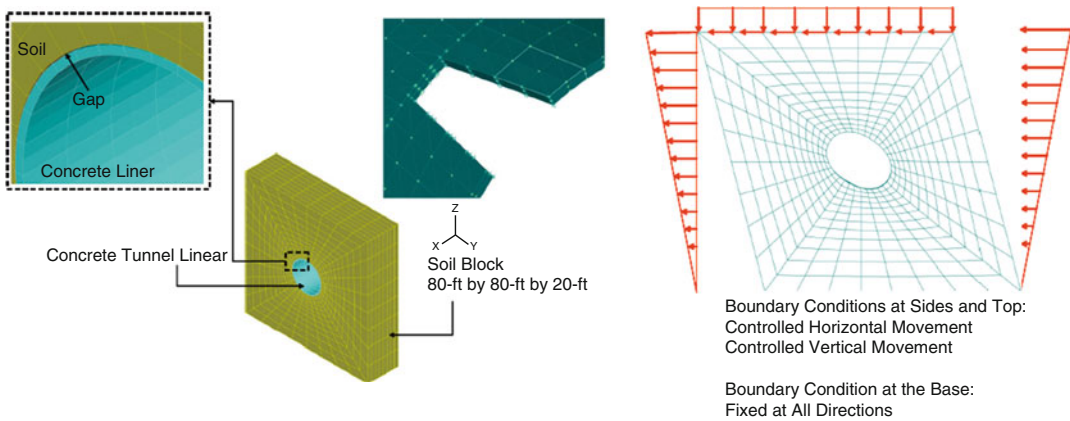
junctions, tunnels crossing distinct geologic media, and local restraints of any type (Hashash et al. 2001).

The most common solution to these interface problems is the use of flexible joints. The first step in the design process is to use closed-form solutions or numerical analyses to determine the required allowable rotation and differential movements in the longitudinal, transverse directions. The joints also must be designed to support static and dynamic earth and water loads before and during the earthquake while remaining watertight. If a continuous design is implemented, very large forces and moments are likely to be generated (Hashash et al. 2001).

Tunnel portals and vent structures have part of their structure above ground; therefore, the



Seismic Design of Tunnels, Fig. 21 FEM lining models for trapezoidal and rectangular geometric configurations boundary conditions (Kramer et al. 2007)



Seismic Design of Tunnels, Fig. 22 Details of mesh development and boundary conditions (Kramer et al. 2007)

seismic design should account for inertial effects or potential pounding. It is always preferable that those structures are isolated from the tunnel through flexible joints (Hashash et al. 2001).

For soil-rock transitions, it is recommended that the tunnel not be cast directly against the rock. Instead, there should be some over-excavation filled with soil or aggregate. If that is not possible in bored tunnels, then a flexible lining can be installed (Hashash et al. 2001).

Precast Tunnel Lining

When a tunneling machine is used to excavate a tunnel, the lining usually is erected in segments. Single pass precast concrete segmental lining systems are often employed. The segment joint

connections must therefore be designed to accommodate anticipated deformations. The joint behavior can be kept elastic or allow inelasticity if proper detailing is performed. Numerical analyses are often used to evaluate the seismic performance of segmental lining. Kramer et al. (2007) performed 3D finite-element analyses, Fig. 21, to compute radial and circumferential joint behavior during seismic ovaling and wave propagation for the Silicon Valley Rapid Transit (SVRT) Project in San Jose, California. These analyses incorporate inelastic constitutive soil behavior, cracked concrete properties, and no-tension, frictional segment joint surfaces (Fig. 22). Kramer et al. (2007) provide guidelines for precast lining design in high seismicity zones.

Seismic Retrofit of Existing Facilities

When considering seismic retrofit of an existing tunnel structure, the retrofit strategy depends on the structure damage mode. If there is a gross stability problem, the seismic retrofit strategies involve strengthening the structure itself or the surrounding geologic materials.

One concern for a circular tunnel is the contact quality between the liner and the surrounding geologic media. Strengthening this interface includes replacing the lining, increasing the lining thickness by adding reinforced concrete, or adding reinforcing bars or internal steel liner. Increasing lining thickness is not always a good solution, as it tends to attract more load. Methods that increase ductility as well as strength are more effective (Power et al. 1996). Adding circumferential joints along the tunnel axis can also reduce the stresses and strains induced by longitudinal propagating waves. However, the value of adding joints must be weighed against the expected performance of the liner without joints. It is important to verify in the retrofit design that there will be no additional water leakage and that the joints will not become weak spots (Power et al. 1996).

In cut-and-cover structures, some seismic retrofit strategies include increasing ductility of reinforced concrete lining, adding confinement at existing columns, and adding steel plate jackets at joints.

Summary

The performance of the underground facilities during recent seismic events showed that the underground structures have suffered appreciably less damage than surface structures. However, the failure of even one of these underground structures can be detrimental to the proper post-earthquake operation of a tunnel network in urban areas. Thus, evaluation of underground structures to seismic shaking is necessary.

A performance-based approach is recommended for seismic evaluation of underground structures. Underground structures seismic response is controlled by the deformation of the soil or rock medium in which they are embedded and their

seismic evaluation is different from aboveground structures.

Seismic evaluation of underground structures includes evaluation for permanent and transient ground deformations. Pseudo-static and dynamic analysis approaches can be used to estimate the deformation of underground structures due to transient ground deformations and are presented.

Additional seismic design issues were discussed, including vertical ground shaking and response, interaction of temporary and permanent structures, impact of superstructure and adjacent structures, tunnel joints, and seismic retrofit of existing facilities.

Cross-References

- [Conditional Spectra](#)
- [Dynamic Soil Properties: In Situ Characterization Using Penetration Tests](#)
- [Liquefaction: Countermeasures to Mitigate Risk](#)
- [Nonlinear Dynamic Seismic Analysis](#)
- [Nonlinear Finite Element Analysis](#)
- [Nonlinear Seismic Ground Response Analysis of Local Site Effects with Three-dimensional High-fidelity Model](#)
- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- [Probabilistic Seismic Hazard Models](#)
- [Seismic Actions Due to Near-Fault Ground Motion](#)
- [Site Response for Seismic Hazard Assessment](#)
- [Site Response: 1-D Time Domain Analyses](#)
- [Site Response: Comparison Between Theory and Observation](#)
- [Soil-Structure Interaction](#)

References

- Anastasopoulos I, Gerolymos N, Drosos V, Kourkoulis R, Georgarakos T, Gazetas G (2007) Nonlinear response of deep immersed tunnel to strong seismic shaking. *J Geotech Geoenviron Eng* 133(9): 1067–1090. doi:10.1061/(ASCE)1090-0241(2007)133:9(1067)

- Anderson DG, Martin GR, Lam I, Wang JN (2008) Seismic analysis and design of retaining walls, buried structures, slopes and embankments, vol NCHRP 611: National Cooperative Highway Research Program, Transportation Research Board
- Ashford SA, Boulanger RW, Donahue JL, Stewart JP (2011) Geotechnical quick report on the Kanto Plain Region during the March 11, 2011, Off Pacific Coast of Tohoku Earthquake, Japan GEER Association report no. GEER-025a: Geotechnical Extreme Events Reconnaissance (GEER)
- Coppersmith KJ, Youngs RR (2000) Data needs for probabilistic fault displacement hazard analysis. In: Cello G, Tondi E (eds) *The resolution of geological analysis and models for earthquake*. Pergamon Press, Oxford
- Cording EJ, Martin G, Parker H (2011) Century city area tunneling safety and fault investigations: tunnel advisory panel. Los Angeles County Metropolitan Transportation Authority, Los Angeles
- EC8 (2002) Eurocode 8: design of structures for earthquake resistance. European Committee for Standardization (CEN), Brussels
- Elnashai AS, Gencturk B, Kwon O-S, Al-Qadi IL, Hashash Y, Roesler JR, et al. (2010) The Maule (Chile) earthquake of February 27, 2010 consequence assessment and case studies: Mid-America Earthquake Center
- Hamburger RO, Hooper JD (2011) Performance-based seismic design. *Mod Steel Constr* 4:36–39
- Hashash YMA, Park D (2001) Non-linear one-dimensional seismic ground motion propagation in the Mississippi embayment. *Eng Geol* 62(1–3):185–206
- Hashash YMA, Tseng WS, Krमतot A (1998) Seismic soil-structure interaction analysis for immersed tube tunnels retrofit. Paper presented at the proceedings of the 1998 conference on geotechnical earthquake engineering and soil dynamics III. Part 2 (of 2), Seattle, 3–6 Aug 1998
- Hashash YMA, Hook JJ, Schmidt B, Yao JI-C (2001) Seismic design and analysis of underground structures. *Tunn Undergr Space Technol* 16:247–293
- Hashash YMA, Park D, Yao JIC (2005) Ovaling deformations of circular tunnels under seismic loading, an update on seismic design and analysis of underground structures. *Tunn Undergr Space Technol* 20(5):435–441. doi:10.1016/j.tust.2005.02.004
- Hashash YMA, Karina K, Koutsoftas D, O'Riordan N (2010) Seismic design considerations for underground box structures. Paper presented at the 2010 earth retention conference – earth retention conference 3, Bellevue, 1–4 Aug 2010
- Hashash YMA, Groholski D, Musgrove M, Park D, Phillips C, Tsai C-C (2011) DEEPSOIL V5.0, manual and tutorial. Board of Trustees of University of Illinois at Urbana-Champaign, Urbana
- Huo H, Bobet A, Fernández G, Ramírez J (2006) Analytical solution for deep rectangular structures subjected to far-field shear stresses. *Tunn Undergr Space Technol* 21(6):613–625
- Iida H, Hiroto T, Yoshida N, Iwafuji M (1996) Damage to the Daikai subway station. *Soils and Foundations* 283–300. Special issue on geotechnical aspects of the January 17, 1995 Hyogoken-Nambu Earthquake, Tokyo, Japan
- Kramer GJE, Sedarat H, Kozak A, Liu A, Chai J (2007) Seismic response of precast tunnel linings. Paper presented at the proceedings of the 2007 rapid excavation and tunneling conference, Toronto. Publisher and copyright holder: Society for Mining, Metallurgy & Exploration (SME). www.smenet.org
- Lysmer J, Udaka T, Tsai CF, Seed HB (1975) FLUSH: a computer program for approximate 3-D analysis of soil-structure interaction problems. Earthquake Engineering Research Center, Berkeley
- Matasovic N (1993) Seismic response of composite horizontally-layered soil deposits. PhD thesis, University of California, Los Angeles. Retrieved from <http://www.lib.umi.com/pqdd2/search/do?query=au%28Matasovic%2C%20Neven%2C%29%20and%20da%281993%29>
- Newmark NM (1967) Problems in wave propagation in soil and rock. Paper presented at the international symposium on wave propagation and dynamic properties of earth materials, Albuquerque, NM
- Peck RB, Hendron AJ, Mohraz B (1972) State of the art in soft ground tunneling. Paper presented at the rapid excavation and tunneling conference, New York
- Penzien J (2000) Seismically induced racking of tunnel linings. *Earthq Eng Struct Dyn* 29(5):683–691. doi:10.1002/(sici)1096-9845(200005)29:5<683::aid-eqe932>3.0.co;2-1
- Power MS, Rosidi D, Kaneshiro J (1996) Volume III Strawman: screening, evaluation, and retrofit design of tunnels. National Center for Earthquake Engineering Research, Buffalo
- Schmidt B, Hashash YMA (1999) Preventing tunnel flotation due to liquefaction. Paper presented at the proceedings of the second international conference on earthquake geotechnical engineering, Lisbon
- Schnabel PB, Lysmer JL, Seed HB (1972) SHAKE: a computer program for earthquake response analysis of horizontally layered sites. Earthquake Engineering Research Center, Berkeley
- Sedarat H, Kozak A, Hashash YMA, Shamsabadi A, Krमतot A (2009) Contact interface in seismic analysis of circular tunnels. *Tunn Undergr Space Technol* 24:482–490
- St John CM, Zahrah TF (1987) Aseismic design of underground structures. *Tunn Undergr Space Technol* 2(2):165–197
- Wang JN (1993) Seismic design of tunnels: a state-of-the-art approach, Monograph 7. Parsons Brinckerhoff Quade & Douglas, New York
- Wang WL, Wang TT, Su JJ, Lin CH, Seng CR, Huang TH (2001) Assessment of damage in mountain tunnels due

to the Taiwan Chi-Chi earthquake. *Tunn Undergr Space Technol* 16(3):133–150

Youngs RR, Arabasz WJ, Anderson RE, Ramelli AR, Ake JP, Slemmons DB, et al. (2003) A methodology for probabilistic fault displacement hazard analysis (PFDHA). *Earthq Spectra* 19(1):191–219. doi:10.1193/1.1542891

Seismic Design of Waste Containment Systems

Edward Kavazanjian Jr.
School of Sustainable Engineering and the Built Environment, Arizona State University,
Tempe, AZ, USA

Synonyms

Equivalent linear analysis; Geosynthetic; Landfill; Performance-based design; Site response; Solid waste; Waste containment systems

Introduction

Maintaining the integrity of waste containment systems subject to seismic loading is not only important with respect to protecting human health and the environment from wastes and waste by-products, it is also important with respect to the availability of facilities essential for post-earthquake recovery, i.e., of landfills for proper disposal of the large volumes of earthquake-generated waste. Important considerations in the seismic design of waste containment systems include the response of the waste mass itself to seismic excitation, the stability of the waste mass (including its foundation), and the integrity of the engineered components of the waste containment system. Engineered components of waste containment systems include base and side-slope liner (barrier) systems, leachate and gas collection and removal systems, and final cover systems. Liner and cover systems include both natural soil and geosynthetic elements that serve as barriers to advective transport

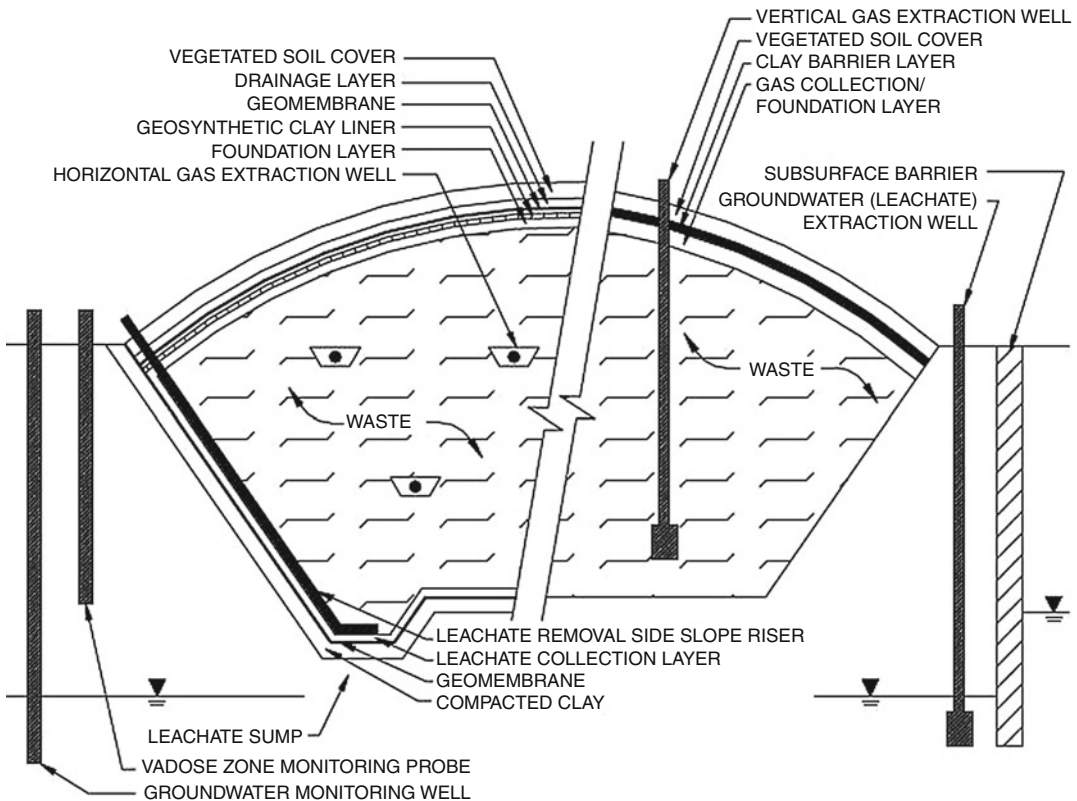
of liquids and gas, as drainage layers for landfill leachate and surface water infiltration, and as protection for other components of the waste containment system. Geosynthetic waste containment system elements, including geomembrane and geosynthetic clay liner (GCL) barrier layers, geonet drainage layers, and geotextile filter, separation, and cushion (geomembrane protection) layers, are typically only found in containment systems constructed within the last 25 years, appearing with increasing frequency over that time span.

Figure 1 illustrates the components of municipal solid waste (MSW) landfill waste containment systems. MSW landfills are by far the most common type of engineered waste containment system, followed (with respect to frequency of occurrence) by cover systems (caps) for uncontrolled dumps and hazardous waste sites and then by hazardous waste landfills. Caps and containment systems for hazardous waste landfills employ similar elements to MSW landfills. The performance of ancillary facilities at a waste disposal site subject to seismic loading, including leachate and gas treatment facilities, surface water control systems, access roadways, and landfill monitoring systems, is also an important consideration but will not be addressed herein.

Following a brief review of the performance of landfills in earthquakes, this entry summarizes the state of knowledge on waste mass seismic response and the properties of the waste and geosynthetic materials relevant to seismic analysis and design. Current approaches to the seismic design of landfills are then discussed, culminating in a discussion of the value of performance-based seismic design for these important facilities.

Seismic Performance of Waste Containment Systems

In general, the seismic performance of waste containment systems subject to strong ground motion from earthquakes has been acceptable, i.e., has not resulted in a harmful discharge of waste or waste by-products to the environment.



Seismic Design of Waste Containment Systems, Fig. 1 Containment system components for modern (*left*) and older (*right*) landfills (Kavazanjian et al. 1998)

In the epicentral region of the 1989 M 6.9 Loma Prieta earthquake, several MSW landfills with 2H:1V slopes rising up to 60 m above grade withstood earthquake motions with peak ground accelerations estimated to be on the order of 0.6 g with only minor cracking in the soil cover, e.g., observed cracks on the order of a 50 mm or less (Buranek and Prasad 1991; Johnson et al. 1991; Orr and Finch 1990). However, there were observations of more significant cracking in the soil cover on the very steep north slope of the Operating Industries, Inc. (OII) Landfill (average inclination, 1.5H:1V; maximum inclination, 1.3H:1V) in the 1987 M 6.1 Whittier Narrows earthquake and the 1994 M 6.7 Northridge earthquake when they were subject to peak ground accelerations estimated to be less than half the 0.6 g level experienced by landfills in the Loma Prieta event (Kavazanjian et al. 2013b; Matasovic et al. 1998). Furthermore, neither the

OII Landfill nor any of the landfills subject to strong ground shaking in the Loma Prieta earthquake employed geosynthetic containment system elements (Kavazanjian et al. 2013b; Matasovic et al. 1998). Tensile rupture of a geomembrane barrier layer was observed at two separate locations at the Chiquita Canyon Landfill in the Northridge earthquake, one location at the crest of a lined side slope and a second location near the crest of a lined side slope (EMCON 1994). Figure 2a shows the cover soil cracking observed at the OII Landfill, and Fig. 2b shows one of the geomembrane tears observed at the Chiquita Canyon Landfill following the Northridge earthquake.

Geosynthetic waste containment system elements are of particular concern with respect to seismic loading because they create the potential for planes of weakness along which slippage and instability can occur (due to a relatively low



Seismic Design of Waste Containment Systems, Fig. 2 Damage to landfills observed following the Northridge earthquake: (a) cracking in soil cover on the

benches of the OII Landfill (Courtesy of Raymond B. Seed); (b) tear in the side-slope liner at the Chiquita Canyon Landfill (Courtesy of Robert M. Koerner)

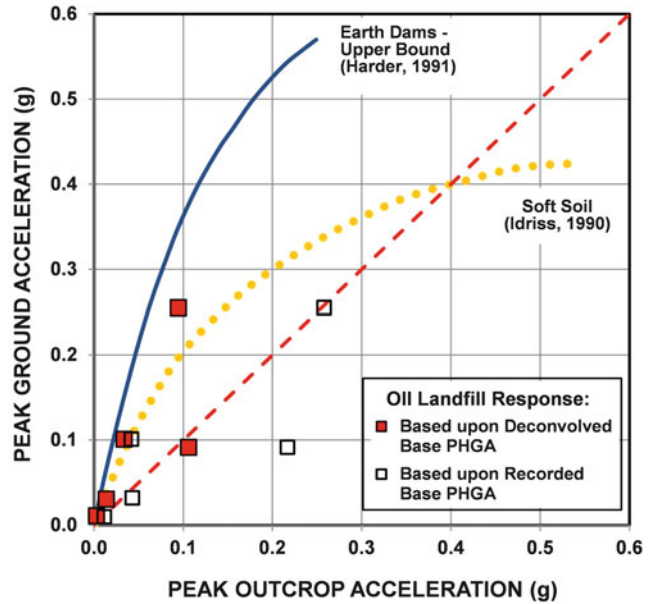
interface or in-plane shear resistance) and because of their potential for tensile rupture which could result in breaching of barrier, filter, and protective layers and disruption of drainage layers. Stability failures of landfills have occurred under static loading due to the low interface shear strength of a geosynthetic waste containment system element (Koerner and Soong 2000; Mitchell et al. 1990; Seed et al. 1990) and the tensile rupture of the side-slope geomembranes at the Chiquita Canyon Landfill were cited previously. Furthermore, there are no case histories of cover systems with geomembrane elements subject to free-field ground motions in excess of 0.2 g, and cover systems are of particular concern due to the potential for amplification of seismic motions by the waste mass. There is also a potential for hidden damage to containment system elements beneath the surface of the landfill, where damage cannot be directly observed or readily detected by the monitoring systems employed in current practice.

Waste Mass Seismic Response

Overview: The seismic response of the waste mass is an important consideration in waste containment system seismic design as it controls the loading on the containment system elements as well as the global stability of the waste mass. Because there is often a significant impedance contrast between the waste mass and the underlying foundation material (soil or rock), landfill seismic response is susceptible to amplification of free-field motions. Amplification of earthquake ground motions is reported to have occurred at the one landfill at which strong ground motions have been recorded to date, OII Landfill (Kavazanjian et al. 2013b), and Bray and Rathje (1998) have conducted analyses suggesting that the amplification potential of municipal solid waste landfills is similar to the amplification reported by Harder (1991) between the base and crest of earth dams in the transverse direction. Figure 3 compares the amplification of peak ground acceleration from the base to the

Seismic Design of Waste Containment Systems,

Fig. 3 Amplification of peak ground acceleration at the OII Landfill in the Northridge earthquake (Kavazanjian et al. 2013b)



crest of the OII Landfill (as determined by back analysis) to the amplification of the transverse peak ground acceleration from the base to the crest of earth dams as reported by Harder (1991).

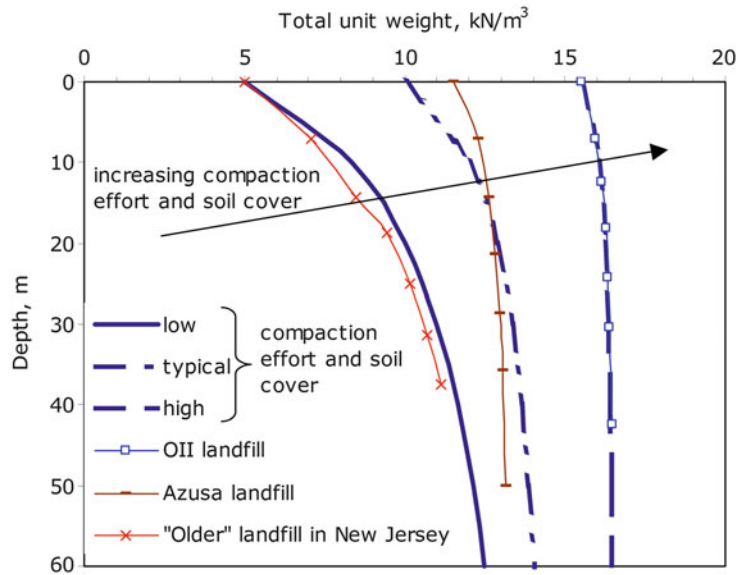
Waste Properties: Seismic response of a waste mass is controlled by waste mass properties, the waste thickness, and the impedance contrast between the base of the landfill and the waste mass (which depends upon the properties of the waste mass and the underlying foundation material) as well as the characteristics of the free-field ground motion. For landfills with geosynthetic elements in the liner and cover system, slip along the interface of a geosynthetic element may also affect the seismic response. However, slip at geosynthetic interfaces is usually ignored in practice when evaluating landfill seismic response. Analyses in which the potential for slip at the interface is ignored are termed “decoupled” analyses as landfill seismic response is decoupled from the geosynthetic interface behavior and potential displacement at the interface. A decoupled analysis is generally considered conservative with respect to predicting landfill seismic performance, though there are relatively rare situations where this has been shown not to be the case (Bray and Rathje 1998).

Most seismic response analyses in landfill practice are conducted using the equivalent linear method. Waste mass properties of importance in an equivalent linear seismic response analysis include the waste total unit weight (or mass density), the small-strain shear modulus, and equivalent linear shear modulus reduction and damping curves. Poisson’s ratio may also be of interest if two-dimensional equivalent linear response analyses are to be conducted. Small strain shear modulus is generally established based upon shear wave velocity and unit weight or mass density. Waste mass shear strength and the interface shear strength of geosynthetic elements (or in the case of a geosynthetic clay liner, the in-plane shear strength) are essential properties in seismic analysis of waste containment systems for waste mass stability and cover system assessments. The properties of natural soil materials used in landfill construction are also important in a landfill seismic stability assessment but are not discussed herein.

Figure 4, from Zekkos et al. (2006), provides a family of typical unit weight profiles for MSW landfills that depend upon the amount of cover soil and compaction effort provided by the landfill operator during waste placement. Due to difficulties in measuring the waste unit weight in the

Seismic Design of Waste Containment Systems,

Fig. 4 Typical MSW total unit weight vs. depth profiles (Zekkos et al. 2006)



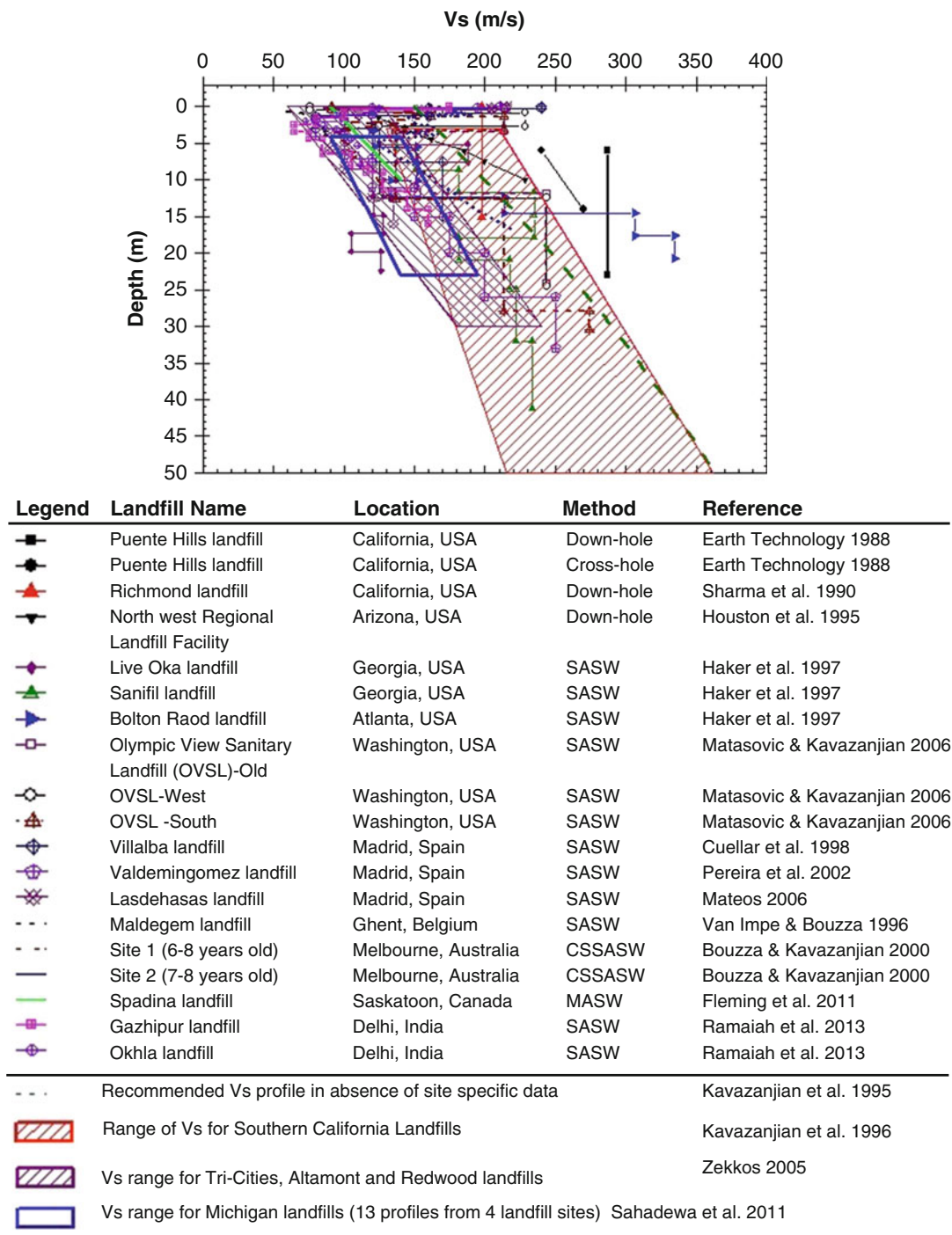
field, Fig. 4 or some other relationship for typical unit values is usually used in landfill seismic analysis and design rather than direct measurement, even for existing facilities.

Shear wave velocity and the small-strain modulus of MSW have been measured in the field and in the laboratory. Zekkos et al. (2008) and Yuan et al. (2011) report shear wave velocity and small-strain modulus values measured in laboratory testing of MSW reconstituted at three different ratios of refuse (MSW particles >20 mm, i.e., retained on a 20-mm sieve) to soil-sized (MSW particles <20 mm, i.e., passing a 20-mm sieve) material (note that these investigators used the same MSW and same composition ratios but different laboratory devices). However, unlike unit weight, waste mass shear wave velocity is readily measured in the field, e.g., using noninvasive surface wave techniques. Field data is more reliable than laboratory testing on reconstituted specimens, and case histories of site-specific shear wave velocity measurements for landfills are being reported in the literature with increasing frequency. Figure 5 provides a summary of shear wave velocity profiles for MSW landfills worldwide as compiled by Ramaiah (2013).

Equivalent linear shear modulus reduction and damping curves for MSW are available from

three sources: (1) from back analysis of strong ground motion records from the OII Landfill alone (e.g., Idriss et al. 1995), in some cases supplemented with cyclic laboratory testing of waste (Matasovic and Kavazanjian 1998); (2) from large-diameter cyclic triaxial testing by Zekkos et al. (2008) on reconstituted waste specimens from the Tri Cities Landfill for three different composition ratios; and (3) from large-scale laboratory simple shear testing by Yuan et al. (2011) on reconstituted waste specimens from the Tri Cities Landfill for the same three composition ratios employed by Zekkos et al. (2008). Figure 6 presents the equivalent linear shear modulus reduction (Fig. 6a) and damping (Fig. 6b) curves from Zekkos et al. (2008) from large-scale triaxial testing, termed the UCB data, and from Yuan et al. (2011) from large-scale simple shear testing, termed the ASU data. These two data sets are relatively consistent with each other and with the OII field data and represent the best available information on these property relationship.

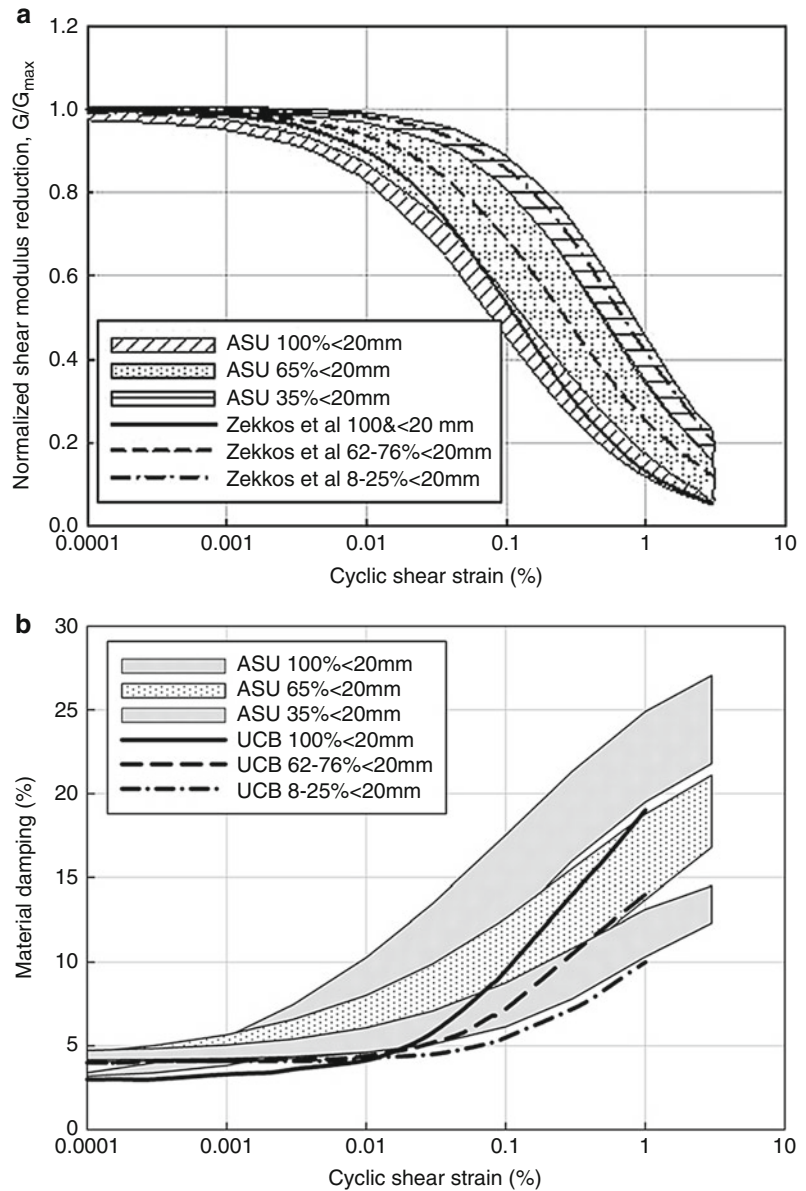
Poisson's ratio values for MSW have been derived from field measurements of shear and compressional wave velocity at the OII Landfill (Matasovic and Kavazanjian 1998) and from laboratory measurements of axial and radial strain (Zekkos 2005). Poisson's ratio can also be



Seismic Design of Waste Containment Systems, Fig. 5 Shear wave velocity at MSW landfills worldwide (Ramaiah 2013)

Seismic Design of Waste Containment Systems,

Fig. 6 Equivalent linear property relationships for waste specimens reconstituted at different composition ratios (Yuan et al. 2011): (a) shear modulus reduction; (b) damping



inferred from the field measurements of the lateral earth pressure at rest in MSW reported by Dixon et al. (1999). The data on Poisson's ratio from these studies is scattered, both within and among the studies, but it appears that values on the order of 0.2–0.4 are appropriate for unsaturated MSW.

Due to difficulties associated with conventional laboratory and field testing, studies of MSW shear strength have employed back

analyses of landfill failures under static loading to develop shear strength envelopes. Kavazanjian et al. (1996) employed back analysis of both landfill failures and steep stable landfill slopes to develop a bilinear strength envelope in which MSW was assigned a cohesion of 25 kPa at low normal stresses and a friction angle of 33° at higher confining pressures. Based upon back analyses of failed landfill slopes, Eid et al. (2000) proposed that typical MSW shear strength could

be characterized by a linear failure envelope with a friction angle of 35° and a cohesion intercept and that the lower- and upper-bound shear strength should be characterized using zero and 50-kPa cohesion, respectively. Based upon laboratory triaxial and direct testing on relatively large-sized specimens (e.g., 150 mm-diameter triaxial specimens), Bray et al. (2009) suggested that MSW shear strength was dependent on the normal stress and characterized by a cohesion of 15 kPa and a friction angle that was equal to 36° at a normal stress less than or equal to one atmosphere and that decreased by 5° for every log cycle of normal stress. Both Bray et al. (2009) and Kavazanjian et al. (2013b) have noted the anisotropic structure and shear strength of MSW, with the strengths reported above applicable to the horizontal plane, the weakest plane in a waste mass due to the preferred alignment of the long axes of waste constituents in this direction, and with substantially greater shear strength on non-horizontal planes that cut across the long axes of the constituents. Based upon analysis of landfill performance in the Northridge earthquake, Augello et al. (1995) suggested that the shear strength of MSW subject to seismic loading was greater than the static strength. Bray et al. (2009) suggest that the dynamic shear strength of MSW is at least 20 % greater than the static shear strength. However, in most seismic evaluations, the static shear strength has been found to be sufficient to demonstrate landfill seismic stability.

The data on material properties for waste materials other than typical MSW is rather sparse. Hazardous waste in modern landfills is generally stabilized and containerized prior to placement in a landfill. The containers are then stacked and backfilled with soil. In older hazardous waste landfills, the waste may have been containerized, typically in partially filled steel drums, without stabilization. Cone penetrometer testing at a 20–30-year-old hazardous waste landfill where the waste was placed in partially filled steel drums without stabilization showed no evidence that the drums were still intact or of voids (as would be expected at the top of a partially filled drum) (Matasovic et al. 2005).

Furthermore, hazardous waste disposed of in landfills is often a hazardous material (e.g., pesticide residue, organic solvents) or mixed waste (e.g., potentially radioactive construction and demolition debris) in a soil matrix. Hence, it is common to assume that the properties of the waste mass at a hazardous waste site are the same as the properties of the backfill soil (or of the native soil at a hazardous material spill site) (Kavazanjian and Matasovic 2001).

Geosynthetic Properties: The interface shear strength of geosynthetic materials and the in-plane internal strength of GCLs can vary widely, depending upon the type of geosynthetic material, the nature of the material in contact with the geosynthetic material, and for GCLs the internal reinforcement and moisture content (Zornberg et al. 2005; McCartney et al. 2009). Furthermore, many geosynthetic materials have significantly different peak and large-displacement shear strengths.

Hydrated unreinforced or poorly reinforced GCLs can have extremely low in-plane shear strengths, in some cases as low as the hydrated strength of their sodium bentonite core (e.g., a friction angle, ϕ , equal to 4° at high normal stresses). However, such materials are rarely used in landfill construction. Hydrated reinforced GCLs can have a substantial peak internal shear resistance, sometimes in excess of $\phi = 30^\circ$, and still maintain a relatively high strength when subject to large deformation (i.e., $\phi > 15^\circ$), depending on the quality of the reinforcement.

Smooth geomembranes in contact with cohesive soils (e.g., compacted clay) can have very low shear strength ($\phi =$ on the order of 10°), but textured geomembranes can provide substantially higher interface strengths, even when subject to a large displacement (ϕ sometimes on the order of 16 – 20° for the large displacements). Granular materials in contact with smooth geomembranes may have an interface strength exceeding 70 % of the shear strength of the granular soil, while the shear strength of a granular material in contact with a textured geomembrane may approach the shear strength of the soil. Nonwoven geotextile placed against a smooth geomembrane will have a relatively low shear

strength but can have a peak interface strength exceeding that of most natural soils if placed against a textured geomembrane. Woven geotextile placed against a smooth geomembrane will typically have a higher interface strength than a nonwoven geotextile against a smooth geomembrane. Woven geotextiles are not typically placed against textured geomembranes. Geonets placed upon a geomembrane (textured or smooth) may have a low interface strength ($\phi = 10^\circ$) but can be heat bonded to a nonwoven geotextile and then placed against the geomembrane if a higher interface strength is desired.

Minimum required interface strengths are typically established by design analyses and then specified in design documents. Laboratory testing is often conducted during design to demonstrate that the required strengths are achievable with available materials. The geosynthetic materials used in construction are then generally subject to conformance testing prior to the start of construction and to quality assurance testing during construction to demonstrate that the specific materials used in construction can provide the specified interface shear strength.

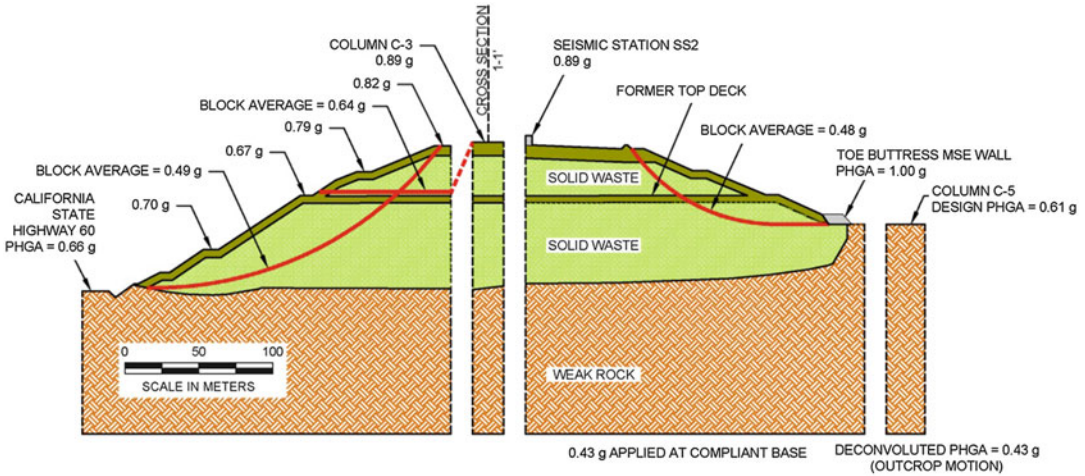
Other geosynthetic properties of importance to seismic design include the tensile strength and the strain at yield and at failure of the geosynthetic elements of the containment system. For advanced analyses, the shear and axial stiffness of geosynthetic material elements may be required.

Design Analyses

Decoupled Analysis: Most seismic analyses of landfill systems conducted in practice today employ a decoupled approach in which the response of the landfill to the seismic loading is considered separately from integrity assessments for the waste mass and containment system components. The seismic response analysis is conducted assuming that the containment system components and waste mass maintain their integrity, that there is no relative displacement at geosynthetic interfaces, and that no instabilities develop. The most common type

of seismic response analysis is a one-dimensional equivalent linear analysis of a representative column or columns for the waste mass. For important projects, nonlinear one-dimensional analysis and two-dimensional equivalent linear analyses are sometimes performed. Nonlinear two-dimensional analyses are primarily used for research purposes. The integrity assessment is usually based upon the calculated seismic response, permanent seismic displacement calculated in a Newmark-type seismic displacement analysis, and empirical rules on allowable displacement for the various elements of the containment system. The Newmark displacement analysis is typically conducted using shear stress time histories from the seismic response analysis and a yield acceleration calculated using conventional limit equilibrium analysis. The standard of practice is to use the large-displacement static shear strength of any interface engaged in the limit equilibrium analysis. If more than one interface is aligned with a slip surface engaged in the limit equilibrium analysis, the large-displacement static strength of the interface with the lowest static peak strength is employed.

Due to the interrelationship between the yield acceleration, shear stress time history, and calculated seismic displacement, multiple potential sliding mechanisms must be checked in a landfill seismic stability assessment. Potential sliding mechanisms include veneer failures of the landfill cover along interfaces with containment system elements (both geosynthetic and natural soil) and between the cover soil and the waste, circular surfaces through the waste mass and/or the foundation, and sliding block surfaces along planes of weakness in waste, including along interfaces with geosynthetic and natural soil containment system elements and other potential planes of weakness. Figure 7 provides an example of a landfill seismic stability analysis in which multiple potential sliding surfaces were evaluated (Kavazanjian et al. 2013b). In this case, the horizontal planes of weakness were continuous layers of daily and interim covers hypothesized to be present within the waste mass. The calculated seismic displacement on each of these failure surfaces is considered independently in the integrity assessment.



Seismic Design of Waste Containment Systems, Fig. 7 Potential sliding mechanisms assumed in seismic stability assessment for the OII Landfill: g-values

represent the peak average acceleration for each potential sliding mechanism (Kavazanjian et al. 2013b)

Due to the conservatism associated with the decoupled approach and the use of large-displacement static shear strength, the calculated permanent seismic displacement from this type of analysis is considered merely an index of seismic performance and not the expected permanent seismic displacement along the hypothesized sliding surface. A calculated seismic displacement of 150 mm or less is typically considered indicative of minimal permanent seismic displacement in the design earthquake, and calculated displacements of up to 1 m are considered indicative of limited displacement. Calculated displacements of greater than 1 m are considered indicative of the potential for large, possibly uncontrolled displacements in the design earthquake.

Table 1 summarizes typical values of acceptable calculated seismic displacement for non-geosynthetic components of landfill containment systems from a decoupled analysis, along with the anticipated duration for the interim and final repairs of these components. For the geosynthetic elements of waste containment systems, 150 mm of calculated displacement is generally considered to be indicative of no damage, and up to 3 m of displacement may be considered acceptable for geosynthetic elements in which

damage can be readily detected in a post-earthquake inspection and readily repaired. It should be noted that these general guidelines are based primarily upon back analyses of the performance of slopes, embankments, and landfills in earthquakes using conventional decoupled analysis. If a more sophisticated seismic analysis is conducted, e.g., a two-dimensional nonlinear coupled analysis of seismic response and displacement, these guidelines may not be appropriate, particularly with respect to damage to the geosynthetic elements of the containment system.

Coupled Analysis: Coupled analysis of landfill seismic response and waste containment system integrity is rare and complex due to difficulties in modeling slip at interfaces and geosynthetic element stress-strain behavior. Bray and Rathje (1998) developed a one-dimensional model for landfill response that included slip at the interface. Kavazanjian et al. (2012) describe a fully coupled two-dimensional model for performance-based seismic design of waste containment systems that accounts for both slip at an interface and geosynthetic element stress-strain behavior. The interface model allows for a post-peak decrease in the interface shear strength. Geosynthetic

Seismic Design of Waste Containment Systems, Table 1 MSW landfill seismic design criteria and performance standards for decoupled analysis (Kavazanjian et al 1998)

Cover system component	Design criteria and performance standard	Interim remediation to restore compliance	Repair to pre-earthquake condition
Final cover			
Soil monocover on side slopes	150 mm of soil deformation; partial failure contained on site	3 months to strip vegetation and regrade and recompact areas of cracking	12 months to restore vegetation
Landfill gas control			
Collection wells	Up to 25 % of wellheads broken	1 month to route headers around broken wellheads	12 months to repair/replace broken wellheads
Headers	Up to 25 % of header pipes cracked or broken	1 month to bypass broken header pipes	3 months to repair/replace broken headers
Vacuum pumps	Power loss; no structural damage	None required	1 month to restore off-site power
Leachate transmission pipes	Acceptable breakage of pipes with double containment	1 month to bypass broken pipes	3 months to repair broken pipes
Surface water management			
Conveyance systems (bench channels, down drains, culverts)	Cracking and up to 300 mm of displacement	2 months to completely restore surface pathways	9 months to replace/rebuild surface pathways
Sedimentation basin	Minor cracking of concrete	2 weeks to 1 month to patch the cracks	9 months to rebuild the basin (if needed)
Access roads	300 mm displacement (cracking)	2 months to patch the cracks	12 months for full repair

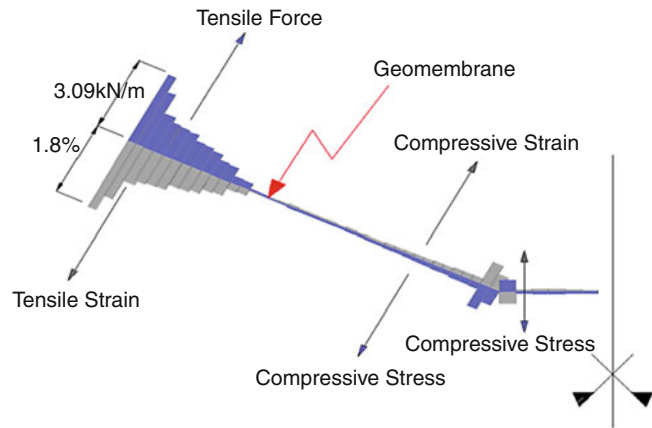
elements of the waste containment system are modeled as structural beam elements with zero moment of inertia (to allow for buckling) and with a parabolic stress–strain model, allowing for explicit calculation of forces and strains in the geosynthetic elements of the containment system. Parametric studies using this model suggest that in some cases the geosynthetic elements of a waste containment system may not sustain significant damage even though a decoupled Newmark analysis results in a calculated displacement of greater than 150 mm. In particular, when the interface shear strength on the top of a geosynthetic element is less than the interface shear strength on the bottom of the geosynthetic element and there are no irregularities (penetrations, seams) in the element, the interface may be able to sustain large sliding displacements without incurring damage to the geosynthetic element, i.e., without exceeding its

tensile strength. However, a back analysis of the geomembrane tears observed at two separate locations at the Chiquita Canyon Landfill (cited previously) has suggested that damage can occur to geosynthetic waste containment system elements when the calculated Newmark displacement is less than 150 mm and when there are irregularities in the containment system, e.g., due to strain concentrations around patches in the geomembrane at locations where coupons were removed for destructive testing as part of the construction quality assurance (CQA) program (Kavazanjian et al. 2013).

Performance-Based Design: Figure 8 illustrates the strain in the geomembrane for a landfill side-slope liner as predicted by the performance-based seismic design model described by Kavazanjian et al. (2012). While this model has yet to be fully validated, back analysis of the performance of the Chiquita Canyon Landfill in

Seismic Design of Waste Containment Systems,

Fig. 8 Strain induced in a side-slope liner system geomembrane by seismic loading (Kavazanjian et al. 2012)



the Northridge earthquake (Kavazanjian et al. 2013a) indicates that strains due to the anchoring of the geomembrane at the crest of the slope and strain concentration factors due to seams in the geomembrane (Giroud 2005) at locations where coupons were removed for CQA testing must be considered in order to predict the tears observed in the liner following the earthquake. These recent findings suggest that, at a minimum, new guidelines need to be developed for acceptable locations for removal of coupons for CQA testing and for minimizing other irregularities in the geosynthetic elements of the containment system. However, ultimately performance-based design using fully coupled models that consider interface and geosynthetic material behavior to explicitly predict strain and forces in geosynthetic elements of modern waste containment systems is likely to become the standard of practice, at least for critical facilities in areas of high seismicity.

Summary

Seismic design of waste containment systems is an important consideration with respect to both maintaining the protection afforded by the containment system for human health and the environment and providing safe and secure capacity for waste disposal during earthquake response and recovery operations. Important engineering considerations with respect to the seismic design of waste containment systems include the seismic

response of the waste mass and methods for evaluating the integrity of containment system components subject to seismic loading. Observations of landfill performance in earthquakes indicate that the waste mass may amplify free-field ground motions and that containment system elements, including geosynthetic liner system components and soil cover, are susceptible to earthquake-induced damage. Waste mass seismic response depends to a large extent upon the waste mass material properties, including the small-strain stiffness, strain-softening behavior, and internal damping in the waste. For landfills with geosynthetic liner and cover systems, the interface and in-plane shear strength of the geosynthetic components is also an important consideration with respect to seismic response. Most containment system performance analyses employed in engineering practice are decoupled analyses in which waste mass seismic response is calculated using the equivalent linear method and the performance of liner system components is assessed based upon the calculated seismic displacement from a Newmark-type analysis. The seismic performance of containment system components is then assessed based upon empirical rules for allowable calculated displacement. While more sophisticated fully coupled performance-based analyses in which the forces, strains, and displacements induced by the earthquake in containment system components are explicitly calculated are possible, such analyses have yet to make their way into engineering practice.

Cross-References

- [Dynamic Soil Properties: In Situ Characterization Using Penetration Tests](#)
- [Geotechnical Earthquake Engineering: Damage Mechanism Observed](#)
- [Nonlinear Dynamic Seismic Analysis](#)
- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- [Seismic Design of Dams](#)
- [Site Response: 1-D Time Domain Analyses](#)

References

- Augello AJ, Matasovic N, Bray JD, Kavazanjian E Jr, Seed RB (1995) Evaluation of solid waste landfill performance during the Northridge earthquake. In: *Earthquake design and performance of solid waste landfills*. Geotechnical special publication no. 54. ASCE, Reston, VA, pp 17–50
- Bray JD, Rathje EM (1998) Earthquake-induced displacements of solid-waste landfills. *J Geotech Geoenviron Eng ASCE* 124(3):242–253
- Bray JD, Zekkos D, Kavazanjian E Jr, Athanasopoulos GA, Reimer M (2009) Shear strength of municipal solid waste. *J Geotech Geoenviron Eng ASCE* 135(6):709–722
- Buranek D, Prasad S (1991) Sanitary landfill performance during the Loma Prieta earthquake. In: *Proceedings of the second international conference on recent advances in geotechnical earthquake engineering and soil dynamics*. University of Missouri, Rolla, pp 1655–1660
- Dixon N, Jones DRV, Whittle RW (1999) Mechanical properties of household waste: in situ assessment using pressuremeters. In: *Proceedings of the Sardinia'99: seventh international waste management and landfill symposium*, vol III, pp 453–460
- Eid HT, Stark TD, Douglas WD, Sherry PE (2000) Municipal solid waste slope failure. 1: waste and foundation properties. *J Geotech Geoenviron Eng, ASCE* 126(5):397–407
- EMCON (1994) Northridge earthquake seismic evaluation, Chiquita Canyon landfill. Technical report prepared for Laidlaw Environmental Systems by EMCON Associates, San Jose
- Giroud J P (2005) Quantification of geosynthetic behavior. *Geosynt Int* 12(1):2–27, Special Issue on the Giroud Lectures
- Harder LS Jr (1991) Performance of earth dams during the Loma Prieta earthquake. In: *Proceedings of the second international conference on recent advances in geotechnical earthquake engineering and soil dynamics*. University of Missouri, Rolla, pp 1673–1691
- Idriss IM, Fiegel G, Hudson MB, Mundy PK, Herzog R (1995) Seismic response of the operating industries landfill. *Earthquake design and performance of solid waste landfills*. Geotechnical special publication no. vol 54. ASCE, Reston, VA, pp 83–118
- Johnson ME, Lundy J, Lew M, Ray ME (1991) Investigation of sanitary landfill slope performance during strong ground motion from the Loma Prieta Earthquake of October 17, 1989. In: *Proceedings of the second international conference on recent advances in geotechnical earthquake engineering and soil dynamics*. University of Missouri, Rolla, pp 1701–1706
- Kavazanjian E Jr, Matasovic N (2001) Seismic design of mixed and hazardous waste landfills. In: *Proceedings of the fourth international conference on recent advances in geotechnical earthquake engineering and soil dynamics*. University of Missouri-Rolla, Rolla (on CD ROM)
- Kavazanjian E Jr, Matasovic N, Stokoe KH II, Bray JD (1996) In-Situ shear wave velocity of solid waste from surface wave measurements. In: *Proceedings of the 2nd international congress environmental geotechnics*, Osaka, Balkema, vol 1, pp 97–104
- Kavazanjian E Jr, Matasovic N, Caldwell J (1998) Damage criteria for solid waste landfills. In: *Proceedings of the sixth US national conference on earthquake engineering*, EERI (on CD ROM)
- Kavazanjian E Jr, Arab MG, Matasovic N (2012) Performance based design for seismic design of geosynthetics-lined waste containment systems. In: *Proceedings of the second international conference on performance-based earthquake geotechnical engineering*, State of the Art – State of the Practice Volume, Keynote Lecture No. 19, University of Catania, p 23 (on CD ROM)
- Kavazanjian E Jr, Arab M, Matasovic N (2013a) Performance of Two Geosynthetics-Lined Landfills in the Northridge Earthquake. In: *Proceedings of the seventh international conference on case histories in geotechnical engineering*, Missouri University of Science and Technology, p 14 (on CD ROM)
- Kavazanjian E Jr, Matasovic N, Bachus RC (2013b) The 11th Peck lecture: pre-design geotechnical investigation for the OII superfund site landfill. *J Geotech Geoenviron Eng, ASCE* 139(10):1839–1863
- Koerner RL, Soong T-Y (2000) Stability assessment of ten large landfill failures. In: *Advances in transportation and geoenvironmental systems using geosynthetics*. Geotechnical special publication no. 103. ASCE, Reston, VA, pp 1–38
- Matasovic N, Kavazanjian E Jr (1998) Cyclic characterization of OII landfill solid waste. *J Geotech Geoenviron Eng ASCE* 124(3):197–210
- Matasovic N, Kavazanjian E Jr, Anderson RL (1998) Performance of solid waste landfills in earthquakes. *Earthquake Spectra EERI* 14(2):319–334

- Matasovic M, Kavazanjian E Jr, De A, Dunn RJ (2005) - CPT-based seismic stability assessment of a hazardous waste site. *Int J Soil Dynam Earth Eng* 26(2):201–208
- McCartney JS, Zornberg JG, Swan RH (2009) Analysis of a large database of GCL-Geomembrane interface shear strength results. *J Geotech Geoenviron Eng ASCE* 134(2):209–223
- Mitchell JK, Seed RB, Seed HB (1990) Kettleman hills waste landfill slope failure. I: liner-system properties. *J Geotech Eng ASCE* 116(4):647–668
- Orr WR, Finch MO (1990) Solid waste landfill performance during Loma Prieta Earthquake. In: *Geotechnics of waste fills: theory and practice*, vol 1070, Special technical publication. ASTM International, Conshohocken, pp 22–30
- Ramaiah BJ (2013) Static and seismic stability assessment of municipal solid waste landfills. Ph.D dissertation, Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas
- Seed RB, Mitchell JK, Seed HB (1990) Kettleman hills waste landfill slope failure. II: stability analyses. *J Geotech Eng* 116(4):669–690
- Yuan P, Kavazanjian E Jr, Chen W, Seo B (2011) Compositional effects on the dynamic properties of municipal solid waste. *Waste Manag* 31:2380–2390
- Zekkos Zeccos D P (2005) Evaluation of static and dynamic properties of municipal solid-waste. PhD thesis, Department of Civil and Environmental Engineering, University of California, Berkeley
- Zekkos D, Bray JD, Kavazanjian E Jr, Matasovic N, Rathje E, Reimer M, Stokoe KH III (2006) Unit weight of municipal solid waste. *J Geotech Geoenviron Eng ASCE* 132(10):1250–1261
- Zekkos D, Bray JD, Riemer MF (2008) Shear modulus and material damping of municipal solid waste based on large-scale triaxial testing. *Can Geotech J* 45:45–58
- Zornberg JG, McCartney JS, Swan RH Jr (2005) Analysis of a large database of GCL internal shear strength results. *J Geotech Geoenviron Eng ASCE* 131(3):1–14

Seismic Event Detection

Jens Havskov¹ and Gerardo Alguacil²

¹Department of Earth Science, University of Bergen, Bergen, Norway

²Instituto Andaluz de Geofísica, University of Granada, Granada, Spain

Synonyms

Earthquakes; Seismic events; Seismic signals; Seismogram; Volcanic events

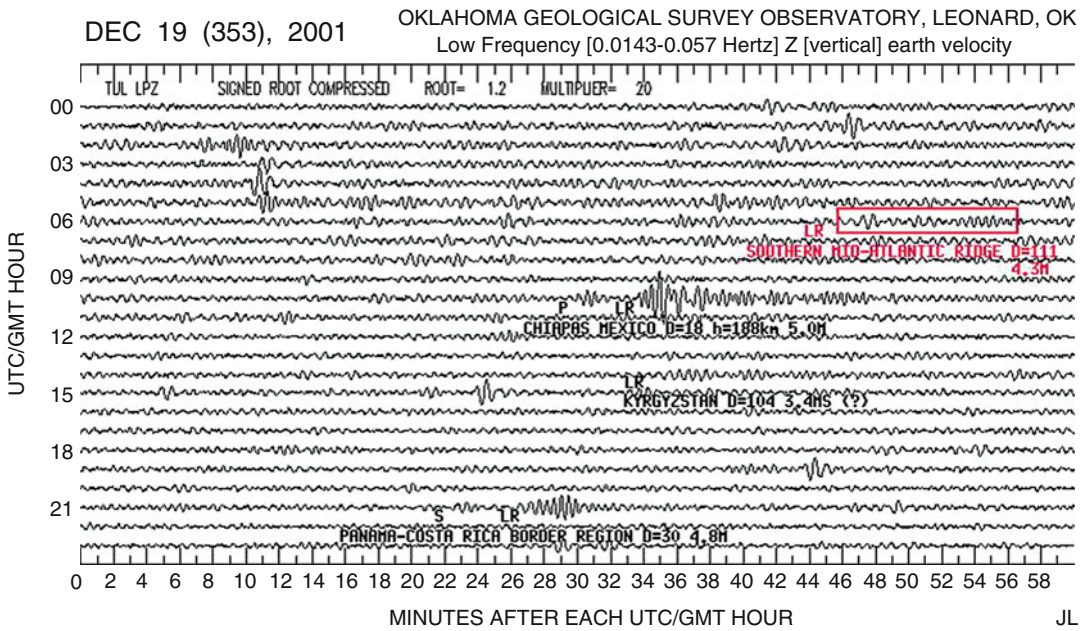
Introduction

Seismic stations record signals continuously; however, the main interest for the user is the recordings of seismic events. The events will normally be contained in a very small fraction of the actually recorded signal (Fig. 1).

As it is seen in this 24 h seismogram, there are only a few sections with real seismic events and a lot of other wiggles which is just noise. The more or less continuous noise seen throughout the seismogram is the earth's background noise (see below and Figs. 2 and 4). In the old days when all the signals were recorded on paper (on seismograms), the events were detected easily by eye, particularly for small local events with high-frequency content. This method is sometimes also used with today's digital recordings which can be displayed on screen or printed on paper. But the majority of event detection is done automatically, either in real time to limit the amount of data recorded or off-line when large data sets have been collected, e.g., from field experiments. Over the years a few simple and very many sophisticated algorithms have been developed to automatically detect a "real" signal in the presence of noise. Few, if any, have been developed that can beat the human eye, which still is used to evaluate the efficiency of the detector. The use of a trigger or a detector can have two purposes:

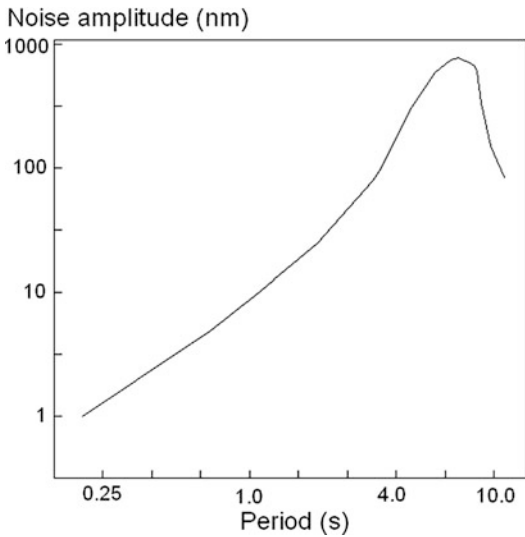
1. Determine the onset of real seismic events in a large data set. These onset times will be used for further processing.
2. Extract out the waveform time segments of a larger data set which contains seismic events, either in a field recorder to limit the amount of data recorded or to extract out just the useful events for processing.

In this entry, the most common event detection methods will be described. An overview of detection methods is found in Sharma et al. (2010).



Seismic Event Detection, Fig. 1 A seismogram for a 24 h period. Detected events are indicated with the name of the region where the earthquake occurred, D is

the distance from the station to the event in degrees, h is the depth of the earthquake, and M is the magnitude (Figure from www.okgeosurvey1.gov/seis/20011219.gif)



Seismic Event Detection, Fig. 2 Typical amplitude of the microseismic background noise as a function of period(s) (The data for the plot are taken from Brune and Oliver 1959)

Examples of Seismic Signals

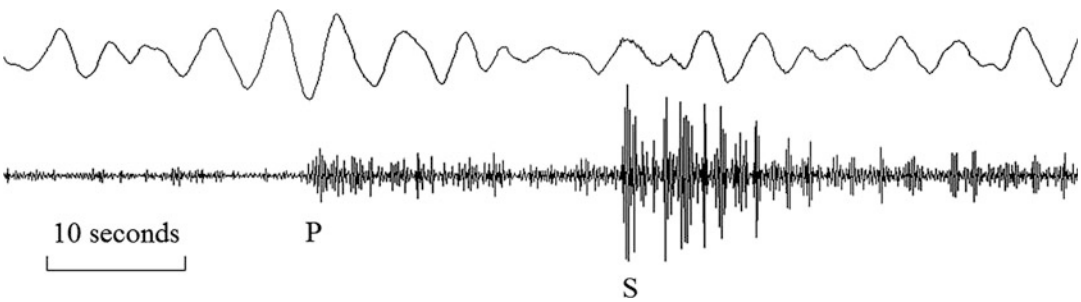
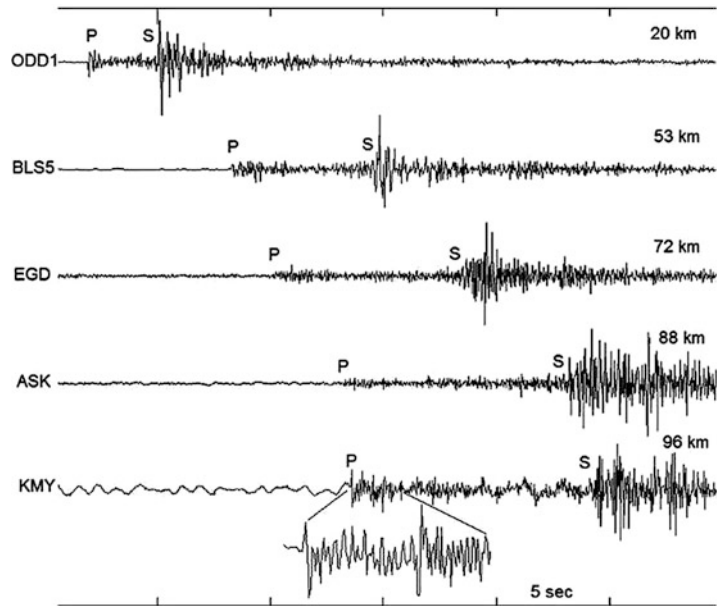
Detecting real seismic events in a continuous data stream can be quite a challenge, particularly if there is nonstationary noise like noise from cars passing by or wind gusts and in some cases it is virtually impossible to distinguish between man-made noise and real seismic events.

In addition, there is an ever present so-called microseismic background noise (see Figs. 4 and 1). This noise is caused by standing waves in the ocean and wave action near the coasts. It is present all over the world, strong near the coast, and weaker far from the coast. It has a maximum amplitude at a period of about 4–6 s. Figure 2 shows a typical curve with the noise as a function or period.

Seismic events can have quite a range of appearances and frequency content depending on the type of seismic sensor used and the

Seismic Event Detection,

Fig. 3 Seismogram of a local event ($M = 1.5$) at 5 stations. Sensors are short period and the traces are not filtered. Each trace is auto scaled. The first arrival phase P and second arrival phase S are indicated. The epicentral distances (distance from the earthquake to the station measured along the surface) are given above the traces to the *right* (The figure is from Havskov and Ottemöller 2010)



Seismic Event Detection, Fig. 4 A magnitude 1.5 event recorded on a broadband station. The *top* trace is the original trace and the *bottom* trace has been filtered from 5 to 10 Hz

distance to the event and its magnitude. Essentially there are three types of seismic sensors: short-period measuring ground velocity (frequencies above 1 Hz), broadband measuring ground velocity (frequencies typically from 0.01 to 50 Hz), and accelerometers measuring ground acceleration from 0 to 100 Hz. Some examples will be given of seismic events.

Local small earthquake recorded with short-period sensors (Fig. 3). The signals are high frequency, and P- and S-phases are clearly seen although the P-phase is getting close to the noise level for station ASK.

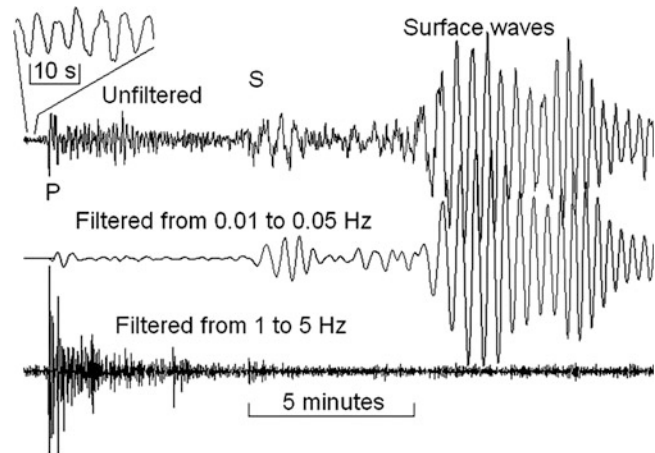
Small event recorded on a broadband station (Fig. 4). The high background noise level, the

microseismic noise, is the harmonic signal dominating the record. It has a period of 4.5 s. A small high-frequency event will “drown” in this noise and can practically not be seen on the original record. By filtering the signal from 5 to 10 Hz, the local event is clearly seen. This illustrates that event triggering often will have to work on filtered traces. So how did humans detect this kind of event before digital recording? Simply by recording the signals in two different frequency bands, short period above 1 Hz and long period below 0.1 Hz, and thereby filtering out the noise around the microseismic peak (Fig. 2).

Distant event recorded on a broadband station (Fig. 5). This magnitude 6.2 event is recorded at

Seismic Event Detection,

Fig. 5 Recording of an earthquake at a large distance (4,200 km). The insert above left shows a zoom of the noise before the *P*



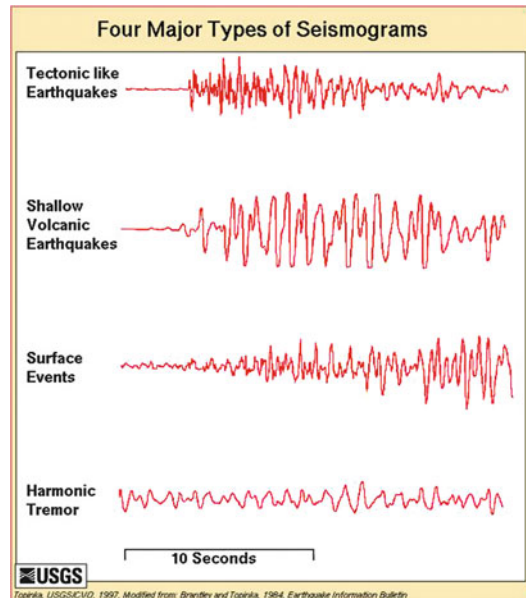
a distance 4,200 km. It is seen that a large distant event has much lower frequencies than small local events, partially due the source nature of earthquakes and partly due to filtering out the high frequencies along the path.

The unfiltered trace shows, before the *P*, the microseismic noise. For smaller events the *P* might drown in the noise. The microseismic noise is filtered away in the middle trace and the phases appear very clear. The bottom trace is filtered so the signal looks like it would have been recorded on a short-period seismograph. In this case a detector would not be able to find the true extent of the seismic event since the long-period *S*- and surface waves are not seen.

Volcanic events (Fig. 6). Volcanoes present a special problem since the signals from the events often are very emergent. Figure 6 shows the most common types.

The tectonic event is like any other earthquake and is easy to detect. The shallow events are very harmonic and monochromatic and of sufficient amplitude to be easy to detect. The surface event can be very difficult to detect automatically since it is very emergent and might be mistaken for noise. The harmonic tremor is a continuous series of harmonic events and cannot be detected.

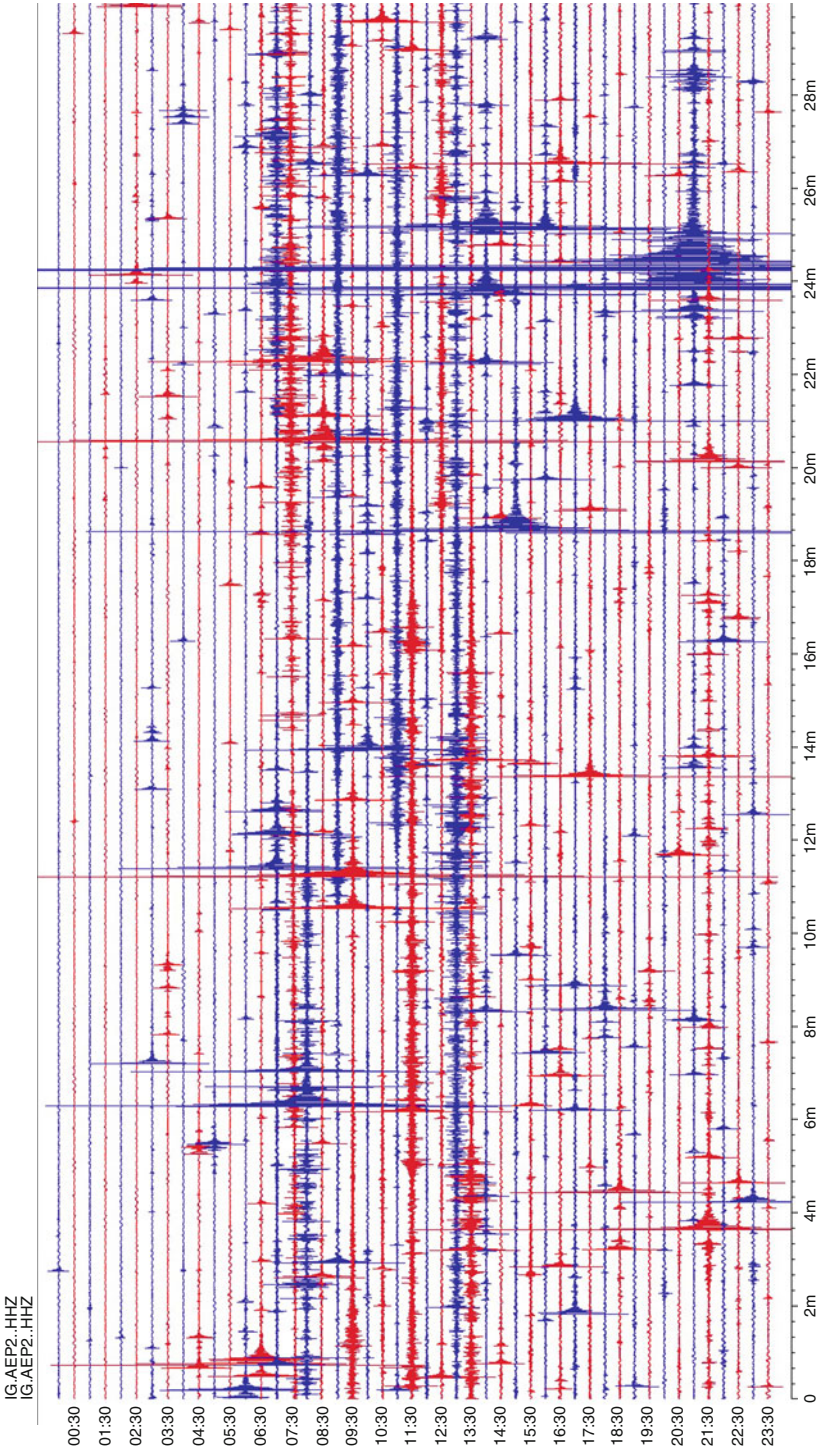
Many events close in time: Figure 7 shows swarm activity of small earthquakes in a 24 h period. The human eye can easily detect the individual events while an automatic trigger would have problems with separating the events.



Seismic Event Detection, Fig. 6 Major types of seismic signals from volcanoes (From <http://vulcan.wr.usgs.gov/Images/Gif/Monitoring/Seismic/quakes.gif>)

Trigger Methods

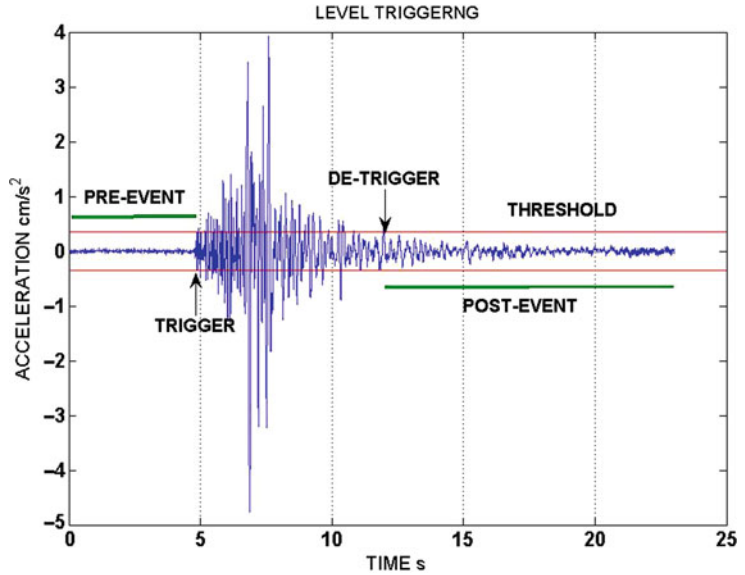
A trigger or detector is an algorithm that checks the signal for variations that could indicate an event. There are several trigger algorithms available, some very sophisticated using neural networks (e.g., Romeo et al. 1995) and pattern recognition (e.g., Köhler et al. 2010). In the hands of an expert, they can significantly improve the event detections/false trigger ratio,



Seismic Event Detection, Fig. 7 Swarms of earthquakes. A 24 h record on 5 February 2013 of a temporary broadband station during the seismic swarm of Torreperegil (Southern Spain)

Seismic Event Detection,

Fig. 8 An example of level trigger. The plot shows a horizontal component of the accelerogram of a small local earthquake. The trigger level is marked with two horizontal lines. In this case the de-trigger level is the same as the trigger level. Usually the triggers of the three components are combined, and an event is declared if at least two or three channels are triggered within a given time window (e.g., 1 s)



particularly for a given type of seismic events. However, the sophisticated adjustments of operational parameters to the actual signal and seismic noise conditions, at each seismic site, which these triggers require, have often been proven unwieldy and subject to error. Hence, for practical purposes, only two types of triggers are widely used, the level trigger and the *STA/LTA* trigger (see below), and only those will be described here.

The level trigger: This trigger is also called the amplitude threshold trigger and simply searches for any amplitude exceeding a preset threshold. The start of the event is declared when this threshold is reached and the end of the event is declared when the level reaches the de-trigger level, usually smaller than the trigger level. When this type of trigger is installed in a recorder, that only shall record the real events, the recording window will be extended with a pre-event time before the trigger and a post-event time after the trigger (Fig. 8). This is to ensure that the whole event is recorded.

The level trigger has some obvious problems. If DC (a constant voltage) is present in the signal and not removed before triggering, it can prevent triggering or keep the instrument in continuous trigger mode. If no averaging or filtering is done,

the level trigger will also trigger on a single high value (a spike), which obviously is not a seismic signal.

This algorithm is often used in strong-motion seismic instruments (recording large ground accelerations), where high sensitivity is not an issue and where consequently man-made and natural seismic noise is not critical. The level trigger is now largely replaced by the *STA/LTA* trigger (see next section), since computer power no longer is an issue. However, many instruments still have the level option.

The short-term average – long-term average trigger (STA/LTA) is the most frequently used trigger algorithm. A single channel of seismic signal is typically processed as follows: The signal is band-pass filtered, and the absolute average *STA* (short-term average) over the *STA* time window is determined. Typically, the *STA* time window is 0.5 s for local earthquakes. A short *STA* time makes the trigger more sensitive to short-term variations in the signal, while a longer *STA* is better at averaging out short-term fluctuations and it can be compared to a low-pass filter. If there are spikes in the signal, a long *STA* must be used in order to average out the interference. This will of course reduce the sensitivity to short-lasting signals, but if a longer-lasting signal is the

objective, *STA* can be as long as the duration of the main P-wave train. The most common mistake in setting the *STA* is to use a too short value.

The same filtered signal is also used to calculate the *LTA* (long-term average) over the *LTA* time window, which is typically 50–500 s. Thus, *LTA* will give the long-term background signal level, while the *STA* will respond to short-term signal variations. The ratio between *STA* and *LTA* is constantly monitored, and once it exceeds a given threshold, the trigger level, the start of an event is declared for that trace. The *LTA* should be short enough to adapt to slow changes of the background noise and long enough to avoid reducing the sensitivity to triggering on low-amplitude emergent signals. This can be a problem with very emergent signal where the *STA* might increase fast enough to prevent triggering (see the surface event in Fig. 6). Since the natural background noise usually changes very slowly, it is generally better to use a too long than a too short *LTA* time. If *LTA* is not frozen during triggering (see next section), the *LTA* time must be at least as long as the event duration to prevent premature de-triggering. Using a too short *LTA* time combined with a high trigger ratio might even prevent triggering completely since the *LTA* is adjusted upward so fast that *STA/LTA* never exceeds the trigger ratio.

Once the event starts, the *LTA* is usually frozen, so that the reference level is not affected by the event signal. The end of the event is declared when the *STA/LTA* ratio reaches the de-trigger level. Trigger levels and de-trigger levels of 4.0 and 2.0, respectively, are typical values. In order to get the complete event, the recording will start pre-event time before the trigger time. Likewise, to avoid the truncation of the signal, recording continues some time after the de-triggering, the post-event time. The maximum time of recording can usually be limited by a maximum recording time setting, and it is also sometimes possible to discard triggers lasting less than a certain time, the minimum trigger time.

STA and *LTA* must be calculated on signals without a DC component in order to reflect the real signal. The DC is normally removed by

filtering, which is desirable also for other reasons (see below). The *STA/LTA* trigger algorithm is well suited to cope with fluctuations of natural seismic noise, which are slow in nature. It is less effective in situations where man-made or natural seismic noise of a bursting or spiky nature (e.g., wind gusts) is present. At sites with high, irregular seismic noise, the *STA/LTA* trigger usually does not function well, meaning there are too many false triggers.

A typical implementation will now be described. The trigger must be running continuously in the computer, so the *STA* and *LTA* values are calculated as running averages:

$$\begin{aligned} STA_i &= STA_{i-1} + \frac{|x_i| - STA_{i-1}}{NSTA} \\ LTA_i &= LTA_{i-1} + \frac{|x_i| - LTA_{i-1}}{NLTA} \\ R_i &= \frac{STA_i}{LTA_i} \end{aligned} \quad (1)$$

where x_i is the signal (filtered or unfiltered); *STA*, the short-term average; *LTA*, the long-term average; *R*, the *STA-LTA* ratio; and *NSTA* and *NLTA*, the number of points in the *STA* and *LTA* windows, respectively. In this case, x is the unfiltered signal. This works if the signal has no DC component, but if x has a large DC component, this will be transferred to the *STA* and *LTA* and they might be so large that *R* never reaches the trigger ratio.

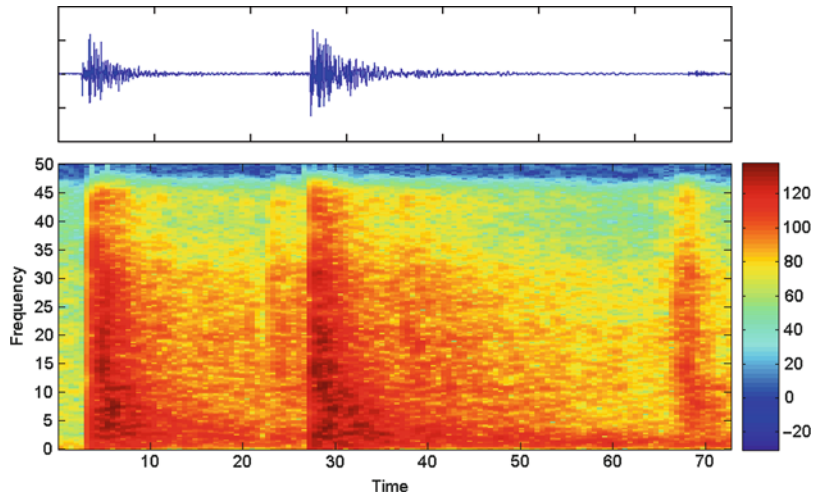
There are two common variations of the trigger algorithm Eq. 1. One is to use the squared x_i instead of the absolute value. Using the squared x_i , makes the trigger more sensitive to changes. The other is the Allen trigger (Allen 1978), in which the absolute value of x_i is replaced by the characteristic function c_i . In a simplified form, c_i is calculated as

$$c_i = x_i^2 + k \cdot (x_i - x_{i-1})^2 \quad (2)$$

where k is a constant. The second term is proportional to the squared first time derivative of the signal. The characteristic function is sensitive to

Seismic Event Detection, Fig. 9 Below:

A spectrogram of the signal plotted above, with three local earthquakes closely spaced in time. It is apparent that the amplitude of high frequencies decays faster than lower frequencies after the first arrival of each event. The colors indicate spectral amplitudes



both amplitude variations and frequency changes.

All the above described trigger methods have in common that they are *not-looking-forward* algorithms, i.e., the present value of the trigger parameters depends only on present and past values of the signal. This makes them easy to implement with simple recursive, real-time operations.

If the seismic events are very close in time (Fig. 7), it is obvious that the STA/LTA would produce many failures. If one event is not finished before the next event starts, the first event would simply be prolonged so correct start times of events would only be obtained when a new trigger occurs after the previous has finished. A more sophisticated trigger might also monitor the frequency content of the signal. Since the end of the event usually has a lower-frequency content than the initial P-wave, this could be used to declare the start of a new event before the previous event is finished. This can be illustrated with a spectrogram which is a time-frequency representation of the spectral content of a signal. Figure 9 shows an example with three events occurring within 70 s.

A human analyst works in a quite different way: He detects an event looking at the entire record and evaluates the past, present, and subsequent values of the trace (or something like the envelope of it) to decide the presence of an event

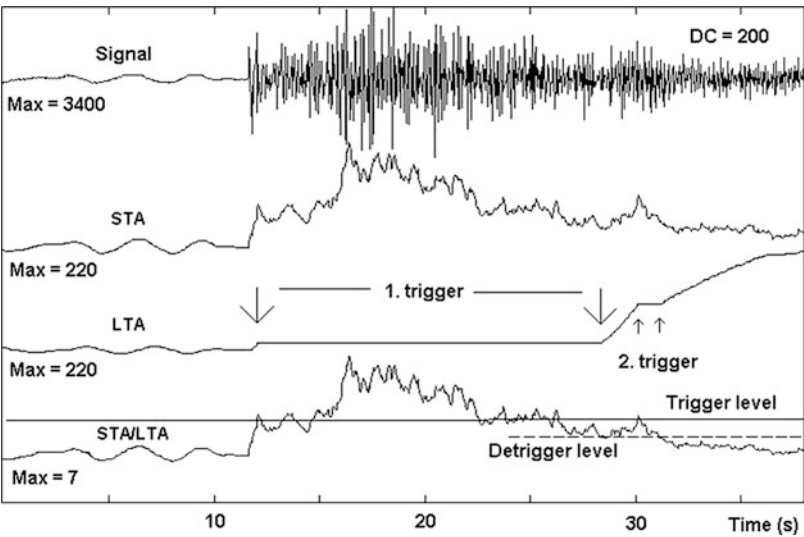
and its onset time. So the human analyst would have no problem detecting the events in Fig. 7.

Figure 10 shows an example of how the algorithm STA/LTA works for the standard trigger using the absolute value.

The original signal is a small high-frequency earthquake with a medium-level microseismic background noise, which is reflected in the slow variations of the STA and more smoothed in the LTA. At about time 12 s, the ratio gets above the threshold of 3 (1. trigger). It immediately gets below again, but since the de-trigger ratio is 2, the trigger remains activated. In the above example, the LTA is frozen once an event has been declared, which can be seen from the flat LTA level. A few seconds after the de-triggering, the ratio again rises above the trigger level and remains in trigger state for about 1 s (2. trigger). This signal has a good signal-to-noise ratio, so why has the trigger not been activated much longer? The first reason is simply that the DC has not been removed before calculating STA and LTA. Since STA and LTA before triggering are about 220, the event will de-trigger when STA is 440 and the maximum ratio is 7 (Fig. 10 and Table 1). The DC of this signal is only 200 as compared to the maximum value of 3,400; yet if it is not removed, it will have a large influence on the trigger performance.

When the DC is removed, the figure corresponding to Fig. 10 will look almost the

Seismic Event Detection, Fig. 10 The seismic trigger. *Top* trace is the original signal; the second trace, the *STA*; the third trace, the *LTA*; and the last trace, the *STA/LTA* ratio. The *STA* window is 0.4 s and the *LTA* window is 50 s. The trigger ratio is 3 and the de-trigger ratio 2. All amplitudes are counts (the raw number generated by the recorder) (Figure from Havskov and Alguacil (2010))



Seismic Event Detection, Table 1 Effect of DC and filter on *STA*, *LTA*, and *STA/LTA*. The units of the numbers are in counts

	DC not removed	DC removed	Filtered 5–10 Hz
<i>STA</i> and <i>LTA</i> before trigger	220	40	3
Maximum <i>STA</i>	1,500	1,450	600
Maximum <i>STA/LTA</i>	7	36	200
Duration of trigger (s)	15	50	105

same, but since both noise, *STA* and *LTA*, are smaller due to the removal of the DC, *STA/LTA* will be much higher for the earthquake signal, and the event will remain in trigger state for 50 s instead of 15 s (Table 1). The performance can be further improved by filtering. The signal has a typical microseismic background noise superimposed, and this gives the main contribution to the *STA* and *LTA* before triggering. By using a band-pass filter from 5 to 10 Hz, the performance is further improved (Table 1) and the trigger lasts for 105 s.

So, the conclusion is that every trigger must have an adjustable band-pass filter in front of the trigger algorithm, which has both the function of removing the DC and making the trigger algorithm sensitive to the frequency band of interest.

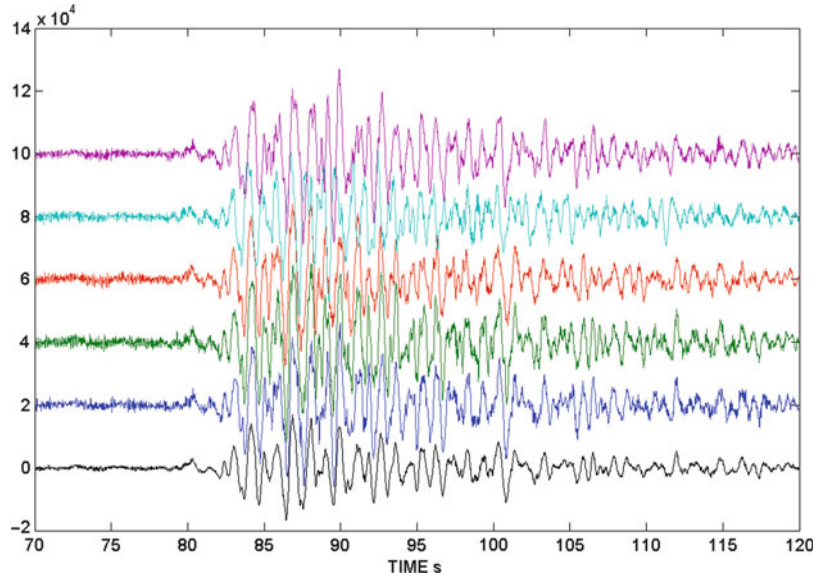
This is particularly important with broadband seismometers, where small earthquake signals often are buried in dominant 0.2–0.3 Hz seismic noise and a trigger without a filter simply would not work except for the largest events (Fig. 4). Some recorders allow several trigger parameter sets to be used simultaneously. This is needed if, e.g., a broadband station has to trigger on micro-earthquakes, teleseismic P-waves, and surface waves which each requires separate filters, *STA*, and *LTA*.

For more detail of how to set and use *STA/LTA* triggers, see Trnkoczy, information sheet 8.1 in New Manual of Seismological Observatory Practice (NMSOP) (Bormann 2012).

Improved Triggering Using Several Stations, the Network Trigger

Whatever trigger method is used with single channels or single three-component stations, there is always likely to be quite a lot of false triggers, all depending on the local conditions and how sensitive the trigger has been set up. When several stations within a single network are processed together, the number of false triggers can be significantly reduced setting up a requirement of concurrent triggering. This means that for a network trigger to be declared, there must be a minimum number of triggers within a given time window, the array

Seismic Event Detection,
Fig. 11 A volcanic event recorded on an array of vertical sensors. The bottom trace (black) is the sum of the individual channels (color), once they have been time aligned using the P-wave relative propagation times. The improvement on the signal-to-noise ratio is apparent



propagation window. The array propagation window is calculated as the time the waves take to propagate across the network. If, e.g., the network is 100 km in diameter and assuming the case that triggering takes place on the S-waves only and the event is outside the network, then the array propagation window should be $100/V_s$ where V_s is the S-wave velocity. If some stations trigger on P and others on S, the array propagation window should be even larger by the S-P time so in general the array propagation time is set to a value larger than $100/V_s$. In case there is swarm activity (Fig. 7), a network trigger might not help a lot since the events come so close in time that it would be hard to associate events, that is, if the events are correctly triggered in the first place.

Triggering Using a Seismic Array

A seismic array consists of multiple sensors in a small enough area that signals are correlated, and the seismic array is therefore also sometimes called a seismic antenna.

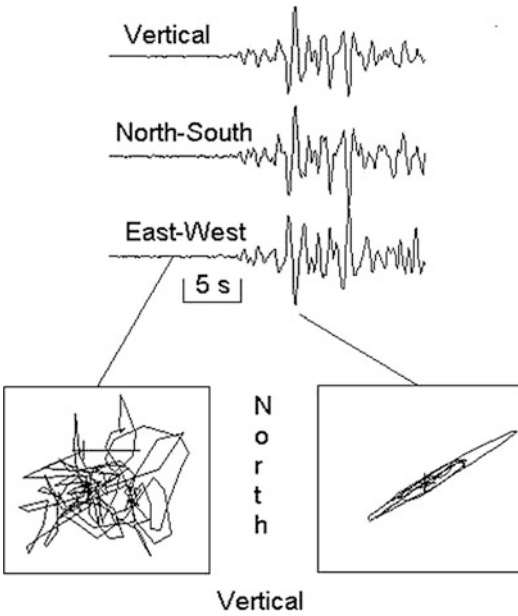
The signal-to-noise ratio of a seismic signal is improved by summing the signals of each station in the array using a delay corresponding to ray travel time across the array so the signals get lined up. In this way the coherent seismic signal

will be enhanced and the supposedly incoherent noise will cancel out. Running a seismic trigger algorithm on the summed signal will therefore be more sensitive and more reliable than on any of the single stations. In a way this is similar to the network trigger where the same event must be present in all channels. An example is shown in Fig. 11.

Triggering Using Three Channels

Most seismic stations today have three channel sensors oriented vertically, north–south and east–west. As a first small improvement in the triggering, it can be required that all three channels trigger on the seismic event. Since external disturbances affect all three channels, this requirement might only improve the detection marginally. With three channels, it is possible to make a polarization analysis or in other words determine the particle motion of the ground (see Fig. 12).

For simplicity, the polarization is only shown in the north-vertical plane. It is seen that the P-wave is not polarized in the noise and highly polarized in the P-wave signal. By continuously calculating the degree of polarization, it is possible to distinguish a P-wave from background noise which would have no polarization.



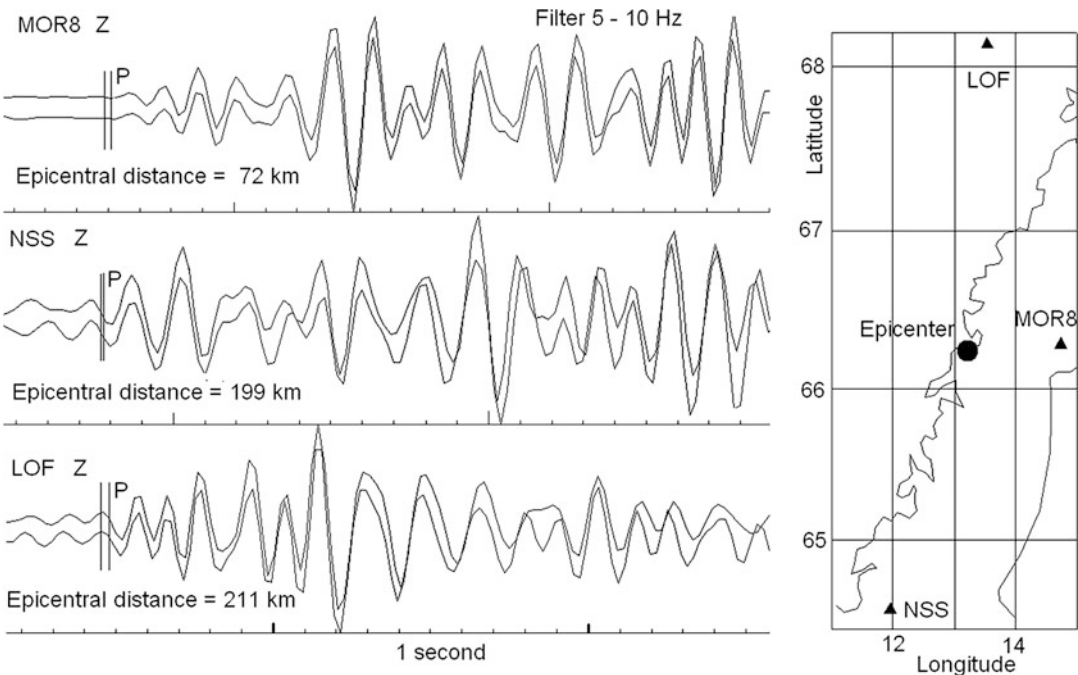
Seismic Event Detection, Fig. 12 *Top:* the three components of ground velocity for a local earthquake. *Bottom:* the particle motion in the north-vertical plane at two different times as indicated on the figure. To the *left* is the analysis of the noise and to the *right* the analysis of the P-wave signal

This property has been used to improve the detection of P- and S-waves (Roberts et al. 1989; Withers et al. 1998).

Triggering Using a Known Waveform

Many seismic events have very similar waveforms if they originate from the same or nearby sites. This could also be the case for volcanic events even though they might not be from exactly the same area. A “known” waveform shape can then be used to detect unknown arrivals by cross correlating the known wavelet with the signal of unknown events. An example of similar events is seen in Fig. 13. Despite the large distances, the signal shapes are nearly identical for the two events.

The only trigger parameter would then be the minimum correlation needed to declare an event. Using this technique, it is often possible to detect events which otherwise could not be detected. When using a seismic array, this method is particularly powerful (Gibbons and Ringdal 2006).



Seismic Event Detection, Fig. 13 *Left:* the first couple of seconds of the P-waves for the two events occurring in the same area as recorded with three different stations.

The *top* trace event has magnitude 1.8 and the *bottom* trace event magnitude 1.6 (Figure from Havskov and Ottemöller (2010))

Summary

Today, huge quantities of continuous seismic data are recorded, and automatic detection of the seismic events is indispensable. Over the years very many algorithms have been developed to detect “real” seismic events; however the overall most used and reliable detectors are based on the standard STA/LTA method. Combined with a network trigger, this becomes a simple and robust event detector, in particular considering that the number of stations is continuously increasing so more use of the network trigger can be done, making the requirement for the single-channel detector less stringent.

Cross-References

- [Passive Seismometers](#)
- [Principles of Broadband Seismometry](#)
- [Seismometer Arrays](#)
- [Seismic Noise](#)

References

- Allen RV (1978) Automatic earthquake recognition and timing from single traces. *Bull Seismol Soc Am* 68:1521–1532
- Bormann, P. (ed) (2012) *New Manual of Seismological Observatory Practice (NMSOP-2)*, IASPEI, GFZ German Research Centre for Geosciences, Potsdam; nmsop.gfz-potsdam.de
- Brune JN, Oliver J (1959) The seismic noise at the earth's surface. *Bull Seismol Soc Am* 49:349–353
- Gibbons SJ, Ringdal F (2006) The detection of low magnitude seismic events using array-based waveform correlation. *Geophys J Int* 165:149–66
- Havskov J, Alguacil G (2010) *Instrumentation in earthquake seismology*, 3rd edn. Springer, Dordrecht, 358 pp
- Havskov J, Ottemöller L (2010) *Routine data processing in earthquake seismology*. Springer, Dordrecht, 347 pp
- Köhler A, Ohrnberger M, Scherbaum F (2010) Unsupervised pattern recognition in continuous seismic wavefield records using self-organizing maps. *Geophys J Int* 182:1619–1630
- Roberts RG, Christofferson A, Cassidy F (1989) Real time event detection, phase identification and source location using single station 3 component seismic data and a small PC. *Geophys J* 97:471–480
- Romeo G, Mele F, Morelli A (1995) Neural networks and discrimination of seismic signals. *Comput Geosci* 21:279–288
- Sharma BK, Kumar A, Murthy VM (2010) Evaluation of seismic event detection algorithms. *J Geol Soc India* 75:533–538
- Withers M, Aster R, Young C, Beiriger J, Harris M, Moore S, Trujillo J (1998) A comparison of select trigger algorithms for automated global seismic phase and event detection. *Bull Seismol Soc Am* 88:95–106

Seismic Fragility Analysis

Murat Altug Erberik
Civil Engineering Department, Middle East
Technical University, Ankara, Turkey

Synonyms

Damage probability matrix; Fragility curve; Fragility surface; Limit state; Numerical simulation; Seismic capacity; Seismic demand; Uncertainty

Introduction

Seismic fragility can be defined as the proneness of a structural component or a system to fail to perform satisfactorily under a predefined limit state when subjected to an extensive range of seismic action. In accordance with the above definition, seismic fragility analysis can be regarded as a probabilistic measure for seismic performance assessment of structural components or systems. There are two different end products of seismic fragility analysis: damage probability matrix and fragility curve.

Damage probability matrix (DPM) is a table that provides discrete values of damage state probabilities for specified levels of ground motion intensities. Each column of DPM stands for a constant level of ground motion intensity whereas each row of DPM denotes the probability of being in a predefined damage state. Hence, any element of DPM represents the probability of

experiencing a certain damage state for a specific level of ground motion intensity.

Fragility curves are continuous functions that represent the probability of exceeding predefined limit (or performance) states for specific levels of ground motion intensity. Fragility curves can be developed for structural and nonstructural components, or they can be generated for structural systems and assemblies. In the field of earthquake engineering, the most common application is to generate fragility curves for building structures. Depending on the characteristics of the building stock in the region of interest, the fragility curves can be derived for many different structural types (i.e., steel, reinforced concrete, or masonry buildings; low-, mid-, or high-rise construction; frame, wall, or mixed structures; etc.). Other than building structures, fragility curves have also been generated for infrastructures (like bridges, dams), lifeline systems (like transportation, power, water supply networks), unrestrained equipment, etc.

Fragility curves can be used for different purposes. The main purpose is to determine the seismic performance of new or existing structural systems. In the former case, the fragility information is obtained for optimal design and in the latter case for condition assessment. Fragility curves can also be employed in order to assess the efficiency of different intervention techniques on existing structural systems. When the curves are derived to represent a certain class of structural systems, they can also be used for regional seismic damage and loss estimation studies. Such studies are very popular nowadays since they can be employed for pre-earthquake mitigation efforts and postearthquake decision-making processes.

There is a close relationship between fragility curves and DPMs so that the information obtained from fragility curves can be converted to construct DPMs or vice versa. Figure 1 demonstrates how to generate DPM from a given set of fragility curves. The vertical axis, simply named as “probability,” actually stands for the probability of exceeding a limit state. The horizontal axis represents the complete range of ground motion intensity levels (ILs). Different

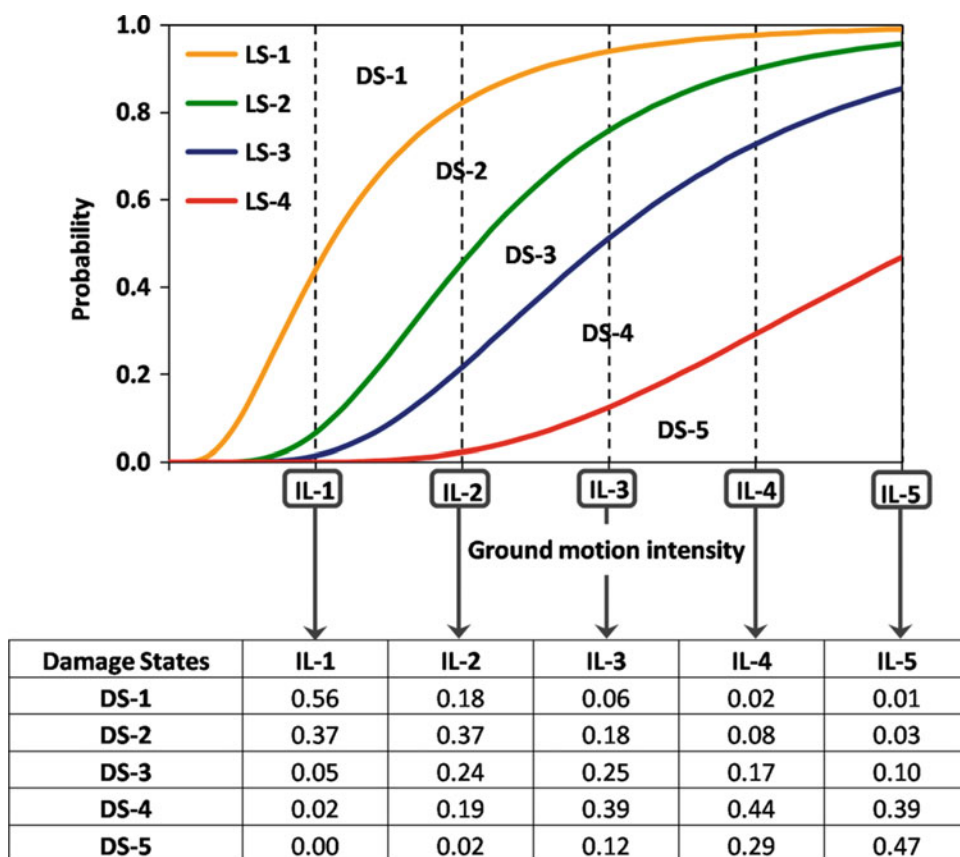
ground motion intensity parameters can be used in the generation of fragility curves. As seen from the figure, four limit states (LSs) have been assigned, which in turn means that there exist five damage states (DSs). Attainment of limit states is an important phase of fragility curve generation and is discussed in following sections. The columns of DPM are obtained by intersecting the fragility curve set with vertical lines (dashed lines in Fig. 1) at specific intensity levels and calculating the damage state probabilities, which are corresponding portions between any two limit states in these vertical alignments.

Different Approaches for Seismic Fragility

There are four main resources to generate the seismic fragility of a structural system or component: expert opinion, field data, experimental data, and analytical simulations.

One of these main data sources for the derivation of fragility curves is based on the opinion of the experts from the field of earthquake engineering that are invited to make a prediction about the probable damage state of a structure or a class of structures under different levels of seismic action. The fragility curves generated by this judgment-based data are called as “judgmental fragility curves.” Until recently, expert opinion was the main source to generate fragility information especially for a wide range of structures and performance limits in the United States by the introduction of the ATC-13 document (Applied Technology Council 1985) and HAZUS earthquake loss estimation methodology (National Institute of Building Sciences 1999). However, the main issue that impairs the use of expert opinion is the subjectivity and bias of the collected data since it depends on the unique experience of the experts involved and it is very difficult to quantify this subjectivity or eliminate the bias.

Fragility curves can also be derived by using the structural damage data obtained from field observations after earthquakes. The fragility curves obtained by this approach are called as



Seismic Fragility Analysis, Fig. 1 Conversion from a set of fragility curves to DPM (Prepared by using the data from Erberik and Elnashai 2004)

“empirical fragility curves.” Although it is difficult to construct empirical fragility curves from field observations, the obtained fragility curves are invaluable in the sense that they reflect the exact behavior of a structure or a class of structures. The difficulty comes from the fact that for empirical fragility curve generation, an extensive field data from many earthquakes with different magnitudes should be gathered. For instance, Rossetto and Elnashai (2003) gathered a huge database that is composed of 340,000 buildings inspected after 19 different earthquakes and locations in order to construct the empirical fragility curves for European-type reinforced concrete frame buildings. Furthermore, it is not an easy task to estimate the spatial distribution of earthquake intensity at different locations where the buildings under consideration reside. Another

issue is the nonuniformity of field data in terms of ground motion intensities and damage state definitions since it is being collected from many different resources. In this case, it becomes crucial to convert all field data to a standard format in order to use in the generation of empirical fragility curves.

Experimental data can be employed in the generation of fragility curves by using a similar approach as in the case of empirical fragility curve generation based on field data. However, in order to obtain sufficient data for fragility curve generation, it may be required to conduct a large number of laboratory tests with different levels of loading intensities and material properties. Hence, this way of generating fragility curves is generally an expensive solution and therefore not popular when compared to the

generation of fragility curves based on analytical simulations.

The most commonly used data source for quantification of seismic fragility comes from the analytical simulations since most of the issues encountered in the case of judgmental, empirical, and experimental-based approaches are handled by the use of numerical data obtained through analytical modeling. It is possible to generate response statistics for any type of structure or class of structures subjected to a wide range of ground motion intensities. However, in this case, there are two main issues: computational effort and limitations in analytical modeling due to idealization of actual structures. Although the computing technology is very powerful nowadays, generation of analytical fragility curves can take considerable time and effort depending on the level of detail in analytical modeling. In some cases, it can even become unfeasible such as when it is required to run a dynamic analysis thousands of times on a detailed finite element model of a special structure. Hence, this means that it is inevitable to introduce some simplifications and idealizations to the analytical model. But then, the issues arising from limitations of analytical modeling begin to play an important role in seismic fragility analysis. This may cause the deviation of the response of the analytical model from the actual behavior of the structure under concern. As a result, analytical approach to quantify seismic fragility always possesses a trade-off between accuracy and computational effort. In this regard, there are different alternatives to be used as the analysis method for seismic fragility analysis. These can be listed as linear static, linear dynamic, nonlinear static, and nonlinear dynamic methods from the simplest to the most complex approach. Among these, linear static method is not generally used in seismic fragility analysis due to the dynamic nature of the actual response. However, nonlinear static analysis is an alternative when it is difficult to carry out series of dynamic analyses to generate fragility functions. Linear dynamic analysis can sometimes be used, especially when the structural system under concern exhibits brittle behavior under seismic action and does not go far

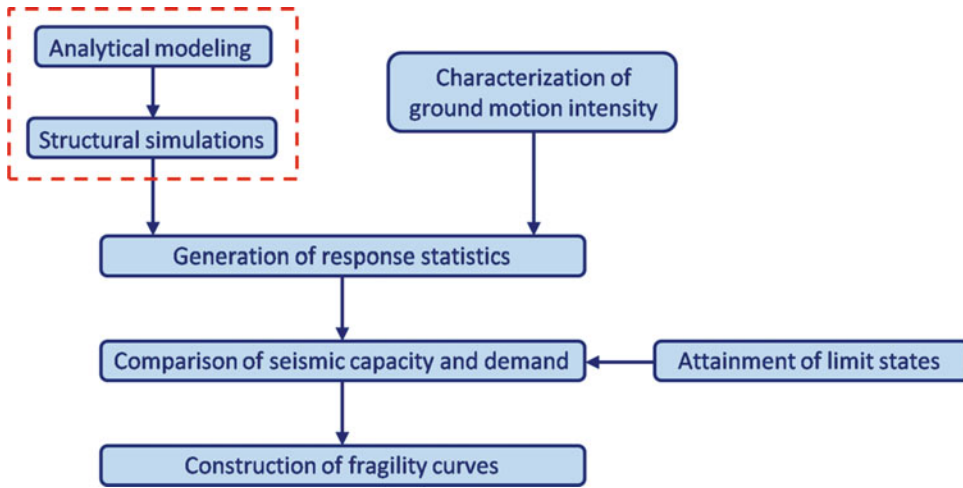
beyond the elastic range of response, like unreinforced masonry structures. The best method is always nonlinear dynamic approach, but in practice, this may not be the case due to computational difficulties as discussed above.

Components of Seismic Fragility Analysis

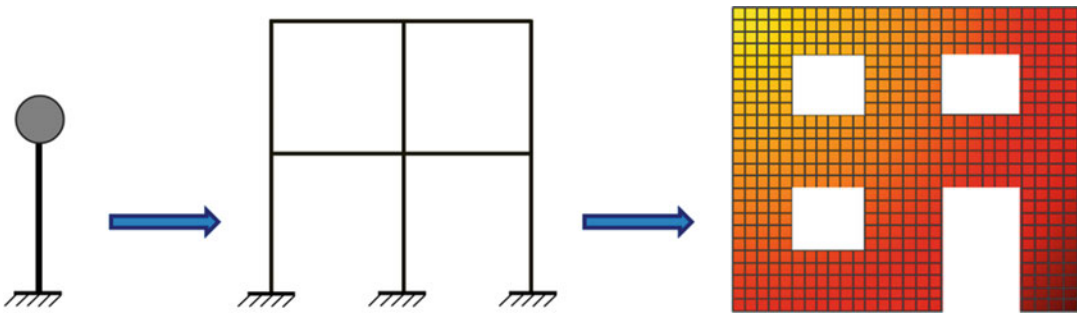
Basically, seismic fragility analysis is the comparison of seismic capacity and demand and to estimate whether the seismic capacity is exceeded for a well-defined performance level when the structural system is subjected to specified levels of ground motion intensity. Due to the probabilistic nature of seismic fragility analysis, both seismic capacity and demand are defined by probability functions in terms of certain random variables to quantify all the uncertainties involved in the process.

In the field of earthquake engineering, there is no consensus on the best method for seismic fragility analysis. However, the general framework of seismic fragility analysis can be more or less presented as shown in Fig. 2. Each element of the flowchart defines a different component of seismic fragility analysis procedure. The elements enclosed by boxes with dashed lines can be regarded as a single component in the analysis. Each component of seismic fragility analysis is explained in detail in the following paragraphs.

The first component is the development of the analytical model and generation of structural simulations to account for structural variability. At this stage, the major challenge is the selection of the analytical model. Depending on the level of idealization and simplification, a wide range of models from equivalent single-degree-of-freedom (SDOF) systems to detailed finite element (FE) meshes can be employed in order to carry out the structural simulations (Fig. 3). Since a huge number of analyses are required to construct the response statistics and the fragility curves, the use of a simple model (like a SDOF model) becomes very advantageous in terms of computational effort. However, it should be noted that such a simple model can only realize



Seismic Fragility Analysis, Fig. 2 General framework of seismic fragility analysis



Seismic Fragility Analysis, Fig. 3 Different levels of idealization for analytical modeling of structural systems from the simplest to the most sophisticated

the global behavior of the structural system to some extent and it is blind to local structural characteristics or construction details. Hence, such models are generally used to generate the fragility curves of a class of structures (for instance, mid-rise reinforced concrete frame structures) in order to predict regional damage or loss. On the other hand, it is not possible to use simplified models for special structures, in which the structural and construction details are important and should be reflected in the seismic fragility analysis. A good example for this case is historical masonry structures, which are unique and possess their own structural characteristics. In such cases, finite element mesh modeling seems to be the best solution, but then the problem is to carry out numerous analyses (especially

if they are dynamic in nature) on this finite element mesh within a reasonable period of time. Briefly, the problem of selecting the appropriate modeling strategy is multidimensional, and one has to make the decision by considering all the pros and cons.

After the selection of the analytical model, the next step is to generate the structural simulations. Due to the probabilistic nature of seismic fragility analysis, some of the major structural parameters within the analytical model are considered as random variables with appropriate probability density functions assigned to them. Normal or lognormal distributions are commonly used for convenience. These can be mechanical properties like stiffness or strength to account for the material variability or geometric properties like

length, height, or cross-sectional dimensions. In order to generate the population of analytical simulations, a sampling method is required. The most popular method is Monte Carlo sampling (Metropolis and Ulam 1949). It is a robust and a straightforward method to generate structural simulations, but the disadvantage is that it requires a very large sample size, which makes the method unfeasible to use in the case of detailed and complex analytical models. This shortcoming opens a door to constrained sampling methods to reduce the sample size, like the Latin hypercube model (McKay et al. 1979).

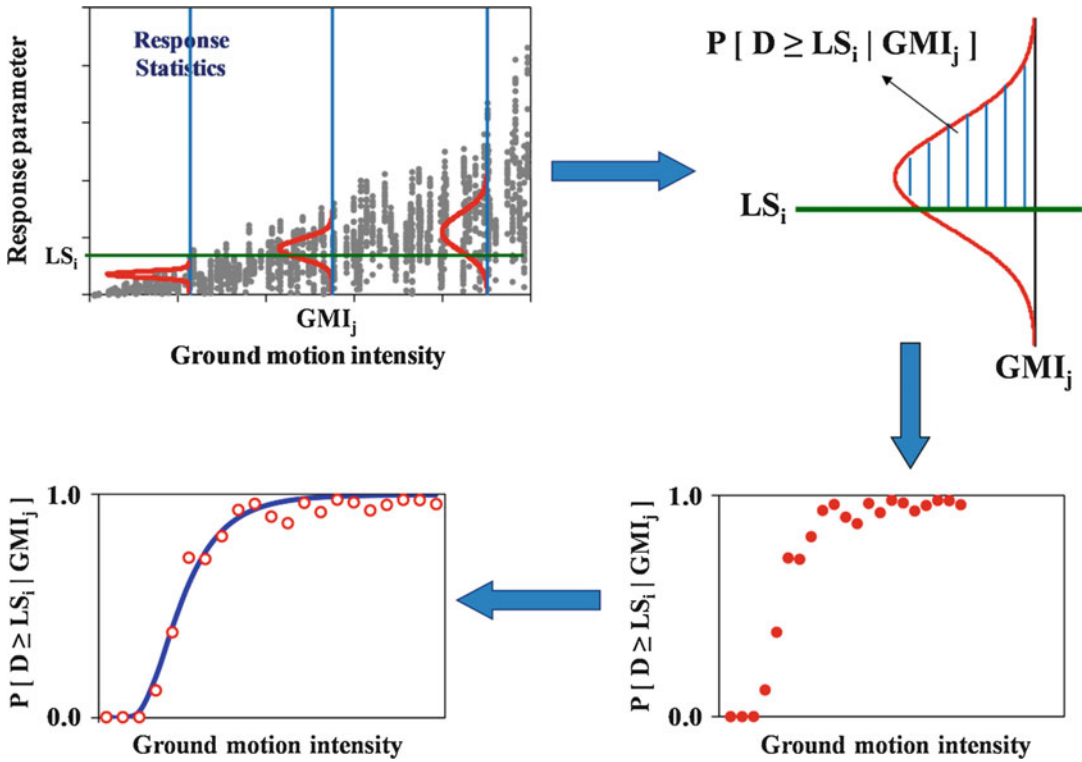
Since dynamic analysis is an indispensable tool for seismic fragility analysis, characterization of ground motion intensity becomes an important component to generate fragility information. To conduct dynamic analyses, there are two options: to employ (1) actual or (2) synthetic ground motion records. If it is possible to find a sufficiently large number of records from different earthquakes in the region of interest, then the first option seems to be the best solution to reflect the effect of regional seismicity and local geologic characteristics. Ground motion records are generally selected to cover the whole range of intensity levels. However, this may induce the need to scale ground motion records to fill the gaps of seismic intensity levels, especially for high intensities. Scaling ground motion records is a different research topic in earthquake engineering, and there are many different methods that one can employ. The second option, using synthetic ground motion records, becomes crucial if there exists an insufficient number of actual ground motion records in the study region. Generating synthetic ground motion records is another huge area of research in earthquake engineering, and there are many different approaches in the literature. However, it should be noted that in order to generate reliable synthetic ground motion records in a region, geologic and seismological parameters should be well studied and known beforehand.

Another challenge in seismic fragility analysis is the selection of the ground motion intensity parameter since it is difficult to determine a unique parameter that fully characterizes

earthquake ground motions. Descriptive parameters like Modified Mercalli Intensity (MMI) had been used in the past, but they are no more popular due to the fact that they are judgmental and subjective measures. In addition, they do not correlate well with damage. The most commonly used group of parameters are single-valued (or peak-valued) parameters like peak ground acceleration (PGA) and peak ground velocity (PGV). These parameters can be easily obtained from the ground motion time history. However, single-valued parameters are insensitive to frequency content and duration of the ground motion record, which is extremely important for obtaining the seismic performance of a structural system in some cases. To overcome the shortcomings of single-valued parameters, spectral parameters like spectral acceleration (S_a) and spectral displacement (S_d) can also be employed since these are enhanced parameters which are functions of both the ground motion characteristics and SDOF system properties. Although rare, there are also cases where involved parameters like effective peak acceleration, Housner's spectral intensity, and input energy have been used as the intensity parameter in seismic fragility analysis.

After generating the structural simulations and selecting the ground motion records, the next step is to carry out a large number of analyses in order to build up the response statistics. These are scattered plots in which the vertical axis stands for a response parameter like displacement, drift, or force, whereas the horizontal axis represents the ground motion intensity parameter (see Fig. 4a). All the uncertainty and the variability arising from the ground motion characterization and the analytical simulation is reflected in this plot. Although scattered, the general trend in this plot is that as the ground motion intensity increases, seismic demand characterized by the response parameter also increases.

The response statistics can be employed to generate fragility curves if some limit states are defined as a function of the response parameter under consideration. This is another important stage in seismic fragility analysis because the definition of a limit state directly affects the



Seismic Fragility Analysis, Fig. 4 Schematic representation of the fragility curve generation procedure (Adopted from Ugurhan et al. 2011) (a) response statistics, (b) probability of exceeding a limit state at a

specific ground motion intensity, (c) discrete fragility information for a certain limit state, (d) continuous fragility function

corresponding fragility curve. There are many different ways to attain the limit states. Limit state values can be directly obtained from the previous studies in the literature if the structural type under consideration had been studied before. However, in most of the cases, the local and special characteristics of the structural system requires the attainment of new limit states. In this case, the best method is to rely on field observations or experimental findings if any. In the absence of such data, analytical approaches can be used to determine the limit states. For instance, nonlinear static analysis, in particular pushover method, is a good candidate for this task. The instantaneous capacity of the structural system and the progression of damage can be monitored in this type of analysis.

Generally, two or more limit states are defined in seismic fragility analysis. For instance, in

FEMA-273 document (Federal Emergency Management Agency 1997), three limit states were defined as “Immediate Occupancy,” “Life Safety,” and “Collapse Prevention” in terms of interstory drift. The limit states are generally considered as deterministic parameters. However, there are cases in literature in which limit states are also considered in a probabilistic manner to account for the variability in seismic capacity.

Each vertical line of scattered data in response statistics stands for a specific ground motion intensity level with its own statistical distribution. If the limit states are also added to the same plot, then it becomes possible to determine the probabilities of exceeding the limit states at specific ground motion intensity levels. This is illustrated in Fig. 4b, in which LS_i represents the i^{th} limit state and GMI_j represents the j^{th} intensity

level. Hence, the shaded area in the given distribution denotes the probability of exceeding the i^{th} limit state at the j^{th} ground motion intensity level, i.e., $P[D \geq LS_i | GMI_j]$. The value obtained represents one of the data points in Fig. 4c. If this process is repeated for different intensity levels for the whole range, the discrete fragility information shown in Fig. 4c can be obtained for a certain limit state. Then, repeating the process for different limit states reveals a set of fragility information. The last step is to convert this discrete information into a continuous function by curve fitting. The continuous fragility functions are generally represented by cumulative lognormal distribution (Fig. 4d), although other forms of equations (like exponential function) have also been used in the literature. The shape of the fragility curve is dependent on the overall uncertainties involved in the generation process. Less variability means a steeper fragility curve. In the limit, the fragility curve with no uncertainty (i.e., deterministic case) turns out to be a step function.

Additional Information on Seismic Fragility Analysis

The topic of seismic fragility analysis has been a very popular and active research area in the field of earthquake engineering for the last two decades. Hence, there are many additional issues to be discussed related to this topic. Some of these are mentioned in the following subsections.

Different Analysis Approaches in the Literature

In the previous section, a fragility curve generation procedure that makes use of time history analyses was introduced. However, it should be stated that this is not the only alternative to generate fragility curves. For instance, instead of using time history analysis, incremental dynamic analysis (IDA) can also be used for the generation of fragility curves. In this approach, the structural system under consideration is subjected to a series of nonlinear time history analyses, in which the intensity of the ground motion is

increased until collapse at each run by using scaling methods (Vamvatsikos and Cornell 2002). Hence, it becomes possible to monitor the complete system response under seismic action from linear elastic behavior to collapse by using IDA approach, which makes it a good candidate for seismic fragility analysis.

In literature, there are two different nonlinear static procedures that have replaced dynamic analyses in the generation of fragility curves, named as capacity spectrum method (CSM) and displacement coefficient method (DCM). The CSM was first developed by Freeman et al. (1975) and then became more publicized thanks to the introduced in the ATC-40 document (Applied Technology Council 1996). It has been very popular since then in performance based earthquake engineering. The CSM method is based on the idea to compare a nonlinear capacity curve (obtained from a pushover analysis) in acceleration-displacement response spectrum (ADRS) format with an elastic demand spectrum that is reduced to account for equivalent damping. The intersection point is obtained by trial and error and it is called as “performance point”. The second method, DCM, was first introduced by the FEMA-273 document (Federal Emergency Management Agency 1997). Just like the CSM, it is based on constructing the pushover curve of the structural system under concern. Then, this information is converted into equivalent SDOF response. The final step is to modify the linear elastic response of this equivalent SDOF system by the help of some empirical coefficients that account for MDOF response, inelastic behavior, degradation, and P- Δ effect. The final product is an estimate of the maximum displacement of the structural system, called as “the target displacement.” By using both methods, it is much easier to generate fragility curves since it is only a matter of repeating the pushover analysis rather than carrying out complicated nonlinear time history analyses.

There is also a different class of analysis methods, called as response surface methods, which are employed to reduce the large number of computations in seismic fragility analysis. In these methods, a response function is generated

in terms of the major random variables (ground motion intensity parameter, geometric or mechanical properties) to predict the seismic response of structural systems without the need to carry out complex dynamic analyses. The other advantage of these methods comes from the fact that both seismic demand and capacity can be directly incorporated with their inherent joint probability distributions in the response surface, which can only be achieved by making qualitative physical assumptions in the other analysis approaches.

Quantification of Uncertainty in Seismic Fragility Analysis

Seismic fragility analysis requires the treatment of uncertainties involved in the process since both seismic demand and capacity possess different sources of uncertainty. The uncertainties may either arise from inherent randomness in nature (aleatoric uncertainty), which cannot be reduced, or from lack of knowledge or data (epistemic uncertainty). From another point of view, the uncertainties in seismic fragility analysis may belong to ground motion characteristics and the demand on the structural system or the structural capacity under seismic action, performance level identification, and the limitations in analytical modeling. All sources of uncertainty should be quantified and reduced (if possible) in the generation of fragility curves.

Most of the researchers have stated that the uncertainties arising from ground motion characteristics outweigh the uncertainties due to structural system characteristics and analytical modeling. But, of course, this may not be always the case, depending on the specific properties of seismic action and the structural system under concern.

The reliability of the generated fragility curves can be tested by plotting the confidence intervals for selected levels of confidence with upper and lower bounds. If the uncertainties arising from the components of seismic fragility analysis are significant, the confidence intervals will eventually be large, indicating a poor estimate of performance that is represented by the fragility information.

Combined Fragility Curves

Fragility curves for individual structures can be combined together to yield the fragility information of a class of structures with similar properties. This is especially required for regional damage and loss estimation, to obtain the fragilities of different construction types such as reinforced concrete frame structures, unreinforced masonry structures, etc. In this case, the combined fragility curve can be obtained by using (Shinozuka et al. 2000)

$$F_C(X) = \sum_{i=1}^M P_i F_i(X) \quad (1)$$

where

$$P_i = \frac{N_i}{\sum_{i=1}^M N_i} \quad (2)$$

In these equations, F_C stands for the combined fragility, whereas F_i represents the fragility curve of the i^{th} individual structure and X denotes any ground motion intensity parameter used in the generation of the fragility curve. Parameter N_i is the number of structures that the fragility information F_i belongs to, and the summation in the denominator of Eq. 2 represents the total number of structures in that specific class for which the information will be combined.

Fragility Surfaces

In some cases, it becomes difficult to obtain the seismic fragility of a structural system by using a single ground motion intensity parameter since there is no such unique parameter that is perfectly correlated with seismic damage as mentioned above. Then, it is possible to use two different ground motion intensity parameters together in the development of fragility information, which is called as fragility surface rather than fragility curve. In this case, there are two horizontal axes that define the ground motion intensity and one vertical axis that defines the probability of exceeding a certain limit state. This means the response statistics obtained can be plotted as a surface, not a curve. The dual ground motion intensity parameters for fragility surfaces can be

selected as magnitude and distance, PGA and PGV, spectral acceleration or displacement at two different vibration periods of the structural system (for instance, the first and the second vibration mode), etc.

Summary

Seismic fragility analysis can be regarded as a probabilistic measure for seismic performance assessment of structural components or systems. This analysis concept is very popular since it can be used in different research areas of earthquake engineering. Due to its probabilistic nature, it requires the use of random variables to characterize both seismic demand and capacity. In addition to this, there exist challenges such as the idealization of structure through an analytical model, selection of the ground motion intensity parameters and the analysis approach, attainment of the limit states, and quantification of the involved uncertainties. That is why the researchers in this field have used many different approaches and methods to obtain their own seismic fragility information. But the main goal is always to predict the performance of structural systems with a quantified level of confidence under different levels of seismic action.

Cross-References

- [Analytic Fragility and Limit States \[P\(EDP|IM\)\]: Nonlinear Dynamic Procedures](#)
- [Analytic Fragility and Limit States \[P\(EDP|IM\)\]: Nonlinear Static Procedures](#)
- [Empirical Fragility](#)
- [Estimation of Potential Seismic Damage in Urban Areas](#)
- [Seismic Loss Assessment](#)

References

Applied Technology Council (1985) Report ATC-13: earthquake damage evaluation data for California. Applied Technology Council, Palo Alto

- Applied Technology Council (1996) Report ATC-40: seismic evaluation and retrofit of concrete buildings. Applied Technology Council, Redwood City
- Erberik MA, Elnashai A (2004) Fragility analysis of flat-slab structures. *Eng Struct* 26:937–948
- Federal Emergency Management Agency (1997) NEHRP guidelines for the seismic rehabilitation of buildings. FEMA Publication 273. Federal Emergency Management Agency, Washington, DC
- Freeman SA, Nicoletti JP, Tyrell JV (1975) Evaluations of existing buildings for seismic risk – a case study of Puget Sound Naval Shipyard, Bremerton, Washington. In: *Proceedings of the US National conference on earthquake engineering*, Berkeley, pp 113–122
- McKay MD, Conover WJ, Beckman RJ (1979) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 22:239–245
- Metropolis N, Ulam S (1949) The Monte Carlo method. *J Am Stat Assoc* 44(247):335–341
- National Institute of Building Sciences (1999) HAZUS technical manual. Prepared for Federal Emergency Management Agency, Washington, DC
- Rossetto T, Elnashai A (2003) Derivation of vulnerability functions for European-type RC structures based on observational data. *Eng Struct* 25:1241–1263
- Shinozuka M, Feng MQ, Lee J, Naganuma T (2000) Statistical analysis of fragility curves. *J Eng Mech* 126(12):1224–1231
- Ugurhan B, Askan A, Erberik MA (2011) A methodology for seismic loss estimation in urban regions based on ground-motion simulations. *Bull Seismol Soc Am* 101(2):710–725
- Vamvatsikos D, Cornell CA (2002) Incremental dynamic analysis. *Earthq Eng Struct D* 31(3):491–514

Seismic Fragility of Aging Offshore Platforms

Viswanath Kammula¹ and Srinivas Sriramula²

¹Structural Engineer-in-Training, Soscia Engineering Limited, Toronto, Canada

²Lloyd's Register Foundation (LRF) Centre for Safety and Reliability Engineering, School of Engineering, University of Aberdeen, Fraser Noble Building, Aberdeen, UK

Synonyms

Aging infrastructure; Fragility analysis; Offshore structures; Probabilistic assessment; Seismic risk

Introduction

The importance of infrastructure to both the fabric of society and its economy is becoming increasingly more apparent. The robustness of infrastructure systems can be judged by their capacity to accommodate change over time. However, the rapid rate at which our infrastructure is aging requires new solutions for providing a resilient infrastructure that can last for future generations.

Earthquakes are paramount among the natural hazards impacting the infrastructure. The past several decades have witnessed a series of costly and damaging earthquakes. The events would have even more devastating effects if a major earthquake were to hit an old and unprepared infrastructure. The challenge that engineers today face is not only on how to construct new infrastructure with proper resistance to earthquakes but rather to update older structures that do not have the proper defense should a natural disaster occur. Failing to modernize older infrastructure could result in very serious consequences.

Offshore platforms are one of the key infrastructure systems housing the facilities for exploration and production of oil and gas. In addition to acting as accommodation for offshore workers, these platforms contain a range of equipment from power generation to oil processing. Assessing the reliability of these structural systems is highly complicated due to the complexity of interaction between various elements. It is possible to classify the offshore platforms further as floating and fixed offshore platforms (Sadeghi 2007), with usage depending on the depth of water at the installation location. These structures are popular in the waters of California, the Gulf of Mexico, Nigeria, the North Sea, and the Persian Gulf.

This entry will focus on a very specific and popular type of offshore platform, the fixed steel jacket platform (FSJ). According to Ferreira (2003), there are around 7,500 FSJ platforms distributed around the world. These platforms have been popular because of their relatively simple design and the availability of oil at shallow waters in the initial days of exploration.

However, as the resources in shallow water have started to extinguish, the exploration activities have moved to deeper waters. Nonetheless, there are many fixed steel jacket platforms that are still in service.

When the FSJ platforms were originally designed, the design life was considered as 25 years. However, due to the continued availability of hydrocarbons at some locations, the operators are looking at various options to extend the design life. For example, in the United Kingdom Continental Shelf (UKCS) sector of the North Sea, out of the 288 offshore installations, about 50 % of the fixed platforms have exceeded their design life (Stacey et al. 2008). When it comes to extending the life of these structures, there are various key limit states one needs to consider, depending on the geographical location. For example, the effects of wave fatigue on offshore structures are a key consideration in one zone, whereas the structural response under earthquakes is a serious limit state at a different location.

Considering the complexity and popularity of these structures, this entry looks at ways to characterize the probabilistic response of FSJ platforms under earthquake loads. The seismic vulnerability curve or seismic fragility curve is one of the most important elements in the assessment of seismic damage and the evaluation of the performance of different structural systems. The fragility curves are used to represent the probabilities of structural damage due to earthquakes as a function of ground motion indices such as peak ground acceleration (PGA), spectral acceleration (S_a), and spectral displacement (S_d). The characterization of earthquake excitation and the identification of different damage levels have an important role in developing the fragility curves and understanding the seismic performance of these structures.

This entry provides details of the development of fragility curve for a typical FSJ offshore platform using the SAC-FEMA technique. The following sections provide mathematical details of the SAC-FEMA technique, an overview on the reference structure and establishment of dynamic characteristics. Then, the ground motions used in

fragility analysis and the analysis results are discussed. The entry concludes with suggestions to formulate a better inspection and maintenance framework based on the summary of findings.

Seismic Fragility Analysis of Structures

Theory of Fragility Analysis

Introduction

Seismic damage and loss estimations are essential for disaster planning and formulating damage-reduction and insurance policies. Regional loss estimation can be mathematically described through Eq. 1, as shown below (Kwon 2007):

$$P[Loss] = \sum_s \sum_{LS} \sum_d P[Loss | D = d] \cdot P[D = d | LS] \cdot P[LS | IM = s] \cdot P[IM = s] \quad (1)$$

where $P[\bullet]$ = probability of events in the brackets. The term *Loss* refers to direct or indirect losses from a seismic event; *IM* is an intensity measure of a seismic hazard such as spectral acceleration or peak ground acceleration (PGA); and *s* is a realization of the intensity measure. $P[LS | IM = s]$ is a conditional probability of attaining structural limit state (LS), and $P[D = d | LS]$ is a conditional probability of attaining damage (minor, moderate, or major), which is a qualitative measure. The loss of a region is determined based on the damage state. Among these elements, the term, $P[LS | IM = s]$, refers to strong, motion-shaking severity influencing the probability of reaching or exceeding a specified performance limit state. This term is referred as *vulnerability* or *fragility*. The strong, motion-shaking severity may be expressed as an intensity, peak ground parameters (PGA, PGV, or PGD), or spectral ordinates (S_a , S_v , or S_d) corresponding to a fundamental structural period.

Analysis Techniques

Among others, some practical approaches to perform fragility analysis are the SAC/FEMA

method, effective fragility analysis method, and methods based on response surface (Pinto et al. 2004). The SAC-FEMA method, which provides the basis for the FEMA-350 (2000) for the seismic design and assessment of steel moment-resisting frames, is the most practical approach among these options. The SAC-FEMA method is efficient when the response is predominantly dominated by the first mode and when failure is influenced by a single scalar variable, for instance, the interstory drift. However, when failure involves several random variables, then one would need to use methods based on effective fragility analysis techniques or response surfaces.

Mathematical Details of SAC-FEMA Technique

The SAC-FEMA method (Pinto et al. 2004) is one of the popular methods to perform vulnerability analysis. In the SAC-FEMA approach, the seismic hazard $P[IM = s]$ is defined in terms of spectral acceleration ordinates, calculated at the fundamental period of the structure. In this method the failure occurs when the maximum demand over the duration of the seismic excitation exceeds the corresponding capacity. The seismic hazard is combined with the drift demand to define drift hazard as follows:

$$H_D(d) = \int P[D \geq d | S_a = x] | dH(x) | \quad (2)$$

where $|dH(x)|$ means the absolute value of the derivative of the site's spectral acceleration hazard curve times dx . The seismic demand is related to the hazard as follows:

$$\hat{D} = a S_a^b \quad (3)$$

where $\hat{D}$ is the median value of demand *D*. The constants, *a* and *b*, are determined from a regression analysis of structural demands calculated from the dynamic analyses. The term β_D , standard deviation of the natural logarithm of demand *D*, is calculated about the median value of demand *D*. Ideally, at each spectral acceleration value, there would be multiple demand values, as two ground motions with same spectral acceleration could result in different drifts. Hence, the

median value over each set of demands is corresponding to a spectral ordinate value that needs to be calculated. Also every set is associated with a unique dispersion. To account for the dispersion, a random variable ϵ , with unit median and dispersion equal to β_D , is introduced. Equation 4 then updates to:

$$D = aS_a^b \epsilon \quad (4)$$

However, for simplicity, only one demand is used for a given spectral acceleration, and it is assumed that this value would be the median demand value, and the dispersion is assumed to be the same across all the spectral acceleration values. The demand D is assumed to be log-normally distributed about the median, with standard deviation of the natural logarithm equal to β_D . For a given intensity level, the probability of reaching a structural demand, d , can be calculated using cumulative distribution function of log-normal distribution, given as:

$$P[d \leq D | S_a = x] = 1 - \Phi\left(\frac{\ln(d/ax^b)}{\beta_{D|S_a}}\right) \quad (5)$$

Equations 4 and 5 result in a closed form solution of drift hazard as shown below:

$$H_D(d) = P[D \geq d] = H(S_a^d) \exp\left[\frac{1}{2} \frac{k^2}{b^2} \beta_{D|S_a}^2\right] \quad (6)$$

The drift hazard should be combined with the drift capacity, C , to evaluate the annual probability of the performance level not being met, P_{PL} .

$$P_{PL} = \int P[C \leq d] |dH_D(d)| \quad (7)$$

The drift capacity C is assumed to follow log-normal distribution with a median value of $\hat{C}$ and dispersion β_C . Therefore, the probability of the capacity, C , being lower than the demand, d , is expressed as

$$P[C \leq d] = \Phi\left(\frac{\ln(d/\hat{C})}{\beta_C}\right) \quad (8)$$

Substituting Eqs. 6 and 8 into 7, the probability of annual failure is given as

$$P_{PL} = H\left(s_a^{\hat{C}}\right) \exp\left[\frac{1}{2} \frac{k^2}{b^2} \left(\beta_{D|S_a}^2 + \beta_C^2\right)\right] \quad (9)$$

Previous studies have shown that P_{PL} is dictated by the hazard and not by the uncertainties in the demand and capacity. The above procedure, summarized in Eq. 9, fully couples hazard, demand, and capacity. However, as highlighted in Kwon (2007), the derivation of vulnerability curves is more advantageous than calculating fully coupled risks.

Application to a Typical Fixed Steel Jacket Platform

Analytical Modeling

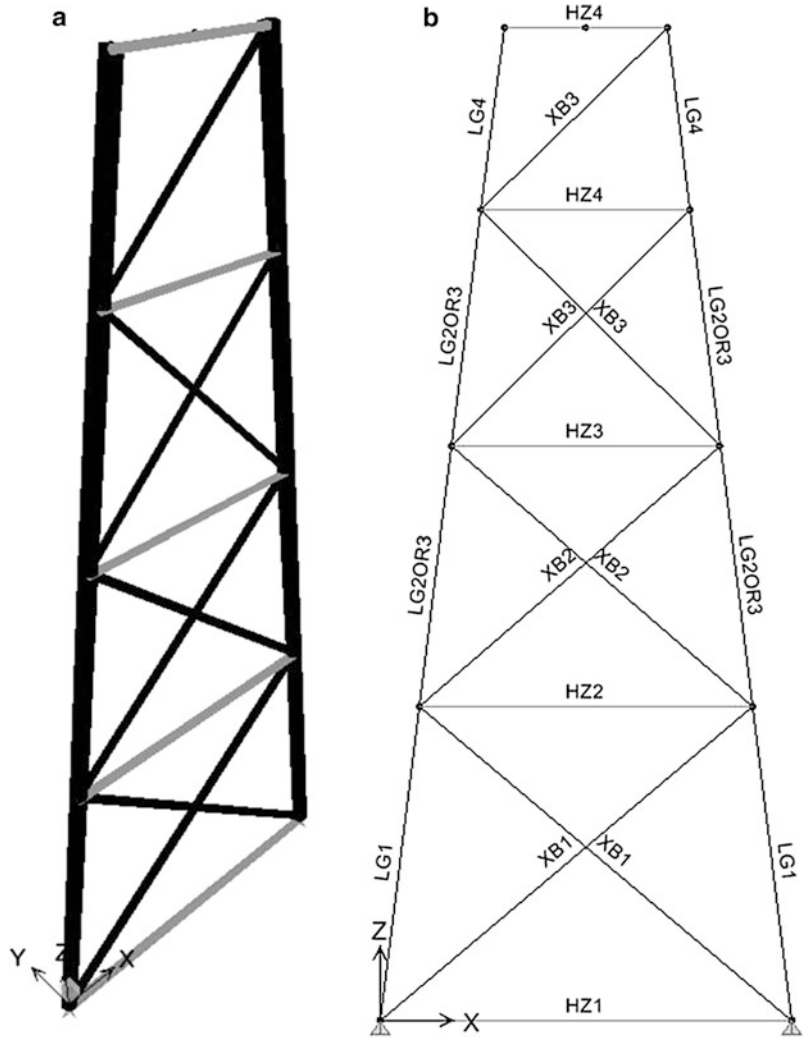
Fragility analysis was performed on a typical FSJ offshore platform that was designed according to the API guidelines and presented by Nordal et al. (1987). It is assumed that all the lateral forces are transferred to the foundation mainly by the diagonal braces. Furthermore, at each level, the horizontal X-braces transfer the forces among the four vertical frames. In this study, the performance of the platform is estimated through fragility analysis when subjected to seismic forces. A 2-D model of the X-braced vertical frame on grid line A is developed in ETABS (2008); the section sizes are available in Nordal et al. (1987). A screenshot of this analytical model is shown in Fig. 1.

The numerical model of the structure is developed based on the following assumptions:

- All the connections are assumed to be pinned connections.
- All the members are modeled as truss elements.
- Rigid diaphragm condition is assumed only for the level 4, where the mass is defined.
- All supports are assigned pin conditions.

Seismic Fragility of Aging Offshore Platforms,

Fig. 1 Analytical model in ETABS (a) 3D view (b) elevation



- The primary lateral force-resisting members are the braces.
- Nonlinearity effects are not modeled.

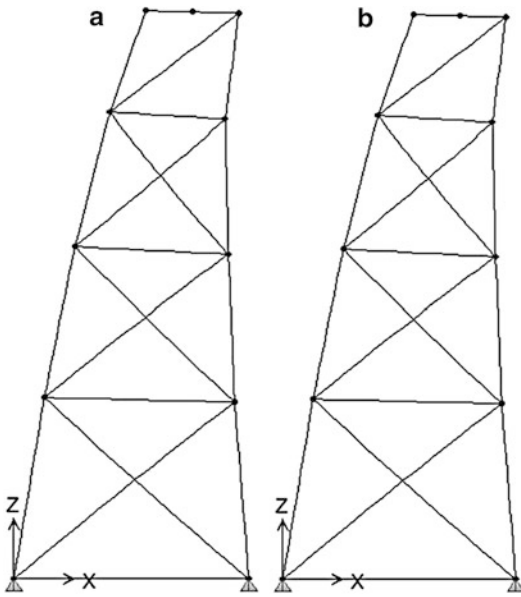
Based on the work of Stear and Bea (1997), the total deck load being supported is taken as 51 MN, which would be shared by the four frames in the y -direction. Hence, a mass of 12.75 MN (12,750 kN) is assigned to the top node of the frame. In addition, as discussed in the modeling assumptions, rigid diaphragm condition (all the nodes in a given floor are constrained together) is assigned for the top level.

Evaluation of Dynamic Properties

Using the above defined mass and section sizes, a modal analysis is performed to obtain the fundamental period. First two mode shapes of this structure along with their corresponding periods are shown in Fig. 2. The fundamental period was also determined using static pushover analysis to verify the period obtained using the modal analysis. The fundamental period obtained using both the analyses are in good agreement.

Selection of Ground Motions

For demonstrative purposes, ground motions are selected from the Pacific Earthquake Engineering



Seismic Fragility of Aging Offshore Platforms, Fig. 2 Mode shapes of the frame (a) Mode 1 ($T = 3.71$ s) (b) Mode 2 ($T = 0.123$ s)

Center (PEER) New Generation Attenuation (NGA) database. However, it is recommended to select the ground motions based on the site location, if available. The database is composed of 3,551 ground motion recordings that represent over 160 seismic events (including aftershock events) ranging in magnitude from M4.2 to M7.9 (Chiou et al. 2008). A total of 30 far-field ground motions are selected from the PEER NGA database. The PGA range for the final selected suite of ground motions, listed in Table 1, is 0.2 g to 1.8 g with an average of 0.5 g. The ground motion suite includes earthquakes that occurred from 1971 to 1999. The spectral acceleration at the fundamental period of the structure, 3.71 s corresponding to 5 % damping, is obtained by constructing the elastic response spectra of each ground motion (Seismosoft 2011). For instance, for the first ground motion in Table 1, Loma Prieta earthquake (A01090), the pseudo-acceleration elastic spectra is shown in Fig. 3.

Calculation of Story Drift

For a linear elastic system, the peak story drift can be obtained from the elastic pseudo-spectral

acceleration using the following equation (Chopra 2010):

$$A = \omega_n^2 D = \left(\frac{2\pi}{T_n} \right)^2 D \quad (10)$$

where ω_n is the natural frequency of the structure, D is the peak deformation of the structure due to earthquake, T_n is the natural period of the structure, and A is the pseudo-spectral acceleration corresponding to T_n . The peak displacement of the top node, obtained above, is divided with total story height to obtain the peak drifts. The graph of peak drift versus spectral acceleration is shown in Fig. 4 for the ground motions listed above.

In general, the pseudo-spectral acceleration is very low because of the higher fundamental period. When compared to buildings, the percentage increment in the mass must be higher than the percentage increment in stiffness for the offshore structures. Having obtained the seismic hazard expressed in spectral acceleration and the demand expressed in interstory drift, a reliability analysis can now be performed. The limit states are discussed in the following section.

Formulation of Limit States

The performance limit states are of prime importance in the fragility analysis and need to be carefully determined. The three performance levels that were used for the offshore steel-braced frame are given in Table 2. However, more investigation is required to ascertain the applicability of these limit states to offshore structures. The limits corresponding to each level are also given in Table 2.

Fragility Analysis Results

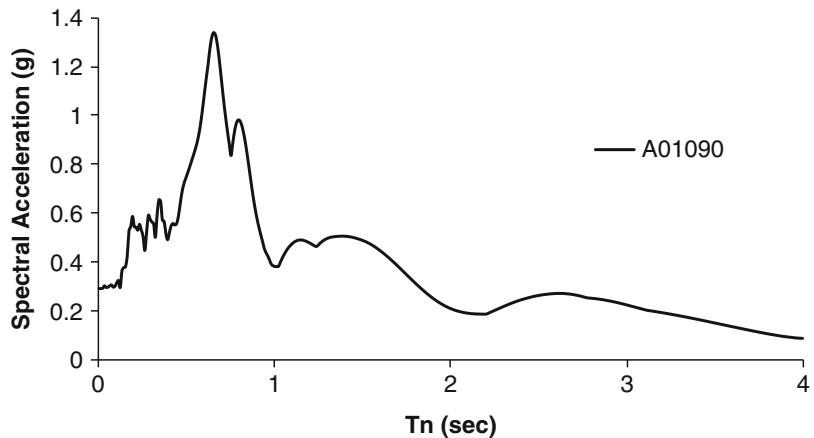
Using the outlined procedure in section “[Mathematical Details of SAC-FEMA Technique](#),” the fragility curves are developed for the three limit states discussed in the previous section and are shown in Fig. 5. It can be inferred that the probability of reaching the higher limit state, characterized with drift 2 %, is almost 100 % for a seismic hazard as low as 1.2 g, which signifies the non-robust seismic design of the structure.

Seismic Fragility of Aging Offshore Platforms, Table 1 Considered ground motions (PEER: NGA Database 2008)

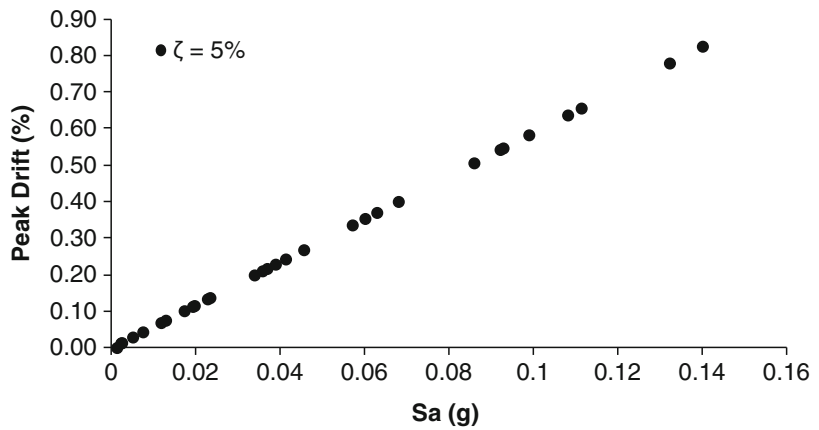
No	Name	PGA, g	Earthquake name	Year	Station name
1	A01090	0.294	Loma Prieta	1989	Foster City – APEEL 1
2	ABRD130	0.313	Whittier Narrows-01	1987	Brea Dam (Downstream)
3	AING000	0.299	Whittier Narrows-01	1987	Inglewood – Union Oil
4	AND250	0.244	Loma Prieta	1989	Anderson Dam (Downstream)
5	ATAR090	0.644	Whittier Narrows-01	1987	Tarzana – Cedar Hill
6	CH10270	0.269	Loma Prieta	1989	Oakland – Outer Harbor Wharf
7	CHY014W	0.229	Chi-Chi, Taiwan	1999	CHY014
8	CHY080N	0.218	Chi-Chi, Taiwan-03	1999	CHY080
9	G03090	0.367	Loma Prieta	1989	Gilroy Array #3
10	Go2000	0.363	Loma Prieta	1989	Gilroy Array #2
11	Go2090	0.321	Loma Prieta	1989	Gilroy Array #2
12	H-Z14090	0.274	Coalinga-01	1983	Parkfield – Fault Zone 14
13	HDA165	0.269	Loma Prieta	1989	Hollister Diff. Array
14	HDA255	0.279	Loma Prieta	1989	Hollister Diff. Array
15	HEC090	0.337	Hector Mine	1999	Hector
16	LAC180	0.316	Northridge-01	1994	LA – City Terrace
17	CHY028-E	0.653	Chi-Chi, Taiwan	1999	CHY028
18	CHY028-N	0.795	Chi-Chi, Taiwan	1999	CHY028
19	H-PVY045	0.592	Coalinga-01	1983	Pleasant Valley P.P. – yard
20	PET090	0.662	Cape Mendocino	1992	Petrolia
21	TAR360	0.990	Northridge-01	1994	Tarzana – Cedar Hill A
22	TCU074-E	0.597	Chi-Chi, Taiwan	1999	TCU074
23	ORR090	0.568	Northridge-01	1994	Castaic – Old Ridge Route
24	ORR291	0.268	San Fernando	1971	Castaic – Old Ridge Route
25	ORR360	0.514	Northridge-01	1994	Castaic – Old Ridge Route
26	PEL090	0.231	San Fernando	1971	LA – Hollywood Stor FF
27	PEL180	0.174	San Fernando	1971	LA – Hollywood Stor FF
28	PEL360	0.358	Northridge-01	1994	LA – Hollywood Stor FF
29	TCU071-N	0.380	Chi-Chi, Taiwan-03	1999	TCU071
30	TAR090	1.779	Northridge-01	1994	Tarzana – Cedar Hill A

Seismic Fragility of Aging Offshore Platforms,

Fig. 3 Pseudo-acceleration elastic response spectra ($\zeta = 5\%$)



Seismic Fragility of Aging Offshore Platforms, Fig. 4 Peak drift versus spectral acceleration



Seismic Fragility of Aging Offshore Platforms, Table 2 Structural performance levels and damage, expressed in terms of drift (Source: ASCE 2007)

Elements	Type	Structural performance levels		
		Collapse prevention	Life safety	Immediate occupancy
Braced steel frames	Primary	Extensive yielding and buckling of braces Many braces and their connections may fail	Many braces yield or buckle but do not totally fail Many connections may fail	Minor yielding or buckling of braces
	Secondary	Same as primary	Same as primary	Same as primary
	Drift	2 % transient or permanent	1.5 % transient 0.5 % permanent	0.5 % transient negligible permanent

The slope of the fragility curve decreases with the limit state which implies that relatively higher hazard increment is necessary to reach the probability of failure as the limit state is increased.

Summary

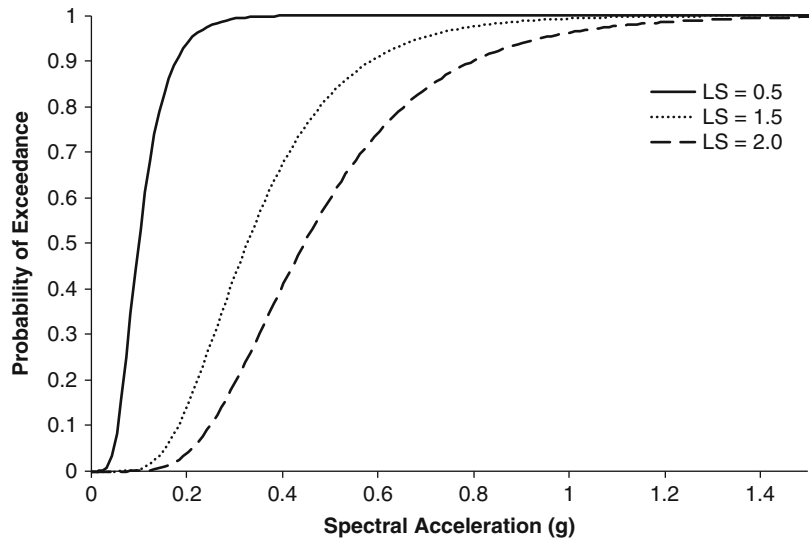
In this entry, the seismic performance of a typical fixed steel jacket platform is evaluated through fragility analysis. The analysis has shown that the considered offshore platform cannot withstand even the seismic hazard corresponding to the collapse prevention performance level. This undermines the robustness of such structures which are aging, for the ever-increasing seismic hazard. Even though the results were shown on a simple elastic system, the authors believe that the implications could be more serious when an inelastic system was used, as the drift demand would be much higher albeit with a reduction in

force demand. Though the fragility curves are derived where the demand is expressed in terms of interstory drift, similar curves can be derived when the demand is expressed in terms of acceleration, residual drift, etc.

The procedure described herein to perform fragility analysis is simple and practical. Hence, when assessing the performance of similar FSJ platforms, practicing engineers can implement it easily. The above results may also be used to prioritize the retrofiting strategy of aging platforms. Fragility curves provide a quick, yet, sufficient insight in to the performance robustness of the structures and help to determine the members that need the utmost attention. This is also ideal for planning the inspection and maintenance strategies for these structures. It is to be noted that other significant limit states also influence such decisions for FSJ platforms; the presented approach is recommended for characterizing the seismic effects along with appropriate consideration of other limit states.

Seismic Fragility of Aging Offshore Platforms,

Fig. 5 Seismic fragility curves for interstory drift



Acknowledgments Sriramula's work within the Lloyd's Register Foundation Centre for Safety and Reliability Engineering at the University of Aberdeen is supported by Lloyd's Register Foundation, a UK-registered charity and sole shareholder of Lloyd's Register Group Ltd.

Cross-References

- Analytic Fragility and Limit States [P(EDP|IM)]: Nonlinear Static Procedures
- Modal Analysis
- Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers
- Seismic Collapse Assessment
- Seismic Fragility Analysis
- Seismic Risk Assessment, Cascading Effects

References

- ASCE (2007) Seismic rehabilitation of existing buildings. ASCE/SEI Standard 41-06, Reston
- Chiou B, Darragh R, Gregor N, Silva W (2008) NGA project strong-motion database. *Earthq Spectra* 24(1):23–44
- Chopra AK (2010) Dynamics of structures—theory and applications to earthquake engineering. Prentice-hall, New Jersey 07458, U.S.A.
- ETABS “Version 9.5.0” (Software) (2008) Computers and Structures, Berkeley. www.csiberkeley.com
- FEMA 350 (2000) Recommended seismic design criteria for new steel moment-frame buildings. NEHRP Program
- Ferreira DD (2003) Fiscal treatment decommissioning and bonds: anticipating impacts of financial assurance requirements for offshore decommissioning, a decision model for the oil industry. Doctoral dissertation, State University Campinas, Brazil, pp 164–168
- Kwon O (2007) Probabilistic seismic assessment of structure, foundation, and soil interacting systems. PhD dissertation, University of Illinois at Urbana-Champaign
- Pacific Earthquake Engineering Research Center (PEER): NGA Database (2005) The Regents of the University of California (Regents). <http://peer.berkeley.edu/nga/>. Accessed 10 Nov 2013
- Nordal H, Cornell CA, Karamchandani A (1987) A structural system reliability care study of an eight-leg steel jacket offshore production platform. In: Proceedings of The Marine Structural Reliability Symposium, SNAME, Arlington, VA, October 1987.
- Pinto PE, Giannini R, Franchin P (2004) Seismic reliability analysis of structures. IUSS Press, Pavia
- Sadeghi K (2007) An overview of design, analysis, construction and installation of offshore petroleum platforms suitable for cyprus oil/gas fields. *GAU J Soc Appl Sci* 2(4):1–16
- SeismoSignal “Version 4.3.0” (Software) (2011) Seismosoft, Pavia. www.seismosoft.com
- Stacey A, Birkinshaw, M, Sharp JV (2008) Life extension issues for ageing offshore installations. In: American Society of Mechanical Engineers, proceedings of the international conference on offshore mechanics and arctic engineering, vol 5, pp 199–216
- Stear JD, Bea RG (1997) Earthquake analysis of offshore platform, screening methodologies. Technical Report to Joint Industry Project Sponsors, Project Phase III, Department of Civil Engineering, University of California at Berkeley

Seismic Instrument Response, Correction for

Jens Havskov¹ and Gerardo Alguacil²

¹Department of Earth Science, University of Bergen, Bergen, Norway

²Instituto Andaluz de Geofísica, University of Granada, Granada, Spain

Synonyms

Accelerometer; Amplitude response; Digitizer; Fourier transform; Phase response; Poles and zeros; Response function; Seismometer

Recording seismic signals, whether in digital or analog form, gives a number in a computer or an amplitude on paper (Fig. 1). For velocity sensors this number is proportional to ground velocity above a certain frequency, and for accelerometers the number is proportional to ground acceleration.

What is really needed is the true motion of the ground in displacement, velocity, or acceleration. So considering Fig. 1, the signal to the right is recorded, and the signal to the left is what is needed. They are obviously different. Seismologists most often use displacement while engineers prefer acceleration. In most cases the correction is not as easy as just multiplying with a constant to get the ground motion for two reasons:

- The seismic sensor is not recording the desired ground motion magnitude: displacement, velocity, or acceleration and a transformation must be done.
- The response of the seismic sensor to the ground motion, as well as the response of the recorder, is in general frequency dependent, so a frequency-dependent correction must be made (Fig. 1).

It is thus essential to know how the seismic sensor and recorder modify the input signal, the ground motion, to produce the recorded output signal. If, e.g., for given harmonic ground

amplitude displacement $X(\omega)$, the output amplitude $Y(\omega)$ can be calculated as

$$Y(\omega) = X(\omega) \cdot A(\omega) \quad (1)$$

where $A(\omega)$ is the displacement amplitude response, then the displacement, $X(\omega)$, of the ground can simply be calculated as

$$X(\omega) = A(\omega)/Y(\omega) \quad (2)$$

$A(\omega)$ has traditionally been called the magnification, since for an analog recorder $A(\omega)$ gives how many times the signal is magnified at different frequencies.

In this document, a description of the response function for different kinds of sensors will be described, as well as methods used for making a correction for the instrument response to arrive at the true ground motion, whether in time domain or frequency domain.

The Elements of a Seismic Station

A seismic station consists of a sensor connected to a recorder. The sensor has an output in volts proportional to the ground velocity (velocity sensor) or to the ground acceleration (an accelerometer), and this output is generally only constant in a specific frequency range. The proportionality constant is called the generator constant G and has units of $V/(m \cdot s^{-1})$ and V/g (g is the gravity acceleration) for the two types of sensors, respectively. The sensor output signal enters the digital recorder which will have some filter built in at high frequency to avoid aliasing. The digitizer will give out a number (called counts) proportional to the input voltage, and this defines the digitizer sensitivity in terms of counts/V.

Frequency Response of a Linear System

A seismic sensor and associated electronics in the seismic recorder is assumed to behave as a linear



Seismic Instrument Response, Correction for, Fig. 1 The input is the ground moving with a maximum displacement 0.02 mm. This “signal” goes through an

accelerometer and becomes a new signal with a maximum number (count) of 1,250,000. Note also that the frequency content is different for input and output

system. The linearity means that there is a linear relationship between input signal and output signal. If the input signal is $x(t)$ and the output $y(t)$, then multiplying $x(t)$ with a constant will result in an output signal multiplied with the same constant. For example, if the ground velocity is doubled, then the output from the seismometer is also doubled. If two signals of different frequency and amplitude are input, then also two signals with the same frequencies (with different amplitude and phase) are output. The frequency-dependent relation between the output and the input is called the frequency response function. For a more complete description, see Scherbaum (2007).

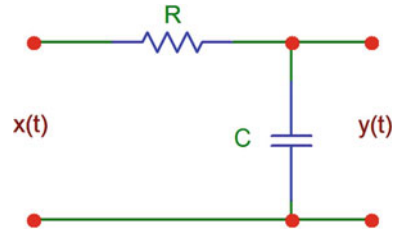
A very simple linear system is the RC filter consisting of a resistor R and a capacitor C . An RC low-pass filter is seen in Fig. 2.

This circuit lets low frequencies pass without attenuation, while high frequencies are attenuated due to the frequency-dependent impedance of the capacitor. For a sine wave signal of frequency f , the capacitor impedance is the relation between the voltage amplitude at the capacitor terminals and the amplitude of the current flowing through the capacitor.

The impedance Z_c of a capacitor seen by the input for a sine wave signal of frequency f is

$$Z_c = \frac{1}{2\pi fC} = \frac{1}{\omega C} \quad (3)$$

where C is the capacitance (F) and $\omega = 2\pi f$. The voltage over the capacitor is delayed by one fourth cycle (90°) with respect to the current through it. Considering that the RC filter is a frequency-dependent voltage divider, using a monochromatic signal of angular frequency ω and amplitude $X(\omega)$, $x(t) = X(\omega)\cos(\omega t)$, the



Seismic Instrument Response, Correction for, Fig. 2 RC low-pass filter. R is the resistor and C the capacitor. Input is $x(t)$ and output $y(t)$

output signal amplitude $Y(\omega)$ can be written as (e.g., Havskov and Alguacil 2010)

$$Y(\omega) = \frac{1}{\sqrt{1 + \omega^2 R^2 C^2}} X(\omega) = \frac{1}{\sqrt{1 + \omega^2 / \omega_0^2}} X(\omega) \quad (4)$$

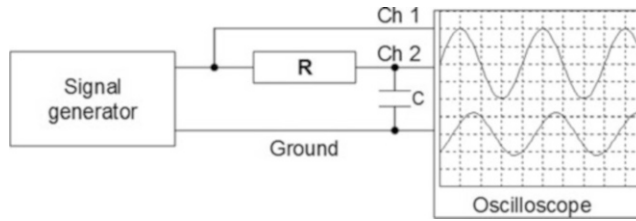
where $\omega_0 = 1/RC$. ω_0 is also called the corner or cutoff frequency of the filter, and for $\omega = \omega_0$, the amplitude has been reduced to $1/\sqrt{2} = 0.707$ (see Fig. 4). From Eq. 4 the amplitude frequency response function of the filter $A(\omega)$ (since only amplitudes are considered) is

$$A(\omega) = \frac{1}{\sqrt{1 + \omega^2 / \omega_0^2}} = \frac{Y(\omega)}{X(\omega)} \quad (5)$$

If $A(\omega)$ is completely known, then the amplitude of the harmonic input signal $X(\omega)$ can be calculated from the measured signal as

$$X(\omega) = Y(\omega) / A(\omega) \quad (6)$$

The amplitude response can be measured very simply as shown in Fig. 3. By varying the frequency, both input and output amplitudes can be



Seismic Instrument Response, Correction for, Fig. 3 Measuring the amplitude and phase response function of an RC filter. The signal from a signal generator

goes directly to channel 1 (Ch1, *top trace*) on the oscilloscope and to channel 2 (Ch2, *bottom trace*) through the filter so both input and output is measured

measured at different frequencies to produce the amplitude response function.

From Fig. 3, it is seen that the output signal not only has been changed in amplitude but also has been delayed a little relative to the input signal; in other words, there has been a phase shift; see definition below.

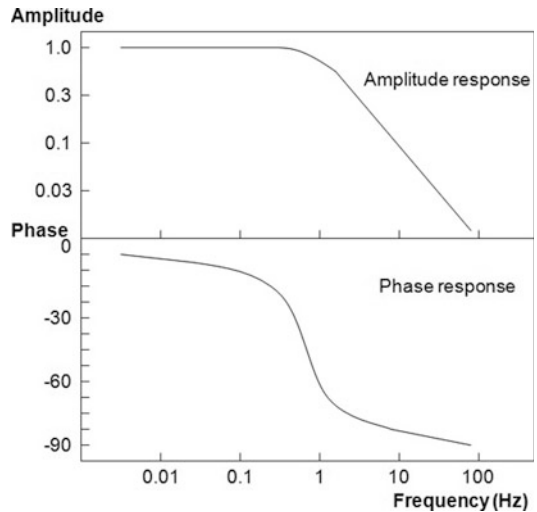
In this example, the phase shift is negative (see definition in Eq. 7). The complete frequency response of the filter therefore consists of both the amplitude response function and the phase response function $\Phi(\omega)$; see Fig. 4. Considering a general input harmonic waveform $x(\omega, t) = X(\omega) \cdot \cos(\omega t)$ at frequency ω , the output can be written as

$$y(\omega, t) = X(\omega) \cdot A(\omega) \cdot \cos(\omega t + \Phi(\omega)) \quad (7)$$

The phase shift is here defined as a quantity being added to the phase as seen above. Thus comparing Fig. 4 and Eq. 7, it is seen that the phase shift is negative. This is the most common way of defining the phase shift, but the opposite sign is sometimes seen, and it may then be called phase delay or phase lag. So it is very important to know which definition has been used.

Using Eq. 7 to correct the amplitude and phase of a single monochromatic signal in time domain is easy, but to make the correction for a whole signal with different frequencies is not simple.

This can be considerably simplified if the complex representation of harmonic signals is used. Instead of writing $\cos(\omega t)$, the real part of the exponential function can be used:



Seismic Instrument Response, Correction for, Fig. 4 Amplitude and phase response of an RC filter with a corner frequency of 1 Hz. The phase response (or phase shift) is given in degrees

$$e^{i\omega t} = \cos(\omega t) + i \sin(\omega t) \quad (8)$$

Equation 7 can now be written as

$$\begin{aligned} y(\omega, t) &= X(\omega)A(\omega)e^{i(\omega t + \Phi(\omega))} \\ &= X(\omega)A(\omega)e^{i\omega t}e^{i\Phi(\omega)} \end{aligned} \quad (9)$$

$y(\omega, t)$ is now a complex number of which the real part is the actual output. This can be further simplified considering that any complex number Z can be written as

$$Z = a + ib = \sqrt{a^2 + b^2}e^{i\Phi} = |Z|e^{i\Phi} \quad (10)$$

where $\Phi = \tan^{-1}(b/a)$ which also follows from Eq. 8. The complex frequency response $T(\omega)$ can then be defined as

$$T(\omega) = A(\omega)e^{i\Phi(\omega)} = |T(\omega)|e^{i\Phi(\omega)} \quad (11)$$

and Eq. 9 can be written as

$$y(\omega, t) = X(\omega)T(\omega)e^{i(\omega t)} \quad (12)$$

Now only one complex function, $T(\omega)$, that includes the phase shift can therefore completely describe the instrument frequency response. So far only monochromatic signals have been used and the corrections can easily be made with both the real and complex representations Eqs. 7 and 12, so it might be hard to see why the complex representation of the response function is needed. Later it is going to be shown how an observed seismogram is corrected, not only for one amplitude at a time, but actually for the whole signal (containing a range of frequencies), which is possible with digital data. Obviously this cannot be done with one amplitude at a time, but will have to be done with the amplitude and frequency content of the whole signal, in other words, making spectral analysis. It will hopefully then be clear why a complex representation is needed.

Frequency Response of Seismic Sensors

There are two important types of seismic sensors:

1. Velocity sensor (seismometer). This sensor has an output proportional to velocity in a limited frequency band; see Fig. 5. This is both the case for the traditional passive seismometers like geophones (for frequencies down to about 1 Hz) and modern broadband seismometers (down to frequencies of 0.003–0.03 Hz).
2. Accelerometer. This sensor usually has an output proportional to ground acceleration from 0 Hz (DC) to some cutoff frequency (see Fig. 6). In some cases there will also be a lower cutoff, e.g., at 0.1 Hz.

Velocity Sensor

The complex response function for a velocity sensor for displacement, velocity, and acceleration is the following (Havskov and Alguacil 2010).

Displacement

$$T_d(\omega) = \frac{Z(\omega)}{U(\omega)}G = \frac{i\omega^3 G}{\omega_0^2 - \omega^2 + i2\omega\omega_0 h} \quad (13)$$

Velocity

$$T_v(\omega) = \frac{\omega^2 G}{\omega_0^2 - \omega^2 + i2\omega\omega_0 h} \quad (14)$$

Acceleration

$$T_a(\omega) = \frac{-i\omega G}{\omega_0^2 - \omega^2 + i2\omega\omega_0 h} \quad (15)$$

where G is the generator constant, ω_0 is the natural frequency of the sensor (the frequency at which it will oscillate if no damping), and h is the damping coefficient (how the motion of the mass is damped; see Fig. 5). It is seen that the only difference between the respective three types of response curves is the factor $i\omega$.

For the velocity response, the amplitude and phase response are (Fig. 5)

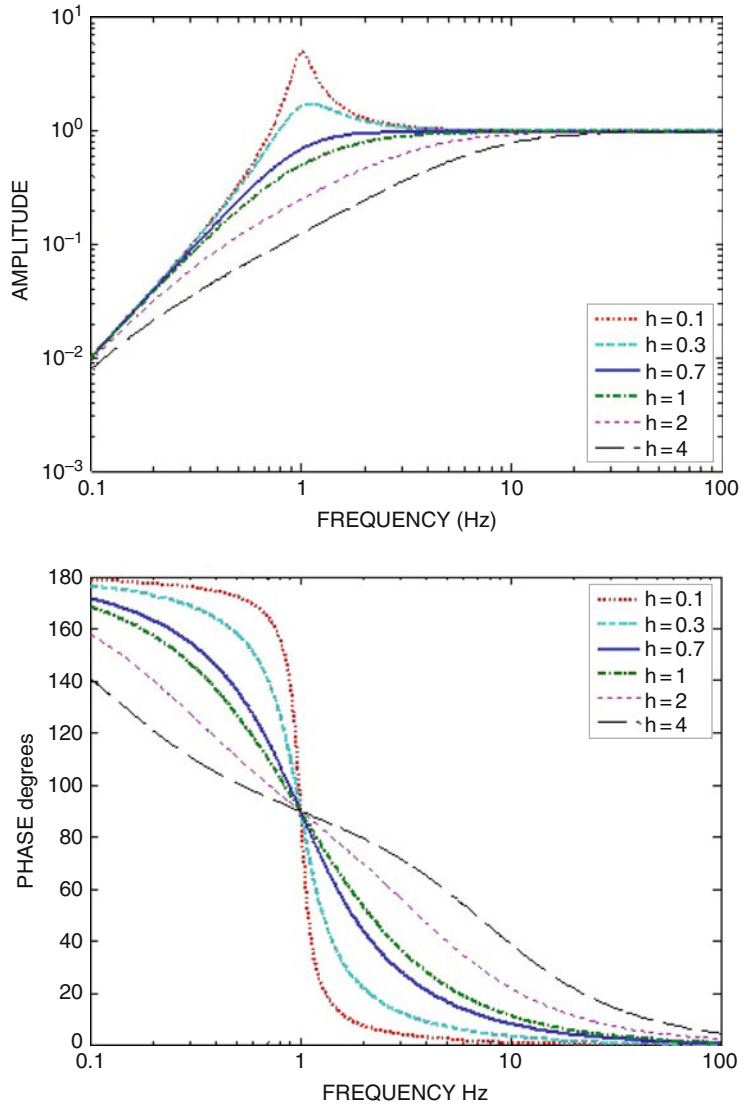
$$A_v(\omega) = |T_v(\omega)| = \frac{G\omega^2}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4h^2\omega^2\omega_0^2}} \quad (16)$$

$$\Phi_v(\omega) = a \tan \left(\frac{\text{Im}(T_v(\omega))}{\text{Re}(T_v(\omega))} \right) = a \tan \left(\frac{-2h\omega\omega_0}{\omega_0^2 - \omega^2} \right) \quad (17)$$

It is seen that as the damping decreases, the sensor response to velocity becomes constant for frequencies above the natural frequency. It is also seen that if there is little damping, the sensor will have a high gain at the natural frequency, so damping is usually set to 0.707.

Seismic Instrument Response, Correction for, Fig. 5

The amplitude and phase response functions for a velocity seismometer with a natural frequency of 1 Hz. Curves for various levels of damping h are shown. It is assumed that the generator constant is 1



Accelerometer

The acceleration response curve for an FBA accelerometer is (Havskov and Alguacil 2010)

$$T_a(\omega) = \frac{G/\omega_f^2}{\omega_f^2 - \omega^2 + i2\omega\omega_f h} \quad (18)$$

where ω_f is the effective natural (angular) frequency with feedback. The amplitude of this response has a flat value of G from DC up to

this frequency (usually above 50 Hz). The instrument correction is then very easy since it is assumed that the feedback natural frequency of the sensor is very high compared to the frequencies of interest so the response for acceleration is simply approximated by multiplying with the generator constant

$$T_a(\omega) = G \quad (19)$$

The velocity and displacement response is then:

Velocity

$$T_v(\omega) = G \cdot i\omega \quad (20)$$

Displacement

$$T_d(\omega) = -G \cdot \omega^2 \quad (21)$$

So for engineering purposes, where acceleration is needed, it is very easy to correct for the accelerometer response since it is just a question of multiplying with a constant, while seismologist wanting displacement must make the transformation (corresponding to integrating the signal twice).

In practice, many accelerograms, especially – but not only – those recorded with earlier accelerographs, are affected by offset or baseline shifts that may influence the estimation of characteristic parameters of strong ground motion (Boore 2001). The source of this long-period noise may be true ground motions (rotation, tilt, local deformation) or instrumental (mechanical nonlinearity, electronic effects) and is not yet fully understood (Akkar and Boore 2009). Prior to any integration to compute the ground velocity or displacement, the accelerogram should be corrected from these effects.

In many practical cases, this correction may consist simply in applying a high-pass filter with a suitable cutoff frequency, as it will be shown below. In some cases, a particular processing scheme adapted to an individual record is required. In fact, most accelerogram databases (e.g., Internet Site for European Strong-motion Data, www.isesd.hi.is) include a “processed” version of each accelerogram, which usually consists of baseline correction, band-pass filtering, and conversion to ground motion units.

It is interesting to note that whether a velocity sensor or an accelerometer is used, it is possible to calculate displacement, velocity, and acceleration from both types of sensors; however, as is shown later, in practice, this will be frequency limited due to noise in the system.

Complete Frequency Response

For passive sensors, the above equations represent exactly the seismometer response, and for many broadband sensors, this is also the case up to a particular frequency. A recording of a seismic signal also involves the seismic recorder which also has a frequency response. Modern recorders have little amplitude change or phase shift in the main frequency band used, so the frequency dependence of the recorder is often ignored. Some recorders might modify the signal enough for a correction to be made. In order to completely describe the response for active sensors and possibly also include the response of the recorder, a more general representation of the response function must be used. It turns out that $T(\omega)$ for all systems made from discrete mechanical or electrical components (masses, springs, coils, capacitors, resistors, semiconductors, etc.) can be represented exactly (e.g., Scherbaum 2007) by rational functions of $i\omega$:

$$T(\omega) = c \frac{(i\omega - z_1)(i\omega - z_2)(i\omega - z_3).....}{(i\omega - p_1)(i\omega - p_2)(i\omega - p_3).....} \quad (22)$$

where c is the combined normalization constant for nominator and denominator polynomials, z are the zeros (or roots) of the nominator polynomial, while the zeros of the denominator polynomial (poles) are p . Using Eq. 22 to represent $T(\omega)$ is the so-called poles and zeros representation, which has become the most standard way.

For example, for the seismometer response function for velocity, $T_v(\omega)$ can be written as

$$T_v(\omega) = \frac{-(i\omega - 0)(i\omega - 0)}{(i\omega - p_1)(i\omega - p_2)} \quad (23)$$

where

$$p_1 = -\omega_0 \left(h + \sqrt{h^2 - 1} \right) \quad (24)$$

$$p_2 = -\omega_0 \left(h - \sqrt{h^2 - 1} \right)$$

So in addition to the poles p_1 and p_2 , the seismometer response function has a double zero at $z = 0$ and the normalization constant is -1 . Note that since h usually is smaller than 1, the poles are usually complex. Complex poles always appear as conjugate pairs. For the displacement response, the equation is

$$T_d(\omega) = \frac{(i\omega - 0)(i\omega - 0)(i\omega - 0)}{(i\omega - p_1)(i\omega - p_2)} \quad (25)$$

and there is thus one more zero. For the standard accelerometer, the displacement response is simply $(i\omega)^2$ which corresponds to two zeros.

For active sensors, there might be more poles and zeros representing the filtering inherent in active sensors.

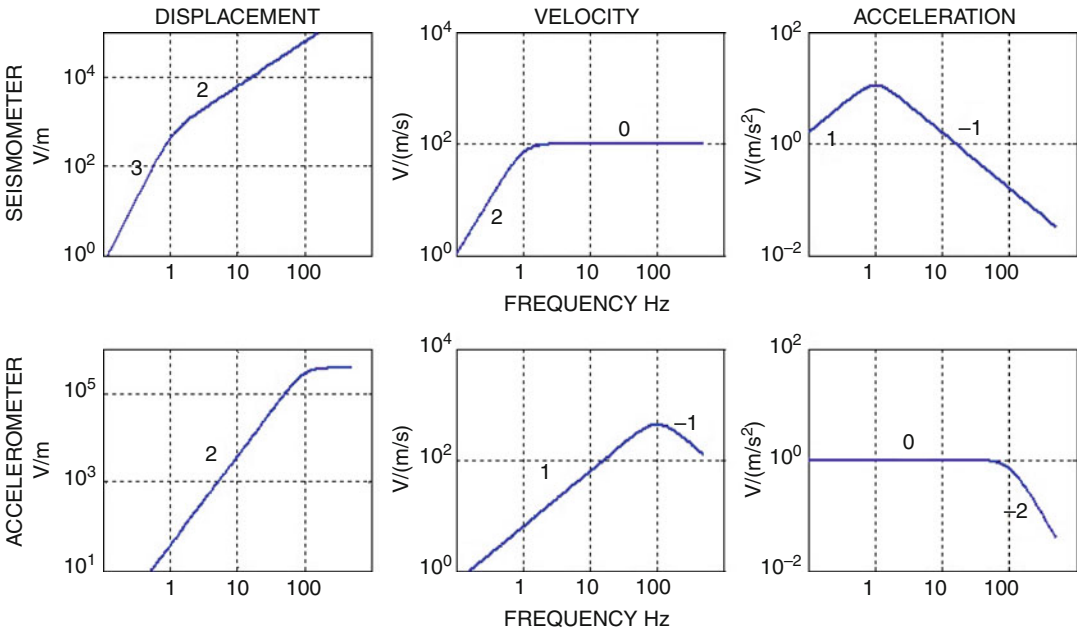
The frequency dependence of a recorder can equally be represented by poles and zeros, so it is

just a question of adding to the equation in the pole and zero representation.

In addition to the frequency-dependent elements, a complete seismograph or accelerograph also has the frequency-independent elements, the generator constant, and the digitizer gain which has to be multiplied with the normalization constant.

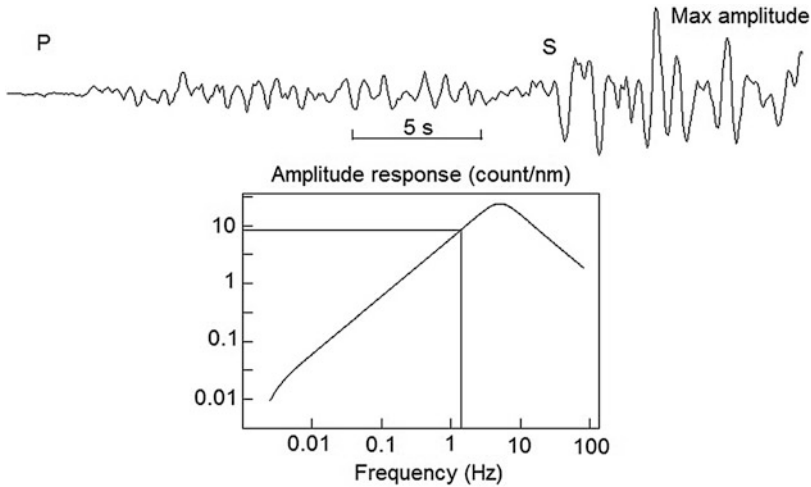
Example

A seismometer has a generator constant $G = 300 \text{ V/ms}^{-1}$ and the digitizer has a sensitivity of 100,000 counts/V. The velocity sensitivity of the seismograph in the passband (flat for velocity) is then $100,000 \text{ counts/V} \cdot 300 \text{ V/ms}^{-1} = 3 \cdot 10^7 \text{ counts/ms}^{-1}$. This number must then be multiplied with the normalization constant for the rest of the system which is the sensor and possibly some filters. For displacement, the amplitude displacement increases with frequency (Fig. 6), but that is



Seismic Instrument Response, Correction for, Fig. 6 Amplitude frequency response of a 1 Hz velocity sensor (top) and an accelerometer (bottom) with a natural frequency of 100 Hz. From left to right, the figures show the sensor response for ground displacement, velocity, and acceleration, respectively. The axes are logarithmic and the horizontal axes show frequency in Hz. The asymptotic

slope for each segment is indicated. Note how one curve translates into another by just adding or subtracting one unit of the slope corresponding to a multiplication or division by frequency. The response curves assume a generator constant $G = 100 \text{ V/(m/s)}$ for the seismometer and $G = 1 \text{ V/(m/s}^2\text{)}$ for the accelerometer. Please note that the vertical scales are different



Seismic Instrument Response, Correction for, Fig. 7 Top: Seismogram of a local earthquake; the position of the maximum amplitude is indicated. The frequency of the maximum amplitude is 1.4 Hz. The signal

is recorded on a broadband station and the amplitude is therefore proportional to ground velocity. Bottom: The amplitude displacement response function. The gain corresponding to the frequency of 1.4 Hz is indicated

taken care of with the $i\omega$ factor so the normalization constant still has to be multiplied with the same number. For example, at 10 Hz, the gain for displacement will be $2\pi \cdot 10\text{s}^{-1} \cdot 3 \cdot 10^7 \text{ counts/ms}^{-1} = 1.88 \cdot 10^9 \text{ counts/m}$.

How to Correct for the Response

In the simplest case, there is only a need to correct for a single amplitude reading. In Fig. 7, a signal from a local earthquake ($M = 5.2$) recorded on a broadband station is seen. Figure 7 shows the corresponding amplitude response curve for displacement. Notice that around 8 Hz, there is a filter cutting off the high frequencies. This filter is part of the sensor electronics and will be represented by poles and zeros in the response function. In the seismogram the maximum amplitude is 1,105,000 counts at a frequency of 1.4 Hz. From the response curve the gain of the system is 8 counts/nm or $8 \cdot 10^9 \text{ counts/m}$. So the maximum ground displacement amplitude is $1,105,000/8 = 138,125 \text{ nm}$. In practical seismology, nm is mostly used for ground motion since it is a more convenient unit than m. In this example, there was no need for a complex response function and the phase

shift was not considered since only one particular amplitude was used.

In the more general case, it is desirable to make instrument correction for the complete seismogram. The seismogram seen in Fig. 7 is proportional to ground velocity in its passband, and a general procedure is therefore needed to calculate the corresponding displacement seismogram. Obviously it is not possible to deal with one frequency at a time and the amplitude and frequency content of the whole signal must be used. In other words, a spectral analysis must be used.

The complex frequency response function $T(\omega)$ is now assumed known and the complex spectra of the input and output signals $x(t)$ and $y(t)$ can be defined as $X(\omega)$ and $Y(\omega)$, respectively, using the definition for the Fourier spectrum (e.g., Oppenheim and Schaffer 1989). Knowing the complete complex output spectrum, the complete complex input spectrum can be obtained:

$$X(\omega) = Y(\omega)/T(\omega) \quad (26)$$

If $T(\omega)$ is a seismic instrument response, the instrument-corrected ground motion spectrum is now obtained, and since $X(\omega)$ is complex, this also includes the correction for phase.

The separate amplitude and phase spectra of the seismic signal can then be obtained as

$$A_{\text{signal}}(\omega) = \sqrt{\text{Re}(X(\omega))^2 + \text{Im}(x(\omega))^2} \quad (27)$$

$$\Phi_{\text{signal}}(\omega) = \tan^{-1} \left(\frac{\text{Im}(X(\omega))}{\text{Re}(X(\omega))} \right) \quad (28)$$

Thus in practice, the ground displacement spectrum would be calculated by taking the complex Fourier spectrum of the output signal and dividing it with the complex displacement response function and finally taking the absolute part of the complex ground displacement spectrum. This is basically what is done to calculate source spectra for earthquakes (Havskov and Ottemöller 2010).

All of this could of course easily have been done with the corresponding noncomplex equations since, so far, the phase information has not been used and usually phase spectra for earthquake signals are not used. However, the next step is to obtain the true ground displacement time domain signal instead of the frequency domain signal. With the knowledge about Fourier transforms, this is now easy, since, when all the frequency domain coefficients or the spectrum has been calculated, the corresponding time domain signal can be generated by the inverse Fourier transform. Since the corrected signal consists of a sum of cosine signals with different phases, each of them delayed differently due to the instrument response, the shape of the output signal will depend on the phase correction. So now life is easy, since the corrected complex displacement spectrum is already corrected for phase and the ground motion can be obtained as

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} Y(\omega)/T(\omega) \cdot e^{i\omega t} d\omega \quad (29)$$

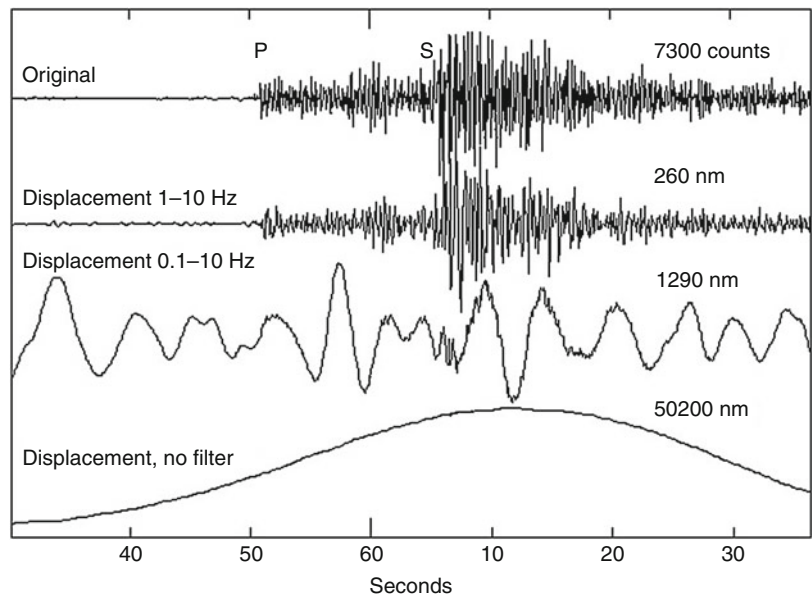
While it is possible to use only half of the positive frequencies for making the amplitude spectrum, both positive and negative frequencies must be used for the inverse transformation. Normalization constants of the Fourier transform can be defined in different ways, but if the same routine

is used for both forward and inverse, the normalization constants will cancel out. In practice, since the signal consists of discrete numbers, the spectrum also is done with discrete numbers; see, e.g., Oppenheim and Schaffer (1989).

Response Correction in Practice

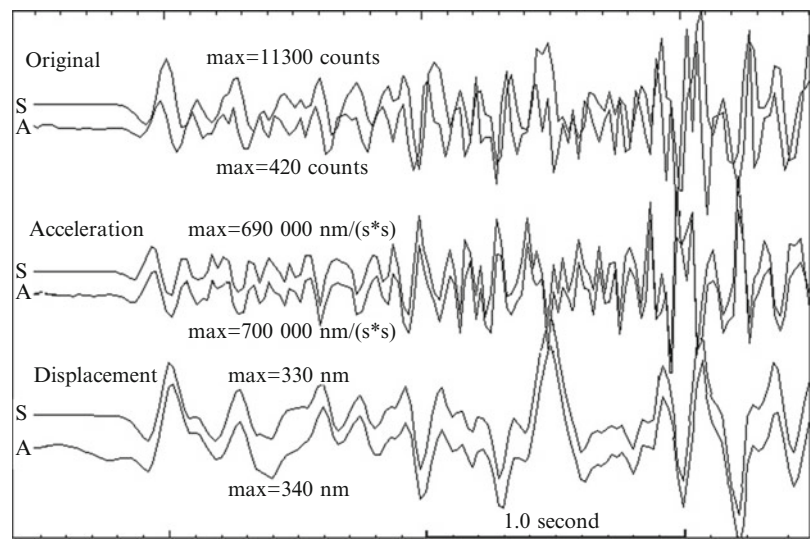
In theory it is possible to recover the ground motion (displacement, velocity, or acceleration) at any frequency knowing the instrument response. In practice, one has to be careful to only do this in the frequency band where the instrument records real ground motion and not just electronic noise, since the instrument correction then becomes unstable and the output has nothing to do with the real seismic signal. Figure 8 shows an example.

The figure shows the influence of filtering, when estimating the ground displacement signal. In the frequency band 1–10 Hz, the signal looks very much like the original signal although a bit more low frequency, since it is converted to displacement and can nearly be considered an integration of the original signal. In the 0.1–10 Hz range, the earthquake signal almost disappears in the microseismic background noise. Why is this signal considered seismic noise and not instrumental generated noise? First, the earthquake signal has about the same amplitude as above; second, it “looks” like seismic background noise; and third, the amplitude at 1,290 nm is at a period of 5 s (peak amplitude of microseismic noise) which looks reasonable compared to worldwide observations. Note that this is how the earthquake signal would have looked being recorded on a broadband sensor, hardly noticeable. The last trace shows the calculation of the displacement without filtering so the lowest frequency used is $1/T$, where T is the length of the window, here 80 s (only 65 s shown) so $f = 0.0125$ Hz. The amplitude is now more than 50,000 nm and the signal looks “funny.” The large amplitude obviously cannot be right since the microseismic noise has its largest amplitude around 3–8 s and it was 1,290 nm. So this is a clear case of trying to make a displacement signal at frequencies lower than where seismic signals exist in the data.



Seismic Instrument Response, Correction for, Fig. 8 Instrument corrections in different filter bands. The top trace is the original recording of a small earthquake with a 1 Hz seismometer. The three bottom traces have been converted to displacement using different

filters. The third trace from the top shows harmonic waves at a frequency of about 0.2–0.3 Hz which is typical microseismic background noise originating in the ocean. The amplitudes to the right are maximum amplitudes (Figure from Havskov and Alguacil (2010))



Seismic Instrument Response, Correction for, Fig. 9 Acceleration and displacement. The seismogram in the figure is the first few seconds of a P-wave. On the site there is also an accelerometer installed (A) next to the seismometer (S). The top traces show the original records in counts. The signal from the seismometer is similar to

the accelerometer signal but of lower dominant frequencies, and the amplitudes are different. The middle traces show the two signals converted to accelerations and the bottom traces converted to displacement (frequency band 1–20 Hz) (Figure from Havskov and Alguacil (2010))

The ratio of the displacement gain for a 1 Hz seismometer at 1 Hz and 0.0125 Hz is $1/0.0125^3 = 5 \cdot 10^5$. In other words, if the gain at 1 Hz is 1.0, the signal must be multiplied by 1.0 to get the displacement, while at 0.0125 Hz, it must be multiplied by $5 \cdot 10^5$. So any tiny amount of instrumental noise present at low frequencies will blow up in the instrument correction. In the above example, it seems that the displacement signal can be recovered down to 0.1 Hz with a 1.0 Hz sensor.

Engineers often think accelerometers are the only instrument to use and seismologist normally will only use the weak motion velocity sensor (seismometer). However, knowing the response curve and the limitation in the data, it does not really matter which instrument is used; see Fig. 9. In this example, a seismic signal has been recorded on a 1 Hz seismometer and an accelerometer at the same site, and using the respective response function, both signals have been converted to displacement and acceleration. The only limitation is that a seismometer may get saturated by a strong motion and an accelerometer will not be sensitive enough for recording weak motions (e.g., from distant earthquakes).

The corrected signals are now very similar and of the same amplitude. This example clearly demonstrates that, with modern instruments and processing techniques, both accelerometers and seismometers can be used to get the same result.

Summary

Recorded seismic signals represent a number proportional with the ground motion. The most common seismic stations have a velocity sensor with an output proportional to ground velocity in some frequency band and/or an acceleration sensor with output proportional with ground acceleration in combination with a digital recorder. All these elements will have some frequency dependence and it has been shown:

- How to calculate the system frequency response
- How to correct for both the sensor and the recorder in order to obtain the true ground

motion, whether in displacement, velocity, or acceleration

- How the noise influences the correction making it necessary to limit the correction to a particular frequency range

Cross-References

- ▶ [Passive Seismometers](#)
- ▶ [Principles of Broadband Seismometry](#)
- ▶ [Recording Seismic Signals](#)
- ▶ [Seismic Accelerometers](#)
- ▶ [Sensors, Calibration of](#)

References

- Akkar S, Boore D (2009) On baseline corrections and uncertainty in response spectra for baseline variations commonly encountered in digital accelerograph records. *Bull Seism Soc Am* 99:1671–1690. doi:10.1785/0120080206
- Boore D (2001) Effect of baseline corrections on displacements and response spectra for several recordings of the 1999 Chi-Chi, Taiwan, Earthquake. *Bull Seism Soc Am* 91:1199–1211
- Havskov J, Alguacil G (2010) Instrumentation in earthquake seismology. Springer, Dordrecht, 358 pp
- Havskov J, Otemöller L (2010) Routine data processing in earthquake seismology. With sample data, exercises and software. Springer, Dordrecht, 347 pp
- Oppenheim AV, Schaffer RW (1989) Discrete-time signal processing. Prentice-Hall, Englewood Cliffs, 870 p
- Scherbaum F (2007) Of poles and zeros, fundamentals of digital seismology, revised 2nd edn. Springer, Dordrecht, 271 pp

Seismic Loss Assessment

Nilesh Shome

Model Development, Risk Management
Solutions, Newark, CA, USA

Synonyms

Damage; Insurance; Loss; Nonstructural component; Risk; Structural component; Vulnerability

Introduction

It is observed that earthquakes, particularly those of large magnitudes, can cause significant damage to structures over a large region surrounding the rupture zone of earthquakes. The top three costliest earthquakes in the world during 1980–2012 are the Mw 9.0 Tohoku-Oki earthquake (2011), Mw 6.8 Kobe earthquake (1995), and Mw 6.7 Northridge earthquake (1994) causing overall economic losses when occurred about \$210B, \$100B, and \$44B respectively (III 2013). The structures those are damaged due to earthquakes are not only buildings, bridges, etc., but also utilities like water, gas and sewer lines, and lifelines like train tracks and roads. The earthquakes also cause landslides due to movement of grounds or liquefaction due to loss of strength of the soil underneath the structures. In addition, earthquakes can cause fire and tsunami leading to significant damage to structures. All the historical losses illustrate that the potential for earthquake loss can be significant in a specific region or to portfolio of buildings for insurance companies. Hence it is important for the federal and state governments or for the insurance companies to estimate accurately the seismic losses to the built environment or to the insured properties to manage earthquake risks.

The performance of buildings nowadays is expressed in terms of monetary loss due to damage, downtime, and number of injuries including fatalities due to earthquakes. See, for example, performance-based earthquake engineering (PBEE) framework developed by Pacific Earthquake Engineering Research (PEER) center (Cornell and Krawinkler 2000) or that developed by ATC-58 (2011). The loss assessment of buildings is often carried out using a vulnerability function (also known as damage function), which defines the distribution of the ratio of the building loss to the building value as a function of ground-motion (GM) intensities. In this chapter, different approaches of development of building vulnerability functions (VF) will be discussed. This will be followed by calculations of the risk of building loss due to earthquakes by integrating the seismic hazard results with the building VFs.

This provides monetary losses to buildings in terms of the average annual loss (AAL) and losses at various return periods.

In 1985, Applied Technology Council (ATC) first published a comprehensive report (ATC-13 1985) to estimate the earthquake loss of existing buildings by developing a suite of VFs based on expert opinions. ATC considered this approach since very limited earthquake damage or loss data were available at that time. The report developed earthquake loss estimates as a function of Modified Mercalli Intensity (MMI) for different facility classes (e.g., low-rise wood frame). Recently, Global Earthquake Model (GEM) project has also followed a similar approach (Jaiswal et al. 2013) for developing VFs.

Professor Karl Steinbrugge based on engineering investigations, data collection, and field studies of earthquake-damaged buildings developed the first, well-accepted VFs providing an estimate of the functional relationships between building damage and earthquake intensities (Steinbrugge 1982). These functions provided a firsthand insight to the damageability of buildings to the insurance companies in the 1990s. Since engineers were interested primarily in the life safety of buildings till the end of the twentieth century, there was limited interest among the engineers to compile the damage data of all the slightly and moderately damaged buildings for assessing performance of buildings in earthquakes. Northridge earthquake (1994) was the first earthquake that provided a large number of insurance claims data, which are used for performance assessment of single-family low-rise wood-frame buildings in high seismic zones of California, for example, Los Angeles and San Francisco (Wesson et al. 2004).

In 1999, the Federal Emergency Management Agency (FEMA) first published a comprehensive methodology to calculate *analytically* earthquake damage of model buildings for a given peak response to estimate potential losses from earthquakes (HAZUS 2003). Later on, the assembly-based vulnerability (ABV) was developed by Porter et al. (2001), and this was adopted by the Consortium of Universities for Research in Earthquake Engineering (CUREE) project (2001)

for developing vulnerability functions for wood-frame buildings. This is a significantly improved analytical approach over the HAZUS. The CUREE approach, however, did not consider more realistic representations of structures (e.g., 2D/3D nonlinear FEM model as has been considered lately in the latest PEER or ATC projects), has limited set of fragility functions compared to those considered lately in the ATC-58 project (2011), and also did not consider the correlation of response or damage between different components of buildings for accurate estimation of variability of total loss of buildings.

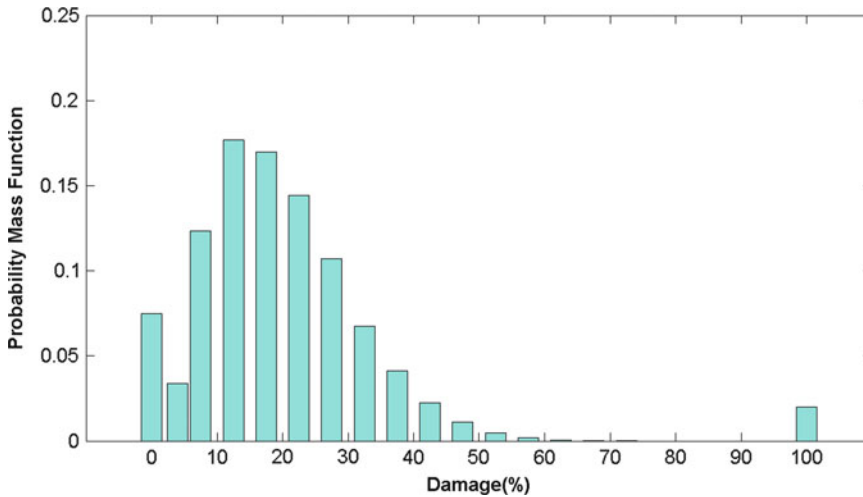
Recently, researchers are focused on developing analytical VFs based on detailed nonlinear analysis of buildings. For example, PEER (2011) considered 3D models of 40+ story tall buildings (Shome et al. 2014; Ramirez and Miranda 2009; Shome and Bazzurro 2009), and many others have considered 2D plane frame model for loss estimation of buildings. In addition, ATC-58 (2011) has developed a number of approaches – simple to very detailed and complex – for loss assessment of buildings. Majority of the researchers has adopted a *story-based loss assessment procedure* to take into account the variation in loss over the height, which is particularly important for mid- to high-rise buildings. All these researchers have considered a large number of ground-motion records for nonlinear response analysis of buildings, and a few of those made appropriate correction to the response results because of possible introduction of bias due to scaling of those records for nonlinear analysis (Haselton 2009). It is observed that standard approach of scaling does not take into account dependence of building response on the spectral shape and this introduces bias in the response results. In this context, the framework developed by Jayaram et al. (2012) will be discussed in detail. This approach considered story-based loss assessment, took into account the bias in the response results due to spectral shape, and considered correlation of response, damage, and repair cost of different components of buildings at different stories.

It is well known that there is significant uncertainty in the estimation of loss results and this

uncertainty can be divided into two different categories: (1) aleatory and (2) epistemic (Benjamin and Cornell 1970). Aleatory uncertainty is the inherent variability in the physical system; it is stochastic and cannot be reduced by improving the current approach of loss estimation. Epistemic uncertainty, on the other hand, is associated with lack of knowledge; it can be subjective and is reducible with additional information. There are significant epistemic uncertainties associated with the loss estimates due to uncertainty in the estimation of different parameters, and this should be considered in the decision-making process. FEMA-355F (2000) and ATC-58 (2011) have recommended typical uncertainties of a large number of parameters based on extensive research, and Jayaram et al. (2012) have developed a comprehensive approach to estimate epistemic uncertainty of VFs based on the uncertainty of those parameters.

Methodology

The seismic loss estimation of buildings is dependent on the lateral load resistance characteristics of buildings. Since building codes are being improved over the years, load resistance is dependent on the age of buildings. In addition, the resistance also depends on: (1) construction class (e.g., steel moment-resisting frame), (2) occupancy class (e.g., residential), and (3) region (e.g., high seismic regions in California). Although lateral resistance ideally should not depend on the occupancy class, it is generally observed that commercial buildings are generally well designed and constructed than residential buildings leading to better lateral resistance capacity of those buildings. The loss estimation of portfolio of buildings uses this type of generic information (e.g., HAZUS 2003) by grouping the buildings with similar load resistance capacities. The loss estimation of individual buildings (e.g., ATC-58 2011), on the other hand, uses the detailed building information to estimate accurately the lateral load resistance capacity of buildings for calculating seismic losses. The seismic loss estimation approaches that will be discussed



Seismic Loss Assessment, Fig. 1 Illustrative five-parameter Beta distribution

in this chapter can be applied easily for loss estimation of both portfolios and individual buildings. The VFs are developed based on mainly three approaches: (1) empirical approach which is generally expert-opinion-based development, (2) observation-based approach which is based on building loss data, and (3) analytical approach. All these approaches as well as the advantages and disadvantages will be discussed here.

Earthquakes not only cause damage to structural and nonstructural components of buildings but also damage building contents, such as furnitures, and cause closure of buildings. The closure incurs losses due to the expenses to the occupants for temporary rents or hotel stay (called additional living expenses, ALE) or loss of income to businesses (called business interruption, BI, loss). The losses to contents and due to closure of buildings will not be discussed in this chapter. This chapter will focus only on the loss due to *property damage of buildings from shaking of earthquakes*.

Loss Distribution

Since there is significant uncertainty (aleatory and epistemic) in estimating building losses at a given intensity of GM, VFs provide the loss distribution parameters as a function of intensities. It is generally assumed that the loss distribution can be represented by parametric distribution

model, and Beta distribution is widely used for building loss calculations following ATC-13 recommendation (1985). The parameters of the distribution can be estimated from the mean and variance of the sample losses (Benjamin and Cornell 1970). The insurance claims data, however, show that the Beta distribution combined with Dirac delta functions at 0 % loss and 100 % loss to model the probabilities of zero (F_0) and complete (F_I) loss fits the observed loss distribution data better. In addition, insurance companies generally pay to completely replace a property (i.e., 100 % loss) beyond a particular threshold (α) of loss (e.g., 70 % loss). Hence, more accurate representation of distribution requires three additional parameters, which are F_0 , F_I , and α , in addition to the conventional mean and variance. The 5-parameter distribution for losses is shown in Fig. 1 for illustration. Note that this type of distribution may be more appropriate to model the insurance losses. The methodology for estimating the five parameters of the distribution from loss data will be discussed in detail later on.

Seismic Risk

Seismic risk assessment of buildings requires estimating the probability of losses for all the possible future earthquakes over a specified period of time or estimating losses for a scenario earthquake. The scenario earthquake can be a

historical earthquake, or maximum credible earthquake (MCE), or standard design basis earthquake (DBE), or more frequent service level earthquake (SLE). The standard risk management of buildings entails estimation of average annual loss (AAL) for all the possible earthquakes to determine the insurance premium of buildings. This is the amount of modeled premium that an insurer needs to collect in order to cover the average loss over time. The other risk quantity that is commonly employed is estimation of probability of exceeding certain loss for all the possible earthquakes. When this is estimated for different amount of losses, the result is known in the insurance industry as the exceedance probability or EP curve. The scenario loss for a given earthquake, on the other hand, is useful for deterministic seismic risk management by estimating the Probable Maximum Loss (PML). PML is the largest possible loss which may occur, in regard to a particular risk, given the worst combination of circumstances. Traditionally, PML is the expected shake damage loss given the maximum size earthquake. There is, however, considerable uncertainty over the severity of loss that might arise under a broad variety of events. So an improved PML is a conservative deterministic upper bound, which is typically 90 % upper confidence bound on the portfolio loss from a risk-based MCE, for example, 2,475-year return period event on a fault. All these risk calculations require distribution of building loss as a function of intensity of GM and this is obtained by developing the VFs of buildings. The following section will discuss different approaches of development of VFs.

Empirical Approach: Expert-Opinion-Based Vulnerability Functions

There is significant advantage in developing expert-opinion-based VFs. This approach can be followed to develop VFs relatively quickly and easily without requiring any loss data or developing any analytical approach. Researchers over the years have followed primarily two approaches for developing VFs: (1) Delphi method which was

followed by Applied Technology Council (ATC) in the ATC-13 (1985) and (2) Cook's approach followed recently in PAGER and GEM project (Jaiswal et al. 2013). Although the information generated from these approaches may not be very reliable, this type of approach will continue to be used in other projects in the future because of ease in developing VFs when information or resource is limited.

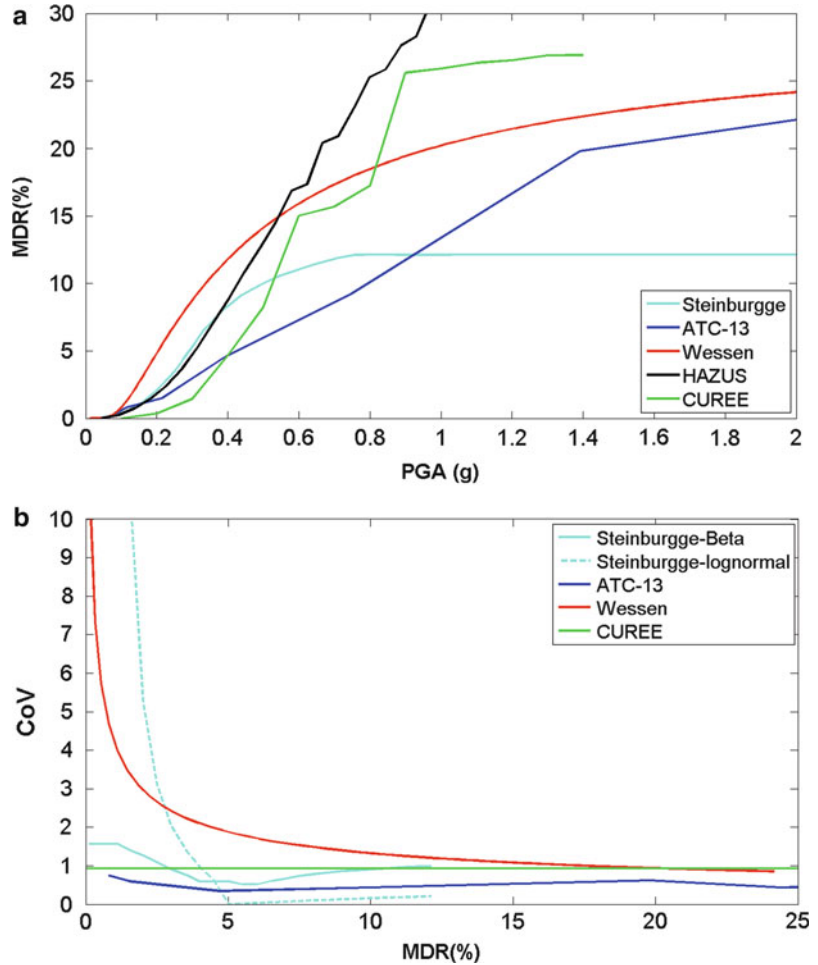
ATC-13 (1985) Approach

One of the first systematic attempts of developing building vulnerability was carried out by ATC-13 for assessing seismic risk in the state of California for the Seismic Safety Commission (SSC). The project was funded by the Federal Emergency Management Agency (FEMA). Since very limited earthquake loss data was available at that time and there was no acceptable analytical approach available for estimating building loss calculations, ATC derived VFs based on experience and judgment of the experts – structural engineers, builders, and the like – by estimating the expected percentage of damage of typical buildings for a specific construction type when these buildings are subjected to a Modified Mercalli Intensity (MMI). Based on personal knowledge and experience, the experts responded to a formal questionnaire with their best estimates of percent physical damage or damage ratio.

In order to develop the VFs, ATC established an advisory project engineering panel (PEP) providing necessary feedback in order to develop consensus of damage or loss estimates of different facility classes. This was modeled after the Delphi method for expert-opinion solicitations, which was originally developed for the Air Force by RAND Corporation in the 1950s. The tasks in Delphi method consist of formulating questionnaires, getting answers to those from experts, iterating the answers one or more times, and aggregating the responses by following a suitable statistical approach. The questionnaire asked each expert to provide low, best, and high estimates of the damage ratio for different construction classes or engineering facilities at different MMI. In addition, the experts were also asked to provide their level of experience with the

Seismic Loss Assessment,

Fig. 2 Vulnerability functions of single-family wood-frame buildings in California. (a) Mean damage ratio (*MDR*). (b) Coefficient of variation (*CoV*)



construction class and the degree of uncertainty in their estimates. Subsequently, two other sets of questionnaires were developed in order to reach a consensus on the estimation of different damage estimates. The questionnaire results were processed to develop damage probability matrices (DPM) by fitting Beta distribution to those results. In DPM, the damage to a construction class for a given intensity is described by a series of damage states (DS) and the probability that the DS will occur at that intensity. The VF of single-family wood-frame buildings in California following this approach is shown in Fig. 2 for illustration.

Because it was the first systematic attempt to develop VFs, ATC-13 quickly became the standard for assessing earthquake risk of buildings.

Catastrophe modelers and risk engineers adopted the ATC-13 damage curves until the 1994 Northridge earthquake. It will be shown later on that there is significant difference in the results based on ATC-13 and that observed in the Northridge earthquake. In addition, the ATC-13 approach is subjective. The damage functions from this approach are based exclusively on the opinion of experts. For this reason, it is difficult to calibrate or modify those in order to incorporate new data or new design (e.g., adjustment for new seismic code or other seismic regions). These functions can also be biased as observed in the Northridge earthquake. Hence, in the recent years, researchers have focused on developing VFs either based on observed building-level damage data (Wesson et al. 2004) or from

analytical approaches. This approach combines the experimental damage data and the observed damage data of different components of buildings (Jayaram et al. 2012).

GEM/PAGER Approach

Recently, US Geological Survey (USGS) has adopted Cook's model for developing VFs for the PAGER system and for the GEM project (Jaiswal et al. 2013). This approach involves identifying experts, constructing seed and target questions, eliciting expert judgments, and combining multiple judgments. In this approach, unlike the Delphi approach (adopted by ATC-13), the judgment of scoring is calibrated objectively and classical statistical test is employed for hypothesis testing of the scores. This involves assigning weight to an expert based on two separate scores: (1) *calibration score* which measures the accuracy of the scores by the experts based on background reference density functions of the scores and (2) *information score* which represents the degree with which the distribution of the expert is concentrated with respect to some background distribution. The product of these two scores is the overall score of each of the experts. Since the weights can be recomputed objectively with new observation or addition of new experts, the weights can be recomputed in the future for updating the VFs.

Observation-Based Approach: Vulnerability Functions from Building Damage Data

The most direct approach of developing the VFs would be based on building damage data observed in historical earthquakes. This approach requires damage data of large number of buildings for different construction classes, height, age, and other characteristics, which are important in defining lateral load resistance of buildings. If the damage data is based on detailed inspection of buildings following earthquakes, VFs can be developed accurately. But this requires significant amount of time and resources

and so this type of data is not available in sufficient quantity for developing the VFs. On the other hand, insurance claims data (e.g., those from Northridge earthquake) can provide a large quantity of damage data, which can be used for estimating VFs. But these data typically contain limited information about the buildings (typically provides construction class – primarily materials of construction, e.g., steel, etc., height, and age).

The earliest attempt in developing VFs following this approach was by Steinbrugge (1982). The VFs were developed based on limited damage data for a handful of earthquakes. Since insurance industry was not active in collecting damage data before the 1990s and engineers were focused on studying a few important damaged buildings following earthquakes, Steinbrugge applied his experience and judgment in order to develop the functions based on limited damage data that he collected over the years from small number of earthquakes. The developed VF for single-family wood-frame buildings is shown in Fig. 2. The results show that the vulnerability of residential wood-frame buildings is significantly higher than that predicted by ATC-13 at lower intensities. This function was used extensively by the insurance companies in the 1980s–1990s as a guideline for managing earthquake risks. Later on, this function will be compared with that based on extensive damage data from Northridge earthquake.

In this section, Northridge insurance claims data, which lacks detailed information on the lateral load resistance characteristics of buildings, will be used for developing the VF. The other type of data, which would be a preferred one, is the damage data that provides the detailed lateral load resistance characteristics of buildings. An example on this would be the data of fractured connections of 167 welded steel moment-frame (SMF) buildings collected in the SAC joint venture project. This data was collected in a very detailed and systematic way following Northridge earthquake, and those data were used for developing empirical VFs for building loss calculations in FEMA-351 Rapid Method (Bonowitz and Maison 2003). The project inspected about 18,000 connections and

found that 13 % of those had some kind of damage. Although this approach provided very accurate information about the performance of welded connections of SMF buildings, seismic loss estimation requires repair cost of those connections as well as other structural and nonstructural components of those buildings. Note that the value of the nonstructural components can be more than 70 % of total building value and so the repair cost of these components is significant, particularly at low intensities. Since damage of nonstructural components were not collected, the empirical VF developed in SAC project is primarily for the structural components, and the VF will provide a low estimate of damage at low intensities. In addition, the uncertainty in the damage or loss of buildings at a given intensity cannot be estimated from this limited data, and this information is essential for seismic risk calculations.

Since insurance claims data lacks detailed building information, the VFs developed from this data are for a broad category of buildings, like low-rise wood-frame buildings constructed in the era of moderately developed building code (e.g., 1950–1975 in high seismic regions in California). This would lead to higher uncertainty in the estimation of losses relative to the losses for a specific building. The other issue with the insurance claims data is that the data captures losses only after applying the deductibles (known as gross loss). Hence the building loss is partially known from insurance claims data (formally known as “censored” data), and this missing information is critical for accurate estimation of building vulnerabilities at low intensities.

Development of Vulnerability Functions Based on Claims Data

In this section, insurance claims data will be used for developing VFs, which estimates the probability distribution of building loss as a function of ground-motion intensity. This approach is illustrated here by using the Northridge claims data, which is even today the best source of a large number of high-quality insurance claims data for earthquakes. The data were available only by zip codes and provided the claim amount, value of

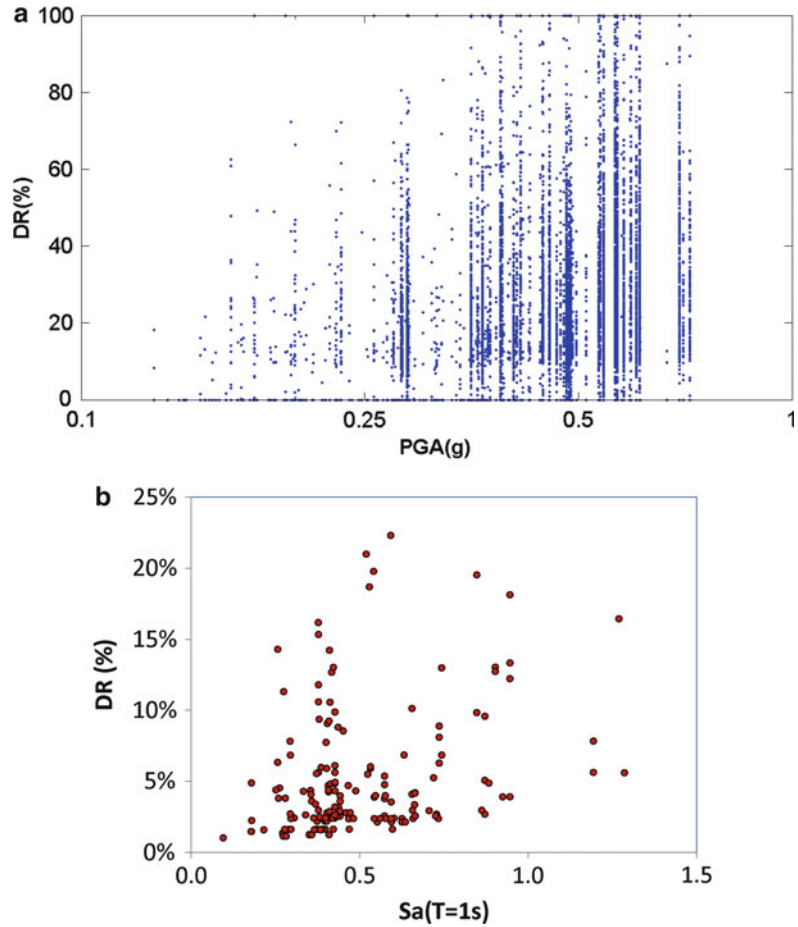
the structure, year of construction of the buildings, and the construction class. The scatter plot of insurance loss data for single-family wood-frame dwellings (SFD) constructed after 1975 is shown in Fig. 3a, and the same for steel moment-frame (SMF) buildings based on detailed inspection of beam-column connections is shown in Fig. 3b for visual comparison of the trend and uncertainty in these two data sets. The figure illustrates that the uncertainty conditioned on earthquake intensity (here PGA for SFD and $S_a(T = 2 \text{ s})$ for SMF) is very high. The vertical stripe of data in the Fig. 3a represents the variation of loss ratio in a specific zip code since Northridge data is available only at the zip code level. Note that although the insurance claims data provides losses for content and additional living expenses (ALE), those are not considered in this section. The building loss data here includes damage to the structural as well as nonstructural components of buildings.

Since the Northridge losses are available by zip code, Wesson et al. (2004) estimated loss distribution by zip code by fitting a 2-parameter Gamma distribution to the claims data. Since insurance data does not capture the losses below deductible, the maximum likelihood approach was followed for estimating the parameters with the constraint that the fraction of losses below deductible must be equal to the observed fraction of no loss. The parameters of the distribution for each of those zip codes are then fitted independently against the intensity measure to get the distribution of loss as a function of intensity measure. The VFs based on this approach are shown in Fig. 2. It is observed that there is significant difference in the results between the vulnerability functions based on Northridge insurance claims data and that estimated earlier in ATC-13 following expert opinion or by Steinbrugge based on limited damage data.

A more direct approach for the estimation of the distribution parameter would be to estimate those directly from the data by grouping those as a function of intensity. In the earlier approach, since each of the parameters of the distribution was fitted independently, the correlation between the parameters was not preserved. In addition, the

Seismic Loss Assessment,

Fig. 3 Distribution of damage ratio (ratio of claims amount or repair cost and building replacement cost) distribution as a function earthquake intensity as observed in Northridge earthquake. (a) Post-75 single-family wood-frame buildings. (b) Steel-frame buildings inspected in SAC project



regression analysis gave equal weight to all the parameters for each of the zip codes irrespective of the number of data points in the zip codes. Since Beta distribution is commonly employed for describing the earthquake loss distribution and a five-parameter Beta distribution is deemed to be the most suitable for modeling seismic loss distribution, the fitting of five-parameter distribution model to the loss data will be discussed here. The five parameters of the distribution are: mean, variance, probability of no loss (F_0), probability of complete loss (F_1), and the threshold (α) beyond which losses are considered to be 100 % by insurance companies. The parameters of the probability distribution can be estimated by using two methodologies, namely, the method of moments and the maximum likelihood approach. Since the insurance loss data is available after applying the insurance limits and deductibles,

the method of moments is not suitable for this type of data. Hence *maximum likelihood approach* will be followed to estimate the parameters (see Shome et al. 2012 for details).

In general, the building VFs based on observed damage/loss data from one region cannot be directly extrapolated to other regions since the building design and construction practice varies among different regions. Since the frequency of large-magnitude earthquakes is small and the occurrence of those events near large population center is even rarer, it is not possible to get a large number of damage data for different construction classes in order to develop VFs. Also this approach needs high-quality data, i.e., building-level loss results like those obtained in 1994 Northridge earthquake. The number of data should be high for estimating VFs with high confidence.

Maximum Likelihood Estimation (MLE) of Parameters

The objective here is to estimate the parameters of the five-parameter Beta distribution for the actual loss paid by the building owners for repairing buildings (i.e., distribution for ground-up (GU) loss). The GU loss is computed by adding the gross loss, the deductible, and the losses beyond the limits if applied. But this cannot be done for the cases when the gross loss equals zero. This is a typical case of left-censored data set where the ground-up losses are accurate above deductible, but the data

below deductible is not available and is reported as zero. If F_0 is simply estimated as the fraction of data points with zero ground-up loss, F_0 will be severely overestimated when the deductibles are large particularly at low intensities. Note that the data is right censored as well because losses beyond the insurance policy limits are truncated to the limit. This is, however, much rarer since it only occurs at very large damage ratios induced by very high-intensity GM.

The density function for damage ratio $f(D_i)$ at a given intensity equals the following:

$$f(D_i) = \begin{cases} F_0 & \text{for } D_i = 0 \\ F_0 + (1 - F_0 - F_1)B_{PDF}(D_i, a, b) & \text{for } 0 < D_i < \alpha \\ [F_1 + (1 - F_0 - F_1)(1 - B_{CDF}(\alpha, a, b))] & \text{for } D_i \geq \alpha \end{cases} \quad (1)$$

where:

α = The threshold beyond which losses are treated as 100 %. Note that this is governed by the underwriting policy of an insurance company.

$B_{CDF}(D_i, a, b)$ = Beta cumulative distribution function (CDF) corresponding to sample damage ratio D_i (=the ratio of the loss to the building's value).

$B_{PDF}(\cdot)$ = Beta probability density function (PDF).

a and b = Parameters for the Beta distribution of the sample damage ratios for a GM intensity bin.

Hence the likelihood of observing damage ratios D_i s equals:

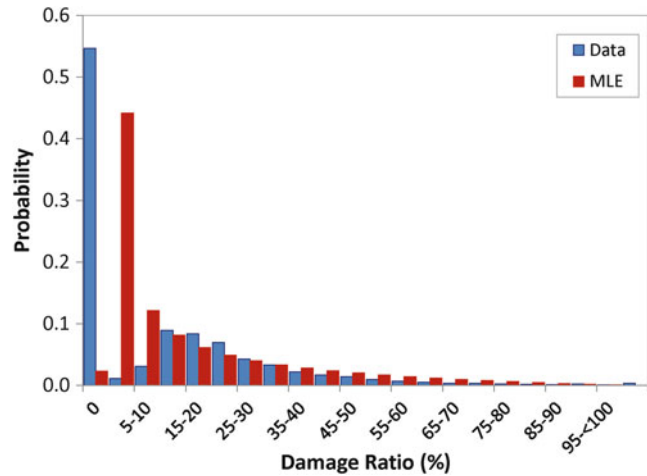
$$L = \left[\prod_{D_i=0} F_0 + (1 - F_0 - F_1)B_{CDF}(ded_i, a, b) \right] \left[\prod_{D_i \geq \alpha} F_1 + (1 - F_0 - F_1)(1 - B_{CDF}(\alpha, a, b)) \right] \left[\prod_{D_i \in (0, \alpha)} F_0 + (1 - F_0 - F_1)B_{PDF}(D_i, a, b) \right] \quad (2)$$

where the parameters, F_0 and F_1 , correspond to a range of GM intensities for which loss data is accumulated for fitting and these parameters are estimated by maximizing L . The first term in the likelihood estimate corresponds to data points where the observed damage ratio equals zero. This includes likelihood of damage ratio less than deductible ded_i , but reported as zero in the claims dataset, and the true zero damage. The second term corresponds to data points where the observed damage ratio equals 1. This corresponds to the likelihood of observing complete loss and the losses above the threshold, α . The third term corresponds to the rest of the data points where the damage ratio is more than deductible, but less than α . The parameters of the distribution F_0 , F_1 , a , and b are obtained by maximizing the likelihood of observing all the damages, D_i , which is shown in Eq. 2 as the product of likelihood of the damages.

The results of fit of the five-parameter distribution to the Northridge loss data at high-intensity GM are shown in Fig. 4. The claims data used for fitting the distribution is from a number of insurance companies and the information of the threshold (α) is not available from those companies. Hence for simplicity, it is assumed that α equals to 100 % damage.

Seismic Loss Assessment,

Fig. 4 Comparison of distribution of loss at high-intensity ground motion between the data and MLE fit



The results show that the fitting captures well the objective of probability mass at zero and 100 % damage. Note that it would be difficult to estimate reliably all the parameters of the distribution when data is limited. In those cases, it is advisable to bring in other sources of data to define the loss distribution. For example, one can use building collapse results – analytical and/or experimental – to define F_I of a specific building or a class of buildings.

Analytical Approach of Development of Vulnerability Functions

There is limited building damage data from earthquakes and that too from small number of construction classes. In addition, building damage characteristics based on the damage data from one region may not be extrapolated to other regions. Hence, VFs for majority of construction classes in most of the regions in the world can be developed only from an analytical approach. Hence HAZUS (2003) first developed a comprehensive analytical methodology to estimate the building seismic losses. HAZUS adopted capacity spectrum method to estimate response of *generic buildings* and estimated only *mean* loss based on damage of different building components. Later on, CUREE-Caltech (2001) developed the assembly-based vulnerability (ABV) methodology for developing *building-specific* VFs. In this

approach, the damage is calculated for the individual building assemblies, e.g., partition wall, based on the response results from *nonlinear time-history analysis* of the model buildings to estimate *mean and variance* of damage of the model buildings as a function of GM intensity. The individual buildings in this approach were represented by deterministic simplified structural models, where building components were idealized as springs and masses. Recently, Jayaram et al. (2012) have developed a systematic simulation approach based on the PEER-PBEE loss assessment framework to develop building-specific VFs from component-level damage at each story of buildings. In this approach, the total building loss is estimated by using the response from detailed nonlinear time-history analysis of buildings. The approach took into account the dependence of response on S_a at multiple periods, the correlation of losses between different components to improve the estimation of uncertainty of losses, and on captured the effects of epistemic and aleatory uncertainties of different parameters considered in loss calculations, such as GMs, structural response parameters, loss costs, etc. All these steps help to get the unbiased loss due to earthquakes and quantify accurately the uncertainty in the loss estimates.

The analytical loss calculation in general consists of (1) repair cost of different building components associated with a damage state or cost

due to demolition, (2) damage state given a demand parameter (EDP), and (3) EDP given an intensity of GM. A brief description of these parameters is given below and the details can be found in HAZUS (2003).

Building Components

In order to carry out the seismic loss calculations, building components are divided broadly into two categories: structural and nonstructural. The loss estimation approach either estimates damage of the individual components (e.g., CUREE-Caltech 2001) or damage of a group of components or subsystems (e.g., HAZUS 2003). The structural component (SD) resists gravity, earthquake, wind, and other types of loads. This includes components like columns, beams, load-bearing walls, etc. For buildings planned by design professionals, the structural components are typically designed and tailored to building-specific configurations and loading conditions. The nonstructural components include a large variety of architectural, mechanical, and electrical components, which can be further divided into as either “drift-sensitive” (NSD) (e.g., non-load-bearing partition walls, exterior curtain walls, etc.) or “acceleration-sensitive” (NSA) components (e.g., suspended ceilings, HVAC systems, etc.). These components are usually neither designed by the professionals nor tailored to specific buildings even when designed. These are tested under some generic load conditions to satisfy the acceptability criteria of the building codes under some idealized conditions. Hence the uncertainty in the performance of structural components is relatively low when designed and constructed by professionals compared to those of the nonstructural components. Drift is used to predict the damage states of the SD and the NSD subsystems or components. Acceleration, on the other hand, is used to predict the damage states of the NSA components.

Damage States

The damage state of each subsystem or component represents a consequence in terms of repair

for the type and the severity associated with that state. This is predicted based on the fragility functions of the subsystem or component. The damage states are defined in HAZUS (2003) and these are none, minor (DS1), moderate (DS2), extensive (DS3), and complete damage (DS4). For example, moderate structural damage of concrete shear-wall buildings is defined by the diagonal cracks on most of the shear-wall surfaces and large diagonal cracks with spalling of concrete at the wall ends of some shear walls. The moderate damage of the NSD components or subsystems represents large and extensive cracks of the partition walls requiring repair and repainting as well as replacement of a few walls. The moderate damage state of the NSA subsystem or component represents extensive falling of the suspended-ceiling tiles with disconnected and/or buckled ceiling support framing (T-bars) at a few locations.

Repair Costs

The loss from earthquakes is associated with the repair due to damage of different components or subsystems as well as building replacement cost associated with demolition and collapse. The uncertainty in repair cost arises primarily due to lack of information about the component. The cost of repair works is generally significantly higher than the cost of new construction, and this is more significant for the structural components than the nonstructural components. For example, the repair of beam-column connections of the structural component requires removing the suspended ceilings, inspection of the joints which may require non-destructive testing, closure of part of the floor of the building, etc. All these expenses have to be incurred on top of the repair cost of the connections. Insurance claims data also suggest relatively higher uncertainty in repair cost at low intensities. In order to estimate the repair cost, it is generally assumed that the insurer will pay the cost of repair or replacement of the damaged buildings with the materials of like kind and quality without any deduction for depreciation, which is called *replacement cost method* for valuation.

HAZUS-Based Vulnerability

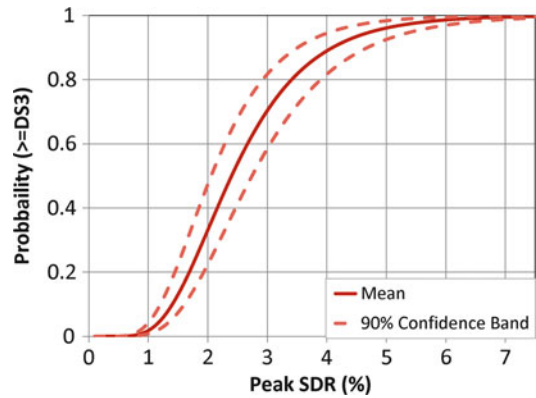
The HAZUS methodology is developed to estimate losses for different federal, state, and local agencies for planning earthquake risk mitigation, emergency preparedness, and many other uses. Although the methodology can be used to estimate losses for individual buildings, it is developed to estimate losses of portfolio of buildings distributed over a large region. The methodology provides an estimate of the probability of mean loss for none, slight, moderate, extensive, and complete damage of building components at a given level of ground shaking. The building damage characteristic is defined by the following curves: (1) fragility curves that describe the probability of reaching or exceeding different states of damage given the peak-building response and (2) building capacity (pushover) curves that are used to determine the peak-building response for the demand spectrum.

Fragility Functions

The fragility curve is defined by a median value of the engineering demand parameter (EDP), which is, for example, spectral displacement in HAZUS and maximum interstory drift ratio (IDR) in Jayaram et al. 2012, and by the variability associated with the estimation of that damage state for that EDP. The median EDP corresponds to the threshold of the damage state. The curve is assumed to follow the cumulative lognormal distribution function as shown below:

$$P(DS \geq ds_i | EDP) = \Phi\left(\frac{\ln(EDP/\mu)}{\beta}\right) \quad (3)$$

where μ denotes the median value of EDP at which building component or subsystem reaches the threshold of the damage state ds_i , β denotes the corresponding dispersion, and $\Phi(\cdot)$ is the cumulative density function for standard normal distribution. The dispersion considers (1) uncertainty in the damage state threshold, (2) variability in the capacity (response) of buildings, and (3) uncertainty in building response or EDP due to the spatial variability of ground motion.



Seismic Loss Assessment, Fig. 5 Mean fragility functions and its uncertainty for extensive damage of nonstructural drift-sensitive subsystem (NSD)

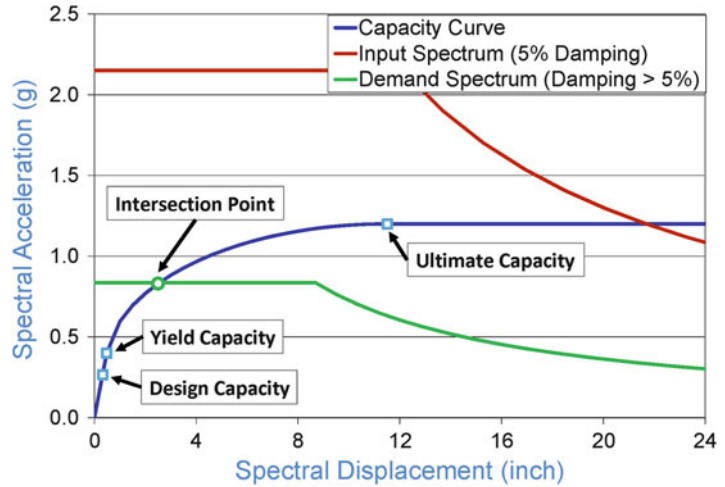
The component fragility functions are estimated from the results of laboratory experiments or from observed damage data. Since there is limited observed damage data or experimental data, past studies recommended fragility functions primarily based on engineering judgment (e.g., HAZUS). Recently, ATC-58 (2011) has made significant effort to develop a suite of fragility functions so that loss calculations can be carried out in day-to-day building design work. Also Network for Earthquake Engineering Simulation (NEES) Program is actively pursuing experimental research (<http://www.nees-nonstructural.org/>) to develop fragility functions for different nonstructural components of buildings. All these research provide a range of fragility functions giving an estimate of the mean and the standard deviation of the fragility functions. The mean fragility function for extensive damage of the generic NSD subsystem, which is an assembly of a number of components such as partitions, exterior glazing systems, etc., is shown in Fig. 5. The epistemic uncertainty in the estimation of the fragility function for that damage state is shown by the 90 % confidence band.

Building Response Calculations

The building responses in HAZUS are estimated based on the *capacity spectrum method* (CSM). The building capacity curve (also known as a

Seismic Loss Assessment,

Fig. 6 Example building capacity curve, demand spectrum, and building response for wood-frame buildings in California for input spectrum from M7 earthquake on Hayward fault at a site close to the fault



pushover curve) describes the lateral load resistance of a building as a function of its characteristic lateral displacement. The curve is derived from the results of static-equivalent base shear and roof displacement and is defined by three control points: (1) design capacity, (2) yield capacity, and (3) ultimate capacity as shown in Fig. 6. It is assumed that the building capacity will have a range of possible properties and those are lognormally distributed as a function of the ultimate capacity. The peak-building response is the intersection of the building capacity curve and the response spectrum, which defines the seismic demand. Since buildings go beyond its elastic limit, the demand spectrum defined by the elastic damping is reduced due to the higher effective damping. The effective damping includes both the elastic damping and the hysteretic damping associated with the post-yield cyclic response of buildings. Figure 6 illustrates the capacity curve, demand spectrum curve, and the peak-building response at the point of intersection of those curves. The main drawback of CSM is that the MDOF structures are idealized by an equivalent SDOF system using an increased period and damping in order to estimate the seismic demand of the MDOF structures by the capacity spectrum approach. In addition, this method is a nonlinear static procedure, whereas the nonlinear time-history analysis generally provides a more accurate estimation of building

response particularly at higher intensities. Hence this approach may not be suitable for mid-rise and high-rise buildings and for types of buildings at high intensities.

Building Loss Calculations

The fragility curves define all the four damage states for each of the three building subsystems based on the EDPs estimated from the intersection of the capacity curve and the seismic demand curve. It is assumed that the nonstructural damage states are independent of structural damage states. The repair cost is estimated based on the damage of different subsystems of buildings. Note that although the results from HAZUS loss estimation methodology are intended to assess the socioeconomic impact of earthquakes in a region, the result has a number of limitations. Earthquake impact on a society is a very complex process and the economic consequences cannot be considered in a simplistic approach. The dollar losses are calculated for all the three coverages: (1) building repair and replacement costs, (2) contents losses, and (3) additional living expenses (ALE) for residential occupancies or business interruption (BI) losses for commercial or industrial occupancies. In this chapter, only building repair loss calculations are discussed for brevity. This loss considers repair and replacement due to both structural and nonstructural damages, and this is calculated as the product of the probability

of a component in a given damage state and the mean repair costs for that damage state, summed over all the building subsystems and the damage states as given below:

$$L_i = RC_i \cdot \sum_{k=1}^3 \sum_{j=2}^5 P_{ijk} \cdot DR_{ijk} \quad (4)$$

where RC_i is the replacement cost of an entire model building i (e.g., wood frame), P_{ijk} is the probability of model building i and subsystem k (e.g., NSD) being in damage state j (e.g., moderate), and DR_{ijk} is the mean damage ratio for building i and subsystem k in damage state j . The ratio of loss of building L_i and the replacement cost of RC_i is the mean damage ratio (MDR) of the building i for the intensity of GM at which the losses are calculated. This calculation is repeated at different intensities to develop the VF(s) for seismic loss calculations of an individual building or a portfolio of buildings. Note that although HAZUS considers only mean repair cost, simulation-based approach, which will be discussed later on, considers the uncertainty in repair cost as well to calculate the uncertainty in the VFs. Figure 2 shows the VF following HAZUS approach of residential light wood-frame buildings constructed during the period of 1940–1975 following a modern building code (e.g., Uniform Building Code) in the high seismic zone of California for a suite of response spectrum for different magnitude of earthquakes at different distances based on the disaggregation of USGS hazard results in Los Angeles.

Assembly-Based Vulnerability (ABV)/CUREE-Caltech Approach

In this approach (CUREE 2001), the nonlinear building response was calculated based on nonlinear time-history analysis of model buildings representing the best estimate of mass, stiffness (load-deformation characteristics), damping, and building properties required for the analysis. The buildings are modeled as a collection of assemblies of components, such as a gypsum wallboard partition, a floor diaphragm, beams, etc., and each of the assembly is represented by a fragility

function to estimate the probability of that assembly in different damage states as a function of structural response. Test data from the CUREE Project, experimental data from other projects, and observation of damage from different earthquakes are used to develop the fragility functions. The losses were estimated for repair of damage of different assemblies following standard construction cost-estimation principles. This approach estimates mean and uncertainties (which is also improvement over HAZUS) in loss estimate by capturing the uncertainties in: ground motion, structural characteristics, component damageability, and repair costs. This approach followed a simulation approach in order to quantify the (aleatory) uncertainty in the loss results. The losses are calculated for individual components of the buildings instead of a group of components like in HAZUS to get the building loss.

The structural response which is calculated from the building simulation results is used as an input to the fragility function to simulate the damage state of the component assemblies. The components are idealized as an assembly of springs and masses to calculate the response. The masses represent the building components that have significant weight, such as the walls, ceiling, floor, etc., and the springs represent building elements, such as the walls, beams, etc., that resist deformation. The idealization took into account for strength and stiffness degradation of the components, along with many other complex nonlinearities, to represent accurately the spring element. The response parameters used in this approach are maximum interstory drift ratio (IDR), peak floor acceleration (PFA), peak transient horizontal shear strain, peak member forces, and residual drift ratio.

The total repair cost is calculated by summing the cost of repairing the damaged assemblies. The cost of repairing a damaged assembly includes labor costs, material requirements, debris removal, and equipment rental. The cost calculation includes the cost of repair of the undamaged assemblies (or line of sight) due to damage of an assembly. The cost also considers the case of collapse of the building. The cost is simulated from the distribution of the cost for repair of each damaged assembly. The simulation

provides total repair or replacement cost at a given level of seismic shaking intensity. The process is repeated many times at a given intensity to estimate the probability distribution of cost for calculating the building-specific VF. The plot of MDR and coefficient of variation (CoV) of typical wood-frame buildings of average size as a function of PGA based on CUREE-Caltech project is shown in Fig. 2. The plot shows that the prediction of MDR and uncertainty are significantly less than those observed in Northridge earthquake.

PEER-PBEE-Based Approach

Recently, Jayaram et al. (2012) have developed a generic procedure for developing the building VFs by following a systematic simulation approach based on the PEER performance-based

earthquake engineering (PBEE) framework. The steps involved in the procedure are quantifying ground-motion hazard using a *vector of spectral accelerations*; predicting building response parameters such as story drifts, floor accelerations, and residual drifts under the quantified hazard; estimating structural collapse and demolition; and calculating *story-wise losses* based on the responses. The total building loss is calculated by summing the story losses. The procedure captures the effects of epistemic and aleatory uncertainties in different parameters considered in loss calculations, such as ground motions, structural response parameters, loss costs, etc., in order to quantify the uncertainty in the final loss estimate.

The mathematical framework of the development of the mean vulnerability function based on this approach is given below:

$$E(L|S_a(T_0)) = \iiint L \cdot dG(L|DM) || dG(DM|EDP) || dG(EDP|S_a(\mathbf{T})) || dG(S_a(\mathbf{T})|S_a(T_0)) \quad (5)$$

where L is the building loss; $S_a(T_0)$ is the spectral acceleration at a reference period; $E(L|S_a(T_0))$ is the mean loss at $S_a(T_0)$; $dG(L|DM)$ denotes the derivative of the probability of exceedance of the building loss given a damage measure (DM); $dG(DM|EDP)$ is the derivative of the probability of exceedance of the DM given an EDP (e.g., story drift ratio); $dG(EDP|S_a(\mathbf{T}))$ is the derivative of the probability of exceedance of the EDP given a vector of spectral acceleration, $S_a(\mathbf{T})$; and $dG(S_a(\mathbf{T})|S_a(T_0))$ is the derivative of the probability of exceedance of spectral accelerations at multiple periods, $S_a(\mathbf{T})$, given $S_a(T_0)$. Note that the reference period, denoted here T_0 , is not necessarily the fundamental period of the analytical

model of a structure. The fundamental period is a good choice when the structural performance is reasonably elastic and dominated by the first mode, but it is not a good choice for tall buildings in high seismic regions. The response of these buildings is nonlinear when damaged during earthquakes, and the response is determined by multiple frequencies. The equation shows that the structural response is predicted based on a vector of spectral accelerations, but the vulnerability is represented as a function of scalar intensity measure for simplicity of use, but without sacrificing any accuracy.

The variance of the vulnerability function can be estimated as follows:

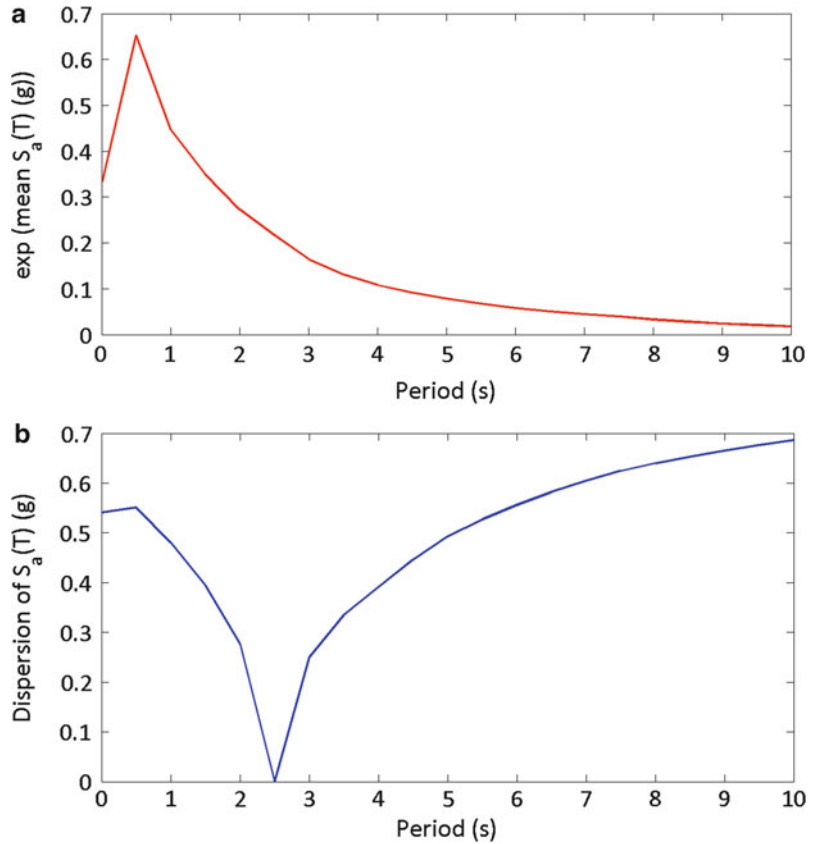
$$Var(L|S_a(T_0)) = \iiint (L - E(L|S_a(T_0)))^2 dG(L|DM) || dG(DM|EDP) || dG(EDP|S_a(\mathbf{T})) || dG(S_a(\mathbf{T})|S_a(T_0)) \quad (6)$$

The above integrals Eqs. 5 and 6 can be evaluated conveniently by using Monte Carlo Simulation

(MCS). When the number of variables is large, particularly when the intensity measure is of

Seismic Loss Assessment,

Fig. 7 Response spectrum conditioned on $S_a(T_0) = 0.22$ g at $T_0 = 2.5$ s. (a) Exponential of the conditional mean of $\ln S_a(T)$. (b) Conditional standard deviation of $\ln S_a(T)$



large dimension (i.e., vector of S_a), it is efficient and convenient to use MCS. The MCS approach involves simulating all the random variables in Eqs. 5 and 6 (which are $S_a(T)$, EDP , DM , and L) and then computing the mean and the variance of L for a wide range of $S_a(T_0)$ values. The MCS approach provides flexibility and allows consideration of accurate but complex models for different parameters as well as correlation between these parameters. It also allows for the rigorous consideration of collapse and demolition.

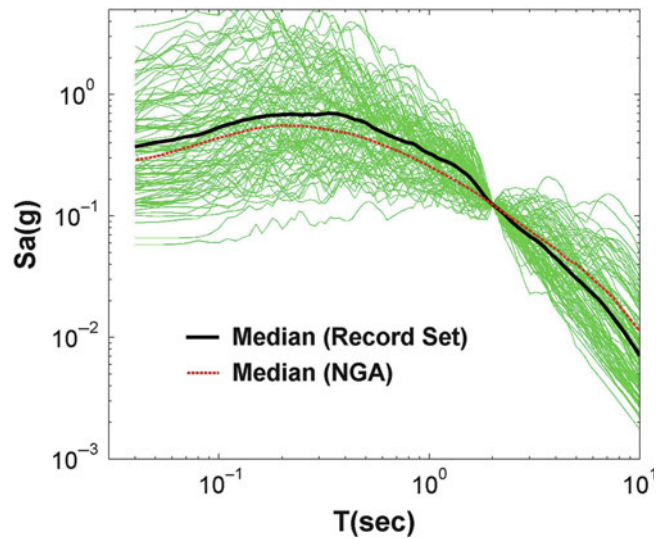
Conditional Spectral Accelerations ($S_a(T)|S_a(T_0)$)

Although it is quite common to scale a suite of ground-motion records to $S_a(T_0)$ for nonlinear analysis of buildings for estimating EDPs, studies have shown that this introduces bias in the results – particularly for mid-rise to high-rise buildings (e.g., Haselton 2009). This is due to the fact that the response of these buildings depends on S_a at periods other than T_0 . Hence it

is important to select the records in such a way that the distribution of $S_a(T)$ at all the periods conditioned on $S_a(T_0)$ of the scaled records is appropriate. The mean and variance of a typical conditional spectrum in Los Angeles are shown in Fig. 7. It is observed that variance is zero at the conditioning period T_0 , but the variance increases rapidly as the period of S_a moves further from T_0 . In general, the risk of exceeding $S_a(T_0)$ at a site is caused by earthquakes of several possible combinations of M - R . Since the shape of response spectrum depends on the M - R of earthquakes, there are different possible shapes of response spectrum conditioned on $S_a(T_0)$ due to the different combinations of M - R of earthquakes that can strike a site. The distribution of M - R conditioned on $S_a(T_0)$ at a building site can be obtained by deaggregation of hazard for the site, and this different shaped response spectra should be considered in the simulation of building response to improve accuracy. In the simulation of loss

Seismic Loss Assessment,

Fig. 8 Response spectra for an event ($M_w = 7$ and $R = 10$ km) scaled to the median $S_a(T_0) = 0.12$ g of a set of 98 records and the median spectrum for the set. The median is compared with the average of the four NGA attenuation models (Bozorgnia et al. 2014). The median spectra are closely matched by carefully selecting the set of records



results, first the M and R pairs are sampled based on the distribution from the deaggregation of hazard results, and subsequently, spectral acceleration at multiple periods, $S_a(T)$, conditioned on $S_a(T_0)$, M , and R , is simulated. An example of simulated response spectrum is shown in Fig. 8.

Calculation of EDP

The EDPs required in this approach for loss calculations are interstory drift ratio (SDR), peak floor accelerations (PFA), and residual drift ratio (ResDR). The SDR values are used to predict the damage of SD and NSD components at each story level, the PFA values are used to predict those of the NSA components, and ResDR is used to determine if a building would have to be demolished after an earthquake. The EDPs are first calculated based on nonlinear time-history analysis of buildings for a suite of ground-motion records. But it is nearly impossible to select records that satisfy conditional distribution of S_a at multiple periods (which is can be more than ten for tall buildings) (see, e.g., Jayaram et al. 2012). Hence it is efficient and also accurate to fit regression models to predict independently the different EDPs at different stories and use these fitted models subsequently to estimate the EDPs for the simulated response spectra, which have the right conditional distribution of $S_a(T)$.

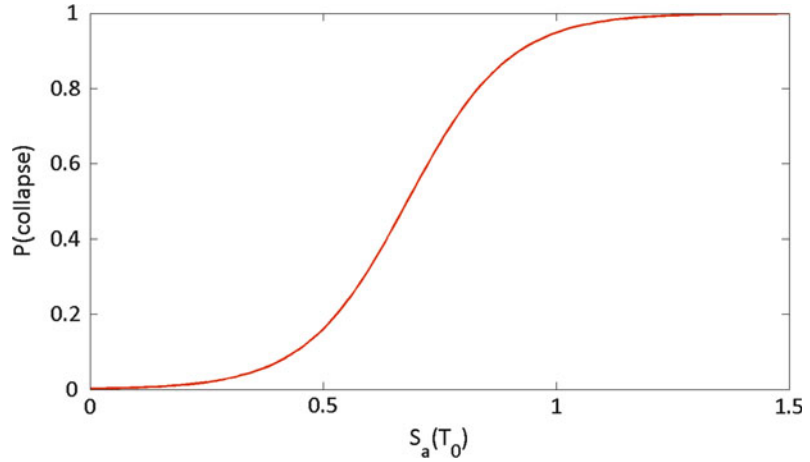
Hence the purpose of computing these EDPs by nonlinear time-history analysis for different recorded ground motions is to develop regression models to predict EDPs as functions of $S_a(T)$ while carrying out simulations.

During earthquakes, there is a possibility of collapse of structures. Two different types of structural collapses can be observed during the analysis of buildings. The first type of collapse, *simulated collapse*, is the collapses observed in the analyses. This occurs due to excessive lateral drifts and deterioration that lead to loss of the gravity load resistance of structures. This is modeled using a collapse probability curve, which provides the probability of collapse of buildings as a function of $S_a(T_0)$. The other type is referred to as *virtual collapse*, which captures those cases when a structure in nonlinear analyses undergoes a very large drift without collapse due to limitations in numerical modeling. Both these collapses are considered here for developing vulnerability functions. The collapse probability curve is developed using logistic regression, which is commonly used for regressing binary data (Shome and Cornell 2000), and can be expressed as follows:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \cdot S_a(T_0) + \varepsilon \quad (7)$$

Seismic Loss Assessment,

Fig. 9 Collapse fragility curve for a modern 20-story steel moment-frame building in Los Angeles



where p = probability of collapse, β_0 and β_I = regression coefficients, and ε = regression error. The collapse fragility curve for a 20-story modern steel-frame building is shown in Fig. 9.

In addition to fitting the EDPs at each story for estimating the mean and standard deviation of the loss results conditioned on $S_a(T)$, the correlation between different EDPs at the same story and between EDPs at different stories needs to be estimated to jointly simulate the EDPs. The extent of these correlations is computed using the regression residuals. It is assumed that the EDPs follow multivariate lognormal distribution.

Damage Measure and Repair Cost

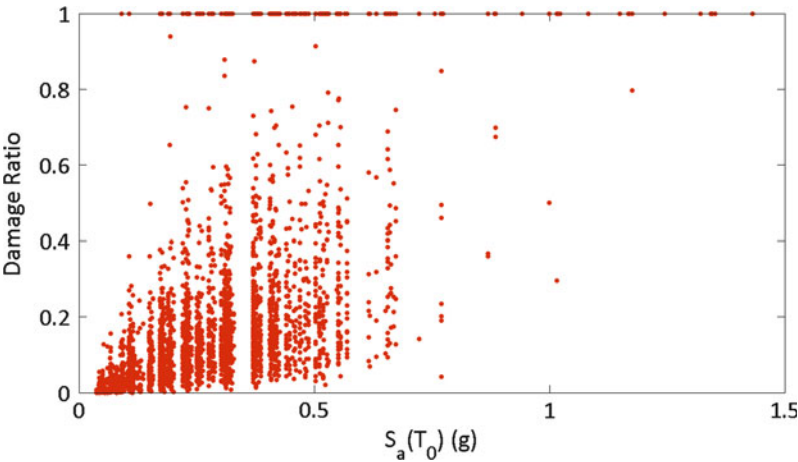
In this approach, building component losses are simulated based on the repair costs associated with different damage states, which are obtained from the component fragility functions. Since the damage states are defined to be discrete, the simulation of correlated discrete damage states is a difficult task. It is shown by Shome et al. (2014) that the correlation between discrete damage states can be easily established by representing the fragility functions using a “damage-capacity” formulation. In this approach, instead of defining the correlation between the discrete damage states, the correlation between the damage capacity of different components is defined. The damage capacities of all the components can be jointly simulated from a multivariate lognormal

distribution (the multivariate nature is assumed). The distribution is defined by the means and dispersions of the capacity based on the fragility functions and the correlation between those capacities.

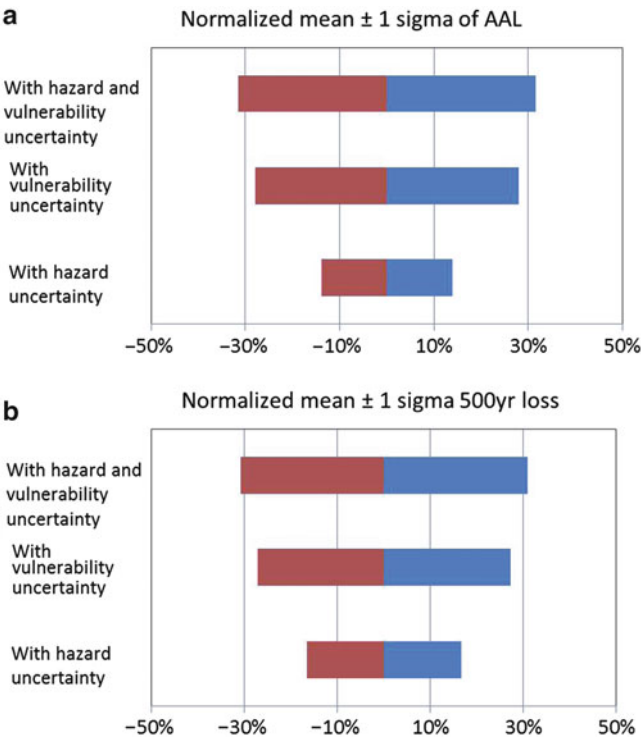
The repair costs conditioned on DM are also modeled as a multivariate lognormal random variable with the median and dispersion and correlations of repair costs between different damage states over the height of buildings. The losses of different components across different damage states are expected to be correlated because of common factors involved in repairs, such as the contractor hired for the repairs and the material costs. The parameters for the distribution of repair cost of buildings are based on the study by Ramirez and Miranda (2009) on the cost of contractors for a large number of building projects. Shome et al. (2014) have shown how the information of repair cost of the contractors can be simulated to estimate the lognormal parameters of repair cost for a specific building or a group of buildings.

The total loss is computed by summing up the sample components losses at each story. This is normalized by the building value to get the damage ratio for the building. The scatter plot of simulation of a 20-story modern steel building in Los Angeles, California, is shown in Fig. 10 as a function of $S_a(T_0)$. The mean and the CoV of the damage ratios are then computed by binning the data points into $S_a(T_0)$ bins to obtain the building VF.

Seismic Loss Assessment,
Fig. 10 Damage ratios of a modern 20-story SMF in Los Angeles for all the simulated $S_a(T)$ vectors corresponding to a wide range of $S_a(T_0)$ values



Seismic Loss Assessment,
Fig. 11 Impact of different sources of epistemic uncertainties in the loss results of a modern 20-story steel moment-frame building in Los Angeles. (a) Average annual loss (AAL). (b) 500-year loss



Seismic Loss Results and Uncertainty in Those Results

The VFs developed in the previous sections are used to carry out seismic risk assessments of buildings. The VFs are integrated with the hazard curves to compute the AAL and the return period or EP losses. As discussed, the methodology

developed by Jayaram et al. (2012) can be used to account for epistemic uncertainties in the vulnerability functions by developing a suite of functions representing the uncertainty. The losses are then sampled from these functions to compute the epistemic uncertainty of different loss matrices used for building risk assessment like AAL, EP loss, etc. Figure 11 illustrates the normalized

64 % confidence interval (CI) of the AAL and the 500-year return period loss of a modern 20-story steel moment-frame building in Los Angeles. The relative extents of contributions of the epistemic uncertainties in vulnerability and hazard to the total epistemic uncertainty would, however, depend on whether the loss assessment is carried out for a single building or a portfolio of buildings. For a single building, as seen in the figure, the VF uncertainty has a large contribution to the total uncertainty. For a portfolio of buildings, however, the uncertainty in losses across different locations will be mostly independent and will average out. The hazard uncertainty, however, is highly correlated across multiple locations and will not average out. Therefore, for a portfolio, hazard uncertainty will have higher contribution to the total uncertainty.

Summary

Seismic loss assessment of buildings plays an important role in managing earthquake risk for public agencies, insurance companies, as well as individual stakeholders. The loss assessment procedure has evolved over the years – from expert opinion to very rigorous analytical approach. Since the analytical procedure is very complex requiring accurate information of a large number of parameters, there is significant uncertainty in accurate estimation of building losses. Hence loss estimation based on observed damage characteristics of buildings in historical earthquakes only can provide accurate loss results. This chapter has discussed in detail the development of building vulnerability functions (VF) based on observed loss data. The data, particularly those that are from insurance companies, can be censored because of insurance deductibles. This is traditionally dealt with maximum likelihood approach and this approach has been described here. But this approach requires a large number of very high-quality building-specific loss data, which is generally not available. The other disadvantage of data-based vulnerability functions is that the

results cannot be extrapolated from one region to the another because of regional building design and construction practices, which are generally unique. Hence this chapter described development of building vulnerability functions based on expert opinion. It is found that this approach can provide significantly biased results and so this is used rarely nowadays. This approach, however, would be used for assessing seismic risk in the future when time and resources are limited.

Over the years, different analytical approaches have been developed and only a few of those have been discussed in this chapter due to limitations in space. This chapter discussed in detail development of vulnerability functions following capacity spectrum method by HAZUS, assembly-based vulnerability, and a comprehensive approach following Jayaram et al. (2012). Analytical approaches have de facto become standard in the recent years for seismic risk calculations. Although these approaches require a large number of parameters, researchers are actively estimating those based on laboratory experiment and observed damage data. It is well known that there is significant uncertainty in the estimation of those parameters vis-à-vis VFs based on those parameters, but the uncertainty in VF can be easily estimated based on the analytical approaches. The importance of consideration of epistemic and aleatory uncertainty and the methodology in order to estimate those in building vulnerability has also been discussed here.

Cross-References

- ▶ [Analytic Fragility and Limit States \[P\(EDPI IM\)\]: Nonlinear Dynamic Procedures](#)
- ▶ [Empirical Fragility](#)
- ▶ [Insurance and Reinsurance Models for Earthquake](#)
- ▶ [Seismic Risk Assessment, Cascading Effects](#)
- ▶ [Seismic Vulnerability Assessment: Masonry Structures](#)
- ▶ [Seismic Vulnerability Assessment: Reinforced Concrete Structures](#)

References

- ATC-13 (1985) Earthquake damage evaluation data for California. Applied Technology Council, Redwood City
- ATC-58 (2011) Guidelines for seismic performance assessment of buildings (75 % draft). Applied Technology Council, Redwood City
- Benjamin JR, Cornell CA (1970) Probability, statistics and decisions for civil engineers. McGraw-Hill, New York
- Bonowitz D, Maison BF (2003) Northridge welded steel moment-frame damage data and its use for rapid loss estimation. *Earthquake Spectra* 19(2):335–364
- Bozorgnia Y, Abrahamson NA, Atik LA et al (2014) NGA-West2 research project. *Earthquake Spectra*. doi:10.1193/072113EQS209M
- Cornell CA, Krawinkler H (2000) Progress and challenges in seismic performance assessment. *PEER Center News* 3(2), <http://peer.berkeley.edu/news/2000spring/>
- CUREE (2001) Improving loss estimation for woodframe buildings: CUREE-Caltech woodframe project. In: Porter KA, Beck JL, Seligson HA, et al (eds) *Element 4, economic aspects*. Consortium of Universities for Research in Earthquake Engineering, Richmond
- FEMA 355F (2000) State of the art report on performance prediction and evaluation of steel moment-frame buildings. Federal Emergency Management Agency, Washington, DC
- Haselton CB (ed) (2009) Evaluation of ground motion selection and modification methods: predicting median interstory drift response of buildings, PEER report 2009/01. PEER GSM, Pacific Engineering Research Center, University of California, Berkeley
- HAZUS-MH MR3 (2003) Multi-hazard loss estimation methodology: earthquake model. Department of Homeland Security, Washington, DC
- III (2013) Earthquakes: risk and insurance issues. Insurance Information Institute. http://www.iii.org/issue_updates/earthquakes-risk-and-insurance-issues.html. Accessed 8 Nov 2013
- Jaiswal KS, Wald DJ, Perkins D et al (2013) Estimating structural collapse fragility of generic building typologies using expert judgment. In: 11th international conference on structural safety and reliability (ICOSSAR), New York
- Jayaram N, Shome N, Rahnema M (2012) Development of earthquake damage functions for tall buildings. *Earthquake Eng Struct Dyn* 41(11):1495–1514
- PEER-2011/05 (2011) Case studies of the seismic performance of tall buildings designed by alternative means. In: Moehle J, Bozorgnia Y, Jayaram N et al (eds) *Task 12 report for the tall buildings initiative*. Pacific Earthquake Engineering Research Center Report, Berkeley
- Porter KA, Kiremidjian AS, LeGrue JS (2001) Assembly-based vulnerability of buildings and its use in performance evaluation. *Earthquake Spectra* 17(2):291–312
- Ramirez CM, Miranda E (2009) Building-specific loss estimation methods & tools for simplified performance-based earthquake engineering, report no 171. The John A. Blume Earthquake Engineering Center, Stanford University, Stanford
- Shome N, Bazzurro P (2009) Comparison of vulnerability of a new high-rise concrete moment frame structure using HAZUS and nonlinear dynamic analysis. In: 10th international conference on structural safety and reliability (ICOSSAR), Osaka
- Shome N, Cornell CA (2000) Structural seismic demand analysis: consideration of collapses. In: 8th ASCE specialty conference on probabilistic mechanics and structural reliability, St Luis
- Shome N, Jayaram N, Rahnema M (2012) Uncertainty and spatial correlation models for earthquake losses. In: 15th world conference on earthquake engineering (WCEE), Lisbon
- Shome N, Jayaram N, Krawinkler H et al (2014) Loss estimation of tall buildings designed for the PEER tall building initiative project. *Earthquake Spectra* (accepted for publication)
- Steinbrugge KV (1982) Earthquakes, volcanoes and tsunamis – an anatomy of hazards. Skandia America Group, New York
- Wesson RL, Perkins DM, Leyendecker EV et al (2004) Losses to single-family housing from ground motions in the 1994 Northridge, California, Earthquake. *Earthquake Spectra* 20:1021–1045

Seismic Monitoring of Volcanoes

Glenn Thompson

School of Geosciences, University of South Florida, Tampa, FL, USA

Synonyms

Volcano-seismic monitoring

Introduction

Seismicity beneath a volcano usually increases before an eruption because magma and volcanic gas must first force their way up through fractures and passageways. When magma and volcanic gases or fluids move, they will either cause rocks to break or cracks to vibrate. When rocks break, high-frequency earthquakes are triggered. When cracks vibrate, either low-frequency

earthquakes or a continuous shaking called volcanic tremor, which can last from minutes to days, occurs.

Volcanic earthquakes often occur in swarms, which are clusters in time and space of similar earthquakes without an obvious mainshock. Volcano seismologists look for changes in the rate, size, and location of earthquakes and for the occurrence of swarms and tremor to forecast eruptions and to evaluate whether a volcanic eruption is intensifying or ending. Volcanic hazards including explosive eruptions, rockfall, pyroclastic flows, and lahars also cause ground vibrations and can be identified by their seismic signatures. This makes it possible, in principle, to detect them as they happen and issue warnings.

This entry discusses volcano-seismic monitoring from the viewpoint of a scientist leading a seismic monitoring program at a volcano observatory. Seismic monitoring records continuous, high-sample-rate data on the internal state of the volcano. Other monitoring techniques require manual labor to collect or process data, rely on daylight and good weather conditions, only detect volcanic activity once it has reached the surface, or have a low sample rate. For these reasons, seismic monitoring is the backbone of most volcano observatories. A volcano-seismic monitoring program comprises of a seismic network, a telemetry system, data acquisition, and alarm, analysis, and archival systems. At large observatories, personnel may include electronics engineers, volcano seismologists, seismic analysts, software developers, and network administrators. At small observatories, one person may cover most or all of these roles.

A volcano observatory has to be able to process large quantities of streaming data, detect changes in the volcanic system immediately, and respond without having to do a lot of manual analysis. This requires high levels of automation and systems engineering. While top priority is interpreting real-time data using established knowledge from the scientific field of volcano seismology, other important roles are troubleshooting data problems and conducting (or promoting) research that may lead to a better

understanding of the volcano or lead to improved monitoring tools.

Much of the equipment and software used for volcano-seismic monitoring were originally developed for regional earthquake monitoring. The main difference is that the software needs to be augmented because of the diverse range of seismic signals recorded at volcanoes, many of which elude standard earthquake detection and location techniques. These signals, and the framework in which they are commonly interpreted, are discussed in the next section. The sections that follow describe volcano observatories including the operations room which is the command center during a crisis, the design of volcano-seismic networks, the real-time seismic monitoring and data analysis tools found in most observatories, and the challenges of maintaining this infrastructure and managing data effectively. In the final section, other monitoring techniques employed at observatories are briefly discussed.

Volcano-Seismic Signals

Volcano seismologists are most interested in signals which are anomalous, because they have high amplitudes, characteristic waveforms, or unusual frequency behavior. Given the variety of volcanoes around the world, with staggeringly different eruptive styles, magma compositions, and viscosities, it is perhaps remarkable that some common types of seismic signals are observed. These signals can be broken down into three broad categories: volcanic earthquakes ("events"), continuous signals, and surface signals. Each of these is discussed below. Any volcano may exhibit some or all of these at some time and perhaps may also exhibit exotic signals unique to that volcano. Classification is important because each type of signal may represent a different physical mechanism. However, classification is problematic.

By analyzing the spatial and temporal patterns between different signal types and volcanic activity, a greater understanding of a particular

volcano may be revealed. There are common patterns of behavior, gleaned from analyzing the seismic data from many volcanic eruptions. These are encapsulated in the Generic Volcanic Earthquake Swarm Model (McNutt 1996), a sequence observed at many volcanoes, the main features of which are volcano-tectonic earthquake swarms, followed by low-frequency earthquake swarms, tremor, and eruption.

For a more detailed discussion of volcano-seismic signals and many other topics mentioned in this chapter, Wasserman (2012) is highly recommended.

Volcanic Earthquakes

Volcano-Tectonic Earthquakes

Volcano-tectonic (VT) earthquakes are tectonic earthquakes that occur near active volcanoes. The physical mechanism is shear failure and slip on a fault plane triggered by magma ascent or the relaxation that occurs after magma is erupted. VT earthquakes have clear P- and S-waves (if recorded with a good signal-to-noise ratio) and high-frequency content (>5 Hz). VT earthquakes can be located in the same way as other tectonic earthquakes, i.e., by using the differential travel times between the P and S phases across a seismic network. However, most VT earthquakes are small ($M_L = -0.5$ – 1.5), and poor signal-to-noise ratio often prevents identification of phases, so in practice many (perhaps most) VT earthquakes cannot be located. Velocity models for volcanoes are often poorly determined, making absolute depths unreliable. Trends in relative depths are a useful diagnostic tool, however, and may indicate the rise of magma toward the surface. VT earthquakes frequently occur in swarms that consist of many similar-sized events and do not occur in mainshock-aftershock sequences where one single event dominates.

Long-Period Earthquakes

Long-period (LP) earthquakes are unique to volcanic regions. They have emergent onsets and a narrow frequency range with a peak frequency typically from 1 to 4 Hz. Tornillos are LP

earthquake events with a particularly monotonic appearance and an exponentially decaying tail. LP earthquakes lack discernable P or S phases; consequently, most are not located. Evidence from Soufrière Hills Volcano (Montserrat) that they are often associated with venting from the surface of the dome and often trigger rockfall suggests they may originate at depths of less than 1 km. LP earthquake focal mechanisms reveal a volumetric component which is evidence of a fluid phase. LP earthquake swarms in volcanic systems are often associated with eruptions or intrusions and are believed to be due to processes such as pressure-induced oscillations of fluid-filled cracks in magmatic and hydrothermal systems.

Hybrid Earthquakes

Hybrid earthquakes have a high-frequency P-wave onset, typical of a VT earthquake, followed by a long-period tail. They may represent triggering of an LP earthquake by a VT earthquake. They typically occur in swarms and may be indicative of magma intrusion or extrusion.

Low-Frequency Earthquakes

Collectively, long-period and hybrid earthquakes are referred to as low-frequency earthquakes. There may be a continuum between long-period and hybrid earthquakes.

Deep Long-Period Earthquakes

Some volcanoes produce low-frequency earthquakes (e.g., 20–40 km depth), particularly in the early stages of unrest. These have been called deep long-period (DLP) earthquakes. These have emergent P and S phases, are rich in frequencies below 5 Hz, and are inferred to represent movement of deep-seated magma and associated fluids in the mid-to-lower crust. They look like VT earthquakes with the high frequencies filtered out, perhaps because they occur in a highly attenuating region.

Explosion Quakes and Very-Long-Period Signals

Explosion quakes are signals that accompany Strombolian or other (larger) explosive eruptions.

These signals are identified by the occurrence of an airwave which is caused by expanding gas accelerated at the vent exit. This wave mainly travels through the air with the typical speed of sound (330 m/s at 20 °C).

With the advent of broadband volcanic seismology, very-long-period (VLP) signals have been observed in seismograms from some volcanoes. Many of these are broadband versions of explosion quakes. VLP signals may be produced by the rapid expansion of a large gas volume at shallow depth within the conduit. The gas expansion might result from shallow gas coalescence and expansion or from expansion of a gas slug formed at greater depth.

Continuous Signals

Swarms

Earthquake swarms are sequences of earthquakes closely clustered in space and time without a single mainshock. Volcanic earthquakes often occur in swarms, whereas nonvolcanic earthquakes usually follow a mainshock-aftershock sequence. The Gutenberg-Richter law describes the relative frequency of occurrence of earthquakes of different magnitudes, and this is encapsulated in a parameter called the b-value (see Sanchez et al., “► [Frequency-Magnitude Distribution of Seismicity in Volcanic Regions](#),” this volume). Volcanic earthquake swarms typically have b-values higher than one (most of the energy is in small earthquakes), whereas mainshock-aftershock sequences typically have $b \leq 1$ (most of the energy is in a single mainshock). Volcanic earthquakes occur in hot, highly heterogeneous material containing many small faults. Nonvolcanic earthquakes tend to occur in more homogeneous material with failure on a single larger fault, which then loads adjacent faults, causing many of them to fail producing aftershocks. Low-frequency earthquake swarms often contain one or more families of repeating earthquakes. Each family is identified by a unique waveform and is the result of the same nondestructive source process being activated repeatedly in the same location.

Benoit and McNutt (1996) examined the reports of over 600 swarms to compile the Global Volcanic Earthquake Swarm Database. They identify three main types of volcanic earthquake swarm. Type 1 swarms begin before an eruption and have a mean (and mode) duration of 8 days. Type 2 swarms begin coincident with, or during, an eruption. Type 3 swarms (39 % of records) were not associated with eruptions and have a mean duration of 3.5 days and a mode of 1.5 days. The most common depth for swarms is 2–3 km. Their study does not distinguish between different types of volcanic earthquakes. They suggest it may also be biased by the underreporting of swarms not associated with eruptions.

Tremor

Volcanic tremor is a narrowband (usually 1–4 Hz), continuous vibration thought to be due to sustained subsurface movement of magma or volatiles and is often observed before explosive eruptions. It may last from a few minutes to months in duration. Tremor has similar spectral characteristics as low-frequency earthquakes. Harmonic tremor shows one or many regularly spaced overtones in addition to a fundamental frequency. Sometimes spectral peaks in harmonic tremor glide upward (or downward) in frequency over as little as a few minutes. Eruption tremor, a continuous vibration coincident with explosive eruptions, has a wider frequency range (0.5–10 Hz). Dome collapses, lahars, weather storms, and telemetry problems can all produce signals that could be confused with tremor.

Tremor may be the result of continuous excitation of the source that produces low-frequency earthquakes. The resonance of a fluid-filled conduit is one model for the origin of tremor, as it is for low-frequency earthquakes. Interface waves traveling along the crack or conduit wall can produce overtones, like an organ pipe. Gliding lines could be caused by a change in the sound speed of the fluid (e.g., due to a change in bubble density) or a change in a length of the section of conduit that is resonating (e.g., a change in nucleation depth).

Tremor may also be a superposition of low-frequency earthquakes. There are many observations of low-frequency earthquake swarms merging into volcanic tremor. Modeling has shown that harmonic tremor could appear when overlapping earthquake signals occur at regular time intervals that differ by less than 2 % (Powell and Neuberg 2003). As this regular rate of earthquakes gradually increases (or decreases), gliding spectral lines are produced. In this model, tremor is a low-frequency earthquake swarm.

Surface Signals

Surface signals are those generated by hazardous surface processes such as rockfall, pyroclastic flows, and lahars. Since these processes occur at the surface, they mainly generate surface waves.

Rockfall and Pyroclastic Flow Signals

Actively growing lava domes are highly unstable, and blocks can be observed falling almost continuously, disintegrating into smaller blocks and plumes of hot ash. Most of these rockfalls are small, but many generate detectable seismic signals. If larger parts of the dome collapse, or if there is an explosive component destabilizing the blocks, a pyroclastic flow may be generated. Pyroclastic flows are more energetic, produce vigorously convecting ash clouds, behave more like a fluid (less friction), and are therefore able to travel farther and at higher speeds. There is a continuum from the smallest rockfalls to the largest pyroclastic flows, and they mostly are spawned from the part of the dome that is actively growing. Pyroclastic flows can also be generated by the collapse of an eruption column.

Rockfall signals are emergent, contain a wide range of frequencies (1–10 Hz, peaking 3–4 Hz), and are dominated by surface waves. They have been located by exploiting the amplitude distribution of rockfalls signals across the seismic network. On Soufrière Hills Volcano (Montserrat), some rockfall and many pyroclastic flows have a long-period precursor, and it is possible that this is an explosive degassing signal. At Unzen Volcano (Japan), pyroclastic flow signals are found to comprise three parts. First, a dome collapse is

observed simultaneously with a small LP signal. Second, free-falling blocks impact the slope below (and fragment) simultaneous with a 0.5 Hz signal. Third, the fragmented material generates a rockfall signal.

Dome collapse signals are the superposition of many rockfall events, which may be occurring simultaneously on different flanks of the volcano. Dome collapses may last minutes to hours. From studying dome collapse signals, it is possible to tell how long dome collapses lasted and which phases of the collapse were most energetic. Several major dome collapses at Soufrière Hills Volcano (Montserrat) and Merapi Volcano (Indonesia) have been triggered by intense rainfall.

Lahar Signals

The composition of lahars varies from muddy water to highly erosive, dense mixtures of wet ash, rocks, and boulders that set like concrete. Lahars usually occur during or immediately after heavy rainfall. Barclay et al. (2006) found that about 2 cm per hour of rain falling on unconsolidated materials was sufficient to trigger lahars on Soufrière Hills Volcano (Montserrat). However, hot volcanic material can generate lahars by mixing with crater lakes or with snow and ice on glaciated volcanoes. Volcanic activity may also melt the base of a glacier and cause an outbreak flood that briefly rivals the force of a major river. These are particularly common in Iceland and are known in Icelandic as *jökulhlaups*.

Lahar signals are tremor-like signals that can be distinguished from pyroclastic flow signals by their duration (tens of minutes to hours), higher-frequency content (6–10 Hz, up to 100 Hz in some cases), and slower speeds; they show amplitude peaks on different stations perhaps minutes apart as they travel down valley. To improve the detection of lahar signals, seismic stations, video cameras, and tripwires can be added at various positions adjacent to (or within) the valley.

Difficulties of Event Classification

Classification of volcano-seismic signals is difficult. The signal recorded is a convolution of the source signature, propagation and site effects,

and the instrument response, superimposed on noise which varies spatially and temporally. The source signature may be the result of complex interactions between a multiphase fluid and an unknown geometry of dikes and conduits. Further from the source, the signal-to-noise ratio decreases, and increased scattering and separation between P, S, and surface waves lengthen the signal, making it less clear. Inelastic attenuation may be greatest at shallow depths where poorly consolidated, highly fractured material results in higher frequencies being filtered out. Attenuation may also be significant at depths of 20 km or more, because of high heat flow and partial melt. Similar-looking signals at two different volcanoes, or even at the same volcano, might be caused by different physical mechanisms.

Classification is also subjective. Different seismic analysts may disagree on the classification of a particular event. Discrepancies in classification also appear in published literature: an LP in Lahr et al. (1994) looks similar to a hybrid in Luckett et al. (2007). Classification might artificially separate signals that lie along a continuum. The classification scheme may also evolve during volcanic unrest as a wider variety of signals are recorded. For these reasons it is helpful to reassess the event classes used before interpreting trends in their rates of occurrence or locations.

Terminology also differs. Some of the terms describe the frequency content of the signal and others imply a physical mechanism. VT earthquakes are also called high-frequency events and A-type events. LP earthquakes are also called B-type events. Collectively, long-period and hybrid events are called low-frequency (LF) events, because they are both dominated by low-frequency coda. Low-frequency seismicity includes low-frequency events and volcanic tremor.

Interpreting Volcanic Seismicity

A common pattern of volcanic seismicity is described by the Generic Volcanic Earthquake Swarm Model (McNutt 1996). Background seismicity varies from volcano to volcano and can only be established for a particular volcano through a long-established seismic monitoring

record. The first sign of unrest is often an increase in the rate of VT earthquakes, indicating rock fracture due to changes in stresses caused by rising magma. In some cases it may be the occurrence of DLP earthquakes.

When magma reaches shallower depths, volatiles exsolve, and LP earthquakes are recorded. Volatiles cause a reduction in the acoustic velocity and an increase in the impedance contrast of magma cavities, trapping energy and leading to longer, lower frequency signals. As the volatile content increases, more energy is trapped, and this may lead to LP earthquake events called *tornillos*. Narrowband tremor may be generated by continuous vesiculation as magma rises further or by the boiling of groundwater. If groundwater is heated rapidly, a phreatic explosion may result.

At any time in the sequence, magma rise may stall. This may be the end of the unrest, or it may recommence as more magma is injected at depth adding more heat and volatiles to the system. Tremor due to groundwater boiling may subside as the system dries out, leading to a period of relative quiescence. For basaltic systems, an eruption may then occur without further warning.

For andesitic to rhyolitic systems, hybrid earthquake swarms often occur due to repeated shear failure of viscous magma as it gets close to the surface. Hybrid earthquake swarms are frequently associated with growth of a lava dome. If the events merge into a continuous tremor, harmonic tremor and gliding spectral lines might be observed and are the strongest signs that an explosive eruption may be about to occur. During an explosive eruption, violent ground shaking typically manifests as a broadband tremor signal. Explosion earthquakes may be recorded, identifiable by the shockwave that travels through the air and is coupled back into the ground.

Explosive eruptions are short-lived, and hopefully before they occur, any vulnerable populations will have been evacuated. But extrusive eruptions may last for years or decades, with varying levels of activity, spawning rockfall, pyroclastic flows, and lahars. Monitoring these long-lived eruptions presents challenges of its own, since communities are often in close

proximity. Escalations in activity are often preceded by short sequences of VT earthquakes, low-frequency earthquakes, and tremor, perhaps suggesting a batch of new magma rising toward the surface. However, volcano-seismic data have also been used to track the rate, energy, and location of debris flows and estimate extrusion rates.

An important question to answer once an eruption is underway is when will it stop. The first sign may be the cessation of low-frequency seismicity, indicating that volatiles are no longer present in the system, which may be confirmed with gas flux measurements. More significant may be the occurrence of VT earthquakes at depth with fault-plane solutions consistent with magma withdrawal (Roman et al. 2006). Geodetic data may also indicate deflation of the volcanic edifice.

A more detailed discussion of the interpretation of volcano-seismic data is beyond the scope of this entry. There are several excellent summaries of our evolving understanding of volcano-seismology including Chouet and Matoza (2013), Chouet (2003), and McNutt (2000, 2002, 2005).

Observatories

Volcano observatories vary greatly in their level of sophistication. The simplest observatory may be a hut on the flank of a volcano, with a single seismograph recording on paper on a revolving drum. A modern observatory, however, will have a network of seismic stations with data telemetered to the observatory using a variety of communication systems such as FM radio, satellite, and cellular networks. The observatory may be on the flanks of the volcano, within a few miles of the active vent. Or it may be tens or hundreds of miles away, colocated with a university or a government agency. Many observatories monitor a single volcano, but some monitor several. The Alaska Volcano Observatory has operated seismic networks on as many as 32 volcanoes. The Japan Meteorological Agency monitors 47 volcanoes with real-time seismic data. This poses unique challenges: how to monitor so many

volcanoes in parallel, how to interpret the data without being able to see the volcanoes, how to maintain so many networks, and how to manage all the data.

Participation in a volcanic crisis, is an opportunity to help society and also witness some of nature's most spectacular phenomena. Having the right team is crucial. The leader of a volcano observatory must be experienced in volcano monitoring and be committed to the job. The work can be stressful, the hours long, and hard decisions have to be made in the heat of the moment, e.g., regarding evacuations. Weekly meetings play a vital role at many observatories by pulling the team together, integrating data from many monitoring techniques, identifying technical problems, and prioritizing work. All observatory staff must conduct themselves professionally and have the authority to perform the roles assigned to them. Fostering good communications within the observatory and with the authorities and the public builds trust and enhances public safety.

The backbone of a volcano observatory is the seismic monitoring program, and the heart of the observatory is the operations room. It is from there that seismologists track live data streams, coordinate field teams and warn them of hazardous activity, and alert the authorities (and the public) to escalations in activity. When volcanoes are at background level, the operations room may be unoccupied. Periodic data checks, perhaps coupled with automatic alarm systems, may be enough to keep seismologists abreast of significant changes in seismicity. When unrest begins, periodic checks may become more frequent. At some point the operations room is activated and manned 24 h a day.

These different stages of volcanic activity, from background to unrest to impending eruption, greatly impact the level of seismic monitoring that can be done. At background level, seismologists may be able to conduct research. When a crisis is underway, it will be difficult to do anything more than interpret available real-time data using operational tools that are in place. Part of the job of an observatory seismologist is to engineer the infrastructure of the seismic

monitoring program and operations room to be able to respond effectively during a crisis. This infrastructure includes seismic networks and software systems used to monitor volcanic seismicity. These are the subjects of the next two sections.

Volcano-Seismic Networks

Around 550 volcanoes have erupted in historical times, and about 200 of these are seismically monitored. The best networks tend to be around volcanoes that pose a particularly high threat to large population centers, e.g., Etna and Vesuvius (Italy), Rainier and St. Helens (continental USA), and Kīlauea (Hawaii). Many volcanoes are monitored only by a single station, often as part of a regional network. Nevertheless, by tracking the number and cumulative energy of different types of seismic signals recorded each day on a single station, a volcanic eruption can be anticipated. It may then be possible to rapidly deploy sufficient additional stations to locate earthquakes and forecast where and when an eruption may take place.

The most commonly used seismometers for volcano monitoring are short-period sensors. These have a corner frequency of 0.5–2 Hz and come in single (vertical) and three component varieties. Short-period seismometers are usually deployed with analog telemetry. Both have a dynamic range (the ratio between the largest and smallest amplitudes that can be represented) of only a few thousand. This results in signal amplitudes often being “clipped,” which greatly diminishes the monitoring and research value of the data. Three-component portable broadband seismometers, available since the late 1980s, use a force-feedback circuit and typically have a corner frequency of 30–120 s and a dynamic range of several million, allowing them to record signals on-scale even right up on the volcano summit. They have to be coupled with 24-bit digital telemetry systems, which have a similar dynamic range.

Analog-telemetered data are time stamped at the observatory by a time signal from a GPS clock, but modern field digitizers have an input

for a GPS clock and can time stamp data on site (which is more accurate). On-site data recording enables data to be retrieved or retransmitted later if there is a communication outage. Two-way telemetry allows data packets to be resent automatically if they are not received intact and allows troubleshooting and reconfiguring of stations from the observatory without the expense and delay of a site visit.

While a digital broadband network offers many advantages over an analog short-period network, the equipment is more expensive and power requirements are 2–3 times higher, meaning that additional solar panels and batteries are needed. Broadband sensors also require precise leveling. Many analog networks have been upgraded to digital telemetry to improve dynamic range, and most new networks use digital telemetry, but it is still common to find analog telemetry at volcano observatories.

Designing a volcano-seismic monitoring network is complicated and involves multiple trade-offs. For the same budget a volcano observatory may be able to install a few digitally telemetered broadband stations or many more analog-telemetered short-period stations. The best choice depends on the goal.

For a volcano that is far removed from population centers, the goal might be to detect eruptions so that aviation authorities can be warned of hazardous ash clouds, requiring a minimal level of monitoring. For a densely populated region around a frequently active volcano, or one capable of devastating eruptions, higher quality monitoring is needed to provide as much lead time as possible. The magnitude detection threshold may vary from 2 for a regional network to below 0 for a dense volcano-seismic network. Moran et al. (2008) identified four levels of volcano-seismic monitoring and the number of short-period, broadband, strong motion, and infrasound sensors needed within different radii to make those levels of monitoring achievable. Their findings are summarized in Table 1.

A key consideration for locating earthquakes accurately is to minimize the azimuthal gap. A ring of stations, which encloses most of the volcanic-earthquake epicenters, is ideal with a

Seismic Monitoring of Volcanoes, Table 1 Recommended instrumentation to achieve different levels of volcano-seismic monitoring from Moran et al. (2008)

Level	Goal	Recommendation
1	Minimal monitoring/eruption detection Detect $M > 1.5$ earthquakes; crudely locate $M > 3$ earthquakes	Site a total of five seismic stations within 200 km, including two within 50 km of the volcanic center
2	Limited monitoring/unrest detection Detect $M > 1$ earthquakes; crudely locate $M > 2$ earthquakes; determine event type; detect energetic seismic tremor	Site a total of five seismic stations within 50 km, including two within 10 km of the volcanic center
3	Basic real-time monitoring Detect $M > 0.5$ earthquakes; accurately locate $M > 1$ earthquakes; determine event type; detect seismic tremor; on-scale recording of energetic seismicity on at least one station; detect very-long-period events Detect changes in travel time; detect broad-scale changes in seismic velocity Use fault-plane solutions and b-values to determine generalized stress fields near the volcanic center	Site six to eight seismic stations within 20 km of the volcanic center, including two or three stations with at least one three-component sensor and at least one broadband station, within 5 km
4	Advanced real-time monitoring Detect and accurately locate $M > 0$ earthquakes; determine event type; detect and crudely locate seismic tremor; on-scale recordings of energetic seismicity on multiple stations; detect and crudely locate very-long-period events and other very-low-frequency seismicity Determine detailed source properties of tornillos; construct 3-D velocity models (provided local seismicity is sufficient) Detect explosions and possible infrasonic precursors to explosions at restless and (or) frequently active volcanoes Detect detailed stress-field changes by calculating well-constrained fault-plane solutions and (or) moment tensors, mapping b-values at high spatial resolution, and detecting changes in S-wave-splitting directions over time	Site 12 to 20 seismic stations within 20 km of the volcanic center, including at least six broadband stations, as many as possible within 5 km; at least one strong-motion station; and at least two infrasonic stations (with at least two infrasonic sensors per station) at erupting, restless, and (or) frequently active remote volcanic centers

few stations closer to the volcanic center where they can help detect smaller seismic signals and be used to locate summit events with greater precision and improve depth resolution. For stratovolcanoes, all the earthquakes may be within 1–2 km of the volcanic center, whereas for calderas the seismicity might be diffused over an area 15–20 km in radius. It is useful also to have real-time data from at least one more distant station (which may be part of a regional network) to help constrain the depths of deeper events, and to discriminate more effectively between regional and local volcanic earthquakes. These steps make it possible to locate earthquake reasonably well relative to each other, but how well

they match the true locations of the earthquake will depend on the velocity model and to a lesser degree, on the algorithm used to locate the events. The sophistication of the velocity model may vary from a simple one-dimensional regional model with constant velocity layers to 3-D models determined from seismic tomography.

Site selection depends on the geology, topography, accessibility, and noise. It is preferable to install seismometers in solid bedrock, but volcanoes are typically comprised of layers of ash, flow deposits, and boulders. Repeaters may be required on ridges to rebroadcast signals or boost them if transmitting over tens of kilometers, but represent additional expense and

Seismic Monitoring of Volcanoes, Table 2 Some of the best volcano-seismic monitoring networks. The numbers in parentheses indicate number of borehole instruments. All of these networks surpass level 4 as defined in Table 1 of Moran et al. (2008)

Volcano	Stations within 5 km			Stations within 20 km		
	Broadband	Short period	Total	Broadband	Short period	Total
St. Helens (USA)	2	7	9	2	16 (4)	18 (4)
Soufrière Hills (Montserrat)	7 (1)	0	7 (1)	9 (1)	6 (4)	15 (5)
Piton de la Fournaise (Reunion)	14	6	20	16	17	33
Halemaumau, (Hawaii)	12	5	17	19	14	33
Stromboli (Italy)	18	0	18	18	0	18
Etna (Italy)	?	?	13	?	?	34
Vesuvius (Italy)	?	?	17	?	?	21
Erebus (Antartica)	6	6	12	6	6	12

potential points of failure. Sites that require helicopter access will be expensive to maintain, so often it is better to find sites that can be easily reached with four-wheel-drive vehicles or on foot. Sites should be far from traffic and other human noise and also away from any tall obstacles that may be vibrated by the wind, such as trees, cliffs, and radio towers. Burying sensors a few feet below the surface helps suppress high-frequency wind noise and low-frequency noise due to temperature and pressure variations.

Among the best volcano-seismic networks are those used to monitor Piton de la Fournaise (Reunion), Halemaumau (Hawaii), St. Helens (USA), and Soufrière Hills (Montserrat) volcanoes (Table 2). According to the criteria in Table 1, these are all level four networks. A map of the digital seismic network that has been operational in Montserrat since 2006 is shown in Fig. 1. This utilizes a mixture of Guralp CMG-40T broadband and Mark L4-C short-period seismometers, coupled with Guralp DM24 digitizers and FreeWave Spread Spectrum serial port radios. One of the broadband stations has a CMG-3 T seismometer in a 30-m borehole and less than 4-km from the dome. Four more sites feature a dilatometer and 2-Hz seismometer in 200-m boreholes, coupled with surface GPS. The Piton de la Fournaise network also uses CMG-40T and L4-C seismometers, coupled with Kinometrics Q330 digitizers and a combination of wireless internet and FM radio telemetry.

Real-Time Monitoring

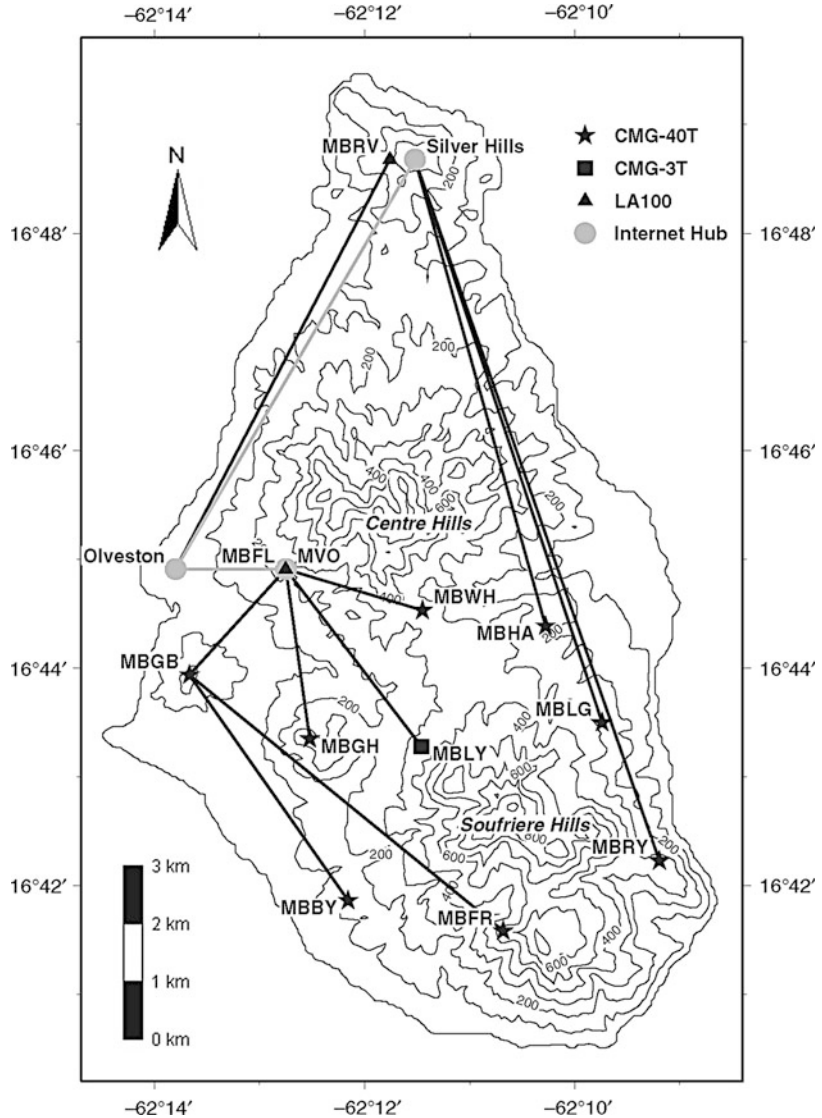
Volcanic activity may escalate suddenly, and the ability to rapidly identify anomalous volcano seismicity is critical. Real-time data visualization systems allow scientists to rapidly assess multiple parameters derived from seismic data such as hypocenters, magnitudes, event rates, spectral variations in tremor, etc. Alarm systems alert scientists to large events and the occurrence of high amplitude tremor and volcanic earthquake swarms. Data-rich websites and remote desktop connections allow scientists to respond rapidly to alarm: This is convenient and enhances safety. The US Geological Survey, through its Earthquake and Volcano Hazards Programs, has played a pivotal role in disseminating free, open-source software to aid in volcano-seismic monitoring. This section describes common techniques and software used to examine continuous seismic data and events in more detail.

Continuous Data

Digital Helicorder Plots
The most basic form of real-time data display is the helical drum recorder (often called “helicorders” or “drums”). A pen etches a trace on a smoked sheet (or draws an ink trace on a blank sheet) of paper wrapped around a cylindrical drum. Helical drum recorders have played a vital role in volcano-seismic monitoring, allowing rapid visualization of seismic amplitudes and event identification of signal types.

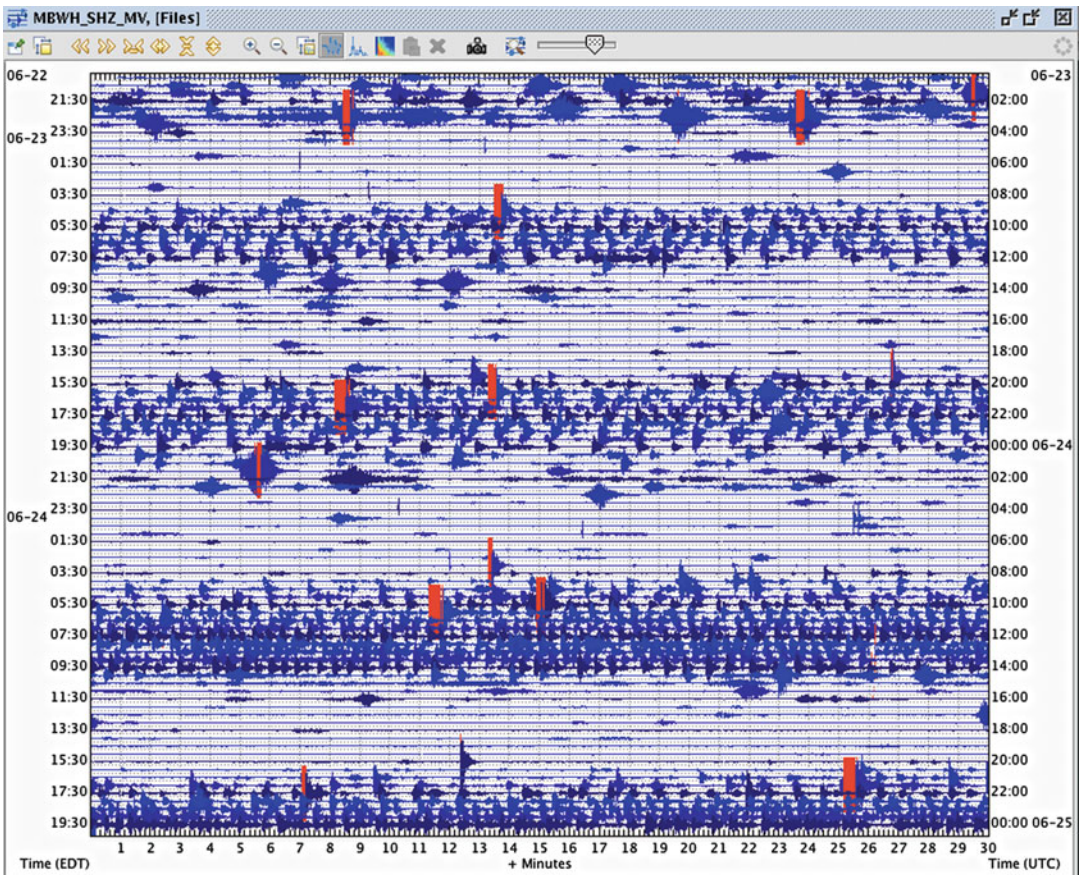
Seismic Monitoring of Volcanoes, Fig. 1

The Montserrat seismic network from April 2006 (Modified after Lockett et al. (2007)). Telemetry paths are shown as *lines*. The observatory is marked with MVO. Repeater sites are marked with a *gray circle*. Short-period stations are marked with a *triangle*. Broadband stations are marked with a *star*, except for the borehole broadband, MBLY, marked with a *square*. Not shown are the four 200-m borehole short-period stations, operated by the CALIPSO consortium since 2002



Helicorders were ubiquitous until the 1990s and are still used at many observatories today, but they have numerous drawbacks. They require considerable maintenance and provide limited dynamic range. Adjacent traces often overlap, making them hard to read. The data are not amenable to other forms of analysis, and the sheets require storage space. So there has long been a desire to replace helicorders with a software equivalent. Seismic Waveform Analysis and Real-time Monitor (SWARM) ([http://volcanoes.](http://volcanoes.usgs.gov/software/swarm/index.php)

[usgs.gov/software/swarm/index.php](http://volcanoes.usgs.gov/software/swarm/index.php)) is an excellent solution that allows data to be plotted dynamically: The user can select the time range and the scale; the data can be filtered, and short segments of data can be highlighted and replotted as spectra or spectrograms. Many observatories have now replaced large collections of helicorders with multiple screens showing SWARM displays from different seismic channels. Figure 2 shows 48 h of seismicity at the Soufrière Hills Volcano in June 1997.



Seismic Monitoring of Volcanoes, Fig. 2 Screenshot of the software Swarm, showing a digital helicorder plot from station MBWH channel SHZ for 48 h from June 23, 1997. Remarkable cyclic hybrid swarms are visible.

In June 25, a moderate dome collapse sent pyroclastic flows the northern flanks of the Soufrière Hills Volcano, claiming 19 lives

RSAM Plots

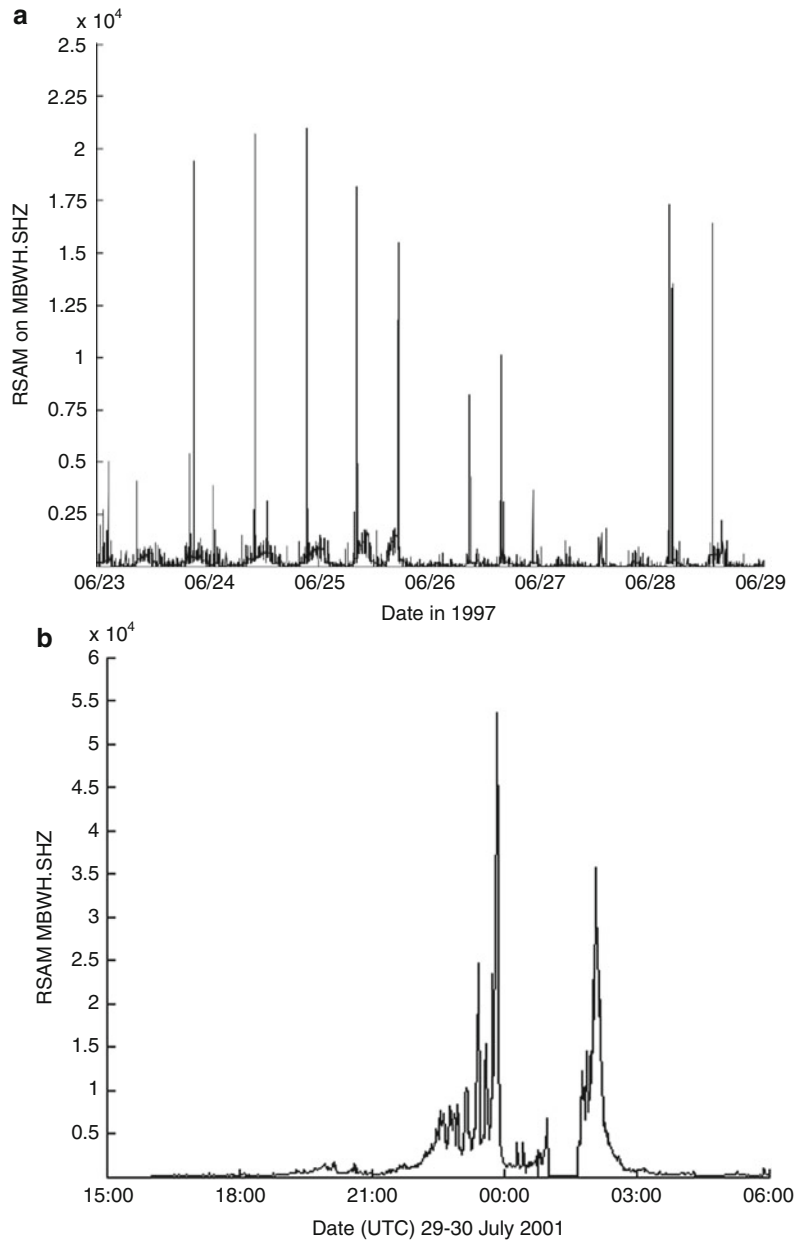
Continuous data are often downsampled to one sample per minute, which makes it easier to identify long-term trends in continuous seismic amplitude. The Real-time Seismic Amplitude Measurement system (Endo and Murray 1991) took the average amplitude of the seismic signal in each 1-min time window, and recorded these data into a file. RSAM data do not discriminate between different types of seismicity.

RSAM data show cyclic seismicity clearly (Fig. 3a). Changes in earthquake activity associated with dome-building episodes (Fig. 3b),

weather, and instrumental difficulties are recognized as distinct patterns in RSAM datasets. RSAM data for dome-building episodes gradually develop into exponential increases that terminate just before the time of magma extrusion. Volcanic earthquakes and rockfall show up as isolated spikes on RSAM plots for seismic stations close to the edifice, but seldom for more distant stations. Weather-related noise shows up as low-level, long-term disturbances on all seismic stations, regardless of distance from the volcano. The RSAM system proved valuable in providing up-to-date information on seismic activity for three Mount St. Helens eruptive

Seismic Monitoring of Volcanoes, Fig. 3 (a)

Plot of RSAM data with one sample per minute for station MBWH channel SHZ from June 22 to 29, 1997. The cyclic hybrid swarms mentioned in Fig. 2 are shown clearly. The spikes represent discrete events, e.g., pyroclastic flows. (b) RSAM plot showing the major dome collapse of the Soufrière Hills Volcano (Montserrat) which occurred on July 29, 2001



episodes from 1985 to 1986 and in numerous eruptions at other volcanoes since. Exponential increases in RSAM data commonly precede explosive eruptions. Inspired by RSAM, many other parameters have been computed on 1-min (or 10-min) timescale. For example, RSEM provides a relative measurement of energy on a 1-min (or 10-min) timescale.

Computing the median (rather than mean) amplitude in each time window provides a measurement of tremor amplitude less biased by events or spikes in the data - but is no longer strictly 'RSAM' data. Corrected for the instrument response and geometrical spreading, and then integrating the data, produces reduced displacement. Since 1996 the Alaska Volcano

Observatory has recorded an instrument corrected seismic amplitude and the peak frequency for each vertical-component seismic channel, on a 10-min timescale.

Spectrograms

Changes in tremor spectra may indicate different flow regimes or changes in source parameters such as geometry, sound speed, or ascent rate (Thompson et al. 2002). Harmonic tremor and gliding spectral lines are frequently followed by eruptions. Spectral monitoring can help differentiate between VT and low-frequency earthquakes, or tremor; wind noise; and electronic noise.

The Alaska Volcano Observatory (AVO) began experimenting with high-resolution spectrograms in 1996. For up to eight stations, a spectrogram is plotted underneath a normalized seismic trace. The time resolution is about 10 s and the frequency resolution 0.1 Hz. Each image file (Fig. 4) displays a 10-min seismic trace and spectrogram for up to eight seismic-data channels. These are ordered in increasing distance from the volcano. The data are instrument corrected and color coded, so that ground motions at one station can be easily compared with those at another station, even at a different volcano.

These web-based spectrograms have been a core AVO monitoring tool since 1998, and a seismologist reviews them twice a day. A convenient web interface (<http://www.aeic.alaska.edu/spectrograms/mosaicMaker.php>) shows mosaics of 10-min spectrogram image files (Fig. 5) and enables 12 h of data from more than 20 volcanoes to be reviewed in just 30–45 min. Similar web-based spectrograms are used for monitoring volcanic seismicity in the Cascades and Hawaii.

Tremor Alarms

Tremor is common precursor to eruptions, and strong tremor frequently accompanies eruptions. Observatory seismologists therefore need to be aware when tremor is being recorded. The earliest widely used volcano-seismic alarm system was part of the RSAM system (Endo and Murray

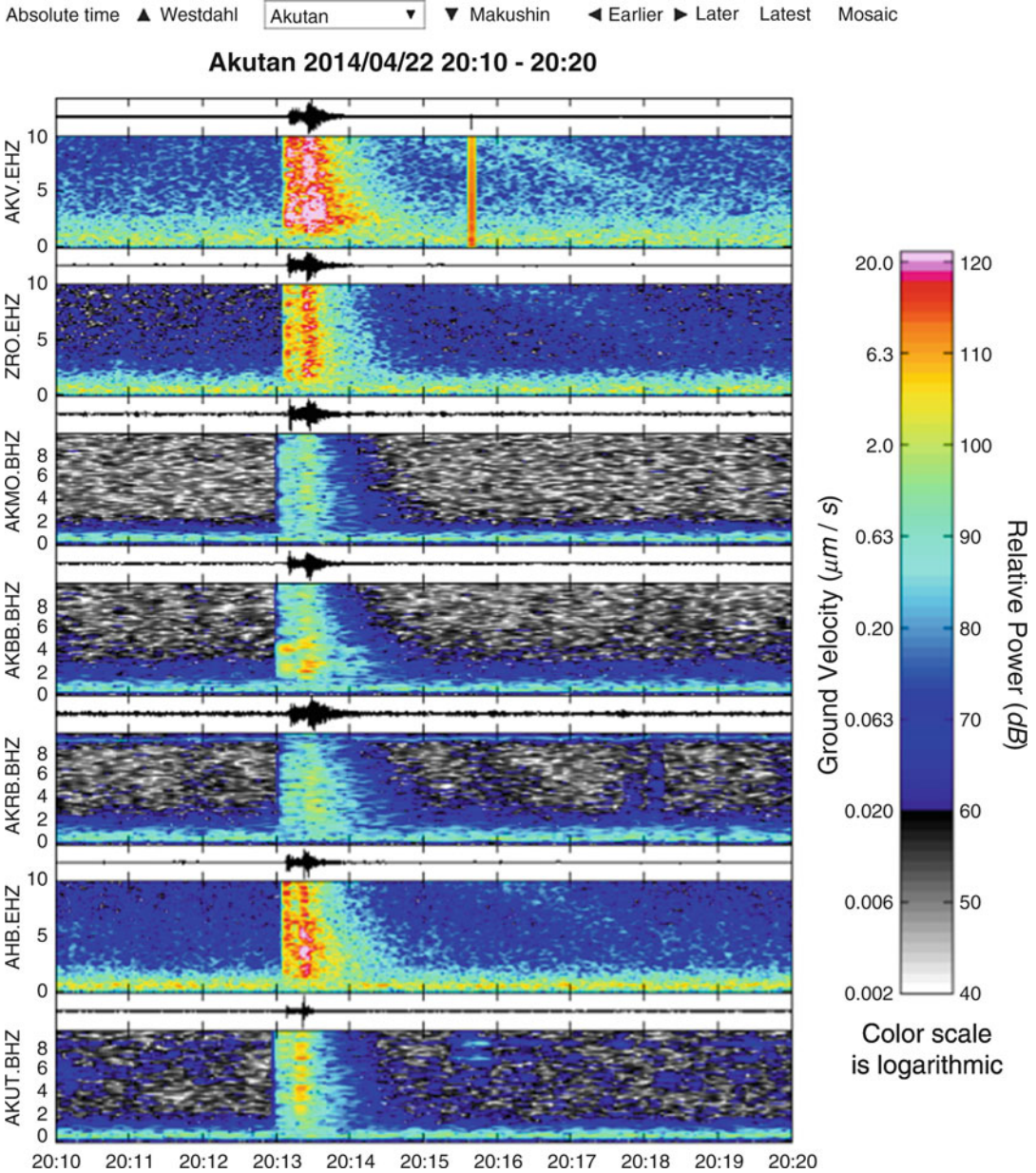
1991), and this is now incorporated in Earthworm (discussed below). The seismologist can choose which stations to monitor and define an amplitude threshold and a duration for each station. Both must be exceeded to trigger a station. The seismologist also must define the number of stations that must trigger simultaneously to declare an alarm. There is also a mechanism which reduces the false alarms due to regional or teleseismic earthquakes: A far away station can be chosen which prevents an alarm being declared when it exceeds an amplitude threshold. The tremor alarm is typically configured for a moderate amplitude and a duration of several minutes (the RSAM system also included an event alarm, which would typically be configured for high amplitude and duration of a few to tens of seconds).

The RSAM tremor alarm system was used to monitor the Soufrière Hills Volcano (Montserrat) from 1999 to 2003. Figure 6 shows the number of alarms issued over this period.

Events

Real-Time Seismic Event Catalogs

During a volcanic crisis, it is vital to have near-real-time information about the rates, sizes, and locations of seismic events, and for this a real-time event catalog is needed. Providing there are at least four stations, event detection software automates the process of capturing anomalous signals in large volumes of continuous data. Each channel of seismic data is detrended, filtered, and rectified, and a running short-term average (typically 1 to a few seconds) and long-term average (typically a few tens of seconds) are computed. Where the STA (short-term average) to LTA (long-term average) ratio exceeds a threshold, a candidate P or S arrival is declared. Association is the process whereby candidate arrivals' "picks" are grouped together based on similar arrival times and geographic location to declare an event. Location techniques typically use a 3-D grid search to minimize the difference between measured and theoretical travel times, given a particular velocity model. For volcanic earthquakes, it is most common to compute a



Seismic Monitoring of Volcanoes, Fig. 4 Seismic traces and corresponding spectrograms for the vertical components of seven seismic stations near Akutan Volcano, corresponding to 10 min of data on April 22, 2014. A regional earthquake is visible clearly around 20:13

UTC. The hottest colors shown (pink) suggest an S-wave ground velocity of 0.02 mm/s. Near-real-time spectrogram plots like this have been linked to the AVO internal website since 1996. The menu at the top of the screen allows easy navigation to other volcanoes or time periods

duration magnitude due to the prevalence of clipped signals from short-period analog telemetry. For on-scale recordings, local magnitude is often also computed. To visualize catalogs, it is

helpful to view event rate and energy rate plots (Fig. 7) and hypocenter plots (Fig. 8).

There are some caveats with event catalogs. Discrete events are only one aspect of volcanic



Seismic Monitoring of Volcanoes, Fig. 5 A spectrogram mosaic, part of the same application shown in Fig. 4. Three hours of data are shown. Regional earthquakes are

visible at 20:13 (Fig. 4) and 22:15. A calibration signal occurs on the top station (AKV) at 22:03. The apparent data dropouts are due to data latency

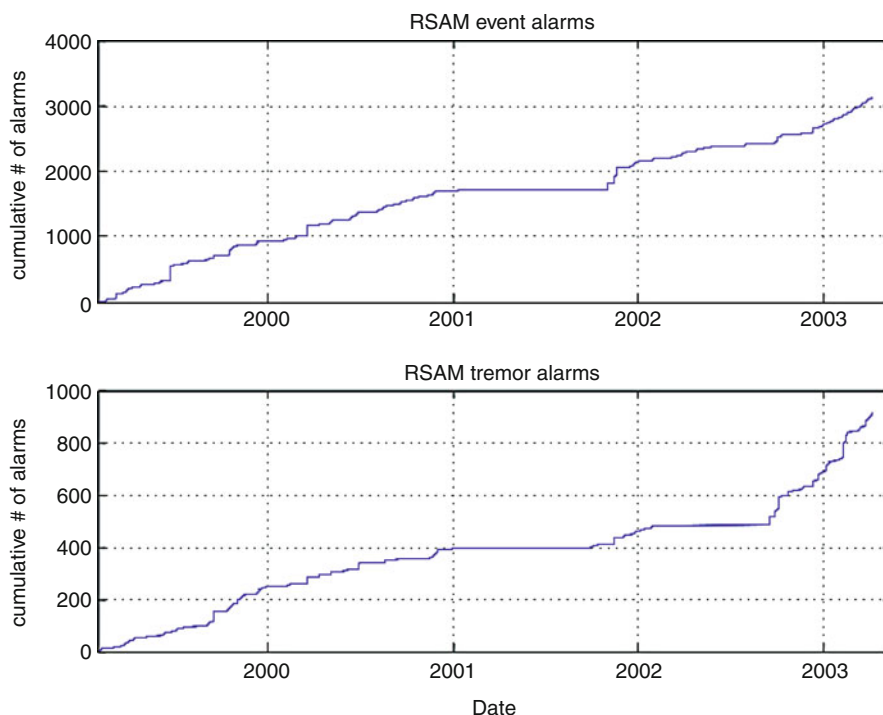
seismicity. Signals such as precursory earthquake swarms and tremor episodes, or those generated by explosive eruptions, dome collapses, pyroclastic density currents, and lahars, are of the greatest interest from a hazard perspective. Such signals generally evade standard earthquake-detection schemes based on comparing the short-term and long-term signal averages and are not systematically cataloged. The vast majority of volcano-seismic signals lack identifiable phases and so cannot be located with differential travel-time techniques. Earthquake catalogs therefore present only a narrow view of volcanic seismicity. Many volcano-seismic signals are emergent or long in duration, so STA/LTA detectors break down. Events may be masked by high background signals (e.g., tremor, wind, or electronic noise). Station outages also increase the

magnitude at which earthquakes become detectable. Times when an event catalog suggests there is little seismicity may actually be times of high seismicity. Real-time catalogs also suffer from poorly resolved locations and magnitudes – also a review by an analyst later corrects for these.

Swarm Alarms

Earthquake swarms, like tremor, are a common precursor to volcanic eruptions. A real-time event catalog can be used as the basis for detecting earthquake swarms.

Okmok Volcano (Alaska) erupted ash to a height of 15 km on July 12, 2008, after less than 5 hours of precursory seismicity which included an earthquake swarm. Concerned by this, the Alaska Volcano Observatory developed a swarm alarm system. The system identified the



Seismic Monitoring of Volcanoes, Fig. 6 RSAM event alarms (*top panel*) and tremor alarms (*bottom panel*) issued at Montserrat Volcano Observatory between February 1999 and May 2003. In total there were about 4,000 alarms, an average of almost three alarms per day.

The alarm thresholds configured for each station were raised throughout 2000 and 2001 as activity increased. Event alarms typically corresponded to pyroclastic flows and regional earthquakes, and tremor alarms to dome collapses (i.e., a series of pyroclastic flows)

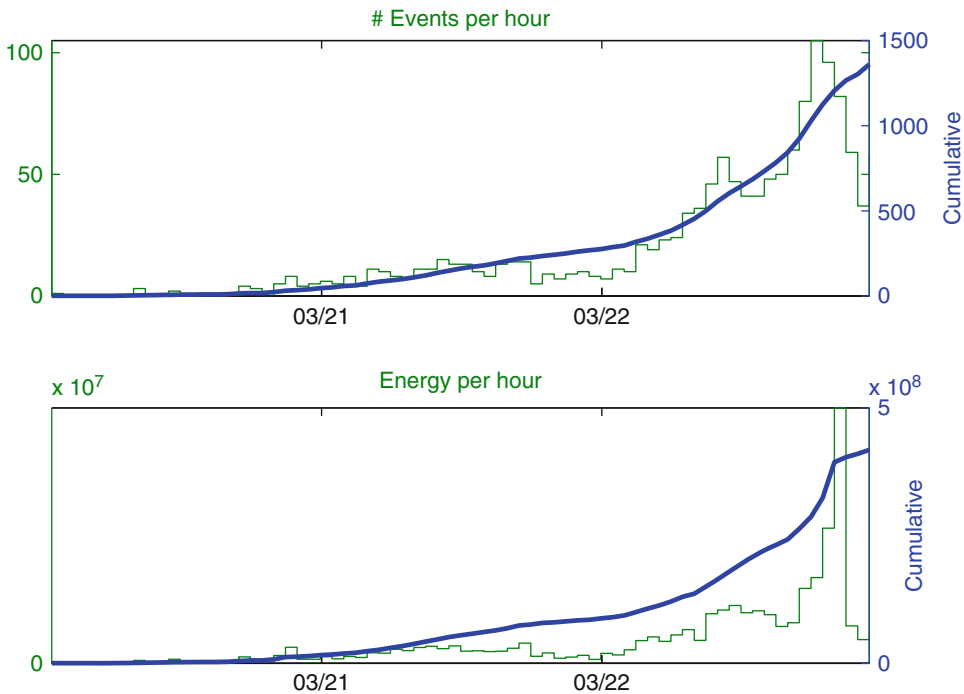
start and end of five swarms that occurred between February and April 2009, as well as significant escalations in the rate of earthquakes and energy release during those swarms (Thompson and West 2010).

Software

Earthworm (<http://folkworm.ceri.memphis.edu/ew-doc/>), is the most widely used system for earthquake and volcano-seismic monitoring today. It is an open-source, data acquisition, and earthquake detection framework developed by the US Geological Survey. It comprises of modules that communicate via messages on shared memory rings. It is used in the National Earthquake Information Center, tsunami warning centers, and regional seismic networks. Earthworm is in widespread use at volcano observatories

today, in conjunction with other tools that expand its real-time monitoring capabilities. Two reasons for this are that all US volcano observatories use them, and the USGS Volcano Disaster Assistance Program deployed them in many countries at the request of foreign governments.

Earthworm is not the only widely used seismic monitoring systems: Alternatives are SeisComP3 (<https://www.seiscomp3.org/wiki/doc>) and Antelope (<http://www.brrt.com/software.html>). Antelope is an excellent framework for developing new monitoring and research tools, but requires a commercial license except for US higher education institutions. SeisComP3 provides a rich set of graphical user interfaces for global and regional seismic monitoring and is particularly suited for tsunami early warning. However,



Seismic Monitoring of Volcanoes, Fig. 7 Two common ways of examining trends in detected events are (upper panel) event rate (also called ‘counts’- the number of detected events per unit time) plots, and (lower panel)

energy release rate plots (computed from magnitude data). The data shown are from Redoubt volcano, Alaska, from 2009/03/20 to 2009/03/23 and are produced using the GISMO toolbox. Energy units are arbitrary

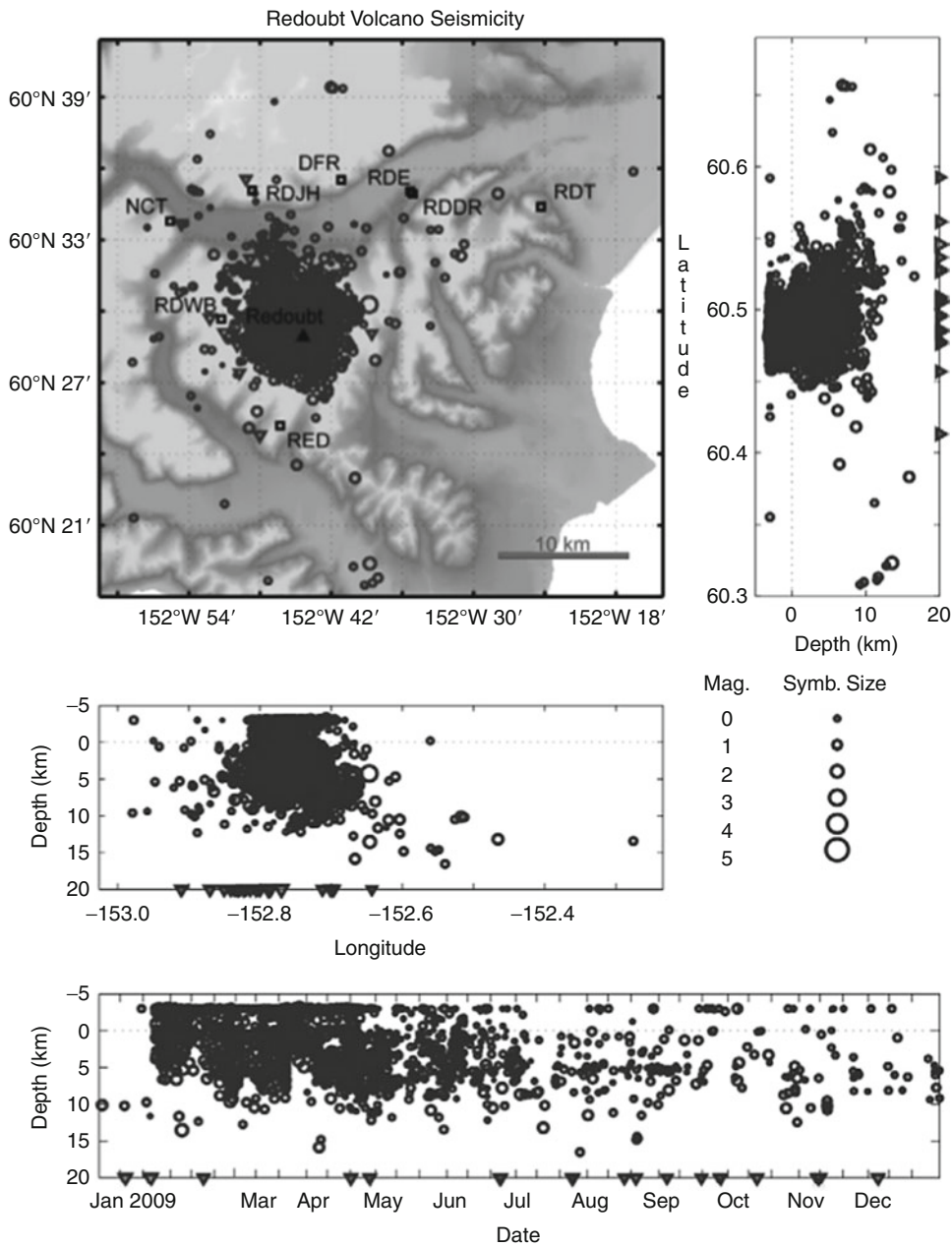
neither currently has adaptations for volcano-seismic monitoring.

Earthworm, SeisComP3, and Antelope are capable of generating real-time earthquake catalogs within a few minutes of event occurrence, and they work well for regional earthquakes and volcano-tectonic earthquakes. However, other volcano-seismic signals lack clear P- and S-waves required to compute even an approximate location. Fortunately, Earthworm includes another mode of event detection called “subnet triggering,” which determines only an approximate event time, and does not attempt to locate an event. This approach is employed by many volcano observatories. Custom code can be added to monitor these events and trigger other calculations. For example, at the Montserrat Volcano Observatory from 2000 to 2003, real-time measurements of the amplitude, energy, and frequency content of new events were made, and

used to regenerate plots on a private web site used by observatory scientists.

Earthworm has the ability to import data from a wide variety of field instrumentation and data servers and includes a variety of earthquake detection, location, and magnitude algorithms. There are modules which compute RSAM data and create daily helicorder and spectrogram plots. Tremor and swarm alarm systems are now included in Earthworm. Other modules can harvest messages about detected events, arrivals, locations, and magnitudes to create a real-time event catalog, serve up plots of event rates and Google Maps of epicenters, and serve data in QuakeML format.

SWARM can either draw data from a SeedLink server, an Earthworm wave server, or an FDSN (International Federation of Digital Seismograph Networks) web service. Another option is a Winston wave server, which emulates



Seismic Monitoring of Volcanoes, Fig. 8 A common way of presenting 3-D hypocenter data is as a set of 2-D slices through the data. In this plot, generated by VolPlot, a program used for many years at the Alaska Volcano Observatory, earthquake hypocenters from Redoubt Volcano in 2009 are shown. The main panel shows epicenters

in map view, and adjacent panels show depth versus latitude and depth versus longitude. These help scientists identify spatial relationships in hypocenter data. At the bottom, depth versus time is shown, which helps scientists recognize if earthquakes are getting closer to the surface. The size of the symbols indicates magnitude

an Earthworm wave server, but stores data in a MySQL database, providing rapid access and a deep data archive. Winston also provides a web interface that allows plotting of helicorder images and RSAM data (RSAM data are precomputed on import to Winston).

Data Analysis

Some analyses are not performed in real time, because it cannot be fully automated or just because the tools to do it well have not yet been developed; but they nevertheless provide timely information that can affect the way scientists interpret activity.

Analyst-Reviewed Event Catalogs

A real-time event catalog is valuable, but it cannot approach the quality of an analyst-reviewed catalog. A seismic analyst will periodically review the detected events, classify them, and delete false events. They might also pick P and S phases, run software to locate events, and compute magnitudes. Such a catalog is essential for research, but can be difficult to produce in a timely manner as volcanic unrest intensifies and decisions need to be made on timescales of minutes and hours, rather than days. During earthquake swarms (and major aftershock sequences), human analysts may be overwhelmed, increasing latency further just when it matters most.

While Antelope and SeisComP3 include an analyst review capability, Earthworm does not. So it is commonly paired with Seisan (<http://seis.geus.net/software/seisan/seisan.pdf>), which is an excellent free software package for processing and editing event catalogs. Seisan can estimate fault-plane solutions, moment tensors, and b-values too. Seisan includes some adaptations for volcano-seismic analysis: It supports event classes such as VT, LP, hybrid, and rockfall and can generate event-rate plots. Event waveforms are typically examined as spectra as well as time series. However, automated event classification is highly desirable because different real-time processing schemes could then be applied to

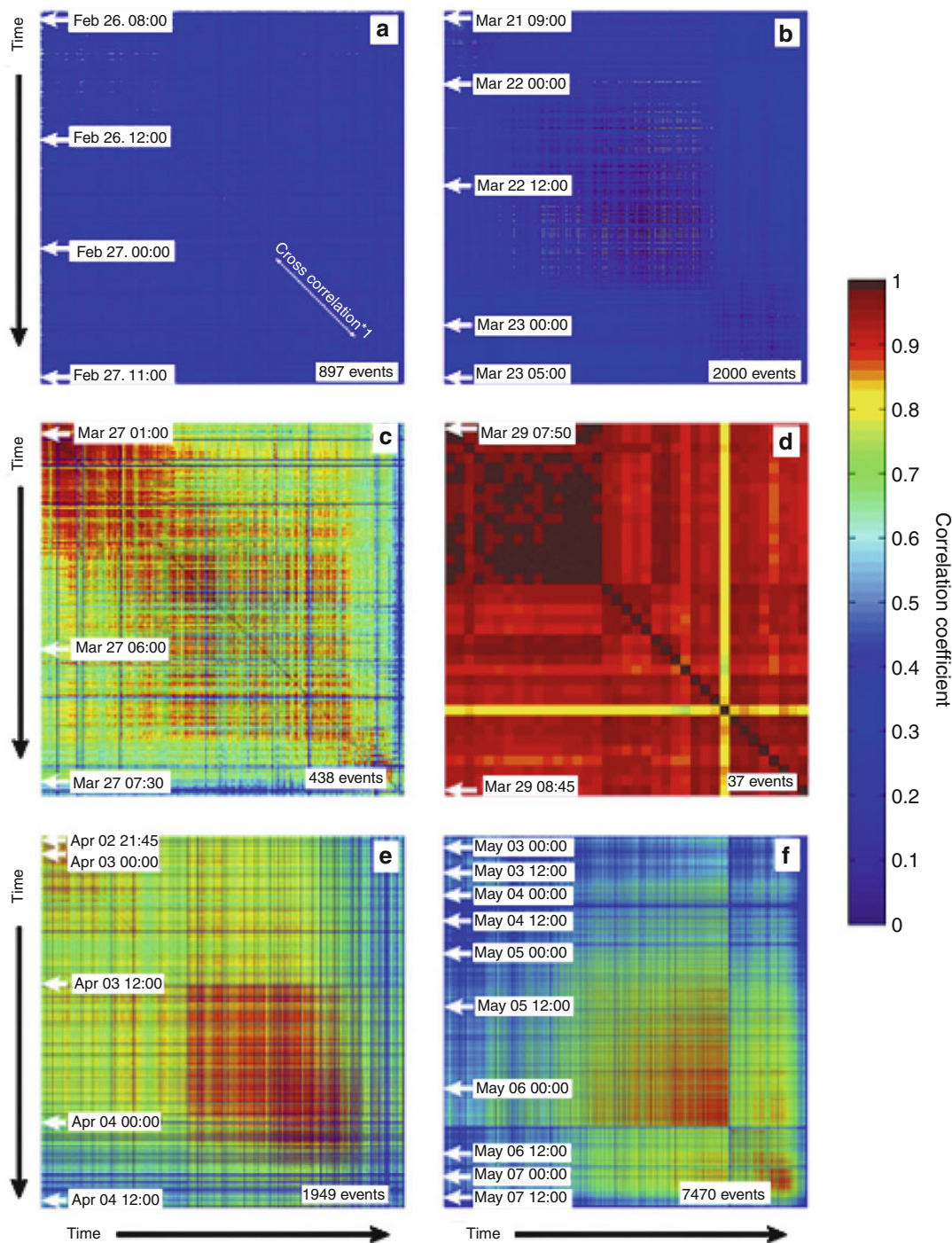
different event classes. For example, at the Montserrat Volcano Observatory from 2000 to 2003, rockfall signals were directed into an automated rockfall location system.

Automated techniques for classifying events have been tried at some observatories, but remain rare. Many researchers have had success with frequency-based analysis, artificial neural networks, and hidden Markov chains, but typically for only certain types of events recorded at a particular volcano on a particular station. Neither Earthworm, Antelope, or SeisComP3 currently include an auto-classification module.

Swarm Analysis

While event detection software captures many (perhaps most) volcanic earthquake signals, additional events can be found by using a match filter. This technique is particularly useful for tracking repeating, low-frequency earthquakes, since the emergent onsets of this type of event make them notoriously difficult to trigger using an STA/LTA detector. A known event waveform (or part of it) is cross-correlated against continuous data to find additional events in the same “family.” Each family represents a source which is repeatedly activated in a volume that is small compared to the wavelength of the correlated signal. A number of volcanic processes are thought to be able to generate repeating earthquakes, including repeated resonances of fluid-filled cracks, a propagating crack tip that is driven by intruding magma, or even repeated shear failure within a body of magma. Additional analyses of the earthquake locations and the waveform frequency spectra are needed in order to distinguish between these various models. Double-difference relocation can be used to image the spatial evolution of the swarm more clearly.

A related technique is to cross-correlate waveforms from all events in a catalog against each other, to discover all event families. The MATLAB toolbox “GISMO” includes tools to automate this process. During unrest at Redoubt Volcano in 2009, this was used in near real time to monitor the evolution of swarms (Buurman et al. 2012) (Fig. 9).



Seismic Monitoring of Volcanoes, Fig. 9 Cross-correlation plots for six swarms during the crisis at Redoubt Volcano in 2009 (reproduced from Buurman et al. (2012)). Each pixel represents of pair of waveforms that have been cross-correlated. Color represents the maximum cross-correlation coefficient. Time progresses from

top to bottom and left to right in each panel. The top-left to lower-right diagonal represents the autocorrelation of each waveform. The swarms began on: (a) February 26, (b) March 21, (c) March 27, (d) March 29, (e) April 2, and (f) May 2

Locating Phaseless Seismic Signals

The tremor alarm systems mentioned previously are very simple and they lack the ability to locate tremor. Some observatories have exploited seismic data to track the location of tremor and map debris-flow trajectories, which could help mitigate hazards.

Phaseless signals, once identified (via an STA/LTA algorithm or visual inspection), can be approximately located by exploiting the manner in which seismic wave amplitudes decay with increasing distance from the source, due to geometrical spreading and attenuation (Jolly et al. 2002). Alternatively, they can be located by cross-correlating data to determine travel-time differences and then using traditional differential travel-time techniques to locate the source (assuming the wave type is known). The Waveform Envelope Clustering and Correlation (WECC) system (Wech and Creager 2008) reverses this approach. It continuously locates envelopes of continuous waveform data, and if the location errors are small, it declares an event. Designed to locate the episodic tremor associated with the slow slip, it is now being applied to volcanoes.

Detecting Lahars

Lahar signals may appear as high-frequency tremor signals on a volcano-seismic network, but may be difficult to detect and locate without designing a network for lahar monitoring. High-frequency surface waves attenuate rapidly with increasing distance, so the amplitude of high-frequency signals can be indicative of how close a flow is to a seismometer. More dilute flows are also richer in higher frequencies. An existing volcano-seismic network can be adapted for lahar monitoring by adding stations close to the flow channel at various points, enabling lahar signals to be more easily resolved and crudely located using the times of the peak amplitudes on each station, and also the amplitude distribution across the seismic network. Trip wires can be positioned across the valley at different points and at different heights, and the arrival time and location are

known precisely when the wire is cut. Disadvantages are that trip wires only work once and a dilute flow may leave the wire intact. Sophisticated lahar-monitoring systems have been installed at many volcanoes including Rainier, Merapi, and Nevado del Ruiz.

Imaging Spatial Changes in Volcanic Seismicity

There are many techniques for imaging magma storage regions. A volume devoid of volcano-tectonic earthquakes may indicate a large magma storage body, incapable of shear failure. Seismic tomography can reveal volumes with low P- or S-wave propagation speeds. B-value mapping (see Sanchez et al., “► [Frequency-Magnitude Distribution of Seismicity in Volcanic Regions](#),” this volume) can reveal volumes incapable of sustaining large failures, which are often inferred as being associated with magma. In the region around Uturuncu Volcano (Bolivia), the presence of a sill at around 15–20-km depth has been inferred by an S-wave shadow zone. Roman et al. (2006) found that changes in the orientation of fault-plane solutions could be used (retrospectively) to differentiate between volcano-tectonic earthquakes related to the injection of new magma and post-eruptive relaxation.

Ambient noise tomography is another technique, which can be used to detect velocity changes. In active source seismology, an impulsive source is generated at one location and recorded at another. The resulting seismogram is the Green’s function of the path between the sites. If this were done repeatedly, seismic velocity changes along the travel path could be detected. At a volcano, a drop in velocity could be indicative of increased heat flow, faulting, fluid injection, or expansion of the volcanic edifice. MSNoise (<http://www.msnoise.org/>) is a package which has been developed to automate the computation of Green’s functions between all station pairs on a daily basis, allowing velocity changes as small as 0.1 % to be identified. At Piton de la Fournaise Volcano, the locations of dike injections and eruptive fissures have been inferred.

Detecting Explosions

Cleveland Volcano (Alaska) has been erupting frequently in recent years, generating ash clouds that threaten aviation travelling between North America and Asia, and, without a functioning seismic station within a few tens of kilometers, many of these eruptions were previously missed or only recognized later in satellite images. De Angelis et al. (2012) describe a system which exploits data from a seismic network on Okmok Volcano, operated by the Alaska Volcano Observatory, to detect explosions at Cleveland Volcano (about 120 km away). An STA/LTA detector creates candidate arrivals. For those that fall within a 2-min sliding window, the differential travel times are computed by cross-correlating envelopes of the seismograms and then inverted for apparent slowness. If this is consistent with a sound speed of 340 ± 30 m/s in the direction of Cleveland, an alarm is sent by email. An alternative to the Cleveland explosion alarm mentioned above looks for a coherent signal across an infrasound array instead. The advantage is that infrasound sensors record explosion signals less ambiguously, and the technique could be used to detect explosions in other regions where installation of a local seismic network is unrealistic.

Technical Challenges

One of the major challenges for observatories and seismic networks is monitoring the state of health of a seismic network. There is only one chance to gather data, and in a crisis, data not collected in real-time have little value. Seismic and repeater stations need to be engineered to withstand weather variations which may include heavy rain, snow and ice accumulation, strong winds, and months of little solar energy. Spare equipment needs to be available so that stations can be fixed quickly when a seismic station component breaks. Vandalization by humans, damage by animals, and theft are other considerations. During unrest, ash build-up on solar panels can quickly cause power loss at a seismic station.

Problems will usually be apparent with tools volcano seismologists use every day. RSAM plots may reveal diurnal cycles in the data which may indicate undercharging of batteries or anthropogenic noise. Spectrograms and seismograms may show a flat signal (power loss), a one-sided signal (stuck seismometer), large offset (seismometer not level), white noise (loss of seismic signal), or spikes and dropouts (interference). These can cause havoc with automated detection and alarm systems, leading to corrupted event catalogs and false alarms, so one strategy is to identify and eliminate affected stations from automated processing schemes. SeisNetWatch (<http://www.isti.com/products/seisnetwatch/>) is used by many seismic network operators to monitor all ip-addressible nodes in a network, including field digitizers, to monitor data latency and other state-of-health parameters.

The volume of data flowing into an observatory can be overwhelming. A single channel of 24-bit, three-component data recorded at 100 Hz requires 620 MB of storage per day (uncompressed). Twenty channels of data, a typical amount for a small volcano observatory, are about 1.5 TB per year. Although this would fit on a single hard drive today, it is about 100 times the capacity of a hard drive available 20 years ago (the storage capacity of hard drives increases by a factor of about 10 every decade). Continuous digital volcano-seismic data before the mid 1990s are rare but, remarkably, the Pacific Northwest Seismic Network recently recovered continuous data saved to tape during the 1980 eruption of St. Helens Volcano.

To improve the reliability of data acquisition and processing systems, several measures can be taken. To protect against power failures, all mission-critical computers can be connected to uninterruptible power supplies and configured to automatically reboot and restart critical processes once power is restored. No one wants to arrive at an observatory on a Monday morning and find that the alarm computer suffered a power outage on Friday night and has not been running since, especially if a significant event occurred. Another measure is to run all mission-critical computers in parallel, providing automatic failover or at

least have pre-configured spare computers ready to plug in whenever a primary computer failed. All potential points of failure can be monitored by a diagnostic alarm system. Finally, a daily checklist helps catch any other problems. All of these measures were taken at the Montserrat Volcano Observatory in 2000, leading to an improvement in the data capture rate and public safety.

Once captured, data need to be managed effectively to preserve them for further analysis and research. Given the capacity of modern hard drives, it is now possible to keep many years of data “online.” Copying data to an off-site location is also becoming more common. Many observatories now transmit real-time data to the IRIS Data Management Center, which not only serves as a backup, but outsources the resources needed to disseminate the data more easily to the research community.

Summary

Volcano-seismologists forecast volcanic activity by analyzing the rates, energy release and spatial distribution of different types of characteristic seismic signals recorded in the vicinity of volcanoes. The most commonly identified signals are VT, LP or hybrid earthquakes. Explosive and effusive eruptions can cause rockfalls and pyroclastic flows, which also generate characteristic seismic signals. Volcanic earthquakes frequently occur in swarms. Swarms and tremor are both common precursors to escalations in volcanic activity. While there is still much that is not understood – the origin of LP earthquakes for example - no other technique provides such detailed information about the internal state of a volcano, or about debris flows as they occur, or provides such a detailed chronology of an eruption.

Volcano-seismic monitoring provides high-sample-rate data, 24 hours a day. Short-period seismometers and analog telemetry are gradually being phased out and replaced with modern broadband seismometers and field digitizers. Two-way telemetry allows observatory staff to diagnose and fix problems sometimes without

an expensive field visit or lengthy data outage. Free, community supported software is available for data acquisition, event detection and location (e.g. Earthworm), real-time data visualization (e.g. SWARM), event processing and catalog production (e.g. Seisan), instead of each observatory reinventing the wheel. Archiving data - including continuous waveform data, derived data such as RSAM, event catalogs and station metadata – is a crucial task that has become much easier thanks to the expanding capacity of hard drives, faster internet, and data management centers such as the IRIS DMC.

Real-time monitoring may rely on an operations room which is manned 24 h a day. Or it may rely on automated alarm systems that alert observatory staff to escalations in seismicity (and data outages) outside of normal office hours. Research and development is often driven by the need for better real-time monitoring tools – automated classification of volcano-seismic signals, for example. Ambient noise tomography and detection of multiplets are examples of techniques that might become routine at volcano observatories as computational power continues to increase. As automated monitoring becomes increasingly sophisticated, observatory seismologists will need to spend less time on troubleshooting and development and more time analyzing and interpreting volcanic-seismicity, and collaborating on research.

Cross-References

- ▶ [Earthquake Location](#)
- ▶ [Earthquake Swarms](#)
- ▶ [Frequency-Magnitude Distribution of Seismicity in Volcanic regions](#)
- ▶ [Principles of Broadband Seismometry](#)
- ▶ [Rockfall Seismicity Accompanying Dome-Building Eruptions](#)
- ▶ [Seismic Anisotropy in Volcanic Regions](#)
- ▶ [Seismic Event Detection](#)
- ▶ [Seismic Tomography of Volcanoes](#)
- ▶ [Very-Long-Period Seismicity at Active Volcanoes: Source Mechanisms](#)

- ▶ Volcanic Eruptions, Real-Time Forecasting of
- ▶ Volcano-Tectonic Seismicity of Soufriere Hills Volcano, Montserrat

References

- Barclay J, Johnstone J, Matthews A (2006) Meteorological monitoring of an active volcano: implications for eruption prediction. *J Volcanol Geotherm Res* 150(4):339–358. doi:10.1016/j.jvolgeores.2005.07.020
- Benoit JP, McNutt SR (1996) Global volcanic earthquake swarm database and preliminary analysis of volcanic earthquake swarm duration. *Annali Di Geofisica* 39(2):221–229
- Buurman H, West ME, Thompson G (2012) The seismicity of the 2009 redoubt eruption. *J Volcanol Geotherm Res*. doi:10.1016/j.jvolgeores.2012.04.024
- Chouet B (2003) Volcano seismology. *Pure Appl Geophys* 160(3):739–788. doi:10.1007/PL00012556
- Chouet BA, Matoza RS (2013) A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. *J Volcanol Geotherm Res* 252:108–175. doi:10.1016/j.jvolgeores.2012.11.013
- De Angelis S, Fee D, Haney MM, Schneider DJ (2012) Detecting hidden volcanic explosions from Mt. Cleveland Volcano, Alaska with infrasound and ground-coupled airwaves. *Geophys Res Lett* 39:1–6. doi:10.1029/2012GL053635
- Endo ET, Murray TL (1991) Volcanology a volcano monitoring and prediction tool. *Bull Volcanol* 53:533–545
- Jolly AD, Thompson G, Norton G (2002) Locating pyroclastic flows on Soufriere Hills Volcano, Montserrat, West Indies, using amplitude signals from high dynamic range instruments. *J Volcanol Geotherm Res* 118(3–4):299–317. doi:10.1016/S0377-0273(02)00299-8
- Lahr JC, Chouet BA, Stephens CD, Power JA, Page RA (1994) Earthquake classification, location, and error analysis in a volcanic environment: implications for the magmatic system of the 1989–1990 eruptions at redoubt volcano, Alaska. *J Volcanol Geotherm Res* 62(1–4):137–151. doi:10.1016/0377-0273(94)90031-0
- Luckett R, Baptie B, Ottemöller L, Thompson G (2007) Seismic monitoring of the Soufriere Hills Volcano, Montserrat. *Seismol Res Lett* 78(2):192–200. doi:10.1785/gssrl.78.2.192
- McNutt SR (1996) Seismic monitoring of volcanoes: A review of the state-of-the-art and recent trends. In: Scarpa R, Tilling R (eds) *Monitoring and mitigation of volcano hazards*, Chapter 3, Springer-Verlag, Berlin, pp 99–146
- McNutt SR (2000) Volcanic seismicity. In: Houghton HB, McNutt SR, Rymer H, Stix J (eds) *Chapter 63 of encyclopedia of volcanoes*. Academic Press, San Diego CA, 1015–1033
- McNutt SR (2002) Volcano seismology. In: Lee WHK, Kanamori H, Jennings PC (eds), *Chapter 25 of international handbook of earthquake and engineering seismology*, IASPEI, Palo Alto, CA, 81A: 383–406
- McNutt SR (2005) Volcanic seismology. *Annu Rev Earth Planet Sci* 33(1):461–491. doi:10.1146/annurev.earth.33.092203.122459
- Moran SC, Freymueller JT, LaHusen RG, McGee KA, Poland MP, Power JA, White RA (2008) Instrumentation recommendations for volcano monitoring at U.S. Volcanoes under the national volcano early warning system: scientific investigations report 2008 – 5114, p 47. Retrieved from <http://pubs.usgs.gov/sir/2008/5114/>
- Powell TW, Neuberg JW (2003) Time dependent features in tremor spectra. *J Volcanol Geotherm Res* 128(1–3): 177–185. doi:10.1016/S0377-0273(03)00253-1
- Roman DC, Neuberg J, Luckett RR (2006) Assessing the likelihood of volcanic eruption through analysis of volcanotectonic earthquake fault–plane solutions. *Earth Planet Sci Lett* 248(1–2):244–252. doi:10.1016/j.epsl.2006.05.029
- Thompson G, McNutt SR, Tytgat G (2002) Three distinct regimes of volcanic tremor associated with the eruption of Shishaldin Volcano, Alaska 1999. *Bull Volcanol* 64(8):535–547. doi:10.1007/s00445-002-0228-z
- Thompson G, West ME (2010) Real-time detection of earthquake swarms at redoubt Volcano, 2009. *Seismol Res Lett* 81(3):505–513. doi:10.1785/gssrl.81.3.505
- Wassermann J (2012) Volcano seismology. In: Peter B (ed) *IASPEI new manual of seismological observatory practice 2 (NMSOP-2)*, second. Potsdam : Deutsches GeoForschungsZentrum GFZ, Potsdam, pp 1–77. doi:10.2312/GFZ.NMSOP-2_ch13
- Wech AG, Creager KC (2008) Automated detection and location of Cascadia tremor. *Geophys Res Lett* 35(20), L20302. doi:10.1029/2008GL035458

Seismic Network and Data Quality

Peter H. Voss¹ and Lars Ottemöller²

¹Geological Survey of Denmark and Greenland – GEUS, Copenhagen K, Denmark

²Department of Earth Science, University of Bergen, Bergen, Norway

Synonyms

Earthquake monitoring; Seismometer sensitivity; Instrument response; Ambient noise; Data latency

Introduction

Measuring of ground motion provides the most essential observations made from earthquakes and the basis for research within seismology and earthquake engineering. The measurements are done by seismic stations that are equipped with sophisticated instruments and are placed at specific selected locations. The good quality of these data is obviously important but can only be achieved if many different factors are considered. Often, the approach is adjusted to the scientific or monitoring needs and the available funding. This section deals with systems of seismic stations, so-called seismic networks and data quality-related issues.

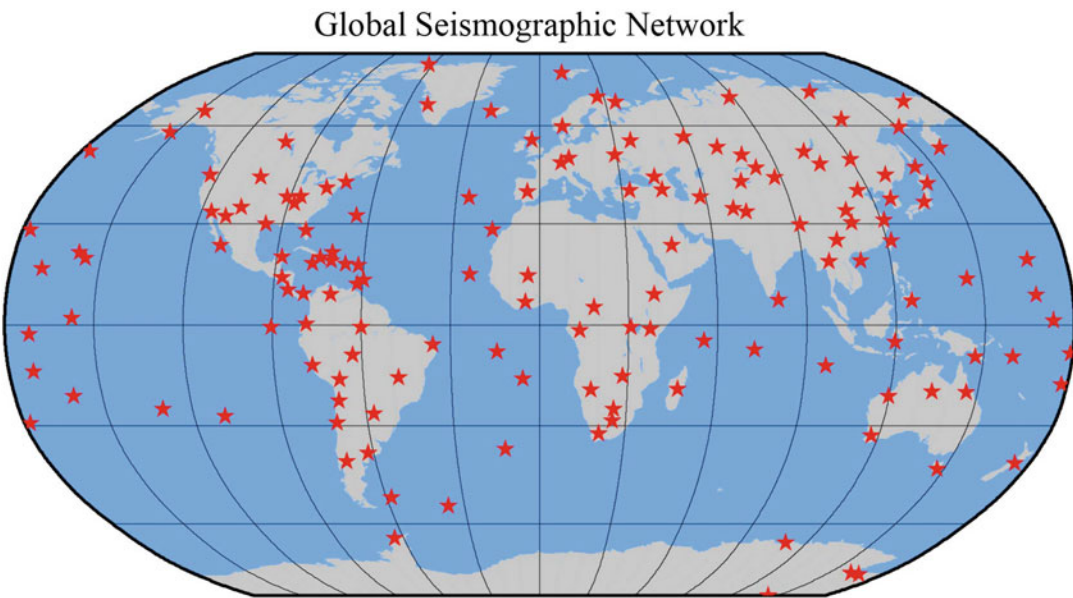
Seismic Networks

Seismic networks are systems of individual seismic stations with the purpose to monitor ground motion. The size and configurations of seismic networks depend on the focus area and type of signals that are to be recorded. The data provided by seismic networks (seismograms; see Fig. 3) can be used for many different applications from

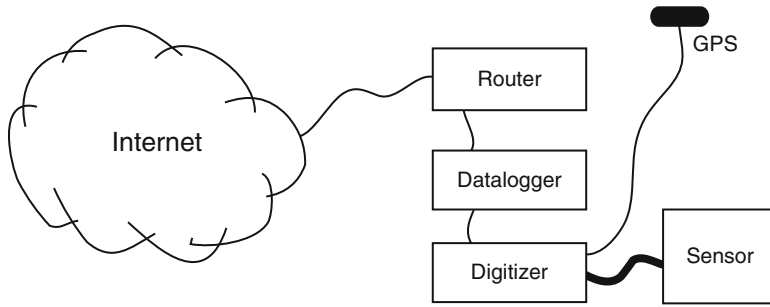
earthquake monitoring services to research of the Earth interior. Examples of seismic networks are global seismograph networks (Fig. 1) that are operated to primarily detect large earthquakes and other events globally, regional or national seismic networks to detect events within a country, microseismic arrays to detect induced earthquakes at a water reservoir or hydrocarbon field, and dense monitoring at volcanoes to measure the ongoing activity. Individual seismic stations are commonly part of several networks as sharing of data is common practice. Seismic networks can be either permanent or temporary (e.g., <http://www.usarray.org/>), i.e., deployed for a shorter limited time period often linked to very specific research projects. Normally, less effort is put into the installation of temporary networks.

Seismic Stations

The main elements of a seismic station are the vault and housing of the equipment, the power supply to the equipment, the sensor, the digitizer with a timing device and recorder, and communication. Figure 2 shows how these units can be connected at a seismic station, and Fig. 5 shows



Seismic Network and Data Quality, Fig. 1 Map of the Global Seismograph Network (GSN)



Seismic Network and Data Quality, Fig. 2 Equipment used at a typical seismic station. The analog signal from the sensor is converted to a computer-readable format in the digitizer that also time stamp the data with a GPS; a datalogger is storing and handling the transmission the

data. A router and Internet connection are standard equipment for the data transmission. Some units will have integrated datalogger and digitizer or integrated digitizer and sensor. The units will naturally also need power supply

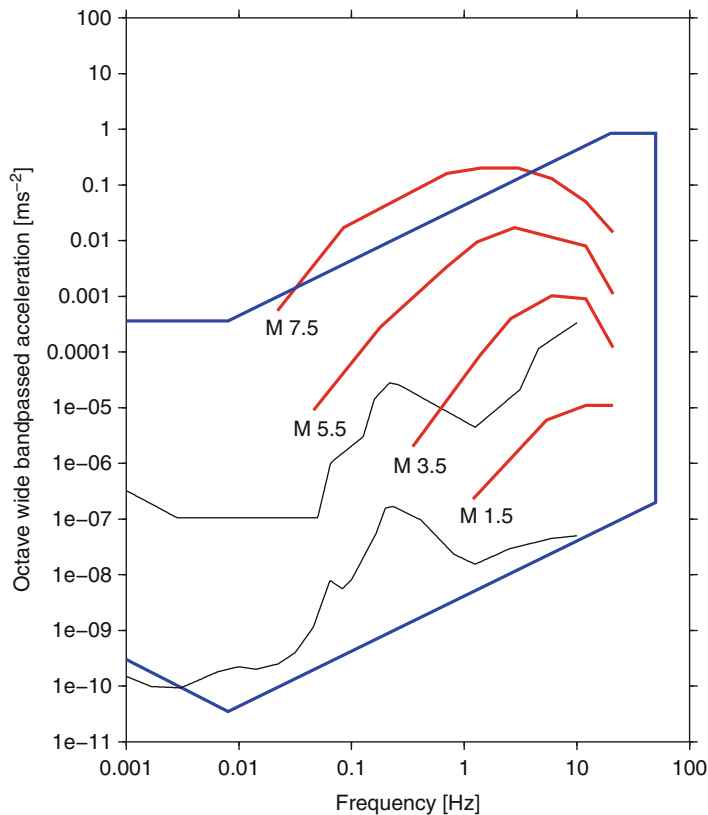
an example of a seismic station. All these elements will have an effect on the data quality and have to be considered carefully. When building a seismic station, there are normally financial constraints, and one has to find the optimum solution also considering the main monitoring purpose. The main choice for the sensor is between seismometer, also referred to as weak motion sensor, and accelerometer, also called strong motion sensor. The difference is that seismometers are more sensitive, particularly at lower frequencies, and the accelerometer will have a higher clip level than the seismometer. At some seismic stations, both a broadband seismometer and an accelerometer are installed to improve the dynamic range of the seismic station.

The main parameters describing the seismometer are the sensitivity, the natural frequency, and the damping. The response of the seismometer is given by its transfer function (impulse response) that has an amplitude and phase-delay part. Seismometers are available with different natural periods. Traditionally, long-period and short-period sensors were used, where the maximum gain would have been at 20 s and 1 s, respectively. Today, broadband sensors are the most common and typically have the high sensitivity between 120 s and 50 Hz. Seismometers are able to record the lowest possible expected ground noise levels within a given frequency band (limited by the instrument self-noise), and this is the most important technical specification of the

sensor. Accelerometers remain on-scale even for the largest earthquakes and are the most common instrument for earthquake engineering applications.

The main considerations when choosing a digitizer are the dynamic range and the timing accuracy. An important setting on the digitizer is the sampling rate, which according to the Nyquist theorem allows to resolve frequencies less than half the sampling frequency. A normal sampling rate at a local seismic network is 100 Hz. Information of earthquake signals above 50 Hz is therefore not recorded, due to the Nyquist theorem. The sampling rate is thus set according to the expected signal frequencies but also restricted by limitations in data storage and communication. Communication to a seismic station can be achieved in many different ways (e.g., cell phone, satellite, the Internet, digital radios) and is an important element as this allows data to be available close to real time. Most seismic stations are able to send data in this way to one or several central recording systems, and this is what today defines a seismic network.

Before one selects instrumentation for a seismic station, it should be considered if the instrument will be able to properly record the earthquakes it is meant for. A way to verify these requirements is to display the instrument sensitivity and clip levels in frequency–amplitude diagrams, with the expected response of likely earthquakes. Figure 3



Seismic Network and Data Quality, Fig. 3 Frequency–amplitude plot for octave-wide band passes of ground-motion acceleration. The blue polygon represents the sensitivity limits of an STS-2 broadband seismometer, where the frequency limit at 50 Hz is given by the Nyquist frequency for a sample rate of 100 Hz. The lower acceleration amplitude limit in the band 0.001–50 Hz is the minimum sensitivity of the sensor,

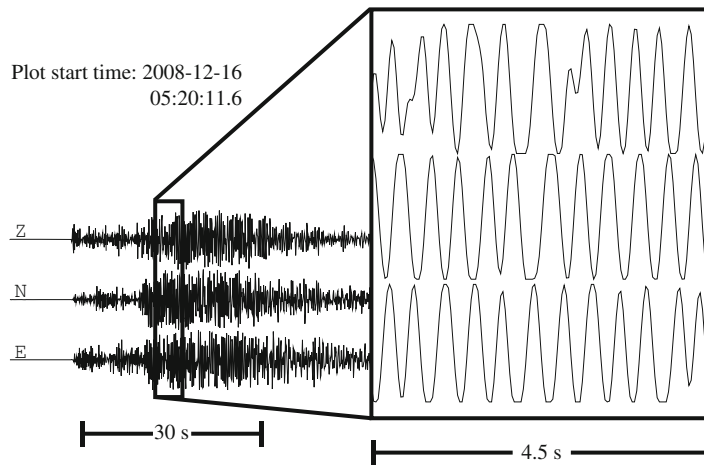
and the upper acceleration amplitude limit in the band 0.001–50 Hz is the clip level for the sensor. Sensor limits are scaled down to account for the band passing of the data. The red lines show the possible acceleration levels from different magnitude regional earthquakes that can be recorded at a distance of approximately 100 km. The two black lines show the USGS high- and low-noise models (Peterson 1993) (See also Clinton and Heaton 2002)

shows such a diagram for an STS-2 seismometer with regional earthquakes of magnitudes 1.5, 3.5, 5.5, and 7.5. Note that a magnitude 7.5 is above the clip level in this case, so another type of sensor should be considered for earthquakes of this strength. Figure 4 shows an example of a clipped seismic signal from a regional earthquake.

Data Exchange

The exchange of seismic data has been practiced since the installation of the very first stations as little could be done based on single stations. Prior

to the exchange of digital data through the Internet, seismologists would mail the seismograms on paper or microfilm to each other. Today, it is still common practice (with some exceptions) to exchange and share data between the now enormous number of seismic stations. This sharing of data is essential for research projects as well as for seismic monitoring services as earthquakes often occur in between different networks and only a combination of the networks allows good determination of even the basic earthquake parameters. Data (including waveform data and derived data) can be exchanged in many different ways from email to data downloads from a server and receiving data in near real time. There are a



Seismic Network and Data Quality, Fig. 4 Seismogram showing a magnitude 4.2 earthquake recorded at a distance of 87 km on the GID station in Denmark. The channels represent individual sensors oriented in the vertical, north-south, and east-west directions, given by

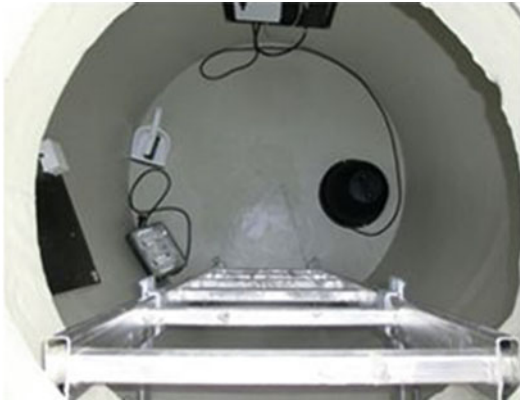
the Z, N, and E, respectively. The signal is clipped since the signal exceeded the digitizer clip level. The instrument is the short-period 4.5 Hz GeoSIG GBV-315 (Data are from the Geological Survey of Denmark and Greenland)

number of different public domain proprietary protocols for the near real-time exchange of data between the stations and data centers and also for the exchange between data centers. Most of these allow for data exchange in near real time and the recovery of missing data that may occur due to transmission errors. The public domain SeedLink protocol is one of the most used for real-time data transmission, especially in between data centers (Hanka et al. 2000). The SeedLink protocol transmits data in 512 byte packages, where station, channel, location, or network codes can be requested. The SeedLink protocol ensures that data are not lost in case of disconnected transmissions for shorter periods. This is also common in most commercial software solutions for data transmission in seismic networks. The largest data center receiving and providing seismic waveform data globally in a number of ways is the Incorporated Research Institutions for Seismology (IRIS) Data Management System (DMS) (see <http://www.iris.edu/hq/programs/dms>). It is also becoming possible to operate distributed data centers as, for example, done in Europe through the ORFEUS European Integrated Data Archive (EIDA) initiative (<http://www.orfeus-eu.org/eida/eida.html>). Exchange of parametric data, the data that contain information

on earthquake (e.g., phase travel times and amplitudes), is often carried out between neighboring networks. On a global scale, the main center for the parametric data is the International Seismological Center (ISC) (see <http://www.isc.ac.uk>).

Data Processing

The evaluation or processing of the data recorded by the seismic network normally takes place at the central recording site. Depending on the purpose of a seismic network, different automatic processing routines are applied to the data when it comes in. For most networks, the main focus is on detecting seismic events. This is done by checking the seismic signal and detecting changes from the background noise in the signal at individual stations. The information from a number of stations is combined to detect, locate, and estimate magnitude of the seismic events using the entire network. This is done for networks with different dimensions and monitoring purpose requiring different approaches. The outcome from the automatic processing is normally inspected before dissemination to the public in so-called earthquake early warning systems. This type of warning or information dissemination is



Seismic Network and Data Quality, Fig. 5 Example of a seismic station installed in southern Norway. The vault housing (with the seismometer and digitizer) is constructed from prefabricated concrete rings. (*left*).

Communication and power are available from a nearby hut (*right*). (Photo taken by Ole Meyer, University of Bergen)

an important responsibility of seismic monitoring systems. At a later stage, the data are further processed through manual interactive data processing. For both the automatic and manual data processing, there are a number of public domains as well as commercial software packages. Figure 6 shows phase readings done with the SEISAN software package (<http://seisan.info/>).

Metadata

The meta- or descriptive data are an essential part of the seismic data as without these the data is not complete. Data are identified by a so-called SCNL code (giving stations, component, network, and location) and the timing of the data. The station code represents the locality of the seismic station, the component code reflects the instrumentation, the seismic network code identifies the operator of the station, and the location code allows distinguishing between two instruments of the same type at one station. The metadata associated to the data streams will include information such as the location of the station (latitude, longitude, elevation), the orientation of the sensors (azimuth and dip), the sample rate, and the instrument response.

The main standard for seismic data is the Standard for the Exchange of Earthquake Data

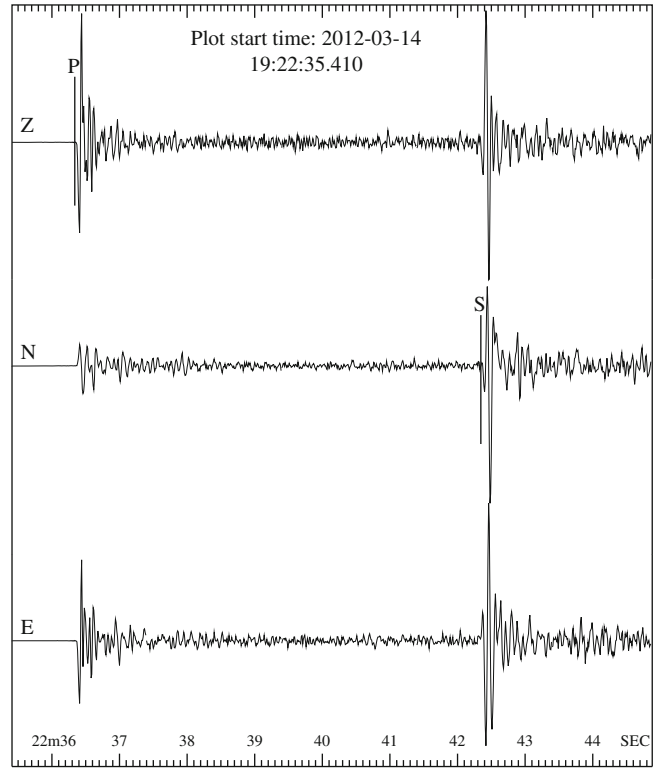
(SEED). One part of SEED refers to the waveform data, known as miniSEED; the other part of SEED, known as the dataless SEED, contains the metadata. The network codes for SEED are organized by the International Federation of Digital Seismograph Networks (FDSN, <http://www.fdsn.org/>). In addition to each component of the seismic data, the complete instrument description makes the data complete. This information makes it possible to convert the output recorded at the seismic station into true ground motion. Incorrect metadata is often the source to failed analysis of earthquake records.

Virtual Seismic Networks

Virtual seismic networks or virtual nets are the type of network where data from seismic stations in different seismic networks are collected and processed as a new “virtual” seismic network. The virtual seismic networks are often seen in regions in between two or more seismic networks where a joint processing of data from the region will improve the detection capabilities. An example of such virtual seismic network is the Greenland Ice Sheet Monitoring Network (<http://www.glisn.info/>), where data from nine different seismic networks are collected and processed jointly to improve the detection capability of glacial

Seismic Network and Data Quality,

Fig. 6 Seismogram displaying an earthquake recorded at a distance of about 50 km. The 9.5 s of data shows the three channels of the seismic station BLS5 in Norway. The channels represent individual sensors oriented in the vertical, north–south, and east–west directions, given by the Z, N, and E, respectively. The P and S phases are seen at 19:22:36.5 and 19:22:42.5, respectively (The data is from the National Norwegian Seismic Network. Data are unfiltered)



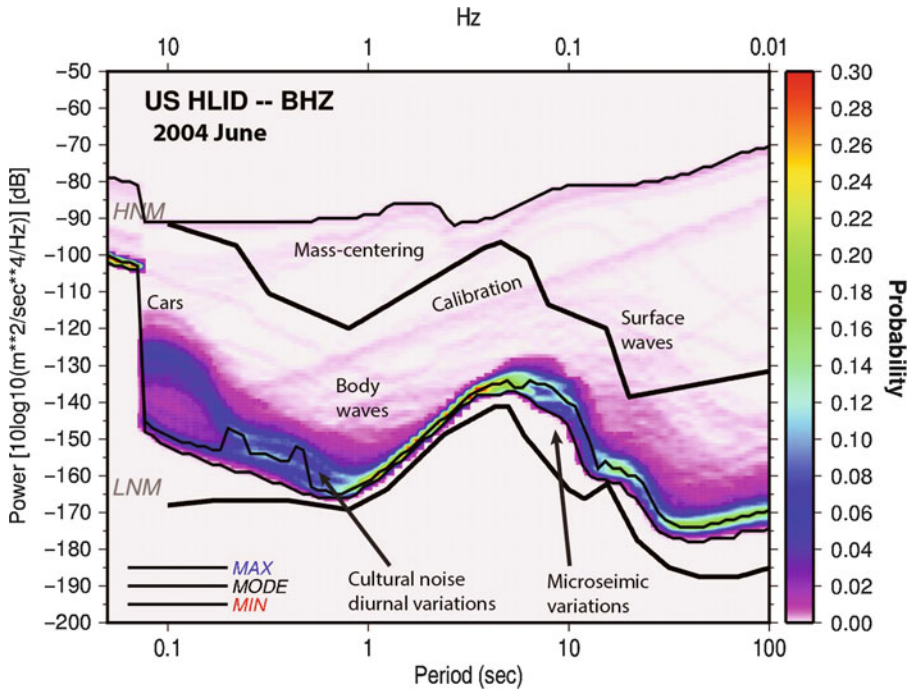
earthquakes in Greenland. Other virtual seismic networks collect data after quality or instrument type; see, e.g., the `_REALTIME` and `_STS-1` virtual networks operated by IRIS (<http://www.iris.edu/vnets>). A precondition for such networks to operate is the use of common standards for real-time data transmission as described above.

Seismic Noise and Data Quality

With seismic waveform data, a number of parameters can be used to quantify data quality. The main parameters are ambient noise levels, timing accuracy, and data completeness. The noise levels at a seismic station depend on a number of factors including the true ambient ground noise, the vault construction, and the instrumentation. Increased noise levels at a seismic station result in a lower signal to noise ratio, which means that a seismic station is worse in detecting and observing seismic events.

When talking about “data quality,” the question often is whether it is sufficient or insufficient. This depends on the application of the data. For example, a seismic station may perform well for local earthquake recording, but not provide sufficient data for global earthquake monitoring, if the used instruments are designed only to measure seismic signals above 5 Hz or if the ambient noise level is very high at the lower frequencies. If data quality is sufficient, it means that the collected data provides the information needed to answer the questions asked prior to the collection of the data. On the other hand, if data quality is insufficient, it means that the collected data cannot answer the questions at hand. Data quality is therefore a term used to describe if data is adequate for its intended use. The border between sufficient and insufficient data quality is not sharp since data are often collected to answer multiple questions or since the data quality might change over time during the collection of the data.

Improving data quality is first of all achieved by reducing the noise in the data, both for seismic



Seismic Network and Data Quality, Fig. 7 Ambient noise probability density functions (PDF) of data from the HLID station show how cars, calibrations, and other sources are clearly distinguished from the natural ambient noise (McNamara and Boaz 2005). Auto mass re-center and calibration signals are due to the station operator re-leveling the sensor and checking the amplitudes,

respectively. Body wave and surface wave signals show the total power of all the earthquake records in the period of the PDF. LNM and HNM are the low- and high-noise models, respectively. Note the peak in the noise at about 5 s period; this is due to the ocean-generated noise (See also http://geohazards.cr.usgs.gov/staffweb/mcnamara/PDFweb/Noise_PDFs.html)

stations designed to record local or distant earthquakes. The ambient or microseismic noise has both natural and man-made sources. Typical man-made sources are roads, machinery, airports, railways, etc., and within some constraints it is often possible to stay away from these. The main natural source of seismic noise is ocean waves, and while stations can be deployed at some distance from the coastline, it will not be possible to get away from this noise completely. Wind and atmospheric changes also result in increased noise, and this can mostly be avoided by building better seismic vaults, by putting sensors in pressure-tight enclosures, and also by increasing the depth of installation. The seismic instrumentation, both seismometer and digitizer, also has self-noise, and this can be minimized when choosing the equipment but also during the

installation process. The instruments should always be well grounded to avoid static electricity. Also ambient variations in the magnetic field should be avoided since it can limit the quality (Forbriger et al. 2010).

The most common way to present the microseismic noise at a station is to compute acceleration power density spectra and to display the probability of noise levels over a longer time period on probability density plots (McNamara and Buland 2004; McNamara and Boaz 2005). An example is shown in Fig. 7. This type of plot allows the identification of the different noise sources that act at different frequencies. However, one needs to be aware that the “real” ground noise is only resolved in a specific frequency range depending on the equipment.

The noise at frequencies greater than 2 Hz is related to cultural sources. The local maxima at around 4–8 s and 10–16 s are related to standing waves and waves breaking near the shore, respectively; see Fig. 7. Even more long-period noise is related to atmospheric changes. There can be strong differences at individual sites between day and night related to the changes in cultural activity but also between different seasons. For example, in northern Europe the noise caused by the ocean is much higher during the winter compared to the summer.

Borehole sensors are often used to improve data quality, since borehole sensors are not as affected by noise generated at the surface, as sensors placed at or near the surface. The effect is documented by Carter et al. (1991) stating that 945 m below the surface, noise levels between 15 and 40 Hz are reduced by 10 dB. For periods of between 30 s and 100 s, Hutt et al. (2002) report that the noise reduction is most prominent at the first few tens of meters. The disadvantage of borehole installations with respect to surface installations is that the instruments required are more expensive to buy and install and servicing the sensor is more difficult.

Network Maintenance and Quality Control

Seismic networks require a significant amount of maintenance work and attention to detail. Monitoring of the data quality provides an essential tool in the identification of problems with seismic stations. Being able to identify problems and respond quickly is key to successful operation of seismic networks. Recording of changes to instrumentation is also important.

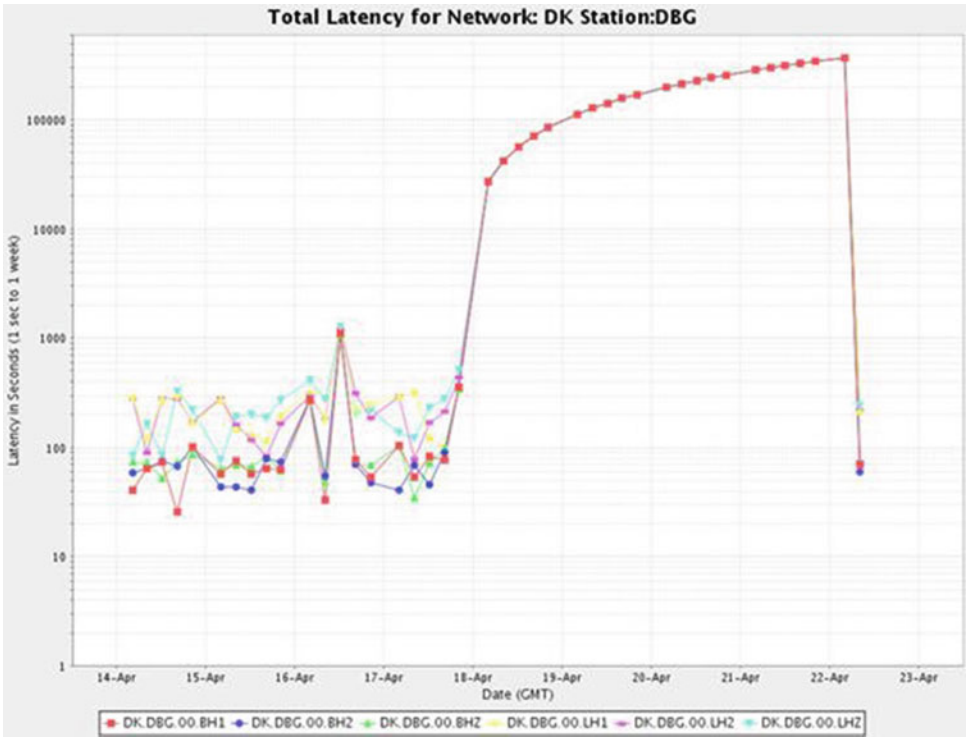
Data quality can be looked at while the data is coming in, later to evaluate the performance of a network or at the end of a temporary deployment. Seismic networks make use of tools to monitor parameters for quality control. One such system is QUACK (Quality Analysis Control Kit) available at IRIS (<http://www.iris.washington.edu/servlet/quackquery/welcome.do>). QUACK provides information on several parameters such as availability,

data completeness, gaps (lack of data in a time interval), noise levels, and data latency (data arriving late). Most seismic networks record data continuously, and measuring data completeness is an efficient tool to evaluate the overall network performance. Checking for gaps is typically done to identify issues with the communication or equipment.

One parameter that can be checked in near real time is the data latency. An increase in latency will indicate unstable communication. If the latency increases continuously, it will indicate that the connection to the station is lost or worse the recorder, digitizer, or other units at the station are broken. Figure 8 shows an example on latency where the connection to the seismic station was down for a few days. Monitoring of latency will also allow evaluating the quality of communication during regular operation.

It is important that seismometers are installed with correct horizontal orientation so that the north and east components are oriented within the true geographic system. It is also possible to have different orientation, but this needs to be recorded with the station's metadata. The orientation of sensors can be computed retrospectively from the recordings of seismic events. For example, the "LDEO Seismology Project: Waveform Quality Center" measures misorientation by comparing computed to observed polarization of surface waves (<http://www.ldeo.columbia.edu/%7Eekstrom/Projects/WQC.html>). They also attempt to measure changes in gain of seismometers over time from recorded data. To record and possibly avoid changes in gain, regular calibration of the seismic equipment should be part of the network routine operation.

When plotting the noise spectra color scaled over time, one obtains spectrograms that are useful in evaluating the changes of noise. Failure of instrumentation will be seen easily as it will look like an apparent sudden change in noise levels. At the ORFEUS Data Center, PSD plots are generated in yearly intervals showing frequency bands around 0.01, 0.05, 0.5, and 2 Hz (http://www.orfeus-eu.org/data/data_quality.html). An example showing a 1-year PSD from the broadband station SFJD in Greenland is seen in Fig. 9.

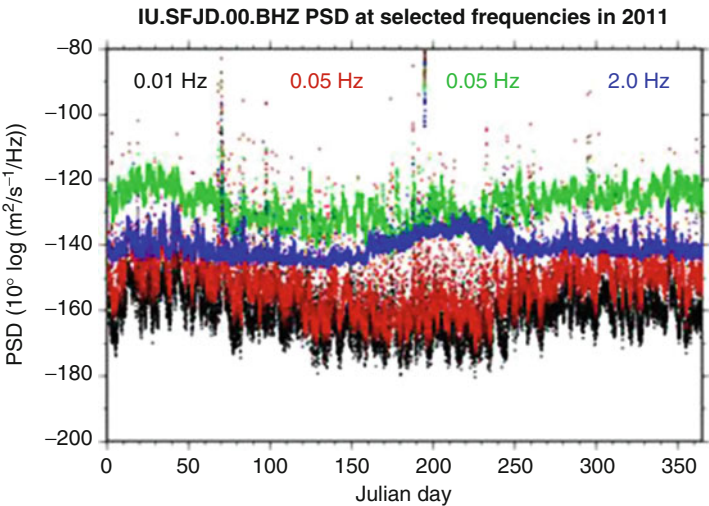


Seismic Network and Data Quality, Fig. 8 Latency of data streams from the DBG station in Greenland. The latency is the time between the last sample in the data

record and GMT. The connection to the station was lost on April 18th and regained on the 22nd (This figure is generated at <http://www.iris.edu/servlet/quackquery/>)

Seismic Network and Data Quality,

Fig. 9 Power spectral density (PSD) plot of data from the SFJD station in Greenland, covering the year 2011. The *black, red, green, and blue* lines show the frequencies 0.01 Hz, 0.05 Hz, 0.5 Hz, and 2.0 Hz, respectively (With permission from Dr. Sleeman, ORFEUS. This figure is obtained from <http://www.orfeus-eu.org/>)



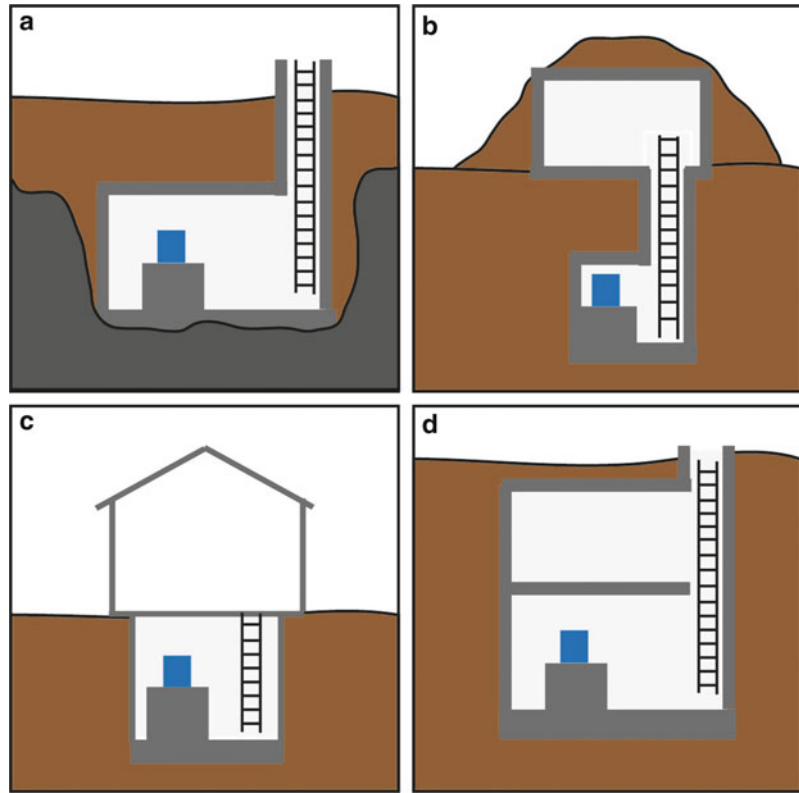
Site Selection and Vault Construction

The choice of site, selected to record seismic data, is essential to obtain data with a high

quality. One can install instruments of the highest quality at sites where a high ambient noise level will make the data useless for the purpose they were intended for. Before deciding on a specific

Seismic Network and Data Quality,

Fig. 10 Example of BB vaults from the GEOFON network. (a) Underground bunker vault for remote recording, (b) wide and shallow borehole type, and (c and d) simple bunker construction. Note that b–d allow onsite recording since there is a separate recording room (See also Fig. 7.55 by W. Hanka in NMSOP, Vol.1, Bormann (Ed.), 2002)



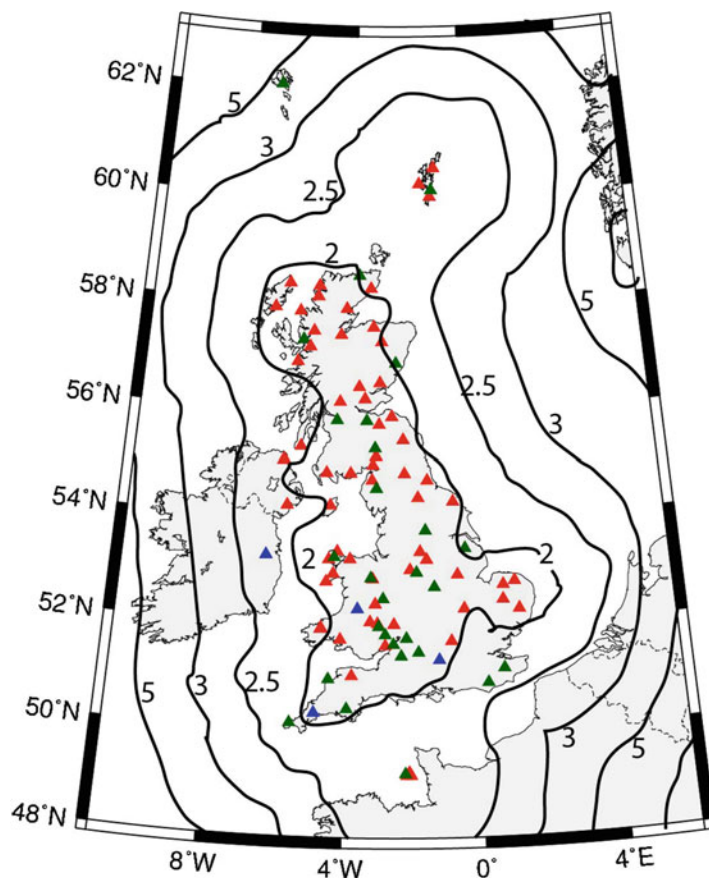
site, it is common practice to carry out a site survey where all relevant criteria are considered and noise measurements are done for a few days to identify potential noise sources. If the source of the seismic signal that one wishes to record is known, for example, a small area that is known to be seismically active, one should aim to get as close to the source as possible. One should also aim to have stations installed with the highest azimuthal coverage possible. It is of course important to stay away from the man-made sources of noise mentioned above. Site selection is often dependent on other factors than the ambient noise, such as access to power and communication as well as the physical access to the site. One might therefore not always be able to use the site with the lowest noise level, since it does not meet these requirements.

Constructing good vaults is also very important, and possibilities range from boreholes to

relatively deep substantial vaults to rather shallow and simple installations (Fig. 10). The coupling of the sensor to bedrock is very important in all types of installation. If the budget is not limited, one would attempt to build a pier placed on bedrock several meters below the surface. Around the pier, a room would be constructed that would be isolated from the pier. It is essential to achieve rather stable temperatures and pressure around the sensor. Another room may be placed on top to house the remaining equipment. In a simpler setup, one can use a concrete structure with a smaller diameter that has the sensor at the bottom, a few meters below the surface. Care has to be taken to either drain water away or avoid water ingress completely. Many such designs have been made taking into account local conditions and available materials. The type of vault chosen should also be based on the specifications of the equipment that will be used.

Seismic Network and Data Quality,

Fig. 11 Detection capability of the UK seismograph network. Unit is magnitude on the Richter scale. *Green* triangles are broadband stations, *red* triangles are short-period stations, and *blue* triangles are broadband stations operated by other agencies (<http://www.earthquakes.bgs.ac.uk/>)



Detection Capability

If detection of earthquakes is the main purpose of a seismic network, knowledge on the detection capability of the network is essential when the network is designed and evaluated. The detection capability depends on the level of the background noise, the type of instrumentation, and the configuration of the network. The detection capability is mostly improved by lower background noise in the frequency range of interest, the distribution of stations, and their distance to the area of interest. An example of detection capability is shown in Fig. 11. Here it is assumed that the amplitude threshold for the detections of an earthquake must be twice the background noise level and that the earthquake is recorded on at least four stations (http://www.earthquakes.bgs.ac.uk/monitoring/detection_capability.html).

References

- Carter JA, Barstow N, Pomeroy PW, Chael EP, Leahy PJ (1991) High-frequency seismic noise as a function of depth. *Bull Seismol Soc Am* 81:1101–1114
- Clinton J, Heaton T (2002) Potential advantages of a strong-motion velocity meter as opposed to a strong motion accelerometer. *Seismol Res Lett* 73:332–342
- Forbriger T, Widmer-Schmid R, Wielandt E, Mark Hayman M, Ackerley N (2010) Magnetic field background variations can limit the resolution of seismic broad-band sensors. *Geophys J Int* 183(1):303–312. doi:10.1111/j.1365-246X.2010.04719.x
- Hanka W, Heinloo A, Jaekel KH (2000) Networked seismographs: GEOFON real-time data distribution. *ORFEUS Newsl* 2(3):1–24
- Havskov J, Alguacil G (2010) Instrumentation in earthquake seismology. Springer, Dordrecht, p 349
- Havskov J, Ottemöller L (2010) Routine data processing in earthquake seismology with sample data, exercises and software. Springer, New York, p 380
- Hutt CR, Bolton HF, Holcomb LG (2002) Seismograph networks. In: Lee WHK, Jennings P, Kisslinger C,

- Kanamori H (eds) International handbook of earthquake & engineering seismology, part 1. Academic Press, p 1200
- McNamara DE, Boaz RI (2005) Seismic noise analysis system, power spectral density probability density function: stand-alone software package, United States geological survey open file report, No. 2005–1438, p 30
- McNamara DE, Buland RP (2004) Ambient noise levels in the continental United States. *Bull Seismol Soc Am* 94(4):1517–1527
- Peterson J (1993) Observations and modeling of background seismic noise. Open file report 92–302, U.S. Geological Survey, Albuquerque
- Scherbaum F (2001) Of poles and zeros. *Fundamentals of digital seismology*, vol 15, 2nd edn, Modern approaches in geophysics. Kluwer, Boston, p 269

Further Reading

The most complete coverage of these topics is probably given in the *New Manual of Seismological Observatory Practice* (<http://http/nmsop.gfz-potsdam.de>), giving an overview, the theoretical background, as well as examples and information sheets. Textbooks on seismic instrumentation are Havskov and Alguacil (2010) and Scherbaum (2001). A textbook on data processing for earthquake seismology is Havskov and Ottemöller (2010). Stein and Wyss (2003) and Lay and Wallace (1995) also have relevant chapters.

Seismic Noise

Peter Bormann

Formerly GFZ German Research Center for Geosciences, Potsdam, Germany

Synonyms

Ambient noise; Cultural noise; Digitization noise; Instrumental noise; Man-made noise; Microseismic noise; Microseisms; Self-noise; Signal-generated noise

Introduction

Seismic signals are usually transient waveforms radiated from a localized natural or man-made

seismic source. They can be used to locate the source, to analyze source processes, and to study the structure of the medium of wave propagation. In contrast, the term “seismic noise” designates undesired components of ground motion that do not fit in our conceptual model of the signal under investigation. What we identify and treat as noise in seismic studies depends on the available data, on the aim of our study, and on the method of analysis. Accordingly, data treated as noise in one context may be considered as useful signals in other applications. For example, short-period noise (frequency >0.2 Hz) can be used for investigating the potential amplification of earthquake-generated ground motions due to local geological site conditions, also termed *microzonation* (e.g., Parolai 2012), and generally, both short- and long-period noise (frequency < 0.1 Hz) can also be used for tomographic studies of inhomogeneities in the Earth’s crust (e.g., ► [Noise-Based Seismic Imaging and Monitoring of Volcanoes](#)).

According to Bormann and Wielandt (2013), disturbing noise in seismic records in a wider sense may comprise:

- Ambient vibrations due to natural sources (like ocean microseisms, wind, etc.)
- Man-made vibrations (from industry, traffic, etc.)
- Secondary signals resulting from Seismic Wave Propagation in inhomogeneous medium (due to *wave scattering* and therefore also termed *signal-generated noise*)
- Effects of gravity (like Newtonian attraction of atmosphere, horizontal accelerations due to surface tilt)
- Signals resulting from the sensitivity of ► [Broadband Seismometers](#) to ambient conditions (like temperature, air pressure, magnetic field, etc.)
- Signals due to technical imperfections or deterioration of the seismic ► [Passive Seismometers](#) (corrosion, leakage currents, defective semiconductors, etc.)
- Intrinsic ► [Seismometer Self-Noise and Measuring Methods](#) of the Seismic

Instrumentation (like Brownian noise, electronic and quantization noise)

- Artifacts from data processing

However, only the first four items are microseismic noise in a proper sense and elaborated in the following whereas the other ones relate to *seismometry* and *data processing* and are referred to only en passant.

The study of earthquakes or the imaging of the Earth with seismic wave arrivals requires ► [Seismic Event Detection](#) above background noise. Levels of natural ambient noise may vary by 60 dB (i.e., a factor of 1,000 in amplitude) depending on location, season, time of day, and weather conditions. This corresponds to differences in detection thresholds for seismic arrivals by about three magnitude units. Therefore, prior investigation of noise levels at potential recording sites is of utmost importance, as is environmental shielding and proper installation of seismic Sensors so as to reduce the influence of various sources of natural and man-made (cultural) noise (see sections “[Causes and Basic Characteristics of Ambient Noise Observed on Land](#)” to “[Improving the Signal-to-Noise Ratio SNR by Data Processing](#)” below; Trnkoczy et al. 2011; Forbriger 2012).

The essay will brief on:

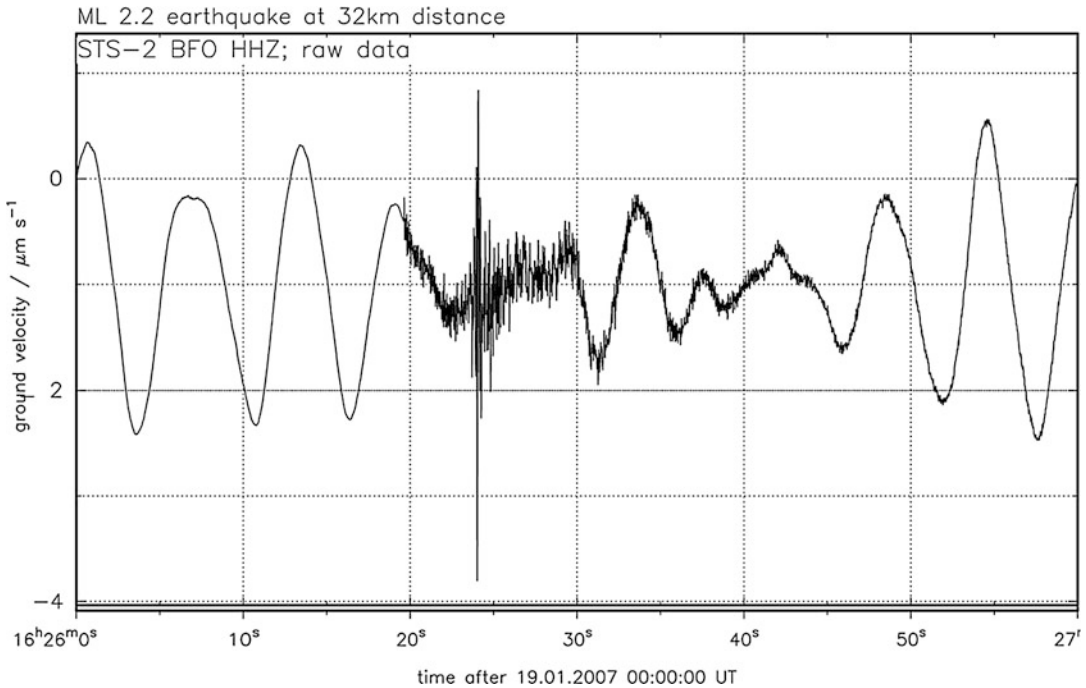
- Different mathematical representations of noise and signals
- Origins of natural and man-made ambient noise
- Basic features of noise in the wide frequency range of interest for the recording of local micro-earthquakes to normal modes of the Earth ($0.2 \text{ mHz} < f < 100 \text{ Hz}$)
- Influence of filter procedures and the sensor environment on the appearance of signals and noise in seismic records
- Task-dependent procedures for signal-to-noise ratio (SNR) improvement of earthquake and explosion records

More than 98 % of all seismological recordings are made on land, despite the growing

interest in ► [Ocean Bottom Seismometer](#) (OSB) installations in order to improve the global ray path coverage for tomographic studies of the Earth and for regional studies of seismicity and Earth structure in the ocean environment. But OBS deployment, operation, and maintenance are rather expensive and logistically demanding, and the noise conditions on the ocean floor are usually worse than on land (Shinohara et al. 2011; Webb 2002). Therefore, this essay will deal, also because of strict space limitations, only with microseismic noise conditions on land. More details about such noise in both environments as well as on the specifics of seismic signals in different bandwidth ranges and related noise-dependent local and global detection thresholds can be found in Bormann and Wielandt (2013).

With respect to historical aspects of the recording and treatment of microseismic noise, it has to be considered that most of the early twentieth-century seismographs by Wiechert, Mainka, Galizyn, Bosch-Omori, Milne-Shaw, and others were mechanical medium-period broadband sensors of low magnification of the ground motion. Only the more sensitive ones with 100–500 times magnification were already able to record the ever-present ground motions with peak amplitudes around 6 s which constitute the dominating background noise for any seismic measurement with maximum amplitudes in the μ micrometer range (Fig. 1). They were termed microseisms. Such recordings were first reported by Algue in 1900. Wiechert (1904) wrote that these microseisms are caused by ocean waves on coasts. Later it was found that it is necessary to discriminate between: (a) the smaller primary ocean microseisms with periods between 10 and 20 s, typically around $14 \pm 2 \text{ s}$ and (b) the secondary or *double frequency microseism* which is related to the main noise peak around $6 \pm 2 \text{ s}$ (Figs. 1 and 2).

Other types of seismic as well as cultural and instrumental noise became more or less disturbing only when the sensitivity of electromagnetic seismographs, and later of electronically amplified seismographs, was raised so much that they were able to resolve even ground



Seismic Noise, Fig. 1 STS-2 velocity broadband record of the 6 s secondary ocean microseisms at the Black Forest Observatory (BFO), Germany, superposed by high-

frequency waves of a local earthquake of magnitude $M_L = 2.2$ that occurred 32 km away (Figure by courtesy of Thomas Forbriger for NMSOP Chapter 4 © IASPEI)

motion amplitudes in the size range of molecules and atoms (10^{-9} to 10^{-10} m).

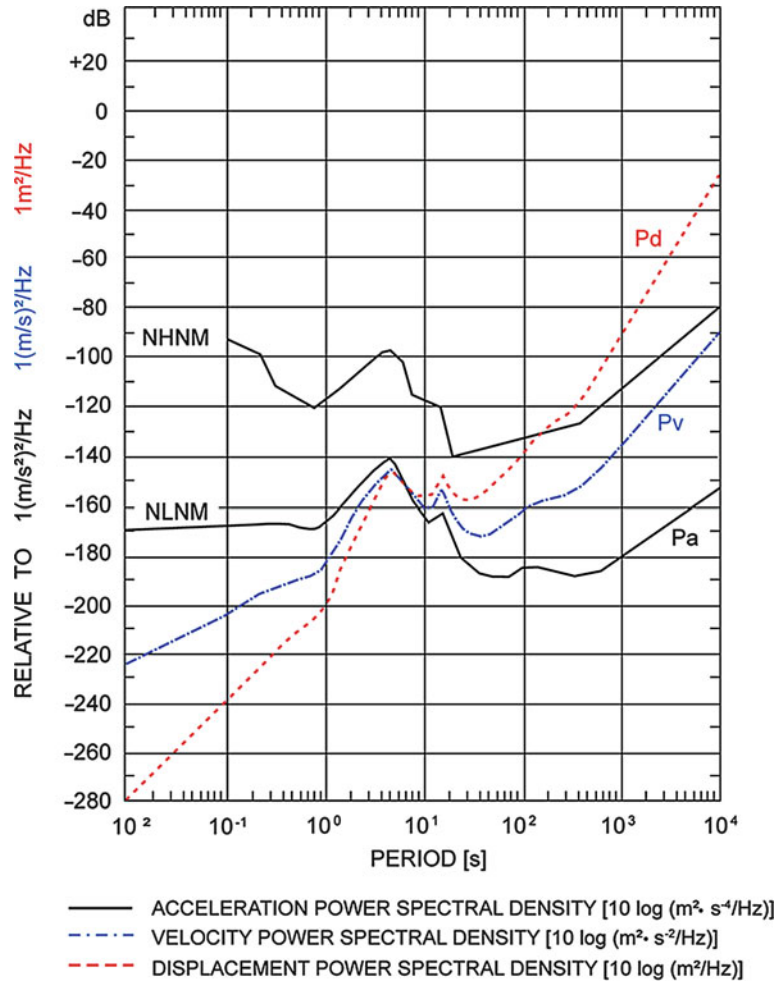
Noise conditions at seismic recording sites vary globally in the wide range of periods of seismological interest between some upper and lower level, termed the New High Noise Model (NHNM) and the New Low Noise Model (NLNM) according to Peterson (1993). These noise models are expressed in dB of *power spectral density* (see section “[Nature and Theoretical Representation of Signals and Noise](#)”) of acceleration related to $1 \text{ (m/s}^2\text{)}^2/\text{Hz}$ (Fig. 2). However, the actual noise at any recording site corresponds nowhere in the whole range of periods to either the NHNM or the NLNM. Some sites and seismographs are more affected by long-period tilt-noise at a period above 30 s, and this differently in each record component, while others are mostly disturbed by high-frequency ($f > 1\text{ Hz}$) man-made noise.

According to Fig. 2, variations in noise acceleration power are typically within about

30–80 dB, but they are much larger in velocity power (>130 dB) and displacement power (>250 dB). The latter is much more than the dynamic range of common modern digital seismographs with 24 bit *analog-to-digital converter* (ADC) and a dynamic range of 144 dB. In order to resolve, at good sites, at least the ambient noise with a few bits, even at the NLNM level, and still record without clipping also seismic signals from events in a reasonably large range of magnitudes and epicentral distance, current broadband seismographs used are *velocity sensors* such as the STS-1 and STS-2 ► [Recording Seismic Signals](#). Their dynamic range is given in Fig. 3 with respect to the NLNM when measured in a constant relative frequency bandwidth of 1/3 octave and represented by “average peak” amplitudes (see section “[Approximate Conversion of Power Densities into Recording Amplitudes](#)”). In contrast, classical analog records on paper or film had only a dynamic range of

Seismic Noise,

Fig. 2 Envelope curves of acceleration noise power spectral density P_a (in units of dB relative to $1 \text{ (m/s}^2\text{)}^2/\text{Hz}$) as a function of noise period (according to Peterson 1993). They define the new global high (NHNM) and low noise models (NLNM) which are currently the accepted standard curves for generally expected limits of microseismic noise depending on period. No single station site, however, meets in the whole period range the NLNM or NHNM. However, some exceptional seasonal, diurnal, or local site noise conditions may exceed these limits. Also shown are for the NLNM the related curves calculated for the displacement and velocity power spectral density P_d and P_v in units of dB relative to $1 \text{ (m/s}^2\text{)}^2/\text{Hz}$ and $1 \text{ m}^2/\text{Hz}$ (Modified version of Fig. 2 from P. Bormann 1998 © Springer; with kind permission from Springer Science + Business Media)



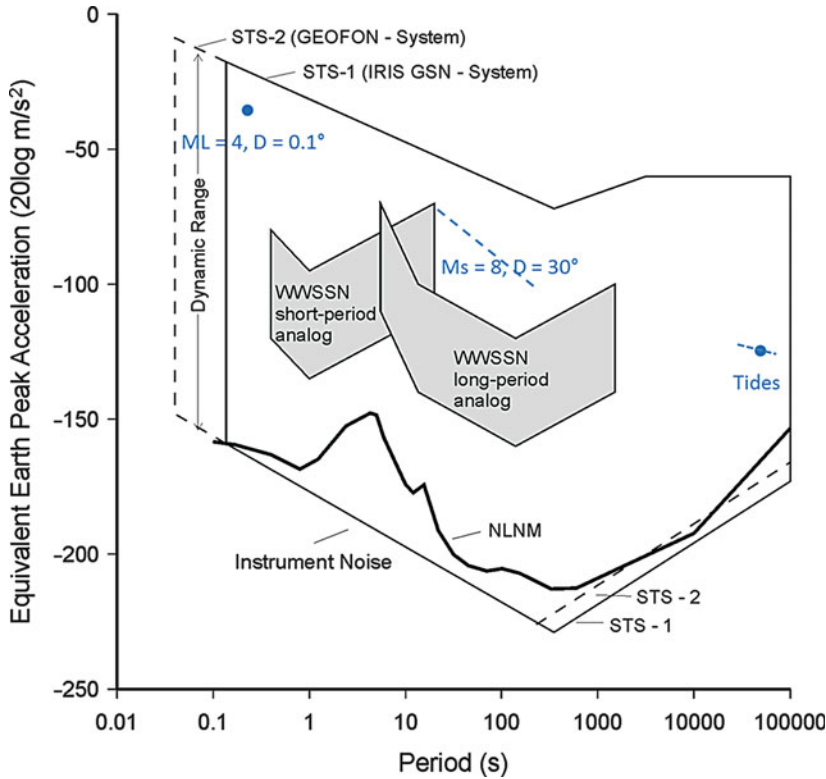
some 40 dB. They could cover only a very limited range of noise and record signal amplitudes between about 1 mm and 100 mm. Covering a range of 144 dB instead would correspond to covering an amplitude range between 1 mm and 16 km! However, what the actual usable dynamic range of a seismograph means has been discussed in Sect. 4.2.2 of Bormann and Wielandt (2013).

The reasons for the main types of natural and cultural microseismic noise observed on land and strategies for their avoidance or reduction in seismic records are discussed in the following sections after a brief discussion on the principal difference between seismic signals and noise, their mathematical representation, and the correct way of comparison.

Nature and Theoretical representation of Signals and Noise

Fourier Transformation of Continuous and Sampled Transient Signals

Seismograms contain noise and the desired signals. Sometimes the signal is completely masked by the noise. One of the main tasks in applied seismology is to ensure a good signal-to-noise ratio (SNR) both by selecting recording sites with low ambient noise and by data processing. The success of the latter largely depends on our understanding of the ways in which seismic signals and noise differ. The basic mathematical tool for this purpose is the harmonic or *Fourier analysis*, that is, the decomposition of a signal into sine waves. Depending on the type of signal



Seismic Noise, Fig. 3 A representation of the bandwidth and dynamic range of conventional analog World Wide Standard Seismograph Network (WWSSN) short- and long-period and digital *broadband seismographs* (STS1 and STS2 with good shielding - see Wielandt 2011 and Forbriger 2012) with respect to the equivalent Earth peak acceleration generated by solid Earth tides due to the attraction forces exerted by the Sun and the Moon, by a magnitude 8 earthquake at about 3,300 km distance or by a local earthquake of magnitude 4 at about 10 km distance (see ► Earthquake Magnitude Estimation). The depicted lower bound is determined by the *Instrumental Self-noise*

(also Sect. 4.3.2 in Bormann and Wielandt 2013). The scale is in decibels (dB) relative to $(1 \text{ m/s}^2)^2$. Noise is measured in a constant relative bandwidth of one third octave and represented by “average peak” amplitudes equal to 1.253 times the root-mean-square (RMS) amplitude. NLNM relates to the global New Low Noise Model according to Peterson (1993; see Fig. 2). The requirement to be fulfilled for presenting seismic signal, noise, and clip levels in one diagram has been outlined in Sect. 4.3.3 of Bormann and Wielandt (2013). Copy of Fig. 4.7 in Bormann and Wielandt (2013), compiled by using data of W. Hanka and J. M. Steim; © IASPEI

(transient or stationary, continuous or sampled) different mathematical formulations must be used which have been outlined in detail in Sect. 4.2 of Bormann and Wielandt (2013). The essence is that according to the *Fourier theorem*, any arbitrary *transient function* $f(t)$ in the *time domain* can be represented by an equivalent function $F(\omega)$ in the *frequency domain*. The mathematical relationship between the two domains is defined by the *Fourier integral transformation*:

$$f(t) = (2\pi)^{-1} \int_{-\infty}^{\infty} F(\omega) \exp(i\omega t) d\omega \quad (1)$$

$$F(\omega) = |F(\omega)| \exp(i\phi(\omega)) = \int_{-\infty}^{\infty} f(t) \exp(-i\omega t) dt \quad (2)$$

where ω (in units radian/s) = $2\pi\nu$ is the circular frequency with ν = common frequency

(in units s^{-1}), $|F(\omega)| = \text{amplitude spectrum}$, and $\phi(\omega) = \text{phase spectrum}$. However, the Fourier integral transformation can only be applied to *transient signals* (i.e., signals that disappear or decay after some time so that they have a finite energy). For other signals, the integral may diverge. Moreover, the signals should have been radiated by localized finite sources thus being coherent signals with a defined phase relationship. In the case of digital data, the discrete transformation has to be applied. It also requires a sampled signal of finite duration. In contrast, signals which do not vary much within the interval of time over which earthquake signals are typically analyzed are mathematically treated as *stationary signals*, that is, signals whose statistical properties do not change with time. Such signals require a different mathematical treatment.

Normally, the detailed waveform of stationary signals is of no interest but rather subjected to efforts to remove the disturbing signal from the record. The exception is the above-mentioned application in *microzonation studies* (or *noise tomography*). However, it is of general interest to know how strong such stationary signals are in different parts of the frequency spectrum. This information is contained in a quantity named *power spectral density (PSD)*.

Transient signals such as earthquake seismograms may have a complicated waveform which is determined by the earthquake source, the structure of the Earth, and the properties of the seismograph. The amplitudes and phases of the harmonic components of the signal are not random numbers but follow certain mathematical relationships. Such waveforms or spectra are sometimes called deterministic, in contrast to the “stochastic” or “random” waveform of noise. However, in a strict sense, even microseismic noise is deterministic as well. But the difference is that its sources and propagation paths are not known well enough to predict or interpret the waveform, and therefore it is assumed that the spectral amplitudes and phases are random numbers. Again, this is a mathematical simplification that may or may not be appropriate.

The concept of frequency filtering (e.g., high-, low-, or band-pass filtration) has its roots in analog signal processing and is therefore intimately connected to the *Fourier integral transformation*. Practical computer procedures of the *discrete Fourier transformation*, however, do not use integrals and are often only approximations to the corresponding analog procedures. Supplementary information on different aspects of the Fourier transformation can also be found in Sect. 5.2.3 of Wielandt (2011).

A general problem is that frequency filtering reduces the amplitudes of harmonic components of a signal in part of the frequency spectrum, usually in order to suppress noise. Examples are given in section “[Frequency Filtering](#).” But filtering may also change the waveform of the desired signal in the time domain, the more so the narrower bandwidth. Systematic differences between various magnitude scales as well as magnitude saturation effects are the consequence of filtering the input signals with different responses in different and often rather limited frequency ranges (see ► [Earthquake Magnitude Estimation](#)).

Spectral Analysis of Stationary Signals (Noise)

The analysis of stationary noise usually aims at assessing “how strong” the noise is or whether it will prevent the detection of a specific earthquake signal. Frequency filters may help to suppress noise in one part of the spectrum but still preserve signals in another part. Therefore it is necessary to consider the spectral distribution rather than the total strength of both types of signals. However, the formal measures of “strength” for transient and stationary signals are incompatible and cannot easily be compared. Mathematically speaking, transient signals have a finite energy and zero power (in the average over all times) while stationary signals have an infinite energy but a finite power. The same holds for the spectral densities. For a quantification of stationary noise, the concept of *power spectral density (PSD)* is therefore appropriate while transient signals are more adequately described by their *energy spectral density (ESD)*.

The *Fourier integral transformation* as formulated in Eqs. 1 and 2 has the mathematical property (known as Rayleigh's or Parseval's theorem)

$$\int_{-\infty}^{\infty} f(t)^2 dt = \int_{-\infty}^{\infty} \frac{|F(\omega)|^2}{2\pi} d\omega = \int_{-\infty}^{\infty} |F(2\pi\nu)|^2 d\nu \quad (3)$$

The left side represents, in data processing, the total energy of a signal $f(t)$ which is not the physical energy but proportional to it. The right side represents the energy as an integral of $E = |F(2\pi\nu)|^2$ over all frequencies ν . Thus, according to the above definition, the integrand is the *energy spectral density* E of the signal.

Note that negative frequencies must be included in the integration. For real signals, the ESD is a symmetric function, $E(-2\pi\nu) = E(2\pi\nu)$. Therefore, by multiplying by 2, the integration can be carried out over positive angular frequencies only. This convention is normally used in engineering and other practical applications; the ESD is then given as twice the “mathematical” value. The same applies to the PSD. It is important to clarify which convention is being used.

Power is energy per time, usually expressed as the average power over at least one cycle of an approximately periodic signal. The concept of power is mainly applied to stationary signals such as sine waves or seismic noise whose average power is constant or varies slowly. A Fourier transform in a strict sense is not defined for such signals, but the signal over any finite time interval of length T can be Fourier analyzed. When both sides of Eq. 3 are divided by T , the left side represents the average power, and $P = E/T = |F(2\pi\nu)|^2/T = |F(\omega)|^2/T$ on the right side is then the *power spectral density* $P(\omega)$ for the interval T . But the PSD is not equivalent to a Fourier transform! The signal cannot be reconstructed from it with an inverse transformation because the phase information is lost. The smoothed and decimated PSD is however normally more useful because it contains all essential information in a concentrated form and eliminates arbitrary normalization factors.

The time derivative $\dot{f}(t)$ of a signal $f(t)$ has the Fourier transform $j\omega F(\omega)$. Thus, if the signal was originally measured as a displacement and the Fourier transform of the velocity is of interest then it is obtained by multiplying the original Fourier transform with $j\omega = j \cdot 2\pi\nu$ and the Fourier transform of the acceleration is then $-\omega^2 F(\omega)$. Going back from acceleration or velocity to displacement is achieved by dividing with the appropriate factor. The ESD and PSD are positive quadratic functions of the Fourier amplitudes. Therefore, differentiating the signal brings a factor ω^2 , integration a factor ω^{-2} , etc. A PSD of velocity can thus easily be converted into that of displacement or acceleration.

Representing PSDs in Decibels

As shown in Fig. 2, it has become a standard to measure microseismic noise as a PSD of acceleration, that is, in $(\text{m/s}^2)^2$ per Hz or m^2s^{-3} . But “power” in this context is not the physical power in watts but simply the mean square amplitude of ground acceleration. As in acoustics, the power ratio $r = (a_2/a_1)^2$ between two signals of amplitude a_1 and a_2 is often expressed in decibels (dB). This is a logarithmic measure that expresses a ratio as a difference of logarithms. The difference in dB is $10 \log_{10}[(a_2/a_1)^2] = 20 \log_{10}[a_2/a_1]$. When expressing the power spectral density P_a of acceleration relative to the metric unit $1 (\text{m/s}^2)^2/\text{Hz}$, it holds that

$$\mathbf{P}_a[\text{dB}] = 10 \log_{10} \left[P_a / 1 (\text{m/s}^2)^2 / \text{Hz} \right] \quad (4)$$

In Fig. 2, however, a comparison has been made of the respective numerical values for physically incommensurable quantities such as acceleration power density with velocity and displacement power density. Therefore, Eqs. 2 and 3 and the following two text lines do not imply that the decibel units on both sides of the equal sign mean the same. They only look the same because the incommensurable reference levels were omitted. For this reason, three decibel scales for acceleration power, velocity power, and displacement

power have been given in Fig. 2. This clarification notwithstanding, it holds that numerically

$$\mathbf{P}_v[\text{dB}] = \mathbf{P}_a[\text{dB}] + 20 \log(\omega^{-1}) \quad (5)$$

and

$$\begin{aligned} \mathbf{P}_d[\text{dB}] &= \mathbf{P}_a[\text{dB}] + 40 \log_{10}(\omega^{-1}) \\ &= \mathbf{P}_v[\text{dB}] + 20 \log(\omega^{-1}) \end{aligned} \quad (6)$$

Consequently, for $\omega = 1$ (period $T = 2\pi/\omega = 6.28$ s) it holds that $\mathbf{P}_a = \mathbf{P}_v = \mathbf{P}_d$ (see Fig. 2) although only the numbers are equal! Also, $(\mathbf{P}_d - \mathbf{P}_a) = 2 \times (\mathbf{P}_v - \mathbf{P}_a) = \text{constant}$ for any given period, negative for $T < 2\pi$ s, and positive for $T > 2\pi$ s.

Cross-correlation and Coherence of Seismic Signals and Noise

The *cross-correlation* of two continuous stationary signals $f(t)$ and $g(t)$ at the time lag τ is defined as the expectation value (practically, the mean value) of the sliding dot product $f(t) g(t + \tau)$:

$$(f \star g)(\tau) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} f^*(t) g(t + \tau) dt, \quad (7)$$

where f^* denotes the complex conjugate of f .

Properly normalized, the cross-correlation is a measure of the similarity between the two signals and is called their *coherence*. Like the power spectral density, it can be calculated via the Fourier transformation and specified as a function of frequency, in which case the name *coherence spectrum* is more adequate. Each signal is fully coherent with itself (*auto-correlation* with the coherence equaling 1). Independent random signals are incoherent (their coherence equals 0) although strictly this is only true in the limit of an infinite duration and after removing the mean. More interesting is the case that signals are partially coherent. Then, with some (frequency-dependent) fraction of the power, a common signal may be hidden in them, for example, a seismic signal masked by instrumental noise of the sensors.

Seismic signals depend not only on time but also on the location where they are recorded. It is

therefore meaningful to ask whether the signal recorded at one location is coherent with the same signal recorded at another location. This aspect of coherence is usually expressed as a “coherence length” over which the recorded signals have a specified coherence. The coherence length is only a fraction of a wavelength for omnidirectional noise (noise propagating in all directions) but can be much longer when noise originates in a limited source area and propagates essentially in one direction towards the station (Fig. 5). In a plane wave, the coherence length is theoretically infinite.

Comparing Spectra and Amplitudes of Transient and Stationary Signals

Bandwidth and Amplitude

Transient signals may be impulsive and oscillatory, and stationary signals either random signals (noise) or monochromatic sine waves. A typical seismogram belongs to the “oscillatory” category. The seismic signal near the focus of an earthquake should be impulsive. Ambient noise is often a superposition of all types, the “random” components being due to a large number of independent sources (traffic, wind, waves in the ocean), “transient” components being caused by identifiable sources such as passing cars, and nearly monochromatic components originating from machinery such as sawmills and pumps.

When the PSD of a stationary signal is constant over the bandwidth of a seismograph or a filter then, when reducing the filter bandwidth, the signal power is proportional to the bandwidth because it is the integral of a constant power density over the bandwidth. The rms amplitude of the signal, which is the square root of the power, must therefore be proportional to the square root of the bandwidth. This explains why a “spectral amplitude density” cannot be generally defined. Such a definition would imply that the amplitude is generally proportional to the bandwidth, which it is not (see Fig. 4.6 in Bormann and Wielandt 2013). It may nevertheless be so in special cases. Consequences of bandwidth differences of seismographs or filter

procedures applied for the evaluation of seismograms have been discussed in detail in Sect. 4.5 of Bormann and Wielandt (2013).

Approximate Conversion of Power Densities into Recording Amplitudes

According to Aki and Richards (1980), the *maximum* amplitude of a wavelet $f(t)$ near $t = 0$ can be *roughly approximated* by the product of what they call the “amplitude spectral density” $|F(\omega)|$ (see Eq. 2) and the bandwidth of the wavelet, i.e.,

$$f(t)_{t=0} = |F(\omega)|2(v_u - v_l) \quad (8)$$

with v_u and v_l being the upper and lower corner frequencies of the band-passed signal. Such an approximation is only possible for signals whose energy is maximally concentrated in time. Likewise, if the power spectral density P of noise is defined as

$$\int_{-\infty}^{\infty} f(t)^2 dt = \int_{-\infty}^{\infty} \frac{|F(\omega)|^2}{2\pi} d\omega = \int_{-\infty}^{\infty} |F(2\pi\nu)|^2 d\nu \quad (9)$$

then the *mean square amplitude* of noise in the time domain is

$$\langle f^2(t) \rangle = 2P(v_u - v_l) \quad (10)$$

This is of course a simplified version of the general rule that the power spectral density (PSD) must be integrated over the passband of a filter to obtain the power (or *mean square amplitude*) at the output of the filter. The square root of this power is then the *root mean square* (RMS) or effective amplitude

$$a_{\text{RMS}} = [2P \times (v_u - v_l)]^{1/2} \quad (11)$$

However, if the power spectral density is defined according to the engineering approach only by its positive frequencies as $P_e = 2P$, then

$$a_{\text{RMS}} = [P_e \times (v_u - v_l)]^{1/2} \quad (12)$$

Stationary signals must be characterized by their PSD, and *specifying noise by its RMS amplitudes is meaningless without definition of the bandwidth*. The values given by the NLNM and NHNM in Fig. 2 follow the engineering power (P_e) convention, and the following related formulas as well.

Often it is necessary to represent the amplitude of noise (whether RMS or average-peak) in a *constant relative bandwidth RBW* over the whole frequency range. This means that the bandwidth B around each frequency ν_0 is a fixed fraction or multiple of ν_0 . The lower and upper band limits, f_l and f_u , are chosen so that f_0 is the geometric mean of the two (i.e., ν_0 appears in the middle between ν_l and ν_u on a logarithmic frequency scale). When the relative bandwidth is n octaves or m decades (n and m normally being fractions, not integers), then the following relationships hold:

$$\begin{aligned} \nu_l &= 2^{-n/2} \nu_0, \quad \nu_u = 2^{n/2} \nu_0, \quad \mathbf{RBW} \\ &= 2^{n/2} - 2^{-n/2} \quad \text{and} \quad B \\ &= (2^{n/2} - 2^{-n/2}) \nu_0 \end{aligned} \quad (13)$$

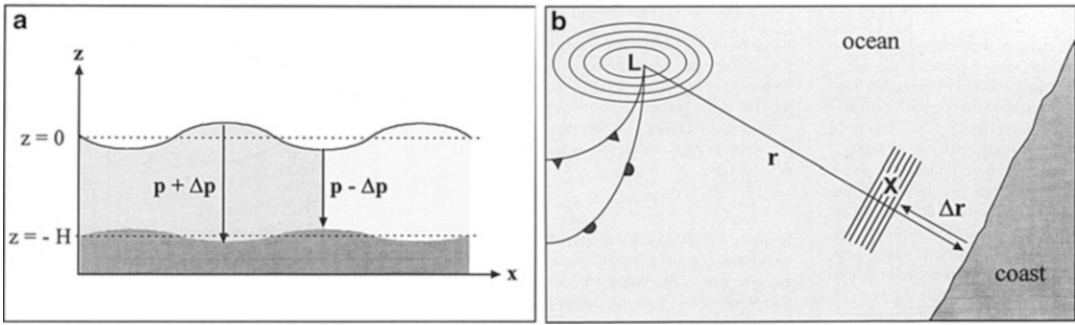
and (12) can then be written as

$$\begin{aligned} a_{\text{RMS}} &= [P_e \times \nu_0 \times (2^{n/2} - 2^{-n/2})]^{1/2} \\ &= (P_e \times \nu_0 \times \mathbf{RBW})^{1/2} \end{aligned} \quad (14)$$

Octaves can be converted easily into decades and vice versa by using the relation

$$m = n \cdot \log_{10}(2) = n \cdot 0.3010 \quad (15)$$

Typical response curves of short-period narrow-band analog seismographs for recording transient teleseismic body waves (i.e., waves that have travelled more than 2000 km either as longitudinal waves, oscillating in the propagation direction, or transverse waves, oscillating perpendicular to the direction of propagation) typically have bandwidths between 1 to 2 octaves. Different authors have used an integration bandwidth of 1/2 to 1/3 octave (a standard bandwidth in acoustics) for computing RMS amplitudes from PSD. According



Seismic Noise, Fig. 4 Schemes for the generation of (a) primary and (b) secondary microseisms (for explanations see text). L cyclone low-pressure area, X area of interference where standing waves with half the period of ocean waves develop, r distance between L and the coast, Δr

distance between the coast and X , p water pressure, Δp water pressure fluctuation proportional to wave amplitude (Copy of Fig. 12 from Friedrich et al. 1998 © Springer; with kind permission from Springer Science + Business Media)

to the above relationships, 1/3 octave RMS amplitudes will be only about 82 %, 70 %, and 50 % of the respective RMS amplitudes calculated for 1/2, 2/3, or 4/3 octave bandwidth, respectively.

There is a 95 % probability that the instantaneous amplitude of random noise with a Gaussian amplitude distribution will lie within a range of $\pm 2a_{\text{RMS}}$. Peterson (1993) showed that both broadband and long-period noise amplitudes follow closely a Gaussian probability distribution. This is true for largely natural ambient noise, often not, however, for short-period and broadband noise in urban and industrialized areas. There the noise is often dominated by transient or periodic signals (Groos and Ritter 2009 and section “Microseismic Noise in Urban Environments”). Yet, in case of a Gaussian probability distribution, the absolute peak amplitudes of the narrowband filtered signal envelopes should follow a Rayleigh distribution. In the case of an ideal Rayleigh distribution, the theoretical average peak amplitudes (APA) are $1.253 a_{\text{RMS}}$. Therefore, RMS amplitudes in 1/6 decade bandwidth correspond approximately to average peak amplitudes in 1/3 octave bandwidth.

Causes and Basic Characteristics of Ambient Noise Observed on Land

Signals which have a similar waveform and polarization, so that they can interfere

constructively, are termed coherent. This is usually the case for seismic signals generated and radiated by a common source process. The degree of coherence is defined by the ratio between the cross-correlation and the autocorrelation of the time series. The cross-correlation is squared and divided by both autocorrelations to render the result independent of the signal amplitudes. It may vary between 0 and 1. For noise, it shows distinct frequency dependence. Spatial coherence may be rather high for long-period ocean microseisms. Accordingly, the correlation radius, i.e., the longest distance between two seismographs for which the noise recorded in certain spectral ranges is still correlated, increases with the noise period. It may be several kilometers for $f < 1$ Hz but drops to just a few tens of meters or even less for $f > 50$ Hz. For random noise, it is usually not larger than a few wavelengths.

Primary and Secondary Microseisms

Primary ocean microseisms are generated only in shallow waters in coastal regions. Here the wave energy can be converted directly into seismic energy either through vertical pressure variations (Fig. 4a) or by the smashing surf on the shores. Therefore, primary ocean microseisms have the same period as the dominating water waves ($T \approx 10$ to 16 s). Haubrich et al. (1963) compared the spectra of microseisms and of swell at the beaches and could demonstrate a close

relationship between the two data sets. The relative stability in amplitude for continental sites around the globe suggests persistent, stable sources.

Contrary to the primary ocean microseisms, the secondary ocean microseisms could be explained by Longuet-Higgins (1950) and Hasselmann (1963) as being generated by the superposition of ocean waves of equal period traveling in opposite directions, thus generating standing gravity waves of half the period. These standing waves cause nonlinear pressure perturbations that propagate without attenuation to the ocean bottom and there couple into an elastic wave of much larger wavelength and higher phase velocity if the propagation directions are nearly opposite. The area of interference X (Fig. 4b) may be off-shore where the forward propagating waves, generated by a low-pressure area L, superpose with the waves traveling back after being reflected from the coast. But it may also be in the deep ocean when the waves, excited earlier on the front side of the low-pressure zone, interfere later with the waves generated on the back-side of the propagating cyclone. Thus, secondary ocean microseisms are mostly recorded with periods between 5 and 8 s and are the result of energetic ocean waves between the 10 and 16 s period. Horizontal and vertical noise amplitudes of marine microseisms are similar. The particle motion is dominantly of Rayleigh-wave type, i.e., elliptical polarization of the particle motion in the vertical propagation plane. But some energy is present also as body waves, and heterogeneous Earth structure may even couple some energy into Love waves, i.e., surface waves that oscillate parallel to the Earth surface and perpendicular to the direction of propagation with amplitudes that decay exponentially with depth below the surface.

Medium-period ocean/sea microseisms experience low attenuation. They may therefore propagate hundreds of kilometers inland. Since they are generated in relatively localized source areas, they have, when looked at from afar, despite the inherent randomness of the source process, a rather well-developed coherent part, at least in the most energetic and prominent component.

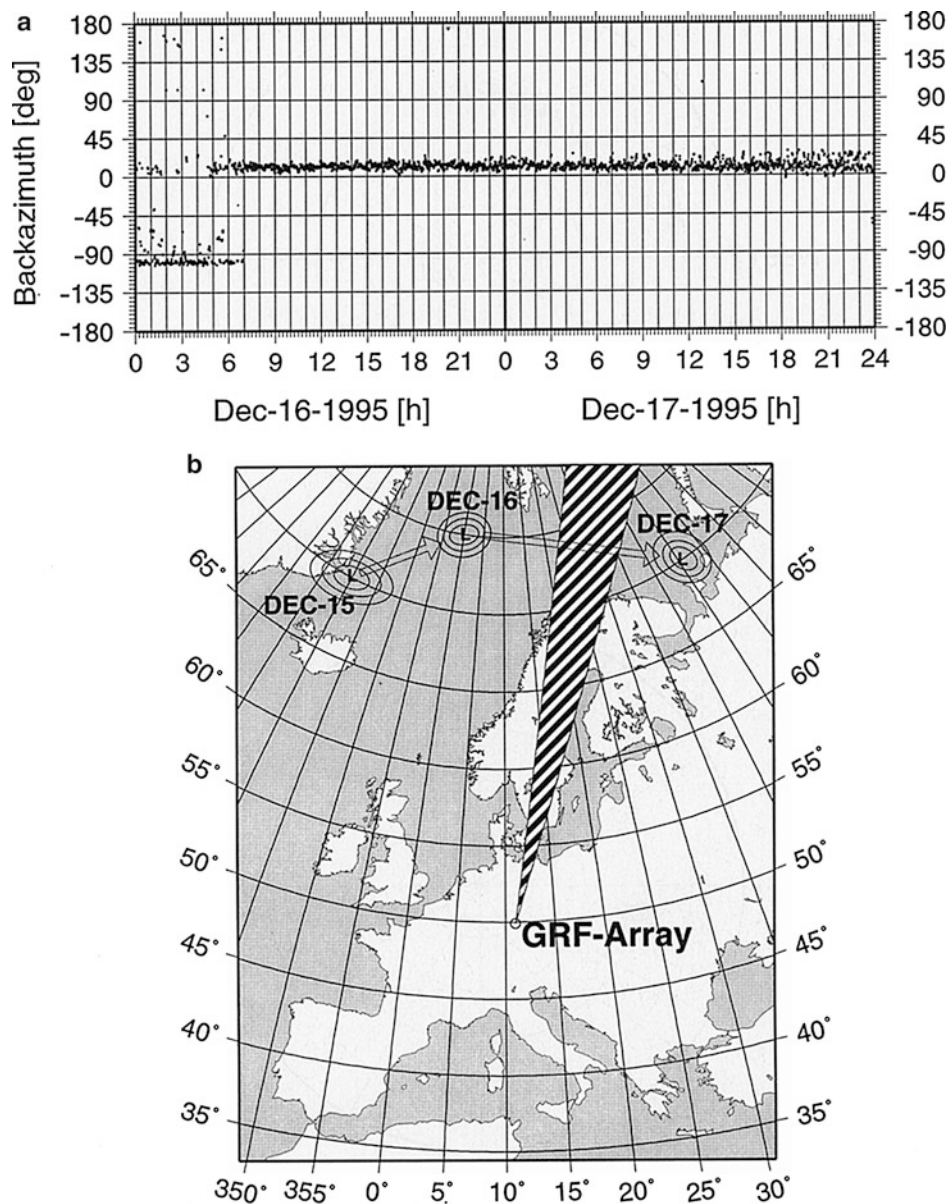
This allows to locate the source areas and track their movement by means of seismic arrays (e.g., Cessaro, 1994; Friedrich et al. 1998, Fig. 5). This possibility has already been used decades ago by some countries, e.g., in India, for tracking approaching monsoons with seismic networks under the auspices of the Indian Meteorological Survey. While near-shore areas may be the source of both primary and secondary microseisms, the pelagic sources of secondary microseisms meander within the synoptic region of peak storm wave activity.

Note that the noise peak of secondary microseisms has a shorter period when generated in shallower inland seas or lakes ($T \approx 2\text{--}4$ s) instead of in deep oceans because of generally shorter wavelengths and periods of the water waves in shallower and smaller water bodies. Also, off-shore interference patterns largely depend on coastal geometries and the latter may allow the development of internal resonance phenomena in bays, fjords, or channels (Fig. 6), which affect the detailed spectrum of microseisms. In fact, certain coastlines may be distinguished by unique “spectral fingerprints” of microseisms.

Long-Period Noise

Ground noise on land observed at frequencies between 0.2 and 50 mHz, i.e., lower than the microseism peak, is usually associated with atmospheric pressure fluctuations. At these frequencies, vertical component seismometers react to changes in gravity as the mass of the atmosphere above a site changes with atmospheric pressure. If seismometers are not fully sealed, they may also experience apparent accelerations due to buoyancy effects on the seismometer mass. Pressure fluctuations may also cause strong longperiod horizontal component noise by distorting the case of the seismometer and/or of the walls of the vault.

Long-period seismometers are also sensitive to temperature changes. Therefore, careful installations aim at sealing the sensor from both pressure and temperature fluctuations (Forbriger 2012). The vertical component noise levels between 0.2 and 1.7 mHz can be reduced by



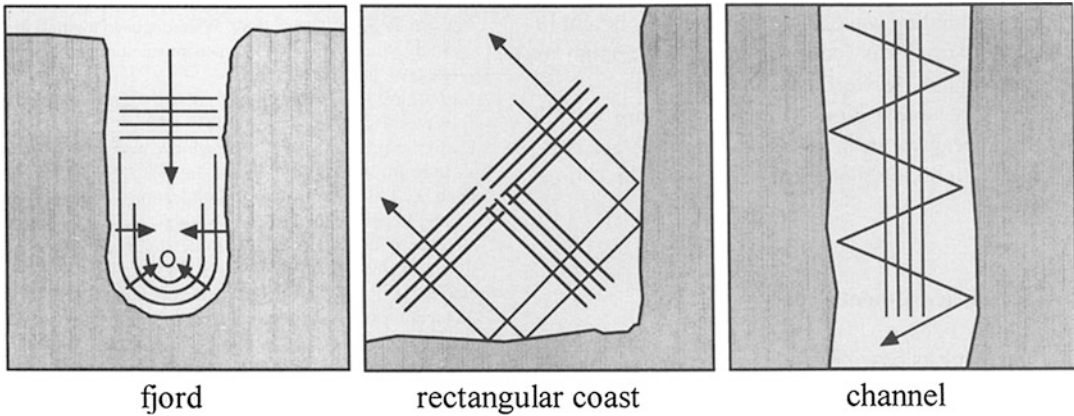
Seismic Noise, Fig. 5 An example of good spatial coherence of medium-period secondary ocean microseisms at larger distance from the source area which, in this case, allows rather reliable determination of the direction from the station towards the source (backazimut angle = angle measured clockwise from the north) by means of array analysis (see ► Seismometer Arrays). (a) Shows how the

backazimuth determination changed from one day to the next, while (b) shows the location of the two storm areas and of the GRF seismic array. Observations by at least two arrays permit localization and tracking of the noise-generating low-pressure areas (Copy of Fig. 7 from Friedrich et al. 1998 © Springer; with kind permission from Springer Science + Business Media)

more than 10 dB by subtracting the scaled, locally recorded pressure signal from the vertical acceleration record, thereby improving the SNR for normal mode observations (e.g., Fig. 2.23 in

Bormann et al. 2012 and Fig. 4.26 in Bormann and Wielandt 2013).

Turbulence in the atmospheric boundary layer produces pressure fluctuations at the Earth's



Seismic Noise, Fig. 6 Examples for coastline geometries that provide suitable interference conditions for the generation of secondary microseisms (Copy of Fig. 13

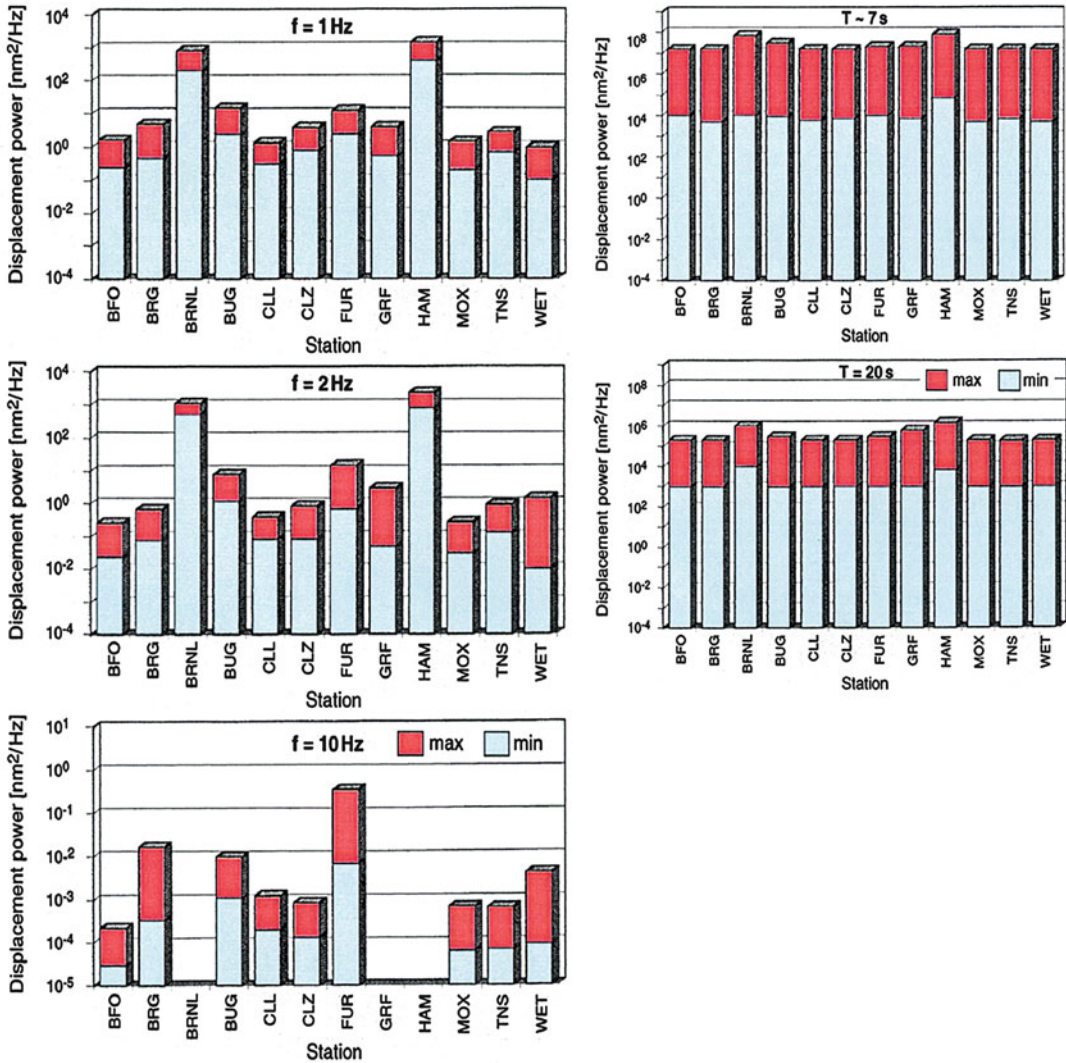
from Friedrich et al. 1998 © Springer; with kind permission from Springer Science+Business Media)

surface that cause significant deformation to depths of a few tens of meters below the surface. This noise source is a primary reason why broadband seismometers at permanent stations are often installed beneath the Earth's surface in shallow boreholes (see Sect. 7.4.5 and Fig. 7.59 in Trnkoczy et al. (2011)). For instruments on the surface, short-wavelength atmospheric pressure fluctuations produce a large tilt signal that makes the horizontal components particularly noisy. Tilt couples gravity into the horizontal components but much less into the vertical component (Chap. 5, Sect. 5.3.3 and Figs. 5.8 and 5.9 in Wielandt 2011). The ratio increases with the period and may reach a factor of up to 300 (about 50 dB). Horizontal component local tilt noise may also be caused by traffic and wind pressure.

At very long periods, deformation of the ground, of the seismometer vault, or of the building due to insolation gradients (uneven heating by sunshine) may become noticeable as well. Generally, recording sites poorly coupled to the ground or on poorly consolidated ground will tend to be noisier, both at long and short periods. Other reasons for increased long-period noise may be air circulation in the seismometer vault itself or underneath the sensor cover. Special care in seismometer installation and shielding is therefore required in order to reduce drifts and long-period environmental noise (Wielandt 2011; Trnkoczy et al. 2011; Forbriger 2012).

Short-Period Noise (0.5–50 Hz)

Ambient noise at frequencies higher than those associated with the microseisms peak may have natural causes such as wind (wind friction over rough terrain, trees, and other vegetation or built-up objects swinging or vibrating in the wind), rushing waters (waterfalls or rapids in rivers and creeks), etc. Wind-generated noise is broadband, ranging from about 0.5 Hz up to about 15–60 Hz. But the dominant sources of high-frequency noise are man-made (power plants, factories, rotating or hammering machinery, road and rail traffic, etc.; see Sects. 7.1 and 7.2 in Trnkoczy et al. 2011). Wind-generated noise couples mostly into *surface-wave* modes; cultural noise, however, couples at least partly into *body waves* that can propagate also to greater depth. Cultural noise in records is in principle avoidable, although it is often impractical to place seismic station sites at sufficient distances away from cities or highways. Some rules of thumb suggest that it may be necessary to site short-period recording stations of high magnification as far as 25 km from power plants or rock-crushing machinery, 15 km from railways, 6 km from highways, and 1 km or more from smaller roads. And with respect to moving waters, it is recommended to place stations away from moving water between some 60 km for very large waterfalls and dams to 15 km for smaller rapids (Table 2.1 in Willmore 1979, <http://www.seismo>.



Seismic Noise, Fig. 7 Comparison of the minimum and maximum levels of short-period and long-period ambient noise power observed at stations of the German Regional

Seismic Network (GRSN) (Color version of Fig. 9 from Bormann et al. 1997 © Springer; with kind permission from Springer Science + Business Media)

com/msop/msop79/sta/tab_2.1.gif or Table 7.5 in Trnkoczy et al. 2011). However, noise travels further across competent rock than through alluvial filled valleys. Therefore, “safe” distances may be from 1/2 to 2/3 of the values above, depending on propagation path.

Most of the short-period natural and man-made sources are distributed, stationary, or moving. Their contributions, coming from various directions, superpose to a rather complex, more or less stationary random noise field.

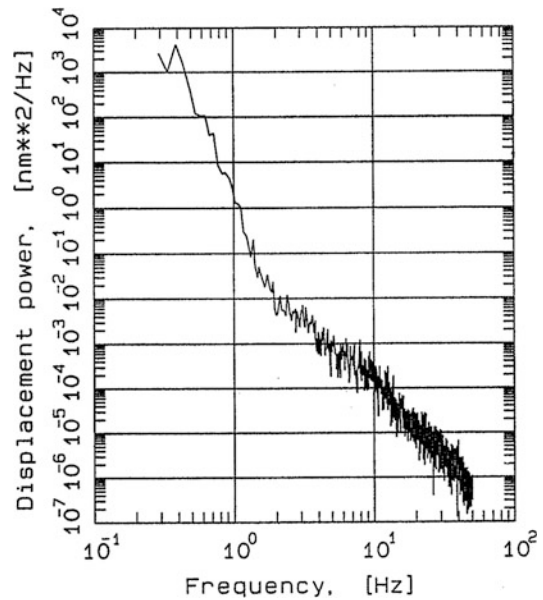
The particle motion of short-period noise is therefore more erratic than for long-period ocean noise and may vary more (by about 30 dB) than long-period noise (by about 10 dB) from station to station within a seismic network. Very pronounced is the seasonal variability of the second-order ocean microseisms by about 30 dB (Fig. 7).

Nevertheless, polarization analysis (the investigation of the different oscillation directions of the various types of seismic waves in space), averaged over moving time-windows, sometimes

reveals preferred azimuths of the main axis of horizontal particle motion, which hints at localized noise sources. Moreover, when the vertical component particle motion in three-component records is clearly developed and averaged, then this indicates the presence of fundamental Rayleigh-wave type polarization. A rather popular and cost-effective microzonation method is based on this assumption. It derives information about the fundamental resonant frequency of the soft-soil cover and estimates local site amplification of ground motion from the peak in the horizontal to vertical component spectral noise ratio (H/V method, e.g., Parolai 2012).

According to the above, every recording site has different noise characteristics depending on its distance from the ocean, industrial and settlement areas, major infrastructure facilities, the wind climate of the site, and the depth of burial of the sensor. At some locations, a seasonal cycle is evident due to variations in wind or water flow in nearby rivers. For example, according to Fyen (1990), spring runoff raises noise levels between 0.5 and 15 Hz at the NORESS array by up to 15 dB. In settlement and industrial areas, a diurnal cycle may dominate due to variable human activity during the day (Fig. 12).

High-frequency noise spectra on land during quiet intervals resemble each other at quieter sites. The rather featureless displacement power spectrum follows a power-law dependence in frequency, proportional to $\approx f^{-2}$ (Fig. 8). This spectral appearance changes drastically due to the strong impact of cultural noise, especially in the frequency band between 1 and 10 Hz (Figs. 13 and 14). Wind noise depends on the strength of the wind and the character of the site. Recorded noise power near 1 Hz in a surface vault may increase by 10–20 dB, as compared to records on calm days, when wind speed reaches about 5 m/s. Short-period wind noise becomes detectable in records at surface sites far from cultural sources at wind velocities above about 3 m/s and at 4 m/s for subsurface sensors as well (Fig. 9). The primary mechanism by which wind couples into high-frequency noise is probably by the direct action of the wind on trees, bushes, and other structures, although at lower frequency the



Seismic Noise, Fig. 8 Displacement power spectral density calculated from 6 moving windows, 50% overlapping, of short-period noise records, 4,096 samples long each, i.e., from a total record length of about 80 s at a rather quiet site in NW Iran (Copy of Fig. 4.6 in Bormann (2002) © IASPEI)

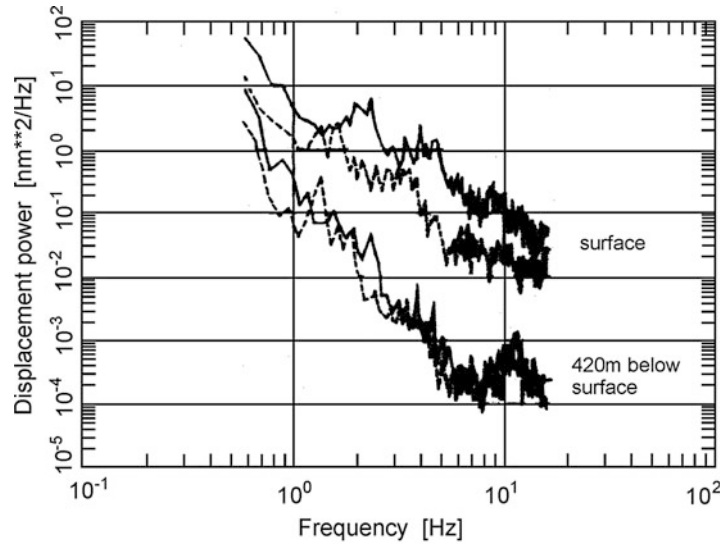
pressure fluctuations associated with wind can directly drive motions of the ground.

Generally, noise levels increase with higher wind speeds. However, there is apparently no linear relationship between wind speed and the amplitude of wind-generated noise. Wind noise increases dramatically at wind speeds greater than 3–4 m/s and may reach down to several hundred meters depth below the surface at wind speeds > 8 m/s (Young et al. 1996). But generally, the level and variability of wind noise is much higher at or near the surface and is reduced significantly with depth (Fig. 9). Moreover, at wind speeds below 3–4 m/s, the background noise may be omnidirectional with relatively large coherence length at frequencies below 15 Hz, but the coherence length is strongly reduced at higher wind speeds with increased air turbulence and local wind pressure fluctuations (Withers et al. 1996).

Comparisons of noise spectra from surface vaults and subsurface installations in mines (Fig. 9) or boreholes (Fig. 10) have repeatedly

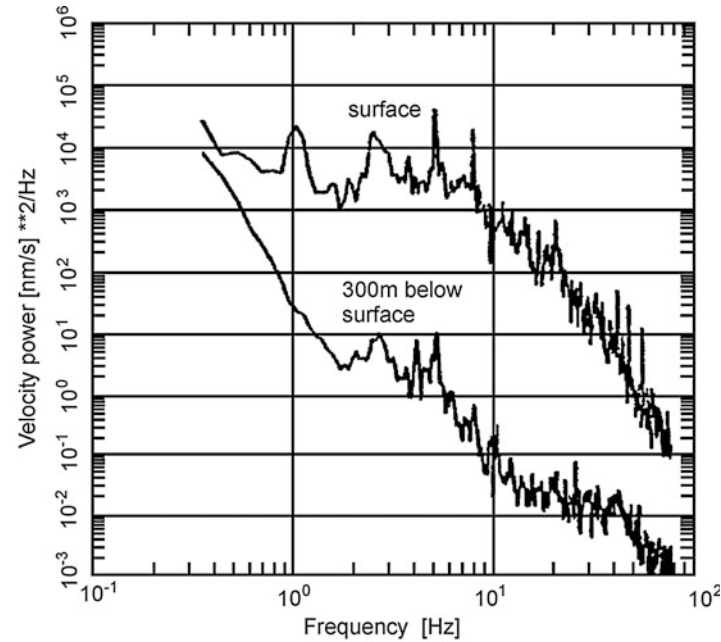
Seismic Noise,

Fig. 9 Displacement power noise spectra measured at the surface (*upper curves*) and at 420 m below the surface in a disbanded salt mine at Morsleben, Germany (*lower curves*), on a very quiet day (*hatched lines*) and on a day with light wind on the surface (wind speed about 4 m/s; *full lines*) (Copy of Fig. 4.24 in Bormann (2002) © IASPEI)



Seismic Noise,

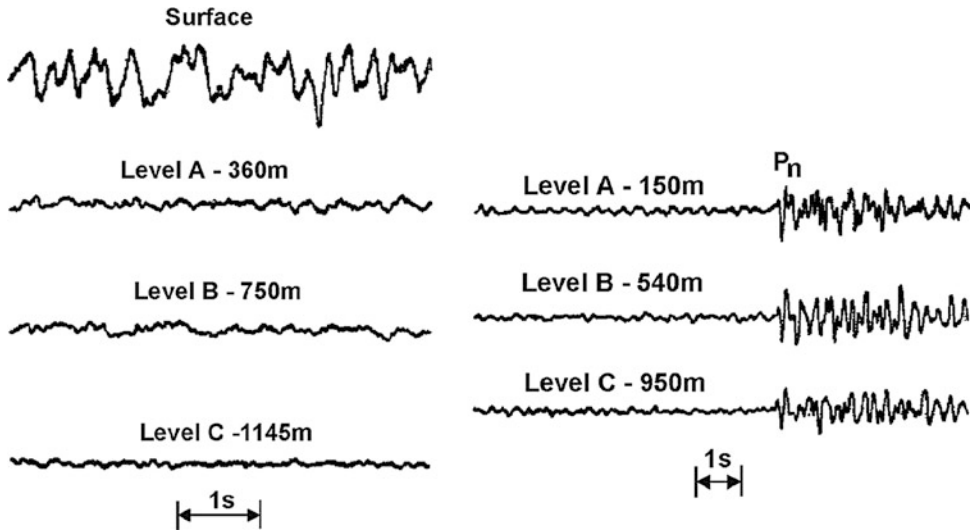
Fig. 10 Velocity power density spectra as obtained for noise records at the surface (*top*) and at 300 m depth in a borehole (*below*) near Gorleben, Germany (Figure by courtesy of M. Henger; copy of Fig. 4.23 in Bormann (2002) © IASPEI)



shown significant improvements in SNR for high-frequency phases recorded at depths, sometimes even as shallow as a few meters. Large reductions in noise level may occur already in the first 5 m between the weathered rock of the surface and the more competent rock in the borehole. Note, however, that for periods longer than 2 s the noise reduction is only marginal, even for

sensor deployments at several hundred meters depth.

In case of Fig. 10, the noise power at 300 m depth was reduced, as compared to the surface, by about 10 dB, at $f = 0.5$ Hz, 20 dB at 1 Hz and 35 dB at 10 Hz. The surface-wave nature of ambient noise (including ocean noise) is the reason for the exponential and frequency-dependent



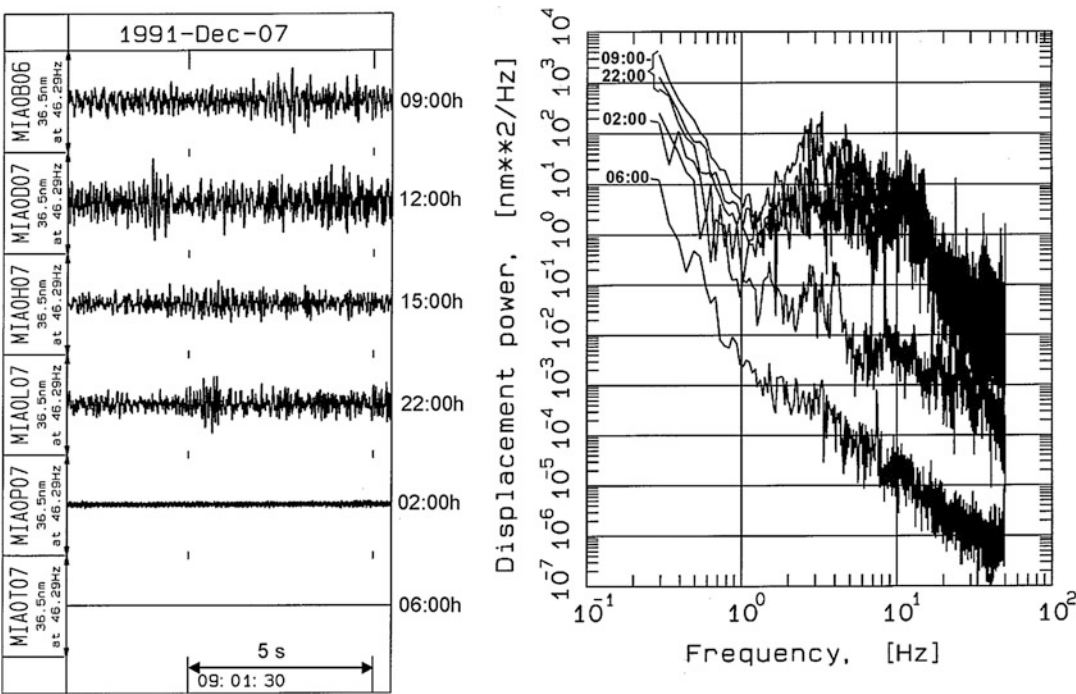
Seismic Noise, Fig. 11 Records of short-period noise (left) and signals (right) at the surface and at different depth levels of a *downhole seismometer array*. Note the significantly larger Pn amplitude closer to the surface. It is

due to the surface amplification effect which may reach a factor of two for vertical incident waves but is less for incidence angles $\gg 0^\circ$ (Fig. copied from Bormann, 1966; modified from Broding et al. 1964)

decay of noise amplitudes with depth. This is not the case for body waves (Fig. 11). Since the penetration depth of surface waves increases with wavelength, high-frequency noise attenuates more rapidly with depth. But both noise reduction and signal behavior with depth depend also on local geological conditions (e.g., section “[Installations in Subsurface Mines, Tunnels, and Boreholes](#)”). Because of the surface-wave character of short- and medium-period noise, the horizontal propagation velocity of ambient microseismic noise depends on frequency. It is close to the shear-wave velocity in the uppermost crustal layers, which is about 2.5–3.5 km/s for outcropping hard rock and about 300–650 m/s for unconsolidated sedimentary cover. This is rather different from the apparent horizontal propagation velocity of P waves and all other steeply emerging teleseismic body-wave onsets.

Broadband spectral levels fall also with increasing borehole depth, but narrow spectral lines due to cultural sources and or waveguide trapping may show less depth dependence. This may result in very different noise decay and signal-to-noise ratio behavior with depth (see section “[Installations in Subsurface Mines, Tunnels, and Boreholes](#)”).

Cultural noise may vary strongly, up to about 30 or even 50 dB at some higher frequencies between sites in different environments (Fig. 14) but also from day to day or with the time of the day at a given site in a busy urban environment (e.g., Fig. 12). Industrial and traffic noise levels are usually lowest on weekends but increase during holidays because of higher traffic. Sources such as power plants, transformer stations, etc. generate narrowband noise, producing energetic spectral lines at 50 or 60 Hz as well as related harmonics and subharmonics (e.g., at 30 Hz, 25 Hz, 12.5 Hz) depending on location (Fig. 13). The NORESS array data included lines at 6, 12, and 17 Hz due to reciprocating saws in a nearby sawmill, and records at Moxa (MOX) station in Germany have been spoiled over decades by a pronounced 2 Hz peak due to a heavy steam engine 16 km away driving a steel-press for rails. It may be possible to remove by frequency filtering such narrowband noise at frequencies that differ from the dominating signal frequency (see section “[Frequency Filtering](#)”). Other types of machinery, however, may also produce narrowband peaks in the noise spectra, but with time-variable frequencies as motor speed may vary. These noise sources are more difficult to remove.



Seismic Noise, Fig. 12 Comparison of relatively quiet sections of vertical component noise records (*left*, without strong transients) and related power spectra (*right*) at a reference site in the district capital town of Miane in NW Iran. The measurements were made at different times of the day (Copy of Fig. 7.15 from Bormann in Trnkoczy et al. (2002) © IASPEI)

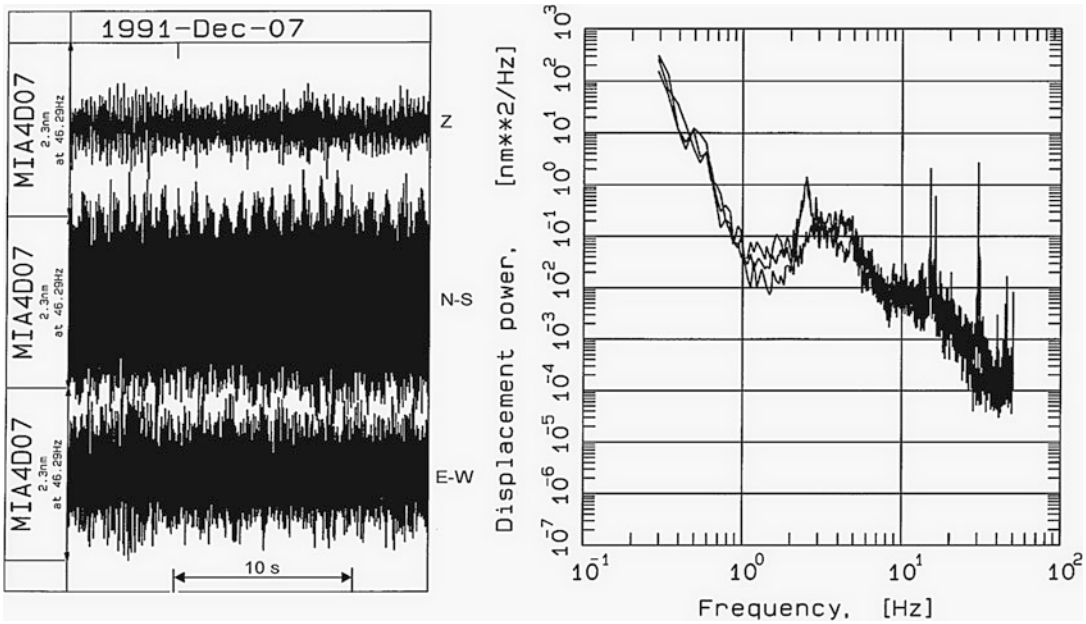
Microseismic Noise in Urban Environments

In urban, densely populated, and industrialized environment both natural and cultural noise components may play a major role in a wide range of frequencies. Measuring and analyzing urban seismic noise (USN) requires broadband recordings and data analysis. Figure 14 compares the noise power density spectrum measured at the former seismological station in the city of Hamburg, Germany, with the spectrum measured on a hard rock site in a quiet Spa village some 70 km away, which is now the alternative station site in the GRSN.

From Fig. 14, the following conclusions can be drawn: While in a remote quiet urban setting the microseismic noise spectrum may decay at night almost monotonously towards higher frequencies, the cultural noise activity may increase the daytime PSD for frequencies above 1 Hz by some 10 dB. In contrast, in a very busy

industrialized city such as Hamburg, the microseismic noise level may be 20–40 dB higher for frequencies of about 0.5–30 Hz and may be reduced during night time by less than 10 dB only between 1 Hz and 10 Hz. In the metropolitan city of Berlin, the “cultural noise” patterns differ from that of HAM (see Fig. 7.31 in Trnkoczy et al. 2011) and may again be different in towns with more pronounced diurnal variability of human and industrial activities than in megacities, as in Fig. 12.

Groos and Ritter (2009) stated that the generally high variability of USN does not allow to characterize a sample time series of USN comprehensively by single measures such as the standard deviation of a noise series or the PSD at a given frequency. For the metropolitan city of Bucharest, Romania, they calculated therefore long-term spectrograms of up to 28 days duration from broadband seismic



Seismic Noise, Fig. 13 Noise records and related power spectra near to a transformer house and power line. Note the monochromatic spectral lines around 13, 30, and 50–60 Hz, either induced by the AC current frequency

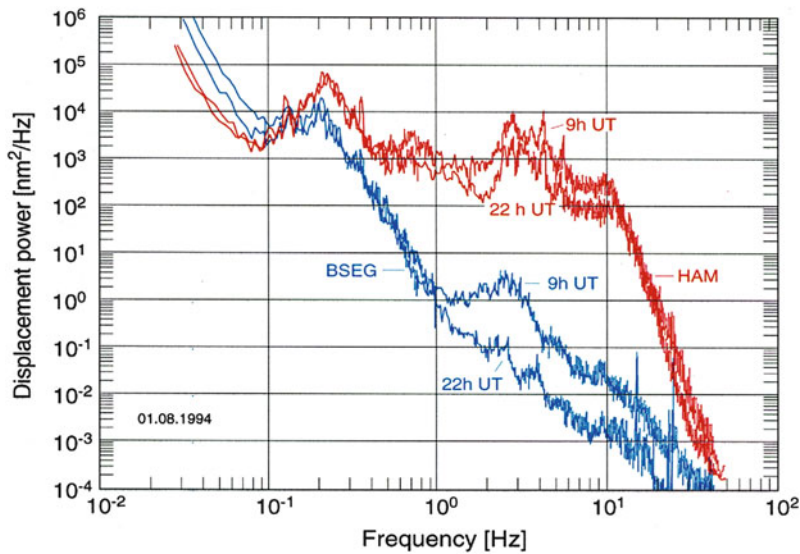
and its lower harmonics and/or caused by the vibration of the transformer (Copy of Fig. 7.19 from Bormann in Trnkoczy et al. (2002) © IASPEI)

recordings. Spectrograms, or sonograms, are visual 2D representations of the spectrum of frequencies contained in a seismic, sound, or other signal as they vary with time or with some other variable. The aim was to identify the frequency-dependent behavior of the time-variable processes that contribute to USN. Based on the spectral analysis of the data in eight frequency ranges between 8 mHz and 45 Hz, they proposed a time-domain classification which allows to identify both Gaussian distributed noise time series as well as time series that are dominated by transient or periodic signals due to traffic, rotating machinery, etc. In the case of Bucharest, only some 40 % of the analyzed time series are characterized by Gaussian distributed random noise. The most common deviations are due to large-amplitude transient signals. Moreover, significant variations of the statistical noise properties during daytime were found in the wide frequency range between 0.04 and 45 Hz, pointing to a broadband human influence on USN. The authors recommend the use of such automatically derived broadband

information about ambient noise for selecting suitable time windows for H/V noise studies (Parolai 2012) or ambient noise tomography.

Signal and SNR Variations due to Local Site Conditions

Compared to hard rock sites, both noise and signals may be amplified when recorded on soft soil cover. This signal amplification may partly or even fully outweigh the higher noise observed on such sites. Signal strength observed for a given event may vary strongly (up to a factor of about 10–30) within a given array or station network, even if its aperture is much smaller than the epicentral distance to the event (<10–20%), so that differences in backazimuth and amplitude-distance relationship are negligible (Figure 15 and Fig. 4.36 in Bormann and Wielandt 2013). Also, while one station of a network may record events rather weakly from a certain source area, the same station may do as well as other stations



Seismic Noise, Fig. 14 Noise power spectra at HAM (upper two curves, red) and BSEG (lower two curves, blue) determined from vertical-component records at day-time and nighttime. Note that the noise reduction is almost negligible for periods longer than 2 s. The drastic reduction at higher frequencies has been achieved both by the reduced cultural noise sources at Bad Segeberg, as compared to the big city of Hamburg, and by installation of the broadband seismograph in an outcropping local hardrock

salt-anhydrite diapir with very high contrast of its *acoustic impedance* (= product of wave velocity in the medium times density of the medium) against that of the surrounding alluvial softrock layers. This prevents much of the short-wavelength noise energy to penetrate into this local hard rock anomaly (Color version of Fig. 18 from Bormann et al. 1997 © Springer; with kind permission from Springer Science + Business Media)

(or even better) for events from another region, azimuth, or distance (e.g., station GWS in Fig. 16 left and right, respectively).

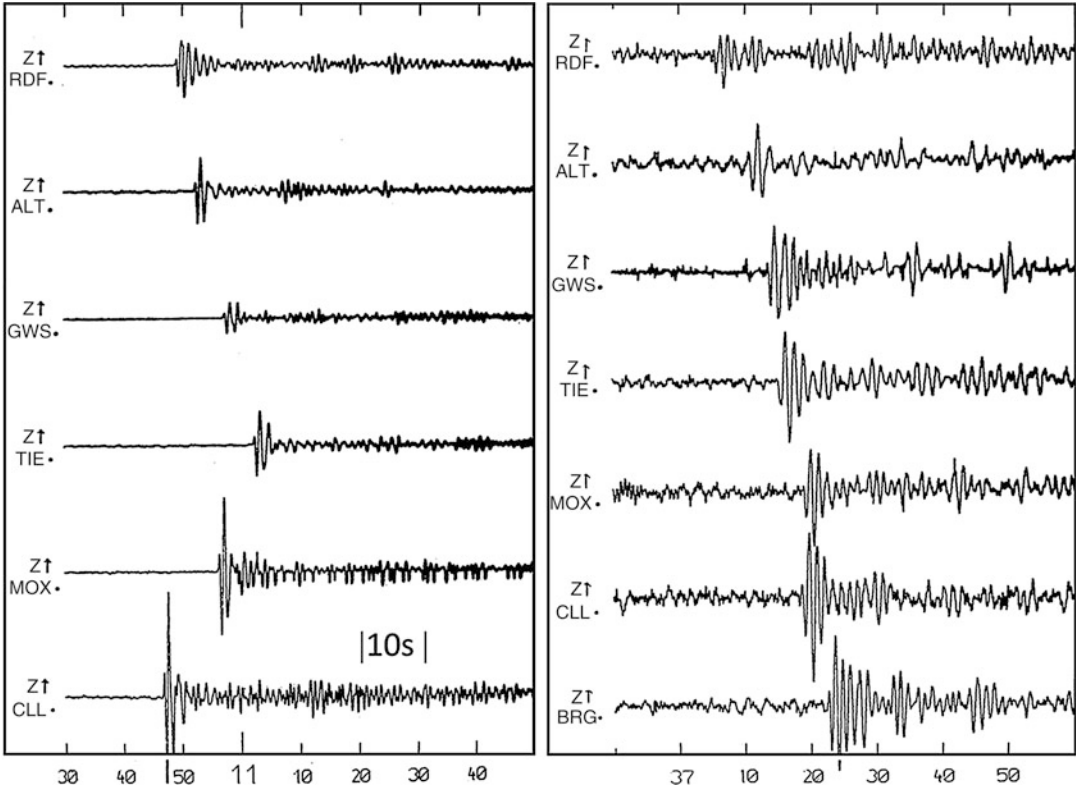
Figure 16 compares, for regional and teleseismic events, the short-period P-wave amplitude ratio (left) and SNR (right) of two stations of the German Regional Seismic Network (GRSN). In the same azimuth range, but at different epicentral distances, station BRG may record both > 3 times larger as well as > 3 times smaller amplitudes than station MOX. This corresponds to differences in station magnitude up to one unit! The same applies to the differences in SNR. Therefore, optimal site selection cannot be made only on the basis of noise measurements. Also, the signal conditions at possible alternative sites should be compared.

However, differences in local SNR conditions may become negligible in long-period recordings (Fig. 4.39 in Bormann and Wielandt 2013) and thus play a lesser role in site selection for broadband records. Thermal shielding and reduction of

tilt noise then have the highest priority (see Forbriger 2012).

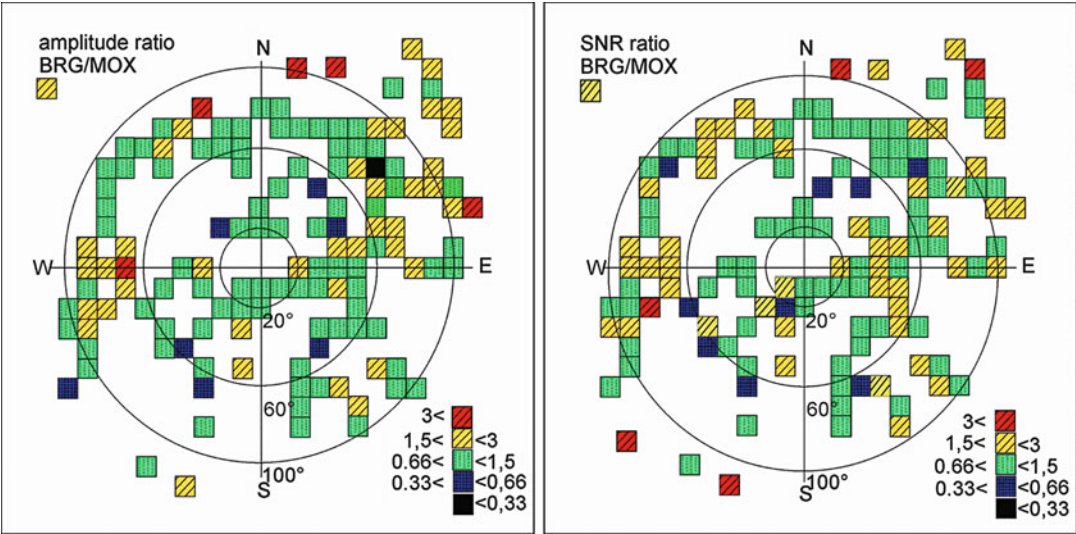
Installations in Subsurface Mines, Tunnels, and Boreholes

As shown in Figs 13–15, short-period microseismic noise is strongly reduced with the depth of sensor installations in mines and boreholes. Withers et al. (1996) found that for rather high frequencies monitoring, as in very local weak motion and industrial seismology measurements between 10 and 20 Hz, the SNR could be improved between 10 and 20 dB and for frequencies between 23 and 55 Hz as much as 20–40 dB by deploying a short-period sensor at only 43 m below the surface. However, depending on geological conditions and related peculiarities of the local noise field as well as the decreasing free-surface amplification effect of incident body waves, the SNR does not necessarily increase steadily with depth. A striking example have been short-period signal and noise measurements



Seismic Noise, Fig. 15 Short-period records of underground nuclear explosions at the test sites of Semipalatinsk (*left*, $D \approx 41^\circ \pm 1^\circ$) and Nevada (*right*, $D \approx 81^\circ \pm 1^\circ$) at stations of the former East German seismic network. Note

the differences in signal amplitudes both amongst the stations for a given event and for the same station, when comparing events in different backazimuth and distance (Copy of Fig. 4.35 in Bormann (2002) © IASPEI)



Seismic Noise, Fig. 16 Pattern of the relative short-period P-wave amplitudes (*left*) and of the related SNR (*right*) at station BRG normalized to those of station MOX

(170 km apart) in a distance-azimuth polar diagram (Modified version of Fig. 7 from Bormann et al. 1992 © Elsevier; with permission from Elsevier)

in the early 1960s with 1 Hz borehole seismometers in two abounded deep oil drills in the USA (Douze, 1964). In the Texas hole, the average ratio of seismic signal in the borehole to that measured at the surface decreased to 1/10th at about 1,500 m depth and then increased again to ½ of the surface value at 3,000 m depth. In contrast, in an Oklahoma borehole, it dropped to about 1/3 at about 1,000 m depth and then remained roughly constant thereafter. Accordingly, despite the noise reduction with depth, there was no SNR improvement (on average) in the Texas borehole down to about 1,000 m depth, but then the SNR increased to a factor of about 15 at 3,000 m depth. Contrary to this, the SNR increased by a factor of 3 in the Oklahoma borehole within the first 800 m but then remained roughly constant (ranging between 1 and 5) up to 3,000 m depth (see Fig. 4.40 in Bormann and Wielandt 2013).

Nevertheless, generally a significant SNR improvement is to be expected within the first few hundred meters depth. This applies particularly to borehole installations of long-period and broadband sensors which benefit already at much shallower depth greatly from the very stable temperature conditions and strongly reduced atmospheric pressure fluctuations and related tilt noise, also in hard rock tunnels. In mines or boreholes, a depth of 100 m below surface is generally sufficient to achieve most of the practicable reduction by -20 to -30 dB of long-

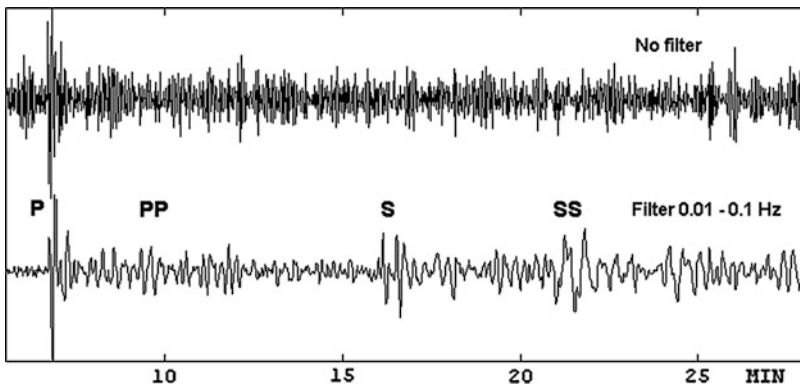
period noise with periods between 30 s and 1,000 s (see Fig. 7.59 in Sect. 7.4.5 of Trnkoczy et al. 2011). Since the cost of drilling and installation increases greatly with depth, no permanent seismic borehole installations deeper than 100–150 m have yet been made. In any event, the borehole should be drilled through the soil or cover of weathered rock and penetrate well into the compacted underlying rock formations. For more information about tunnel and long-period borehole installations, see Sects. 7.4.3 and 7.3.5 in Trnkoczy et al. 2011.

Thus, differences in the frequency spectrum, horizontal wave-propagation velocity, degree of coherence, and depth dependence between (short- and medium-period) microseismic noise and seismic waves allows to improve the signal-to-noise-ratio (SNR) by installing seismic sensors either at reasonable depth below the surface or by way of data processing.

Improving the Signal-to-Noise Ratio by Data Processing

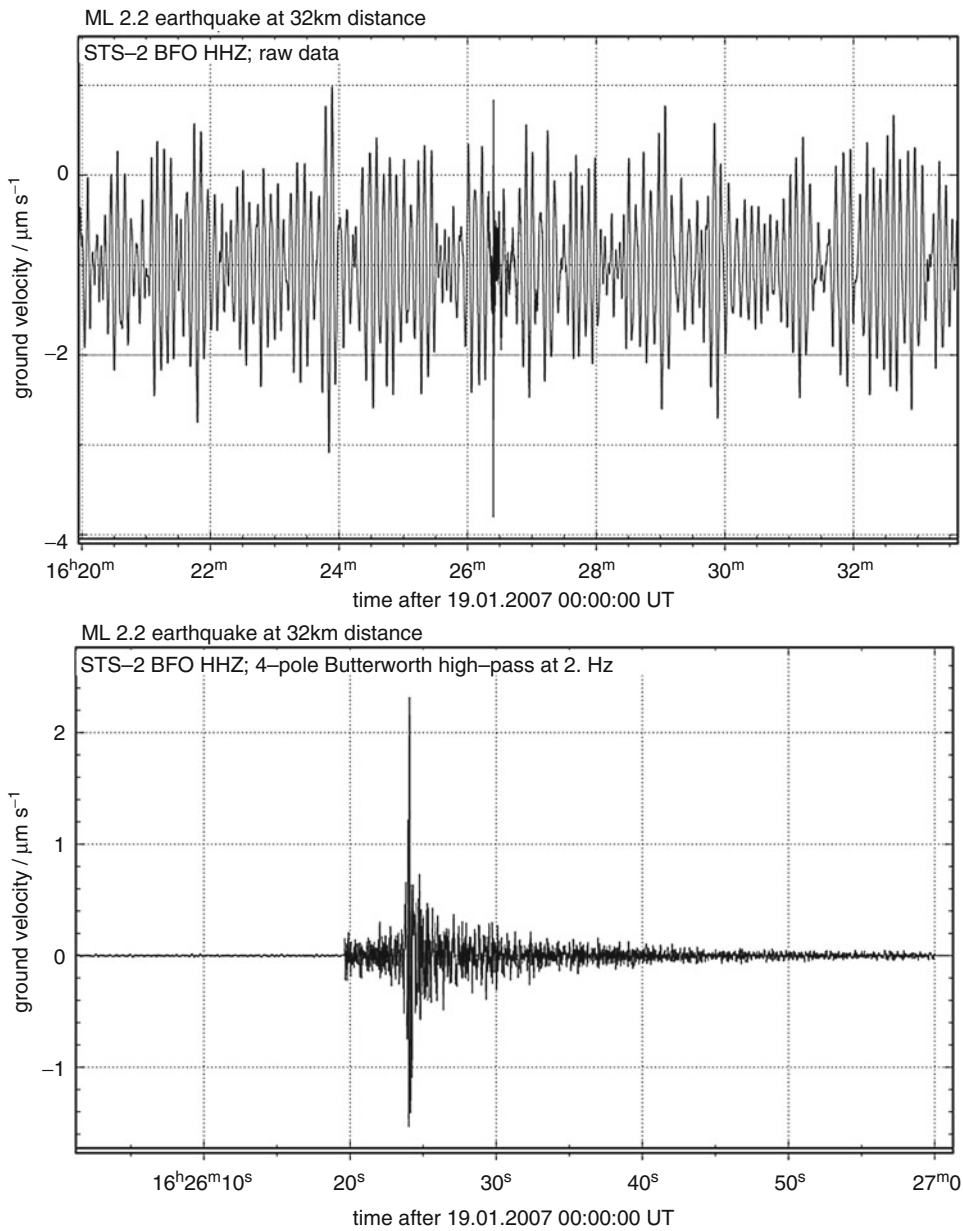
Frequency Filtering

If the dominating frequency content of the seismic signal differs from that of the disturbing noise, then band-pass, high-pass, or low-pass filtering may significantly improve the SNR (Figs. 17 and 18).



Seismic Noise, Fig. 17 *Top:* Velocity broadband record in the period range between 0.5 and 360 s of a teleseismic earthquake. Only the P-phase might be identified. *Bottom:* On the long-period band-pass filtered trace the signal-to-

noise ratio is much improved and several later phases are clearly recognizable (Figure by courtesy of J. Havskov for NMSOP editions © IASPEI)



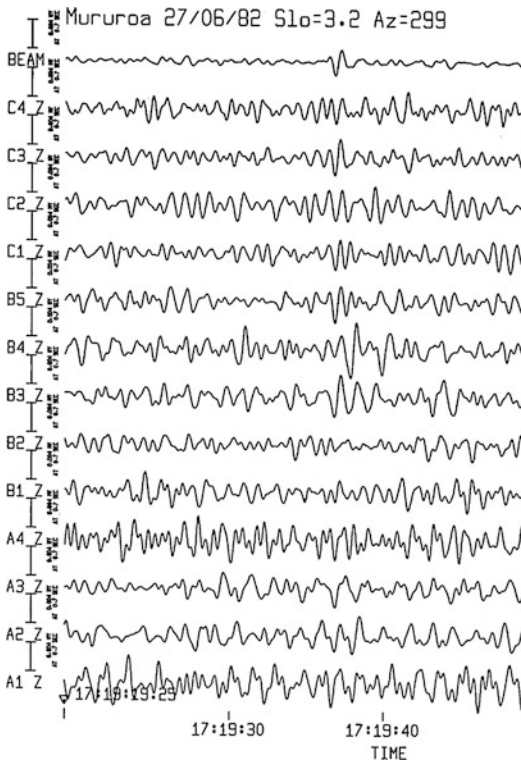
Seismic Noise, Fig. 18 Isolation of a local $ML = 2.2$ earthquake (*bottom trace*) that occurred 32 km away from the Black Forest Observatory (BFO), Germany, by 2-Hz high-pass filtering of the raw data of a velocity broadband

STS-2 record (*top trace*), thus eliminating the dominating 6 s ocean microseisms (Figure by courtesy of Thomas Forbriger for NMSOP Chapter 4 © IASPEI)

Velocity Filtering and Beamforming

Often the dominant signal frequencies may coincide with that of strong noise. Then frequency filtering does not help to improve the SNR. On the other hand, the horizontal propagation

velocity of noise, being dominantly surface waves of Rayleigh-wave type, is much lower than that of P waves and also lower than that of teleseismic S waves with a steep angle of incidence. Then frequency-wavenumber (f-k)



Seismic Noise, Fig. 19 Detection of a weak underground nuclear explosion in the 10 kt range at the Mururoa Atoll test site ($D = 145^\circ$) by beam forming (*top trace*). No signal is recognizable in any of the 13 individual record traces from stations of the Gräfenberg array, Germany (*below*) (From Buttke 1986 © Schweizerbart <http://www.schweizerbart.de/>)

filtering is suitable to improve the SNR in the beam trace of seismic array records (Schweitzer et al. 2011). Assuming that the noise within the array is random while the signal is coherent, even a simple direct summation of the n sensor outputs would already produce some modest SNR improvement. When the direction and velocity of travel of a signal through an array is known, it is possible to compensate for the differences in arrival time at the individual sensors and then sum up all the n record traces (beam forming). This increases the signal amplitude by a factor n while the random noise amplitudes increase in the beam trace only by $\sqrt{n}$, thus improving the SNR by $\sqrt{n}$. Figure 19 compares the (normalized) individual records of 13 stations of the Gräfenberg array, Germany, with the beam

trace. A weak underground nuclear explosion at a distance of 143.6° , not recognizable in any of the single traces, is very evident in the beam trace.

Noise Prediction Error, Polarization, and Other Multi-parameter Methods of Filtering

In near real time, it is possible to use a moving time-window to determine the characteristics of a given noise field by means of cross- and autocorrelation of array sensor outputs. This allows the prediction of the expected random noise in a subsequent time interval. Subtracting the predicted noise time series from the actual record results in a much reduced noise level. Weak seismic signals, originally buried in the noise but not predicted by the noise “forecast” of the prediction-error filter (NPEF), may then stand out clearly. NPEFs have several advantages as compared to frequency filtering:

- No assumptions on the frequency spectrum of noise are required since actual noise properties are determined by the correlation of array sensor outputs.
- While frequency differences between signal and noise are lost in narrowband filtering, they are largely preserved in the case of the NPEF. This may aid signal identification and onset-time picking.
- Signal first-motion polarity is preserved in the NPEF whereas it is no longer certain after narrow-band or zero-phase band-pass filtering.

Three-component recordings allow to reconstruct the ground particle motion and to determine its polarization which may differ between different types of seismic waves and noise. One great advantage of polarization filtering is that it is independent of differences in the frequency and velocity spectrum of signal and noise and thus can be applied in concert with other procedures for SNR improvement.

Early examples of NPEF and polarization filtering have been given in Bormann and Wieland (2013). For a briefing on modern three-component broadband procedures of SNR

improvement, signal detection, phase identification, and event location that have been implemented in procedures for fully automatic near real-time multi-parameter signal analysis, see Sect. 11.9 in Bormann et al. (2014). These procedures apply cross-correlation methods, use differences in the complexity and polarization properties between signal and noise as well as a compressed signal representation by a sequence of n-dimensional feature vectors which increase the robustness of the procedures. Some of these versions also adapt the signal detector to temporal/seasonal variations in the nature and power of ambient noise.

Summary

The essay defines the term of “seismic noise” and explains why coherent seismic signals and incoherent random noise are not commensurate, thus requiring different mathematical treatment and analysis. Also described are the different causes and ways of analysis of ambient microseismic noise as recorded at stations on land in the very wide frequency range between some mHz and 100 Hz. The article sketches some essential procedures of noise reduction by way of proper sensor shielding and installation, also at depth, as well as feature and problem adapted filtering aimed at improved signal-to-noise ratio and thus signal detection and identification. Many references for further readings are given, amongst them to relevant IASPEI Manual articles accessible via <http://nmsop.gfz-potsdam.de>.

Cross-References

- [Downhole Seismometers](#)
- [Noise-Based Seismic Imaging and Monitoring of Volcanoes](#)
- [Ocean-Bottom Seismometer](#)
- [Seismic Event Detection](#)
- [Seismic Instrument Response, Correction for](#)
- [Seismic Network and Data Quality](#)
- [Seismometer Self-Noise and Measuring Methods](#)

- [Seismometer Arrays](#)
- [Sensors, Calibration of](#)
- [Site Response: 1-D Time Domain Analyses](#)
- [Site Response: Comparison Between Theory and Observation](#)
- [Stochastic Ground Motion Simulation](#)

References

- Aki K, Richards PG (1980) Quantitative seismology – theory and methods, vol II. Freeman, San Francisco
- Bormann P (1998) Conversion and comparability of data presentations on seismic background noise. *J Seismol* 2(1):37–45
- Bormann P (2002) Chapter 4: Seismic signals and noise. In: Bormann P (ed) IASPEI new manual of seismological observatory practice (NMSOP), vol 1. GeoForschungsZentrum, Potsdam, 34 pp
- Bormann P, Wielandt E (2013) Chapter 4: Seismic signals and noise. In: Bormann P (ed) New manual of seismological observatory practice (NMSOP-2) (2012). IASPEI, GFZ German Research Centre for Geosciences, Potsdam, 63 pp. doi:10.2312/GFZ.NMSOP-2_CH4. <http://nmsop.gfz-potsdam.de>
- Bormann P, Wylegalla K, Strauch W, Baumbach M (1992) Potsdam seismological station network: processing facilities, noise conditions, detection threshold and localization accuracy. *Phys Earth Planet Int* 69:311–321
- Bormann P, Wylegalla K, Klinge K (1997) Analysis of broadband seismic noise at the German Regional Seismic Network and search for improved alternative station sites. *J Seismol* 1:357–381
- Bormann P, Engdahl ER, Kind R (2012) Chapter 2: Seismic waves and Earth models. In: Bormann P (ed) New manual of seismological observatory practice (NMSOP-2) (2012). IASPEI, GFZ German Research Centre for Geosciences, Potsdam, 105 pp. doi:10.2312/GFZ.NMSOP-2_CH2. <http://nmsop.gfz-potsdam.de>
- Bormann P, Klinge K, Wendt S (2014) Chapter 11: Data analysis and seismogram interpretation. In: Bormann P (ed) New manual of seismological observatory practice (NMSOP-2) (2012). IASPEI, GFZ German Research Centre for Geosciences, Potsdam, doi:10.2312/GFZ.NMSOP-2_CH11. <http://nmsop.gfz-potsdam.de>
- Broding RA, Bentley-Llewellyn NJ, Hearn DP (1964) A study of three dimensional seismic detection system. *Geophysics* 29(2):221–249
- Buttkus B (ed) (1986) Ten years of the Gräfenberg array: defining the frontiers of broadband seismology. *Geologisches Jahrbuch E-35*, Hannover, 135 pp
- Cessaro RK (1994) Sources of primary and secondary microseisms. *Bull Seismol Soc Am* 84(1):142–148

- Douze EJ (1964) Signal and noise in deep wells. *Geophysics* 29(5):721–732
- Forbriger Th (2012) IS 5.4: Recommendations for seismometer deployment and shielding. In: Bormann P (ed) New manual of seismological observatory practice (NMSOP-2) (2012). IASPEI, GFZ German Research Centre for Geosciences, Potsdam, 10 pp. doi:10.2312/GFZ.NMSOP-2_IS_5.4. <http://nmsop.gfz-potsdam.de>
- Friedrich A, Klinge K, Krüger F (1998) Ocean-generated microseismic noise located with the Gräfenberg array. *J Seismol* 2:47–64
- Fyen J (1990) Diurnal and seasonal variations in the microseismic noise observed at the NORESS array. *Phys Earth Planet In* 63:252–268
- Groos JC, Ritter JRR (2009) Time domain classification and quantification of seismic noise in an urban environment. *Geophys J Int* 179:1213–1231
- Hasselmann KA (1963) A statistical analysis of the generation of microseisms. *Rev Geophys* 1:177–209
- Haubrich RA, Munk WH, Snodgrass FE (1963) Comparative spectra of microseisms and swell. *Bull Seismol Soc Am* 53:27–37
- Longuet-Higgins MS (1950) A theory for the generation of microseisms. *Phil Trans R Soc Lond A* 243:1–35
- Parolai St (2012) Chapter 14: Investigation of site response in urban areas by using earthquake data and seismic noise. In: Bormann P (ed) New manual of seismological observatory practice (NMSOP-2) (2012). IASPEI, GFZ German Research Centre for Geosciences, Potsdam, 38 pp. doi:10.2312/GFZ.NMSOP-2_CH14. <http://nmsop.gfz-potsdam.de>
- Peterson J (1993) Observations and modeling of seismic background noise. U.S. Geological Survey Open-File Report, 93–322, 95 pp
- Schweitzer J, Fyen J, Mykkeltveit S, Gibbons SJ, Pirli M, Kühn D, Kværna T (2011) Chapter 9: Seismic arrays. In: Bormann P (ed) (2012) New manual of seismological observatory practice (NMSOP-2) (2012). IASPEI, GFZ German Research Centre for Geosciences, Potsdam, 80 pp. <http://nmsop.gfz-potsdam.de>
- Shinohara M, Suyehiro K, Shiobara H (2011) Chapter 7, section 7.5: Marine seismic observation. In: Bormann P (ed) (2012) pp 112–135. <http://nmsop.gfz-potsdam.de>
- Trnkoczy A, Bormann P, Hanka W, Holcomb LG, Nigbor RL (2002) Chapter 7: Site selection, preparation and installation of seismic stations. In: Bormann P (ed) IASPEI New Manual of Seismological Observatory Practice (NMSOP), vol 1. GeoForschungsZentrum Potsdam, Potsdam, 108 pp
- Trnkoczy A, Bormann P, Hanka W, Holcomb LG, Nigbor RL, Shinohara M, Suyehiro K, Shiobara H (2011) Chapter 7: Site selection, preparation and installation of seismic stations. In: Bormann P (ed) New manual of seismological observatory practice (NMSOP-2) (2012). IASPEI, GFZ German Research Centre for Geosciences, Potsdam. doi:10.2312/GFZ.NMSOP-2_CH7. <http://nmsop.gfz-potsdam.de>
- Webb SC (2002) Seismic noise on land and on the sea floor. In: Lee WHK, Kanamori H, Jennings PC, Kisslinger C (eds) International handbook of earthquake and engineering seismology, Part A. Academic, Amsterdam, pp 305–318
- Wiechert E (1904) Verhandlungen der zweiten internationalen Seismologischen Konferenz. Beitr Geophys, Ergänzungsband II, pp 41–43
- Wielandt E (2011) Chapter 5: Seismic sensors and their calibration. In: Bormann P (ed) New manual of seismological observatory practice (NMSOP-2) (2012). IASPEI, GFZ German Research Centre for Geosciences, Potsdam, 51 pp. doi:10.2312/GFZ.NMSOP-2_CH5. <http://nmsop.gfz-potsdam.de>
- Withers MM, Aster RC, Young CJ, Chael EP (1996) High-frequency analysis of seismic background noise as a function of wind speed and shallow depth. *Bull Seismol Soc Am* 86(5):1507–1515
- Young CJ, Chael EP, Withers MW, Aster RC (1996) A comparison of the high-frequency (>1Hz) surface and subsurface noise environment at three sites in the United States. *Bull Seismol Soc Am* 86(5):1516–1528

Seismic Reliability Assessment, Alternative Methods for

Hossein Ebrahimian and Raffaele De Risi
Department of Structures for Engineering and Architecture, University of Naples Federico II, Naples, Italy

Synonyms

Bayesian method; Fragility curve; Limit state; Mean annual frequency of exceedance; Monte Carlo simulation; PEER performance-based approach; Reliability assessment; Second-moment reliability methods; Simulation-based reliability; System reliability

A Perspective into the Seismic Reliability Assessment

The treatment of structural behavior under seismic actions can be considered as one of the most challenging aspects of structural performance assessment and design. It encompasses numerous

sources of uncertainty associated with the seismic action, structural response, and structural capacities. The various parameters involved in the seismic performance assessment problem can be generically designated in terms of seismic *demand* (D) and *capacity* (C) of the structure. The seismic demand and capacity can be characterized as a functional (i.e., function of function) of a set of *uncertain parameters* (also denoted as “random variables”). Hence, probabilistic methods are necessary in order to take into account the uncertainties in demand and capacity within the performance assessment or design.

Traditionally, seismic-design codes have addressed the uncertainties by allowing some degree of conservatism in evaluating demand and capacity at the level of structural components; nevertheless, the link to the overall performance of the structure remains unclear. Strictly speaking, the reliability of the structure to withstand future events remains more-or-less unknown to the designer while employing various established code-based approaches. Therefore, in the past decades, several research efforts have been carried out and substantial progress has been made towards the inclusion of various sources of uncertainty into structural performance assessment and design frameworks.

This entry introduces the main sources of uncertainty in demand and capacity, as well as their treatments through alternative reliability methods. In this regard, three general families of methods are briefly summarized in this entry: the second-moment reliability methods, the simulation-based reliability approaches, and finally the probabilistic performance-based methodology developed within the joint efforts of Pacific Earthquake Engineering Research Center (PEER). The closing section provides a brief discussion about future challenges in seismic reliability assessment.

An Introduction to Seismic Reliability Assessment

Seismic structural reliability is concerned with the probabilistic performance evaluation of

structures under seismic actions. This issue is addressed by means of defining a prescribed set of *limit states*, which describe specific stages of structural behavior and its consequences (e.g., states associated with consequences in terms of costs, loss of lives, impact on the environment, or serviceability; see Bozorgnia and Bertero 2004). Reaching a limit state denotes that an unacceptable behavior for the structure occurs. This condition is generically referred to as *failure*. The limit state exceedance is verified by comparing a measurable quantity, associated with the structural response, denoted as the *demand* D (also referred to as the engineering demand parameter (EDP), e.g., maximum inter-story drift ratio, maximum chord rotation, maximum component shear, etc.), with the corresponding limit value, denoted as the *capacity* C (e.g., yielding rotation, ultimate rotation, shear strength, etc.). As a result, the inequality $D \leq C$, also called *safety margin formulation*, defines acceptable performance respect to the current limit state in mathematical terms (Pinto et al. 2007). This formulation is generally expressed as *limit state function* (or *performance function*) $G = C - D$. This type of formulation is also employed in the *load and resistance factor design* (LRFD) format (Galambos et al. 1982), which is familiar to engineers as the safety checking equation. As an alternative to the safety margin formulation, the *safety factor formulation* is introduced as $G = \ln(C/D)$ (DNV 1992). For both the *safety margin formulation* and the *safety factor formulation*, failure is defined as $G \leq 0$.

As mentioned before, the evaluation of the structural performance in terms of demand and capacity may encompass several sources of uncertainty. In such context, the performance with respect to a given limit state is evaluated in terms of the probability that the limit state function will be less than zero (*probability of failure* P_F) in a specified reference time period. Identifying $\mathbf{X} = [X_1, X_2, \dots, X_n]$ to be the group of n uncertain parameters involved in the reliability problem, the probability of failure may be determined by the following integral:

$$P_F = P[G(\mathbf{X}) \leq 0] = \int_{G(\mathbf{X}) \leq 0} f_{\mathbf{X}}(\mathbf{X}) d\mathbf{x} \quad (1)$$

where $f_{\mathbf{X}}$ is the joint probability distribution for uncertain parameters $\mathbf{X}$. This means that one can evaluate the probability of failure by integrating the joint density function over the failure domain defined as $G(\mathbf{X}) \leq 0$.

Conventionally, the *generalized reliability* (or *safety*) index β associated with a given limit state function is defined by one-to-one mapping relationship with failure probability as follow:

$$\beta = -\Phi^{-1}(P_F) \quad (2)$$

where $\Phi(\cdot)$ is the standard normal Cumulative Distribution Function (CDF). Equation 2, however, does not mean to imply that uncertain parameters are jointly normal. The inverse of the standard normal CDF simply provides a convenient one-to-one mapping between the computed probability of failure and a reliability index.

The evaluation of the integral in Eq. 1 can be computationally difficult; some examples are as follows: $f_{\mathbf{X}}$ is often not well-defined because of the incompleteness of the statistical information available; $G(\mathbf{X})$ may have a nonlinear form; the computation of the multifold integral can be very difficult if the number of uncertain parameters is high. Various methods have been proposed for solving the integral form in Eq. 1. These approaches range from the classical moment methods for structural reliability (e.g., first-order second-moment reliability method) to the simulation-based approaches (i.e., Monte Carlo family of methods), and also the PEER approach, which is quite different compared to the other two techniques. In this entry, alternative methods for estimating the probability of failure are described.

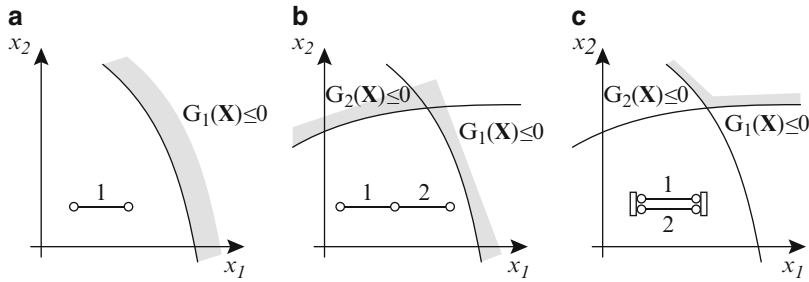
Types of Uncertainty

There are alternative ways for classifying the sources of uncertainty. Arguably, such classification depends to some extent on the reliability method adopted. Some classical reliability

methods provide the following classification of uncertainties in seismic structural engineering: (1) *inherent randomness* due to the intrinsic variability in material mechanical properties as well as environmental actions (i.e., loads or seismic excitation); (2) *statistical uncertainties* due to the lack of knowledge associated with the estimation of the parameters of probability distributions from observational samples of limited size; and finally (3) *model uncertainties* due to the imperfections of the mathematical models used to describe complex physical phenomena (i.e., model describing load or element capacity) (see also Der Kiureghian 1996; Ditlevsen and Madsen 1996). Some other reliability methods (mostly simulation based) classify the sources of uncertainty into two general categories: those related to the representation of the seismic action and those related to the structural modeling (see, e.g., Papadimitriou et al. 2001). Another useful classification is based on the definition of the *aleatory* and *epistemic* uncertainties (Cornell et al. 2002; Der Kiureghian and Ditlevsen 2009). These two categories are also known as *inherent* randomness (i.e., uncertainty impossible to mitigate) and *statistical* uncertainties (i.e., uncertainties that can be mitigated with reasonable effort), respectively (McGuire 2004). On the other hand, in the Bayesian probabilistic framework, the various sources of uncertainty are treated in a unified manner. In subsequent sections, the uncertainties in the seismic reliability problems are divided (if required) into those related to the representation of the seismic action, also referred to as the record-to-record (RTR) variability, and other sources of uncertainty. As mentioned above, the set of uncertain parameters involved in the seismic performance assessment or design problem are denoted hereafter by the vector of parameters $\mathbf{X}$.

Structural Component and Structural System Reliability

A *structural component* is a basic structural element whose performance is defined by a single limit state function $G(\mathbf{X}) = 0$ (Fig. 1a). In the hyperspace of the designated uncertain parameters $\mathbf{X}$, the limit state function represents the



Seismic Reliability Assessment, Alternative Methods for, Fig. 1 Reliability problems for the special case of only two uncertain parameters x_1 and x_2 : (a) Component,

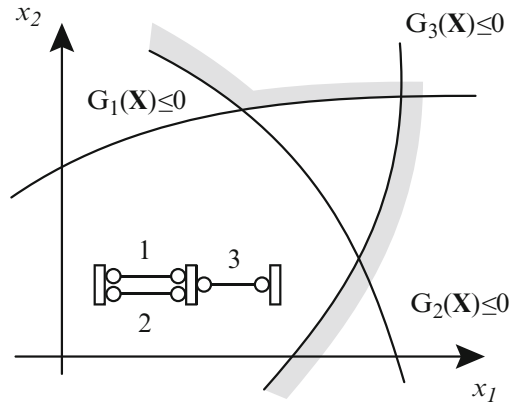
(b) series system, (c) parallel system. The gray shaded area represents the side of the failure domain

boundary (i.e., *limit state surface*) between safe and failure zones (the latter is also called the *failure domain*), as shown in Fig. 1a for the special case of only two uncertain variables (Ditlevsen and Madsen 1996).

A *structural system* is an assemblage of components, whose performance is described by individual component limit state function. Therefore, in a structural system reliability problem, the failure domain may be described as the union and/or intersection of several limit state surfaces. In particular, a *series system* (also known as the *weakest-link*) reliability problem is defined as the union of the failure domains (Fig. 1b), and a *parallel system* reliability problem is defined as the intersection of the failure domains (Fig. 1c). In analytic terms, the probability of failure for series and parallel systems with n components is defined, respectively, as

$$P_{F, \text{series}} = P \left[\bigcup_{i=1}^n (G_i(\mathbf{X}) \leq 0) \right] \quad P_{F, \text{parallel}} = P \left[\bigcap_{i=1}^n (G_i(\mathbf{X}) \leq 0) \right] \quad (3)$$

A cut-set is made as a combination of series and parallel systems generating a *generalized system* reliability problem. A cut-set (Fig. 2) is any set of components whose joint failure represents a failure of the system (Ditlevsen and Madsen 1996; Au and Beck 2003b; Jalayer et al. 2007a). Analytically, the probability of failure for a cut-set system is evaluated as



Seismic Reliability Assessment, Alternative Methods for, Fig. 2 Reliability problems for the special case of only two uncertain parameters x_1 and x_2 : cut-set system. The gray area represents the side of the failure domain

$$P_{F, \text{cut-set}} = P \left[\bigcup_{j=1}^N \bigcap_{i=1}^{n_j} (G_i(\mathbf{X}) \leq 0) \right] \quad (4)$$

where N defines the associated number of the subsystem that is connected in series to the other subsystems and n_j is the number of components connected in parallel belonging to the j th subsystem.

Seismic Reliability Assessment, a Time Variant Nonlinear Problem

If the limit state function reveals time-dependent properties, then the reliability problem can be referred to as *time variant*. Hence, given the stochastic nature of the seismic excitation, the seismic reliability assessment can be classified as a

time-variant problem. Therefore, the failure event constitutes the outcrossing of the time-variant limit state surface. This represents the so-called *first-passage* (or *first-excursion*) problem (e.g., Au and Beck 2001b). Moreover, in many practical cases, the limit state function is nonlinear. Therefore, the seismic reliability problems can be classified, in general, as time-variant nonlinear problems.

The Second-Moment Reliability Methods

In general, the n^{th} moment of a mono-variate distribution (i.e., a distribution with a single scalar random variable) is defined as the expected value of the variable raised to the power of n . Using moments of finite order for describing the uncertainty in given parameter is a well-known problem in statistics referred to as the *moment methods*. A comprehensive review of the *moment methods* in the structural reliability analysis is presented in Zhao and Ono (2001). The second-moment reliability methods can be dated back to the 1960s (some historical notes can be found in Ditlevsen and Madsen 1996). Since then, this family of methods has been refined and extended significantly. At present, they represent one of the most important and widespread approaches for seismic reliability evaluation of structures. As reflected by their title, the second-moment reliability methods employ the first two moments of the vector of uncertain parameters $\mathbf{X}$ (i.e., expected value $\mathbb{E}[\mathbf{X}]$ and covariance $\Sigma_{\mathbf{X}}$).

Linear (First-Order) Limit State Function and Its First Second Moments

The simplest mathematical form for a performance function is the linear (first-order) polynomial:

$$G(\mathbf{X}) = a_0 + a_1 X_1 + \dots + a_n X_n = a_0 + \mathbf{a}^T \mathbf{X} \quad (5)$$

where the coefficients a_0 and $\mathbf{a} = [a_1, \dots, a_n]$ are constants. It is possible to write the two central

moments of $G(\mathbf{X})$ in terms of the first two (central) moments of the vector of uncertain parameters $\mathbf{X}$:

$$\begin{aligned} \mu_G &= \mathbb{E}[G(\mathbf{X})] = a_0 + \mathbf{a}^T \mathbb{E}[\mathbf{X}] = a_0 + \mathbf{a}^T \boldsymbol{\mu}_{\mathbf{X}} \\ \sigma_G^2 &= \mathbb{E}[(G(\mathbf{X}) - \mu_G)^2] \\ &= \mathbf{a}^T \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})(\mathbf{X} - \boldsymbol{\mu}_{\mathbf{X}})^T] \mathbf{a} = \mathbf{a}^T \Sigma_{\mathbf{X}} \mathbf{a} \end{aligned} \quad (6)$$

One of the simplest forms of reliability measure was proposed by Cornell (1969), in which the second-moment reliability index is defined as the ratio between the mean value and the standard deviation of the performance function G (i.e., the so-called “Cornell index”):

$$\beta = \frac{\mu_G}{\sigma_G} \quad (7)$$

Assuming that the unacceptable performance is reached when $G \leq 0$, it is possible to establish a one-to-one mapping between the Cornell index and the probability of failure, as already shown in Eq. 2:

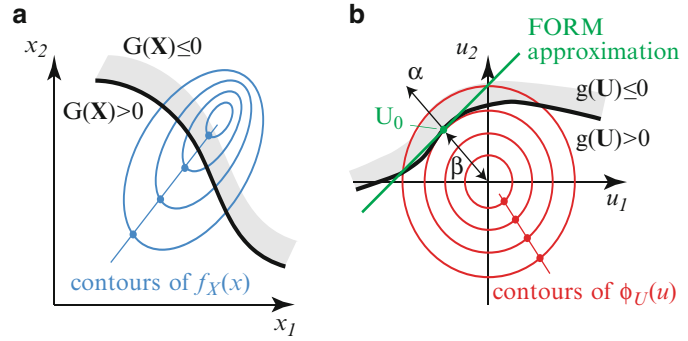
$$\begin{aligned} P_F &= P[G(\mathbf{X}) \leq 0] \\ &= P\left[\frac{G(\mathbf{X}) - \mu_G}{\sigma_G} \leq \frac{0 - \mu_G}{\sigma_G}\right] \\ &= \Phi\left(-\frac{\mu_G}{\sigma_G}\right) = \Phi(-\beta) \end{aligned} \quad (8)$$

This mapping is exact if the variables are normal. Otherwise, Eq. 8 returns an approximate value of the failure probability.

First-Order Second-Moment (FOSM) Method in Component Reliability

In the case of a nonlinear limit state surface, the reliability index fails to be constant under different but mechanically equivalent formulations of the performance function. The issue of a lack of invariance was first recognized by Ditlevsen (1973) and later resolved by Hasofer and Lind (1974). In order to overcome this issue, they proposed the transformation of the uncertain

Seismic Reliability Assessment, Alternative Methods for, Fig. 3 The limit state function and the joint-normal PDF of the uncertain parameters in (a) nonstandard space of correlated uncertain parameters and (b) standard space of the uncorrelated uncertain parameters



parameters $\mathbf{X}$ into a set of uncorrelated standard normal deviates (variables) $\mathbf{U}$ (i.e., $\boldsymbol{\mu}_U = \mathbf{0}$, $\boldsymbol{\Sigma}_U = \mathbf{I}$). To this aim, the basic random variables, $\mathbf{X}$, can be transformed into a set of uncorrelated standard variates $\mathbf{U}$ through the linear transformation:

$$\mathbf{U} = \mathbf{L}_X^{-1} \mathbf{D}_X^{-1} (\mathbf{X} - \boldsymbol{\mu}_X) \quad \text{or} \quad \mathbf{X} = \mathbf{D}_X \mathbf{L}_X \mathbf{U} + \boldsymbol{\mu}_X \quad (9)$$

where $\mathbf{D}_X$ is diagonal matrix of the standard deviations of the basic random variables and $\mathbf{L}_X$ is the lower triangular matrix obtained as the Cholesky decomposition of the correlation coefficient matrix $\mathbf{R}_X$ (i.e., $\mathbf{R}_X = \mathbf{L}_X \mathbf{L}_X^T$). It is then possible to define the joint-normal PDF $\phi_U(\mathbf{u})$ and $g(\mathbf{U}) = G[\mathbf{X}(\mathbf{U})]$ as the joint-normal distribution of the uncertain parameters and the limit state function, respectively, transformed in the space of the standard uncorrelated variables (Fig. 3). By considering a linear performance surface (see Eq. 5), $g(\mathbf{U})$ can be expressed as

$$\begin{aligned} g(\mathbf{U}) &= G[\mathbf{X}(\mathbf{U})] = a_0 + \mathbf{a}^T (\mathbf{D}_X \mathbf{L}_X \mathbf{U} + \boldsymbol{\mu}_X) \\ &= a_0 + \mathbf{a}^T \boldsymbol{\mu}_X + \mathbf{a}^T \mathbf{D}_X \mathbf{L}_X \mathbf{U} = b_0 + \mathbf{b}^T \mathbf{U} \end{aligned} \quad (10)$$

It can be easily demonstrated that the first and second moments of this performance function are $\mu_g = b_0$ and $\sigma_g^2 = \mathbf{b}^T \mathbf{b}$. Substituting these two terms in Eq. 7, the reliability index in the standard space can be obtained as follows:

$$\beta = \frac{\mu_g}{\sigma_g} = \frac{b_0}{\sqrt{\mathbf{b}^T \mathbf{b}}} \quad (11)$$

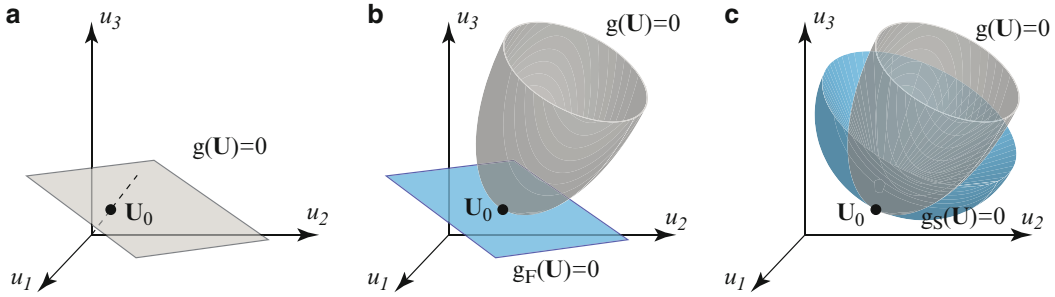
The reliability index in Eq. 11 can be interpreted as the distance from the origin to the hyperplane (in standard normal space) defined by $b_0 + \mathbf{b}^T \mathbf{U} = 0$. In other words, the reliability index for a linear limit state function is numerically equivalent to the distance from the origin to the limit state hyperplane in the space of the standard normal variables. Using this geometric interpretation, Hasofer and Lind (1974) defined the reliability index as the minimum distance between the origin of standard normal space and the function $g(\mathbf{U})$ (Fig. 4a).

The point $\mathbf{U}_0$ belonging to the limit state surface that has the minimum distance from the origin of standard normal space is called *design point* (or *performance point*). The latter is the most likely failure point. Strictly speaking, $\mathbf{U}_0$ is the vector of uncertain parameters that has the maximum contribution towards the failure probability.

As also mentioned before, the limit state function is generally nonlinear. Therefore, the function $g(\mathbf{U})$ usually cannot be characterized by a first-order polynomial. An approximate way to solve this problem is to replace the nonlinear function with a first-order Taylor series expansion. In other words, the performance surface, in the neighborhood of the design point $\mathbf{U}_0$, is approximated by the tangent hyperplane at $\mathbf{U}_0$ (Fig. 4b). In analytical terms, the approximate limit state function $g_F(\mathbf{U})$ becomes

$$g_F(\mathbf{U}) = g(\mathbf{U}_0) + \nabla g^T(\mathbf{U}_0) \cdot (\mathbf{U} - \mathbf{U}_0) \quad (12)$$

where $\nabla g(\mathbf{U}_0)$ is the gradient of g evaluated at the performance point and the subscript F represents



Seismic Reliability Assessment, Alternative Methods for, Fig. 4 Different types of limit state surface and their approximations: (a) linear, (b) nonlinear with linear

approximation, $g_F(\mathbf{U}) = 0$ at performance point, (c) nonlinear surface approximated with a paraboloid, $g_S(\mathbf{U}) = 0$

the first-order approximation to g . In particular, the first two moments of $g(\mathbf{U})$ can be calculated as follows (by using the relationships in Eq. 6):

$$\begin{aligned}\mu_g &= g(\mathbf{U}_0) + \nabla g^T(\mathbf{U}_0)(\boldsymbol{\mu}_{\mathbf{U}} - \mathbf{U}_0) \\ &= -\nabla g^T(\mathbf{U}_0)(\boldsymbol{\mu}_{\mathbf{U}} - \mathbf{U}_0)\sigma_g^2 \\ &= \nabla g^T(\mathbf{U}_0)\boldsymbol{\Sigma}_{\mathbf{U}}\nabla g(\mathbf{U}_0) \\ &= \nabla g^T(\mathbf{U}_0)\nabla g(\mathbf{U}_0)\end{aligned}\quad (13)$$

where $g(\mathbf{U}_0)$ is equal to zero since $\mathbf{U}_0$ is on the limit state function, and the mean is equal to zero ($\boldsymbol{\mu}_{\mathbf{U}} = 0$) and the covariance is equal to one ($\boldsymbol{\Sigma}_{\mathbf{U}} = \mathbf{I}$) by definition. For a nonlinear performance function, the first-order reliability index can be defined as equal to the distance of the linearized failure surface (at the performance point $\mathbf{U}_0$) to the origin of the standardized uncorrelated variables (Fig. 3b). With reference to Eq. 12, it is possible to calculate the first two moments of the linearized limit state function (according to Eq. 13) and obtain the following reliability index (in literature also referred to as first-order second-moment reliability index β_{FOSM} or Hasofer-Lind reliability index β_{HL}):

$$\beta = \frac{\mu_g}{\sigma_g} = \frac{-\nabla g^T(\mathbf{U}_0) \cdot \mathbf{U}_0}{\sqrt{\nabla g(\mathbf{U}_0)^T \nabla g(\mathbf{U}_0)}} = \boldsymbol{\alpha}^T \cdot \mathbf{U}_0 \quad (14)$$

where $\boldsymbol{\alpha}$ is the unit normal (directed towards the failure region) on the hyperplane that is tangent to the limit state surface at the performance

point $\mathbf{U}_0$ (Fig. 3b). The components of $\boldsymbol{\alpha}$ are called the sensitivity factors of the reliability index β with respect to uncertain parameters $\mathbf{U}$. The smaller is α_i , the smaller is the contribution of the related uncertain parameter U_i to the failure probability assessment.

Clearly, the determination of β requires the finding of the design point $\mathbf{U}_0$. Therefore, one has to solve the following constrained optimization problem:

$$\begin{cases} \beta = \min|\mathbf{U}| \\ g(\mathbf{U}) = 0 \end{cases} \quad (15)$$

There are many algorithms that are able to solve this kind of problems (see, e.g., Liu and Der Kiureghian 1991). The most common algorithm used in structural reliability is the one developed by Rackwitz and Flessler (1978). A comprehensive description of the algorithm and relevant examples can be found in Pinto et al. (2007), while a first review of the efficiency of the available algorithms was presented in Liu and Der Kiureghian (1991). Once β is determined, the probability of failure can be approximated by Eq. 8.

Extension to System Reliability

It is noteworthy that in order to compute a system reliability problem, one must be able to compute (or approximate) the probability of a union of N events (series system) or the probability of an intersection of N events (parallel systems). The probability of failure for these two cases

can be approximated as (Hohenbichler and Rackwitz 1983)

$$P_{F,\text{series}} \approx 1 - \Phi_N(\boldsymbol{\beta}, \mathbf{R}) \quad P_{F,\text{parallel}} \approx \Phi_N(-\boldsymbol{\beta}, \mathbf{R}) \quad (16)$$

where Φ_N is the multi-normal CDF, $\boldsymbol{\beta}$ is the N -vector of the reliability indexes associated with the structural components, and $\mathbf{R}$ is the $N \times N$ correlation matrix, in which the correlation coefficients, $\rho_{ij} = \alpha_i^T \alpha_j$, quantify the correlation between failure modes i and j . In general, the evaluation of $\Phi_N(\boldsymbol{\beta}, \mathbf{R})$ and $\Phi_N(-\boldsymbol{\beta}, \mathbf{R})$ in Eq. 16 is not trivial; however, an efficient Monte Carlo simulation algorithm has been developed for this purpose (Ambartzumian et al. 1998). The previous results can be combined to estimate the probability of failure of a general system (Ditlevsen and Madsen 1996).

FORM and SORM Method in Component Reliability

For the reliability index presented so far, only the two moments of each uncertain parameter have been considered. However, in many applications, information beyond the second moments is available. The methods that make use of that additional information are called *full distribution reliability methods*. If the probability distribution is available for some or all the basic variables, it would be sensible to incorporate it in the reliability analysis. If the joint probability distribution for the vector of uncertain parameters in the original space is not normal and the variables are not independent, the multifold integral presented in Eq. 1 becomes very hard to solve. In fact, several efficient methods for approximating the integral take advantage of the special properties of the standard normal distribution of uncorrelated variables. Therefore, the transformation of independent/dependent non-normal uncertain parameters to independent joint-normal uncertain parameters is quite important.

Historically, the transformation approach was developed as an extension of the FOSM and is

therefore often called *advanced* or *extended* FOSM, since the limit state function is approximated with a first-order approximation, while the probability distributions are no longer approximated only with the first two central moments. In particular, if the limit state function is approximated with its first-order Taylor expansion, these methods are known as FORM (*First-Order Reliability Methods*), and the analytical relations to calculate the reliability index are similar to those reported from Eqs. 12 to 15. If the limit state function is approximated with its second-order Taylor expansion, these methods are known as SORM (*Second-Order Reliability Methods*). In both cases, the Taylor expansion is done at the *design point* $\mathbf{U}_0$.

A SORM analysis reduces the error associated with the non-flatness of the limit state function at the design point in a FORM approximation. In particular, this approach fits a second-order surface (usually a paraboloid) at the performance point as shown in Fig. 4c. A parabolic approximation to the limit state surface at the design point is

$$g_S(\mathbf{U}) = g(\mathbf{U}_0) + \nabla g^T(\mathbf{U}_0) \cdot (\mathbf{U} - \mathbf{U}_0) + \frac{1}{2}(\mathbf{U} - \mathbf{U}_0)^T \mathbf{H}(\mathbf{U}_0)(\mathbf{U} - \mathbf{U}_0) \quad (17)$$

where $\mathbf{H}(\mathbf{U}_0)$ is the Hessian matrix of g evaluated at $\mathbf{U}_0$ and the subscript S denotes the approximation to the second order of g . In this case, it is more difficult to express the first two moments of the limit state function (for more details, see Pinto et al. 2007); instead, the probability of failure is approximated in terms of β and the $n-1$ principal curvatures κ_i of the paraboloid in the standardized uncorrelated n -space defined by vector $\mathbf{U}$. The exact expression of this probability is given in terms of single-fold integral defined by Tvedt (1990) which can be solved numerically in lack of closed-form solution. A simpler approximation was derived on the basis of asymptotic analysis by Breitung (1984):

$$P_F \approx \Phi(-\beta) \prod_{i=1}^{n-1} \frac{1}{\sqrt{1 + \beta \kappa_i}} \quad (18)$$

where β is calculated as the minimum distance between the origin of the standard normal space and the paraboloid.

Reliability Methods in Design Codes: The Safety Factors in LRFD

Until the 1960s, most design codes did not provide an explicit implementation of reliability concepts. The recommended methods mainly involved stress checks in critical sections, which aimed to verify that the unit stresses under the action of working, or service, loads do not exceed predesignated allowable values (i.e., values prescribed by the building code to provide a factor of safety against attainment of some limiting stress, such as the minimum specified yield stress or the stress at which the buckling occurs); this design philosophy is also known as the working stress design. The issue of uncertainty in the various parameters was implicitly addressed by employing safety factors much larger than unity on the material strength. During the 1960s and 1970s, reliability analyses were addressed more explicitly by the codes, giving rise to reliability-based design. However, because professional design engineers found the application of the reliability-based methods somewhat difficult to access, a *semi-probabilistic* code-based approach known as the *load and resistance factor design* (LRFD) format was proposed (Galambos et al. 1982):

$$R/\gamma_R \geq \gamma_S \cdot S \quad \text{or} \quad R_d \geq S_d \quad (19)$$

where R_d and S_d are the design resistance and load, respectively; R and S are the nominal resistance and nominal action (due to loads); and γ_R and γ_S are the associated resistance and load safety factors. The codes aim to link these coefficients with the target (admissible) reliability and with the amount of uncertainty in each uncertain parameter. The two safety factors in Eq. 19 have the general form

$$\gamma_R = \frac{\mu_R}{R} (1 + \alpha'_R \beta \delta_R) \quad \text{or} \quad \gamma_S = \frac{\mu_S}{S} (1 + \alpha'_S \beta \delta_S) \quad (20)$$

where μ_R and μ_S denote the mean resistance and load; α'_R and α'_S are the sensitivity coefficients (function of the sensitivity factors presented in Eq. 14) for the resistance and load corresponding to R and S , respectively; β is the FOSM reliability index associated to the target reliability (obtained from the inverse of the standard normal cumulative distribution); and finally, δ_R and δ_S are the coefficient of variation (i.e., the standard deviation divided by the mean) of the resistance and load. For example, assuming that the nominal values are equal to the mean values, it can be observed that the safety factor is only a function of α , β , and δ .

Clearly, the safety factors depend on the chosen limit states. For the ultimate limit state, for instance, most of design codes provide safety factors that are calibrated with respect to nominal values. These nominal values are chosen as the value with a predefined probability of being exceeded; typically a lower (usually 5 %) percentile is selected for the resistance and an upper (usually 95 %) percentile is selected for the load effect. The European standard (CEN 2002) and ASCE standard (ASCE 2010) both define the acceptable values of the reliability index. In particular, the European standard defines minimum values that are established as a function of the consequences related to the reaching of a given limit state (with β varying from 3.3 to 4.3 with respect to 50 years of service period), whereas the ASCE standard defines the acceptable reliability index (which is related to a specified failure probability) as a function of the structural damage and occupancy category of the construction (with β varying from 2.5 to 4.5 for 50 years of service period). This approach, however, allows for reliability checking at the member level rather than at the system level. As mentioned before, establishing the link between component reliability and system reliability is not a trivial problem.

Alternative Methods Based on Time-Invariant Reliability Methods

When an equivalent linear static procedure or a response spectrum analysis is performed, the

reliability analysis can be carried out in a straightforward manner using the time-invariant reliability methods. However, in the case of dynamic nonlinear analyses, the problem becomes much more complicated. In fact, in the latter case, it is generally difficult to express the limit state of interest directly in terms of time-invariant basic load and resistance uncertain parameters. Therefore, the combination of the different source of uncertainties becomes an intricate problem to tackle. Nowadays, it is possible to combine the reliability methods and the finite element analysis to overcome this problem (Der Kiureghian 1996). Currently, useful specialized software is available to address this issue (Der Kiureghian et al. 2006). Alternatively, a few *hybrid* approaches have been recently proposed in order to address the time variance in reliability problem.

These methods combine the RTR variability (taken into account by alternative methods like the PEER approach, which is discussed subsequently in this entry) with other source of uncertainties (treated by current FORM/SORM reliability methods). Typical hybrid approaches consist of the confidence interval approach (Cornell et al. 2002; Ellingwood and Kinali 2009) and the mean estimate approach (Cornell et al. 2002). Other relevant examples can be found in Ibarra and Krawinkler (2005), Haselton and Deierlein (2007), Dolšek (2009), Liel et al. (2009), Vamvatsikos and Fragiadakis (2010), and Celarec and Dolšek (2013). In these studies, many approximations are presented; for instance, if the limit state function does not have a defined functional form, the required gradients of the linearized limit state function can be obtained through perturbation of individual uncertain parameters in a series of sensitivity analyses.

Simulation-Based Reliability Methods

The general integral form in Eq. 1 can also be estimated by employing various simulation techniques. These techniques are especially efficient when the limit state function is non-differentiable

or has several design points that make significant contributions to the failure probability (Faber 2012; Pinto et al. 2007). Although a large variety of simulation techniques may be found in the literature, they generally have their origin in the so-called *Monte Carlo* (MC) method (Rubinstein 1981; Fishman 1996). The general principles of standard MC simulation technique are outlined in the following section. To have a further insight into the basis of simulation techniques, the probability integral in Eq. 1 is rewritten as follows:

$$P_F = \int_{\Omega_{\mathbf{x}}} I_F(\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} = \mathbb{E}[I_F(\mathbf{X})] \quad (21)$$

where $\Omega_{\mathbf{x}}$ denotes the entire sample space of $\mathbf{X}$ and $I_F(\mathbf{x})$ is an indicator function that equals one when $G(\mathbf{x}) \leq 0$ and zero otherwise. The integral in Eq. 21 is the expected value of the indicator function denoted as $\mathbb{E}[I_F(\mathbf{x})]$. In other words, P_F is the expected value of $I_F(\mathbf{x})$ based on the joint PDF $f_{\mathbf{x}}(\mathbf{x})$. It is noteworthy that throughout section “[Simulation-Based Reliability Methods](#),” $p(\mathbf{x})$ might be used in place of $f_{\mathbf{x}}(\mathbf{x})$ interchangeably. This is done in order to homogenize and to simplify the probability notations within different equations. Hence, the term $p(\mathbf{x})$ is used in cases where the conditioning on a preposition is defined, the use of $f_{\mathbf{x}}$ might create further confusions, or in the robust reliability expressions defined subsequently in section “[Robust Reliability Assessment Using Simulation-Based Approaches](#).” However, the marginal PDFs are expressed generally by $f_{\mathbf{x}}$. By employing the MC method, various realizations of the uncertain parameters $\mathbf{x}$ (i.e., $\mathbf{x}_k$, $k = 1: N_S$) can be independently sampled so that the multidimensional integral in Eq. 21 is estimated as follows:

$$P_F = \mathbb{E}[I_F(\mathbf{X})] \cong \frac{1}{N_S} \sum_{k=1}^{N_S} I_F(\mathbf{x}_k) = \frac{N_F}{N_S} \quad (22)$$

where N_S is the total number of simulations and N_F is the number of realizations for which $I_F(\mathbf{x}_k) = 1$ (i.e., the realizations leading to failure).

It can be shown that Eq. 22 is an unbiased estimator for P_F since

$$\mathbb{E}(\mathbb{E}[I_F(\mathbf{X})]) = \frac{1}{N_S} \sum_{k=1}^{N_S} \mathbb{E}[I_F(\mathbf{x}_k)] = \frac{N_S P_F}{N_S} = P_F \quad (23)$$

The variance of the estimator can be calculated by considering that $I_F(\mathbf{x}_k)$'s are independently distributed Bernoulli variables with variance equal to $P_F(1-P_F)$:

$$\begin{aligned} \text{VAR}(\mathbb{E}[I_F(\mathbf{X})]) &= \text{VAR}\left(\frac{1}{N_S} \sum_{k=1}^{N_S} I_F(\mathbf{x}_k)\right) \\ &= \frac{N_S P_F (1 - P_F)}{N_S^2} \\ &= \frac{P_F (1 - P_F)}{N_S} \end{aligned} \quad (24)$$

This equation reveals that the uncertainty in the estimator decreases with increasing N_S . Therefore, in order to accurately quantify the likelihood of rare events (with small P_F), which is a characteristic feature in seismic assessment, N_S should be increased. Besides the integral form in Eq. 21 for the structural reliability formulation, many problems in probabilistic inference require the calculation of multidimensional integrals or summations over very large outcome spaces. In the general form, they can be expressed as

$$R = \int_{\Omega_{\mathbf{x}}} h(\mathbf{x}) f_{\mathbf{x}}(\mathbf{x}) d\mathbf{x} = \mathbb{E}[h(\mathbf{X})] \quad (25)$$

where R is a performance measure of interest and $h(\mathbf{x})$ is a response quantity of interest. The general idea of MC integration is used to approximate the expectation. Hence, a set of N_S samples can be drawn independently from distribution $p(\mathbf{x})$, and the expectations can be approximated by a finite sum:

$$R = \mathbb{E}[h(\mathbf{X})] \cong \frac{1}{N_S} \sum_{k=1}^{N_S} h(\mathbf{x}_k) \quad (26)$$

In this procedure, the analytic integration is replaced with summation over a suitably large set of samples. Generally, the approximation can be made as accurate as needed by increasing N_S .

Fundamentals of MC Method

The MC family of methods employs various sampling techniques in order to generate realizations of the prescribed uncertain parameters. The standard MC simulation uses the random sampling approach in order to generate a large amount of realizations of uncertain parameters. The probability of failure can then be estimated by directly implementing Eq. 22 regardless of the complexity of the problem. Checking whether the structure has failed for each sample usually requires a structural analysis. To simulate $k = 1:N_S$ outcomes of the joint density function (corresponding to Eqs. 21–22 or Eqs. 25–26) in standard MC simulation, a random number, denoted as $z_{j,k}$, between 0 and 1 is generated for each of the components in $\mathbf{x}_k$ (i.e., $x_{j,k}$, $j = 1:n$). Assuming that X_j 's are independent, the random numbers $z_{j,k}$ are mapped to $x_{j,k}$ by

$$x_{j,k} = F_{X_j}^{-1}(z_{j,k}) \quad (27)$$

where F_{X_j} is the marginal Cumulative Distribution Function (CDF) associated with the uncertain parameter X_j . Generally, the standard MC simulation is not justifiable computationally for adequate estimation of very small failure probabilities as a very large number of simulations are required.

The Latin hypercube sampling technique (LHS; Helton and Davis 2003) can be employed in order to reduce the number of simulations, N_S , in addition to achieving an acceptable level of accuracy for the statistical characteristics of response. The LHS is a special type of MC simulation that uses the stratification of the theoretical CDFs of uncertain parameters. Stratification divides the CDF curve into N_S equal intervals on the probability scale (i.e., 0.0 to 1.0). A sample is then randomly drawn from each interval or “stratification” of the input CDFs based on the

technique of “sampling without replacement.” According to this technique, once a sample is taken from a designated stratification, this stratification is not sampled again.

As a result, the samples based on LHS reflect more accurately the distribution of values in the input probability distribution. LHS offers great benefits in terms of increased sampling efficiency and faster runtimes (due to fewer N_S that is required). Its roles in different aspects of reliability engineering have been described by, e.g., Novák et al. (1998) and Olsson et al. (2003). Stein (1987) has shown that the LHS reduces the variance of the response function compared to the crude Monte Carlo method.

Various methodologies were proposed for sampling based on LHS method; however, the most efficient strategy to perform sample selection, which deals also with samples of the tails of the PDF, is the sampling of interval mean values denoted as LHS mean (Huntington and Lyrintzis 1998). In this context, consider that the range of probability distribution F_{X_j} between 0.0 and 1.0 is divided into N_S equidistant intervals each with equal probability of $1/N_S$. The expression $\xi_{j,k}$ is the value of the uncertain parameter associated with the k th stratification with the probability k/N_S where $k = 1:N_S$:

$$\xi_{j,k} = F_{X_j}^{-1}(k/N_S) \quad (28)$$

The sampled data $x_{j,k}$ within the interval $[\xi_{j,k-1}, \xi_{j,k}]$ can be derived as

$$\begin{aligned} x_{j,k} &= \frac{\int_{\xi_{j,k-1}}^{\xi_{j,k}} x f_{X_j}(x) dx}{\int_{\xi_{j,k-1}}^{\xi_{j,k}} f_{X_j}(x) dx} \\ &= N_S \int_{\xi_{j,k-1}}^{\xi_{j,k}} x f_{X_j}(x) dx, \quad j = 1, \dots, n, \\ &\quad k = 1, \dots, N_S \end{aligned} \quad (29)$$

where f_{X_j} is the PDF of the uncertain parameter X_j . In spite of high efficiency of LHS technique, there are generally two issues concerning

statistical correlation (Vořechovský and Novák 2009): (1) diminishing undesired and spurious correlation between uncertain parameters generated during sampling procedure, particularly in the case of a very small number of simulations, and (2) introducing the prescribed statistical correlations between pairs of uncertain parameters defined by the target correlation. Hence, in order to impose a prescribed correlation into the sampling scheme, an optimization problem for minimizing the difference between the target correlation and the actual correlation (estimated from samples) should be solved. For this purpose, a stochastic optimization approach called simulated annealing (SA) has been recently proposed (Vořechovský and Novák 2009; Golafshani et al. 2011) for application of LHS together with the SA methodology in structural reliability assessment).

The main drawback of the standard MC simulation methods is that it is not computationally efficient for problems dealing with small probabilities of failure (i.e., $P_F \leq 10^{-3}$). The main reason is due to the fact that the number of samples and hence the number of structural analyses, which are required to achieve a given accuracy, is inversely proportional to P_F . It requires information from rare samples that lead to failure, and on average, many samples are required before one such failure sample occurs. For instance, in case of achieving the P_F in order of 10^{-4} (which is typical in seismic reliability assessment) with a coefficient of variation around 10 %, nearly 10^6 simulations are required. Essentially, for rare events, the chance of generating samples from $f_{\mathbf{X}}(\mathbf{x})$ for which $I_F(\mathbf{x})$ equals one is extremely small (i.e., success rate is slow); therefore, many samples are needed to fulfill this requirement.

To overcome this obstacle, different techniques have been proposed in order to reduce the number of simulations. One of the most commonly applied techniques is namely the *importance sampling* method, which is described in the next section.

Importance Sampling Simulation Technique

The underlying idea of the importance sampling simulation (Rubinstein 1981; Shinozuka 1983;

Schuëller and Stix 1987; Au and Beck 1999) is to carry out Monte Carlo simulation with samples having a higher rate of falling in the failure region since these samples contribute only to the evaluation of P_F . In this class of methods, the integral is written in the following form:

$$P_F = \int_{\Omega_X} I_F(\mathbf{x}) \frac{p(\mathbf{x})}{p_{IS}(\mathbf{x})} p_{IS}(\mathbf{x}) d\mathbf{x} = \mathbb{E} \left[I_F(\mathbf{X}) \frac{p(\mathbf{X})}{p_{IS}(\mathbf{X})} \right] \quad (30)$$

where $P_{IS}(\mathbf{x})$ is denoted the importance sampling density function to increase the probability of sampling from the failure region. Applying Monte Carlo simulation to the integral, P_F is then estimated by

$$P_F \cong \frac{1}{N_S} \sum_{k=1}^{N_S} I_F(\mathbf{x}_k) \frac{p(\mathbf{x}_k)}{p_{IS}(\mathbf{x}_k)} \quad (31)$$

The efficiency of this method relies on a proper choice of the importance sampling distribution, which inevitably requires some knowledge about failure zone. Many past studies successfully applied importance sampling to time-invariant reliability problems where the number of uncertain parameters is not too large (see, e.g., Schuëller and Stix 1987; Papadimitriou et al. 1997; Der Kiureghian and Dakessian 1998; Au and Beck 1999).

For instance, Schuëller and Stix (1987) proposed the selection of an importance sampling density function as an N -dimensional joint-normal PDF with uncorrelated components, where the mean value equals the design point as obtained from FORM analysis and standard deviation values corresponding to the standard deviations of the components of $\mathbf{X}$. Although design points are a good notion to characterize the important region in $\mathbf{X}$ and hence importance sampling densities centered on them are often a good choice, they are not a necessary ingredient for forming a good importance sampling density. Hence, adaptive or iterative schemes and in particular the kernel density estimators, which do not involve the notion of design points, have attracted

considerable attention (see Au and Beck (1999) for a complete discussion and suggestion of an efficient kernel method).

Au and Beck (2001b) have also developed a very efficient importance sampling method for the first-excursion problem for linear dynamical systems under Gaussian stochastic excitation. However, for time-dependent problems, which are often characterized by a large number of uncertain parameters with complexity arising from their dynamic nature, the application of importance sampling is much more difficult (Au and Beck 2003a).

Hence, Au and Beck (2001a) proposed a *subset simulation* technique that is capable of efficiently computing the small failure probabilities encountered in engineering reliability analysis of general dynamical systems.

Calculation of Failure Probability Using the Subset Simulation

Subset simulation is an efficient and adaptive stochastic simulation algorithm for estimating small failure probabilities in high dimensions (Au and Beck 2001a, 2003b; Zuev et al. 2012). The underlying idea is to express the small failure probability as a product of larger probabilities conditional on some intermediate events. This allows converting the simulation of a rare event into a sequence of simulations of more frequent events (see also Jalayer and Beck 2008; Jalayer et al. 2010).

In this methodology, the failure region is modeled as the last element in a sequence of embedded failure regions (i.e., decreasing nested sequence of failure regions) $F = F_m \subset \dots \subset F_2 \subset F_1$. F_1 is the first element in the failure sequence (i.e., largest failure region), and $F = F_m$ is the target failure region and the last element in the failure sequence. Therefore, by considering the product rule in probability, the failure probability can be derived as follows:

$$\begin{aligned} P_F &= P(F_m) = P(F_m|F_{m-1})P(F_{m-1}) \\ &= P(F_1) \prod_{i=1}^{m-1} P(F_{i+1}|F_i) \end{aligned} \quad (32)$$

According to Eq. 32, the rough idea of subset simulation is to estimate the failure probability as a product of a sequence of conditional probabilities denoted as $\{P(F_{i+1}|F_i): i = 1:m-1\}$ and $P(F_1)$. Hence, although P_F is small, the conditional probabilities can be made sufficiently large by appropriately adopting m .

Various simulation techniques can be used for sampling from PDFs conditional on a failure region F_i . Subset simulation proceeds in advance by simulating N_S samples $\mathbf{x}^{(0)}$ from $p(\mathbf{x})$ by MC simulation (note that superscript “0” denotes that the samples correspond to the first level):

$$P(F_1) = \int_{\Omega_x} I_{F_1}(\mathbf{x})p(\mathbf{x})d\mathbf{x} \cong \frac{1}{N_S} \sum_{k=1}^{N_S} I_{F_1}(\mathbf{x}_k^{(0)}) \quad (33)$$

From these samples, one can readily obtain new samples that are distributed as $p(\mathbf{x}|F_1)$ (i.e., these samples lie in F_1). These samples can be used to estimate $P(F_2|F_1)$ by sample averaging similar to what has already done in Eq. 33. These samples provide seeds for simulating more samples according to $p(\mathbf{x}|F_2)$ and hence for estimating $P(F_3|F_2)$. Repeating this process, the conditional probabilities of the higher-conditional levels can be computed until the failure region of interest F_m is attained. In the i th conditional level, where $1 \leq i \leq m-1$, let $\mathbf{x}^{(i)}$ be the samples with distribution $p(\mathbf{x}|F_i)$; the conditional probability can be expressed as

$$P(F_{i+1}|F_i) = \int_{\Omega_x} I_{F_{i+1}}(\mathbf{x})p(\mathbf{x}|F_i)d\mathbf{x} \cong \frac{1}{N_S} \sum_{k=1}^{N_S} I_{F_{i+1}}(\mathbf{x}_k^{(i)}) \quad (34)$$

The subset simulation is shown to be especially efficient for modeling rare failure events (i.e., when the probability of failure is very small). The statistical properties of each conditional failure probability as well as the P_F are derived in Au and Beck (2001a, 2003b).

Considering that the target failure region can be stated as $F = \{\mathbf{x}: D(\mathbf{x}) > C(\mathbf{x})\}$, the choice

of the sequence of embedded failure regions, F_i , is a key issue to be addressed. It can significantly affect the efficiency of the subset simulation procedure. If F is introduced with a single parameter, the sequence of intermediate failure regions F_i can be generated by adaptively varying that parameter. The (scalar) demand-to-capacity ratio, which is suitable within the structural reliability assessment, is defined as (Au and Beck 2003b; Jalayer et al. 2007a):

$$Y(\mathbf{x}) = \max_{j=1}^{N_{\text{mech}}} \min_{i=1}^{n_j} \frac{D_{i,j}(\mathbf{x})}{C_{i,j}(\mathbf{x})} \quad (35)$$

where N_{mech} is the number of potential failure mechanisms and n_j is the number of components in the j th mechanism (cf. Eq. 4). This generalized performance variable can be interpreted as the component demand-to-capacity ratio that brings the system closer to the desired limit state. It is interesting to note that the performance variable in Eq. 35 takes into account a range of potential failure mechanisms (*cut-sets*) as noted in section “[Structural Component and Structural System Reliability](#).” As a result, the target failure region can be rewritten as $F = \{\mathbf{x}: Y(\mathbf{x}) > 1\}$, and the sequence of embedded intermediate failure regions can be generated as $F_i = \{\mathbf{x}: Y(\mathbf{x}) > y_i\}$ where $0 < y_1 < \dots < y_m = 1$ are the intermediate thresholds. Au and Beck (2003b) proposed an adaptive methodology for choosing y_i values so that the estimated conditional probabilities are equal to a fixed value. They found that a fixed value equal to 0.1 yields good efficiency.

This section attempts to briefly introduce the main concepts of the subset simulation as an advanced stochastic simulation method for estimation of small probabilities corresponding to rare failure events. However, a detailed introductory description of methodology can be found in the chapter “[Subset Simulation Method for Rare Event Estimation: An Introduction](#)” of the general section “[Reliability and Robustness](#)” of the Encyclopedia of Earthquake Engineering.

Robust Reliability Assessment Using Simulation-Based Approaches

The characterization of uncertainties can be treated in two levels: (1) prior probability distributions for the uncertain parameters based on available information and/or (qualitative) professional judgment and (2) the results of in situ tests, various types of inspections, pseudo-dynamic health-monitoring tests, and the consideration of the corresponding measurement errors that can be used to update the prior distributions of uncertain parameters. The treatment of the latter issue is tackled by the Bayesian updating procedure (e.g., Box and Tiao 1973; Jaynes 2003).

In the *Bayesian* approach, the probability is always conditional on the amount of information available. In this probabilistic framework, probability represents the degree of belief in a certain outcome based on the amount of information. The Bayesian approach was not assessed objectively by many statisticians until late last century because of the absence of a strong rationale behind the theory (see Jaynes 2003; Zuev et al. 2012 for a brief summary of development corresponding to the Bayesian approach). However, the seminal work of the physicist Cox (1961) expounded by the physicist Jaynes (2003) has significantly enhanced the Bayesian probability theory as a convenient mathematical language for inference and uncertainty quantification. The Bayesian approach usually leads to high-dimensional integrals that often can be evaluated neither analytically nor numerically by straightforward quadrature. Nevertheless, the development of new algorithms as well as increasing computing power has led to an explosive growth of using Bayesian approach in different branches of science.

In the *robust* (updated) reliability approach, the plausibility of all the possible structural models conditional on the amount of information available is taken into account in a Bayesian framework (for more details, see Jaynes 2003; Papadimitriou et al. 2001; Beck and Au 2002; Jalayer et al. 2010; Zuev et al. 2012). The robust failure probability then can be calculated (in comparison to Eq. 21) by the following integral:

$$\begin{aligned} P_F &= p(F|D, M) = \int_{\Omega_{\mathbf{x}}} p(F|\mathbf{x}, D, M) p(\mathbf{x}|D, M) d\mathbf{x} \\ &= \int_{\Omega_{\mathbf{x}}} I_F(\mathbf{x}) p(\mathbf{x}|D, M) d\mathbf{x} \end{aligned} \quad (36)$$

where D denotes some test data and M is the set of possible structural models used to specify (both the structural and the probabilistic) modeling assumptions in the analysis; $p(\mathbf{x}|D, M)$ is the posterior PDF of the model parameters, $\mathbf{x}$, based on the observed data and assumed structural model; and $P(F|\mathbf{x}, D, M)$ is the failure probability given the model parameters defined by $\mathbf{x}$, which can be reduced to a deterministic index function, I_F . This index is a function of the model parameters which equals one if failure occurs and zero otherwise. The conditioning on M is included here to stress the fact that all probabilities involved in model updating are always conditional on the choice of the modeling assumptions.

The Bayesian framework is used herein to provide a rigorous method for updating the plausibility of each of the models in representing the structure that is quantified by the probability distribution over the vector of model parameters or the posterior PDF, $p(\mathbf{x}|D, M)$:

$$p(\mathbf{x}|D, M) = p_D(\mathbf{x}) = c^{-1} p(D|\mathbf{x}, M) p(\mathbf{x}|M) \quad (37)$$

where $p(\mathbf{x}|M)$ is the prior probability distribution for $\mathbf{x}$ specified by M , which reflects the relative plausibility of each model before utilizing the data D , $p(D|\mathbf{x}, M)$ is the (updated) probability distribution (also known as the likelihood function) for observed data D based on a model specified by the model parameters $\mathbf{x}$, and c^{-1} is a normalizing constant. Thus, the robust failure probability in Eq. 36 can be shown as

$$P_F = c^{-1} \int_{\Omega_{\mathbf{x}}} I_F(\mathbf{x}, M) p(D|\mathbf{x}, M) p(\mathbf{x}|M) d\mathbf{x} \quad (38)$$

Similarly, the general performance measure of Eq. 25 can also be rewritten based on the robust reliability concept as follows:

$$R_D = \int_{\Omega_{\mathbf{x}}} h(\mathbf{x})p(\mathbf{x}|D, M)d\mathbf{x} = \int_{\Omega_{\mathbf{x}}} h(\mathbf{x})p_D(\mathbf{x})d\mathbf{x} \quad (39)$$

Substituting Eq. 37 into 39, R_D can also be expressed as

$$R_D = c^{-1} \int_{\Omega_{\mathbf{x}}} h(\mathbf{x})p(D|\mathbf{x}, M)p(\mathbf{x}|M)d\mathbf{x} \quad (40)$$

Equations 36 and 38 (similarly, Eqs. 39 and 40) suggest two ways of evaluating the robust reliability by simulation. Equations 36 and 39 suggest estimating the P_F or R_D as the average of I_F or h over samples simulated from p_D , while Eqs. 38 and 40 indicate that the integral and the normalizing constant should be estimated, individually and then combined.

Particular difficulties are encountered in the evaluation of P_F or R_D based on the expressions in Eq. 36 or 39, respectively. The updated probability density function p_D is known only up to a multiplicative constant (see Eq. 37), since the normalizing constant c^{-1} is usually given by a high-dimensional integral that is also difficult to evaluate (Beck and Au 2002; Jalayer et al. 2010). Hence, the application of MC simulation or importance sampling is not generally feasible in this case, because these methods cannot simulate independent samples from p_D .

Additionally, when using importance sampling (see Eq. 30), it is necessary to choose a sampling density that is concentrated in the aforementioned small zone; otherwise similar problems, as in the MC simulation, will arise. However, this task is not trivial to accomplish since information about the zone, where $p(D|\mathbf{x}, M)$ is concentrated, is not directly available (see Beck and Au 2002 for more details). Strictly speaking, the posterior PDF p_D occupies a much smaller volume than that of the prior PDF $p(\mathbf{x}|M)$, so samples in its high probability region cannot

be generated efficiently by sampling from the prior PDF using direct MC method. Of course, if the posterior PDF p_D is evaluated first from Eq. 37, then there would be no problem in using the standard MC from Eq. 36. However, in this case, the factor c^{-1} needs to be calculated.

According to the drawbacks of the current MC methods as well as the importance sampling techniques in solving robust reliability problems, Markov chain Monte Carlo (MCMC) simulation techniques, in particular, the Metropolis-Hastings (MH) algorithm (Metropolis et al. 1953; Hastings 1970), are discussed briefly in the next section. It offers a feasible and powerful way for simulating samples according to an arbitrary distribution (when the target PDF is known only up to a scaling constant), at the expense of introducing dependence among the samples.

Generating Samples in Robust Reliability Based on Adaptive MCMC Simulation Technique

The goal of MCMC is to design a Markov chain such that the stationary distribution of the chain is exactly the distribution that we are interested in sampling from (i.e., the target distribution). The idea is to use specially designed methods for setting up the transition function, such that no matter how each chain is initialized, it will converge to the target distribution. The MH algorithm is normally used to generate samples according to an unscaled PDF when the target PDF is known only up to a scaling constant, i.e., there is no need to know the normalizing constant (see Eq. 33) in advance. The fact that this procedure allows us to sample from un-normalized distributions is one of its major attractions especially in case of Bayesian model updating. The MH algorithm was originally developed by Metropolis and his coworkers (Metropolis et al. 1953) in statistical physics and later generalized by Hastings (1970) in Bayesian statistics (see Fishman 1996 for comprehensive discussions on the MCMC methods). However, its potential use for solving reliability problems in structural engineering has been only recently demonstrated (see Au and Beck 1999, 2001a, 2003b; Beck and Au 2002).

The MH algorithm can be used to generate samples according to the target updated PDF p_D . Using the Markov chain samples generated from the MH procedure, P_F or R_D is estimated as the average over the samples, which is the same approach as the usual MC method, except that the samples are simulated from a Markov chain instead of being independent and identically distributed. In order to reduce the initial effect of the choice of the transient probability distributions (e.g., non-normalized target PDF) on the estimate, the first few samples are often not included in the sample averaging.

In the MH procedure, a proposal PDF is chosen to generate a candidate point that is conditional on the previous state of the sampler. The next step is to either accept the proposal or reject it. However, as noted previously, the updated posterior PDF p_D is concentrated in a small zone, and direct adaptation using a proposal PDF, which varies with a vastly different length scale from that of the target PDF, will not be effective (see Beck and Au 2002 for more details). Therefore, a sequence of intermediate proposal PDFs that vary gradually between the prior PDF and the target PDF p_D are introduced (Au and Beck 1999; Beck and Au 2002). According to the proposed methodology, however, the proposal PDF is chosen as the kernel sampling density constructed using the Markov chain samples from the previous simulation level (for the first simulation level, the prior PDF is used as the proposal PDF). Thus, the adaptation is done from one simulation level to the next. This methodology is especially useful in the subset simulation described previously in section “Calculation of Failure Probability Using the Subset Simulation.” Accordingly, a modified MH algorithm was developed by Au and Beck (2003b) that can obtain samples from a posterior PDF conditioned on a failure region F_i .

PEER Performance-Based Approach

In recent years, an effective foundation for the development of probabilistic Performance-Based Earthquake Engineering (PBEE) for the design

and assessment of building structures has been developed by the Pacific Earthquake Engineering Research Center (PEER; Cornell and Krawinkler 2000; Moehle and Deierlein 2004). There are several stages to this process: (1) calculation of ground motion hazard by quantifying uncertainty in ground motion with a probabilistic model for a parameter related to ground motion and known as the intensity measure (IM); (2) estimation of the uncertainty in structural response D (e.g., force and deformation engineering parameters), for each IM level; (3) estimation of the uncertainty in damage measure DM (i.e., physical states of damage) given response level D ; and (4) estimation of the uncertainty in the decision variable DV as the resulting consequences (e.g., financial losses, fatalities, and business interruption) given damage measure.

Each stage of the process is performed and executed (more or less) independently and then linked back together, as expressed in the following integral form:

$$\lambda(DV) = \iiint G[DV|DM] \cdot \left| \frac{dG[DM|D]}{dDM} \right| \cdot \left| \frac{dG[D|IM]}{dD} \right| \cdot \left| \frac{d\lambda(IM)}{dIM} \right| \cdot dIM \cdot dD \cdot dDM \quad (41)$$

where $G[Y|X]$ denotes generically the conditional Complementary Cumulative Distribution Function (CCDF) of Y given a certain value of X and $\lambda(Y)$ denotes the mean annual exceedance rate (mean annual frequency) of Y . As it can be observed from Eq. 41, the PEER framework enjoys a modular structure and benefits from the (hypothetical) conditional independence between the main parameters (i.e., the conditional independence of $DV|DM$ from D and IM and other ground motion parameters, $DM|D$ from IM and other ground motion parameters, and $D|IM$ from other ground motion parameters, such as, but not limited to, magnitude and distance).

This (hypothetic) Markovian independence between various intermediate parameters is one of the main factors that distinguish the PEER approach from the two other approaches

discussed in this entry, namely, the classic and the simulation-based reliability methods. In these approaches, the probability of failure (which can generically represent a decision variable DV exceeding a certain threshold value) is calculated directly and without considering intermediate parameters such DM , D , and IM . Another important factor that distinguishes the PEER framework equation (Eq. 41) from the other two approaches, discussed herein, is that Eq. 41 yields a rate of exceedance and not a probability. This is while the classic and the simulation-based reliability methods lead to a probability of exceedance. In other words, the rate of exceedance reported in Eq. 41 needs to be translated into probability by assuming an underlying probabilistic model (see Der Kiureghian 2005).

In this section, the second stage of the PEER integral expression dedicated to estimating the conditional probability of exceeding the engineering demand parameter D given IM , denoted as $G[D|IM]$, is discussed in detail. The estimation of the mean annual frequency of exceeding a specific IM values, denoted as $\lambda(IM)$, is the main focus of the chapter “► Probabilistic Seismic Hazard Models.” Finally, an outlook into the whole framework is provided in the chapter entitled “► Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers”.

Taking into account other sources of uncertainty imposes an overall increase in the dispersion for the failure probability (Jalayer 2003; Jalayer and Cornell 2003), while an eventual shift in the median failure probability is not envisioned in this method (see the “hybrid” methods discussed in section “Alternative Methods Based on Time-Invariant Reliability Methods”).

In the PEER PBEE methodology, the evaluation of structural performance is defined by probabilistic performance objectives, which can be expressed in terms of the mean annual frequency (MAF) of exceeding a specified limit state (also referred to as the limit state frequency) denoted as λ_{LS} . Therefore, it is worth noting that the usual outcome of this framework is an estimate of the rate of exceeding a designated limit state threshold rather than a probability.

The main advantage of this methodology is that it evaluates the limit state frequency, λ_{LS} , by decoupling the ground motion hazard and nonlinear dynamic analyses via a site- and structure-specific intermediate variable known as the ground motion intensity measure, IM (see also Shome et al. 1998; Shome and Cornell 1999; Cornell and Krawinkler 2000; Luco 2002; Jalayer 2003; Jalayer and Cornell 2009). The ground motion IM serves as link between seismic hazard analysis, typically provided by seismologists, and structural analysis conducted by engineers.

This IM approach is appealing because it allows the first two analysis stages in the PEER PBEE methodology to be performed (almost) independently. The benefit of this approach is that the number of analyses needed can be substantially reduced because most of the uncertainties are concentrated in ground motion hazard in terms of the MAF of exceeding a certain level of IM , λ_{IM} , which can directly be estimated from probabilistic seismic hazard analysis (PSHA; Cornell 1968; McGuire 2004) for the area where the structure is located.

The conditional probability of D given IM can be predicted through alternative nonlinear dynamic analysis procedures. Herein, three distinct nonlinear dynamic procedures are described: (1) the so-called *cloud analysis* method, in which the structure is subjected to a set of (as-recorded) ground motions (Cornell et al. 2002; Jalayer and Cornell 2003; Elefante et al. 2010), and subsequently, the distribution of D given IM (D and IM are denoted as *cloud data*) is directly obtained by performing a logarithmic linear regression on the cloud data; (2) the incremental dynamic analysis (Vamvatsikos and Cornell 2004) where a selected suite of ground motions are incrementally scaled to different levels of the IM ; and (3) multiple-stripe analysis (MSA; Jalayer 2003; Jalayer and Cornell 2009), where analysis is performed at a specified set of IM levels known as *stripes*, each of which may be performed for a different suite of ground motions (i.e., structural dynamic analyses for multiple stripes of IM through selected set of records for each IM level). It is noteworthy that although IDA uses the same set of ground motions for all

IM levels, the MSA approach allows the re-selection of ground motions at each IM level to be consistent with the seismicity of the designated site (based on the disaggregation of seismic hazard; see, e.g., Baker and Cornell 2005). Thus, the distribution of D given IM is estimated for multiple IM levels.

The structural fragility curve for a range of IM values is a useful probabilistic structural response quantity that can directly be estimated from the aforementioned nonlinear dynamic analysis procedures. The structural fragility for a given limit state, LS , can be defined as the conditional probability of exceeding the limit state capacity for a given level of IM (Ellingwood 2001; Jalayer 2003; Baker and Cornell 2005; Jalayer and Cornell 2009; Baker 2014). If the considered limit state corresponds with the collapse condition, the associated fragility curve is known as the collapse capacity. Subsequently, this result can be combined with a ground motion hazard, i.e., λ_{IM} , to compute the MAF of exceeding any given limit state (Shome and Cornell 1999; Jalayer 2003; Jalayer and Cornell 2003; Ibarra and Krawinkler 2005; Haselton and Deierlein 2007).

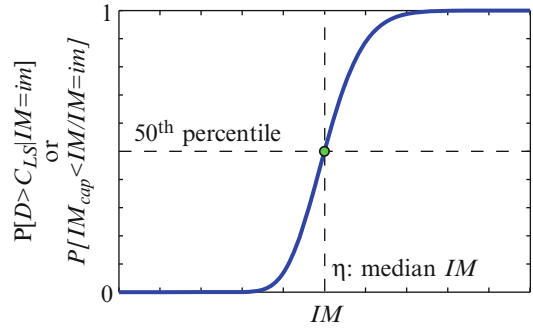
Mean Annual Frequency of Exceeding a Limit State

The MAF of exceeding a specified limit state, denoted as λ_{LS} , can be obtained from two approaches: (a) the IM -based approach (also known as the fragility/hazard format) and (b) the D -based (i.e., demand-based) approach, which in the literature is also called the engineering demand parameter (EDP)-based approach (Cornell and Krawinkler 2000). The two approaches are expressed by the two following equations, respectively:

$$\lambda_{LS} = \int_{im} P[D > C_{LS} | IM = im] |d\lambda_{IM}(im)| \quad (42)$$

$$\lambda_{LS} = \int_{im} P[IM_{cap} < IM | IM = im] |d\lambda_{IM}(im)| \quad (43)$$

where C_{LS} is the capacity associated with the given LS , IM_{cap} is an uncertain parameter



Seismic Reliability Assessment, Alternative Methods for, Fig. 5 Typical fragility curve

representing the capacity of structure in terms of IM values (distribution of IM values) that result in a D level equal to C_{LS} , and the probability terms denote the structural fragilities for the desired limit state. The typical shape of the fragility curve is presented in Fig. 5.

The fragility curve can be estimated by cloud analysis, MSA, or IDA.

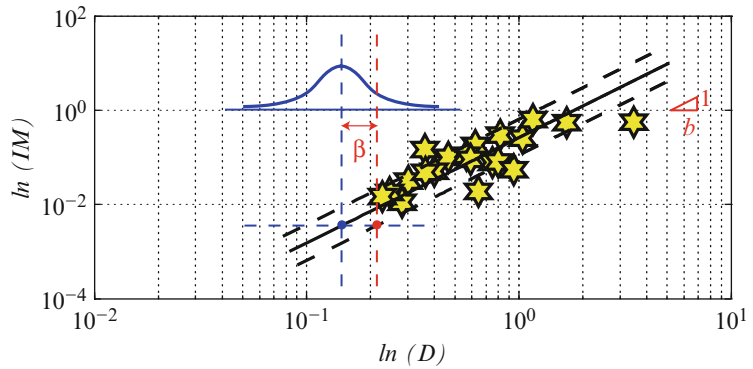
Cloud Analysis Approach

The probability term presented in Eq. 42 can be derived by assuming D to be a lognormal variable. This is a common assumption that has been confirmed as reasonable in many past studies (e.g., Porter et al. 2007; Jalayer and Cornell 2009):

$$P[D > C_{LS} | IM = im] = 1 - \Phi\left(\frac{\ln C_{LS} - \ln \eta_{D|IM}(im)}{\beta_{D|IM}(im)}\right) \quad (44)$$

where Φ is the standardized Gaussian CDF and η and β are conditional median and standard deviation (dispersion) of the natural logarithm of D given IM . In order to estimate the statistical properties of the cloud data, i.e., conditional median and dispersion, conventional linear regression (using least squares method) is used in the natural logarithmic scale, as shown in Fig. 6. Thus,

Seismic Reliability Assessment, Alternative Methods for,
Fig. 6 Typical logarithmic linear regression on cloud data



$$\ln \eta_{D|IM}(im) = \ln a + b \ln(im),$$

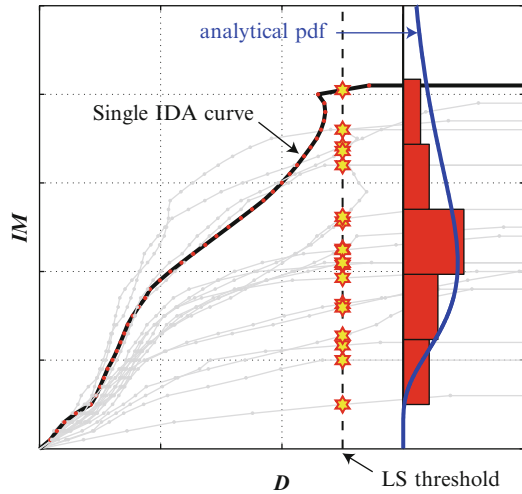
$$\beta_{D|M} = \sqrt{\frac{\sum_{j=1}^{N_{\text{cloud}}} \left(\ln \frac{C_{LS}}{a \cdot im_j^b} \right)^2}{N_{\text{cloud}} - 2}} \quad (45)$$

where a and b are the coefficients of the logarithmic linear regression and N_{cloud} is the number of the cloud data.

Rare events may exert a very large demand D (i.e., even cause instability in the nonlinear dynamic analysis) that can affect the general trend in the linear regression of the cloud data. In such cases, a logistic regression (Agresti 2002) can generally be used. In this methodology, binary values of 0 (for non-collapse data) and 1 (for collapse cases) are assigned for each IM , and the logistic regression allows the construction of the associated fragility curve based on these binary data (Baker and Cornell 2005; Elefante et al. 2010).

IDA Approach

In a great number of research works (see, e.g., Jalayer 2003; Vamvatsikos and Cornell 2004; Baker and Cornell 2005; Ibarra and Krawinkler 2005; Zareian and Krawinkler 2007; Haselton and Deierlein 2007; Jalayer and Cornell 2009; Baker 2014), the collapse capacity is estimated by repeatedly scaling a ground motion using the IDA procedure until the ground motion causes collapse of the structure; hence, each ground motion can be associated with a single IM value



Seismic Reliability Assessment, Alternative Methods for,
Fig. 7 Typical IDA curves and fitting of the empirical distribution of the IM related to the overcoming of limit state

associated with collapse. By repeating this process for a set of ground motions, one can obtain a set of IM values corresponding to the onset of collapse. Subsequently, the probability of collapse for a specific IM equal to im can be estimated as the fraction of records for which collapse occurs at a level lower than the predefined value, im . A lognormal distribution is often fitted to the empirical distribution of IM levels that cause the structure to collapse. In Fig. 7, the black dashed line represents the LS threshold, and the red stars illustrate the

IM_{cap} values that cause the onset of the limit state. On the same diagram, the histogram of the IM_{cap} data and the related empirical distribution are shown as well.

Multiple-Stripe Analysis Approach

In cases where different records are used at each IM level, the IDA method (section “[IDA Approach](#)”) cannot be used to estimate the probability of collapse. A substitute for the conventional IDA approach should be utilized in this particular case. Therefore, instead of an IM value associated with the onset of collapse for each record (i.e., the IDA approach), the fraction of records that cause collapse at each IM level can be taken into account. The latter can be obtained from individual stripes by calculating the number of accelerograms for which the limit state has been exceeded with respect to the total number of records (Baker and Cornell 2005; Ebrahimian 2012; Baker 2014).

As a result, the empirical distribution of the collapse capacity for a set of IM values can be obtained. Once the ratio of the number of collapse cases to the total number of ground motions is calculated for each IM level, a fragility curve can be generated through the logistic regression. The observed fraction of collapse and the estimated

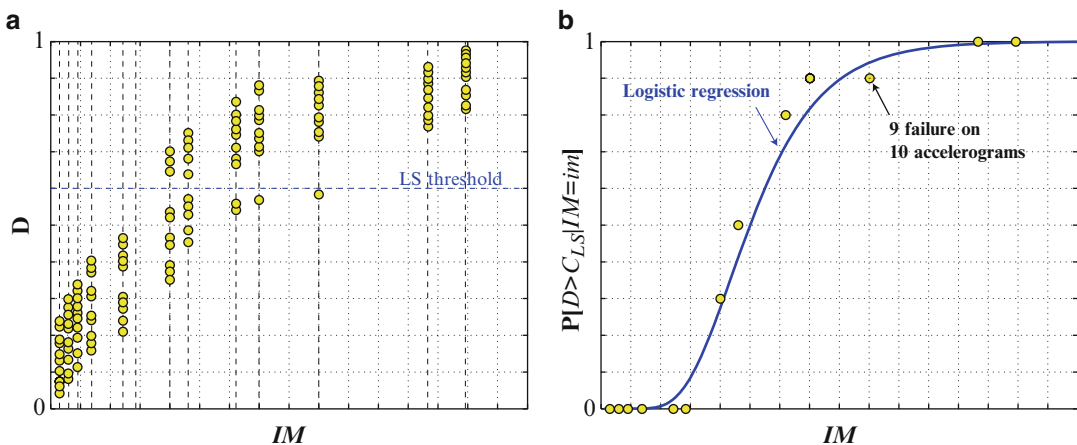
fragility function are shown schematically in Fig. 8 (Ebrahimian 2012).

Demand and Capacity Factor Design (DCFD) Format (IM-Based Version)

The IM -based closed-form solution for the annual frequency of exceeding a desired limit state, λ_{LS} , has been derived by Jalayer (2003) (see also Jalayer and Cornell 2003) under a set of simplifying assumptions. It is used as a performance-oriented design procedure proposed in the SAC/FEMA Steel Project (Cornell et al. 2002) as

$$\lambda_{LS} = \lambda_{IM}(\eta_{IMcap}) \exp\left(\frac{1}{2}k^2\beta_{IMcap}^2\right) \quad (46)$$

where η_{IMcap} and β_{IMcap} are the parameters of the estimated lognormal distribution assuming that the IM_{cap} capacity is a lognormal variable and k reflects the steepness of the hazard curve, λ_{IM} , in the vicinity of the point η_{IMcap} (the hazard curve is approximated in the region of interest by the power-law relationship as $\lambda_{IM}(\eta_{IMcap}) = k_0(\eta_{IMcap})^{-k}$). The first term in Eq. 46 shows the first-order approximation to the limit state probability, and the exponential expression is a magnifying factor that accounts for the



Seismic Reliability Assessment, Alternative Methods for, Fig. 8 (a) Typical MSA output and (b) fragility curve obtained through the logistic regression applied on the MSA results

sensitivity of the limit state probability to the randomness in the IM_{cap} .

The closed-form expression for the limit state frequency can represent an alternative interpretation for the seismic design or assessment considering an IM -based design criterion. A certain design criterion is to check whether the MAF of exceeding a certain limit state (limit state frequency) is less than or equal to an allowable annual frequency, λ_o . Thus,

$$\lambda_{LS} = \lambda_{IM}(\eta_{IM_{cap}}) \exp\left(\frac{1}{2}k^2\beta_{IM_{cap}}^2\right) \leq \lambda_o \quad (47)$$

After some simple rearrangements with the objective of allocating the parameters corresponding to the capacity to one side, Eq. 47 takes the following form (Jalayer 2003; Jalayer and Cornell 2003):

$$im^{\lambda_o} \leq \eta_{IM_{cap}} \exp\left(-\frac{1}{2}k^2\beta_{IM_{cap}}^2\right) \quad (48)$$

where im^{λ_o} is the indicator IM for a hazard level equal to the allowable annual frequency, λ_o . The right-hand side of this expression represents the *factored capacity* (expressed in terms of the adopted IM). Therefore, this expression illustrates a design criterion based on the fragility/hazard format in terms of the IM -based factored capacity being greater than or equal to the IM for a given allowable annual rate.

Summary and Future Challenges

A summary of general methodologies used in the seismic reliability assessment has been presented. In particular, the historical evolution of each method that allows for accounting uncertainties in demand and capacity is briefly illustrated. It is noted that considerable progress has been made in the development of these methodologies, and great efforts have been performed to implement them with the aim of loss reduction and risk mitigation. Considering that the seismic reliability of structures is a specific time-dependent problem with many sources of

uncertainty, applying any of the aforementioned methods is not a trivial task.

Arguably, the first-order and second-order reliability methods cannot be adopted in a straightforward manner when solving time-dependent nonlinear reliability problems. On the other hand, the PEER approach works best in cases where the primary source of uncertainty is reflected in the RTR variability (Der Kiureghian 2005). Therefore, in the recent years, various hybrid methods (section “[Alternative Methods Based on Time-Invariant Reliability Methods](#)”) have been proposed (see, e.g., Liel et al. 2009). Other approaches include those developed by Vamvatsikos and Fragiadakis (2010) and Celarec and Dolšek (2013) where the PEER approach is mixed with other reliability methods (FOSM and Latin hypercube sampling). As a result, the treatment of various sources of uncertainty in both aforementioned reliability methods deserves further research and investigations.

It is noteworthy that both the modern construction codes and the professional community of engineers around the world are becoming more inclined towards the application of nonlinear dynamic analyses. This raises the professional interests to the PEER approach, since it provides a more accessible way of taking into account and propagating the RTR variability in design and assessment problems. Moreover, this approach leads directly to the estimation of the economic losses, which are viewed as essential risk metrics in decision making.

In order to take into account various sources of uncertainty (e.g., modeling, RTR, etc.) in the PEER approach while maintaining a low number of structural analyses, Jalayer et al. (2013) proposed a method inspired from the concept of robust reliability discussed in section “[Calculation of Failure Probability Using the Subset Simulation](#).” Conditioned on a prescribed analytical fragility model, they suggest using the standard MC simulation (with few samples) in order to create various structural model realizations based on the modeling uncertainties present in the problem. A suite of recorded ground motions employed to represent the RTR variability is then applied to each of the above-mentioned structural

models in order to calculate a desired structural response. In other words, the number of structural response realizations is equal to the number of ground motion records present in the suite of records. These structural response realizations are then used as data in order to both update the joint probability distribution for the parameters of the adopted analytic fragility model and also to directly obtain the robust fragility.

In the simulation-based approach, consideration of RTR variability requires the adoption of a suitable stochastic ground motion model (see the chapter “► [Stochastic Ground Motion Simulation](#)”) conditional on seismic source parameters (see Au and Beck 2003b; Jalayer and Beck 2008). Accordingly, the consideration of RTR variability together with other sources of uncertainties will provide a complete probabilistic treatment of the structural response (see, e.g., Jalayer et al. 2007b where RTR variability and modeling uncertainty are considered together in the seismic reliability analysis of RC frames based on the subset simulation). One of the main challenges in applying this approach is to ensure that the stochastic ground motion model provides a realistic description of the characteristics of the ground motions expected to happen at the building site (see, e.g., Atkinson and Silva 2000; Rezaeian and Der Kiureghian 2010). Moreover, there are difficulties in the application of advanced simulation-based algorithms (e.g., subset simulation) so that these techniques cannot be performed in a straightforward and practical manner. However, the fast and incredible increase in computational power is going to render the simulation-based methods ever more accessible for resolving high-dimensional reliability problems.

In comparison to the PEER approach, the advanced simulation-based methods have the advantage of being able to directly estimate losses without passing through multiple stages (e.g., hazard, demand, damage, losses). Undoubtedly, the application of advanced simulation-based techniques is another major concern in the future of reliability methods due to the usually high dimension of the vector of uncertain parameters.

The Bayesian network methodologies have been developed during the past 25 years mostly in the field of artificial intelligence (Russell and Norvig 2003). This method has recently been applied in order to solve high-dimensional structural reliability problems such as infrastructure seismic risk assessment for spatially distributed systems (see, e.g., Grêt-Regamey and Straub 2006; Straub and Der Kiureghian 2010). The Bayesian networks rely on graphical visualization of the uncertain parameters and their correlation structure. They are arguably going to become efficient means of resolving systemic reliability problems due to their propensity for being automatized and also due to the fact that they lend themselves quite well to the implementation of advanced simulation-based methods.

References

- Agresti A (2002) *Categorical data analysis*. Wiley, New York, 710 pp
- Ambartsumian R, Der Kiureghian A, Ohanian V, Sukiasian H (1998) Multinomial probability by sequential conditioned importance sampling: theory and application. *Probab Eng Mech* 13(4):299–308
- ASCE American Society of Civil Engineers (2010) *ASCE/SEI 7–10, Minimum design loads for buildings and other structures*, New-York
- Atkinson G, Silva W (2000) Stochastic modelling of California ground motions. *Bull Seismic Soc Am* 90(2):255–274
- Au SK, Beck JL (1999) A new adaptive importance sampling scheme for reliability calculations. *Struct Saf* 21(2):135–158
- Au SK, Beck JL (2001a) Estimation of small failure probabilities in high dimensions by subset simulation. *Probab Eng Mech* 16(4):263–277
- Au SK, Beck JL (2001b) First excursion probabilities for linear systems by very efficient importance sampling. *Probab Eng Mech* 16(3):193–207
- Au SK, Beck JL (2003a) Important sampling in high dimensions. *Struct Saf* 25(2):139–163
- Au SK, Beck JL (2003b) Subset simulation and its application to seismic risk based on dynamic analysis. *J Eng Mech* 129(8):901–917
- Baker JW (2014) Efficient analytical fragility function fitting using dynamic structural analysis. *Earthq Spectra* doi: <http://dx.doi.org/10.1193/021113EQS025M>
- Baker JW, Cornell CA (2005) Vector-valued ground motion intensity measures for probabilistic seismic demand analysis. Report No. 150, John A. Blume Earthquake Engineering Center, Stanford, 321 pp

- Beck JL, Au SK (2002) Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation. *J Eng Mech* 128(4):380–391
- Box GEP, Tiao GC (1973) Bayesian inference in statistical analysis. Wiley, New York
- Bozorgnia Y, Bertero VV (2004) Earthquake engineering: from engineering seismology to performance-based engineering. CRC Press, ISBN 0-8493-1349-9
- Breitung K (1984) Asymptotic approximations for multinormal integrals. *J Eng Mech* 110(3):357–366
- Celarec D, Dolšek M (2013) The impact of modelling uncertainties on the seismic performance assessment of reinforced concrete frame buildings. *Eng Struct* 52:340–354
- CEN (2002) European Committee for Standardisation, Eurocode-Basis of Structural Design, EN 1990:2002
- Cornell CA (1968) Engineering seismic risk analysis. *Bull Seismol Soc Am* 58(5):1583–1606
- Cornell CA (1969) A Probability-based structural code. *ACI J* 66(12):974–985
- Cornell CA, Krawinkler H (2000) Progress and challenges in seismic performance assessment. *PEER Cent News* 3(2):1–3
- Cornell CA, Jalayer F, Hamburger RO, Foutch DA (2002) Probabilistic basis for 2000 SAC Federal Emergency Management Agency steel moment frame guidelines. *J Struct Eng* 128(4):526–533
- Cox RT (1961) The algebra of probable inference. Johns Hopkins Press, Baltimore
- Der Kiureghian A (1996) Structural reliability methods for seismic safety assessment: a review. *Eng Struct* 18(6):412–424
- Der Kiureghian A (2005) Non-ergodicity and PEER's framework formula. *Earthq Eng Struct Dyn* 34(13):1643–1652
- Der Kiureghian A, Dakessian T (1998) Multiple design points in first and second-order reliability. *Struct Saf* 20(1):37–49
- Der Kiureghian A, Ditlevsen O (2009) Aleatory or epistemic? Does it matter? *Struct Saf* 31(2):105–112
- Der Kiureghian A, Haukaas T, Fujimura K (2006) Structural reliability software at the University of California, Berkeley. *Struct Saf* 28(1):44–67
- Det Norske Veritas (1992) Structural reliability analysis of marine structures, classification notes no 30.6
- Ditlevsen O (1973) Structural reliability and the invariance problem. Research report no 22, Solid Mechanics Division, University of Waterloo, Waterloo
- Ditlevsen O, Madsen HO (1996) Structural reliability methods. Wiley, Chichester
- Dolšek M (2009) Incremental dynamic analysis with consideration of modeling uncertainties. *Earthq Eng Struct Dyn* 38(6):805–825
- Ebrahimian H (2012) Assessment of existing offshore platforms in Persian Gulf region for life extension or operative conditions. Ph.D. dissertation, Department of Civil and Environmental Engineering, Sharif University of Technology, Tehran, 363 pp
- Elefante L, Jalayer F, Iervolino I, Manfredi G (2010) Disaggregation-based response weighting scheme for seismic risk assessment of structures. *Soil Dyn Earthq Eng* 30(12):1513–1527
- Ellingwood BR (2001) Earthquake risk assessment of building structures. *Reliab Eng Syst Saf* 74(3):251–262
- Ellingwood BR, Kinali K (2009) Quantifying and communicating uncertainty in seismic risk assessment. *Struct Saf* 31(2):179–187
- Faber MH (2012) Statistics and probability theory: in pursuit of engineering decision support, vol 18. Springer, New York
- Fishman GS (1996) Monte Carlo: concepts, algorithms, and applications. Springer, New York
- Galambos TV, Ellingwood B, MacGregor JG, Cornell CA (1982) Probability based load criteria: assessment of current design practice. *J Struct Div* 108(5):959–977
- Golafshani AA, Ebrahimian H, Bagheri V, Holmas T (2011) Assessment of offshore platforms under extreme waves by probabilistic incremental wave analysis. *J Constr Steel Res* 67(5):759–769
- Grêt-Regamey A, Straub D (2006) Spatially explicit avalanche risk assessment linking Bayesian networks to a GIS. *Nat Hazards Earth Syst Sci* 6(6):911–926
- Haselton CB, Deierlein GG (2007) Assessing Seismic Collapse Safety of Modern Reinforced Concrete Moment-Frame Buildings, PEER Report 2007/08, Pacific Engineering Research Center. University of California, Berkeley
- Hasofer AM, Lind NC (1974) Exact and invariant second-moment code format. *J Eng Mech Div* 100(1):111–121
- Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1):97–109
- Helton JC, Davis FJ (2003) Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliab Eng Syst Saf* 81(1):23–69
- Hohenbichler M, Rackwitz R (1983) First-order concepts in system reliability. *Struct Saf* 1(3):177–188
- Huntington DE, Lyrantzis CS (1998) Improvements to and limitations of Latin hypercube sampling. *Probab Eng Mech* 13(4):245–253
- Ibarra LF, Krawinkler H (2005) Global collapse of frame structures under seismic excitations. John A. Blume Earthquake Engineering Center, Report no 152, Department of Civil and Environmental Engineering, Stanford University, Stanford, 324 pp
- Jalayer F (2003) Direct probabilistic seismic analysis: implementing non-linear dynamic assessments. Ph.D. dissertation, Department of Civil and Environmental Engineering, Stanford University, Stanford, 244 pp
- Jalayer F, Beck JL (2008) Effects of two alternative representations of ground-motion uncertainty on probabilistic seismic demand assessment of structures. *Earthq Eng Struct Dyn* 37(1):61–79
- Jalayer F, Cornell CA (2003) A technical framework for probability-based demand and capacity factor design (DCFD) seismic formats. Pacific Earthquake Engineering Research Center (PEER) 2003/08

- Jalayer F, Cornell CA (2009) Alternative non-linear demand estimation methods for probability-based seismic assessments. *Earthq Eng Struct Dyn* 38(8):951–972
- Jalayer F, Franchin P, Pinto PE (2007a) A scalar damage measure for seismic reliability analysis of RC frames. *Earthq Eng Struct Dyn* 36(13):2059–2079
- Jalayer F, Franchin P, Pinto PE (2007b) Considering structural modelling uncertainty in seismic reliability analysis of RC frames: use of advanced simulation methods. In: *Proceedings of the conference: COMPDYN 2007*, paper ID: 1218, Crete, Greece, 13–16 June 2007
- Jalayer F, Iervolino I, Manfredi G (2010) Structural modeling uncertainties and their influence on seismic assessment of existing RC structures. *Struct Saf* 32(3):220–228
- Jalayer F, De Risi R, Elefante L, Manfredi G (2013) Robust fragility assessment using Bayesian parameter estimation. Paper presented at the Vienna Congress on Recent Advances in Earthquake Engineering and Structural Dynamics 2013 (VEESD 2013), Vienna, 28–30 Aug
- Jaynes ET (2003) *Probability theory: the logic of science*. Cambridge University Press, New York
- Liel AB, Haselton CB, Deierlein GG, Baker JW (2009) Incorporating modeling uncertainties in the assessment of seismic collapse risk of buildings. *Struct Saf* 31(2):197–211
- Liu P-L, Der Kiureghian A (1991) Optimization algorithms for structural reliability. *Struct Saf* 9(3):161–177
- Luco N (2002) Probabilistic seismic demand analysis, SMRF connection fractures, and near-source effects. Probabilistic seismic demand analysis, SMRF connection fractures, and near-source effects. Ph.D. dissertation, Department of Civil and Environmental Engineering, Stanford University, Stanford, 260 pp
- McGuire RK (2004) *Seismic hazard and risk analysis*. Earthquake Engineering Research Institute, Berkeley, 240 pp
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21:1087–1092
- Moehle J, Deierlein GG (2004) A framework methodology for performance-based earthquake engineering. In: *Proceedings of the 13th world conference on earthquake engineering*, Vancouver, 1–6 Aug 2004, 13 pp
- Novák D, Teplý B, Keršner Z (1998) The role of Latin Hypercube Sampling method in reliability engineering. In: Shiraishi N, Shinozuka M, Wen YK (eds) *The 7th international conference on structural safety and reliability, ICOSSAR 97*, vol 2. Balkema, Rotterdam (Kyoto, Japan)
- Olsson A, Sandberg G, Dahlblom O (2003) On Latin hypercube sampling for structural reliability analysis. *Struct Saf* 25(1):47–68
- Papadimitriou C, Beck JL, Katafygiotis LS (1997) Asymptotic expansions for reliability and moments of uncertain systems. *J Eng Mech* 123(12):1219–1229
- Papadimitriou C, Beck JL, Katafygiotis LS (2001) Updating robust reliability using structural test data. *Probab Eng Mech* 16(2):103–113
- Pinto PE, Giannini R, Franchin P (2007) *Seismic reliability of structures*. IUSS Press, Pavia
- Porter K, Kennedy R, Bachman R (2007) Creating fragility functions for performance-based earthquake engineering. *Earthq Spectra* 23(2):471–489
- Rackwitz R, Flessler B (1978) Structural reliability under combined random load sequences. *Comput Struct* 9(5):489–494
- Rezaeian S, Der Kiureghian A (2010) Simulation of synthetic ground motions for specified earthquake and site characteristics. *Earthq Eng Struct Dyn* 39(10):1155–1180
- Rubinstein RY (1981) *Simulation and the Monte Carlo method*. Wiley, New York
- Russell SJ, Norvig P (2003) *Artificial intelligence: a modern approach*, 2nd edn. Prentice-Hall, Upper Saddle River
- Schuëller GI, Stix R (1987) A critical appraisal of methods to determine failure probabilities. *Struct Saf* 4(4):293–309
- Shinozuka M (1983) Basic analysis of structural safety. *J Struct Eng* 109(3):721–740
- Shome N, Cornell CA (1999) Probabilistic seismic demand analysis of nonlinear structures. Report no RMS-35, Department of Civil and Environmental Engineering, Stanford University, Stanford, 320 pp
- Shome N, Cornell CA, Bazzurro P, Carballo JE (1998) Earthquakes, records, and nonlinear responses. *Earthq Spectra* 14(3):469–500
- Stein M (1987) Large sample properties of simulations using Latin hypercube sampling. *Technometrics* 29(2):143–151
- Straub D, Der Kiureghian A (2010) Bayesian network enhanced with structural reliability methods: methodology. *J Eng Mech* 136(10):1248–1258
- Tvedt L (1990) Distribution of quadratic forms in normal space-application to structural reliability. *J Eng Mech* 116(6):1183–1197
- Vamvatsikos D, Cornell CA (2004) Applied incremental dynamic analysis. *Earthq Spectra* 20(2):523–553
- Vamvatsikos D, Fragiadakis M (2010) Incremental dynamic analysis for estimating seismic performance sensitivity and uncertainty. *Earthq Eng Struct Dyn* 39(2):141–163
- Vořechovský M, Novák D (2009) Correlation control in small-sample Monte Carlo type simulations I: a simulated annealing approach. *Probab Eng Mech* 24(3):452–462
- Zareian F, Krawinkler H (2007) Assessment of probability of collapse and design for collapse safety. *Earthq Eng Struct Dyn* 36(13):1901–1914
- Zhao Y-G, Ono T (2001) Moment methods for structural reliability. *Struct Saf* 23(1):47–75
- Zuev KM, Beck JL, Au S-K, Katafygiotis LS (2012) Bayesian post-processor and other enhancements of Subset Simulation for estimating failure probabilities in high dimensions. *Comput Struct* 92:283–296

Seismic Resilience

Domenico Asprone and Gaetano Manfredi
Department of Structures for Engineering and
Architecture, University of Naples “Federico II”,
Naples, Italy

Synonyms

City resilience; Community resilience; Disaster resilience; Ecosystem resilience; Engineering resilience; Urban resilience

Introduction

Resilience has several meanings in academic discourse. It is derived from the Latin term *resilire*, which means “to bounce back.” It is used in multiple scientific contexts to identify the capability to recover, absorb shocks, and restore equilibrium after a perturbation. First, the concept of resilience was introduced in the nineteenth century in physics to indicate the ability of materials to withstand impulsive loads without suffering damages. Then, resilience was also used in medicine (Pfeiffer 1929) and psychology (Werner 1971; Garmezy 1973).

Recently, resilience is triggering increasing interest in other scientific contexts, referred to communities, urban systems, and built environment, as the capability to recover from natural and human-induced disasters. The advent of the concept of resilience in this context is the result of an increasing need for a response to new and intense threats to modern societies. Increasing interdependence and complexity of contemporary cities along with more severe events induced by climate change is making modern societies ask for prevention, preparation, reduced impacts and damages, and rapid recovery; that is resilience. This urgent need is pushing scientific community to discuss about the best approach to resilience against disasters and, first of all, to define disaster resilience (see “► [Earthquake Disaster Recovery: Leadership and Governance](#)”;

“► [Earthquakes and Their Socio-economic Consequences](#)”).

Contemporary cities can be interpreted as complex systems, composed of dynamic relationships between physical environment, i.e., infrastructural systems (e.g., utility and transportation networks) and more in general all lifelines, natural environment and social environment, consisting of communities and their internal relationships. Hence, according to a general definition, cities can be considered resilient if able to cope with extreme events without suffering devastating losses and damages to their physical systems or reduced quality of life for the inhabitants (Godschalk 2003). However, a comprehensive definition is still not available, given the complexity in defining the properties of urban systems and the response of cities to extreme events.

What are the real operations taking place in urban systems? What about the dynamic equilibrium at the basis of the urban system operations? What is meant by limited damages and preservation of functionality for urban systems after extreme events? Does the optimal response of urban systems to extreme events, i.e., the “resilient” response, depend on the type of extreme event? These are just some of the questions that make the concept of resilience exploding with different and multidisciplinary meanings.

Defining Resilience

The extreme events that urban systems can be subjected into can be divided into four categories (O’Brien et al. 2006), and each of them may need different approaches to define a resilient response:

- Natural events, such as earthquakes, tsunamis, floods, etc.
- Technological events, or human-induced events, such as accidents on transport networks, industrial accidents, terrorist attacks, etc. In developed countries, these events are extensively considered in risk management strategies, aimed at public awareness of risks,

risk mitigation, and building capacity to withstand and recover from extreme events.

- Humanitarian emergencies, i.e., man-made or natural events, hitting very vulnerable and weak populations, such as droughts, famines, epidemics, wars, etc.
- Events induced by climate changes, i.e., events that may still be counted as natural events (floods, landslides, etc.) or as humanitarian emergencies (drought, heat waves, etc.) but that are induced by recent climate changes and can affect unprepared populations with unexpected intensity. In addition, for the climate-induced events, the most effective mitigation techniques are indirect, undertaken on a global scale, through the reduction of greenhouse gases.

Hence, resilience of urban systems against different events varies, and different strategies can be implemented to build resilient cities, in terms of risk mitigation actions, shock preparedness, and recovery capability from damages.

Given the multiplicity of points of view to disaster resilience, numerous definitions of resilience are available in literature; an excellent review of these definitions is presented by Zhou and coworkers (Zhou et al. 2010), who elaborated the most of the following list:

- Holling (1973, 1986), Holling et al. (1995)

Resilience is defined as the amount of disturbance that can be sustained by a system before a change in system control or structure occurs. It could be measured by the magnitude of disturbance the system can tolerate, still persisting in its pre-disturbance state

- Timmerman (1981)

Resilience is the ability of human communities to withstand external shocks or perturbations to their infrastructure and to recover from such perturbations

- Pimm (1984)

Resilience is the speed with which a system returns to its original state following a perturbation

- Pimm (1984), Holling et al. (1995), Gunderson et al. (1997)

Resilience of an ecological system relates to the functioning of the system, rather than the stability

of its component populations, or even the ability to maintain a steady ecological state

- Wildavsky (1991)

Resilience is the capacity to cope with unanticipated dangers after they have become manifest, learning to bounce back

- Dovers and Handmer (1992)

Re-active and pro-active resilience of society can be distinguished, based on the major difference between ecosystems (that react to disturbances) and societies (that can plan in advance, due to human capacity for anticipation and learning)

- Adger (1997, 2000)

Social resilience could be measured through proxies of institutional change and economic structure, property rights, access to resources, and demographic change

- Horne and Orr (1998)

Resilience is a fundamental quality of individuals, groups and organizations, and systems as a whole to respond productively to significant change that disrupts the expected pattern of events without engaging in an extended period of regressive behavior

- Mallak (1998)

Resilience is the ability of an individual or organization to expeditiously design and implement positive adaptive behaviors matched to the immediate situation, while enduring minimal stress

- Milette (1999)

Local resiliency with regard to disasters means that a locale is able to withstand an extreme natural event without suffering devastating losses, damage, diminished productivity or quality of life and without a large amount of assistance from outside the community

- Comfort (1999)

The capacity to adapt existing resources and skills to new systems and operating conditions

- Milette (1999), Geis (2000), Chen et al. (2008)

In the context of disaster management, resilience is used to describe the ability to resist or adapt to stress from hazards, and the ability to recover quickly

- Adger (2000), Kimhi and Shamai (2004)

Social resilience is understood as having three properties: resistance, recovery and creativity, in which (1) resistance relates to a social entity's efforts to withstand a disturbance and its consequences, and can be understood in terms of the degree of disruption that can be accommodated without social entity undergoing long-term change; (2) Recovery relates to an entity's ability to pull through the disturbance, and can be understood in terms of the time taken for an entity to recover from a disruption. (3) Creativity is represented by a gain in resilience achieved as part of the recovery process, and it can be attained by adapting to new circumstances and learning from the disturbance experience

- Carpenter et al. (2001)

The Resilience Alliance consistently refers to social-ecological systems (SES) and defines their resilience by considering three distinct dimensions: (1) the amount of disturbance a system can absorb and still remain within the same state or domain of attraction; (2) the degree to which the system is capable of self-organization; (3) the degree to which the system can build and increase the capacity for learning and adaptation

- Paton et al. (2000)

Resilience describes an active process of self-righting, learned resourcefulness and growth—the ability to function psychologically at a level far greater than expected given the individual's capabilities and previous experiences

- UN/ISDR (2009)

The ability of a system, community or society exposed to hazards to resist, absorb, accommodate to and recover from the effects of a hazard in a timely and efficient manner, including through the preservation and restoration of its essential basic structures and functions. Resilience means the ability to “resile from” or “spring back from” a shock. The resilience of a community in respect to potential hazard events is determined by the degree to which the community has the necessary resources and is capable of organizing itself both prior to and during times of need.

- Bruneau et al. (2003)

An analysis of seismic resilience at four levels: (1) technical, how physical systems perform when subjected to earthquake forces; (2) organizational, the ability to respond to emergencies and carry out critical functions; (3) social, the capacity to reduce the negative social consequences of loss of critical

services; and (4) economic, the capacity to reduce both direct and indirect economic losses. Resilience has four dimensions: (1) robustness, strength to withstand a given level of stress without loss of function; (2) redundancy, the extent to which elements and systems are substitutable; (3) resourcefulness, the capacity to identify problems, establish priorities and mobilize resources; and (4) rapidity, the capacity to meet priorities and achieve goals in a timely manner. A resilient system has: (1) reduced probability of failures; (2) reduced consequences from failures; and (3) reduced time to recovery

- Kendra and Wachtendorf (2003)

The ability to respond to singular or unique events

- Cardona (2003)

The capacity of the damaged ecosystem or community to absorb negative impacts and recover from these

- Pelling (2003)

The ability of an actor to cope with or adapt to hazard stress

- Rockstrom (2003)

Strategies of social resilience building include manageable strategies, such as institutional development, land reform, land tenure, diversification, marketing, human capacity building, and unmanageable ones, such as relief food, cereal banks, social networks, virtual water imports

- Rose (2004, 2007)

Resilience includes inherent resilience (ability under normal circumstances) and adaptive resilience (ability in crisis situations due to ingenuity or extra effort)

- Aguirre (2006)

A resilient social entity absorbs, responds and recovers from the shock; and improvises and innovates in response to disturbances

- Maguire and Hagan (2007)

In broad terms, social resilience is the capacity of a social entity (e.g., a group or community) to bounce back or respond positively to adversity

- Kang et al. (2007)

Resilience is the ability of the system to recover once hazard has occurred and can be measured by the duration of an unsatisfactory condition.

- Asprone and Manfredi (2014)

An extreme event and the resulting changes, moving urban systems to new dynamic equilibrium, represent a phase in urban life cycle; resilience represents the sustainability of this phase, from the economic, social and environmental point of view, for all the present and future actors, directly and indirectly involved in the recovery process.

Resilience of Ecosystems and Engineering Resilience

To apply the concept of resilience to complex systems, such as cities, two approaches can be followed: the resilience of ecosystems (a) and the engineering resilience (b). In the first, proposed and developed by Holling (1973, 1986, 2001), resilience can be defined as the ability of a system in dynamic equilibrium, subject to external shocks, to move back to a dynamic equilibrium state. On the contrary, engineering resilience, developed by Pimm and other authors (Pimm 1984; Bruneau et al. 2003) can be defined as the ability of a system to absorb an external shock and quickly return to the initial state.

Apparently the first definition may be more complete and suitable for urban systems; in fact, moving from the fact that a complex system in dynamic equilibrium (as the urban system, which consists of physical and social subsystems linked by a dynamic network of relationships) can present different equilibrium states (i.e., can “work”) in various configurations, it can be concluded that a positive response to a negative external shock can be also represented by a new equilibrium state, different from the previous one. For example, looking at the terrorist attack on the World Trade Center in New York on 11 September 2001, it can be said that the city of New York had a “resilient” response. New York quickly recovered from the social and economic effects induced by the event, even if a new equilibrium was reached in a different configuration of the physical system, i.e., without rebuilding the World Trade Center towers and relocating the activities that took place there, that is, re-thinking that space (i.e., “ground zero” site).

Furthermore, the social value of the towers, representing a crucial symbol for the collective identity of the city of New York, has been preserved by reconfiguring the city in a different dynamic equilibrium. The towers’ values still exist and their physical absence was recovered from the social and cultural point of view. Nevertheless, engineering resilience is also extremely meaningful. In fact, one could argue that a complex and dynamic system, as the city, is always able to reach a state of equilibrium after a shock, because the ability of cities to adapt to changes is extremely high. But the new post-event dynamic equilibrium could be “worse” than the previous one. In this case only with an engineering resilience approach a “negative” response can be appreciated; for example, quality and performance indicators of the urban system can be used for this scope. Hence, aiming at synthesizing the different approaches, it can be concluded that the urban system is resilient if, after a shock, it can reach a dynamic equilibrium state, even if different from the previous one, but the urban system is really resilient if, at the same time, certain indicators of quality and performance of the system return to pre-shock values. This concept was also introduced by Dalziel and McManus (2004), which affirmed the need to introduce metrics of resilience. Hence, the questions the scientific community is currently dealing with refer to the indicator system that should be used. Are the indicators currently in use to assess city sustainability suitable for the scope? Which quality and performance indices can describe the “effectiveness” of the response to external shocks?

A further approach to the concept of resilience of cities leads to the definition of the social resilience. It moves from the centrality of communities in urban systems, i.e., the predominance of the social over the physical system; according to this approach, social resilience is defined as the ability of communities to deal with external shocks, managing the changes induced on infrastructures, on the external environment, and on the economic and social systems (Adger 1997, 2000). Social resilience, according to Adger, can be measured by three characteristics: resistance

to external shocks (a), the ability to recover from external shocks (b), and creativity (c), that is, the ability to adapt to new circumstances (Adger 2000; Kimhi and Shamai 2004). With a similar approach, Lorenz (2010) affirms that social resilience consists of the adaptive capacity (a), that is, the ability to change in order to withstand the external shocks; the coping capacity (b), that is, the ability to preserve and give continuity to the system of relationships, given the external shocks; and the participative capacity (c), that is, the self-organization ability, aimed at coping with external shocks. Hence, the approach to social resilience affirms the centrality of communities, able to manage the other physical elements and determine resilient urban systems.

In all the approaches so far analyzed, however, resilience is perceived as the ability of the city to have a “positive” response, when exposed to an external shock, as an extreme event. The main issue is the need to define a “positive” response: the ability to return to the previous equilibrium configuration or even a different reconfigured equilibrium state. And, in a complex and dynamic system, what is an equilibrium state? Furthermore, is resilience to be reached separately both in the social and the physical systems, or does social system resilience entail physical system resilience? Thus, is the physical system resilience condensed in community resilience? Hence, is community, representing the only decision maker for urban management, the only master of a city’s destiny, the key to a resilient city?

Quantifying Resilience

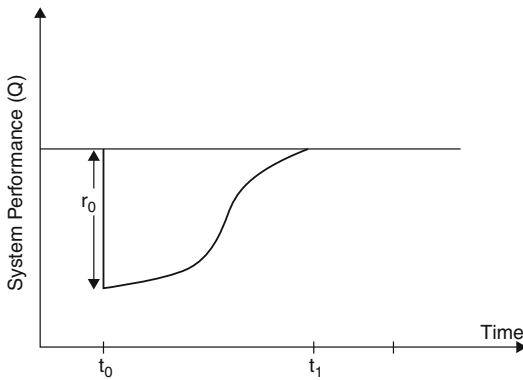
The increasing interest in resilience requires methodological frameworks to measure and assess it. Measuring disaster resilience would help understand and improve resilience of urban systems against risks and implement the most effective strategies to “bounce back” from disasters. Aimed at this goal, different studies have been developed, proposing measurement frameworks of disaster resilience and other properties related to resilience.

The most important methodologies available in literature can be divided into two categories: (a) the physical resilience approach and (b) the social-economic resilience approach. In the former, attention is focused on performances of physical systems, e.g., single structures, urban lifelines, and transportation systems. In this case, resilience is measured as the capability of the physical components and systems to recover their functionality. Mainly, these methods are developed and proposed within engineering community. In the latter, attention is focused on social systems, and resilience is measured as the capability of communities to recover a good life quality level. Mainly, these methods are proposed in social sciences community.

Physical Resilience

Bruneau et al. (2003) define resilience as characterized by four main properties: robustness, rapidity, redundancy, and resourcefulness (4 R’s), to be managed and computed as proxies of resilience. *Robustness* is related to the “strength, or the ability of elements, systems, and other units of analysis to withstand a given level of stress or demand without suffering degradation or loss of function.” *Rapidity* is “the capacity to meet priorities and achieve goals in a timely manner in order to contain losses and avoid future disruption.” *Redundancy* refers to the availability of substitutable elements or systems in the aftermath of a disruption, and *resourcefulness* is the capacity to mobilize material and human resources. Within this approach, different methods have been proposed, whose final scope is to compute resilience as the ability to cope with degradation in system performance $Q(t)$, over time. Numerically, resilience R is often computed as the area underneath function $Q(t)$, divided by the time to restore the pre-event performance (Fig. 1):

$$R = \frac{\int_{t_0}^{t_1} Q(t) dt}{t_1 - t_0} \quad (1)$$



Seismic Resilience, Fig. 1 Physical resilience

Being t_0 the time of the event and t_1 the time of the total recovery of the pre-event performance. This approach has been applied to buildings (Bruneau and Reinhorn 2007), bridges (Decò et al. 2013), road networks (Arcidiacono et al. 2012), and urban infrastructure systems (Ouyang and Dueñas-Osorio 2012; Franchin and Cavalieri 2013), using different performance functions $Q(t)$.

Further works are increasingly conducted on this topic and made available in literature, focusing on different urban systems and different performance functions $Q(t)$; however, the most of the works in recent literature share the theoretical scheme in Eq. 1 to compute resilience. Recently, Cavallaro et al. (2014) and Franchin and Cavalieri (2013) applied this approach to social-physical graphs (merging physical networks and social components) using as performance $Q(t)$ the efficiency of the network in the “social” nodes, aiming at measuring the capability of the physical systems to serve their end users.

Social-Economic Resilience

Studies aimed at computing resilience from a social perspective focus on economic, demographic, and institutional variables, in time and space. In example, economic growth and the

distribution of income among people are fundamental aspects of resilience (Adger 2000) and are often used to compute resilience. Attitude to mobility and migration or amount of young people is also related to resilience (Ruitenbeek 1996; Adger 2000). Social memory of past changes and impacts (Olick and Robbins 1998) also relates to the capacity of communities to adapt and cope with disasters, that is, resilience. Hence, different authors refer to this kind of variables to estimate community resilience, in terms of preparedness and coping capacity to disasters. Specific indicators have been also developed, moving from social-economic variable. This is the case of the Disaster Deficit Index (DDI), proposed by Cardona et al. (2008), measuring country resilience against disasters from a macroeconomic perspective:

$$DDI = \frac{L_R}{R_E} \quad (2)$$

being L_R the maximum expected direct economic impact of possible disasters and R_E the available internal and external resources that can be made available to face disasters.

A recent attempt to integrate physical and social-economic perspectives of resilience has been done with the PEOPLES Resilience Framework (Renschler et al. 2010), linking different resilience dimensions (technical, organizational, societal, and economic) and resilience properties (robustness, redundancy, resourcefulness, and rapidity) as proposed by Bruneau et al. (2003).

Summary

An increasing research interest in scientific community is focusing on disaster resilience of urban environment, from different scientific perspectives, including those from risk engineering and social sciences communities. A comprehensive approach to resilience is still not available, but the need for a general methodology to manage and measure urban resilience is urged due to the

increasing interdependence and complexity of contemporary cities along with more severe events induced by climate change.

This entry deals with the most recent approaches to urban disaster resilience. It moves from the description of a large variety of definitions of disaster resilience, as proposed by researchers of different disciplines. The differences between engineering resilience and resilience of ecosystems are discussed. Thus, different approaches to compute and measure resilience are described, focusing on physical resilience, i.e., resilience of infrastructures and lifelines, and social-economic resilience, i.e., resilience of social systems and communities.

Cross-References

- [Earthquake Disaster Recovery: Leadership and Governance](#)
- [Earthquakes and Their Socio-economic Consequences](#)

References

- Adger WN (1997) Sustainability and social resilience in coastal resource use. CSERGE working paper series, Centre for Social and Economic Research on the Global Environment, University of East Anglia, Norwich and University College London
- Adger WN (2000) Social and ecological resilience: are they related? *Prog Hum Geogr* 24(3):347–364
- Aguirre B (2006) On the concept of resilience. Disaster Research Center, University of Delaware, Delaware
- Arcidiacono V, Cimellaro GP, Infuso A, Reinhorn AM, Road network's disaster resilience assessment methodology (2012) Bridge maintenance, safety, management, resilience and sustainability. In: Proceedings of the sixth international conference on bridge maintenance, safety and management, Como (Italy), pp 1912–1919
- Asprone D, Manfredi G (2014) Linking disaster resilience and urban sustainability: a global approach for future cities. *Disasters J* 39(1):s96–s111
- Bruneau M, Reinhorn AM (2007) Exploring the concept of seismic resilience for acute care facilities. *Earthq Spectra* 23(1):41–62
- Bruneau M, Chang S, Eguchi R, Lee G, O'Rourke T, Reinhorn A, Shinozuka M, Tierney K, Wallace W, von Winterfeldt D (2003) A framework to quantitatively assess and enhance seismic resilience of communities. *Earthq Spectra* 19:733–752
- Cardona OD (2003) The notions of disaster risk: conceptual framework for integrated management. Information and indicators program for disaster risk management. Inter-American Development Bank, Manizales
- Cardona OD, Ordaz MG, Marulanda MC, Barbat AH (2008) Estimation of probabilistic seismic losses and the public economic resilience – an approach for a macroeconomic impact evaluation. *J Earthq Eng* 12(S2):60–70
- Carpenter S, Walker B, Anderies JM, Abel N (2001) From metaphor to measurement: resilience of what to what? *Ecosystems* (N Y, Print) 4(8):765–781
- Cavallaro M, Asprone D, Latora V, Manfredi G, Nicosia V (2014) City ecosystem resilience analysis in case of disasters. *Comput-Aided Civ Infrastruct Eng* 29(8):608–625
- Chen SC, Ferng JW, Wang YT, Wu TY, Wang JJ (2008) Assessment of disaster resilience capacity of hillslope communities with high risk for geological hazards. *Eng Geol* 98(3–4):86–101
- Comfort L (1999) Shared risk: complex systems in seismic response. Pergamon, New York
- Dalziel EP, McManus ST (2004) Resilience, vulnerability, and adaptive capacity: implications for system performance. Stos, Switzerland: 1st International Forum for Engineering Decision Making (IFED), 17 pp
- Decò A, Bocchini P, Frangopol DM (2013) A probabilistic approach for the prediction of seismic resilience of bridges. *Earthq Eng Struct Dyn* 42(10):1469–1487
- Dovers SR, Handmer JW (1992) Uncertainty, sustainability and change. *Glob Environ Chang* 2(4):262–276
- Franchin P, Cavalieri F (2013) A framework for physical simulation of critical infrastructures, accounting for interdependencies and uncertainty. In: 11th international conference on structural safety & reliability ICOSSAR, Taylor & Francis, New York
- Garnezy N (1973) Competence and adaptation in adult schizophrenic patients and children at risk. In: Dean SR (ed) Schizophrenia: the first ten Dean award lectures. MSS Information Corporation, New York
- Geis DE (2000) By design: the disaster resistant and quality-of-life community. *Nat Hazard Rev* 1(3):106–120
- Godschalk D (2003) Urban hazard mitigation: creating resilient cities. *Nat Hazard Rev* 4:136–143
- Gunderson LH, Holling CS, Pritchard L, Peterson GD (1997) Resilience in ecosystems, institutions and societies. Discussion paper 95. Beijer International Institute of Ecological Economics, Stockholm
- Holling CS (1973) Resilience and stability of ecological systems. *Annu Rev Ecol Syst* 4:1–23
- Holling CS (1986) The resilience of terrestrial ecosystems: local surprise and global change. In: Clark WC, Munn RE (eds) Sustainable development of the

- biosphere. Cambridge University Press, Cambridge, pp 292–317
- Holling CS (2001) Understanding the complexity of economic, ecological, and social system. *Ecosystems* (N Y, Print) 4:390–405
- Holling CS, Schindler DW, Walker BW, Roughgarden J (1995) Biodiversity in the functioning of ecosystems: an ecological synthesis. In: Perrings C, Ma'ler KG, Folke C, Holling CS, Jansson BO (eds) *Biodiversity loss: economic and ecological issues*. Cambridge University Press, Cambridge, pp 44–83
- Horne JF, Orr JE (1998) Assessing behaviors that create resilient organizations. *Employ Relat Today* 24(4):29–39
- Kang B, Lee SJ, Kang DH, Kim YO (2007) A flood risk projection for Yongdam dam against future climate change. *J Hydro Environ Res* 1(2):118–125
- Kendra MJ, Wachtendorf T (2003) Elements of resilience after the world trade center disaster: reconstructing New York city's emergency operation center. *Disasters* 27(1):37–53
- Kimhi S, Shamai M (2004) Community resilience and the impact of stress: adult response to Israel's withdrawal from Lebanon. *J Community Psychol* 32(4):439–451
- Lorenz DF (2010) The diversity of resilience: contributions from a social science perspective. *Nat Hazards* 2010:1–18
- Maguire B, Hagan P (2007) Disasters and communities: understanding social resilience. *Aust J Emerg Manag* 22(2):16–20
- Mallak L (1998) Resilience in the healthcare industry. In: Paper presented at the seventh annual engineering research conference, Banff, 9–10 May
- Miletti D (1999) *Disasters by design: a reassessment of natural hazards in the United States*. Joseph Henry Press, Washington, DC
- O'Brien G, O'Keefe P, Rose J, Wisner B (2006) Climate change and disaster management. *Disasters* 30(1):64–80
- Olick JK, Robbins J (1998) Social memory studies: from “collective memory” to historical sociology of mnemonic practices. *Annu Rev Sociol* 24:105–140
- Ouyang M, Dueñas-Osorio L (2012) Time-dependent resilience assessment and improvement of urban infrastructure systems. *Chaos* 22(3):art no 033122
- Paton D, Smith L, Violanti J (2000) Disasters response: risk, vulnerabilities and resilience. *Disaster Prev Manag* 9(3):173–179
- Pelling M (2003) *The vulnerability of cities: natural disasters and social resilience*. Earthscan, London
- Pfeiffer K (1929) *Untersuchungen über die Resilienz der durch die Prothesen beanspruchten Gewebe und ihre Bedeutung für die Okklusion der Prothesen*. Berichthaus, Zürich
- Pimm SL (1984) The complexity and stability of ecosystems. *Nature* 307:321–326
- Renschler C, Frazier A, Arendt L, Cimellaro GP, Reinhorn AM, Bruneau M (2010) Framework for defining and measuring resilience at the community scale: the PEOPLES resilience framework. MCEER technical report –MCEER-10-006, pp 91, University at Buffalo (SUNY), The State University of New York, Buffalo
- Rockstrom J (2003) Resilience building and water demand management for drought mitigation. *Phys Chem Earth* 28:869–877
- Rose A (2004) Defining and measuring economic resilience to disasters. *Disaster Prev Manag* 13:307–314
- Rose A (2007) Economic resilience to natural and man-made disasters: multidisciplinary origins and contextual dimensions. *Environ Hazards* 7:383–398
- Ruitenbeek HJ (1996) Distribution of ecological entitlements: implications for economic security and population movement. *Ecol Econ* 17:49–64
- Timmerman P (1981) *Vulnerability, resilience and the collapse of society: a review of models and possible climatic applications*. Institute for Environmental Studies, University of Toronto, Toronto
- United Nations International Strategy for Disaster Reduction (UNISDR) (2009) *Global assessment report on disaster risk reduction: risk and poverty in a changing climate*. UNISDR, Geneva
- Werner E (1971) *The children of Kauai. A longitudinal study from the prenatal period to age ten*. University of Hawaii Press, Honolulu
- Wildavsky A (1991) *Searching for safety*. Transaction, New Brunswick
- Zhou HJ, Wang JA, Wan JH et al (2010) Resilience to natural hazards: a geographic perspective. *Nat Hazards* 53(1):21–41

Seismic Response Prediction of Degrading Structures

Ching Hang Ng¹, Nopdanai Ajavakom² and Fai Ma³

¹US Nuclear Regulatory Commission, Office of Nuclear Reactor Regulation, Washington, DC, USA

²Department of Mechanical Engineering, Chulalongkorn University, Bangkok, Thailand

³Department of Mechanical Engineering, University of California, Berkeley, Berkeley, CA, USA

Synonyms

Degrading structures; Hysteresis; Nonlinear response; System identification

Introduction

All structures degrade when acted upon by cyclic forces such as those associated with earthquakes, high winds, and sea waves. Development of a practical model of degrading structures that would match experimental observations is an important task. Furthermore, identification and prediction of deterioration is a problem of considerable practical significance. Under cyclic excitation, degradation manifests itself in the evolution of the associated hysteresis loops. Theoretical research in internal friction in the last few decades has noticeably increased the conceptual understanding of hysteresis (Kojic and Bathe 2005; Krasnoselskii and Pokrovskii 1989). Practical issues related to internal friction, however, have not been adequately addressed. The lack of a practical theory of hysteretic evolution is at present a major barrier to successful design of structures against performance deterioration.

In the past 30 years, cyclic performance testing of structural joints and subassemblies around the world has generated a substantial amount of experimental data on load–displacement traces. In the same period, the generalization of the Bouc–Wen differential model of hysteresis permits curve fitting of practically any hysteretic trace with a suitable choice of its 13 control parameters (Bouc 1967; Wen 1976). Using system identification techniques, it appears highly feasible to utilize the generalized differential model of hysteresis and the extensive database of experimental hysteretic traces to deduce a working model for degrading structures. A fundamental objective of this research project is to do just that.

Many questions naturally spring to mind. What are the advantages of the differential model over other empirical models of hysteresis? How can an empirical model of hysteresis be used for prediction of nonlinear system response? What are the factors that would affect the precision of prediction? These questions will be answered along the way. Two principal tasks in connection with hysteretic evolution will be addressed. First, a robust identification algorithm will be used to generate

hysteretic models of a deteriorating structure from its experimental load–displacement traces. Second, it will be shown that a hysteretic model obtained by system identification can be used to predict the future performance of the same deteriorating structure. The organization of this entry is as follows. In section “[Differential Model of Hysteresis](#),” the smoothly varying Bouc–Wen differential model of hysteresis in both its classical (non-degrading) and generalized forms is described. A robust identification algorithm is constructed in section “[System Identification](#)” to generate hysteretic models of a deteriorating structure from its experimental load–displacement traces. This algorithm is based upon the generalized Bouc–Wen model and differential evolution, streamlined through global sensitivity analysis. The model thus generated can account for degradation and pinching effects, which are prominent features of real-life structural deformation. To obtain experimental data for model validation, cyclic performance tests of simple joints and subassemblies are reported in section “[Prediction of Performance](#).” Using an experimental load–displacement trace, a working hysteretic model is identified. It will be shown that the hysteretic model obtained by identification may be used for predicting the nonlinear response of the same structure when driven by other cyclic loads. Finally, the requirements for accurate prediction of system response will be discussed. The terms “Bouc–Wen model” and “differential model” will be used interchangeably throughout the entry.

Differential Model of Hysteresis

When a structure is subjected to severe cyclic loading, the hysteresis loops associated with the structural response are memory dependent. That means the evolution of hysteresis loops depends not only on the instantaneous deformation but also on the history of deformation. This memory nature of degradation makes modeling and analysis challenging. The generalized Bouc–Wen model is one of the widely used empirical models capable of modeling the memory effects.

Suppose the equation of motion of a multi-degree-of-freedom system can be decoupled and, along the direction of the generalized coordinate x , the system is governed by

$$m\ddot{x} + c\dot{x} + r(x, z) = f(t) \quad (1)$$

where m and c are, respectively, the mass and damping coefficients, z is an imaginary hysteretic displacement, and $r(x, z)$ is the total restoring force. It is assumed that the excitation $f(t)$ is cyclic. In the development of differential model, the restoring force $r(x, z)$ is separated into an elastic (linear) component and a hysteretic (nonlinear) component by

$$r(x, z) = \alpha kx + (1 - \alpha)kz \quad (2)$$

where k is the stiffness coefficient and $0 \leq \alpha \leq 1$ is a weighting parameter. Obviously, the restoring force is purely hysteretic if $\alpha = 0$; it is purely elastic if $\alpha = 1$. A diagrammatic representation of the system is shown in Fig. 1.

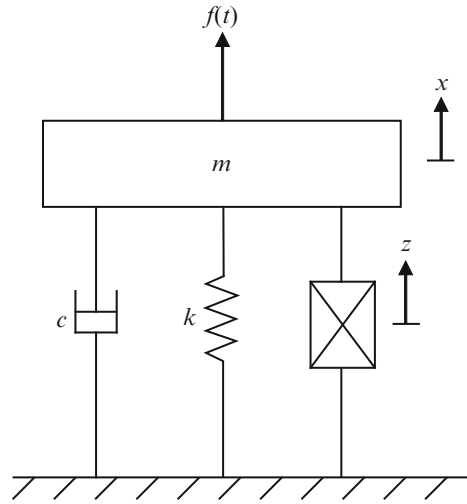
Hysteresis loops may be generated if the hysteretic displacement z and the total displacement x are connected by the nonlinear differential equation (Bouc 1967; Wen 1976)

$$\dot{z} = A\dot{x} - \beta|\dot{x}||z|^{n-1}z - \gamma\dot{x}|z|^n \quad (3)$$

There are five unspecified loop parameters A , α , β , γ , and n in Eqs. 2 and 3, which together represent the classical Bouc–Wen model. The parameters A , β , and γ are basic hysteresis shape parameters. The parameter n controls the sharpness of yield. Over the years, the original Bouc–Wen model has been extended, and new parameters have been added to fit hysteretic shapes arising from deteriorating systems. The result is a contemporary model with thirteen control parameters given by

$$\dot{z} = h(z) \left\{ \frac{A\dot{x} - v \left(\beta|\dot{x}||z|^{n-1}z + \gamma\dot{x}|z|^n \right)}{\eta} \right\} \quad (4)$$

In the above expression, v and η are degradation shape functions (Baber and Wen 1981), and $h(z)$



Seismic Response Prediction of Degrading Structures, Fig. 1 Schematic diagram of a hysteretic degrading system

is a pinching shape function. In general, degradation depends on the response duration and severity. A convenient measure of such combined effect is the energy

$$E(t) = \int_0^t (1 - \alpha)kz\dot{x}dt \quad (5)$$

dissipated through hysteresis from initial time $t = 0$ to present time t . Since

$$\varepsilon(t) = \int_0^t z \dot{x}dt \quad (6)$$

is proportional to $E(t)$, it may also be used as a measure of response duration and severity. Both degradation shape functions v and η are assumed to depend linearly on ε as the system evolves:

$$v(\varepsilon) = 1 + \delta_v \varepsilon \quad (7)$$

$$\eta(\varepsilon) = 1 + \delta_\eta \varepsilon \quad (8)$$

Two unspecified degradation parameters δ_v and δ_η are thus introduced. Under cyclic excitation, the pinching of hysteresis loops is often observed.

For example, pinching may be associated with slippage of longitudinal reinforcement in reinforced concrete or with X-braced steel frames driven by high-shear loads. Baber and Noori (1986) introduced a slip-lock element behaving quite similarly to a hardening nonlinear spring, with the special characteristics that the “slip” zone stiffness is nearly zero, while the “lock” zone stiffness is infinite. The “slip” zone stiffness is initially zero and increases continuously as the system degrades with time. Thus, the pinching shape function $h(z)$ takes the form (Foliente 1995)

$$h(z) = 1 - \zeta_1 e^{-[z \operatorname{sgn}(\dot{x}) - q z_u]^2 / \zeta_2^2} \quad (9)$$

where sgn is the signum function and z_u is the ultimate value of z given by

$$z_u = \left(\frac{A}{v(\beta + \gamma)} \right)^{1/n} \quad (10)$$

The two functions $\zeta_1(\varepsilon)$ and $\zeta_2(\varepsilon)$ control the progress of pinching and are written as

$$\zeta_1(\varepsilon) = \zeta_s \left[1 - e^{(-p \varepsilon)} \right] \quad (11)$$

$$\zeta_2(\varepsilon) = (\psi + \delta_\psi \varepsilon)(\lambda + \zeta_1) \quad (12)$$

Six pinching parameters ζ_s , q , p , ψ , δ_ψ , and λ are therefore present. Altogether there are thirteen loop parameters of hysteresis: A , α , β , γ , n , δ_v , δ_η , ζ_s , q , p , ψ , δ_ψ , and λ . This generalized model of hysteresis possesses all the important features observed in real structures, which include strength degradation, stiffness degradations, and pinching of the successive hysteresis loops.

Compared with other parametric models, a differential model of hysteresis has many advantages (Foliente 1995; Song and Der Kiureghian 2006). The primary one is its ability to generate a large variety of realistic hysteresis loops. Another advantage is the coupling of the equation of motion (1) to either Eqs. 3 or 4 to form an overall differential system. This greatly facilitates any theoretical and numerical manipulations.

Nonetheless, it should be noted that the differential model is one-dimensional and cannot account for interactions between loadings in different directions. There are generalizations in which the differential model is augmented to a multi-degree linear system (Roberts and Spanos 1990). In such generalizations, however, it is also assumed that the system could be decoupled. Since no new physics is obtained, these generalizations would not be pursued herein. Finally, it must be emphasized that the extended Bouc–Wen model of hysteresis is an empirical model. As such, it is not derivable from the fundamental postulates of mechanics and the exact physical meanings of its thirteen parameters are not fully understood. The probable role played by each parameter is summarized in Table 1. Also contained in Table 1 are the sensitivity rankings (Ma et al. 2004) of the control parameters. As will be explained in the next section, in system identification, it is required to estimate only a subset of the 13 unspecified parameters.

System Identification

In analysis, the response of a system is sought if the system model and excitation are known. This is sometimes termed a forward problem. In this interpretation, the inverse problem of finding a system model given the excitation and response is called system identification. System identification in the time domain involves the determination of unspecified parameters of an assumed system model. This can always be formulated as an optimization problem. In the present context, suppose the differential model of hysteresis is adopted and a set of measured excitation–response data from cyclic performance tests of an inelastic structure is given. How can the loop parameters of hysteresis be estimated from the measured data? For each choice of the parameters, the response of the degrading structure subjected to the given excitation can be obtained by numerical simulation. The calculated response data can then be compared to the measured data to see if there are large errors. Obviously, the assumed loop parameters

Seismic Response Prediction of Degrading Structures, Table 1 Parameters of the generalized differential model of hysteresis

Parameter	Description	Local sensitivity ranking (highest = 1)	Global sensitivity ranking (highest = 1)
α	Ratio of linear to nonlinear response	1	2
A	Basic hysteresis shape control	Not varied	Not varied
β	Basic hysteresis shape control	5	4
γ	Basic hysteresis shape control	6	5
n	Sharpness of yield	8	7
δ_v	Strength degradation	12	9
δ_η	Stiffness degradation	4	8
ζ_s	Measure of total slip	2	1
q	Pinching initiation	9	6
p	Pinching slope	3	10
ψ	Pinching magnitude	7	3
δ_ψ	Pinching rate	11	12
λ	Pinching severity	10	11

provide a good fit if the errors are small. Thus, this amounts to estimating the 13 control parameters of the differential model of hysteresis when a load–displacement trace is given. The optimization problem can be stated as the determination of the parameter vector

$$\mathbf{p} = (A, \alpha, \beta, \gamma, n, \delta_v, \delta_\eta, \zeta_s, q, p, \psi, \delta_\psi, \lambda) \quad (13)$$

such that the objective function

$$g(\mathbf{p}) = \frac{1}{N} \sum_{j=1}^N [x(t_j) - \hat{x}(t_j|\mathbf{p})]^2 \quad (14)$$

is minimized. In the above expression, t_j is a sequence of time instants and $x(t_j)$ is the given system displacement at t_j , where $j = 1, 2, \dots, N$. On the other hand, $\hat{x}(t_j|\mathbf{p})$ is the system displacement at t_j calculated from Eqs. 1 and 4 when the parameter vector is equal to $\mathbf{p}$. The chosen objective function is simply the mean-square error in the displacement. Minimization of the objective function is subjected to the constraint that all parameters in $\mathbf{p}$ with the exception of γ are positive (Ma et al. 2004).

Early studies of parametric identification of hysteresis typically employed the non-degrading classical differential model containing only five parameters (Kyprianou et al. 2001; Ni

et al. 1998). In the few studies that involved degradation (Furukawa and Yagawa 1997; Sues et al. 1988; Zhang et al. 2002), either a restricted Bouc–Wen model containing less than thirteen parameters or a restricted identification algorithm was used. For example, Zhang et al. (2002) used gradient-based local-search algorithms to estimate some of the hysteretic control parameters for degrading structures. However, the generalized differential model of hysteresis is highly nonlinear and any gradient-based method tends to be trapped near local minima and therefore fails to converge. A robust and efficient identification algorithm is needed to estimate all thirteen unspecified parameters of the differential model of hysteresis (Ma et al. 2006).

Reduction of Parameters

In order to streamline the identification of the control parameters of differential hysteresis, the generalized Bouc–Wen model has been reexamined (Ma et al. 2004). Two significant issues have been uncovered. First, it was discovered that the unspecified parameters of the differential model are functionally dependent. One of the 13 control parameters can be eliminated through suitable transformations in the parameter space. The number of unspecified parameters can thus be reduced from 13 to 12 without any loss of

generality. Elimination of a parameter will appreciably accelerate the convergence of any identification algorithm. As explained by Ma et al. (2004), a convenient way to eliminate one parameter is to map the parameter A into 1. Henceforth, $A = 1$ will be assumed, and there remain twelve loop parameters in the differential model of hysteresis.

The second issue uncovered is the existence of insensitive parameters in the generalized Bouc–Wen differential model: the variations of three or four control parameters in the generalized model do not appreciably alter the computed hysteretic evolution of the system. Through local and global sensitivity analyses (Iman and Helton 1991; Ma et al. 2004), the twelve remaining parameters are ranked in order of decreasing sensitivity, as shown in Table 1. The method of global analysis employed is a probabilistic method recently expounded by Sobol (1993). It can account for the mutual interactions of the twelve parameters. On an overall basis, δ_v , δ_ψ , and λ are the least sensitive parameters. In the identification of hysteresis, these parameters are likely to slow down numerical convergence. One extreme measure to deal with this problem is to set the insensitive parameters to constants in the beginning; at any rate they need not be estimated with high precision. It is decided in this investigation that a two-stage procedure will be adopted to streamline the identification process. In the first stage, the three least sensitive parameters δ_v , δ_ψ , and λ are fixed, and a crude value of the nine-parameter vector

$$\mathbf{p}_1 = (\alpha, \beta, \gamma, n, \delta_\eta, \zeta_s, q, p, \psi) \quad (15)$$

is first estimated by minimization of the objective function in Eq. (14). This crude value of $\mathbf{p}_1$, together with the fixed values of δ_v , δ_ψ , and λ , is then used as seeds in a second-stage identification to estimate the optimal value of the twelve-parameter vector

$$\mathbf{p}_2 = (\alpha, \beta, \gamma, n, \delta_v, \delta_\eta, \zeta_s, q, p, \psi, \delta_\psi, \lambda) \quad (16)$$

In the identification of hysteresis, this two-stage procedure economizes on both computer core memory and computing time.

Nonlinear Optimization Algorithms

The goal of an optimization problem is to find a vector $\mathbf{p}^*$ in the search space S so that certain quality criterion is satisfied, namely, the error norm $g(\mathbf{p})$ in Eq. 14 is minimized. The vector $\mathbf{p}^*$ is a solution to the minimization problem if $g(\mathbf{p}^*)$ is a global minimum in S . For the constrained nonlinear optimization problem associated with the identification of differential hysteresis, the error surface defined by the objective function and constraints can exhibit many local minima or can even be multimodal. For this reason, different solution techniques will have dramatically different performance. The primary consideration in evaluating an optimization algorithm may be convergence speed or the minimum error achieved. Secondary consideration may be consistency, robustness, computational efficiency, or tracking capabilities.

Solution techniques are classified into local or population search. Basic local-search methods rely on iterative descent toward a minimum in which the direction of descent depends either on the gradient $\nabla g(\mathbf{p})$ or the Hessian matrix $\nabla^2 g(\mathbf{p})$ of the objective function. There are drawbacks of most gradient-based techniques, which include difficulties in calculating the gradient and frequent trapping of the iterates near a local minimum (Bertsekas 1999). In view of that, Zhang et al. (2002) used a non-gradient method based upon the Nelder and Mead simplex method to minimize the objective function $g(\mathbf{p})$ in Eq. 14. The simplex method explores the search space S either by reflecting, contracting, or expanding away from the worst vertex or shrinking toward the best vertex. An appropriate sequence of such movements converges to the nearest local minimum. This downhill simplex method requires only function evaluations and not derivatives. However, even this method does not have desirable convergence properties.

Population-based methods, also known as evolutionary computations, search the entire solution space S by maintaining a group of candidate vectors. Evolutionary techniques are inspired by natural evolution and adaptation with the essence of survival of the fittest. During the iterative process, new candidate vectors are

generated from existing ones by means of variation and selection such that the set of candidates would move toward increasingly favorable regions of the search space. Variation is achieved by first combining several candidate vectors to form a new one and then performing a random modification of the components of the resulting vector. Selection is introduced by comparing the fitness of the candidate vectors; the smaller the objective function $g(\mathbf{p})$, the fitter the vector $\mathbf{p}$. The fitter vectors will be selected to survive and become members of a new generation. Popular evolutionary techniques include genetic algorithms (Goldberg 1989), evolution strategies (Schwefel 1995), and differential evolution (Kyprianou et al. 2001; Lampinen and Storn 2004). Instead of performing sequential search as in the von Neumann architecture, evolutionary techniques employ massively parallel schemes with many computational elements connected by links of various weights. As a consequence, these techniques tend to be relatively robust because the risk of being trapped near a local minimum is significantly reduced. The trade-off is that evolutionary techniques are relatively slow compared with gradient methods. Genetic algorithms require conversion of floating-point values into bit strings and are less efficient for problems in which the parameters are real valued. For most floating-point problems, evolution strategies are more complicated and computationally involved than differential evolution. Differential evolution emerges as the best evolutionary method for use in the identification of hysteretic parameters of the generalized Bouc–Wen model.

Differential Evolution

Differential evolution is a relatively new and efficient approach for minimizing real-valued nonlinear and non-differentiable continuous-space functions (Price and Storn 1997). Let K be the dimension of the parameter vector $\mathbf{p}$ in Eq. 13. Suppose each generation in differential evolution consists of a population of P vectors. New vectors are generated in successive generations through certain mechanisms such that the vectors are getting closer and closer to the optimal value.

Basic mechanisms of differential evolution can be described as follows:

Initialization. In the beginning, the population can be initialized with random values chosen within boundaries specified by constraints if there is no prior knowledge about the system. Alternatively, a uniform probability distribution of the initial parameters may be assumed.

Mutation. Suppose all P vectors in generation G have been constructed. Denote these vectors by $\mathbf{v}_{i,G}$, where $i = 1, 2, \dots, P$ and every vector has dimension K . For each population member $\mathbf{v}_{i,G}$, a mutated vector $\mathbf{w}$ is generated by adding the weighted difference between two population members to a third member such that

$$\mathbf{w} = \mathbf{v}_{r_3,G} + F(\mathbf{v}_{r_1,G} - \mathbf{v}_{r_2,G}) \quad (17)$$

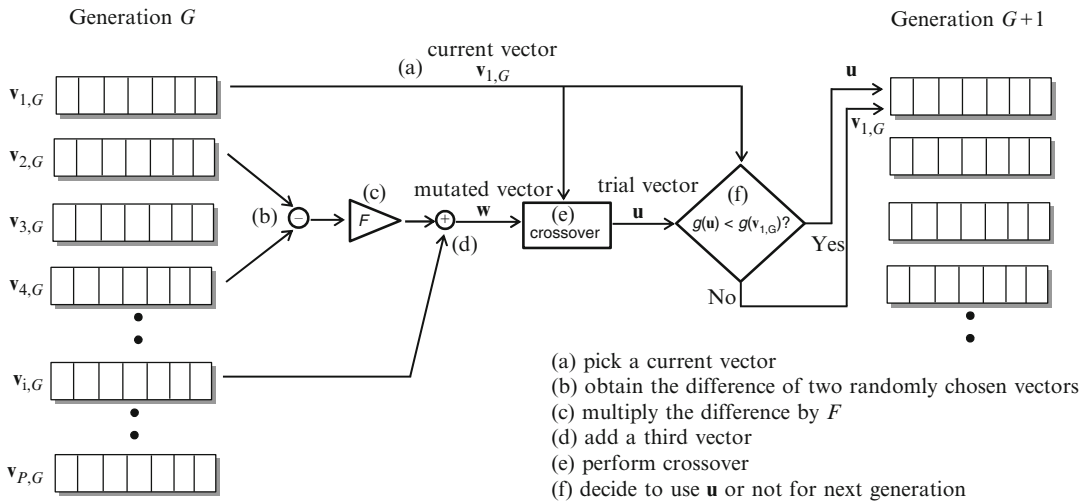
In the above expression, $0 < r_1, r_2, r_3 \leq P$ are three different integers chosen randomly, and $0 < F \leq 1$ is a preselected real-valued scaling factor. A common variation of mutation is to use the fittest vector $\mathbf{v}_{\text{best},G}$ of the current generation in place of $\mathbf{v}_{r_3,G}$ in Eq. 17.

Crossover. In order to increase the diversity of the population, the operation of crossover between $\mathbf{v}_{i,G}$ and $\mathbf{w}$ is performed. First, independent uniform random numbers $z_1, z_2, \dots, z_K$ are generated. With a preselected crossover constant $0 \leq CR \leq 1$, a trial vector $\mathbf{u} = [u_1, u_2, \dots, u_K]^T$ is constructed by means of

$$u_j = \begin{cases} w_j & \text{if } z_j < CR \\ (\mathbf{v}_{i,G})_j & \text{otherwise} \end{cases} \quad (18)$$

The constant CR determines whether each component of the vector $\mathbf{u}$ will originate from the mutated vector $\mathbf{w}$ or from the current vector $\mathbf{v}_{i,G}$. In essence, CR controls the probability that a crossover would occur for each component of the trial vector $\mathbf{u}$.

Selection. The selection process is performed by a one-to-one competition in which the fitness of a vector is determined by its objective value, which is calculated by Eq. 14 with a fourth-order Runge–Kutta method for numerical integration. The trial vector $\mathbf{u}$ is then compared to the



Seismic Response Prediction of Degrading Structures, Fig. 2 Basic mechanisms of differential evolution

vector $v_{i,G}$. If the vector u yields a smaller objective value, it replaces $v_{i,G}$ as $v_{i,G+1}$ in generation $G+1$. Otherwise, the original vector $v_{i,G}$ will be kept. The process of mutation, crossover, and selection of the differential evolution algorithm is summarized in Fig. 2. The process is repeated P times to construct all vectors of generation $G+1$. As evolution progresses, new generations of vector members become fitter and fitter as the objective values decrease.

In the differential evolution algorithm, three control parameters P , F , and CR must be preselected. Lampinen and Storn (2004) indicated that convergence of differential evolution is not particularly dependent upon the choice of these control parameters. It suffices to choose F and CR as multiples of 0.1 and P as a multiple of 10. Finally, a convergence criterion is needed to determine if the evolution process has converged. A common procedure is to set a desired objective value or to establish a threshold based on the relative fitness between the best and the worst vectors in each generation. Another method is to require a minimum percentage of improvement in fitness in order to continue the iteration process. One may also limit the number of generations or the computing time or both. Each convergence criterion has its advantages and drawbacks. Thus, a combination of convergence criteria is often used.

Based upon the exposition in this section, a robust algorithm using the latest theory of differential evolution is constructed for the identification of differential hysteresis. As explained before, a two-stage procedure is adopted whereby a crude value of the vector p_1 in Eq. 15 is first obtained before the optimal value of p_2 in Eq. 16 is computed. In simulations reported in this entry, the population size P is set to 100, while both the scaling factor F and the crossover constant CR are set to 0.5. The computing time is dependent on the amount of input data used for identification. The runtime can be reduced appreciably if some of the insensitive parameters are fixed and parallel computing is utilized.

Prediction of Performance

With a robust and relatively efficient algorithm devised in section “[System Identification](#),” a differential model of hysteresis of any degrading system may be identified from a given load–displacement trace. Suppose a hysteretic model is generated with a given load–displacement trace. It will be shown that, under fairly broad conditions, the model predicts the response of the same structure when driven by other cyclic loads. This can be carried out experimentally by building a number of identically configured structural joints

and subjecting them to different cyclic loads. One set of load–displacement trace is chosen to generate a hysteretic model of the joint by system identification. Subsequently, this identified model is used to compute the displacements of the joint under other cyclic loads. The computed displacements will be compared with experimental measurements. A reasonable match will validate the identified model and will demonstrate its prediction capability. In order to reduce the front-end costs of this investigation, only wood joints are used in the experiments. It must be emphasized that steel and concrete structures may be used as easily. Although a limited set of experimental data will be presented in this entry, many different structural joints have been built and tested to support any observations made herein.

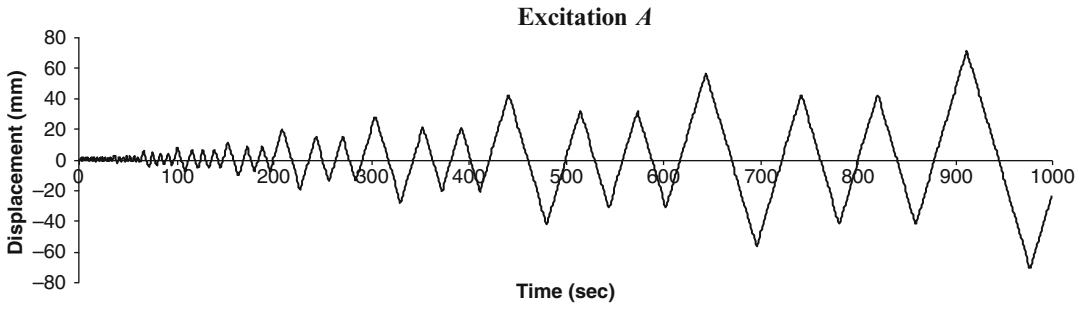
Test Setup

Cyclic tests of degrading structures are performed in the Forest Products Laboratory located at the Richmond Field Station of the University of California, Berkeley. Twenty identical specimens of a simple *T*-connection consisting of two wood members joined by plywood gusset plates, as shown in Fig. 3, are constructed. The two wood members are made out of laminated veneer lumber (LVL), which is a highly predictable engineered wood product with relatively small material variability. Each LVL member is a uniform beam with a square cross section of 3.5×3.5 in. and an elastic modulus of 2.0×10^6 psi. The LVL lumber is sawn to consistent sizes and is virtually free from warping and splitting. The two LVL members are connected on both sides by plywood gusset plates of 0.5 in. thickness. Six 2-in. metal screws are arranged in three rows to fasten the plates to the vertical beam. The two plates are also connected securely to the horizontal beam with four metal screws and two retrofit bolts. Cyclic loads are applied parallel to the lower beam at a height 36 in. above the lower beam. The arrangement is such that the *T*-connection will fail through detachment of vertical LVL member from the gusset plates.

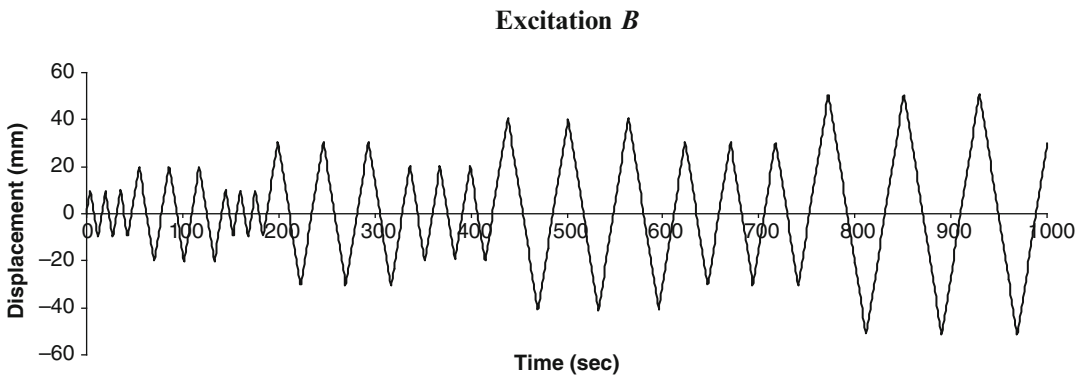


Seismic Response Prediction of Degrading Structures, Fig. 3 Reaction frame test system with a specimen

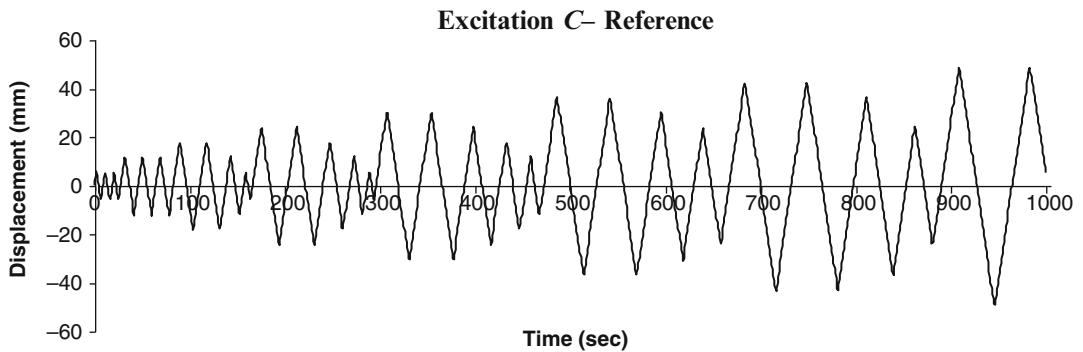
Common seismic testing protocols are employed to drive the *T*-connections. It is claimed that such protocols would simulate damage accumulation and degradation of the structural joint under real-life loads. Three of the cyclic excitations used in the experiments are shown in Figs. 4, 5, and 6. Excitation A is proposed by the Consortium of Universities for Research in Earthquake Engineering (CUREE), developed exclusively for wood testing (Krawinkler et al. 2001). The CUREE protocol is designed to model ordinary ground motion typical of most far-field locations. It is characterized by a pattern of a primary cycle followed by a number of trailing cycles that are 75 % of the previous primary cycle. Each primary cycle, which has higher amplitude, initiates a new level of damage to the structure. Damage then accumulates and changes the structural response in the trailing cycles. In addition, the number of reversed cycles is limited in order to induce failure modes resembling those in actual structures.



Seismic Response Prediction of Degrading Structures, Fig. 4 Excitation of CUREE



Seismic Response Prediction of Degrading Structures, Fig. 5 Excitation of Forintek Canada Corporation



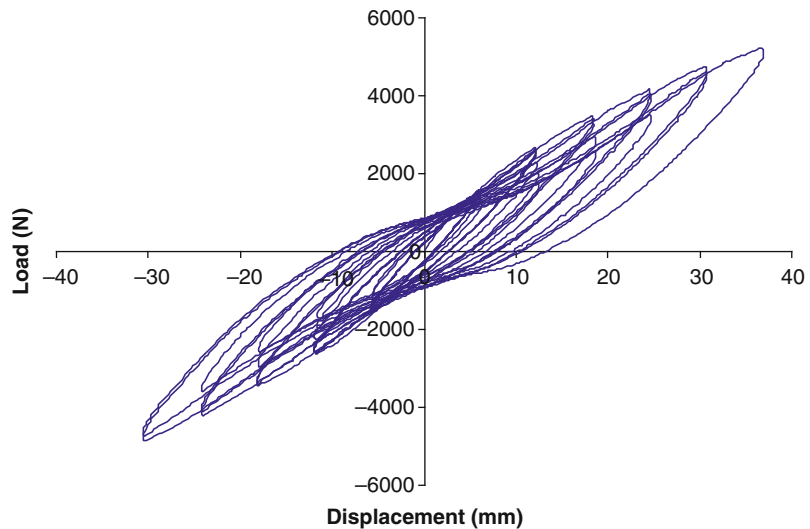
Seismic Response Prediction of Degrading Structures, Fig. 6 Excitation of Forest Products Laboratory

Excitation *B* is devised by Forintek Canada Corporation (FCC), a national wood products research institute in Canada. Compared with excitation *A*, equal-amplitude primary and trailing cycles are used in this excitation. The FCC protocol consists of a sequence of cyclic

groups, each composed of three identical cycles. The amplitude of each group is taken as a percentage of the nominal yield slip and is increased stepwise, with inter-spaced decreasing cycles until the specimen fails. Excitation *C* was developed at the Forest Products Laboratory for the

Seismic Response Prediction of Degrading Structures,

Fig. 7 Hysteresis loops generated by excitation *C* in 500 s



experiments reported herein. This protocol features trailing cycles that are decreasing in amplitude, which are different from the equal-amplitude trailing cycles in the CUREE and FCC excitations. In addition, excitation *C* does not have the series of low-amplitude cycles featured in the starting interval of excitation *A*.

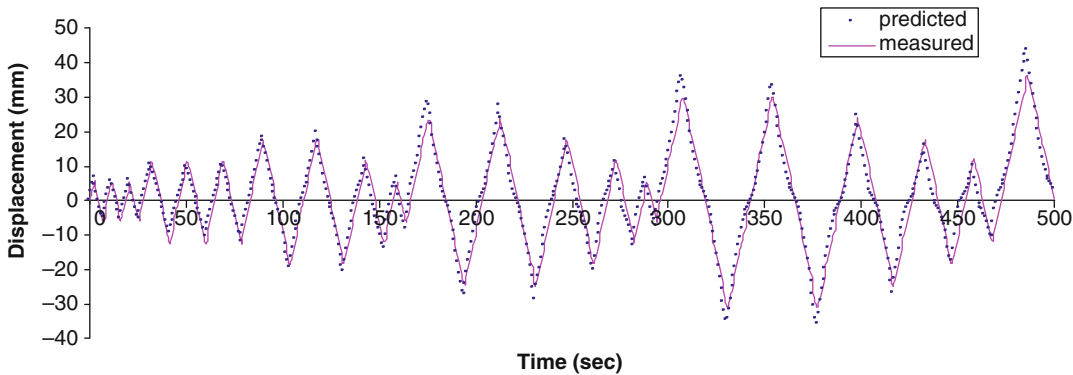
Model Identification and Validation

Each specimen of the identically configured *T*-connections is driven by a cyclic excitation and 1,000 s of load–displacement data are recorded. Using any of these load–displacement traces, a model of hysteretic evolution can be identified. Two issues will now be examined. (1) Can a hysteretic model identified with a given cyclic load predict the future structural response if the *same* cyclic load continues beyond the duration used for identification? (2) Can a hysteretic model identified with a given cyclic load predict the structural response due to a *different* cyclic load?

In Fig. 7, the load–displacement trace associated with excitation *C* for the first 500 s is shown. As reflected in the hysteresis trace, the gusset joint is degrading in both strength and stiffness. There is also some pinching of the hysteretic loops, which is probably caused by slipping of the connecting screws. The load–displacement trace of Fig. 7 is used as input to the identification

algorithm devised in the last section. Recall that $A = 1$, and there remain twelve hysteretic parameters to be estimated of the gusset-plate connection. Once these twelve parameters are identified, a hysteretic model of the *T*-connection is obtained. This hysteretic model can now be used to compute the displacement history of the gusset joint associated with excitation *C*. As illustrated in Fig. 8, the computed displacement history closely matches the experimental measurements. This validates the identified hysteretic model over the first 500 s.

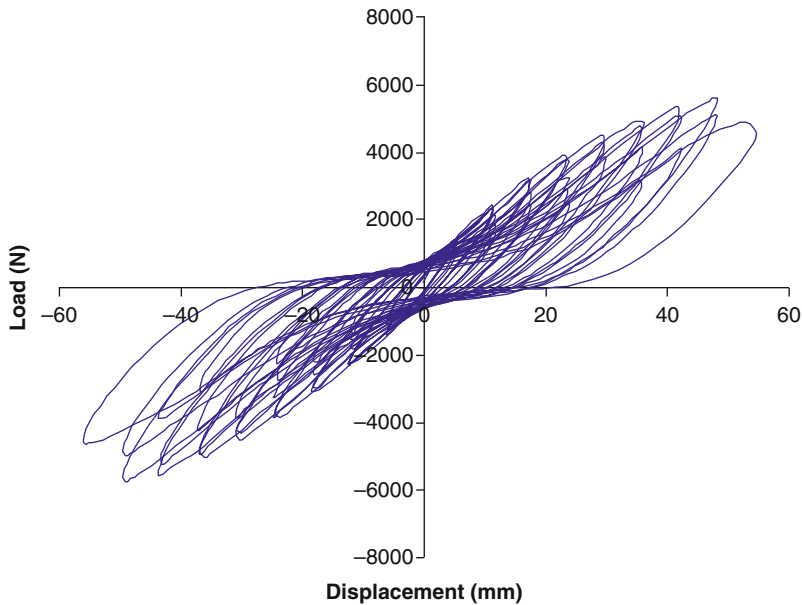
A different hysteretic model of the *T*-connection is generated when the duration of input data changes. In Fig. 9, the load–displacement trace associated with excitation *C* for $0 < t < 1,000$ s is shown. Compared with Fig. 7, pinching is more prominent after the first 500 s. Thus, the twelve control parameters estimated with input data of Fig. 9 are different from those associated with Fig. 7. The new hysteretic model must be validated again for a duration of 1,000 s because the dominant characteristics of degradation are different between the two intervals $0 < t < 500$ s and $0 < t < 1,000$ s. A hysteretic model identified with input data over $0 < t < 500$ s will not produce a displacement history closely matching the experimental data over $500 < t < 1,000$ s. In short, system identification cannot produce



Seismic Response Prediction of Degrading Structures, Fig. 8 Comparison of computed and measured displacements under excitation *C*

Seismic Response Prediction of Degrading Structures,

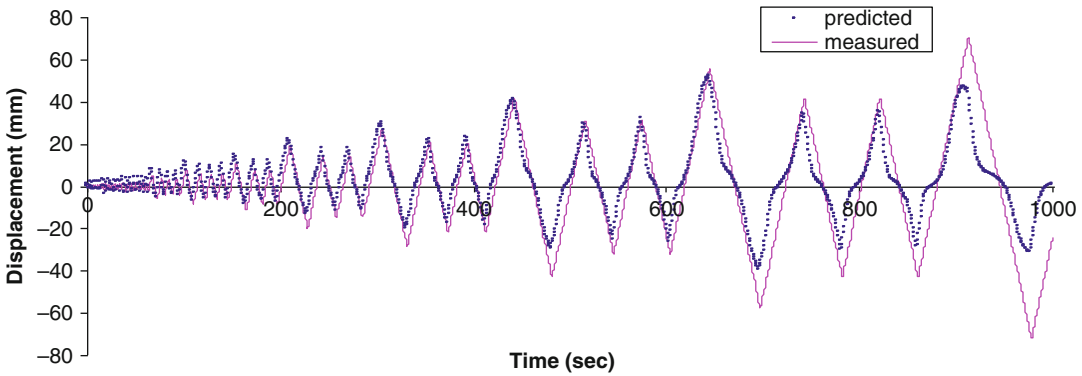
Fig. 9 Hysteresis loops generated by excitation *C* in 1,000 s



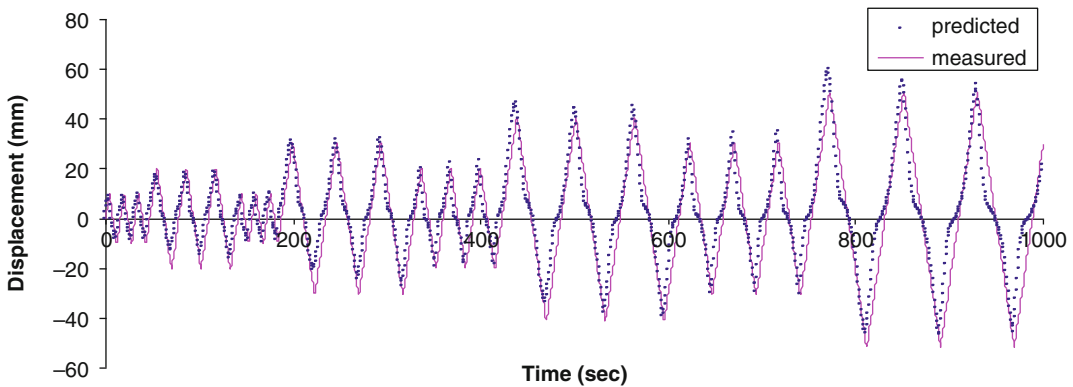
features and trends not already contained in the input data. If the duration of input used for identification already contains all features of hysteretic evolution, then a hysteretic model identified with a given excitation can accurately predict the future response when the *same* excitation continues beyond the duration used for identification. This basically answers the first question brought up at the beginning of this section.

In identifying a hysteretic model of the *T*-connection, the experimental load–displacement trace associated with excitation *C* has been used

directly as input for identification. It can be argued that such input data are corrupted by Gaussian noise, and a more accurate model can be identified if noise is suppressed with a filter. To this end, the input was first taken through a median filter and then a least-squares low-pass filter. Such a filtering process is known to be effective in smoothing out random fluctuations caused by Gaussian noise (Papoulis and Pillai 2002; Zhang et al. 2002). However, it has been found that a better match than what is shown in Fig. 8 is not achieved with the noise-filtered data. It appears that differential evolution is not sensitive to a moderate level of



Seismic Response Prediction of Degrading Structures, Fig. 10 Predicted and measured displacements under excitation A in 1,000 s



Seismic Response Prediction of Degrading Structures, Fig. 11 Predicted and measured displacements under excitation B in 1,000 s

Gaussian noise. Subsequently, noise-filtering is not used on any of the load–displacement traces.

Can a hysteretic model obtained by identification using a given excitation predict the structural response due to a *different* cyclic load? This is the second question brought up earlier. Using 1,000 s of the load–displacement trace associated with excitation C, a hysteretic model of the T-connection has been identified. Keep in mind that other cyclic loads do not play any part in the identification of hysteresis. The identified model is then used to compute the displacement of the gusset joint under excitation A of Fig. 4. However, load–displacement data associated with excitation A have already been collected experimentally using another specimen of the identically configured gusset joints. Thus, the

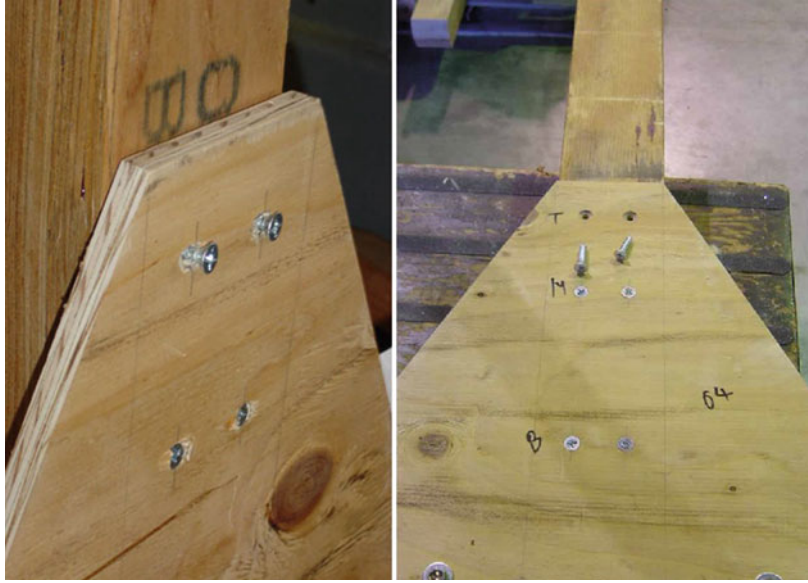
computed (or predicted) displacement can be compared with the measured displacement. As shown in Fig. 10, there is reasonably good match. A similar observation can be made with other cyclic loads. As illustrated in Fig. 11, there is also reasonably good match between the computed (or predicted) and experimental data when excitation B of Fig. 5 is used. This demonstrates that a hysteretic model identified with a given excitation may be used for response prediction under different excitations. However, the accuracy of prediction tends to diminish as time increases.

Requirements for Accurate Prediction

As demonstrated in previous sections, a hysteretic model obtained by identification can be used

Seismic Response Prediction of Degrading Structures,

Fig. 12 Failure configuration of *T*-connection under excitation *C* over 1,000 s



to predict future performance of a system. Large errors, however, can arise in many cases. For example, the prediction of response associated with excitation *A* is no longer satisfactory after 900 s, as shown in Fig. 10. Perhaps this is not surprising. As time progresses, damage accumulates and the *T*-connection begins to fail. There is, however, a divergence in the failure configuration due to different excitations.

As shown in Fig. 12, a typical failure configuration under excitation *C* over 1,000 s is that the top two screws in the vertical member are appreciably loosened due to wood crushing. Disassembly of the gusset joint reveals that both screws are broken into two pieces under excitation *C*. If 1,000 s of the load–displacement trace associated with excitation *C* is used for identification, a hysteretic model is generated that anticipates the breaking of the top two screws. Under excitation *A*, however, it has been observed that all four screws in the top two rows are broken over 1,000 s, as shown in Fig. 13. If a model identified with excitation *C* is used for response prediction when the structure is driven by excitation *A*, it fails to predict that all four screws in the top two rows would be broken over 1,000 s. This explains the discrepancy between the predicted and measured displacements under excitation



Seismic Response Prediction of Degrading Structures, Fig. 13 Failure configuration of *T*-connection under excitation *A* over 1,000 s

A in Fig. 10. If the failure mechanism can be controlled, the accuracy in response prediction will be increased.

The observations made in experimenting with the elementary *T*-connection can be extended to complex assemblies or structures. A complex structure consists of many different elementary connections. Since each elementary connection may have several failure configurations, the totality of failure configurations is very large for a complex structure. If structural degradation can be controlled so as to result in a failure configuration identical to what is contained in the input data used for identification, then nonlinear degrading response of a complex structure can be predicted. Otherwise, prediction is only reliable up to a point at which the input data and the existing complex structure progress into different failure configurations. A study of the influence of failure configurations on the reliability of response prediction will be worthwhile in a subsequent course of investigation.

Summary

A basic objective of this entry is to advance the methodology for predicting the performance of nonlinear deteriorating structures. Using the generalized Bouc–Wen differential model, it is shown how system identification can be utilized to predict the response of a degrading structure well beyond its linear range. Important observations reported in the paper are summarized in the following statements.

1. A robust identification algorithm based upon the generalized Bouc–Wen model and the theory of differential evolution can be used to generate practical models of hysteresis of degrading structures. Differential evolution is not sensitive to a moderate level of input noise.
2. If the duration of input used for identification already contains all features of hysteretic evolution, then a hysteretic model identified with a given cyclic load can accurately predict the future structural response when the same cyclic load continues beyond the duration used for identification.
3. A hysteretic model identified with a given cyclic load may be used to predict the nonlinear response under different cyclic loads. As damage accumulates, the precision of predicted response decreases, and the reliability of prediction ultimately depends on the degree of similarity between failure configurations under different cyclic loads.

In the absence of a fundamental theory of degradation, the response prediction of degrading structures is indeed a challenging task. System identification of hysteretic evolution provides a brute-force procedure for prediction that has the potential of allowing a closer representation of reality. The research reported herein addresses various aspects of an identification methodology for predicting the performance of real-life deteriorating structures well beyond their linear ranges. Among other things, it is hoped that this entry would point to directions along which further research efforts should be made. This entry could also contribute to the development of a general formulation of the mechanisms of degradation in the long term.

Cross-References

- [Nonlinear System Identification: Particle-Based Methods](#)
- [Online Response Estimation in Structural Dynamics](#)
- [System and Damage Identification of Civil Structures](#)
- [Vibration-Based Damage Identification: The Z24 Bridge Benchmark](#)

References

- Baber TT, Noori MN (1986) Modeling general hysteresis behavior and random vibration application. *ASME J Vib Acoust Stress Reliab Des* 108:411–420
- Baber TT, Wen YK (1981) Random vibration of hysteretic degrading systems. *ASCE J Eng Mech* 107(6):1069–1087

- Bertsekas DP (1999) Nonlinear programming, 2nd edn. Athena Scientific, Belmont
- Bouc R (1967) Forced vibration of mechanical systems with hysteresis. In: Proceedings of 4th conference on nonlinear oscillations, Prague, Czechoslovakia, p 315
- Foliente GC (1995) Hysteretic modeling of wood joints and structural systems. *ASCE J Struct Eng* 121:1013–1022
- Furukawa T, Yagawa G (1997) Inelastic constitutive parameter identification using an evolutionary algorithm with continuous individuals. *Int J Numer Methods Eng* 40:1071–1090
- Goldberg DE (1989) Genetic algorithms in search, optimization & machine learning. Addison-Wesley, Reading
- Iman RL, Helton JC (1991) The repeatability of uncertainty and sensitivity analyses for complex probabilistic risk assessments. *Risk Anal* 11:591–606
- Kojic M, Bathe KJ (2005) Inelastic analysis of solids and structures. Springer, Berlin
- Krasnoselskii MA, Pokrovskii AV (1989) Systems with hysteresis. Springer, Berlin
- Krawinkler H, Parisi F, Ibarra L, Ayoub A, Medina R (2001) Development of a testing protocol for woodframe structures. CUREE-Caltech woodframe project report W-02, Consortium of Universities for Research in Earthquake Engineering, Richmond
- Kyprianou A, Worden K, Panet M (2001) Identification of hysteretic systems using the differential evolution algorithm. *J Sound Vib* 248:289–314
- Lampinen J, Storn R (2004) Differential evolution. In: Onwubolu GC, Babu BV (eds) New optimization techniques in engineering. Springer, Berlin, pp 125–166
- Ma F, Zhang H, Bockstedte A, Foliente GC, Paevere P (2004) Parameter analysis of the differential model of hysteresis. *ASME J App Mech* 71:342–349
- Ma F, Ng CH, Ajavakom N (2006) On system identification and response prediction of degrading structures. *Struct Control Health Monit* 13(1):347–364
- Ni YQ, Ko JM, Wong CW (1998) Identification of nonlinear hysteretic isolators from periodic vibration tests. *J Sound Vib* 217:737–756
- Papoulis A, Pillai SU (2002) Probability, random variables, and stochastic processes, 4th edn. McGraw-Hill, New York
- Price K, Storn R (1997) Differential evolution. *Dr Dobb's J* 22:18–24
- Roberts JB, Spanos PD (1990) Random vibration and statistical linearization. Wiley, New York
- Schwefel HP (1995) Evolution and optimum seeking. Wiley, New York
- Sobol IM (1993) Sensitivity estimates for nonlinear mathematical models. *Math Model Comput Exp* 1:407–414
- Song J, Der Kiureghian A (2006) Generalized Bouc-Wen model for highly asymmetric hysteresis. *ASCE J Eng Mech* 132(6):610–618
- Sues RH, Mau ST, Wen YK (1988) System identification of degrading hysteretic restoring forces. *ASCE J Eng Mech* 114:833–846
- Wen YK (1976) Method for random vibration of hysteretic systems. *ASCE J Eng Mech* 102(2):249–263
- Zhang H, Foliente GC, Yang Y, Ma F (2002) Parameter identification of inelastic structures under dynamic loads. *Earthq Eng Struct Dyn* 31(5):1113–1130

Seismic Risk Assessment, Cascading Effects

Paolo Gasparini and
Alexander Garcia-Aristizabal
Center for the Analysis and Monitoring of
Environmental Risk (AMRA), Naples, Italy

Synonyms

Domino effects; Short-term hazard and risk assessment; Time-dependent risk; Triggering effects

Introduction

Risk is conveniently assessed as a function of the probability that a certain event will occur and of the extent of the damage caused to humans, environment, and objects. Conceptually, it is the result of the operation

$$\text{Risk} = \text{Hazard} \times \text{Damage} = \text{Hazard} \times \text{Vulnerability} \times \text{Value at Risk} \quad (1)$$

where the *Hazard* can be understood as the probability that a given adverse event of a given intensity (hereinafter referred to as *intensity measure*, IM) will occur in a certain area within a defined time interval; the *Vulnerability*, a concept interpreted and applied in various ways, here is considered in its physical dimension as the probable damage to an element at risk given a level of intensity (IM) of the hazard; and the *Value at Risk* term is considered as a measure of the total potential loss due to an adverse event in a given area.

Note that both *Hazard* and *Vulnerability* are generally defined in probabilistic terms, while *Value at Risk* has some defined units of measure (monetary cost, number of human lives, etc.).

Seismic risk assessment methods have shown an important development over the last 40 years reaching an outstanding level of acceptance by the scientific community as a valid framework for sound risk mitigation and decision-making. In general the seismic risk assessment uses seismic hazard information combined with fragilities of structures and facilities in order to estimate the probabilities of damages and to measure expectancies of losses using a predefined metric of loss.

Different past disasters have highlighted the fact that natural or man-made events can trigger other events, leading to a significant increase of fatalities and damages. As a consequence, there is a growing demand from risk managers to implement multi-type hazard and risk assessments as well as assessing cascading effects. Consequently, an increasing number of scientists studying natural hazards and risk assessment are turning their interest towards the multi-hazard risk and multi-risk assessment practice. The transition from the single- to the multi-risk assessment represents a process in which both the complexity and the data requirements are significantly increased. This transition implies a shift from the “hazard-centered” perspective that characterizes the single-risk assessment to a “territorial-centered” one.

The cascading events are one of the fundamental concepts of the multi-risk assessment. It is worth noting that a different terminology is used in the practice of risk evaluation when addressing the problem of chains of events. In particular, the term “domino effect” is generally used in the field of industrial accidents, whereas the term “cascading effects” is mainly used in ecology and in natural hazards in the context of the multi-risk assessment. The objective of this entry is to discuss the role of cascading effects triggered by earthquakes and its importance for a holistic view of the seismic risk assessment process.

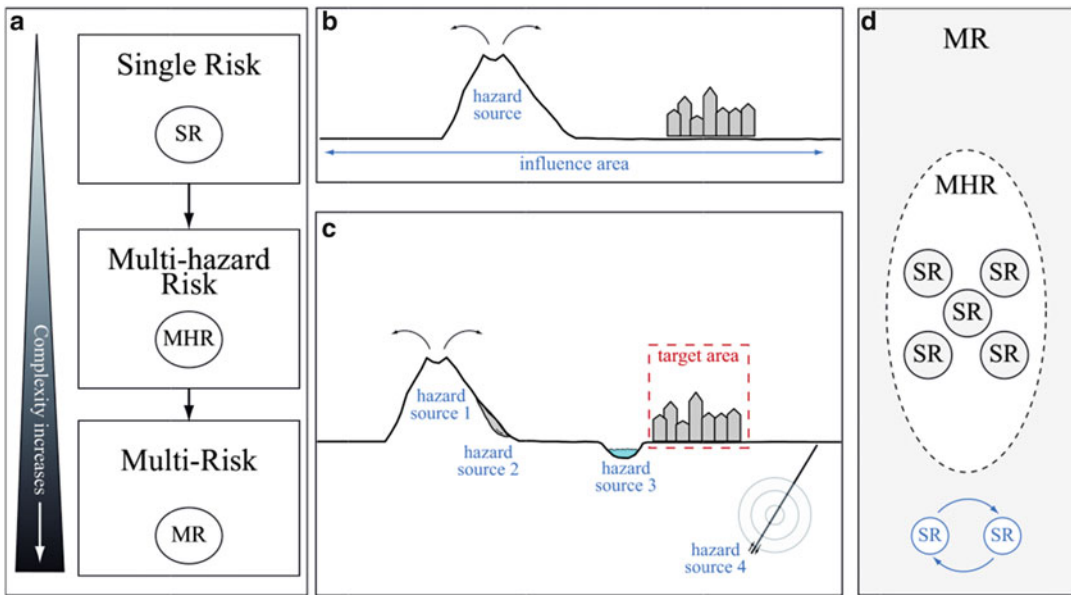
Multi-risk as a Framework for the Assessment of Cascading Effects

The transition from single- to multi-risk assessment represents a process in which both the complexity and the data requirements are significantly increased (Fig. 1a). This transition implies a shift from a “hazard-centered” perspective that characterizes the single-risk assessment, to a “territorial-centered” one. In fact, from a multi-risk perspective, the first element to be defined is the target area of interest, which is in general terms the piece of territory composed of elements at risk that are vulnerable in different ways to various sources of hazard.

The evolution from the single-risk to the multi-risk analysis consists of three different and complementary elements: the single-risk assessment (SR), the multi-hazard risk assessment (MHR), and the multi-risk assessment (MR).

Figure 1 represents the transition from the SR to the MR assessment (Fig. 1a). The hazard-centered perspective of the SR is represented in Fig. 1b; in this case, the first element of the analysis is generally the hazard source followed by the definition of the impact area and the assessment of the potential effects. The change towards a territorial perspective is represented in Fig. 1c. In the MHR, different independent hazard sources affecting a given common area of interest are considered; in this case, the first element of the analysis is the definition of the piece of territory which is the target area for the analysis; the target area includes the elements at risk exposed to adverse events. Note that this change in perspective is common to both the MHR and the MR assessments. Once the target area has been identified, the next element of the analysis is the identification of the potential hazard sources that may cause harm to the set of exposed elements in the target area.

Figure 1d illustrates how each element in this transition is a subset in the higher levels of the sequence. In fact, a set of SR analyses compose an MHR analysis, and when the MHR assessment is complemented with the analysis of interactions



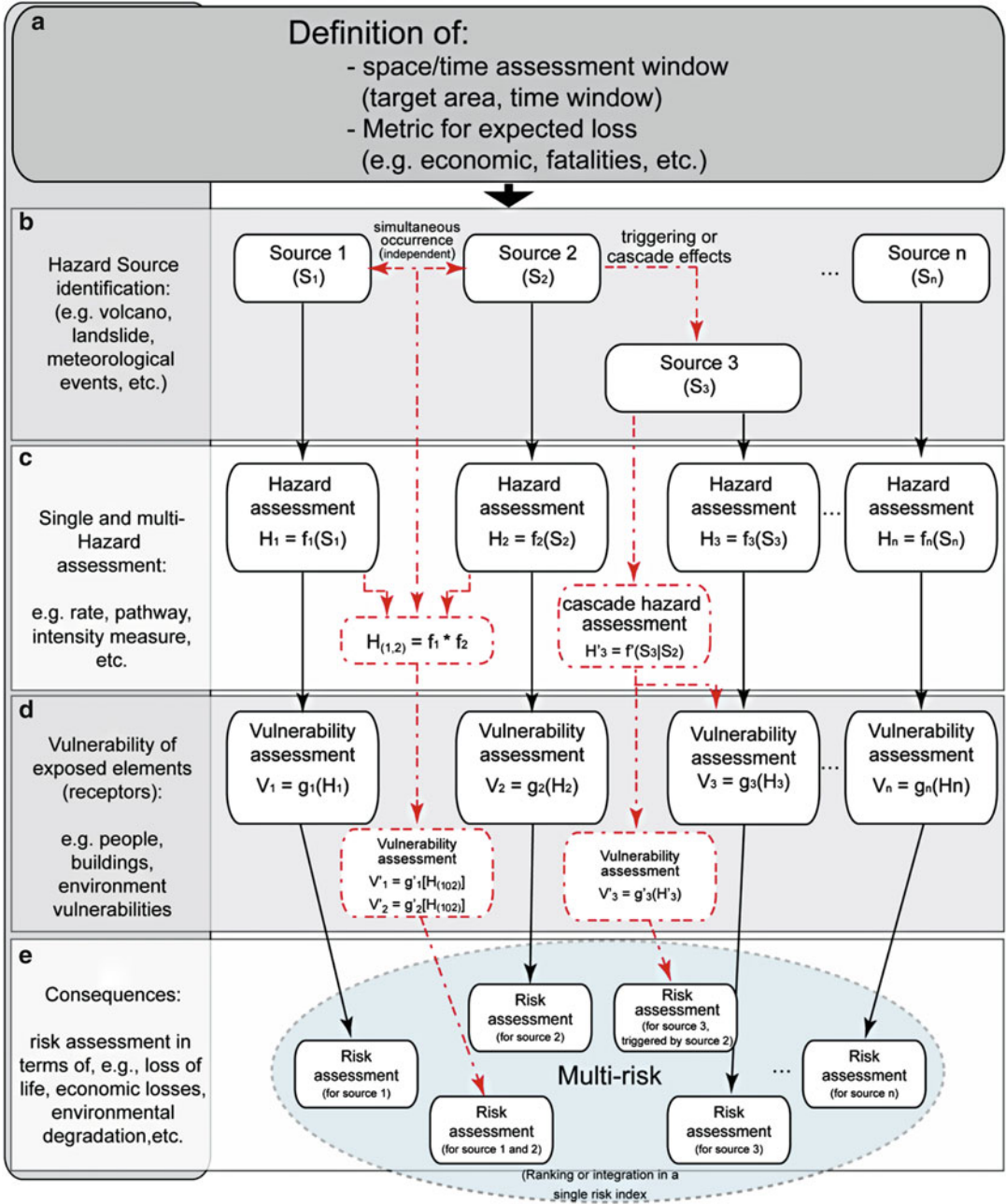
Seismic Risk Assessment, Cascading Effects, Fig. 1 Representation of the transition from the single- to the multi-risk assessment

among events, then the MR level is reached. This sequence is of course valid if the SR and MHR analyses are done coherently in a multi-risk framework, following some basic (and often non-simple) elements of harmonization.

The multi-risk assessment (MR) is then a generalization of the MHR assessment in the sense that it integrates also the possible interactions or cascading effects (Fig. 1c, d); therefore, the main objective of the MR assessment level is to assess the effects of possible interactions or cascading effects among the different hazardous events. Considering long- and short-term assessments, Marzocchi et al. (2012) identify the following main steps of the MR assessment procedure:

1. Definition of the space-time window for the risk assessment and the metric for evaluating the risks
2. Identification of the risks impending on the selected area
3. Identification of selected hazard scenarios covering all possible intensities and relevant hazard interactions
4. Probabilistic assessment of each scenario
5. Vulnerability and exposure assessment for each scenario, taking into account the vulnerability of combined hazards
6. Loss estimation and multi-risk assessment

In this framework, a set of scenarios considering adverse events from different sources are identified along with the potential interactions. Each chain of adverse events is defined in a series/parallel sequence of happenings and is generally represented in an “event tree.” Each branch of the event tree is quantified by a probabilistic analysis considering the sequence of the events and the vulnerability and the exposed values of the specified targets. This concept is illustrated in Fig. 2, where the possible kinds of interactions are identified using discontinuous red lines. For example, a sequence defined for independent hazards interacting at the vulnerability (or damage) level is represented by the red discontinuous line in the sequence at the left side of Fig. 2. Likewise, an example of interaction at the hazard level is presented with source 3 triggering source 2 (see the path defined with the red discontinuous line at the right in Fig. 2). Note that



Seismic Risk Assessment, Cascading Effects, Fig. 2 Multi-risk framework considering possible interactions at the hazard and the vulnerability levels (Reproduced from Marzocchi et al. 2012, with permission from Springer®)

in this case, the interaction at the vulnerability level may or may not be present. Details about the kinds of interactions are more deeply presented in the following sections of this entry.

Within this context of interactions and triggering mechanisms, earthquake-triggered events have drawn much attention in the last decades because of the severe damages that the triggered

events may cause. Sometimes, the losses caused by the secondary events may exceed the losses caused by the initial, triggering earthquake itself. For example, in a number of earthquake-triggered landslides, the losses caused by the triggered mass movements have been comparable (if not higher) to the losses caused by the direct impact of ground shaking. An outstanding case is the Huascarán rock-debris avalanche triggered by the M7.7 Peru earthquake in 1970 (e.g., Plafker et al. 1971), in which the impact of the avalanche counts for more than 50 % of the total casualties.

The kind of events that can be triggered and the possible interactions that can be analyzed vary in a wide spectrum of events including natural (as other earthquakes, landslides, tsunamis, volcanic unrest and eruptions, etc.) and nonnatural (as Natech) events. A non-exhaustive description of known examples is presented in the section [“Examples of Cascading Effects Caused by Earthquakes.”](#)

Cascading Effects: An Overview

From the risk assessment point of view, the cascades of events occurring after an earthquake may represent an important source of loss amplifications. In this context, we consider as “*loss amplification*” all the losses not caused directly by the earthquake (as the triggering event) but caused by other secondary events occurring as a consequence of the initiating triggering event. Therefore, the loss amplification can be a factor resulting from comparing the total losses caused by the chain of events triggered by an earthquake with the direct losses caused by the initial earthquake itself.

The cascading effects result as a consequence of interactions generated by cause/effect relationships among different phenomena. The nature of these interactions may be described by a wide set of phenomenological relationships, and it makes difficult to set a generalized procedure for the quantification of cascading effects. For this reason and to simplify the setting of the problem for quantitative purposes, a generic taxonomy of the

possible kinds of interactions can be defined. Following the framework presented in Marzocchi et al. (2012), Garcia-Aristizabal et al. (2013), and Garcia-Aristizabal et al. (forthcoming-2015), we can consider the following two major sets of interactions: (1) interactions at the hazard level and (2) interactions at the vulnerability level. This generic taxonomy of possible interactions is qualitatively described in the following paragraphs, and a more quantitative definition is presented in the last section of this entry.

Interactions at the Hazard Level

In this case the interaction problem is understood as the assessment of possible chains of adverse events in which the occurrence of a given initial “triggering” event entails a modification of the probability of occurrence of a secondary event. Even if this typology of problem can be assessed on a long-term basis, its utility can be highlighted in short-term problems. It can be seen that the typology of interactions that can be grouped under this name are in fact phenomena in which the physical process of interest is a pure triggering mechanism in which a first occurring event produces a perturbation that, when acting over a given system, may bring it to unstable conditions forcing it to find a new equilibrium. Reaching this new equilibrium may imply the occurrence of an event that in this case may be said as *triggered* by the initial triggering one.

The link between the intensity of the triggering event (e.g., the ground shaking caused by an earthquake) and the intensity of the triggered event (e.g., a volume of mass moving in a slope) is governed by complex physical mechanisms that are intrinsically related to the specific triggering and triggered events. This fact and the ubiquitous random effects that may affect these processes make of probabilistic approaches the most promising way for the quantitative characterization of such interactions. Therefore, assessing interactions at the hazard level in probabilistic terms implies to quantify the conditional probabilities of occurrence of the triggered event, given the occurrence of the triggering event (with given intensity).

Interactions at the Vulnerability Level

In this case the problem fundamentally intends to assess the effects that the simultaneous occurrence of two or more events (not necessarily linked) may have on the response of a given typology of exposed elements. In this case, the process of interest is the additive effect of different events acting on the same elements. More specifically, this kind of interaction is referred to the case in which the occurrence of one event (the first one occurring in time) may alter the response of the exposed elements against another event (that may be of the same kind as the former but also a different kind of hazard). In this case, it is assumed that two or more events act on a system (e.g., a structure) and that the additive or cumulated effect produces a change in the response of the system with respect to the conditions before the occurrence of the former event. In practice, it implies that two or more hazardous events act simultaneously or in a short-time window (in general short enough that the system cannot be repaired) on the same structure.

The cases that can be grouped under this kind of interaction are of different nature; in general, the physical processes of interest are those related with the response of the system (in this case, an exposed element) to the loads caused by different events taking in account their additive or cumulated effects. Note that the different events may be of the same nature (i.e., same kind of hazards, as, e.g., two earthquakes shaking the same structures in a short-time window) or coming from different kinds of phenomena (e.g., the shaking caused by an earthquake as the first event, followed by loads caused by river flooding or strong wind). Note that the different events causing additive loads can be themselves the result of a common triggering event or may be independent events, and this relationship is important for the quantitative analyses.

As in the previous case, the quantitative assessment of this kind of interactions relies on probabilistic analyses. It may be understood as a multidimensional development of fragility functions for the different typologies of exposed elements. In general, it involves the determination of

the conditional probabilities of having certain damage state given the simultaneous or cumulative solicitation caused by different combinations of intensities of the different events considered. Note that this information can be represented in different ways, according with the nature of the acting events. For instance, if the acting events are not causing cumulated damages in the same elements (e.g., seismic loads from one side and the loading caused by snow accumulation or flooding water on the other), then the conditional probabilities define a multidimensional fragility surface. Conversely, if the different acting events produce cumulated damages in the same elements (i.e., when the first event pre-damages a given element, as, e.g., two earthquakes acting one after the other in a short time interval), the conditional probabilities are generally referred as *time-dependent* or *damage-dependent* fragilities (or vulnerabilities).

Examples of Cascading Effects Caused by Earthquakes

Many examples of cascading effects caused by earthquakes can be cited as, for example, the well-known cases of tsunamis triggered by earthquakes or earthquake-triggered landslides. As described in the previous section, there may be different typologies of interactions, and the physical processes involved in these interactions are extremely case dependent. In this section, we describe some outstanding cases of interactions (mainly at the hazard level) that highlight the potential loss amplifications that these triggering effects may produce.

Earthquake-Earthquake Triggering

Earthquake mainshocks cause aftershocks to occur, which in turn activate their own subsequent aftershock sequences, resulting in a cascade of triggering that may increase the damage caused to the exposed elements with respect to the effects directly produced by the mainshock alone. For this reason, contrary to the normal practice in seismic hazard and risk assessment, the cascade of seismic events triggered during a

seismic sequence must be of interest, at least in the short term.

A cascade of earthquakes triggering earthquakes causes the seismicity to develop in complex patterns (see, e.g., Marsan and Lengliné 2008). Different physical mechanisms have been proposed to study mainly the direct triggering problem; these include static stress changes (i.e., the difference in the stress field just before an earthquake to shortly after the seismic waves have decayed, see, e.g., Hill et al. 2002; Manga and Brodsky 2006), quasi-static stress changes (e.g., associated with slow viscous relaxation of the lower crust and upper mantle beneath the epicenter of a large earthquake, see, e.g., Hill et al. 2002), dynamic stress changes (i.e., induced by the seismic waves from a large earthquake, see, e.g., Hill et al. 2002; Manga and Brodsky 2006), and fluid flow (e.g., migrating pore fluids and pore pressure changes as a mechanism controlling the timing of aftershocks, see, e.g., Harris 1998 and references therein). The distinction between static and dynamic stress transfer mechanisms breaks down at distances less than a fault length from a given earthquake. That region, the *near field*, is generally the spatial domain of aftershock occurrences (Hill et al. 2002).

The evident complexity of the triggering problem explains the objective difficulty to efficiently discriminate between direct and indirect triggering in the observed data, limiting the possibility of implementing pure deterministic approaches for the analysis of the events in a sequence. Therefore, the main efforts, as found in literature, attempt to identify and separate earthquakes between mainshocks and aftershocks, with the most sophisticated methods employing stochastic approaches. Our discussion in this section is restricted to the description of stochastic modeling of aftershock sequences.

The statistical properties of the aftershock sequences have long been of interest to study the earthquake generation processes. It has resulted in the development of a number of point process models (a type of random process for which any realization consists of a set of isolated points in time or space) to describe the standard activity of earthquake sequences.

Considering in particular the *trigger models* (Vere-Jones and Davies 1966), it is assumed that the conditional probability of a shock occurring at a time t after a given triggering mainshock is proportional to a *decay function* $\lambda(t)$. In general, it considers that the triggering events occur by following a Poisson process with a constant rate, while the number of triggered events (i.e., the group of shocks initiated by a single triggering event) decays by following a prescribed function. Regarding the nature of $\lambda(t)$, Vere-Jones and Davies (1966) considered in detail two functions:

- An exponential decay, $\lambda(t) = pe^{-pt}$, ($p > 0$)
- An inverse power-law decay, $\lambda(t) = pc^p / (c + t)^{p+1}$, ($c > 0$)

(where the model parameters arise from the Omori formula). Vere-Jones and Davies (1966) suggested that the inverse power-law type of decay function gives a better fit than the negative exponential. Considering a restricted form of the trigger model, Ogata (1988) proposed the epidemic-type aftershock sequences (ETAS) model. ETAS is a stochastic point process in which each earthquake has some magnitude-dependent ability to trigger its own Omori law-type aftershocks. The total occurrence rate can be described, in time, as the superposition of a background uncorrelated seismicity μ_0 and the events triggered by another earthquake:

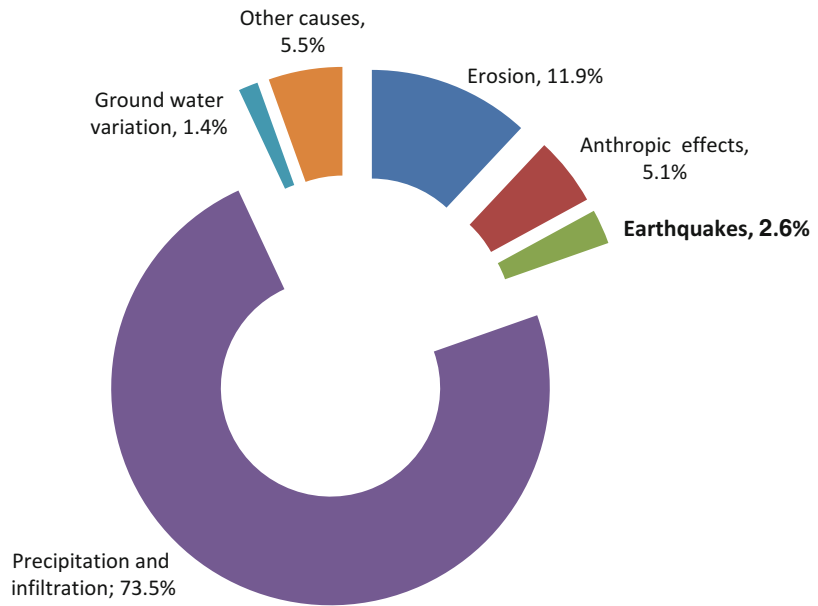
$$\lambda(t) = \mu_0 + \sum_{i:t_i < t} \lambda_i(t) \quad (2)$$

where $\lambda_i(t)$ is the rate of aftershocks induced by an event occurred at time t_i with magnitude M_i , defined as (Ogata 1988):

$$\lambda_i(t) = \frac{k}{(c + t - t_i)^p} e^{\alpha(M_i - M_c)} \quad (3)$$

for $t > t_i$. The parameter k measures the productivity of the aftershock activity; α defines the relation between triggering capability and magnitude M_i of a triggering event; c measures the incompleteness of the catalog in the earliest part of each cluster; p controls the temporal decay of

Seismic Risk Assessment, Cascading Effects, Fig. 3 Percentage of landslide events in Italy classified by the triggering mechanism (Source of the data: CNR-GNDCI – database of areas affected by landslides and floods in Italy)



triggered events; and M_c is the completeness magnitude of the data catalog. A space-time generalization of ETAS was later introduced by Ogata (1998). The ETAS model is widely used in different ways, as, for example, to describe short-term seismicity, in the forecasting of after-shock sequences (useful in short-term hazard studies) and also as a declustering method to separate the background seismicity from the triggered one.

From the perspective of considering cascading effects in the seismic risk assessment, the stochastic description of the seismic sequences (in particular by the ETAS model) has important consequences in probabilistic seismic hazard assessments, since it permits to take into account a more realistic behavior of the seismicity to calculate aftershock hazard curves, which is of fundamental importance for short-term hazard assessments.

It is worth noting here that considering the sequences of earthquakes in a chain of triggering effects can be regarded as an interaction at the hazard level (according to the taxonomy provided in this entry). Nevertheless, it may require also taking into account the cumulative effect of the different earthquakes on the exposed elements (e.g., the physical structures), which requires to consider also a possible interaction at the

vulnerability level by using damage-dependent fragilities.

Earthquakes as Triggering Mechanism of Landslides

Strong earthquakes may cause widespread landslides and other ground failure phenomena as, for example, liquefaction. A comprehensive historical review of outstanding past examples of earthquake-triggered landslides can be found in Keefer (2002) and references therein. In many cases the loss of lives and properties caused by the landslides exceed the losses caused by the earthquake itself (see, e.g., Chen et al. 2012). The occurrence of landslides can be associated with different triggering mechanisms as intense precipitation, erosion, ground shaking caused by earthquakes, groundwater variations, nonnatural effects of anthropic activity, etc. For example, let us consider the database from the “National Group for Prevention of Hydrological Hazards” of areas affected by landslides in Italy (CNR-GNDCI, <http://avi.gndci.cnr.it>); the most frequent kind of events triggering landslides in Italy are (see Fig. 3): the precipitation and infiltration (73.5 %), erosion (11.9 %), anthropic activity (5.1 %), earthquakes (2.6 %), and groundwater variation (1.4 %).

Taking as reference the available landslide catalog in Italy, about a 3 % of the total landslide occurrences in this area can be associated with earthquake-triggering mechanisms (Fig. 3). It is worth noting that this example is representative for the Italian area, and therefore the proportion of events for the different triggering phenomena may considerably change from one region to the other in function of different factors as, for example, the amount of precipitation, topography and land use, size and location of earthquakes, etc.

From the cascading effects assessment perspective, the occurrence of earthquakes in landslide-prone areas may increase the likelihood of landslide occurrence due to different physical mechanisms, as, for example, the inertial forces during the ground shaking itself or the rapid increase in pore water pressure induced by the cyclic stresses and subsequent reduction of the soil shear strength. Furthermore, earthquakes may reduce the soil strength for longer periods, causing the affected area to be more susceptible to post-earthquake rainfall-triggered landslides until the soil has regained its original strength.

Several models have been proposed for assessing the stability of slopes during earthquakes; an interesting overview of methods is provided in Jibson (2011). These methods are conveniently grouped into (1) pseudo-static analyses, (2) stress-deformation analyses, and (3) permanent displacement analyses. One of the most popular models in the latter group is that proposed by Newmark (1965), where the landslide is modeled as a rigid block sliding on an inclined surface. Movement of the block occurs when the sum of the static and dynamic driving forces exceed the shear resistance of the block. Permanent deformation occurs when the induced accelerations exceed some critical acceleration. The sliding block model assumes (1) a rigid-plastic behavior of the boundary between the surface and the sliding block, (2) the existence of a well-defined slip surface, (3) negligible loss in shear resistance during shaking, and (4) permanent displacement occurs only if dynamic stresses exceed shear resistance. A displacement can take place then during a cyclic load whenever the shear stress exceeds the base resistance, and

the block displacement ends when the block-base relative velocity equals zero. Displacement steps are summed over the entire duration of the acceleration-time history to obtain the total displacement. The method has undergone several modifications and improvements, and several relations between seismic ground-motion parameters and computed landslide displacements have been proposed. A particularly interesting application for quantitative hazard assessment of earthquake-triggered landslides is presented in Miles and Ho (1999), in which hazard analyses are performed combining Newmark's approach and Boore's stochastic ground-motion simulations. Likewise, stochastic methods for probabilistic earthquake-triggered landslide hazard assessment have been also proposed. For example, Jibson et al. (2000) performed a comparative analysis of the displacements modeled using the Newmark's permanent-deformation analysis and the landslide observations after the 1994 Northridge, California earthquake and used this information to construct a probability curve relating predicted displacement to probability of failure. Likewise, Refice and Capolongo (2002) presented a procedure that takes into account uncertainties and the spatial variability of geological, geotechnical, geomorphological, and seismological parameters in the Newmark slope stability model and applied it for probabilistic earthquake-triggered landslide hazard assessment.

Earthquakes Triggering Volcanic Eruptions

Deformation of the surrounding rocks is required for the magma movement before volcanic eruptions; therefore, small to moderate local earthquakes (usually called *volcano-tectonic* seismicity) normally precede and accompany eruptive activity. However, volcanic unrest episodes and volcanic eruptions have been observed within days after the occurrence of big earthquakes occurring outside the volcanic area. Evidences of tectonic earthquakes preceding volcanic activity can be found in different tectonic setting and on different time and distance scales (see, e.g., Eggert and Walter 2009; Gasparini 2013). Well-known cases of this

possible triggering effect may include, among many others, the 1990 M7.8 Luzon earthquake and the reawakening of Mount Pinatubo in 1991 (e.g., Bautista et al. 1996); the 1960 M9.5 Chile earthquake, which was followed about 38 h later by an eruption in the Puyehue-Cordón Caulle volcano (e.g., Barrientos 1994); Mount Vesuvius having the largest eruptions shortly after earthquakes in the Apennines (Nostro et al. 1998); and eruptions of different volcanoes in Kamchatka (some of them not considered active) following earthquakes in different periods (e.g., Karymsky and Dzenzur volcanoes erupting after an M8.5 earthquake in 1923, the Karpinsky volcano and the Tao-Rusyr Caldera erupting few days after an M9.0 earthquake, and the Karymsky and Akademia Nauk volcanoes erupting few days after a M7.1 earthquake, Walter 2007 and references therein).

Approximately 0.4 % of explosive volcanic eruptions occur within a few days of large, distant earthquakes, which is a value much greater than expected by chance (Manga and Brodsky 2006; Linde and Sacks 1998). All the available examples are suggestive of a triggering connection, but quantifying that connection is not straightforward. Linde and Sacks (1998) analyzed the historical record of large earthquakes and explosive eruptions and found that within a day or two of large earthquakes, there are many more eruptions within a range of about 750 km than would otherwise be expected. Therefore, they concluded that eruptions occur in the vicinity of a large earthquake more often than would be expected by chance. Additionally, it is well known that volcanoes separated by hundreds of kilometers frequently erupt in unison; the characteristics of such eruption pairs are also consistent with the hypothesis that the second eruption is also triggered by the earthquakes associated with the first (Linde and Sacks 1998). There is also strong evidence that large earthquakes can affect the unrest condition of volcanoes producing transient enhancement of microseismic and hydrothermal activities as well as ground deformation (see Gasparini 2013).

As for the earthquake-earthquake interactions, several mechanisms have been proposed to

plausibly explain the volcanic eruption triggering by large earthquakes, being the most widely discussed those associated with transient (or dynamic), viscoelastic (or quasi-static), and permanent (or static) stress and strain changes. In fact, earthquakes can stress magmatic systems either through static stresses (e.g., the offset of the fault generates a permanent deformation in the crust), viscoelastic stresses (e.g., a change in the stress state in the elastic crust associated with slow, viscous relaxation of the lower crust and upper mantle beneath the epicenter of a large earthquake), or dynamic stresses from the seismic waves (Hill et al. 2002; Manga and Brodsky 2006).

Dynamic stresses are proportional to the seismic wave amplitude, which is influenced by different factors as directivity, radiation pattern, and crustal structure. However, the significant difference in dependence on distance is a robust feature distinguishing static and dynamic stresses. The difference in decay rate helps to identify which stresses may be important for triggering eruptions. Static stress changes die off rapidly with distance (as $\sim 1/r^3$, with r the distance from the epicenter); therefore, their influence is limited to distances of a few fault lengths (near field). Quasi-static stress changes die off more slowly with distance (as $\sim 1/r$ or $1/r^2$), similarly to the dynamic stress changes that die off as $\sim 1/r$ for surface waves and $\sim 1/r^2$ for body waves (e.g., Hill et al. 2002; Manga and Brodsky 2006).

To provide a comparative example, Table 1 lists the order of magnitude of earthquake stresses

Seismic Risk Assessment, Cascading Effects, Table 1 Static and dynamic stress changes of an M8.0 earthquake at different distances from the source (From Manga and Brodsky 2006)

M8.0 earthquake	Distance from the source		
	Stress (MPa)		
	10 ² km	10 ³ km	10 ⁴ km
Static stress changes	1.0 × 10 ⁻¹	1.0 × 10 ⁻⁴	1.0 × 10 ⁻⁷
Dynamic stress changes	3.0	6.0 × 10 ⁻²	1.0 × 10 ⁻³

Seismic Risk Assessment, Cascading Effects, Table 2 Order of magnitude of other sources of forcing (From Manga and Brodsky 2006)

Phenomena	Stress MPa
Solid earth tides	10^{-3}
Ocean tides	10^{-2}
Hydrological loading	10^{-3} to 10^{-1}
Glacier loading	10^1 to 10^2

at different distances from the source, and Table 2 lists other external stresses that have been also proposed to influence volcanoes (example from Manga and Brodsky 2006). In the near field, both static and dynamic stresses may exhibit significant values; conversely, in the far field only the dynamic stresses are larger than other potential triggering stresses proposed; for instance, for distances greater than 10^3 km, the static stress changes are at least one order of magnitude lower than the other forcing sources presented in Table 2.

Several models have been proposed to explain the different triggering mechanisms. Here we summarize the main examples presented in Hill et al. 2002 and in Manga and Brodsky 2006 (and references therein) for the static and dynamic stress triggering. The reader is invited to consult these references for more detailed information.

Static stress triggering: Static stress changes can be a factor triggering volcanic eruptions at intermediate distances (few hundred kilometers). Explanations for this kind of interaction are based on pressure changes in a magma body induced by the isotropic, compressional component of the stress field in the vicinity of the volcano. Increased compressional stress in the crust surrounding an already fully charged magma chamber may squeeze magma upward. Likewise, a decrease in compressional stress can promote vesiculation, bubble formation, and unclamping of conduits above the magma body (Hill et al. 2002).

Dynamic stress interactions: Several models have been proposed to explain the phenomenon of dynamic triggering by distant earthquakes but the physical processes remain a matter of discussion. Considering in particular those mechanisms in which fluids are active in the triggering

process, examples of the models proposed include (Hill et al. 2002):

- Excitation of bubbles in a gas-saturated fluid by passing seismic waves, which acts to increase fluid pressure through advective overpressure and/or rectified diffusion (see, e.g., Hill et al. 2002; Manga and Brodsky 2006).
- Hydraulic surge, a model in which the seismogenic crust is separated from the underlying plastic crust and an embedded magma body by an impermeable transition zone; a hydraulic surge occurs when the large dynamic strains associated with seismic waves from a large, distant earthquake disrupt the impermeable seal.
- Rupturing blockages in confined aquifers, in which passing seismic waves disrupt blockages in fractures formed by precipitates in a confined aquifer allowing local adjustments in fluid pressure within the aquifer.
- Sinking crystal plume model, in which relatively dense, loosely held masses of crystals can accumulate on the ceiling and walls of a crystallizing magma body and might be dislodged by passing seismic waves. As crystal-rich plumes sink, less dense, crystal-poor magma will be forced upward initiating a convective cycle. Volatiles in the ascending plume will form bubbles and a pressure increase through advective overpressure.
- Relaxing magma body: A magma body may reach a partially crystallized state by either (1) gradual cooling of a previously molten volume or (2) partial melting of a previously solid volume. The solid matrix will partially support accumulating tectonic stresses in the region. Strong seismic waves may disrupt the solid matrix thereby redistributing stress to the surrounding crust as the magma body deforms through viscous relaxation.

Different observations worldwide indicate then that the triggering effect between earthquakes and volcanic eruptions may exist under certain conditions. Stress changes, in particular, seem to be a plausible mechanism capable of triggering volcanic eruptions in a wide range of

distances and times. Given the complexity of the problem and of the interactions, until now the problem is generally treated statistically. New observations in currently well-monitored volcanoes, statistical significance of the correlations, and physics of the triggering mechanisms are important research topics that need to be addressed in order to better constrain this kind of hazard interaction.

Earthquakes Triggering Industrial Accidents

Damages in industrial plants are known to have triggered a number of severe accidents caused mainly by the loss of containment of hazardous substances. Industrial accidents can be triggered by a number of causes, the most common of which are internal system failures and/or human errors (Rasmussen 1995). However, industrial facilities may also be subject to accidents triggered by external factors as natural events. Analyzing different databases of industrial accidents in Europe and the USA, Kraussmann et al. (2011) outlined that between 2 % and 5 % of the recorded accidents were triggered by natural hazards (so-called Natech accidents), being this probably a low bound because of a reporting bias towards accidents with severe consequences.

From the perspective of cascading effects, earthquakes are one of the possible natural hazards triggering Natech accidents. Earthquakes may cause damages in industrial facilities by the direct impact caused by the ground shaking and also by secondary effects as liquefaction. Considering the results presented by Kraussmann et al. (2011), pipework, pipelines, and tanks for the atmospheric storage of chemicals appear to be particularly vulnerable to excitation and damage by the earthquake forces. In the analyzed cases, about 27 % of the accidents presented only structural damages, whereas about 73 % of the cases exhibited damage with release of hazardous materials. These results seem to suggest that chemical equipment is much more likely to suffer damage with loss of containment than not (Kraussmann et al. 2011). Therefore, scenarios of cascading effects with earthquakes as triggering mechanism causing Natech accidents

(because of damages in industrial facilities) can be followed by other triggered hazards as, for example, the release of hazardous materials, explosions, and fires.

A recent outstanding example of cascading effects leading to industrial accidents triggered by earthquakes is the case of the Tohoku earthquake, Japan. On March 11, 2011, the Tohoku-Oki earthquake generated a tsunami 130 km off the coast of Miyagi Prefecture, northeast Japan. The tsunami first reached the Japanese mainland about 20 min after the earthquake and ultimately affected a 2,000 km stretch of Japan's Pacific coast (e.g., Mori et al. 2011). The tsunami caused nuclear accidents, primarily the damages at three reactors in the Fukushima Daiichi Nuclear Power Plant complex. Accounting for all of the factors relevant during the Tohoku disaster, Khazai et al. (2011) analyzed the cascading effects produced by the main earthquake and both the tsunami and the nuclear power plant incidents and produced a causal map depicting the direct and indirect interactions between the various events.

Considering Cascading Effects in the Seismic Risk Assessment

Taking as reference the general concepts described in Eq. 1, in more quantitative terms, risk is often quantified using an expression of the form (e.g., Cornell and Krawinkler 2000; Marzocchi et al. 2012):

$$\lambda(\ell) = \int_D \int_{I_m} \underbrace{F(\ell|D)}_{\text{damgae \& loss}} \underbrace{dG(D|I_m)}_{\text{hazard}} d\lambda(I_m) \quad (4)$$

where $\lambda(I)$ represents a rate of loss, $F(I|D)$ is the probability that a given loss is reached given the damage state D , $G(D|I_m)$ is the vulnerability (or fragility) term, $\lambda(I_m)$ is the hazard, measuring the exceedance rates of given intensity measures (I_m) characterizing the hazard, I is the loss in one specific metric, and D is a given damage state.

Assessing interactions, or cascading effects, in a multi-hazard problem consists on the identification of the possible interactions that are likely to happen and that may result in an amplification of the expected losses in a given area of interest. This process may be performed in the following steps:

1. Identification and structuring of the plausible scenarios of cascading effects
2. Identification of the kind of interactions according to the proposed taxonomy
3. Quantification of the scenarios of interest

Assuming that the value of the exposed elements remains constant in a given time window of interest, the amplification of the losses may be caused by an increase in the occurrence probability of the hazard or by an increase in the probability of damage by a change in the fragility of the exposed elements. In some cases both elements may be considered in a single scenario. In the first case, the analysis of interactions updates the hazard term in Eq. 4; in the second case, the damage term in Eq. 4 is updated.

Identification and Structuring Scenarios of Cascading Effects

An important initial step towards the assessment of cascading effects is the identification and structuring of the possible scenarios. The term “scenario” is used in a wide range of fields and so different interpretations can be found in practical applications. In general, a scenario may be considered as a plausible and consistent representation of an event or series of actions and events. It must be plausible because it must fall within the limits of what might conceivably happen and must be consistent in the sense that the combined logic used to construct a scenario must not have built-in inconsistencies.

Different strategies can be adopted in order to exhaustively identify a complete set of scenarios. In general, they are based on different complementary approaches ranging from event-tree-like to fault-tree-like strategies. In many applications, an adaptive method

combining both kinds of approaches is applied in order to ensure an exhaustive exploration of scenarios. A wide description of practical methods for scenario identification and structuring can be found, for example, in Haimes (2009).

Taxonomy of Interactions

Taking as reference the works of Marzocchi et al. (2012), Garcia-Aristizabal et al. (2013), Selva (2013), and Garcia-Aristizabal et al. (forthcoming-2015), it is possible to define a generalized taxonomy of interactions considering two possible kinds of interactions, namely, (1) interactions at the hazard level, in which the occurrence of a given initial *triggering* event entails a modification of the probability of occurrence of a secondary event and (2) interaction at the vulnerability (or damage) level, in which the main interest is to assess the effects that the occurrence of one event (the first one occurring in time) may have on the response of the exposed elements against another event (that may be of the same kind as the former but also a different kind of hazard). Implicitly, a combination of both kinds of interactions is another possible case; therefore, in the discussion of the interactions at the vulnerability level, both dependent and independent hazards can be considered as possible cases.

Triggering Effects: Interactions at the Hazard Level

In this case, the interaction problem is understood as the assessment of possible “chains” of adverse events in which the occurrence of a given initial *triggering* event entails a modification of the occurrence probability of a secondary event. Even if this typology of problems can be assessed in a long-term basis, their utility can be highlighted in short-term assessments, in which specific scenarios conditioned to the occurrence of given specific events can be assessed and compared among them.

Let us first consider two different threatening events, whose occurrences are E_1 and E_2 . In general, the probability of E_2 occurring (generically hazard H_2) can be written as (Marzocchi et al. 2012):

$$H_2 = p(E_2) \\ = p(E_2|E_1)p(E_1) + p(E_2|\bar{E}_1)p(\bar{E}_1) \quad (5)$$

where, p represents a probability or a distribution of probability and $\bar{E}_1$ means that the event E_1 does not occur. Equation 5 represents the total probability of E_2 occurring considering generically the occurrence or nonoccurrence of E_1 . However, the quantification of risk generally requires the hazard assessed in function of a specific intensity measure (IM) characterizing the hazard, and therefore Eq. 5 can be generalized in order to consider the specific contribution of different IM values. Therefore, the occurrence of the triggered event 2 (E_2) given the occurrence of the triggering event 1 (E_1) representing this kind of interaction can be assessed as (Garcia-Aristizabal et al. 2013):

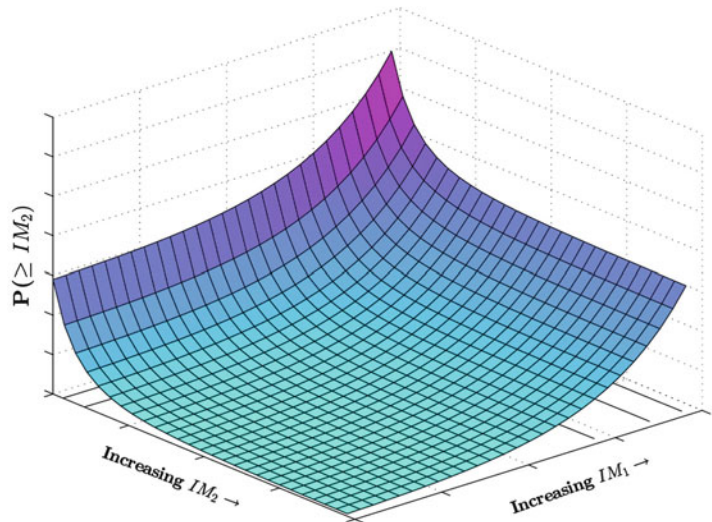
$$p(\geq IM_2^i) = \sum_j p(\geq IM_2^{i,j}) \quad (6)$$

that explicitly conditioned to E_1 can be written as (e.g., Garcia-Aristizabal et al. 2013; Selva 2013; Garcia-Aristizabal et al. forthcoming-2015):

$$p(\geq IM_2^i) = \sum_j p(\geq IM_2^i | IM_1^j) p(IM_1^j) \quad (7)$$

For $j = 1, 2, 3, \dots, n$, for the n (exhaustive and mutually exclusive) classes of IM defined for the triggering event. IM_1 is the intensity measure defined for the triggering event, and IM_2 is the intensity measure defined for the triggered event. Note that Eq. 7 is a generalization of Eq. 5. A key element in this formalization is $p(\geq IM_2^{i,j})$, which we call the *interaction term*. When $i > 1$ and $j > 1$, $p(\geq IM_2^{i,j})$ is represented by a $i \times j$ matrix that defines a probability surface, as shown in Fig. 4. When $i = 1$ or $j = 1$, $p(\geq IM_2^{i,j})$ is a row or column vector, and in the hypothetical limit case in which the problem does not consider the intensity measures of both the triggered and triggering events, the problem is simplified as the case described by Eq. 5 (i.e., considering the “occurrence” and “nonoccurrence” of the event 1). Considering, for example, the earthquake-triggered landslide case, the earthquake is represented by E_1 and the triggered landslides are the E_2 ; IM_1 can represent the ground-motion intensity and IM_2 the volume of the mass sliding. Therefore, $p(\geq IM_2^i | IM_1^j)$ represents the probability of exceeding a given sliding mass volume given a certain ground shaking intensity, and $p(\geq IM_2^i)$ represents the *total* probability of exceeding a given sliding mass volume considering the full range of possible ground shaking intensities taken into account from the triggering earthquake hazard.

Seismic Risk Assessment, Cascading Effects, Fig. 4 Probability surface represented by $p(\geq IM_2^{i,j})$, for $i > 1$ and $j > 1$



Additive Effects of Different Events: Interactions at the Vulnerability Level

This perspective of the cascading effect problem fundamentally intends to assess the effects that the simultaneous occurrence of two or more events (not necessarily linked among them) may have for the final risk assessment. In this case, the action of different hazards is considered and combined at the vulnerability (or damage) level, and the main interest is to assess the effects that the occurrence of one event (the first one occurring in time) may have on the response of the exposed elements against another event (that may be of the same kind as the former but also a different kind of hazard). Examples of this kind of interactions have been presented in literature, for example, in Lee and Rosowsky (2006), Zuccaro et al. (2008), Marzocchi et al. (2012), Garcia-Aristizabal et al. (2013), and Selva (2013).

In practice, the idea in this case is to quantify how the expected damages in the target area (caused by a given hazard) can be modified if another hazardous event acts on the exposed elements simultaneously or in a short-time window (in general short enough that the system cannot be repaired). In the case of two hazards having additive load effects (i.e., they act simultaneously over the exposed element), the fragility function will depend on the intensities IM_1 and IM_2 of the two hazards, and then it will represent a fragility surface. The probability that a given damage state (D_k) is reached given the occurrence of the i -th value of IM_1 and the j -th value of IM_2 can be represented as (for details see Lee and Rosowsky 2006; Garcia-Aristizabal et al. 2013; Selva 2013):

$$p(D_k) = \sum_i \sum_j \left[P(D_k | IM_1^i \cap IM_2^j) p(IM_1^i \cap IM_2^j) \right] \quad (8)$$

The conditional probability $P(D_k | IM_1^i \cap IM_2^j)$, hereinafter referred to as $P(D_k | IM_1^i, IM_2^j)$ for simplicity, is the probability that the damage state D_k is reached at given levels of loads IM_1 and IM_2 (due to hazard events 1 and 2, respectively) acting simultaneously. Therefore, to calculate $p(D_k)$

considering any value of the IM of the events 1 and 2, we need to consider two cases, namely, (1) when events 1 and 2 are independent and (2) when there is dependence between events 1 and 2.

Independent Events In the simpler case, we consider the two hazards as independent events, in the sense that the occurrence of one does not change the probability of the other occurring. In that case, the probability that a given damage state is reached for all the possible values of IM_1 and IM_2 can be defined as (e.g., Garcia-Aristizabal et al. 2013):

$$p(D_k) = \sum_i \sum_j \left[P(D_k | IM_1^i, IM_2^j) p(IM_1^i) p(IM_2^j) \right] \quad (9)$$

which is an expression coherent with the framework defined by Lee and Rosowsky (2006) analyzing the effects of the combined seismic and snow loads on wood structures. In this case, the fragility function $P(D_k | IM_1^i, IM_2^j)$ will represent a probability surface as that shown in Fig. 5. The example presented in Fig. 5 shows the combined effect of seismic (in pga) and volcanic ash (in Kpa) loads acting simultaneously on roof structures in Naples, Italy. The data used to generate this figure was derived from the work presented in Zuccaro et al. (2008).

Dependent Events If the occurrence of one event (E_1) does affect the probability of the other occurring (E_2), then the events are dependent. This is the case described in the interactions at the hazard level. In this case, the term $p(IM_2^j \cap IM_1^i) = p(IM_2^j | IM_1^i) p(IM_1^i)$ and then the probability that a given damage state D_k is reached for the case of dependent events can be defined as:

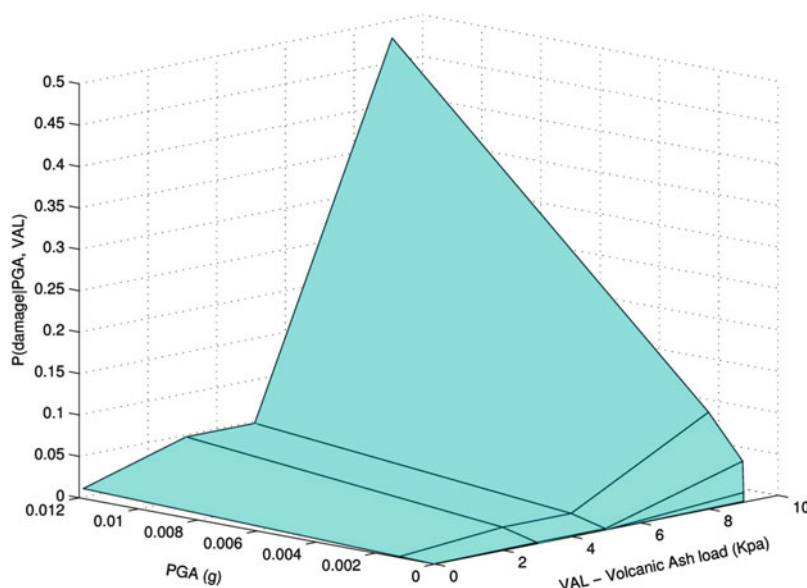
$$p(D_k) = \sum_i \sum_j \left[P(D_k | IM_2^j, IM_1^i) p(IM_2^j | IM_1^i) p(IM_1^i) \right] \quad (10)$$

In this case, $P(D_k | IM_2^j, IM_1^i)$ is again a fragility surface as that represented in Fig. 5, and the

Seismic Risk

Assessment, Cascading Effects, Fig. 5

Fragility surface for the case of seismic and volcanic ash acting simultaneously on the roofs in Naples, Italy (Derived from the data in Zuccaro et al. 2008). Note that when the ash load tends to zero, the fragility function converges towards the pure seismic fragility



hazard term takes into account the possible dependence between events 1 and 2. For example, considering again the earthquake-triggered landslide case, assessing the “collapse” damage state would require the fragility surface $P(D_k | IM_2^l, IM_1^l)$ for assessing the probability of collapse given the combined effect of the ground shaking intensity (IM_1) and, for example, the depth of the deposit caused by the mass movement (IM_2). On the other hand, the hazard term in this case would be calculated as shown in the example presented for the *interactions at the hazard level*, in which the probability of exceeding a given IM_2 value is assessed considering the full range of likely intensities of the triggering ground motions caused by the earthquakes.

Damage-Dependent Vulnerabilities in Cascading Effects

The damage-dependent vulnerabilities (or fragilities) can be considered as a special case of the interaction at the vulnerability level. In fact, the development of fragility surfaces, as presented in the previous section for that kind of interactions, is intended for general cases in which the additive effect of different loads may act simultaneously over the exposed element, which does not necessarily imply a pre-damage of the element. For example, the action of the first occurring hazard

may just result in a modification of the response of the exposed element (e.g., a structure) to another typology of hazard without causing a permanent deformation or damage that remains in time. Examples of this case can be snow or volcanic ash deposited over a roof; this extra weight in the structure may cause it to behave differently if it is shaken by the ground motion caused by an earthquake (compared with the normal structure’s behavior). However, in many cases this interaction may imply that the first acting event produces permanent deformations in the structure (i.e., causing some damages) implying some modification to its response to new solicitations and the need to develop damage-dependent fragility functions.

The most frequent development of damage-dependent fragility functions is found for seismic loads (i.e., to assess the structure’s response to new earthquakes after it has been damaged by the past earthquakes). Different works tackling the problem of the response of structures pre-damaged by earthquakes can be found in literature and mainly associated with the assessment of post-mainshock time-dependent risk. Examples can be found in (Bazzurro et al. 2004; Yeo and Cornell 2005; Ryu et al. 2011; Luco et al. 2011) and references therein. In general, the fragility curves used for the post-mainshock

risk assessment accounts for any damage caused by the mainshock. This typically increases the probabilities of collapse for the considered range of potential (future) IM values, and the amount of increase depends of course on the extent of the mainshock damage, which is commonly discretized into damage states (DS) as, for example, none, slight, moderate, extensive, and complete (Luco et al. 2011). Following Ryu et al. (2011, and references therein), the fragility for a mainshock-damaged structure can be computed as:

$$p(D_a > d_a | IM_a = im_a, D_m = d_m) \\ = \int P(D_a > d_a | IM_a = im_a, EDP_m = edp_m) \times \\ f(EDP_m = edp_m | D_m = d_m) dedp_m \quad (11)$$

where D_a represents the post-aftershock damage state, D_m represents the post-mainshock damage state, EDP_m , the engineering demand parameter (e.g., drift), represents the mainshock building response, and IM_a represents the ground-motion intensity of an aftershock (Ryu et al. 2011). The first term in Eq. 11 can be computed using the fragility for the intact building or that for a mainshock-damaged building whose mainshock response (EDP_m) is edp_m . The second term in Eq. 11 can be computed using the assumed distribution of mainshock response given post-mainshock damage state (for details see Ryu et al. 2011; Luco et al. 2011).

Summary

Different past disasters have highlighted the fact that natural or man-made events can trigger other events, leading to a non-negligible increase of fatalities and damages. As a consequence, there is an increasing interest on implementing multi-type hazard and risk assessments as well as assessing cascading effects, which are one of the fundamental concepts of the multi-risk assessment. As for most of the quantitative risk assessment for natural hazards, the seismic risk assessment is generally carried out separately

without considering the effects of possible chains of events triggered by earthquakes. From the risk assessment point of view, the cascades of events occurring after an earthquake may represent a non-negligible source of loss amplifications. For quantitative purposes, we can consider the cascading effects analyzing two major sets of interactions: (1) interactions at the hazard level and (2) interactions at the vulnerability level. In this entry we present and discuss different examples of scenarios of cascading effects triggered by earthquakes.

Cross-References

- [Analytic Fragility and Limit States \[P\(EDP|IM\)\]: Nonlinear Dynamic Procedures](#)
- [Analytic Fragility and Limit States \[P\(EDP|IM\)\]: Nonlinear Static Procedures](#)
- [Early Earthquake Warning \(EEW\) System: Overview](#)
- [Earthquake Mechanism and Seafloor Deformation for Tsunami Generation](#)
- [Earthquake Mechanisms and Stress Field](#)
- [Earthquake Swarms](#)
- [Economic Impact of Seismic Events: Modeling](#)
- [Empirical Fragility](#)
- [Probabilistic Seismic Hazard Models](#)
- [Seismic Fragility Analysis](#)
- [Seismic Monitoring of Volcanoes](#)

References

- Barrientos SE (1994) Large thrust earthquakes and volcanic eruptions. *Pure Appl Geophys* 142(1): 225–237
- Bautista BC, Bautista MLP, Stein RS, Barcelona ES, Punongbayan RS, Laguerta EP, Rasdas AR, Ambubuyog G, Amin EQ (1996) Relationship of regional and local structures to Mount Pinatubo activity. In: *Fire and mud; eruptions and lahars of Mount Pinatubo, Philippines*. Philippine Institute of Volcanology and Seismology/University of Washington Press United States, Quezon City/Philippines
- Bazzurro P, Cornell CA, Menun C, Motahari M (2004) Guidelines for seismic assessment of damaged buildings. In: *Proceedings of the 13th world conference on earthquake engineering, Vancouver*

- Chen XL, Zhou Q, Ran H, Dong R (2012) Earthquake-triggered landslides in southwest China. *Nat Hazards Earth Syst Sci* 12:351–363. doi:10.5194/nhess-12-351-2012
- Cornell CA, Krawinkler H (2000) Progress and challenges in seismic performance assessment. *PEER Center News* 3(2), URL <http://peer.berkeley.edu/news/2000spring/index.html>
- Eggert S, Walter TR (2009) Volcanic activity before and after large tectonic earthquakes: observations and statistical significance. *Tectonophysics* 471:14–26
- Garcia-Aristizabal A, Marzocchi W, Di Ruocco A (2013) Probabilistic framework for multi-hazard assessment. Technical report D3.4, MATRIX project (New methodologies for multi-hazard and multi-risk assessment methods for Europe), grant no. 265138. URL <http://matrix.gpi.kit.edu/Deliverables.php>
- Garcia-Aristizabal A, Gasparini P, UHINGA G (forthcoming-2015) Multi-risk assessment as a tool for decision-making. In: Pauleit S, Jorgensen G, Kabisch S, Gasparini P, Fohlmeister S, Simonis I, Yeshitela K, Coly A, Lindley S, Kombe WJ (eds) *Future Cities*, vol 4 Climate change and urban vulnerability in Africa: a multidisciplinary approach. Springer Netherlands, ISBN 978-3-319-03984-8
- Gasparini P (2013) The physics of volcanoes since the 1960s. *Rivista Nuovo Cimento* 36:461–500. doi:10.1393/ncr/i2013-10093-0
- Haimes Y (2009) Risk modeling, assessment, and management, 3rd edn. Wiley, Hoboken
- Harris RA (1998) Introduction to a special section: stress triggers, stress shadows, and implications for seismic hazards. *J Geophys Res* 103: 24,347–24,358
- Hill DP, Pollitz F, Newhall C (2002) Earthquake-volcano interactions. *Physics Today* 55(11):41–47
- Jibson RW (2011) Methods for assessing the stability of slopes during earthquakes – a retrospective. *Eng Geol* 122:43–50
- Jibson RW, Harp EL, Michael JA (2000) A method for producing digital probabilistic seismic landslide hazard maps. *Eng Geol* 58(3–4):271–289. doi:10.1016/S0013-7952(00)00039-9
- Keefer DK (2002) Investigating landslides caused by earthquakes – a historical review. *Surv Geophys* 23:473–510
- Khazai B, Daniell JE, Wenzel F (2011) The March 2011 Japan earthquake: analysis of losses, impacts, and implications for the understanding of risks posed by extreme events. *TECHNIKFOLGENABSCHÄTZUNG, Theorie und Praxis*. N3, 20 Dec 2011. S. 22–33. <http://www.itas.fzk.de/tatup/113/khual1a.htm>
- Kraussmann E, Renzi E, Campedel M, Cozzani V (2011) Industrial accidents triggered by earthquakes, floods and lightning: lessons learned from a database analysis. *Nat Hazards*. doi:10.1007/s11069-011-9754-3
- Lee KH, Rosowsky DV (2006) Fragility analysis of woodframe buildings considering combined snow and earthquake loading. *Struct Saf* 28(3):289–303. doi:10.1016/j.strusafe.2005.08.002
- Linde AT, Sacks IS (1998) Triggering of volcanic eruptions. *Nature* 395:888–890. doi:10.1038/27650
- Luco N, Gerstenberger MC, Uma SR, Ryu H, Liel AB, Raghunandan M (2011) A methodology for post-mainshock probabilistic assessment of building collapse risk. In: *Proceedings of the ninth pacific conference on earthquake engineering building an earthquake-resilient society*, Auckland
- Manga M, Brodsky E (2006) Seismic Triggering of Eruptions in the Far Field: Volcanoes and Geysers. *Annual Review of Earth and Planetary Sciences* 34:263–291. doi:10.1146/annurev.earth.34.031405.125125
- Marsan D, Lengliné O (2008) Extending earthquakes' reach through cascading. *Science* 319(5866): 1076–1079. doi:10.1126/science.1148783
- Marzocchi W, Garcia-Aristizabal A, Gasparini P, Mastellone ML, Di Ruocco A (2012) Basic principles of multi-risk assessment: a case study in Italy. *Nat Hazards* 62(2):551–573. doi:10.1007/s11069-012-0092-x
- Miles SB, Ho CL (1999) Rigorous landslide hazard zonation using Newmark's method and stochastic ground motion simulation. *Soil Dyn Earthquake Eng* 18:305–323
- Mori N, Takahashi T, Yasuda T, Yanagisawa H (2011) Survey of 2011 Tohoku earthquake tsunami inundation and run-up. *Geophys Res Lett* 38:L00G14. doi:10.1029/2011GL049210
- Newmark NM (1965) Effects of earthquakes on dams and embankments. *Geotechnique* 15(2):139–160
- Nostro C, Stein RS, Cocco M, Belardinelli ME, Marzocchi W (1998) Two-way coupling between Vesuvius eruptions and southern Apennine earthquakes, Italy, by elastic stress transfer. *J Geophys Res B Solid Earth* 103(B10):24487–24504
- Ogata Y (1988) Statistical models for earthquake occurrences and residual analysis for point processes. *J Am Stat Assoc* 83:9–27
- Ogata Y (1998) Space-time point-process models for earthquake occurrences. *Ann Inst Stat Math* 50:379–402
- Plafker G, Erickson GE, Concha JF (1971) Geological aspects of the 31 May 1970, Peru earthquake. *Bull Seis Soc Am* 61:543–578
- Rasmussen K (1995) Natural events and accidents with hazardous materials. *J Hazard Mater* 40:43–54
- Refice A, Capolongo D (2002) Probabilistic modeling of uncertainties in earthquake-induced landslide hazard assessment. *Comput Geosci* 28(6):735–749. doi:10.1016/S0098-3004(01)00104-2
- Ryu H, Luco N, Uma SR, Liel AB (2011) Developing fragilities for mainshock-damaged structures through incremental dynamic analysis. In: *Proceedings of the ninth pacific conference on earthquake engineering building an earthquake-resilient society*, Auckland

- Selva J (2013) Long-term multi-risk assessment: statistical treatment of interaction among risks. *Nat Hazards* 67(2):701–722
- Vere-Jones D, Davies RB (1966) A statistical survey of earthquakes in the main seismic region of New Zealand. *N Z J Geol Geophys* 9(3):251–284. doi:10.1080/00288306.1966.10422815
- Walter TR (2007) How a tectonic earthquake may wake up volcanoes: stress transfer during the 1996 earthquake–eruption sequence at the Karymsky Volcanic Group, Kamchatka. *Earth Planet Sci Lett* 264(3–4):347–359
- Yeo GL, Cornell CA (2005) Stochastic characterization and decision based under time-dependent aftershock risk in performance-based earthquake engineering. Technical report 2005/13. Pacific Earthquake Engineering Research Center (PEER)
- Zuccaro G, Cacace F, Spence R, Baxter P (2008) Impact of explosive eruption scenarios at Vesuvius. *J Volcanol Geotherm Res* 178(3):416–453

Seismic Robustness Analysis of Nuclear Power Plants

Manuel Pellissetti and Ulrich Klapp
AREVA GmbH, Erlangen, Germany

Synonyms

Seismic margin assessment; Seismic PRA; Seismic probabilistic risk analysis; Seismic probabilistic safety analysis; Seismic PSA; SMA

Introduction

Nuclear Safety

Commercial nuclear power plants (NPP) are industrial facilities that represent a significant hazard potential. Consequently, the pertinent safety requirements are particularly stringent. Nuclear safety is assured by deterministic and probabilistic analysis. In deterministic analysis, the safety systems are shown to withstand design basis accidents (DBAs), typically using conservative assumptions. In probabilistic safety analysis (PSA), the frequency of accident progressions leading to core damage (PSA level 1) and large

early release (PSA level 2) is quantified and compared to target values that are judged as low enough.

Consideration of Earthquakes in Nuclear Safety

Earthquakes are external hazards on which the duty holder of a nuclear facility, e.g., a nuclear power plant, has no or only little control.

Effects to be expected at a site from an earthquake especially include vibrations induced in structures, systems, or components (SSC) through the civil structures of the plant. These vibrations could affect the plant safety functions directly, e.g., when the induced seismic loads would exceed the capacity of safety relevant equipment items. Indirect failure modes such as mechanical interaction, dropped loads, release of hazardous substances, seismic-induced fire or flooding, impairment of operator access, or unavailability of evacuation and access routes may also be of concern (see IAEA 2003a, §2.11).

Current design practice for a nuclear facility ensures an adequate seismic design from a deterministic perspective.

A safe shutdown earthquake (SSE) is selected based on the results of a seismic hazard analysis of the site. The SSE represents a site-specific extreme seismic event. The attribute “extreme” is justified because the SSE is defined so conservatively that it can be expected with a very high level of confidence that it will not be exceeded during the operational lifetime of the plant. SSCs necessary to transfer the plant into a safe shutdown state are identified and designed to remain functional under SSE loading. Similarly, SSCs whose seismic-induced failure could lead to interaction effects on safety equipment are identified and designed for integrity or stability as required by safety analysis. The demonstration that the design requirements are met consists in mechanical analysis or shaking table testing of the selected systems, structures, and components (SSC), subject to vibrations induced by the SSE.

The described provisions ensure that even in the case of an extreme seismic event, represented by the SSE, the main safety functions (reactivity

control, residual heat removal, confinement of radioactive substances) are met.

Motivation for Evaluating Seismic Robustness

What was described above is – in very simplified terms – the procedure for ensuring a safe initial design of an NPP, as far as seismic loads are concerned. Besides that, nuclear safety is also concerned with ensuring continuously that it is safe to operate the plant, e.g., by means of periodic safety evaluations or if new insights are gained.

Indeed, during the plant's lifetime, new information becomes available which may challenge the initial design basis of the plant. This new information may especially be of seismological nature (e.g., newly discovered seismogenic structures, newly installed seismological networks, new paleoseismological evidence or actual occurrence of damaging earthquakes in the site vicinity). However, there may also be new technical findings such as the indication that certain SSC may be more vulnerable than initially expected (see IAEA 2009, §2.10).

In consequence the nuclear industry and regulatory authorities are frequently faced with the question whether these changes can be accommodated within the seismic capacity of the original design or whether modifications are necessary to maintain an adequate level of safety (see ASCE 2000, p. 41).

When assessing the safety of a nuclear facility, the plant performance in seismic events exceeding the ground motion characteristics of the SSE (typically associated with mean return periods ≤ 10.000 years, see below) is also of interest. See, for example, §2.39 in IAEA (2003a): *“Seismic design should be carried out [...] to provide margins for seismic events that are beyond the design basis and to prevent potential small deviations in plant parameters from giving rise to severely abnormal plant behaviour ('cliff edge' effects). [...]”*

Last but not least, the severe accidents which occurred at Japan's Fukushima-Daiichi plants in response to the Great Tohoku Earthquake of March 2011 have stimulated an increased

interest in studies quantifying the actual margin of failure (“Seismic Margin Assessment” – SMA) or providing an insight into the risk associated to earthquakes (“Seismic Probabilistic Safety Analysis” – SPSA, often also referred to as “Seismic Probabilistic Risk Assessment” – SPRA).

Seismic Hazard Assessment

Before introducing the two main approaches for assessing the robustness of NPPs, it is necessary to briefly introduce the notion of seismic hazard.

The assessment of the site-specific seismic hazard – i.e., the characterization of the ground motions to be “expected” at a particular site – is a critical ingredient of both the (initial) design against seismic loads and periodic safety reviews during the plant's lifetime.

Seismic events are unpredictable (stochastic) as far as three of their most basic characteristics are concerned: timing, i.e., when they occur; magnitude; and location of the epicenter. The unpredictability of the latter two properties implies that the ground motion characteristics are also unpredictable. As in many fields of science and engineering, the difficulties associated with the unpredictability can be alleviated – to some extent – by adopting a probabilistic approach: indeed the unpredictability is confined to individual realizations of the stochastic phenomenon, whereas ensemble properties (statistics) are now predictable.

For the purpose of NPP safety, the results of a seismic hazard assessment are essentially pairs of data. Each pair consists of:

- One data item characterizing ground motion (such as the peak ground acceleration, or a response spectrum)
- One data item characterizing how rare it is to observe an event leading to such a ground motion (e.g., return period or annual probability of exceedance)

The latter data item is typically governing the specification of seismic loads. The annual probability of exceedance is prescribed at an

adequately low level, which varies from country to country; typical values of this probability are $10^{-4}/a$ and $10^{-5}/a$. The corresponding ground motion then represents the SSE.

The need to characterize events with extremely low probabilities of exceedance introduces an obvious data problem, as there cannot be – by definition – direct observations within the range of interest. Extensive efforts – which are far beyond the scope of this essay – have been ongoing for decades in order to predict extreme seismic events, mainly by modeling the processes from the source to the origin, i.e., the source mechanisms, the propagation (attenuation), and the site response. For an overview and further reading on seismic hazard refer, for example, to Chen and Scawthorne (2002). The level of uncertainty associated with these models – both due to limited observations for calibrating the model and due to the existence of concurrent models – remains very high.

This large degree of uncertainty associated with seismic hazard estimates is also one of the main reasons for the significant interest in the seismic robustness.

Two Approaches for Quantifying Seismic Robustness

There exist essentially two approaches for quantifying the robustness of NPPs to seismic-induced ground motions, namely, seismic margin assessment (SMA) and seismic probabilistic safety analysis (seismic PSA). It should be noted that in the United States, the term “probabilistic risk analysis” (PRA) is used instead of PSA.

In **SMA** the robustness is determined **explicitly** in terms of the margin between the design acceleration, used for determining the seismic loads on structures and equipment in the design calculations, and the plant capacity, i.e., the limit load that the plant can at most sustain and hence still be brought to a safe state and maintained in that state.

In contrast, in **seismic PSA** the margin is not quantified explicitly. Rather, for a comprehensive range of characteristic ground motions of increasing severity, the corresponding probability of

a core damage is evaluated. Core damage is mainly defined as the failure to meet the core cooling criteria and the associated overheating of the fuel rod cladding.

By taking into account the frequency of occurrence of the characteristic ground motions – expressed by the so-called hazard curve – and the corresponding core damage probabilities, one obtains a concise measure of the seismic risk in the form of the seismic-induced core damage frequency. In this quantity, the robustness is **implicitly** contained. Indeed, for a given hazard curve, a plant with a larger margin will have a smaller core damage frequency than a plant with a smaller margin.

Conversely, the opposite can be said about the methods as far as the notion of risk is concerned: in SMA the risk is contained implicitly, whereas the result of seismic PSA is an explicit expression of risk.

In both approaches – SMA and seismic PSA – the focus is on the **plant-level** robustness, i.e., it is essential that the safeguard systems of the plant be able to fulfill their safety function, even if individual equipment items may have failed. The analysis of the plant-level robustness takes into account the multiple lines of defense and the safety concepts – such as redundancy and diversity – implemented in the engineered safeguards. Ultimately, however, the plant-level robustness can be traced back to the seismic margin of **individual** systems, structures, and components (SSC). Hence, before discussing seismic robustness in terms of plant-level safety functions, the concept of seismic margin is introduced for individual SSC.

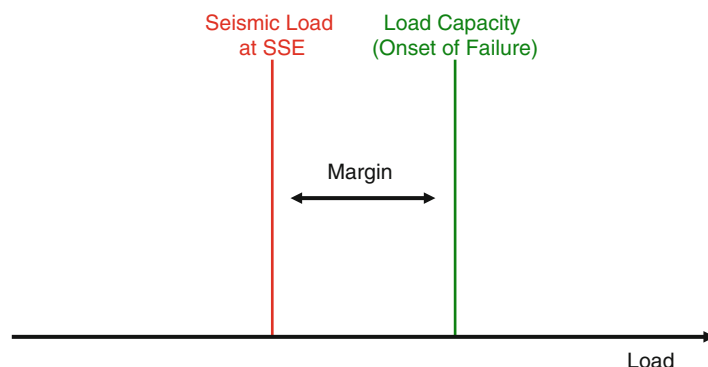
Seismic Margin of Systems, Structures, and Components

Load Versus Capacity

Considering the deterministic approach applied for the seismic design of nuclear facilities, a straightforward approach to quantify seismic margins would be to compare (for each important SSC) the load experienced by the SSC in case of

Seismic Robustness Analysis of Nuclear Power Plants,

Fig. 1 Margin definition in terms of seismic load



SSE with the load capacity, i.e., with the maximum load the SSC can still handle without failing to meet its requirements (stability, integrity, or operability) (Fig. 1).

However, a statement like “the mechanical analysis of the control rod drive mechanism reveals a seismic margin of x kN” would not be a very comprehensive expression of the existing margin, because it looks only at the quantities in the final step of the analysis chain that needs to be undergone in the seismic design of typical NPP equipment. Indeed, for the standard equipment item, the mechanical analysis is preceded – at least – by the analysis of the response of the floor on which the equipment is mounted.

The above statement is hence an incomplete estimate of the margin, because it neglects additional margins accumulated in the previous links of the analysis chain, such as the dynamic analysis of the civil structures.

Furthermore, it would not be a suitable basis for comparing the margins of different components relative to each other. For these reasons, it has proved useful to define the margin in terms of one characteristic ground motion parameter.

Expressing Margins in Terms of Ground Motion Parameters

Typically an SSE is defined by free-field ground motion response spectra. Seismic margin above the SSE is then most commonly quantified in terms of one parameter representative of the ground response spectra, the peak ground acceleration (PGA) being the most frequently used parameter. Alternative parameters are, e.g., the

peak spectral accelerations or the intensity on the Medvedev–Sponheuer–Karnik (MSK) scale. Frequently, the use of PGA as the representative parameter for expressing seismic margin and seismic fragility is criticized, on the grounds that it is a poor indicator of structural damage. Nevertheless, the PGA has gained the broadest acceptance as the characteristic margin parameter, probably due to its simple physical interpretation that can be grasped also by decision makers not familiar with more sophisticated notions of earthquake engineering or seismology (refer to IAEA 2003b, p. 6).

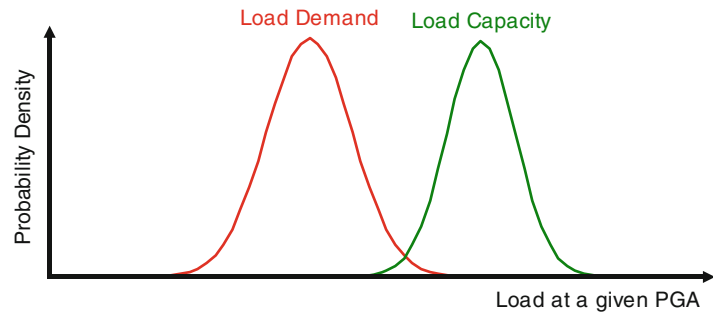
It should be noted that there are ways for compensating the limitation of PGA as a unique descriptor of a seismic event, either on the hazard definition – relying on notions such as the intensity or cumulative absolute velocity (CAV; refer to EPRI (2006), Campbell and Bozorgnia (2011)) in the hazard curve development – or on the structural response (→spectral shape factor in the fragility definition below).

Variability Due to Randomness and Uncertainty

The load to which an SSC is subjected during an earthquake of a certain PGA is clearly nondeterministic: indeed, there are an infinite number of possible ground motions showing the same PGA, each of which would induce different loads. This inherently aleatory characteristic of seismic ground motions is denoted as randomness.

Moreover, the majority of safety-related SSC is not directly subjected to ground motion but to vibrations induced through the plant’s civil

Seismic Robustness Analysis of Nuclear Power Plants,
Fig. 2 Nondeterministic character of load demand and capacity



structures. Thus, there is nondeterminism – both inherent (randomness) and epistemic (uncertainty) – associated to effects like soil-structure interaction, damping, load combination, etc. Finally, also the load capacity is nondeterministic, considering, e.g., that material properties cannot be fully controlled.

The nondeterministic character of both load demand and seismic capacity is typically modeled by random variables, as illustrated in Fig. 2.

Seismic failure will occur when the load demand D exceeds the capacity C ; the probability of failure P_F is then

$$P_F = P(D > C)$$

In view of the clearly nondeterministic nature of load demand and capacity, a conceptually sound definition of margin must be formulated in a probabilistic framework. A concept that has found widespread acceptance for this purpose is the definition of a HCLPF (pronounced as “hay-clipf”) capacity. It is an implicit definition: if a component has a HCLPF capacity of x , then the probability of failing due to a demand of value x is low (5 %), with a high confidence (95 %). For the reasons stated in the previous section, the parameter in terms of which the HCLPF capacity is typically defined is the PGA.

Having captured the nondeterministic nature of the quantities governing failure, by means of the HCLPF capacity, it is now straightforward to quantify margin, e.g., as the distance between the HCLPF capacity and the load demand prescribed in the design of the NPP (i.e., the SSE).

More details on the HCLPF concept, including the approaches for its quantification in practical problems, are presented in section “Sources of Seismic Margin” below.

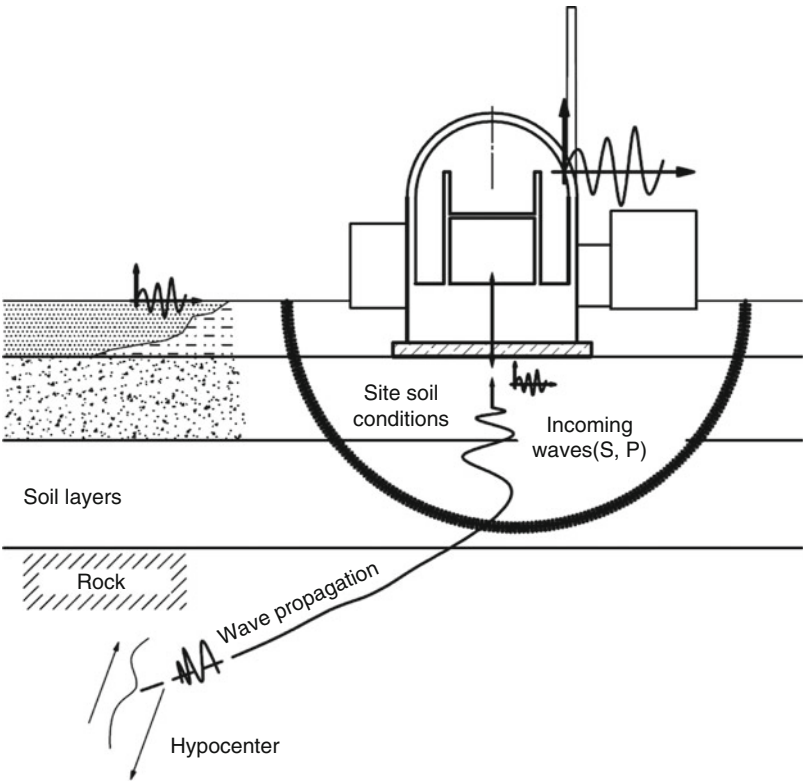
Sources of Seismic Margin

In principle seismic margin could be introduced on a deterministic basis specifying design basis loads exceeding those determined for the SSE or introducing penalizing stress utilization criteria. However, an overly conservative approach would induce significant extra cost in the entire life cycle of the plant (equipment design and procurement, civil construction, maintenance of heavy equipment items, decommissioning, etc.) and possibly detrimental effects on SSC performance in normal plant operation. In practice there are already various conservatisms in the seismic design of SSC. These conservatisms can be assessed and quantified when performing an SMA or an SPSA to show that adequate margin is available. Seismic margin especially results from the following sources:

- *Definition of the safe shutdown earthquake:*
 SSE ground motion response spectra are enveloping design spectra, i.e., the spectra have usually been subject of smoothing and broadening. In particular this introduces a substantial artificial increase of energy content of the excitation.
- *Conservatism in the determination of the seismic demand:*
 Earthquakes induce oscillations of a chain of different oscillators at a site: starting from

**Seismic Robustness
Analysis of Nuclear
Power Plants,**

Fig. 3 Earthquakes and their impact on civil structures and equipment (Translated from Sadegh-Azar and Hartmann 2011)



the soil, the base plate, and the upper floors of buildings up to smaller pieces of equipment installed in instrument racks or connected to flexible piping (Fig. 3). These different oscillating subsystems show more or less significant interaction effects; for example, the heavy weight of the reactor building has a damping effect on ground oscillations. Similarly heavy equipment installed inside the containment has a damping effect on the vibrations of the supporting floors. A realistic description of this complex oscillation behavior would therefore require modeling the entire system and considering also nonlinear effects such as plastic deformation.

Since such a global analysis is usually not practicable, the overall system is normally split into several subsystems analyzed separately. The following example illustrates this approach:

Subsystem	Excitation	Response
Ground + buildings	Ground response spectrum and corresponding time histories	Floor response spectra (secondary spectra) and corresponding time histories
Large equipment (e.g., primary circuit), civil substructures	Floor response spectra (secondary spectra) and corresponding time histories	Tertiary spectra and corresponding time histories
Equipment (e.g., pipes)	Tertiary spectra and corresponding time histories	Quaternary spectra and corresponding time histories
Built-in components (e.g., valve)	Quaternary spectra and corresponding time histories	Loads

Such an approach induces several sources of conservatism:

- Excitation characteristics and physical parameters considered for the different models are chosen conservatively to ensure that the conclusions made are valid without high uncertainty. The resulting level of conservatism may be moderate considering one model only. However, combining several models in a calculation chain, these conservatisms are multiplied and result typically in high factors of safety.
- Especially with respect to input characteristics, bounding assumptions are typically made. For example, many relevant codes require smoothing, broadening, or even increasing of input spectra. Similarly, the combination of load cases which can physically not occur simultaneously is often applied for the sake of simplification. Finally, the orthogonal vibration components are typically combined in a conservative manner.
- Interaction effects and in particular damping effects between the subsystems are neglected. The resulting conservatisms are especially important if the substructure is excited with a frequency near its resonance frequency, i.e., when the resulting vibrations are most important.
- Realistic anchorage characteristics allowing small displacements - significantly influencing the vibration behavior of the anchored system (shifting of natural frequencies) and thus avoiding resonant situations - are neglected.
- Other nonlinear effects such as plastic deformation and friction are typically also not modeled, and the corresponding energy dissipation is neglected or considered with simplified approaches only.
- *Conservatisms in the determination of the resistance to seismic loads:*

These conservatisms result especially from:

- Application of safety factors following the relevant codes.
- Use of conservative material properties (typically 95 % fractiles are used). In case of civil structures, the time-dependent concrete strengthening is normally also neglected.
- Neglecting plastic deformation capability.
- Finally, seismic margin results from stress utilization < 1 , i.e., if the resulting stresses are smaller than the code allowable stresses. This is particular true when seismic loads are not the governing loads for an SSC.

Quantification of Seismic Margin: Fragility Curves and HCLPF Values

Capacity and Fragility

In order to quantify the abovementioned sources of conservatism and derive an expression of the overall margin, the concept of seismic fragility as introduced by Kennedy and Ravindra (1984) has proven useful.

Seismic fragility is defined as a conditional failure probability, where the condition is represented by a seismic event with a given value of the characteristic ground motion parameter, denoted by a . In the sequel it will be assumed that this ground motion parameter is given by the peak ground acceleration, i.e., the zero period acceleration in the ground response spectrum. In this case the fragility is defined as follows:

$$Fr(a) = P[\text{failure} | PGA = a]$$

The definition of fragility is flexible in the sense that the meaning of the term *failure* is arbitrary. In the context of seismic PSA, it relates mainly to components of safety-relevant systems or to structures in which such components are installed; the entire set of these components and structures is abbreviated as SSC (systems, structures, and components). Quantitatively, failure

occurs if one of the parameters representing the loads experienced by the analyzed SSC exceeds its limit state (e.g., the tensile stress in a reinforcement bar exceeds the tensile strength of the utilized steel). The definition of the limit states for an SSC requires the identification of the relevant failure modes.

It is useful to express the fragility in terms of a random variable representing the capacity, denoted by A , and defined in terms of the same ground motion parameter used for the fragility definition. Obviously, failure occurs if the capacity is lower than the peak ground acceleration of the assumed seismic event. The definition of the fragility can then be extended as follows:

$$Fr(a) = P[\text{failure} | PGA = a] = P[A < a] \\ = F_A(a)$$

The right portion of the above equation emphasizes that the fragility is equivalent to the cumulative density function of the capacity A . This implies that defining the random capacity A is equivalent to defining the fragility.

Log-Normal Capacity Model

The most widely used model for the seismic capacity is given by

$$A = \tilde{A} \varepsilon_R \varepsilon_U$$

where $\tilde{A}$ is the median of the capacity, while ε_R and ε_U are log-normally distributed with unit median and logarithmic standard deviations of β_R and β_U , respectively. The random variables ε_R and ε_U model the variability due to randomness and due to uncertainty, respectively. In order to preserve the distinction between these two different sources of variability, the variables ε_R and ε_U are introduced in two distinct steps of the analysis. To begin with, it is assumed that only the variability to randomness exists, whereas there is no variability due to uncertainty, i.e., $\varepsilon_U = 1$.

At this stage, the fragility is then (due to the properties of the log-normal distribution of ε_R)

$$Fr(a) = \Phi \left(\frac{\ln \left(\frac{a}{\tilde{A}} \right)}{\beta_R} \right)$$

where Φ is the standard normal cumulative distribution function.

In the second step of the fragility analysis, the assumption that $\varepsilon_U = 1$ is relaxed. Instead, ε_U is assumed to be log-normally distributed with logarithmic standard deviation β_U .

Rewriting the capacity as

$$A = \tilde{A} \varepsilon_R \varepsilon_U = (\tilde{A} \varepsilon_U) \varepsilon_R$$

and replacing $\tilde{A}$ with $\tilde{A} \varepsilon_U$ in the cumulative probability density leads to

$$Fr(a) = \Phi \left(\frac{\ln \left(\frac{a}{\tilde{A} \varepsilon_U} \right)}{\beta_R} \right)$$

Due to the presence of ε_U , the fragility at a given PGA a itself turns also into a random variable. It can be shown that for a given probability level Q , the fractile of this random variable is

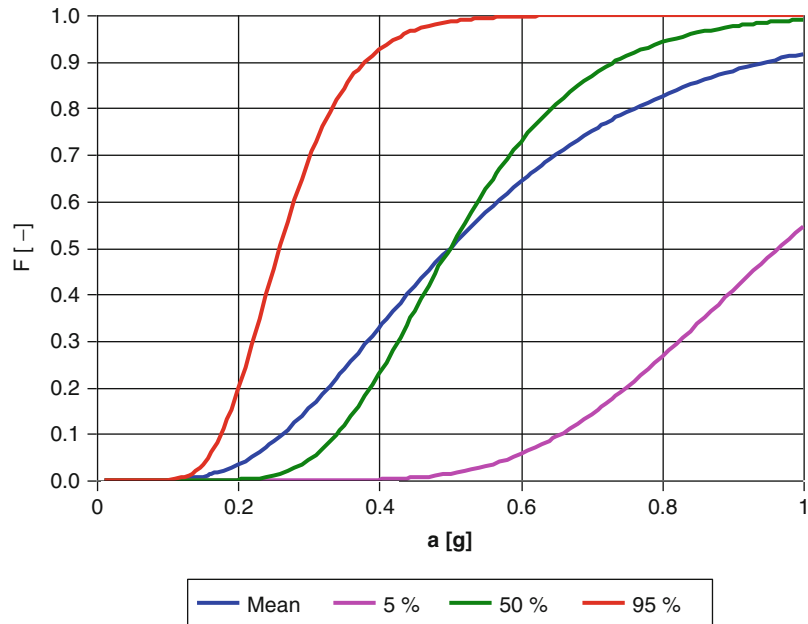
$$Fr(a, Q) = \Phi \left(\frac{\ln \left(\frac{a}{\tilde{A}} \right) + \beta_U \Phi^{-1}(Q)}{\beta_R} \right)$$

Interpreting Q as a subjective probability, or equivalently as a confidence level, one may use the above equation to obtain the fragility curves associated with specific confidence levels. This is demonstrated in Fig. 4 for the fragility model with the parameters $\tilde{A} = 0.5$ g, $\beta_R = 0.3$, and $\beta_U = 0.4$.

The median (50 % confidence level) fragility curve is identical to the one obtained by neglecting uncertainty (i.e., $\beta_U = 0$).

Seismic Robustness Analysis of Nuclear Power Plants,

Fig. 4 Example of fragility curve in linear scale



The mean fragility curve is obtained by averaging over the random fragility $F_A(a)$ for each value of the ground motion parameter a , i.e., the averaging is done along the vertical direction.

It can be shown that the mean fragility has the following simple form

$$F_A(a) = \Phi \left(\frac{\ln \left(\frac{a}{A} \right)}{\beta_C} \right) \quad \beta_C = \sqrt{\beta_R^2 + \beta_U^2}$$

The above model is called the **composite fragility model**, because it considers random variability and variability due to uncertainty in a compact form. It is used for point estimates (best estimate) of the seismic risk.

Methods for Deriving Fragility Curves of SSC

For the quantification of seismic fragility, one can generally distinguish between approaches relying on the assessment of conservatism in **existing** seismic verification studies and approaches requiring **new** dynamic analyses.

For cost reasons, where possible, preference is given to the former. Within this category, the most widely used approach for deriving fragility

curves is based on **separation of variables**, as described in detail in Kennedy and Ravindra (1984) and in EPRI (1994).

The basic concept underlying this approach is the following expression for the capacity A as a product of the (deterministic) value of the PGA, a_{SSE} , adopted in the seismic design of the SSC under consideration, and the scaling factor F :

$$A = a_{SSE} \cdot F$$

The scaling factor F may hence be viewed as the maximum scalar, by which the design ground motion can be multiplied ("scaled") without producing failure. Alternatively, the factor F may however also be viewed as a safety factor, considering that it is the ratio between the actual capacity A and the acceleration used in the design, a_{SSE} :

$$F = \frac{A}{a_{SSE}}$$

It should be noted that the scaling (safety) factor F is a random variable, since the capacity A is also a random variable.

The essence of the separation of variables approach is to break down the scaling (safety) factor F into a product of “partial” safety factors. This factorization process extends over two levels, since some of these partial safety factors are yet again factorized.

For structures, the factor of safety F can be modeled as the product of three random variables:

$$F = F_S F_\mu F_{RS}$$

where F_S is the strength factor, F_μ is the inelastic energy absorption factor, and F_{RS} is the structural response factor.

The strength factor, F_S , represents the ratio of ultimate strength (or strength at loss of function) to the stress calculated for a_{SSE} . In calculating the value of F_S , the non-seismic portion of the total load acting on the structure is subtracted from the strength as follows:

$$F_S = \frac{S - P_N}{P_T - P_N}$$

where S is the strength of the structural element for the specific failure mode, P_N is the normal operating load (i.e., dead load, operating temperature load, etc.), and P_T is the total load on the structure (i.e., sum of the seismic load – due to a_{SSE} – and the normal operating load).

The inelastic energy absorption factor (ductility factor), F_μ , accounts for the fact that an earthquake represents a limited energy source, and many structures or equipment items are capable of absorbing substantial amounts of energy beyond yield without loss of function.

The structural response factor, F_{RS} , is based on the recognition of the fact that in the design analyses, structural response is computed using specific (often conservative) deterministic response parameters for the structure. Since many of these parameters are nondeterministic (often with wide variability), the actual response may differ substantially from the calculated response for a given peak ground acceleration. F_{RS} is modeled as a product of factors influencing the response bias and variability:

$$F_{RS} = F_{SA} F_{GMI} F_\delta F_M F_{MC} F_{EC} F_{SSI}$$

where:

F_{SA} = spectral shape factor representing bias and variability in ground motion and associated ground response spectra

F_{GMI} = ground motion incoherence factor that accounts for the fact that a traveling seismic wave does not excite a large foundation uniformly

F_δ = damping factor representing bias and variability in response due to the difference between actual damping and damping assumed in design

F_M = modeling factor accounting for limitations regarding the resolution and calibration of the model adopted for representing the structural mechanics

F_{MC} = mode combination factor accounting for bias and variability in the estimated response due to the method used in combining dynamic modes of response

F_{EC} = earthquake component combination factor accounting for bias and variability in response due to the method used in combining earthquake components

F_{SSI} = factor that accounts for the effect of soil-structure interaction

Similarly, for equipment and other components, the overall factor of safety is composed of a capacity factor, F_C ; a structure response factor, F_{RS} ; and an equipment response (relative to the structure) factor, F_{RE} . Thus,

$$F = F_C F_{RE} F_{RS}$$

The individual safety factors F_i are assumed to follow a log-normal distribution. Their characteristic parameters are their median values F_i – representing the (often conservative) **bias** – and the logarithmic standard deviations $\beta_{i,R}$ and $\beta_{i,U}$, representing the **variability** due to aleatory and epistemic nondeterminism, respectively. The most widely used method for estimating these characteristic parameters is the approximate second moment procedure.

The overall median safety factor $\widetilde{F}$ may be found by scaling the ground response spectrum so long, until the limit state corresponding to the analyzed failure mode is reached. All input variables must be set to their median values in this process for the scaling factor to correspond to the median safety factor.

Alternatively, the sources of conservatism of the design calculations – compared to median-centered calculations – may be identified individually for each partial safety factor. For each of the response factors, the response quantity of interest (e.g., stress) is calculated (or estimated) anew, but only with the corresponding input variables set to their median values. The ratio of this newly computed response with the one from the design calculations gives the corresponding median partial safety factor. For the median capacity factor, the margin between the median strength and the stress obtained in the design calculation is quantified. The overall median safety factor is then simply given by the product of the individual safety factors.

In order to estimate the logarithmic standard deviations $\beta_{i,R}$ and $\beta_{i,U}$, it is necessary to perform an additional calculation of the safety factor, in which all model variables are median centered, with the exception of the input variables corresponding to the analyzed partial safety factor. The latter ones are instead perturbed by a multiple κ of their standard deviation. The input variables should be perturbed to the side which leads to a smaller safety factor (since we are more interested in quantifying the effect of variables in the more unfavorable cases). The resulting partial safety factor is denoted here as $F_{\kappa\sigma_i}$.

The variability parameter (either randomness or uncertainty) of the analyzed safety factor is then

$$\beta_i = \frac{1}{|\kappa|} \ln \left| \frac{F_{\kappa\sigma_i}}{\widetilde{F}} \right|$$

For several safety factors it is common practice to use generic variability parameters, as suggested in EPRI (1994).

Finally, the logarithmic standard deviations of the individual safety factors are combined using square root of sum of squares (SRSS), in order to obtain the corresponding variability parameter of the global safety factor, β_R and β_U , respectively.

Alternative, more refined methods can be used for deriving fragility curves, if necessary. For instance, the simplifying assumptions adopted in the separation of variables method can be relaxed by applying Monte Carlo simulation to the structural dynamic analysis. In this method, all parameters assumed to contribute significant nondeterminism to the overall capacity are sampled simultaneously, and the model used for structural dynamic analysis is perturbed correspondingly. Repeating this exercise for several levels of the PGA leads to a discrete set of points of the fragility curve, which can be interpolated or extrapolated as needed.

The HCLPF Capacity

As already indicated in earlier sections, the HCLPF value or HCLPF capacity is a crucial concept in the context of seismic robustness of NPPs. It is defined – implicitly – as the value of the PGA for which there is a *high confidence* (95 %) that the *probability of failure* does not exceed 5 % (i.e., it is *low*).

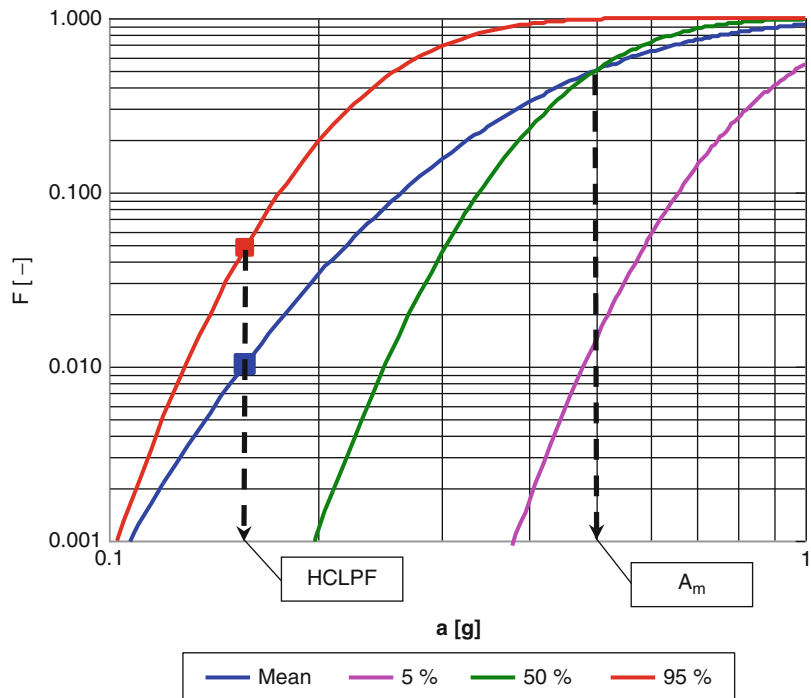
Based on the above definition, the HCLPF capacity can be related to the median capacity and the variability parameters as follows:

$$HCLPF = A_m \exp(-1.65(\beta_R + \beta_U))$$

where -1.65 is the 5 % fractile of the standard normal distribution.

The HCLPF capacity can be approximated by the 1 % fractile of the composite mean fragility curve. This definition of the HCLPF in the composite model is conservative, because it is guaranteed to be equal or smaller than the HCLPFs of all corresponding non-composite fragility models, i.e., those in which β_R and β_U are kept separately (for a specific β_C , there are infinitely many possible combinations of β_R and β_U).

Seismic Robustness Analysis of Nuclear Power Plants,
Fig. 5 Example of fragility curve in logarithmic scale



The HCLPF capacity has then the following simple form:

$$HCLPF = A_m \exp(-2.33\beta_C)$$

where -2.33 is the 1 % fractile of the standard normal distribution.

The above-described relationship of the HCLPF value with the fragility curve is exemplified in Fig. 5 for an SSC with a median capacity of 0.5 g and variability parameters β_R and β_U equal to 0.3 and 0.4, respectively.

Simplified HCLPF Estimation Using the CDFM Method

The previous section described the formal definition of the HCLPF capacity based on the corresponding fragility curve. It is clear that once a full-scope fragility analysis has been performed, the HCLPF capacity can also be easily obtained.

Within SMA it is, however, not necessary to have a full description of the fragility curve; the HCLPF capacity is sufficient. As a simplified alternative to estimate the HCLPF value,

a semi-probabilistic approach has been proposed. It is referred to as “conservative deterministic failure margin” (CDFM) method; compare EPRI (1991). The CDFM method can be briefly summarized by the following four steps:

1. For the targeted HCLPF capacity (e.g., the Review Level Earthquake in an SMA), the floor responses (demand) are computed using the 84 % fractile (i.e., an 84 % probability of non-exceedance) of the ground response spectrum. Damping is to be assumed conservatively.
2. The strength parameters are evaluated at the 2 % fractile, i.e., the probability of exceedance of the selected strength is at the 98 % probability level.
3. The inelastic energy absorption factor, F_μ , should be estimated conservatively, i.e., at a 84 % probability of exceedance level.
4. If, based on the above, the following inequality is satisfied, then the targeted HCLPF capacity has been demonstrated:

$$\text{Demand/Capacity} \leq F_\mu$$

Analysis of Seismic Robustness at Plant Level

Seismic Margin Assessment (SMA)

Goals and Approaches

The main goal of an SMA is to identify the seismic capacity of the plant, i.e., the maximum level of seismic ground motion for which the plant can still be safely shut down. In addition, it is the goal of any SMA to identify the weaker components, i.e., the components limiting the seismic capacity of the plant.

There are essentially two approaches that have gained widespread acceptance and that will be described in more detail in the sequel: a success-path-based approach (“EPRI method”) and a fault-tree-based approach (“NRC method”).

Stated in simple terms, the difference of the approaches is the following:

The success-path approach identifies the components **necessary to avoid** a core damage; the plant capacity is that of the **weakest** of these components.

The fault-tree approach identifies the combinations (“minimum cutsets,” abbreviated MCS) of component failures **sufficient to cause** a core damage. For each of these MCS, the capacity is that of the strongest of these components. The plant capacity is then equal to the minimum capacity of any of the MCS capacities.

Having stated the differences between the two approaches, this introductory section deserves to be concluded by noting a fundamental, unifying feature of the two approaches, namely, that they are essentially **deterministic**. This property becomes particularly obvious when comparing the plant-level HCLPF capacity with a **target** value of the PGA. For instance the US Nuclear Regulatory Commission requires 1.67 times the PGA of the design basis earthquake; refer to NUREG (1993). In the European Utility Requirements, a minimum capacity of 1.4 times the PGA of the SSE is recommended; refer to EUR organisation (2001). The result of this comparison is a **binary and hence deterministic expression**, indicating either sufficient margin or lack thereof, but no intermediate answers.

Success-Path-Based Approach

This approach is also referred to as the EPRI (Electric Power Research Institute)-approach, because it is best described in EPRI NP-6041.

The basic concept underlying this approach is that of a success path, indicating the transition from a disturbance to the normal plant operation – caused by a design-exceeding earthquake – to a safe state, in which the fundamental safety functions (reactivity control, residual heat removal, and radioactivity confinement) are durably ensured.

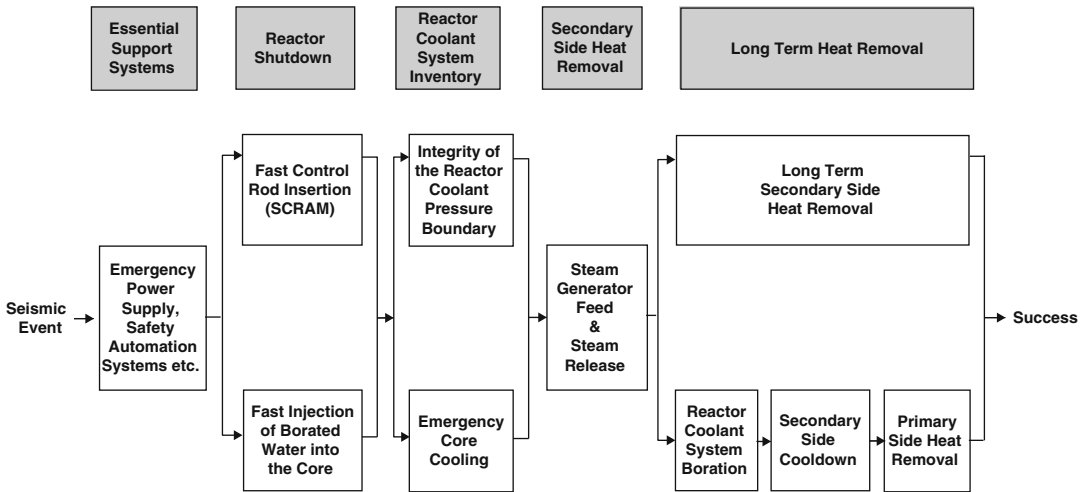
An example of a success path is shown for a pressurized water reactor (PWR) in Fig. 6.

It is important to note that a reasonable success-path definition should take into account the seismic design of the plants’ safety systems. The preferred success paths should be those along which the safety functions are fulfilled by seismically designed systems.

The success-path definition should make sure that the most likely scenarios of plant disturbance after a design-exceeding seismic event are covered. At least two success paths – leading to a durable safe state of the plant – need to be defined, and one of the two success paths should include the mitigation of a small break loss of coolant accident (LOCA), unless even a small break LOCA can be excluded, e.g., due to a particularly high quality of the supports of small-bore primary piping such as measurement lines.

The above mentioned Fig. 6 describes the success path at the level of the safety functions. The goal is to ensure that all the safety functions along the success path have a sufficient margin over the SSE. This requirement obviously needs to be cascaded down to the components making up the systems fulfilling the safety functions. For example, in order to ensure steam generator (SG) feed following the earthquake, it is necessary for the emergency feedwater pump to remain operational.

The requirement of sufficient margin does not only apply to the so-called “frontline” systems, i.e., the fluid systems specifically performing the respective safety function, but also to the so-called support systems, such as electrical power supply, heating/ventilation/air-conditioning (HVAC), and instrumentation and control (I&C).



Seismic Robustness Analysis of Nuclear Power Plants, Fig. 6 Example of a success path to a safe shutdown state for a pressurized water reactor

The result of cascading the requirement for sufficient margin to the individual components consists in a list of components which are needed to fulfill the safety functions credited in the success-path definition. This list is referred to as Seismic Equipment List (SEL) or also Safe Shutdown Equipment List (SSEL).

The metric used to measure the margin is the HCLPF value introduced in earlier chapters. Since the success-path-based approach is meant to be a predominantly deterministic approach, the CDFM method is preferred.

Considering that – as a rough estimate – the size of the SEL is in the order of 10^3 components, it should become clear that it is not feasible to perform dedicated HLCLF calculations for all components of the SEL. Instead, a walkdown-based **screening process** is to be implemented. Since this screening process is common to all approaches for seismic robustness and risk assessment, it is discussed in the final chapter of this essay on practical issues.

A critical step preceding the screening process is the definition of the so-called Review Level Earthquake (RLE). It consists in a site-specific characterization of a seismic ground motion (typically a response spectrum) that is more severe than the ground motion considered in the design of the plant (i.e., the SSE).

The result of the screening process is a, typically large, set of components which are judged – based on a structured and standardized walkdown procedure – to have a generic HCLPF capacity **equal or larger** than the RLE. For the remaining components, individual HCLPF values need to be derived.

The fact that the walkdown-based screening provides only a lower bound of the **generic** HCLPF capacity, instead of a sharp value, implies that the RLE should be set high enough, so that some components for which a non-generic HCLPF is evaluated are below the RLE. Only in this way, a sharp plant HCLPF can be determined, whereas otherwise one obtains only a lower bound, because the generic HCLPFs can only be stated as being equal or higher than the RLE.

For those components for which an individual HCLPF is to be evaluated, the responses due to the RLE are needed. These may be either obtained by performing new floor response spectra and equipment response calculations based on the RLE or by scaling existing responses to the ground motion of the SSE.

Once both generic and individual HCLPF capacities have been evaluated, the overall capacity of the plant results as the lowest HCLPF of the considered equipment. This implies that both success paths must be sufficiently robust.

As a concluding remark, the deterministic nature of the success-path-based approach is emphasized, both at the individual component level (CDFM used for HCLPF estimation) and at the system modeling level (no systematic generation of minimum cutsets; see the following section).

Fault-Tree-Based Approach

This approach is also referred to as the NRC approach, because it has been proposed and developed for the NRC in NUREG (1985, 1986).

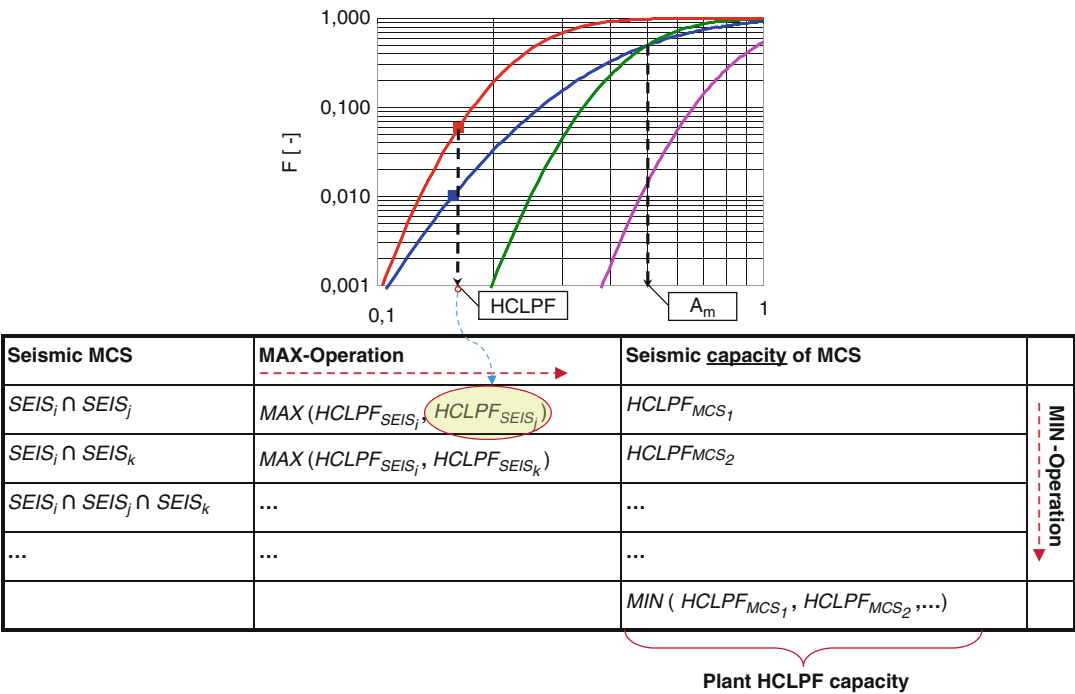
In the fault-tree-based approach, several aspects are unchanged with respect to the success-path-based approach, such as the walkdown-based screening process. The main difference compared to the success-path-based approach is that a fault-tree-based system analysis is relied upon in order to determine **minimum cutsets**, i.e., combinations of failures leading to core damage. The attribute “minimum” indicates that no individual failure event can be removed from the set without altering its property of leading to core damage. The **minimum** cutsets have

the highest risk relevance, because their probability is higher than non-minimum cutsets.

The components screened out on the basis of the walkdown observations are not included in the system modeling, i.e., only components which cannot be generically stated as having a HCLPF capacity equal or higher than the RLE are modeled in the fault trees. Individual HCLPF capacities of these components are to be estimated, either based on a fragility analysis or at least based on the CDFM method.

Performing a cutset analysis for the event trees judged as relevant for the SMA (typically at least a seismic-induced transient, such as loss of off-site power, and a small break LOCA) results in a list of minimum cutsets to be further processed as follows.

For each of the cutsets, a characteristic HCLPF capacity is obtained by taking the **maximum** HCLPF capacities of the component failure forming the cutset. From the resulting list of cutset HCLPFs, the **minimum** is then taken, in order to obtain the fault-tree-based plant HCLPF capacity. Fig. 7 is meant to illustrate the process.



Seismic Robustness Analysis of Nuclear Power Plants, Fig. 7 Min-Max operation in fault-tree-based SMA

Individual seismic-induced failure events are denoted by $SEIS_i$, $SEIS_j$, $SEIS_k$, etc.

If the fault-tree-based plant HCLPF capacity is higher than the RLE, then the plant HCLPF can only be judged as being at least equal to the RLE, since the components screened out on the basis of the walkdown observations were not included in the system modeling.

If a PSA model is available, then the effort for performing the system modeling is greatly reduced, since the seismic-induced failures of the non-screened components merely need to be added to the existing fault trees. A major benefit of reusing an existing PSA model is also the possibility to analyze mixed cutsets, i.e., combinations of seismic-induced and non-seismic-induced failures, such as random component failures or human (operator) errors.

As a final remark it is emphasized that – even though the probabilities of the cutsets are not used in the quantification of the plant HCLPF capacity – the fault-tree-based approach to SMA resembles a **probabilistic** approach, because cutsets are generated, i.e., a systematic exploration of the combinations of individual component failures is performed. Nevertheless, the subsequent steps of the procedure and – in particular – the final result are **deterministic**, since a single number (the HCLPF capacity) is used to characterize the capacity of each individual SSC.

Seismic PSA

Basic Notions

While sharing numerous aspects with SMA, seismic probabilistic safety analysis (PSA) is distinct from its deterministic counterpart (SMA) in several respects. The most prominent difference is – of course – due to the attribute “probabilistic,” which implies nondeterminacy for all the possible states of the system (i.e., the model of the NPP). Indeed, for a seismic event of a given PGA, a component may fail or may not fail, and both possibilities are assigned a **probability**. This probability is called fragility and has been introduced at length in section “[Quantification of Seismic Margin: Fragility Curves and HCLPF Values](#).” The same property of nondeterminacy

applies to the event “seismic-induced core damage,” which may or may not occur in case of a seismic event of a given PGA.

The event “seismic-induced core damage” is a central notion of both seismic PSA and SMA and hence a unifying aspect. Clearly, since both methodologies (seismic PSA and SMA) aim at analyzing the plant-level resistance to seismic events, one is – in general – not concerned with individual failures, but only with combinations of failure leading to core damage.

However, there is a significant difference: in the seismic PSA, it is assumed that the event “seismic-induced core damage” can – at least theoretically – occur due to seismic events of **any** intensity (and hence PGA).

This is where the concept of seismic hazard at the NPP site comes into the picture: it is the characterization of the seismic ground motions “to be expected” at the site (see section “[Motivation for Evaluating Seismic Robustness](#)”). For the purpose of seismic PSA, the most important characterization is afforded by the so-called hazard curve, in which the probability of observing a seismic event exceeding a PGA equal to a during a predefined time interval (typically one year) is expressed as a function of a . Thus, the hazard curve is a monotonously decreasing function $H(a)$.

In Fig. 8, a set of hazard curves is shown, each curve being associated with a probability level (50 %, 5 %, and 95 %). This characterization of the site hazard in terms of a population of hazard curves, rather than a single hazard curve, reflects the uncertainties in the hazard estimation procedures.

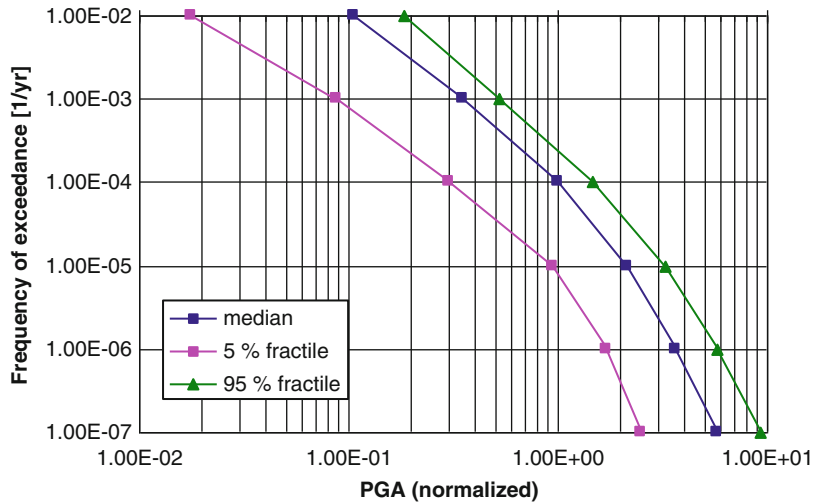
Recall that for each level a of the PGA, core damage may or may not occur and denote the corresponding probability by $F_A^{CDF}(a)$. The probability that a seismic event leading to core damage occurs during the time interval to which the hazard curve refers (typically one year) can be shown to be:

$$CDF = \int_{a=0}^{\infty} F_A^{CDF}(a) \frac{dH}{da} da$$

The acronym CDF denotes core damage frequency, which is – in quantitative terms – almost

Seismic Robustness Analysis of Nuclear Power Plants,

Fig. 8 Hazard curves associated with different confidence levels



identical to the annual probability of core damage.

In the practical implementation of seismic PSA, the evaluation of the CDF is however more involved. The reason is that the conditional probability of core damage $F_A^{CDF}(a)$ cannot be obtained in a closed form, because of the enormous number of component failure combinations potentially leading to core damage.

This is where the specific application, seismic PSA, becomes reminiscent of its more general “parent,” namely, level 1 PSA for internal events. In level 1 PSA risk-relevant accident progression sequences are modeled by event trees, and the potential combinations of component failures leading to unavailability of safety systems called upon during the accident progression sequences are modeled by fault trees.

These event trees and fault trees represent a (very complex) Boolean expression of the event “core damage.”

A well-established procedure for evaluating the probability of this event is to express the (complex) Boolean expression as a union of (much simpler) intersections of component failures; these intersections are called “minimum cutsets” (MCS).

For each minimum cutset, MCS_i , the contribution to the core damage frequency is

$$CDF^{(i)} = \int_{a=0}^{\infty} F_A^{(i)}(a) \frac{dH}{da} da$$

where $F_A^{(i)}(a)$ is the probability of the MCS_i , conditional on a given PGA a , and $H(a)$ is the frequency of occurrence of a seismic event with a PGA equal or higher than a .

The probability of the MCS_i is

$$F_A^{(i)}(a) = \prod_{j=1}^{n^{(i)}} f_j^{(i)}(a) \prod_{k=1}^{m^{(i)}} p_k^{(i)}$$

where $f_j^{(i)}(a)$ is the j th fragility appearing in MCS_i and $p_k^{(i)}$ is the failure probability of the k th non-seismic basic event in MCS_i . The integers $n^{(i)}$ and $m^{(i)}$ are the number of fragilities and the number of non-seismic basic events in MCS_i .

In practice the above integral is approximated by

$$CDF^{(i)} \approx \sum_{l=1}^n f\left(a_{l-\frac{1}{2}}\right)(-\Delta H_l)$$

where a_0 and a_n are the limits of the risk-relevant PGA range and the remaining a_i 's are selected at convenient points, e.g., the evaluations of the inverse hazard function at the values

10^{-1} , 10^{-2} , etc. The corresponding mid-interval fragilities $f(a_{l-\frac{1}{2}})$ are logarithmically interpolated,

$$f(a_{l-\frac{1}{2}}) = 10^{(\log_{10} f(a_{l-1}) + \log_{10} f(a_l))/2}$$

The probability of occurrence associated with each interval is $\Delta H_l = H(a_l) - H(a_{l-1})$. The total CDF is then obtained by evaluating the union of the individual MCS.

Correlation Between Seismic Failures

The issue of mutual dependence (or alternatively, correlation) between seismic-induced failures relates to the question whether the seismic-induced failure of a component, e.g., an emergency diesel generator (EDG), is more probable, under the condition that another component (e.g., the EDG of another safety train) has experienced seismic-induced failure. Several studies indicate that in seismic PSA, the effect of correlations on the seismic-induced core damage frequency (CDF) may be very significant; refer to Pellissetti and Klapp (2011) and to references 1 through 4 therein.

The most straightforward and – in general – most conservative approach consists in assuming full correlation between identical and “seismically similar” components of different, redundant safety trains, i.e., components connected by “AND”-logic. Therefore, with this model the redundancy represented by the different trains of the plant is not credited, as far as seismic-induced failures are concerned. Furthermore, for components of a single train combined under the same “OR” gate in the fault tree, the failures are assumed to be independent.

In this approach failures are thus assumed to be either fully correlated or totally uncorrelated; it is hence a binary model (black/white, zero/one) which cannot accommodate gradually varying levels of correlation. In fact, for redundant components connected by “AND”-logic, this model negates the existence of distinct failure events for the individual redundant trains and will hence also be referred to as “deterministic correlation model.” This binary, deterministic model is

conservative with respect to the probability of the top event, because correlation is detrimental (i.e., the failure probability increases) for intersection of events and beneficial for unions of events (Fig. 9).

An alternative approach, described in NUREG (1990), consists in modeling the dependencies between seismic failures in terms of the correlation coefficients between the variables characterizing the seismic demand and capacity of the SSCs. The adopted model makes full use of the possible (positive) range of correlation coefficients (between 0 and 1).

The separation of the correlation coefficients of demand and capacity reflects the difference in their respective cause: the correlation of the demand variables is due to “similarity” of the excitation experienced by distinct components (e.g., due to spatial proximity), whereas the correlation of the capacity variables is due to “similarity” in terms of the physical properties of the components (e.g., due to the fact that they have been manufactured by the same supplier). The word “similarity” is placed between quotes because it is to be understood in the statistical sense.

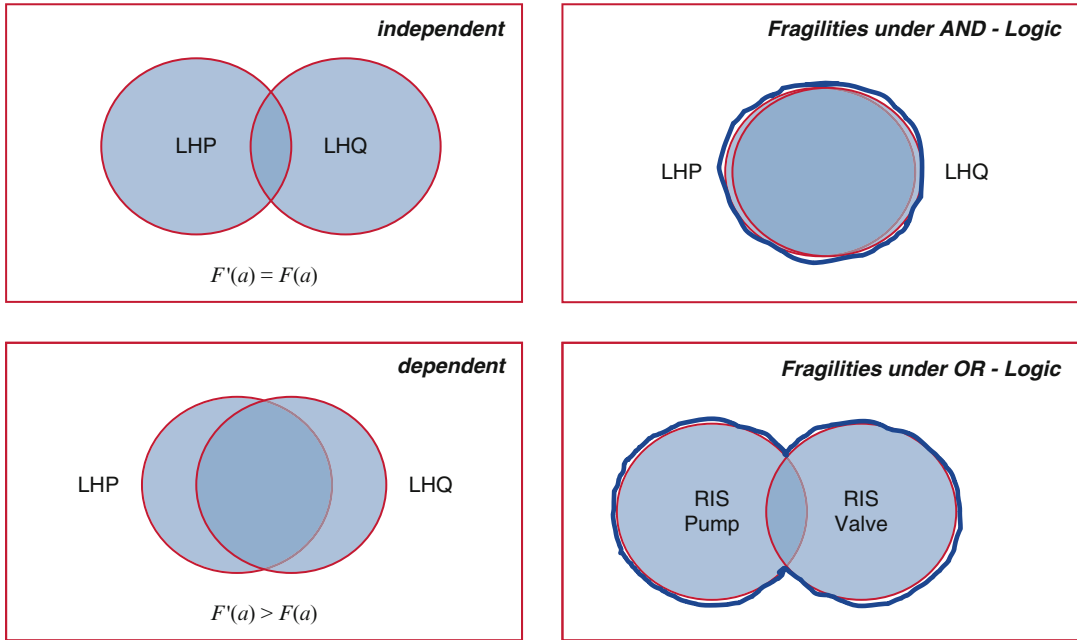
In this approach failure is defined in terms of the following limit-state function:

$$X = \ln\left(\frac{R}{S}\right)$$

where R and S represent the response and the strength for a given failure criterion. Failure takes place if

$$R > S \quad \Leftrightarrow \quad X > 0$$

It is convenient to model both R and S by log-normal random variables with the parameters μ_R , β_R and μ_S , β_S . According to the definition of log-normal variables, the respective companion normal variables are then $C_R = \ln R$ (with mean μ_R and standard deviation β_R) and $C_S = \ln S$ (with mean μ_S and standard deviation β_S). In view of its definition, X is then normally distributed and its mean and standard deviation are



Seismic Robustness Analysis of Nuclear Power Plants, Fig. 9 Schematic representation (Venn diagrams) of dependence between individual seismic-induced failures

$$\mu_X = \mu_R - \mu_S, \quad \beta_X = \sqrt{\beta_R^2 + \beta_S^2}$$

The probability of failure is then given by

$$P[X > 0] = 1 - \Phi\left(\frac{-\mu_X}{\beta_X}\right) = \Phi\left(\frac{\mu_X}{\beta_X}\right)$$

where Φ is the cumulative probability density function of the standard normal distribution. Considering two individual components, one may then define the random variables R_1, S_1 and R_2, S_2 , respectively, and X_1, X_2 according to the above definition of X . Assuming no cross-correlation between R_i and S_j , after some arithmetics, the correlation coefficient for X_1, X_2 may be expressed as

$$\rho_{X_1 X_2} = \frac{1}{\beta_{X_1} \beta_{X_2}} (\beta_{R_1} \beta_{R_2} \rho_{R_1 R_2} + \beta_{S_1} \beta_{S_2} \rho_{S_1 S_2})$$

$$\text{where } \beta_{X_i} = \sqrt{\beta_{R_i}^2 + \beta_{S_i}^2}.$$

It should be noted that in the above expression, the correlation coefficients $\rho_{R_1 R_2}$ and $\rho_{S_1 S_2}$ relate to

the companion normal variables of R_i and S_j , i.e., $\ln(R_i)$ and $\ln(S_j)$. The advantage of calibrating the correlation model in terms of the companion normal variables is that the meaningful range of the correlation coefficient is always $[-1, 1]$. This is not the case for log-normal random variables. It should be noted that according to its definition, the correlation coefficient $\rho_{X_1 X_2}$ depends on both the correlation coefficients and the standard deviations; this is necessary to prevent a pair of variables with a high correlation but with very small standard deviation from introducing a high overall correlation. With the above definitions, the joint probability of failure of the two components $Q_{1 \cap 2}$ is defined in terms of the following two-dimensional integral:

$$\begin{aligned} Q_{1 \cap 2} &= P[(X_1 > 0) \wedge (X_2 > 0)] \\ &= \int \int_{0 \leq x_1, x_2 \leq \infty} f_{X_1 X_2}(x_1, x_2) dx_1 dx_2 \end{aligned}$$

In the above equation, $f_{X_1 X_2}$ is the bivariate normal probability density function (PDF):

$$f_{X_1 X_2}(x_1, x_2) = \frac{1}{2\pi\beta_{X_1}\beta_{X_2}\sqrt{1-\rho_{X_1 X_2}^2}} \cdot \exp\left(-\frac{1}{2(1-\rho_{X_1 X_2}^2)}\left(\frac{(x_1-\mu_{X_1})^2}{\beta_{X_1}^2} + \frac{(x_2-\mu_{X_2})^2}{\beta_{X_2}^2} - 2\rho_{X_1 X_2}\frac{(x_1-\mu_{X_1})(x_2-\mu_{X_2})}{\beta_{X_1}\beta_{X_2}}\right)\right)$$

Accordingly, the probability of simultaneous failure of n components has the form

$$\begin{aligned} Q_{1 \cap \dots \cap n} &= P[(X_1 > 0) \wedge \dots \wedge (X_n > 0)] \\ &= \int_{0 \leq x_1, \dots, x_n \leq \infty} f_X(x) dx \end{aligned}$$

The application of the above correlation model in the context of fault-tree-based seismic PSA – using general purpose PSA software – is exemplified in Pellissetti and Klapp (2011).

Operator Actions

Earthquakes not only may impact the reliability of safety systems but also influence operator performance:

- The workload for the shift team may be higher due to a failure of operational systems.
- Spurious signals, e.g., caused by relay chatter, may make appropriate decision-making more difficult.
- Accessibility of buildings may be degraded (e.g., due to a failure of lighting or automatic locking systems) and may hence lead to smaller grace times for the necessary actions.
- Finally, earthquakes may have a direct impact on human behavior (shocks).

Past seismic PSA experiences have shown that the choice of probabilities for seismic-induced operator actions can have a significant impact on the overall PSA result. Thus, a short overview of seismic human reliability models discussed in the literature is presented in the following.

According to the **shock model**, it is conservatively assumed that all human actions fail if the peak ground acceleration exceeds a certain limit. For instance, Swiss regulations define this limit at

two times the PGA of the DBE; refer to Klügel (2007).

In the **time-dependent model** described in NUREG (1987), the seismic-induced human error probability (HEP) is expressed as a function of the time t between the earthquake and the action:

$$HEP(t) = HEP_{t=1 \text{ min}} \cdot t^{-a}$$

In the **ramp model** presented by Yokobayashi et al. (1998), the operator performance is modeled by a linear relationship between the seismic-induced HEP and the PGA until a maximum ground acceleration level L is reached.

Finally, it is worth mentioning a **model based on the CAV** (cumulative absolute velocity) value presented by Klügel (2007). No adverse effect on operator performance is expected if the CAV is below 0.16 gs. In case of stronger earthquakes,

$$HEP = HEP_{Baseline} + HEP_{GF, direct} + HEP_{Shock}$$

$HEP_{Baseline}$ is the HEP used in the PSA for internal events. $HEP_{GF, direct}$ describes structural failure of equipment (e.g., instruments, main control room boards) as well as accessibility restrictions leading to a guaranteed failure (GF) of the considered action. Finally, HEP_{Shock} models a failure of all operator actions due to operator shock, depending on the CAV: shock-induced failure is assumed with probability one if the CAV is greater than 0.16 gs and strong-motion duration exceeds a threshold value; otherwise, the probability of shock-induced failure is zero.

Uncertainty Analysis

In the earlier sections on seismic hazard analysis and – in greater detail – on fragility analysis, two sources of nondeterminism have been introduced

for the parameters relevant in a seismic PSA, i.e., the frequency of occurrence of a specific PGA (hazard) and the capacity of a component or structure (again in terms of the PGA).

These two sources of nondeterminism are random variability and (epistemic) uncertainty.

Random variability is the nondeterminism due to the inherently unpredictable nature of earthquakes and, consequently, of the earthquake-induced vibrations of the ground, the NPP civil structures, and the NPP equipment. The degree of scatter induced by random variability is termed as irreducible, since additional collection of data can improve its statistical representation, but not the scatter itself (expressed for instance by the variance or standard deviation).

Uncertainty, on the other hand, denotes the nondeterminism due to limited accuracy of models used to characterize the structural response. This lack of accuracy can be caused by inherent limitations of the model when representing the phenomena it is meant to capture (including issues such as model resolution). Also, it can be caused by limited data availability for model calibration. Another type of nondeterminism described by the term “uncertainty” is the subjective variability associated with the process of selecting a model from a set of concurrent models involving different experts. This source of uncertainty is particularly pronounced in probabilistic seismic hazard analysis.

The uncertainty in the seismic hazard is typically characterized by deriving different hazard curves for different confidence levels. In the following figure this is shown in the upper left subfigure; in this case the confidence levels are expressed by multiples of the standard deviation. For a defined value of the PGA, the uncertainty is schematically highlighted by the blue, double-sided vertical arrow.

As discussed in section “[Log-Normal Capacity Model](#),” the uncertainty associated with the seismic capacity of SSC is characterized by the uncertainty variability parameter β_U . Similarly as for the hazard curve, the uncertainty expressed by β_U can be shown graphically in terms of the fragility curves associated with the various confidence levels, as shown in the lower left

subfigure below. In this subfigure, too, for a defined value of the PGA the uncertainty in the fragility is schematically highlighted by a blue, double-sided vertical arrow.

Finally, in the right portion of the figure, it is indicated how the uncertainties in the seismic hazard and in the fragility propagate – across the convolution integral introduced previously – to the core damage frequency.

The result of this uncertainty analysis – or uncertainty propagation – is then a distribution function of the core damage frequency, as shown in the lower right subfigure. The uncertainty in the core damage frequency is here emphasized by a blue, double-sided *horizontal* arrow (Fig. 10).

Practical Issues

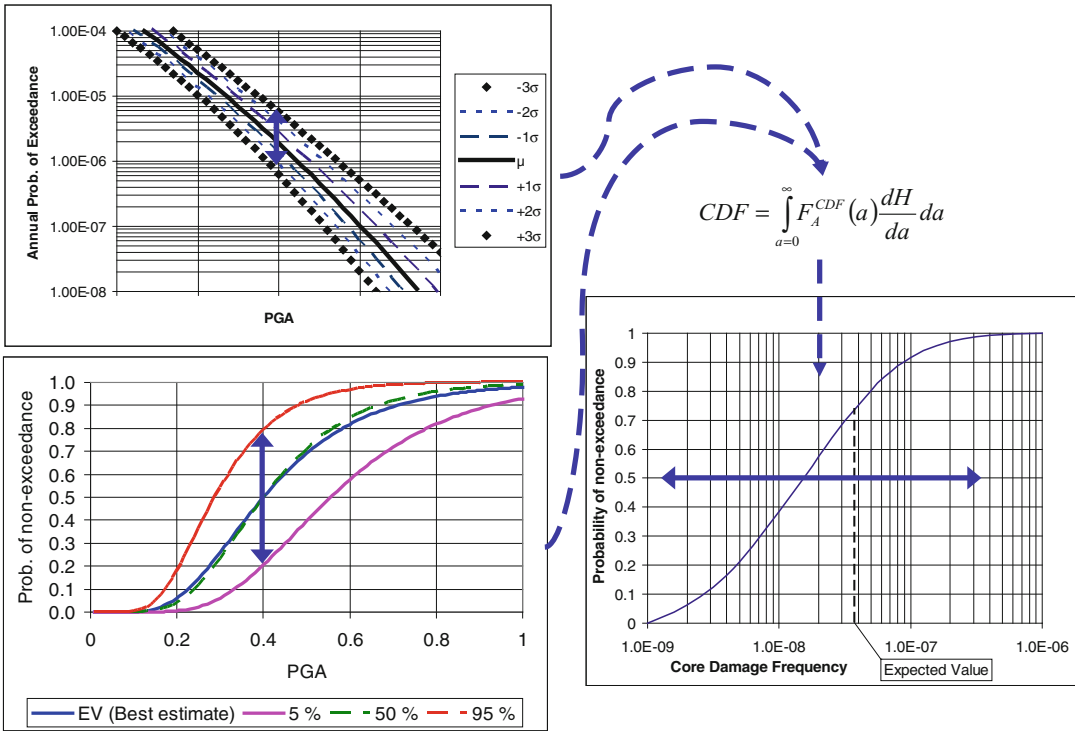
Seismic Equipment List

The assembly of the so-called Seismic Equipment List (SEL) is a major task in both SMA and seismic PSA. To begin with, this list includes all those systems, structures, and components (SSC) which contribute to bringing the plant to a safe shutdown state after a seismic event (this is why in EPRI (1991), this list is referred to as Safe Shutdown Equipment List (SSEL)).

For success-path-based SMA, EPRI (1991) provides detailed guidance on how to assemble the SEL. For fault-tree-based SMA and seismic PSA, it is a common situation that a PSA model already exists for the level 1 PSA of internal events. The basic events of this level 1 PSA model represent a good starting point for the SEL in this case; however, the list is to be extended by including passive components (e.g., tanks) and structures, which are typically not modeled in level 1 PSA for internal events.

Walkdown-Assisted Use of Generic Data and Screening

The number of components in the initial SEL (in the order of magnitude of 10^3) prevents a consistent degree of detail to be applied in the analysis of all of them. It is hence a common feature of SMA and seismic PSA that the initial SEL has to undergo a more or less extensive



Seismic Robustness Analysis of Nuclear Power Plants, Fig. 10 Schematic illustration of uncertainty quantification in seismic PSA

screening process in order to reduce the number of equipment items on the list. The main purpose of doing so is to enable the analysts to focus the majority of their subsequent efforts on equipment with relatively low capacity and consequently with the highest risk relevance.

The screening process takes its legitimacy from extensive earthquake experience accumulated over the years. A crucial ingredient of the process is the execution of a seismic walkdown during which it is verified, whether the equipment in the experience database is representative of the components encountered in the plant under investigation.

Summary

Seismic robustness of nuclear power plants (NPP) denotes the ability of the plant to withstand severe seismic ground motion – possibly even beyond the seismic loads against which the

plant is designed – without experiencing the degradation of equipment to a point that the plant cannot be brought and maintained in a safe shut-down state.

By necessity, seismic robustness assessment focuses on the systems, structures, and components (SSC) that are essential to the goal of a safe shutdown in the wake of an extreme seismic event. For these SSC the actual capacity – or a lower bound thereof – is evaluated; in order to ensure a common measure of this capacity for all SSC, it is quantified in terms of a characteristic ground motion parameter, typically the peak ground acceleration.

Based on the above, evaluating the seismic robustness requires both system analysis – accounting for the response of a plant to a seismic-induced accident from a process engineering point of view – and structural analysis of the essential SSC.

There are two approaches that have been found to be viable for fulfilling the above

requirements, namely, seismic margin assessment (SMA) and seismic probabilistic safety analysis (seismic PSA; in the United States, the term seismic PRA is preferred, with “risk” replacing “safety”).

The main difference between the two methods is that SMA is essentially a deterministic approach, leading to an estimate of the seismic robustness of the plant in terms of a single number, namely, the plant-level capacity.

In contrast, seismic PSA is an explicitly probabilistic approach, analyzing those combinations of component failures that prevent the plant from being transferred to a safe shutdown state, thus resulting in a damage to the reactor core. Compounding the probabilities of these combinations with the frequency of occurrence of seismic events of various intensity results in an estimate of the seismic-induced core damage frequency. A small value of the latter indicates a large seismic robustness of the plant.

Cross-References

- [Analytic Fragility and Limit States \[P\(EDPI IM\)\]: Nonlinear Dynamic Procedures](#)
- [Code-Based Design: Seismic Isolation of Buildings](#)
- [Seismic Risk Assessment, Cascading Effects](#)
- [Site Response for Seismic Hazard Assessment](#)
- [Structural Reliability Estimation for Seismic Loading](#)

References

- American Society of Civil Engineers (ASCE) (2000) Seismic analysis of safety-related nuclear structures and commentary. ASCE 4-98, Reston
- Campbell KW, Bozorgnia Y (2011) Evaluation of the use of cumulative absolute velocity (CAV) in the shutdown of nuclear power plants after an earthquake. In: Transactions, SMiRT 21, New Delhi
- Chen W-F, Scawthorn C (eds) (2002) Earthquake engineering handbook. CRC Press, Boca Raton
- Electric Power Research Institute (EPRI) (1991) A methodology for assessment of nuclear power plant seismic margin (revision 1). Report EPRI NP-6041-SL, Palo Alto
- Electric Power Research Institute (EPRI) (1994) Methodology for developing seismic fragilities. Report TR-103959, Palo Alto
- Electric Power Research Institute (EPRI), U.S. Department of Energy (2006) Program on technology innovation: use of cumulative absolute velocity (CAV) in determining effects of small magnitude earthquakes on seismic hazard analyses. Report 1014099, Palo Alto/Germantown
- EUR Organisation (2001) European utility requirements for LWR nuclear power plants
- International Atomic Agency (IAEA) (2003a) Seismic design and qualification for nuclear power plants. Safety guide NS-G-16, Vienna
- International Atomic Agency (IAEA) (2003b) Seismic evaluation of existing nuclear power plants. Safety series 28, Vienna
- International Atomic Agency (IAEA) (2009) Evaluation of seismic safety for existing nuclear installations. Safety guide no. NS-G-2.13, Vienna
- Kennedy RP, Ravindra MK (1984) Seismic fragilities for nuclear power plant risk studies. Nucl Eng Des 79:47–68
- Klügel J (2007) An improved methodology for the evaluation of human error probabilities in a seismic PSA. In: Transactions, SMiRT 19, Toronto
- Pellisetti MF, Klapp U (2011) Integration of correlation models for seismic failures into fault tree based seismic PSA. In: Transactions, SMiRT 21, New Delhi
- Sadegh-Azar H, Hartmann H-G (2011) Grundlagen der seismischen Auslegung von Kernkraftwerken und Einfluss der Boden-Bauwerk Wechselwirkung. Bauingenieur (D-A-CH-Mitteilungsblatt)
- U.S. Nuclear Regulatory Commission (NUREG) (1985) An approach to the quantification of seismic margins in nuclear power plants. NUREG/CR-4334, UCID-20444, Washington, DC
- U.S. Nuclear Regulatory Commission (NUREG). (1986). *Recommendations to the Nuclear regulatory commission on trial guidelines for seismic margin reviews of nuclear power Plants*. NUREG/CR-4482, UCID-20579, Washington, DC
- U.S. Nuclear Regulatory Commission (NUREG) (1987) Relay Chatter and operator response after a large earthquake. NUREG/CR-4910, Washington, DC
- U.S. Nuclear Regulatory Commission (NUREG) (1990) Procedures for the external event core damage frequency analyses for NUREG-1150. NUREG/CR-4840, SAND88-3102, Washington, DC
- U.S. Nuclear Regulatory Commission (NUREG). (1993). Subject: policy, technical, and licensing issues pertaining to Evolutionary and Advanced Light-Water Reactor (ALWR) designs. SECY 93-087, Washington, DC
- Yokobayashi M et al (1998) Consideration of the effect of human error in a seismic PSA. In: Proceedings of the 4th international conference on probabilistic safety assessment and management, Berlin

Seismic Sources from Landslides and Glaciers

Ewald Brückl¹, Daniel Binder² and Stefan Mertl³

¹Department of Geodesy and Geoinformation,
TU Wien, Vienna, Austria

²Zentralanstalt für Meteorologie und
Geodynamik, Vienna, Austria

³Mertl Research GmbH, Vienna, Austria

Synonyms

Earthquakes generated by landslides and glaciers; Seismic activity of landslides and glaciers

Introduction

With few exceptions, the discovery and analysis of global earthquakes with landslides and glaciers as sources did not begin until the twenty-first century. This may come as a surprise considering that the magnitudes of such earthquakes have reached and even exceeded $M = 5$. In this essay, we describe the specific characteristics of these seismic signals and their source locations. We describe the source mechanisms and discuss the information that can be derived from interpretation of the data. In the case of landslides, the source process represents a severe hazard, and the same may be true for glaciers.

Although seismic signals from landslides and glaciers are observed at intermediate scales, we will jump from magnitudes of $M \sim 5$ to $M < 0$. The deployment of seismic stations on the target landslide or glacier is necessary to observe such weak signals. Systematic research on these microearthquakes also did not start until the twenty-first century. We offer an overview on landslide and glacier processes relevant to the generation of microearthquakes. Using selected case studies, we present the diversity of seismic signals generated by landslide and glacier sources and emphasize the close relations to geodetic and hydrometeorological data. We relate the seismic data to landslide and glacier

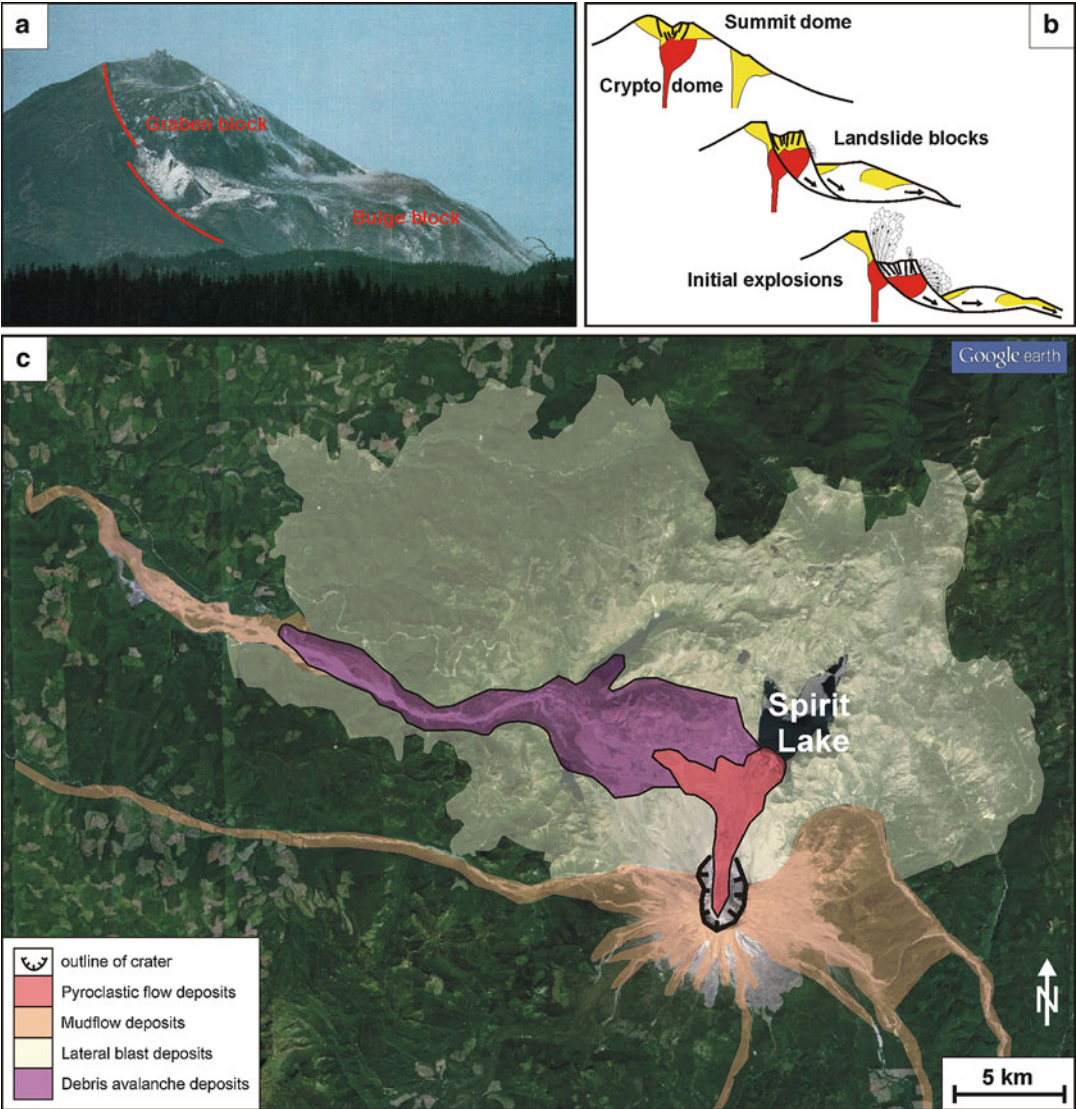
processes as far as possible at the current state of knowledge. The potential usage of this seismic monitoring data for hazard estimation is also addressed.

A New Class of Earthquakes

Seismic Signals from the 1980 Mount St. Helens Eruption and Landslide

Mount St. Helens (46.2°N , 122.2°W) has been the most active volcano in the Cascade Range during the Holocene. Its most impressive active phase in historical times started with a series of small earthquakes in March 1980. During the first half of the following May, seismic activity increased dramatically, and a bulge over 100 m grew outward at the mountain's north flank. The cataclysmic Plinian-style eruption, VEI 5, occurred on May 18, 1980. Within 15–20 s of a magnitude 5.1 earthquake, the volcano's bulge and summit slid away in a huge landslide (Fig. 1a). Volcanic explosions interacted with the landslide process (Fig. 1b). The landslide, comprising a volume of 2.8 km^3 , rushed first down the slope to the north and bent later to the west-northwest along the North Fork Toutle River with a speed briefly exceeding 200 km/h. The landslide covered an area of nearly 60 km^2 , with the most distal deposits traveling about 25 km (Fig. 1c). The height of Mount St. Helens was reduced from about 2950 to 2550 m (Brantley and Myers 2000).

Seismic signals from the May 18, 1980, eruption and landslide of Mount St. Helens have been recorded by seismic stations at epicentral distances from 23° up to 133° (Fig. 2a). The seismograms show strong surface waves in the period range of 100–260 s corresponding to $M_s = 5.9$. The spectra fall off very rapidly at periods shorter than 75 s. This makes these recordings significantly different from typical tectonic earthquake spectra (Fig. 2b). A cutoff period of ~ 5 s would be expected for a tectonic earthquake of similar magnitude. The low-frequency content of the Mount St. Helens recordings suggests a very slow source process. Kanamori and Given (1982) analyzed the



Seismic Sources from Landslides and Glaciers, Fig. 1 The Mount St. Helens eruption and landslide 18 May 1980; (a) the bulge at the north flank; (b) development of the landslide during eruption; (c) landslide

deposit area (Modified after Voight 1981; Moore and Albee 1981, <http://pubs.usgs.gov/gip/msh/mudflows.html>)

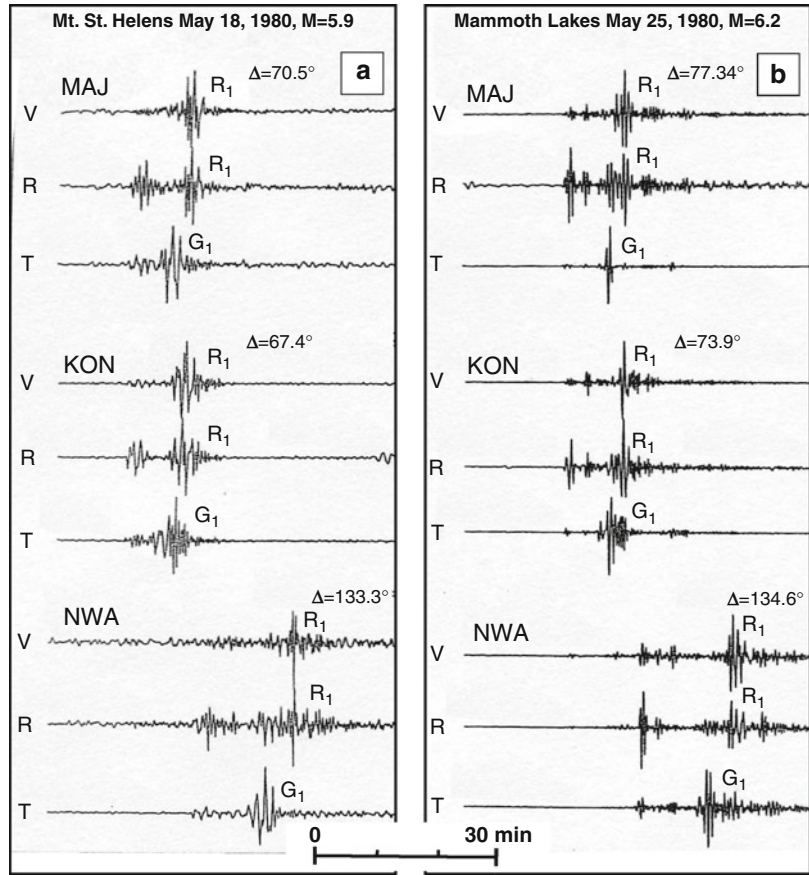
radiation pattern of the Love and Rayleigh waves and concluded that a subhorizontal, about southward-directed single force is an adequate source mechanism and describes the radiation pattern much better than a conventional DC. The earthquake attributed to the Mount St. Helens landslide resembled the first example of a new class of global earthquakes.

The Single-Force Seismic Source

Earthquakes of the new class produce long-period surface waves detectable on a global scale with wavelengths at least ten times the extent of the source area. The spatial migration of the source during the event cannot be resolved by long period data, therefore we consider a single force $F(t)$ varying in time but fixed in space as

Seismic Sources from Landslides and Glaciers,

Fig. 2 Recordings from seismic observatories in Japan (*MAJ*), Norway (*KON*), and Australia (*NWA*), 30 s high-cut filtered of (a) Mount St. Helens eruption and landslide, (b) Mammoth Lakes earthquake (Modified after Kanamori and Given 1982); V, R, and T are the vertical, radial and transverse components, R1 and G1 mark the first Rayleigh and Love phases



the seismic source. We refer to $\mathbf{F}(t)$ as force history. $\mathbf{F}(t)$ is the reaction force to the acceleration or deceleration $\mathbf{b}(t)$ times a mass $\mathbf{M}$ on the Earth's surface (Eq. 1).

$$\mathbf{F}(t) = -\mathbf{b}(t)\mathbf{M} \quad (1)$$

The acceleration is caused by the drop of the frictional force at a sliding plane from static equilibrium to its value during sliding. The excess gravitational forces result in the acceleration. When the slope of the landslide trajectory reduces, frictional forces exceed the driving gravitational forces, and the landslide decelerates. The momentum $\mathbf{I}(t)$ and velocity of the landslide $\mathbf{v}(t)$ follow from integration of Eq. 1 over time (Eq. 2).

$$\mathbf{I}(t) = \mathbf{v}(t)\mathbf{M} = \int_0^t \mathbf{F}(s)ds \quad (2)$$

We have $\mathbf{I}(T) = \mathbf{v}(T) = 0$, where T is the duration of the landslide. This condition imposes a constraint on the derivation of the force history $\mathbf{F}(t)$. Kanamori and Given (1982) found that the data of the Mount St. Helens eruption can be fitted by a sinusoidal $\mathbf{F}(t)$ with a period of 240 s.

CSF: Centroid Single Force

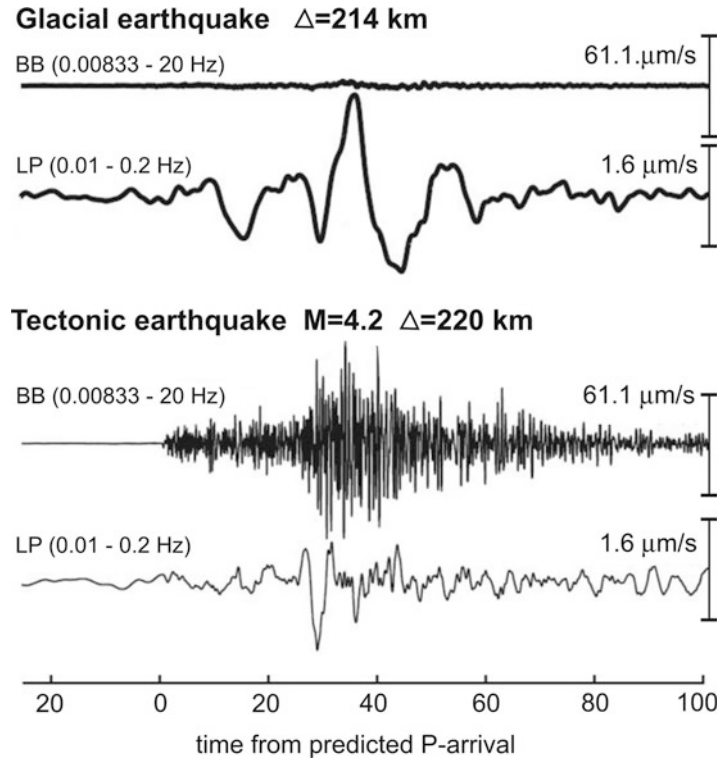
The displacement multiplied by the landslide mass at time t follows from an integration of $\mathbf{I}(t)$ (Eq. 2). The total displacement vector $\mathbf{D}$ times $\mathbf{M}$ is given by Eq. 3.

$$\mathbf{D}\mathbf{M} = \int_0^T \mathbf{I}(t)dt \quad (3)$$

The product of $\mathbf{D}$ and $\mathbf{M}$ can be derived from seismic data. Satellite images or other topographic information may also be used to break

Seismic Sources from Landslides and Glaciers,

Fig. 3 Recordings of the Alaska glacial earthquake and a M 4.2 earthquake (Modified after Ekström et al. 2003)



down the product into its factors **D** and **M**. Equations 1, 2, and 3 are valid for any surface mass accelerated or decelerated by either gravitational, inertial, or frictional forces, stress relief, or volcanic processes. Kawakatsu (1989) showed that **D M** represents the spatiotemporal centroid single force vector $\text{CSF} = -\text{D M}$, the analogue of the CMT for moment tensor seismic sources. He also provided the formalism for normal-mode inversion.

Global Glacial Earthquakes

A search for earthquakes of the new class in Arctic and Antarctic areas in data provided by global networks was carried out by Ekström et al. (2003) for the period 1999–2001 and extended to the period 1993–2008 by Tsai and Ekström (2007) and Nettles and Ekström (2010). About 243 earthquakes were found in Greenland, 14 in Antarctica, and 1 in Alaska. The one in Alaska (Fig. 3) exemplarily shows the difference

of the new-class earthquake to a nearby tectonic earthquake of similar magnitude and may explain why the new class of earthquakes has so long been undiscovered. CSF inversion was applied to the events in Greenland and Antarctica by Tsai and Ekström (2007) and Nettles and Ekström (2010), resulting in an improved location with an uncertainty of approximately 20 km. These events were clustered at or near eight fast-moving outlet glaciers with speeds >800 m/year (Fig. 4). An antisymmetric boxcar function (25 s of constant acceleration before the centroid focal time followed by 25 s of constant deceleration) was assumed in order to derive reactive forces transmitted to the Earth's crust by the displacement of a hypothetical mass. The magnitudes of centroid single force vectors were in the range $10 \cdot 10^{12} \leq \text{CSF} \leq 200 \cdot 10^{12}$ m kg, and the magnitudes were $4.6 \leq M_s \leq 5.1$. Most CSF vectors point in the opposite direction of the general flow of the glaciers.

The frequency of glacial earthquakes in Greenland is subject to seasonal variations and

Seismic Sources from Landslides and Glaciers,
Fig. 4 Glacial earthquakes in Greenland (Modified after Tsai and Ekström (2007) and Nettles and Ekström (2010))



synchronous with maximum glacier flow velocities in July–August and a corresponding minimum in January and February (Nettles et al. 2008). A power law does not properly describe the magnitude/CSF – frequency relation. It seems that each outlet glacier has its typical earthquake, which depends on glacier size, hydrological conditions, and the calving rate (Tsai and Ekström 2007).

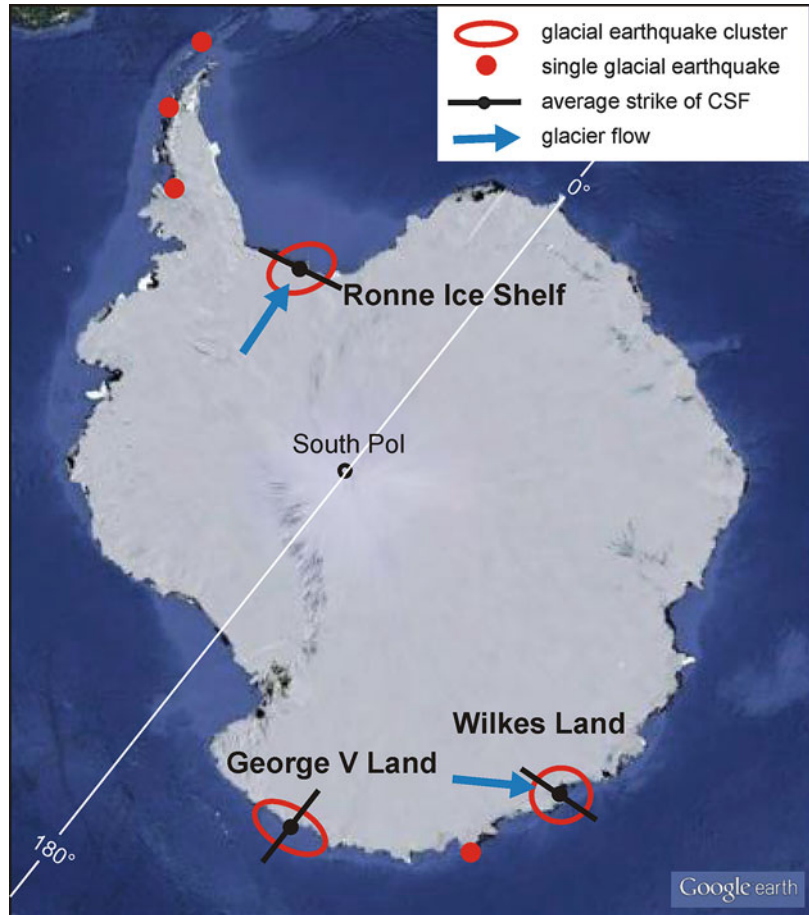
Chen et al. (2011) detected 13 more new-class glacial earthquakes in Antarctica with epicenters at or near the Ronne Ice Shelf, Ninnis Glacier, and Vanderford Glacier (Fig. 5). At the Ronne Ice Shelf, the direction of the CSF vector was normal to the ice flow direction and parallel to the ice

front; force direction was parallel to the rift propagation direction. At Vanderford and Ninnis glaciers, CSF and local ice flow directions coincided with only one exception. The magnitudes of the glacier earthquakes in Antarctica were $4.2 \leq M_s \leq 4.9$.

Global Landslide Earthquakes

Ekström and Stark (2013) applied the detection and location method developed for the glacial earthquakes to known large landslides or areas where a potential for giant landslides can be assumed. The landslide force history $\mathbf{F}(\mathbf{t})$ was

Seismic Sources from Landslides and Glaciers,
Fig. 5 Glacial earthquakes in Antarctica (Modified after Nettles and Ekström (2010) and Chen et al. (2011))



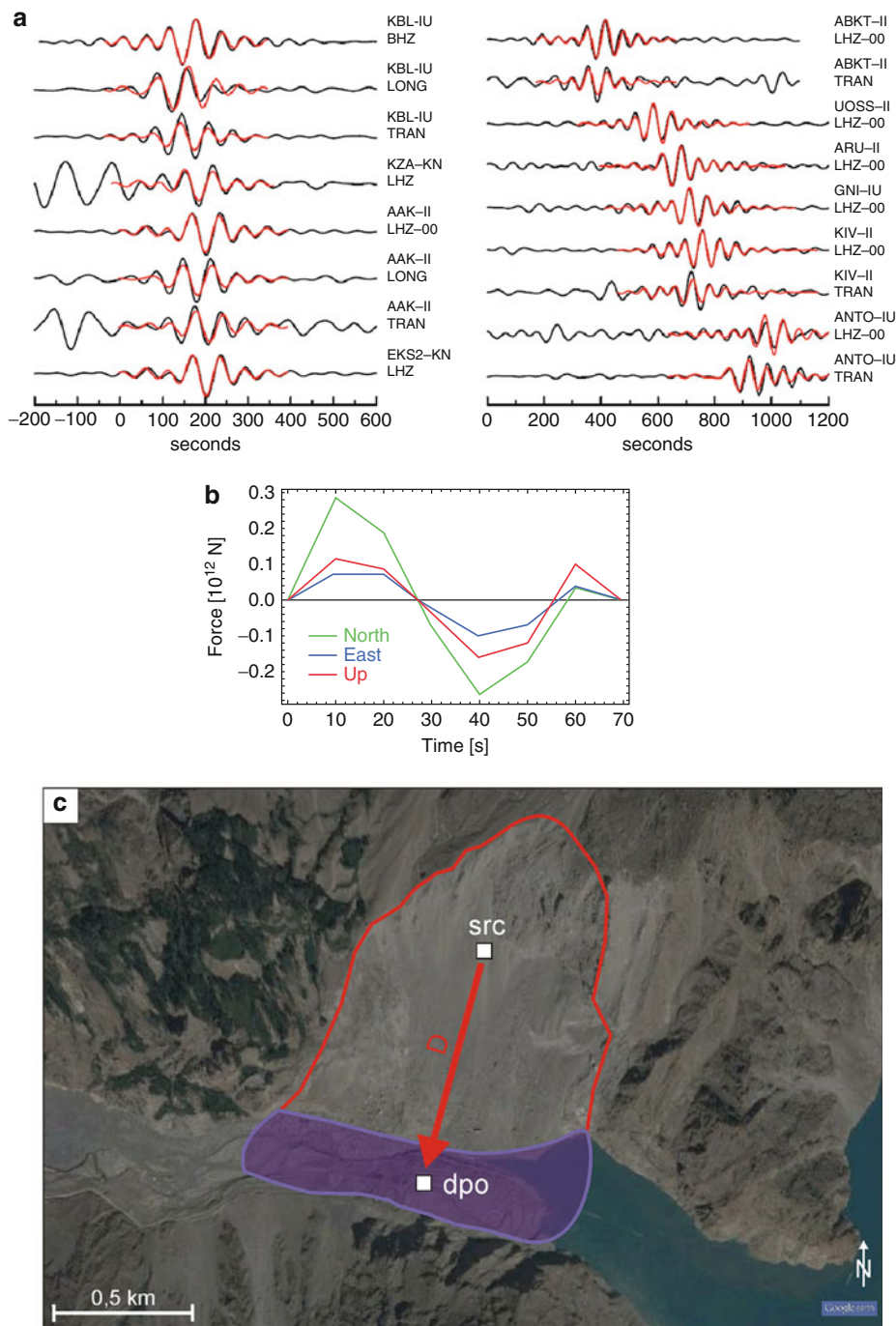
discretized into a sequence of overlapping isosceles triangles with a half-width of 10–15 s. Momentum, displacement, and CSF were derived from $\mathbf{F}(t)$. Satellite images and other topographic information were used to separate the factors D and M out of the product CSF. Figure 6 shows, as an example, the Hunza-Attabad landslide in North Pakistan with a total mass of $140 \cdot 10^9$ kg. Ekström and Stark (2013) analyzed seismic recordings of 29 giant landslides in the Aleutian and Coast Ranges, the Rocky Mountains, the Andes, the Alpine – Himalayan orogeny, Taiwan, New Guinea, and Antarctica, with the Mount St. Helens 1980 landslide included. The magnitudes of these events range from $M_s = 4.6$ – 5.6 . The CSF can be calculated from the data provided by Ekström and Stark (2013). CSF or $\log_{10}(\text{CSF})$ significantly

correlates with the landslide mass M and the magnitude M_s (Fig. 7).

Mechanisms of Giant-Single Force Sources

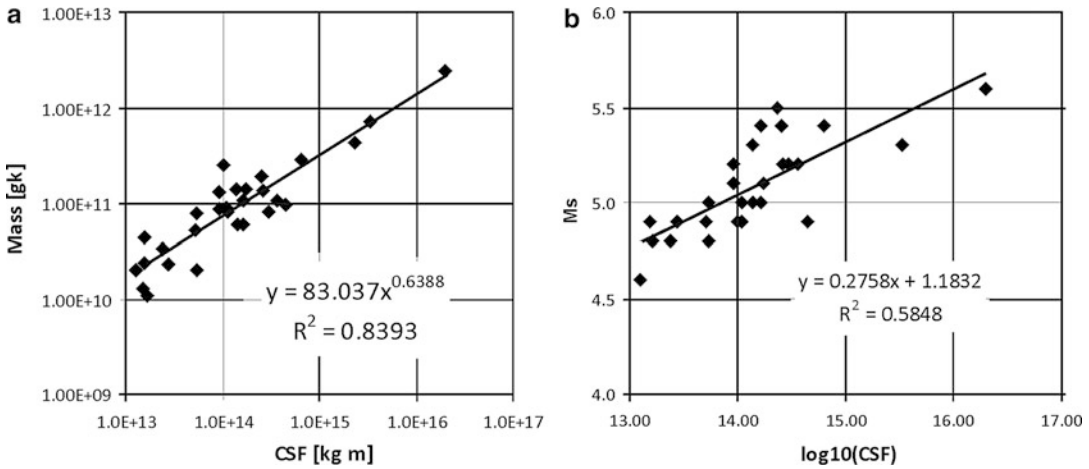
Landslide: Rotating Slider Block Model

Planar slider block models have been proven to support understanding of earthquake processes considerably. When considering the application of a planar slider block to earthquake processes, the slider block is loaded by tectonic strain, and after strain release the slider block comes to rest. For landslides, the situation is different; on an inclined planar sliding plane the slider block accelerates constantly and the travel distance is not confined. A rotational slider block overcomes

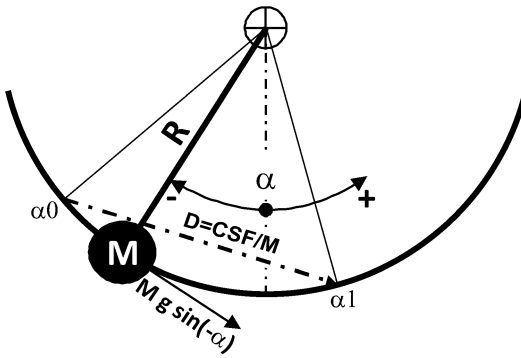


Seismic Sources from Landslides and Glaciers, Fig. 6 Hunza-Attabad landslide 04 January 2010 (Modified after Ekström and Stark 2013); (a) observed and synthetic traces; (b) landslide force history; (c) source (red outline) and deposition area (blue) of the landslide;

src and dpo are centroids of the landslide mass M before and after the slide, D the trajectory derived from satellite image; given $CSF = M D = 14 \cdot 10^{13}$ kg m the landslide mass is $M \sim 14 \cdot 10^{10}$ kg; note the damming of the Hunza river



Seismic Sources from Landslides and Glaciers, Fig. 7 Relations between landslide data (Ekström and Stark 2013); (a) Mass – CSF, (b) M_s – $\log_{10}(\text{CSF})$



Seismic Sources from Landslides and Glaciers, Fig. 8 Rotational slider block model (for description of symbols see section “Landslide: Rotating Slider Block Model”)

this problem and confines the travel distance (e.g., Brückl and Parotidis 2005).

Figure 8 shows the geometry of a rotational slider block. The dynamic system of a landslide is approximated by a mathematical pendulum with the length R and the mass M concentrated to a point. The position of the mass is defined by the angle α times R . Before the landslide happens at $\alpha = \alpha_0$, the driving gravitational force $Mg \sin(-\alpha)$ is in equilibrium with the frictional force $\mu_0 Mg \cos(\alpha)$, which always acts in opposition to the direction of movement (Eq. 4). A decrease of the friction coefficient from μ_0 to μ initiates the landslide. The excess driving force is

compensated for by the inertial force $M \frac{d^2\alpha}{dt^2} R$ (Eq. 5). We only look at one half oscillation of the slider block.

$$Mg \sin(-\alpha_0) - Mg \cos(\alpha_0)\mu_0 = 0 \quad (4)$$

$$\begin{aligned} Mg \sin(-\alpha) - Mg(\cos \alpha)\mu \\ = -MR \frac{d^2\alpha}{dt^2} \end{aligned} \quad (5)$$

We consider a first-order approximation for small amplitudes ($\sin(\alpha) \sim \alpha$, $\cos(\alpha) \sim 1$) and travel distance $D \sim R(\alpha - \alpha_0)$. Introduction of these approximations and subtraction of Eq. 4 from Eq. 5 gives

$$(R/g)d^2\alpha/dt^2 + (\alpha - \alpha_0) = \mu_0 - \mu = \Delta\mu \quad (6)$$

The solution of the inhomogeneous differential equation for the initial conditions $\alpha = \alpha_0$ and $d/dt(\alpha) = 0$ is given by Eqs. 7 and 8.

$$\alpha(t) = \alpha_0 + \Delta\mu(1 - \cos(\omega_0 t)) \quad (7)$$

$$\omega_0 = 2\pi/T_0 = \sqrt{g/R} \quad (8)$$

The position angles α of the slider block at start time t_0 , stop time $t_1 = T_0/2$, and the time of maximum velocity $t_m = T_0/4$ are α_0 , $\alpha_1 = \alpha_0 + 2\Delta\mu$, and $\alpha_m = (\alpha_0 + \alpha_1)/2 = \mu = -\text{atan}(\text{CSF}_Z/\text{CSF}_h)$. The travel distance D and

maximum magnitude of the landslide force (positive for $t = t_0$, negative for $t = t_1$) are given by Eqs. 9 and 10.

$$D = \text{CSF}/M = R(\alpha_1 - \alpha_0) = 2 R \Delta\mu$$

$$= 2 \Delta\mu g (T_0/2\pi)^2 \quad (9)$$

$$F_{\max} = M g \Delta\mu = g \text{CSF}/(2R)$$

$$= \pi^2 \text{CSF} / \left(2(T_0/2)^2 \right) \quad (10)$$

The half-period $T_0/2$ of the rotational slider block is the duration of the slide or the landslide force history T . We introduce a calibration factor $k < 1$ (Eq. 11) to account for the observation of the landslide force history $F(t)$ not starting and ending with the extreme value predicted by the slider block model.

$$T_0/2 = k T \quad (11)$$

A calibration factor of $k = 0.68$ brings $F_{\max}$, calculated in Eq. 10, into a nearly perfect 1:1 relation with the $F_{\max}$ derived from the inverted force–time function (Ekström and Stark 2013; Fig. 9a). Figure 9b shows the dependence of α_m and α_0 on CSF. Over the range of CSF (approximately three decades), the average sliding angle reduces from $\sim 25^\circ$ to $\sim 10^\circ$. There is also a systematic increase of ratio $(\alpha_0 - \alpha_m)$ to α_0 with CSF from $\sim 28\%$ to $\sim 55\%$.

Calving Glacier: Toppling of an Iceberg

The large outlet glaciers of Antarctica and Greenland lose mass mainly by calving. One of the best-monitored sites where calving takes place is the front of the Jakobshavn Glacier near Ilulissat, West Greenland. On 28 May 2008, a gigantic sequence of calving events was filmed by a team of Chasing Ice (<http://earthsky.org/earth/video-largest-glacier-calving-ever-caught-on-film>). An area of the size of Manhattan calved successively over a time span of 75 min. One can observe a great number of slabs of about 1000 m in height toppling from unstable to stable equilibrium, thus pushing the iceberg mélange of

previously calved icebergs away from the ice front out to sea. This toppling of large icebergs has been considered a candidate for a seismic source of global scale (Tsai et al. 2008).

Figure 10 illustrates the geometry, kinematics, and acting forces during the toppling of an iceberg, leaving it in unstable equilibrium after calving from the front of an ice stream. The calving front may or may not be grounded. The side lengths of the cross section are H (the height before toppling) and L . The width of the iceberg is W . With $H > L$, the iceberg is initially in an unstable position. Toppling of the iceberg implies not only a rotation of $\sim 90^\circ$ but also acceleration off the glacier front. This inertial force causes a reaction force at the contact area of the iceberg with the glacier, which is then transmitted via the glacier ice to the glacier bed and represents the main part of the glacial seismic source. A first-order approximation of CSF:

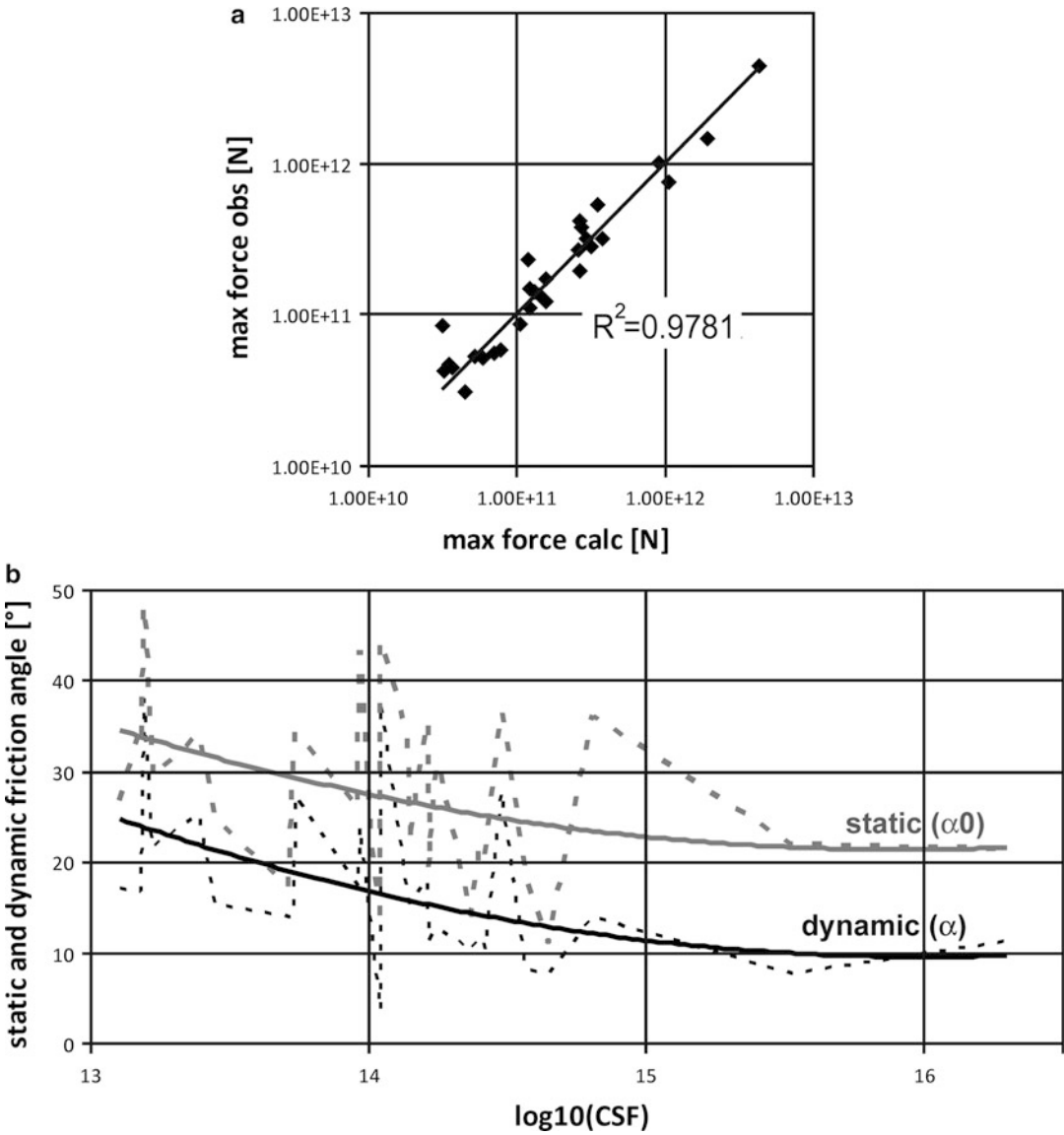
$$\text{CSF} \sim M(H - L)/2, \text{ with } M$$

$$= H L W d_i \quad (12)$$

The density of ice is d_i and that of seawater d_w . The characteristic time scale of toppling can be estimated from the period of an iceberg interpreted as a mathematical pendulum:

$$T \sim 2\pi \sqrt{d_i H / ((d_w - d_i)g)} \quad (13)$$

Equations 12 and 13 do not take into consideration the remaining kinetic energy the iceberg has after reaching its stable position at the horizontal travel distance $(H - L)/2$. Furthermore, the friction between the iceberg and the glacier ice front along with the interaction of the iceberg with seawater and the mélange of smaller icebergs and fragments in front of the calving front is not taken into account. However, assuming the very reasonable dimensions $H = 800$ m, $L = 100$ m, and $W = 400$ m, $\text{CSF} \sim 10^{13}$ kg m (Eq. 12), $T \sim 150$ s (Eq. 13), and $M_s = 4.8$ according to the relation between $\log_{10}(\text{CSF})$ and M_s (Fig. 7b).



Seismic Sources from Landslides and Glaciers, Fig. 9 Relations derived from the rotational slider block model; (a) maximum force observed – maximum force calculated; (b) α , α_0 – $\log_{10}(\text{CSF})$

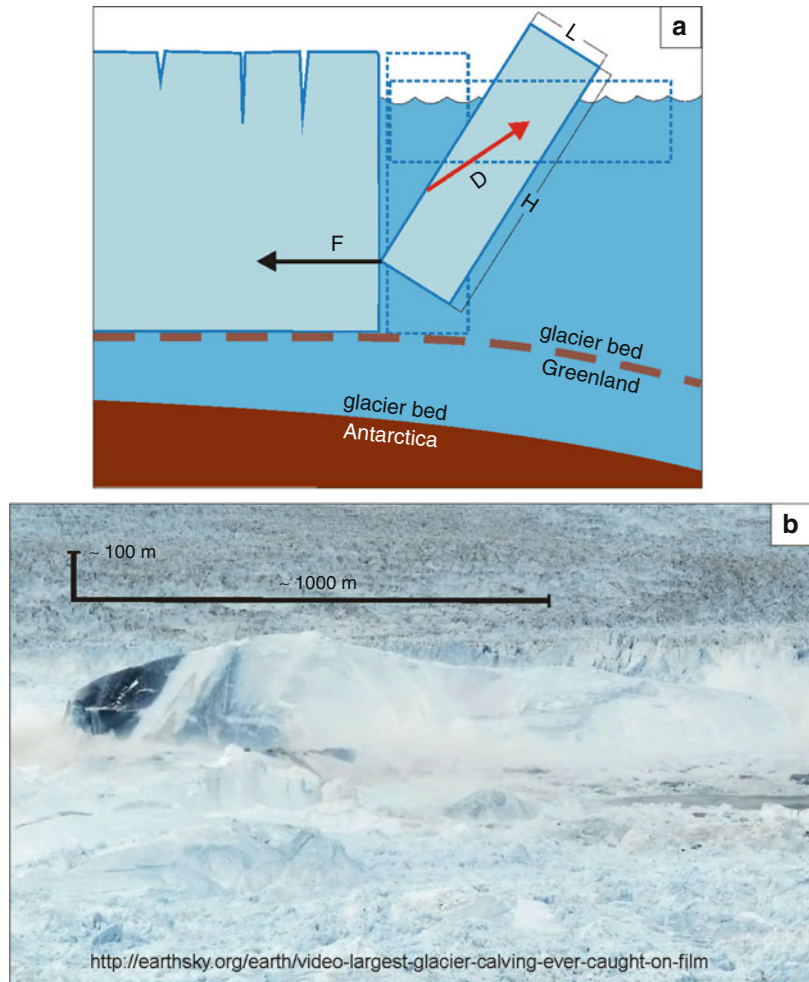
Microearthquakes Generated by Landslides

The dynamic collapse of a slope within several tens of seconds is only representative of how a landslide can end in the most extreme cases. However, the majority of landslides do not develop into dynamic and catastrophic failure but reach more or less stable final states after

phases of varying levels of activity. A deeper understanding of landslide processes and evolution is of fundamental importance for early warning and hazard estimation. Seismic activity connected to these processes can be observed by local seismic networks deployed on the landslide mass. We term microearthquakes generated by landslides LMEs (landslide microearthquakes) in the following.

Seismic Sources from Landslides and Glaciers,

Fig. 10 Toppling (capsizing) of an iceberg acting as single force seismic source; (a) kinematics and acting force (toppling may also occur by counterclockwise rotation); (b) toppling of an iceberg during the 28 May 2008 calving event at Jakobshavn glacier

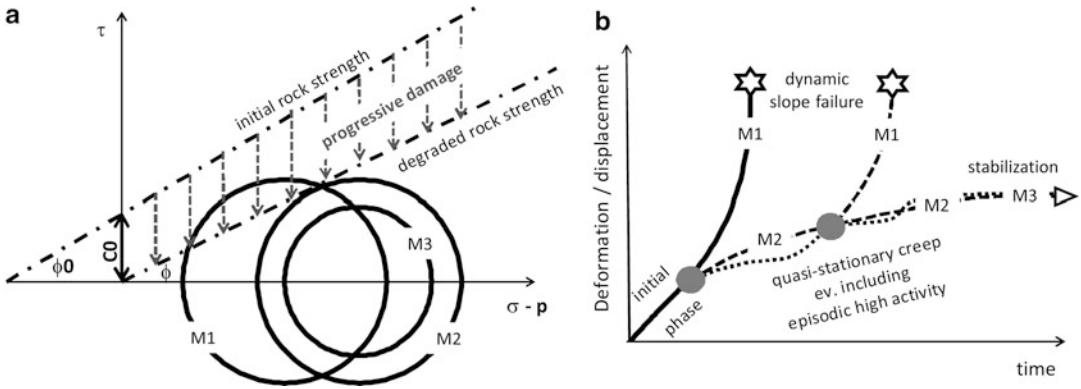


Landslide Processes and Evolution

We assume brittle rheology of the rock material constituting a landslide as a precondition for seismic activity. However, under some circumstances, the landslide mass may behave as ductile, and even a brittle deformation will not necessarily act as a seismic source. In the following, we present a simplified scheme of landslide processes and possible evolutions to dynamic failure or stabilization (Fig. 11). This schema should provide a backbone for the interpretation of the seismic data provided by the case studies presented in sections “Pre- and Postfailure Processes at Steep Slopes” to “Flow”.

Slopes of a given inclination and height can only exist if the stability is guaranteed by the initial strength of the rock material. The possible

physical states are depicted in Fig. 11a in a Mohr diagram by the three Mohr's circles ($M1$, $M2$, $M3$) and Mohr-Coulomb failure criteria representing the initial and degraded strengths of the rock mass forming the slope. The Mohr's circles symbolize the states of stress, which are relevant for the stability of the slope. The initial rock strength is given by a Mohr-Coulomb failure criterion for the shear stress, with C_0 and ϕ_0 the initial cohesion and the initial friction angle. Increasing pore pressure p shifts the Mohr's circles to the left and nearer to the initial rock strength. In an extreme case of high pore pressure, a dynamic rupture could be induced; however, this is not the only situation that reduces the slope stability. Subcritical crack growth (Brückl and Parotidis 2005) supported by deep chemical



Seismic Sources from Landslides and Glaciers, Fig. 11 Landslide processes and evolution; (a) reduction of rock mass strength; (b) temporal evolution of displacement

or thermomechanical weathering (Gischig et al. 2011) may progressively damage the rock material and degrade the rock strength. Ultimately, cohesion (C_0) is completely lost and the friction angle diminished considerably from ϕ_0 to ϕ . This failure criterion of the degraded rock strength intersects Mohr's circle M1 and is tangential to M2, which means that in the first case a dynamic failure has already occurred. In the second case, the slope is in an equilibrium state of driving (gravitational) and resistive (frictional) forces. Creep of the whole slope may reduce the inclination and therefore the driving forces, and this situation is symbolized in Mohr's circle M3. It is well below the failure criterion of the degraded rock. The slope comes to an ultimate stabilization by this process.

The temporal evolution of the deformation or displacement of a landslide mass is schematically shown in Fig. 11b. The initial phase corresponds to the progressive damage from initial to degraded rock strength (Fig. 11a), where the whole rock mass of the landslide is degraded and deformed. However, deformation concentrates successively to one basal or several main sliding zones or surfaces. Toward the end of the initial phase, the further development of the slope comes to a bifurcation. In the case where Mohr's circle M1 represents the critical state of stress in the slope, the initial phase leads directly to a dynamic slope failure. In the case of Mohr's circle M2-type critical state of stress, a phase of

quasistationary creep, eventually interrupted by episodic phases of high creep or sliding velocities, follows. However, we cannot exclude that episodic high activity is the precursor for a later bifurcation between quasistationary creep and dynamic failure, as indicated in Fig. 11b by a second grey dot. The longer the quasistationary phase lasts, the more the driving forces and the radius of the Mohr's circle reduce, so that in the final state of stress represented by M3 the slope stabilizes. Decrease of the pore pressure p by improved drainage of the slope due to increased permeability of the rock mass may support stabilization.

Rockfall

Rockfall is a phenomenon occurring on slopes where, at least locally, the inclination equals or exceeds that of talus of the same rock material. It may be the consequence of weathering but also an indication of increased activity in a landslide. The seismic signature of rockfall may be a sequence of many irregular impulsive signals moving down the slope. Figure 13 shows seismic recordings from the tumbling of a medium-sized boulder (~ 2 m diameter) at the Gradenbach site (Fig. 12), images of the boulder itself and the site of heaviest impact, and the seismic location of that place (Brückl et al. 2013).

As a second example, the seismic signature of a rockfall at Steinlehen (Fig. 12) is documented in Fig. 14. This rockfall was more of a small rock



Seismic Sources from Landslides and Glaciers, Fig. 12 Monitoring sites in the European Alps

avalanche, which left behind a path resolved by ground-based InSAR monitoring. The whole event lasted around 30 s, and location applied to several time windows allowed for the tracing of the path of this rockfall. There is fair agreement between both tracing methods (Weginger 2012).

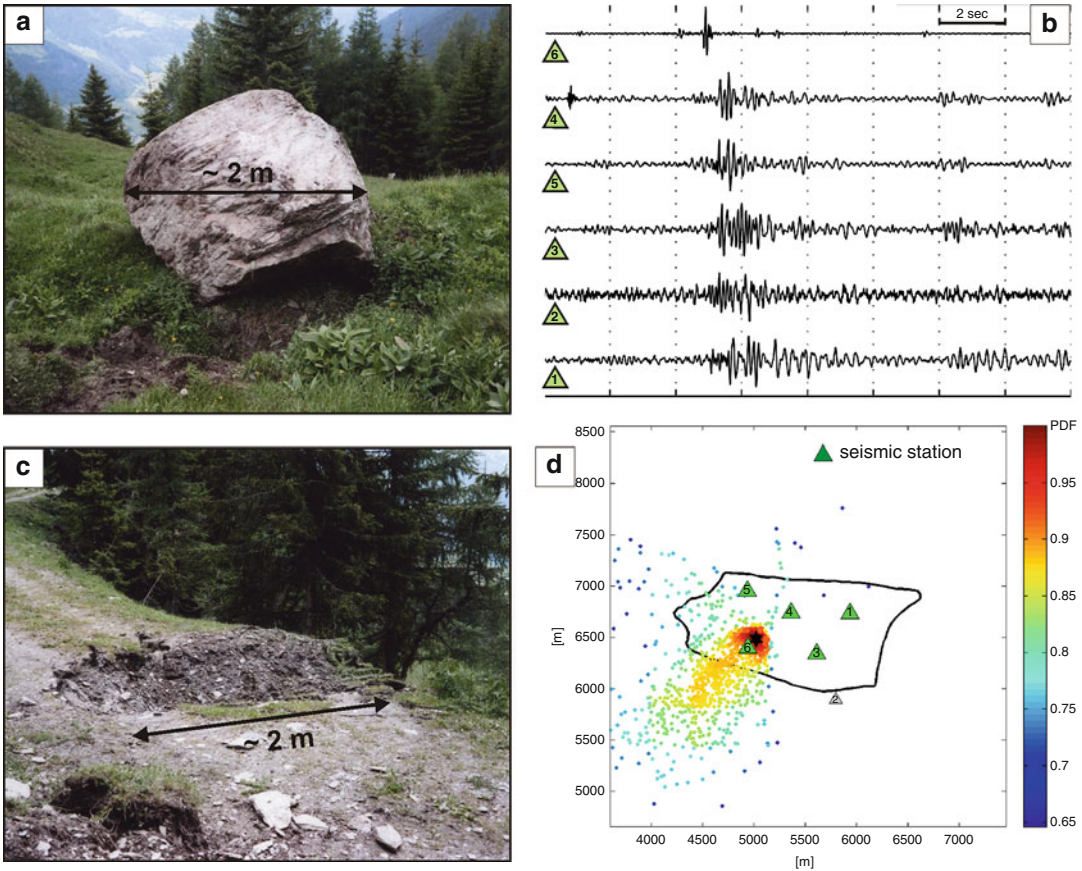
Pre- and Postfailure Processes at Steep Slopes

We denote “steep slopes” as slopes that are steeper than the initial angle of friction or the inclination of talus built up by the same rock material. The rock material of these slopes needs cohesion to balance the gravitational forces. In the case of cohesion dropping to zero because of progressive damage (Fig. 11a), the slope will immediately transit from this initial phase to dynamic failure (Fig. 11b). Identification and monitoring of precursors of this kind of dynamic rupture would be of great scientific and practical importance.

The retreat of cliffs along the British and French Channel Coasts by sea erosion-driven rockfalls is an ongoing process. Amitrano et al. (2005) report seismic monitoring of a cliff collapse at Mesnil-Val on the Normandy coast, France. Precursory seismic activity was recorded

about 1 h before the collapse. Frequency of detected seismic signals and the seismic energy obey inverse Omori’s laws. The data suggests that a warning based on increased activity could be given ~10 min before the collapse and a reliable prediction of the rupture time about 1 min in advance, provided data acquisition and continuous updating of calculations are done in real time.

The village of Randa, Matter Valley, Swiss Alps (Figs. 12, 15), was threatened by rock- and icefalls several times in its history. We anticipate a look from rock to ice and icefalls because of the similar phenomenology behind these processes in rock and ice. Icefalls were released by the Weisshorn hanging glacier five times in the last 35 years. The break-offs in 1973 and 2005 were monitored by geodetic and seismic methods. The velocity of the glacier front developed according to an inverse Omori law (Faillettaz et al. 2011) toward the instant of rupture. Log-periodic oscillations appeared superimposed on the general trend. Starting 30 days before rupture, three phases characterized by different cumulative size-frequency distribution of the signal energies and interevent waiting-time distribution behavior were identified and a correlation of seismic activity with the log-periodic oscillations established.



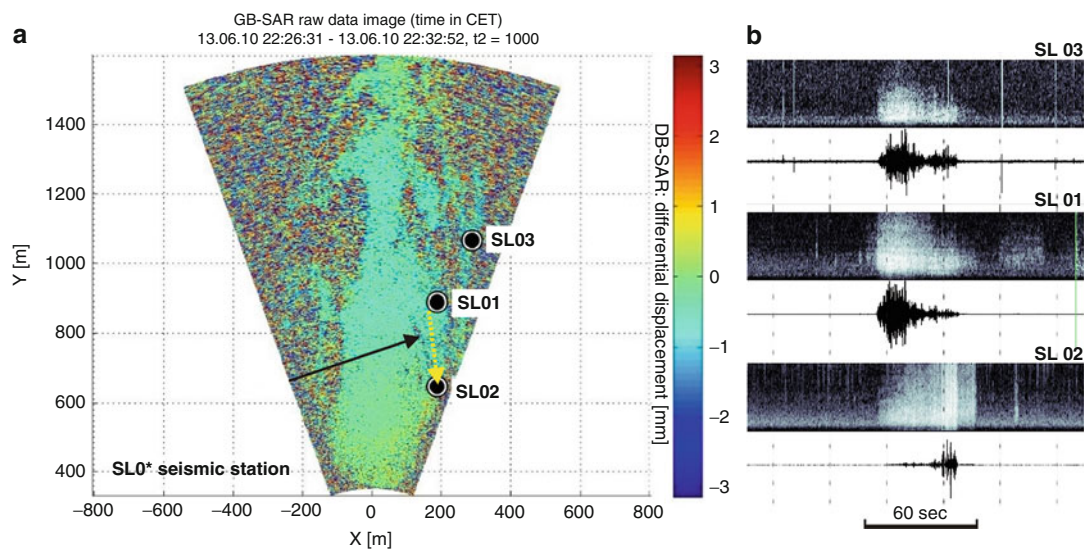
Seismic Sources from Landslides and Glaciers, Fig. 13 Rockfall at the Gradenbach landslide (Brückl et al. 2013); (a) main boulder; (b) recordings by the

local network; (c) main impact; (d) location by the seismic data (normalized PDF) and actual position of the impact (star)

These phases represent the development from random crack generation over crack concentration at a fracture zone to the final rupture. The findings could be the basis of an efficient in-time warning.

From the detour to ice and a hanging glacier, we revert to rock and landslides. On 18 April and 9 May 1991, two successive rockslides took place near Randa (Fig. 15). The failure volumes involved about 20 and 10 10^6 m^3 massive para- and orthogneisses (Eberhardt et al. 2004). The average slopes of the failure planes were roughly 60° and 50° (Fig. 16). The rock masses dropped down 200 and 300 m and produced M_s 4.9 and 4.8 earthquakes detected by Ekström and Stark (2013). Deformation of the slope above the actual scarp continued after the rockfall sequence of the

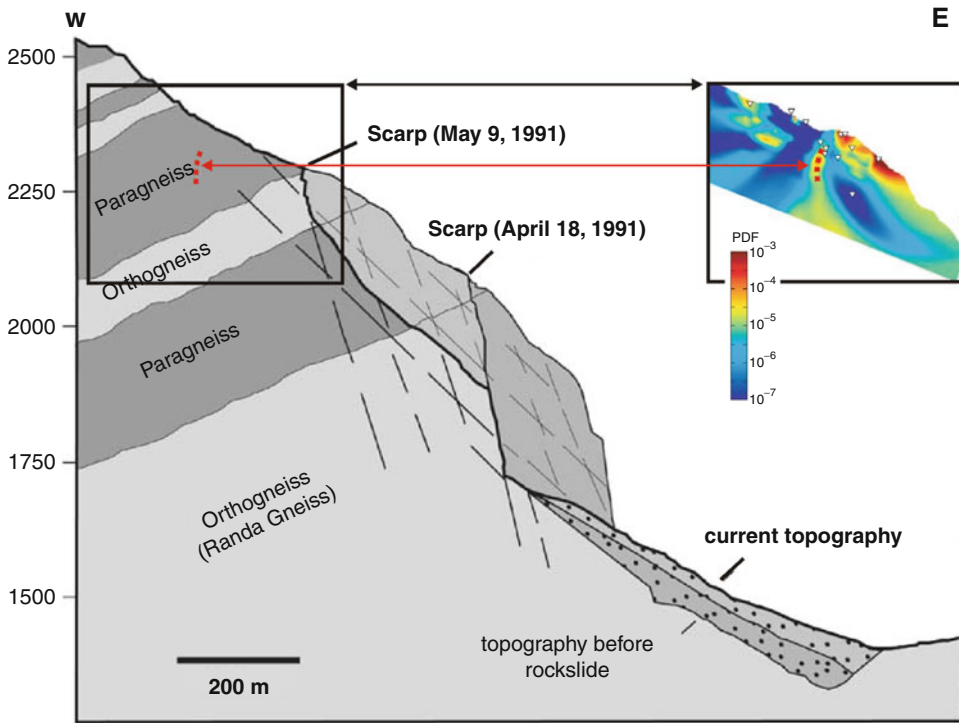
year 1991 with $\sim 10\text{--}20 \text{ mm/year}$ (2004–2008). A seismic monitoring campaign was started in January 2002 and continued until July 2004 (Spillmann et al. 2007), where a monitoring network was deployed above the scarp (Fig. 15). Two types of waveforms classified as LMEs were observed: 94 single events comparable to the recordings of small detonations and 129 multiple events with a more complex source time function. Frequency content of the signals reached up to 100 Hz for the surface sensors and up to 200 Hz for the borehole sensors. Most of the seismic activity was concentrated on a volume of low seismic velocities ($\leq 1500 \text{ m/s}$) near the scarp, but also deeper-reaching zones of increased seismic activity were identified, possibly outlining the scarp of a



Seismic Sources from Landslides and Glaciers, Fig. 14 Rockfall at the Steinlehn landslide (Weginger 2012); (a) differential ground based InSAR image shows the trace of a rockfall (*black arrow*); *yellow arrow* represents the path of the rockfall from the seismic data shown in (b)



Seismic Sources from Landslides and Glaciers, Fig. 15 View from east to the glacier break-off and rockslide sites near Randa; location of the profile in Fig. 16 is shown by a *dashed red line*



Seismic Sources from Landslides and Glaciers,
Fig. 16 Cross section through the Randa rockslide site
 (Modified after Eberhardt et al. 2004); cumulative PDF of

LMEs is shown within the detail marked by *black rectangle* (Modified after Spillmann et al. 2007); the location of a possible next scarp is marked by *red dots*

future rock slide (Fig. 16). Hydrological and meteorological conditions had no impact on the seismic activity and the geodetically observed displacement rates. Most magnitudes were in the range $-2.0 < M_w < -0.5$. It is questionable if the observed seismic activity and geodetically observed displacements are postfailure relaxation processes or if they belong to the initial phase of a third, larger rock slide at Randa.

DSGSD: Deep-Seated Gravitational Slope Deformation

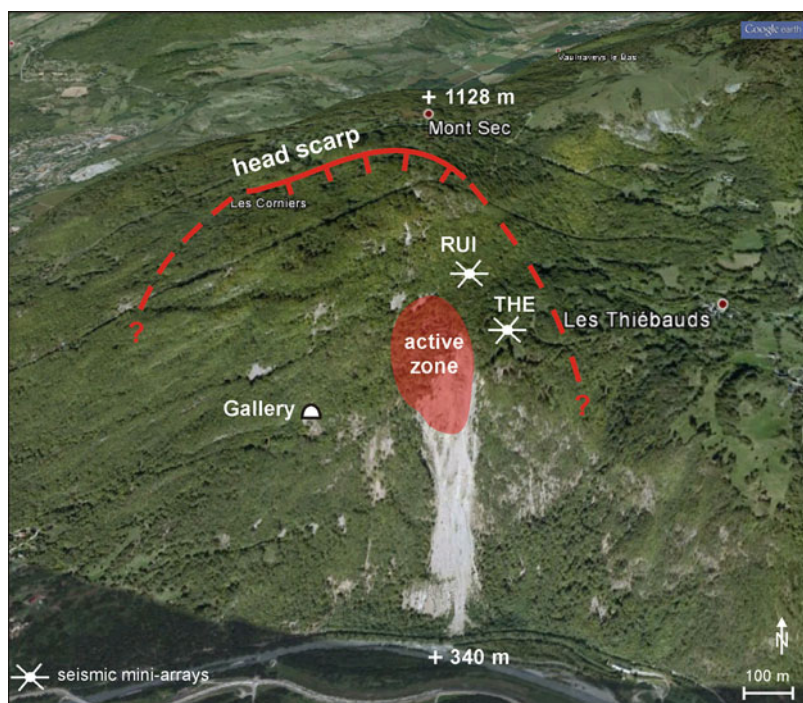
Deep-seated gravitational slope deformation (DSGSD) is a phenomenon frequently observed in the hard rock slopes of alpine regions. Typically, the deformation comprises the whole flank of a valley, and the maximum thickness of the deformed rock mass exceeds 100 m. The mechanisms of DSGSD are manifold, and deformation styles such as “Bergzerreiung,” “Talzuschub,” sagging or Sackung (Zischinsky 1969), and kink

band slumping (Goodman and Kieffer 2000) can be summarized under DSGSD. The initiation of DSGSD at alpine slopes mostly dates back to the retreat of the ice age glaciers, when the lateral support of the valley flank by the ice pressure was lost. This conception implies a long initial phase of mass movement, during which progressive damage degraded the rock strength almost to its residual value (see Fig. 11a; Brckl and Parotidis 2005). DSGSDs may be in a quasistationary creep phase or may exhibit episodic high deformation rates in parts of the slope or its entirety interrupting longer phases of little activity (Fig. 11b). There is currently no general answer as to whether these processes eventually lead to stabilization or a transition to dynamic rupture.

In the following, we consider the seismic signals recorded at three DSGSD examples in crystalline rocks of the European Alps: Schilienne (Helmstetter and Garambois 2010), Steinlehen (Weginger 2012), and Gradenbach (Brckl

Seismic Sources from Landslides and Glaciers,

Fig. 17 View from south to the Séchillienne DSGSD and the active zone (area colored in red); the locations of the 2 seismic micro-arrays RUI and THE are marked by symbols



et al. 2013; Mertl 2015). The locations of these landslides are shown in Fig. 12; Figs. 17, 18, and 19 supply information about the morphology, the seismic monitoring network, and partly the geology and structure of these landslides. The average slopes of these DSGSDs vary between 28° (Gradenbach) and 35° (Steinlehn).

At the Séchillienne DSGSD, a presently very active zone is located on top of the steeper lower part of the slope (marked in Fig. 17). The velocity of this active zone reached 1.4 m/year in 2008. Until April 2009, rockfalls were the most frequent seismic sources related to the dynamics recorded at the Séchillienne DSGSD. The cumulative frequency of >3000 events obeys a power law distribution over 2.5 decades of the amplitude range with an exponent corresponding to a b-value of $b = 1.1$.

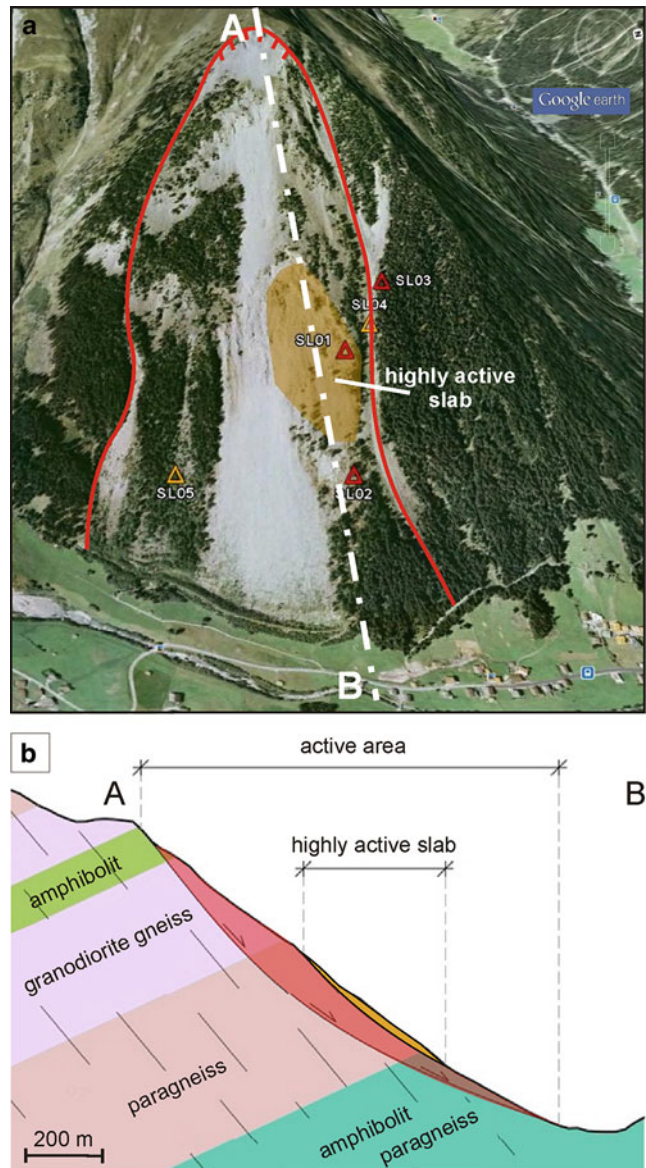
Figure 20 shows typical waveforms of LMEs at the Séchillienne DSGSD. One type corresponds to impulsive sources similar to small explosions, while another event type shows mostly emergent onsets and represents source-time functions lasting ~5 s, either relatively broadband or low frequency. The LMEs cluster either in the

high-velocity area or the head scarp. It is also possible that some hypocenters are located at or near a basal shear zone. Rockfall and LME frequencies precede accelerations of the active zone by 4–5 days. Some but not all rainfall events trigger increased rockfall and seismic activity (Helmstetter and Garambois 2010).

Seismic monitoring campaigns at Steinlehn and Gradenbach had the opportunity to observe episodic phases of high landslide velocities interrupting longer periods of quasistationary creep. During June 2010, the Steinlehn DSGSD was monitored by a ground-based InSAR system installed opposite the Steinlehn DSGSD. An acceleration phase of the active slab was observed with total displacements reaching up to 0.9 m. Up to five seismic monitoring stations were operating at the Steinlehn slope during this period (Fig. 18a). Besides the rockfall (see Fig. 14), Weginger (2012) identified three classes (n, In, i in Fig. 21a) of LMEs that had their sources within the area of the accelerated slab. This classification was based on characteristic properties of the seismic signals (e.g., duration,

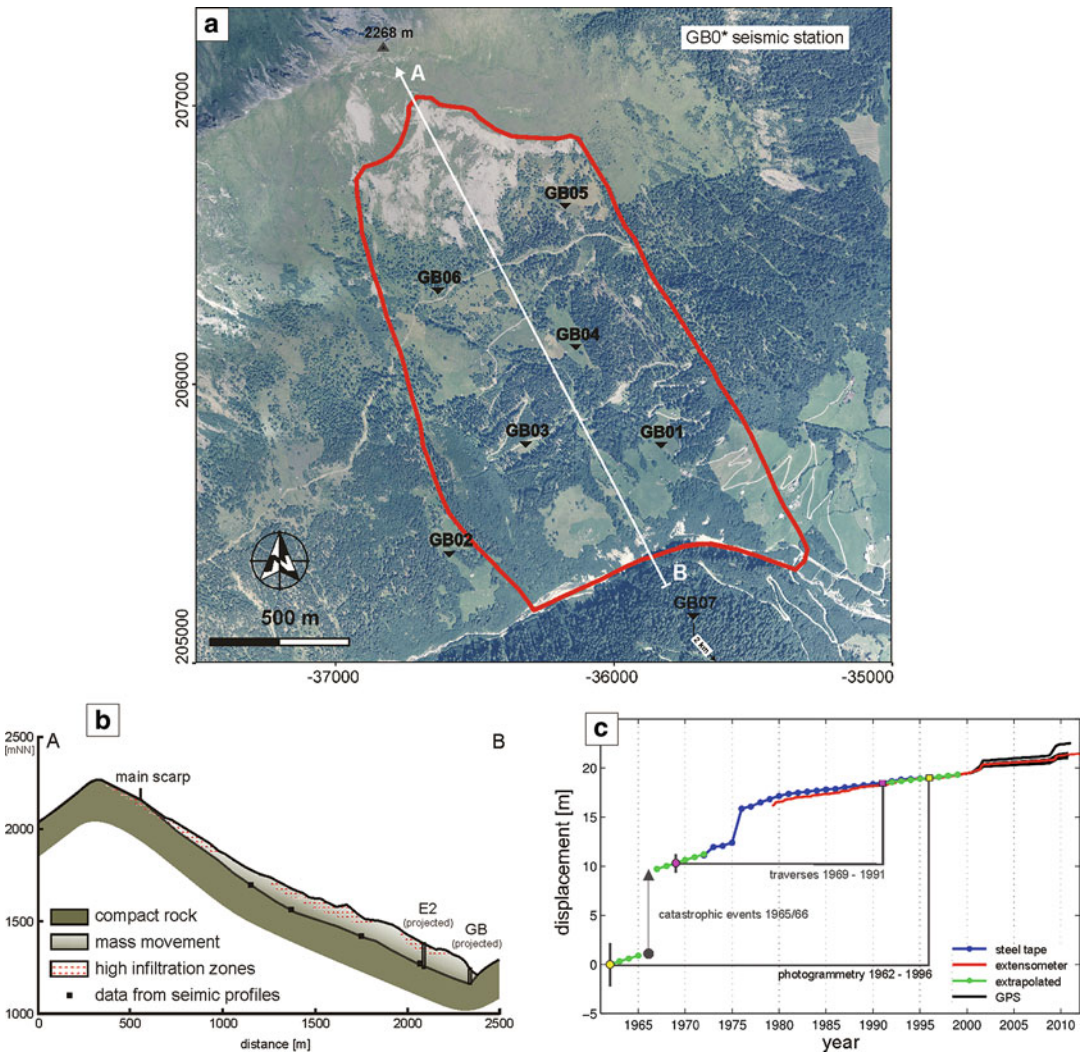
Seismic Sources from Landslides and Glaciers,

Fig. 18 Steinlehn DSGSD (Weginger 2012); (a) view from east to the slope with contour of the active zone; triangles mark the locations of the seismic monitoring stations, the extent of the slab accelerating during the observation period is colored in ochre; (b) geological cross section A-B (Modified after Zangerl et al. 2007)



frequency spectral characteristic, amplitude, envelopes, cross correlations, polarization). Figure 21b shows velocity and displacement of the highly active slab for the period of 10–28 June 2010. The high-velocity phase starts on 20 June and lasts until 25 June. The cumulative frequencies for LME types n, ln, and i rise from 13 June onward, and some rockfalls also occur (Fig. 21c). A sudden increase of the cumulative rockfall frequency accompanies the onset of the high-velocity phase between 20 and

21 June, and rockfall ceases completely thereafter. The cumulative frequencies of the events n, ln, i increase more or less continuously with the highest increments during the high-velocity phase. No extraordinary rainfall was reported to precede or accompany the period of high velocities of the active slab, and no external trigger of this phase is known so far. The incipient seismic and rockfall activity on 13 June was most likely the only precursor of the high-velocity phase in the following week.

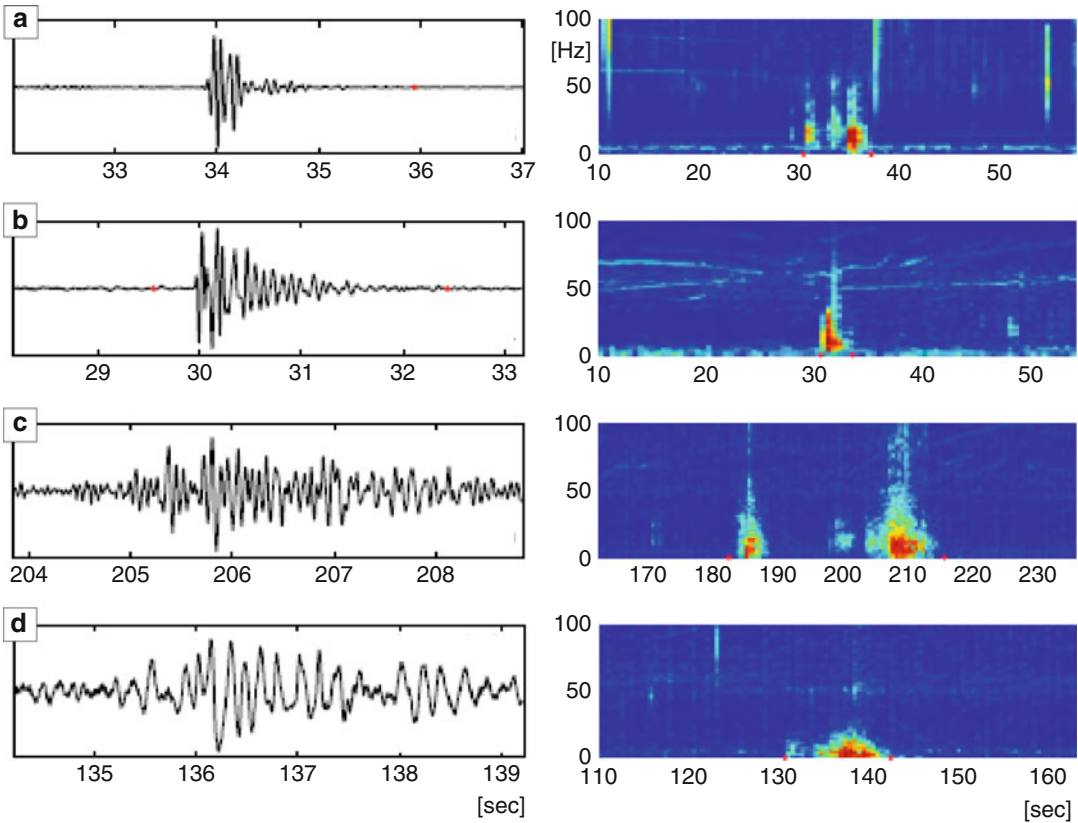


Seismic Sources from Landslides and Glaciers, Fig. 19 Gradenbach DSGSD (Brückl et al. 2013): (a) Orthophoto with contour of the active zone and seismic stations GB01–GB07; the main scarp, the highly fractures zone in the uppermost area of the mass movement, and the erosional zone at the toe of the mass movement are well visible in the orthophoto; (b) cross section A-B; the depth

of the basal sliding surface or zone was determined by drilling (boreholes E2 and GB) and refraction seismic measurements (c) displacement history derived from terrestrial geodetic measurements 1969–1991, photogrammetric evaluation of the aerial photos 1962 and 1996, and GPS since 1999

The episodic high-sliding velocity phase at the Gradenbach DSGSD occurred during spring of 2009. In contrast to the Steinlehn event, the whole landslide became active due to infiltration of melting snow and precipitation (Brückl et al. 2013). Five different LME types (AA, A, B, D, tremor; Fig. 22a) besides rockfall (rf) were identified using criteria comparable to those used

for the data from the Steinlehn site (Mertl 2015). The cumulative frequencies of these events behave distinctly. Event types A and D start with the onset of snow melt (Fig. 22b–d) and high infiltration. The tremor type was only observed at that time. Thereafter, a quiet phase that lasted over 1 month followed. The beginning of the episodic phase of high creep velocity was



Seismic Sources from Landslides and Glaciers, Fig. 20 Typical waveforms and spectrograms from LMEs recorded at the Séchilienne DSGSD (Modified

after Helmstetter and Garambois 2010); (a) impulsive LME; (b) small explosion; (c) emergent and broad band LME; (d) emergent and low frequency LME

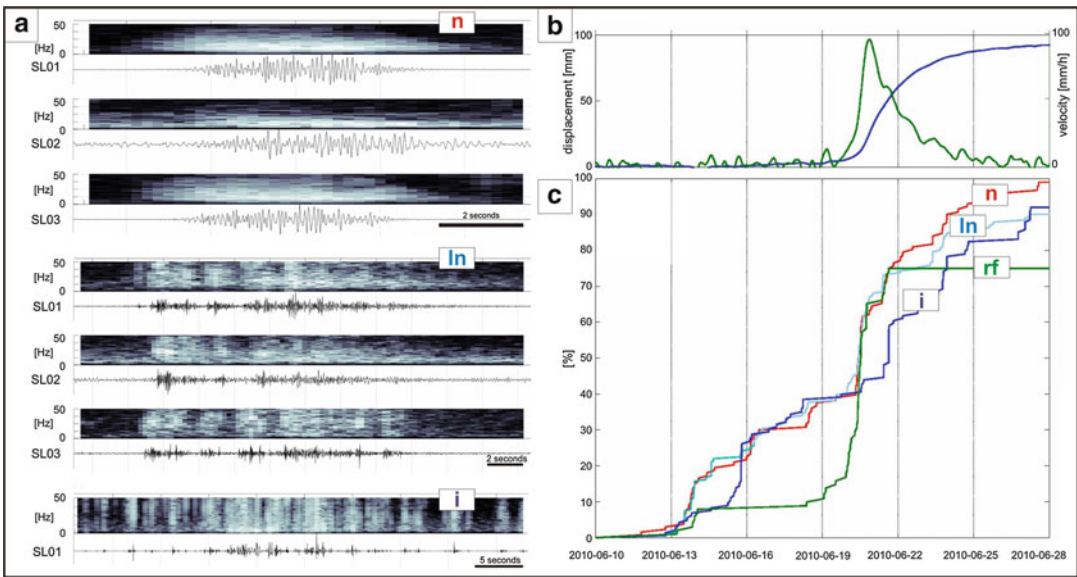
accompanied by type B microearthquakes and some rockfall. The frequency of type A and D events rose 1–2 weeks later and lasted for the rest of the episodic high-creep phase. Magnitudes of the LMEs range from -2.8 to -0.9 with a b -value of $b = 1.7$.

Though the geodetically observed movements covered the whole landslide mass, the spatial distribution of the seismic activity during the episodic high-creep phase is concentrated to limited areas (Fig. 23). Event types A, D, and tremor, which coincide with the onset of snow melt and high infiltration, are most likely very shallow events. The waveforms and coincidence in time with the onset of high creep velocities may be indications that the hypocenters of B-type LMEs are at the level of the basal sliding zone. Asperities must have been overcome before the whole

landslide mass could increase in velocity (Mertl 2015).

Flow

A flow is a spatially continuous movement in which surfaces of shear are short lived, closely spaced, and not usually preserved. Whereas in DSGSD the original structure of the rock mass is partly preserved, in a flow it is completely destroyed and mixed up. The ratio of mean thickness to length of a flow is <0.01 and significantly lower than for DSGSDs. When considering the scheme of landslide processes and evolution presented in Fig. 11a, it is evident that the initial rock (or soil) strength has been completely reduced to its residual values. Changing pore pressure may keep the flow either active or temporarily inactive (Mohr's circles M2 and



Seismic Sources from Landslides and Glaciers, Fig. 21 LMEs observed at the Steinlehn DSGSD, 10–28 June 2010 (Weginger 2012); (a) typical waveforms and spectrograms of LME types n (emerging signal, few seconds length), ln (irregular sequence of n-type signals), i

(irregular sequence of impulses over a period of several seconds); (b) displacement and velocity; (c) development of cumulative frequency of event types n, ln, i, and rf (rockfall)

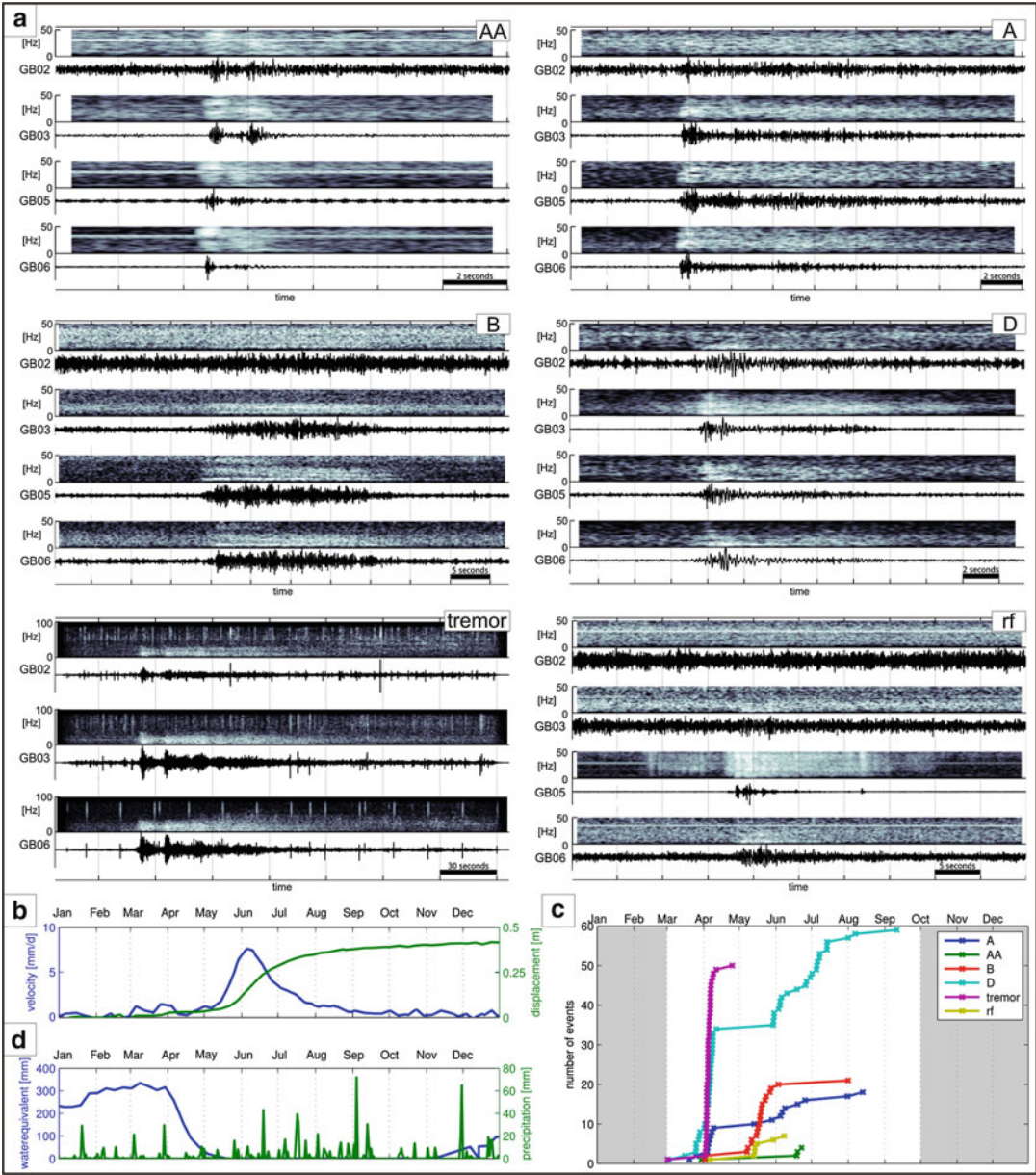
M3 in Fig. 11a). Flow here represents quasistationary creep including episodic high activity, depending on infiltration and the pore pressure (Fig. 11b). The distribution of velocities in space and time within the landslide mass justifies its macroscopic description as a ductile medium. However, as the following two examples will prove, brittle deformations or interactions between pore fluid and the rock/soil mass also exist and resemble seismic sources.

The Super-Sauce landslide is a flow-type mass movement situated in the Barcelonnette basin in the southern French Alps (Fig. 12). It developed from heterogeneous soft Jurassic black marks (Fig. 24). The average slope of the flow is $\sim 23^\circ$. The landslide triggered around 1960, ruling out a causal connection to the retreat of ice age glaciers. From 1996 to 2007, the velocities covered the range of 2–30 mm/day.

Seismic monitoring at the Super-Sauce landslide was carried out between 14 and 24 July 2008 (Walter et al. 2012). Besides rockfalls, 34 LMEs were located and labeled as “slide

quakes.” They have an impulsive onset, last 2–5 s, and cover a wide frequency band (10–80 Hz). Later-arriving phases have lower frequencies (Fig. 25a). The epicenters are concentrated around the high-velocity area of the landslide. Magnitudes between $-3.2 \leq M \leq -1.3$ were detected following a power law distribution with a b-value of $b = 0.84$. The hypocenters were probably located within the uppermost few meters and not at the basal sliding zone or surface. The highest magnitudes of slide quakes followed few hours after a strong rainfall event; the temporal distribution was otherwise uniform. Single miniarrays also detected very high-frequency events (5–150 Hz) with durations of 2–20 s (Fig. 25b). These LMEs were concentrated around a stable crest within the high-velocity zone of the flow and related to fissure development in the uppermost layer of the flow. Laboratory tests proved that the flow material can have brittle rheology.

The second example of a flow is the Slumgullion landslide (Fig. 26). The landslide occurs in

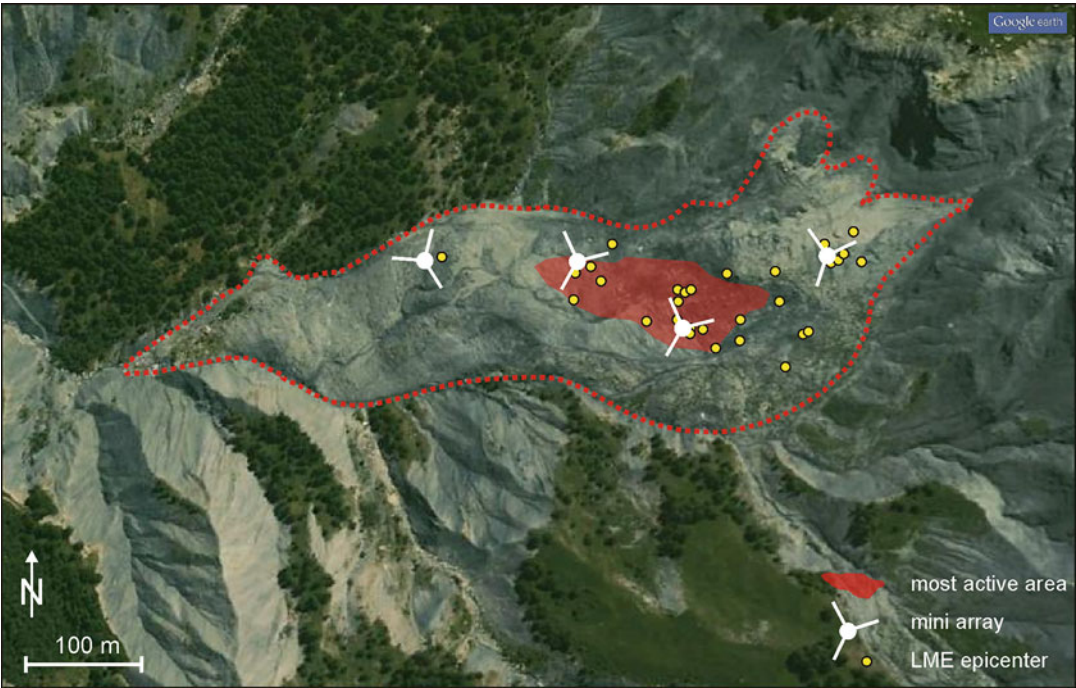
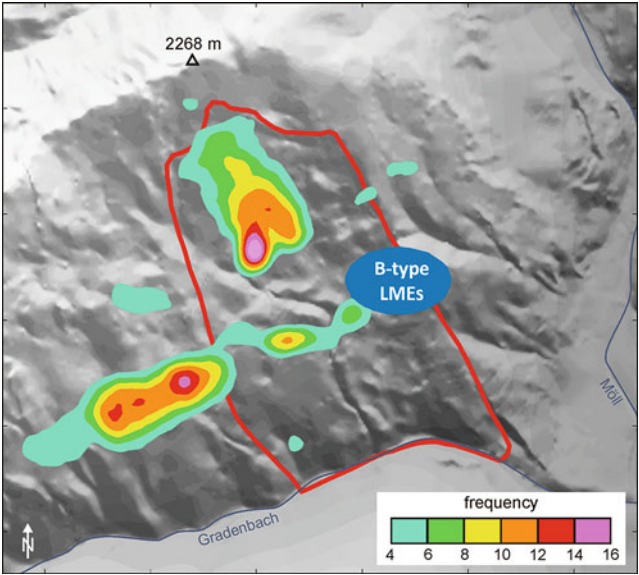


Seismic Sources from Landslides and Glaciers,
Fig. 22 LMEs observed at the Gradenbach DSGSD, March–October 2009 (Brückl et al. 2013; Mertil 2015); (a) typical waveforms and spectrograms of LME types AA (impulsive onset with short lower frequency coda), A (like AA but with long coda), B (emergent onset, low frequency, typical duration 20–30 s), D (emergent onset

but significantly higher frequency content and shorter duration than B), tremor (emergent low frequency signal, duration >1 min, superimposed by irregular spikes), rf (rockfall); (b) displacement and velocity; (c) snow cover water equivalent and precipitation; (d) development of cumulative frequency of event types

Seismic Sources from Landslides and Glaciers,

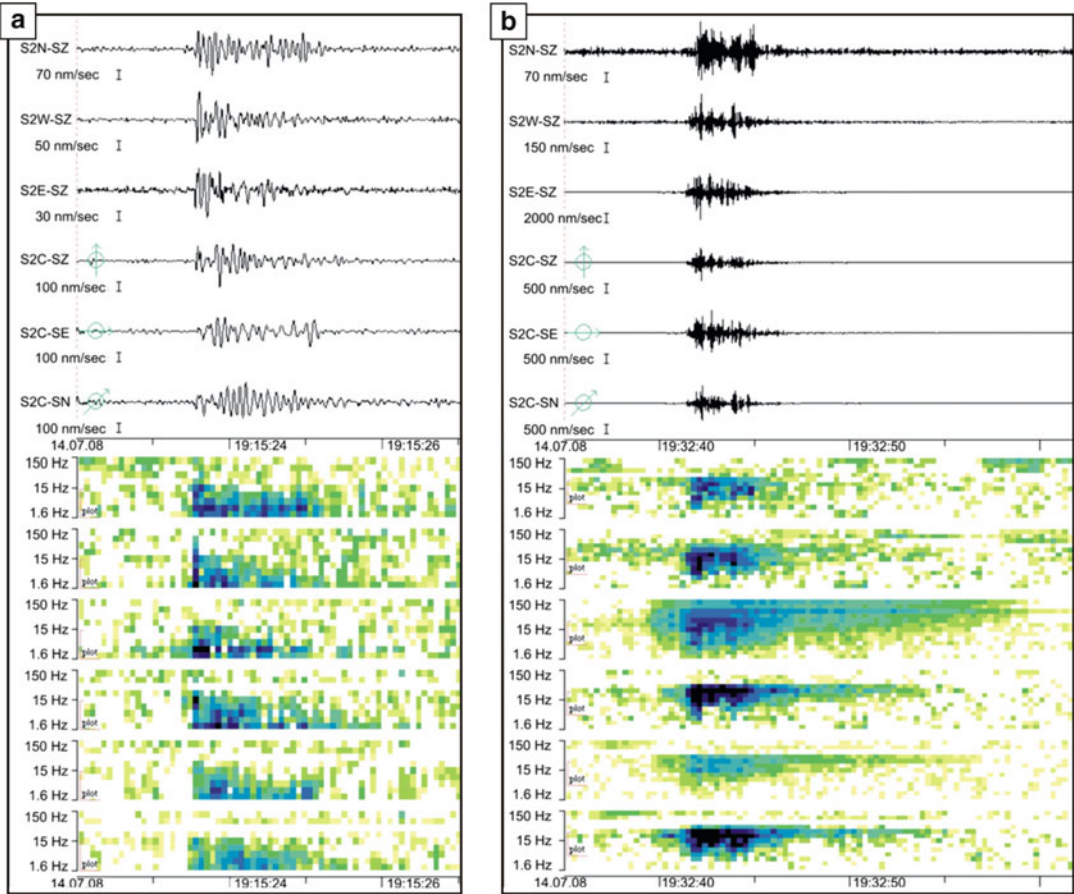
Fig. 23 Spatial distribution of seismic activity March–October 2009 at the Gradenbach DSGSD; colored contour plot visualizes the frequency of possible epicenters with normalized PDF > 0.7, event types A, AA, D; *blue* area marks area of event type B epicenters (Brückl et al. 2013; Mertl 2015)



Seismic Sources from Landslides and Glaciers, Fig. 24 Super-Sauce flow, *red dotted line* outlines landslide area (Walter et al. 2012)

Tertiary volcanic rocks and comprises sandy, silty clay with scattered patches of boulder debris, clay, and pond and stream sediments. The landslide consists of an active upper part,

which moved over an older, inactive lower part in its path. The average slope of the active part is 13°. The Slumgullion moves primarily by sliding along discrete bounding faults. Additionally,

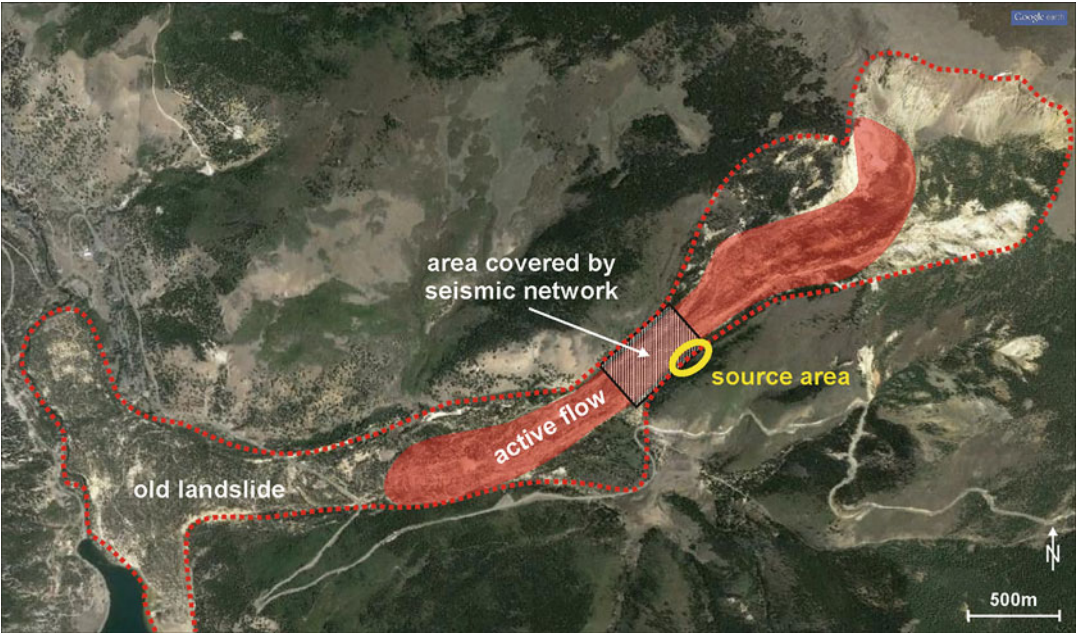


Seismic Sources from Landslides and Glaciers, Fig. 25 Typical waveforms and sonograms of LMEs observed at Super-Sauce debris/mud flow (Modified after Walter et al. 2012); (a) “slide quakes”; (b) “fissure development”

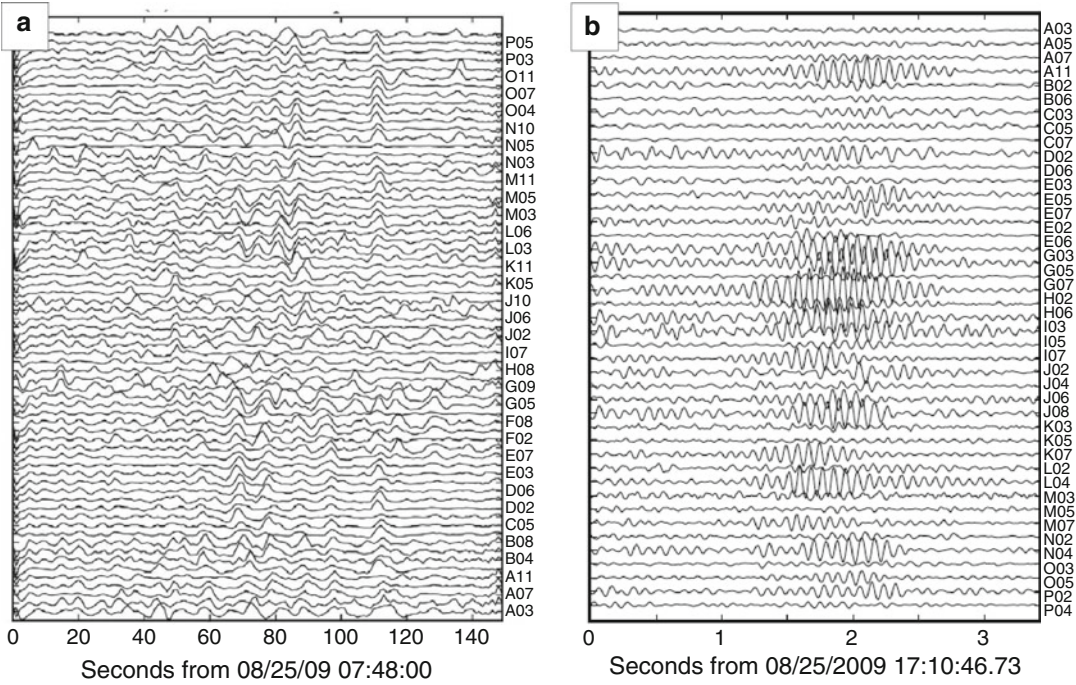
differential internal processes take place at faults and fractures within the landslide body.

During 18–26 August 2009, a seismic monitoring campaign was carried out at the Slumgullion landslide (Gomberg et al. 2011). Geodetic and hydrometeorological monitoring supplemented the campaign. During the monitoring campaign, the displacement trend was ~ 10 mm/day. One stronger period of rainfall occurred during the observation period. Several seismic signals were detected and related to LMEs: “tremor” and “harmonic slide quakes” (Fig. 27a, b). The dominant energy of tremor is distributed broadly above 30–50 Hz and

diminishes toward 125 Hz. High amplitudes of tremor envelopes last < 10 s and follow irregularly in intervals of several tens of seconds. Harmonic slide quakes last ~ 2 s, their fundamental frequency is ~ 12 Hz, and no harmonics are visible throughout the whole duration. This LME type occurs more frequently during daytime, and an increase in sliding velocity is also observed during the day. Tremor and harmonic slide quakes most probably originate at the left side-bounding strike-slip faults of the landslide. The stick-slip behavior of the side-bounding strike-slip faults is explained by a transient dilatant strengthening.



Seismic Sources from Landslides and Glaciers, Fig. 26 Slumgullion landslide (37° 59' 30" N, 107° 15' 25" W), San Juan Mountains, Colorado, USA (Gomberg et al. 2011)



Seismic Sources from Landslides and Glaciers, Fig. 27 Typical LME waveforms observed at Slumgullion flow (Modified after Gomberg et al. 2011); (a) envelopes of tremor; (b) harmonic slide quake

Microearthquakes Generated by Glaciers

Ductile and Brittle Deformations of Glaciers

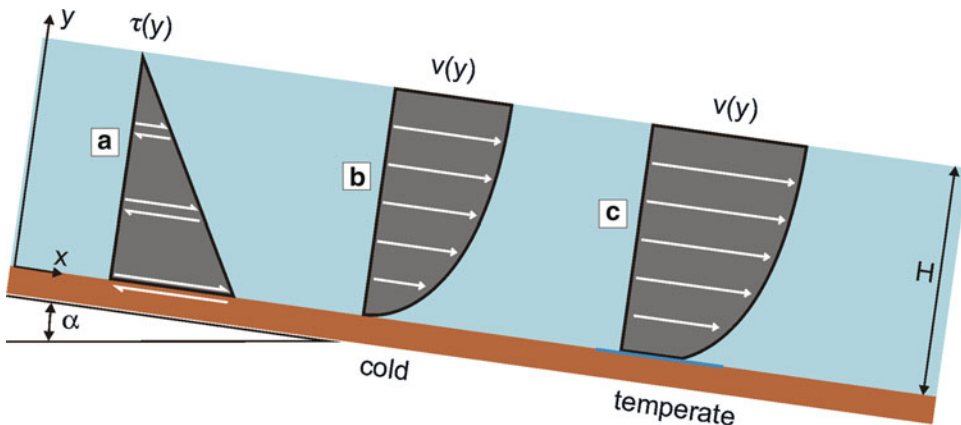
In contrast to landslides, which are intrinsic transient phenomena, glaciers or ice sheets can reach a long-lasting stationary state in steady climatic conditions. Snow is transformed to ice and added to the glacier in the accumulation area and removed in the ablation area. Glacier flow transports ice from the accumulation area to the ablation area, thereby keeping the total mass as well as the thickness of the glacier constant under stationary conditions. The glacier flow is driven by gravitational forces, controlled by the rheology of the ice and physical conditions that allow for sliding of the glacier ice over its bed (Cuffey and Paterson 2010).

We consider an infinite slab of ice with constant thickness H resting on a plane glacier bed with the slope α . Shear stress parallel to the glacier bed is proportional to $\sin(\alpha)$ and increases linearly with the overburden (Fig. 28a). According to Glen's law, the corresponding shear strain rate is proportional to the n th power of the shear stress, with $n \sim 3$.

Integration yields the velocity profile, which is an $(n + 1)$ th-order parabola. For cold glaciers, the ice temperature is below melting point at the base, and the ice may be frozen to the glacier

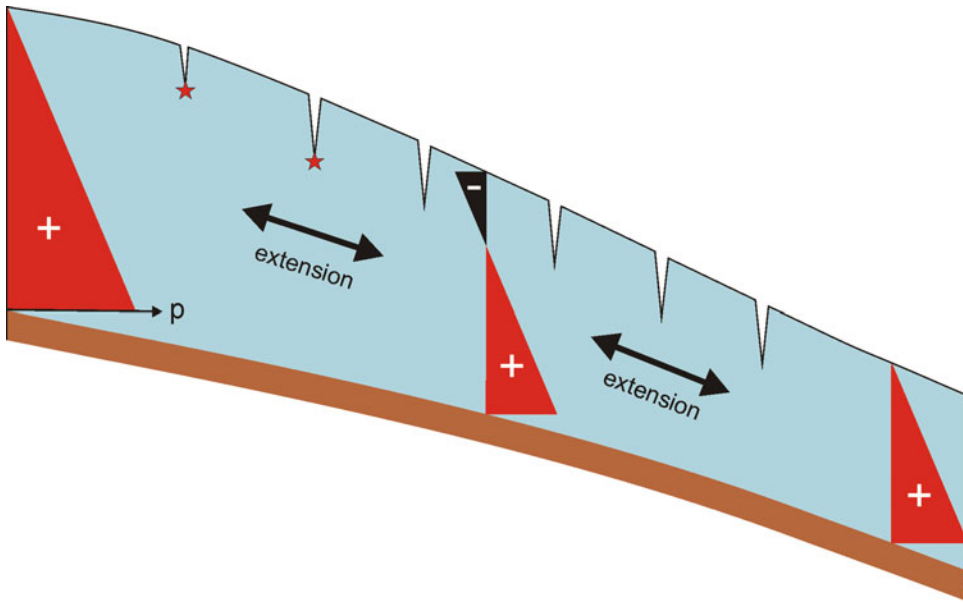
bed (Fig. 28b). For temperate glaciers, a thin coat of water between ice and the bedrock is assumed, and sliding occurs over the thin viscous water layer (Fig. 28c). Sliding at the glacier bed is controlled by obstacles, which are overcome either by enhanced flow due to stress concentration at the obstacles or regelation. All processes related to glacier flow described so far are either ductile (internal deformation of the ice, viscous flow of the water coat at the glacier bed, enhanced flow over obstacles) or correspond to another steady-state process (regelation) and cannot produce earthquakes.

However, the rheology of ice changes significantly in the case of the hydrostatic pressure component p of the state of stress becoming negative. In this case, the ice behaves in a brittle manner and may radiate seismic energy. First, we consider tensile crack nucleation and propagation without the influence of water. Glacier thickness adapts to varying slopes of the glacier bed by keeping the basal shear stress approximately constant. Therefore, the thickness is inversely proportional to $\sin(\alpha)$. Neglecting accumulation or ablation over a relatively short distance of an increasing slope, the velocity must increase to keep the flux of ice constant (Fig. 29). Longitudinal tensile stress must develop to extend the glacier ice moving from the gentle to the steeper slope. Surface crevasses develop



Seismic Sources from Landslides and Glaciers, Fig. 28 Planar slab of ice resting on a sloping plane; (a) shear stress parallel to surface; (b) velocity profile of a

cold glacier – no basal sliding; (c) velocity profile of a temperate glacier – basal sliding



Seismic Sources from Landslides and Glaciers, Fig. 29 Crevasse formation due to thickness adaption to increasing slope; stars mark possible locations of GMEs

to the depth at which the hydrostatic pressure equals the longitudinal tensile stress component. On alpine glaciers, surface crevasse depths seldom exceed 25–30 m. After reaching the zone of constant steeper slope, the tensile stress component disappears and the crevasses close by ductile deformation.

Water significantly influences the mechanics of landslides and glaciers. The fundamental difference between landslides and glaciers with respect to the effect of water is that buoyancy is lower than hydrostatic pressure in landslides and higher in glaciers due to the comparative densities of water ($\sim 1000 \text{ kg m}^{-3}$), rock ($\sim 2500 \text{ kg m}^{-3}$), and ice ($\sim 900 \text{ kg m}^{-3}$). Therefore, a water column of sufficient height can introduce a tensile stress regime and lead to crack development and propagation at all glacier depths. Possible scenarios are shown in Fig. 30.

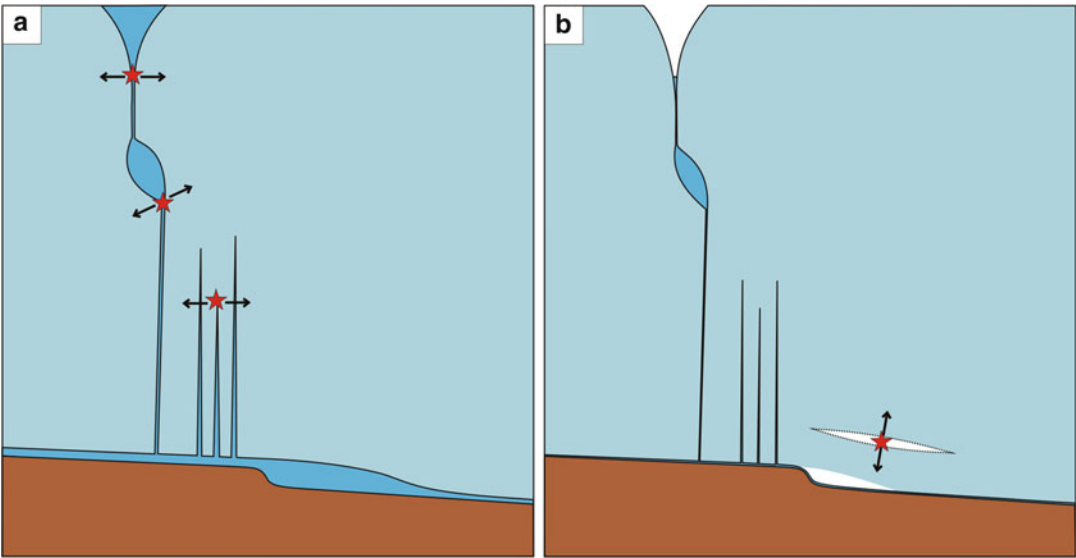
Thus far, we have only considered hard-bedded glaciers where the glacier bed consists of hard rock and is not deformed by the glacier flow. However, it is well known that fast-moving outlet glaciers in Alaska, Antarctica, and Greenland rest on a basal layer of soft glacial till, which makes up a considerable part of the basal sliding

(Alley et al. 1987). The mechanical strength of glacial till obeys the Mohr-Coulomb failure criterion and may be comparable to the masses of flow-type landslides treated in section “Flow.” Although steady shear deformation may dominate, experience from seismic monitoring on landslides shows that stick-slip deformation may also occur. At the glacier beds of polythermal and temperate glaciers, there may also be asperities or “sticky spots,” which induce stick-slip basal sliding under the otherwise nonseismic deformation conditions at the glacier bed.

In the following, we term seismic activity connected to the various processes of glacier flow and observed by local networks GMEs (glacial microearthquakes).

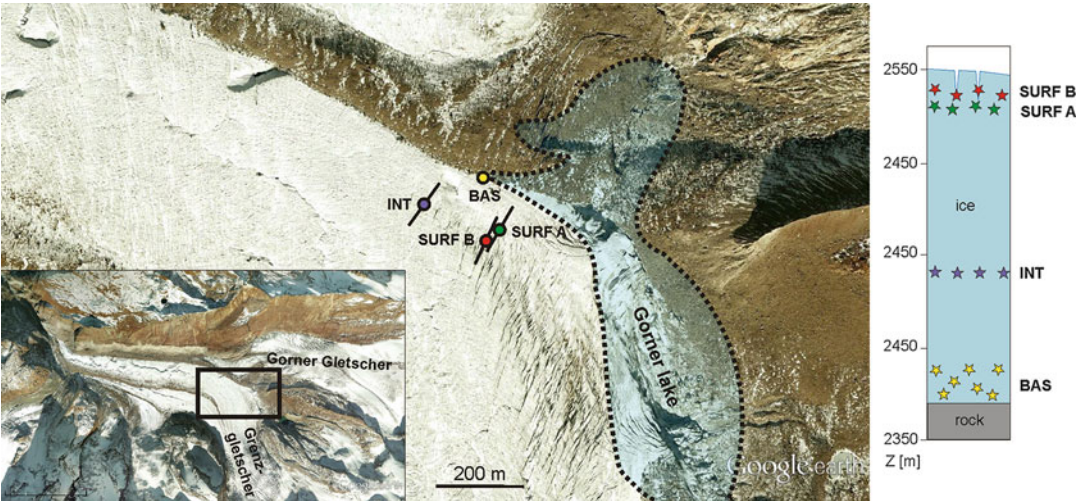
Glacial Lake Outburst Floods at an Alpine Glacier

As an example of seismic activity of an alpine glacier, we review observations made at the polythermal Gornergletscher, Monte Rosa Massif, Switzerland (Fig. 12). At the confluence of Gornergletscher and Grenzgletscher, there is the ice-dammed Gorner lake (Fig. 31). The lake



Seismic Sources from Landslides and Glaciers, Fig. 30 Hydraulically induced fracturing in a glacier (Modified after Walter 2009; Walter et al. 2010); stars mark possible locations of GMEs; (a) opening of internal and basal crevasses due to high water pressure; meltwater

influx from active moulin or basal channels (b) closure of internal and basal crevasses due to decreasing water pressure; the collapse of a basal cavity generated during the high water pressure phase acts as seismic source



Seismic Sources from Landslides and Glaciers, Fig. 31 GMEs observed at the Gornergletscher before and during the GLOFs 2004 and 2006 (Walter 2009); (a) location of the GME clusters; fault plane strike of the

SURF A, SURF B, and INT GME clusters agree with the orientation of crevasses; dotted line outlines approximately the Gorner lake at the maximum water levels; (b) focal depths of the different GME clusters

drains during annual glacial lake outburst floods (GLOFs), which were monitored in the years 2004 and 2006 with seismic and geodetic networks (Walter 2009; Walter et al. 2010). During

the 2004 and 2006 field seasons, about 35000 and 50000 events were detected. Generally, diurnal variations in the seismicity were observed, with the peak in the early afternoon and much lower

GME activity during the nighttime and morning hours. GME clusters were recorded at three different depth levels: SURF_A and SURF_B at the depth of surface crevasses, INT at ~ 100 m depth, and BAS just above the glacier bed (Fig. 31b). The SURF_A, SURF_B, and INT GMEs have magnitudes between $-2.3 \leq M_w \leq -1.6$; the BAS GMEs reach $M_w = -0.7$. With the exception of BAS GMEs, fault planes are steeply dipping, and strike directions of the interpreted fault planes follow the general trend of surface crevasses (Fig. 31a). The radiation pattern of the SURF_A and INT LMEs can best be explained by the generation of tensile cracks, but the SURF_B events more likely result from a DC mechanism. The BAS GMEs occur on subhorizontal fault planes and are well explained by tensile cracks. The source mechanisms of the SURF_A, INT, and BAS cluster correspond to the processes schematically outlined in Figs. 29 and 30. The BAS events occurred mainly during otherwise quiescent night and early morning hours. This observation supports the idea of a collapsing basal cavity due to the reduced melt-water access and basal water pressure (Fig. 30b). The occurrence of DC-source mechanisms with the SURF_B GMEs may be due to the particular dynamic situation near the Gorner lake and the GLOF.

Arctic Glacier Prone to Surges

The temperate Bering Glacier flows from the St. Elias mountain range to its terminus on the south-central coast of Alaska and has a history of dramatic surges. Seismic monitoring was carried out between the equilibrium line altitude and the terminus (Fig. 32a) and covered the early melting period from 20 April until 19 June 2007 (West et al. 2010). Yet another surge started in 2008 (Burgess et al. 2012).

Around 160,000 events were detected during the observation period. Location of the GMEs was not attempted. The waveforms of the GMEs (Fig. 32b) show a bimodal distribution and range from high-frequency signals (20–35 Hz) to low-frequency signals (6–15 Hz). The high-frequency GMEs have impulsive source functions and were interpreted as brittle failures.

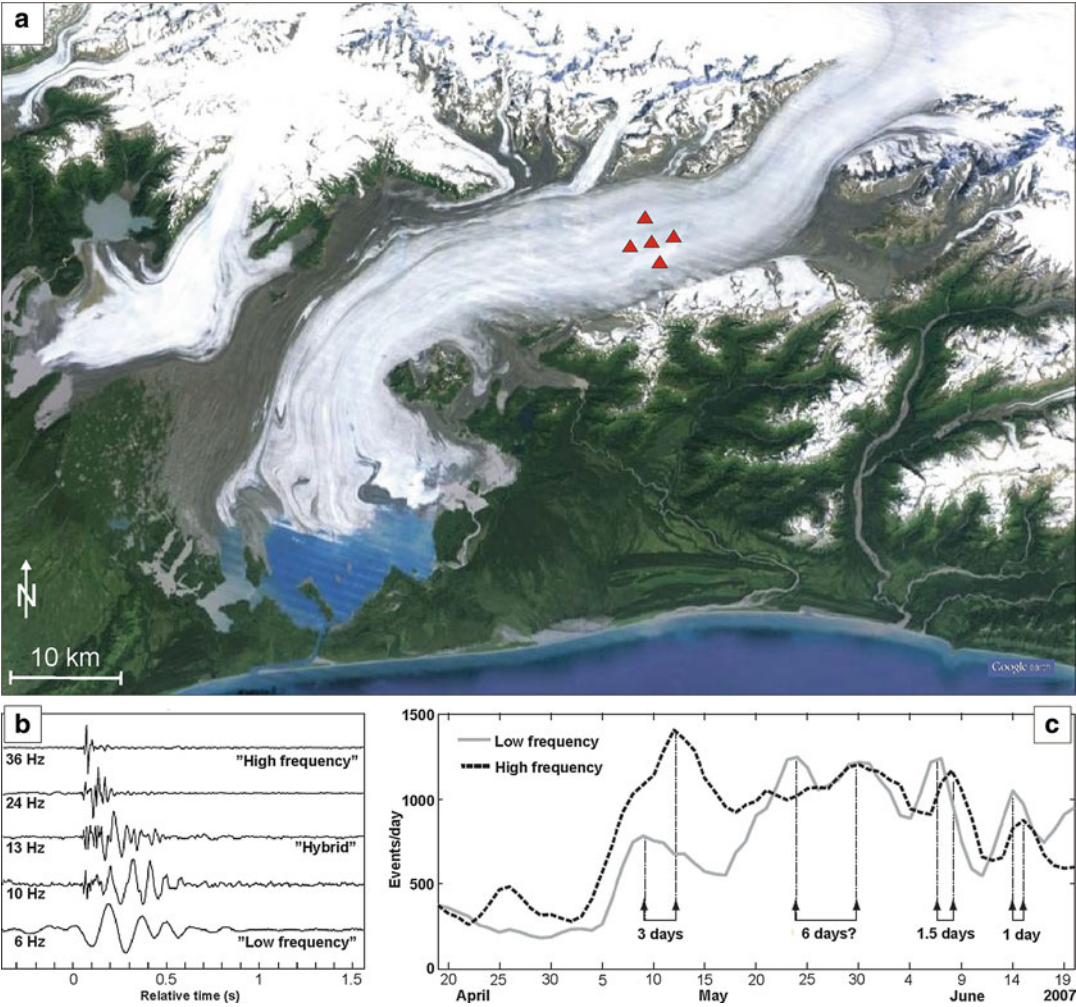
In an analogy to waveforms of volcanic earthquakes, the low-frequency tremor GMEs were explained as a fluid-driven crack model with a resonant water-filled cavity as the seismic source. Hybrid GMEs were interpreted as hydrofracturing followed by the rush of water into the new opening.

Relative maxima of the daily rate of low-frequency events precede the corresponding maxima of the high-frequency events by 1–6 days (Fig. 32c). The interpretation of this pattern is that the peaks of the low-frequency events reflect the subglacial rush in of water leading to a decoupling of the glacier from its bed and consequently enhanced flow velocities. The enhanced glacier motion is accompanied by an increase of brittle failures. Velocity data would be essential to further develop this interpretation.

Stick-slip Movement of an Antarctic Ice Stream

Bindenschadler et al. (2003) reported on major West Antarctic ice streams discharging in sudden periods of rapid motion, in particular Whillans Ice Stream (WIS) (Fig. 33a), which has an area of $200 \text{ km} \times 100 \text{ km}$ with an average ice thickness of ~ 600 m. During December 2010 and January 2011 and again during December 2011, Winberry et al. (2013) deployed a GPS and seismic network on the WIS in order to monitor the slip nucleation and the subsequent period of rapid motion (Fig. 33b). Figure 33c shows the displacements for a time span of ~ 1 h, covering the end of a preceding interevent period, the nucleation phase, the rupture, the main-slip phase, and the transition to the following interevent period. Clear evidence for the nucleation phase is shown only at one GPS station (station 1), where the rupture occurs within ~ 5 min at all stations according to the rupture propagation speed of >500 m/s. The main-slip phase lasts ~ 30 min and terminates almost simultaneously at all stations.

GMEs recorded at the individual stations correlate well in time and frequency domain with the ice stream velocities (Fig. 34). At station 1, seismicity increases at the beginning of the



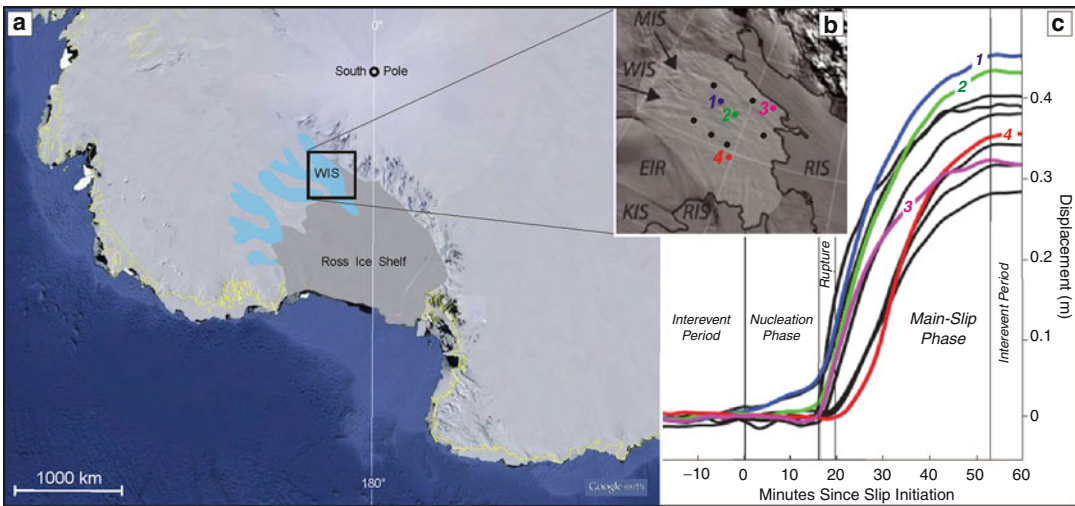
Seismic Sources from Landslides and Glaciers,
Fig. 32 Seismic monitoring at the Bering glacier (Modified after West et al. 2010); (a) location of seismic

stations (red triangles); (b) waveforms; (c) temporal evolution of daily rates of high and low frequency event

nucleation phase but only begins at the other stations around the rupture time. Two different styles of GMEs or microseismicity can be distinguished. One type (Fig. 34a) is built by a sequence of fewer than a hundred to several thousand individual impulsive events. The wide distances between the seismic stations did not allow for a precise location; however, the relative timing of P and S wave arrivals and the lack of crevasses is consistent with source locations at or near the glacier bed. The individual events can be grouped into families with a high degree of similarity between each other. The same event

families persist in subsequent high-slip-rate phases. This observation is an indication of the temporal stability of subsurface asperities representing the source regions of the individual events. Other locations show in general a similar behavior but are characterized by lower frequencies, indicating larger source areas. The lack of microseismic activity or GMEs at some locations may be explained by the presence of more continuous fault gouge and higher water pressure at the glacier bed providing basal lubrication.

The second style of microseismicity is the emergent tremor (Fig. 34b). It is continuous and



Seismic Sources from Landslides and Glaciers, Fig. 33 Geodetic and seismic monitoring of an episodic high flow velocity phase at the Whillan Ice Stream (WIS); (a) location map; (b) zoom of the monitoring area with

locations of monitoring-stations; (c) displacements versus time since slip initiation; numbers and colours assign displacement data to stations in (b) (Modified after Winberry et al. 2013)

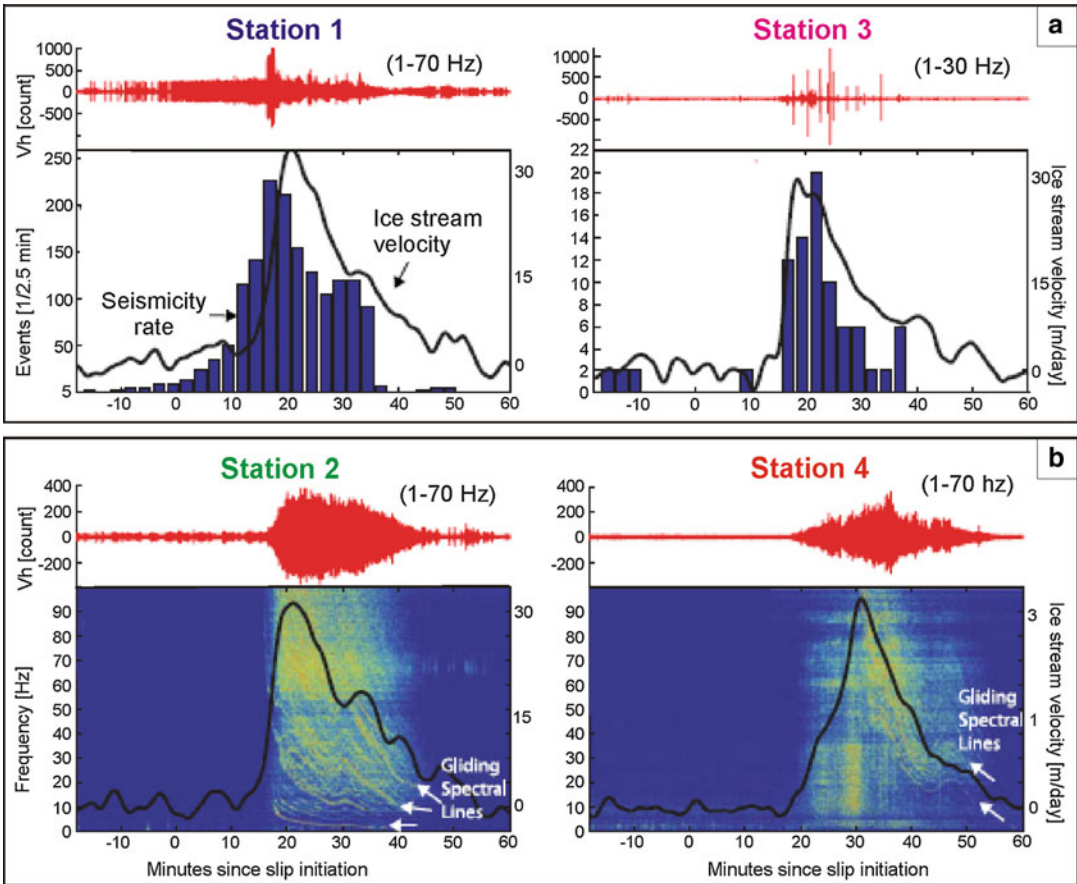
cannot be resolved into individual events. The overall spectral power of this tremor correlates well with the glacial flow velocities. Seismic energy during the tremor is concentrated in the horizontal components. A source, radiating shear energy from near the glacier bed, is consistent with this observation. The spectra reveal coherent gliding spectral lines, some of which are overtones of fundamental frequencies. Periodic stick–slip failure of small asperities could be the source mechanism of these seismic signals. Winberry et al. (2013) argue that the highly repeatable WIS system represents an excellent natural laboratory which could provide insight into friction and sliding complementary to laboratory or other field studies (e.g., Gombert et al. 2011).

Summary

A new class of global earthquakes ($M_s > 4.5$) was discovered and related to giant and catastrophic landslides or phenomena occurring near the front of fast-moving outlet glaciers in Greenland and Antarctica. The unique

characteristics of these earthquakes are the depletion of short-period energy, the elongated source time functions, and their radiation patterns, which can be explained by a single surface force better than by seismic moments. Efficient detection and location methods were developed and force histories $F(t)$ along with the centroid single force (CSF) derived. The CSF is the product of the total moving mass M and the displacement D of the centroid. It represents a major constraint on the whole source process.

Originally, the new-class global earthquakes with sources on glaciers were interpreted as stick–slip flow behavior of large portions of outlet glaciers in Greenland and Antarctica. However, the postulated movements of the ice masses could not to date be confirmed geodetically. An alternative source mechanism which generates sufficient CSF and provides the appropriate duration of the source function is the capsizing (or toppling) of large icebergs at the front of the outlet glaciers during major calving events. The basic mechanics of this process are well understood, and correlations with glacier flow, seasonal variations of iceberg production, and glacier retreat support the explanation of the



Seismic Sources from Landslides and Glaciers, Fig. 34 Waveforms observed at stations 1–4 at WIS; (a) impulsive, high frequency waveforms; (b) low frequency (Modified after Winberry et al. 2013)

majority of these earthquakes coming about through this source process.

Related to landslide-generated global earthquakes, additional information from satellite images or local reconnaissance helps to separate CSF into mass and total displacement. The inclination of the CSF vector corresponds to the Fahrböschung; the duration of the landslide force history constrains the sliding velocity. Application of a rotational slider block model allows, e.g., for estimates of the reduction of the friction coefficient from static to dynamic conditions.

Microearthquakes related to the processes in landslide masses (LMEs) that evolve in a more steady form or those in a post- or prefailure state have magnitudes of $M < 0$. Seismic networks

deployed on or very near to the landslide mass are necessary to capture these signals. A rich variety of waveforms can be observed and classified by attributes derived from the frequency content, the duration, or the shape of the envelope. Currently, no definite assignment of the different waveforms to landslide types and processes can be established. Impulsive events have been observed on slopes with steep or medium inclinations ($>20^\circ$); tremor or harmonic LMEs dominate the seismic activity of gently sloping flows. Most seismic activity on landslides is rather shallow. However, the location of a few events at the basal sliding zone cannot be excluded. Rockfall and LME frequency correlate frequently with precipitation and precede geodetically observed acceleration. An increase

in seismic activity preceding acceleration without discernible external triggers has also been observed. The general pattern of rockfall and LME activity suggests that the DSGSD- and flow-type landslides are near their yield stress not only at the basal sliding zone but also in their entire moving mass. Further progress in the understanding of LMEs would be supported by denser seismic networks including borehole seismometers as a standard. Long-term seismic monitoring covering seasonal variations and episodic high-velocity phases combined with complementary monitoring (e.g., geodetic, meteorological, hydrological, geomorphologic) are preconditions for testing refined hypotheses about LMEs and their relation to landslide processes and evolution.

Monitoring of glacier flow-related microearthquakes (GMEs) by seismic networks deployed on glaciers and supplemented by geodetic and meteorological devices provides a deeper insight into the stress distribution within the glacier ice, glacier hydraulics, and processes at the glacier bed or within a basal layer of soft sediments. The latter aspect can be exemplarily studied at the Whillans Ice Stream (WIS) in Antarctica. The ice stream exhibits a stick-slip behavior, controlled by tides with a CSF corresponding to $M_s = 5.6$. The duration of this process is above the seismically observable range, but GMEs generated by this process yield unique information about the spatial distribution and mechanical behavior of asperities or “sticky spots” in the basal layer of soft sediments. Research in this field does not only support the understanding of glacier sliding but addresses generally the friction behavior of fault gouges, e.g., the basal sliding zones of landslides or active tectonic faults.

Finally, we touch on how the potential monitoring and analyzing of landslide- and glacier-generated earthquakes contribute to prediction and early warning of catastrophic events. Global landslide or glacial earthquakes always represent the finite state of a development and may message a disaster. Prediction and early warning is not provided by the detection of these earthquakes; however, instantaneous location and

quantification by seismic methods may support mitigation measures and warnings against subsequent hazards like damming of the valley (see Fig. 6c) or increased density of icebergs. LMEs have been proven to indicate a dangerous acceleration of a landslide, usually earlier than the geodetically observable displacement. GMEs precede glacier break-off and glacial lake outburst floods (GLOFs). The results from basic research on LMEs and GMEs lead to a recommendation of seismic monitoring as an essential component of an integrated early warning system for landslides and glacial hazards.

Cross-References

- [Earthquake Location](#)
- [Earthquake Mechanism Description and Inversion](#)
- [Mechanisms of Earthquakes in Iceland](#)
- [Non-Double-Couple Earthquakes](#)
- [Rockfall Seismicity Accompanying Dome-Building Eruptions](#)
- [Volcanic Eruptions, Real-Time Forecasting of](#)

References

- Alley RB, Blankenship DD, Bentley CR, Rooney ST (1987) Till beneath Ice Stream B. Till deformation: evidence and implications. *J Geophys Res* 92(B9):8921–8929
- Amitrano D, Grasso JR, Senfaute G (2005) Seismic precursory patterns before a cliff collapse and critical point Phenomena. *Geophys Res Lett* 32, L08314. doi:10.1029/2004GL022270
- Bindschadler RA, King MA, Alley EB, Anandakrishnan S, Padman L (2003) Tidally controlled stick-slip discharge of a West Antarctic ice stream. *Science* 301(5636):1087–1089. doi:10.1126/science.1087231
- Brantley SR, Myers B (2000) Mount St. Helens – from the 1980 eruption to 2000. U.S. Geological Survey Fact Sheet, FS-036-00, 2 p. <http://pubs.er.usgs.gov/publication/fs03600>
- Brückl E, Parotidis M (2005) Prediction of slope instabilities due to deep-seated gravitational creep. *Nat Hazards Earth Syst Sci* 5:155–172. SRef-ID: 1684-9981/nhess/2005-5-155
- Brückl E, Brunner FK, Lang E, Mertl S, Müller M, Sary U (2013) The Gradenbach observatory-monitoring deep-seated gravitational slope deformation by geodetic,

- hydrological, and seismological methods. *Landslides*. doi:10.1007/s10346-013-0417-1
- Burgess EW, Forster RR, Larsen CF, Braun M (2012) Surge dynamics on Bering Glacier, Alaska, in 2008–2011. *Cryosphere* 6:1251–1262. doi:10.5194/tc-6-1251-2012
- Chen X, Shearer PM, Walter F, Fricker HA (2011) Seventeen Antarctic seismic events detected by global surface waves and a possible link to calving events from satellite images. *J Geophys Res* 116, B06311. doi:10.1029/2011JB008262
- Cuffey KM, Paterson WSB (2010) *The physics of glaciers*, 4th Edition. Academic Press, ISBN 978012333694614, p 704
- Eberhardt E, Stead D, Coggan JS (2004) Numerical analysis of initiation and progressive failure in natural rock slopes – the 1991 Randa rockslide. *Int J Rock Mech Min Sci* 41:69–87
- Ekström G, Stark CP (2013) Simple scaling of catastrophic landslide dynamics. *Science* 339(6126):1416–1419. doi:10.1126/science.1232887
- Ekström G, Nettles M, Abers GA (2003) Glacial earthquakes. *Science* 302:22–624. doi:10.1126/science.1088057
- Faillietaz J, Funk M, Sornette D (2011) Icequakes coupled with surface displacements for predicting glacier break-off. *J Glaciol* 57(203):453–460
- Gischig V, Moore JR, Evans KF, Amann F, Loew S (2011) Thermomechanical forcing of deep rock slope deformation: 2. The Randa rock slope instability. *J Geophys Res* 116:F04011. doi:10.1029/2011JF002007
- Gomberg J, Schulz W, Bodin P, Kean J (2011) Seismic and geodetic signatures of fault slip at the Slumgullion Landslide Natural Laboratory. *J Geophys Res* 116, B09404. doi:10.1029/2011JB008304
- Goodman RE, Kieffer DS (2000) Behaviour of rock in slopes. *J Geotech Geoenviron* 128(8):675–684
- Helmstetter A, Garambois S (2010) Seismic monitoring of Séchilienne rockslide (French Alps): analysis of seismic signals and their correlation with rainfalls. *J Geophys Res* 115, F03016. doi:10.1029/2009JF001532
- Kanamori H, Given JW (1982) Analysis of long-period seismic waves excited by the May 18, 1980, eruption of Mount St. Helens – a terrestrial monopole? *J Geophys Res* 87:5422–5432. doi:10.1029/JB087iB07p05422. ISSN: 0148–0227
- Kawakatsu H (1989) Centroid single force inversion of seismic waves generated by landslides. *J Geophys Res* 94:12,363–12,374. doi:10.1029/89JB01118. ISSN: 0148–0227
- Mertl S (2015) Characterization of local seismic events on the deep-seated events gravitational slope deformation Gradenbach. PhD thesis, Vienna University of Technology, p 176
- Moore J, Albee W (1981) Topographical and structural changes, March–July 1980 – photogrammetric data. In: Lipman P, Mullineaux D (eds) *The 1980 eruptions of Mount St. Helens* Washington. Geological Survey Professional Paper 1250, pp 123–134
- Nettles M, Ekström G (2010) Glacial earthquakes in Greenland and Antarctica. *Annu Rev Earth Planet Sci* 38:467–491. doi:10.1146/annurev-earth-040809-152414
- Nettles M, Larsen TB, Elósegui P, Hamilton GS, Stearns LA, Ahlström AP, Davis JL, Andersen ML, de Juan J, Khan SA, Stenseng L, Ekström G, Forsberg R (2008) Step-wise changes in glacier flow speed coincide with calving and glacial earthquakes at Helheim Glacier, Greenland. *Geophys Res Lett* 35, L24503. doi:10.1029/2008GL036127
- Spillmann T, Maurer H, Green AG, Heincke B, Willenberg H, Husen S (2007) Microseismic investigation of an unstable mountain slope in the Swiss Alps. *J Geophys Res* 112, B07301. doi:10.1029/2006JB004723
- Tsai VC, Ekström G (2007) Analysis of glacial earthquakes. *J Geophys Res* 113:F03S22. doi:10.1029/2006JF000596
- Tsai VC, Rice JR, Fahnestock M (2008) Possible mechanisms for glacial earthquakes. *J Geophys Res* 113, F03014. doi:10.1029/2007JF000944
- Voight B (1981) Time scale for the first moments of the May 18 eruption. In: Lipman P, Mullineaux D (eds) *The 1980 eruptions of Mount St. Helens* Washington. Geological Survey Professional Paper 1250, pp 68–86
- Walter F (2009) Seismic activity on Gornegletscher during Gornese outburst floods. PhD thesis, DISS. ETHNo 18184, ETH Zürich, p 135
- Walter F, Dreger DS, Clinton JF, Deichmann N, Funk M (2010) Evidence for near-horizontal tensile faulting at the base of Gornegletscher, a Swiss Alpine Glacier. *Bull Seismol Soc Am* 100(2):458–472. doi:10.1785/0120090083
- Walter M, Arnhardt C, Joswig M (2012) Seismic monitoring of rockfalls, slide quakes, and fissure development at the Super-Sauze mudslide, French Alps. *Eng Geol* 128:12–22
- Weginger S (2012) Detektion und Lokalisierung seismischer Signale zur Überwachung der Massenbewegung Steinlehen. Master thesis, Vienna University of Technology, p 96
- West ME, Larsen C, Truffer M, O’Neel S, LeBlanc L (2010) Glacier microseismicity. *Geology* 38(4):319–322. doi:10.1130/G30606.1
- Winberry JP, Anandakrishnan S, Wiens DA, Alley RB (2013) Nucleation and seismic Tremor associated with the glacial earthquakes of Whillans Ice Stream, Antarctica. *Geophys Res Lett* 40:312–315. doi:10.1002/grl.50130
- Zangerl C, Eberhardt E, Schönlaub H, Anegg J (2007) Deformation behavior of deep-seated rockslides in crystalline rock. In: *Rock mechanics: proceedings of the 1st Canada-US rock mechanics symposium*, Vancouver, 27–31 May 2007. doi:10.1201/NOE0415444019-c112
- Zischinsky U (1969) Über Sackungen. *Rock Mech* 1:30–52

Seismic Strengthening Strategies for Existing (Code-Deficient) Ordinary Structures

Kosmas-Athanasios Stylianidis
Civil Engineering Department, Aristotle
University of Thessaloniki, Thessaloniki, Greece

Synonyms

Intervention procedures to upgrade existing ordinary structures which are designed and executed according to older codes

Introduction

Scope of the Entry

This entry presents basic principles, criteria, and procedures in regard to the conceptual design of seismic strengthening of existing ordinary structures and its treatment in current codes. In this context, certain code provisions constitute a significant part of the present entry.

Although it can be read by nonexperts, this entry is mainly addressed to structural engineers involved in the design of structures and interventions in buildings.

Notes on Terminology Used

The terms used in the entry are standard and familiar to engineers; nevertheless some clarifications are in order:

Seismic: The entry is restricted to seismic upgrading only. Certain aspects of the methodology may also apply to upgrading against other actions. However, it is pointed out that the main characteristic of a seismic strengthening strategy is the fact that an earthquake is an accidental load. In this sense, it is possible to define target performance levels, dependent on the level of the design earthquake, which in general can be related to the accepted risk. This probability leads to a compromise between “desirable” and “feasible,” i.e.,

instead of a practically unfeasible (due to cost or disturbance of use) high target, the adoption of a lower, yet feasible, target is more pragmatic. This option is typically not available in the case of upgrading against permanent loads.

Intervention: The term structural intervention implies any operation that results in the desired modification of existing mechanical characteristics of a member or a structure and has as a consequence the modification of its response.

Repair: The term implies the process of intervention to a structure damaged by any cause that reinstates the mechanical characteristics of its structural members to their pre-damage level and restores its original structural capacity.

Strengthening: The term implies the process of intervention to a structure with or without damage, which increases the capacity and/or ductility of a member or the entire structure to a level higher than that prescribed in the original design.

Strategy: It is a high-level plan to achieve certain objectives, usually under uncertain conditions.

Technique: It is a procedure to be implemented to complete a task (see also Wikipedia 2014).

Code Deficient: The term refers to buildings designed according to older seismic codes or without any code. It is typically the case that contemporary codes introduce significantly higher demands than older codes and this is the main reason that upgrading is needed.

Ordinary Structures: This term includes mostly above surface R/C building structures. Underground, special structures (bridges, chimneys, etc.) and historical buildings fall beyond the scope of this entry. The same principles can be applied to structures made of different structural materials, such as steel, masonry, and wood, after suitable adaptation.

Sources

The entry is primarily based on two sources, the European Standard EN 1998–3 (2005) and the Greek Code of Structural Interventions (2012), which standardize the existing knowledge on upgrading of ordinary structures.

Assessment

General

The assessment of an existing structure is not strictly mandatory, but it is valuable, since it reveals the weaknesses of the structural system and indicates the strengthening measures to be taken. In this sense, the assessment is integrated in the design strategy and is strongly recommended.

The assessment of existing structures follows the steps below:

- Collection of data (investigation of structural history)
- Analysis (see Section [Methods of Analysis for Assessment and Redesign](#) of the present entry)
- Verification against limit states (see [Section Compliance Criteria for Assessment and Redesign](#) of the present entry)

Scope

- (a) The purpose of the assessment of an existing structure is the evaluation of its available bearing capacity and the verification that the minimum mandatory requirements imposed by the existing codes are met.
- (b) To estimate the available bearing capacity of the structure, the data from the structural history survey should be taken into account (see Greek Code of Structural Interventions (2012), Chap. 3, and European Standard EN 1998-3 (2005), Chap. 3).
- (c) The designer should prescribe and supervise a series of investigating works in order to document and justify the assumptions on which the assessment will be based.
- (d) The process of assessment differs depending on the existence or otherwise of damage in the building assessed.
- (e) In case of no damage, the result of the assessment, depending on the selected redesign objective, will dictate the decision for potential retrofit.
- (f) In the presence of damage, due to any past actions (prescribed or not by the standards applicable up to then), the assessment process includes two phases:
 - (i) First, the structure is assessed as it is, taking damage into account. Depending on

the selected redesign objective, the result of the assessment might lead to a decision for intervention (repair and/or retrofit).

- (ii) In case that intervention is required, the structure is assessed at its pristine condition, i.e., assuming that damage will be repaired. Depending on the selected redesign objective, the result of this assessment will lead to the decision for repair only or for repair and retrofit.

Collection of Data

The collection of the data required for the assessment should be governed by the following principles:

- (a) The data required for the assessment of the bearing capacity of existing structures should be cross-verified and properly calibrated, wherever possible.
- (b) It is recommended that the program of field and laboratory investigations be designed, executed, and supervised by the designer of the assessment, according to the specific design requirements.
- (c) The reliability of the data collected should be properly taken into account in the assessment of the existing structure and the development of the intervention strategy.

Data Reliability Level (DRL)

General

- (a) The reliability level of data (DRL) related to actions or resistances indicates the adequacy of the information regarding the existing building and is taken into account in the assessment and redesign. In existing structures, the numerical values of the data required for the assessment and redesign are usually characterized by a higher uncertainty than in the case of new structures.
- (b) DRL is not necessarily the same for the entire building. Individual DRLs for the various subcategories of information can be determined. For the selection of methods of analysis, the most unfavorable among the individual DRLs are used.

- (c) The concept of DRL is also applied in assessing the completeness of the survey of the structure and infill walls, especially in case of hidden elements. The effects of uncertainties can be taken into account in actions or resistances depending on the case (e.g., uncertainty in the thickness of the flooring of the slab will be taken into account in actions; uncertainty in the thickness of the slab itself will be considered mainly in the resistances).

Description of DRLs

- (a) According to the European Standard EN 1998-3 (2005), three knowledge levels of data are adopted: (KL) 1–3 (limited, normal, full). According to the Greek Code of Structural Interventions (2012), Chap. 3, these levels correspond to high, satisfactory, and tolerable data reliability level.
- (b) Regarding the self-weight, the characteristic value considered must be the most unfavorable value that is consistent with the geometry of the structure and/or applies for such structures.
- (c) Regarding the resistances, their values can be determined from the dimensions, reinforcement, and material characteristics that lead to the justification of prior behavior of the structure. So, for example, a strength value that corresponds to the ultimate resistance of a cross section under the existing acting loads can be used. Similarly, dimensions of inaccessible foundations can be estimated with the assumption that they correspond to the ultimate soil bearing capacity under acting loads, etc.

Impact of DRL on the Assessment and Redesign Depending on the reliability of the data:

- (a) The appropriate safety factors γ_f for certain actions with uncertain values are selected (see Greek Code of Structural Interventions 2012, Chap. 4). Such may be the case for the representative values of some indirect actions (e.g., soil pressure) and the weight of inaccessible infill walls or coating/plastering.

In certain cases of high uncertainty (and if it is considered that the influence of the magnitude of the corresponding action is significant), consideration of two “reasonably extreme” representative values ($S_{k, \min}$ and $S_{k, \max}$) is recommended.

- (b) The appropriate safety factors γ_m are selected according to the data for existing materials (see also Greek Code of Structural Interventions 2012, Chap. 4). Material data are the dimensions of members and strengths of concrete and reinforcing steel, as well as the actual reinforcement detailing, anchoring, starter bars etc. that determine the resistances.

Assessment Principles

General

Assessment of existing structures follows the principles listed below:

- (a) When the existing structural system is expected to contribute to the redesigned structural system solely by resisting vertical loads, its assessment may be performed based on simple, yet conservative, methods.

In this case, the accuracy of the assessment method used should be adjusted to the desired goal. For instance, an approximate, yet conservative, assessment method is sufficient to demonstrate the adequacy of the existing structural system against vertical loads. When the existing structural system is apparently inadequate and is expected to be fully demolished, its assessment is not necessary.

- (b) However, when the existing structural system is expected to contribute to the redesigned structural system by resisting both vertical and seismic loads, it should be assessed based on the principles below. For the assessment of the structure against vertical loads, it is possible to use the methods prescribed by Eurocode 2 (EN 1992-1-1 2004), appropriately adapted:

- (i) The assessment is made using analytical methods (see Section [Methods of Analysis for Assessment and Redesign](#)

of the present entry). Especially in the case of structures without damage, for which the approved design study (which has been applied) is available, the assessment could be based on it.

- (ii) The numerical models to be used for the assessment may represent the entire structure or individual members. Different numerical models may be used, depending on the type of the imposed actions. In general, the types of numerical models should be selected consistently with the calculation methods to be applied. Advanced methods of analysis should be combined with detailed numerical models.
- (iii) It is recommended that the accuracy of the methods used be compatible with the accuracy of the data. High-quality numerical analysis is reliable only if it is based on equally reliable data.
- (iv) The use of empirical–analytical or purely empirical methods is allowed only in cases covered by relevant special provisions issued by the competent public authority. Such specific provisions can be issued, provided that they refer to a building stock with common, known, features and that they are always preceded by a proper investigation.
- (v) In cases of structures that have already suffered damage or deterioration, the applied assessment method must be able to interpret, as a rough approximation, both the type and the location of this significant damage. In structures of high importance, where damage has been identified, parametric analyses may be required to achieve the interpretation of damage based on their type and location.

The possible interpretation of damage, in terms of type and location, constitutes an acceptance criterion for the analytical methods used. Possible parameters may include non-visible geometrical data, mechanical characteristics that have not been investigated, random

combinations of actions assumed to have been applied in the past, etc.

- (vi) For analysis, limit state control, verification of the adopted behavior factor, control of the imposed displacements, and local ductility indices, Sections [Methods of Analysis for Assessment and Redesign](#) and [Compliance Criteria for Assessment and Redesign](#) of the present entry are applied. In the case of masonry walls, paragraph [Consideration of Masonry Infill Walls](#) is applied.
- (vii) In many cases, a quick assessment of the loss of bearing capacity of a damaged or deteriorated structure may be useful and/or necessary. This estimate can be made based on the intensity and extent of damage, as derived according to well-established (refined or approximate) methods.

Consideration of Masonry Infill Walls

- (a) It is not permitted to consider masonry infill walls as part of the system that carries non-seismic actions (such as gravity loads). To calculate the internal forces of the structure due to non-seismic actions, numerical models are used that do not include masonry infill walls or do not impose stresses (induced by vertical loads) to them.
- (b) It is recommended to consider masonry infill walls as part of the seismic action resisting system.
- (c) It is mandatory to consider masonry infill walls as part of the system resisting seismic actions, when this decision has an adverse effect on the results obtained for the structural system at a global or local level.

The assessment of whether the influence of infill walls is favorable or unfavorable has to be made by the designer; however, the difficulty of the assessment has to be noted, particularly in the case that analysis data and calculations are not available. As a result, the above assessment will be on the safety side, if the masonry infills are introduced in advance to the numerical analysis models.

Redesign

General

The redesign of existing structures follows these steps:

- Conceptual and preliminary design
- Analysis (see Section [Methods of Analysis for Assessment and Redesign](#) of the present entry)
- Verification against limit states (see Section [Compliance Criteria for Assessment and Redesign](#) of the present entry)

Conceptual and Preliminary Design

- (a) According to the criteria and the types of structural interventions, to be presented in Section [Selection Criteria and Types of Structural Interventions](#), an intervention strategy is devised and the type and extent of interventions is decided.

Decisions on the appropriate strategy and the subsequent type of interventions for each case should be made on the basis of all information obtained during the assessment stage of the existing structure. The perception of the overall behavior of the building and the identification of its weaknesses, such as inadequate strength or stiffness or ductility, the unfavorable structural system, individual weaknesses, etc., should be dominant in the decision-making process.

Regardless of the analysis method that will be eventually adopted for the redesigned structure, inelastic static analysis may provide substantial assistance in identifying these weaknesses. Furthermore, with the aid of the above method, it is feasible to preliminarily select the characteristics of the types of intervention that will be prioritized.

- (b) In all cases this selection should be justified (compared with other possible options), while the anticipated post-intervention behavior of the building should also be described qualitatively.
- (c) A preliminary estimate of the dimensions and strength of the materials used and the modified stiffness of the structural elements undergoing intervention should be made.

- (d) Preliminary estimate is made of the ductility class that the structure will fall into after the intervention, or (in case of application of inelastic static analysis) preliminary estimate is made of either the amplitude of the target displacement or the tolerable rotation angle of all structural members after intervention.

Assessment and Redesign Objectives

General

- (a) In order to satisfy broader socioeconomic needs, various “performance levels” (target behaviors) are stipulated under relevant prescribed design earthquakes.
- (b) The objectives of the assessment or redesign (Table 1; see Greek Code of Structural Interventions 2012, Chap. 2.2) are combinations of both a performance level and a seismic action, given an “acceptable probability of exceedance within the technical life cycle of the building” (design earthquake). The European Standard EN 1998–3 (2005) specifies that these levels should be defined in the National Annex (for each country).
- (c) In the present entry, objectives that refer solely to the structural system are prescribed. In contrast, no objectives are set for the nonstructural system.

The relevant fundamental requirements, set by Chap. 2.1 of European Standard EN 1998–3 (2005), are satisfied through Table 1. In the case of two reassessment objectives, the possible pairs are B1 and A2 or C1 and B2.

The term “structural system” is used here in the classical sense and corresponds to the system bearing vertical loads. Accordingly, the term “nonstructural system” corresponds to the system that does not participate in bearing vertical loads. It is noted that the above conditions are not associated with the terms “primary” and “secondary” structural elements that are used in subsequent paragraphs.

- (d) The objectives of the assessment or redesign are not necessarily identical. The objectives of redesign may be higher than those of the assessment.

Seismic Strengthening Strategies for Existing (Code-Deficient) Ordinary Structures, Table 1 Assessment or redesign objectives of the structure

Probability of exceedance of seismic action within a conventional life cycle of 50 years (%)	Performance level		
	Damage limitation (DL)	Significant damage (SD)	Near collapse (NC)
10	A1	1	C1
50	A2	2	C2

- (e) The minimum acceptable assessment or redesign objectives for the structural system of existing buildings are defined ad hoc by the public authority. In special cases, the public authority may designate additional objectives for assessment or redesign of the nonbearing system as well. In this case, the same authority also defines the criteria for meeting the respective objectives.
- (f) In any case, the reassessment objective (assessment or redesign) is chosen by the project owner, provided that it is equal to or higher than the above minimum acceptable objectives. In defining these objectives, the following criteria (among others) should be taken into account:
- Importance of the building (e.g., temporary structure, ordinary residential house, area of public gathering, area of crisis management, high-risk facility)
 - Available financial resources during the given period
- (g) The owner of the project or the public authority has to define the time frame within which the relevant interventions will be conducted, wherever required.
- (h) A nominal life cycle equal to the conventional lifetime of 50 years is generally accepted, regardless of the estimated “actual” remaining life of the building. An exception to this rule is permitted only under special circumstances where the remaining lifetime is fully guaranteed, based on the judgment and approval of the public authority; in such a case, the seismic actions are modified accordingly.

It is noted that according to Table 1, design objective B1 is the objective normally set for new structures.

The adoption of an assessment or redesign objective associated with a probability of exceedance of the seismic action of 50 % will generally lead to more frequent, more extensive, and more severe damage compared to a corresponding objective associated with a probability of exceedance of seismic action equal to 10 %.

The probability of exceedance of 50 % (maximum tolerable) in 50 years corresponds to an average return period of about 70 years, while a probability of exceedance of 10 % in 50 years corresponds to an average return period of approximately 475 years.

Structural Performance Levels

The performance levels of the structure are defined as follows [see also Chap. 2.1 of European Standard EN 1998–3 (2005)]:

- (a) “Damage Limitation” (A) is a condition wherein the structure is only lightly damaged, with reinforced concrete elements prevented from significant yielding and retaining their strength and stiffness properties. Nonstructural components, such as partitions and infills, may show distributed cracking, but the damage could be economically repaired. Permanent drifts are negligible. The structure does not need any repair measures. No building operation is interrupted during and after the design earthquake, with the possible exception of minor importance functions. A few hairline cracks may occur in the structure.
- (b) “Significant Damage” (B) is a condition wherein the structure is significantly damaged, with some residual lateral strength and stiffness, and vertical elements are capable of sustaining vertical loads. Nonstructural components are damaged, although partitions and infills have not failed out of plane. Moderate permanent drifts are present. The structure can sustain aftershocks of moderate intensity. The damage to the structure, expected to occur during the design earthquake, is

repairable but possibly uneconomic, without causing loss or serious injury of people.

- (c) “Near Collapse” (C) is a condition wherein the structure is heavily damaged, with low residual lateral strength and stiffness, although vertical elements are still capable of sustaining vertical loads. Most nonstructural components have collapsed. Large permanent drifts are present. The structure is near collapse and would probably not survive another earthquake, even of moderate intensity. The damage to the structure is in general unrepairable. Injuries of certain individuals due to structural damage or falling elements of the nonbearing structure or other objects are not excluded.

The term “unrepairable damage” refers to serious or severe damage, for which strengthening (and not just repair) or replacement or substitution of the component or the entire structure is required.

General Principles for Pre- and Post-earthquake Intervention Decisions

Selection Criteria and Types of Structural Interventions

- (a) Based on the conclusions drawn during the assessment of the structure, and the nature, extent, and intensity of the damage or deterioration (if any), intervention-related decisions are made, aiming to (a) meet the basic requirements of the seismic code, (b) minimize the cost, and (c) serve the needs of the society.
- (b) The selection of the type of structural intervention should be made primarily on the basis of general cost- and time-related criteria, the availability of the resources required, architectural or other needs, etc. In this selection, the financial (or other) value of the structure should also be taken into consideration, both prior and subsequent to the intervention.

Such general criteria include the following:

- The cost, both initial and long term (i.e., the cost of maintenance and possible future damage or deterioration), compared

to the importance and age of the building under consideration.

- The available quality of the work (it is very important that intervention measures be compatible with available resources and available quality of work).
 - The availability of an adequate quality control system.
 - The use of the building (possible consequences of the intervention works to the use of the building).
 - The design, from an aesthetic point of view (the intervention scheme may vary between a fully invisible intervention and a deliberately distinctive set of new or added members).
 - The conservation of the architectural identity and integrity of historical buildings and the consideration of the degree of reversibility of the interventions.
 - The duration of works.
- (c) The selection of the type, technique, scale, and timing of the intervention should be based on technical criteria related to the observed current state of the building, as well as to a provision to maximize the ability of the structure to absorb seismic energy (ductility) after the intervention.

Such technical criteria include the following:

- All identified serious deficiencies must be remedied.
- All identified serious damage (and deterioration) in primary structural members must be repaired properly.
- In the case of highly irregular buildings (mainly in terms of distribution of their overstrength, both in plan and in height of the building), structural regularity should be improved to the maximum possible extent.
- All resistance requirements in critical regions of primary structural members (i.e., the required resistance and plastic deformation capacity) must be satisfied after the intervention (distinguishing between primary and secondary members).
- Where possible, the increase of local ductility in critical regions should be pursued.

Particular provision should be taken, to the largest extent possible, so that the local repair and strengthening do not diversely affect the available ductility within the critical region.

- In special cases, the durability of both new and original structural members and the potential acceleration of the deterioration should be taken into consideration.

Types of Intervention and Their Consequences

- (a) Based on the foregoing criteria and the results of the assessment of the structure, appropriate types of intervention should be ad hoc selected for individual structural members or the entire building and the nonstructural system (if required), always taking into account the side effect of the interventions on the foundations. This selection is part of an intervention strategy, which aims to improve the seismic behavior of the building by modifying or controlling the basic parameters that affect its seismic behavior. In order to achieve a reduction of seismic risk, strategies of technical or managerial nature, or combination of the two, can be adopted.

A number of technical and managerial strategies are indicatively given herein:

Technical strategies:

- Enhancement of the building strength
- Enhancement of the building stiffness
- Enhancement of the deformation capacity of the structural members
- Reduction of seismic demand

Managerial strategies:

- Limitation or change of use of the building.
- Partial or global demolition (i.e., of a number of storeys).
- “Rigid body” transfer of the entire structure to another location.
- Decision for “no intervention”. In such a case, a reduction of the nominal life cycle of the structure can be accepted, under the condition that upon expiry of this period, demolition of the structure is guaranteed.

Some types of interventions in structural elements associated with specific strengthening strategies of technical nature are listed below:

- The enhancement of strength and stiffness is alternatively achieved by selective or large-scale strengthening of structural members or by the addition of new elements that can resist either partially or fully the seismic actions (e.g., reinforced concrete shear walls, steel trusses, infill walls, etc.). In this case, particular attention should be given to the design of the foundation due to the increase of both the structural mass and the seismic loads.
- The enhancement of post-elastic deformation capacity (ductility) is achieved by improving the confinement of existing members, e.g., with external connectors, strips of steel or fiber-reinforced polymers, etc.
- The remedy of critical deficiencies refers to alleviating those features that lead to unfavorable seismic behavior. Indicatively:
 - Modification of the structural system (abolition of certain expansion joints, replacement of sensitive members, alteration actions aiming at a more regular and ductile system)
 - Addition of special links to connect the brittle masonry and surrounding member, whenever this is permitted by the strength of masonry
 - Local or global modification of members with or without damage
 - Full replacement of inadequate members or members that have suffered extensive damage
 - Redistribution of demand (e.g., through external prestressing)
- The reduction of seismic demand is achieved by reducing the mass of the structure and modifying the structural system toward a favorable shift of the fundamental period of the structure [e.g., through seismic isolation systems or absorption of seismic energy; see

Chap. 10 of European Standard EN 1998–1 (2004)].

- (b) In cases where, for the redesign objective set, the seismic behavior of nonbearing structural members might endanger the lives of the occupants (or third persons) or might have consequences to stored goods, measures should be taken to repair or strengthen the particular members.

In such cases, local or global collapse should be prevented by:

- Appropriate links to the structural members or by taking supportive measures to prevent possible fallout of parts of those members
 - The improvement of the mechanical characteristics of nonstructural members
- (c) The potential impact of repairs and strengthening of nonstructural members should be taken into account.
 - (d) The side effects of all structural interventions on the local and global capacity of the building to absorb seismic energy should be taken into account.

The enhancement of strength usually leads to a reduction in ductility, unless special measures are taken (e.g., in reinforced concrete elements, the increase of the tensile reinforcement should be in principle accompanied by a sufficient increase of the compression reinforcement and of confinement).

Methods of Analysis for Assessment and Redesign

General

- (a) The action effects and/or the required plastic rotations of all structural members of the building under the design earthquake and other combinations of actions are derived using appropriate analysis methods.

To determine the internal forces and displacements, it is permissible to ignore proximity to other buildings.

- (b) The selection of the appropriate method of analysis should be based on the importance of the building and its potential damage or

deterioration, as well as on the available data regarding the cross sections and strength of its structural members.

Whenever possible, it is recommended to calibrate such methods through comparison with the behavior of buildings that have already been studied with the particular methods.

- (c) Where appropriate, additional partial factors γ_{sd} will be applied to account for the uncertainties related to the numerical analysis models.

Consideration of Masonry Infill Walls

Consideration of the masonry infill walls in the redesigned structure can be carried out, subject to the conditions of paragraph Consideration of Masonry Infill Walls

As part of the redesign process, it is desirable to make every effort to mitigate the potential deficiencies imposed by the masonry infills. Addition or upgrading of masonry infills can be used for the improvement and strengthening of existing buildings, subject to conditions of paragraph Consideration of Masonry Infill Walls

Methods of Analysis

For the assessment and redesign of a building, one of the following analysis methods may be used. The field of application of each analysis method depends on satisfying a number of conditions, primarily related to regularity (see Greek Code of Structural Interventions 2012, Chap. 5):

- (a) Elastic (equivalent) static analysis with global (q) or local (m) behavior (or ductility) factors, regardless of the data reliability level.
- (b) Elastic dynamic analysis with global (q) or local (m) behavior (or ductility) factors, regardless of the data reliability level.
- (c) Inelastic static analysis. In this case, it is recommended to ensure, as a minimum, a “satisfactory” data reliability level.
- (d) Inelastic dynamic (response history) analysis. In this case, it is again recommended to ensure, as a minimum, a “satisfactory” data reliability level.

- (e) In special cases, solely for the assessment of existing buildings, it is permitted to estimate the demand approximately, without detailed analysis involving a finite element model of the entire building.
- (f) Apart from the above analytical methods, solely for the assessment of existing buildings, in special cases and for specific objectives, it is possible to use empirical methods.
- (g) It is permitted to apply the elastic methods provided that all the following conditions apply:

- (i) The failure index (λ) of each primary member is in general lower than 2.5.

The adopted threshold value of the failure index (λ) generally implies that the available strength of each primary structural member is at least 40 % of the demand resulting from an elastic seismic analysis without reducing the seismic action, i.e., for $q = 1$.

- (ii) The average failure index ($\bar{\lambda}_k$) in each storey does not exceed 1.50 times the average failure index of the storey above or below.

It is deemed that the average failure index ($\bar{\lambda}_k$) identifies the regularity in the resistance (strength) along the building height, whereas its adopted threshold value ensures that no weak, in flexure and/or shear, intermediate storey exists.

- (iii) The failure index (λ) of each primary structural member that is located on one side of the building, for a given direction of seismic action, does not exceed 1.50 times the average failure index (λ) of a primary member that is located on any other side of the same storey.

It is deemed that with this provision, issues of torsionally sensitive storeys are addressed.

Principal (or Primary) and Secondary Structural Members

The individual parts of the structure of a building and the individual structural elements (members) affecting the stiffness and demand distribution within the building, or the members that are

loaded due to lateral displacements of the building, can be distinguished during assessment or redesign into “principal” (or “primary”) and “secondary.”

In general, the structural members or substructures that contribute to the strength and stability of a building under seismic loading will be characterized as principal. The remaining structural elements or substructures will be characterized as secondary. See also related Chap. 4.2.2 of European Standard EN 1998–1 (2004).

The main consequence of classifying a structural member (or substructure) as a secondary is that for these members, different performance criteria apply, i.e., they are permitted to undergo larger displacements and suffer higher damage compared to the primary elements.

In cases where damage limitation after the earthquake has been set as the assessment or redesign objective, the above distinction between primary and secondary data is not permitted.

For the masonry infill walls, which do not carry vertical loads (see paragraph Consideration of Masonry Infill Walls), the distinction between primary and secondary members does not apply. Where those members are considered as part of the seismic action resisting system, they are addressed and verified separately.

Compliance Criteria for Assessment and Redesign

General

- (a) Compliance with the selected performance level for assessment and redesign is achieved by adoption of the seismic action, method of analysis, verification, and detailing procedures, appropriate for the different structural materials (concrete, steel, masonry) within the scope of the code used.
- (b) Except when using the q-factor approach, compliance is checked by making use of the seismic action as defined in Table 1.
- (c) For the verification of structural elements, a distinction is made between “ductile” and “brittle” ones. Except when using the q-factor approach, the former is verified by checking

that demands do not exceed the corresponding capacities in terms of deformations. The latter is verified by checking that demands do not exceed the corresponding capacities in terms of strengths (see Greek Code of Structural Interventions (2012), Chap. 9).

- (d) Alternatively, a q -factor approach may be used, where use is made of a seismic action (response spectrum) reduced by a q -factor. In safety verifications all structural elements are verified by checking that demands due to the reduced seismic action do not exceed the corresponding capacities in terms of strength.
- (e) For the estimation of the capacities of ductile or brittle elements, where these will be compared with demands for safety verifications, mean value properties of the existing materials are used as directly obtained from in situ tests and from additional sources of information. Nominal properties are used for new or added materials.
- (f) Some of the existing structural elements can be designated as “secondary seismic” (see Chapter Principal (or Primary) and Secondary Structural Members), in accordance with the definitions in European Standard EN 1998–1 (2004), Chap. 4.2.2 (1)P, (2) and (3). “Secondary seismic” elements have to be verified according to the same compliance criteria as primary seismic ones, but using less conservative estimates of their capacity than for elements considered as “primary seismic.”
- (b) The partial safety factors of the existing and added materials should take into account the geometrical uncertainties, the dispersion of material properties, the relevant information available on site, as well as any uncertainties due to the nature of works and the difficulties of effective quality control (see Greek Code of Structural Interventions 2012, Chaps. 4, 7–9).
- (c) Where appropriate, additional factors γ_{Rd} are applied to account for the uncertainties arising from the numerical modeling of the resistance in critical (or noncritical) regions.
- (d) In cases of structural interventions against seismic actions, the damage limitation verification is made in accordance with appropriate provisions (see Greek Code of Structural Interventions 2012, Chap. 9). The damage limitation verification generally includes the primary and secondary structural members, infills, and appendages.

Safety Verifications

- (a) The available resistance in the critical regions of all structural members (i.e., the resistance quantities and/or the tolerable plastic rotations) should be calculated on the basis of rational numerical models, which are widely accepted by the international scientific community, especially in terms of force transfer between existing and added materials or members (see Greek Code of Structural Interventions (2012), Chap. 6 for the numerical models, Chap. 7 for the determination of the behavior of structural members, and Chap. 8 for the design of the interventions).
- (b) The local available ductility in critical regions
- (c) The available secondary resistance mechanisms at large relative displacements
- (d) The potential consequences of the brittleness of a limited number of structural members on the ductility of the entire structure

Verification of the Adopted Behavior Factor

After the verifications above, it is required to approximately reevaluate the predefined behavior factor for the repaired–strengthened building, taking into account all the criteria favoring energy absorption, such as:

- (a) The sequence of failure of horizontal and vertical structural members
- (b) The type of failure in critical regions of each structural member (i.e., the ratio of the ultimate shear force to the effective shear at the time of flexural failure, as imposed by capacity design)
- (c) The local available ductility in critical regions
- (d) The available secondary resistance mechanisms at large relative displacements
- (e) The potential consequences of the brittleness of a limited number of structural members on the ductility of the entire structure

It is pointed out that in existing structures the requirements of capacity design, limitation of the axial force, local confinement, etc. have not been in general met. The implication of this is the difficulty in assessing a global behavior factor.

Summary

This entry deals with the seismic strengthening strategies for existing (code-deficient) ordinary structures. Reference to two regulatory texts that are compatible with each other, namely, the European Standard EN 1998–3 (2005) and the Greek Code of Structural Interventions (2012), has been selected as a basis for the presentation of the issue. These texts condense existing knowledge concerning the upgrading of ordinary structures and at the same time determine the procedure of the normative treatment of the issue.

First, clarifications on basic terms are given. Then, the general principles that rule the two aspects on which the issue of upgrading is based, i.e., assessment and redesign, are analyzed. Subsequently, the assessment and redesign objectives, which in general differ from those of the design of new structures, are presented.

The possibility of a variety of redesign objectives is introduced. High targets are proposed for a small, yet critical in terms of safety, group of buildings. On the contrary, there is a possibility, under certain conditions, to adopt lower targets for common existing buildings, in comparison to the targets for new buildings, with a view to a realistic treatment of the need for upgrading.

Furthermore, the general principles for pre- and post-earthquake intervention decisions are set, the methods of analysis and the conditions of their application are briefly presented, and, finally, the compliance criteria for assessment and redesign are described.

- [Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions](#)
- [Seismic Analysis of Steel–Concrete Composite Buildings: Numerical Modeling](#)
- [Strengthening Techniques: Code-Deficient R/C Buildings](#)

References

- Greek Earthquake Planning and Protection Organisation – EPPO (2012) Greek code of structural interventions – harmonized text http://ecpfe.oasp.gr/sites/default/files/files/%CE%9A%CE%91%CE%9D%CE%95%CE%A0%CE%95_EN2013_FINAL.pdf
- European Standard EN 1992-1-1 (2004) Design of concrete structures – general rules and rules for buildings
- European Standard EN 1998-1 (2004) Design of structures for earthquake resistance – general rules, seismic actions and rules for buildings
- European Standard EN 1998-3 (2005) Design of structures for earthquake resistance – assessment and retrofitting of buildings
- Wikipedia, the free Encyclopedia (2014) Terms “strategy” and “technique”

Seismic Strengthening Strategies for Heritage Structures

Dina D’Ayala¹ and Sara Paganoni²

¹Department of Civil, Environmental and Geomatic Engineering, University College London, London, UK

²ZiegertlRoswaglSeiler Architekten Ingenieure, Berlin, Germany

Cross-References

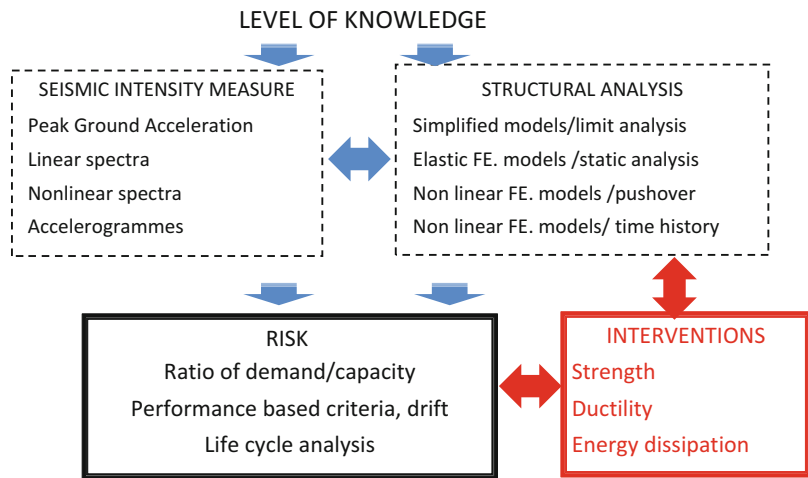
- [Assessment of Existing Structures Using Inelastic Static Analysis](#)
- [Behavior Factor and Ductility](#)
- [Equivalent Static Analysis of Structures Subjected to Seismic Actions](#)
- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- [Reinforced Concrete Structures in Earthquake-Resistant Construction](#)

Introduction

The global seismic behavior of historic masonry buildings is highly influenced by the integrity of the connections among vertical and horizontal structural elements, to ensure the so-called box behavior. This, providing the transfer of inertial and dynamic actions from elements working in flexure out-of-plane to elements working in in-plane shear, leads to a global response best suited to the strength capacity of the constitutive

Seismic Strengthening Strategies for Heritage Structures,

Fig. 1 Correlation between knowledge level, analytical tools, risk representation, and interventions. credits: this publication D'Ayala Paganoni 2014



materials and hence enhanced performance and lower damage level. Notwithstanding the importance of connections' integrity, analytical checks of existing connecting elements, or design of new elements to strengthen existing connections, are generally based on qualitative rules or simplified overall checks, rather than rigorous analytical approach. The "Guidelines for Earthquake Resistant Non-Engineered Construction" were first published by the International Association for Earthquake Engineering (IAEE) in 1986 specifically with the objective of improving the seismic safety of non-engineered housing constructions.

A wide range of techniques and products for the seismic strengthening of heritage buildings are reported in the scientific literature and used in current practice to ensure the enhancement of existing connections. Heritage buildings require far more attention, especially when dealing with issues such as the compatibility between the chemical and mechanical properties of the strengthening system and the parent material. Many strengthening techniques, after an initial success and a strong commercial promotion, underperformed and showed unexpected drawbacks when put to the test of real seismic loading outside the controlled conditions of the laboratory environment. On the other hand, some existing strengthening systems can provide highly flexible applications and meet the expected requirements in terms of performance; some of these systems in fact draw on traditional

reinforcement techniques, with the addition of innovative materials and a deeper insight in the laws governing the dynamics of structures.

The choice of the most suitable strengthening system for a given building, however, is determined not only by the set of constraints of the specific project but, very importantly, by the framework used for its assessment. The EC8 part 3 or ASCE41-06, dealing with assessment, repair, and strengthening, introduce the concept of knowledge level as the determining factor for the choice of alternative assessment procedures involving diverse levels of resources. The diagram in Fig. 1 shows that there is a strict correlation not just between the representation of the seismic input and the type of analysis that can be carried out with it but also and most importantly between the input and the way risk is quantified and measured and between such measures and the principle on which possible strengthening interventions operate. Hence the level of knowledge and availability of data to carry out the assessment becomes critical in determining and fully designing the most appropriate strengthening solution.

Moreover, although European and national codes (e.g., Italian Ministry of Cultural Heritage and Activities 2006) suggest the use of various systems for the strengthening of connections, for example, ring beams, no detailed reference is made to specific procedures for the dimensioning and checks of such elements. The only indication

in this sense can be found in Section 6.1 Retrofit Design Procedure for existing building of Eurocode 8 ([EN 1998-3:2005](#)), which states that the design process should cover:

1. Selection of techniques and/or materials, as well as of type and layout of intervention.
2. Preliminary sizing of additional structural parts.
3. Preliminary calculation of stiffness of strengthened elements.
4. Analysis of strengthened structure by linear or nonlinear analysis. The typology of analysis is chosen depending on the level of knowledge regarding the geometry detailing and materials of the structure.
5. Safety verifications for existing, modified, and new structural elements carried out by checking that the demand at three different limit states – damage limitation, significant damage, and near collapse – is lower than the structural capacity.

In the safety verifications, mean values of mechanical properties of existing materials derived from in situ tests and other sources of information, like available documentation or relevant sources, shall be used, taking into account the confidence factors (CFs) specified in 3.5 of Eurocode 8 ([EN 1998-3:2005](#)). Conversely, for new materials, nominal properties shall be used without modification by confidence factor. The code also states that in case the structural system, comprising both existing and new structural elements, can be made to fulfill the requirements of [EN1998-1-2004](#), the verifications may be carried out in accordance with the provisions therein.

This last sentence indicates that for systems such as reinforced concrete (RC) ring beams or corner confinement, reference can be made to the specifications for RC members in the relevant sections of EC8 and other Eurocodes. However, this leaves open the problem of quantifying the interaction between original and new structural elements, and the assessment of the global seismic performance of the strengthened structure will still be affected by a large number of uncertainties.

The last 15 years have seen a steep increase in the use of new technologies for strengthening heritage buildings. Such fast development, as well as the high level of expertise and financial resources required for their application, often results in lack of standardization. Innovative technologies have not been extensively applied and validated in real-life situations yet, and the retrofit of a complex, precious building by means of unconventional systems is a difficult task that goes beyond the standard conservation practice. In fact, looking at the current scientific literature, it is clear that many projects of restoration and upgrade of monumental buildings are carried out by research organizations within the framework of specific projects or by large enterprises that specialize in the production and design of strengthening devices.

On the other hand, the lack of appropriate standards and procedures can be blamed for the incorrect application of some innovative strengthening techniques, notwithstanding recognized effectiveness and other benefits. One exception to this situation is the case of fiber-reinforced polymers for which guidelines and standards have been produced for selected countries (see, for instance, [CNR-DT 200/04](#), [CNR-DT 200R/13](#)).

In the following an overview of methods to structurally strengthen the masonry elements of historic buildings is given, pointing out the advantages and pitfalls of each system and the fundamental physical parameters and design process required. Available design procedures are reviewed and referenced, when available, to the writers' knowledge. The systems reviewed are applicable mainly to stone or brick masonry structures with timber floors and roof or with vaulted structures. The concepts underlying these methods are also usually applied to the strengthening of earth structures, although some of the details of the specific applications might differ due to the specific characteristics of the parent material.

The effective strengthening of a heritage building also relies on a correct redistribution of the inertia forces among vertical elements. For this to occur, the floor structures need to be acting

as a whole and ensure diaphragm action. While there is a wealth of research and methods to assess and strengthen historic and traditional floor structures to ensure they develop such action, these are not currently included in this chapter.

Base isolation is also increasingly becoming a proposed option for the retrofitting of historic buildings; however, actual implementations are still very rare and very expensive. This technique is also not treated in this current edition.

In this current edition, only methods suggested in the Eurocodes or methods related to recent technological developments are included. Many traditional “vernacular” methods of strengthening historic building exist in earthquake-prone countries and some have been studied in detail. These are not addressed here either.

Strengthening systems are reviewed depending on the type of enhancement achieved, i.e., whether there is an increase in strength, a control of deformation and displacement, or a dissipation of energy.

Connections Between Vertical and Horizontal Macroelements

The Eurocode 8 states that to improve connection between intersecting walls, use should be made of cross-bonded bricks or stones. The connection can be made more effective in different ways (EN 1998-3:2005):

- I. Through construction of a reinforced concrete belt
- II. By addition of steel plates or meshes in the bed joints
- III. Through insertion of inclined steel bars in holes drilled in the masonry and grouting thereafter
- IV. Through post-tensioning

The addition of steel ties, along or transversely to the walls, external or within holes drilled in the walls, is an efficient means of connecting walls and improving the overall behavior of masonry buildings (EN 1998-3:2005).

In the following we review:

- Stiffening of the wall system by corner and T-junctions confinement of wall panels achieved by inserting vertical concrete columns or steel meshes, fiber-reinforced horizontal strips, and polypropylene meshes
- Stiffening of the wall and floor system and connection to the horizontal structures by ring beams
- Connection of the walls and floor system by anchorage systems
- Improvement of the performance of the wall system by including energy absorbers and dissipating devices to connect the walls

The design of effective connections between various structural elements of a masonry structures is a critical step for the achievement of a good structural response in case of seismic events. Current codes provide for carefully designed and detailed connections; for instance, the Eurocode states that: “Floor systems and the roof should be provided with in-plane stiffness and resistance and with effective connection to the vertical structural systems (EN 1998-1:2004).”

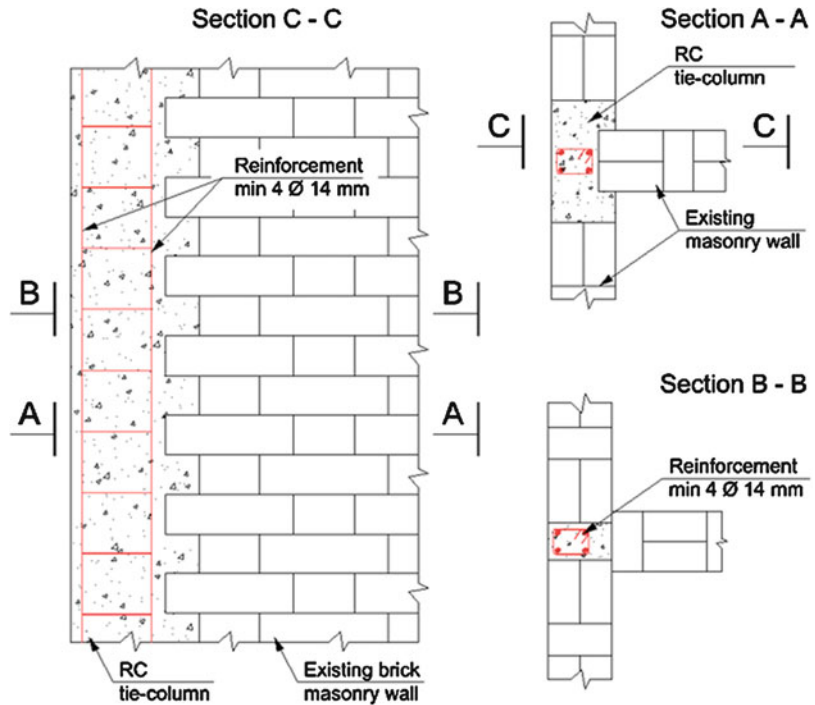
This rule is also applicable to interventions on existing structures as the Eurocode indeed states that: “The connection between the floors and walls shall be provided by steel ties or reinforced concrete ring beams (EN 1998-1:2004)” and “Specifically for masonry structures: non-ductile lintels should be replaced, inadequate connections between floor and walls should be improved, out-of-plane horizontal thrusts against walls should be eliminated (EN 1998-3:2005).”

Moreover, in relation to the connections between walls and floors, EN 1998-3:2005 states that: if existing tie-beams . . . are damaged, they should be repaired or rebuilt. If there are no tie-beams in the original building structure, such beams should be added.

Improvement of connections between vertical and horizontal structures can be achieved through a variety of methods, which are described in the following.

Seismic Strengthening Strategies for Heritage Structures,

Fig. 2 Example of column tie (From <http://www.know2do.org/Retrofit/BrickMasonry/RSBrickMasonryP.htm>)



Confinement at Walls Junctions: Strength Enhancement

Column Ties

By confining plain masonry walls with vertical elements placed at all corners and wall intersections, as well as along the vertical frame of large openings, the seismic performance of a masonry building is improved as a result of the enhanced integrity of the structural system (Tomažević, 1999). This effectively changes an unreinforced masonry structure in a confined masonry structure. The disruption is significant and can affect very substantially the heritage value of the building. Moreover, such a technique is suitable only when horizontal RC tie beams in the bearing walls at floor level (see this IS section “Confinement at Wall to Floor Junctions: Ring Beams”) and stiff, monolithic floor diaphragms are in place; otherwise, the effect of the vertical confinement is scarce, or null, as shown by analysis conducted by Karantoni and Fardis (1992). Creation of corner confinement by column ties is not recommended for stonework masonry (Tomažević 1999).

For the construction of the confinement elements, all the bricks in the intervention area are removed; the concrete of the existing tie beams is also removed to allow for the connection between existing and new reinforcement. Vertical rebars and stirrups are placed and concrete cast (Fig. 2).

Tie columns should be as thick as the wall where they are located, although in many cases of repair of existing buildings, they are of reduced dimensions due to on-site constraints. Furthermore, they may sometimes be replaced by sets of reinforcing bars placed in holes drilled in the masonry and connected to the surrounding substratum by stirrups.

The confinement prevents disintegration and improves ductility and energy dissipation of URM buildings but has limited effect on the ultimate load resistance (Chuxian et al. 1997). However, the real confinement effect mainly depends on the relative stiffness between the masonry wall and the surrounding resulting concrete frame. Before cracking, the confinement effect can in general be neglected (Chuxian et al. 1997; Karantoni and Fardis 1992). For very squat URM walls (geometrical aspect ratio

of 0.33 and double fixed boundary conditions), the confinement increased the cracking load by a factor of 1.27 and the ultimate lateral capacity by a factor of 1.2 (Chuxian et al. 1997). For walls with higher aspect ratio, the confinement increased the ultimate capacity by a factor of 1.5. In addition, the confinement improved the lateral deformations and energy dissipation by more than 50 %.

The dimensioning procedure for column ties can be derived from the prescriptions specific to confined masonry structures. The process can be summarized as follows:

- Calculation of the appropriate combination of static and seismic loads acting on the structural elements (EN 1991-1-1:2002 and EN 1998-1:2004)
- Dimensioning and verification of the column ties for both static and seismic loading following the prescriptions provided for confined masonry structures (EN 1996-1-1:2005 and EN 1998-1:2004)

Special care should be taken in realizing the connection between the column ties and the horizontal structures in order to ensure a monolithic behavior of the resulting RC frame.

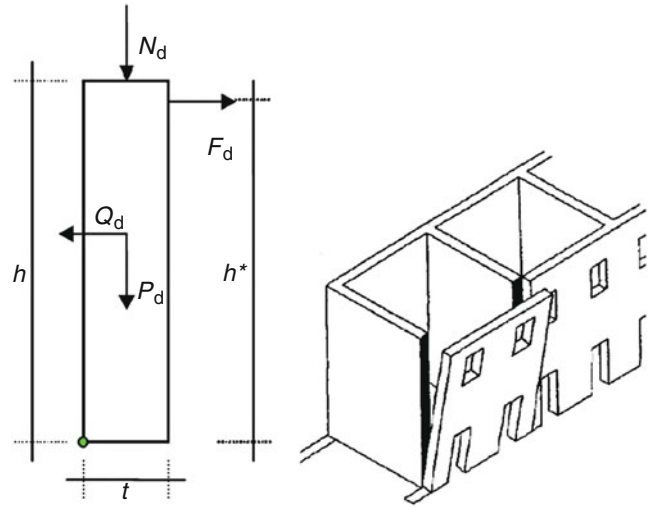
A much cheaper solution to confining masonry walls is achieved with Welded Wire Mesh (WWM). This consists in deploying a steel mesh on the two sides of a wall and around the corners, connecting it with bolts and then concrete over, by either shotcreting or other forms of plastering. The anchorage of the mesh to the foundation and horizontal structures is fundamental to achieve proper action transfer during shaking. Although less destructive than the construction of column ties, the system seriously affects the breathability of the walls and may cause deterioration due to moisture entrapment. The thickness of this reinforced plaster may vary between 25 and 100 mm depending on the regularity of the walls underneath. It cannot be applied in the presence of historic plasters or other valuable finishes. Its effectiveness highly depends on the relative stiffness between the original wall and the added layers.

Confinement by Fiber-Reinforced Plastic Strips

The application of fiber-reinforced polymer (FRP) material to the strengthening of masonry structures has to an extent mimicked application to RC structures. As a result, there are numerous examples in the literature relating to the confinement of masonry columns (e.g., Di Ludovico et al. 2010; Corradi et al. 2007). In the case of junctions between walls, horizontal strips of FRPs bonded at various levels along a masonry panel and anchored to the side walls can be used to restore corner connections, thus preventing overturning of façade walls. Optimal application is achieved when the whole perimeter of the structure is confined. Application of such techniques entails the removal of existing plaster and other superficial finish as a strong bond of the FRP strips to the masonry is necessary to ensure effectiveness of the system. The strips also need to be protected from UV radiations to prevent material deterioration. Their implementation might cause substantial loss of heritage value, and it is not recommended when valuable finishes are in place.

For the confinement to be effective, strips should be laid both horizontally and vertically. Tests by Hamoush et al. (2001) on walls strengthened with vertical and horizontal strips showed the effectiveness of the FRP systems as out-of-plane flexural strengthening elements, while Tumialan et al. (2001) have investigated the potential of near surface-mounted GFRP rods embedded into epoxy-based paste in the bed joints as shear enhancement. Cyclic testing of walls strengthened with vertical and horizontal GFRP or CFRP strips was performed by Marcari et al. (2003) and by Krevaikas and Triantafillou (2006). Those tests have highlighted a general decrease in strength and ductility of the confined members with the increase of the aspect ratio of the confined cross section. This is due to a reduction in ultimate strain at failure when increasing the aspect ratio. Glass fibers are more effective than carbon fibers, owing to lower stiffness. However, its effectiveness in connections between orthogonal walls has not been tested to the authors' knowledge. Issues of debonding should be thoroughly investigated on a case by

Seismic Strengthening Strategies for Heritage Structures, Fig. 3 Static scheme for the calculation of the tensile capacity of FRP strengthening to prevent overturning of front wall (CNR-DT 200/04)



case basis as failure mode and extent will be highly affected by the integrity and mechanical characteristics of the parent material external strata in relation to the composite characteristics.

FRP confinement is dimensioned by verifying:

1. The tensile capacity of the composite material
2. The capacity of the anchorage area

In detail this is done by (CNR-DT 200/04):

1. Checking that:

$$F_d \leq 2F_{Rd} = 2(A_f \cdot f_{fd}) \quad (1)$$

where

F_d is the design action applied to the FRP stripes by the front wall undergoing overturning caused by seismic equivalent action Q_d ; this is calculated by:

$$F_d = \frac{1}{2h^*} (Q_d h - N_d t - P_d t) \quad (2)$$

Dimensions and weight involved in the overturning mechanism are shown in Fig. 3:

F_{Rd} : design tensile capacity of the FRP strengthening

f_{fd} : design ultimate strength of the FRP strengthening

h^* : distance between level where FRPs are bonded and bottom hinge

A_f : FRP area

2. Checking the rip-off of FRP stripes from side walls:

$$F_d < 2F_{pd} = 2(A_f \cdot f_{pd}) \quad (3)$$

where

F_d is calculated as above.

F_{pd} is maximum anchorage capacity of FRP to one of the two side walls.

f_{pd} is design debonding strength of FRPs.

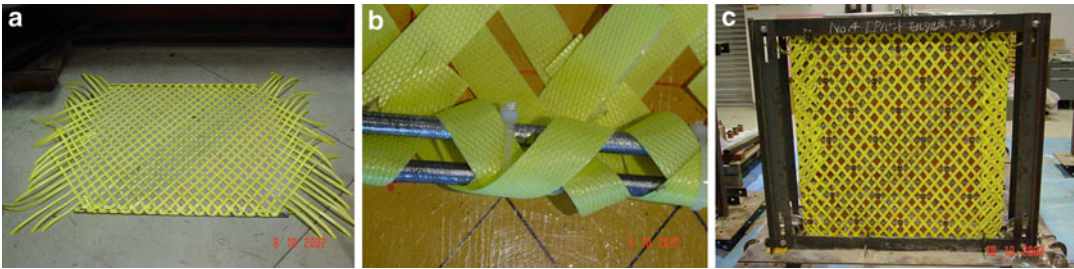
A_f is FRP area.

The second check is generally more demanding than the first, but it is not needed when full confinement of the structure is achieved by wrapping FRP stripes around the whole perimeter of the building and a suitable anchorage length is ensured (CNR-DT 200/04). Anchorage length should be at least 300 mm; otherwise, mechanical fixings can be used; however, the guideline does not specify suitable types and the procedure for sizing and positioning the fixings.

Additional calculations of the strengthened structure are necessary to ensure that the strengthened structure will be able to withstand other actions, such as the increased in-plane shear loading, now transferred from the walls



Seismic Strengthening Strategies for Heritage Structures, Fig. 4 Corner confinement of masonry walls by use of SRP strips anchored with SRP connectors (Photo Paolo Casadei 2008, product FIDSTEEL)



Seismic Strengthening Strategies for Heritage Structures, Fig. 5 Procedure for installation of PP meshing: (a) PP-band mesh, (b) detail of top/bottom

connection, and (c) retrofitted wall before application of final layer of mortar (Mayorca and Meguro 2004)

orthogonal to the walls parallel to the action. This should be done according to the specific indications of codes dealing with masonry structures in general.

Of course in order to prevent damage in the historic building, the capacity of the strengthening system needs to be less than the shearing capacity of any masonry course in the façade wall. Given the absence of ductility in the fiber composite, this system also presents the fundamental drawback of brittle failure, modest ductility, and relying on the damaging process of the masonry for energy dissipation. An improved version in this respect is provided by SRP strips, strips reinforced with steel, which show a greater ductility (see Fig. 4).

Confinement by Polypropylene (PP) Mesh

PP meshing uses common packaging straps (PP-bands) to form a mesh, which is then used to encase masonry walls, preventing both collapse and the fall of debris during earthquakes. PP-bands are commonly used for packaging and are therefore cheap and readily available while

the retrofitting technique is simple and suitable for local builders. PP meshing was developed by Meguro Lab, Tokyo University, and has had application in Kashmir and Nepal.

The retrofitting installation procedure is as follows (Mayorca and Meguro 2004):

1. The PP-bands are arranged in a mesh and connected at their crossing points (Fig. 5a).
2. Two steel rods are placed at the edges of the mesh; these bars are used to anchor the mesh at the foundation and at the top edge of the wall (Fig. 5b).
3. The walls are cleaned and, if possible, the paint is removed. Any loose brick is removed and replaced.
4. Six millimeter diameter holes are drilled through the wall at approximately 250–300 mm distance. The holes are cleaned with water spray or air.
5. The meshes are installed on both sides of the wall and wrapped around the corners and wall edges. An overlapping length of approximately 300 mm is needed.

6. Wire is passed through the holes and used to connect the meshes on both wall sides. In order to prevent the wires from cutting the PP-band mesh, a plastic element is placed between the band and the wire. Connectors are placed in proximity of the wall intersections and of the wall edges.
7. The top and bottom edges of the mesh are glued to the foundation and top of the wall by epoxy resin. The epoxy is used to connect the bars and the wall and it is not directly applied to the mesh. The bands, which are rolled around the bars, transfer their load through friction.
8. The overlapping parts of mesh are glued together so as to ensure the continuity of the strengthening (Fig. 5c).
9. A layer of mortar is laid on top of the mesh to protect it from UV radiation and rain and also to provide further bond.

Shear tests on walls strengthened by PP mesh (Mayorca and Meguro 2004) showed that although the strengthening does not increase the peak strength, it contributes to improving the structural performance after crack occurrence. The strengthened walls exhibit larger post-peak strength, while the mesh helps to spread the diagonal cracks over a wide region. The presence of connectors and the mortar layer, ensuring bond to the masonry, are critical to the performance of the wall.

Confinement at Wall-to-Floor Junctions: Ring Beams

A ring beam is a structural element built on top of the masonry structure to the purpose of:

- Creating a continuous connection between the roof structure and walls to better distribute the vertical loads
- Improving the connection between orthogonal walls and the three-dimensional behavior of structure

Reinforced concrete (RC) ring beams have been widely used for the retrofit of masonry structures from the aftermath of the Friuli, Italy

1976 earthquake onwards. The construction of RC ring beams involves:

- Consolidating the top part of the existing masonry by injections
- Drilling vertical holes and inserting steel rods for the vertical connection between the masonry and new element
- Building the formwork and placing the reinforcement of the beam
- Casting the concrete

Such system improves the seismic capacity of structures by:

- Distributing the vertical loads
- Transferring the horizontal loads from the floors to the bearing walls
- Connecting the bearing walls so as to create a boxlike behavior and prevent out-of-plane failure

However, during a seismic event, the high stiffness of RC beams may induce out-of-plane bending of the portions of wall between the restrained floors (Borri et al. 2009) with consequent bulging of walls and activation of out-of-plane damage mechanisms. Furthermore, such technique is often time-consuming, not cost-effective, and adds mass to the structure, which increases the earthquake-induced inertia forces and consequently requires strengthening at the basis of the walls.

All these collateral effects represent major drawbacks for heritage buildings due to the large deformability and scarce cohesion of historic masonry as opposed to the high stiffness and weight of reinforced concrete. Moreover, the construction of a ring beam has a high impact from the aesthetic point of view and may require the removal of large portions of the original material of the walls and of part of the structure of the roof.

Indeed, several surveys carried out in European historic centers after earthquakes (e.g., Spence and D'Ayala 1999; D'Ayala and Paganoni 2011) highlighted how RC beams with poor connections to the underlying material

negatively affected the buildings rather than improving its structural response. Damage generally consisted of a shear failure at the interface between roof ring beam and wall, with more or less extensive damage to the masonry and eventual collapse of walls.

Nevertheless, the use of RC ring beam is allowed for the connection between walls and roof structure (Italian Ministry of Cultural Heritage and Activities 2006), providing that the dimensions of the ring beam itself do not cause an excessive increase of mass. However, conversely to steel ring beam, the use of RC beams for the strengthening of the connection between walls and intermediate floor structures is advised against (Ministero dei Beni Architettonici e Culturali, Italia, 2006), as it is far too invasive and deeply affects the performance of the structural system.

The 2007 edition of the California Historical Building Code (2007), instead, generally recommends the use of tie beams but does not specify the use of reinforced concrete.

In the scientific literature, references to procedures for the sizing of reinforced concrete ring beams can be hardly found, and it is highly likely that in common practice a large majority of additional concrete elements are simply dimensioned by the rule of thumb, using standard geometry and amount of reinforcement of ring beams in RC frames.

An improvement on RC ring beams are reinforced masonry ring beams, combining good quality brickwork (generally made of solid bricks) to steel or composite reinforcement bonded together by a binder-like grout (Ministero dei Beni Architettonici e Culturali, Italia, 2007).

In the case of steel reinforcement, the brickwork of the ring beam is laid leaving an internal cavity where the reinforcement bars and stirrups are positioned before casting the mortar-crete.

The creation of a reinforced masonry ring beam aims to:

- Provide a continuous connection between the roof structure and walls to better distribute the vertical loads.

- Improve the connection between orthogonal walls so as to enhance the three-dimensional behavior of structure.

The advantage of the technique in respect to RC ring beams is that stiffness and mass are closer to the original masonry. A RM ring beam has also good vertical deformability, and therefore it spreads the vertical load to the masonry beneath. Furthermore, it has a lower aesthetic impact, as the additional element can match the appearance of the original masonry.

RM beams are realized by:

- Building two wallets of brickwork or stonework on top of the existing wall, leaving a cavity between the two
- Placing reinforcement bars within the cavity
- Casting the concrete within the cavity, as to create a beam

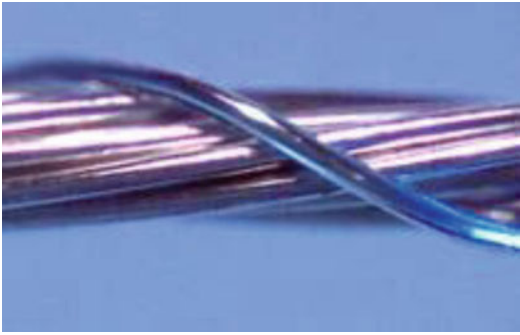
Steel plate and transversal bricks connect the outer wallets.

In some cases, consolidation of the masonry under the ring beam by injections is needed to ensure a sufficient shear capacity and similar stiffness.

Like in the case of RC ring beams, examples of the sizing of RM ring beams can be hardly found in the scientific literature.

A systematic dimensioning procedure for RM ring beams could be derived from the prescriptions specific to RM structures. The process can be detailed as follows:

- Calculation of the appropriate combination of static and seismic loads acting on the structural elements (EN 1991-1-1:2002 and EN 1998-1:2004).
- Dimensioning and verification of the beam for both static and seismic loading following the prescriptions provided for reinforced masonry structures (EN 1996-1-1:2005 and EN 1998-1:2004). This is done because a RM ring beam is in fact a bearing element and should therefore be calculated for vertical loading as beams are. Furthermore, the RM ring beams should be dimensioned for tensile load as the confining

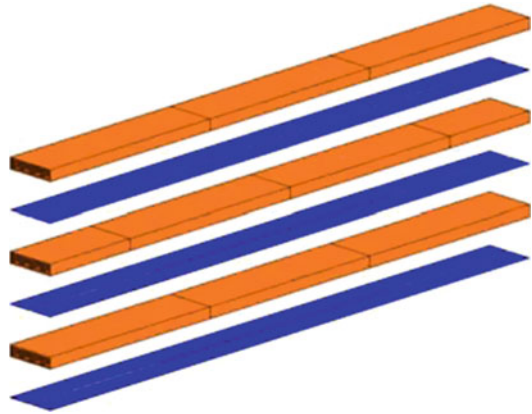


Seismic Strengthening Strategies for Heritage Structures, Fig. 6 Example of steel cord used in SRG (Borri et al. 2007)

action of the beam consists indeed in resisting the thrust generated by walls that separate under horizontal actions; this is done by ignoring the cross-sectional area of concrete and masonry and sizing the steel reinforcement so that it can bear the whole in-plane action.

- Dimensioning and checks of metallic connectors for both static and seismic loading (EN 1993-1:2005).
- Dimensioning and checks of connection of the metallic fasteners to the concrete element of the beam for both static and seismic loading (EN 1992-1-1:2004 – Section 8 Detailing of Reinforcement; DD CEN/TS 1992-4-1:2009).
- Dimensioning and checks of connection of the metallic fasteners to the masonry of walls for both static and seismic loading.
- Dimensioning and checks of connection of the metallic fasteners fixed to the timber elements of the roof for both static and seismic loading (EN 1995-1-1:2004 – Section 8 Connections with Metal Fasteners).

Borri et al. (2009) propose a system combining FRPs and steel-reinforced grout (SRG), the latter made of high strength steel wires forming cords (Fig. 6) embedded in a cementitious grout. The ring-beam system, called LATLAM, is obtained by overlapping several layers of bricks and laminates embedded within a polymeric matrix or a cementitious grout (Fig. 7). By superposition of blocks and composite sheets, it is possible to achieve a structural element of any size and length,



Seismic Strengthening Strategies for Heritage Structures, Fig. 7 Assembly of LATLAM ring beam (Borri et al. 2007)

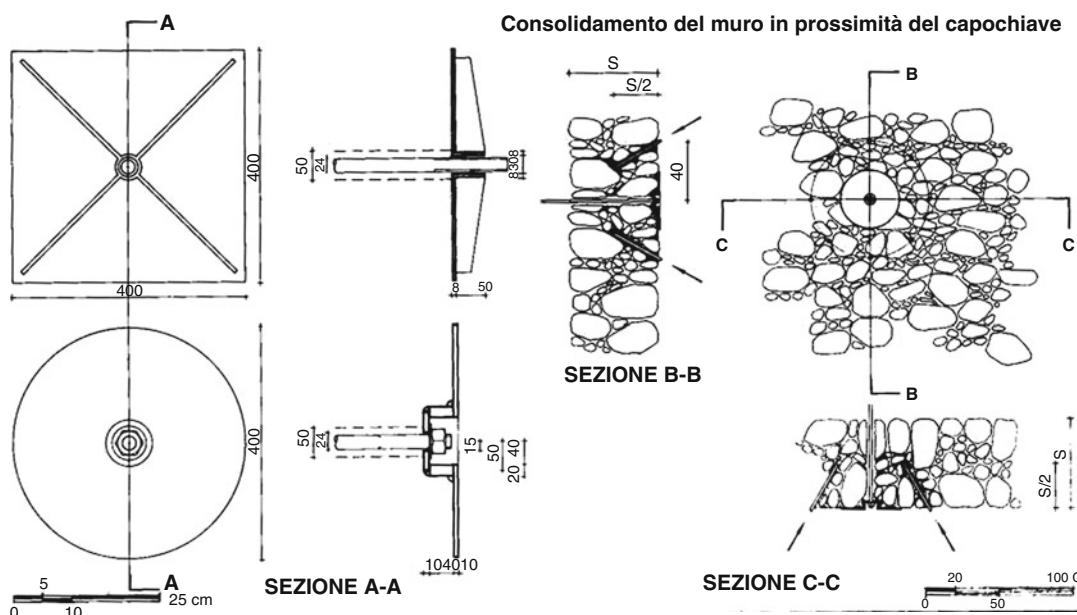
similarly to glue-laminated timber beams. The assembly acts as a single structural element and combines the compressive strength of masonry units with the good tensile properties of composite materials, avoiding the problem of the increase of mass typical of RC ring beams. A simplified model for the evaluation of the ultimate strength of a LATLAM section under bending stresses is available in Borri et al. (2007). The collapse is assumed to occur by crushing of masonry.

Strengthening by Way of Ties: Displacement Control

The insertion of ties to connect walls to walls, walls to floor, or walls to vaults is a traditional system, still in large use nowadays.

Post earthquake observation, experimental evidence and computational results, all show that crossties installed at the intersection of perpendicular sets of walls are able to prevent the overturning of whole façades without interfering with the original structural layout (Tomaževič 1999) (D'Ayala and Yeomans 2004; D'Ayala and Paganoni 2011).

For the ties to be effective, proper anchorage within or against the masonry is necessary. This might be achieved, either by using end plates or by using grouted ends. The tie is usually a passive element which becomes active when cracks open between orthogonal walls or timber beams tend to slide off their seat.



Seismic Strengthening Strategies for Heritage Structures, Fig. 8 Examples of metallic ties with end plates (Giuffrè 1993)

The dimensioning procedure of metallic ties connecting timber joist to walls can be derived from the prescriptions specific to steel structures. The process can be detailed as follows:

- Calculation of the appropriate combination of static and seismic loads acting on the structural elements (EN 1991-1-1:2002 and EN 1998-1:2004).
- Dimensioning and checks of the ties following the prescriptions provided for steel structures (EN 1993-1-1:2005).
- Dimensioning and checks of connection between metallic elements and timber elements of the horizontal structure (EN 1995-1:2004).
- Dimensioning and checks of connection of the metallic elements to masonry. Depending on the type of connection – by binder or by mechanical locking – checks should be performed in a different way as discussed in the next two subsections.

In the case of ties connecting masonry to masonry point c is not relevant. However, it

might be important to proceed to a local consolidation of the masonry by use of mortar grout before implementing the tie. The major pitfall of cross-ties is the possibility of pullout damage at the head of the anchorage due to the different deformabilities of anchor and masonry. The use of high-ductility systems recommended by codes would overcome this issue, thus achieving the objective of protecting culturally valuable finishes and preserving life and safety.

Ties with End Plates

The dimensioning process of metallic ties with end plates (Fig. 8) consists in:

- Sizing the cross section of the metallic rod to resist the axial load deriving from the pulling action of the portion of wall constrained by the tie in case of out-of-plane action. The minimum diameter of the rod is calculated by (Tomažević 1999):

$$D_{\min} = \sqrt{\frac{H_{u, \text{seg}}}{n} \frac{4}{\pi} \frac{1}{f_y}} \quad (4)$$

where

$H_{u,seg}$: ultimate seismic resistance of a critical part of a building,

n : number steel ties,

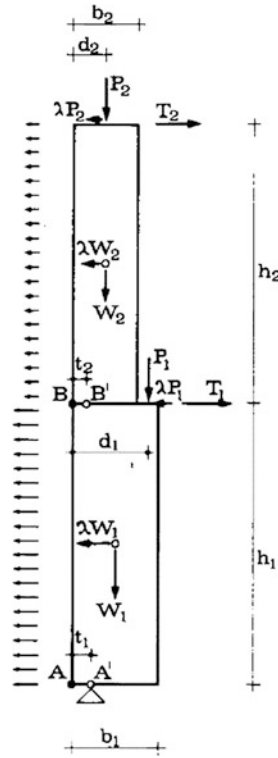
f_y : yield stress of steel.

The horizontal load that activates the overturning mechanism in a portion of wall, $H_{u,seg}$, needs to be equilibrated by the action of cross-ties and can be calculated by limit analysis. The position of hinges and the collapse load factor, for the portion of wall involved in the out-of-plane mechanism, are calculated so as to satisfy rotational equilibrium and the distribution of stress assumed in the masonry section at collapse as shown in Fig. 9. In deciding the static scheme for the calculation of the mechanism, the type of connections should be considered as they influence the constraints of the ideal beam that represents the wall: for instance, a wall with no positive connections to the floor structures can be modeled as a cantilever, while the positive effect of well-connected horizontal structures should be accounted for by using a simply supported beam scheme.

- Sizing the end plate to prevent punching failure. This is done by verifying that the tensile strength of the parent material is sufficient to withstand the 45° component of the axial load acting on the tie. The rupture surface to consider for the calculation of the acting stress is a conical/pyramid surface with sides inclined at 45° in respect to the plain of the wall (Fig. 10).

Ties Without End Plates

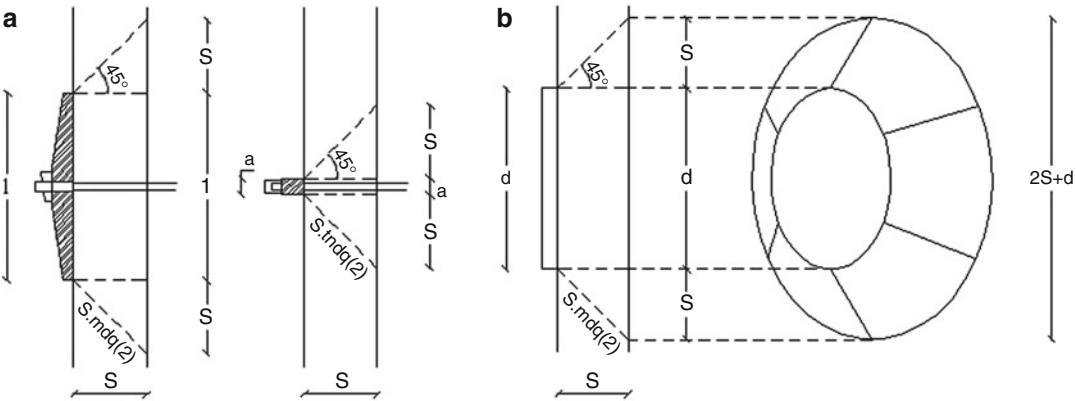
Anchors without end plates (Fig. 11) consist of a metallic profile that is embedded in the masonry and can transmit loads to the substratum by mechanical locking, friction, bond, or a combination of these three (Eligehausen et al. 2006), rather than use of an element such as a plate, a key, or a peg. Mechanical locking can be obtained, for instance, by undercut, namely, shaping the end part of the hole and introducing an anchor with an end shape so as to result larger than the rest of the shaft. Friction systems, instead, consist of torsion or displacement-



Seismic Strengthening Strategies for Heritage Structures, Fig. 9 Static scheme of multistory wall panel for determination of hinge position and collapse load multiplier (D'Ayala, Speranza, 2003)

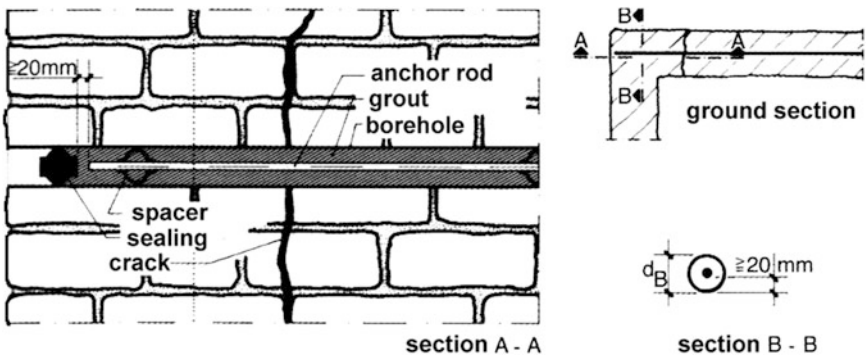
expansion anchors, which are designed to introduce a radial pressure between the anchor and the hole. Finally bond systems rely on the use of a binding agent, which is either injected or released through a capsule system (Fig. 12).

A fairly common commercial product consists of metallic profiles shaped as a coil; such profiles can be dry screwed in the masonry, thus mainly relying on a mechanical and frictional mechanism, or injected with resins/grout, in which case they work through the bond established between masonry and binder. Another popular system, increasingly used in conservation, is formed by steel sections provided with a fabric sleeve; profiles are installed in holes and then grout/resin is injected in the sleeve, so that the sleeve molds itself to the spaces and voids in the wall and creates a system combining mechanical locking and bond. A further advantage of this latter system is the

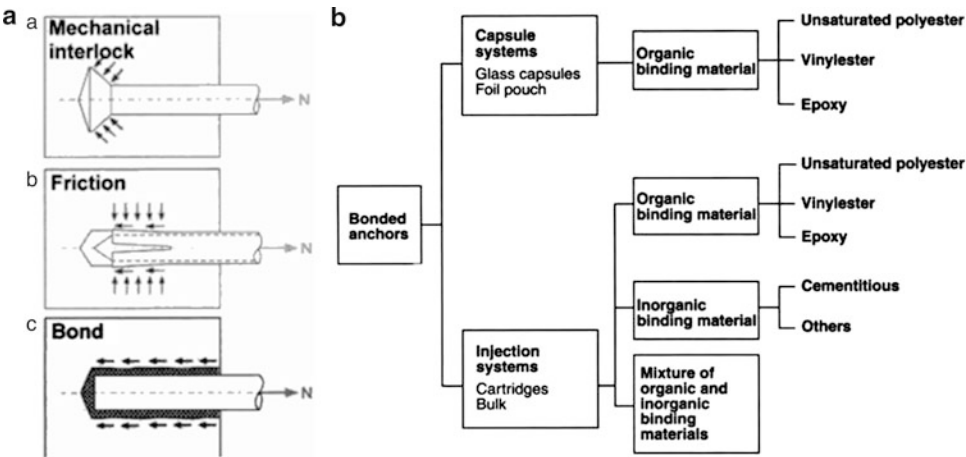


Seismic Strengthening Strategies for Heritage Structures, Fig. 10 Rupture surface for the calculation of tensile strength acting on parent material and sizing of

the end plate in case of: (a) rectangular key or (b) circular end plate (<http://posterremoto.altervista.org/Dati/i.html>, last accessed 7th March 2011)



Seismic Strengthening Strategies for Heritage Structures, Fig. 11 Grouted anchor injection scheme (Gigla, Wenzel 2000)



Seismic Strengthening Strategies for Heritage Structures, Fig. 12 Post-installed anchors: (a) systems for the transmission of load to substratum. (b) Typologies of binding materials for bond anchors (Eligehausen et al. 2006)

control over the diffusion of the binder within the parent material, while the grout can expand to the cavity around the ties, particularly useful for rubble double leaf masonry.

A variety of materials, ranging from mineral binders to polymers, can be used as binder. However, materials based on polymers, such as epoxy or polyester resins, which are often used for concrete structures, should undergo careful assessment before use in historical masonry, due to possible issues regarding the mechanical and physical compatibility. Much more common for the repair of historical buildings are mineral binder systems based on cement or hydraulic lime with the addition of admixtures and fillers or aggregate. To inject bore holes, usually pure water/binder systems are used with typical w/b values of 0.8–1.0. However, the w/b ratio has to be adjusted according to the volume to be injected and to the moisture content of the substrate.

Post-installed anchors are a very common method for the strengthening of historic masonry since they allow for a wide range of applications, from restoring the through-thickness cohesion of multilayered masonry to the increase of tensile and shear capacity of panels; furthermore, they can be used at the joints between structural elements either in the form of short connectors or as longitudinal elements. These latter have the same function as metallic crossties, but, rather than relying on an end plate for the connection to the wall, pullout loads are transmitted by a shear mechanism – for friction and bond anchors – or through mechanical locking within the masonry itself.

The diameter of the anchors is chosen as a function of the expected pullout force applied, the tensile strength of the masonry, and the bond strength of the grout to masonry interface. Bonded anchors can be passive or active, i.e., prestressed or posttensioned. Passive anchors are used for up to 4 m in length; prestressed anchors can be used up to 35 m in length (Gigla and Wenzel 2000).

The installation of anchors is carried out by:

- Drilling holes in the parent material; due to the weakness and preciousness of the parent

material, dry/wet diamond rotary drilling rather than percussive drilling is recommended.

- Removing all cores from the bore hole and checking the depth. Removing dust and debris.
- Placing anchors and injecting the binder.

The anchors utilized are usually stainless steel, threaded rods, or special prestressing bars. The diameter of the hole drilled to accommodate the anchor is a function of the strength differential between parent material and grout and this and the steel itself. The bond of the anchor inside the masonry unit depends on the type and properties of the grouting material and on the type of masonry block units.

The failure mechanisms of injection anchors in masonry are discussed in Gigla (2004, 2010), based on an extensive test program of over 500 pullout tests. The studies consider five different failure modes including failure mechanisms related to the (a) poor strength of the injected grout, (b) exceeded tensile strength of surrounding parent material, (c) bond failure between the outer surface of the grout and the parent material, (d) a combination of b and c, and (e) failure of the steel or grout in tension (under-designed).

Recommendations for the design of injection anchors are given in Gigla (2004). The design value for the bond strength is computed as follows:

$$X_{A,d} = \frac{\Phi_f}{\gamma_m} \left(\frac{f_{G,c}^2}{500} + X_{B,w} \right) \quad (5)$$

$X_{A,d}$: Design value of bond strength, as a function of the compressive strength of grout and independent of bond length.

Required minimum bond length: $L_b = 150$ mm inside monolithic stone, $L_b = 190$ mm in bed or head joints of brick, and $L_b = 430$ mm in bed and head joints of blockwork.

$f_{G,c}$: Compressive strength of grout. Minimum value: $f_{G,c} = 16.6$ N/mm². Maximum value covered: $f_{G,c} = 38.7$ N/mm². A minimum

bending tensile strength of $f_{G,ct} = f_{G,c}/8 = 2.0 \text{ N/mm}^2$ is required.

$f_{G,ct}$: Flexural strength of grout obtained from standard.

Φ_f : Reduction factor for bond in bed or head joints, $\Phi_f = 0.5$.

$X_{B,W}$: Term indicating the increase of bond strength inside water-absorptive stone material. $X_{B,W} = 0.15 \text{ N/mm}^2$.

γ_M : Partial factor for property, recommended: $\gamma_M = 1.35$

The term $X_{B,W}$ considers the water absorption of the masonry substrate. Pullout studies revealed that materials with higher water absorption capacity yielded higher pullout strength than materials with low or no water absorption capacity.

Furthermore, the anchor capacity can be designed by first computing $R_{A,d}$ which refers to the bond capacity between grout and rod. The equation considers the ratio of bed and head joints of the borehole surface across bond length and limits $R_{A,d}$ to bond inside full stone sections ($A_B/A_{G,d} = 1$ completely in stone; $A_B/A_{G,d} < 1$ in joint):

$$R_{A,d} = X_{A,d} \cdot \frac{A_B}{A_{G,d}} \cdot A_{A,d} \quad (6)$$

with $R_{A,d}$, design capacity of the injected anchor; A_B , portion of the cylindrical surface of injected mortar grout made of stones or units; $A_{G,d}$, total cylindrical surface of injected mortar grout; and $A_{A,d}$, interface of tensile element and injected mortar plug (cylindrical surface of steel bar), calculated with nominal bar diameter and bond length. The design strength of the anchor shall be smaller than the tensile capacity of the surrounding masonry so as to cause failure by yielding of the metallic element and prevent brittle failure of the masonry. This threshold can be computed as

$$F \leq \frac{1.9 \cdot f_{B,t} \cdot L_b \cdot \pi \cdot d_B \cdot (h_s^2 - d_B^2)}{\gamma_M \cdot \tan(\varphi) \cdot (d_B^2 + h_s^2)} \quad (7)$$

where

F : anchor tensile force

$f_{B,t}$: tensile strength of surrounding stone

d_B : borehole diameter

h_s : minimal distance to edge of surrounding masonry

L_b : bond length

$\tan(\varphi)$: tangent of angle of friction for force transmission between anchor's grout and bore-hole, about 50° in water-absorptive material and about 60° in non-water-absorptive material

γ_M : partial factor of safety for stone tensile strength; recommendation, 1.5

While these values are appropriate for an axial pullout test, they do not apply to failures of anchors loaded in shear or bending as a result of relative movements between the structural elements that they reconnect. Furthermore, as the family of dowel anchors also includes short fixings such as pins and nail-like fixings, failures connected to loads transmitted from other elements and in proximity to edges or openings should also be analyzed, these being highly relevant to various typologies of strengthening, such as confinement or end plates, where the use of fixings is required but rarely regulated.

Whereas the dimensioning procedure of bonded anchors in concrete has been extensively studied and commented (Eligehausen et al. 2006) and European guidelines do exist (EOTA TR029 2010; EOTA TR045 2013; DD CEN/TS 1992-4-1:2009), anchors in masonry lack specifically dedicated codes or recommendations. Modes of failure and hence the procedure for dimensioning could be partly derived from DD CEN/TS 1992-4-1:2009. Possible modes of failure due to axial load may be bond between metal and binder or between binder and parent material; cone pullout of parent material, due either to the failure of bond in the mortar joints or to tensile failure of masonry units; and failure of steel, although this is highly unlikely unless the metallic section is severely under-designed. Possible modes of failure due to loading transversal to the anchor axis may be crushing failure of either the binder or the parent material, edge failure of masonry when the load is applied in proximity of openings and corners, pryout failure of parent material, or failure of steel for combined bending and shear action. Each of these failures is considered by the guidelines and the elements of the anchorage assembly dimensioned accordingly.

Other factors considered in the guidelines are the performance of anchors when undergoing fatigue loading and the dimensioning criteria for seismic loading. For anchors embedded in concrete, ductile failure is achieved by controlling the dimensioning of steel so as to avoid the fragile failure of the substratum, and this concept should be extended to masonry structures. However, prescriptions in [EOTA TR029](#), [EOTA TR045](#), and [DD CEN/TS 1992-4-1:2009](#) rely on a series of parameters, such as bond strength or minimum distance to edge, that should be either provided by the producer of the anchors or derived experimentally. In case of anchors for masonry, due to the lack of standardization as well as the variability of parent materials and masonry fabrics, not every producer is able to provide all the required information and extensive tests covering the full range of parameters needed are missing from the scientific literature.

Energy Dissipation Systems

Systems like crossties have been and are still commonly applied in rehabilitation practice throughout Europe for providing connection at the joints of perpendicular sets of masonry walls where out-of-plane damage is most likely to occur. Nonetheless pullout damage at the head of the anchorage and increased in-plane diagonal cracking may affect valuable finishes and precious frescoes.

The current codes encourage the use of these conventional stiffness-based systems ([EN 1998 Eurocode 8](#); Italian Ministry of Cultural Heritage and Activities [2006](#)) because innovative techniques drawing on performance-based principles (Priestley [2000](#)), despite their effectiveness in new structures, rarely meet the requirements of reversibility and low impact required for historic structures.

As widely experienced in many applications carried out in recent years, an effective alternative to conventional strengthening techniques is the use of passive energy control and dissipation techniques.

Energy dissipation systems indeed comply with the concept expressed in [EC 8 \(EN 1998-3:2005\)](#) that:

The selection of the type, technique, extent and urgency of the intervention shall be based on the structural information collected during the assessment of the building. The following aspects should be taken into account:

[..]

- d) Increase in the local ductility supply should be pursued where required;
- e) The increase in strength after the intervention should not reduce the available global ductility.

In the field of cultural heritage, such techniques must aim to enhance the seismic performance of structures as well as limit the aesthetic impact on the building and protect elements that, despite not being relevant to the structural behavior, have cultural and historic value.

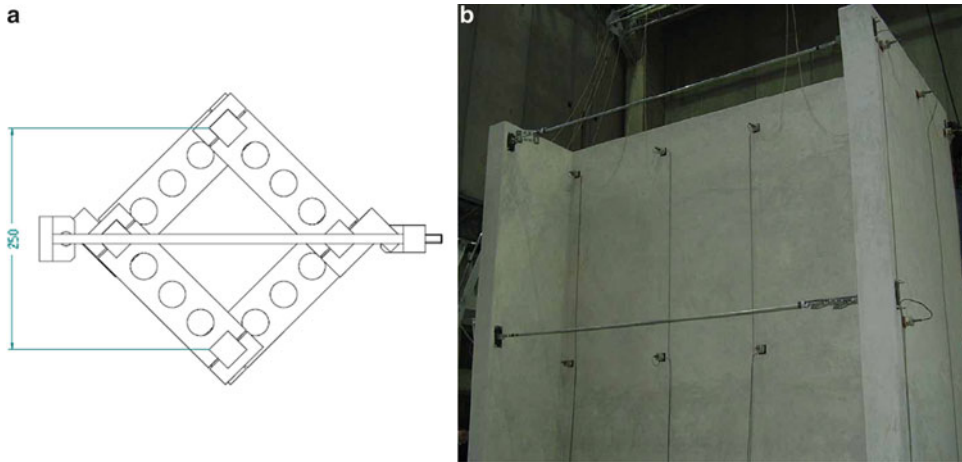
Energy control systems can either provide the structure with additional dissipation capacity and/or reduce the amount of ground input energy transferred to the structure. However, base isolation systems require heavy interventions on the bearing structure (base cut, new foundation structure, etc.) and are therefore rarely suitable for the retrofit of heritage buildings.

Energy dissipation devices can instead be placed in a number of key locations within the structure where relative displacements between members are expected to occur; thus, from the point of view of intrusiveness, they do not differ from many other standard techniques. A suitable position for the installation of dissipative devices is at the wall-to-floor interface of masonry buildings, where relative motion can occur without impairing the global structural integrity.

Energy Absorbers

Benedetti ([2004](#)) developed a series of energy-absorbing devices for existing masonry buildings drawing on experimental results that had showed that the more energy is absorbed through damage by the noncritical elements of the structure, the later global failure occurs. The main requirements set for the devices were:

- Activation of the device for very low level of damage and cracking
- Sensitivity to both in-plane and out-of-plane movements



Seismic Strengthening Strategies for Heritage Structures, Fig. 13 RAG energy absorber: (a) prototype (Benedetti 2004); (b) setup in series with metallic crosstie installed on a masonry specimen (Benedetti 2004)

- Limitation of forces transmitted to the parent material at the fixings of the dissipative devices, so as to avoid localized damage and detachment of the devices

The RAG energy absorber (Fig. 13a) consists of four arms hinged at their end so as to form a square element. Hinges are made of lead cylinders press-fitted into the arms. When two (or four) of the corners of the device, which are connected to the masonry wall, displace relatively to each other as a consequence of damage in the parent material, the hinges deform plastically in torsion (in-plane action of the device) or bending (out-of-plane action of the device), thanks to the low yield strength of lead. Conversely, the arms are made of a stiffer material so as to remain in the elastic range. Besides a full characterization of the device, RAG energy absorbers were also tested inserted in full-scale masonry specimens on shaking table in an out-of-plane configuration (Fig. 13b), where the devices were used in series with traditional metallic crossties to connect two parallel walls. The holes that can be observed in the arms of the RAG device aim at reducing the weight of the element and avoiding a delay in the activation of the yielding mechanism. A pretensioning element was positioned in series with the anchor to further ensure the lack of any initial deformation in the crosstie.

Shock Transmission Units

STUs, or viscous dampers, consist of a hollow cylinder filled with fluid, this typically being silicone based. As the damper piston rod and piston head receive an impact, the fluid is forced to flow through orifices either around or through the piston head. The resulting differential in pressure across the piston head can produce very large forces that resist the relative motion of the damper, while the input energy is dissipated in form of heat due to friction between the piston head and fluid particles flowing at high velocities. Conversely, since this type of devices is velocity dependent, slow movements such as thermal expansions are allowed.

The viscous devices are characterized by a nominal strength: for high load rates and for input forces below their capacity, they show high stiffness; above it they feature a perfect plastic behavior, with dissipation of hysteretic energy in case of cyclic loads. The main drawback of this type of devices is the risk of leakage of the fluid, which involves the need for regular inspections and further costs involved with the maintenance of the dissipative system. Mandara and Mazzolani (2001) investigated two different methods for the dimensioning of viscous devices. In the Plastic Threshold Approach (PTA), devices are conceived and sized to limit the magnitude of force transmitted across connected

members to a maximum value, determined according to the design resistance of structural elements involved. Beyond this threshold, the hysteretic energy dissipation takes place, while below the threshold the behavior of structural members is virtually rigid, which ensures the maximum degree of redundancy to the structure under serviceability load conditions. According to the Optimal Viscous Approach (OVA), the interaction between connected members is controlled by the viscous properties of the devices, which are dimensioned so as to minimize the magnitude of the force acting on them independently from its value. Contrary to PTA, the connection between elements is never fully rigid, so that energy can be dissipated under moderate intensity earthquakes too.

Shape Memory Alloys

Shape Memory Alloys (SMAs) are metallic materials that show special thermomechanical properties due to a reversible transformation between two crystalline configurations, austenite and martensite, without degradation of the crystal structure.

For some SMAs, the phase transformation is temperature-correlated, while for others, such as nitinol (NiTi SMA), the phase change can be stress induced at room temperature.

The stress–strain curve, measured during a monotonic tension test on a NiTi SMA wire, shows:

- (a) An almost linear portion corresponding to the elastic deformation of the material in its austenitic phase.
- (b) A loading plateau due to the transformation in martensite.
- (c) A new elastic phase that initiates after the complete transition to the martensitic phase (whose upper boundary is called maximum superelastic strain).
- (d) A final plateau, related to the true yielding of the alloy, which ends with failure.
- (e) Stress removal causes a reverse phase transformation, where the material goes back to the austenitic phase, stable for lower load's values.
- (f) Strains are almost completely recovered. This property is known as superelasticity.

These characteristics make SMAs particularly suitable for use in seismic dampers, especially considering that some alloys, such as nitinol, have very good corrosion-resistance characteristics.

Shape Memory Alloy Devices (SMADs), which consist of metallic ties and groups of wires of SMAs, were developed within the framework of the ISTECH project (Indirli et al. 2001) and patented by Fip Industriale, Padova, Italy. Their application has featured in high-profile projects such as the repair and strengthening of the Upper Basilica of St. Francis in Assisi, San Feliciano cathedral in Foligno, and San Serafino Church in Montegranaro, Italy. In the Upper Basilica of St. Francis, 47 SMADs were used to connect the roof to the tympanum of the transept; the SMADs were connected on one end to a threaded bar attached to an anchorage plate in the façade and on the other end to a plate bolted to a counter plate embedded in the RC rib that was cast to stiffen the existing concrete roof (dating back to the 1950s). SMADs were three-plateau self-balanced devices of three different sizes, with design capacity ranging from 17 to 52 kN and maximum allowable displacement between ± 8 and ± 25 mm.

Dissipative Anchor Devices

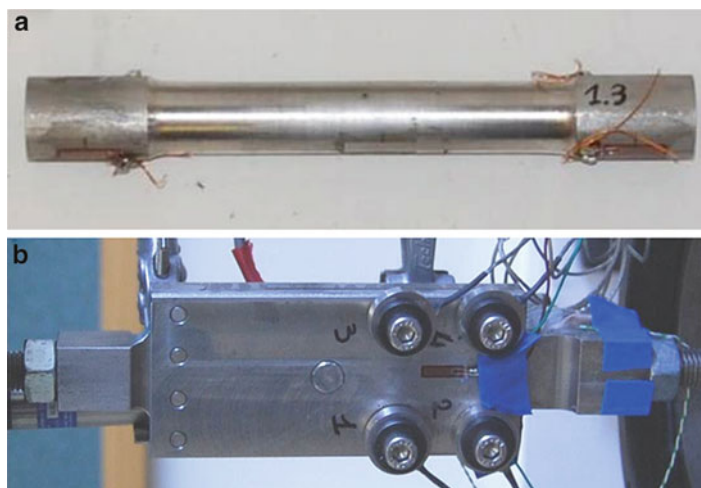
Given the difficulty of ensuring ductile failures of anchors embedded in masonry, D'Ayala and Paganoni (2014) in collaboration with CINTEC International developed two prototypes of dissipative anchor devices to address the problem of out-of-plane mechanisms.

The devices are conceived to be inserted at the joint between perpendicular walls, as part of longitudinal steel anchors grouted within the thickness of the walls. Devices are designed to work in tension, similarly to traditional crossties, which experience axial loading when one wall panel tilts outwards as a consequence of ground acceleration. This type of installation ensures a low impact on the aesthetics of the building as it does not affect the finishes.

One device relies on the plastic behavior of steel and consists of a metallic element designed to achieve lower capacity than the standard tie to

Seismic Strengthening Strategies for Heritage Structures,

Fig. 14 Dissipative anchor devices: (a) hysteretic and (b) friction based (D'Ayala, Paganoni 2014)



which it is connected (Fig. 14a). The second device (Fig. 14b) is made of a set of small plates; bolts are used to apply a perpendicular pressure creating friction and adjusting the level of slip load beyond which relative sliding occurs. Specially designed stops ensure that the sliding motion remains in the desired range.

While the metallic profiles improve the box-like behavior of the building, contributing to an increase in stiffness that improves the structural response to low excitations, the devices allow small relative displacements between orthogonal sets of walls; additionally, for higher horizontal loads, they dissipate part of the energy input into the structure so that problems of localized damage can be avoided. Therefore, the design focuses on control of displacements and reduction of accelerations and stress concentration.

The two dissipative devices have been validated by cyclic pseudo-static and dynamic tests on the isolated devices and by pullout tests on the devices anchored in series to a conventional post-installed anchor embedded in low-shear capacity masonry (D'Ayala and Paganoni 2014).

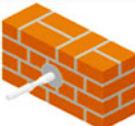
Experimental results confirmed that both the yielding and frictional elements are able to provide the anchorage with ductility and prevent damage to occur in either the grouted sleeve or the masonry, while traditional anchors, which were tested for comparison purposes, display a

high stiffness and fail at the interface between grout and parent material or in the masonry units. In the case of friction devices, it was observed that, due to tolerances of the pieces composing the assembly, a locking phenomenon occurred for the higher levels of perpendicular pressures; consequently, load-deflection curves feature additional stiffness in respect to the flat branch typical of a friction mechanism. Such behavior will need to be corrected so that the device can provide homogeneous performance over the required range of perpendicular forces.

The dimensioning procedure is based on the idea that a strengthening system can be divided into subcomponents to which one type of failure controlled by a single parameter can be associated. These parameters can be used in the formulae prescribed by design codes to calculate the capacities of components and hence determine the hierarchy of failure. The parameter values can be obtained via a number of sources: producers' specifications, recommended or limit values provided by codes, and, if any, laboratory and/or on-site tests.

Table 1 exemplifies the procedure for the calculation of the tensile capacity of the dissipative anchoring devices in series with a grouted metallic anchor. Maximum capacity is reached if one of the components fails or when the dissipative device is activated. Further checks can also be

Seismic Strengthening Strategies for Heritage Structures, Table 1 Design of dissipative anchors: parameters that identify the tensile capacity, value range achieved during experimental validation, and range prescribed/recommended by design codes (From D'Ayala and Paganoni 2014)

Performance parameters	Achievable range	Expected range																								
1a) F_{yield} : yielding capacity of hysteretic dissipative device [kN]	$F_{yield}=33$ kN (for hysteretic device of size suitable to coupling with M16 threaded bar)	$F_{yield}=27.8$ kN; calculated as: $F_{yield}=f_{y,yield}A_{yield}$ with $f_{y,yield}$ yielding strength of steel of hysteretic element and A_{yield} net cross sectional area of hysteretic element (EN 1993-1-1:2005)																								
1b) $F_{\perp}$: slip-load of frictional dissipative device [kN]	Considering: ✓ $F_{\perp}$: initial value imposed on devices. Variations recorded during tests are not considered; ✓ Slip load is given as range of values between maximum and minimum recorded values at constant level of $F_{\perp}$. <table><tr><th>$F_{\perp}$ [kN]</th><th>$F_{\parallel}$ min [kN]</th><th>$F_{\parallel}$ Max [kN]</th></tr><tr><td>12.5</td><td>3.25</td><td>14.5</td></tr><tr><td>15</td><td>5.7</td><td>18.3</td></tr><tr><td>17.5</td><td>6.65</td><td>22.4</td></tr></table>	$F_{\perp}$ [kN]	$F_{\parallel}$ min [kN]	$F_{\parallel}$ Max [kN]	12.5	3.25	14.5	15	5.7	18.3	17.5	6.65	22.4	Calculated as: $F_{\parallel}=\Phi n F_{\perp}$ with Φ coefficient expressing the ratio between $F_{\perp}$ and $F_{\parallel}$, $n=2$ number of frictional surfaces and $F_{\perp}$ applied perpendicular pressure. <table><tr><th>$F_{\perp}$ [kN]</th><th>$F_{\parallel} (\Phi=0.15)$</th><th>$F_{\parallel} (\Phi=0.55)$</th></tr><tr><td>12.5</td><td>3.75</td><td>13.75</td></tr><tr><td>15</td><td>4.5</td><td>16.5</td></tr><tr><td>17.5</td><td>5.25</td><td>19.25</td></tr></table>	$F_{\perp}$ [kN]	$F_{\parallel} (\Phi=0.15)$	$F_{\parallel} (\Phi=0.55)$	12.5	3.75	13.75	15	4.5	16.5	17.5	5.25	19.25
$F_{\perp}$ [kN]	$F_{\parallel}$ min [kN]	$F_{\parallel}$ Max [kN]																								
12.5	3.25	14.5																								
15	5.7	18.3																								
17.5	6.65	22.4																								
$F_{\perp}$ [kN]	$F_{\parallel} (\Phi=0.15)$	$F_{\parallel} (\Phi=0.55)$																								
12.5	3.75	13.75																								
15	4.5	16.5																								
17.5	5.25	19.25																								
2) F_{steel} : tensile capacity of metallic bar at yielding [kN] 	$F_{steel}=71$ kN (for M16 threaded bar - values stated by producer)	$F_{steel}=71$ kN; calculated as: $F_{steel}=f_y A$ with f_y yielding strength of steel and A net cross sectional area of metallic profile (EN 1993-1-1:2005)																								
3) $f_{b,ab}$: bond strength anchor/binder [MPa] calculated on cylindrical surface of embedded bar 	Calculated as: $f_{b,ab}=F_{sb,bond}/A_{steel}$ with $F_{sb,bond}$ recorded load at failure and A_{steel} cylindrical lateral surface calculated as: $A_{steel}=\pi l d_{pitch}$ with l embedment length and d_{pitch} pitch diameter of steel bar. For pull-out tests of M16 threaded bars from 550 mm long grouted socks: $f_{b,ab}=2.07$ MPa (CoV 4%)	$f_{b,ab}=$ 3.4 MPa – design value suggested in BS 5268-2 for tested binder, bar diameter and type of bar 2 MPa – design value suggested in EN 1996-1-1:2005 for tested binder and type of application																								
4) $f_{b,b/p}$: bond strength binder/parent material [MPa] calculated on cylindrical surface of grouted socket 	Calculated as: $f_{b,b/p}=F_{b/p,bond}/A_{hole}$ with $F_{b/p,bond}$ recorded load at failure and A_{hole} inner cylindrical surface of drilled hole of length l . For pull-out tests with vertical load on masonry specimens σ_d : <table><tr><th>l [mm]</th><th>σ_d [MPa]</th><th>$f_{b,b/p}$ [MPa]</th></tr><tr><td rowspan="2">350</td><td>0.70</td><td>0.67 (CoV 8%)</td></tr><tr><td>0.07</td><td>0.57 (CoV 18%)</td></tr><tr><td rowspan="2">220</td><td>0.10</td><td>0.26 (CoV 34%)</td></tr><tr><td>0.05</td><td>0.4</td></tr></table>	l [mm]	σ_d [MPa]	$f_{b,b/p}$ [MPa]	350	0.70	0.67 (CoV 8%)	0.07	0.57 (CoV 18%)	220	0.10	0.26 (CoV 34%)	0.05	0.4	Calculated as: $f_{b,b/p}=f_{vk}=f_{vk,0}+0.4\sigma_d$ with $f_{vk,0}$ initial shear strength (calculated through experimental results) and σ_d vertical load (EN 1996-1-1:2005). <table><tr><th>σ_d [MPa]</th><th>$f_{b,b/p}$ [MPa]</th></tr><tr><td>0.7</td><td>0.52</td></tr><tr><td>0.07</td><td>0.27</td></tr><tr><td>0.10</td><td>0.08</td></tr><tr><td>0.05</td><td>0.06</td></tr></table>	σ_d [MPa]	$f_{b,b/p}$ [MPa]	0.7	0.52	0.07	0.27	0.10	0.08	0.05	0.06	
l [mm]	σ_d [MPa]	$f_{b,b/p}$ [MPa]																								
350	0.70	0.67 (CoV 8%)																								
	0.07	0.57 (CoV 18%)																								
220	0.10	0.26 (CoV 34%)																								
	0.05	0.4																								
σ_d [MPa]	$f_{b,b/p}$ [MPa]																									
0.7	0.52																									
0.07	0.27																									
0.10	0.08																									
0.05	0.06																									
5) $f_{masonry}$: Shear strength of parent material [N/mm ²] 	This type failure, although expected, did not occur during experimental campaigns	Calculated as: $f_{masonry}=f_{vk}=f_{vk,0}+0.4\sigma_d$ (EN 1996-1-1:2005). In the tested case it would be expected: 0.52 MPa 0.27 MPa The failure surface, A_f , is a truncated cone with smallest base corresponding to the drilled hole, apothem inclined at 45° and height equal to the wall thickness																								
6) $f_{masonry}$: Tensile strength of parent material [N/mm ²] 	A "wrench" failure occurs instead of the expected "cone pull-out" failure. Failure surface, A_f , develops along vertical joints. $f_{masonry}=f_{nt}=0.67$ MPa (from wrench test)	No mention about this type of failure has been found in the technical literature or design codes.																								

performed in terms of displacement and ductility, as discussed in the following.

The demand in terms of tensile capacity of metallic anchors is calculated as:

$$F_D = M \cdot a_i \quad (8)$$

where

M : mass of structure that bears on the i th anchor of the strengthening system. M depends on the geometry and construction arrangement of the building, including horizontal structures, and on the layout of the set of anchors to be designed.

a_i : horizontal acceleration experienced by the mass M . An estimate of the natural period of the system can be used to determine the correct design spectral ordinate and the distribution of amplification over the height of the structure. An accurate estimate of the acceleration at the height of the anchor can be obtained either by using the first natural modal shape or the procedure proposed by Miranda and Taghavi (2005).

The reference acceleration is calculated as function of the three limit states defined in EN 1998-3:2004, so that the design demand is:

- F_{DNC} : near collapse, calculated for a seismic action with a probability of exceedance of 2 % in 50 years
- F_{DSD} : significant damage, calculated for a seismic action with probability of exceedance of 10 % in 50 years
- F_{DDL} : damage limitation, calculated for a seismic action with probability of exceedance of 20 % in 50 years

All the subcomponents of the strength-only portion of the anchor assembly are brittle or, in the case of the grouted steel elements, are not supposed to experience large deformations; therefore, they are dimensioned in terms of strength for near collapse limit state, according to EN 1998-3:2004. The minimum capacity in the assembly must be:

$$\text{Min}(F_{\text{steel}}, F_{a/b \text{ bond}}, F_{b/p \text{ bond}}, F_{\text{masonry}}) > F_{DU} \quad (9)$$

Capacities are calculated using information in the relevant row in Table 1. It is important to notice that while certain values in Table 1 are generally applicable, others are not: the standardization of steel production ensures highly repeatable performance, whereas bond strength, for instance, largely varies depending on the substratum. Therefore, it is advisable that tests be performed each time; if this cannot be done, lower limit values provided by the code and reported in Table 1 can be used.

Table 1 does not include the connections' anchor/dissipative devices and the stops of the friction device. However, these components should also be dimensioned to resist F_{DNC} as also required by EN 15129:2009.

The dissipative elements of the devices, either hysteretic or frictional, are designed to be activated at the threshold of damage limitation, when cracks start opening and allow for the dissipative elements to become active. When the hysteretic devices enter the plastic field, or the friction devices start sliding, the connection between wall panels is still maintained, but the pullout of the head of the anchorage is prevented and drift controlled.

Hence, from the point of view of force design, depending on which dissipative device is used, the following requirements apply:

$$\begin{aligned} F_{\text{yield}} &= F_{DD} \text{ or} \\ F_{//} &= F_{DD} \end{aligned} \quad (10)$$

where F_{yield} is the yielding capacity of the hysteretic device and $F_{//}$ is the slip load of the frictional device.

The dissipative devices should also comply with requirements for interstorey drift of buildings undergoing seismic action. The chosen value of maximum allowable drift, $d_r = 0.003$, for damage limitation is taken from OPCM (2005) which provides drift limits for damage limitation in walls of masonry buildings. This limit is also in line with the expected drift stated in FEMA 356 (BSSC 2000) for unreinforced masonry buildings at the limit state of immediate occupancy.

For historic buildings, the allowable drift is increased by using the reduction factor $\nu = 0.4$, which is taken from [EN 1998:2004](#) and accounts for the fact that devices are designed to be used in heritage structures, which fall in the importance category III of Eurocode 8 ([EN 1998:2004](#)).

It is therefore

$$d_r = \Delta_e < 0.003 \cdot h/\nu \quad (11)$$

where

Δ_e : the elongation of the device before yielding, in case of the hysteretic device, or before activation of the friction mechanism, in case of the frictional device,

h : interstorey height, or vertical distance of installation of anchors. A standard distance of 3 m has been assumed in the calculations, but anchors might need to be spaced more closely along the height of the wall to prevent substratum failures.

The first threshold of the hysteretic device, identified at 0.5 % elongation of the dissipative element, coincides with 46 % of the hysteretic device load capacity. The second threshold, at 5 % elongation and 72 % maximum load, should instead be used to verify the capacity of the dissipative device for the limit state of significant damage, so that dissipation of energy is ensured during low-to-medium seismic excitations. Beyond this limit, the device has a further margin of safety given by buckling, meaning that the device can reach the limit of near collapse with damage, but not complete failure, and it could still be substituted, as it has been proved by experimental testing and in response to the requirements of [EN 1998:2004](#).

In the case of a frictional device, the drift limit is ensured by default because: before activation of the friction mechanism, deformations are negligible and, beyond activation of sliding, the device displacement is limited by the assembly stops. The device can therefore perform for all limit states, as long as the connections and stops in the assembly are designed to resist up to the state of near collapse.

Summary

In the last decades, a number of technical solutions for the improvement of structural connections have been developed in response to the increasing demand for strengthening systems specifically designed for the protection of heritage assets from earthquake-induced damage.

Notwithstanding recommendations from seismic standards and conservation codes of practice, and the availability of alternative technical solutions on the market, much emphasis when choosing strengthening solutions of connections is still placed on their capacity to enhance the overall strength of the system, rather than improve its ductility or ability to dissipate energy. When connections between vertical elements, and between vertical and horizontal elements, as well as vaults and domes have sufficient capacity, vertical and horizontal loads are better distributed, out-of-plane mechanisms of damage are prevented, and, hence, failure of floors and roofs can be avoided, substantially reducing the probability of collapse and fatalities in an earthquake.

Historically, builders mastered the art of careful detailing of joints from a process of trial and error; today's engineers possess the necessary insight into physical and mechanical laws governing the dynamics of structures to control the process of structural upgrading at the design level. Thus, when connections, in existing buildings, lack adequate capacity, further elements can be added to the original structure to achieve the desired performance.

The review of modern techniques to enhance connections presented in the previous section highlights aspects specific to each applications as well as advantages and pitfalls typical of each system. A large majority of the discussed methods are applicable both at the local and global scale for repair and upgrade of single structural elements or of whole structures.

In respect to the past, strengthening systems for structural connections rely on a more accurate design, modern structural systems, and innovative and more durable materials, for example, stainless steel or titanium, which substituted iron in cross-ties, with considerable advantages

Seismic Strengthening Strategies for Heritage Structures, Table 2 Performance parameters typical of each strengthening technique

Typology of strengthening		Performance parameters
Ring beams		Flexural, tensile, and shear capacity of beam (kNm/kN)
		Capacity of fixings between ring beam and parent material (crushing or pryout of parent material, shearing off of fixings, shearing of parent material, etc.) (kN)
		Horizontal sliding shear capacity of the masonry underneath the level of the fixings of the ring beam into the parent material (—)
Wall plates		Flexural, tensile, and shear capacity of wall plates (kNm/kN)
		Capacity of fixings between wall plates and parent material (crushing or pryout of parent material, shearing off of fixings, shearing of parent material, etc.) (kN)
		Horizontal sliding shear capacity of the masonry underneath the level of the fixings of the ring beam into the parent material (—)
Steel connectors for floor/roof structures		Tensile capacity (kN) as minimum depending on
Anchoring with end plates		Combined tensile and bending strength of connector/tie (N/mm ²)
Anchoring without end plates and with grouting		Tensile capacity of fixings in timber elements (kN)
Nailing		Bond strength (for bonded systems) connector/binder (N/mm ²)
		Bond strength (for bonded systems) binder/parent material (N/mm ²)
		Tensile capacity of fixings (frictional or mechanical fixings) in parent material (kN)
		Tensile strength of parent material for cone pullout or punching shear (N/mm ²)
		Reduction of tensile capacity due to distance to edge (%)
		Reduction of tensile capacity due to distance between anchors, nails, and connectors (%)
		Compressive strength of parent material under action of end plate (N/mm ²)
		Flexural capacity of end plate (kN)
		Shear capacity (kN) as minimum depending on
		Shear strength of anchor (N/mm ²)
		Tensile strength of anchor for combined axial and bending load (N/mm ²)
		Bearing capacity of parent material (N/mm ²)
		Bearing capacity (for bonded system) of binder (N/mm ²)
		Shear capacity of fixings in timber elements (kN)
		Reduction of shear capacity due to distance to edge (%)
		Reduction of shear capacity due to distance between anchors (%)
Corners confinement	Column ties	Flexural, tensile, and shear capacity of columns (kNm/kN)
		Capacity of fixings between ring beam and parent material (crushing or pryout of parent material, shearing off of fixings, shearing of parent material, etc.) (kN)
	FRPs	Tensile capacity of FRPs (kN)
		Bond strength between FRP/parent material (N/mm ²) and the tensile and compressive capacity of the parent material (N/mm ²)
	Polymer mesh, PP mesh, steel mesh	Tensile capacity (kN) as minimum depending on
		Tensile strength of mesh (N/mm ²)
		Tensile strength of fixings (N/mm ²)
		Bond strength fixing/binder (N/mm ²)
		Bond strength binder/parent material (N/mm ²)
		Shear strength of parent material for cone pullout (N/mm ²)

(continued)

Seismic Strengthening Strategies for Heritage Structures, Table 2 (continued)

Typology of strengthening		Performance parameters
Energy dissipation systems		Shear capacity (kN) as minimum depending on
		Shear strength of parent material at the interface strengthened/unstrengthened material
		Shear strength of fixings (N/mm^2)
		Bearing capacity of binder (N/mm^2)
		Bearing capacity of parent material (N/mm^2)
	Yielding anchor devices	Yielding load of device (kN)
		Allowable displacement after yielding (mm)
		Dissipated energy (kJ)
		Change in material properties during cyclic loading (e.g., hardening of metal, progressive crushing of parent material)
		Buckling of device
		Ultimate tensile capacity (kN) as minimum depending on
		Tensile strength of device (N/mm^2)
		Other checks same as traditional anchors
		Shear capacity (kN) as minimum depending on
		Shear strength of device (N/mm^2)
		Other checks same as traditional anchors
	Frictional anchor devices	Load for activation of frictional mechanism (kN)
		Ultimate tensile load of device (kN)
		Ultimate allowable displacement for sliding (mm)
		Dissipated energy (kJ)
		Change in material properties during cyclic loading (e.g., hardening of metal, deterioration of frictional surfaces, effect of thermal expansion, presence of debris on frictional surfaces, progressive crushing of parent material)
		Ultimate tensile capacity (kN) as minimum depending on
		Tensile capacity of device (kN)
		Other checks as traditional anchors
		Shear capacity (kN) as minimum depending on
		Checks as per traditional anchors
Energy dissipation system	RAG energy absorbers	Axial yielding load that leads to yielding of lead pins (kN)
		Allowable displacement after yielding (mm)
		Dissipated energy (kJ)
		Change in material properties during cyclic loading (e.g., hardening of metal, progressive crushing of parent material)
		Ultimate tensile capacity (kN) as minimum depending on
		Tensile strength of anchor (N/mm^2)
		Tensile strength of RAG device
		Tensile strength of parent material to punching failure (N/mm^2)
		Compressive strength of parent material under action of end plate (N/mm^2)
		Flexural capacity of end plate (kN)
	RETE energy absorbers	Yielding load of the lead wire (kN)
		Ultimate tensile load of device (kN)
		Allowable displacement after yielding (mm)
		Dissipated energy (kJ)
		Change in material properties during cyclic loading (e.g., hardening of metal, progressive crushing of parent material)

(continued)

Seismic Strengthening Strategies for Heritage Structures, Table 2 (continued)

Typology of strengthening	Performance parameters
Shock Transmission Units (STUs) and viscous dampers	1st threshold – thermal deformations
	Allowable displacement (mm)
	Transmitted load for low-rate loading (kN)
	2nd threshold – low seismic excitation, connections expected to perform as pinned connections
	Allowable displacements (mm)
	Capacity (kN)
	Capacity of connectors between STUs and parent material at 1st threshold (kN)
	3rd threshold – higher seismic excitation, connections expected to have hysteretic response
	Allowable displacement (mm)
	Transmitted load (kN)
	Dissipated energy for hysteretic response (kJ)
Shape Memory Alloy Devices (SMADs)	1st threshold – thermal deformations
	Allowable displacement (mm)
	Transmitted load for low-rate loading (kN)
	2nd threshold – low seismic excitation, connections expected to perform as pinned connections
	Allowable displacements (mm)
	Capacity (kN)
	Capacity of connectors between STUs and parent material (e.g., pullout or tensile failure of fixings) (kN)
	3rd threshold – higher seismic excitation, connections expected to have hysteretic response
	Allowable displacement (mm)
	Transmitted load (kN)
	Dissipated energy for hysteretic response (kJ)
	ith threshold: same parameters as per 2nd and 3rd threshold repeated for the higher plateaus of Shape Memory Alloys

in terms of issues related to material deterioration (e.g., expansion due to corrosion, failures due to reduced resisting section).

Some of the techniques described have undergone revision after observation of their performance in the aftermath of major earthquakes proved that they are unsuited to certain applications. For instance, RC ring beams need to be especially well connected to bearing walls; otherwise, they tend to create shear failure in walls and collapse; furthermore, it has been shown that they are not a viable solution in case of multileaf masonry not well bonded.

Recent strengthening solutions are evolving to include issues of environmental compatibility

while remaining economically feasible, especially for vulnerable communities in developing countries.

The most important shift for future implementation is the increased awareness and technological research for economic solutions for ductility and energy dissipation-based devices, which can allow the improvement of performance of entire building stock in historic centers of earthquake-prone countries.

The performance of each system can be verified by determining the parameters and carrying out the checks summarized in Table 2. The parameters listed refer, when available, to dimensioning procedures prescribed by codes and

guidelines and otherwise from the available technical literature included in this review. Some of the systems in the table are not codified; therefore, the dimensioning procedure is not standardized or sometimes not reported for intellectual property reasons; in these cases parameters are suggested by the authors on the basis of their expertise and of the available information on the behavior of the various devices.

The major advantage of Table 2 is that parameters have been homogenized, using a common nomenclature, to facilitate the comparison of various techniques.

Cross-References

- [Passive Control Techniques for Retrofitting of Existing Structures](#)
- [Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions](#)
- [Strengthening Techniques: Masonry and Heritage Structures](#)

References

- Benedetti D (2004) Increasing available ductility in masonry buildings via energy absorbers. *Shaking table tests*. *Eur Earthq Eng* 18(3):3–16
- Borri A, Castori G, Grazini A, Giannantoni A (2007) Performance of masonry elements strengthened with steel reinforced grout. In: *Fiber-reinforced polymer reinforcement for concrete structures – FRPRCS-8*, Patras, 16–18 July
- Borri A, Castori G, Grazini A (2009) Retrofitting of masonry building with reinforced masonry ring-beam. *Constr Build Mater* 23(5):1892–1901
- Building Seismic Safety Council (BSSC) (2000) *Prestandard and commentary for the seismic rehabilitation of buildings*, FEMA, 356. Federal Emergency Management Agency, Washington, DC
- California Historical Building Code (2007) *California code of regulations*. Title 24, part 8. California Building Standards Commission, International Code Council, Washington, DC
- Casadei P, Agneloni E (2008) *Advance composites applications on historical structures in Italy: Case studies and future developments* 6th International Conference on Structural Analysis of Historical Construction
- Chuxian S, Guiqiu L, Wenchao W (1997) The design of brick masonry structure with concrete column. In: *XI brick and block masonry conference*, Shanghai
- CNR-DT 200 R1/2013. *Guide for the Design and Construction of Externally Bonded FRP Systems for Strengthening Existing Structures*, ROMA – CNR October 10th 2013 – release of May 15th 2014
- Corradi M, Grazini A, Borri A (2007) Confinement of brick masonry columns with CFRP materials. *Compos Sci Technol* 67:1772–1783
- D'Ayala D, Paganoni S (2011) Assessment and analysis of damage in L'Aquila historic city centre after 6th April 2009. *Bull Earthq Eng* 9(1):81–104
- D'Ayala D, Paganoni S (2014) Protocol for testing and design of dissipative devices for out-of-plane damage *Struct Build* 167(1):26–40
- D'Ayala D, Speranza E (2003) Definition of Collapse Mechanisms and Seismic Vulnerability of Historic Masonry Buildings. *Earthquake Spectra* 19(3):479–509. doi:10.1193/1.1599896
- D'Ayala D, Yeomans D (2004) Assessing the seismic vulnerability of late Ottoman buildings in Istanbul. In: *IV international seminar on structural analysis of historical constructions – SAHC04*, Padua
- DD CEN/TS 1992-4-1:2009. *Design of fastenings for use in concrete*
- Di Ludovico M, D'Ambra C, Prota A, Manfredi G (2010) FRP confinement of tuff and clay brick columns, experimental study and assessment of analytical models. *J Compos Constr* 14(5):583–596
- Eligehausen R, Mall R, Silva JF (2006) *Anchorage in concrete construction*. Ernst Sohn, Berlin
- EN 1991-1:2002. *Eurocode 1- Actions on structures*
- EN 1992-1:2004. *Eurocode 2- Design of concrete structures*
- EN 1993-1:2005. *Eurocode 3- Design of steel structures*
- EN 1995-1:2004 + A1:2008. *Eurocode 5- Design of timber structures*
- EN 1996-1:2005. *Eurocode 6- Design of masonry structures*
- EN 1998-1:2004. *Eurocode 8- Design of structure for earthquake resistance. Part 1: General rules, seismic actions and rules for buildings*
- EN 1998-3:2005. *Eurocode 8- Design of structure for earthquake resistance. Part 3: Assessment and retrofitting of buildings*
- EN 15129:2009. *Anti-seismic devices*
- EOTA, European Organisation for Technical Assessments (2010) TR 029, *Design of bonded anchors*, Sept 2010. Downloaded at www.eota.eu. Accessed 8 Jan 2014
- EOTA, European Organisation for Technical Assessments (2013) TR 045, *Design of metal anchors for use in concrete under seismic actions*. Downloaded at www.eota.eu. Accessed 8 Jan 2014
- Gigla B (2004) Bond strength of injection anchors as supplementary reinforcement inside historic masonry. In: *XIII international brick and block masonry conference*, Amsterdam, pp 1–10
- Gigla B (2010). Comparison of failure of injection anchors as supplementary reinforcement inside masonry and concrete. In: *VIII international masonry conference*, Dresden, pp 2019–2028

- Gigla B, Wenzel F (2000) Design recommendations for injection anchors as supplementary reinforcement of historic masonry. In: XII international brick and block masonry conference, Madrid, pp 691–706
- Giuffrè A (1993) Sicurezza e conservazione dei centri storici: il caso di Ortigia. Laterza, Roma-Bari
- Hamoush SA, McGinley MW, Mlakar P, Scott D, Murray K (2001) Out-of-plane strengthening of masonry walls with reinforced composites. *J Compos Constr* 5(3):139–145
- Indirli M, Castellano MG, Clemente P, Martelli A (2001) Demo-application of shape memory alloy devices, the rehabilitation of S. Giorgio Church Bell-Tower. In: VI international symposium on smart structures and materials – SPIE 2001, Newport Beach
- Ministero dei Beni Architettonici e Culturali, Italia, Guidelines for evaluation and mitigation of seismic risk to cultural heritage : (2007) Cangemi Editore S.p.A.
- Karantoni F, Fardis M (1992) Effectiveness of seismic strengthening techniques for masonry buildings. *ASCE* 118(7):1884–1902
- Krevaikas TD, Triantafillou T (2006) Masonry confinement with fiber-reinforced polymers. *ASCE J Compos Constr* 9(2):128–135
- Mandara A, Mazzolani FM (2001) Energy dissipation devices in seismic up-grading of monumental buildings. In: III international seminar on structural analysis of historical constructions – SAHC01, Guimarães
- Marcari G, Manfredi G, Pecce M (2003) Experimental behaviour of masonry panels strengthened with FRP sheets. In: Tan KH (ed) 6th international conference on fibre-reinforced plastics for reinforced concrete structures, World Scientific, Singapore, pp 1209–1218
- Mayorca P, Meguro K (2004) Proposal of an efficient technique for retrofitting unreinforced masonry dwellings. In: XIII world conference on earthquake engineering, Vancouver
- Miranda E, Taghavi S (2005) Approximate floor acceleration demands in multistory buildings. I: formulation. *J Struct Eng* 131:203–211
- OPCM (2005) No. 3431, May 3, 2005. Official Bulletin no. 107, May 10, 2005 (in Italian). *Gazzetta Ufficiale – Serie Generale* n. 107
- Priestley MJN (2000) Performance based seismic design. In: XII world conference of earthquake engineering, Auckland
- Spence R, D'Ayala D (1999) Damage assessment and analysis of the 1997 Umbria-Marche earthquakes. *Struct Eng Int* 9(3):229–233
- Tomazevic M (1999) Earthquake-resistant design of masonry buildings. Imperial College Press, London
- Tumialan G, Huang P-C, Nanni A, Silva P (2001) Strengthening of masonry walls by FRP structural repointing. In: Burgoyne CJ (ed) 5th international conference on fibre-reinforced plastics for reinforced concrete structures, Thomas Telford, Cambridge, pp 1033–1042

Seismic Tomography of Volcanoes

Ivan Koulakov^{1,2} and Nikolay Shapiro³

¹Trofimuk Institute of Petroleum Geology and Geophysics, SB RAS, Novosibirsk, Russia

²Novosibirsk State University, Novosibirsk, Russia

³Institut de Physique du Globe de Paris
Laboratoire de Sismologie, Paris, France

Synonyms

Ambient noise tomography; Body-wave tomography; Magma sources; Seismic observations on volcanoes; Seismic properties of crust; Seismic tomography

Introduction

Volcano tomography is a branch of geophysics oriented to studying the deep structures beneath volcanoes by means of seismic tomography. **Seismic tomography** is a method for reconstruction of continuous distribution of seismic parameters in 1D, 2D, 3D, or 4D (space and time) using the characteristics of seismic waves traveling between sources and receivers. Seismic parameters to be found in tomographic inversion are in most cases velocities of P and S seismic waves (P and S velocities). For volcanoes, one of the key parameters appears to be the V_p/V_s ratio which can be used to evaluate the content of fluids and melts. Besides the velocity distributions, seismic tomography may provide the information on the anisotropy of seismic parameters which helps studying regional stresses and space-oriented geological structures. In some tomography studies, the target parameter might be the attenuation of P or S waves which may also give important information on magma sources beneath volcanoes.

At all stages of the human history, understanding the mechanisms causing volcano eruptions was one of the most intriguing problems.

Until relatively recent time, scientists had only indirect ways to explore processes beneath volcanoes, mainly by making analogies with geological signatures of ancient magmatic systems. During the last decades, the advancement of geophysical methods made possible direct observations of deep structures beneath presently active volcanoes which give precious information about general mechanism of working the volcanic systems. Among all geophysical methods, seismic tomography is one of the most powerful and effective approaches to look at great depths below the volcanoes. Using powerful artificial and natural sources of seismic signal makes it possible getting the information from depths where the main magma sources are located.

Observation Schemes

The configurations and characteristics of observation systems in seismology depend on the properties and sizes of the target objects, as well as on the type of seismic sources. In case of volcano seismology, the parameters of the target volume depend on the type of volcanoes. For example, the subduction-related volcanoes usually have complex multilevel structures of feeding system which may cover the depths down to 100–200 km. In the case of the hot-spot volcanoes, the initial source of magmatism is related to mantle plumes which may propagate throughout the mantle; however, the magma reservoirs and the related seismicity are fairly shallow in this case (only the first kilometers of depth). When designing an experiment for tomography studies of volcanoes, one should clearly figure out the target properties and define the configuration of the network according to this. In reality, there are many limitations, such as insufficient amount of instruments and difficulties in access to some areas around a volcano that make impossible deploying an ideal network. In the following section, the major components of the observation systems are described, and some recommendations to optimize the designing of the seismic networks are given.

Seismic Sources

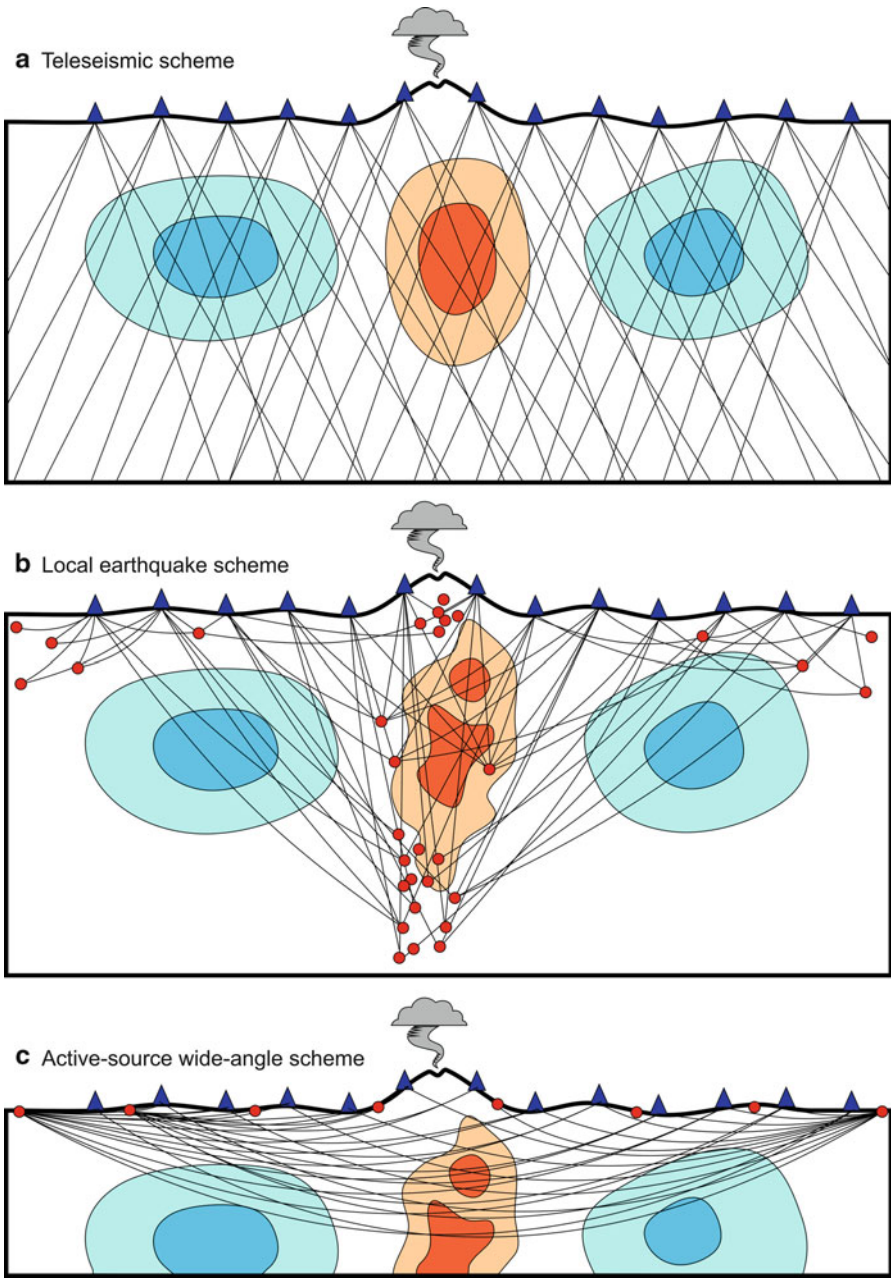
Sources used for seismic imaging of volcanoes can be divided in three categories: human-made sources, earthquakes, and ambient seismic noise. Fully controlled artificial sources are expected to provide the highest accuracy of tomographic images. At the same time, their implementation on volcanoes is very sophisticated and expensive. Another shortcoming of artificial sources is that they cannot be produced at depth. In regions with well-distributed volcanic or local tectonic seismicity, earthquakes may give a distribution of sources plausible for high-resolution seismic tomography. However, many volcanoes do not generate sustained seismicity. In this case, the recently developed methods based on correlations of ambient seismic noise become suitable for imaging volcanic edifices.

Human-Made Seismic Sources

Human-made seismic sources include explosions, vibrators, air-gun shots, and weight (or mass) drops and are usually called **active sources**. There are some onshore experiments which use chemical explosions of sufficiently large magnitude to generate seismic signal propagating to the distances of dozens kilometers and depths of a few kilometers. The problem of such explosions is that they are fairly expensive and thus their number is strongly limited. Furthermore, chemical explosions are not ecologically friendly and they are prohibited in most volcanic areas. In most cases, the surveys based on such sources are used for 2D profiling (Fig. 1c).

An alternative method of active-source onshore generation might be using vibrators which are more ecologically safe compared to explosions. However, for the purposes of volcano tomography, the vibrators should be sufficiently powerful and heavy to enable the required propagation of seismic rays. In reality, the transportation of powerful vibrators appears to be not possible in most volcanic areas; therefore, they are not widely used in practice.

In case if a volcano is located close to the seashore, it is possible to use underwater air guns as sources of seismic signal (Fig. 2a). The advantage is that air-gun shots are rather cheap



Seismic Tomography of Volcanoes, Fig. 1 Observation schemes used in volcano tomography. Blue triangles depict seismic stations; red dots are sources; black lines are seismic rays

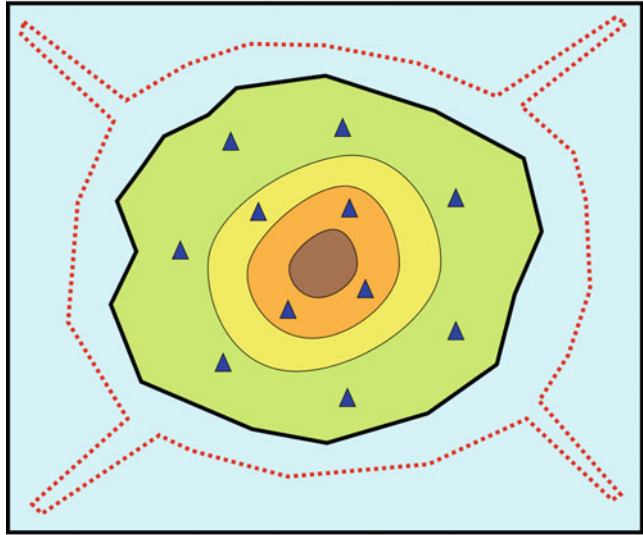
and sufficiently powerful. They can be frequently generated along the ship cruise that gives a huge data base with a clear technology of processing. The problem is that these sources are limited by the shore line and not suitable for most volcanoes.

The major advantage of the active sources is that their parameters are known. The problem is that they are located close to the surface, and seismic signal propagate along the nearly horizontal ray paths. Another problem, which is

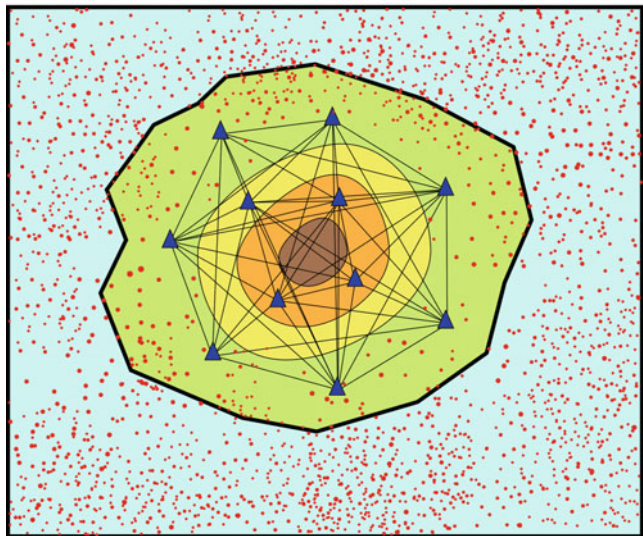
Seismic Tomography of Volcanoes,

Fig. 2 Schemes for active-source experiments using air-gun shots (**a**) and for ambient noise tomography (**b**). *Blue triangles* depict seismic stations; *red dots* are sources

a Active-source scheme (airgun shots)



b Scheme for ambient noise tomography



mostly actual for the air-gun shots, is that they generate very weak S waves which cannot be picked and used for tomography studies.

Earthquakes

Volcanic and tectonic earthquakes located in the volcanic area (including some events located outside the perimeter of the network) are regularly used as sources for seismic tomography of volcanoes (Fig. 1b). The major problem with earthquakes is that their location parameters (coordinates and origin times) are not exactly

known and must be determined simultaneously with velocity models. This leads to a coupled problem with significant trade-off between velocity and source parameters. However, an important advantage of natural sources compared to the artificial sources is that they are located at some depths which allow illuminating the target object from below at different directions. For many cases, the amount of earthquakes is much larger than that of the artificial explosions. In addition, earthquakes in most cases produce clear P and S waves which can be used for exploring P and

S velocities. In turn, both P and S velocities give much more constraints on the petrophysical state of rocks below volcanoes than just a single P model.

Ambient Seismic Noise

Ambient seismic noise is the permanent motion of the Earth surface that is not related to earthquakes or specific controlled sources (see also “► [Seismic Noise](#)”). When averaged over long time series, this wavefield can be approximately considered as produced by randomly and homogeneously distributed sources (Fig. 2b). Cross-correlation between records of such random wavefields at two stations yields the impulsive response of the media (Green’s function) between these two points (e.g., Gouédard et al. 2008). In the present form, every recorder can be converted into a “virtual” source recorded at all other recorders resulting in $N(N-1)/2$ source-receiver pairs for a network of N stations, as shown in Fig. 2b. Shortcomings of noise-based approach are (1) dependence on noise sources distribution that in many cases is not perfectly homogeneous possibly preventing successful reconstruction of impulsive responses for a part of station pairs; (2) need of long continuous records (months or years); and (3) the fact that reconstructed virtual sources are located at the surface and are dominated by surface waves.

Receivers

Networks for active and passive surveys on volcanoes use generally the same instruments as for similar seismological studies in other regions. Tomography inversion needs numerous stations in a network (more than ten) which are usually deployed temporarily. Such networks consist of portable seismic stations which are composed of three major components: sensor, recorder, and power supply. In some cases, there is also a device for telemetric transition of signal.

Seismic sensor is a kind of “microphone” which converts the ground movements into electric signal. Modern sensors which are used for volcano surveys are three-component broadband instruments which allow recording oscillations of more than 30–40 s period (see also “► [Principles](#)

of Broadband Seismometry”). It should be noted that for the body-wave tomography based on local seismicity such broad frequency capacity is not really required: the signal from local earthquakes is clearly seen at frequencies of more than 1 Hz.

Seismic recorder is a tool which digitizes the analogue electrical signal coming from the sensor and records it onto a flash or another kind of memory. In each recorder, there should be also a GPS receiver which synchronizes the absolute time. In some cases, sensors and recorders are combined in one unit.

All recorders and most broadband sensors consume the electric power, and an important part of the seismic station is **power supply**. It can include chemical rechargeable or single-used batteries, solar panels, or wind generators. Using each type of power source is determined by the conditions of real experiments. For example, in Kamchatkan volcanoes, solar panels cannot be used due to large amount of snow during the most part of the year. Furthermore, there is a risk of destruction of any part located outside the ground by wild animals or vandals. In this case, an only possibility to conserve the station is to place the station deep into the ground together with batteries.

Some Hints on the Network Design

Design of seismological networks for performing tomography studies has some particular features. Here some hints which should be taken into account while deploying seismic networks are given.

- In case of using refracted rays from active (artificial) sources, when both sources and receivers are located near the surface, a typical depth of seismic ray propagation is 1/10–1/15 of the source-receiver distance (Fig. 1c). Thus, to study the depths of ~10 km beneath the volcano, the sources and receivers should be located in the opposite sides of the volcano at distances of 50–70 km.
- The spacing between stations normally has the same order as the minimum size of resolved anomalies in the tomographic inversion.

Further decrease or increase of the resolution depends on the distribution of events. If there are a very few events, densification of the network would hardly bring to the resolution increase. In contrary, in case of a large amount of sources evenly distributed throughout the study area, the resolved patterns can be smaller than the station spacing.

- For passive sources (local earthquakes), the depth of seismic model resolved from tomographic inversion cannot be larger than the maximum depths of sources.
- For uncontrolled sources (earthquakes), a critical parameter for tomography is a number of picks per events. If this number is equal to 4, one can theoretically locate the sources but cannot get any information about the velocity model. In practice, the data are noisy, and to get a stable location of source, more than four picks are required. The information on the velocity structure can be extracted only if the number of picks is more than 10. The higher this value is, the more velocity parameters can be retrieved and the higher resolution model will be obtained.
- Using out-of-network events (within reasonable limits) can appear to be useful for tomography as was shown by Koulakov (2009b). Although the location accuracy for such events is fairly low, they improve the ray coverage and help in increasing the spatial resolution beneath the network area.
- In all cases, prior the deployment of the network, it is useful to simulate a series of synthetic models with possible configurations of network, sources, and target objects that are expected in the real case. This can help to optimize the distribution of seismic stations in the real experiment.

Seismic Tomography Methods Based on Body Waves

Basic Principle of Body-Wave Seismic Tomography

In the ray-based approach which is used in most body-wave tomography algorithms, seismic

waves are produced by point sources. Along this way, seismic rays accumulate the information on the inner structure of the Earth. The purpose of seismic tomography is to decipher this information and to reconstruct the structure inside the Earth. Most simple and most popular way for many practical applications, including volcano tomography, is the kinematic approach based on using travel times of seismic rays. Travel times of seismic rays can be represented as an integral along the ray paths

$$T = \int_{\gamma} \frac{dS}{V(x, y, z)} \quad (1)$$

where γ is the ray path and V is seismic velocity. Here, an infinite frequency of the wave is presumed which allows presenting the ray as a curved line. Actually, this is often not true in seismology where the wavelengths are in some cases compatible with the size of the target objects. For these cases, there are methods of finite-frequency tomography which are based on “fat” rays representing a banana-shaped area having a rather complex sensitivity patterns (e.g., Husen and Kissling 2001). Although there are some successful applications of this method in global or in teleseismic tomography (Montelli et al. 2004), it was not yet widely implemented for volcano tomography. Up to today, most of the practical applications for studying the volcanoes are based on the ray-based kinematic approach.

In general, the velocity field in Eq. 1 is a function which is represented by infinite number of mathematical points. In practice, to solve the inverse problem, i.e., to derive velocity distribution based on given travel times, the velocity field should be parameterized with a finite number of parameters. There are different methods of parameterization, such as using regular or irregular cells, nodes with linear interpolation, harmonic functions with unknown coefficients, etc.

Another problem of solving the inverse problem based on the representation in Eq. 1 is that this problem is nonlinear. It means that one should find the velocity distribution based on the ray paths which are strongly affected by the

unknown velocity distribution. The direct solution of this nonlinear problem is almost never used for practical applications. Actually, in most cases, this problem is reduced to the linear representation. It is assumed that the unknown velocity distribution is close to the given reference model, and in this case, the problem in Eq. 1 is reduced to the linear integral

$$\Delta t = - \int_{\gamma_0} \frac{\Delta v dS}{V_0^2(x, y, z)} \quad (2)$$

where Δv is unknown velocity anomalies, Δt is time residuals, V_0^2 is velocity in the reference model, and γ_0 is the ray path in the given reference model. Then, the continuous representation in Eq. 2 can be reduced into the discrete system of linear equations:

$$\Delta t_i = \sum_j A_{ij} \Delta v_j \quad (3)$$

where Δv is the unknown value of j th velocity parameter and A_{ij} is the first-derivative matrix which represents an effect of a unit velocity change of j th parameter on travel time i th ray. In the case of cell parameterization, each element of this matrix is just the length of j th ray in i th cell. In the case of uncontrolled sources, this matrix also includes the elements for relocation of sources (four parameters, dx , dy , dz , and dt , for each source). In practical tomography studies, where this matrix may appear to be very large, the system in Eq. 3 is solved using the iterative LSQR algorithm (Paige and Saunders 1982) which proceeds only nonzero elements in the matrix and performs the calculations line by line without keeping the entire matrix in the memory. The direct solving of the system as it stands in Eq. 3 normally fails because of strong instability of the solution due to noise in the data. To avoid this problem, the inversion should be regularized using some additional terms. The most popular way of regularization is the amplitude damping which is realized by adding a diagonal matrix block with zero data vector. Another way is damping a gradient between all pairs of

neighboring parameters (nodes or cells) that allows controlling the smoothness of the solution.

The quality of the solution and optimal values of the inversion parameters are estimated based on synthetic modeling which should simulate the conditions of a real experiment as close as possible. The synthetic travel times are computed in an artificial model and perturbed by a realistic noise. Then, the model parameters are “forgotten,” and the reconstruction procedure is performed identically as in the case of observed data processing. This step is especially important for the volcano tomography where the distributions of sources and events are often far from optimal.

There are different seismic tomography schemes which are used for different purposes in geosciences. Many of them are used for studying the volcano feeding systems on scales from the entire mantle (thousands kilometers) to the uppermost layers (hundreds meters) (e.g., Koulakov 2013). Here, only some examples of tomography studies on scales from first to dozens kilometers are given. They provide the information about magma sources in the crust and uppermost mantle that are directly responsible for eruptions of active volcanoes.

Teleseismic Tomography

Seismic tomography is a relatively young method which started actively developed from the late seventies after the pioneer work by Aki et al. (1977). The method proposed in this study is called **teleseismic tomography**, and it uses the relative time delays from distant (teleseismic) earthquakes recorded by seismic stations located in the study area (Fig. 1a). There were several attempts of using this method for studying the volcanoes. For example, Ellsworth and Koyanagi (1977) have implemented the teleseismic scheme for studying the crustal structure beneath Kilauea volcano in Hawaii. Three years later, Sharp et al. (1980) have published the tomographic model for Mt. Etna. Among later studies, one can mention the work by Stauber et al. (1988) who presented tomographic models of Newberry Volcano, Oregon, based on teleseismic delays.

It should be noted that during the last decades, the teleseismic tomography is almost not used

any more for studies of crustal structures beneath volcanoes because of several reasons. The first problem is that nearly vertical seismic rays in the study area cause poor vertical resolution which leads to strong vertical smearing of the retrieved anomalies. Another problem, which seems to be more serious, is related to the fact that the teleseismic waves are usually of low frequency (around 1 Hz and lower). Thus, the large wavelengths strongly limit the resolution capacity of this method. Theoretically, it cannot provide robust images of anomalies smaller than 8–15 km. Magma reservoirs are usually expected to be smaller in size, and thus, they are hardly resolvable by teleseismic tomography.

Local Earthquake Tomography

Another method, which is called **local earthquake tomography (LET)**, uses the data of arrival times of P and S waves from earthquakes located within the study area or slightly outside (Fig. 1b). Because of using the uncontrolled sources, the solution in this case is reduced to the coupled problem of simultaneous determination of source parameters and velocity models. Active volcanoes usually produce a lot of seismicity related to processes in magmatic reservoirs and tectonic displacements in the crust. Therefore, deploying stations for the LET experiment is much cheaper than organizing active-source works with the use of explosions or air-gun shots. Furthermore, the earthquakes in passive source schemes are distributed at some depths that enable much more favorable observation system than in the case of active-source studies. In addition, seismic records from earthquakes usually provide the data on P and S waves that allows getting more information on petrophysical state of rocks in volcanoes than just a single P-velocity model which is usually resulted from the active-source studies. The problem of LET studies is an uncontrolled distribution of sources which is not always optimal to get optimal ray coverage in the target areas. An important difficulty of LET is the uncertainty of the source locations and the trade-off between source and velocity parameters. Another technical problem of LET is the necessity of fairly long

deployment of a large number of seismic stations (for more than several months) which is required to accumulate sufficient information on the seismicity in the volcano area. This needs solving some logistical issues related to providing power supply and hiding the equipment.

The first LET algorithm was developed and implemented to real seismological data by Thurber (1983). One year later, the same algorithm was used for studying the crustal structure beneath Kilauea volcano (Thurber 1984). Among the later studies based on this approach, one can single out the studies of Hengill-Grensdalur volcanic complex in Iceland (Foulger and Toomey 1989) and of Mount St. Helens (Lees 1992). There are more than ten different LET studies of Mt. Etna which appears to be one of the best studied volcanoes in the world (e.g., Villasegnor et al. 1998; Chiarabba et al. 2000; Patane et al. 2006). Note that the work by Patane et al. (2006) was a pioneer study presenting the 4D tomography model which revealed temporal variations of seismic structure beneath the volcano having possible relations to the eruption activity (see also “► [Seismic Monitoring of Volcanoes](#)”). During the last decades, the LET methods have become the most popular for studying crustal structure beneath volcanoes. Nowadays, there are dozens of successful studies of volcanoes in different parts of the world. In the final part of this paper, two examples of recent studies of Mt. Spurr in Alaska (Koulakov et al. 2013a) and of Klyuchevskoy volcano group in Kamchatka (Koulakov et al. 2011, 2013b) will be presented. Both of these studies were performed using the LET code LOTOS (Koulakov 2009a).

Studies with Human-Made Sources

Human-made sources or active sources are used in many seismic studies of volcanoes. The first experiments of three-dimensional active-source investigations of volcanoes were made in two relict calderas located in the Cascade Range, western USA. One of them is the potentially active Newberry shield volcano with the caldera of ~8 km diameter. In the active seismic experiment, Achaer et al. (1988) used three chemical explosions and a dense receiver network covering the caldera. They obtained a rather clear

low-velocity anomaly at about 3 km depth which was interpreted as a magma chamber. Another study was made in the same year in the Medicine Lake Caldera by Evans and Zucca (1988). They deployed 120 seismometers and made eight chemical explosions which were recorded by all stations. The tomography inversion of travel time data revealed a low-velocity pattern beneath the caldera which was also interpreted as a magma source.

Performing 3D experiments using onshore explosions appeared too expensive and not providing sufficient spatial resolution with the available number of explosions. Therefore, many onshore experiments were mostly organized along one or several profiles (Fig. 1c). For example, Zollo et al. (1996) used four chemical explosions on a profile passing through the summit of the Vesuvius volcano (Italy). Their P-velocity model reveals a rather clear high-velocity anomaly beneath the volcano. A similar experiment, but with five explosions on two crossing profiles, was performed by Aoki et al. (2009) for the Asama volcano in Japan. Similarly, as in the previous case, they found a clear high-velocity body beneath the volcano.

It should be noted that logistically the experiments based on chemical explosions are very difficult. Because of high expenses, the number of shots is strictly limited that makes it problematic to illuminate the target objects from different directions. Furthermore, the 2D approximations are often too rough for the volcanic areas where the expected structures are strongly 3D. In addition, the explosion type sources are not ecologically safe, and in most volcanic areas, they are prohibited now.

Good solutions for areas which are close to the seashore are the underwater air-gun shots which can be used to generate high-quality and sufficiently powerful seismic signal and can surround the target area (Fig. 2a). Nowadays, this scheme is widely used in experiments in many volcanic areas. Among many different studies, three examples are selected here:

- A very interesting site for seismic investigations is the volcanic Deception Island in

Antarctic which was studied by Zandomenighi et al. (2009). This island has a circular shape and is suitable for performing air-gun shots inside and outside the circle. Using onshore and offshore seismometers made this observation system suitable for high-resolution tomography studies. The inversion of travel time data revealed the seismic structures down to ~ 2 km which appeared to be linked with the main geological structures. The circular rim of the caldera is associated with high-velocity lineaments. Inside the caldera, a low-velocity anomaly is observed.

- A network with onshore and offshore stations and 4,414 air-gun shots were used by Paulato et al. (2010) to build a tomographic model beneath Montserrat Island in Lesser Antilles. They found that the location of the Island is associated with high P velocity. Especially sharp positive anomaly is located beneath the Soufriere volcano.
- The high-resolution 3D structure beneath the Tenerife Island was studied by Garcia-Yeguas et al. (2012) based on a large amount of air-gun shots and fairly dense onshore seismic network consisting of 140 stations. The main volcanic structure of the Las Cañadas-Teide-Pico Viejo Complex (CTPVC) is characterized by a high P-wave velocity body, similar to many other stratovolcanoes. Furthermore, reduced P-wave velocities are found in a small confined region in CTPVC and are more likely related to hydrothermal alteration, as indicated by the existence of fumaroles.

Imaging and Monitoring of Volcanoes Based on Ambient Seismic Noise

General Principle, Data Processing, and Inversion Approaches

New methods for probing the Earth's interior using noise records only emerged during the last decade (e.g., Campillo et al. 2011). Origin of seismic noise strongly depends on the considered spectral range (see also "► Seismic Noise"). At high frequencies (>1 Hz), the noise is strongly dominated by local sources that are often related

to strong winds or the anthropogenic activity. Seismic noise at intermediate periods (between 1 and 20 s) that is most often used for studies of volcanoes is dominated by natural sources. In particular, the two main picks in the seismic noise spectra in this so-called microseismic band (1–20 s) are related to forcing from oceanic gravity waves. The interaction between these oceanic waves and the solid Earth is governed by a complex nonlinear mechanism (Longuet-Higgins 1950), and the noise excitation depends on many factors such as the intensity of the oceanic waves but also the intensity of their interferences as well as the seafloor topography. Overall, the generation of seismic noise is expected to be strongly modulated by strong oceanic storms and, therefore, to have a clear seasonal and nonrandom pattern. However, when averaged over long time series, distribution of sources of microseismic noise can be approximated as random and homogeneous. In this case, cross-correlation of noise records between two stations yields the impulsive response (Green's function) between these two points (e.g., Gou  ard et al. 2008). In the case of a uniform spatial distribution of noise sources, the cross-correlation of noise records converges to the complete Green's function of the medium, including all reflection, scattering, and propagation modes. However, in the case of the Earth, most of ambient seismic noise is generated by atmospheric and oceanic forcing at the surface. Therefore, the surface-wave part of the Green's function is most easily extracted from the noise cross-correlations.

Seismic surface waves, in contrast with body waves, are waves that propagate in a waveguide near the Earth's surface. There are two types of **surface waves**: **Rayleigh** (vertically polarized waves) and **Love** (horizontally polarized waves transverse to the direction of motion). Their depth sensitivity depends on frequency, with a fair approximation being about a third of a wavelength. As a result, the surface waves are strongly dispersive with longer periods (wavelengths) propagating faster than shorter periods because of the general increase of seismic speed with depth in the Earth.

Using surface waves extracted from correlations of seismic noise for imaging of the Earth's interior is called **ambient noise surface-wave tomography**. This approach generally consists of three steps (e.g., Ritzwoller et al. 2011). First, noise cross-correlations between every pair of stations in the array are computed, and interstation phase or group travel times as a function of period are measured from the resulting waveforms. Second, two-dimensional maps of the speed of Rayleigh or Love waves as a function of frequency are produced. At the last step, these maps are inverted for a 3D distribution of shear speeds.

Applications of Ambient Noise Seismic Tomography (ANST) to Studies of Volcanoes

After recent emergence of the ambient noise surface-wave tomography (e.g., Shapiro et al. 2005), it became clear that this method might be very useful for studying structures within volcanoes because of its relatively low cost (especially comparing with active-source studies) and because of good coverage of the shallow part of edifices compared with earthquake-based body-wave imaging. First "volcanic" application of ANST was performed at Piton de la Fournaise located at La R  union Island (Brennguier et al. 2007). This hot-spot volcano is very active with on average more than one eruption per year during the past two centuries. The noise-based tomography was based on correlations of 18 months of records by 21 short-period vertical-component stations of the permanent volcano monitoring network. Rayleigh wave group velocities could be measured between 2 and 5 s of period. Inversion of these dispersion curves resulted in a shear velocity model extending between the surface and the depth of 1 km below sea level (while average elevation of the volcano is ~ 2 km). The main result of this study was imaging of an intrusive high-velocity body below the main crater moving westward from the surface to 1 km below sea level. This body is located above the shallow magma reservoir and coincides with the region through which the magma ascends during the summit eruptions.

Following the pioneering study of Brennguier et al. (2007), the ANST was applied to image

subduction zone volcanoes: Okmok in Aleutian Islands (Masterlark et al. 2010), Lake Toba in Sumatra (Stankiewicz et al. 2010), and Asama in Japan (Nagaoka et al. 2012). The first study used 12 stations from a permanent monitoring network, and the second study used 40 stations from a temporary deployment, while the third study used 81 sensors combined from permanent and temporary networks. In all these studies, the authors reported very similar results: imaging of low-velocity anomalies corresponding to magma reservoirs at depths below 4 km.

Noise-Based Volcano Monitoring

One of the advantages of using continuous seismic noise records to characterize the earth materials is that a measurement can easily be repeated (see also “► [Seismic Monitoring of Volcanoes](#)”). First application of this idea to volcano monitoring was done by Sens-Schönfelder and Wegler (2006) who used the repetitive waveforms of seismic noise cross-correlations to track changes in the subsurface material properties on mount Merapi in Indonesia during 2 years. This study was based on high-frequency (>1 Hz) waves and mainly revealed seasonal variations caused by changes of hydrological conditions in a very shallow layer. Brenguier et al. (2008) used noise cross-correlations at intermediate periods (between 1 and 10 s) to detect decreases in seismic velocity a few weeks before eruptions that suggests preeruptive inflation of the volcanic edifice of the Piton de la Fournaise. A comprehensive review of recent advancements in the noise-based volcano monitoring is provided by Brenguier et al. (2011) (see also “► [Tracking Changes in Volcanic Systems with Seismic Interferometry](#)”).

Current Limitations and Prospective

Application of seismic noise cross-correlation methods for studies of volcanoes is a rapidly growing field. However, these methods have some important limitations that are mainly related to the properties of the seismic noise recorded at the Earth surface. So far, the basic theoretical foundation of the cross-correlation method assumes a random and homogeneous

distribution of the wavefield sources, while properties of a real seismic noise may deviate from this simplified model. This implies that applications of the noise-based imaging or monitoring should be accompanied by studies of properties of recorded seismic noise to characterize the degree of its spatial and temporal inhomogeneity and to determine the minimal time required for good convergence of noise cross-correlations. Development of advanced methods for the preprocessing of the noise time series in order to improve the quality of the extracted signals and to decrease the convergence time is a very active research area.

Up to date, the noise-based seismic tomography was based on surface waves that are most easily extracted from noise cross-correlations but have a limited resolution in deep layers. Most recent results demonstrate that, when working with large and dense arrays of seismometers, it is possible to indentify body waves in correlation waveforms and to measure their travel times (e.g., Boué et al. 2013; Lin et al. 2013). This gives a hope that noise-based body-wave tomography of volcanoes could be developed in the future.

Examples of Tomography Studies of Volcanoes

Tomography of Mt. Spurr

Mount Spurr is an active predominantly andesitic volcano in Alaska located at the northeastern end of the Aleutian arc. Before a single eruption in 1953, it was considered as a dormant volcano. In 1992, a series of three explosive eruptive events occurred on June 27, August 18, and September 16–17 in a small composite cone forming the Crater Peak on the south flank of Mount Spurr. The structure beneath Mt. Spurr in a period of time nearly corresponding to this eruption was studied by Power et al. (1998) using the data collected from 1991 to 1993 by 7–11 seismic stations of the Alaska Volcano Observatory (AVO).

The most recent episode of Mt. Spurr unrest occurred in 2004–2006 which accompanied with

strong seismicity and fumarolic activity. However, this episode did not result at any magma eruption, and thus, it was called as “lost eruption.” For 1 month in June 2005 corresponding to the middle stage of this unrest, the AVO deployed a fairly dense seismic network consisting of 26 broadband three-component stations which operated in addition to 11 short-period stations of the permanent network. During this period, 512 crustal earthquakes were located in the area of Mt. Spurr using P and S picks recorded by portable and permanent stations. The final catalog contained 5,960 P- and 4,973 S-wave arrival times with an average of 20 picks per event. These data were used to perform the tomographic inversion presented in Koulakov et al. (2013a).

The result of tomographic inversion including the 3D distributions of P and S velocities, V_p/V_s ratio, and accurate source locations is presented in Fig. 3. The horizontal section at 5 km depth demonstrates a prominent anomaly of high V_p/V_s ratio which reaches the value of 1.9 located beneath the northern rim of the volcano. This pattern coincides with the distribution of shallow seismicity. In the deeper section at 22 km depth, this anomaly is located at the same place, but it appears to be less intensive than another zone of high V_p/V_s ratio which is collocated with a cluster of deep crustal earthquakes. In vertical section, these two anomalies look like two fingers of high V_p/V_s ratio separated by 8 km of lateral distance. One of the two anomalies nearly reaches sea level beneath the northern border of the caldera, but is separated from the surface with a zone of low V_p/V_s ratio. The top of the second anomaly appears to be much deeper and may reach a depth of 20 km. The earthquakes that occurred during the 3-month span show distinct clusters near the tops of each of these fingers which support the active behavior of these structures. It is important that for both these finger-shaped anomalies, the high V_p/V_s correspond to higher P and lower S velocities.

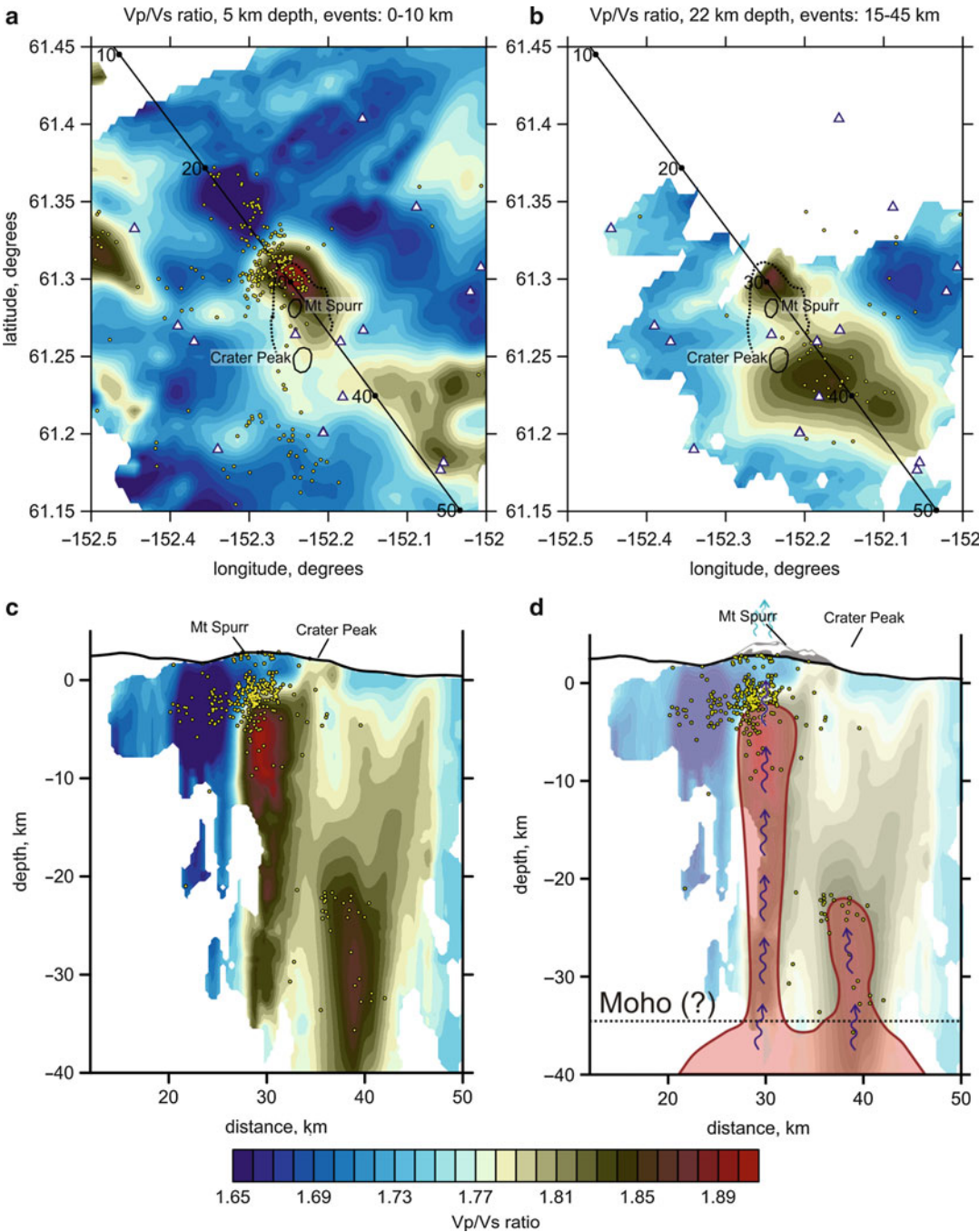
The observed finger-like features are suggestive of pathways through the crust and possibly correspond to complexes of conduits that transport liquid volatiles and/or melts from the mantle.

The combination of high V_p and low V_s suggests the presence of liquid fluids and/or melts and the rock composition corresponding to the low crustal or even mantle levels. The existence of very low values of V_p/V_s above the NW anomaly might be an indicative for aggregate transition occurred between the top of the plume and the surface. It is important that the NW finger location generally coincides with the presence of fumaroles and active volatile flux during the most recent eruption in 2004–2005. It can be proposed that this liquid fluids brought by the NW conduit are transformed to gases at shallow depths which can escape to the surface without producing any stresses. This explains lack of any explosive activity in this location during the 2004–2005 unrest of Mt. Spurr (Fig. 3).

Another conduit located below the Crater Peak seems to be blocked at ~ 20 km depth. This fact is supported by the occurrence of deep seismicity just above this anomaly. In this blocked channel, the accumulation of high pressure is more probable which makes it more dangerous for a potential explosive eruption. Note that the explosive eruption of Mt. Spurr in the SE rim in 1992 occurred just above this anomaly.

Tomography of the Klyuchevskoy Volcano Group

The Klyuchevskoy volcano group is located in Kamchatka, in the far east of Russia. It includes 13 active and dormant volcanoes of various types and compositions located in a relatively small area of approximately 100×60 km. The main volcano of the group, Klyuchevskoy Sopka, has the elevation of $\sim 4,800$ m, and it is the highest volcano in continental Eurasia. Another volcano of the group, Bezymianny, located at approximately 10 km distance from the Klyuchevskoy volcano is an explosive-type andesitic volcano. One of the world's largest explosive eruptions in the twentieth century occurred here in 1957. The Tolbachik complex is situated in the southwestern part of the Klyuchevskoy group, and it consists of several shield and stratovolcanoes of predominantly basaltic compositions. Plosky Tolbachik, which is currently active, is a typical Hawaiian-type volcano with fissure eruptions and



Seismic Tomography of Volcanoes, Fig. 3 Distribution of the V_p/V_s ratio in two horizontal and one vertical section. Dotted line depicts the limits of the caldera rim. Triangles indicate the locations of stations. Yellow dots mark the source locations. Bottom-right plot represents

the interpretation of the results. Blue wavy arrows indicate hypothetical migration of fluids and their escape to the atmosphere in 2004–2006 unrest (Reproduced from Koulakov et al. 2013a, JRL)

calderas. The existence of so many different volcanoes located close to each other can most likely be explained by complex processes of fractioning, mixing, and melting in magmatic reservoirs located under the volcanic group at different depth levels.

During the last several decades, the volcano-related seismicity in the Klyuchevskoy group has been monitored by the Kamchatkan Branch of Geophysical Survey. After 1999, a network consisting of 17 digital telemetric stations was installed in the region of the Klyuchevskoy group. From 1999 to 2009, it provided the information on more than 80 thousand events and the corresponding half million of P and S arrival times. The tomographic inversions were performed based on selected yearly subsets. The temporal variations of seismic structure are especially interesting in the context of the volcanic activity during the considered time. In particular, in 2005, there were simultaneous eruptions of the Klyuchevskoy and Bezymianny volcanoes that give a possibility to explore the changes of the crust related to the volcanic activity at stages of preparation, activation, and relaxation (see also “[Seismic Monitoring of Volcanoes](#)”).

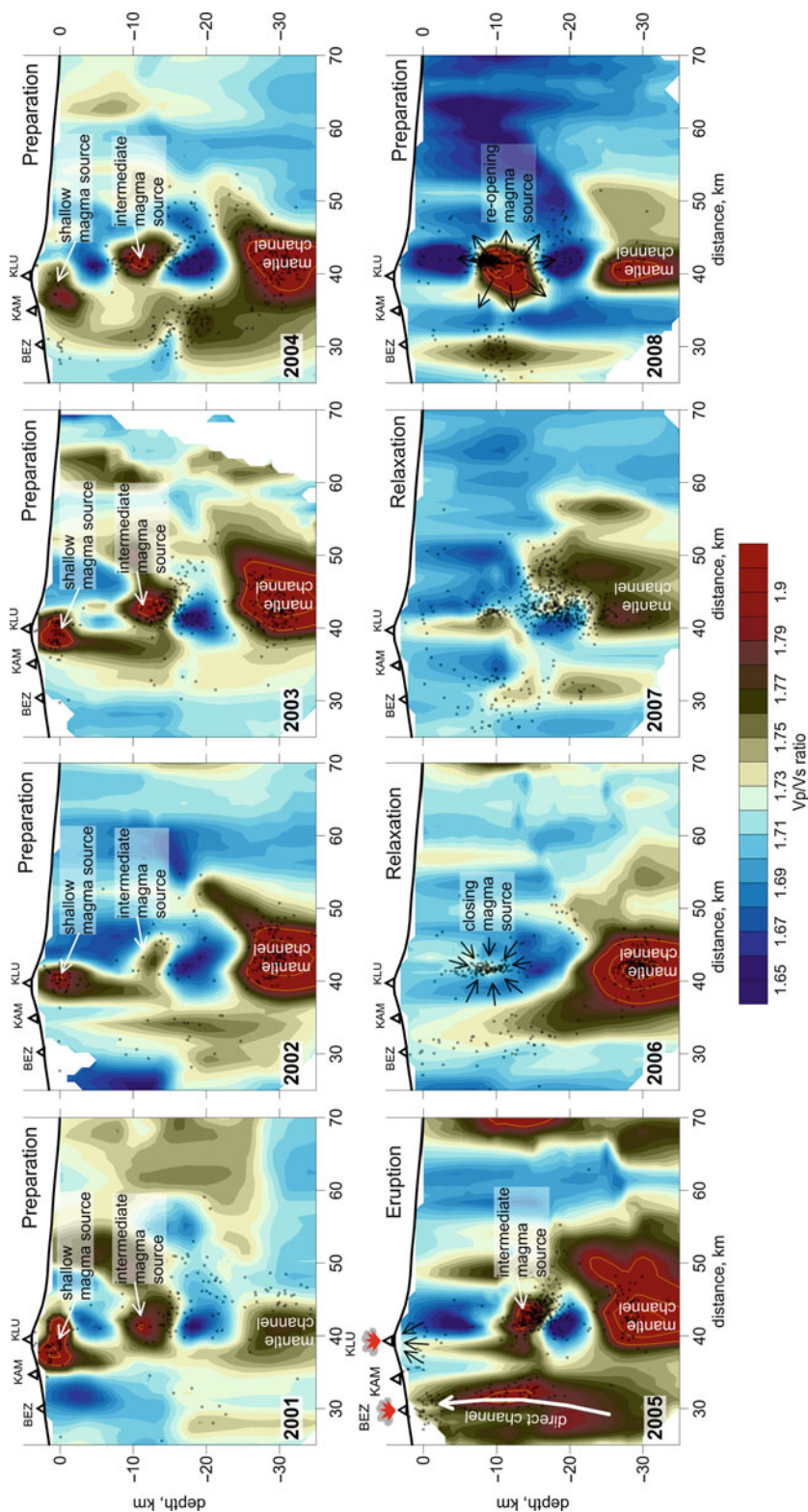
The results of tomographic inversion corresponding to the year 2004 just before the activation phase were presented in a separate study by Koulakov et al. (2011). Similar structures are observed in the precedent 2003 year (Koulakov 2013). The key feature of these models is a strong anomaly of a high V_p/V_s ratio, reaching values of 2.2, located below 25 km depth. Within this anomaly, the P velocity appears to be higher than average, whereas the S velocity is strongly low. This combination of seismic anomalies is a possible indicator for a mantle conduit beneath the Klyuchevskoy group with rocks brought from deep levels and high content of fluids and melts. It was proposed that this conduit is the original magma source that feeds all of the volcanoes in the Klyuchevskoy group.

The tomography result corresponding to 2004 reveals complex contrasted structures in the crust which can be associated with magma sources. The distributions of high values of V_p/V_s ratio

reveal at least two levels of magma sources in the crust: one is located at a depth of 10–12 km and another is just below the Klyuchevskoy volcano close to the surface. This may explain the diversity of the volcanic styles in the group. The distribution of earthquakes appears to link different levels of magma storage that may trace the paths of magma migration.

Time variations of seismic structure beneath the Klyuchevskoy group in a time period from 1999 to 2009 were studied by Koulakov et al. (2013b) and presented in Fig. 4. A common feature which is seen in all yearly windows is the anomaly with an extremely high V_p/V_s ratio located below 25 km depth, which was previously mentioned for the year 2004 and interpreted as a mantle conduit feeding all volcanoes of the group. For the crust, this study demonstrates considerable temporal changes of seismic structures. In particular, prior the 2005 eruptions of the Klyuchevskoy and Bezymianny volcanoes, the seismic structures in the crust look generally similar and contain two anomalies of high V_p/V_s ratio beneath the Klyuchevskoy volcano at a depth of 10–12 km and close to the surface. As mentioned above, these anomalies are interpreted as two levels of magma sources. In 2005, the average value of the V_p/V_s ratio increases, and it may be explained by a considerable increase of the fluid and melt content in the crust. In this year, a strong anomaly of high V_p/V_s ratio is observed beneath the Bezymianny volcano, and it appears to link the volcano with mantle sources. In other years, this anomaly was not observed. The shallow anomaly of high V_p/V_s anomaly, which was previously observed beneath the Klyuchevskoy volcano, disappears in 2005. The deeper anomaly at 10–12 km depth is still presented. In the years after the eruption, 2006 and 2007, the average V_p/V_s ratio lowers, and all the patterns at 10–12 km depth that were interpreted as intermediate magma sources in the crust are not observed anymore. Three years later, in 2008, a similar anomaly at the same location starts reappearing again.

The observed variations of the seismic properties are probably related to changes in the stress and deformation fields, migration of fluids, and



Seismic Tomography of Volcanoes, Fig. 4 Cross sections of V_p/V_s from tomographic inversions for each year from 2001 to 2008 for a profile passing through Klyuchevskoy (KLU) and Bezmianny (BEZ) volcanoes. In areas of high V_p/V_s ratios, yellow lines show contours of $V_p/V_s = 2.0, 2.1$, and 2.2 . Dots show earthquakes less than 0.5 km from the profile, and triangles mark the locations of the volcanoes near the profile. The eruptions of Klyuchevskoy and Bezmianny volcanoes in 2005 are highlighted. The main phases and structural elements, which are discussed in the text, are indicated (Reproduced from Koulakov 2013, JVGR)

aggregate transformations in the crust. The mantle conduit, which is visible as an anomaly of high V_p/V_s ratio below 25 km depth at all time periods, should have strong mechanical, thermal, and/or chemical influence upon the bottom of the crust and cause strong seismicity. These stresses lead to the origin of fractures and upward penetration of the fluids to the crust. Reaching intermediate levels, these fluids lead to melting of the overheated rocks. Due to decompression, some of the fluids are transformed to gases that increase the pressure and cause additional fractures. These avalanche-type processes finally lead to the eruptions. During the eruption, most of the fluids escape to the atmosphere, and it causes a decrease in the melt content in the intermediate magma sources in the post-eruption period. Several years later, a new portion of fluids comes from the mantle conduit, and melting of rocks in intermediate reservoirs starts again. An important conclusion that can be drawn from the 4D tomography study is that the magmatic sources appear to be dynamic systems with rapidly variable properties. According to this concept, the magma sources are composed of sponge-type materials with the capacity to change the proportion of the molten partition depending on the presence of fluids. Propagation of cracks and fluids may occur rather rapidly, and this may abruptly change the properties of the “sponge” material over a period of years or even months or weeks.

Summary

This paper presents an overview with the description of several seismic tomography schemes which are used for studying volcanoes. It describes several examples of practical works including two featured case studies of Mt. Spurr and Klyuchevskoy volcano group. It was demonstrated that seismic tomography gives rich material which allows better understanding of the behavior of magmatic systems. A variety of the tomographic solutions show that there is no uniform structure: each volcano has some particular features and appears to be unique.

During the last years, there appear more and more dense networks on different volcanoes of the world designed for passive and active experiments. There are also new algorithms actively developed during the last years, such as scattering tomography or full waveform inversion, which might appear very effective for studying volcanoes. This gives a hope that increasing amount of data and implementing new algorithms will lead in the nearest time to a real breakthrough in understanding the causes of volcanic activity.

Acknowledgments The contribution of Ivan Koulakov is supported by the Russian Scientific Foundation (grant #14-17-00430).

References

- Achauer U, Evans JR, Stauber DA (1988) High-resolution seismic tomography of compressional wave velocity structure at Newberry Volcano, Oregon Cascade Range. *J Geophys Res Solid Earth* (1978–2012) 93(B9):10135–10147
- Aki K, Christofferson A, Husebye ES (1977) Determination of the three-dimensional seismic structure of the lithosphere. *J Geophys Res* 82(2):277–296
- Aoki Y, Takeo M, Aoyama H, Fujimatsu J, Matsumoto S, Miyamachi H, Yamawaki T (2009) P-wave velocity structure beneath Asama Volcano, Japan, inferred from active source seismic experiment. *J Volcanol Geotherm Res* 187(3):272–277
- Boué P, Poli P, Campillo M, Pedersen H, Briand X, Roux P (2013) Teleseismic correlations of ambient seismic noise for deep global imaging of the Earth. *Geophys J Int* 194(2):844–848. doi:10.1093/gji/ggt160
- Brenguier F, Shapiro NM, Campillo M, Nercissian A, Ferrazzini V (2007) 3-D surface wave tomography of the Piton de la Fournaise volcano using seismic noise correlations. *Geophys Res Lett* 34:L02305. doi:10.1029/2006GL028586
- Brenguier F, Shapiro N, Campillo M, Ferrazzini V, Duputel Z, Coutant O, Nercissian A (2008) Toward forecasting volcanic eruptions using seismic noise. *Nat Geosci* 1(2):126–130
- Brenguier F, Clarke D, Aoki Y, Shapiro NM, Campillo M, Ferrazzini V (2011) Monitoring volcanoes using seismic noise correlations. *C R Geosci* 343:633–638. doi:10.1016/j.crte.2010.12.010
- Campillo M, Roux P, Shapiro NM (2011) Correlations of seismic ambient noise to image and to monitor the Solid Earth. In: Gupta HK (ed) *Encyclopedia of solid earth geophysics*. Springer, Dordrecht, pp 1230–1235

- Chiarabba C, Amato A, Boschi E, Barberi F (2000) Recent seismicity and tomographic modeling of the Mount Etna plumbing system. *J Geophys Res Solid Earth* (1978–2012) 105(B5):10923–10938
- Ellsworth WL, Koyanagi RY (1977) Three-dimensional crust and mantle structure of Kilauea Volcano, Hawaii. *J Geophys Res* 82(33):5379–5394
- Evans JR, Zucca JJ (1988) Active high-resolution seismic tomography of compressional wave velocity and attenuation structure at Medicine Lake Volcano, Northern California Cascade Range. *J Geophys Res Solid Earth* (1978–2012) 93(B12):15016–15036
- Foulger GR, Toomey DR (1989) Structure and evolution of the Hengill-Grensdalur Volcanic Complex, Iceland: geology, geophysics, and seismic tomography. *J Geophys Res Solid Earth* (1978–2012) 94(B12):17511–17522
- García-Yeguas A, Koulakov I, Ibáñez JM, Rietbrock A (2012) High resolution 3D P wave velocity structure beneath Tenerife Island (Canary Islands, Spain) based on tomographic inversion of active-source data. *J Geophys Res* 117, B09309, doi:10.1029/2011JB008970
- Gouéard P, Stehly L, Brenguier F, Campillo M, Colin de Verdière Y, Larose E, Margerin L, Roux P, Sanchez-Sesma FJ, Shapiro NM, Weaver RL (2008) Cross-correlation of random fields: mathematical approach and applications. *Geophys Prospect* 56:375–393
- Husen S, Kissling E (2001) Local earthquake tomography between rays and waves: fat ray tomography. *Phys Earth Planet In* 123(2):127–147
- Koulakov I (2009a) LOTOS code for local earthquake tomographic inversion. Benchmarks for testing tomographic algorithms. *Bull Seismol Soc Am* 99(1):194–214. doi:10.1785/0120080013
- Koulakov I (2009b) Out-of-network events can be of great importance for improving results of local earthquake tomography. *Bull Seismol Soc Am* 99(4):2556–2563. doi:10.1785/0120080365
- Koulakov I (2013) Studying deep sources of volcanism using multiscale seismic tomography. *J Volcanol Geotherm Res* 257:205–226. doi:10.1016/j.jvolgeores.2013.03.012
- Koulakov I, Gordeev EI, Dobretsov NL, Vernikovsky VA, Senyukov S, Jakovlev A (2011) Feeding volcanoes of the Klyuchevskoy group from the results of local earthquake tomography. *Geophys Res Lett* 38: L09305. doi:10.1029/2011GL046957
- Koulakov I, West M, Izbekov P (2013a) Fluid ascent during the 2004–2005 unrest at Mt. Spurr inferred from seismic tomography. *Geophys Res Lett*. doi:10.1002/grl.50674.
- Koulakov I, Gordeev EI, Dobretsov NL, Vernikovsky VA, Senyukov S, Jakovlev A, Jaxybulatov K (2013b) Rapid changes in magma storage beneath the Klyuchevskoy group of volcanoes inferred from time-dependent seismic tomography. *J Volcanol Geotherm Res* (available online). doi:10.1016/j.jvolgeores.2012.10.014
- Lees JM (1992) The magma system of Mount St. Helens: non-linear high-resolution P-wave tomography. *J Volcanol Geotherm Res* 53(1):103–116
- Lin FC, Li D, Clayton RW, Hollis D (2013) High-resolution 3D shallow crustal structure in Long Beach, California: application of ambient noise tomography on a dense seismic array. *Geophysics* 78(4):Q45–Q56. doi:10.1190/geo2012-0453.1
- Longuet-Higgins MS (1950) A theory of the origin of microseisms. *R Soc Lond Philos Trans Ser A* 243:1–35
- Masterlark T, Haney M, Dickinson H, Fournier T, Searcy C (2010) Rheologic and structural controls on the deformation of Okmok volcano, Alaska: FEMs, InSAR, and ambient noise tomography. *J Geophys Res* 115:B02409. doi:10.1029/2009JB006324
- Montelli R, Nolet G, Dahlen FA, Masters G, Engdahl ER, Hung SH (2004) Finite-frequency tomography reveals a variety of plumes in the mantle. *Science* 303(5656):338–343
- Nagaoka Y, Nishida K, Aoki Y, Takeo M, Ohminato T (2012) Seismic imaging of magma chamber beneath an active volcano. *Earth Planet Sci Lett* 333–334:1–8. doi:10.1016/j.epsl.2012.03.034
- Paige CC, Saunders MA (1982) LSQR: an algorithm for sparse linear equations and sparse least squares. *ACM Trans Math Soft* 8:43–71
- Patanè D, Barberi G, Cocina O, De Gori P, Chiarabba C (2006) Time-resolved seismic tomography detects magma intrusions at Mount Etna. *Science* 313(5788):821–823
- Paulatto M, Minshull TA, Baptie B, Dean S, Hammond JOS, Henstock T, Voight B (2010) Upper crustal structure of an active volcano from refraction/reflection tomography, Montserrat, Lesser Antilles. *Geophys J Int* 180(2):685–696
- Power JA, Villasenor A, Benz HM (1998) Seismic image of the Mount Spurr magmatic system. *Bull Volcanol* 60(1):27–37
- Ritzwoller MH, Lin FC, Shen W (2011) Ambient noise tomography with a large seismic array. *Compte Rendus Geosci*. doi:10.1016/j.crte.2011.03.007
- Sens-Schönfelder C, Wegler U (2006) Passive image interferometry and seasonal variations of seismic velocities at Merapi Volcano, Indonesia. *Geophys Res Lett* 33:L21302
- Shapiro NM, Campillo M, Stehly L, Ritzwoller M (2005) High resolution surface wave tomography from ambient seismic noise. *Science* 307:1615–1618
- Sharp ADL, Davis PM, Gray F (1980) A low velocity zone beneath Mount Etna and magma storage. *Nature* 287:587–591
- Stankiewicz J, Ryberg T, Haberland C, Fauzi, Natawidjaja D (2010) Lake Toba volcano magma chamber imaged by ambient seismic noise tomography. *Geophys Res Lett* 37:L17306. doi:10.1029/2010GL044211
- Stauber DA, Green SM, Iyer HM (1988) Three-dimensional P velocity structure of the crust below Newberry Volcano, Oregon. *J Geophys Res Solid Earth* (1978–2012) 93(B9):10095–10107

- Thurber CH (1983) Earthquake locations and three-dimensional crustal structure in the Coyote Lake area, central California. *J Geophys Res Solid Earth* (1978–2012) 88(B10):8226–8236
- Thurber CH (1984) Seismic detection of the summit magma complex of Kilauea volcano, Hawaii. *Science* 223(4632):165–167
- Villasenor A, Benz HM, Filippi L, De Luca G, Scarpa R, Patanè G, Vinciguerra S (1998) Three-dimensional P-wave velocity structure of Mt. Etna, Italy. *Geophys Res Lett* 25(11):1975–1978
- Zandomenighi D, Barclay AH, Almendros J, Ibáñez JM, Wilcock WSD, Ben-Zvi T (2009) Crustal structure of deception island volcano from p-wave seismic tomography: tectonic and volcanic implications. *J Geophys Res* 114:B06310
- Zollo AEA, Gasparini P, Virieux J, Le Meur H, De Natale G, Biella G, Vilardo G (1996) Seismic evidence for a low-velocity zone in the upper crust beneath Mount Vesuvius. *Science* 274(5287):592–594

Seismic Vulnerability Assessment: Lifelines

Kyriazis Pitilakis and Sotiris Argyroudis
Department of Civil Engineering, Aristotle
University, Thessaloniki, Greece

Synonyms

Fragility curves; Infrastructures; Seismic risk assessment; Transportation networks (railway, roadway, harbor, airports); Utilities (gas, oil, water, wastewater, energy supply)

Introduction

Main Features of Lifelines

Lifelines refer to the complex system and network assets of connected components, usually interacting with other components and systems, which are performing vital functions that are essential to sustain the life and the growth of a community, such as producing, transporting, and distributing goods or services. Their global value for the society and economy is permanently increased in our modern, technologically advanced, highly demanding, and fragile world.

They constitute by themselves a set of critical facilities for the proper and safe functioning of the societies. In case of a strong earthquake motion, their physical damages and the consequent disruption of the services they provide may contribute seriously to the global economic loss. At the same time, their repairation cost may be very high, reaching in several cases 10 % or 15 % of the initial construction cost of the whole system to which they belong.

The term *lifeline* is ambiguous. Lifelines may be distinguished in two major categories: (i) Utility systems including potable water, natural gas, oil, electric power distribution, wastewater, or communication systems and (ii) transportation systems comprising roadways, railways, airport, and port facilities. Sometimes the terms infrastructures or critical facilities are used instead, at least for some of them.

Compared to buildings, lifeline systems present three distinctive features: (a) spatial variability and topology and exposure to different geological and geotechnical hazards, (b) wide variety of component typologies and material used, and (c) specific functionality requirements, which make them highly hierarchical networks. For practical reasons they are generally categorized in links (e.g., pipelines, roads) and nodes (e.g., tanks, power substations). Moreover, for their vulnerability assessment, an important parameter is the presence of synergies between components within the same system (intra-dependencies) or between different systems (interdependencies). Due to these synergies, the physical damage of a component of a system interacting with another one may affect seriously the second system's performance and functionality. Therefore, it is essential to define the taxonomy of each interacting lifeline system that describes the individual components of each system and their role in the network as well as the way other systems affect its performance.

Due to their spatial extent, lifeline systems are exposed to variable seismic ground motions, often presenting important incoherency (shaking effects), and to geotechnical hazards expressed in terms of permanent ground deformations, resulted from fault crossing, landslides, and

liquefaction (e.g., lateral spreading, settlements, buoyancy effects). Various parameters may be used to describe and characterize the strong ground motion and the induced phenomena. The selection of appropriate parameter (i.e., intensity measure), which is efficiently correlated with the response and the expected damages of each exposed elements, depends on the technical and response characteristics of each element.

As an example, a motorway network presents the following features: It includes multiple elements at risk like bridges, overpasses, interchanges, tunnels, culverts, embankments, cuts, slopes, retaining walls, signage and markings, electrical systems, rest areas, and buildings (e.g., tolls, maintenance, traffic management). These structures are usually built with different design methods, construction techniques, and materials. The seismic design requirements, and therefore their vulnerability, depend on the type and the location of the structure, the seismicity, and the tectonic features of the area. In particular, roadways are crossing different geological and topographical conditions such as valleys, rivers, canals, gulfs, or mountainous regions. Obviously, the network components can be exposed to various seismic ground motions in terms of amplitude, frequency, or duration, as well as to different geotechnical hazards (subsidence, liquefaction, landslides, faults, etc.). In this context, the design of a long bridge crossing an active seismic fault is much more demanding than a single span bridge in the same area. On the other hand, nowadays, the design level (and subsequently the resistance to earthquakes) is higher than of 30 years ago, making old structures more vulnerable, introducing a new factor, the aging or time effects. Furthermore, the hierarchy (importance) of roads depends on their traffic capacity, i.e., highways and major arterials and secondary and local urban roads. The importance of the road segments, serving different part of the urban fabric, depends also on the land use and the density of the population. The functional role of roads is related to the specific served areas such as industrial or residential areas, hospital facilities, and transport hubs (e.g., airports and major harbors). Therefore, the importance of each

individual infrastructure and part of the network is variable and at the same time an essential factor in the seismic risk management. Finally, the synergies that are presented in a road network and can affect its overall functionality in case of earthquake include the supply of electric power for the signaling and lighting and the proximity to buildings and utility networks (e.g., water or gas pipes) that may induce the interruption of road traffic.

Experience from Past Earthquakes

Several devastating earthquakes occurred during the last century, causing extensive losses in terms of fatalities, injuries, damages, and disruptions. Knowledge, experience, and lessons learned from past earthquakes have significantly contributed to disaster risk reduction efforts across the globe.

The 1971 San Fernando earthquake in USA ($M_L 6.6$), an event that caused catastrophic damage to almost all types of lifelines, motivated the evolution of the lifeline earthquake engineering as well as many new efforts to better understand the causes of these failures and identify ways to mitigate future earthquake damage and disruption (e.g., the establishment of the Technical Council of Lifeline Earthquake Engineering/TCLEE, in ASCE). Other strong and damaging earthquakes that contributed to the progress of lifeline earthquake engineering are the 1978 Miyagiken-oki (Japan, $M_s 7.7$), the 1985 Michoacan (Mexico, $M_w 8.3$), the 1989 Loma Prieta (USA, $M_w 6.9$), the 1990 Luzon (the Philippines, $M_s 7.8$), the 1993 Kushiro-Okai (Japan, $M_w 7.6$), the 1993 Hokkaido-Nansei-Okai (Japan, $M_w 7.7$), the 1994 Northridge (USA, $M_w 6.7$), the 1995 Hyogoken-Nanbu (Kobe) (Japan, $M_w 6.8$), the 1999 Chi-Chi (Taiwan, $M_w 7.6$), the 1999 Kocaeli (Turkey, $M_w 7.4$), and the 2001 El Salvador ($M_w 7.6$) events. More recent earthquakes which caused serious damages and losses to utility systems and transportation infrastructures, even in developed countries, include the 2004 Niigata ken Chuetsu ($M_w 6.8$) and 2007 Niigata-Chuetsu Oki ($M_w 6.6$) earthquakes in Japan, the 2008 Wenchuan (China, $M_w 7.9$), the 2009 L'Aquila (Italy, $M_w 6.3$), the 2010 Maule

(Chile, M_w 8.8), and recently in 2011 the Christchurch (New Zealand, M_w 6.3) and the mega 2011 Tohoku (Japan, M_w 9.0) earthquake and the tsunami associated with this earthquake.

The experience from major earthquakes proved that most lifeline elements are vulnerable, and the mid- and long-term effects, mainly in terms of financial losses, may be very important. A typical example of long-term effects is the port of Kobe in Japan, one of the largest container cargo ports in the world, which suffered major damages, mostly due to liquefaction-induced phenomena, as a result of the 1995 M_w 6.8 Hyogoken–Nanbu earthquake. A year after almost all port infrastructures have been reconstructed, but it is remarkable that in 1998, 3 years after the disaster, cargo traffic remained at roughly half of pre-disaster levels. For the same earthquake, the direct repair cost for the Hanshin Expressway was \$4.6 billion (NCEER 1995), while the total cost was actually much higher if the \$3.4 million of daily income from tolls and the losses from business interruption and traffic delays was taken into account. In the 1994 M_w 6.7 Northridge earthquake, the repair cost for damaged bridges was evaluated at \$190 million, almost the 10 % of the total cost of the whole transportation system (\$1.8 billion, Basöz and Kiremidjian 1998).

Interactions between systems may have a significant impact exceeding direct consequences. Loss of power or communication can severely impact the emergency response, especially when the disruptions concern critical facilities, such as hospitals, command centers, etc. Loss of power may have severe indirect effects due to the synergies between the lifelines and the dependence of all networks on the power supply system. In 1989 M_w 6.9 Loma Prieta earthquake, the damages that occurred in the water network greatly affected the fire-fighting capacity in the area of San Francisco, Oakland, and Berkeley. Representative cases of interactions have been also reported after the 1994 M_w 6.7 Northridge and 1995 M_w 6.8 Hyogoken–Nanbu earthquakes, when the reparation of the water and gas networks have been delayed due to traffic congestion, street blocking, damaged buildings, and

water that flowed into gas pipelines. A large area was also burned down in Kobe due to the disruption and the inefficiency of the fire-fighting network. Considering that modern societies and built environment are relying on the good performance of their interconnected infrastructures, it is clear that an efficient rescue policy and an optimum recovery strategy after a strong earthquake require the evaluation of the interactions between different lifeline systems.

In addition to ground shaking, ground failures (i.e., landslides, liquefaction, later spreading, fault ruptures), if they occur, may produce even more severe damages and losses to lifelines and infrastructures. As an example, the 2008 M_w 7.9 Wenchuan earthquake in China triggered more than 15,000 landslides of various types within an area of 50,000 km² and causing more than 20,000 fatalities representing one-quarter of the total fatalities. The landslides produced extensive damage to housing settlements, irrigation channels, highways, bridges, and other infrastructures. The city of Wenchuan and many other towns were isolated from the rest of the country, and the rescue and relief efforts were greatly affected due to blocking and damages in road network (Tang et al. 2011). During the 2010 M_w 7.1 followed by the 2011 M_w 6.3 Christchurch earthquake sequences in New Zealand, the ground and slope failures, mainly due to liquefaction, produced extensive damages to the water, electricity, and road networks rendering many of them inoperable (O'Rourke et al. 2012). On the contrary the gas system performed rather well. A large number of slope failures, landslides, and rock falls occurred during the 2004 M_w 6.8 Niigata ken Chuetsu earthquake on steep natural slopes and artificial cuts in the area, resulting in major damage to road and railway infrastructures (cuts, embankments, pavements, tunnels, abutments, retaining walls), as well as to other lifeline components (water pipes, electric power poles). Similar damages and disruptions, related to geotechnical failures, were recorded after the subsequent 2007 M_w 6.6 Niigata-Chuetsu Oki earthquake in Japan and in several recent events in Europe (e.g., 2003 M_w 6.4 Lefkas earthquake in Greece, the 2009

Seismic Vulnerability Assessment: Lifelines,

Fig. 1 Damages to road, telecommunication, and gas system in 1994 Northridge ($M_w 6.7$) earthquake (Photo Credit: US Geological Survey)



Seismic Vulnerability Assessment: Lifelines,

Fig. 2 Damaged equipment in Adapazari substation in 1999 Kocaeli ($M_w 7.4$) earthquake (Photo Credit: Kandilli Observatory and Earthquake Research Institute/KOERI)



$M_w 6.3$ L'Aquila earthquake in Italy, the 1999 $M_w 7.4$ Kocaeli earthquake in Turkey) and worldwide. In Figs. 1, 2, 3, 4, 5, 6, 7, and 8, some representative damages for utility and transportation networks are shown.

Several utility systems and networks are now outdated and lack proper seismic design and protection. For example, there are parts of the water and wastewater systems in several cities built more than a century ago. Hence, their material properties and strength are seriously affected from aging effects and are consequently more vulnerable and exposed to earthquake risk. Currently, most of urban infrastructures are not designed to sustain severe seismic hazard and complex systemic threats, and it is difficult to anticipate what may be the complex socio-economic losses due to potential strong earthquakes.

The poor performance of elements at risk in a network is likely to make the whole system much more vulnerable and susceptible to strong earthquakes.

Finally, the short- and long-term economic impact due to earthquake damages of lifelines can be significant. In a recent study by CATDAT (Daniell and Vervaeck 2013), the disaggregation of direct economic losses of 61 selected major earthquakes from 1900 to 2012 showed that around 30 % of the direct losses came from infrastructure losses in transport, communications, pipelines, energy supply systems, etc.

Lifeline Earthquake Engineering Research

In the last two decades, several important projects and research efforts related to the vulnerability and risk assessment of different lifelines at



Seismic Vulnerability Assessment: Lifelines, Fig. 3 Lateral deformation of train tracks at the Arahama station due to surface waves in 2007 Niigata-Chuetsu Oki ($M_w6.6$) earthquake (Photo Credit: US Geological Survey)

regional or urban scale have been undertaken. They resulted in the development of relevant methods and guidelines. A short reference to the most important contributions is following.

United States

In 1992, the Multidisciplinary Center for Earthquake Engineering Research (MCEER) initiated the Highway Project, funded by the Federal Highway Administration (FHWA). The project uniquely examined the impact of earthquakes on the highway system as an integrated network, rather than a collection of individual road segments, bridges, embankments, tunnels, etc. The purpose of the project was to ensure the usability of highways following earthquakes, by improving performance of all interconnected components. Overall goals were to deepen the understanding of the seismic hazard impact to highways and to improve and develop analysis



Seismic Vulnerability Assessment: Lifelines, Fig. 4 Damages to road embankments in 2007 Niigata-Chuetsu Oki ($M_w6.6$) earthquake (Photo Credit: US Geological Survey)

methods, screening procedures and additional tools, retrofit technologies, design criteria, and other approaches to reduce seismic vulnerability of existing and future highway infrastructure. REDARS (Risks due to Earthquake DAMAGE to Roadway Systems) is a public-domain software package developed under this MCEER project that accounts for how earthquake damage affects post-event traffic flows and travel times and estimates losses from these travel-time and traffic-flow impacts (<http://mceer.buffalo.edu/research/redars/>).

In 1996, the PEER Lifelines Program (<http://peer.berkeley.edu/lifelines>) was initiated, under the coordination of the Pacific Earthquake Engineering Research Center, funded by the California Department of Transportation and the Pacific Gas and Electric Company. The aim was to improve the seismic safety and reliability of lifeline systems with particular focus on the characterization of ground motions, local soil and site response, and the performance of bridge structures and electric substation equipment.

In 1997, the Federal Emergency Management Agency (FEMA) released the first edition of HAZUS methodology and software that

**Seismic Vulnerability
Assessment: Lifelines,**

Fig. 5 Collapse of a section of the Nishinomiya in 1995 (M_w 6.8) Hyogoken–Nanbu earthquake (Photo by K. Pitilakis)

**Seismic Vulnerability
Assessment: Lifelines,**

Fig. 6 Damage to Trans-European Motorway due to fault rupture in 1999 Kocaeli (M_w 7.4) earthquake (Photo Credit: Kandilli Observatory and Earthquake Research Institute/Photo by D. Kalafat)

**Seismic Vulnerability
Assessment: Lifelines,**

Fig. 7 Blocking of Pan-American Highway due to landslide in 2001 El Salvador (M_w 7.6) earthquake (Photo Credit: US Geological Survey/Photo by E.L. Harp)



estimates potential building and infrastructure losses from earthquakes, built using GIS technology. The current version is HAZUS-MH V2.0, including also losses from floods and hurricanes.

It was released in 2004 (www.fema.gov/hazus). HAZUS-MH loss estimates reflect the state-of-the-art scientific and engineering knowledge and can be used to inform decision makers at all



Seismic Vulnerability Assessment: Lifelines, Fig. 8 Failure of port infrastructures in 1995 ($M_w 6.8$) Hyogoken–Nanbu earthquake (Photo by K. Pitilakis)

levels of government by providing a reasonable basis for developing mitigation, emergency preparedness, response, and recovery plans and policies. The main merit of the HAZUS platform is that it provided for the first time an unparalleled set of fragility models for all the main components and systems of the built environment. However, it must be recognized that many of these models have been derived solely on expert judgment and the level of knowledge at that time. One effect of the sheer size of the HAZUS framework and its set of tools is that it established itself very soon as the reference for all studies in the sector worldwide.

In 1998, FEMA and the American Society of Civil Engineers entered into a cooperative agreement to establish the American Lifelines Alliance (ALA). ALA's objectives were to facilitate the creation, adoption, and implementation of design and retrofit guidelines and other national consensus documents to improve the performance of utility and transportation systems (electric power, telecommunication, water, wastewater, oil, natural gas, rail, and shipping ports) in natural hazard events, including earthquakes. ALA has not attempted to focus efforts on all lifeline systems and components but has instead chosen to place priority on specific topics relevant to selected lifeline systems where a need for improved hazard mitigation practices is identified. The primary function was to bridge the gap between hazard mitigation practices for buildings and lifeline systems; therefore a close collaboration with lifeline

stakeholders, owners/operators, and recognizing regulatory bodies was established to facilitate use of developed guidelines (<http://www.americanlifelinesalliance.com/>).

ERGO (the new name for the Multi-Hazard Assessment, Response, and Planning software previously known as mHARP and MAEvis) started somewhat later than HAZUS and was the product of the research efforts carried out at the Mid-America Earthquake Center in collaboration with the National Center for Supercomputing Applications (NCSA). It is an open-source software and incorporates many of the design concepts and capabilities motivated by NCSA efforts to develop “cyberenvironments” that span scientific disciplines and that can rapidly evolve to incorporate new research results (Elnashai et al. 2008). An important aspect of ERGO is its extensibility, both in terms of analysis/features modules and of visualization/representation (GIS) modules. The framework has been designed to implement the consequence-based risk management (CRM) paradigm supported by the NCSA and MAE center.

Europe

Until now, no coordinated action for the development of a comprehensive methodology and software tool at European level is made. However, important efforts have been undertaken through research projects mainly funded by European Union. Brief description of the most important is following.

RISK-UE (2001–2004) project involved the assessment of earthquake scenarios based on the analysis of the global impact of one or more plausible earthquakes at city scale, within a European context. The primary aim of these scenarios was to increase awareness within the decision-making centers of a city. The developed methodology for creating earthquake scenarios focused on the distinctive features of European cities with regard to current and historical buildings and lifelines, as well as on their functional and social organization, in order to identify weak points within the urban system. The approach was applied to 7 European cities: Barcelona, Bitola, Bucharest, Catania, Nice, Sofia, and Thessaloniki (Mouroux and Le Brun 2006).

LESS-LOSS (2003–2007) was a European Integrated Project focusing on risk mitigation for earthquakes and landslides that relied on the active participation of 46 partners from both academia and industry. The project addressed research issues on seismic engineering, earthquake risk and impact assessment, landslides monitoring, mapping and management strategies, improved disaster preparedness and mitigation of geotechnical hazards, development of advanced methods for risk assessment, methods of appraising environmental quality, and relevant pre-normative research. Separate sections were devoted on earthquake disaster scenario predictions and loss modeling for urban areas and infrastructures with emphasis on water and gas systems (Faccioli 2007).

SYNER-G (2009–2013, www.syner-g.eu) is the most recent European funded research project that developed an integrated, general methodology and a comprehensive framework for the systemic vulnerability and risk analysis of buildings, lifelines, and infrastructures. The proposed methodology and tools encompass in an integrated fashion all aspects in the chain, from hazard to the vulnerability assessment of components and systems and to the socioeconomic impacts of an earthquake, accounting for most relevant uncertainties within an efficient quantitative simulation scheme and modeling interactions between the multiple component systems. It systematically integrates the most advanced fragility or

vulnerability functions to assess the vulnerability of physical assets. The increasing impact due to interdependencies and intra-dependencies between different components and among different interacting systems is treated in a comprehensive way, providing specifications for each network and infrastructure. The proposed socioeconomic model integrates social vulnerability into the physical systems modeling approaches providing to decision makers with a dynamic platform to capture post disaster emergency issues like shelter demand and health impact decisions (Pitilakis et al. 2014a, b).

Other Efforts Worldwide

The Central American Probabilistic Risk Assessment (CAPRA) platform was developed in partnership with Central American governments, the support of the Central American Coordination Centre for Disaster Prevention (CEPREDENAC), the Inter-American Development Bank (IDB), the International Strategy of United Nations for Disaster Reduction (UN-ISDR), and the World Bank. CAPRA is an information platform to enhance decision-making in risk management across various sectors, such as emergency management, territorial planning, public investment, and the financial sectors. Through the application of probabilistic risk assessment principles to threats like hurricanes, earthquakes, volcanic activity, floods, tsunamis, and landslides, CAPRA allows to measure and compare different types of risks and to develop sector-specific applications for risk management (Cardona et al. 2012; <http://www.ecapra.org/>). It is not directly designed for lifeline vulnerability and risk assessment, but it can be used for this purpose as well.

Key Elements of Seismic Risk Assessment

Risk assessment is the process used to determine risk management priorities by evaluating and comparing the level of risk to specific standards, either defined by a code or a set of target risk levels and criteria. The estimation of losses, both material and immaterial, is the ultimate target of the risk assessment. Then, the risk management procedure should define pre-seismic (retrofitting

and strengthening actions), coseismic, and post-seismic strategies and policies. The risk assessment of lifelines and infrastructures follows the general scheme of seismic risk assessment:

$$[\text{seismic risk}] = [\text{seismic hazard}] \times [\text{vulnerability}] \\ \times [\text{exposure} - \text{elements at risk}]$$

The complexity of elements at risk, their variability from one place and one country to another, and, till recently, the lack of well-validated damage and loss data from strong earthquakes make the vulnerability assessment of each particular component and of the network as a whole a challenging task. Adding to that the spatial extent of lifelines, the synergies between different systems, and the several sources of uncertainties that are inherent in the various aspects (e.g., typology, damage states), models (e.g., seismic hazard, spatial correlation, fragility curves), and finally tools and methods for estimation of losses, it is obvious that the risk assessment of lifelines is indeed a very complex and challenging issue. Therefore, multidisciplinary task and combined efforts are really needed to reach reliable and comprehensive estimates.

This difficulty is amplified due to lack of well-documented data and the partial or poor knowledge of the geometric and other features of the systems; in some cases it is also due to the confidentiality of data (i.e., communication or oil networks). The situation however has been improved the last decade. Important strong earthquakes provided valuable good quality data, while the public awareness and the reported huge direct and indirect losses associated to lifeline damages drew the attention of the scientific community, the governmental authorities, and the insurance sector. Moreover, the development of geographical information systems (GIS) as well as remote sensing technologies offered an excellent platform for the implementation of efficient and innovative techniques to screen, capture, store, manipulate, analyze, manage, and present geographical data.

Few terms employed in this chapter are described in the following. Inventory and

taxonomy/typology definitions of all elements at risk and systems are the necessary initial steps to describe the elements and networks exposed to seismic risk; at the same time, the development of detailed inventories is the most expensive task in terms of financial cost and time. The estimation of seismic hazard and site-specific seismic ground motion characteristics provides appropriate seismic intensity measures (IM), which will be used later in the vulnerability and loss assessment modeling. Vulnerability is the propensity of damage of an element (building, bridge, pipeline, roadway, oil tank, etc.) or a network of specific elements (e.g., water or gas system) to a given seismic intensity defined in the seismic hazard analysis. It is commonly assessed based on fragility models, which estimate the probability of damage for a given seismic intensity. Fragility functions are constructed with respect to the taxonomy and typological characteristics of each element at risk. The expected damages and losses of the entire network are estimated through a systemic analysis, which takes into account the physical damages of the various parts and segments of the system, and in some cases, the interdependencies between different components within the system and the interdependencies between systems, in order to estimate the serviceability or flow capacity of the damaged network. The results are commonly provided in terms of performance indicators (PI), which are describing the network functional or the expected economic or socioeconomic losses. Efficient risk management is based on the evaluation of the resulting performance indicator values.

Seismic Hazard for Distributed Networks

Lifelines are in most cases spatially distributed systems. Therefore, seismic hazard should meet the specific features of each system considering also its spatial variability. The vulnerability analysis and risk assessment must be evaluated according to the precise typological characteristics of the components and networks, taking also into account the models used to describe vulnerability, usually in terms of fragility curves and relationships. Moreover, due to the spatial extent of utility and transportation networks, the seismic

hazard should describe the spatial variability of ground motion considering local soil conditions, topographic effects, and geotechnical hazards; site-specific seismic hazard analyses are always necessary. Traditional seismic hazard analysis, while effective in translating the hazard into a probabilistic formulation, is limited in the extent to which it can incorporate spatial coherency of the form needed for estimation of loss to spatially distributed portfolios. The extension of seismic risk analysis to multiple systems of spatially distributed infrastructures presents new challenges in the characterization of the seismic hazard input, particularly with respect to the spatial correlation structure of the ground motion residuals (i.e., residuals from empirical intensity models) that form the basis for the systemic risk analysis.

A general procedure entitled “Shakefield” has been recently established in the frame of SYNER-G project (Pitilakis et al. 2014b), which allows for the generation of samples of ground motion fields both for single-scenario-type events and for stochastically generated sets of events needed for probabilistic seismic risk analysis (Weatherill et al. 2014). For a spatially distributed infrastructure of vulnerable elements, the spatial correlation of the ground motion fields for different measures of the ground motion intensity is incorporated into the simulation procedure. This is extended further to consider spatial cross-correlation between different measures of ground motion intensity.

The consideration of hazard from permanent ground deformation (PGD) is essential in modeling the seismic risk to lifeline systems. For pipelines and similar systems with linear elements, fragility models are generally given in terms of PGD, as they are mostly vulnerable to permanent displacement of the ground rather than transient shaking. Four primary causes of permanent ground displacements are commonly considered: liquefaction-induced lateral spread, liquefaction-induced settlement, slope displacement, and coseismic fault rupture. In addition to strong shaking, another transient effect that poses a potential risk to lifeline systems is the transient ground strain (PGS). Several models are available for the estimation of PGD; some of them are

intended to relate the degree of deformation and the probability of the geotechnical hazard occurring to the intensity of the ground motion (Weatherill et al. 2014). However, it should be noted that most theoretical and empirical models relating PGD to strong shaking require a level of geotechnical detail that may be impractical to obtain for a spatially distributed set of sites. HAZUS methodology (FEMA 2003) provides a rather simple “baseline” model that can be implemented in the widest variety of applications.

When the scenario type or probabilistic site-specific seismic hazard is available, the damage assessment of a lifeline system is evaluated using appropriate fragility or vulnerability functions relating the probability of damages to seismic intensity measures. The data and parameters needed to perform this analysis are shortly described in the following.

Taxonomy, Inventory, and Typology

The key assumption of the vulnerability assessment of lifelines is that the structures and components of the systems, having similar structural characteristics (e.g., a bridge of a given typology), being in similar geotechnical conditions (e.g., a segment of a pipeline crossing the same soil conditions and exposed to the same geotechnical hazards), are expected to perform in the same way for a given seismic excitation. Within this context, expected damages are directly related to the structural properties of the elements at risk. *Taxonomy* and *typology* are thus fundamental descriptors of a system that are derived from the inventory of each element and system. Geometry, material properties, morphological features, age, seismic design level, anchorage of the equipment, soil conditions, and foundation details are among usual typology descriptors/parameters. Buildings, bridges, pipelines (gas, fuel, water, wastewater), tunnels, road embankments, power substations, harbor facilities, and road and railway networks have their own specific set of typologies and different taxonomy.

The taxonomy of any lifeline network is thus an essential step for identifying, characterizing, and classifying all types of lifeline elements according to their specific typology and their

distinctive geometric, structural, and functional features. It allows the classification of the network elements in an ordered classification system adequate for the seismic risk assessment of each component and the network (or system) as a whole. The taxonomy is specific for each lifeline system and it needs a detailed typology definition of all elements at risk. For that, an adequate inventory process is essential. However, several difficulties arise in the collection and archiving of the data related to the absence of well-organized archives in public and often in private organizations managing the systems, the oldness of the networks (e.g., water network), and the high cost to draw up complete inventories. To this respect remote sensing techniques and GIS offer a useful and indispensable instrument and platform to enhance with relatively low cost the collection of the data and implement any inventory inquiries. Different types of satellite data can be used to extract information and parameters needed to compile or update inventories for seismic risk assessment. Examples include the width of roads, the geometry or material of buildings, building aggregates, infrastructures like bridges and channels, land use characteristics, and others.

The inventory of a specific structure in a region and the capability to create classes of structural types (e.g., with respect to material, geometry, and design code level) are among the main challenges when carrying out a general seismic risk assessment, for example, at a city or region scale, where it is practically impossible to perform this assessment separately for each single structure. It is necessary to classify all elements at risk, in “as much as possible” homogeneous classes and subclasses presenting more-or-less similar response characteristics to ground shaking. Thus, the derivation of appropriate fragility curves for any type of structure depends entirely on the creation of a reasonable taxonomy that is able to classify the different kinds of structures and infrastructures in any system exposed to seismic hazard. Several uncertainties are inherent to each taxonomy system, which are inevitably propagated through the generic fragility (vulnerability) curves to the final loss estimation. The most coherent and

comprehensive taxonomy presently available in Europe is produced in SYNER-G project (Table 1, Pitilakis et al. 2014a). Detailed classifications are also provided in HAZUS (FEMA 2003) and ALA (2001).

Each component is further classified according to its specific features regarding its seismic behavior. For instance, the main typological features of gas pipelines include the material type, material strength, diameter, wall thickness and smoothness of coating, type of connection and joints, as well as the nominal design and actual flow. Among them, material and the connection types (joints) together with the soil type are the most critical parameters for their seismic response. Table 2 outlines common types of pipelines based on the above classification criteria.

Another example is storage tanks, which are usually categorized according to their material type (steel or reinforced concrete), construction type (at grade or elevated), anchorage (anchored or unanchored), roof type and capacity, shape factor (height-on-diameter ratio), and amount of content in the tank (full, half full, empty). Among them, material type, construction type, and anchorage are considered as the most important. Table 3 presents the classification of tanks in USA according to ALA (2001).

For a systematic and detailed description of all classes and subclasses of all elements at risk in most lifeline systems and networks in the European context, the reader should refer to Pitilakis et al. (2014a).

Damage Assessment

The damage patterns and mechanisms are different for each lifeline system and element (component), and they are strongly depending on the typology of each structure (i.e., materials, geometry, structural types). For example, the damage mechanisms of pipelines are generally classified in the form of breaks or leaks; material type (i.e., brittle or ductile) and joints (i.e., flexible or rigid) are also factors affecting the seismic response and damages of pipelines. In case of tunnels, damage patterns include lining cracks (longitudinal or transverse) and spalling, wall

Seismic Vulnerability Assessment: Lifelines, Table 1 SYNER-G infrastructure taxonomy (Pitilakis et al. 2014a)

System	Component (and subcomponents)
BDG: buildings	Force-resisting mechanism (FRM1): moment-resisting frame structural wall, flat slab, bearing walls, precast, confined masonry
	FRM material (FRMM1): concrete, masonry
	Plan (P): regular, irregular
	Elevation (E): regular/irregular geometry
	Cladding (C): regular infill vertically, irregular infill vertically, bare
	Detailing (D): ductile, non-ductile, with tie rods/beams, without tie rods/beams
	Floor system (FS): rigid, flexible
	Roof system (RS): peaked, flat, gable end walls
	Height level (HL): Low rise, mid-rise, high rise, tall
	Code level (CL): none, low, moderate, high
EPN: electric power network	EPN01: electric power grid
	EPN02: generation plant
	EPN03: substation
	EPN04: distribution circuits
	EPN05–09: substation macro-components (autotransformer line; line without transformer; bar-connecting line; bars; cluster)
	EPN10–23: substation micro-components (circuit breaker; lightning arrester or discharger; horizontal disconnect switch or horizontal sectionalizing switch; vertical disconnect switch or vertical sectionalizing switch; transformer or autotransformer; current transformer; voltage transformer; box or control house; power supply to protection system; coil support; bar support or pothead; regulator; bus; capacitor tank)
	EPN24: transmission or distribution line
GAS: natural gas system	GAS01: production and gathering facility (onshore, offshore)
	GAS02: treatment plant
	GAS03: storage tank
	GAS04: station (compression; metering/pressure reduction; regulator; metering)
	GAS05: pipeline
	GAS06: SCADA
OIL: oil system	OIL01: production and gathering facility (onshore, offshore)
	OIL02: refinery
	OIL03: storage tank farm
	OIL04: pumping plant
	OIL05: pipeline
	OIL06: SCADA
WSS: water-supply network	WSS01: source (springs, rivers, natural lakes, impounding reservoirs, shallow or deep wells)
	WSS02: treatment plant
	WSS03: pumping station
	WSS04: storage tank
	WSS05: pipe
	WSS06: tunnel
	WSS07: canal
	WSS08: SCADA system
WWN: wastewater network	WWN01: wastewater treatment plant
	WWN02: pumping (lift) station
	WWN03: pipe
	WWN04: tunnel
	WWN05: SCADA system

(continued)

Seismic Vulnerability Assessment: Lifelines, Table 1 (continued)

System	Component (and subcomponents)
RDN: road network	RDN01: bridge (material, type of deck, deck structural system, pier to deck connection, type of pier to deck connection; type of section of the pier, spans, type of connection to the abutments, skew, bridge configuration, foundation type, seismic design level)
	RDN02: tunnel
	RDN03: embankment (road on)
	RDN04: trench (road in)
	RDN05: unstable slope (road on or running along)
	RDN06: road pavement (ground failure)
	RDN07: bridge abutment
RWN: railway network	RWN01: bridge
	RWN02: tunnel
	RWN03: embankment (track on)
	RWN04: trench (track in a)
	RWN05: unstable slope (track on or running along)
	RWN06: track
	RDN07: bridge abutment
HBR: harbor	RWN08: station
	HBR01: waterfront components (gravity retaining structures; sheet pile wharves; piers; breakwaters mooring and breasting dolphins)
	HBR02: earthen embankments (hydraulic fills and native soil material)
	HBR03: cargo handling and storage components (cranes, tanks, etc.)
	HBR04: buildings (sheds, warehouse, offices, etc.)
FFS: fire-fighting system	HBR05: liquid fuel system (as per the OIL system)
	FFS01: fire-fighter station
	FFS02: pumping station
	FFS03: storage tank
	FFS04: fire hydrant
	FFS05: pipe

deformation, bending and buckling of reinforcing bars, obstruction of the opening, or pavement cracks. In subway stations damages to columns may also occur. Damages to quay walls are related to excessive lateral pressures or decrease of shear strength of the foundation soil that can cause sliding, deformation, and tilting of the walls. Settlements and lateral movement of the backfill materials and cracking of apron pavements can be also induced. Damage to electric power substations includes failures of the various subcomponents (e.g., disconnect switches, circuit breakers, transformers) or damage of the building. Earth structures such as highway and railway embankments can spread laterally and settle, resulting in opening of cracks in the road pavement or displacement of the railway tracks. The list of possible damage patterns is unlimited.

Therefore, classification of damage and the subsequent definition of specific damage states are important in the vulnerability assessment as the seismic intensity is correlated to the expected damage level through the fragility or vulnerability functions. Again, the form of the fragility functions depends on the typology of the element at risk. For common structures (e.g., buildings, bridges) and other not extended elements (e.g., cranes, tanks, substations), the fragility curves describe the response and damage level of particular subcomponents (e.g., columns, transformers) or of the entire structure. For linear elements of extended networks such as gas pipelines, the fragility functions describe the number of expected damages along a certain length (i.e., per km). Examples and further details are given in the next sections.

Seismic Vulnerability Assessment: Lifelines, Table 2 Common types of materials and connections for buried pipelines

Material type	Connection type
Asbestos cement (AC)	Arc welded
Cast iron (CI)	Bell and spigot
Ductile iron (DI)	Cemented
Concrete (C)	Riveted
Polyvinyl chloride (PVC)	Rubber gasket
Welded steel (WS)	Gas welded
Medium-density polyethylene (MDPE)	
High-density polyethylene (HDPE)	

Damage States

In seismic risk assessment, the performance levels of a structure, for example, a reinforced concrete building, belonging to a specific class (Pitilakis et al. 2014a) can be defined through damage thresholds called limit states. A limit state defines a boundary between two different damage conditions often referred to as *damage states*. The thresholds are related to functionality and serviceability levels while they are usually defined based on engineering judgment and common sense. They are also strongly depending on the model applied for the analysis and derivation of the fragility functions. Uncertainties related to this stage of the analysis are normally referred as epistemic uncertainties (see Pinto 2014). Different damage criteria have been proposed depending on the typologies of elements at risk and the approach used for the derivation of fragility curves. The most common way to define earthquake consequences is a classification in terms of the following damage states: *no damage*, *slight/minor*, *moderate*, *extensive*, and *complete*. This qualitative approach requires an agreement on the meaning and the content of each damage state. The number of damage states is variable and is related with the functionality of the components and/or the repair duration and cost. In this way the total losses of the system (economic and functional) can be estimated. In particular, physical damages are related to the expected serviceability level of the component (i.e., fully or partial operational or inoperative) and the

Seismic Vulnerability Assessment: Lifelines, Table 3 Typology of tanks (ALA 2001)

Unanchored redwood tank (5×10^4 – 5×10^5 gal)
Unanchored post-tensioned circular concrete tank ($>1 \times 10^6$ gal)
Unanchored steel tank with integral shell roof (1×10^5 – 2×10^6 gal)
Unanchored steel tank with wood roof (1×10^5 – 2×10^6 gal)
Anchored steel tank with integral steel roof (1×10^5 – 2×10^6 gal)
Unanchored steel tank with integral steel roof ($>2 \times 10^6$ gal)
Anchored steel tank with wood roof ($>2 \times 10^6$ gal)
Anchored reinforced (or prestressed) concrete tank (5×10^4 – 1×10^6 gal)
Elevated steel tank with no seismic design
Elevated steel tank with nominal seismic design
Open-cut reservoir
Fiberglass tanks

corresponding functionality (e.g., power availability for electric power substations, number of available traffic lanes for roads, flow or pressure level for water system). These correlations provide quantitative measures of the component’s performance and can be applied for the definition of specific *performance indicators (PIs)*, which are introduced in the systemic analysis of each network. Therefore, the comparison of a demand with a capacity quantity, the consequence of a mitigation action, or the accumulated consequences of all damages (usually referred as the “impact”) can be evaluated.

Methods for deriving *fragility curves* generally describe damages on a discrete damage scale. In the empirical procedures, the scale is used in survey efforts to produce post-earthquake damage statistics, and sometimes it is rather subjective (i.e., there is often a discrepancy between the damage levels that any two different inspectors would assign for the same incident). In analytical procedures the scale is related to limit state of selected mechanical properties that are described by appropriate indices, such as the displacement capacity or the storey drift in the case of buildings or simple drift in pier bridges. For other elements at risk, the definition of the performance levels or the limit states may be more vague and follows

other criteria related, for example, in the case of pipelines, to the limit strength characteristics of the material used in each typology. The definition and consequently the selection of the damage thresholds, i.e., limit states, are among the main, yet, unavoidable sources of uncertainties.

Intensity Measures

An important issue related to the fragility curve construction and implicitly to the risk assessment is the selection of an appropriate earthquake *intensity measure (IM)* that characterizes the strong ground motion that best correlates with the response of each element, for example, building, pipeline, or harbor facilities like cranes. Several measures of the intensity of ground motion (IMs) have been developed. Each intensity measure may describe different characteristics of the motion, some of which may be more adverse for the structure or the system under consideration. The use of a particular IM in seismic risk analysis should be guided by the extent to which the measure corresponds to damage to the components of a system. Optimum intensity measures are defined in terms of practicality, effectiveness, efficiency, sufficiency, robustness, and computability (Mackie and Stojadinovic 2003).

In general, IMs are grouped in two general classes: empirical intensity measures and instrumental intensity measures. With regard to the empirical IMs, different macroseismic intensity scales could be used to identify the observed effects of ground shaking over a limited area. Instrumental IMs are, by far, more accurate and representative of the seismic intensity characteristics and the severity of ground shaking. For example, for bridges the best descriptor is a spectral response value at a specific period (i.e., $T = 1.0$ s). For other lifeline components, it may be the peak ground acceleration (e.g., buildings, tanks, electric power substations), peak ground velocity (e.g., pipelines), or even the permanent ground deformations (e.g., pipes, embankments, roadways, railways). The correlation between damages of specific elements at risk and intensity measures is not simple and never unique. Several other descriptors like peak ground strain, Arias intensity, cumulative absolute velocity, and other

parameters of the ground motion have been also used for different structures composing a lifeline system. More recently, it is proposed to use two descriptors instead of one, and hence the fragility curve is transformed in fragility surfaces (Seyedi et al. 2010; Douglas et al. 2014).

The selection of the adequate intensity parameter is also related to the approach that is followed for the derivation of fragility curves and the typology of elements at risk. The identification of the proper IM is determined from different constraints, which are first of all related to the adopted hazard model, but also to the element at risk under consideration and the availability of data and fragility functions for all different exposed assets. Empirical fragility functions are usually expressed in terms of the macroseismic intensity defined according to different macroseismic scales (EMS, MCS, and MM). Analytical or hybrid fragility functions are, on the contrary, related to instrumental IMs, which are related to parameters of the ground motion (PGA, PGV, PGD) or of the structural response (spectral acceleration S_a or spectral displacement S_d , for a given value of the period of vibration T). When the vulnerability of elements due to ground failure is examined (i.e., liquefaction, fault rupture, landslides), permanent ground deformation (PGD) is the most appropriate IM.

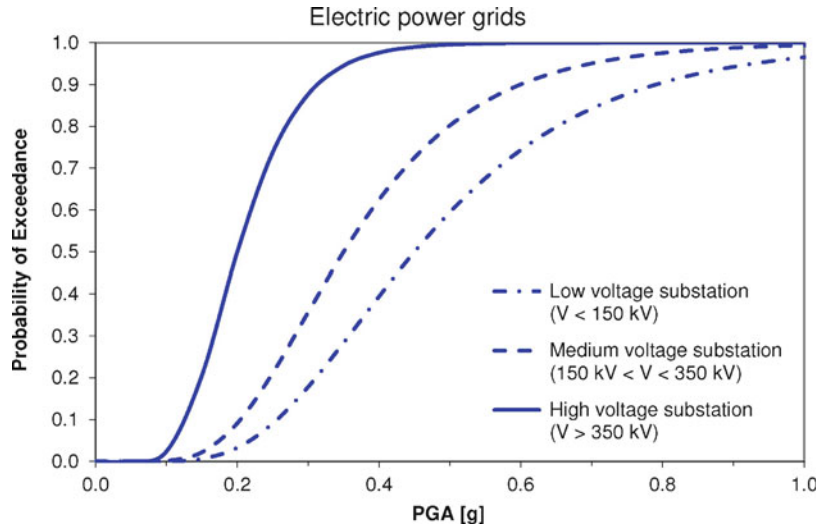
Vulnerability Assessment and Fragility Curves

The fundamental tool in seismic risk assessment of lifeline components is the *fragility curves* which describe the probability that a structure will reach or exceed a certain damage state for a given ground motion intensity. An extensive review of available fragility functions and state-of-the-art methods for vulnerability assessment of buildings and lifeline components can be found in Pitilakis et al. (2014a). Fragility curves are usually represented by two-parameter (median and log-standard deviation) cumulative lognormal distributions. Several approaches are used to establish fragility functions. They are grouped in the following five categories:

- Empirical fragility curves, based on post-earthquake surveys and observations of actual

Seismic Vulnerability Assessment: Lifelines,

Fig. 9 Empirical fragility curves for power grids made up of substations of different voltages based on data from US west coast earthquakes (Dueñas-Orsorio et al. 2007)



damage. They are specific to particular sites and seismotectonic, geological, and geotechnical conditions, as well as the properties of the damaged structures. Consequently, the use of these functions in different regions is always questionable (Figs. 9 and 10).

In case of pipelines, the empirical fragility functions relate the repair rates (RR) expressed as repairs/km with the peak ground velocity (PGV) or permanent ground deformation (PGD) (Fig. 11). The curves may be further adapted to the material properties and geometry of the pipelines and soil conditions. In the last two decades, the increased density of high-quality strong ground motion records in different soil conditions, in combination with new technologies such as geographical information systems (GIS) and remote sensing technologies (e.g., LiDAR) capable to measure more accurately ground movements, contributed significantly to the development and verification of such relationships (O'Rourke et al. 2012).

- Analytical fragility curves, based on numerical simulations of structural models under increasing earthquake loads (Figs. 12 and 13). Analytical methods, validated with large-scale experimental data and observations from recent strong earthquakes, have become more popular in recent years.

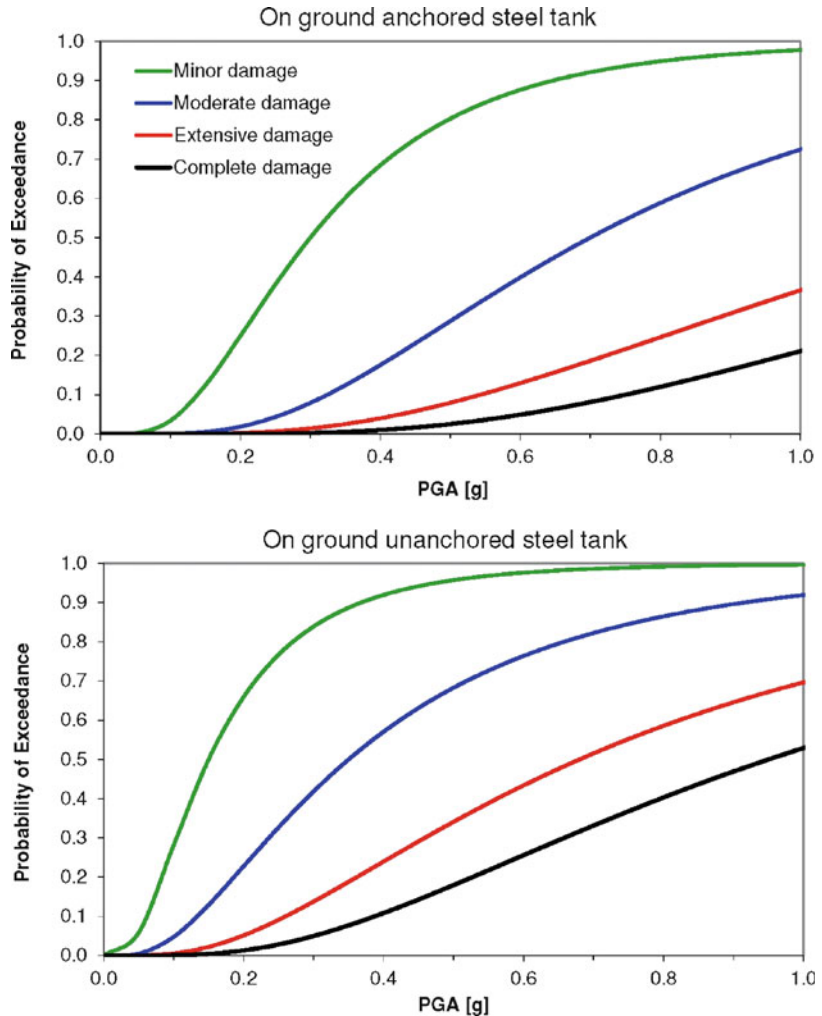
The main reason is the considerable improvement of computational tools, methods, and skills, which allows comprehensive parametric studies covering most common typologies to be undertaken. Moreover, several of the associated uncertainties, e.g., material properties, are better controlled.

- Judgmental or expert elicitation fragility curves, using questionnaires by which the experts are queried on the probability of a component being in a certain damage state for a given intensity (Fig. 14). They are versatile and relatively fast to establish, but their reliability is questionable because of their dependence on the experiences of the experts consulted.
- Hybrid fragility curves, which combine any of the above-mentioned techniques in order to compensate for their respective drawbacks.
- Fragility curves based on a fault-tree analysis, where complex components (e.g., substations, pumping plants, hospitals) are broken down into subcomponents and the global fragility is obtained based on the relationships between the subcomponents and their individual fragilities (Fig. 15).

An example of damage estimation for “on-ground unanchored steel tanks” due to ground shaking and “pipelines” due to ground failure is given in Table 4. In the first case, the exceedance probabilities of each damage state are estimated

Seismic Vulnerability Assessment: Lifelines,

Fig. 10 Empirical fragility curves for on-ground steel tanks subjected to ground shaking (ALA 2001)



based on the fragility curves in Fig. 10, and then the punctual probabilities of each damage state are obtained. In the second case, the repairs per kilometer are estimated for a given value of PGD based on the curves in Fig. 11. The estimated repairs should be modified according to the length of the pipe segment under study.

Systemic Analysis

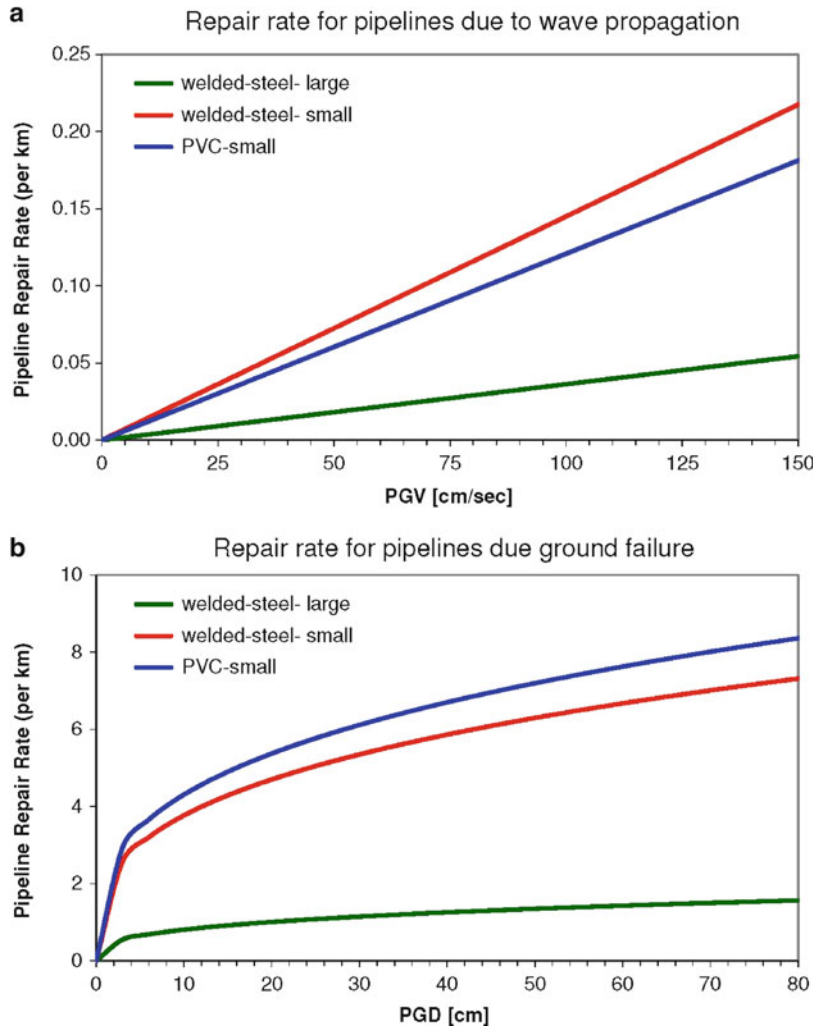
The majority of the methodologies that have been developed worldwide for the seismic risk assessment of lifelines and infrastructures refer to single systems, without considering interactions, cascading failures, and complex impacts. The study of interdependent infrastructures is challenging due to heterogeneous quality and

insufficient data availability and the need to account for their spatial and temporal aspects of complex supply–demand operation (Satumtira and Duenas-Osorio 2010). The various approaches available in the literature depend on the simulation method, modeling objectives, scale of analysis, availability of input data, and end-user type or needs. A comprehensive categorization scheme describing different levels is provided in Modaressi et al. (2014):

- Vulnerability analysis: This level considers only the potential physical damages of the components of the systems, with no consideration of functionality of either the elements or the whole system.

Seismic Vulnerability Assessment: Lifelines,

Fig. 11 Empirical fragility functions for common pipeline typologies provided by ALA (2001): (a) pipelines subjected to wave propagation (PGV), (b) pipelines subjected to permanent ground deformation (PGD)



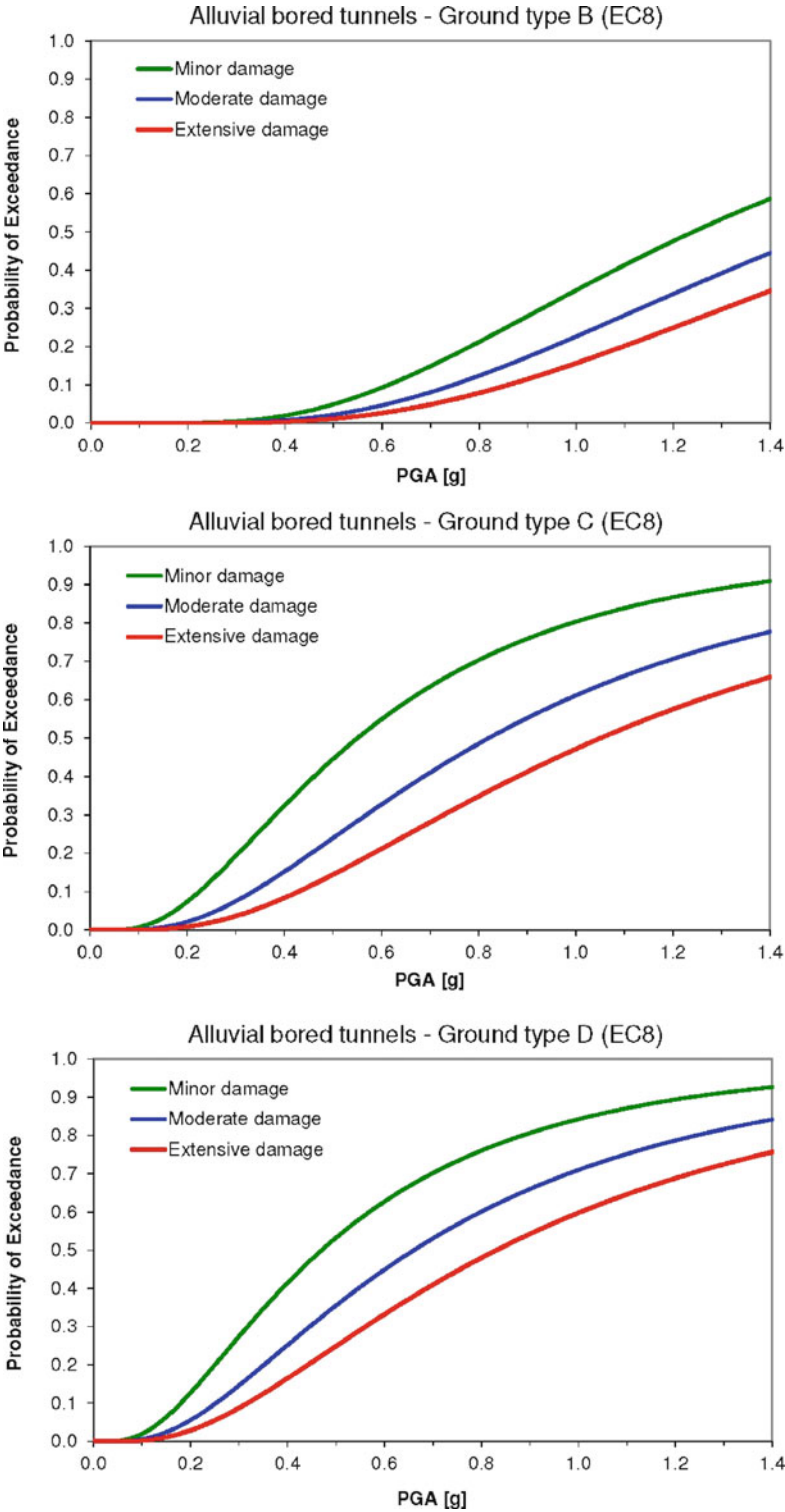
- Connectivity analysis: This level describes the probability of the demand nodes to be connected to functioning supply nodes through undamaged paths. In this approach the damaged components are removed from the network and the adjacency matrix is updated accordingly, thus pointing out the nodes or areas that are disconnected from the rest of the system. This qualitative approach is used for all utility networks (water, electricity, gas) and the road transportation system.
- Capacitive analysis: This level describes the ability of the system to provide to the users the required functionality through a quantitative approach. For utility networks, graph algorithms and flow equations can be used to

estimate capacitive flows from sources (e.g., generators, reservoirs) to sinks (i.e., distribution nodes), based on the damages sustained by the network components (from total destruction to slight damages reducing the capacity).

- Fault-tree analysis: This level of analysis concerns critical infrastructures, where multiple conditions are necessary for the systems to ensure its task. This type of approach aims to evaluate the remaining operating capacity (residual operation capacity) of objects such as health-care facilities. The system is broken down into structural, nonstructural, or human components, each one of them being connected with logic operators.

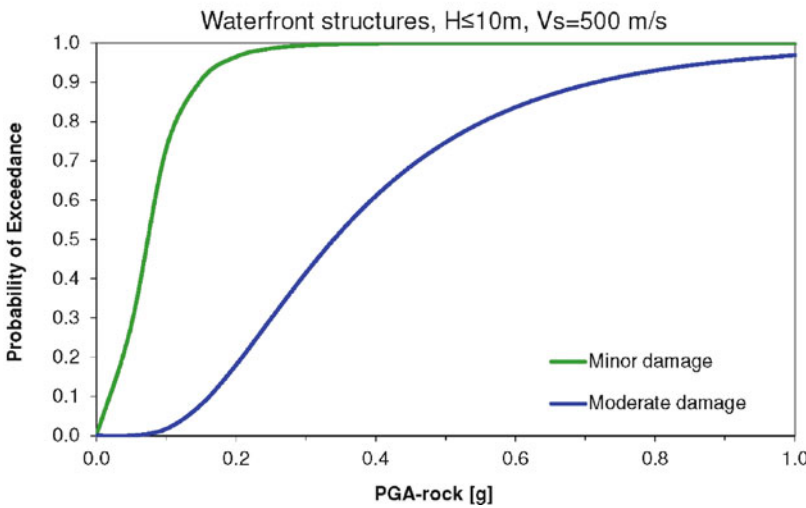
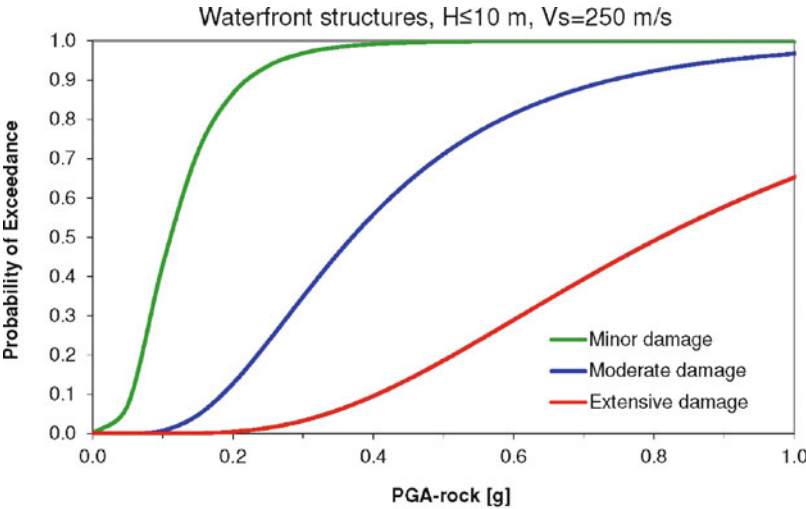
Seismic Vulnerability Assessment: Lifelines,

Fig. 12 Analytical fragility curves for alluvial bored tunnels due to ground shaking, classified to ground type *B*, *C*, and *D* according to Eurocode 8 (Argyroudis and Pitilakis 2012)



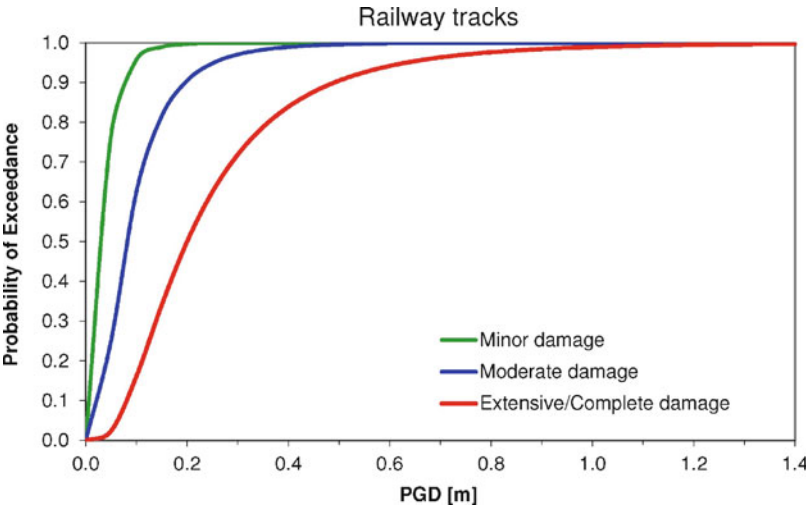
Seismic Vulnerability Assessment: Lifelines,

Fig. 13 Analytical fragility curves for gravity waterfront structures due to ground shaking, classified according to the wall height (H) and the soil foundation conditions (Vs values) (Kakderi and Pitilakis 2010)



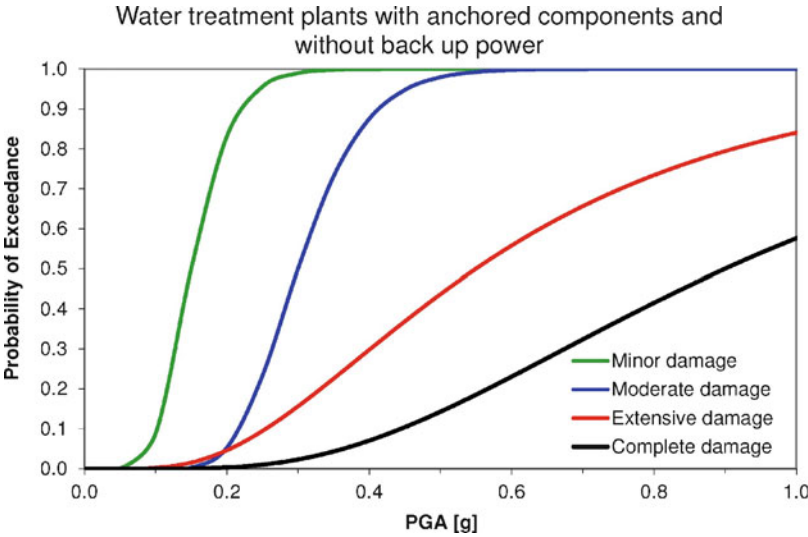
Seismic Vulnerability Assessment: Lifelines,

Fig. 14 Expert judgment fragility curves for railway tracks subjected to permanent ground deformation (Argyroudis and Kaynia 2014)



Seismic Vulnerability Assessment: Lifelines,

Fig. 15 Fault-tree based fragility curves for water treatment plant with anchored components subjected to ground shaking (Pitilakis et al. 2014a)



Seismic Vulnerability Assessment: Lifelines, Table 4 Example of damage estimation for steel tanks and pipelines

	On-ground unanchored steel tanks subjected to ground shaking (PGA = 0.4 g)				
Damage state	No damage	Minor	Moderate	Extensive	Complete
Probability of exceedance	–	0.919	0.571	0.240	0.108
Probability of occurrence	0.081	0.348	0.331	0.132	0.108
	(=1–0.919)	(=0.919–0.571)	(=0.571–0.240)	(=0.240–0.108)	
	Pipelines subjected to ground failure (PGD = 40 cm)				
Material diameter	Welded steel – large		Welded steel – small		PVC – small
Repairs/km	1.26		5.86		6.70

The performance of each network (e.g., utility or accessibility losses) is commonly measured through appropriate *performance indicators (PIs)*, which, if combined with direct losses from physical damages, can yield a first partial estimate of the overall socioeconomic impact of an earthquake. Performance indicators, at the component or the system level, depend on the type of analysis that is performed. Connectivity analysis gives access to indices such as the connectivity loss (measure of the reduction of the number of possible paths from sources to sinks). Capacitive modeling yields more elaborate performance indicators at the distribution nodes (e.g., head ratio for water system, voltage ratio for electric buses) or for the whole system (e.g., system serviceability index comparing the customer demand satisfaction before and after the seismic event). The fault-tree analysis method is generally used for the derivation of fragility

curves for specific components that comprise a set of subcomponents (e.g., health-care facilities, water treatment plants).

The importance of the interconnection between different systems is a more recent acquisition that targets two or, rarely, more systems (Pitilakis et al. 2014b). Several classifications have been proposed to categorize the types of interactions. The most common are physical, demand, and geographic interactions (Rinaldi et al. 2001). Physical interaction describes physical reliance on material flow from one infrastructural system to another as, for example, the supply of power to various network facilities by electric generators. Demand interactions correspond to a supply–demand from a given component to another system. An example is the number of casualties that should be treated by health-care facilities after an earthquake. Finally, geographic interactions describe the way that a local

environmental event affects components across multiple infrastructural systems due to physical proximity. For instance, the collapse of buildings in city centers can induce the blockage of adjacent roads due the debris accumulation.

A comprehensive methodological framework for the assessment of physical as well as socio-economic seismic vulnerability and risk of buildings, lifelines, and infrastructures at urban and regional level considering inter-element and intra-systems dependencies has been developed in SYNER-G project. The reader is referred to Pitilakis et al. (2014b) for more details and applications of the proposed framework.

Seismic Risk Assessment of Lifelines:

Examples

Representative examples of seismic risk assessment studies for lifelines are given in the following, for different scales of analysis.

Medium- to High-Voltage Electric Power Network of the Sicily Region, Southern Italy (Regional Scale)

A power flow analysis is performed for the electric power network of Sicily, which is composed of 181 nodes and 220 transmission lines (Cavalieri et al. 2014). The nodes, i.e., the buses, are subdivided into 175 demand or load nodes and six supply nodes, five of which are power plants and one is the balance node (or slack bus), which is coinciding with the generation node providing the highest power. The load nodes (two for transmission/distribution and one for distribution substations) deliver power to users. In total, 390 municipalities are served by the network. All transmission lines are overhead lines and considered as non-vulnerable elements. They are classified into high- (HV), medium- (MV), and low-voltage (LV) lines (Fig. 16a). The vulnerable elements are the components within substations, called micro-components. A probabilistic evaluation of the performance of network is carried out by means of Monte Carlo simulation by sampling seismic events for 18 faults taken from the Italian DISS database employing the truncated Gutenberg and Richter recurrence model for the source activity.

The distribution of performance losses is shown in Fig. 16c as the mean annual frequency (MAF) of exceedance for the system serviceability index (SSI). SSI provides a global scalar measure of the system performance. It is defined as the ratio of the sum of the real power delivered from load buses after an earthquake to that before the earthquake. For each bus inside the substations, voltage ratio (VR) is defined as the ratio of the voltage magnitude in the seismically damaged network to the reference value for non-seismic, normal conditions. Figure 16b displays a contour map of the expected values of VR, averaged on the whole simulation for each demand node. It can be seen that the reduction in voltage due to induced damage is less than the tolerated threshold of 10 %, allowing the power demand delivery everywhere in the island.

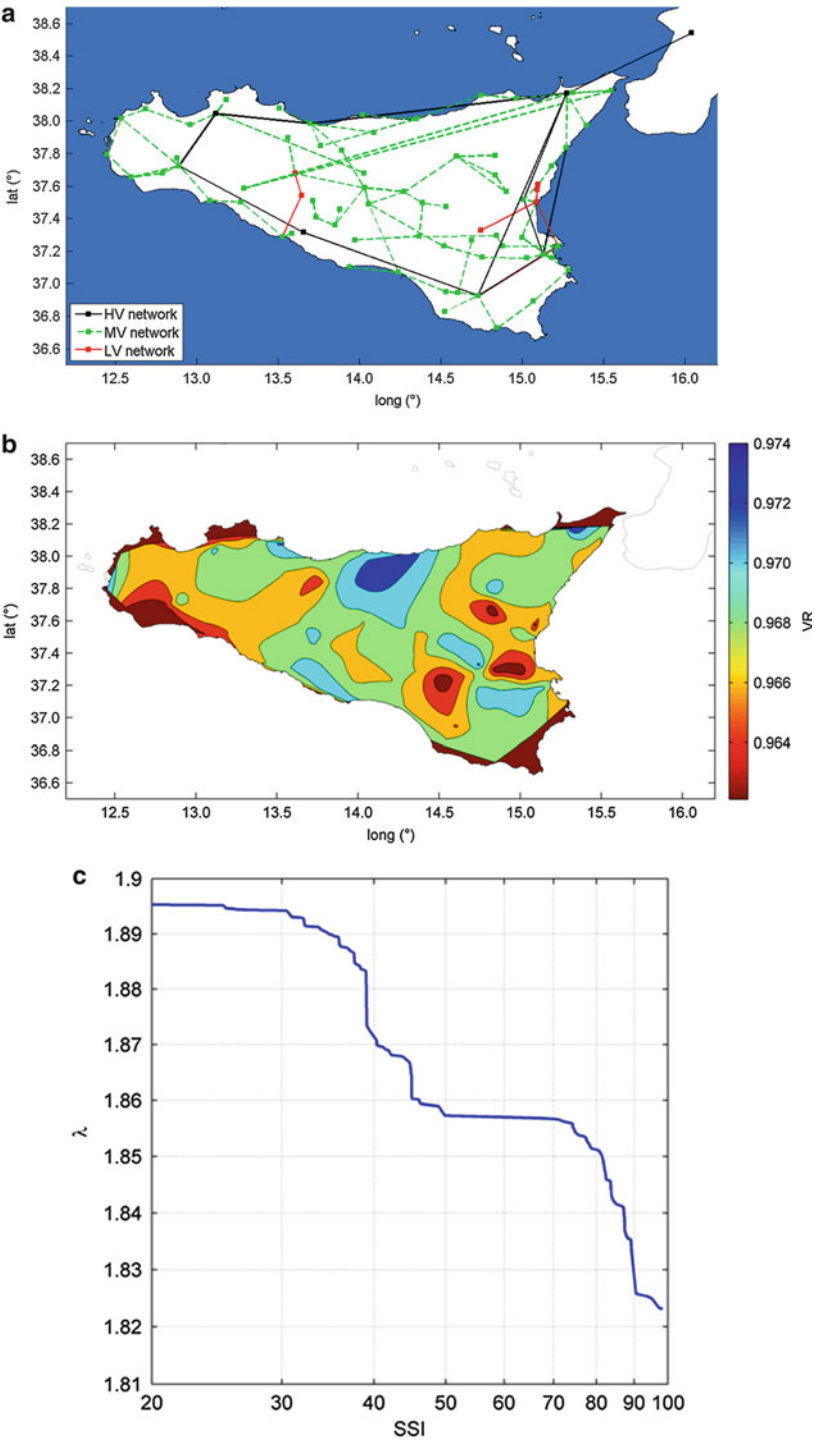
Water Network of Thessaloniki, Greece (City Level)

A connectivity analysis is performed for the main water system (WSS) of Thessaloniki in Greece considering the interaction of electric power network (EPN) with pumping stations (Pitilakis et al. 2014b). If a pump serving a source node is not fed by the reference EPN node due to damage in EPN substation, then the pump itself is considered out of service and the relative WSS node is removed from the system. A probabilistic evaluation of the performance of networks is carried out by means of Monte Carlo simulation by sampling seismic events for five seismic zones with $M_{\min} = 5.5$ and $M_{\max} = 7.5$. Pipeline damage is evaluated for each simulation considering both wave propagation and ground failure due to liquefaction. The network is analyzed for each sampled event and the results are aggregated all over the sampled events, in order to numerically obtain the marginal distribution of performance losses (Fig. 17). The interaction can be important; as an example the water connectivity loss is increased from 1 to 1.8 % for $\lambda = 0.001$ (corresponding to mean return period $T = 1,000$ years) when the connections of water pumping stations to EPN are included in the analysis.

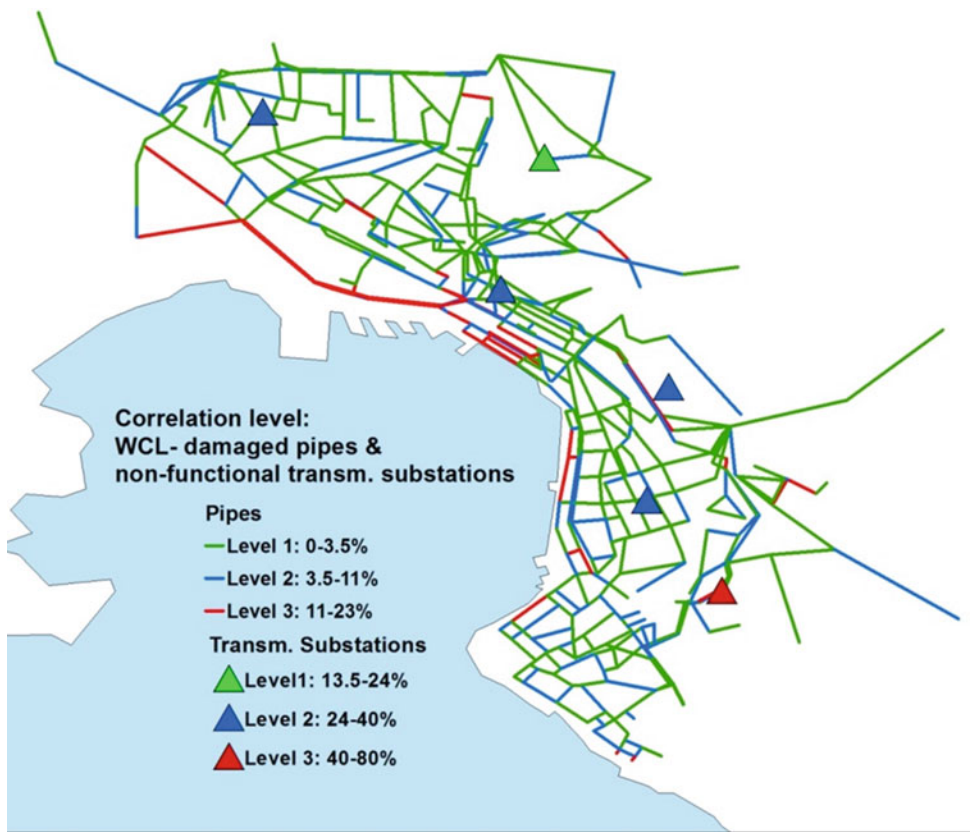
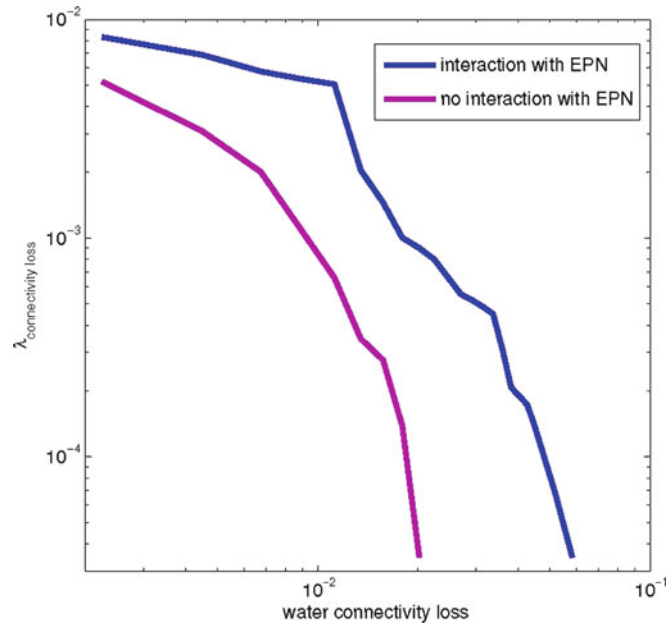
Figure 18 shows the level of correlation between the water connectivity loss and damages in pipes as well as the nonfunctional EPN

Seismic Vulnerability Assessment: Lifelines,

Fig. 16 Seismic risk assessment of electric power network in Sicily, Italy (Source: Cavalieri et al. 2013). (a) Transmission lines, classified by voltage. (b) Contour map of expected values of voltage ratio (VR). (c) Mean annual frequency (λ) curve for electric power system serviceability index (SSI)



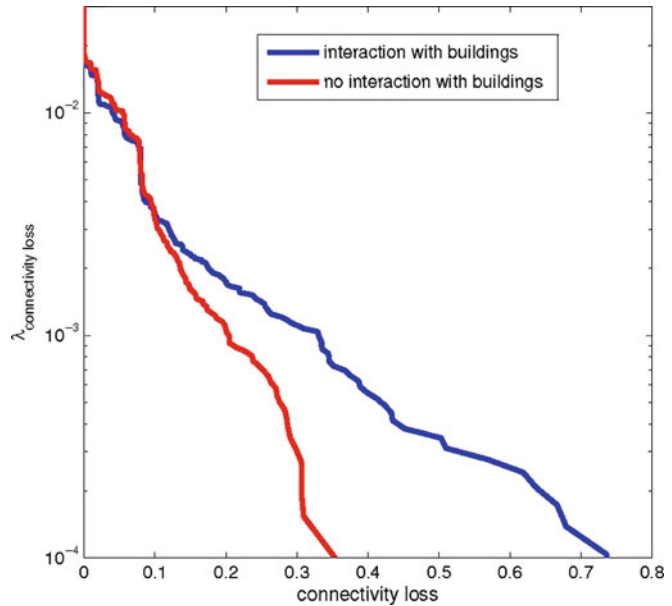
Seismic Vulnerability Assessment: Lifelines,
Fig. 17 Mean annual frequency (λ) curve for water connectivity loss with and without interaction with electric power network (EPN)



Seismic Vulnerability Assessment: Lifelines, Fig. 18 Correlation of damaged pipes and nonfunctional EPN transmission stations to water network connectivity

Seismic Vulnerability Assessment: Lifelines,

Fig. 19 Mean annual frequency (λ) curve for road network connectivity loss with and without interaction with collapsed buildings

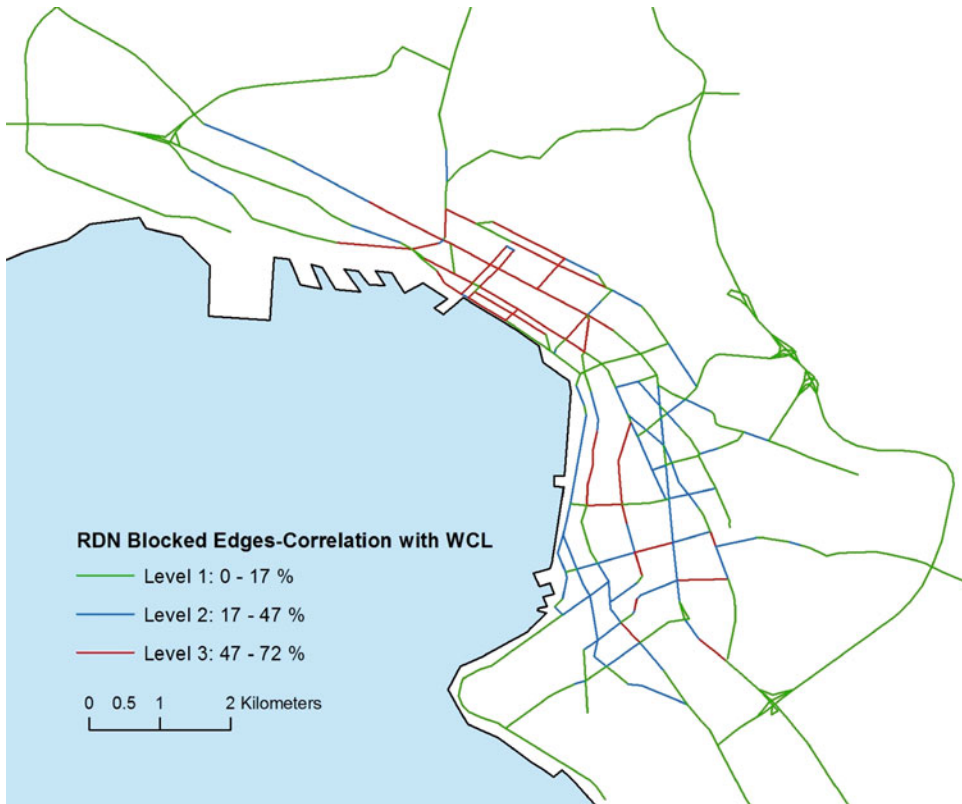


substations supplying the water pumping stations. The most correlated pipes are concentrated along the coastline where liquefaction susceptibility is high and therefore damages due to permanent ground displacement are expected. A higher level of correlation is obtained for the EPN transmission substations. The highest value of 80 % is attributed to a component in the south-east part of the city, where several pumping stations (connected to EPN) are located.

Road Network of Thessaloniki, Greece (City Level)
The main network of the urban area is considered in this case study, together with the ring road and the main exits of the city where the majority of bridges and overpasses are located. In particular, 594 nodes and 674 edges are included in the simulation. The nodes are subdivided into 15 external nodes, 127 traffic analysis zone (TAZ) centroids, and 452 simple intersections. Edges are assumed to be the only vulnerable components in the network. They are classified into road pavements and bridges, with fragility models expressed in terms of permanent ground deformation (PGD) due to liquefaction and peak ground acceleration (PGA) for ground shaking, respectively (Pitilakis et al. 2014b).

Road closures are estimated due to soil liquefaction and bridge damages. Moreover, the interaction with collapsed buildings that can induce road blockages is considered (Pitilakis et al. 2014b). A probabilistic evaluation of the network's performance is carried similarly to the one described in the previous example.

The interaction with building collapses can be important especially for return periods higher than 500 years ($\lambda = 0.002$). As an example the connectivity loss is increased from 20 % to 33 % for $\lambda = 0.001$ ($T = 1,000$ years) when the building collapses are included in the analysis (Fig. 19). Figures 20 and 21 show the level of correlation between the connectivity loss and the distribution of blockages due to building collapses and damage in bridges and road pavements, respectively. Relatively higher correlation factors are found for edges blocked by building collapse, demonstrating the importance of this failure mechanism in the analysis. A few road segments near the coastline which are subjected to ground failure due to liquefaction are also highly correlated to the network connectivity. The high risk of failure for bridges is attributed to their typology characteristics (old, simple span bridges) and the high values of PGA.



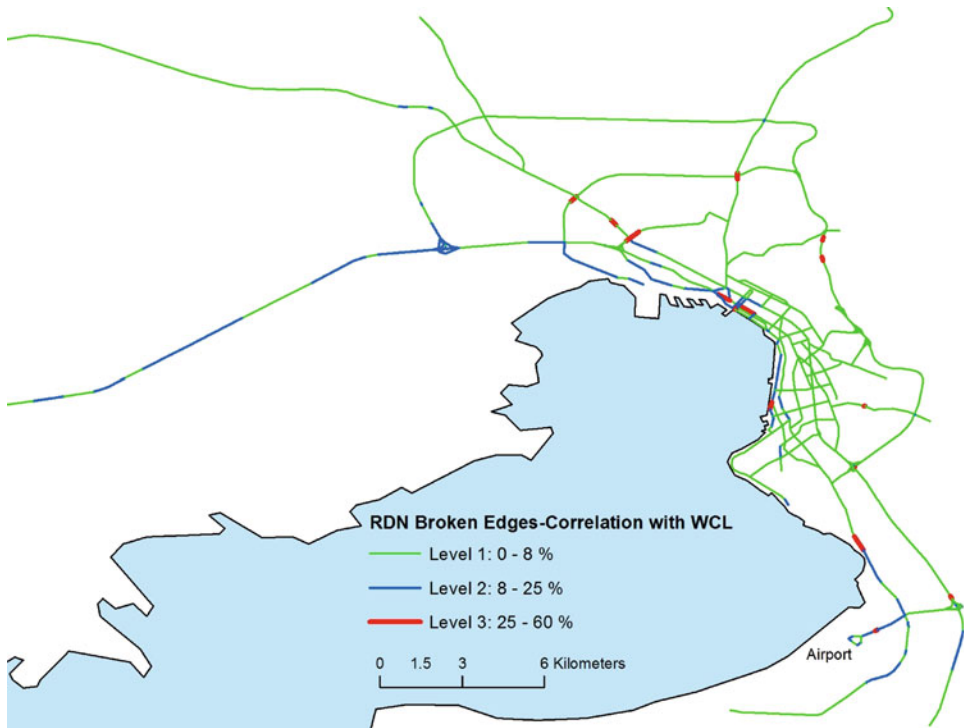
Seismic Vulnerability Assessment: Lifelines, Fig. 20 Correlation of edges blocked by buildings' collapse to road network connectivity

Port System of Thessaloniki, Greece (Infrastructure Level)

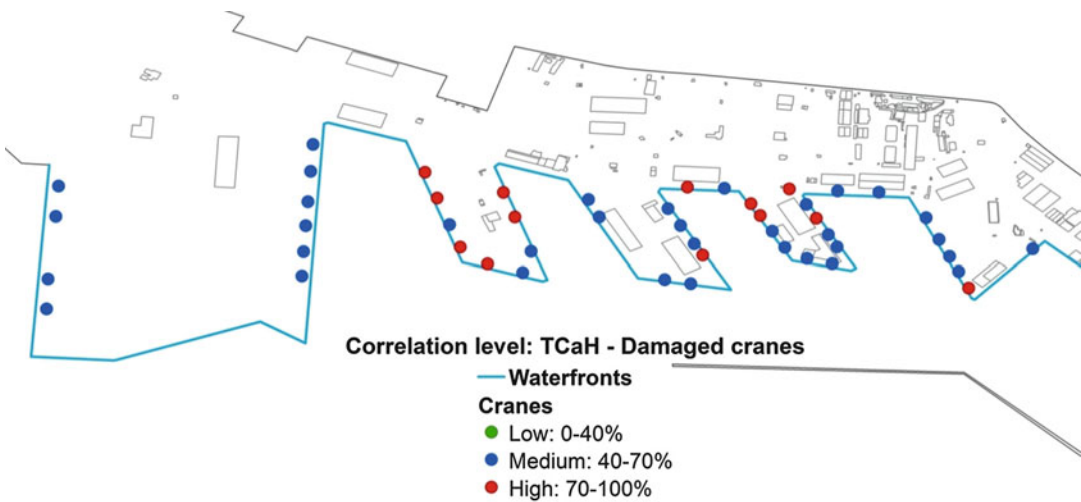
The port covers an area of 1,550,000 m² and trades approximately 16,000,000 t of cargo annually, having a capacity of 370,000 containers and six piers with 6,500 m length. In this case study, waterfront structures, cargo handling equipment, power supply system, roadway system, and buildings are examined. In particular, waterfront structures of a total 6.5 km length, 48 crane nodes, and two terminals (one container and one bulk cargo) are considered. The interactions accounted for in the analysis are the supply of EPN to cranes and the road closures due to building collapses. A probabilistic evaluation of the performance of networks is carried out by means of Monte Carlo simulation by sampling seismic events for five seismic zones affecting the city of Thessaloniki and the harbor with $M_{\min} = 5.5$ and

$M_{\max} = 7.5$. The performance of the port is described through the total cargo or containers handled in a predefined time frame per terminal and for the whole port system (Pitilakis et al. 2014b).

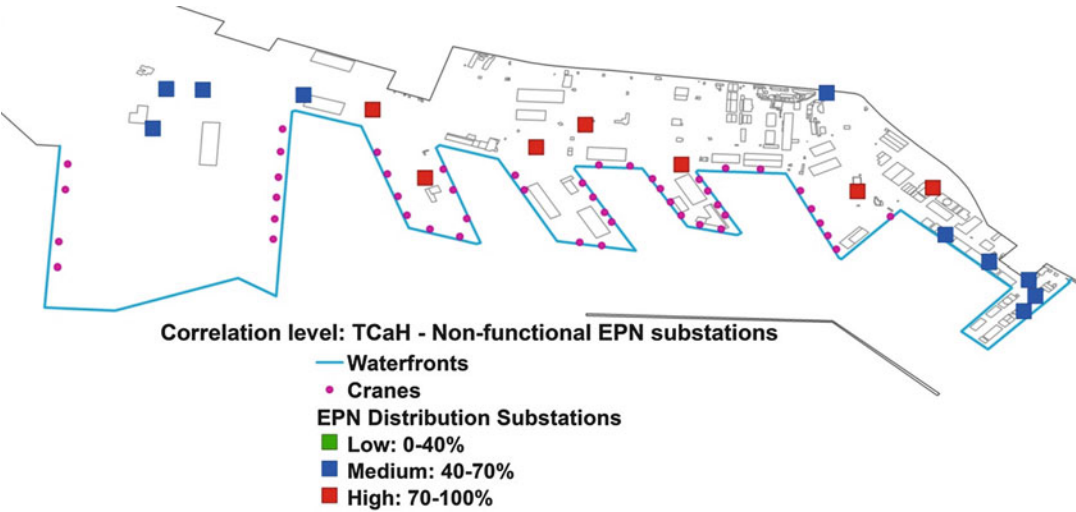
Figures 22 and 23 show the level of correlation between the total cargo handled per day (TCaH) and the distribution of damages in cranes and non-functionality of electric power distribution substations, respectively. In this way the most critical components can be defined in relation with their contribution to the performance loss of the system. All cranes have medium (40–70 %) to high (over 70 %) levels of correlation, indicating their great importance to the functionality of the overall port system. A higher level of correlation is estimated for the EPN distribution substations, with 40 % of the components having values greater than 70 %.



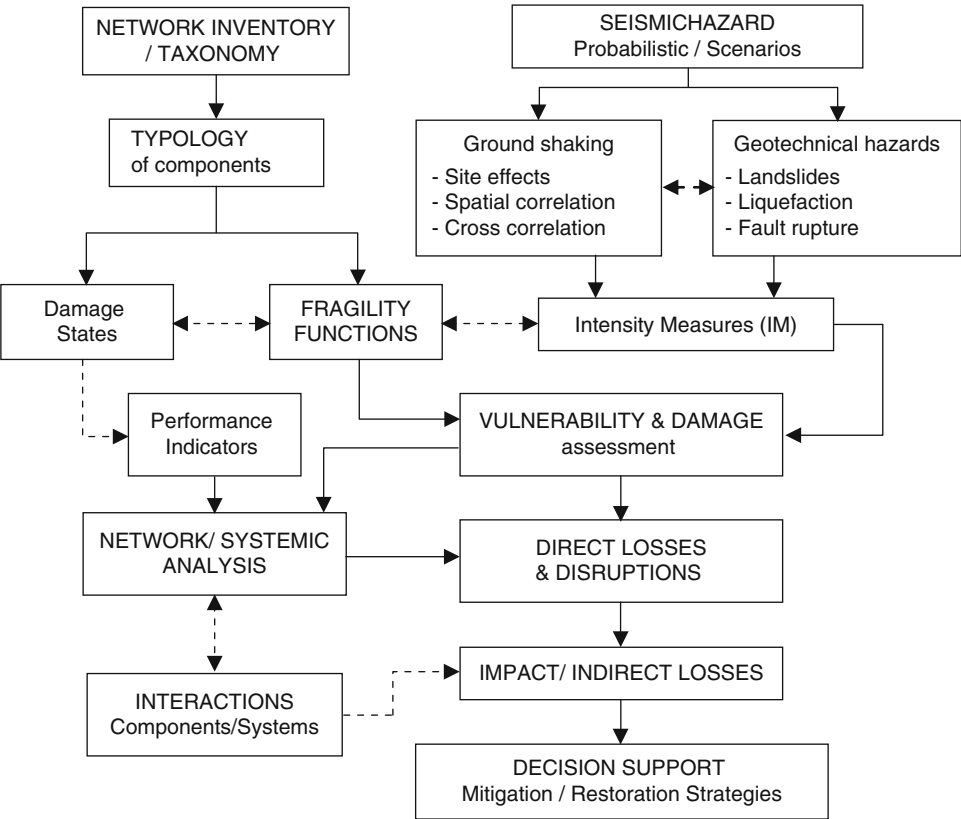
Seismic Vulnerability Assessment: Lifelines, Fig. 21 Correlation of broken edges (bridges due to ground shaking or road segments due to liquefaction) to road network connectivity



Seismic Vulnerability Assessment: Lifelines, Fig. 22 Correlation of damaged cranes to port performance (PI = TCaH)



Seismic Vulnerability Assessment: Lifelines, Fig. 23 Correlation of nonfunctional electric power distribution substations to port performance (PI = TCaH)



Seismic Vulnerability Assessment: Lifelines, Fig. 24 General layout for the seismic risk assessment of lifelines

Summary

Lifelines are spatially distributed systems that provide essential services to any modern society. They play also an important role for emergency response and restoration in the aftermath of disastrous earthquakes; in general they are vital for the resilience of society. They are often grouped to transportation and utility systems and comprise multiple components, which are exposed to different ground shaking effects and geotechnical hazards. Research efforts and studies undertaken the last 20 years, and after numerous devastating earthquakes that caused extensive losses and disruptions to lifelines and infrastructures, contributed to important improvement of knowledge and expertise in the seismic vulnerability and risk assessment of lifelines. The main objective of these efforts has been the development of methodologies and tools for the estimation of probable direct and indirect losses (physical, performance, economic, social) due to future earthquakes in order to develop efficient emergency response and mitigation strategies. A general layout for the seismic risk assessment of lifelines is outlined in Fig. 24. Inventory and taxonomy/typology definitions of all elements at risk and systems are the necessary initial steps to describe the elements and networks exposed to seismic risk. The estimation of site-specific seismic ground motion and other geotechnical hazards and the selection–estimation of appropriate seismic intensity measures (IM) is the basis of the seismic hazard analysis. Vulnerability is the expected response of an element or a network to a given seismic intensity. It is commonly assessed based on fragility models, which estimate the probability of exceeding certain damage states for given seismic intensity. Fragility functions are adapted with respect to the typological characteristics of each element at risk. The expected damage and loss of the entire network are estimated through appropriate systemic analysis, which takes into account the physical damage and the relations and interactions between the different components and systems. The interdependencies among different systems may considerably increase the overall impact. The results, which are commonly

provided in terms of performance indicators that describe the network functional losses or the expected economic or socioeconomic losses, provide the means for an efficient mitigation or recovery planning. Obviously, several sources of uncertainties are inherent in the various definitions, models, tools, and methods for estimation of losses, which make the risk assessment of lifelines a complex and challenging topic.

Cross-References

- [Damage to Infrastructure: Modeling](#)
- [Earthquake Risk Mitigation of Lifelines and Critical Facilities](#)
- [Empirical Fragility](#)
- [Liquefaction: Performance of Building Foundation Systems](#)
- [Probabilistic Seismic Hazard Models](#)
- [Seismic Actions Due to Near-Fault Ground Motion](#)
- [Seismic Fragility Analysis](#)
- [Seismic Loss Assessment](#)
- [Seismic Risk Assessment, Cascading Effects](#)
- [Seismic Vulnerability Assessment: Reinforced Concrete Structures](#)
- [Spatial Variability of Ground Motion: Seismic Analysis](#)
- [Strengthening Techniques: Bridges](#)
- [Uncertainty Theories: Overview](#)

References

- American Lifelines Alliance [ALA] (2001) Seismic fragility formulations for water systems. Part 1 – guideline. ASCE-FEMA, Washington, DC, 104 pp
- Argyroudis S, Kaynia AM (2014) Fragility functions of highway and railway infrastructure. In: Ptilakis K, Crowley H, Kaynia AM (eds) SYNER-G: typology definition and fragility functions for physical elements at seismic risk, vol 27, Geotechnical, geological and earthquake engineering. Springer, Dordrecht. doi:10.1007/978-94-007-7872-6_10
- Argyroudis S, Ptilakis K (2012) Seismic fragility curves of shallow tunnels in alluvial deposits. *Soil Dyn Earthq Eng* 35:1–12
- Basöz N, Kiremidjian AS (1998) Evaluation of bridge damage data from the Loma Prieta and Northridge, California earthquake. Technical report MCEER-98-0004. State University of New York, Buffalo

- Cardona OD, Ordaz MG, Reinoso E, Yamín LE, Barbar AH (2012) CAPRA – comprehensive approach to probabilistic risk assessment: international initiative for risk management effectiveness. In: Proceedings of the 15th world conference of earthquake engineering, Lisbon, 24–28 Sept 2012
- Cavalieri F et al (2013) Application and validation study to an electric power network in Italy. In: Pitilakis K, Argyroudis S (eds) Systemic seismic vulnerability and loss assessment: validation studies SYNER-G. Reference report 6, Publications Office of the European Union, doi: 10.2788/16706
- Cavalieri F, Franchin P, Pinto PE (2014) Application to selected transportation and electric networks in Italy. In: Pitilakis K et al (eds) SYNER-G: systemic seismic vulnerability and risk assessment of complex urban, utility, lifeline systems and critical facilities. Methodology and applications. Springer, Dordrecht. ISBN 978-94-017-8834-2
- Daniell J, Vervaeck A (2013) CATDAT Damaging earthquakes in 2012 – the year in review. <http://earthquake-report.com/2013/01/07/damaging-earthquakes-2012-database-report-the-year-in-review/>
- Douglas J, Seyedi DM, Ulrich T, Modaressi H, Foerster E, Pitilakis K, Pitilakis D, Karatzetou A, Gazetas G, Garini E, Loli M (2014) Evaluation of seismic hazard for the assessment of historical elements at risk: description of input and selection of intensity measures. *Bull Earthquake Eng*. doi:10.1007/s10518-014-9606-0
- Dueñas-Osorio L, Craig JJ, Goodno BJ (2007) Seismic response of critical interdependent networks. *Earthq Eng Struct* 36(2):285–306
- Elnashai A, Hampton S, Lee JS, McLaren T, Myers JD, Navarro C, Spencer B, Tolbert N (2008) Architectural overview of MAEvis-HAZTURK. *J Earthq Eng* 12(S2):92–99. doi:10.1080/13632460802013610
- Faccioli E (ed) (2007) Prediction of ground motion and loss scenarios for selected infrastructure systems in European urban environments: LESSLOSS report No. 2007/08, ISBN: 978-88-6198-012-9. IUSS Press, Pavia
- Federal Emergency Management Agency [FEMA] (2003) Multi hazard loss estimation methodology: earthquake model – HAZUS-MH MR3 technical manual. FEMA, Washington, DC
- Kakderi K, Pitilakis K (2010) Seismic analysis and fragility curves of gravity waterfront structures. In: Fifth international conference on recent advances in geotechnical. *Earthquake Engineering and Soil Dynamics and Symposium in Honour of Prof. I. M. Idriss*, 6.04a, San Diego, CA, May 24–29
- Mackie K, Stojadinovic B (2003) Seismic demands for performance-based design of bridges. PEER Report 2003/16. Pacific Earthquake Engineering Research Center, University of California, Berkeley
- Modaressi H, Desramaut N, Gehl P (2014) Specification of the vulnerability of physical systems. In: Pitilakis K et al (eds) SYNER-G: systemic seismic vulnerability and risk assessment of complex urban, utility, lifeline systems and critical facilities. Methodology and applications. Springer, Dordrecht. ISBN 978-94-017-8834-2
- lifecycle systems and critical facilities. Methodology and applications. Springer, Dordrecht. ISBN 978-94-017-8834-2
- Mouroux P, Le Brun B (2006) Risk-UE project: an advanced approach to earthquake risk scenarios with application to different European towns. In: Oliveira CS, Roca A, Goula X (eds) Assessing and managing earthquake risk. Springer, Netherlands, pp 479–508. doi:10.1007/978-1-4020-3608-8_23
- NCEER (1995) The Hanshin-Awaji earthquake of January 17, 1995: performance of lifelines, technical report NCEER-95-0015 (ed: Shinozuka M), State University of New York, Buffalo
- O'Rourke TD, Jeon SS, Toprak S, Cubrinovski M, Jung JK (2012) Underground lifeline system performance during the Canterbury earthquake sequence. In: Proceedings of the 15th world conference on earthquake engineering, Lisbon
- Pinto P (2014) Modeling and propagation of uncertainties. In: Pitilakis K, Crowley H, Kaynia AM (eds) SYNER-G: typology definition and fragility functions for physical elements at seismic risk, vol 27, Geotechnical, geological and earthquake engineering. Springer, Dordrecht. doi:10.1007/978-94-007-7872-6_2
- Pitilakis K, Crowley E, Kaynia A (eds) (2014a) SYNER-G: typology definition and fragility functions for physical elements at seismic risk, vol 27, Geotechnical, geological and earthquake engineering. Springer, Heidelberg. ISBN 978-94-007-7872-6
- Pitilakis K, Franchin P, Khazai B, Wenzel H (eds) (2014b) SYNER-G: systemic seismic vulnerability and risk assessment of complex urban, utility, lifeline systems and critical facilities. Methodology and applications, Geotechnical, geological and earthquake engineering. Springer, Heidelberg. ISBN 978-94-017-8834-2
- Rinaldi SM, Peerenboom JP, Kelly TK (2001) Identifying, understanding, and analyzing critical infrastructure interdependencies. *IEEE Contr Syst Mag* 21(6):11–25
- Satamtira G, Duenas-Osorio L (2010) Synthesis of modeling and simulation methods on critical infrastructure interdependencies research. In: Gopalakrishnan K, Peeta S (eds) Sustainable and resilient critical infrastructure systems. Springer, Berlin/Heidelberg. doi:10.1007/978-3-642-11405-2
- Seyedi DM, Gehl P, Douglas J, Davenne L, Mezher N, Ghavamian S (2010) Development of seismic fragility surfaces for reinforced concrete buildings by means of nonlinear time-history analysis. *Earthq Eng Struct* 39:91–108
- Tang C, Zhu J, Qi X (2011) Landslide hazard assessment of the 2008 Wenchuan earthquake: a case study in Beichuan. *Can Geotech J* 48:128–145
- Weatherill G, Esposito S, Iervolino I, Franchin P, Cavalieri F (2014) Framework for seismic hazard analysis of spatially distributed systems. In: Pitilakis K et al (eds) SYNER-G: systemic seismic vulnerability and risk assessment of complex urban, utility, lifeline systems and critical facilities. Methodology and applications. Springer, Dordrecht. ISBN 978-94-017-8834-2

Seismic Vulnerability Assessment: Masonry Structures

Dina D'Ayala and Viviana Novelli
Department of Civil, Environmental and
Geomatic Engineering, University College
London, London, UK

Introduction

The assessment of the vulnerability of the building stock of an urban center is an essential prerequisite to its seismic risk assessment. The other two ingredients are the expected hazard over given return periods and the distribution and values of the assets constituting the building stock. All three elements of the seismic risk assessment are affected by uncertainties of aleatory nature, related to the spatial variability of the parameters involved in the assessment, and epistemic nature, related to the limited capacity of the models used to capture all aspects of the seismic behavior of buildings and to describe them in simple terms, suitable for this type of analysis. Hence it should always be kept in mind that the computation of a risk level is highly probabilistic and that to accurately represent the risk, the expected values should always be accompanied by a measure of the associated dispersion. A very preliminary estimate of the seismic capacity of the local building stock can be obtained by consulting requirements included in seismic standards and code of practices in force at the time of construction of such buildings. This information together with a temporal and spatial record of the growth of the urban center can provide a definition of classes of buildings assumed to have different capacities. This information can be obtained by looking at past and present cadastral maps with ages of buildings and knowing the historical development and enforcement of codes at the site. In general however for a correct assessment of the seismic risk, a more detailed inventory and classification should be considered, the extent of which is a function of the economic and

technical resources available and of the dimensions and building density of the area under investigation, as well as the diversity within the building stock.

Seismic vulnerability analysis of masonry structures has traditionally been carried out using “empirical” (or statistical) methods. These can be divided in categorization methods, which classify buildings into typologies characterized by propensity to damage, or inspection and rating methods, wherein scores are attributed to each significant characteristic of the buildings thought to have a bearing on its overall vulnerability (UNDP/UNIDO 1985). Several of these methods have been developed in the past 30 years, and the more robust and commonly used are reviewed in section “[Empirical Methods](#).”

Recent dramatic increase in computational powers and consequent improvement and refinement in numerical modelling of relatively complex structures, using either static or dynamic approaches (e.g., Fajfar and Gašperšič 1996; Vamvatsikos and Cornell 2005), has facilitated greater exploitation of analytical approaches for vulnerability assessment purposes. Analytical methods need experimental validation of the various parameters used to define the vulnerability, and so far relatively little experimental work on the behavior of masonry structures has been undertaken by the international community, albeit a wide variety of existing unreinforced masonry typologies, both historic and modern, exists. The framework within which analytical methods can be applied and their limitation are presented in section “[Analytical Methods](#).”

An attempt to exploit positive aspects of the two classes of methods above is made with the so-called hybrid methods. These combine numerical input/output from analytical models with statistical and probabilistic data to define exposure and vulnerability distribution. This allows to reduce the analytical burden while grounding results in a geographical context. Applicability and reliability of such approaches are constrained by the ability to define certain mechanical and structural characteristics in numerical terms and by the need to define a common approach for treating the various sources of uncertainty

associated with vulnerability, exposure, and hazard. This is further discussed in section “[Hybrid Methods](#).”

The review and discussion of existing methods sets out the rationale for a combined method, called FaMIVE (Failure Mechanism Identification and Vulnerability Evaluation), and developed by the first author and her co-researchers, in the past decade. The hypotheses and analytical approach to derive capacity curves and fragility functions are presented in detail in section “[Development of Analytical Methods: FaMIVE](#).” Results obtained from the FaMIVE method are presented in section “[Damage Thresholds and Performance Indicators](#)” in terms of damage limit states and drift and compared with the recommendation of Eurocode 8 (EC8) (CEN 2005) and experimental evidence to discuss issues of validation and calibration.

Vulnerability Approaches

Empirical Methods

The damage probability matrix (DPM) method has been the most common vulnerability assessment method used since the 1970s. Based on field observations, it expresses in a discrete form the conditional probability of obtaining a damage level D_j , due to a ground motion of macro intensity I_i ,

$$P = \left[D = \frac{D_j}{I_i} \right].$$

The method relies on the wealth of observed damage data available from past earthquakes and the correlation of these, on the one hand, with assigned shaking macroseismic intensity at the site of observation and, on the other, with construction materials and methods in different geographical and seismic regions. First proposed by Whitman et al. (1973) after the San Fernando earthquake of 1971, specific applications of this method to masonry structures are numerous (Braga et al. 1982; Corsanego and Petrini 1990; Di Pasquale et al. 2005) and still widely used in developing countries and regions with extensive

historic seismicity records (Askan and Yucemen 2010; Zobin et al. 2010). Notwithstanding its popularity, the DPM has major limitations: discrete definition of damage levels/states and dependence on a specific seismic and architectural context. Hence, it may not be applicable to different geographic locations, in the absence of direct damage data.

Some of these limitations are overcome by the Vulnerability Index Method (VIM) (Benedetti et al. 1988), based on the vulnerability index I_v . The vulnerability index is a summation of weighted parameters, each associated with a constructional or structural feature of the typology, which affects its vulnerability. In this way the definition of vulnerability relates only to the mechanical characteristics of the building, while damage data from past earthquakes are used to calibrate the vulnerability functions, by relating I_v to observed global damage levels for buildings of the same typology and hence extending applicability in regions having experienced the same level of macroseismic intensity or peak ground acceleration (PGA).

The most substantial improvement of the VIM over the DPM is that the former provides a continuous vulnerability function, while the latter uses discrete vulnerability classes expressed in terms of expected damage. This means that possible interventions to shift the vulnerability level of a structural typology are readily quantifiable with I_v , while they are only recorded in the DPM if the intervention results in a shift of damage class. Moreover, the correlation of VIM with an instrumental measure of seismic action allows for application to and comparison between different seismic zones, even though in the last decade the reliability of correlating PGA and damage has been put into question by various authors (Elenas 2000; Wu et al. 2003). Giovinazzi and Lagomarsino (2004) have proposed a direct correlation of the two methods by introducing a definition of damage states and DPM as function of the I_v using the definition of the EMS-98 macroseismic scale (Grünthal 1998). This approach has been further developed in the European RISK-UE project for larger number of building typologies and vulnerability classes (Lagomarsino and Giovinazzi 2006).

The EMS-98 scale (Grünthal 1998) is in itself a DPM approach, where vulnerability classes are defined across different construction typologies. The uncertainty inherent in the attribution of a typology to a class and of its response to a seismic event in terms of damage level is dealt with by assuming a central value and a fuzzy set membership, so that both vulnerability and damage can be expressed in descriptive (few, many, most) rather than numerical terms. A method to compute total probability of damage of a given typology spreading over more vulnerability classes is proposed in Jaiswal et al. (2011).

A first attempt at deriving generic fragility curves was made by Spence et al. (1992) with the Parameterless Scale Intensity (PSI), aimed at providing fragility curves for different building typologies and at quantifying benefits from upgrading and strengthening interventions of the building stock. The advantage of this approach is that damage grades and fragility curves are independent of macroseismic intensity scales, as they are defined relative to each other, on the assumption that for each building type, it is possible to define the level of the scale corresponding to the median of the fragility curve for level of damage D3 (structural damage). Limitations of applicability are however related to the extensive need for observations to calibrate each fragility curve and to the difficulty to correlate observed or expected damage into construction and structural characteristics in a deterministic way.

Analytical Methods

Analytical methods present the advantage of framing the problem of seismic vulnerability of masonry structures in structural engineering terms, defining a direct relationship among construction characteristics, structural response to seismic action, and damage effects.

The development of attenuation equations for specific seismic regions and corresponding derivation of seismic hazard maps in terms of spectral ordinates, as opposed to macroseismic intensity or PGA, has given impetus to the development of analytical methods. These methods tend to feature more detailed and transparent

vulnerability assessment algorithms with direct physical meaning. In the development of analytical fragility functions for masonry structures, two methods can be identified: (i) correlation between damage index and damage thresholds and (ii) correlation between acceleration/displacement capacity curves and spectral demand curves, following the HAZUS99 (FEMA 2001) or N2 method (Fajfar and Gašperšič 1996). These two procedures, in line with the performance-based design for new built, define the seismic assessment of existing building stock and prediction of losses. The required analysis steps are (i) classification of buildings by typology and seismic design, (ii) definition of damage states, (iii) assignment of capacity curves, (iv) definition of demand spectra (associated with return periods and performance targets), and (v) evaluation of building response in terms of performance points. The two procedures, applied to large sets of buildings, allow to derive fragility curves and damage scenarios for given sites and as a function of ground motion parameters which are linked with hazard levels. The damage levels are directly related to the demand parameter, which is expressed in terms of displacement or drift. As the fragility curves are developed on the basis of lognormal distribution, once the typologies have been defined, it is not necessary to have a detailed knowledge of the building stock. All required are the parameters defining the capacity curve for each typology, various damage thresholds, and the number of buildings belonging to each typology. Once capacity curves are developed, then the range of behavior and variability of fragility within a building class or stock can be analyzed by parametric analysis, providing important insight for retrofitting (see D'Ayala 2005). Analytical approaches to define seismic vulnerability of masonry buildings are becoming more and more popular, as improved engineering knowledge on the behavior of masonry structures increases confidence on the reliability of such models. In the past decade, a relatively significant number of procedures aimed at defining reliable analytical vulnerability functions for masonry structures in urban context have been proposed (Lang and Bachmann 2004; Erberik 2008; Borzi

et al. 2008). Although they share similar conceptual hypotheses, they differ substantially in modelling/numerical complexity and treatment of uncertainties. Two main limitations are common to all: limited geographic applicability and limited number of failure mechanisms. Specifically, for the latter, most of the analytical models only consider in-plane or frame-like behavior, disregarding overturning and out-of-plane mechanisms, often occurring at lower levels of shaking and hence substantially affecting the seismic vulnerability in an urban context. Analytical vulnerability assessment methods for existing masonry buildings feature in many risk assessment integrated systems; however, their choices of capacity curves are predetermined. Among nonlinear analysis methods, the collapse-mechanism or limit-equilibrium method presents the advantage of requiring few input parameters and allowing consideration of different failure modes. Procedures using this approach are based on collapse load multipliers which identify the occurrence of different possible mechanisms for given typology and structural characteristics. Bernardini et al. (1990) developed a numerical routine (VULNUS), combined with a fuzzy set theory, in recognition of the limited knowledge associated with many of the parameters. Only one in-plane and one simple out-of-plane mechanism were initially considered as possible failures. D'Ayala and Speranza (2003) further enhanced this approach by developing the FaMIVE method, a mechanical approach based on a suite of 12 possible failure mechanisms directly correlated with in situ observed damage. The FaMIVE algorithm produces vulnerability functions for different building typologies and quantifies the effect of strengthening and repair intervention on reduction of vulnerability. The method has been applied in several locations worldwide and recently in Abruzzo, Italy, following the 2009 L'Aquila earthquake (D'Ayala and Paganoni 2011).

Hybrid Methods

Hybrid methods combine post-earthquake damage statistics with analytically derived nonlinear behavior using pushover analysis (e.g., Kappos et al. 1998; Barbat et al. 1996). Recently, Kappos

et al. (2008) have applied it for estimating direct losses for masonry buildings in Thessaloniki (Greece). This procedure is very useful when the damage data collection of an area of interest with a specific intensity is only partially available. However, the use of different data sources derived from different procedures, for which a direct cross-correlation is not readily available, might not necessarily result in reduced uncertainty of the output.

An application of this approach to masonry structure within the Risk-UE project shows the importance of knowledge of the building stock for reliable seismic vulnerability assessment, as increased level of uncertainties is compounded both on capacity and demand curves (Lagomarsino and Giovinazzi 2006). Application made using code prescription in terms of capacity curves provides a good first approximation but needs extensive in situ calibration as shown by Erberick (2008).

A summary of the methods reviewed above is presented in Table 1. While this is by no means an exhaustive list of the many applications of seismic vulnerability assessment for masonry structures available in literature, it concentrates on procedures that have been specifically developed for vulnerability assessment at territorial scale. For this reason procedures aimed at the assessment of single buildings are not included. The choice of the most suitable procedure is highly dependent on the resources available for the data collection, the computational expertise available, and ultimately the scale and aim of the study. Empirical procedures can be used for fairly large-scale studies to define damage scenarios; however, if the purpose of the study is to identify within a district or urban center specific buildings in need of strengthening, so as to increase their seismic capacity, then a suitable analytical procedure should be preferred.

Development of Analytical Methods: FaMIVE

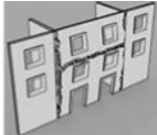
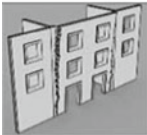
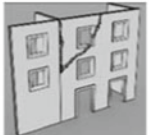
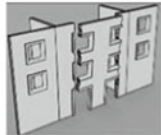
The seismic vulnerability assessment of unreinforced masonry historic buildings can be

Seismic Vulnerability Assessment: Masonry Structures, Table 1 Summary of procedures for the seismic vulnerability assessment of masonry structures

Method	References	Data requirement	Data collection	Assessment type	Approach	Demand input	Output
VULNUS	Bernardini et al. (1990)	Geometry, material parameters, structural details	On-site observation or systematic survey	Analytical	Mechanism method	PGA, response spectrum	Lateral force, lateral drift, vulnerability index
FaMIVE	D'Ayala and Speranza (2003); D'Ayala (2005)	Geometry, material parameters, structural details	On-site observation or systematic survey	Analytical	Capacity spectrum-based method, mechanism	PGA, response spectrum	Lateral force, lateral drift, damage scale, vulnerability index
SP-BELA	Borzi et al. (2008)	Structural description	Random generation	Analytical	Pushover-based method	PGA, response spectrum	Lateral force, lateral drift
CSBM	RISK-UE Kappos et al. (2008)	Structural description	Systematic survey and random generation	Analytical	Capacity spectrum-based method	PGA, response spectrum	Lateral force, lateral drift
VIM	Benedetti et al. (1988); Lagomarsino and Giovinazzi (2006)	Typological description	On-site observation	Empirical	Vulnerability index method	Macroseismic intensity, PGA	Phenomenological damage scale
DPM	Whitman et al. (1973)	Typological description	On-site observation	Empirical	Damage probability matrix	Macroseismic intensity	Phenomenological damage scale
AeDES	Baggio et al. (2009)	Typological description	On-site observation	Empirical	Vulnerability index method	Macroseismic intensity	Phenomenological damage scale
PSI	Spence et al. (1992)	Typological description	Exposure database	Empirical, statistical	Damage probability function	Macroseismic intensity	Phenomenological damage scale
Hybrid Methods	Kappos et al. (1998); Barbat et al. (1996)	Geometry, material parameters	On-site observation	Empirical/analytical, statistical basis	Hybrid methods	Macroseismic intensity	Damage scale, lateral forces

Seismic Vulnerability Assessment: Masonry Structures,

Fig. 1 Mechanisms for computation of limit lateral capacity of masonry façades

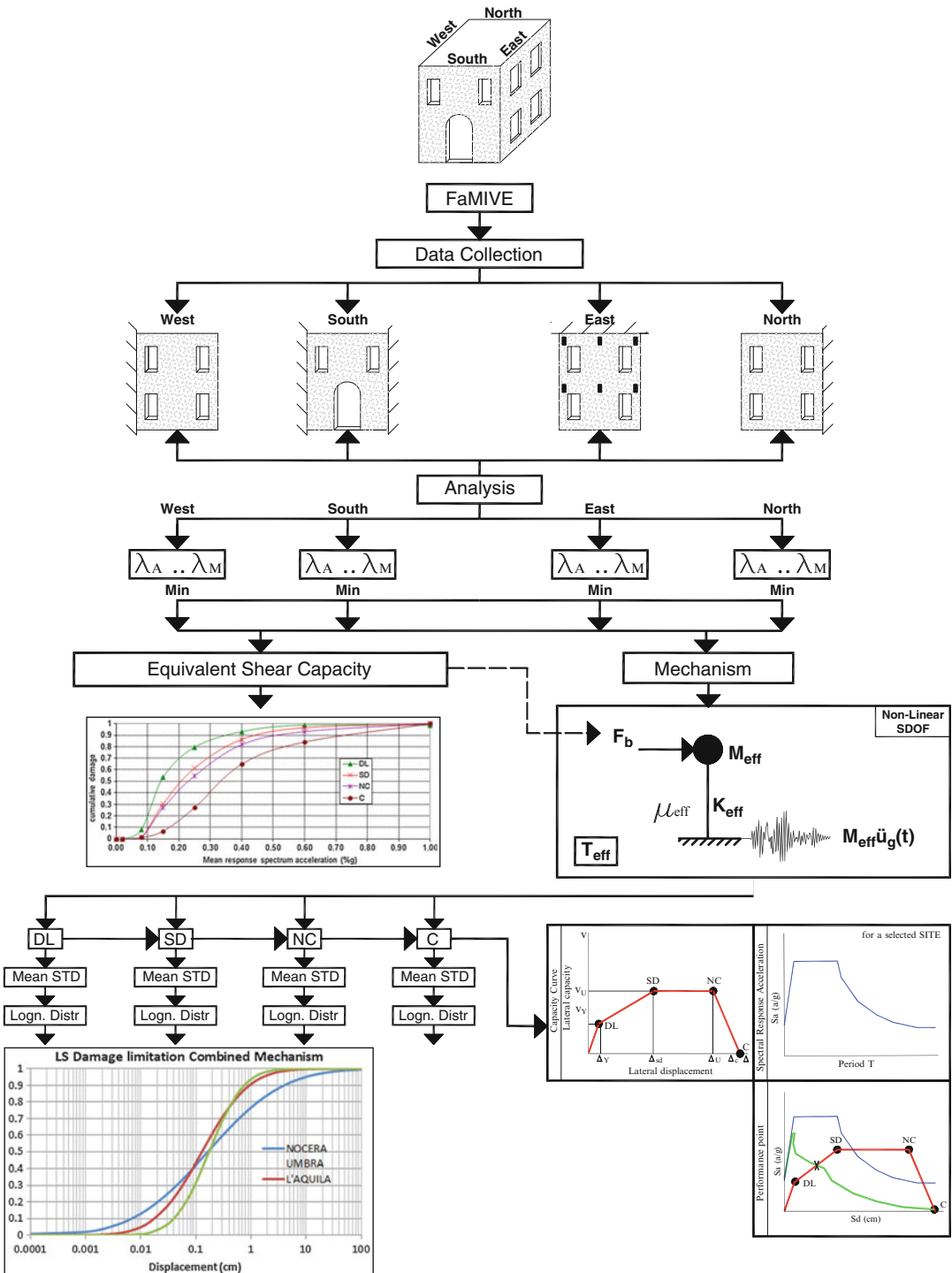
Combined Mechanisms			
			
B1: façade overturning with one side walls	B2: façade overturning with two side walls	C: overturning with diagonal cracks involving corners	F: overturning constrained by rings beams or ties
In plane Mechanisms			
			
H1: diagonal cracks mainly in piers	H2: diagonal cracks mainly in spandrel	M1: soft storey due to shear	M2: soft storey due to bending
Out of Plane Mechanism			
			
A: façade overturning with vertical cracks	D: façade overturning with diagonal crack	E: façade overturning with crack at spandrels	G: façade overturning with diagonal cracks

performed with the Failure Mechanisms Identification and Vulnerability Evaluation (FaMIVE) analytical method (D'Ayala and Speranza 2003; D'Ayala 2005). The FaMIVE method uses a nonlinear pseudo-static structural analysis with a degrading pushover curve to estimate the performance points by way of a variant of the N2 method (Fajfar and Gašperšič 1996), included in EC8 part 3 (CEN 2005). It yields as output collapse load multipliers which identify the occurrence of possible different mechanisms for a given masonry construction typology, given certain structural characteristics.

It is based on computing the collapse load factor for 12 different failure modes, shown in Fig. 1. Each mode of failure corresponds to different constraints and conditions between the façade and the rest of the structure; hence a collapse mechanism can be univocally defined and its collapse load factor computed for each mechanism developing

over either part or the whole façade. As shown in the flowchart of Fig. 2, the program FaMIVE first calculates the collapse load factor for each façade in a building and then, taking into account geometric and structural characteristics and constraints, identifies the one which is most likely to occur considering the combination of the largest portion mobilized with the lowest collapse load factor at building level.

The FaMIVE algorithm produces vulnerability functions in terms of ultimate lateral capacity for different building typologies and quantifies the effect of strengthening and repair intervention on reduction of vulnerability. In its current version, it also computes capacity curves and performance points and outputs fragility curves for different seismic scenarios in terms of intermediate and ultimate displacements or ultimate acceleration. Within the FaMIVE database, capacity curves and fragility functions are available for



Seismic Vulnerability Assessment: Masonry Structures, Fig. 2 Flowchart setting out the rationale of the FaMIVE procedure

various unreinforced masonry typologies, from adobe to concrete blocks, for a number of reference typologies studied at sites in Italy, Spain, Slovenia, Turkey, Nepal, India, Iran, and Iraq. The procedure has been validated against the EMS-98 vulnerability classes and recently used to produce capacity curves and fragility curves for use in the USGS PAGER environment. The mechanism's characteristics are used to derive an equivalent nonlinear single degree-of-freedom capacity curve to be compared to a spectrum demand curve and eventually define performance points as illustrated in the flow-chart in Fig. 2.

In order to derive fragility curves, limit state performance criteria correlated to damage states need to be defined. This step is fraught with uncertainties, as very limited consolidated evidence exists to perform such correlation over a wide range of building typologies and shaking levels. While robust database of damage states exist in literature, no attempt has been so far made to record permanent drift and corresponding ground shaking in a consistent way so as to provide empirical evidence for capacity curves. As an alternative, a number of authors have worked on correlating performance indicators and damage indicators on experimentally obtained capacity curves, by way of shaking table tests or pushover tests.

Definition of Capacity Curve

Capacity curves can be derived for each façade on the basis of the following steps. The first step is to calculate the lateral effective stiffness for each wall and its tributary mass. As the floor and roof structures in traditional masonry buildings have relatively modest in-plane stiffness, it is assumed that redistribution of inertia forces among walls will be ineffective and torsional effects are not considered. The effective stiffness of each external wall is calculated on the basis of the type of mechanism attained, the geometry of the wall and layout of opening, the constraints to other walls and floors, and the portion of other walls involved in the mechanism:

$$K_{\text{eff}} = k_1 \frac{E_t I_{\text{eff}}}{H_{\text{eff}}^3} + k_2 \frac{E_t A_{\text{eff}}}{H_{\text{eff}}} \quad (1)$$

where H_{eff} is the height of the portion involved in the mechanism; E_t is the estimated modulus of the masonry as it can be obtained from experimental literature for different masonry typologies; I_{eff} and A_{eff} are the second moment of area and the cross-sectional area, respectively, calculated taking into account extent and position of openings and variation of thickness over height; and k_1 and k_2 are constants which can assume values equal or greater than 0, depending on edge constraints and whether shear and/or flexural stiffness are relevant for the specific mechanism. For instance, for in-plane mechanisms, the inclusion of shear stiffness in Eq. 1 will depend on the slenderness of the pier and the relative stiffness between piers and spandrels of the façade being assessed.

The tributary mass Ω_{eff} is calculated following the same approach, and it includes the portion of the elevation activated by the mechanisms plus the mass of the horizontal structures involved in the mechanism:

$$\Omega_{\text{eff}} = V_{\text{eff}} \delta_m + \Omega_f + \Omega_r \quad (2)$$

where V_{eff} is the solid volume of the portion of wall involved in the mechanism, δ_m is the density of the masonry Ω_f , Ω_r are the masses of the horizontal structures involved in the mechanism. Effective mass and effective stiffness are used to calculate a natural period T_{eff} , which characterizes an equivalent single degree-of-freedom (SDoF) oscillator:

$$T_{\text{eff}} = 2\pi \sqrt{\frac{\Omega_{\text{eff}}}{K_{\text{eff}}}} \quad (3)$$

The mass is applied at the height of the center of gravity of the collapsing portion with respect to the ground, and a linear acceleration distribution over the wall height is assumed. The elastic limit acceleration S_{ay} is also computed depending on the failure mechanism identified; for failure

mechanisms involving flexural strain limit (assuming no tensile capacity in the material), for instance, S_{ay} will be the value of lateral acceleration that, combined with gravitational load resultant, will cause a triangular distribution of compression stresses at the base of the overturning portion, just before the onset of cracking:

$$\begin{aligned} S_{ay} &= \frac{t_b}{6h_0} g \text{ with corresponding displacement} \\ S_{de} &= \frac{S_{ay}}{4\pi^2} T_{\text{eff}}^2 \end{aligned} \quad (4)$$

where t_b is the effective thickness of the wall at the base of the overturning portion, h_0 is the height to the ground of the center of mass of the overturning portion, and T_{eff} is the natural period of the equivalent single degree-of-freedom (SDoF) oscillator. The maximum lateral capacity S_{au} is defined as

$$S_{au} = \frac{\lambda_c}{\alpha_1} g \quad (5)$$

where λ_c is the collapse load factor of the mechanism chosen, calculated by FaMIVE, and α_1 is the proportion of total mass participating to the mechanism. This is calculated as the ratio of the mass of the façade and sides or internal walls and floor involved in the mechanism Ω_{eff} to the total mass of the involved macroelements (walls, portion of floors, and portion of roof supported by the wall). The displacement corresponding to the damage threshold of structural damage, S_{ds} is

$$3S_{dy} \leq S_{ds} \leq 6S_{dy} \quad (6)$$

as suggested by (Tomazevic 2007). The range in Eq. 6 is useful to characterize masonry fabric of diverse levels of regularity and its integrity at ultimate conditions, with the lower bound better describing the behavior of adobe, rubblestone,

and brickwork in mud mortar, while the upper bound can be used for massive stone, brickwork set in lime or cement mortar, and concrete blockwork. The near collapse condition is determined by the displacement S_{du} identified by the condition of loss of vertical equilibrium which, for overturning mechanisms, can be computed as a lateral displacement at the top or for in-plane mechanism by the loss of overlap of two units in successive courses as quantifiable with one of the two alternative expressions in Eq. 7:

$$S_{du} \geq t_b/3 \quad \text{or} \quad S_{du} \geq l_u/2 \quad (7)$$

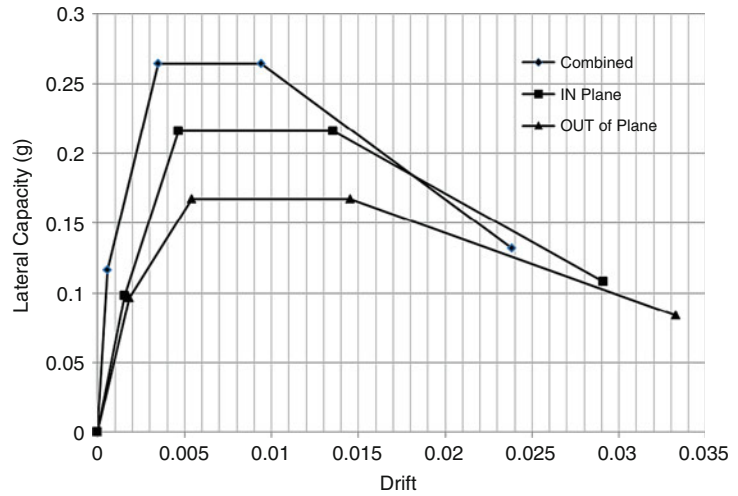
where t_b is the thickness at the base of the overturning portion and l_u is the typical length of units forming the wall. In the case of in-plane mechanism, the geometric parameter used for the elastic limit is, rather than the wall thickness, the width of the slender pier.

The threshold points identified by Eqs. 4, 5, 6, and 7 can be associated to corresponding states of damage. Specifically **DL**, *damage limitation*, corresponds to the elastic lateral capacity threshold (S_{de} , S_{ay}) defined by Eq. 4; **SD**, *significant damage*, corresponds to the peak capacity threshold (S_{ds} , S_{au}) defined by Eqs. 5 and 6; and **NC**, *near collapse*, corresponds to incipient or partial collapse threshold (S_{du} , S_{au}) defined by either conditions of Eq. 7.

The procedure's approach allows a direct analysis of the influence of different parameters on the resulting capacity curves, whether these are geometrical, mechanical, or structural. By way of example, Figs. 3 and 4 show a comparison of average capacity curves obtained by grouping the results using different criteria for the same sample of buildings in L'Aquila. In Fig. 3 the average curves represent the behavior according to the classes of mechanisms as per Fig. 1. In Fig. 4 the capacity curves are obtained by considering different structural typologies, as classified by the WHE-PAGER project (Jaiswal et al. 2011) and shown in Table 2. It can be seen

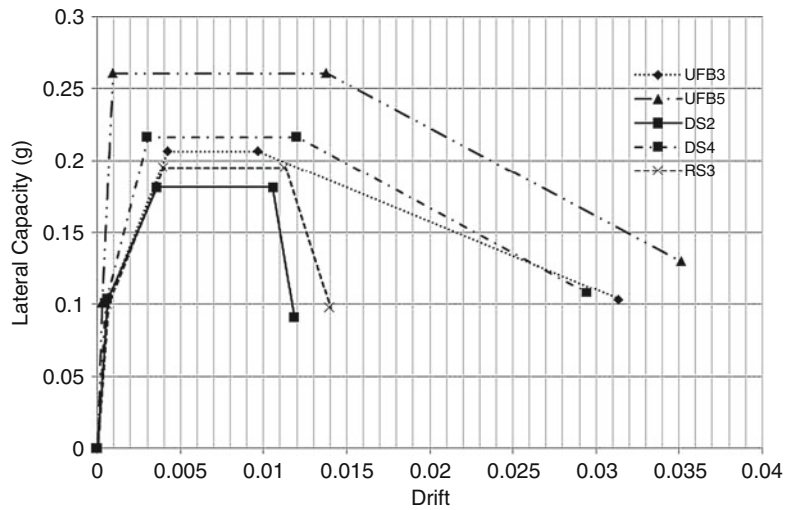
Seismic Vulnerability Assessment: Masonry Structures,

Fig. 3 Average capacity curves for sample grouped by collapse-mechanism classes



Seismic Vulnerability Assessment: Masonry Structures,

Fig. 4 Average capacity curves for sample grouped by structural typology



that the correlation between mode of failure and structural typology is qualitatively good but not univocal, and the grouping affects both ultimate lateral capacity and drift at all damage states.

Such results bring in sharper focus the limitation and inaccuracy of using idealized models and average curves without adequately considering the inherent aleatoric variation associated with any given site where the assessment is conducted and the importance of a detailed knowledge of the local construction characteristics when sampling the building representative of

the building stock. The lack of such knowledge determines a non-negligible epistemic error.

In Figs. 3 and 4 the capacity curves show a fourth point, which represents the collapse of the façade corresponding to a lateral displacement S_{dc} and a lateral acceleration S_{ac} as computed in Eq. 8:

$$S_{dc} \geq t_b/2 \text{ or } S_{dc} \geq l_u, \quad S_{ac} = \frac{S_{au}}{2} \quad (8)$$

Such relationship are estimated on the basis of experimental work available in literature.

Seismic Vulnerability Assessment: Masonry Structures, Table 2 Structural typologies classification according to PAGER (Jaiswal et al. 2011)

Load-bearing material	PAGER code	Description
Stone masonry	RS3	Local fieldstones with lime mortar
	RS4	Local field stones with cement mortar, vaulted brick roof, and floors
	DS2	Rectangular cut stone masonry block with lime mortar
	DS3	Rectangular cut stone masonry block with cement mortar
	DS4	Rectangular cut stone masonry block with reinforced concrete floors and roof
	MS	Massive stone masonry in lime or cement mortar
Brickwork or blockwork	UFB1	Unreinforced brick masonry in mud mortar without timber posts
	UFB3	Unreinforced brick masonry in lime mortar. Timber flooring
	UFB4	Unreinforced fired brick masonry, cement mortar. Timber flooring
	UFB5	Unreinforced fired brick masonry, cement mortar, but with reinforced concrete floor and roof slabs
	UCB	Unreinforced concrete block masonry with lime or cement mortar

Damage Thresholds and Performance Indicators

In the previous section capacity curves have been determined as polylines characterized by four notable points which represent as many damage thresholds. A method for assessing the overall behavior by use of a global performance indicator is the computation of the performance point. In order to calculate the performance point, it is necessary to intersect the capacity

curve derived above with the demand spectra for different return periods in relation to the performance criteria considered. Two broadly equivalent approaches for the derivation of the nonlinear demand spectra exist: the N2 method (Fajfar 1999) included in the EC8 and the capacity spectrum method (CSP) (FEMA 2001). The two methods differ essentially in the way the nonlinear demand spectrum is arrived at: the N2 method uses a reduction factor R , function of the structure expected ductility μ , while the CSP uses a fictitious damping factor derived from the hysteresis loop of the structure. In the following the N2 method will be used to illustrate the derivation of performance points.

The original version of the N2 method was derived for elastic-perfectly plastic SDoF oscillator, and the relationship between the reduction factor R and the ductility μ , used to determine the nonlinear response spectrum, is dependent on the bilinear shape of the capacity curve. Alternative relationships between the reduction factor, the ductility, and the natural period of the system, R - μ - T , have been developed to determine the inelastic displacement ratios and the inelastic response spectra for structures characterized by quadri-linear backbone curves, characterized by a plateau of reduced capacity and further ductility, after a negative slope branch (see Miranda 2001; Dolsek and Fajfar 2004; Vamvatsikos and Cornell 2005).

In the case of the capacity curves derived with the FaMIVE procedure however, as mentioned above, the behavior beyond the near collapse performance point is fraught with uncertainty and hence the descending branch of the curve is ignored, and the simplified relationship for bilinear curves is used.

To calculate the coordinates of the performance point in the displacement-acceleration space, the intersection of the capacity curve with the nonlinear demand spectrum for an appropriate level of ductility μ can be determined as shown in Eq. 8, given the value of ultimate lateral capacity of the equivalent SDoF oscillator, S_{au} :

$$\begin{aligned}
& \text{if } T^* < T_c \\
& \text{if } S_{au} \geq S_a(T^*) \Rightarrow S_d^*(\mu) = \frac{T_c^2(S_{ae}(T^*) - S_a(T^*))^2}{(\mu - 1)^2} * \frac{g\mu}{4\pi^2 S_a(T^*)} \\
& \text{if } S_a(T_c) < S_{au} < S_a(T^*) \Rightarrow S_d^*(\mu) = \frac{T_c^2(S_{ae}(T^*) - S_{au})^2}{(\mu - 1)^2} * \frac{g\mu}{4\pi^2 S_{au}} \\
& \text{if } S_{au} \leq S_a(T_c) \Rightarrow S_d^*(\mu) = \frac{gT_c^2(S_{ae}(T^*))^2}{4\pi^2 \mu S_{au}} \\
& \text{if } T^* \geq T_c \\
& \text{if } S_{au} \geq S_a(T^*) \Rightarrow S_d^*(\mu) = \frac{gT_c^2(S_{ae}(T^*))^2}{4\pi^2 \mu S_a(T^*)} \\
& \text{if } S_{au} < S_a(T^*) \Rightarrow S_d^*(\mu) = \frac{gT_c^2(S_{ae}(T^*))^2}{4\pi^2 \mu S_{au}}
\end{aligned} \tag{9}$$

where two different formulations are provided for values of ultimate lateral capacity S_{au} greater or smaller than the nonlinear spectral acceleration $S_a(T_c)$ associated with the corner period T_c marking the transition from constant acceleration to constant velocity section of the parent elastic spectrum. Figure 5 shows an example of calculation of performance points for median capacity curves for some different types of mechanisms (see Fig. 1).

Derivation of Fragility Functions

Once capacity curves and limit state points are identified on them, fragility curves for different limit states are obtained by computing the median and standard deviation values of the performance point displacements for each index building in a given set of buildings surveyed on-site or created through randomization of the input parameters. The median and standard deviation are used to derive equivalent lognormal distributions. To this end the median displacement for each limit state can be calculated as:

$$\hat{S}_{ds_i} = e^{\bar{\mu}_i} \text{ with } \bar{\mu}_i = \frac{1}{n} \sum_{j=1}^n \ln S_{ds_i}^j \tag{10}$$

and the corresponding standard deviation as:

$$\begin{aligned}
\beta_{ds_i} &= e^{\bar{\mu}_i + \frac{1}{2}\sigma_i^2} \sqrt{e^{\sigma_i^2} - 1} \text{ with } \sigma_i \\
&= \sqrt{\frac{\sum_{j=1}^n \left(\ln S_{ds_i}^j - \ln \bar{S}_{ds_i}^j \right)^2}{n}} \tag{11}
\end{aligned}$$

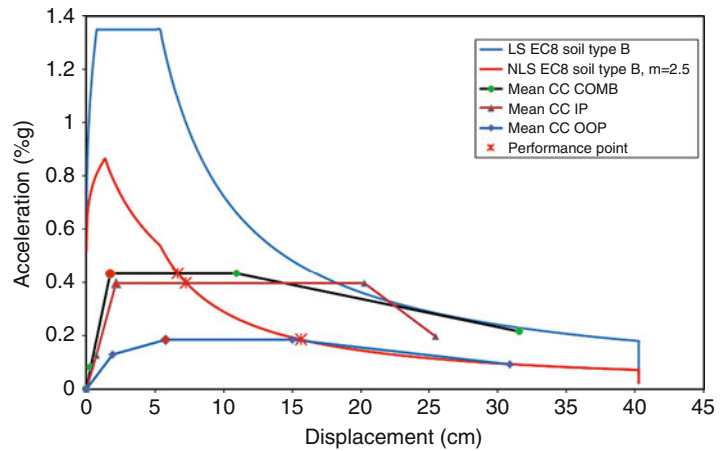
Figures 6 and 7 show the set of fragility curves obtained for the damage limit states of damage limitation and significant damage as computed for the two Italian sites of Nocera Umbra and Serravalle considering separately the three types of structural behavior. Once a structural typology has been assigned, the values of the mechanical characteristics are the same across the two samples, while the structural details are accounted for directly in the three classes of mechanisms. Hence the variability observed between samples in each chart can be related directly to geometric differences and masonry fabric, i.e., to the local aspects of the construction practice and architectural layout. Curves on the left of the diagrams relate to corresponding capacity curves with a stiffer elastic branch in the case of damage limitation and with lower ductility in the case of significant damage. However the distribution does not bare consistency across the three classes of mechanism for the two sites.

Conducting Seismic Vulnerability Assessment

As seen in the previous section, an analytical method based on limit state analysis and mechanism approach has the advantage of employing specific structural models to assess the seismic behavior of masonry buildings of different typologies, represented through capacity curves and performance points, while at the same time

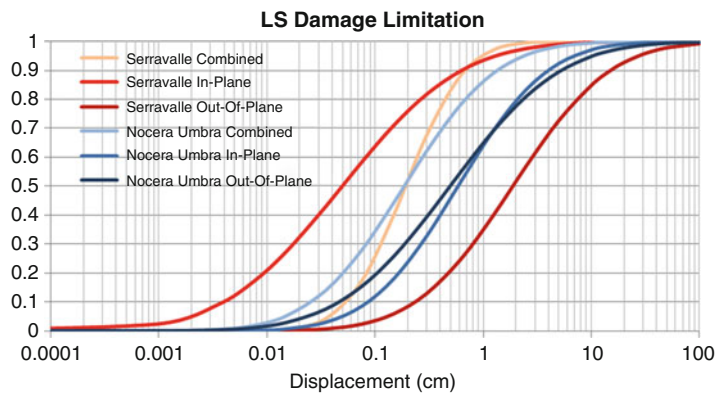
Seismic Vulnerability Assessment: Masonry Structures,

Fig. 5 Example of calculation of performance points for median capacity curves for types of mechanisms triggered (*COMB* combined mechanism, *IP* in-plane mechanism, *OOP* out-of-plane mechanism; see also Fig. 1)



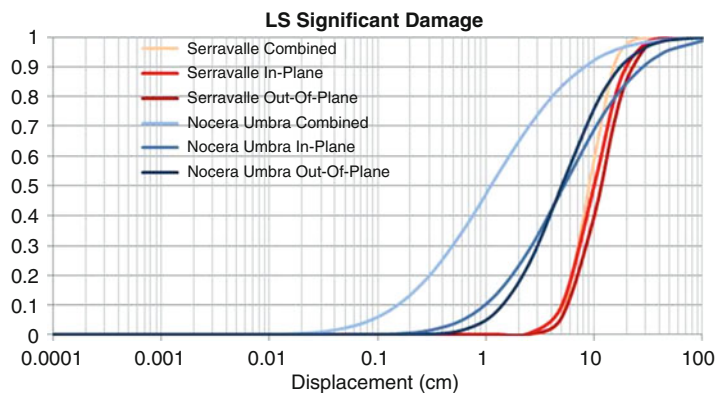
Seismic Vulnerability Assessment: Masonry Structures,

Fig. 6 Fragility curves for limit state of damage limitation for the three classes of collapse mechanisms for two different Italian sites



Seismic Vulnerability Assessment: Masonry Structures,

Fig. 7 Fragility curves for limit state of significant damage for the three classes of collapse mechanisms for two different Italian sites



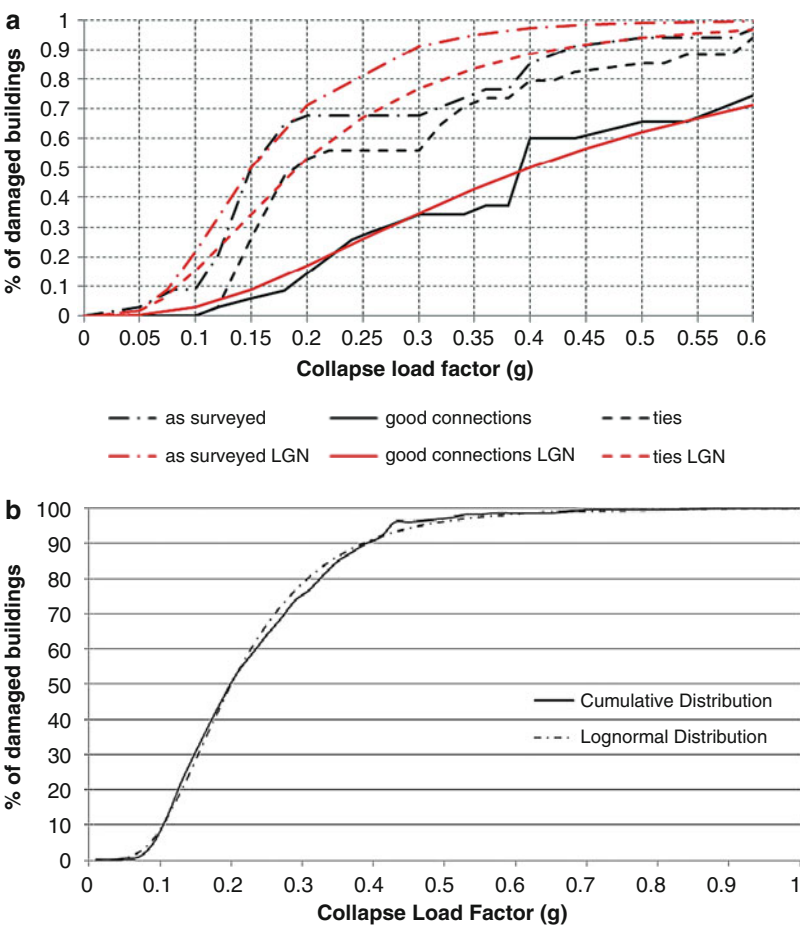
being applicable to relatively large numbers of buildings. Hence procedures of this type, if properly developed, are particularly suitable for the seismic assessment of building aggregates in historical centers (Novelli and D'Ayala 2014) or

when a group of heritage structures at territorial scale are considered (D'Ayala and Ansal 2012).

A robust approach for seismic vulnerability assessment at territorial scale needs to address a number of issues, mainly related to a correct

Seismic Vulnerability Assessment: Masonry Structures,

Fig. 8 Fragility curves for two different samples – (a) 30 buildings and (b) 200 buildings – represented as actual distribution and lognormal distribution (Eqs. 10 and 11)



representation of the existing exposure so as to obtain meaningful fragility and vulnerability functions, while explicitly accounting for the uncertainties related to modelling limitation, the inherent randomness of the sample, and the randomness of the response. Given a territorial distribution, uncertainties are also present about the seismic action, relating to record to record, attenuation, soil characteristics, etc. Hence specific attention and due consideration should be given to the following aspects of a process leading to a seismic vulnerability assessment at territorial scale and identification of rehabilitation strategies.

Such process can be articulated in 10 steps:

- Consideration on the urban fabric and study area

- Identification of typologies and construction details
- Size and heterogeneity of sample to be surveyed
- Data collection
- Data analysis
- Safety and conservation requirements
- Seismic hazard
- Definition of performance points
- Definition of fragility functions
- Rehabilitation decisions

A detailed discussion of each step is beyond the scope of this document. However some observations in relation to specific applications of FaMIVE can help in highlighting some of the issues.

In relation to the sample size and heterogeneity, for instance, Fig. 8a, b shows the resulting

distributions for two different cases. The first sample included 30 buildings, which were considered relatively homogenous in terms of structural typology and materials but differing in geometry and quality of connections. The second one is made up of more than 200 buildings of varied characteristics. As it can be seen, the assumption of a lognormal model to describe the fragility cumulative distribution is more reliable as the size and heterogeneity of the sample increases.

Figure 8a also shows that results are highly affected by data collection and data analysis: the three cumulative distributions are obtained for the same sample of buildings when the presence of anchors or good wall connection is in turn ignored or considered. Because the collection and interpretation of this type of data is very sensitive to the expertise of the surveyor and the analyst, results can be highly affected by either incorrect interpretation of this information on the basis of the survey or use of a modelling tool that cannot properly simulate the behavior of the buildings when these constraints are in place.

The requirement of the codes and the limitations imposed by conservation principles also affect the way in which the fragility functions can be derived and used.

The reliability of the results obtained in the previous section can be considered within the framework set out in the EC 8 (CEN 2005), whereby the reliability associated to the results of a seismic assessment of a structure is expressed as a function of the level of knowledge and quantified by means of the confidence factor. Hence this can be considered a measure of the epistemic uncertainty. EC 8 (CEN 2005) recognizes three levels of knowledge (limited, normal, and full) and three fields of knowledge (geometry, construction details, and materials). As data used in the FaMIVE approach are collected by on-site visual inspection with some measurement and in situ accurate observation of construction details, while only very limited in situ nondestructive test on materials are performed and material characteristics are otherwise assigned based on literature or surveyor experience, then the level of knowledge is

superior to KL1, *limited*, but not quite equal to KL2, *normal*. For this level of knowledge, a static nonlinear analysis, such as the limit state mechanism approach, leading to a capacity curve is deemed appropriate. Hence according to the recommended values, the confidence factor CF should be in the range 1.2 to 1.35 depending on how closer the actual knowledge can be considered to the reference level identified by KL2. The confidence factor is then used in EC8 to reduce the capacity values as obtained from the assessment.

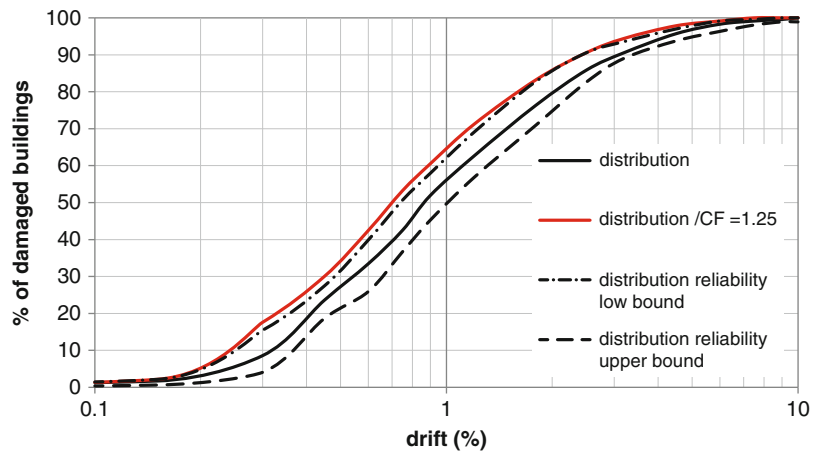
The FaMIVE procedure uses a measure of uncertainty of the input data to determine the uncertainty of the output. Depending on whether data, in each section of the data collection form, has been collected and measured directly on-site, or collected on-site and confirmed by existing drawings or photograph, or collected from photographic evidence only, three levels of uncertainty are considered, as high, medium, and low, respectively, to which three confidence ranges of the value given for a parameter can be considered corresponding to 10 % variation, 20 % variation, and 30 % variation. The parameter value attributed during the survey is considered central to the confidence range so that the interval of existence of each parameter is defined as $\mu \pm 5\%$, $\mu \pm 10\%$, and $\mu \pm 15\%$, depending on highest or lowest uncertainty. The uncertainty applied to the output parameters, specifically lateral acceleration and limit states' displacement, is calculated as a weighted average of the uncertainty of each section of the data form, with minimum 5 % confidence range to maximum 15 % confidence range.

In Fig. 9 the cumulative distribution for drift at the structural damage state threshold is compared with the curve obtained assuming a confidence factor of 1.25 in agreement with EC8 knowledge level and with the confidence boundaries computed by FaMIVE according to the procedure outlined above.

The major limitation of using confidence factors in the way suggested by EC8 is the possible underestimation of the overall capacity of the sample and hence an appropriate representation of epistemic uncertainty.

Seismic Vulnerability Assessment: Masonry Structures,

Fig. 9 Cumulative distribution for drift at structural damage threshold: comparison with curve obtained for an EC8 confidence factor of 1.25 and with FaMIVE confidence boundary



FaMIVE can also be used to foresee the shift in seismic performance of the building stock of a historic center if certain interventions are implemented to improve the seismic behavior. Among possible interventions, the improvement of cohesion of the masonry through grouting of multi-leaf walls with lime-based grout and the introduction of ties connecting walls together are seen as meeting the conservation principles of like for like, minimum intervention, and, in the case of ties, reversibility. Figure 10a, b shows the change in the median capacity curves when these interventions are considered: it can be seen that the response is different depending on the typology; in general both interventions cause an increase in lateral capacity. In some cases also a change in ductility and/or in the behavior up to structural damage as the intervention causes a change in the prevalent failure mechanisms.

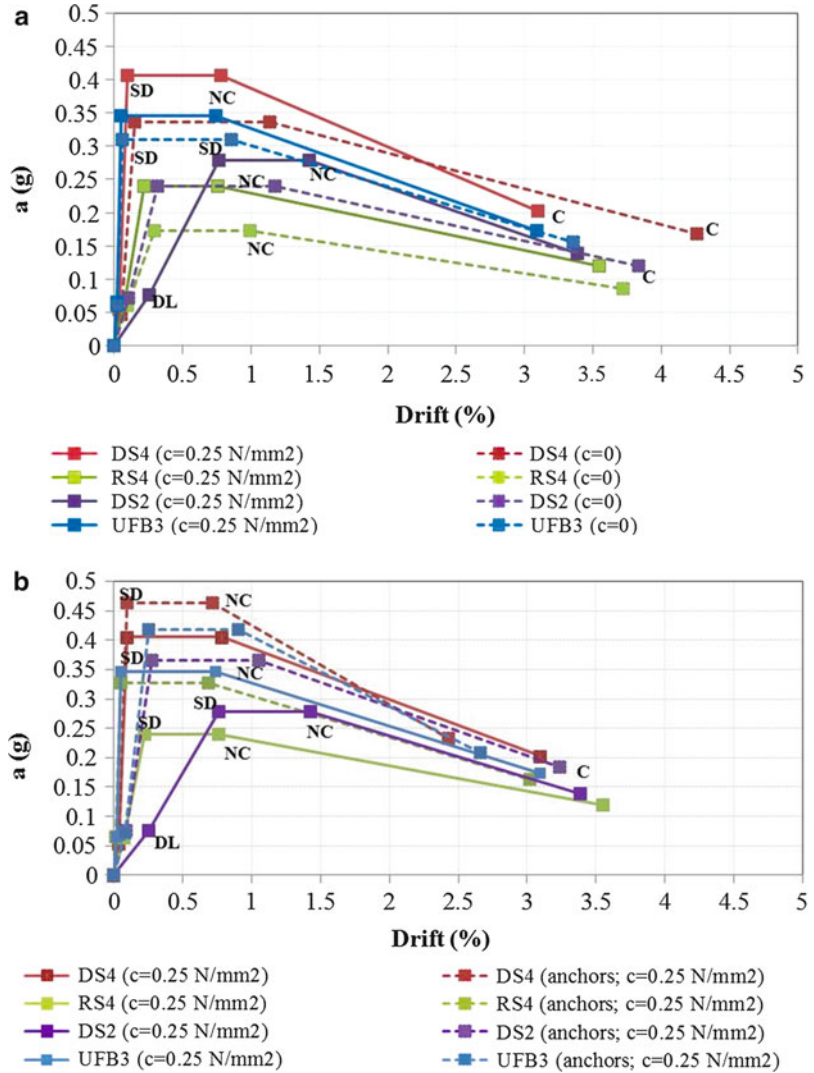
Summary

Traditionally seismic vulnerability of historic buildings has been evaluated using empirical methods based either on damage observation or construction data. In the last decade, analytical methods for assessment of masonry structures have increased in number and quality. However, most of them assume only in-plane failure behavior reducing the walls to frame-like

behavior. This approach has serious limitations and it is nonconservative, as it overlooks out-of-plane behavior which might occur for structures with insufficient out-of-plane wall capacity. Nevertheless, recommendations of the EC 8 (CEN 2005) only consider basic in-plane behavior, either controlled by flexure or controlled by shear, and set limit drifts for this performance criterion. The application of the FaMIVE procedure to a number of cases in regions of different seismicities shows the relevance of considering separately different failure mechanisms to identify realistic risks of failure. A procedure for the computation of the performance point given performance criteria and hazard return period is introduced to show the variability of performance points for different failure behaviors and how they affect probability of failure computed as cumulative. The random uncertainty associated with this type of analysis is presented in terms of fragility curves that can be modelled as lognormal distribution. In the same way, the epistemic uncertainty associated with the modelling procedure is also evaluated using the same approach. Results show that the reliability approach set out in FaMIVE correctly accounts for epistemic uncertainty and it is of the same order of magnitude as the confidence factor suggested by EC8, although this only accounts for negative effects.

Seismic Vulnerability Assessment: Masonry Structures,

Fig. 10 Comparison between the capacity curves for (*DS2*), (*DS4*), (*UFB3*), and (*RS4*) for (a) the condition of cohesion, $c = 0$ and $c = 0.25 \text{ N/mm}^2$, and (b) with the addition of ties to the condition of cohesion $c = 0.25 \text{ N/mm}^2$



Cross-References

- Analytic Fragility and Limit States [P(EDPI IM)]: Nonlinear Static Procedures
- Empirical Fragility
- Masonry Structures: Overview
- Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers
- Seismic Analysis of Masonry Buildings: Numerical Modeling
- Seismic Loss Assessment

- Seismic Vulnerability Assessment: Reinforced Concrete Structures
- Strengthening Techniques: Masonry and Heritage Structures

References

- Askan A, Yucemen MS (2010) Probabilistic methods for the estimation of potential seismic damage: application to reinforced concrete buildings in Turkey. *Struct Saf* 32(4):262–271

- Baggio C, Bernardini A, Colozza R, Corazza L, Della Bella M, Di Pasquale G, Dolce M, Goretti A, Martinelli A (2009) Manuale per la compilazione della scheda di 1° livello di rilevamento danno, pronto intervento e agibilità per edifici ordinari nell'emergenza post-sismica (AeDES). Editrice Italiani nel Mondo srl, Roma
- Barbat AH, Yépez Moya F, Canas JA (1996) Damage scenarios simulation for seismic risk assessment in urban zones. *Earthq Spectra* 12(3):371–394
- Benedetti D, Benzoni G, Parisi MA (1988) Seismic vulnerability and risk evaluation for old urban nuclei. *Earthq Eng Struct Dyn* 16:183–201
- Bernardini A, Gori R, Modena C (1990) An application of coupled analytical models and experiential knowledge for seismic vulnerability analyses of masonry buildings. In: Koridze A (ed) *Engineering aspects of earthquake phenomena*, vol 3. Omega Scientific, Oxon, pp 161–180
- Borzi B, Crowley H, Pinho R (2008) Simplified pushover-based earthquake loss assessment (SP-BELA) method for masonry buildings. *Int J Architect Herit* 2(4):353–376
- Braga F, Dolce M, Liberatore D (1982) A statistical study on damaged buildings and ensuing review of the MSK – 76 SCALE. In: *Proceedings of the 7th European conference on earthquake engineering*, Athens
- CEN – EN 1998-3 (2005) Eurocode 8: design of structures for earthquake resistance, Part 3: strengthening and repair of buildings. Comité' Européen de Normalisation, Brussels
- Corsanego A, Petrini V (1990) Seismic vulnerability of buildings – work in progress. In: *Proceedings of the workshop II on seismic risk vulnerability and risk assessment*, Trieste, pp 577–598
- D'Ayala D (2005) Force and displacement based vulnerability assessment for traditional buildings. *Bull Earthq Eng* 3:235–265, Springer
- D'Ayala D, Ansal A (2012) Nonlinear push over assessment of heritage buildings in Istanbul to define seismic risk. *Bull Earthq Eng* 10(1):285–306
- D'Ayala D, Paganoni S (2011) Assessment and analysis of damage in L'Aquila historic city centre after 6th April 2009. *Bull Earthq Eng* 9(1):81–104
- D'Ayala D, Speranza E (2003) Definition of collapse mechanisms and seismic vulnerability of historic masonry buildings. *Earthq Spectra* 19:479–509
- Di Pasquale G, Orsini G, Romeo RW (2005) New developments in seismic risk assessment in Italy. *Bull Earthq Eng* 3(1):101–128
- Dolsek M, Fajfar P (2004) Inelastic spectra for infilled reinforced concrete frames. *Earthq Eng Struct Dyn* 33:1395–1416
- Elenas A (2000) Correlation between seismic acceleration parameters and overall structural damage indices of buildings. *Soil Dyn Earthq Eng* 20(1–4):93–100
- Erberik MA (2008) Generation of fragility curves for Turkish masonry buildings considering in-plane failure modes. *Earthq Eng Struct Dyn* 37:387–405
- Fajfar P (1999) Capacity spectrum method based on inelastic demand spectra. *Earthq Eng Struct Dyn* 28(9):979–993
- Fajfar P, Gašperšič P (1996) The N2 method for the seismic damage analysis of R.C. buildings. *Earthq Eng Struct Dyn* 25(1):31–46
- Federal Emergency Management Agency (FEMA) (2001) HAZUS99 technical manual. Service Release 2. FEMA, Washington, DC
- Giovinazzi S, Lagomarsino S (2004) A macroseismic method for the vulnerability assessment of buildings. In: *Proceedings of the 13th world conference on earthquake engineering*, Vancouver, Paper No. 896
- Grünthal G (ed) (1998) *Cahiers du Centre Européen de Géodynamique et de Séismologie: volume 15 – European macroseismic scale 1998*. European Center for Geodynamics and Seismology, Luxembourg
- Jaiswal K, Wald D, D'Ayala D (2011) Developing empirical collapse fragility functions for global building types. *Earthq Spectra* 27:775
- Kappos AJ, Stylianidis KC, Pitilakis K (1998) Development of seismic risk scenarios based on a hybrid method of vulnerability assessment. *Nat Hazards* 17(2):177–192
- Kappos AJ, Panagopoulos G, Penelis GG (2008) Development of a seismic damage and loss scenario for contemporary and historical buildings in Thessaloniki, Greece. *Soil Dyn Earthq Eng* 28(10–11):836–850
- Lagomarsino S, Giovinazzi S (2006) Macroseismic and mechanical models for the vulnerability and damage assessment of current buildings. *Bull Earthq Eng* 4(4):415–443
- Lang K, Bachmann H (2004) On the seismic vulnerability of existing buildings: a case study of the city of Basel. *Earthq Spectra* 20(1):43–66
- Miranda E (2001) Estimation of inelastic deformation demands of SDOF systems. *J Struct Eng (ASCE)* 127(9):1005–1012
- Novelli V, D'Ayala DF (2014) The PERPETUATE procedure for the seismic vulnerability assessment at territorial scale. Application to the Casbah of Algiers. *Bull Earthq Eng* 12:1–24
- Spence R, Coburn AW, Pomonis A (1992) Correlation of ground motion with building damage: the definition of a new damage-based seismic intensity scale. In: *Proceedings of the tenth world conference on earthquake engineering*, Madrid, vol 1, pp 551–556
- Tomažević M (2007) Damage as a measure for earthquake-resistant design of masonry structures: Slovenian experience. *Can J Civil Eng* 34(11):1403–1412
- UNDP/UNIDO Project RER/79/015 (1985) Post-earthquake damage evaluation and strength assessment of buildings under seismic condition, vol 4. UNDP, Vienna
- Vamvatsikos D, Cornell C (2005) Direct estimation of seismic demand and capacity of multidegree-of-freedom systems through incremental dynamic

- analysis of single degree of freedom approximation. *J Struct Eng* 131(4):589–599
- Whitman RV, Reed JW, Hong ST (1973) Earthquake damage probability matrices. In: *Proceedings of the 5th world conference on earthquake engineering*, Rome, Italy, pp. 2531–2540
- Wu Y-M, Teng T-I et al (2003) Relationship between peak ground acceleration, peak ground velocity, and intensity in Taiwan. *Bull Seismol Soc Am* 93(1):386–396
- Zobin VM, Cruz-Bravo AA, Ventura-Ramirez F (2010) Microzonation of seismic risk in a low-rise Latin American city, based on the macroseismic evaluation of the vulnerability of residential buildings: Colima city, Mexico. *Nat Hazard Earthq Syst Sci* 10:1347–1358

Seismic Vulnerability Assessment: Reinforced Concrete Structures

Flavia De Luca¹ and Gerardo M. Verderame²

¹Department of Civil Engineering, University of Bristol, Bristol, UK

²Department of Structures for Engineering and Architecture (DiSt), University of Naples Federico II, Naples, Italy

Synonyms

Axial load failure; Damage measure; Flexure; Nonlinear modeling; PBEE; RC structures; Shear

Introduction

Seismic vulnerability can be defined as the degree of loss to a given element at risk (e.g., buildings) resulting from the occurrence of an earthquake event (Coburn and Spence 2002). The development of methodologies for seismic vulnerability assessment is an essential tool for seismic risk management and for prioritizing pre-earthquake strengthening of the built environment.

Seismic risk can be approached through different methodologies. Most of them aim to deconstruct the problem into the typical elements defining any kind of risk: (i) hazard,

(ii) vulnerability, and (iii) exposure, and they are based on the total probability theorem.

Available seismic vulnerability methods, i.e., empirical, analytical, or hybrid (e.g., Calvi et al. 2006), developed in the last 30 years, differ because of the nature of tools and data employed. In the following only analytical approach to seismic vulnerability is considered, given the significance that such approach had in the last years. Analytical vulnerability methods stand at the basis of current worldwide codes and guidelines for seismic design and assessment of structures.

Therefore, next-generation codes (e.g., FEMA P-58-1 2012) are proceeding toward an explicit quantification of seismic risk through the Performance-Based Earthquake Engineering (PBEE) framework. Section “[Performance Earthquake Engineering Framework](#)” of this entry describes how in PBEE seismic risk assessment is deconstructed through the total probability theorem. PBEE’s deconstruction of seismic risk in hazard, structural, damage, and loss analysis emphasizes how structural and damage analyses are strictly related to vulnerability assessment in strict sense.

Notwithstanding the fact that PBEE is a generalized methodology that can be applied for any kind of structure, the case of reinforced concrete (RC) structures is specifically addressed in the following. As far as the relationship between intensity measure and engineering demand parameter arises in PBEE, it is the phase in which it is necessary to focus on the nonlinear behavior of RC structural elements (section “[Behavior of RC Elements and Non Structural Elements](#)”), in order to approach analytical modeling (section “[Analytical Modeling of RC Structures](#)”), and qualify and quantify components damage measures (section “[Damage Measures](#)”). A review of experimental and analytical results collected in the last decades is provided herein. It aims at the quantification of the behavior of RC structures and, in turn, of RC elements. Most significant results on vulnerability assessment of RC structures are collected together, and they allow emphasizing future challenges and needs (section “[Future Challenges and](#)

Concluding Remarks) that scientific community and practitioners should take on in the next years.

The focus in the following is the quantification of nonlinear parameters of RC elements aimed at a reliable modeling of their behavior. Emphasis is also given to consolidated and more recent literature results that have been included in codes and guidelines.

Performance Earthquake Engineering Framework

“PBEE implies design, evaluation, construction, monitoring the function and the maintenance of engineered facilities whose performance under common and extreme loads respond to the diverse needs and objectives of owners-users and society. It is based on the premise that performance can be predicted and evaluated with quantifiable confidence to make, together with the client, intelligent and informed trade-offs based on life cycle considerations rather than construction costs alone” (Krawinkler and Miranda 2004). Guidelines and codes, since the 1990s, have partially implemented, in their general framework, the basic concepts of PBEE in various forms, resulting in the attempt to tie design criteria to a performance level, usually that of collapse prevention. The Pacific Earthquake Engineering Research (PEER) Center has focused for several years on the development of procedures, knowledge, and tools for a comprehensive seismic performance assessment of buildings and bridges. The efforts made have resulted in a general framework that is now shared by the earthquake engineering community all over the world. Different publications describe the basic steps of PBEE methodology (e.g., Cornell and Krawinkler 2000; Krawinkler 2002; Moehle 2003; Porter 2003; Deierlein et al. 2003; Krawinkler and Miranda 2004; Moehle and Deierlein 2004), and a comprehensive document for a punctual methodological description of last enhancements of PBEE is available in FEMA P-58-1 document (2012).

Sharing a common vision and approach allows placing further enhancements and progresses in

this PBEE consolidated framework, thus aiming at the final challenge to contribute effectively to the reduction of losses and the improvement of safety (Cornell and Krawinkler 2000).

The basis of performance assessment can be summarized in a single equation (see Eq. 1). This equation suggests a generic structure for coordinating, combining, and assessing the many considerations implicit in performance-based seismic assessment.

$$\lambda(DV) = \iiint |G(DV|DM)dG(DM|EDP)dG(EDP|IM)d\lambda(IM)| \quad (1)$$

Based on total probability theorem, Eq. 1 allows deconstructing the problem in four steps: (i) hazard analysis, (ii) structural analysis, (iii) damage analysis, and (iv) loss analysis. Each step carries out a specific generalized variable: *intensity measure (IM)*, *engineering demand parameter (EDP)*, *damage measure (DM)*, and *decision variable (DV)*. The key issue of PBEE methodology is to identify and quantify *DV* of primary interest to the decision makers with consideration to all important uncertainties. *DVs* have been defined in terms of different quantities, such as repair costs, downtime, and casualty rates.

First member of Eq. 1 is a probabilistic description of the *DV*, such as the mean annual frequency of the *DV* exceeding a specified value. $\lambda(DV)$ might be the mean annual frequency of the direct dollar loss (repair cost) exceeding 50 % of the replacement cost of the facility. The terms $dG(EDP|IM)$, $dG(DM|EDP)$, and $G(DV|DM)$ or their derivatives on the right side of Eq. 1 are conditional probabilities relating one quantity to another, while $d\lambda(IM)$ is the derivative of the hazard curve, relating ground motion intensity measure to its mean annual frequency of exceedance.

Aimed at facilitating probability calculation, but mostly aimed at compartmentalizing discipline-specific knowledge, the above PBEE framework should choose intermediate variables (*EDP* and *DM*), so that conditional probabilities

are independent of one another (Deierlein et al. 2003). A brief intro to each step of PBEE is provided herein. It is a general methodological framework description that does not refer yet specifically to RC structures.

Hazard Analysis

Hazard analysis allows describing earthquake hazard in a probabilistic manner. The result of this phase is a hazard curve, which shows the variation of the selected *IM* versus mean annual frequency of exceedance (*MAF*). Probabilistic seismic hazard analysis, for a specific site, passes through four main steps: (i) characterization of the seismic sources (e.g., identified faults or geographical areas), in which it is generally assumed that the occurrence of an event at one source does not affect the occurrence of events at other sources; (ii) characterization of magnitude distribution, generally obtained through a truncated exponential distribution or through characteristic magnitude distribution estimated for specific faults; (iii) ground motion estimation evaluated through prediction equation of the considered *IM*; and (iv) probability analysis for the evaluation of the hazard curve (see also McGuire 2004). Hazard analysis can be performed with different intensity measures, provided the appropriate prediction equation. Notwithstanding the fact that advanced intensity measures have been considered in literature (e.g., Tothong and Luco 2007) and prediction equations for these *IMs* are available (e.g., Tothong and Cornell 2006; De Luca et al. 2014a), still the employment of peak ground acceleration (*PGA*) and, more commonly, of spectral acceleration at the fundamental period of the structure, *Sa(T)*, represents the basic practice.

Hazard analysis is generally performed for the evaluation of a target spectrum finally aimed at ground motion selection, if the assessment is made through nonlinear dynamic analysis. It is worth to note that according to FEMA P-58-1 (2012), hazard analysis for vulnerability assessment and loss assessment can be performed with three different approaches according to the final aim of the study: intensity-based, scenario-based, and time-based.

Notwithstanding the fact that hazard analysis is a common step for each structural typology to which PBEE methodology is applied, it is worth noting that it is also the step in which most of the variability of the whole problem is considered. The so-called record-to-record variability is the source of the most significant variability in risk evaluation (e.g., Goulet et al. 2007). On the other hand, such source of variability is also a common aspect for each structural typology, and it is not an aspect that strictly characterizes RC more than other structural typologies (e.g., steel, masonry, etc.).

Structural Analysis

Structural analysis is the phase of PBEE in which structural simulations are performed for the evaluation of *EDPs* at given *IMs*. At this step, a structural model is built up for the estimation of *EDPs*, such as internal member forces and local or global deformations, including structural collapse. The computational model is built up taking into account uncertainties in model parameters. A proper nonlinear model of a structure is the result of a reliable analytical model of the elements of the structure considered and of the analysis of experimental behavior of the elements.

The choice of *EDPs* depends on the performance target and on the type of system of interest. The general approaches for the evaluation of $dG(EDP/IM)$ is to perform incremental dynamic analyses or other kinds of nonlinear analyses; see also FEMA P-58-1 (2012).

Damage Analysis

Damage analysis is the third step of PBEE methodology. It relates the *EDPs* to *DMs*. The *DMs* include quantitative descriptions of damage to structural elements, nonstructural elements, and contents. This quantification must be in sufficient detail to enable subsequent quantification of the necessary repairs, disruption of function, and safety hazards (Moehle and Deierlein 2004). The *EDPs* considered in structural analysis are the input to a set of fragility functions that model the probability of various level of physical damage, conditioned on structural response. Physical damage is described at a detailed level, defined relative to particular repair efforts required to

restore the component to its undamaged state. The current approach is a component-based approach (e.g., Aslani 2005; FEMA P-58-1 2012). The study of experimental behavior of components is still a key issue for reliable evaluation of damage measures.

Loss Analysis

The last step of a PBEE loss assessment is the probabilistic estimation of performances through *DVs*. *DVs* represent the outcome parameters of PBEE and allow a transformation of engineering evaluations in terms of variables of interest for stakeholders (e.g., dollars, deaths, downtime); see Porter (2003). Different studies are available in literature for the evaluation of direct and indirect losses (e.g., Mitrani-Reiser 2007), and results of such studies are now the basis of the loss assessment framework provided by FEMA P-58-1 (2012). It is worth noting that loss analysis works on repair costs and downtime that, in turn, are strictly related to the social and economical environment in which they are evaluated.

Behavior of RC Elements and Nonstructural Elements

The basis of a reliable vulnerability analysis for any structural material passes through a solid modeling approach. PBEE assessment framework naturally points to nonlinear dynamic analysis as the reference structural analysis tool for RC structures. Thus, it is necessary to consider reliable nonlinear hysteretic models of all components of RC structures. Structural modeling of RC elements is based on the observation and interpretation of experimental tests and post-earthquake evidence from which it is possible to carry out nonlinear models of element behavior. It is important to state that a significant difference has to be made between elements with and without seismic detailing. In fact, existing elements are susceptible of various modes of failure, while capacity design in seismic detailed elements prevents brittle modes of failure.

Seismic vulnerability assessment has to account for all typical modes of failure that RC

elements can show; so, in the following, behavior of RC elements is described considering the case of existing elements that represent the most general situation in vulnerability assessment problems. This section is primarily focused on the vulnerability of RC frame structures; thus the behavior of columns and beam-column joints is specifically addressed. At the end of this section, behavior of masonry infills is also considered. Masonry infills are considered because of the significant effect they can have on vulnerability of RC frame structures. In fact, they provide on one hand the strength and stiffness increase and on the other hand, the local interaction between infill and RC frame.

RC Members

The specification of nonlinear structural component models in the form of monotonic backbone curves and hysteretic rules for RC structures is one of the key targets for the earthquake engineering community worldwide. Monotonic backbone curves and hysteretic rules should be able to capture the experimental behavior of RC elements. Any RC element component model is defined as function of specific parameters, such as shear span ratio, materials' strengths, longitudinal and transversal reinforcement ratios, and axial load ratio (e.g., CEB 1996; Elwood 2004; Elwood and Moehle 2005a, b; Zhu et al. 2007; Haselton et al. 2008; Biskinis and Fardis 2010a, b). Specific combinations of the above ruling parameters end up to the different definitions of wall, beam, or column. Furthermore, the structural difference between walls, beams, and columns is also made by the structural typology in which the above prismatic RC members are included, and by the structural role they play with respect to loads they are subjected to (e.g., gravity loads, earthquake loads, etc.).

In particular a prismatic RC member that carries gravity loads through shear and bending and that is designed under the assumption of uniaxial load with zero axial force is referred as a beam in RC frame structures. Again, a prismatic RC member in which the ratio between the two nondominant dimensions (i.e., width and depth of the cross section) is higher than four is typically defined as wall. This latter geometrical

assumption implies that a “wall” resists lateral forces mainly in one direction, and it can be designed for such unidirectional resistance by assigning flexural resistance to the two far ends of the section and shear resistance to the web (Fardis 2009). The classification of an RC prismatic elements as “column” is the more general possible; it covers the widest range of parameters ruling RC component modeling (e.g., variable axial load ratio, biaxial load, variable reinforcement ratios). Therefore, in the following, specific attention is given to modeling issues related to RC columns rather than beams or walls.

Notwithstanding the fact that columns are characterized by biaxial loading and by the widest variety of ruling parameters, experimental data more commonly available address the case of typical columns in RC frame structures. More specifically, most test specimens have rectangular cross section, symmetric reinforcement, and nonzero axial load, and rarely they are characterized by biaxial loading. The latter is the reason why most reliable nonlinear structural component models are conceived for columns under uniaxial loading. It is worth noting that recent experimental and modeling efforts have been made for the characterization of biaxial component models (e.g., Bousias 1993; Bousias et al. 2002; Di Ludovico et al. 2013), and some guidelines now provide recommendations for their employments. On the other hand, biaxial modeling efforts are still in a preliminary phase, and most common PBEE nonlinear analysis applications are performed discarding biaxial nonlinear component models. Thus, in the following, the only uniaxial behavior of elements is described.

Failure Mode Classification of RC Elements

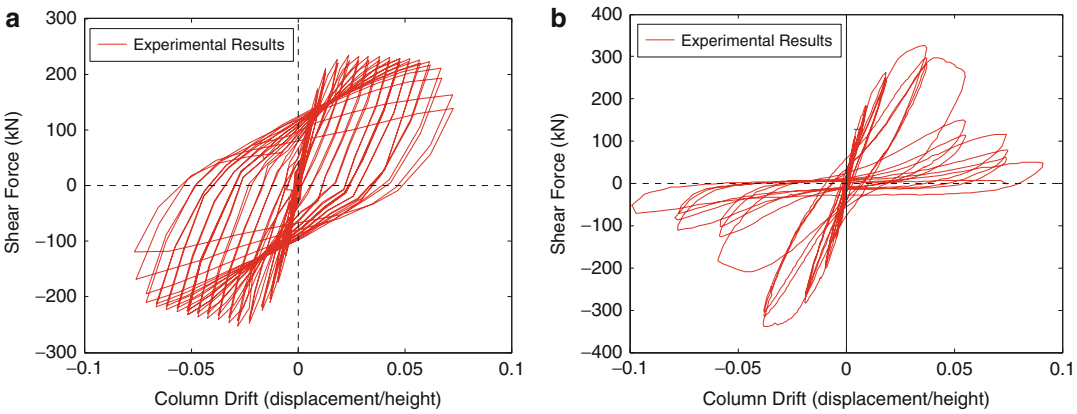
Columns are generally subjected to the contemporary presence of axial load, shear forces, and bending moment. The interaction between those internal forces strictly affects the response of structural elements and consequently their mode of failure. Interaction effects, i.e., the influence of shear forces on flexural deformations and that of normal stress resultants on shear deformations, are important. In general, they cannot be neglected,

especially for inelastic cyclic loading for non-capacity-designed elements. Flexural response can be affected by the interaction with shear also in linear elastic phase, when plastic shear capacity of the element is not attained. Conversely, in the case of inelastic response, shear capacity can decrease after yielding as a result of increasing ductility demand in the element. In fact, damage caused by cyclic degradation in flexural response decreases post-cracking shear capacity mechanisms (e.g., aggregate interlock, loss of anchorage of transverse reinforcement, etc.) resulting in a decrease of shear capacity.

In Fig. 1, two experimental tests on rectangular RC columns from PEER database by Berry et al. (2004) are shown. The first column, Fig. 1a, is characterized by flexural-dominated behavior, while the second column, Fig. 1b, is characterized by post-yielding interaction between flexure and shear.

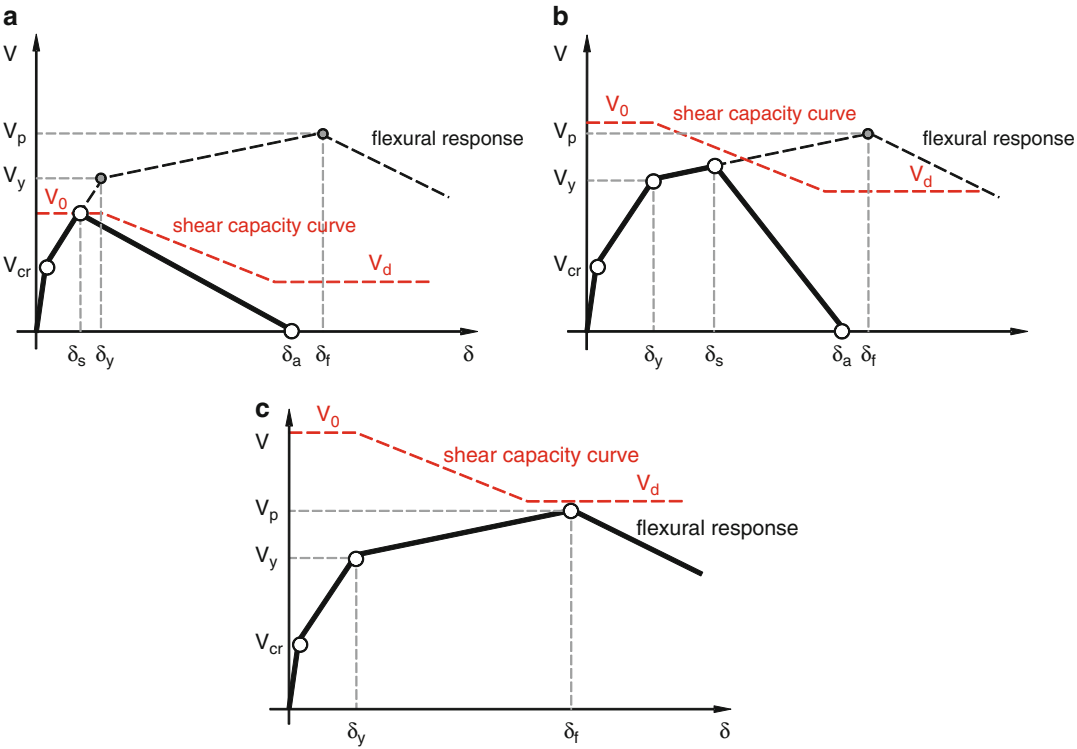
The above observations allow indentifying three different failure modes for columns resulting from flexure–shear interaction, as shown, schematically, in Fig. 2. The analytical classification of different modes of failure can be made only considering degrading shear capacity models which account for shear capacity degradation after yielding.

1. Shear failure mode, S – initial shear capacity (V_o), i.e., not affected by cyclic degradation, is lower than the plastic shear capacity (V_p), and flexural response is modified by a preemptive shear failure in the elastic response phase of the element. Deformation capacity (i.e., drift, chord rotation) is very limited. Subsequent response of the element is characterized by a strictly degrading behavior with significant shear strength reduction and a consequently loss of axial load-carrying capacity; see Fig. 2a.
2. Flexure–shear failure mode, FS – V_o is higher than V_p , and flexural response is characterized by an inelastic phase. However, if degraded shear strength (V_d), the lower red dotted branch in Fig. 1, is lower than V_p , the inelastic flexural response is modified by a shear failure that occurs at an intermediate shear strength



Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 1 Experimental results for Berry et al. (2004) database. Experimental tests by (a) Legeron

and Paultre (2000) and (b) Pujol (2002) (Adapted from Haselton et al. 2008)



Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 2 Collapse modes of RC elements subjected to axial load, shear, and bending

moment: (a) shear failure mode, S; (b) flexure–shear failure mode, FS; (c) flexural failure mode, F

between V_0 and V_d . The corresponding deformation capacity is higher with respect to that of case (1). The behavior of the element succeeding this kind of shear failure is strictly

degrading up to the loss of axial load-carrying capacity; see Figs. 1b and 2b.

3. Flexural failure mode, $F - V_d$ is higher than V_p , and flexural response is not affected by

interaction with shear. Thus, the element is characterized by inelastic deformation dominated by flexure and its typical damage, such as local buckling of longitudinal reinforcement and concrete spalling. This damage typically corresponds to the ultimate flexural capacity (θ_f). The softening branch of shear force–displacement envelope ends up to the loss of vertical load-carrying capacity; see Figs. 1a and 2c. This is the failure mode typical of capacity-designed elements in which shear–flexure hierarchy is controlled during the design process (see also De Luca and Verderame 2013).

The classification of failure mode is made on the basis of the ratio between plastic shear capacity V_p and degrading shear strength V_n . Several models have been developed to represent the degradation of shear strength with increasing inelastic deformations (Watanabe and Ichinose 1992; Aschheim and Moehle 1992; Priestley et al. 1994; Biskinis et al. 2004; Sezen and Moehle 2004).

In particular, the shear-degrading model proposed by Sezen and Moehle (2004) estimates column shear strength as the summation of shear carried by concrete (V_c) and shear carried by transverse reinforcement through a 45° truss model (V_s). This shear strength model is calibrated on 51 experimental column tests, and it has the form shown in Eq. 2.

$$\begin{aligned} V_n &= k \cdot V_o = k \cdot (V_c + V_w) \\ &= k \cdot \left[\left(\frac{0.5\sqrt{f_c}}{L_v/d} \cdot \sqrt{1 + \frac{P}{0.5 \cdot \sqrt{f_c} \cdot A_g}} \right) \cdot \right. \\ &\quad \left. 0.8A_g + \frac{A_{sw} \cdot f_{yw} \cdot d}{s} \right] \end{aligned} \quad (2)$$

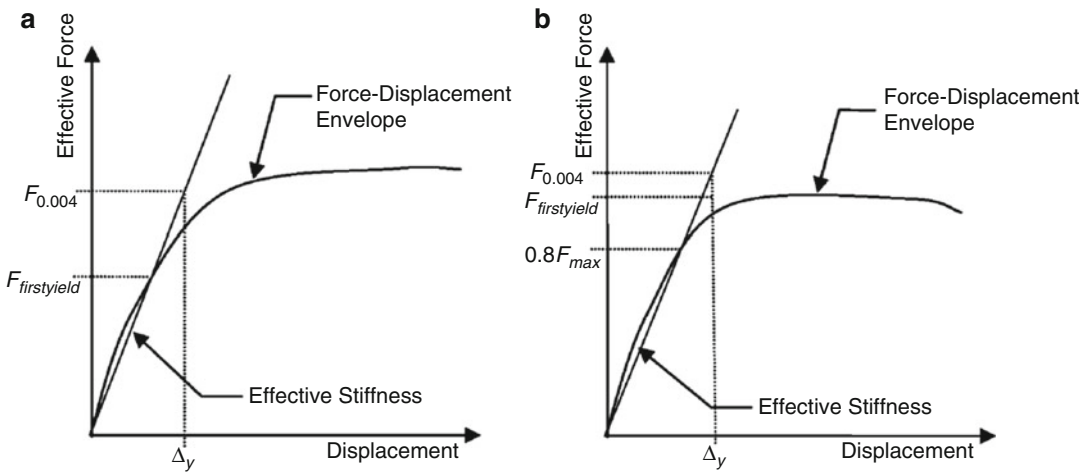
Shear strength contribution carried by concrete depends on concrete compressive strength, (f_c); shear span, (L_v); section depth, (d); axial load, (P); and gross area section, (A_g), while shear strength contribution carried by reinforcements depends on transverse reinforcement area (A_{sw}) within a spacing (s) in the loading direction,

yielding steel strength (f_y), and section depth (d), computed as distance from the extreme compression fiber to centroid of tension reinforcement. This model takes into account shear strength degradation after yielding through a degradation coefficient (k); k is defined to be equal to 1.0 for displacement ductility less than 2, to be equal to 0.7 for displacement ductility exceeding 6, and to vary linearly for intermediate displacement ductilities. The degradation coefficient k is applied to both V_c and V_s , under the assumption that concrete component degrades due to increased cracking and degradation of the aggregate interlock mechanism, while steel component likewise degrades due to hoop opening and bond degradation. This shear capacity model is adopted in ASCE/SEI 41 guidelines (2007).

It is worth noting that the classification of column failure modes based only on shear strength and V_p can be still not adequate. Other column parameters may also influence the observed failure mode (Zhu et al. 2007). These considerations are reflected by ASCE/SEI 41 – supplement 1 provisions (Elwood et al. 2007) as follows: the expected failure is first evaluated based on the ratio between V_p and the shear capacity, and then, when classifying a column as flexure-controlled, further parameters are considered, namely, the transverse reinforcement ratio, assuming 0.002 as the lower limit, as in Zhu et al. (2007), and the tie spacing-to-section depth ratio, assuming 0.5 as the upper limit.

Characteristic points of shear force–displacement envelope (i.e., moment–chord rotation), for each failure mode, can be identified schematically through the representation shown in Fig. 2. Monotonic and cyclic deformation capacities to be attributed to each characteristic point can be found in literature. It is worth to note that if in the elastic phase the distinction between monotonic and cyclic deformation capacity can be unnecessary (pre-yielding phase), such distinction becomes significant in the inelastic phase (post-yielding), especially if those deformation capacities are employed in the analytical modeling of the element.

Monotonic deformation capacities are employed for the definition of the force–deformation envelope



Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 3 Definition of yield displacement and effective stiffness from test data for

(a) yielding columns and (b) columns that did not yield (From Elwood and Eberhard 2009)

of the element. Successively, it is necessary to assign cyclic degrading rules. So, for example, a specific degradation percentage of peak force (conventional failure) is attained at a deformation capacity lower with respect to that considered on the monotonic envelope. When cyclic degrading rules are not modeled, the conventional failure has to be fixed at a displacement capacity that implicitly accounts for it (i.e., a value on the monotonic envelope that corresponds to the lower capacity degraded by cyclic effects). It is worth noting that the latter solution is obviously an approximated way for the estimation of a reliable deformation capacity. Cyclic degradation depends on the load path, and it cannot be accurately captured analytically with monotonic models. Still, for some specific failure modes, the monotonic degraded solution can be more straightforward with respect to the calibration of fulfilling cyclic models.

Yielding Displacement in RC Members

For the estimation of a reliable inelastic backbone of an RC element, the estimation of the yielding point is quite relevant. In literature, different empirical or hybrid models for the estimation of yielding drift capacity based on experimental observations are available. Direct estimations of yielding drift (Elwood and Eberhard 2009; Biskinis and Fardis 2010a) are generally based

on the assumption that yielding drift is the sum of three different components: a flexural component, a shear component, and a fixed end rotation component due to longitudinal bar slip; see Eq. 3. Alternatively yield drift can be evaluated indirectly from the estimation of yield stiffness (Elwood and Eberhard 2009; Biskinis and Fardis 2010a).

$$\delta_y = \delta_{y, \text{flex}} + \delta_{y, \text{shear}} + \delta_{y, \text{slip}} \quad (3)$$

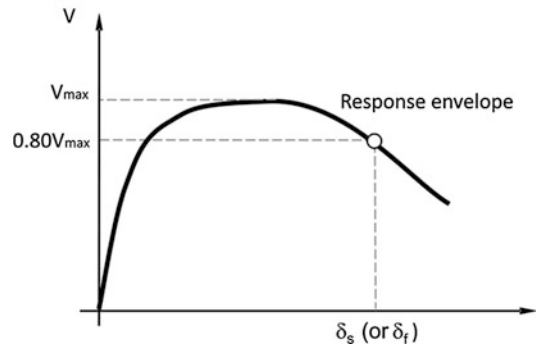
Yield drift obtained through Eq. 3 is generally evaluated as the corner for an ideal envelope approximating a bilinear force deformation response of the element and not as the drift corresponding to the first yielding in the reinforcement steel (or equivalently in the concrete) at the first section of the element. Figure 3 shows the procedure by Elwood and Eberhard (2009) for the definition of yielding displacement on the envelope of measured lateral load–displacement relationship corrected for P-delta. Elwood and Eberhard's procedure covers both the cases in which yield force is attained in the envelope (Fig. 3a) and cases in which the calculated yield force is not attained (Fig. 3b); in fact, it provides an estimate of effective stiffness also in the case of columns that do not yield (see Fig. 3b).

Eurocode 8 part 3 (CEN 2005) recommends a three-component model for the estimation of yielding drift, while ASCE/SEI 41 (2007) provides an effective stiffness to be computed as function of the axial load to which the element is subjected. Experimental databases on which the above formulations are calibrated are often characterized by modern code-conforming elements with proper seismic detailing. In the case of nonconforming elements, such formulations have still shown a fair agreement; emphasizing that seismic detailing does not affect strictly yielding capacity (Biskinis and Fardis 2010a). It is worth noting that nonconforming definition changes according to the specific code considered; on the other hand, this definition is generally referred to lack of detailing in transversal reinforcements, smooth bar presence, and lack of confinement (e.g., no 135° hooks, poor stirrup spacing, etc.). Analogously, database of nonconforming elements with smooth bars have shown a fair agreement with empirical formulations based on conforming elements with ribbed bars (e.g., Ricci et al. 2013).

Ultimate Drift in RC Members

Other characteristic point of RC member force–displacement response is the ultimate drift capacity for each failure mode considered. Ultimate drift capacity is generally evaluated at the drift characterizing a 20 % loss of the maximum shear strength (FIB 2003) attained in the element; see Fig. 4. According to the failure mode, different empirical ultimate drift capacity formulations are available in the literature. Most experimental databases are characterized by cyclic tests and few monotonic tests. Thus, most of those empirical formulations account implicitly for cyclic degradation.

Zhu et al. (2007) provide an ultimate drift capacity model for two collapse modes. They consider a database of 125 nonconforming columns. Tests are divided in two different sub-databases according to a binary classification approach: Zone S columns (shear-dominated failures) and Zone F columns (flexural-dominated failures). Zone F is composed of 85 tests, while Zone S is composed of 40 tests that include both



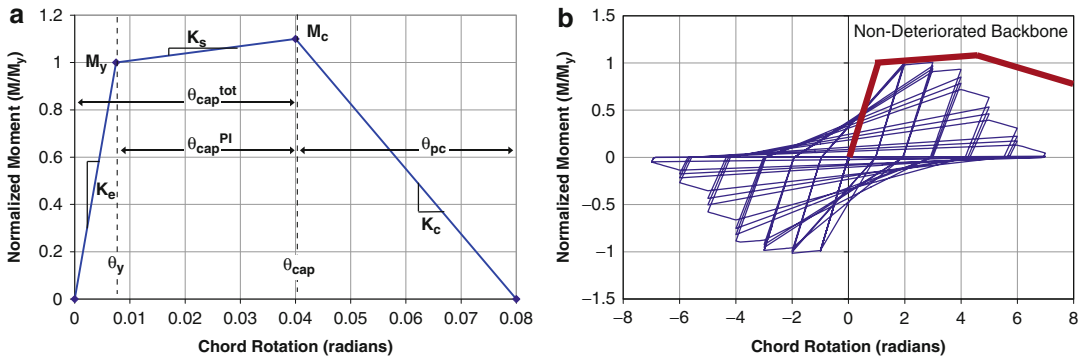
Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 4 Conventional criterion for the evaluation of ultimate drift capacity independently of the failure mode observed

shear- and flexure–shear-dominated failures. Classification is made on the basis of three column parameters: ratio of plastic shear demand to strength ratio (V_p/V_n), the aspect ratio (a/d), and the transverse reinforcement ratio (ρ''). The two sub-databases are employed for the estimation of the median ultimate drift capacity (20 % degradation of maximum shear strength) in the two cases, with characterization of logarithmic standard deviations. Equations 4 and 5 show the median estimates for Zone S and Zone F ultimate drift capacity, respectively. Zone S ultimate drift is function of hoop spacing to depth ratio (s/d), aspect ratio (a/d), and normalized axial force (v). Zone F ultimate drift depends on longitudinal reinforcement ratio (ρ_l), mechanical transversal reinforcement ratio ($\rho''f_{yt}/f_c$), s/d , and v .

$$\delta_s = 2.02\rho'' - 0.025\frac{s}{d} + 0.013\frac{a}{d} - 0.031v \quad (4)$$

$$\delta_f = 0.049 + 0.716\rho_l + 0.120\frac{\rho''f_{yt}}{f_c} - 0.042\frac{s}{d} - 0.070v \quad (5)$$

An alternative solution for the characterization of ultimate drift capacity for flexure–shear mode of failure (i.e., after yielding) is provided by Elwood and Moehle (2005a) based on a



Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 5 (a) Monotonic and (b) cyclic behavior of component model used in the calibration study by Haselton et al. (2008) (From Haselton et al. 2008)

database of 51 lightly transversal-reinforced columns (Sezen and Moehle 2004); see Eq. 6. This ultimate drift depends on the maximum nominal shear stress (v), ρ'' , and v .

$$\delta_s = \frac{3}{100} + 4\rho'' - \frac{1}{40} \frac{v}{\sqrt{f_c}} - \frac{1}{40} v \geq \frac{1}{100} \quad (6)$$

For flexural-dominated behavior, an empirical ultimate drift capacity is provided by Biskinis and Fardis (2010b). It represents the evolution of the first formulation provided by Panagiotakos and Fardis (2001). This formulation is based on a database of 1,352 uniaxial columns with good detailing and continuous bars (299 monotonic and 1,053 cyclic tests). According to the loading of the test, Fardis and his coauthors provide two empirical formulations for ultimate drift capacity, one for monotonic loading (see Eq. 7) and one for cyclic loading (see Eq. 8). Biskinis and Fardis' formulations depend on v , f_c , shear span ratio (L_v/h), longitudinal mechanical reinforcement ratio in compression and tension (ω' and ω , respectively), confinement effectiveness factor (α), transversal reinforcement ratio (ρ_w or ρ''), transversal reinforcement yielding strength (f_{yw} or f_{yt}), and diagonal reinforcement ratio (ρ_d).

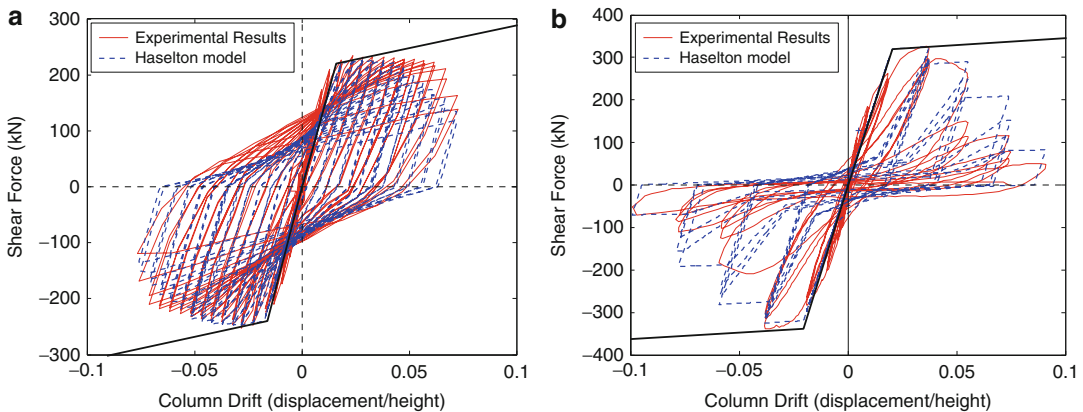
$$\delta_f = 0.028 \cdot (0.3^v) \left[\frac{\max(0.01; \omega')}{\max(0.01; \omega)} f_c \right]^{0.225} \times \left[\min \left(9; \frac{L_v}{h} \right) \right]^{0.35} 25^{\frac{\alpha \rho_w f_{yw}}{f_c}} 1.25^{100 \rho_d} \quad (7)$$

$$\delta_f = 0.016 \cdot (0.3^v) \left[\frac{\max(0.01; \omega')}{\max(0.01; \omega)} f_c \right]^{0.225} \left[\min \left(9; \frac{L_v}{h} \right) \right]^{0.35} 25^{\frac{\alpha \rho_w f_{yw}}{f_c}} 1.25^{100 \rho_d} \quad (8)$$

Considering a database of 48 cyclic tests, Biskinis and Fardis (2010b) provide a correction factor to Eq. 8 that accounts for poor detailing (i.e., older columns), and it is equal to (1/1.20). For substandard elements with smooth bars (based on 31 tests), the same authors provide an additional correction factor, equal to 0.95, to be applied to Eq. 8 (see also Verderame et al. 2010).

Flexural Component Model

For flexural-dominated behavior, Haselton et al. (2008) provide a complete calibration based on Berry et al. (2004) database for the specification of a complete nonlinear structural component model based on the beam-column element model developed by Ibarra et al. (2005). Figure 5 shows the monotonic and cyclic behavior of the component model employed by Haselton et al. (2008). Parameters to be considered for the employment of Haselton et al. (2008) model are (1) M_y , yielding moment; (2) θ_y , yielding rotation; (3) θ_{cap} , monotonic chord rotation at onset of strength loss (i.e., capping); (4) K_s , hardening stiffness; (5) θ_{pc} or K_c , post-capping stiffness; (6) λ , normalized hysteretic energy degradation capacity (i.e., cyclic degradation



Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 6 Example of Haselton et al.'s calibration for test specimens 193 and 212 (see Fig. 1)

characterized by (a) flexural failure mode and (b) flexure–shear failure mode (Adapted from Haselton et al. 2008)

parameter); and (7) c , exponent term to model rate of deterioration (i.e., cyclic parameter). This model accounts for cyclic degradation directly through the modeling of hysteretic behavior. It is significantly different with respect to the other models considered and also more complete. On the other hand, drift capacity thresholds that implicitly account for cyclic degradation cannot be evaluated on the monotonic backbone provided by this model. In fact, the monotonic backbone by Haselton without the degrading parameters λ and c is incomplete and nonconservative for RC elements capacity evaluation through static methods.

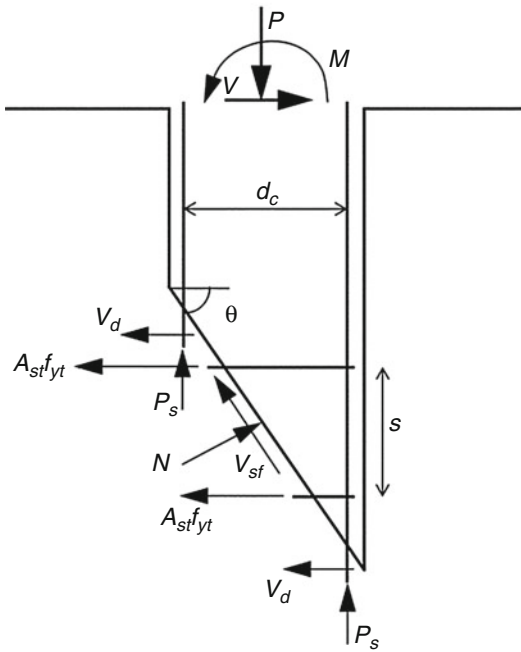
It is worth to note that Haselton et al. make a calibration effort for most of columns of Berry et al. database considering both columns that failed in flexural mode and columns characterized by flexure–shear mode. In Fig. 6, an example of the calibration made by Haselton et al. (2008) is shown for the two experimental test specimens provided in Fig. 1. The comparison of the two examples of calibration emphasizes that this component model is very well fitted for columns characterized by flexural behavior and less accurate for flexure–shear-dominated columns.

Loss of Vertical Carrying Capacity

Some of the experimental tests characterized by shear-dominated failure have shown, after the

attainment of ultimate deformation, a loss of vertical carrying capacity. In particular, Elwood and Moehle (2005b), on the basis of 12 tests characterized by flexure–shear mode of failure, observed that the drift ratio at axial load failure (δ_a) of a shear-damaged column is inversely proportional to the magnitude of the axial load and directly proportional to the amount of transverse reinforcement. Based on the classical shear friction model, Elwood and Moehle (2005b) propose a model for the drift at axial failure for shear-damaged columns. The axial load on a shear-damaged column is assumed to be supported by a combination of compression of the longitudinal reinforcement and force transfer through shear friction on an idealized shear failure plane (see Fig. 7).

The effective coefficient of friction from the classical shear friction equation is related to the drift ratio at axial failure using the results from 12 full-scale pseudo-static column tests. The effective coefficient of friction along the critical shear failure plane can be calculated using equilibrium and subsequently related to the drift ratio at axial failure. Elwood and Moehle found that the effective coefficient of friction calculated by ignoring the contribution of longitudinal reinforcement provided good agreement with the experimental data. According to the above observations, the empirical formulation provided in Eq. 9 is proposed for the evaluation of drift at



Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 7 Free body diagram of column after shear failure (From Elwood and Moehle 2005b)

loss of axial load-carrying capacity, where θ is equal to 65° , d_c is the depth of the column core (centerline to centerline of transverse reinforcement), and A_{st} and s are area and spacing of transverse reinforcement, respectively.

$$\delta_a = \frac{4}{100} \frac{1 + \tan(\theta)^2}{\tan(\theta) + P \left(\frac{s}{A_{st} f_{yt} d_c \tan(\theta)} \right)} \quad (9)$$

Elwood and Moehle's model is successively recalibrated by Zhu et al. (2007). Zhu et al. employed a database of 28 column specimens. All tests are unidirectional pseudo-static tests and were terminated after loss of axial load capacity. All the considered columns experienced flexural yielding prior to shear failure. Zhu et al. provide Eq. 10 for the evaluation of drift ratio at loss of vertical load-carrying capacity (axial failure), in which μ is the effective coefficient of friction evaluated according to Eq. 11. The same ruling parameters are assumed with respect to Eq. 9, and θ is equal, again, to 65° . Both Eqs. 9 and 10 are

based on the same model, but the most recent study (i.e., Eq. 10 by Zhu et al.) is based on a larger database, and it also provides a coefficient of variation equal to 0.35 for the median estimate of Eq. 10.

$$\delta_a = 0.184 \exp(-1.45\mu) \quad (10)$$

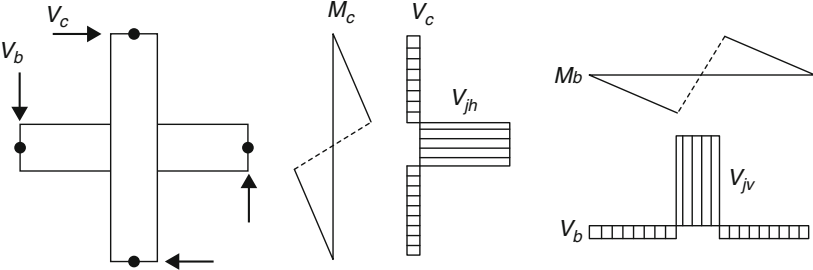
$$\mu = \frac{\frac{P}{A_{st} f_{yt} d_c / s} - 1}{\frac{P}{A_{st} f_{yt} d_c / s} - \tan \theta} \quad (11)$$

Beam-Column Joints

The performance of beam-column joints has been identified as a critical issue in the seismic resistance of RC moment-resisting frames (RC MRF). In RC MRF structures under severe ground motions, beam-column connections are subjected to moment reversals across the joint due to the adjacent beams and columns. As a result, the joint regions undergo significant horizontal (and vertical) shear forces whose magnitudes are much larger than those in the adjacent members (CEB 1996); see Fig. 8.

As a result, beam-column joints are susceptible to shear failure which generally involves a brittle process. Such brittle shear failure must be avoided through appropriate design to ensure ductile response of the frame. Despite the importance of shear design of RC beam-column joints, the research community has not yet developed a commonly accepted approach for the determination of the shear strength of RC beam-column joints, probably due to the complexity of joint behavior.

Many reinforced concrete (RC) buildings constructed without transverse steel shear reinforcement in the beam-column joint region still widely exist in seismically active regions, since the transverse reinforcement requirements for the design of beam-column joints were not addressed in earlier code provisions. Such unreinforced joints are considered vulnerable to brittle shear failure under earthquake shaking due to insufficient shear reinforcement in the joint region. These problems have been highlighted, in recent past, by the damage observed after



Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 8 Shear forces within the joint. M_c , V_c , M_b , and V_b are column and beam moment and

shear, respectively. V_{jh} and V_{jv} are horizontal and vertical joint shears (Adapted from Fardis 2009)

Seismic Vulnerability Assessment: Reinforced Concrete Structures, Table 1 γ_n for joint shear strength evaluation

$\gamma_n(MPa)^{0.5}$	Interior joint with transverse beams	Interior joint without transverse beams	Exterior joint with transverse beams	Exterior joint without transverse beams	Knee joint with or without transverse beams
ρ''					
<0.003	1.0	0.8	0.7	0.5	0.3
≥ 0.003	1.7	1.2	1.2	1.0	0.7

devastating earthquakes in different countries; moreover, many tests have proven the poor seismic performance of unreinforced joints, especially of exterior joints.

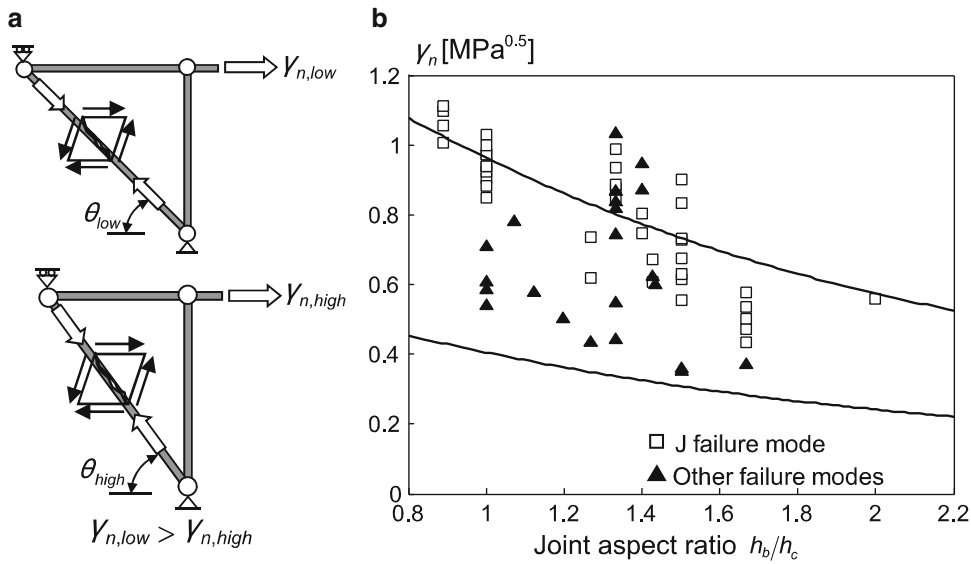
Design recommendations for beam–column joints in code provisions have been considerably revised since the ACI-ASCE Committee 352 (ACI 352R-02 2002) published its first seismic design guidelines in 1976. Still, nominal joint shear strengths proposed in current code provisions are only applicable to joints satisfying the minimum requirements of transverse reinforcement in the joint region. Some analytical models (Vollum 1998; Hwang and Lee 1999; Lowes and Altoontash 2003; Wong 2005) for predicting joint shear strength have been developed based on the strut-and-tie (SAT) idealization incorporating Mohr’s strain compatibility and softening concrete behavior, e.g., the modified compression field theory (MCFT) proposed by Vecchio and Collins (1986). However, LaFave and Shin (2005) indicated that the strength prediction using the MCFT tends to underestimate the shear strength of lightly reinforced and that of unreinforced joints.

Therefore, current code provisions and available analytical models may be inappropriate to predict the shear strength of unreinforced joints; see also Park and Mosalam (2013) for details.

Alternatively, ASCE/SEI 41 (2007) provides recommendations for the shear strength and joint shear stress–strain relationship of unreinforced joints for seismic rehabilitation purposes based on the pre-standard developed in FEMA 273 (FEMA 1997) and FEMA 356 (FEMA 2000). According to ASCE 41, nominal joint shear strength (V_n) is defined according to Eq. 12; it is a function of a coefficient γ_n , joint width (b_j), column depth (h_c), and concrete tensile strength ($\sqrt{f_c}$)

$$V_n = \gamma_n \sqrt{f_c} b_j h_c \tag{12}$$

The values of γ_n for joint shear strength calculation are presented in Table 1. To investigate the relevance of the strength recommendations in ASCE 41, Park and Mosalam (2012) collected 62 previous unreinforced exterior or corner joint test data. The evaluation of the joint shear



Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 9 Effect of the joint aspect ratio: (a) SAT idealizations for two joint aspect ratios and (b)

evaluation of the database against joint aspect ratio (Adapted from Park and Mosalam 2012)

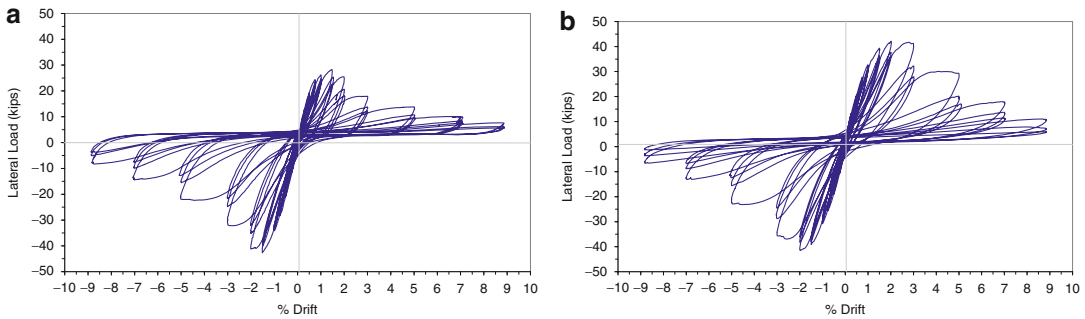
strength for the collected database reveals that ASCE 41 may underestimate the shear strength of unreinforced exterior joints. This parametric study shows that the shear strength of unreinforced exterior joints is affected by the joint aspect ratio, defined as the ratio of beam to column cross-sectional heights. The effect of the joint aspect ratio can be explained by the SAT idealization where a steeper diagonal strut is developed in the unreinforced joint with a high aspect ratio; see Fig. 9. Consequently, this steeper diagonal strut results in less effective shear resistance to equilibrate the horizontal joint shear force. Hence, the shear strength of unreinforced exterior joints decreases with the increase of the joint aspect ratio. Similar results are reported by Kim and LaFave (2007), Vollum and Newman (1999), and Bakir and Boduroğlu (2002).

Another key parameter, joint shear demand calculated from beam longitudinal bars and frame geometry, has been suggested by Anderson et al. (2008) based on results of unreinforced interior joint tests by Walker (2001) and Alire (2002). Depending on the joint shear demand, two types of joint failure have been commonly

recognized by other researchers (Wong 2005; Kim and LaFave 2007):

1. Joint failure prior to beam longitudinal reinforcement yielding (referred to as J type failure)
2. Joint failure after beam longitudinal reinforcement yielding (referred to as BJ-type failure)

The joint shear strength in most code provisions is pertinent to joint failure independent of beam longitudinal bars yielding or not, whereas the maximum joint shear stress for the BJ-type failure is directly determined from the joint shear demand dictated by the yielding of the beam longitudinal reinforcement. Conversely, experimental results show that the joint shear strength of unreinforced joints increases with the beam longitudinal reinforcement ratio. This evidence can be explained as follows: (i) increasing the beam longitudinal reinforcement ratio leads to the increase of the horizontal joint shear force without the yielding of the beam longitudinal bars, i.e., larger horizontal joint shear force is imposed with less deterioration of bond resistance around the beam longitudinal bars in the



Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 10 Experimental results of unreinforced exterior joint: (a) effect of beam-column

joint failure on the sub-assembly; (b) effect of anchorage failure (Adapted from Pantelides et al. 2002)

joint region; (ii) this more stable bond resistance produces a wider diagonal strut which can carry the larger horizontal joint shear force. Some strength models use either ductility factor (Park 2002; Hakuto et al. 2000) to predict the joint shear strength of BJ type of joint failure. A similar approach is adopted to predict the shear strength of columns (Sezen and Moehle 2004).

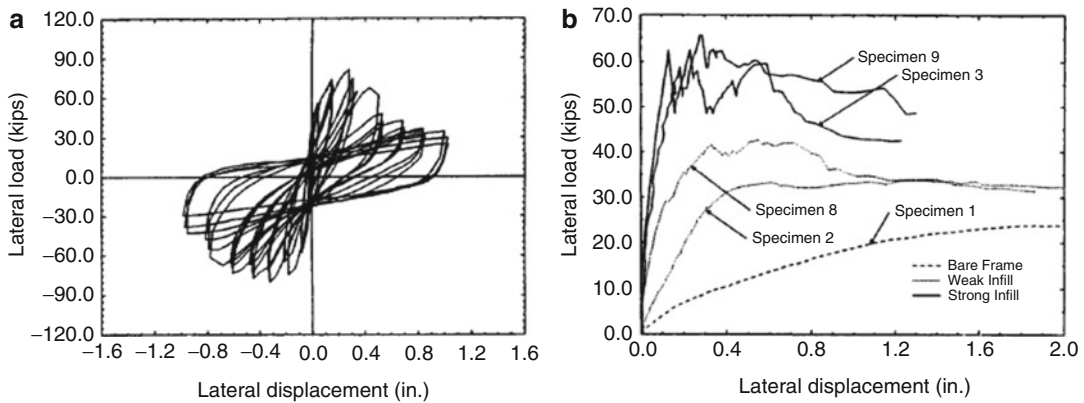
In general, experimental results on beam-column joints without transverse steel shear reinforcement emphasize a strictly degrading behavior caused, on one hand, by the shear behavior of the joint panels and, on the other hand, by the bond-slip behavior of longitudinal reinforcements anchored in it; see Fig. 10. Experimental research on the seismic performance of the beam-column joints (Walker 2001; Alire 2002; Pantelides et al. 2002) has revealed that the joint shear stress-strain response, typically, has a degrading envelope and a highly pinched hysteresis. Moreover, the common practice of terminating the beam's bottom reinforcement within the joints makes the bottom reinforcement prone to pullout under a seismic excitation. Insufficient beam bottom bar anchorage precludes the formation of bond stresses necessary to develop the yield stress in the beam's bottom reinforcement.

Finally, as described in section “**Loss of Vertical Carrying Capacity**,” axial failure of a shear-damaged column can occur by sliding along an inclined crack plane, with resistance provided by

transverse reinforcement clamping the crack and longitudinal reinforcement supporting axial force directly. Axial failure of a joint may be viewed similarly (Hassan 2011). After developing joint shear failure, the axial load will be supported by shear friction on the diagonal shear failure plane and the axial capacity of column reinforcing bars. Two axial capacity models designed for unconfined joints were developed. However, this result is obtained based on a small dataset.

Infills

The practice of realizing RC frames with masonry infill panels is very common in European countries. Masonry infill walls affect the strength and stiffness of infilled RC frame structures. The general approach to discard their contribution in design and assessment could be acceptable when proper capacity design rules are employed, since code recommendations account implicitly for their contribution through proper acceptance criteria and drift limitations (CEN 2004; ASCE/SEI 41 2007; DM14/01/2008 2008). In addition, their contribution to overall strength and stiffness is less significant in modern designed structures (e.g., Dolšek and Fajfar 2001). On the other hand, assessment of RC structures often involves substandard buildings in which the presence of masonry infill can strictly affect the structural behavior. Typical two-layer hollow clay brick infills employed in the Mediterranean area and very common in existing RC buildings cannot be neglected.



Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 11 Example of (a) cyclic- and (b) monotonically loaded specimens of RC-infilled frames (From Mehrabi et al. 1996)

Their contribution represents the first capacity source to earthquake shaking in this kind of buildings (e.g., De Luca et al. 2014b; Manfredi et al. 2014), and, at the same time, they can lead to preemptive brittle failures due to local interaction between infill and RC frame (e.g., Verderame et al. 2011).

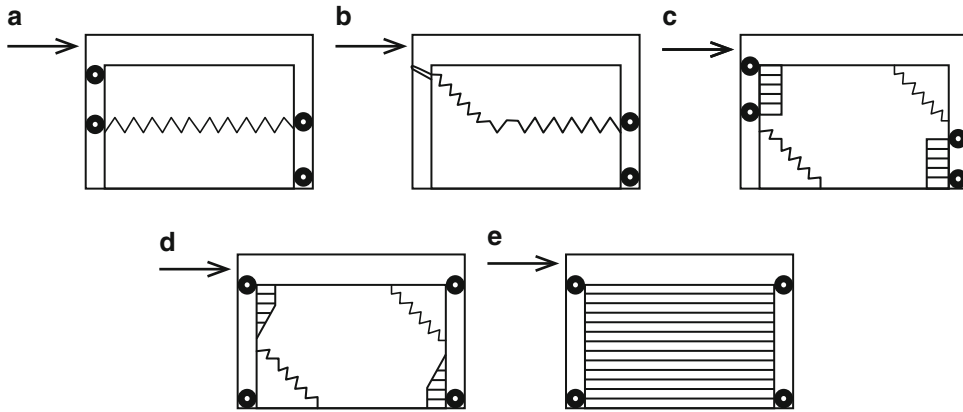
Notwithstanding the above considerations, it is worth noting that these elements are considered as *nonstructural* in the design process, and they are not subjected to the same acceptance and quality controls of RC elements. Thus their behavior can be significantly affected by specific executive practice and very different material properties.

Experimental investigations of the last decades (e.g., Mehrabi et al. 1996; Mosalam 1996; Colangelo 2003; Calvi et al. 2004, among others) emphasized a considerable reduction in the response of infilled frames under reversed cyclic loading. This behavior is due to the rapid degradation of stiffness, strength, and low energy dissipation capacity, resulting from the brittle and sudden damage of the unreinforced masonry (URM) infill walls. In Fig. 11 an example of cyclic (Fig. 11a) and monotonic (Fig. 11b) behavior of masonry infill frames from the experimental campaign by Mehrabi et al. (1996) is shown.

Based on both experimental and analytical results of the last decades, different failure mode classifications of masonry-infilled frames

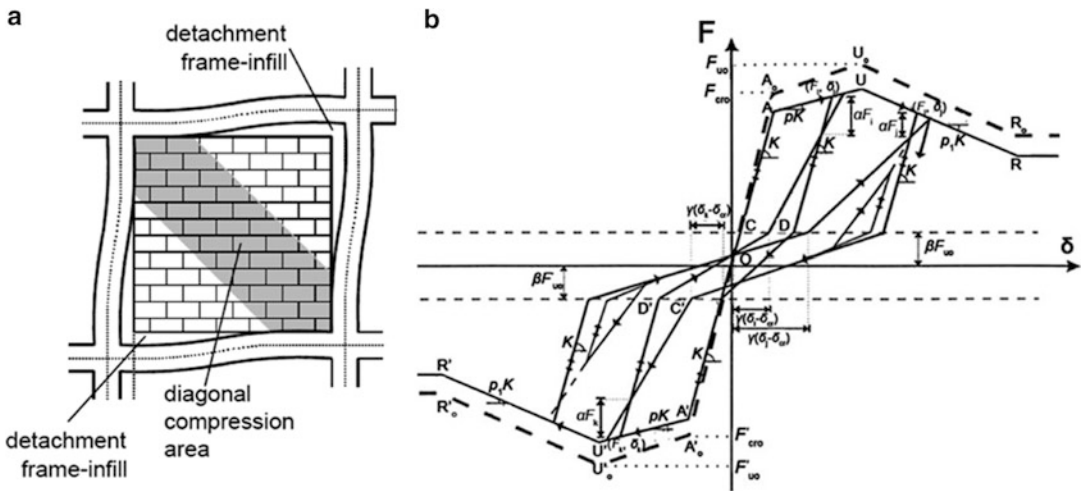
have been proposed. Infill failure can be classified into five distinct modes (e.g., Shing and Mehrabi 2002; Calvi et al. 2004; Asteris et al. 2011). Infill failure graphical examples according to Shing and Mehrabi (2002) are shown in Fig. 12.

- *Sliding shear mode* represents horizontal sliding shear failure through bed joints of a masonry infill. This mode is associated with infill of weak mortar joints and a strong frame.
- *Diagonal cracking mode* is seen in the form of a crack across the compressed diagonal of the infill panel and often takes place with simultaneous initiation of the sliding shear mode. This mode is associated with a weak frame or a frame with weak joints and strong members infilled with a rather strong infill.
- *Diagonal compression mode* represents the crushing of the infill within its central region. This mode is associated with a relatively slender infill, where failure results from out-of-plane buckling of the infill.
- *Corner crushing mode* represents the crushing of the infill in at least one of its loaded corners. This mode is usually associated with infilled frames consisting of a weak masonry infill panel surrounded by a frame with weak joints and strong members.
- *Frame failure mode* is seen in the form of plastic hinges developing in the columns or



Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 12 Failure mode according to the classification made by Shing and Mehrabi (2002), (a)

midheight crack, (b) diagonal crack, (c and d) corner crushing, (e) horizontal slip (Adapted from Calvi et al. 2004)

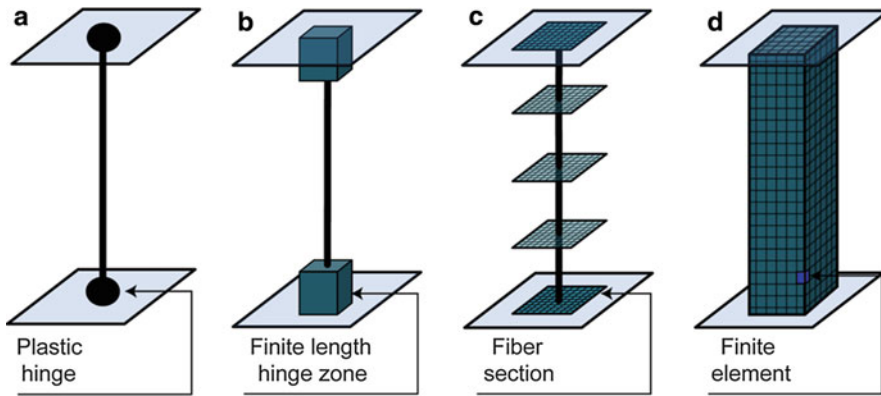


Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 13 (a) Strut model analogy of infilled frame (From Asteris et al. 2011); (b) Fardis and Panagiotakos (1997) analytical model for infills (From Fardis 2009)

the beam–column connections. This mode is associated with a weak frame or a frame with weak joints and strong members infilled with a rather strong infill.

Analytical modeling of infills can be carried out according to different macro-modeling approaches. Equivalent strut macro-model approach is the simplest way to model the global interaction between infill and RC elements; see Fig. 13a; it is also explicitly suggested by codes (e.g., ASCE/SEI 41 2007). On the other hand, the

characterization of the strut can be made according to different formulations. In literature are available different overviews of the different analytical approaches (e.g., Crisafulli 1997; Chrysostomou and Asteris 2012, among others). Some analytical approaches are able to account for all the failure modes considered above (e.g., Bertoldi et al. 1993), other macro-models does not account explicitly for all the five modes of failure but show a fair agreement with experimental results (e.g., Fardis and Panagiotakos 1997); see Fig. 13b.



Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 14 Idealized models of RC elements (Adapted from Deierlein et al. 2010)

Local interaction between infills and RC frame can be captured by multiple strut macro-models (e.g., Crisafulli 1997) or by single strut macro-models with ends located at the edge of the column or of the beam (e.g., ASCE/SEI 41 2007). Notwithstanding the fact that it can be quite hard to distribute strength among the different struts of the analytical model, still modeling local interaction allows to check if brittle failures are likely to occur in the RC elements (e.g., Verderame et al. 2011).

Analytical Modeling of RC Structures

Inelastic analytical structural component models can be differentiated by the way that plasticity is distributed through the member cross sections and along its length. Modeling approaches can be divided in three main groups: (i) lumped plasticity models, (ii) distributed plasticity models, and (iii) finite element micro-modeling. The main difference between lumped plasticity, distributed plasticity, and finite element models resides in the parameters employed for the definition of inelastic behavior. In the first case, moment–rotation relationships are employed, while, in the other two options, moment–curvature or stress–strain relationships are assumed, and it is necessary to pass through a numerical integration step.

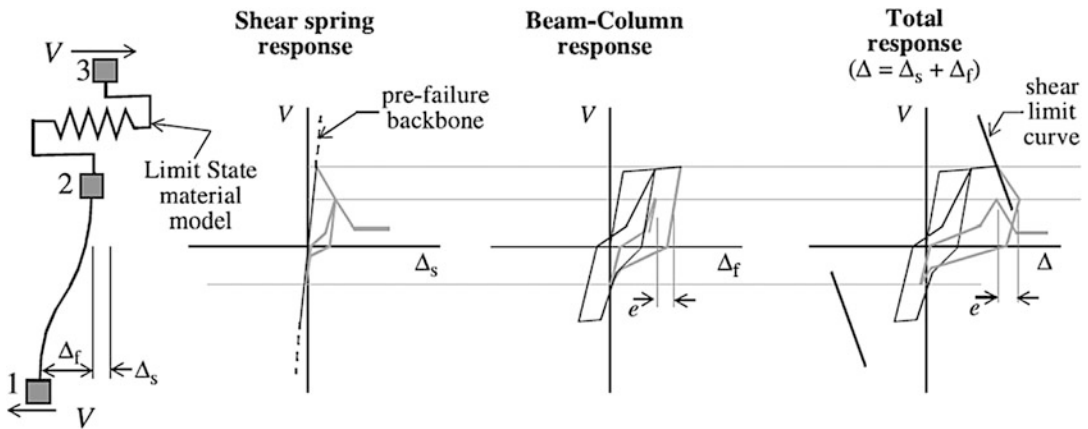
The analysis of experimental behavior of structural elements in section “[Behavior of RC Elements and Nonstructural Elements](#)” allowed

the definition of nonlinear structural component models for RC buildings, in terms of $M-\theta$. It was emphasized that component models are generally defined on the basis of experimental results, and often curvature is not the easiest parameter to be employed for the description of all the modes of failures characteristic of RC element. Thus, most commonly, PBEE applications employ lumped plasticity models. On the other hand, it is possible to find PBEE loss and vulnerability assessment applications on RC structures that employ fiber models for low-intensity levels (e.g., Mitrani-Reiser 2007).

In the following, a brief description of both analytical modeling solutions is provided. A more complete overview of modeling approaches is provided in Deierlein et al. (2010). In Fig. 14, the comparison of different idealized model types for simulating the inelastic response of elements is shown.

Lumped plasticity models concentrate the inelastic deformations at the end of the element, such as through a rigid-plastic hinge or an inelastic spring with hysteretic properties (Fig. 14a). By concentrating the plasticity in zero-length hinges with moment–rotation model parameters, these elements have relatively condensed numerically efficient formulations.

Distributed plasticity models are different according to the way the element is discretized, and numerical integration is carried out resulting in finite length hinge models (Fig. 14b) or fiber



Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 15 Shear spring in series model using limit state material model (From Elwood 2004)

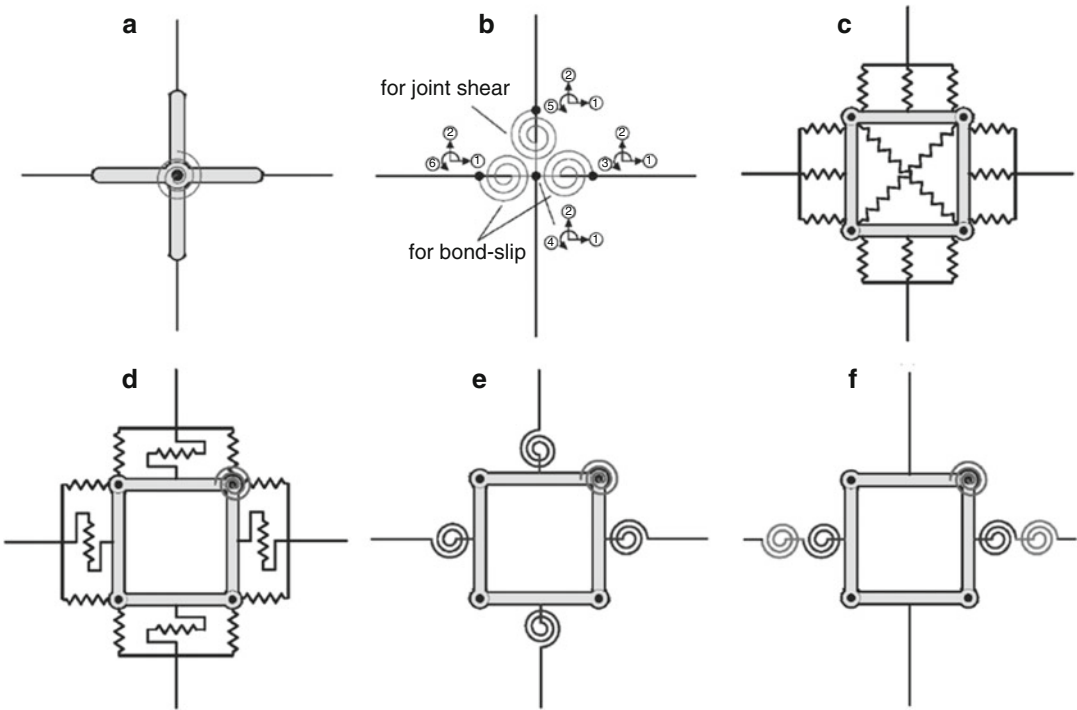
model (Fig. 14c). In the finite length hinge model, cross sections in the inelastic hinge zones are characterized through either nonlinear moment–curvature relationships or explicit fiber-section integrations that enforce the assumption that plane sections remain plane. The inelastic hinge length may be fixed or variable, as determined from the moment–curvature characteristics of the section together with the concurrent moment gradient and axial force. The fiber formulation models distribute plasticity by numerical integrations through the member cross sections and along the member length. In this latter case, stress–strain relationship is defined and plane-sections-remain-plane assumption is enforced; see Deierlein et al. (2010). Distributed fiber formulations do not generally report plastic hinge rotations, but instead report strains in concrete cross-section fibers. The calculated strain demands can be quite sensitive to the moment gradient, element length, integration method, and strain hardening parameters on the calculated strain demands. These modeling approaches allow capturing the interaction between axial load and biaxial components of bending moment at section level, under cyclic loading.

Finally, finite element models (Fig. 14d) represent a detailed micro-modeling solution in which material constitutive relationships are assigned to each element. It is obvious that this

latter approach represents the most detailed way to represent the elements; on the other hand, it is computationally demanding, and it asks for the definition of numerous input parameters that, in turn, need to be calibrated.

While distributed plasticity formulations capture variations of stress and strain through the section and along the member in more detail, important local behaviors, such as strength degradation due to local buckling of steel reinforcing bars or nonlinear interaction of flexural and shear, are difficult to capture without sophisticated and numerically intensive models. Phenomenological concentrated hinge/spring models may be better suited to capture the nonlinear degrading response of members through calibration (see Fig. 15), using member test data, on experimental moment rotations and hysteresis curves (Deierlein et al. 2010); see Fig. 6. It is worth noting that even when opting for lumped plasticity models, it is possible to refine the analytical modeling choice in different ways, and the accuracy of results is consequently affected; see, for example, Yavari et al. (2009).

Beam–column joint modeling is still at an early and less mature stage with respect to member modeling (i.e., beam and columns), and practical available solutions are to consider rotational springs capable to describe the beam–column joint behavior that is governed by shear and



Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 16 Existing beam-column joint models: (a) Alath and Kunnath (1995), (b) Biddah and Ghobarah (1999), (c) Youssef and Ghobarah (2001),

(d) Lowes and Altoontash (2003), (e) Altoontash (2004), and (f) Shin and LaFave (2004) (From Celik and Ellingwood 2008)

bond-slip deformation; see some modeling scheme examples in Fig. 16 as collected by Celik and Ellingwood (2008). Alath and Kunnath (1995) modeled the joint shear deformation with a rotational spring model with degrading hysteresis (Fig. 16a). Biddah and Ghobarah (1999) modeled the joint shear and bond-slip deformations with separate rotational springs (Fig. 16b). Youssef and Ghobarah (2001) proposed a joint element in which two diagonal translational springs connecting the opposite corners of the panel zone simulate the joint shear deformation, and 12 translational springs at the panel zone interfaced simulate all other modes of inelastic behavior (Fig. 16c). Lowes and Altoontash (2003) proposed a 4-node 12° of freedom joint element that explicitly represents three types of inelastic mechanisms of beam-column joints under reversed cyclic loading (Fig. 16d). Successively Altoontash (2004) simplified the model by Lowes and Altoontash introducing a four

zero-length rotational springs located at beam-column joint interfaces (Fig. 16e). Shin and LaFave (2004) represented the joint by rigid elements located at the edges of the panel zone and rotational springs embedded in one of the four hinges linking adjacent rigid elements (Fig. 16f).

Damage Measures

The nature and amount of structural damage depends necessarily on the quality of the materials that compose structural and nonstructural elements, and on the configuration and type of structural systems. In the early years of modern earthquake engineering, damage definition was basically approached in qualitative terms (e.g., through the definition of probable localization of such damage in a structure). This kind of approach relied fundamentally on the observation of damaged structures after seismic events or it

was based on experimental tests. The first studies for damage quantification of RC structural members date back to late 1980s (e.g., Park and Ang 1985; Fardis et al. 1992). The availability of experimental data allowed more and more detailed approaches to damage analysis up to the most recent PBEE applications.

Most recent studies on damage analysis employ component-based approaches (e.g., Aslani 2005; Mitrani-Reiser 2007). The PBEE damage analysis overpasses the approach to global estimate of damage implicit in previous damage scales (e.g., ATC-13 1985; Grunthal 1998; NIBS 1999) and switch to a component-based damage analysis that allows a more accurate definition of losses. On the other hand, component-based approach requires a significant amount of data, such as experimental data, earthquake experience, expert opinion, and some combination of these.

For component-based approaches, it is necessary to define fragility functions for various damage states for each structural and nonstructural component in typical RC frames. In literature, there are available different benchmark studies for damage assessment of RC buildings (e.g., Aslani 2005; Mitrani-Reiser 2007). FEMA P-58-1 (2012) provides a list of fragility specifications for structural components, nonstructural components, and contents.

The functional form most commonly used for components' fragility functions is the lognormal distribution that asks for the definition of median and logarithmic standard deviation. Therefore, for each component, it is necessary to define median capacity and logarithmic standard deviation ($\sigma_{\log}$) in terms of EDP for each damage state.

In section "Examples of Damage Analysis", there is a brief overview of component-based damage analyses for RC structural components and infills according to different authors. This overview is aimed at providing an example of damage evaluation made for the elements whose behavior was considered in section "Behavior of RC Elements and Non Structural Elements." Notwithstanding the fact that quantitative approaches to component damage analysis rely on experimental databases, they are still based on a limited

number of data, and they cannot be considered fully reliable yet, since they need further numerical refinement.

Finally, in section "Code Limit States," a brief overview of code acceptance criteria for RC elements at different limit states is provided, and compared with the information provided in section "Examples of Damage Analysis." The two sections allow a comparison between next-generation performance-based codes (i.e., section "Examples of Damage Analysis") with respect to current performance-based codes (i.e., section "Code Limit States").

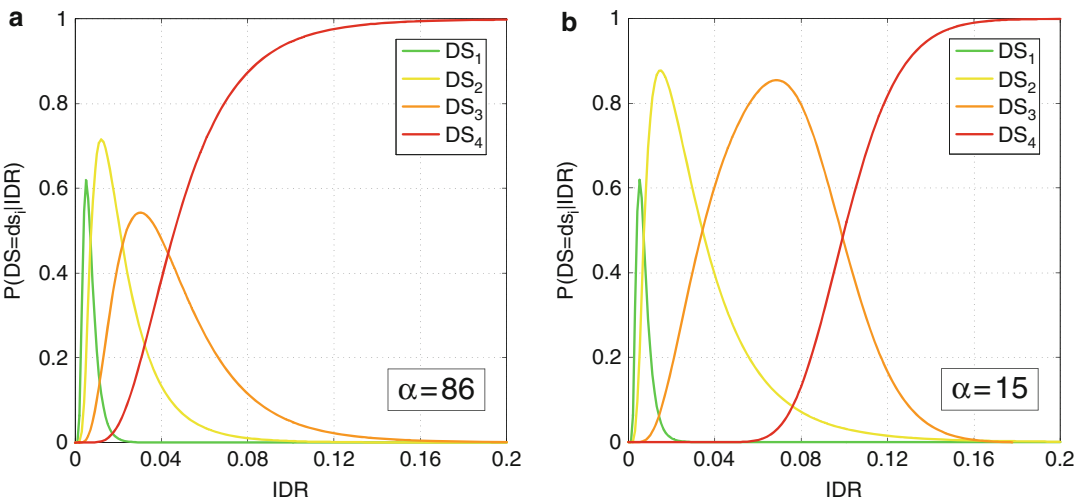
Examples of Damage Analysis

Various damage indices or EDP thresholds are used to quantify damage of RC structural members. As an example, Mitrani-Reiser (2007) employed for beam and columns' damage states the fragility curves developed by Beck et al. (2002) and, in turn, based on Williams and Sexsmith (1997) studies. Beck et al.'s fragility functions use a deformation damage index (DDI) that can be evaluated from chord rotations obtained in structural analyses. This approach allows the definition of four damage states for RC members. On the other hand, Aslani (2005) provides fragility functions for RC columns with light transversal reinforcement in terms of IDR, evaluated from experimental data, and based on the component modeling provided in section "Behavior of RC Elements and Non Structural Elements." For both approaches, four damage states are defined resulting in damage descriptions compared in Table 2. Notwithstanding the fact that Aslani refers specifically to lightly reinforced RC columns (i.e., non-ductile columns), similar damage descriptions are provided in Table 2. In both approaches, the typical lognormal shape is adopted for the fragility curve, so that for each damage state, a median EDP and a $\sigma_{\log}$ are estimated.

Aslani estimates median and $\sigma_{\log}$ for each damage states, based on a database of 92 lightly reinforced columns. The preliminary result of this estimation leads to a significant dispersion for all damage states; so an additional parameter (α) is considered in the regression to reduce the

Seismic Vulnerability Assessment: Reinforced Concrete Structures, Table 2 Comparison of damage states for columns

Williams and Sexsmith (1997) – ductile columns		Aslani (2005) – non-ductile columns	
Damage state	Damage description	Damage state	Damage description
None	None or small number of light cracks, either flexural (90°) or shear (45°)		
Light	Widespread light cracking or a few cracks >1 mm or light shear cracks tending to flatten toward 30°	DS1: light cracking	Visible cracks; crack widths smaller than 0.3 mm. Light repair to improve aesthetic appearance
Moderate	Significant cracking, e.g., 90° cracks >2 mm; 45° cracks >2 mm; 30° cracks >1 mm	DS2: severe cracking	Wider and deeper crack widths, more extensive compared to DS1
Severe	Very large flexure or shear cracks, usually accompanied by limited spalling of cover concrete	DS3: shear failure	Deterioration of shear capacity after yielding (that occurs at earlier stages of loading compared to ductile columns) leads to shear failure
Collapse	Very severe cracking and spalling of concrete; buckling, kinking, or fracture of rebar	DS4: axial failure	Loss of vertical carrying capacity. This damage state has possible disastrous consequences, if there is no possibility to redistribute vertical load to other members



Seismic Vulnerability Assessment: Reinforced Concrete Structures, Fig. 17 Probability of being at each damage state for non-ductile RC columns characterized by different α parameter

resulting $\sigma_{\log}$ of the fragilities and to adjust the estimated median IDR at each damage states. The parameter $\alpha = P/(A_g \rho'' f_c)$ is function of axial load (P), gross area section (A_g), concrete compressive strength (f_c), and transversal reinforcement ratio (ρ''). In Fig. 17, the probability of being in a specific damage state for lightly reinforced

columns is shown, and the effect of α is emphasized through the comparison of the curves in Fig. 17a, b. It is worth noting that α affects strictly shear failure and axial failure as it can be expected by the observation of the parameters ruling equations shown in section “Behavior of RC Elements and Non Structural Elements.”

Seismic Vulnerability Assessment: Reinforced Concrete Structures, Table 3 Damage state definition for interior and exterior beam–column joints according to Aslani (2005)

Damage state	Damage description
DS1: sever cracking in the beam	Wide and extensive cracking of the beam
DS2: severe cracking in the column	Wide and extensive cracking of the column
DS3: severe cracking in the joint	Severe cracking at the beam–column joint
DS4: joint spalling	Spalling of the concrete cover of the beam–column joint
DS5: loss of vertical carrying capacity	The joint collapses under its gravity load

For beam–column connections, Aslani considers different fragility curves for interior beam–column connections and exterior beam column connections. Damage state definitions are identical in the two cases, and they are shown in Table 3, while fragility functions, based on experimental data, can be really different. The numerical estimation for joints is again affected by high uncertainties.

Fragility functions for masonry infill panels are not available in the damage and loss studies cited up to this point. On the other hand, experimental tests (e.g., Mehrabi et al. 1996; Colangelo 2003) and post-earthquake damage observations (e.g., Ricci et al. 2011; De Luca et al. 2014b; Manfredi et al. 2014) have shown that this kind of elements can have a significant impact on both damage and loss analysis of RC structures, especially in the Mediterranean area, in which this kind of structural typology is quite common. Notwithstanding the fact that damage observations based on experimental tests are available in literature and macro-modeling approaches and component modeling backbones as well are available in literature, still characteristic parameters for infill fragilities are characterized by significant differences according to different authors (e.g., Gu and Lu 2005; Özcebe et al. 2012; Colangelo 2013). The point is that in the case of damage analysis of infills, authors do not even agree on damage states' qualitative definitions. The damage degrees are distinguished on the basis of the description of physical damage in

terms of cracking, crushing, etc. and the feasibility of repair. As an example, Table 4 provides damage description at different damage states according to different authors; see Colangelo (2013). Significant differences can be found in damage descriptions and, as consequences, also in the fragility function parameters estimated from each description (see also Colangelo 2013).

Despite the significant differences that can be found in literature formulations for component fragilities of infills, it is worth noting that widely employed damage scales for RC structures that interpret global damage of buildings often refer to infill damage. An example is the EMS 98 scale (Grunthal 1998) in which the first three grades of global damage are mainly characterized by damage description of masonry infills; see, for example, De Luca et al. (2014b) and Manfredi et al. (2014), in which infill damage analysis is employed as analytical proxy for damage classification of the whole RC building. On the other hand, EMS 98 damage scale is not a component-based approach, but the single most damaged element represents the damage classification of the whole structure.

Code Limit States

Current PBEE approach is also the result of the huge effort made in ATC-58 project. The purpose of this project is to develop next-generation seismic design procedures that will provide a more reliable means of predicting and designing performances of structures. On the other hand, PBEE has been already implemented in codes providing target performance levels. As an example, FEMA 356 (FEMA 2000) and successively ASCE/SEI 41 (2007) do not attempt to quantify the probability of achieving a given performance level or to quantify losses; but, both FEMA 356 and ASCE/SEI 41 address component-level and system-level damage states, and they relate damage to life safety and post-earthquake operability. The damage states given in FEMA 356 tend to be qualitative and open to multiple interpretations. In Table 5, target performance levels according to FEMA 356 and ASCE/SEI 41 are shown.

ASCE/SEI 41 (2007) succeeded FEMA 356 as document for the seismic assessment of existing

Seismic Vulnerability Assessment: Reinforced Concrete Structures, Table 4 Damage state definition for masonry infills according to Colangelo (2013)

Damage state	Damage description (Colangelo 2013)	Damage description (Gu and Lu 2005)	Damage description (Özcebe et al. 2012)
DS1	Onset of cracking in the bricks, associated with the first noticeable reduction of stiffness	Minor cracking and falling of plaster; only local repair needed with function maintained; maximum strength	Negligible cracks; maximum strength (i.e., base shear)
DS2	Moderate cracks before attaining the maximum strength (i.e., base shear)	Continual diagonal cracking and flaking; repairable damage; 30 % reduction of the maximum strength	Appreciable damage; maximum stress in the infill
DS3	Extensive cracks with tensile splitting and falling of the outer layer of a few bricks; repairable damage	Loss of wall integrity; 70 % reduction of the maximum strength	Ultimate strain in the infill
DS4	So many broken bricks that repair is unreasonable; reconstruction needed		

buildings; it provides a performance-based engineering framework whereby deformation and force demands for different seismic hazards are compared against deformation and force capacities for various performance levels. Elwood et al. (2007) provided an update to concrete provisions in ASCE/SEI 41 given new data available in literature and considering that practitioners observed that previous concrete provisions tend to err on the conservative side. In their update to concrete provisions, Elwood et al. provide modeling parameters and acceptance criteria modifications for columns, beam–column joints, etc. Regarding columns, the classification described in section “RC Members” is considered. As an example, for columns, drift limits at the different acceptance criteria change according to the classification of the expected mode of failure.

European code for the assessment of RC elements (CEN 2005) provides, as collapse prevention acceptance criteria for RC members, the value of the ultimate rotation capacity according to Biskinis and Fardis (2010b); as life safety acceptance criteria, 75 % of the ultimate rotation capacity according to Biskinis and Fardis (2010b); and, as immediate occupancy acceptance criteria, the value of yielding rotation

according to Biskinis and Fardis (2010a). All these chord rotation capacities are intended as member capacity thresholds if shear capacity evaluated according to Biskinis et al. (2004) is not attained, and a preemptive brittle failure (computed through a force-based approach) does not occur before the attainment of such chord rotations.

Regarding acceptance criteria for infills, codes (e.g., CEN 2004), in general, do not provide specific drift limits for infill elements, since it is not taken for granted the fact that infills are explicitly modeled. On the other hand, immediate occupancy and operational acceptance criteria are evaluated to implicitly account for damage to masonry infills.

In FEMA 356 and ASCE/SEI 41, four nonstructural performance levels are classified (see Table 5). The limit for exterior walls differs from the limit for heavy partitions (light partitions may be drywall partitions with studs, for instance) for both occupancy and life safety performance level. For the occupancy level for heavy partitions, the limit is 0.5 %, the same as Eurocode 8 (CEN 2004), while for exterior walls, the limit is 1 %. For life safety level, the above occupancy limits double. The Italian code (DM14/1/2008 2008) also prescribes 0.5 % for

Seismic Vulnerability Assessment: Reinforced Concrete Structures, Table 5 Damage control and building performance level extracted from Table C1-2 in FEMA 356 (FEMA 2000) and ASCE/SEI 41 (2007)

	Collapse prevention level	Life safety level	Immediate occupancy level	Operational level
Overall damage	Severe	Moderate	Light	Very light
General	Little residual stiffness and strength, but load-bearing columns and walls function. Large permanent drifts. Some exits blocked. Infills and unbraced parapets failed or at incipient failure. Building is near collapse	Some residual strength and stiffness left in all stories. Gravity-load-bearing elements function. No out-of-plane failure of walls or tipping of parapets. Some permanent drift. Damage to partitions. Building may be beyond economical repair	No permanent drift. Structure substantially retains original strength and stiffness. Minor cracking of facades, partitions, and ceilings as well as structural elements. Elevators can be restarted. Fire protection operable	No permanent drift. Structure substantially retains original strength and stiffness. Minor cracking of facades, partitions, and ceilings as well as structural elements. All systems important to normal operation are functional
Nonstructural components	Extensive damage	Falling hazards mitigated but many architectural, mechanical, and electrical systems are damaged	Equipment and contents are generally secure but may not operate due to mechanical failure or lack of utilities	Negligible damage occurs. Power and other utilities are available, possibly from standby sources

occupancy, while 0.5 % times 2/3 is prescribed to ensure operation for civil protection. Moreover, with regard to existing buildings, in the commentary to the Italian code (CM 617 2009), both limits are reduced if the analysis model includes the infills. In such a case, the limits become the same as for the masonry buildings, which are equal to 0.3 % and 0.2 % to ensure occupancy and operation, respectively.

Future Challenges and Concluding Remarks

The overview of the seismic assessment framework for RC building provided in this entry has been organized according to the most recent state of the art and guidelines that provide a vision for next-generation codes for RC structures. On the other hand, there are still different aspects on which research and, in turn, code standards should focus.

In particular, behavior of RC elements and their modeling still requires significant experimental and numerical efforts. Behavior of

non-ductile RC elements (e.g., characterization of axial load failure) should be better characterized on the basis of higher number of experimental tests in order to better characterize both modeling issues and damage analysis issues in terms of median and logarithmic standard deviation. Same considerations can be made on beam-column joints and masonry infills.

Finally, masonry infills represent a relevant issue for all countries in the Mediterranean area in which this kind of construction practice is very common. Despite the huge efforts made in the last decades, the characterization of nonstructural component modeling and damage is still challenging considering that they are not subjected to the same design and control process to which new RC elements are subjected to.

Performance evaluation of RC structure according to the methodological framework of FEMA P-58-1 (2012) provides a significant enhancement toward a more reliable performance evaluation for stakeholders and decision makers. Still, some procedures need to be codified in simpler and more user-friendly tools to be employed by practitioners and professional

engineers in any structural performance evaluations as current and consolidated practice.

Summary

The framework in which modern design and assessment of reinforced concrete structures are placed is the well-known Performance-Based Earthquake Engineering (PBEE) framework. The most recent methodological organization of PBEE tools is represented by FEMA P-58-1 document. The brief analysis of the main steps of PBEE provides the basis to introduce where, in such framework, the specific structural material – in this case RC – plays a role. In particular, when the relationship between intensity measure and engineering demand parameter arises in PBEE, it is the phase in which it is necessary to focus on the behavior of RC structural elements, in order to approach analytical modeling, and quantify components damage measures.

Seismic vulnerability assessment has to account for all the typical modes of failure that RC elements can show. Thus, behavior of RC elements (e.g., beams, columns, beam–column joints) is described considering the case of existing elements representing the most general situation for vulnerability assessment problems. The behavior of masonry infills is also considered, given the significant effect they can have on vulnerability in terms of strength and stiffness increase, and in terms of occurrence of brittle failures caused by local interaction of infill and RC frame. Critical review of behavior of RC structures and, in turn, of RC elements, allows emphasizing future challenges and needs that the scientific community and practitioners should take on in the next years.

Cross-References

- [Analytic Fragility and Limit States \[P\(EDP|IM\)\]: Nonlinear Dynamic Procedures](#)
- [Analytic Fragility and Limit States \[P\(EDP|IM\)\]: Nonlinear Static Procedures](#)

- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- [Seismic Collapse Assessment](#)
- [Seismic Reliability Assessment, Alternative Methods for](#)
- [Seismic Risk Assessment, Cascading Effects](#)
- [Site Response for Seismic Hazard Assessment](#)

References

- ACI 352R-02 (2002) Recommendations for design of beam-column connections in monolithic reinforced concrete structures. American Concrete Institute, Farmington Hills
- Alath S, Kunnath SK (1995) Modeling inelastic shear deformations in rc beam-column joints. In: Engineering mechanics proceedings of 10th conference, University of Colorado at Boulder, Boulder, 21–24 May, vol 2. ASCE, New York, pp 822–825
- Alire DA (2002) Seismic evaluation of existing unconfined RC beam–column joints. MSc thesis, University of Washington
- Altoontash A (2004) Simulation and damage models for performance assessment of reinforced concrete beam-column joints. PhD dissertation, Department of Civil and Environment Engineering, Stanford University, Stanford
- American Society of Civil Engineers (ASCE) (2007) Seismic rehabilitation of existing buildings, ASCE/SEI 41-06. American Society of Civil Engineers, Reston, VA
- Anderson M, Lehman D, Stanton J (2008) A cyclic shear stress-strain model for joints without transverse reinforcement. *Eng Struct* 30:941–954
- ATC Applied Technology Council (1985) ATC-13, earthquake damage evaluation data for California. Applied Technology Council, Redwood City, CA, pp 492
- Aschheim AM, Moehle JP (1992) Shear strength and deformability of RC bridge columns subjected to inelastic displacements. Technical Report No UCB/EERC 92/04, University of California at Berkeley, Berkeley
- Aslani H (2005) Probabilistic earthquake loss estimation and loss disaggregation in buildings. PhD thesis, Stanford University
- Asteris PG, Kakaletsis DJ, Chrysostomou CZ, Smyrou EE (2011) Failure modes in infilled frames. *Electron J Struct Eng* 11(1):11–20
- Bakir PG, Boduroğlu HM (2002) A new design equation for predicting the joint shear strength of monotonically loaded exterior beam-column joints. *Eng Struct* 24:1105–1117
- Beck JL, Porter KA, Shaikhutdinov R, Au SK, Moroi T, Tsukada Y, Masuda M (2002) Impact of seismic risk on lifetime property values, final report for CUREE-Kajima phase IV project. Consortium of Universities for Research in Earthquake Engineering, Richmond

- Berry M, Parrish M, Eberhard M (2004) PEER structural performance database user's manual. Pacific Engineering Research Center, University of California, Berkeley, 38 pp. Available at <http://nisee.berkeley.edu/spd/> and <http://maximus.ce.washington.edu/~peera1/>. Accessed 23 Feb 2014
- Bertoldi SH, Decanini LD, Gavarini C (1993) Telai tamponati soggetti ad azioni sismiche, un modello semplificato: confronto sperimentale e numerico. In: *Proceeding of 6° Convegno Nazionale ANIDIS*, vol 2, 13–15 Perugia, Italy, 815–824 (in Italian)
- Biddah A, Ghobarah A (1999) Modelling of shear deformation and bond slip in reinforced concrete joints. *Struct Eng Mech* 7(4):413–432
- Biskinis D, Fardis MN (2010a) Deformations at flexural yielding of members with continuous or lap-spliced bars. *Struct Concr* 11(3):127–138
- Biskinis D, Fardis MN (2010b) Flexure-controlled ultimate deformations of members with continuous or lap-spliced bars. *Struct Concr* 11(2):93–108
- Biskinis DE, Roupakias GK, Fardis MN (2004) Degradation of shear strength of reinforced concrete members with inelastic cyclic displacement. *ACI Struct J* 101(6):773–783
- Bousias SN (1993) Experimental and analytical study of RC columns in cyclic biaxial bending with axial force. Doctoral thesis, Civil Engineering Department, University of Patras, Patras
- Bousias SN, Panagiotakos TB, Fardis MN (2002) Modeling of RC members under cyclic biaxial flexure and axial force. *J Earthq Eng* 6(3):213–238
- Calvi GM, Bolognini D, Penna A (2004) Seismic performance of masonry-infilled RC frames—benefits of slight reinforcements. Invited lecture to “Sismica 2004 – 6° Congresso Nacional de Sismologia e Engenharia Sísmica”, Guimarães, 14–16 Apr
- Calvi GM, Pinho R, Magenes G, Bommer JJ, Restrepo-Vélez LF, Crowley H (2006) Development of seismic vulnerability assessment methodologies over the past 30 years. *ISST J Earthq Technol Paper No 472* 43(3):75–104
- CEB (1996) RC frames under cyclic loading. T. Telford, London, for Comité Euro-international du beton, Lausanne
- Celik OC, Ellingwood BR (2008) Modeling beam–column joints in fragility assessment of gravity load designed reinforced concrete frames. *J Earthq Eng* 12(3):357–381
- CEN (2004) EN 1998-1 Eurocode 8: design of structures for earthquake resistance-part 1: general rules, seismic actions and rules for buildings. European Committee for Standardization, Brussels
- CEN (2005) Eurocode 8: design of structures for earthquake resistance – part 3: assessment and retrofitting of buildings. European Standard EN 1998-1:2005-Comité Européen de Normalisation, Brussels
- Chrysostomou CZ, Asteris PG (2012) On the in-plane properties and capacities of infilled frames. *Eng Struct* 41:385–402
- CM 617 (2009) Istruzioni per l'applicazione delle nuove norme tecniche per le costruzioni. Ministero delle Infrastrutture e dei Trasporti, Rome (in Italian)
- Coburn A, Spence R (2002) Earthquake protection, 2nd edn. Wiley, Chichester
- Colangelo F (2003) Experimental evaluation of member-by-member models and damage indices for infilled frames. *J Earthq Eng* 7(1):25–50
- Colangelo F (2013) Drift-sensitive non-structural damage to masonry infilled reinforced concrete frames designed to Eurocode 8. *Bull Earthq Eng*. doi:10.1007/s10518-013-9503-y
- Cornell A, Krawinkler H (2000) Progress and challenges in seismic performance assessment. *PEER News*, Apr
- Crisafulli FG (1997) Seismic behaviour of reinforced concrete structures with masonry infills. PhD thesis, University of Canterbury, Christchurch
- De Luca F, Verderame GM (2013) A practice-oriented approach for the assessment of brittle failures in existing RC elements. *Eng Struct* 48:373–388
- De Luca F, Ameri G, Iervolino I, Bindi D (2014a) Toward validation of simulated accelerograms via prediction equations for nonlinear SDOF response. *Boll Geofis Teor Appl* 55(1):81–101
- De Luca F, Verderame GM, Gómez-Martínez F, Pérez-García A (2014b) The structural role played by masonry infills on RC building performances after the 2011 Lorca, Spain, earthquake. *Bull Earthq Eng* 12(5):1999–2026
- Deierlein GG, Krawinkler H, Cornell CA (2003) A framework for performance-based earthquake engineering. In: *Proceedings of 2003 pacific earthquake conference on earthquake engineering*, University of Canterbury, Christchurch
- Deierlein GG, Reinhorn AM, Willford MR (2010) Nonlinear structural analysis for seismic design, NEHRP Seismic Design Technical Brief No. 4, produced by the NEHRP Consultants Joint Venture, a partnership of the Applied Technology Council and the Consortium of Universities for Research in Earthquake Engineering, for the National Institute of Standards and Technology, Gaithersburg, NIST GCR 10-917-5
- Di Ludovico M, Verderame G, Prota A, Manfredi G, Cosenza E (2013) Experimental behavior of nonconforming RC columns with plain bars under constant axial load and biaxial bending. *J Struct Eng* 139(6):897–914
- DM 14/1/2008(2008) Norme tecniche per le costruzioni. Ministero delle Infrastrutture e dei Trasporti, Rome (in Italian)
- Dolšek M, Fajfar P (2001) Soft storey effects in uniformly infilled reinforced concrete frames. *J Earthq Eng* 5(1):1–12
- Elwood K (2004) Modelling failures in existing reinforced concrete columns. *Can J Civ Eng* 31:846–859
- Elwood KJ, Eberhard M (2009) Effective stiffness of reinforced concrete columns. *ACI Struct J* 106(4): 476–484

- Elwood K, Moehle JP (2005a) Drift capacity of reinforced concrete columns with light transverse reinforcement. *Earthq Spectra* 21(1):71–89
- Elwood K, Moehle JP (2005b) Axial capacity model for shear-damaged columns. *ACI Struct J* 106(S45):578–587
- Elwood KJ, Matamoros AB, Wallace JW, Lehman DE, Heintz JA, Mitchell AD, Moore MA, Valley MT, Lowes LN, Comartin CD, Moehle JP (2007) Update to ASCE/SEI 41 concrete provisions. *Earthq Spectra* 23(3):493–523
- Fardis MN (2009) Seismic design, assessment and retrofitting of concrete buildings based on EN-Eurocode 8. Springer, Dordrecht
- Fardis MN, Panagiotakos TB (1997) Seismic design and response of bare and infilled reinforced concrete buildings – part II: infilled structures. *J Earthq Eng* 1(3):473–503
- Fardis MN et al (1992) Damage measures and failure criteria during seismic response. In: Cooperative research program on the seismic response of reinforced concrete structures. First phase, Final report, II, National Laboratory for Civil Engineering, Lisbon
- Federal Emergency Management Agency (FEMA) (1997) FEMA 273, guidelines for the seismic rehabilitation. Federal Emergency Management Agency, Washington, DC
- Federal Emergency Management Agency (FEMA) (2000) FEMA 356, prestandard and commentary for the seismic rehabilitation of buildings. Federal Emergency Management Agency, Washington, DC
- Fédération internationale du béton (FIB). Task Group 7.1 (2003) Seismic assessment and retrofit of reinforced concrete buildings: state-of-art report (vol 24). International Federation for Structural Concrete (ed). FIB-Féd. Int. du Béton
- Federal Emergency Management Agency (FEMA) P-58-1 (2012) Seismic performance assessment of buildings volume 1 – 558 methodology. Federal Emergency Management Agency, Washington, DC
- Goulet CA, Haselton CB, Mitrani-Reiser J, Beck JL, Deierlein GG, Porter KA, Stewart JP (2007) Evaluation of the seismic performance of a code-conforming reinforced-concrete frame building – from seismic hazard to collapse safety and economic losses. *Earthq Eng Struct Dyn* 36(13):1973–1997
- Grunthal G (ed) (1998) European Macroseismic Scale 1998 (EMS-98), Cahiers du Centre Européen de Géodynamique et de Séismologie, 15, Luxembourg
- Gu X, Lu Y (2005) A fuzzy-random analysis model for seismic performance of framed structures incorporating structural and non-structural damage. *Earthq Eng Struct Dyn* 34(10):1305–1321
- Hakuto S, Park R, Tanaka H (2000) Seismic load tests on interior and exterior beam-column joints with substandard reinforcing details. *ACI Struct J* 97(1):11–25
- Haselton CB, Liel AB, Taylor Lange S, Deierlein GG (2008) Beam-column element model calibrated for predicting flexural response leading to global collapse of RC frame buildings, PEER report 2007/03
- Hassan WM (2011) Analytical and experimental assessment of seismic vulnerability of beam-column joints without transverse reinforcement in concrete buildings. PhD dissertation, University of California, Berkeley
- Hwang S, Lee H (1999) Analytical model for predicting shear strengths of exterior reinforced concrete beam-column joints for seismic resistance. *ACI Struct J* 96(5):846–858
- Ibarra LF, Medina RA, Krawinkler H (2005) Hysteretic models that incorporate strength and stiffness deterioration. *Earthq Eng Struct Dyn* 34:1489–1511
- Kim J, LaFave JM (2007) Key influence parameters for the joint shear behavior of reinforced concrete (RC) beam-column connections. *Eng Struct* 29:2523–2539
- Krawinkler H (2002) A general approach to seismic performance assessment. In: Proceedings of international conference on advances and new challenges in earthquake engineering research, ICANCEER, Hong Kong, 19–20 Aug
- Krawinkler H, Miranda E (2004) Performance-based earthquake engineering, chapter 9. In: Bozorgnia Y, Bertero VV (eds) *Earthquake engineering: from engineering seismology to performance-based engineering*. CRC Press, Boca Raton
- LaFave M, Shin M (2005) Discussion of ‘Modeling reinforced-concrete beam-column joints subjected to cyclic loading’, by Lowes, L. N. and Altoontash, A. *ASCE J Struct Eng* 131(6):992–993
- Legeron F, Paultre P (2000) Behavior of high-strength concrete columns under cyclic flexure and constant axial load. *ACI Struct J* 97(4):591–601
- Lowes LN, Altoontash A (2003) Modeling reinforced-concrete beam-column joints subjected to cyclic loading. *ASCE J Struct Eng* 129(12):1686–1697
- Manfredi G, Prota A, Verderame GM, Ricci P, De Luca F (2014) 2012 Emilia earthquake, Italy: reinforced concrete buildings response. *Bull Earthq Eng* 12(5):2275–2298
- McGuire RK (2004) Seismic hazard and risk analysis. Report MNO-10. Earthquake Engineering Research Institute Publication, Oakland
- Mehrabi AB, Shing PB, Schuller MP, Noland JL (1996) Experimental evaluation of masonry-infilled RC frames. *J Struct Eng ASCE* 122(3):228–237
- Mitrani-Reiser J (2007) An ounce of prevention: probabilistic loss estimation for performance-based earthquake engineering. PhD dissertation, California Institute of Technology (Caltech), Pasadena
- Moehle JP (2003) A framework for performance-based earthquake engineering. In: Proceedings of ATC-15-9 workshop on the improvement of building structural design and construction practices, Maui, June
- Moehle JP, Deierlein GG (2004) A framework for performance-based earthquake engineering.

- In: Proceedings of 13th world conference on earthquake engineering, Paper No 679, Vancouver
- Mosalam KM (1996) Experimental and computational strategies for the seismic behavior evaluation of frames with infill walls. PhD dissertations, Cornell University, Ithaca
- National Institute of Building Science (NIBS) (1997, 1999 and 2002) Earthquake loss estimation methodology. HAZUS. Technical manuals, vols 1, 2, 3. Federal Emergency Management Agency (FEMA), Washington, DC. <http://www.fema.gov/hazus/>
- Özcebe S, Crowley H, Bal IE (2012) Distinction between no and slight damage states for existing RC buildings using a displacement-based approach. In: Proceedings of the 15th world conference on earthquake engineering, paper 5126, 24–28, Lisbon, Portugal
- Panagiotakos TB, Fardis MN (2001) Deformations of RC members at yielding and ultimate. *ACI Struct J* 98(2):135–148
- Pantelides CP, Hansen J, Nadauld J, Reaveley LD (2002) Assessment of reinforced concrete building exterior joints with substandard details, PEER 2002/18, May 2002
- Park R (2002) A summary of results of simulated seismic load tests on RC beam–column joints, beam and columns with substandard reinforcing details. *J Earthq Eng* 6(2):147–174
- Park YJ, Ang AMS (1985) Mechanistic seismic damage model of reinforced concrete. *ASCE J Struct Eng* 111:722–739
- Park S, Mosalam KM (2012) Parameters for shear strength prediction of exterior beam–column joints without transverse reinforcement. *Eng Struct* 36:198–209
- Park S, Mosalam KM (2013) Experimental investigation of nonductile RC corner beam–column joints with floor slabs. *ASCE J Struct Eng* 139(1):1–14
- Porter KA (2003) An overview of PEER's performance-based earthquake engineering methodology. In: Conference on Applications of Statistics and Probability in Civil Engineering (ICASP9), Civil Engineering Risk and Reliability Association (CERRA), San Francisco, 6–9 July
- Priestley MJN, Verma R, Xiao Y (1994) Seismic shear strength of reinforced concrete columns. *J Struct Eng* 120(8):2310–2329
- Pujol S (2002) Drift capacity of reinforced concrete columns subjected to displacement reversals. PhD thesis, Purdue University
- Ricci P, De Luca F, Verderame GM (2011) 6th April 2009 L'Aquila earthquake, Italy: reinforced concrete building performance. *Bull Earthq Eng* 9(1):285–305
- Ricci P, Verderame G, Manfredi G (2013) ASCE/SEI 41 provisions on deformation capacity of older-type reinforced concrete columns with plain bars. *ASCE J Struct Eng* 139(12):04013014
- Sezen H, Moehle JP (2004) Shear strength model for lightly reinforced concrete columns. *ASCE J Struct Eng* 130(11):1692–1703
- Shin M, LaFave JM (2004) Testing and modeling for cyclic joint shear deformations in rc beam–column connections. In: Proceedings of the thirteenth world conference on earthquake engineering, Vancouver, 1–6 Aug 2004, Paper No 0301
- Shing PB, Mehrabi AB (2002) Behaviour and analysis of masonry-infilled frames. *Prog Struct Eng Mater* 4:320–331
- Tothong P, Cornell CA (2006) An empirical ground motion attenuation relation for inelastic spectral displacement. *Bull Seismol Soc Am* 96: 2146–2164
- Tothong P, Luco N (2007) Probabilistic seismic demand analysis using advanced ground motion intensity measures. *Earthq Eng Struct Dyn* 36:1837–1860
- Vecchio FJ, Collins MP (1986) The modified compression-field theory of reinforced concrete elements subjected to shear. *ACI Struct J* 83(2): 219–231
- Verderame GM, Ricci P, Manfredi G, Cosenza E (2010) Ultimate chord rotation of RC columns with smooth bars: some considerations about EC8 prescriptions. *Bull Earthq Eng* 8(6):1351–1373
- Verderame GM, De Luca F, Ricci P, Manfredi G (2011) Preliminary analysis of a soft-storey mechanism after the 2009 L'Aquila earthquake. *Earthq Eng Struct Dyn* 40(8):925–944
- Vollum RL (1998) Design and analysis of exterior beam column connections. PhD dissertation, Imperial College of Science Technology and Medicine–University of London
- Vollum RL, Newman JB (1999) Strut and tie models for the analysis/design of external beam–column joints. *Mag Concr Res* 51(6):415–425
- Walker SG (2001) Seismic performance of existing RC beam–column joints. MSc thesis, University of Washington
- Watanabe F, Ichinose T (1992) Strength and ductility of RC members subjected to combined bending and shear. In: Concrete shear in earthquake. Elsevier Applied Science, New York, pp 429–438
- Williams MS, Sexsmith RG (1997) Seismic assessment of concrete bridges using inelastic damage analysis. *Eng Struct* 19(3):208–216
- Wong HF (2005) Shear strength and seismic performance of non-seismically designed reinforced concrete beam–column joints. PhD dissertation, Department of Civil Engineering, The Hong Kong University of Science and Technology, August
- Yavari S, Elwood KJ, Wu C (2009) Collapse of a nonductile concrete frame: evaluation of analytical models. *Earthq Eng Struct Dyn* 38(2):225–241
- Youssef M, Ghobarah A (2001) Modelling of RC beam–column joints and structural walls. *J Earthq Eng* 5(1):93–111
- Zhu L, Elwood K, Haukaas T (2007) Classification and seismic safety evaluation of existing reinforced concrete columns. *J Struct Eng* 133(9):1316–1330

Seismometer Arrays

Johannes Schweitzer
NORSAR, Kjeller, Norway

Synonyms

Arrays; Beamforming; fk-analysis; Seismic arrays

Introduction

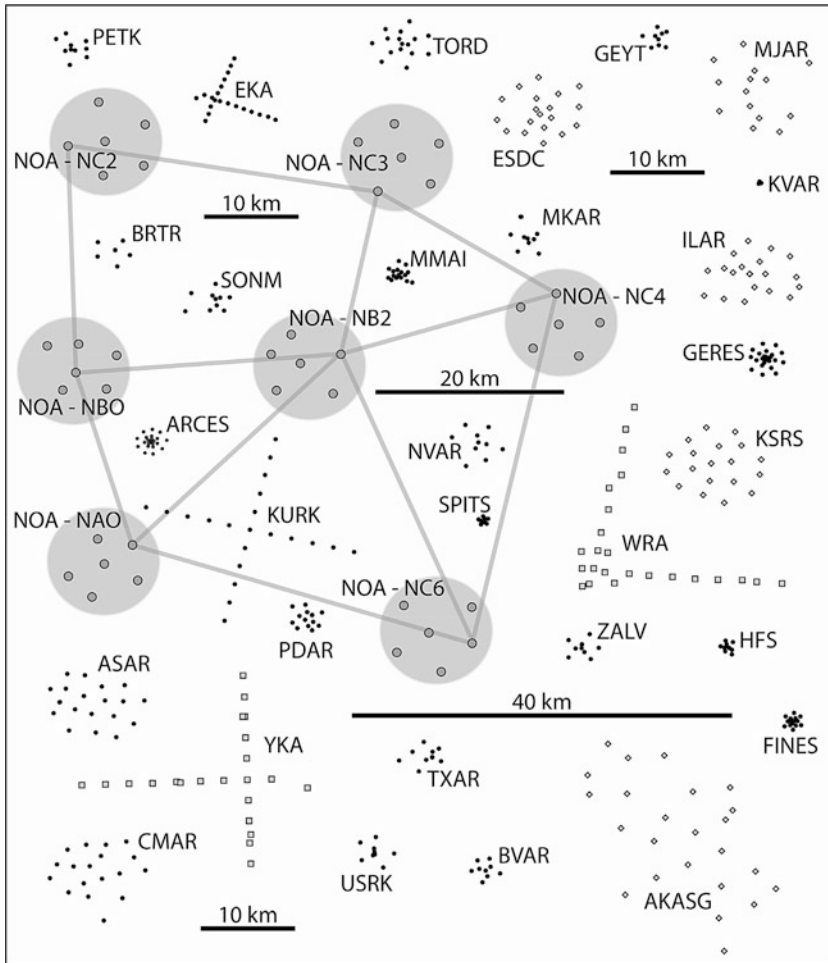
Definition and Purpose

In seismology, the term “array” (of seismometers) has been known for more than 50 years. Today this phrase is often used for any group of seismic stations in the sense as defined in Macmillan’s online dictionary, where an array is described as “a number of pieces of equipment of the same type, connected together to do a particular job” (Macmillan 2013). In this chapter, the term “seismometer arrays” is used with a more restricted definition, as it was originally introduced to seismology. A seismometer array can be defined as (Schweitzer et al. 2012): “A seismic array is a set of seismometers deployed so that characteristics of the seismic wavefield at a specified reference point, within or close to the array, can be inferred by analyzing the waveforms recorded at the different sites. A seismic array differs from a local network of seismic stations mainly by the techniques used for data analysis. Thus, in principle, a network of seismic stations can be used as an array, and data from an array can be analyzed as data from a network. The size of an array is defined by its aperture, which is the largest horizontal distance between two sensors of the array. In practice, the geometry and the number of seismometer sites of an array are determined by the intended scientific purpose and economic limits.” The main purpose to install a seismic array is threefold. At first, an array can be steered as an antenna to amplify the signals of interest by stacking (summing) the seismic wave energy recorded at

the different array sites after applying appropriate phase delays. With these so-called “beamforming” techniques, arrays show superior signal detection capabilities with respect to single 3-component (3C) seismic stations. The second main purpose to install a seismic array is the capability to estimate the station-to-seismic source azimuth (backazimuth, BAZ) and the apparent velocity of unknown seismic signals crossing the array. Thirdly, a single seismic array often provides enough information about the incoming seismic signals that an automatic algorithm can be used to estimate a first (preliminary) solution of the seismic source location.

History of Seismometer Arrays

The history of seismometer array installations is going back to the 1950s. At that time, the idea of installing arrays of sensors to improve the signal-to-noise ratio (SNR) of a seismic onset was adopted from radio astronomy, radar, acoustics, and sonar. Since then, classified arrays deployed to monitor nuclear test activities at teleseismic distances have been built worldwide. In 1958, “The Conference of Experts to study the methods of detecting violations of a possible agreement on the suspension of nuclear tests” was held under the auspices of the United Nations in Geneva. This conference was followed up by several initiatives for improving the quality of seismic stations worldwide. Many of these classified arrays became known in the 1990s and are today part of the International Monitoring System (IMS) for the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO) (see e.g., Douglas 2002; Dahlman et al. 2009) as primary or auxiliary stations, e.g., AKASG (Malin, Ukraine), ASAR (Alice Springs, Australia), BRTR (Keskin, Turkey), CMAR (Chiang Mai, Thailand), ESDC (Sonseca, Spain), ILAR (Eielson, Alaska), and KURK (Kurchatov, Kazakhstan). Many of these arrays have quite diverse geometries (Fig. 1) and in some cases comprise different installations for short-period (frequencies above 0.5 Hz) and long-period (frequencies below 0.1 Hz) signals as, e.g., the Belbaşı array nearby the Keskin array. Similar to arrays of seismometers, the



Seismometer Arrays, Fig. 1 Seismic arrays equipped with short-period or broadband seismic sensors in operation in December 2013 as part of the International Monitoring System (IMS) for the Comprehensive Nuclear-Test-Ban Treaty Organization (CTBTO). The code name

with which it is registered in the international registry of seismic stations at the International Seismological Centre (<http://www.isc.ac.uk/registries/>) is provided for each array. All array maps are plotted in the same scale (Courtesy of S. J. Gibbons, NORSAR)

IMS includes also arrays of infrasound sensors and hydrophones, and similar data analysis techniques are applied as for data from seismic arrays.

To the best of knowledge, the first experimental seismic array with more than four elements and openly available data was established in February 1961 by the United Kingdom Atomic Energy Agency (UKAEA) on the Salisbury Plain (UK), followed in December 1961 by the Pole Mountain array (PMA, Wyoming, USA), in June 1962 by the Eskdalemuir array (EKA, Scotland, UK), and in December 1963 by the Yellowknife array (YKA, Canada). This array type

(known as UK array) is orthogonal linear or L-shaped, with apertures between 10 and 25 km. Later, arrays of the same type were built in Australia (Warramunga), Brazil (Brasilia), and India (Gauribidanur).

In the 1960s, arrays with very different apertures and geometries were tested, from small circular ones with apertures of a few kilometers to huge arrays with apertures of up to 200 km. The largest arrays were the Large Aperture Seismic Array (LASA) in Montana (USA), opened in 1965 and in operation until 1978 with 525 seismometer sites, and the original Norwegian

Seismic Array (NORSAR) in southern Norway, consisting of 132 sites over an aperture of approximately 100 km with altogether 198 seismometers, which became fully operational in spring 1971. The original NORSAR array was reduced in 1976 to seven subarrays and was assigned the new code name NOA.

LASA, NORSAR, and the UKAEA arrays had narrowband short-period seismometers (for signal frequencies around 1 Hz) at all sites and additional long-period seismometers (for signal periods around 20 s) at selected sites in their original configurations. In Germany, a new array type was planned and installed in the 1970s with an aperture of about 110 km. The Gräfenberg-Array (GRF) was installed on the limestone plateau of the Franconian Jura as the world's first seismometer array equipped entirely with broadband sensors (for frequencies between 0.01 and 8 Hz). Since then, the short-period and long-period sensors of many arrays were or will be in the near future exchanged with broadband seismometers.

In the 1980s, the geometry of the so-called regional arrays was developed. This is often called a NORES-type array design and has array sites located on concentric rings (each with an odd number of sites) spaced at log-periodic intervals. It is now used for the design of most modern, small aperture arrays; only the number of rings and the aperture (diameter of the outermost ring of sites) differ from installation to installation.

Another approach to seismometer arrays was developed in the 1990s. In parts of the world, the networks of seismometer stations became so dense that data from these single station networks could be combined and analyzed as data from a seismic array. Examples are the J-array in Japan, the German Regional Seismic Network, the Californian array, the Kyrgyz Network (K-Net), and the ongoing USArray project with semi-temporary stations. Most of the known array techniques could be applied to analyze data from these and other networks.

Array installations demonstrated during the last decades of the twentieth century that seismic arrays could facilitate detection and

characterization of seismic signals that was superior to that of single three-component (3C) stations. Today, many of the seismic stations of the IMS for the CTBTO are arrays (e.g., Dahlman et al. 2009).

During the last two decades, temporary, very small aperture arrays (apertures usually below 1 km) were used to investigate the distribution of (mostly) S-wave velocities below and nearby the array installation. Knowing S-wave velocities of the uppermost layers is essential for seismic hazard mitigation; for details see, (e.g., Schweitzer et al. 2012).

Further details about different array configurations can be found in the literature (e.g., Barber 1958; Haubrich 1968; Harjes and Henger 1973; Mykkeltveit et al. 1983). As examples of seismic arrays, Fig. 1 shows maps of all primary and auxiliary IMS arrays in operation in December 2013; note the huge variety in geometries, number of sites, and apertures.

Basics of Array Seismology

Basic Requirements for Seismometer Arrays

The observed apparent-velocity range and the dominant frequency content of seismic signals are quite different for different types of signals (i.e., local, regional, or teleseismic observations, body waves, or surface waves). Thus, the geometry, aperture, and instrumentation of seismic arrays have to be adjusted with respect to the scientific requirements of an array installation.

Proper analysis of array data depends on a stable, high-precision relative timing of all array elements. This is required because most of the parametric information calculated using an array involves the measurement of (usually very small) time differences (phase shifts) between the seismic signals recorded at the different sensors. Most array-data analysis algorithms assume that seismic energy crosses an array as a plane wave. From the theory of signal processing, it is known that signals can be constructively summed up as long as the time shift between the signals is not larger than about $\frac{1}{4}$ of the dominant signal period T . This rule applies also to the case of

seismic signals observed with an array. Many different effects can influence the arrival time of seismic onsets and thereby disturb the plane-wave approximation and should be taken in account when analyzing array data:

- Seismic waves usually propagate with spherical (in the case of body waves) or circular geometry (in the case of surface waves). Therefore, the aperture of an array should be small enough that the plane-wave approximation is still valid, i.e., the (theoretical) arrival-time differences with respect to the plane-wave approximation should be so small that they do not influence the analysis results.
- Since the fundamental work of Hans Benndorf (1870–1953) and Emil Wiechert (1861–1921) during the first decade of the twentieth century, it is known that in the case of a body wave, there is a direct relation between the observed (apparent) propagation velocity of a seismic wave, the ray parameter (the derivative of the travel-time curve), and the seismic velocity at the turning point of the observed seismic wave within the Earth. With an array of seismometers, this (apparent) velocity can be directly measured as long as all sensors are located in one horizontal plane.

In the case that the sensors are not located on one horizontal plane, the theoretical time shifts between the array sensors do not only depend on the propagation direction and (apparent) velocity of the plane wave but also on the wavelength-dependent seismic velocities below the array sites. If travel-time effects due to array topography become larger than about one fourth of the signal period, it becomes important for the accuracy of data analysis results to take this effect in account. However, analysis algorithms then become rather complicated, and if one plans for a new array, one should try to avoid this additional complexity and locate all seismometer sites on one horizontal plane.

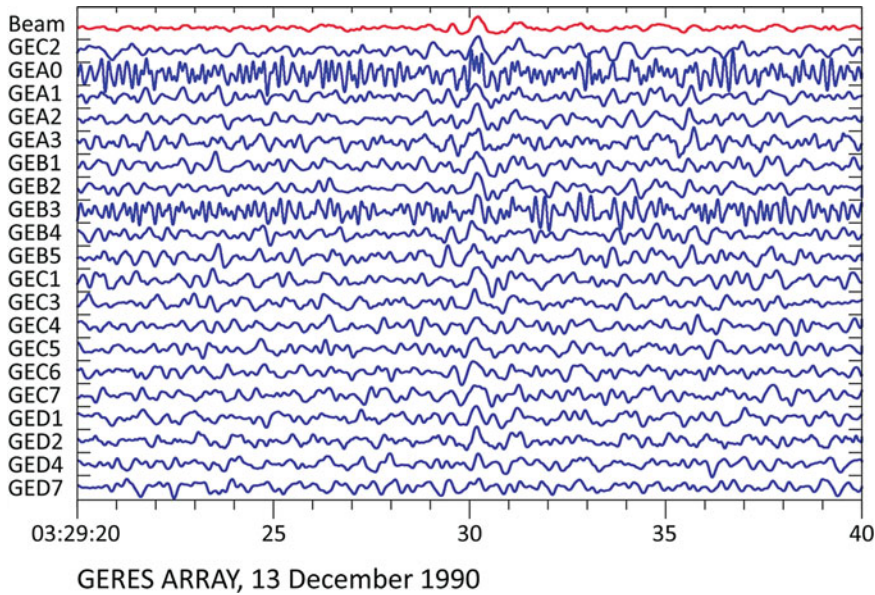
- The seismic wavefield is always influenced by lateral heterogeneities in the Earth, which can disturb the plane-wave approximation due to velocity inhomogeneities. In particular local

lateral heterogeneities can have large effects. To avoid this, many arrays are built on homogeneous geological units, but this is not everywhere possible, and uniform geology at the Earth's surface does not guarantee that deeper structures are similarly homogeneous. Therefore, it is necessary to investigate recorded data from each seismic array for systematic effects which may disturb the plane-wave approximation.

- There are many examples of how timing-system errors at single array sites can produce difficulties for array-data analysis. However, stable and correct relative timing can be achieved by installing at the array a central clock, which distributes a common time signal to all digitizers at all array sites. Even in the case that this central timing system fails and has some offset to absolute time, all array-analysis algorithms can still be applied. So, whenever it is possible, a central timing system should be considered for new array installations.

The previously mentioned “beamforming” techniques delay (phase shift) and stack the signals from the different sensors. Due to constructive interference of the signals, the SNR is then enhanced, whereas the (random and uncorrelated) background noise is suppressed. One can show that a seismic array can theoretically improve the SNR by a factor of $\sqrt{M}$, where M is the number of array elements. Figure 2 shows an example from the GERES array in Germany. The tiny onset of a P phase, which was reflected at the Earth's core, was recorded with vertical sensors at 20 sites of the array. Figure 2 shows these data in blue and on top the array beam in red. On many single traces the signal is hardly discernible, but it is clearly visible on the beam trace due to the drastically improved SNR. To achieve the theoretical factor of $\sqrt{M}$, most array-data processing techniques require not only high clock stability of the recording systems at the different array sites and best knowledge about possible deviations from the plane-wave approach but also high signal coherency across the array.

Waveforms of seismic signals can be influenced by interference with wave energy



Seismometer Arrays, Fig. 2 Vertical seismograms (blue) with the onset of a tiny P onset reflected from the Earth's core (PcP) and recorded at 03:29:29.5 on 13 December 1990 with the GERES array after an earthquake in the Tyrrhenian Sea at an epicentral distance of

about 1,070 km. The top trace (red) shows the array beam of the shown vertical records after applying “delay-and-sum” processing. All data are Butterworth band-pass filtered between 0.8 and 4 Hz and equally scaled

scattered at heterogeneities along the entire ray path, and as a result, waveforms observed at different seismometer sites may significantly differ. Since high signal coherency is required for seismic arrays, such waveform-altering effects should be minimized by avoiding array locations close to known lateral inhomogeneous structures. This constitutes an additional reason to prefer installing seismic arrays at one relatively homogeneous geological unit, almost transparent to seismic waves. Since such local site effects are signal frequency dependent, signal coherency over the array also becomes frequency dependent, and this may impose constraints on array geometry, spatial extent, and instrumentation.

As mentioned earlier, seismometer arrays can provide estimates of the station-to-event azimuth (backazimuth, BAZ) and the apparent velocity of seismic signals. These estimates are important both for event location purposes and for signal identification and classification, e.g., as P, S, local, regional, or teleseismic phases. The slowness resolution of an array – i.e., how accurately

the propagation direction and the apparent velocity of a wave front can be measured – improves with increasing array aperture. However, the apparent velocity of a seismic onset changes with epicentral distance and the shape of a seismic waveform can change drastically due to interference with other seismic phases; in addition, signal coherency diminishes with increasing sensor separation. Therefore, when building an array, one has to find a balance between coherency and theoretical slowness resolution.

Finally, some remarks about the instrumentation of arrays. Traditionally, arrays were equipped with vertical sensors at all sites and only a few sites with additional horizontal sensors. This was mostly related to the additional costs for more sensors, data transmission, and data storage. However, many studies have shown that arrays equipped with more 3C sensors are superior in analyzing S-type onsets. Since costs for data transmission and storage have drastically dropped, one should always consider installing 3C instruments at all sites for new array installations.

The Array-Transfer Function

In signal-processing theory, the process of steering a seismic array to a specific target, i.e., optimizing the array beam for a given BAZ and apparent velocity of a seismic wavefield, can be described as a linear filter, which allows only signals with these characteristics to pass. In the case of an array, it is a two-dimensional filter, which is defined in the frequency-wavenumber (fk) space. The wavenumber k of a signal is defined as $k = 2\pi f/v$ or $k = 2\pi/\lambda$, where f is the frequency, λ is the wavelength, and v is the apparent velocity of the signal. Filter-transfer functions are mathematical descriptions of the filter characteristics (bandpass filter). The two-dimensional array filter is described by the sharpness of its main lobe (maximum of passing signals) and the position and relative height of eventually existing side maxima (side lobes). The narrower the main lobe and the smaller are eventual side lobes, the better the array's performance as a filter for signals with very specific BAZ and apparent velocity. For more details on the theory of seismic arrays, one may look in, e.g., Capon (1973), Aki and Richards (1980), Buttkus (2000), Johnson and Dudgeon (2002), Schweitzer et al. (2012), and the citations therein. By comparing array-transfer functions, the quality of an array of seismometers as a frequency-wavenumber filter can be discussed. The amount of literature about general criteria used to evaluate array-transfer functions for a given array geometry is huge, and some suggestions for more details are, e.g., in Harjes and Henger (1973) or Schweitzer et al. (2012) and the references therein. However, some general rules about transfer-function characteristics of seismometer arrays can be formulated as follows (from Harjes and Henger 1973; and Schweitzer et al. 2012):

1. The aperture a of an array defines the resolution of the array for small wavenumbers k . The larger the aperture is, the smaller the wavenumbers that can be measured with the array. The upper limit for the longest wavelength λ that can be resolved by array techniques is approximately similar to the aperture of the array. The array behaves like a single station for signals with $\lambda \gg a$.

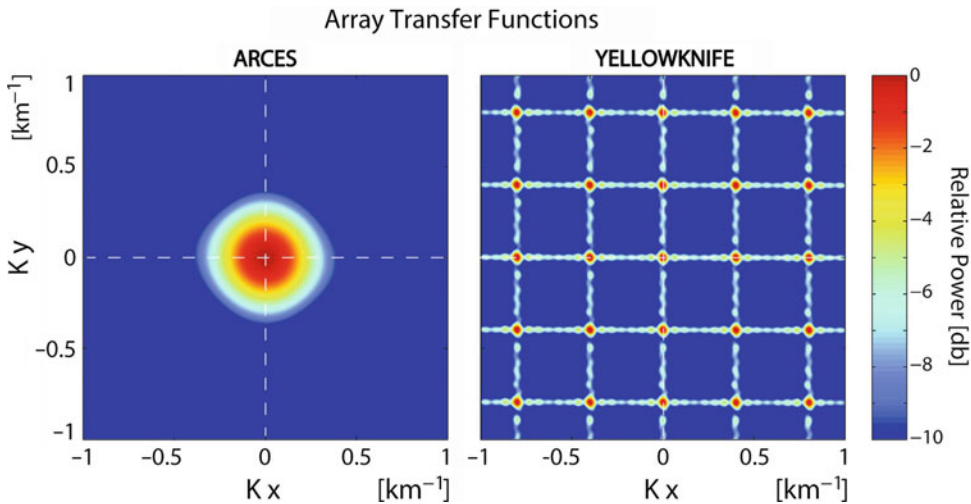
2. The number of sites controls the quality of the array as a wavenumber filter, i.e., its ability to suppress seismic energy crossing simultaneously the array with a different slowness than the one on which the array is steered.
3. The distances between the seismometers define the position of the side lobes of the array-transfer function and the largest resolvable wavenumber: the smaller the mean distance, the smaller the wavelength of a resolvable seismic phase will be (for a given apparent velocity).
4. The geometry of the array defines the azimuth dependence of the aforementioned points.

Figure 3 shows two examples of such array-transfer functions, the ARCES array in northern Norway and the Yellowknife array in northern Canada for a 1 Hz signal. The geometry and aperture of these two arrays is included in Fig. 1. The ARCES shows no differences for signals from different azimuths, and the side lobes of the transfer function are far away (outside the plot) from the main lobe. However, because of the small aperture of only 3 km, this array cannot distinguish between waves with small wavenumber differences, as can be seen in the relatively wide main lobe of the transfer function. In contrast, in the case of Yellowknife, the main lobe is very narrow because of the much larger aperture of the array of about 25 km. This results in a much higher resolution in measuring small apparent-velocity differences. However, the array shows resolution differences for the different azimuths: the many side lobes of the transfer function along north-south and east-west oriented lines are caused by its cross-shaped geometry and the relative large distances between the single array sites.

For further reading, reviews on array theory, together with quite comprehensive citation lists, see, e.g., Douglas (2002), Rost and Thomas (2002, 2009), and Schweitzer et al. (2012).

Data Analysis Algorithms for Seismometer Arrays

One standard analysis technique for array data is the earlier described beamforming algorithm,



Seismometer Arrays, Fig. 3 The array-transfer functions of the circular ARCES array (*left*) and of the cross-shaped Yellowknife array (*right*) for a 1 Hz signal as

relative power (*color coded*) of the array response normalized with its maximum in [db] (Modified from Schweitzer et al. 2012)

also referred to as “delay-and-sum” processing. In seismic prospecting, “beamforming” is called “stacking.” An extension of this concept is the so-called “double-beam” technique, in which not only data of seismometer arrays are stacked but also the array beams of different seismic sources closely located to each other (Krüger et al. 1993).

Another standard array-data analysis technique is frequency-wavenumber analysis (fk-analysis). This method, originally developed for narrowband filtered one-component data, determines the slowness vector (i.e., BAZ and apparent velocity) for a single signal at a single frequency. Fk-analysis can be performed either in the frequency domain or the time domain (see Schweitzer et al. 2012). Fk-analysis applies systematically different phase shifts (or delay times) to a selected time window of the different array-site data and calculates the power of the seismic wave. By searching for the maximum power, the algorithm estimates the corresponding BAZ and apparent velocity of the signal. A variety of fk-analysis algorithms have been developed, which mostly differ in the method applied to find this maximum: e.g., maximum-likelihood method (Levin 1964; Capon et al. 1967, 1973), maximum-entropy method (Burg 1964), multiple

signal classification (MUSIC) algorithm (Schmidt 1986), and various adaptive algorithms for the estimation of spectral power density (e.g., Goldstein and Archuleta 1991). Later, fk-analysis was expanded to wider frequency bands and 3C data (Kværna and Doornbos 1986), the inversion for spherical waveforms (Almendros et al. 1999), and arrays with incoherent data (Gibbons et al. 2008, 2012).

One can also directly measure signal-arrival times at the different array sites and invert the observed travel-time differences for the best fitting plane wave. The (relative) arrival times can be either determined by an analyst or by automatic arrival-time-picking algorithms, as, e.g., by correlation analysis (Cansi 1995).

One widely applied technique to analyze data observed with an array is the VElocity SPectrum Analysis (VESPA) process (Davies et al. 1971), also called *vespagram*. This algorithm is very similar to the time-domain version of the fk-analysis, but it measures the power of seismic signals from a constant BAZ with different apparent velocities as function of time to investigate apparent-velocity changes for the different onsets in a seismogram. Later, the *vespagram* concept was expanded by calculating the observed power from different azimuths for a constant, specific

apparent velocity. In this case, the vespagram is a very useful tool to investigate the BAZ of the seismic phases, their precursors, and in particular the scattered energy of seismic waves arriving after seismic onsets.

Seismological Research with Seismometer Arrays

There are numerous publications from authors using data of seismometer arrays, which cannot be discussed in detail herein. In the following, only some general topics are named to illustrate the diversity of array-seismology applications.

Seismic array data have been used to detect and investigate all kinds of small amplitude phases: seismic onsets observed in front of seismic phases, which had traversed the Earth's core (PKP), were often interpreted as regular seismic phases, which led to many, quite controversial velocity models of the lower part of the outer core of the Earth. Using array observations it could be shown that these precursors are scattered energy from lateral heterogeneities in the Earth's interior and not regular seismic phases. There exist numerous array-data studies of the global and local structure of the lower mantle, the laterally heterogeneous lowermost part of the mantle directly above the core-mantle boundary (known as D'' region), the core-mantle boundary, and the inner-core boundary.

Also strong-motion instruments have been installed in array configurations at the Earth's surface, as well as in boreholes to study near-field effects of earthquakes, and arrays were used to track the aftershock activity after larger earthquakes or to monitor earthquake swarms. In combination with cross-correlation techniques, this became a very powerful tool to lower the detection thresholds for seismic events in monitored areas. Similar is the quite recent approach to track the earthquake slip process itself with array data observed at teleseismic distances. For such studies data from seismic networks are combined to virtual arrays and analyzed.

Seismic arrays have been used to measure dispersion curves of surface-wave velocities.

This technique became recently quite important in context with seismic risk and hazard studies. Since ambient noise mostly consists of surface-wave energy, one can use local dispersion-curve data, measured with temporary, very small aperture arrays, to invert for the near-surface S-velocity structure; for further details see, e.g., Schweitzer et al. (2012).

Seismic arrays are also used to investigate the nature and source regions of microseisms and to locate and track volcanic tremor for analyzing complex seismic wavefield properties in volcanic areas.

The capability of seismometer arrays to locate seismic events was already mentioned at the beginning of this chapter. The data analysis of arrays can be automated for signal detection and fk-analysis of the detected onsets. Then, BAZ observations can be used together with additional travel-time constraints to automatically group the onsets to events. The observed apparent velocities can be used to classify the type of seismic onset, together with other signal characteristics, as, e.g., the dominant frequency. This way, P- and S-type onsets from local and regional events can be automatically identified with the use of small aperture, regional arrays and the associated seismic event can be located. A detailed description of such an automatic regional event location algorithm can be found in (Schweitzer et al. 2012).

Apparent-velocity observations from seismic arrays of at least 10 km aperture can be directly inverted to an epicentral distance for the first arriving P-type onsets from seismic events at teleseismic distances (i.e., from about 22° to about 100° epicentral distance). Knowing the epicentral distance, the observed BAZ can then be used to define the epicentral coordinates. This algorithm cannot be applied to local or regional events because measurable apparent velocities of the P-wave onsets can no longer directly be inverted to epicentral distance. At distances beyond ~90°, these derivatives become again very small for P waves, and in the Earth's shadow zone (distance >100°), the interpretation of the different core-phase onsets is also quite difficult and limits the location capabilities

of a seismic array. The described event location technique has been in use at least since the 1960s, and a quick look in the bulletins of the International Seismological Centre shows the huge amount of reported teleseismic event locations made with, e.g., the Large Aperture Seismic Array (LASA) in Montana, USA, the Yellowknife Array (YKA) in Northern Canada, the Gräfenberg Array (GRF) in Bavaria, Germany, or the large Norwegian Seismic Array (NOA) in southern Norway.

For further reading, more details on the research topics mentioned herein and for many more array applications in seismology, together with quite comprehensive citation lists, see, e.g., Gibbons et al. (2008, 2012), Rost and Thomas (2002, 2009), and Schweitzer et al. (2012).

Summary

A seismometer array is a set of seismic sensors deployed so that characteristics of the seismic wavefield at a specified reference point can be inferred by analyzing together the waveforms recorded at the different array sites. Seismometer arrays have been shown to be superior to single, 3C stations: they are able to lower the detection threshold for seismic onsets due to beamforming; they are able to measure the apparent velocity of seismic onsets and their propagation direction; they can be used to locate seismic events in an automated data processing scheme; and they are very successful tools to investigate the Earth's interior on local, regional, or global scale.

Cross-References

- [Earthquake Location](#)
- [Earthquake Swarms](#)
- [Principles of Broadband Seismometry](#)
- [Recording Seismic Signals](#)
- [Seismic Event Detection](#)
- [Seismic Network and Data Quality](#)
- [Seismic Noise](#)
- [Time History Seismic Analysis](#)

References

- Aki K, Richards PG (1980) Quantitative seismology – theory and methods, vol II. Freeman and Company, San Francisco, pp 609–625. ISBN 0-7167-1059-5
- Almendros J, Ibáñez JM, Alguacil G, Del Pezzo E (1999) Array analysis using circular-wave-front geometry: an application to locate the nearby seismo-volcanic source. *Geophys J Int* 136:159–170
- Barber NS (1958) Optimum arrays for direction finding. *N Z J Sci* 1:35–51
- Burg JP (1964) Three-dimensional filtering with an array of seismometers. *Geophysics* 29:693–713
- Buttkus B (2000) Spectral analysis and filter theory in applied geophysics. Springer, Berlin, xv + 667 pp. ISBN 3-540-62674-3
- Cansi Y (1995) An automatic seismic event processing for detection and location: P.M.C.C. method. *Geophys Res Lett* 22:1021–1024
- Capon J (1973) Signal processing and frequency-wavenumber spectrum analysis for a large aperture seismic array. In: Bolt BA (ed) *Methods in computational physics*, 13 geophysics. Academic, New York, xiii + 473 pp. ISBN 0-12-460813-2, 1–59
- Capon J, Greenfield RJ, Kolker KJ (1967) Multidimensional maximum-likelihood processing of a Large Aperture Seismic Array. *Proc IEEE* 55(2):192–211
- Dahlman O, Mykkeltveit S, Haak H (2009) Nuclear test ban – converting political visions to reality. Springer, Dordrecht. doi:10.1007/978-1-4020-6885-0, xviii + 277 pp. ISBN 978-1-4020-6883-6
- Davies D, Kelly EJ, Filson JR (1971) The VESPA process for the analysis of seismic signals. *Nature* 232:8–13
- Douglas A (2002) Seismometer arrays – their use in earthquake and test ban seismology. In: Lee WHK, Kanamori H, Jennings PC, Kisslinger C (eds) *Handbook of earthquake and engineering seismology Part A*. Academic, Amsterdam, xxiii + 933 pp. ISBN 0-12-440652-1, 357–367
- Gibbons SJ (2012) The applicability of incoherent array processing to IMS seismic arrays. *Pure Appl Geophys*. doi:10.1007/s00024-012-0613-2
- Gibbons SJ, Ringdal F, Kverna T (2008) Detection and characterization of seismic phases using continuous spectral estimation on incoherent and partially coherent arrays. *Geophys J Int* 172:405–421. doi:10.1111/j.1365-246X.2007.03650.x
- Goldstein P, Archuleta RJ (1991) Deterministic frequency-wavenumber methods and direct measurements of rupture propagation during earthquakes using a dense array: theory and methods. *J Geophys Res* 96:6173–6185
- Harjes H-P, Henger M (1973) Array-Seismologie. *Z Geophys* 39:865–905
- Haubrich RA (1968) Array design. *Bull Seismol Soc Am* 58:977–991

- Johnson DH, Dudgeon DE (2002) Array signal processing: concepts and techniques, Prentice Hall signal processing series. PTR Prentice Hall, Upper Saddle River, XIII + 533 pp. ISBN 0-13-048513-6
- Krüger F, Weber M, Scherbaum F, Schlittenhardt J (1993) Double beam analysis of anomalies in the core-mantle boundary region. *Geophys Res Lett* 20: 1475–1478
- Kværna T, Doornbos DJ (1986) An integrated approach to slowness analysis with arrays and three-component stations. *NORSAR Sci Rep*, 2-85/86, 60–69
- Levin MJ (1964) Maximum-likelihood array processing. Lincoln Laboratory report, 31 Dec 1964
- Macmillan (2013) Array. <http://www.macmillan-dictionary.com/dictionary/british/array>. Accessed 08 Dec 2013
- Mykkeltveit S, Åstebøl K, Doornbos DJ, Husebye ES (1983) Seismic array configuration optimization. *Bull Seismol Soc Am* 73:173–186
- Rost S, Thomas C (2002) Array seismology: methods and applications. *Rev Geophys* 40(3):1008. doi:10.1029/2000RG000100
- Rost S, Thomas C (2009) Improving seismic resolution through seismic arrays. *Surv Geophys* 30:271–299. doi:10.1007/s10712-009-9070-6
- Schmidt RO (1986) Multiple emitter location and signal parameter estimation. *IEEE Trans Ant Prop AP-34*:276–280
- Schweitzer J, Fyen F, Mykkeltveit S, Gibbons SJ, Pirli M, Kühn D, Kværna T (2012) Seismic arrays. http://ebooks.gfz-potsdam.de/pubman/item/escidoc_43213:7. doi:10.2312/GFZ.NMSOP-2_ch9; 80 pp. In: Bormann P (ed) New manual of seismological observatory practice (NMSOP-2), 2nd (rev) edn. Deutsches GeoForschungsZentrum GFZ, Potsdam. doi:10.2312/GFZ.NMSOP-2. Accessed 08 Dec 2013

Seismometer Self-Noise and Measuring Methods

Adam T. Ringler¹, Reinoud Sleeman², Charles Robert Hutt¹ and Lind S. Gee¹

¹Albuquerque Seismological Laboratory, U.S. Geological Survey, Albuquerque, NM, USA

²Seismology Division, Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands

Synonyms

Coherence analysis; Instrument noise; Seismometer testing

Introduction

Seismometer self-noise is usually not considered when selecting and using seismic waveform data in scientific research as it is typically assumed that the self-noise is negligibly small compared to seismic signals. However, instrumental noise is part of the noise in any seismic record, and in particular, at frequencies below a few mHz, the instrumental noise has a frequency-dependent character and may dominate the noise. When seismic noise itself is considered as a carrier of information, as in seismic interferometry (e.g., Chaput et al. 2012), it becomes extremely important to estimate the contribution of instrumental noise to the recordings.

Noise in seismic recordings, commonly called seismic background noise or ambient Earth noise, usually refers to the sum of the individual noise sources in a seismic recording in the absence of any earthquake signal. Site noise (e.g., cultural sources, nearby tilt signals, etc.) and noise introduced by the sensitivity of an instrument to non-seismic signals (e.g., temperature and pressure variations, magnetic field changes, etc.) both contribute to the ambient seismic noise levels. The background noise ultimately defines a lower limit for the ability to detect and characterize various seismic signals of interest. Background noise levels have also been found to introduce a systematic bias in arrival times because the amplitude of the seismic phase must rise above the station's noise levels (Röhm et al. 1999). The upper limit of useful signals is governed by the clip level of the recording system (the point at which a recording system's output is no longer a linearly time-invariant representation of the input).

Site noise can be reduced by careful site selection (e.g., hard rock far from strong noise sources) and by emplacing instruments in good vaults or boreholes. It is also possible to reduce sensitivity to non-seismic signals by thermal insulation and appropriate shielding such as pressure chambers (Hanka 2000). At quiet sites with well-installed instrumentation, instrument noise may be the dominant noise source

(Berger et al. 2004); this is especially true for long-period seismic data (>100 s period) on very broadband instruments (e.g., Streckeisen STS-1 seismometer). The interpretation of such data only makes sense if the instrumental noise level is known. Also, research on noise levels in seismic recordings, the effect of noise reduction by the installation technique, and the nature and contribution of different noise sources to the recordings require knowledge of instrumental self-noise to rule out bias from the instrumentation self-noise.

A number of tests have been developed, under the assumption that instrument self-noise is approximately constant as a function of time (e.g., Evans et al. 2010; Holcomb 1989, 1990; Sleeman et al. 2006). Only recent studies have started to look at the potential time dependence of instrument self-noise (e.g., Sleeman and Melichar 2012). Understanding the self-noise of a given piece of recording equipment helps station operators to better identify sites that can take advantage of low-self-noise instruments. This knowledge also allows a network operator to provide higher-quality data with limited resources by making better use of their high-quality instruments. Having a rough understanding of an instrument's self-noise also gives a first-order diagnostic for determining if the recording system, as installed, is performing satisfactorily.

In order to estimate the self-noise of a seismometer, it is necessary to remove non-instrumental noise signals (e.g., earthquakes and ambient Earth noise) from the data. This is often accomplished by using coherence analysis techniques. In its simplest form, one can select a quiet time period at a low-noise site and attribute all recorded noise to the sensor (one-sensor method). However, it can be difficult to find stations with sufficiently low site noise to evaluate high-quality broadband seismometers. This is especially true at periods between the primary and secondary microseism (approximately 4–22 s period). In such cases, more sophisticated techniques, where one removes coherent signals using colocated instruments, are required. This can be done using a second instrument or two additional instruments (two- and three-instrument methods). For passive sensors, it

is possible to obtain an estimate of the self-noise by locking the mass of the instrument.

This is a general overview on the various methods currently used for estimating the instrument self-noise when using one, two, or three sensors. As there is no universally “best” method for all types of instruments and test conditions, some of the advantages and disadvantages of each method are discussed. Variants of these methods, such as rotating horizontal components of Earth motions to maximize the coherence and correcting for misalignments, are also included. Finally, the test setup used for estimating self-noise for the various methods is discussed and examples of corresponding test results are given.

Methods

To discuss the various methods currently in use for estimating the self-noise of a seismometer, a mathematical framework common to all the methods is developed. The system under test is assumed to be a linear time-invariant system (LTI) making the system completely determined by its impulse response (Scherbaum 2007).

Basic Assumptions and Conventions

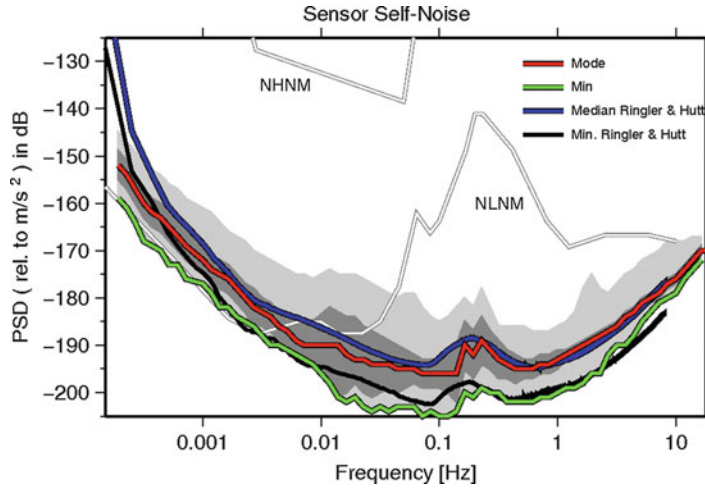
Let x_i denote an input seismic signal, h_i denote the seismometer's impulse response, and n_i denote the self-noise of instrument i . Then the instrument's output signal y_i can be modeled, in the time domain, as

$$y_i = h_i * (x_i + n_i) \quad (1)$$

where “*” denotes convolution (Holcomb 1990; Sleeman et al. 2006). It is possible to write this in the frequency domain as

$$Y_i = H_i \cdot (X_i + N_i), \quad (2)$$

where capital letters denote the Fourier transforms of the corresponding lowercase letter time domain terms of index i . With the assumption that the self-noise of two different instruments is incoherent, the term $N_{ij} = N_i N_j$ is zero for $i \neq j$. It is



Seismometer Self-Noise and Measuring Methods, Fig. 1 Self-noise estimates of the vertical component of the STS-2 seismometer for the minimum noise (green) and the mode (red) of sensor noise measured at the Conrad Observatory, Austria. The light gray and dark gray bands

depict the 95 % and 68 % percentile power spectral density (PSD) estimates. Finally, median (blue) and minimum (black) self-noise estimates for the STS-2 are taken from the data at the Albuquerque Seismological Laboratory (ASL)

also assumed that the self-noise and the input signal are incoherent, $X_i N_j$ is zero for $i \neq j$. The cross power between instrument i and j is denoted by P_{ij} so for $i \neq j$ we can write the cross power as

$$P_{ij} = H_i X_i \overline{H_j X_j} \quad (3)$$

where the bar denotes the complex conjugate. The coherence between instruments i and j is given by

$$\gamma^2 = \frac{|P_{ij}|^2}{P_{ii} P_{jj}} \quad (4)$$

Finally, it is assumed that all of the instruments in the test have common output units so that they have similar ground motion units (e.g., m/s or m/s²) after removing the instrument response. This assumption can generally be made true by using an “omega correction” (multiplying or dividing by $\omega = 2\pi f$, where f is the frequency) in the frequency domain (Stearns 1975).

Noise levels are often compared to absolute Earth noise models such as Peterson’s New Low-Noise Model (NLNM) and New High-Noise Model (NHNM), which are global seismic

noise models derived from the 74 Global Seismographic Network (GSN) stations along with one additional station (Peterson 1993). This NLNM represents the lower envelope of 2,000 data records and represents an approximation of the lowest observed seismic noise levels, whereas the NHNM represents the upper envelope of the same 2,000 data records and represents an approximation of the highest observed seismic noise levels (Fig. 1).

Single-Sensor Method

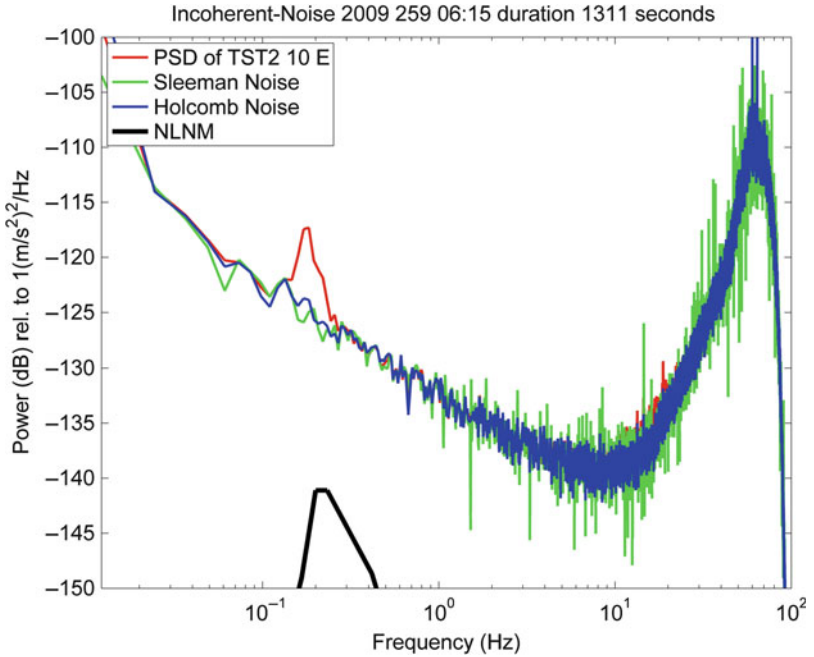
When testing instruments at locations where the site noise is well below the self-noise of the instrument, it is possible to attribute the power in a given frequency band entirely to the instrument’s self-noise (Fig. 2). This often occurs when testing strong-motion accelerometers in a quiet vault or lower-grade sensor in almost any good site (Evans et al. 2010). In such cases the simple relation is obtained:

$$N_{ii} = \frac{P_{ii}}{H_i \overline{H_i}} \quad (5)$$

assuming that $X_i \ll N_{ii}$. Even when this assumption is not satisfied across the entire frequency

Seismometer Self-Noise and Measuring Methods,

Fig. 2 Self-noise estimates for a strong-motion accelerometer using the single-sensor method (red), the Sleeman (Sleeman et al. 2006) three-sensor method (green), and the Holcomb (Holcomb 1989) two-sensor method (blue). For reference the New Low-Noise Model (NLNM) is included (black)



band, it is often possible to get an initial estimate of the instrument's self-noise in a specific frequency band (e.g., outside the primary and secondary microseism bands for state-of-the-art accelerometers). The single-sensor method can also be used to get an upper bound on the self-noise of an instrument. If this upper bound is obtained at a relatively quiet location, it might characterize the self-noise sufficiently well to identify the suitability of an instrument at a station with potentially much higher site noise. This method gives a first approximation of an instrument's self-noise, which can be valuable when testing time is limited or there are insufficient resources to use multiple sensors.

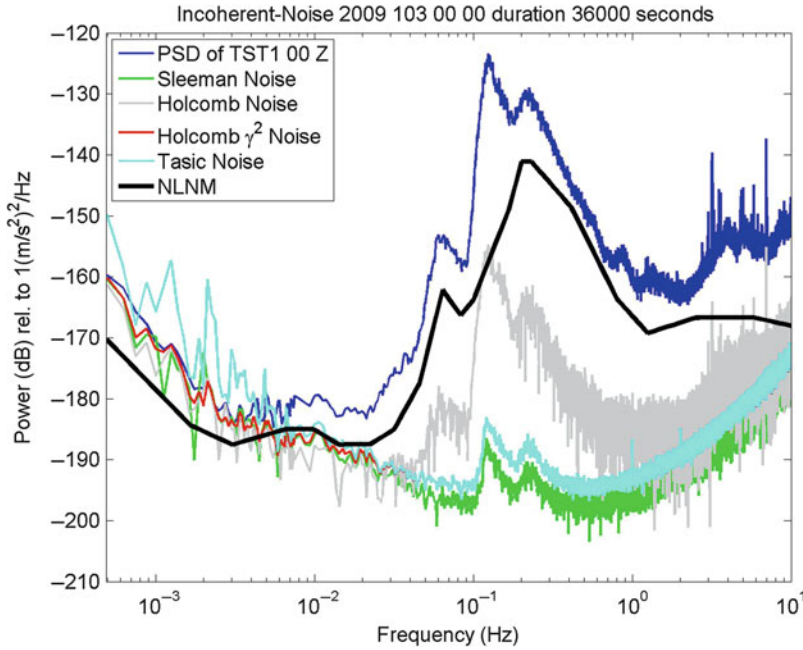
Two-Sensor Methods

Estimating self-noise using two sensors has been the traditional method for noise testing broadband seismometers for many years (Holcomb 1989). As very few locations have background noise levels below that of high-quality broadband sensors across a wide frequency range, it becomes important to remove the local background noise (Fig. 3). Assuming that the colocated sensors are recording similar seismic

signals (same ground motion), $X_i = X_j$, it is possible to derive the self-noise of instrument i from Eq. 2. Recalling the assumption that $N_{ij} = 0$ for $i \neq j$ we have

$$N_{ii} = \frac{P_{ii}}{|H_i|^2} - \frac{P_{ij}}{H_i H_j^*}. \quad (6)$$

Under the assumption that the sensors both have well-known responses, the instrument corrected output, from the two sensors, should only differ in instrumental self-noise. By estimating the coherent signal between the two records, the coherence signal is removed resulting in the incoherent signal which is attributed to the self-noise. It can be seen from this estimate of the self-noise that it is critical to have well-described transfer functions for both instruments i and j . Since the transfer functions are used in the calculation, errors in the transfer function will produce errors in the self-noise estimates of the instrument. This method was originally proposed by Holcomb (1989), who later characterized error sources (Holcomb 1990). In the latter work, Holcomb also suggested alternative two-sensor methods, under the assumptions that the two sensors



Seismometer Self-Noise and Measuring Methods, Fig. 3 Comparison of five different methods for estimating the self-noise using vertical broadband sensor data: the one-sensor (direct power spectral density (PSD)) method (blue), the three-sensor Sleeman (Sleeman et al. 2006) method (green), the two-sensor Holcomb (Holcomb 1989) method (light gray), the two-sensor γ^2 Holcomb

method (red), and the two-sensor Tasič (Tasič and Runovc 2012) method (cyan) (for the different methods, see text). The cyan and red lines overlay at frequencies less than approximately 0.01 Hz. For reference the New Low-Noise Model (NLNM) is the solid black line. The elevated noise in the microseism band is caused by misalignment

under test have equal noise and a high signal-to-noise ratio (ratio of input signal to instrument self-noise). In this case, it is possible to derive the following estimate for the self-noise of instrument i :

$$N_{ii} = \frac{\left(P_{ii} \frac{(1 - \gamma^2)}{\gamma^2} \right)}{H_i \overline{H_j}}. \quad (7)$$

Using a related approach, Tasič and Runovc (2012) develop a different method for estimating the self-noise using two instruments. In their approach the self-noise of instrument i is given by

$$N_{ii} = \frac{P_{ii}(1 - \gamma^2)}{(H_i \overline{H_j})} \quad (8)$$

where again they assume the signal-to-noise ratio is large.

Three-Sensor Method

Using three colocated sensors, it is possible to estimate the self-noise of each instrument while minimizing errors in the estimate due to uncertainty in the transfer functions (Sleeman et al. 2006). This method has become the preferred approach for estimating self-noise for broadband sensors, even though it requires additional resources and setup; it typically “sees through” site noise more deeply to extract lower estimates of instrument noise.

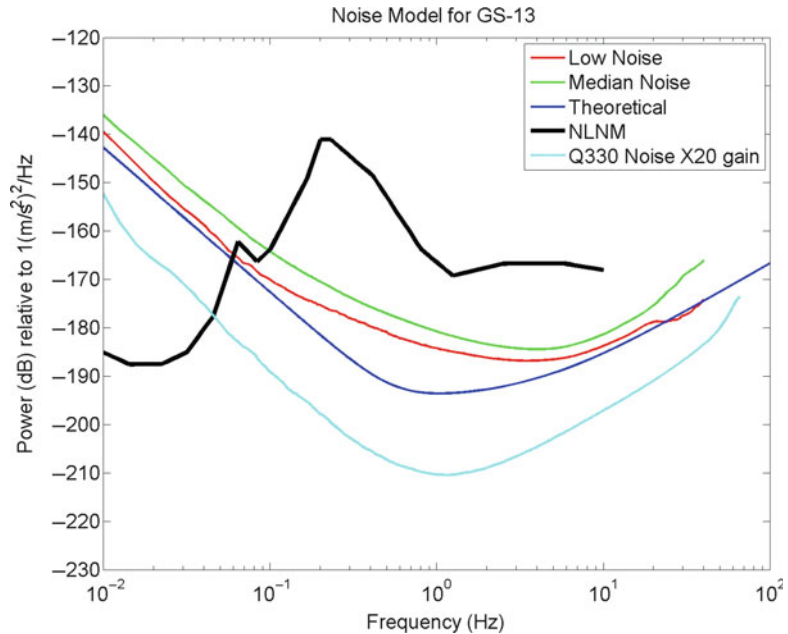
Using Eq. 2 and the assumption that $X_i = X_j = X_k$, one can estimate the self-noise of instrument i as

$$N_{ii} = \frac{\left(P_{ii} - P_{ij} \frac{P_{ik}}{P_{jk}} \right)}{H_i \overline{H_l}}. \quad (9)$$

The term P_{ik}/P_{jk} can be viewed as a “relative transfer function” H_{ij} between instruments i and j . The three-sensor method is related to the

Seismometer Self-Noise and Measuring Methods,

Fig. 4 Self-noise estimate of the Geotech GS-13 geophone using a theoretical self-noise model (blue) versus the estimated low self-noise model of the Geotech GS-13 (red) from 11 self-noise measurements; the median self-noise from these 11 tests is also shown (green). The self-noise of the digitizers used, Quanterra Q330HRs, is also shown after correction with the response of the Geotech GS-13 (cyan). For reference the New Low-Noise Model (NLNM) is included (black)



two-sensor method developed by Tasič and Runovc (2012) when one compares relative transfer functions. It is important to note that the transfer function of instrument i is only used to convert the noise estimate to units of ground motion; thus, errors in the transfer functions do not propagate to the self-noise estimate before deconvolution with the individual instrument (Ringler et al. 2011). Because the three-sensor method is not as sensitive to errors in the transfer functions, it extracts the self-noise with a potentially higher accuracy than the two-sensor method (Fig. 3).

Other Techniques

When instrument designers are selecting components for building a seismometer, they often have noise estimates for each electronic component as well as estimates of fundamental noise contributions (e.g., Brownian noise). Through modeling these individual noise sources, it is possible to estimate the total self-noise of a seismometer without having a working prototype. A modeling example for Brownian noise is described by Aki and Richards (2002), using the physical parameters of a simple gravitational pendulum. Such methods often give a first approximation of the instrumental noise to be aimed for during design (Fig. 4).

Such theoretical methods for estimating the self-noise of an instrument have been applied to a number of seismometers in conjunction with their digitizers (Rodgers 1994). However, when such methods are used, a number of assumptions must be made (e.g., that the pendulum obeys the small-angle approximation and all electronic components perfectly match their specified noise levels).

For seismometers that are not controlled by feedback but having self-noise well below any available test site's ambient noise, such as geophones at high frequency (e.g., GeoTech GS-13), it is possible to estimate the self-noise using a "locked mass" test (Havskov and Alguacil 2004). This test is performed by locking the mass of the instrument and recording the output of a single instrument. This test provides a lower noise-bound estimate but may not fully characterize the self-noise of the instrument because it lacks noise contributions from moving hinges and other mechanical sources.

Testing and Analysis

Characterizing the self-noise of a seismometer requires attention to detail in the test setup,

Seismometer Self-Noise and Measuring Methods,

Fig. 5 Three colocated Geotech GS-13 short-period seismometers setup for test to be sensitive to vertical ground motion. All three are sitting on a granite slab supported at three points and are in the Albuquerque Seismological Laboratory (ASL) underground vault. On the back right is a reference Streckeisen STS-2 seismometer in a steel bell jar (not used in the test)



careful selection of data windows to avoid transients, identification of potentially non-seismic noise sources, and methods for selecting data to analyze. In both the two- and three-sensor methods, it is necessary to make sure all instruments truly are recording the same ground motion.

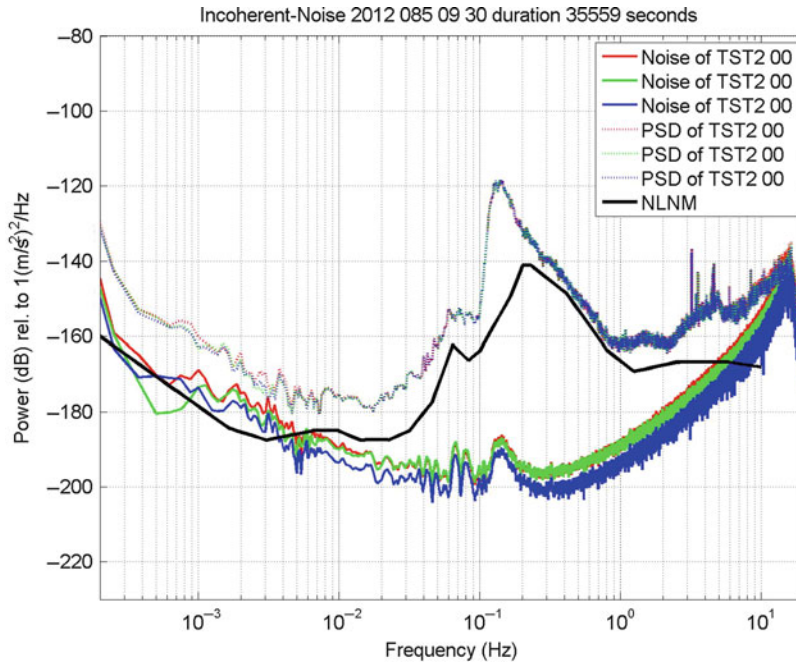
Test Setup

Both the two- and three-sensor methods assume that all instruments are recording the same ground motion. In order for this to be true, it is critical that the instruments be colocated and co-aligned. Furthermore, since local tilt signals can vary widely over even a few inches, one must take care to insure that all three instruments are measuring similar tilt signals so that they can be removed using coherence analysis. Locating all instruments on a stiff slab (e.g., granite or gabbro) with a three-point suspension is usually sufficient for this purpose (Fig. 5).

It is also important to isolate the instruments from non-seismic noise sources (e.g., temperature variations, pressure variations, locally induced tilt). These noise sources can increase the self-noise estimate of the individual instruments as they generally do not respond coherently to such sources (Anderson et al. 2012). Misalignment of the horizontal components can

also contribute to a decrease in signal coherency and increase the estimate of the self-noise. Such relative errors in misalignment can be seen by elevated incoherent noise-level estimates in the microseism band (Fig. 6), so it is possible to reduce such misalignment errors by rotating the instrument outputs numerically to maximize coherence (Tasič and Runovc 2012). Recent work by Gerner and Bokelmann (2013) shows that the leakage of microseism signals into the self-noise can be eliminated successfully by numerical rotation of the three-component traces and using the three-sensor method. Their study optimally aligns two RefTek 151-60A sensors with a third one and shows that self-noise estimates outside the microseism bands are not compromised by misalignment errors.

Finally, if one wants to isolate the self-noise of a seismometer, it is critical that the digitizer and other recording equipment have self-noise levels well below that of the sensor. This can be a problem on low-gain instruments, such as strong-motion accelerometers. Verifying that the digitizer's self-noise levels are below that of the seismometer requires a terminated-input test of the digitizer, using a terminating resistor similar to the output impedance of the sensors to be tested. When the digitizer's noise level is not below the seismometer noise level for any of the



Seismometer Self-Noise and Measuring Methods, Fig. 6 Self-noise estimates from three horizontal very broadband Metrozet M-2166 seismometers (flat to velocity from 0.0028 to 10 Hz) using Sleeman (Sleeman et al. 2006) three-sensor self-noise estimate (solid lines). Single-sensor estimates are in dotted lines. In order to minimize incoherent tilt noise, all three instruments were

installed with the same orientation on the same baseplate and the baseplate is supported at three points only. The relatively small increase in the self-noise in the secondary microseism frequency band (at approximately 0.125–0.25 Hz) suggests all three instruments are well aligned to one another

frequency range of interest, digitizer preamps generally can be used for low-noise amplification of the sensor signal to a level above the digitizer noise level.

Data Selection

Data selection is currently an important item in the development of algorithms and techniques to estimate self-noise. This requirement makes it necessary to estimate self-noise during only “quiet” time periods (e.g., during nighttime and periods with low pressure-induced tilt noise). No current agreement exists as to whether data selection should be applied or not, or which criteria should be used for data selection. In particular, when features in the data are not fully understood, the debate on data selection is ongoing and not yet entirely resolved, though a best-case result seems to require such minimization (Ringler et al. 2011). For example, the

occurrences of periodic pulsing in the data are not yet understood (though many have been named “popcorn noise,” “spherics,” and so forth and are tentatively explained in various ways). Some of these noise sources could properly be considered either as part of the self-noise of the instrument or as transient external signals that would artificially elevate the self-noise if included (Sleeman and Melichar 2012). The discussion on data selection of time segments becomes even more controversial when limited amounts of data are available. By using multiple quiet time segments, it is possible to understand the self-noise of an instrument in at least a best-case scenario (Ringler and Hutt 2010). One such example of a best-case scenario, along with a self-noise estimate using the median of multiple tests, is shown in Fig. 4. This example also suggests that even when using coherency analysis techniques, it is important to do such tests in a

low-noise environment. The selection of time segments is additionally controversial when limited amounts of test data exist, because the assumption that the self-noise of a seismometer is time invariant is only approximately true, as seen in the relatively large gray bands in Fig. 1, which depict the 95 % and 68 % of the self-noise of the Streckeisen STS-2 seismometer as recorded under very stable conditions at the Conrad Observatory (Vienna, Austria) over an entire year.

A synthetic-data experiment by Sleeman and Melichar (2012) shows that the three-sensor technique can reliably extract instrumental noise even for high “seismic-signal-to-instrumental-noise (SNR)” ratios, but is in practice limited to lower ratios due to misalignments or mis-leveling between the sensors. As discussed earlier, small alignment or leveling differences between the sensors decrease the coherency between the recordings and thus affect the noise estimate. It was found that an alignment error of 0.2° in any axis allows the technique to extract self-noise for SNR values up to 60 dB, which was confirmed in their associated experiment using real data. As this order of misalignment between corresponding axes in similar colocated sensors is in agreement with the precision of best-practices manufacturing, one should only use data taken during “quiet” time periods (SNR below 60 dB, e.g., during night and periods with minimal pressure-induced tilt). Even during these times, microseism noise can make it difficult to estimate the self-noise of broadband sensors in the period bands of 4–8 s and 18–22 s for instruments (Ringler et al. 2011).

Analysis Parameters

As in any time series analysis, appropriate time window lengths, number, and overlapping must be used to resolve the low frequencies of interest while minimizing variance. For very broadband seismometers, it is often of interest to understand the self-noise for periods up to several thousands of seconds, which corresponds to a total window length of at least 8 h divided into overlapping sub-windows, each of about an hour (e.g., Table 1 of Evans et al. 2010).

To directly compare results between noise tests, it is important that the tests be conducted using similar methods (e.g., settling time, installation methods, data selection, and spectral processing parameters).

Expected Results and Caveats

A discussion on the expected self-noise results for various types of seismometers now follows. This discussion is in no way exhaustive but simply discusses some of the more common situations and what is generally expected for various test scenarios.

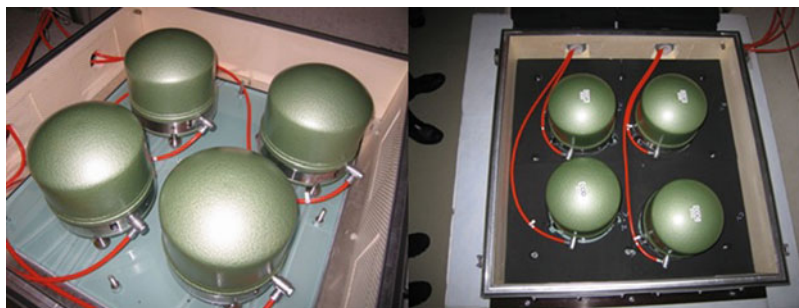
Strong-Motion Accelerometers

The noise levels of strong-motion accelerometers make estimating the self-noise of such sensors easier than typical low-noise broadband sensors (Fig. 2). Since the user community is generally only interested in strong motions with frequencies from a few Hz to a few tens of seconds period, it is often possible to resolve the self-noise using relatively short testing intervals.

As noted earlier, one must be careful to make sure that the digitizer noise is below that of the sensor. Since strong-motion accelerometers are generally sampled at higher rates than broadband sensors (≥ 100 sps), it is necessary to identify the self-noise to much higher frequencies than for broadband sensors (Cauzzi and Clinton 2013). At these higher frequencies, it becomes more difficult to isolate the instrument from potentially non-seismic noise sources such as electrical power grid frequencies (e.g., 50 Hz in Europe and 60 Hz in North America) noise. At these higher frequencies, the coherency between recordings from colocated sensors presumably decreases, but this decrease in coherency is typically compensated by decreasing ambient noise and increasing instrument self-noise.

Broadband Seismometers

After installing a broadband seismometer, it is important to let the instrument fully settle down from mechanical stresses that can build up during transport and adjust to temperature before



Seismometer Self-Noise and Measuring Methods, Fig. 7 Colocated Streckeisen STS-2 sensors at the Conrad Observatory, Austria. Sensor casings are co-aligned by parallel grooves in the underlying glass plate. Thermal

insulation consists of thin layers of neoprene around the sensors, which also reduces the noise contribution due to air convection around the sensors

attempting to estimate the self-noise. What constitutes an instrument that is “fully settled” is not well understood and currently a topic of some controversy. After the instrument has been allowed to settle, one can estimate the self-noise by using long-running quiet time periods (Hutt et al. 2010). Local changes in wind and pressure can introduce incoherent elevated noise levels on horizontal channels at long periods (e.g., >100 s period) as well as short (e.g., <1 s period).

Estimates Using Other Data Selection Techniques

Although the process of self-noise measuring is not yet mature, the methods and techniques are rapidly improving. As manufacturers’ abilities to reduce sensor noise levels improves, methods for measuring the self-noise will also need to improve. A better understanding of the sensitivity of seismometers to non-seismic phenomena and techniques to thermally isolate the sensors (e.g., Fig. 7) or correct the data from these sources will be needed. Although there has been initial work on this subject (e.g., Forbriger et al. 2010; Zürn et al. 2007) with promising results, there is still much to be done before these methods become widely used by the seismological community.

Other Geophysical Instruments

The above methods for estimating seismometer and accelerometer self-noise can be applied more generally to other instruments and recording

systems. As the only assumptions are that all instruments will have a common input signal and output units, it is also possible to estimate the self-noise of digitizers (a “terminated noise” test), infrasound sensors (Hart et al. 2007), tilt meters, gravimeters, and rotational seismic sensors.

In all cases, the general approach for estimating the self-noise will be similar, though the details may differ (e.g., test setup logistics, frequency band of interest, etc.). For a discussion on applying the three-sensor self-noise technique to digitizers, the reader is referred to Kromer (2006) and Sleeman et al. (2006). Details on applying self-noise techniques to other geophysical instruments are beyond the scope of this entry.

Summary

With the suite of various methods for testing seismometer self-noise, it is possible to get an initial estimate of the self-noise of a seismometer even in less-than-ideal testing situations. For instruments with relatively high self-noise, simple (one-sensor) self-noise methods may prove sufficient.

Using additional high-quality sensors and more sophisticated techniques (two- or three-sensor methods) can help to eliminate common signals when there is concern of the local background noise being above the sensor’s self-noise.

This will almost always be the case when testing broadband seismometers, but will not necessarily be the case when testing accelerometers though at least microseisms are above the noise floor of the state-of-the-art accelerometers even at an inland continental site.

Equally important to the two- and three-sensor methods is that the common seismic signals be coherent among all instruments in the test. As discussed, this can be done by installing the instruments in a quiet vault on a granite slab or other stiff material to increase the coherence among sensors. Care should also be taken to reduce the influence on instruments under test from local non-seismic noise sources such as pressure or temperature variations and that adequate thermal settling time be provided before testing.

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US government.

Cross-References

- [Seismic Network and Data Quality](#)
- [Seismic Noise](#)
- [Sensors, Calibration of](#)

References

- Aki K, Richards PG (2002) Quantitative seismology. University Science, Sausalito
- Anderson KE, Anderson JF, Anthony RE, Chaput J, McMahon ND, Morton EA, Aster R (2012) A site comparison between shallow vault-deployed and direct burial broadband seismometers, IRIS 2012 Workshop Abstracts. http://www.iris.edu/hq/iris_workshop2012/scihi/WebPages/0070.html
- Berger J, Davis P, Ekström G (2004) Ambient earth noise: a survey of the global seismographic network. *J Geophys Res* 109, B11307
- Cauzzi C, Clinton J (2013) A high- and low-noise model for high-quality strong-motion accelerometer stations. *Earthq Spectra* 29:85–102
- Chaput JA, Zandomenighi D, Aster RC, Hunter K, Kyle PR (2012) Imaging of Erebus volcano using body wave seismic interferometry of Strombolian eruption coda. *Geophys Res Lett* 39, L07304
- Evans JR, Followill F, Hutt CR, Kromer RP, Nigbor RL, Ringler AT, Steim JM, Wielandt E (2010) Method for calculating self-noise spectra and operating ranges for seismographic inertial sensors and recorders. *Seismol Res Lett* 81:640–646
- Forbriger T, Sidmer-Schnidrig R, Wielandt E, Hayman M, Ackerley N (2010) Magnetic field background variations can limit the resolution of seismic broad-band sensors. *Geophys J Int* 183:303–312
- Gerner A, Bokelmann G (2013) Instrument self-noise and sensor misalignment. *Adv Geosci* 36:17–20
- Hanka W (2000). Which parameters influence the very long period performance of a seismological station? GEOFON Program. <http://geofon.gfz-potsdam.de/geofon/manual/welcome.html>
- Hart D, Merchant B, Chael E (2007) Seismic and infrasound sensor testing using three-channel coherence analysis. In: 29th monitoring research review: ground-based nuclear explosion monitoring technologies. LA-UR-07-5613, 935–944
- Havskov J, Alguacil G (2004) Instrumentation in earthquake seismology. Springer, Dordrecht
- Holcomb LG (1989) A direct method for calculating instrument noise levels in side-by-side seismometer evaluations. U.S. Geological Survey Open File Report, 89–214, 34 pp
- Holcomb LG (1990) A numerical study of some potential sources of error in side-by-side seismometer evaluations. U.S. Geological Survey Open File Report, 90–406, 41 pp
- Hutt CR, Evans JR, Followill F, Nigbor RL, Wielandt E (2010) Guidelines for standardized testing of broadband seismometers and accelerometers. U.S. Geological Survey Open File Report, 2009–1295, 62 pp
- Kromer RP (2006) Technology report evaluation of the Kinematics/Quanterra Q330HR Remote Seismic System for IRIS/GSN Q330HR/GainX1 Configuration, 32286. http://www.iris.edu/hq/files/programs/gsn/documents/IRIS_Progress_Report2_Q330HR.pdf
- Peterson J (1993) Observations and modeling of seismic background noise. U.S. Geological Survey Open File Report, 93–322, 95 pp
- Ringler AT, Hutt CR (2010) Self-noise models of seismic instruments. *Seismol Res Lett* 81:972–983
- Ringler AT, Hutt CR, Evans JR, Sandoval LD (2011) A comparison of seismic instrument noise coherence analysis techniques. *Bull Seismol Soc Am* 101:558–567
- Rodgers PW (1994) Self-noise spectra for 34 common electromagnetic seismometer/preamplifier Pairs. *Bull Seismol Soc Am* 84:222–228
- Röhm AHE, Trampert J, Paulssen H, Snieder RK (1999) Bias in reported seismic arrival times deduced from the ISC Bulletin. *Geophys J Int* 137:163–174
- Scherbaum F (2007) Of poles and zeros: fundamentals of digital seismology, 2nd edn. Springer, Dordrecht
- Sleeman R, Melichar P (2012) A PDF representation of the STS-2 self-noise obtained from one year of data recorded in the Conrad Observatory, Austria. *Bull Seismol Soc Am* 102:587–597

- Sleeman R, van Wettum A, Trampert J (2006) Three-channel correlation analysis: a new technique to measure instrumental noise of digitizers and seismic sensors. *Bull Seismol Soc Am* 96:258–271
- Stearns SD (1975) *Digital signal analysis*. Hayden Book Company, Rochelle Park
- Tasić I, Runovc F (2012) Seismometer self-noise estimation using a single reference instrument. *J Seismol* 16:183–194
- Zürn W, Exß J, Steffen H, Kroner C, Jahr T, Westerhaus M (2007) On reduction of long-period horizontal seismic noise using local barometric pressure. *Geophys J Int* 171:780–796

Seismometer, Extended Response

Dieter Stoll

Lennartz Electronic GmbH, Tübingen, Germany

Synonyms

Feedback system; Inverse filtering; Negative impedance converter

Introduction

Until the advent of the broadband seismometer (Wielandt and Streckeisen 1982), electrodynamic geophones were used for picking up short-period (1 s and less) earth movement, and LP (long period) seismometers were used for anything with lower frequency than one second. Owing to their bulk and weight, LP seismometers are not practical for use in the field; they typically require a controlled observatory-type environment. Most types of geophones, on the other hand, are by their very nature field worthy, compact, and cheap. This is because they are mass-produced for use in seismic exploration where they are typically deployed in large numbers (hundreds or thousands). The downside is that their natural frequency is usually 4.5 Hz or more. Some geophones designed for scientific use have a natural frequency of 1 Hz, but this low-frequency limit comes at the expense of

significant bulk and weight and an unwelcome increase in mechanical complexity.

Thus, combining the ruggedness of the short-period geophone with the extended frequency response of an LP seismometer has long been a goal for researchers and instrument makers alike.

Geophone Frequency Response

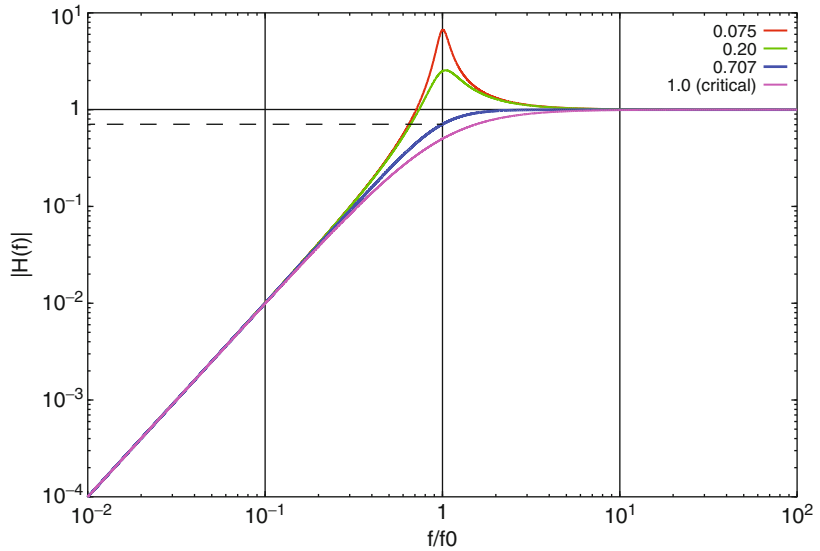
In its simplest form, a geophone consists of a mass suspended by a spring. To pick up the relative motion between the mass (which, due to inertia, tends to remain steady) and the frame (which will follow the motion of the Earth's surface), a moving coil system is typically used. A coil that is attached to the suspended mass moves in a magnetic field provided by a permanent magnet attached to the frame, thus providing an electrical output voltage that is proportional to the relative velocity between mass and frame.

If the suspended mass were an ideal oscillator (i.e., without any damping forces at play), a geophone would not pick up Earth motion very faithfully. In fact, a single impulse would cause it to oscillate forever at its natural frequency. A certain amount of damping is always provided by mechanical damping (friction) and electromagnetic damping, with the latter being responsible for the major part of damping. Electromagnetic damping is a consequence of Lenz's law (in a closed circuit moving inside a magnetic field, a current is induced which in turn generates a magnetic field opposed to the original magnetic field, thus resulting in a force counteracting the movement of the conductor), and is proportional to the sum of the coil resistance and the external load on the coil (the input resistance of the amplifier or recorder).

In a plain geophone (without any active electronics involved), the minimum amount of damping is given by the open circuit damping (typically in the range 0.25–0.5, but some commercially available geophones go down to 0.02 (Rodgers 1993)), meaning there is no external load connected. With no external load connected,

Seismometer, Extended Response,

Fig. 1 Geophone frequency response for different damping factors (normalized to be 1.0 for $f \gg f_0$)



the circuit is open, no current flows through the coil, and no additional damping is contributed by the coil. The other extreme is a shorted coil – in this case, the external load is $0 \, \Omega$, and only the intrinsic resistance (inductive reactance, to be more precise) limits the amount of current flowing. Consequently, this setup provides the maximum possible amount of damping (this is why classical geophones should be transported only with the output coil(s) shorted – the high damping will help inhibit extraneous mass movement).

The modulus of the complex transfer function of an electrodynamic geophone with respect to ground velocity depends on damping as follows (Bormann 2012, Chapter 5.2):

$$H_V(\omega) = \frac{\omega^2}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4h^2\omega^2\omega_0^2}} \quad (1)$$

where h is the damping factor expressed as the ratio of actual to critical damping (when $h = 1$), ω is angular frequency, and ω_0 denotes the natural angular frequency.

Figure 1 shows the modulus of a geophone's transfer function to ground velocity over a wide range of frequencies for different damping

factors, ranging from open circuit (very low damping, strong resonance peak) to 1.0. Usually, the desired value is 0.707 of critical damping (thick line). At this setting, the modulus of the transfer function at the natural frequency is 0.707 ($0.5\sqrt{2}$).

Since open circuit damping is lower than the desired value of 0.707, an external damping resistor (often referred to as a “shunt”) is used to achieve higher damping. An inevitable consequence of shunting the geophone coil is that its output decreases (because the external resistor and the geophone coil resistance form a voltage divider). In the flat part of the transfer function, the geophone's output is proportional to

$$\left(\frac{R_d}{R_d + R_{\text{coil}}} \right) \quad (2)$$

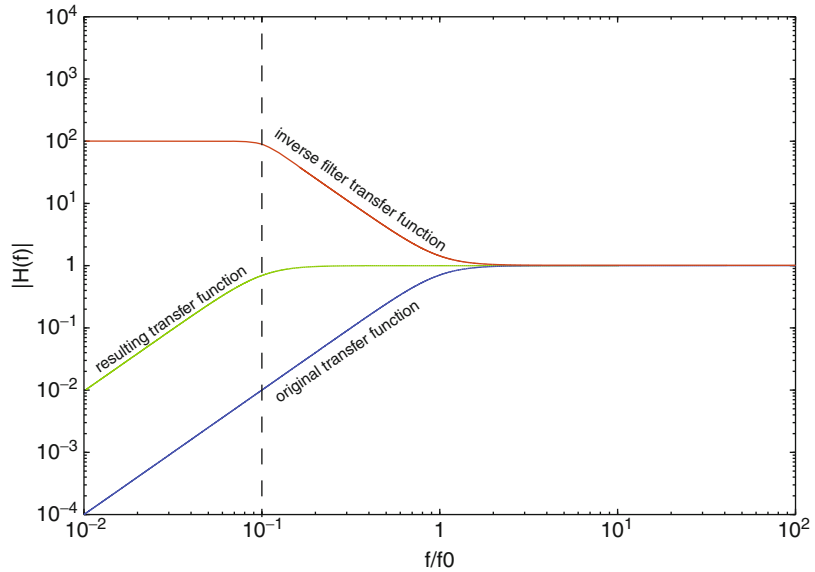
where R_d is the damping (shunt) resistance, and R_{coil} is the coil resistance.

This fact has been deliberately omitted from Fig. 1; all curves have been normalized to 1.0 for frequencies much higher than the natural frequency.

For frequencies $\ll f_0$, the geophone's output rolls off at $-12 \, \text{dB/octave}$ (proportional to f^2). For frequencies $\gg f_0$, the transfer function is flat.

Seismometer, Extended Response,

Fig. 2 Schematic operation of inverse filter, new $f_0 = 0.1 \times$ original f_0



Methods of Frequency Response Extension

Extending the low-frequency response of a geophone without increasing its bulk and weight can be achieved with several methods. The most common ones will be briefly outlined here.

Using an Inverse Filter

An electronic filter can be implemented to compensate the geophone's roll-off at low frequencies (Havskov and Alguacil 2004). In real life, the filter's transfer function cannot increase indefinitely towards lower frequencies. A new fictitious natural frequency is defined by the point where the filter's transfer function returns to being flat. Figure 2 shows the schematic function of an inverse filter (Lippmann 1982).

Intriguing as it may seem at first sight, this approach is plagued with many practical problems. Noise appearing in electronic circuits is not necessarily "white," i.e., equally distributed over a wide frequency range. There is one particularly annoying type of noise called $1/f$ noise, which, as its name implies, means it gets increasingly stronger the lower the frequency gets. This phenomenon occurs below a certain corner frequency

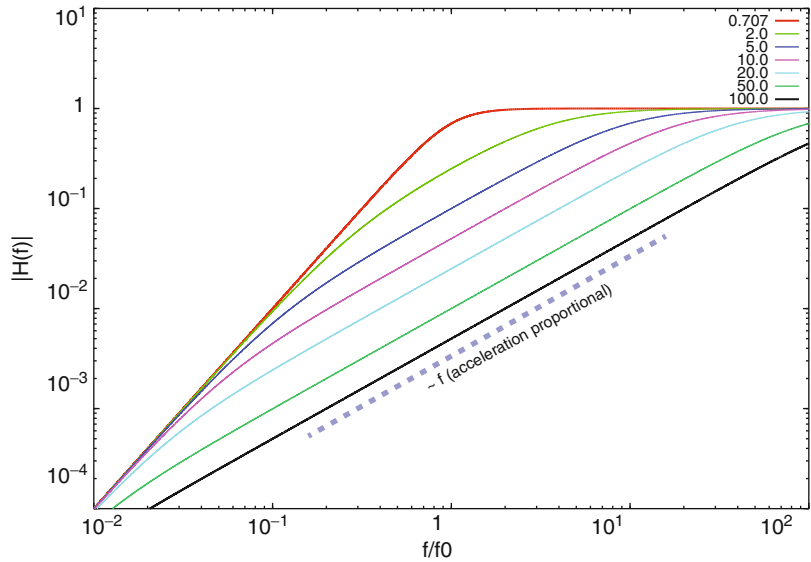
which, unfortunately, typically is found to be not less than a few Hz – a typical operation amplifier such as the OP-27 has a $1/f$ corner frequency of 2.7 Hz (Analog Devices 2006). Since the inverse filter method requires an increasing amount of amplification towards lower frequencies, intrinsic $1/f$ noise will be amplified overproportionally. Further, maintaining thermal stability of low-frequency circuits is not trivial. Thus, signal-to-noise ratio for low frequencies will be problematic. Also, the signal chain may take a prohibitively long time to recover from even a brief overload. Inverse filtering was applied in the 1960s–1980s but, owing to its drawbacks and the advent of superior methods, is no longer in practical use today.

Positive Feedback

As outlined previously, the bare-bones version of a geophone consists of a coil suspended by one or more springs, and moving in a magnetic field. When motion occurs, a voltage is induced across the coil's electrical terminals. However, fancier versions exist where a second coil is present. This coil is not a movable coil but is present so as to be able to feed an external current through it which, again through induction, will exert a force on the

Seismometer, Extended Response,

Fig. 3 Geophone frequency response for different(over) damping factors



suspended mass. This procedure is usually applied just for calibration (hence the name “calibration coil”) but can also be applied for positive feedback.

Hypothesizing the presence of such a coil, positive feedback could be applied by integrating the output signal (thus rendering it proportional to displacement) and feeding the integrated voltage back through the feedback coil (Lippmann 1982). When applied with the correct polarity, the feedback signal will “nudge” the mass so that the original movement will be amplified, thus counteracting the decay of sensitivity. The effect is similar to using a softer spring, leading to a smaller restoring force and, consequently, to a lower natural frequency. Like the inverse filter method, positive feedback has practical stability problems and is not in practical use nowadays.

Negative Feedback

Rather than integrating the output signal, this method differentiates it (thus rendering it proportional to acceleration of the suspended mass) and feeds it back through the second coil choosing the polarity so that it will not amplify but attenuate the mass movement. Ideally, the geophone mass would be damped so strongly that it does not perform any movement relative to the frame at

all; in other words, the mass is accelerated in exactly the same way as the Earth, and the current required to keep it steady with respect to the frame is proportional to acceleration for frequencies higher than the new corner frequency (it is not feasible to extend the bandwidth to arbitrarily low frequencies). Practical limitations of this method are two coils required (thus excluding cheap mass-produced exploration geophones) and stability problems of feedback electronics due to inductive coupling of signal and feedback coils (Lippmann 1982).

Broadband sensors also use negative feedback, but they use a capacitive transducer to pick up mass movement, thus avoiding the stability issues. However, it is a nontrivial undertaking to fit a capacitive pickup in an off-the-shelf exploration geophone, so this method, while promising, does not lend itself well to commercializing.

Response Extension Using NIC (Negative Impedance Converter)

Recalling Eq. 1, let us now extend the series of damping factors in Fig. 1 towards higher values, and we get this (Fig. 3).

It is evident that the higher the damping factor gets, the wider is the frequency range in which the output is proportional to acceleration. If a way can be found to achieve such high damping factors, the geophone will effectively be transformed into an accelerometer with a wide frequency range, albeit at the expense of output. As can be seen from Fig. 3, output decreases by several orders of magnitude in a certain frequency range.

Achieving Overdamping

Neglecting the mechanical contribution to damping (which is a reasonable simplification for typical geophones if a very high damping is required), we can limit our consideration to electrical damping which is proportional to the combined resistances of the generator coil and the external shunt resistor (Havskov and Alguacil 2004):

$$h = \frac{CDR}{R_{coil} + R_{shunt}} \quad (3)$$

where CDR is the total resistance needed to get $h = 1$ (critical damping resistor).

As we can see, the maximum amount of damping that can be achieved in a bare-bones geophone is when the signal coil is shorted ($R_{shunt} = 0$). In this case, only the intrinsic coil resistance is responsible for damping. Loading the coil with any external resistance at all will only decrease damping. In order to arrive at damping factors higher than short-circuit damping, a negative resistance would have to be used – clearly a feat that cannot be achieved using passive electronics. An NIC (negative impedance converter) provides exactly that functionality (Lippmann 1982; Ulmann 2005). In its simplest form, it is just an operational amplifier and two resistors.

A practical value for the negative resistance is 80 % of the coil resistance (Ulmann 2005). Completely compensating the coil resistance with close to 100 %, negative resistance is not practical for reasons of stability.

Practical Considerations

Linearity

Conventional geophones are plagued with linearity issues on account of the inherent nonlinearity of the suspension spring(s). Positive feedback would magnify this type of problems, whereas negative feedback (by overdamping) will actually alleviate them. Since, in a first-order approximation, the mass will remain steady, the suspension springs will be flexed only very little. Since spring nonlinearity increases with the amount of flexing or elongation, this is a very efficient means of suppressing spring-induced nonlinearity.

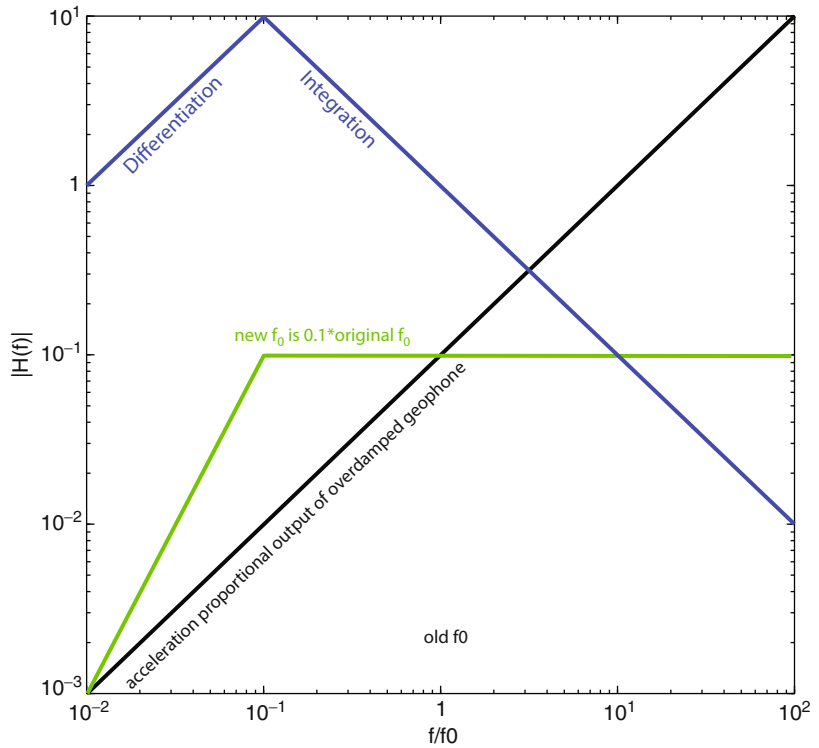
Amplification

Since, as seen in Fig. 3, output all across the acceleration proportional passband is significantly lower than the velocity proportional signal of the non-overdamped geophone, additional amplification is called for. However, it is not convenient to amplify the acceleration proportional signal since this would create dynamic range problems. Applying linear amplification by, say, a factor of 1,000 would lead to impractically large amplitudes for the high frequencies. It is a much more practical approach to apply amplification only after the acceleration proportional signal has been converted back to velocity proportional (see next section).

Converting the Output Signal Back to Velocity

The obvious solution of simply integrating the acceleration proportional signal back to velocity is not a very practical one. As can be seen in Fig. 3, the low-frequency roll-off of the standard damped seismometer is no longer seen in the overdamped transfer function, but there is still quite an amplitude gap to cover. The difference in amplitude is largest around the original corner frequency and decreases towards higher and lower frequencies, but it is especially towards lower frequencies that trouble is to be expected if one were to attempt a simple integration without any frequency band limitation. As an

Seismometer, Extended Response, Fig. 4 Design of integrating/differentiating filter



example, let's take a look at $0.01 f/f_0$. The output of the standard damping seismometer is down to 10^{-4} of its plateau output, and the overdamped (acceleration proportional) output is down even further, already off scale. Bringing these low frequencies back to "flat" would lead to extremely high amplification factors, leading to the same type of problems that were described above in "Using an Inverse Filter." Rather than simulating an extremely broadband seismometer, a band-limited approach is a lot more promising. To this end, a filter is needed that integrates frequencies above its corner frequency and differentiates those below. Figure 4 (after Lippmann 1982) schematically shows its functional principle. With a frequency proportional rise for $f < f_{0\text{NEW}}$, and a frequency proportional decay for $f > f_{0\text{NEW}}$, the result of multiplying the frequency proportional acceleration signal with the filter's transfer function will be exactly the transfer function of a geophone with a virtual natural frequency of $f_{0\text{NEW}}$. However, in order to arrive at an output

level comparable to the non-overdamped geophone, the filter must provide more amplification than shown in our schematic example. The exact amount of peak amplification depends on the ratio between the original and the new virtual natural frequency.

A filter with the desired characteristics (6 dB/octave roll-off on either side of a given frequency) is not difficult to implement – it is basically a combination of a first-order high-pass and a first-order low-pass filter, with both corner frequencies being set to the desired new f_0 and of course an amplification stage.

Environmental Influences

Except for a few broadband models, seismometers will typically be used in field scenarios, i.e., they have to work under adverse environmental conditions, and will have to maintain their characteristics over a wide range of temperatures.

The negative feedback principle outlined above basically amounts to measuring one small

physical entity (in our case, the current flowing through the generator coil when the coil starts moving), then adding a counteracting entity of similar magnitude (in our case, a current that will create a force to keep the coil steady). Effectively two similar entities are subtracted from each other. Consequently, if one or both of the two entities are susceptible to changes in environmental parameters, the result of the subtraction will also be susceptible to changes. The coil resistance is a “real” resistor consisting of many windings of copper wire whose value will increase with temperature. The negative impedance is implemented by active circuitry which will not be immune to temperature changes but will certainly react differently from the purely ohmic coil resistance. The following example from Lippmann (1982) illustrates the problem, focusing only on the coil.

A coil resistance of $600\ \Omega$ at $20\ ^\circ\text{C}$ will change to $640\ \Omega$ at $40\ ^\circ\text{C}$. If our negative resistance is $520\ \Omega$, the effective damping resistance will change from $80\ \Omega$ ($20\ ^\circ\text{C}$) to $120\ \Omega$ at $40\ ^\circ\text{C}$ – a 50 % increase, leading to drastic changes in the sensor’s characteristics.

Clearly, temperature compensation needs to be applied. Lippmann (1982) describes the measures and results.

Electronic Noise

Electronic noise is the prime limiting factor for how much the response can be extended since it defines the smallest detectable ground velocity amplitude. Electronic noise is generally frequency dependent. As outlined previously in the “Inverse Filter” entry, below a certain threshold frequency (which depends on the particular active electronic element, e.g., an operational amplifier), noise tends to be proportional to $1/f$. In other words, lower frequencies tend to exhibit more noise.

In an extended response seismometer using negative feedback and a subsequent integrating/differentiating filter, electronic noise can be considered separately for the acceleration proportional part (i.e., the combination of geophone and NIC) and for the filter/amplifier part. Lippmann (1982) reveals that the contribution of the

latter part is close to negligible, compared to the NIC’s contribution.

The only thing of relevance to the practitioner, though, is the composite noise, expressed not in electrical but in seismological units, i.e., in equivalent ground velocity or acceleration. The following slightly edited diagram from Wielandt (1991) shows the equivalent RMS velocity for two different types of Lennartz (2013) seismometers – the LE-3Dlite type based on physical 4.5 Hz geophones and converted to 1 Hz seismometers and the LE-3D/5 s type based on physical 2 Hz geophones and converted to 0.2 Hz (5 s) seismometers. For reference, the NLNM (New Low Noise Model, Peterson (1993)) and the noise level at the BFO (Black Forest Observatory) site are also given. Note that the noise levels have been calculated for frequency bins of constant relative bandwidth. For the 1/3 octave bandwidth shown here, the relative bandwidth is 23.2 %. For example, the 10 Hz bin is 2.32 Hz wide, whereas the 50 Hz bin is 11.6 Hz wide. Using constant relative bandwidth bins is best suited for comparing and converting noise measurements (Bormann 2012, Chapter 4.1) (Fig. 5).

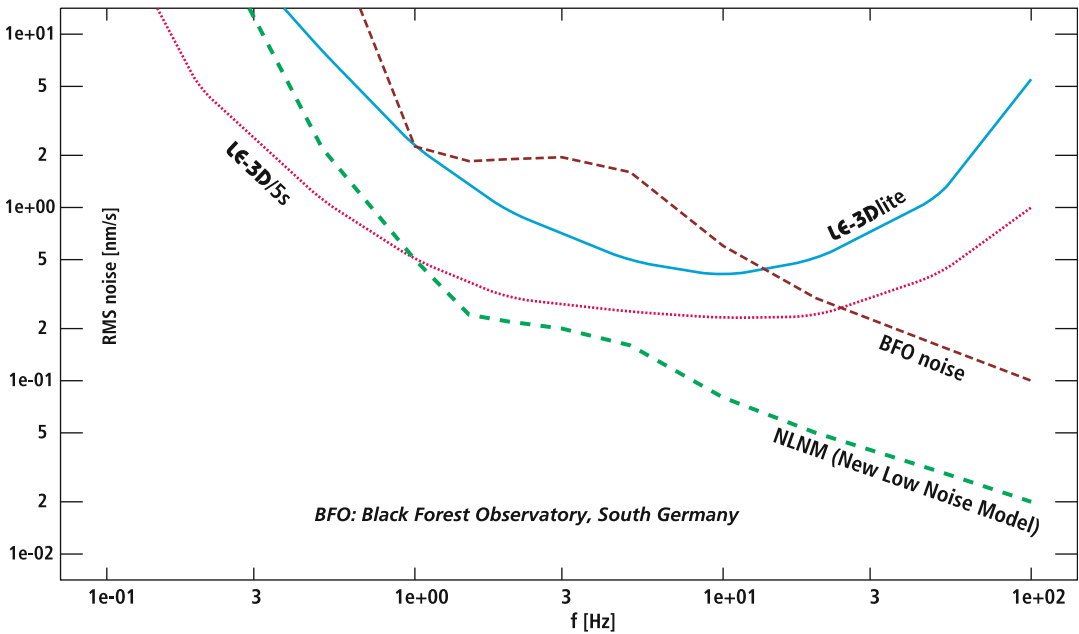
It turns out that both seismometer types are able to resolve ground noise at a very quiet site like BFO over a wide range of frequencies. The vast majority of sites will exhibit much more pronounced noise than BFO, thus rendering the extended response seismometers useable over their full frequency range.

A Practical Implementation

The photo shows a classical, purely mechanical seismometer next to an electronically enhanced seismometer built around a 4.5 Hz geophone. In both cases, natural frequency is one hertz, and both are single-component instruments (Fig. 6).

Summary

Range-extended seismometers based on robust, readily available, and comparatively cheap



Seismometer, Extended Response, Fig. 5 Lε-xD Self Noise in 1/10 decade (1/3 octave)

Seismometer, Extended Response, Fig. 6 Classical Geophone (left) vs. Range-Extended Seismometer (right), both of them single-component, 1 Hz units



exploration-grade geophones provide a viable alternative to clumsy and delicate mechanical seismometers for short- and intermediate-period applications. Unlike typical broadband sensors,

range-extended seismometers do not contain physical elements tuned to very low frequencies and do not respond to very low frequencies. Consequently they usually require no shielding from

pressure and temperature changes and provide extended high-frequency response, typically up to and above 100 Hz. Inherent limitations such as electronic noise can be mitigated by proper design and implementation.

Cross-References

- [Principles of Broadband Seismometry](#)
- [Recording Seismic Signals](#)
- [Seismic Noise](#)
- [Seismometer Self-Noise and Measuring Methods](#)
- [Sensors, Calibration of](#)

References

- Analog Devices (2006) OP27 datasheet, downloadable from http://www.analog.com/static/imported-files/data_sheets/OP27.pdf
- Bormann P (ed) (2012) New manual of seismological observatory practice (NMSOP-2), IASPEI. GFZ German Research Centre for Geosciences, Potsdam. <http://nmsop.gfz-potsdam.de>
- Havskov J, Alguacil G (2004) Instrumentation in earthquake seismology. Springer, Dordrecht
- Lennartz electronic (2013) LE-xD seismometer family (Document 990-0003), PDF document downloadable from <http://www.lennartz-electronic.de>
- Lippmann E (1982) Umwandlung eines elektrodynamischen Seismometers in einen Beschleunigungsaufnehmer und Entwicklung eines Filters zur Simulation eines niederfrequenten Seismometers. Diploma thesis, Ludwig-Maximilians-Universität München
- Peterson J (1993) Observations and modeling of seismic background noise. US geological survey open file report 93-322
- Rodgers P (1993) Maximizing the signal-to-noise ratio of the electromagnetic seismometer: the optimum coil resistance, amplifier characteristics, and circuit. B Seismol Soc A 83:561–582
- Ulmann B (2005) Overdamping geophones using negative impedances. http://www.vaxman.de/publications/teach_gp.pdf
- Wielandt E (1991) Messungen der kurzperiodischen Bodenunruhe in Schiltach (BFO). Internal report, published in German only. PDF of original text and English translation available from the author of the present article
- Wielandt E, Streckeisen G (1982) The leaf-spring seismometer: design and performance. B Seismol Soc A 72:2349–2367

Selection of Ground Motions for Response History Analysis

Anastasios G. Sextos

Division of Structural Engineering, Department of Civil Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece

Synonyms

Earthquake records; Ground motion; Record scaling; Response history analysis

Introduction

The evolution in computational power and the parallel processing capabilities of modern engineering software make nowadays the use of complicated structural analysis methods an attractive alternative for the design and assessment of structures. In contrast to the past, when the elastic static analysis was almost exclusively used for the seismic design of structures, the state of practice has progressively moved toward dynamic-elastic, nonlinear-static (i.e., single mode or multi-modal “pushover”), and even nonlinear response history analysis. The latter, capturing more efficiently the hierarchy of failure mechanisms, the energy dissipation, the force redistribution among the structural members, and contact issues (such as gap, impact, sliding, and uplift), is deemed preferable in cases of significant material or geometrical nonlinearities and, as such, is used for the design of seismically isolated buildings and bridges or the assessment of existing structures with various degrees of damage. Elastic response history analysis is also extensively used, primarily for structures whose response is dominated by higher modes (mostly tall and irregular buildings and towers) or structures of high importance that are typically designed to remain elastic even for long return-period earthquake intensities (i.e., industrial facilities, power plants, dams, critical administrative buildings, etc).

In all cases, the main task of the design procedures is to achieve more predictable and reliable levels of safety and operability against different levels of seismic intensity, a framework known as performance-based design and assessment. Despite the above major advances made in terms of structural analysis, the reliability of the analysis output and the subsequent structural performance prediction strongly depend on the decisions made for the selection of the seismic input which is used as ground excitation. Research has shown that among all possible sources of uncertainty stemming from structural and soil material properties, the modeling approximations, and the design and analysis assumptions as well as the earthquake-induced ground motion, the latter has by far the highest effect on the variability observed in the structural response (Elnashai and McClure 1996; Padgett and Desroches 2007; Shome et al. 1998). Therefore, the selection of a “reliable” suite of earthquake ground motions constitutes an important prerequisite for the reliability of the structural analysis procedure as a whole.

Along these lines, numerous computational methods and tools have been developed for (a) selecting suites of earthquake records from available strong ground motion record databases (b) generating synthetic and artificial ground motions or (c) modifying existing ground motions until they present desirable target characteristics.

Objectives of Ground Motion Selection

Currently more than 40 methods exist for selecting, modifying, and scaling earthquake ground motions. These methods can be grouped conceptually by their objective as follows based on the Ground Motion Selection and Modification (GMSM) Program of the Pacific Earthquake Engineering Research (PEER) Center (Haselton 2009) in the following broad four categories:

- *Predict the Probability Distribution* (mean and dispersion) of *Structural Response* (i.e., engineering demand parameter, EDP) from ground motions that comply with an

earthquake scenario of *given magnitude, M , source-to-site distance, R , and, in some cases, faulting type or soil class (i.e., $V_{s,30}$) at the site of interest.*

- *Predict the Median or Mean of Structural Response* from ground motions that are selected (or generated) to *match a median target response spectrum* for a given M-R pair, obtained from a ground motion prediction equation (or attenuation relationship). Dispersion of structural response might also be of interest; however, minimizing the standard error of response quantities is not explicitly envisaged.
- *Predict the Probability Distribution* (mean and dispersion) of *Structural Response* using ground motions that satisfy a given *spectral acceleration at the fundamental period* of the structure, $S_a(T_1)$, that has resulted from an associated M and R (as well as fault mechanism and soil class or $V_{s,30}$) scenario. Notably, a priori knowledge of the structural properties (i.e., its fundamental period T_1) is a prerequisite.
- *Predict the Median or Mean of Structural Response* for a given set of $S_a(T_1)$, M , and R . In this case, the scenario refers to a Maximum Considered Earthquake $S_a(T_1)$ and not a MCE response spectrum.

A fifth objective can also be distinguished to *minimize the structural response discrepancy* by considering a threshold confidence level for the standard error of the response quantities, additional to conventional spectral matching (Katsanos and Sextos 2013). This involves consideration of structural analysis results within the GSM process.

Methodologies for Ground Motion Selection and Scaling

Independently of the ultimate objective of the GSM process, the various methodologies for developing earthquake ground motion sets for (linear or nonlinear) dynamic analysis of structures can also be classified based on the concept

of selection and the procedure of modulating or scaling ground motions. Naturally, method objectives and selection and scaling procedures are not fully correlated. The following classes of methods are listed below. A detailed review of most important methods is also performed elsewhere (Katsanos et al. 2010):

Selection by M and R and Scaling to a Target Intensity Measure

The simplest GSM method involves the formation of set (bin) of motions that satisfy, as close as possible but without explicit constraints, preliminary magnitude, M , source-to-site distance, R , and often seismotectonic or soil class criteria. It is recalled that the total number of sets N_{tot} that can be formed, from m , potentially eligible, records out of a larger group of s records, can be calculated by the following factorial formula of the binomial coefficient:

$$N_{\text{tot}} = \binom{s}{m} = \frac{s!}{m!(s-m)!} \quad (1)$$

Once eligible ground motions are selected, their accelerations are multiplied by a scaling factor to match a target intensity measure (IM), typically being the peak ground acceleration (PGA) or the spectral acceleration at the fundamental period of the structure, $S_a(T_1)$. In the latter case, it is evident that all ground motion spectra will have identical ordinates at period T_1 and different spectral accelerations in all other periods. More elaborative IMs have also been proposed, involving the spectral shape and the structural characteristics. Such IMs are expected to result into a relatively more accurate prediction of the seismic demand (Baker and Cornell 2005; Luco and Cornell 2007; Tothong and Luco 2007); however, the approximate definition of the seismic scenario is a significant limitation.

Selection and Scaling to a Target Uniform Hazard Spectrum

In the light of the above criteria, it is also common to envision matching of the response spectra of the selected acceleration time series with

a target, Uniform Hazard Spectrum (UHS) (Kramer 1996; American Society of Civil Engineers 2005; McGuire 2004), which is determined from (a) a ground motion prediction relationship, (b) a seismic hazard assessment for the site of interest, or (c) the seismic code provisions. In this case, ground motions are scaled with a single or individual scaling factors so that their individual spectrum, or the mean of their response spectra closely match and in fact exceed, in terms of “shape,” the ordinates of the target UHS, typically within a given range of periods around the fundamental period of the structure. When a uniform scaling factor is sought, it may be determined through the following expression:

$$sf_{\text{avg}} = \left\{ \min \left(\frac{Sa_{\text{avg}}(T_i)}{Sa_{\text{target}}(T_i)} \right) \right\}^{-1}, \quad i = 1 \text{ to } N \quad (2)$$

where T_i is the sample period and N is the size of the sample into which the prescribed period range is discretized.

Quite commonly, all eligible suites are ranked according to their “goodness-of-fit” to the target spectrum, as quantified by the normalized root-mean-square-error, δ , between the scaled average Sa_{avg} , and the target spectrum Sa_{trt} (Iervolino et al. 2010b):

$$\delta = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N \left(\frac{Sa_{\text{avg}}(T_i) - Sa_{\text{trt}}(T_i)}{Sa_{\text{trt}}(T_i)} \right)^2} \quad (3)$$

A wide variety of similar expressions have also been used in the literature. Further discussion on the efficiency of various spectral matching indicators can be found elsewhere (Beyer and Bommer 2007; Jayaram et al. 2011; Kottke and Rathje 2008; Naeim et al. 2004).

In addition to spectral shape, these methods may consider other earthquake, site, or ground motion parameters in selecting ground motions. Finally, the mean or the maximum (depending on the number of ground motions within the set) of the response quantities is used as the design value.

Seismic Codes of Practice

Most contemporary seismic codes and design recommendations, such as Eurocode 8 for buildings (CEN 2004) and bridges (CEN 2005), ASCE standards 7-10 (ASCE/SEI 2010) and FEMA P-750 (FEMA 2009a) as well as various national norms (New Zealand Standards, Italian Code, and Greek Seismic Code), describe relatively similar procedures that are based on spectral matching for the selection of earthquake ground motions in the framework of dynamic analysis of structures. In most cases, seismic motions can be represented by real, artificial, or simulated records, typically complying with the aforementioned preliminary criteria of earthquake magnitude, distance, the seismotectonic environment, and the local soil conditions. Differences among the codes can be summarized as follows:

Duration Strong-motion duration is not explicitly considered as an additional selection parameter in most documents inclusive of ASCE standards 7-10 (ASCE/SEI 2010) and the Eurocodes.

Simulated and Artificial Ground Motions Use of simulated ground motions is permitted in a number of seismic codes; however, it is restricted only to cases of inadequate number of real accelerograms in others, i.e., FEMA P-750 and Eurocode 8 Part 2 for bridge design (CEN 2005).

Near-Field Considerations Near-field effects are typically either ignored or considered in a quantitative manner in terms of direction of fault rupture and velocity pulses, an example being FEMA P-750 (FEMA 2009a). More detailed distinction is made in FEMA P695 (FEMA 2009b) by forming distinct far-field and near-field record sets. ASCE standards 7-10 prescribe different design quantities for sites within 3 miles (5 km) of the active fault that controls the hazard. In this case, each pair of ground motion components shall be rotated to the fault-normal and fault-parallel directions of the causative fault. Next, it shall be appropriately scaled so that the

average of the fault-normal components is not less than the MCE response spectrum for the period range of spectral matching (as described below).

Spectral Matching A distinction is typically made between two-dimensional and three-dimensional analysis, the latter involving selection and application of pairs of records for linear or nonlinear dynamic analysis. In principle, the procedure is similar among most important seismic codes and recommendations and involves the following steps: (a) the 5 %-damped elastic spectra are derived for each component of the eligible horizontal motions selected; (b) the mean of the individual (EC8-Part 1) or the square root of the sum of squares (SRSS spectra) of the individual elastic spectra ordinates is determined; (c) the mean spectrum of the ground motions is compared with the target code-prescribed spectrum; and (d) records are scaled so that the spectral ordinates of the mean spectrum (either from individual or SRSS spectra) exceed a lower bound of the target spectral acceleration (Table 1) within a prescribed period range (also depicted in Table 1).

Scaling Scaling of the individual records toward spectral matching can be performed either with the use of a uniform scaling factor (i.e., EC8-Part2, ASCE 7-10) or a record-dependent, individual scaling factor (FEMA P-750). Other codes (i.e., EC8-Part1) do not provide specific guidelines regarding the scaling of seismic records in order to establish the required compliance with the target design spectrum. New Zealand Standards (Standards New Zealand (SNZ) 2004) on the contrary prescribe two distinct scale factors k_1 and k_2 , the first minimizing the difference between recorded and target response spectrum in a least mean square scheme over the period range of interest, while the latter ensures that the energy in the spectrum of the strongest ground motion exceeds the energy of the target spectrum.

Design Quantities Structural analysis results are processed in a statistical manner, and the

Selection of Ground Motions for Response History Analysis, Table 1 Earthquake records selection and spectral matching criteria prescribed in the seismic codes and guidelines studied here

Seismic codes and guidelines	Selection criteria	Ensemble spectrum	Spectral matching period range	Lower bound of mean spectral acceleration	Scaling factor
EC8 Part 1 (buildings)	Seismotectonic features, soil type	Mean of individual spectra	$0.2-2.0 T_1$	$0.90 \cdot Sa(T)_{\text{design}}$	Not specified
EC8 Part 2 (bridges)	Source mechanism, M, R	Mean of SRSS spectra	$0.2-1.5 T_1$	$1.30 \cdot Sa(T)_{\text{design}}$	Uniform
FEMA P-750	Source mechanism, M, R	Mean of SRSS spectra	$0.2-1.5 T_1$	$1.17 \cdot Sa(T)_{\text{design}}$	Individual
ASCE/SEI 7-10	Source mechanism, M, R	Mean of SRSS spectra	$0.2-1.5 T_1$	$1.00 \cdot Sa(T)_{\text{design}}$	Uniform
NZS 1170.5	Seismotectonic features, soil type	Individual record scaling and matching	$0.4-1.3 T_1$	Matching as nearly as practicable	k_1 and k_2 factors
Taiwan seismic code	Source mechanism, M, R	Mean of individual spectra	$0.2-1.5 T_1$	$0.90 \cdot Sa(T)_{\text{design}}$	Not specified

design quantities are defined as either the mean (in case of seven records or more) or the maximum (in case of three records) of the engineering demand parameters of interest (EC8, FEMA P-750). Certain codes solely provide the option of obtaining the maximum response out of the three response history analyses (CPA 2011; Standards New Zealand (SNZ) 2004). The most refined to date provision to derive seismic design values is prescribed in ASCE 7-10 clearly distinguishing between (a) *force response parameters* that shall be multiplied for each ground motion analyzed by a factor I_e/R , where I_e is the importance factor and R is the Response Modification Coefficient, and (b) *drift response quantities*, which shall be multiplied by C_d/R , C_d being the deflection amplification factor. It is noted herein that the above methods refer to the design of new buildings according to modern seismic codes and do not apply to the probabilistic assessment of the nonlinear response of existing buildings (Jalayer and Cornell 2009). It is also noted that despite the simplification and applicability of the above code-based procedures, significant limitations also exist (Sextos et al. 2010).

More details on comparative seismic code provisions on ground motion selection and scaling can be found elsewhere (Hachem et al. 2010).

Selection and Scaling to a Target Conditional Mean Spectrum

An alternative group of methods for ground motion selection and scaling is similar to the aforementioned spectral matching, but the Uniform Hazard target Spectrum, the Conditional Mean Spectrum (CMS) (Baker 2011), is used instead. The main reason behind this substitution (Kottke and Rathje 2008; Lin et al. 2013a; Wang 2010) is that the UHS is shown to be unsuitable to serve the main purpose of dynamic analysis, which is to excite the structure under consideration with ground motions having a specified spectral acceleration at a given period. In fact, UHS compatible ground motions are often associated with large-amplitude spectral values in a wide range of periods. Furthermore, UHS can hardly be considered as a spectrum of a single earthquake event as it rather represents an envelope of spectra corresponding to different seismic events and sources. On the contrary, a Conditional Mean Spectrum (CMS) represents the expected (i.e., mean) response spectrum, *conditioned* on the occurrence of a target spectral acceleration at the period of interest, typically, though not exclusively, the fundamental period of the structure T_1 .

Similarly to the UHS, the target CMS is calculated for the scenario $S_a(T_1)$, magnitude, M , and source-to-site distance, R , together with

other preliminary criteria such as fault type or soil, if desirable. In simpler words, the CMS represents the expected response spectrum for the defined ground motion scenario, which is based on a target $S_a(T_1)$ at a single structural period, in contrast to the UHS, which represents equally rare $S_a(T)$ values at many periods (including T_1) simultaneously (Haselton 2009). To develop this new target spectrum, Probabilistic Seismic Hazard Assessment can be used to determine the spectral acceleration $S_a(T_1)$ that corresponds to the target probability of exceedance at the site of interest, denoted as $S_a(T_1)^*$. De-aggregation can then be used to estimate the mean values of magnitude, source-to-site distance, and “epsilon” (“ ϵ ”) ($\bar{M}$, $\bar{R}$, $\bar{\epsilon}$) that lead to an acceleration equal to $S_a(T_1)^*$. The definition of parameter “epsilon” is provided in the following section. The way in which spectral matching to the target CMS is achieved is similar to the procedures described above.

Selection and Scaling to a Proxy (“ ϵ ”) of the Target Conditional Mean Spectrum

A more advanced approach category of methods for ground motion selection and scaling to a target Conditional Mean Spectrum utilizes the parameter “epsilon” (“ ϵ ”) as a proxy of the desirable CMS spectral shape (Baker and Cornell 2006). It is recalled that “epsilon” was first defined by engineering seismologists as the number of standard deviations by which a given spectral acceleration, expressed in logarithmic terms, differs from the mean logarithmic spectral acceleration provided by a ground motion prediction (attenuation) equation. In other words, “ ϵ ” is derived by subtracting the predicted mean logarithmic spectral acceleration at given period T_1 ($\ln\{S_a(T_1)\}$) from the corresponding value ($\ln\{S_a(T_1)\}$) of the record under examination and then dividing by the logarithmic standard deviation estimated by the attenuation relationship. In practical terms, this implies that a record with “ ϵ of 1.5 at 0.5 s” has a spectral acceleration at a period $T_1 = 0.5$ s that is 1.5 standard deviations higher than the predicted mean spectral value.

An advantage of the “ ϵ ” parameter is that it is determined with respect to the unscaled record

and does not change in case of record scaling. On the other hand, it is also noticeable that for a given ground motion record, “ ϵ ” is clearly a function of the period of interest and depends on the particular ground motion prediction model used, since different attenuation relationships lead to different mean and standard deviation of $\ln\{S_a(T)\}$. Therefore, it is important to ensure that the ground motion prediction model used to compute ϵ is the same model used in the ground motion hazard assessment. This dependence of ϵ to attenuation relationships is perhaps the most important drawback in the use of this parameter.

Ground motion selection methods (Tothong and Luco 2007) that use “ ϵ ” as a proxy of the CMS spectral shape ensure that the value of “ ϵ ” at the fundamental period of the building, denoted as $\epsilon(T_1)$, is as close as possible to the target $\epsilon(T_1)$ of the ground motion scenario. This practically implies that the record-to-record variability at the fundamental period of the structure is reduced, and in turn, the discrepancy in the structural response of the corresponding SDOF system is also lower. Some concerns have been expressed regarding the appropriate choice of CMS and the conditioning period, primarily related to the importance of higher modes of vibration of MDOF systems and the anticipated period elongation under strong ground motions (Katsanos et al. 2014). However, it has been shown that risk-based assessments are relatively insensitive to the choice of conditioning period provided that the ground motions are carefully selected to ensure hazard consistency (Lin et al. 2013a).

Available Computational Tools and Databases

Strong Ground Motion Databases

Among numerous strong ground motion databases in Japan, Taiwan, and Europe (European Strong Ground Motion database, www.isesd.hi.is), the PEER-NGA Next Generation Attenuation strong-motion database is a continuously developing project currently consisting of 3,551 publicly available, three-component seismic records (i.e., about 10,650 individual earthquake

acceleration time series) that have been recorded during 173 shallow crustal earthquakes from active tectonic regions worldwide. The corresponding seismic events, which have been recorded primarily in California, range in magnitude from 4.2 to 7.9 and cover epicentral distances in the range 0.2–600 km. Apart from the magnitude and the distance, the earthquake database contains basic information about the seismic source including date and time of the event, hypocenter location, faulting mechanism, seismotectonic environment, and others. Detailed data about 1,600 strong-motion stations are also provided (i.e., site characterizations, surface geology, shallow subsurface conditions, the location of the instrument inside the structure's installation place). Furthermore, each acceleration time history has been corrected for the response of the strong-motion instrument itself and filtered out the noise included while it can also be automatically scaled online.

Software and Tools for Ground Motion Selection

Given the above extensive repository of earthquake records and the fact that the most common earthquake record selection procedures involve spectral matching, recent work evolved to develop computational tools for quantifying and/or optimizing spectrum compatibility to a code prescribed or CMS (Youngs et al. 2007). REXEL (Iervolino et al. 2010a, 2011, 2012) was the first, all-in-one, software introduced for this purpose and facilitates the search for suites of waveforms compatible to target spectra that are either user defined or automatically generated according to Eurocode 8 and the recently issued Italian seismic code. An alternative web-based software for earthquake record selection is SeleQ (Dias 2010), offering various filtering options. More recently, the Integrated System for Structural Analysis and Record Selection (ISSARS) software has been developed (Katsanos and Sextos 2013), retrieving dynamically ground motions from the PEER-NGA database, to form suites of records that not only comply with specific criteria but also ensure, through numerical analyses of the structure that

run at the background, a target level of dispersion of structural response quantities.

Accepted Knowledge on Ground Motion Selection for Response History Analysis Purposes

- Structural response is inherently probabilistic in nature due to the variability among ground motions.
- A major challenge in choosing the most appropriate method for ground motion selection and scaling is to understand the purpose of the dynamic analysis for which the ground motions are sought. A major distinction is between design and assessment purposes. Careful selection of methods, intensity measures, and engineering demand parameters is necessary.
- An additional challenge, at least for a design viewpoint, is to keep structural response discrepancy low.
- Numerous GSM methods exist in the literature with distinct advantages and drawbacks. The direct comparison of these methods is not always feasible.
- Most GSM methods adopted in contemporary seismic codes are rather simplified compared to the breadth of the existing methods available in the literature. Still, they require significant effort, deep understanding of the physical problems and the parameters involved as well as specialized computational tools in order to overpass their inherent limitations.

Summary

Linear and nonlinear dynamic analysis of structures is becoming increasingly popular in structural design and assessment practice. Along these lines, the selection and scaling of the appropriate set of earthquake ground motions, required in the framework of dynamic analysis, has become of paramount importance due to the significant sensitivity of structural response to the assumptions made in forming the necessary set of earthquake

records. A large number of ground motion selection and modification (GMSM) methods have been proposed in the literature; still though, the major progress made during the last decade is not yet reflected in modern seismic codes. Along these lines, the present entry aims at briefly presenting, in the simplest possible terms but not simpler than necessary, the objectives and fundamental concepts of GMSM methods along with the current seismic code framework and the computational tools developed to facilitate code-prescribed procedures.

Cross-References

- [Assessment of Existing Structures Using Response History Analysis](#)
- [Conditional Spectra](#)
- [Engineering Characterization of Earthquake Ground Motions](#)
- [Nonlinear Dynamic Seismic Analysis](#)
- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)
- [Probabilistic Seismic Hazard Models](#)
- [Response-Spectrum-Compatible Ground Motion Processes](#)
- [Seismic Collapse Assessment](#)
- [Spatial Variability of Ground Motion: Seismic Analysis](#)
- [Stochastic Ground Motion Simulation](#)
- [Time History Seismic Analysis](#)

References

- American Society of Civil Engineers (2005) Seismic design criteria for structures, systems, and components in nuclear facilities, Structural Engineering Institute, Working Group for Seismic Design Criteria for Nuclear Facilities, ASCE/SEI 43-05, Reston, VA, 81 pp
- ASCE/SEI (2010) Minimum design loads for buildings and other structures: ASCE standard 7-10. American Society of Civil Engineers/Structural Engineering Institute, Reston
- Baker JW (2011) The conditional mean spectrum: a tool for ground motion selection. *J Struct Eng* 137(3):322–331. doi:10.1061/(ASCE)ST.1943-541X.0000215
- Baker JW, Cornell CA (2005) A vector-valued ground motion intensity measure consisting of spectral acceleration and epsilon. *Earthq Eng Struct Dyn* 34:1193–1217. doi:10.1002/eqe.474
- Baker JW, Cornell CA (2006) Spectral shape, epsilon and record selection. *Earthq Eng Struct Dyn* 35(9):1077–1095. doi:10.1002/eqe.571
- Beyer K, Bommer JJ (2007) Selection and scaling of real accelerograms for bi-directional loading: a review of current practice and code provisions. *J Earthq Eng* 11:13–45. doi:10.1080/13632460701280013
- CEN (2004) European Standard EN 1998-1. Eurocode 8: design of structures for earthquake resistance, part 1: general rules, seismic actions and rules for buildings, committee for standardization. Design, vol 3. European Committee for Standardization, Brussels
- CEN (2005) European Standard EN 1998-2. Eurocode 8: design of structures for earthquake resistance – part 2 bridges, committee for standardization, vol 3. European Committee for Standardization, Brussels
- CPA (2011) Seismic design code and commentary for buildings, construction and planning agency. Ministry of Interior Affairs, Taipei (in Chinese), pp 4–51
- Dias J (2010) SelEQ: a web-based application for the selection of earthquake ground motions for structural analysis. In: The 14th European conference on earthquake engineering, Ohrid, 30 Aug–3 Sept
- Elnashai AS, McClure DC (1996) Effect of modelling assumptions and input motion characteristics on seismic design parameters of RC bridge piers. *Earthq Eng Struct Dyn* 25(5):435–463
- FEMA (2009a) NEHRP recommended seismic provisions for new buildings and other structures, FHMA 750. Building Seismic Safety Council, Washington, DC
- FEMA (2009b) Quantification of building seismic performance factors, FEMA P695. Federal Emergency Management Agency, Washington, DC
- Hachem MM, Mathias NJ, Wang YY, Fajfar P, Tsai K-C, Ingham JM, ... Francisco S (2010) An international comparison of ground motion selection criteria for seismic design. In: Codes in structural engineering, developments and needs for international practice, joint IABSE – fib conference. Dubrovnik, pp 237–250
- Haselton CB (2009) Evaluation of ground motion selection and modification methods: predicting median interstory drift response of buildings, PEER report, 2009/01
- Iervolino I, Galasso C, Cosenza E (2010a) New features of REXEL 2. 61 beta, a tool for automated record selection. In: Bulletin of earthquake engineering, vol 8. Ohrid, 30 Aug–3 Sept, pp 339–362. doi:10.1007/s10518-009-9146-1
- Iervolino I, Galasso C, Cosenza E (2010b) REXEL: computer aided record selection for code-based seismic structural analysis. *Bull Earthq Eng* 8:339–362. doi:10.1007/s10518-009-9146-1
- Iervolino I, Galasso C, Paolucci R, Pacor F (2011) Engineering ground motion record selection in the Italian

- ACcelerometric Archive. Bull Earthq Eng 9(6):1761–1778. doi:10.1007/s10518-011-9300-4
- Iervolino I, Galasso C, Chioccarelli E (2012) REXEL 3.3: closing the loop of computer aided record selection. In: The 15th world conference on earthquake engineering, vol 10, Lisbon
- Jalayer F, Cornell CA (2009) Alternative nonlinear demand estimation methods for probability-based seismic assessments. Earthq Eng Struct Dyn 38(8):951–972
- Jayaram N, Lin T, Baker JW (2011) A computationally efficient ground-motion selection algorithm for matching a target response spectrum mean and variance. Earthq Spectra 27(3):797–815. doi:10.1193/1.3608002
- Katsanos EI, Sextos AG (2013) ISSARS: an integrated software environment for structure-specific earthquake ground motion selection. Adv Eng Softw 58:70–85. doi:10.1016/j.advengsoft.2013.01.003
- Katsanos EI, Sextos AG, Manolis GD (2010) Selection of earthquake ground motion records: a state-of-the-art review from a structural engineering perspective. Soil Dyn Earthq Eng 30(4):157–169. doi:10.1016/j.soildyn.2009.10.005
- Katsanos EI, Sextos AG, Elnashai AS (2014) Prediction of inelastic response periods of buildings based on intensity measures and analytical model parameters. Eng Struct 71:161–177. doi:10.1016/j.engstruct.2014.04.007
- Kottke AR, Rathje EM (2008) A semi-automated procedure for selecting and scaling recorded earthquake motions for dynamic analysis. Earthq Spectra 24(4):911–932
- Kramer SL (1996) Geotechnical earthquake engineering. Prentice Hall, Upper Saddle River, N.J., 653 pp
- Lin T, Haselton CB, Baker JW (2013a) Conditional-spectrum-based ground motion selection. Part II: intensity-based assessments and evaluation of alternative target spectra. Earthq Eng Struct Dyn. doi:10.1002/eqe
- Lin T, Haselton CB, Baker JW (2013b) Conditional-spectrum-based ground motion selection. Part I: hazard consistency for risk-based assessments. Earthq Eng Struct Dyn. doi:10.1002/eqe
- Luco N, Cornell CA (2007) Structure-specific scalar intensity measures for near-source and ordinary earthquake ground motions. Earthq Spectra 23(2):357–392. doi:10.1193/1.2723158
- McGuire RK (2004) Seismic hazard and risk analysis, Earthquake Engineering Research Institute, Berkeley, 240 pp
- Naeim F, Alimoradi A, Pezeshk S (2004) Selection and scaling of ground motion time histories for structural design using genetic algorithms. Earthq Spectra 20(2):413–426
- Padgett JE, Desroches R (2007) Sensitivity of seismic response and fragility to parameter uncertainty. J Struct Eng 133(12):1710–1718
- Sextos AG, Katsanos EI, Manolis GD (2010) EC8-based earthquake record selection procedure evaluation: validation study based on observed damage of an irregular R/C building. Soil Dyn Earthq Eng 1–15. doi:10.1016/j.soildyn.2010.10.009
- Shome N, Cornell CA, Bazzurro P, Carballo EJ (1998) Earthquakes, records and nonlinear responses. Earthq Spectra 14(3):469–500
- Standards New Zealand (SNZ) (2004) NZS 1170.5:2004 – Structural design actions. Earthquake actions. Standards New Zealand, Wellington
- Tothong P, Luco N (2007) Probabilistic seismic demand analysis using advanced ground motion intensity measures. Earthq Eng Struct Dyn 36:1813–1835. doi:10.1002/eqe
- Wang G (2010) A ground motion selection and modification method capturing response spectrum characteristics and variability of scenario earthquakes. Soil Dyn Earthq Eng 1–15. doi:10.1016/j.soildyn.2010.11.007
- Youngs RR, Power MS, Wang G, Makdisi F, Chin CC (2007) Design ground motion library (DGML) – tool for selecting time history records for specific engineering applications. In: Proceedings of SMIP07 seminar on utilization of strong-motion data, Sacramento

Sensitivity of First-Excursion Probabilities for Nonlinear Stochastic Dynamical Systems

Marcos Valdebenito and Hector A. Jensen
Department of Civil Engineering, Santa Maria University, Casilla, Valparaiso, Chile

Synonyms

First Excursion; Stochastic Dynamics; Sensitivity; Simulation Methods

Introduction

Quantification of the performance of structural systems subject to dynamic loading is of paramount interest in several fields of engineering and particularly in the case of earthquake engineering. Knowledge on the performance of a structure during seismic events allows taking design decisions that ensure its serviceability and safety throughout its life. Nonetheless, quantification

of performance is a challenging task as there is always uncertainty on future loadings that affect a structure during its lifetime. Structural reliability has emerged as a discipline that allows accounting for the unavoidable effects of uncertainty over performance. Thus, probability theory is used to describe the uncertainty associated with different relevant parameters that affect performance by means of random variables, random fields, and/or stochastic processes. In this manner, uncertainty is *propagated* from these *input* parameters to the responses of interest such as displacements, accelerations, forces, etc.

A particularly useful way to measure the effects of uncertainty in the dynamic response of structural systems is the so-called first-excursion probability. This probability is widely used in stochastic structural dynamics and measures the chances that one or more structural responses exceed a prescribed threshold level within the duration of a dynamical excitation (Soong and Grigoriu 1993). First-excursion probability estimation is particularly challenging as characterization of uncertain loading usually comprises stochastic processes whose discrete representation can involve hundreds or even thousands of random variables. Similarly, the number of possible failure criteria involved can be extremely large as well, i.e., there can be several responses of interest that must be controlled at a large number of discrete time instants. Hence, several different techniques have been proposed in order to estimate first-excursion probabilities. Among these, methods based on simulation (such as the Monte Carlo method and its more advanced variants) have been shown to be the most appropriate approach to compute these probabilities (Schuëller et al. 2004).

Although first-excursion probability provides a most useful way to rationally account for the effects of uncertainty on structures subject to stochastic loading, it is certainly not the only metric that should be taken into account when designing a system. In fact, it is also of interest analyzing the *sensitivity* of the probability with respect to variations in the properties of the

structural system. For example, determining the variation in probability due to a change in the size of a structural member can provide useful information to increase the safety level or to identify the most influential design parameters. Nonetheless, estimation of the sensitivity of first-excursion probabilities for dynamical systems is a challenging task as it comprises not only taking into account the uncertainty in input parameters but also assessing how performance is affected due to variations in properties of a structure.

This contribution presents an approach for assessing probability sensitivity of systems subject to stochastic excitation with emphasis on structures whose response is nonlinear. The approach combines state-of-the-art simulation strategies with a series of approximation concepts. Salient features of the approach are the capability of considering problems involving a large number of random variables (in the order of thousands), the possibility of estimating sensitivity with respect to several variables simultaneously (scalability), and a high numerical efficiency achieved by integrating an advanced simulation algorithm with local approximations of the functions modeling the structural performance.

Formulation of the Problem

Structural Model

Let $f(t)$ be a scalar representing loading acting over a structure during a time span $t \in [0, T]$. This load is modeled at discrete time instants of analysis $t_k = (k - 1)\Delta t$, $k = 1, \dots, n_T$ where Δt is the time step and n_T is the number of time points considered (clearly, $\Delta t = T/(n_T - 1)$). The loading $f(t)$ is uncertain and is characterized by means of a stochastic process using an appropriate representation (Schuëller 1997), e.g., Karhunen-Loève (KL) expansion, polynomial chaos (PC) expansion, etc. Thus, the loading can be represented as $f(t_k, \mathbf{z})$ where $\mathbf{z}$ is a vector of random variables of dimension n_z whose associated probability density function $f_z(\mathbf{z})$ depends on the characteristics of the stochastic process. Note that depending on the specific situation under

analysis, load could also be represented as a -vector-valued stochastic process (i.e., a vector whose entries are stochastic processes) instead of a scalar. However, in order to simplify the presentation of this contribution and with no loss of generality, the vector-valued case is not explored further.

In addition to the stochastic loading $f(t, z)$, consider a vector $y \in \Omega_y \subset \mathbb{R}^{n_y}$ of dimension n_y , grouping the *design* variables of the structural system. This vector can group those variables that can be altered during the design process (such as cross section of structural members, material properties, etc.). Moreover, consider a classically damped structural system represented by an appropriate model (e.g., a finite element model (Bathe 1996)) comprising a total of n degrees of freedom. Then, the differential equation describing the response of the structure subjected to the stochastic excitation $f(t, z)$ is (see, e.g., Chopra, 1995)

$$\begin{aligned} \mathbf{M}(\mathbf{y})\ddot{\mathbf{x}}(t, \mathbf{y}, \mathbf{z}) + \mathbf{C}(\mathbf{y})\dot{\mathbf{x}}(t, \mathbf{y}, \mathbf{z}) + \mathbf{K}(\mathbf{y})\mathbf{x}(t, \mathbf{y}, \mathbf{z}) \\ + \mathbf{f}_{NL}(\mathbf{y}, \mathbf{x}(t, \mathbf{z}), \dot{\mathbf{x}}(t, \mathbf{y}, \mathbf{z})) = \mathbf{g}\mathbf{f}(t, \mathbf{z}) \end{aligned} \quad (1)$$

where $\mathbf{x}$ is the displacement response vector of dimension n ; $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are the mass, damping, and stiffness matrices of dimension $n \times n$; $\mathbf{f}_{NL}(\mathbf{x}, \mathbf{y}(t, \mathbf{y}, \mathbf{z}), \dot{\mathbf{x}}(t, \mathbf{y}, \mathbf{z}))$ is a vector of dimension n representing the effect of nonlinear forces of the structure due to, e.g., special devices for energy dissipation, yielding, nonlinear behavior, etc.; and $\mathbf{g}$ is the vector of dimension n that couples the stochastic excitation $f(t, z)$ with the degrees of freedom of the structure. Note that mass, damping, stiffness, and nonlinear forces are a function of the vector of design variables $\mathbf{y}$.

First-Excursion Probability

The performance of the structural system in view of the stochastic excitation is characterized by means of n_r responses of interest $r_i(t, \mathbf{y}, \mathbf{z})$, $i = 1, \dots, n_r$, $t \in [0, T]$ measuring, e.g., displacements, accelerations, stresses, etc. For example, in applications associated with earthquake

engineering, a response of interest r_i could be the interstory drift displacement, which can be calculated as a linear combination of the displacement vector, i.e., $r_i(t, \mathbf{y}, \mathbf{z}) = \boldsymbol{\gamma}_i^T \mathbf{x}_i(t, \mathbf{y}, \mathbf{z})$, where $\boldsymbol{\gamma}_i$ is a vector of size n whose entries are 0 and 1. For design purposes, the responses of interest r_i , $i = 1, \dots, n_r$ are checked against allowable threshold levels r_i^* , $i = 1, \dots, n_r$. In a deterministic design framework, the objective is verifying that these responses do not exceed their prescribed thresholds in order to avoid undesirable situations (such as loss of serviceability or collapse). However, when uncertainties are explicitly taken into account, the aforementioned condition cannot be always satisfied, i.e., there is the chance that the responses surpass their prescribed thresholds thus leading to an undesirable situation.

In order to account for the effects of uncertainties and potential undesirable performance of a structure during its lifetime, reliability offers the means for quantifying the level of safety associated with a structural system. A criterion widely used for characterizing safety of a structure is the first-excursion probability (see, e.g., Soong and Grigoriu, 1993). This probability measures the chances that uncertain structural responses exceed in magnitude prescribed thresholds within a specified time interval. That is, first-excursion probability measures the chances of occurrence of the following event F (which is termed in the sequence as *failure* event):

$$F = D_N(\mathbf{y}, \mathbf{z}) \geq 1 \quad (2)$$

where $D_N(\mathbf{y}, \mathbf{z})$ is the so-called normalized demand (Au and Beck 2001) and is defined as

$$D_N(\mathbf{y}, \mathbf{z}) = \max_{i=1, \dots, n_r} \left(\max_{t \in [0, T]} \left(\frac{|r_i(t, \mathbf{y}, \mathbf{z})|}{r_i^*} \right) \right) \quad (3)$$

The normalized demand represents the maximum of the quotient between the structural responses of interest and their corresponding threshold levels. Clearly, whenever $D_N(\mathbf{y}, \mathbf{z})$ exceeds 1, there is failure as the response exceeds its maximum acceptable value. In this context, note

the word *failure* is not a synonym of collapse. It is intended to denote unacceptable system performance which can range from partial damage states (e.g., loss of serviceability) to collapse depending on the specific application under study.

The probability of occurrence of the failure event can be defined by means of the following classical probability integral:

$$\begin{aligned} P_F(\mathbf{y}) &= P[D_N(\mathbf{y}, \mathbf{z}) \geq 1] = \int_{D_N(\mathbf{y}, \mathbf{z}) \geq 1} f_z(\mathbf{z}) d\mathbf{z} \\ &= \int_{\mathbf{z} \in \Omega_z} I_F(\mathbf{y}, \mathbf{z}) f_z(\mathbf{z}) d\mathbf{z} \end{aligned} \quad (4)$$

In the above equation, $P[\cdot]$ denotes probability of occurrence of the argument between brackets, $P_F(\mathbf{y})$ represents the probability of failure (i.e., probability of occurrence of the event F), and $I_F(\mathbf{y}, \mathbf{z})$ denotes the indicator function which is equal to 1 in case the normalized demand is equal or larger than 1 and 0 otherwise.

The evaluation of the failure probability integral is a challenging task as it usually involves a high number of dimensions (for applications of practical interest, n_z can be in the order of thousands) and the normalized demand function $D_N(\mathbf{y}, \mathbf{z})$ can be evaluated point-wise only through (numerically demanding) FE analyses. A possible means for evaluating probability integrals is applying simulation methods. Among different available *simulation methods*, Monte Carlo simulation (MCS) (Metropolis and Ulam 1949) is the most general technique. However, MCS is numerically demanding for estimating low failure probabilities (which are typical in engineering applications). In order to circumvent this issue, *advanced simulation techniques* have

been developed, allowing to estimate small failure probabilities at affordable numerical costs (Schuëller et al. 2004).

Sensitivity of First-Excursion Probability

As shown in Eq. 4, the probability of failure is a function of the design variable vector $\mathbf{y}$. Such dependence can be understood as follows: modifications on the design vector do affect the response of the structure and, consequently, also affect the probability of exceeding the prescribed thresholds. In consequence, for decision making and risk analysis, it is important to evaluate the probability of failure $P_F(\mathbf{y})$ and its sensitivity with respect to changes in the design vector $\mathbf{y}$. That is, besides computing the value of the failure probability, it is also of relevance estimating how much the probability changes due to a modification of the design variable vector. A classical measure for sensitivity is calculating the gradient of the quantity of interest. However, within the context of nonlinear dynamics, the estimation of such quantity may not be feasible as the gradient may not exist due to the nonsmooth normalized demand defined in Eq. 3 (see, e.g., Kang et al. 2006). In order to circumvent this difficulty, an approximate representation of the failure probabilities that is differentiable is constructed, and then, the gradient of this approximation is estimated. Details on these approximations are discussed below.

Estimation of First-Excursion Probability Sensitivity

From a mathematical viewpoint, estimating the gradient of the first-excursion probability implies solving the following limit:

$$\frac{\partial P_F(\mathbf{y})}{\partial y_q} = \lim_{\Delta y_q \rightarrow 0} \frac{P_F(\mathbf{y} + \mathbf{v}(q)\Delta y_q) - P_F(\mathbf{y})}{\Delta y_q}, \quad q = 1, \dots, n_y \quad (5)$$

In Eq. 5, $\mathbf{v}(q)$ is a vector of dimension n_y with all entries equal to zero, except by the q -th entry,

which is equal to one. Introducing the definition of failure probability (see Eq. 4) in Eq. 5 yields

$$\frac{\partial P_F(\mathbf{y})}{\partial y_q} = \lim_{\Delta y_q \rightarrow 0} \frac{P_F[D_N(\mathbf{y} + \mathbf{v}(q)\Delta y_q, \mathbf{z}) \geq 1] - P[D_N(\mathbf{y}, \mathbf{z}) \geq 1]}{\Delta y_q}, \quad q = 1, \dots, n_y \quad (6)$$

Note the limit (as well as the partial derivative) in the above equation may not exist as the normalized demand function $D_N(\mathbf{y}, \mathbf{z})$ may be nonsmooth with respect to both the design parameter vector $\mathbf{y}$ and the uncertain variable vector $\mathbf{z}$. In order to avoid this issue and still obtain a sensitivity measure of the first-excursion probability, approximate representations for the normalized demand function and the excursion probability are introduced. These approximations – which were proposed in Jensen et al. (2009) and Valdebenito and Schuëller (2011) – are discussed in the following. The first approximation comprises an approximate representation of the normalized demand function, i.e.,

$$\begin{aligned} D_N(\mathbf{y} + \Delta \mathbf{y}, \mathbf{z}) &\approx \tilde{D}_N(\mathbf{y} + \Delta \mathbf{y}, \mathbf{z}) \\ &= D_N(\mathbf{y}, \mathbf{z}) + \sum_{q=1}^{n_y} a_q \Delta y_q \end{aligned} \quad (7)$$

where $\tilde{D}_N(\mathbf{y}, \mathbf{z})$ is the approximate normalized demand function, $\Delta \mathbf{y}$ is a certain perturbation of

the design variable vector, and $a_q, q = 1, \dots, n_y$ are real, constant coefficients. The procedure to determine these coefficients is described in section “Numerical Implementation.” The second approximation involves an approximate representation of the probability that the normalized demand exceeds a threshold level b , i.e.,

$$P[D_N(\mathbf{y}, \mathbf{z}) \geq b] \approx e^{\psi_0 + \psi_1(b-1)}, \quad b \in [1 - \epsilon, 1 + \epsilon] \quad (8)$$

where ϵ is a small constant and where ψ_0, ψ_1 are real coefficients. The issue of how to calculate them is analyzed in section “Numerical Implementation.” A thorough discussion on the applicability of these two approximations can be found in Jensen et al. (2009) and Valdebenito and Schuëller (2011).

Using the approximations introduced in Eqs. 7 and 8, it can be shown (see Valdebenito and Schuëller (2011)) that the sought gradient can be approximated as follows:

$$\frac{\partial \tilde{P}_F(\mathbf{y})}{\partial y_q} = \lim_{\Delta y_q \rightarrow 0} \frac{P[\tilde{D}_N(\mathbf{y} + \mathbf{v}(q)\Delta y_q, \mathbf{z}) \geq 1] - P[D_N(\mathbf{y}, \mathbf{z}) \geq 1]}{\Delta y_q} \quad (9)$$

$$= \lim_{\Delta y_q \rightarrow 0} \frac{P[D_N(\mathbf{y}, \mathbf{z}) \geq 1] - a_q \Delta y_q - P[D_N(\mathbf{y}, \mathbf{z}) \geq 1]}{\Delta y_q} \quad (10)$$

$$= \lim_{\Delta y_q \rightarrow 0} \frac{e^{\psi_0 - \psi_1 a_q \Delta y_q} - e^{\psi_0}}{\Delta y_q} \quad (11)$$

$$= -\psi_1 a_q e^{\psi_0} \quad (12)$$

$$= -\psi_1 a_q \tilde{P}_F(\mathbf{y}) \quad (13)$$

In the above equations, $\partial \tilde{P}_F(\mathbf{y}) / \partial y_q$ represents the partial derivative of the approximate

representation of the failure probability with respect to $y_q, q = 1, \dots, n_y$ and $\tilde{P}_F(\mathbf{y})$ is the estimate for the failure probability in Eq. 4 which is calculated using an advanced simulation method. Thus, for estimating the gradient of the probability, it is necessary to determine the probability $\tilde{P}_F(\mathbf{y})$ and the coefficients ψ_1 and $a_q, q = 1, \dots, n_y$. A procedure for determining this probability and these coefficients is discussed in the following section.

Numerical Implementation

General Remarks

This section provides details on the procedure to calculate the different coefficients and probability involved in the estimation of the first-excursion probability sensitivity according to the approximate formula proposed in Eq. 13. First, it is explained how excursion probabilities are estimated. Then, the sensitivity of the first-excursion probability with respect to a normalized threshold level is analyzed, and finally, an approximate representation of the normalized demand function is discussed.

First-Excursion Probability Estimations

As already pointed out, estimating a first-excursion probability implies solving the integral in Eq. 4. In this contribution, this probability is evaluated by means of the so-called advanced simulation methods (Schuëller et al. 2004). In particular, in this contribution, subset simulation (SS) (Au and Beck 2001) is applied to estimate first-excursion failure probabilities.

In SS, the failure domain F (see Eq. 2) is defined as a sequence of subsets (or *intermediate* failure events) F_i , $i = 1, \dots, m$ such that $F_1 \supset F_2 \supset \dots \supset F_m = F$. Thus, the failure probability is cast as a product of conditional failure probabilities, i.e.,

$$P_F = P[F_m] = P[F_1] \prod_{i=1}^{m-1} P[F_{i+1}|F_i] \quad (14)$$

where $P[F_{i+1}|F_i]$ is the probability of occurrence of the event F_{i+1} conditioned on the event F_i . In this way, a small failure probability is expressed as the product of larger, conditional probabilities which can be calculated using, e.g., Monte Carlo simulation (MCS). The practical implementation of SS requires an efficient algorithm for generating samples of the uncertain parameters ($\mathbf{z}$) conditioned on an intermediate failure event, such as the modified Metropolis algorithm (Au and Beck 2001). Details on the algorithmic implementation of SS can be found in Au and Beck (2001).

A salient feature of SS is that it *populates* the space of uncertain parameters by means of successive subsets F_i , $i = 1, \dots, m$, each of which is of

a rarer occurrence than the previous one, i.e., $P[F_i] < P[F_{i-1}]$. Hence, a full run of SS does not only provide the probability of occurrence associated with a normalized threshold $b = 1$ (see Eq. 4), but actually for a range of the normalized threshold levels (Au and Beck 2001).

Sensitivity of First-Excursion Probability with Respect to Threshold

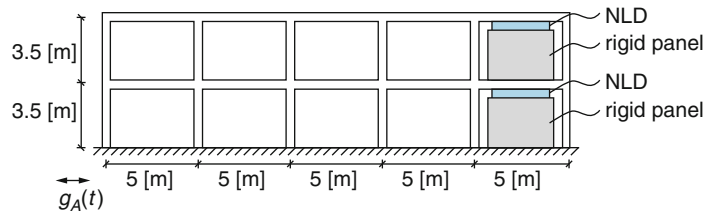
The approximation introduced in Eq. 8 allows estimating the probability that the normalized demand function $D_N(\mathbf{y}, \mathbf{z})$ exceeds a prescribed threshold b , i.e., it is sought to estimate the curve relating normalized thresholds b and their corresponding first-excursion probabilities. Note that this curve can be calculated as a byproduct of reliability analysis applying advanced simulation methods such as SS. Thus, the coefficients ψ_0 and ψ_1 can be estimated with no additional numerical efforts once a reliability analysis has been carried out. In fact, when applying SS, the sought coefficients can be calculated with the samples of the normalized demand generated at the last stage of SS. The main idea is generating the curve probability versus normalized threshold in a discrete manner using samples already available. Then, the sought coefficients are calculated in a least square sense considering the analytic approximation of Eq. 8. Details on how to implement this procedure are discussed in Valdebenito and Schuëller (2011).

Approximate Representation of Normalized Demand Function

The approximate representation in Eq. 7 suggests that changes in the normalized demand function $D_N(\mathbf{y}, \mathbf{z})$ due to changes in the design variable vector $\mathbf{y}$ can be explained through a linear relation. Such relation is clearly not captured exactly by such a simple expression due to a number of reasons: higher-order terms involving $\mathbf{y}$ are ignored, no interaction between $\mathbf{y}$ and $\mathbf{z}$ is considered, the nonlinear nature of the normalized demand function (see Eq. 3) is not captured appropriately, etc. However, in case the coefficients a_q , $q = 1, \dots, n_y$ are chosen appropriately, it could be expected that $\tilde{D}_N(\mathbf{y} + \Delta\mathbf{y}, \mathbf{z})$ approximates $D_N(\mathbf{y} + \Delta\mathbf{y}, \mathbf{z})$ sufficiently well. That is, the coefficients a_q represent an *average sensitivity* on how

Sensitivity of First-Excursion Probabilities for Nonlinear Stochastic Dynamical Systems,

Fig. 1 Example – 2-story RC frame structure including nonlinear devices (NLDs)



the design variables affect the normalized demand. Numerical validation (see Jensen et al. 2009; Valdebenito and Schuëller 2011) has shown such an assumption is appropriate within the scope of the problems studied in this contribution.

The procedure applied to calculate the coefficients a_q , $q = 1$, n_y is quite straightforward. Samples of the uncertain variable vector $\mathbf{z}$ are taken from the last stage of SS. It is expected that for these samples the normalized demand function is close to 1. Then, for each of these samples, the value of the normalized demand is reevaluated considering perturbed values of the design variable vector. Then, the sought coefficients are estimated in a least square sense considering the sampled data and the analytic model of Eq. 7. For details on the implementation of this procedure, the reader is referred to Jensen et al. (2009) and Valdebenito and Schuëller (2011).

Example

Description of the Problem

In order to illustrate the application of the procedure for failure probability sensitivity estimation, the following example is considered. It involves a two-story reinforced concrete (RC) frame which includes nonlinear hysteretic devices (NLDs). Figure 1 illustrates the elevation of the model. The frame is excited by a horizontal ground acceleration of 15 [s] duration, which is modeled as a stochastic process. The failure event

takes place whenever displacement of each floor of the building exceeds a prescribed threshold within the duration of the stochastic ground acceleration. The design variables of the problem refer to the dimensions of the columns of the RC frame.

The RC frame possesses a Young's modulus equal to 2×10^{10} [N/m²]. Each of its floors (of mass 1.5×10^5 [kg]) is supported by six columns of square cross section (side length equal to 0.5 [m]) and a height of 3.5 [m]. The beams of the frame are rigid in the axial direction, so each floor can be described by a single horizontal degree of freedom (DOF); thus, the model involves a total of two DOFs. It is assumed that the columns and beams remain linear within the duration of the stochastic ground acceleration and classical modal damping of 5 % for all modes is assumed.

In order to improve the safety of the frame, two hysteretic NLDs are included. The restoring force ($F_{R,i}(\cdot)$) associated with the i -th NLD is described by the following model (Pradlwarter and Schuëller 1993):

$$F_{R,i}(t) = k_d(\Delta_i(t) - q_i^1(t) + q_i^2(t)), \quad i = 1, 2 \quad (15)$$

where $\Delta_i(\cdot)$ is the relative displacement between the $(i, i - 1)$ -th floors, k_d is the stiffness of the NLD, and $q_i^1(\cdot)$ and $q_i^2(\cdot)$ denote the plastic elongations of the NLD, which are governed by the following equations:

$$\dot{q}_i^1(t) = \dot{\Delta}_i(t)H(\dot{\Delta}_i(t)) \left[H(\tau_i(t) - \Delta_p) + H(\tau_i(t) - \Delta_y) \frac{\tau_i(t) - \Delta_y}{\Delta_p - \Delta_y} H(\Delta_p - \tau_i(t)) \right], \quad i = 1, 2 \quad (16)$$

$$\dot{q}_i^2(t) = -\dot{\Delta}_i(t)H(-\dot{\Delta}_i(t)) \left[H(-\tau_i(t) - \Delta_p) + H(-\tau_i(t) - \Delta_y) \frac{-\tau_i(t) - \Delta_y}{\Delta_p - \Delta_y} H(\Delta_p + \tau_i(t)) \right], \quad i = 1, 2 \quad (17)$$

where $\tau_i(t)$ is an auxiliary variable defined as $\tau_i(t) = \Delta_i(t) - q_i^1(t) + q_i^2(t)$, $i = 1, 2$, and Δ_y and Δ_p are the yielding and plastic displacements, respectively. The numerical values considered for the parameters of the NLD are $k_d = 10^8$ [N/m], $\Delta_p = 6 \times 10^{-3}$ [m], and $\Delta_y = 0.8\Delta_p$.

The ground acceleration ($g_A(t)$) is modeled as a filtered white noise of 15 [s] duration. The ground acceleration is calculated as $g_A(t) = \alpha^T \dot{p}(t)$; the vectors α^T and $p(t)$ are defined as

$$\alpha^T = \langle \Omega_1^2, 2\xi_1\Omega_1, -\Omega_2^2, -2\xi_2\Omega_2 \rangle \quad (18)$$

$$\dot{p}(t) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\Omega_1^2 & -2\xi_1\Omega_1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \Omega_1^2 & 2\xi_1\Omega_1 & -\Omega_2^2 & -2\xi_2\Omega_2 \end{pmatrix} p(t) + \begin{pmatrix} 0 \\ \omega(t)e(t) \\ 0 \\ 0 \end{pmatrix} \quad (19)$$

where $\Omega_1 = 15$ [rad/s], $\xi_1 = 0.8$, $\Omega_2 = 0.3$ [rad/s], and $\xi_2 = 0.995$ are the filter parameters; $\omega(t)$ denotes a white noise signal; and $e(t)$ is an envelope function:

$$e(t) = \begin{cases} t^2/16 & 0[s] \leq t < 4[s] \\ 1 & 4[s] \leq t < 10[s] \\ e^{-(t-10)^2} & 10[s] \leq t \leq 15[s] \end{cases} \quad (20)$$

A time discretization step equal to $\Delta t = 0.01$ [s] is used to model the ground acceleration. Thus, the discrete representation of the white noise signal is $\omega(t_k) = \sqrt{2\pi S/\Delta t} z_k$, $k = 1, \dots, 1501$, where $S = 10^{-3}$ [m²/s³] is the spectral density of the white noise and z_k , $k = 1, \dots, 1501$ are independent, identically distributed standard Gaussian variables.

The failure event is formulated as a first-excursion problem during the time of analysis; the structural responses to be controlled are the two interstory drift displacements and the roof

Sensitivity of First-Excursion Probabilities for Nonlinear Stochastic Dynamical Systems, Table 1 Estimates of first-excursion probability sensitivity (cov: coefficient of variation)

Proposed approach		Finite differences	
$\frac{\partial \tilde{P}_F(y)}{\partial y_1}$	$\frac{\partial \tilde{P}_F(y)}{\partial y_2}$	$\frac{\partial \tilde{P}_F(y)}{\partial y_1}$	$\frac{\partial \tilde{P}_F(y)}{\partial y_2}$
(cov)	(cov)		
−2.24	−1.11	−2.49	−1.19
(20.7)%	(24.8)%		

displacement. The threshold values are chosen equal to 0.2 % of the floor height for the interstory drift displacements and 0.1 % of the frame height for the roof displacement. The design variables refer to the cross sections of the columns of the RC frame, more specifically to the second moment of area of the cross section. These cross sections are grouped into two design variables linking the columns of the first and second floor, respectively.

Results

The sensitivity of the first-excursion probabilities is estimated using the approach described previously. In particular, SS is applied considering 2,000 samples of the uncertain ground acceleration at each simulation stage. The resulting first-excursion probability estimate is equal to $\tilde{P}_F = 10^{-3}$. Hence, a total of 6,000 samples of the uncertain variables are required in order to estimate the sought probability. Recall that these samples are also used to estimate the coefficients ψ_0 and ψ_1 . In order to calibrate the approximate model of Eq. 7, 200 perturbed designs are analyzed. Hence, the estimation of the probability sensitivity demands only 200 extra structural analyses. In order to illustrate the variability associated with the approach for estimating first-excursion probability sensitivity, a total of 100 independent runs were generated. The results in terms of the mean of these runs and their corresponding coefficient of variation are shown in Table 1. In addition and in order to validate the results obtained in Table 1, the probability sensitivity is estimated using a central finite difference estimator. In order to ensure the probability

estimates used in the finite difference scheme are sufficiently accurate, the average of 100 independent runs is considered. The results associated with the finite differences are presented in Table 1 as well. It can be observed that the results presented are in good agreement, indicating the approach for sensitivity estimation reported herein provides appropriate results.

Summary

This contribution has presented an approach for sensitivity analysis of first-excursion probabilities associated with nonlinear dynamical systems. The basis of the approach is combining advanced simulation methods for probability estimation with a series of local approximations involving the normalized demand function.

A salient feature of the approach reported herein is that it is numerically efficient. This is achieved as the proposed local approximations take advantage of the results already available from probability estimation.

The results presented in this contribution indicate that the proposed approach for sensitivity is applicable for problems involving a large number of uncertain parameters; results presented elsewhere (see, e.g., Valdebenito and Schuëller 2011) indicate that the approach is also capable of including a considerable number of design variables.

Cross-References

- [Reliability Estimation and Analysis](#)
- [Robust Design Optimization for Earthquake Loads](#)
- [Stochastic Analysis of Nonlinear Systems](#)
- [Structural Optimization Under Random Dynamic Seismic Excitation](#)
- [Structural Reliability Estimation for Seismic Loading](#)
- [Structural Seismic Reliability Analysis](#)
- [Subset Simulation Method for Rare Event Estimation: An Introduction](#)

References

- Au S, Beck J (2001) Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilist Eng Mech* 16(4):263–277
- Bathe K (1996) *Finite element procedures*. Prentice Hall, New Jersey
- Chopra A (1995) *Dynamics of structures: theory and applications to earthquake engineering*. Prentice Hall, Englewood Cliffs
- Jensen H, Valdebenito M, Schuëller G, Kusanovic D (2009) Reliability-based optimization of stochastic systems using line search. *Comput Method Appl Mech Eng* 198(49–52):3915–3924
- Kang B-S, Park G-J, Arora J (2006) A review of optimization of structures subjected to transient loads. *Struct Multidiscip Optim* 31(2):81–95
- Metropolis N, Ulam S (1949) The Monte Carlo method. *J Am Stat Assoc* 44(247):335–341
- Pradlwarter HJ, Schuëller GI (1993) Equivalent linearization – a suitable tool for analyzing MDOF systems. *Probabilist Eng Mech* 8(2):115–126
- Schuëller G (1997) A state-of-the-art report on computational stochastic mechanics. *Probabilist Eng Mech* 12(4):197–321
- Schuëller G, Pradlwarter H, Koutsourelakis P (2004) A critical appraisal of reliability estimation procedures for high dimensions. *Probabilist Eng Mech* 19(4):463–474
- Soong T, Grigoriu M (1993) *Random vibration of mechanical and structural systems*. Prentice Hall, Englewood Cliffs
- Valdebenito M, Schuëller G (2011) Efficient strategies for reliability-based optimization involving non linear, dynamical structures. *Comput Struct* 89(19–20): 1797–1811

Sensors, Calibration of

Jens Havskov¹ and Gerardo Alguacil²

¹Department of Earth Science, University of Bergen, Bergen, Norway

²Instituto Andaluz de Geofísica, University of Granada, Granada, Spain

Synonyms

Accelerometer; Response function; Seismometer; Seismometer damping

Introduction

The output of a seismic sensor, a seismometer or an accelerometer, is a time-varying voltage, which is related to the ground motion by a differential equation in the time domain or by a transfer function in the frequency domain. This transfer function or response function is characterized by a number of parameters, which are assumed to be constant, at least in the short term.

A seismic sensor will have calibration information given by the manufacturer. These specifications will be used to correct the seismic signal from the sensor to produce the true ground motion. If, e.g., for a given harmonic ground velocity $X(\omega)$, the output from the sensor is $Y(\omega)$, the amplitude response or transfer function $A(\omega)$ is defined as the ratio $A(\omega) = Y(\omega)/X(\omega)$, where ω is the frequency in radian/s. So, if $A(\omega)$ is known, the input (ground motion) can then be calculated as

$$X(\omega) = Y(\omega)/A(\omega) \quad (1)$$

The sensor will, in addition to changing the amplitude of the ground signal, also change the phase of the signal, so an additional function will be needed for the phase response. The phase and amplitude response functions can be combined in one complex function ($A(\omega)$ would then be complex); however, here, for simplicity, they will be treated separately.

The amplitude and phase response functions are usually calculated from the sensor parameters but can also be obtained by direct measurements.

With time, the sensor might degrade or it might develop a fault so there is a need to be able to calibrate the sensor. This means determining the instrument calibration parameters (parametric calibration) or determining $A(\omega)$ directly for each frequency of interest (empirical calibration).

Modern sensors can be very complicated in terms of electronics and mechanical construction so it might not be possible to determine all relevant parameters; however, some parameters can be obtained with simple tests. In this section,

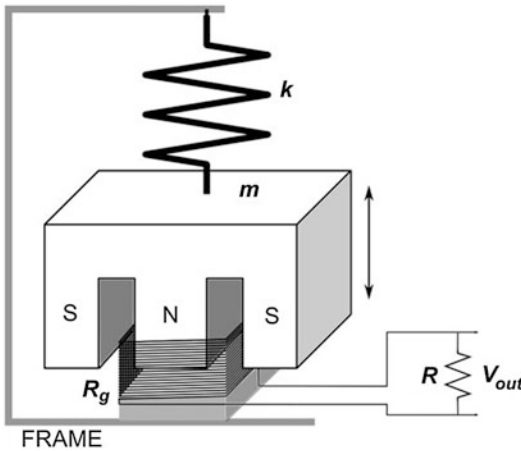
basic parameters will be described, some tests used to obtain them as well as methods for obtaining $A(\omega)$ directly.

Seismic Sensors

Seismic sensors can be divided into two kinds: sensors measuring the ground velocity (velocity sensor or seismometer) and sensors measuring the ground acceleration (accelerometers). A seismometer can be passive, meaning there is no electronic parts or active with an electronic circuit integrated. All accelerometers for seismic recording are active sensors. Piezoelectric accelerometers are passive (although they may have a conditioning circuit built-in) and are widely used for structural vibrations monitoring, but seldom in seismology, due to their poor sensitivity at low frequency and low dynamic range.

Passive Seismometers: A passive seismometer consists of a swinging mass with a coil moving in magnetic field. It is also called an electromagnetic sensor. The swinging system has resonance frequency ω_0 . An example is seen in Fig. 1.

When the mass is moving, the magnetic field will vary in the coil. An output voltage proportional to the velocity of the mass relative to the ground will then be produced. The proportionality constant is called the generator constant G and has the unit V/ms^{-1} . But the relative motion of the mass depends on the frequency of the ground motion. Thus the sensor will not be equally sensitive to ground motion for all frequencies. Qualitatively the response of the sensor can be understood as follows. If the ground moves with a very fast sinusoidal motion, the mass remains stationary in an inertial frame, and thus the ground sinusoidal velocity is measured directly. With the ground moving very slowly, the mass would have time to follow the ground motion, so there would be little relative motion and the gain would be low. At the resonance frequency, the mass could get a new push at the exact right time, so the mass would move with a larger and larger amplitude, only limited by the damping of the motion. The sensor amplitude response $A(\omega)$,



Sensors, Calibration of, Fig. 1 A model of an electro-magnetic sensor. The mass m is also a magnet suspended by a spring k . The coil resistance is R_g , the damping resistor is R , and the voltage output is V_{out} . The mass motion is damped by the current through the coil (fixed to the frame) and the external damping resistor

which is the output voltage of the sensor as a function of the input ground velocity, can be obtained as (e.g., Lay and Wallace 1995)

$$A(\omega) = \frac{G\omega^2}{\sqrt{(\omega_0^2 - \omega^2)^2 + 4h^2\omega^2\omega_0^2}} \quad (2)$$

where h is the damping, ω is the frequency, and ω_0 is the natural frequency of the swinging system. Figure 2 shows examples of $A(\omega)$ with different damping.

It is seen that, as the damping decreases, the gain gets a larger and larger peak at the natural frequency. The flat curve represents a damping of $h = 0.707$ and is the desired value. The damping is achieved by shunting the signal coil with a damping resistor and thereby draining energy out of the swinging system.

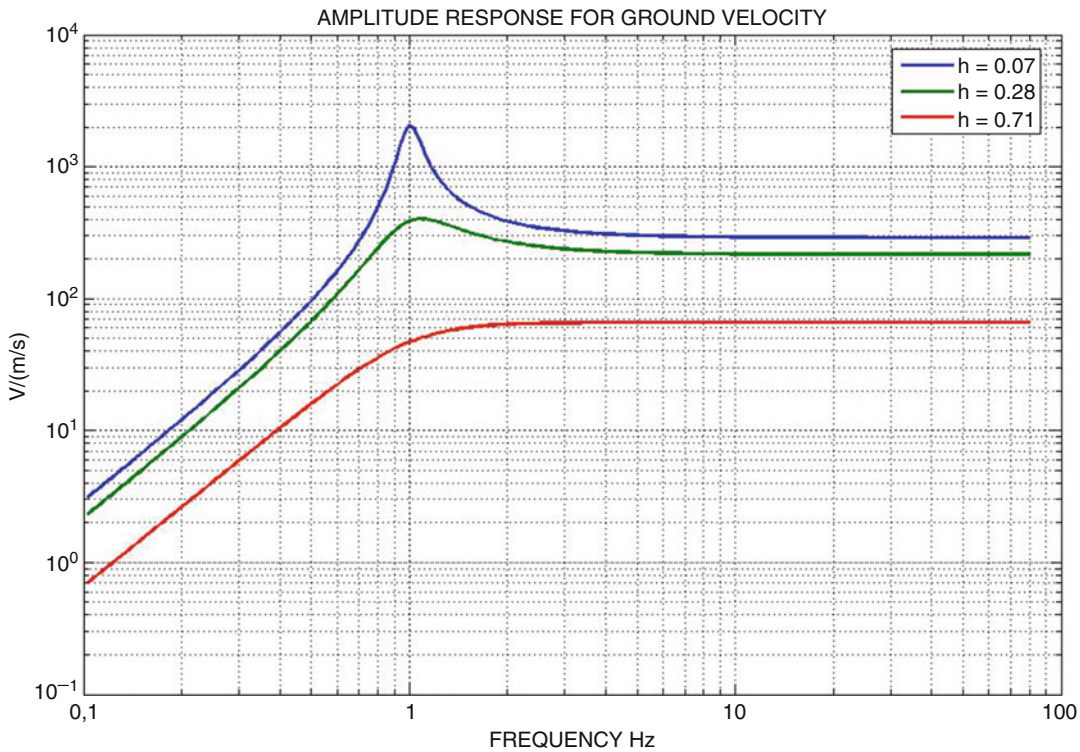
Active Sensors: The heart of the active sensor is a device measuring the ground acceleration, the so-called force balanced accelerometer (FBA) (Fig. 3). The FBA has a feedback coil, which can exert a force equal and opposite to the inertia force due to the acceleration. The displacement transducer sends a current to this force coil

through a resistor R in a negative feedback loop. The polarity of the current is such that it opposes any motion of the mass, and it will try to prevent the mass from moving at all with respect to the frame. A small permanent acceleration on the mass will therefore result in a small permanent current, and a large acceleration will need a large current. The current is in fact proportional to the ground acceleration, so the voltage over the resistor gives a direct measure of acceleration. This is how nearly all accelerometers work, and in practice the only constant of importance is the generator constant G in units of V/g, where g is the gravity acceleration.

The FBA principle is now the heart of nearly all modern strong motion and broadband sensors (BB, sensors recording in a large frequency band like 0.01–50 Hz). By connecting an integrating circuit after the output, the sensor can give out a voltage proportional to velocity like for passive sensors. However, due to the mechanical-electrical qualities of the sensor, there is in practice a low frequency limit for the flat velocity response. For lower frequencies, the amplitude response decreases proportional to frequency squared. This means that in practice, a BB sensor will have the equivalent of a free period and a damping, and its amplitude response can be approximated by Eq. 2, although a more exact model of its response should include additional parameters that influence the high-frequency behavior. In practice Eq. 2 must then be multiplied with a function which often represents a filter, like a Butterworth filter that can be represented by a simple function of filter frequency and number of poles of the filter; see Havskov and Alguacil (2010). An estimate for the filter parameters can be obtained by empirical calibration; see later.

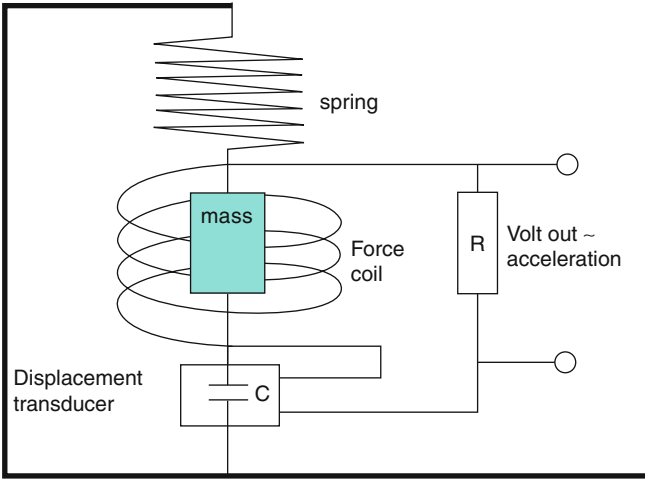
The sensor parameters to measure can therefore be summarized as instrument natural frequency, damping, and generator constant. With these parameters, the amplitude response function Eq. 2 can be calculated.

Alternatively the amplitude response function can be determined directly by measuring the output of the sensor with a controlled input.



Sensors, Calibration of, Fig. 2 Amplitude response to ground velocity of a typical passive sensor with a natural frequency of 1 Hz for different damping values

Sensors, Calibration of, Fig. 3 Simplified principle behind the force balanced accelerometer. The displacement transducer normally uses a capacitor C, whose capacitance varies with the displacement of the mass. A current, proportional to the displacement transducer output, will force the mass to remain stationary relative to the frame (Figure from Havskov and Alguacil 2010)



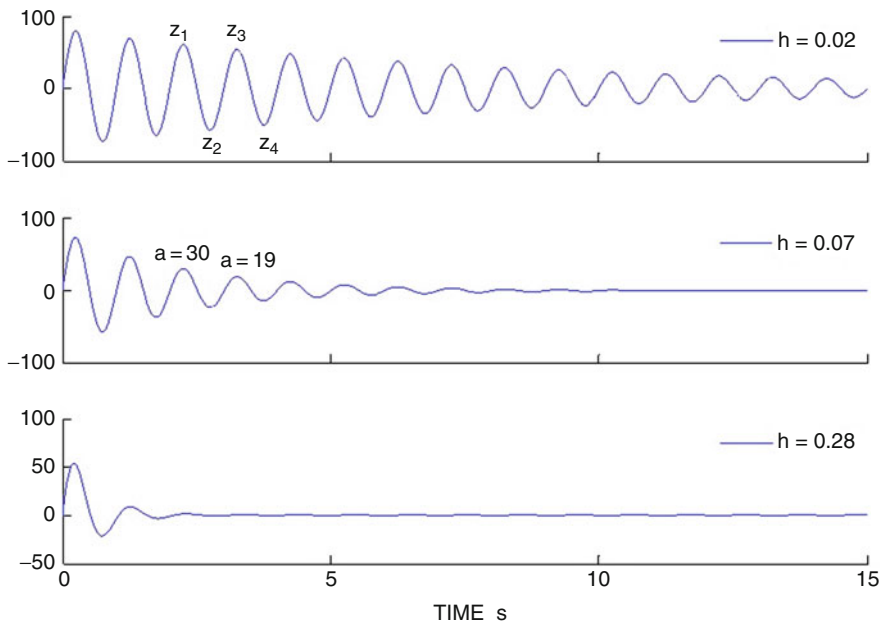
Determining Sensor Parameters

Natural Frequency

The frequency with which the seismometer mass is oscillating depends on the damping; more

damping makes it swing at a lower frequency ω_1 . The relation is (Havskov and Alguacil 2010)

$$\omega_1 = \sqrt{\omega_0^2(1 - h^2)} \tag{3}$$



Sensors, Calibration of, Fig. 4 Free swinging of three typical 1 Hz seismometers with different open-circuit damping h . Traces have been generated synthetically. The decaying extremes (peak amplitudes) are labeled z_1 ,

z_2 , etc. (top trace). On the middle trace, the amplitudes of two maximums following each other are given in an arbitrary scale (Figure modified from Havskov and Alguacil 2010)

The natural frequency is therefore best measured without any external damping. The free period can be determined in several ways depending on the type of sensor.

Passive Sensor: Give the sensor an impulse or a step to make it swing, a small push or tilt will do. For a very few sensors, it is possible to see the mass swing, else the output signal has to be observed on an oscilloscope or recorded. Measure the time of several swings and calculate the average or measure the frequency from the recorded signal. Many short-period sensors (natural frequency higher than 1 Hz) have too high open-circuit damping (damping without a damping resistor) to observe more than a few swings of the signal (Fig. 4), and the signal must be recorded to measure the period.

BB Sensor: The response of a BB sensor is controlled mainly by its internal feedback loop. The user can only measure its “apparent free period”; however, this is difficult since the damping is fixed at 0.7 so there will be almost no swinging (see Fig. 9). Most BB sensors have a calibration input, which can be used to produce

a calibration pulse from which the “apparent free period” and damping can be obtained; see section on calibration pulse later.

If the open-circuit damping is high, say larger than 0.15 (corresponding to about 1 % change in period), the measured free period should be corrected for damping.

Damping

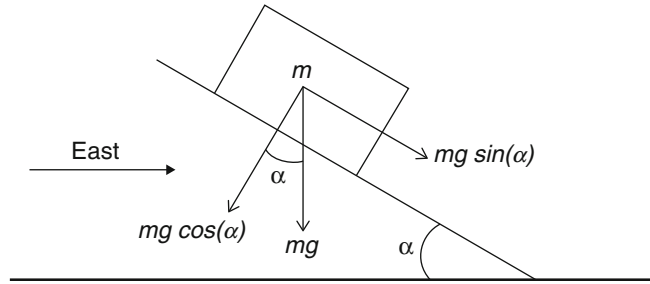
The open-circuit damping h_m is caused by the friction in the mechanical system and can be determined from the signals from the free swinging system as shown in Fig. 4. Using the extreme amplitudes z_1 and z_n where z_1 is extreme 1 and z_n is extreme $n + 1$ (so $n = 1$ for two following extremes (Fig. 4)), the open-circuit damping can be determined as (Havskov and Alguacil 2010)

$$h_m = \frac{\ln(z_1/z_n)}{\sqrt{n^2\pi^2 + \ln^2(z_1/z_n)}} \quad (4)$$

Measuring z_1 and z_2 , h_m can now be determined directly. In Fig. 4, two following maxima ($n = 2$)

Sensors, Calibration of,

Fig. 5 Tilting an accelerometer to determine generator constant. It is assumed that the sensor horizontal direction is toward the east (Figure from Havskov and Alguacil 2010)



have approximate amplitudes 30 and 19, respectively. This gives a damping of

$$h_m = \frac{\ln(30/19)}{\sqrt{2^2\pi^2 + \ln^2(30/19)}} = 0.0725 \quad (5)$$

A sensor in operation should always have a damping resistor connected in order to get the desired total damping of 0.707 so as to get the most flat response. The additional damping caused by the electrical damping is called h_e , and the corresponding total resistance of the coil and damping resistor is R_T . The damping resistor cannot be determined experimentally as above, since with total damping of 0.707, there would be too little overswing. It can be shown (Havskov and Alguacil 2010) that the ratio between two electrical damping constants is inversely related to the ratio of the damping resistances:

$$R_{T2} = R_{T1} \frac{h_{e1}}{h_{e2}} \quad (6)$$

R_{T1} and R_{T2} are two different total damping resistances and h_{e1} and h_{e2} the corresponding damping coefficients caused by the resistances, respectively. In addition comes the open-circuit damping, so, the total damping with R_{T1} is h_1 :

$$h_1 = h_{e1} + h_m \quad (7)$$

and writing Eq. 6 in terms of the total damping gives

$$R_{T2} = R_{T1} \frac{h_{e1}}{h_{e2}} = R_{T1} \frac{h_1 - h_m}{h_2 - h_m} \quad (8)$$

When $h_2 = 0.707$, the corresponding $R_{T2} = R_{0.707}$ is

$$R_{0.707} = R_{T1} \frac{h_1 - h_m}{0.707 - h_m} \quad (9)$$

so the procedure to determine $R_{0.707}$ is:

1. Determine h_m as described above.
2. Connect a damping resistor R_{T1} of “suitable size,” meaning that the overswing is substantially less than when measuring h_m but large enough to get an accurate determination of h_1 .
3. Use Eq. 9 to determine $R_{0.707}$.

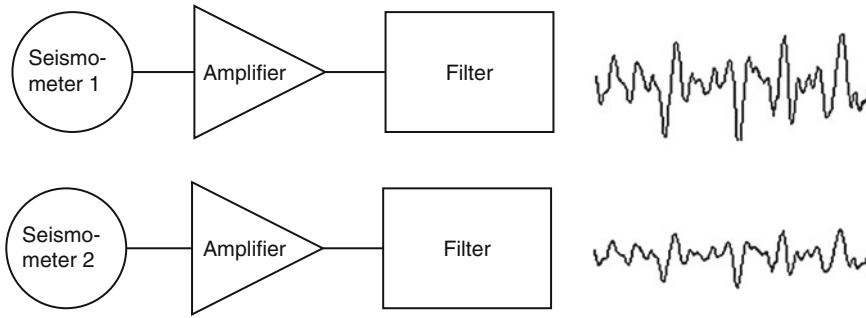
For BB sensors or other active sensors, there is no damping resistor, and the damping is preset from the factory. These sensors have a buffered low-impedance voltage output, so an external load resistor would not influence the sensor at all.

Generator Constant

In order to determine the generator constant experimentally, a known input to the sensor must be used or the generator constant must be calculated from other known measurable parameters.

Accelerometer: For a static ground acceleration, an accelerometer has an output proportional to this acceleration. An accelerometer cannot distinguish any difference between an inertial force due to ground acceleration and an equivalent gravity force. By tilting the accelerometer, the effective force on the three components can be determined for different tilt angles and the sensitivity determined; see Fig. 5.

The accelerometer is tilted an angle α . The force in the horizontal direction is now $mg \sin(\alpha)$,



Sensors, Calibration of, Fig. 6 Comparison of signals recorded from two different seismometers with the same recording equipment. The filters and amplifiers are identical and filter out signals below the seismometers natural

frequencies. The ratio of the output amplitudes (right) indicates the ratio of the sensor generator constants (Figure from Havskov and Alguacil 2010)

while in the vertical direction it is $mg\cos(\alpha)$. The voltage output for the vertical component is V_z , and for the horizontal component it is V_h . Considering that in the horizontal position, the output is supposed to be zero and that the vertical force has been *decreased* from mg to $mg\cos(\alpha)$, the generator constant can be calculated as

$$G_z = \frac{-V_z}{g(1 - \cos(\alpha))} \quad \text{and} \quad G_h = \frac{V_h}{g \sin(\alpha)} \quad (10)$$

V_z will be negative since the vertical force is decreased, and V_h will be positive since the force is in the east direction. This method is very simple and can determine the generator constant accurately, but will not give a dynamic calibration. The horizontal components can be tested for symmetry by inclining in the opposite direction. In the case that the instrument output is not zero in the horizontal position, this offset should be adjusted before any measurement is made; alternatively it may be subtracted from the output for each measurement.

Passive Velocity Seismometer: The generator constant can be calculated from

$$G = \sqrt{2m\omega_0 h_e R_T} \quad (11)$$

where m is the mass of the moving mass and h_e is the electrical damping corresponding to the total resistance R_T . In the previous section it has been described how R_T , h_e , and ω_0 can be determined.

All Sensors: A shaking table is a platform that moves with a controlled motion. Setting it up to move with a harmonic motion with maximum amplitude A (velocity or acceleration) at a frequency where the response curve is flat, the output from the sensor will also be a harmonic function with amplitude B . The generator constant is then

$$G = B/A \quad (12)$$

Unfortunately a shaking table is rarely available, so other methods must generally be used.

The ground is always moving so if the motion is known, it is equivalent to having a shaking table moving both horizontally and vertically. The problem is then to determine this motion in order to have an absolute measure. This is done by using a well-calibrated sensor. The experiment then consists of setting up the two sensors close to each other and assuming that they are subject to the same ground motion (Fig. 6). The data is filtered so as to use a frequency band where both sensors are measuring the signal in the flat part of their instrument response function, e.g., from 2 to 5 Hz using a BB sensor as a reference and a 1 Hz sensor to be calibrated.

The generator constant can then be determined as

$$G = G_R B/A \quad (13)$$

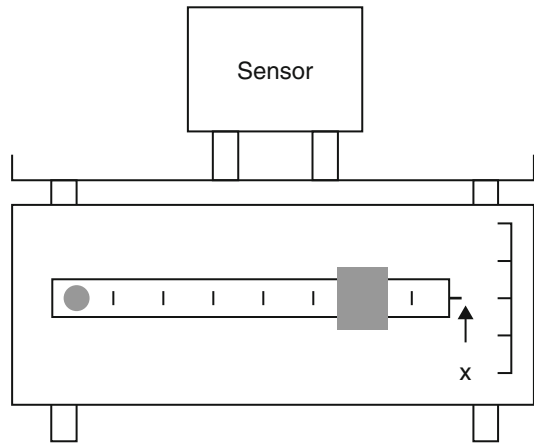
where G_R is the generator constant of the reference sensor, A is the amplitude of the reference sensor, and B is the amplitude of the unknown

sensor. The amplitude can be picked for any part of the two seismograms provided that they are identified as the same “swing”; see Fig. 6. It is here assumed that both sensors are of the same type like comparing a velocity sensor to a velocity sensor or an accelerometer to an accelerometer. Note that many accelerometers are not sensitive enough to be able to use this method, so the tilt method should be used.

Calibration of Velocity Sensors by Stepwise Motion: The main idea behind this method is to (1) move the sensor to a known distance (like 1 mm) in its direction of sensitivity, (2) record the signal, and (3) correct the signal for the known displacement frequency response (response function relating sensor output to displacement) to get the displacement. Theoretically, the displacement is now calculated, which can be compared to the actual displacement, from which the generator constant can be calculated. The method is described in detail by Wielandt in chapter “► [MEMS Sensors for Measurement of Structure Seismic Response and Their Application](#)” in NMSOP (Bormann 2002), which claims that the method works well for broadband sensors (accuracy down to 1 %) and even 10 Hz seismometers. Although the method sounds simple, it is not trivial to correct for the instrument response down to DC, particularly for SP instruments.

Another problem is how to move the sensor, horizontally or vertically, at a controlled distance. A simple instrument that has been used for vertical motion is a mechanical balance (Fig. 7). Placing the sensor on the mechanical balance table, the seismometer can be moved at a controlled distance as measured on the balance arm vertical scale. In principle, it is enough to make one displacement, but usually several are made to take average measurements.

A simple portable calibration table which operates on this principle is commercially available from Lennartz Electronic (www.lennartz-electronic.de). It moves the vertical axis a known displacement. It also allows to tilt the sensor a known angle, which is sensed as an apparent step in acceleration along the horizontal axis, thus permitting the calibration of horizontal components as well.



Sensors, Calibration of, Fig. 7 Moving a sensor at a controlled distance vertically with a mechanical balance. The ratio between the motion at x (where we can measure it) and on the balance table can be determined by placing a mass m_1 at x that will balance a mass m_2 on the balance table. The ratio of the two motions is then m_1/m_2 (Figure from Havskov and Alguacil 2010)

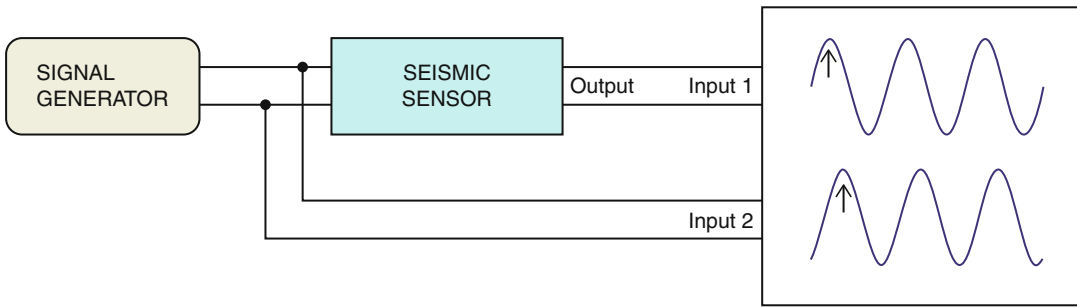
It should be noted that some of the triaxial BB seismometers have the three sensors arranged symmetrically, forming an orthogonal frame with its active axes inclined 54.7° with respect to horizontal, so each sensor is sensitive to vertical motions. Their outputs are then electronically combined to generate the conventional two horizontal and one vertical outputs. If the individual sensor outputs are available, they can be calibrated with only the vertical step motion.

Measuring the Complete Response Curve $A(\omega)$

The alternative to obtain the individual sensor parameters and then calculating $A(\omega)$ is to obtain $A(\omega)$ directly. This is an empirical calibration and can only be done by providing a known input to the sensor $X(\omega)$ and measuring the output $Y(\omega)$, from which the response function can be obtained as

$$A(\omega) = Y(\omega)/X(\omega) \quad (14)$$

Shaking Table: Again a shaking table is the simplest way. The shaking table can move



Sensors, Calibration of, Fig. 8 Setup for measuring the instrument response using the calibration input (Figure from Havskov and Alguacil 2010)

vertically and horizontally in a controlled fashion so the ground input is exactly known. All that has to be done is to measure the output, as a function of frequency, and divide by the known input to get the response function. It is not as simple as it sounds. Making a precise shaking table, particularly for horizontal motion, without introducing tilt, is complicated and expensive, so shaking tables are rarely available for the general user and only used in special laboratories.

Sensor Calibration Input: Many sensors have a calibration input. For passive seismometers, this is a calibration coil around the mass, and applying a current to the coil results in a force on the mass. Most active sensors do not have a calibration coil, but an equivalent test can be made when a voltage is applied to the feedback loop to produce a force on the mass. In both cases it is possible to accelerate the moving mass by applying a voltage to the calibration input. By using a sine wave, the sensors experience the equivalent of a ground acceleration, and, by varying the frequency, the complete frequency-dependent amplitude response can be determined, since the input $X(\omega)$ is known and $Y(\omega)$ is measured. This method assumes that the calibration circuit is accurate. An experimental setup is seen in Fig. 8.

A signal generator sends out a sine wave with a constant amplitude V_0 . This voltage is applied to the sensor calibration input and exerts a force on the mass proportional to V_0 . For the velocity sensor with a calibration coil, the amplitude of the force f_i will be

$$f_i = K_c V_0 / R_c \quad (15)$$

where K_c is the calibration coil motor constant (N/A) and R_c is the resistance of the calibration coil. For the active sensor

$$f_i = K_a V_0 \quad (16)$$

where K_a is the calibration input sensitivity (N/V). With a sensor mass of m , the equivalent ground acceleration amplitude is then f_i/m , the equivalent ground velocity is $f_i/m\omega$, and the equivalent ground displacement amplitude is $f_i/m\omega^2$. If the amplitude of the seismometer voltage output is V_s , then the velocity amplitude response (V/ms^{-1}) for this frequency is

$$A(\omega) = \frac{V_s m \omega}{f_i} \quad (17)$$

and similarly for displacement response. By varying the frequency and measuring both input and output signals, a complete response curve can be obtained.

When the calibration coil is used to calibrate a passive sensor, an undesirable effect may be present: a spurious coupling between the calibration coil and the signal coil. This coupling affects mainly the higher frequencies and may be approximately corrected for by subtracting the signal output when the same current is injected with the sensor mass locked (Sauter and Dorman 1986; Steck and Prothero 1989).

From Fig. 8, it is seen that the output signal not only has been changed in amplitude but also has been delayed a little relative to the input signal; in

other words, there has been a phase shift. In this example, the phase shift is positive; see definition in Eq. 18. The complete frequency response of the sensor therefore consists of both the amplitude response function and the phase response function $\Phi(\omega)$. Considering a general input harmonic waveform $x(\omega, t) = X(\omega) \cdot \cos(\omega t)$ at frequency ω , the output can be written as

$$y(\omega, t) = X(\omega) \cdot A(\omega) \cdot \cos(\omega t + \Phi(\omega)) \quad (18)$$

The phase shift is here defined as a quantity being added to the phase as seen above. Thus comparing Fig. 8 and Eq. 18, it is seen that the phase shift is positive.

The phase response function can therefore also be measured using this harmonic drive method. It can similarly be obtained using a shaking table.

Knowing the complete response function $A(\omega)$, it is now possible to get an indication if the response function can be represented by Eq. 2 or there are additional filters to include.

Using the Ground Motion as a Shaking Table: The experiment consists of setting up two sensors close to each other and assuming that they are subject to the same ground motion. If the known and unknown sensors have recorded signals of the ground motion z_1 and z_2 , respectively, and corresponding spectra are $Z_1(\omega)$ and $Z_2(\omega)$, the input ground motion (displacement, velocity or acceleration) is

$$X_1(\omega) = \frac{Z_1(\omega)}{A_1(\omega)} \quad (19)$$

where A_1 is the response function of the known sensor 1. The unknown response function A_2 is then

$$A_2(\omega) = A_1(\omega) \frac{Z_2(\omega)}{Z_1(\omega)} \quad (20)$$

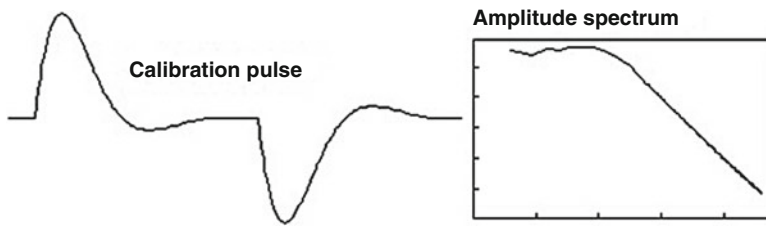
Note that if A_1 is in displacement, then A_2 will also be in displacement and similarly if A_1 is in velocity or acceleration. It is, e.g., possible to use an accelerometer to calibrate a seismometer or

vice versa. Thus, by measuring the signals of the known and unknown sensor, it is in principle simple to determine the response function of the unknown sensor. In order for the method to work, the sensor self-noise must be well below the signal levels generated by the ground motion. That will clearly limit the usefulness of the method at lower frequencies, where, e.g., for a geophone, the output signal level will be low. So, contrary to normal sensor installation, these tests should be made in a high ground noise environment like the top floor of a building, although some care is required to avoid air-coupled acoustic noise, which may affect both sensors in different ways (Pavlis and Vernon 1994).

If the spectra $Z_1(\omega)$ and $Z_2(\omega)$ of the two sensors outputs are contaminated with instrumental noise (a noncausal output), a simple ratio Z_2/Z_1 may yield a biased or unreliable estimate of the response. By using only the correlated part of the signals, i.e., the part due to the common input of both sensors, the ground noise in this case, a more reliable estimate, will be obtained. This may be done by using a known relation between input and output of linear and causal systems: the input-output cross-spectrum is the product of the transfer function and the input power spectrum (see, e.g., Ljung 1999). Consider a linear system whose input is the output of seismometer 1 (Z_1) and its output is the output of seismometer 2 (Z_2). This system would have a transfer function $A_2/A_1 = P_{21}(\omega)/P_{11}(\omega)$, or Z_2/Z_1 if these two spectra were noise-free and generated by the same input. Therefore, a more robust estimation of the unknown response A_2 is obtained using

$$A_2(\omega) = A_1(\omega) \frac{P_{21}(\omega)}{P_{11}(\omega)} \quad (21)$$

where $P_{11}(\omega)$ represents the power spectrum of the output of sensor 1 and $P_{21}(\omega)$ is the cross-spectrum between the outputs of sensors 1 and 2. The advantage of this equation is that, ideally, it cancels out the contributions of uncorrelated components (i.e., self-noise) of the input and output signals. This method has been further



Sensors, Calibration of, Fig. 9 Calibration pulse from a velocity sensor. *Left:* The recorded calibration pulse due to an input step in acceleration. *Right:* The log-log

spectrum of the calibration pulse. The sensor natural frequency ω_0 is indicated. The damping is 0.707 (Figure from Havskov and Alguacil 2010)

improved using three channels of data, resulting in more robust estimate of the relative response function (Sleeman et al. 2006). However, small misalignments between the three sensors may prevent the background noise to be fully canceled out, and a leak of such noise is then evaluated as sensor self-noise. Gerner and Bokelmann (2013) suggest a technique to numerically fix this effect by searching for the optimal alignment.

Calibration Pulse: It is common practice to generate a calibration pulse by applying a step current into the calibration coil of a passive sensor. For a BB sensor an equivalent test can be made when a voltage is applied to the feedback loop to produce a force on the mass. As the applied force is proportional to the coil current or the voltage, this input signal is equivalent to a ground step in acceleration (see also section above). The output pulse can be recorded (Fig. 9), and this pulse can be used in several ways to obtain sensor parameters.

From signal theory (e.g., Scherbaum 2007) it is known that the Fourier transform of the impulse response of a linear system is the frequency response function. However, what is generated here is a step. If the amplitude spectrum of the calibration pulse is $C(\omega)$, the response function will then be $A(\omega) = \omega C(\omega)$. This corresponds to the response function for acceleration since the input was an acceleration step so the velocity response for the velocity sensor is then $\omega^2 C(\omega)$. Thus, multiplying the spectrum in Fig. 9 with ω^2 will give the velocity response for $h = 0.707$ seen in Fig. 2. It is also possible to obtain the approximate natural frequency ω_0 as the frequency where the spectral amplitude (Fig. 9) has decreased from the flat level by a factor

$1/\sqrt{2} = 0.707$ or the frequency for which the low-frequency and high-frequency asymptotes intersect. It is then in principle simple to get the frequency response for the sensor by just doing spectral analysis of the calibration pulse; however, ground and electronic noise might make it inaccurate. In practice the signal-to-noise ratio is improved by summing several calibration pulses with their onsets well aligned. The background noise is decreased in this way, while the signal due to the calibration pulses is increased.

Summary

Calibration of seismic sensor can be done in many ways. The calibration can determine the frequency response of a sensor by determining the sensor parameters natural frequency, damping, and generator constant from which the frequency response can be calculated. The alternative is to experimentally determine the complete frequency response of the sensor using, e.g., a shaking table or the sensor calibration input. In most cases it will be possible for the general user to get a good idea about the response function using one of the methods described in this paper.

Cross-References

- [Passive Seismometers](#)
- [Principles of Broadband Seismometry](#)
- [Recording Seismic Signals](#)
- [Seismic Accelerometers](#)
- [Seismometer Self-Noise and Measuring Methods](#)

References

- Bormann P (ed) (2002) IASPEI New manual of seismological observatory practice (NMSOP). GeoForschungsZentrum Potsdam, Potsdam
- Gerner A, Bokelmann G (2013) Instrument self-noise and sensor misalignment. *Adv Geosci* 36:17–20. doi:10.5194/adgeo-36-17-2013
- Havskov J, Alguacil G (2010) Instrumentation in earthquake seismology. Springer, Dordrecht, p 358
- Lay T, Wallace TC (1995) Modern global seismology. Academic, San Diego, p 521. ISBN 0-12-732870-X
- Ljung L (1999) System identification: theory for the user, 2nd edn. Prentice Hall PTR, Englewood Cliffs, p 609
- Pavlis GL, Vernon FL (1994) Calibration of seismometers using ground noise. *Bull Seismol Soc Am* 84:1243–1255
- Sauter AW, Dorman LM (1986) Instrument calibration of ocean bottom seismographs. *Mar Geophys Res* 8:265–275
- Scherbaum F (2007) Of poles and zeros, fundamentals of digital seismology, revised 2nd edn. Springer, Dordrecht, p 271
- Sleeman R, van Wietum A, Trampert J (2006) Three-channel correlation analysis: a new technique to measure instrumental noise of digitizers and seismic sensors. *Bull Seismol Soc Am* 96(1):258–271
- Steck L, Prothero WA (1989) Seismic calibration using the simplex algorithm. *Bull Seismol Soc Am* 79:1618–1628

Site Response for Seismic Hazard Assessment

Gaetano Elia
School of Civil Engineering and Geosciences,
Newcastle University, Newcastle Upon
Tyne, UK

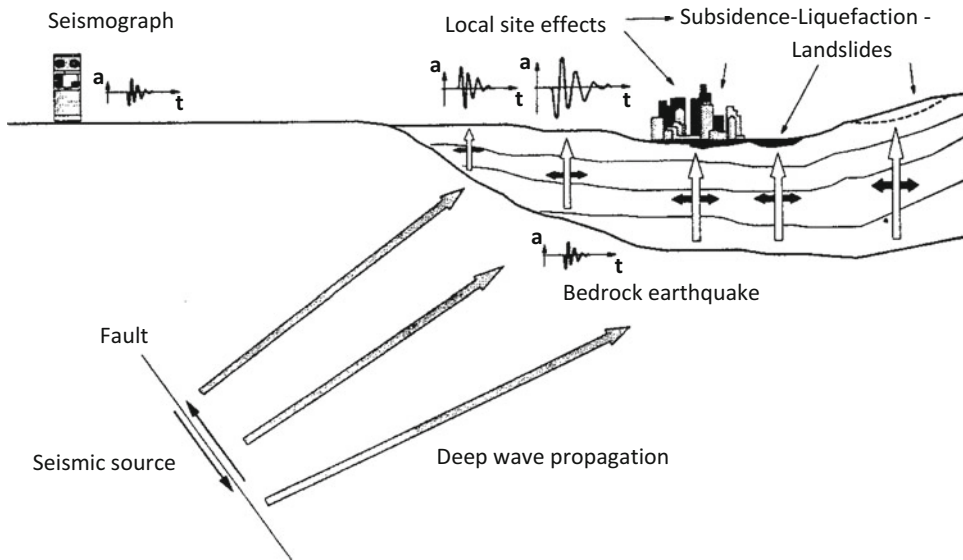
Introduction

Lessons learned worldwide from historical (e.g., Niigata, Japan, and Alaska in 1964) and recent strong earthquakes (e.g., L'Aquila in 2009, Chile in 2010, Tohoku, Japan, and Christchurch in 2011, among others) have distinguished site amplification and soil liquefaction as two of the main causes of damage to man-made and natural structures during seismic events. As illustrated in Fig. 1, the ground shaking observed at surface during an earthquake depends on the seismic source characteristics and focal mechanism, the

deep wave propagation from the fault to the bedrock, and the local soil conditions. The first two phenomena are commonly studied by seismologists, geologists, and geophysicists, while the third one falls in the geotechnical earthquake engineering field, being strongly related to the mechanical behavior of soils subjected to dynamic loading. The term “site effects” refers to the overall set of modifications of the bedrock motion, in terms of amplitude, frequency content, and duration, during its (almost) vertical propagation through shallow geological deposits. Despite the traveling path of the propagating waves within the surface soils is often less than 100 m, local site conditions can produce significant ground motion amplification, slope instability, excessive building settlements (i.e., subsidence), and liquefaction in loose and saturated granular deposits, thus playing a crucial role in building and infrastructure damage (Fig. 1).

Ground response numerical analyses are, therefore, used by geotechnical engineers to predict the free-field motion, to determine permanent soil deformations (leading to subsidence and liquefaction), and to evaluate the risk of instability of earth structures (e.g., earthquake-induced landslides). They may also include the evaluation of basin and topographic effects on ground motion (e.g., King and Tucker 1984; Bard and Bouchon 1985; Geli et al. 1988; Bard 1994; Ashford et al. 1997; Bouckovalas and Papadimitriou 2005; Semblat et al. 2005; Pagliaroli et al. 2011). In addition, the results of these studies can be incorporated into microzonation and probabilistic seismic hazard analyses (e.g., Tsai 2000; Bazzurro and Cornell 2004; Papaspiliou et al. 2012).

Site response analyses have been traditionally performed using a one-dimensional (1D) frequency-domain numerical scheme based on the equivalent viscoelastic approach. This approach has been extensively adopted in the last 30 years, and it is widely accepted in the engineering practice, although its limitations are well known. Being based on a total stress formulation, it disregards the buildup of excess pore water pressures in the soil deposit. Additionally, the adopted equivalent viscoelastic material properties cannot properly represent the soil



Site Response for Seismic Hazard Assessment, Fig. 1 Wave propagation from seismic source to ground surface and related geotechnical problems (Modified from Lanzo and Silvestri 1999)

behavior under cyclic loading for high seismic intensities at bedrock. Finally, the 1D scheme cannot take into account site effects related to surface and buried complex morphologies, i.e., topographic and valley effects. Time-domain schemes are nowadays available to solve the wave propagation problem in a more realistic way, accounting for the solid–fluid interaction by means of a coupled effective stress formulation. In those schemes, the behavior of the soil can be described using either simple or sophisticated nonlinear soil constitutive models of different level of complexity. In addition, time-domain analyses, usually performed with finite element codes, can also describe two- (2D) and three-dimensional (3D) complex geometries to model topographic and basin effects. Nevertheless, these nonlinear analyses are seldom adopted by nonexpert users because the calibration of advanced soil constitutive models can be challenging and the code usage protocols are often unclear or poorly documented in the literature.

This entry begins with a short overview of available methods for site characterization and evaluation of soil dynamic properties (section “[Soil Dynamic Properties and Measurement Techniques](#)”), factors that are essential for the

assessment of site response effects. Ground response analyses are then discussed in the section “[Ground Response Analysis](#),” describing linear and nonlinear approaches for the study of one-dimensional wave propagation problems in free-field conditions. Soil–structure interaction and earthquake-induced ground failure problems, such as soil liquefaction, landslides, and retaining structure instability, are outside the scope of the entry and, therefore, are not discussed here. A review is given in Kramer (1996), Kramer and Stewart (2004), and Semblat and Pecker (2009). Finally, the section “[Future Challenges](#)” provides an overview of future challenges in the field of geotechnical earthquake engineering.

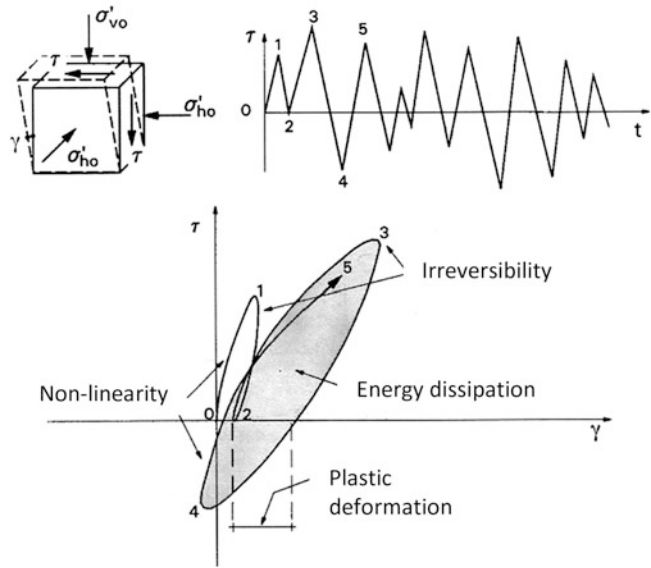
Soil Dynamic Properties and Measurement Techniques

Soil Dynamic Properties

As schematically indicated by Fig. 1, the propagation of waves from bedrock to ground surface is commonly considered as a vertical process. According to Snell’s law (cf. Richart et al. 1970), in fact, waves traveling from higher-velocity materials into lower-velocity materials

Site Response for Seismic Hazard Assessment,

Fig. 2 Mechanical behavior of a soil element subjected to an irregular simple shear loading history (Modified from Lanzo and Silvestri 1999)



are refracted closer to the normal to the interfaces. Therefore, earthquake waves propagating upward through horizontal layers characterized by lower velocities and densities (such as in typical top soil deposits) are refracted closer to a vertical path. Moreover, saturated soils subjected to earthquake loading behave essentially in undrained conditions, given the rapidity of the seismic action. The volumetric deformations induced by P-waves are, consequently, negligible with respect to the distortional deformations associated with S-waves. For the above reasons, many of the methods of ground response analysis presented in the section “[Ground Response Analysis](#)” simulate the seismic event as a loading process induced by SH-waves only, characterized by a vertical traveling path associated with soil particle motion in the horizontal plane. This phenomenon requires the analysis of the mechanical behavior of soils under simple cyclic shear loading conditions. Considering a generic soil element within the deposit in geostatic conditions (when only the vertical, σ'_{v0} , and the horizontal, σ'_{h0} , effective stresses are applied), the earthquake action induces an additional simple shear stress $\tau(t)$ changing irregularly with time (Fig. 2).

The corresponding stress–strain ($\tau - \gamma$) curve under this cyclic loading history is typically

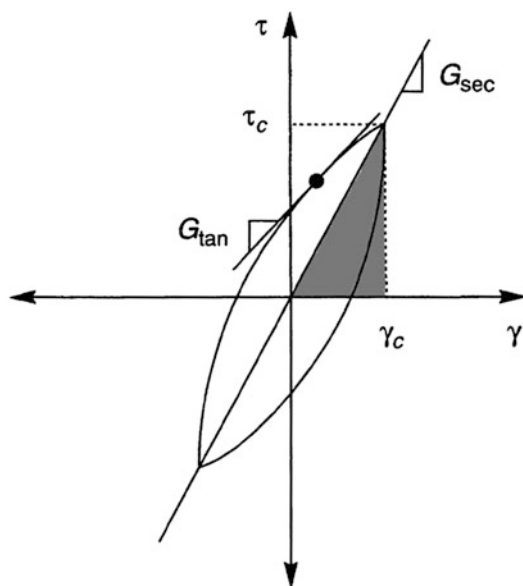
characterized by state dependency, early irreversibility, nonlinearity, buildup of excess pore pressures, decrease of nominal stiffness, and related hysteretic dissipation (e.g., Sangrey et al. 1969; Hardin and Drnevich 1972; Castro and Christian 1976; Vucetic and Dobry 1991). Under a symmetric cyclic loading condition, the hysteresis loop in the $\tau - \gamma$ plane can be effectively described by two parameters: the shear modulus and the damping ratio. As shown in Fig. 3, the tangent shear modulus $G_{\tan}$ represents the soil stiffness for a specific point of the loop, and it changes continuously throughout the cycle.

An overall indication of the average soil stiffness is, instead, represented by the secant shear modulus $G_{\sec}$, defined as:

$$G_{\sec} = \frac{\tau_c}{\gamma_c} \quad (1)$$

where τ_c and γ_c are the maximum shear stress and shear strain amplitudes, respectively. The area included in the hysteresis loop is a measure of the energy dissipated by the soil during the cycle and is described by the damping ratio D :

$$D = \frac{W_D}{4\pi W_S} = \frac{A_{\text{loop}}}{2\pi G_{\sec} \gamma_c^2} \quad (2)$$



Site Response for Seismic Hazard Assessment, Fig. 3 Evaluation of shear moduli and damping ratio from hysteresis loop (Modified from Kramer 1996)

where W_D is the dissipated energy (equal to the area of the hysteresis loop A_{loop}) and W_S is the maximum strain energy represented by the shaded area in Fig. 3.

Starting from a maximum value, i.e., G_0 or G_{max} , the soil stiffness tends to decrease with increasing shear strains. Its variation with cyclic shear strain amplitude (γ_c) is described graphically by a normalized modulus reduction curve $G/G_0 - \gamma_c$. Conversely, the energy dissipation provided by the soil, increasing with loop amplitude due to plasticity effects, is typically depicted by a $D - \gamma_c$ curve. Therefore, the mechanical characterization of the soil stiffness and damping requires both the evaluation of G and D at very low strains and the way in which the two properties change with cyclic shear strain amplitude. Laboratory tests have shown that soil stiffness and damping are also influenced by other factors, such as mean effective confining pressure, plasticity index (PI), overconsolidation ratio (OCR), and number of imposed cycles (N). In particular, extensive laboratory investigations on the cyclic response of normally consolidated and slightly

overconsolidated reconstituted clays presented by Vucetic and Dobry (1991) indicated that (Fig. 4):

- The values of OCR and effective consolidation stress have almost no effect on the position and shape of G/G_0 and D curves.
- The plasticity index PI is the key factor controlling the dimensionless parameters G/G_0 and D .
- The number of cycles N does affect both the value of G/G_0 (due to the degradation of the shear modulus with N) and the $D - \gamma_c$ curve at high cyclic strains (i.e., after the volumetric threshold).

It should be noted that the modulus reduction and damping curves reported in Fig. 4 for $PI = 0$ are nearly identical to the average curves commonly used for sands (Seed and Idriss 1970). This suggests that the curves shown in Fig. 4 can be used for both fine- and coarse-grained soils. Modulus reduction and damping ratio curves are influenced by the mean effective stress for cohesionless and low-plasticity soils, but this influence decreases with increasing plasticity index, being generally not significant for $PI \geq 30$ (Ishibashi 1992).

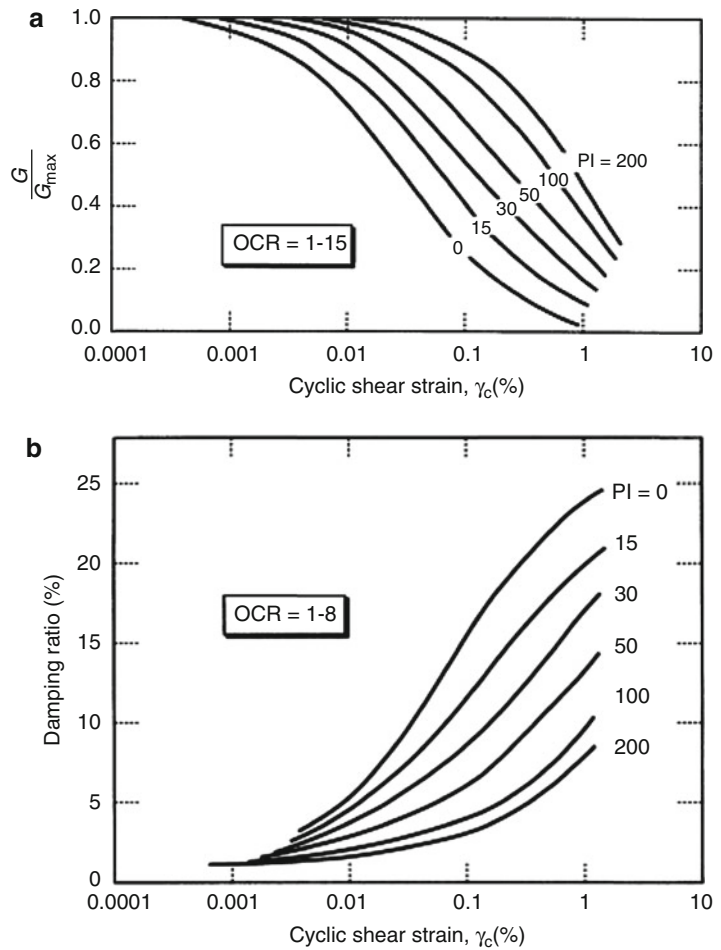
Measurement Techniques

The measurement of soil dynamic properties is a crucial task in the solution of any geotechnical earthquake engineering problem. Particularly for the evaluation of local site effects, the characterization of the soil deposit above bedrock in terms of variation of G_0 and D with depth and $G/G_0 - \gamma_c$ and $D - \gamma_c$ curves is essential. A variety of field and laboratory techniques are available, each oriented toward the measurement of low-strain properties and characterized by different advantages and limitations. A complete review of the existing techniques is outside the scope of this entry, and only the most significant ones are discussed in the following.

Low-strain field tests induce seismic waves in the soil and measure the velocities at which these waves propagate. The maximum shear modulus can be computed using the measured shear wave

Site Response for Seismic Hazard Assessment,

Fig. 4 Nonlinear behavior as function of plasticity index in terms of: (a) normalized shear modulus and (b) damping ratio (Vucetic and Dobry 1991)



velocity (V_s) and the soil density (ρ) through the following equation based on the elasticity theory:

$$G_{max} = \rho V_s^2 \quad (3)$$

Shear wave velocities can be measured in situ by several seismic tests, including cross-hole and down-hole, seismic cone penetrometer, suspension logger, SASW (spectral analysis of surface waves), and MASW (multichannel analysis of surface waves). A review of these test methods is given in Woods (1994) and Kramer (1996). Their accuracy can be sensitive to procedural details, soil conditions, and interpretation techniques. Fig. 5 shows the layouts and principles of three established geophysical tests: the continuous surface wave (CSW) method, the down-hole

test, and the cross-hole technique. In the CSW test (Fig. 5a), a mechanical, servo-hydraulic, or electromagnetic vibrator applies a single-frequency sinusoidal force to the ground surface. Rayleigh waves traveling through the soil are detected by a series of geophones (usually two) displaced at a range of distances from the source. By changing the input frequency, a profile of phase velocity against wavelength is obtained, and, consequently, a stiffness profile with depth can be computed. Although less economical than an SASW test, the CSW method has been proved to provide better data, as background noise can be easily recognized and filtered. Likewise the CSW test, the MASW approach uses a multiple of equally spaced receivers (usually 12 to 60) that are deployed on the surface along a survey line.

Each receiver is connected to a common multichannel recording instrument (i.e., a seismograph). This is the most significant difference between the CSW and the MASW techniques, as CSW is usually based on a two-receiver approach. Also the MASW method generally uses an impulsive source, such as a sledgehammer, to produce surface waves, whereas the CSW technique makes use of a frequency-controlled vibrator.

Down-hole and cross-hole tests are alternative low-strain techniques which require one or more borings. In a down-hole test, a vibration source is placed on the ground surface adjacent to a borehole. The arrival of seismic energy is detected at depth either by geophones secured against the borehole sides or by geophones within a seismic CPT (Fig. 5b). The test is repeated changing the depth of the geophones (typically at 1 m intervals) to plot the shear wave travel time as function of depth. The average shear velocity can be computed by knowing the distance between the source and the receiver. The cross-hole test makes use of more than one boring (usually two, less commonly three as this latter option is more expensive): a source is placed in one boring and a receiver is placed at the same depth in each of the other boreholes (Fig. 5c). An impulsive disturbance is applied at the source and the travel time to each receiver is measured. A borehole verticality survey is required in order to calculate the actual distance between the boreholes at each test depth (usually 1 m intervals). The wave propagation velocity is, in fact, computed by knowing the distances between receivers. The use of two sets of receivers (three boreholes) avoids the issue of trigger accuracy, but increases the cost of the test. While CSW and down-hole tests allow to determine the shear modulus for distortion in vertical plane (G_v) as the source produces a vertically polarized horizontally traveling shear wave, the cross-hole technique allows to calculate also the shear modulus in the horizontal plane (G_h) by using horizontally polarized shear wave sources.

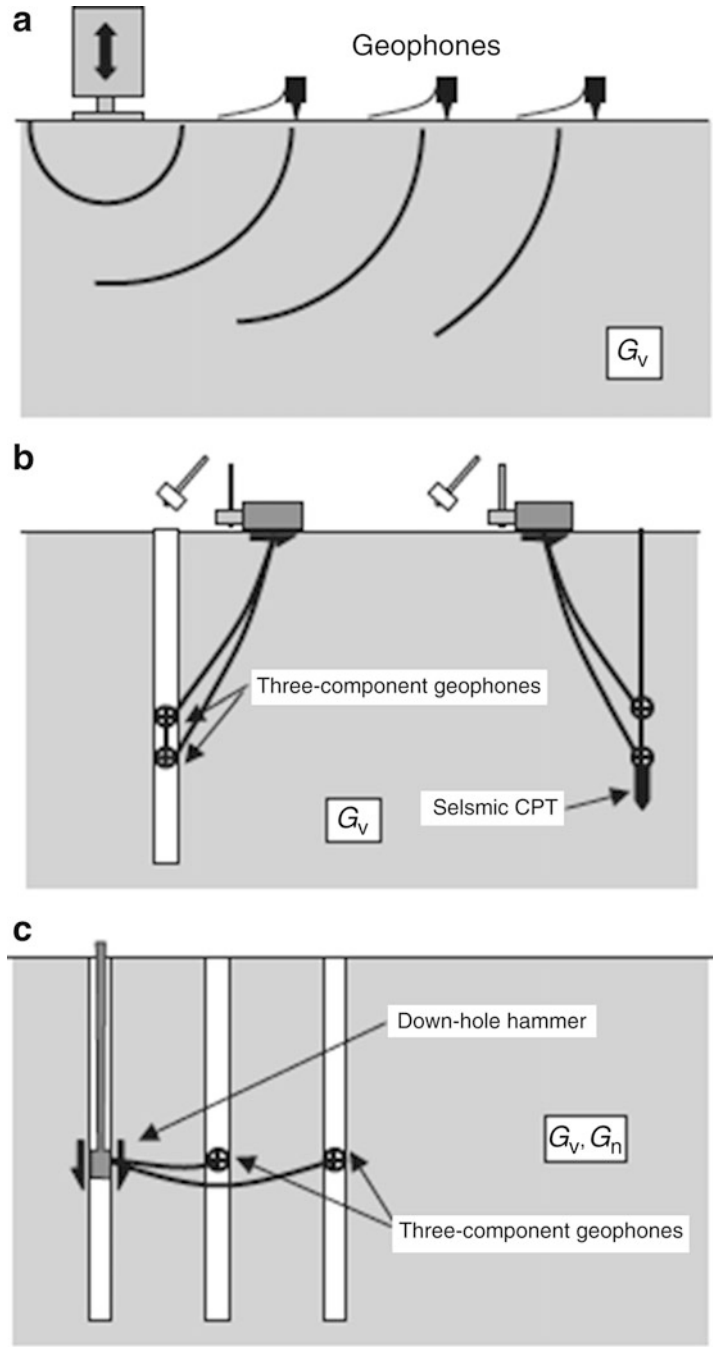
The standard penetration test (SPT) and the cone penetration test (CPT), although originally developed for the measurement of soil properties

mobilized at large strains, can be indirectly used to determine shear modulus profiles in situ by using empirical correlations between penetration resistance (N_{SPT}) or tip resistance (q_c) values and G_{max} (a review is given in Kramer 1996). These estimates are affected by high uncertainties and should be used very cautiously, given the scatter in the data on which they are based and the variability in the results obtained by different correlations. Such correlations should be adopted only for preliminary estimates of G_{max} in the framework of simplified approaches.

Finally, standard spectral ratio (SSR) and horizontal-to-vertical spectral ratio (HVSr) methods for the determination of the deposit fundamental frequencies are becoming increasingly popular not only in the research field but also in professional practice (i.e., for microzonation studies). Site amplification factors can, in fact, be inferred, at least in the linear strain range, using the SSR technique described by Field and Jacob (1993). HVSr amplifications obtained experimentally can, instead, be used to validate numerical model results (e.g., SESAME 2004; Lanzo et al. 2011).

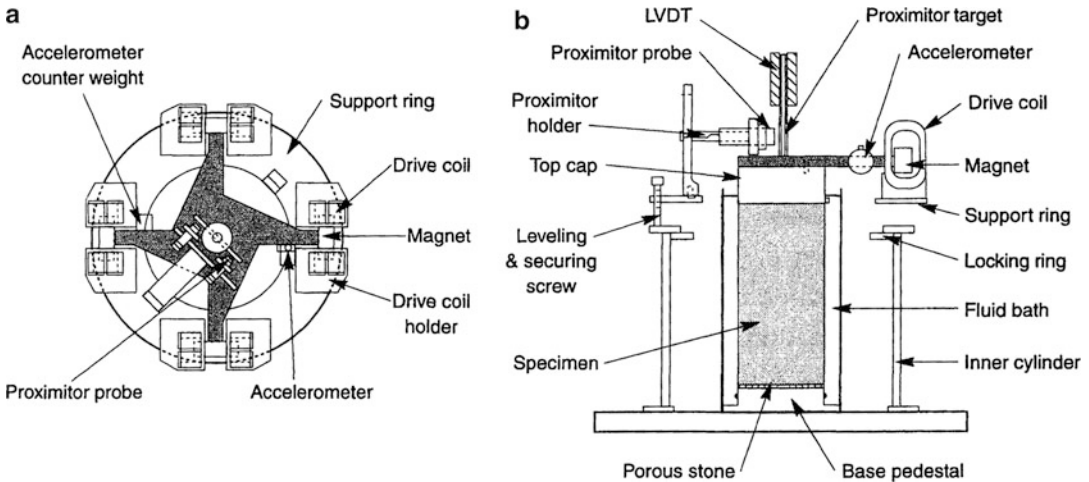
In general, field tests have the advantage to describe the dynamic properties of the soil as it is in situ. Laboratory testing methods are, on the contrary, usually performed on relatively small specimens that can be disturbed by the sampling technique and may not be representative of the larger body of soil from which they are retrieved. Nevertheless, cyclic and dynamic laboratory tests are complementary to field methods as they can provide the description of soil nonlinearity over a wide range of shear strains in terms of $G/G_0 - \gamma_c$ and $D - \gamma_c$ curves. Typical low-strain element techniques are the resonant column (RC) and the bender element (BE) tests. A schematic view of the resonant column apparatus is presented in Fig. 6: a solid or hollow cylindrical specimen is subjected to harmonic torsional or axial loading by an electromagnetic loading system. The system usually applies a harmonic load for which the frequency and amplitude can be controlled. The fundamental frequency of the specimen can be identified by gradually increasing the loading frequency. Given the mass polar moment of

Site Response for Seismic Hazard Assessment, Fig. 5 Field seismic tests: (a) continuous surface wave (CSW), (b) down-hole, and (c) cross-hole (Clayton 2011)



inertia of the loading head and the specimen mass and dimensions, the shear modulus of the soil can be calculated, assuming linear elasticity (i.e., using Eq. 3). Repeating the test with increasing loading amplitude, the variation of secant shear modulus with shear strain can also be measured.

Damping can be determined using the half-power bandwidth method or from the logarithmic decrement by placing the sample in free vibration (ASTM 2007). This technique has been modified to allow cyclic torsional shear testing to strain levels above those typically achieved during a



Site Response for Seismic Hazard Assessment, Fig. 6 Schematic drawing of a typical resonant column apparatus: (a) top view and (b) profile view (Kramer 1996)

conventional resonant column test. It is well established that the large strain rates related to the high frequencies applied in an RC test can affect the measured small-strain shear modulus and damping ratio. Dry cohesionless soils usually do not exhibit rate effects, while high strain rates can produce an increase of shear stiffness at small strains in fine-grained plastic soils leading to a G_0 overprediction with respect to cyclic simple shear tests (e.g., Lo Presti et al. 1997; Cavallaro et al. 2003). Moreover, the laboratory results indicate that some energy is always dissipated by the soil, even at very low strains, so that the damping ratio is never equal to zero. This mechanism cannot be justified by plasticity theory, as no hysteretic dissipation of energy associated to the development of plastic deformations takes place at strain levels within the elastic domain. Therefore, the initial damping ratio D_0 observed in resonant column experiments can be probably attributed to material viscous effects and/or the inertia of the resonant column apparatus (e.g., Meng and Rix 2003; Lo Presti et al. 2007).

Another laboratory technique that allows the measurement of the shear wave velocity of a soil sample is the bender element testing method. Bender elements consist of two piezo-ceramic plates bonded together in such a way that application of a voltage causes one plate to expand and the other to contract. This generates seismic waves in

the soil sample in which the device is embedded. At the same time, a lateral disturbance of a bender element produces a voltage, thus allowing the detection of incoming waves. Therefore, BE can be used as both transmitters and receivers of S-waves within a soil sample (typically triaxial). By measuring the time required for the wave to travel from the source to the receiver and knowing the distance between each, the shear wave velocity of the specimen can be measured.

Finally, laboratory tests able to measure the dynamic soil properties at high strain levels have been derived from conventional tests by adding cyclic loading capabilities to the testing apparatus. Examples are represented by cyclic triaxial tests with local strain measurements and cyclic simple shear tests.

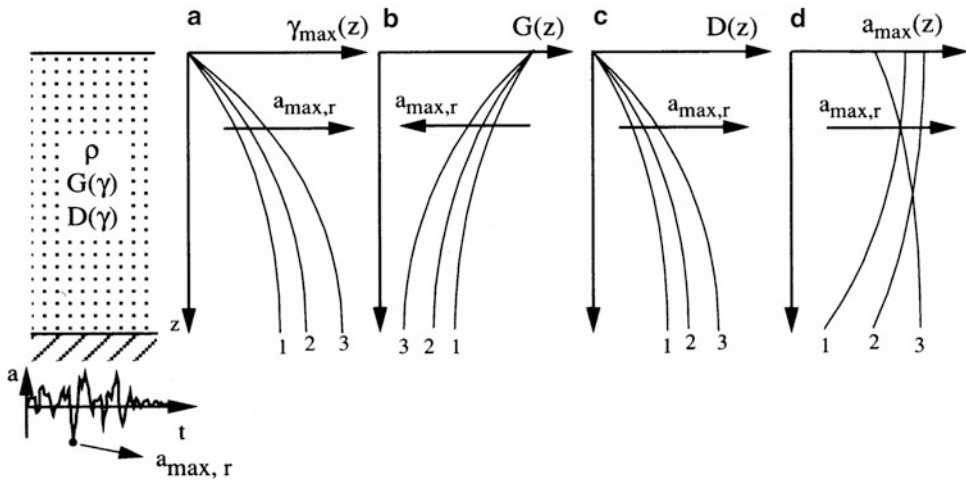
Table 1 summarizes the range and applicability of the most common field and laboratory cyclic and dynamic techniques, with particular reference to the induced shear strain level, characteristic frequency, and dynamic soil properties which can be obtained from the tests.

Ground Response Analysis

Analytical solutions have been developed to solve one-dimensional wave propagation problems, assuming that the input motion is harmonic and

Site Response for Seismic Hazard Assessment, Table 1 Range and applicability of field and laboratory dynamic tests (Modified from Vinale et al. 1996)

Test type			Shear strain γ (%)	Frequency f (Hz)	Shear modulus	Damping ratio
Field	Standard	SPT	–	–	$N_{\text{SPT}} \rightarrow V_S \rightarrow G_0$	–
		CPT			$q_c \rightarrow V_S \rightarrow G_0$	
	Geophysics	Down-hole	$<10^{-3}$	$10 \div 100$	$V_S \rightarrow G_0$	Possible
		Cross-hole			$V_S \rightarrow G_0$	
		SASW			$V_R \rightarrow V_S \rightarrow G_0$	
Lab	Cyclic	Triaxial	$>10^{-2}$	$0.01 \div 1$	$\sigma - \varepsilon \rightarrow E \rightarrow G$	Hysteresis loop $\rightarrow D$
		Simple shear	$>10^{-2}$	$0.01 \div 1$	$\tau - \gamma \rightarrow G$	
		Torsional shear	$10^{-4} \div 1$	$0.01 \div 1$	$\tau - \gamma \rightarrow G, G_0$	
	Dynamic	Bender elements	$<10^{-3}$	>100	$V_S \rightarrow G_0$	Possible
		Resonant column	$10^{-5} \div 10^{-1}$	>10	$f_r \rightarrow G, G_0$	H-p or log. decrement $\rightarrow D$



Site Response for Seismic Hazard Assessment, Fig. 7 Profiles with increasing bedrock acceleration of (a) maximum shear strain, (b) shear modulus, (c) damping

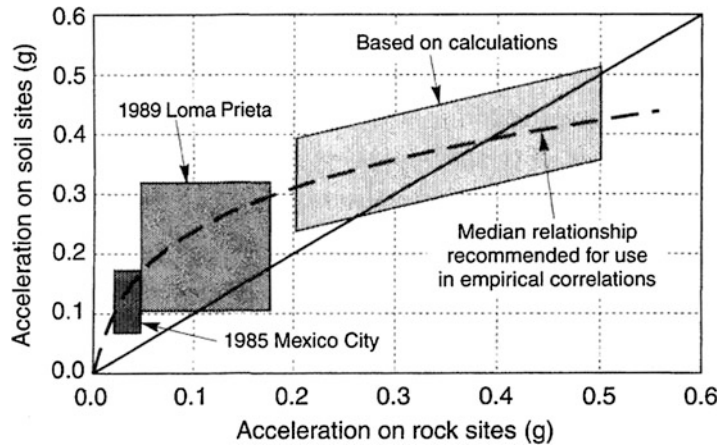
ratio, and (d) maximum acceleration mobilized within the soil deposit (Lanzo and Silvestri 1999)

the soil is a homogeneous elastic or viscoelastic material (e.g., Roesset 1977). They make use of amplification functions to describe the dynamic response of a soil deposit (considered as a single-degree-of-freedom system) in the frequency domain. This approach is limited to the analysis of linear systems, as it relies on the principle of superposition. When the soil deposit is layered, heterogeneous (i.e., G_0 changes with depth), and

nonlinear (i.e., G and D are functions of the shear strain induced by the earthquake), the amplification function is no more constant (as in the analytical solutions) but site specific. The effects of soil nonlinearity on wave propagation are qualitatively illustrated in Fig. 7, where a homogeneous deposit overlying a horizontal bedrock is subjected to an input motion characterized by increasing values of the maximum acceleration ($a_{\text{max},r}$).

Site Response for Seismic Hazard

Assessment, Fig. 8 Peak acceleration on soft soil sites as function of bedrock maximum acceleration (Idriss 1990)



Increasing the peak ground acceleration at bedrock:

- The mobilized shear strain increases (Fig. 7a).
- The shear modulus reduces and the damping ratio increases correspondingly (Fig. 7b, c).
- For low input energy, the peak acceleration along the soil profile increases moving from the bedrock to the ground surface (profiles 1 and 2 in Fig. 7d).
- For high input energy, the peak acceleration along the profile can reduce because the nonlinear behavior reduces the soil shear stiffness, thus preventing from the transmission of high frequencies, while the corresponding increase of damping ratio reduces the displacement and acceleration amplitude peaks (profile 3 in Fig. 7d).

In terms of amplification factor (given as the ratio between the max acceleration at surface to the max acceleration at bedrock), nonlinearity effects can lead to a reduction of the amplification at surface with increasing acceleration levels at bedrock. On the basis of data recorded during Mexico City (1985) and San Francisco Bay area (1989) earthquakes, and of additional ground response analyses, Idriss (1990) related peak accelerations on soft soil deposits to those on rock sites, as reported in Fig. 8. For acceleration levels lower than about 0.4 g, the peak acceleration at soft soil surface is likely to

be higher than on rock. At higher acceleration levels, on the contrary, nonlinearity can prevent from the development of peak accelerations at surface as large as those observed on rock sites.

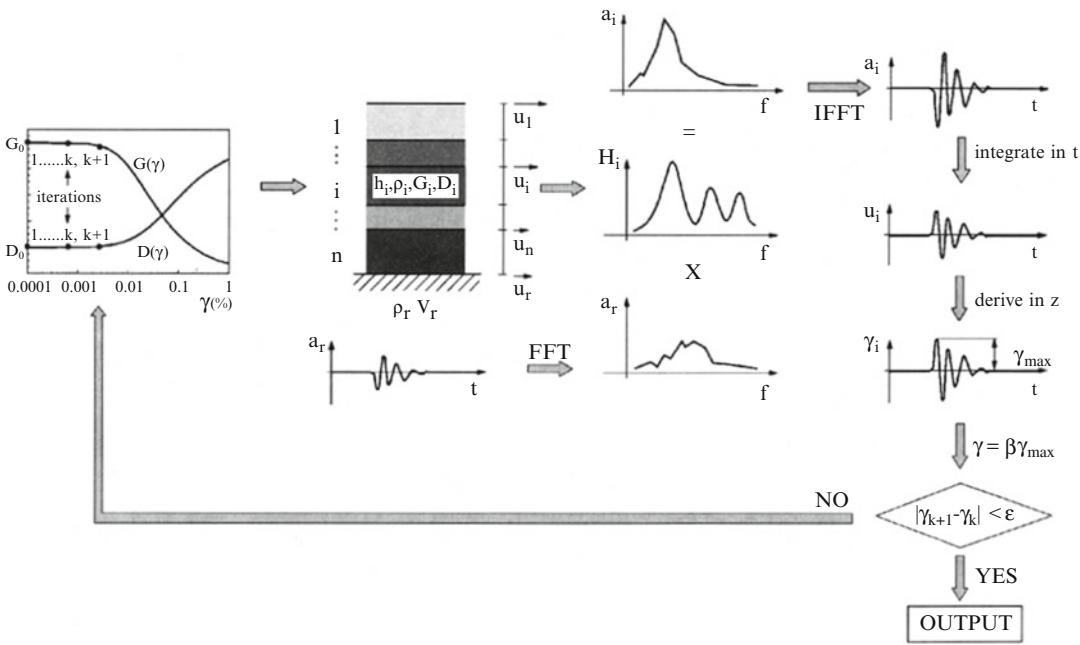
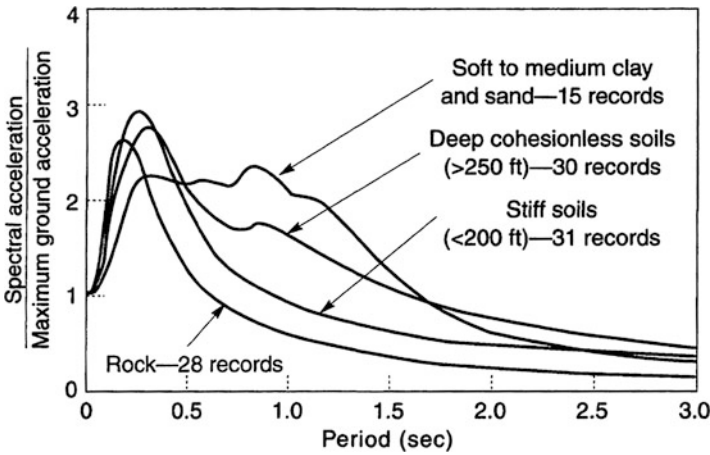
Local site conditions not only influence the peak acceleration amplitudes but can also strongly affect the frequency content of surface motions. Fig. 9 shows the effects of local soil conditions on the shape of the normalized response spectra computed from ground motions recorded on different sites: for periods above 0.5 s, spectral accelerations for soil sites are higher than those for rock sites. The figure clearly indicates that deep and soft soil deposits enhance the transmission of low frequencies (high periods). The results also show that the use of a single response spectrum shape for all site conditions is not appropriate. This evidence has been incorporated in a large number of seismic codes worldwide which propose the use of different spectral shapes for different subsoil conditions.

Equivalent Linear Approach

Nonlinear soil behavior can be approximated by an equivalent linear characterization of soil dynamic properties. The method makes use of the exact continuum solution of wave propagation in horizontally layered viscoelastic materials subjected to vertically propagating transient motions (e.g., Roesset 1977). It models the nonlinear variation of soil shear modulus and

Site Response for Seismic Hazard Assessment,

Fig. 9 Normalized response spectra (5 % damping) for different local site conditions (Seed et al. 1976)



Site Response for Seismic Hazard Assessment, Fig. 10 Equivalent linear algorithm (Modified from Lanzo and Silvestri 1999)

damping with shear strain through a sequence of linear analyses with iterative update of stiffness and damping parameters. For a given soil layer, G and D are assumed to be constant with time during the shaking. Therefore, an iterative procedure is needed to ensure that the properties used in the linear dynamic analyses are consistent with the level of strain induced in each layer by the input motion.

The iterative procedure, illustrated in Fig. 10, operates as follows:

1. Initial estimates of G and D are made for each layer at low-strain values.
2. These initial values are used in a linear frequency-domain dynamic analysis where the acceleration time history at bedrock is transformed through a fast Fourier transform

(FFT) algorithm and the transfer functions between the bedrock and each layer (H_i in Fig. 10) are computed.

3. The inverse fast Fourier transform (IFFT) algorithm is employed to obtain acceleration, displacement, and shear strain time histories at each layer.
4. The effective shear strain (γ) in each layer is determined from the maximum shear strain ($\gamma_{\max}$) as follows:

$$\gamma = \beta \gamma_{\max} \quad (4)$$

where $\beta = (M - 1)/10$ is usually related to the expected earthquake magnitude M .

5. From this effective shear strain, new equivalent linear values of G and D are chosen for the next iteration.
6. Steps 2–5 are repeated until a convergence criterion is satisfied.

The most widely used computer software for one-dimensional ground response analysis, based on the equivalent linear approach described above, is SHAKE (Schnabel et al. 1972) and its modified version SHAKE91 (Idriss and Sun 1992). For two-dimensional geometries, the equivalent linear approach has been implemented in a number of codes such as FLUSH (Lysmer et al. 1975) and QUAD4M (Hudson et al. 1994), among others. Equivalent linear methods are extensively adopted in engineering practice for their simplicity, flexibility, and low computational requirements. Nevertheless, their limitations are well known. The analysis is performed adopting a total stress approach. This means that excess pore water pressures induced by the earthquake cannot be predicted and the displacements due to consolidation processes cannot be calculated. The model employed to describe the mechanical behavior of soils is viscoelastic. Therefore, the method cannot predict permanent soil displacements or cumulated strains at the end of the analysis. Even though the iteration process allows to approximate nonlinear soil behavior, the approach is still a linear method of analysis. The strain-compatible soil properties are constant throughout the duration of the earthquake. The method is thus not

capable of representing changes in soil stiffness and hysteretic damping during the seismic action.

For problems where strain levels remain low (stiff soil profile and/or relatively weak input motions), the equivalent linear method can produce reasonable estimates of ground response. For high seismic intensities at bedrock, nonlinear time-domain analyses should be preferred as they are likely to provide better results (section “[Verification and Parametric Studies](#)”).

Nonlinear Approach

In recent years, the use of fully coupled effective stress formulations, based on Biot’s theory for solid–fluid interaction (Biot 1941; Zienkiewicz et al. 1999), is becoming increasingly popular as the theoretical basis for the dynamic analysis of geotechnical structures (e.g., Dewoolkar et al. 2001; Elgamal et al. 2002a; Aydingun and Adalier 2003; Dakoulas and Gazetas 2005; Liu and Song 2005; Madabhushi and Zeng 2007; Sica et al. 2008; Alyami et al. 2009; Shahrouar et al. 2010; Elia et al. 2011; Kontoe et al. 2011; Elia and Rouainia 2013, 2014). Specifically for ground response analysis, nonlinear approaches are able to solve the wave propagation problem by direct numerical integration in the time-domain, accounting for the development of plastic deformations and buildup of excess pore water pressures within the soil deposit induced by the earthquake. In those schemes, the mechanical behavior of the soil can be described using either simple or sophisticated nonlinear constitutive models of different level of complexity. In terms of spatial discretization, most numerical approaches are modelling the problem in 1D or 2D plane strain conditions. Nonlinear three-dimensional (3D) ground response analyses are rarely found in the literature, being only performed for critical structures, such as tunnels, bridges, and earth dams (e.g., Elgamal 1992; Stamos and Beskos 1995; Elgamal et al. 2008; Ou 2009; Hatzigeorgiou and Beskos 2010).

A summary of the available computer software, employing equivalent linear and nonlinear approaches to solve the wave propagation problem in one- and two-dimensional conditions, is reported in Table 2.

Site Response for Seismic Hazard Assessment,

Table 2 Ground response analysis software (Modified from AGI 2005)

Geometry	Software	Type of analysis	
1D	SHAKE (Schnabel et al. 1972)	TS	EL
	SHAKE91 (Idriss and Sun 1992)		
	PROSHAKE (EduPro Civil Systems 1998)		
	SHAKE2000 (www.shake2000.com)		
	EERA (Bardet et al. 2000)		
	TESS (Pyke 1992)	TS	NL
	NERA (Bardet and Tobita 2001)		
	DEEPSOIL (Hashash 2009)		
	DESR_2 (Lee and Finn 1978)	ES	
	DESRAMOD (Vucetic 1986)		
	D-MOD_2 (Matasovic 1995)		
	D-MOD2000 (GeoMotions 2007)		
	SUMDES (Li et al. 1992)		
	CYCLIC 1D (www.soilquake.net)		
2D	QUAD4 (Idriss et al. 1973)	TS	EL
	QUAD4M (Hudson et al. 1994)		
	FLUSH (Lysmer et al. 1975)		
	QUAKE/W (GeoSlope 2002)		
	SPECTRA (Borja and Wu 1994)	TS	NL
	DYNAFLOW (Prevost 2002)		
	GEFDYN (Aubry and Modaressi 1996)	ES	NL
	TARA-3 (Finn et al. 1986)		
	FLAC 2D (Itasca 2002)		
	PLAXIS 2D (www.plaxis.nl)		
	SWANDYNE II (Chan 1995)		
	OpenSees (McKenna and Fenves 2001)		

TS total stress, *ES* effective stress, *EL* equivalent linear, *NL* nonlinear

Verification and Parametric Studies

The effectiveness of the different numerical approaches, briefly described in the previous sections, to simulate the complex wave propagation process has been tested over the last decades using real vertical array data and/or centrifuge test results. In these cases, in fact, the input motion is reasonably well defined, as it is directly recorded on rock or it is imposed during the test. In addition, validation studies have been conducted in several international benchmark projects, such as ESG-IASPEI/IAEE 1992 (Turkey Flat and Ashigara Valley arrays), ESG-IASPEI/IAEE 2006 (Grenoble basin), Turkey Flat 2008, and E2VP 2010 (EuroSeisTest, Greece).

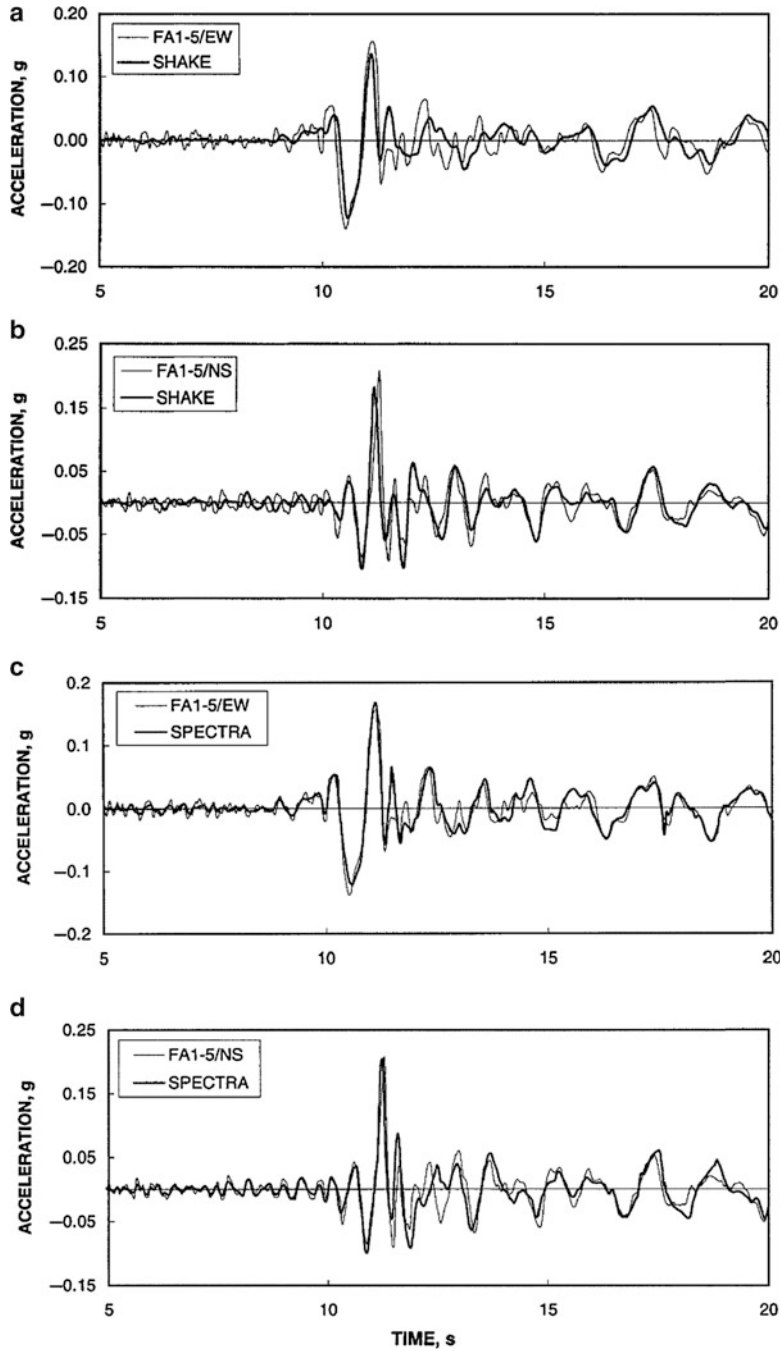
As an example, Borja et al. (1999) presented a verification study using the recordings from accelerometer arrays installed at the Lotung Large-Scale

Seismic Test (LSST) site in Taiwan. The site was established in 1985 to study the dynamic behavior of two scaled-down nuclear plant containment structures constructed by the Electric Power Research Institute (EPRI) and the Taiwan Power Company.

Borja and coworkers studied the ground response at Lotung employing both a standard linear equivalent method (SHAKE) and an advanced nonlinear approach, where a bounding surface soil constitutive model (formulated in terms of total stresses) was implemented into the three-dimensional finite element (FE) code SPECTRA. Fig. 11a, b compares the recorded ground surface accelerations along the EW and NS directions (i.e., FA1-5/EW and FA1-5/NS) with the corresponding predictions obtained through SHAKE. The peak values of both EW and NS accelerations recorded at ground surface

Site Response for Seismic Hazard Assessment,

Fig. 11 Comparison of ground surface accelerations recorded at Lotung (EW and NS components) and predictions by SHAKE and SPECTRA (Borja et al. 1999)

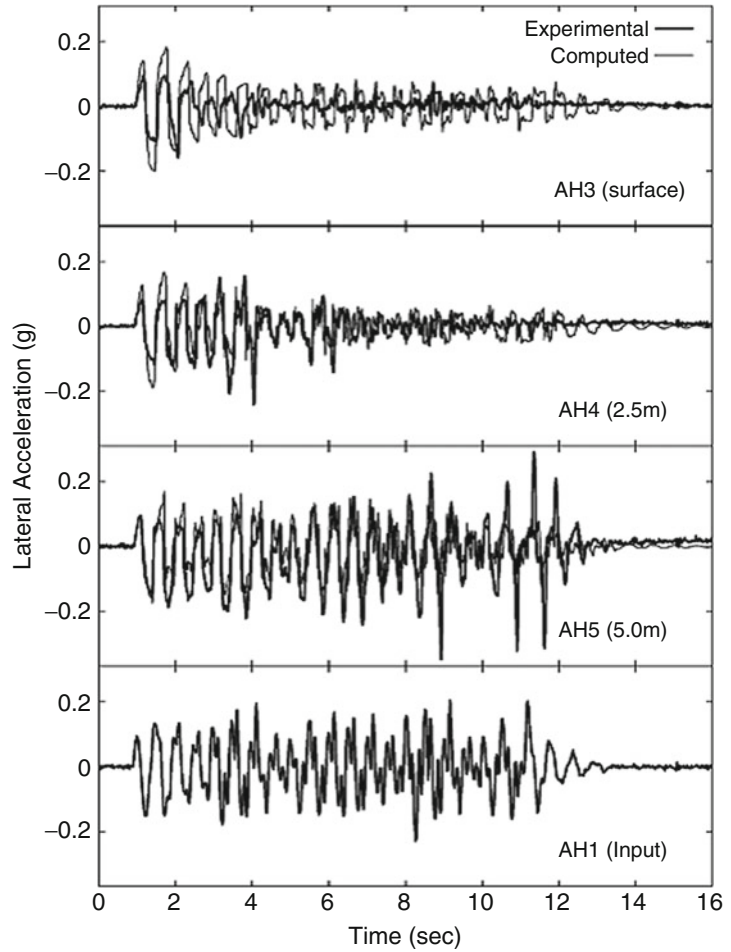


were slightly underpredicted. Moreover, a time shift in the acceleration peak between the recorded and computed NS motions can be observed. On the contrary, Fig. 11c, d shows that peak accelerations were predicted by the nonlinear model within 10–20 % error in both

directions. An excellent match also in terms of zero crossings was obtained using the advanced nonlinear time-domain approach. The advantage of the described approach lies in its capacity to extend the analysis to three dimensions, thus allowing the simultaneous application of the

Site Response for Seismic Hazard Assessment,

Fig. 12 VELACS model 1 recorded and computed acceleration time histories (Elgamal et al. 2002b)



three components of the earthquake (one vertical and two horizontal) at the base of the FE column.

Another example of verification study, this time using centrifuge testing data, is provided by Elgamal et al. (2002b). The authors adopted an advanced plasticity-based constitutive model to simulate the cyclic mobility response mechanism typically observed in medium-dense sands subjected to cyclic loading. The model was implemented into an unnamed fully coupled finite element code to predict the dynamic response observed during different centrifuge tests (VELACS project, Arulanandan and Scott 1993). In particular, Fig. 12 shows the comparison between the computed acceleration time histories and the ones recorded for VELACS model 1 (representing a level site composed of Nevada

sand and subjected to a 2Hz harmonic base excitation). Good agreement was achieved between the computed and recorded responses at different depths within the soil deposit in terms of accelerations, although the numerically predicted settlement due to liquefaction was generally smaller than observations. In addition, the work by Elgamal and coworkers demonstrates how, under level ground conditions, shear strains can be relatively small with minor cyclic mobility effects, while mildly sloping ground may result in large cyclic shear strain accumulation leading to unbounded flow failure. The importance of the dominant excitation frequency on post-liquefaction soil response is also highlighted.

The two examples described above are representative of the state of the art in the field of

nonlinear ground response analysis, but other verification studies are reported by Kramer and Stewart (2004). Although a direct comparison with recorded ground response during seismic actions is invaluable, the performance of advanced numerical approaches can also be usefully evaluated through parametric analyses. These numerical studies can, in fact, provide helpful guidelines to nonexpert users of time-domain nonlinear schemes by clarifying the importance of the different ingredients required to perform sophisticated simulations. One of the most controversial factors affecting FE dynamic simulations is the calibration of the viscous damping, usually needed in nonlinear analyses when the adopted soil constitutive model is unable to provide enough hysteretic energy dissipation through plasticity. Viscous damping is typically introduced in FE codes by means of the Rayleigh formulation, whose damping matrix is defined as follows:

$$[C] = \alpha_R [M] + \beta_R [K] \quad (5)$$

where $[M]$ and $[K]$ are the mass and the stiffness matrix of the system, respectively. The coefficients α_R and β_R are obtained considering the following relationship with the damping ratio D :

$$\begin{Bmatrix} \alpha_R \\ \beta_R \end{Bmatrix} = \frac{2D}{\omega_m + \omega_n} \begin{Bmatrix} \omega_m \omega_n \\ 1 \end{Bmatrix} \quad (6)$$

where ω_m and ω_n are the angular frequencies related to the frequency interval $f_m \div f_n$ over which the viscous damping is equal to or lower than D . The amount of Rayleigh damping to be introduced in an FE dynamic analysis is difficult to quantify a priori, but, at the same time, its magnitude can play a crucial role on the results of the simulations (Woodward and Griffiths 1996). Different possible calibration procedures have been proposed in the literature to identify the interval $f_m \div f_n$. A well-established one (e.g., Hudson et al. 1994) suggests to select f_m as the first natural frequency of the deposit f_I , while f_n is assumed equal to n times f_m , where n is the closest odd integer

larger than the ratio f_p/f_I between the predominant frequency of the input earthquake motion (f_p) and the fundamental frequency of the soil deposit (f_I). This latter assumption was based on the evidence that the higher modes of a shear beam are odd multiples of the fundamental mode of the beam. Recently, Kwok et al. (2007) used linear frequency-domain solutions to provide guidelines on the specification of viscous damping adopted in four 1D nonlinear codes (namely, D-MOD_2, DEEPSOIL, OpenSees, and SUMDES). The idea was to assume the results of 1D linear equivalent analyses as target solutions for the nonlinear approaches, thus guiding the user in the calibration of the viscous damping parameters. The main conclusions of this parametric investigation can be summarized as follows:

- The target damping ratio should be set to the small-strain material damping.
- The two target frequencies should be set, as a first approximation, to the first mode site frequency and five times the site frequency (i.e., $n = 5$).
- More generally, the target frequencies should be established through an iterative process by which linear time-domain and frequency-domain solutions are matched.

Similarly to Kwok et al. (2007), Amorosi et al. (2010) performed a parametric investigation using two different finite element codes (i.e., PLAXIS 2D and SWANDYNE II) and compared the results of 2D FE viscoelastic simulations with 1D linear equivalent solutions (i.e., EERA). In order to provide a useful framework for finite element users, some of the factors potentially influencing the numerical results were critically discussed. In particular, the amount of viscous damping adopted in FE viscoelastic analyses, the spatial and time discretization, and the nature of boundary conditions were examined. The investigation showed that:

- The traditionally adopted procedures for the calibration of the Rayleigh coefficients (e.g., Hudson et al. 1994) can lead to large

overestimation of the peak ground acceleration with respect to the EERA solutions. Therefore, a novel calibration procedure was proposed based on the linear equivalent amplification function and the frequency content of the input motion. It was suggested that the first target frequency (f_m) should be set to the first mode site frequency significantly excited by the seismic motion, while the second (f_n) should be selected equal to the frequency where the amplification function gets lower than one.

- As regards the effect of boundary conditions adopted in the FE simulations, tied nodes at the lateral boundaries of the mesh should be preferred, as they are properly representative of a 1D condition and allow to obtain a perfect match with EERA. When using Lysmer and Kuhlemeyer (1969) boundaries, the 2D mesh should be characterized by a width to height ratio between five and eight to avoid spurious wave reflections at the vertical boundaries.
- In terms of spatial discretization, the FE mesh should always satisfy the condition that the spacing between the nodes, Δl_{node} , must be smaller than approximately one-tenth to one-eighth of the wavelength associated with the maximum frequency component f_{max} of the input wave.
- The time-step algorithm adopted in the FE code can introduce some numerical damping in the integration of the governing equations of motion, and, therefore, the time-stepping coefficients need to be selected carefully in order not to lose accuracy.

Two alternative soil damping formulations for small and large strains have been recently implemented by Phillips and Hashash (2009) in the nonlinear one-dimensional site response analysis code DEEPSOIL. The first one introduces an approach to construct a frequency-independent viscous damping matrix which reduces the overdamping at high frequencies and, therefore, the filtering at those frequencies. A good match with frequency-domain analysis results can be obtained with this new approach, which, nonetheless, requires the calculation of the eigenvalues

and eigenvectors of the matrix $[M]^{-1}[K]$. The second formulation introduces a reduction factor that modifies the extended Masing loading/unloading stress-strain relationship (Masing 1926) to match measured modulus reduction and damping curves simultaneously over a wide range of shear strains. This modified hysteretic model allows to reduce the typically observed overestimation of damping predicted by advanced soil constitutive models for large strains (see next section). The proposed models have been implemented in a total stress nonlinear code and, therefore, cannot predict accumulation of excess pore water pressures during the shaking.

Future Challenges

While during recent years important advances have been made in the field of geotechnical earthquake engineering, some challenges still remain to be addressed.

In situ and laboratory characterization of soil dynamic properties represents a crucial aspect of any seismic design of geotechnical structures. Techniques for more accurate characterization of in situ soil shear wave velocity and damping ratio are required (e.g., Hall and Bodare 2000). Although expensive and time consuming, advanced laboratory cyclic and/or dynamic tests are essential for the correct understanding of soil behavior during seismic excitations and for the proper calibration of sophisticated constitutive assumptions introduced in nonlinear ground response analyses. At the same time, these tests should be interpreted within the framework of plasticity theory, as irreversible deformations in soil samples can develop also for small imposed strains.

With respect to site response analysis, the use of nonlinear time-domain approaches should be encouraged as a viable alternative to the standard linear equivalent method. It can, in fact, better predict soil deformation, degradation of stiffness, and accumulation of excess pore water pressures throughout the shaking. The issue of the accurate prediction of hysteretic damping with advanced soil constitutive models still remains controversial, as these models can significantly

underpredict the damping ratio for small strains and overestimate it in the large strain range. Some adjustments in the mathematical formulation of advanced soil models have been proposed recently (e.g., Phillips and Hashash 2009; Seidalinov and Taiebat 2014) to address this point, but more research work is needed. In general, clear guidance and protocols should be provided to nonexpert users of time-domain codes in order to improve the reliability of nonlinear ground response analysis results.

Another open challenge in the area of seismic risk assessment is related to the selection of appropriate input motions for nonlinear ground response simulations. The correct definition of the design seismic actions, based on seismic hazard and site response analyses, is, indeed, essential to fully implement the performance-based design approach proposed by code prescriptions (e.g., Eurocode 8). The options available to engineers in terms of input acceleration time series are represented by artificial, synthetic, and real accelerograms. Spectrum-compatible artificial records are characterized by excessive number of cycles of strong motion, and consequently they possess unreasonably high-energy content. It is now widely accepted that the use of these artificial records is problematic and not suitable particularly for nonlinear analyses (e.g., Bommer and Acevedo 2004). Synthetic accelerograms generated from seismological source models are highly sensitive to the definition of earthquake source parameters, which involve a significant degree of expert judgment. On the contrary, real earthquake accelerograms represent a more viable option for providing input to dynamic analyses, being more realistic than artificial records and easier to obtain than synthetic accelerograms. In any case, current practice for selecting and scaling ground motions for linear and nonlinear response-history analyses is based largely on engineering judgment. Very few systematic studies provide impartial guidance to geotechnical engineers regarding appropriate methods to use in site-specific applications. The majority of the research has been focused in the last years on structural response more than on site response (e.g., NIST 2011). Practitioners often select

ground motion records based only on distance, site soil conditions, and magnitude of the characteristic event expected to dominate the seismic hazard. However, many other factors, such as directivity of the rupture and presence of basin, affect the ground motion intensity and frequency content, which ultimately govern the nonlinear response and damage in geotechnical and structural systems (e.g., Frankel et al. 2011). Despite this, current practice is not adequately equipped to fully incorporate near-fault directivity, basin, and duration effects in the design process. Methods to implicitly consider inelastic demands by amplifying the design spectra do not provide a reliable basis for representing realistic ground input motions (Kalkan and Kunnath 2006).

More research is still needed to address all these issues, requiring a better integration between the work of geotechnical engineers, seismologists, geologists, and structural engineers. Finally, an ongoing transfer of research knowledge into practice represents a significant requisite to support performance-based earthquake engineering design and quantitative seismic risk assessment.

Summary

Experience from past strong earthquakes worldwide has demonstrated the relevance of local soil conditions on seismic ground response. The changes in amplitude, frequency content, and duration of the seismic motion during its propagation in soil deposits, commonly referred to as site effects, have a crucial impact on buildings and infrastructure response during earthquakes. Numerical methods allow geotechnical engineers to quantify the effects of soil deposits on the wave propagation process from bedrock to ground surface. These methods can be divided into: (i) frequency-domain analyses (using the linear equivalent method) and (ii) time-domain schemes (usually performed with finite element codes). The benefits and limitations of the two approaches are extensively discussed in this entry, together with a short description of the available methods for site

characterization and evaluation of soil dynamic properties. An overview of future challenges in the field of geotechnical earthquake engineering is also presented at the end of the entry.

List of symbols

$a_{\max}$	Earthquake maximum acceleration
A_{loop}	Area of the hysteresis loop
D	Damping ratio
f	Frequency
f_l	Fundamental frequency of the soil deposit
f_p	Predominant frequency of the input earthquake motion
f_r	Resonance frequency
G_{sec}	Secant shear modulus
G_{tan}	Tangent shear modulus
G_0 or $G_{\max}$	Initial shear modulus
M	Earthquake magnitude
N	Number of imposed cycles
N_{SPT}	SPT number
OCR	Overconsolidation ratio
PI	Plasticity index
q_c	CPT test tip resistance
V_R	Rayleigh wave velocity
V_S	Shear wave velocity
W_D	Energy dissipated in one hysteresis loop
W_S	Maximum strain energy in one hysteresis loop
α_R	Rayleigh damping coefficient
β	Ratio of effective shear strain to maximum shear strain
β_R	Rayleigh damping coefficient
γ	Shear strain
γ_c	Cyclic shear strain
ρ	Soil density
ω	Angular frequency

References

- Alyami M, Rouainia M, Wilkinson SM (2009) Numerical analysis of deformation behaviour of quay walls under earthquake loading. *Soil Dyn Earthq Eng* 29(3):525–536
- Amorosi A, Boldini D, Elia G (2010) Parametric study on seismic ground response by finite element modelling. *Comput Geotech* 37(4):515–528
- Arulanandan K, Scott RF (eds) (1993) Proceedings of VELACS symposium. A.A., Balkema, Rotterdam
- Ashford SA, Sitar N, Lysmer J, Deng N (1997) Topographic effects on the seismic response of steep slopes. *Bull Seismol Soc Am* 87(3):701–709
- Associazione Geotecnica Italiana (2005) Aspetti Geotecnici della Progettazione in Zona Sismica. Patron Editore, Bologna
- ASTM (2007) Standard test methods for modulus and damping of soils by resonant-column method. ASTM D4015-07. ASTM International, West Conshohocken
- Aubry D, Modaressi A (1996) GEFDYN – Manuel Scientifique. Ecole Centrale Paris, Châtenay-Malabry
- Aydingun O, Adalier K (2003) Numerical analysis of seismically induced liquefaction in earth embankment foundations. Part I. Benchmark model. *Can Geotech J* 40(4):753–765
- Bard PY (1994) Effects of surface geology on ground motion: recent results and remaining issues. In: Proceedings of the X European conference on earthquake engineering, vol 1, Vienna
- Bard PY, Bouchon M (1985) The two-dimensional resonance of sediment-filled valleys. *Bull Seismol Soc Am* 75(2):519–541
- Bardet JP, Tobita T (2001) NERA – a computer program for Nonlinear Earthquake site Response Analyses of layered soil deposits. Department of Civil Engineering, University of Southern California, Los Angeles
- Bardet JP, Ichii K, Lin CH (2000) EERA – a computer program for Equivalent-linear Earthquake site Response Analyses of layered soils deposits. University of Southern California, Department of Civil Engineering, Los Angeles. User manual
- Bazzurro P, Cornell CA (2004) Nonlinear soil-site effects in probabilistic seismic hazard analysis. *Bull Seismol Soc Am* 94:2110–2123
- Biot MA (1941) General theory of three-dimensional consolidation. *J Appl Phys* 12:155–164
- Bommer JJ, Acevedo AB (2004) The use of real earthquake accelerograms as input to dynamic analysis. *J Earthq Eng* 8(4):1–50
- Borja RI, Wu WH (1994) Vibration of foundations on incompressible soils with no elastic region. *J Geotech Eng ASCE* 120(9):1570–1592
- Borja RI, Chao HY, Montáns FJ, Lin CH (1999) Nonlinear ground response at Lotung LSST site. *J Geotech Geoenviron Eng ASCE* 125(3):187–197
- Bouckovalas GD, Papadimitriou AG (2005) Numerical evaluation of slope topography effects on seismic ground motion. *Soil Dyn Earthq Eng* 25:547–558
- Castro G, Christian JT (1976) Shear strength of soils and cyclic loading. *J Geotech Eng Div ASCE* 102(GT9):887–894
- Cavallaro A, Lanzo G, Pagliaroli A, Maugeri M, Lo Presti DCF (2003) A comparative study on shear modulus and damping ratio of cohesive soil from laboratory tests. In: Di Benedetto H et al (eds) Deformation characteristics of geomaterials. Swets & Zeitlinger, Lisse, pp 257–265

- Chan AHC (1995) User manual for DIANA-SWANDYNE II. School of Engineering, University of Birmingham, Birmingham
- Clayton CRI (2011) Stiffness at small strain: research and practice. *Géotechnique* 61(1):5–37
- Dakoulas P, Gazetas G (2005) Seismic effective-stress analysis of caisson quay walls: application to Kobe. *Soils Found* 45(4):133–147
- Dewoolkar MM, Ko H-Y, Pak RYS (2001) Seismic behaviour of cantilever retaining walls with liquefiable backfills. *J Geotech Geoenviron Eng ASCE* 127(5):424–435
- EduPro Civil Systems Inc. (1998) ProShake – ground response analysis program. Redmond, Washington. User's manual
- Elgamal A (1992) Three-dimensional seismic analysis of La Villita dam. *J Geotech Eng ASCE* 118(12):1932–1958
- Elgamal A, Parra E, Yang Z, Adalier K (2002a) Numerical analysis of embankment foundation liquefaction countermeasures. *J Earthq Eng* 6(4):447–471
- Elgamal A, Yang Z, Parra E (2002b) Computational modelling of cyclic mobility and post-liquefaction site response. *Soil Dyn Earthq Eng* 22:259–271
- Elgamal A, Yan L, Yang Z, Conte J (2008) Three-dimensional seismic response of humboldt bay bridge-foundation-ground system. *J Struct Eng ASCE* 134(7):1165–1176
- Elia G, Rouainia M (2013) Seismic performance of earth embankment using simple and advanced numerical approaches. *J Geotech Geoenviron Eng ASCE* 139(7):1115–1129
- Elia G, Rouainia M (2014) Performance evaluation of a shallow foundation built on structured clays under seismic loading. *Bull Earthq Eng* 12(4):1537–1561
- Elia G, Amorosi A, Chan AHC, Kavvas M (2011) Fully coupled dynamic analysis of an earth dam. *Géotechnique* 61(7):549–563
- Field EH, Jacob KH (1993) The theoretical response of sedimentary layers to ambient seismic noise. *Geophys Res Lett* 20(24):2925–2928
- Finn WDL, Yogendrakumar M, Yoshida N, Yoshida H (1986) TARA-3: a program for nonlinear static and dynamic effective stress analysis. Soil Dynamics Group, University of British Columbia, Vancouver
- Frankel A, Stephenson W, Carver D, Odum J, Williams R, Rhea S (2011) Probabilistic seismic hazard maps for seattle: 3D sedimentary basin effects, nonlinear site response, and uncertainties from random velocity variation. In: *Proceedings of the 4th IASPEI/IAEE international symposium: effects of surface geology on seismic motion*, University of California, Santa Barbara
- Geli L, Bard PY, Jullien B (1988) The effect of topography on earthquake ground motion: a review and new results. *Bull Seismol Soc Am* 78(1):42–63
- GeoMotions (2007) D-MOD2000 – a computer program package for seismic response analysis of horizontally layered soil deposits, earthfill dams, and solid waste landfills. GeoMotions LLC, Washington. User's manual
- GeoSlope Int. Ltd. (2002) QUAKE/W for finite element dynamic earthquake analysis. User's Guide, Calgary
- Hall L, Bodare A (2000) Analyses of the cross-hole method for determining shear wave velocities and damping ratios. *Soil Dyn Earthq Eng* 20:167–175
- Hardin B, Drnevich VP (1972) Shear modulus and damping in soils: measurements and parameter effects. *J Soil Mech Div ASCE* 98:603–624
- Hashash YMA (2009) DEEPSOIL V 3.7, Tutorial and user manual. 2002–2009. University of Illinois at Urbana-Champaign, Urbana
- Hatzigeorgiou GD, Beskos DE (2010) Soil-structure interaction effects on seismic inelastic analysis of 3-D tunnels. *Soil Dyn Earthq Eng* 30:851–861
- Hudson M, Idriss IM, Beikae M (1994) QUAD4M: a computer program to evaluate the seismic response of soil structures using finite element procedures and incorporating a compliant base. Center for Geotechnical Modeling, University of California, Davis
- Idriss IM (1990) Influence of local site conditions on earthquake ground motions. In: *Proceedings of IV U.S. national conference on earthquake engineering*, vol 1, Palm Springs
- Idriss IM, Sun JI (1992) SHAKE91: a computer program for conducting equivalent linear seismic response analyses of horizontally layered soils deposits. Center for Geotechnical Modeling, University of California, Davis
- Idriss IM, Lysmer J, Hwang R, Seed HB (1973) QUAD-4: a computer program for evaluating the seismic response of soil structures by variable damping finite element procedures. Report No EERC 73-16, Earthquake Engineering Research Center, University of California, Berkeley
- Ishibashi I (1992) Discussion to “effect of soil plasticity on cyclic response”, by M. Vucetic and R. Dobry. *J Geotech Eng ASCE* 118(5):830–832
- Itasca Consulting Group Inc. (2002) FLAC – Fast Lagrangian Analysis of Continua. Minneapolis. User's manual
- Kalkan E, Kunnath SK (2006) Effects of fling step and forward directivity on seismic response of buildings. *Earthq Spectra* 22(2):367–390
- King JL, Tucker BE (1984) Observed variations of earthquake motion across a sediment-filled valley. *Bull Seismol Soc Am* 74(1):137–151
- Kontoe S, Zdravkovic L, Potts DM, Menkiti CO (2011) On the relative merits of simple and advanced constitutive models in dynamic analysis of tunnels. *Géotechnique* 61:815–829
- Kramer SL (1996) *Geotechnical earthquake engineering*. Prentice Hall, Upper Saddle River, New Jersey
- Kramer SL, Stewart JP (2004) Geotechnical aspects of seismic hazards. In: Bozorgnia Y, Bertero VV (eds) *Earthquake engineering – from engineering seismology to performance-based engineering*. CRC Press LLC, Boca Raton

- Kwok AOL, Stewart JP, Hashash YMA, Matasovic N, Pyke R, Wang Z, Yang Z (2007) Use of exact solutions of wave propagation problems to guide implementation of nonlinear seismic ground response analysis procedures. *J Geotech Geoenviron Eng ASCE* 133(11):1385–1398
- Lanzo G, Silvestri F (1999) *Risposta Sismica Locale (Teoria ed esperienze)*. Hevelius Edizioni srl, Benevento
- Lanzo G, Silvestri F, Costanzo A, d'Onofrio A, Martelli L, Pagliaroli A, Sica S, Simonelli A (2011) Site response studies and seismic microzoning in the Middle Aterno valley (L'Aquila, Central Italy). *Bull Earthq Eng* 9(5):1417–1442
- Lee MKW, Finn WDL (1978) DESRA-2, dynamic effective stress response analysis of soil deposits with energy transmitting boundary including assessment of liquefaction potential, Soil mechanics series no 38. Department of Civil Engineering, University of British Columbia, Vancouver
- Li XS, Wang ZL, Shen CK (1992) SUMDES: a nonlinear procedure for response analysis of horizontally layered sites subjected to multi-directional earthquake loading. Department of Civil Engineering, University of California, Davis
- Liu H, Song E (2005) Seismic response of large underground structures in liquefiable soils subjected to horizontal and vertical earthquake excitations. *Comput Geotech* 32(4):223–244
- Lo Presti DCF, Jamiolkowski M, Pallara O, Cavallaro A, Pedroni S (1997) Shear modulus and damping of soils. *Géotechnique* 47(3):603–617
- Lo Presti DFC, Pallara O, Mensi E (2007) Characterization of soil deposits for seismic response analysis. In: Ling HI et al (eds) *Soil stress-strain behavior: measurement, modeling and analysis*. Springer, Dordrecht, The Netherlands, pp 109–157
- Lysmer J, Kuhlemeyer RL (1969) Finite dynamic model for infinite media. *J Eng Mech Div ASCE* 95(EM4):859–877
- Lysmer J, Udaka T, Tsai C-F, Seed HB (1975) FLUSH: a computer program for approximate 3-D analysis of soil-structure interaction problems. Report EERC 75-30, University of California, Berkeley
- Madabhushi SPG, Zeng X (2007) Simulating seismic response of cantilever retaining walls. *J Geotech Geoenviron Eng ASCE* 133(5):539–549
- Masing G (1926) *Eigenstressungen und verfestigung beim messing*. In: *Second international congress on applied mechanics*, Zurich, pp 332–335
- Matasovic N (1995) D-MOD_2 A computer program for seismic response analyses of horizontally layered soil deposits, earthfill dams and solid waste landfills. GeoSyntec Consultants, Huntington Beach
- McKenna F, Fenves GL (2001) *The OpenSees command language manual*, version 1.2. Pacific Earthquake Engineering. Research Center, University of California, Berkeley
- Meng J, Rix GJ (2003) Reduction of equipment-generated damping in resonant column measurements. *Géotechnique* 53(5):503–512
- NIST (2011) Selecting and scaling earthquake ground motions for performing response-history analyses. NIST GCR 11-917-15, National Institute of Standards and Technology, Gaithersburg
- Ou JH (2009) Three-dimensional numerical modelling of interaction between soil and pore fluid. PhD thesis, School of Engineering, University of Birmingham
- Pagliaroli A, Lanzo G, D'Elia B (2011) Numerical evaluation of topographic effects at the Nicastro ridge in Southern Italy. *J Earthq Eng* 15(3):404–432
- Papaspiliou M, Kontoe S, Bommer JJ (2012) An exploration of incorporating site response into PSHA-part II: sensitivity of hazard estimates to site response approaches. *Soil Dyn Earthq Eng* 42:316–330
- Phillips C, Hashash YMA (2009) Damping formulation for nonlinear 1D site response analyses. *Soil Dyn Earthq Eng* 29:1143–1158
- Prevost JH (2002) DYNFLOW – a nonlinear transient finite element analysis program. Version 2002. Release 01.A, Department of Civil Engineering & Operation Research, Princeton University, Princeton
- Pyke RM (1992) TESS: a computer program for nonlinear ground response analyses. TAGA Engineering Systems and Software, Lafayette
- Richart FE, Hall JR, Woods RD (1970) *Vibrations of soils and foundations*. Prentice-Hall, Englewood Cliffs
- Roesset JM (1977) Soil amplification of earthquakes. In: Desai CS, Christian JT (eds) *Numerical methods in geotechnical engineering*. McGraw-Hill, New York, pp 639–682
- Sangrey DA, Henkel DJ, Esrig MI (1969) The effective stress response of a saturated clay soil to repeated loading. *Can Geotech J* 6(3):241–252
- Schnabel PB, Lysmer J, Seed HB (1972) SHAKE: a computer program for earthquake response analysis of horizontally layered sites. Report no EERC72-12, Earthquake Engineering Research Center, University of California, Berkeley
- Seed HB, Idriss IM (1970) Soil moduli and damping factors for dynamic response analyses. Report EERC 70-10, Earthquake Engineering Research Center, University California, Berkeley
- Seed HB, Ugas C, Lysmer J (1976) Site dependent spectra for earthquake resistant design. *Bull Seismol Soc Am* 66(1):221–243
- Seidalinov G, Taiebat M (2014) Bounding surface SANICLAY plasticity model for cyclic clay behavior. *Int J Numer Anal Methods Geomech* 38(7):702–724
- Semblat JF, Pecker A (2009) *Waves and vibrations in soils: earthquakes, traffic, shocks, construction works*. IUSS Press, Pavia
- Semblat JF, Kham M, Parara E, Bard PY, Pitilakis K, Makra K, Raptakis D (2005) Seismic wave amplification: basin geometry vs soil layering. *Soil Dyn Earthq Eng* 25:529–538

- SESAME (2004) Guidelines for the implementation of the H/V spectral ratio technique on ambient vibrations measurements, processing and interpretation. SES-AME European Research Project. WP12 – Deliverable D23.12
- Shahrour I, Khoshnoudian F, Sadek M, Mroueh H (2010) Elastoplastic analysis of the seismic response of tunnels in soft soils. *Tunn Undergr Space Technol* 25: 478–482
- Sica S, Pagano L, Modaressi A (2008) Influence of past loading history on the seismic response of earth dams. *Comput Geotech* 35(1):61–85
- Stamos AA, Beskos DE (1995) Dynamic analysis of large 3-D underground structures by the BEM. *Earthq Eng Struct Dyn* 24(6):917–934
- Tsai C-CP (2000) Probabilistic seismic hazard analysis considering nonlinear site effect. *Bull Seismol Soc Am* 90:66–72
- Vinale F, Mancuso C, Silvestri F (1996) *Dinamica dei terreni. Manuale di Ingegneria Civile Cremonese*. Zanichelli/ESAC, Roma
- Vucetic M (1986) Pore pressure buildup and liquefaction at level sand sites during earthquakes. Rensselaer Polytechnic Institute, Troy
- Vucetic M, Dobry R (1991) Effects of the soil plasticity on cyclic response. *J Geotech Eng Div ASCE* 117(1): 89–107
- Woods RD (1994) Geophysical characterization of sites. Balkema, Rotterdam
- Woodward PK, Griffiths DV (1996) Influence of viscous damping in the dynamic analysis of an earth dam using simple constitutive models. *Comput Geotech* 19(3): 245–263
- Zienkiewicz OC, Chan AHC, Pastor M, Schrefler BA, Shiomi T (1999) *Computational geomechanics (with special reference to earthquake engineering)*. Wiley, Chichester

Site Response: 1-D Time Domain Analyses

Neven Matasovic
Geosyntec Consultants, Huntington Beach, CA,
USA

Synonyms

1-D nonlinear seismic site response analysis; 1-D nonlinear site response analysis; One-dimensional (1-D) time domain analysis

Introduction

One-dimensional (1-D) time domain analysis is one of several currently available methods to evaluate the influence of local site conditions on input (i.e., bedrock) ground motions. However, this method is favored by Geotechnical Earthquake Engineers as it takes into account the unique geotechnical characteristic of a given site, including soil nonlinearity and hysteretic behavior and porewater pressure generation and dissipation. It can also be applied for a wide range of shaking intensities, and it is relatively easy to implement in practice as it is coded in commercially available software.

Through back analysis of numerous case histories, the 1-D time domain analysis has been shown to work well when analyzed soil deposits are horizontally layered, there is a significant impedance contrast within the profile, and when material (model) parameters are established through a reasonable site investigation and characterization efforts.

There are many circumstances under which a 1-D time domain site response analysis is performed. Such an analysis is commonly performed in the following situations: (i) when soil conditions cannot be reasonably categorized into one of the standard site conditions, (ii) when empirical site factors for the site are not available (e.g., the International Building Code Site Class F), (iii) when special ground conditions govern the design (e.g., soil liquefaction, seismic settlement and/or lateral spreading, or slope stability), (iv) for any case where the objective is to obtain ground motions considered to be more representative of the local geologic and seismic site conditions than motions obtained from other approaches, and/or (v) when calibration of a more advanced, multidimensional models with or without structural elements is required.

In general, site response analysis methods can be classified based upon the domain in which calculations are performed (frequency domain or time domain), the sophistication of the constitutive model employed (linear, equivalent linear, and/or nonlinear), whether the effects of

porewater pressure generation are neglected or not (total-stress and effective-stress analyses, respectively), and the dimensionality of the space in which the analysis is performed (one-dimensional, 1-D; quasi 2-D, 2-D, and 3-D). Other considerations in classifying site response analysis methods include modeling of cyclic reduction and cyclic degradation in a total-stress mode. This chapter focuses upon time domain nonlinear total- and effective-stress 1-D site response analysis which has significant advantages over its frequency domain linear and equivalent-linear counterparts in areas of very high seismicity (e.g., $\text{PGA} \geq 0.4 \text{ g}$) and/or when the effects of soft and/or potentially liquefiable soils in the profile cannot be ignored.

Theoretical Background

Dynamic Response Model

In 1-D time domain analysis, seismic response of a horizontally layered soil deposit is computed by solving the dynamic equation of motion. The dynamic equation of motion is commonly written as

$$[M] \{\ddot{u}\} + [C] \{\dot{u}\} + [K] \{u\} = -[M] \{\ddot{u}_g\} \quad (1)$$

where $[M]$, $[C]$, and $[K]$ are the mass matrix, viscous damping matrix, and nonlinear stiffness matrix, respectively; $\{u\}$, $\{\dot{u}\}$, and $\{\ddot{u}\}$ are the displacements, velocities, and accelerations of the mass $[M]$ relative to the base; and $\{\ddot{u}_g\}$ is the acceleration of the base.

The stiffness matrix $[K]$ is derived from the nonlinear (and effective-stress) soil constitutive model selected to represent a cyclic response of soil. In principle, all damping in the soil can be captured through the hysteretic loops of the constitutive model used to model cyclic soil behavior. However, most available soil constitutive models cannot properly represent soil damping at low strains. Therefore, it is necessary to add damping through the use of velocity proportional viscous damping, as discussed in detail below.

The dynamic equation of motion can be solved by numerical integration. The numerical integration calls for temporal discretization (i.e., system of coupled equations is discretized temporally), hence the term “solution in time domain.” The dynamic equation is solved by a time-stepping scheme. Examples of time-stepping schemes include Wilson’s theta (θ) algorithm (Clough and Penzien 1993) and several variations of the Newmark’s β algorithms (Newmark 1959).

In order to solve the dynamic equation of motion, it is necessary to properly discretize the domain of interest, which in this case is the 1-D soil column. Two different approaches for discretization of the soil column are available: (i) lumped mass discretization and (ii) finite element discretization.

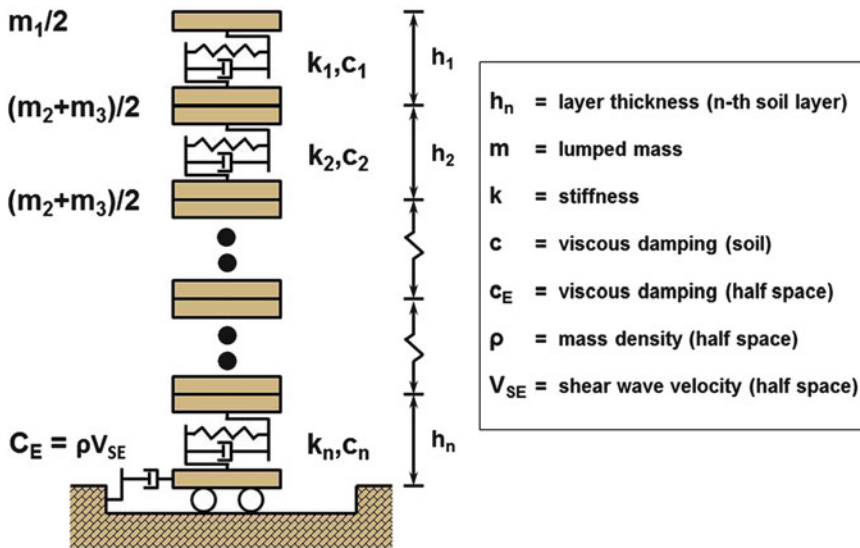
Figure 1 shows a lumped mass discretization (i.e., lumped mass model) that of a horizontally layered soil deposit. Soil mass (m) is lumped at the layer interfaces (layers 1 through n). In each layer, soil stiffness is represented by a nonlinear spring. In addition to hysteretic damping inherent to a nonlinear spring, a viscous damping coefficient (c) is also part of a model shown in Fig. 1, as schematically represented by dashpots.

The finite element discretization requires solving the dynamic equation of motion (Eq. 1) by means of an explicit time marching integration algorithm. In this type of discretization, mass can be distributed over layer thickness.

The thickness of the layers in both lumped mass and finite element discretization have an impact on calculated site response. The layer thickness determines the maximum and minimum frequency that can be propagated through a soil column. If the layer is too thick, important components of the input ground motion may be filtered, and thus the calculated ground response may be underestimated. If layer thickness is too small, representation of the actual site conditions might be inadequate (e.g., soil lens is modeled as a soil layer).

Nonlinear Constitutive Models with Hysteretic Damping

One-dimensional nonlinear total-stress site response analysis is generally conducted using



Site Response: 1-D Time Domain Analyses, Fig. 1 Lumped mass discretization for 1-D dynamic response model (Matasovic 1993)

relatively simplified constitutive models of soil layers. These models evolved from the early stress–strain relationships of Ramberg and Osgood (1943) and Kondner and Zelasko (1963). The hyperbolic model, introduced by Duncan and Chang (1970) for axial soil behavior, was based upon the basic formulation of the above-cited shear stress and strain behavior models. It was accompanied by sets of generic material properties and hence allowed for an elegant and simple way to capture soil nonlinearity at small axial strains. All three models provided the basis for constitutive models that are presently in use. These include models by Pyke (1979), Matasovic (1993), Matasovic and Vucetic (1993), and Darendeli (2001) which, in addition to better simulation of nonlinear stress–strain behavior, allow for simulation of cyclic loading and reloading in accordance with certain rules. The stress–strain relationship in these models is generally established by: an initial loading curve; a series of rules that describe the backbone curve (see Fig. 2 for definition of backbone curve); and unloading–reloading behavior rules required to establish cyclic loops. The most widely used rules are the Masing rules (Masing 1926) and extended Masing rules

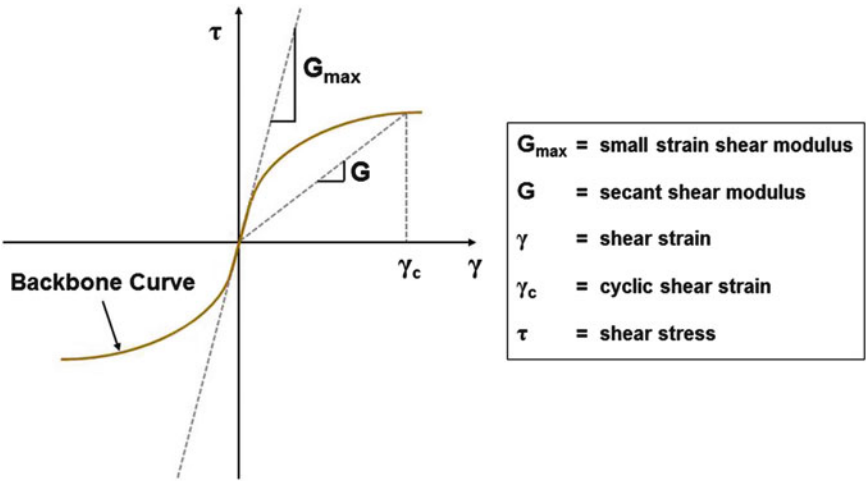
(Pyke 1979; Vucetic 1990). Recently, Phillips and Hashash (2009) developed a set of rules that compensate for one of the most common deficiency of nonlinear models with hysteretic damping – overestimation of damping at large shear strains.

Bounding surface plasticity models may be used to simulate nonlinear stress–strain behavior in one dimension. Example of such models and their application in site response analysis can be found in Borja et al. (2002).

Viscous Damping Models

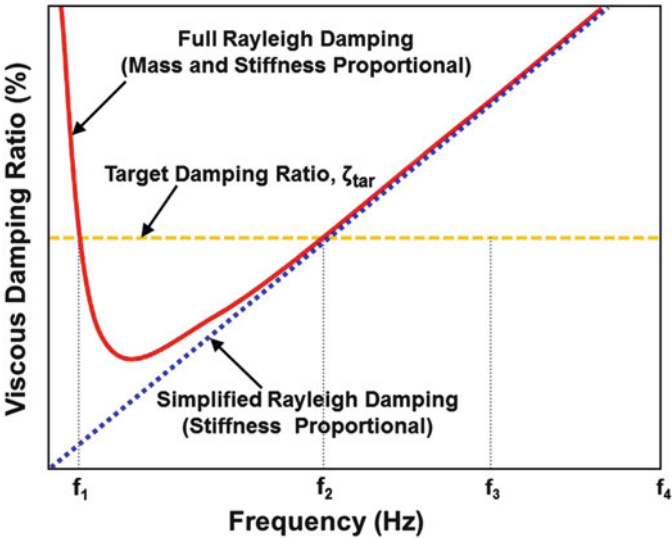
The viscous damping models are an essential part of nonlinear site response programs. This is because: (i) viscous damping is a necessary input into a linear-elastic solution that is typically an integral part of a time domain site response analysis program; (ii) viscous damping readily increases numerical stability of several numerical schemes used to solve Eq. 1; and (iii) damping calculated by a hyperbolic model at very low strains (strains less than 0.1 %) is very small compared to the values obtained by laboratory measurements. Viscous damping compensates for that difference.

Given the above, most engineers introduce a small amount of viscous damping into site



Site Response: 1-D Time Domain Analyses, Fig. 2 Backbone curve as stress–shear strain relationship for monotonic loading

Site Response: 1-D Time Domain Analyses, Fig. 3 Schematic illustration of viscous damping change with frequency



response analysis. The value of viscous damping introduced is typically small (0.5–5 %).

The most commonly used formulation for evaluation of the viscous damping coefficient (c) is the Rayleigh damping (model). The viscous damping coefficient can be evaluated as

$$c = \alpha_R \mathbf{m} + \beta_R \mathbf{k} \tag{2}$$

where α_R and β_R are the Rayleigh damping coefficients (Rayleigh and Lindsay 1945)

and $\mathbf{m}$ and $\mathbf{k}$ are elements of the mass and stiffness matrices, respectively.

Figure 3 illustrates how Rayleigh damping, expressed through c , changes with frequency. The viscous damping ratio can be brought closer to a constant value of the target damping ratio (ζ_{tar}) by specifying c at only one frequency (e.g., at f_2 in Fig. 3), which is termed the *simplified* Rayleigh damping formulation, and/or at two frequencies (at f_1 and f_2), which is termed the *full* Rayleigh damping formulation.

Kwok et al. (2007) recommended, for most practical applications, use of the *full* Rayleigh damping formulation in nonlinear (total-stress) site response analysis whereby the first frequency is equal to the fundamental frequency of the soil column, while the second frequency is equal to five times the fundamental frequency.

Time Domain Analysis with Porewater Pressure Change

General

The cyclic loading of saturated soils is accompanied by porewater pressure change. This change includes simultaneous porewater pressure generation and dissipation. If the generated porewater pressures are sufficiently large, the soil stiffness and strength are significantly reduced, and ultimately, in some soils, liquefaction can occur.

In nonlinear site response analysis with porewater pressure generation (i.e., in effective-stress analysis), the response of the soil to cyclic loading accounts for the generation of excess porewater pressure (PWP) during cyclic shearing of the soil as well as dissipation of these excess porewater pressures during and after the cyclic loading. The representation of dissipation/redistribution of porewater pressure influences soil stiffness (modulus) and strength (shear stress) during shaking, thus resulting in a more realistic simulation of site response. The PWP dissipation/redistribution is discussed later.

The influence of PWP changes during cyclic loading is incorporated in soil constitutive models in two ways: (i) semiempirical PWP generation models used in combination with total-stress soil models and (ii) advanced effective-stress models whereby the model formulation is in terms of effective stress, and the PWP change is computed as the change between total stresses (or loads) and effective stresses via the soil constitutive model.

Semiempirical Porewater Pressure Generation Models

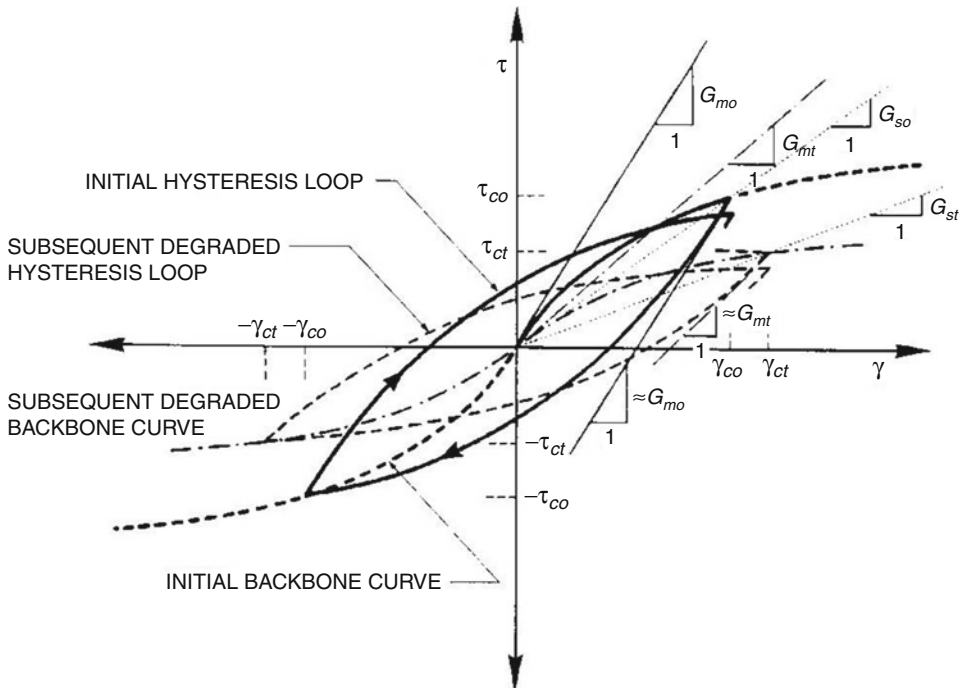
In this class of porewater pressure generation models, PWP generation in a soil profile is

computed in a semiempirical way. At the beginning of shaking (i.e., at time $t = 0$), stress-strain relationships of the soil are identical to that of the total-stress models since PWP is zero. As shaking progresses, PWP is generated and cyclic degradation (of clay microstructure) starts. Subsequently, the effects of porewater pressure generation and, in some models, of cyclic degradation are included by degradation of soil strength and stiffness. Different models for degradation of soil strength and stiffness are used in some of the models. For example, Matasovic (1993) and Matasovic and Vucetic (1995b) proposed degradation index functions that degrade strength and stiffness of sands at different rates, while the concept of the degradation index (Idriss et al. 1978) is generally used to degrade strength and stiffness of soft clays.

A number of PWP generation models have been developed starting with Martin and Seed (1978) and Martin et al. (1975). A more recent example of the semiempirical porewater pressure model for saturated sand is Dobry et al. (1985) model. This model was developed based upon strain-controlled cyclic direct simple shear and cyclic triaxial testing. The model was later modified by Vucetic (1986) to allow for quasi-two-dimensional shaking and further by Matasovic (1993) to more accurately model porewater pressure-induced degradation of shear modulus and shear stress. The porewater pressure generation models described thus require the use of an equivalent number of cycles to represent earthquake shaking. Polito et al. (2008) introduced an energy-based model (GMP model) for the generation of porewater pressure based on a large number of laboratory tests, which does not require the development of an equivalent number of cycles.

With exception of the modified Dobry et al. (1985) model, there is limited information available to guide the user in selecting the appropriate PWP model parameters.

The effect of cyclic degradation on soil stiffness and strength is illustrated in Fig. 4 for the MKZ constitutive model (Matasovic 1993; Matasovic and Vucetic 1993). The initial hysteresis loop shown in the figure refers to the first cycle of cyclic loading (i.e., at time $t = 0$).



Site Response: 1-D Time Domain Analyses, Fig. 4 Stress-strain behavior modeling illustrating stiffness degradation due to porewater pressure buildup (Matasovic 1993)

The subsequent degraded hysteretic loop refers to any subsequent cycle (i.e., at time t) for which enough porewater pressure has built up to degrade both initial shear modulus G_{mo} and initial shear stress τ_{co} at corresponding shear strain γ_{co} .

An example of a porewater pressure model for clay is the Matasovic and Vucetic (1995a) model. This model was developed based upon the results of cyclic simple shear testing. It should be noted that porewater pressure in clay is of much lower intensity than in sand and that in overconsolidated clay, both positive and negative (suction) porewater pressures may develop (e.g., Matasovic and Vucetic 1992).

Advanced Effective-Stress-Based Models

Another class of soil constitutive models used in site response analysis is effective-stress models. In these models, the formulation of the

constitutive law is developed in effective-stress space, and porewater pressures are computed as the difference between effective stresses and total stresses in the domain of interest. Examples of plasticity-based constitutive models include Roscoe and Schofield (1963), Pestana (1994), and Elgamal et al. (2001). These advanced constitutive models are capable of simulating complex soil behavior under a variety of loading conditions. Key elements of these models include yield surfaces, flow rules, and hardening (or softening) laws. A review of advanced constitutive models with application in site response analysis is provided in Potts and Zdravković (1999).

Generic material parameters for advanced constitutive models are often not available. Evaluation of material parameters for these models requires significant expertise and detailed site-specific soil properties.

Porewater Pressure Dissipation and Redistribution Models

The layers in a soil column have finite, saturated hydraulic conductivities. During ground shaking, even though loading is relatively rapid, it is possible that porewater dissipation and redistribution can occur due to differences in hydraulic gradients and hydraulic conductivities between layers.

An early model for porewater pressure dissipation and redistribution was introduced by Martin and Seed (1978). Input parameters include (effective-stress-dependent) constrained rebound modulus and (saturated) hydraulic conductivity. A porewater pressure dissipation model for composite soil deposits (alternating sand and clay layers) was developed by Matasovic and Vucetic (1995a).

Porewater pressure redistribution occurs and is often modeled in accordance with the principles of Terzaghi's theory of consolidation.

Time Domain Analysis Computer Programs

There are about 25 time domain analysis computer programs that are available globally, either commercially or through direct contact with the author(s). However, only a handful of these programs have a provision to perform an effective-stress analysis, including 2-D and 3-D programs which can be used for analysis of 1-D geometry.

Kwok et al. (2007) performed a comprehensive survey of available time domain analysis computer programs in the United States and found that only five 1-D nonlinear programs were used at the time, including OpenSees (Ragheb 1994) which is primarily used for analysis of 2-D and 3-D problems. Other programs identified by Kwok et al. (2007) include DEEPSOIL (Hashash and Park 2001), D-MOD_2 (Matasovic 2006; subsequently updated with graphical user interface as D-MOD2000), SUMDES (Wang 1990), and

TESS (Pyke 2000), with last three programs also having a provision for effective-stress analysis. The study compared the programs in total-stress mode and found that, when input is properly controlled, most of the software provided similar results.

Matasovic and Hashash (2012) performed a survey of the United States Department of Transportation (DOT) practices. Part of the survey was an inquiry on what site response analysis programs are used, both by DOT-s and their consultants. The survey revealed that the most used program was D-MOD2000 (Matasovic and Ordonez 2007) followed by DEEPSOIL (Hashash et al. 2011). However, D-MOD2000 was the only program cited to have been used in effective-stress mode. Coincidentally, these two programs are the only time domain analysis programs available in United States that have a graphical user interface and the effective-stress analysis provision.

Calibration and Benchmarking Studies

A number of individuals and groups have conducted back-analysis exercises to evaluate site response analysis procedures (and programs) versus measured response to strong ground shaking. Most of these studies include use of recordings from downhole vertical arrays.

Researchers have used a range of inverse analysis techniques that include ad hoc system identification (e.g., Zeghal and Elgamal 1993; Assimaki and Steidl 2007) and evolutionary soil models (e.g., Tsai and Hashash 2007) to calibrate soil constitutive models in site response analysis actual (in situ) soil behavior.

A number of benchmarking exercises have also been conducted to evaluate site response analysis tools. A recent key exercise is the PEER benchmarking exercise for total-stress site response analyses (Kwok et al. 2007). Another interesting benchmarking exercise was the evaluation of site response at Turkey Flat (e.g., Kramer 2009). The interesting outcome of

this exercise is in highlighting the challenges encountered in computing the response at a well-constrained relatively stiff soil site.

Summary

One-dimensional time domain analysis is a powerful tool for evaluation of site response of horizontally layered soil deposits. Compared to other types of 1-D analysis, it has a provision to simulate cyclic soil behavior in a more realistic manner than its frequency domain counterparts, and it can accommodate models for seismically induced porewater pressure buildup. Compared to its 2-D and 3-D counterparts, the advantage of this type of analysis lies in simple constitutive models for which material parameters can be readily evaluated, or generic sets of material parameters are available to practicing engineers. Furthermore, compared to its 2-D and 3-D counterparts, 1-D analysis is far better calibrated and validated. Consequently, consistent with current trends, it is anticipated that the role of 1-D analysis will evolve into calibration of 2-D and 3-D models.

One-dimensional time domain analysis is not perfect. Further calibration and validation of 1-D models are required both in effective-stress mode and at high levels of shaking (>0.4 g). There are also several issues related to the application of this type of analysis in engineering practice and research, as summarized by Hashash et al. (2010). Future trends will certainly address these issues, yet also include improvement of graphical user interfaces that are essential not only for effective use of the programs, but for evaluation and validation of the results.

Cross-References

- [Selection of Ground Motions for Response History Analysis](#)
- [Site Response: Comparison Between Theory and Observation](#)

References

- Assimaki D, Steidl J (2007) Inverse analysis of weak and strong motion downhole array data from the Mw 7.0 Sanriku-Minami earthquake. *Soil Dyn Earthq Eng* 27:73–92
- Borja RI, Duvernay BG, Lin CH (2002) Ground response in Lotung: total stress analyses and parametric studies. *J Geotech GeoEnviron Eng* 128(1):54–63
- Clough RW, Penzien J (1993) *Dynamics of structures*, 2nd edn. McGraw Hill, London
- Darendeli MB (2001) Development of a new family of normalized modulus reduction and material damping curves. PhD thesis, University of Texas at Austin, Austin
- Dobry R, Pierce WG, Dyvik R, Thomas GE, Ladd RS (1985) Pore pressure model for cyclic straining of sand. Research Report, Civil Engineering Department, Rensselaer Polytechnic Institute, Troy, 56 p
- Duncan JM, Chang C-Y (1970) Nonlinear analysis of stress and strain in soils. *J Soil Mech Found Div* 96(SM5):1629–1653
- Elgamal A, Lai T, Yang Z, He L (2001) Dynamic soil properties, seismic downhole arrays and applications in practice. In: *Proceedings of the 4th international conference on recent advances in geotechnical earthquake engineering and soil dynamics*, San Diego
- Hashash YMA, Park D (2001) Nonlinear one-dimensional seismic ground motion propagation in the Mississippi embayment. *Eng Geol (Amst)* 62(1–3):185–206
- Hashash YMA, Phillips C, Groholski D (2010) Recent advances in non-linear site response analysis. In: *Proceedings of the 5th international conference on recent advances in geotechnical earthquake engineering and soil dynamics*, San Diego
- Hashash YMA, Groholski DR, Phillips CA, Park D, Musgrove M (2011) DEEPSOIL 4.0, User Manual and Tutorial. 98 p
- Idriss IM, Dobry R, Singh RD (1978) Nonlinear behavior of soft clays during cyclic loading. *J Geotech Eng Div ASCE* 104(12):1427–1447
- Kondner RL, Zelasko JS (1963) A hyperbolic stress–strain formulation of sands. In: *Proceedings of the 2nd Pan American conference on soil mechanics and foundation engineering*, Sao Paulo, pp 289–324
- Kramer SL (2009) Analysis of Turkey flat ground motion prediction experiment – lessons learned and implications for practice. In: *SMIP09 seminar on utilization of strong-motion data*, San Francisco, pp 1–22
- Kwok O-LA, Stewart JP, Hashash YMA, Matasovic N, Pyke R, Wang Z, Yang Z (2007) Use of exact solutions of wave propagation problems to guide implementation of nonlinear ground response analysis procedures. *ASCE J Geotechn Geoenviron Eng* 133(11):1385–1398
- Martin PP, Seed HB (1978) APOLLO, a computer program for the analysis of pore pressure generation and dissipation in horizontal sand layers during cyclic

- earthquake loading. Report No. EERC 78–21, Earthquake Engineering Research Center, University of California, Berkeley
- Martin GR, Finn WDL, Seed HB (1975) Fundamentals of liquefaction under cyclic loading. *J Geotech Eng Div ASCE* 101(GT5):423–438
- Masing G (1926) Eigenspannungen und Verfestigung beim Messing. In: *Proceedings of the 2nd international congress on applied mechanics, Zürich*, pp 332–335
- Matasovic N (1993) Seismic response of composite horizontally layered soil deposits. PhD dissertation, Civil and Environmental Engineering Department, University of California, Los Angeles, 452 p
- Matasovic N (2006) D-MOD_2 – a computer program for seismic response analysis of horizontally layered soil deposits, earthfill dams, and solid waste landfills. User's Manual, GeoMotions, LLC, Lacey, 20 p (plus Appendices)
- Matasovic N, Hashash YMA (2012) Site response analysis in transportation engineering practice – a TRB Survey. In: *Proceedings of the GeoCongress 2012, Oakland, CD-ROM paper*, pp 1789–1798
- Matasovic N, Ordonez GA (2007) D-MOD2000 – a computer program package for seismic response analysis of horizontally layered soil deposits, earthfill dams, and solid waste landfills. User's Manual, GeoMotions, LLC, Lacey, Washington, 182 p. <http://www.geomotions.com>
- Matasovic N, Vucetic M (1992) A pore pressure model for cyclic straining of clay. *Soils Found J JSSMFE* 32(3):156–173
- Matasovic N, Vucetic M (1993) Cyclic characterization of liquefiable sands. *ASCE J Geotech Eng* 119(11):1805–1822
- Matasovic N, Vucetic M (1995a) Generalized cyclic degradation-pore pressure generation model for clays. *ASCE J Geotech Eng* 121(1):33–42
- Matasovic N, Vucetic M (1995b) Seismic response of soil deposits composed of fully saturated clay and sand layers. In: Kenji Ishihara AA (ed) *Proceedings of the IS-Tokyo'95, the first international conference on earthquake geotechnical engineering*, Balkema, Rotterdam
- Newmark NM (1959) A method of computation for structural dynamics. *J Eng Mech Div* 85:67–94
- Pestana JM (1994) A unified constitutive model for clays and sands. ScD thesis, MIT, Cambridge, MA
- Phillips C, Hashash YMA (2009) Damping formulation for nonlinear 1D site response analyses. *Soil Dyn Earthq Eng* 29(7):1143–1158
- Polito CP, Green RA, Lee JH (2008) Pore pressure generation models for sands and silty soils subjected to cyclic loading. *J Geotech Geoenviron* 134(10):1490–1500
- Potts DM, Zdravković L (1999) *Finite element analysis in geotechnical engineering: theory*. Thomas Telford, London
- Pyke RM (1979) Nonlinear soil models for irregular cyclic loadings. *J Geotech Eng Div, ASCE* 105(GT6):715–726
- Pyke RM (2000) TESS: a computer program for nonlinear ground response analyses. TAGA Engineering Systems & Software, Lafayette, <http://www.tagasoft.com>
- Ragheb AM (1994) Numerical analysis of seismically induced deformations in saturated granular soil strata. PhD dissertation, Department of Civil Engineering, Rensselaer Polytechnic Institute, Troy
- Ramberg W, Osgood WR (1943) Description of stress-strain curves by three parameters. Technical Note 902, National Advisory Committee for Aeronautics, Washington, DC
- Rayleigh JWS, Lindsay RB (1945) *The theory of sound*. Dover, New York
- Roscoe KH, Schofield AN (1963) Mechanical behavior of an idealized 'wet' clay. In: *Proceedings of the 2nd European conference on soil mechanics*, vol 1, Wiesbaden, pp 47–54
- Tsai C-C, Hashash YMA (2007) A novel framework integrating downhole array data and site response analysis to extract dynamic soil behavior. *Soil Dyn Earthq Eng* 28(3):181–197
- Vucetic M (1986) Pore pressure buildup and liquefaction of level sandy sites during earthquakes. PhD dissertation, Rensselaer Polytechnic Institute, Troy, 616 p
- Vucetic M (1990) Normalized behavior of clay under irregular cyclic loading. *Can Geotech J* 27:29–46
- Wang ZL (1990) Bounding surface hypoplasticity model for granular soils and its applications. PhD dissertation, University of California at Davis
- Zeghal M, Elgamal AW (1993) Lotung sites: downhole seismic data analysis. Electric Power Research Institute, Palo Alto

Site Response: Comparison Between Theory and Observation

Takaji Kokusho
Department of Civil & Environmental
Engineering, Chuo University, Tokyo, Japan

Synonyms

1/4 wavelength formula; Two-layer system; Downhole array; Multilayer system; Multi-reflection theory; Nonlinear soil properties; Seismic site amplification; Soil damping model; Surface array; S-wave velocity

Introduction

Kanai et al. (1959) discovered from earthquake observations that a horizontal component in site response can be explained by multi-reflection of SH-waves propagating vertically in soil layers. Horizontal displacement of the vertically propagating SH-wave is expressed as

$$u = Af(z - V_s t) + Bf(z + V_s t) \quad (1)$$

where z = vertical coordinate upward positive; t = time; V_s = S-wave velocity; $f()$, $g()$ = an arbitrary functions; and A , B = amplitudes of upward and downward waves, respectively. Site response governed by this equation is largely dependent on the composition of soil layers, S-wave velocities, and soil damping ratios of those layers and their strain-dependent variations. In the following, site response observations in array systems are compared with the SH-wave multi-reflection theory to clarify its applicability in terms of dominant frequencies and spectrum amplifications with special emphasis on soil damping mechanism and strain-dependent soil nonlinearity.

Basic Mechanism on Site Amplification

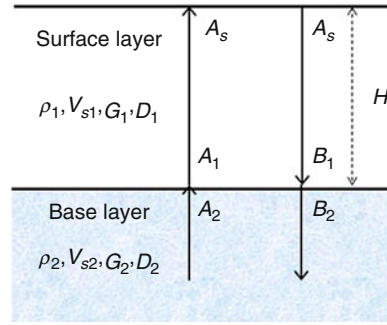
For a harmonic wave with angular frequency ω and amplitudes A , B , Eq. 1 is replaced by

$$u = Ae^{i(\omega t - kz)} + Be^{-i(\omega t + kz)} \quad (2)$$

Here, k is the wave number expressed as

$$k = \omega/V_s = \omega/(G/\rho)^{1/2} \quad (3)$$

Basic mechanisms of site amplification may be simplified by a two-layer model as illustrated in Fig. 1, where a surface layer with the thickness H overlies a base layer of infinite thickness. Pertinent properties are: ρ_1 , ρ_2 = soil density; V_{s1} , V_{s2} = S-wave velocity; G_1 , G_2 = shear modulus; and D_1 , D_2 = damping ratio of the surface



Site Response: Comparison Between Theory and Observation, Fig. 1 A simplified two-layer model to calculate site amplification

and base layer, respectively. The wave equations in the surface and base layers are expressed as

$$\begin{aligned} u_1 &= A_1 e^{i(\omega t - k_1 z)} + B_1 e^{i(\omega t + k_1 z)} \\ u_2 &= A_2 e^{i(\omega t - k_2 z)} + B_2 e^{i(\omega t + k_2 z)} \end{aligned} \quad (4)$$

where A_1 , B_1 = amplitudes of upward and downward waves in the surface layer and A_2 , B_2 = those in the base layer.

Utilizing the boundary conditions that at $z = 0$, $u_1 = u_2$ and $G_1 \frac{\partial u_1}{\partial z} = G_2 \frac{\partial u_2}{\partial z}$ and at $z = H$, $G_1 \partial u_1 / \partial z = 0$, and also introducing an impedance ratio α as

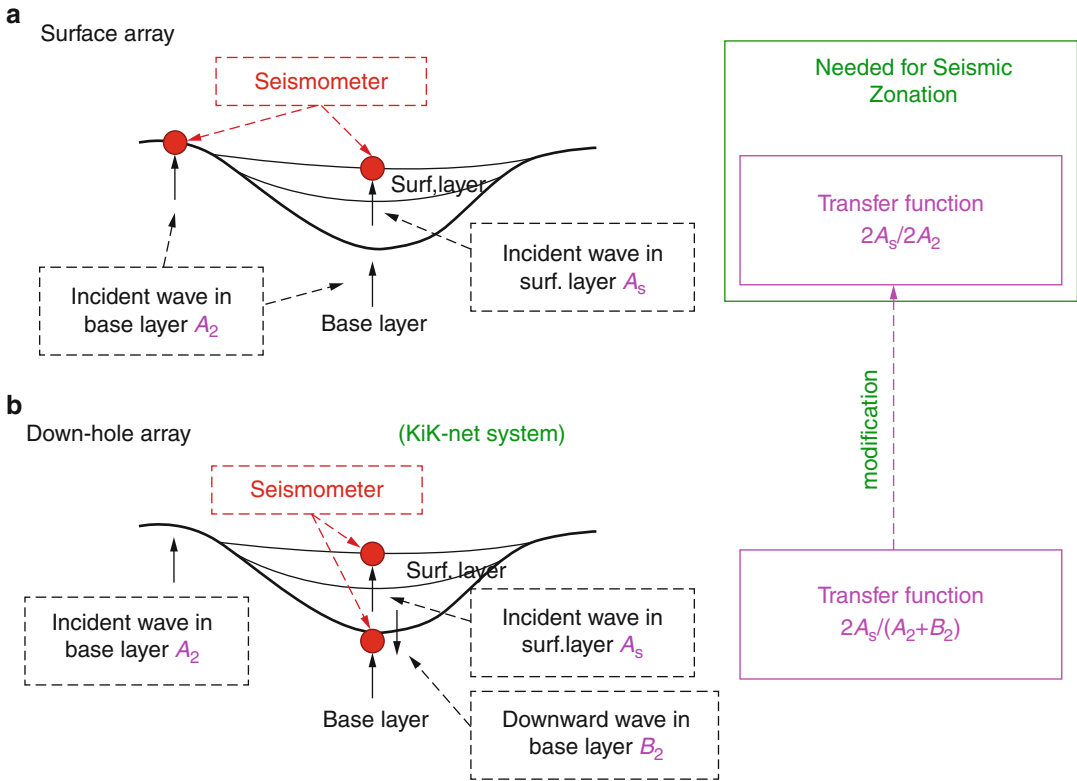
$$\alpha = \frac{k_1 G_1}{k_2 G_2} = \frac{\rho_1 V_{s1}}{\rho_2 V_{s2}}, \quad (5)$$

the transfer function $(2A_s)/(2A_2)$ is obtained as

$$\frac{2A_s}{2A_2} = \frac{2}{(1 + \alpha)e^{ik_1 H} + (1 - \alpha)e^{-ik_1 H}} \quad (6)$$

Here, $2A_s$ is the wave amplitude to be observed at the ground surface, and $2A_2$ is that at the base layer if it were outcropped and free from the overburden. The site amplification to be used in micro-zonation mapping is determined by Eq. 6.

There are two types of site response monitoring: (a) surface array and (b) downhole array as illustrated in Fig. 2. In the surface array (a), ground motion is monitored at two different



Site Response: Comparison Between Theory and Observation, Fig. 2 Two types of earthquake observation array systems to measure site amplification between ground surface and base layer

surface locations with different geologies, overlying soft layer and outcropping stiff base layer. If the soft layer is underlain by the same base layer and the upward wave in the base layer A_2 is assumed the same at the two places, the site amplification in the soft soil site with respect to outcropping base layer is given by Eq. 6.

In the downhole array (b), surface and downhole seismometers can evaluate the site amplification exactly at the same location, and its transfer function can be formulated as

$$\frac{2A_s}{A_2 + B_2} = \frac{2}{e^{ik_1H} + e^{-ik_1H}} \quad (7)$$

Here, B_2 is the amplitude of the downward wave in the base layer, which is influenced by the dynamic response of the surface layer. In order

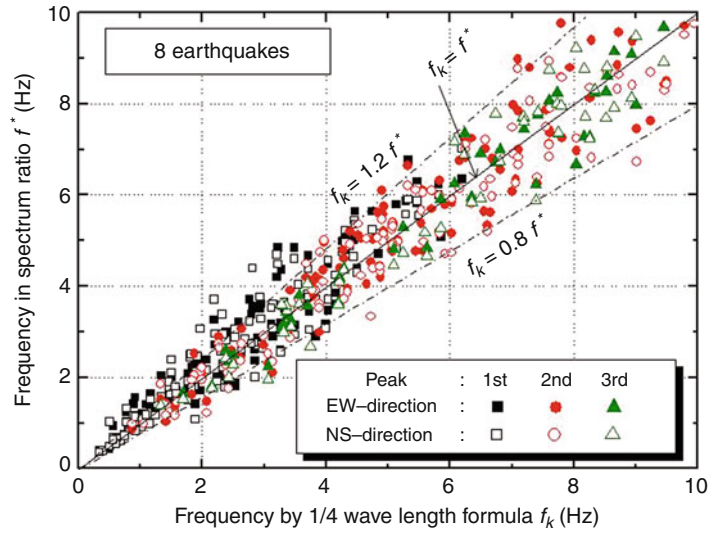
to derive the transfer function for micro-zonation $(2A_s)/(2A_2)$, Eq. 7 cannot directly be used, but some modification is necessary to extract the outcropping base motion $2A_2$ from observed base motion $(A_2 + B_2)$.

In a two-layer system, the resonant frequency can be computed by the next equation

$$f = \omega/2\pi = (2n - 1) \frac{V_{s1}}{4H} \quad (8)$$

where n is the order of resonance. The most important is the first-order resonant frequency $n = 1$, and the equation $f_1 = V_{s1}/(4H)$ is named as the 1/4 wavelength formula. In many cases, the site amplification in real site conditions with multilayer systems can be simplified by a

**Site Response:
Comparison Between
Theory and Observation,
Fig. 3** Frequency by $1/4$
wavelength formula f_k
compared to peak
frequency of observed
spectrum ratio f^* for main
shocks of eight strong EQs



two-layer model consisting of a soft surface layer underlain by a stiff base layer. In this case, too, the $1/4$ wavelength formula can be expressed in an extended form as

$$f_j = 1 / \left[4 \sum_{i=1}^j (H_i / V_{si}) \right] \quad (9)$$

Here, j is the number of surface layers involved to create a specific resonant frequency. In Fig. 3, frequencies f_j obtained from the above equation are compared with peak frequencies f^* in the spectrum ratio of observed motions at a number of downhole array sites in Japan during recent strong earthquakes. The figure shows that the value f^* are mostly within 0.8–1.2 times f_k , indicating that the extended $1/4$ wavelength formula Eq. 9 may be useful to roughly estimate the resonant frequency by simplifying actual multilayer soil systems by the two-layer system. Also note in the diagram that Eq. 9 holds for not only the first- but also second- and third-order peak frequencies, which are generated by combinations of base layers of different depths and corresponding overlying layers.

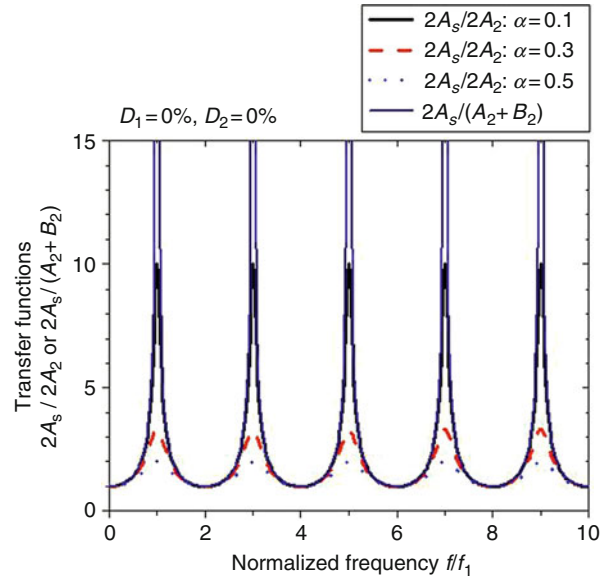
In Fig. 4, two types of transfer functions of the two-layer system in Fig. 1 calculated by Eqs. 6 and

7 are compared in the same diagram. It is apparent from the two equations that the properties of the base layer are included in $(2A_s)/(2A_2)$ in terms of impedance ratio α but not in $2A_s/(A_2 + B_2)$. When $\alpha = \rho_1 V_{s1} / \rho_2 V_{s2} = 0$ in Eq. 6, indicating $\rho_2 V_{s2} \rightarrow \infty$ (rigid base layer), Eqs. 6 and 7 are identical. Hence, in the downhole array transfer function, the base layer is equivalent to a rigid base with the prescribed motion $A_2 + B_2$ when no radiation damping occurs, resulting in infinite amplification in resonant frequencies if there is no soil damping in the surface layer, $D_1=0$. In contrast, $(2A_s)/(2A_2)$ produces a certain finite peak values corresponding to α even for $D_1=0$, because energy radiation into the base layer is possible. It is noteworthy that in a multilayer system, the base layer where the downhole seismometer is installed also serves as the rigid base, whose properties have nothing to do with the downhole array amplification (Schnabel et al. 1972).

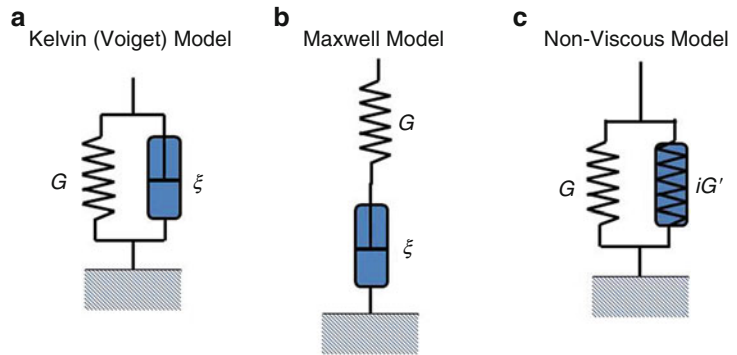
Soil Damping Models and Site Amplification

In order to incorporate soil damping in site amplifications, the variables included in Eqs. 6 and 7, k ,

Site Response:
Comparison Between
Theory and Observation,
Fig. 4 Comparison to two
 types of transfer function
 $2A_s/2A_2$ and $2A_s/(A_2 + B_2)$
 in a two-layer system



Site Response:
Comparison Between
Theory and Observation,
Fig. 5 Three mechanical
 models for soil damping



V_s , and G , should be replaced by complex wave number k^* , complex S-wave velocity V_s^* , and complex shear modulus G^* , respectively, so that Eq. 3 is modified as

$$k^* = \omega/V_s^* = \omega/(G^*/\rho)^{1/2} \quad (10)$$

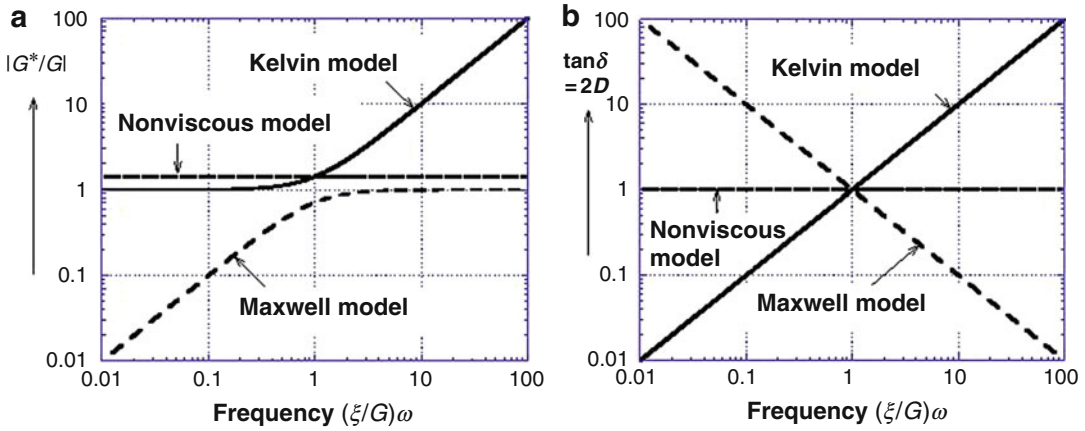
Three mechanical models for soil damping can be incorporated in site amplification: (a) Kelvin (Voigt) model, (b) Maxwell model, and (c) nonviscous model as illustrated in Fig. 5, consisting of a spring with shear modulus G and a dashpot with viscosity ζ . For model (c), the dashpot is replaced by a nonviscous

(time independent) dashpot with an imaginary constant iG' . For each model, shear stress τ versus shear strain γ relationship is expressed as

$$\tau = G^* \gamma \quad (11)$$

where G^* is complex shear modulus considering soil damping. The complex shear modulus G^* normalized by G is formulated for each damping models a, b and c using $i = \sqrt{-1}$ as follows (Ishihara 1996):

$$G^*/G = 1 + i(\zeta/G)\omega \quad (12a)$$



Site Response: Comparison Between Theory and Observation, Fig. 6 Variations of normalized shear modulus $|G^*/G|$ and $\tan \delta$ along with normalized angular frequency $(\xi/G)\omega$

$$G^*/G = \frac{i\omega}{1/(\xi/G) + i\omega} \quad (12b)$$

$$G^*/G = 1 + iG'/G \quad (12c)$$

In these equations, the normalized complex shear modulus can be expressed as

$$G^*/G = |G^*/G|e^{i\delta} \quad (13)$$

where δ is the phase lag angle between stress–strain relationships of the models, and $\tan \delta$ is called a loss coefficient. In Fig. 6, the variations of normalized complex shear modulus $|G^*/G|$ and $\tan \delta$ are shown versus dimensionless angular frequency $(\xi/G)\omega$ for models (a) and (b), and constant values are shown for model (c). The damping ratio is correlated with the loss coefficient as follows (Ishihara 1996):

$$D = (\tan \delta)/2 \quad (14)$$

Hence, from Eqs. 12a, b, and c:

$$D = \tan \delta/2 = \omega\xi/(2G) \quad (15a)$$

$$D = \tan \delta/2 = G/(2\omega\xi) \quad (15b)$$

$$D = \tan \delta/2 = G'/(2G) \quad (15c)$$

From Eq. 2, the equation for a wave including soil damping propagating unidirectionally toward $+z$ is written as

$$u = Ae^{i(\omega t - k^*z)} \quad (16)$$

If k^* in this equation is substituted by Eq. 10, and G^* in Eq. 10 is further substituted by Eqs. 12a, b, and c, Eq. 16 can be transformed as follows:

$$u = Ae^{-\beta z}e^{i\omega(t-z/V_s)} \quad (17)$$

Here, β is a positive real number called a wave attenuation coefficient by internal damping, because it determines how the wave attenuates as it propagates by distance z . If δ is small, β is correlated to the damping ratio D as

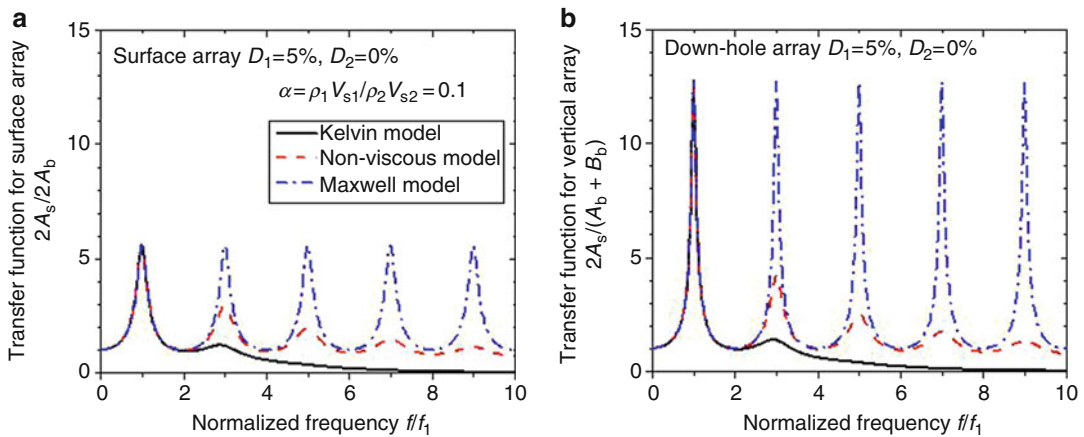
$$\beta = \omega D/V_s \quad (18)$$

Substituting D in Eq. 15a, b and c into Eq. 18 gives β for the corresponding damping models:

$$\beta = \omega^2\xi/(2\rho V_s^3) \quad (19a)$$

$$\beta = \rho V_s/(2\xi) \quad (19b)$$

$$\beta = \omega G'/(2\rho V_s^3) \quad (19c)$$



Site Response: Comparison Between Theory and Observation, Fig. 7 Transfer functions of a two-layer model using three types of soil damping models: (a) $2A_s/$

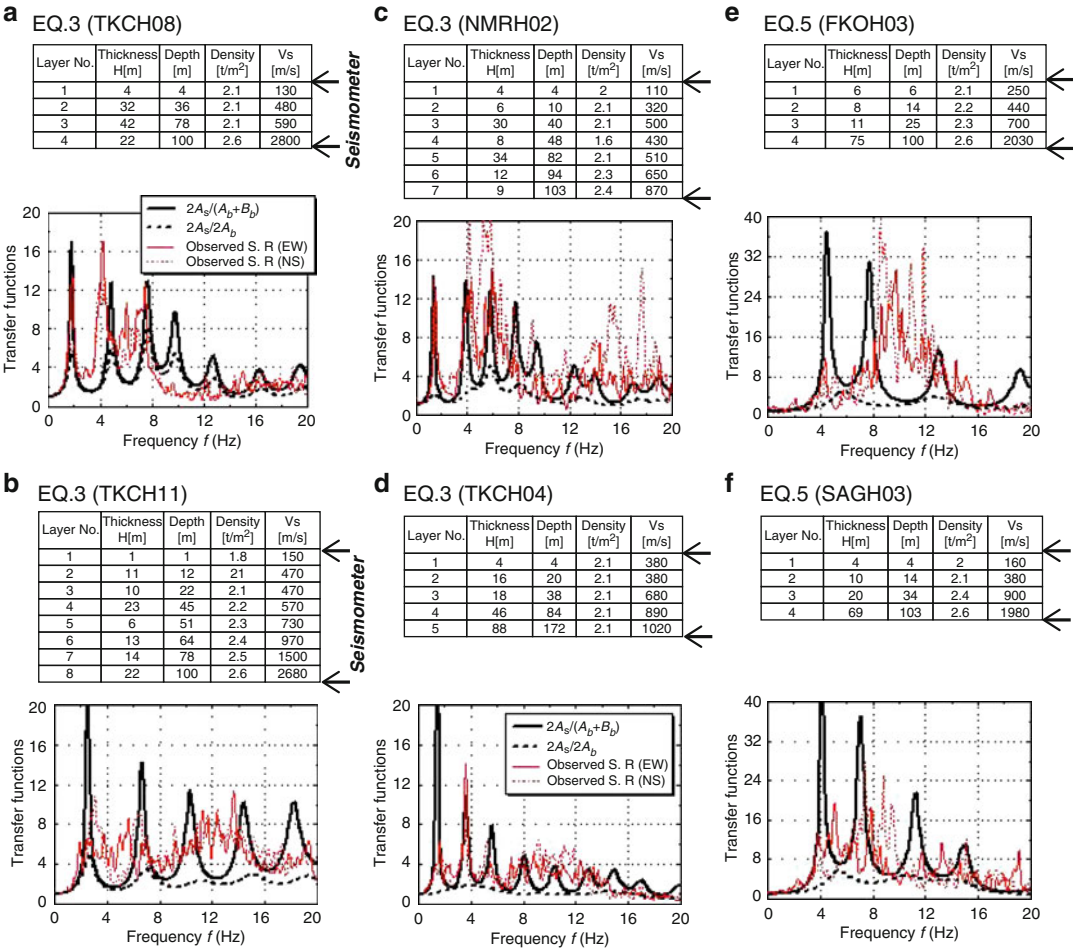
$2A_b$ ($\alpha = 0.1$) for surface array and (b) $2A_s/(A_b + B_b)$ for downhole array

In Fig. 7, transfer functions of the two-layer model in Fig. 1 are shown for the three damping models assuming the damping ratios in the surface and base layers $D_1 = 5\%$ and $D_2 = 0\%$, respectively. Figure 7a is for the surface array $(2A_s)/(2A_b)$ with the impedance ratio $\alpha = (\rho_1 V_{s1})/(\rho_2 V_{s2}) = 0.1$, and Fig. 7b is for the downhole array $2A_s/(A_b + B_b)$. The transfer functions are identical in the first peak but obviously differ in higher-order peaks corresponding to the different damping models. In the Kelvin model, the amplification in higher-order peaks tend to reduce more drastically than in the nonviscous model, and in the Maxwell model, it is unchanged at all peaks. These differences arise from the different formulations of β in Eqs. 19a, b, and c with respect to ω . In cyclic loading tests on soil elements, it is widely accepted that soil damping is almost frequency independent (Hardin 1965; Hardin and Drnevich 1972). Hence, in most engineering practice, the soil damping is assumed nonviscous as in Eq. 15c. In this case, the peak amplification of the transfer function becomes lower for higher-order peaks because the wave attenuation coefficient β is proportional to ω as shown in Eq. 19c.

In Figs. 8a–e, some typical Fourier spectrum ratios calculated between surface and base

records at six downhole array sites where PGA exceeds 0.2 g during strong earthquakes are depicted with thin lines in EW and NS directions. For each site, the soil profile with density ρ and S-wave velocity V_s is tabulated in Table 1 together with installation levels of seismometers indicated with arrows. Using the properties in the table, transfer functions for each site, $2A_s/(A_b + B_b)$ for the downhole array and $(2A_s)/(2A_b)$ for the surface array, are calculated and superposed in the diagram with thick curves, where A_b and B_b = amplitudes of upward and downward waves in the base layer. In the calculation, the nonviscous damping model is used as normal engineering practice, and a uniform damping ratio $D = 2.5\%$ is assumed tentatively for all layers. If the observed spectrum ratios are compared with corresponding transfer functions for downhole arrays, $2A_s/(A_b + B_b)$, a fairly well correspondence in peak frequencies can be recognized between the two in most sites despite some minor effect of soil nonlinearity. This indicates the applicability of one-dimensional soil models in these sites to a certain extent.

If $2A_s/(A_b + B_b)$ is compared with $(2A_s)/(2A_b)$ at each site, the difference in the peak values between the two transfer functions is obviously large.



Site Response: Comparison Between Theory and Observation, Fig. 8 Transfer functions of a multilayer system $2A_s/(A_b + B_b)$ compared with $2A_s/2A_b$ and

observed spectrum ratios at six downhole array sites during a strong earthquake

However, the coincidence in peak frequencies is almost perfect in (a) and good in (b) but gets poorer in (c) and (d) and very poor in (e) and (f). The reason may be gleaned by examining the soil profiles. In (a) and (b), the V_s -value at the downhole seismometer is much larger than the upper layers, and the seismometer depth is not so deep from the boundary of clear V_s -contrast. In (c) and (d), the V_s -value at the base layer is not so different from the upper layers, and the seismometer depth is not so deep from a boundary of major V_s -contrast. In (e) and

(f), though the V_s -value at the base layer is much larger than the upper layers, the depth of seismometer is too deep from the upper boundary of clear V_s -contrast to properly detect the response of the upper layers. This observation tells us a significance of choosing appropriate seismometer depth in deploying a downhole array system considering site specific soil profiles.

Another observation may be made on amplification values at individual peaks of observed spectrum ratios compared with the downhole array transfer functions $2A_s/(A_b + B_b)$.

Site Response: Comparison Between Theory and Observation, Table 1 Soil profiles and properties at two vertical array sites corresponding to spectrum ratios in Figs. 9 and 10

For Fig. 9				
IBUH03	Soil profiles and properties			
Layer no.	Thickness h (m)	Depth H (m)	Density (t/m ³)	V_s (m/s)
1	2	2	21	60
2	16	18	21	90
3	10	28	20	190
4	12	40	21	320
5	12	52	21	210
6	24	76	21	310
7	56	132	21	430
8	21	153	22	520
For Fig. 10				
NMRH02	Soil profiles and properties			
Layer no.	Thickness h (m)	Depth H (m)	Density (t/m ³)	V_s (m/s)
1	4	4	2.0	110
2	6	10	2.1	320
3	30	40	2.1	500
4	8	48	1.6	430
5	34	82	2.1	510
6	12	94	2.3	650
7	9	103	2.4	870

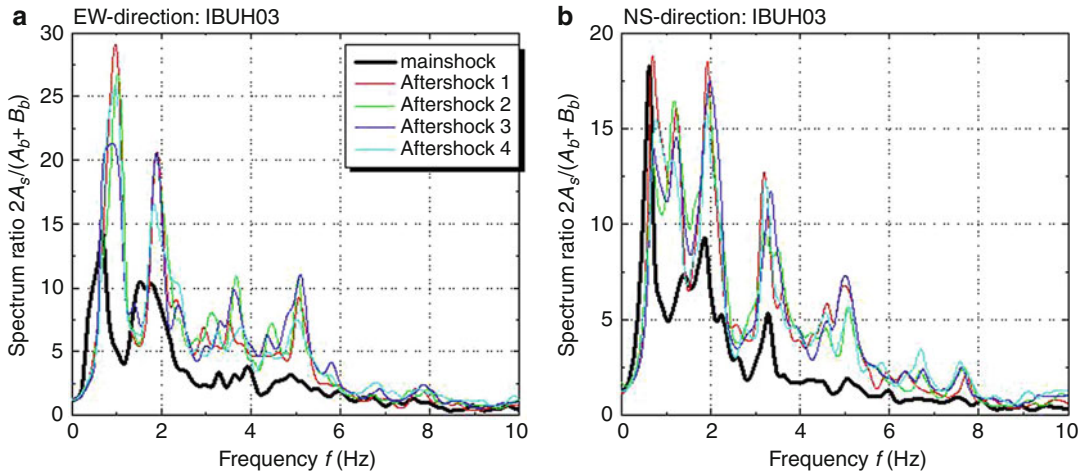
As already mentioned, the damping ratio is tentatively set as 2.5 % in calculating $2A_s/(A_b + B_b)$. At a glance, the peak amplification values of the transfer functions are almost monotonically decreasing with increasing peak frequencies, presumably reflecting the damping characteristics of the non-viscous model used here, as previously explained. In contrast, the peak values of observed spectrum ratios seem to be quite site specific and show no consistent increasing or decreasing trend. This indicates that the soil damping is not perfectly nonviscous (frequency independent) but to some extent frequency dependent (the damping ratio D decreases with increasing frequency f) in a similar manner as the Maxwell model. It is clear that the Kelvin model, for which the peak amplification tends to decrease more drastically with increasing f , has even less applicability than the nonviscous model. Thus, the frequency-independent damping found in laboratory soil tests may not perfectly fit the actual performance of site amplification based on earthquake observations.

One major reason for this deviation may be wave scattering in heterogeneous soils. The wave scattering effect on frequency dependency of damping in elastic waves propagating in heterogeneous earth crust is an important topic in seismology (e.g., Frankel and Clayton 1986; Wu 1982). The effect tends to be more conspicuous with decreasing strain in a stiff rock site. The damping ratio D by wave scattering may be expressed as

$$D = D_r(f/f_r)^{-m} \quad (20)$$

where f_r = a reference frequency, D_r = reference damping ratio, and $D = D_r$ at $f = f_r$. The power constant m is a positive number, and $m = 1$ corresponds to the Maxwell model, as obviously seen in Eq. 15b. In general, the damping ratio may be expressed as the sum of the frequency-independent hysteretic part D_0 and the frequency-dependent part as

$$D = D_0 + D_r(f/f_r)^{-m} \quad (21)$$



Site Response: Comparison Between Theory and Observation, Fig. 9 Fourier spectrum ratios $2A_s/(A_b + B_b)$ observed at a very soft soil site during main shock and aftershocks in directions EW (a) and NS (b)

However, the frequency-dependency of damping is normally ignored in geotechnical engineering practice, because hysteretic nature of soil damping becomes more dominant during strong earthquakes as soil gets softer and strain-dependent nonlinearity becomes more dominant.

Soil Nonlinearity

Seismically induced shear strain γ is calculated from particle velocity $\dot{u}$:

$$\gamma = -\dot{u}/V_s \quad (22)$$

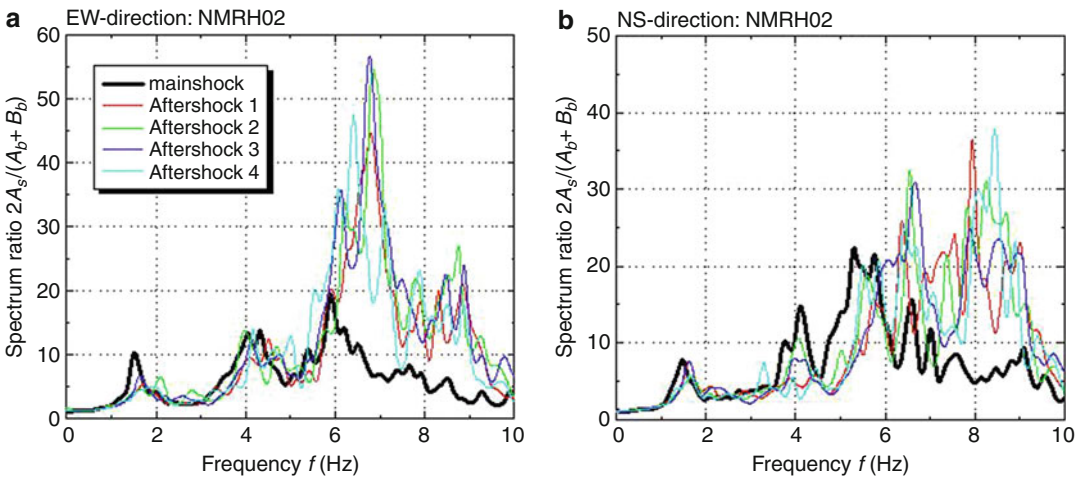
if upward propagating wave is chosen in Eq. 1, for instance. This indicates that the shear strain tends to be greater for strong earthquakes with larger particle velocity amplitudes and in soft soils with smaller V_s -values. Because stress-strain behavior of soil is nonlinear and hence V_s and D are highly strain dependent, the site response during strong earthquakes may greatly differ from that during small earthquakes, particularly in soft soil sites. The soil nonlinearity effect is normally evaluated using the equivalent linear approximation method in engineering practice (Schnabel et al. 1972).

In each of Figs. 9 and 10, Fourier spectrum ratios for a strong main shock are compared with

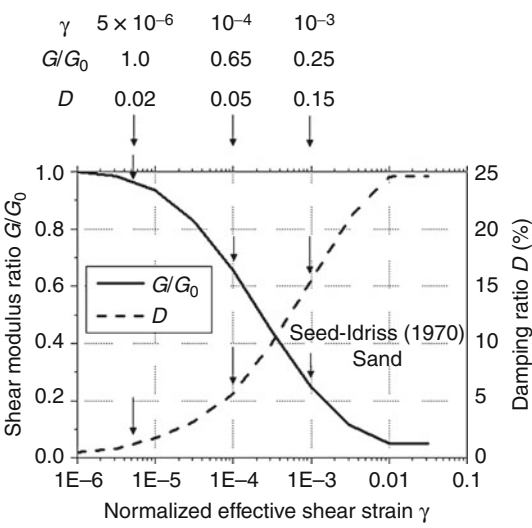
those for weaker aftershocks in directions EW (a) and NS (b) using earthquake records obtained in the same downhole arrays. As indicated in Table 1, the soil condition for the former site is very soft ($V_s < 100$ m/s at top 18 m), while that of the latter site are relatively stiff ($V_s > 320$ m/s deeper than 4 m). Thick curves for the main shock are obviously different with thin curves for the aftershocks with respect to peak values and peak frequencies. The difference of peak values is greater in Fig. 9 than in Fig. 10 because of the softer soil conditions, and the difference in the same site tends to be greater in higher-order peaks than in the first peak because softer layers near the ground surface tends to contribute more.

Basic effects of soil nonlinearity on site amplification can be examined using the two-layer system shown in Fig. 1. Three levels of induced equivalent strain amplitude are assumed, and corresponding shear modulus degradation G/G_0 and equivalent damping ratio D (nonviscous) are determined based on empirical curves (Fig. 11) often used in engineering practice (Seed and Idriss 1970). Figures 12a and b depict the transfer functions $2A_s/(A_b + B_b)$ for the downhole array and $(2A_s)/(2A_b)$ for the surface array calculated for the three levels of nonlinear soil properties.

Obviously, soil nonlinear properties have great effects on the peak frequencies and the peak amplifications not only in $2A_s/(A_b + B_b)$ as



Site Response: Comparison Between Theory and Observation, Fig. 10 Fourier spectrum ratios $2A_s/(A_b + B_b)$ observed at a stiff soil site during main shock and aftershocks in directions EW (a) and NS (b)



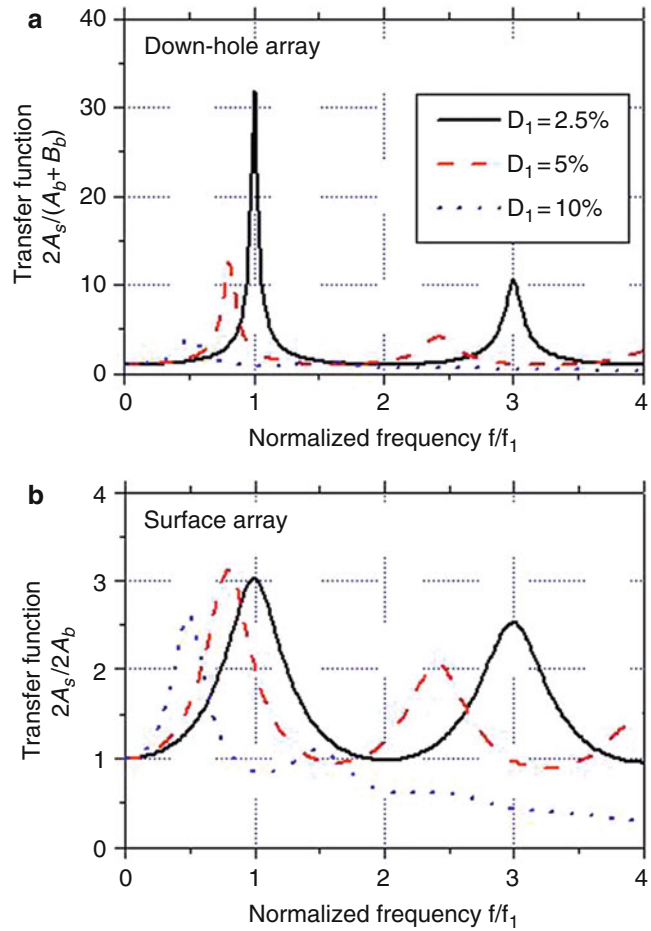
Site Response: Comparison Between Theory and Observation, Fig. 11 Soil nonlinearity curves (G/G_0 - γ , D - γ) used in calculating amplifications in a two-layer system

recognized in actual site amplifications in Figs. 9 and 10 but also in $(2A_s)/(2A_b)$. However, the difference in the peak amplifications due to strain level is less pronounced in $(2A_s)/(2A_b)$ than in $2A_s/(A_b + B_b)$ for the first peak in particular. It is because the radiation damping effect presented by the impedance ratio α affects $(2A_s)/(2A_b)$ in Eq. 6, whereas no effect of α is involved in

$2A_s/(A_b + B_b)$ as indicated in Eq. 7. Under the paramount effect of radiation damping associated with α , the difference in the amplification $(2A_s)/(2A_b)$ due to strain-dependent nonlinear properties becomes less conspicuous. Furthermore, the impedance ratio, $\alpha = \rho_1 V_{s1}/\rho_2 V_{s2}$, which becomes smaller with degraded modulus or degraded S-wave velocity V_{s1} in the surface layer, tends to give larger amplification compensating the effect of increased damping ratio in the surface layer during strong earthquakes. Thus, the difference in soil nonlinearity between the main shock and aftershocks seems to have smaller influence on the amplification in $(2A_s)/(2A_b)$ than in $2A_s/(A_b + B_b)$ as indicated from the comparison of Fig. 12a, b.

The Fourier spectrum ratio corresponding to the transfer function $(2A_s)/(2A_b)$ cannot be obtained directly from downhole array records, but the peak value can be calculated from $2A_s/(A_b + B_b)$. Figure 13 shows an example how this calculation is carried out. First, a transfer function $2A_s/(A_b + B_b)$ is calculated at a downhole array site based on the multi-reflection theory. Among soil properties needed, the S-wave velocities of individual layers are given by in situ logging tests if strain-dependent soil nonlinearity is not so significant, and soil densities are judged from previous experiences.

**Site Response:
Comparison Between
Theory and Observation,
Fig. 12** Transfer functions
of a two-layer system (a)
 $2A_s/(A_b + B_b)$ for
downhole array and (b)
 $2A_s/2A_b$ for surface array in
different induced strain
levels



The damping ratio D is tentatively assumed 2.5 % in all layers and also postulated to be nonviscous or frequency independent.

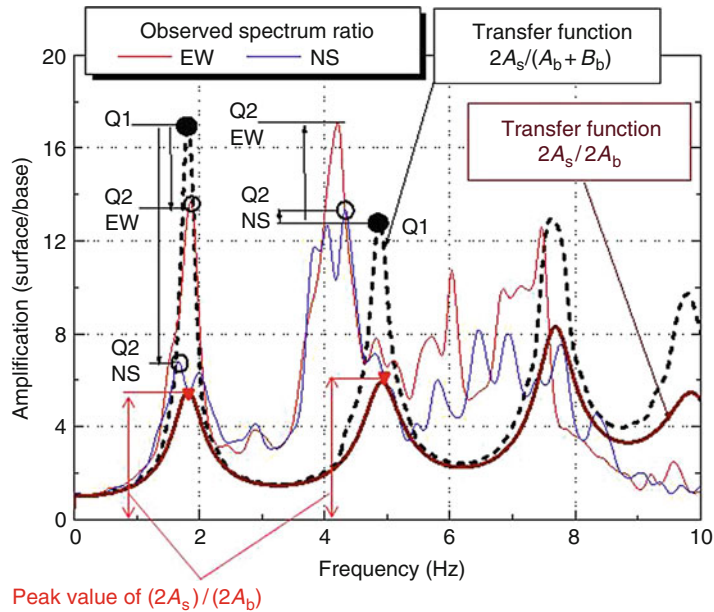
Then, the calculated transfer function $2A_s/(A_b + B_b)$ is compared with the corresponding spectrum ratio obtained from downhole array records. If a peak in the transfer function can be found at about the same frequency in the spectrum ratio of recorded motions, it is identified as the corresponding peak, and the damping ratio assumed as $D = 2.5\%$ previously is modified to have the same peak value, by using the equation $D = Q_1/Q_2 \times 2.5(\%)$, where Q_1 is the peak value of the calculated transfer function and Q_2 is that of spectrum ratio based on the actual records as indicated in Fig. 13. Not only the first peak but also the higher-order peaks are

compared in this manner if possible, and the values of D in the two directions, EW and NS, are averaged for individual peaks. Then, the transfer function $(2A_s)/(2A_b)$ is computed using the modified damping ratio D based on the same multilayer system. In this way, the peak amplifications for $(2A_s)/(2A_b)$ between surface and base, to be used in seismic zonation study for surface soil resting on the common base layer, can be obtained using the downhole array records.

In Fig. 14, the peak amplification values in the first peak of the spectrum ratios are compared between small aftershocks (in the horizontal axis) where shaking is relatively smaller ($\text{PGA} \leq 0.1 \text{ g}$) and main shock (in the vertical axis) with $\text{PGA} \approx 0.1 \sim 2.4 \text{ g}$, based on vertical array records during strong earthquakes and

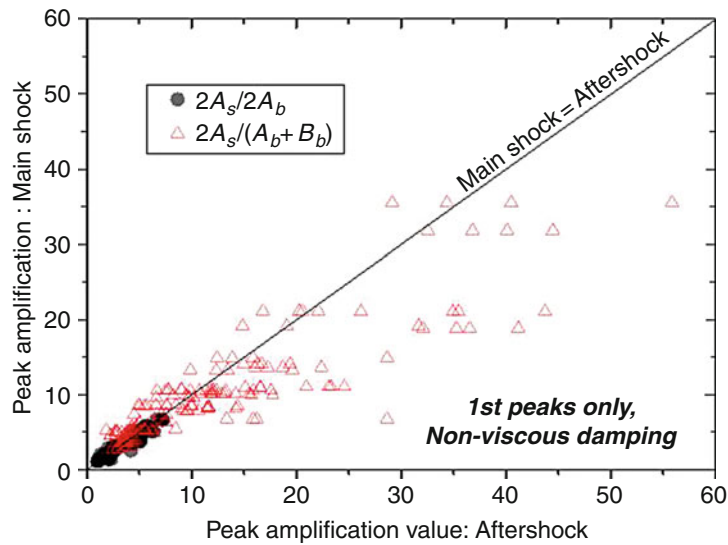
Site Response:
Comparison Between
Theory and Observation,

Fig. 13 How to obtain spectrum peak $2A_s/2A_b$ from $2A_s/(A_b + B_b)$ of vertical array observed motions at the same site



Site Response:
Comparison Between
Theory and Observation,

Fig. 14 Comparison of spectrum peak amplifications between aftershocks and main shocks of downhole arrays $2A_s/(A_b + B_b)$ and surface arrays in $2A_s/2A_b$



associated aftershocks recorded at many vertical array sites in Japan (Kokusho and Sato 2008; Kokusho 2013). The triangular symbols for downhole arrays $2A_s/(A_b + B_b)$ are dispersed in a wide range, and the majority is plotted around or below the diagonal line (main shock = aftershock), indicating that during strong shaking the site amplification in terms of $2A_s/(A_b + B_b)$

possibly reduces due to soil nonlinearity. In contrast, the solid circular symbols corresponding to $(2A_s)/(2A_b)$ in surface arrays, though the amplification values are small, concentrate near the diagonal line, indicating that the effect of soil nonlinearity is less dominant in the first peak site amplification, as demonstrated in Fig. 12 by using a simplified two-layer model.

Thus, the soil nonlinearity affects the site amplification considerably with respect to peak frequency and peak amplification. However, for the amplification of the first peaks in the surface array, the nonlinear effect is obviously minor. This may simplify the seismic zonation procedure considering strong shaking.

Summary

1. Two types of site amplification monitoring can be defined and implemented in earthquake observations, surface array and downhole array. For micro-zonation mapping, the amplification for the surface array is needed.
2. For reproduction of site amplification observed in the field, it is possible to simplify a complicated multilayer system to a two-layer system composed of a surface layer of a certain thickness underlain by a base layer of infinite thickness. Resonant frequencies of the multilayer system can roughly be evaluated using the $1/4$ wavelength formula.
3. In the downhole array, the installation level of the downhole seismometer serves as a rigid boundary, and soil properties below have nothing to do with the amplification of the ground surface with respect to the base motion.
4. Among three soil damping models, Kelvin, Maxwell, and nonviscous, the nonviscous model is normally used for site amplification evaluation in engineering practice. In this case, the peak amplification of the transfer function becomes lower for higher-frequency peaks because the wave attenuation coefficient β is proportional to the frequency.
5. There are evidences from earthquake observations to suspect that some degree of frequency-dependent damping similar to Maxwell damping due to a wave-scattering mechanism exists in the field. However, it is normally ignored in geotechnical engineering, because hysteretic (nonviscous) nature of soil damping becomes more dominant during strong earthquakes as soil gets softer and more strain-dependent nonlinear.

6. Site amplification observations show that soil nonlinearity affects the site amplification considerably with respect to peak frequency and peak amplification. However, for the amplification of the first peak in the surface array, the nonlinear effect is obviously minor, which may simplify the seismic zonation procedure even for strong earthquakes.

Cross-References

- [Downhole Seismometers](#)
- [Dynamic Soil Properties: In Situ Characterization Using Penetration Tests](#)
- [Site Response: 1-D Time Domain Analyses](#)

References

- Frankel A, Clayton RW (1986) Finite difference simulations of seismic scattering: implications for the propagation of short-period seismic waves in the crust and models of crustal heterogeneity. *J Geophys Res* 91(B6):6465–6489
- Hardin BO (1965) The nature of damping in sands. *Proc Soil Mech Found Div, ASCE* 91(SM1):63–97
- Hardin BO, Drnevich VP (1972) Shear modulus and damping in soils: measurement and parameter effects. *J Soil Mech Found Div, ASCE* 98(SM6):603–624
- Ishihara K (1996) The representation of stress–strain relations in cyclic loading, Chapter 3. In: Ishihara K - (ed) *Soil behaviour in earthquake geotechnics*, Oxford engineering science series, 46. Clarendon, Oxford, pp 16–39
- Kanai K, Tanaka T, Yoshizawa S (1959) Comparative studies of earthquake motions on the ground and underground. *Bulletin of the Earthquake Research Institute, Tokyo University*, Vol. 37, 53–87.
- Kokusho T (2013) Site amplification formula using average vs in equivalent surface layer based on vertical array strong motion records. *ICEGE, Istanbul*
- Kokusho T, Sato K (2008) Surface-to-base amplification evaluated from KiK-net vertical array strong motion records. *Soil Dyn Earthq Eng* 28:707–716
- Schnabel PB, Lysmer J, Seed HB (1972) SHAKE – a computer program for earthquake response analysis of horizontally layered sites, Report EERC 72-12. University of California, Berkeley
- Seed HB, Idriss IM (1970) Soil Moduli and damping factors for dynamic response analysis, Report EERC 70-10. University of California, Berkeley
- Wu R-S (1982) Attenuation of short period seismic waves due to scattering. *Geophys Res Lett* 9(1):9–12

Social Media Benefits and Risks in Earthquake Events

Hamish McLean

School of Humanities, Griffith University,
Nathan, QLD, Australia

Synonyms

Citizen seismologists; Disaster communication analysis; Earthquake education; Enhanced situational awareness; Post-disaster recovery; Psychological first aid; Real-time warnings on Twitter

Introduction

The rapidly evolving social media platforms, with an estimated 1.9 billion users worldwide, offer a myriad of communication benefits and risks in the context of a disaster. Social media generally refers to internet-based technologies that enable people to interact and share resources and information using either text or multimedia applications (Lindsay 2011; Dabner 2011). Advances in mobile devices allow access to anyone who has the ability to connect online (Abbasi et al. 2012). For example, the microblogging platform Twitter allows followers to track what an account holder is doing and thinking in real time within the confines of 140 characters (Kaigo 2012). Tweets can be sent from a variety of platforms ranging from cell phones to computers. Other examples of social media platforms today are Facebook, YouTube, Qzone, Pinterest, Instagram, and Flickr. This entry describes how social media, in particular Twitter, can be used for a variety of applications before, during, and after an earthquake, both inside and outside the area of impact. These include detection, warnings, connecting to survivors, situational awareness, notifying responders of where help is needed, and galvanizing humanitarian aid. Importantly, the increasing participation of “citizen seismologists” via social media is filling the information gap with field observations

immediately after an earthquake (Young et al. 2013) for both first responders and survivors. In fact, social media, in particular Twitter, may be the only immediate source of data from locations with limited sensors or other scientific instruments. Obviously, not all earthquakes are reported on social media as many events are in remote areas or undersea and in countries with limited social media access or the magnitude is too small to be felt. A unique benefit of social media is that it is user generated – disaster agencies, seismologists, and other parties do not have to motivate citizens to tweet – they will do it anyway, potentially by the thousands in a significant earthquake. The challenge is how to transform the rapidly spreading flood of real-time information, some of it inaccurate, into reliable, useful, and valuable data. Part of the solution is to train Twitter users to tweet messages that can be more easily analyzed both manually and automatically using a crisis-specific syntax (Starbird and Stamberger 2010). This approach, among others discussed in this entry, will help guide disaster response, galvanize ongoing humanitarian efforts, and add value to the expanding body of earthquake sciences gathered since the development of the modern seismograph in the late 1800s.

Sourcing Information in Disasters

People facing a disaster seldom act on one source of information. Hunting for firsthand local information, they will “channel swap” across the traditional media, go online to news websites and social media, and contact family and friends. Without proof from a variety of trusted sources, people will believe there is no immediate threat or that the situation does not apply to them. The delay in taking action – known as milling – can increase the risk of death or injury from a hazard before steps are taken for self-protection. The milling effect is increased if warnings are vague and conflicting across various channels or the credibility of the source is in doubt. Under these circumstances, social media becomes a double-edged sword. The speed of social media can reduce the decision-making lag time while, at

the same time, increasing confusion and uncertainty if the information is wrong or conflicting. It is generally accepted that trust in social media information remains well below that of traditional media. The 2014 Edelman Trust Barometer found that 47 % of people trusted social media sources, while 65 % trusted the traditional media. In practice this was reflected in the 2011 Japan earthquake. Although most people turned to social media as their most reliable source of information, one third described the “lack of trust in formation as the greatest problem associated with social media use in the disaster” (Perry et al. 2012, p. 15). For disaster agencies, the challenge of earning trust and building credibility and relationships can be overcome by actively engaging online before the disaster. As noted by the US-based Centers for Disease Control and Prevention (CDC): “Organisations need to be regular users of social media before the crisis. If not, social media users will go to other sources and groups with whom they already have relationships for information” (2012, p. 268). Engaging with social media before a disaster can also build preparedness and education for survival. For example, more than 1.3 million people in New Zealand took part in the “ShakeOut” earthquake drill in 2012, driven by a social media campaign. Many organizations took the drill realistically. For example, a hospital in the country’s North Island shut off its power and water for 24 h to simulate the impact of an earthquake. Portable cooking, lighting, and other equipment maintained hospital services (Civil Defence Emergency Management 2012). Similar events are held in other countries to prepare communities for an earthquake. Although disaster agencies are using social media to educate communities and disseminate information, they have been historically more reluctant to rely on social media data in their incident command systems. The reasons include fears about misinformation (Williams et al. 2012), the speed and spread of the information which makes validation difficult (Gowing 2009), and a lack of understanding about how they can make use of social media (Duffy 2012; Palen 2008). For example, Tapia et al. (2013) argue that a major issue for

disaster agencies is how social media data can be effectively incorporated into time-critical decision-making processes. As Tapia et al. (2013, p. 770) note, “while data quality continues to be a barrier, what is far more important to organizational use is the serving of this data at the appropriate time, in the appropriate form to the appropriate person and the appropriate level of confidence.”

Online Support on the Ground

Driven by developing mobile technology, social media is taking on an increasing role in connecting people in disasters. Part of the reason is that people reach out to both their “offline” and “online” communities during a crisis or disaster (Dutta-Bergman 2006). In other words, they parallel their physical world with their virtual world to garner “social support and gather information, and vice versa.” In turn, on a much wider scale, this online convergence builds and strengthens community resilience through “people power” (Duffy 2012). A bank of social capital is developed by the exchange of information “during difficult times” to build relationships between people (Kaigo 2012). Therefore, social media in disaster impact areas is frequently driven by the community to share knowledge and as a form of empowerment toward recovery. The CDC (2012) notes that the public uses social media on a greater scale in the hours after a disaster than official agencies. Williams et al. (2012), who have produced a practical guide to community-based social media in disasters based on lessons from a series of tornadoes and floods in the United States, found that post-disaster social media was generated and driven by citizens rather than emergency agencies. In most cases, community-managed social media was the primary source of information for those impacted by the disaster. For example, a University of Missouri Extension Facebook site – Branson Tornado Info – attracted 14,000 followers within 12 h of a tornado in February 2012. One victim posted: “For the first few days after the storm, this Facebook page was our main source of

Social Media Benefits and Risks in Earthquake Events, Table 1 The expanding role of social media in earthquake events

2008 Sichuan (China) earthquake	The first alert of the severity of the earthquake was reported on Twitter (Moore 2008; Cellan-Jones 2008)
2010 Haiti earthquake	First major earthquake in the age of social media. The purpose-built Ushahidi crisis map became an emergency reporting system to locate people texting for help. Social media-generated significant humanitarian aid (Meier 2012; MacLeod 2010)
2011 Christchurch (NZ) earthquake	Social media became “the Church or meeting hall” for people to support each other and share information (Mathewson 2012)
2011 Japan earthquake and tsunami	Social media platforms, particularly Facebook, Twitter, and Japan’s own social media site Mixi, provided information and connected families when telephone communication was damaged or became congested (Wallop 2011)
2011 Turkey earthquake	Facebook was used to coordinate donations and aid requests. People trapped in rubble tweeted for help (Turgut 2011)
2013 Ya’an (China) earthquake	Social media platforms become a place for mourning when people turned their profile photographs to gray in remembrance of the victims (Hui 2013)

information. Volunteers here answered our questions about where to go to get help, what resources were available and what we needed to do next” (Williams et al. 2012, p. 18). In the context of earthquakes, social media has played an important and increasingly sophisticated role, as the following Table 1 demonstrates:

Twitter Faster than Earthquakes

It is well documented that the real-time speed and user-input microblogging capability of Twitter make it one of the most useful social media tools for disaster management agencies to gain a rapid snapshot of the earthquake aftermath and

the level of response and resources that may be required. Earthquake alerts are now within the scope of Twitter. Crooks et al. (2013) found that the velocity of Twitter can be used as a warning system in large-scale events. Tweets and re-tweets spread from the epicenter faster than the physical effects to distant locations. For example, tweets about the 2011 Virginia earthquake were read in New York 30 s before it was felt there, “showing that information moves faster through networks than the earthquakes themselves” (Perry et al. 2012, p. 6). Automatic earthquake warnings sourced from Twitter data are being developed using algorithms to analyze keywords, the number of words, and their context (Sakaki et al. 2010). Importantly, Tweets can include locations, which is fundamental in sensing earthquake events (Sakaki et al. 2010). Another real-time online method of detecting earthquakes in a general location is to trace the IP addresses of visitors to earthquake information websites. For example, surges in visitor traffic to the European-Mediterranean Seismological Centre from people wanting information about what they had felt can provide a snapshot within 2 min into the location and potential damage from an earthquake (Bossu et al. 2011).

Greater Situational Awareness

Tweets within the first minutes of an earthquake are mostly generated from around the epicenter and provide potentially useful situational awareness for both emergency responders, seismologists, and, importantly, survivors. In the 2011 Christchurch, New Zealand, earthquake, the first tweet was within 30 s, the first photo was within 4 min, and the first video was uploaded to YouTube in 40 min. It is widely accepted that information such as this supplements, rather than replaces, data sourced from scientific instruments. As Earle et al. (2011, p 709) note: “The qualitative descriptions contained in the tweets are available at the same time as the seismically-derived earthquake parameters and sometimes provide a responding seismologist with a quick indication of the severity of the earthquake

effects.” Although Twitter’s 140 characters may be limiting, the narratives often produce a consensus on the intensity of the earthquake because “citizens tend to report very similar experiences” (Young et al. 2013, p. 19). Hashtags are quickly formulated to spread information. For example, within minutes of the 2011 Christchurch, New Zealand, earthquake, the hashtags #eqnz and #chch helped to share images and videos of the damage (Edmond 2013). Within 2 min of a tremor in Victoria, Australia, more than 100 tweets were posted, giving an indication of strength and reports of minor damage (Anderson 2012). In the 2011 Japan earthquake, Twitter was more effective in providing information in devastated areas than traditional media and websites (Kaigo 2012). It has led the Japanese authorities to consider making social media networks part of the country’s emergency call system (Dugan 2012). Scanning other social media platforms can also strengthen situational awareness. There are a myriad of social media platforms that can convey video, photographs, audio, and written accounts of the earthquake event. For example, the following chart outlines key benefits of Twitter, YouTube, Facebook, and blogs (Table 2).

Social Media Clutter, Misinformation, and Rumors

Social media is flooded with information in a disaster event. For example, an estimated 2.3 million tweets mentioned Haiti or the Red Cross in the 24 h following the 2010 Haiti earthquake. In Japan, more than 2000 tweets were posted every second of the day following the 2011 earthquake and tsunami (Meier 2013). Significant numbers of tweets are re-tweeted, adding to the social media “clutter” and potential spread of misinformation in the aftermath of a disaster. To make sense of the situation and to gather reliable data, a number of social mapping and analysis projects are under development. They include machine-learning approaches for classifying and extracting “informative” Twitter messages to augment situational awareness (Imran et al. 2013)

Social Media Benefits and Risks in Earthquake Events, Table 2 The benefits of various social media platforms for communicating during a disaster

Twitter (140 characters)	Instant messaging One-to-many receivers (followers) Monitoring first impressions of the shaking in real time Issuing warnings and alerts rapidly Integrating mass or interpersonal communication Initiating situational awareness from the field Channels to dispel rumors and correct information Interactive mapping Rapid updating of traditional media Alternative method of seeking aid for survivors Linking to more detailed information on other platforms Ability to “snowball” information by re-tweeting
YouTube (and others)	Broadcasting live vision from a location Providing a channel to group videos Upload vision/audio for traditional and online media consumption Facilitating updates
Facebook	Offers two-way communication Connecting people inside and outside the disaster area Providing more information than Twitter Immediate updates Delivering vision/audio/images Linking to other agencies and sources of information Accessible to anyone with an email address
Blogs	Rapid updates Allows discussion Space for more information Linking to other agencies and sources of information Providing opportunities to share stories and experiences

and volunteers, such as micromappers.com, to quickly filter social media data during a disaster using apps to tag tweets and photographs. Advances in technology are providing dividends. Another project, the Artificial Intelligence for Disaster Response, reports that 40–80 % of tweets containing disaster information can be detected automatically, with an 80–90 %

accuracy rate on whether the tweet was from an eyewitness. To obtain more structured firsthand accounts from social media users, crowdsourcing approaches are utilized by dedicated not-for-profit and government-based earthquake reporting and information sites. For example, the US Geological Survey actively seeks contributions through its “Did You Feel It” project by “asking people where they were, what they observed and what they experienced during the earthquake” (Young et al. 2013, p. 2). Earthquake-Report.com, which describes itself as the “best independent earthquake reporting site in the world,” utilizes multiple social media platforms to share firsthand accounts of earthquakes in real time along with merging data from scientific sources. Crowdsourcing questions from the impact area include location, scale of intensity, and a brief description of the experience, including damage.

Squashing Viral Rumors and False Information

People power – termed the Wikipedia effect – is a self-correcting social media. Many of the rumors and misinformation in a disaster are identified and corrected by social media users. Traditional journalists, unable to compete with the speed of social media, have assumed the mantle of “fact-checkers” to validate information. For example, fake information and images distributed on Twitter during Hurricane Sandy in 2012 were quickly ousted by other social media users and traditional media outlets. Although traditional media outlets have published inaccurate information sourced from social media, it is often quickly corrected. Twitter and Facebook were used extensively by the Queensland Police Service during the South East Queensland, Australia, floods in 2011 to identify rumors and respond with factual information. Those platforms became an important resource for traditional media. As Bruns et al. (2012, p. 8) observed: “Additionally, @QPSMedia also played a crucial role in enabling affected locals and more distant onlookers to begin the difficult process of making

sense and coming to terms with these events, even while they were still unfolding.”

Psychological Support for Survivors

Social media plays a key role in the earthquake recovery phase. Unlike the limitations of the one-way traditional media model, social media’s two-way interaction helps rebuild communities and bring together families, friends, and neighbors. Importantly, social media provides “psychological first aid” where people “reported feeling a sense of connectedness and usefulness, felt supported by others and felt encouraged by the help and support being given to people” (Taylor et al. 2012, p. 25). In the weeks and months following the 2011 Christchurch earthquake, Dabner (2011) found that online discussion provided support and information, with one participant describing it as a lifeline “that helped her (and therefore her children) cope with aftershocks by realizing normally (sic) would eventually return” (2011, p. 10). Due to the level of destruction, the role of churches was resumed through social media, with one researcher observing: “Social media was really a way for people to feel like they weren’t being forgotten or like they were part of a larger community. As far as someone sitting at home alone at 10 pm, they were not able to go out for a cuppa. That’s where social media really kicked in” (Chapman-Smith 2012).

Summary

Social media continues to rapidly evolve as a useful tool in earthquake communication. Developing technology will increase the accuracy of information from “citizen seismologists” to enhance situational awareness, improve warnings, coordinate aid and recovery, and galvanize humanitarian relief efforts. Impacted communities will increasingly turn to social media as a way to communicate lifesaving information, gather support, and empower each other in the recovery process.

Cross-References

- [Community Recovery Following Earthquake Disasters](#)
- [Emergency Response Coordination Within Earthquake Disasters](#)
- [Resilience to Earthquake Disasters](#)

References

- Abbasi M, Kumar S, Filho A, Liu H (2012) Lesson learned in using social media for disaster relief – ASU Crisis Response Game. <http://www.public.asu.edu/~huanliu/papers/SBP12Game.pdf>
- Anderson M (2012) Earthquake social media response. <http://www.cfaconnect.net.au/news/earthquake-social-media-response.html>
- Bossu R, Gilles S, Mazet-Roux G, Roussel F, Kamb L (2011) Flash sourcing, or rapid detection and characterization of earthquake effects through website traffic analysis. *Ann Geophys* 54(6):716–727
- Bruns A, Burgess J, Crawford K, Shaw F (2012) #qldfloods and @QPSMedia: Crisis Communication on Twitter in the 2011 South East Queensland Floods. Media Ecologies Project. ARC Centre of Excellence for Creative Industries & Innovation, Brisbane, Australia
- Cellan-Jones R (2008) Twitter and the China earthquake, BBC blog, 12 May 2008. http://www.bbc.co.uk/blogs/legacy/technology/2008/05/twitter_and_the_china_earthqua.html
- Centres for Disease Control (2012) Crisis and emergency communication. http://emergency.cdc.gov/cerc/pdf/CERC_2012edition.pdf
- Chapman-Smith B (2012) Social media took ‘church’ role in quakes. http://www.nzherald.co.nz/technology/news/article.cfm?buffer_share=a6e0e&c_id=5&objectid=10828401
- Civil Defence Emergency Management (2012) New Zealand ShakeOut, more than 1.3 million involved. <http://www.scoop.co.nz/stories/AK1209/S00584/new-zealand-shakeout-more-than-1-3-million-involved.htm>
- Crooks A, Croitoru A, Stefanidis A, Radzikowski K (2013) #Earthquake: Twitter as a distributed sensor system. *Trans GIS* 17(1):124–147
- Dabner N (2011) Breaking ground in the use of social media: a case study of University earthquake response to inform education design with Facebook. www.educ.utas.edu.au/users/aflick/.../wg332011_submission_12.doc
- Duffy N (2012) Using social media to build community resilience. *Aust J Emerg Manag* 27(1):40–45
- Dugan L (2012) Would you place a 911 call on Twitter? http://www.mediabistro.com/alltwitter/911-call-on-twitter_b27752
- Dutta-Bergman MJ (2006) Community participation and Internet use after September 11: complementarity in channel consumption. *J Comput Mediat Commun* 11(2):659–673
- Earle PS, Bowden D, Guy M (2011) Twitter earthquake detection: earthquake monitoring in a social world. *Ann Geophys* 54(6):708–715
- Edmond S (2013) Social media pre, during and post-earthquake. <http://blog.potentia.co.nz/social-media-quake/>
- Gowing N (2009) Skyful of lies and black swans. The new tyranny of shifting information power in crises. The Reuters Institute of Technology, Oxford
- Hui N (2013) Social media’s role in Ya’an earthquake aftermath is revealing. <http://www.chinafile.com/social-medias-role-yaan-earthquake-aftermath-revealing>
- Imran M, Elbassuoni S, Castillo C, Diaz F, Meier P (2013) Extracting information nuggets from disaster-related messages in social media. In: Proceedings of the 10th international ISCRAM conference, Baden
- Kaigo M (2012) Social media usage during disasters and social capital: Twitter and the Great East Japan Earthquake. *Keio Commun Rev* 34:19–35
- Lindsay BR (2011) Social media and disasters: current uses, future options, and policy considerations. Congressional Research Services Report 7-5700. <http://fpc.state.gov/documents/organization/174191.pdf>
- MacLeod L (2010) New media vital in breaking Haiti earthquake story. http://www.bbc.co.uk/worldservice/worldagenda/2010/01/100122_worldagenda_haiti_monitoring.shtml
- Mathewson N (2012) Social media excel after quakes. <http://www.stuff.co.nz/the-press/news/christchurch-earthquake-2011/7517088/Social-media-excel-after-quakes>
- Meier P (2012) How crisis mapping saved lives in Haiti. <http://voices.nationalgeographic.com/2012/07/02/crisis-mapping-haiti/>
- Meier P (2013) Automatically extracting disaster-relevant information. <http://irevolution.net/2013/04/01/auto-extracting-disaster-info/>
- Moore M (2008) China earthquake brings out citizen journalists. <http://www.telegraph.co.uk/news/worldnews/asia/china/1950212/China-earthquake-brings-out-citizen-journalists.html>
- Palen L (2008) Online social media in crisis events. *EDUCAUSE Q* 31(3):76–78
- Perry B, Shaw R, Takeuchi Y (2012) Utilisation of social media in the East Japan earthquake and tsunami and its effectiveness. *J Nat Disaster Sci* 34(1):3–18
- Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international WWW conference (WWW2010), Raleigh
- Starbird K, Stamberger J (2010) Tweak the Tweet: leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting. In: Proceedings of the 7th international ISCRAM conference, Seattle
- Tapia A, Moore C, Johnson N (2013) Beyond the trustworthy Tweet: a deeper understanding of microblogged

- data use by disaster response and humanitarian relief organisations. In: Proceedings of the 10th international ISCRAM conference, Baden
- Taylor M, Well G, Howell G, Raphael B (2012) The role of social media as psychological first aid as a support to community resilience building. A Facebook study for Cyclone Yasi Update. *Aust J Emerg Manag* 27(1):20–26
- Turgut P (2011) Turkey's earthquake: social media to the rescue. <http://world.time.com/2011/10/24/turkeys-earthquake-social-media-to-the-rescue/>
- Wallop H (2011) Japan earthquake: how Twitter and Facebook helped. <http://www.telegraph.co.uk/technology/twitter/8379101/Japan-earthquake-how-Twitter-and-Facebook-helped.html>
- Williams R, Williams G, Burton D (2012) The use of social media in disaster recovery. <http://extension.missouri.edu/greene/documents/PlansReports/using%20social%20media%20in%20disasters.pdf>
- Young JC, Wald D, Earle P, Shanley L (2013) Transforming earthquake detection and science through citizen seismology. Woodrow Wilson International Centre for Scholars, Washington, DC

Soil-Structure Interaction

Christos Vrettos
Division of Soil Mechanics and Foundation
Engineering, Technical University of
Kaiserslautern, Kaiserslautern, Germany

Synonyms

Dynamic soil-structure interaction; Seismic soil-structure interaction; SSI

Introduction

Seismic waves propagating through the soil impinge upon structures founded on the soil surface or embedded into it. Displacements are then produced both in the structure and in the soil. The mutual dependency of the displacements is called soil-structure interaction, abbreviated as SSI. Consequently, the motion occurring at the base of the structure is different compared to the free-field motion (motion in the absence of the

structure). Soil-structure interaction characteristics depend on several factors:

- Intensity, wavelength, and angle of incidence of the seismic waves
- Soil stratigraphy
- Stiffness and hysteretic damping of the particular soil layers
- Geometry and rigidity of the foundation
- Embedment depth of the structure
- Inertia characteristics, slenderness, and natural vibration period (eigenperiod) of the superstructure
- Presence of nearby structures

Various effects are associated to this phenomenon:

- A building founded on compliant ground has different vibrational characteristics, for example, higher natural period compared to the same building on rigid base (solid rock). The softer the soil, the larger the difference.
- A part of the vibrational energy emanating from the compliant structure foundation is transmitted into the surrounding soil through wave radiation in the unbounded soil medium and hysteretic energy dissipation. Such effects do not occur in a rigidly supported structure.
- Due to the compliance of the foundation, the motion at the foundation base contains rocking and torsional components in addition to the translational components.

The mechanisms governing soil-structure interaction can be divided into two distinct interactions: inertial and kinematic interaction.

Kinematic interaction is the deviation of the soil response from the free-field motion due to the resistance of the stiffer foundation to conform to the distortions of the soil imposed by the traveling seismic waves. It is commonly expressed in terms of frequency-dependent transfer functions relating the disturbed motion at the interface foundation/soil to the free-field motion.

Inertial interaction arises as the structure responds to the soil motion induced by kinematic interaction at the foundation level. Inertial forces

are developed in the structure being transmitted to the compliant soil. Frequency-dependent impedance functions are used to represent the stiffness of the foundation/soil system and the associated radiation damping.

The relative impact of each contribution is a function of the characteristics of the incoming waves, the foundation geometry and rigidity, and the soil conditions.

The analysis is particularly challenging due to the semi-infinite extent of the soil medium, the nonlinearity of the soil behavior, the inherent variability of the soil stratigraphy, and the dependency of the response on frequency. Several procedures of different degrees of complexity have been proposed during the past five decades. A historical overview is given by Kausel (2010). The book by Wolf (1985) provides a rigorous and comprehensive treatment of the topic including applications to seismic problems.

Even with the computational facilities available today, such analyses are associated with a major effort, both for modeling the soil-structure system and carrying out the calculation. In particular during the early design stage, parametric studies are necessary in order to assess the influence of the various parameters and optimize the system for purposes of cost estimation. This necessitates the application of simplified methods that capture the essential features of the system response. The next sections provide a brief overview with emphasis on such simplified methods.

Soil-Foundation-Structure Analysis Models

Two general approaches are commonly used for the analysis of soil-structure interaction problems.

Direct Approach

The soil and the structure are treated together in a combined analysis by modeling them using finite elements or finite differences in two or three dimensions. This offers the advantage that inelastic behavior, particularly for the soil, can be taken

into account by the step-by-step numerical integration of the equations of motion within a time-domain algorithm. A drawback is the necessity to specify the input motion at the base of the model, where it is not known a priori. Since the design seismic motion is usually given at the free surface or at outcropping rock, a deconvolution is necessary to obtain the compatible bedrock motion. Often the bedrock is located at large depths thus prohibiting the modeling of the entire soil layer, and some artificial boundary is defined at a shallower depth. The deconvolution then involves an iterative procedure. For convenience, the deconvolution is often carried out using algorithms that are based on 1D vertical shear wave propagation, thus requiring an adjustment of the model parameters in order to achieve compatible solutions between 1D and 2D analyses. Attention is further required in the selection of appropriate boundary conditions at the side boundaries of the discretized domain to avoid spurious reflections that would contaminate the results. The composite soil-structure model is finally subjected to the previously determined base rock motion, and the evolution in time of displacement and stresses is computed.

Substructure Approach

The underlying calculation method comprises three steps. At first the seismic motion acting at the foundation level is determined assuming a rigid but massless foundation. This is referred to as foundation input motion (FIM), and for an embedded structure, it will include both translational and rotational components. In the second step, the complex-valued frequency-dependent impedances for the foundation/soil system are determined. The real part of the impedance function represents a linear spring and the imaginary part, a viscous dashpot accounting for the energy radiation into the soil medium. Finally, the structure supported by the frequency-dependent springs and dashpots is subjected to the foundation input motion computed in the first analysis step.

While impedance functions are sufficient for rigid foundations, distributed springs and dashpots placed around the foundation are used for

nonrigid embedded foundations when the distribution of sectional forces is sought. In this case, due to the vertical variation of ground motion, the imposed differential ground displacements vary over the height of the basement walls.

The validity of this approach – often called superposition theorem – is shown by Kausel and Roesset (1974). The main advantage of the method is that each step can be handled independently and with different algorithms. Further, it allows an insight into the contributions from each analysis step and is particularly suitable for parametric studies.

The application of the principle of superposition requires linear behavior. Inelastic behavior is implemented using equivalent linearization by selecting soil modulus and radiation damping to correspond to the likely effective strain level the soil will experience under the specific loading. This is achieved by means of an iterative procedure. Superposition is shown to be a reasonable approximation even when inertia forces produce large strains in the vicinity of the foundation, since shear strains due to kinematic interaction effects are usually significant in deeper soil regions.

Inertial Interaction

Shallow Foundations

The illustration of the concepts is made on the basis of a simple structure-soil system being composed by a linear structure of height h , mass m , lateral stiffness k , and damping ratio β_{str} that is connected to a rigid foundation of radius r resting on the surface of a homogeneous elastic half-space. The half-space is used to represent the unbounded soil medium and is characterized by its shear modulus G , Poisson's ratio ν , and mass density ρ . Mass and moment of inertia of the foundation are neglected for simplification. The compliance of the soil is modeled by two frequency-dependent springs placed underneath the rigid foundation: a horizontal translation spring of stiffness K_x and a rotational spring of stiffness K_θ . Energy dissipation in the soil due to friction within the material (hysteretic damping)

and wave radiation in the unbounded medium is modeled by a pair of frequency-dependent dashpots with coefficients C_x and C_θ attached parallel to the respective springs. This model may be viewed as a single- or multistory building after an appropriate reduction of the degrees of freedom.

Springs and dashpots for each degree of freedom j can be condensed to complex-valued impedances that are expressed in two equivalent forms:

$$\tilde{K} = K_j + i\omega C_j = K_j(1 + i2\beta_j) \quad (1)$$

where ω is the circular frequency of the excitation, i is the imaginary unit, and β_j is a damping coefficient that is related to the viscous dashpot coefficient of a simple oscillator by

$$\beta_j(\omega) = \frac{\text{Im}(\tilde{K}_j)}{2\text{Re}(\tilde{K}_j)} = \frac{\omega C_j}{2K_j} \quad (2)$$

The use of β_j has the advantage that at resonance of the compliant system β_j corresponds to the percentage of critical damping.

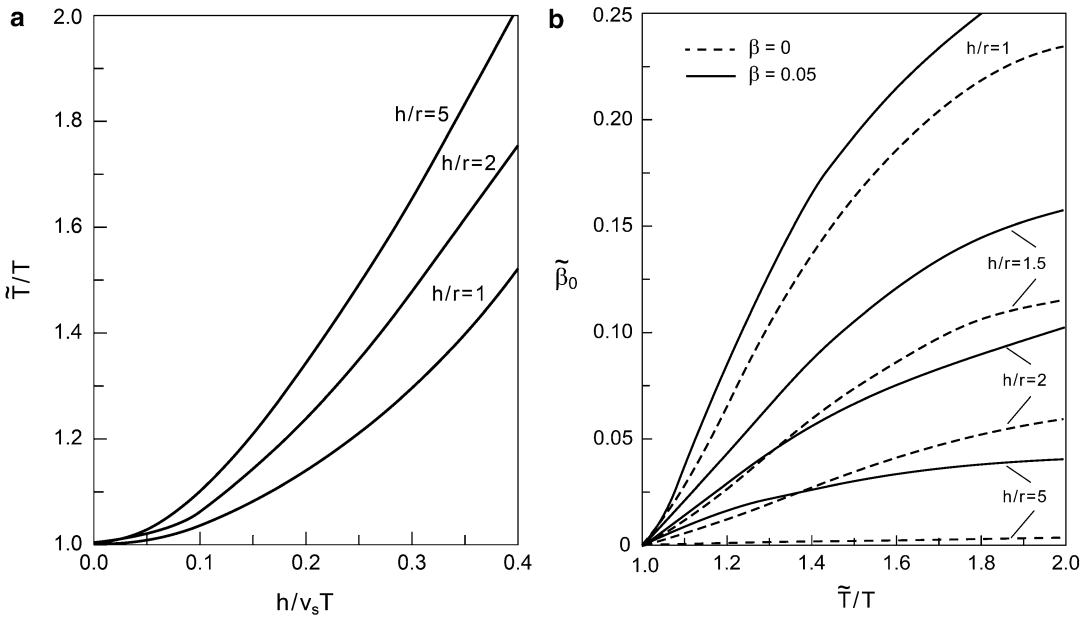
The undamped natural vibration period of the structure in its fixed-base condition is

$$T = 2\pi\sqrt{\frac{k}{m}} \quad (3)$$

For the case of a compliant base, it can be shown that the respective natural period is (Veletsos and Meek 1974)

$$\tilde{T} = T\sqrt{1 + \frac{k}{K_x} + \frac{kh^2}{K_\theta}} \quad (4)$$

Hence, the period of the flexibly supported structure is higher than that on rigid base. Since the spring stiffnesses are in general frequency dependent, an iterative procedure is necessary to evaluate the period $\tilde{T}$. A reasonable approximation consists in using the spring values corresponding to the fixed-base natural period, and even simpler is to use the static values of the springs.



Soil-Structure Interaction, Fig. 1 Effects of soil-structure interaction: (a) elongation of natural period in dependency on the ratio of structure-to-soil stiffness $\bar{s}$; (b) increase in effective damping in dependency on the natural period ratio $\tilde{T}/T$. Curves are for mass ratio $\bar{m} = 0.15$

and different values of the slenderness ratio $\bar{h}$. Poisson's ratio $\nu = 0.45$ (Adapted from Veletsos (1977). In: Hall WJ (ed) Structural and geotechnical mechanics, 1st edn, © 1977. Reprinted by permission of Pearson, Inc., Upper Saddle River, NJ)

The dimensionless parameters controlling the period lengthening are

$$\text{Stiffness ratio structure-to-soil } \bar{s} = \frac{h}{T v_s} \quad (5)$$

$$\text{Slenderness ratio } \bar{h} = \frac{h}{r} \quad (6)$$

$$\text{Mass ratio } \bar{m} = \frac{m}{\rho \pi r^2 h} \quad (7)$$

with

$$v_s = \sqrt{G/\rho} \quad (8)$$

denoting the shear wave velocity in the soil.

The stiffness ratio will be larger for stiff structural systems such as shear walls and smaller for flexible systems such as moment frames. For soil and weathered rock sites, this term is typically smaller than 0.1 for flexible systems such as moment frames and between approximately 0.1

and 0.5 for stiff systems such as shear wall and braced frame structures. The period lengthening variation with the stiffness ratio is shown in Fig. 1a for typical values of the parameters involved.

For the overall effective damping ratio of the system, several approaches have been proposed in the literature differing in the degree of approximation involved. Usually products of damping ratios are neglected as higher-order terms. The most widespread among these solutions – that also entered design codes – is that derived by Veletsos and Meek (1974). Assuming structural damping of viscous nature, the overall effective damping becomes

$$\tilde{\beta} = \beta_0 + \frac{\beta_{\text{str}}}{(\tilde{T}/T)^3} \quad (9)$$

where β_0 represents the contribution from the soil-structure interaction – being referred to as foundation damping – that includes both material

and radiation damping (Veletsos 1977). The respective expression is written here in the more general form

$$\beta_0 = \left(\frac{T}{\tilde{T}} \right)^3 \left| k \frac{\tilde{T}}{T} \left(\frac{\beta_x}{\tilde{K}_x} + \frac{\beta_\theta}{\tilde{K}_\theta} h^2 \right) \right| \quad (10)$$

From Eq. 9 it is evident that the effectiveness of the structural damping is reduced by soil-structure interaction as the period ratio $\tilde{T}/T$ increases. This may lead to a decrease in overall damping unless this reduction is compensated by the increase in the foundation damping. In practice, effective damping is taken higher than the structural damping, the value 5 % used in the development of design provisions being considered as a lower bound. Figure 1b shows the significant increase of the foundation damping with decreasing slenderness ratio h/r : rocking motion that is characterized by small radiation damping dominates the response of slender structures, whereas for squat structures the prevailing motion is horizontal translation that radiates energy into the soil more efficiently.

Observations based on data from instrumented buildings confirmed the analytical findings. For the majority of structures, the stiffness ratio h/Tv_s will be less than 0.5 and the mass ratio will range between 0.1 and 0.2 with a typical average of 0.15 (Stewart et al. 2003). The case studies analyzed revealed that the governing parameter for inertial interaction is the stiffness ratio and that these effects can be neglected for values less than 0.1.

Impedances for Shallow Foundations

Frequency-dependent springs and dashpots for shallow foundations have been determined in the last decades by several authors for different geometries and soil stratigraphies. In most cases radiation damping is expressed in terms of the dashpot coefficient C_j , as given in Eq. 1. The stiffness K_j at zero frequency is referred to as the static foundation stiffness and is denoted by K_j^0 . The effects of frequency on the spring values for the particular vibrational mode j are then given by stiffness modifiers such that

$$K_j = K_j^0 k_j \quad (11)$$

Exact closed-form solutions are available only for perfectly rigid circular foundations and relaxed boundary conditions at the soil-foundation interface, i.e., normal stresses are neglected for swaying and shear stresses for rocking. These solutions are

Horizontal translation

$$K_x^0 = \frac{8}{2 - \nu} Gr \quad (12)$$

Rocking

$$K_\theta^0 = \frac{8}{3(1 - \nu)} Gr^3 \quad (13)$$

These expressions may be used for square foundations – and also for rectangular foundations with aspect ratio less than 3 – by replacing the radius by an equivalent value that yields the same footprint area for swaying and equal moments of inertia for rocking.

A review of available solutions for foundation impedances is presented by Pais and Kausel (1988) and Gazetas (1991) and the update by Mylonakis et al. (2006). Approximate expressions and graphs are compiled for various configurations and for all six modes of vibration. They include static values for rectangular foundations, stiffness modifiers, and expressions for the radiation damping.

We restrict here the presentation of results to swaying and rocking motion for rectangular foundations with footprint area $2a \times 2b$ with $a \geq b$ with the x -axis running parallel to the longer foundation side. The subscripts θ_x and θ_y in the impedances indicate rotation around the x - and y -axis, respectively. The weak coupling between translational and rocking mode is neglected. The frequency dependency is captured by the dimensionless parameter

$$a_0 = \frac{\omega b}{v_s} \quad (14)$$

and the foundation aspect ratio is denoted by

$$\ell = \frac{a}{b} \geq 1 \quad (15)$$

The approximate expressions obtained by Pais and Kausel (1988) are displayed in the following.

Surface Foundations

The static solutions are:

$$\text{Swaying } K_x^0 = \frac{Gb}{2-v} [6.8\ell^{0.65} + 2.4] \quad (16)$$

$$\text{Swaying } K_y^0 = \frac{Gb}{2-v} [6.8\ell^{0.65} + 0.8\ell + 1.6] \quad (17)$$

$$\text{Rocking } K_{\theta_x}^0 = \frac{Gb^3}{1-v} [3.2\ell + 0.8] \quad (18)$$

$$\text{Rocking } K_{\theta_y}^0 = \frac{Gb^3}{1-v} [3.73\ell^{2.4} + 0.27] \quad (19)$$

The frequency-dependent stiffness modifiers are:

$$\text{Swaying } k_x = 1 \quad (20)$$

$$\text{Swaying } k_y = 1 \quad (21)$$

$$\text{Rocking } k_{\theta_x} = 1 - \left[\frac{0.55a_0^2}{0.6 + \frac{1.4}{\ell^3} + a_0^2} \right] \quad (22)$$

$$\text{Rocking } k_{\theta_y} = 1 - \left[\frac{(0.55 + 0.01\sqrt{\ell})a_0^2}{2.4 - \frac{0.4}{\ell^3} + a_0^2} \right] \quad (23)$$

The viscous damping coefficients accounting for radiation damping as determined from the dashpot coefficients using Eq. 2 are:

$$\text{Swaying } \beta_x = \left[\frac{4\ell}{K_x^0/Gb} \right] \left[\frac{a_0}{2k_x} \right] \quad (24)$$

$$\text{Swaying } \beta_y = \left[\frac{4\ell}{K_y^0/Gb} \right] \left[\frac{a_0}{2k_y} \right] \quad (25)$$

$$\text{Rocking } \beta_{\theta_x} = \left[\frac{(4/3)\bar{v}\ell a_0^2}{(K_{\theta_x}^0/Gb^3) \left[\left(2.2 - \frac{0.4}{\ell^3} \right) + a_0^2 \right]} \right] \left[\frac{a_0}{2k_{\theta_x}} \right] \quad (26)$$

$$\text{Rocking } \beta_{\theta_y} = \left[\frac{(4/3)\bar{v}\ell^3 a_0^2}{(K_{\theta_y}^0/Gb^3) \left[\left(\frac{1.8}{1 + 1.75(\ell - 1)} \right) + a_0^2 \right]} \right] \left[\frac{a_0}{2k_{\theta_y}} \right] \quad (27)$$

where

$$\bar{v} = \sqrt{2(1-v)/(1-2v)} \quad \bar{v} \leq 2.5 \quad (28)$$

is the ratio of compressional wave velocity to shear wave velocity in the soil.

It should be mentioned that the exact curves for the stiffness modifiers and the damping factors have in general a smooth wavy form, and the expressions given above consist approximations to these curves.

Key features of the system behavior are:

Dynamic modifiers for translational stiffness are almost unity independent of the foundation aspect ratio, whereas rocking modifiers are significantly reduced with frequency in a very weak dependency on the foundation aspect ratio.

Radiation damping for the horizontal translational mode is only modestly influenced by the direction of vibration or the foundation aspect ratio. For rocking on the other hand, the damping is strongly affected by the aspect ratio and the direction of vibration, increasing with the foundation aspect ratio. At low frequencies damping in rocking motion is smaller compared to that in horizontal translation due to interference phenomena; it only overweighs translational damping at higher frequencies and for elongated foundations when excited in the direction of the longer foundation side. Hence, translational foundation movement may often be predominant with respect to radiation damping.

Embedded Foundations

The references cited above contain also information for embedded foundations. Embedment increases static foundation stiffness. According to the review by Pais and Kausel (1988), dynamic stiffness modifiers remain largely unaffected. The dynamic analyses for obtaining such impedances usually assume a perfect contact between the soil and the basement walls, a situation that seldom occurs in reality. This yields higher damping values as observed from the actual response of buildings. A practical, conservative approach consists in considering the embedment effects only for the static stiffness and applying the dynamic modifiers of surface foundations. Alternatively, one may use the formulae given by Gazetas (1991) and by Mylonakis et al. (2006) that consider an effective height of the contact zone along the perimeter of the embedded foundation.

Soil Layering

Impedance functions for multilayered soils can only be determined with specialized software that is not easily accessible to practicing engineers. Available algorithms are mostly based on finite element procedures incorporating efficient consistent boundaries for the proper energy radiation at the domain boundaries.

A particular case constitutes a soil layer of finite depth on rock where a cutoff frequency exists, below which there is no radiation damping. The respective formulae given in the above references may be used for a two-layer system when the shear wave velocity in the top layer is less than half of that of the underlying stratum. Impedances for square foundations on uniform or nonuniform soil layer overlying a half-space are tabulated by Wong and Luco (1985).

Parameters for Soil Behavior

The expressions given above assume linear elastic or viscoelastic soil behavior. However, for moderate or strong seismic excitations, the nonlinearity of the soil must be taken into account. Hence, the values of the shear modulus entering the equations for the SSI effects must be adjusted to reflect the strain level in the ground

associated with the stipulated design ground motion. In critical projects seismic site response analyses are carried out with the soil properties being determined from special dynamic laboratory tests on undisturbed samples. First-order estimates for the strain-compatible values are given in some code provisions. Typical values as recommended by Eurocode 8, Part 5 (CEN 2004), are tabulated below in terms of their small-strain amplitude values G_0 and v_{s0} in dependency on the effective ground acceleration defined as the spectral acceleration at the plateau of the response spectrum divided by 2.5. Guide values for the hysteretic soil damping are also given.

The small-strain values of the soil shear modulus or the shear wave velocity may be determined by a variety of methods, the choice depending on the variability of the soil conditions, available knowledge on the material behavior, and the importance of the structure. These methods include: (i) empirical relationships in terms of the SPT blow count or of the tip resistance of the CPT tests, (ii) geophysical field methods based on wave propagation, and (iii) dynamic laboratory tests. An overview of the testing procedures and available design equations is summarized by Kramer (1996).

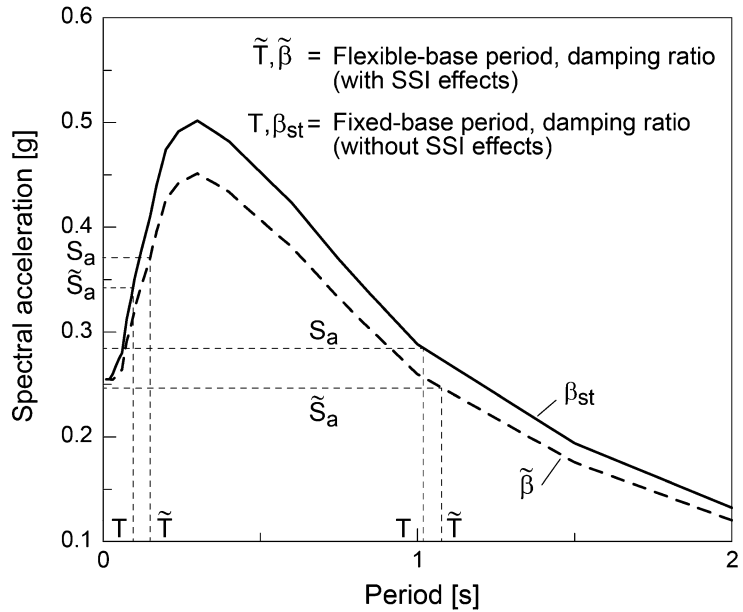
	Effective spectral ground acceleration [g]		
	0.10	0.20	0.30
G/G_0	0.80 (± 0.10)	0.50 (± 0.20)	0.36 (± 0.20)
v_s/v_{s0}	0.90 (± 0.07)	0.70 (± 0.15)	0.60 (± 0.15)
Damping ratio	0.03	0.06	0.10

Adaption in Design Codes and Implication for the Design

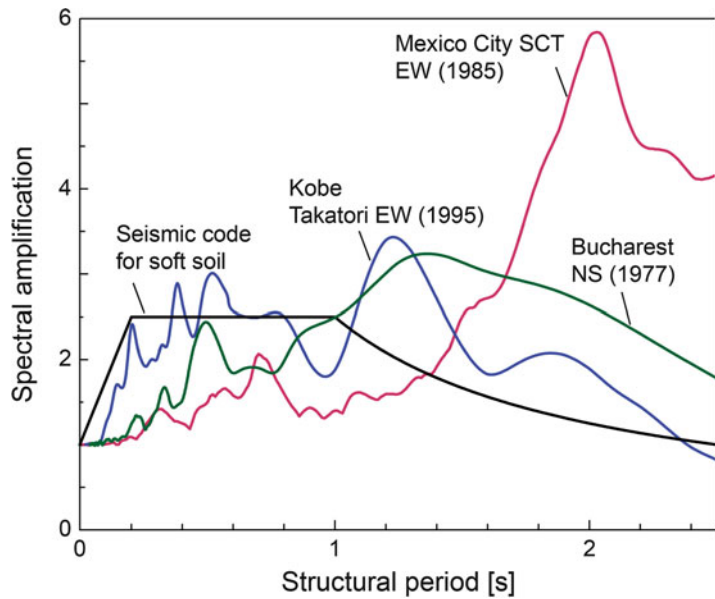
The implications of inertial SSI for design are illustrated in Fig. 2 with reference to the acceleration response spectrum used for evaluating seismic base shear forces in buildings. Idealized envelope spectra in modern codes initially increase with period, attaining a plateau value, and start decreasing monotonically after a certain period that is in the order of 0.4 to 1.0 s.

Soil-Structure

Interaction, Fig. 2 Effect of natural period elongation and foundation damping on a typical acceleration design spectrum (Adapted from Stewart et al. (2003) by permission of the Earthquake Engineering Research Institute)

**Soil-Structure**

Interaction, Fig. 3 Ratio of spectral acceleration to peak ground acceleration for 5 % structural damping for some severe earthquakes with long-period components compared to that of a typical code for soft soil (Adapted from Gazetas and Mylonakis (1998) by permission of the American Society of Civil Engineers)



For buildings with periods larger than about 0.5 s, consideration of period elongation and flexible base damping will lead to a reduction of the base shear demand. Hence, in most cases, SSI effects are neglected in the frame of conservative design.

However, there are various seismic environments with recorded response spectra exhibiting

their peak at periods greater than 1.0 s. Spectra from some prominent records are contrasted to a typical design spectrum for soil in Fig. 3. SSI phenomena in these earthquakes had detrimental effects as revealed by analyses linking site conditions and building natural periods to observed damage. In the 1985 Mexico City earthquake, for example, due to SSI effects, the natural period of

10–12 story buildings founded on soft clay was altered from about 1.0–1.5 s to nearly 2.0 s, thus coinciding with the peak of the response spectrum at the particular site. The associated phenomena are elucidated among others by Gazetas and Mylonakis (1998).

Hence, proper assessment of both the anticipated seismic input and the prevailing soil conditions is an indispensable prerequisite in any SSI analysis. In modern seismic codes the site characterization for deep soil deposits is based almost exclusively on the near-surface region of the soil (often the top 30 m), disregarding the depth of the underlying rock. The representative average shear wave velocity to this depth in this deposit is used as parameter for the classification; cf. Dobry et al. (2000).

Pile Foundations

Single Pile

Consider a pile horizontally loaded at its head at the ground surface. The deformed shape of the pile extends down to a so-called active (or effective) length below which it becomes negligible. This length depends on the pile diameter, the elastic modulus of the soil, the ratio of pile modulus to soil modulus, and the fixity conditions. Expressions for static and dynamic loading are given by Gazetas (1991). For static loads this length is of the order of 10–20 pile diameters, while for dynamic loading this length will be greater due to the wave propagation. With respect to flexural response, the pile can be modeled without significant error as an infinite-long beam when its length is greater than the active length. Two models are commonly used for the analysis: elastic continuum theory or Winkler spring models (Pender 1993).

Following the same principles as for shallow foundations, the horizontally loaded soil-pile system may be represented by three impedances corresponding to swaying, rocking, and cross-swaying-rocking. Consideration of the latter is necessary since the reference level is located at the pile head and the resultant of the reactions acts at a specific depth thus inducing a bending moment at the pile head. Expressions synthesized

from results by various authors are summarized by Gazetas (1991). The static stiffnesses are expressed in terms of the pile diameter d and Young's moduli of the soil and the pile E and E_p , respectively:

$$\text{Swaying } K_x^0 = dE \left(\frac{E_p}{E} \right)^{0.21} \quad (29)$$

$$\text{Rocking } K_\theta^0 = 0.15d^3E \left(\frac{E_p}{E} \right)^{0.75} \quad (30)$$

$$\text{Cross-swaying-rocking } K_{x\theta}^0 = -0.22d^2E \left(\frac{E_p}{E} \right)^{0.50} \quad (31)$$

The dynamic modifiers are approximately equal to unity:

$$k_x = k_\theta = k_{x\theta} \approx 1 \quad (32)$$

The expressions for the radiation damping β_j as defined by Eq. 2 are given in dependency on the dimensionless frequency

$$a_0 = \frac{\omega d/2}{v_s} \quad (33)$$

as follows:

$$\text{Swaying } \beta_x = 0.35 a_0 \left(\frac{E_p}{E} \right)^{0.17} \quad (34)$$

$$\text{Rocking } \beta_\theta = 0.11 a_0 \left(\frac{E_p}{E} \right)^{0.20} \quad (35)$$

$$\text{Cross-swaying-rocking } \beta_{x\theta} = 0.27 a_0 \left(\frac{E_p}{E} \right)^{0.18} \quad (36)$$

Pile Groups

Building foundations are always constructed as groups of piles. In evaluating the dynamic stiffness of a pile group, the interactions between the piles must be taken into consideration, just like in the case of static loading. However, the cross-interaction of individual piles is strongly

dependent on frequency, thus precluding description by simple explicit formulae. The rigorous solution methods available are based on the thin-layer method (Kaynia and Kausel 1982; Waas and Hartmann 1984). Fortunately, a remarkable simple solution procedure was discovered by Dobry and Gazetas (1988) that is straightforward to implement, thus facilitating the assessment of the associated SSI effects with a very good accuracy. The respective interaction coefficients between the piles are given in terms of pile spacing, excitation frequency, and the wave velocity through the soil between the piles. The values for the overall stiffness and damping of the pile group are then assembled using the respective values of the single piles and these interaction factors.

Kinematic Interaction

Shallow Foundations

Kinematic interaction is induced by the presence of a stiff foundation that forces the foundation motions to deviate from the free-field motions. The associated phenomena are due to (i) base-slab averaging of inclined or incoherent seismic waves, and (ii) embedment of the foundation.

Base-Slab Averaging

Seismic waves impinging at directions other than vertical arrive at different points along the foundation at different times giving rise to the so-called wave passage effects. The apparent propagation velocity of the waves is in the order of 1.5–3.5 km/s and is controlled by the wave propagation in the underlying rock. In addition to this, ground motion is in most cases inherently incoherent resulting from inhomogeneities along the travel path from the source to the site.

Studies conducted hitherto mainly address the wave passage problem that is amenable to analytical treatment. They show that the slab due to its stiffness and flexural rigidity averages the free-field displacement pattern by reducing the translational motions and at the same time introducing rotational motions. The latter include rocking in the presence of inclined SV-, P-, or

Rayleigh waves, and torsion in the presence of SH- or Love waves. The torsion of symmetrical buildings observed in earthquakes is a consequence of obliquely incident seismic waves. Further, the modification of the seismic motion depends on the frequency content of the seismic motion with high-frequency components being filtered out by the slab when the respective apparent wavelength is shorter than an effective length of the foundation slab (the diameter for circular foundations).

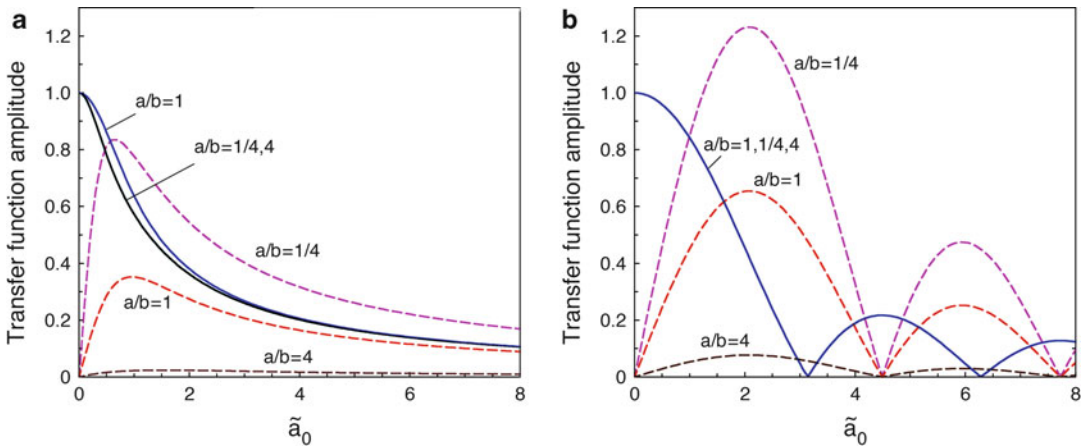
Kinematic interaction effects are expressed in terms of transfer functions relating the amplitude of the foundation input motion to that of the free-field motion. The system considered consists of a rectangular foundation with area $2a \times 2b$, $a \geq b$ that is subjected to harmonic SH waves of circular frequency ω with particle motion in the direction of the x -axis impinging on the foundation at an angle α_v with the vertical and propagating along the positive y -axis. The transfer functions derived by Veletsos et al. (1997) include both coherent and incoherent seismic motions. They are given in dependency on the dimensionless parameter

$$\tilde{a}_0 = \frac{\omega b_e}{v_s} \sqrt{\kappa^2 + \sin^2 \alpha_v \left(\frac{b}{b_e}\right)^2} \quad (37)$$

where $b_e = \sqrt{ab}$ is the half-side length of an equivalent square foundation, v_s is the shear wave velocity, and κ is a ground motion incoherence parameter. The curves shown in Fig. 4 represent the two limiting cases with $\kappa = 0$ and $\alpha_v = 0$, respectively. The transfer functions for torsional motions are referred to the foundation edge being the product of foundation half-width b and rotational angular distortion.

Lateral transfer functions are for both types of wave motion only very weak dependent on the aspect ratio a/b suggesting that the governing parameter is the foundation area. The induced torsional component, however, is very sensitive both to the aspect ratio and the type of wave motion.

Recent observations on buildings indicate that the apparent value of κ (denoted by κ_a) is nearly



Soil-Structure Interaction, Fig. 4 Amplitude of transfer functions between free-field and foundation input motion for rectangular foundations subjected to obliquely incident shear waves: (a) vertically incident, incoherent

waves; (b) non-vertically incident, coherent waves. The *solid lines* are for the horizontal motion and the *dashed lines* for the induced torsional component. Curves computed from expressions in Veletsos et al. (1997)

proportional to the small-strain shear wave velocity v_s yielding roughly $\kappa_a = 0.2$ at a typical value $v_s = 250$ m/s, Kim and Stewart (2003).

Foundation Embedment

Embedment effects result from the scattering of incoming waves. Rocking motions develop due to nonuniformly distributed tractions against the side walls. Assessment is made by means of transfer functions relating the base-slab translational and rocking motions to the free-field motions. An accurate numerical solution for cylindrical foundations subjected to coherent shear wave motions is provided by Day (1977). The embedment depth e is normalized with respect to the radius of the foundation r , and the frequency dependency is captured by the dimensionless parameter $a_0 = \omega r / v_s$. Figure 5 shows typical patterns for vertically propagating waves.

Piles

Piles embedded in a soil stratum respond to incident vertical shear waves in dependence on their flexural rigidity in relation to the stiffness of the surrounding soil. The incoming wave field is modified, the displacement at the pile head differs from that of the free field, and pile-head rotation is induced. The displacement reduction depends on the ratio of pile modulus to soil modulus, the

slenderness of the pile, and the frequency of excitation with high-frequency components being filtered out especially by relatively short, rigid piles.

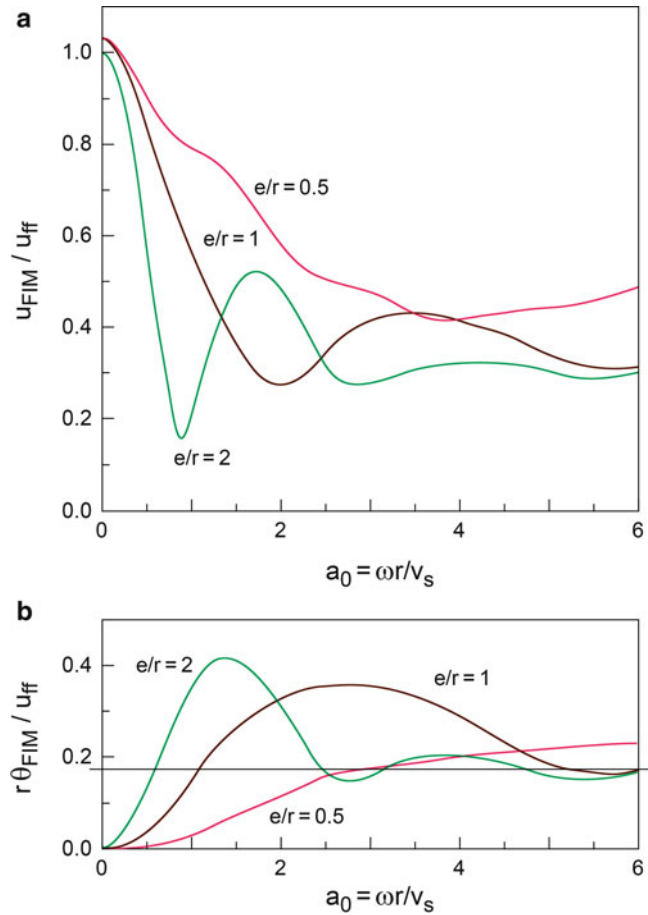
Analyses of kinematic interaction that include variable stratigraphy are carried out mostly by using the Winkler spring model. The ends of the springs and dashpots that capture SSI effects are connected to the free field where the soil response is imposed. The latter is computed independently. The accuracy of this simplified approach depends on the selection of the springs and dashpots that are obtained adopting physically justified approximations (Pender 1993).

Attention deserves the bending moment induced in the pile during the passage of seismic waves. The maximum value occurs, as expected, at soil layer interfaces, strongly increasing with the contrast in shear wave velocity between the bottom and top layer (Nikolaou et al. 2001).

The numerical study by Fan et al. (1991) using the continuum model by Kaynia and Kausel (1982) for pile groups excited by vertically propagating shear waves provides graphs showing the effects of pile rigidity relative to that of the soil, pile slenderness, pile spacing, number of piles, and pile-head fixity conditions. An idealized general shape of the frequency dependence of the kinematic response is defined in terms of a

Soil-Structure Interaction,

Fig. 5 Amplitude of transfer functions between free-field and foundation input motion for cylindrical, embedded foundation subjected to vertically incident coherent shear waves for different normalized embedment depths. (a) horizontal translation; (b) rocking component (Redrawn from Day (1977))



displacement factor relating the pile-head displacement to that of the free field. This factor is approximately unity at low frequencies with the pile closely following the ground movement; in the medium frequency range, it decreases with frequency and beyond a distinct frequency fluctuates around a constant value of 0.2–0.4. The difficulty consists in defining the transition frequencies for the particular system layout.

It must be realized though that there is no simple means for evaluating kinematic interaction for pile groups. In noncritical situations, however, this difficulty may be circumvented by neglecting kinematic interaction. This is justified by findings that kinematic effects for pile groups are similar to those for individual piles, in particular for horizontal translation and to a lesser extent for torsional and rocking vibration modes.

Concluding Remarks

Despite its inherent complexity, the theory of linear soil-structure interaction and the implications in structural performance are now well understood. Refinements, optimization, and validation studies are subjects of ongoing research. The variability in the stratigraphy of soil deposits, the nonlinearity of the soil behavior, the frequency dependency of the response, and the limited availability of specialized software for the analysis make the proper assessment of the SSI effects still a difficult task, requiring physical insight when applying such concepts. It should be self-evident that the effective implementation in an integrated structural design asks for a close collaboration between structural and geotechnical engineers.

Summary

The main effects of soil-structure interaction on the seismic response of structures founded on compliant ground are presented. The modeling concepts to capture the associated modification of the building natural period and the energy dissipation due to radiation damping are highlighted by reference to relatively simple structures. Both kinematic and inertial actions are treated. Available expressions for the dynamic impedance functions are summarized both for shallow foundations and piles. A brief account is given of the implications in seismic design provisions for buildings.

Cross-References

- [Building Codes and Standards](#)
- [Earthquake Response Spectra and Design Spectra](#)
- [Engineering Characterization of Earthquake Ground Motions](#)
- [Response Spectrum Analysis of Structures Subjected to Seismic Actions](#)
- [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)
- [Selection of Ground Motions for Response History Analysis](#)
- [Site Response: 1-D Time Domain Analyses](#)
- [Substructuring Methods for Finite Element Analysis](#)

References

- CEN-European Committee for Standardization (2004) Eurocode 8: design of structures for earthquake resistance part 5: foundations, retaining structures and geotechnical aspects
- Day SM (1977) Finite element analysis of seismic scattering problems. PhD thesis, University of California, San Diego
- Dobry R, Gazetas G (1988) Simple method for dynamic stiffness and damping of floating pile groups. *Géotechnique* 38:557–574
- Dobry R, Borchardt RD, Crouse CB, Idriss IM, Joyner WB, Martin GR, Power MS, Rinne EE, Seed RB (2000) New site coefficients and site classification system used in recent building seismic code provisions. *Earthq Spectra* 16(1):41–67
- Fan K, Gazetas G, Kaynia A, Kausel E (1991) Kinematic seismic response of single piles and pile groups. *J Geotech Eng, ASCE* 117(12):1860–1879
- Gazetas G (1991) Foundation vibrations. In: Fang HY (ed) *Foundation engineering handbook*, 2nd edn. Van Nostrand Reinhold, New York, pp 553–593
- Gazetas G, Mylonakis G (1998) Seismic soil structure interaction: new evidence and emerging issues. In: *Geotechnical earthquake engineering and soil dynamics III*, ASCE, Reston, pp 1119–1174
- Kausel E (2010) Early history of soil-structure interaction. *Soil Dyn Earthq Eng* 30(9):822–832
- Kausel E, Roesset JM (1974) Soil-structure interaction problems for nuclear containment structures. In: *ASCE power division specialty conference, electric power and the civil engineer*, Boulder, pp 469–498
- Kaynia AM, Kausel E (1982) Dynamic stiffness and seismic response of pile groups. In: *Research report R82-03*, Massachusetts Institute of Technology
- Kim S, Stewart JP (2003) Kinematic soil-structure interaction from strong motion recordings. *J Geotech Geoenviron Eng, ASCE* 129(4):323–335
- Kramer SL (1996) *Geotechnical earthquake engineering*. Prentice Hall, Upper Saddle River
- Mylonakis G, Nikolaou A, Gazetas G (2006) Footings under seismic loading: analysis and design issues with emphasis on bridge foundations. *Soil Dyn Earthq Eng* 26:824–853
- Nikolaou S, Mylonakis G, Gazetas G, Tazoh T (2001) Kinematic pile bending during earthquakes: analysis and field measurements. *Géotechnique* 51(5):425–440
- Pais A, Kausel E (1988) Approximate formulas for dynamic stiffnesses of rigid foundations. *Soil Dyn Earthq Eng* 7(4):213–227
- Pender M (1993) Aseismic pile foundation design analysis. *Bull NZ Nat Soc Earthq Eng* 26(1):49–160
- Stewart JP, Kim S, Bielak J, Dobry R, Power MS (2003) Revisions to soil-structure interaction procedures in NEHRP design provisions. *Earthq Spectra* 19(3):677–696
- Veletsos AS (1977) Dynamics of structure-foundation systems. In: Hall WJ (ed) *Structural and geotechnical mechanics*. Prentice-Hall, Englewood Cliffs, pp 333–361
- Veletsos AS, Meek JW (1974) Dynamic behavior of building-foundation systems. *Earthq Eng Struct Dyn* 3:121–138
- Veletsos AS, Prasad AM, Wu WH (1997) Transfer functions for rigid rectangular foundations. *Earthq Eng Struct Dyn* 26(1):5–17
- Waas G, Hartmann HG (1984) Seismic analysis of pile foundations including pile-soil-pile interaction. In: *Proceedings, 8th World conference on earthquake engineering*, San Francisco, vol V, pp 55–62
- Wolf JP (1985) *Dynamic soil-structure interaction*. Prentice-Hall, Englewood Cliffs
- Wong HL, Luco JE (1985) Table of impedance functions for square foundations on layered media. *Soil Dyn Earthq Eng* 4(2):64–81

Source Characterization for Earthquake Early Warning

Aldo Zollo, Gaetano Festa, Antonio Emolo and Simona Colombelli

Department of Physics, University of Naples Federico II, Federico II – AMRA S.c.ar.l, Napoli, Italy

Synonyms

Earthquake early warning; Earthquake ground motion; Earthquake source observation; Real-time location; Real-time magnitude

Introduction

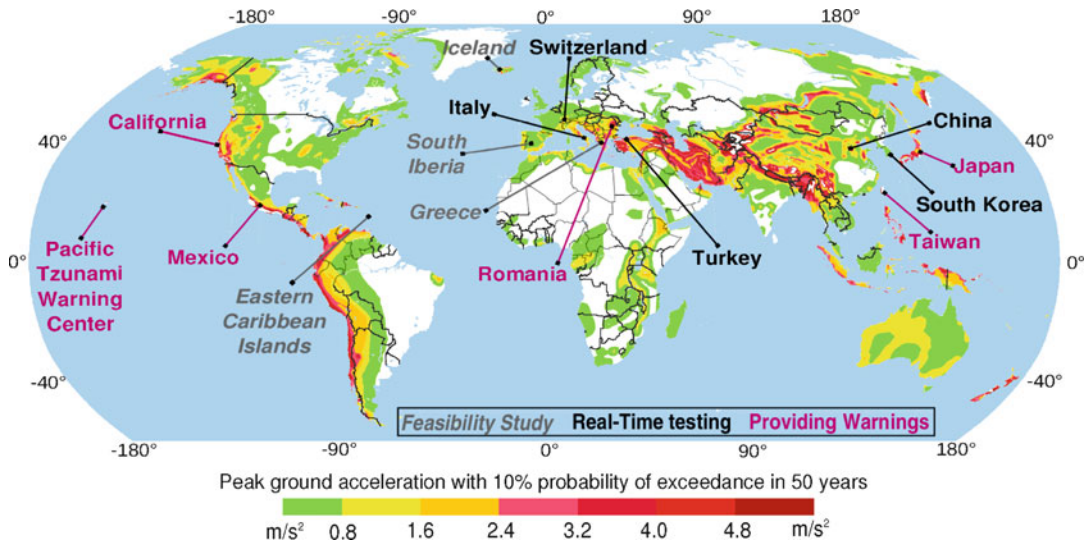
Earthquake Early Warning Systems (EEWS) are real-time, seismic monitoring infrastructures that are able to provide a rapid notification of the potential damaging effects of an impending earthquake. This objective is achieved through the fast telemetry and processing of data from dense instrument arrays deployed in the source region of the event of concern (regional EEWS) or surrounding/at the target infrastructure (front-detection or site-specific EEWS).

A regional EEWS is based on a dense sensor network covering a portion or the entire area that is threatened by earthquakes. The relevant source parameters (event location and magnitude) are estimated from the early portion of recorded signals (initial P-waves) and are used to predict, with a quantified confidence, a ground-motion intensity measure at a distant site where a target structure of interest is located. Site-specific (or on-site) EEWS consist of a single sensor or an array of sensors deployed in the proximity of the target structure that is to be alerted and whose measurements of amplitude and predominant period on the initial P-wave motion are used to predict the ensuing peak ground motion (mainly related to the arrival of S- and surface waves) at the same site. Front-detection EEWS is essentially a variant of the on-site approach, where a

barrier-shaped, accelerometric network is deployed between the source region and the target site to be protected. The alert is issued when two or more nodes of the array record a ground acceleration amplitude larger than a default threshold value. For typical regional distances, the peak acceleration at the barrier nodes is expected to be associated with the S-wave train, so that the distance between the network and the target is set to maximize the lead time (i.e., the time available for warning before the arrival of strong ground shaking at the target sites), which is, in this case, the travel time of S-waves from the barrier to the target site.

EEWS have experienced a very rapid improvement and a wide diffusion in many active seismic regions of the world in the last three decades (Fig. 1). They are operating in Japan, Taiwan, Mexico, and California. Many other systems are under development and testing in other regions of the world such as in Italy, Turkey, Romania, and China. Most of these existing EEWS essentially operate in the two different configurations described above, i.e., regional and on-site, depending on the source-to-site distance and on the geometry of the considered network with respect to the source area. The “front-detection” EEWS such as the barrier-type, Seismic Alert System (Espinosa-Aranda et al. 2011) for Mexico City can be particularly advantageous when the only potential seismic sources are at some distance from the strategic target to be protected.

The regional EEWS approach is based on the detection of the initial P-wave signal at a number of near-source stations, typically 4 to 6. Several methodologies have been proposed for the real-time estimation of the earthquake location and magnitude and are now implemented in the EW algorithms, such as ElarmS (Allen et al. 2009), Virtual Seismologist (Cua et al. 2009), and PRESTo (Satriano et al. 2010) presently running in California, Switzerland, and Southern Italy, respectively. In the framework of EU REAKT (Strategies and Tools for Real-time Earthquake Risk Reduction, FP7:ENV2011.1.3.1-1) and international collaboration projects, testing of PRESTo early warning platform is performed in



Source Characterization for Earthquake Early Warning, Fig. 1 The map shows the distribution of Earthquake Early Warning Systems around the world, with a color indicating the status of the system. In *purple*, the

operative systems, which are providing warnings to public users. In *black*, the systems which are currently under real-time testing. *Gray* color is finally used for those countries where feasibility studies are currently being doing

Romania, Greece, Turkey, Spain, and South Korea. The real-time magnitude estimation is generally inferred from the measurement of peak displacement amplitude and/or the predominant period measured in the first few seconds of the recorded P-signal, typically 3–4 s. Although the saturation of the P-wave parameters has been observed for $M > 6.5$ –7 earthquakes, several methodologies making use of longer time windows of the P-wave and/or the S-wave to update magnitude estimates have been shown to be efficient in minimizing the problem of magnitude underestimation (Colombelli et al. 2012b). The source location and magnitude estimations, which are continuously updated by adding new station data, as the P-wave front propagates through the regional EW network, are then used to predict the severity of ground shaking at sites far away from the source, by using regional-specific, ground-motion prediction equations.

The on-site early warning approaches are generally aimed at estimating the expected peak ground shaking, associated with S- and surface waves, directly from the recorded early P-wave signal. This is achieved through the use of empirical regressions between measurements

performed on the initial P-wave signal and the final peak ground motion. Wu and Kanamori (2005) first showed that the maximum amplitude of a high-pass filtered vertical displacement, measured on the initial 3 s of the P-wave (named P_d), can be used to estimate the peak ground velocity (PGV) at the same site, through a power-law relationship. The main advantage is that this relationship does not require an independent estimate of the magnitude as for regional EEWs. Although initially observed for near-source records (distances < 30 km), further analyses on independent datasets have confirmed that $\log PGV$ vs. $\log P_d$ scaling is still valid at relatively large distances (distances < 300 km) (Zollo et al. 2010; Colombelli et al. 2012a). Most of the currently operating on-site EEWs are threshold-based, alert methodologies: the alert is issued as the measured initial P-wave peak amplitude overcomes a given threshold which is arbitrarily set according to the predicted S-wave peak ground-motion amplitude. Since small magnitude earthquakes may have very large amplitudes driven by high-frequency spikes, such a basic threshold system can produce frequent false alarms. A more robust approach is to combine

the P-wave peak (which scales with distance and magnitude) and P-wave predominant period (which scales with the magnitude), into a single proxy to be used for on-site warning (Wu and Kanamori 2005). Following this idea, Zollo et al. (2010) and Colombelli et al. (2012a) have proposed a threshold-based EW method based on the real-time measurement of the period (τ_c) and peak displacement (P_d) parameters at stations located at increasing distances from the earthquake epicenter. The measured values of early warning parameters are compared to threshold values, which are set for a given minimum magnitude and instrumental intensity. At each recording site an alert level is assigned based on a decisional table with four levels defined upon threshold values of the parameters P_d and τ_c . Given a real-time, evolutionary estimation of earthquake location from first P arrivals, the method also furnishes an estimation of the extent of potential damage zone as inferred from continuously updated averages of the period parameter and from mapping of the alert levels determined at the near-source accelerometer stations.

P-wave-based, regional, and on-site EW methods can be integrated in a unique alert system (as actually done, e.g., in the new version of PRESTo, e.g., PRESTo Plus, Zollo et al. 2014), which can be used in the very first seconds after a moderate-to-large earthquake to determine the earthquake location and magnitude and to map the most probable damaged zone, using data from receivers located at increasing distances from the source.

Methodologies for regional earthquake early warning assume a point-source model of the earthquake source and isotropic wave amplitude attenuation. These assumptions may be inadequate to describe the earthquake source of large earthquakes and wave amplitude attenuation effects, and they can introduce significant biases in the real-time estimation of earthquake location and magnitude. This issue is critically related to the EEWS performances in terms of expected lead time and of uncertainties in predicting the peak ground motion at the site of interest. Within this context, new developments have been

proposed, such as the strategy of expanding the P-wave time window for the real-time signal processing, the 2D mapping of the potential damage zone, and the use of continuous GPS measurements and methodologies to estimate fault rupture extent in real time by classifying stations into near source and far source. These innovative aspects of early warning will be discussed in the present review, with a specific focus on methods for rapid and reliable source characterization for early warning applications.

Methodology

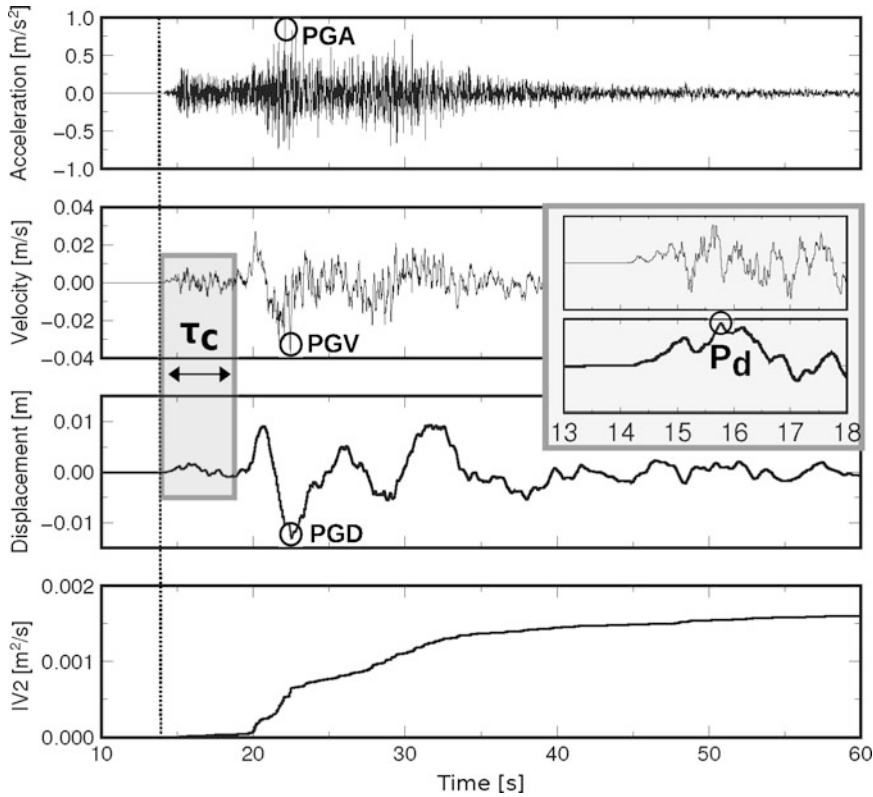
Point-Source Characterization for an Earthquake Early Warning System

In EEWS the strong motion is generally synthesized by a single parameter (in most cases the peak ground velocity, PGV , the peak ground acceleration, PGA , or the peak ground displacement, PGD , Fig. 2), which can be directly related to the damage that a building or an infrastructure may undergo because of the earthquake. Two possible approaches can be explored for the prediction/estimation of ground-motion parameters at a given site.

A first possibility is to relate the ground shaking to simplified macroscopic description of the source, yielding ground-motion prediction equations. In such a case, indicating with PGX the selected ground-motion parameter, the simplest attenuation relationship relates the logarithm of PGX with the earthquake-to-site distance R and the magnitude M :

$$\log PGX = f(M, R) \quad (1)$$

By definition of magnitude, a linear function of M is in most cases adequate to describe the influence of the earthquake size on the ground motion. The distance effect is instead accounted for by two terms describing the decay of the amplitude owing to geometrical spreading and inelastic processes within the upper crust. Nevertheless, more complex ground-motion prediction equations may be used containing high-order terms, focal depth dependence, and site effects. It is worth to



Source Characterization for Earthquake Early Warning, Fig. 2 Schematic illustration of early warning parameters. From *top to bottom*, an example of the vertical component of acceleration, velocity, displacement, and integral of squared velocity (IV2) signals. The ground-motion parameters PGA, PGV, and PGD are measured as the absolute maximum along the signal, using acceleration, velocity, and displacement records, respectively.

The *gray small box* shows a zoom on the first few seconds of P-wave on velocity (*top*) and displacement (*bottom*) records. The initial peak displacement (P_d) is measured as the absolute maximum of displacement waveform on the early portion of P-wave (typically 2–4 s) while the period parameter τ_c is measured from the ratio between initial displacement and velocity waveforms in the same time window

note that while magnitude is an ensemble measure for earthquake size, the definition of the distance requires a specific metric, which is sensitive to the ratio between the source-to-site distance and the earthquake size. At distances significantly larger than the source size, a point-source approximation for the earthquake can be generally assumed, and R refers to the epicentral or the hypocentral distance. In the near-source range, instead, finite-fault effects may be relevant and a different metric for the distance could be required. Anyhow, whatever choice of attenuation relationships and distance metrics are done, the ground-motion prediction requires the knowledge of earthquake location and size. This model

is used by regional early warning systems, for which the characterization of the source is performed by a network installed in the source vicinity.

A complementary approach is based on empirical relationships between a ground-motion parameter P_y measured in the early portion of the P-wave train and the final PGX . This is physically grounded on the first-order approximation that $\log P_y$ has the same magnitude and distance dependence of $\log PGX$, being differences concentrated only on static and possibly frequency-dependent effects. In such a case, estimation of source parameters is hidden in the common dependence and the uncertainty may be

significantly reduced by avoiding the estimation of magnitude and distance. Additionally, this approach does not require a seismic network to constrain the source parameters and can be efficiently used for single stations. Such a model is the one implemented in on-site early warning systems.

Finally both models can be combined together, with source parameters estimated by a regional network and ground motion locally verified at a specific site.

Real-Time Location

A main concern for any early warning system is the reliable estimation of earthquake hypocenter in real time. Recently, Satriano et al. (2008) have developed an evolutionary approach aimed at constraining the earthquake location, which starts when the first station is triggered by the seismic event and is updated as time passes. This technique is based on the equal differential time (EDT) formulation and provides a probabilistic density function for the earthquake location in a three-dimensional space accounting, at each time step, for information from both triggered and not-yet-triggered stations. In particular, with a single, initial, recorded arrival time, the hypocenter has to belong to the Voronoi cell containing the triggered station, which is created using the travel times to the not-yet-triggered stations. As more and more stations identify the seismic event, the location is constrained in the volume defined by the intersection of the Voronoi cells for the remaining not-yet-triggered stations and the EDT surfaces evaluated for all pairs of already-triggered stations.

Let us consider a seismic network (Fig. 3a) with N operational recording stations $S_1, \dots, S_N$ and a gridded volume V containing both the network and the earthquake source. Preliminarily, we computed the travel times from each grid point (i, j, k) in the volume V to each station of the network. If the earthquake hypocenter is at the node (i^*, j^*, k^*) of the searching grid, the classical EDT formulation prescribes that the difference between the theoretical travel times ttt_m and ttt_n from the event source to two stations S_m and S_n of the network is exactly equal to the difference

between the observed arrival times ot_m and ot_n at the same stations, since they share the same earthquake origin time:

$$(ttt_m - ttt_n)_{i^*, j^*, k^*} = ot_m - ot_n, \quad \text{with } m \neq n \quad (2)$$

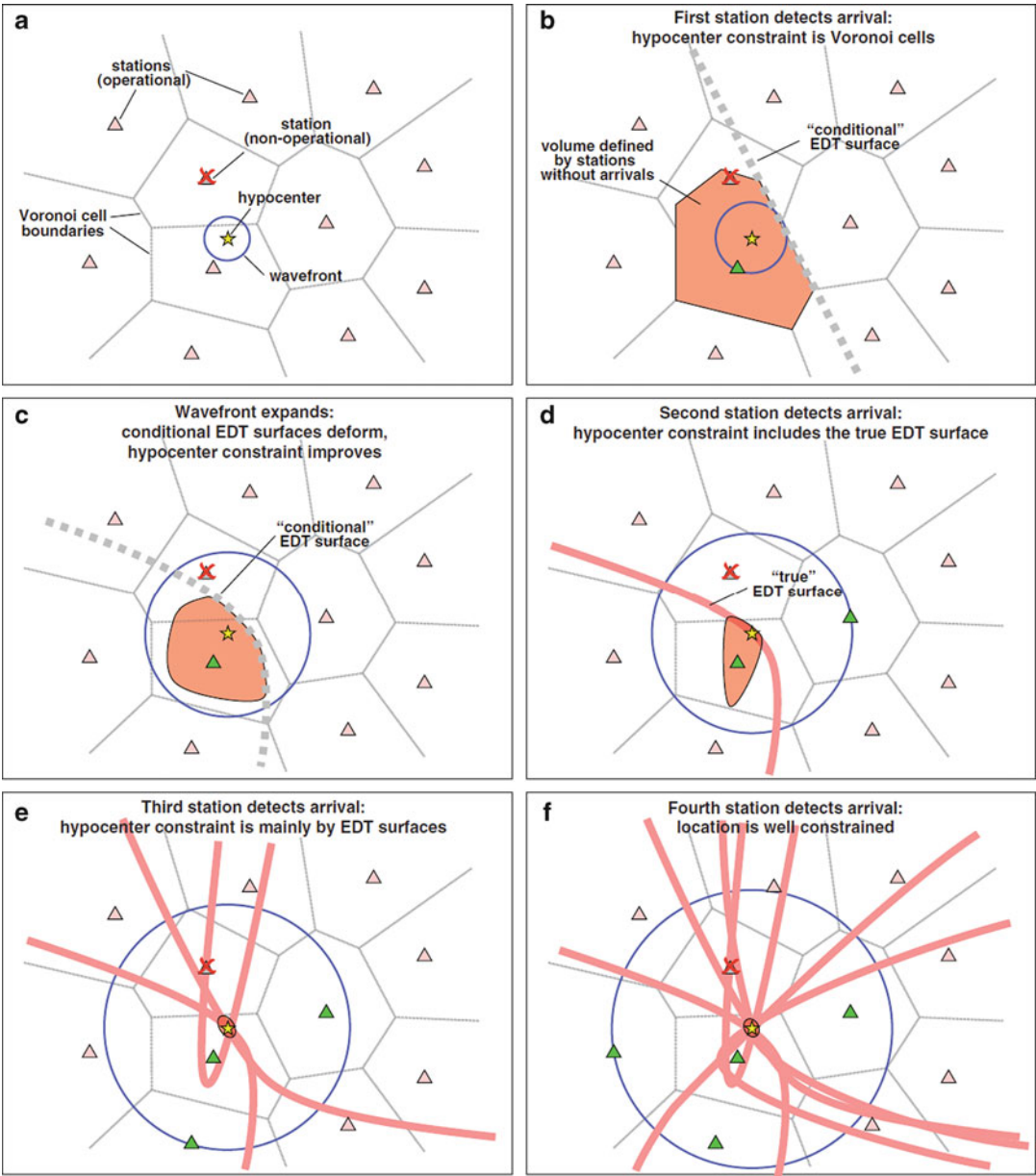
If a homogeneous velocity model is assumed, the previous equation defines a 3D hyperbolic surface whose symmetry axis passes through the two stations. With N_T triggered stations, we obtain $N_T(N_T - 1)/2$ surfaces and the hypocenter is then located at the region crossed by the maximum number of EDT surfaces.

This description is modified in the evolutionary approach introduced by Satriano et al. (2008) in which, at each time step, EDTs are evaluated not only for pairs of triggered stations but also for those pairs where only one station has already triggered. This means that when the first station, let us say S_n , is triggered by the earthquake at the time t_n , it is immediately possible to limit the hypocentral position (Fig. 3b) from the EDT surfaces defined considering that any operational but not-yet-triggered receiver S_l will identify the seismic arrival at a later time t_l (with $t_l \geq t_n$) such that

$$(ttt_l - ttt_n)_{i^*, j^*, k^*} \geq 0, \quad \text{with } l \neq n \quad (3)$$

The equality in the previous equation corresponds to the set of points in the volume of interest for which the travel time at the first station is equal to the travel time at any remaining, not-yet-triggered, receivers (conditional EDT). As a consequence, the inequality delimitates the region in the volume V , bounded by the conditional EDT, that must contain the earthquake hypocenter. In the case of a homogeneous propagation medium, the region bounded at the time t_n by the conditional EDTs evaluated for all the stations corresponds to the Voronoi cell for the receiver S_n . As the time goes on, the source volume is bounded by the system of equations

$$(ttt_l - ttt_n)_{i^*, j^*, k^*} \geq t_{\text{cur}} - t_n, \quad \text{with } l \neq n \quad (4)$$



Source Characterization for Earthquake Early Warning, Fig. 3 Sketch of the evolutionary earthquake location approach. For simplicity of the presentation, it is shown only a plane view of the epicentral location. (a) The Voronoi cells associated with each station (operational and not) of the network are a priori known. (b) As the first station triggers, it is possible to identify a volume that is bounded by the conditional EDT surface on which the travel time for the first receiver is equal to the travel times for each not-yet-triggered station. This volume is likely to contain the true earthquake location. (c) As the time passes, further information from station that

has not yet triggered is gained and the EDT surface bends around the first triggered station. As a consequence, the source volume decreases. (d) As the second station triggers, a true EDT surface is defined. The intersection between this surface and the conditional EDT surfaces limits the likely source volume whose dimension is continuing to decrease. (e) Two more true EDT surfaces become available when the third station triggers, thus better constraining the earthquake location. (f) As more and more stations trigger, the earthquake location converges to the standard EDT solution which is based on true EDT only (After Satriano et al. (2008))

provided that, at the current time t_{cur} , only the station S_n has triggered at the previous time t_n . As t_{cur} increases, the hypocentral volume becomes smaller since the conditional EDT surfaces fold around the station S_n (Fig. 3c).

It is possible to prescribe a probability density function (pdf) for the hypocentral volume associated with each inequality in the previous equations and for each grid point. We assign a value $p_{n,l}(i, j, k) = 1$ if the inequality is satisfied and a value $p_{n,l}(i, j, k) = 0$ if it is not. Summing up over all the stations, we obtain, for each grid point, a non-normalized PDF $P(i, j, k)$ whose maximum value is $(N-1)$ for those grid points for which all inequalities in the equation are satisfied.

When the second station (and, as the time progresses, further receivers) triggers, the equation is evaluated for all the possible pairs of triggered and not-yet-triggered stations. Then, true EDT surfaces are defined for each pair (n, m) of triggered stations computing, for each grid point, the quantity

$$q_{n,m}(i, j, k) = \exp \left\{ \frac{[(ttt_n - ttt_m)_{i,j,k} - (ot_n - ot_m)]^2}{2\sigma^2} \right\},$$

with $n \neq m$

(5)

where the expression in square brackets corresponds to the standard EDT and σ represents the uncertainty associated with the arrival time picking and travel-time computation. The quantity $q_{n,m}(i, j, k)$ varies between 0 and 1. We then sum the $q_{n,m}$'s with the updated $p_{n,l}$'s to obtain a new $P(i, j, k)$ which now has a maximum value equal to

$$P_{\text{max}} = (N - N_T)N_T - \frac{N_T(N_T - 1)}{2} \quad (6)$$

being N_T the number of triggered stations. Once we have evaluated P , we can define the function

$$Q(i, j, k) = \left(\frac{P(i, j, k)}{P_{\text{max}}} \right)^N \quad (7)$$

which varies in the range $[0, 1]$ and represents the PDF for hypocentral location at the grid cell (i, j, k) .

As the time increases and more and more stations trigger, the number of not-yet-triggered stations becomes smaller and smaller, and the earthquake location converges toward the hypocenter which would be obtained through the standard EDT approach using data from all operational stations of the network (Figs. 3d–f). Tests performed on both synthetic and real data have shown that when a dense seismic network (i.e., with a mean station spacing of about 10 km) is deployed around the fault zone, a location accuracy is achieved within 1–3 s after the first arrival detection.

Real-Time Magnitude Estimation and Earthquake Rupture

Magnitude estimation for early warning applications is based on empirical relationships relating the earthquake size with parameters measured in the early portion of the P- and S-wave trains. These parameters are generally associated to the low-frequency content of the data, which is sensitive to the seismic moment, and can be related to the maximum amplitude, the dominant frequency, or the energy released by the event. Associated proxies are peak values, predominant period and integrated measurements, respectively. Several authors showed that the initial portion of recorded P-waves carries information about the event magnitude, both through its frequency content and amplitude (Allen and Kanamori 2003; Kanamori 2005; Zollo et al. 2006; Wu and Zhao 2006; Böse et al. 2007; Wu and Kanamori 2008). A review of the common used parameters is detailed in the next section; here we want to note that peak and energy estimates are dependent on the source-to-site distance, while predominant period is pretty insensitive to the epicentral location (Allen and Kanamori 2003). In all cases, when dealing with dense regional networks surrounding the fault that generated the earthquake, location is generally available before or at the same time the first estimates of magnitude are performed. Therefore, it is not a

disadvantage to use distance-dependent parameters as compared to distance-independent ones. All of the parameters are estimated from measures in the early portion of the signal: with the goal of issuing an early warning for earthquakes with magnitude larger than 5.5, the time scale in which performing the measurements is few seconds (2–4 s generally).

The definition of a time window t_0 for the measurements corresponds to image a specific area on the fault plane, delimited by the corresponding isochrone. It is defined as the set of points whose radiation arrives at a given station at the same time t_0 . Hence, the portion of the fault enlightened by the first few seconds of signal depends both on the relative location of the station as compared to the fault and on the phase, either P or S. Specifically, for the same time window, S-waves image a larger area on the fault, because their speed is closer to the rupture speed as compared to the P-waves, and the regions explored by the different stations do not overlap as much as P-waves. We then expect that inclusion of early portion of S-waves may significantly constrain the estimation of the earthquake magnitude, as compared to the only use of P-waves. This approach is possible for close stations (with epicentral distances smaller than 30 km) for which the analysis of S-waves does not significantly affect the lead time for early warning. Additionally, isochrone mapping enhances that few seconds of P/S-waves correspond to an earthquake size of magnitude 6–6.5. This indicates that the early portion of the signal captures almost the whole rupture process up to this magnitude, while effective prediction is possible beyond that limit. Using a kinematic description of the rupture, we can argue that early warning parameters up to magnitude 6–6.5 change because both the ruptured area and the total average slip increase. Beyond that threshold, these parameters image almost the same portion of fault and any change with magnitude is only ascribed to any increase of the average slip in that region.

From this consideration we can retain that scaling of early warning parameters may be different for different magnitude ranges, and a

two-slope behavior is expected around the deterministic threshold. Hence, the use of a single relationship for a broad magnitude range should be statistically checked to avoid under/over estimations of the magnitude at the limits of the investigated range. Additionally, if any scaling occurs for events with magnitude larger than 6, this indicates a different initiation process for earthquakes that reach different sizes. Specifically, since the scaling of parameters is observed up to magnitude 7.5, this implies that when we look at the earthquake rupture on a given space scale characteristic of a magnitude 6.0 event, we can probabilistically forecast if this rupture will soon stop or it will grow up to a larger space scale. Beyond that limit, standard regression laws saturate and a different approach is required to capture the earthquake size for very big-size events. As a final comment, since the parameters have a different sensitivity to the slip increase, the uncertainty associated to the magnitude estimation may change. Generally, peak parameters are more sensitive to slip changes, the slope of the scaling is larger, and the uncertainty in the magnitude estimation is smaller as compared to predominant period estimations.

We now discuss a real-time approach for magnitude estimation based on a probabilistic evolutionary approach (Lancieri and Zollo 2008). This approach generalizes standard averages coming from single-station estimates of magnitude and can include also a priori probability density functions for magnitude characterization. As discussed before, we assume a linear scaling between the logarithm of the early warning parameter Py and the final magnitude

$$\log Py^{\text{theo}} = A + BM + C(R/R_0) \quad (8)$$

where C is a function of the distance R and R_0 a reference distance. Coefficients are generally different for P- and S-waves. For low-frequency estimators that depend on the distance, the predominant effect comes from geometrical spreading and we can assume $C(R) = K \log R$. After defining the duration of the window for P- and S-waves, the computation of Py is performed by isolating the corresponding time windows after

the P- and S-wave arrivals. P-wave arrival is automatically picked, while S-wave arrival is computed using a theoretical travel-time database derived by a 1D or 3D velocity model suitable for the area of investigation. Since S-wave estimates are significantly different from P-wave estimates, we need to be sure that the selected P-wave window is not contaminated by the following S-wave, to avoid any bias in the magnitude estimation. If the S-wave arrival is expected in the P-wave window, that P-wave measurement is discarded. For a single P_y measurement at a given station and for a given phase, the magnitude can be estimated using the Bayes theorem:

$$P(M|P_y) = \frac{P(P_y|M)P_a(M)}{\int_M P(P_y|M)P_a(M)} \quad (9)$$

In the above formula the probability $P_a(M)$ represents the a priori probability on the magnitude, the probability $P(P_y|M)$ comes from the regression relationship, and, assuming a Gaussian distribution for $\log(P_y)$, this writes

$$P(P_y|M) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left[-\frac{1}{2\sigma^2} (\log P_y - \log P_y^{\text{theo}})^2 \right] \quad (10)$$

(σ is the standard deviation of the fit). Finally the denominator is a normalization factor. This approach can be directly plugged into a real-time evolutionary magnitude estimation where the probability is updated as new information comes from additional stations or phases. Without any information, we assume an a priori magnitude distribution coming from the long-term earthquake catalog inspection (generally a Gutenberg-Richter distribution). To build up an evolutionary scheme, assuming to have computed the probability density function for the magnitude after $n-1$ measurements of P_y , this function is then used as the a priori distribution for the n th measurement. It is worth to note that this scheme allows accounting for magnitude saturation. In this case the probability

distribution $P(P_y|M)$ is assumed uniform beyond the magnitude saturation threshold.

Parameters for Magnitude Estimation

In the context of real-time applications, different amplitude and period parameters have been proposed to get independent estimates of the earthquake size. We present here an overview of the parameters used by the different systems for the magnitude estimation.

Following an original idea of Nakamura (1988), Allen and Kanamori (2003) first proposed the use of a period parameter, measured on the first few seconds of P-wave signal, to infer the size of the ongoing earthquake. The predominant period is measured on the vertical component of velocity record and is defined as

$$T_i^p = 2\pi \sqrt{X_i/D_i} \quad (11)$$

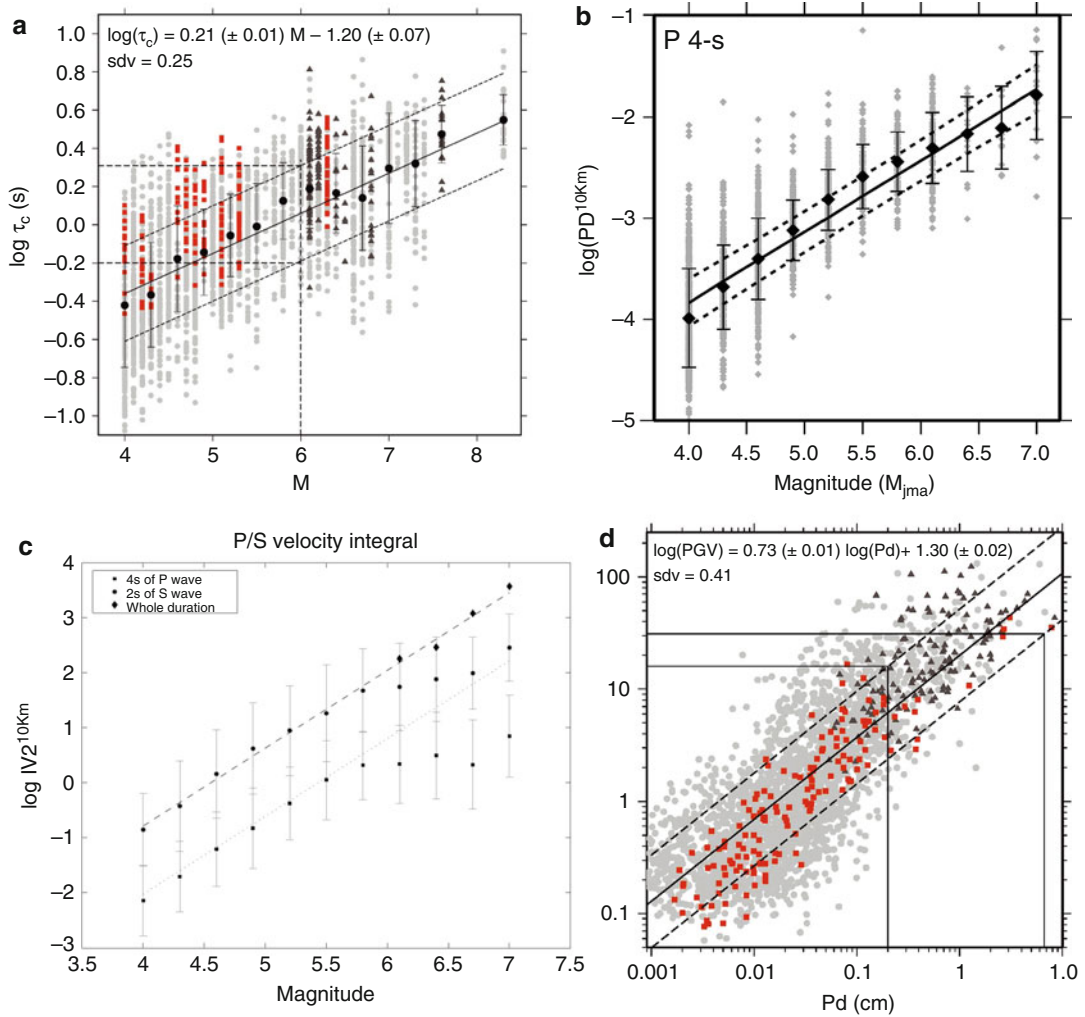
where

$$X_i = \alpha X_{i-1} + x_i^2 \quad (12)$$

$$D_i = \alpha D_{i-1} + \left(\frac{dx}{dt} \right)_i^2 \quad (13)$$

where T_i^p is the predominant period at the i th sample, x_i is the recorded ground velocity, X_i is the smoothed squared ground velocity, D_i is the smoothed squared velocity derivative, and α is a smoothing constant. They showed that an empirical log-linear correlation exists between the event magnitude and the maximum of T_i^p (named $T_{i \text{ max}}^p$) within 2–4 s after the P-wave arrival. Using a dataset of Californian earthquakes, they derived two linear relations between $T_{i \text{ max}}^p$ and magnitude, for small earthquakes (in the magnitude range 3–5) and for larger earthquakes (magnitude range 5–7.3). In particular, for the upper magnitude class they found a log-linear relationship of the form:

$$M = 7.0 \log(T_{i \text{ max}}^p) + 5.9 \quad (14)$$



Source Characterization for Earthquake Early Warning, Fig. 4 Empirical scaling relationships between early warning parameters and magnitude. (a) The average period as a function of magnitude (After Zollo et al. (2010)); (b) the initial peak displacement as a

function of magnitude (After Lancieri and Zollo (2008)); (c) the correlation between P_d and PGV (After Zollo et al. (2010)); and (d) the scaling of the integral of squared velocity (IV2) with magnitude (After Festa et al. (2008))

A similar period parameter has been proposed by Kanamori (2005), who defined the average period (τ_c) of the first seconds of P-wave signal as

$$\tau_c = 2\pi \sqrt{\frac{\int_0^{\tau_0} u^2(t) dt}{\int_0^{\tau_0} v^2(t) dt}} \quad (15)$$

where u and v are displacement and velocity, respectively; the integrals are computed over a time window $(0, \tau_0)$ starting from the P-wave arrival and τ_0 is generally set at 3 s.

Using a database of Taiwan, Japan, and Italy earthquakes (magnitude range $4 < M < 8.3$), Zollo et al. (2010) determined the relationship between average period and magnitude (Fig. 4a). By measuring τ_c in seconds, they found

$$\log(\tau_c) = 0.21(\pm 0.01)M - 1.19(\pm 0.08) \quad (16)$$

Both period parameters ($T_{i \max}^p$ and τ_c) are empirically related to the event magnitude but are pretty independent of the distance.

An alternative estimate of the earthquake size can be obtained using amplitude parameters. Zollo et al. (2006) showed that the low-pass filtered, peak displacement amplitude of initial P- and S-wave seismic signals correlates with the earthquake magnitude (Fig. 4b). The P- and the S-peak amplitudes are measured in a short time window (2–4 s) after the arrival times of P- and S-waves, respectively. The initial peak amplitude can be measured on the single vertical component or on the modulus of the displacement vector, as proposed by Lancieri and Zollo 2008. In both cases the functional dependence of the peak amplitude follows the more general relationship, including also the dependence on the distance (Eq. 8), with the reference distance $R_0 = 10$ km. Hereinafter, we will adopt the initial P-peak amplitude of the vertical component and will refer to it as P_d .

Finally, Festa et al. (2008) investigated the scaling of the early-radiated energy, with the final size of the event. The radiated energy can be inferred from the squared velocity integral (IV2), which is measured on the initial portion of P- and S-wave signals and is defined as follows:

$$IV2_c = \int_{t_c}^{t_c + \Delta t_c} v_c^2(t) dt \quad (17)$$

where the subscript c refers to the P- or S-phase, t_c is the corresponding first arrival, and v_c is the particle velocity measured on the seismograms. Moreover Δt_c is the length of the signal along which the analysis is performed. They found an evident log-linear scaling of IV2 with magnitude for both P- and S-wave data, up to $M = 5.8$ (Fig. 4c). Beyond this value, the early energy increases less with respect to the final magnitude. Thus, they suggested that early-radiated energy can be used to discriminate whether the event has a magnitude larger or smaller than 5.8, and only in the latter case it allows for real-time magnitude

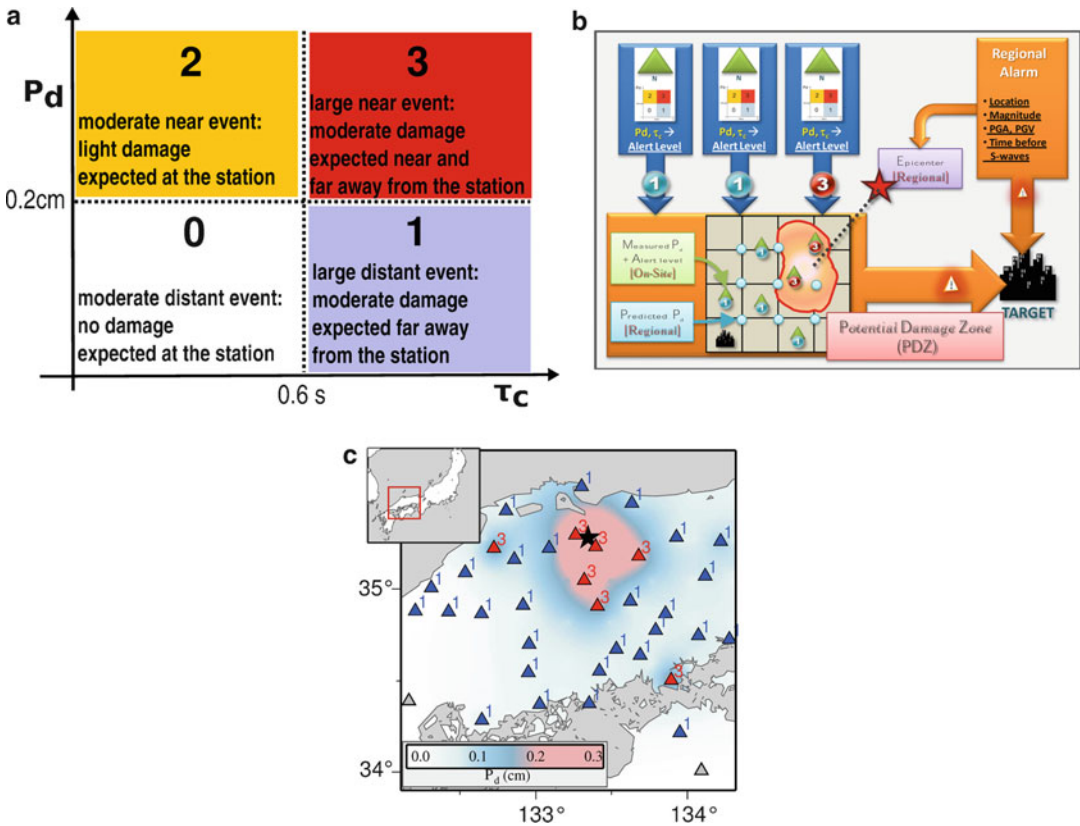
estimation. For larger magnitudes a saturation effect of IV2 prevents from a correct evaluation of the event size. The saturation effect disappears when the velocity integral is evaluated along the whole signal duration, showing a robust log-linear correlation up to larger magnitudes (~ 7).

In addition to the event magnitude, another relevant and complementary piece of information to be provided in real time is the estimate of the expected ground shaking at target sites. This latter represents an important aspect of the practical implementation of an EEWS and is crucial for the decision-making processes and the prompt activation of security actions and emergency procedures. With this aim, an empirical correlation between the initial peak displacement (P_d) and the final observed peak ground velocity (PGV) can be used. Analyzing a database of earthquake records from Japan, Taiwan, and Italy (2009 Mw 6.3 L'Aquila mainshock and aftershocks) and considering a maximum recorded distance of 60 km, Zollo et al. (2010) obtained the following regression relation (Fig. 4d):

$$\log(PGV) = 0.73(\pm 0.01)\log(P_d) + 1.30(\pm 0.02) \quad (18)$$

where PGV is in cm/s and P_d is in centimeters.

Amplitude and period parameters thus provide complementary information about the ongoing earthquake, being them related to the expected ground shaking at recording sites and to the earthquake magnitude, respectively. With this in mind, Wu and Kanamori (2008) proposed an original, on-site alert-level scheme, based on the combination of measured P_d and τ_c parameters. The idea is to rapidly distinguish the case of a small/large and close/faraway event from the observed values of P_d and τ_c at each recording site. Following this idea, Zollo et al. (2010) proposed a threshold-based approach to EEWS which is aimed at the setup of local alert levels based on a decision table. The key element of the method is the real-time, simultaneous measurement of initial peak displacement (P_d) and period parameter (τ_c) in a 3 s window after the first P-arrival time and on the use of the initial peak displacement as a proxy for the PGV .



Source Characterization for Earthquake Early Warning, Fig. 5 Alert levels and threshold values for observed early warning parameters (After Zollo et al. (2010)). (a) P_d versus τ_c diagram showing the chosen threshold values and the regions delimiting the different alert levels. Level 3 = damage expected nearby and far away from the station; level 2 = damage expected only nearby the station; level 1 = damage expected only far away from the station; level 0 = no expected damage (After Zollo et al. (2010)). (b) Conceptual scheme for the potential damage zone definition. Measured and predicted Pd

values over the area are interpolated and the PDZ is obtained by delimiting the $P_d = 0.2$ isoline. (c) Example of potential damage zone resulting from the interpolation of measured and predicted P_d values. Gray triangles represent the stations triggered by the earthquake, while red and blue triangles show the alert level recorded at each station, as soon as 3 s of signal after the P-picking are available. The color transition from light blue to red delimits the potential damage, PDZ (After Colombelli et al. (2012a))

From the analysis of strong-motion data of Japan, Taiwan, and Italy, Zollo et al. (2010) calibrated the threshold values, for the definition of four alert levels (0, 1, 2, 3) (Fig. 5a). The threshold values have been established according to the P_d vs. PGV and to the τ_c vs. M empirical relationships (Fig. 4a, d). The threshold values correspond to a minimum magnitude $M = 6$ and to an instrumental intensity $I_{MM} = 7$, assuming that the peak ground velocity provides the instrumental intensity through the relationship of Wald et al. (1999). The alert-level scheme can be

interpreted in terms of potential damaging effects nearby the recording station and far away from it. For example, following the scheme of Fig. 5a, the maximum alert level (level 3, i.e., $\tau_c \geq 0.6$ s and $P_d \geq 0.2$ cm) corresponds to an earthquake with predicted magnitude $M > 6$ and with an expected instrumental intensity (at the site) $I_{MM} \geq 7$. This means that the earthquake is likely to have a large size and to be located close to the recording site and a high level of damage is therefore expected either nearby or far away from the recording station. On the contrary, in

case of a recorded alert level equal to 0 ($\tau_c < 0.6$ s and $P_d < 0.2$ cm), the event is likely to be small and far from the site, thus no damage is expected either close or far away from the station. The application of this method to a series of large Japanese events ($M > 6$) (Colombelli et al. 2012a) confirmed that the threshold-based approach is a robust strategy to rapidly predict the expected damage at recording sites, showing a very good matching between the real-time assigned local alert levels and the final, observed peak ground velocity, carried by the later arrival of S-waves.

Extended Source Approach

Expansion of the P-Wave Time Window for Large Magnitude Events

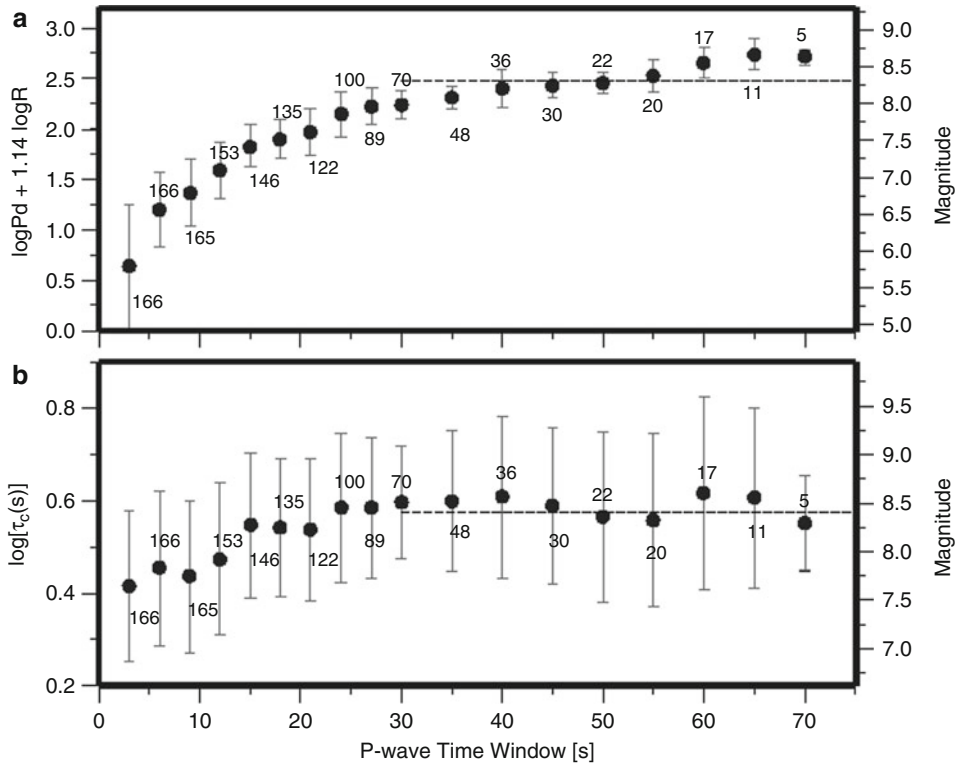
There are no concerns, in fact, on the effectiveness of EEWS for the real-time characterization of small and moderate events, for which the fracture process is concluded within a few seconds. Anyhow, regression relationships between peak displacement, predominant period, and other integrated parameters based on 2–4 s of P-wave saturate beyond a magnitude threshold that can be estimated into the range 7–7.5. This saturation is owed to the limited window range that images a too small portion of the fault plane. When an earthquake extends for hundreds of kilometers on the fault, its dynamics is expected to be controlled by large asperities that the rupture can meet during its propagation also far away from the hypocenter. Another aspect to be carefully checked is the filter used to get a reliable low-frequency signal. Generally the high-pass cutoff frequency is 0.075 Hz that is comparable with the corner frequency of an M 7 event. This means that up to magnitude 7, the cutoff filter still includes the ensemble radiation emitted by the fault while for larger magnitudes the selected frequency range extracts only the constructive interference coming from specific regions of the fault. Investigation of large magnitude events hence requires a progressive extension of the P-wave time window, with update of early warning parameters at individual stations. Since at fixed time close stations experienced a large time

window when compared to farther stations, an efficient procedure should manage in real time both the increasing of the time window at single stations and different time windows, when combining information coming from different stations.

The analysis of strong-motion data of the recent Mw 9.0, 2011 Tohoku-Oki mega-thrust earthquake confirmed the necessity of considering larger time windows to overcome the problem of parameter saturation. Colombelli et al. (2012b) proposed an evolutionary approach to early warning, in which amplitude and period parameters (P_d and τ_c , respectively) are measured in progressively increasing P-wave time windows. The real-time measurements are no more done within a fixed portion of signal but starting from the P-wave arrival and expanding the time window up to the arrival of the S-waves at each recording site. With such an approach, the standard methodologies and the empirical regression relationships can be extended to very large earthquakes, provided that appropriate time windows are selected for the measurements. With this approach data from different stations require a specific weight depending on the available P-wave time window (PTW). Specifically, a weight proportional to the square of PTW was shown to provide a good compromise between the window length and the number of stations. As shown in Fig. 6a, that peak displacement increases with PTW, indicating a final level of 8.5. Final underestimation may be related in this case to the regression relationship, validated for smaller magnitude events and shorter distances, to the cutoff frequency, or to the specific dynamics of the event, which exhibited a frequency-dependent radiation (Fig. 6b). On the other hand, predominant period does not monotonically increase with larger time window, indicating that its use in real time could be sensitive to the specific event to be analyzed.

Real-Time Estimation of Fault Rupture Extent Using Strong-Motion Data

Even if we improve the magnitude estimation for very large earthquakes, distance metrics might be inadequate, mostly for targets that are close to the causative fault but far from the epicenter. In such a



Source Characterization for Earthquake Early Warning, Fig. 6 Real-time evolution of average values of peak displacement (P_d) (a) and predominant period (τ_c) (b) as a function of the P-wave time window used. Error bars are computed as the standard deviation associated to each value; the gray number close to each point represents the number of stations used for each considered time window.

Both parameters exhibit saturation when a 25–30 s PTW is used; the saturation level is shown by the *gray-dashed lines*. For each plot, the corresponding magnitude scale is also represented; this has been derived based on the coefficients of Eq. 1 and on the τ_c vs. M relationship determined by Zollo et al. (2010) (After Colombelli et al. (2012b))

case ground-motion prediction could be significantly underestimated. To this end, Yamada et al. (2007) proposed a method to classify recording stations in near- and far-source distance. If a dense seismic network is available, this subdivision can be used to infer information about the fault geometry. Analyzing strong-motion data from past events, the authors found that the combination of vertical acceleration and horizontal velocity produces the best performance for stations classification. Their discriminant is given by

$$F_i = 6.046 \log Z a_i + 7.885 \log H v_i - 27.091 \quad (19)$$

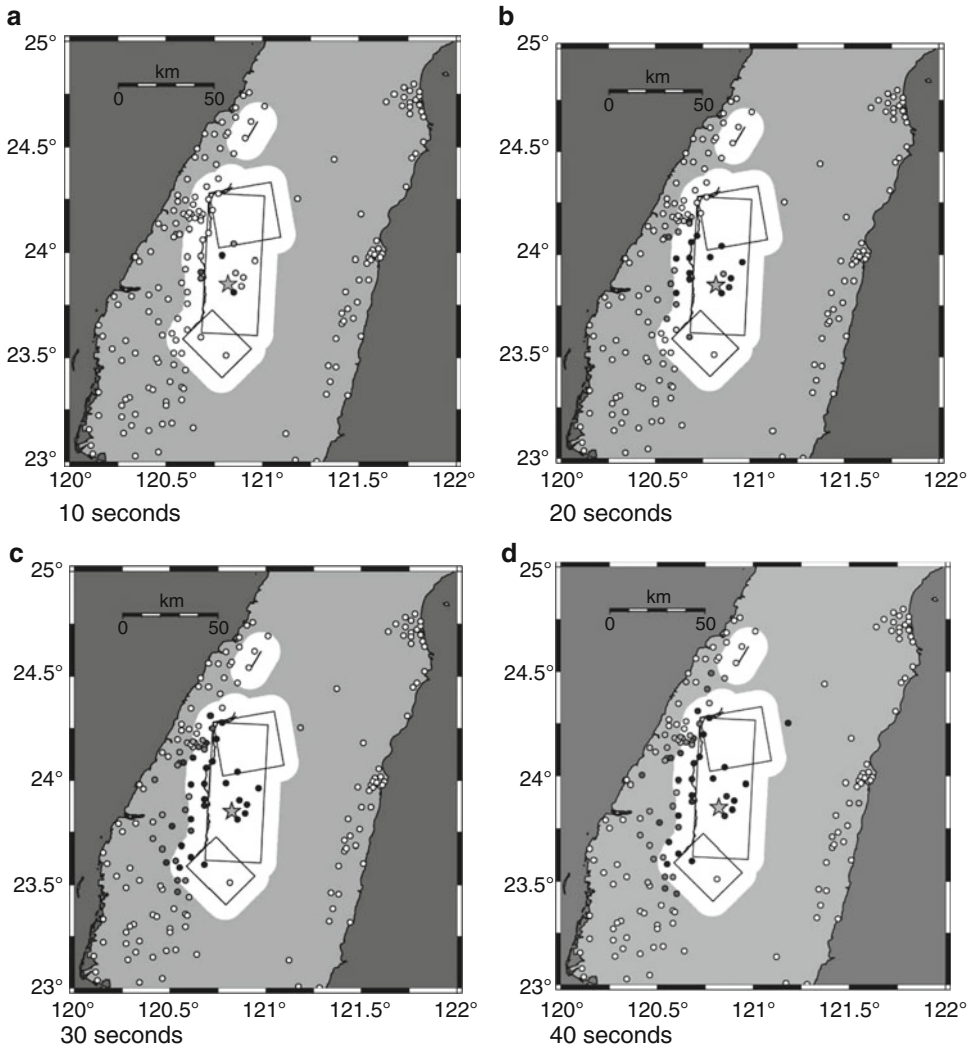
where $Z a_i$ and $H v_i$ represent the peak acceleration on the vertical component (in cm/s^2) and the

square root of the sum of the squares of the peak velocity on the horizontal components, recorded at the i th station, respectively. Thus, the quantity

$$P_i = \frac{1}{1 + e^{-F_i}} \quad (20)$$

provides the probability that the i th station is in the near- or far-source region. Specifically, the recording site is classified as a near-source station if the probability is greater than 1/2.

For real-time application, peak values used for recording site classification are computed from incoming data every 10 s for each station and then used in the discriminant function. As example, in Fig. 7 the procedure is applied to the case of



Source Characterization for Earthquake Early Warning, Fig. 7 Snapshots showing the recording stations identified as near-source sites in the case of the 1999 M7.0 Chi-Chi (Taiwan) earthquake. The *star* marks the earthquake epicenter. *Circles* represent the

recording stations. *White zones* correspond to the source area and *rectangle s* identify the surface projection of the causative faults. The darker is the marker, the higher is the probability that the corresponding site is located in near source (After Yamada et al. (2007))

the 1999, M7.0, Chi-Chi (Taiwan) earthquake. As time goes on, the near-fault stations identify quite well the fault extent.

In a recent paper, Yamada (2014) proposed an improved version of this discriminant. This approach identifies the fault rupture geometry by classifying stations into near-source and far-away source and provides reasonably good estimates of the extent of the near-source area.

Potential Damaged Zone

As for the point-source analysis, on-site approaches can be used to predict finite source effects, such as directivity and azimuthal changes on the ground motion directly analyzing the early motion at the single sites, without explicitly estimating any source parameters. The idea is to use the threshold-based approach to compute threshold levels at single stations and to interpolate the

levels to define the potential damage zone associated with the earthquake (Kanamori 2005; Zollo et al. 2010).

At a single site, the real-time measure of the P-wave parameters (P_d and τ_c) thus provide a rough but rapid alert notification, although no information is given about the accurate earthquake size and location. Considering that the same measure can be performed at different nodes of a dense array of stations, deployed nearby and far away from the earthquake epicenter, the mapping of the recorded maximum alert levels (3 or 2) provides a preliminary estimation of the extent of the potential damage zone (PDZ), i.e., the area where the damages are expected to be equal or greater to those predicted by the level VII according to the instrumental intensity scale.

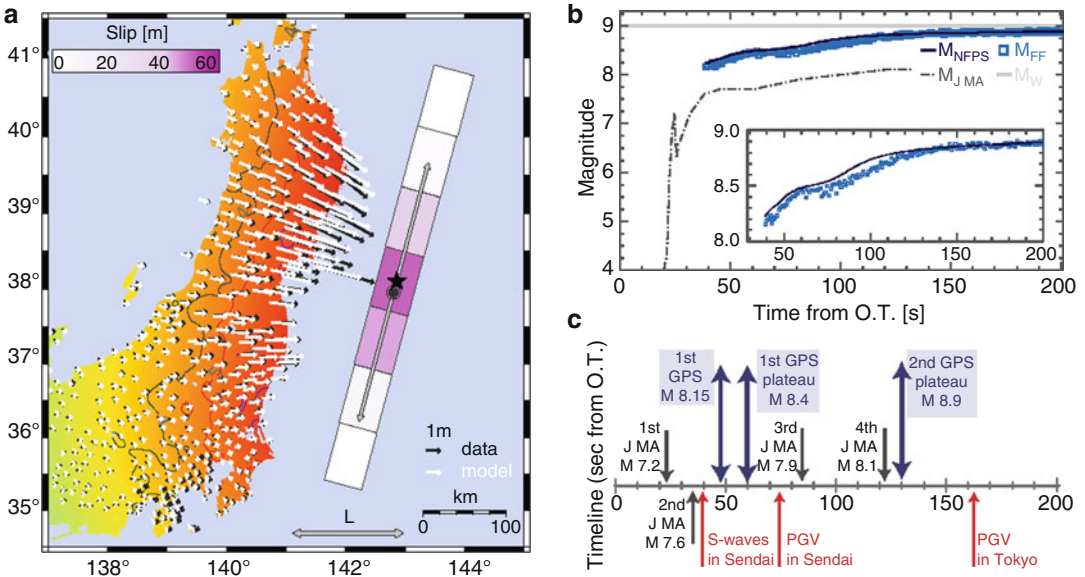
The same authors proposed that a rough but very rapid estimation of the PDZ extent can also be made from the updated averages of the period parameter τ_c which is recorded at strong-motion sites in the near-source region. Fixing the parameter P_d at its threshold value and using the progressively updated estimations of the period parameter, the empirical equation

$$\log P_d = A + B \log \tau_c + C \log R \quad (21)$$

(R is the hypocentral distance; A, B, C are determined by the regression analysis) can be used to determine the radius of the area within which the strong ground motion is expected to produce an instrumental intensity $I_{MM} > 7$. The off-line application of the method showed a very consistent match between the rapidly predicted (within a few sec from the first recorded P-wave) and observed damage zone, the latter being mapped from detailed macroseismic surveys a few days after the event. A more refined technique for the real-time mapping of the PDZ has been proposed by Colombelli et al. (2012a), who tested the method on data from ten $M > 6$ Japanese earthquakes occurred in the period 2000–2009 and recorded at the K-Net/Kik-Net Japan accelerometric arrays. The method is based on the recursive use of the ground-motion prediction equation for P_d (Eq. 21), with real-time, updated estimate of the earthquake location and

characteristic period τ_c . The area covered by stations is divided into cells, using a prefixed spatial grid, which is needed to fill the gaps where stations are not available. At those stations where the first 3 s of signal after the P-picking are available, an alert level is locally assigned, based on P_d and τ_c measurements. At the same time, the event location is obtained by using the available P-picks and a real-time location method. Furthermore, the expected P_d value can be predicted at each node of the grid, through Eq. 21 and the τ_c averaged over the available measurements at the considered time step. Measured and predicted P_d values are thus interpolated; the area within which the highest level of damage is expected can be delimited by the isoline corresponding to $P_d = P_d^{th} = 0.2$ cm.

The predicted PGV values at stations and at grid nodes can be computed using the equation relating $\log(PGV)$ and $\log(P_d)$ (Eq. 1 in Zollo et al. 2010) and finally converted into an instrumental intensity measure using the relationship of Wald et al. (1999). An interpolation is performed between all intensity values to produce a real-time, continuously updated, intensity map (Fig. 5b,c). This procedure is repeated every time step (typically 1 s); as the waves propagate within the area and trigger other stations, the event location is refined by using P-pickings, average τ_c value is updated, and more data are used for the interpolation procedure. The studied cases for Japan earthquakes displayed a very good matching between the rapidly predicted earthquake potential damage zone inferred from initial P-peak displacement amplitudes and the instrumental intensity map, the latter being mapped after the event, using peak ground velocity and/or acceleration, or from field macroseismic surveys. The performance of the method has been evaluated by defining *successful*, *missed*, and *false alarms* in terms of observed versus predicted instrumental intensities for all the analyzed events and by counting their relative percentage. A very high percentage (88 %) of alert levels has been correctly assigned, and most of maximum alert levels correspond to the area within which the highest level of damage has been observed.



Source Characterization for Earthquake Early Warning, Fig. 8 (a) Example of snapshot of the GPS-based strategy for the Mw 9.0, 2011 Tohoku-Oki earthquake. The background color represents the predicted intensity distribution using the final magnitude value and the distance from the finite fault. The purple color scale shows the slip distribution on the seven-patch slip model. The total length estimate (L) is also plotted as vectors on the fault plane. Black vectors represent the observed horizontal offset while white vectors show the static displacement resulting from the inversion algorithm. (b) Magnitude estimate as a function of time from the origin time. The magnitude is obtained by using the static displacement

Summary

An approach that may be more robust is the inversion for the final slip on the fault plane, which allows consideration of the contributions from the entire fault plane and may provide a more realistic estimate of both size and potential damage of the ongoing event.

One way to recover the slip distribution on the fault plane is to use the permanent ground deformation, which is directly related to the earthquake magnitude. The static component of ground motion could, in principle, be obtained using dynamic range, accelerometric sensors, which are able to record unsaturated signals in a broad range of frequencies (0 ÷ 100 Hz). Accelerometric records are integrated twice to

provided by GPS data using the near-field, point source approximation (dark blue solid line) and resulting from the slip inversion (small blue squares). For comparison, the evolution of magnitude estimate provided by the JMA early warning system is also shown as a dotted gray line, and the continuous gray line represents the real moment magnitude value. (c) Warning timeline for the Tohoku-Oki earthquake showing when the GPS information is available with respect to the time at which the strongest shaking occurs in the Sendai and Tokyo regions and with respect to the JMA warnings. After Colombelli et al (2013)

obtain displacement time series. Unfortunately, for near-field records, this operation may introduce artificial effects and long-period drifts (Boore et al. 2002), which are usually removed by applying a high-pass filter. The application of the filter, while removing the artificial distortions, reduces the low-frequency content of the recorded waveforms, resulting in the complete loss of the low-frequency energy radiated by the source and of the static offset, which is the most relevant piece of information for a large earthquake.

GPS stations are able to provide a direct and evolutionary measurement of the permanent ground deformation, i.e., of the resulting co-seismic displacement after the dynamic vibration has finished. With the increasing diffusion of

high-rate 1 Hz GPS stations, the seismological community has begun looking at GPS data as a valid complement to the high-frequency information provided by seismic data.

Once the permanent static offset is extracted from GPS displacement time series, a real-time static slip inversion scheme can be used to infer the slip distribution on the fault plane for both the rapid determination of the event size and for the near-real-time estimation of the rupture area (Fig. 8). Many authors have recently started applying GPS data to EEW (Allen and Ziv 2011; Crowell et al. 2012; Wright et al. 2012; Ohta et al. 2012; Colombelli et al. 2013); they showed that 1 Hz GPS data provide a rapid and remarkably robust magnitude estimate and can be used for the real-time estimation of the rupture area, which, in turns, would allow for an improved prediction of the earthquake damaging potential.

Another possibility would be a fast kinematic inversion of the rupture process, searching for the fault geometry, the focal mechanism, and the slip distribution on the fault plane. A first approach was proposed by Dreger et al. (2005), who combined preliminary information on location and magnitude with scaling laws for source size to determine both the focal mechanism and a kinematic rupture model from inversion of strong-motion integrated displacement waveforms. The approach revealed useful for estimating the ground shaking in the near-fault domain for the 2003, M 6.5, San Simeon, California, earthquake. Nevertheless, computational times were estimated however of the order of minutes, making the results available in near real time but not in real time.

Cross-References

- [Early Earthquake Warning \(EEW\) System: Overview](#)
- [Earthquake Magnitude Estimation](#)
- [Engineering Characterization of Earthquake Ground Motions](#)
- [Seismic Actions Due to Near-Fault Ground Motion](#)

References

- Allen RM, Kanamori H (2003) The potential for earthquake early warning in southern California. *Science* 3:685–848
- Allen RM, Ziv A (2011) Application of real-time GPS to earthquake early warning. *Geophys Res Lett* 38, L16310. doi:10.1029/2011GL047947
- Allen RM, Brown H, Hellweg M, Khainovski O, Lombard P, Neuhauser D (2009) Real-time earthquake detection and hazard assessment by ElarmS across California. *Geophys Res Lett* 36, L00B08. doi:10.1029/2008GL036766
- Boore DM, Stephens CD, Joyner WB (2002) Comments on baseline correction of digital strong-motion data: examples from the 1999 hector mine, California, earthquake. *Bull Seism Soc Am* 92:1543–1560. doi:10.1785/0120000926
- Böse M, Ionescu C, Wenzel F (2007) Earthquake early warning for Bucharest, Romania: novel and revised scaling relations. *Geophys Res Lett* 34, L07302. doi:10.1029/2007GL029396
- Colombelli S, Amoroso O, Zollo A, Kanamori H (2012a) Test of a threshold-based earthquake early warning using Japanese data. *Bull Seismol Soc Am* 102:1266–1275. doi:10.1785/0120110149
- Colombelli S, Zollo A, Festa G, Kanamori H (2012b) Early magnitude and potential damage zone estimates for the great Mw 9 Tohoku-Oki earthquake. *Geophys Res Lett* 39, L22306. doi:10.1029/2012GL053923
- Colombelli S, Allen RM, Zollo A (2013) Application of real-time GPS to earthquake early warning in subduction and strike-slip environments. *J Geophys Res* 118:3448–3461. doi:10.1002/jgrb.50242
- Crowell BW, Bock Y, Melgar D (2012) Real-time inversion of GPS data for finite fault modeling and rapid hazard assessment. *Geophys Res Lett* 39, L09305. doi:10.1029/2012GL051318
- Cua G, Fisher M, Heaton T, Wiemer S (2009) Real-time performance of the virtual seismologist earthquake early warning algorithm in southern California. *seism. Res Lett* 80:740–747. doi:10.1785/gssrl.80.5.740
- Dreger DS, Gee L, Lombard P, Murray MH, Romanowicz B (2005) Rapid finite-source analysis and near-fault strong ground motions: application to the 2003 M_w 6.5 San Simeon and 2004 M_w 6.0 Parkfield Earthquakes. *Seism Res Lett* 76:40–48. doi:10.1785/gssrl.76.1.40
- Espinosa-Aranda JM, Cuéllar A, Rodríguez FH, Frontana B, Ibarrola G, Islas R, García A (2011) The seismic alert system of Mexico (SASMEX): progress and its current applications. *Soil Dyn Earth Eng* 31:2154–2162. doi:10.1016/j.soildyn.2010.09.011
- Festa G, Lancieri M, Zollo A (2008) Magnitude estimation from early radiated energy. *Geophys Res Lett* 35, L22307. doi:10.1029/2008GL035576
- Kanamori H (2005) Real-time seismology and earthquake damage mitigation. *Ann Rev Earth Planet Sci* 33: 195–214

- Lancieri M, Zollo A (2008) A Bayesian approach to the real time estimation of magnitude from the early P- and S-wave displacement peaks. *J Geophys Res* 113, B12302. doi:10.1029/2007JB005386
- Nakamura Y (1988) On the urgent earthquake detection and alarm system (UrEDAS) In: Proceedings 9th world conference on earthquake engineering, Tokyo-Kyoto, Japan vol 7, pp 673–678
- Ohta Y et al (2012) Quasi real-time fault model estimation for near-field tsunami forecasting based on RTK-GPS analysis: application to the 2011 Tohoku-Oki earthquake (Mw 9.0). *J Geophys Res* 117:B02311. doi:10.1029/2011JB008750
- Satriano C, Lomax A, Zollo A (2008) Real-time evolutionary earthquake location for seismic early warning. *Bull Seism Soc Am* 98(3):1482–1494
- Satriano C, Elia L, Martino C, Lancieri M, Zollo A, Iannaccone G (2010) PRESTo, the earthquake early warning system for southern Italy: concepts, capabilities and future perspectives. *Soil Dyn Earthq Eng* 31(2):137–153. doi:10.1016/j.soildyn.2010.06.008
- Wald DJ, Quitoriano V, Heaton T, Kanamori H (1999) Relationships between peak ground acceleration, peak ground velocity and modified mercalli intensity in California. *Earthq Spectra* 15:557–564
- Wright TJ, Houlié N, Hildyard M, Iwabuchi T (2012) Real-time, reliable magnitudes for large earthquakes from 1 Hz GPS precise point positioning: the 2011 Tohoku-Oki (Japan) earthquake. *Geophys Res Lett* 39, L12302
- Wu YM, Kanamori H (2005) Rapid assessment of damage potential of earthquake in Taiwan from beginning of P-waves. *Bull Seism Soc Am* 95:1181–1185. doi:10.1785/0120040193
- Wu YM, Kanamori H (2008) Development of an earthquake early warning system using real-time strong motion signals. *Sensors* 8:1–9
- Wu YM, Zhao L (2006) Magnitude estimation using the first three seconds P-wave amplitude in earthquake early warning. *Geophys Res Lett* 33, L16312. doi:10.1029/2006GL026871
- Yamada M (2014) Estimation of fault rupture extent using near-source records for earthquake early warning. In Wenzel F, Zschau J (eds) *Early warning for geological disasters. Advanced technologies in earth sciences*. Springer, Berlin/Heidelberg, p 29ff. doi:10.1007/978-3-642-12233-0_2
- Yamada M, Heaton T, Beck J (2007) Real-time estimation of fault rupture extent using near-source versus far-source classification. *Bull Seism Soc Am* 97:1890–1910. doi:10.1785/0120060243
- Zollo A, Lancieri M, Nielsen S (2006) Earthquake magnitude estimation from peak amplitudes of very early seismic signals on strong motion. *Geophys Res Lett* 33, L23312. doi:10.1029/2006GL027795
- Zollo A, Amoroso O, Lancieri M, Wu YM, Kanamori H (2010) A threshold-based earthquake early warning using dense accelerometer networks. *Geophys J Int* 183:963–974
- Zollo A, Colombelli S, Elia L, Emolo A, Festa G, Iannaccone G, Martino C, Gasparini P (2014) An integrated regional and on-site earthquake early warning system for southern Italy: concepts, methodologies and performances. In: Wenzel F, Zschau J (eds) *Early warning for geological disasters - scientific methods and current practices*. Springer-Verlag Berlin Heidelberg, Dordrecht

Spatial Filtering for Structural Health Monitoring

G. Tondreau¹ and A. Deraemaeker²

¹Postdoctoral Researcher, Building Architecture and Town Planning (BATir), Brussels, Belgium

²FNRS Research Associate, Building Architecture and Town Planning (BATir), Brussels, Belgium

Synonyms

Control charts; Damage localization; Damage detection; Sensor networks; Spatial filtering; Structural health monitoring (SHM)

Introduction

Assessing the integrity of the structures in real time is a very important topic for which many methods have been developed in the last decades. Today, structural health monitoring (SHM) is gaining more and more attention: in the case of bridges, the maximum loads tend to increase (increase of the vehicle weights), while most of the structures are coming to the end of their theoretical lifetime. In addition, exceptional events such as collisions or earthquakes can cause more severe and fast deteriorations. Optimal maintenance calls for an early detection of small damage in structures, as it is well known that limited and frequent repairs are much less costly than major repairs or total rebuilding after

collapse. Current monitoring practice consists in scheduled maintenances including visual inspections, ultrasounds, eddy current, magnetic field, or radiography techniques (Hellier 2003). All these experimental methods require however that the vicinity of the flaw is known and that the proximity to be inspected is accessible. Moreover, these local inspections are labor intensive and therefore very expensive. A major problem is that traditional monitoring is noncontinuous which means that if a critical damage occurs between two inspections, it might lead to catastrophic structural failure. One of the most relevant examples is the I-35W Mississippi River Bridge case (Rofidal 2007): this bridge collapsed in August 2007 killing 13 people and injuring 145, despite annual inspection.

A general trend for new structures and bridges is a lighter and more slender design, which tends to increase the levels of vibrations under ambient excitation. While these levels of vibrations need to be controlled as they could be detrimental to the lifetime of the structure, they can also be used for the continuous monitoring of the structure without disruption or decrease of functionality. The basic idea is that the occurrence of damage alters the structural parameters which in turn affect the vibration characteristics. Vibration SHM of civil engineering structures relies on ambient vibrations, as artificial excitation of such large structures is often unpractical.

Based on this basic concept, many vibration-based SHM techniques have been developed in the last decades using mainly eigenfrequencies, damping ratios, or mode shapes (Doebling et al. 1998). The reason of this popularity is the ease of measuring modal parameters or frequency responses on real structures thanks to recent advances in sensors and sensing systems and in the development of efficient operational modal analysis techniques (Reynders and De Roeck 2008; Reynders et al. 2012). Such advances are so important that more and more very large bridges are instrumented with larger and larger sensor networks. China has been the driving force in this direction with the massive instrumentation of bridges in the Hong Kong area, the largest one

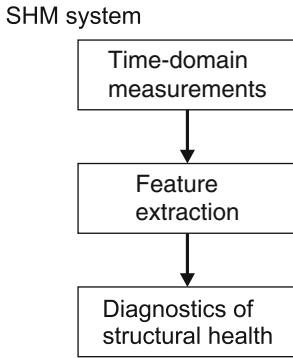
being the Stonecutters Bridge with more than 1,500 sensors among which are 58 accelerometers and 853 dynamic and static strain sensors (Ni et al. 2012). In Europe, the Messina bridge project, designed to be the largest cable-stayed bridge in the world would include a very large monitoring system with more than 3,000 sensors (De Neumann et al. 2011).

These technological advances have opened the way to real-time automated SHM of bridges. A major problem related to the use of such very large sensor networks is to find adequate techniques to post-process the data: intelligent methods are needed in order to take advantage of the enormous amount of information provided by these large networks. In fact, operational modal analysis is not yet fit for automated modal analysis using very large sensor networks.

With that perspective, rather than identifying online the full set of modal properties of the structure, an alternative is to condense the measured data while keeping the information about the potential damage occurring in the structure. The technology presented in this entry is spatial filtering, which consists in using a linear combiner to condense the information from very large networks of sensors into one or just a few “virtual sensors.” Such virtual sensors can be designed in order to react strongly to damage while being insensitive to environmental changes or even to react to a damage occurring in a specific location along the bridge.

Rytter (Rytter 1993) has proposed a hierarchical decomposition of the SHM process in four levels, which has been widely accepted in the SHM literature: detection (level 1), localization (level 2), quantification (level 3) of the damage, and prediction of the remaining service life of the damaged structure (level 4). As the level of SHM increases, the knowledge about the damage increases and, usually, the complexity of the method increases as well. The method based on spatial filtering can deal with levels 1 and 2 by comparing data measured in the current unknown state with data measured from the structure assumed to be undamaged.

The general scheme of the method is shown in Fig. 1. It is divided in three parts: (i) the



Spatial Filtering for Structural Health Monitoring, Fig. 1 General scheme of vibration-based SHM

measurement of raw time-domain data, including data reduction, (ii) the transformation of the data into information using feature extraction, and (iii) the diagnostics of the structural health based on the monitored features.

The first section of this entry deals with measurement of raw time-domain data and details the spatial filtering technique for data reduction. The second section is devoted to feature extraction, and the third section deals with the diagnostics, both for damage detection and localization. The last section presents an experimental illustration of the SHM technique on a 3.78 m long steel I-beam tested in the laboratory.

Measurement of Raw Time-Domain Data

The first building block of the SHM system is the sensor network. As stated earlier, a current trend is to implement very large sensor networks on large civil engineering structures. In such cases, it is often necessary to perform data reduction in order to decrease the power consumption and the bandwidth needed to transmit the data and to facilitate the data storage and post-processing. For SHM applications, an optimal reduction is one that significantly reduces the amount of data while keeping most of the information about the damage. The technique presented in this entry for data reduction is spatial filtering.

Spatial and Modal Filtering

Consider a structure excited with an ambient force $f(t)$ equipped with a network of n sensors whose time-domain output is denoted by $y_k(t)$ as shown in Fig. 2. The dynamic time-domain response at each sensor can be decomposed into a sum of contributions of the N mode shapes excited by the ambient force:

$$y_k(t) = \sum_{i=1}^N a_i(t) \phi_{ki} \quad (1)$$

where $a_i(t)$ is the modal amplitude of mode i and ϕ_{ki} is the projection of mode shape i on sensor k . The application of spatial filtering with coefficients α_k leads to a single sensor output $g(t)$:

$$g(t) = \sum_{k=1}^n \alpha_k y_k(t) = \sum_{i=1}^N \sum_{k=1}^n \alpha_k \phi_{ki} a_i(t) \quad (2)$$

The general scheme of spatial filtering can be used for different purposes. A first idea is to condense the information into the modal coordinates of the undamaged structure, which allows to reduce the data $y_k(t)$ from a very large network of n sensors to a limited set of $a_i(t)$ time series from N modal sensors. This idea is motivated by the fact that the vibration of structures typically involves only a few mode shapes which are excited in a given frequency band of interest.

To design a modal filter, the vector of the linear combiner α_k must be orthogonal to all the modes of the structure in a frequency band of interest, except mode l :

$$\sum_{k=1}^n \alpha_k \phi_{ki} = \delta_{li} \quad (3)$$

Equation 2 then reduces to

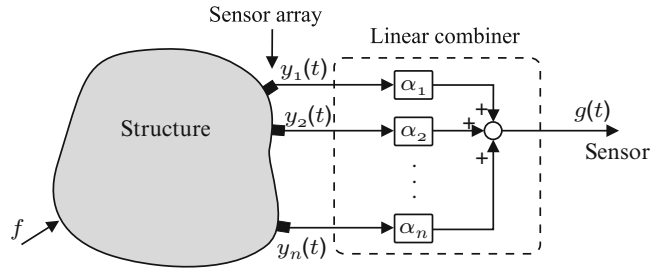
$$g(t) = a_l(t) \quad (4)$$

Equation 3 can be written in a matrix form:

$$[C]^T \{\alpha\} = e_l, \quad (5)$$

Spatial Filtering for Structural Health Monitoring,

Fig. 2 Principle of spatial filtering on a network of n sensors



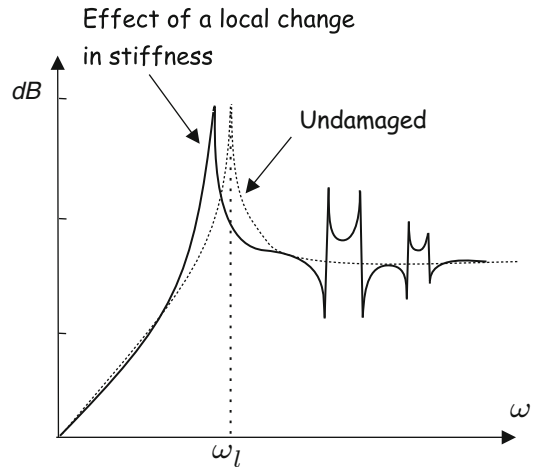
where $\{\alpha\} = \{\alpha_1 \dots \alpha_n\}^T$ and $e_l = \{0 \ 0 \dots 1 \dots 0 \ 0\}^T$ (all components set to 0 except the l th component) and $[C]$ is a matrix whose columns correspond to the N mode shapes projected on the n sensors of the array. Equation 5 can be solved only if there are at least as many sensors as there are mode shapes ($n \geq N$). In this case, matrix $[C]$ is rectangular and the system of equations is overdetermined (several solutions exist which satisfy Eq. 5). The minimum norm solution is usually adopted by computing the pseudo-inverse (regularized with singular value decomposition) of $[C]^T$ (Deraemaeker et al. 2008).

The modal filter can be tuned to any of the N mode shapes in the frequency band of interest. In the frequency domain, the FRF of the modal filter tuned on mode l is given by

$$G(\omega) = \frac{b_l}{(\omega_l^2 - \omega^2 + 2j\xi_l\omega_l\omega)}, \quad (6)$$

where b_l depends on the excitation level and position. It corresponds to the FRF of a single degree of freedom system which presents a single peak at frequency ω_l .

Using modal filtering, the amount of data from a large network of n sensors can be significantly reduced to just a few modal filters. Such modal filters are virtual sensors which measure the amplitude of vibration of each mode separately. Typically, for a bridge excited by the ambience, only a few (up to ten) mode shapes are relevant, and the information can be drastically reduced. The computation of the linear combiner coefficients α_k is based on Eq. 5 which requires the knowledge of matrix $[C]$. This matrix should be built using experimentally identified mode shapes in order to avoid the need for



Spatial Filtering for Structural Health Monitoring, Fig. 3 Effect of a structural change on the modal filter tuned on mode l (Deraemaeker et al. 2008)

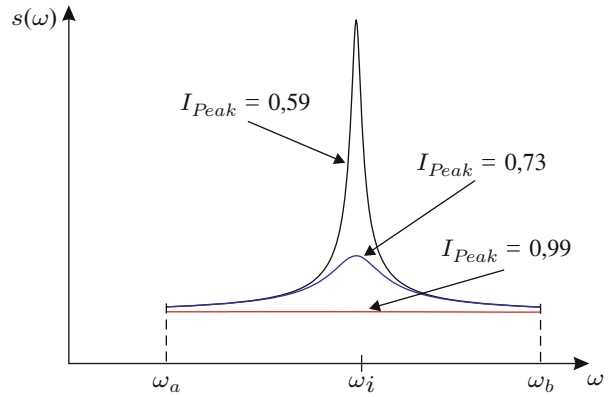
a numerical model of the structure to be monitored. This can be achieved thanks to operational modal analysis, using, for instance, stochastic subspace-based methods (Reynders and De Roeck 2008).

Effect of Damage on Modal Filters

Suppose now that damage initiates in the structure. This damage will alter the stiffness matrix, affecting the eigenfrequencies and mode shapes of the structure. The change of mode shapes will be reflected in matrix $[C]^T$ so that Eq. 5 will now be violated. In other terms, the damage will alter the mode shapes, and the coefficients of the linear combiner will not be tuned anymore. This will result in the reappearance of the filtered peaks, as illustrated in Fig. 3. This is the central idea of vibration-based SHM based on modal filters, as detailed in Deraemaeker et al. (2008).

Spatial Filtering for Structural Health Monitoring,

Fig. 4 Example of I_{Peak} values for an increasing peak



From Data to Information

The raw time-domain output of modal filters is not exploitable as such, as the changes in the time-domain response due to damage will be very small. The transformation in the frequency domain is an important step which allows to enhance these small changes by focusing on the frequency bands away from the main peak where the filtered peaks will reappear. In the case of ambient vibrations, the input force is not known, and the power spectral density (PSD) $S_{gg}(\omega)$ of $g(t)$ should be computed. This quantity is directly related to the amplitude of the FRF $G(\omega)$ as follows (Ewins 1984):

$$S_{gg}(\omega) = |G(\omega)|^2 S_{ff}(\omega) \quad (7)$$

This equation shows that if peaks are filtered in the FRF $G(\omega)$, they will be filtered in $S_{gg}(\omega)$ as well. When damage occurs, the reappearance of spurious peaks should be monitored based on the power spectral density of the output of the modal filter $S_{gg}(\omega)$.

Feature Extraction

Because the spurious peaks are expected to appear around the initial eigenfrequencies of the structure, the strategy consists in extracting one feature in each frequency band around them. Let $s(\omega)$ be the frequency dependent amplitude in the frequency range (ω_a, ω_b) (Fig. 4). For ambient vibrations, $s(\omega)$ is the PSD $S_{gg}(\omega)$.

The frequency band is typically defined by $\omega_a = 0.95\omega_i$ and $\omega_b = 1.05\omega_i$, where ω_i is the angular eigenfrequency. A peak indicator is then computed in this frequency interval:

$$I_{Peak} = \frac{2\sqrt{3RVF}}{\omega_b - \omega_a}, \quad (8)$$

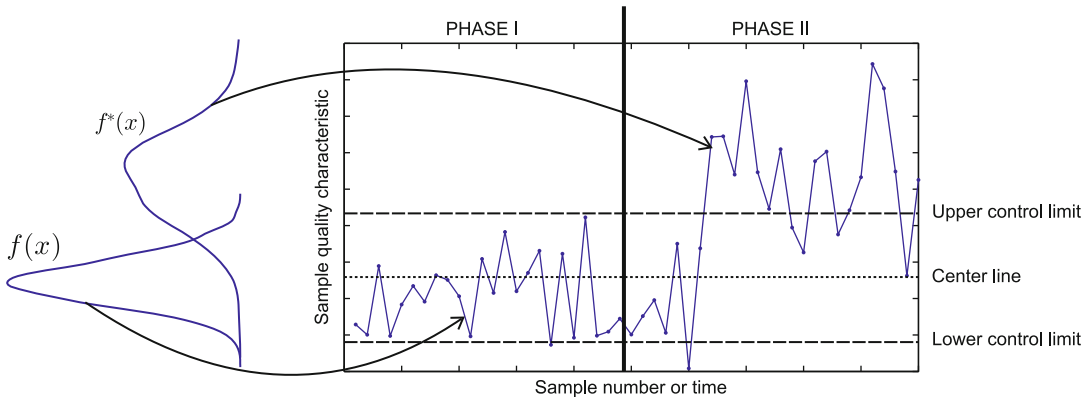
where RVF is the root variance frequency defined by

$$RVF = \sqrt{\frac{\int_{\omega_a}^{\omega_b} (\omega - FC)^2 s(\omega) d\omega}{\int_{\omega_a}^{\omega_b} s(\omega) d\omega}}, \quad (9)$$

and FC is the frequency center defined by

$$FC = \frac{\int_{\omega_a}^{\omega_b} \omega s(\omega) d\omega}{\int_{\omega_a}^{\omega_b} s(\omega) d\omega} \quad (10)$$

Theoretically, I_{Peak} is equal to 1 if $s(\omega)$ is constant and decreases when the peak grows. Figure 4 gives an example of I_{Peak} values computed between ω_a and ω_b when a spurious peak grows around ω_i . The advantage of that feature is that it is very sensitive to the peak growth but not to the level of the excitation force, which is particularly needed for ambient vibrations. More details on the peak indicator computation can be found in Deraemaeker and Worden (2010).



Spatial Filtering for Structural Health Monitoring, Fig. 5 A typical control chart

Diagnostics

The last building block of the SHM system is the diagnostics. It consists in assessing, based on the monitored features (here the peak indicators) whether the structure is healthy or damaged, and possibly in giving indication on the location of damage.

Statistical Analysis of the Features

When the excitations are random, the peak indicators behave like random variables. They will therefore follow a statistical distribution which can be inferred from several undamaged samples. Many tools have been developed to detect a change in that statistical distribution such as outlier analysis or hypothesis testing. In this contribution, control charts (Montgomery 2009; Ryan 2000) are presented. This tool of statistical quality control plots the features or quantities representative of their statistical distribution as a function of the samples. Different univariate or multivariate control charts exist but all these control charts are based on the same principle which is summarized in Fig. 5.

In phase I, a set of samples are collected and analyzed to infer statistical characteristics of the process when it is assumed to be in control (i.e., when the structure is undamaged). The aim of this step is to compute the control limits (upper control limit UCL and/or lower control limits LCL) between which the feature should be included if

the process stays in control. Those limits are governed by the statistical distribution $f(x)$ of the quality characteristic and the probability $1 - \gamma$ that any in-control sample will fall inside the control limits. There are control limits that can be computed to detect a shift of the mean value of the process or a shift of the variance of the process.

Once a set of reliable control charts has been established (phase I), the process is under monitoring (phase II). The process state is unknown (it might be in or out of control), and if a sample falls outside the control limits previously computed, it is considered as an abnormal value, and a warning is triggered. Phase I fixes the probability of type I (false alarms) and type II (missing alarms) errors. Because the control limit values are based on the number of samples in the in-control set of data, the statistical distribution $f(x)$, and the γ value, the statistical analysis must be done very carefully.

Typically, one can find two families of control charts in the literature: the univariate control charts and the multivariate control charts. The first family will be used if there is only one feature to be monitored, while the second one is used when several features are monitored at the same time. In the present application, this means that if one checks the appearance of only one spurious peak around one given natural frequency, the univariate control chart will be applied on that feature, while a multivariate

control chart will be used if one checks the spurious peaks around several eigenfrequencies in each modal filter. Finally, there are two categories of control charts in each family: the Shewhart control chart and the time-weighted control charts. The first category monitors each sample independently while the second category considers the previous samples to monitor the current sample, which allows to detect smaller shifts. It has been found that the best results are obtained when the Shewart control charts are applied, because the time weighted control charts increased too much the number of type I errors. For this reason, only the univariate and the Hotelling T^2 control charts are presented.

Individual control chart Consider that only one feature x following a normal distribution is monitored (e.g., one I_{peak} value in each modal filter). The individual control chart will monitor that individual feature x . The control limits are

$$\begin{aligned} \text{Upper Control Limit: } UCL &= \bar{x} + 3 \frac{\overline{MR}}{d_2} \\ \text{Lower Control Limit: } LCL &= \bar{x} - 3 \frac{\overline{MR}}{d_2} \end{aligned} \quad (11)$$

If the number of samples in phase I is n , then

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (12)$$

$$\overline{MR} = \frac{1}{n} \sum_{i=2}^n |x_i - x_{i-1}|, \quad (13)$$

and $d_2 = 1.128$. In fact, $\frac{\overline{MR}}{d_2}$ is an estimate of the standard deviation σ of x assumed to follow a normal distribution in phase I. Equation 11 is therefore based on a choice of $\gamma = 0.027$. The individual control chart is designed to detect a shift of $\bar{x}$.

Hotelling T^2 control chart If two or more features are monitored at the same time, monitoring these two quantities independently by applying two or more univariate control charts can be very misleading, especially if those features are correlated. On the opposite, the Hotelling T^2 control chart is designed for the monitoring of

several features simultaneously. Consider p features following a p -normal distribution. The Hotelling T^2 control chart monitors the Mahalanobis distance T^2 :

$$T^2 = (x - \bar{x})^T \Sigma^{-1} (x - \bar{x}), \quad (14)$$

where Σ is the $p \times p$ estimated covariance matrix of features, x is the current $p \times 1$ feature vector, and $\bar{x}$ is the $p \times 1$ vector of estimated mean values of x vectors (only the undamaged samples are considered to obtain Σ and $\bar{x}$). Since the Mahalanobis distance is always positive, only the upper control limit UCL based on an F distribution is considered:

$$UCL = \frac{p(m+1)(m-1)}{(m^2 - mp)} F_{\gamma, p, m-p}, \quad (15)$$

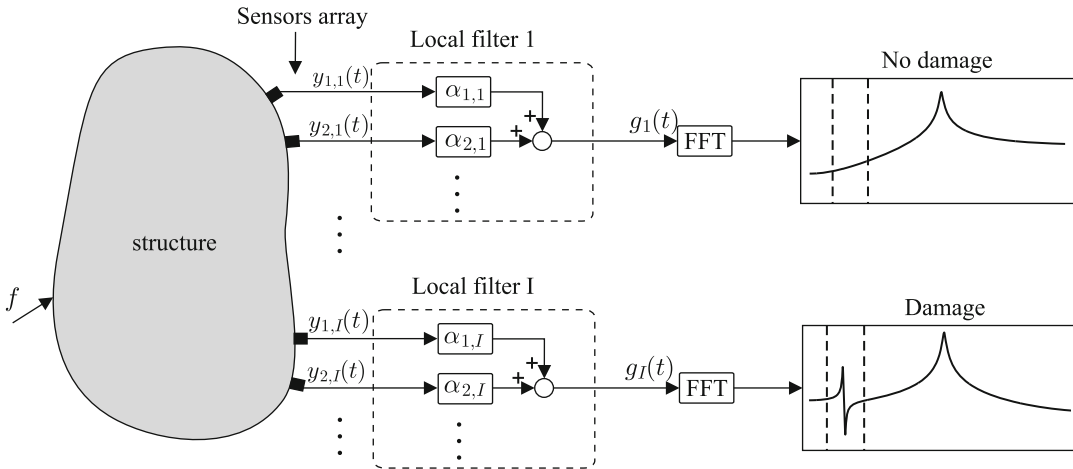
where p is the number of variables, m is the number of samples in the set of data in phase I, and γ is such that there is a probability of $1 - \gamma$ that any in-control sample will fall between the control limits. Like the individual control chart, the Hotelling T^2 control chart detects a change of $\bar{x}$.

Damage Detection and Localization

When condensing all the sensors into a single virtual modal sensor as shown in Fig. 2, the statistical analysis can only give an indication on the deviation from the normal condition on the structure as a whole, leading to damage detection. The methodology can be extended to damage localization: consider now that the n sensors installed on the structure are grouped in several smaller sensor networks, each consisting of m sensors. Modal filters can be built for each of these local sensor networks resulting in independent *local modal filters* (Fig. 6).

If the local network l contains sensors $y_{1,l}, \dots, y_{m,l}$, the output of its modal filter tuned to mode l is given by

$$g_l(t) = \sum_{k=1}^m \alpha_{k,l} y_{k,l}(t), \quad (16)$$



Spatial Filtering for Structural Health Monitoring, Fig. 6 Principle of damage localization using local modal filters

where the $\alpha_{k,I}$ coefficients are computed in order to satisfy the following condition:

$$\sum_{k=1}^m \alpha_{k,I} \phi_{(k,I)i} = \delta_{li}, \quad (17)$$

where $\phi_{(k,I)i}$ is the k th ($k = 1, \dots, m$) component of the i th mode shape projected on the I th local sensor network. If a damage occurs under spatial filter I and if the sensor responses are locally sensitive to damage, the mode shape will only be altered in that spatial filter. As a result, only the spatial filter I will have spurious peaks, indicating the location of the damage.

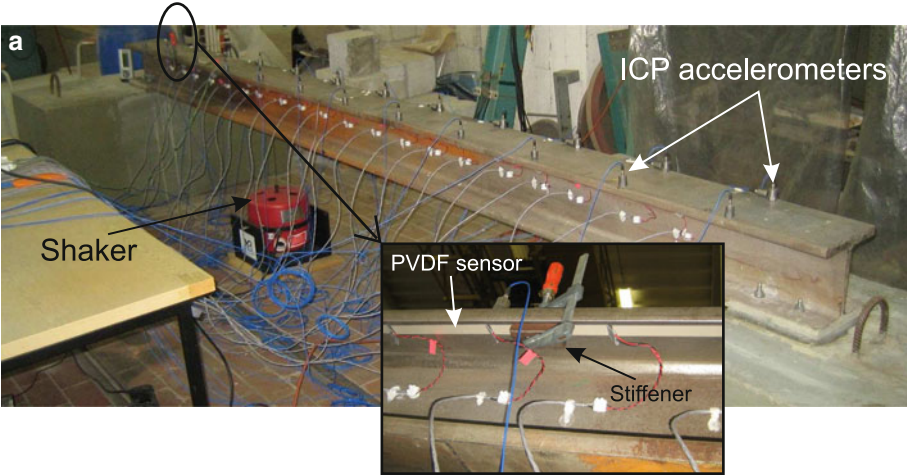
The efficiency of the approach relies therefore on a very strong assumption: damage in a local filter will cause a local change of the mode shape which is limited to the very close vicinity of the damage location. The fulfillment of this requirement depends on the type of measured quantity which is considered. Two different approaches coexist in the literature: in Mendrok and Uhl (2010), accelerometers are used, while in Tondreau and Deraemaeker (2013), dynamic strain sensors are used. These two approaches have been compared in Tondreau and Deraemaeker (2011), showing that the method based on strain sensors has a higher sensitivity to damage and better

localization capabilities. This highlights the importance of the choice of the type of sensor which is part of the first building block of the SHM system. Applications of spatial filtering techniques have been limited so far to accelerometers or dynamic strain sensors.

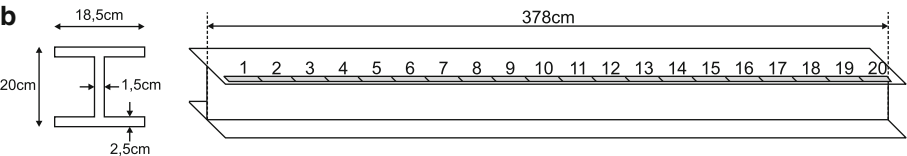
Illustrative Example

Description of the Case Study

The experimental application consists in a 3.78 m long steel I-beam which is bolted on two big concrete cubes. The structure is excited with a Modal 110 electrodynamic shaker from MB Dynamics, and a network of 20 $13 \text{ mm} \times 170 \text{ mm} \times 50 \mu\text{m}$ low-cost PVDF sensors have been fixed with double-coated tape, providing a continuous measurement of the dynamic strains along the beam between sensors 1 and 20. A National Instrument PXIe-1082 data acquisition system is used to measure the sensor responses with a sampling frequency of 6,400 Hz, as well as to generate a band-limited white noise between 0 and 500 Hz (not measured) which drives the shaker. Figure 7 shows the experimental setup as well as the definition of the PVDF sensors (accelerometers installed for preliminary tests can also be seen, but are not used in the present study).



Test setup.



Artist front view: definition of the PVDF sensors.

Spatial Filtering for Structural Health Monitoring, Fig. 7 Experimental setup: 3.78 m steel I-beam equipped with 20 dynamic strain sensors (PVDF) for damage localization

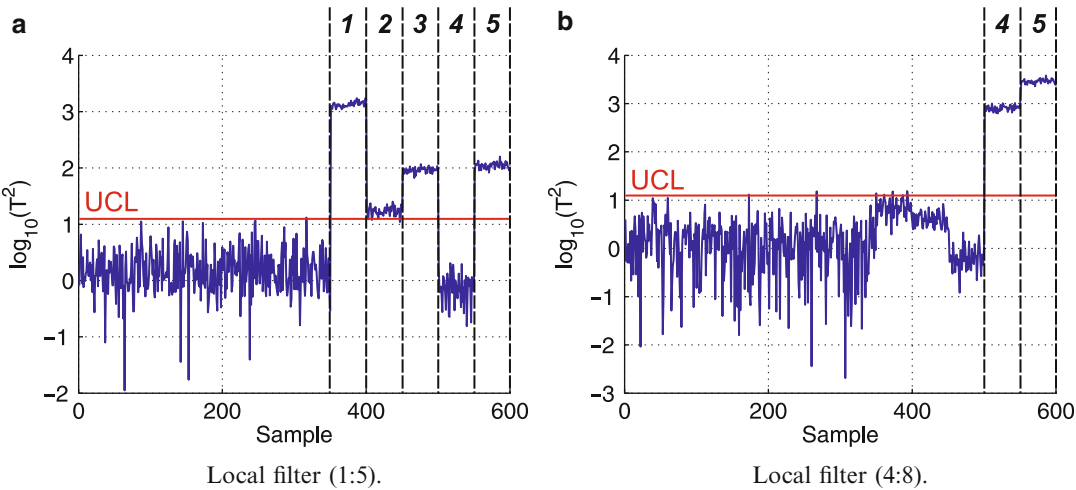
The damage is introduced by fixing a very small steel stiffener (35 mm × 65 mm × 17 mm) directly against one of the PVDF sensors (Fig. 7a) in order to induce a local change of stiffness at that position. It has been checked that such a local change of stiffness induces a local change of strain similar to what happens with damage. The network of 20 sensors is split in five local filters of five sensors, with a small overlap: (i) [1:5], (ii) [4:8], (iii) [8:12], (iv) [12:16], and (v) [16:20]. The damage scenarios are described in Table 1. For the undamaged case, the measurement is performed 350 times in order to infer the statistical properties of the peak indicators. For each damaged case, 50 measurements are performed. Each measurement is referred to as a statistical sample.

The modal filters tuned on the two first bending mode shapes of the beam at 64 and 230 Hz are applied for each of the five local filters separately. The feature vector therefore consists in two peak indicators in each local filter (appearance of peak

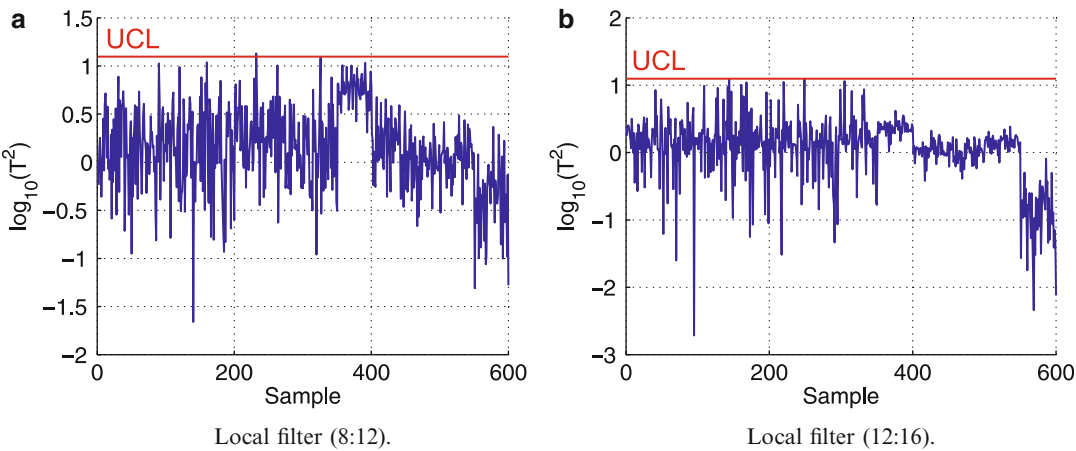
Spatial Filtering for Structural Health Monitoring, Table 1 Damage scenarios

Case	Samples	Location of damage (sensor)	Location of damage (local filter)
1	1–350	None	None
2	351–400	1	[1:5]
3	401–450	2	[1:5]
4	451–500	3	[1:5]
5	501–550	4	[1:5] and [4:8]
6	551–600	5	[1:5] and [4:8]

at 64 Hz for the modal filter tuned to 230 Hz, and appearance of peak at 230 Hz for the modal filter tuned to 64 Hz). Note that the peak indicator used in this example is slightly different from the peak indicator presented in section “[Spatial and Modal Filtering](#)” but shares similar properties. As the feature vector is multivariate, the Hotelling T^2 control chart has been applied to automate the damage localization in each local filter. The first 200 undamaged samples have been considered to



Spatial Filtering for Structural Health Monitoring, Fig. 8 Automated damage detection in local filter (1:5) and local filter (4:8)



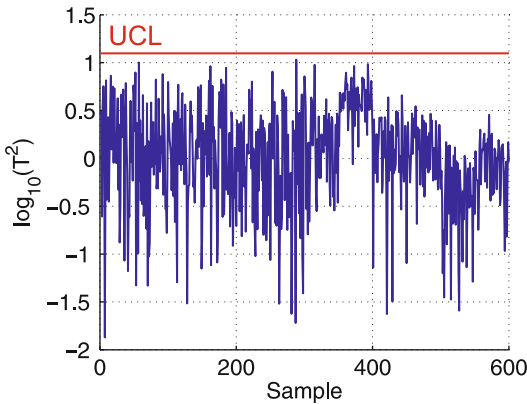
Spatial Filtering for Structural Health Monitoring, Fig. 9 Automated damage detection in local filter (8:12) and local filter (12:16)

estimate the covariance matrix, as well as to compute the control limit (γ is fixed to 0.25 %).

Figures 8, 9, and 10 show the Hotelling T^2 control chart. There is only one missing alarm in local filter [1:5] for a damage at sensor four. However, that missing alarm is compensated thanks to the overlapping of the local filters. Indeed, the damage at sensors four and five is correctly located in local filter [4:8]. The results show that the method has successfully, and automatically, localized all the damage cases.

Summary

There is a strong incentive for the development of online automated SHM techniques for large civil infrastructures. The objective of such systems is to be able to assess the structural integrity of safety critical civil infrastructure in real time. This is particularly important to detect the onset of damage due to aging or more severe damage due to accidental event such as an earthquake or a collision.



Spatial Filtering for Structural Health Monitoring, Fig. 10 Automated damage detection in local filter (16:20)

With the deployment of very large sensor networks on structures, alternatives to modal identification techniques can be interesting when the focus lies in fast and efficient damage detection and localization. This is the aim of the method presented in this chapter. The three important ingredients are (i) the use of a linear combiner to perform data reduction in the time domain through modal filtering, (ii) the transformation of the time-domain output of modal filters to the frequency domain and the subsequent feature extraction to detect the appearance of spurious peaks, and (iii) the use of control charts to automate the damage detection and localization process.

The fully integrated and automated methodology allows to process efficiently data from large sensor networks and to condense it into very limited information for diagnostics in the form of control charts. The efficiency of this technology has been illustrated on a laboratory experiment of a 3.78 m long steel I-beam.

Cross-References

- [Operational Modal Analysis in Civil Engineering: An Overview](#)
- [Stochastic Structural Identification from Vibrational and Environmental Data](#)

- [System and Damage Identification of Civil Structures](#)
- [Vibration-Based Damage Identification: The Z24 Bridge Benchmark](#)

References

- De Neumann S, Andersen JE, Enckell M, Vullo E (2011) Messina bridge – structural health monitoring system. In: Proceedings IABSE-IASS conference. London
- Deraemaeker A, Worden K (2010) New trends in vibration based structural health monitoring. CISM courses and lectures, vol 520. Springer Wien, New York
- Deraemaeker A, Reynders E, De Roeck G, Kullaa J (2008) Vibration-based structural health monitoring using output-only measurements under changing environment. *Mech Syst Signal Pr* 22:34–56
- Doebeling SW, Farrar C, Prime MB (1998) A summary review of vibration-based damage identification methods. *Shock Vib Dig* 30:91–105
- Ewins DJ (1984) Modal testing: theory and practice. Research Studies Press, Letchworth
- Hellier C (2003) Handbook of nondestructive evaluation. McGraw-Hill, New York
- Mendrok K, Uhl T (2010) Experimental verification of the damage localization procedure based on modal filtering. *Struct Health Monit Int J* 20:1615–1630
- Montgomery DC (2009) Statistical quality control: a modern introduction. Wiley, New York
- Ni YQ, Wong KY (2012) Integrating bridge structural health monitoring and condition-based maintenance management. In: Proceedings of the CSHM-4. Berlin
- Reynders E, De Roeck G (2008) Reference-based combined deterministic-stochastic subspace identification for experimental and operational modal analysis. *Mech Syst Signal Pr* 22:617–637
- Reynders E, Houbrechts J, De Roeck G (2012) Fully automated (operational) modal analysis. *Mech Syst Signal Pr* 29:228–250
- Rofidal K (2007) U.S. coast guard responds following the collapse of the I-35W bridge in Minnesota. *USCG Reservist* 54:26–29
- Ryan TP (2000) Statistical methods for quality improvement. Wiley, New York
- Rytter A (1993) Vibration based inspection of civil engineering structures. PhD thesis, University of Aalborg
- Tondreau G, Deraemaeker A (2011) Comparison of damage localization based on modal filters using strain measurements and acceleration measurements. In IWSHM2011, Stanford, Sept 2011
- Tondreau G, Deraemaeker A (2013) Local modal filters for automated data-based damage localization using ambient vibrations. *Mech Syst Signal Pr* 39: 162–180

Spatial Variability of Ground Motion: Seismic Analysis

Mounir Khaled Berrah¹, Djawad Zendagui² and Mustapha Djafour²

¹Ecole Nationale Polytechnique, Algiers, Algeria

²Risam Laboratory, Faculty of Technology, University Abou Bakr Belkaïd, Tlemcen, Algeria

Synonyms

Coherency; Dynamic; Pseudo-static; Seismic analysis; Spatial variability; SVGM

Introduction

The seismic design of structures considers that all ground supports are subjected simultaneously to identical seismic excitation. This assumption is not valid for extended structures since it has been recognized that seismic motion exhibits no negligible variability. This variability, called spatial variability of ground motion (SVGGM), must hence be considered in the design of extended structures.

Many questions arise: what is SVGGM and how does one model it? Does SVGGM increase or decrease internal forces of a structure? Is there any general conclusion that can be drawn while considering SVGGM?

This entry aims at answering these questions. The first part of this entry addresses the description of SVGGM through models in terms of coherency function. Subsequently, the response of structures subjected to SVGGM is derived. At this stage, the concept of pseudo-static and dynamic responses is introduced. To understand the effect of the SVGGM on structures, two case studies are presented: a single degree of freedom (SDOF) system and a multi-degree of freedom (MDOF) system, subjected to SVGGM.

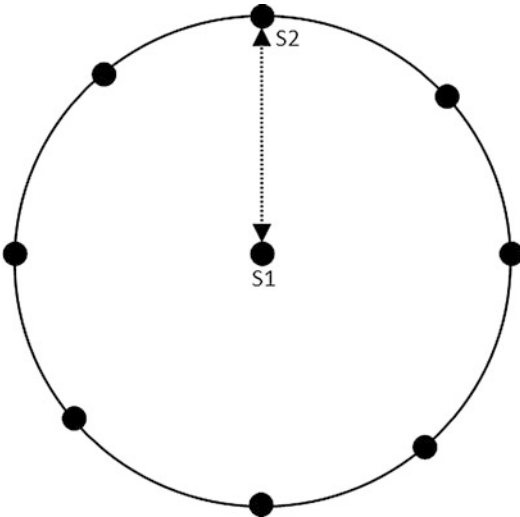
Spatial Variability of Seismic Motion (SVGGM): What It Is and How to Describe It

Seismic Motion and Its Measurement

The seismic motion is regarded as one of the most important and unknown load which acts on a structure. This huge motion results from mainly two causes: (a) explosive volcanic eruptions which are very common in areas of volcanic activity and (b) tectonic activity associated with plate margins and faults. The majority of earthquakes worldwide are of this second type. Thus, the main cause of earthquakes is either the sudden movement of various plate boundaries or when plates scrape against each other. Some earthquakes are also caused by old plate boundaries or faults. The point from which starts this sudden movement is the epicenter of the earthquake. This point could be between 10 and 400 km below the surface of the earth.

Following this sudden movement, many waves radiate from the epicenter and propagate through the earth. It has been recognized that both P and S waves, known as body waves, emanate from the source and travel with a velocity that exceeds 4 km/s in earth. When they arrive at the free field, they are followed by both Love and Rayleigh waves. These latter, known as surface waves, travel only at the surface. The movements associated with body as well as surface waves are well known and are mainly decomposed into compressional and shear movement. At the free field, the seismic motion could be decomposed into two parts: (a) motion caused by body waves and (b) motion caused by surface waves.

The seismic motion is recorded by seismometers which are deployed in seismic areas. The number of seismometers is growing from year to year denoting the need for a better understanding of the seismic motion. While some of them are deployed in a single scheme, others are deployed in grouped schemes. This latter scheme forms what is called the dense seismic array. A typical dense array configuration is presented



Spatial Variability of Ground Motion: Seismic Analysis, Fig. 1 Typical dense array configuration

in Fig. 1 where in general a central station S1 is surrounded by other stations such as station S2 with a separating distance.

Many arrays have been and are being deployed around the world to assess the characteristics of seismic ground motions. Among them SMART-1 array “Strong Motion ARray in Taiwan,” located in Lotung, in the north-east corner of Taiwan deployed in 1980 remains one of the most important. Other dense seismic arrays have been deployed (Zerva 2009). The deployment of these arrays enhanced greatly our knowledge on how the SVGM affects structure and what the main parameters used to describe the SVGM are.

What Is SVEGM?

Let’s consider that the medium is described using the Cartesian coordinate $\mathbf{x} = (x, y, z)$ and $\mathbf{u}(\mathbf{x}, t) = \{u(\mathbf{x}, t) \ v(\mathbf{x}, t) \ w(\mathbf{x}, t)\}^T$ is the seismic motion measured along these axis. Hence, $v(\mathbf{x}, t)$ is the vertical motion, whereas $u(\mathbf{x}, t)$ and $w(\mathbf{x}, t)$ are the horizontal motions. At any point of the soil, both horizontal and vertical motions could be recorded. For instance, let’s consider two points A and B at the free surface. Recorded earthquake along a particular axis at two points is obviously different in terms of amplitude and frequency. This difference is well accepted if

the separating distance between those points is somewhat important. However, data from dense seismic array show that even if the separating distance between two points is not very important (sometimes less than 50 m), the recorded motion at both points exhibits some differences.

The effect of this difference, called spatial variability of earthquake motion (SVEGM), on extended structures is addressed in this entry. Other sources of variability such as the result from relative surface fault motion for recording stations located on either side of a causative fault, soil liquefaction, and landslides (Zerva and Zervas 2002) are not considered herein. Finally, although SVEGM has been extensively studied for horizontal component and at free surface, some dense seismic arrays have been instrumented to record motion with respect to vertical component and/or at depth.

What causes SVEGM? It has been recognized that SVEGM is caused by several effects (Der Kiureghian 1996) (Fig. 1):

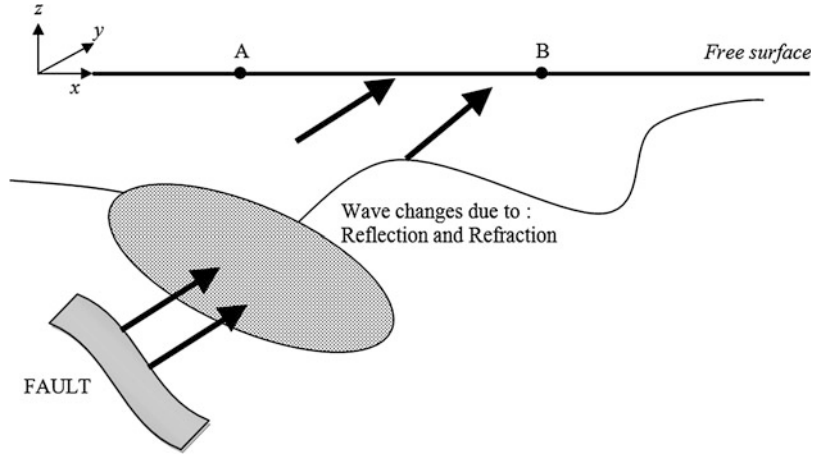
1. Incoherence effects due to scattering in the heterogeneous ground and extended source effects.
2. Traveling-wave effects, in which nonvertical waves reach different points on the ground surface at different times, producing a time shift between the motions at those points; this effect is also known as “wave passage effect.”
3. Site effects due to the variation by the filtering effects of overlying soil columns; this effect is known as “site effect.”

Although the attenuation of seismic motion increases the variability of seismic motion, it is not usually included since its effect on structures will not be visible below a separating distance of a dozen kilometers, which is greater than the usual separating distance for structures.

How to Measure and Describe SVEGM

For the purpose of simplicity, two points A and B located at the free surface with $z = 0$ are considered. Thus, coordinates of both points are, respectively, $\mathbf{x}_A = (x_A, 0, 0)$ and $\mathbf{x}_B = (x_B, 0, 0)$. Even if the motion could be recorded along the

Spatial Variability of Ground Motion: Seismic Analysis,
Fig. 2 Propagation of seismic waves



three axes, only the horizontal motion which is along the x axis, i.e., $u(\mathbf{x}_A, t)$ and $u(\mathbf{x}_B, t)$, is considered herein (Fig. 2).

The SVEGM is described by models in terms of correlation or coherence functions. This latter, which is the most used, is the ratio of the cross-spectral density function of the motion at two points separated by a horizontal offset to the square root of the product between the auto-spectral density functions at the two points:

$$\rho_{AB}(\omega) = \frac{S_{AB}(\omega)}{\sqrt{S_A(\omega)S_B(\omega)}} \quad (1)$$

where

$$L = x_B - x_A,$$

$S_{AB}(\omega)$ is the cross-spectral density function of the motion along the x axis for the two points A and B.

$S_A(\omega)$ and $S_B(\omega)$ are the auto-spectral density functions of the motion along the x axis at, respectively, points A and B.

The coherency function is often written as

$$\rho_{AB}(\omega) = |\rho_{AB}(\omega)| \exp[i\theta_{AB}(\omega)] \quad (2)$$

with

$$\theta_{AB}(\omega) = \tan^{-1} \left(\frac{\text{Im}(S_{AB}(\omega))}{\text{Re}(S_{AB}(\omega))} \right) \quad (3)$$

where

$\text{Im}(S_{AB}(\omega))$ and $\text{Re}(S_{AB}(\omega))$ stands for imaginary and real parts of $S_{AB}(\omega)$.

$|\rho_{AB}(\omega)|$ is the lagged coherency and $\text{Re}(S_{AB}(\omega))$ as unlagged coherency. It is worth noting that the coherency function is complex, and lagged coherency is usually called coherence function. This latter describes the similarity of the waveforms at two stations without taking into account the difference in the arrival times of the waves.

If the subscript A and B are skipped and replaced by the separating distance L , then the coherency function is rewritten as

$$\rho(L, \omega) = |\rho(L, \omega)| \exp[i\theta(L, \omega)] \quad (4)$$

This coherency function measures the coherency along the x axis and at the free surface. It could be measured at depth by, for instance, considering that both points A and B are located at the same depth, i.e., $y = -h$, which leads to the following coherency function

$$\rho(h, L, \omega) = |\rho(h, L, \omega)| \exp[i\theta(h, L, \omega)] \quad (5)$$

Models in terms of coherency functions have been developed during last decades (Abrahamson et al. 1991; Der Kiureghian 1996; Zendagui et al. 1999). They could be divided into three categories: empirical, semiempirical, and physically based models. An exhaustive list of these models could be found in Zerva (2009). The common feature of these models is that the

coherency function decreases with separating distance and frequency.

Generation of Variable Seismic Excitation and Response Spectra

The generation of variable seismic excitation has been extensively studied. The method developed by Shinozuka (1971) remains the most popular. This method has been extended and modified by others to take into account nonstationarity of processes (Deodatis 1996). A spectral-representation-based simulation algorithm has been developed by Deodatis (1996). The simulation uses the coherency function and allows the generation of acceleration as well as displacement time histories. This latter is necessary while performing non linear structural analysis. A recent method developed by Benmansour et al. (2012) solves the problem of integrability of seismic motion to avoid the use of baseline correction.

Alternatively, Berrah and Kausel (1992) developed an interesting method that modifies a particular response spectrum to take into account SVGM.

Structure Subjected to SVGM: General Derivation

Consider a structure with N DOFs. For the analysis of such a system, consider that the displacement vector contains two parts: (1) V includes the N DOFs of the superstructure, and (2) v_s contains the N_g components of support displacements. The equation of dynamic equilibrium is written as

$$\begin{bmatrix} m & m_s \\ m_s^T & m_{ss} \end{bmatrix} \begin{Bmatrix} \ddot{V} \\ \ddot{v}_s \end{Bmatrix} + \begin{bmatrix} c & c_s \\ c_s^T & c_{ss} \end{bmatrix} \begin{Bmatrix} \dot{V} \\ \dot{v}_s \end{Bmatrix} + \begin{bmatrix} k & k_s \\ k_s^T & k_{ss} \end{bmatrix} \begin{Bmatrix} V \\ v_s \end{Bmatrix} = \begin{Bmatrix} 0 \\ R(t) \end{Bmatrix} \quad (6)$$

The mass, damping, and stiffness matrices can be determined from properties of the structure, while the support motions $\ddot{v}_s(t)$, $\dot{v}_s(t)$, and $v_s(t)$ must be specified. It is desired to determine the displacements V in the superstructure DOF and the support forces $R(t)$.

To write the governing equations in a form similar to that of the case of a single excitation, the displacements are separated into two parts:

$$\begin{Bmatrix} V \\ v_s \end{Bmatrix} = \begin{Bmatrix} u_s \\ v_s \end{Bmatrix} + \begin{Bmatrix} U \\ 0 \end{Bmatrix} \quad (7)$$

In this equation, u_s is the vector of structural displacements due to static application of prescribed support displacements v_s at each time instant. The two parts are related as

$$\begin{bmatrix} k & k_s \\ k_s^T & k_{ss} \end{bmatrix} \begin{Bmatrix} u_s \\ v_s \end{Bmatrix} = \begin{Bmatrix} 0 \\ R_s(t) \end{Bmatrix} \quad (8)$$

$R_s(t)$ are the support forces necessary to statically impose displacements v_s that vary with time; obviously, u_s varies with time and is therefore known as the vector of quasi-static displacements. Observe that $R_s(t) = 0$ in the case of identical support ground motion. The displacements of the superstructure U are known as dynamic displacements.

With the total structural displacements split into quasi-static and dynamic displacements, consider hence the first of the two partitioned equations (Eq. 6):

$$m\ddot{V} + m_s\ddot{v}_s + c\dot{V} + c_s\dot{v}_s + kV + k_sv_s = 0 \quad (9)$$

Substituting Eq. 7 and transferring all terms involving v_s and u_s to the right-hand side lead to

$$m\ddot{U} + c\dot{U} + kU = P_{eff}(t) \quad (10)$$

where the vector of effective earthquake forces is

$$P_{eff}(t) = -(m\ddot{u}_s + m_s\ddot{v}_s + c\dot{u}_s + c_s\dot{v}_s + ku_s + k_sv_s) \quad (11)$$

Equation 8 gives

$$ku_s + k_sv_s = 0 \quad (12)$$

This relation also enables us to express the quasi-static displacement u_s in terms of the specified support displacements v_s :

$$u_s = [\Lambda]v_s \quad (13)$$

where $[\Lambda] = -k^{-1}k_s$ is the influence matrix.

Equation 13 can be written in a different form as

$$u_s = \sum_{i=1}^{N_s} \Lambda_i v_{si}(t) \quad (14)$$

where Λ_i is the i th column of the influence matrix $[\Lambda]$.

Substituting Eqs. 12 and 13 in Eq. 11, one gets

$$P_{eff}(t) = -(m[\Lambda] + m_s)\ddot{v}_s(t) - (c[\Lambda] + c_s)\dot{v}_s \quad (15)$$

For many practical applications, further simplification of P_{eff} is possible in two stages. First, the damping term is usually small relative to the inertia term and can be dropped. Second, for structures with mass idealized as lumped at the DOFs, the mass matrix is diagonal, implying that m_s is a null matrix and m is diagonal. With these simplifications, P_{eff} is expressed as

$$P_{eff}(t) = -m[\Lambda]\ddot{v}_s(t) \quad (16)$$

By using Eq. 14, P_{eff} can be expressed as

$$P_{eff}(t) = -\sum_{i=1}^{N_s} m\Lambda_i \ddot{v}_{si}(t) \quad (17)$$

The dynamic response U can be computed by modal analysis, as the superposition of the modal contributions

$$U(t) = \sum_{n=1}^N \phi_n q_n(t) \quad (18)$$

where ϕ_n are the natural modes and q_n are scalar multipliers called modal coordinates. Instead of Eq. 18, the modal equation can be expressed for the n th mode as

$$\ddot{q}_n + 2\xi_n \omega_n \dot{q}_n + \omega_n^2 q_n = -\sum_{i=1}^{N_s} \Gamma_{ni} \ddot{v}_{si}(t) \quad (19)$$

where

$$\Gamma_{ni} = \frac{L_{ni}}{M_n}, \quad L_{ni} = \phi_n^T m \Lambda_i \quad \text{and} \quad M_n = \phi_n^T m \phi_n \quad (20)$$

The solution of Eq. 19 can be written as

$$q_n(t) = \sum_{i=1}^{N_s} \Gamma_{ni} D_{ni}(t) \quad (21)$$

where $D_{ni}(t)$ is the displacement response of the n th-mode SDOF system to support acceleration $\ddot{v}_{si}(t)$.

The displacement response of the structure (Eq. 7) contains two parts:

1. The dynamic displacements are obtained by combining Eqs. 18 and 21:

$$U(t) = \sum_{i=1}^{N_s} \sum_{n=1}^N \Gamma_{ni} \phi_n D_{ni}(t) \quad (22)$$

2. The quasi-static displacements u_s are given by Eq. 14.

Combining the two parts gives the total displacements of the structure

$$V(t) = \sum_{i=1}^{N_s} \Lambda_i v_{si}(t) + \sum_{i=1}^{N_s} \sum_{n=1}^N \Gamma_{ni} \phi_n D_{ni}(t) \quad (23)$$

The equivalent static forces in structural DOFs are given by

$$F = kV + k_s v_s \quad (24)$$

Substituting Eq. 7 for V and using Eq. 12 give

$$F = kU(t) \quad (25)$$

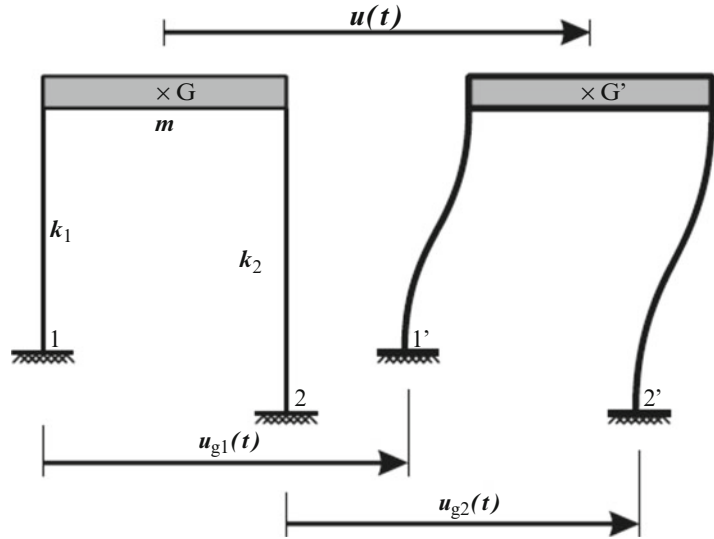
These forces depend on the dynamic displacements only given by Eq. 22. Therefore,

$$F(t) = \sum_{i=1}^{N_s} \sum_{n=1}^N \Gamma_{ni} k \phi_n D_{ni}(t) \quad (26)$$

The equivalent static forces along support DOFs are also given by the last term on the left-hand side of Eq. 6:

$$F_s = k_s^T V + k_{ss} v_s \quad (27)$$

Spatial Variability of Ground Motion: Seismic Analysis, Fig. 3 A SDOF system subjected to SVG



Substituting Eq. 7 and using Eq. 8 for the quasi-static support forces $R_s(t)$ give

$$F_s(t) = k_s^T U(t) + R_s(t) \quad (28)$$

Application 1: SDOF System Subjected to SVG

Derivation of the Relative Displacement

Consider a simple multi-support structure represented by a SDOF system with two supports 1 and 2 excited by, respectively, $u_{g1}(t)$ and $u_{g2}(t)$ (Fig. 3). The columns are assumed to be weightless and inextensible in the vertical (axial) direction, and the resistance to girder displacement provided by each column is represented by its spring constant k_1 and k_2 . For the purpose of simplicity, it is assumed that the slab, which has a mass m , is infinitely rigid and that damping is neglected. The total displacement of the mass from the initial position G to the final position G' is $u(t)$

The forces acting on the slab:

- Inertial forces: $f_I = m\ddot{u}$.
- Elastic force on column (1): $f_{e1} = k_1(u - u_{g1})$
- Elastic force on column (2): $f_{e2} = k_2(u - u_{g2})$.

Hence, the equation of motion is an expression of the equilibrium of these forces given by

$$m\ddot{u} + (k_1 + k_2)u = k_1 u_{g1}(t) + k_2 u_{g2}(t) \quad (29)$$

The natural frequency is defined by

$$\omega = \sqrt{\frac{k_1 + k_2}{m}} \quad (30)$$

The quasi-static displacement is thus

$$u_s(t) = \frac{k_1 u_{g1}(t) + k_2 u_{g2}(t)}{k_1 + k_2} \quad (31)$$

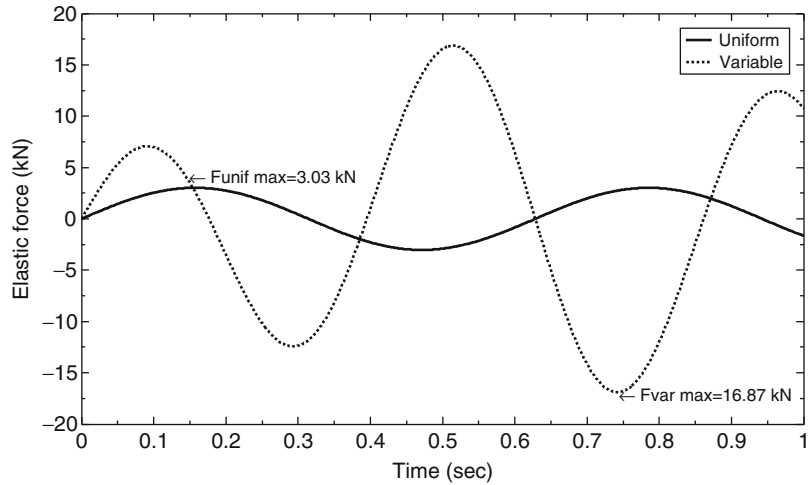
At this stage, it is considered that the support motions are harmonic variable both in amplitude and phase:

$$\begin{cases} u_{g1}(t) = A_1 \sin \bar{\omega}_1 t \\ u_{g2}(t) = A_2 \sin \bar{\omega}_2 t \end{cases} \quad (32)$$

The total dynamic and quasi-static displacements are hence

$$u(t) = \frac{k_1}{k_1 + k_2} \frac{A_1}{\left(1 - \left(\frac{\omega_1}{\omega}\right)^2\right)} \sin \bar{\omega}_1 t + \frac{k_2}{k_1 + k_2} \frac{A_2}{\left(1 - \left(\frac{\omega_2}{\omega}\right)^2\right)} \sin \bar{\omega}_2 t \quad (33)$$

Spatial Variability of Ground Motion: Seismic Analysis, Fig. 4 Variation of elastic forces at column 1 for case S-1



$$u_s(t) = \frac{k_1}{k_1 + k_2} A_1 \sin \bar{\omega}_1 t + \frac{k_2}{k_1 + k_2} A_2 \sin \bar{\omega}_2 t \quad (34)$$

Finally, the dynamic response could be derived as follows:

$$U(t) = u(t) - u_s(t) \quad (35)$$

Case Study

As the purpose of this entry is to see how SVGM affects a structure, it will be considered hereafter two cases: a uniform seismic motion at supports 1 and 2 equal to u_{g1} and a variable seismic motion, i.e., u_{g1} at support 1 and u_{g2} at support 2. The natural frequency is equal to 24 rad/s . The results are presented in terms of elastic forces of the column (1). For the case of spatially variable ground motion, two cases will be considered:

$$\text{Case S-1} \quad \begin{cases} u_{g1} = 0.1 \sin 10t \\ u_{g2} = 0.1 \sin 15t \end{cases} (m) \quad (36)$$

$$\text{Case S-2} \quad \begin{cases} u_{g1} = 0.1 \sin 20t \\ u_{g2} = 0.1 \sin 25t \end{cases} (m) \quad (37)$$

The frequencies of excitation have been chosen below and above the natural frequency. It is interesting to see that for both cases (Eqs. 36 and 37), SVGM (dashed line) exhibits greater

value than uniform load (continuous line) (Figs. 4 and 5). Thus, for these cases, it seems that SVGM induces greater values of elastic forces than uniform seismic motion.

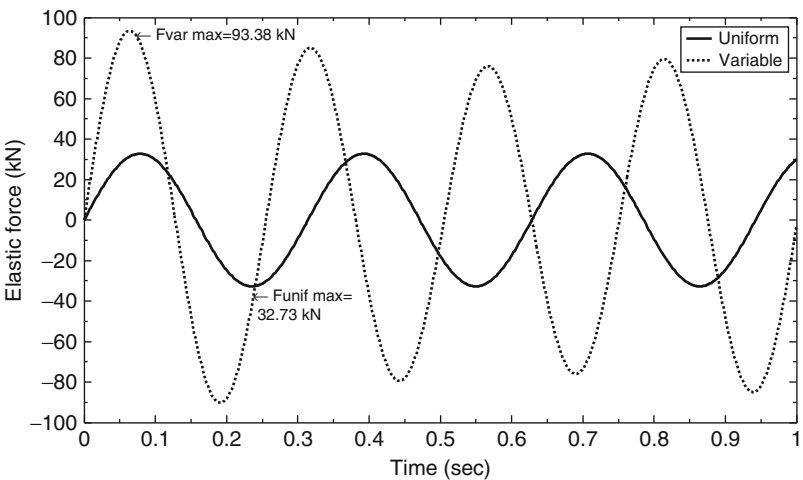
What about the variation of the quasi-static and dynamic components for both cases? Results show great difference on the part of both components depending on frequency of excitations. Hence, by analyzing the variations of the two components, quasi-static and dynamic (Figs. 6 and 7), it is found that for case S-1 the amplification caused by variable motion is mainly due to the importance of the quasi-static component. In this case, the frequency of the structure is larger than those of the imposed displacement. The SDOF system could be considered herein as a rigid structure. However, for case S-2, the frequency of the structure is close to the frequency of the excitation. In this case, the dynamic component is the main cause of the amplification of the response.

Application 2: MDOF System Subjected to SVGM

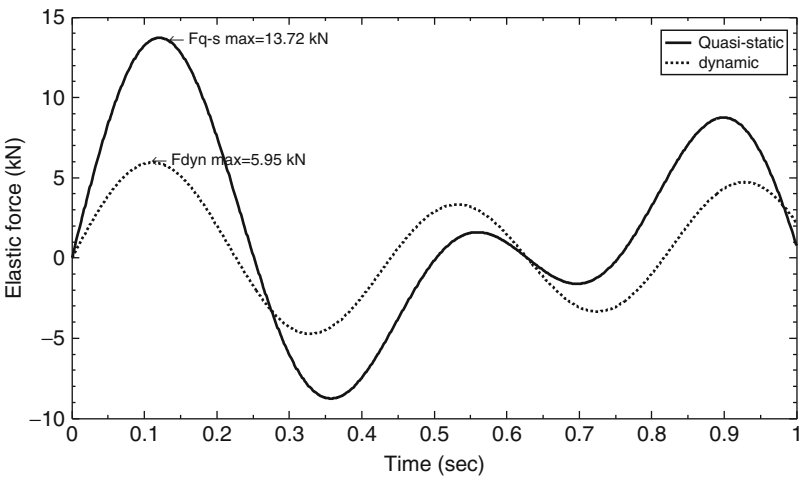
Derivation of the Relative Displacements

In this part, responses of a MDOF system under uniform and variable seismic input will be computed. It is assumed herein that the MDOF is composed of five masses connected by five springs (Fig. 8). The motion, in terms of

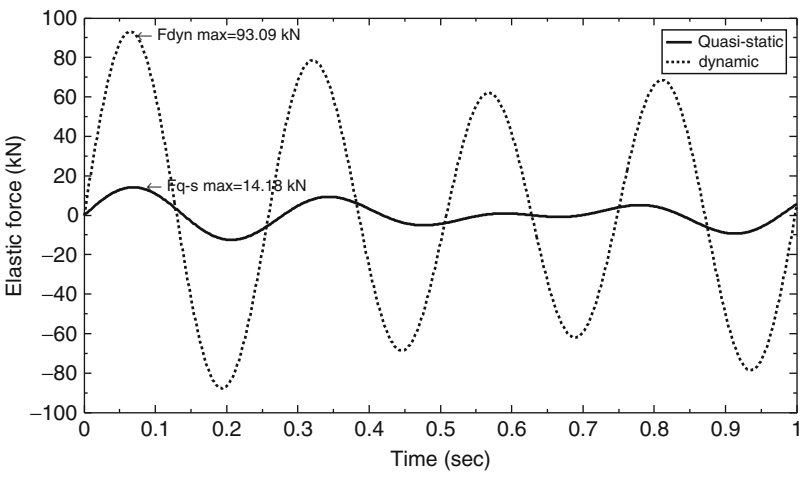
Spatial Variability of Ground Motion: Seismic Analysis, Fig. 5 Variation of elastic forces at column 1 for case S-2

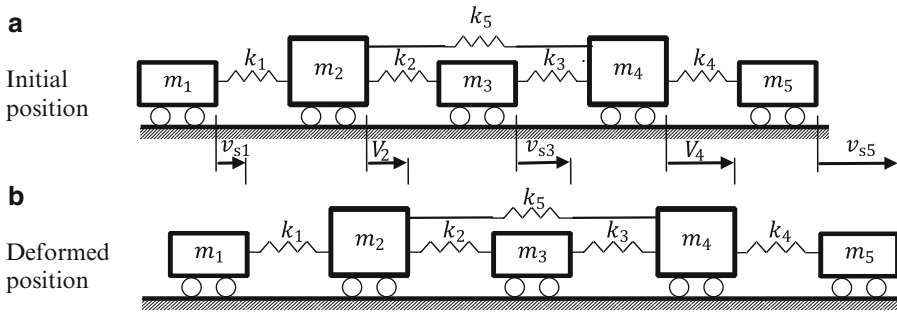


Spatial Variability of Ground Motion: Seismic Analysis, Fig. 6 Variation of the pseudo-static and dynamic components of the elastic forces at column 1 for case S-1



Spatial Variability of Ground Motion: Seismic Analysis, Fig. 7 Variation of the pseudo-static and dynamic components of the elastic forces at column 1 for case S-2





Spatial Variability of Ground Motion: Seismic Analysis, Fig. 8 MDOF system under SVG

displacements, is imposed at the outer masses and the middle one, v_{s1} , v_{s3} et v_{s5} . Hence, only two masses do not have imposed displacements and their displacements are noted: V_2 et V_4 . Thus, the displacement vector could be decomposed

into two parts $\{v_s\} = \langle v_{s1} \ v_{s3} \ v_{s5} \rangle^T$ and $\{V\} = \langle V_2 \ V_4 \rangle^T$.

The equation of dynamic equilibrium for all the DOFs is written as

$$\begin{bmatrix} m_1 & & & & \\ & m_2 & & & \\ & & m_3 & & \\ & & & m_4 & \\ & & & & m_5 \end{bmatrix} \begin{Bmatrix} \ddot{v}_{s1} \\ \ddot{V}_2 \\ \ddot{v}_{s3} \\ \ddot{V}_4 \\ \ddot{v}_{s5} \end{Bmatrix} + \begin{bmatrix} k_1 & -k_1 & & & \\ -k_1 & k_1 + k_2 + k_5 & -k_2 & & \\ & -k_2 & k_2 + k_3 & -k_3 & \\ & & -k_3 & k_3 + k_4 + k_5 & -k_4 \\ & & & -k_4 & -k_4 \end{bmatrix} \begin{Bmatrix} v_{s1} \\ V_2 \\ v_{s3} \\ V_4 \\ v_{s5} \end{Bmatrix} = \begin{Bmatrix} R_1 \\ 0 \\ R_3 \\ 0 \\ R_5 \end{Bmatrix} \quad (38)$$

By considering only DOFs 2 and 4,

$$\begin{bmatrix} m_2 & 0 \\ 0 & m_4 \end{bmatrix} \begin{Bmatrix} \ddot{V}_2 \\ \ddot{V}_4 \end{Bmatrix} + \begin{bmatrix} k_1 + k_2 + k_5 & -k_5 \\ -k_5 & k_3 + k_4 + k_5 \end{bmatrix} \begin{Bmatrix} V_2 \\ V_4 \end{Bmatrix} = \begin{Bmatrix} k_1 v_{s1} + k_2 v_{s3} \\ k_3 v_{s3} + k_4 v_{s5} \end{Bmatrix} = \begin{bmatrix} k_1 & k_2 & 0 \\ 0 & k_3 & k_4 \end{bmatrix} \begin{Bmatrix} v_{s1} \\ v_{s3} \\ v_{s5} \end{Bmatrix} \quad (39)$$

The pseudo-static component is obtained by neglecting the inertial forces

$$0 + \begin{bmatrix} k_1 + k_2 + k_5 & -k_5 \\ -k_5 & k_3 + k_4 + k_5 \end{bmatrix} \begin{Bmatrix} u^s_2 \\ u^s_4 \end{Bmatrix} = \begin{bmatrix} k_1 & k_2 & 0 \\ 0 & k_3 & k_4 \end{bmatrix} \begin{Bmatrix} v_{s1} \\ v_{s3} \\ v_{s5} \end{Bmatrix} \quad (40)$$

Thus,

$$\{u^s\} = \begin{Bmatrix} u^s_2 \\ u^s_4 \end{Bmatrix} = \begin{bmatrix} k_1 + k_2 + k_5 & -k_5 \\ -k_5 & k_3 + k_4 + k_5 \end{bmatrix}^{-1} \begin{bmatrix} k_1 & k_2 & 0 \\ 0 & k_3 & k_4 \end{bmatrix} \begin{Bmatrix} v_{s1} \\ v_{s3} \\ v_{s5} \end{Bmatrix} \quad (41)$$

If $k_i = k$

$$\{u^s\} = \begin{Bmatrix} u^s_2 \\ u^s_4 \end{Bmatrix} = \frac{1}{8} \begin{bmatrix} 3 & 4 & 1 \\ 1 & 4 & 3 \end{bmatrix} \begin{Bmatrix} v_{s1} \\ v_{s3} \\ v_{s5} \end{Bmatrix} \quad (42)$$

The equation of movement could be rewritten in terms of the dynamic component $\{U\} = \{V\} - \{u^s\}$:

$$\begin{aligned}
& \begin{bmatrix} m_2 & 0 \\ 0 & m_4 \end{bmatrix} \begin{Bmatrix} \ddot{U}_2 \\ \ddot{U}_4 \end{Bmatrix} + \begin{bmatrix} k_1 + k_2 + k_5 & -k_5 \\ -k_5 & k_3 + k_4 + k_5 \end{bmatrix} \begin{Bmatrix} U_2 \\ U_4 \end{Bmatrix} \\
&= \begin{bmatrix} m_2 & 0 \\ 0 & m_4 \end{bmatrix} \begin{bmatrix} k_1 + k_2 + k_5 & -k_5 \\ -k_5 & k_3 + k_4 + k_5 \end{bmatrix}^{-1} \\
& \begin{bmatrix} k_1 & k_2 & 0 \\ 0 & k_3 & k_4 \end{bmatrix} \begin{Bmatrix} \ddot{v}_{s1} \\ \ddot{v}_{s3} \\ \ddot{v}_{s5} \end{Bmatrix}
\end{aligned} \quad (43)$$

$$\begin{Bmatrix} v_{s1} = 0.1 \sin 30t \\ v_{s3} = 0.1 \sin 35t \\ v_{s5} = 0.1 \sin 30t \end{Bmatrix} (m) \quad (47)$$

Case M-3 variability in terms of temporal delay and frequency close to those of the system:

$$\begin{Bmatrix} v_{s1} = 0.1 \sin 30t \\ v_{s3} = 0.1 \sin 30(t - 0.03) \\ v_{s5} = 0.1 \sin 30(t - 0.06) \end{Bmatrix} (m) \quad (48)$$

Case M-4 variability in terms of temporal delay and frequency greater than of the system:

$$\begin{Bmatrix} v_{s1} = 0.1 \sin 50t \\ v_{s3} = 0.1 \sin 50(t - 0.03) \\ v_{s5} = 0.1 \sin 50(t - 0.06) \end{Bmatrix} (m) \quad (49)$$

If $m_i = m$ and $k_i = k$, then

$$\begin{aligned}
& m \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{Bmatrix} \ddot{U}_2 \\ \ddot{U}_4 \end{Bmatrix} + k \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix} \begin{Bmatrix} U_2 \\ U_4 \end{Bmatrix} \\
&= -\frac{m}{8} \begin{bmatrix} 3 & 4 & 1 \\ 1 & 4 & 3 \end{bmatrix} \begin{Bmatrix} \ddot{v}_{s1} \\ \ddot{v}_{s3} \\ \ddot{v}_{s5} \end{Bmatrix}
\end{aligned} \quad (44)$$

The natural modes are

$$\{\phi_1\} = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} \text{ and } \{\phi_2\} = \begin{Bmatrix} 1 \\ -1 \end{Bmatrix} \quad (45)$$

Case Study

In this part, elastic forces at the spring k_1 obtained by considering uniform imposed displacements will be compared to those obtained when imposing variable displacements v_{s1} , v_{s3} et v_{s5} . For the sake of simplicity, it will be considered that the uniform displacement is set equal to v_{s1}

For the purpose of study, let's consider that

$$\begin{aligned}
m &= 1t, \quad k = 600kN/m \Rightarrow \\
\omega_1 &= 34.64rad/s \text{ et } \omega_2 = 49rad/s
\end{aligned}$$

For the case of spatially variable ground motion, four cases will be considered:

Case M-1 variability in terms of maximum amplitude:

$$\begin{Bmatrix} v_{s1} = 0.10 \sin 30t \\ v_{s3} = 0.15 \sin 30t \\ v_{s5} = 0.30 \sin 30t \end{Bmatrix} (m) \quad (46)$$

Case M-2 variability in terms of phases:

It is worth noting that these cases model incoherence effect (Eqs. 46 and 47) and wave passage effect (Eqs. 48 and 49). Results obtained in terms of elastic forces at spring 1 for the four above cases are presented in Figs. 9, 10, 11, and 12.

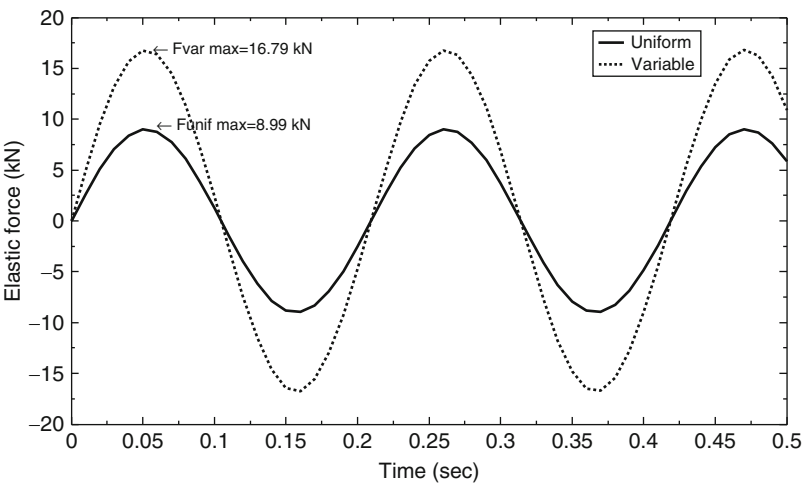
It is observed that SVGm caused by incoherence effect leads to values of elastic forces greater than those obtained for uniform loading (Figs. 9 and 10). This could be readily observed while varying amplitude (Eq. 46, Fig. 9) or frequency (Eq. 47, Fig. 10). In turn, SVGm caused by wave passage effect does not yield greater values of elastic forces than uniform loading for the case where the frequency of excitation is close to the natural frequency of the structure (Eq. 48, Fig. 11). However, this conclusion is not valid for case M-4 where the excitation frequency is greater than the natural frequency of the system (Fig. 12).

Summary

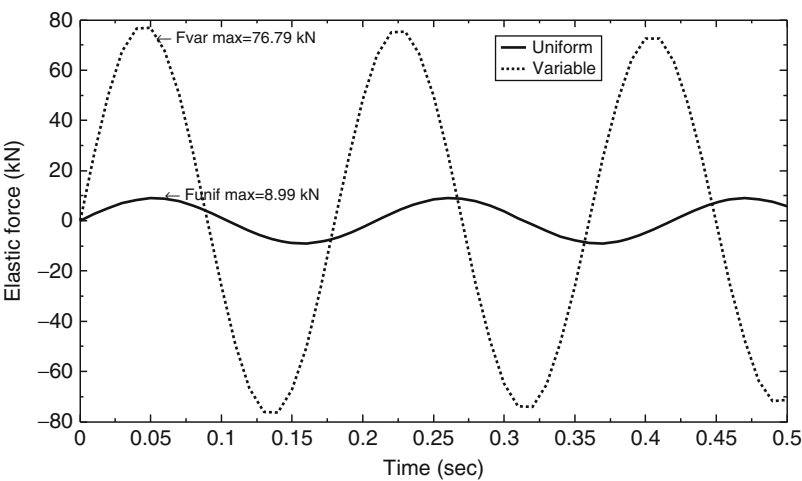
This entry aims at describing the modeling of SVGm and its effects on structures. The total *force* is divided into two components: pseudo-static and dynamic forces. After deriving a general formulation for the case of MDOF systems, two cases have been considered: SDOF and MDOF systems.

It has been shown that SVGm could increase or decrease the elastic forces for both cases.

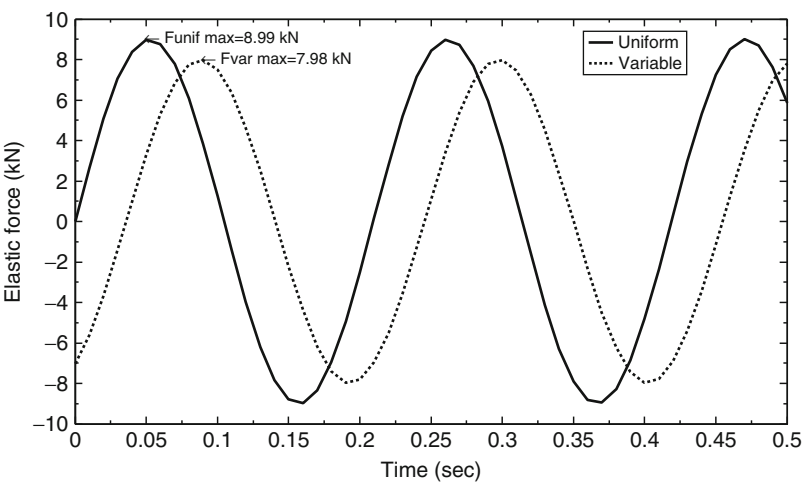
Spatial Variability of Ground Motion: Seismic Analysis, Fig. 9 Variation of elastic forces at spring 1 for case M-1



Spatial Variability of Ground Motion: Seismic Analysis, Fig. 10 Variation of elastic forces at spring 1 for case M-2

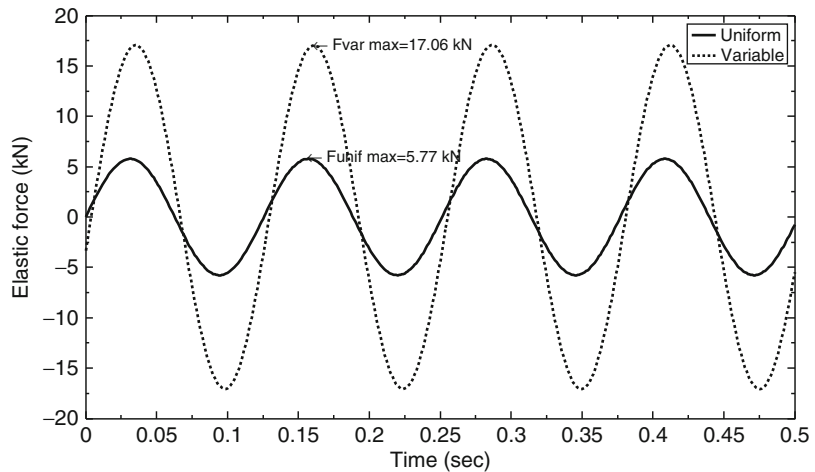


Spatial Variability of Ground Motion: Seismic Analysis, Fig. 11 Variation of elastic forces at spring 1 for case M-3



Spatial Variability of Ground Motion: Seismic Analysis,

Fig. 12 Variation of elastic forces at spring 1 for case M-4



Cross-References

- [Design of Cable-Supported Bridges: Control Strategies](#)
- [Earthquake Location](#)
- [Earthquake Magnitude Estimation](#)
- [Engineering Characterization of Earthquake Ground Motions](#)
- [Physics-Based Ground-Motion Simulation](#)
- [Random Process as Earthquake Motions](#)
- [Response-Spectrum-Compatible Ground Motion Processes](#)
- [Selection of Ground Motions for Response History Analysis](#)
- [Stochastic Ground Motion Simulation](#)

References

- Abrahamson N, Schneider J, Stepp J (1991) Empirical spatial coherency functions for applications to soil-structure interaction analyses. *Earthquake Spectra* 7(1):1–27
- Benmansour N, Djafour M, Bekkouche A, Zendagui D (2012) Seismic response evaluation of bridges under differential ground motion: a comparison with the new Algerian provisions. *Eur J Environ Civil Eng* 16(7):1–19
- Berrah MK, Kausel E (1992) Response spectrum analysis of structures subjected to spatially varying motions. *Earthq Eng Struct Dyn* 21(6):461–470
- Deodatis G (1996) Non-stationary stochastic vector processes: seismic ground motion applications. *Probab Eng Mech* 11(3):149–167
- Der Kiureghian A (1996) Coherence model for spatially varying ground motions. *Earthq Eng Struct Dyn* 25(1):99–111
- Shinozuka M (1971) Simulation of multivariate and multi dimensional random processes. *J Acoust Soc Am* 49(1):357–367
- Zendagui D, Berrah M, Kausel E (1999) Stochastic de-amplification of spatially varying ground motions. *Soil Dyn Earthq Eng* 18(6):395–471
- Zerva A (2009) Spatial variation of seismic ground motions. Modeling and engineering applications. CRC Press, Boca Raton
- Zerva A, Zervas V (2002) Spatial variation of seismic ground motions: an overview. *Appl Mech Rev* 55(3):271–296

Spectral Finite Element Approach for Structural Dynamics

Priscilla Brandão Silva, Danilo Beli and José Roberto de França Arruda
Department of Computational Mechanics,
Faculty of Mechanical Engineering, UNICAMP,
Campinas, SP, Brazil

Synonyms

Forced response; Spectral element; Structural dynamics; Wave propagation

Electronic supplementary material: The online version of this chapter (doi:[10.1007/978-3-642-35344-4_284](https://doi.org/10.1007/978-3-642-35344-4_284)) contains supplementary material, which is available to authorized users.

Introduction

Earthquake disaster in the past has motivated studies on earthquake-resistant buildings to save lives and resources (Newmark and Hall 1982). With this objective, physical models and numerical methods are used to predict the dynamic behavior of structures. The numerical methods provide quantitative analyses of physical phenomena. They can be used to design structures with geometry and materials for an adequate dynamic behavior under seismic excitation. The most frequently used in structural dynamics are the finite element method (FEM) and the boundary element method (BEM).

Based on wave propagation, the spectral finite element or spectral element method (SEM) was introduced by Beskos in 1978, organized and seeded by Doyle (1997) in the 1990s. It allows calculating relatively complex structures with different boundary conditions and discontinuities using simple theories. It combines important characteristics of FEM, dynamic stiffness method (DSM), and spectral analysis (Lee 2009).

The dynamic stiffness matrices and shape functions used in SEM are exact within the scope of the underlying physical theory, and the method allows a reduced number of degrees of freedom. The matrices are depended on frequency, but using spectral analysis, the dynamic response can be easily composed by wave superposition. Harmonic, random, or damped transient excitations can be decomposed using the discrete Fourier transform (DFT). The discrete frequencies are used to calculate the spectral matrix and discrete responses. Then, the complete dynamic response is computed by the sum of frequency components (inverse DFT). As FEM, SEM uses the assembly of a global matrix using elementary matrices and spatial discretization. However, differently from FEM, only discontinuities and locations where loads are applied need to be meshed (Ahmida and Arruda 2001).

It can be a useful tool because it joins DSM with spectral analysis to produce accurate solutions in both frequency and time domains at a low computational cost due to a dramatic system

reduction, making it more adequate for uncertainty and optimization analyses. Furthermore, it can easily be associated with other numerical methods such as FEM and BEM. The limitations of SEM are general nonlinear analysis, or when exact wave solutions are not available, as it occurs in geometrically complex structures. When using SEM it is important to be careful with the time–frequency transformations to avoid leakage and aliasing errors (Lee 2009).

This chapter presents spectral element for straight and curved frame elements. First, the equations of motion are developed and physical and kinematic assumptions are presented. Then the spectrum relation is established and the spectral element is derived. Two simple examples – a two- and a three-dimensional frame structures, excited by a real earthquake signal – are analyzed. The purpose is to present the methodology and concepts of dynamic structural analysis via the spectral finite element method.

Classical Straight Frame

Equations of Motion

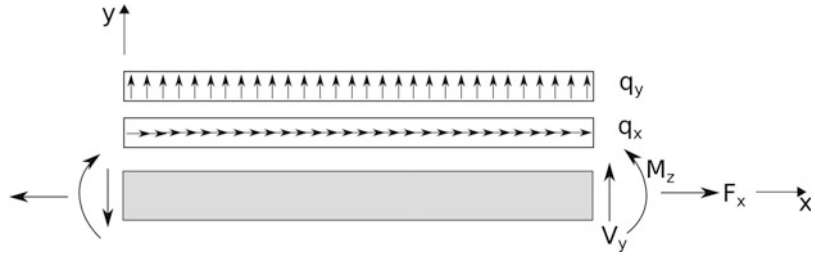
Consider a straight, long, and slender beam, as shown in Fig. 1. A Cartesian reference frame is placed such that the x -axis is along neutral axis of the beam. Thus, for a rectangular cross section, if h is the beam thickness, the position y will vary in the range $-h/2 \leq y \leq h/2$. Along the z -axis is the beam width.

Here, it is considered a plane frame element under flexural moments and axial forces. In straight beams, flexural and longitudinal behaviors are decoupled, and they may be modeled separately, the former via the Euler–Bernoulli beam theory and the latter via the elementary rod theory.

The Euler–Bernoulli theory assumes the vertical deflection is constant across any cross section and that cross sections remain plane after deformation, which means that the shear deformation is neglected. On the other hand, the classical rod theory considers only axial displacements and neglects lateral contractions due to Poisson's effect. Under these assumptions, the displacement field for a straight beam is given by:

Spectral Finite Element Approach for Structural Dynamics,

Fig. 1 Straight frame element model



$$\begin{aligned}\bar{u}(x, y) &= u(x) - y\theta_z(x) \\ \bar{v}(x, y) &= v(x)\end{aligned}$$

where $\bar{u}$ is the displacement in the x -direction (axial displacement), $\bar{v}$ is the displacement in the y -direction (vertical displacement), and θ_z is the rotation with respect to the z -axis.

Then, from differentiation of the displacement field, axial and transverse shear strains at a given cross section of the beam are expressed:

$$\begin{aligned}\epsilon_{xx} &= \frac{\partial \bar{u}}{\partial x} = \frac{\partial u}{\partial x} - y \frac{\partial \theta_z}{\partial x} \\ \epsilon_{xy} &= \frac{1}{2} \left(\frac{\partial \bar{u}}{\partial y} + \frac{\partial \bar{v}}{\partial x} \right) = \frac{1}{2} \left(-\theta_z + \frac{\partial v}{\partial x} \right).\end{aligned}$$

The Euler–Bernoulli beam theory neglects the shear deformation, although not the shear force. This implies that the transverse shear strain must vanish, which yields:

$$\epsilon_{xy} = 0 \rightarrow \theta_z = \frac{\partial v}{\partial x}.$$

The expression for the cross-section rotation can be used to write the axial strain, which is the only nonzero strain in the Euler–Bernoulli beam model, in terms of the vertical and the longitudinal displacements, as follows:

$$\epsilon_{xx} = \frac{\partial u}{\partial x} - y \frac{\partial^2 v}{\partial x^2}.$$

Moreover, neglecting all stresses except for the predominant axial stress, one can write:

$$\sigma_{xx} = E\epsilon_{xx} = E \frac{\partial u}{\partial x} - yE \frac{\partial^2 v}{\partial x^2}.$$

Once the stress and strains are determined, Newton's laws or variational formulations, such

as Hamilton's principle, may be used to determine the equations of motion for this classical beam model, which, in the absence of damping, is given by:

$$\begin{aligned}EA \frac{\partial^2 u}{\partial x^2} - \rho A \frac{\partial^2 v}{\partial t^2} &= q_x(x, t) \\ \frac{\partial^2}{\partial x^2} \left(EI \frac{\partial^2 v}{\partial x^2} \right) + \rho A \frac{\partial^2 v}{\partial t^2} &= q_y(x, t)\end{aligned}$$

where E is the Young's modulus, I is the area moment of inertia with respect to the z -axis, ρ is the material density, A is the cross-section area, and $q_x(x, t)$ and $q_y(x, t)$ are, respectively, the longitudinal and vertical distributed loads along the beam length. The boundary conditions associated to this beam problem are:

$$\begin{aligned}u &= 0 \quad \text{or} \quad F_x = EA \frac{\partial u}{\partial x} \\ v &= 0 \quad \text{or} \quad V_y = -\frac{\partial}{\partial x} \left(EI \frac{\partial^2 v}{\partial x^2} \right) \\ \theta_z &= 0 \quad \text{or} \quad M_z = EI \frac{\partial^2 v}{\partial x^2}\end{aligned}$$

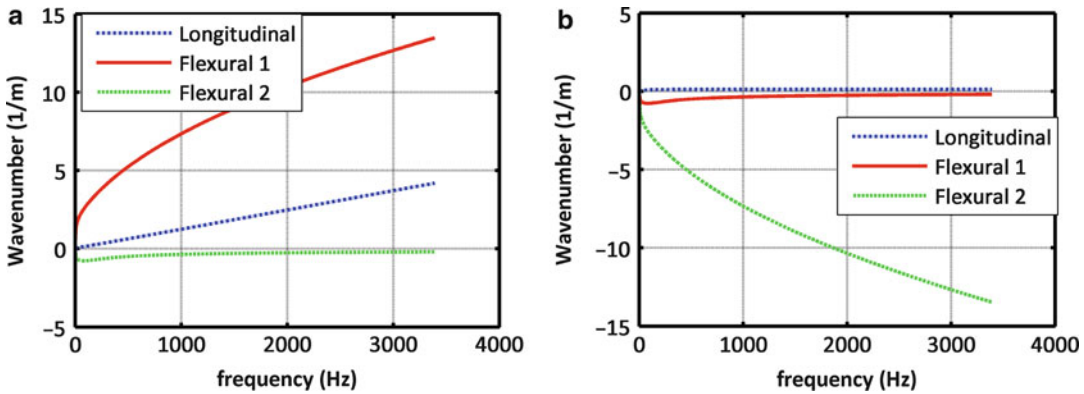
The term EI is usually called the *bending stiffness*.

If some internal viscoelastic damping η is present, the equations of motion for the straight frame are modified to:

$$\begin{aligned}EA \frac{\partial^2 u}{\partial x^2} - \eta A \frac{\partial u}{\partial t} - \rho A \frac{\partial^2 u}{\partial t^2} &= q_x(x, t) \\ \frac{\partial^2}{\partial x^2} \left(EI \frac{\partial^2 v}{\partial x^2} \right) + \eta A \frac{\partial v}{\partial t} + \rho A \frac{\partial^2 v}{\partial t^2} &= q_y(x, t).\end{aligned}$$

Spectral Analysis

Assuming a harmonic solution for the displacements of the beam (or applying the Fourier transform), i.e., by writing:



Spectral Finite Element Approach for Structural Dynamics, Fig. 2 Spectrum relations of a damped frame element: (a) real part of wavenumbers and (b) imaginary part of wavenumbers

$$u(x, t) = \sum_n \hat{u}_n(x, \omega_n) e^{i\omega_n t},$$

$$v(x, t) = \sum_n \hat{v}_n(x, \omega_n) e^{i\omega_n t}$$

in the homogenous case and considering constant properties along the beam, the equations of motion, in the frequency domain, are:

$$\frac{d^2 \hat{u}}{dx^2} + k_x^2 \hat{u} = 0, \quad k_x = \pm \sqrt{\frac{\omega^2 \rho A - i\omega \eta A}{EA}}$$

$$\frac{d^4 \hat{v}}{dx^4} - \beta^4 \hat{v} = 0, \quad \beta^2 = \pm \sqrt{\frac{\omega^2 \rho A - i\omega \eta A}{EI}}.$$

The second homogeneous equation, relative to the transverse displacement, can be split into a product of two terms which must vanish. Thus, it yields:

$$\frac{d^2 \hat{v}}{dx^2} + \beta^2 \hat{v} = 0$$

$$\frac{d^2 \hat{v}}{dx^2} - \beta^2 \hat{v} = 0.$$

Then, considering a second Fourier transformation, now in the spatial domain, $\hat{u}(x, \omega) = \sum_m \tilde{u}(k_m, \omega) e^{-ik_m x}$, $\hat{v}(x, \omega) = \sum_m \tilde{v}(k_m, \omega) e^{-ik_m x}$, gives:

$$-k_1^2 + k_x^2 = 0 \rightarrow k_1 = \pm k_x$$

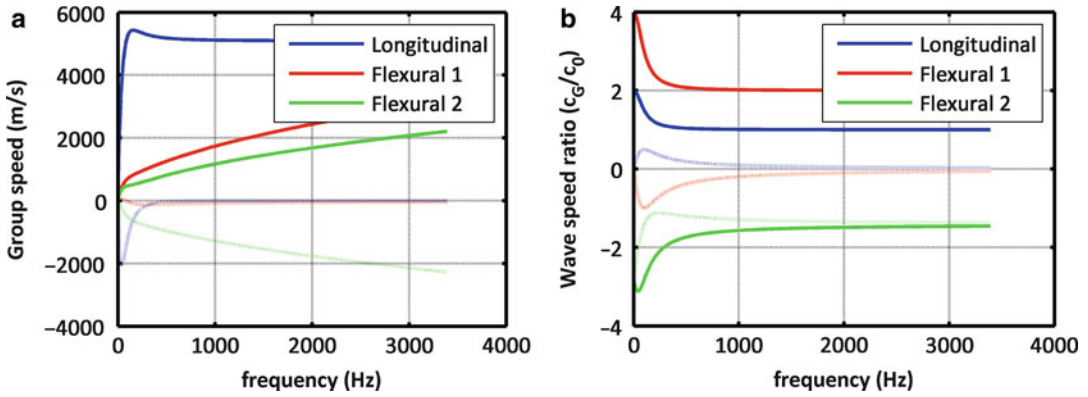
$$-k_2^2 + \beta^2 = 0 \rightarrow k_2 = \pm \beta$$

$$-k_3^2 - \beta^2 = 0 \rightarrow k_3 = \pm i\beta$$

where k_1, k_2, k_3 are the wavenumbers. They are function of the angular frequency ω , and this relationship constitutes the spectrum or dispersion relation, which is plotted in Fig. 2. Notice that when some damping is present, longitudinal and flexural modes are complex waves; thus, they exhibit a decaying behavior while propagating. In the undamped case, there are two real wavenumbers (the longitudinal and first flexural waves), which are propagating waves, and an imaginary wavenumber (the second flexural wave), which is a purely spatial decaying wave, the so-called evanescent wave. Phase and group speeds, c_0 and c_G , respectively, can be also derived (see Fig. 3), and they are expressed as:

$$c_0 = \frac{\omega}{k}, \quad c_G = \frac{d\omega}{dk}.$$

In the undamped case, unlike in the case of longitudinal motion, the phase and group speeds related to flexural waves depend on the cross-section properties, given by the area and the second moment of inertia. The group speed is twice the phase speed and both depend on frequency and, hence, are dispersive. It should be noted that, with the Euler–Bernoulli assumptions, the wave speeds tend to infinity as the frequency increases,



Spectral Finite Element Approach for Structural Dynamics, Fig. 3 Wave speeds of a frame element, real part in solid line, and imaginary part in dotted line: (a) group speeds and (b) speed ratios between group and phase speeds

which is not reasonable. This theory is, therefore, only suitable for low frequencies.

The displacement solution may be expressed in the spectral form as:

$$u(x, t) = \sum_n (A_1^+ e^{-ik_{xn}x} + A_1^- e^{ik_{xn}x}) e^{i\omega_n t}$$

$$v(x, t) = \sum_n (A_2^+ e^{-i\beta_n x} + A_3^+ e^{-\beta_n x} + A_2^- e^{i\beta_n x} + A_3^- e^{\beta_n x}) e^{i\omega_n t},$$

i.e., as a superposition of waves propagating in opposite directions.

It is worth recalling here that the cross-section rotation is related to the transverse displacement by the following expression:

$$\theta_z = \frac{\partial v}{\partial x}.$$

Thus, it may be also written as a superposition of waves, as follows:

$$\theta_z(x, t) = \sum_n (-iA_2^+ e^{-i\beta_n x} - A_3^+ e^{-\beta_n x} + iA_2^- e^{i\beta_n x} + A_3^- e^{\beta_n x}) \beta_n e^{i\omega_n t}.$$

From the classical theories for beams and rods, the internal loads related to the frame may be derived from the displacement field and are given by:

$$F_x = EA \frac{\partial u}{\partial x}$$

$$V_y = -\frac{\partial}{\partial x} \left(EI \frac{\partial^2 v}{\partial x^2} \right).$$

$$M_z = EI \frac{\partial^2 v}{\partial x^2}$$

Thus, they may be expressed in spectral form as:

$$F_x(x, t) = EA \sum_n k_{xn} (-iA_1^+ e^{-ik_{xn}x} + iA_1^- e^{ik_{xn}x}) e^{i\omega_n t}$$

$$V_y(x, t) = EI \sum_n \beta_n^3 (-iA_2^+ e^{-i\beta_n x} + A_3^+ e^{-\beta_n x} + iA_2^- e^{i\beta_n x} - A_3^- e^{\beta_n x}) e^{i\omega_n t}$$

$$M_z(x, t) = EI \sum_n \beta_n^2 (-A_2^+ e^{-i\beta_n x} + A_3^+ e^{-\beta_n x} - A_2^- e^{i\beta_n x} + A_3^- e^{\beta_n x}) e^{i\omega_n t}.$$

Displacements and loads evaluated at an arbitrary position along the frame may be expressed in matrix form, as follows:

$$\hat{\mathbf{q}} = \begin{bmatrix} \hat{u} \\ \hat{v} \\ \hat{\theta}_z \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 \\ 0 & -i\beta & -\beta & 0 & i\beta & \beta \end{bmatrix} \begin{bmatrix} \Lambda_x^+ & 0 \\ 0 & \Lambda_x^- \end{bmatrix} \begin{bmatrix} A_1^+ \\ A_2^+ \\ A_3^+ \\ B_1^- \\ B_2^- \\ B_3^- \end{bmatrix}$$

$$\hat{\mathbf{f}} = \begin{bmatrix} \hat{F}_x \\ \hat{V}_y \\ \hat{M}_z \end{bmatrix} = E \begin{bmatrix} -iAk_x & 0 & 0 & -iAk_x & 0 & 0 \\ 0 & -iI\beta^3 & I\beta^3 & 0 & iI\beta^3 & -I\beta^3 \\ 0 & -I\beta^2 & I\beta^2 & 0 & -I\beta^2 & I\beta^2 \end{bmatrix} \begin{bmatrix} \Lambda_x^+ & 0 \\ 0 & \Lambda_x^- \end{bmatrix} \begin{bmatrix} A_1^+ \\ A_2^+ \\ A_3^+ \\ B_1^- \\ B_2^- \\ B_3^- \end{bmatrix}$$

where

$$\Lambda_x^+ = \begin{bmatrix} e^{-ik_x x} & 0 & 0 \\ 0 & e^{-i\beta x} & 0 \\ 0 & 0 & e^{-\beta x} \end{bmatrix},$$

$$\Lambda_x^- = \begin{bmatrix} e^{ik_x(x-L)} & 0 & 0 \\ 0 & e^{i\beta(x-L)} & 0 \\ 0 & 0 & e^{\beta(x-L)} \end{bmatrix}.$$

Here, the reference for wave amplitudes related to negative-going waves, i.e., those propagating to the left or in the negative sense of the x -axis, is placed at the end of an element of frame ($x = L$). Thus, instead of A_1^-, A_2^-, A_3^- , one has B_1^-, B_2^-, B_3^- . This is done for the purpose of numerical conditioning in the case where there are highly evanescent waves.

From Mace et al. (2005) and Mead (1973), the classification of waves into positive- or negative-

going waves consists in analyzing the sense of decreasing wave amplitudes or, if the amplitude remains constant, the sense of the time-averaged power transmission. The latter can be evaluated by regarding the sign of $\Re\{i\omega \mathbf{f}_L^T \mathbf{q}_L\}$: if it is negative, the corresponding wave is a positive-going wave; otherwise, it is a negative-going wave. Alternatively, the sign of the group speed may be used, which may be shown to follow the time-averaged power transmission sign.

Spectral Element for a Classical Frame

Consider an element of classical frame (Euler–Bernoulli beam combined with an elementary rod) with length L . The displacements at both ends of the spectral element can be expressed as a function of the wave amplitudes as follows:

$$\begin{bmatrix} \hat{u}(0) \\ \hat{v}(0) \\ \hat{\theta}_z(0) \\ \hat{u}(L) \\ \hat{v}(L) \\ \hat{\theta}_z(L) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & -e^{-ik_x L} & 0 & 0 \\ 0 & 1 & 1 & 0 & e^{-i\beta L} & e^{-\beta L} \\ 0 & -i\beta & -\beta & 0 & i\beta e^{-i\beta L} & \beta e^{-\beta L} \\ e^{-ik_x L} & 0 & 0 & -1 & 0 & 0 \\ 0 & e^{-i\beta L} & e^{-\beta L} & 0 & 1 & 1 \\ 0 & -i\beta e^{-i\beta L} & -\beta e^{-\beta L} & 0 & i\beta & \beta \end{bmatrix} \begin{bmatrix} A_1^+ \\ A_2^+ \\ A_3^+ \\ B_1^- \\ B_2^- \\ B_3^- \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mathbf{q}}_0 \\ \hat{\mathbf{q}}_L \end{bmatrix} = \begin{bmatrix} \Phi_q^+ & \Phi_q^- \Lambda_L^+ \\ \Phi_q^+ \Lambda_L^+ & \Phi_q^- \end{bmatrix} \begin{bmatrix} \mathbf{A}^+ \\ \mathbf{B}^- \end{bmatrix}.$$

Analogously, the external loads applied to the frame may be written as:

$$\begin{bmatrix} -\hat{F}_x(0) \\ -\hat{V}_y(0) \\ -\hat{M}_z(0) \\ \hat{F}_x(L) \\ \hat{V}_y(L) \\ \hat{M}_z(L) \end{bmatrix} = E \begin{bmatrix} iAk_x & 0 & 0 & iAk_x e^{-ik_x L} & 0 & 0 \\ 0 & iI\beta^3 & -I\beta^3 & 0 & -iI\beta^3 e^{-i\beta L} & I\beta^3 e^{-\beta L} \\ 0 & I\beta^2 & -I\beta^2 & 0 & I\beta^2 e^{-i\beta L} & -I\beta^2 e^{-\beta L} \\ -iAk_x e^{-ik_x L} & 0 & 0 & -iAk_x & 0 & 0 \\ 0 & -iI\beta^3 e^{-i\beta L} & I\beta^3 e^{-\beta L} & 0 & iI\beta^3 & -I\beta^3 \\ 0 & -I\beta^2 e^{-i\beta L} & I\beta^2 e^{-\beta L} & 0 & -I\beta^2 & I\beta^2 \end{bmatrix} \begin{bmatrix} A_1^+ \\ A_2^+ \\ A_3^+ \\ B_1^- \\ B_2^- \\ B_3^- \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mathbf{F}}_0 \\ \hat{\mathbf{F}}_L \end{bmatrix} = \begin{bmatrix} -\Phi_f^+ & -\Phi_f^- \Lambda_L^+ \\ \Phi_f^+ \Lambda_L^+ & \Phi_f^- \end{bmatrix} \begin{bmatrix} \mathbf{A}^+ \\ \mathbf{B}^- \end{bmatrix}.$$

With the general method of computing $\mathbf{A}^+$ and $\mathbf{B}^-$ from the displacement equation and replacing it in the expression for external loads, one can write:

$$\mathbf{K}_\infty = \frac{E(1+i)}{2} \begin{bmatrix} iAk_x & 0 & 0 \\ 0 & 2iI\beta^3 & (1+i)I\beta^2 \\ 0 & (1+i)I\beta^2 & 2I\beta \end{bmatrix}.$$

$$\begin{bmatrix} \hat{\mathbf{F}}_0 \\ \hat{\mathbf{F}}_L \end{bmatrix} = \mathbf{K}(\omega) \begin{bmatrix} \hat{\mathbf{q}}_0 \\ \hat{\mathbf{q}}_L \end{bmatrix},$$

$$\mathbf{K} = \begin{bmatrix} -\Phi_f^+ & -\Phi_f^- \Lambda_L^+ \\ \Phi_f^+ \Lambda_L^+ & \Phi_f^- \end{bmatrix} \begin{bmatrix} \Phi_q^+ & \Phi_q^- \Lambda_L^+ \\ \Phi_q^+ \Lambda_L^+ & \Phi_q^- \end{bmatrix}^{-1}.$$

Throw-Off Spectral Element for a Straight Frame
A semi-infinite element of beam can be easily modeled via SEM by assuming the wave amplitudes related to waves propagating to the left (negative-going waves) are zero, i.e., considering there is no reflection. In this case, the displacement and force vectors at the left side of the element are given by:

$$\begin{aligned} \hat{\mathbf{q}}_0 &= \Phi_q^+ \mathbf{A}^+ \\ \hat{\mathbf{F}}_0 &= -\Phi_f^+ \mathbf{A}^+. \end{aligned}$$

Thus, the spectral matrix for the throw-off element of frame is:

$$\hat{\mathbf{F}}_0 = \mathbf{K}_\infty(\omega) \hat{\mathbf{q}}_0, \quad \mathbf{K}_\infty = -\Phi_f^+ (\Phi_q^+)^{-1}$$

Curved Beam

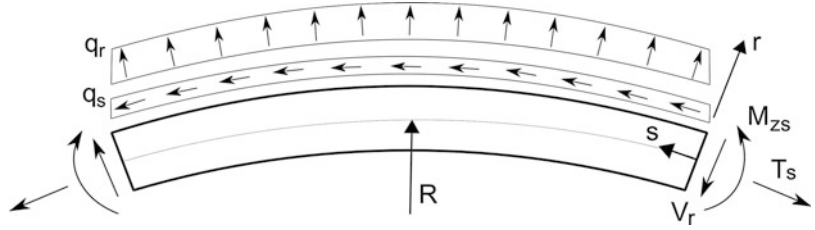
Equations of Motion

Consider a slender in-plane curved beam as shown in Fig. 4. The circumferential reference frame is placed such that the s -axis is along the neutral axis of the beam, here supposed to be equidistant of the upper and lower surfaces of the beam. The beam thickness h varies in the range $-h/2 \leq r \leq +h/2$ along the radial r -axis and its width lies along the z -axis. Based on the Euler–Bernoulli beam theory, which neglects shear deformation, the kinematics of a curved beam can be written as:

$$\begin{aligned} \bar{u}_s(r, s) &= u_s(s) + r \left(\frac{\partial u_s(s)}{\partial s} - \frac{v_r(s)}{R} \right) \\ \bar{v}_r(r, s) &= v_r(s) \end{aligned}$$

where $\bar{u}_s$ is the displacement in the s -direction (tangential displacement), $\bar{v}_r$ is the displacement in the y -direction (radial displacement), $\bar{\psi}$ is the rotation with respect to the z -axis, and R is the radius of the centerline.

**Spectral Finite Element
Approach for Structural
Dynamics, Fig. 4** Curved
beam element model



Then, from differentiation of the displacement field, considering a significant strain in s -direction, a strain at a given cross section of the curved beam is expressed:

$$\varepsilon_{ss} = \frac{\partial u_s}{\partial s} - \frac{v_r}{R} + r \left(\frac{\partial^2 v_r}{\partial s^2} + \frac{1}{R} \frac{\partial u_s}{\partial s} \right).$$

Moreover, neglecting all stresses except for the predominant circumferential stress, one can write:

$$\sigma_{ss} = E\varepsilon_{ss} = E \left[\frac{\partial u_s}{\partial s} - \frac{v_r}{R} \right] + rE \left(\frac{\partial^2 v_r}{\partial s^2} + \frac{1}{R} \frac{\partial u_s}{\partial s} \right).$$

Then, as a result of the application of Newton's laws or variational formulations, the equations of motion for the undamped curved beam are derived:

$$\begin{aligned} & \frac{EI}{R} \left(\frac{1}{R} \frac{\partial^2 u_s}{\partial s^2} - \frac{\partial^3 v_r}{\partial s^3} \right) + \frac{EA}{R} \left(\frac{\partial v_r}{\partial s} + R \frac{\partial^2 u_s}{\partial s^2} \right) \\ & - \rho A \frac{\partial^2 u_s}{\partial t^2} = q_s(s, t) \\ & \frac{EI}{R} \left(\frac{\partial^3 u_s}{\partial s^3} - R \frac{\partial^4 v_r}{\partial s^4} \right) - \frac{EA}{R} \left(\frac{v_r}{R} + \frac{\partial u_s}{\partial s} \right) \\ & + \rho A \frac{\partial^2 v_r}{\partial t^2} = q_r(s, t) \end{aligned}$$

where $q_s(s, t)$ and $q_r(s, t)$ are, respectively, tangential and radial distributed loads along the curved beam length. The boundary conditions associated to this problem are:

$$\begin{aligned} u_s = 0 & \quad \text{or} \quad T_s = EA \left(\frac{\partial u_s}{\partial s} - \frac{v_r}{R} \right) \\ v_r = 0 & \quad \text{or} \quad V_r = -EI \left(\frac{1}{R} \frac{\partial^2 u_s}{\partial s^2} + \frac{\partial^3 v_r}{\partial s^3} \right) \\ \psi = 0 & \quad \text{or} \quad M_{zs} = EI \left(\frac{1}{R} \frac{\partial u_s}{\partial s} + \frac{\partial^2 v_r}{\partial s^2} \right). \end{aligned}$$

If some internal viscoelastic damping η is present, the equations of motion are modified to:

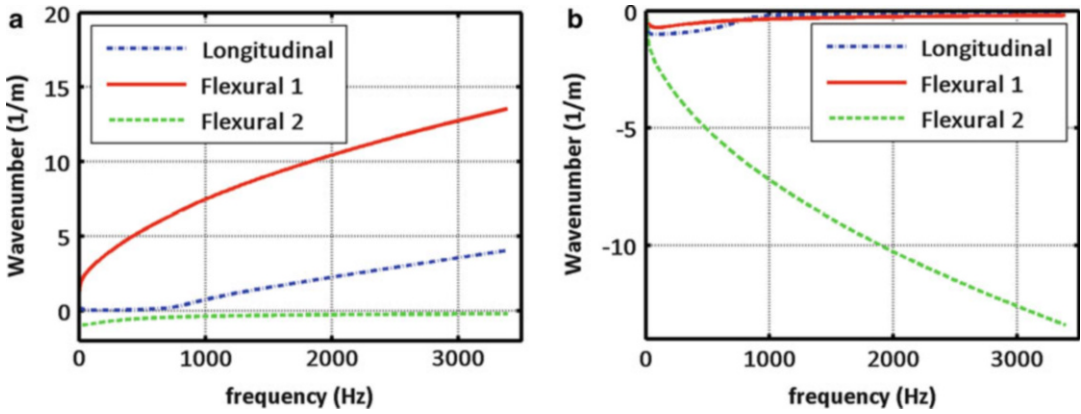
$$\begin{aligned} & \frac{EI}{R} \left(\frac{1}{R} \frac{\partial^2 u_s}{\partial s^2} + \frac{\partial^3 v_r}{\partial s^3} \right) + \frac{EA}{R} \left(-\frac{\partial v_r}{\partial s} + R \frac{\partial^2 u_s}{\partial s^2} \right) \\ & - \eta A \frac{\partial u_s}{\partial t} - \rho A \frac{\partial^2 u_s}{\partial t^2} = q_s(s, t) \\ & \frac{EI}{R} \left(\frac{\partial^3 u_s}{\partial s^3} + R \frac{\partial^4 v_r}{\partial s^4} \right) + \frac{EA}{R} \left(\frac{v_r}{R} - \frac{\partial u_s}{\partial s} \right) \\ & + \eta A \frac{\partial v_r}{\partial t} + \rho A \frac{\partial^2 v_r}{\partial t^2} = q_r(s, t). \end{aligned}$$

Spectral Analysis

As before, in the straight beam case, two Fourier transforms can be successively applied to the displacement solution of a curved beam, which allows writing:

$$\begin{aligned} u_s(s, t) &= \sum_n \hat{u}_{sn}(s, \omega_n) e^{i\omega_n t} = \sum_n \sum_m \tilde{u}_s(k_m, \omega) e^{-ik_m s} e^{i\omega_n t} \\ v_r(s, t) &= \sum_n \hat{v}_{rn}(s, \omega_n) e^{i\omega_n t} = \sum_n \sum_m \tilde{v}_r(k_m, \omega) e^{-ik_m s} e^{i\omega_n t}. \end{aligned}$$

In the homogenous case and considering constant properties along the beam, the displacement solution is substituted in the equations of motion, which yields, in the frequency domain,



Spectral Finite Element Approach for Structural Dynamics, Fig. 5 Spectrum relations of a damped curved beam element: (a) real part of wavenumbers and (b) imaginary part of wavenumbers

$$\begin{bmatrix} -\frac{EI k^2}{R^2} - E A k^2 + i \eta A \omega + \omega^2 \rho A & \frac{i E I k^3}{R} + \frac{i E A k}{R} \\ \frac{i E I k^3}{R} + \frac{i E A k}{R} & E I k^4 + \frac{E A}{R^2} - i \eta A \omega - \omega^2 \rho A \end{bmatrix} \begin{bmatrix} \tilde{u}_s \\ \tilde{v}_r \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

A nontrivial solution for this system of equations is obtained by setting the determinant to vanish, which results in a sixth-order polynomial equation in k , the characteristic equation, or spectrum relation:

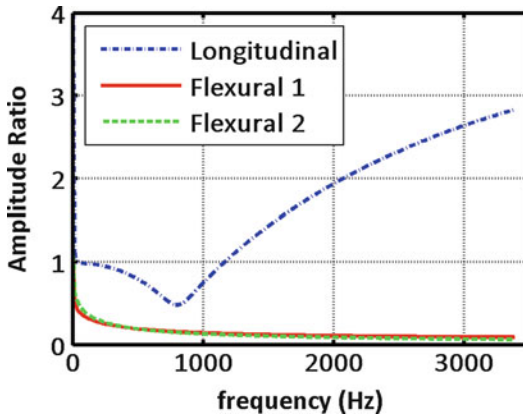
$$k^6 - \left[k_0^2 + \frac{2}{R^2} \right] k^4 - \left[\beta_0^4 + k_0^2 \frac{1}{R^2} - \frac{1}{R^4} \right] k^2 + \left[k_0^2 - \frac{1}{R^2} \right] \beta_0^4 = 0$$

where $\beta_0^4 = \frac{\rho A \omega^2 + i \eta A \omega}{EI}$ and $k_0^2 = \frac{\rho A \omega^2 + i \eta A \omega}{EA}$.

This equation provides three pairs of wavenumbers $\pm k_{s1}$, $\pm k_{s2}$, $\pm k_{s3}$. One pair is linked to extensional waves and the other two pairs to flexural waves (Fig. 5).

Then, the displacement solution may be expressed in the spectral form as:

$$\begin{aligned} u_s(s, t) &= \sum_n \left(\bar{A}_{s1}^+ e^{-ik_{s1}n s} + \bar{A}_{s2}^+ e^{-ik_{s2}n s} + \bar{A}_{s3}^+ e^{-ik_{s3}n s} + \bar{A}_{s1}^- e^{ik_{s1}n s} + \bar{A}_{s2}^- e^{ik_{s2}n s} + \bar{A}_{s3}^- e^{ik_{s3}n s} \right) e^{i\omega_n t} \\ v_r(s, t) &= \sum_n \left(A_{s1}^+ e^{-ik_{s1}n s} + A_{s2}^+ e^{-ik_{s2}n s} + A_{s3}^+ e^{-ik_{s3}n s} + A_{s1}^- e^{ik_{s1}n s} + A_{s2}^- e^{ik_{s2}n s} + A_{s3}^- e^{ik_{s3}n s} \right) e^{i\omega_n t}, \\ \psi(s, t) &= \frac{\partial v_r(s, t)}{\partial s} = \sum_n \left(-ik_{s1}n A_{s1}^+ e^{-ik_{s1}n s} - ik_{s2}n A_{s2}^+ e^{-ik_{s2}n s} - ik_{s3}n A_{s3}^+ e^{-ik_{s3}n s} + ik_{s1}n A_{s1}^- e^{ik_{s1}n s} + ik_{s2}n A_{s2}^- e^{ik_{s2}n s} + ik_{s3}n A_{s3}^- e^{ik_{s3}n s} \right) e^{i\omega_n t}, \end{aligned}$$



Spectral Finite Element Approach for Structural Dynamics, Fig. 6 Amplitude ratios for a damped curved beam element

i.e., as a superposition of waves propagating in opposite directions. From the homogeneous system of equations for $\tilde{u}_s$, $\tilde{v}_r$, amplitude ratios for each wave mode are obtained (Fig. 6):

$$\alpha_{sm} = \frac{\bar{A}_{sm}}{A_{sm}} = -\frac{EI k_{smn}^4 + EAR^{-2} - i\eta A\omega - \rho A\omega^2}{i(EIk_{smn}^3 + EAk_{smn})/R}.$$

Expressions for the internal loads in curved beams have been derived in terms of the displacement field and are rewritten below:

$$\begin{aligned} T_s &= EA \left(\frac{\partial u_s}{\partial s} - \frac{v_r}{R} \right) \\ V_r &= -EI \left(\frac{1}{R} \frac{\partial^2 u_s}{\partial s^2} + \frac{\partial^3 v_r}{\partial s^3} \right) \\ M_{zs} &= EI \left(\frac{1}{R} \frac{\partial u_s}{\partial s} + \frac{\partial^2 v_r}{\partial s^2} \right). \end{aligned}$$

Thus, substituting the spectral solutions for the displacement field, it yields:

$$\begin{aligned} T_s(s, t) &= EA \sum_n \left(-ik_{s1n} \alpha_{s1} A_{s1}^+ e^{-ik_{s1n}s} - ik_{s2n} \alpha_{s2} A_{s2}^+ e^{-ik_{s2n}s} - ik_{s3n} \alpha_{s3} A_{s3}^+ e^{-ik_{s3n}s} \right) e^{i\omega_n t} \\ &\quad + ik_{s1n} \alpha_{s1} A_{s1}^- e^{ik_{s1n}s} + ik_{s2n} \alpha_{s2} A_{s2}^- e^{ik_{s2n}s} + ik_{s3n} \alpha_{s3} A_{s3}^- e^{ik_{s3n}s} \\ &\quad - \frac{EA}{R} \sum_n \left(\frac{A_{s1}^+ e^{-ik_{s1n}s} + A_{s2}^+ e^{-ik_{s2n}s} + A_{s3}^+ e^{-ik_{s3n}s}}{A_{s1}^- e^{ik_{s1n}s} + A_{s2}^- e^{ik_{s2n}s} + A_{s3}^- e^{ik_{s3n}s}} \right) e^{i\omega_n t} \\ V_r(s, t) &= -\frac{EI}{R} \sum_n \left(k_{s1n}^2 \alpha_{s1} A_{s1}^+ e^{-ik_{s1n}s} + k_{s2n}^2 \alpha_{s2} A_{s2}^+ e^{-ik_{s2n}s} + k_{s3n}^2 \alpha_{s3} A_{s3}^+ e^{-ik_{s3n}s} \right) e^{i\omega_n t} \\ &\quad - k_{s1n}^2 \alpha_{s1} A_{s1}^- e^{ik_{s1n}s} - k_{s2n}^2 \alpha_{s2} A_{s2}^- e^{ik_{s2n}s} - k_{s3n}^2 \alpha_{s3} A_{s3}^- e^{ik_{s3n}s} \\ &\quad - EI \sum_n \left(\frac{ik_{s1n}^3 A_{s1}^+ e^{-ik_{s1n}s} + ik_{s2n}^3 A_{s2}^+ e^{-ik_{s2n}s} + ik_{s3n}^3 A_{s3}^+ e^{-ik_{s3n}s}}{-ik_{s1n}^3 A_{s1}^- e^{ik_{s1n}s} - ik_{s2n}^3 A_{s2}^- e^{ik_{s2n}s} - ik_{s3n}^3 A_{s3}^- e^{ik_{s3n}s}} \right) e^{i\omega_n t} \\ M_{zs}(s, t) &= \frac{EI}{R} \sum_n \left(-ik_{s1n} \alpha_{s1} A_{s1}^+ e^{-ik_{s1n}s} - ik_{s2n} \alpha_{s2} A_{s2}^+ e^{-ik_{s2n}s} - ik_{s3n} \alpha_{s3} A_{s3}^+ e^{-ik_{s3n}s} \right) e^{i\omega_n t} \\ &\quad + ik_{s1n} \alpha_{s1} A_{s1}^- e^{ik_{s1n}s} + ik_{s2n} \alpha_{s2} A_{s2}^- e^{ik_{s2n}s} + ik_{s3n} \alpha_{s3} A_{s3}^- e^{ik_{s3n}s} \\ &\quad + EI \sum_n \left(\frac{k_{s1n}^2 A_{s1}^+ e^{-ik_{s1n}s} + k_{s2n}^2 A_{s2}^+ e^{-ik_{s2n}s} + k_{s3n}^2 A_{s3}^+ e^{-ik_{s3n}s}}{-k_{s1n}^2 A_{s1}^- e^{ik_{s1n}s} - k_{s2n}^2 A_{s2}^- e^{ik_{s2n}s} - k_{s3n}^2 A_{s3}^- e^{ik_{s3n}s}} \right) e^{i\omega_n t} \end{aligned}$$

Displacements and loads evaluated at an arbitrary position along the curved beam may be expressed in matrix form, as follows:

$$\hat{\mathbf{q}}_{CB} = \begin{bmatrix} \hat{u}_s \\ \hat{v}_r \\ \hat{\psi} \end{bmatrix} = \begin{bmatrix} \alpha_{s1} & \alpha_{s2} & \alpha_{s3} & \alpha_{s1} & \alpha_{s2} & \alpha_{s3} \\ 1 & 1 & 1 & 1 & 1 & 1 \\ -ik_{s1} & -ik_{s2} & -ik_{s3} & ik_{s1} & ik_{s2} & ik_{s3} \end{bmatrix} \begin{bmatrix} \Lambda_s^+ & 0 \\ 0 & \Lambda_s^- \end{bmatrix} \begin{bmatrix} A_{s1}^+ \\ A_{s2}^+ \\ A_{s3}^+ \\ B_{s1}^- \\ B_{s2}^- \\ B_{s3}^- \end{bmatrix}$$

$$\hat{\mathbf{f}}_{CB} = \begin{bmatrix} \hat{T}_s \\ \hat{V}_r \\ \hat{M}_{zs} \end{bmatrix} = \begin{pmatrix} C_1 \begin{bmatrix} -ik_{s1} & -ik_{s2} & -ik_{s3} & ik_{s1} & ik_{s2} & ik_{s3} \\ k_{s1}^2 & k_{s2}^2 & k_{s3}^2 & -k_{s1}^2 & -k_{s2}^2 & -k_{s3}^2 \\ -ik_{s1} & -ik_{s2} & -ik_{s3} & ik_{s1} & ik_{s2} & ik_{s3} \end{bmatrix} \Gamma_s \\ + C_2 \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ ik_{s1}^3 & ik_{s2}^3 & ik_{s3}^3 & -ik_{s1}^3 & -ik_{s2}^3 & -ik_{s3}^3 \\ k_{s1}^2 & k_{s2}^2 & k_{s3}^2 & -k_{s1}^2 & -k_{s2}^2 & -k_{s3}^2 \end{bmatrix} \end{pmatrix} \begin{bmatrix} \Lambda_s^+ & 0 \\ 0 & \Lambda_s^- \end{bmatrix} \begin{bmatrix} A_{s1}^+ \\ A_{s2}^+ \\ A_{s3}^+ \\ B_{s1}^- \\ B_{s2}^- \\ B_{s3}^- \end{bmatrix}$$

where

$$\Lambda_s^+ = \begin{bmatrix} e^{-ik_{s1}s} & 0 & 0 \\ 0 & e^{-ik_{s2}s} & 0 \\ 0 & 0 & e^{-ik_{s3}s} \end{bmatrix}, \quad \Lambda_s^- = \begin{bmatrix} e^{ik_{s1}(s-s_0)} & 0 & 0 \\ 0 & e^{ik_{s2}(s-s_0)} & 0 \\ 0 & 0 & e^{ik_{s3}(s-s_0)} \end{bmatrix},$$

$$\Gamma_s = \text{diag}[\alpha_{s1} \quad \alpha_{s2} \quad \alpha_{s3} \quad \alpha_{s1} \quad \alpha_{s2} \quad \alpha_{s3}],$$

$$C_1 = \begin{bmatrix} EA & 0 & 0 \\ 0 & -EI/R & 0 \\ 0 & 0 & EI/R \end{bmatrix}, \quad C_2 = \begin{bmatrix} -EA/R & 0 & 0 \\ 0 & -EI & 0 \\ 0 & 0 & EI \end{bmatrix}.$$

It should be noticed that the reference for the wave amplitudes related to negative-going waves is placed at the end of the beam element, i.e., at $s = s_0$.

Spectral Element

Consider an element of curved beam with circumferential length s_0 . The displacements at both ends of the spectral element can be expressed as a function of the wave amplitudes as follows:

$$\begin{bmatrix} \hat{u}_s(0) \\ \hat{v}_r(0) \\ \hat{\psi}(0) \\ \hat{u}_s(s_0) \\ \hat{v}_r(s_0) \\ \hat{\psi}(s_0) \end{bmatrix} = \begin{bmatrix} \alpha_{s1} & \alpha_{s2} & \alpha_{s3} & \alpha_{s1}e^{-ik_{s1}s_0} & \alpha_{s2}e^{-ik_{s2}s_0} & \alpha_{s3}e^{-ik_{s3}s_0} \\ 1 & 1 & 1 & e^{-ik_{s1}s_0} & e^{-ik_{s2}s_0} & e^{-ik_{s3}s_0} \\ -ik_{s1} & -ik_{s2} & -ik_{s3} & ik_{s1}e^{-ik_{s1}s_0} & ik_{s2}e^{-ik_{s2}s_0} & ik_{s3}e^{-ik_{s3}s_0} \\ \alpha_{s1}e^{-ik_{s1}s_0} & \alpha_{s2}e^{-ik_{s2}s_0} & \alpha_{s3}e^{-ik_{s3}s_0} & \alpha_{s1} & \alpha_{s2} & \alpha_{s3} \\ e^{-ik_{s1}s_0} & e^{-ik_{s2}s_0} & e^{-ik_{s3}s_0} & 1 & 1 & 1 \\ -ik_{s1}e^{-ik_{s1}s_0} & -ik_{s2}e^{-ik_{s2}s_0} & -ik_{s3}e^{-ik_{s3}s_0} & ik_{s1} & ik_{s2} & ik_{s3} \end{bmatrix} \begin{bmatrix} A_{s1}^+ \\ A_{s2}^+ \\ A_{s3}^+ \\ B_{s1}^- \\ B_{s2}^- \\ B_{s3}^- \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mathbf{q}}_0 \\ \hat{\mathbf{q}}_{s_0} \end{bmatrix} = \begin{bmatrix} \Phi_{qs}^+ & \Phi_{qs}^- \Lambda_{s_0}^+ \\ \Phi_{qs}^+ \Lambda_{s_0}^+ & \Phi_{qs}^- \end{bmatrix} \begin{bmatrix} \mathbf{A}_s^+ \\ \mathbf{B}_s^- \end{bmatrix}.$$

Analogously, the external loads applied may be written as:

$$\begin{bmatrix} -\hat{T}_s(0) \\ -\hat{V}_r(0) \\ \hat{M}_{zs}(0) \\ \hat{T}_s(s_0) \\ \hat{V}_r(s_0) \\ -\hat{M}_{zs}(s_0) \end{bmatrix} = \begin{bmatrix} C_1 \\ +C_2 \end{bmatrix} \begin{bmatrix} ik_{s1} & ik_{s2} & ik_{s3} & -ik_{s1}e^{-ik_{s1}s_0} & -ik_{s2}e^{-ik_{s2}s_0} & -ik_{s3}e^{-ik_{s3}s_0} \\ -k_{s1}^2 & -k_{s2}^2 & -k_{s3}^2 & k_{s1}^2e^{-ik_{s1}s_0} & k_{s2}^2e^{-ik_{s2}s_0} & k_{s3}^2e^{-ik_{s3}s_0} \\ -ik_{s1} & -ik_{s2} & -ik_{s3} & ik_{s1}e^{-ik_{s1}s_0} & ik_{s2}e^{-ik_{s2}s_0} & ik_{s3}e^{-ik_{s3}s_0} \\ -ik_{s1}e^{-ik_{s1}s_0} & -ik_{s2}e^{-ik_{s2}s_0} & -ik_{s3}e^{-ik_{s3}s_0} & ik_{s1} & ik_{s2} & ik_{s3} \\ k_{s1}^2e^{-ik_{s1}s_0} & k_{s2}^2e^{-ik_{s2}s_0} & k_{s3}^2e^{-ik_{s3}s_0} & -k_{s1}^2 & -k_{s2}^2 & -k_{s3}^2 \\ ik_{s1}e^{-ik_{s1}s_0} & ik_{s2}e^{-ik_{s2}s_0} & ik_{s3}e^{-ik_{s3}s_0} & -ik_{s1} & -ik_{s2} & -ik_{s3} \end{bmatrix} \Gamma_s \begin{bmatrix} A_{s1}^+ \\ A_{s2}^+ \\ A_{s3}^+ \\ B_{s1}^- \\ B_{s2}^- \\ B_{s3}^- \end{bmatrix} + \begin{bmatrix} -1 & -1 & -1 & -e^{-ik_{s1}s_0} & -e^{-ik_{s2}s_0} & -e^{-ik_{s3}s_0} \\ -ik_{s1}^3 & -ik_{s2}^3 & -ik_{s3}^3 & ik_{s1}^3e^{-ik_{s1}s_0} & ik_{s2}^3e^{-ik_{s2}s_0} & ik_{s3}^3e^{-ik_{s3}s_0} \\ k_{s1}^2 & k_{s2}^2 & k_{s3}^2 & -k_{s1}^2e^{-ik_{s1}s_0} & -k_{s2}^2e^{-ik_{s2}s_0} & -k_{s3}^2e^{-ik_{s3}s_0} \\ e^{-ik_{s1}s_0} & e^{-ik_{s2}s_0} & e^{-ik_{s3}s_0} & 1 & 1 & 1 \\ ik_{s1}^3e^{-ik_{s1}s_0} & ik_{s2}^3e^{-ik_{s2}s_0} & ik_{s3}^3e^{-ik_{s3}s_0} & -ik_{s1}^3 & -ik_{s2}^3 & -ik_{s3}^3 \\ -k_{s1}^2e^{-ik_{s1}s_0} & -k_{s2}^2e^{-ik_{s2}s_0} & -k_{s3}^2e^{-ik_{s3}s_0} & k_{s1}^2 & k_{s2}^2 & k_{s3}^2 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{F}}_0 \\ \hat{\mathbf{F}}_{s_0} \end{bmatrix} = \begin{bmatrix} -\Phi_{fs}^+ & -\Phi_{fs}^- \Lambda_{s_0}^+ \\ \Phi_{fs}^+ \Lambda_{s_0}^+ & \Phi_{fs}^- \end{bmatrix} \begin{bmatrix} \mathbf{A}_s^+ \\ \mathbf{B}_s^- \end{bmatrix}.$$

With the general method for computing $\mathbf{A}_s^+$ and $\mathbf{B}_s^-$ from the displacement equation and replacing it in the expression for external loads, one can write:

$$\begin{bmatrix} \hat{\mathbf{F}}_0 \\ \hat{\mathbf{F}}_{s_0} \end{bmatrix} = \mathbf{K}_{CB}(\omega) \begin{bmatrix} \hat{\mathbf{q}}_0 \\ \hat{\mathbf{q}}_{s_0} \end{bmatrix},$$

$$\mathbf{K}_{CB}(\omega) = \begin{bmatrix} -\Phi_{fs}^+ & -\Phi_{fs}^- \Lambda_{s_0}^+ \\ \Phi_{fs}^+ \Lambda_{s_0}^+ & \Phi_{fs}^- \end{bmatrix} \begin{bmatrix} \Phi_{qs}^+ & \Phi_{qs}^- \Lambda_{s_0}^+ \\ \Phi_{qs}^+ \Lambda_{s_0}^+ & \Phi_{qs}^- \end{bmatrix}^{-1}.$$

Rod Under Torsion (Shaft)

Consider a rod element subjected to a distributed torsion moment (torque) m_x as shown in Fig. 7. Due to the applied load, at a position x , the cross section rotates about the x -axis by an angle $\theta_x(x)$. Here, it is assumed that the cross section remains plane while twisting. For wavelengths which are greater than ten times the cross-section dimensions, the Saint-Venant theory is also valid (Petyt 2010). In this case, the kinematics of a rod under torsion may be approximated as:

$$\begin{aligned} u(x, y, z) &= \psi(y, z) \frac{\partial \theta_x(x)}{\partial x} \\ v(x, y, z) &= -z \theta_x(x) \\ w(x, y, z) &= y \theta_x(x) \end{aligned}$$

where $\psi(y, z)$ is a function which represents the warping of the cross section. In this chapter, only rods of circular cross section are considered. In this case, there is no warping ($\psi = 0$) and the nonzero strain components are:

$$\begin{aligned} \epsilon_{xy} &= \frac{1}{2} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \rightarrow \epsilon_{xy} = -\frac{z}{2} \frac{\partial \theta_x}{\partial x}, \quad \gamma_{xy} = -z \frac{\partial \theta_x}{\partial x} \\ \epsilon_{xz} &= \frac{1}{2} \left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial y} \right) \rightarrow \epsilon_{xz} = \frac{y}{2} \frac{\partial \theta_x}{\partial x}, \quad \gamma_{xz} = y \frac{\partial \theta_x}{\partial x}. \end{aligned}$$

The associated stresses are given by:

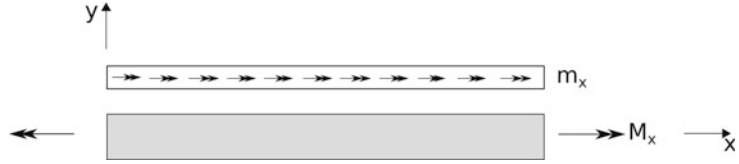
$$\begin{aligned} \sigma_{xy} &= G \gamma_{xy} \rightarrow \sigma_{xy} = -zG \frac{\partial \theta_x}{\partial x} \\ \sigma_{xz} &= G \gamma_{xz} \rightarrow \sigma_{xz} = yG \frac{\partial \theta_x}{\partial x} \end{aligned}$$

where G is the shear modulus. In cylindrical coordinates, the resultant stress is:

$$\sigma_{xr} = G \gamma_{xr} \rightarrow \sigma_{xr} = rG \frac{\partial \theta_x}{\partial x}.$$

By applying Newton's laws or variational formulations, the equation of motion for this element is derived:

Spectral Finite Element Approach for Structural Dynamics, Fig. 7 Rod element model under torsion (shaft)



$$\frac{\partial}{\partial x} \left[GJ \frac{\partial \theta_x}{\partial x} \right] - \eta J \frac{\partial \theta_x}{\partial t} - \rho J \frac{\partial^2 \theta_x}{\partial t^2} = -m_x$$

where θ_x is the rotation with respect to the x -axis, J is the torsion constant or the polar moment of area with respect to the x -axis for concentric circular rods (round shafts), η is the damping (viscous) coefficient per unit volume, and ρ is the mass density.

The associated boundary conditions in this case are the rotation θ_x and the internal torque $M_x = GJ \frac{\partial \theta_x}{\partial x}$.

Spectral Analysis

As it was done in previous sections, by applying two Fourier transforms (time and space), the displacement field is written as spectral solutions, in the form:

$$\begin{aligned} \theta_x(s, t) &= \sum_n \hat{\theta}_{xn}(x, \omega_n) e^{i\omega_n t} \\ &= \sum_n \sum_m \tilde{\theta}_x(k_{xnm}, \omega) e^{-ik_m s} e^{i\omega_n t}. \end{aligned}$$

Then, assuming geometric and material properties constant along the rod length and substituting this expression in the homogeneous differential equation of motion, the spectrum relation is obtained:

$$\begin{aligned} (-GJk_{xx}^2 - i\omega\eta J + \omega^2 \rho J) \theta_x &= 0, \\ k_{xx} &= \pm \sqrt{\frac{\omega^2 \rho J - i\omega\eta}{GJ}}. \end{aligned}$$

Thus, the rotation θ_x and the internal moment M_x evaluated at an arbitrary point along the rod may be expressed as:

$$\begin{aligned} \theta_x(x, t) &= \sum_n (A_{xx1}^+ e^{-ik_{xxn}x} + A_{xx1}^- e^{ik_{xxn}x}) e^{i\omega_n t} \\ M_x(x, t) &= \sum_n ik_{xxn} GJ (-A_{xx1}^+ e^{-ik_{xxn}x} + A_{xx1}^- e^{ik_{xxn}x}) e^{i\omega_n t}. \end{aligned}$$

Setting the reference for negative-going waves at $x=L$, the rotation and the internal load are written in matrix form:

$$\begin{aligned} \hat{q}_{xx} = \hat{\theta}_x &= [1 \ 1] \begin{bmatrix} e^{-ik_{xx}x} & 0 \\ 0 & e^{ik_{xx}(x-L)} \end{bmatrix} \begin{bmatrix} A_{xx1}^+ \\ B_{xx1}^- \end{bmatrix} \\ \hat{f}_{xx} = \hat{M}_x &= ik_{xx} GJ [-1 \ 1] \begin{bmatrix} e^{-ik_{xx}x} & 0 \\ 0 & e^{ik_{xx}(x-L)} \end{bmatrix} \begin{bmatrix} A_{xx1}^+ \\ B_{xx1}^- \end{bmatrix}. \end{aligned}$$

Spectral Element

The spectral element of a shaft element in torsion of length L is straightforwardly obtained by writing the rotations with respect to the x -axis at both ends as a function of the wave amplitudes, as follows:

$$\begin{bmatrix} \hat{\theta}_x(0) \\ \hat{\theta}_x(L) \end{bmatrix} = \begin{bmatrix} 1 & e^{-ik_{xx}L} \\ e^{-ik_{xx}L} & 1 \end{bmatrix} \begin{bmatrix} A_{xx1}^+ \\ B_{xx1}^- \end{bmatrix},$$

i.e.,

$$\begin{bmatrix} \hat{q}_{xx0} \\ \hat{q}_{xxL} \end{bmatrix} = \begin{bmatrix} \Phi_{qxx}^+ & \Phi_{qxx}^- A_L^+ \\ \Phi_{qxx}^+ A_L^+ & \Phi_{qxx}^- \end{bmatrix} \begin{bmatrix} A_{xx}^+ \\ B_{xx}^- \end{bmatrix}.$$

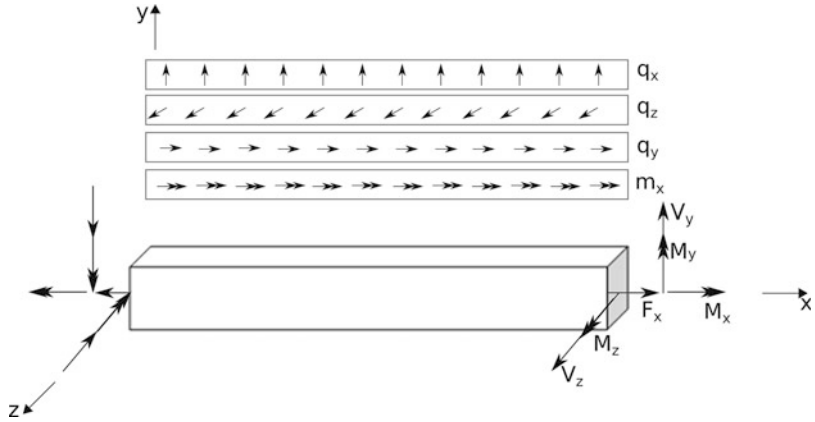
The applied torque may be also evaluated at both ends:

$$\begin{bmatrix} -\hat{M}_x(0) \\ \hat{M}_x(L) \end{bmatrix} = ik_{xx} GJ \begin{bmatrix} 1 & -e^{-ik_{xx}L} \\ -e^{-ik_{xx}L} & 1 \end{bmatrix} \begin{bmatrix} A_{xx1}^+ \\ B_{xx1}^- \end{bmatrix}$$

or

$$\begin{bmatrix} \hat{F}_{xx0} \\ \hat{F}_{xxL} \end{bmatrix} = \begin{bmatrix} -\Phi_{fxx}^+ & -\Phi_{fxx}^- A_L^+ \\ \Phi_{fxx}^+ A_L^+ & \Phi_{fxx}^- \end{bmatrix} \begin{bmatrix} A_{xx}^+ \\ B_{xx}^- \end{bmatrix}.$$

**Spectral Finite Element
Approach for Structural
Dynamics, Fig. 8** 3D
frame member



Thus, by writing the wave amplitudes as function of the rotation vector and substituting it into the expression for external applied moments, the spectral element matrix is obtained:

$$\begin{bmatrix} \hat{F}_{xx0} \\ \hat{F}_{xxL} \end{bmatrix} = \mathbf{K}_{xx}(\omega) \begin{bmatrix} \hat{q}_{xx0} \\ \hat{q}_{xxL} \end{bmatrix},$$

$$\mathbf{K}_{xx}(\omega) = \begin{bmatrix} -\Phi_{fxx}^+ & -\Phi_{fxx}^- \Lambda_L^+ \\ \Phi_{fxx}^+ \Lambda_L^+ & \Phi_{fxx}^- \end{bmatrix} \begin{bmatrix} \Phi_{qxx}^+ & \Phi_{qxx}^- \Lambda_L^+ \\ \Phi_{qxx}^+ \Lambda_L^+ & \Phi_{qxx}^- \end{bmatrix}^{-1},$$

$$\mathbf{K}_{xx}(\omega) = GJk_{xx} \sin(k_{xx}L) \begin{bmatrix} \cos(k_{xx}L) & -1 \\ -\cos(k_{xx}L) & 1 \end{bmatrix}.$$

3D Frame Structure

A 3D frame structure is a combination of one-dimensional elements in the 3D space that are capable to support axial, bending, and torsion efforts (Fig. 8). Thus, using the expressions derived in the previous sections, a spectral element for a 3D frame element can be easily built. It is composed of six degrees of freedom per node (three translations and three rotations), which allow an axial deformation and torsion along the main axis and bending in the two transverse planes. Its displacement field may be expressed as:

$$\begin{aligned} \bar{u}(x, y, z, t) &= u(x, t) & \bar{\theta}_x(x, y, z, t) &= \theta_x(x, t) \\ \bar{v}(x, y, z, t) &= v(x, t) & \bar{\theta}_y(x, y, z, t) &= \theta_y(x, t) = -\frac{\partial w(x, t)}{\partial x} \\ \bar{w}(x, y, z, t) &= w(x, t) & \bar{\theta}_z(x, y, z, t) &= \theta_z(x, t) = \frac{\partial v(x, t)}{\partial x} \end{aligned}$$

It is important to emphasize at this point that the expression for the rotation w.r.t. the y-axis requires a negative sign as a result of the assumption that translations are positive in the sense of the Cartesian coordinates and also the respect of the right-hand rule for the coordinate system. By superposing the spectral solutions for a beam element under axial, flexural, and torsion deformations, the displacement solution vector is given by:

$$\hat{\mathbf{q}}_{3D}(x) = \begin{bmatrix} \hat{u}(x) \\ \hat{v}(x) \\ \hat{w}(x) \\ \hat{\theta}_x(x) \\ \hat{\theta}_y(x) \\ \hat{\theta}_z(x) \end{bmatrix} = [\Phi_q^+ \quad \Phi_q^-] \begin{bmatrix} \Lambda_x^+ & 0 \\ 0 & \Lambda_x^- \end{bmatrix} \begin{bmatrix} \mathbf{A}^+ \\ \mathbf{B}^- \end{bmatrix},$$

where:

$$\Phi_q^+ = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & i\beta_z & \beta_z & 0 \\ 0 & -i\beta_y & -\beta_y & 0 & 0 & 0 \end{bmatrix},$$

$$\Phi_q^- = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -i\beta_z & -\beta_z & 0 \\ 0 & i\beta_y & \beta_y & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{\Lambda}_x^+ = \begin{bmatrix} e^{-ik_x x} & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{-i\beta_y x} & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{-\beta_y x} & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{-i\beta_z x} & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{-\beta_z x} & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{-ik_{xx} x} \end{bmatrix},$$

$$\mathbf{\Lambda}_x^- = \begin{bmatrix} e^{ik_x(x-L)} & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{i\beta_y(x-L)} & 0 & 0 & 0 & 0 \\ 0 & 0 & e^{\beta_y(x-L)} & 0 & 0 & 0 \\ 0 & 0 & 0 & e^{i\beta_z(x-L)} & 0 & 0 \\ 0 & 0 & 0 & 0 & e^{\beta_z(x-L)} & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{ik_{xx}(x-L)} \end{bmatrix}.$$

The corresponding internal loads acting on a 3D frame member are:

$$\begin{aligned} F_x &= EA \frac{\partial u}{\partial x} & M_x &= GJ \frac{\partial \theta_x}{\partial x} \\ V_y &= -\frac{\partial}{\partial x} \left(EI_z \frac{\partial^2 v}{\partial x^2} \right) & M_y &= -EI_y \frac{\partial^2 w}{\partial x^2} \\ V_z &= -\frac{\partial}{\partial x} \left(EI_y \frac{\partial^2 w}{\partial x^2} \right) & M_z &= EI_z \frac{\partial^2 v}{\partial x^2} \end{aligned}$$

$$\begin{aligned} \hat{\mathbf{f}}_{3D}(x) &= \begin{bmatrix} \hat{F}_x(x) \\ \hat{V}_y(x) \\ \hat{V}_z(x) \\ \hat{M}_x(x) \\ \hat{M}_y(x) \\ \hat{M}_z(x) \end{bmatrix} \\ &= [\Phi_f^+ \quad \Phi_f^-] \begin{bmatrix} \mathbf{\Lambda}_x^+ & 0 \\ 0 & \mathbf{\Lambda}_x^- \end{bmatrix} \begin{bmatrix} \mathbf{A}^+ \\ \mathbf{B}^- \end{bmatrix}, \end{aligned}$$

with:

which, in vector form, yields:

$$\Phi_f^+ = \begin{bmatrix} -ik_x EA & 0 & 0 & 0 & 0 & 0 \\ 0 & -iEI_z \beta_y^3 & EI_z \beta_y^3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -iEI_y \beta_z^3 & EI_y \beta_z^3 & 0 \\ 0 & 0 & 0 & 0 & 0 & -ik_{xx} GJ \\ 0 & 0 & 0 & EI_y \beta_z^2 & -EI_y \beta_z^2 & 0 \\ 0 & -EI_z \beta_y^2 & EI_z \beta_y^2 & 0 & 0 & 0 \end{bmatrix},$$

$$\Phi_f^- = \begin{bmatrix} -ik_x EA & 0 & 0 & 0 & 0 & 0 \\ 0 & iEI_z \beta_y^3 & -EI_z \beta_y^3 & 0 & 0 & 0 \\ 0 & 0 & 0 & iEI_y \beta_z^3 & -EI_y \beta_z^3 & 0 \\ 0 & 0 & 0 & 0 & 0 & ik_{xx} GJ \\ 0 & 0 & 0 & EI_y \beta_z^2 & -EI_y \beta_z^2 & 0 \\ 0 & -EI_z \beta_y^2 & EI_z \beta_y^2 & 0 & 0 & 0 \end{bmatrix}.$$

A spectral element of frame with length L is straightforwardly obtained by writing displacement and external loads at both ends, as follows:

$$\begin{bmatrix} \hat{\mathbf{q}}_{3D0} \\ \hat{\mathbf{q}}_{3DL} \end{bmatrix} = \begin{bmatrix} \Phi_q^+ & \Phi_q^- \Lambda_L^+ \\ \Phi_q^+ \Lambda_L^+ & \Phi_q^- \end{bmatrix} \begin{bmatrix} \mathbf{A}^+ \\ \mathbf{B}^- \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mathbf{F}}_{3D0} \\ \hat{\mathbf{F}}_{3DL} \end{bmatrix} = \begin{bmatrix} -\hat{\mathbf{f}}_{3D}(x=0) \\ \hat{\mathbf{f}}_{3D}(x=L) \end{bmatrix}$$

$$= \begin{bmatrix} -\Phi_f^+ & -\Phi_f^- \Lambda_L^+ \\ \Phi_f^+ \Lambda_L^+ & \Phi_f^- \end{bmatrix} \begin{bmatrix} \mathbf{A}^+ \\ \mathbf{B}^- \end{bmatrix}.$$

Then, the spectral element matrix of a 3D frame member is readily obtained:

$$\begin{bmatrix} \hat{\mathbf{F}}_{3D0} \\ \hat{\mathbf{F}}_{3DL} \end{bmatrix} = \mathbf{K}_{3D}(\omega) \begin{bmatrix} \hat{\mathbf{q}}_{3D0} \\ \hat{\mathbf{q}}_{3DL} \end{bmatrix},$$

$$\mathbf{K}_{3D}(\omega) = \begin{bmatrix} -\Phi_f^+ & -\Phi_f^- \Lambda_L^+ \\ \Phi_f^+ \Lambda_L^+ & \Phi_f^- \end{bmatrix} \begin{bmatrix} \Phi_q^+ & \Phi_q^- \Lambda_L^+ \\ \Phi_q^+ \Lambda_L^+ & \Phi_q^- \end{bmatrix}^{-1}.$$

Spectral Element Modeling

In previous sections, it has been shown how to formulate spectral element matrices for straight and curved frame members. However, in real engineering systems, those structures do not appear isolated; they are usually connected to other basic structures involving or not discontinuities. In this section, it will be shown that the spectral element method can be efficiently used to compute the forced response of coupled systems in the frequency and in the time domains. This is particularly useful in earthquake engineering, as it allows simulating the effect of seismic ground motion at a very low computational cost compared to conventional computational analyses via FEM or BEM. This is possible as SEM allows building exact element matrices at each frequency. Therefore, mesh refinement is not

required with the increase in frequency, and it is possible to build spectral models with very few degrees of freedom. On the other hand, as it makes use of matrix formulations as FEM does. The assemblage of spectral elements is quite simple via the direct stiffness method (Doyle 1997). It is also important to note that the global problem is solved in the frequency domain. For every discrete frequency, an assemblage process identical to that of the static case is performed and the global response vector is obtained. The discrete Fourier transform is used to obtain the spectral representation of the input data, and the inverse DFT allows one to recover the response of the coupled system in the time domain.

Global Dynamic Stiffness Matrix Assembling Procedure

Consider a frame element as shown in Fig. 9. Two types of coordinate system are defined: a global one, defined by the orthogonal X , Y , and Z directions, and a local one, defined by x , y , and z axes with the x -axis parallel to the main axis (neutral axis) of the frame member.

The spectral element matrix, as formulated previously for frame member, is written with respect to the local reference frame, as follows:

$$\hat{\mathbf{K}}_{xyz}(\omega) \hat{\mathbf{q}}_{xyz}(\omega) = \hat{\mathbf{F}}_{xyz}(\omega).$$

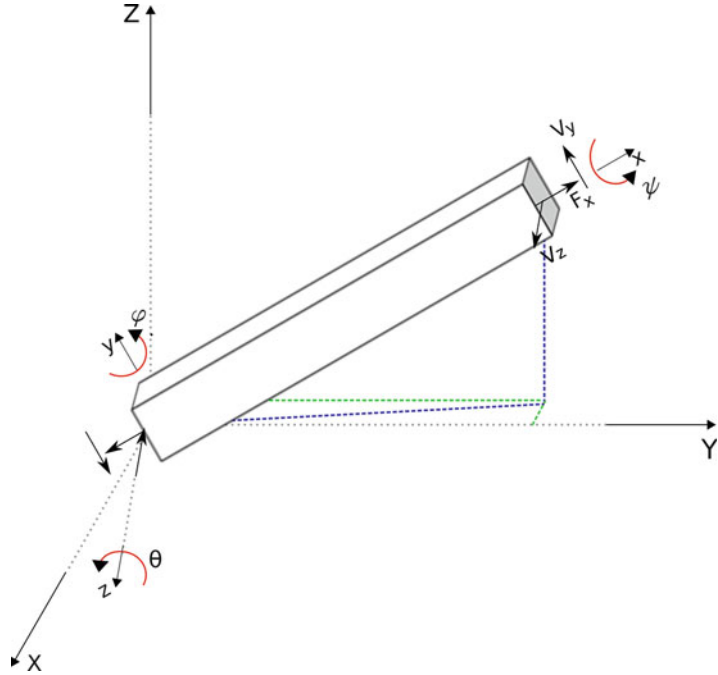
Load and displacement vectors in the global reference frame may be related to those defined in the local coordinate system by application of successive rotations from the local axes to the global axes:

$$\begin{aligned} \hat{\mathbf{q}}_{XYZ} &= \mathbf{T} \hat{\mathbf{q}}_{xyz} \\ \hat{\mathbf{F}}_{XYZ} &= \mathbf{T} \hat{\mathbf{F}}_{xyz} \end{aligned}$$

with:

$$\mathbf{T} = \begin{bmatrix} \mathbf{R}_x(\psi) \mathbf{R}_y(\varphi) \mathbf{R}_z(\theta) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_x(\psi) \mathbf{R}_y(\varphi) \mathbf{R}_z(\theta) & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{R}_x(\psi) \mathbf{R}_y(\varphi) \mathbf{R}_z(\theta) \end{bmatrix},$$

Spectral Finite Element Approach for Structural Dynamics, Fig. 9 Frame element relative to local and global coordinate systems



where $\mathbf{R}_x$, $\mathbf{R}_y$, and $\mathbf{R}_z$ are rotation matrices, therefore unitary ($\mathbf{R}^T \mathbf{R} = \mathbf{I}$), defined as:

$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \psi & \sin \psi \\ 0 & -\sin \psi & \cos \psi \end{bmatrix},$$

$$\mathbf{R}_y = \begin{bmatrix} \cos \varphi & 0 & -\sin \varphi \\ 0 & 1 & 0 \\ \sin \varphi & 0 & \cos \varphi \end{bmatrix},$$

$$\mathbf{R}_z = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Applying the transformation matrix to the local system, the equilibrium system of equations for a frame member in the global reference frame is written:

$$\hat{\mathbf{K}}_{XYZ}(\omega) \hat{\mathbf{q}}_{XYZ}(\omega) = \hat{\mathbf{F}}_{XYZ}(\omega),$$

with:

$$\hat{\mathbf{K}}_{XYZ}(\omega) = \mathbf{T} \hat{\mathbf{K}}_{xyz}(\omega) \mathbf{T}^T.$$

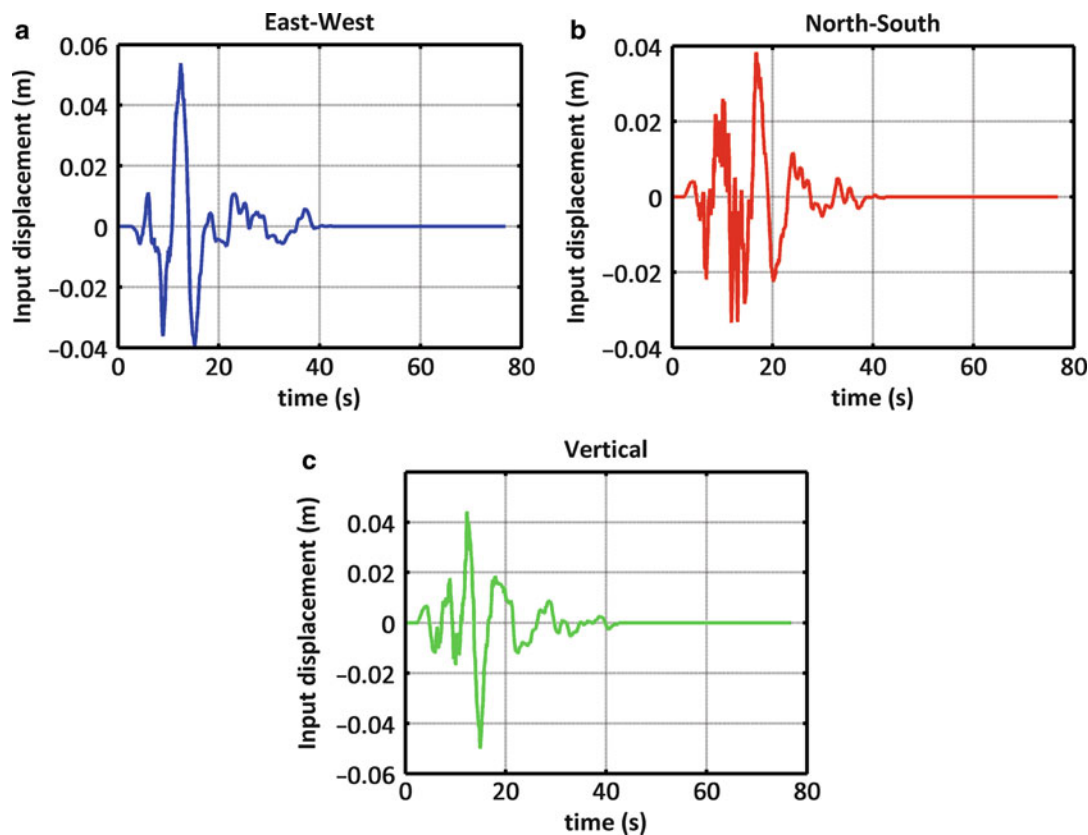
The transformation of coordinate system is performed for each structural member. This allows one to enforce kinematic compatibility and equilibrium at all nodes connecting structural members. Then, the global dynamic system is constructed following classic direct stiffness procedure:

$$\hat{\mathbf{K}}_G(\omega) = \sum_i \mathbf{L}^{(i)T} \hat{\mathbf{K}}_{XYZ}^{(i)}(\omega) \mathbf{L}^{(i)},$$

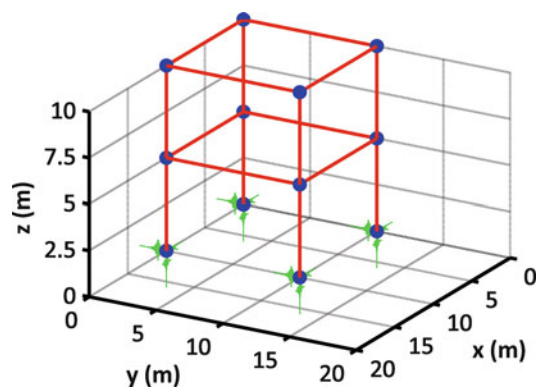
where $\mathbf{L}^{(i)}$ is the incidence matrix linking the local vector of degrees of freedom of a structural member i to the global vector of degrees of freedom.

For each discrete frequency, the global problem to be solved is:

$$\hat{\mathbf{K}}_G \hat{\mathbf{q}}_G = \hat{\mathbf{F}}_G,$$



Spectral Finite Element Approach for Structural Dynamics, Fig. 10 Seismic records using as input displacements in numerical experiments



Spectral Finite Element Approach for Structural Dynamics, Fig. 11 3D frame model

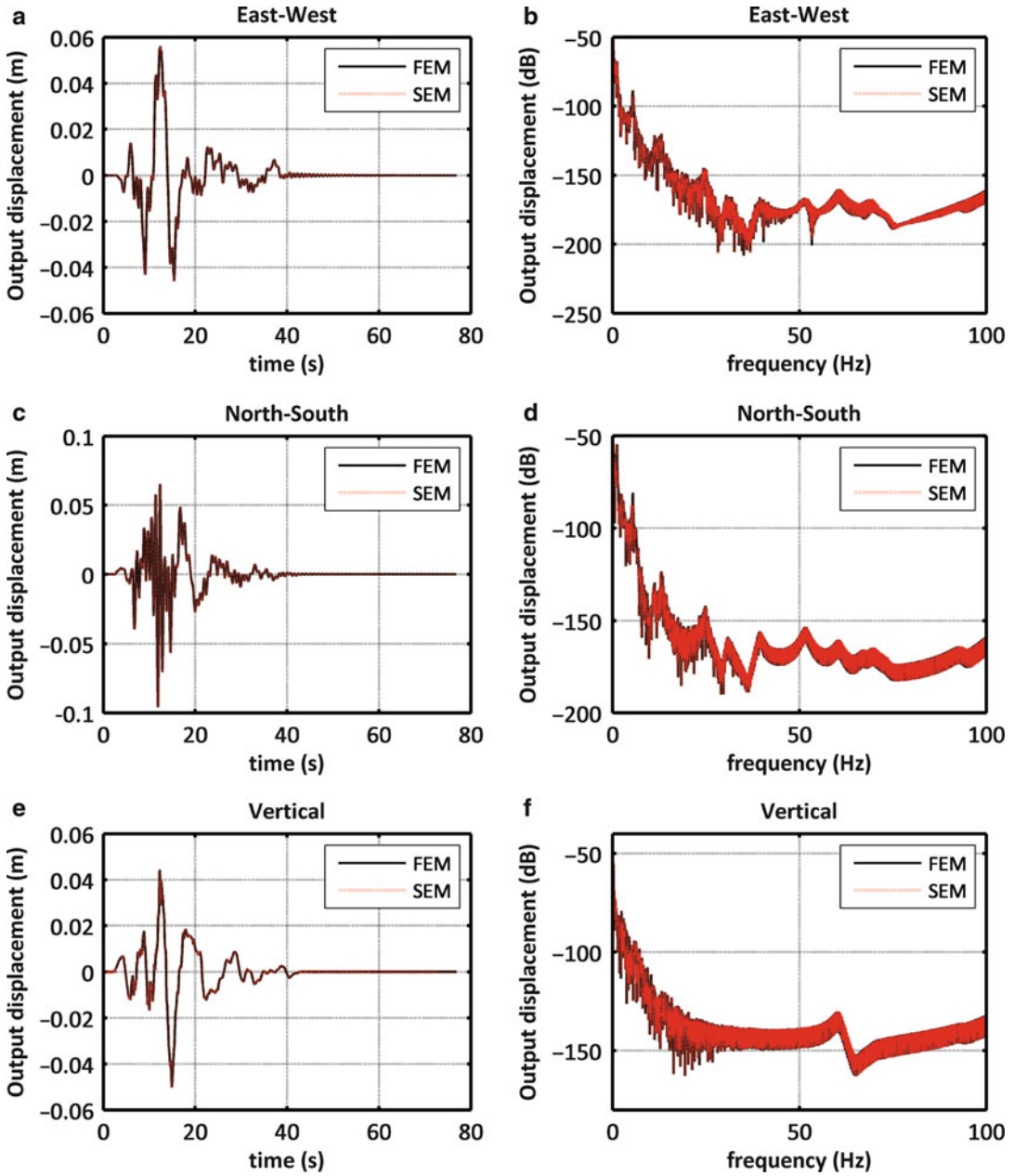
subjected to:

$$\begin{aligned}\hat{\mathbf{F}}_F &= \hat{\mathbf{F}}_0, \\ \hat{\mathbf{q}}_P &= \hat{\mathbf{q}}_0,\end{aligned}$$

where the subscripts F and P are used, respectively, to denote sets of degrees of freedom where Neumann (prescribed loads) and Dirichlet (prescribed displacements/rotations) boundary conditions are applied.

The solution of the global system is possible by rewriting it as:

$$\begin{bmatrix} \hat{\mathbf{K}}_{FF} & \hat{\mathbf{K}}_{FP} \\ \hat{\mathbf{K}}_{PF} & \hat{\mathbf{K}}_{PP} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{q}}_F \\ \hat{\mathbf{q}}_0 \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{F}}_0 \\ \hat{\mathbf{F}}_P \end{bmatrix},$$



Spectral Finite Element Approach for Structural Dynamics, Fig. 12 Response of the 3D frame structure to the earthquake measured at $(x = 10, y = 10)$ m

which allows one to get:

$$\begin{aligned}\hat{\mathbf{q}}_F &= \hat{\mathbf{K}}_{FF}^{-1}(\hat{\mathbf{F}}_0 - \hat{\mathbf{K}}_{FP}\hat{\mathbf{q}}_0) \\ \hat{\mathbf{F}}_P &= (\hat{\mathbf{K}}_{PP} - \hat{\mathbf{K}}_{PF}\hat{\mathbf{K}}_{FF}^{-1}\hat{\mathbf{K}}_{FP})\hat{\mathbf{q}}_0 + \hat{\mathbf{K}}_{PF}\hat{\mathbf{K}}_{FF}^{-1}\hat{\mathbf{F}}_0.\end{aligned}$$

Numerical Examples

Earthquake Signal

Seismic records from the Loma Prieta earthquake which struck the San Francisco Bay area of

California in 1989 are used as input displacements applied to the bottom nodes of the framework. The input signal in time domain has been obtained from an online database available in Silva et al. (2000) and it is shown in Fig. 10.

3D Frame

Here, the dynamic response of a 3D framework subjected to seismic ground motion is analyzed. The model is presented in Fig. 11. Each frame member is made of steel with mass density $\rho = 7,800 \text{ Kg.m}^{-3}$, Young's modulus $E = 210 \text{ GPa}$, structural loss factor $\eta = 0.04$, and Poisson's ratio $\nu = 0.3$. All members have a circular cross section of radius 0.15 m.

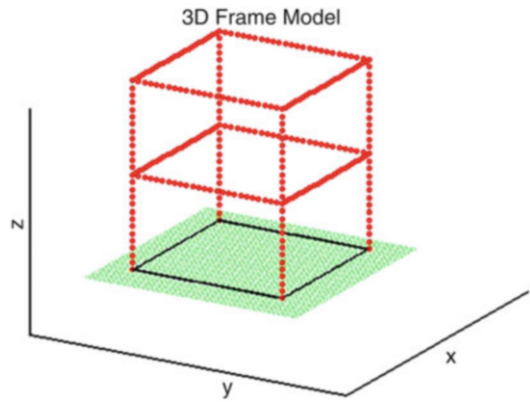
The response of the structure at $(x = 10, y = 10) \text{ m}$ along the three orthogonal directions is shown in Fig. 12, in the time and in the frequency domains. It is worth pointing out that the SEM has yielded a great reduction of the problem size compared to classic FEM. The response obtained from the SEM using 16 elements is in good agreement with the one obtained by solving a classic FE problem composed of 1200 FE elements. An animation of the framework dynamics during the passage of seismic waves is simulated via the SEM and it is shown in Video 1.

2D Frame

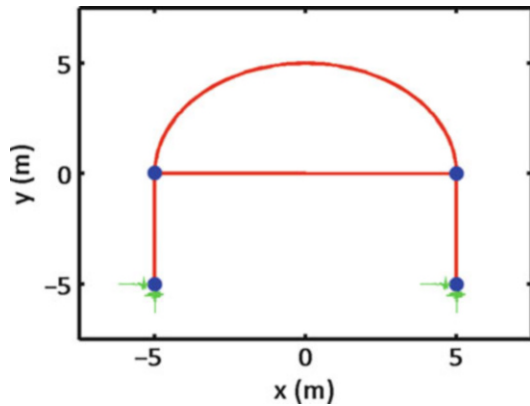
A 2D frame structure modeled using straight and curved beam elements is shown in Fig. 13. The frame has a square cross section $0.20 \text{ m} \times 0.20 \text{ m}$ and steel as material, with Young's modulus $E = 210 \text{ GPa}$, mass density $\rho = 7,800 \text{ kg/m}^3$, and structural damping $\eta = 0.02$.

As in the previous section, the structure is excited at the base nodes by the Loma Prieta earthquake signal. Because of the in-plane configuration, only the east–west and vertical displacement data are used.

The output displacements at the node $(x = 5, y = 0) \text{ m}$ are shown in Fig. 14. The results are validated with FEM results obtained using

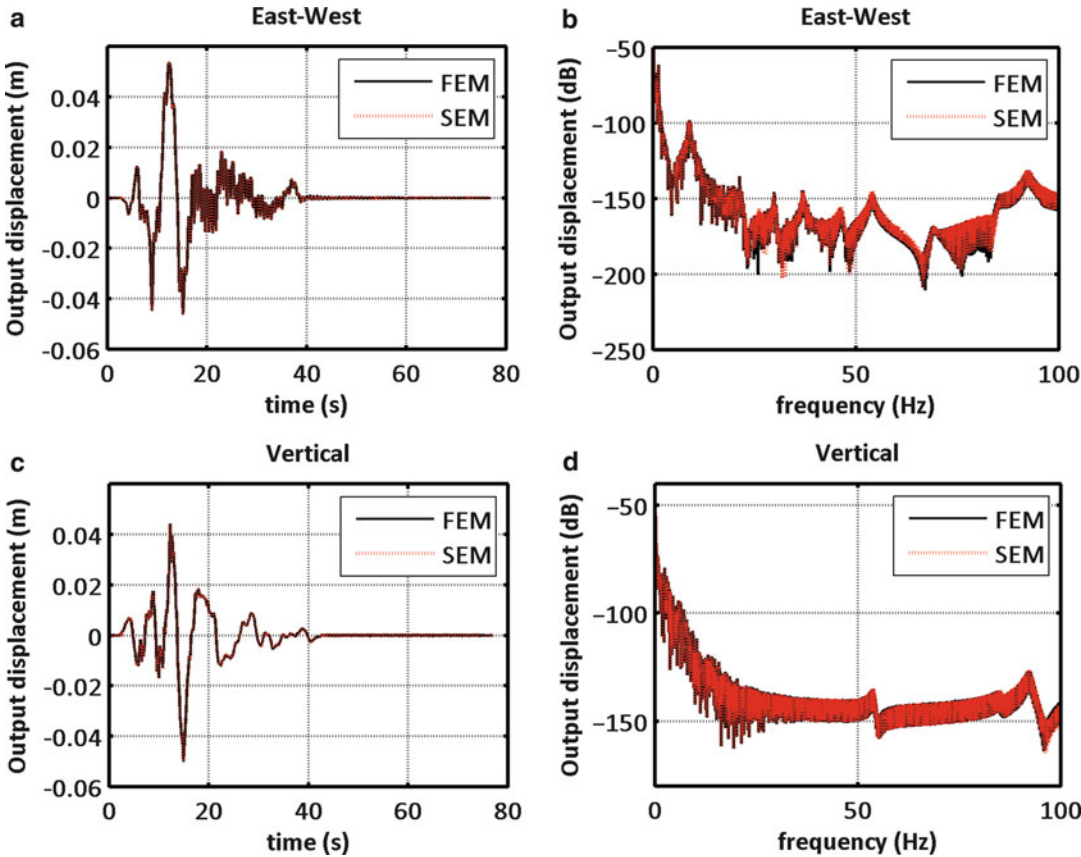


Spectral Finite Element Approach for Structural Dynamics, Video 1 Animation of a 3D frame under seismic ground motion

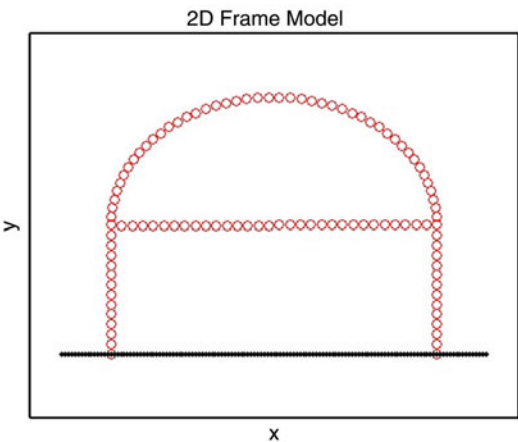


Spectral Finite Element Approach for Structural Dynamics, Fig. 13 2D frame model

ANSYS®. As in the previous numerical problem, fewer degrees of freedom were required to solve the problem via the spectral element method compared to the classic FEM, which results in CPU time savings. The 2D frame motion caused by the earthquake passage was computed via the SEM and it is available in Video 2. The methodology presented in this chapter can be used to derive other elements, such as plates and shells (Doyle 1997), and to study more complex structures under dynamic excitation.



Spectral Finite Element Approach for Structural Dynamics, Fig. 14 Response of the 2D frame structure to the earthquake measured at $(x = 5, y = 0)$ m



Spectral Finite Element Approach for Structural Dynamics, Video 2 Animation of a 2D frame under seismic ground motion

Summary

Elementary matrix of straight and curved beams, rod under axial force, and torsion were developed using spectral finite element approach. This method can be used to analyze the structural dynamic behavior. From the perspective of earthquake engineering, SEM allows to design structures subject to seismic excitation. The approach has advantages such as low computational cost, accuracy in frequency/time response, and simple modeling of complex structures. The equations of motion, spectral relation, and the spectral elementary matrix were derived. The simple elements were combined into a 3D frame element and a global matrix was assembled. Time and frequency responses of 3D and 2D frame structures under a

known seismic ground motion were computed with SEM and results were validated using well-known FEM commercial software.

Cross-References

- [Nonlinear Dynamic Seismic Analysis](#)
- [Nonlinear Finite Element Analysis](#)
- [Online Response Estimation in Structural Dynamics](#)
- [Operational Modal Analysis in Civil Engineering: An Overview](#)
- [Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach](#)
- [Robust Design Optimization for Earthquake Loads](#)
- [Steel Structures](#)
- [Stochastic Analysis of Linear Systems](#)
- [Stochastic Finite Elements](#)
- [Structural Optimization Under Random Dynamic Seismic Excitation](#)
- [Structures with Nonviscous Damping, Modeling, and Analysis](#)
- [Substructuring Methods for Finite Element Analysis](#)
- [Time History Seismic Analysis](#)

References

- Ahmida KM, Arruda JRF (2001) Spectral element-based prediction of active power flow in Timoshenko beams. *Int J Solids Struct* 38:1669–1679
- Doyle JF (1997) Wave propagation in structures, 2nd edn. Springer, New York
- Lee U (2009) Spectral element method in structural dynamic. Wiley, Singapore
- Mace BR, Duhamel D, Brennan MJ, Hinkle L (2005) Finite element prediction of wave motion in structural waveguides. *J Acoust Soc Am* 117(5):2835–2843
- Mead DJ (1973) A general theory of harmonic wave propagation in linear periodic systems with multiple coupling. *J Sound Vib* 27(2):235–260
- Newmark NM, Hall WJ (1982) Earthquake spectra and design. Earthquake Engineering Research Institute, Berkeley
- Petyt M (2010) Introduction to finite element vibration analysis, 2nd edn. Cambridge University Press, Cambridge
- Silva W, Bartling D, Spitzer C (2000) PEER strong motion database. Regents of the University of California. <http://peer.berkeley.edu/smcat/search.html>. Accessed 9 July 2014

Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures

George Vasdravellis¹ and
Theodore L. Karavasilis²

¹Institute for Infrastructure and Environment,
Heriot-Watt University, Edinburgh, UK

²School of Engineering, University of Warwick,
Coventry, UK

Synonyms

Mitigation of residual drifts; Optimized seismic dampers; Seismic resilience; Steel self-centering frames

Introduction

Conventional steel moment-resisting frames (MRFs) are currently designed to form a global plastic mechanism under the design basis earthquake (DBE) by developing plastic hinges at the ends of the beams and the bases of the columns. This design methodology offers many advantages, including collapse prevention and initial economy; however, plastic hinges in structural members involve difficulty to inspect and repair damage and local buckling as well as residual drifts. The socioeconomic losses associated with damage and residual drifts are repair costs, increased downtime, and possibly demolition due to the complications associated with large residual drifts (McCormick et al. 2008).

A challenge of modern earthquake engineering is the development, standardization, and practical implementation of resilient minimal-damage structures with the inherent potential to overcome the socioeconomic losses related to earthquake damage by minimizing or avoiding inelastic deformations and residual drifts. Examples of minimal-damage steel structures include frames equipped with rate-dependent passive dampers (Karavasilis et al. 2011; Whittle et al. 2012),

frames with steel energy dissipation devices (Karavasilis et al. 2012), and self-centering moment-resisting frames (SC-MRFs) with posttensioned (PT) connections.

SC-MRFs exhibit softening force–drift behavior while eliminating inelastic deformations and residual drifts under strong earthquakes as the result of gap openings developed in beam-to-column interfaces and self-centering capability due to elastic pretensioning elements (e.g., high-strength steel bars) which clamp beams to the columns. PT connections use carefully designed energy dissipation devices which are activated when gaps open and can be classified into yielding devices which dissipate energy through inelastic deformations and devices which dissipate energy through friction. Yielding devices were proposed as angles bolted on the top and bottom flanges of the beam and on the column flanges, dissipating energy through inelastic bending (Ricles et al. 2001, 2002; Garlock et al. 2005); buckling-restrained steel bars placed between the beam flanges and welded on the beam and column, dissipating energy through axial deformations (Christopoulos et al. 2002); reduced flange plates welded around a square hollow section column and bolted on the beam flanges (Chou et al. 2006); and reduced-section or cross-shaped steel plates placed below the bottom flange of the beam (Chou and Lai 2009). Friction-based devices were proposed as friction-bolted surfaces placed on the top and bottom flanges of the beam (Rojas et al. 2004; Kim and Christopoulos 2008a), on the web of the beam (Tsai et al. 2008), or on the bottom flange of the beam (Wolski et al. 2009).

A new steel PT beam–column connection using web hourglass shape steel pins (WHPs) as EDs has been recently developed by the authors. The development phases included (a) large-scale experimental evaluation (Vasdravellis et al. 2013a) of the proposed connection, (b) detailed finite element models (FEM) and associated parametric studies (Vasdravellis et al. 2013b), and (c) numerical simulations and seismic analysis of SC-MRFs equipped with PT connections with WHPs (Dimopoulos et al. 2013). A more recent study evaluated the

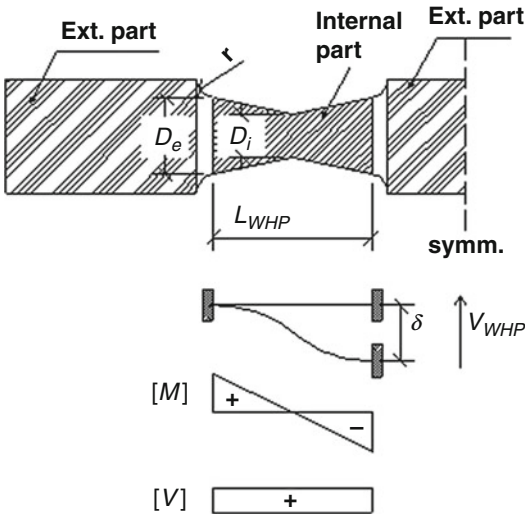
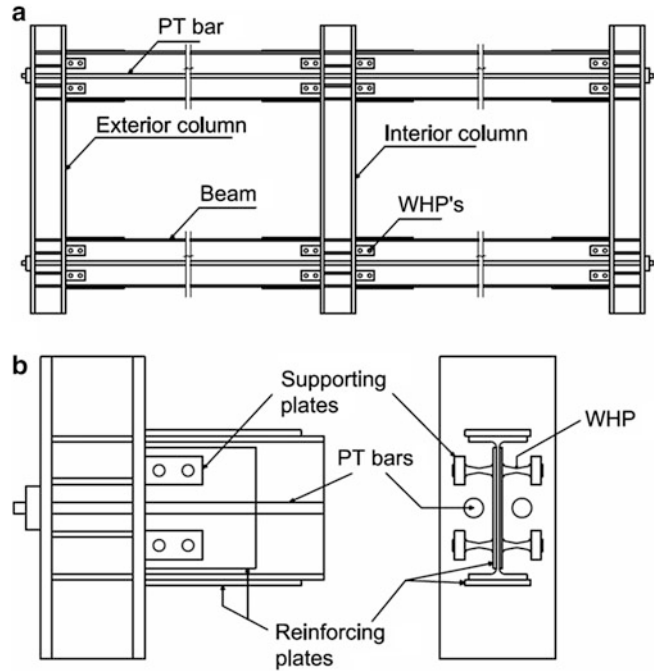
hysteresis and fracture capacity of WHPs made of high performance steel, i.e., high-strength steel and stainless steel (Vasdravellis et al. 2014). It was found that the connection isolates inelastic deformations in WHPs, avoids damage in other connection parts as well as in beams and columns, and eliminates residual drifts. The WHPs do not interfere with the composite slab and are very easy to replace without bolting or welding, and so, the connection enables nondisruptive repair and rapid return to building occupancy in the aftermath of a strong earthquake. This chapter presents the development of the PT connection through the experimental evaluation and the main results of the numerical analyses.

Steel Posttensioned Connection with Web Hourglass Pins

Figure 1 shows a schematic representation of an SC-MRF incorporating the proposed PT connection and the details of an exterior PT connection. Two high-strength steel bars located at the mid-depth of the beam, one at each side of the web, pass through holes drilled on the column flanges. The bars are posttensioned and anchored to the exterior column flange and hence clamp the beam to the column. Four cylindrical steel pins (WHPs) are inserted in aligned holes drilled on the web of the beam and on strong supporting plates. The supporting plates are welded to the column flanges and are of greater thickness to provide fixed support boundary conditions to the WHPs. Energy dissipation is provided by the inelastic bending of the WHPs, which are symmetrically placed (close to the top and bottom beam flange) to provide increased lever arm and hence increased internal moment resistance. As shown in Fig. 2, the WHPs are designed to have an hourglass shape to provide enhanced energy dissipation and fracture capacity (Kobori et al. 1992). Both sides of the beam web are reinforced with steel plates to increase the contact surface of the WHPs with the web. In that way, possible ovalization of the holes drilled on the web and the reinforcing plates under the WHP-bearing forces will be negligible, and

Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures,

Fig. 1 (a) SC-MRF incorporating the proposed PT connection; (b) exterior PT connection details



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 2 Geometry of half a WHP, assumed static system, and internal force diagrams

pinching behavior under cyclic deformations can be avoided. The connection includes beam flange-reinforcing plates to avoid excessive early yielding in the beam flanges under high

PT bar forces. In addition, the panel zone is strengthened with horizontal stiffeners and continuity plates along the web of the column.

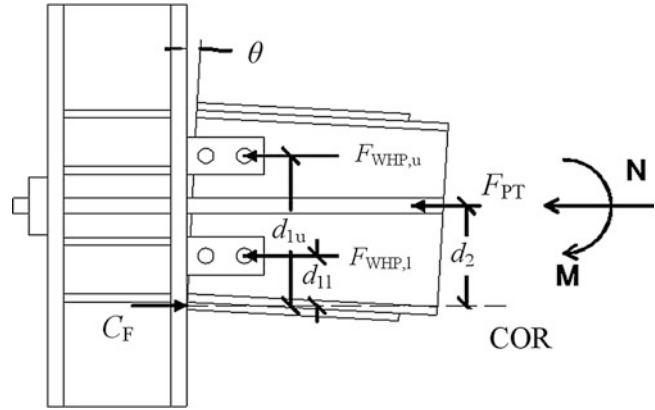
Connection Behavior and Simplified Analysis

Flexural Behavior

The connection behavior is characterized by gap opening and closing in the beam–column interface as a result of the re-centering force in the PT bars. Figure 3 shows the gap-opening mechanism in the connection where d_{1u} and d_{1l} are the distances of the upper and lower WHPs from the center of rotation (COR), respectively; d_2 is the distance of the PT bars from the COR; F_{PT} is the total force in both PT bars; $F_{WHP,u}$ and $F_{WHP,l}$ are the forces in the upper and lower WHPs, respectively; and C_F is the compressive force on the beam–column bearing surface. It is assumed that the COR is located at the inner edge of the beam flange-reinforcing plate. This assumption was verified during the connection experiments described later.

Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures,

Fig. 3 Gap-opening mechanism in the proposed PT connection



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures,

Fig. 4 Force versus displacement cyclic behavior of one WHP

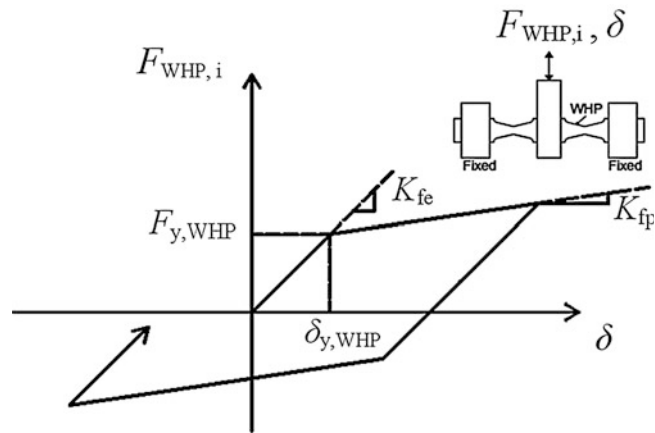
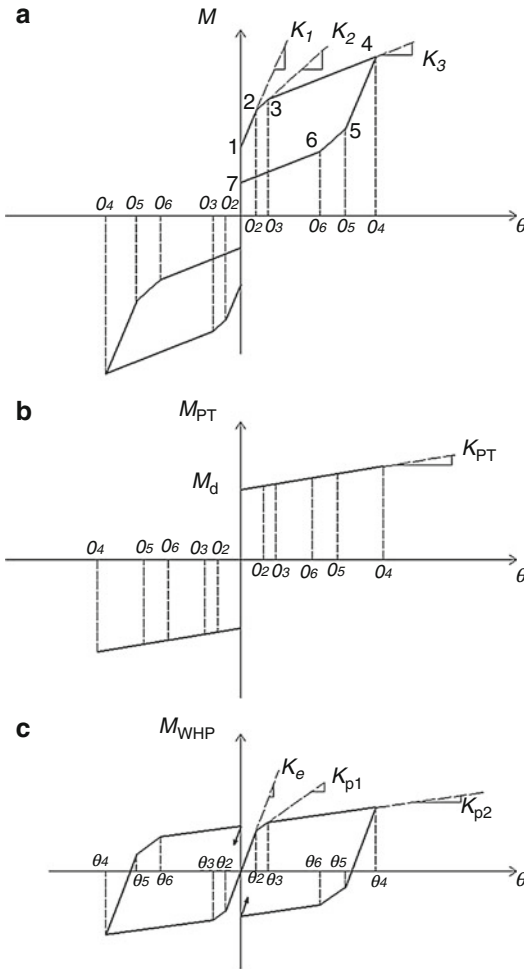


Figure 4 shows the assumed bilinear elastoplastic force–displacement ($F_{WHP,i}$ – δ) behavior of one of the four ($i = 1$ to 4) WHPs. Figure 5a plots the assumed cyclic moment–rotation (M – θ) behavior of the connection. M is the sum of the moment contributions from the PT bars, M_{PT} , and the WHPs, M_{WHP} . Figure 5b shows the nonlinear elastic moment contribution of both PT bars. In Fig. 5b, M_d is the decompression moment, i.e., equal to $F_{PT,i}d_2$, where $F_{PT,i}$ is the total initial posttensioning force in both PT bars. Figure 5c shows the moment contribution of the four WHPs. The M_{WHP} – θ curve changes slope two times since the upper WHPs yield at a rotation θ_2 and the lower WHPs yield afterwards at rotation θ_3 . Point 1 in Fig. 5a corresponds to the decompression moment, M_d , of the connection. After decompression, gap

opens, and the connection behavior becomes nonlinear elastic. Following the procedure described in Garlock et al. (2007), the stiffness K_1 between points 1 and 2 is equal to the sum of the rotational stiffness contribution of the PT bars, K_{PT} , and the rotational stiffness contribution of the WHPs, K_e :

$$K_1 = K_{PT} + K_e = \frac{K_{PT,a}K_b}{K_{PT,a} + K_b}d_2^2 + 2K_{fe}(d_{1l}^2 + d_{1u}^2) \quad (1)$$

where K_{fe} is the elastic stiffness of the force–displacement relationship of one WHP shown in Fig. 4, $K_{pt,a}$ is the total axial elastic stiffness of both PT bars, and K_b is the axial elastic stiffness of the beam. At point 2, the upper WHPs yield and M continues to increase with slope K_2 :



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 5 Conceptual cyclic moment-rotation behavior of the proposed PT connection

$$K_2 = K_{PT} + K_{p1}$$

$$= \frac{K_{PT,a}K_b}{K_{PT,a} + K_b}d_2^2 + 2(K_{fe}d_{1l}^2 + K_{fp}d_{1u}^2) \quad (2)$$

where K_{fp} is the post-elastic stiffness of the force-displacement relationship of one WHP shown in Fig. 4. At point 3, the lower WHPs yield and M continues to increase with slope K_3 :

$$K_3 = K_{PT} + K_{p2}$$

$$= \frac{K_{PT,a}K_b}{K_{PT,a} + K_b}d_2^2 + 2K_{fp}(d_{1l}^2 + d_{1u}^2) \quad (3)$$

When loading is reversed, the connection begins to unload, from point 4 to point 5, ideally with the same elastic stiffness, K_1 . After point 5, the connection unloads with stiffness equal to K_2 between points 5 and 6 and with stiffness K_3 between points 6 and 7 until the gap closes. The behavior of the connection is symmetric, since the WHPs are placed symmetrically to the beam centerline. The moment in the connection, M , is calculated as:

$$M = M_{PT} + M_{WHP} \quad (4)$$

Following the procedure described by Garlock et al. (2007), M_{PT} is given by:

$$M_{PT} = M_d + \frac{K_{PT,a}K_b}{K_{PT,a} + K_b}d_2^2\theta \quad (5)$$

According to Fig. 3, M_{WHP} can be obtained as:

$$M_{WHP} = F_{WHP,u}d_{1u} + F_{WHP,l}d_{1l} \quad (6)$$

Plastic Design of WHPs

Plastic analysis is used to calculate the force in the WHPs as they deform due to gap opening and closing. The assumed static system for half a WHP is shown in Fig. 2. The yield strength, V_{WHP} , of half a WHP is controlled either by the plastic moment of resistance, M_{pl} , or the plastic shear resistance, V_{pl} , (Eurocode 3 2003), i.e.:

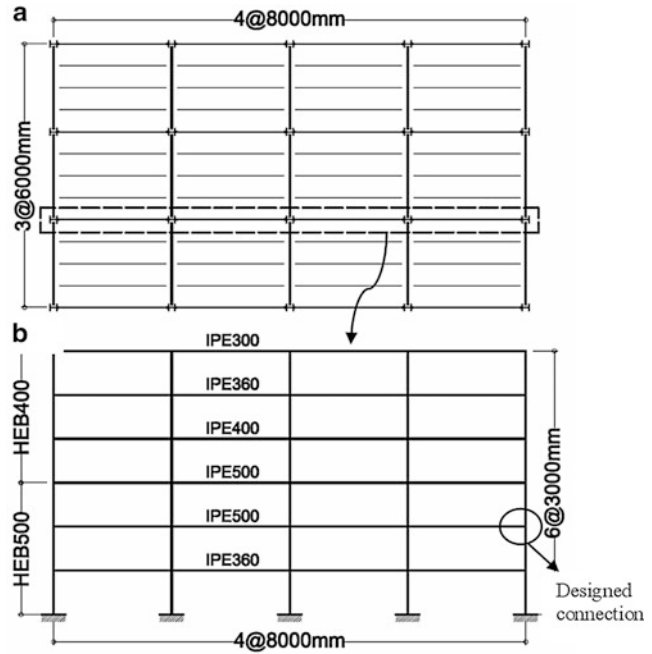
$$M_{pl} = \frac{D_e^3}{6}f_y \quad (7)$$

$$V_{pl} = \frac{0.9\pi D_i^2 f_y}{\sqrt{3}} \quad (8)$$

where f_y is the yield strength of the WHP material, D_e is the equivalent external diameter (to be defined later), and D_i is the diameter at the mid-length of half a WHP, as indicated in Fig. 2. The factor 0.9 in Eq. 8 accounts for the relation between the average shear stress and the maximum shear stress in a circular section. The internal WHP part is connected to the external WHP part using a radius of 5 mm to avoid stress

Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures,

Fig. 6 Prototype steel building: (a) plan view and (b) elevation and cross-sections of the interior steel MRF used in this study



concentration and early fracture. It is assumed that D_e , which controls the WHP bending resistance, is equal to the diameter at the start of the round-shaped part with radius r . Plastic analysis assumes that the plastic moment of resistance should be reached at the ends before the plastic shear resistance is reached at the mid-length of half a WHP. By equilibrium, the aforementioned condition can be written as:

$$V_{WHP} = \frac{2M_{pl}}{L_{WHP}} < V_{pl} \quad (9)$$

where L_{WHP} is the clear length of the bending parts of half a WHP. From Eqs. 7 to 9, any combination of f_y , D_e , D_i , and L_{WHP} can be chosen to provide the desired V_{WHP} . The yield force of a WHP, $F_{y,WHP}$, is then calculated as:

$$F_{y,WHP} = 2V_{WHP} \quad (10)$$

The virtual work method along with analytical integration was applied to derive the elastic stiffness K_{fe} of a WHP:

$$K_{fe} = 2 \frac{9\pi D_e^3 D_i E G}{(40 E D_e^2 L_{WHP} + 48 G L_{WHP}^3)} \quad (11)$$

where E is the modulus of elasticity and G the shear modulus of the material. The WHP post-yield stiffness K_{fp} is assumed to be 2 % the initial stiffness, i.e., $K_{fp} = 0.02K_{fe}$.

Design of Prototype SC PT Connection with WHPs

Prototype Building

Figure 6a shows the plan view of the prototype building used for this study. The study focuses on the interior steel MRF shown in Fig. 6b. The yield stress of structural steel is equal to 300 MPa. The design seismic action, referred to herein as DBE, has a return period equal to 475 years and is expressed by the elastic design spectrum of Eurocode 8 (2004) with peak ground acceleration equal to 0.3 g and soil B. The interior MRF is first designed as a conventional MRF according to EC3 (2003) and EC8. The behavior (or strength reduction) factor q is equal to 6.5. Eurocode 8 imposes a serviceability limit on the peak story drift, θ_{max} , under the frequent earthquake (FOE) with a return period equal to 95 years. The FOE has intensity equal to 40 % the intensity of the DBE, and the limit on θ_{max} is 0.5 % and

0.75 % for non-ductile and ductile nonstructural components, respectively. A strength-based design under the DBE was first performed. Beams and columns from this design had to be increased to satisfy the limit on $\theta_{\max}$ under the FOE. The final sections are shown in Fig. 6b and were found iteratively, i.e., by decreasing the value of q , designing the MRF under the DBE, and then checking drifts under the FOE. The conventional MRF has story drifts equal to 0.64 % under the FOE, 1.6 % under the DBE, and 2.4 % under the maximum considered earthquake (MCE). The MCE has intensity equal to 150 % the intensity of the DBE.

The sections of the conventional MRF are used for designing the proposed PT connection (see next section). In that way, a SC-MRF using the proposed PT connections would have the same initial stiffness and period of vibration with the conventional MRF.

Performance-Based Design of PT Connection with WHPs

The design focuses on the exterior connection of the 2nd floor of the prototype steel MRF shown in Fig. 6b. This connection is designed as a PT connection with WHPs using the methodology presented by Garlock et al. (2007) that adopts two performance objectives and associated structural limit states, namely, (1) immediate occupancy under the DBE by avoiding damage in beams and columns while permitting gap opening and (2) collapse prevention under the MCE by avoiding PT bar yielding and beam local buckling while permitting minor yielding in beams and columns.

The design methodology proposed by Garlock et al. (2007) uses beam flange-reinforcing plates to avoid early beam flange yielding and buckling. However, the connection can experience sudden loss of strength and stiffness due to beam web local buckling when deformed beyond the MCE drift. In this work, the use of the structural details (drilled holes on the beam flanges and beam web stiffeners) proposed by Kim and Christopoulos (2009) in a connection designed according to the procedure described by Garlock is also examined with the aim to further delay beam local buckling

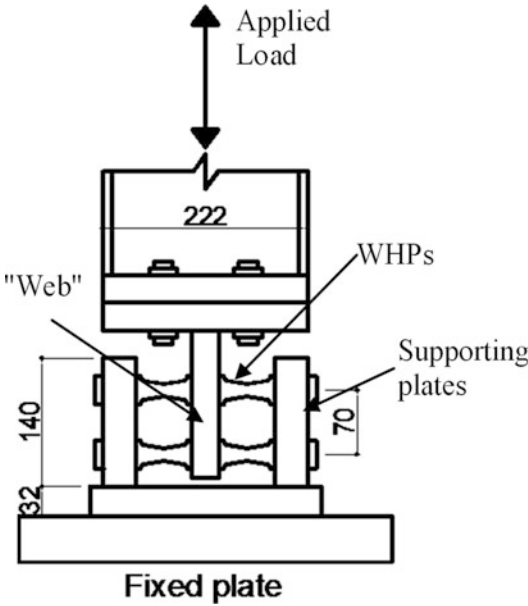
for drifts higher than the MCE drifts. Such a design approach is slightly different than the one proposed by Kim and Christopoulos (2008b) where the connection is designed to form a ductile plastic hinge for drifts approximately equal to the DBE drifts.

The design of the PT connection is based on drifts equal to 0.64 % under the FOE, 1.6 % under the DBE, and 2.4 % under the MCE according to the prototype building design. These drifts are significantly lower than those used in previous works which are based on the IBC 2 % drift limit under the DBE (Rojas et al. 2004; Garlock et al. 2007). To initiate the design procedure, the moment corresponding to point 2 in Fig. 5a, referred to herein as M_{IGO} , is set equal to 66 % of the nominal plastic moment of resistance of the beam section. In addition, M_d is set equal to 69 % of M_{IGO} . These M_{IGO} and M_d values provide an energy dissipation factor, β (Garlock et al. 2007), equal to 0.31. The posttensioning force and the length of the reinforcing plates are then designed. PT bars are designed to avoid yielding for $\theta \leq 0.07$ rad.

Hysteretic Characterization of WHPs

Component tests on the WHPs were conducted to assess their energy dissipation capacity and ductility under cyclic loading. Figure 7 shows the experimental setup for the tests. A pair of WHPs was tested to reproduce their behavior in the proposed PT connection. Two supporting plates were welded on a fixed plate. A plate simulating the web of the beam was bolted to a thick plate attached to the actuator. The thicknesses of the plates were equal to those of the plates designed for the large-scale PT connection tests discussed later.

Four coupon tests were conducted to identify the material properties of the WHPs. The results of the coupon tests are summarized in Table 1. The WHPs were made of 1020-Grade carbon steel, with an average yield strength 557 MPa, ultimate strength 598 MPa, and elongation at fracture equal to 20 %. This type of steel was chosen to achieve the required strength while keeping the sizes of the WHPs relatively small.



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 7 Component test setup for WHPs (dimensions in mm)

The hourglass shape of the WHPs was designed to be consistent with their bending moment diagram shown in Fig. 2. Such a design results in almost uniform distribution of inelastic deformations along the length of the WHPs and, hence, in enhanced energy dissipation and fracture capacity. For the tested configuration, the design resulted in WHPs with external diameter $D_e = 20$ mm, internal diameter $D_i = 14$ mm, and $L_{whp,i} = 40$ mm, while the external part had a diameter equal to 30 mm.

Figure 8a shows the deformed WHPs during the component test. WHPs were capable of sustaining repeated large inelastic cycles without fracture up to displacements associated with connection rotation of 0.07 rad. Figure 8b plots the force–displacement hysteresis of the WHPs. WHPs provide a yielding force equal to 150 kN which is in excellent agreement with the 147 kN WHP yielding force calculated by the plastic analysis procedure presented previously. The ultimate strength of the WHPs at the initiation of fracture is 160 kN. Slight pinching (short flat region) is observed at the points where the force

Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Table 1 Material properties of WHPs from coupon tests

Coupon no.	Yield stress (MPa)	Tensile stress (MPa)	E (GPa)	% Elongation at fracture
1	551	595	207	17
2	550	594	209	20
3	564	599	210	23
4	563	604	208	20
Average	557	598	208	20

changes sign due to slight ovalization of the holes of the supporting plates under the bearing forces induced by the WHPs.

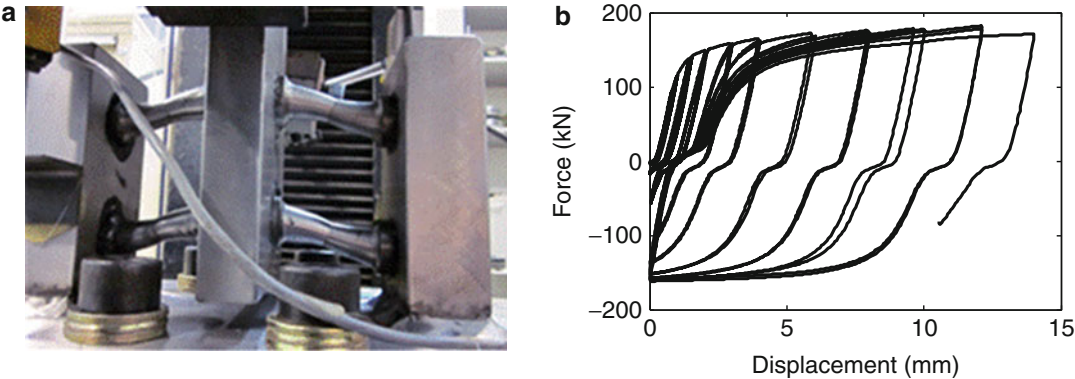
Large-Scale Experimental Program

Test Setup

Experiments on PT connections equipped with WHPs were conducted in the test setup shown in Fig. 9. The test specimens were based on the prototype second-floor exterior beam-to-column connection (see Fig. 6) at 0.6 scale. A strong 310UC158 column was used to minimize the column deformations since they do not influence the behavior of the main connection components (beam, PT bars, and WHPs). Two additional steel members were welded to the column to form a truss system which increases the horizontal stiffness of the test setup (310UC158 horizontal member and 200UC52 diagonal member in Fig. 9). The whole system was bolted on the strong floor. The displacement history was applied vertically by a hydraulic actuator positioned at a distance of 1,800 mm from the inner face of the column.

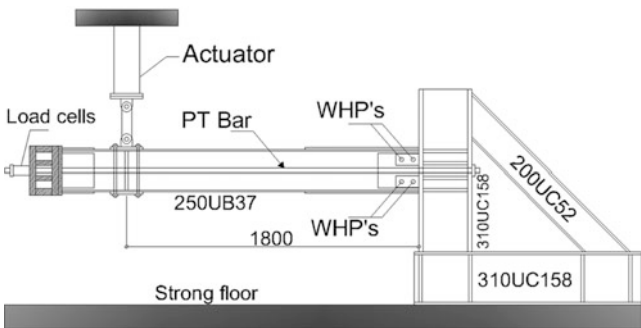
Test Specimens

Two connection specimens (SC-WHP1 and SC-WHP2) were designed according to the procedure described earlier. The resulting beam and column sizes and details of the connection specimens are summarized in Table 2. The specimens are identical except for the beam web stiffeners and beam flange drilled holes used in specimen SC-WHP2.



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 8 WHPs test results: (a) deformed WHPs and (b) WHPs hysteresis

Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 9 PT connection test setup (dimensions in mm)



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Table 2 Specimen details

	SC-WHP1	SC-WHP2
Beam section	250UB37	250UB37
Column section	310UC158	310UC158
$F_{PT,i}$ (kN) ^a	518	504
L_{rp} (mm) ^b	700	700
Web stiffeners length (mm)	n/a	540
Drilled holes diameter (mm)	n/a	27

^a $F_{PT,i}$: Initial posttensioning force
^b L_{rp} : Length of reinforcing plates on the beam flanges

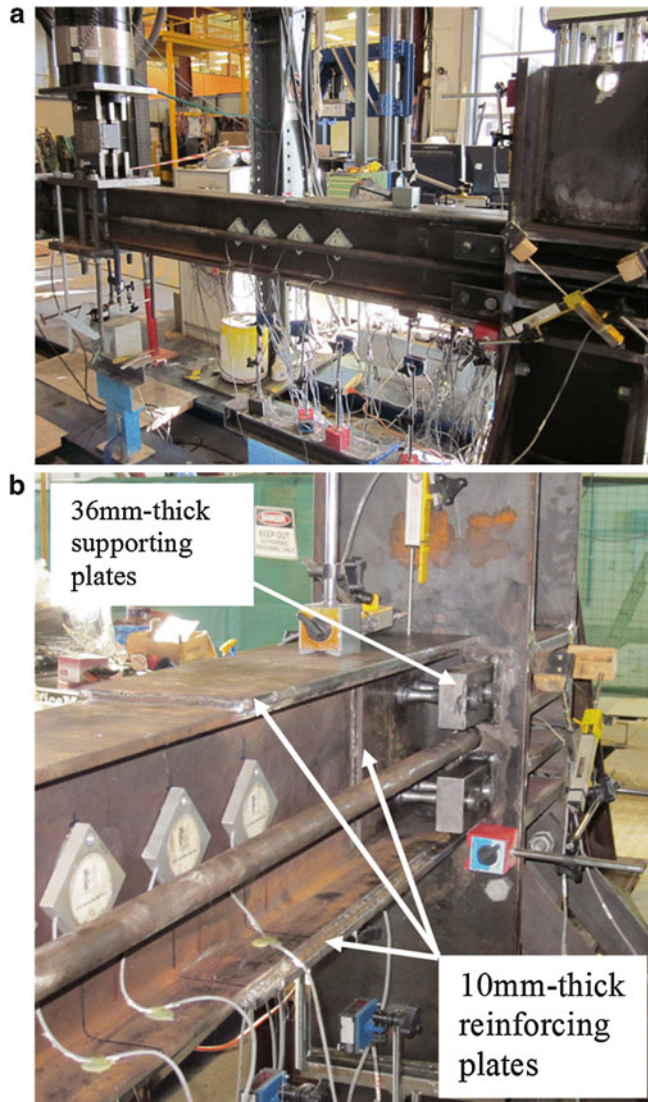
Figure 10a shows the SC-WHP1 specimen, while Fig. 10b shows a close-up view of its connection region. Beam flanges are reinforced with 700 mm long and 10 mm thick plates, while the web of the beam is reinforced with 10 mm

thick plates as shown in Fig. 10b. The thickness of the WHP-supporting plates is 36 mm, their width is 80 mm, and a weld radius of 6 mm was used for the welding of all the plates in the connection.

Figure 11 shows the SC-WHP2 specimen which is identical to the SC-WHP1 apart from four 540 mm long longitudinal stiffeners which are welded on the web and four 27 mm holes drilled on the flanges immediately after the end of the reinforcing plate. The longitudinal stiffeners have thickness 10 mm, start 270 mm before and are extended 270 mm beyond the end of the reinforcing plates. The longitudinal stiffeners aim to resist the web buckling in the beam which typically occurs after the end of the reinforcing plates due to the increased bending moment and compressive force developed in this section as the gap opening increases. The drilled holes aim to reduce the moment

Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures,

Fig. 10 SC-WHP1 specimen: (a) overview and (b) close-up view of the connection region

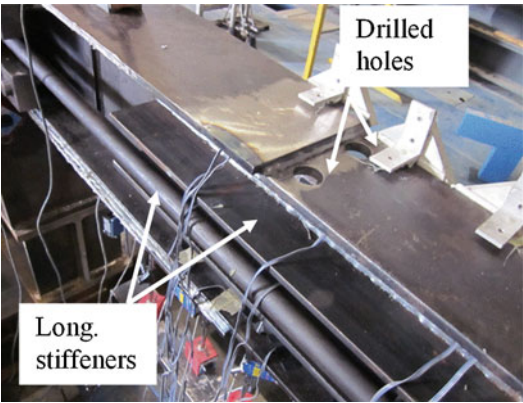


capacity of the beam in this region in order to allow for the formation of a plastic hinge under large drifts. A dog-bone detailing could be used instead.

Material Properties

The beam and column sections were made of steel with nominal yield strength equal to 300 MPa. The reinforcing plates, WHPs supporting plates, and stiffeners were made of steel with nominal yield strength equal to 350 MPa. Tensile tests were carried out on

coupons cut from the flanges and webs of the beams. Table 3 summarizes the mean actual yield stress, tensile stress, and modulus of elasticity resulting from the material tests. The coupon tests showed that the mean yield stresses are 361 and 388 MPa for the flanges and the web of the beam, respectively, indicating considerable overstrength of the steel material. The material of the PT bars has nominal yield strength equal to 930 MPa, tensile stress 1,050 MPa, and elongation capacity 6 %, according to the specifications of the supplier.



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 11 Details of SC-WHP2 specimen

Test Procedure

The loading protocol for the two specimens consists of cyclic vertical displacements of increasing amplitude imposed in a quasi-static fashion with a speed approximately equal to 21 mm/min. The AISC (2005) loading protocol was used. This protocol consists of three initial sets of six cycles at 6.75, 9, and 13.5 mm displacements, four subsequent cycles at 18 mm, and six sets of two cycles at 27, 36, and 54 mm. These displacements correspond to drifts equal to 0.00375, 0.005, 0.0075, 0.01, 0.015, 0.02, and 0.03. The specimens were further imposed to drifts equal to 0.06 and, then, up to 0.09–0.10 to identify the connection failure mode.

Experimental Results

Specimen SC-WHP1

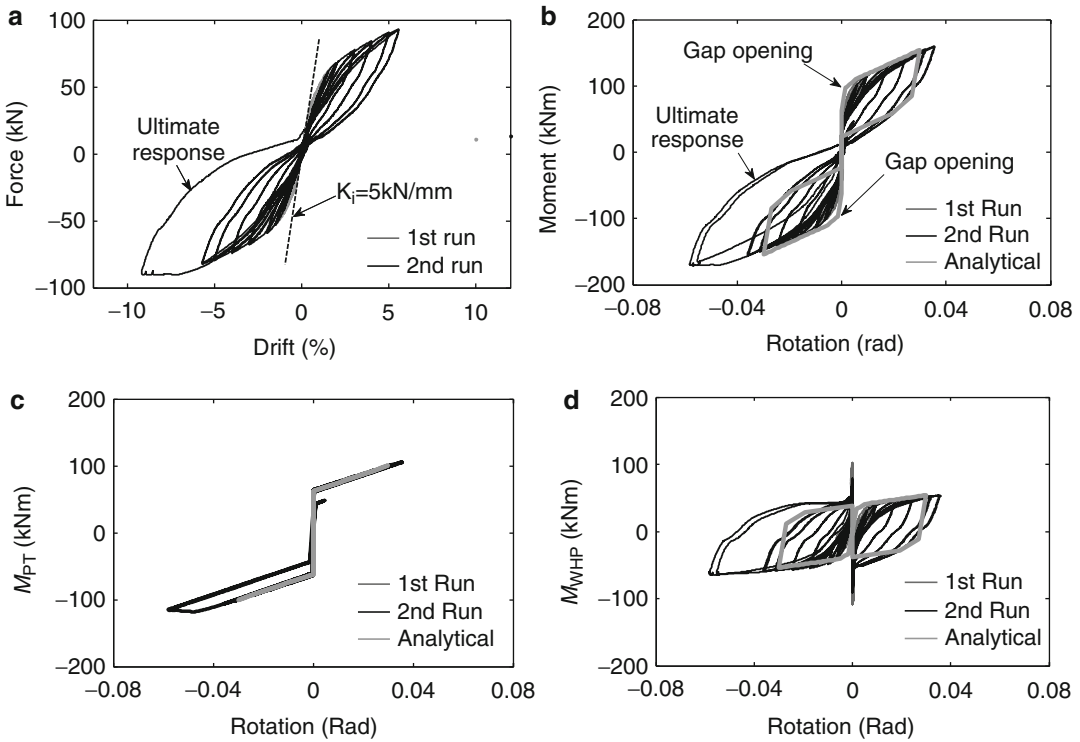
Figure 12a shows the force–drift hysteresis of the SC-WHP1 specimen, while Fig. 12b shows the associated moment–rotation hysteresis. In Fig. 12a, the drift is calculated as the ratio of the vertical displacement at the tip of the actuator to the horizontal distance of the tip of the actuator to the inner face of the column. The rotation is calculated as the gap-opening angle at the beam–column interface measured during the tests. In Fig. 12b, the moment is calculated as the product of the actuator force and the distance

Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Table 3 Material test results

Coupon	Sample no.	Yield stress (MPa)	Tensile stress (MPa)	Modulus of elasticity (GPa)
Flange	1	354	535	208
	2	370	536	207
	3	359	530	208
	Average	361	534	208
Web	1	385	547	205
	2	392	548	199
	3	387	545	176
	Average	388	547	193

of the actuator tip to the inner face of the column. Figure 12a shows that the connection maintains self-centering capability for drifts up to 6 %. The initial stiffness of the connection is 5 kN/mm and is equal to the theoretical stiffness of a welded (fully restrained) connection. During the last cycle of 9 % drift, the connection lost its self-centering capability and experienced a residual drift equal to 3 %.

To assess the reparability of the proposed connection after a strong earthquake, the SC-WHP1 specimen was initially imposed to displacements up to the DBE level (i.e., 1.6 % drift). The test was then interrupted, and damaged WHPs were substituted by new ones. The substitution was easily accomplished according to the following steps: (a) a temporary support was placed below the beam close to the beam–column interface to accommodate the self-weight of the beam; (b) the PT force was relaxed; (c) the damaged WHPs were taken out with the aid of a wooden hammer; and (d) new WHPs were placed in the connection. Since no welding or bolting is needed for the WHPs replacement, the proposed connection can be repaired with minimal disturbance to building use or occupation in the aftermath of a major earthquake. After the substitution of the WHPs, the test was restarted. Figure 12a plots superimposed loops from the tests before and after the replacement of WHPs. During the first cycles, the connection loses part of its initial



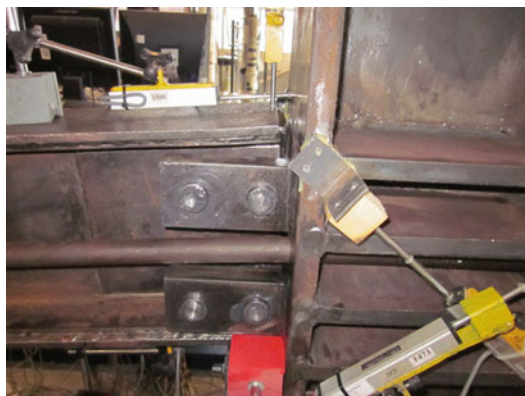
Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures,
Fig. 12 Hysteretic response of SC-WHP1 specimen and comparison with analytical predictions

stiffness as a result of the unavoidable local yielding of the beam edges at the beam–column interface. Nevertheless, Fig. 12a shows that the connection gains its stiffness after drifts of 1 %. In fact, the two loops are perfectly superimposed to each other, demonstrating that the proposed connection can be repaired after a strong earthquake without compromising its stiffness, strength, and energy dissipation capacity.

Figure 12b plots superimposed moment–rotation loops from the tests before and after replacement of WHPs along with the analytical moment–rotation prediction using the analytical expressions of the simplified procedure described earlier. The moment contribution of the PT bars, M_{PT} , is calculated from the posttension force recorded by the load cells and is plotted in Fig. 12c, while the contribution of the WHPs, M_{WHP} , is found by subtracting M_{PT} from the total moment of the connection and is plotted in Fig. 12d.

Figure 12b shows that the connection achieves rotations larger than 0.035 rad with full self-centering capability. The experimental moments at decompression and gap opening are $M_d = 62$ kNm and $M_{IGO} = 95$ kNm, respectively. These values are slightly smaller than the design values of $M_d = 66$ kNm and $M_{IGO} = 100$ kNm. However, the predicted analytical curve captures well the cyclic behavior in terms of both stiffness and strength. The analytical prediction is also in good agreement with both the M_{PT} – θ and the M_{WHP} – θ loops, as shown in Fig. 12c, d, respectively.

Displacement measurements at the beam–column interface indicated a vertical translation of the beam axis with respect to its initial position of about 4 mm. This translation took place gradually during the testing procedure and is attributed to a gradual loss of friction between the beam and the column interface. Shear tabs could be welded on the column flange and bolted on the beam web using slotted holes to

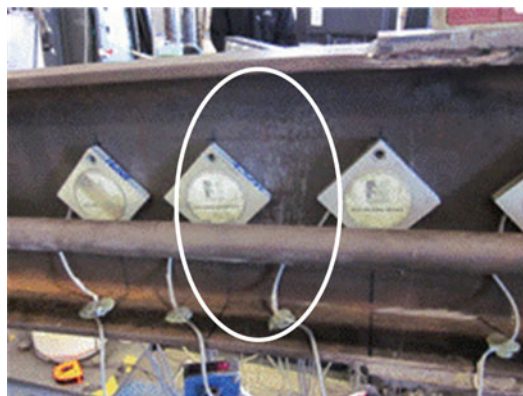


Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 13 Gap opening at 6 % imposed drift of the SC-WHP1 specimen

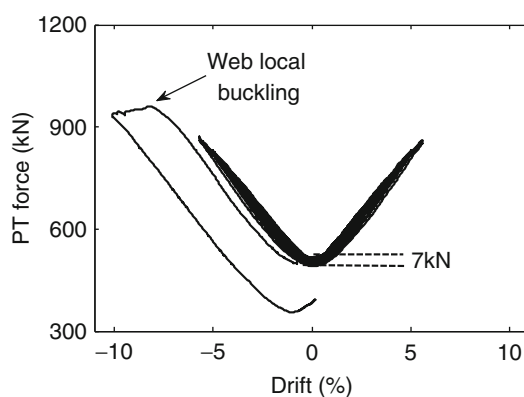
accommodate gravity loads and eliminate the vertical beam displacement without influencing the connection moment behavior. Another alternative to accommodate gravity loads could be the use of a seat angle positioned below the bottom flange of the beam.

Figure 13 shows a close-up view of the gap-opening angle corresponding to 6 % drift. Apart from the intended damage in the WHPs and local yielding in the beam edges at the beam–column interface, no other evidence of damage was observed for drifts lower or equal to 6 %. Yielding in the beam flanges at the beam–column bearing surface was evident from the first cycles of the test. In fact, it was visually verified that the extreme edges of the beam at the beam–column interface became rounded during the test. As a result, the location of the COR was not constant during the experimental process. At the initial cycles, the COR was located approximately at the center of the reinforcing plate, while at larger drifts, it was stabilized at the inner side of the reinforcing plates.

During the last cycle of 9 % drift, the specimen failed due to local buckling of the web of the beam immediately after the end of the reinforcing plates. Figure 14 shows a close-up view of the buckled region. Web buckling took place at 7 % drift, while local buckling of the beam flanges initiated at the same time. Strain gauge measurements did not indicate



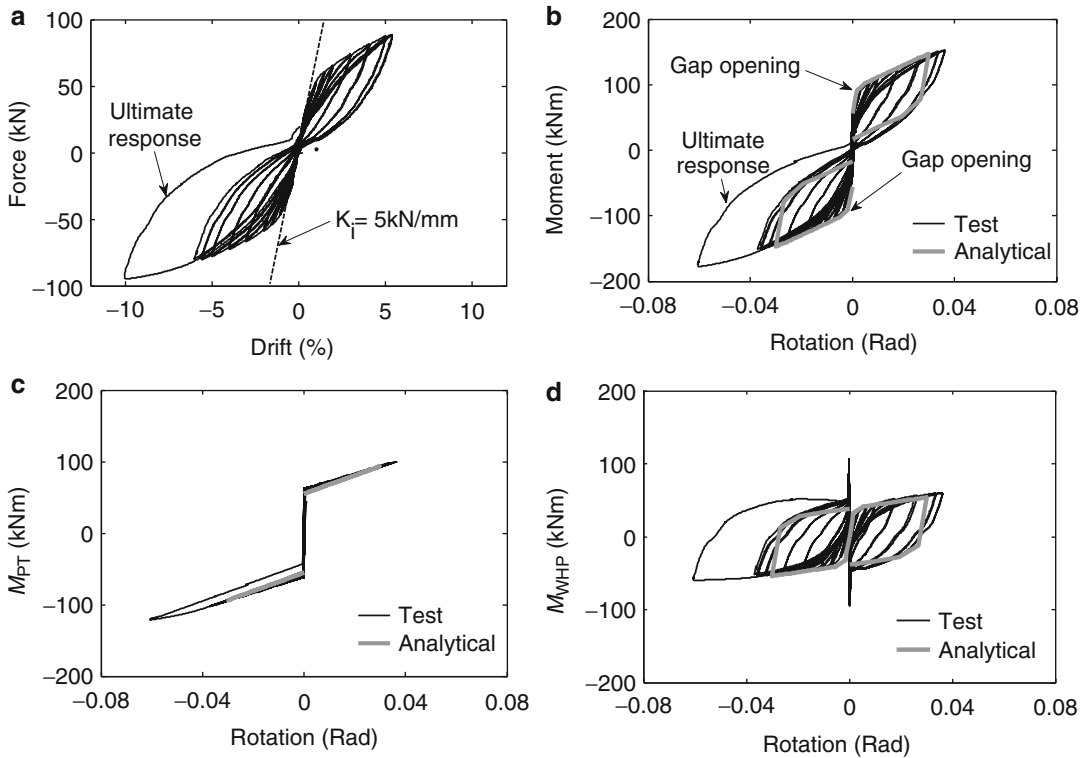
Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 14 Web buckling of the SC-WHP1 specimen



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 15 PT force versus drift for the SC-WHP1 specimen

yielding at any point throughout the beam section when local buckling initiated.

Figure 15 plots the variation of the force in the PT bars with respect to drift. The PT force increases with gap opening and returns to its initial value of 504 kN when gap closes. The loss in posttensioning during the test was about 7 kN. The maximum force in the PT bars has values of 850 kN and 970 kN at 6 % and 9 % drift, respectively. These values show that the PT bars remained elastic during the test since their nominal yield limit is 1,227 kN. A drop in the PT force due to beam local buckling is observed during the last cycle.



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures,
Fig. 16 Hysteretic response of specimen SC-WHP2 and comparison with analytical predictions

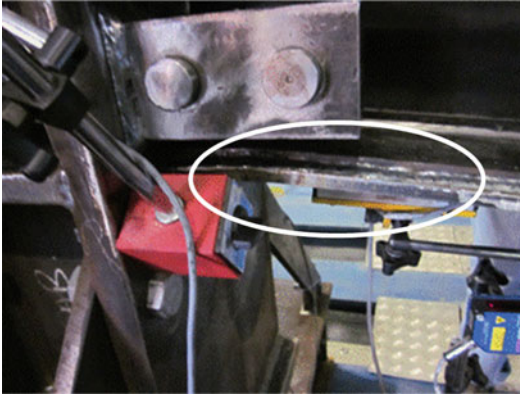
Specimen SC-WHP2

Figure 16a shows the cyclic force–drift behavior of the SC-WHP2 specimen. The behavior is stable with full self-centering capacity up to 6 % drift similar to specimen SC-WHP1. A final experimental loop of 10 % drift caused the connection to fail, and a residual drift of 4 % was present at the end of the test.

The moment versus relative rotation hysteresis is plotted in Fig. 16b. The experimental M_d value was 60 kNm, while M_{IGO} was 92 kNm. These values are somewhat smaller than the corresponding values of the SC-WHP1 specimen due to the slightly smaller initial posttensioning force (504 instead of 518 kN). The simplified design procedure predicts well the cyclic envelope with the predicted values for M_d and M_{IGO} being 64 and 98 kNm, respectively. Figure 16c–d shows that the analytical predictions are in good

agreement with the individual experimental moment contributions M_{PT} and M_{WHP} .

The detailing employed for specimen SC-WHP2 resulted in a different failure mode. No web or flange buckling was observed in the beam section after the end of the beam flange-reinforcing plates. In addition, no plastic hinge was developed in this region. The strain gauges indicated that the yield strain was not reached in any point throughout the beam section at the end of the beam flange-reinforcing plate. Figure 17 shows that the SC-WHP2 specimen failed due to excessive yielding in the beam flanges at the beam–column interface under large bearing forces. Local yielding at the bearing interface of specimen SC-WHP2 was more pronounced than in specimen SC-WHP1 and resulted in a gradual reduction in posttensioning force. Figure 18 shows that the loss of force in the PT bars was



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 17 Local yielding in the beam flange at the bearing interface of the SC-WHP2 specimen

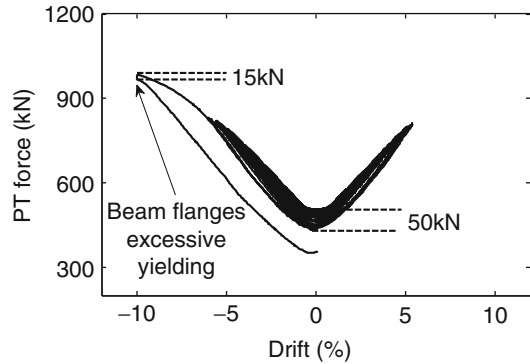
50 kN during the test, with the largest amount attributed to loading cycles associated with drifts larger than the MCE drift. A reduction of 15 kN in the force of the PT bars was observed as a result of the local failure at the final loading cycle.

Finite Element Simulations and Parametric Study

This section presents FEM models which can be used to reliably assess the design and behavior of the connection. A detailed nonlinear FEM model was developed in Abaqus (Dassault Systems 2010). The FEM model was calibrated against experimental results and found capable to trace the nonlinear cyclic behavior of the connection and capture all possible local failure modes. The calibrated FEM model was used to conduct a series of simulations to study the effect of different parameters on the connection behavior.

Models for the WHPs

The cyclic behavior of the WHPs was first simulated. The FEM model geometry reproduced the actual geometry of the test setup of the component tests that were carried out to characterize the behavior of the WHPs. Figure 19 shows the setup used for the WHPs characterization tests along

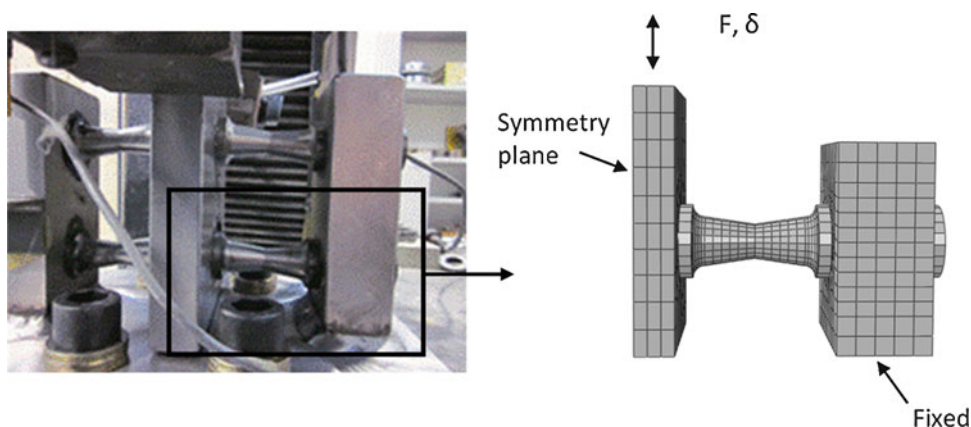


Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 18 PT force versus drift for the SC-WHP2 specimen

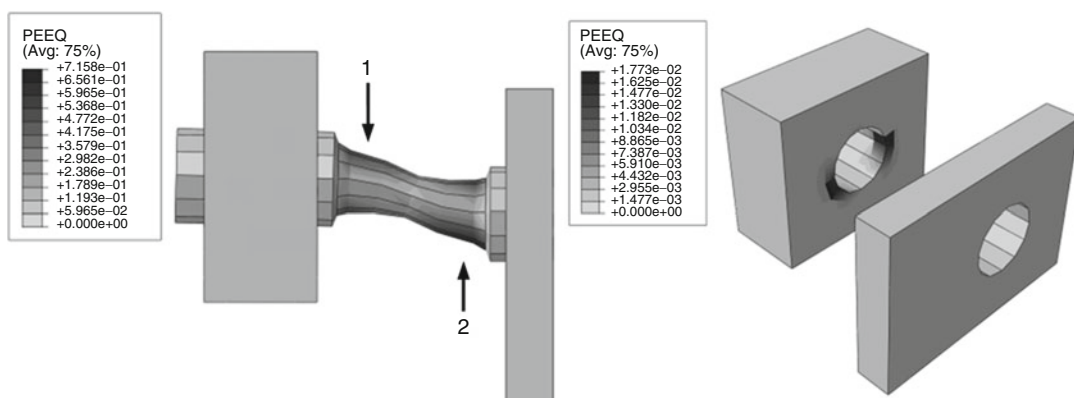
with the FEM model and its boundary conditions. Since the actual geometry was symmetric, only one fourth of the specimens was modeled to decrease computational time.

Figure 20a shows the deformed shape of the WHP and the equivalent plastic strain (*PEEQ*) distribution at a displacement of 12 mm. Figure 20a shows that the hourglass shape geometry results in uniform distribution of plastic deformation along the length of the internal part of the WHP, while the external parts of the WHP and the supporting plates are essentially elastic. The *PEEQ* distribution in the WHP's supporting plates is plotted separately in Fig. 20b which shows minor plastic concentrations. These results agree with the experimental test observations where a negligible ovalization of the holes of the supporting plates was evidenced.

Figure 21 compares the force–displacement hysteresis from FEM analysis with the experimental hysteresis from the WHPs characterization tests. The FEM model can trace the nonlinear cyclic behavior of the WHPs with good accuracy. The model captures the pinching behavior at zero force which is observed in the experimental hysteretic curve as the result of the negligible ovalization of the holes of the supporting plates. Figure 21 shows that the experimental hysteresis of the WHPs deteriorates after a displacement amplitude of 12 mm. The WHP failed due to



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures,
Fig. 19 The FEM model used for the WHPs



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures,
Fig. 20 Deformed shape of the WHP and equivalent plastic strain (*PEEQ*) contour plot

ductile fracture at the section close to the internal support plate under a displacement amplitude of 12 mm.

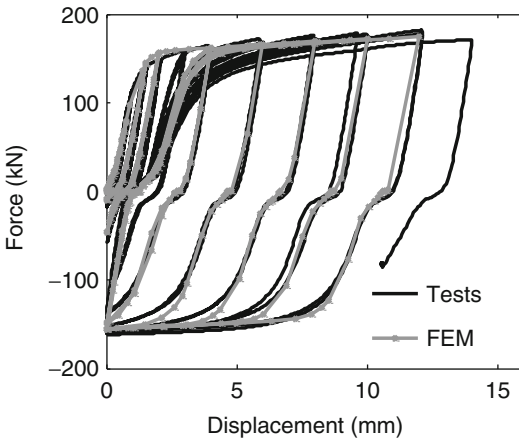
Models for the Connection

A three-dimensional FEM model was developed to simulate the behavior of the connections as shown in Fig. 22. The geometry of the tests was reproduced in full detail. The column, the PT bars, the WHPs, and the plates were modeled using C3D8R solid elements. The beam was modeled using solid elements with incompatible modes (C3D8I). C3D8I are first-order elements that are enhanced by incompatible modes to

improve their bending behavior. The mesh was refined in regions where severe plastic deformations or buckling phenomena were expected to occur, i.e., close to the beam–column interface and at the end of the beam flange-reinforcing plates. A coarser mesh was used for regions that were expected to remain essentially elastic, i.e., the anchor block, the tip of the beam, and the top and bottom parts of the column. The WHPs and their supporting plates were modeled using the verified model for the WHPs described in the previous section.

The vertical load was applied as an imposed displacement, U_2 , at a distance equal to 1,800 mm

from the connection face as shown in Fig. 22a. The analysis consisted of several steps. In the first step, the contact interactions were established to ensure that numerical problems due to contact formulation will not be encountered during the next steps. The posttensioning force was applied during the second analysis step by imposing an axial displacement at the free ends of the PT bars



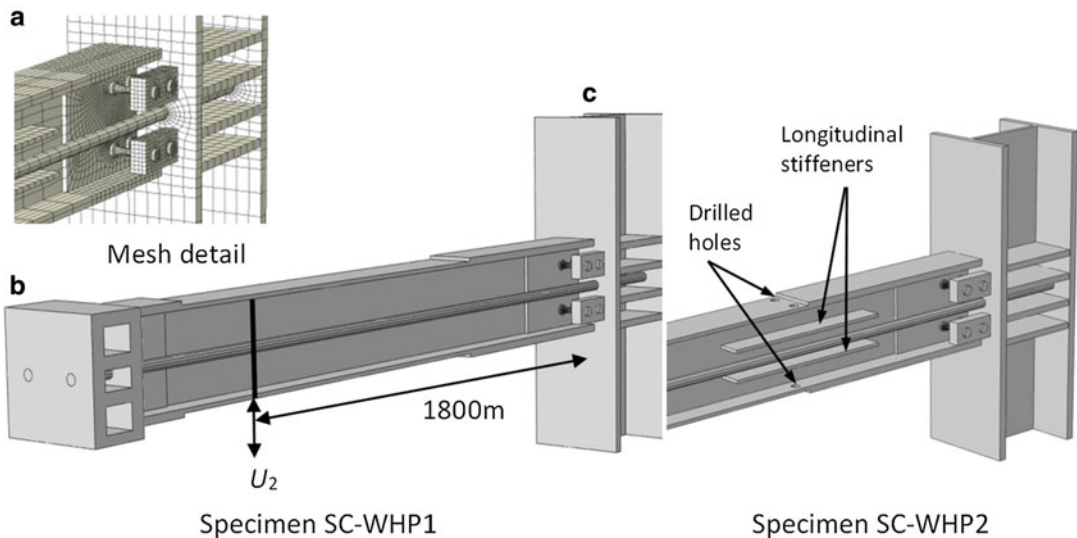
Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 21 Comparison of the experimental and the numerical force–displacement hysteretic behavior of the WHPs

capable of producing the desired $F_{PT,i}$ in the beam. In the subsequent steps, the cyclically displacement history was applied. Displacement-controlled nonlinear analysis was performed.

Assessment of the Finite Element Model

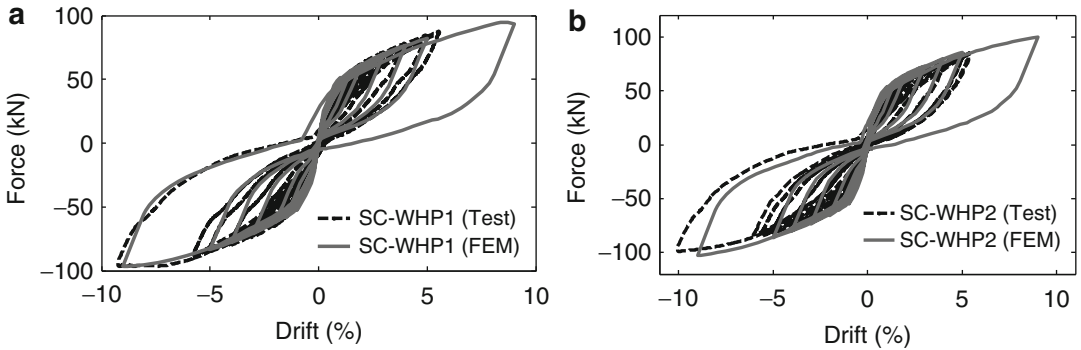
Figure 23 plots the force–drift hysteresis from FEM analyses along with the experimental hysteresis of the two connection specimens. The FEM model can capture well the overall cyclic behavior of the connections. The predicted values for the M_d and M_{IGO} are in very good agreement with the corresponding experimental ones, while the initial and post-elastic stiffness are almost identical. The variation of F_{PT} from experiments is compared with the FEM prediction in Fig. 24. The FEM simulations and experiments are in reasonably good agreement. The numerical model predicts a slightly smaller loss in F_{PT} than that measured one in the experiments for specimen SC-WHP1, while the loss in F_{PT} is more accurately predicted for specimen SC-WHP2.

Following the loading protocol of the experimental program, the FEM models were pushed to large drifts to investigate all possible global and local failure modes. Figure 23 shows that the two



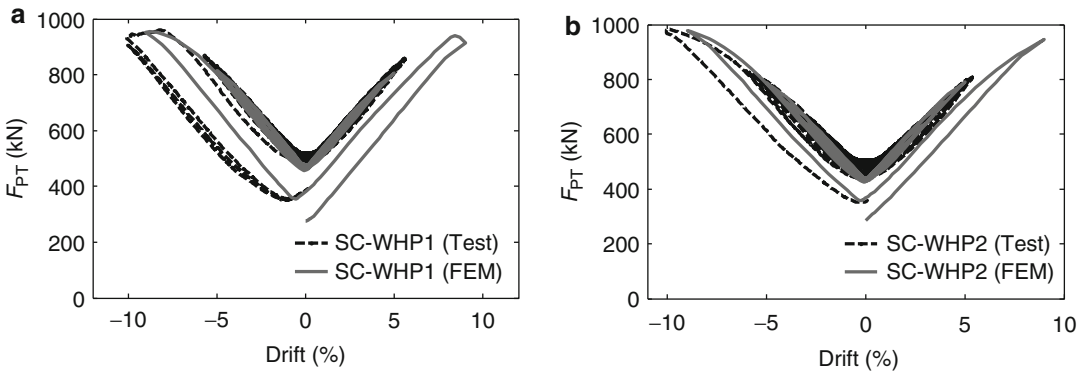
Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 22 FEM model details: (a) FEM

discretization, (b) model for specimen SC-WHP1, and (c) model for specimen SC-WHP2



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 23 Force–drift hysteresis of the two

connection specimens tested in Kim and Christopoulos (2008b) and comparison with FEM results: (a) SC-WHP1 and (b) SC-WHP2



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 24 F_{PT} -drift plots for the two connection

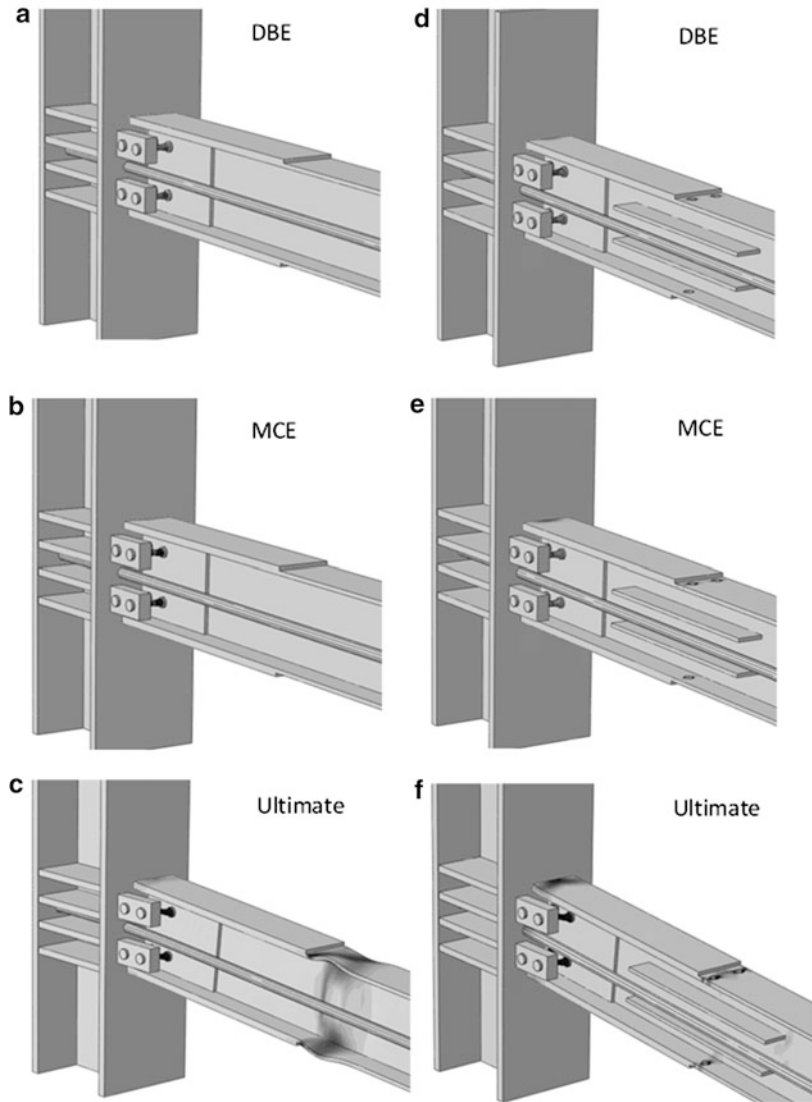
specimens tested in Kim and Christopoulos (2008b) and comparison with FEM results: (a) SC-WHP1 and (b) SC-WHP2

large last cycles of 9 % drift for specimen SC-WHP1 and 10 % drift for specimen SC-WHP2 resulted in residual inelastic drifts and loss of self-centering capability. These large drifts resulted in different failure modes of the two connection specimens, consistent with the experimental observations. The FEM model for specimen SC-WHP1 failed due to web buckling immediately after the end of the reinforcing plates. The FEM model for specimen SC-WHP2 failed due to yielding in the region of the drilled holes on the beam flange and in the bearing interface, while the web of the beam remained elastic until the end of the analysis. This can be verified by the *PEEQ* value distribution in the deformed configurations of the connections (Fig. 25) for

three drift levels: DBE (i.e., Design Basis Earthquake (Charney and Downs 2004)) = 1.6 %, MCE = 2.4 %, and Ultimate = 9–10 %. In this figure, darker areas indicate larger plastic strain concentration. The FEM models captured the failure modes observed in the experimental program. Figure 25 shows that under the DBE drift, the connection damage is isolated in the WHPs. Under the MCE drift, damage is concentrated in the WHPs and in the beam flanges in the beam–column interface. Damage in the beam–column interface is more evident in the connection SC-WHP2. Under the ultimate drift, damage in the connection is spread along the beam web immediately after the end of the reinforcing plates and in the beam flanges at the

Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures,

Fig. 25 Contour plots of the equivalent plastic strain ($PEEQ$) in the connection at different drift levels: SC-WHP1 (*left*) and SC-WHP2 (*right*)

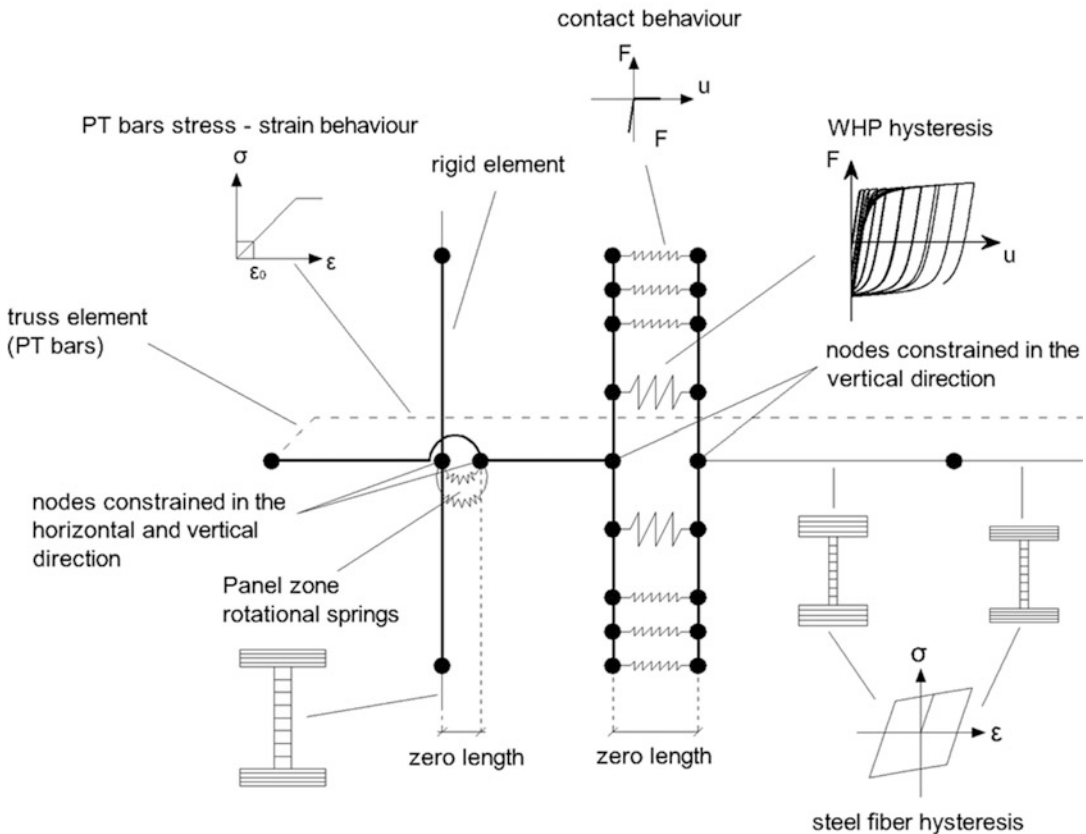


bearing interface. Figure 25f shows that damage in the beam flanges is more evident in the SC-WHP2 model.

The comparisons between the FEM analyses and experiments show that the proposed FEM model is capable of reproducing the inelastic response of the tested PT connections up to the ultimate deformation levels and to capture accurately all possible failure modes. Therefore, it is a reliable tool for the simulation of the hysteretic behavior of PT steel connections (Figs. 23 and 24).

Seismic Analysis of SC-MRFs Using PT Connections with WHPs**Modeling of PT Connection with WHPs for Nonlinear Analysis in OpenSees**

A model for the PT connection with WHPs and the associated beams and columns was developed in OpenSees (Mazzoni et al. 2006) as shown in Fig. 26. The beams and columns were modeled using the nonlinear force-based beam-column fiber element which can strictly satisfy equilibrium and accurately capture the distribution of

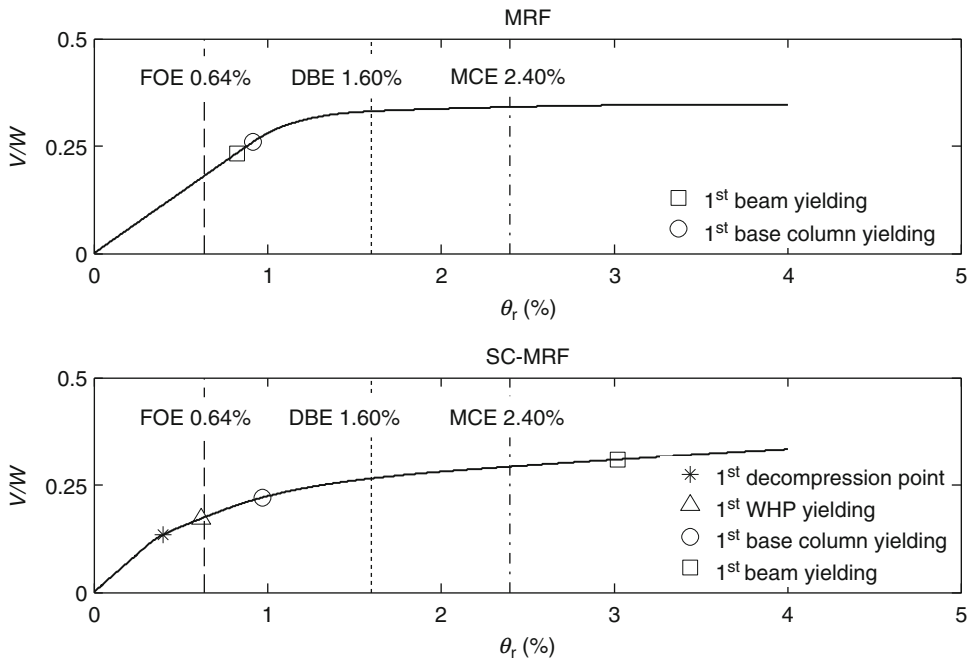


Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures,
Fig. 26 Model for an exterior PT connection with WHPs and associated columns and beam

inelasticity along the depth of the section and along the length of the physical member. For the beam, two fiber elements with cross-sections having different flange thickness were used to account for the beam flange-reinforcing plates. Each fiber was associated with uniaxial bilinear elastoplastic stress-strain behavior (Steel01 in OpenSees) with a post-yield stiffness ratio equal to 0.01.

Rigid elastic beam-column elements were used to model the beam-column interface where gap opening and closing take place. To accurately capture the gap-opening mechanism in the beam-column interface, three zero-length contact spring elements were placed at equal spaces along the beam flange thickness. These contact springs were associated with an elastic compression – no tension force-displacement behavior (ENT material in OpenSees). A value

of the compression stiffness equal to 20 times the axial stiffness of the beam K_b was assigned to these contact springs. Larger values for this stiffness were found to produce practically the same results but with higher computational cost, i.e., more iterations to achieve convergence and equilibrium in nonlinear analysis. To capture the hysteretic energy dissipation capacity of the connection, two zero-length hysteretic springs were placed at the exact locations of WHPs along the depth of the beam web. These springs were associated with a smooth hysteretic Giuffre-Menegotto-Pinto model with isotropic hardening (Steel 02 material in OpenSees). To account for panel zone shear deformations and possible yielding, the panel zone was modeled using the Scissors model which introduces four additional rigid elastic beam-column elements and two nodes in the center of the panel zone



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures,
Fig. 27 Base shear coefficient – roof drift behavior from nonlinear monotonic (pushover) static analysis

connected with two zero-length rotational springs. These springs are associated with bilinear elastoplastic hysteretic rules (Steel01 material in OpenSees) with properties calculated to reflect the contribution of the column web (including doubler plates) and the column flanges in the force–shear deformation panel zone behavior. This simple panel zone model has been found to produce identical results to those of the more computationally expensive Krawinkler panel zone model (Charney and Downs 2004). PT bars were modeled using a truss element running parallel to beam centerline axis and connected to the exterior nodes of the panel zones of the exterior columns of the SC-MRF. The truss element has a cross-section area A_{PT} equal to that of both PT bars. To account for posttensioning, an initial strain equal to $F_{PT,i}/(A_{PT} \cdot E_{PT})$ was first assigned to the truss element where E_{PT} is the modulus of elasticity of the PT bar material. However, posttensioning results in shortening of the beams which in turn decreases the posttensioning force. To account for this decrease, the initial strain in the truss element

was increased to ensure that the posttensioning force in the PT bars will be equal to $F_{PT,i}$ after the beam shortening.

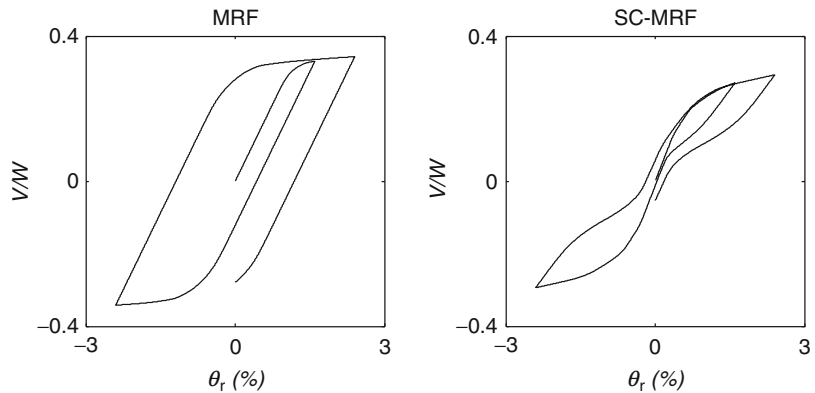
The accuracy of the developed model in OpenSees was assessed by comparing the force–deformation and moment–rotation responses of the connection with the experimental results. Detailed comparisons can be found in (Dimopoulos et al. 2013), where it is shown that the numerical model is in acceptable agreement with the experimental response and can be considered as reliable.

Nonlinear Monotonic and Cyclic Static Analysis

Figure 27 shows the base shear coefficient (V/W) – roof drift (θ_r) behavior of the conventional MRF and the SC-MRF from nonlinear monotonic (pushover) static analysis. V is the base shear and W is the seismic weight. An inverted triangular force distribution along with roof displacement control was used in these analyses. The MRF and the SC-MRF have comparable base shear strengths and comparable

Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures,

Fig. 28 Base shear coefficient – roof drift behavior from nonlinear cyclic (push–pull) static analysis



initial stiffness. The pushover curves are plotted along with points associated with structural limit states and vertical lines corresponding to roof drifts expected under the FOE, DBE, and MCE. The structural limit states for the conventional MRF are beam yielding and base column yielding and occur at θ_r equal to 0.82 % and 0.92 %, respectively. The conventional MRF avoids damage under the FOE but experiences significant damage under the DBE. The structural limit states for the SC-MRF are decompression in a PT connection, WHP yielding, base column yielding, and beam yielding. Figure 27 shows that the beams of the SC-MRF are damage-free for θ_r equal or lower than 3 %, i.e., drifts higher than the MCE. Damage in the SC-MRF is experienced at the column bases that yield at θ_r equal to 0.97 %. No PT bar yielding is observed. The first decompression occurs at θ_r equal to 0.4 % while WHPs yield at θ_r equal to 0.62 % which is almost equal to the FOE drift. Decompression does not involve damage, while yielding of the WHPs is acceptable under low drifts since WHPs can be easily replaced without bolting or welding. The conventional MRF experiences softening at θ_r equal to 1.25 %, while the SC-MRF shows a more gradual softening behavior. In particular, the SC-MRF shows softening due to decompression in the PT connections at low drifts and further softening due to plastic deformations at the column bases and yielding of a large number of WHPs at θ_r equal to 1 %. Figure 28 shows the $V/W - \theta_r$ behavior of the MRF and the SC-MRF from nonlinear cyclic

(push–pull) static analysis. The first cycle of the analysis is performed up to the DBE drift while the next cycle up to the MCE drift. The SC-MRF shows full re-centering capability under the DBE, adequate energy dissipation, and a small residual drift under the MCE due to plastic deformations at the column bases. The conventional MRF shows large energy dissipation capacity due to plastic deformations at the beam ends and at the column bases and the possibility of experiencing large residual drifts under the DBE and MCE.

Nonlinear Dynamic Time History Analyses

Ground Motions

A set of 20 earthquake ground motions recorded on ground type B were used in 2D nonlinear dynamic time history analyses to evaluate the performance of the SC-MRF and the performance of the conventional MRF. None of the ground motions exhibit near-fault forward directivity effects. The ground motions were scaled to the DBE level using the scaling procedure of Somerville (1997). Table 4 provides the scale factors and information on the 20 earthquake ground motions.

Modeling for Nonlinear Dynamic Analysis

Two-dimensional nonlinear analytical models of the conventional MRF and the SC-MRF were developed for nonlinear dynamic analyses in OpenSees. Nonlinear beam–column fiber elements were used for the beams and columns

Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Table 4 Properties of the ground motions used for nonlinear dynamic analyses

Earthquake	Station	Component	Magnitude (M_w)	Distance (km)	Scale factor		
					FOE	DBE	MCE
Imperial Valley 1979	Cerro Prieto	H-CPE237	6.53	15.19	0.82	2.05	3.08
Loma Prieta 1989	Hollister – S & P	HSP000	6.93	27.67	0.29	0.72	1.08
Loma Prieta 1989	Woodside	WDS000	6.93	33.87	1.40	3.49	5.24
Loma Prieta 1989	WAHO	WAH090	6.93	17.47	0.48	1.20	1.80
Manjil 1990	Abbar	ABBAR-T	7.37	12.56	0.28	0.70	1.05
Cape Mendocino 1992	Fortuna – Fortuna Blvd	FOR000	7.01	15.97	0.99	2.47	3.71
Cape Mendocino 1992	Rio Del Overpass – FF	RIO360	7.01	14.33	0.50	1.25	1.88
Landers 1992	Desert – Hot Springs	LD-DSP000	7.30	21.78	0.95	2.37	3.56
Northridge 1994	LA – W 15th St	W15090	6.69	25.60	1.14	2.86	4.29
Northridge 1994	Moorpark – Fire Sta	MRP180	6.69	16.92	0.78	1.94	2.91
Northridge 1994	N Hollywood – Cw	CWC270	6.69	7.89	0.53	1.33	2.00
Northridge 1994	Santa Susana Ground	5108-360	6.69	1.69	0.78	1.95	2.93
Northridge 1994	LA – Brentwood VA	0638-285	6.69	12.92	0.85	2.12	3.18
Northridge 1994	LA – Wadsworth VA	5082-235	6.69	14.55	0.62	1.54	2.31
Kobe 1995	Nishi-Akashi	NIS090	6.90	7.08	0.48	1.19	1.79
Kobe 1995	Abeno	ABN090	6.90	24.85	1.00	2.49	3.74
ChiChi 1999	TCU105	TCU105-E	7.62	17.18	0.96	2.39	3.59
ChiChi 1999	CHY029	CHY029-N	7.62	10.97	0.53	1.32	1.98
ChiChi 1999	CHY029	CHY041-N	7.62	19.83	0.56	1.40	2.10
Hector 1999	Hector	HEC090	7.13	10.35	0.42	1.04	1.56

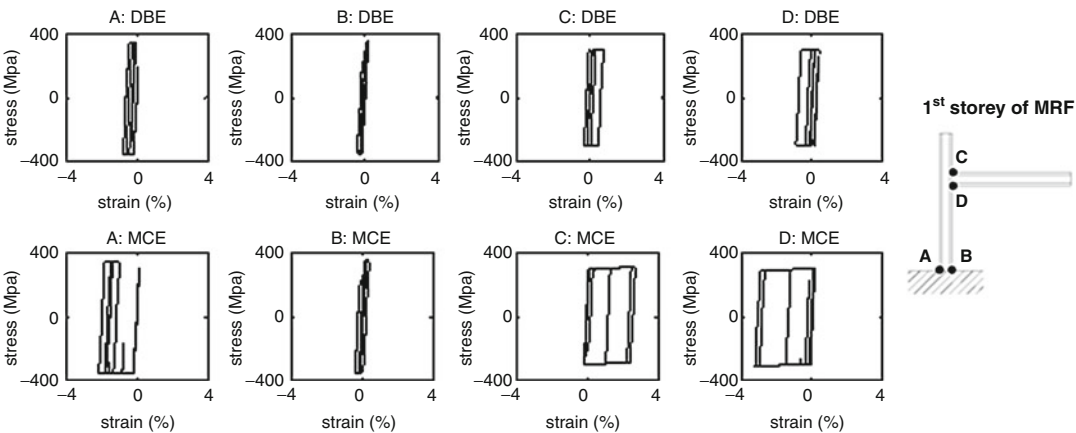
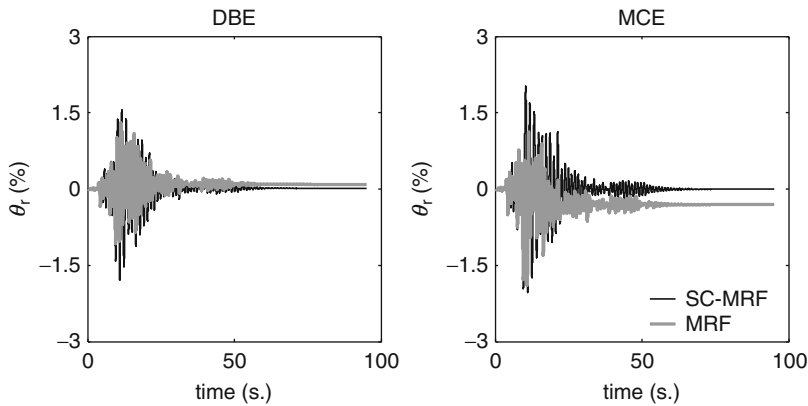
and the Scissors model (Charney and Downs 2004) for the panel zones of the conventional MRF and the SC-MRF. A diaphragm constraint is imposed on each floor level of the conventional MRF. Stiff truss elements were used to connect the internal nodes of each bay of the SC-MRF to allow beam shortening due to posttensioning, gap opening in PT connections, and to consider the use of a discontinuous slab proposed by Chou et al. (2009). The model of Fig. 26 was used to model the PT connections and the associated beams and columns of the SC-MRF. Each dynamic analysis was extended well beyond the actual earthquake time (the ground motions were padded with zeros) to allow for damped free vibration decay and accurate residual drifts calculation.

Seismic Response Results

Figure 29 compares the roof drift time histories of the conventional MRF and the SC-MRF under the 5082-235 ground motion scaled to the DBE and MCE. Near the end of the time histories, the SC-MRF oscillates around the origin, indicating negligible residual drift, while the conventional MRF experiences residual drifts. The peak roof displacements of the MRF and the SC-MRF are similar. Figure 30 shows the stress–strain hysteresis at points A and B (extreme column base flange fibers), and C and D (extreme beam flange fibers) of the first story of the conventional MRF under the 5082-235 ground motion scaled to the DBE and MCE. Figure 31 presents similar information for the SC-MRF. The stress–strain hysteresis immediately after the end of the beam

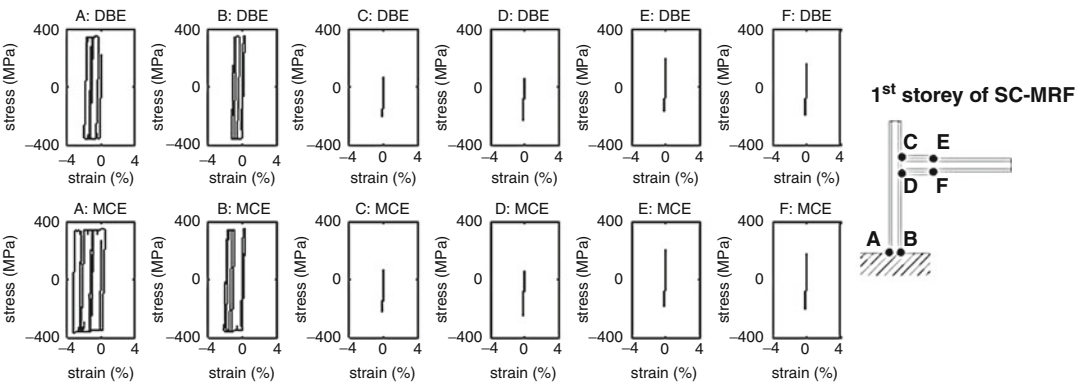
Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures,

Fig. 29 Comparison of the roof drift time histories under the 5082-235 ground motion scaled to the DBE and MCE



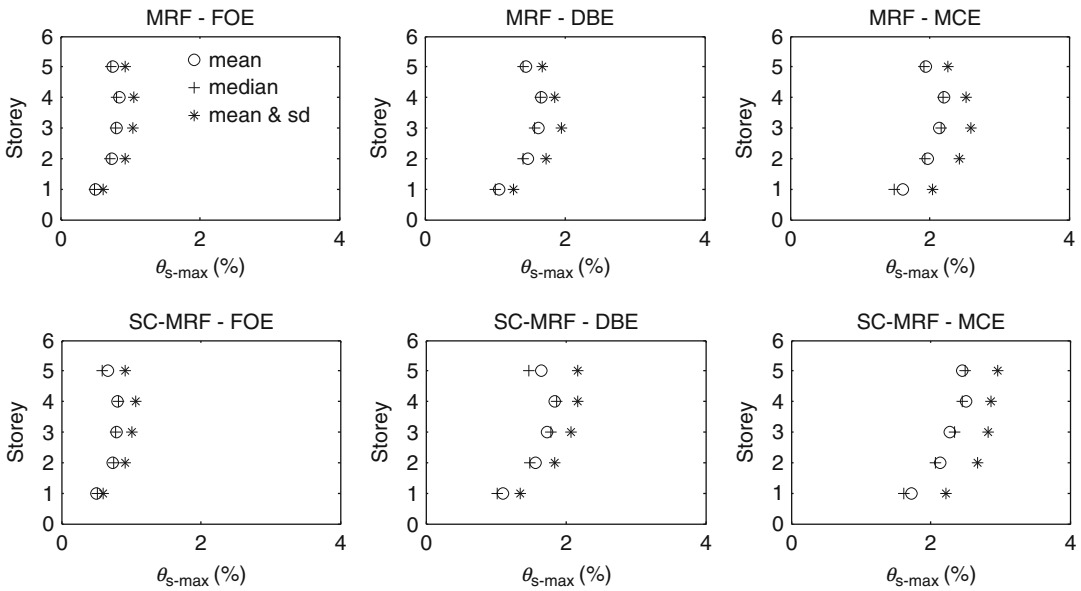
Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 30 Stress-strain hysteresis at points A, B, C,

and D of the conventional MRF under the 5082-235 ground motion scaled at the DBE and MCE



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 31 Stress-strain hysteresis at points A, B, C,

D, E, and F of the SC-MRF under the 5082-235 ground motion scaled at the DBE and MCE



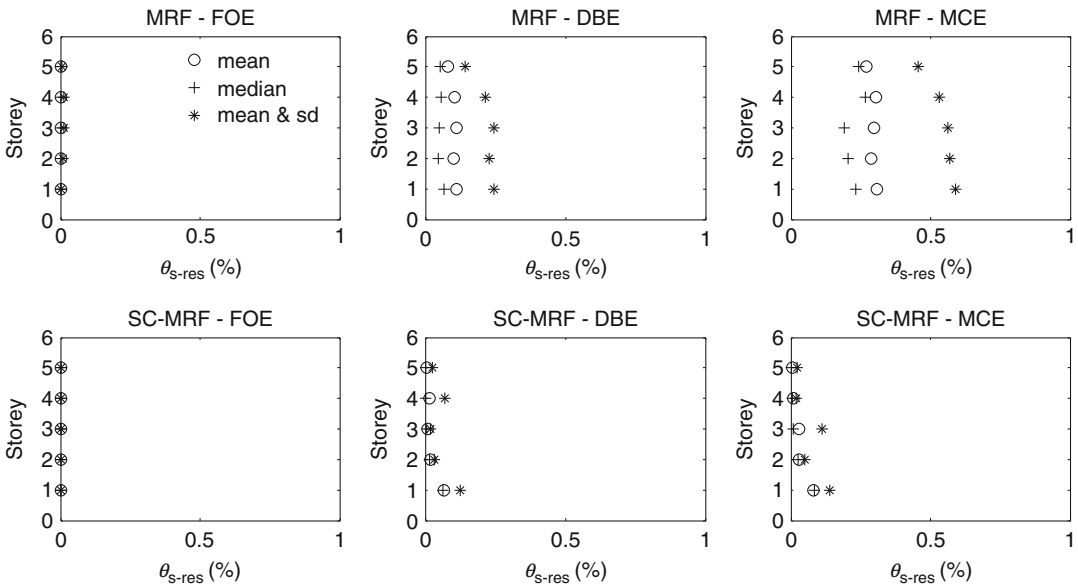
Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 32 Statistics of peak story drifts of the

conventional MRF and the SC-MRF under 20 earthquake ground motions scaled to the FOE, DBE, and MCE

flange-reinforcing plate (points E and F) of the SC-MRF is also shown. The SC-MRF eliminates beam plastic deformations under both the DBE and MCE, while the conventional MRF experiences appreciable beam damage. Figures 28 and 29 show that the column bases of the SC-MRF experience larger plastic deformations than those of the column bases of the conventional MRF.

Figure 32 shows μ , $\mu + \sigma$, and median $\theta_{s-\max}$ values under the earthquake ground motions of Table 4 scaled to the FOE, DBE, and MCE. The μ , $\mu + \sigma$ and median height-wise $\theta_{s-\max}$ distributions show identical shapes. The MRF has the largest $\theta_{s-\max}$ in the fourth story with μ values equal to 0.75 % under the FOE, 1.65 % under the DBE, and 2.2 % under the MCE, i.e., close to the design values of 0.64 % under the FOE and 1.6 % under the DBE and smaller than the design value of 2.4 % under the MCE. The SC-MRF has the largest $\theta_{s-\max}$ in the fourth story with mean values equal to 0.75 % under the FOE, 1.8 % under the DBE, and 2.5 % under the MCE, i.e., slightly larger than the DBE and MCE design ones. Figure 33 shows μ , $\mu + \sigma$, and median values of the residual story drifts, $\theta_{s-\text{res}}$. $\theta_{s-\text{res}}$ values show a

uniform height-wise distribution for the conventional MRF and large dispersion compared to that of $\theta_{s-\max}$. The largest $\theta_{s-\text{res}}$ of the conventional MRF occurs in the first story with mean values equal to 0.1 % under the DBE and 0.3 % under the MCE. The associated $\mu + \sigma$ $\theta_{s-\text{res}}$ values are equal to 0.25 % under the DBE and 0.6 % under the MCE. The latter $\theta_{s-\text{res}}$ values indicate that repair of damage in the conventional MRF would be costly and disruptive after the DBE and not financially viable after the MCE (Mc Cormick et al. 2008). These results highlight the need for Eurocode 8 to include residual deformations as an additional seismic performance parameter. The SC-MRF practically eliminates residual story drifts apart from the first story that has μ and $\mu + \sigma$ $\theta_{s-\text{res}}$ values equal to 0.1 % and 0.15 % under both the DBE and MCE. The latter $\theta_{s-\text{res}}$ values are lower than the global sway imperfections defined in EC3 (2003) and so it can be assumed that there will be no need for these residual drifts to be straightened out. Figure 33 shows small $\mu + \sigma$ $\theta_{s-\text{res}}$ values in the third story of the SC-MRF due to modest yielding in the beam ends.



Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures, Fig. 33 Statistics of residual story drifts of the

conventional MRF and the SC-MRF under 20 earthquake ground motions scaled to the FOE, DBE, and MCE

Conclusions

A new steel posttensioned connection was developed for use in self-centering moment-resisting frames. The connection performance was validated through a large-scale experimental program and extensive numerical analyses, including seismic analysis of steel frames equipped with the proposed connection. The main conclusions of this research are:

(a) *Based on the experimental results:*

- The proposed connection has stable self-centering behavior, enhanced energy dissipation capacity, and strength and stiffness comparable to those of a welded connection.
- The proposed connection eliminates residual deformations and avoids beam damage for drifts lower or equal to 6 %.
- A simplified analytical procedure using plastic analysis and simple mechanics was found to accurately predict the connection behavior.
- Repeatable tests on a connection specimen were conducted along with replacing

damaged WHPs. These tests showed that WHPs can be easily replaced without welding or bolting, and hence, the proposed connection can be repaired with minimal disturbance to building use or occupation in the aftermath of a major earthquake.

- Web local buckling at the connection region after the beam flange-reinforcing plates can be avoided by using web stiffeners. However, this detailing results in excessive local yielding at the beam–column interface due to high bearing forces.
- To accommodate the gravity loads of the frame and to facilitate the replacement procedure of the damaged WHPs, it is necessary to include a vertical loading transfer system, e.g., a shear tab with slotted holes or a seat angle below the beam.
- It should be emphasized that the tests described herein were performed on a 300 mm deep beam section. For a given rotation, the demand on the WHPs and PT bars is expected to increase for deeper sections. The response of the proposed

connection should be verified by using deeper beam sections.

(b) *Based on the finite element simulations:*

- The developed nonlinear FEM models can be reliably used to assess the design of the proposed connection as they are capable to trace the hysteretic behavior and predict the local failure modes both of the individual WHPs and the connection when subjected to either monotonic or cyclic loading.
- The FEM analyses confirm that for drifts equal or lower than those expected under the design basis earthquake (DBE), damage in a carefully designed connection is concentrated in the WHPs which are components that can be very easily replaced without welding or bolting.

(c) *Based on the seismic analyses of steel frames with the proposed connection:*

- The proposed model for the PT connection with WHPs and the associated beams and columns has been calibrated against experimental results and found to accurately simulate the hysteretic behavior of the PT connection.
- Nonlinear static monotonic (pushover) analysis shows that the conventional MRF and the SC-MRF have comparable base shear strength and initial stiffness. The conventional MRF experiences significant damage in beams at the DBE drift. On the other hand, the SC-MRF has damage-free beams for drifts even higher than the MCE drift.
- Nonlinear static cyclic (push–pull) analysis shows that the SC-MRF has full re-centering capability and adequate energy dissipation capacity under the DBE.
- Seismic analyses show that the conventional MRF and the SC-MRF have comparable peak story drifts. In particular, the conventional MRF has slightly lower peak story drifts than the SC-MRF. For both frames through the mean peak story drifts are close to the design values.
- Seismic analyses show that the SC-MRF practically eliminates residual story drifts

apart from the first story which sustains small residual drifts due to plastic deformations at the column bases. The mean plus one standard deviation value of the first story residual drift of the SC-MRF is equal to 0.15 % under both the DBE and MCE which is considered small and does not need to be straightened out.

- Seismic analyses show that the mean plus one standard deviation value of the maximum residual story drift of the conventional MRF is 0.25 % under the DBE and 0.6 % under the MCE. These values indicate that repair of damage in the conventional MRF would be costly and disruptive after the DBE and not financially viable after the MCE.
- Seismic analyses show that the beams of the SC-MRF do not exhibit any yielding even under the MCE, while significant inelastic deformations are developed in the beams of the MRF under both the DBE and MCE. On the other hand, the column bases of the SC-MRF experience larger inelastic deformations than those of the conventional MRF.

Summary

This contribution presents the experimental and analytical validation on a new self-centering steel posttensioned connection using web hourglass shape steel pins (WHPs) as energy dissipation devices. The connection isolates inelastic deformations in WHPs, avoids damage in other connection parts as well as in beams and columns, and eliminates residual drifts. WHPs do not interfere with the composite slab and can be very easily replaced without bolting or welding, and so, the connection enables nondisruptive repair and rapid return to building occupancy in the aftermath of a strong earthquake. The experimental results are first presented followed by parametric studies using nonlinear finite element models. A simplified nonlinear model for the connection and the associated beams and columns that consist of nonlinear beam–column

elements and hysteretic and contact zero-length spring elements appropriately placed in the beam–column interface was also developed in OpenSees. The model was calibrated against experimental results and found to accurately simulate the connection behavior. A prototype building was selected and designed as a conventional steel moment-resisting frame (MRF) according to Eurocode 8 or as a self-centering steel MRF (SC-MRF) using the connection with WHPs. Seismic analyses results show that the conventional MRF and the SC-MRF have comparable peak story drifts and highlight the inherent potential of the SC-MRF to eliminate damage in beams and residual drifts. It is shown that repair of damage in the conventional MRF will be costly and disruptive after the design basis earthquake and not financially viable after the maximum considered earthquake due to large residual drifts.

Cross-References

- [Seismic Analysis of Steel Buildings: Numerical Modeling](#)
- [Seismic Collapse Assessment](#)
- [Seismic Strengthening Strategies for Existing \(Code-Deficient\) Ordinary Structures](#)

References

- Charney FA, Downs WM (2004) Connections in steel structures V. ESSC/AISC workshop, Amsterdam, 3–4 June
- Chou CC, Lai YJ (2009) Post-tensioned self-centering moment connections with beam bottom flange energy dissipators. *J Constr Steel Res* 65(10–11):1931–1941
- Chou CC, Chen JH, Chen YC, Tsai KC (2006) Evaluating performance of post-tensioned steel connections with strands and reduced flange plates. *Earthq Eng Struct Dyn* 35(9):1167–1185
- Chou C, Tsai K, Yang W (2009) Self-centering steel connections with steel bars and a discontinuous composite slab. *Earthq Eng Struct Dyn* 38:403–422. doi:10.1002/eqe.
- Christopoulos C, Filiatrault A, Uang CM, Folz B (2002) Posttensioned energy dissipating connections for moment-resisting steel frames. *J Struct Eng* 128(9):1111–1120
- Dassault Systems (2010) Abaqus theory manual. Dassault Systèmes Simulia, Providence
- Dimopoulos AI, Karavasilis TL, Vasdravellis G, Uy B (2013) Seismic design, modelling and assessment of self-centering steel frames using post-tensioned connections with web hourglass shape pins. *Bull Earthq Eng* 11(5):1797–1816. doi:10.1007/s10518-013-9437-4
- EC3 Eurocode 3 (2003) Design of steel structures
- EC8 Eurocode 8 (2004) Design of structures for earthquake resistance
- Garlock M, Ricles JM, Sause R (2005) Experimental studies of full-scale posttensioned steel connections. *J Struct Eng* 131(3):438–448
- Garlock M, Sause R, Ricles JM (2007) Behavior and design of posttensioned steel frame systems. *J Struct Eng* 133(3):389–399
- Karavasilis TL, Ricles JM, Sause R, Chen C (2011) Experimental evaluation of the seismic performance of steel MRFs with compressed elastomer dampers using large-scale real-time hybrid simulation. *Eng Struct* 33(6):1859–1869
- Karavasilis TL, Kerawala S, Hale E (2012) Model for hysteretic behaviour of steel energy dissipation devices and evaluation of a minimal-damage seismic design approach for steel frames. *J Constr Steel Res* 70:358–367
- Kim HJ, Christopoulos C (2008a) Friction damped posttensioned self-centering steel moment-resisting frames. *J Struct Eng* 134(11):1768–1779
- Kim HJ, Christopoulos C (2008b) Seismic design procedure and seismic response of post-tensioned self-centering steel frames. *Earthq Eng Struct Dyn* 38(3):355–376
- Kim H, Christopoulos C (2009) Numerical models and ductile ultimate deformation response of post-tensioned self-centering moment connections. *Earthq Eng Struct Dyn* 38:1–21. doi:10.1002/eqe.
- Kobori T, Miura Y, Fukuzawa E, Yamada T, Arita T, Takenaka Y, Miyagawa N, Tanaka N, Fukumoto T (1992) Development and application of hysteresis steel dampers. In: *Earthquake engineering, tenth world conference*. Balkema, Rotterdam
- Mazzoni S, McKenna F, Scott M, Fenves G (2006) Open system for earthquake engineering simulation (OpenSees). User Command Language Manual, Pacific Earthquake Engineering Research Center, University of California, Berkeley
- McCormick J, Aburano H, Ikenaga M, Nakashima M (2008) Permissible residual deformation levels for building structures considering both safety and human elements. 14th WCEE, Beijing
- Ricles J, Sause R, Garlock M, Zhao C (2001) Posttensioned seismic-resistant connections for steel frames. *J Struct Eng* 127(2):113–121
- Ricles J, Sause R, Peng SW, Lu LW (2002) Experimental evaluation of earthquake resistant posttensioned steel connections. *J Struct Eng* 128(7):850–859
- Rojas P, Ricles JM, Sause R (2004) Seismic performance of post-tensioned steel moment resisting frames with friction devices. *J Struct Eng* 131(4):529–540

- Somerville P (1997) Development of ground motion time histories for phase 2 of the FEMA/SAC steel project, Report No. SAC/DB-97/04, Sacramento
- Tsai KC, Chou CC, Lin CL, Chen PC, Jhang SJ (2008) Seismic self-centering steel beam-to-column moment connections using bolted friction devices. *Earthq Eng Struct Dyn* 37:627–645
- Vasdravellis G, Karavasilis TL, Uy B (2013a) Large-scale experimental validation of steel post-tensioned connections with web hourglass pins. *J Struct Eng* 139(6):1033–1042
- Vasdravellis G, Karavasilis TL, Uy B (2013b) Finite element models and cyclic behavior of self-centering steel post-tensioned connections with web hourglass pins. *Eng Struct* 52:1–16
- Vasdravellis G, Karavasilis TL, Uy B (2014) Design rules, experimental evaluation and fracture models for high-strength and stainless-steel hourglass shape energy dissipation devices. *J Struct Eng (ASCE)*. doi:10.1061/(ASCE)ST.1943-541X.00010
- Whittle J, Williams MS, Karavasilis TL, Blakeborough T (2012) A comparison of viscous damper placement methods for improving seismic building design. *J Earthq Eng* 16:540–560
- Wolski M, Ricles JM, Sause R (2009) Experimental study of a self-centering beam-column connection with bottom flange friction device. *J Struct Eng* 135(5): 479–488

Steel Structures

Roberto Leon

The Charles Edward Via, Jr. Department of Civil and Environmental Engineering, Virginia Tech, Blacksburg, VA, USA

Synonyms

Metal structures; Steel connections; Steel construction

Introduction

The use of metals as the principal construction material for modern buildings and bridges dates back to the beginning of the Industrial Revolution (Bodsworth 2001; Ashton 1968). Three metallic materials have been commonly used in the construction industry: cast and wrought iron, steel,

and aluminum. Cast iron, which was first manufactured as early as fifth century BC, is typically considered to be too brittle and difficult to join but received widespread application in construction through the late eighteenth century. The use of cast iron is uncommon in today's construction industry, but some ductile cast irons continue to be used in specialized applications such as water pipes. The use of cast iron was rapidly replaced by wrought iron, which was easier to work with, leading to the construction of the first large metal structure, the Iron Bridge at Coalbrookdale by A. Darby in 1780 (www.greatbuildings.com/buildings/Iron_Bridge_at_Coalbrookdale.html). Extensive use of iron continued through the first half of the nineteenth century, culminating in iconic structures such as the Crystal Palace in London (http://en.wikipedia.org/wiki/The_Crystal_Palace).

Wrought iron began to be replaced by steel in civil structures when the latter became commercially available in large quantities in the late nineteenth century AD through the development of the Siemens and Bessemer processes and the development of rolling technologies (Bodsworth 2001; Ashton 1968; Halmos 2000). Initially steel was rolled into flat plates and small sections (angles and channels or L- and C-sections) that were assembled into larger structural sections through riveting (Fig. 1), just as it had been done for wrought iron. Riveted, or built-up sections, were used through the first half of the twentieth century.

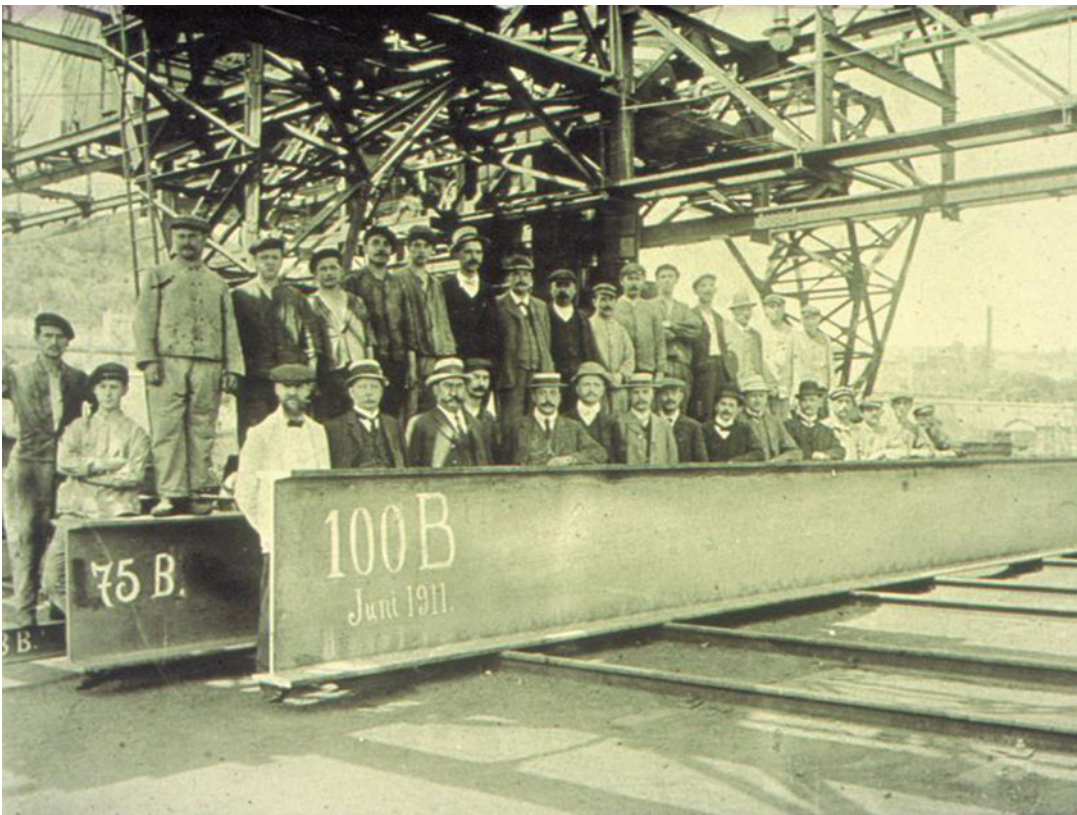
Steel rolling, which began to produce large sections at the turn of the twentieth century (Fig. 2), led to the development of optimal W- (or H-) sections for carrying flexure and axial loads. Advanced fabrication techniques, recycling, and improvements in energy consumption during production have resulted in extremely efficient and sustainable construction practices in the steel industry in the twenty-first century. Much of today's structural steel is manufactured in minimills that utilize scrap steel as their source materials; some estimates are that up to 95 % of steel is recycled in the USA.

The use of aluminum, which became common in the 1940s when the need to build light

Steel Structures,

Fig. 1 Riveted built-up steel sections in a bridge truss (thebridgehunter.
areavoices.com,

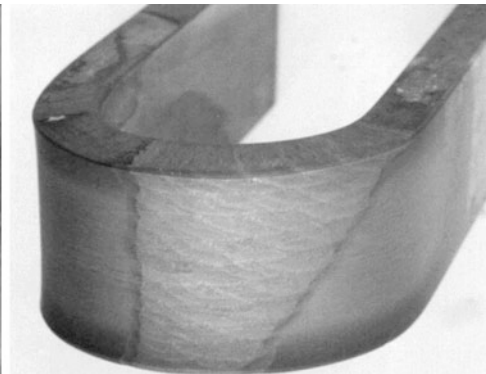
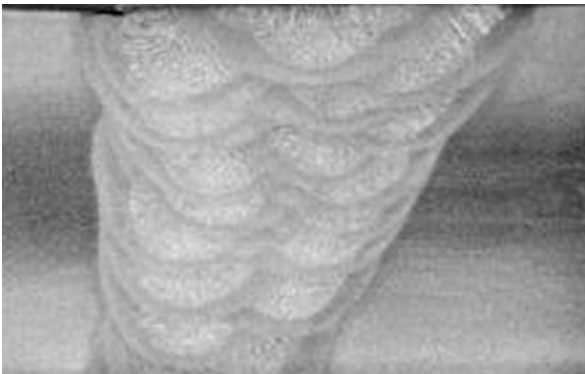
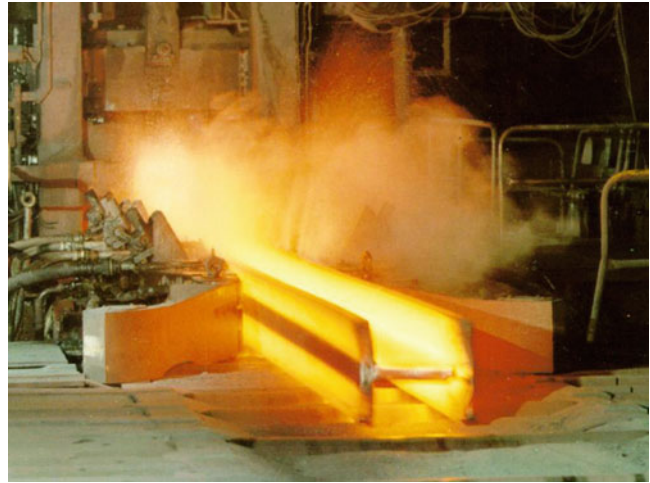
© RTLeon)



Steel Structures, Fig. 2 Rolling of deep sections (75 and 100 cm) in 1911. Note that flanges are not as wide as modern sections (© J.C. Gerardy, ArcelorMittal)

Steel Structures,

Fig. 3 Modern rolling of large steel wide-flange section showing selective cooling (*darker area*) to improve metal performance (© J.C. Gerardy, ArcelorMittal)



Steel Structures, Fig. 4 Carefully executed weld (*left*) showing numerous passes and good workmanship; bend test from this weld showing excellent deformation

capacity as the weld does not fracture even if bent at 180° (© J.C. Gerardy, ArcelorMittal)

airplanes made its strength-to-weight ratio very attractive, is also limited in today's construction industry due to cost and welding issues. Its use is limited mostly to facades, finishes, and nonstructural elements.

Steel for Seismic Applications

A large number of improvements through better metallurgy (heat treatments (Tylecote 1976) and the development of alloy steels with large deformation capacity, Fig. 3) and joinery (welding and high-strength bolts, Fig. 4) in the twentieth century have made steel the preferred construction

material to withstand the large cyclic loads imposed by earthquakes.

The use of steel in multistory framed buildings began in the 1880s in Chicago, notably with the Home Insurance Building (Condit 1968) and similar skyscraper buildings. The use of steel frames was still in its maturing stage when the excellent behavior of steel structures, when subjected to large ground motions, became evident after the San Francisco earthquake and fire of 1906 (Kurzman 2001). These events highlighted the lateral resistance (strength) and deformation capacity (ductility) of steel structures (Hamburger and Meyer 2006), but also their susceptibility to fire unless suitably protected (Fig. 5).



Steel Structures, Fig. 5 Damaged buildings from the 1906 San Francisco earthquake and fire. The two tall steel frames were under construction and were not damaged.

The structure of the burned building in the center of the photograph was undamaged (en.wikipedia.org)

Most structural members were made up of riveted plates and connections for larger structures (Fig. 6) and riveted smaller rolled sections for buildings (Fig. 7). Structures that contained steel columns and beams encased in masonry or concrete survived the earthquake and fire relatively unscathed (Fig. 5). The superior performance of steel frames was enshrined in the report by ASCE (1906), which stated: *The well-designed steel frame offers the best solution of the question of an earthquake proof building, as all the stresses can be cared for.*

This and similar observations by many others led to the widespread utilization of such systems in the earthquake-prone areas of the Western USA in the first half of the twentieth century. These structures consisted of rolled steel sections joined by connections made by riveted steel

angles or plates. Initially, the encasement was made from masonry rubble inside masonry facades, but this was later replaced by encasement with lightly reinforced concrete (Fig. 7). Today, the synergistic use of steel in tension and concrete in compression has led to the development of composite steel–concrete construction particularly for high-rise construction (Viest et al. 1997).

Three important developments have occurred in the steel construction industry since the 1950s. The first development is the change of joining methods, as rivets have been replaced by both high-strength bolts and welds. The use of industrialized welding, which was developed in the early twentieth century for the shipping industry, led to much simpler, stronger, and stiffer connections (Fig. 8). These connections were considered

Steel Structures,

Fig. 6 Typical built-up riveted members and connection for an elevated railway structure (Engineering News, 1914)

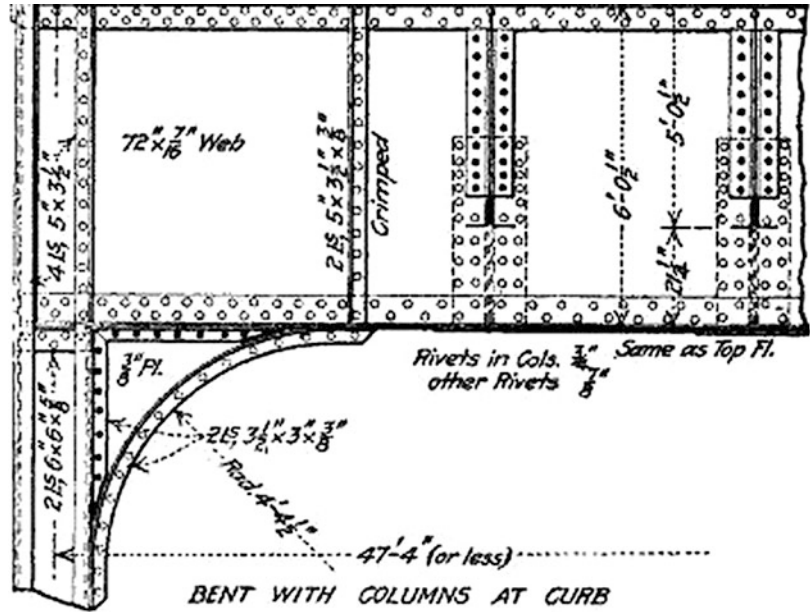
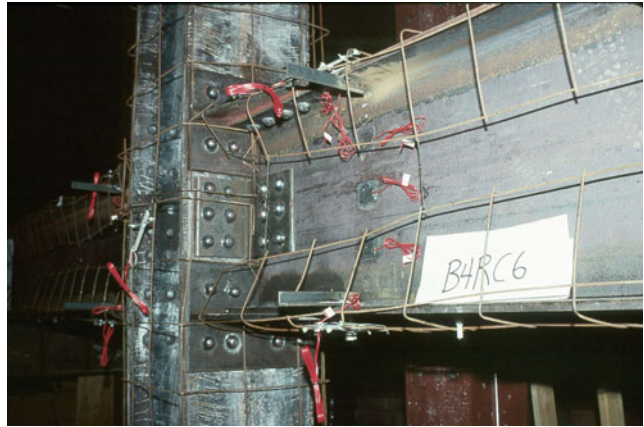
**Steel Structures,**

Fig. 7 Riveted building beam-to-column connection before encasement (© RTLeon)



for many years to be the most reliable structural system in seismic areas. Failures in welded connections observed after the 1994 Northridge (Fig. 9) and 1995 Kobe earthquakes (FEMA 355E 2000a) have led to radical changes in the specifications for such joints. While both joining methods can be used in seismic construction, the use of welding requires extensive planning, inspection, and QA/QC procedures to ensure desirable ductile performance (FEMA 355B 2000b). In the aftermath of those failures, design requirements for bolted connections have also been increased even though few bolted

connection failures have been observed after earthquakes, except for the case of under-designed brace connections. Welded connections are still preferred by many designers because of their perceived superior strength and stiffness characteristics and ease of design when compared to bolted connections.

The second development is that most floors in steel structures have become composite ones, with the use of shear studs to connect floor beams and girders to concrete slab cast in metal deck (Viest et al. 1997). The use of the metal deck as both formwork and reinforcement, the

Steel Structures,

Fig. 8 Modern welded connection to the beam flanges; bolted connection to the beam web is used for erection (FEMA 355E, 2000)

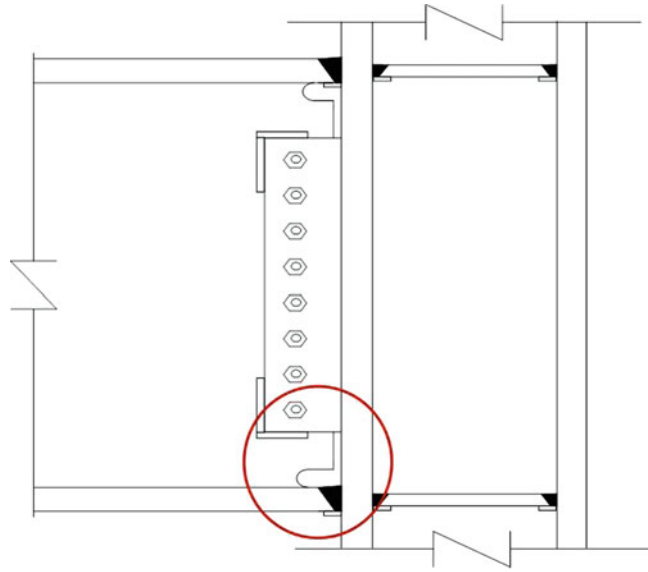
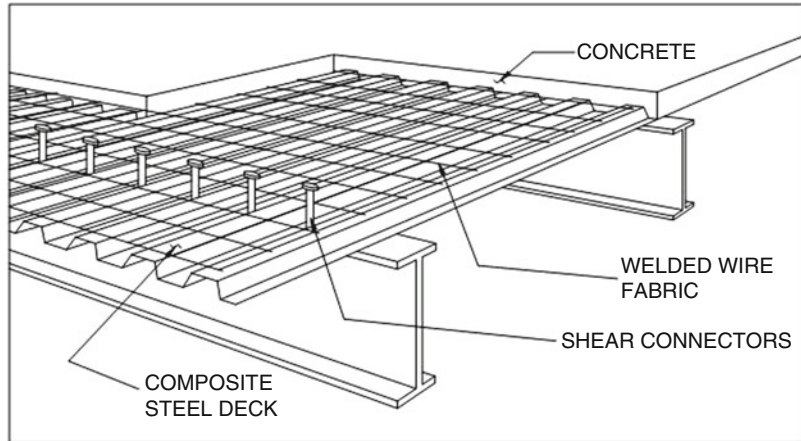
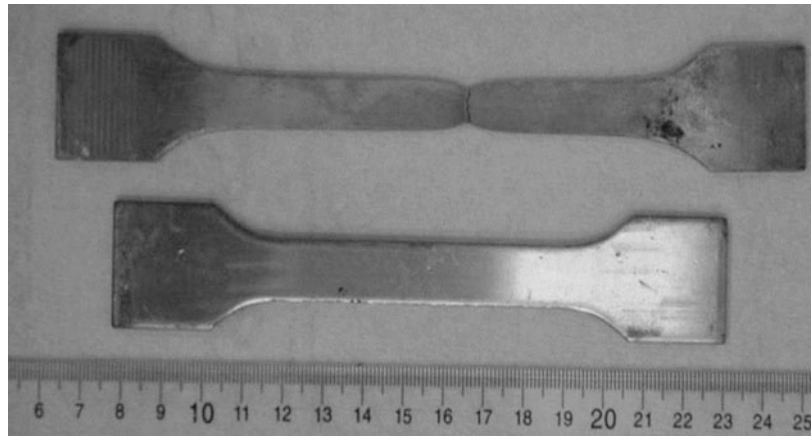
**Steel Structures,**

Fig. 9 Laboratory simulation of connection weld brittle failure as observed in the 1994 Northridge earthquake (The photo shows the area circled in red in Fig. 7 (© RTLeon))



ease of welding the shear studs that act as connectors between the steel beam and the concrete floor, and the tremendous increases in strength and stiffness have made composite floor construction a most economical system (Fig. 10). Floors of this type provide sufficient in-plane stiffness, such that all parts of the lateral force-resisting system can be assumed to work together (rigid floor diaphragm assumption), so that all lateral load-resisting systems can act concurrently. This characteristic is important as many frames in the USA utilize lateral load systems engaging only partially the perimeter frames.

The third important change has been the development of numerous new structural systems and proprietary connections that have given designers a very wide choice of technologies for use in seismic design. The development of such systems has been made possible by both extensive experimental testing and advanced simulation tools brought forth by the use of computers and improvements in analytical methods. Some of these systems are discussed under *Structural Systems* later in this article. Striking evidence of the advantages of these new systems can be seen in the excellent performance of steel *eccentrically braced frames* (EBF) in the recent

Steel Structures,**Fig. 10** Typical composite floor system (© RTLeon)**Steel Structures,****Fig. 11** Ductility of steel as shown by the $>25\%$ elongation of the original coupon (*bottom*) to the fractured one (*top*) (© RTLeon)

Christchurch earthquakes (Clifton et al. 2011) as compared to that of modern reinforced *special moment-resisting reinforced concrete frame* (SMRF-CR) and shear wall structures (Leon et al. 2014).

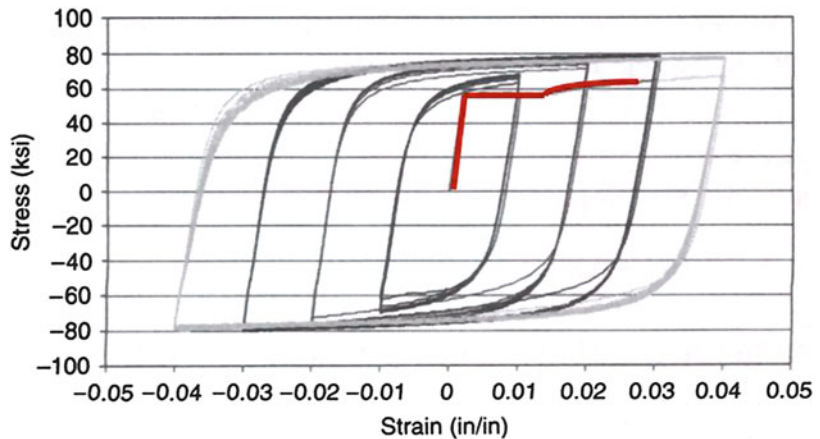
Steel Properties

For structural design purposes, steel is an iron-carbon alloy that contains many additional alloying elements such as Mn, Mo, Cr, Ni, and V. A large variety of mechanical properties can be achieved with steels by either varying its alloys (primarily carbon which constitutes 0.2–0.6 % of steel) and/or subjecting the material to different heat treatments that change its grain microstructure.

Steel is an ideal material for seismic design because both of its isotropic and homogeneous properties and its ability to undergo large local plastic deformations before failure. Figure 11 shows a typical mild steel test coupon elongating in excess of 15 % over a 2 in. length and more than 50 % locally. The stress–strain (or load–deformation) behavior for this type of steel is shown by the red line in Fig. 12. After an initial stiff elastic response, the steel reaches its yield point (knee in the curve), the steel deforms plastically (i.e., undergoing large deformations with little small increments of load, shown by the flat portion of the curve), and picks up additional resistance as the deformation becomes large (last part of the curve, termed the strain-hardening region). The curve for a steel coupon subjected to large deformation reversals such as those experienced

Steel Structures,

Fig. 12 Cyclic stress–strain behavior of a mild structural steel
(© Alan Pense, Lehigh University)



locally during a large seismic event is shown by the rest of the curves in Fig. 12. If loaded cyclically, the steel will show fat, stable, and hardening hysteresis (stress–strain) loops, evidence of good energy dissipation capacity. Failure by ductile fracture will ideally occur after many cycles of deformation. For seismic design, steel will perform best when deformed in shear.

Steels for structural use are classified as carbon steels, high-strength low-alloy steels, and alloy steels. For design purposes, these steels can be assumed to have a density of 7.85 g/cm^3 , a modulus of elasticity of 210 GPa, and a Poisson's ratio of 0.3. Carbon steels are classified based on the percentage of carbon. Mild carbon steels (0.15–0.29 % C) with yield points in the range of 220–250 MPa and tensile strengths of 400–500 MPa are the most common structural carbon steels. Typically, an increase in carbon percent raises the yield point and increases hardness, but reduces ductility and makes welding more difficult. These drawbacks can be minimized by heat treatments.

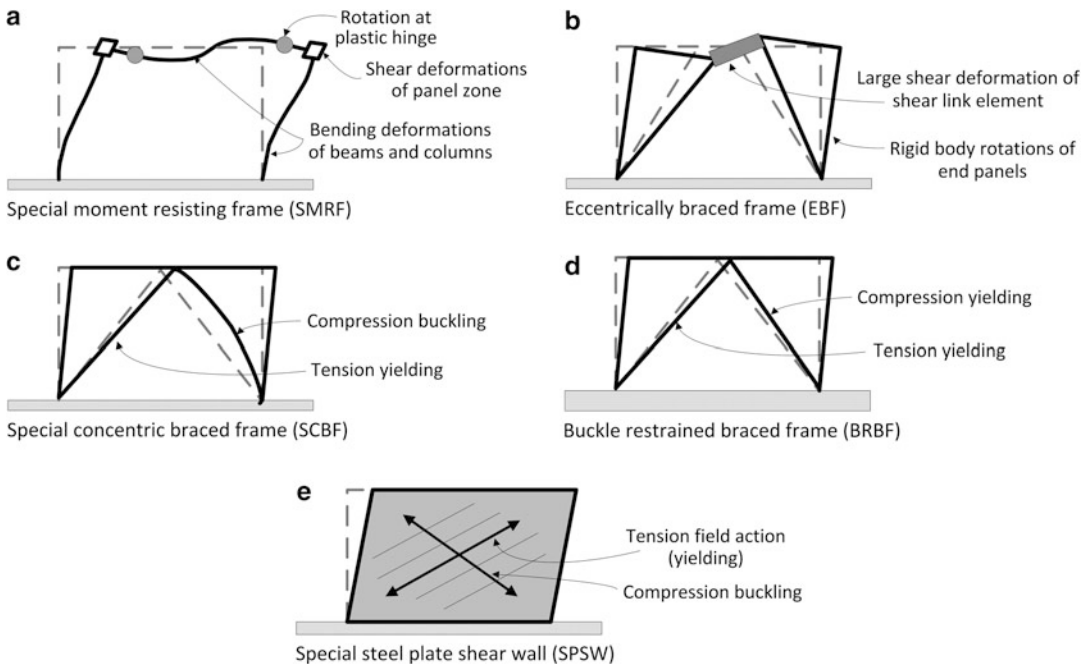
The stress–strain behavior of a mild carbon steel is characterized by a sharp yield point and large ultimate strains (percent elongation $>20\%$). High-strength low-alloy steels exhibit higher yield points (345–480 MPa) and somewhat larger tensile strengths (450–600 MPa) without an appreciable drop in ductility. Alloy steels are heat-treated steels (mostly quenched and tempered) with yield points of 550–760 MPa, but lower tensile to yield ratios and ductility than for other structural steels.

Other important engineering properties needed for seismic design are notch toughness and weldability. Notch toughness refers to the ability of the material to resist fracture propagation from an existing defect when subjected to dynamic loads. Fracture propagation is resisted through local plastic deformations and is primarily a function of temperature and heat treatment. Weldability is primarily a matter of obtaining a structural joint free of any undesirable defects by utilizing appropriate electrode materials and weld procedure specifications (WPS).

Structural Systems

As noted earlier, one of the great developments in seismic construction has been the promotion of a large variety of steel structural systems. These systems are classified as *special*, *intermediate*, and *ordinary* based on the amount of seismic detailing present. A *special* system will require great care in design and detailing to obtain the large deformation capacity needed to activate the plastic deformations and energy dissipation characteristic of these systems. The benefit of a special system is lower design lateral forces, which results in savings in both material and construction costs.

At its most basic, lateral force-resisting systems can be visualized as a continuum, with a pure frame system (Fig. 13a) at one extreme and a wall system at the other end (Fig. 13e). In these



Steel Structures, Fig. 13 Deformation mechanisms for typical steel seismic force-resisting systems (© RTLeon)

figures, only the primary structural elements are shown; secondary structural elements, such as out-of-plane braces, floor slabs, chord, collectors, and diaphragms, are not shown. Also not shown are *nonstructural elements*, such as partition walls, which are assumed not to contribute significantly to lateral resistance and should be designed so as not to interact with the primary structural members. If not properly isolated, nonstructural elements can have deleterious effects on frame behavior.

The *moment-resisting frame* (MRF), a combination of slender beams and columns, was the first type of steel structural system used and which showed excellent seismic behavior (Fig. 5). In a pure moment-resisting frame (MRF, Fig. 13a), the lateral deformations are to be accommodated primarily by bending of the beams and columns, shear deformation of the panel zone, and the formation of *plastic hinges*, or areas of concentrated plasticity, in the beams. The yielding of the beam steel and the formation of a plastic hinge are shown by the flaking of the whitewash in the critical section of the beam in Fig. 14. All of these deformation mechanisms can

be very ductile when properly designed, but pure frame structures tend to be rather flexible. To eliminate or limit damage to nonstructural elements and contents under both small earthquakes and large wind loads, special moment-resisting frames need to meet some maximum deformation criteria (typically $\approx 3\%$ drift under earthquake loads and 0.25% under wind loads) and thus are said to be *drift controlled*. It is clear that rigid and strong connections that limit the shear deformations of the panel zones and promote beam yielding are desirable (note thick end plate and large bolts in the connection in Fig. 14), as long as the local buckling of the beam flanges and web can be delayed. In addition, mechanisms that can reduce the strength demand at the welds at the joint without substantially decreasing stiffness, such as reduced beam sections (RBS, Fig. 15), are desirable. In an RBS, the steel in the beam flanges is cut to intentionally weaken the beam and promote the formation of a plastic hinge away from the welded connection.

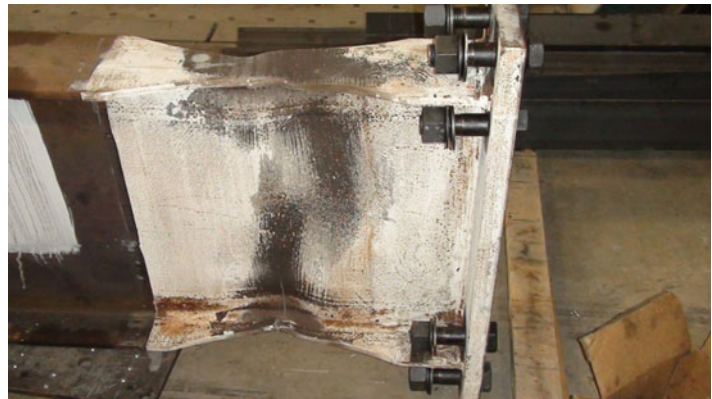
At the other extreme, one can think of a wall structure composed of a steel frame infilled with a steel plate (Fig. 13e) as the analogue to a

Steel Structures,

Fig. 14 End-plate connection showing plastic hinging followed shortly by local buckling of the flanges (© RTLeon)

**Steel Structures,**

Fig. 15 End-plate connection showing reduced beam section (RBS) (© RTLeon)



reinforced concrete shear wall system. For economy reasons, in steel structures it is desirable to use a rather thin plate. When displaced laterally, the plate will buckle in compression along one diagonal, allowing a tension field to develop in the other diagonal (Fig. 16). This creates a large tensile strut mechanism that transfers the horizontal forces similarly to a braced frame (Fig. 13c). This force transfer is exactly the inverse of a concrete shear wall structure, where a compression strut will form, as the concrete is strong in compression but weak in tension. The great advantage of a steel shear wall over a concrete one is that whereas the behavior of a concrete shear wall will degrade rapidly due to an X-crack pattern that develops with cycling, the steel shear wall performance will not degrade substantially due to the

post-buckling strength of metal plates. Steel-plated shear walls are relatively modern structures, with a few implementations as far back as the late 1970s, but have become more accepted only in the last decade or so.

In between these two limits, one can visualize a large number of alternatives. Starting at the bottom of Fig. 13, one can visualize a system where the struts in the walls are replaced by discrete, strong, and rigid braces (or diagonal elements, Fig. 17). This system is known as a *buckling-restrained brace frames* (BRBF, Fig. 13d), because the braces in compression are so stout that they will not buckle but yield. The buckling in these systems is limited to an interior core of the brace, with the exterior portion being disconnected and acting only to restrict the

Steel Structures,

Fig. 16 Thin plate under lateral loads showing tension-field action from *top right to bottom left* and buckling in the opposite direction (© M. Kurata)

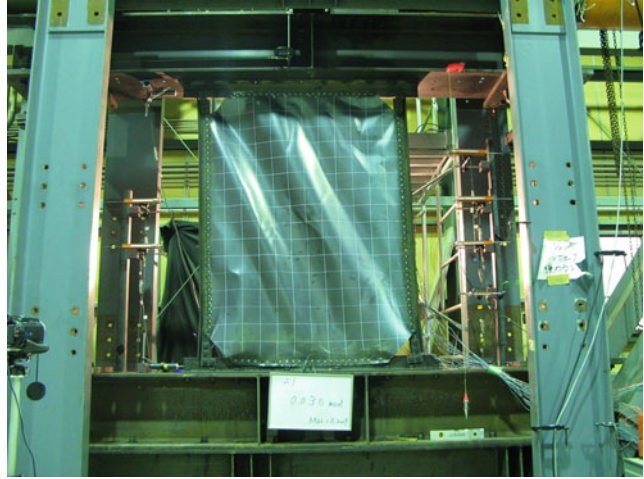
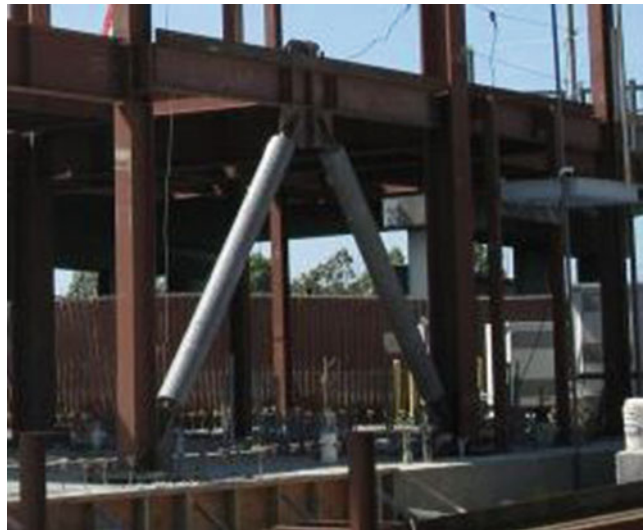
**Steel Structures,**

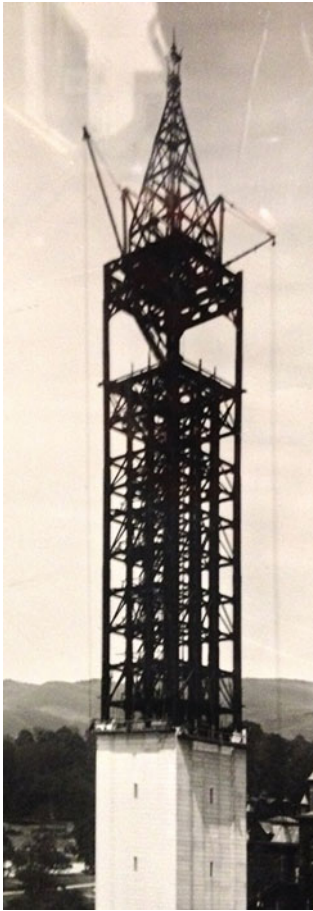
Fig. 17 Use of large BRBs in a modern structure (Berkeley Animal Shelter) (© ci. berkeley.ca.usa)



buckling of the brace. Many of these buckling-restrained braces are proprietary devices. BRBF systems originated in Japan in the early 1990s and are a very common type of structural system in the Western USA today.

The next system is similar, except that the braces are not so stiff and they will buckle slightly in compression. This system is known as a *special concentrically braced frame* (SCBF, Fig. 13c). In this system, the brace in tension is assumed to carry most of the force (generally about 70 %) as the compression brace will buckle and be able to provide only its post-buckling resistance. *Concentrically braced*

frames (CBF) originated when engineers began to stiffen MRF for wind loads in the 1910s using light diagonals, as in the Sather Tower in Berkeley, CA (Fig. 18). Modern CBF are much stiffer than SMRF, but generally less ductile as the buckling of the braces can lead to a concentration of deformations in a few floors. It is not clear when CBF began to be used in seismic systems (i.e., going from CBF to SCBF), but seismic design recommendations for braced frames began to appear in the 1960s. The original CBF had slender X-shaped braces in which only the brace in tension provided resistance. There are a large number of variations of this



Steel Structures, Fig. 18 Sather Tower under construction (1914)

system (Fig. 19), each with its own design requirements and behavior characteristics.

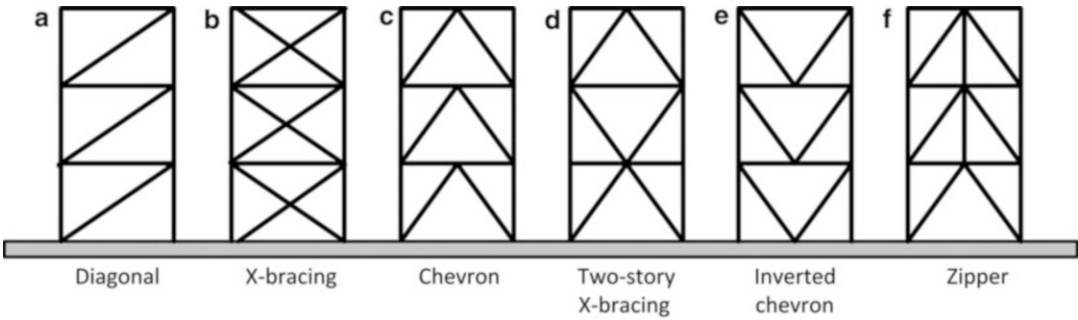
Finally, one can think of beginning to move the brace in each bay horizontally, such that a hybrid structure between a pure moment frame and a braced frame is achieved. This is known as an *eccentrically braced frame* (Fig. 13b). In these structures, the central portion of the beam, known as the link, will deform and yield in shear if the link is short and in a combination of flexure and shear if the link is longer. Eccentrically braced frames have performed extremely well in recent earthquakes (Fig. 20), with numerous such structures providing superior performance during the 2011 Christchurch earthquake (Clifton et al. 2011).

Basis of Design

The seismic design of steel structures is predicated on having little or no structural damage under a service-level earthquake, moderate but rapidly repairable damage under a moderate earthquake, and not to collapse and cause loss of lives in the event of the design-level earthquake. Until recently, design specifications typically addressed only the design-level earthquake through a series of prescriptive design provisions intended to address ultimate strength design (ULD) criteria (limit states). However, the costs and time related to both loss of contents and functionality under smaller events and the desire to provide some freedom from prescriptive building codes have begun to drive the design process toward a performance-based design (PBD) approach. A pure PBD approach will be completely non-prescriptive, giving the structural engineer complete freedom in determining both the limit states and how to comply with them. In the short term (next 5–15 years), it is likely that a hybrid model for seismic design with elements of both ULS and PBD will emerge.

Whether a ULD or a PBD approach is used, the basis for design of steel structures remains the need to provide ductility, redundancy, robustness, and resiliency. Ductility comes from a synergistic use of the very ductile behavior of steel at the material level to create elements, connections, and structural systems capable of sustaining large inelastic deformations without appreciable stiffness or strength degradation. From the design standpoint, most of our ductility is ensured through a series of prescriptive requirements that have been developed through observed performance in past earthquakes (Fig. 19), advanced analytical studies (Fig. 20), and complex laboratory testing (Figs. 14, 15, and 16).

Redundancy is the ability of the structure to redistribute forces as inelastic action occurs in order to efficiently activate all major lateral load resistance systems. In its simplest form, redundancy can be defined as providing multiple load paths in a structure. Redundancy is primarily a function of the number and ductility of the lateral



Steel Structures, Fig. 19 Different braced frame configurations (© RTLeon)

Steel Structures, Fig. 20 Damaged eccentrically brace frame, showing yielding of the shear link element (© RTLeon)



load-resisting systems employed, the system's rational 3D configuration, the ductility of connections, the strength of diaphragms and collector elements, and the detailing of the structure. Importantly, redundancy implies activating not only the lateral load-resisting system but also the gravity load system and other nonstructural elements that can contribute to the strength and stiffness of the system.

Robustness refers to the ability of the structural system to limit any failure to a relatively small part of the structure so that the consequences of the failure are not disproportionate to the initial failure. For example, a common way of determining the robustness of a moment frame is to remove an interior column in a lower floor (something that may happen during an earthquake) and assess the stability of the system under gravity loads (to make sure that the structure does not “pancake” and crush its

occupants). If the structure does not collapse, regardless of its damaged state, it will be considered robust. Clearly, a client can establish more stringent criteria for robustness in her/his structure if deemed necessary.

Resiliency refers to the ability of the structure to fulfill its intended functions after a large event. For example, a level-four trauma hospital would be resilient if it can continue to function uninterrupted after the design-level earthquake. While it is possible today to provide that level of performance for the structural system, it is rather more difficult to do it for all the other systems (from electricity, water, oxygen, and the myriad of IT systems) present in such hospitals. For a large store, it may mean being able to get back in business within a few days and at a minimum of cost. Resiliency can be quantified in many ways, including the ability to resist ground motions significantly larger than the design one, the

costs to repair the structure, the time needed to carry out such repairs, and the indirect costs associated with business interruption.

Connections

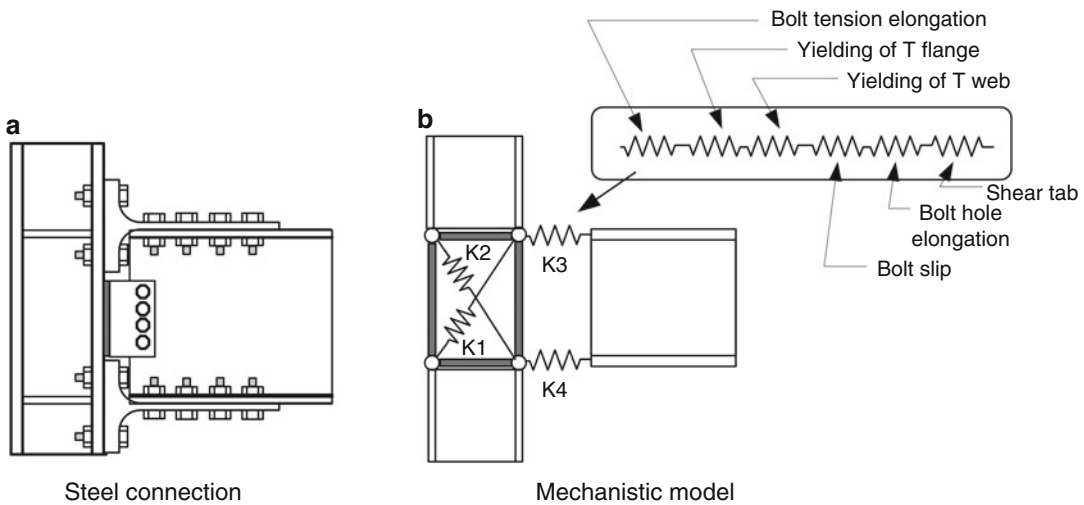
In order to satisfy ductility, redundancy, robustness, and resiliency criteria, steel structures in seismic areas need to be designed primarily to avoid any type of brittle failure modes, which can result in sudden losses of local or global strength, stiffness, and stability. In general, it is possible to design and fabricate steel members (beams, columns, braces, and walls) to avoid these type of failures through close attention to proper material selection, using compact sections (i.e., keeping a small enough ratio of the width of an element to its length.) and providing lateral bracing to prevent out-of-plane buckling. The potential weak links in steel structures are the connections, which must transfer very large forces through complex force paths. The design of connections for seismic loads is a complex design issue as it is difficult to predict failure modes and their interactions.

From the standpoint of steel as a material, brittle failures can occur because of a combination of poor selection of or unmatched materials, large stress concentrations due to poor design or execution, and large triaxial state of stresses. These conditions can often arise in connections. Problems with welded connections were evidenced by many failures in the 1994 Northridge earthquake (Fig. 9). In the welds to the beam flanges in these connections, large triaxial stresses arise from a combination of multidirectional forces, residual stresses from welding, and the inevitable imperfections in the welds. These conditions lead to brittle material behavior because shear forces, the source of most deformations in steel, are small when compared to normal forces. Even if only minor overall damage to the structures was observed in structures with failed connections after the Northridge earthquake, a major effort was launched in the USA to better understand connection behavior.

As part of that effort, known as the SAC Project (FEMA 355B 2000b), extensive experimental studies were conducted on bolted and welded connections. The probability of brittle failure in connections can be minimized by (1) proper attention to selecting appropriate base materials and welding consumables with high toughness; (2) careful consideration of surface preparation, preheating, and welding sequences; (3) inspection; and (4) nondestructive testing. This amounts to the need for a very comprehensive quality assurance/quality control (QA/QC) plan throughout all phases of the construction projects (materials, fabrication, and erection). In the USA, these requirements are now embodied in the required weld procedure specifications (WPS).

Connections with bolts are usually less susceptible to material-type problems, but require more care in the fabrication as slight misalignments and exceeding tolerances can lead to severe problems during erection. In addition, the force transfer mechanisms in connections with bolts are less obvious and require more detailed computations than those with welds. Finally, in general, bolted connections will be less rigid than welded ones, requiring that this semirigid behavior be included in the analyses, substantially complicating the design process. Bolted connections, such as the end plates shown in Figs. 14 and 15, when properly detailed, can provide excellent performance.

From the structural standpoint, ductility should arise in “plastic hinges” or zones of concentrated plasticity primarily in beams. The behavior of a plastic hinge can be visualized as that of a nonlinear rotational spring. The ability of these hinges to rotate depends primarily on delaying the onset of any local or global buckling of the section. Local buckling (Fig. 14) is mostly dependent on the slenderness (width/thickness) ratio of the flanges and webs, and codes prescribe strict limits to ensure that this type of buckling does not occur until large rotations are achieved. Global buckling of a beam or column, which in this case implies large out-of-plane displacements, is also controlled by slenderness criteria,



Steel Structures, Fig. 21 (a) T-stub bolted connection and (b) corresponding component model (© RTLeon)

in the form of length/moment of inertia ratios. Global buckling can be minimized by the addition of bracing along the member length, but this leads to additional costs and complications in the construction process. Global buckling or collapse of the structure can occur when large lateral deformations arise and the gravity forces result in large additional second-order effects. Global buckling is addressed in design by specifying strict drift (lateral deformation) limits and conducting advanced analyses that ensure overall stability.

As the results of the research began to filter into practice, a number of new design technologies became popular. For example, for the successful design of a T-stub connection (Fig. 21a), it is necessary to first determine all the possible ductile and brittle failure modes and prioritize them from most brittle to most ductile. One approach to this task is the component approach, in which each deformation mechanism in a joint is identified and individually quantified through a series of small laboratory component tests and associated analytical studies. These tests are carefully designed to measure one deformation component at a time. Each of these components is then represented by a spring with either linear or nonlinear characteristics (Fig. 21b). These

springs are arranged in series or in parallel and the overall moment-rotation ($M-\theta$) curve derived with the aid of simple computer programs that conduct the analysis of the spring system. In this example, the K1 and K2 springs model the panel zone deformation due to shear, while springs K3 and K4 model the bending deformations of the T-stubs. Springs K3 and K4 are made up of the contributions of several other springs that model different deformation components. With the aid of this approach, it is possible to achieve designs that meet and exceed current performance requirements. Care in the design of these connections can lead to very successful designs (Fig. 22).

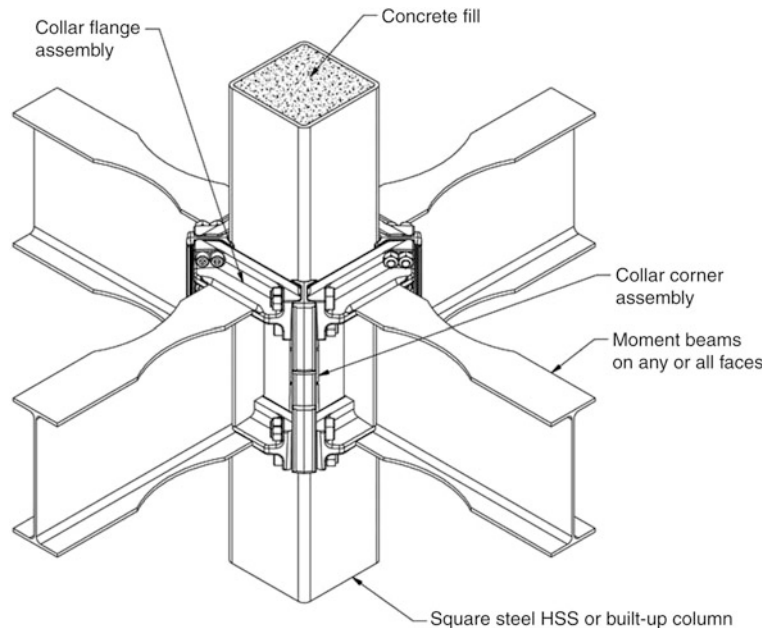
Because the design of connections is a difficult and time-consuming task, in the USA, the idea of “prequalifying” connections has arisen (AISC 358 2010). In this case, an extensive experimental campaign is conducted to test a range of beam and column sizes with a particular type of connection. If the results indicate reliable ductile cyclic behavior to rotations around 4 %, after careful examination of all the results, the connection will become “prequalified.” This implies that a series of simple, predefined design steps is all that is necessary to design that type of connection.

Steel Structures,

Fig. 22 Successful design
of a T-stub connection
(© RTLeon)

**Steel Structures,**

Fig. 23 Innovative,
proprietary connection
(ConXTech, www.conxtech.com)



An interesting and important consequence of the “prequalification” process is that a number of proprietary connections have entered the market. In these cases, the manufacturer of the connection will provide all the necessary connection design data and manufacture all proprietary pieces. For example, the ConXL connection (Fig. 23, ConXtech Inc.) consists of a concrete-filled tube HSS column, RBS beams, and a series of steel castings that make the field assemblage of the connection a fast and simple operation.

Summary

Properly designed modern steel structures can provide superior performance when subjected to very large earthquake motions. The key to good performance is in strict attention to detail in every phase of the construction process: (a) initial selection of the structural form; (b) ductile design of members, particularly with respect to plastic hinge formation; (c) careful attention to connection strength and ductility; (d) frequent

interaction between the designer and fabricator to ensure proper fabrication procedures and tolerances; and (e) proper erection, along with extensive QA/QC for any field welding and bolting needed.

Cross-References

- [Seismic Strengthening Strategies for Existing \(Code-Deficient\) Ordinary Structures](#)

References

- AISC 358 (2010) Prequalified connections for special and intermediate moment frames for seismic applications. American Institute of Steel Construction, Chicago
- ASCE (1906) The effects of the San Francisco earthquake of April 18th, 1906, on Engineering Constructions: report of Committee on Fire and Earthquake Damage To Buildings. Trans ASCE 57(2):208–263
- Ashton TS (1968) Iron and steel in the industrial revolution. Augustus M. Kelley, New York
- Bodsworth C (ed) (2001) British iron and steel AD1800–2000 and beyond, Book 472. IOM Communications, London
- Clifton C et al (2011) Steel structures damage from the Christchurch earthquake of February 22, 2011. Bull N Z Soc Earthq Eng 44(4):297–318
- Condit CW (1968) American building. The University of Chicago Press, Chicago
- FEMA 355E (2000a) State of the art report on past performance of steel moment-frame buildings. FEMA, Washington, DC
- FEMA 355B (2000b) State of the art report on welding and inspection. FEMA, Washington, DC
- Halmos GT (2000) Roll forming handbook. CRC Press, Boca Raton
- Hamburger R, Meyer JD (2006) The performance of steel-frame buildings with infill masonry walls in the 1906 San Francisco earthquake. Earthq Spectra, EERI 22(2):S43–S68
- Kurzman D (2001) Disaster! The Great San Francisco earthquake and fire of 1906. Harper Perennial, New York
- Leon et al (2014) Performance of RC beam-column joints in the Christchurch 2010–2011 earthquakes. In: SP-296: symposium honoring James O. Jirsa's contributions in structural concrete: a time to reflect, CD. ACI, Farmington Hills
- Tylecote RF (1976) A history of metallurgy. Metals Society, London
- Viest et al (1997) Composite construction design for buildings. McGraw-Hill/ASCE, New York, 416 pp

Stochastic Analysis of Linear Systems

Giuseppe Muscolino

Dipartimento di Ingegneria Civile, Informatica, Edile, Ambientale e Matematica Applicata, Università degli Studi di Messina, Messina, Italy

Synonyms

Evolutionary frequency response function; Evolutionary power spectral density function; Gaussian zero-mean random models of seismic accelerations; Non-geometric spectral moments; Stochastic analysis

Introduction

The stochastic analysis of structural vibrations deals with the description and characterization of structural loads and responses that are modeled as stochastic processes. The probabilistic characterization of the input process could be extremely complex in time domain where the probability density functions depend on the autocorrelation functions which experimentally have to be specified over given set points. Since this approach is difficult to be used in applications, stochastic vibration analysis of structural linear systems subjected to Gaussian input processes is quite often performed in the frequency domain by means of the spectral analysis. This analysis is a very powerful tool for the analytical and experimental treatment of a large class of physical as well as structural problems subjected to random excitations. The main reasons are (a) the spectrum has an immediate physical interpretation as a power-frequency distribution; (b) the spectrum provides information on the stochastic structure of the process; and (c) the spectrum may be estimated by fairly simple numerical techniques which do not require any specific assumption of the structure of the process.

In the framework of earthquake engineering, the stationary non-white input models were suggested first. These models, which account for site properties and for the dominant frequency in ground motion, fail to reproduce the time-varying intensity typical of real earthquakes ground-motion accelerograms. In order to overcome this drawback, the so-called quasi-stationary (or uniformly modulated) random processes have been introduced (see, e.g., Shinozuka and Sato 1967; Jennings et al. 1969; Hsu and Bernard 1978). These processes are constructed as the product of a stationary zero-mean Gaussian random process by a deterministic function of time; for this reason they are also called separable nonstationary stochastic processes.

Furthermore, a time-varying frequency content is observed in actual accelerogram records. This nonstationary frequency is prevalently due to different arrival times of the primary, secondary, and surface waves that propagate at different velocities through the Earth's crust. To take into account both the simultaneous amplitude and frequency non-stationarity, Spanos and Solomos (1983) proposed a non-separable model introducing a particular *evolutionary power spectral density* (EPSD) function; Fan and Ahmadi (1990) proposed a generalization of the Kanai-Tajimi filter model with time-dependent coefficients; Conte and Peng (1997) defined the ground-motion accelerations as the sum of a finite number of pairwise independent uniformly modulated zero-mean Gaussian stochastic process, the so-called sigma-oscillatory process.

Once the problem is formulated from a mathematical point of view, the further step deals with the evaluation of the structural response to perform the prediction of the safety of structural systems. In this framework, the maximum absolute peak of stationary or nonstationary stochastic responses may be useful in design information of several engineering situations (see, e.g., Lin 1976; Lutes and Sarkani 2004; Muscolino and Palmeri 2005; Li and Chen 2009). Approximate procedures to calculate the statistics of the maximum absolute peak of the response have been proposed. These procedures lead to the probabilistic assessment of structural failure as a function of barrier

crossing rates, distribution of peaks, and extreme values. The latter quantities can be evaluated, for stationary input process, as a function of the well-known *geometric spectral moments* (GSMs) introduced by Vanmarcke (1972). For stationary stochastic response processes, the GSMs are defined as the geometric moments of the one-sided *power spectral density* (PSD) of the response process. Application of spectral methods to nonstationary random processes is more difficult than for the stationary ones; indeed for nonstationary processes the geometric approach fails (Di Paola 1985; Muscolino 1991). To perform the structural reliability in the latter cases, the so-called *nongeometric spectral moments* (NGSMs) have been introduced (Michaelov et al. 1999a, b).

In this study the before outlined topics will be addressed in order to evaluate the spectral characteristics of the structural response that are useful to perform the reliability assessment of linear systems subjected to stationary or nonstationary mono-/multi-correlated excitations.

Spectral Representation of Stochastic Processes

To carry out the spectral analysis, it is necessary to determine the spectral properties of the involved functions. These properties may be determined through the *Fourier-Stieltjes transform* (Priestley 1999).

Let us consider now a zero-mean stationary stochastic process, $F(t)$. This process is characterized by the feature that its *statistical moments* do not change over time and generally arise from any "stable" system which has achieved a "steady state." Moreover, the probabilistic structure of a stationary process is invariant under a shift of the time origin. This is a consequence of the fact that a sample of the process will almost certainly not "decay" to zero at infinity. Then the stationary processes possess infinite energy. Since a stationary process possesses infinite energy, its k th sample, $F^{(k)}(t)$, cannot be represented by the Fourier transform. In fact in this case, the Dirichlet condition is not satisfied. It follows that the spectral representation of a sample of a stationary

stochastic process can be performed only by the *Fourier-Stieltjes integral* (Priestley 1999):

$$F^{(k)}(t) = \int_{-\infty}^{+\infty} \exp(i\omega t) dN^{(k)}(\omega) \quad (1)$$

where $i = \sqrt{-1}$ is the imaginary unit and $N^{(k)}(\omega)$ is the k th sample of the complex stochastic process $N(\omega)$, satisfying the condition

$$E\langle dN(\omega_1) dN^*(\omega_2) \rangle = \delta(\omega_1 - \omega_2) S_{NN}(\omega_1) d\omega_1 d\omega_2 \quad (2)$$

where $\delta(\bullet)$ is the Dirac delta, the symbol $E\langle \bullet \rangle$ means stochastic average, and the asterisk indicates the complex conjugate quantity. Notice that in Eq. 2 sometimes the Kronecker delta is introduced instead of the Dirac delta; this is to avoid the inconsistency of this relationship for $\omega_1 = \omega_2$ (Spanos and Solomos 1983). The relationship (2) shows that the stochastic process $N(\omega)$ is a process with orthogonal increments, in the sense that its increments $dN(\omega_1)$ and $dN(\omega_2)$ at any two distinct points ω_1 and ω_2 are uncorrelated random variables. Furthermore, in Eq. 2 $S_{NN}(\omega)$, which is a real and symmetric function, $S_{NN}(-\omega) = S_{NN}(\omega)$, is the *power spectral density* (PSD) function of the process $N(\omega)$. According to the theory of stationary stochastic process, the autocorrelation function, $R_{FF}(\tau)$, of the zero-mean stationary stochastic process $F(t)$ is a real function given as (Lin 1976; Lutes and Sarkani 2004; Li and Chen 2009)

$$\begin{aligned} R_{FF}(\tau) &= E\langle F(t + \tau) F(t) \rangle \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp[i(\omega_1(t + \tau) - \omega_2 t)] \\ &\quad E\langle dN(\omega_1) dN^*(\omega_2) \rangle \end{aligned} \quad (3)$$

which in virtue of Eq. 2 leads to

$$R_{FF}(\tau) = E\langle F(t + \tau) F(t) \rangle = \int_{-\infty}^{\infty} \exp(i\omega\tau) S_{FF}(\omega) d\omega \quad (4)$$

In this equation $S_{FF}(\omega)$ is the PSD function of the stationary process $F(t)$.

In postulating the stationarity of the stochastic process, very strong assumptions regarding the structure of the process are made. Once these assumptions are dropped, the process can become nonstationary in many different ways. In the framework of the spectral analysis of nonstationary processes, Priestley (see, e.g., Priestley 1999) introduced the *evolutionary power spectral density* (EPDS) function. The EPDS function has essentially the same type of physical interpretation of the PSD function of stationary processes. The main difference is that whereas the PSD function describes the power-frequency distribution for the whole stationary process, the EPDS function is time dependent and describes the local power-frequency distribution at each instant time. The theory of EPDS function is the only one which preserves this physical interpretation for the nonstationary processes. Moreover, since the spectrum may be estimated by fairly simple numerical techniques, which do not require any specific assumption of the structure of the process, this model, based on the EPDS function, is nowadays the most adopted model for the analysis of structures subjected to nonstationary processes as the seismic motion due to earthquakes.

In the Priestley spectral representation of nonstationary processes, a sample of the nonstationary stochastic process is defined by the *Fourier-Stieltjes integral* as follows:

$$F^{(k)}(t) = \int_{-\infty}^{+\infty} \exp(i\omega t) a(\omega, t) dN^{(k)}(\omega) \quad (5)$$

where $a(\omega, t)$ is a slowly varying complex deterministic time-frequency modulating function which has to satisfy the condition $a(\omega, t) \equiv a^*(-\omega, t)$ and $N(\omega)$ is an orthogonal process satisfying the condition (2). In Eq. 2 $S_{NN}(\omega)$ is the PSD function of the so-called “embedded” stationary counterpart process, $N(\omega)$ (Michaelov et al. 1999a). It follows that the autocorrelation function of the zero-mean Gaussian nonstationary random process $F(t)$ can be obtained as

$$\begin{aligned}
 R_{FF}(t_1, t_2) &= E\langle F(t_1)F(t_2) \rangle \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \exp[i(\omega_1 t_1 - \omega_2 t_2)] a(\omega_1, t) \\
 &\quad a^*(\omega_2, t_2) E\langle dN(\omega_1) dN^*(\omega_2) \rangle d\omega_1 d\omega_2
 \end{aligned} \quad (6)$$

It is a real function which, in virtue of Eq. 2, leads to

$$\begin{aligned}
 R_{FF}(t_1, t_2) &= \int_{-\infty}^{\infty} \exp[i\omega(t_1 - t_2)] a(\omega, t_1) a^*(\omega, t_2) \\
 S_{NN}(\omega) d\omega &= \int_{-\infty}^{\infty} \exp[i\omega(t_1 - t_2)] S_{FF}(\omega, t_1, t_2) d\omega
 \end{aligned} \quad (7)$$

where

$$S_{FF}(\omega, t_1, t_2) = a(\omega, t_1) a^*(\omega, t_2) S_{NN}(\omega) \quad (8)$$

According to the Priestley evolutionary process model (Priestley 1999), the function

$$S_{FF}(\omega, t) = |a(\omega, t)|^2 S_{NN}(\omega) \quad (9)$$

is the so-called EPSPD function of the nonstationary process $F(t)$. In the previous equations the symbol $|\cdot|$ denotes the modulus of the function in brackets. The processes characterized by the EPSPD function $S_{FF}(\omega, t)$ are called *fully nonstationary* or *non-separable* random process, since both time and frequency content change, and they cannot be decoupled. If the modulating function is a time-dependent function, $a(\omega, t) \equiv a(t)$, the nonstationary process is called *quasi-stationary* or *uniformly modulated* or *separable* random process. In this case the time content change is independent by the frequency content change; indeed, the EPSPD function assumes the following expression:

$$S_{FF}(\omega, t) = a^2(t) S_{NN}(\omega) \quad (10)$$

In the stochastic analysis the one-sided PSD is generally used; the latter can be suitably defined in the Priestley representation by the following equation:

$$G_{FF}(\omega, t_1, t_2) = \begin{cases} a(\omega, t_1) a^*(\omega, t_2) G_{NN}(\omega) \equiv 2S_{FF}(\omega, t_1, t_2), & \omega \geq 0; \\ 0, & \omega < 0 \end{cases} \quad (11)$$

where $G_{NN}(\omega)$ ($G_{NN}(\omega) = 2S_{NN}(\omega)$, $\omega \geq 0$; $G_{NN}(\omega) = 0$, $\omega < 0$) is the one-sided PSD function of the stationary counterpart of the input process $F(t)$. In this case the autocorrelation function of the process $F(t)$ is given by the following relationship:

$$\begin{aligned}
 \bar{R}_{FF}(t_1, t_2) &= \int_{-\infty}^{\infty} \exp[i\omega(t_1 - t_2)] a(\omega, t_1) \\
 &\quad a^*(\omega, t_2) G_{NN}(\omega) d\omega
 \end{aligned} \quad (12)$$

Note that since the one-sided PSD function $G_{NN}(\omega)$ is not symmetric, the corresponding autocorrelation function, $\bar{R}_{FF}(t_1, t_2)$, is a complex function (Di Paola 1985), whose real part coincides with the function defined in Eq. 6:

$\text{Re}\{\bar{R}_{FF}(t_1, t_2)\} \equiv R_{FF}(t_1, t_2)$. It can be easily proved that the complex function (12) is the autocorrelation function of a complex process $\bar{F}(t)$ defined as (Di Paola and Petrucci 1990)

$$\bar{F}(t) = \sqrt{2} \int_0^{+\infty} \exp(i\omega t) a(\omega, t) dN(\omega) \quad (13)$$

The real part of $\bar{F}(t)$ is proportional to the process $F(t)$, while the imaginary part of $\bar{F}(t)$ is a nonstationary process having stationary counterpart proportional to Hilbert transform of the real part of the stationary counterpart of the process itself (Di Paola 1985; Di Paola and Petrucci 1990; Muscolino 1991). The complex process, $\bar{F}(t)$, which generates the complex

autocorrelation function (12), has been called *pre-envelope process* by Di Paola (1985).

Reliability of Linear Structural Systems Subjected to Stochastic Excitations

The structural systems are conceived and designed to survive to natural actions. If the excitations are modeled as random processes, the dynamic responses are random processes too, and the structural safety needs to be evaluated in a probabilistic sense. Among the models of failure, the simplest one, which is also the most widely used in practical analyses, is based on the assumption that a structure fails as soon as the response at a critical location exits a prescribed safe domain for the first time. The probability of failure, in this case, coincides with the first passage probability, i.e., the probability that the absolute value of the random response process $X(t)$ of a selected structural response (e.g., strain or stress at a critical point) will exceed a specified safety bound, b , within a specified time interval (Lin 1976). In random vibration theory, the problem of probabilistically predicting this event is termed *first passage problem*. Unfortunately, this is one of the most complicated problems in computational stochastic mechanics. The solution of this problem has not been derived in exact form, even in the simplest case of the stationary response of a single-degree-of-freedom (SDoF) linear oscillator under zero-mean Gaussian white noise (Lin 1976; Lutes and Sarkani 2004; Muscolino and Palmeri 2005). Hence, a large number of approximated techniques have been proposed in literature, which differ in generality, complexity, and accuracy. In the framework of approximate methods, the time-dependent reliability of the structure, based on the first passage failure criterion, can be expressed, for a symmetric barrier, as (Lutes and Sarkani 2004)

$$L_{|X|}(b, t) = L_{|X|}(b, 0) \exp \left[- \int_0^t \eta_X(b, \rho) d\rho \right] \quad (14)$$

where $\eta_X(b, t)$ is the so-called *hazard function* and $L_{|X|}(b, 0)$ is the reliability at time $t = 0$, which for the nonstationary case can be assumed as unity.

For narrow-band zero-mean stationary Gaussian process, the hazard function has been derived by Vanmarcke (1972) as

$$\eta_X(b) = \frac{1}{\pi} \sqrt{\frac{\lambda_{2,X}}{\lambda_{0,X}}} \left[\frac{1 - \exp \left(-b \delta_X^{1.2} \sqrt{\frac{\pi}{2\lambda_{0,X}}} \right)}{\exp \left(\frac{b^2}{2\lambda_{0,X}} \right) - 1} \right] \quad (15)$$

with

$$\delta_X = \sqrt{1 - \frac{\lambda_{1,X}^2}{\lambda_{0,X} \lambda_{2,X}}} \quad (16)$$

In this equation δ_X is the so-called *bandwidth parameter* of the process $X(t)$ (Vanmarcke 1972, 1975). This parameter measures the variation of the narrowness of the stochastic process $X(t)$. Usually, a stochastic process with $0 \leq \delta_X \leq 0.35$ is called a narrow-band stochastic process. Finally, in the previous equations $\lambda_{i,X}$ ($i = 0, 1, 2$) are the so-called *geometric spectral moments* (GSMs), introduced by Vanmarcke (1972) as the geometric moments of the one-sided PSD of the response process:

$$\lambda_{i,X} \equiv \lambda_{i,X}^G = \int_0^\infty \omega^i G_{XX}(\omega) d\omega \quad (17)$$

In this equation the apex G emphasizes the geometric evaluation of the GSMs. Notice that $\lambda_{0,X}$ coincides with the variance of the zero-mean process $X(t)$, $\lambda_{0,X} \equiv \sigma_X^2 = E\langle X^2(t) \rangle$; $\lambda_{2,X}$ coincides with the variance of the zero-mean process $\dot{X}(t)$, $\lambda_{2,X} \equiv \sigma_{\dot{X}}^2 = E\langle \dot{X}^2(t) \rangle$; while $\lambda_{1,X}$ does not coincide with the cross-covariance of the processes $X(t)$ and $\dot{X}(t)$.

For narrow-band zero-mean nonstationary Gaussian process, the hazard function has been derived by Corotis et al. (1972) as

$$\eta_X(b, t) = \frac{1}{\pi} \sqrt{\frac{\lambda_{2,X}(t)}{\lambda_{0,X}(t)}} \left[\frac{1 - \exp\left(-b \delta_X^{1+d}(t) \sqrt{\frac{\pi}{2\lambda_{0,X}(t)}}\right)}{\exp\left(\frac{b^2}{2\lambda_{0,X}(t)}\right) - 1} \right] \quad (18)$$

where d ($d = 0$ or $d = 0.2$) is an empirical parameter and $\delta_X(t)$ is the bandwidth parameter which in the nonstationary case is time dependent. Because of the non-stationarity of random process, this parameter involves complex functions, and it is defined by Michaelov et al. (1999a, b) as

$$\delta_X(t) = \sqrt{1 - \frac{\text{Re}\{\lambda_{1,X}(t)\}^2}{\lambda_{0,X}(t)\lambda_{2,X}(t)}} \quad (19)$$

The evaluation of the time-dependent quantities $\lambda_{i,X}(t)$ is conceptually more complicated than for stationary processes; indeed for these processes the geometric approach fails for $i = 1$ and $i = 2$ (Di Paola 1985; Di Paola and Petrucci 1990; Muscolino 1991; Michaelov et al. 1999a, b), that is

$$\begin{aligned} \lambda_{i,X}(t) &\neq \lambda_{i,X}^G(t) \\ &= \int_0^\infty \omega^i G_{XX}(\omega, t) d\omega, \quad i = 1, 2 \end{aligned} \quad (20)$$

The physical inconsistency on the evaluation of the geometric GSMs, in the nonstationary case, as the moments of the one-sided EPSD function was pointed out by Corotis et al. (1972). In fact they discovered that for the case of the transient response of an oscillator subjected to stationary Gaussian white noise processes, the second GSM does not exist because it is unbounded. At same time in the stationary case, this GSM, which is the limit of the transient as the time approaches infinity, is finite. The first that considered the problem of spectral characteristics from a nongeometric point of view was Di Paola (1985). The basic idea

of this approach is to establish a time-domain interpretation of the SM. In order to do this, the pre-envelope covariances as the covariances of structural systems subjected to a complex-valued random process (the so-called pre-envelope process) have been introduced (Di Paola and Petrucci 1990). The real part of this process is proportional to the original nonstationary process, while the imaginary part is an auxiliary random process related to the real part in such a way that the complex process exhibits power in the positive frequency range only. Since the use of complex pre-envelope process is not very intuitive, Michaelov et al. (1999a, b) evaluated the pre-envelope covariances as a function of the EPSD of the response and recalled them as *nongeometric spectral moments* (NGSMs). It has to be emphasized that the NGSMs contain more information than the “conventional” covariances. Indeed, the NGSMs have been proved to be more appropriate for describing nonstationary process and can be effectively employed in structural reliability applications (Di Paola 1985; Di Paola and Petrucci 1990; Muscolino 1991; Michaelov et al. 1999a, b).

It has to be emphasized that in the framework of nonstationary analysis of structures, other time-dependent parameters, very useful in describing the time-variant spectral properties of the stochastic process, are (i) the mean frequency, $v_X^+(t)$, which evaluate the variation in time of the mean up-crossing rate of the time axis, and (ii) the central frequency, $\omega_{C,X}(t)$, which scrutinizes the variation of the frequency content of the stochastic process with respect to time. The two functions introduced before can be evaluated as a function of NGSMs and have been defined, respectively, as (Michaelov et al. 1999a, b)

$$v_X^+(t) = \frac{1}{2\pi} \sqrt{\frac{\lambda_{2,X}(t)}{\lambda_{0,X}(t)}}; \quad \omega_{C,X}(t) = \frac{\text{Re}\{\lambda_{1,X}(t)\}}{\lambda_{0,X}(t)} \quad (21)$$

In stationary case the mean frequency, v_X^+ , and the central frequency, $\omega_{C,X}$, are not time dependent. Moreover, in the latter case, because of the $\text{Im}\{\lambda_{1,X}\} = 0$, it follows $\text{Re}\{\lambda_{1,X}\} \equiv \lambda_{1,X}$.

Response of Single-Degree-of-Freedom (SDoF) Oscillators

Fundamental of Deterministic Analysis

The theory of deterministic linear systems plays a fundamental role in the dynamic analysis of structures subjected to stochastic excitations. For this reason in this section the fundamental of deterministic analysis of SDoF subjected to deterministic excitation is synthetically reviewed. Particular care has been devoted to the state-space approach. This approach is the best suited for the development of formulations in the framework of random vibrations. In fact, its adaptability to numerical method of solution of differential equations and its extension to multi-degree-of-freedom (MDoF) systems are very straightforward.

The equation of motion of a linear oscillator with mass, m ; viscous damping, c ; stiffness, k ; and subjected to the excitation $f(t)$ and at rest at initial time of motion, $t = t_0$, can be written as

$$\begin{aligned} m \ddot{u}(t) + c \dot{u}(t) + k u(t) &= f(t); \\ u(t_0) &= 0, \quad \dot{u}(t_0) = 0 \end{aligned} \quad (22)$$

or in canonical form as follows:

$$\begin{aligned} \ddot{u}(t) + 2\xi_0\omega_0 \dot{u}(t) + \omega_0^2 u(t) &= F(t); \\ u(t_0) &= 0, \quad \dot{u}(t_0) = 0 \end{aligned} \quad (23)$$

where $u(t)$ is the displacement response of the mass, $\omega_0 = \sqrt{k/m}$ is the natural circular frequency, $\xi_0 = c/2\sqrt{mk}$ is the damping ratio, and $F(t) = f(t)/m$; a dot over a variable denotes differentiation with respect to time t . In state variables the equation of motion of the oscillator, in canonical form, can be written as a set of two first-order differential equations:

$$\dot{\mathbf{y}}(t) = \mathbf{D}_0 \mathbf{y}(t) + \mathbf{v}_0 F(t); \quad \mathbf{y}(t_0) = \mathbf{0} \quad (24)$$

where

$$\mathbf{y}(t) = \begin{bmatrix} u(t) \\ \dot{u}(t) \end{bmatrix}, \quad \mathbf{D}_0 = \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -2\xi_0\omega_0 \end{bmatrix}, \quad \mathbf{v}_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (25)$$

Denoting by $\mathbf{y}_p(t)$ the particular solution vector, the solution of Eq. 24 can be evaluated as (Borino and Muscolino 1986)

$$\mathbf{y}(t) = \mathbf{y}_p(t) + \mathbf{\Theta}_0(t - t_0) [\mathbf{y}(t_0) - \mathbf{y}_p(t_0)] \quad (26)$$

where $\mathbf{\Theta}_0(t)$ is the so-called transition matrix:

$$\mathbf{\Theta}_0(t) = \exp(\mathbf{D}_0 t) = \begin{bmatrix} -\omega_0^2 g_0(t) & h_0(t) \\ -\omega_0^2 h_0(t) & \dot{h}_0(t) \end{bmatrix} \quad (27)$$

with

$$\begin{aligned} g_0(t) &= -\frac{1}{\omega_0^2} \exp(-\xi_0\omega_0 t) \left[\cos(\bar{\omega}_0 t) + \frac{\xi_0\omega_0}{\bar{\omega}_0} \sin(\bar{\omega}_0 t) \right]; \\ h_0(t) &= \dot{g}_0(t) = \frac{1}{\omega_0} \exp(-\xi_0\omega_0 t) \sin(\bar{\omega}_0 t); \\ \dot{h}_0(t) &= \exp(-\xi_0\omega_0 t) \left[\cos(\bar{\omega}_0 t) - \frac{\xi_0\omega_0}{\bar{\omega}_0} \sin(\bar{\omega}_0 t) \right] \end{aligned} \quad (28)$$

and $\bar{\omega}_0 = \omega_0 \sqrt{1 - \xi_0^2}$ is the damped natural circular frequency. Notice that the contribution of the last term in the right member of Eq. 26 decreases in the time because the transition matrix satisfies the following condition:

$$\lim_{t \rightarrow \infty} \mathbf{\Theta}_0(t) = \mathbf{0} \quad (29)$$

Alternatively, by applying the so-called *parameter variation method*, the solution of Eq. 24 can be written in integral form as follows:

$$\mathbf{y}(t) = \mathbf{\Theta}_0(t - t_0) \mathbf{y}(t_0) + \int_{t_0}^t \mathbf{\Theta}_0(t - \tau) \mathbf{v} F(\tau) d\tau \quad (30)$$

For quiescent systems, because the relationship $\mathbf{y}(t_0) = \mathbf{0}$ is satisfied, Eqs. 26 and 30 become

$$\begin{aligned} \mathbf{y}(t) &= \mathbf{y}_p(t) - \mathbf{\Theta}_0(t - t_0) \mathbf{y}_p(t_0) \\ &= \int_{t_0}^t \mathbf{\Theta}_0(t - \tau) \mathbf{v} F(\tau) d\tau \end{aligned} \quad (31)$$

Notice that in this case the first element of vector $\mathbf{y}(t)$, written in integral form, coincides with the well-known Duhamel integral. Starting by the integral form solution in state variables,

it is possible to derive a very powerful unconditionally stable numerical procedure for the evaluation of the structural response (see, e.g., Borino and Muscolino 1986).

Stochastic Response

Mathematically strictly speaking, as a consequence of the introduction of the

one-sided PSD, the input process is a complex one. It follows that the response processes, $u(t)$, is a complex function too. After some algebra, for the quiescent oscillator (Eq. 23), the NGSMs can be evaluated, in time domain, as (Di Paola 1985; Di Paola and Petrucci 1990, Muscolino 1991)

$$\begin{aligned}\lambda_{0,uu}(t) &\equiv E\langle u(t)u^*(t) \rangle = \iint\limits_{00}^t h(t-\tau_1)h(t-\tau_2)\bar{R}_{FF}(\tau_1, \tau_2) d\tau_1 d\tau_2; \\ \lambda_{1,uu}(t) &= -i \iint\limits_{00}^t h(t-\tau_1)\dot{h}(t-\tau_2)\bar{R}_{FF}(\tau_1, \tau_2) d\tau_1 d\tau_2; \\ \lambda_{2,uu}(t) &\equiv E\langle \dot{u}(t)\dot{u}^*(t) \rangle = \iint\limits_{00}^t \dot{h}(t-\tau_1)\dot{h}(t-\tau_2)\bar{R}_{FF}(\tau_1, \tau_2) d\tau_1 d\tau_2\end{aligned}\quad (32)$$

where $\bar{R}_{FF}(t_1, t_2)$ is the complex autocorrelation function defined in Eq. 12. Moreover, the presence of the imaginary unit in the second of Eq. 32 inverts the roles of the real and imaginary parts of $\lambda_{1,uu}(t)$ with respect to the variances $\lambda_{0,uu}(t)$ and $\lambda_{2,uu}(t)$; furthermore, while $\lambda_{0,uu}(t)$ and $\lambda_{2,uu}(t)$ are real functions, $\lambda_{1,uu}(t)$ is a complex one. Substituting Eq. 12 into Eq. 32, the following relationships are obtained:

$$\begin{aligned}\lambda_{0,uu}(t) &= \int_0^\infty Z_0^*(\omega, t)Z_0(\omega, t)G_{NN}(\omega)d\omega; \\ \lambda_{1,uu}(t) &= -i \int_0^\infty Z_0^*(\omega, t)\dot{Z}_0(\omega, t)G_{NN}(\omega)d\omega; \\ \lambda_{2,uu}(t) &= \int_0^\infty \dot{Z}_0^*(\omega, t)\dot{Z}_0(\omega, t)G_{NN}(\omega)d\omega\end{aligned}\quad (33)$$

where

$$\begin{aligned}Z_0(\omega, t) &= \int_0^t h(t-\tau)\exp(i\omega\tau)a(\omega, \tau) d\tau; \\ \dot{Z}_0(\omega, t) &= \int_0^t \dot{h}(t-\tau)\exp(i\omega\tau)a(\omega, \tau) d\tau\end{aligned}\quad (34)$$

Notice that the function $Z_0(\omega, t)$ is the so-called *evolutionary frequency response function* of the oscillator (Li and Chen 2009). Remarkably, since the integrals (Eq. 34) are convolution integrals of Duhamel's type, they can be interpreted as the response, in terms of state variables, of the quiescent oscillator, at time $t = 0$, subjected to the deterministic complex function $f(\omega, t) = \exp(i\omega t) a(\omega, t)$. By introducing the state variables, the *evolutionary frequency response vector function* can be defined as

$$\mathbf{Y}_0(\omega, t) = \begin{bmatrix} Z_0(\omega, t) \\ \dot{Z}_0(\omega, t) \end{bmatrix}\quad (35)$$

It follows that relationships (33) can be rewritten in compact form as follows:

$$\begin{aligned}\Sigma_{uu}(t) &= \begin{bmatrix} \lambda_{0,uu}(t) & i\lambda_{1,uu}(t) \\ -i\lambda_{1,uu}^*(t) & \lambda_{2,uu}(t) \end{bmatrix} \\ &\equiv \int_0^\infty G_{NN}(\omega) \mathbf{Y}_0^*(\omega, t)\mathbf{Y}_0^T(\omega, t)d\omega\end{aligned}\quad (36)$$

This matrix coincides with the so-called *pre-envelope covariance matrix* introduced by Di Paola and Petrucci (1990) which is a complex

matrix. As a conclusion, for input processes characterized by one-sided EPSP function, the cross-covariance matrix (Eq. 36) of an oscillator is a complex matrix whose elements are the NGSMs.

Generalizing Eq. 32, it can be easily proved that the complex cross-correlation function matrix of the zero-mean response process, which collects the cross-correlation of the pre-envelope response processes, can be evaluated as

$$\mathbf{R}_{uu}(t_1, t_2) = \int_0^\infty G_{NN}(\omega) \mathbf{Y}_0^*(\omega, t_1) \mathbf{Y}_0^T(\omega, t_2) d\omega \quad (37)$$

Finally, according to the Priestley evolution-ary process model (Priestley 1999), this complex function matrix $\mathbf{R}_{uu}(t_1, t_2)$ can be also rewritten as

$$\mathbf{R}_{uu}(t_1, t_2) = \int_0^\infty \exp[i\omega(t_1 - t_2)] \mathbf{G}_{uu}(\omega, t_1, t_2) d\omega \quad (38)$$

where

$$\mathbf{G}_{uu}(\omega, t_1, t_2) = G_{NN}(\omega) \mathbf{Y}_0^*(\omega, t_1) \mathbf{Y}_0^T(\omega, t_2) \quad (39)$$

Consequently the NGSMs of the oscillator response are the elements of the pre-envelope covariance matrix, given as

$$\Sigma_{uu}(t) \equiv \mathbf{R}_{uu}(t, t) \equiv \int_0^\infty \mathbf{G}_{uu}(\omega, t) d\omega \quad (40)$$

Let us assume now the modulating function $a(\omega, t) = U(t)$. Starting from the nonstationary formulation, it is possible to deduce the formulation in the case of stationary input. This result is obtained by performing the limit as $t \rightarrow \infty$ into Eq. 34:

$$\begin{aligned} \lim_{t \rightarrow \infty} Z_0(\omega, t) &= \int_0^\infty h_0(t - \tau) \exp(i\omega\tau) d\tau = \exp(i\omega t) H_0(\omega); \\ \lim_{t \rightarrow \infty} \dot{Z}_0(\omega, t) &= \int_0^\infty \dot{h}_0(t - \tau) \exp(i\omega\tau) d\tau = i\omega \exp(i\omega t) H_0(\omega) \end{aligned} \quad (41)$$

where $h_0(t)$ and $\dot{h}_0(t)$ are the functions defined in Eq. 28 and $H_0(\omega)$ is the *frequency response function* of the oscillator defined as

$$\begin{aligned} H_0(\omega) &= \int_0^\infty h_0(\tau) \exp(-i\omega\tau) d\tau \\ &= \frac{1}{\omega_0^2 - \omega^2 + i2\xi_0\omega\omega_0} \end{aligned} \quad (42)$$

By substituting Eq. 41 into Eq. 35 and the results into Eq. 36, the following relationship is obtained:

$$\begin{aligned} \Sigma_{uu} &= \int_0^\infty G_{NN}(\omega) \mathbf{H}_0^*(\omega) \mathbf{H}_0^T(\omega) d\omega \\ &\equiv \begin{bmatrix} \lambda_{0,uu} & i\lambda_{1,uu} \\ -i\lambda_{1,uu}^* & \lambda_{2,uu} \end{bmatrix} \end{aligned} \quad (43)$$

where

$$\mathbf{H}_0(\omega) = \begin{bmatrix} H_0(\omega) \\ i\omega H_0(\omega) \end{bmatrix} \quad (44)$$

The elements of this matrix Σ_{uu} are the GSMs, introduced by Vanmarcke (1972), that is

$$\begin{aligned} \lambda_{i,uu} &\equiv \lambda_{i,uu}^G \\ &= \int_0^\infty \omega^i |H_0(\omega)|^2 G_{NN}(\omega) d\omega; \quad i = 0, 1, 2 \end{aligned} \quad (45)$$

Response of Multi-degree-of-Freedom (MDOF) Systems

Fundamental of Deterministic Analysis

Let us consider the equation of motion of a linear quiescent n -degree-of-freedom (n -DoF) classically

damped structural system whose dynamic behavior is governed by the equation of motion:

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{f}(t) \quad (46)$$

where $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are the $(n \times n)$ mass, damping, and stiffness matrices of the structure; $\mathbf{u}(t)$ is the $(n \times 1)$ vector of displacements, having for i th element $u_i(t)$; and $\mathbf{f}(t)$ denotes the external load vector. Under the assumption of classically damped system, the equation of motion can be decoupled by applying the modal analysis. To this aim let us introduce the modal coordinate transformation:

$$\begin{aligned} \mathbf{u}(t) &= \Phi \mathbf{q}(t) = \sum_{j=1}^m \phi_j q_j(t) \Rightarrow u_i(t) \\ &= \sum_{j=1}^m \phi_{ij} q_j(t) \end{aligned} \quad (47)$$

In this equation, $\Phi = [\phi_1 \ \phi_2 \ \dots \ \phi_m]$ is the modal matrix, of order $n \times m$, collecting the m eigenvectors ϕ_j , normalized with respect to the mass matrix $\mathbf{M}$, solutions of the following eigenproblem:

$$\mathbf{K}^{-1}\mathbf{M}\Phi = \Phi\Omega^{-2}; \quad \Phi^T\mathbf{M}\Phi = \mathbf{I}_m \quad (48)$$

In this equation Ω is a diagonal matrix listing the undamped natural circular frequency ω_j , $\mathbf{I}_m$ is the identity matrix of order m , and the apex T means transpose operator. Once the modal matrix Φ is evaluated, by applying the coordinate transformations (47) to Eq. 46, the following set of decoupled second-order differential equations is obtained:

$$\ddot{\mathbf{q}}(t) + \Xi \dot{\mathbf{q}}(t) + \Omega^2 \mathbf{q}(t) = \Phi^T \mathbf{f}(t) \quad (49)$$

in which Ξ is a generalized damping matrix given by

$$\Xi = \Phi^T \mathbf{C} \Phi \quad (50)$$

For classically damped structures the modal damping matrix Ξ is a diagonal matrix listing the

quantities $2\zeta_j\omega_j$, ζ_j being the modal damping ratio. It follows that the j th differential Eq. 49 can be written as

$$\ddot{q}_j(t) + 2\zeta_j\omega_j \dot{q}_j(t) + \omega_j^2 q_j(t) = \phi_j^T \mathbf{f}(t) \quad (51)$$

In the state space, Eq. 51 can be written in the first-order form as

$$\dot{\mathbf{y}}_j(t) = \mathbf{D}_j \mathbf{y}_j(t) + \mathbf{V}_j \mathbf{f}(t) \quad (52)$$

where

$$\begin{aligned} \mathbf{y}_j(t) &= \begin{bmatrix} q_j(t) \\ \dot{q}_j(t) \end{bmatrix}; \quad \mathbf{D}_j = \begin{bmatrix} 0 & 1 \\ -\omega_j^2 & -2\zeta_j\omega_j \end{bmatrix}; \\ \mathbf{V}_j &= \begin{bmatrix} \mathbf{0} \\ \phi_j^T \end{bmatrix} \end{aligned} \quad (53)$$

Stochastic Response for Mono-correlated Stochastic Input Processes

Let us assume now that the forcing term is a mono-correlated zero-mean Gaussian random process vector given by the relationship:

$$\mathbf{f}(t) = \mathbf{b}\bar{F}(t) \quad (54)$$

where $\mathbf{b}$ is the $(n \times 1)$ vector of spatial distribution of loads and $\bar{F}(t)$ is a zero-mean Gaussian nonstationary random process. It follows that the j th differential Eq. 51 can be written as

$$\ddot{q}_j(t) + 2\zeta_j\omega_j \dot{q}_j(t) + \omega_j^2 q_j(t) = p_j \bar{F}(t), \quad j = 1, 2, \dots, m \quad (55)$$

where

$$p_j = \phi_j^T \mathbf{b} \quad (56)$$

After very simple algebra it can be shown that the j th NGSMs, $\lambda_{j,u_i}(t)$ ($i = 0, 1, 2$), of the i th nodal displacement response, $u_i(t)$, are given as a function of modal NGSMs, $p_j \lambda_{j,k\ell}(t)$ ($\ell, k = 1 \dots, m$), by the following relationships:

$$\begin{aligned}
\lambda_{0,u_i u_i}(t) &= \sum_{k=1}^m \sum_{\ell=1}^m p_k p_\ell \phi_{ik} \phi_{i\ell} \lambda_{0,k\ell}(t); \\
\lambda_{1,u_i u_i}(t) &= \sum_{k=1}^m \sum_{\ell=1}^m p_k p_\ell \phi_{ik} \phi_{i\ell} \lambda_{1,k\ell}(t); \quad (57) \\
\lambda_{2,u_i u_i}(t) &= \sum_{k=1}^m \sum_{\ell=1}^m p_k p_\ell \phi_{ik} \phi_{i\ell} \lambda_{2,k\ell}(t)
\end{aligned}$$

It has to be emphasized that the zeroth NGSM, $\lambda_{0,u_i u_i}(t)$, and the second-order NGSM, $\lambda_{2,u_i u_i}(t)$, are real functions that coincide with the variances of the response in terms of displacement and velocity, respectively, while the first-order NGSM, $\lambda_{1,u_i u_i}(t)$, is a complex quantity whose real part can be evaluated as the cross-covariance

between the response process and the response velocity process of the same linear system subjected to a nonstationary input whose stationary counterpart is proportional to its Hilbert transform (Di Paola 1985; Langley 1986; Di Paola and Petrucci 1990; Muscolino 1991). As shown in Eq. 57 the nodal NGSMs can be evaluated as a function of $\lambda_{r,k\ell}(t)$, $r = 0, 1, 2$, which are the so-called time-dependent modal NGSMs “purged” by p_j factors. After some algebra, these quantities, which are complex ones, can be evaluated, in time domain, for quiescent structural systems as (Di Paola 1985; Di Paola and Petrucci 1990; Muscolino 1991)

$$\begin{aligned}
\lambda_{0,k\ell}(t) &= \int_0^t \int_0^t h_k(t-\tau_1) h_\ell(t-\tau_2) \bar{R}_{FF}(\tau_1, \tau_2) d\tau_1 d\tau_2; \\
\lambda_{1,k\ell}(t) &= -i \int_0^t \int_0^t h_k(t-\tau_1) \dot{h}_\ell(t-\tau_2) \bar{R}_{FF}(\tau_1, \tau_2) d\tau_1 d\tau_2; \quad k = 1, 2, \dots, m; \quad \ell = 1, 2, \dots, m; \\
\lambda_{2,k\ell}(t) &= \int_0^t \int_0^t \dot{h}_k(t-\tau_1) \dot{h}_\ell(t-\tau_2) \bar{R}_{FF}(\tau_1, \tau_2) d\tau_1 d\tau_2
\end{aligned} \quad (58)$$

where $\bar{R}_{FF}(t_1, t_2)$ is the complex autocorrelation function defined in Eq. 12. Notice that the presence of the imaginary unit in the second term of the second of Eq. 58 inverts the roles of the real and imaginary parts of $\lambda_{1,k\ell}(t)$ with respect to the cross-covariance $\lambda_{0,k\ell}(t)$ and $\lambda_{2,k\ell}(t)$; furthermore for $k = \ell$ while $\lambda_{0,k\ell}(t)$ and $\lambda_{2,k\ell}(t)$ become real quantities, $\lambda_{1,k\ell}(t)$ remains a complex one.

It can be easily proved that the “purged” NGSMs, given in Eq. 58, can be evaluated as a function of the statistics of the response of a dummy oscillator whose motion is governed by

$$\ddot{z}_j(t) + 2\zeta_j \omega_j \dot{z}_j(t) + \omega_j^2 z_j(t) = \bar{F}(t) \quad (59)$$

where $\bar{F}(t)$ is a complex process whose imaginary part is a process having stationary counterpart proportional to Hilbert transform of the real part of the stationary counterpart of the complex

process itself; it follows that $z_j(t) = q_j(t)/p_j$ is a complex process too. Notice that, with the position $p_j = 1$, Eq. 59 coincides with Eq. 55. Then by substituting the complex function $\bar{R}_{FF}(t_1, t_2)$, defined in Eq. 12, into Eq. 58, after very simple algebra, it is possible to evaluate the “purged” NGSMs as

$$\begin{aligned}
\lambda_{0,k\ell}(t) &= \int_0^\infty Z_k^*(\omega, t) Z_\ell(\omega, t) G_{NN}(\omega) d\omega; \\
\lambda_{1,k\ell}(t) &= -i \int_0^\infty Z_k^*(\omega, t) \dot{Z}_\ell(\omega, t) G_{NN}(\omega) d\omega; \\
\lambda_{2,k\ell}(t) &= \int_0^\infty \dot{Z}_k^*(\omega, t) \dot{Z}_\ell(\omega, t) G_{NN}(\omega) d\omega
\end{aligned} \quad (60)$$

where, for $j = k, \ell$, the following positions have been made:

$$\begin{aligned} Z_j(\omega, t) &= \int_0^t h_j(t - \tau) \exp(i\omega\tau) a(\omega, \tau) d\tau; \\ \dot{Z}_j(\omega, t) &= \int_0^t \dot{h}_j(t - \tau) \exp(i\omega\tau) a(\omega, \tau) d\tau \end{aligned} \quad (61)$$

with

$$h_j(t) = \frac{1}{\bar{\omega}_j} \exp(-\zeta_j \omega_j t) \sin(\bar{\omega}_j t) \quad (62)$$

and $\bar{\omega}_j = \omega_j \sqrt{1 - \zeta_j^2}$ ($j = k, \ell$) is the damped circular frequency of the j th dummy oscillator. Moreover, introducing the state variable, the *modal evolutionary frequency response function vector* is defined as

$$\mathbf{Y}_j(\omega, t) = \begin{bmatrix} Z_j(\omega, t) \\ \dot{Z}_j(\omega, t) \end{bmatrix}; \quad j = k, \ell \quad (63)$$

Then, relationships (33) can be rewritten in compact form as follows:

$$\begin{aligned} \Sigma_{k\ell}(t) &= \begin{bmatrix} \lambda_{0,k\ell}(t) & i\lambda_{1,k\ell}(t) \\ -i\lambda_{1,k\ell}^*(t) & \lambda_{2,k\ell}(t) \end{bmatrix} \\ &\equiv \int_0^\infty G_{NN}(\omega) \mathbf{Y}_k^*(\omega, t) \mathbf{Y}_\ell^T(\omega, t) d\omega \end{aligned} \quad (64)$$

which coincides with the cross-modal *pre-envelope covariance matrix* which is a complex matrix (Di Paola and Petrucci 1990). As a conclusion for input processes characterized by one-sided PSD function, the cross-covariance matrix (Eq. 64) between the k, ℓ th dummy oscillators is a complex matrix whose elements are the “purged” NGSMs. It can be easily proved that the complex cross-correlation function matrix of the zero-mean modal “purged” response process, which collects the cross-correlation of the

pre-envelope response processes, according to the Priestley evolutionary process model (Priestley 1999), can be evaluated as

$$\begin{aligned} \mathbf{R}_{k\ell}(t_1, t_2) &= \int_0^\infty G_{NN}(\omega) \mathbf{Y}_k^*(\omega, t_1) \mathbf{Y}_\ell^T(\omega, t_2) d\omega \\ &\equiv \int_0^\infty \exp[i\omega(t_1 - t_2)] \mathbf{G}_{k\ell}(\omega, t_1, t_2) d\omega \end{aligned} \quad (65)$$

where

$$\mathbf{G}_{k\ell}(\omega, t_1, t_2) = G_{NN}(\omega) \mathbf{Y}_k^*(\omega, t_1) \mathbf{Y}_\ell^T(\omega, t_2) \quad (66)$$

is the cross-modal EPSP function matrix between the k, ℓ th dummy oscillators. By means of the modal transformation (Eq. 47), the nodal autocorrelation and the EPSP function matrices of the displacement response, $u_i(t)$, can be evaluated, after very simple algebra, respectively, as follows:

$$\begin{aligned} \mathbf{R}_{u_i u_i}(t_1, t_2) &= \sum_{k=1}^m \sum_{\ell=1}^m p_k p_\ell \phi_{ik} \phi_{i\ell} \mathbf{R}_{k\ell}(t_1, t_2), \\ \mathbf{G}_{u_i u_i}(\omega, t_1, t_2) &= \sum_{k=1}^m \sum_{\ell=1}^m p_k p_\ell \phi_{ik} \phi_{i\ell} \mathbf{G}_{k\ell}(\omega, t_1, t_2) \end{aligned} \quad (67)$$

Consequently the nodal displacement NGSMs are the elements of the nodal pre-envelope covariance matrix, given as

$$\begin{aligned} \Sigma_{u_i u_i}(t) &\equiv \mathbf{R}_{u_i u_i}(t, t) \equiv \int_0^\infty \mathbf{G}_{u_i u_i}(\omega, t) d\omega \\ &= \sum_{k=1}^m \sum_{\ell=1}^m p_k p_\ell \phi_{ik} \phi_{i\ell} \Sigma_{k\ell}(t) \end{aligned} \quad (68)$$

Let us assume now the modulating function $a(\omega, t) = U(t)$. Starting from the nonstationary formulation, it is possible to deduce the formulation in the case of stationary input. This result

is obtained by performing the limit as $t \rightarrow \infty$ into Eq. 61:

$$\begin{aligned}\lim_{t \rightarrow \infty} Z_j(\omega, t) &= \int_0^{\infty} h_j(t - \tau) \exp(i\omega\tau) d\tau \\ &= \exp(i\omega t) H_j(\omega); \\ \lim_{t \rightarrow \infty} \dot{Z}_j(\omega, t) &= \int_0^{\infty} \dot{h}_j(t - \tau) \exp(i\omega\tau) d\tau \\ &= i\omega \exp(i\omega t) H_j(\omega)\end{aligned}\quad (69)$$

where $h_j(t)$ is the functions defined in Eq. 62 and $H_j(\omega)$ is the *frequency response function* of the j th modal oscillator:

$$\begin{aligned}H_j(\omega) &= \int_0^{\infty} h_j(\tau) \exp(-i\omega\tau) d\tau \\ &= \frac{1}{\omega_j^2 - \omega^2 + i2\zeta_j\omega_j\omega}\end{aligned}\quad (70)$$

By substituting Eq. 69 into Eq. 63 and the results into Eq. 64, the following relationship is obtained:

$$\begin{aligned}\Sigma_{k\ell} &= \int_0^{\infty} G_{NN}(\omega) \mathbf{H}_k^*(\omega) \mathbf{H}_\ell^T(\omega) d\omega \\ &\equiv \begin{bmatrix} \lambda_{0,k\ell} & i\lambda_{1,k\ell} \\ -i\lambda_{1,k\ell}^* & \lambda_{2,k\ell} \end{bmatrix}\end{aligned}\quad (71)$$

where

$$\mathbf{H}_j(\omega) = \begin{bmatrix} H_j(\omega) \\ i\omega H_j(\omega) \end{bmatrix} \quad j = k, \ell \quad (72)$$

The elements of the matrix $\Sigma_{k\ell}$ are the cross GSMs, introduced by Vanmarcke (1972), that is,

$$\begin{aligned}\lambda_{i,k\ell} &\equiv \lambda_{i,k\ell}^G \\ &= \int_0^{\infty} \omega^i H_k^*(\omega) H_\ell(\omega) G_{NN}(\omega) d\omega; \quad i = 0, 1, 2\end{aligned}\quad (73)$$

Stochastic Response for Multi-correlated Stochastic Input Processes

In earthquake engineering it is common to assume that the entire base of a structure is subjected to a uniform ground motion. This hypothesis inherently implies that the ground motion is a result of spatially uniform motion. Thus it is certainly true when the base dimensions of the structure are small relative to the seismic vibration wavelengths. From an analytical point of view, this hypothesis leads to mono-correlated zero-mean Gaussian models of the ground-motion acceleration. It follows that this model of earthquake excitations is undoubtedly advantageous, because it substantially facilitates the stochastic analysis of dynamic problem. Indeed, in this case the stochastic analysis requires the knowledge of only one PSD of the input process.

The assumption of uniform ground motion is inappropriate for structures, such as long span bridges, which require accounting for spatial variability of the support motions. The main sources of ground-motion spatial variability are the loss of coherency of seismic waves with distance, the difference in the arrival times of waves at separate supports, and the difference in soil conditions underneath the supports. In these cases the multi-correlated model of the seismic excitation is more appropriate.

In order to take into account of the spatial variability of earthquake-induced ground motions, let us consider an n -DoF structural system subjected to an N support motion. It follows that the stochastic forcing vector process has to be modeled as a multi-correlated zero-mean Gaussian random vector process:

$$\mathbf{f}(t) = \mathbf{B}\bar{\mathbf{F}}(t) \quad (74)$$

where $\mathbf{B}$ is the $(n \times N)$ matrix of spatial distribution of loads and $\bar{\mathbf{F}}(t)$ is a zero-mean Gaussian nonstationary multi-correlated random vector process of order N . Following the Priestley spectral representation of nonstationary processes (Priestley 1999), this vector can be defined by the *Fourier-Stieltjes integral* as follows (Di Paola and Petrucci 1990):

$$\bar{\mathbf{F}}(t) = \sqrt{2} \int_{-\infty}^{+\infty} \exp(i\omega t) \mathbf{A}(\omega, t) d\mathbf{N}(\omega) \quad (75)$$

where $\mathbf{A}(\omega, t)$ is a slowly varying complex deterministic time-frequency modulating diagonal function matrix, of order $N \times N$, which has to satisfy the condition $\mathbf{A}(\omega, t) \equiv \mathbf{A}^*(-\omega, t)$, while $\mathbf{N}(\omega)$ is an orthogonal vector process satisfying the following conditions:

$$\begin{aligned} E\langle d\mathbf{N}(\omega_1) d\mathbf{N}^{*T}(\omega_2) \rangle \\ = \delta(\omega_1 - \omega_2) \mathbf{G}_{\mathbf{NN}}(\omega_1) d\omega_1 d\omega_2 \end{aligned} \quad (76)$$

This relationship shows that the stationary counterpart of the multi-correlated stochastic process is a vector process, $\mathbf{N}(\omega)$ (of order N), with orthogonal increments. Furthermore, $\mathbf{G}_{\mathbf{NN}}(\omega)$ is an $(N \times N)$ Hermitian matrix function which describes the one-sided PSD function matrix of the so-called “embedded” stationary counterpart vector process, $\mathbf{N}(\omega)$. After some algebra it can be proved that the autocorrelation function matrix of the zero-mean Gaussian nonstationary random vector process $\bar{\mathbf{F}}(t)$ can be obtained as

$$\begin{aligned} \bar{\mathbf{R}}_{\mathbf{FF}}(t_1, t_2) &= E\langle \bar{\mathbf{F}}(t_1) \bar{\mathbf{F}}^{*T}(t_2) \rangle = \int_0^{\infty} \exp[i\omega(t_1 - t_2)] \mathbf{A}(\omega, t_1) \mathbf{G}_{\mathbf{NN}}(\omega) \mathbf{A}^*(\omega, t_2) d\omega \\ &= \int_{-\infty}^{\infty} \exp[i\omega(t_1 - t_2)] \mathbf{G}_{\mathbf{FF}}(\omega, t_1, t_2) d\omega \end{aligned} \quad (77)$$

where

$$\mathbf{G}_{\mathbf{FF}}(\omega, t_1, t_2) = \mathbf{A}(\omega, t_1) \mathbf{G}_{\mathbf{NN}}(\omega) \mathbf{A}^*(\omega, t_2) \quad (78)$$

According to the Priestley evolutionary process model (Priestley 1999), the EPSD function matrix of the nonstationary multi-correlated process $\bar{\mathbf{F}}(t)$ is given as

$$\mathbf{G}_{\mathbf{FF}}(\omega, t) = \mathbf{A}(\omega, t) \mathbf{G}_{\mathbf{NN}}(\omega) \mathbf{A}^*(\omega, t) \quad (79)$$

In the framework of stochastic seismic analysis, the main distinct phenomena that give rise to the spatial variability of earthquake-induced ground motions are (i) the *incoherence effect* associated to loss of coherency of seismic waves due to the differential superpositioning of waves arriving from an extended source, (ii) the *wave-passage effect* due to difference in the arrival times of waves at separate supports, (iii) the *attenuation effect* due to gradual decay of wave amplitudes with distance due to energy

dissipation in the ground medium, and (iv) the *site-response effect* associated to spatially varying of the local soil profiles. These effects are taken into account by the so-called coherency function, $\gamma_{\mathbf{NN},sr}(\omega)$, defined as

$$\gamma_{\mathbf{NN},sr}(\omega) = \frac{G_{\mathbf{NN},sr}(\omega)}{\sqrt{G_{\mathbf{NN},rr}(\omega) G_{\mathbf{NN},ss}(\omega)}} \quad (80)$$

where $G_{\mathbf{NN},sr}(\omega)$ is the s th and r th element of the matrix $\mathbf{G}_{\mathbf{NN}}(\omega)$ that is the one-sided PSD at the s th and the r th supports. The coherency function is usually written as

$$\gamma_{\mathbf{NN},sr}(\omega) = \rho_{sr}(\omega) \exp(-i\omega d_{sr}/c) \quad (81)$$

in which $\exp(-i\omega d_{sr}/c)$ is a measure of the wave-passage delay due to the apparent velocity of waves, c represents the velocity of the seismic waves through the ground (c decreases as the distance between the support points increases or the soil is softer), d_{sr} is the distance between the r th and s th support points, and $\rho_{sr}(\omega)$ is the real frequency-dependent spatial correlation function

between the two support points. Several models of the coherency function have been proposed in literature (see, e.g., Harichandran and Vanmarcke 1986; Zerva 1991). Moreover, the one-sided cross-PSD of ground accelerations in a particular direction between surface points r and s , $G_{\mathbf{NN},sr}(\omega)$, may be written as the product of the coherence function by the target one-sided PSD of ground acceleration $G_0(\omega)$ as follows:

$$G_{\mathbf{NN},sr}(\omega) = \begin{cases} \gamma_{\mathbf{NN},sr}(\omega) G_0(\omega), & \text{for } r \neq s \\ G_0(\omega), & \text{for } r = s \end{cases} \quad (82)$$

Let us evaluate now the NGSMs of the structural response. In the case of multi-correlated excitations, the forcing vector $\mathbf{f}(t)$ assumes the expression (Eq. 74); it follows that the differential equation (49), governing the motion in modal subspace, can be written as

$$\ddot{\mathbf{q}}(t) + \mathbf{\Xi} \dot{\mathbf{q}}(t) + \mathbf{\Omega}^2 \mathbf{q}(t) = \mathbf{P} \bar{\mathbf{F}}(t) \quad (83)$$

where

$$\mathbf{P} = \mathbf{\Phi}^T \mathbf{B} \quad (84)$$

After very tedious algebra it can be shown that the j th NGSMs, $\lambda_{j,u_i u_i}(t)$ ($j = 0, 1, 2$), of the j th nodal displacement response, $u_i(t)$, are given as a function of modal NGSMs, $p_{jv} \lambda_{j,k\ell rs}(t)$ ($\ell, k = 1, \dots, m$; $r, s = 1, \dots, N$), by the following relationships:

$$\begin{aligned} \lambda_{0,u_i u_i}(t) &= \sum_{k=1}^m \sum_{\ell=1}^m \sum_{r=1}^N \sum_{s=1}^N p_{kr} p_{\ell s} \phi_{ik} \phi_{i\ell} \lambda_{0,k\ell rs}(t); \\ \lambda_{1,u_i u_i}(t) &= \sum_{k=1}^m \sum_{\ell=1}^m \sum_{r=1}^N \sum_{s=1}^N p_{kr} p_{\ell s} \phi_{ik} \phi_{i\ell} \lambda_{1,k\ell rs}(t); \\ \lambda_{2,u_i u_i}(t) &= \sum_{k=1}^m \sum_{\ell=1}^m \sum_{r=1}^N \sum_{s=1}^N p_{kr} p_{\ell s} \phi_{ik} \phi_{i\ell} \lambda_{2,k\ell rs}(t) \end{aligned} \quad (85)$$

where p_{jv} is the j th and v th element of the matrix $\mathbf{P}$ ($j = k, \ell$; $v = r, s$). Once again it has to be emphasized that the zeroth NGSM, $\lambda_{0,u_i u_i}(t)$, and the second-order NGSM, $\lambda_{2,u_i u_i}(t)$, are real functions that coincide with the variances of the

response in terms of displacement and velocity, respectively, while the first-order NGSM, $\lambda_{1,u_i u_i}(t)$, is a complex quantity (Di Paola 1985; Langley 1986; Di Paola and Petrucci 1990; Muscolino 1991).

As shown in Eq. 85 the nodal NGSMs can be evaluated as a function of $\lambda_{j,k\ell rs}(t)$, $j = 0, 1, 2$, which are the so-called time-dependent modal NGSMs “purged” by p_{jv} factors. After some algebra, these quantities, which are complex ones, can be evaluated, in frequency domain, for quiescent structural systems, at time $t = 0$, as (Di Paola 1985; Di Paola and Petrucci 1990; Muscolino 1991):

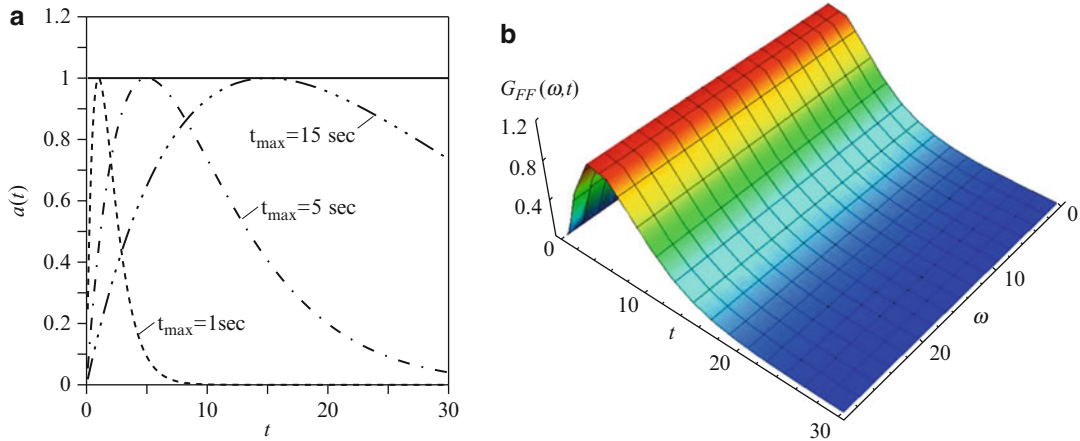
$$\begin{aligned} \lambda_{0,k\ell rs}(t) &= \int_0^\infty Z_{kr}^*(\omega, t) Z_{\ell s}(\omega, t) G_{\mathbf{NN},sr}(\omega) d\omega; \\ \lambda_{1,k\ell rs}(t) &= -i \int_0^\infty Z_{kr}^*(\omega, t) \dot{Z}_{\ell s}(\omega, t) G_{\mathbf{NN},sr}(\omega) d\omega; \\ \lambda_{2,k\ell rs}(t) &= \int_0^\infty \dot{Z}_{kr}^*(\omega, t) \dot{Z}_{\ell s}(\omega, t) G_{\mathbf{NN},sr}(\omega) d\omega \end{aligned} \quad (86)$$

where for $j = k, \ell$, the following positions have been made:

$$\begin{aligned} Z_{jv}(\omega, t) &= \int_0^t h_j(t - \tau) \exp(i\omega\tau) a_{vv}(\omega, \tau) d\tau; \\ \dot{Z}_{jv}(\omega, t) &= \int_0^t \dot{h}_j(t - \tau) \exp(i\omega\tau) a_{vv}(\omega, \tau) d\tau; \\ j &= k, \ell; \quad v = r, s \end{aligned} \quad (87)$$

with $h_j(t)$ the function defined in Eq. 62 and $a_{vv}(\omega, t)$ the v th element of the diagonal matrix $\mathbf{A}(\omega, t)$. Moreover, introducing the state variable, the *modal evolutionary frequency response function vector* is defined as

$$\mathbf{Y}_{jv}(\omega, t) = \begin{bmatrix} Z_{jv}(\omega, t) \\ \dot{Z}_{jv}(\omega, t) \end{bmatrix}; \quad j = k, \ell; \quad v = r, s \quad (88)$$



Stochastic Analysis of Linear Systems, Fig. 1 Hsu and Bernard (1978) model of the uniformly modulated nonstationary excitation: (a) modulating function for different value of the time instant, $t_{\max}$, in

Then relationships (33) can be rewritten in compact form as follows:

$$\begin{aligned} \Sigma_{k\ell rs}(t) &= \begin{bmatrix} \lambda_{0,k\ell rs}(t) & i\lambda_{1,k\ell rs}(t) \\ -i\lambda_{1,k\ell rs}^*(t) & \lambda_{2,k\ell rs}(t) \end{bmatrix} \\ &\equiv \int_0^\infty \mathbf{Y}_{kr}^*(\omega, t) G_{\mathbf{NN},sr}(\omega) \mathbf{Y}_{\ell s}^T(\omega, t) d\omega \end{aligned} \quad (89)$$

where $\Sigma_{k\ell rs}(t)$ coincides with the cross-modal *pre-envelope covariance matrix* (Di Paola and Petrucci 1990). The one-sided PSD $G_{\mathbf{NN},sr}(\omega)$ can be evaluated as a function of the coherence function introduced in Eq. 81. Notice that according to the Priestley evolutionary process model (Priestley 1999), the complex cross-correlation function matrix of the zero-mean modal “purged” response process can be evaluated as

which it takes the maximum value; (b) EPSP function $G_{FF}(\omega, t) = |a(t)|^2 G_{\mathbf{NN}}(\omega)$ with $G_{\mathbf{NN}}(\omega) = 1 \text{ cm}^2/\text{sec}^3$ and $\alpha = 1/5$

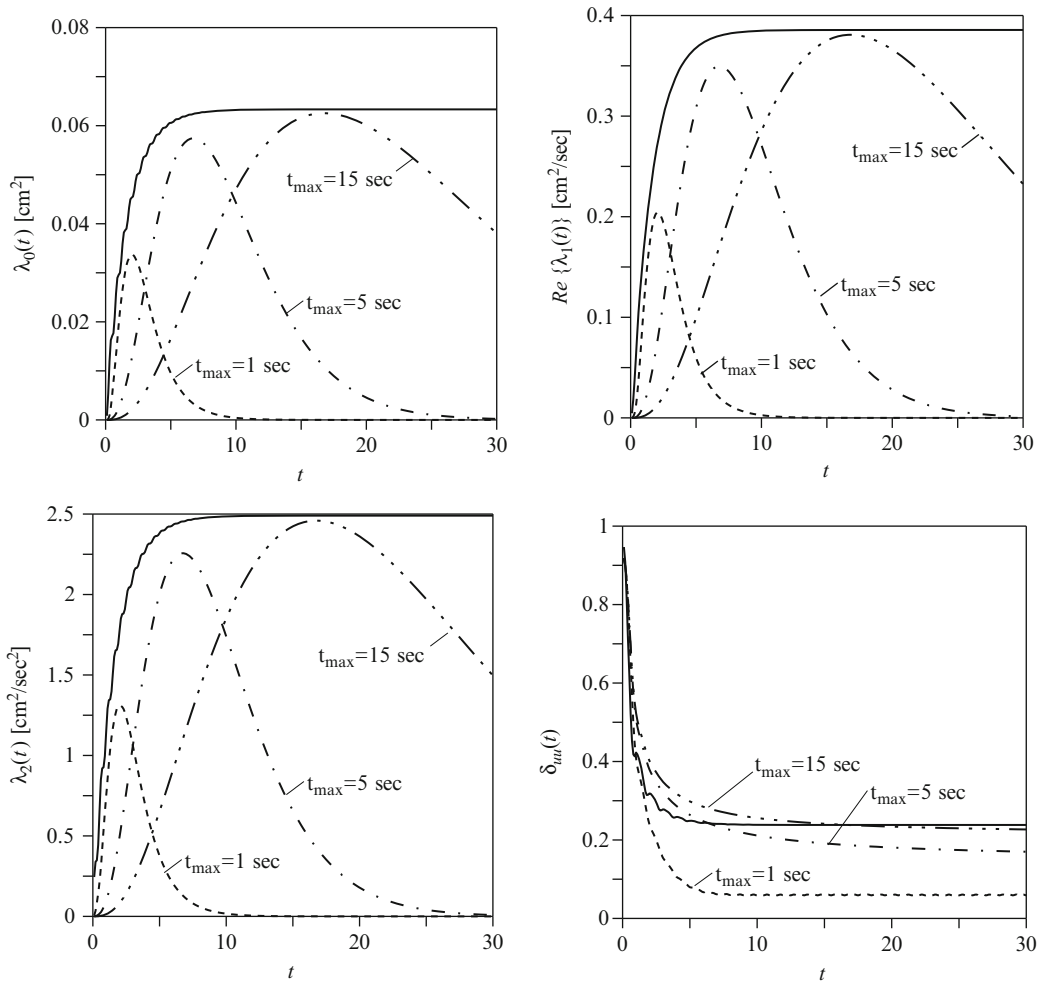
$$\begin{aligned} \mathbf{R}_{k\ell rs}(t_1, t_2) &= \int_0^\infty \mathbf{Y}_{kr}^*(\omega, t_1) G_{\mathbf{NN},sr}(\omega) \mathbf{Y}_{\ell s}^T(\omega, t_2) d\omega \\ &\equiv \int_0^\infty \exp[i\omega(t_1 - t_2)] \mathbf{G}_{k\ell rs}(\omega, t_1, t_2) d\omega \end{aligned} \quad (90)$$

where

$$\mathbf{G}_{k\ell rs}(\omega, t_1, t_2) = \mathbf{Y}_{kr}^*(\omega, t_1) G_{\mathbf{NN},sr}(\omega) \mathbf{Y}_{\ell s}^T(\omega, t_2) \quad (91)$$

is the cross-modal EPSP function matrix between the k, ℓ th dummy oscillators at the r th and s th support points. By means of the modal transformation (47), the nodal autocorrelation and the EPSP function matrices of the displacement response, $u_i(t)$, can be evaluated, after very simple algebra, respectively, as follows:

$$\begin{aligned} \mathbf{R}_{u_i u_i}(t_1, t_2) &= \sum_{k=1}^m \sum_{\ell=1}^m \sum_{r=1}^N \sum_{s=1}^N p_{kr} p_{\ell s} \phi_{ik} \phi_{i\ell} \mathbf{R}_{k\ell rs}(t_1, t_2), \\ \mathbf{G}_{u_i u_i}(\omega, t_1, t_2) &= \sum_{k=1}^m \sum_{\ell=1}^m \sum_{r=1}^N \sum_{s=1}^N p_{kr} p_{\ell s} \phi_{ik} \phi_{i\ell} \mathbf{G}_{k\ell rs}(\omega, t_1, t_2) \end{aligned} \quad (92)$$



Stochastic Analysis of Linear Systems, Fig. 2 NGSM functions, $\lambda_{j,uu}(t)$ ($j = 0, 1, 2$) and bandwidth parameter, $\delta_{uu}(t)$, of the transient response of an oscillator ($\omega_0 = 2\pi$ rad/sec, $\xi_0 = 0.05$) in the stationary case

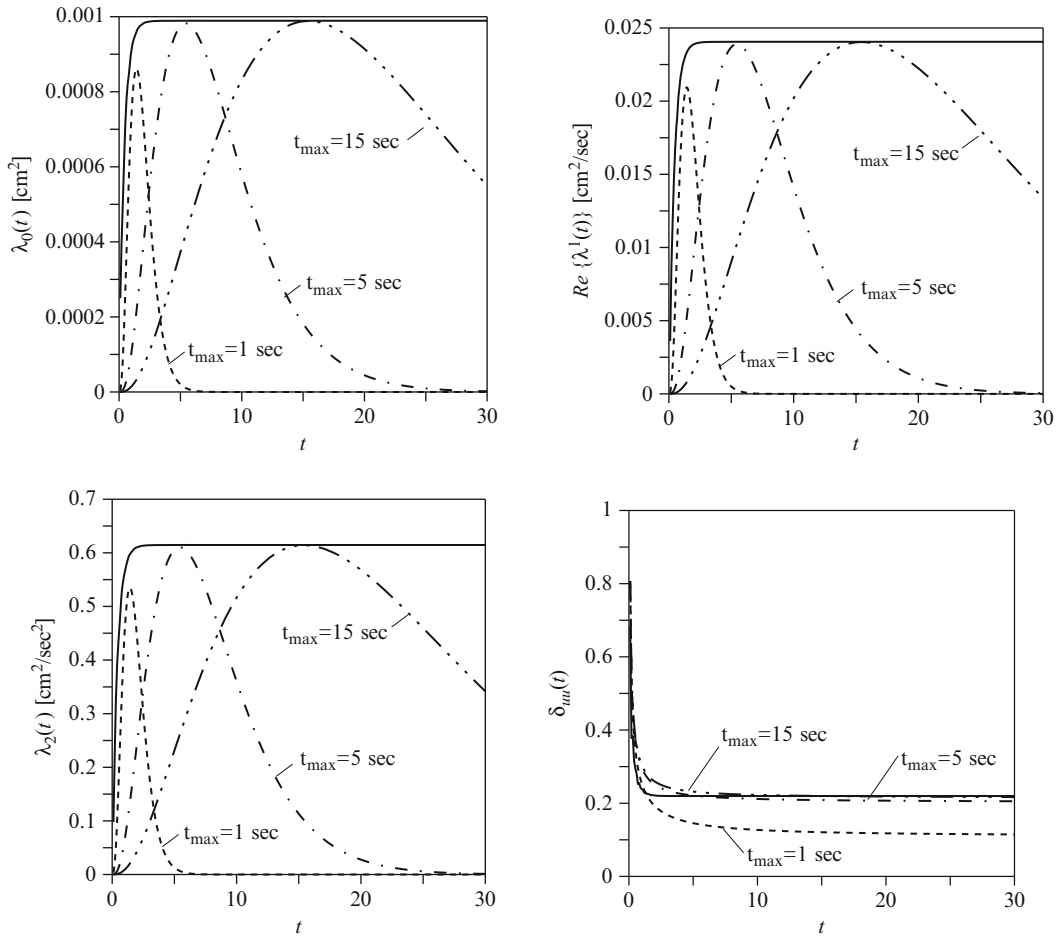
(solid line) and adopting the Hsu and Bernard (1978) model, for different value of the time instant, $t_{\max}$, in which the modulating $a(t)$ function take the maximum value (dashed lines)

Consequently the nodal displacement NGSMs are the elements of the nodal pre-envelope covariance matrix, given as

$$\begin{aligned} \Sigma_{u_i u_i}(t) &\equiv \mathbf{R}_{u_i u_i}(t, t) \equiv \int_0^\infty \mathbf{G}_{u_i u_i}(\omega, t) d\omega \\ &= \sum_{k=1}^m \sum_{\ell=1}^m \sum_{r=1}^N \sum_{s=1}^N p_{kr} p_{\ell s} \phi_{ik} \phi_{i\ell} \Sigma_{k\ell rs}(t) \end{aligned} \quad (93)$$

Calculating Nonstationary Stochastic Responses

In this section, in order to evidence the main differences between the stochastic responses of structural systems subjected to both uniformly and fully nonstationary models of seismic excitations, the spectral characteristics of the response of two SDoF oscillators are evaluated.



Stochastic Analysis of Linear Systems, Fig. 3 NGSMs, $\lambda_{j,u}(t)$ ($j = 0, 1, 2$) and bandwidth parameter, $\delta_u(t)$, of the transient response of an oscillator ($\omega_0 = 8\pi$ rad/sec, $\xi_0 = 0.05$) in the stationary case

(solid line) and adopting the Hsu and Bernard (1978) model, for different value of the time instant, $t_{\max}$, in which the modulating $a(t)$ function take the maximum value (dashed lines)

Comparison Between Steady-State and Nonstationary Responses of SDoF Systems

Let us consider an oscillator, whose differential equation governing the motion is written in canonical form in Eq. 23, forced by the uniformly modulate Gaussian zero-mean nonstationary process $F(t)$, defined as

$$F(t) = a(t) N(t) \quad (94)$$

where $N(t)$ is a Gaussian white noise process, with one-sided PSD $G_{NN}(\omega) = 1 \text{ cm}^2/\text{sec}^3$, and

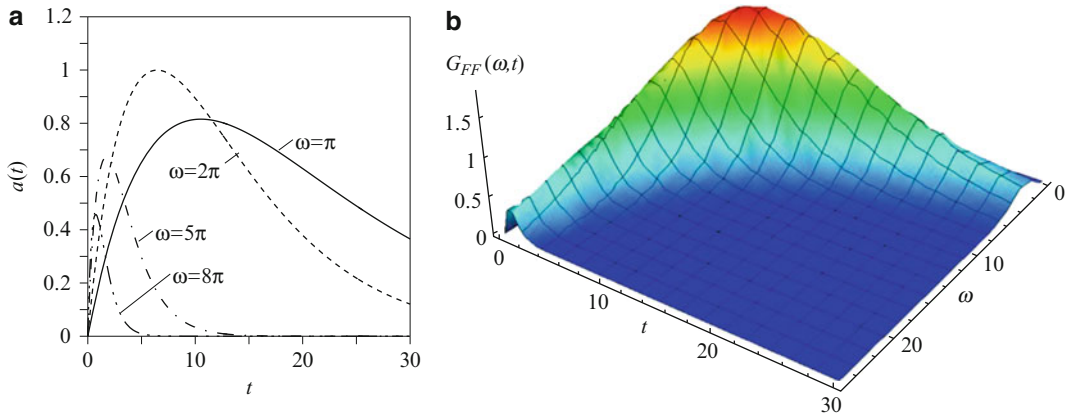
$a(t)$ is the normalized to one modulating function proposed by Hsu and Bernard (1978):

$$a(t) = \alpha \exp(1)t \exp(-\alpha t) U(t) \quad (95)$$

which takes its maximum value at time, $t_{\max} = 1/\alpha$, and $U(t)$ is the unit step function defined as

$$U(t - t_0) = \begin{cases} 0, & t \leq t_0; \\ 1, & t > t_0 \end{cases} \quad (96)$$

In Fig. 1 the modulating function (95) is depicted for different value of $\alpha = 1/t_{\max}$.



Stochastic Analysis of Linear Systems, Fig. 4 Spanos and Solomos (1983) model of the fully nonstationary excitation model: (a) sections of the normalized to one

modulating function $a(\omega, t)$ at different values of abscissa ω ; (b) EPSD function $G_{FF}(\omega, t) = |a(\omega, t)|^2 G_{NN}(\omega)$ with $G_{NN}(\omega) = 1 \text{ cm}^2/\text{sec}^3$

In the same figure the EPSD function $G_{FF}(\omega, t) = |a(t)|^2 G_{NN}(\omega)$ of the quasi-stationary (separable) Hsu and Bernard (1978) model is also shown assuming $\alpha = 1/5$ and $G_{NN}(\omega) = 1 \text{ cm}^2/\text{sec}^3$.

In Fig. 2 the NGSM, $\lambda_{j,uu}(t)$ ($j = 0, 1, 2$), and bandwidth parameter, $\delta_{uu}(t)$, functions of the response of an oscillator with natural circular frequency $\omega_0 = 2\pi \text{ rad/sec}$ and damping ratio $\xi_0 = 0.05$ for the Hsu and Bernard (1978) model of the uniformly modulated nonstationary excitation are depicted. These NGSMs are compared to the transient NGSMs of the stationary case ($a(t) = U(t)$), for different value of the time instant, $t_{\max}$. In Fig. 3 the comparisons are performed for an oscillator with natural circular frequency $\omega_0 = 8\pi \text{ rad/sec}$ and damping ratio $\xi_0 = 0.05$. These figures shows that when the response approaches to its steady-state condition, defined by the steady-state time $t_{SS} \geq 3/(\xi_0 \omega_0)$, the maximum values of the NGSM functions coincide with the corresponding values of the NGSM for stationary input processes. If this condition is not satisfied, the stationary approximation of input process leads to overestimated results. In fact, for the first oscillator for which the steady-state time is $t_{SS} \approx 9.55 \text{ sec}$, the NGSMs with $t_{\max} < t_{SS}$ possess maximum values lesser than the corresponding stationary case. Similar results

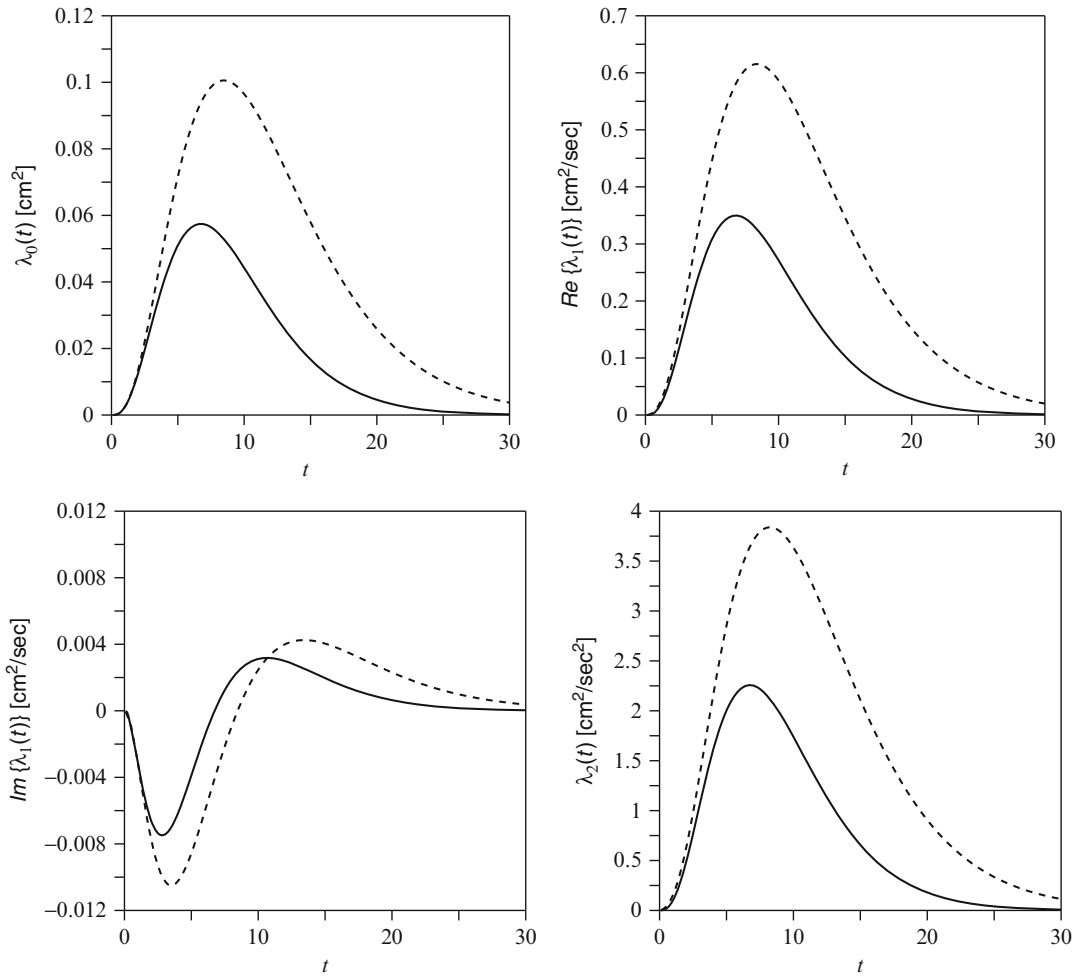
are obtained for the second oscillator for which $t_{SS} \approx 2.39 \text{ sec}$. Moreover, the analysis of the bandwidth parameters shows that if $t_{\max} < t_{SS}$ the parameter $\delta_{uu}(t)$ is narrower than the value obtained in the relative stationary case.

Comparison Between Uniformly and Fully Nonstationary Responses of SDoF Systems

Let us consider now the Spanos and Solomos (1983) model of the fully nonstationary Gaussian zero-mean process. For this model the normalized to one evolutionary modulating function can be written as

$$a(\omega, t) = \varepsilon(\omega) t \exp(-\alpha(\omega) t) U(t) \quad (97)$$

Selecting the following parameters $\varepsilon(\omega) = \omega \sqrt{2}/5 \pi a_{\max}$ and $\alpha(\omega) = 0.15/2 + \varepsilon^2(\omega)/4$, the normalizing to one coefficient is $a_{\max} = 1.34$. The unitary maximum is reached at $\omega = 1.937 \pi \text{ rad/sec}$ and at $t = 6.667 \text{ sec}$. In Fig. 4, the section of the modulating function at different values of abscissa ω is depicted together with the one-sided EPSD function $G_{FF}(\omega, t) = |a(\omega, t)|^2 G_{NN}(\omega)$ with $G_{NN}(\omega) = 1 \text{ cm}^2/\text{sec}^3$. This figure evidences the frequency dependence of this model of nonstationary input process.

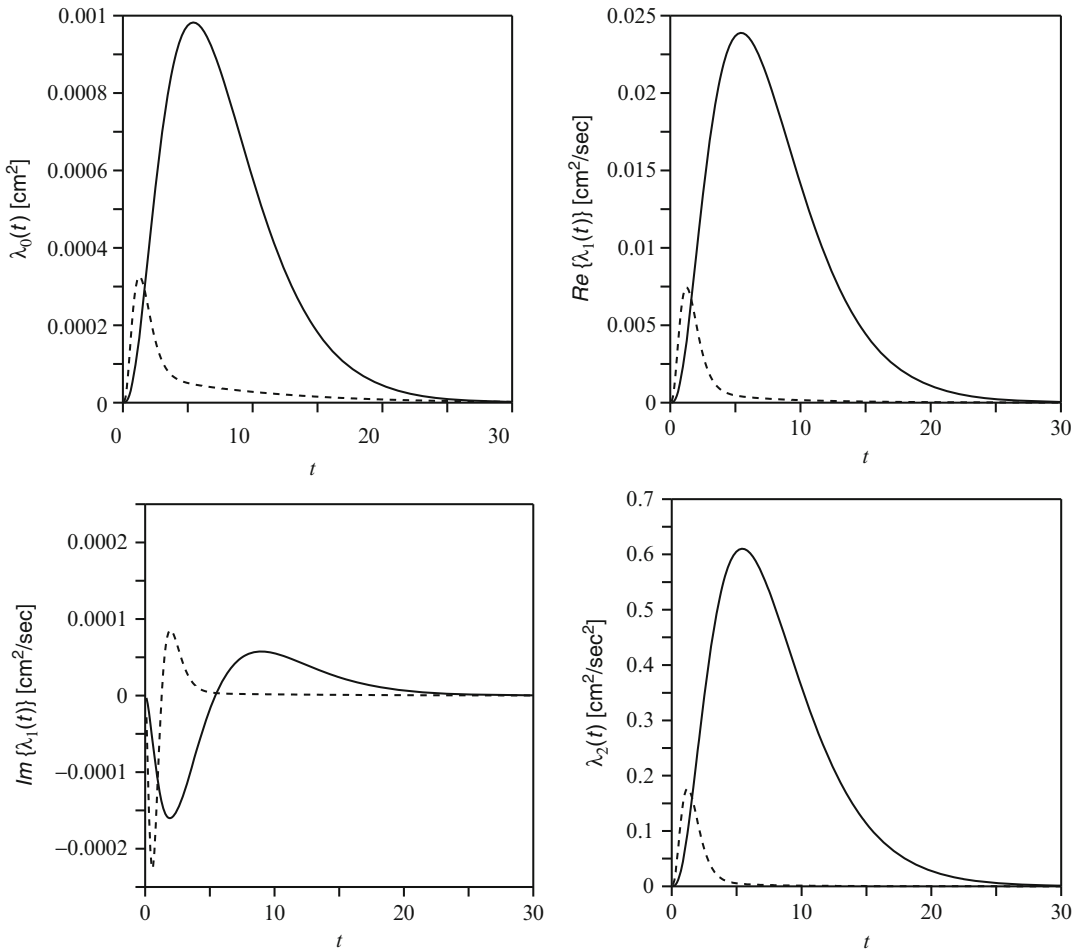


Stochastic Analysis of Linear Systems, Fig. 5 NGSMs of the response of an oscillator with $\omega_0 = 2\pi$ rad/sec, $\xi_0 = 0.05$ for the Hsu and Bernard (1978) model (solid line) and for the Spanos and Solomos (1983) model (dashed line)

In Figs. 5 and 6 the NGSMs, $\lambda_{j,uu}(t)$ ($j = 0, 1, 2$), of the two oscillators with natural circular frequency $\omega_0 = 2\pi$ rad/sec and $\omega_0 = 8\pi$ rad/sec and damping ratio $\xi_0 = 0.05$, obtained by means of the two modulating functions before described, are depicted and compared. Analyzing these figures it can be evidenced that the temporal variation of the frequency content of the EPSD function, often neglected for mathematic convenience, has substantial effects on the structural response. In fact the maximum values of the NGSM functions depend on the form of modulating function as well as the dynamic

characteristics of the structural systems. Moreover, in some cases, the quasi-stationary modelling of the modulating function can lead to very unconservative results (see Fig. 5) of the stochastic structural response.

In Figs. 7 and 8 the mean frequencies, $v_{uu}^+(t)$, and the normalized time-varying central frequencies, $\omega_{C,uu}(t)/\omega_0$, defined in Eq. 21, of the response of the two oscillators before analyzed for the Hsu and Bernard (1978) and Spanos and Solomos (1983) modulating functions of the nonstationary zero-mean Gaussian input process, are depicted.



Stochastic Analysis of Linear Systems, Fig. 6 NGSMs of the response of the oscillator with $\omega_0 = 8\pi$ rad/sec, $\xi_0 = 0.05$ for the Hsu and Bernard (1978) model (solid line) and for the Spanos and Solomos (1983) model (dashed line)

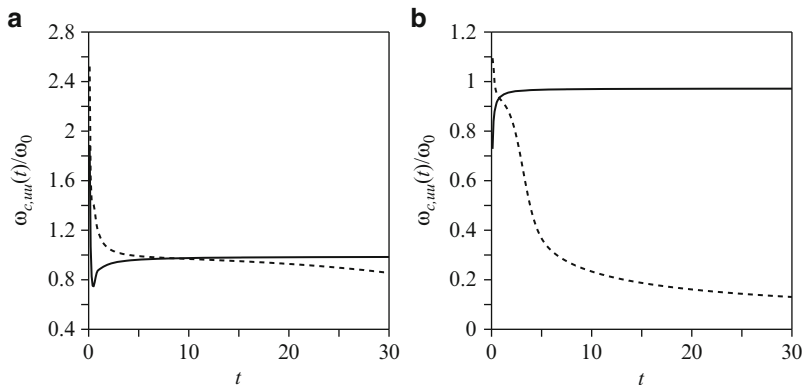
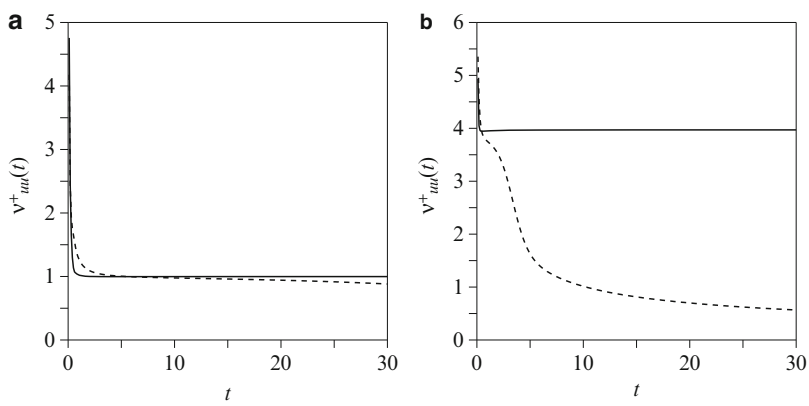
Figure 7 shows that for the Hsu and Bernard (1978) modulating function, the mean frequencies, $v_{uu}^+(t)$, get to an asymptotic value very close to the natural frequency of the oscillators. This behavior is not verified for the oscillator with higher natural circular frequency subjected to the Spanos and Solomos (1983) model of the nonstationary input process, where the mean frequency decreases evidencing the frequency dependence of the structural response. Similar results are obtained for the normalized time-varying central frequencies, $\omega_{C,uu}(t)/\omega_0$, depicted in Fig. 8.

Summary

The dynamic behavior of structural systems subjected to uncertain dynamic excitations can be performed through the stochastic analysis, which requires the probabilistic characterization of both input and output processes. The characterization of output processes can be extremely complex, when nonstationary and/or non-Gaussian input processes are involved. However, in several cases the approximate description of the dynamic structural response based on its spectral characteristics may be sufficient.

Stochastic Analysis of Linear Systems,

Fig. 7 Mean frequencies of the response of the oscillator for the Hsu and Bernard (1978) model (solid line) and for the Spanos and Solomos (1983) model (dashed line); (a) oscillator with $\omega_0 = 2\pi$ rad/sec and $\xi_0 = 0.05$; (b) oscillator with $\omega_0 = 8\pi$ rad/sec and $\xi_0 = 0.05$



Stochastic Analysis of Linear Systems, Fig. 8 Time-varying central frequency, normalized by the natural circular frequency, of the oscillator for the Hsu and Bernard (1978) model (solid line) and for the Spanos and Solomos

(1983) model (dashed line); (a) oscillator with $\omega_0 = 2\pi$ rad/sec and $\xi_0 = 0.05$; (b) oscillator with $\omega_0 = 8\pi$ rad/sec and $\xi_0 = 0.05$

In this study a unitary approach to evaluate the spectral characteristics of the structural response, to perform the reliability assessment, of classically damped linear systems subjected to stationary or nonstationary mono-/multi-correlated zero-mean Gaussian excitations, is described.

The main steps of the described approach are (i) the use of modal analysis to decouple the equation of motion; (ii) the determination, in state variable, of the evolutionary frequency response vector functions and of the evolutionary power spectral density function matrix of the structural response; and (iii) the evaluation of the nongeometric spectral moments as well as the spectral characteristics of the stochastic response of linear systems subjected to stationary

or nonstationary mono-/multi-correlated zero-mean Gaussian seismic excitations.

Cross-References

- [Probability Density Evolution Method in Stochastic Dynamics](#)
- [Stochastic Analysis of Nonlinear Systems](#)
- [Stochastic Ground Motion Simulation](#)

References

Borino G, Muscolino G (1986) Mode-superposition methods in dynamic analysis of classically and non-classically damped linear systems. Earthq Eng Struct Dyn 14:705–717

- Conte JP, Peng B-F (1997) Fully nonstationary analytical earthquake ground-motion model. *J Eng Mech (ASCE)* 123:15–24
- Corotis RB, Vanmarcke EH, Cornell CA (1972) First passage of nonstationary random processes. *J Eng Mech (ASCE)* 98:401–414
- Di Paola M (1985) Transient spectral moments of linear systems. *SM Arch* 10:225–243
- Di Paola M, Petrucci G (1990) Spectral moments and pre-envelope covariances of nonseparable processes. *J Appl Mech (ASME)* 57:218–224
- Fan FG, Ahmadi G (1990) Nonstationary Kanai-Tajimi models for El Cento 1940 and Mexico City 1985 earthquakes. *Probab Eng Mech* 5:171–181
- Harichandran RS, Vanmarcke EH (1986) Stochastic variation of earthquake ground motion in space and time. *J Eng Mech (ASCE)* 112:154–174
- Hsu T-I, Bernard MC (1978) A random process for earthquake simulation. *Earthq Eng Struct Dyn* 6:347–362
- Jennings PC, Housner GW, Tsai C (1969) Simulated earthquake motions for design purpose. In: *Proceedings of the 4th world conference earth engineering*, Santiago, A-1, pp 145–160
- Langley RS (1986) Structural response to non-stationary non-white stochastic ground motion. *Earthq Eng Struct Dyn* 14:909–924
- Li J, Chen JB (2009) *Stochastic dynamics of structures*. Wiley, Singapore
- Lin YK (1976) *Probabilistic theory of structural dynamics*. Krieger, Huntington
- Lutes LD, Sarkani S (2004) *Random vibrations – analysis of structural and mechanical vibrations*. Elsevier, Boston
- Michaelov G, Sarkani S, Lutes LD (1999a) Spectral characteristics of nonstationary random processes – a critical review. *Struct Saf* 21:223–244
- Michaelov G, Sarkani S, Lutes LD (1999b) Spectral characteristics of nonstationary random processes – response of a simple oscillator. *Struct Saf* 21:245–267
- Muscolino G (1991) Nonstationary pre-envelope covariances of nonclassically damped systems. *J Sound Vib* 149:107–123
- Muscolino G, Palmeri A (2005) Maximum response statistics of MDOF linear structures excited by non-stationary random processes. *Comput Method Appl Mech Eng* 194:1711–1737
- Priestley MB (1999) *Spectral analysis and time series*. Academic, London
- Shinozuka M, Sato Y (1967) Simulation of nonstationary random process. *J Eng Mech (ASCE)* 93:11–40
- Spanos P, Solomos GP (1983) Markov approximation to transient vibration. *J Eng Mech (ASCE)* 109:1134–1150
- Vanmarcke EH (1972) Properties of spectral moments with applications to random vibrations. *J Eng Mech (ASCE)* 98:425–446
- Vanmarcke EH (1975) On the distribution of the first-passage time for normal stationary random processes. *J Appl Mech (ASME)* 42:215–220
- Zerva A (1991) Effect of spatial variability and propagation of seismic ground motions on the response of multiply supported structures. *Probab Eng Mech* 6:212–221

Stochastic Analysis of Nonlinear Systems

Annibale Luigi Materazzi and Marco Breccolotti
Department of Civil and Environmental
Engineering, University of Perugia, Perugia, Italy

Synonyms

Earthquake engineering; Hysteresis;
Nonlinearity; Random vibrations

Introduction

The present chapter is devoted to the probabilistic analysis of the random response of nonlinear structural systems exposed to random excitation with special attention to earthquake action. The random system response may be due to random excitation, to random system properties, to random boundary conditions. The nonlinear character of the response is mainly due to the nonlinear materials properties and to the effect of large displacements (the so-called P- Δ effect).

The first pioneering studies on this topic took place in the 1960s and 1970s (VanMarcke et al. 1970; Iwan 1973; Atalik and Utku 1976; Spanos 1976), when equivalent linearization techniques were used, taking advantage of the availability of first digital computers.

The attention is limited herein to systems having deterministic properties, including deterministic boundary conditions, considering those types of nonlinearity that can really occur during earthquakes.

After a brief review of the types of nonlinear behavior that can be expected, the available methods for the analysis of the response are discussed.

They can be classified into analytical and simulation methods. Among the first ones are considered in the following: the Fokker-Planck-Kolmogorov (FPK) equation, the equivalent linearization, the perturbation method, and the

stochastic averaging. The simulation methods based on Monte Carlo techniques are then presented.

Types of Nonlinearity in Earthquake Engineering

During severe earthquakes most structures undergo great amplitude time-varying displacements which can induce inelastic behavior in the structural members.

In the case of reinforced concrete structures, several phenomena take place, like cracking and crushing, together with yielding and strain hardening of steel. Also the bond between steel and concrete may be imperfect, as great shear forces must be exchanged and slippage can occur. Moreover, the reversal of the sign of displacements, due to the structural vibration, leads to the phenomenon of the hysteresis and to the subsequent dissipation of energy.

This apparently undesirable behavior is deliberately taken into account by the modern strategies for aseismic design of reinforced concrete structures, based on the concept of the capacity design, as it allows the reduction of the inertia forces acting upon the structures by means of energy dissipation.

The modeling of the interaction of concrete and steel under severe earthquakes was studied in depth in the past (Park et al. 1972, Takeda et al. 1970, Popov et al. 1972). Among the other aspects, the hysteretic behavior after steel yielding, the stiffness deterioration due to cracking of concrete, the strength deterioration and softening consequent to cumulative severe large deformations, and the pinching behavior due to shear, along with bond deterioration, were investigated.

Simple hysteretic models that can be used in the practice are presented in Fig. 1. Model (a) represents an elastic-perfectly plastic behavior, while model (b) is a variant allowing for strain hardening. Model (c) (Masing 1926) takes into account the stiffness deterioration, using smooth curves, and model (d) is the degrading stiffness model due to Takeda.

The Masing model is also suitable to represent the friction dissipation which takes place in masonry structures under earthquakes.

An interesting and powerful nonlinear model is represented by the Bouc-Wen equation (Bouc 1967; Wen 1976).

The restoring force F is given by

$$F = \alpha k_0 x + (1 - \alpha) Q Z \quad (1)$$

where k_0 is the initial system stiffness, α is the ratio between the post- and the pre-yielding stiffness, x is the displacement, Q is the yield strength, and Z is a nondimensional parameter which takes into account the hysteresis, satisfying the following nonlinear first-order differential equation:

$$q \frac{dZ}{dt} = A \dot{x} - \beta |\dot{x}| Z |Z|^{n-1} - \tau \dot{x} |Z|^n \quad (2)$$

where β , τ , A , and n are the dimensionless parameters which control the shape of the hysteresis loop, q is the yield displacement, and $\dot{x}$ is the velocity.

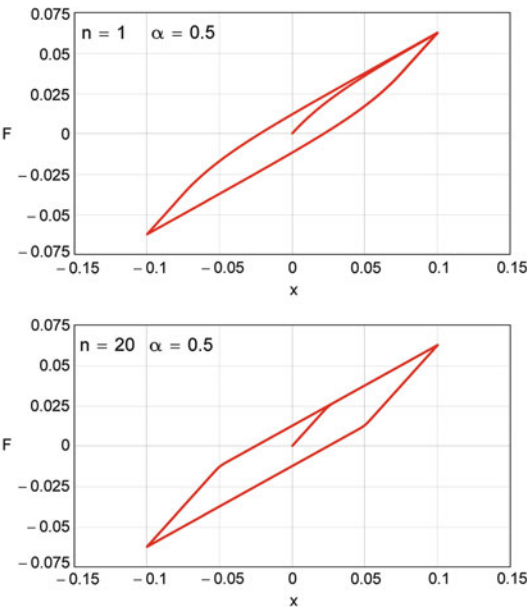
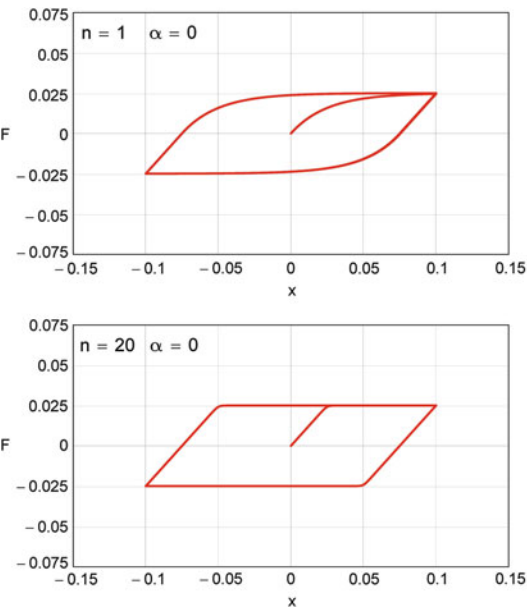
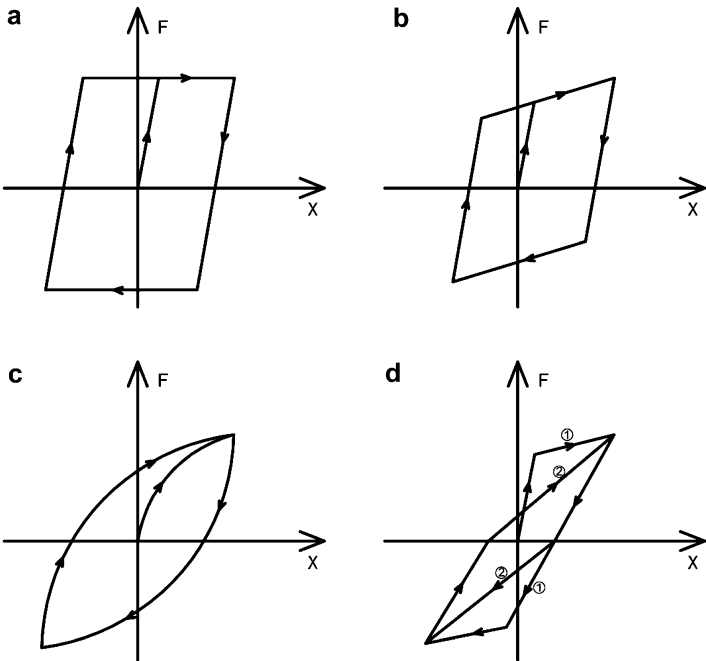
Typical force-deformation hysteresis loops generated using Eqs. 1 and 2 are shown in Fig. 2. In the examples the dynamical system is acted upon by a sinusoidal motion having amplitude 0.1 m and frequency 1.0 Hz. The other parameters of the model are $q = 0.025$ m, $\beta = \tau = 0.5$, $k_0 = 1$, and $A = 1$. Thus, by changing the different parameters of the Wen's model, many hysteretic behaviors can be obtained such as the Masing type, the elastic-perfectly plastic, and the elastic plastic with hardening.

Analytical Techniques

The solution of the stochastic differential equations that govern the motion of nonlinear dynamical systems has been first achieved by means of analytical techniques. Exact and approximate solutions can be obtained by these techniques depending on the complexity of the problem. The stochastic response of nonlinear single degree of freedom system with elastic behavior and Gaussian external excitation can be generally

Stochastic Analysis of Nonlinear Systems,

Fig. 1 (a) Elastic-perfectly plastic model; (b) bilinear strain hardening model; (c) Masing-type model; (d) degrading stiffness model due to Takeda



Stochastic Analysis of Nonlinear Systems, Fig. 2 Different hysteresis loops from the Bouc-Wen equation

solved exactly through the FPK equation. Nonlinear elastic multi-degree of freedom can be solved exactly via FPK equation only for very restrictive conditions on the stochastic external excitation. Thus, approximate solutions have

been searched for nonlinear multi-degree of freedom systems and for inelastic systems. In the next paragraphs a brief description of the FPK equation and of some approximate methods is provided.

Fokker-Planck-Kolmogorov Equation

Introduction

The theory of stochastic processes began in the nineteenth century when physicists were trying to show that heat in a medium is essentially a random motion of the constituent molecules. At the end of that century, some researches began to adopt more direct mathematical models of random disturbances instead of considering random motion as due to collisions between objects having a random distribution of initial positions and velocities. In this context several physicists, among which Fokker (1914) and Planck (1915), developed partial differential equations, which were versions of what was subsequently called the Fokker-Planck equation, to study the theory of Brownian motion.

The theory of the Fokker-Planck equation was made considerably more general by Kolmogorov (1931). He assumed the process to be continuous with respect to time and Markovian, i.e., a process for which the future probability density conditional on the present and past is actually independent of the past. On introducing further assumptions, Kolmogorov was able to show that the probability density of the process obeys a partial differential equation of the Fokker-Planck type. He also gave another partial differential equation which the future probability density, conditional on the present state, obeys with respect to the present state. The latter equation is called Kolmogorov's first equation, and the Fokker-Planck-type equation is called Kolmogorov's second equation. Sometimes the Fokker-Planck equation is called the Fokker-Planck-Kolmogorov (FPK) equation.

Solution of the Fokker-Planck-Kolmogorov Equation for an SDOF Elastic Nonlinear Second-Order System

Let us consider the nonlinear nonconservative stochastic system

$$m \frac{d^2x}{dt^2} + c \frac{dx}{dt} + \frac{\partial G(x)}{\partial x} = \xi(t) \quad (3)$$

that describes the displacement x of a body with mass m , damping c , and potential energy $G(x)$

subjected to a (scalar) white-noise forcing function $\xi(t)$ with second moment rate b .

By placing $y = m \frac{dx}{dt}$ the system may be rewritten in the state space as a system of first-order differential equations:

$$\begin{aligned} \frac{dx}{dt} &= \frac{y}{m} \\ \frac{dy}{dt} &= -\frac{\partial G(x)}{\partial x} - \frac{c}{m}y + \xi(t) \end{aligned} \quad (4)$$

This is a special case of the system:

$$\begin{aligned} \frac{dx}{dt} &= \frac{\partial H}{\partial y} \\ \frac{dy}{dt} &= -\frac{\partial H}{\partial x} - c \frac{\partial H}{\partial y} + \xi(t) \end{aligned} \quad (5)$$

where the symbol H represents the total energy (potential plus kinetic):

$$H = G(x) + \frac{1}{2m}y^2 \quad (6)$$

The steady-state probability density $p(x,y)$ for the system of Eq. 5 can be determined by solving the appropriate stationary FPK equation:

$$\begin{aligned} \left[-\frac{\partial}{\partial x} \left(\frac{\partial H}{\partial y} p \right) + \frac{\partial}{\partial y} \left(\frac{\partial H}{\partial x} p \right) \right] + \frac{\partial}{\partial y} \left(c \frac{\partial H}{\partial y} p \right) \\ + \frac{1}{2} b \frac{\partial^2 p}{\partial y^2} = 0 \end{aligned} \quad (7)$$

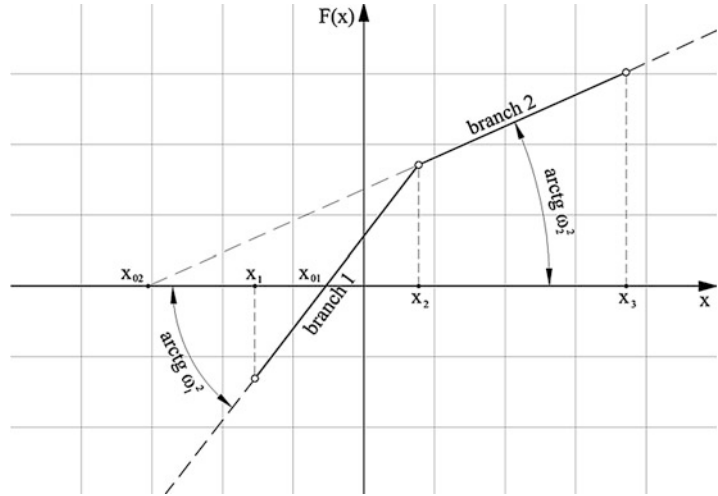
If p is any function of the energy H , the terms in the square brackets vanish, and therefore if it can be found a particular function $p(H)$ which makes the remaining two terms vanish, $p(H)$ is a solution of Eq. 7.

By substituting $p(x, y) = p(H(x, y))$ and assuming null the terms in the square bracket, Eq. 7 becomes

$$\frac{\partial}{\partial y} \left(c \frac{\partial H}{\partial y} p(H) \right) + \frac{1}{2} b \frac{\partial^2}{\partial y^2} p(H) = 0 \quad (8)$$

Stochastic Analysis of Nonlinear Systems,

Fig. 3 Parameters describing a continuous piecewise-linear restoring force



The integration with respect to y yields

$$c \frac{\partial H}{\partial y} p(H) + \frac{1}{2} b \frac{\partial}{\partial y} p(H) = L(x) \quad (9)$$

where $L(x)$ is an arbitrary function.

As boundary conditions for the stationary FPK equation can be assumed, that p is zero at infinity. Thus, the left side of Eq. 9 vanishes when $y \rightarrow \infty$ and hence $L(x) = 0$. Equation 9 simplifies to

$$c \frac{\partial H}{\partial y} p(H) + \frac{1}{2} b \frac{\partial p}{\partial H} \frac{\partial H}{\partial y} = 0 \quad (10)$$

Since $\frac{\partial H}{\partial y}$ cannot be identically null (otherwise H would be independent of y and hence so will be p), then Eq. 10 yields

$$\frac{\partial p}{\partial H} + \frac{2c}{b} p = 0 \quad (11)$$

whose general solution representing the steady-state probability density is

$$p(x, y) = C \exp \left[-\frac{2c}{b} H(x, y) \right] \quad (12)$$

being C a normalizing constant.

For the case of an SDOF system with unit mass and continuous linear piecewise restoring force, a closed form solution of the FPK equation can, thus, be determined. This model can be

used, for instance, to determine the response of cracked reinforced concrete beams with cracks that alternatively open and close (Breccolotti et al. 2008).

In this case the motion of the vibrating system can be described by the equation

$$\ddot{x} + \beta_1 \dot{x} + F_i(x) = f(t) \quad (13)$$

where $F_i(x) = \omega_i^2(x - x_{0i})$ for $x_i \leq x \leq x_{i+1}$, $i = 1, 2, \dots$ (see Fig. 3), ω_i and x_{0i} are positive constants, and $f(t)$ is a Gaussian stochastic process having null mean and uniform power spectrum S_0 over the entire frequency domain.

The following boundary conditions have to be satisfied:

$$w_i(x_{i+1}, \dot{x}, \ddot{x}) = w_{i+1}(x_{i+1}, \dot{x}, \ddot{x}) \quad (14)$$

$$\sum_i \int \int \int w_i(x, \dot{x}, \ddot{x}) dx d\dot{x} d\ddot{x} = 1 \quad (15)$$

$$\omega_i^2(x_{i+1} - x_{0i}) = \omega_{i+1}^2(x_{i+1} - x_{0i+1}) \quad (16)$$

that represent the continuity of the probability density function of the response, the normalizing condition, and the continuity of the piecewise-linear restoring force characteristics, respectively.

It can be demonstrated that the solution of the FPK equation corresponding to this system, representing the stationary probability density function (PDF) of the response, can be written in the form

$$\begin{aligned} w_i(x, \dot{x}) &= C_i \cdot e^{-\left(\frac{\omega_i^2}{s_0} \beta_1 x^2 + \frac{\beta_1}{s_0} \dot{x}^2 - \frac{2s_0 \omega_i^2}{s_0} \beta_1 x\right)} \\ &= C_i \cdot e^{-\frac{\omega_i^2}{s_0} \beta_1 x^2 + \frac{2s_0 \omega_i^2}{s_0} \beta_1 x} \cdot e^{-\frac{\beta_1}{s_0} \dot{x}^2} \\ &= w_{i,x}(x) \cdot w_{i,\dot{x}}(\dot{x}) \end{aligned} \quad (17)$$

It can be noted, as demonstrated by Lin and Cai (1995), that for an SDOF system having nonlinear stiffness, linear damping, and exposed to a Gaussian white-noise excitation, the stationary displacement (x) and the velocity ($\dot{x}$) are independent random variables.

Solution of the Fokker-Planck-Kolmogorov Equation for an nDOF Elastic Nonlinear Second-Order System

Several solutions of the FPK equation for nonlinear elastic systems with n degrees of freedom (nDOF) subjected to random excitation have been developed in the past (Piszczec and Nizioł 1986). Let us consider the following set of n equations for an nDOF system, characterized by elastic nonlinear properties:

$$\ddot{x}_i + \beta_i \dot{x}_i + \frac{1}{m_i} \frac{\partial U}{\partial x_i} = f_i(t) \quad (18)$$

where $f_i(t)$ are independent and uncorrelated white noises with null mean and spectral density S_i and β_i are positive constants.

The term $U(x_1, x_2, \dots)$ represents the potential energy.

This system of differential equations can be solved assuming the validity of the following conditions:

$$\frac{S_i m_i}{2\beta_i} = k \quad i = 1, 2, \dots \quad (19)$$

being k a positive constant. In this case the solution of the FPK equation is

$$\begin{aligned} w(x_1, x_2, \dots, \dot{x}_1, \dot{x}_2, \dots) \\ = C \exp \left[-\frac{1}{k} \left(\frac{1}{2} \sum_{i=1}^n m_i \dot{x}_i^2 + U \right) \right] \end{aligned} \quad (20)$$

This form of the probability density function is known as the Maxwell-Boltzmann distribution, being the expression inside the parentheses the total mechanical energy of the system.

Equivalent Linearization

The very restrictive conditions that make available exact solutions for the stochastic dynamical systems motivated the development of approximate solution techniques. Methods such as the equivalent linearization have been developed to generate first-order approximate solutions. These techniques can also be applied in some cases to single-degree-of-freedom nonlinear oscillator with hysteretic behavior.

The adaptation of the classical equivalent linearization technique of the deterministic theory to systems subjected to random excitations was independently developed by Booton (1953) and Caughey (1953). Later on other researchers have extended the method to encompass approximate solutions of the stationary random response of multi-degree-of-freedom nonlinear oscillators. In this case the mathematical equation that describes the response of the system is

$$M\ddot{\underline{x}} + C\dot{\underline{x}} + K\underline{x} + \underline{f}(\underline{x}, \dot{\underline{x}}) = \underline{G} \quad (21)$$

where M , C , and K are the constant mass, damping, and stiffness square matrices with dimension n , $\underline{f}(\underline{x}, \dot{\underline{x}})$ is an n -vector function of the dependent variable $\underline{x}$ and its derivative, and

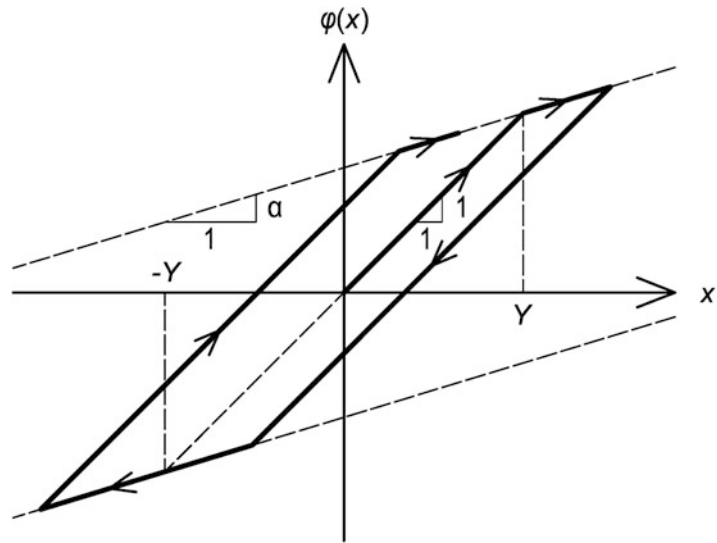
$$\underline{G}^T = (g_1, \dots, g_n) \quad (22)$$

with n stationary Gaussian processes g_i .

The principle of the method is the replacement of the nonlinear dynamical system described by Eq. 21 by an auxiliary linear system for which the exact analytic solution is known. The replacement is made so as to be optimum with respect to some measure of the difference between the original and the auxiliary system.

Stochastic Analysis of Nonlinear Systems,

Fig. 4 Bilinear hysteretic restoring force



This optimum auxiliary linear system, which will be called *equivalent*, is defined as

$$M\ddot{\underline{x}} + (C + C_e)\dot{\underline{x}} + (K + K_e)\underline{x} = \underline{G} \quad (23)$$

where C_e and K_e are t -independent matrices. These matrices must be such that the difference $\underline{d}$ between systems of Eqs. 21 and 23, defined by

$$\underline{d} = \underline{f}(\underline{x}, \dot{\underline{x}}) - C_e\dot{\underline{x}} - K_e\underline{x}, \quad (24)$$

is minimized for every $\underline{x}$ which belongs to the class of solutions of the system (23).

Since the excitation $\underline{G}$ of the linear system is Gaussian, it is well known that the response $\underline{x}$ will be Gaussian as well. Therefore, the matrices C_e and K_e must be such that the difference $\underline{d}$ is minimized for every stationary Gaussian random vector $\underline{x}$.

The equivalent linearization technique is thus composed of the following steps:

1. Identification of a class of approximate solution functions and the parameters defining each member of the class.
2. Selection of the norm of the difference vector $\underline{d}$.
3. Selection of the averaging operator G .
4. Determination of the matrices C_e and K_e of the equivalent linear system in terms of the identification parameters of $\underline{x}$.

5. Solution of the equivalent linear system to obtain equations for the specification of the identification parameters of $\underline{x}$.

The application of the method can be conveniently shown by analyzing the hysteretic stochastic dynamical single DOF system described by the following differential equation (Iwan and Lutes 1968):

$$\ddot{x} + 2\beta_0\omega_0\dot{x} + \omega_0^2\varphi(x) = \frac{n(t)}{m} \quad (25)$$

where m denotes the mass, ω_0 is the undamped natural circular frequency for small-amplitude vibrations, β_0 is the fraction of critical damping for small-amplitude vibrations, and $\varphi(x)$ is the bilinear hysteretic restoring force, shown in Fig. 4, having a unit slope for small amplitudes and a second slope α . The excitation $n(t)$ is a stationary random function with a uniform power spectral density S_0 and a Gaussian probability distribution.

The equivalent linear system

$$\ddot{x} + 2\beta_{eq}\omega_{eq}\dot{x} + \omega_{eq}^2x = \frac{n(t)}{m} \quad (26)$$

may be identified by choosing the parameters ω_{eq} and β_{eq} so as to minimize the mean-squared

difference between the nonlinear inelastic equation of motion (25) and the linear equation (26).

For the case of bilinear hysteretic oscillator with “small” nonlinearity, a solution for the parameters ω_{eq} and β_{eq} has been found by Caughey (1960) under the following assumptions:

1. The response of the nonlinear system is contained within a narrow frequency band.
2. The probability density of the amplitude of this narrowband response follows the Rayleigh distribution.

These assumptions lead to the following expressions for ω_{eq} and β_{eq} :

$$\left(\frac{\omega_{eq}}{\omega_0}\right)^2 = 1 - \left[\frac{8(1-\alpha)}{\pi}\right] \int_1^\infty \left(2^{-3} + \frac{1}{\lambda z}\right) (z-1)^{\frac{1}{2}} e^{-\frac{2}{z}} dz \quad (27)$$

and

$$\beta_{eq} = \beta_0 \left(\frac{\omega_0}{\omega_{eq}}\right) + \left(\frac{\omega_0}{\omega_{eq}}\right)^2 \frac{(1-\alpha)}{\sqrt{\pi\lambda}} \operatorname{erfc}\left(\lambda^{-\frac{1}{2}}\right) \quad (28)$$

where

$$\lambda = \frac{2\sigma_x^2}{Y^2}$$

In general Eq. 27 must be evaluated numerically. However, when $\lambda \gg 1$, the asymptotic expansion shown in Eq. 29 can be used:

$$\left(\frac{\omega_{eq}}{\omega_0}\right)^2 = \alpha + \left[\frac{8(1-\alpha)}{\pi}\right] \left(0.6043\lambda^{-\frac{3}{4}} - 0.2451\lambda^{-\frac{5}{4}} - 0.1295\lambda^{-\frac{7}{4}}\right) \quad (29)$$

After finding ω_{eq} and β_{eq} , the RMS levels of response of the equivalent linear system can be obtained from

$$\sigma_x^2 = \omega_{eq}^2 \sigma_x^2 = \frac{\pi S_0}{4m^2 \beta_{eq} \omega_{eq}} \quad (30)$$

Perturbation Method

The random response of slightly nonlinear vibrating systems can also be obtained by applying the classical perturbation method (Crandall 1963). The method is based on the assumption that the nonlinearity is small enough to allow the solution of the stochastic differential equation of motion to be expressed as a power series. If the following SDOF system is considered

$$\ddot{x} + 2\zeta\omega_0\dot{x} + \omega_0^2 x + \varepsilon\eta(x, \dot{x}) - f(t) = 0 \quad (31)$$

where η is a nonlinear function and ε is a sufficiently small parameter, the solution can be expressed as a power series in ε :

$$x(t) = x_0(t) + \varepsilon x_1(t) + \varepsilon^2 x_2(t) + \dots \quad (32)$$

Substituting Eq. 32 into Eq. 31 and grouping the different terms having the same power of ε , a set of linear equation in $x_0, x_1, x_2, \dots$ can be obtained. If the external excitation $f(t)$ is weakly stationary, the steady-state response to Eq. 31 can be constructed from the following solutions of the set of linear equations:

$$\begin{aligned} x_0(t) &= \int_{-\infty}^{\infty} f(t-\tau)h(\tau)d\tau \\ x_1(t) &= \int_{-\infty}^{\infty} \eta[x_0(\tau), \dot{x}_0(\tau)]h(\tau)d\tau \\ x_2(t) &= \int_{-\infty}^{\infty} \eta[x_1(\tau), \dot{x}_1(\tau)]h(\tau)d\tau \\ &\dots \end{aligned} \quad (33)$$

being h the impulse response function corresponding to $\varepsilon = 0$.

Stochastic Averaging

The method of stochastic averaging, firstly introduced by Landau and Stratonovich (1962), has proved to be a very useful tool for deriving approximate solutions to problems involving the dynamical response of lightly damped systems to broadband random excitation. It is based on the principle that the rate of change with respect to time of the oscillator's total energy is equal to the power input due to random excitation, minus the power dissipation due to the damping mechanism.

In lightly damped structures the slowly varying energy can be treated as a constant over an appropriate period of oscillation, and oscillatory terms can be approximated by their time averages over one period of oscillation. Furthermore, under broadband random excitation, the relaxation time of the oscillator response is much greater than the correlation time of the excitation. Thus, it is possible to model the power input due to the excitation as a nonzero mean component plus an additional, fluctuating component with the character of white noise.

Thanks to these properties, the method can also be applied to strongly nonlinear stiffness oscillators and to certain nonlinearities of the hysteretic kind.

The essence of the standard stochastic averaging method is embodied in a limit theorem due to Stratonovich (1963) and Khasminskii (1966):

$$\dot{X} = \varepsilon^2 f(X, t) + \varepsilon g(X, t, Y(t)) \quad (34)$$

where $X(t)$ is an n -vector stochastic process, usually representing the response, and $Y(t)$ is an m -vector stochastic excitation process. If the elements of $Y(t)$ are broadband processes, with zero means, and the vectors f and g satisfy certain requirements (which are almost invariably met in practice), then it can be shown that $X(t)$ may be uniformly approximated over a time interval of order $O(\varepsilon^{-1})$ by an n -dimensional Markov process, which satisfies the Ito equation:

$$dX = \varepsilon^2 m(X)dt + \varepsilon \sigma(X)dW \quad (35)$$

The symbol $W(t)$ denotes an n -vector of independent Wiener (or Brownian) processes, with unit variance, and m and σ are, respectively, the “drift vector” and the “diffusion matrix.”

If ε is small, then the elements of $X(t)$ must be slowly varying, with respect to time. The equations of motion which occur in random vibration problems can be written in state space form, involving displacement and velocity response variables. However, these variables are usually rapidly fluctuating with respect to time. Evidently a transformation of variables is required to cast the equations of motion into the form of Eq. 34.

It has been shown that the stochastic averaging method is applicable to an oscillator with bilinear restoring force-displacement characteristic when the energy dissipation due to hysteresis is relatively low. In this case, after a preliminary averaging operation, the equation of motion could be cast into the standard form of Eq. 34, enabling the stationary response distribution to be determined together with statistics such as the standard deviation of the displacement and the average yielding rate.

Nevertheless, the standard stochastic averaging technique cannot be used for examining the effect of strongly nonlinear restoring forces since, to $O(\varepsilon^2)$, this effect vanishes. In these cases it is possible to combine the equivalent linearization method with stochastic averaging and treat the equivalent frequency as amplitude dependent.

An alternative approach to the analysis of nonlinear oscillators based on a consideration of the energy envelope has been developed by Roberts and Spanos (1986).

Consider an oscillator with the following equation of motion:

$$\ddot{x} + \varepsilon^2 h(x, \dot{x}) + G(x) = \varepsilon z(t) \quad (36)$$

with ε assumed small enough to ensure that $x(t)$ is $O(\varepsilon^0)$. The symbol $G(x)$ denotes an arbitrary nonlinear stiffness term, while the energy envelope $V(t)$ may be defined as the sum of the kinetic and the potential energy:

$$V(t) = \frac{\dot{x}^2}{2} + U(x) \quad (37)$$

and

$$U(x) = \int_0^x G(\xi) d\xi \quad (38)$$

For the special case where the damping is linear $\varepsilon^2 h(x, \dot{x}) = 2\zeta\omega_0\dot{x}$ and the nonlinear stiffness has the power-law form

$$G(x) = k|x|^v \text{sgn}(x), \quad (39)$$

it is possible to obtain analytical solution for the transition density function $p(V, t|V_1, t_1)$ of $V(t)$. By introducing the nondimensional energy variable

$$\chi(t) = \frac{V(t)}{\gamma k \sigma^{v+1}}, \quad (40)$$

the following transition density $p(\chi, t|\chi_1, t_1) = p(\chi, \tau|\chi_1)$ can be found:

$$p(\chi, \tau|\chi_1) = \frac{1}{(1-q)} \left(\frac{\chi}{\chi_1 q} \right)^{\frac{\rho}{2}} \exp \left\{ -\frac{\chi + q\chi_1}{1-q} \right\} I_{\rho} \left\{ \frac{2\sqrt{\chi\chi_1 q}}{1-q} \right\} \quad (41)$$

being

$$\rho = \frac{1}{\alpha} - 1 \quad (42)$$

$$q = e^{-2\alpha\zeta\omega_0 t} \quad (43)$$

and with $I_{\rho}(\cdot)$ as the modified ρ - order Bessel function of the first kind.

Path Integral Solution

The FPK equation that describes the evolution of the response's probability density (PD) of a nonlinear system excited by an external white noise can be solved numerically by path integral (PI) solution procedures. In essence the PI method is a stepwise calculation of the joint probability density function of a set of state space variables describing a white-noise-excited nonlinear dynamic system. Among the first efforts to develop the PI method into numerical tools are those of Wehner and Wolfer (1983), Sun and Hsu (1990), and Naess and Johnsen (1993).

The PI method has been proved to provide extremely accurate results for the tail behavior of the joint probability density function of the state space vector and thus for the estimation of extreme responses of nonlinear dynamical systems excited by forces, external or parametric, that can be approximated as white noise, filtered

white-noise processes, and combined normal and Poisson white noise (Pirrotta and Santoro 2011).

The PI method is based on the fact that the state space vector, Y_t say, obtained as a solution of a stochastic differential equation is a Markov vector process. This makes it possible to use a time stepping procedure to obtain the joint probability density function $p(y, t)$ of Y_t as a function of time t by exploiting the fundamental equation:

$$p(y, t) = \int_{-\infty}^{\infty} p(y, t|y', t') p(y', t') dy' \quad (44)$$

where $p(y, t|y', t')$ denotes the conditional probability density function of Y_t given that $Y_{t'} = y'$. For small time increments $\Delta t = t - t'$, $p(y, t|y', t')$ will be referred to as the short-time transition probability density function. It can be demonstrated that for a numerical solution of a stochastic differential equation, the short-time transition probability density function can always be given as an analytical, closed-form expression. Hence, if an initial probability density function, $p_0(y) = p(y, t = 0)$, is given, then Eq. 44 can be invoked repeatedly to produce the time evolution of $p(y, t)$. If the stochastic differential equation has an invariant measure there exists a stationary probability density function $p_s(y)$, then, eventually, assuming that $p_0(y) \neq p_s(y)$, $p(y, t)$ will approach this stationary probability density function. The number of times Eq. 44 has to be repeatedly used to reach the stationary situation depends, of course, on the dynamic system and on the specified initial probability density function $p_0(y)$.

Monte Carlo Simulation Methods

Generality

The probabilistic structure and the statistical moments of the response of any type of nonlinear mechanical systems can be evaluated using simulation techniques, like the Monte Carlo method. This method operates in the time domain by repeating a great number of times deterministic

analyses, each one consisting in a step-by-step nonlinear analysis of the structure subjected to an earthquake record.

In order to evaluate properly the statistics of the response, each earthquake record must belong to the same “family,” in the sense that all the records must be compatible with the same power spectral density (PSD) function.

The major advantage of Monte Carlo simulation is that it can deal with almost any type of nonlinearity, maybe using the same commercially available packages that are used for the deterministic analyses. Its only disadvantage is that it is time-consuming, as many analyses are requested in order to have reliable estimates of the statistical properties of the response.

The number of needed runs may be many thousands, even if 300–400 analyses can give acceptable results in practical applications.

Probably the most important and delicate part of Monte Carlo simulation is the generation of realizations of the stochastic process or field that represents the earthquake.

In principle the earthquakes may be modeled as nonstationary non-Gaussian random fields, even if the null-mean Gaussian approximation is considered acceptable by most authors.

If the analyzed system is a building, it is reasonable, due to the reduced area of ground that it occupies, to model the process as one-dimensional. If the hypothesis that the earthquake induces ground acceleration in one direction only may be accepted, the process is univariate.

In the case of structures that have a notable extension over the ground, like long span bridges, the earthquake input is not the same in different part of the system and the process must be considered as multidimensional. If only one-direction component of the ground acceleration is considered, it is also univariate. Otherwise it is multivariate.

Univariate One-Dimensional Stochastic Processes

The basic method to generate a realization of a univariate, one-dimensional (1V-1D) stationary stochastic process $f_0(t)$ with zero mean and one-sided PSD function $G_{f_0}(\omega)$ was proposed by Shinozuka (Shinozuka 1972).

The process may be simulated by the series

$$f(t) = \sqrt{2} \sum_{n=0}^{N-1} A_n \cdot \cos(\omega_n t + \Phi_n) \quad (45)$$

where

$$A_n = \sqrt{G_{f_0}(\omega_n) \Delta\omega}, \quad n = 0, 1, 2, \dots, N-1 \quad (46)$$

$$\omega_n = n\Delta\omega, \quad (47)$$

$$\Delta\omega = \frac{\omega_u}{N} \quad (48)$$

ω_u is the upper cutoff frequency and the phase angle Φ_n is the realization of a random variable uniformly distributed over the interval $[0, 2\pi]$.

Shinozuka and Deodatis (1991) showed that it must be

$$A_0 = 0 \text{ and } S_{f_0}(\omega_0 = 0) = 0 \quad (49)$$

to ensure that the realization of the process is ergodic.

Univariate Multidimensional Stochastic Processes

For the sake of simplicity, let us consider the special case of a univariate two-dimensional stochastic process (1V-2D), which may represent the random field of the one-direction ground acceleration interesting a great-extension structure frozen in time.

The process $f_0(x_1, x_2)$, having cross-PSD $G_{f_0}(k_{1n_1}, k_{2n_2})$, can be simulated by the series

$$f(x_1, x_2) = \sqrt{2} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} \left[A_{n_1 n_2} \cos(k_{1n_1} x_1 + k_{2n_2} x_2 + \Phi_{n_1 n_2}^{(1)}) + \tilde{A}_{n_1 n_2} \cos(k_{1n_1} x_1 - k_{2n_2} x_2 + \Phi_{n_1 n_2}^{(2)}) \right] \quad (50)$$

where

$$G_{f_0}(k_1, k_2), \tilde{A}_{n_1 n_2} = \sqrt{G_{f_0}(k_{1n_1}, -k_{2n_2}) \Delta k_1 \Delta k_2} \quad (51)$$

$$k_{1n_1} = n_1 \Delta k_1, \quad k_{2n_2} = n_2 \Delta k_2, \quad (52)$$

$$\Delta k_1 = \frac{k_{1u}}{N_1}, \quad \Delta k_2 = \frac{k_{2u}}{N_2} \quad (53)$$

$$n_1 = 0, 1, 2, \dots, N_1 - 1, \quad n_2 = 0, 1, 2, \dots, N_2 - 1 \quad (54)$$

and

$$A_{0n_2} = A_{n_1 0} \text{ for } n_1 = 0, 1, 2, \dots, N_1 - 1 \text{ and } n_2 = 0, 1, 2, \dots, N_2 - 1 \quad (55)$$

$$\tilde{A}_{0n_2} = \tilde{A}_{n_1 0} \text{ for } n_1 = 0, 1, 2, \dots, N_1 - 1 \text{ and } n_2 = 0, 1, 2, \dots, N_2 - 1 \quad (56)$$

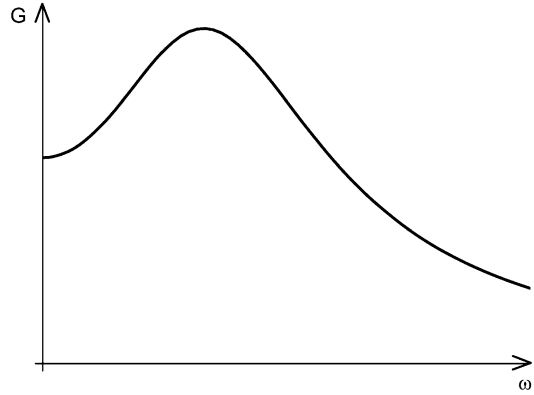
k_{1u} and k_{2u} are the upper cutoff wave numbers and the phase angles $\Phi_{n_1 n_2}^{(1)}$ and $\Phi_{n_1 n_2}^{(2)}$ are the realization of two statistically independent random variables uniformly distributed over the interval $[0, 2\pi]$.

The extension to the case of 1D-3V process $f_0(x_1, x_2, t)$, including also the time variability, is straightforward.

Earthquake Power Spectra

A well-known PSD of the ground motion available in literature is the so-called Kanai-Tajimi spectrum. It is based on the hypothesis that the ground acceleration during earthquakes may be considered as a filtered band-limited white-noise process expressed by the following function:

$$G(\omega) = \frac{[1 + 4\xi_g^2 \cdot (\omega/\omega_g)^2]}{[1 - (\omega/\omega_g)^2]^2 + 4\xi_g^2 \cdot (\omega/\omega_g)^2} G_0 \quad (57)$$



Stochastic Analysis of Nonlinear Systems, Fig. 5 Kanai-Tajimi power spectrum

where G_0 is the constant value of the spectrum of the white-noise process and ω_g and ξ_g are the predominant ground frequency and the ground damping.

The values of $\omega_g = 4\pi$ and $\xi_g = 0.60$ have been suggested as representative of earthquakes on firm ground.

The typical aspect of the Kanai-Tajimi PSD is shown in Fig. 5.

Anyway in most cases it is better to derive the PSD from the response spectrum.

For this purpose let us consider the following SDOF deterministic dynamical system:

$$\ddot{x}(t) + 2\xi\omega\dot{x}(t) + \omega x(t) = f_0(t) \quad (58)$$

acted upon by a 1V-1D stationary stochastic process $f_0(t)$ which represents the ground acceleration. If the process is Gaussian, its probabilistic structure is fully represented by its PSD G_{f_0} .

The random dynamic analysis of the system in the frequency domain allows finding the PSD of the response $x(t)$, which is the system displacement:

$$G_x(\omega) = G_{f_0}(\omega) |H(\omega)|^2 \quad (59)$$

The variance of the response is represented by the area under the PSD function:

$$\sigma_x^2 = \int_0^\infty G_{f_0}(\omega) |H(\omega)|^2 d\omega \quad (60)$$

where $|H(\omega)|^2$ is the squared frequency-domain transfer function between loading and system response:

$$|H(\omega)|^2 = \frac{1}{(\omega_n^2 - \omega^2)^2 + 4\xi^2 \omega_n^2 \omega^2} \quad (61)$$

with ω_n the natural frequency of the system.

The variance of the pseudo-acceleration response is

$$\begin{aligned} \sigma_{\ddot{x}}^2 &= \omega_n^4 \sigma_x^2 = \omega_n^4 \int_0^\infty G_{f_0}(\omega) |H(\omega)|^2 d\omega \\ &\approx G_{f_0}(\omega_n) \omega_n \left(\frac{\pi}{4\xi} - 1 \right) \end{aligned} \quad (62)$$

The ordinates $\ddot{x}_{s,p}$ of the acceleration response spectrum are the maxima of the response $\ddot{x}(t)$ of an oscillator whose frequency ω varies from 0 to some stipulated maximum value. The parameter “s” is the duration of the strong motion and “p” is the non-exceedance probability of the maxima.

The solution of the equation

$$\ddot{x}_{s,p} = r_{s,p} \sigma_{\ddot{x}} \quad (63)$$

implies the evaluation of the peak factor $r_{s,p}$ by solving the corresponding “first-passage” problem.

First of all, it must be reminded that the moment of order i of the generic one-sided PSD $S(\omega)$ is

$$\lambda_i = \int_0^\infty \omega^i S(\omega) d\omega \quad (64)$$

while a measure where the spectral mass is concentrated along the frequency axis is

$$\Omega = \sqrt{\lambda_2/\lambda_0} \quad (65)$$

parameter that resembles the root mean square of a random variable.

A measure of the spread of the PSD function is

$$\delta = \sqrt{1 - \lambda_1^2/\lambda_0\lambda_2} \quad (66)$$

Given the above, an approximate expression for the peak factor $r_{s,p}$ is given by

$$r_{s,p} = \sqrt{2 \log \left\{ 2n \left[1 - \exp \left(-\delta_{\ddot{x}}(s) \sqrt{\pi \log 2n} \right) \right] \right\}} \quad (67)$$

where

$$n = \frac{\Omega_{\ddot{x}}(s) s_0 / 2\pi}{-\log p} \quad (68)$$

$$s_0 = s \cdot \exp \left[-2 \left\{ \sigma_{\ddot{x}}^2(s) / \sigma_{\ddot{x}}^2(s/2) - 1 \right\} \right] \quad (69)$$

$\Omega_{\ddot{x}}(s)$ and $\delta_{\ddot{x}}(s)$ are the spectral moments of the response.

Then the PSD function is

$$\begin{aligned} G_{f_0}(\omega_n) &\approx \frac{1}{\omega_n \left(\frac{\pi}{4\xi_s} - 1 \right)} \\ &\sqrt{\frac{\omega_n^2 S_v(\omega_n)_{s,p}^2}{r_{s,p}^2} - \int_0^{\omega_n} G_{f_0}(\omega) d\omega} \end{aligned} \quad (70)$$

where ξ_s is the time-dependent damping for the duration s

and S_v is the velocity response spectrum.

As the PSD $G_{f_0}(\omega)$ appears on both sides of Eq. 65, an iterative computational procedure must be used.

Summary

The present chapter is devoted to the probabilistic analysis of the random response of nonlinear structural systems exposed to the random excitation due to earthquakes. The system response is nonlinear due to the nonlinear character of the

material properties and to the effects of large structural displacements.

The attention is focused on systems having deterministic properties, including deterministic boundary conditions, taking into account those types of nonlinearity that can really occur during earthquakes.

After a brief review of the types of nonlinear behaviors that can be expected, some available methods for the analysis of the response are discussed.

They can be classified into analytical and simulation methods. Among the first ones are presented: the Fokker-Planck-Kolmogorov (FPK) equation, the equivalent linearization, the perturbation method, and the stochastic averaging. The simulation methods based on Monte Carlo techniques are then presented.

Cross-References

- [Plastic Hinge and Plastic Zone Seismic Analysis of Frames](#)
- [Reinforced Concrete Structures in Earthquake-Resistant Construction](#)
- [Stochastic Analysis of Linear Systems](#)

References

- Atalik T, Utku S (1976) Stochastic linearization of multi-degree of freedom nonlinear systems. *Earthquake Eng Struct Dyn* 4(4):411–420
- Booton R (1953) The analysis of nonlinear control systems with random inputs. Polytechnic Inst, Brooklyn
- Bouc R (1967) Forced vibrations of mechanical systems with hysteresis. *Proceedings of the Fourth Conference on Nonlinear Oscillation*, Prague
- Breccolotti M, Materazzi A, Venanzi I (2008) Identification of the nonlinear behavior of a cracked RC beam through the statistical analysis of the dynamic response. *Struct Control Health Monitor* 15:416–435
- Caughey T (1953) Response of nonlinear systems to random excitation. *Lecture Notes California Institute of Technology*, Pasadena, Calif
- Caughey T (1960) Random excitation of a system with bilinear hysteresis. *J Appl Mech* 27:649–652
- Crandall S (1963) Perturbation techniques for random vibration of nonlinear systems. *J Acoust Soc Am* 35(11):1700–1705
- Fokker A (1914) The median energy of rotating electrical dipoles in radiation fields. *Annalen Der Physik* 43:810–820
- Iwan WD (1973) A generalization of the concept of equivalent linearization. *Int J Nonlinear Mech* 8:279–287
- Iwan WD, Lutes LD (1968) Response of the bilinear hysteretic system to stationary random excitation. *J Acoust Soc Am* 48(3):545–552
- Khasminskii R (1966) A limit theorem for the solution of differential equations with random right-hand sides. *Theory Probab Appl* 11:390–405
- Kolmogorov A (1931) Ober analytische Methoden in der Wahrscheinlichkeitsrechnung. *Math Ann* 104:415–458
- Landau P, Stratonovich R (1962) *Theory of stochastic transitions of various systems between different states*. Moscow University, Vestnik
- Lin Y, Cai G (1995) *Probabilistic structural dynamics advanced theory and applications*. McGraw-Hill, New York
- Masing G (1926) *Eigenspannungen und Verfestigung beim Messing* (in German). In *Proceedings of the 2nd international congress of applied mechanics*, Zurich, pp 332–335
- Naess A, Johnsen J (1993) Response statistics of nonlinear, compliant offshore structures by the path integral solution method. *Probab Eng Mech* 8(2): 91–106
- Park R, Kent D, Sampson R (1972) Reinforced concrete members with cyclic loading. *ASCE J Struct Div* 98(7):1341–1360
- Pirrotta A, Santoro R (2011) Probabilistic response of nonlinear systems under combined normal and Poisson white noise via path integral method. *Prob Eng Mech* 26(1):26–32
- Piszczec K, Nizioł J (1986) *Random vibration of mechanical system*. Ellis Horwood, Chichester
- Planck M (1915) *Sgr. preuss. Akad. Wiss.* p 512
- Popov E, Bertero V, Krawinkler H (1972) Cyclic behavior of three r.c. flexural members with high shear. In: *EERC report 72-5. Earthquake Engineering Research Center*, University of California, Berkeley
- Roberts J, Spanos P (1986) Stochastic averaging: an approximate method of solving random vibration problems. *Int J Nonlinear Mech* 21(2):111–134
- Shinozuka M (1972) Monte Carlo solution of structural dynamics. *Comput Struct* 2(5/6):855–874
- Shinozuka M, Deodatis G (1991) Simulation of stochastic processes by spectral representation. *Appl Mech Rev* 44(4):191–204
- Spanos P-TD (1976) *Linearization techniques for nonlinear dynamical systems*, EERL 76-04. California Institute of Technology, Pasadena, Calif
- Stratonovich R (1963) *Topics in the theory of random noise*. Gordon & Breach, New York
- Sun J, Hsu C (1990) The generalized cell mapping method in nonlinear random vibration based upon short-time Gaussian approximation. *J Appl Mech* 57: 1018–1025

- Takeda T, Sozen M, Nielson N (1970) Reinforced concrete response to simulated earthquakes. *Proc ASCE J Struct Div* 96(ST12):2257–2573
- VanMarcke E, Yanev P, De Estrada M (1970) Response of simple hysteretic systems to random excitation, s.l., Research report R70-66, Department of Civil Engineering, MIT
- Wenhner M, Wolfer W (1983) Numerical evaluation of path-integral solutions to Fokker–Planck equations. *Phys Rev A* 27(5):2663–2670
- Wen Y (1976) Method for random vibration of hysteretic systems. *ASCE J Eng Mech* 102:2299–2325

Stochastic Finite Elements

Carsten Proppe

Institut für Technische Mechanik, Karlsruhe
Institute of Technology, Karlsruhe, Germany

Introduction

The finite element method (FEM) has become the dominant computational method in structural engineering. In general, the input parameters in the standard FEM assume deterministic values. In earthquake engineering, at least the excitation is often random. However, considerable uncertainties might be involved not only in the excitation of a structure but also in its material and geometric properties. A rational treatment of these uncertainties needs a mathematical concept similar to that underlying the standard FEM. Thus, FEM as a numerical method for solving boundary value problems has to be extended to stochastic boundary value problems. The extension of the FEM to stochastic boundary value problems is called stochastic finite element method (SFEM).

The first developments of the SFEM can be traced back at least to Cornell (1970), who studied soil settlement problems, and to Shinozuka (Astill et al. 1972), who combined FEM with Monte Carlo simulation for reliability analysis of structures with random excitation, random material properties, or random geometric properties. He introduced random fields and discretized them based on spectral representation theory

(Shinozuka 1971) of wide-sense homogeneous random fields.

The expression SFEM came in use in the early 1980s (Contreras 1980; Baecher and Ingra 1981). Der Kiureghian and Ke (1988) defined SFEM as “a finite element method which accounts for uncertainties in the geometry or material properties of a structure, as well as the applied loads” where “the uncertainties are usually spatially distributed over the region of the structure and should be modelled as random or stochastic fields.” The distinguishing feature of an SFEM is that it involves the discretization of the random field and the computation of solution statistics.

Discretization methods for random fields replace the random field by a finite set of random variables. They can be broadly classified into two groups: direct discretization schemes and series expansion techniques. Direct discretization schemes either assign the values of the random field at a given set of nodes to the finite set of random variables (point discretization schemes) or compute the values from local averages of the random field over a spatial domain (local averaging schemes).

Possible sets of nodes for point discretization schemes are the nodes of the finite element mesh (Hisada and Nakagiri 1981; Liu et al. 1986), the centroids of the finite elements (Der Kiureghian and Ke 1988), or the Gauss points used by the integration rules of the finite elements (Brenner and Bucher 1995). If the set of nodes for the point discretization scheme of the random field corresponds to the set of nodes of the finite elements, the shape functions for the representation of the random field may differ from those for the displacements (Liu et al. 1986). Point discretization schemes yield a positive definite correlation matrix that is easily computed. The distribution function of the random variables is the same as for the random field. Nearly all schemes can be applied to Gaussian as well as non-Gaussian random fields. The point discretization schemes that do not use different shape functions have the disadvantage that the FE mesh depends on the correlation structure of the random field and that the shape and size of all these elements should be the same. In general, the discretization produces a

huge number of random variables and leads to inefficient numerical procedures.

In contrast, local averaging schemes may yield accurate results even on rather coarse FE meshes. They can be based directly on averages of the random parameter field (Vanmarcke and Grigoriu 1984) or on integrals computed over the domain of a finite element, e.g., for the determination of the element stiffness matrix (Deodatis 1991). However, direct averaging yields random variables whose distribution functions are difficult to obtain (except for Gaussian random fields) and the approximation of non-rectangular elements may lead to a nonpositive definite covariance matrix (Matthies et al. 1997), and averaging based on element integration introduces again a dependence of the FE mesh on the correlation structure.

Series expansion techniques represent the random field by a series involving countably infinite random variables and a complete set of deterministic functions. The discretization is obtained by truncation of the series expansion. Lawrence (Lawrence 1987) considers a series expansion with an a priori set of orthogonal deterministic base functions that are multiplied by random variables. In the Karhunen-Loève (KL) expansion, the set of orthogonal base functions is obtained as eigenfunctions of a homogeneous Fredholm integral equation of second kind that involves the covariance kernel of the random field. The KL expansion is optimal in the sense that it reduces the mean-square error resulting of its truncation. In principle, it can be applied to homogeneous as well as inhomogeneous random fields and to Gaussian as well as non-Gaussian random fields. However, in practical cases, the KL expansion is applied to Gaussian random fields almost exclusively because the random variables in the expansion are then independent and standard normally distributed. In general, the eigenfunctions have to be computed numerically and the accuracy of the numerical solution influences the accuracy of the random field representation. For strongly correlated random fields with smooth covariance kernel, few terms of the series expansion are sufficient to represent the random field.

However, according to Stefanou and Papadrakakis (2007), homogeneity and ergodicity of sample functions generated by the KL expansion are questionable. The spectral representation method approximates a homogeneous Gaussian random field by a superposition of harmonics with fixed frequencies and random amplitudes or random frequencies and amplitudes. Due to the central limit theorem, the generated samples are Gaussian, if a sufficient number of harmonics (at least 128, according to Stefanou (2009)) are superposed.

In some cases, non-Gaussian random fields can be represented by nonlinear memoryless transformations of Gaussian random fields (Yamazaki and Shinozuka 1988). The memoryless transformation introduces a compatibility condition between the marginal distribution function and the autocorrelation function of the non-Gaussian random field. If this condition is not satisfied, only an approximation of the non-Gaussian random field by a nonlinearly transformed Gaussian random field is possible. Iterative procedures have been proposed (Deodatis and Micaletti 2001; Yamazaki and Shinozuka 1988) that calibrate the power spectral density of the Gaussian random field in order to approximately match the marginal distribution function and the autocorrelation function of the non-Gaussian random field. The underlying Gaussian random field can then be discretized by any of the methods described above.

A series expansion of a non-Gaussian random field in terms of independent random variables is obtained by projection on a set of orthogonal polynomials. This method is known as polynomial chaos expansion and is discussed extensively in Ghanem and Spanos (1991), where Hermite polynomials and Gaussian random variables are considered. In Xiu and Karniadakis (2002), other families of orthogonal polynomials and corresponding distributions of the random variables are discussed. The polynomial chaos expansion is convergent in mean-square sense. Field and Grigoriu (2004) critically discuss polynomial chaos expansions and show that the accuracy of the PC approximation is not always improved as additional terms are retained and

that the polynomial chaos expansion might be computationally demanding due to the large number of expansion coefficients that have to be computed.

Once the random field involved in the stochastic boundary value problem has been discretized, a solution method has to be adopted in order to solve the boundary value problem numerically. The choice of the solution method depends on the required statistical information of the solution. If only the first two statistical moments of the solution are of interest (*second moment analysis*), the perturbation method can be applied. However, if a *full probabilistic analysis* is necessary, Galerkin schemes can be utilized or one has to resort to Monte Carlo simulations eventually in combination with a von Neumann series expansion.

The perturbation method starts with a Taylor series expansion of the solution, the external loading, and the stochastic stiffness matrix in terms of the random variables introduced by the discretization of the random parameter field. The unknown coefficients in the expansion of the solution are obtained by equating terms of equal order in the expansion. From this, approximations of the first two statistical moments can be obtained. The perturbation method is computationally more efficient than direct Monte Carlo simulation. However, higher-order approximations will increase the computational effort dramatically, and therefore accurate results are obtained for small coefficients of variation only.

In the spectral SFEM, the random parameter fields are discretized by a KL or a polynomial chaos expansion, the solution is expanded with Hermite polynomials, and a Galerkin approach is applied to solve for the unknown expansion coefficients. The theoretical foundation has been laid in Deb et al. (2001) and Babuška et al. (2005), where local and global polynomial chaos expansions for linear elliptic boundary value problems with stochastic coefficients were investigated and where a priori error estimates have been proved for a fixed number of terms of the KL expansion.

Instead of a Galerkin projection, several authors employed collocation schemes for the

determination of the unknown coefficients in the approximation scheme (Acharjee and Zabarar 2007; Babuška et al. 2007; Baroth et al. 2007; Huang et al. 2007). This leads to nonintrusive algorithms that allow to combine the solution procedure with repetitive runs of a finite element (FE) solver for deterministic problems. A nonintrusive algorithm based on least squares regression has been presented recently in Berveiller et al. (2006).

In the following, discretization methods for the random parameter field are illustrated, and a mathematical theory for the approximate solution of stochastic elliptic boundary value problems involving a discretized random parameter field that is represented as a superposition of independent random variables is outlined. In the random domain, global and local polynomial chaos expansions are employed. The relation between local approximations of the solution and Monte Carlo simulation is considered, and reliability assessment is briefly discussed. Finally, an example serves to illustrate the different solution procedures.

Discretization of Random Fields

Let D be a convex bounded open set in $\mathbb{R}^n$ and $(\Omega, \mathcal{F}, P)$ be a complete probability space, where Ω is the set of outcomes, $\mathcal{F}$ the σ -field of events, and $P : \mathcal{F} \rightarrow [0 : 1]$ a probability measure. A function $\alpha : D \times \Omega \rightarrow \mathbb{R}^m$ is a *random field*, if $\alpha(x, \omega)$ is a random variable for any $x \in D$. In the following, we consider scalar valued random fields ($m = 1$).

The finite dimensional distribution of order q of α at $x_1, x_2, \dots, x_q \in D$ is the probability of the set $\cap_{i=1}^q \{\alpha(x_i, \omega) \leq \alpha_i\}$. The random field is *homogeneous* (in strict sense), if the finite dimensional distributions are invariant under a space shift and thus depend only on the space lag.

Suppose that the random field α is square integrable on $D \times \Omega$ and denote by $E[X] = \int_{\Omega} x dP$ the expectation of the random variable $X(\omega)$. Then, the mean, correlation, and covariance are given by $E[\alpha(x, \omega)]$, $R(x, y) = E$

$[\alpha(x, \omega)\alpha(y, \omega)]$, and $C(x, y) = E[(\alpha(x, \omega) - E[\alpha(x, \omega)])(\alpha(y, \omega) - E[\alpha(y, \omega)])]$, respectively. A random field is *homogeneous* (in weak sense), if the mean is constant and the correlation depends only on the space lag $y - x$.

Bochner's theorem allows to introduce the spectral distribution $S(v)$ of a weakly homogeneous random field with continuous correlation function by

$$R(x) = \int_{\mathbb{R}^n} \exp(ix \cdot v) dS(v). \quad (1)$$

If S is absolutely continuous, the Radon-Nikodym derivative $s(v)$ is called *spectral density function*: $dS(v) = s(v) dv$.

Spectral Representation

If α is a weakly homogeneous mean-square continuous random field, it can be represented by

$$\alpha(x, \omega) = \int_{\mathbb{R}^n} \exp(ix \cdot v) dW(v), \quad (2)$$

where the random field $W(v)$ has mean zero and satisfies $E[dW(v) dW(\mu)^*] = \delta(v - \mu) dS(v)$. The asterisk denotes complex conjugation and δ is the Dirac δ -distribution.

This distribution can be used to approximately represent a homogeneous Gaussian random field by a superposition of harmonics. For a real-valued random field, one obtains the representation

$$\alpha(x, \omega) = \int_{\mathbb{R}^n} (\cos(v \cdot x) dU(v) + \sin(v \cdot x) dV(v)), \quad (3)$$

where $E[dU(v)^2] = E[dV(v)^2] = dS(v)$. Starting from a partition of the wave number domain, the increments $dU(v)$ and $dV(v)$ are approximated by

$$\begin{aligned} \Delta U(v_i) &= \sqrt{2\Delta S} \cos \phi_i; \\ \Delta V(v_i) &= \sqrt{2\Delta S} \sin \phi_i; \\ \Delta S &= S(v_{i+1}) - S(v_i) \end{aligned} \quad (4)$$

where ϕ_i are random variables uniformly distributed on $[0, 2\pi]$. A representation involving a finite number of random variables is thus

$$\sum_{i=1}^M \sqrt{2\Delta S} \cos(v_i \cdot x + \phi_i). \quad (5)$$

As a consequence of the central limit theorem, it converges for $M \rightarrow \infty$ to a Gaussian random field with the same mean value and autocorrelation structure as the target Gaussian random field.

Karhunen-Loève Expansion

Due to the properties of the covariance function, the operator $T : L^2(D) \rightarrow L^2(D)$,

$$T_u = \int_D C(x, y) u(y) dy, \quad (6)$$

is compact and self-adjoint and thus admits a spectrum of decreasing nonnegative eigenvalues $\{\lambda_i\}_{i=1}^\infty$. The corresponding eigenfunctions $\{\phi_i(x)\}_{i=1}^\infty$ are orthonormal in $L^2(D)$. The random variables given by

$$\xi_i(\omega) = \frac{1}{\sqrt{\lambda_i}} \int_D (\alpha(x, \omega) - E[\alpha](x)) \phi_i(x) dx \quad (7)$$

are uncorrelated (but in general not independent), have zero mean and unit variance, and allow to represent the random field by the KL expansion

$$\alpha(x, \omega) = E[\alpha](x) + \sum_{i=1}^\infty \sqrt{\lambda_i} \xi_i(\omega) \phi_i(x) \quad (8)$$

that converges in $L^2(D \times \Omega)$ (Loève 1977). Conditions for stronger convergence properties are given in Babuška et al. (2005). The KL expansion is usually truncated by retaining only the first M terms. In order to keep the computational effort small, a fast decay of the spectrum of Eq. 6 is important. It is shown in Todor and Schwab (2006) that fast eigenvalue decay corresponds to smoothness of the covariance function. Moreover, for a decreasing correlation length, the number M of retained terms increases if the accuracy of

the approximation is kept constant. The KL expansion reduces to the spectral representation method for homogeneous random fields defined over an infinite domain (Huang et al. 2001).

For a prescribed, uniformly bounded random field $\alpha(x, \omega)$, the random variables $\xi_i(\omega)$ in Eq. 8 would be dependent non-Gaussian random variables whose joint distribution function is very difficult to identify. If, on the other hand, independent but bounded distributions are prescribed for $\xi_i(\omega)$, the random field $\alpha(x, \omega)$ is not necessarily bounded for $N \rightarrow \infty$. Thus, one is left with Gaussian distributions for $\xi_i(\omega)$ and $\alpha(x, \omega)$, with transformations of Gaussian random fields or with some situations, where nonnegative distributions for $\xi_i(\omega)$ lead to meaningful (e.g., Erlang) distributions for $\alpha(x, \omega)$.

The KL expansion is optimal in the sense that the error measured in $L^2(D \times \Omega)$ resulting from a truncation after M terms is smaller than for any other linear combination of M functions. It is a representation for homogeneous as well as nonhomogeneous random fields. However,

several authors (Grigoriu 2006; Stefanou and Papadrakakis 2007; Sudret and Der Kiureghian 2000) observed problems regarding the homogeneity of samples generated from the truncated expansion.

Kriging

Application of Kriging to the discretization of random fields has been introduced by Li and Der Kiureghian (1993). The random field $\alpha(x, \omega)$ is approximated by a linear function of M nodal values $\alpha_i(\omega) = \alpha(x_i, \omega)$

$$\tilde{\alpha}(x, \omega) = \phi_0(x) + \sum_{i=1}^M \alpha_i(\omega) \phi_i(x). \quad (9)$$

The functions $\phi_i(x)$, $i = 0, \dots, M$ are determined by minimizing in each point x the variance of the error $\alpha(x, \omega) - \tilde{\alpha}(x, \omega)$ under the condition that $\tilde{\alpha}(x, \omega)$ is an unbiased estimator of $\alpha(x, \omega)$. This yields

$$\begin{aligned} E[\alpha(x, \omega)] &= \phi_0(x) + \sum_{i=1}^M E[\alpha_i(\omega)] \phi_i(x), \\ \sum_{i=1}^M \phi_i(x) \text{Cov}[\alpha_i(\omega) \alpha_j(x)] &= \text{Cov}[\alpha(x, \omega) \alpha_j(\omega)], \quad j = 1, 2, \dots, M, \end{aligned} \quad (10)$$

with the covariance operator

$$\text{Cov}[X, Y] = E[(X - E[X])(Y - E[Y])]. \quad (11)$$

Li and Der Kiureghian (1993) introduced a spectral decomposition of the nodal covariance matrix. They showed that the maximum error of the KL expansion is not always smaller than the error of Kriging for a given number of retained terms. The point-wise variance error estimator of the KL expansion for a given order of truncation is smaller than the error of Kriging in the interior of the discretization domain but larger at the boundaries. Note however that the

KL expansion provides the lowest mean error over the domain.

Polynomial Chaos Expansion

In contrast to the KL expansion, the polynomial chaos expansion does not need an a priori knowledge of the covariance structure of the random field. A functional representation of the random field in terms of a vector of basic random variables ξ ,

$$\alpha(x, \omega) = f(x, \xi), \quad (12)$$

is necessary. Such a representation is given for the solution field of a stochastic boundary value

problem, if the random parameter field has been discretized.

A nonlinear expansion of the functional $f(x, \xi)$ is obtained by projecting it on a set of polynomials $\{\Gamma_p\}$ the basic random variables ξ . The space spanned by $\{\Gamma_p\}$ is called the p th homogeneous chaos. The polynomial Γ_p , called the polynomial chaos of order p , is a polynomial of order p that is orthogonal to all polynomials with order less than p .

Assuming symmetry of the polynomials, which is always possible (Ghanem and Spanos 1991), the random field can be approximated by

$$\begin{aligned} \tilde{\alpha}(x, \omega) = & \phi_0(x) + \sum_{i_1=1}^N \phi_{i_1}(x) \Gamma_1(\xi_{i_1}) \\ & + \sum_{i_1=1}^N \sum_{i_2=1}^{i_1} \phi_{i_1 i_2}(x) \Gamma_2(\xi_{i_1}, \xi_{i_2}) \\ & + \sum_{i_1=1}^N \sum_{i_2=1}^{i_1} \sum_{i_3=1}^{i_2} \phi_{i_1 i_2 i_3}(x) \Gamma_3(\xi_{i_1}, \xi_{i_2}, \xi_{i_3}) + \dots \end{aligned} \quad (13)$$

or more briefly by

$$\tilde{\alpha}(x, \omega) = \sum_{i=1}^M \Phi_i(x) \Psi_i(\xi). \quad (14)$$

The expansion is convergent in the mean-square sense. For Gaussian random variables, the polynomial chaos expansion is described in

more detail in Ghanem and Spanos (1991). Xiu and Karniadakis (2002) extended the approach to general families of orthogonal polynomials, the Wiener-Askey chaos.

Field and Grigoriu (2004) pointed out some limitations of polynomial chaos approximations. They demonstrated that the convergence rate of polynomial chaos approximations may be slow, that the accuracy of the polynomial chaos approximation is not always improved by adding terms, that higher-order moments may be inaccurate, and that the polynomial chaos approximations of homogeneous non-Gaussian processes may not be homogeneous.

Transformation Techniques for Non-Gaussian Random Fields

Transformation techniques for non-Gaussian random fields seek to represent the non-Gaussian random field as a nonlinear transformation of a Gaussian random field:

$$\alpha(x, \omega) = F^{-1}(\Phi(g(x, \omega))), \quad (15)$$

where Φ is the standard Gaussian cumulative distribution function, F is the non-Gaussian marginal cumulative distribution function of $\alpha(x, \omega)$, and $g(x, \omega)$ is the underlying Gaussian random field.

The transformation imposes a correlation structure to $\alpha(x, \omega)$, namely,

$$R(\xi) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F^{-1}(\Phi(u)) F^{-1}(\Phi(v)) \phi(u, v, R_g(\xi)) du dv, \quad (16)$$

where $\phi(u, v, R_g(\xi))$ denotes the joint Gaussian probability density function. If the correlation structure of $\alpha(x, \omega)$ does not match the prescribed values, one has to resort to nonlinear transformations that match the target marginal distribution and/or the correlation structure approximately (Grigoriu 1998).

The nonlinear transformation technique can be combined with any series expansion schemes described above for the underlying Gaussian

random field. It allows to calculate analytically many important quantities, such as crossing rates and extreme value distributions.

Phoon et al. (2002, 2005) used the KL expansion for the simulation together with an iterative mapping scheme to fit the target marginal distribution function of non-Gaussian random fields. The method allows to simulate homogeneous as well as nonhomogeneous random fields.

Stochastic Linear Elliptic Boundary Value Problems

Consider the following model problem with stochastic operator and deterministic input on $\bar{D} \times \Omega$: find $\bar{D} \times \Omega \rightarrow \mathbb{R}$, such that P -almost surely

$$\begin{aligned} -\nabla \cdot \alpha(x, \omega) \nabla u(x, \omega) &= f(x) \quad \text{on } D, \\ u(x, \omega) &= 0 \quad \text{on } \partial D. \end{aligned} \quad (17)$$

It is assumed that the deterministic input function $f(x)$ is square integrable and that the random field $\alpha: D \times \Omega \rightarrow \mathbb{R}$ is bounded and coercive, i.e., there exists positive constants $a_{\min}, a_{\max}$, such that

$$P(\omega \in \Omega : a_{\min} < \alpha(x, \omega) < a_{\max} \quad \forall x \in D) = 1 \quad (18)$$

and that the random field has a continuous and square-integrable covariance function.

We are interested in the probability that a functional $F(u)$ of the solution $u(x, \omega)$ exceeds a threshold F_0 , i.e., we want to evaluate the integral

$$P_F = \int_{\Omega} \chi_{(F_0, \infty)}(F(u(x, \omega))) dP(\omega), \quad (19)$$

where $\chi_I(\cdot)$, the indicator function, assumes the value 1 in the interval I and vanishes elsewhere.

The variational formulation of the stochastic boundary value problem necessitates the introduction of the Sobolev space $H_0^1(D)$ of functions having generalized derivatives in $L^2(D)$ and vanishing on the boundary ∂D with norm $\|u\|_{H_0^1(D)} = \left(\int_D |\nabla u|^2 dx \right)^{1/2}$, the space $L_P^2(\Omega)$ of square-integrable random variables, and the tensor product space $H_0^1(D) \otimes L_P^2(\Omega)$ of $H_0^1(D)$ -valued random fields with finite second order moments, equipped with the inner product

$$(u, v)_{H_0^1(D) \otimes L^2(\Omega)} = \int_{\Omega} \int_D \nabla u(x, \omega) \cdot \nabla v(x, \omega) dx dP(\omega). \quad (20)$$

The variational formulation of the stochastic linear elliptic boundary value problem Eq. 17 then

reads: find $u \in H_0^1(D) \otimes L_P^2(\Omega)$, such that for all $v \in H_0^1(D) \otimes L_P^2(\Omega)$,

$$\int_{\Omega} \int_D \alpha(x, \omega) \nabla u \cdot \nabla v dx dP(\omega) = \int_{\Omega} \int_D f(x) v(x, \omega) dx dP(\omega). \quad (21)$$

The assumptions on the random field $\alpha(x, \omega)$ guarantee the continuity and coercivity of the bilinear form in Eq. 21, and thus, the existence and uniqueness of a solution to Eq. 21 follows from the Lax-Milgram lemma.

Numerical Solution of the Stochastic Boundary Value Problem

In a first step, the random parameter field is discretized and replaced by a finite sum of random variables. Assume that a suitable approximation is given by a linear combination of continuous and independent random variables $\xi_i(\omega)$ with zero mean and unit variance,

$$\alpha_M(x, \omega) = E[\alpha](x) + \sum_{i=1}^M \xi_i(\omega) \phi_i(x), \quad (22)$$

where $\Gamma_i = \xi_i(\Omega)$ are bounded intervals in $\mathbb{R}$.

Under these assumptions, the stochastic variational problem involving the random field $\alpha_M(x, \omega)$ has the following deterministic equivalent: find $u \in H_0^1(D) \otimes L_P^2(\Gamma)$, such that

$$\begin{aligned} \int_{\Gamma} \int_D \alpha_M(x, y) \nabla_x u(x, y) \cdot \nabla_x v(x, y) dx p(y) dy \\ = \int_{\Gamma} \int_D f(x) v(x, y) dx p(y) dy, \end{aligned} \quad (23)$$

for all $v \in H_0^1(D) \otimes L_P^2(\Gamma)$, where $p \in L^\infty(\Gamma)$ is the joint probability density functions of the random variables $\xi_i(\omega)$, $i = 1, 2, \dots, M$; see Babuška et al. (2005). Here, $\Gamma = \prod_{i=1}^M \Gamma_i \subset \mathbb{R}^M$. This variational formulation is now discretized on finite dimensional approximation spaces. For $H_0^1(D)$, a family of standard finite element approximation spaces $X_h \subset H_0^1(D)$ of continuous piecewise linear functions in a regular

triangulation T_h of D with mesh parameter h is considered. Following Deb et al. (2001), discontinuous finite elements are applied on the domain Γ , which is partitioned into elements $\gamma = \prod_{i=1}^M (a_i^\gamma, b_i^\gamma)$, with $(a_i, b_i) \subset \Gamma_i$. Elements of $Y_k^q \subset L_p^2(\Gamma)$ are functions that are polynomials of degree at most q (i.e., $y_1^{q_1} y_2^{q_2} \dots y_M^{q_M} \in Y_k^q$ if $q_1 + q_2 + \dots + q_M \leq q$) when restricted to each element $\gamma \subset \Gamma$. The parameter $k = (k_1, k_2, \dots, k_M)$, with $k_i = \max_{\gamma \in \Gamma} |b_i^\gamma - a_i^\gamma|$, represents the mesh parameter. If the partition consists of a single element only and the degree q is varied, global approximations are obtained.

Denote with $N_i(x)$, $i = 1 \dots N$ a basis of X_h and with $\psi_k(y)$, $k = 1 \dots P$, a basis of Y_k^q . The solution $u(x, y)$ is approximated on $X_h \otimes Y_k^q$ by

$$\sum_{i=1}^N \sum_{k=1}^P u_{ik} N_i(x) \psi_k(y). \quad (24)$$

In order to determine the unknown coefficients u_{ik} , this representation is inserted into Eq. 23 together with test functions $v(x, y) = N_j(x) \psi_l(y)$, $j = 1, 2, \dots, N$, $l = 1, 2, \dots, P$, yielding

$$\begin{aligned} \sum_{j=1}^P \int_{\Gamma} (K^{(0)} + \sum_{s=1}^M K^{(s)} y_s) \psi_j(y) \psi_l(y) p(y) dy u_j \\ = \int_{\Gamma} \psi_l(y) p(y) dy, \end{aligned} \quad (25)$$

where

$$\begin{aligned} K_{ij}^{(0)} &= \int_D Ex \nabla N_i(x) \cdot \nabla N_j(x) dx, \\ K_{ij}^{(s)} &= \int_D \phi_s(x) \nabla N_i(x) \cdot \nabla N_j(x) dx, \\ f_i &= \int_D f(x) N_i(x) dx, \quad i, j = 1, 2, \dots, N, \end{aligned} \quad (26)$$

and u_j is the $N \times 1$ dimensional matrix obtained from u_{ij} for fixed value of j . In structural mechanics, the matrices $K^{(s)}$, $s = 0, 1, \dots, M$ can be

interpreted as finite element stiffness matrices with a spatial variation of Young's modulus.

If the basis $\{\psi_k\}_{k=1}^P$ consists of discontinuous finite element base functions, the system of equations Eq. 25 decouples and can be treated separately for each element γ . Moreover, due to the presence of terms in one single variable y_s , $s = 1, 2, \dots, M$ on the left-hand side of Eq. 25, it is possible to construct the basis $\{\psi_k\}_{k=1}^P$ in a way that the problem decouples into problems that have the same size as the deterministic FE problem (Babuška et al. 2005). To see this, consider the polynomial basis $\{\psi_i^\gamma\}_{i=1}^{P_\gamma}$ in one element γ only. This basis can be constructed by multiplying polynomials in one single variable y_s , $s = 1, 2, \dots, M$. Let $P(y)$ be the monomial basis $(1, y, y^2, \dots, y^q)^T$. We want to find a transformation matrix $S_\gamma^{(s)}$, such that the transformed basis $\psi_\gamma^{(s)}(y) = S_\gamma^{(s)} P(y)$ satisfies

$$\begin{aligned} \int_{a_s^\gamma}^{b_s^\gamma} \psi_i^{(s)}(y) \psi_j^{(s)}(y) p_s(y) dy &= \delta_{ij} \text{ and} \\ \int_{a_s^\gamma}^{b_s^\gamma} \psi_i^{(s)}(y) \psi_j^{(s)}(y) y_s p_s(y) dy &= \lambda_i^{(s)} \delta_{ij}. \end{aligned} \quad (27)$$

Defining the matrices $A_\gamma^{(s)}$ and $B_\gamma^{(s)}$ by

$$\begin{aligned} A_{\gamma ij}^{(s)} &= \int_{a_s^\gamma}^{b_s^\gamma} P_i(y) P_j(y) y p_s(y) dy \\ B_{\gamma ij}^{(s)} &= \int_{a_s^\gamma}^{b_s^\gamma} P_i(y) P_j(y) p_s(y) dy, \end{aligned} \quad (28)$$

it is easily seen that $\lambda_i^{(s)}$ are the eigenvalues and $S_\gamma^{(s)}$ is the matrix of the eigenvectors of the generalized eigenvalue problem

$$A_\gamma^{(s)} S_\gamma^{(s)T} = B_\gamma^{(s)} S_\gamma^{(s)T} \Lambda, \text{ with } S_\gamma^{(s)} B_\gamma^{(s)} S_\gamma^{(s)T} = I. \quad (29)$$

Now, by introducing multiindices $j = (j_1, j_2, \dots, j_M)$ for the basis function $\psi_{ji} = \prod_{s=1}^M \psi_{j_s}^{(s)}(y_s)$ on each element γ , we find from Eq. 25

$$\left(K_\gamma^{(0)} + \sum_{s=1}^M K_\gamma^{(s)} \right) u_j = f_\gamma \int_\gamma \psi_{ji}(y) p(y) dy, \quad 1 \leq j_i \leq q + 1, \quad i = 1, 2, \dots, M \quad (30)$$

due to the orthogonality properties Eq. 27 of this basis.

As can be seen from this equation, the parallelization of the algorithm for computing the local chaos approximations is easily possible due to the fact that:

1. The approximations on each element γ are independent.
2. The linear systems for the expansion coefficients decouple due to the choice of biorthogonal polynomials instead of Hermite polynomials.

As a consequence, any degree of parallelization (from coarse grained to fine grained) is possible, depending on the number of processors at disposal. Note, however, that the introduction of biorthogonal polynomials is possible only if the random variables ξ_i are independent and that approximations with biorthogonal polynomials require in general an upper limit q for the polynomial degree in each single variable y_s , $s = 1, \dots, M$, i.e., $y_1^{q_1} y_2^{q_2} \dots y_M^{q_M}$ is an approximation function, if $q_s \leq q$, $s = 1, 2, \dots, M$. For the same level q , this truncation leads to much more expansion coefficients than requiring $q_1 + q_2 + \dots + q_M \leq q$, especially if M is large.

Relationship with Monte Carlo Simulation

There is a close correspondence between Monte Carlo simulation and local polynomial approximations. The equations for Monte Carlo simulation can be obtained if instead of the Galerkin approximation, a collocation method is applied with respect to Γ , leading to

$$\left(\mathbf{K}^{(0)} + \sum_{s=1}^M \mathbf{K}^{(s)} y_s^j \right) u(y^j) = f, j = 1, 2, \dots, N, \quad (31)$$

with N sampling points y^j , $j = 1, 2, \dots, N$, where finite element solutions are obtained.

If the vectors $u(y^j)$ are computed by means of Eq. 31, it is not obvious how to interpolate the solution in probability space for other values of y . A simple interpretation in terms of a local approximation would be a partition of Γ into N subdomains, where each subdomain contains the nearest neighbors of the sampling point y_j . In each subdomain, $u(y)$ is then approximated by a constant value, namely, the value of the sampling point associated with the subdomain. If the number of sampling points tends to infinity, the error of the approximation vanishes.

This approach can be related to Latin hypercube sampling, if the number of subdomains γ is very high and the approximations are not computed for every subdomain of the partition. For computing n samples according to the Latin hypercube sampling method, Γ is partitioned into n^N subdomains of equal size. Then, the $n \times N$ matrix Π containing random permutations of $1, \dots, n$ and an $n \times N$ matrix G of independent and uniformly over $[0, 1]$ distributed random numbers are constructed, and the sampling plan $S = (\Pi - R)/n$ is established. Mapping the elements of the sampling plan to Γ via $\xi_{ij} = P_G^{-1}(S_{ij})$, $i = 1, \dots, n$, $j = 1, \dots, N$, where $P_G^{-1}(\cdot)$ is the inverse of the standard Gaussian distribution, one computes the solution only in those subdomains, in which the sample $\xi_i = (\xi_{i1}, \dots, \xi_{iN})$, $i = 1, \dots, n$ falls. In this way, a hybrid method, i.e., a combination of sampling techniques with approximation techniques, is obtained that leads to a considerable reduction of the sampling variance.

Evaluation of Response Quantities

Once the algebraic problem is solved and the approximation coefficients have been determined, an expression of the displacement field that depends on the input random variables has been obtained. This expression can be considered as a response surface. This response surface has local character (Proppe 2008) and depends on the size and location of the elements γ , if a partition of Γ is adopted.

Approximations for the moments of $u(x, \omega)$ can be computed by evaluating this expression

with respect to the input random variables. Computation of approximations for the distribution of $u(x, \omega)$ or the failure probability Eq. 19 is a more complex task, and resort to Monte Carlo simulation via the obtained expression for $u(x, \omega)$ seems to be the easiest way to accomplish it.

For solving reliability problems, it is very helpful to consider the approximation for the displacements as a local response surface, which leads to a functional relationship between the input random variables and $u(x, \omega)$. It is then possible to compute the most probable point of failure (MPP), i.e., the point ξ with $F(u(x, \xi)) = F_0$ with lowest Euclidean norm, and to refine the grid at its vicinity. In this way, it is possible to solve the reliability problem with a high degree of accuracy in an adaptive manner.

The MPP may also be useful for the evaluation of the integral in Eq. 19 by means of variance-reduced Monte Carlo simulation (importance sampling). To this end, a sampling density $\tilde{p}(y)$ is introduced by shifting the original probability density function $p(y)$ of the random variables ξ_i , $i = 1, \dots, M$, to the previously obtained MPP, and Eq. 19 is approximated by

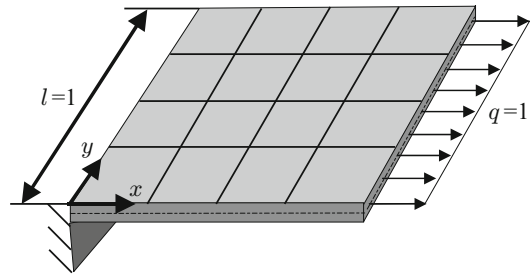
$$P_F \approx \sum_{j=1}^N \chi(F_{0,\infty})(F(u(x, y^j))) \frac{p(y^j)}{\tilde{p}(y^j)} \tilde{p}(y^j), \quad (32)$$

where the sampling points y^j , $j = 1, 2, \dots, M$ are generated according to $\tilde{p}(y)$ and $u(x, y^j)$ is computed from the approximation on the element that contains y^j .

The accuracy of the MPP is influenced by the FE mesh, the truncation level M of the random parameter field, the partition of Γ , and the choice of the ansatz functions both in spatial and random domain. These parameters can be gradually adapted such that the MPP is computed with a prescribed accuracy.

Example

In order to illustrate the stochastic finite element techniques, consider a clamped thin square plate under uniform in-plane tension (cf. Ghanem and



Stochastic Finite Elements, Fig. 1 Thin square plate under uniform in-plane tension

Spanos 1991 and Fig. 1). The product of Young's modulus and the thickness of the plate is assumed to be an isotropic Gaussian random field with covariance function

$$C(x_1, y_1; x_2, y_2) = \sigma \exp\left(-\frac{|x_1 - x_2|}{l_c} - \frac{|y_1 - y_2|}{l_c}\right), \quad (33)$$

standard deviation $\sigma = 0.2$, correlation length $l_c = 1$, and unit mean value. The Poisson's ratio is 0.3. The plate has unit length and the external excitation is deterministic and of unit magnitude. The longitudinal displacement of one of the free corners is considered in the following.

Approximation of Higher-Order Moments: Skewness and Kurtosis

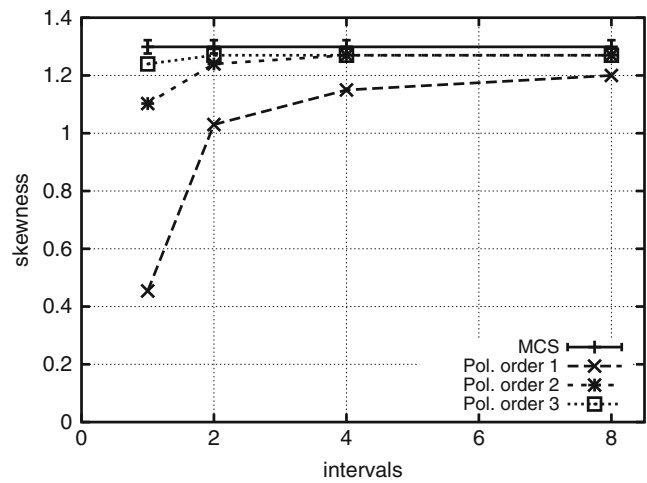
The random field has been represented by two random variables ($M = 2$). Figures 2 and 3 display the approximations obtained with polynomials up to third order for the skewness and the kurtosis of the displacement of the free corner with increasing number of intervals of equal probability. Reference values have been obtained by Monte Carlo simulation with $30 \cdot 10^6$ samples. For one interval, the result corresponds to a global approximation. It can be seen that global approximations are rather inexact for the skewness and kurtosis and that it is sufficient to split the intervals into two parts in order to improve the results significantly.

Hybrid Method

Figures 4 and 5 display relative (with respect to the estimated values) 95 % confidence intervals

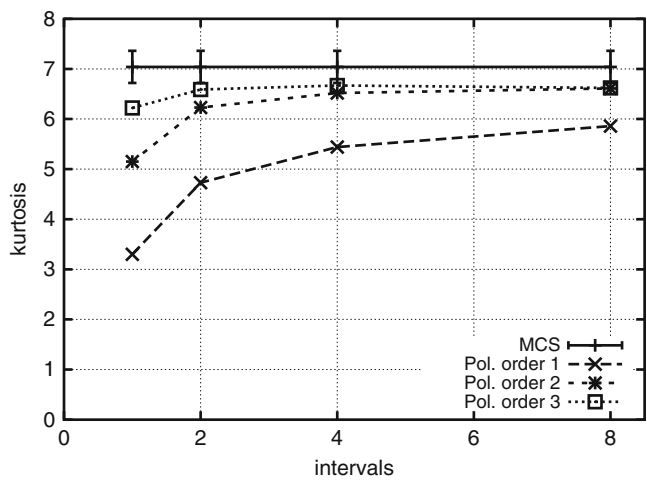
Stochastic Finite Elements,

Fig. 2 Approximation of the skewness



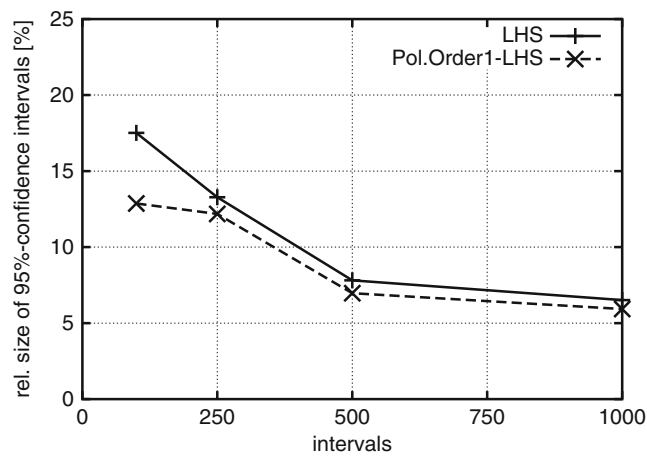
Stochastic Finite Elements,

Fig. 3 Approximation of the kurtosis



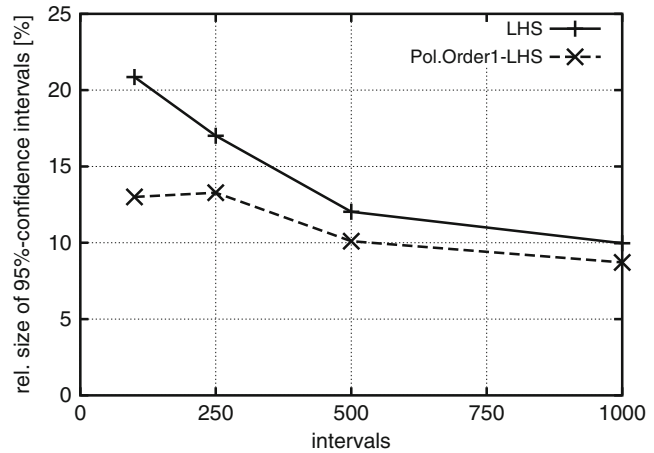
Stochastic Finite Elements,

Fig. 4 Approximation of the skewness by Latin hypercube sampling techniques



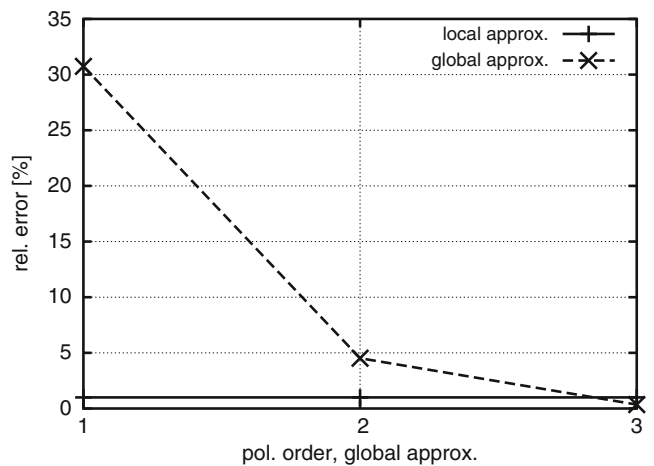
Stochastic Finite Elements,

Fig. 5 Approximation of the kurtosis by Latin hypercube sampling techniques



Stochastic Finite Elements,

Fig. 6 Accuracy of estimates for the MPP



for the estimation of the skewness and kurtosis obtained from Latin hypercube sampling and Latin hypercube sampling combined with linear approximation in the sampling intervals. A three-term ($M = 3$) representation of the random field by truncated Gaussian random variables has been employed. For a low to moderate number of intervals, the confidence intervals obtained by the hybrid method are narrower than those of the standard Latin hypercube sampling.

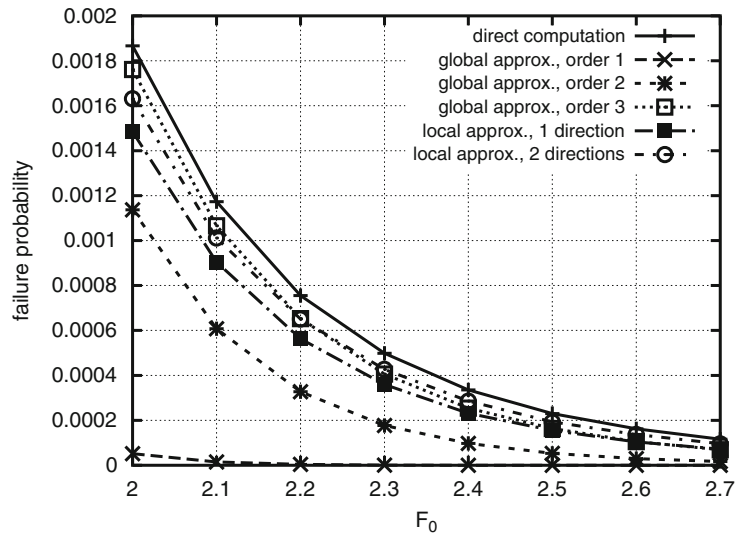
Reliability Assessment

For most reliability approximation techniques, it is necessary to compute the MPP. In order to obtain this point, a constrained optimization problem has to be solved. This implies expensive function calls (e.g., FE solutions) for the

computation of the displacements. These computations can be bypassed, if the approximations obtained from the numerical solution of the stochastic boundary value problem are employed as response surfaces. Figure 6 compares the relative error (i.e., the relative Euclidean distance) between the “true” MPP computed with FE calls and approximations obtained with local and global response surfaces of the displacement. The random field has been represented by four random variables ($M = 4$). The functional $F(u)$ is the value of the displacement field at the free corner and the threshold value F_0 has been set to 2.0. In order to obtain local approximations, an approximation of the MPP has been computed with a linear global approximation, and an interval of length 1σ has been inserted at that point for

Stochastic Finite Elements,

Fig. 7 Prediction of the failure probability



the coordinate with largest partial derivative of the longitudinal displacement for the free corner at the MPP. Linear polynomials have been applied for local approximations. From Fig. 6, one can see that this method yields very accurate approximations for the MPP, while for global approximations, polynomials of third degree are necessary in order to achieve a comparable accuracy.

From this fact, a corresponding approximation quality of the failure probability can be deduced, if variance-reduced sampling techniques which rely on the MPP together with the abovementioned response surface techniques are employed. This can be clearly seen from Fig. 7, where the probability of failure obtained with various approximation techniques has been plotted over the threshold value F_0 . Importance sampling at the predicted MPP with 30 batches of 10,000 samples has been employed for each failure probability estimate. For global approximations, again polynomials of third degree are necessary in order to obtain accurate predictions. On the other hand, local approximations with linear polynomials lead already to quite accurate results, if only the principle direction is partitioned.

In Fig. 8, the relative error of the failure probability for $F_0 = 2.3$ is displayed over the corresponding number of deterministic FE runs.

This allows to compare the efficiency of local and global approximations. From Fig. 8, it is evident that local approximations have considerable advantages over global approximations. This behavior is even more pronounced for smaller values of F_0 .

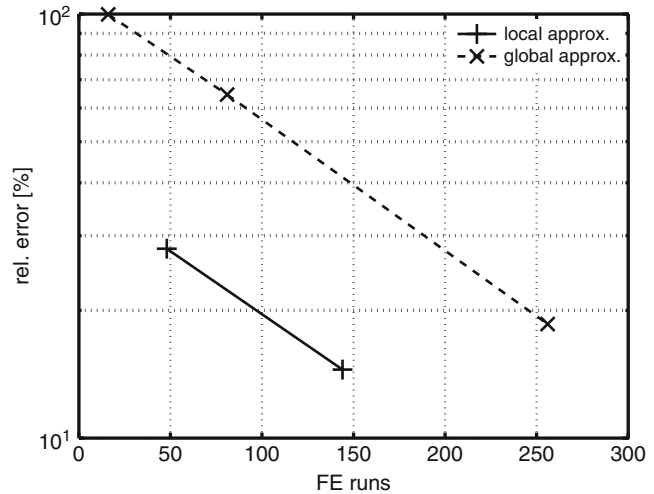
Finally, for $F_0 = 2.7$ (corresponding to a probability of failure of 8.2×10^{-5}) and a tolerance of 2 % for the Euclidean norm of the MPP, the procedure described in section “[Relationship with Monte Carlo Simulation](#)” yields $M = 4$ and a partition of Γ into 48 elements leading to an overall effort of 768 deterministic FE runs for the reliability estimation problem. This is considerably lower than 10^4 runs for importance sampling and 10^5 runs for a direct Monte Carlo simulation.

Concluding Remarks

For the SFE solution of stochastic boundary value problems, a mathematical theory is available that is in many aspects comparable to its deterministic counterpart, the finite element method. However, for random fields with short correlation lengths (requiring a high number M of random variables), the solution methods become inefficient, due to the series expansion of the solution. This is also true for most other SFE approximations, be they global or local, and recourse to efficient sampling

Stochastic Finite Elements,

Fig. 8 Efficiency of global and local approximation techniques for the prediction of the failure probability



techniques combined with efficient model reduction might be the only way to circumvent this problem.

A fundamental question that still has to be addressed in detail concerns the error of the solution due to the discretization of the random parameter field. Beyond this aspect of verification, the validation of the random field model itself, either from experimental data or from information pertaining to the microscale, remains an important issue (cf. the critique of the SFE method raised, e.g., in Ostoja-Starzewski 2011).

References

- Acharjee S, Zabaraz N (2007) A non-intrusive stochastic Galerkin approach for modeling uncertainty propagation in deformation processes. *Comput Struct* 85(5–6):244–254
- Astill CJ, Noseir SB, Shinozuka M (1972) Impact loading on structures with random properties. *J Struct Mech* 1:63–77
- Babuška IM, Tempone R, Zouraris GE (2005) Solving elliptic boundary value problems with uncertain coefficients by the finite element method: the stochastic formulation. *Comput Methods Appl Mech Eng* 194(1):1251–1294
- Babuška I, Nobile F, Tempone R (2007) A stochastic collocation method for elliptic partial differential equations with random input data. *SIAM J Numer Anal* 45:1005–1034
- Baecher GB, Ingra TS (1981) Stochastic FEM in settlement predictions. *J Geotech Eng Div* 107:449–463
- Baroth J, Bresslotte P, Chauvière C, Fogli M (2007) An efficient SFE method using Lagrange polynomials: application to nonlinear mechanical problems with uncertain parameters. *Comput Methods Appl Mech Eng* 196:4419–4429
- Berveiller M, Sudret B, Lemaire M (2006) Stochastic finite element: a non intrusive approach by regression. *Rev Eur Mécanique Numérique* 15:81–92
- Brenner CE, Bucher CG (1995) A contribution to the SFE-based reliability assessment of nonlinear structures under dynamic loading. *Probab Eng Mech* 10:265–273
- Contreras H (1980) The stochastic finite-element method. *Comput Struct* 12:341–348
- Cornell CA (1970) First order uncertainty analysis in soils deformation and stability. In: *Proceedings of first international conference of statistics and probability in soil and structural engineering*, Hong Kong
- Deb MK, Babuška IM, Oden JT (2001) Solution of stochastic partial differential equations using Galerkin finite element techniques. *Comp Methods Appl Mech Eng* 190:6359–6372
- Deodatis G (1991) Weighted integral method I: stochastic stiffness matrix. *J Eng Mech* 117:1851–1864
- Deodatis G, Micaletti RC (2001) Simulation of highly skewed non-Gaussian stochastic processes. *Trans ASCE, J Eng Mech* 127:1284–1295
- Der Kiureghian A, Ke JB (1988) The stochastic finite element method in structural reliability. *Probab Eng Mech* 3:83–91
- Field RV Jr, Grigoriu M (2004) On the accuracy of the polynomial chaos approximation. *Probab Eng Mech* 19(1–2):65–80
- Ghanem RG, Spanos PD (1991) *Stochastic finite elements: a spectral approach*. Springer, New York
- Grigoriu M (1998) Simulation of stationary non-Gaussian translation processes. *J Eng Mech* 124:121–126

- Grigoriu M (2006) Evaluation of Karhunen-Loève, spectral and sampling representations for stochastic processes. *J Eng Mech* 132:179–189
- Hisada T, Nakagiri S (1981) Stochastic finite element method developed for structural safety and reliability. In: *Proceedings of third international conference on structural safety and reliability*, Trondheim, pp 395–408
- Huang SP, Quek ST, Phoon KK (2001) Convergence study of the truncated Karhunen-Loève expansion for simulation of stochastic processes. *Int J Numer Methods Eng* 52:1029–1043
- Huang S, Mahadevan S, Rebba R (2007) Collocation-based stochastic finite element analysis for random field problems. *Probab Eng Mech* 22:194–205
- Lawrence M (1987) Basis random variables in finite element analysis. *Int J Numer Methods Eng* 24:1849–1863
- Li CC, Der Kiureghian A (1993) Optimal discretization of random fields. *J Eng Mech* 119:1136–1154
- Liu WK, Belytschko T, Mani A (1986) Probabilistic finite elements for nonlinear structural dynamics. *Comput Methods Appl Mech Eng* 56:61–81
- Loève M (1977) *Probability theory*. Springer, Berlin
- Matthies HG, Brenner CG, Bucher CG, Guedes Soares C (1997) Uncertainties in probabilistic numerical analysis of structures and solids – stochastic finite elements. *Struct Saf* 19:283–336
- Ostoj-Starzewski M (2011) Stochastic finite elements: where is the physics? *Theor Appl Mech* 38:379–396
- Phoon KK, Huang HW, Quek ST (2002) Simulation of second-order processes using Karhunen-Loève expansion. *Comput Struct* 80:1049–1160
- Phoon KK, Huang SP, Quek ST (2005) Simulation of strongly non-Gaussian processes using Karhunen-Loève expansion. *Probab Eng Mech* 20:188–198
- Proppe C (2008) Estimation of failure probabilities by local approximation of the limit state function. *Struct Saf* 30(4):277–290
- Shinozuka M (1971) Simulation of multivariate and multidimensional random processes. *J Acoust Soc Am* 49:357–367
- Stefanou G (2009) The stochastic finite element method: past, present and future. *Comput Methods Appl Mech Eng* 198:1031–1051
- Stefanou G, Papadarakakis M (2007) Assessment of spectral representation and Karhunen-Loève expansion methods for the simulation of Gaussian stochastic fields. *Comput Methods Appl Mech Eng* 196:2465–2477
- Sudret B, Der Kiureghian A (2000) Stochastic finite element methods and reliability – state of the art. Technical report, UCB/SEMM-2000/08, Department of Civil & Environmental Engineering, University of California, Berkeley
- Todor RA, Schwab C (2006) Convergence rates for sparse chaos approximations of elliptic problems with stochastic coefficients. Research Report No. 2006–05, Seminar für Angewandte Mathematik, ETH Zürich
- Vanmarcke E, Grigoriu M (1984) Stochastic finite element analysis of simple beams. *J Eng Mech* 109:1203–1214
- Xiu D, Karniadakis G (2002) The Wiener-Askey polynomial chaos for stochastic differential equations. *SIAM J Sci Comput* 24:619–644
- Yamazaki F, Shinozuka M (1988) Digital generation of non-Gaussian stochastic fields. *Trans ASCE, J Eng Mech* 114:1183–1197

Stochastic Ground Motion Simulation

Sanaz Rezaeian¹ and Xiaodan Sun²

¹U.S. Geological Survey, Golden, CO, USA

²Southwest Jiaotong University, Chengdu, China

Synonyms

Earthquake ground motion simulation; Nonstationary stochastic process; Stochastic models; Strong ground motion simulation; Synthetic accelerograms

Introduction

Strong earthquake ground motion records are fundamental in engineering applications. Ground motion time series are used in ► [response-history dynamic analysis](#) of structural or geotechnical systems. In such analysis, the validity of predicted responses depends on the validity of the input excitations. Ground motion records are also used to develop ground motion prediction equations (GMPEs) for intensity measures such as spectral accelerations that are used in ► [response-spectrum dynamic analysis](#). Despite the thousands of available strong ground motion records, there remains a shortage of records for large-magnitude earthquakes at short distances or in specific regions, as well as records that sample specific combinations of source, path, and site characteristics. The limited number of recordings has become problematic in the emerging field of

► [performance-based earthquake engineering](#) (PBEE), which considers the entire spectrum of structural response, from linear to grossly nonlinear and even collapse, and thereby requires ground motions with various levels of intensity for various earthquake design scenarios (e.g., a design scenario can be defined by the earthquake magnitude, distance, and site conditions). To obtain the desired ground motions for the purposes of PBEE, it is common engineering practice to scale or modify acceleration time series that were recorded during previous earthquakes to represent certain ground motion characteristics for the design of structural or geotechnical systems. However, scaling and modification methods can significantly alter other ground motion characteristics and result in unrealistic earthquake ground motion time series. Synthetic ground motions can be used instead to replace or supplement recorded motions when scarcity of previously recorded motions becomes a problem, provided they accurately capture characteristics of real earthquake ground motions and their natural variability.

Synthetic ground motions can be based on deterministic or stochastic simulations. A determinist model is one in which variables are uniquely determined, and the model performs the same way for a given set of initial conditions. Conversely, in a stochastic model, randomness is present and variables are not described by unique values, but rather by probability distributions. In earthquake engineering, deterministic ground motion simulation is commonly referred to as “physics-based” ground motion simulation. These simulation models synthesize the earthquake source by defining a source model (e.g., kinematic or dynamic rupture models) and describe the seismic wave travel path by defining a material model (e.g., seismic velocity model); then, they utilize numerical methods (e.g., finite element or finite difference methods) to estimate the solution to the wave propagation equation. In other words, these models explicitly incorporate the physics of the propagation of seismic waves. They produce realistic synthetic ground motions at low frequencies (typically $<1\text{ Hz}$). But they are computationally intensive due to the fine

discretization of the medium and require a thorough knowledge of the environment (i.e., earthquake source, crustal structure, material properties, basin and local site effects); this information is not available for many regions or is not accessible to engineers for well-studied regions. On the other hand, stochastic simulation models are simple and powerful tools that incorporate what is known about ground motion, source, path, and site characteristics into simple functional forms, and as a result are widely used to simulate earthquake ground motions. There is a specific entry dedicated to ► [physics-based ground motion simulation](#) in this encyclopedia by Taborda and Roten that provides more details on deterministic ground motion simulation models. Here, we focus on stochastic ground motion simulation.

A word of caution may be in order here about the terminology “physics-based.” The use of this term for deterministic models is mostly out of convenience and should not imply that stochastic models are not based on physics. In fact, some experts disagree with this terminology because many stochastic models are valid representations of the physics of earthquakes, and although they do not necessarily solve the mathematical problem of wave propagation, they integrate the physical characteristics of the earthquake source, path effects, and site effects implicitly through their formulations and parameters.

Both deterministic and stochastic simulations can be used for response-history dynamic analysis, but only stochastic simulations can be utilized for stochastic dynamic (i.e., random vibration) analysis, because the latter analysis method requires a random process model of the earthquake ground motion. Synthetic ground motions are particularly useful for nonlinear dynamic analysis due to the scarcity of recorded motions for large-magnitude earthquakes that are capable of causing nonlinear responses. Two approaches are available for nonlinear dynamic analysis of structures subjected to earthquakes: (1) nonlinear response-history analysis by the use of a selected set of ground motion time series and (2) nonlinear stochastic dynamic analysis by the use of a stochastic representation of the ground motion.

Well-developed methods and tools are available for nonlinear response-history analysis; a set of recorded or synthetic ground motions or a collection of both can be used for this type of analysis. Nonlinear stochastic methods are not as developed, but research in that direction is continuing (e.g., Au and Beck 2003; Der Kiureghian and Fujimura 2009). Several stochastic ground motion simulation models exist that can satisfy the requirements of these nonlinear stochastic methods. These methods typically require the input excitation to be represented in a discrete form and in terms of a finite number of random variables. For example, in Der Kiureghian and Fujimura (2009), it is necessary to represent the input excitation in the following form

$\mu(t) + \sum_{i=1}^n s_i(t)u_i$, where $\mu(t)$ is the deterministic mean of the random process, n is the discretization of the time series and provides a measure of the resolution, and u_i and $s_i(t)$ are, respectively, representatives of standard normal random variables and deterministic basis functions of time.

Regardless of the method of simulation, the resulting synthetic time series should properly represent the main characteristics of real earthquake ground motions. For most engineering applications, the intensity, duration, and frequency content of the input ground motion are the characteristics that control the response of structural and geotechnical systems. In addition to the duration of motion, which results from variation of the ground motion intensity over time, variation of the frequency content of the motion is also an important characteristic that influences the response, particularly in nonlinear analysis. Consequently, two important characteristics of real earthquake ground motions that separate them from a simple stochastic process (e.g., a ► white-noise signal) are temporal and spectral nonstationarities. Temporal nonstationarity refers to the nonstationarity of ground motion in the time domain (i.e., variation of intensity with time). Spectral nonstationarity refers to the nonstationarity of ground motion in the frequency domain (i.e., variation of frequency

content with time). Whereas most simulation methods account for temporal nonstationarity, which controls most common intensity measures and duration of motion, many ignore the spectral nonstationarity, which can be very important in nonlinear dynamic analysis.

Many synthetic ground motion models exist, but not all represent the natural variability of real earthquake ground motions. Some models underestimate the natural variability of ground motions because they underestimate the variability in their model parameters. On the other hand, if correlations between the model parameters are not taken into account or are underestimated, the variability of the resulting ground motions can be overestimated compared to real earthquake ground motions. In probabilistic analysis, it is important to capture the natural variability of ground motions in the simulations to accurately quantify the variation of response due to the variation in input excitation.

In summary, stochastic ground motion simulations have many applications. They can be used in linear and nonlinear response-history dynamic analysis as well as in stochastic dynamic analysis, or they can be used to develop GMPEs for various intensity measures. Stochastic models are generally simpler, more widely used, and more accurate at high frequencies than deterministic (physics-based) models. A good stochastic ground motion model must represent the temporal and spectral nonstationary characteristics of real earthquake ground motions and must properly represent their natural variability. More details on stochastic ground motion models, their classification, formulation, parameters, and practicality are given below. Discussions on adjustment of models for special situations including multicomponent, near-fault, and multi-station simulations are also provided.

Classification of Ground Motion Simulation Models

There are two main categories of models for generating synthetic ground motions, source-based models and site-based models.

Source-based models explicitly describe the fault rupture process at the source, the propagation of the resulting seismic waves through the ground medium, and the effects of local site conditions on the seismic waves to generate a time series at a specific site. Site-based models describe the ground motion time series as it is observed at a specific site by fitting a stochastic process to previously recorded motions with known earthquake and site characteristics; in this way, they implicitly account for the source, path, and site effects. Due to their need for empirical recordings of ground motions, site-based models are sometimes referred to as empirical models, but note that this terminology is not strictly correct because many source-based models also use data to empirically identify their parameters.

Source-Based Models

Source-based models can be deterministic or stochastic. An early review of source-based models is presented by Zerva (1988). Deterministic (physics-based) simulations were briefly described in the Introduction and have their own entry in this encyclopedia. They can produce realistic accelerograms at low frequencies (typically <1 Hz) but often need to be adjusted for high frequencies by combining with a stochastic component, resulting in hybrid models. For many years, source-based simulations have been utilized to develop GMPEs for locations where earthquakes are rare and the collection of recorded ground motions is sparse, for example, in the stable continental regions of the Central and Eastern United States (e.g., Somerville et al. 2001); research is continuing to improve these simulations and the resulting GMPEs.

One family of widely used synthetic ground motions are the source-based stochastic simulations based on the work of David Boore and a number of other researchers in the past several decades (e.g., Boore 2003; Beresnev and Atkinson 1998; Motazedian and Atkinson 2005; Boore 2009). This simulation method roots in the work of McGuire and Hanks (1980), which identifies the Fourier amplitude spectrum (FAS) of a ground motion considering the source, path, and

site effects and then combines it with a random phase spectrum. This method assumes ground motion to be a band-limited white Gaussian noise with finite duration. The white-noise is then adjusted to have the following FAS:

$$Y(M_0, f, R) = S(M_0, f) \times P(R, f) \times G(f) \times I(f)$$

where $S(M_0, f)$ is the source spectrum with f denoting the frequency and M_0 representing the seismic moment; R is the distance from the earthquake source to the site and $P(R, f)$ accounts for the wave propagation effects, including the geometrical attenuation with distance, which is usually described by a three-segmented function, and the energy dissipation (combining intrinsic and scattering attenuation); $G(f)$ accounts for the effects of local site conditions, including near-surface amplification and reduction of high frequencies that result from the path-independent loss of energy; and finally, $I(f)$ is the indicator of the instrument or the type of motion (i.e., acceleration, velocity, or displacement). More details are provided in Boore (2003).

This method was first developed to model far-field ground motions where the earthquake source can be considered as a point (Boore 1983), resulting in “point-source stochastic models.” For simulation of ground motions closer to the earthquake source, the method was improved to consider the rupture progress on a finite fault (Beresnev and Atkinson 1998), resulting in “finite-fault stochastic models.” In finite-fault modeling, the fault is discretized into many subfaults, and each subfault is treated as a point-source. The ground motion from each subfault is modeled using the point-source stochastic model with the amplitude spectrum described in the above equation. The total ground motion at a site is the superposition of the contributions of all subfaults with a proper time lag considering the difference from the triggering time of the subfault and from the travel time between the subfault and the site. Using finite-fault modeling, the synthetic motion is able to capture some near-fault characteristics of ground motion, such as rupture directivity and hanging-wall effect.

Most improvements in source-based stochastic modeling focus on enhancing the source spectrum, $S(M_0, f)$. The originally used source spectrum was according to the ω -square model with single corner frequency (Aki 1967)

$$S(M_0, f) = \frac{CM_0}{1 + \left(\frac{f}{f_c}\right)^2}$$

where C is a constant related to the radiation pattern and f_c is the corner frequency inversely controlled by the dimension of the source. By applying the above equation in finite-fault modeling, however, the sum of acceleration amplitude spectrum at high-frequency range ($f > f_c$) strongly depends on the discretization scheme of the fault. The proper subfault size is required to be within 5–15 km in order to obtain a reliable synthesis (Beresnev and Atkinson 1998). To solve this problem, Motazedian and Atkinson (2005) made a conceptual improvement by introducing a “dynamic corner frequency” in the source spectrum

$$S_{ij}(f) = \frac{CM_{0ij}H_{ij}}{1 + \left(\frac{f}{f_{cij}}\right)^2}$$

where f_{cij} is the dynamic corner frequency for the ij -th subfault and H_{ij} is the scaling factor to compensate the high-frequency spectral loss from using a dynamic corner frequency.

The dynamic corner frequency of the ij -th subfault depends on the cumulative number of subfaults ruptured by the time the subfault is triggered. Therefore, each subfault will have a different corner frequency in the source spectrum. Generally, a subfault that is triggered in the early part of the rupture will have a higher corner frequency than a subfault triggered in the late part. This describes the difference in the frequency contents of the generated motions from different subfaults. The most significant advantage of using a dynamic corner frequency is that the high-frequency spectral amplitude is no longer dependent on the size of the subfault; thus a subfault can be as small as 1 km, which allows the finite-fault modeling to be used for

small- or intermediate-magnitude events. One of the latest improvements to the source-based stochastic simulation is the work of Boore (2009) that allows selection of different functions for calculating the scaling factor H_{ij} .

Source-based stochastic models are simple and practical compared to deterministic (physics-based) models. Availability of source codes for these models has increased their use in recent years. For example, EXSIM_dmb is a simulation software that is developed by Boore and can be downloaded from his website (http://www.daveboore.com/software_online.html), and EXSIM_beta is a simulation software that is developed by Motazedian and can be found on his website (<http://http-server.carleton.ca/~dariush/>); both EXSIM_dmb and EXSIM_beta are based on stochastic finite-fault modeling with dynamic corner frequency. One of the disadvantages of source-based stochastic models is that they assume stationarity of the frequency content with time, which is usually not the case in real earthquake ground motions as will be discussed in more detail in the next section. Also, care should be taken in using these models for probabilistic analyses because they usually underestimate the variability in ground motions by fixing their parameters instead of assigning a probability distribution to them.

In general, source-based models tend to heavily employ seismological principles to describe the source mechanism and wave travel path and depend on physical parameters that vary significantly from region to region. This limits their use in regions where seismological data are lacking. This is more critical for deterministic (physics-based) models than it is for stochastic models that use simpler parameters and models for the source, path, and site effects. In the current practice, most engineers prefer using methods of scaling and spectrum matching of recorded motions instead of incorporating source-based models. This is partly due to the lack of understanding the seismological principles underlying these models and the fact that these models require a thorough knowledge of the source, wave path, and site characteristics, which typically are not available to a design engineer.

In this regard, source-based stochastic models have an advantage compared to deterministic (physics-based) models; but the user is still required to make certain assumptions depending on the region in order to define the models for the source $S(M_0, f)$, path $P(R, f)$, and site $G(f)$. More recently developed site-based stochastic models can be more useful in engineering applications because their relatively simple formulations facilitate time-efficient simulations (similar to source-based stochastic models) and their parameters are expressed directly in terms of earthquake and site characteristics such as magnitude and distance that describe a design scenario and are readily available to a design engineer.

Site-Based Models

Site-based models are stochastic and represent the earthquake ground motion time series by a stochastic process. A large number of site-based stochastic ground motion models have been developed. Formal reviews are presented by several authors including Liu (1970), Shinozuka and Deodatis (1988), and Conte and Peng (1997). The papers by Rezaeian and Der Kiureghian (2008, 2010, and 2012) provide reviews and are representative of the more recent work. As previously mentioned, a good synthetic ground motion model must represent both the temporal and the spectral nonstationary characteristics of earthquake ground motions. Whereas temporal nonstationarity can be easily modeled in site-based stochastic simulations by time-modulating a stationary ► [stochastic process](#), spectral nonstationarity is not as easy to model. Nevertheless, this spectral nonstationarity is of particular importance in nonlinear dynamic analysis because of the moving resonance effect of nonlinear structures (i.e., when the natural period of a structure varies during the ground shaking period due to inelastic behavior and gets closer to the varying frequency content of the ground motion) and should not be ignored. In general, for a stochastic model to be of practical use in earthquake engineering, it should be parsimonious, i.e., it must have as few parameters as possible. Preferably, the model parameters should provide physical insight into the characteristics

of the motion. Furthermore, the model should refrain from complicated analysis, involving extensive processing of recorded motions for parameter identification.

Site-based stochastic ground motion simulation models can be classified into four categories:

Filtered white-noise processes

Processes obtained by passing a white-noise signal through a filter. These models are appealing due to the efficient digital simulation of sample functions for a filtered white-noise process (i.e., generating realizations). Many models of this type use subsequent modulation in time to achieve temporal nonstationarity, but these processes have essentially time-invariant frequency content. This type of simulation can be improved by varying the filter parameters in time to achieve nonstationarity of the frequency content (e.g., Rezaeian and Der Kiureghian 2008).

Filtered Poisson processes

Processes obtained by passing a train of Poisson pulses through a linear filter. Through modulation in time, these processes can possess both temporal and spectral nonstationarities (e.g., Lin 1986). However, matching with recorded ground motions is difficult.

Auto-regressive-moving-average (ARMA) models

By allowing the model parameters to vary with time, these models (e.g., Conte et al. 1992) can have both temporal and spectral nonstationarities. However, it is difficult to relate the model parameters to any physical aspects of the ground motion.

Time-varying spectral representation

Instead of working in the time domain, as is done in the previous three categories, these models work in the frequency domain and use various forms of spectral representation (e.g., Conte and Peng 1997). The focus in these models is in developing a time-varying spectral representation by matching to a target recorded ground motion. These models require extensive processing of the target motion. Virtually all these models assume the ground motion to be a zero-mean Gaussian process.

$$a(t) = q(t, \alpha) \times f(t, \lambda)$$

where $a(t)$ is the acceleration time series with t denoting time; $q(t, \alpha)$ is a deterministic nonnegative time-modulating function with α representing a vector of parameters; and $f(t, \lambda)$ is a filtered white-noise process with λ representing a vector of parameters. Several models have been proposed for the $q(t, \alpha)$ modulating function, e.g.,

- The piecewise function by Housner and Jennings (1964)
- The double exponential function proposed by Shinozuka and Sato (1967)
- The gamma function proposed by Saragoni and Hart (1974)

These models or their modified versions have been used by many other modelers. In general, a time-modulating function gradually increases from zero to achieve a nearly constant intensity that represents the “strong shaking” phase of an earthquake ground motion and then gradually decays back to zero (see Fig. 1). The parameters α control the intensity and shape of the function. Although it is not a common practice, two or more modulating functions can be combined to simulate ground motions that have more than a single strong shaking phase, a feature possibly caused by multiple sub-events.

The conventional “filtered white-noise” process $f(t, \lambda)$ is the stationary response of a linear time-invariant filter subjected to a white-noise process. White-noise $w(t)$ is a stationary random process in time that has a zero mean and a constant spectral density for all frequencies. The response of a linear filter to a white-noise process may be calculated by using the Duhamel convolution integral, and hence the general formulation of a filtered white-noise process can be written in the following form:

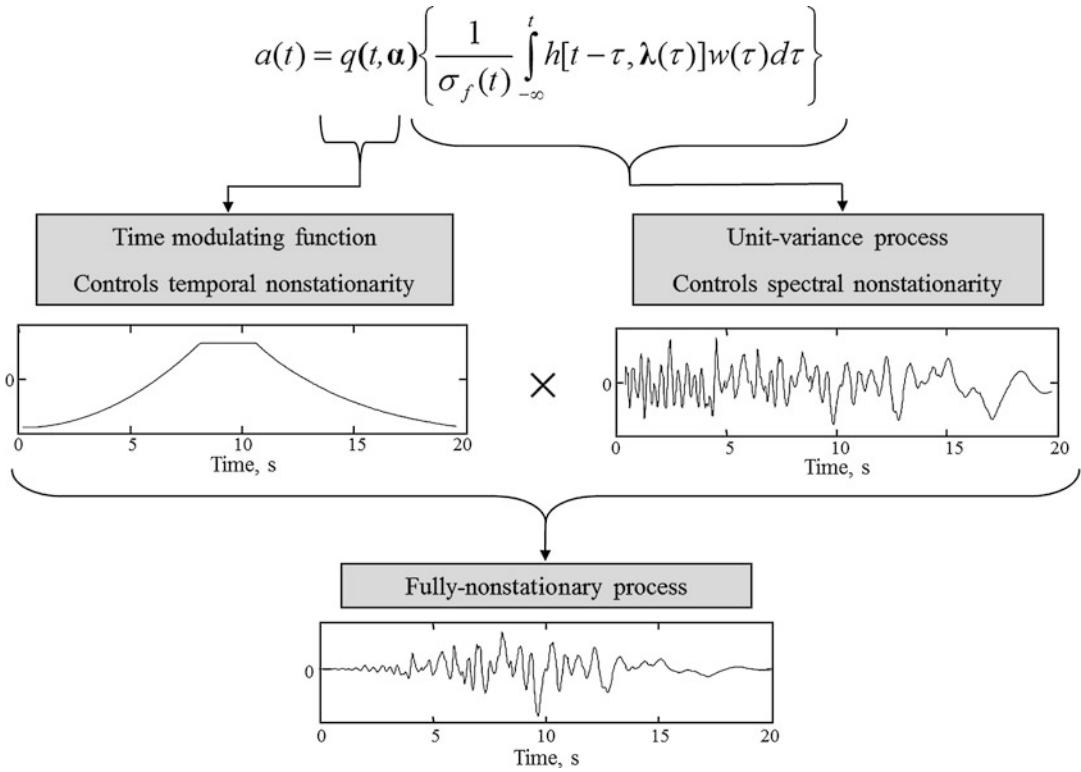
$$f(t, \lambda) = \int_{-\infty}^t h(t - \tau, \lambda) w(\tau) d\tau$$

where $h(t, \lambda)$ denotes the impulse response function or IRF of a linear filter and the parameters λ can be the natural frequency and damping of the

Simulating a Single Ground Motion Time Series Using a Site-Based Stochastic Model

Model Formulation

Regardless of the specific formulation, every site-based model is essentially a stochastic process with a certain number of parameters. As an example, the general formulation for a site-based stochastic model based on the filtered white-noise process is given below (Rezaeian and Der Kiureghian 2008)



Stochastic Ground Motion Simulation, Fig. 1 Development of a fully nonstationary stochastic process according to Rezaeian and Der Kiureghian (2008)

filter or other parameters that shape the filter response. Different filters are used and various tricks are incorporated by modelers to make the resulting process “fully nonstationary” and representative of real earthquake ground motions. One approach taken by Rezaeian and Der Kiureghian (2008) is illustrated in Fig. 1, where the filtered white-noise process is normalized by its standard deviation $\sigma_f(t)$ to separate the temporal from the spectral characteristics; and the filter parameters are functions of time, $\lambda(t)$, to achieve spectral nonstationarity. Like all other models that use a time-modulating function, temporal nonstationarity is achieved by $q(t, \alpha)$. In this figure, the nonstationary characteristics of ground motions are demonstrated using the trace of an example stochastic process.

Model Parameters

Given a target accelerogram (i.e., a recorded ground motion), the model parameters α and

$\lambda(t)$ are identified by fitting certain characteristics of the stochastic process to those of the target accelerogram. The approaches to parameterization and model fitting are entirely different from model to model. Currently, there are no standards on the exact characteristics that modelers should match to identify their parameters for a target motion. However, all these different characteristics are designed to capture the intensity, duration, and frequency content of the ground motion at certain points in time or in frequency. In some models, this is done over the entire duration of the recorded ground motion and its entire frequency spectrum; in this case, the parameters are identified using the following generalized formulation:

$$\hat{\mathbf{p}} = \underset{\mathbf{p}}{\operatorname{argmin}} \int_0^{t_n} \left[f_{obj}(\text{recorded}) - f_{obj}(\text{simulation}) \right] dt$$

where $\mathbf{p}$ is a vector of model parameters and t_n denotes the entire duration of the recorded ground

motion. Linear or nonlinear optimization methods are used to minimize the difference between an objective function for the recorded motion, $f_{obj}(\text{recorded})$, and the same function for the simulation, $f_{obj}(\text{simulation})$, resulting in the vector of identified parameters $\hat{\mathbf{p}}$ for that particular recorded motion. The misfit of the objective functions can be characterized by their difference, as is shown in the above equation, or by their absolute difference, or by their square root difference. Many objective functions are used in the literature by various modelers. Objective functions such as the expected cumulative energy of the process or the variance of the process can be used to identify the parameters of the time-modulating function; objective functions such as the zero-level up-crossing rate can be used to identify the frequency parameters (e.g., Yeh and Wen 1990; Rezaeian and Der Kiureghian 2008).

Once a set of model parameters has been identified, $\hat{\mathbf{p}}$, the model formulation is used to simulate realizations of the ground motion. These realizations are all different due to the stochasticity of the model (e.g., the white-noise process in the above example), but they all have the same model parameters and expected characteristics similar to those of the target motion. In this way, the target accelerogram may be regarded as a single realization of the ground motion process for a specified set of model parameters, while the simulated motions may be regarded as other random samples of the process for the same set of model parameters.

Post-Processing for Long Spectral Periods (Low Frequency)

In general, site-based stochastic ground motion models fail to match the response spectrum associated with the target accelerogram in the long spectral period range and tend to overestimate the structural response (typically beyond 2–4 s depending on the model). This is because a stochastic model developed for an acceleration process cannot guarantee zero velocity and displacement residuals (final values at the end of the record) upon integration of a sample realization. This shortcoming of site-based models has been recognized by some modelers who “post-process” the simulations by extending

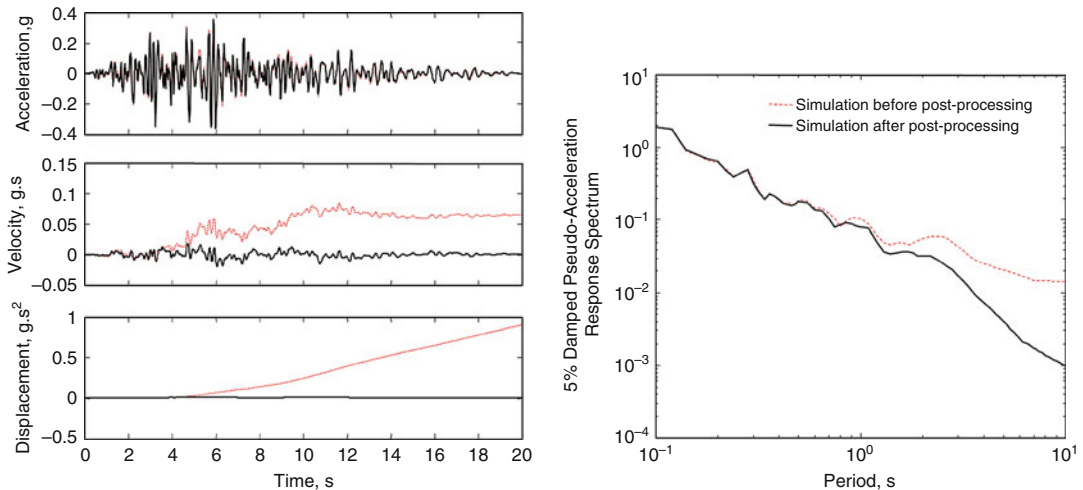
baseline correction methods or high-pass filtering methods used for recorded accelerograms to simulated motions in order to overcome this disadvantage (e.g., Papadimitriou 1990; Liao and Zerva 2006). For example, Rezaeian and Der Kiureghian (2008) use a high-pass filter representing a critically damped oscillator to post-process the stochastically simulated ground motion to obtain zero velocity and displacement residuals and appropriate response spectral values for periods as long as 5–10 s. An example of a simulated ground motion before and after post-processing is shown in Fig. 2. Even though the difference between the acceleration processes is insignificant, the integration over time results in unacceptably high nonzero velocity and displacement residuals as well as high spectral intensities at long periods before post-processing. After the post-processing, the residuals are zero and the spectral content is lower at long periods.

Simulating a Suite of Ground Motions for a Specified Design Scenario

Here, a design scenario refers to a set of earthquake and site characteristics such as earthquake magnitude, source-to-site distance, and soil conditions of the site. In the broader context of PBEE, an ensemble of ground motions that represents all possible ground shakings for a specified design scenario is of interest, not just ground motion simulations with characteristics similar to those of a previously observed motion. A good stochastic ground motion model must properly represent the natural variability of real earthquake ground motions. It also must be practical for use by a practicing engineer.

Variability

The variability among an ensemble of simulated ground motions for a specified design scenario comes from two sources: (1) the randomness of ground motions for a specified set of model parameters (the stochasticity mentioned in the discussion of the model parameters) and (2) the randomness of the model parameters for a specified design scenario. The former is due to the stochastic nature



Stochastic Ground Motion Simulation, Fig. 2 An example stochastic simulation before and after post-processing using a high-pass filter with a cutoff frequency of 0.2 Hz

of the model and it is achieved when fitting to and simulating a target accelerogram. The latter is referred to as “parametric uncertainty.” Both should be modeled properly to reproduce in the synthetic motions the variability present in real earthquake ground motions.

The underestimation of parametric uncertainty results in the underestimation of the natural variability of ground motions. This has been a major problem in the current practice of seismic hazard analysis and the generation of synthetic ground motions. A few exceptions are some recent site-based models (e.g., Rezaeian and Der Kiureghian 2010), where model parameters are fitted to a database of ground motions and are randomized at the time of simulation to achieve variability similar to the sample database. Most source-based models account for the variability in ground motions by manually varying the source parameters; whether the realizations follow the “true” underlying probability distribution for each parameter is the question in these models.

To properly account for parametric uncertainty, a good model needs to randomly generate parameters according to their underlying probability distributions and also to account for the correlations between the parameters. Probability distributions for parameters and their correlations can be estimated empirically or based on the

physical meaning of a parameter. If correlations between the parameters are ignored, the variability in the resulting ensemble of ground motions may be overestimated because the parameter realizations are not well constrained.

Practicality

There is a need for generating synthetic motions that correspond to a specified design scenario, but most site-based stochastic models limit their scope to generating synthetic motions similar to a target recorded ground motion. In such models, no attempt is made to select a different but appropriate set of model parameters for the specified design scenario. Moreover, for a model to be practical and to be used in the engineering community, it should be expressed in terms of variables that are readily available to a design engineer.

In practice, GMPEs are widely used because they predict measures of ground motion intensity in terms of common earthquake and site characteristics that define a design scenario to a practicing engineer. Typical intensity measures modeled by GMPEs are peak ground motion values (i.e., peak ground acceleration, velocity, and displacement) and elastic response spectra at specified oscillator periods and damping ratios. A few GMPEs for inelastic response spectra have also been developed. These GMPEs are useful for

linear response-spectrum analysis or crude nonlinear analysis, but not for response-history analysis, as they do not predict ground motion time series. Such simplified analysis methods have proven to be adequate for simple code-based design purposes, but not for PBEE. Stochastic ground motion models can complement GMPEs by generating ground motion time series in terms of similar earthquake and site characteristics used by GMPEs. This is done by developing predictive relations for the model parameters in terms of commonly used earthquake and site characteristics.

To adjust a stochastic model for practical use, the model parameters are identified for a database of recorded ground motions. Then, through regression analyses and empirical modeling methods, predictive relations for each parameter in terms of earthquake and site characteristics are developed to relieve the design engineer from the task of predicting the parameter values for a design scenario (e.g., Sabetta and Pugliese 1996; Stafford et al. 2009; Rezaeian and Der Kiureghian 2010). A generic form of predictive relations for a model parameter p is given below

$$\Phi^{-1}[F_P(p)] = \mu(\text{Earthquake, Site, } \boldsymbol{\beta}) + e$$

where Φ^{-1} denotes the inverse of the standard normal cumulative distribution function, F_P denotes the marginal cumulative distribution function estimated for parameter p , μ is the conditional mean of $\Phi^{-1}[F_P(p)]$ given the earthquake and site characteristics of interest, and $\boldsymbol{\beta}$ is the vector of regression coefficients. The regression error e is a zero-mean normally distributed random variable with standard deviation σ representing the uncertainty of the parameter. The earthquake and site characteristics are typically the earthquake magnitude, source-to-site distance, faulting mechanism, and various proxies to describe the site soil conditions such as V_{S30} (i.e., the time-averaged shear wave velocity in the top 30 m of soil). These regression variables change depending on the region and the parameter under consideration.

Given the predictive relations as described above, one can calculate the mean of each model parameter given a set of earthquake and site

characteristics. The error can then be randomized to generate the realizations of the model parameter. To generate the realizations of multiple parameters, correlations between the parameters should be taken into account in the simulation.

Extensions of Site-Based Stochastic Models

Multiple-Component Simulation

Earthquake ground motions are multidimensional. Neglecting the rotational components, three translational components are usually expressed in two orthogonal horizontal directions and one vertical direction. For proper dynamic analysis of complex structural and geotechnical systems, for example, asymmetric structures that are vulnerable to torsion or structural components that are sensitive to vertical ground shaking, it is important to simulate multicomponent ground motion time series. The vast majority of stochastic models are restricted to simulating a single horizontal ground motion component, but the research to simulate multiple components is continuing and under development (e.g., bidirectional ground motion simulation by Rezaeian and Der Kiureghian (2012)).

When modeling and simulating multiple ground motion components, differences as well as similarities and dependencies between the components must be taken into account. Because the ground motion components are generated from the same earthquake source and from seismic waves that travel through the same ground medium, high correlations between model parameters of the components are expected. Correlations and coherencies between the components can be estimated empirically by analyzing a large number of recorded ground motion components.

For stochastic ground motions, in addition to cross-correlations between the model parameters, dependencies between the stochastic components (e.g., the seed white-noise processes) should also be taken into account. This is a difficult task, but Penzien and Watabe (1975) showed that there is a unique set of orthogonal axes, referred to as “principal axes,” along which the three translational

components of ground motion may be considered uncorrelated. These are the major, intermediate, and minor principal axes along which the ground motion components have intensities in decreasing order and the stochastic processes (e.g., the seed white-noise processes) are statistically independent. Although the stochastic processes can be assumed statistically independent, there can still be high correlations between the model parameters of the two ground motion components (see Rezaeian and Der Kiureghian 2012). Based on examination of few real accelerograms, Penzien and Watabe (1975) suggested that the major principal axis usually points in the general direction of the epicenter, the intermediate principal axis is horizontal and perpendicular to the major principal axis, and the minor principal axis is almost vertical. The concept of principal axes has been the basis for many subsequent studies on multicomponent stochastic modeling of ground motions, but more studies are required to confirm the direction of the principal components. If simulation is done along the horizontal principal axes, an orthogonal pair of acceleration time series $a_1(t)$ and $a_2(t)$ can easily be rotated to the horizontal direction of interest using a rotation matrix

$$\begin{bmatrix} a_{1,\theta}(t) \\ a_{2,\theta}(t) \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} a_1(t) \\ a_2(t) \end{bmatrix}$$

where $a_{1,\theta}(t)$ and $a_{2,\theta}(t)$ are the counterclockwise rotations by angle θ .

Near-Fault Simulation

Ground motions close to fault ruptures (within about 30 km), i.e., ► **near-fault** ground motions, can have distinct characteristics such as the directivity effect in the fault-normal direction and the fling step in the fault-parallel direction. These characteristics can have a strong influence on the structural response and are usually modeled separately and then added to a site-based stochastic ground motion simulation.

A site near a fault may experience forward directivity, which occurs when the fault rupture propagates toward the site with a velocity almost equal to the shear wave velocity. The resulting ground motion time series is characterized by the

presence of a two-sided, long-period, large-amplitude velocity pulse that typically shows as a long-period pulse in the response spectrum and may impose extreme demands on structures. Because near-fault recordings that contain directivity pulses are relatively scarce and because scaling and modification methods cannot create a pulse in a record that does not initially contain a pulse, simulated near-fault ground motions are needed. But only a few stochastic models currently exist that simulate the forward directivity effect (e.g., Mavroeidis and Papageorgiou 2003; Dabaghi et al. 2011). This is usually done by modeling a long-period velocity pulse that is a function of the earthquake design scenario, including the relative location of the site to the fault rupture. The simulated velocity pulse is added to the derivative of the simulated acceleration time series resulting in a simulated pulse-like near-fault ground motion.

Not all near-fault ground motions contain a pulse. For example, backward directivity occurs when the fault rupture propagates away from the site. The resulting ground motion tends to be of low intensity and long duration. A valid near-fault ground motion model should take into account both the pulse-like and non-pulse-like cases.

The fling step is caused by the permanent displacement of the fault and is usually characterized by a one-sided velocity pulse in the fault-parallel direction. Modeling of the fling step is not as well developed in stochastic ground motion modeling, but its effect on structural response is not as extreme as the forward directivity effect. The fling step is difficult to model because information is usually lost in the typical processing of recorded ground motions.

Spatial Simulation (Multiple Stations)

Most stochastic ground motion models are developed for single-station simulations. To simulate ground motions from a single earthquake event at multiple stations that are located close to each other, the spatial coherency between records should be taken into account. This coherency usually depends on the frequency content of the ground motions and separation distance between the stations and can be estimated from the power spectral density functions of recorded motions.

Several models exist in the literature that can be used to adjust a ground motion time series for spatial variation (e.g., Liao and Zerva 2006; Konakli and Der Kiureghian 2012, Konakli et al. 2014). These models can be applied to a stochastic simulation of ground motion at a single station to generate consistent simulations at nearby stations.

Validation of Ground Motion Simulation Models

Although many stochastic ground motion simulation models exist, no standard validation techniques are available. There are almost as many validation approaches undertaken as there are simulation models. Comparisons against empirical data and trusted models in engineering practice aid in identifying the limitations of the proposed simulation models and may encourage implementation of simulations in practice. The next step in the field of ground motion simulation that has been gaining attention in recent years is development of standard techniques for ground motion simulation validation for use in various engineering applications (see, e.g., the efforts of the Southern California Earthquake Center ground motion simulation validation technical activity group http://collaborate.scec.org/gmsv/Main_Page).

Summary

Stochastic ground motion simulations are used to replace or supplement recorded earthquake ground motions when recorded motions are sparse. Stochastic ground motion models are categorized into source-based and site-based models. Definitions and examples for each category are given. Source-based simulations explicitly model the source spectrum, path, and site effects. Site-based stochastic models are classified into four groups depending on their formulations. Regardless of the formulation, the model must exhibit the characteristics of real earthquake ground motions including temporal and spectral nonstationarities. Site-based models simulate a

recorded accelerogram by identifying the model parameters through optimization and fitting into certain target characteristics of real earthquake ground motions. The parameters are then used in the formulation of the stochastic process. Post-processing is usually required for site-based stochastic models to ensure zero residuals and proper modeling of long-period spectral values. Simulating a suite of ground motions for a specified design scenario is discussed, where the natural variability of ground motions must be modeled properly by accounting for parametric uncertainty and correlations between the model parameters. Practicality of a model is also discussed, where the model parameters are expressed in terms of earthquake and site characteristics that define a design scenario and are readily available to a design engineer. Finally, adjusting site-based stochastic ground motion simulations for special cases including multicomponent, near-fault, and multi-station simulations is discussed. Currently, many research efforts are directed toward developing ground motion simulation validation procedures for various engineering applications.

Acknowledgment We would like to thank Mayssa Dabaghi, Morgan Moschetti, Ricardo Taborda, and the anonymous reviewer of this entry for their review and valuable feedback. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the US government.

Cross-References

► [Physics-Based Ground-Motion Simulation](#)

References

- Aki K (1967) Scaling law of seismic spectrum. *J Geophys Res* 72:1217–1231
- Au SK, Beck JL (2003) Subset simulation and its applications to seismic risk based on dynamic analysis. *J Eng Mech ASCE* 129:901–917
- Beresnev IA, Atkinson G (1998) Stochastic finite-fault modeling of ground motions from the 1994 northridge, California, earthquake (I): validation on rock sites. *Bull Seismol Soc Am* 88(6):1392–1401
- Boore DM (2003) Simulation of ground motion using the stochastic method. *Pure Appl Geophys* 160:635–676

- Boore DM (1983) Stochastic simulation of high-frequency ground motions based on seismological models of the radiated spectra. *Bull Seismol Soc Am* 73:1865–1894
- Boore DM (2009) Comparing stochastic point-source and finite-source ground-motion simulations: SMSIM and EXSIM. *Bull Seismol Soc Am* 99:3202–3216
- Conte JP, Peng BF (1997) Fully nonstationary analytical earthquake ground-motion model. *J Eng Mech ASCE* 12:15–24
- Conte JP, Pister KS, Mahin SA (1992) Nonstationary ARMA modeling of seismic motions. *Soil Dyn Earthq Eng* 11:411–426
- Dabaghi M, Rezaeian S, Der Kiureghian A (2011) Stochastic simulation of near-fault ground motions for specified earthquake and site characteristics. In: *Proceedings of the 11th international conference on applications of statistics and probability in civil engineering (ICASP11)*, Zurich, pp 759–760
- Der Kiureghian A, Fujimura K (2009) Nonlinear stochastic dynamic analysis for performance-based earthquake engineering. *Earthq Eng Struct* 38:719–738
- Housner GW, Jennings PC (1964) Generation of artificial earthquakes. *J Eng Mech Div ASCE* 90:113–150
- Konakli K, Der Kiureghian A (2012) Simulation of spatially varying ground motions including incoherence, wave-passage and differential site-response effects. *Earthq Eng Struct* 41:495–513
- Konakli K, Der Kiureghian A, Dreger D (2014) Coherency analysis of accelerograms recorded by the UPSAR array during the 2004 Parkfield earthquake. *Earthq Eng Struct* 43:641–659
- Liao S, Zerva A (2006) Physically compliant, conditionally simulated spatially variable seismic ground motions for performance-based design. *Earthq Eng Struct* 35:891–919
- Lin YK (1986) On random pulse train and its evolutionary spectral representation. *Probabilist Eng Mech* 1:219–223
- Liu SC (1970) Synthesis of stochastic representations of ground motions. *Bell Sys Tech J* 49:521–541
- Mavroeidis GP, Papageorgiou AS (2003) A mathematical representation of near-fault ground motions. *Bull Seismol Soc Am* 93(3):1099–1131
- McGuire RK, Hanks TC (1980) RMS accelerations and spectral amplitude of strong ground motion during the San Fernando, California, earthquake. *Bull Seismol Soc Am* 70(5):1907–1919
- Motazedian D, Atkinson GM (2005) Stochastic finite-fault modeling based on a dynamic corner frequency. *Bull Seismol Soc Am* 95(3):995–1010
- Papadimitriou K (1990) Stochastic characterization of strong ground motion and application to structural response. *Earthquake Engineering Research Laboratory, California Institute of Technology, Pasadena, Report No. EERL 90-03*
- Penzien J, Watabe M (1975) Characteristics of 3-dimensional earthquake ground motions. *Earthq Eng Struct* 3:365–373
- Rezaeian S, Der Kiureghian A (2008) A stochastic ground motion model with separable temporal and spectral nonstationarities. *Earthq Eng Struct* 37:1565–1584
- Rezaeian S, Der Kiureghian A (2010) Simulation of synthetic ground motions for specified earthquake and site characteristics. *Earthq Eng Struct* 39:1155–1180
- Rezaeian S, Der Kiureghian A (2012) Simulation of orthogonal horizontal ground motion components for specified earthquake and site characteristics. *Earthq Eng Struct* 41:335–353
- Sabetta F, Pugliese A (1996) Estimation of response spectra and simulation of nonstationary earthquake ground motions. *Bull Seismol Soc Am* 86:337–352
- Saragoni GR, Hart GC (1974) Simulation of artificial earthquakes. *Earthq Eng Struct* 2:249–267
- Shinozuka M, Deodatis G (1988) Stochastic process models for earthquake ground motion. *Probabilist Eng Mech* 3:114–123
- Shinozuka M, Sato Y (1967) Simulation of nonstationary random process. *J Eng Mech ASCE* 93:11–40
- Somerville P, Collins N, Abrahamson N, Graves R, Saikia C (2001) Ground motion attenuation relations for the Central and Eastern United States – Final report, 30 June 2001: Report to U.S. Geological Survey for award 99HQGR0098, 38 p
- Stafford PJ, Sgobba S, Marano GC (2009) An energy-based envelope function for the stochastic simulation of earthquake accelerograms. *Soil Dyn Earthq Eng* 29:1123–1133
- Yeh CH, Wen YK (1990) Modeling of nonstationary ground motion and analysis of inelastic structural response. *Struct Safety* 8:281–298
- Zerva A (1988) Seismic source mechanisms and ground motion models, review paper. *Probabilist Eng Mech* 3:64–74

Stochastic Structural Identification from Vibrational and Environmental Data

Minas D. Spiridonakos and Eleni N. Chatzi
Department of Civil, Environmental and
Geomatic Engineering, ETH Zurich, Institute of
Structural Engineering, Zurich, CH, Switzerland

Synonyms

ARMA model; Environmental conditions; Output-only identification; Polynomial chaos expansion; Structural identification; Vibration response

Introduction

In recent years, Structural Health Monitoring (SHM) of engineering structures has become an important area of research. This is a broad term which by large also deals with the online detection and identification of structural damage and therefore holds a critical significance for the case of large-scale civil structures where a failure incidence is connected not only to high cost losses but potentially to human loss in the worst-case scenario.

Vibration-based methods are today the fastest growing research area in the SHM field (Fassois and Kopsaftopoulos 2013). However, while such methods are already used for industrial mechanical structures, these are still far from being successfully implemented in large-scale civil structures such as high-rise buildings and bridges. The main reason for this discrepancy is the fact that these structures do not operate in the controlled environment of an industry or a laboratory facility but rather they are affected by a number of uncontrollable environmental conditions, such as temperature, temperature spatial gradients due to solar irradiation, temporal temperature gradient due to thermal capacity, humidity, and others (Sohn 2007). The shortcomings of vibration-based SHM methods have been recognized already in one of the first review papers on the subject (Doebling et al. 1998) exposing the advantages and capabilities of several methodologies in this area.

For the aforementioned reason, one is faced with the problem of identifying a dynamic model that is able to represent a given structure for an extended spectrum of operational conditions – if not its entire life cycle. Toward this end, the combined vibrational response and environmental data, which are normally available in current SHM systems, have to be utilized along with appropriate models that can account for the variability of structural dynamics due to environmental conditions. The approach detailed herein treats this issue through a functional representation that introduces this dependence into the model's parameters.

Structures Operating Under Varying Conditions and Their Responses

A linear viscously damped structure may be described by the following lumped parameter, time-varying ordinary differential equation:

$$\mathbf{M}(t)\ddot{\mathbf{x}}(t) + \mathbf{C}(t)\dot{\mathbf{x}}(t) + \mathbf{K}(t)\mathbf{x}(t) = \mathbf{u}(t), \quad \text{for } t \in [t_0, t_f] \quad (1)$$

where t designates analog time with t_0 denoting the start and t_f the end of the structure's lifetime or inspection interval, $\mathbf{x}(t)$ the structural displacements response vector, $\mathbf{u}(t)$ the applied excitation forces vector, and $\mathbf{M}(t)$, $\mathbf{C}(t)$, $\mathbf{K}(t)$ the mass, damping, and stiffness matrices of the structure, respectively.

The variability of structural response may be usually attributed to a small number of parameters related with either inherent properties of the structure or exogenous random variables, e.g., environmental and loading conditions, geometry, mass distribution, and so on. Gathering these L parameters in a single vector $\boldsymbol{\xi}(t) = [\xi_1(t), \xi_2(t), \dots, \xi_L(t)]^T$ the previous model may be rewritten as

$$\mathbf{M}(\boldsymbol{\xi}(t))\ddot{\mathbf{x}}(t) + \mathbf{C}(\boldsymbol{\xi}(t))\dot{\mathbf{x}}(t) + \mathbf{K}(\boldsymbol{\xi}(t))\mathbf{x}(t) = \mathbf{u}(t) \quad (2)$$

Focusing on environmental conditions and assuming that they are characterized by slow (compared to the structural dynamics) variation, the previous model may be rewritten as

$$\mathbf{M}(\boldsymbol{\xi})\ddot{\mathbf{x}}(t) + \mathbf{C}(\boldsymbol{\xi})\dot{\mathbf{x}}(t) + \mathbf{K}(\boldsymbol{\xi})\mathbf{x}(t) = \mathbf{u}(t), \quad \text{for } t \in [t_a, t_b] \subset [t_0, t_f] \quad (3)$$

The latter may be considered as a local time-invariant model which represents the structure when monitored during a small time interval. Within this interval the input parameter vector,

ξ and as a consequence the mechanical properties of the structure, may be considered to be “fixed” to constant values. Then, the input parameter vector ξ may be considered as a realization of the random vector Ξ with joint probability density function (pdf) $f_{\Xi}(\xi)$. In this way the initial linear time-varying (LTV) model of Eq. 1 may be written as a linear time-invariant (LTI) model:

$$\mathbf{M}(\Xi)\ddot{\mathbf{x}}(t) + \mathbf{C}(\Xi)\dot{\mathbf{x}}(t) + \mathbf{K}(\Xi)\mathbf{x}(t) = \mathbf{u}(t), \quad \text{with } \Xi(\omega) \in \Omega \quad (4)$$

which is however characterized by parameters that are depending on the input variable vector Ξ . In the equation above, Ω designates the event space of the random vector Ξ , and ω an elementary event of this space.

The above continuous-time model obviously has a discrete-time counterpart, which for the simpler single-input single-output case may be approximated by a scalar difference equation of the following form (Andersen 1997):

$$\begin{aligned} x(t) + a_1(\Xi)x[t-1] + \dots + a_{n_a}(\Xi)x[t-n_a] \\ = b_0(\Xi)u[t] + b_1(\Xi)u[t-1] + \dots \\ + b_{n_b}(\Xi)u[t-n_b], \quad \text{for } t = 1, \dots, N \end{aligned} \quad (5)$$

in which t designates normalized discrete time (absolute time normalized by the sampling period T_s), N the signals' length in samples, $x[t]$, $u[t]$ the discretized versions of the observed vibration response and excitation force, respectively, $a_i(\Xi)$, $b_i(\Xi)$ the model parameters depending on Ξ , and finally n_a , n_b the equation orders.

Structural identification is an inverse problem which concerns the estimation of an appropriate mathematical representation of the structural dynamics based on vibrational measurements. In the case of ambient loading, the excitation may not be measured with accuracy, and thus, the input sequence $u[t]$ in the theoretical model of Eq. 5 has to be replaced by an unknown excitation

sequence $e[t]$. The latter is typically considered to be a normally independently distributed (NID) sequence, that is, a white noise process, with zero mean and constant variance σ_e^2 . Therefore, the analytical model of Eq. 5 is approximated by a model of the following form:

$$\begin{aligned} x(t) + a_1(\Xi)x[t-1] + \dots + a_{n_a}(\Xi)x[t-n_a] \\ \underbrace{= e[t] + c_1(\Xi)e[t-1] + \dots + c_{n_c}(\Xi)e[t-n_c]}_{\text{MA part}}, \\ e[t] \sim \text{NID}(0, \sigma_e^2), \quad \text{for } t = 1, \dots, N \end{aligned} \quad (6)$$

which has the form of a conventional ARMA model (Ljung 1999), with the important difference that it is characterized by stochastic parameters which depend on the input variable vector Ξ . These parameters may actually be represented by a deterministic mapping describing their relation with the input random variables. Specifically, assuming that the ARMA model parameters $a_i(\Xi)$ and $c_i(\Xi)$ have finite variance, they admit the following representation (Soize and Ghanem 2004):

$$\begin{aligned} a_i(\Xi) &= \sum_{j=1}^{\infty} a_{i,j} \varphi_{\mathbf{d}(j)}(\Xi) \quad (i = 1, \dots, n_a), \\ c_i(\Xi) &= \sum_{j=1}^{\infty} c_{i,j} \varphi_{\mathbf{d}(j)}(\Xi) \quad (i = 1, \dots, n_c) \end{aligned} \quad (7)$$

with $a_{i,j}$ and $c_{i,j}$ designating unknown deterministic coefficients of projection onto the probability space of Ξ and $\varphi_{\mathbf{d}(j)}$ are multivariate basis functions that are orthonormal with respect to the joint pdf of Ξ , that is,

$$E[\varphi_{\alpha}(\Xi)\varphi_{\beta}(\Xi)] = \begin{cases} 1, & \text{for } \alpha=\beta \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

and $\mathbf{d}(j)$ is the multi-indices of the multivariate basis. Each probability density function may be

associated with a well-known family of orthogonal polynomials. For instance, a normal distribution is associated with Hermite polynomials while a uniform distribution with Legendre. A list of the most common probability density functions along with the corresponding orthogonal polynomials and the relations for their construction may be found in Xiu and Karniadakis (2002). For the case of independent random variables Ξ_j the expansion of Eq. 7 is called polynomial chaos (PC) expansion (Blatman and Sudret 2010), and the resulting models may be referred to as PC-ARMA models.

Of course, for purposes of practicality, the infinite series of PC expansion of Eq. 7 must be truncated by selecting an appropriate functional subspace consisting of a finite number of terms. A usual approach lies in the selection of a functional subspace consisting of polynomial basis functions with total maximum degree P , that is, $|d(j)| = \sum_{\ell=1}^L d(j, \ell) \leq P$ for all j . In this case the dimensionality of the functional subspace is equal to

$$p = \frac{(L + P)!}{L!P!} \quad (9)$$

where L is the number of random variables and P the maximum basis total degree. For instance, for a random input vector of dimension equal to three ($L = 3$) and maximum degree equal to two ($P = 2$), the multi-indices vectors are:

	l		
	1	2	3
$d(1)$	0	0	0
$d(2)$	1	0	0
$d(3)$	0	1	0
$d(4)$	0	0	1
$d(5)$	2	0	0
$d(6)$	0	2	0
$d(7)$	0	0	2
$d(8)$	1	1	0
$d(9)$	1	0	1
$d(10)$	0	1	1

In this way, the resulting PC-ARMA model is fully parametrized in terms of a finite number

of deterministic coefficients of projection $a_{i,j}$ and $c_{i,j}$ while a specific PC-ARMA model structure is fully defined by the model orders n_a, n_c , and the basis total degree P .

At this point it should be noted that similar models have been used for the structural identification and damage detection issues in a number of recent studies (Functionally Pooled ARX models; see Kopsaftopoulos and Fassois 2013 and the references therein) and automatic control within the Linear Parameter Varying (LPV) models framework (Tóth 2010, Zhao et al. 2012).

PC-ARMA Model Identification

The complete PC-ARMA structural identification problem may be stated as follows: Given a set of dynamic vibration response signals obtained from the structure under study for various measured environmental conditions, select the PC-ARMA model structure and estimate the corresponding model parameters that best fit the available records. The issues of PC-ARMA model parameter estimation and model structure selection are discussed in the following sections.

Model Parameter Estimation

As already mentioned, the estimation of a PC-ARMA model refers to the determination of the coefficients of projection assembled in the parameter vector

$$\theta = [a_{1,1} a_{1,2} \cdots a_{n_a,p} c_{1,1} c_{1,2} \cdots c_{n_c,p}]^T$$

or a subset of θ selected by the procedure described in the next section. Its estimation is based on the availability of structural vibration response data and measurements of the environmental conditions of interest. In order to identify a stochastic model that will be able to describe the structure's dynamics for the wide range of its operational spectrum, data from a sufficient number of operating conditions should be available.

Considering K such datasets have been acquired during an equal number of experiments conducted under different input random vector realizations

$\xi_k = [\xi_{1k} \xi_{2k} \dots \xi_{Lk}]^T$, for $k = 1, 2, \dots, K$. Based on the validity of the assumptions described in the previous section, it is considered that the recorded

vibration response signals $\mathbf{x}_k^N = \{x_k[1], x_k[2], \dots, x_k[N]\}^T$ follow the PC-ARMA model of Eqs. 6 and 7:

$$\begin{aligned} x_k[t] &+ \sum_{j=1}^p a_{1,j} \varphi_{\mathbf{d}(j)}(\xi_k) x_k[t-1] + \dots + \sum_{j=1}^p a_{n_a,j} \varphi_{\mathbf{d}(j)}(\xi_k) x_k[t-n_a] \\ &= e_k[t] + \sum_{j=1}^p c_{1,j} \varphi_{\mathbf{d}(j)}(\xi_k) e_k[t-1] + \dots + \sum_{j=1}^p c_{n_c,j} \varphi_{\mathbf{d}(j)}(\xi_k) e_k[t-n_c], \end{aligned} \quad (9a)$$

for $t = 1, \dots, N$, and $k = 1, \dots, K$

$$e_k \sim \text{NID} \left(0, \sigma_{e_k}^2 \right) \quad (9b)$$

$$E\{e_i[t]e_j[t-\tau]\} = \begin{cases} \sigma_{e_i}^2 \delta_{i,\tau} & \text{for } i=j \\ 0 & \text{for } i \neq j \end{cases} \quad (9c)$$

with $\delta_{i,j}$ designating the Kronecker delta function. In the relationships above, the uncorrelatedness of the residual series between different vibration response records has been added to the model assumptions.

The estimation of the PC-ARMA model parameter vector θ may be based on the minimization of the Prediction Error (PE) criterion. The PE criterion consists of the sum of squares of the model's one-step-ahead prediction errors for the complete set of datasets. As it may be easily shown, the one-step-ahead prediction errors

coincide with the model's residual sequence $e_k[t]$ (Ljung 1999), and thus, θ has to be estimated through the following minimization problem:

$$\begin{aligned} \theta &= \arg \min_{\theta} \left\{ \sum_{k=1}^K \sum_{t=1}^N \frac{(x_k[t] - \hat{x}_k[t|t-1])^2}{\gamma_k^2} \right\} \\ &= \arg \min_{\theta} \left\{ \sum_{k=1}^K \sum_{t=1}^N \frac{\hat{e}_k[t]^2}{\gamma_k^2} \right\} \end{aligned} \quad (10)$$

with $\hat{x}_k[t|t-1]$ denoting the model's one-step-ahead prediction, γ_k user defined weights, and argmin minimizing argument with respect to the indicated quantity. Toward this end, Eq. 9 may be rewritten in a regression type form as

$$\begin{aligned} x_k[t] &= - \sum_{i=1}^{n_a} \sum_{j=1}^p a_{i,j} \varphi_{\mathbf{d}(j)}(\xi_k) x_k[t-i] + \sum_{i=1}^{n_c} \sum_{j=1}^p c_{i,j} \varphi_{\mathbf{d}(j)}(\xi_k) e_k[t-i] + e_k[t] \\ &= \underbrace{\begin{bmatrix} -\varphi_{\mathbf{d}(1)}(\xi_k) x_k[t-1] \\ -\varphi_{\mathbf{d}(2)}(\xi_k) x_k[t-1] \\ \vdots \\ -\varphi_{\mathbf{d}(p-1)}(\xi_k) x_k[t-n_a] \\ -\varphi_{\mathbf{d}(p)}(\xi_k) x_k[t-n_a] \\ \varphi_{\mathbf{d}(1)}(\xi_k) e_k[t-1] \\ \varphi_{\mathbf{d}(2)}(\xi_k) e_k[t-1] \\ \vdots \\ \varphi_{\mathbf{d}(p-1)}(\xi_k) e_k[t-n_c] \\ \varphi_{\mathbf{d}(p)}(\xi_k) e_k[t-n_c] \end{bmatrix}}_{\varphi[\xi_k, t]^T} \cdot \underbrace{\begin{bmatrix} a_{1,1} \\ a_{1,2} \\ \vdots \\ a_{n_a,p-1} \\ a_{n_a,p} \\ c_{1,1} \\ c_{1,2} \\ \vdots \\ c_{n_c,p-1} \\ c_{n_c,p} \end{bmatrix}}_{\theta} + \hat{e}_k[t] \end{aligned} \quad (11)$$

or by stacking all time instants in a single vector

$$\underbrace{\begin{bmatrix} x_k[1] \\ x_k[2] \\ \vdots \\ x_k[N] \end{bmatrix}}_{x_k^N} = \underbrace{\begin{bmatrix} \varphi[\xi_k, 1]^T \\ \varphi[\xi_k, 2]^T \\ \vdots \\ \varphi[\xi_k, N]^T \end{bmatrix}}_{\Phi(\xi_k)} \cdot \underbrace{\begin{bmatrix} a_{1,1} \\ a_{1,2} \\ \vdots \\ a_{n_{a,p-1}} \\ a_{n_{a,p-1}} \\ c_{1,1} \\ c_{1,2} \\ \vdots \\ c_{n_{c,p-1}} \\ c_{n_{c,p}} \end{bmatrix}}_{\theta} + \underbrace{\begin{bmatrix} e_k[1] \\ e_k[2] \\ \vdots \\ e_k[N] \end{bmatrix}}_{e_k^N} \quad (12)$$

Finally, by pooling all the available datasets, the following regression model is obtained:

$$\underbrace{\begin{bmatrix} x_1^N \\ x_2^N \\ \vdots \\ x_K^N \end{bmatrix}}_X = \underbrace{\begin{bmatrix} \Phi(\xi_1) \\ \Phi(\xi_2) \\ \vdots \\ \Phi(\xi_K) \end{bmatrix}}_{\Phi(\xi)} \cdot \theta + \underbrace{\begin{bmatrix} e_1^N \\ e_2^N \\ \vdots \\ e_K^N \end{bmatrix}}_E \quad (13)$$

where $\Phi(\xi)$ is the regression matrix. Thus, the minimization problem of Eq. 7 may be rewritten as follows:

$$\theta = \arg \min_{\theta} \{E^T \cdot \Gamma \cdot E\} \quad (14)$$

where $\Gamma = I_N \otimes \gamma$ with $\gamma = [\gamma_1 \cdots \gamma_K]^T$ denoting the vector of weights, I_N the $N \times N$ identity matrix, and $\otimes$ the Kronecker product.

Due to the nonlinear dependence of the a priori, unknown residual sequence E on the parameter vector θ , the last equation leads to a nonlinear-weighted least squares problem, which has to be tackled by nonlinear optimization methods. However, nonlinear least squares techniques are sensitive to the initial parameter values and if no accurate estimates are available, the minimization procedure is very likely to converge to a local minimum. In order to avoid potential inaccurate convergence problems associated with arbitrary initial estimates, initial values for the coefficients of projection may be obtained by identifying conventional ARMA models for each of the K data

sets and expanding their parameters onto the PC basis (Zhao et al. 2012).

However, in the simpler PC-AR case, the regression model of Eq. 13 is linear, and thus, θ may be estimated by the linear weighted least squares (WLS) estimator:

$$\hat{\theta} = (\Phi(\xi)^T \cdot \Gamma \cdot \Phi(\xi))^{-1} (\Phi(\xi)^T \cdot \Gamma \cdot Y) \quad (15)$$

The simplest and obvious choice for the weights γ_k is to set them equal to unity, with this selection leading to the ordinary least squares (OLS) estimator. However, it has been shown for the similar FP-ARX models (Kopsaftopoulos and Fassois 2006) that the weights that lead to consistent and asymptotically efficient estimates are those equal to the standard deviation of the residual sequence, that is,

$$\gamma_k = \sigma_{e_k}, \quad \text{for } k = 1, 2, \dots, K \quad (16)$$

Due to the fact that the true residual standard deviation is typically unknown, it has to be replaced by an appropriate estimate, typically obtained via the OLS estimation which is used as a first stage. The WLS estimator may then be realized in an iterative manner till convergence of the PE criterion is achieved.

Model Structure Selection

Model structure selection is the optimization procedure during which models corresponding to various candidate “structures” are estimated and weighted according to an appropriately selected fitness criterion. Model “fitness” may be considered in terms of various criteria, such as the residual sum of squares (RSS)

$$\text{RSS} = \sum_{k=1}^K \sum_{t=1}^N e_k[t]^2 \quad (17a)$$

or the equivalent mean squared error (MSE)

$$\text{MSE} = \frac{\text{RSS}}{NK} \quad (17b)$$

However, the limitation of using such criteria is that they may monotonically decrease with

increasing model structure and as a result may lead to overfitting. For this reason, caution has to be exercised and a compromise has to be made between model accuracy and model structure complexity.

The model structure selection procedure for a PC-ARMA model concerns the determination of the set of 3 integers (n_a, n_c, P) for obtaining the best fitting model, and a combinatorial approach, that is, estimating PC-ARMA models for various integer sets, may be used. However, a simpler approach that may lead to sparse PC basis by selecting only a subset of the initial PC functional subspace may be described as follows (Poulimenos and Fassois 2006):

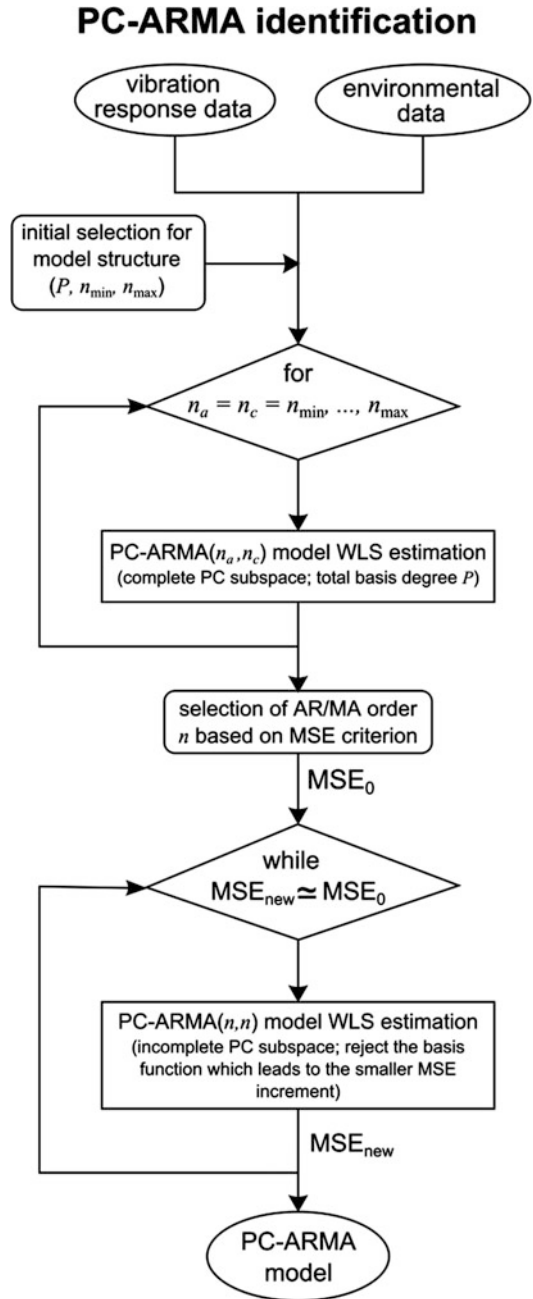
Phase I. AR/MA orders selection. In order to “decouple” the selection of the model orders (n_a, n_c) from that of functional subspaces, their interaction has to be minimized. For this reason a high-degree P and the complete set of PC basis functions may be initially adopted. When employing these, AR/MA model order selection may be achieved through trial-and-error techniques based on the values of the fitness function.

Phase II. PC basis subspace selection. The aim of this phase is the optimization of the PC functional subspace, in the sense of reducing model complexity without significantly reducing model accuracy. This may be accomplished by dropping, one at a time, the PC basis functions whose removal does not cause significant loss of the criterion employed.

A flowchart of the complete PC-ARMA identification procedure is given in Fig. 1.

Validation

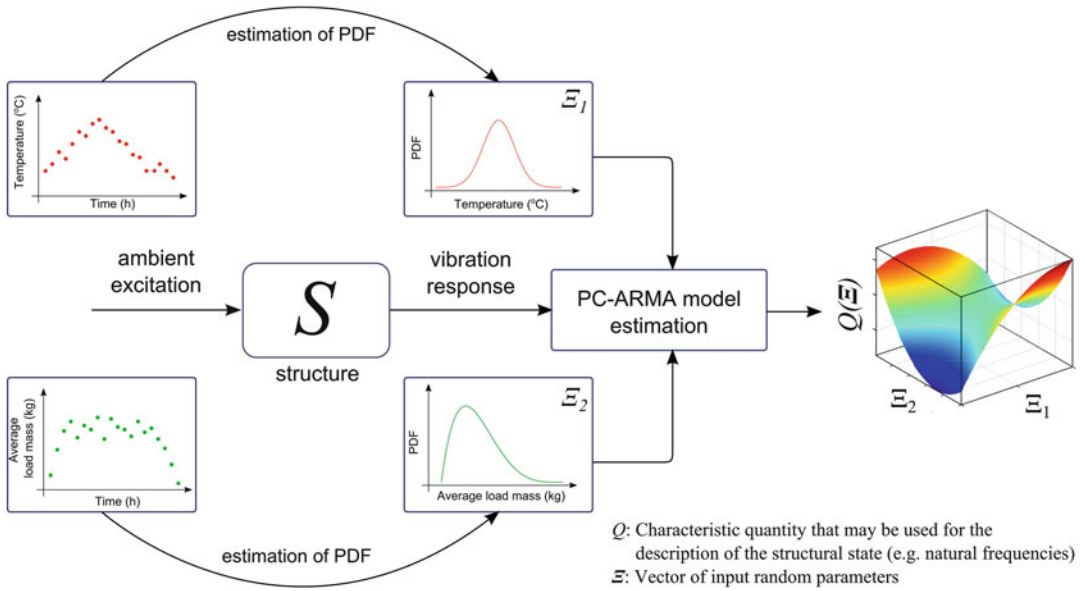
Once a PC-ARMA model has been obtained, it must be validated with respect to the assumptions behind the identification method. A standard validation procedure in the simple ARMA case involves the testing of Gaussianity and, in particular, the whiteness of the identified model’s one-step-ahead prediction error sequence. In the PC-ARMA case this assumption has to be validated for each of the recorded datasets, while the



Stochastic Structural Identification from Vibrational and Environmental Data, Fig. 1 Flowchart of the PC-ARMA model identification procedure

additional assumption of cross-uncorrelatedness between the datasets has to be validated.

The validation procedure of the estimated PC-ARMA model may be based on the typical



Stochastic Structural Identification from Vibrational and Environmental Data, Fig. 2 The complete stochastic structural identification framework based on PC-ARMA models

cross-validation principle which presumes the separation of the processed signal into an estimation set and a smaller validation set of data (Ljung 1999).

Summary of the Method

- Step 1.** Collect K datasets of vibration response and environmental variables measurements from the structure under study for a range of operating conditions.
- Step 2.** Fit a probability density function to the environmental variables based on the available data and select the corresponding family of PC basis functions.
- Step 3.** Select a high total degree for the PC basis P ; estimate PC-ARMA models of various orders (e.g., considering common AR/MA order, from $n_a = n_c = n_{\min}$ up to $n_a = n_c = n_{\max}$); and select the final orders based on the “fitness” criterion (e.g., MSE).
- Step 4.** Estimate PC-ARMA (n_a, n_c) models by rejecting each time a number of PC basis function from the complete PC subspace. Reject the functions whose removal does not imply significant reduction of the “fitness” criterion value.

The complete stochastic structural identification framework based on PC-ARMA models is illustrated in the abstract schematic representation of Fig. 2. It is noted that the estimated PC-ARMA model parameters may be used for the calculation of the structure’s modal characteristics, such as the modal frequencies and damping ratios, as a function of the input variable vector Ξ for the entire event space Ω .

Numerical Case Study

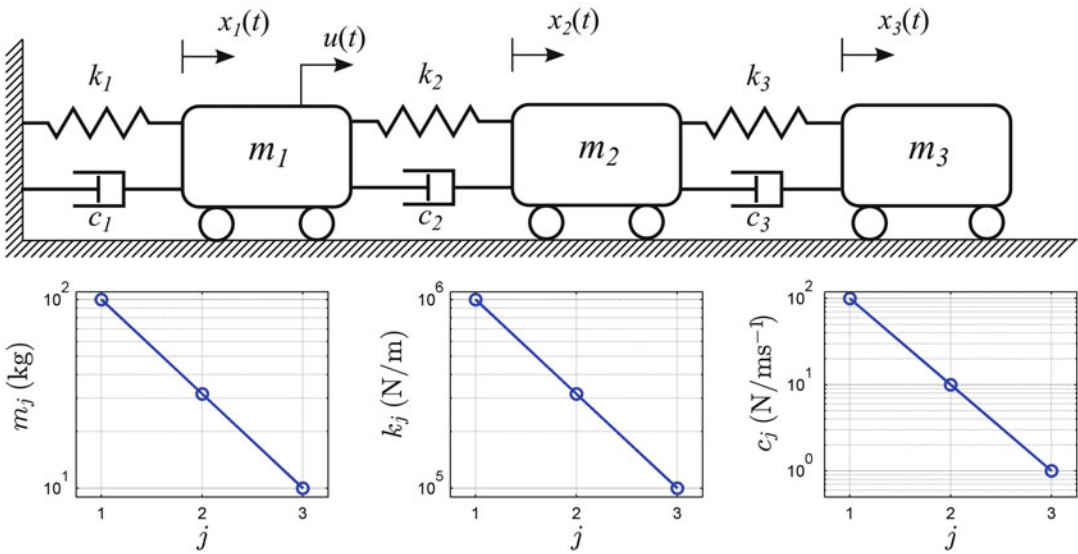
The modeling of a 3-DOF system (Fig. 3) is presently considered for demonstrating the workings of the PC-ARMA identification framework. The mechanical properties of the system are:

$$m_1 = 10^2, \quad m_2 = 10^{1.5}, \quad m_3 = 10 \text{ (kg)}$$

$$k_1 = 10^6, \quad k_2 = 10^{5.5}, \quad k_3 = 10^5 \text{ (N/m)}$$

$$c_1 = 100, \quad c_2 = 10, \quad c_3 = 10, \text{ (N/ms}^{-1}\text{)}$$

The system is considered to operate under varying conditions of temperature and humidity.



Stochastic Structural Identification from Vibrational and Environmental Data, Fig. 3 Three degree of freedom system and its properties

The system's stiffness coefficients are considered to be linearly depended on the temperature through the following relationship:

$$k'_i = k_i[1 - (T - T_{\text{ref}})k_T]$$

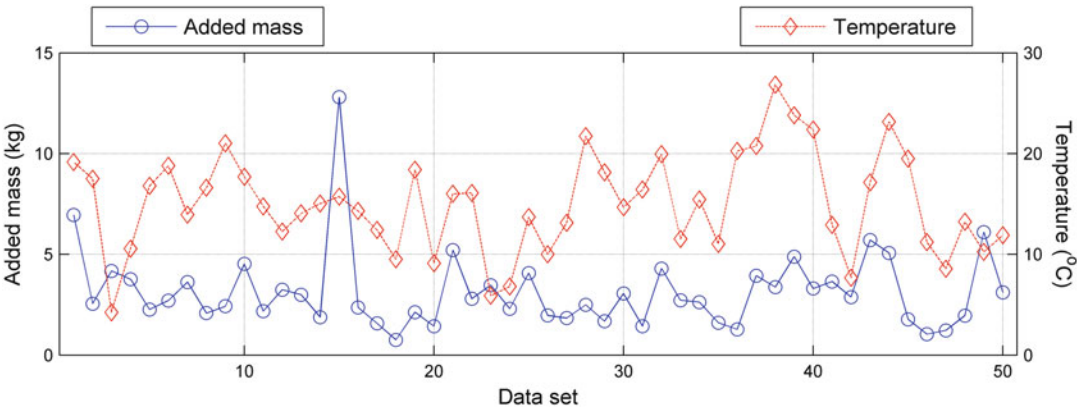
with $k_T = 5 \times 10^{-3}$ denoting the thermal coefficient and $T_{\text{ref}} = 20^\circ\text{C}$ the reference temperature, while humidity is modeled as mass increment due to trapped humidity. For the sake of simplicity, in this test case study, it is assumed that the temperature is uniform with no spatial gradients or temporal variations within each experiment, while the mass load is evenly distributed over the system.

In order to identify an inverse model for this system under various operating conditions, the system's response (acceleration of mass m_3) is simulated for various operating condition scenarios. Specifically, 50 samples of the added mass and temperature variables are drawn, by means of the Latin hypercube sampling (LHS) method, from their theoretical pdfs which are considered to correspond to a log-normal distribution with mean value equal to 1 and variance equal to 0.25 for the mass load, that is, $\Xi_1 \sim \ln N(1, 0.25)$, and a normal distribution with mean value equal to

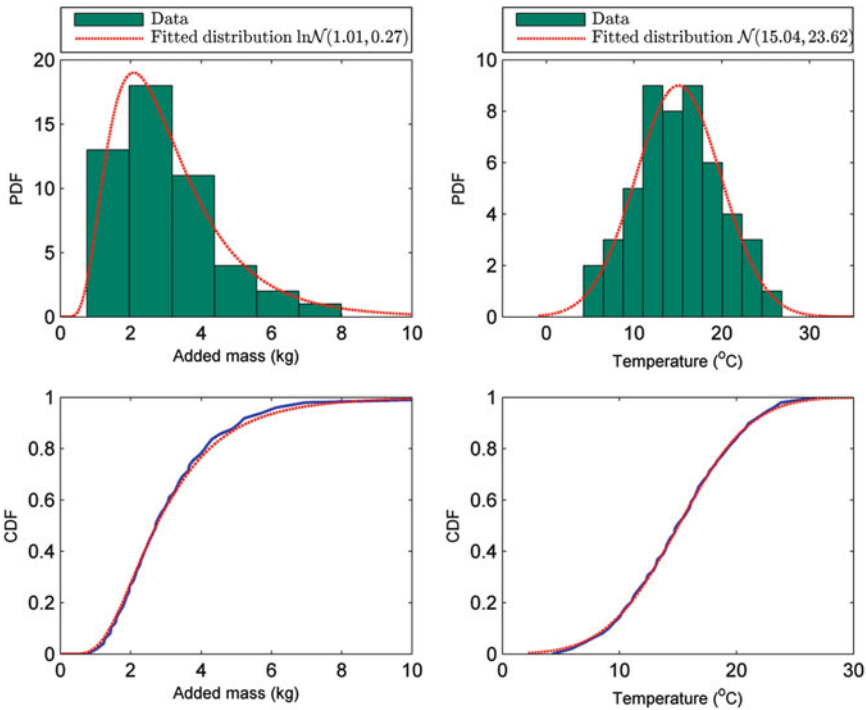
15 and variance equal to 25 for the temperature, that is, $\Xi_2 \sim N(15, 25)$. The 50 samples of the input variables are shown in Fig. 4.

In a real-world structural identification application, where no information is available regarding the true pdfs of the input random vector, someone could use maximum likelihood estimation fitting of the environmental condition data values to a parametric distribution. The results of such a fitting of the data onto pdfs are shown in Fig. 5. Based on this fitting and after transforming the pdf of the mass load into a normal distribution by using the natural logarithm, the Hermite polynomials may be selected for the construction of the multivariate PC basis functions.

For each set of these "environmental" conditions, a white noise realization of the excitation force $u(t) \sim N(0, 10^4)$ is used for the excitation of the system. System's response is obtained by using the Runge-Kutta algorithm for 1,800 points with sampling frequency equal to 64 Hz. After discarding the first 1,000 samples in order to avoid transient effects, the rest of the data are separated into an estimation set consisting of 500 samples and a validation set of 300 samples.



Stochastic Structural Identification from Vibrational and Environmental Data, Fig. 4 Input variable values for the 50 simulation experiments conducted

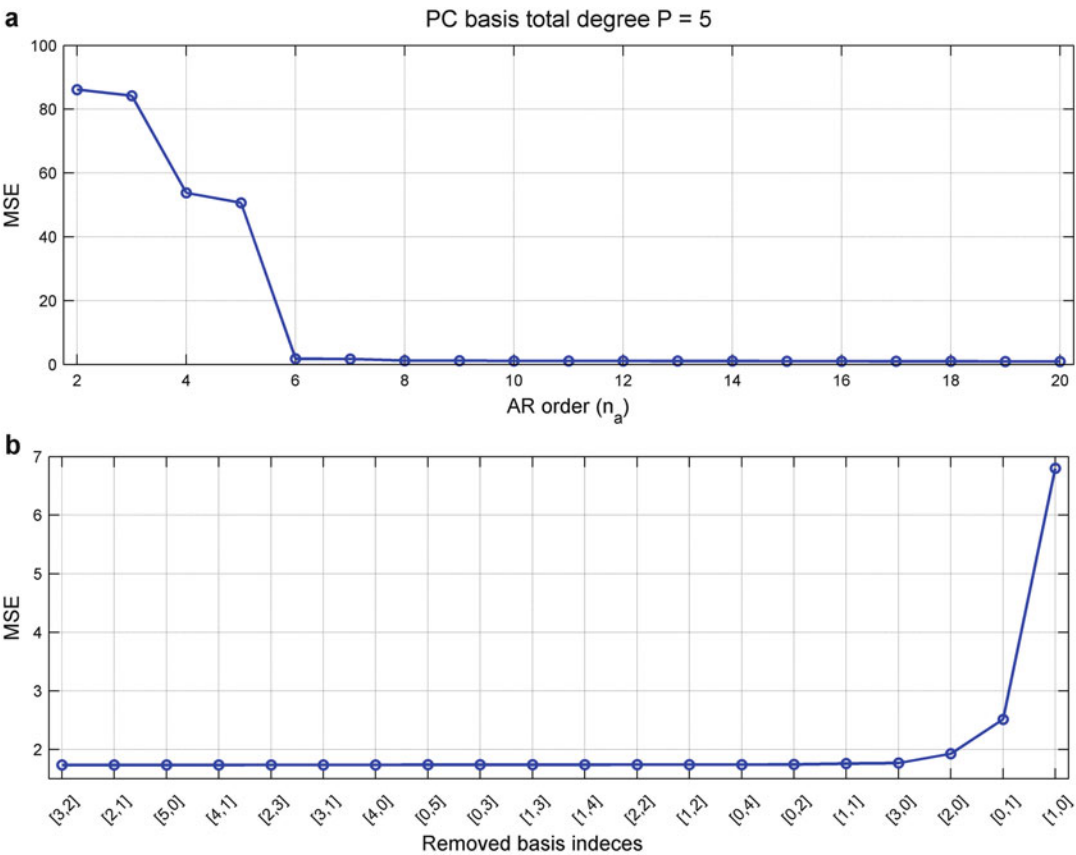


Stochastic Structural Identification from Vibrational and Environmental Data, Fig. 5 The empirical density and distribution functions calculated based on the 50 realizations of the input variables, compared to the

corresponding probability density and cumulative distribution functions estimated by the maximum likelihood method

The estimation dataset (50 sets $\times$ 500 samples) is used for the identification of the system through simple PC-AR models in what follows. PC-AR model structure selection is based on the

MSE criterion and the procedure described in section 2.2. The maximum total degree of the PC multivariate basis is initially selected equal to 5, which using Eq. 9 with $P = 5$, and



Stochastic Structural Identification from Vibrational and Environmental Data, Fig. 6 PC-AR model structure selection results: (a) MSE criterion values for WLS-based estimated PC-AR models with $n_a = 2, 3,$

$\dots, 20$ and $P = 5$, and (b) MSE criterion values obtained from a PC-AR(6) model by dropping, one at a time, the PC basis functions of the initial subspace

$L = 2$ gives a total number of 21 candidate basis functions. These functions are firstly used for the AR order selection by estimating PC-AR models of various orders ($n_a = 2, 3, \dots, 20$) by means of the WLS estimator. The results illustrated in Fig. 6a show an AR order equal to six as appropriate for capturing the dynamics of the simulated system. At the second stage of the structure selection procedure, a sparse PC basis is sought. As shown in Fig. 6b, this comes down to a basis of only 4 basis functions which are given in Table 1. It may be observed that the basis indicates a linear relationship of the PC-AR model parameters with respect to temperature and quadratic with respect to humidity (mass load).

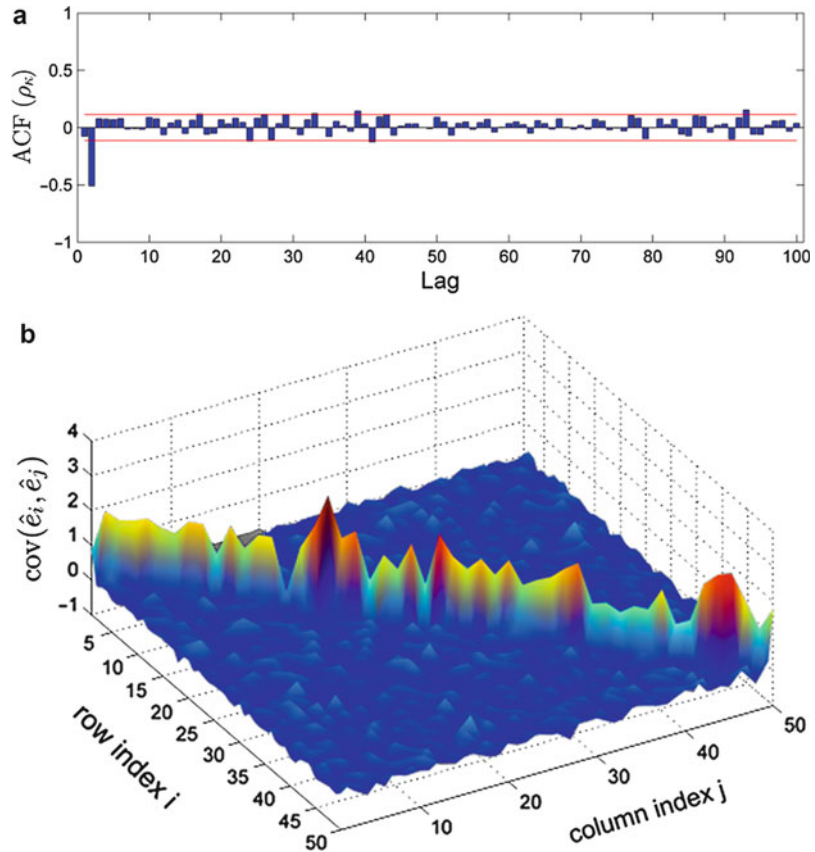
Stochastic Structural Identification from Vibrational and Environmental Data, Table 1 Multivariate indices of the finally selected PC basis functions of the PC-AR (6) model

	<i>l</i>	
	1	2
<i>d</i> (1)	0	0
<i>d</i> (2)	1	0
<i>d</i> (3)	0	1
<i>d</i> (4)	2	0

The estimated PC-AR(6) model is successfully validated by calculating the one-step-ahead prediction errors for the validation dataset (50 sets $\times$ 300 samples). Indicative results for the autocorrelation function of the resulting

Stochastic Structural Identification from Vibrational and Environmental Data,

Fig. 7 PC-AR(6) model validation results. (a) The autocorrelation function for the prediction error sequence of the validation set (simulation experiment 2; time lags 100; 95 % confidence limits) and (b) the covariance matrix of the prediction error matrix E

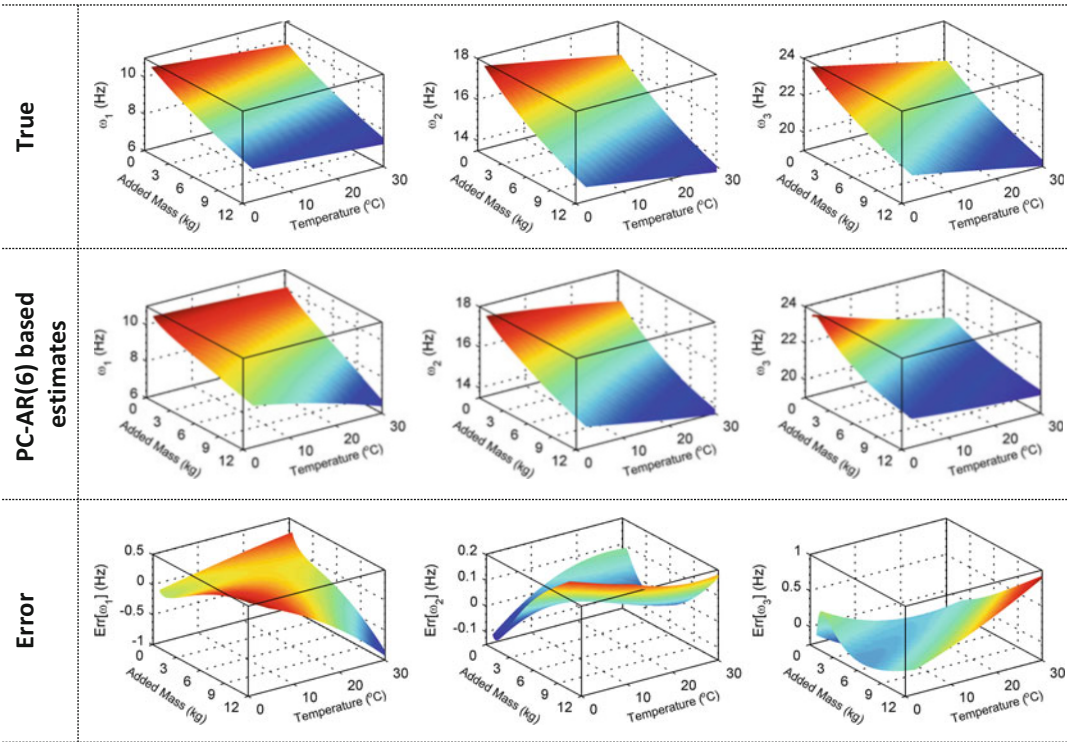


prediction error sequence corresponding to the second simulation experiment are shown in Fig. 7a, while also the assumption of uncorrelatedness between the various prediction error sequences may be qualitatively validated by visually inspecting the residual covariance matrix which is shown in Fig. 7b.

Finally, the estimated coefficients of projection may be used for calculating the PC-AR(6)-based estimates of the system natural frequencies and thus for constructing a surface that shows the dependence of the natural frequencies of the 3-DOF system on the two input random variables. The results are plotted against the theoretical curves in Fig. 8. It may be observed that there is a good agreement between the theoretical and the estimated natural frequencies with the highest errors, which are below 1 Hz, occurring at extreme values of the input variables which were not included in the estimation set.

Summary

Mechanical properties, and thus the dynamic behavior of an engineering structure, are significantly affected by environmental and operational conditions. A complete identification procedure has to lead to a consistent representation of the structural dynamics for a great percentage, if not the whole, range of the structure's operational spectrum. This may be achieved either by a multi-model approach, that is, estimating local models for every new set of environmental conditions data or using functional models that incorporate into their parameters the structure's dependency on a set of random input variables which correspond to the measured environmental conditions. In this way, after the estimation (training) of the functional model on a given set of measurements, its parameters may be extrapolated for any new set of environmental conditions. Presently, the functional models used are of the



Stochastic Structural Identification from Vibrational and Environmental Data, Fig. 8 PC-AR(6)-based natural frequency estimates compared to the theoretical

curves. The natural frequencies are plotted versus the input variables, that is added mass and temperature

ARMA form with parameters that are expanded on polynomial chaos basis. The latter has the important feature of being orthogonal to the probability density function of the input variable vector, providing in this way the best convergence rate with increasing functional subspace.

Cross-References

- [Operational Modal Analysis in Civil Engineering: An Overview](#)
- [Response Variability and Reliability of Structures](#)
- [Stochastic Analysis of Linear Systems](#)
- [System and Damage Identification of Civil Structures](#)
- [Uncertainty Quantification in Structural Health Monitoring](#)
- [Vibration-Based Damage Identification: The Z24 Bridge Benchmark](#)

References

- Andersen P (1997) Identification of civil engineering structures using vector ARMA models. Aalborg University, Department of Mechanical Engineering
- Blatman G, Sudret B (2010) An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probabilist Eng Mech* 25:183–197
- Doebeling S, Farrar C, Prime M (1998) A summary review of vibration-based damage identification methods. *Shock Vib Dig* 30:91–105
- Fassois SD, Kopsaftopoulos FP (2013) Statistical time series methods for vibration based structural health monitoring. In: Ostachowicz W, Guemes A (eds) *New trends in structural health monitoring*. Springer, Wien/London, pp 209–264
- Kopsaftopoulos FP, Fassois SD (2006) Identification of stochastic systems under multiple operating conditions: the vector dependent FP-ARX parameterization. In: *Proceedings of the 14th Mediterranean conference on control and automation*. Ancona
- Kopsaftopoulos FP, Fassois SD (2013) A functional model based statistical time series method for vibration based damage detection, localization, and magnitude estimation. *Mech Syst Signal Process* 39:143–161

- Ljung R (1999) *System identification: theory for the user*, 2nd edn. Prentice Hall, Upper Saddle River
- Poulimenos AG, Fassois SD (2006) Parametric time-domain methods for non-stationary random vibration modelling and analysis: a critical survey and comparison. *Mech Syst Signal Process* 20:763–816
- Sohn H (2007) Effects of environmental and operational variability on structural health monitoring. *Philos Trans R Soc A* 365:539–560
- Soize C, Ghanem R (2004) Physical systems with random uncertainties: chaos representations with arbitrary probability measure. *SIAM J Sci Comput* 26:395–410
- Tóth R (2010) Identification and modeling of linear parameter-varying systems. *Lecture notes in control and information sciences*, vol 403. Springer, Berlin
- Xiu D, Karniadakis GE (2002) The Wiener–Askey polynomial chaos for stochastic differential equations. *SIAM J Sci Comput* 24:619–644
- Zhao Y, Huang B, Su H, Chu J (2012) Prediction error method for identification of LPV models. *J Process Control* 22:180–193

Strengthened Structural Members and Structures: Analytical Assessment

Georgia E. Thermou
Department of Civil Engineering, Aristotle
University of Thessaloniki, Thessaloniki, Greece

Synonyms

Assessment; Buildings; Interface; Jacketing; Repair; Retrofitting; Shear transfer; Strengthening; Structural interventions; Structures

Introduction

The catastrophic earthquakes that struck several countries over the last 20 years brought to light the susceptibility of the existing building stock (Fig. 1). Old-type reinforced concrete (R/C) structures are characterized by insufficient reinforcement detailing (lack of stirrups for ensuring a certain ductility level, indirect supports, insufficient anchorages of bars), nonuniform distribution of stiffness and/or mass along the height of the building, insufficient foundation system, poor

quality of materials, and various other weaknesses such as increased loading due to change of use and corrosion of reinforcement.

The majority of multistory R/C buildings in southern Europe were built in the first half of the twentieth century. Structures were designed for gravity loads only by implementing the allowable stress design philosophy which did not allow any control of the mode of failure and the corresponding deformation capacity of the individual members. Taking Greece as an example, the first seismic code was introduced in 1959, and R/C walls were introduced in construction in the 1960s (often without extending up to the foundation of the building). The modern seismic codes were introduced more than 20 years later, in the mid-1980s. The multistory R/C buildings of the 1950s represent the cutting edge of the construction technology for gravity load-designed frame buildings. Information regarding the material, detailing, and geometrical characteristics of representative low- to medium-rise (up to 8 stories) R/C buildings real structures found in the urban areas built between the 1920s and 1960s is presented in Table 1 (Thermou and Palaioxorinou 2013). The lack of a continuous vertical load path along the height of the buildings is a common feature of the buildings of that era. There is no typical floor since the dimensions of the columns and beams change from story to story. Often in-plan column layout does not follow a grid pattern, hence leading to indirect supports. Representative typical floor plan layouts are presented in Fig. 2.

The result of these systematic deficiencies of existing buildings is a decreased level of seismic protection, increased seismic vulnerability, and hence extensive damage expected in future seismic excitations. This is a rather alarming issue considering the socioeconomic impact of severely damaged buildings or collapses in future strong ground motions. The recommended solution for this category of buildings, which comprises the vast majority of the existing stock, is retrofitting (strengthening) with a view to upgrading their seismic capacity and meeting the current standards for seismically designed structures. Adopting global intervention (GI)



Strengthened Structural Members and Structures: Analytical Assessment, Fig. 1 Damage patterns in old-type R/C members and structures; typical multistory R/C residential buildings with pilotis in Greece (Source of figures: author’s personal files)

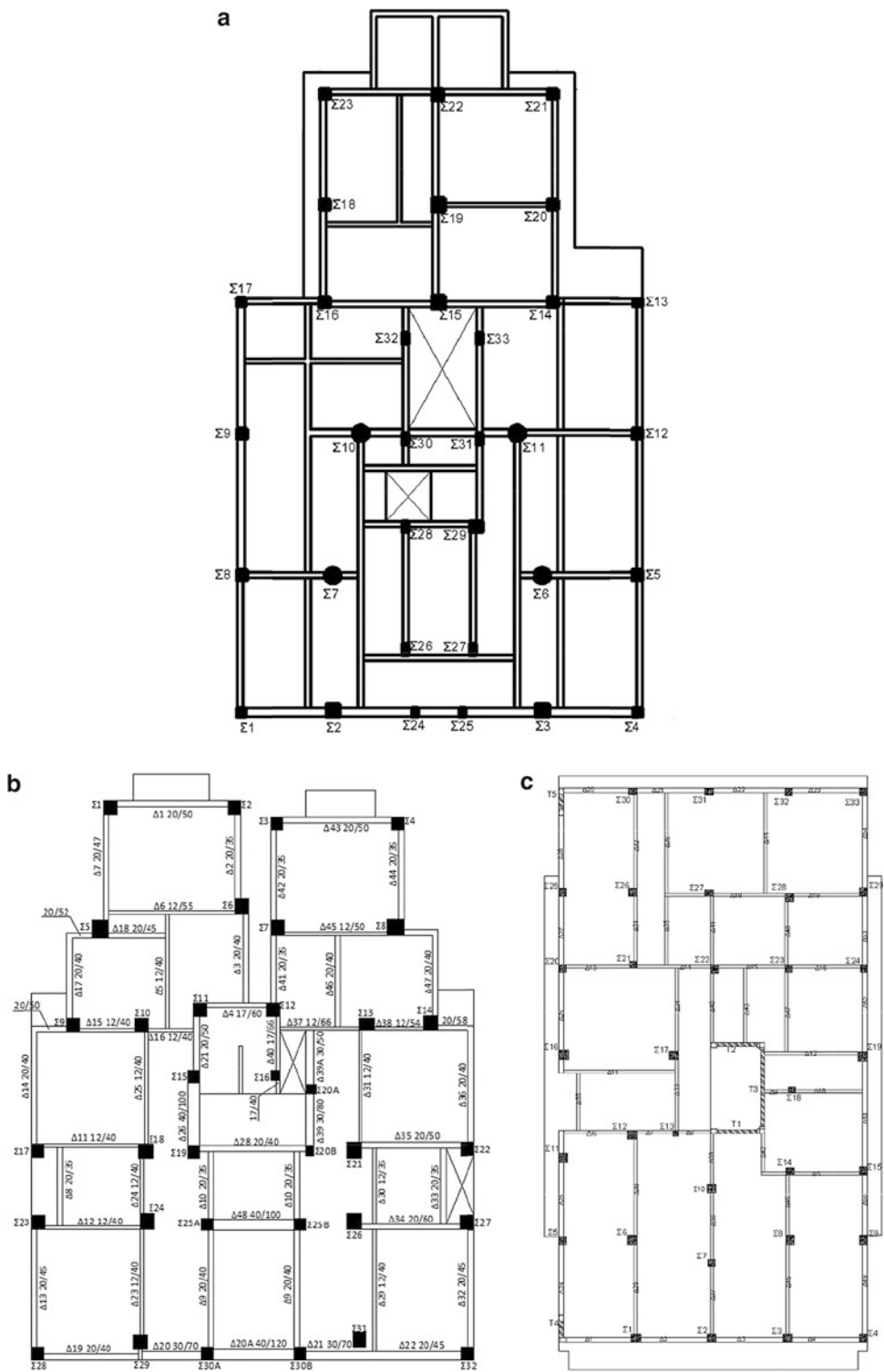
Strengthened Structural Members and Structures: Analytical Assessment, Table 1 Information of the typical construction practice followed in the 1920s–1960s in Greece (Thermou and Palaioxorinou 2013)

Construction practice characteristics	
Concrete grade	B120 ÷ B160 ^a ($f_{ck} \cong 8\text{--}10\text{ MPa}$)
Steel grade (longitudinal and transverse reinforcement)	Smooth StI ($f_{yk} = 220\text{ MPa}$) ^b
Typical column cross section dimensions	250 ÷ 600 mm
Typical diameters of column longitudinal reinforcement bars	Ø14 ÷ Ø20
Typical column transverse reinforcement	Ø6/250 ÷ 300 mm ^c
Typical column longitudinal reinforcement ratio	7‰ ÷ 9‰ < 10‰
Typical beam cross section dimensions	150 × 300 ÷ 300 × 600 mm
Typical diameters of beam longitudinal reinforcement bars	Ø10 ÷ Ø18
Typical beam transverse reinforcement	Ø6/200 ÷ 250 mm ^c
Typical thickness of masonry infill walls	100 mm for internal, 200 mm for external
Typical wall cross section dimensions	Length: 1.5 ~ 3.6 m, thickness: 150 ~ 250 mm, boundary elements: length 200 mm, 4Ø12, web: #Ø8/250
Anchorage/lap splice construction practice	Longitudinal reinforcement with hooks with arbitrary lengths
	Stirrups anchored with 90° hooks
	Unconfined lap splices

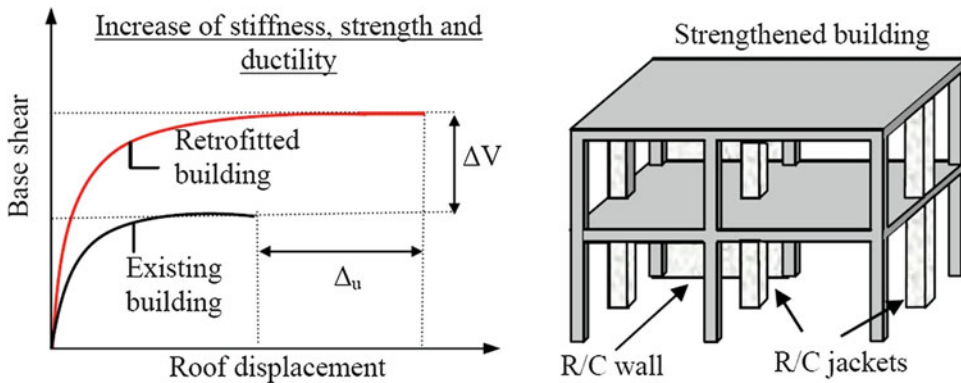
^aAs per Greek Royal Decree of 18/2/1954. Later concrete grade B225 ($f_{ck} \cong 14\text{ MPa}$) was introduced in construction
^bAs per DIN 1045 (1936). Later longitudinal reinforcement steel grade was changed to StIII ($f_{yk} = 400\div420\text{ MPa}$)
^cDiameter Ø8 was rarely applied

methods (e.g., R/C jacketing, R/C infill walls) leads to the modification of strength and stiffness (Fig. 3). They are used in systems with a high flexibility to sway, in torsionally unbalanced systems, or where strengthening of the existing building is required (e.g., increase of base shear strength in buildings with an open first story,

Fig. 1). GI methods cannot be used separately from local intervention (LI) methods since they are used to accommodate premature failure modes (Thermou et al. 2012a). LI methods as, for example, the intervention methods with composite materials are considered not to alter the stiffness or flexural strength of the retrofitted



Strengthened Structural Members and Structures: Analytical Assessment, Fig. 2 Representative ground floor plan layouts of the R/C multistory buildings in Greece of the (a) 1920s, (b) 1950s, and (c) 1960s



Strengthened Structural Members and Structures: Analytical Assessment, Fig. 3 Retrofitting strategy: stiffness, strength, and ductility increase (Thermou and Elnashai 2006)

members, but only affect their post-yielding deformation capacity through confinement and by suppressing premature modes of failure (e.g., brittle shear or lap splice failure).

The decision on the extent of the interventions (i.e., which floors, which elements of the floor, and which design parameters are going to be modified) and its impact on the modification of selected global response indices may be assessed through analytical modeling of the strengthened structure. This implies that there should sufficient information relative to the effectiveness of the intervention methods adopted at member level. The addition of a new concrete layer (e.g., flexural strengthening of beams) or a new R/C member (e.g., R/C jacketing, R/C infill walls) in various retrofit techniques entails issues related to the connection between existing and newly cast concrete. The response of the composite member, and subsequently of the whole structure, depends largely on the response characteristics of the interface, since slip takes place and shear transfer mechanisms are mobilized. Estimating response indices such as stiffness and strength in a reliable manner requires the development of analytical models that will consider the phenomena that take place along interfaces due to slip and their interaction.

This contribution aims to present the framework for the analytical assessment of strengthened structures when the intervention methods adopted result in composite members with multiple monolithic phases. The slippage introduced

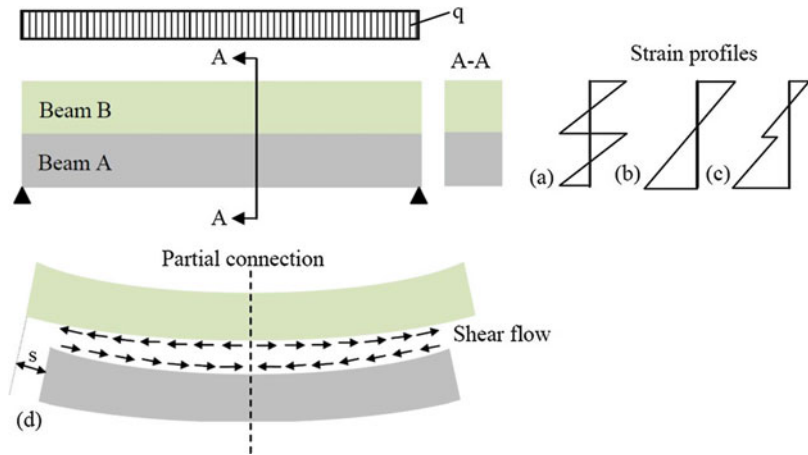
due to the discontinuity in the normal strain along the interfaces is rather determinant for the response at member level. One of the objectives is to highlight the role of the interface characteristics and present the parameters that influence the interconnection between the existing and new member. The shear transfer mechanisms mobilized due to relative slippage of the contact interfaces are presented in the following, and their interaction is discussed. Information is also provided for the influence of cyclic loading on the shear resistance mechanisms. The fundamentals for the development of analytical models which take into account slip at the interfaces are discussed. The interrelation between the effects that the changes at local level may have on global level is identified as a key issue in the assessment of strengthened structures. Moreover, the alternative simplified design approach suggested by codes, where the monolithic factors are utilized as to indirectly account for the adverse effect of slippage on response, is presented. The last section of this entry focuses on the alternative analytical design approaches that could be followed for one of the most popular global intervention techniques, the R/C jacketing technique.

Implications of the Addition of New Concrete Layers or New Elements

The response of a composite beam, as the one shown in Fig. 4, depends on the degree of

Strengthened Structural Members and Structures: Analytical Assessment,

Fig. 4 Strain distribution profiles for (a) no connection; (b) perfect and (c) partial connection between beam A and beam B. (d) Shear flow along the interface



interconnection between beam A and beam B. In one extreme case, the case of zero friction (no connection), beam B could slide relative to beam A without mobilizing any kind of shear resistance (Fig. 4a). In an ideal situation, the composite members (beam A and beam B) would behave as monolithic (perfect connection, Fig. 4b). In real conditions, partial connection between the two bodies is expected, which implies that relative slip, s , between the two bodies is allowed, thus mobilizing the shear transfer mechanisms at the interface (Fig. 4c). The discontinuity in the strain profile corresponds to the slip, s , at the horizontal interfaces (Fig. 4c). Curvature, which is the slope of the strain profile on the cross section, is the same between the various slipping parts. This is a common assumption in the analysis of composite (layered) beams (e.g., glulam timber beams, composite steel beams, etc.). The shear flow, T , at the interface is estimated by

$$T = \frac{V \cdot S}{I} \quad (1)$$

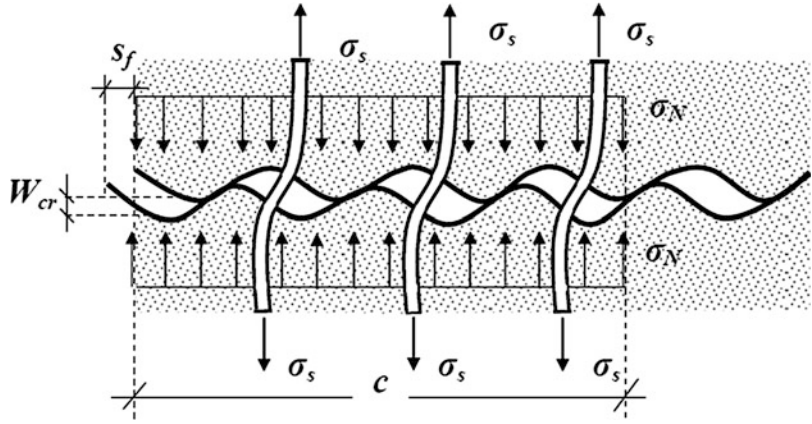
where S is the first moment of area, I is the moment of inertia of the composite cross section, and V is the shear force on the member.

Flexural enhancement of beams with the addition of a new reinforced concrete layer in the tension zone follows the principles presented for the composite beam of Fig. 4. The same approach can also be followed in the case of composite

cross sections where an outer shell is placed around the core of the cross section, as in the case of R/C jacketing of an existing column, or when an infill wall is added to an existing bay by incorporating the existing columns.

The preparation of the interface and the interconnection measures taken between existing and new member play a key role in the interfacial force transfer and relative slip occurring along the contact surface. The bond between the monolithic phases of the composite member can be enhanced by roughening of the substrate surface, addition of dowels placed through holes driven on the substrate surface, and when applicable (e.g., R/C jacketing technique) welding of the new longitudinal bars to the existing ones through deformed reinforcing bars bent into a U-shape. Moreover, experimental evidence has shown that differential shrinkage and stiffness of new-to-old concrete interfaces, as well as the compressive strength of the added concrete to an existing concrete substrate, influence the response of composite system where concrete-to-concrete interfaces exist (e.g., Choi et al. 1999). For example, in the experimental study conducted by Júlio et al. (2006), it was found that the increase of the concrete compressive strength of the new layer compared to that of the existing one leads to an enhancement of the compressive strength of the interface, i.e., taking into account the compressive strength of the weakest concrete seems to be conservative. Recently, Santos and Júlio (2014) proposed

Strengthened Structural Members and Structures: Analytical Assessment, Fig. 5 Slip at a concrete interface crossed by reinforcement



design expressions where the effects of differential shrinkage and differential stiffness at the concrete-to-concrete interface are considered.

Shear Transfer Along Concrete-to-Concrete Interfaces

Shear transfer along interfaces has been a subject of continuous research since the 1970s. A variety of analytical models are available for modeling the main shear transfer mechanisms (friction and dowel resistance) by considering that they act separately or jointly. There are models where all forces are transferred through reinforcement (e.g., Birkeland and Birkeland 1966; Walraven 1981), whereas others that, apart from reinforcement contribution, include a cohesion term (e.g., Mattock and Hawkins 1972; Vecchio and Collins 1986; Tassios 1986). In the modified compression field theory (MCFT) (Vecchio and Collins 1986), the role of the cohesion term is played by the aggregate interlock mechanism, whereas in the model of Tassios (1986), by friction owing to the clamping action of reinforcement normal to the interface. In the rest of the models, the cohesion term corresponds to the friction resistance developed along the interface.

Shear transfer is affected by the roughness of the sliding planes, by the characteristics of the reinforcement, by the compliance of concrete, and by the state of stress in the interface zone.

Mechanisms that resist sliding (slip) are (i) aggregate interlock between contact surfaces, including any initial adhesion of the jacket concrete on the substrate; (ii) friction owing to the clamping action of reinforcement normal to the interface; and (iii) dowel action of any properly anchored reinforcement crossing the sliding plane. Dowel action develops by three alternative mechanisms, namely, by direct shear, by kinking, and by flexure of the bars crossing the contact plane. The relationship that describes the contribution of the individual shear transfer mechanisms is

$$\begin{aligned}\tau_{\text{tot}} &= \tau_{\text{agr}} + \tau_f + \tau_D = \tau_{\text{agr}} + \mu\sigma_N + \tau_D \Leftrightarrow \\ \tau_{\text{tot}} &= \tau_{\text{agr}} + \mu(\sigma_c + \rho\sigma_s) + \tau_D \\ &= \tau_{\text{agr}} + \mu(vf_c + \rho\sigma_s) + \tau_D\end{aligned}\quad (2)$$

where τ_{agr} represents the shear resistance of the aggregate interlock mechanism, μ is the interface shear friction coefficient, σ_N is the normal clamping stress acting on the interface, and τ_D is the shear stress resisted by dowel action, F_D , in cracked reinforced concrete. The clamping stress represents any normal pressure, p , externally applied on the interface, but also the clamping action of reinforcement crossing the contact plane, σ_s is the axial stress of the bars crossing the interface, ρ is the corresponding reinforcement area ratio, $v = N/(\sigma_c f_c) = \sigma_c / f_c$ is the normalized axial load at the interface of

area A_c , and f_c is the concrete compressive strength (Fig. 5). The first two terms in Eq. 2 collectively represent the *contribution of concrete* as they depend on the frictional resistance of the interface planes.

As it is mentioned in Model Code (2010) and the Greek Code for Interventions (KANEPE 2013), in real structures the various mechanisms interact, thereby affecting each other as a function of the shear slip. The aggregate interlock including any initial adhesive decreases at low slip values, whereas the maximum contribution of frictional and dowel resistance occurs at different slips. Therefore, Model Code (2010) as well as the Greek Code for Interventions (KANEPE 2013) introduce interaction factors which consider that the dowels crossing the interface are subject to bending and axial forced simultaneously and that the maximum values of the different shear resistance mechanisms occur at different slips. The interaction factors depend on various parameters such as the magnitude of the expected slip at the interface, the diameter of the reinforcing bars crossing the interface, the concrete compressive strength, cyclic loading, etc. Equation 2 is thus modified according to Model Code (2010) as follows:

$$\tau_{\text{tot}} = \tau_{\text{agr}} + \mu(\sigma_c + \kappa \cdot \rho \sigma_s) + \alpha \cdot \rho \cdot \sqrt{f_y \cdot f_c} \leq \beta \cdot v \cdot f_c \quad (3)$$

where the coefficients κ and α refer to the interaction factors for interface roughness and flexural resistance, respectively, whereas β is a coefficient related to the compressive struts. For the case where dowels are driven into the interface, $\kappa = 0.5$ for roughened interfaces and $\kappa = 0$ for smooth interfaces, whereas in the latter case $\alpha = 1.5$ denoting that the main resistance mechanism is provided by the dowel action. Model Code (2010, Chap. 7) provides detailed information regarding the design of the interface in shear and how the interaction coefficients are modified for different surface roughness.

The Greek Code for Interventions (KANEPE 2013) considers also the interaction between the

shear transfer mechanisms by estimating the shear resistance of the interface as

$$\tau_{\text{tot}} = \beta_F \tau_f + \beta_D \tau_D \quad (4)$$

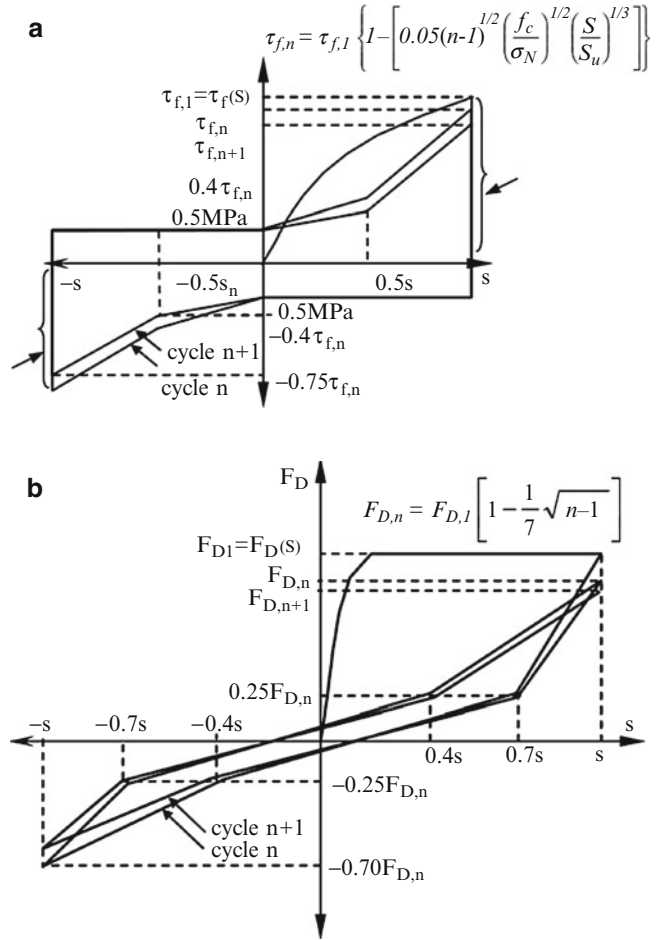
where the β_F and β_D are the interaction factors of the friction and dowel resistance, which are equal to $\beta_F = 0.4$ and $\beta_D = 0.7$ for slip values $s \leq 1.00$ mm. These coefficients can be further decreased to 0.5 when either the expected slip value cannot be predicted or when the external compressive force at the interface is negligible.

The friction resistance, τ_f (Eq. 2), is calculated by considering the coefficient of friction, μ , which depends on the classification of the roughness of the interface. The majority of the experimental studies provide a qualitative description of the roughness of the interface after application of the selected type of treatment (e.g., sandblasting, use of pneumatic chipping device). It is obvious that the characterization by visual inspection only may lead to inaccuracies. Therefore, the roughness of the interface should be quantified by specific indicators. Although several methods have been developed that can successfully quantify the roughness of concrete interfaces (Santos and Júlio 2013), design codes in their majority still provide qualitative descriptions of the interface (e.g., smooth, rough, or very rough). An exception is Model Code (2010) which suggests two methods: the sand area method and the one where the average roughness is estimated (i.e., measuring the average deviation of the profile from a mean line).

From the above, it is seen that describing in detail the mechanisms mobilized along interfaces due to slip and their interaction is a complex mechanical issue which becomes even more complicated considering the response of the interface under cyclic loading conditions where degradation should also be accounted for. The research conducted on the derivation of constitutive models that describe the combined shear force resistance mobilized along interfaces due to sliding both under cyclic imposed deformations is limited. Palieraki et al. (2012) based on the previous studies of Tassios and Vintzileou (1987) and Vintzileou and Tassios (1986, 1987)

Strengthened Structural Members and Structures: Analytical Assessment,

Fig. 6 Response to symmetric cyclic loading: (a) friction and (b) dowel resistance



and after carrying out a new experimental program proposed the degradation rules for frictional and shear resistance, which were also adopted by the Greek Code for Interventions (KANEPE 2013). The frictional and dowel resistance is reduced at each cycle, n , according to (see Fig. 6)

$$\begin{aligned} \tau_{f,n} &= \tau_{f,1} \cdot \tau_{\text{deg}} \\ &= \tau_{f,1} \left\{ 1 - \left[0.05(n-1)^{1/2} \left(\frac{f_c}{\sigma_N} \right)^{1/2} \left(\frac{s}{s_u} \right)^{1/3} \right] \right\} \end{aligned} \quad (5)$$

$$F_{D,n} = F_{D,1} \cdot D_{\text{deg}} = F_{D,1} \left[1 - \frac{1}{7} \sqrt{n-1} \right] \quad (6)$$

where $\tau_{f,1}$ and $F_{D,1}$ are the peak frictional and dowel resistance value, respectively, attained in

the first cycle, n is the number of cycles, s is the slip value, s_u is the ultimate slip value, σ_N is the normal clamping stress acting on the interface, and f_c is the concrete compressive strength.

Analytical Assessment of Strengthened Members and Structures

Analytical Model: Slip at the Interface

The success of repair/strengthening, as quantified by the improvement of strength and stiffness of the retrofitted members, depends on its entirety on the degree of collaboration between the various monolithic phases of the composite member. The response of the retrofitted member is modified proportionally to the slip that takes place along the interface. The calculation of slip

and its effects on the mechanical characteristics of the retrofitted members presuppose the use of analytical models where the interface characteristics need to be explicitly modeled. This comprises a quite challenging subject of continuous research.

In case of R/C jacketing of columns, Thermou et al. (2007, 2012b) have developed an analytical model for predicting the response under monotonic and reversed cyclic loading of structural members with *old-type* detailing, strengthened with R/C jacketing. The analytical model introduces one additional degree of freedom between the existing member and its outer R/C shell, thus allowing slip to take place at the interface between the existing member and the jacket. Shear resistance mechanisms, such as aggregate interlock, friction, and dowel action, are mobilized to resist slip. The magnitude of shear flow sustained along the contact surface is calculated by considering the states of stress of the composite member at a cracked cross section and at a point between successive cracks, in order to introduce in the flexural behavior the effect of the moment gradient (shear force magnitude). A detailed description of the analytical model for R/C jacketed members is presented in the last section of this article. The same approach has also been implemented after the necessary modifications for predicting the response of thin reinforced self-compacting three-jacketed beams (Chalioris et al. 2014). Moreover, Tsioulou and Dritsos (2011) developed an analytical model for predicting the flexural capacity of beams strengthened by the addition of a concrete layer in the tension or compression zone. The algorithm developed aimed at the evaluation of slip and shear stress distribution along the interface.

The analytical models, as the ones previously presented, may provide information relative to the effectiveness of the selected intervention method applied at member level. Stiffness and strength of the retrofitted (composite) members are estimated and assessed. The interrelation between the effects that the changes at local level may have on global level is a key issue in the assessment of strengthened structures. These

may be evaluated rapidly through standard Rayleigh-type or Stodola-type iteration (Clough and Penzien 1993), using secant-to-yield stiffness values for the individual members (Thermou et al. 2012c). The estimated fundamental translational response shape and period may guide definition of retrofit objectives. Additional lateral strengthening of the building (owing to the implicit relationship between stiffness and strength) by means of controlled stiffness addition along the building height may be required. Proportioning the stiffness of the individual floors is determined so as to even out large discrepancies in relative drift between successive floors detected in the fundamental response shape pattern.

Simplified Design Approach: The Use of Monolithicity Factors

A pragmatic design approach commonly adopted by codes of practice considers the monolithic approach for the analysis of composite members making use of properly defined *monolithicity factors* for obtaining the mechanical properties of the strengthened member. The redesign procedure is thus simplified substantially and is used extensively in practice for various intervention methods for different types of structural members (slabs, beams, columns, walls, or footings). The values given so far to these modification factors are empirical or semiempirical. The use of reliable analytical models (e.g., Thermou et al. 2007, 2012b) can be utilized for the derivation of monolithicity factors after extensive study of their sensitivity to the characteristics of the intervention method (e.g., Kappos et al. 2012).

The use of monolithicity factors simplifies substantially design calculations and is applicable to various types of structural members (slabs, beams, columns, walls, foundation elements) and intervention methods (e.g., R/C jacketing, R/C infill walls). These reduction factors are used to obtain the strength and deformation indices of the jacketed members and are applied to the respective properties of monolithic members with identical geometry. The monolithicity factors are defined as follows:

$$K = \frac{\text{Response index of the composite member}}{\text{Response index of the monolithic member with identical geometry}} \quad (7)$$

The various monolithicity factors usually used in design are:

- The monolithicity factors that refer to the deformation capacity indices such as the chord rotations at yield and ultimate, which are defined as

$$\begin{aligned} \text{Chord rotation at yield : } K_{\theta_y} &= \frac{\theta_{y,J}}{\theta_{y,M}}; \\ \text{Chord rotation at ultimate : } K_{\theta_u} &= \frac{\theta_{u,J}}{\theta_{u,M}} \end{aligned} \quad (8)$$

- The strength related monolithicity factors

$$\begin{aligned} \text{Shear strength : } K_v &= \frac{V_{J,\max}}{V_{M,\max}}; \\ \text{Moment at yield : } K_{M_y} &= \frac{M_{y,J}}{M_{y,M}} \end{aligned} \quad (9)$$

- The stiffness monolithicity factor

$$\text{Stiffness at yield : } K_k = \frac{K_{y,J}}{K_{y,M}} \quad (10)$$

where the subscripts J and M correspond to the composite (jacketed) cross section and to the identical monolithic cross section, θ_y , θ_u are the chord rotations at yield and ultimate, $V_{\max}$ is the maximum strength of the cross section, M_y is the yield moment, and K_y is the secant flexural stiffness at yield, defined as the ratio of the yield moment (M_y) to the yield curvature (φ_y).

Eurocode 8, Part 3 (§A.4.2, 2005), and the Greek Code for Interventions (KANEPE 2013) suggest the monolithicity factors presented in Table 2 to be used in the design of R/C members (slabs, columns, beams, and foundation system) when strengthened with additional reinforced concrete layers. In the case of Eurocode 8, Part3 (§A.4.2, 2005), the values of the monolithicity factors for columns of Table 2 should be used under the assumptions of: (i) full composite

action between old and new concrete, (ii) application of full axial load to the jacketed member, and (iii) application of the concrete properties of the jacket over the full section of the element. It is noted that the monolithicity factors for the design approach as proposed by the Greek Code for Interventions (KANEPE 2013) are subject to certain limitations, i.e., that the target strength increase of the jacketed member should not exceed twice that of the original. On the other hand, considering the code minima regarding the percentage of longitudinal reinforcement of the jacket, but also of the entire composite cross section (equal to 1 %), and the pertinent detailing rules (minimum thickness of the jacket is 70 mm), it is seen that the strength of the strengthened member far exceeds twice its original strength. In the case of EC8-Part 3 (2005), there is no restriction related to the increase of resistance of the R/C member due to jacketing.

In case of R/C column jacketing, the sensitivity of the monolithicity factors to the construction materials of the existing cross section (core) and the jacket, as well as to the percentage of longitudinal reinforcement of the jacket for increasing axial load, has been highlighted by Thermou et al. (2007). The analytical model for R/C jacketed columns, which is presented in the last section of this article, was utilized by Thermou et al. (2007, 2014) for performing an extensive parametric study for the derivation of monolithicity factors. For the range of parameters considered in their study, the lower and upper limits for the monolithicity factors were found to be $K_{\theta_y} = 1.17 \sim 4.85$, $K_{\theta_u} = 0.45 \sim 4.13$, $K_{M_y} = 0.32 \sim 0.99$, $K_r = 0.35 \sim 1.02$ and $K_k = 0.19 \sim 0.95$. Lampropoulos et al. (2012) utilized a computational (finite element) model of experimental specimens, and for the specific range of parameters examined, the monolithicity factors received the following values: $K_{\theta_y} = 1.05 \sim 3.00$, $K_{\theta_u} = 0.95 \sim 2.85$, $K_r = 0.70 \sim 0.90$ and

Strengthened Structural Members and Structures: Analytical Assessment, Table 2 Suggested values of the monolithicity factors by EC8-Part 3 (2005) and KANEPE (2013)

Element type	K_i	K_v	K_{My}	K_k	$K_{\theta y}$	$K_{\theta u}$
Slab	KANEPE (2013)	0.95		0.85	1.15	0.85
Column	EC8-Part 3 (2005)	0.90	1.00	—	1.05 ^a	1.00
					1.20 ^b	
	KANEPE (2013)	0.90 ^c	—	0.80 ^c	1.25 ^c	0.75 ^c
Beam	KANEPE (2013)	0.85	—	0.80	1.25	0.75
Foundation	KANEPE (2013)	0.90		0.70	1.30	0.80

^aRoughening^bRest of measures or no measures^cRoughening and connection measures

$K_k = 0.25 \sim 0.75$. The comparison of the code suggested monolithicity factors to the experimental values for monolithicity factors revealed that there is large dispersion in the case of monolithicity factors $K_{\theta y}$, $K_{\theta u}$, and K_k (Thermou et al. 2014). This observation has to be further assessed by considering the limited range of parameters of the experimental database (Thermou et al. 2011) as well as the fact that deformation and stiffness values are difficult to measure experimentally and “ultimate” conditions are not defined in a uniform way in all tests (Thermou et al. 2014). The suggested values for the monolithicity factors to be used in EC8-Part 3 (2005) and KANEPE (2013) apply for the specific properties of the construction materials and level of applied axial load. These limits are defined by the experimental data as: (i) percentage of the longitudinal reinforcement of the existing cross section 0.81 ~ 2.01 %; (ii) percentage of the longitudinal reinforcement of the jacket 0.75–1.64 %; (iii) concrete compressive strength of the existing cross section 23–56 MPa; (iv) concrete compressive strength of the jacket 18–69 MPa; and (v) dimensionless axial load (with the assumption that it is applied to the jacketed cross section) 0–0.23.

Analytical Assessment of R/C Jacketed Columns

Reinforced concrete (R/C) jacketing is one of the most commonly applied methods for the rehabilitation of concrete members. Jacketing is

considered to be a global intervention method if longitudinal reinforcement placed in the jacket passes through holes drilled in the slab and new concrete is placed in the beam-column joint (Fig. 7). The main advantage of the R/C jacketing technique is the uniformly distributed lateral load capacity throughout the structure, thereby avoiding concentrations of lateral load resistance, which occur when only a few shear walls are added. A disadvantage of the method is the presence of beams which may require most of the new longitudinal bars in the jacket to be bundled into the corners of the jacket. Because of the presence of the existing column, it is difficult to provide cross-ties for the new longitudinal bars, which are not at the corners of the jacket.

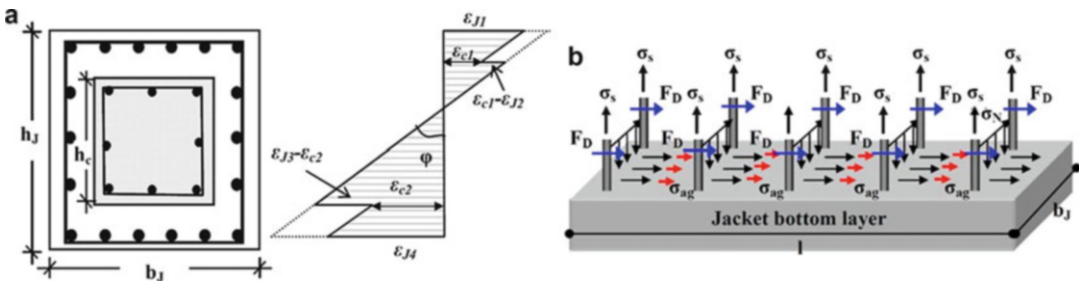
The analytical model presented in section “Analytical Model for R/C Jacketed Columns” explicitly accounts for the slip at the interface between the existing member (core) and the jacket. The transfer of normal and shear stresses in the interface between the external layers of the jacket and the existing member determines the composite action developed by the strengthened members (Thermou et al. 2007). An alternative to this approach is presented in section “Simplified Analytical Expressions for R/C Jacketed Columns” where simplified analytical expressions in combination with the monolithicity factors may be used for designing the retrofit scheme.

Analytical Model for R/C Jacketed Columns

The proposed analytical model for predicting the flexural response of existing R/C members strengthened with concrete jacketing under



Strengthened Structural Members and Structures: Analytical Assessment, Fig. 7 Column R/C jacketing
(Source of figures: author's personal files)



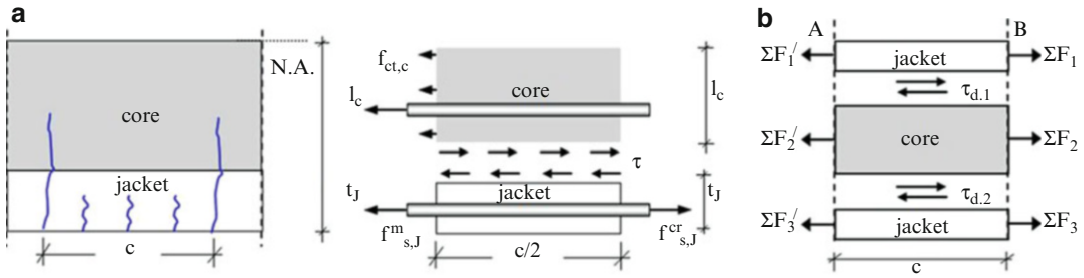
Strengthened Structural Members and Structures: Analytical Assessment, Fig. 8 (a) Strain profile of the jacketed cross section; (b) shear transfer mechanisms at the interface between the jacket and the core

monotonic and cyclic loading conditions introduces a degree of freedom allowing the relative slip at the interface between the existing member and the jacket (Thermou et al. 2007). Slip along the member's length is attributed to the difference in normal strains at the contact interfaces (Fig. 8a). For flexural analysis, the cross section is divided into three layers which bend with the same curvature, φ (Fig. 8a). The two external layers represent the contribution of the jacket, whereas the internal one represents both the core (existing cross section) and the web of the jacket shell. Slip at the interface mobilizes the shear transfer mechanisms such as aggregate interlock, friction due to clamping action, and dowel action provided by the stirrup legs of the jacket and by the dowels placed at the interface between the core and the jacket in case that such a connection measure is taken (Fig. 8b).

According to the analytical model of Thermou et al. (2007) for R/C jacketed members, shear transfer at the interface between the existing member and the jacket takes place between half crack intervals along the length of the jacketed member, as commonly considered in bond analysis. At the initial stages of loading, cracks form only at the external layers (jacket) increasing in number with increasing load, up to crack stabilization. This occurs when the jacket steel stress at the crack, $\sigma_{s,cr}$, exceeds the limit (fib 2010)

$$\sigma_{s,cr} > f_{ctm,J} \frac{1 + \eta \rho_{s,eff}}{\rho_{s,eff}} \quad (11)$$

where $f_{ctm,J}$ is the tensile strength of concrete, $\eta(=E_s/E_c)$ is the modular ratio, and $\rho_{s,eff}$ is the effective reinforcement ratio defined as the total steel area divided by the area of mobilized



Strengthened Structural Members and Structures: Analytical Assessment, Fig. 9 (a) Free body equilibrium in the tension zone of the core of the composite section; (b) section equilibrium between adjacent cracks; (c) crack spacing

concrete in tension, usually taken as a circular domain with a radius of $2.5D_b$ around the bar (Model Code 2010). Using the same considerations in the combined section it may be shown that a number of the external cracks penetrate the second layer (core) of the jacketed member (Fig. 8a). The distance between those cracks, taken as c , is a key element of the proposed methodology (Fig. 9a).

After crack stabilization and assuming that the neutral axis depth is about constant in adjacent cross sections, from the free body equilibrium in the tension zone of the core of the composite section (Fig. 9a), the crack spacing is defined as follows (Thermou et al. 2007):

$$c = \frac{0.64 \cdot b_J l_c f_{ct,c}}{n_c D_{b,c} f_{b,c} + n_J D_{b,J} f_{b,J}} \quad (12)$$

where b_J is the width of the jacketed cross section, l_c is the height of the tension zone in the core of the composite cross section, $f_{ct,c}$ is the tensile strength of concrete core, n_c, n_J are the number of bars in the tension steel layer of the core and the jacket, respectively, $D_{b,c}, D_{b,J}$ are the bar diameter of the core and jacket longitudinal reinforcement, respectively, and $f_{b,c}$ and $f_{b,J}$ are the average bond stress of the core and the jacket reinforcement layer, respectively.

Shear stress demand at the interfaces, $\tau_{d,i}$, is determined by examining the cross section along the height and along a member length equal to the distance between successive cracks (Fig. 9b). The layer force resultant ΣF_i (sum of concrete and steel forces at each layer), for the externally

applied axial load, N_{ext} (considered to be applied to the jacketed section), is used to calculate the vertical shear stress demand in the member, $\tau_{d,i}$ (Fig. 9b). With the assumption that the shear flow, q , reversal takes place at length equal to $c/2$ (where c is the crack spacing), the average stress demand $\tau_{d,i}$ is equal to

$$\tau_{d,i} = \frac{\Sigma F_i}{0.5 c b_J} \quad (13)$$

where ΣF_i is the layer force resultant, b_J is the width of the jacketed cross section, and c is the crack spacing length.

Shear Transfer Mechanisms Under Monotonic and Cyclic Loading

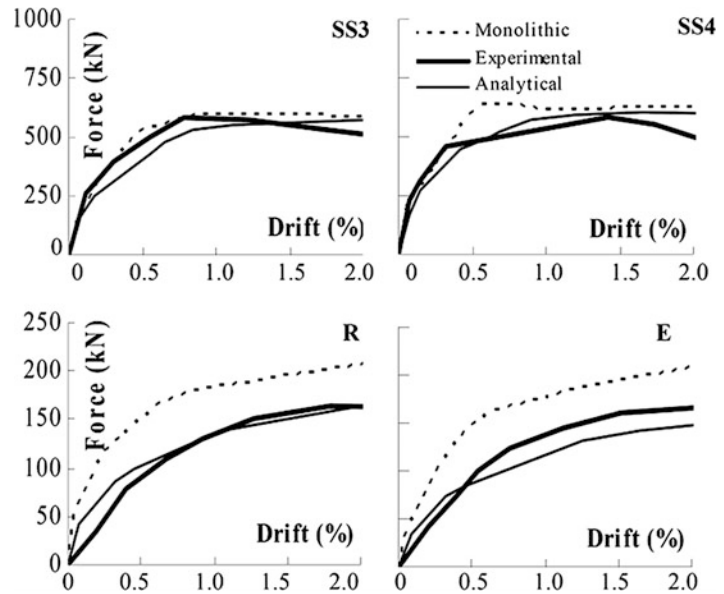
The monotonic and cyclic behavior of interfaces is described by the constitutive model developed by Vassilopoulou and Tassios (2003) where the models of friction and dowel resistance of Tassios and Vintzileou (1987), Vintzileou and Tassios (1986, 1987) are adopted. The interface model accounts for the combined shear force resistance mobilized along interfaces due to sliding both under monotonic and cyclic imposed deformations. This model was further modified by Palieraki et al. (2012) and adopted by the current Greek Code for Interventions (2013) (see Fig. 6). In case of cyclic loading, additional modifications and extensions were applied (Thermou et al. 2012b).

Solution Algorithm

The objective of the calculation algorithm at each loading step was twofold: simultaneous

Strengthened Structural Members and Structures: Analytical Assessment,

Fig. 10 Comparison of moment–curvature response histories with the corresponding experimental ones (SS1 and SS3 specimens from Rodriguez and Park (1994), R and E from Vadoros and Dritsos (2008))



establishment of equilibrium between the shear stress capacity and demand at the interfaces for relative slip, s , and force equilibrium at the cross section. An iterative procedure was followed, and equilibrium is established until convergence is achieved. Due to the complexity of the proposed solution algorithm, a program was necessary to be developed where fiber analysis was considered. The analytical model was set up to predict moment–curvature response curves. More details can be found in Thermou et al. (2012b).

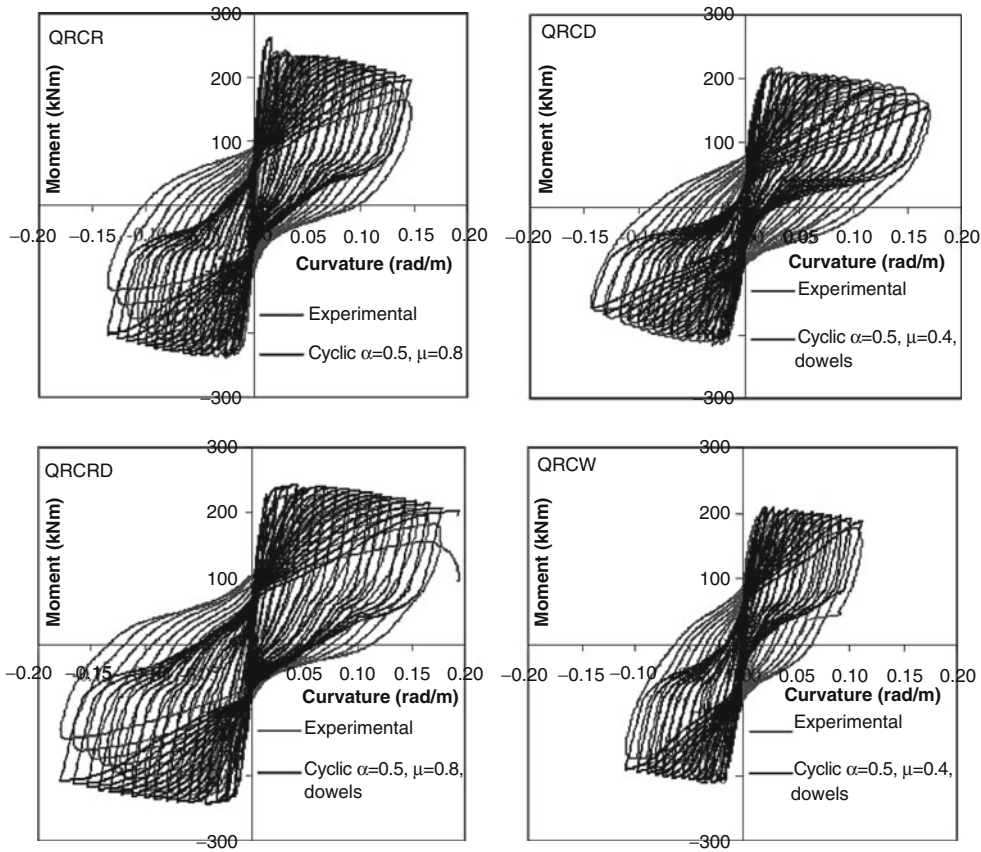
Comparison with Experimental Data

The validity of the proposed analytical model for predicting the flexural response of R/C jacketed members was examined by comparing: (i) for monotonic loading, the analytical lateral load versus lateral displacement curves along with the envelope experimental curves of the recorded lateral load versus lateral displacement hysteretic loops from various experimental studies (Thermou et al. 2007) (the monotonic moment–curvature response curves were derived according to the analytical model and then converted to force–displacement response curves). In Fig. 10, the experimental lateral load versus drift curves of a representative number of test specimens are compared to the corresponding analytical curves; (ii) For cyclic loading, the

moment–curvature histories derived by the analytical model with those for the specimens studied by Bousias et al. (2007). The decision to select this experimental study was based not only on its scope, but also on the fact that it provides detailed test results in terms of moment–curvature curves. The comparison between the experimental and the analytical moment–curvature histories for some of the specimens of Bousias et al. (2007) is presented in Fig. 11. It is observed that the analytical curves reasonably match the strength and stiffness level of the experimental curves at each loading step. When slip is taken into account, pinching is more pronounced in the analytical model, indicating less energy dissipation at each loading cycle, which is an indication of conservatism.

Simplified Analytical Expressions for R/C Jacketed Columns

The composite cross section is considered as monolithic, assuming that there is full connection in the interface between old and new concrete. The monolithic factors may be utilized to modify the response indices in order to indirectly account for the adverse effect of slip on the interface (see section “Simplified Design Approach: The Use of Monolithicity Factors”). A simplified model of a typical jacketed cross section of a



Strengthened Structural Members and Structures: Analytical Assessment, Fig. 11 Comparison of moment–curvature response histories with the

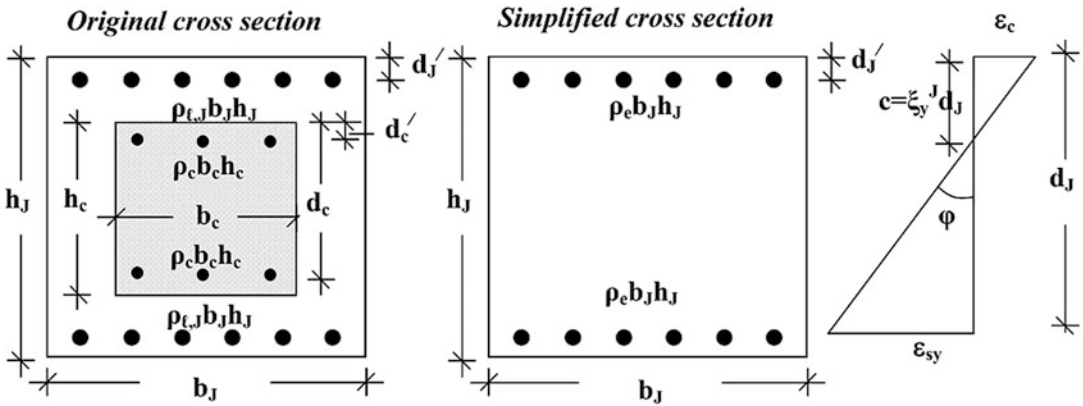
corresponding experimental ones for specimens QR/CR, QR/CD, QR/CRD, and QR/CW from Bousias et al. (2007)

column member is depicted in Fig. 12 (Thermou et al. 2012c). Dimensions of the initial section are b_c and h_c , whereas those of the section after jacketing are b_J , h_J . Compression zone depth c is expressed as a fraction of the depth of the jacketed section, $c = \xi_y^J \cdot d_J$. For calculations of yield moment and flexural stiffness, all reinforcement is considered to act at the location of the added (jacket) reinforcement. Note that according to Steiner's theorem, existing tension longitudinal reinforcement (equal to compression reinforcement, Fig. 12), given by the area ratio ρ_c , contributes to flexural stiffness of the jacketed section through the term $\rho_c b_c d_c (0.5h_c - d_c)^2$; thus, in order to maintain the same contribution in this calculation, the equivalent amount that is transferred to the location of jacket reinforcement

is $\rho_c b_c d_c (0.5h_c - d_c)^2 / (0.5h_J - d_J)^2$. Furthermore, any other existing web longitudinal reinforcement is neglected, as it is considered to have a small influence on post-jacketed flexural strength. Thus, the equivalent longitudinal tensile reinforcement ratio ρ_e , of the jacketed cross section (total reinforcement area divided by the total area of the jacketed member) is given by

$$\rho_e = \rho_J + \rho_c \frac{(0.5h_c - d_c)^2}{(0.5h_J - d_J)^2} \cdot \frac{b_c h_c}{b_J h_J} \quad (14)$$

where $\rho_c (=A_c/b_c h_c)$ and $\rho_J (=A_J/b_J h_J)$ are the tension longitudinal reinforcement ratios of the existing cross section (equal to compression reinforcement) and the jacket, respectively, h_c and h_J are the heights of the existing and jacketed cross



Strengthened Structural Members and Structures: Analytical Assessment, Fig. 12 Simplified model for R/C jacketed members (existing web longitudinal

reinforcement is neglected as it is considered to have a small influence on post-jacketed flexural strength)

sections, respectively, b_c and b_j are the widths of the existing and jacketed cross sections, respectively, and d_c and d_j are the depths of the existing and jacketed cross sections, respectively.

The j th member translational stiffness (secant to yield) is defined as

$$K_{y,j}^J = \frac{12E_c I_{y,j}^J}{h_{st}^3} = \frac{12M_{y,j}^J}{\varphi_{y,j}^J h_{st}^3} \quad (15)$$

The moment at yield, $M_{y,j}^J$, at the center of gravity of the simplified jacketed cross section is estimated equal to

$$M_{y,j}^J = \varphi_{y,j}^J b_j h_j^3 E_c [0.40 \rho_e (\xi_y (-1 - 0.25 \eta_c) + 1.15 \eta_c + 0.10) + 0.25 \xi_y^2 (1 - 0.66 \xi_y)] \quad (16)$$

Hence, the j th member translational stiffness is equal to

$$K_{y,j}^J = \frac{b_{j,j} h_{j,j}^3 E_c}{h_{st}^3} \left\{ 4.8 \rho_{e,j} [1.15 \cdot \eta_c - \xi_{y,j}^J] \cdot (1 + 0.25 \eta_c) + 0.1 \right\} + 3 \left(\xi_{y,j}^J \right)^2 (1 - 0.66 \xi_{y,j}^J) \quad (17)$$

where ρ_e is the equivalent longitudinal tensile reinforcement ratio, $\eta_c (= E_s/E_c)$ is the modular

ratio of steel and concrete, and $\xi_{y,j}^J$ is the normalized depth of compression zone of the R/C jacketed member.

Depending on the magnitude of the axial load ratio, $\nu_j = N_j/b_j h_j f_c'$, yielding may occur when either the tension steel reinforcement reaches yielding or the compressive concrete strain exceeds the limit of linear response in the compressive stress-strain envelope, estimated in the range of $\varepsilon_c = 1.8 \cdot f_c'/E_c$. From the basic cross-sectional equilibrium, the normalized depth of compression zone, $\xi_{y,j}^J$, associated with these two alternative definitions of phenomenological yielding is obtained from:

(a) Upon yielding of tension steel

$$\xi_{y,j}^J = - \left[(2\eta_c - 1) \rho_{e,j} + \frac{\nu_j f_c'}{E_c \varepsilon_{sy}} \right] + \left\{ \left[(2\eta_c - 1) \rho_{e,j} + \frac{\nu_j f_c'}{E_c \varepsilon_{sy}} \right]^2 + 2 \left[(1.10 \eta_c - 0.10) \rho_{e,j} + \frac{\nu_j f_c'}{E_c \varepsilon_{sy}} \right] \right\}^{0.5} \quad (18a)$$

where f_c' is the concrete compressive strength and ε_{sy} is the steel strain at yielding.

(b) At the onset of concrete strain nonlinearity, $\varepsilon_c = 1.8 \cdot f_c'/E_c$

$$\begin{aligned} \zeta_{y,j}^J = & - \left[(2\eta_c - 1)\rho_{e,j} + 0.55v_j \right] \\ & + \left\{ \left[(2\eta_c - 1)\rho_{e,j} + 0.55v_j \right]^2 + 2(1.10\eta_c)\rho_{e,j} \right\}^{0.5} \end{aligned} \quad (18b)$$

From the above, the secant to yield stiffness and the moment at yield may be calculated directly by utilizing Eqs. 15 and 16, respectively, after being multiplied by the adequate monolithicity factors according to Table 2.

References

- Birkeland PW, Birkeland HW (1966) Connections in precast concrete construction. *J ACI* 63(3):345–368
- Bousias S, Biskinis D, Fardis MN, Spathis A-L (2007) Strength, stiffness, and cyclic deformation capacity of the concrete jacketed members. *ACI Struct J* 104(5):521–531
- Chalioris KE, Thermou GE, Pantazopoulou SJ (2014) Behaviour of rehabilitated R/C beams with self-compacting concrete jacketing – analytical model and test results. *Constr Build Mat J* 55(31):257–273
- Choi D-U, Fowler DW, Jirsa JO (1999) Interface shear strength of concrete at early ages. *ACI Struct J* 96(3):343–348
- Clough RW, Penzien J (1993) *Dynamics of Structures*, 2nd edn. McGraw-Hill International Editions, New York, p 738
- Deutsches Institut für Normung [German Standards Institute] (DIN) (1936) *Beton und Stahlbetonbau: Bemessung und Ausführung*. DIN 1045, Berlin [Concrete and reinforced concrete: design and construction]
- Eurocode 8 (2005) *Design of structures for earthquake resistance, Part 3: assessment and retrofitting of buildings*. EN1998-3, European Committee for Standardization (CEN), Brussels
- fib (2013) *Model Code for Concrete Structures 2010*. Ernst & Sohn, Hoboken, New Jersey
- Greek Code for Interventions (KANEPE) (2013) *Earthquake Planning and Protection Organization*, Athens
- Júlio ENBS, Branco FAB, Silva VD, Lourenco JF (2006) Influence of added concrete compressive strength on adhesion to an existing concrete substrate. *Build Environ* 41:1934–1939
- Kappos AJ, Thermou GE, Papanikolaou VK (2012) Investigation of the behaviour of old type detailing R/C columns strengthened with R/C jackets, Report. *Earthquake Planning and Protection Organization* 179 (in Greek)
- Lampropoulos AP, Tsiolou OT, Dritsos SE (2012) Monolithic coefficient values for design when seismically strengthening RC columns with jackets. *J Earthq Eng* 16:1023–1042
- Mattock HA, Hawkins MN (1972) Shear transfer in reinforced concrete- recent research. *PCI J* 17(2):55–75
- Palieraki V, Vintzileou E, Zeris C (2012). Behaviour of interfaces in repaired/strengthened R/C elements subjected to cyclic actions: experiments and modelling. In: *Third international symposium on life-cycle and sustainability of civil infrastructure systems (IALCCE'12)*, Vienna
- Rodriguez M, Park R (1994) Seismic load tests on reinforced concrete columns strengthened by jacketing. *ACI Struct J* 91(2):150–159
- Royal Decree (1954) Design regulation of reinforced concrete building works. Royal Decree (18.2/26.07.1954), Ministry of Public Works, Greece (in Greek)
- Santos PMD, Júlio ENBS (2013) A state-of-the-art review on roughness quantification methods for concrete surfaces. *Constr Build Mat J* 38:912–923
- Santos PMD, Júlio ENBS (2014) Interface shear transfer on composite concrete members. *ACI Struct J* 111(1):113–121
- Tassios TS (1986) Fundamental mechanisms of force-transfer across reinforced concrete critical interfaces. In: *CEB workshop, West Germany*, pp 381–397
- Tassios T, Vintzileou E (1987) Concrete-to-concrete friction. *ASCE J Struct Eng* 113(4):832–849
- Thermou GE, Elnashai AS (2006) Seismic retrofit schemes for R/C structures and local-global consequences. *J Progr Struct Eng MatWiley InterSci* 8(1):1–151
- Thermou GE, Palaioxorinou T (2013) Rapid evaluation of non-ductile multistory RC buildings. In: *Proceedings of ECCOMAS thematic conference “computational methods in structural dynamics and earthquake engineering (COMPdyn 2013)”*, Kos, paper no 1459
- Thermou GE, Pantazopoulou SJ, Elnashai AS (2007) Flexural behavior of brittle R/C members rehabilitated with concrete jacketing. *J Struct Eng* 133(10):1373–1384
- Thermou, GE, Papanikolaou VK, Kappos AJ (2011) Analytical model for predicting the response of old-type columns rehabilitated with concrete jacketing under reversed cyclic loading. In: *Proceedings of ECCOMAS thematic conference “computational methods in structural dynamics and earthquake engineering (COMPdyn 2011)”*, Corfu, paper no 137
- Thermou GE, Pantazopoulou SJ, Elnashai AS (2012a) Global interventions for seismic upgrading of substandard R/C buildings. *J Struct Eng* 138(3):387–401
- Thermou GE, Papanikolaou VK, Kappos AJ (2012b) Cyclic response of R/C jacketed columns including modelling of the interface behavior. In: *Proceedings of 15th world conference on earthquake engineering*, Lisbon, 24–28 Sep 2012, paper no 2855
- Thermou GE, Elnashai AS, Pantazopoulou SJ (2012c) Retrofit Yield Spectra—a practical device in seismic rehabilitation. *Earthq Struct* 3(2):141–168
- Thermou GE, Papanikolaou VK, Kappos AJ (2014) Monolithicity factors for the design of R/C columns strengthened with R/C jackets. In: *Proceedings of*

- 2nd European conference on earthquake engineering and seismology, Istanbul, paper no 766
- Tsioulou OT, Dritsos SE (2011) A theoretical model to predict interface slip due to bending. *Mat Struct* 44:825–843
- Vandoros KG, Dritsos SE (2008) Concrete jacket construction detail effectiveness when strengthening R/C columns. *Constr Build Mater* 22:264–276
- Vassilopoulou I, Tassios P (2003) Shear transfer capacity along a R/C crack under cyclic sliding. In: *Proceedings fib symposium, Technical Chamber of Greece, Athens*, paper no 271
- Vecchio FJ, Collins MP (1986) The modified compression field theory for reinforced concrete elements subjected to shear. *J ACI* 83(2):219–231
- Vintzileou E, Tassios TS (1986) Mathematical models for dowel action under monotonic and cyclic conditions. *Mag Concr Res* 38(134):13–22
- Vintzileou E, Tassios TS (1987) Behavior of dowels under cyclic deformations. *ACI Struct J* 84(1):18–30
- Walraven JC (1981) Fundamental analysis of aggregate interlock. *Struct Div ASCE* 107(ST11):2245–2270

Strengthening Techniques: Bridges

Chris P. Pantelides

Department of Civil and Environmental
Engineering, University of Utah, Salt Lake City,
UT, USA

Synonyms

Bridge; Damage; Design; Earthquakes; Rehabilitation; Repair; Retrofitting; Seismic; Strengthening

Introduction

The goal of seismic design is to prevent life-threatening damage and to allow moderate damage that can be repaired after an earthquake. From a life safety perspective, the most serious bridge damage occurs when the deck collapses, because of inadequate seat details, column failure, or cap beam-column joint failure. The collapse of the link span of the San Francisco-Oakland Bay Bridge during the 1989 Loma Prieta earthquake (Fig. 1a) was caused by failure of two bolted

connections under the span's supporting truss which led to its unseating (Buckle et al. 1990). The bridge failure of the Hanshin elevated expressway during the 1995 Great Hanshin Kobe earthquake (Fig. 1b) is attributed to column failure caused mainly by poor anchorage of the circular hoop ends which were lap spliced in the cover concrete (Park 1996) and inadequate amount and details of the longitudinal and transverse reinforcement (Anderson et al. 1996). This entry presents a summary of observed damage in past earthquakes and seismic strengthening techniques of existing bridges. The stages of evolution of bridge design in the USA from the 1970s including vulnerabilities and typical design details are outlined.

Seismic Bridge Design

The seismic design of bridges is separated into three periods, based on the State of California provisions (Caltrans 2006).

Pre-1971 Design

In this period, it was recognized that earthquakes generate forces proportional to the structure's dead weight. Until 1965, the maximum lateral seismic design force was 6 % of the structural dead weight. Vulnerabilities of bridges built during this period include: (a) column shear failure, (b) column longitudinal reinforcement pullout, and (c) unseating of expansion hinges. Typical design details of the period are 13 mm diameter column ties at 300 mm spacing (regardless of column or bar size), very short seat widths at expansion joints (150–200 mm), inadequate lap splices of column bars from the footing (20 bar diameters), inadequate development of column bars into the footing (20 bar diameters without standard hooks), and lap-spliced column ties in the cover concrete (absence of seismic hooks).

1971–1994 Design

After the 1971 San Fernando earthquake, the importance of ductility and detailing was recognized, and the concept of capacity design was



Strengthening Techniques: Bridges, Fig. 1 Bridge failure: (a) San Francisco-Oakland Bay Bridge (© California Department of Transportation (2003)

(CALTRANS)); (b) Hanshin elevated expressway (<https://www.fhwa.dot.gov/publications/publicroads/96fall/imgs/p96au18.jpg> (FHWA))

introduced in the design code. Vulnerabilities of bridges built during this period include: (a) column shear failure of plastic hinge regions, (b) shear failure of flared columns, and (c) unseating of expansion joint hinges. In retrofitted bridges, vulnerabilities were found at expansion joint hinge restrainers, particularly for skewed bridges. Typical details for this period include: closer spacing and improved column shear detailing (spacing 100–150 mm, but no confinement/anti-buckling requirement for plastic hinges), top reinforcement in footings and pile caps (but no shear reinforcement), column longitudinal splices were prohibited at maximum moment locations, short seat widths at expansion joints (300 mm), no cap beam-column joint reinforcement, and poor column flare details.

Post-1994 Design

After the 1994 Northridge earthquake, a capacity design philosophy was adopted that would ensure ductile flexural failure of the columns while all other bridge components remained elastic. Typical design details include: tight confinement reinforcement in plastic hinge regions (100 mm spacing), long seat widths at expansion joints (600 mm), improved flare column details (gap between top of flare and superstructure), no lap splices in plastic hinge zones, shear reinforcement in footings, and cap/column and footing/column joint reinforcement.

Damage to Bridges in Past Earthquakes

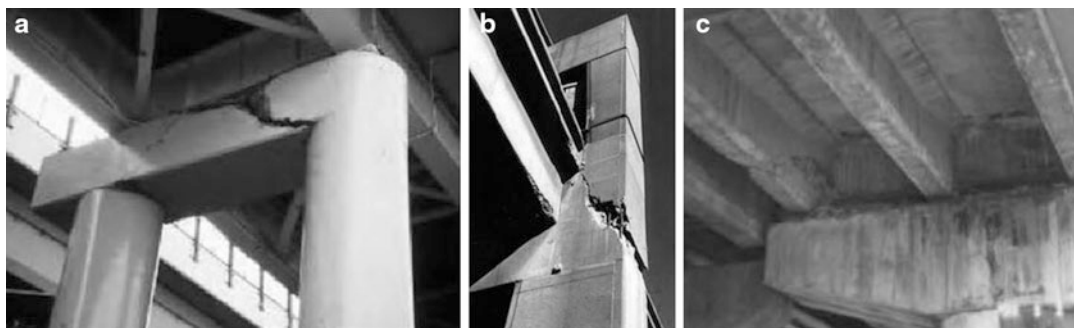
After an earthquake, lessons are learned which influence seismic strengthening of existing bridges as well as the development of new design codes. This section describes damage and its location as observed in bridges from past earthquakes (Caltrans 2006). Deck damage includes complete collapse of a span, as shown in Figs. 1a and 2a, b, or movement of the deck, as shown in Fig. 2c. Cap beam, cap beam-column joint, and girder shear damage are shown in Fig. 3. Damage to columns can be flexural as shown in Fig. 4 or shear damage as shown in Fig. 5; shear damage is brittle and frequently leads to collapse. Soil settlements can occur around columns, and gaps between columns and soil may become large as shown in Fig. 6. Abutment failures are also common, as shown in Fig. 7.

Bridge Seismic Strengthening

There is historic evidence which supports the assertion that seismic strengthening of bridges is effective. Damage in past earthquakes, such as the 1994 Northridge earthquake, has been reported from cable restrainers breaking through the hinge diaphragm; however, no bridge retrofitted after the 1989 Loma Prieta earthquake experienced any serious damage (Yashinsky 1998). Similarly, most of the observed structural



Strengthening Techniques: Bridges, Fig. 2 Deck failure: (a), (b) span collapse (© California Department of Transportation (2006)); (c) span displacement



Strengthening Techniques: Bridges, Fig. 3 Cap beam, cap beam-column joint, and girder shear damage (Cap beam, girder shear damage: © California Department of

Transportation (2006); beam-column joint: © California Department of Transportation (2003))

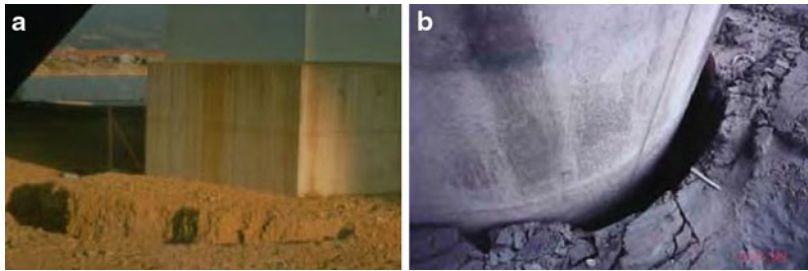


Strengthening Techniques: Bridges, Fig. 4 Column flexural damage (© California Department of Transportation (2006))

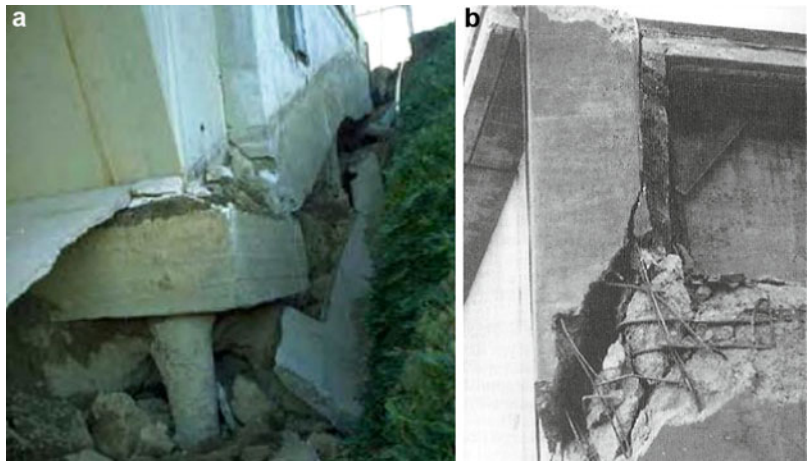


Strengthening Techniques: Bridges, Fig. 5 Column shear damage (© California Department of Transportation (2006))

Strengthening Techniques: Bridges, Fig. 6 Soil settlement and gap around columns (© California Department of Transportation (2006))



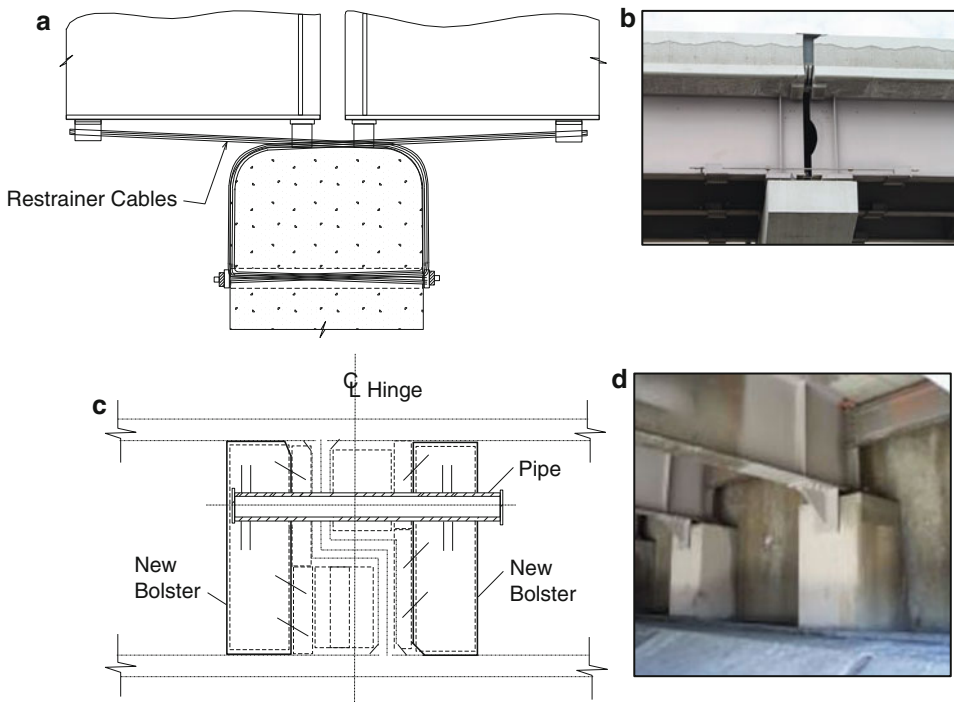
Strengthening Techniques: Bridges, Fig. 7 Abutment failures (© California Department of Transportation (2006))



damage in the 2011 Great East Japan earthquake has reportedly occurred in older bridges that had not yet been retrofitted after the 1995 Great Hanshin Kobe earthquake or only partially so (Marsh et al. 2011).

Several measures are used to strengthen bridges: restrainers, steel girder plate connections,

seat extenders, column strengthening, infill wall between columns, footing and abutment strengthening, and cap beam-column joint strengthening. In addition, retrofit measures such as seismic isolation reduce the demand imposed on a bridge system, and in this sense they improve its seismic performance significantly.



Strengthening Techniques: Bridges, Fig. 8 (a) Cable restrainer, (b) steel bar restrainer, (c) pipe seat extender, (d) bumper blocks

Restrainers/Expansion Joint Seat Width Extenders/Steel Plates/Bumper Blocks

Cable restrainers (Fig. 8a) and longitudinal steel bar restrainers (Fig. 8b) are used to prevent girders from falling off their supports (Yashinsky 1998). The typical cable restrainer is a galvanized 19 mm diameter steel cable with galvanized cold-swaged fittings and 25 mm diameter threaded studs at each end. Loss of support of the superstructure can be achieved using beam seat extenders, pipe seat extenders shown in Fig. 8c, steel plates used to connect two steel girders at expansion joints, and bumper blocks as shown in Fig. 8d (Pantelides et al. 2004).

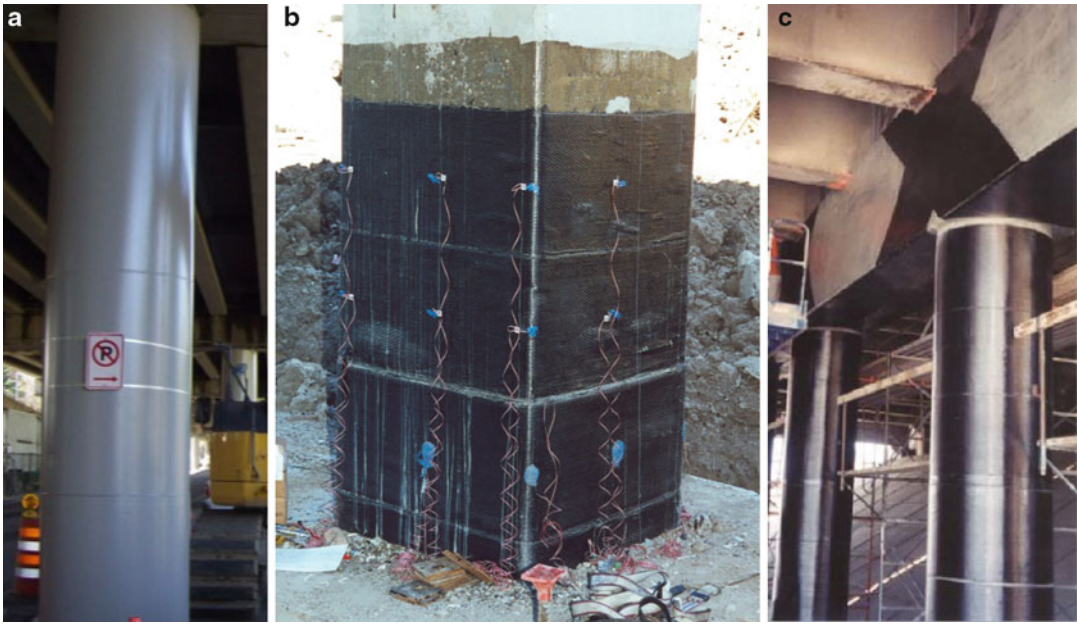
Failure of restrainer cables has been observed in the 1989 Loma Prieta, 1994 Northridge, and 1995 Kobe earthquakes (FHWA 2006). Restrainers ruptured, anchorage plates pulled through concrete diaphragms, fittings pulled away from cables, and anchorage nuts worked loose from the ends of cable units. Restrainers alone may not be the best solution in many cases

because of the need to limit restrainer forces to protect other structural elements, while increasing restrainer forces to accomplish their intended purpose. Alternatives that should be considered when loss of support is a possibility include seat extensions, bearing strengthening, and replacement with conventional or isolation bearings.

Column Jacketing: RC/Steel/Fiber Reinforced Polymer (FRP)

Strengthening of concrete columns can be achieved by reinforced concrete (RC) jackets. This is an intrusive and time-consuming procedure, but in some cases it is necessary. However, the effect of strengthening a column using RC jackets should be examined carefully as this method will increase column flexural strength and stiffness more than a steel jacket, FRP composite jacket, or wire wrap, with potentially undesirable effects on bridge performance.

Steel jacketing of circular columns by circular steel jackets (Fig. 9a), and of rectangular columns by elliptical steel jackets, was found to be



Strengthening Techniques: Bridges, Fig. 9 Column jackets: (a) steel (Photo credit: Washington State Department of Transportation. [http://sdotblog.seattle.gov/2012/](http://sdotblog.seattle.gov/2012/06/21/fauntleroy-expressway-wearing-new-jackets/#sthash.nlxozUnm.dpbs)

[06/21/fauntleroy-expressway-wearing-new-jackets/#sthash.nlxozUnm.dpbs](http://sdotblog.seattle.gov/2012/06/21/fauntleroy-expressway-wearing-new-jackets/#sthash.nlxozUnm.dpbs)), (b) FRP square, (c) FRP circular

effective in enhancing shear strength and flexural ductility (Priestley et al. 1994). Encasing circular columns with a bonded steel jacket was found to inhibit bond failures in lap splices of longitudinal reinforcement in plastic hinge regions (Chai et al. 1991). Rectangular steel jackets are effective in enhancing the performance of shear critical columns. These jackets can improve column ductility by eliminating the brittle shear mode of failure, but it is important to recognize that the failure mode may then shift to a flexural one for which the rectangular jacket can provide only limited assistance.

Seible et al. (1997) used continuous FRP jackets for flexural hinge confinement of circular and rectangular columns, shear strengthening, and lap splice clamping. Full-scale tests (Fig. 9b) have shown that significant increases can be gained in the ductility of bridge column bents strengthened with FRP composites (Pantelides et al. 1999, 2007). FRP composites have been used widely for seismic strengthening of existing bridges (Priestley et al. 1996; Ogata and Osada 2000), as shown in Fig. 9c for the State

Street Bridge on I-80 in Salt Lake City, Utah (Pantelides et al. 2004).

The advantage of using FRP composites stems from their superior resistance to corrosion, high stiffness-to-weight ratio, high strength-to-weight ratio, and ability to control the material's behavior by selecting the proper fiber orientation. FRP composites must be protected against ultraviolet radiation; even though carbon fibers themselves are not affected by moisture, the quality of the epoxy matrix is critical in the performance of the composite. Carbon fiber systems have shown loss of tensile strength at high temperatures due to moisture absorption. Another consideration when using FRP systems is their high initial cost; however, they have a superior life-cycle cost as compared to steel because of their resistance to corrosion. In shear and flexural critical applications, bond of FRP composites to concrete is a critical parameter which must be considered in the design (ACI 440 2002); mechanical anchorage of FRP laminates to concrete has been used successfully to postpone debonding.

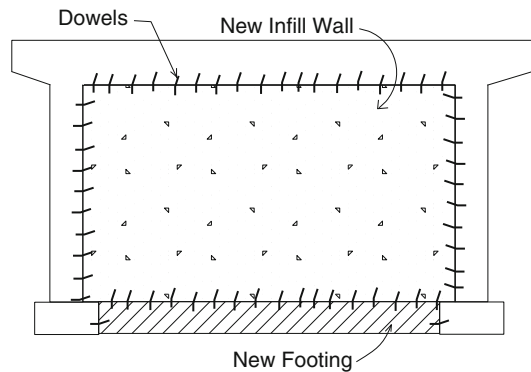
Another effective strengthening technique is to wrap prestressing wire under tension around a column (Hawkins et al. 1999). Rectangular steel columns have been strengthened with steel angles and circular steel columns with circular steel pipes (Nishikawa et al. 1998).

Infill Wall Strengthening

Infill shear walls are reinforced concrete walls cast in place between the columns of a multicolumn bent, as shown in Fig. 10. The reinforced concrete columns can have a circular or rectangular shape. They have been used successfully to increase transverse shear capacity. These walls prevent the formation of plastic hinges in the columns during transverse loading and help overcome deficiencies in the flexural or shear strength of the cap beam. To be effective, infill shear walls should be designed to act compositely with existing members. This is done by providing drilled and bonded dowels in the columns and the bottom of the cap beam, so that shear is transferred at the interfaces through a shear friction mechanism; this may require the existing concrete surfaces to be roughened. A footing should be provided under the infill wall which is tied into the existing column footings. There needs to be sufficient reinforcement to transfer all seismic forces, and the potential for differential settlement between the existing and new footings should be considered.

Foundation Strengthening

Footings that support columns may be structurally unable to resist the forces transmitted from those columns. This usually occurs when there is a lack of reinforcement in the top of the footing. A concrete overlay can be designed to provide increased negative moment capacity in the footing. Existing positive moment reinforcement is utilized along with the additional footing depth provided by the overlay to increase positive moment capacity. For isolated pile caps a grade or link beam is utilized to improve shear capacity, as shown in Fig. 11. In addition, longitudinal reinforcement in reinforced concrete piles may be terminated too close to the top of the pile, resulting in insufficient tensile capacity; in this



Strengthening Techniques: Bridges, Fig. 10 Pier strengthening using infill wall

case, it is desirable to add hold-down anchors or supplemental tension piles. An alternative is to drill holes into the piles from the top of the pile cap and epoxy high strength steel bars into the pile cap and piles (Pantelides et al. 2001); these vertical bars are shown in Fig. 11 (Pantelides et al. 2007).

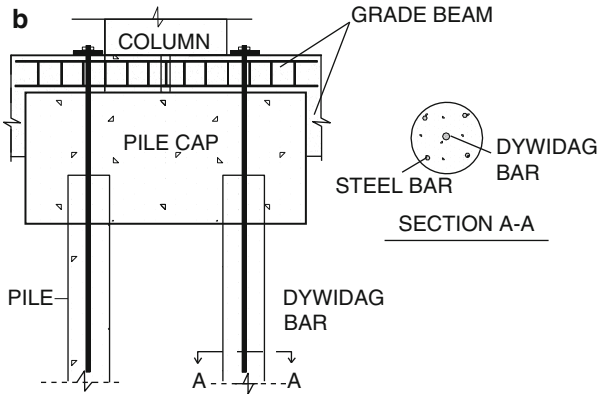
Abutment Strengthening

Strengthening of seat-type abutments in the longitudinal and transverse directions can reduce forces on the piers. Anchor slabs are used to increase the ability of abutments to carry both longitudinal and transverse loads, and resist displacement in either direction, as shown in Fig. 12. Transverse abutment shear keys and anchors can reduce forces on the piers (FHWA 2006).

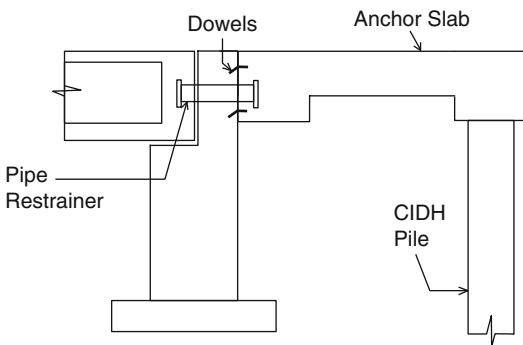
Cap Beam Joints: RC Bolsters/Posttensioning/Steel Jackets/FRP Jackets

Cap beam-column joints are subjected to large stresses and are vulnerable to damage in earthquakes transverse to the bridge. Existing cap beams and joints were not usually designed to behave in a ductile manner, so any strengthening measure must ensure that these elements are either capable of accommodating the ductility demands placed on them or are capable of elastically resisting the forces that will result from plastic hinging in the columns.

RC bolsters are used to improve flexural capacity, with or without prestressing. Another

a

Strengthening Techniques: Bridges, Fig. 11 Footing strengthening using grade beam and pile cap to pile steel bars



Strengthening Techniques: Bridges, Fig. 12 Abutment strengthening using anchor slab

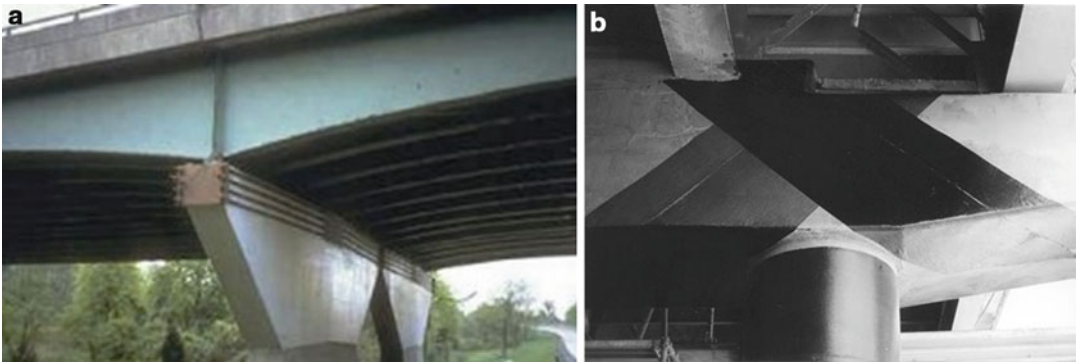
strengthening method is to use external posttensioning as shown in Fig. 13a. Steel jackets bonded to the sides of the joint and anchored to the bent cap are also effective (Thewalt and Stojadinovic 1995). Finally, FRP jackets in the form of an ankle wrap have been tested (Gergely et al. 2000) and used to strengthen column to cap beam-column joints as shown in Fig. 13b (Pantelides et al. 2004). The nominal principal tensile stress developed in the beam-column joint is used to design the carbon FRP composite layers. When strengthening of these joints is required to improve the longitudinal response, transversely prestressed bolsters may be used.

Seismic Isolation: Elastomeric Bearings/ Friction Pendulum Bearings

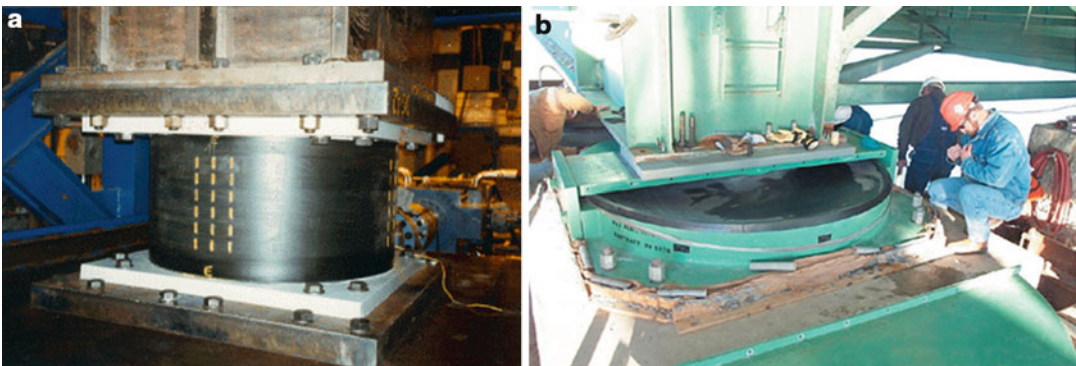
Seismic isolation bearings can shift the natural frequency of the bridge away from the region of dominant earthquake energy; in addition, they increase damping and reduce and redistribute the earthquake-induced lateral forces to levels within the elastic capacities of the substructure and foundation (Mayes et al. 1994). An elastomeric bearing (Fig. 14a) consists of alternating layers of rubber and steel plates and often has a lead core to dissipate seismic energy. A friction pendulum bearing (Fig. 14b) consists of a stainless steel concave dish and articulated slider surfaced with a composite liner. During an earthquake, the slider moves back and forth on the concave dish; the spherical surfaces of the slider and the dish define a motion similar to that of a pendulum. A friction pendulum bearing isolates a bridge from an earthquake through pendulum motion and absorbs energy through friction.

Summary

The entry presents a number of methods for seismic strengthening of existing bridges. A brief review of seismic bridge design evolution and



Strengthening Techniques: Bridges, Fig. 13 Cap beam-column joint strengthening: (a) posttensioning (Photo Credit: VSL International Ltd. <http://vsl-sg.com/sps-images/4-004.jpg>), (b) FRP jacket



Strengthening Techniques: Bridges, Fig. 14 Bridge bearings: (a) elastomeric (Photo Credit Federal Highway Administration. <https://www.fhwa.dot.gov/publications/publicroads/99marapr/images/seismic3.gif>), (b) friction pendulum (Photo credit: Federal Highway Administration)

an overview of damage to structural elements of bridges have been presented. Strengthening techniques covered include restrainers, expansion joint seat width extenders, column jackets, footings, abutments, and cap beam-column joints. Retrofit measures such as seismic isolation reduce the demand imposed on a bridge system thus improving the seismic performance. Hinges and abutments are the two areas where retrofitted bridges continue to experience damage. The 1994 Northridge and 2011 Great East Japan earthquakes have shown that retrofitted bridges did not experience any serious damage. This clearly demonstrates that seismic strengthening of bridges is valuable and should be continued. A unique feature of earthquakes is that they often expose the weaknesses of a structural system. Despite the

progress made in designing new and strengthening existing bridges, it is likely that future earthquakes will expose additional vulnerabilities of existing and retrofitted bridges.

Cross-References

- [Bridge Foundations](#)
- [Resilience to Earthquake Disasters](#)
- [Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions](#)
- [Seismic Resilience](#)
- [Seismic Strengthening Strategies for Existing \(Code-Deficient\) Ordinary Structures](#)
- [Strengthened Structural Members and Structures: Analytical Assessment](#)

References

- American Concrete Institute, ACI Committee 440 (2002) Guide for the design and construction of externally bonded FRP systems for strengthening concrete structures. ACI 440.2R-02, Farmington Hills
- Anderson DL, Mitchell D, Tinawi RG (1996) Performance of concrete bridges during the Hyogo-Ken Nanbu (Kobe) earthquake on January 1995. *Canadian J Civil Eng* 23:714–726
- Buckle IG, Goodson M, Cassano R, Douglas B, Hegemier G, Imbsen R, Jones D, Liu D, Mayes R, North P, Priestley N, Rojahn C, Seible F, Selna L, Viest I (1990) Bridge structures. *Earthq Spectra* 6(S1):151–187
- California Department of Transportation (2003) The race to seismic safety: protecting California's transportation system. Department of Transportation Advisory Board Report, Sacramento, 186 pp
- California Department of Transportation (2006) Visual catalog of reinforced concrete bridge damage. Structure Maintenance and Investigations Division, Sacramento, 350 pp
- Chai YH, Priestley MJN, Seible F (1991) Seismic retrofit of circular bridge columns for enhanced flexural performance. *ACI Struct J* 88(5):572–584
- Federal Highway Administration (2006) Seismic retrofitting manual for highway structures: part 1-bridges. U.S. Department of Transportation, McLean, 656 pp
- Gergely I, Pantelides CP, Reaveley LD (2000) Shear strengthening of R/C T-joints using CFRP composites. *J Compos Constr* 4(2):56–64
- Hawkins NM, Kaspar II, Karshenas M (1999) Seismic retrofit of Poplar Street interchange in East St. Louis. Structures Congress. ASCE, Reston
- Marsh ML, Buckle IG, Yen WP, Dasti S, Frost D, Monzon E (2011) Bridge performance during the Great East Japan Earthquake of March 11, 2011. http://www.pwri.go.jp/eng/ujnr/tc/g/pdf/27/27-1-3_Marsh.pdf
- Mayes RL, Choudhury D, Crooks RS, Jones DM, Knight RP (1994) Seismic isolation retrofit of existing bridges. In: Proceedings of the materials engineering conference. ASCE, Reston, pp 1179–1186
- Nishikawa K, Yamamoto S, Natori T, Terao K, Yasunami H, Terada M (1998) Retrofitting for seismic upgrading of steel bridge columns. *Eng Struct* 20(4–6): 540–551
- Ogata T, Osada K (2000) Seismic retrofitting of expressway bridges in Japan. *Cem Concr Compos* 22:17–27
- Pantelides CP, Gergely I, Reaveley LD, Volny VA (1999) Retrofit of R/C bridge pier with CFRP advanced composites. *J Struct Eng* 125(10):1094–1099
- Pantelides CP, Gergely J, Reaveley LD (2001) In-situ verification of rehabilitation and repair of reinforced concrete bridge bents under simulated seismic loads. *Earthq Spectra* 17(3):507–530
- Pantelides CP, Alameddine F, Sardo T, Imbsen R (2004) Seismic retrofit of State Street bridge on Interstate 80. *J Br Eng* 9(4):333–342
- Pantelides CP, Duffin JB, Reaveley LD (2007) Seismic strengthening of reinforced-concrete multicolumn bridge piers. *Earthq Spectra* 23(3):635–664
- Park R (1996) An analysis of the failure of the columns of a 600 m length of the Hanshin elevated expressway during the Great Hanshin Earthquake of 17 January 1995. *Bull N Z Natl Soc Earthq Eng* 29(2):73–82
- Priestley MJN, Seible F, Xiao Y, Verma R (1994) Steel jacket retrofitting of reinforced concrete bridge columns for enhanced shear strength-part 2: test results and comparison with theory. *ACI Struct J* 91(5):537–551
- Priestley MJN, Seible F, Calvi GM (1996) Seismic design and retrofit of bridges. Wiley, New York, 704 pp
- Seible F, Priestley MJN, Hegemier GH, Innamorato D (1997) Seismic retrofit of RC columns with continuous carbon fiber jackets. *J Compos Constr* 1(2):52–62
- Thewalt CR, Stojadinovic B (1995) Behavior of bridge outrigger knee joint systems. *Earthq Spectra* 11(3):477–509
- Yashinsky M (1998) Performance of bridge seismic retrofits during Northridge Earthquake. *J Br Eng* 3(1):1–14

Strengthening Techniques: Code-Deficient R/C Buildings

Andreas P. Lampropoulos
School of Environment and Technology,
University of Brighton, Brighton, UK

Synonyms

Epoxy grouting; Interfaces; Monolithicity coefficients; RC jacketing; Repair; Shrinkage; Strengthening

Introduction

Nowadays, many buildings need to be upgraded, especially in earthquake-prone areas, because of increased strength demands introduced by modern design codes or damage due to strong earthquakes. As a result, many techniques have been developed to improve the performance of the existing buildings. There is a wide range of applications of fiber-reinforced polymers (FRPs) in the form of rebars, plates, and sheets for the

strengthening of the existing buildings. These materials can be used for flexural and/or shear strengthening of existing structures, and a comprehensive description of this technique can be found in Essay 382,109 “Retrofitting and Strengthening of Contemporary Structures: Materials Used.”

In this entry, special emphasis is given on the technique of reinforced concrete (RC) jacketing. RC jacketing is a widespread technique used for the strengthening of existing columns, beams, and walls. The use of epoxy resins for the crack repair of damaged structural elements is also presented.

In earthquake-prone areas, the majority of the old buildings could be characterized as “code deficient,” and these buildings need to be upgraded.

“Code-deficient” buildings are the buildings that do not satisfy the requirements proposed by the new codes (Eurocode 8 2005; Code of Structural Interventions 2012; fib 2010). There are different factors responsible for structural deficiencies in existing buildings. The structural system of many old buildings was designed with configuration problems. Lack of regularity in geometry, in strength or stiffness, in plan, or in elevation are common characteristics of several existing buildings. Also, in the past, approximations and simplifications were adopted in carrying out structural analysis. Computers were not available, 3D analysis was not feasible, even 2D analysis was rarely used, and (continuous) beams and columns were considered as independent elements. Some critical requirements concerning the behavior of structures under earthquakes, such as ductility, capacity design, and detailing provisions (minimum amount of stirrups, lower limits for compressive strength, upper limit for tensile reinforcement), were also ignored. Also, in the past, the seismic actions used for the design of structures were much lower than the actions currently used for the design of new structures (Dritsos 2012).

As a consequence, the majority of existing buildings are deficient and in need of strengthening. Also, some of the old buildings are already damaged after previous strong earthquakes. The type of intervention that will be used for

strengthening deficient buildings depends on the level of damage and the requirements. In case of lightly damaged concrete elements, where there are no requirements for further upgrading of the structure, the technique of epoxy grouting is widely used. Epoxy grouting can be used to improve the performance and the durability of cracked reinforced concrete elements.

In case of heavily damaged structural elements or structures that need to be upgraded to conform with modern codes, seismic strengthening is required. There are a number of available techniques for the earthquake strengthening of existing buildings, such as jacketing of load-bearing elements, addition of infill walls or bracing systems, and installation of energy dissipation systems. For the jacketing of structural elements, steel, fiber-reinforced polymers (FRPs), and reinforced concrete can be used. The decision for the selection of the most suitable technique is not always easy and depends on a number of parameters such as cost, required time, and disruption of the occupants. Detailed descriptions of a number of different repair and strengthening techniques have been published by Tassios (2009) and Dritsos (2005).

In this entry, the technique of repair of damaged buildings with epoxy grouting and the strengthening with RC jackets are described in detail, and practical recommendations are presented. Special focus is given on the modeling and design of the jacketed elements. Critical parameters for the effectiveness of this technique are the bond between the old and new concrete elements and shrinkage of the new concrete. To take into account these two parameters, finite element method can be used. However, for the practical design of strengthened elements the use of monolithicity coefficients is proposed, which is a simplified procedure. Monolithicity coefficients are used to correlate the strength, stiffness, and deformation at yield and failure of strengthened columns to those of respective monolithic columns, the behavior of which can be determined by following conventional concrete design procedures. Monolithicity coefficients for jackets with different characteristics are presented at the end of this entry.

Strengthening Techniques:

Code-Deficient R/C Buildings, Fig. 1

(a) Preparation of cracks before the use of epoxy resin and (b) repaired columns. <http://www.episkeves.civil.upatras.gr>



Repairing Existing RC Structures Using Epoxy Resins and Bonded Mortar

Epoxy grouting is one of the most commonly used techniques for the repair of cracked structural elements. For the successful application of this technique, the surface of the damaged or cracked section needs to be properly prepared (Fig. 1a). Depending on the type of damage, different epoxy resin types and installation procedures can be used (Fig. 1b).

In this section, the technique of epoxy resin injection for crack sealing and the use of epoxy-bonded concrete replacement for the repair of damaged concrete are described.

Epoxy Resin Injection for Crack Sealing

The resin characteristics and the application methods used for the epoxy resin injection depend on the environmental and concrete temperature, the crack width, the thickness of the concrete member, and the extent of cracking.

Substrate Preparation

The first critical step for the successful application of this method is the preparation of the substrate. Any existing contaminants should be removed to

improve substrate conditions. All the existing coatings along the crack, at intervals of not less than the thickness of the concrete member or 10 cm (ETEP 2012), should be removed.

Port Installation

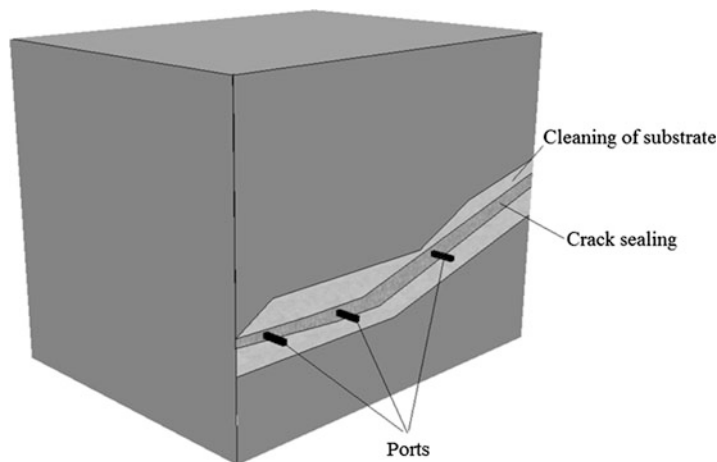
Entry ports (tube-like devices that provide for the successful transfer of the epoxy resin under pressure into the crack) together with “cap” crack sealing should be installed on the face of the crack to ensure containment of the epoxy as it will be injected under pressure into the crack (ACI 2003; Fig. 2). In case of crack penetration throughout a section, the use of cap seals on both sides of the crack elements is suggested (ACI 2003).

The distance of the ports depends on the opening at the face of the crack. The recommended port distance values proposed by ETEP (2012) are presented in Table 1.

Before the cap seal installation, the location of the widest portion of the crack should be marked. The material used for the seal cap should be carefully prepared with accurate batching of components and with a consistent application of the material over the crack. The use of 25 mm wide and 5 mm thick cap seals is proposed by ACI (2003).

Strengthening Techniques: Code-Deficient R/C Buildings,

Fig. 2 Installation of ports for crack sealing with epoxy resins



Strengthening Techniques: Code-Deficient R/C Buildings, Table 1 Recommended port spacing for different crack widths (ETEP 2012)

Crack width (mm)	Port spacing (mm)
0.3–0.5	100
0.5–1	100–135
1–2	135–170
2–3	170–200
3–10	200

Epoxy Resin Injection

Epoxy resins cure to form solids with high strength and relatively high modulus of elasticity. Epoxy resin used for crack injection should be a 100 % solid resin that meets the requirements of specification ASTM C-881 (2010) for type I or IV, grade 1 (low viscosity), class B or C (depending on the ambient temperature). If the purpose of injection is to restore concrete to its original load-bearing capacity, a type IV epoxy should be used. If full restoration of load-bearing capacity is not envisaged, a type I epoxy is sufficient. No solvents or nonreactive diluents are permitted in the resin (USBR 1997). Class B epoxy is used between 4 °C and 16 °C, while class C is used above 16 °C (ASTM C-881 2010).

The injection should start from the bottom in the case of vertical cracks, while for horizontal cracks the starting point should be the widest crack section. The injection should continue until refusal or bleeding of the adjacent port (ACI 2003; ETEP 2012).

Strengthening Techniques: Code-Deficient R/C Buildings, Table 2 Recommended resin viscosity and injection pressure values for different crack widths (ETEP 2012)

Crack width (mm)	Dynamic viscosity of the resin (cps)	Required injection pressure (MPa)
0.3–0.5	1,000	—
	500	0.8
	250	0.4
	130	0.2
0.5–1	1,000	0.8
	500	0.4
	250	0.2
1–2	1,000	0.4
	500	0.2
	250	0.1
2–3	1,000	0.3
	500	0.1
	250	0.005

The injection pressure and the dynamic viscosity of the resin depend on the crack width. In Table 2, the recommended values proposed by ETEP (2012) are presented.

Upon completion of the resin injection, the ports and the cap seals will be removed by heat chipping or grinding (ACI 2003).

Quality Control

The first check should be done by hand 48 h after the injection, to check if the coagulation

(polymerization) of the epoxy resin has been completed.

Then, an additional check is required to ensure that the crack is filled with the resin. Cores should be taken with a diameter of 25–50 mm and depth equal to the thickness of the cracked element and at least equal to 15 cm (ETEP 2012). At least one core is required for 30 m crack length, and the core will be visually inspected to check if the crack is filled to the appropriate level with resin. The cores can also be tested for compressive and tensile loading following ASTM C42 (1999) and (ACI 2003). The removed core should be patched with a non-shrink or expansive cementitious or epoxy grout (ETEP 2012; ACI 2003).

The use of an endoscope is also recommended for quality control. In this method, holes with a diameter 2 mm larger than the diameter of the tube of the endoscope should be drilled, to a depth equal to the thickness of the cracked element and at least 15 cm. At least two points (in two different crack segments) for every 30 m of crack injected (total length of the injected crack segments) should be inspected. There are also non-destructive methods suggested for the inspection of the repaired elements. Use of the ultrasonic pulse velocity method is also an option (ETEP 2012; ACI 2003). At least three measurements should be taken (in three different crack segments) for every 30 m of crack injected (total length of the injected crack segments) (ETEP 2012).

Epoxy-Bonded Mortar or Concrete for Concrete Repair

Epoxy-bonded mortar or concrete is used for repairs involving shallow replacement of concrete. Epoxy-bonded mortar is used for depth of replacement less than 40 mm, while epoxy-bonded concrete is used in depths between 40 and 140 mm (USBR 1997). Shallow concrete replacements are subject to poor curing conditions as a result of moisture loss to evaporation and to capillary absorption by the existing concrete element with subsequent poor bond to the existing structure. The use of epoxy-bonded mortar or concrete can be used to ensure adequate bond between the old and the new material

(USBR 1997). The surface preparation of the old concrete and the materials used for the repair are critical parameters for the effectiveness of the technique.

Old Concrete Preparation

The first step for the old concrete preparation is to saw cut the perimeter of the damaged area to a depth of 25–40 mm in order to provide a retaining boundary for the compaction and consolidation of the repair material. Then, the deteriorated concrete needs to be removed to create a sound substrate for the bonding of the new repair material (USBR 1997). There are several techniques for the damaged concrete removal such as the use of high-pressure hydroblasting, jackhammer, shotblasting, and dry or wet sandblasting (USBR 1997).

Materials Used for Epoxy-Bonded Repair

The epoxy resin used for epoxy-bonded mortar repairs (less than 40 mm depth) should be two component, 100 % solid type, meeting the requirements of specification ASTM C-881 (2010) for type III, grade 2 (medium viscosity), class B or C (depending on the ambient temperature).

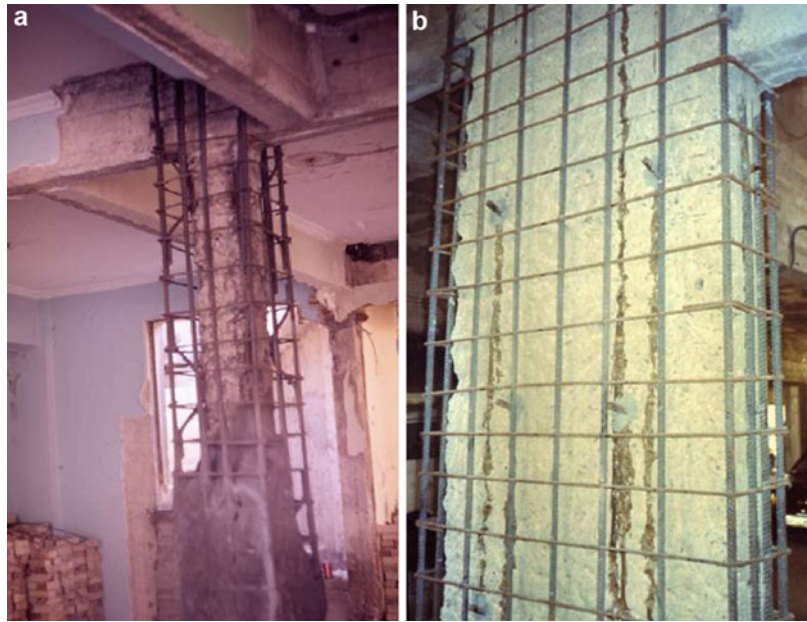
The material used in this technique is conventional concrete and epoxy resin bonding agent. The concrete should be the same as the old concrete but with consistency (slump value) within 40 mm. The resin should be of two components, 100 % solid type, meeting the requirements of specification ASTM C-881 (2010) for type II or type V, class B or C (depending on the ambient temperature). Type II can be used for use in non-load-bearing applications to bond fresh to hardened concrete, while for load-bearing applications, type V should be used (ASTM C-881 2010).

Strengthening Using Additional Layers/Jackets

This technique involves casting of an additional reinforced concrete element in connection to the existing load-bearing elements and is extensively used to upgrade existing buildings (Fig. 3).

Strengthening Techniques: Code-Deficient R/C Buildings,

Fig. 3 Jacketing of existing column and wall. <http://www.episkeves.civil.upatras.gr>



Two crucial issues for the effectiveness of this technique are the interface between the old and the new concrete and shrinkage of the new concrete layer or jacket (Lampropoulos and Dritsos 2011). Concrete shrinkage induces an initial slip at the interface which in some cases can be high enough and even lead to debonding (Beushausen and Alexander 2006, 2007). Also, tensile stresses are developed at the new concrete, which may lead to a reduced concrete compressive strength due to biaxial stress state (Lampropoulos and Dritsos 2010, 2011; Lampropoulos et al. 2012a). Practical recommendations for the strengthening of the additional elements using additional layers or jackets are presented in section “[Interface Preparation](#)”, while in section “[Analysis and Design of Strengthened Elements](#)”, analysis and design methods for strengthened/composite specimens are presented.

Practical Recommendations for Jacketing Procedure

Interface Preparation

In order to guarantee a sufficient connection between contact surfaces, the shear force acting at the interface should be lower than, or equal to, the shear resistance.

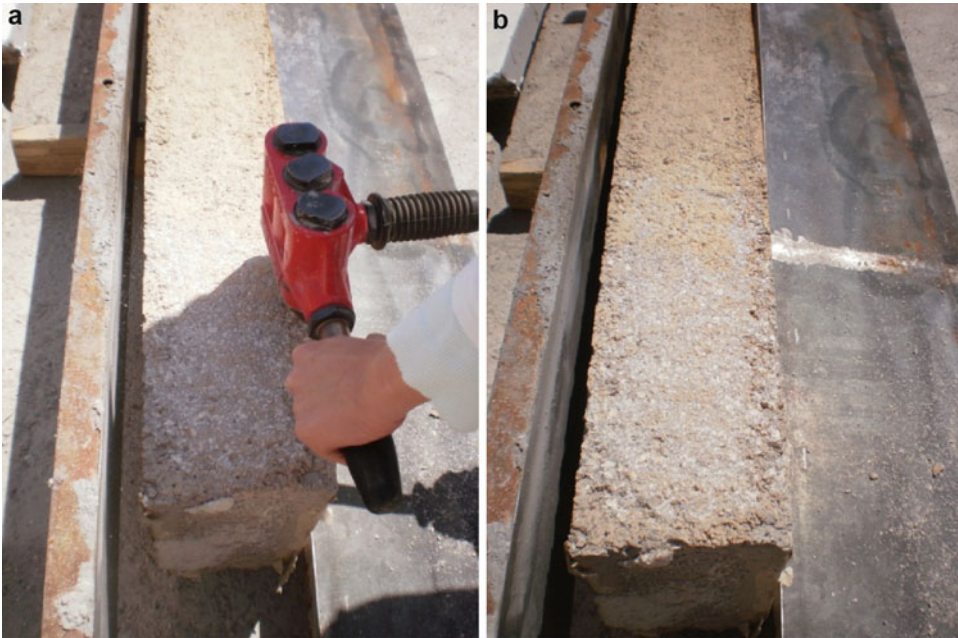
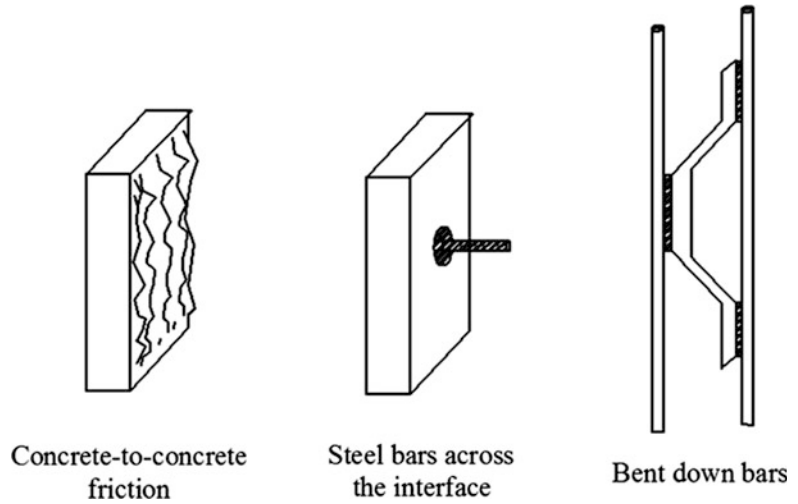
Four mechanisms contribute to the shear resistance at the interface: concrete-to-concrete adhesion, concrete-to-concrete friction (Fig. 4a), the connecting action from either steel bars placed across the interface between the old and the new concrete (Fig. 4b), and bent-down bars welded between the bars of the old and the new concrete (Fig. 4c; Dritsos 2007). These four mechanisms can be subdivided into the two groups of unreinforced and reinforced interfaces, depending on whether or not additional steel is placed across the interface or welded between the bars of the old and the new concrete (Dritsos 2007).

The roughening of the interfaces is crucial for the performance of the strengthened specimens especially in case of unreinforced interfaces. The roughening of the surface of the existing elements can be achieved using jet with water and sand mixture or light air equipment or electric needle (Code of Structural Interventions 2012). In Fig. 5, the use of air chipping hammer for roughening the surface of a beam is presented.

There are two main techniques currently used to characterize the interface roughness. The one of the two is a qualitative approach, wherein the surface of the existing elements is compared to standard surface samples with different roughness

**Strengthening
Techniques:
Code-Deficient R/C
Buildings,**

Fig. 4 Friction, steel bars across the interface, and bent-down bars used in concrete-to-concrete interfaces



Strengthening Techniques: Code-Deficient R/C Buildings, Fig. 5 (a) Roughening procedure and (b) roughened surface

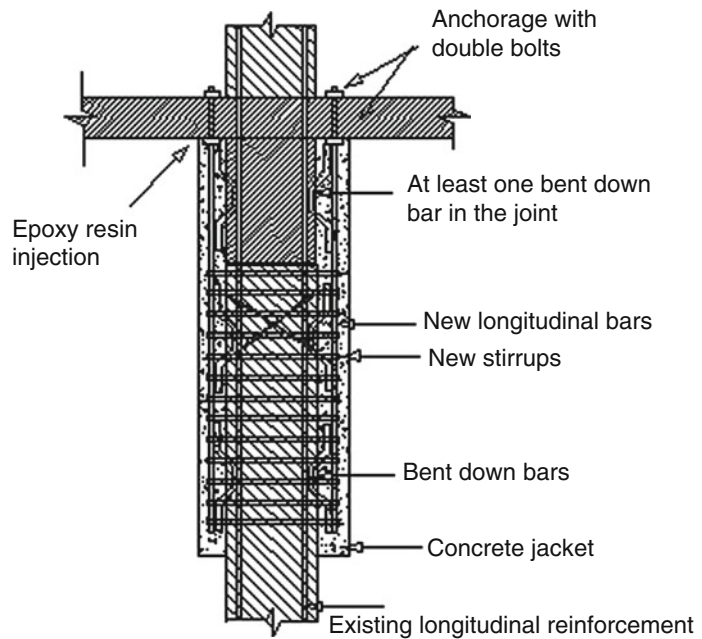
grades. The second one, which is a quantitative approach, is the sand patch test. This method consists in uniformly spreading a standard volume of sand (V) and then measuring the diameter (D) of the covered area from which the roughness is calculated ($=4 V/\pi D^2$). This method presents the advantage of being inexpensive and easy to

perform, both in place and in laboratory, and as all interfaces are horizontal top surfaces, it is easy to be applied (ASTM E965 2003; Santos and Júlio 2010, 2013).

Based on the roughness value, the surface can be classified as very smooth when the roughness is not measurable, smooth when the roughness is

**Strengthening
Techniques:
Code-Deficient R/C
Buildings,**

Fig. 6 Construction detail of a typical jacketed column (Tassios 2009)



lower than 1.5 mm, rough when the roughness is higher than 1.5 mm, and very rough when the roughness is higher than 3 mm (Santos and Júlio 2013; fib 2010).

Apart from the two techniques presented above, there are other “relatively new” techniques used to determine the roughness of the interface (i.e., outflow meter, mechanical stylus, circular texture meter, digital surface roughness meter, microscopy, ultrasonic, slit-island method, roughness gradient method, photogrammetric method, shadow profilometry, air leakage method, processing of the digital image (PDI) method, two-dimensional laser roughness analyzer (2D-LRA) method, and three-dimensional laser scanning method). Most of these methods are based on optical techniques, and their main benefits are the improved accuracy and the fact that they are not destructive methods and can be applied both in the laboratory and in situ (Santos and Júlio 2010).

**Casting of New Concrete and Reinforcement
Details**

For the casting of new concrete, small aggregate size of about 2 mm is normally used because of the lack of space in the jacket or the additional

layer, while use of non-shrinkage concrete is also recommended (Júlio et al. 2003).

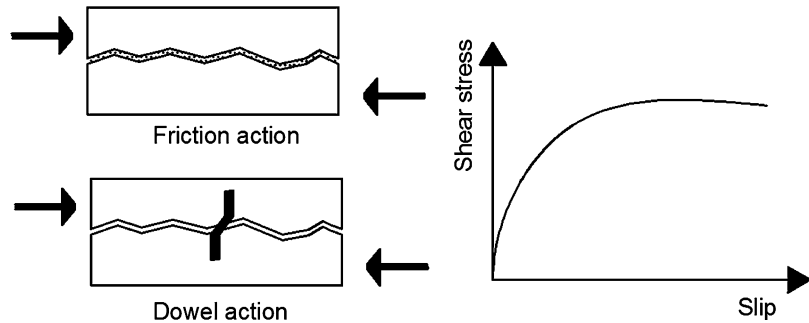
The anchorage of the added longitudinal reinforcement to the existing element is essential to prevent debonding. In case of jacketed columns, continuity between floors is recommended (Júlio et al. 2003). However, it is not always feasible to continue a jacket to the other floors. In such cases, or in order to minimize occupant disturbance, the following construction details are required (Fig. 6) (Tassios 2009):

- (i) Filling of the gap between the jacket and the slab with epoxy resin injection to ensure effective compressive load transfer. The epoxy resin should be injected at least one month after the casting of the jacket.
- (ii) The longitudinal bars of the jacket should be anchored to the slab with double bolts to ensure tensile load transfer.
- (iii) Welding of at least one bent-down bar with each reinforcing bar of the initial column.

In case of “open” jackets, where the jacket cannot be cast on all four sides of the column, special provisions are required to prevent debonding. In three-sided jackets, epoxy grout

Strengthening Techniques:**Code-Deficient R/C**

Buildings, Fig. 7 Typical shear stress-slip model for roughened interface and for interface with dowels



should be used to integrate the stirrups of the jacket into the initial column. In case of strengthening with additional layers only on one or two sides, the connection of the new with the old reinforcement using bent-down bars is essential (Dritsos 2005).

Analysis and Design of Strengthened Elements

To take into account the effect of the interface and concrete shrinkage, finite element analyses can be conducted (Lampropoulos and Dritsos 2011). However, in practice, it is not feasible to perform sophisticated finite element analyses and take into account all these parameters, so the use of monolithicity coefficients is proposed by the codes (Eurocode 8 2005; Code of Structural Interventions 2012). Monolithicity coefficients correlate the characteristics of the strengthened members to those of the respective monolithic ones. This procedure is quite old; however, most of the proposed monolithicity coefficients are empirical, and single values are proposed without accounting for the different values of the normalized axial load, the thickness of the jacket, and the shrinkage of the new concrete parameters that considerably affect the response of the strengthened specimens (Lampropoulos et al. 2012b). In the literature, there are studies (Dritsos 2007; Thermou et al. 2007; Thermou et al. 2014) where monolithicity coefficients values are proposed, and the effect of jacket concrete shrinkage has been ignored. It must be stated that the restrained shrinkage of the jacket concrete significantly influences the response of strengthened columns and cannot be overlooked (Lampropoulos and Dritsos 2010; Lampropoulos

et al. 2012a). Detailed investigations have recently been published where monolithicity coefficient values have been calculated for different values of the normalized axial load and jacket thicknesses (Lampropoulos et al. 2012b; Thermou et al. 2014). The effect of jacket's restraint shrinkage has also been studied (Lampropoulos et al. 2012b), and it has been shown that in some of the examined cases, the "code"-proposed values are not conservative. In the following section (section "Concrete-to-Concrete Interface"), different models for the shear stress calculation at the interface are presented. The proposed method for modeling concrete shrinkage effects is presented in section "Concrete Shrinkage Effect".

Concrete-to-Concrete Interface

There are a number of available models for the calculation of the interface strength between old and new concrete. In most of these models, the shear stress at the interface between the old and new concrete is related to the slip at the interface (Tassios 1986; Tassios and Vintzileou 1987; CEB-FIP 1993; Code of Structural Interventions 2012). In these studies, concrete-to-concrete interfaces with different roughening grades were examined. A typical shear stress-slip model for roughened interfaces with and without steel dowels is presented in Fig. 7.

For the shear resistance due to cohesion, the following model is proposed by CEB-FIP (1993) and Code of Structural Interventions (2012). The distribution of the shear stress with the slip at the interface is considered linear until the limit value for the shear resistance (τ_{fud}) which is the shear stress value for slip equal to 0.01-0.02 mm.

After that point, shear resistance is considered to be constant and equal to the limit value (τ_{fud}). The τ_{fud} value depends on the interface type Eq. 1.

$$\tau_{fud} = \begin{cases} 0.25 \cdot f_{ct}, & \text{smooth interface} \\ 0.75 \cdot f_{ct}, & \text{rough interface} \\ f_{ct}, & \text{shotcreting} \end{cases} \quad (1)$$

where:

τ_{fud} is the shear resistance at the interface,
 f_{ct} is the concrete tensile strength.

The shear stress-slip distribution is presented in Fig. 8.

In case of smooth interfaces, there are also similar models for the shear stress distribution due to friction (Tassios and Vintzileou 1987; CEB-FIP 1993; Code of Structural Interventions 2012; Fig. 9).

According to this model, the shear stress distribution is linearly increasing up to the point of its limit value, while after that point the stress remains constant. The maximum shear stress point is calculated using Eq. 2.

$$\tau_{fud} = 0.4 \cdot \sigma_c, \quad s_{fu} = 0.15 \cdot \sqrt{\sigma_c} \quad (2)$$

where

s_{fu} is the slip at maximum shear strength (τ_{fud}) and σ_c is the design compressive strength at the interface.

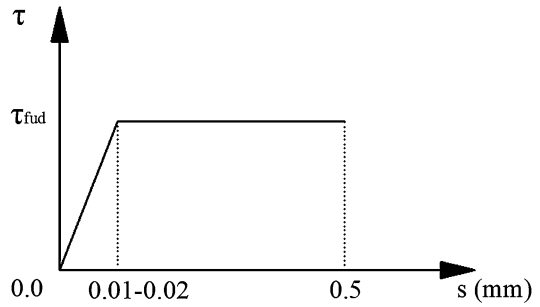
Various models have also been proposed for roughened interfaces. The first model for roughened interfaces is presented in Fig. 10 (Tassios 1986; CEB-FIP 1993).

According to this model, the shear strength of roughened interfaces is linearly increasing up to a slip equal to 0.1 mm. At this point, the shear stress is equal to half of its maximum value. For slip values higher than 0.1 mm and lower than 2.0 mm, the shear resistance can be derived using Eq. 3.

$$\left(\frac{\tau_f}{\tau_{fud}} \right)^4 - 0.5 \cdot \left(\frac{\tau_f}{\tau_{fud}} \right)^3 = 0.3 \cdot s - 0.03$$

where $\tau_{fud} = 0.4 \cdot (f_c^2 \cdot \sigma_c)^{1/3}$

(3)



Strengthening Techniques: Code-Deficient R/C Buildings, Fig. 8 Shear stress-slip model for concrete-to-concrete interfaces due to cohesion (CEB-FIP 1993; Code of Structural Interventions 2012)

A similar model is presented in the Code of Structural Interventions (2012), where the shear stress distribution with the slip is defined using two different equations, before and after half of the maximum slip value Eq. 4 (Fig. 11):

$$\begin{aligned} \frac{s}{s_{fu}} \leq 0.5 &\rightarrow \frac{\tau_f}{\tau_{fud}} \\ &= 1.14 \cdot \sqrt[3]{\frac{s}{s_{fu}}}, \quad \frac{s}{s_{fu}} > 0.5 \rightarrow \frac{\tau_f}{\tau_{fud}} \\ &= 0.81 + 0.19 \frac{s}{s_{fu}} \quad \text{where} \quad \tau_{fud} = 0.4 \cdot (f_c^2 \cdot \sigma_c)^{1/3} \end{aligned} \quad (4)$$

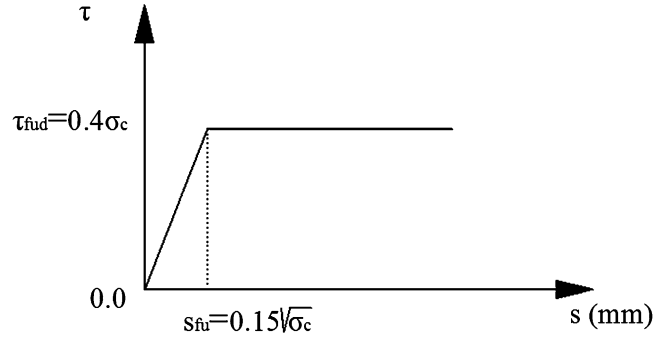
The recommended maximum shear stress at the interface (τ_{fud}) proposed by Model Code 2010 (fib 2010) is in the range of 1.5–2.5 MPa for strengthened interfaces with sandblasting, while for a well-roughened interface with high-pressure hydroblasting, the maximum shear stress is in the range of 2.5–3.5 MPa. These values are valid for concrete with strength lower than C50/60.

Values for the coefficients of friction and cohesion at the interface of strengthened elements have also been proposed. The suggested values according to CEB (1983), together with the coefficient of friction values proposed by Eurocode 2 (1996) and Model Code 2010 (fib 2010) are presented in Table 3.

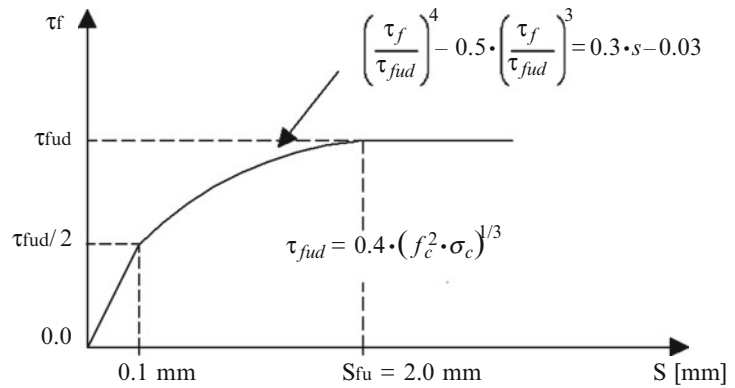
From the values of Table 3, it can be observed that the design values for the coefficient of friction proposed by Eurocode 2 (1996) are lower for

Strengthening Techniques:

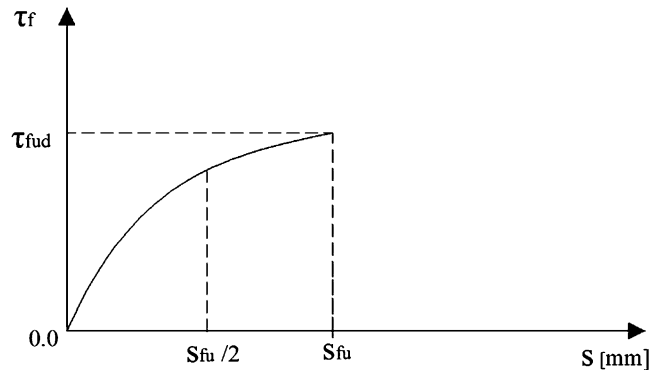
Code-Deficient R/C Buildings, Fig. 9 Shear stress-slip models due to friction for smooth interfaces (Tassios and Vintzileou 1987; CEB-FIP 1993; Code of Structural Interventions 2012)


Strengthening Techniques:

Code-Deficient R/C Buildings, Fig. 10 Shear stress-slip distribution due to friction for roughened interfaces (Tassios 1986; CEB-FIP 1993)


Strengthening Techniques:

Code-Deficient R/C Buildings, Fig. 11 Shear stress-slip distribution due to friction in roughened interfaces (Code of Structural Interventions 2012)



well-roughened interfaces compared to the values proposed by CEB Bulletin No. 162 (CEB 1983) and Model Code 2010 (fib 2010). The coefficient of friction values proposed by Model Code 2010 (fib 2010) are close to the values proposed by CEB (1983).

In the study of Tassios and Vintzileou (1987) and in the Code of Structural Interventions (2012), an analytical relationship for the calculation of coefficients of friction for different

values of the normal to the interface stresses is proposed (Eq. 5).

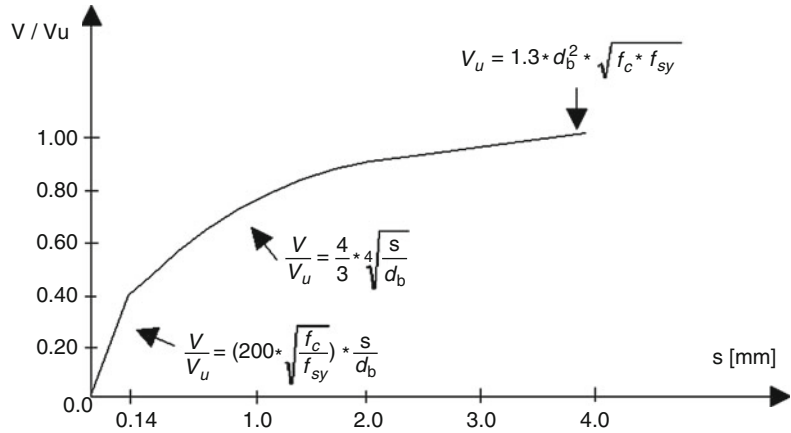
$$\mu = 0.44 * \left(\frac{\sigma_c}{f_c} \right)^{-2/3} \quad (5)$$

Another investigation has been conducted for dowel-reinforced interfaces, and models for the distribution of the shear stress with the friction

Strengthening Techniques: Code-Deficient R/C Buildings, Table 3 Coefficient of friction and cohesion values for strengthened elements with various roughening levels (CEB 1983; Eurocode 2 1996; fib 2010)

	CEB Bulletin No. 162 (1983)		Eurocode 2 (1996)	Model code 2010 (fib 2010)
	Coefficient of friction	Cohesion (MPa)	Coefficient of friction	Coefficient of friction
Well-roughened interface	1.5	1.9	0.9	1–1.4
Roughened interface	0.9	1.7	0.7	0.7–1
Smooth interface	0.7	1.0	0.6	0.5–0.7
Very smooth interface	–	–	0.5	–

Strengthening Techniques: Code-Deficient R/C Buildings, Fig. 12 Shear stress-slip relationship for dowel action (Vintzileou and Tassios 1986)



have been proposed (Vintzileou and Tassios 1986; CEB-FIP 1993; Code of Structural Interventions 2012). In the model proposed by Vintzileou and Tassios (1986), the variation of

the shear stress with the slip is defined by two relationships, depending on the slip values (Eq. 6):

$$\frac{V}{V_u} = \left(200 * \sqrt{\frac{f_c}{f_{sy}}} \right) * \frac{s}{d_b} \quad \text{for } s \leq 0.14 \text{ mm, and } \frac{V}{V_u} = \frac{4}{3} * \sqrt{\frac{s}{d_b}} \quad \text{for } s > 0.14 \text{ mm} \quad (6)$$

where $V_u = 1.3 * d_b^2 * \sqrt{f_c * f_{sy}}$

where:

V	is the shear stress (N),
V _u	is the shear strength (N),
f _{sy}	is the yield stress of steel (MPa), and
d _b	is the diameter of the dowels (mm).

This model is presented in Fig. 12.

A similar bilinear model has been proposed by CEB (1983). The equation for the initial branch

($s \leq 0.14$ mm) is similar to Eq. 6 with the only difference that a coefficient equal to 250 is proposed instead of 200. After the initial elastic branch, there is a second linear one up to the maximum shear resistance.

A different model has been presented in CEB-FIP (1993). According to this model, the shear resistance due to dowel action linear increases up to half of the maximum shear stress.

At this point, the slip equals to 0.05 of the maximum slip and after this point Eq. 7 is used. The

shear stress-slip distribution is presented in Fig. 13.

$$\frac{s}{s_u} = 1.7 \cdot \left(\left[\frac{V}{V_u} \right]^4 - 0.5 \times \left[\frac{V}{V_u} \right]^3 \right) + 0.05 \quad \text{for} \quad 0.05 \leq \frac{s}{s_u} \leq 1 \quad \text{where} \quad s_u = 2.0 \text{ mm} \quad (7)$$

$$V_u = \frac{1.30}{\gamma_{Rd}} \cdot d_b^2 \cdot \left[\sqrt{1 + (1.3 \cdot \varepsilon)^2} - 1.3 \cdot \varepsilon \right] \cdot \sqrt{f_{cd} \cdot f_{yd} \cdot (1 - \zeta^2)} < A_s \cdot f_{yd} / \sqrt{3}$$

where

γ_{Rd}	is considered equal to 1.30,
A_s	is the dowel cross section, and
e	is the load eccentricity.

In the model proposed by the Code of Structural Interventions (2012), the maximum shear strength value is calculated using the same equation proposed by CEB-FIP (1993). The variation of the shear stress with the slip is considered linear for slip equal or lower than 0.005 dB. For any value between 0.005 and 0.05 dB, the shear stress value is calculated using Eq. 8. This model is presented in Fig. 14.

$$s = 0.1 \cdot d_u + 1.80 \cdot d_u \cdot \left(\left[\frac{V}{V_u} \right]^4 - 0.5 \cdot \left[\frac{V}{V_u} \right]^3 \right)$$

$$d_u = 0.05 \cdot d_b \quad (8)$$

The model for the dowel action proposed by Model Code 2010 (fib 2010) is given by Eq. 9

$$V_u = k \cdot A_s \cdot \sqrt{f_{cd} \cdot f_{yd}} \left(\frac{s}{s_u} \right)^{0.5}, \quad (9)$$

$$s_u = 0.1 - 0.2 \cdot d_b$$

where $k \sim 1.6$ for circular reinforcement cross section and concrete strength equal to or lower than C50/60.

There are also models for the reduction of shear stress due to cyclic loading in roughened interfaces and for dowel action (Vintzileou and Tassios 1986, 1987; Tassios and Vintzileou 1987; Code of Structural Interventions 2012). According to these models, there is a shear stress

reduction after each loading cycle. Characteristic diagrams of the shear stress distribution with the slip under cyclic loading are presented in Fig. 15 (Vintzileou and Tassios 1986, 1987; Tassios and Vintzileou 1987).

The first model for the calculation of the reduced shear stress due to cyclic loading is the one proposed by Tassios and Vintzileou (1987) (Eq. 10)

$$\tau_{f,n} = \tau_{f,1} \cdot \left(1 - [0.002 \cdot (n-1) \cdot (s/s_{fu}) / (\sigma_{cn}/f_c)]^{1/3} \right) \quad (10)$$

where

$\tau_{f,n}$	is the shear stress at the interface in the n^{th} loading cycle,
$\tau_{f,1}$	is the shear stress at the interface in the first loading cycle,
N	is the number of cycles, and
σ_{cn}	is the normal to the interface stress in the n^{th} loading cycle.

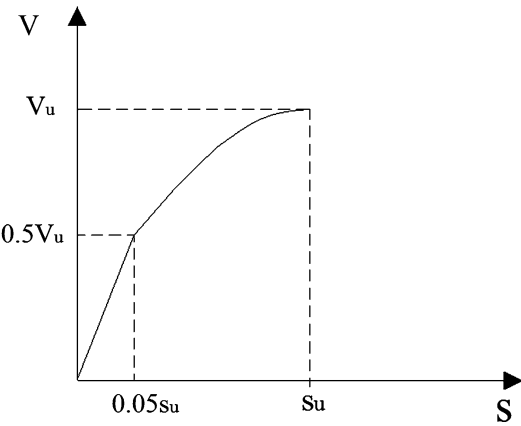
In the Code of Structural Interventions (2012), two models have been presented for the shear stress reduction in smooth (Eq. 11) and roughened interfaces (Eq. 12).

Smooth interfaces:

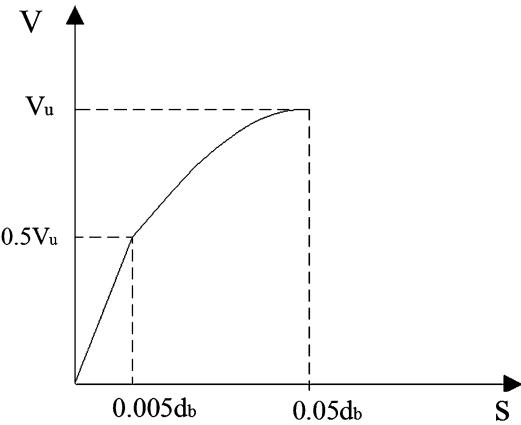
$$\tau_{f,n} = \tau_{f,1} \cdot \left(1 - 0.15 \cdot \sqrt{(n-1)} \right) \quad (11)$$

Roughened interfaces:

$$\frac{\Delta \tau_{f,n}}{\tau_{f,1}} = 0.05 \cdot \left(\frac{f_c}{\sigma_{cn}} \right)^{1/2} \cdot (n-1)^{1/2} \cdot \left(\frac{s}{s_{fu}} \right)^{1/3} \quad (12)$$



Strengthening Techniques: Code-Deficient R/C Buildings, Fig. 13 Shear stress-slip relationship for dowel action (CEB-FIP 1993)



Strengthening Techniques: Code-Deficient R/C Buildings, Fig. 14 Shear stress-slip relationship for dowel action (Code of Structural Interventions 2012)

The following model has been proposed by Vintzileou and Tassios (1986) and Vintzileou and Tassios (1987) for dowel action under cyclic loading (Eq. 13)

$$D_n = D_1 * \left(1 - \frac{1}{7} * \sqrt{(n - 1)}\right) \tag{13}$$

where

D_n	is the maximum shear stress due to dowel action in the n^{th} loading cycle,
D_1	is the maximum shear stress due to dowel action in the 1 st loading cycle.

Concrete Shrinkage Effect

Concrete shrinkage can considerably affect the performance of strengthened elements with additional layers or jackets. Tensile stresses are induced to the new concrete of the layer or jacket while there is also an induced slip at the interface between the old and the new concrete. Denarie and Silfwerbrand (2004) presented a model where the tensile stresses of the new concrete due to restrained concrete shrinkage are proportional to coefficient “R.” Coefficient “R” is characteristic of the connection between the old and the new concrete and the degree of restraint (Eq. 14).

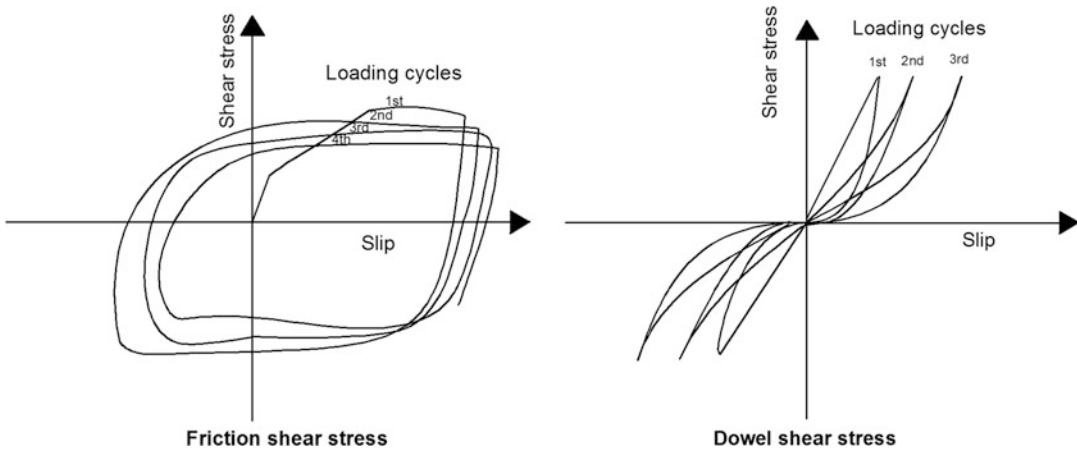
$$\sigma = E * \epsilon_{fsh} * R \tag{14}$$

where

E:	modulus of elasticity of concrete,
ϵ_{fsh} :	free shrinkage strain, and
R:	degree of restraint.

Values for the coefficient R have been proposed by Abbasnia et al. (2005) based on experimental investigations. The risk of debonding and cracking of the new layer of strengthened beams was investigated by Beushausen and Alexander (2006, 2007) where the tensile stress relaxation due to concrete creep was also taken into consideration.

The biaxial stress state in strengthened elements due to the development of restrained shrinkage tensile stresses was numerically investigated, and the importance of the restrained concrete shrinkage in case of jacketed columns was highlighted (Lampropoulos and Dritsos 2010, 2011). An extensive experimental study was also conducted to validate the reliability of the numerical results (Lampropoulos et al. 2012a). In this study, specimens with and without steel plates (with expanded polystyrene) were cast, to simulate the concrete of the jacket under restrained and free shrinkage, respectively. Specimens were stored until testing in a room with a constant temperature and relative humidity, and shrinkage strain at the end of the specimens was measured using micrometer dial



Strengthening Techniques: Code-Deficient R/C Buildings, Fig. 15 Characteristic shear stress-slip diagrams for friction and dowel shear stress due to cyclic loading

Strengthening Techniques: Code-Deficient R/C Buildings,

Fig. 16 Specimens with and without steel plates and strain measuring setup



gauges. The specimens together with the setup for the shrinkage strain measurements are presented in Fig. 16.

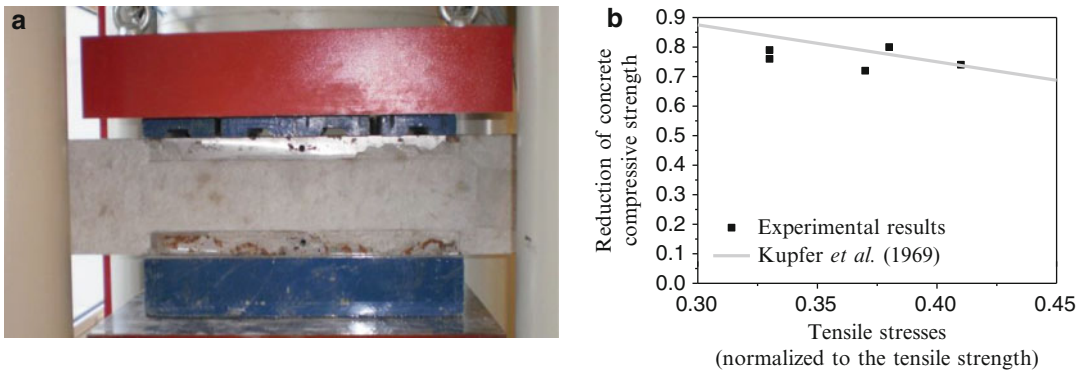
Each pair of specimens (with steel plates or expanded polystyrene cast from the same batch of concrete) was tested in compression at a different age, to obtain results for different values of the shrinkage strain. The specimens with restrained and free concrete shrinkage were tested under compression (Fig. 17a). The ratio of the tensile stress to the tensile strength together with the respective reduction of the concrete strength was calculated. It was found that there is a significant reduction of concrete strength due to restrained concrete shrinkage and the subsequent biaxial stress state, and this is in agreement with the model proposed by Kupfer et al. (1969) (Fig. 17b).

In the tested square prism specimens, the reduction of the concrete strength was found to range from 20 % to 30 %. This reduction was more significant for larger values of tensile stresses caused by the restrained concrete shrinkage (Lampropoulos et al. 2012a).

Modeling of Strengthened Elements and Monolithicity Coefficients for the Design of Strengthened Elements

Finite Element Modeling

For modeling the strengthened elements, the finite element method has been used. Different assumptions have been examined, and the accuracy of the numerical models was validated using experimental data (Lampropoulos and Dritsos 2011).



Strengthening Techniques: Code-Deficient R/C Buildings, Fig. 17 (a) Compressive test setup and (b) Biaxial stress state due to restrained concrete shrinkage (Lampropoulos et al. 2012a)

Interface Between the Old and the New Concrete The interface between the old and the new concrete can be simulated using spring elements or using special contact elements. To determine the characteristics of the interface, the models presented in section “Concrete-to-Concrete Interface” can be used. Strength deterioration due to cyclic loading should also be taken into account (Lampropoulos and Dritsos 2011).

Concrete Shrinkage Concrete shrinkage effect can be simulated by applying volumetric strain to the elements of the jacket. The effect of concrete creep should also be included to take into account the tensile stress relaxation. In a previous study (Lampropoulos and Dritsos 2010), in order to include creep effects, application of a volumetric strain equal to half of the free concrete shrinkage strain was proposed.

Monolithicity Coefficients

The accuracy of the numerical models was validated using a number of different experimental data. The same assumptions were used in a parametric study where columns with different characteristics were examined (Lampropoulos et al. 2012b). Analyses were performed for the strengthened and the respective monolithic specimens (Fig. 18), and monolithicity coefficients were calculated using Eq. 15. Monolithic specimens are specimens with exactly the same characteristics as the examined strengthened

columns, with the only difference that there is full connection (no slip) between the original column and the jacket, and concrete shrinkage is neglected.

The load and the deflection at the points of yield, maximum load, and failure of the specimens were estimated, and Eq. 15 was used to calculate monolithicity coefficient values.

$$K_F = \frac{[F_{\max}]_{STR}}{[F_{\max}]_{MON}}, \quad K_k = \frac{\left[\frac{F_y}{\delta_y}\right]_{STR}}{\left[\frac{F_y}{\delta_y}\right]_{MON}}, \quad (15)$$

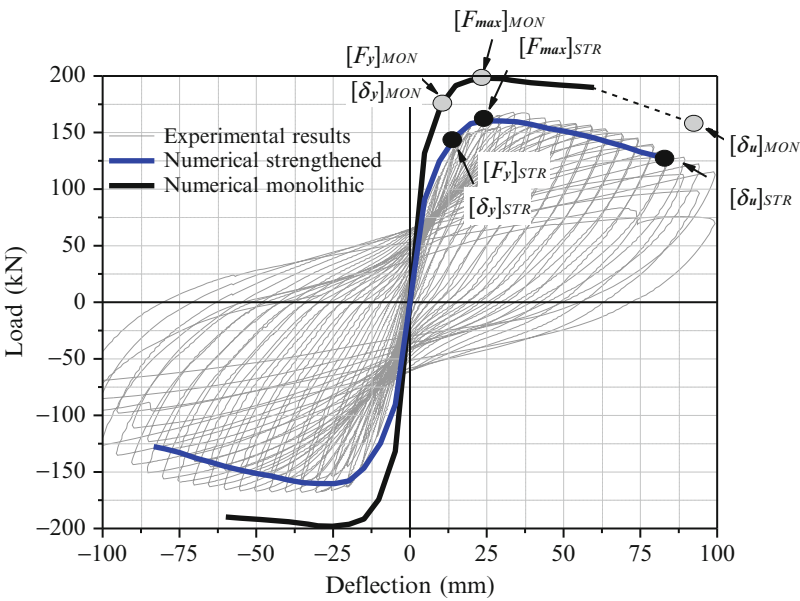
$$K_{\delta, \theta_y} = \frac{[\delta_y]_{STR}}{[\delta_y]_{MO}}, \quad K_{\delta, \theta_u} = \frac{[\delta_u]_{STR}}{[\delta_u]_{MO}}$$

where K_F , K_k , and $K_{\delta, \theta_y, u}$ are monolithicity coefficients for the strength, stiffness, and deflection or rotation angle at yield and failure, respectively, F_y and $F_{\max}$ are the yield and maximum load, and δ_y , δ_u are the deflections at yield and failure. Subscripts *STR* and *MON* indicate strengthened and monolithic specimens, respectively.

The proposed design values are based on the assumption of concrete with free shrinkage strain of 800 microstrains. A roughened interface was assumed without any other connection technique (bent-down bars or dowels) which is a conservative assumption. Different values have been proposed depending on the normalized axial load (v) and the ratio of the concrete cross-sectional area of the jacket (A_{cj}) to the cross-sectional area of

Strengthening Techniques: Code-Deficient R/C Buildings,

Fig. 18 Typical load deflection relationship for the strengthened and the respective monolithic specimens (Lampropoulos and Dritsos 2011, 2012b)



the original column (A_{co}) (A_{cj}/A_{co}). The normalized axial load (v) can be calculated using Eq. 16.

$$v = \frac{N}{A_{co} \cdot f_{co} + A_{cj} \cdot f_{cj}} \tag{16}$$

where N is the axial load, A_{co} is the cross-sectional area of the original column, f_{co} is the concrete strength of the original column, A_{cj} is the cross-sectional area of the concrete jacket, and f_{cj} is the strength of the jacket concrete.

Four- and three-sided jackets have been examined. In case of three-sided jacket, the stirrups of the jacket were merged into the initial column in order to simulate anchorage, which is essential and in practice is ensured by using epoxy grout (Lampropoulos et al. 2012b).

The proposed values for a four-sided jacket with A_{cj}/A_{co} ratio of 1.5 and normalized axial load values of 0.1, 0.2, and 0.4 are presented in Table 4 (Lampropoulos et al. 2012b) together with the values proposed by existing codes (Eurocode 8 2005; Code of Structural Interventions 2012).

Table 5 presents monolithicity coefficient values for a normalized axial load of 0.2 and different A_{cj}/A_{co} ratio values for a four-sided jacket (Lampropoulos et al. 2012b).

Strengthening Techniques: Code-Deficient R/C Buildings, Table 4 Monolithicity coefficient values for a four-sided jacket with different normalized axial load values and an A_{cj}/A_{co} ratio of 1.5 (Lampropoulos and Dritsos 2012b)

	v	Monolithicity coefficients			
		K_F	K_k	K_{δ_y, θ_y}	K_{δ_u, θ_u}
Proposed design values	0.1, 0.2	0.85	0.55	1.50	1.00
	0.4	0.70	0.50	1.30	1.10
Eurocode 8, Part 3 (2005)		0.90	0.95	1.20	1.00
Code of Structural Interventions (2012)		0.90	0.80	1.25	0.80

The proposed values for the monolithicity coefficients for a three-sided jacket with an A_{cj}/A_{co} ratio of 1.0, and normalized axial load values of 0.1, 0.2, and 0.4 are presented in Table 6 (Lampropoulos et al. 2012b).

Table 7 presents proposed monolithicity coefficients values for a three-sided jacket and a normalized axial load value of 0.2 with different A_{cj}/A_{co} ratio values (Lampropoulos et al. 2012b).

Monolithicity coefficients for any case with a different A_{cj}/A_{co} ratio or normalized axial load

Strengthening Techniques: Code-Deficient R/C Buildings, Table 5 Monolithicity coefficient values for a four-sided jacket with different A_{cj}/A_{co} ratio values and a normalized axial load of 0.2 (Lampropoulos and Dritsos 2012b)

	A_{cj}/A_{co}	Monolithicity coefficients			
		K_F	K_k	K_{δ, θ_y}	K_{δ, θ_u}
Proposed design values	0.5	0.90	0.70	1.30	0.95
	1.5	0.85	0.55	1.50	1.00
	4.0	0.75	0.75	1.05	2.85
Eurocode 8, Part 3 (2005)		0.90	0.95	1.20	1.00
Code of Structural Interventions (2012)		0.90	0.80	1.25	0.80

Strengthening Techniques: Code-Deficient R/C Buildings, Table 6 Monolithicity coefficient values for a three-sided jacket with different of normalized axial load values and an A_{cj}/A_{co} ratio of 1.0 (Lampropoulos and Dritsos 2012b)

	v	Monolithicity coefficients			
		K_F	K_k	K_{δ, θ_y}	K_{δ, θ_u}
Proposed design values	0.1, 0.2	0.80	0.55	1.40	1.10
	0.4	0.70	0.35	1.90	1.70
Eurocode 8, Part 3 (2005)		0.90	0.95	1.20	1.00
Code of Structural Interventions (2012)		0.90	0.80	1.25	0.80

value from those presented in Tables 4, 5, 6, and 7 can be estimated by linear interpolation (Lampropoulos et al. 2012b).

The monolithicity coefficient values for the design of strengthened RC columns presented herein depend on the type of the jacket (four or three sided), the normalized axial load value, and the thickness of the jacket. The proposed values are more realistic than using one value to cover all situations, as is the case with present code recommendations (Eurocode 8 2005; Code of Structural Interventions 2012). It was also found that in many cases, monolithicity coefficient

Strengthening Techniques: Code-Deficient R/C Buildings, Table 7 Monolithicity coefficient values for a three-sided jacket with different A_{cj}/A_{co} ratio values and a normalized axial load of 0.2 (Lampropoulos and Dritsos 2012b)

	A_{cj}/A_{co}	Monolithicity coefficients			
		K_F	K_k	K_{δ, θ_y}	K_{δ, θ_u}
Proposed design values	0.5	0.85	0.55	1.70	1.15
	1.5	0.80	0.55	1.40	1.10
	4.0	0.75	0.25	3.00	1.00
Eurocode 8, Part 3 (Eurocode 8 2005)		0.90	0.95	1.20	1.00
Code of Structural Interventions (2012)		0.90	0.80	1.25	0.80

values proposed by codes are not conservative. The proposed values of the present work can also be used when considering other techniques to connect between the original column and the new jacket (such as dowels crossing the interface or bent-down bars welded between the two sets of longitudinal bars), as these values are conservative (only roughening at the interface was considered) and are suitable for design purposes (Lampropoulos et al. 2012b).

Summary

In this entry, the technique of concrete repair with epoxy grouting and the strengthening of existing structures with additional reinforced concrete layers and jackets were described in detail. Practical recommendations were presented for both techniques. For the strengthening of the existing structures using additional layers and jackets, recommendations for the analysis and monolithicity coefficients for the design of the strengthened elements were also presented.

The following general recommendations can be drawn:

- Epoxy resins can be used for the crack sealing and for the replacement of damaged concrete. Due consideration should be given to

successfully apply this method. In this chapter, a step-by-step procedure was presented, with detailed instructions about the epoxy resin injection technique, including details about the substrate preparation, the nozzle placement, and the quality control. The use of epoxy resins for shallow concrete replacements was also described, and information about the suitable material for these applications was presented.

- The technique of strengthening using additional layers and jackets is one of the most commonly used techniques. Two crucial parameters for the effectiveness of this technique are the connection between the old and the new elements and shrinkage of the new concrete.
- Models for the shear stress at the interface for reinforced and unreinforced interfaces and models for the strength deterioration of the interfaces due to cyclic loading were presented.
- For the modeling of the strengthened specimens, it is essential to simulate the interface between the old and the new concrete and the shrinkage effect of the new concrete. The interface can be simulated using contact elements, while concrete shrinkage effect can be simulated by applying a volumetric strain to the elements of the jacket.
- For the design of strengthened elements, monolithic coefficient values can be used. Various values have been proposed for the design of strengthened RC columns, based on the type of the jacket (four or three sided), the normalized axial load value, and the thickness of the jacket.

Cross-References

- [Assessment and Strengthening of Partitions in Buildings](#)
- [Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used](#)
- [Retrofitting and Strengthening Measures: Liability and Quality Assurance](#)
- [Retrofitting and Strengthening of Contemporary Structures: Materials Used](#)
- [Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions](#)
- [Seismic Strengthening Strategies for Existing \(Code-Deficient\) Ordinary Structures](#)
- [Seismic Strengthening Strategies for Heritage Structures](#)
- [Strengthening Techniques: Bridges](#)
- [Strengthening Techniques: Masonry and Heritage structures](#)

References

- Abbasnia R, Godossi P, Ahmadi J (2005) Prediction of restrained shrinkage based on restraint factors in patching repair mortar. *Cem Concr Res* 35:1909–1913
- ACI (2003) Structural Crack repair by epoxy injection, field guide to concrete repair application procedures. ACI Committee E706, RAP Bulletin 1
- ASTM C42/C42M–99 (1999) Standard test method for obtaining and testing drilled cores and sawed beams of concrete. West Conshohocken, Pennsylvania
- ASTM C881/C881M (2010) Standard specification for Epoxy-resin-base bonding systems for concrete. ASTM International, West Conshohocken
- ASTM E965 (2003) Standard test method for measuring pavement macrotexture depth using a volumetric technique. ASTM International, West Conshohocken
- Beushausen H, Alexander MG (2006) Failure mechanisms and tensile relaxation of bonded concrete overlays subjected to differential shrinkage. *Cem Concr Res* 36:1908–1914
- Beushausen H, Alexander MG (2007) Localised strain and stress in bonded concrete overlays subject to differential shrinkage. *Mater Struct* 40:189–199
- CEB Bulletin No. 162 (1983) Assessment of concrete structures and design procedures for upgrading (Redesign), CEB, Paris
- CEB-FIP (1993) Model code 1990. Thomas Telford, London
- Code of Structural Interventions (2012) Team for development of code of interventions on reinforced concrete buildings harmonization team of code of interventions to Eurocodes. Earthquake Planning and Protection of Greece (E.P.P.O.), Athens
- Denarie E, Silfwerbrand J (2004) Structural behaviour of bonded concrete overlays. In: *Proceedings international RILEM workshop on bonded concrete overlays*, Stockholm, pp 37–45
- Dritsos SE (2005) Repair and strengthening of reinforced concrete structures. Dritsos, Patras (in Greek)
- Dritsos SE (2007) Seismic strengthening of columns by adding new concrete. *Bull N Z Soc Earthq Eng* 40(2):49–68
- Dritsos SE (2012) Seismic assessment and retrofit of existing buildings under the eurocodes framework. *Assessment and Retrofitting Framework*, Boku

- ETEP (2012) Recommended technical specifications for retrofitting. Technical Chamber of Greece, Athens (in Greek)
- Eurocode 2 (1996) Design of concrete structures Part 1.3: general rules – precast concrete elements and structures. British Standards Institution; ENV 1992-1-3, London
- Eurocode 8 (2005) European standard EN. Design of structures for earthquake resistance, Part 3, assessment and retrofitting of buildings. CEN Technical Committee CEN/TC250, Brussels
- fib Bulletin No 55 (2010) Model code 2010. Thomas, Telford
- Júlio ES, Branco F, Silva VD (2003) Structural rehabilitation of columns with reinforced concrete jacketing. *Prog Struct Eng Mater* J 5:29–37
- Kupfer H, Hilsdorf H, Rusch H (1969) Behaviour of concrete under biaxial stresses. *Am Concr Inst J* 66(8):656–666
- Lampropoulos AP, Dritsos SE (2010) Concrete shrinkage effect on columns strengthened with concrete jackets. *Struct Eng Int* 20(3):234–239
- Lampropoulos AP, Dritsos SE (2011) Modelling of RC columns strengthened with RC jackets. *J Earthq Eng Struct Dyn* 40(15):1689–1705
- Lampropoulos AP, Tsioulou OT, Dritsos SE (2012a) Biaxial stress due to shrinkage in concrete jackets of strengthened columns. *ACI Mater J* 109(3): 331–340
- Lampropoulos AP, Tsioulou OT, Dritsos SE (2012b) Monolithic coefficient values for design when seismically strengthening RC columns with jackets. *J Earthq Eng* 16(7):1023–1042
- Santos P, Júlio E (2010) Comparison of methods for texture assessment of concrete surfaces. *ACI Mater J* 107(5):433–440
- Santos P, Júlio E (2013) A state-of-the-art review on roughness quantification methods for concrete surfaces. *Construct Build Mater* 38:912–923
- Tassios T (1986) Fundamental mechanisms of force-transfer across reinforced concrete critical interfaces. CEB workshops of Commissions II and IV, Karlsruhe
- Tassios T (2009) Design theory for repair and strengthening. Symmetria Publications, Athens (in Greek)
- Tassios T, Vintzileou E (1987) Concrete to concrete friction. *Jf Struct Eng ASCE* 113(4):832–849
- Thermou GE, Pantazopoulou SJ, Elnashai AS (2007) Flexural behavior of brittle RC members rehabilitated with concrete jacketing. *ASCE J Struct Eng* 133(10): 1373–1384
- Thermou GE, Papanikolaou VK, Kappos AJ (2014) Monolithic factors for the design of R/C columns strengthened with R/C jackets. In: The 2nd European conference on earthquake engineering and seismology, Istanbul
- USBR (1997) United States Department of the Interior Bureau of Reclamation Technical Service Centre, Guide to concrete repair. http://www.usbr.gov/pmts/materials_lab/concrete/
- Vintzeleou E, Tassios T (1986) Mathematical model for dowel action under monotonic and cyclic conditions. *Magazine of Concrete Research* 38(134):13–22
- Vintzeleou E, Tassios T (1987) Behavior of dowels under cyclic deformations. *Am Concr Inst Struct J* 84(1):18–30. <http://www.episkeves.civil.upatras.gr> (Repair and Strengthening of Structures, University of Patras webpage)

Strengthening Techniques: Code-Deficient Steel Buildings

Konstantinos Daniel Tsavdaridis

School of Civil Engineering, University of Leeds, Leeds, UK

Synonyms

Braced frames; Composite materials; Connections; Cyclic behavior; Deformation capacity; Energy dissipation; Fuses; Repair; Seismic retrofitting; Steel buildings; Strengthening techniques

Introduction

The design of steel buildings is often governed by lateral wind loads and not seismic loads. Also, statistics indicate that the number of fatalities during earthquakes due to failure of all types of steel buildings is significantly less compared to other types of buildings. Consequently, much effort has been invested to seismically retrofit buildings having unreinforced masonry walls and reinforced concrete frames. However, recently steel buildings have received significant attention, while this interest mainly stems from the realization, following the 1994 Northridge earthquake, that the welded beam-to-column connections in moment-resisting frames were likely to fail in a brittle manner, prior the development of significant inelastic response, therefore negating the design intent and possibility causing safety hazards.

Recent research has expanded the variety and versatility of the tools available in the structural

engineer's toolbox to meet the seismic performance objectives. This entry provides an overview of how this research is expanding the available options for the seismic strengthening of steel buildings, by reporting on some selected research projects.

Structural strengthening and proving seismic resistance for steel building, but also masonry and reinforced concrete, may be done by first considering the direction of the weak links in the structures. For instance, for a heavy building with large dead load, this would be the major factor that contributes to the increase of lateral seismic load. Therefore, it is reasonable to first consider reducing the overall existing dead load and then provide the necessary strengthening technique for the lateral load-resisting system of the structure.

The use of structural steel in buildings' retrofitting can be often considered economical and efficient because:

- Steel buildings are particularly effective under performance-based design.
- Steel members exhibit ductile behavior beyond elastic limit, hence dissipate considerable amount of energy before damages occur.
- Steel members have higher strength-to-weight and stiffness-to-weight ratios; hence, the buildings attract less base shear under an earthquake.
- A better quality control is practiced in the production of the material as well as the fabrication and erection of them, while ensuring results close to the theoretical predictions.
- Steel can be generally used to retrofit all types of structures without increasing the dead weight dramatically, making the works less intrusive and time consuming.

Code-Deficient Buildings

All buildings can carry their own weight. They can usually carry a bit of snow and a few other floor loads vertically, so even badly built buildings and structures can resist some up-and-down loads. However, buildings and structures are not necessarily resistant to lateral loads, unless this has been taken into account carefully during the

structural engineering design and construction phase with some earthquake-proof measures taken into consideration. It is the side-to-side load which causes the worst damage. Poorly designed buildings often collapse on the first shake. The side-to-side load can be even worse if the shocks come in waves, as taller buildings can vibrate like a huge tuning fork, while each new sway is bigger than the last one, until failure. Usually, significant weight is added in time to such code-deficient steel buildings (i.e., walls, partitions to make more and smaller rooms, etc.) or even due to extreme reinforcing techniques. The more weight there is, and the higher this weight is located in the building, the stronger the building and its foundations must be to withstand the earthquake actions. Many buildings have not been strengthened when such extra weight was added. These buildings are then more vulnerable to even a weak aftershock, perhaps from a different direction or at a different frequency, and can cause collapse. Moreover, in a lot of multistory steel buildings, the ground floor has increased headroom with taller slender columns as well as with more large openings and fewer walls. So, these columns, which carry the largest loads from both the self-weight and the cumulative sideways actions from the seismic event, are vulnerable and they are often the first to fail. It only takes one to fail for the worst disaster; therefore, it is deemed necessary to cautiously strengthen steel buildings with the most appropriate method.

The potential deficiencies are different for different types of steel buildings (i.e., steel moment frames, steel braced frames, steel frames with concrete shear walls, steel frames with infill masonry shear walls). The indicators such as the global strength and stiffness, the configuration, the load path, the component detailing, the diaphragm, and the foundation design demonstrate the performance under seismic actions and the margins for improvement in specific ways; hence, they should be studied carefully before any decision is taken.

Retrofitting of existing code-deficient steel buildings accounts for a major portion of the total cost of hazard mitigation. Therefore, it is

important to identify correctly the structures that need and can accept strengthening or weakening, while the overall cost should be also monitored. If appropriate, seismic retrofitting should be performed through several methods such as increasing the load, deformation, and energy dissipation capacity of the structure (FEMA 356 2000).

Code-Efficient Buildings Resistant to Earthquake

To be earthquake proof, the buildings and their foundations need to be built to be resistant to sideways loads. The lighter the building is, the less the loads are. In steel, especially in high-rise buildings, the sideways resistance mainly comes from the diagonal bracing which must be placed equally in both directions. Where possible, the diagonal bracing should be strong enough to accept tension as well as compression loads; the bolted or welded connections should resist more tension than the ultimate tension value of the brace or much more than the design load. If the sideways load is to be resisted with moment-resisting framing, then great care has to be taken to ensure that the joints are stronger than the beams and that the beams will fail before the columns. Also in such a case, special care should be provided to the foundation-to-first-floor level, avoiding soft-story effects, while the columns should be much stronger than at higher levels. The foundations could be enhanced by having a grillage of steel beams at the foundation level able to resist the high column moments and keep the foundations in place. The main beams should be fixed to the outer columns with full capacity joints, which almost mean hunched connections, and care should be taken to consider the shear within the column at these connections.

When the steel beams are able to yield and bend at their highest stressed points, without losing resistance, while the connections and the columns remain full strength, then the resonant frequency of the whole frame changes, while the energy is absorbed and evenly dissipated across the framing.

The vibration occurred from the shock waves is tend to be damped out. This phenomenon is called “plastic hinging” and is easily demonstrated in steel beams. In extreme earthquake sway, the beams should always be able to form hinges somewhere, while the columns should behave elastically. In this way, the frame can deflect and the plastic hinges can absorb energy, while the resonant frequency of the structure is altered without major loss of strength and inevitable collapse. All floors should be connected to the framing in a robust and resilient way and should be as light as possible. They should possibly span around each column and be fixed to every supporting beam using enough shear connectors (e.g., studs). An effective way of reducing the vulnerability of large buildings is to isolate them from the floors using bearings or dampers; however, this is an expensive process, and it is not applied to low to medium rise buildings which have not been classified as important, due to the content they carry and the occupancy usage.

Nothing can be though guaranteed to behave as such, even in code-based designs; hence, most of the steel buildings and especially those under-designed with older seismic codes can be considered as code-deficient buildings in certain circumstances.

Design Concept for EC8

Eurocode 8 (EC8) follows three general design concepts based on the ductility requirements and capacity design considerations of steel buildings: the concept of the low-dissipative structural behavior of DCL structures, the concept of dissipative structural behavior of DCM and DCH structures satisfying the ductility and capacity design requirement, and the concept of dissipative structural behavior with steel dissipative-controlled zones. In the latter case, when composite action may be considered from Eurocode 4 (EC4) in the presence of the steel and concrete (slab) interaction, specific measures have been stipulated to prevent the contribution of concrete under seismic conditions, hence apply general rules for steel frames.

Introduction to Strengthening Techniques

Preliminary Investigation

It is becoming preferable, both environmentally and economically, to upgrade building structures rather than to demolish them and rebuild them. Engineers assessing structures for increased or special loadings are finding that new methods of analysis using, for instance, computer models are revealing shortcomings under service and ultimate conditions. Under such circumstances, a method has to be found to bring the structure up to the required standard. There is a range of techniques which can be used on structures, but, one must take into consideration that disruption to normal stage must be minimal while work is in progress.

Evaluation and subsequent strengthening of existing structures require a realistic and pragmatic design approach. However, some of the solutions proposed by researchers do not lie within this category and they will be eliminated in the current study. Also, effective communication between the owner, structural engineer, architect, risk analyst, insurance provider, and other stakeholders is paramount to a successful finished solution. In general, in the case where additional load-carrying capacity is required of an existing building, engineers have the option of either **reinforcing** the existing framing or **adding** new framing to replace or supplement the existing.

Where a decision is made to strengthen some parts of an existing facility or a specific structural system or element, the design approach is influenced by a series of factors such as:

- Information about relevant existing conditions which is often limited
- Part of the structure to be strengthened is commonly hidden or obstructed by existing architectural or building services systems that are difficult or costly to remove
- Structural renovation work is typically constrained by the need for continuity of building operations
- Level of ductility of the existing construction may limit its strength

- The susceptibility to local buckling of outstanding flanges as well as the lack of connection ductility

Often, the nonstructural costs will likely exceed the structural costs; therefore, the true costs of a retrofit project are primarily dependent on the number of locations of work than the amount of work done in each location, and thus, this influences the structural design and analysis decisions.

The general approach to strengthen existing structures includes the following aspects:

- Risk assessment and structural vulnerability assessment
- Preliminary analysis
- Consideration of alternatives (structural and nonstructural)
- Detailed design and impact of connections

The common goal is to:

- Protect specific structural elements.
- Provide redundancy to structural systems.
- Strengthen a specific part of the structure.

Assessing Existing Conditions and Strengthening Methods

Introduction

A site visit should also be performed to inspect the building, especially for structures more than 30 years old. Some key things to look for when assessing the existing condition of a steel building are: damage to framing, noticeable corrosion, signs that modifications to the structure that may have been performed without engineering review, unusual deflections in floor framing, cracks in supported slabs, signs of foundation settlement, signs for new rooftop equipment, heavy hung piping loads, folding partitions, rigging, or other suspended loads that may have been added without proper structural engineering review (Schwinger 2014). A valuable resource available to structural engineers working with existing building structures is the *AISC Steel Design Guide 15 – AISC Rehabilitation and*

Retrofit Guide (Brockenbrough 2002). Other publications for further reference are ASCE 41-06 2006, FEMA 274 1997, and FEMA 547 2006.

The strengthening methods can be categorized as follows:

1. Passive against active methods.
2. Strengthening techniques:
 - (a) Reinforcing beams by welding (enlarge section with plates).
 - (b) Reinforcing (or weakening) connections:
 - Framing.
 - Seated angles.
 - Partial-depth end plate.
 - Replace with high-strength fasteners.
 - Add welds at the perimeter of the connection and/or properly clean existing welds.
 - Converting single- to double-shear connections by adding angles or plates.
 - Add web stiffener plate.
 - Add steel cover plates.
 - Enhance column splices.
 - Enhance braced frame connection.
 - (c) Shortening span (provided that there are no fitting issues):
 - Add beams.
 - Add columns and girders.
 - Add diagonal braces.
 - Add walls with openings.
 - Add steel braced frame.
 - Add concrete, masonry, or steel plate walls.
 - Enhance strength and ductility of braced frames.
 - (d) Introducing composite action:
 - Steel (partially or fully) encased with concrete
 - Shear connectors
 - (e) Post-tensioning (or external pre-stressing) of beams and connections (considering eccentricities of brackets on member capacity – some need protection from corrosion, fire, and vandalism).
 - (f) Openings in existing beams (using thermal cutting – plasma arc cutting is faster than

oxy-fuel – while avoiding cuts at areas subjected to high shear):

- Place reinforcement (e.g., stiffeners) before cutting holes.
- (g) Replacement of members (may be economical).
 - (h) Strengthening columns.
 - (i) Convert gravity frame to moment-resisting frame.

Determining Load Capacity of Existing Buildings

Knowing the yielding strength of the steel used in the framing is essential for computing the load capacity; therefore, testing should be performed to ascertain and verify the actual yield strength. One technique is to test the steel to determine its actual yield strength, in hope of finding it to be a higher value than the one that was used in the original design. Another technique applied in existing structures is to analyze the framing using the load and resistance factor design (LRFD) method (AISC 1999); in LRFD usable strength is approximately 1.5 times greater than the older allowable stress design (ASD) service level strength.

Increasing Capacity of Connections

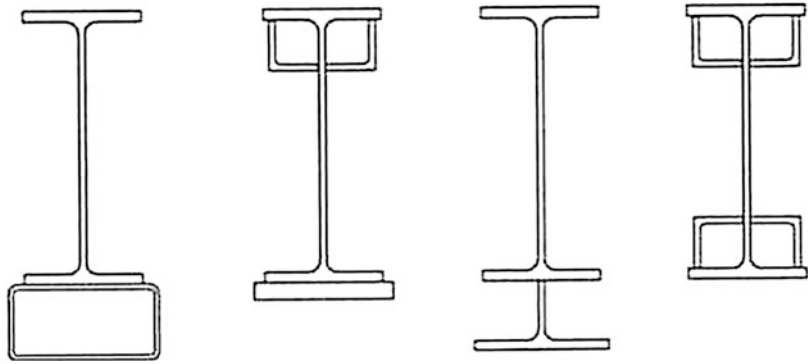
The technique by which existing shear and moment connections can be strengthened is limited only by the imagination of the engineer. Various techniques are going to be presented thereafter, based on well-established but also recent research outcomes obtained from numerous computational analyses and experimental campaigns. It is worth to be aforementioned that the capacities of existing connections must be determined when existing framing is modified or additional capacity is sought.

Increasing Flexural Strength Floor Framing Members

There are two options for reinforcing existing flooring systems to support additional loads: (a) **add** new framing to supplement the existing framing and (b) **reinforce** the existing beams, girders, and connections. The easiest solution is usually that of reinforcing the existing structural

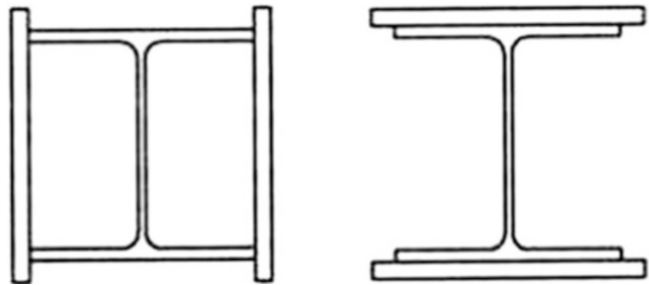
**Strengthening Techniques:
Code-Deficient Steel Buildings,**

Fig. 1 Examples of strengthened beams



**Strengthening Techniques:
Code-Deficient Steel Buildings,**

Fig. 2 Examples of strengthened columns



elements, provided that the floor slab has sufficient capacity to carry the loads. The most efficient way is to weld rectangular high-strength steels (HSS) to the flanges as shown in Fig. 1.

Increasing Axial Load Capacity of Columns

The buckling limit state and its variable slenderness should be determined in order to evaluate the axial load capacity of columns. Column strengthening serves both to reduce slenderness by increasing the radius of gyration of the section and to reduce stress. Column buckling is a mid-height phenomenon; therefore, increasing column stiffness between the supports, not at the supports, is required to increase column capacity. Both methods shown in Fig. 2 are effective; however, the one on the left better increases the weak axis stiffness of an H-shaped section.

Dealing with Weldability Issues

Weldability is verified by mechanical and chemical testing. The former measures ductility and the latter determines the “carbon equivalent” value.

Connecting New Frame to Existing Frame

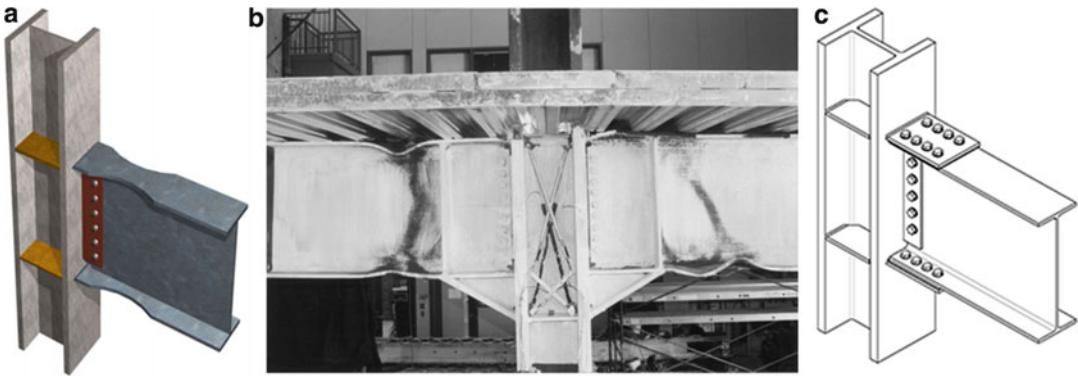
Similarly to the connections, there are numerous ways that new framing can be connected to existing framing. Welding the new steel members to the existing members is a straightforward approach which requires less precision as compared to the bolting process, while drilling new holes through existing steel and bolting in the field. Various details for connecting new framing to an existing one can be found by Schwinger (2014).

Detailed Description of Retrofitting and Strengthening Techniques

Introduction

The performance of steel frames can be synthesized in three very different behaviors:

1. Formation of plastic hinges
2. Local and global instabilities
3. Fracture and structural discontinuity



Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 3 (a) Reduced beam section (RBS) connection (Crawford 2002), (b) welded haunch connection (Uang et al 2000), (c) bolted brackets

These three behaviors and/or combinations of them are likely to occur and govern the capacity of a connection or member with result on the structural continuity and integrity of the system. Overall, it is known from seismic studies that frame capacity is related to two different aspects of frame behavior:

1. The **member** response, as controlled by plastic rotational strength and deformation characteristics (including local and global buckling)
2. The **connection** response, as controlled by bolt fracture, premature brittle weld failure, and panel zone failure

Determining the capacity of columns is difficult as in many situations code-deficient buildings are not designed for large lateral loading.

Steel Connections: Fuses

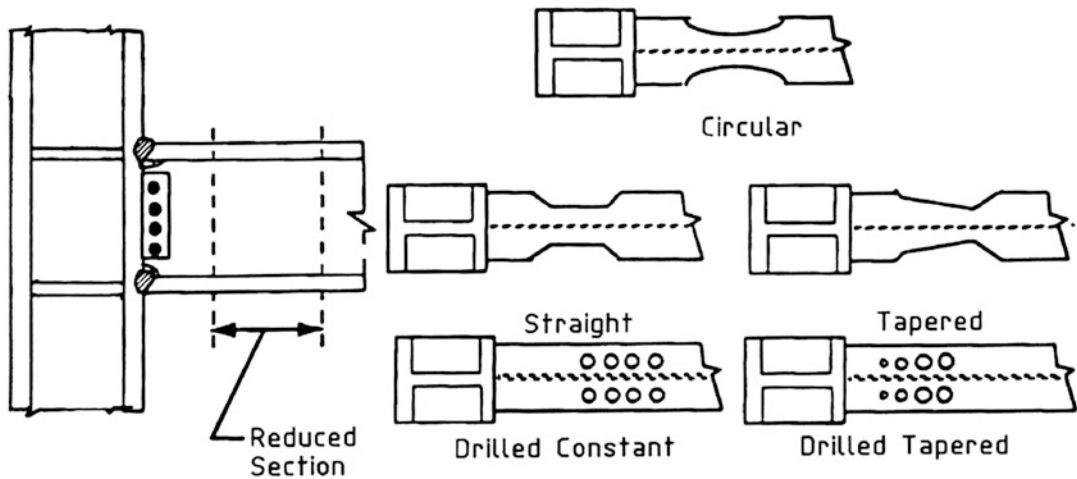
Beam-to-Column Connections: Developing Ductile Behavior (Fuse Concept)

In parallel with the FEMA/SAC steel research program, the National Institute of Standards and Technology (NIST) and the AISC initiated a research project to upgrade existing special moment frames (SMFs) and investigate the effectiveness of two rehabilitation schemes. Modifications to pre-Northridge moment connections to achieve improved seismic performance focused on reducing or eliminating some of the

contribution factors to the brittle fractures. Brittle fractures originated in the beam flange groove welds and often propagated to rupture beam flanges or columns. A cooperative effort by NISC, AISC, the University of California at San Diego, the University of Texas at Austin, and Lehigh University examined three techniques for the retrofit of existing code-deficient steel moment connections trying to force plastic hinging of the beam away from the column face, namely, (a) the reduced beam section (RBS) concept to weaken a portion of the beam near the column so that plastic hinging would occur at the designated location, (b) the addition of a welded haunch to strengthen the steel beam near the welded connection, and (c) the use of bolted brackets to reinforce the connection (Fig. 3). More RBS patterns have been developed by Plumier in 1997 and appeared in different configurations, as it is shown in Fig. 4.

Further analytical research on the same connection complement this work, as design model and guidelines have been recommended. A target plastic rotation capacity of 0.02 rad was selected. In 2004, Engelhardt recommended (Bruneau 2004) the following as potential positive solutions:

- The use of a bottom flange RBS combined with the replacement of top and bottom beam flange groove welds with high toughness weld metal provided plastic rotations on the order of 0.02 to 0.025 radian. The presence of a



Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 4 Various RBS patterns

composite slab had little effect on the performance of this retrofit technique.

- The addition of a welded bottom haunch, with the existing low toughness beam flange groove welds left in-place, resulted in significantly improved connection performance, which was dramatically influenced by the absence or presence of a composite concrete floor slab. In the former case, the specimens developed plastic rotations of 0.015 to 0.025 radian, whereas in the latter case developed plastic hinges in excess of 0.03 radian.
- The use of bolted brackets at the top and bottom flanges provided plastic rotations in excess of 0.03 radian.

Design recommendations have been provided for fully restrained, radius cut RBS connections. A step-by-step procedure is presented, with commentary for various design considerations. A similar procedure is included in FEMA 351 (2000) and AISC 358 (2005), which also provides design guidelines for other prequalified post-Northridge connections such as the bolted flange plate (BFP) moment connections, the bolted unstiffened (BUEEP) and stiffened extended end plate (BSEEP) moment connections, and the so-called ConXtech ConXL and Kaiser bolted bracket (KBB) moment connections.

It is worth to note that conventional beam theory cannot provide a reliable prediction for neither of the above structural systems. Uang et al. (2000) and Yu et al. (2000) proposed a simplified model that considers the interaction of forces and deformation compatibility between the beam and the haunches.

Exhaustive research works have been conducted on RBS connections varying the geometric characteristics of both the beam and the connection assembly itself. More recently, RBS moment-resisting connections have been also investigated by researchers in Europe using European HEA-profile sections (Pachoumis et al. 2008), since so far they have been only investigated by the US design construction practices. The result is the readjustment of the geometric characteristics of the RBS in order to apply to the European profiles. Limitation in using RBS is the shear connection between the top steel flange and the metal decking of the steel-concrete composite (SCC) slab due to the significant width reduction.

More recently, the same concept has been applied to steel frames as a strengthening-weakening technique, while introducing a circular opening (Fig. 5) in the beam's web instead, at a certain distance from the beam-to-column connection, as an effective method to improve the



Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 5 Failure mode of connection with circular web opening (Qingshan et al. 2009)

aseismic behavior of MRFs (Qingshan et al. 2009). The accurate position and size of the circular opening has been investigated through numerical modeling as well as experimental works, while the plastic hinge positions is effectively controlled. Similar studies have demonstrated the effect of various nonstandard web opening shapes (Fig. 6), in enhancing the ductility but also the strength of the connections (Tsavdaridis et al. 2014). Step-by-step procedures have been proposed to determine the most suitable geometries to achieve adequate connection strength, ductility, stiffness, and rotational capacities. Such techniques prove suitable in cases where large plastic rotations are required (i.e., larger than 0.03 rad). Tsavdaridis et al. (2014) have further proposed the use of previously patented novel elliptically based web opening shapes, which can also be used for perforated beams (e.g., cellular and castellated beams) adding numerous advantages from the manufacturing process to their life-span, while they can develop rotational capacities up to 0.05 rad with insignificant strength degradation (Fig. 7).

The so-called reduced web section (RWS) connections have been studied, yet not extensively, assessing local connection models mainly computationally with cyclic (quasi-static), and pseudo-dynamic (PSD) analyses. The results

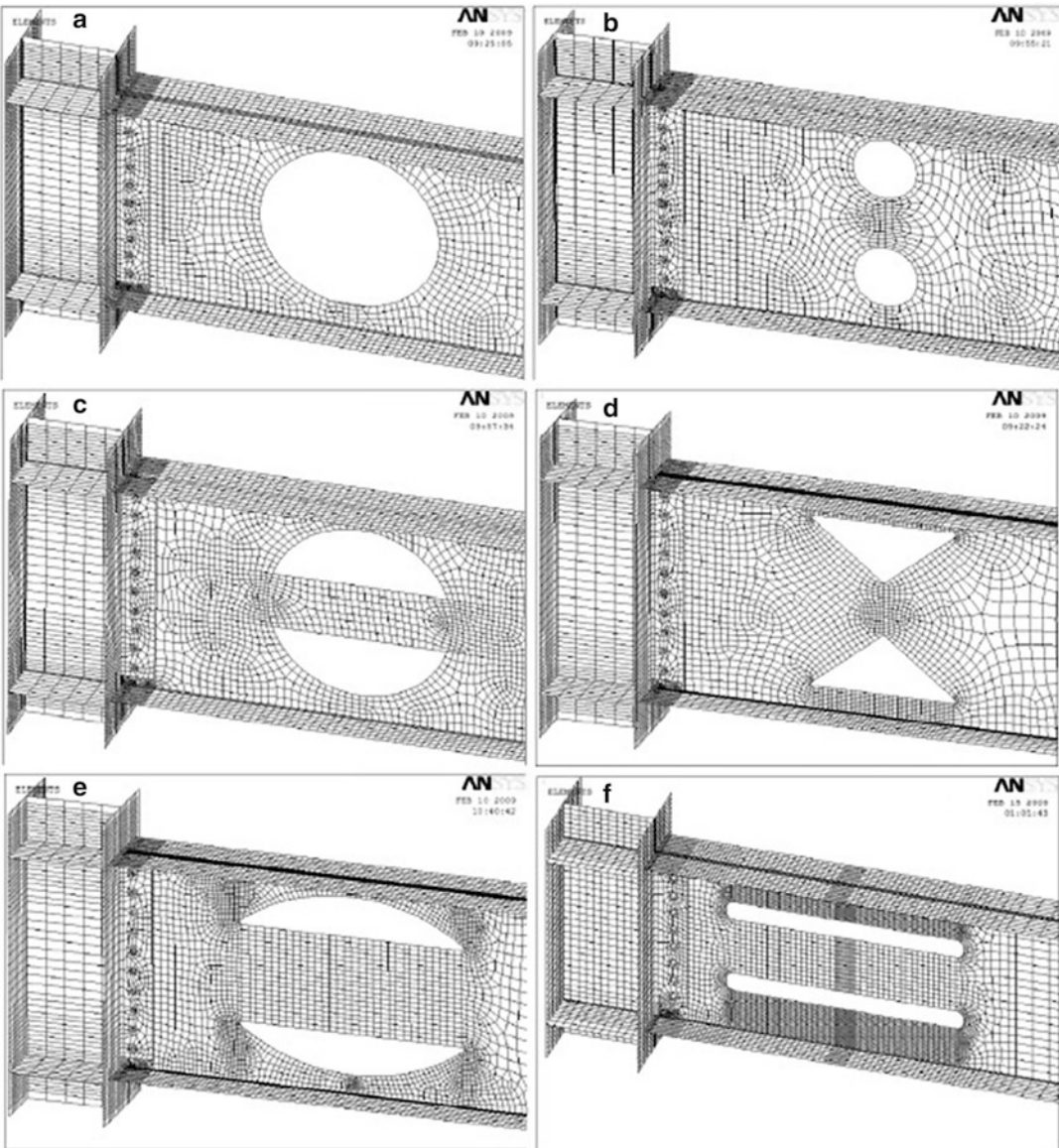
show that the ultimate displacement of the modified buildings increases a lot due to the web openings, and thus, the building ductility is improved greatly. Moreover, brittle weld fractures can be avoided and the maximum plastic zone moved to the weakened areas. RWS connections can easily be applied to the beams of new as well as existing code-deficient buildings as the web cutting does is away from the composite slab that sits on top of the compression steel flange. It is worth to mention that different geometric characteristics and limitations of beams and columns should be used for different RWS connection types. There is a need to bring the attention and propose more experiment physical testings to validate and establish RWS connections in the current European and American practices.

Similarly, welded haunches and bolted brackets are used to move the plastic hinges away from the column, but also strengthen the existing connection and seek to maintain the original flexural capacity of the beam.

Ductile Behavior: Fuses in Bracing Members

The concept of adding ductile fuses in bracing members of steel concentrically braced frames (CBFs) resisting seismic loads is well linked to the RBS technique. Current code provisions require that steel CBFs are designed to exhibit ductile energy dissipation. Limits on brace overall slenderness ratio must be satisfied to achieve ductile inelastic behavior. It is apparent that implementation of this design approach may result in significant increases in design loads for brace connections, beams, and columns (Egloff et al. 2012).

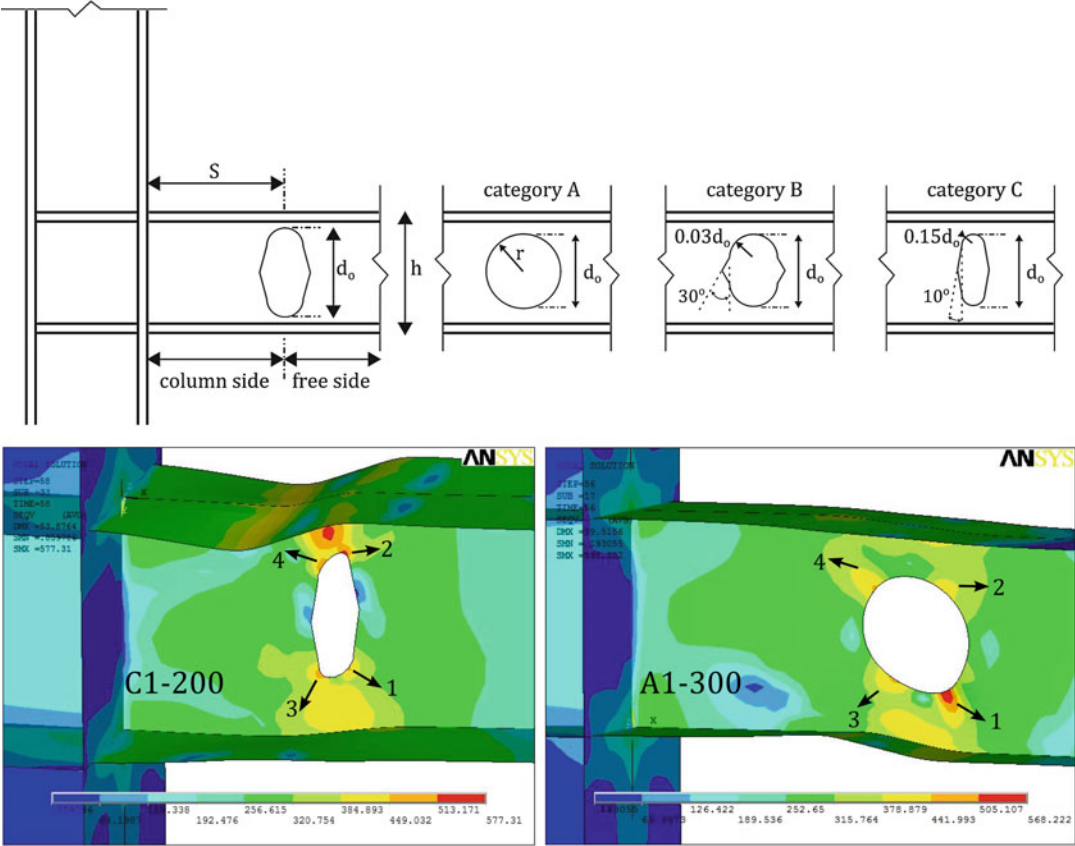
In order to reduce seismic design loads, ductile fuses in bracing members have been recently proposed, as they control their axial resistances. Such behavior can be achieved by locally reducing the brace cross-sectional area or by introducing ductile components that yield in both tension and compression. In the former case, the reduced section might need to be confined to prevent local buckling, or it can be resized (reshaped - restructured) to yield in tension while remaining elastic in compression, which means that the



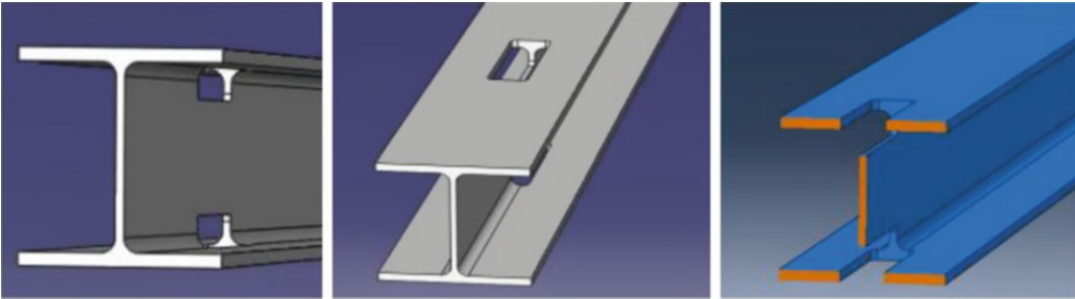
Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 6 Types of perforated beam webs (Hedayat and Celikag 2009)

effect of fatigue loading will be minimized. In the latter case, the overall buckling is eliminated, and hence, the strength degradation is limited due to symmetrical hysteretic behavior. This type of fuse technique can be applied in open and closed profile sections of the bracing members, while it has been noticed that the former ones perform better exhibiting higher ductility. A special fuse (Fig. 8) for controlling the tension resistance of

open bracing members has been proposed by Vincent in 2008. A part of the flange to web intersection is removed to limit the impact on the brace flexural stiffness and buckling resistance. In 2012, Egloff et al. introduced a new local buckling restraining system (LBRS) which includes two cold-formed channels that support the web and the flanges. Moreover, external cover plates can be bolted to the channels in order to



Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 7 Geometric parameters of various novel perforated beam-to-column moment connections and von Mises plastic stresses (Tsavdaridis et al. 2014)

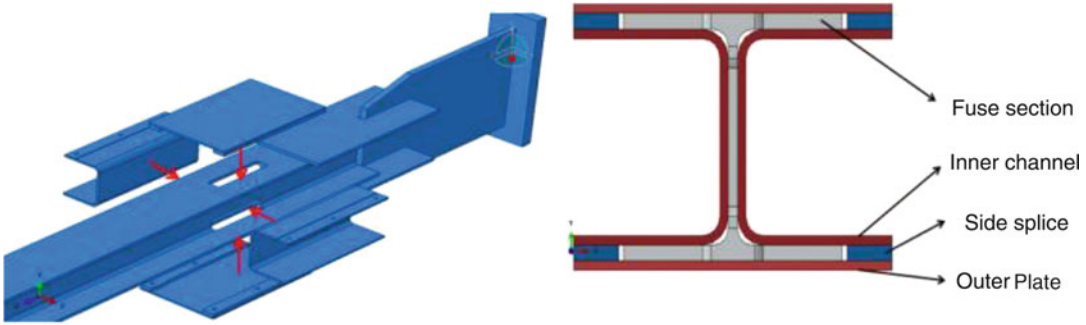


Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 8 Ductile fuse for H-shaped bracing members (Vincent 2008)

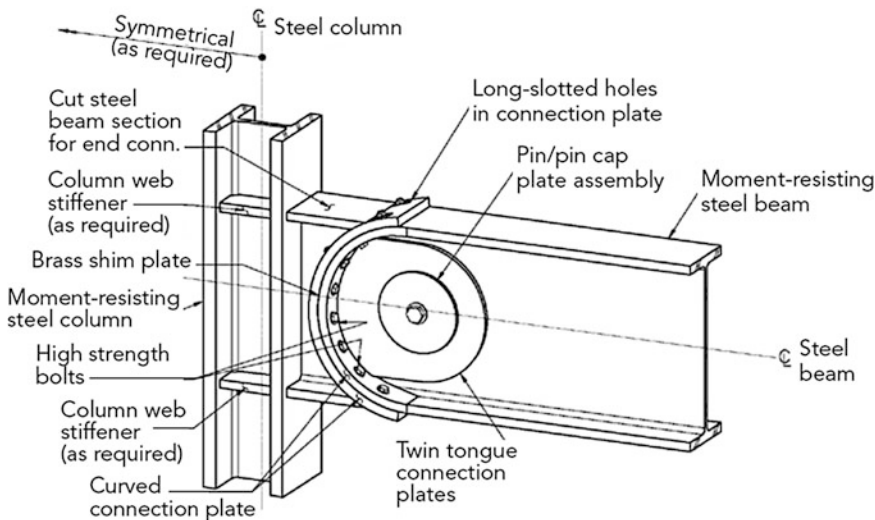
prevent local buckling of the brace flanges. Splice plates can also be used to provide lateral support to the flanges in the fuse (Fig. 9). This LBRS can slip longitudinally with respect to the brace so that it does not attract any axial forces. Further

improvements and the design procedure for the fuse and the fuse LBRS are available in the literature.

Cast Connex, a high-strength steel connector for round hollow structural section brace



Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 9 Proposed fuse local buckling restraining mechanism (Egloff et al. 2012)



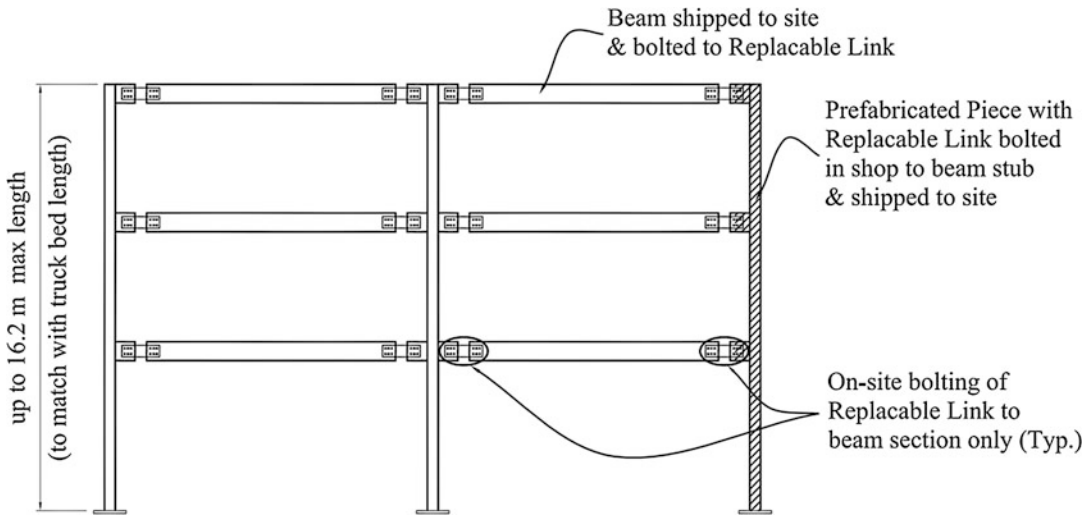
Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 10 Structural details of pin-fuse connection (Coedova and Hamburger 2011)

members in CBF, has also developed a yielding fuse connector for CBF, called the Scorpion Yielding Brace System (SCBF), that relies on the flexural yielding of fingerlike plates that are specially designed to dissipate energy locally and can be used in both architecturally exposed and nonexposed braced bays.

Pin-Fuse Joints

A new type of connection, which begun its prequalification process in 2011, is the pin-fuse connection (Fig. 10) which incorporates a curved plated end connection using slip-critical bolts and a steel pin adjacent to the beam-to-

column joint. The bolts are designed to slip within slotted holes allowing the pin joint to rotate, dissipating the energy through frictional resistance. This joint acts as the fuse for the system, while the rest of the steel frame can be designed to remain elastic. Following an earthquake, and avoiding damages, the frame can be adjusted to its initial position, reducing the potential for permanent residual displacements while both the connection and the frame maintain their structural integrity and reduce the need for costly structural repairs. The simplicity of the pin-fuse connection offers the ability to accommodate braces and dampers.



Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 11 Proposed connections with replaceable link (Shen 2009)

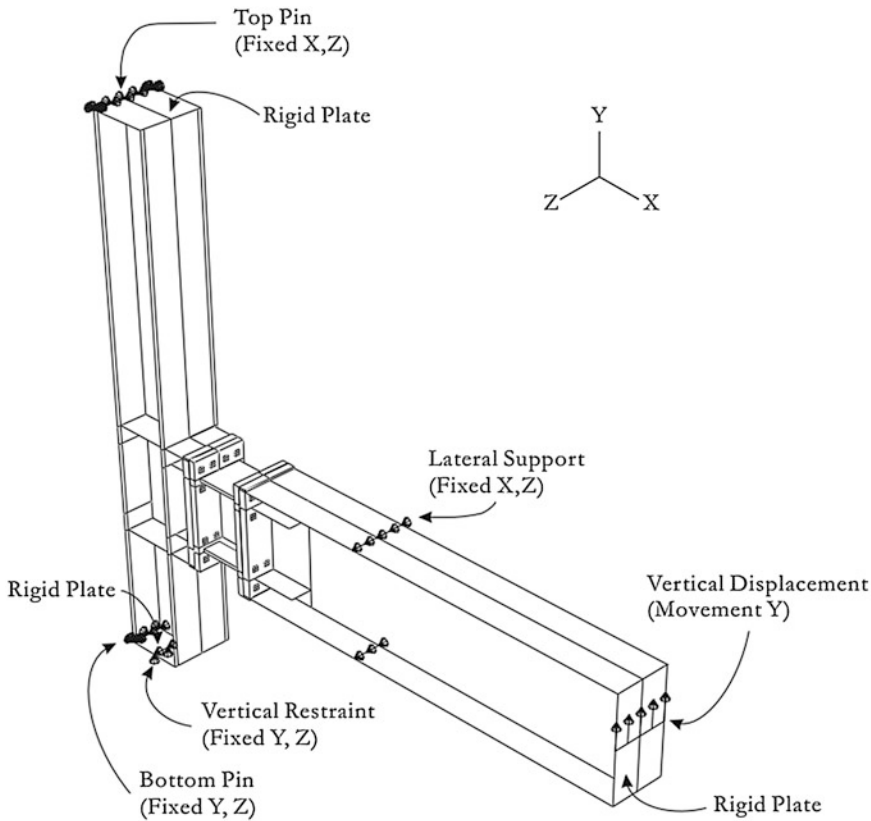
Replaceable Links

Using RBS- and RWS-type connections has been proved very efficient for certain applications; however, there are some drawbacks. As the yielding fuse is a part of the beam, strength design and drift design of the structure are interlinked. For instance, due to increased drift requirements, the capacity of the yielding fuse may be also increased, which then leads to higher demands on the other parts of the structure including columns, floor slabs, connections, and foundations and often resulting in oversized buildings with increased overall cost. Further, significant damage can result in the beam from repeated inelastic deformation and localized buckling (hence crack propagation) during a design-level earthquake. As this cumulative inelastic action of the building cannot be precisely anticipated, it is not trivial to assess the extent of damage on site and the residual capability of the structure to adequately provide the required level of safety for any subsequent loading. In such a case, repair of the beam is not a straightforward procedure and it can be disruptive and costly.

The replaceable link concept (Fig. 11) effectively eliminates the aforementioned concerns instead of reducing the beam section size; dismantlable dissipative elements (Fig. 12)

which can be removed and replaced with smaller flexural capacities are used at the locations of the expected inelastic actions. Consequently, the other structural elements in the frame will remain elastic during an earthquake. This efficient method of repair for MRFs allows for quick inspection and replacement of damaged links while it minimizes the disruption time. Further, the welding of critical elements of beam-to-column connections can be done in the shop while improving construction quality and significantly reducing the initial erection time (Shen 2009).

In particular, two types of replaceable links have been proposed by Shen (2009): (i) H-section with end plates and (ii) back-to-back channels eccentrically bolted to the beam web. The former one is prepared in the shop using complete joint penetration welds. The end plate (flush or extended) is then bolted to the column flange using prestressed high-strength bolts. The latter type of these double channel-built sections intended to act as truss girders in special truss moment frames, and they have been connected using welded reinforcing gusset plates. In certain circumstances, lateral bracing is deemed necessary in the region adjacent to the plastic hinge to achieve large (i.e., 0.06 rad) plastic rotation of the hinge. However, sometimes



Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 12 End plate model boundary conditions (Shen 2009)

large overstrength has been observed in these buildup channel sections.

The end plate links can exhibit 0.04 rad contributing to 90 % of the total story drift and demonstrate higher energy-dissipating capacity than the double channel links, while some strength degradation occurs due to ductile local flange and web buckling. On the other hand, higher story drift can be reached by the double channel links before experiencing strength degradation at 0.06 rad. The degradation has been also caused due to ductile tearing of the flanges and the webs. Overall, double channel links type has been considered to be preferable as it provides a more gradual transmission of forces at the connections via friction. For more stability, further modifications can take place enhancing the connection of the channel webs and the connection of the beam segments.

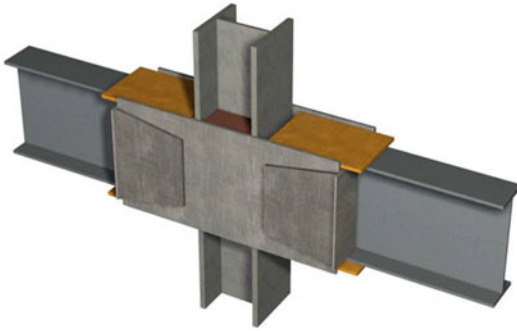
Steel Connections: Stiffeners

Introduction

Forcing the plastic deformation to the beam end away from the connection is a common practice in seismic moment-resisting frames, but in contrast to the fuse concept, this can be also achieved by increasing the relative stiffness of the column and the connection with respect to the beam. Eventually, this is an alternative in effectively controlling the position and intensity of the plastic hinge in the connection zone when such modifications are allowed.

SidePlate™ Connections

A well-promising retrofitting method for upgrading an existing traditional moment-resisting connection is shown in Fig. 13. This concept uses the so-called SidePlate™ retrofit system where the

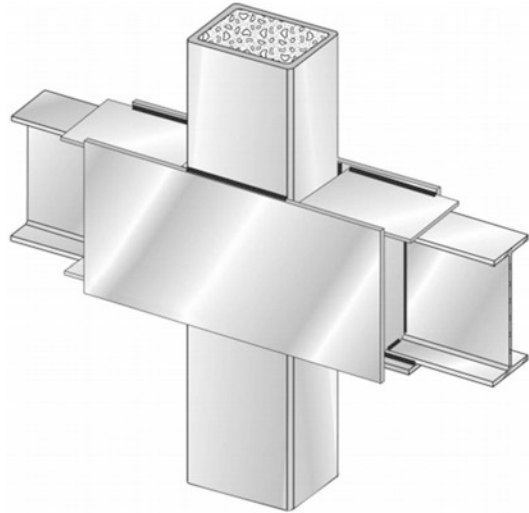


Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 13 SidePlate™ retrofit connection (Crawford 2002)

physical separation between the face of the column flange and the end of the beam, for mitigating the stress concentration, is achieved using parallel full-depth side plates which act as discrete continuity elements to sandwich the connecting beam(s) and column (Crawford 2002). Similar design concept can be used for steel and concrete-filled hollow section columns (Fig. 14).

Whole steel frame is eventually stiffened, and the zone panel deformation is eliminated using this type of connections due to the increased stiffness of the side plates that ultimately provide the three panel zones. This connection system uses all fillet-welded fabrication which predominately carries all shear actions as well as moments through the combination of vertical shear plates and fillet welds. The side plates should be designed with sufficient strength and stiffness to force all significant plastic behavior of the connection system into the beam.

The same system can be used to upgrade construction of deficient buildings. The difference is that an initial hole is required in each side of the plates to permit welding access, while the holes are closed with the same cutoff plate following the completion of the welding process. All new welds are again fillet welds loaded in shear along their length, whereas if there are any existing Complete Joint Penetration (CJP) welds, they are removed by air arcing to eliminate the reliance on through thickness properties and triaxial stress concentrations. More information can be found from FEMA 351 (2000).

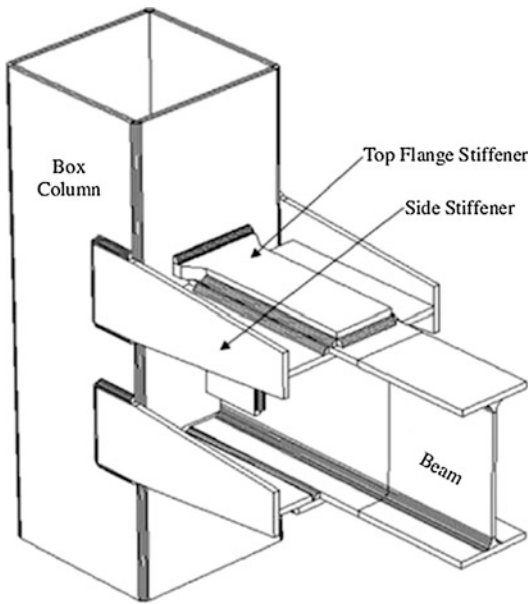


Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 14 Strengthening retrofit concept: concrete-filled hollow section of column (Crawford 2002)

Stiffeners at Connections

In addition to the prequalified connections for SMFs and intermediate moment frames (IMF) presented in AISC (1999) and the SidePlate™ system, research has been focused on the effect of stiffeners on the strain patterns of the welded connection zone.

For example, the effect of both internal and external stiffeners on the behavior of I-beam to hollow-column section connections (Fig. 15) has been initially thoroughly investigated by Chen and Lin (1990). It has been observed that the connections with triangular stiffeners have the lowest rigidity, in contrast to those with side stiffeners which present significantly higher moment rotation capacities, stiffness, and ductility. Moreover, the performance of the retrofitted connections with side stiffeners has been investigated (Ghobadi et al. 2009) and design guidelines proposed. The benefits of using side stiffeners have been also introduced on concrete-filled tubular (CFT) columns connected to I-beams, while stable hysteresis and adequate ductility are provided. Overall, it has been concluded that connections with both column stiffeners and top-flange stiffeners have the highest value of energy dissipation, while the beam top-flange



Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 15 Typical I-beam to box-column connection (Kiamanesh et al. 2010)

stiffener is the most effective one, especially when it is incorporated with the column stiffeners of a hollow section.

Steel Frames: Modifications

Frame Modification at Beam's Mid-Span (Fuse Concept)

A retrofitting method which can be used for new construction as well as a strengthening technique for existing moment-resisting frames has been developed by Leelataviwat et al. (1998). This technique replaces certain beams and is introducing a ductile fuse element in shear at their mid-span instead of modifying the beam-to-column connections (Fig. 16). A braced rectangular opening is created in the web of each girder at the mid-span, to move the plastic deformation away from the critical connection regions, while ensuring the development of a ductile mechanism.

Cabling: Self-Centering Systems

The use of the cables is another promising technique, which can be applied to both slabs and connections. Placement of cables with connections

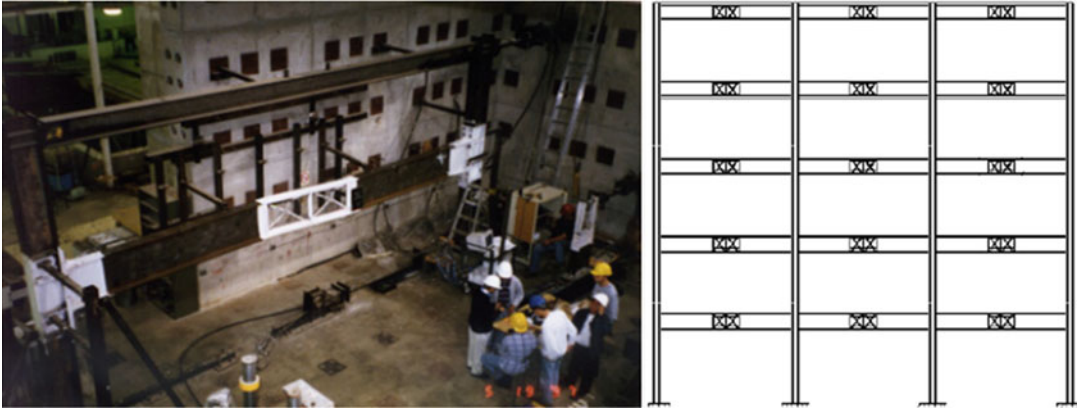
to girders or cables with connections to beams is very important especially for high-rise buildings. Self-centering braces have been designed and built using prestressed aramid fiber strands in conjunction with friction pads or memory alloys. Energy dissipation is implemented using yielding seat angles, friction dampers, or energy-dissipating bars confined in tubes. Researchers have investigated self-centering column bases that use post-tensioned (PT) bars or spring-loaded wedges. Tendons can span over multiple floors, while elastomeric spring dampers and fuse bars can be used to provide energy dissipation.

Furthermore, self-centering structural systems (Fig. 17) have been proposed, for the seismic retrofit of special moment-resisting frames, by Christopoulos et al. (2002). This is a post-tensioned energy-dissipating (PTED) steel frame design, where high-strength bars or tendons provide the post-tension at each floor. Confining steel sleeves have been often used to prevent the energy-dissipating bars from buckling during cyclic inelastic loading. It has been concluded that these economical innovative systems:

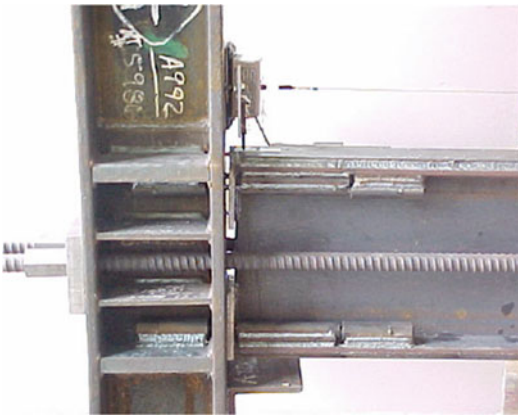
- Incorporate the nonlinear characteristics of yielding structures and, thereby, limit the induced seismic forces and provide additional damping characteristics.
- Encompass self-centering properties allowing the structural system to return to its original position after an earthquake.
- Reduce or eliminate cumulative damage to the main structural elements.

Later, Garlock et al. (2004) has proposed a similar structural system with high-strength steel PT strands, after bolted replaceable top-and-seat angles have been installed (Fig. 18). Here, the vertical shear is supported by both the angles and the friction between the beam and the column, and it is expected to continue to perform if failure of one or more strands occurs. It is proved that this system can achieve greater strength and ductility.

Recently, researchers have designed and experimentally evaluated a new self-centering PT connection using yielding web hourglass shape pins (WHPs) as seismic energy dissipaters. WHPs do



Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 16 Frame modified with mid-span truss opening (Leelataviwat et al. 1998)



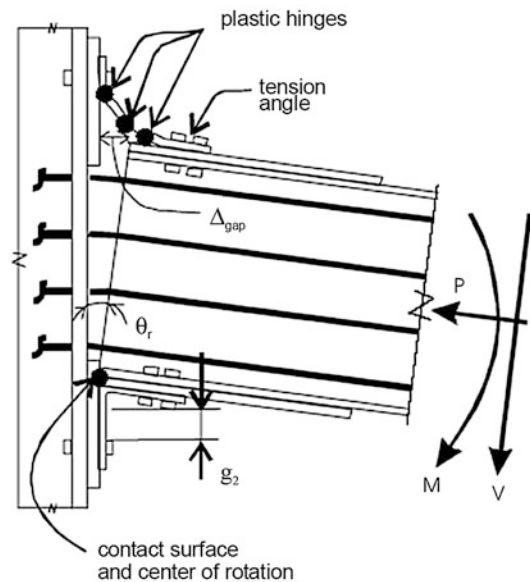
Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 17 PTED system (Christopoulos et al. 2002)

not interfere with the composite slab and can be very easily replaced without the need for welding and bolting and, therefore, can significantly decrease downtime in the aftermath of a strong earthquake. Repeated experiments are described in detail and proved the reparability of the PT connection with WHPs (Vasdravellis et al. 2013).

Structural System: Adding Structural Elements (Walls, Blocks, Bracings)

Introduction

Conventional retrofitting methods include the addition of new structural elements to the system,



Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 18 Post-tensioned moment connection with top and bottom seat angles (Garlock et al. 2004)

enlarging the existing members. Bracings, masonry (Fig. 19), and reinforced concrete (pre- and post-cast infill) shear walls are the most popular and efficient strengthening techniques as they provide lower overall cost and they are easy to use. Braces are more effective due to their much higher ductility, but the shear walls are indeed the most commonly applied method, as they also reduce the demand on the other



Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 19 Ordinary hybrid wall (Adopted by <http://masonryedge.com/site/mim-archives-the-storypole-vol-38-no-4/143-filling-in-the-frame->)

structural members resisting large lateral loads, hence increasing their safety in a simple manner. The actual capacity of bearing walls has been often underestimated or even ignored. However, it can be a major contributor, and it can provide the required capacity, without the need of more complex strengthening techniques (Sarhosis et al. 2014).

Bearing Walls

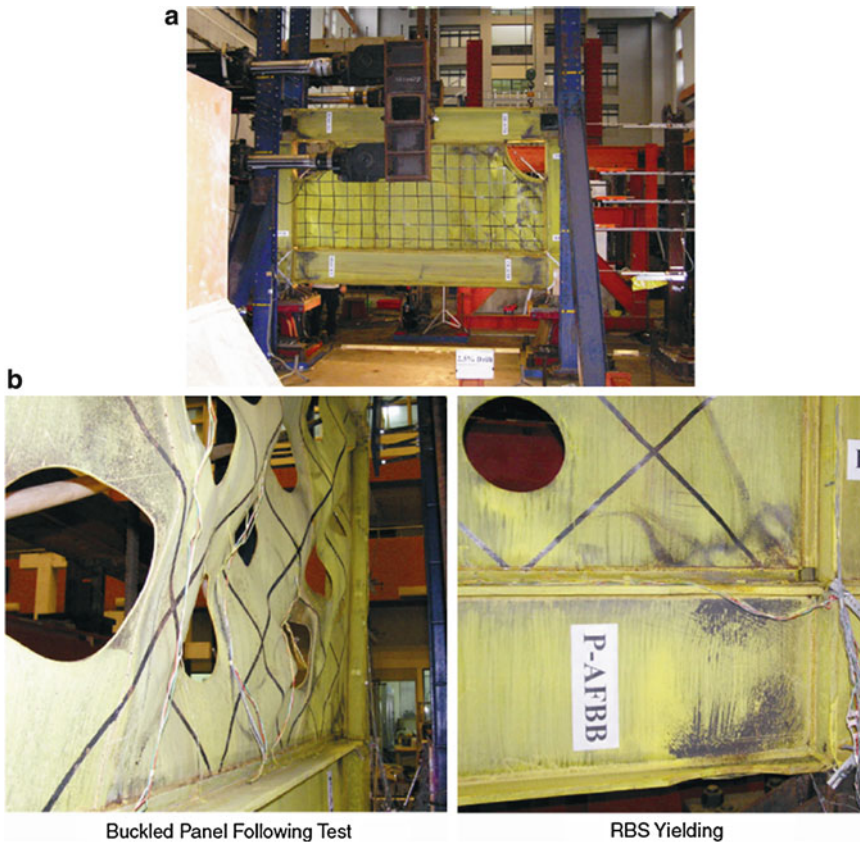
Walls must go equally in both directions, and they must be strong enough to add stiffness to the steel framing system while they are tied in to any framing in order to take load in their weakest direction. Also, they must not fall apart and must remain in place after the worst shock waves, so as to retain strength for the aftershocks.

In particular, three approaches have been identified (Crawford 2002) for enhancing the resilience of a building's bearing walls under the progressive collapse and seismic scenarios. These are the following: (i) backup wall, build a second wall or gravity-carrying frame to support the existing wall; (ii) strong wall, employ fabric retrofit to control the breach area; and (iii) ductile

wall, polyurethane spray to prevent punching shear failure. Moreover, openings can be accommodated in such walls, especially with multistory buildings, while further enhancements are needed.

Steel Plate Shear Walls

In addition to concrete and reinforced concrete walls with SCC beams, research has been initially conducted by Thorburn et al. (1983) later on steel plate shear wall (SPSW) design (Fig. 20a) and retrofitting methods. SPSWs can be used as the primary lateral force resisting system in steel buildings allowing the occurrence of shear buckling. Following buckling, diagonal tension field is developed to transfer the lateral load in the panel, while the forces in beams and columns are reduced. Furthermore, the use of low-yield-strength steel panels and RBS connections as well as light-gauge cold-formed steel plates has been examined as potential applications by Berman and Bruneau (2004). The former one demonstrates an earlier onset of energy dissipation by the panel, while perforated panel specimens (Fig. 20b) can be used to control the



Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 20 (a) SPSW specimen with cutout corners (*right*), (b) buckled panel and RBS yielding of SRW specimen (Vian and Bruneau 2004)

stiffness and overstrength issues using hot-rolled plates. This option is also useful in a retrofit situation, providing access for utilities to penetrate the predesigned system.

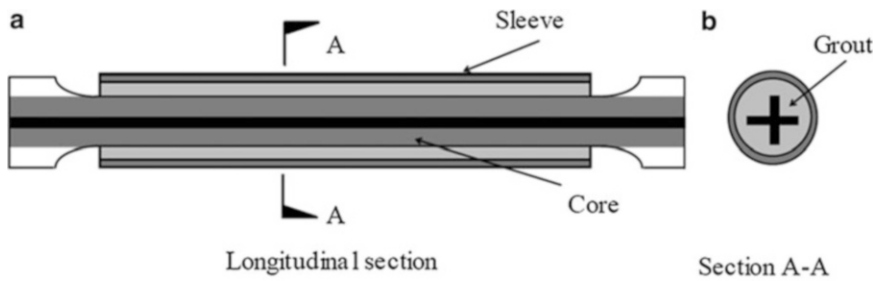
Braced Frames

Braced frames can be constructed of single diagonal, x and k braces, chevron and split braces, and lattice or knee braces, and they can be used in interior cores – so connections could be easily made with wall panels, as well as in the exterior. Composite braced frames are also becoming popular where concrete bracings are supporting steel frames.

A simplified design procedure for suspended zipper frames, initially proposed by Khatib et al. (1988), has been introduced by Leon and Yang (2003) and consists of inverted V-braces

adding zipper columns which connect the intersection point of the braces above the first floor. The zipper columns tie all brace-to-beam intersection points together and force compression braces in a braced bay to buckle simultaneously, hence better distribute the dissipated energy over the height of the building (Bruneau 2004). A suspension system has been proposed later, ensuring that the top story braces are designed to remain elastic, whereas all other compression braces are designed to buckle, while the suspended zipper struts are designed to yield in tension. Therefore, adequate ductility is provided, with superior seismic performance.

Engaging the fuse, the replaceable links, the self-centering, and the braced frame retrofitting concept, Eatherton et al. (2008) proposed a controlled rocking system which virtually eliminates



Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 21 A non-buckling brace

residual drifts and concentrates the majority of structural damage in replaceable fuse elements. The system is consisted of three components: (a) a stiff steel braced frame which remains elastic, but it is not tied to the foundation and hence allowed to rock, (b) a vertical post-tensioning that strands the top of the frame down to the foundation and brings the frame back to the center, and (c) the replaceable fuses that absorb the energy as the frame rocks.

Non-Buckling Braces

Conventional braces tend to buckle under the compression cycle of the seismic load, hence dissipating little energy under compression. This causes pinching of the hysteresis loop and failure of the braces within a few cycles, due to the formation of plastic hinge close to mid-length of the member. The use of non-buckling braces (also known as buckling restrained braces or un-bonded braces) bypasses this problem. In this type of bracing system, the requirements of adequate strength to resist compression, as well as rigidity to avoid buckling, have been addressed separately to a core and a sleeve (Fig. 21).

In the last decade, buckling restrained braced (BRB) frames have received much attention in the United States, as they demonstrate stable hysteretic behavior and excellent low-cycle fatigue life characteristics. However, buckling and cracking of gusset plates are expected in certain cases, similarly to all types of braced frames.

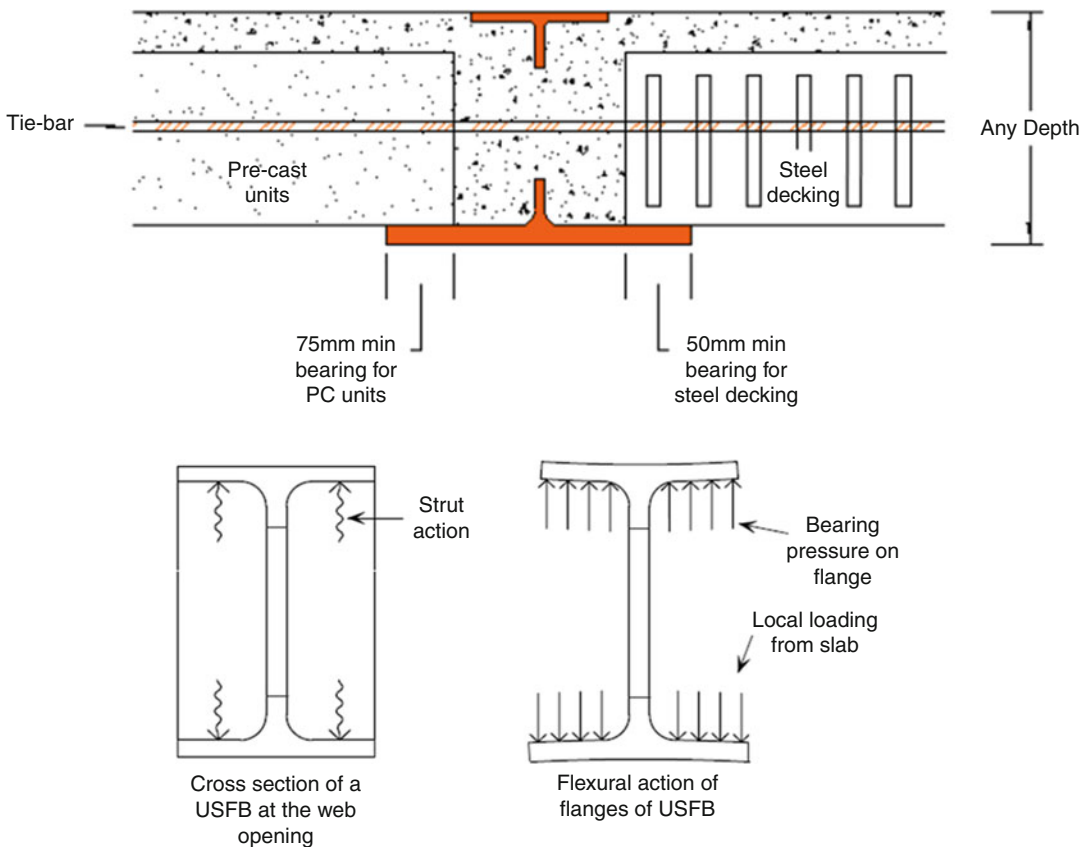
Strengthening Members

Strengthening members by adding plates or encasing/filling them with concrete provides an

effective technique to add strength, and it can be applied for a particular group of members, such as on the ground floor. Thus, concrete encasement of columns (Figs. 14), and floor beams (Fig. 22) as proposed (Tsavdaridis et al. 2013), constitute forms of strengthening techniques for new and existing steel buildings.

Materials

Innovative ways have been explored for the strengthening and rehabilitation of deficient steel buildings, due to the demand to access the specified load capacity and the deterioration as a result of corrosion. In particular, externally bonded fiber-reinforced polymer (FRP) composites can be applied to various structural members such as columns, beams, slabs, and walls in order to improve their structural performance in terms of stiffness, load-carrying and deformation capacity, and ductility, while simultaneously providing environmental durability. Generally, FRPs have been widely used mostly in applications that allow complete wrapping of the member, while attention deemed necessary to avoid brittle shear and de-bonding failures, especially prone when used on steel. In such cases, the actual member can entirely waste the strengthening application, or the composite material might harm the member itself by decreasing its ductility (Buyukozturk et al. 2003). It has been proposed that for buildings with large seismic deficiencies, a combination of conventional and FRP strengthening techniques may prove to be an effective retrofitting solution.



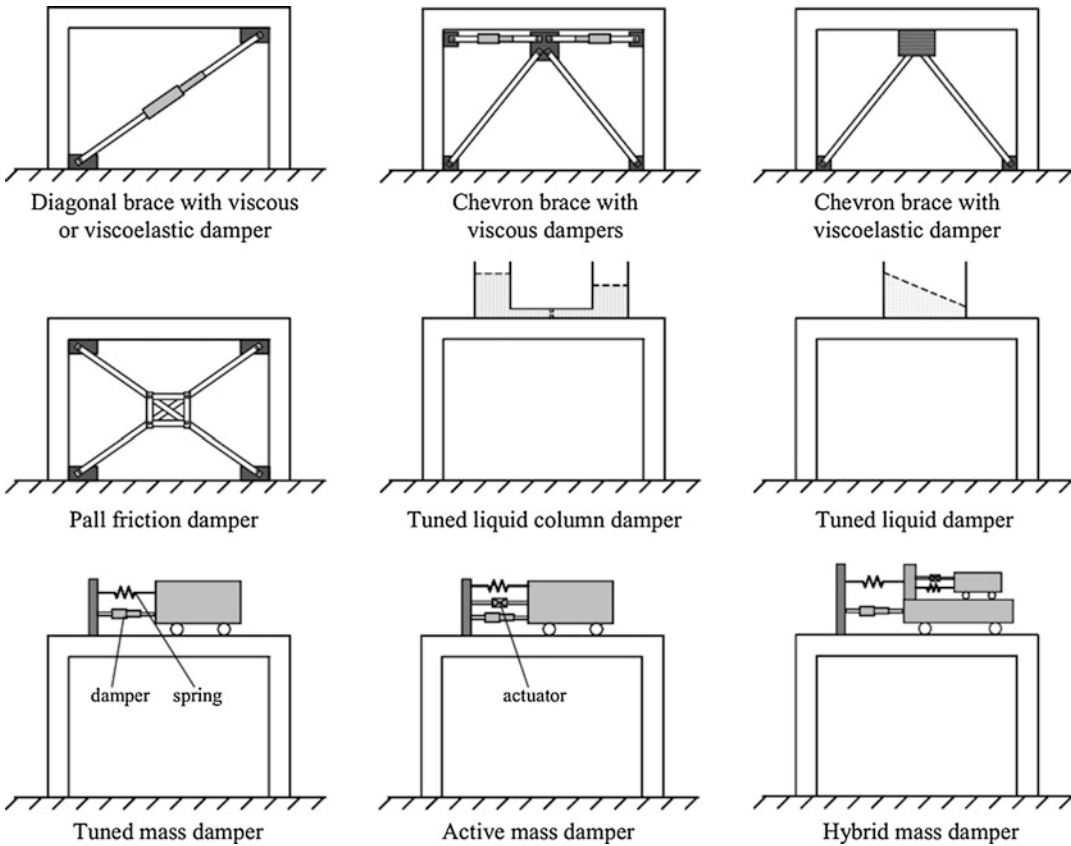
Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 22 Schematic representation of the USFB system and the internal actions (Tsavdaridis et al. 2013)

The preformed high-strength carbon fiber plates currently being used for concrete structures are typically 4 mm thick. To strengthen steel beams, they would need to be at least 20 mm thick, in order to achieve a significant increase in bending moment for the steel or SCC beam. Consequently, new high-modulus carbon fiber-reinforced polymer (CFRP) materials are likely to provide solutions for steel structures' deterioration issue (Schnierch et al. 2005). Time should be allowed for the surface preparation, application of the adhesive, and curing time (usually between 4 and 8 h). Specific surface preparation and detailing are critical to ensure adequate bond interaction between steel and FRP materials, both in the short and long term, and capable of sustaining the high interfacial

stresses necessary to appreciate the full strength of these materials (Lenwari et al. 2006).

Energy Dissipation and Active/Passive Structural Control Systems

A quite different retrofitting method, which can be quite cost-efficient, is the installation of complementary energy dissipation devices in structures as a means of passive, semi-active, or active structural control systems. These are not described thoroughly here, as it is beyond the scope of the present entry. The main objective of structural control is to minimize structural vibrations improving safety and serviceability limits mainly under wind and earthquake actions. Up to date, the majority of passive energy dissipation devices have been found very effective in



Strengthening Techniques: Code-Deficient Steel Buildings, Fig. 23 Supplemental energy dissipation devices (Adopted from MIT.edu website)

controlling the seismic response of steel frames. Further advanced techniques seem very promising such as the introduction of an “inertor” (Smith 2002; Marian and Giaralis 2014), and its combined use with the already well-operated tuned mass dampers, with scope to reduce the size of the mass required to control and dissipate the energy of high-rise buildings. Figure 23 shows the basic principles of various control systems commonly used on building structures.

There is a vast research conducted on energy dissipation devices during the past 20 years, while their use becomes more direct and apparent with the upsurge of technology. However, a diverse background of researchers is required, integrating a number of disciplines, some of

which are not within the domain of traditional civil engineering. In particular, the control theory is elaborated with computer science, data processing, sensing technology and materials science using the knowledge and principles of earthquake (and wind) engineering, structural dynamics, as well as stochastic processes. It is essential to mention though, that the effectiveness of such dissipation devices is predominantly dependent of the deformation capacity of the structure. Consequently, the application of such devices to code-deficient buildings with inadequate seismic detailing or post-earthquake damages should be carefully considered, and perhaps it should be combined with an appropriate strengthening technique with deformation enhancement measures as proposed above.

Ongoing research on special strengthening techniques involving the use of simple and robust active control systems, while emphasizing on the performance of various controllers (Demetriou et al. 2014) as well as new conceptual methods introducing the ability of structures to adapt under dynamic loads (Slotboom et al. 2014), are posing great expectations.

Summary

It is worth to mention that there are factors which inhibit retrofit design when it comes to strengthening existing buildings. Many times important issues are misaddressed due to the complexity of strengthening design concepts, lack of technology understanding or even ignoring it, while uncertainties about the design of the building involved. Therefore, existing buildings should be best approached with a risk-based retrofit scheme in order to concentrate the works where they are actually needed most. In this way, more safe, effective, and cost-efficient steel buildings will stay operated in the future. This inherently requires more skilled designers and engineers, better design tools, and more truly innovative “smart” retrofit techniques to be developed for the seismic strengthening of code-deficient steel buildings.

Cross-References

- [Assessment of Existing Structures Using Inelastic Static Analysis](#)
- [Buckling-Restrained Braces and Their Implementation in Structural Design of Steel Buildings](#)
- [Buildings and Bridges Equipped with Passive Dampers Under Seismic Actions: Modeling and Analysis](#)
- [Passive Control Techniques for Retrofitting of Existing Structures](#)
- [Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers](#)

- [Retrofitting and Strengthening of Contemporary Structures: Materials Used](#)
- [Steel Structures](#)
- [Strengthened Structural Members and Structures: Analytical Assessment](#)
- [Tuned Mass Dampers for Passive Control of Structures Under Earthquake Excitations](#)

References

- AISC 1999 LRFD (1999) Specification for structural steel buildings. American Institute of Steel Construction, Chicago
- AISC 358 (2005) Specification for structural steel buildings. American Institute of Steel Construction, Chicago
- ASCE 41–06 (2006) Seismic rehabilitation of existing buildings. American Society of Civil Engineers – ASCE, Reston
- Brockenbrough RL (2002) AISC Rehabilitation Bernam on and retrofit guide – a reference for historic shapes and specifications, Steel design guide series 15. American Institute for Steel Construction, Chicago
- Bruneau M (2004) Seismic retrofit of steel structures. In: Sociedad Mexicana de Ingeniera Sismica, VIII SNIS, Tlaxcala
- Buyukozturk O, Gunes O, Karaca E (2003) Progress on understanding de-bonding problems in reinforced concrete and steel members strengthened using FRP composites. *Construct Build Mater* 18(1):9–19
- Chen SJ, Lin HY (1990) Experimental study of steel I-beam to box-column moment connection. In: 4th international conference on steel structures and space frames, Singapore, pp 41–47
- Christopoulos C, Filiatrault A, Uang CM (2002) Self-centering post-tensioned energy dissipating (PTED) steel frames for seismic regions. In: Structural systems research project report no. SSRP-2002/06, Department of Structural Engineering, University of California, San Diego, p 292
- Coedova PP, Hamburger RO (2011) Steel connections: proprietary or public domain. *Modern Steel Construct* 51(10):24–31
- Crawford JE (2002) Retrofit methods to mitigate progressive collapse, 2002. Online: available in May 2014
- Demetriou D, Nikitas N, Tsavdaridis KD (2014) Performance of proportional-integral-derivative controllers on structures with variable damping tuned mass dampers. In: 6th World conference on structural control and monitoring, Barcelona
- Eatherton MR, Hajjar JF, Deierlein GG, Krawinkler H, Billington S, Ma X (2008) Controlled rocking of steel-framed buildings with replaceable energy-dissipating fuses. In: The 14th World conference on earthquake engineering, Oct 12–17, Beijing

- Egloff O, Tremblay R, Vincent R, Dussault S (2012) Finite element analysis of ductile fuses for W-shape steel bracing members. In: 15th World conference of earthquake engineering, Lisboa
- FEMA 274 (1997) NEHRP commentary on the guidelines for the seismic rehabilitation of buildings. Report no. FEMA 274. Federal Emergency Management Agency, Washington, DC
- FEMA 351 (2000) Recommended seismic evaluation and upgrade criteria for existing welded steel moment frame buildings. Report No. FEMA 351. Federal Emergency Management Agency, Washington, DC
- FEMA 356 (2000) Prestandard and commentary for the seismic rehabilitation of buildings. In: Report No. FEMA 356. Federal Emergency Management Agency, Washington, DC
- FEMA 547 (2006) Techniques for the seismic rehabilitation of existing buildings. Report No. FEMA 547. Federal Emergency Management Agency, Washington, DC
- Garlock M, Ricles JM, Sauce R (2004) Experimental studies on full-scale post-tensioned steel moment connections. In: 13th World conference of earthquake engineering, Vancouver
- Ghobadi MS, Ghassemieh M, Mazroi A, Abolmaali A (2009) Seismic performance of ductile welded connections using T-stiffener. *J Construct Steel Res* 65:766–775
- Hedayat A, Celikag M (2009) Post-Northridge connection with modified beam end configuration to enhance strength and ductility. *J Construct Steel Res* 65 (7):1413–1430
- Khabit IF, Mahin SA, Pister KS (1988) Seismic behavior of concentrically braced steel frames. Report no. UCB/EERC-88/01. Earthquake Engineering Research Center, University of California, Berkeley
- Kiamanesh R, Abolmaali A, Ghassemieh M (2010) The effect of stiffeners on the strain patterns of the welded connection zone. *J Construct Steel Res* 66(1): 19–27
- Leelataviwat S, Goel SC, Stojadinovic B (1998) Drift and yield mechanism based seismic design and upgrading of steel moment frames. Research report no. UMCEE 98–29. Department of civil and environmental engineering, The University of Michigan, Ann Arbor
- Lenwari A, Thepchatri T, Albrecht P (2006) Debonding strength of steel beams strengthened with CFRP plates. *J Comp Constr* 10(1):69–78
- Leon RT, Yang CS (2003) Special inverted-V-braced frames with suspended zipper struts. Georgia Institute of Technology
- Marian L, Giaralis A (2014) Optimal design of a novel tuned mass-damper-inerter (TMDI) passive vibration control configuration for stochastically support-excited structural systems. *Prob Eng Mech*. <http://www.sciencedirect.com/science/article/pii/S0266892014000216>
- Pachoumis DT, Galousis EG, Kalfas CN, Christitsa AD (2008) Reduced beam section moment connections subjected to cyclic loading: experimental analysis and FEM simulation. *Eng Struct* 31:216–223
- Plumier A (1997) The dog-bone: back to the future. *Eng J* 34(2):61–67
- Qingshan Y, Bo LA, Na Y (2009) Aseismic behaviors of steel moment resisting frames with opening in beam web. *J Construct Steel Res* 65:1323–1336
- Sarhosis V, Tsavdaridis KD, Giannopoulos I (2014) Discrete element modelling of masonry infilled steel frames with multiple window openings subjected to lateral load variations. *The Open Construction and Building Technology Journal* 8:93–103
- Schnerch DA (2005) Strengthening of steel structures with high module carbon fiber reinforced polymer (CFRP) materials. PhD dissertation, North Carolina State University, p 265
- Schwinger C (2014) Quality assurance for structural engineering firms. Online: available in May 2014 <http://www.aisc.org/store/p-1795-quality-assurance-for-structural-engineering-firms.aspx>
- Shen Y (2009) Seismic performance of steel moment-resisting frames with nonlinear replaceable links. Master Thesis, Department of Civil Engineering, University of Toronto
- Slotboom M, Robbemont A, Habraken A, Teuffel P (2014) Safety and redundancy of adaptive buildings structures. In: Civil engineering for sustainability and resilience international conference, CESARE'14, Amman
- Smith MC (2002) Synthesis of mechanical networks: the inerter. *IEEE Trans Autom Contr* 47(10):1648–1662
- Thorburn LJ, Kulak GL, Montgomery CJ (1983) Analysis of steel plate shear walls. In: Structural engineering report no. 107. Department of Civil Engineering, University of Alberta, Edmonton
- Tsavdaridis KD, D'Mello C, Huo BY (2013) Experimental and computational study of vertical shear behaviour of partially encased perforated steel beams. *J Eng Struct* 56:805–822
- Tsavdaridis KD, Faghih F, Nikitas N (2014) Assessment of perforated steel beam-to-column connections subjected to cyclic loading. *J Earthquake Eng* 18(8):1302–1325
- Uang CM, Yu QS, Noel SA, Gross JL (2000) Cyclic testing of steel moment connections rehabilitated with RBS or welded haunch. *J Struct Eng* 126(1):57–68
- Vasdravellis G, Karavasilis TL, Uy B (2013) Large-scale experimental validation of steel post-tensioned connections with web hourglass pins. *J Struct Eng ASCE* 139(6):1033–1042
- Vian D, Bruneau M (2004) Testing of special LYS steel plate shear walls. In: 13th World conference on earthquake engineering, Vancouver, Paper no. 978
- Vincent RB (2008) Minimizing the strength of bracing connections. In: 6th international workshop connections in steel structures VI. Chicago, pp 127–141
- Yu QS, Uang CM, Gross JL (2000) Seismic rehabilitation design of steel moment connection with welded haunch. *J Struct Eng ASCE* 126(1):57–68

Strengthening Techniques: Masonry and Heritage Structures

Elizabeth Vintzileou

Department of Structural Engineering, Faculty of Civil Engineering, National Technical University of Athens, Athens, Greece

Synonyms

Arch; Box action; Deep rejoining; Diaphragm; Dome; Foundation; Grouting; Masonry; Mortar; Repair and strengthening; Stitching; Ties; Vault

Introduction

Masonry structures constitute a large part of the existing structures throughout the world. A significant part of the existing masonry structures belong to the built cultural heritage. Furthermore, many countries with rich cultural heritage are situated in earthquake-prone areas (in Europe, in Asia, in South America). The built cultural heritage includes unique monuments (temples, churches, castles, palaces, bridges, etc.) and (inhabited and leaving) historic centers. Therefore, there is a continuous need for the preservation of this wealth, whose cultural, historical, social, and economic values cannot be overestimated.

On the other hand, the international scientific community is working quite intensively in the field of preservation of the existing stock of structures, and, thus, the knowledge acquired in the last decades allows for interventions to be more rationally selected and better applied.

It is obvious that the vast variety of existing masonry structures, in terms of typology, of materials used for their construction, of their cultural significance, etc., does not allow this text to be of general applicability and to cover all possible cases. Thus, to make this text to be of some use, one has to limit its scope to a part of the existing masonry structures. Unique monuments are usually the subject of specific studies

(exhaustive surveys, in situ and laboratory tests, monitoring, sophisticated analyses, etc.) leading to the selection of adequate intervention techniques by multidisciplinary national or international teams. Thus, they are left out of the scope of this text. Furthermore, structures other than buildings (e.g., towers, bridges, etc.) present peculiarities that need to be treated separately. Therefore, those structures as well are not covered by this text. However, the repair and strengthening techniques, concisely presented and treated herein, are applicable also to the categories of structures that are not explicitly covered.

Prerequisites for the Selection of Adequate Intervention Techniques: Intervention Strategies

Interventions to individual structural members or to parts of an existing masonry structure should be part of a global intervention strategy. The selection of a strategy and, hence, of appropriate intervention techniques (for the repair or strengthening of a structure) is based on a series of investigations and studies that should precede the redesign of the structure, as briefly described herein:

- (a) **Documentation of the history of the structure:** It comprises the date of construction, information on construction phases, events that may have affected the structural system (earthquakes, floods, fires, wars, etc.), previous damages and interventions, etc. This phase is of paramount significance especially for cultural heritage structures.
- (b) **Documentation of the structural system:** It comprises the documentation of both the structural system and its current state. The documentation of the structural system includes the survey of the bearing system (bearing elements, their interconnections, path of forces from the region of their application to the Earth), as well as type and properties of materials, construction type of masonry, etc. The documentation of the

current state of the structure includes the survey of its pathology (survey of damage and decay), as well as qualitative interpretation of damage and decay. This step may require in situ and laboratory investigations (destructive and/or nondestructive, depending on the significance of the structure, its state, its future use, the characteristic to be detected, etc.).

- (c) **Diagnosis and assessment:** The input from steps (a) and (b) serves the needs of diagnosis (i.e., the identification of the causes of the observed pathological image) and assessment (i.e., the margins of safety of the structure in its current state). At this step, the structure is adequately modeled and subjected to normal (self weight, live loads, etc.) and to accidental actions (e.g., seismic action) with the purpose to reproduce the main damages of the structure. This is a step of major significance, as it allows for identification of the causes of damage and of the regions of the structure that are vulnerable, for checking the adequacy of modeling, and for assessing the margins of safety of the existing structure.

The output of steps (a) to (c) allows the engineer to select the strategy of interventions, to choose among the available intervention techniques those techniques that serve better the purpose of repair or strengthening of the structure, and to proceed with dimensioning of the interventions.

Depending on the results of the assessment of the existing structure and taking into account the actions that are expected to act on the structure in the future (e.g., change of use, required seismic upgrading of the structure, etc.), one of the following strategies of interventions may be applied:

- (i) **Local interventions to bearing elements:** This strategy is applicable when the bearing system is in principle adequate, whereas local inadequacies are detected in a limited number of bearing elements. The interventions consist in enhancing the bearing capacity or the deformation capacity or the

stiffness or the interconnection of limited number of elements, without major alteration of the bearing system.

- (ii) **Alleviation of irregularities:** This strategy is applicable, when the irregular distribution of masses, stiffnesses, and/or strengths leads to respective excessive requirements in some regions of the structure. In such a case, the alleviation of irregularities (by adding shear walls, by modifying the arrangement of existing bearing walls, etc.) may constitute the best strategy. It should be noted, however, that this strategy may be of limited applicability in the case of cultural heritage structures.
- (iii) **Global enhancement of bearing capacity:** This strategy is adequate when (a) a large number of bearing elements are severely damaged or (b) large inelastic deformations are expected to be imposed to the structure. In such a case, a global enhancement of the bearing capacity of the structure is advisable (strengthening of existing elements and/or addition of new bearing elements). Needless to say, the resulting enhancement of the global stiffness of the structure has to be taken into account.
- (iv) **Global enhancement of stiffness:** This strategy may be selected when the calculations show that the expected deformations are larger than those the (brittle) masonry elements can sustain. If this is the case, global enhancement of stiffness may be achieved by strengthening of the bearing elements and/or by improving the box action of the structure, e.g., by enhancing the diaphragm action of floors. Thus, the deformations to be imposed to the structure are reduced to a level the masonry elements can sustain.
- (v) **Reduction of masses:** This is a strategy that is desirable for structures to be subjected to seismic actions. In case of masonry structures, it may consist in using lighter materials for roof and floor pavements, as well as by reducing live loads (selection of appropriate use of the structure). It should be noted, however, that this strategy does not

lead to spectacular results, as – in the case of masonry structures – their self weight constitutes the most significant part of the vertical loads.

- (vi) **New bearing system:** When the existing bearing system is inadequate for seismic actions, a new bearing system may be constructed inside the existing structure. The new bearing system (e.g., made of reinforced concrete (RC) or steel) is connected to the existing structure, thus contributing to the resistance to seismic actions. This is a strategy applied in some cases to cultural heritage structures, where extensive interventions to the existing bearing system are invasive and, thus, undesirable. It should be noted that, the dimensioning of the new bearing system will be governed by the compatibility of deformations with those of the existing one, rather than by the required additional bearing capacity.

Repair and Strengthening Techniques

The information included in this chapter regards the most commonly applied techniques for the repair and strengthening of existing masonry structures. It should be noted that some of the techniques presented here, although of frequent use, are not sufficiently documented through experimental work or through monitoring of structures after intervention. The research on intervention techniques, as well as on the production, testing, and use of new materials, is ongoing.

The material included in this chapter is organized as follows: (a) Repair and strengthening techniques for masonry as well as for the connection between masonry elements are first presented. The aim of those techniques is to reinstate the pre-damage bearing capacity of individual elements or to enhance their bearing capacity, so that they may sustain seismic actions in the future. Finally, the repair of damage in the connection between masonry elements as well as its strengthening is examined, since the so-called box action of masonry structures strongly

depends on those connections. Subsequently, (b) techniques for the enhancement of the diaphragm action of floors and roofs as well as for their connection with the vertical elements are presented. The purpose of those techniques is, in most cases, not to enhance the bearing capacity of floors and roofs, but to ensure almost equal displacements of all vertical elements and to prevent diaphragms from moving independently of the vertical elements during an earthquake, thus, causing severe damage to the walls. Finally, (c) some indicative solutions frequently applied to cope with problems such as sliding soil or insufficient dimensions of the foundation are briefly presented.

Repair and Strengthening Techniques for Masonry Walls

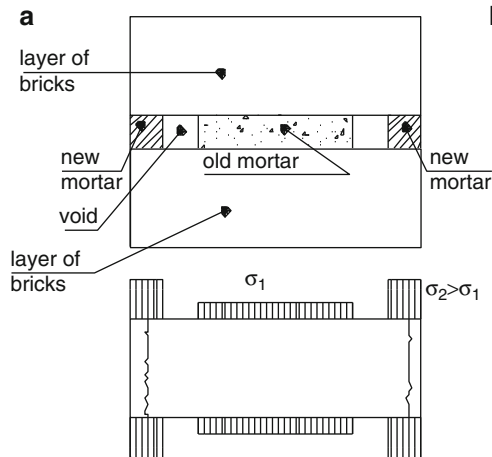
Deep Rejointing

Deep rejointing consists in the removal of the mortar from the joints to a depth at least equal to the thickness of the joints and filling with a mortar of improved properties. The technique should preferably be applied on both faces of masonry. In order to reach a significant enhancement of the mechanical properties of masonry, and taking into account that the contribution of the mortar to the strength of masonry is rather limited, a large part of the existing mortar has to be replaced. Therefore, this technique may be efficient only when applied to masonry of limited thickness (say 300 mm max.). This limits the applicability of deep rejointing to brick or block masonry, excluding stone masonry walls (of minimum thickness around 400 mm). It has to be noted that rejointing may be needed for reasons other than strengthening (e.g., to avoid water permeability of masonry or to prepare masonry for the application of grouts). In those cases, usually, there is no need for deep rejointing.

The application of the new mortar has to be made very carefully, so that the voids created due to the removal of the existing mortar are fully filled. Furthermore, the new mortar should not be excessively strong compared to the existing one. Otherwise, the compressive stresses on the

Strengthening Techniques: Masonry and Heritage Structures,

Fig. 1 (a) Qualitative interpretation of spalling of masonry units due to defective application of deep rejoining, (b) A brick wall tested in compression after a defective application of deep rejoining (Vintzileou 2001)



exterior (limited in thickness) part of masonry may be excessive (Fig. 1a) and lead to spalling of the masonry units (Fig. 1b). As a result, the compressive strength of masonry may be reduced, instead of being increased.

In conclusion, (a) deep rejoining may be efficient in case of brick masonries, 300 mm thick at maximum, and (b) provided that the technique is applied with due care and using adequate materials, (c) in case of stone masonries, it is not advisable to apply deep rejoining both because it is not expected to contribute to the improvement of the mechanical properties of masonry and because it is a quite expensive technique (mainly in workmanship). Even if applied to stone masonries, deep rejoining is considered as a repair technique, reinstating the initial mechanical properties of masonry.

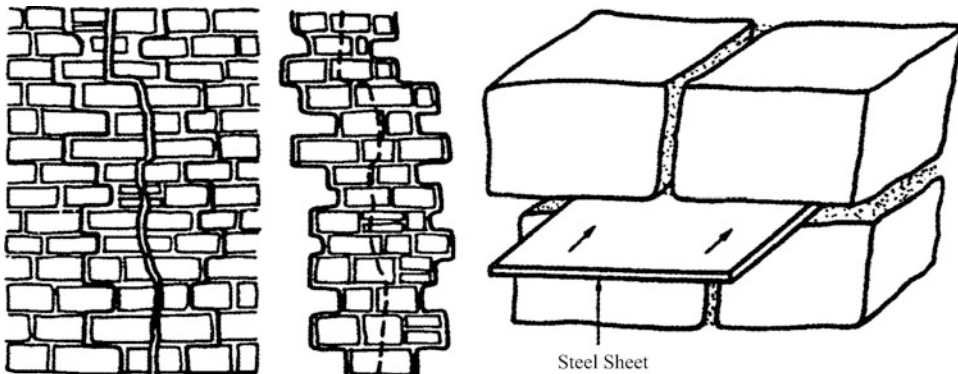
Grouting of Cracks

This repair technique is applicable to cracks having a width that does not exceed 10 mm approximately:

(a) Preparation of masonry: When masonry is plastered, plaster is removed in the vicinity of the crack. In cases where plaster cannot be removed (e.g., when masonry is covered with paintings or frescoes or mosaics or other decorative elements), special care has to be taken for the protection of the decoration of masonry. For example, it should be

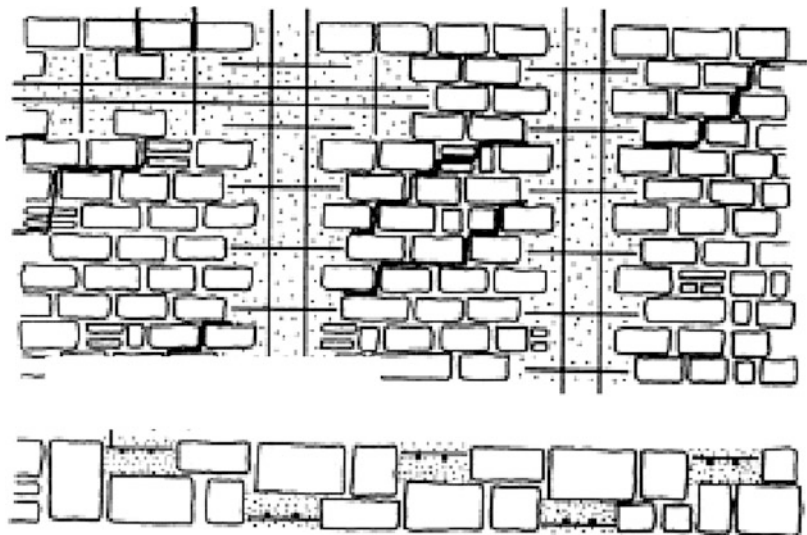
checked whether the frescoes are fixed on masonry. If not, their adequate fixing should precede the application of grout. Furthermore, during the application of grout, special care should be taken for immediate cleaning of the face of masonry, in case of leaking of the grout.

- (b) Holes are drilled to masonry at distances along the crack. The distance of consecutive holes should not exceed the thickness of masonry. Transparent plastic tubes are installed into the holes. Part of the length of the tubes is protruding.
- (c) The crack is sealed along its entire length, using an adequate mortar.
- (d) The materials for the grout (see also section “Generalities”) are mixed and the grout is introduced to the crack from the lower tube until it leaks from the immediately higher up placed tube and so on, until the crack is filled over its entire length. The grout is applied with a pressure not exceeding 1 bar. Depending on the state of the masonry, in order to avoid excessive hydrostatic pressure that might cause further damage to masonry, the crack length to be grouted should be limited within 1 day. For damaged, low strength masonries, this length should be limited to 1 m per day.
- (e) After hardening of the grout, the protruding parts of the tubes are cut and masonry is plastered, if this is the case.



Strengthening Techniques: Masonry and Heritage Structures, Fig. 2 Stitching of wide cracks

Strengthening Techniques: Masonry and Heritage Structures, Fig. 3 Repair of wide cracks by means of reinforced zones (indicative reinforcement)



Stitching of Cracks

In case of almost vertical wide cracks (width exceeding 10 mm), grouting alone cannot lead to reinstatement of the initial strength of masonry. In such a case, stitching of the cracks, prior to grouting, is needed.

Stitching consists in the removal of cracked masonry units along the crack and replacement with new sound masonry units laid in mortar of adequate composition (Fig. 2). If necessary, some of the bed joints along the crack may be locally reinforced. To this purpose, steel plates are inserted in the mortar joints (Fig. 2). The plates are smooth and, hence, of limited bond with the mortar. It is therefore advisable to artificially roughen them.

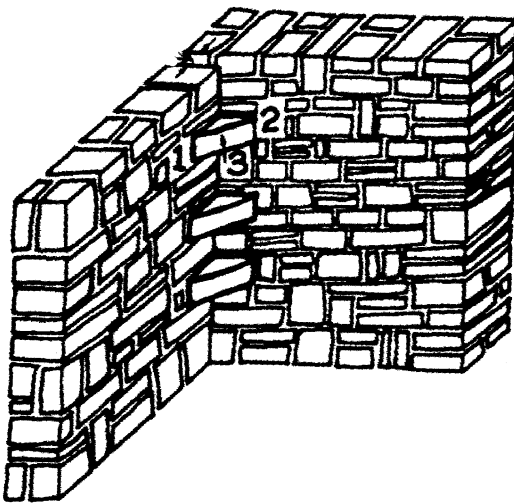
This technique cannot be applied to masonry walls that are cracked due to shear (diagonal or bi-diagonal cracks), as the removal of cracked masonry units along the cracks is not feasible. In those cases, the wall should be either demolished (and reconstructed) or it should be carefully grouted, provided that its construction type and its state allow grouting (see section “Grouting”).

An alternative, applicable exclusively to buildings without any heritage value, would be that of vertical and horizontal reinforced zones (Fig. 3). Grooves are created in the wall, both horizontally and vertically on both sides of masonry. Steel reinforcement is placed and a mortar of adequate composition or concrete is poured.

Stitching of Vertical Cracks Between Perpendicular Walls

The same technique is applicable to large cracks in the connection of walls. In that case, stones or bricks are removed and replaced by intact ones positioned diagonally (Fig. 4), to connect the two walls.

Alternatively, steel plates can be introduced to mortar joints (Fig. 5), provided that the thickness thereof is large enough for accommodating the plates.



Strengthening Techniques: Masonry and Heritage Structures, Fig. 4 Stitching of vertical cracks between perpendicular walls. Stones 1 and 2 are substituted by stone 3

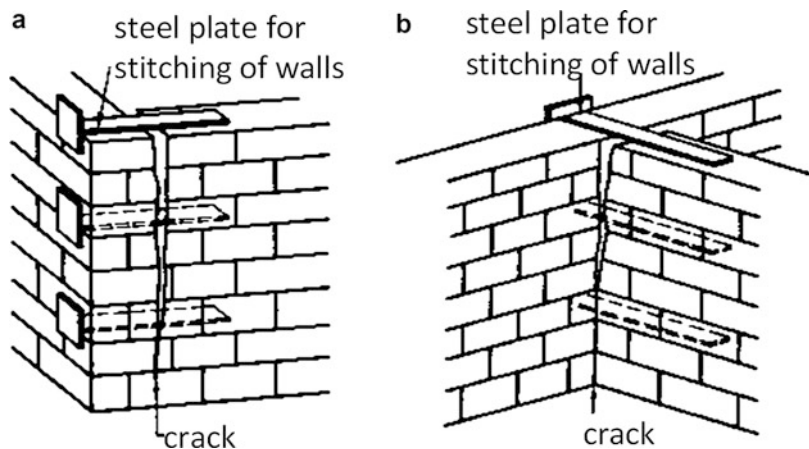
It should be noted that (a) the separation between the walls is not reversible and (b) the steel plates inserted into mortar joints are passive (i.e., they are mobilized in case further relative displacement of the walls takes place). Therefore, the cracks – after stitching – have to be grouted.

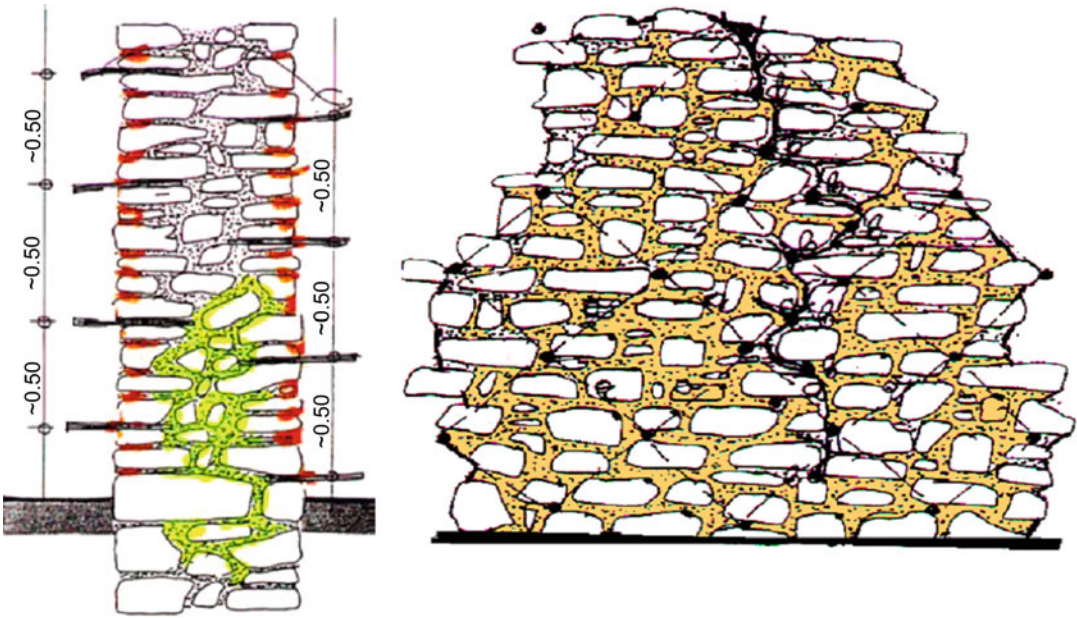
Grouting

Generalities Grouting of masonry (termed also as “homogenization” of masonry) consists in filling all voids and cracks within the mass of masonry using a hydraulic mixture. Organic binders (mainly epoxy resin) are not in use for masonry; the exothermal reaction of their hardening, combined with the (quite often) large voids within the mass of masonry may lead to disintegration of the epoxy resin-based grout before the application of any actions to the building. Furthermore, due to the dust present in the mass of masonry, the bond between the grout and the in situ materials may be defective. Finally, due to durability considerations, organic binders are excluded. It should also be noted that hydraulic grouts, even of low to medium mechanical properties can ensure significant enhancement of the mechanical properties of masonry (e.g., Binda et al. 1994; Valluzzi 2004; Vintzileou and Miltiadou-Fezans 2008), thanks to their high bond with the in situ materials.

A special mention needs to be made here of cement-based grouts (with a cement content exceeding 50 % per weight). Although they

Strengthening Techniques: Masonry and Heritage Structures, Fig. 5 Steel plates accommodated in horizontal mortar joints for stitching cracks





Strengthening Techniques: Masonry and Heritage Structures, Fig. 6 Grid of holes in a masonry wall to be grouted

constitute the first application of grouts to masonry, the experimental work carried out in several countries (Vintzileou and Tassios 1995; Toubakari 2002) has proven that cement-based grouts are not more efficient than ternary grout (cement content of approximately 30 % per weight, lime (and additives)) or hydraulic lime-based grouts. Furthermore, several cases of durability problems due to the use of cement are reported (e.g., efflorescence). Thus, in the recent years, the use of cement-based grouts is very limited, if not fully abandoned.

The Design of the Grout Hydraulic grouts injected into masonry have to comply with a set of requirements, namely, rheological (injectability, i.e., penetrability into fine cracks and voids, according to the design, and sufficient fluidity for the grout to be diffused into masonry, as well as stability), physical (low hydration heat, limited shrinkage, adequate hardening time, and hygroscopic properties), chemical (e.g., resistance to expansion and chemical stability of the products of chemical reactions), and mechanical requirements that are related to the desirable

mechanical properties of the grouted masonry (i.e., strength and deformability characteristics), depending on the actual state of masonry, the actions to be imposed, the overall scheme of interventions, etc.

Therefore, the design of a grout requires sufficient knowledge on the construction type of masonry, on its state, on the in situ materials, and on the aim of the intervention, all these combined with knowledge on material science.

Preparation of Masonry, Mixing of Grout, and Application The preparation of masonry for grouting is similar to that described in section “[Grouting of Cracks](#).” Holes are drilled to masonry at distances, forming a grid. In the general case of application of grouts from both faces of masonry elements, holes are drilled on both faces of masonry. As shown in Fig. 6, the holes on the interior face of masonry are drilled halfway between consecutive holes drilled on the exterior face of masonry. Thus, the resulting grid of holes is dense and it ensures better filling of the internal voids and cracks. The holes should be drilled deep enough (Fig. 7) to reach the core of



Strengthening Techniques: Masonry and Heritage Structures, Fig. 7 Drilling holes in masonry

masonry, thus ensuring that the grout will fill all the internal voids and cracks. Transparent plastic tubes are installed into the holes (Fig. 8) with an inclination toward the bottom of the element (to facilitate the flow of the grout). The (horizontal and vertical) distance between consecutive tubes may vary between 0.5 m and 1.0 m. However, it should in no case exceed the thickness of the elements to be injected.

The plastic tubes are numbered and reported on a drawing (Fig. 8). This is of major importance, because it allows for full control of the application procedure (tubes from which the grout is introduced and those from which the grout overflows are marked during the injection). All cracks are sealed (Fig. 9).

The materials to be used for the production of the grout should be available at the vicinity of the region to be injected, together with the mixer (Fig. 10). The use of high turbulence mixer is recommended, to provide adequate mix of the materials. After mixing of the constituent materials, the grout is transferred to a transparent

tube (Fig. 11), equipped with a device that provides a mechanical stirring of the grout during application (to avoid segregation).

When the preparation of masonry is completed (Fig. 12) and the grout mix is ready, the grout is applied with low pressure from the bottom to the top of masonry. The consumed quantity of grout is recorded. This is a very significant piece of information; it is taken into account when the enhanced mechanical properties of masonry are estimated.

Masonry is cured (kept wet) during the entire procedure that may last for several days, depending on the size and state of the injected element.

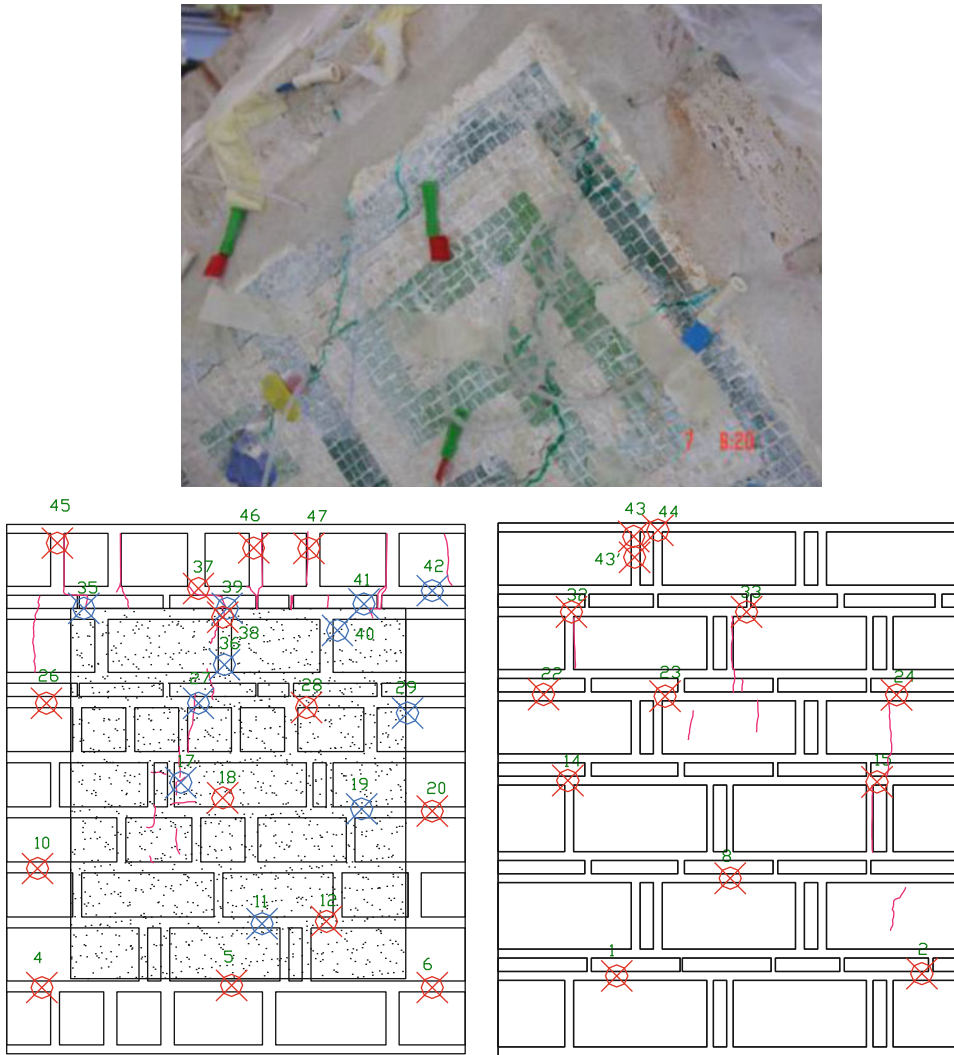
Reinforced Plaster

This is a technique applicable to masonry of limited thickness (not exceeding 0.40 m). It is applied after repair of masonry (by grouting or by stitching of cracks, by rejoining masonry, etc.). Reinforced plaster should preferably be applied on both sides of masonry. Actually, it is beneficial especially against out-of-plane bending, and, therefore, either the interior plaster or the exterior one is mobilized (in tension) depending on the direction of the seismic action.

The plaster is removed and masonry is cleaned from any loose material. The (metallic or nonmetallic) grid of reinforcement is placed on the faces of masonry and fixed on it (using anchors). In order to achieve better bond of the reinforced plaster, it is advisable to remove the mortar from the joints as deep as possible and fill them during the application of the plaster. In case of metallic reinforcement (steel grid), the plaster is usually a cement-based one. In case of nonmetallic reinforcement (e.g., fiber-reinforced wires or bars), the plaster can be a hydraulic lime + pozzolan mortar.

Recently, fiber-reinforced mortars are also applied to masonry.

It should be noted that the use of reinforced plasters may not be advisable in the case of cultural heritage assets, where masonry is quite thick, and the application of the reinforced plaster may be possible only on the interior face of masonry (therefore, in its least efficient position).



Strengthening Techniques: Masonry and Heritage Structures, Fig. 8 Plastic tubes installed into drilled holes. Tubes are numbered and reported to drawings

Local Demolition and Reconstruction

In case of local collapse of walls or in case of partial collapse of the exterior leaf of a double- or three-leaf masonry or in case of significant out-of-plane displacement of a wall (Fig. 13), the only solution is the reconstruction of the part that has collapsed or undergone significant deformations.

Needless to say, reconstruction is usually combined with repair or strengthening techniques described in the previous paragraphs (e.g., grouting of three-leaf masonry, stitching of

large cracks in other than the collapsed regions of the building, etc.).

Strengthening of Walls Using FRP Rods, Laminates, or Sheets

The use of FRPs in strengthening (bearing and infill) masonry is quite extensive in the recent years. FRPs with various types of fibers and in various forms (rods, laminates, or sheets) are used for strengthening masonry walls (Fig. 14) either against out-of-plane (bending) or in-plane (shear) actions.

The use of FRP materials is supported by extensive experimental research (collected, e.g., in Morbin 2001; Rashadul 2008), whereas models for the dimensioning of the intervention are also proposed (e.g., Faella et al. 2004;

ElGawady 2005). The available experimental data prove that the technique is quite efficient in enhancing the bearing capacity of masonry, as well as its ductility, provided that a good bond with the substrate (masonry) is ensured.

The technology of FRP applications is not presented in detail in this text for several reasons, namely, (a) most of the tests available in the Literature refer to monotonic loading of masonry specimens; in a limited number of cases, also repeated loading (without change of sign) or cyclic loading (e.g., Grillo 2003, on concrete masonry) was applied to the specimens. On the other hand, (b) in laboratory tests, masonry walls are subjected either to out-of-plane bending or to in-plane shear. However, during an earthquake, masonry walls are subjected to cyclic actions, as well as to simultaneous in-plane and out-of-plane actions. This is a very adverse condition which is not sufficiently covered by the international Literature. The problem that may arise is



Strengthening Techniques: Masonry and Heritage Structures, Fig. 9 Sealing of cracks



Strengthening Techniques: Masonry and Heritage Structures, Fig. 10 Materials for preparation of the grout mix. High turbulence mixer

a premature de-bonding of the FRPs from the substrate with subsequent loss of the strengthening of the walls. Furthermore, (c) fire protection of the intervention is an issue. Another issue that



Strengthening Techniques: Masonry and Heritage Structures, Fig. 11 Device for mechanical mixing of the grout before application

needs to be very carefully examined is (d) whether, after the occurrence of damage in FRP strengthened masonry, its repair or strengthening is still possible. Actually, as proven by tests, the failure of FRP strengthened walls – even under monotonic actions – may be associated with extensive cracking and disintegration of masonry. Especially in the case of historic masonry, for which the possibility to re-intervene after a seismic event is a prerequisite, this aspect may be governing. Last but not least, (e) there are limited data regarding strengthening of historic masonry using FRP materials. It is questionable whether poor-quality (double-leaf or three-leaf) stone masonry, having very low elastic properties, can be strengthened using high-strength and low-deformability materials. In such a case, prior strengthening of masonry using “traditional” techniques would be required, whereas the issue of compatibility of the in situ and added materials should be carefully examined. For example, the application of FRP sheets, covering large portions of masonry elements, may create problems of building physics (e.g., entrapment of humidity in the mass of masonry walls), thus leading to decay of the covered masonry. Another example would be that of inserting FRP rods in (horizontal and vertical) mortar joints. Even if the practical difficulties in finding

Strengthening Techniques: Masonry and Heritage Structures, Fig. 12 Masonry walls ready for the application of grouts

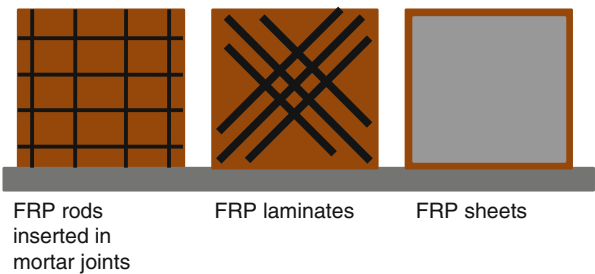




Strengthening Techniques: Masonry and Heritage Structures, Fig. 13 (a) collapse of the corner of a building, (b) partial collapse of the exterior leaf of

a three-leaf masonry wall, (c) significant out-of-plane displacement of a masonry wall

Strengthening Techniques: Masonry and Heritage Structures, Fig. 14 Alternative configurations of FRPs for masonry strengthening (schematic)



continuous bed joints in rubble stone masonries, the fact that grooves should be created to insert vertical FRP rods, and the difficulties in anchoring the rods at the top and at the base of a wall, as well as into transverse walls, are neglected, the problem remains of the behavior of rather soft original mortar joints that are partially filled with a (by orders of magnitude) stronger material. In real buildings, where the wall will be subjected to simultaneous cyclic shear and cyclic out-of-plane

bending, there is a danger of stress concentration in the reinforced portions of the masonry section. The stress concentration may lead to spalling of the masonry units and to subsequent adverse effects on the bearing capacity of the masonry element (comp. with Fig. 1). Moreover, (f) in case of historic buildings, the application of FRPs may be not be acceptable, as it alters the appearance of (unplastered) masonry. Thus, although the application of FRPs seems to be a technically



Strengthening Techniques: Masonry and Heritage Structures, Fig. 15 Typical damages due to out-of-plane bending of walls

promising intervention, it is believed that this technique should be applied with caution for strengthening masonry buildings that are to be subjected to seismic actions. It is believed that the (under investigation) use of inorganic bonding material (mortar) instead of an organic one (resins) may be a sensible alternative.

Enhancement of the Box Action of Masonry Buildings

Introductory Remarks

It is well known that masonry is a brittle material; it reaches its maximum resistance at small deformation value, whereas an abrupt decrease of its resistance takes place for imposed deformations exceeding the deformation corresponding to the maximum resistance. The vulnerability of existing masonry buildings to large deformations imposed by an earthquake is due to various typical construction features, namely, the defective connection of walls among them, the flexibility of floors and roofs, the large openings (especially when located close to the corners of the building), the large distance between transverse walls, etc. All measures taken with the purpose of alleviating those sources of vulnerability contribute to an improved seismic behavior of masonry. Last but not least, all the techniques that improve the box action of masonry buildings lead to an

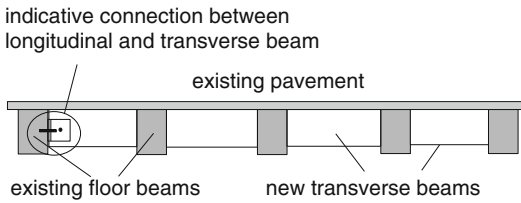
increase of their stiffness and to subsequent reduction of their natural period. In this way, the building “moves” to the left region of the response spectrum and, hence, to lower seismic actions.

Enhancement of the Diaphragm Action of Floors and Roofs

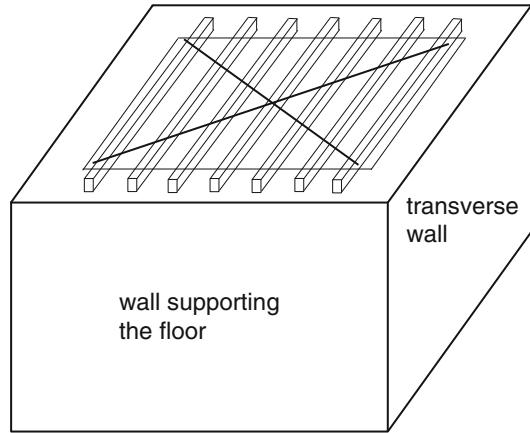
In existing masonry buildings, typically, timber floors and roofs are simply resting on vertical elements (walls or piers). Furthermore, due to the flexibility of floors and roofs in their plane, masonry walls are quite vulnerable to out-of-plane bending (Fig. 15). Under such actions, vertical cracks close to the connections between longitudinal and transverse walls, as well as out-of-plane collapse of portions of walls, constitute typical damages of masonry buildings.

The enhancement of the diaphragm action of floors and roofs leads to almost equal deformations of all vertical elements, thus limiting the out-of-plane vulnerability. Thanks to the overall increase of the stiffness of the building, the imposed deformations are also significantly reduced.

There are several techniques for the enhancement of the diaphragm action of floors and roofs. The Engineer has to select those most appropriate for each specific case and document (through analysis) the efficiency of the selected techniques.



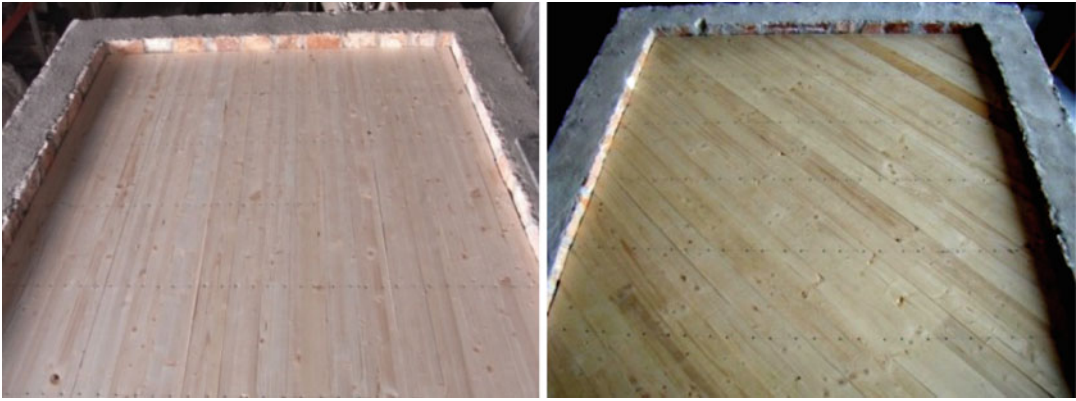
Strengthening Techniques: Masonry and Heritage Structures, Fig. 16 Enhancement of the diaphragm action of a floor, arrangement of transverse beams connected to the existing ones (schematic)



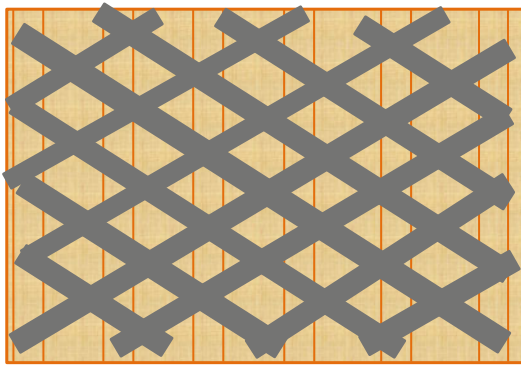
Strengthening Techniques: Masonry and Heritage Structures, Fig. 17 Arrangement of steel ties for the enhancement of the in-plane stiffness of a floor (schematic)

Some of the possible alternatives are briefly presented herein:

- (a) Providing transverse beams to the existing floor: Typically, in existing structures, floors consist of beams spanning in one direction (normally, the smaller of the two). The in-plane flexibility of the floor can be reduced by arranging (timber or steel) beams perpendicular to the existing ones (Fig. 16), thus forming a grid. It should be noted that the cross-sectional dimensions of the transverse beams, their spacing, and the dimensioning of their connections to the existing members are to be determined for each specific case. Similarly, in case of a trussed roof, beams perpendicular to the existing trusses can be arranged.
- (b) Providing a system of bi-diagonal steel ties: Significant enhancement of the in-plane stiffness of a floor (or roof) can be achieved by arranging steel ties along the two diagonals of the floor (Fig. 17). A system of multiple ties may be needed in the case of rectangular floors, in which one in-plan dimension is significantly larger than the other. As in the previous case, the ties (adequately connected to the existing floor beams or to the existing trusses of the roof) are to be adequately dimensioned. It should be noted that, typically, ties should be dimensioned to develop a tensile stress not exceeding 50 % of their yield strength, in order to avoid excessive strains of the steel elements and, hence, avoid significant in-plane deformations of the floor.
- (c) Providing a second pavement: In the frequent case of a timber pavement of floors and roof, significant enhancement of the in-plane stiffness can be reached through a second pavement. This second pavement can be located either on top of the existing one or at the bottom of the beams (to avoid changes in the height of the story). The second pavement should be either perpendicular to the existing one or, preferably, arranged at an angle of 45° (Fig. 18). In this case too, the Designer should check the efficiency of the solution through modeling and analysis of the structure.
- (d) Alternatively, instead of providing a second pavement, the in-plane stiffness of the floor may be enhanced by arranging bi-diagonal FRP laminates (Fig. 19). This alternative solution should be applied with caution: The efficiency of the FRP plates depends on their bond with the existing pavement. The simultaneous flexural (due to vertical loads) and in-plane (due to seismic actions) deformations of the floor may lead to detachment of FRP laminates from the substrate. In such a case, the efficiency of the intervention may be significantly reduced.
- (e) Replacement of the existing floors and roof by reinforced concrete slabs: This alternative, quite popular in previous decades, is



Strengthening Techniques: Masonry and Heritage Structures, Fig. 18 Existing timber pavement and additional timber pavement at an angle of 45°



Strengthening Techniques: Masonry and Heritage Structures, Fig. 19 FRP laminates glued on the existing pavement (schematic)

nowadays of rather limited application. Actually, it is rather invasive in case of cultural heritage buildings and, thus, not accepted by the competent authorities. Furthermore, recent research (Zaopo 2011; Mazzon et al. 2009; Magenes et al. 2012; Mouzakis et al. 2012) on floors stiffened using several alternative techniques, as well as shaking table tests on model buildings, has proven that solutions (a) to (d) provide sufficient in-plane stiffness to the floor, and, hence, the replacement of the existing floors and roofs with RC slabs is not inevitable.

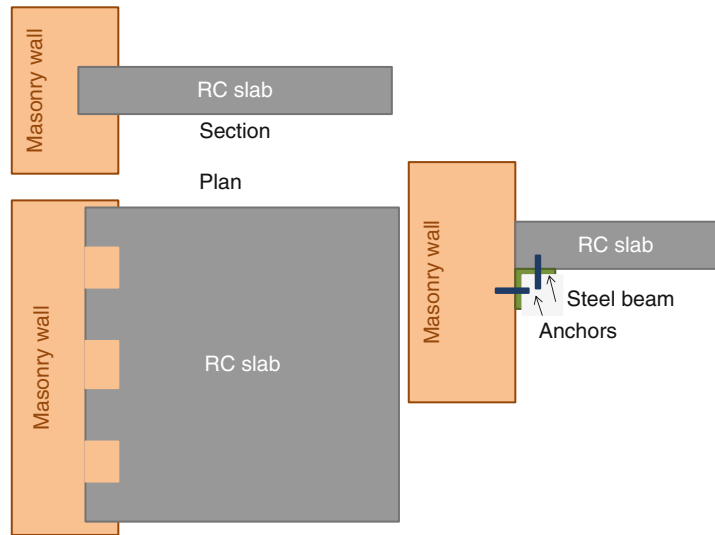
When this alternative is selected, adequate connection of the RC slab with the walls should

be ensured. To this purpose, one of the following two techniques may be applied: (i) Grooves are made in masonry walls at distances (Fig. 20), steel reinforcement is positioned, and the grooves are filled with concrete, during the construction of the slab. The spacing, the dimensions, and the reinforcement of the grooves are determined so that they are able to transfer the shear (due to the design seismic action) from the slab to the walls. The respective resistance of masonry should also be checked, (ii) steel beams are positioned along the masonry walls (Fig. 20), and they are fixed onto masonry. The RC slab is supported by the steel beams. In this case too, the connection of the steel beams to masonry walls should be adequately dimensioned.

Independently of the alternative selected for the enhancement of the in-plane stiffness of floors and roofs, there are two prerequisites for the efficiency of the solution, namely, (i) adequate connection of the diaphragm to masonry walls should be ensured and (ii) the supporting masonry walls should be adequately repaired or strengthened. Actually, a very stiff diaphragm not connected to masonry walls may be displaced as a rigid body during the earthquake, causing damage to the masonry walls. If the stiff diaphragm is connected to masonry walls made of unstrengthened poor-quality masonry, catastrophic damage may occur (Fig. 21).

Figure 22 shows the connection between floor and walls provided to a building model tested on

Strengthening Techniques: Masonry and Heritage Structures, Fig. 20 Alternative connections of RC slabs to masonry walls (schematic)



Strengthening Techniques: Masonry and Heritage Structures, Fig. 21 Catastrophic effects of RC slabs supported by poor-quality masonry (Courtesy of Prof. C. Modena)

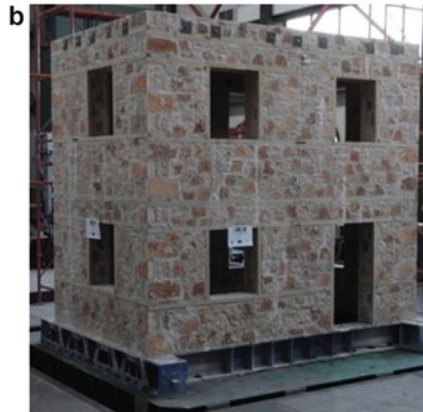
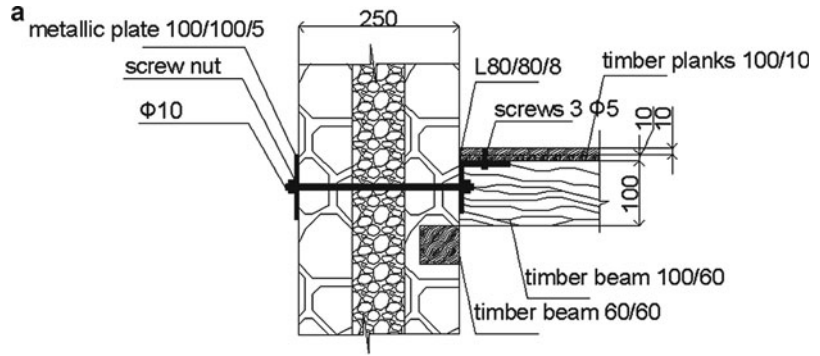
the earthquake simulator (Mouzakis et al. 2012). The diaphragm action of the floor was enhanced as shown in Fig. 18, and the walls (made of three-leaf masonry) were grouted. The connection was proven to be efficient, and no damage was caused to masonry walls even for an input acceleration exceeding 1g.

Improving the Connection Between Floors/Roofs and Vertical Elements

Frequently, in existing masonry buildings, roof and floors are simply resting on masonry walls. Thus, the vertical loads (the self weight of floors

and roof, the weight of pavements, as well as the live loads of the building) are concentrated at intervals. In case of poor-quality masonry, this may lead to local damage. Furthermore, during an earthquake, the floors are moving rather independently from the vertical masonry elements, thus, causing damage to the walls at their top (Fig. 23). Finally, the out-of-plane deformation of the walls is not prevented, due to loose connection with the horizontal elements. For all those reasons, it is desirable to provide better connection between horizontal and vertical elements.

Strengthening Techniques: Masonry and Heritage Structures,
Fig. 22 Detail of the connection between diaphragm and masonry walls. The building model after strengthening



Strengthening Techniques: Masonry and Heritage Structures,
Fig. 23 Damaged masonry walls due to the movement of simply supported roof



The improvement of the connection between horizontal and vertical elements can be achieved as follows:

- (1) Arrangement of a timber or a steel beam in the perimeter of floors and roof (Fig. 24) and

connection of the (timber or steel) beams of the floor with the collector beam. Although the beam shown in Fig. 24 was positioned during construction, it is feasible to provide a collector beam during interventions as well, by creating a recess in masonry walls. Floor

Strengthening Techniques: Masonry and Heritage Structures,
Fig. 24 The timber beams of the floor are connected to a perimeter collector beam



beams or roof trusses have to be connected to the collector beam, which should be connected to the walls (Fig. 22).

- (2) When the roof is replaced or temporarily removed during interventions, a reinforced concrete tie beam can be constructed at the top of the upper floor walls. Usually, the width of the RC tie beam is equal to the thickness of the walls. However, in case there are decorative elements on the facade (to be restored after interventions) or the accommodation of insulation is needed, the width of the RC tie beam can be adequately reduced. The construction of the RC tie beam is also possible after temporary shoring of the roof, if the roof is not to be replaced. The construction of the tie beam with the roof in its place is not encouraged, although it is technically feasible.

The tie beam may be connected to masonry using anchors. The diameter, the length, and the spacing of the anchors should be adequately designed. It should be noted that the construction of RC tie beams in intermediate floors is not recommended because (a) it is a quite invasive intervention and (b) the RC tie beam will inevitably be of limited width. Thus, its connection to the beams of the floor and to the walls should be ensured in a way similar to that described in the

previous paragraph (1), where all the application is dry and (c) durability problems may arise due to the continuous contact of timber elements and concrete.

Improving the Connection Between Walls

The significance of the adequate connection between walls was known to the constructors of masonry buildings. This is proven by the use of better quality, large-size cut stones in the corners of the buildings, even in the case of rather poor general construction quality (Fig. 25).

Nevertheless, the insufficiently stiff in their plane floors and roofs that are simply supported on the walls do not prevent damage due mainly to out-of-plane bending of walls and piers (Fig. 16).

It should be noted that, if enhancement of the diaphragm action of floors and roofs and adequate connection thereof with the walls is provided, further improvement of the connection between walls may not be needed. Actually, in such a case, the walls do behave as vertical slabs fixed along their perimeter, and their vulnerability to out-of-plane actions is significantly reduced. Nevertheless, if further improvement of the connection between walls is needed, steel ties can be used. The use of ties for the connection between walls was quite common during the construction of the existing masonry buildings.



Strengthening Techniques: Masonry and Heritage Structures, Fig. 25 Elaborate connection between walls at the corners of masonry buildings

Strengthening Techniques: Masonry and Heritage Structures, Fig. 26 The metallic tie (of limited length) was not able to prevent the out-of-plane collapse of the wall



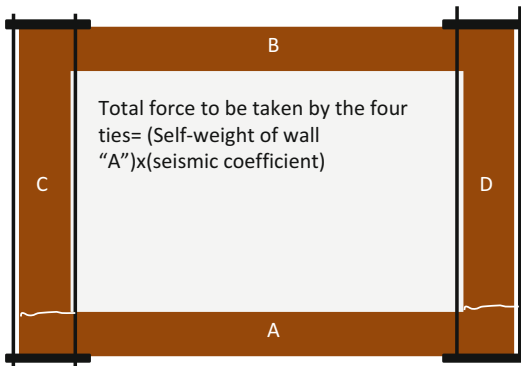
In that case, the ties (of limited length) were arranged within the thickness of one wall and anchored on its transverse wall (Fig. 26).

Steel ties provided as an intervention to improve the connection between walls are not embedded in masonry. They are (preferably) arranged at both sides of a wall or only along the interior face of a wall. They may be arranged in one single level or in more levels, provided that the height of the story allows for more ties to be positioned. Ties may be provided along one or both main axes of the building.

The ties are usually passive (i.e., non-prestressed). Therefore, they are expected to be mobilized only in case differential movement of the two connected walls takes place. In case wall A (Fig. 27) tends to separate from walls C and D, the ties should be able to develop a tensile force sufficient to sustain wall A in its

position. The ties are dimensioned so that their tensile stress does not exceed 50 % of their yield strength, in order to avoid excessive deformations when they are mobilized. Special care should be taken for their anchorage on masonry. The dimensions of the anchoring plates should allow (a) for almost uniform distribution of the compressive stresses on masonry and (b) for sufficiently small compressive stresses on masonry, to avoid local crushing.

The use of FRP ties as an alternative to steel ties is not encouraged. Although FRPs have a very high tensile strength (compared to normal steel) and they may have very high modulus of elasticity (which could prevent the occurrence of large deformations), they are characterized by complete lack of ductility. Therefore, if deformations larger than their deformation at failure occur, their fracture will be imminent.



Strengthening Techniques: Masonry and Heritage Structures, Fig. 27 Arrangement of steel ties to improve the connection between walls (schematic)

Arches, Vaults, and Domes

In old masonry structures, arches (Fig. 28) were constructed to bridge openings in buildings (Fig. 28 a, b, f) and river bridges (Fig. 28g), to transfer loads from domes to masonry piers, and so on, whereas vaults and domes (Fig. 28c, d, e, f) are frequently used to bridge large spans. In most of cases, stone or brick curved elements can safely bear the vertical loads for which they were conceived and constructed. Damage may be due to excessive vertical loading resulting, for example, from inadequate use (Fig. 29) or due to the lack of maintenance and repair (Fig. 28e).

In cases of damage due to excessive loads or to decay, simple repair is sufficient. Repair may include local reconstruction, grouting, or stitching of cracks, as described in section “[Repair and Strengthening Techniques for Masonry Walls](#).”

It should be noted, however, that most of the damages of curved elements are due to the effects of the thrust. It is well known that the loads are transferred from the curved elements to their supports following a thrust line; the thrust line is oblique when it reaches the supports. Therefore, there is a horizontal component of the thrust that causes horizontal displacements to the supporting elements. In the (quite frequent) case of flexible supports, outwards moving vertical elements may be damaged and induce damage to the curved

elements as well. This phenomenon is accentuated in the case of seismic events.

A typical example of the aforementioned damage is presented in Fig. 30: The cupola itself remained undamaged. However, all piers of the drum failed (some of them in shear, others in out-of-plane bending). Due to the out-of-plane deformations of the piers, the supporting arches have failed. The horizontal component of the thrust of the arches has pushed the vertical elements apart, resulting to out-of-plane displacements of the walls exceeding 0.10 m (observe the deformed shape of the monument in Fig. 30).

It is, therefore, of major importance to take measures for the reduction of the effects of thrust, by applying the techniques described in the following.

Ties and Struts

This is a typical efficient measure to reduce the effect of thrust. The technique was applied in masonry structures either during construction or as a remedial measure during previous interventions (Fig. 31). The technique consists in providing at the base of the curved element horizontal ties that can sustain the horizontal component of the thrust. The ties (made of steel or stainless steel or titanium, depending on the case) are anchored to transverse elements. For the maximum acceptable stress in the ties, as well as for their anchorage to masonry, the rules presented in section “[Improving the Connection Between Walls](#)” are valid.

When the structure is subjected to seismic action, there is a cyclic (inwards/outwards) movement of the vertical supporting elements. Therefore, in addition to the ties, elements functioning in compression (struts) should also be provided. In old structures, quite frequently, timber elements were used, able to play the double role of strut and tie (due to their sufficient resistance in buckling). This solution can be applied in modern interventions as well. Alternatively, the solution shown in Fig. 32 (adopted for interventions to Dafni Monastery, Miltiadou et al. 2012) can be chosen: A timber compression element is provided. A hole is drilled at the core of the



Strengthening Techniques: Masonry and Heritage Structures, Fig. 28 Arches, vaults, and domes

timber element and a steel tie is accommodated. The tie is anchored on masonry.

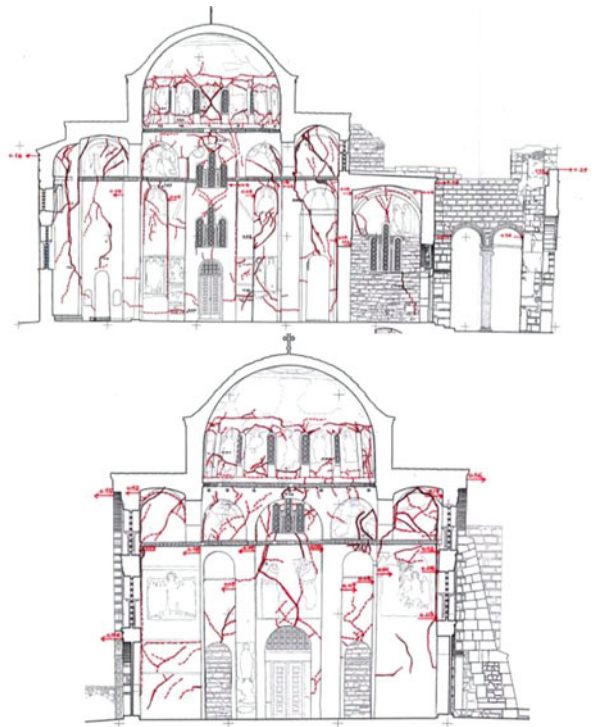
Steel or FRP Ties in the Perimeter of the Curved Elements

This is a technique frequently applied at the base of cupolas, as well as at the base of their

drum (Fig. 33). The ties (made either of steel or FRP) are designed to sustain the horizontal component of the thrust. It has to be noted, however, that they do not lead to any reduction in the flexibility of the system. Therefore, their application is useful but, quite often, not sufficient for the protection of the curved elements.



Strengthening Techniques: Masonry and Heritage Structures, Fig. 29 Partial collapse of the bridge after the passage of a heavy track



Strengthening Techniques: Masonry and Heritage Structures, Fig. 30 The main church of Dafni Monastery (Greece) damaged due to the 1999 earthquake

Buttresses

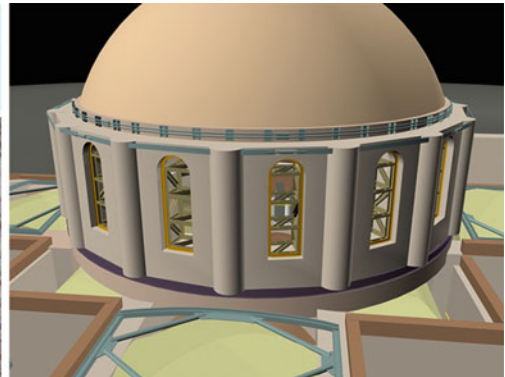
This is a technique applied also in the past (Fig. 34): The flexible vertical elements

supporting a vault or a dome are stiffened by an external buttress. Thus, the out-of-plane deformations of the supporting elements are



Strengthening Techniques: Masonry and Heritage Structures, Fig. 31 Arches with timber ties

Strengthening Techniques: Masonry and Heritage Structures, Fig. 32 The timber strut/steel tie solution was tested in-laboratory and adopted for application to the monument (Dafni Monastery)



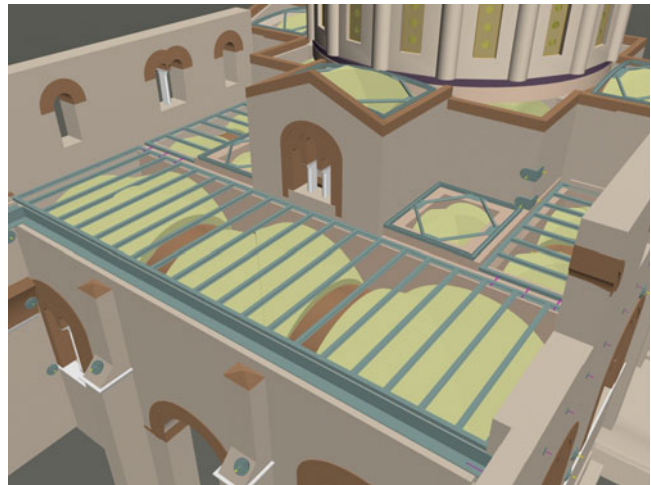
Strengthening Techniques: Masonry and Heritage Structures, Fig. 33 Dafni Monastery: existing steel tie at the base of the cupola, to be replaced by a stainless steel tie



Strengthening Techniques: Masonry and Heritage Structures, Fig. 34 (a) Ossios Lucas Monastery: external stone masonry buttresses were added against the southern facade. Due to architectural constraints, the buttresses do not reach the critical region of the system of arches that support the cupola. (b) Dafni Monastery:

during the extensive interventions of the early twentieth century, a pair of stone masonry buttresses was constructed against the northern wall of the church. This technique was not repeated against the southern wall (where the courtyard is located)

Strengthening Techniques: Masonry and Heritage Structures, Fig. 35 Dafni Monastery: steel diaphragms to be constructed on top of the system of vaults; they will be covered by a timber pavement



significantly reduced, and damage is avoided or significantly reduced.

It is obvious that in many cases, the construction of external buttresses, although a mechanically efficient solution, cannot be adopted either because there is no space for them to be accommodated or because the appearance of the structure is significantly altered.

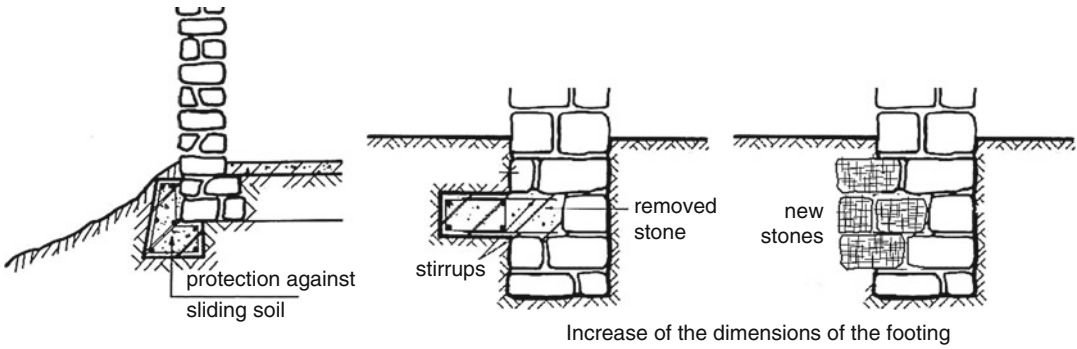
Enhancing the Diaphragm Action of Vaults and Domes

This is an alternative solution applicable when diaphragms (metallic or wooden ones) can be

arranged either under the curved elements or on top of them, when the curved elements are visible from the interior of the building. In this way, the effects of the thrust are minimized and both the curved elements and their supporting vertical elements are protected. This solution was adopted also in the case of Dafni Monastery (Fig. 35).

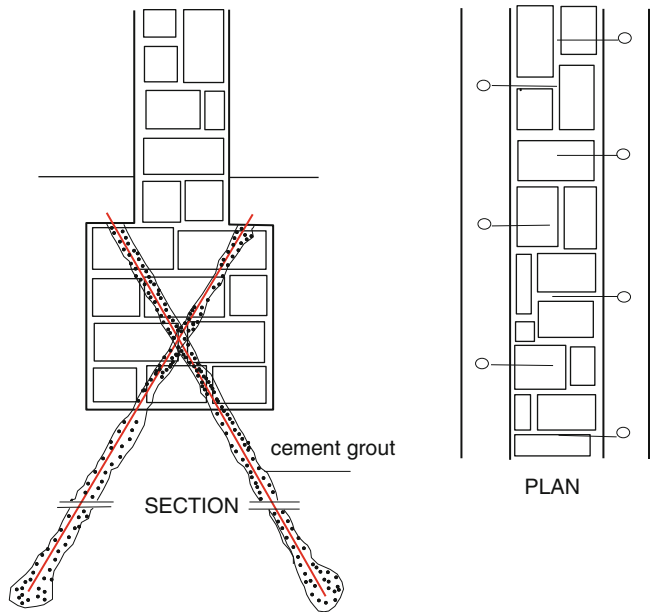
Modification of the Structural System

If the existing bearing system proves to be inadequate to resist seismic actions or if the new use of the structure requires higher performance (e.g., a residential building is given the use of a cultural



Strengthening Techniques: Masonry and Heritage Structures, Fig. 36 Interventions to foundation

Strengthening Techniques: Masonry and Heritage Structures, Fig. 37 Micro-piles



center), there may be a need for drastic interventions, such as construction of additional bearing elements, closing of openings with masonry, etc. In such cases, the construction of a new bearing system in the interior of the existing structure may be envisaged. When this alternative is selected, usually the new bearing system is made of steel. It should be noted, however, that for the new added system to assist the existing one in sustaining seismic actions, the former has to be connected with the original structural members. Moreover, the two systems have to be compatible from the deformation point of view; otherwise, the new added system may cause damage to the

original one (due to pounding). Therefore, in the design of the new system, deformations (rather than forces) may be the governing parameter.

Strengthening of the Foundation

Interventions to the foundation are needed when the mass of the structure is significantly increased due to the interventions or due to the change of use (higher live loads or heavier pavements, etc.). There are also cases of interventions needed when there are damages in the building due to sliding or differential settlements. The interventions to the foundation consist usually in increasing the dimensions of foundation walls using

masonry or reinforced concrete (Fig. 36), depending on the case.

When the properties of the foundation soil are such that a deep foundation is needed, micro-piles (Fig. 37) constitute a good solution. In the case of buildings in regions of archaeological interest, it should be made sure – before the application of micro-piles – that there are no ruins of older structures under the foundation of the building under examination.

Summary

Techniques that are commonly applied either to reinstate or to enhance the bearing capacity of masonry, and masonry components and masonry buildings are briefly presented and commented upon. This text comprises techniques aiming at (a) reinstating the mechanical properties of masonry (such as rejoining, grouting and stitching of cracks, etc.), (b) enhancing the mechanical properties of masonry (such as grouting of masonry, reinforced plaster), (c) improving the overall behavior of a masonry building (such as enhancement of the diaphragm action of floors and roof, improved connection between horizontal and vertical elements, etc.), and (d) improving the foundation conditions of the building. The techniques presented herein are, in principle, applicable to existing masonry buildings independently of their architectural, historical, and cultural values. There are, however, some cases – indicated in the text – where intervention techniques have to be applied with caution, when a heritage building is to be repaired or strengthened.

Cross-References

- [Ancient Monuments Under Seismic Actions: Modeling and Analysis](#)
- [Damage to Ancient Buildings from Earthquakes](#)
- [Masonry Box Behavior](#)
- [Masonry Components](#)
- [Masonry Modeling](#)

- [Post-Earthquake Diagnosis of Partially Instrumented Building Structures](#)
- [Seismic Analysis of Masonry Buildings: Numerical Modeling](#)
- [Seismic Strengthening Strategies for Heritage Structures](#)
- [Seismic Vulnerability Assessment: Masonry Structures](#)

References

- Binda L, Fontana A, Mirabella Roberti G (1994) Mechanical and stress distribution in multiple-leaf walls. In: Proceedings, 10th international brick block masonry conference, Calgary, Canada, pp 51–59
- ElGawady MA (2005) Shear model for URM walls retrofitted with FRP, Annual NZSEE Conference, Wairakei, New Zealand (electronic source)
- Faella C, Martinelli E, Nigro E, Paciello S (2004) Tuff masonry walls strengthened with a new kind of C-FRP sheet: experimental tests and analysis. In: 13th WCEE, Vancouver (Paper no, 923)
- Grillo VE (2003) FRP/steel strengthening of unreinforced concrete masonry piers. MSc thesis. University of Florida. 222 pp
- Magenes G, Penna A, Rota M, Galasco A, Senaldi I (2012) Shaking table test of a full scale stone masonry building with stiffened floor and roof diaphragms. In: 15th WCEE, Lisbon
- Mazzon N, Valluzzi MR, Aoki T, Garbin E, De Canio G, Ranieri N, Modena C (2009) Shaking table tests on two multi-leaf stone masonry buildings. In: 11th Canadian masonry symposium, Toronto
- Miltiadou A, Delinicolis N, Vintzileou E, Mouzakis CH, Dourakopoulos J, Yannopoulos P (2012) Retrofitting of the main church of Dafni Monastery. Hellenic Ministry of Culture (in Greek)
- Morbin A (2001) Strengthening of masonry elements with FRP composites. Diploma Dissertation. Dipartimento di Ingegneria Edile, Università di Padova, Italy, 190 pp
- Mouzakis C, Vintzileou E, Adami C-E, Karapitta L (2012) Dynamic tests on three leaf stone masonry building without timber ties before and after interventions. SAHC, Wrocław
- Rashadul I (2008) Inventory of FRP strengthening methods in masonry structures. Master thesis, Technical University of Catalunya. 131 pp
- Toumbakari E-E (2002) Lime-pozzolan-cement grouts and their structural effects on composite masonry walls. Doctor thesis, Katholieke Universiteit Leuven
- Valluzzi M-R (2004) Consolidamento di murature in pietra. Iniezioni di calce idraulica naturale. Collana Scientifica REIFICERE, GruppoEditoriale Faenza Editrice S.p.a., 128 pp. (in Italian)
- Vintzileou E (2001) The effect of deep rejoining on the compressive strength of brick masonry. *Masonry Int* 15(1):8–12

- Vintzileou E, Miltiadou-Fezans A (2008) Mechanical properties of three-leaf stone masonry grouted with ternary or hydraulic lime based grouts. *Eng Struct* 30(8):2265–2276
- Vintzileou E, Tassios TP (1995) Three-leaf stone masonry strengthened by injecting cement-grouts. *J Struct Eng Struct Division ASCE* 121(5):848–856
- Zaopo N (2011) Valutazione sperimentale dell'efficacia di interventi di miglioramento sismico di solai in legno rinforzati nel piano, Diploma dissertation (supervisor MR Calluzzi), University of Padua, 212 pp. (in Italian)

Structural Design Codes of Australia and New Zealand: Seismic Actions

George P. Kouretzis, Mark J. Masia and Clive Allen

Faculty of Engineering and Built Environment,
The University of Newcastle, Callaghan, NSW,
Australia

Synonyms

Australia; Design; Earthquake; New Zealand;
Seismic actions; Structural

Introduction

The design of structures to resist earthquake effects in Australia and New Zealand follows the pertinent Australian Standard AS 1170.4:2007 and New Zealand Standard NZS 1170.5:2004 provisions. Both Standards are based on the common Australian/New Zealand Standard AS/NZS 1170:2002 on structural design actions; and although they share elements such as the site subsoil classification system, they incorporate certain provisions such as the near-fault factor in NZS 1170.5:2004 to account for the different seismotectonic regime of the two countries. Indeed, Australia is an area of generally low seismicity, with the most catastrophic recent event being the 1989 Newcastle earthquake of magnitude $M = 5.6$, which resulted in 13 casualties and significant damages in the wider Newcastle area. On the other hand, New Zealand is

located on the boundary between the Indo-Australian and Pacific Plate and suffers from frequent strong earthquakes. In fact, New Zealand is one of the first countries to account for seismic actions in a building standard; as early as 1935, and following the 1931 Hawke's Bay earthquake that claimed 256 lives, the NZS 95 provided seismic loads to be considered in design (McRae et al. 2011). Despite constant advances and amendments in the seismic standards over the years, the recent 22 February 2011 Christchurch earthquake of magnitude $M = 6.3$ resulted in 185 casualties, numerous injuries, and extensive damages in the central business district of Christchurch and its eastern suburbs.

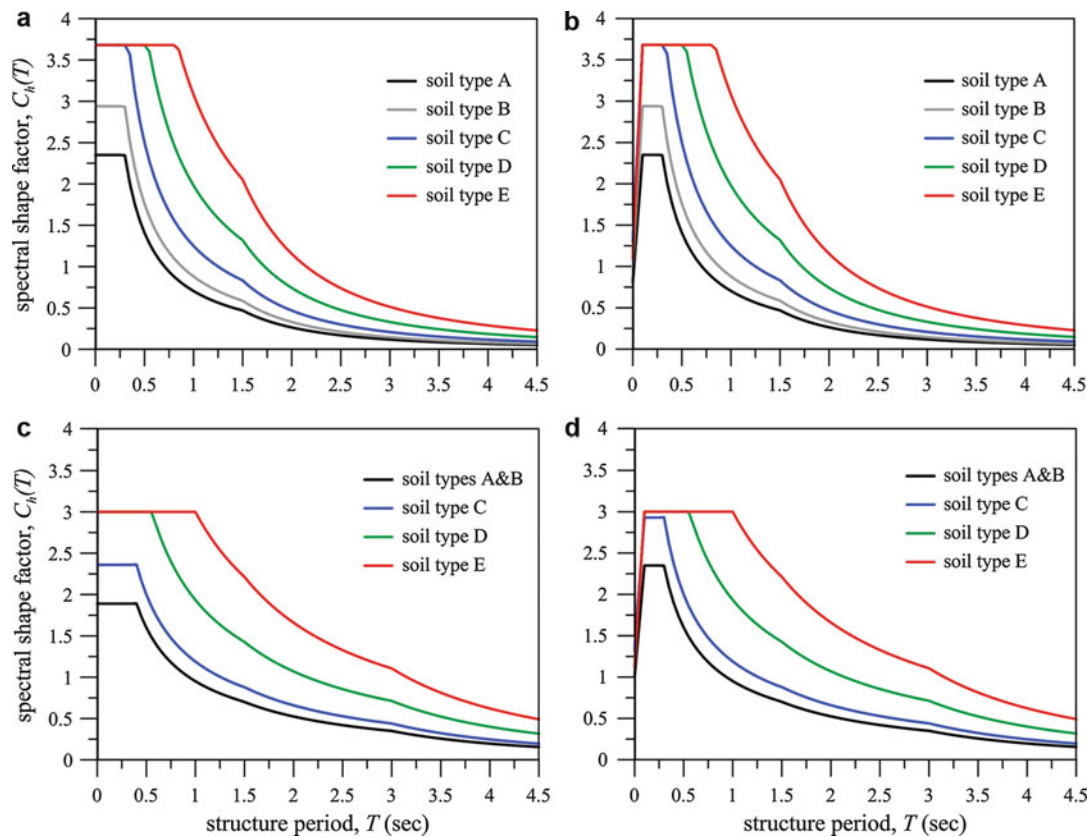
The purpose of this entry is to provide an outline of the seismic design actions currently considered in the abovementioned Standards and a short discussion on the rationale behind their adaptation, in the light of other widely used modern seismic codes provisions, such as the Eurocode 8 EN-1998-1 and the American ASCE/SEI 7-10. It covers the determination of elastic and inelastic design response spectra to be used together with static and dynamic analysis methods, as well as scaling of strong motion recordings according to NZS 1170.5:2004 to perform time history analyses. Emphasis is put on the determination of the input rather than on the provisions about the performance of equivalent static, modal, and time-domain analyses, which can be found elsewhere, including the Standards themselves.

Elastic Response Spectra

The elastic response spectra for horizontal loading, $C(T)$, can be described by the following generic equation, compatible with both Standards:

$$C(T) = C_h(T) \cdot (Z \cdot R) \cdot N(T, D) \quad (1)$$

where $C_h(T)$ is the spectral shape factor, $Z \cdot R$ is the product of the hazard factor Z times the factor R that depends on the annual probability of exceedance of the design earthquake and the



Structural Design Codes of Australia and New Zealand: Seismic Actions, Fig. 1 Spectral shape factor (5 % damping) plots for (a) AS 1170.4:2007/equivalent

static analyses, (b) AS 1170.4:2007/dynamic analyses, (c) NZS 1170.5:2004/equivalent static analyses, (d) NZS 1170.5:2004/dynamic analyses

design limit state (ultimate or serviceability) under consideration, and $N(T, D)$ is the near-fault factor, applicable only to the NZS 1170.5:2004. The parameters comprising the elastic response spectra equation are presented in the following paragraphs.

Spectral Shape Factor for Different Subsoil Classes

The spectral shape factors $C_h(T)$ to be used for equivalent static analyses and for dynamic (modal and time history) analyses are plotted separately in Fig. 1, while the relevant functions are provided in Table 1 for the different site subsoil classes. These spectral shape factors correspond to 5 % structural damping, and a correction factor to convert them to other damping values is not provided as, e.g., in EN-1998-1. However, such

a provision would be of value only for special structures such as tanks containing liquids, which are not within the scope of the specific Australian and New Zealand Standards.

As mentioned in the Introduction, Australian and New Zealand Standards share a common site classification scheme, which is preferably based on the predominant site period (or low amplitude natural period) of the site T_s , estimated as:

$$T_s = 4H_s/V_s \tag{2}$$

where H_s is the depth to the seismic bedrock of the site and V_s is the shear wave velocity of the overlying soil layer. In the absence of sufficient data to estimate the predominant site period, the site can be classified to one of the five subsoil classes (A to E) using (in the order of most preferred to least

Structural Design Codes of Australia and New Zealand: Seismic Actions, Table 1 Spectral shape factor (5 % damping) for different site subsoil classes

Site subsoil class	Equivalent static method			Dynamic analyses methods		
	Structure period, T (sec)	AS 1170.4:2007	NZS 1170.5:2004	Structure period, T (sec)	AS 1170.4:2007	NZS 1170.5:2004
A	$0 < T \leq 0.1$	2.35	1.89	$0 < T \leq 0.1$	$0.8 + 15.5 T$	$1.0 + 1.35(T/0.1)$
	$0.1 < T < 0.3$	0.704/		$0.1 < T < 0.3$	0.704/	2.35
	$0.3 \leq T < 0.4$	$T \leq 2.35$		$0.3 \leq T \leq 1.5$	$T \leq 2.35$	$1.60(0.5/T)^{0.75}$
	$0.4 \leq T \leq 1.5$					
	$1.5 < T \leq 3.0$	$1.056/T^2$	$1.05/T$	$1.5 < T \leq 3.0$	$1.056/T^2$	$1.05/T$
	$3 < T$		$3.15/T^2$	$3 < T$		$3.15/T^2$
B	$0 < T \leq 0.1$	2.94	1.89	$0 < T \leq 0.1$	$1.0 + 19.4 T$	$1.0 + 1.35(T/0.1)$
	$0.1 < T < 0.3$	0.88/		$0.1 < T < 0.3$	0.88/	2.35
	$0.3 \leq T < 0.4$	$T \leq 2.94$		$0.3 \leq T \leq 1.5$	$T \leq 2.94$	$1.60(0.5/T)^{0.75}$
	$0.4 \leq T \leq 1.5$					
	$1.5 < T \leq 3.0$	$1.32/T^2$	$1.05/T$	$1.5 < T \leq 3.0$	$1.32/T^2$	$1.05/T$
	$3 < T$		$3.15/T^2$	$3 < T$		$3.15/T^2$
C	$0 < T \leq 0.1$	3.68	2.36	$0 < T \leq 0.1$	$1.3 + 23.8 T$	$1.33 + 1.60(T/0.1)$
	$0.1 < T < 0.3$	1.25/		$0.1 < T < 0.3$	1.25/	2.93
	$0.3 \leq T < 0.4$	$T \leq 3.68$		$0.3 \leq T \leq 1.5$	$T \leq 3.68$	$2.0(0.5/T)^{0.75}$
	$0.4 \leq T \leq 1.5$					
	$1.5 < T \leq 3.0$	$1.874/T^2$	$1.32/T$	$1.5 < T \leq 3.0$	$1.874/T^2$	$1.32/T$
	$3 < T$		$3.96/T^2$	$3 < T$		$3.96/T^2$
D	$0 < T \leq 0.1$	3.68	3.0	$0 < T \leq 0.1$	$1.1 + 25.8 T$	$1.12 + 1.88(T/0.1)$
	$0.1 < T < 0.3$	1.98/		$0.1 < T < 0.3$	1.98/	3.0
	$0.3 \leq T < 0.56$	$T \leq 3.68$		$0.3 \leq T < 0.56$	$T \leq 3.68$	$2.4(0.75/T)^{0.75}$
	$0.56 \leq T \leq 1.5$			$0.56 \leq T \leq 1.5$		
	$1.5 < T \leq 3.0$	$2.97/T^2$	$2.14/T$	$1.5 < T \leq 3.0$	$2.97/T^2$	$2.14/T$
	$3 < T$		$6.42/T^2$	$3 < T$		$6.42/T^2$
E	$0 < T \leq 0.1$	3.68	3.0	$0 < T \leq 0.1$	$1.1 + 25.8 T$	$1.12 + 1.88(T/0.1)$
	$0.1 < T < 0.3$	3.08/		$0.1 < T < 0.3$	3.08/	3.0
	$0.3 \leq T < 1.0$	$T \leq 3.68$		$0.3 \leq T < 1.0$	$T \leq 3.68$	$3.0/T^{0.75}$
	$1.0 \leq T \leq 1.5$			$1.0 \leq T \leq 1.5$		
	$1.5 < T \leq 3.0$	$4.62/T^2$	$3.32/T$	$1.5 < T \leq 3.0$	$4.62/T^2$	$3.32/T$
	$3 < T$		$9.96/T^2$	$3 < T$		$9.96/T^2$

preferred): borehole log data together with in situ and laboratory tests, seismological methods including strong motion recordings (Nakamura 1989; Bouckovalas et al. 2002), qualitative borehole log descriptions, and geological information for the site. For layered sites, the site period can be estimated via an averaging method described in AS 1170.4:2007 and NZS 1170.5:2004.

The five different subsoil classes are defined in the Standards as follows (AS 1170.4:2007, NZS 1170.5:2004):

Class A: Strong rock. Strong to extremely strong rock, with:

- Unconfined compressive strength greater than 50 MPa, and

Structural Design Codes of Australia and New Zealand: Seismic Actions, Table 2 Maximum depth limits for site subsoil class C (After AS 1170.4:2007, NZS 1170.5:2004)

Soil type and description		Representative undrained shear strength, S_u (KPa)	Representative SPT N -values	Maximum depth of soil (m)
Cohesive soils	Very soft	<12.5	–	0
	Soft	12.5–25	–	20
	Firm	25–50	–	25
	Stiff	50–100	–	40
	Very stiff to hard	100–200	–	60
Cohesionless soils	Very loose	–	<6	0
	Loose dry	–	6–10	40
	Medium dense	–	10–30	45
	Dense	–	30–50	55
	Very dense	–	>50	60
	Gravels	–	>50	100

- (b) An average shear wave velocity over the top 30 m $V_{s,30} > 1500$ m/s, and
- (c) Not underlain by materials having a compressive strength less than 18 MPa or a shear wave velocity less than 600 m/s

Class B: Rock. Rock, with:

- (a) Unconfined compressive strength between 1 MPa and 50 MPa, and
- (b) An average shear wave velocity over the top 30 m $V_{s,30} > 360$ m/s, and
- (c) Not underlain by materials having a compressive strength less than 0.8 MPa or a shear wave velocity less than 300 m/s

A surface layer of no more than 3 m depth of highly weathered or completely weathered rock or soil material may be present in a class B site. It should be mentioned here that, based on the average shear wave velocity over the top 30 m criterion, other seismic codes such as EN-1998-1 or ASCE/SEI 7-10 would classify a site with $V_{s,30} > 360$ m/s to dense/stiff soil deposits, rather than rock (class B according to EN-1998-1, which corresponds to $360 < V_{s,30} < 800$ m/s, and class C according to ASCE/SEI 7-10, which corresponds to $366 < V_{s,30} < 762$ m/s). This suggests that class B in AS 1170.4:2007 and NZS 1170.5:2004 is rather broad, covering a range of subsoil conditions that may not exhibit similar behavior during an earthquake rich in high- to medium-frequency content (Bouckovalas and Kouretzis 2001).

Class C: Shallow soil sites. Sites that are not classified as class A, class B, or class E:

- (a) The predominant site period estimated as above is $T_s \leq 0.6$ s, or
- (b) Soil depth does not exceed the maximum values listed in Table 2.

Class D: Deep or soft soil sites. Sites that are not classified as class A, class B, or class E; and:

- (a) The predominant site period estimated as above is $T_s > 0.6$ s, or
- (b) Soil depth exceeds the maximum values listed in Table 2, or
- (c) Sites that are underlain by less than 10 m of cohesive soil with undrained shear strength $S_u < 12.5$ KPa or cohesionless soil soils with SPT values $N < 6$.

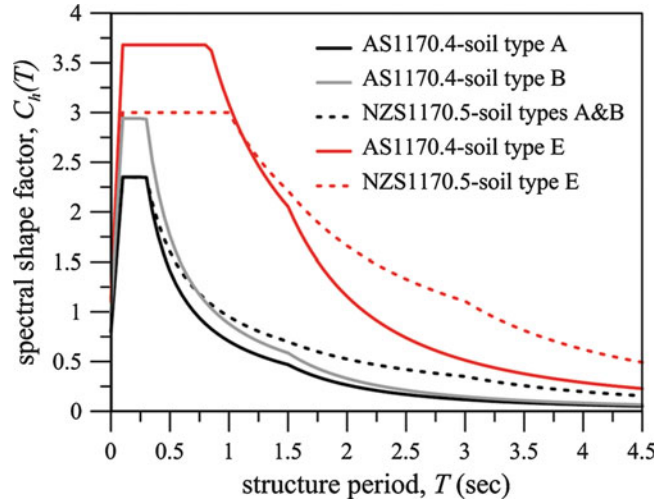
Class E: Very soft soil sites:

- (a) More than 10 m of very soft cohesive soil with undrained shear strength $S_u < 12.5$ KPa, or
- (b) More than 10 m of very loose cohesionless soils with SPT values $N < 6$, or
- (c) More than 10 m depth of soft-loose soils with shear wave velocity values $V_s < 150$ m/s, or
- (d) More than 10 m combined depth of soils with properties described in (a), (b), and (c) above.

A comparison between the spectral shape factors of AS 1170.4:2007 and NZS 1170.5:2004 is

Structural Design Codes of Australia and New Zealand: Seismic Actions,

Fig. 2 Comparison of AS 1170.4:2007 and NZS 1170.5:2004 dynamic spectral shape factor plots for rock (A, A and B) and soft soil (E) types (5 % damping)



attempted in Fig. 2, for the two extreme cases of rock (A and B) and soft soil (E) sites. The New Zealand Standard, which generally applies to earthquakes with higher magnitudes ($M \geq 6.5$), features lower spectral factor values in the low structural period range ($T < 1$ s) and higher spectral factor values in the high period range ($T > 1$ s). This is compatible with the fact that soil nonlinearity effects alter the spectral content of stronger earthquakes with magnitude $M > 5.5$, which in addition have a richer low-frequency content attributed also to near-fault effects.

Hazard Factor and Annual Probability of Exceedance

The hazard factor Z of a particular location corresponds to the peak ground acceleration (in g 's) for site class B (AS 1170.4:2007) or classes A and B (NZS 1170.5:2004), considering a design earthquake with return period of 500 years, i.e., a 10 % probability of exceedance during a 50-year design life. The hazard factors in the Australian Standard have not been updated since its previous version AS 1170.4:1993 and are based on the work of Gaull et al. (1990), who used data of the Australian Geological Survey Organisation dated back to 1856. All areas in Australia are assumed to be seismically active for design purposes, with the hazard factor generally ranging between $Z = 0.05$ and $Z = 0.13$, with the exception of the Meckering region in

Western Australia, where the hazard factor ranges between $Z = 0.14$ and $Z = 0.22$, and the Macquarie Island, located halfway between New Zealand and Antarctica, where the hazard factor is equal to $Z = 0.60$.

In NZS 1170.5:2004, the hazard factor Z corresponds to 0.5 times the spectral acceleration (5 % damping) for a structure period $T = 0.5$ s and soil subclass C, considering a design earthquake with a return period of 500 years (Fig. 1c, d). The minimum value across New Zealand is $Z = 0.13$, to ensure no collapse of structures even in areas of low seismicity, i.e., every structure is designed to survive the 84-percentile strong motion of a magnitude $M = 6.5$ normal-faulting earthquake at source-to-site distance of 20 km, a seismic scenario corresponding to the low-seismicity areas of New Zealand. The hazard factor at the high-seismicity major cities (e.g., Wellington, Napier, Hastings) is of the order of $Z = 0.40$, with the maximum value considered in the Standard being $Z = 0.60$.

The probability factor (AS 1170.4:2007) or return period factor (NZS 1170.5:2004) R , listed in Table 3, is used to scale the hazard factor to the required annual probability of exceedance, for the design limit state under consideration. In AS 1170.4:2007, the annual probability of exceedance is correlated with the design working life and the importance level of the structure, with the

Structural Design Codes of Australia and New Zealand: Seismic Actions, Table 3 Probability factor (AS 1170.4:2007) or return period factor (NZS 1170.5:2004)

Annual probability of exceedance	R
1/2,500	1.8
1/2,000	1.7
1/1,500	1.5
1/1,000	1.3
1/800	1.25
1/500	1.0
1/250	0.75
1/200	0.70
1/100	0.50
1/50	0.35
1/25	0.25
1/20	0.20

latter provided in AS/NZS 1170.0:2002 (Appendix F). The design annual probability of exceedance applies to the ultimate limit state only: According to AS 1170.4:2007, structures of importance levels 1–3 (i.e., all structures except ones carrying critical post-disaster functions and exceptional structures) de facto satisfy the serviceability limit state requirements when designed in accordance to the Standard. A special study needs to be carried out for critical post-disaster structures only (importance level 4), to ensure that they remain operational after a seismic event equivalent to the ultimate limit state design event for ordinary importance level 2 structures. The design annual probability of exceedance for structures and buildings (importance level 2) with a working life of 50 years is 1/500 and corresponds to a 10 % probability of exceedance during the design life of the structure. For major structures affecting crowds and associated with “very great” economic, social, and environmental consequences of failure (importance level 3), the annual probability of exceedance is reduced to 1/1,000 for structures with a design working life of 50 years and to 1/2,500 for structures with a design working life of 100 years.

The required annual probability of exceedance for structures designed according to NZS 1170.5:2004 is provided again in AS/NZS

1170.0:2002 (Table 3.3 of AS/NZS 1170.0:2002); however, in the New Zealand Standard, it is correlated with the design (serviceability or ultimate) limit state: The annual probability of exceedance for the common serviceability limit state SLS1 (requirement for no repairs on structural and nonstructural components) is 1/25 for all structures. The special serviceability limit state SLS2 applies to structures carrying critical post-disaster functions only (importance level 4) and is associated with an annual probability of exceedance of 1/500 for design working life of 50 years. A special hazard study is required to define the SLS2 actions for structures of importance level 4 with design working life of 100 years or more.

As far as the ultimate limit state is concerned, the design annual probability of exceedance is the same as in AS 1170.4:2007, for structures classified to importance levels 2 and 3. Note that the product $Z \cdot R$ in NZS 1170.5:2004 need not exceed $Z \cdot R = 0.70$ for ultimate limit state analyses but must be higher than $Z \cdot R = 0.20$ when an annual probability of exceedance 1/2,500 is considered. These $Z \cdot R$ bounds correspond to the higher and lower seismicity regions of New Zealand, respectively: The upper bound matches the 84-percentile near-fault strong motion due to a $M = 8.1$ event from the major Alpine Fault that runs along the South Island, divided by a margin of safety equal to 1.5 likely to result from applying the code design provisions.

Near-Fault Factor

The near-fault factor $N(T, D)$ is introduced in NZS 1170.5:2004 to account for near-source effects on the strong motion during the activation of a strike-slip fault: a) forward directivity, resulting in high peak velocities and displacements, and b) polarization of the long-period motions in the near-source region, where medium- to long-period pulses tend to be stronger in the direction perpendicular to the strike of the fault. Near-fault effects in NZS 1170.5:2004 are considered for eleven (11) major strike-slip faults classified as “class A” faults according to the Californian Category A fault criteria (Petersen et al. 2000), capable of producing earthquakes with magnitude $M = 7.0$

Structural Design Codes of Australia and New Zealand: Seismic Actions, Table 4 Maximum values of the near-fault factor $N_{\max}(T)$ for different structure periods

NZS 1170.5:2004		Linear interpolation functions	
Structure period, T (sec)	$N_{\max}(T)$	Structure period, T (sec)	$N_{\max}(T)$
≤ 1.5	1.0	≤ 1.5	1.0
2	1.12	$1.5 < T < 4$	$0.24 T + 0.64$
3	1.36		
4	1.60	$4 \leq T < 5$	$0.12 T + 1.12$
≥ 5	1.72	≥ 5	1.72

or greater and having slip rates of 5 mm/year of greater. The near-fault factor $N(T, D)$ must be considered during the derivation of the medium-to long-period band of the elastic response spectra for sites located within a distance D less than 20 km from the traces of the major strike-slip faults listed in NZS 1170.5:2004 and applies to annual probability of exceedance less than 1/250:

$$N(T, D) = N_{\max}(T) \quad \text{for } D \leq 2 \text{ km} \quad (3.1)$$

$$N(T, D) = 1 + (N_{\max}(T) - 1) \cdot (20 - D) / 18 \quad \text{for } 2 \text{ km} < D \leq 20 \text{ km} \quad (3.2)$$

$$N(T, D) = 1.0 \quad \text{for } D > 20 \text{ km} \quad (3.3)$$

where the factor $N_{\max}(T)$ is provided as a function of the structure period T in Table 4 and Fig. 3a. Linear interpolation functions are provided for intermediate period values, as proposed in NZS 1170.5:2004. In addition, a comparison of the effect of near-fault correction (Eqs. 3.1, 3.2, and 3.3) on the spectral shape factors for dynamic analyses is illustrated in Fig. 3b, for different distances from the fault.

Vertical Design Actions

According to AS 1170.4:2007, vertical earthquake actions generally need not to be considered in design. For the design of mechanical and electrical components, the vertical earthquake forces are taken equal to 50 % of the horizontal forces.

However, a provision is included stating that when the dynamic analysis requires the consideration of vertical earthquake forces, both upward and downward motions need to be considered and the vertical spectral shape factor shall be taken equal to

$$C_v(T) = 0.5 \cdot C_h(T_v) \quad (\text{AS 1170.4 : 2007}) \quad (4)$$

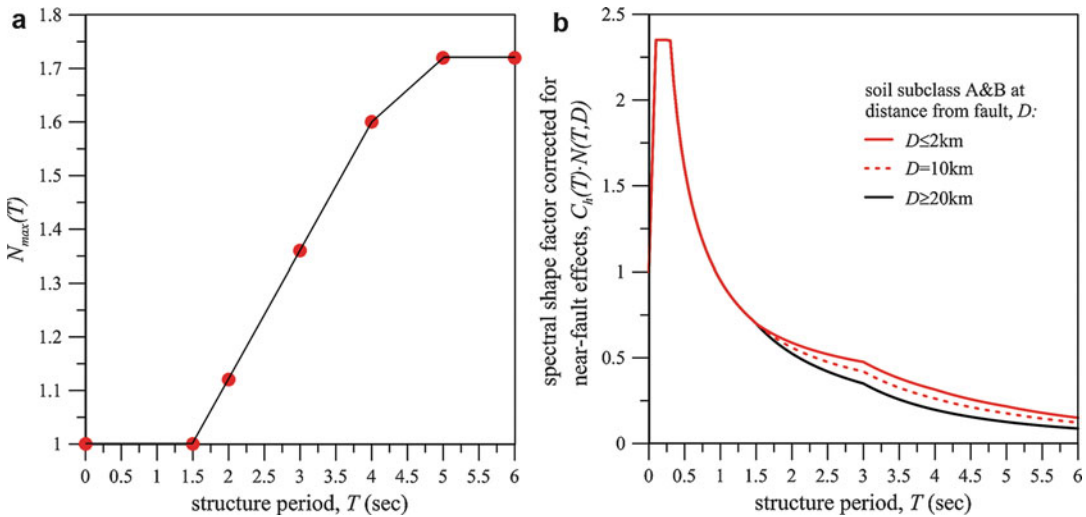
where T_v is the vertical period of vibration. In NZS 1170.5:2007 on the other hand, when vertical design actions need to be considered in the analyses (e.g., time history analyses), the elastic response spectrum for vertical loading is taken equal to 70 % of the corresponding horizontal spectrum, as

$$C_v(T) = 0.7 \cdot C_h(T_v) \quad (\text{NZS 1170.5 : 2004}) \quad (5)$$

where $T_v = 0$ for the design of the structure as a whole, or equal to the vertical period of the element under consideration, for the design of parts and components. Note that the vertical component of strong earthquake motion is usually rich in high-frequency content; thus, peak ground acceleration and spectral values in the near-source region may exceed the horizontal values (NZS 1170.5 Supp 1:2004, Niazi and Bozorgnia 1992; Ambrassey and Simpson 1996). In light of this, NZS 1170.5 Supp 1:2004 recommends that when the distance of the site from the fault trace is less than $D = 10$ km, the vertical design actions be considered equal to the horizontal design actions for structure period $T \leq 0.30$ s. Strong motion recordings from recent earthquakes, such as the Christchurch 2011 earthquake, provided further evidence in support of the above, as will be discussed later in this entry.

Structural Ductility and Structural Performance Factor-Inelastic Design Spectrum

Derivation of the (inelastic) design spectrum requires the correction of the elastic response spectrum to account for the ability of the



Structural Design Codes of Australia and New Zealand: Seismic Actions, Fig. 3 (a) Variation of the near-fault factor $N_{max}(T)$ with the structure period, T , and (b) comparison of spectral shape factors corrected for

near-fault effects for a rock site (soil subclass A and B) at a distance $D \leq 2$ km, $D = 10$ km, and $D > 20$ km from a major active fault (5 % damping)

particular structure to dissipate seismic energy via nonlinear response. Apart from the structural ductility factor, μ , which is related to the level of inelastic demand that can be reliably sustained by the structure, AS 1170.4:2007 and NZS 1170.5:2004 introduce the structural performance factor S_p to reduce the elastic design seismic actions. The structural performance factor is employed to quantify a number of effects that are not explicitly taken into account in state-of-practice structural analysis procedures by simply scaling the design loads. Such effects include (NZS 1170.5 Supp 1:2004):

- (a) Excitation effects: The estimated seismic loads correspond to a peak ground acceleration value, which may be reached only during a single loading cycle and therefore is unlikely to lead to significant damage (“effective” ground acceleration concept).
- (b) Individual structural elements typically feature a higher capacity than modeled during the analysis of the structure, due to higher material strength, strain hardening, strain rate effects, etc.
- (c) The total structural capacity is typically higher than predicted, due to redundancy

effects or the contribution of nonstructural elements (e.g., in fill walls) which is not taken directly into account in typical analysis models.

- (d) The energy dissipation of the structure is typically higher than assumed, due to damping introduced from nonstructural elements and soil-foundation interaction effects.

The ductility factor μ for the ultimate limit state is calculated in accordance with the appropriate material standard, where such data are provided. Contrary to NZS 1170.5:2004, where reference is made to material standards only and a special study is required otherwise, maximum ductility factors for typical structural systems are provided directly in AS 1170.4:2007 (Table 5). Alternatively, for a specific structure, μ and S_p can be determined via a nonlinear static pushover analysis. Note also that AS 1170.4:2007 applies only to structures with ductility factor $\mu \leq 3$, and when a higher ductility factor is considered, the design must be performed in compliance with NZS 1170.5:2004 provisions (Table 5).

While NZS 1170.5:2004 does not refer explicitly to ductility factor values for the ultimate limit state, for serviceability limit state analyses of

Structural Design Codes of Australia and New Zealand: Seismic Actions, Table 5 Ductility factor μ and structural performance factor S_p for different structural systems and specific structure types (After AS 1170.4:2007)

Structural system	Description	μ	S_p
Steel structures	Special moment-resisting frames (fully ductile) ^a	4	0.67
	Intermediate moment-resisting frames (moderately ductile)	3	0.67
	Ordinary moment-resisting frames (limited ductile)	2	0.77
	Moderately ductile concentrically braced frames	3	0.67
	Limited ductile concentrically braced frames	2	0.77
	Fully ductile concentrically braced frames ^a	4	0.67
	Other steel structures not defined above	2	0.77
Concrete structures	Special moment-resisting frames (fully ductile) ^a	4	0.67
	Intermediate moment-resisting frames (moderately ductile)	3	0.67
	Ordinary moment-resisting frames	2	0.77
	Ductile coupled walls (fully ductile) ^a	4	0.67
	Ductile partially coupled walls ^a	4	0.67
	Ductile shear walls	3	0.67
	Limited ductile shear walls	2	0.77
	Ordinary moment-resisting frames in a combination with limited ductile shear walls	2	0.77
	Other concrete structures not listed above	2	0.77
Timber structures	Shear walls	3	0.67
	Braced frames (with ductile connections)	2	0.77
	Moment-resisting frames	2	0.77
	Other wood- or gypsum-based seismic-force-resisting systems not listed above	2	0.77
Masonry structures	Close-spaced-reinforced masonry ^b	2	0.77
	Wide-spaced-reinforced masonry ^b	1.5	0.77
	Unreinforced masonry ^b	1.25	0.77
	Other masonry structures not complying with AS 3700:2001	1	0.77
Specific structure types	Tanks, vessels, or pressurized spheres on braced or unbraced legs	2	1.0
	Cast-in-place concrete silos and chimneys having walls continuous to the foundation	3	1.0
	Distributed mass cantilever structures, such as stacks, chimneys, silos, and skirt-supported vertical vessels	3	1.0
	Trussed towers (freestanding or guyed), guyed stacks, and chimneys	3	1.0
	Inverted pendulum-type structures	2	1.0
	Cooling towers	3	1.0
	Bins and hoppers on braced or unbraced walls	3	1.0
	Storage racking	3	1.0
	Signs and billboards	3	1.0
	Amusement structures and monuments	2	1.0
	All other self-supporting structures not otherwise covered	3	1.0

^aThe design of structures with $\mu > 3$ should be in accordance to NZS 1170.5:2004^bValues from AS 3700:2001

common structures (SLS1), it is provisioned that $1.0 \leq \mu \leq 1.25$, and for critical post-disaster structures (SLS2), it is provisioned that $1.0 \leq \mu \leq 2.0$. Furthermore, the ductility factor for vertical actions is always considered to be $\mu = 1.0$.

Note that in NZS 1170.5:2004, the inelastic horizontal design spectra are not derived by directly multiplying the elastic spectra by $(1/\mu)$. Instead, to account for the transition between equal displacement theory (which is valid for longer structure periods) and equal energy theory

(valid for shorter structure periods), a transition point is defined at $T_1 = 0.7$ s for soil subclasses A to D and at $T_1 = 1.0$ s for soil subclass E, where T_1 is the largest translation period of vibration along the direction under consideration. So, the factor to derive the inelastic response spectra is estimated as:

For subsoil classes A, B, C, and D:

$$k_\mu = \mu \quad \text{for } T_1 \geq 0.7 \text{ s} \quad (6.1)$$

$$k_\mu = 1 + (\mu - 1)T_1/0.7 \quad \text{for } T_1 < 0.7 \text{ s} \quad (6.2)$$

For subsoil class E:

$$k_\mu = \mu \quad \text{for } T_1 \geq 1.0 \text{ s or } \mu < 1.5 \quad (6.3)$$

$$k_\mu = 1.5 + (\mu - 1.5)T_1 \quad \text{for } T_1 < 1.0 \text{ s and } \mu \geq 1.5 \quad (6.4)$$

For the purpose of calculating k_μ , the largest translation period of vibration T_1 is not taken less than 0.4 s.

The structural performance factor S_p in AS 1170.4:2007 is provided in Table 5. In NZS 1170.5:2004, the structural performance factor is taken equal to $S_p = 0.7$ for ultimate limit state analyses, except for low-ductility structures with $1.0 < \mu < 2.0$, where S_p is calculated as:

$$S_p = 1.3 - 0.3\mu \quad (7)$$

For ultimate limit state verification of structures against sliding or overturning, S_p is taken equal to $S_p = 1.0$. For serviceability limit state SLS1 and SLS2 analyses, S_p is taken equal to $S_p = 0.7$, except otherwise dictated by the relevant material standard.

In accordance to the above, the inelastic design response spectrum in AS 1170.4:2007 is defined as

$$C_d(T) = \frac{C(T)S_p}{\mu} \quad (8)$$

and in NZS 1170.5:2004 as

$$C_d(T) = \frac{C(T)S_p}{k_\mu} \quad (\text{for horizontal actions}) \quad (9.1)$$

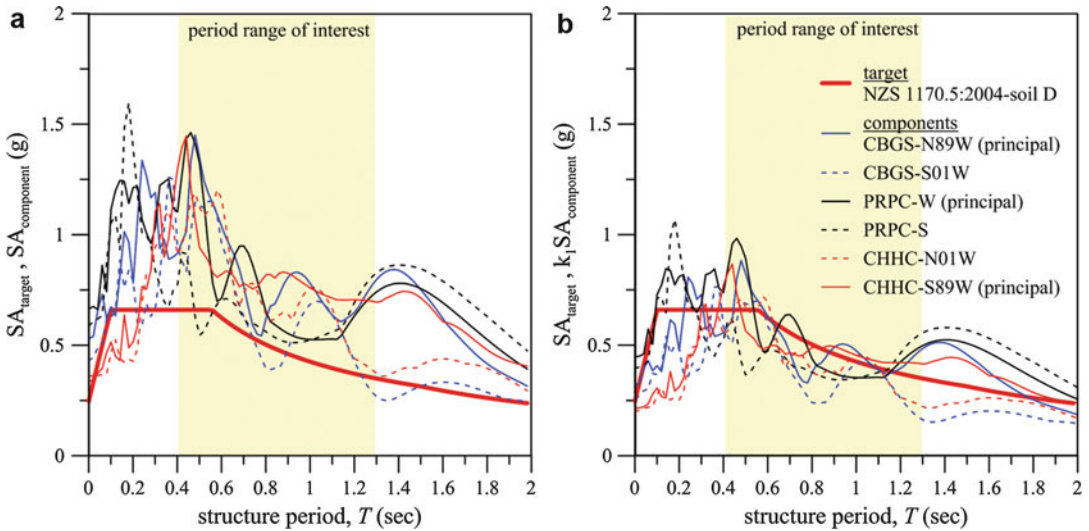
$$C_d(T) = C_v(T)S_p \quad (\text{for vertical actions}) \quad (9.2)$$

Actions for Dynamic Time History Analyses

In addition to actions applicable to common equivalent static and modal dynamic analyses, AS 1170.4:2007 and, mainly, NZS 1170.5:2004 include certain provisions for the determination of the appropriate design actions when dynamic time-domain analyses are to be performed. While in AS 1170.4:2007 there is a vague requirement that the response spectra of the actual acceleration time histories used shall “approximate” the design spectrum of Eq. 8, NZS 1170.5:2004 prescribes a more elaborate and explicit procedure for the selection and scaling of ground motion records.

Ground motion accelerographs to be used for time-domain analyses according to NZS 1170.5:2004 shall consist of both horizontal components of the recording; the vertical component may have to be considered too, for the analysis of structures sensitive to vertical strong ground motion. The above imply that the procedure refers to three-dimensional analyses, yet an adaptation to two-dimensional models is feasible, by considering only one horizontal component of each ground motion record as, e.g., in ASCE/SEI 7-10. A “family” of no less than three records must be employed in each time history analysis.

The selected acceleration records must be consistent with the local seismotectonic regime (expected magnitude, fault rupture mechanism, source-to-site distance, etc.). When the site is located near a major fault, i.e., $N(T, D) > 1$ (Eqs. 3.1, 3.2, and 3.3), then one set of the records must have a forward directivity component. Simulated, artificial accelerographs may be used too, when appropriate real acceleration recordings are not available. Each record shall be scaled by a “record scale factor” k_1 to



Structural Design Codes of Australia and New Zealand: Seismic Actions, Fig. 4 (a) Comparison of the 2011 Christchurch strong motion recordings elastic response spectra (5 % damping, horizontal components) with the design NZS 1170.5:2004 elastic design spectra

for the city of Christchurch and soil subclass D, (b) scaled elastic response spectra of the same strong motion recordings to match the NZS 1170.5:2004 design spectra for largest structure translation period $T_1 = 1$ s and $S_p = 1$

match the target design spectrum over the period range of interest and by a “family scale factor” k_2 to ensure that the energy content of at least one record in the family exceeds that of the design spectrum over the period range of interest. The target design spectrum is defined as

$$SA_{\text{target}} = \left(\frac{1 + S_p}{2} \right) C(T) \quad (10)$$

where the spectral shape factor $C(T)$ is calculated from Eq. 1 and S_p is the structural performance factor defined in the section “**Structural Ductility and Structural Performance Factor-Inelastic Design Spectrum.**”

In order to present the procedure of scaling ground motion records according to NZS 1170.5:2004, the horizontal components of three typical strong ground motion recordings from the 2011 Christchurch $M = 6.3$ earthquake were used. These records were scaled to match the target ultimate limit state spectrum of NZS 1170.5:2004 for a common structure (importance level 2) located in Christchurch, founded on subsoil class D (soft soil). According to NZS 1170.5:2004, for a structure of

importance level 2 in Christchurch with a design life of 50 years, $Z \cdot R = 0.22$ g and $N(T, D) = 1$, i.e., no near-fault effects, need to be taken into account due to possible rupture of the Port Hills Fault that is believed to have caused the 2011 earthquake. Assuming for simplicity that $S_p = 1$, the target spectrum is plotted in Fig. 4.

Strong motion recordings during the Christchurch 2011 earthquake exceeded by far the design peak ground acceleration at the ground surface of class D soft soils implied by NZS 1170.5:2004 (0.246 g) and reached maximum values of 1.67 g in the horizontal direction and 2.20 g in the vertical direction at the Heathcote Valley Primary School Station (HVSC) (www.geonet.org.nz). Here a family of three near-field recordings on soft to very soft soil are used, with peak horizontal ground accelerations ranging from 0.34 to 0.67 g (Table 6).

The elastic response spectra of the horizontal components of these three recordings are plotted in Fig. 4a, in comparison with the design response spectra of NZS 1170.5:2004. A band-pass filter in the frequency ranges of 0.10–0.25 Hz and 24.5–25.5 Hz was applied to the time histories

Structural Design Codes of Australia and New Zealand: Seismic Actions, Table 6 Characteristics of the typical strong motion recordings from the Christchurch 22 February 2011 $M = 6.3$ earthquake (Source: www.geonet.org.nz)

Site code	Subsoil class acc. to NZS 1170.5:2004	Epicentral distance	PGA-horiz.1 (g)	PGA-horiz.2 (g)	PGA-vert. (g)
CBGS	D	7	0.553	0.452	0.360
PRPC	E	6	0.669	0.595	1.88
CHHC	D	6	0.345	0.364	0.601

used to derive the spectra. It is clear that the spectral values of these typical records are above the design spectra across practically the whole range of important structural periods, an indication of the severity of the Christchurch earthquake.

Scaling of ground motion records to match the target design spectrum via the scale factor k_1 was performed while taking into account a structure period range that depends on the largest translational period of the structure in the direction of interest, T_1 . The period range of interest is defined as $0.4 T_1 \leq T \leq 1.3 T_1$, with the product $0.4 T_1$ not taken less than 0.4 s. Notice that this band of significant periods is considerably narrower compared to the corresponding one prescribed in EN-1998-1, which is $0.2 T_1 \leq T \leq 2.0 T_1$, or even ASCE/SEI 7-10, where the corresponding range is $0.2 T_1 \leq T \leq 1.5 T_1$. For the case at hand, it was assumed that $T_1 = 1$ s, and the period range of interest is denoted by the yellow band in Fig. 4.

The best-fit scale factor k_1 of each horizontal ground motion component was determined as the value minimizing the function $\log(k_1 SA_{\text{component}}/SA_{\text{target}})$ over the period range of interest, in a least mean square sense. In other words, the aim was to find via an interactive procedure the k_1 value that minimizes the sum:

$$\int_{0.4T_1}^{1.3T_1} \left[\log \left(\frac{k_1 SA_{\text{component}}}{SA_{\text{target}}} \right) \right]^2 dT = \min \quad (11)$$

The selected recordings must satisfy a “similarity” criterion $0.33 < k_1 < 3.0$. Moreover, in order to verify that each selected record reasonably matches the target spectrum over the period range of interest, the following inequality quantifying the mean (over the period of interest) square difference of the scaled over the target spectral values must be satisfied:

$$D_1 = \sqrt{\frac{1}{(1.3 - 0.4)T_1} \cdot \int_{0.4T_1}^{1.3T_1} \left[\log \left(\frac{k_1 SA_{\text{component}}}{SA_{\text{target}}} \right) \right]^2 dT} \leq \log(1.5) \quad (12)$$

Although not a normative criterion, a stricter fit $D_1 \leq \log(1.3)$ is proposed for most cases, according to NZS 1170.5 Supp 1:2004.

The k_1 and D_1 factors for the records at hand are listed in Table 7, where it is depicted that the specific records are compatible with NZS 1170.5:2004 requirements. The component of each record with the lower k_1 value was nominated as “principal” (Fig. 4, Table 7), and this value of k_1 was used as the record scale factor.

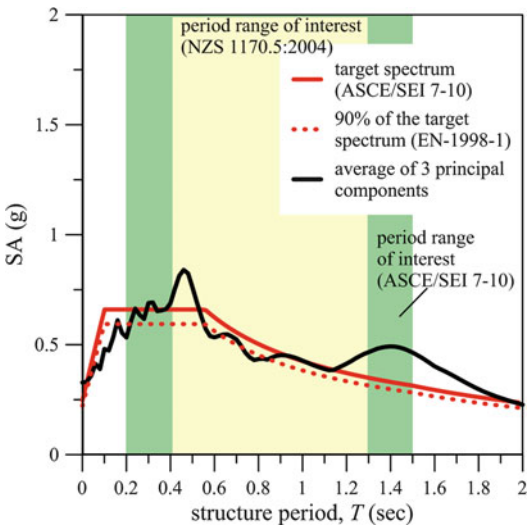
The record family scale factor, k_2 , ensures that the principal component of at least one record spectrum (after being scaled by k_1) exceeds the target spectrum over the period range of interest. It is estimated as the maximum value of the ratio $SA_{\text{target}}/\max(SA_{\text{principal}}) \geq 1.0$ within the period range of interest, where $\max(SA_{\text{principal}})$ is the maximum spectral value between all the principal components of the family, at each period step considered for the derivation of the spectra. To confirm the selection of the principal and

Structural Design Codes of Australia and New Zealand: Seismic Actions, Table 7 Scaling factors for the considered typical strong motion records

Site code	Component	k_1	D_1	k_2
CBGS	N89W (principal)	0.608	0.070	1.05
	S01W	0.789	0.105	
PRPC	W (principal)	0.672	0.081	
	S	0.785	0.097	
CHHC	N01W	0.683	0.064	
	S89W (principal)	0.598	0.058	

secondary components, the record family scale factor must be within the range $1.0 < k_2 \leq 1.3$. If $k_2 > 1.3$, then either a different record must be selected or the selection of the principal/secondary components may be reversed, aiming to minimize the product $k_1 k_2$. This may be the case when all three principal components in a family feature low spectral values within a particular period band, while one of the secondary components is rich in frequency content within the same band; thus, although the scale factor k_1 may be greater for that particular strong motion component, the product $k_1 k_2$ may be lower. In the case at hand however, the family factor k_2 is within the desirable range (Table 7), a fact that can be confirmed visually from Fig. 4.

The above scaling procedure must be followed for all directions of interest, corresponding to different translational periods of the structure, T_1 , and thus a different period range of interest. Although more elaborate than the analogous procedure of EN-1998-1 or ASCE/SEI 7-10 for the time history representation of the seismic action, where it is simply required that no value of the mean elastic spectrum of all time histories used is less than 90 % (EN-1998-1) or 100 % (ASCE/SEI 7-10) of the corresponding value of the elastic response spectrum, the procedure prescribed in NZS 1170.5:2004 results in a more rational scaling of strong motion records, and in a closer match to the design spectrum, as it does not impose restrictions related only to lower-than-target spectral acceleration values found in EN-1998-1 and ASCE/SEI 7-10. As a result of



Structural Design Codes of Australia and New Zealand: Seismic Actions, Fig. 5 Comparison of the mean response spectrum (5 % damping) of the three principal components listed in Table 7 with the NZS 1170.5:2004 target spectrum for soil subclass D and $S_p = 1$. The period of interest band according to NZS 1170.5:2004 ($0.4 T_1 \leq T \leq 1.3 T_1$) and according to ASCE/SEI 7-10 ($0.2 T_1 \leq T \leq 1.5 T_1$) is also drawn

the above, if one embraced the more conservative EN-1998-1 and ASCE/SEI 7-10 provisions regarding the average response spectra from the three principal components, the specific scaling factors applied would not be acceptable, as their average spectrum is below 100 % (ASCE/SEI 7-10) and marginally below 90 % (EN-1998-1) of the target spectrum (Fig. 5).

Finally, note that NZS 1170.5:2004 (as also ASCE/SEI 7-10) bases the accelerograph scaling procedure on spectral values within the range of interest only and does not impose restrictions on the zero period spectral response acceleration values of the strong motion, as, e.g., EN-1998-1.

Summary

Seismic design action definitions in the Australian AS 1170.4:2007 and New Zealand NZS 1170.5:2004 Standards were presented in a concise way, focusing on the determination of the elastic and inelastic design response spectra to be used together with static and dynamic methods

of analysis. A short discussion against provisions of other modern seismic codes, such as EN-1998-1 and ASCE/SEI 7-10, was attempted to point out certain key differences in particular clauses. Although Australian and New Zealand Standards embrace most of the recent developments on earthquake engineering, such as accounting for near-source effects on strong motion (but not topography effects as, e.g., EN-1998-1), any seismic code is not a “static” document, but rather a “dynamic” one, and evolves with lessons learned from recent major earthquakes that are included in future revisions as, e.g., the increase of the hazard factor Z for the Christchurch area from $Z = 0.22$ g to $Z = 0.30$ g in the latest compliance document of the New Zealand Building Code (Department of Building and Housing 2011).

Cross-References

- [Earthquake Response Spectra and Design Spectra](#)
- [Earthquake Return Period and Its Incorporation into Seismic Actions](#)
- [European Structural Design Codes: Seismic Actions](#)
- [Response-Spectrum-Compatible Ground Motion Processes](#)
- [Seismic Actions due to Near-Fault Ground Motion](#)
- [Selection of Ground Motions for Response History Analysis](#)

References

- Ambrassey NN, Simpson KA (1996) Prediction of vertical response spectra in Europe. *Earthq Eng Struct Dyn* 25:401–412
- AS 1170.4:2007. Structural design actions. Part 4: earthquake actions in Australia. Standards Australia
- AS 3700:2001. Masonry structures. Standards Australia
- AS/NZS 1170:2002. Structural design actions. Standards Australia/standards New Zealand
- ASCE/SEI 7-10. Minimum design loads for buildings and other structures. ASCE Standard, American Society of Civil Engineers, Structural Engineering Institute
- Bouckovalas GD, Kouretzis GP, Kalogeras IS (2002) Site-specific analysis of strong motion data from the

- September 7, 1999 Athens, Greece Earthquake. *Nat Hazards* 27:105–131
- Bouckovalas G, Kouretzis GP (2001) Stiff soil amplification effects in the 7 september 1999 Athens (Greece) earthquake. *Soil Dyn Earthq Eng* 21(8): 671–687
- Department of Building and Housing (2011) Compliance document for New Zealand building code, clause B1, structures, Amendment 11, August. New Zealand Government
- EN-1998-1. Eurocode 8: design of structures for earthquake resistance – Part 1: general rules, seismic actions and rules for buildings. European Committee for Standardization
- Gaull BA, Michael-Leiba MO, Rynn JMW (1990) Probabilistic earthquake risk maps of Australia. *Aust J Earth Sci* 37:169–187
- GeoNet-Earthquake Commission-GNS Science. website: www.geonet.org.nz. Accessed Sept 2013
- McRae G, Clifton C, Megget L (2011) Review of NZ building codes of practice. Report to the royal commission of inquiry into the building failure caused by the Christchurch earthquakes ENG.ACA.0016.1
- Nakamura Y (1989) A method for dynamic characteristics estimation of subsurface using microtremors on the ground surface. *RTRI Q Rep* 30(1):25–33
- Niazi M, Bozorgnia Y (1992) Behaviour of near-source vertical and horizontal response spectra at SMART-1 array. *Earthq Eng Struct Dyn* 21:37–50
- NZS 1170.5:2004. Structural design actions. Part 5: earthquake actions-New Zealand. Standards New Zealand
- NZS 1170.5 Supp 1:2004. Structural design actions. Part 5: earthquake actions. New Zealand-commentary. Standards New Zealand
- NZS 95. New Zealand standard code of building by-laws. Wellington, Standards Association of NZ
- Petersen MD, Toppozada TR, Cao T, Cramer CH, Reichle MS, Bryant WA (2000) Active fault near-source zones within and bordering the state of California for the 1997 Uniform Building Code. *Earthq Spectra* 16(1):69–83

Structural Optimization Under Random Dynamic Seismic Excitation

Christian Bucher
Institute of Building Construction and
Technology, Vienna University of Technology,
Wien, Austria

Synonyms

Optimal design; Random vibration; Response surface; Seismic isolation; Structural reliability

Introduction

In civil engineering, the design of structures has always been governed by fulfilling the needs of both safety and economy. Ideally, the designer would want to minimize cost while simultaneously maximizing safety. To a large extent, these objectives are in conflict; therefore, suitable compromises need to be found. A well-established tool for finding this set of best compromises is Pareto optimization. Figure 1 illustrates the trade-off between two conflicting objectives, both of which should be minimized. Reducing one objective automatically implies increasing the other one. Good compromises are to be found on the Pareto front.

Mathematically, this can be expressed as follows. Given a multi-objective minimization problem with m scalar-valued objective functions $f_k(\mathbf{x})$

$$f_k : D_{\mathbf{x}} \subset \mathbb{R}^n \rightarrow \mathbb{R} \quad (1)$$

a point $\mathbf{x}^*$ is called nondominated if (Ngatchou et al. 2008)

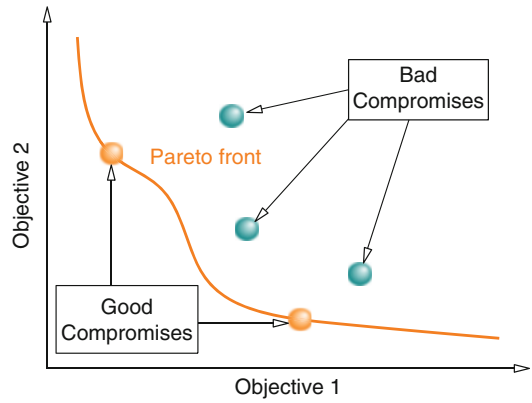
$$\begin{aligned} f_k(\mathbf{x}^*) &\leq f_k(\mathbf{x}) \forall k \forall \mathbf{x} \in D_{\mathbf{x}} \text{ and} \\ \exists \ell : f_\ell(\mathbf{x}^*) &< f_\ell(\mathbf{x}) \forall \mathbf{x} \in D_{\mathbf{x}} \end{aligned} \quad (2)$$

The Pareto set, i.e., the set of all nondominated solutions of an m objective problem can be constructed by assembling all solutions $\mathbf{x}_\alpha$ of all aggregated optimization problems of the form

$$f_\alpha = \sum_{k=1}^m \alpha_k f_k(\mathbf{x}) \rightarrow \text{Min.}! \quad (3)$$

in which α_k are positive coefficients. For more involved problems, alternative method based on multi-objective genetic algorithms (MOGA) such as non-dominated sorting genetic algorithms (NSGA) (Ngatchou et al. 2008) may be more appropriate. Applications of these approaches can be found, e.g., in Liu and Frangopol (2005) and Furuta and Kameda (2006).

Traditionally, however, in structural engineering, the abovementioned compromise has been realized by setting minimum target safety levels



Structural Optimization Under Random Dynamic Seismic Excitation, Fig. 1 Pareto front for the case of two conflicting objectives to be minimized

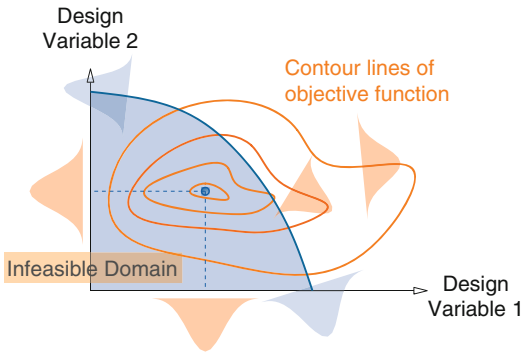
(as specified, e.g., by design codes) and aiming for the economically best solution satisfying the safety requirements. Mathematically, this corresponds to a *constrained* optimization problem. Due to many uncertainties involved in the construction process as well as in the environmental loads acting on a structure, the optimization problem cannot be treated deterministically. Figure 2 shows a simple example of a stochastic optimization problem in which the design variables contain uncertainties (e.g., due to the manufacturing or construction tolerances) and the objective (cost) function contains uncertainties (e.g., due to price fluctuations on the markets), and finally, the safety constraint function contains uncertainties (e.g., due to randomness of environmental loads such as earthquakes or storms).

Optimization Methodology

In a deterministic setting, a constraint optimization problem involves an objective function $f(\mathbf{x})$ and several constraint functions $g_k(\mathbf{x})$, $k = 1 \dots m$. Then the problem can be formulated as

$$\mathbf{x}^* = \operatorname{argmin} f(\mathbf{x}); g_k(\mathbf{x}) \leq 0; k = 1 \dots m \quad (4)$$

Since the cost function f will typically be random, any optimization will have to be formulated on the basis of expected values or probabilities



Structural Optimization Under Random Dynamic Seismic Excitation, Fig. 2 Simple example for a stochastic optimization problem

derived from the cost function. In this way, the objective function R for the optimization can be formulated, e.g., in terms of the mean value $\bar{f}$ and standard deviation σ_f , as

$$R = \bar{f} + k\sigma_f \quad (5)$$

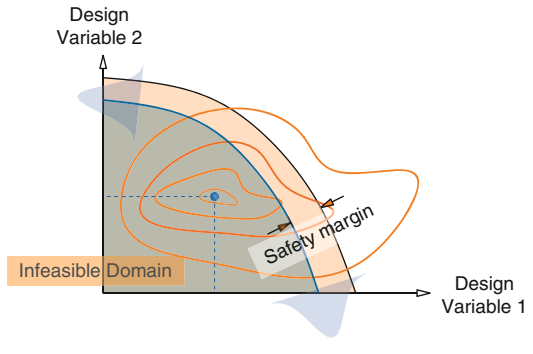
In this form, a multiple of standard deviation of the cost function is added to the mean values to form the objective function. The case $k = 0$ corresponds to game theory (suitable for, e.g., large portfolio management, fair gambling), whereas $k > 0$ implies a so-called *risk-averse* strategy.

Such a strategy is commonly employed when large losses cannot be compensated by long-term averaging (strong asymmetry).

Also, the constraint functions $g_k(\mathbf{x})$ will be random. Therefore, it is helpful to use the probability of constraint violation as new constraint, i.e., subject to the minimization of R to

$$P[g_k(\mathbf{x}^*)] > 0] \leq \varepsilon \quad (6)$$

In safety-related constraints, the probability level ε may be very small, e.g., $\varepsilon = 10^{-5}$. Hence, it is absolutely mandatory to provide computationally efficient methods for computing such small probabilities. This need is even more pronounced as the application of optimization algorithm required the repeated computation of small failure probabilities (whenever the algorithm changes the design variables).



Structural Optimization Under Random Dynamic Seismic Excitation, Fig. 3 Safety margin for constraints in a stochastic optimization problem

In traditional structural design, this is realized by introducing a safety margin (usually expressed in terms of partial safety factors) into the otherwise deterministic formulation of the constraints (cf. Fig. 3).

Computational Tools

The reduction of the large computational effort can be tackled in various ways, such as:

- Reduce and simplify the structural model, e.g., by modal reduction (Clough and Penzien 1993) or proper orthogonal decomposition (Bamer and Bucher 2012).
- Reduce the probabilistic description, e.g., by equivalent linearization (Roberts and Spanos 2003) or tail-equivalent linearization (Fujimura and Kiureghian 2007).
- Approximate the input-output relation of the system by a meta-model (Bucher and Macke 2005).
- Approximate the input-failure probability relation by a meta-model (Gasser and Schuëller 1997).
- Compute the input-failure probability relation by efficient Monte Carlo methods (Bucher 2009a).

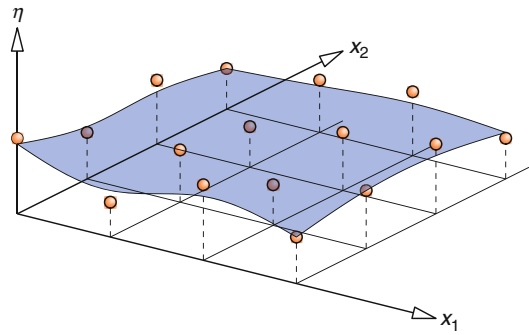
In the following, the response surface method and a specific Monte Carlo method (asymptotic sampling) will be explained in some detail.

Response Surface Method

The response surface method has been a topic of extensive research in many different application areas since the influential paper by Box and Wilson in 1951 (Box and Wilson 1951). A fairly complete review on existing techniques and research directions of the response surface methodology can be found in Hill and Hunter (1966), Mead and Pike (1975), Myers et al. (1989), and Myers (1999). One of the earliest suggestions to utilize the response surface method for structural reliability assessment was made in Rackwitz (1982). Here Lagrangian interpolation surfaces and second-order polynomials are rated as useful response surfaces. Moreover, the importance of reducing the number of basic variables and error checking is emphasized. Polynomials of different orders in combination with regression analysis are proposed in Faravelli (1989), in which fractional factorial designs are utilized to obtain a sufficiently large number of support points. The absolutely essential validation of the chosen response surface model is done by means of analysis of variance.

In Ouyornprasert et al. (1989), it has been pointed out that for reliability analysis, it is most important to obtain support points for the response surface very close to or exactly at the limit state $g(\mathbf{x}) = 0$. This finding has been further extended in Kim and Na (1997) and Zheng and Das (2000). In Brenner and Bucher (1995), the response surface concept has been applied to problems involving random fields and nonlinear structural dynamics. Besides polynomials of different orders, piecewise continuous functions such as hyperplanes or simplexes can also be utilized as response surface models.

Assume that an appropriate response surface model $q(\cdot)$ has been chosen to represent the experimental data. Then, for estimating the values of the parameters θ in the model, the method of maximum likelihood can be utilized. Under the assumptions of a Gaussian distribution of the random error terms ε , the method of maximum likelihood can be replaced by the more common method of least squares (Box and Draper 1987). In the latter case, the parameters θ are determined in such a way that the sum of squares of the differences between the value of



Structural Optimization Under Random Dynamic Seismic Excitation, Fig. 4 Response surface and data points

the response surface $q(\theta; \mathbf{x}^{(k)})$ and the measured response $z^{(k)}$ at the m points of experiment

$$\mathbf{x}^{(k)} = (x_1^{(k)}, \dots, x_n^{(k)})', \quad k = 1, 2, \dots, m \quad (7)$$

becomes as small as possible. In other words, the sum of squares function

$$s(\theta) = \sum_{k=1}^m \left(z^{(k)} - q(\theta; \mathbf{x}^{(k)}) \right)^2 \quad (8)$$

has to be minimized. This corresponds to a minimization of the variance of the random error terms ε . The minimizing choice of θ is called a *least squares estimate* and is denoted by $\hat{\theta}$.

The above regression problem becomes more simple to deal with when the response surface model is linear in its parameters θ . Let us assume that the response surface is given by (Fig. 4)

$$\eta = \theta_1 q_1(\mathbf{x}) + \theta_2 q_2(\mathbf{x}) + \dots + \theta_p q_p(\mathbf{x}) \quad (9)$$

The least squares estimate of the parameter vector is given by

$$\hat{\theta} = (\mathbf{Q}'\mathbf{Q})^{-1}\mathbf{Q}'\mathbf{z} \quad (10)$$

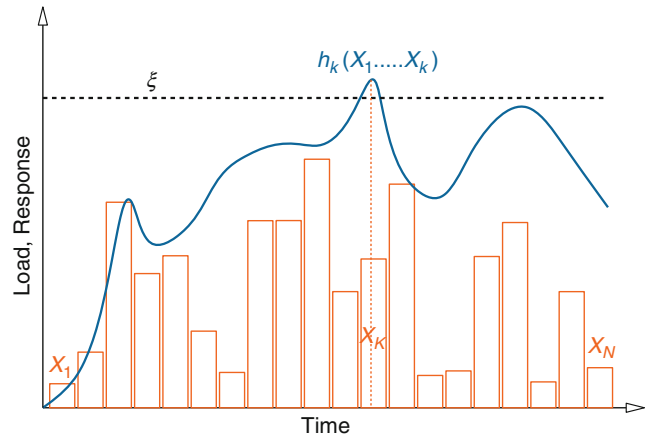
This estimator is unbiased, i.e.,

$$\mathbf{E}[\hat{\theta}] = \theta \quad (11)$$

A detailed discussion is given by (Bucher and Macke 2005).

Structural Optimization Under Random Dynamic Seismic Excitation,

Fig. 5 Schematic sketch of first-passage problem



Monte Carlo Simulation

First-Passage Problem

Generally, the probability of failure P_F in an n -dimensional space of random variables $X_1, \dots, X_n$ can be computed as

$$P_F = \int \cdots \int_{D_F} f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_1 \cdots dx_n \quad (12)$$

In this equation, $f_{X_1, \dots, X_n}(x_1, \dots, x_n)$ denotes the joint probability function of the random variables $X_1, \dots, X_n$, and D_F denotes the failure domain, i.e., the region of the n -dimensional random variable space in which failure occurs. Typically, this is denoted in terms of a scalar limit state function $g(\cdot)$ attaining negative values, i.e., $D_F = \{(X_1, \dots, X_n) \mid g(X_1, \dots, X_n) \leq 0\}$. For applications in time-dependent problems as those arising structural dynamics, the appropriate formulation leads to the first-passage problem. Here the failure domain is typically written as

$$D_F = \left\{ (X_1, \dots, X_n) \mid \max_{1 \leq k \leq N} h_k(X_1, \dots, X_n) \geq \xi \right\} \quad (13)$$

In this equation, $h(\cdot)$ denotes a response quantity of interest, and ξ is a critical threshold value of this response. The random variables X_i usually denote the random excitation (e.g., earthquake or wind) which is discretized in time. A schematic sketch is shown in Fig. 5. Note that due to the

principle of causality, the values of h at point k in time can only depend on the basic variables $X_1 \dots X_k$ (i.e., on those in the past put to the present as expressed by the index k) and not on those in the future.

The generalized safety index (or reliability index) β is defined by

$$\beta = \Phi^{-1}(1 - P_F) = \Phi^{-1}[F_H(h)] \quad (14)$$

Here $\Phi^{-1}(\cdot)$ is the inverse standardized Gaussian distribution function. Without loss of generality, it is assumed that the random variables X_i are Gaussian and that they are independent and identically distributed (i.i.d.) with zero mean and unit standard deviation.

Since the failure probabilities to be computed are usually very small (e.g., $PF = 10^{-5}$), it is not feasible to evaluate the integral in Eq. 12 using standard numerical integration procedures. This is due to the fact that the number of integration points required to perform, e.g., Gaussian integration in dimension n , grows exponentially with n . Monte Carlo methods do not have this dependence on the dimension; however, crude or plain Monte Carlo simulation requires a number of samples roughly larger than $\frac{10}{P_F}$. This, again, may lead to prohibitively large computational effort.

Asymptotic Sampling Method

The asymptotic sampling method as described in Bucher (2009a, b) exploits the asymptotic

behavior of the failure probability expressed in terms of the reliability index β when changing the standard deviation of the basic random variables. If the original standard Gaussian random variables are replaced by variables with non-unit standard deviations $\sigma = \frac{1}{f}$, then the computed reliability index will depend on the choice of f . As a first simple case, consider a linear function of the basic random variables X_k , say

$$g(\mathbf{X}) = \sum_{k=1}^N a_k X_k \quad (15)$$

with arbitrary real-valued coefficients a_k . The random variable $Y = g(\mathbf{X})$ will then be Gaussian with a zero mean and a variance

$$\sigma_Y^2 = \sum_{k=1}^N a_k^2 \quad (16)$$

The distribution function $F_Y(y)$ of this variable will therefore be given by

$$F_Y(\xi) = \Phi\left(\frac{\xi}{\sigma_Y}\right) = \Phi\left(\frac{\xi}{\sqrt{\sum_{k=1}^N a_k^2}}\right) \quad (17)$$

Upon changing the standard deviation of all basic variables from unity to a value of $1/f$, the standard deviation of Y changes by the same amount. Thus, the distribution function changes to

$$F_Y(\xi) = \Phi\left(\frac{\xi f}{\sqrt{\sum_{k=1}^N a_k^2}}\right) \quad (18)$$

from which the generalized reliability index according to Eq. 14 is immediately found as

$$\beta = \frac{\xi f}{\sqrt{\sum_{k=1}^N a_k^2}} \rightarrow \frac{\beta}{f} = \frac{\xi}{\sqrt{\sum_{k=1}^N a_k^2}} \quad (19)$$

This means that for a linear function of Gaussian variables, the scaled reliability index β/f is invariant with respect to the choice of f .

For the general case, this is not true. There is, however, an asymptotic property which ensures similar behavior for many nonlinear cases. As stated by Breitung (1984) and Gollwitzer and Rackwitz (1988), the reliability index asymptotically depends linearly on f or in scaled notation

$$\lim_{f \rightarrow \infty} \frac{\beta(f)}{f} = \text{const.} \quad (20)$$

In order to exploit this asymptotic relation, Bucher (2009b) suggested to utilize the formulation

$$\frac{\beta}{f} = A + \frac{B}{f^2} \quad (21)$$

in which the coefficients A and B are determined from a regression analysis using the values of β as estimated from several Monte Carlo runs with different values of f . It should be mentioned that the asymptotic sampling method is a representative of a very recent family of methods exploiting the dependency of the failure probability on a scaling parameter of the standard deviation of the input variables. A different approach to this is given by Naess and Gaidai (2008) and Naess et al. (2010). For a comparative review and synthesis of these two approaches, it is referred to Qin et al. (2012).

Optimal Design of Seismic Isolation Systems

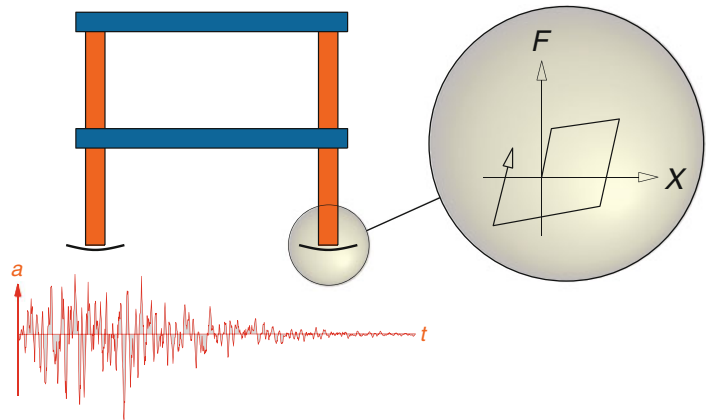
Definition of Problem

In order to ensure structural safety and integrity in earthquake conditions, it may be useful or even necessary to equip structures with protective devices. One possible choice is seismic isolation devices.

The basic scenario in which such devices are utilized is shown in Fig. 6.

In this scenario, a structure is subjected to an earthquake described by the ground acceleration $a(t)$. Its effect on the structure is to be mitigated

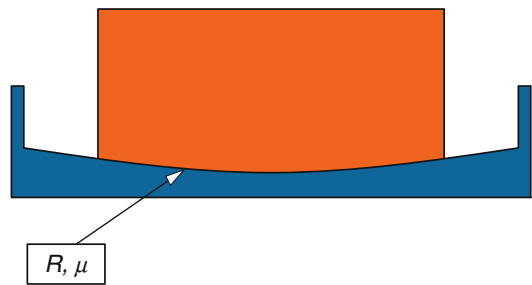
Structural Optimization Under Random Dynamic Seismic Excitation, Fig. 6 Structure and friction-based seismic isolation device



by a device which limits the transfer of forces from the ground to the structure. One such device consists of a combined friction and spring element, in which the spring can also be replaced by a re-centering force due to gravity effects (so-called friction pendulum systems, see, e.g., Roussis and Constantinou (2006)). The device has two characteristic parameters, one is the maximum force transmitted by friction (described by the friction coefficient μ) and the other is the re-centering (or restoring) spring stiffness constant (described by the effective radius of curvature R) (Fig. 7). Both parameters should be chosen such as to ensure the desired protective effect for the structure and the integrity of the isolation device itself.

The question of optimal design for stiffness and friction of seismic isolation systems has been addressed previously in the literature (e.g., Constantinou and Tadjbakhsh (1983), Iemura et al. (2007), Jangid (2005)). The method of equivalent linearization in conjunction with a random process model for the earthquake excitation tuned to the El Centro record was used in Jangid (1996). Here, a full nonlinear dynamic analysis will be used as a basis for the computation of the structural response. The ground motion is modeled as a nonstationary (i.e., evolutionary) random process; hence, the response should be characterized in suitable probabilistic terms.

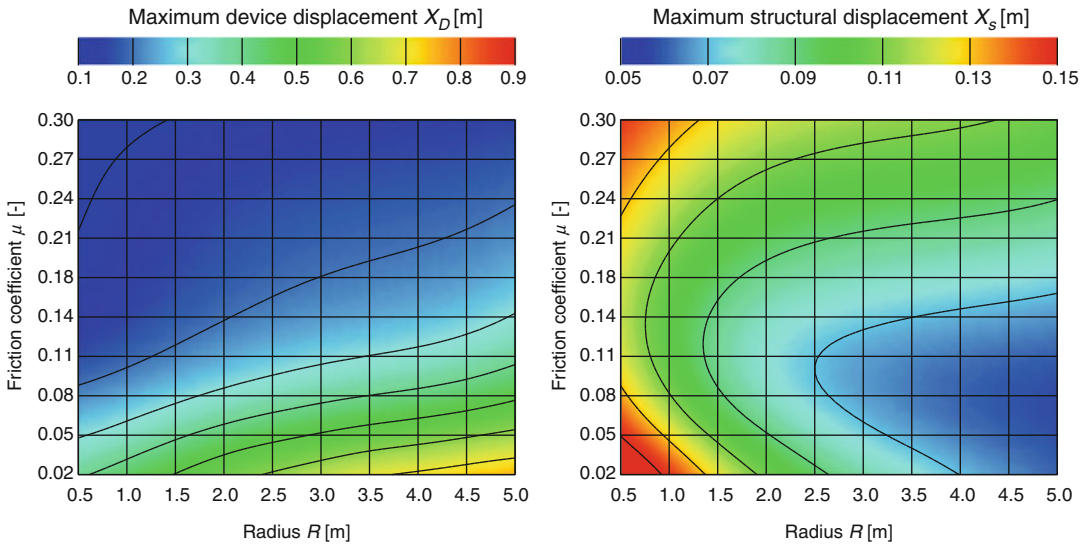
As the main function, if the seismic isolation is to keep the effect of the ground acceleration from the superstructure, it is obvious that a “soft”



Structural Optimization Under Random Dynamic Seismic Excitation, Fig. 7 Simple sketch for friction pendulum system

isolator will be good for that purpose. On the other hand, the isolator itself can have only limited displacements due to constructive reasons. This means that the isolator should be “stiff” enough. Hence, the targets (objectives) of the optimal design of a seismic isolation device are in conflict with each other. As mentioned above, the determination of the Pareto set allows the preselection of suitable candidates for the final decision-making process.

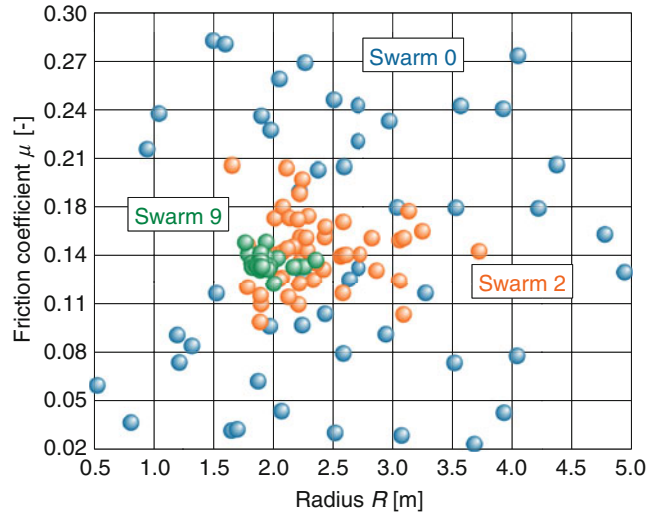
The relevant quantities for the Pareto optimization are the maximum displacement x_D of the friction device (this should be small in order to ensure safety of the supports) and the maximum structural displacement x_S (measured relative to the device, this should be small to ensure seismic isolation). As the quantities are random, characteristic values (i.e., mean value plus three times the respective standard deviation) are chosen for the objectives.



Structural Optimization Under Random Dynamic Seismic Excitation, Fig. 8 Characteristic value of maximum device displacement depending (*left*) and maximum

structural displacement (*right*) depending on design parameters R and μ

Structural Optimization Under Random Dynamic Seismic Excitation, Fig. 9 Convergence of particle swarm optimization method to one point in the Pareto set



Selected Results

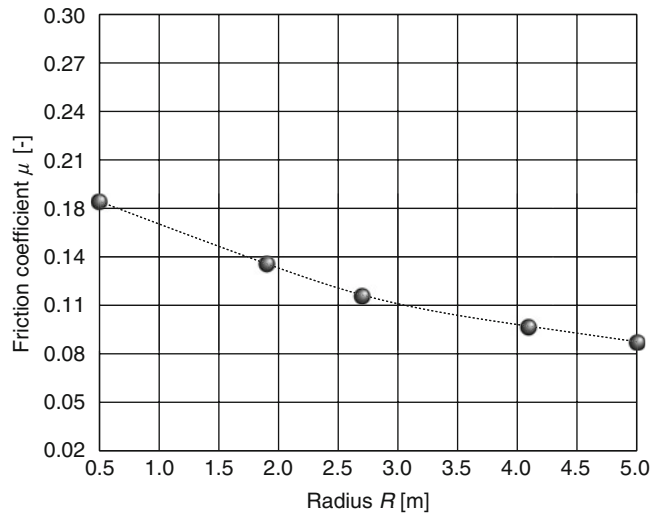
In order to prepare the optimization, the characteristic values for x_D and x_S are computed for 100 randomly selected combinations of the design variables R and μ . At each sample point, the structural system is analyzed for 100 randomly generated earthquakes (using a nonstationary Kanai-Tajimi model (Bucher 2009c)), and the characteristic values for x_D and x_S are computed using sample statistics. These

discrete values are utilized to form response surface approximations which by inherent smoothing largely eliminate the sampling uncertainty inherent in Monte Carlo methods. The resulting response surfaces are shown in Fig. 8.

An aggregated objective function R is then generated as

$$R = \alpha x_D + (1 - \alpha) x_S \quad (22)$$

**Structural Optimization
Under Random Dynamic
Seismic Excitation,
Fig. 10** Final Pareto set



in which α is a real number in the interval $[0, 1]$. Solving the aggregated problem results in one point of the Pareto set. The convergence of this optimization using a particle swarm optimizer is shown in Fig. 9.

Finally, the results of aggregated optimization runs for different values of α yield the Pareto set as shown in Fig. 10.

It turns out that the Pareto set contains very large friction coefficients which are very difficult to realize in practice. Hence, the optimal solutions may not be accessible.

Cross-References

- [Code-Based Design: Seismic Isolation of Buildings](#)
- [Nonlinear Dynamic Seismic Analysis](#)
- [Reliability Estimation and Analysis for Dynamical Systems](#)
- [Robust Design Optimization for Earthquake Loads](#)
- [Stochastic Analysis of Nonlinear Systems](#)

References

- Bamer F, Bucher C (2012) Application of the proper orthogonal decomposition for linear and nonlinear structures under transient excitations. *Acta Mech* 223:2549–2563
- Box GEP, Draper NR (1987) *Empirical model-building and response surfaces*. Wiley, New York
- Box GEP, Wilson KB (1951) On the experimental attainment of optimum conditions. *J R Stat Soc Ser B* 13:1–45
- Breitung KW (1984) Asymptotic approximations for multinormal integrals. *J Eng Mech* 110(3):357–366
- Brenner CE, Bucher C (1995) A contribution to the SFE-based reliability assessment of nonlinear structures under dynamic loading. *Probabilist Eng Mech* 10:265–273
- Bucher C (2009a) Computational analysis of randomness in structural mechanics. In: Frangopol DM (ed) *Structures and infrastructures book series*, vol 3. Taylor & Francis, London
- Bucher C (2009b) Asymptotic sampling for high-dimensional reliability analysis. *Probabilist Eng Mech* 24:504–510
- Bucher C (2009c) Probability-based optimization of friction damping devices. *Struct Saf* 31:500–507
- Bucher C, Macke M (2005) Response surface methodology. In: Nikolaidis E, Ghiocel DM, Singhal S (eds) *Structural reliability handbook*. CRC Press, Boca Raton, p 19, 1–19, 23
- Clough RW, Penzien J (1993) *Dynamics of structures*, 2nd edn. McGraw-Hill, New York
- Constantinou MC, Tadjbakhsh IG (1983) Probabilistic optimum base isolation of structures. *J Struct Eng* 109(3):676–689
- Faravelli L (1989) Response-surface approach for reliability analysis. *J Eng Mech* 115:2763–2781
- Fujimura K, Kiureghian AD (2007) Tail-equivalent linearization method for nonlinear random vibration. *Probabilist Eng Mech* 22:63–76
- Furuta H, Kameda T (2006) Application of multi-objective genetic algorithm to bridge maintenance. In: Ceragioli F, Dontchev A, Futura H, Marti K, Pandolfi L (eds) *System modeling and optimization*.

- IFIP International Federation for Information Processing, Boston, pp 139–148
- Gasser M, Schuëller GI (1997) Reliability-based optimization of structural systems. *Math Method Oper Res* 46:287–307
- Gollwitzer R, Rackwitz R (1988) An efficient numerical solution to the multinormal integral. *Probabilist Eng Mech* 3(2):98–101
- Hill WJ, Hunter WG (1966) A review of response surface methodology: a literature survey. *Technometrics* 8:571–590
- Iemura H, Taghikhany T, Jain S (2007) Optimum design of resilient sliding isolation system for seismic protection of equipments. *Bull Earthquake Eng* 5(1):85–103
- Jangid RS (1996) Optimum damping in a non-linear base isolation system. *J Sound Vibration* 189(4):477–487
- Jangid RS (2005) Optimum friction pendulum system for near-fault motions. *Eng Struct* 27(3):349–359
- Kim S-H, Na S-W (1997) Response surface method using vector projected sampling points. *Struct Saf* 19:3–19
- Liu M, Frangopol DM (2005) Bridge annual maintenance prioritization under uncertainty by multiobjective combinatorial optimization. *Comp-Aided Civil Inf Eng* 20(5):343–353
- Mead R, Pike DJ (1975) A review of response surface methodology from a biometric view-point. *Biometrics* 31:803–851
- Myers RH (1999) Response surface methodology – current status and future directions. *J Qual Technol* 31:30–44
- Myers RH, Khuri AI, Carter JWH (1989) Response surface methodology: 1966–1988. *Technometrics* 31:137–157
- Naess A, Gaidai O (2008) Monte Carlo methods for estimating the extreme response of dynamical systems. *J Eng Mech* 134(8):628–636
- Naess A, Gaidai O, Batsevych O (2010) Prediction of extreme response statistics of narrow-band random vibrations. *J Eng Mech* 136(3):290–298
- Ngatchou PN, Zarei A, Fox WLJ, El-Sharkawi MA (2008) Pareto multiobjective optimization. In: Lee KY, El-Sharkawi MA (eds) *Modern heuristic optimization techniques*. Wiley-Interscience, Piscataway, pp 189–207
- Ouyopornprasert W, Bucher C, Schuëller GI (1989) On the application of conditional integration in structural reliability analysis. In: Ang AHS, Shinozuka M, Schuëller GI (eds) *Proceeding 5th international conference on structural safety and reliability*. San Francisco, pp 1683–1689
- Qin J, Nishijima K, Faber MH (2012) Extrapolation method for system reliability assessment: a new scheme. *Adv Struct Eng* 15(11):1893–1909
- Rackwitz R (1982) Response surfaces in structural reliability. *Berichte zur Zuverlässigkeitstheorie der Bauwerke*. Heft 67, München
- Roberts JB, Spanos P (2003) *Random vibration and statistical linearization*. Dover Publications, Mineola
- Roussis PC, Constantinou MC (2006) Uplift-restraining friction pendulum seismic isolation system. *Earthquake Eng Struct Dyn* 35(5):577–593
- Zheng Y, Das PK (2000) Improved response surface method and its application to stiffened plate reliability analysis. *Eng Struct* 22:544–551

Structural Reliability Estimation for Seismic Loading

Achintya Haldar

Department of Civil Engineering and
Engineering Mechanics, University of Arizona,
Tucson, AZ, USA

Synonyms

Energy dissipation in dynamic analysis; Finite element analysis; Implicit performance function; Nonlinear analysis; Risk management; Seismic loading; Structural reliability; Time-domain dynamic analysis; Time-domain reliability analysis

Introduction

The estimation of structural reliability has matured over the years. It has been accepted by the profession that if the risk in engineering design cannot completely be eliminated, it needs to be managed appropriately. The issue has attracted added significance since the basic design philosophy has changed from human safety to structural safety. This is in response to limit enormous amount of damage caused to infrastructures during recent earthquakes in China, Chile, Haiti, India, Iran, Japan, the USA, and other parts of the world. The word “structure” is used here in a generic sense. It represents real engineered systems including structures in a nuclear power plant, multistory buildings, onshore and offshore structures, soil–pile interaction problems in offshore mooring systems, aerospace structures, computer

packaging systems, systems represented by finite elements, etc. Most of the discussions made here are appropriate for different forms of dynamic including seismic loadings.

The estimation of the probability of failure, expressed in terms of risk or reliability, implied that it needs to be estimated just before failure indicating a very complicated process a structure has to go through before failure. The appropriate risk management requires an acceptable reliability method considering all major loads and load combinations that may act on the structures during their lifetime, modeling the structures appropriately satisfying the underlying physics at the time of failure and analyzing their nonlinear behavior as realistically as possible just before failure.

If the performance functions or limit state functions are explicit, i.e., if they can be expressed as explicit functions of all the random variables present in the formulation and the performance requirements, there are several methods with various degrees of sophistication that are available (Haldar and Mahadevan 2000a). The most commonly used procedures are the first-order and second-order reliability method commonly known as FORM/SORM. They are also referred to as the Rackwitz-Fiessler (1976) algorithm. Although not always necessary, for the most efficient implementation of the procedures require the performance functions are available in explicit forms. The algorithms are iterative in nature and the gradients of performance or limit state functions are required to estimate the coordinates of the most probable failure point (MPFP) and the corresponding reliability index and failure probability. However, their applicability in estimating the reliability of real structural systems could be very limited. For routine applications of FORM/SORM, the derivatives of the performance functions are difficult to obtain even for a simple structure like a two-dimensional small frame. Furthermore, risk analysis methods currently available were developed for relatively simple systems with numerous assumptions which cannot be satisfied for large realistic structural systems. For such systems, the limit states are expected to be implicit, and for

dynamic nonlinear problems, they are also to be a function of time. In response to questions on future research directions in the risk evaluation techniques, Rackwitz (2000) commented that the use of FORM/SORM complemented by Monte Carlo simulations (MCS) would be the next frontier. Several recent studies reflected this idea and simulation has become an integral part of reliability evaluation studies.

Challenges in the Reliability Estimation for Seismic Loading

Reliability estimation for dynamic or seismic loading is evolving. The probability of failure implies that it needs to be estimated just before failure developing various sources of nonlinearities. The finite element method (FEM) is commonly used by the deterministic community to study nonlinear problems where dynamic loadings are applied in time domain. This clearly indicates that an FEM-based general purpose reliability evaluation method, parallel to the deterministic analysis, is necessary.

The most sophisticated deterministic dynamic analysis of structures requires that the load should be applied in time domain. This is one of the major challenges in the reliability analysis for seismic loading. The classical random vibration-based approaches were used in the past for this purpose; however, they did not provide information acceptable to the deterministic community. The classical random vibration-based approaches have numerous limitations including the loads which are applied in the form of power spectral density functions essentially appropriate for linear structural behavior, the uncertainty in the linear or nonlinear structural behavior which may need to be incorporated in approximate ways, several performance-enhancing features currently introduced in structures which cannot be incorporated in appropriate ways, etc. The most severe weakness is that the seismic loading cannot be applied in time domain.

Structural elements (beams, columns, connections, etc.) are generally designed first for

strength according to design guidelines given in codes satisfying some underlying reliability requirements, although unknown to most designers. However, the overall system reliability in strength considering the reliabilities of all the elements remains unknown. It is not addressed in codes since it is difficult to estimate and can only be assessed based on numerous idealistic assumptions (brittle or ductile behavior, dependency of failure of elements, consideration of nonlinear limit states, etc.) (Haldar and Mahadevan 2000a). Moreover, the design satisfying strength requirements may not satisfy the global performance or serviceability requirement, e.g., excessive lateral deflection, a major cause of structural failure for seismic loading. Serviceability requirements also may control the seismic design in most cases and should not be ignored (Huh and Haldar 2002).

Physics-Based Dynamic Behavior: Connection Conditions

Structural members are connected to each other using various types and forms of connections, specifically for steel structures. These connections are routinely modeled as fully restrained (FR). However, extensive experimental studies indicate that they are essentially partially restrained (PR) connections with different rigidities. In a deterministic analysis, PR connections add a major source of nonlinearity by decreasing the overall stiffness of a frame. In a dynamic analysis, it changes the dynamic properties of the structures and adds a major source of energy dissipation. In reliability analysis, it also adds a major source of uncertainty. Thus, the reliability of steel frames with PR connections and subjected to seismic excitation is expected to be quite different than that of frames with FR connections. The implications of the presence of PR connections and the uncertainty in modeling them are essential for the reliability analysis.

Static Versus Dynamic or Seismic Analyses: Improved FEM-Based Formulation

The modeling of structures to extract response information under the static and dynamic application of loadings is expected to be very different.

Under dynamic or seismic loading, the structural response will depend on many factors including the changes in the structural dynamic properties (stiffness, frequency, damping, etc.) and the corresponding response amplification as the loading progresses from linear to nonlinear stages; the sophistication in the physics-based structural modeling, i.e., FR versus PR connections and support conditions; the various sources of energy dissipation intentionally introduced in the system to improve its behavior; etc.

One basic limitation of reliability analysis under seismic loading is the incapability of handling structural behavior appropriately. Because the analysis is based on tracking the uncertainty propagating through the steps of deterministic analysis (Haldar and Mahadevan 2000b), the efficiency of the deterministic FEM is extremely important. It has been reported in the literature (Kondoh and Atluri 1987; Shi and Atluri 1988; Haldar and Nee 1989) that the assumed stress-based FEM has many advantages over the commonly used displacement-based FEM, particularly for nonlinear analysis of frame structures capable of producing large deformation. In the assumed stress-based FEM, the tangent stiffness can be expressed in an explicit form, the stresses of an element can be obtained directly, fewer elements are required to describe a large deformation configuration, and integration is not required to obtain the tangent stiffness. Different sources of nonlinearity, especially due to the presence of partially restrained (PR) connections, can be incorporated without losing the basic simplicity. It is very accurate and efficient in analyzing nonlinear responses. Some of its essential features in the context of the dynamic problems are discussed very briefly in the following sections.

Unified Stochastic Finite Element Method (SFEM)

As discussed earlier, to capture a realistic structural behavior, an FEM-based formulation is desirable. When uncertainty in the variables in the formulation is considered, it leads to the concept of stochastic finite element method (SFEM) (Haldar and Mahadevan 2000b).

Without losing any generality, the limit state function can be expressed in terms of the set of basic random variables $\mathbf{x}$ (e.g., loads, material properties, and structural geometry), the set of displacements $\mathbf{u}$, and the set of load effects $\mathbf{s}$ (except the displacements) such as internal forces, stresses, etc. The displacement $\mathbf{u} = \mathbf{QD}$, where $\mathbf{D}$ is the global displacement vector and $\mathbf{Q}$ is a transformation matrix. The limit state function can be expressed as $g(\mathbf{x}, \mathbf{u}, \mathbf{s}) = 0$. For reliability computation, it is convenient to transform $\mathbf{x}$ into the standard normal space $\mathbf{y} = \mathbf{y}(\mathbf{x})$ such that the elements of $\mathbf{y}$ are statistically independent and have a standard normal distribution. An iterative algorithm can be used to locate the design point or MPFP on the limit state function using the first-order approximation. During each iteration, the structural response and the response gradient vectors are calculated using finite

element models. The following iteration scheme can be used for finding the coordinates of the design point:

$$\mathbf{y}_{i+1} = \left[\mathbf{y}_i^t \alpha_i + \frac{g(\mathbf{y}_i)}{|\nabla g(\mathbf{y}_i)|} \right] \alpha_i \quad (1)$$

where

$$\nabla g(\mathbf{y}) = \left[\frac{\partial g(\mathbf{y})}{\partial y_1}, \dots, \frac{\partial g(\mathbf{y})}{\partial y_n} \right]^t \quad \text{and} \quad \alpha_i = -\frac{\nabla g(\mathbf{y}_i)}{|\nabla g(\mathbf{y}_i)|} \quad (2)$$

To implement the algorithm, the gradient $\nabla g(\mathbf{y})$ of the limit state function in the standard normal space can be derived as (Haldar and Mahadevan 2000b):

$$\nabla g(\mathbf{y}) = \left[\frac{\partial g(\mathbf{y})}{\partial \mathbf{s}} \mathbf{J}_{s,x} + \left(\mathbf{Q} \frac{\partial g(\mathbf{y})}{\partial \mathbf{u}} + \frac{\partial g(\mathbf{y})}{\partial \mathbf{s}} \mathbf{J}_{s,D} \right) \mathbf{J}_{D,x} + \frac{\partial g(\mathbf{y})}{\partial \mathbf{x}} \right] \mathbf{J}_{y,x}^{-1} \quad (3)$$

where $\mathbf{J}_{i,j}$'s are the Jacobians of transformation (e.g., $\mathbf{J}_{s,x} = \partial \mathbf{s} / \partial \mathbf{x}$) and y_i 's are statistically independent random variables in the standard normal space. The evaluation of the quantities in Eq. 3 will depend on the problem under consideration (linear or nonlinear, two or three dimensional, etc.) and the performance functions used. The essential numerical aspect of SFEM is the evaluation of three partial derivatives, $\partial g / \partial \mathbf{s}$, $\partial g / \partial \mathbf{u}$, and $\partial g / \partial \mathbf{x}$ and four Jacobians, $\mathbf{J}_{s,x}$, $\mathbf{J}_{s,D}$, $\mathbf{J}_{D,x}$, and $\mathbf{J}_{y,x}$. They can be evaluated by procedures suggested by Haldar and Mahadevan (2000b) for linear and nonlinear, two- or three-dimensional structures. Once the coordinates of the design point $\mathbf{y}^*$ are evaluated with a preselected convergence criterion, the reliability index β can be evaluated as:

$$\beta = \sqrt{(\mathbf{y}^*)^t (\mathbf{y}^*)} \quad (4)$$

The probability of failure, P_f , can be calculated as:

$$P_f = \Phi(-\beta) = 1.0 - \Phi(\beta) \quad (5)$$

where Φ is the standard normal cumulative distribution function. Equation 5 can be considered as a notational failure probability. When the reliability index is larger, the probability of failure will be smaller. The author and his team published numerous papers to validate the above procedure for static application of the loading.

Dynamic Governing Equation for SFEM

The nonlinear dynamic equilibrium equation can be expressed at time $t + \Delta t$ as:

$$\begin{aligned} & \mathbf{M}^{t+\Delta t} \ddot{\mathbf{D}}^{(k)} + {}^t\mathbf{C}^{t+\Delta t} \dot{\mathbf{D}}^{(k)} + {}^t\mathbf{K}^{(k)} \mathbf{D}^{(k)} \\ &= {}^{t+\Delta t}\mathbf{F}^{(k)} - {}^{t+\Delta t}\mathbf{R}^{(k-1)} - \mathbf{M}^{t+\Delta t} \ddot{\mathbf{D}}_g^{(k)} \end{aligned} \quad (6)$$

where $\mathbf{M}$ is the mass matrix, ${}^t\mathbf{C}$ is the viscous damping coefficient matrix at time t , ${}^t\mathbf{K}^{(k)}$ is the global tangent stiffness matrix of the k^{th} iteration

at time t , ${}^{t+\Delta t}\Delta\mathbf{D}^{(k)}$ is the incremental displacement vector of the k^{th} iteration at time $t + \Delta t$, ${}^{t+\Delta t}\mathbf{F}^{(k)}$ is the external load vector of the k^{th} iteration at time $t + \Delta t$, ${}^{t+\Delta t}\mathbf{R}^{(k-1)}$ is the internal force vector of the $(k-1)^{\text{th}}$ iteration at time $t + \Delta t$, and ${}^{t+\Delta t}\ddot{\mathbf{D}}_g^{(k)}$ is the seismic ground acceleration vector of the k^{th} iteration at time $t + \Delta t$.

The damping matrix ${}^t\mathbf{C}$ in Eq. 6 is considered to be viscous. In a realistic seismic analysis of frames, the amount of damping energy to be generated will depend on the non-yielding and yielding state of the material and the hysteretic behavior if the material yields. For mathematical simplicity, the effect of non-yielding energy dissipation is generally represented by equivalent viscous damping varying between 0.1 % and 7 % of the critical damping (Leger and Dussault 1992). Among many alternatives, Rayleigh-type damping is utilized in this study. It can be represented as:

$${}^t\mathbf{C} = \alpha\mathbf{M} + \gamma{}^t\mathbf{K} \quad (7)$$

where ${}^t\mathbf{K}$ is the tangent stiffness matrix and α and γ are proportional constants which can be evaluated from the natural frequencies of the structure. Since Rayleigh-type damping is proportional to mass and stiffness matrixes, it is more appropriate for nonlinear behavior as it incorporates information on changes in the stiffness values.

Numerical Procedures

The following numerical procedures can be followed to implement SFEM concept for seismic or dynamic loading. The step-by-step direct integration numerical procedure using the Newmark- β method, with parameters $\eta = 1/2$ and $\beta_N = 1/4$, can be used to solve Eq. 6. The iterative scheme can be expressed as:

$${}^t\mathbf{K}_D {}^{t+\Delta t}\Delta\mathbf{D}^{(k)} = {}^{t+\Delta t}\mathbf{F}^{(k)} - {}^{t+\Delta t}\mathbf{R}^{(k-1)} \quad (8)$$

where

$${}^{t+\Delta t}\mathbf{F}^{(k)} = {}^{t+\Delta t}\mathbf{F}^{(k-1)} + {}^{t+\Delta t}\Delta\mathbf{F}_D^{(k)} \quad (9)$$

and ${}^t\mathbf{K}_D$ is the dynamic tangent stiffness matrix and can be shown to be:

$${}^t\mathbf{K}_D = f_1\mathbf{M} + f_2{}^t\mathbf{K} \quad (10)$$

${}^{t+\Delta t}\mathbf{F}^{(k-1)}$ and ${}^{t+\Delta t}\Delta\mathbf{F}_D^{(k)}$ are the modified external force and its incremental vector, respectively. The modified external force vector can be expressed as:

$${}^{t+\Delta t}\mathbf{F}^{(k-1)} = {}^{t+\Delta t}\mathbf{F}^{(k-1)} + {}^{t+\Delta t}\mathbf{P}^{(k-1)} - \mathbf{M}{}^{t+\Delta t}\ddot{\mathbf{D}}^{(k-1)} \quad (11)$$

And ${}^{t+\Delta t}\mathbf{R}^{(k-1)}$ is the internal force vector of the system. The term ${}^{t+\Delta t}\mathbf{P}^{(k-1)}$ in Eq. 11 is the modified force vector contributed by the displacement, velocity, and acceleration at time t and the displacement at time $t + \Delta t$ and can be written as

$$\begin{aligned} {}^{t+\Delta t}\mathbf{P}^{(k-1)} = \mathbf{M} & \left[f_1{}^t\ddot{\mathbf{D}} + f_3{}^t\dot{\mathbf{D}} + f_4{}^t\ddot{\mathbf{D}} - f_1{}^{t+\Delta t}\ddot{\mathbf{D}}^{(k-1)} \right] \\ & + {}^t\mathbf{K} \left[f_5{}^t\mathbf{D} + f_6{}^t\dot{\mathbf{D}} + f_7{}^t\ddot{\mathbf{D}} - f_5{}^{t+\Delta t}\mathbf{D}^{(k-1)} \right] \end{aligned} \quad (12)$$

The incremental external force term ${}^{t+\Delta t}\Delta\mathbf{F}_D^{(k)}$ can be shown to be:

$${}^{t+\Delta t}\Delta\mathbf{F}_D^{(k)} = {}^{t+\Delta t}\Delta\mathbf{F}^{(k)} - \mathbf{M}{}^{t+\Delta t}\Delta\ddot{\mathbf{D}}_g^{(k)} \quad (13)$$

The coefficients, f_i 's, are constants and can be evaluated (Haldar and Nee 1989) in terms of η , α , β_N , γ , and Δt as:

$$\begin{aligned} f_1 &= \frac{1}{\beta_N\Delta t^2} + \frac{\eta\alpha}{\beta_N\Delta t}, \quad f_2 = \frac{\eta\gamma}{\beta_N\Delta t} + 1, \\ f_3 &= \frac{1}{\beta_N\Delta t} + \frac{\eta\alpha}{\beta_N} - \alpha, \\ f_4 &= \left(\frac{1}{2\beta_N} - 1 \right) + \eta\alpha \left(\frac{1}{2\beta_N} - \frac{1}{\eta} \right) \Delta t, \\ f_5 &= \frac{\eta\alpha}{\beta_N\Delta t}, \quad f_6 = \frac{\eta\gamma}{\beta_N} - \gamma, \quad f_7 = \left(\frac{\eta\gamma}{2\beta_N} - \gamma \right) \Delta t \end{aligned} \quad (14)$$

Equation 8 now can be solved using the modified Newton-Raphson method. Once the displacements are obtained, the member forces can be calculated accordingly.

Consideration of PR Connections in the Finite Element Formulation

The presence of partial connection rigidity needs to be incorporated into the deterministic analysis of structures to capture their behavior. In general, the relationship between the moment M , transmitted by the connection, and the relative rotation angle θ is used to represent the flexible behavior. Among the many alternatives (Richard model, piecewise linear model, polynomial model, exponential model, B-spline model, etc.), the Richard four-parameter moment–rotation model is chosen here to represent the flexible behavior of a connection. It can be expressed as (Richard and Abbott 1975):

$$M = \frac{(k - k_p)\theta}{\left(1 + \left|\frac{(k - k_p)\theta}{M_0}\right|^N\right)^{1/N}} + k_p\theta \quad (15)$$

where M is the connection moment, θ is the relative rotation between the connecting elements, k is the initial stiffness, k_p is the plastic stiffness, M_0 is the reference moment, and N is the curve shape parameter. These parameters are identified in Fig. 1.

Although an ordinary beam-column element is used to represent a PR connection element for numerical analyses, its stiffness needs to be updated at each iteration since the stiffness representing the partial rigidity depends on θ . This can be accomplished by updating the Young's modulus as:

$$E_C(\theta) = \frac{l_C}{I_C} K_C(\theta) = \frac{l_C}{I_C} \frac{\partial M(\theta)}{\partial \theta} \quad (16)$$

where l_C , I_C , and $K_C(\theta)$ are the length, the moment of inertia, and the tangent stiffness of the connection element, respectively. $K_C(\theta)$ is calculated using Eq. 15 and can be shown to be:

$$K_C(\theta) = \frac{dM}{d\theta} = \frac{(k - k_p)}{\left(1 + \left|\frac{(k - k_p)\theta}{M_0}\right|^N\right)^{\frac{N+1}{N}}} + k_p \quad (17)$$

The Richard model discussed up to now represents only the monotonically increasing loading portion of the M – θ curves. However, the unloading and reloading behavior of the M – θ curves is also essential for any nonlinear seismic analysis. This subject was extensively addressed in the literature (Colson 1991; El-Salti 1992). They theoretically developed the unloading and reloading parts of the M – θ curves using the Masing rule. A general class of Masing models can be defined with a virgin loading curve expressed as:

$$f(M, \theta) = 0 \quad (18)$$

and its unloading and reloading curve can be described by the following equation:

$$f\left(\frac{M - M_a}{2}, \frac{\theta - \theta_a}{2}\right) = 0 \quad (19)$$

where (M_a, θ_a) is the load reversal point as shown in Fig. 1.

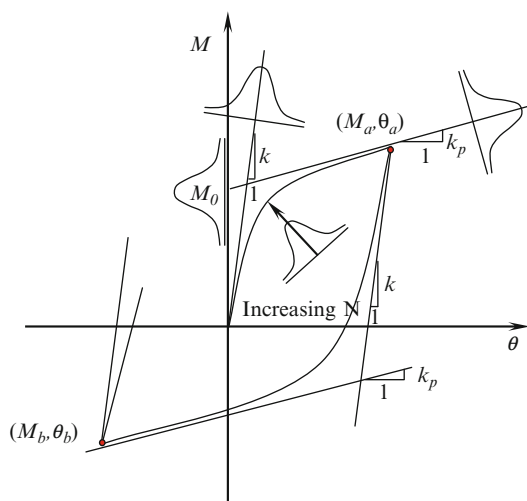
Using the Masing rule and the Richard model represented by Eqs. 15 and 17, the unloading and reloading behavior of a PR connection can be generated as:

$$M = M_a - \frac{(k - k_p)(\theta_a - \theta)}{\left(1 + \left|\frac{(k - k_p)(\theta_a - \theta)}{2M_0}\right|^N\right)^{1/N}} - k_p(\theta_a - \theta) \quad (20)$$

and

$$K_C(\theta) = \frac{dM}{d\theta} = \frac{(k - k_p)}{\left(1 + \left|\frac{(k - k_p)(\theta_a - \theta)}{2M_0}\right|^N\right)^{\frac{N+1}{N}}} + k_p \quad (21)$$

If (M_b, θ_b) is the next load reversal point, as shown in Fig. 1, the reloading relation between M and θ can be obtained by simply replacing (M_a, θ_a) with (M_b, θ_b) in Eqs. 20 and 21. Thus, the proposed method uses Eqs. 15 and 17 when the connection is loading and Eqs. 20 and 21 when



Structural Reliability Estimation for Seismic Loading, Fig. 1 M - θ curve using the Richard model, Masing rule, and uncertainty (Huh and Haldar 2002)

the connection is unloading and reloading. This represents hysteretic behavior at the PR connections.

The basic FEM formulation of the structure remains unchanged, and thus the incorporation of the PR connection can be successfully accomplished.

Pre- and Post-Northridge PR Connections

During the Northridge earthquake of 1994, several connections in steel frames fractured in a brittle and premature manner. A typical connection, shown in Fig. 2, was fabricated with the beam flanges attached to the column flanges by full penetration welds (field welded) and with the beam web bolted (field bolted) to single plate shear tabs (Richard and Radau 1998), denoted hereafter as the pre-NC.

In the post-Northridge design practices, the thrusts were to make the connections more flexible than the pre-NC and to move the location of formation of any plastic hinge away from the connection and to provide more ductility to increase the energy absorption capacity. Several improved connections can be found in the literature including cover-plated connections, spliced beam

connections, side-plated connections, bottom haunch connections, connections with vertical ribs, and connections with a reduced beam sections (RBS) or dog boned (FEMA 350–3 2000). Seismic Structural Design Associates, Inc. (SSDA) proposed a unique proprietary slotted web (SSDA SlottedWebTM) moment connection (Richard et al. 1997), as shown in Fig. 3, denoted hereafter as the post-NC. The author was given access to some of the actual SSDA full-scale test results. Using the four parameters Richard model discussed earlier, Mehrabian et al. (2005) proposed a mathematical model to represent moment-relative rotation (M - θ) curves for this type of connections. Once (M - θ) curves are available, they can be used for the reliability evaluation, as discussed in the previous section. This is very important that this type of performance-enhancing feature is incorporated in any reliability evaluation.

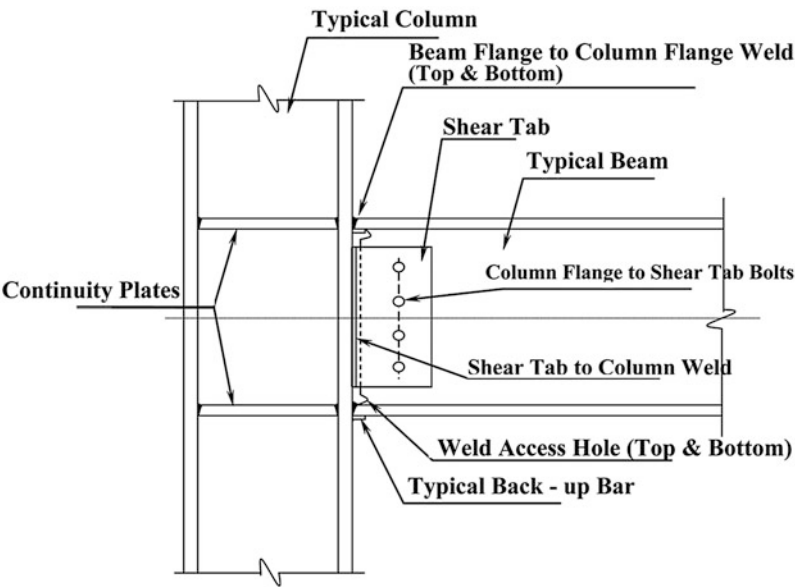
Uncertainties in the Connection Model

A considerable amount of the uncertainty in the connection behavior comes from the uncertainties in the manufacturing and assembling processes and from the mathematical modeling (Haldar and Mahadevan 2000a). Deterministic prediction of connection behavior, based on either empirical formulations or single test data, is likely to overestimate the strength and stiffness. In practice, parameters in a typical M - θ curve are estimated from experimental results using a curve-fitting technique. Therefore, deterministic curves do not account for the scatter in the connection behavior. To consider the uncertainty in modeling the behavior of PR connections, the four parameters in the Richard model can be considered to be the basic random variables as shown in Fig. 1, along with other load- and resistance-related random parameters.

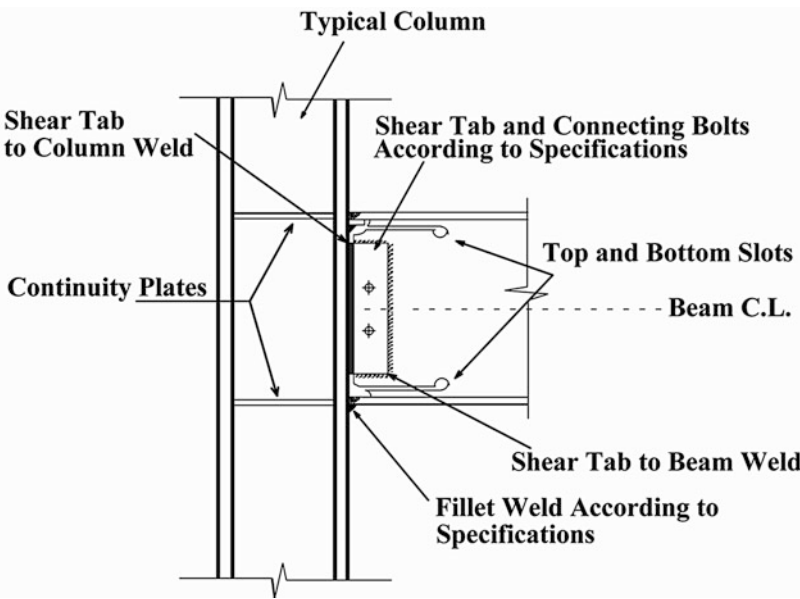
Seismic Reliability Evaluation for Large Structures

Considering all the challenges discussed above, a robust reliability evaluation technique for large

Structural Reliability Estimation for Seismic Loading, Fig. 2 A typical pre-NC (Mehrabian et al. 2005)



Structural Reliability Estimation for Seismic Loading, Fig. 3 A typical post-NC (Mehrabian et al. 2005)



structural systems may not be available at this time. In the presence of such vacuum, one may decide a simple Monte Carlo simulation (MCS) to address the presence of uncertainty. It is observed that one deterministic analysis of a large structure excited by seismic loading applied in time domain satisfying different sources of energy dissipation and satisfying underlying physics may take over 10 h of computer time.

If one decides only say 10,000 cycles of MCS, it will take about 100,000 h or about 11.4 years of continuous running of a computer. Obviously, it will be impractical.

Eliminating the basic MCS as a realistic alternative for large structural systems with relatively moderate probability of failure events, the available computational approaches can be broadly divided into two categories: (i) the

sensitivity-based stochastic finite element method (SFEM) analysis and (ii) the response surface method (RSM). This led to the development of sensitivity-based SFEM formulation (Haldar and Mahadevan 2000b). The other alternative is RSM (Box et al. 1978). The primary purpose of applying RSM in reliability analysis is to approximate the original complex and implicit limit state function using a simple and explicit polynomial (Bucher and Bourgund 1990; Yao and Wen 1996; Khuri and Cornell 1996). The three basic weaknesses of RSM that need to be addressed before applying it for structural reliability evaluations are the following: (1) it cannot incorporate distribution information of random variables even when it is available; (2) if the response surface is not generated in the failure region, it may not be directly applicable or robust; and (3) for large systems, it may not give the optimal sampling points. Thus, a basic RSM-based reliability method may not be acceptable.

At present, a second-order polynomial without and with cross terms are generally used to generate response surfaces. Recently, Li et al. (2001) proposed high-order response surface method (HORSM). The method employs Hermite polynomials and the one-dimensional Gaussian points as sampling points to determine the highest power of each variables. In recent past, several methods with the general objective of approximately developing multivariate expressions for response surface for mechanical engineering applications were proposed. One such method is the high-dimensional model representation (HDMR) (Alis and Rabitz 2001; Rao and Chowdhury 2009; Sobol 2003; Wei and Rahman 2007). It is also referred to as “decomposition method,” “univariate approximation,” “bivariate approximation,” “S-variate approximation,” etc. HDMR captures the high-dimensional relationships between sets of input and output model variables in such a way that the component functions of the approximation are ordered starting from a constant and adding terms such as first order, second order, and so on. The concept appears to be reasonable if higher-order variable correlations are weak,

allowing the physical model to be captured by the first few lower-order terms.

Another major work is known as the explicit design space decomposition (EDSD). It can be used when responses can be classified into two classes, e.g., safe and unsafe. The classification is performed using explicitly defined boundaries in space. A machine learning technique known as support vector machines (SVM) (Basudhar et al. 2008; Layman et al. 2007) is used to construct the boundaries separating distinct classes. The failure regions corresponding to different modes of failure are represented with a single SVM boundary, which is refined through adaptive sampling.

The HORSM, HDMR, and EDSD-SVM approaches use MCS to estimate the underlying reliability. They may not be suitable for reliability evaluation of large structural systems where dynamic loadings need to be applied in time domain and several important features like PR connections, several sources of nonlinearities and energy dissipation features must be explicitly analytically incorporated in the formulation satisfying the underlying physics.

Proposed New Method

The proposed reliability evaluation method for large structural systems is developed in two stages. In the first stage, the two weaknesses of RSM, i.e., the consideration of distributional information of the random variables present in the formulation and identifying the location of the failure region, are addressed by integrating it with FORM/SORM. This approach will lead to a hybrid approach consisting of SFEM, FORM/SORM, and RSM. In the second stage, the efficiency of the method is improved by using several advanced factorial schemes so that the response surface can be generated with fewer sampling points.

Proposed Method: Stage 1

Considering the fact that a higher-order polynomial may result in an ill-conditional system of equations for unknown coefficients and exhibit

an irregular behavior outside of the domain of samples, their utilization in generating RSM has received relatively little attention (Gavin and Yau 2008; Rajashekhar and Ellingwood 1993). For complicated problems considered by the research team of the author, a second-order polynomial, without and with cross terms, is considered to be appropriate. They can be expressed as:

$$\hat{g}(\mathbf{X}) = b_0 + \sum_{i=1}^k b_i X_i + \sum_{i=1}^k b_{ii} X_i^2 \quad (22)$$

$$\hat{g}(\mathbf{X}) = b_0 + \sum_{i=1}^k b_i X_i + \sum_{i=1}^k b_{ii} X_i^2 + \sum_{i=1}^{k-1} \sum_{j>i}^k b_{ij} X_i X_j \quad (23)$$

where X_i ($i = 1, 2, \dots, k$) is the i th random variable and b_0 , b_i , b_{ii} , and b_{ij} are unknown coefficients to be determined. The numbers of coefficient necessary to define Eqs. 22 and 23 are $p = 2k + 1$ and $p = (k + 1)(k + 2)/2$, respectively. The coefficients can be fully defined by estimating deterministic responses at intelligently selected data points called experimental sampling points. The concept behind a sampling scheme can be expressed as:

$$X_i = X_i^C \pm h_i \sigma_{x_i} x_i \quad (24)$$

where X_i^C and σ_{x_i} are the coordinates of the center point and the standard deviation of a random variable X_i , respectively, and h_i is an arbitrary factor that defines the experimental region.

Sampling points are selected around the center point. The selection of the center point and experimental sampling points around it are crucial factors in establishing the efficiency and accuracy of the proposed iterative method. The center point is selected to be the coordinates of the checking points as the iteration continues. In the context of iterative scheme of FORM/SORM, the initial center point $\mathbf{x}_{C_1}$ is selected to be the mean values of the random variable X_i 's. Then, using the responses obtained from the deterministic FEM evaluations for all the experimental sampling points around the center point, the response surface $\hat{g}_1(\mathbf{X})$ can be generated explicitly in terms of the random variables $\mathbf{X}$. Once a closed-form expression for the limit state function is obtained, the coordinates of the checking point $\mathbf{x}_{D_1}$ can be estimated using FORM/SORM, using all the statistical information on the X_i 's, eliminating one major deficiency of RSM. The response can be evaluated again at the checking point $\mathbf{x}_{D_1}$, and a new center point $\mathbf{x}_{C_2}$ can be selected using linear interpolation from the center point $\mathbf{x}_{C_1}$ to $\mathbf{x}_{D_1}$ such that $g(\mathbf{X}) = 0$, i.e.,

$$\mathbf{x}_{C_2} = \mathbf{x}_{C_1} + (\mathbf{x}_{D_1} - \mathbf{x}_{C_1}) \frac{g(\mathbf{x}_{C_1})}{g(\mathbf{x}_{C_1}) - g(\mathbf{x}_{D_1})} \quad \text{if } g(\mathbf{x}_{D_1}) \geq g(\mathbf{x}_{C_1}) \quad (25)$$

$$\mathbf{x}_{C_2} = \mathbf{x}_{D_1} + (\mathbf{x}_{C_1} - \mathbf{x}_{D_1}) \frac{g(\mathbf{x}_{D_1})}{g(\mathbf{x}_{D_1}) - g(\mathbf{x}_{C_1})} \quad \text{if } g(\mathbf{x}_{D_1}) < g(\mathbf{x}_{C_1}) \quad (26)$$

A new center point $\mathbf{x}_{C_2}$ then can be used to develop an explicit performance function for the next iteration. This iteration scheme can be repeated until a preselected convergence criterion of $(\mathbf{x}_{C_{i+1}} - \mathbf{x}_{C_i})/\mathbf{x}_{C_i} \leq \varepsilon$ is satisfied. ε can be considered to be $[0.05]$. In the final iteration, the information on the most recent center point

can be used to formulate the final response surface. FORM/SORM can then be used to calculate the reliability index and the corresponding coordinates of the MPFP.

It is to be noted that since increasing efficiency is a major objective, random variables with low-sensitivity indexes (Haldar and Mahadevan 2000a)

can be considered as deterministic at their mean values without compromising the accuracy of the reliability estimation.

To select experimental sampling points around the center point, saturated design (SD) and central composite design (CCD) are the two most promising schemes. SD is less accurate but more efficient since it requires only as many sampling points as the total number of unknown coefficients to define the response surface. CCD is more accurate but less efficient since a regression analysis is needed to evaluate the unknown coefficients for the response surface. Also, a second-order polynomial with cross terms (Eq. 23) must be used for CCD.

To illustrate the computational effort required for the reliability evaluation of a large structural system, suppose the total number of significant random variables, after making some of them deterministic at their mean values, present in the formulation is $k = 50$. The total number of coefficients necessary to define Eq. 22 will be $2 \times 50 + 1 = 101$ and to define Eq. 23 will be $(50 + 1)(50 + 2)/2 = 1326$. It can also be shown that if Eq. 22 and SD scheme are used to generate the response surface, the total number of sampling points, essentially the total number of deterministic FEM-based time-domain nonlinear response analyses will be 101. However, if Eq. 23 and a full SD scheme are used, the corresponding deterministic analyses will be 1326. If Eq. 23 and CCD scheme are used, the corresponding deterministic analyses will be $2^{50} + 2 \times 50 + 1 > 1.1258999 \times 10^{15}$. Obviously, some of these alternatives may not be meaningful.

Since the proposed algorithm is iterative and the basic SD and CCD require different amounts of computational effort, considering efficiency without compromising accuracy, several schemes can be followed. Among the numerous schemes considered by the research team, one basic and two promising schemes are:

Scheme 0 – SD using second-order polynomial without the cross terms throughout all the iterations.

Scheme 1 – Implement SD using Eq. 22 for the intermediate iterations and SD using Eq. 23 for the final iteration.

Scheme 2 – Implement SD using Eq. 22 for the intermediate iterations and CCD using Eq. 23 for the final iteration.

The efficiency and accuracy of the above schemes will be discussed with the help of an example.

Proposed Method: Stage 2

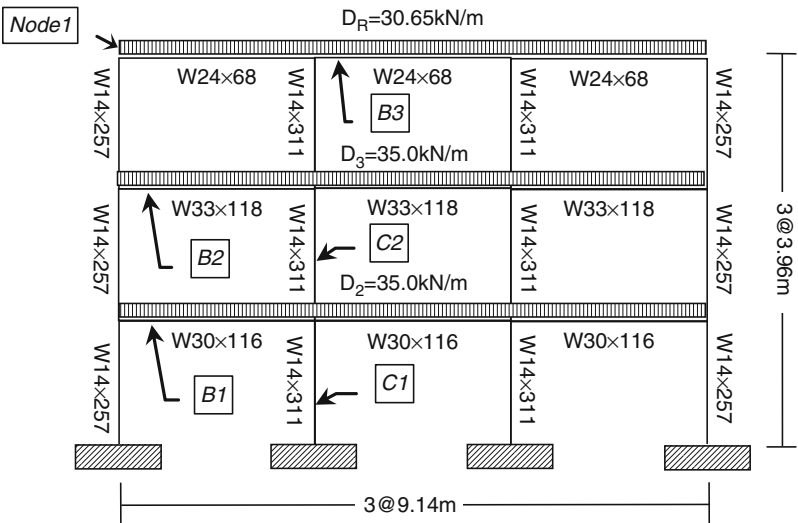
The above two schemes may be major improvements but still may not be able to estimate the reliability of large structural systems. They may require fewer deterministic evaluations but still they could be in thousands. Their efficiency can be improved significantly by reducing the deterministic evaluations in hundreds in the following way.

Scheme M1: To improve the efficiency of Scheme 1, the cross terms (edge points), $k(k - 1)$, are suggested to be added only for the most important variables in the last iteration. Since the proposed algorithm is an integral part of FORM/SORM, all the random variables in the formulation can be arranged in descending order of their sensitivity indexes $\alpha(X_i)$, i.e., $\alpha(X_1) > \alpha(X_2) > \alpha(X_3) \dots \dots > \alpha(X_k)$. The sensitivity of a variable X , $\alpha(X)$, is the directional cosines of the unit normal vector at the design point. In the last iteration, the cross terms are added only for the most sensitive random variables, m , and the corresponding reliability index is calculated. The total number of FEM analyses required for Scheme 1 and M1 are $(k + 1)(k + 2)/2$ and $2k + 1 + m(2k - m - 1)/2$, respectively. For an example, suppose for a large structural system, $k = 50$ and $m = 3$. The total number of required FEM analyses will be 1326 and 245, respectively, for the two schemes; a significant improvement in the efficiency without compromising the accuracy.

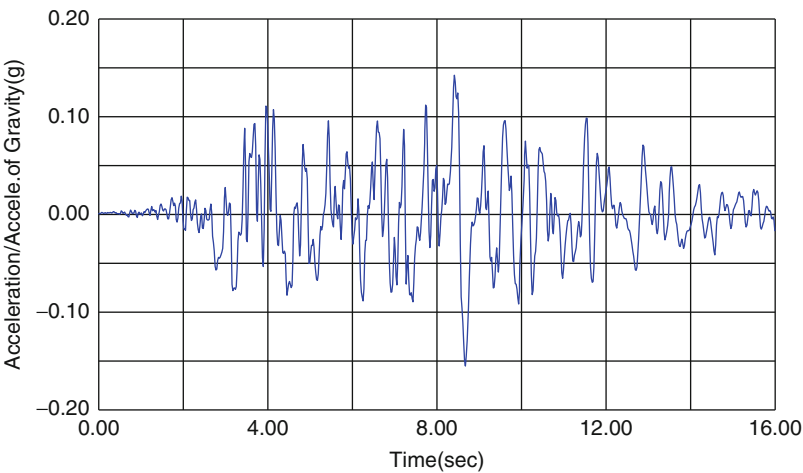
Illustrative Examples

To illustrate the different procedures discussed here for the seismic risk evaluation of a structure, the following illustrative examples are presented.

Structural Reliability Estimation for Seismic Loading, Fig. 4 A three-story three-bay SMRF structure (Haldar et al. 2012)



Structural Reliability Estimation for Seismic Loading, Fig. 5 Earthquake time history used in this study



Reliability of Steel Frames in the Presence of Reinforced Concrete Shear Wall: Seismic Loading is Applied in Time Domain

A three-story three-bay steel frame, as shown in Fig. 4, is considered (Haldar et al. 2012). Section sizes of beams and columns, using A36 steel, are also shown in the figure. The gravity loads acting on the second, third, and roof are given in the figure. A building is supposed to consist of several such frames. It was excited by a seismic time history shown in Fig. 5.

The four parameters of the Richard model are calculated by PRCONN (Richard 1993), a commercially available computer program for both pre-NC and post-NC connections. For the

example under consideration, considering the sizes of columns and beams, three types of connection are necessary. They are denoted as types A, B, and C, hereafter. Four Richard parameters for both pre-NC and post-NC connections are summarized in Table 1.

Limit States of Performance Functions

Both strength and serviceability limit states are used for structural reliability estimation.

Strength limit states: They mainly depend on the failure mode to be considered. Most of the elements in the structural system considered are beam columns, i.e., they are subjected to both axial load and bending moment at the same time.

Structural Reliability Estimation for Seismic Loading, Table 1 Parameters of Richard equation for M-θ curves

Connection assembly type				Connection parameters			
ID		Beam	Column	k ^a	k _p ^a	M ₀ ^b	N
Pre-NC	A	W24 × 68	W14 × 257	2.51 × 10 ⁷	5.56 × 10 ⁵	4.16 × 10 ⁴	1.1
		W24 × 68	W14 × 311				
	B	W33 × 118	W14 × 257	5.08 × 10 ⁷	1.14 × 10 ⁵	6.79 × 10 ⁴	1.1
		W33 × 118	W14 × 311				
	C	W30 × 116	W14 × 257	3.95 × 10 ⁷	9.19 × 10 ⁵	5.65 × 10 ⁴	1.1
		W30 × 116	W14 × 311				
Post-NC	A	W24 × 68	W14 × 257	1.00 × 10 ⁹	4.52 × 10 ⁵	9.64 × 10 ⁴	1.0
		W24 × 68	W14 × 311				
	B	W33 × 118	W14 × 257	2.34 × 10 ⁹	4.52 × 10 ⁵	2.44 × 10 ⁵	1.0
		W33 × 118	W14 × 311				
	C	W30 × 116	W14 × 257	2.14 × 10 ⁹	4.52 × 10 ⁵	2.21 × 10 ⁵	1.0
		W30 × 116	W14 × 311				

^akN · cm/rad^bkN · cm

For design purposes, interaction equations are generally used to consider the combined effect of axial load and bending moment. The interaction equations suggested by the American Institute of Steel Construction's *Load and Resistance Factor Design* manual (AISC 2005) for two-dimensional structures are used for this illustrative example. They can be expressed as:

$$\frac{P_u}{\phi P_n} + \frac{8}{9} \left(\frac{M_{ux}}{\phi_b M_{nx}} \right) \leq 1.0; \quad \text{if } \frac{P_u}{\phi P_n} \geq 0.2 \quad (27)$$

$$\frac{P_u}{2\phi P_n} + \left(\frac{M_{ux}}{\phi_b M_{nx}} \right) \leq 1.0; \quad \text{if } \frac{P_u}{\phi P_n} < 0.2 \quad (28)$$

where ϕ and ϕ_b are the resistance factors, P_u is the required tensile/compressive strength, P_n is the nominal tensile/compressive strength, M_{ux} is the required flexural strength, and M_n is the nominal flexural strength. P_n and M_{nx} can be calculated using AISC's LRFD code (AISC 2005).

Serviceability limit states: For seismic loading, the design may be controlled by the serviceability, e.g., inter-story drift or the overall lateral displacement. The general form of a serviceability limit state can be represented as:

$$g(\mathbf{X}) = \delta_{\text{allow}} - y_{\text{max}}(\mathbf{X}) = \delta_{\text{allow}} - \hat{g}(\mathbf{X}) \quad (29)$$

where δ_{allow} is the allowable inter-story drift or overall lateral displacement specified in codes and $y_{\text{max}}(\mathbf{X})$ is the corresponding maximum inter-story drift or overall lateral displacement estimated from the analysis.

Reliability Evaluations of Frames with Different Connection Conditions and Without RC Shear Wall

The statistical information on all the random variables for the illustrative example is summarized in Table 2. The allowable displacement at the top of the frame, δ_{allow} , is considered to be $h/400$, where h is the height of the frame. For this example, δ_{allow} is 2.97 cm. The corresponding serviceability limit state is defined by Eq. 29 (Huh and Haldar 2011). The probabilities of failure of the frame for the lateral deflection at the top of the frame for serviceability limit state and the strength limit state of the weakest members are first estimated assuming all the connections are of FR type, considered routinely in the profession. The results are summarized in Table 3 using the proposed method. To verify the results, 20,000 MCS for the serviceability and 30,000 MCS for the strength limit states were carried out.

The results clearly indicate that the bare steel frame will not satisfy the serviceability requirement. Then, the reliabilities of the frame are estimated assuming all the connections are post-NC

Structural Reliability Estimation for Seismic Loading, Table 2 Statistical information on the design variables

	Item	Random variable	Mean value	COV	Dist.
Member	All	E(kN/m ²)	1.999 × 10 ⁸	0.06	Ln
		F _y (kN/m ²)	2.482 × 10 ⁵	0.10	Ln
	Column W14 × 257	I _x ^{C1} (m ⁴)	1.415 × 10 ^{−3}	0.05	Ln
		Z _x ^{C1} (m ³)	7.981 × 10 ^{−3}	0.05	Ln
	Column W14 × 311	A ^{C2} (m ²)	5.897 × 10 ^{−2}	0.05	Ln
		I _x ^{C2} (m ⁴)	1.802 × 10 ^{−3}	0.05	Ln
	Beam W33 × 118	I _x ^{B2} (m ⁴)	2.456 × 10 ^{−3}	0.05	Ln
	Beam W30 × 116	I _x ^{B3} (m ⁴)	2.052 × 10 ^{−3}	0.05	Ln
Z _x ^{B3} (m ³)		6.194 × 10 ^{−3}	0.05	Ln	
Seismic load		ξ	0.05	0.15	Type I
		g _e	1.0	0.20	Type I
Connection	Richard model parameter	K ^a	Refer to values in Table 3	0.15	N
		k _p ^a		0.15	N
		M ₀ ^b		0.15	N
		N		0.05	N
Shear wall ^c		E _C (kN/m ²)	2.137 × 10 ⁷	0.18	Ln
		ν	2.137 × 10 ⁷	0.18	Ln

Ln lognormal distribution

^a $\text{kN} \cdot \text{cm/rad}$

^b kN cm

^c $f'_C = 2.068 \times 10^4 \text{ (kN/m}^2\text{)}$

Structural Reliability Estimation for Seismic Loading, Table 3 Reliability evaluation of FR frame

Limit state		Serviceability	Strength limit state	
		Node at 1	Beam (B1)	Column (C1)
MCS	P_f	0.08740	N/A ^a	N/A ^a
	$\beta \approx \Phi^{-1}(1 - P_f)$	1.357	N/A	N/A
	NOS ^b	20,000	30,000	30,000
Proposed algorithm	No. of RV	8	6	6
	Scheme	1	2	2
	β	1.330	4.724	5.402
	Error w.r.t β	1.99 %	N/A	N/A
	TNSP ^c	79	103	103

^aNot a single failure observed for 30,000 cycles of simulation since large reliability indexes are expected in the strength limit state

^bNumber of simulation for deterministic FEM analyses

^cTotal number of sampling points (total number of deterministic FEM analyses)

and pre-NC types and the results are summarized in Table 4. The behavior of the frame in the presence of FR and post-NC for both serviceability and strength limit states are very similar. This was also observed during the full-scale experimental investigations establishing the advanced features of the proposed method. In any case, the lateral stiffness of the frame needs to be increased.

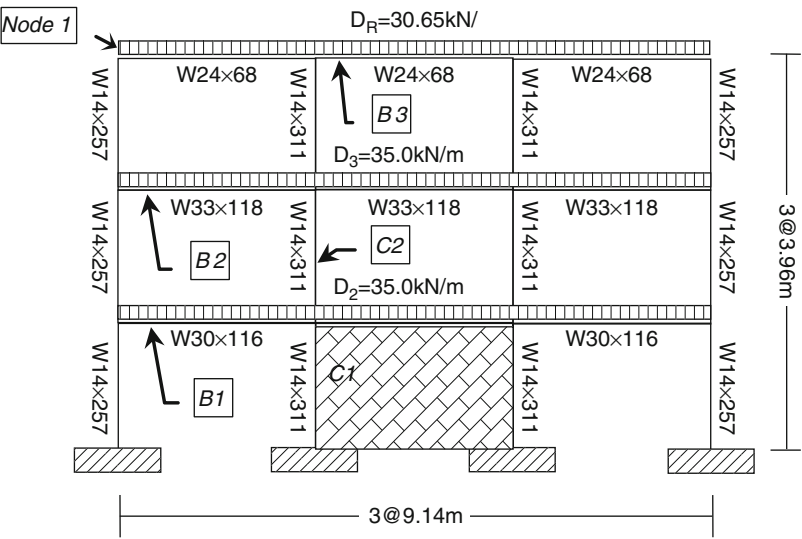
Reliability Evaluations of Frames with Different Connection Conditions with RC Shear Wall

To increase the lateral stiffness, the steel frame is strengthened with a reinforced concrete (RC) shear wall at the first floor level, as shown in Fig. 6. Obviously, the RC shear wall will increase the complexity in the problem.

Structural Reliability Estimation for Seismic Loading, Table 4 Reliability evaluations of frame without and with shear wall

3 × 3 SMRF without shear wall			Connection type		
			FR	Post-NC	Pre-NC
Serviceability limit state (Node 1)		β	1.330	1.329	0.463
		P _f ≈ Φ(−β)	0.09176	0.09192	0.32168
		No. of RV	8	20	20
		TNSP	79	313	313
Strength limit state	Beam	β	4.724	4.756	3.681
		P _f ≈ Φ(−β)	1.156 × 10 ^{−6}	9.873 × 10 ^{−7}	1.162 × 10 ^{−4}
		No. of RV	6	18	18
		TNSP	103	264	264
	Column	β	5.402	5.376	4.154
		P _f ≈ Φ(−β)	3.295 × 10 ^{−8}	3.808 × 10 ^{−8}	1.634 × 10 ^{−5}
		No. of RV	6	18	18
		TNSP	103	264	264
3 × 3 SMRF with shear wall					
Serviceability limit state (Node 1)		β	3.667	3.534	1.685
		P _f ≈ Φ(−β)	0.00012	0.00020	0.04599
		No. of RV	10	22	22
		TNSP	108	366	366
Strength limit state	Beam	β	6.879	6.714	4.467
		P _f ≈ Φ(−β)	3.014 × 10 ^{−12}	9.468 × 10 ^{−12}	3.966 × 10 ^{−6}
		No. of RV	8	20	20
		TNSP	79	313	313
	Column	β	6.879	6.714	4.467
		P _f ≈ Φ(−β)	3.014 × 10 ^{−12}	9.468 × 10 ^{−12}	3.966 × 10 ^{−6}
		No. of RV	8	20	20
		TNSP	79	313	313

Structural Reliability Estimation for Seismic Loading, Fig. 6 Steel frame with RC shear wall in the first floor (Haldar et al. 2012)



However, if it is placed properly, it will increase the lateral stiffness of the frame.

For the steel and concrete dual system, all the steel elements in the frame are modeled as beam-column elements. A four-node plane stress element is introduced for the shear wall in the frame. To consider the presence of RC shear wall, two additional parameters, namely, the modulus of elasticity, E_C , and the Poisson ratio of concrete, ν , are necessary in the deterministic formulation. The tensile strength of concrete is small compared to its compressive strength and cracking may develop at a very early stage of loading. The behavior of an RC shear wall is expected to be significantly different before and after cracking. It was observed that the degradation of the stiffness of the shear walls occurs after cracking and can be considered effectively by reducing the modulus of elasticity of the shear walls (Lefas et al. 1990). The same concept is used in this study. The shear wall is assumed to develop cracks when the tensile stress in concrete exceeds the prescribed value. The rupture strength of concrete, f_r , is assumed to be $f_r = 7.5 \times \sqrt{f'_c}$, where f'_c is the compressive strength of concrete. After the tensile stress of each shear wall exceeds the prescribed tensile stress of concrete, the degradation of the shear wall stiffness is assumed to be reduced to 40 % of the original stiffness (Lefas et al. 1990). The uncertainty in all the variables considered for the bare steel frame will remain the same. However, two additional sources of uncertainty, namely, in E_C and ν , need to be considered (Lee and Haldar 2003), as given in Table 2.

The frame is again excited by the same earthquake time history as shown in Fig. 2. The probabilities of failure for the combined dual system in the presence of FR, post-NC, and pre-NC connections are calculated using the proposed algorithm for the strength and serviceability limit states. The results are summarized in Table 4. The results indicate that the presence of shear wall at the first floor level significantly improves both the serviceability and strength behavior of the steel frame. If the probabilities

of failure need to be reduced further, RC shear walls can be added in the second and/or third floor. Again, this improved behavior can be observed and quantified by carrying out about a hundred deterministic evaluations instead of thousands of MCS. The improved behavior of the frame in the presence of RC shear wall is expected; however, the proposed algorithm can quantify the amount of improvement in terms of probability of failure considering all major sources of uncertainty. This is a significant development. The information will help to make decision as what to do next in the design.

Summary

Structural reliability estimation for seismic loading, specifically for large structural systems, is evolving. An overview of the existing state of the art is given. For wider acceptance, it is necessary to estimate reliability by applying the seismic loading in time domain, considering special features recently being introduced to improve seismic response behavior and explicitly considering all major sources of nonlinearity and uncertainty. A new method is suggested incorporating all these features. Several seismic risk evaluation procedures are illustrated with examples.

Cross-References

- [Nonlinear Dynamic Seismic Analysis](#)
- [Nonlinear Finite Element Analysis](#)
- [Probabilistic Seismic Hazard Models](#)
- [Reliability Estimation and Analysis for Dynamical Systems](#)
- [Response Variability and Reliability of Structures](#)
- [Seismic Reliability Assessment, Alternative Methods for](#)
- [Seismic Risk Assessment, Cascading Effects](#)
- [Stochastic Analysis of Nonlinear Systems](#)

- Stochastic Finite Elements
- Structural Seismic Reliability Analysis
- Time History Seismic Analysis

References

- Alis OF, Rabitz H (2001) Efficient implementation of high dimensional model representations. *J Math Chem* 29(2):127–142
- American Institute of Steel Construction (2005) Manual of steel construction: load and resistance factor design. American Institute of Steel Construction, Chicago
- Basudhar A, Missoum S, Harrison Sanchez A (2008) Limit state function identification using support vector machines for discontinuous responses and disjoint failure domains. *Probab Eng Mech* 23(1):1–11
- Box GP, William GH, Hunter JS (1978) Statistics for experimenters: an introduction to design, data analysis and modeling building. Wiley, New York
- Bucher CG, Bourgund U (1990) A fast and efficient response surface approach for structural reliability problems. *Struct Saf* 7:57–66
- Colson A (1991) Theoretical modeling of semirigid connections behavior. *J Constr Steel Res* 19:213–224
- El-Salti MK (1992) Design of frames with partially restrained connections. PhD dissertation, Department of Civil Engineering and Engineering Mechanics, University of Arizona
- Federal Emergency Management Agency (FEMA) (2000) Seismic design criteria for steel moment-frame structures, FEMA 350–353 and 355A–F
- Gavin HP, Yau SY (2008) High-order limit state function in the response surface method for structural reliability analysis. *Struct Saf* 30(2):162–179
- Haldar A, Mahadevan S (2000a) Probability, reliability, and statistical methods in engineering design. Wiley, New York
- Haldar A, Mahadevan S (2000b) Reliability assessment using stochastic finite element analysis. Wiley, New York
- Haldar A, Nee KM (1989) Elasto-plastic large deformation analysis of PR steel frames for LRFD. *Comput Struct* 31(5):811–823
- Haldar A, Farag R, Huh J (2012) A novel concept for the reliability evaluation of large systems. *Adv Struct Eng* 15(11):1879–1892
- Huh J, Haldar A (2002) Seismic reliability of nonlinear frames with PR connections using systematic RS-M. *Probab Eng Mech* 17(2):177–190
- Huh J, Haldar A (2011) A novel risk assessment method for complex structural systems. *IEEE Trans Reliab* 60(1):210–218
- Khuri AI, Cornell JA (1996) Response surfaces designs and analyses. Marcel Dekker, New York
- Kondoh K, Atluri SN (1987) Large-deformation, elasto-plastic analysis of frames under nonconservative loading, using explicitly derived tangent stiffnesses based on assumed stresses. *Comput Mech* 2(1):1–25
- Layman R, Missoum S, Vande Geest J (2007) Failure prediction and robust design of grafts for aortic aneurysms. In: 48th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and material conference, Honolulu
- Lee SY, Haldar A (2003) Reliability analysis of frame and shear wall structural systems – dynamic loading. *J Struct Eng ASCE* 129(2):233–240
- Lefas D, Kotsosovos D, Ambraseys N (1990) Behavior of reinforced concrete structural walls: strength, deformation characteristics, and failure mechanism. *ACI Struct J* 87(1):23–31
- Leger P, Dussault S (1992) Seismic-energy dissipation in MDOF structures. *J Struct Eng ASCE* 118(5):1251–1269
- Li G, Rosenthal C, Rabitz H (2001) High dimensional model representations. *J Phys Chem A* 105:7765–7777
- Mehrabian A, Haldar A, Reyes AS (2005) Seismic response analysis of steel frames with post-Northridge connection. *Steel Compos Struct* 5(4):271–287
- Rackwitz R (2000) Reliability analysis: past, present and future, Keynote Lecture. In: 8th ASCE joint specialty conference on probabilistic mechanics and structural reliability, University of Notre Dame, South Bend
- Rackwitz R, Fiessler B (1976) Note on discrete safety checking when using non-normal stochastic models for basic variables. Load Project Working Session, MIT, Cambridge
- Rajashekhar MR, Ellingwood BR (1993) A new look at the response surface approach for reliability analysis. *Struct Saf* 12(3):205–220
- Rao BN, Chowdhury R (2009) Enhanced high dimensional model representation for reliability analysis. *Int J Numer Methods Eng* 77(5):719–750
- Richard RM (1993) PRCONN manual. RMR Design Group, Tucson
- Richard RM, Abbott BJ (1975) Versatile elastic–plastic stress–strain formula. *J Eng Mech ASCE* 101(EM4):511–515
- Richard RM, Radau RE (1998) Force, stress and strain distribution in FR bolted welded connections. In: Proceedings of structural engineering worldwide, San Francisco, CA
- Richard RM, Allen CJ, Partridge JE (1997) Proprietary slotted beam connection designs. Modern Steel Construction, Chicago
- Shi G, Atluri SN (1988) Elasto-plastic large deformation analysis of space frames. *Int J Numer Methods Eng* 26:589–615
- Sobol IM (2003) Theorems and examples on high dimensional model representations. *Reliab Eng Syst Saf* 79(2):187–193
- Wei D, Rahman S (2007) Structural reliability analysis by univariate decomposition and numerical integration. *Probab Eng Mech* 22:27–38
- Yao TH-J, Wen YK (1996) Response surface method for time-variant reliability analysis. *J Struct Eng ASCE* 122(2):193–201

Structural Seismic Reliability Analysis

Shankar Sankararaman

Intelligent Systems Division, SGT Inc., NASA
Ames Research Center, Moffett Field, CA, USA

Synonyms

First-order reliability method; Monte Carlo sampling; Probability of failure; Reliability analysis; Uncertainty

Introduction

Perhaps, no other discipline within engineering has to deal with as much uncertainty as the field of earthquake engineering (Der Kiureghian 1996). To begin with, the occurrence of earthquakes in time and space is completely random in nature, and this leads to a large amount of uncertainty while predicting the intensities of ground motions resulting from earthquakes. Further, it is challenging to precisely assess the load-carrying capacity of the structural system of interest, due to the inherent variability across different structural members that constitute the overall structural system. It is necessary to analyze all of these different sources of uncertainty and assess the safety of the structure by accounting for such sources of uncertainty. Structural safety assessment is important both during the design of the structural system and for analyzing its performance while the system is under operation, particularly before and after earthquakes.

Engineering design is usually a trade-off between maximizing safety levels and minimizing cost. It is necessary to account for the uncertainty in anticipated loading conditions and the different sources of uncertainty regarding the structural system while assessing safety during system design. Traditional design approaches simplify the problem by treating the uncertain quantities to be deterministic and account for

the inherent uncertainty through the use of empirical safety factors, also referred to as deterministic safety factors. These empirical safety factors do not provide any information on how the different uncertain quantities influence the overall structural safety. Therefore, it is difficult to design a system with a uniform distribution of safety levels among the different components using empirical safety factors (Haldar and Mahadevan 2000). Further, during the design stage, deterministic safety factors do not provide adequate information to achieve optimal use of the available resources to maximize safety. For these reasons, it is necessary to use probabilistic approaches that rigorously account for the different sources of uncertainty and directly aid in the design of the structural system. Probabilistic approaches directly calculate the probability that the structural system may fail by probabilistically analyzing the load-carrying capacity of the structure and the actual loading on the structure. The safety of the structure is defined in terms of the converse of the probability of failure, which in turn is used to compute the so-called reliability metric that measures the probabilistic reliability of the structure. In fact, probabilistic methods provide a systematic framework to analyze the different sources of uncertainty, quantify their individual contributions to the overall structural safety, and aid in efficient design by choosing design parameters that maximize the reliability of the structure.

During the operation of the structural system, particularly before and after earthquakes, it is important to assess the safety of the structure and make decisions regarding repairs and replacements that are necessary for maintenance, rehabilitation, and structural retrofitting. The quality of the structural components may have degraded over the course of the operation of the structure, and therefore, the load-carrying capacity of the structures may have changed over the course of time. Therefore, it is important to reassess structural safety by recalculating the structural demands and load-carrying limits and reestimating the reliability of the structure.

Since the advantages of probabilistic assessment of structural safety have become evident

during the past two decades, several design guidelines and codes are being revised to incorporate probabilistic analysis. Examples of such revision include the American Institute of Steel Construction Load and Resistance Factor Design (1994) specifications and the European and Canadian structural design specifications. The use of probabilistic analysis in these codes is expected to provide more information about system behavior, the influence of the various uncertain quantities on system performance, and the interaction between the different system components. Therefore, the engineering design community has been increasingly resorting to the use of probabilistic reliability assessment techniques over the past few years. This has further been aided by the advent of high power computing technology that supports the use of advanced computational techniques for probabilistic analysis.

This entry discusses the fundamental concepts of structural reliability analysis in the context of earthquake engineering and provides an overview of the various mathematical tools that may be used to assess the safety and reliability of different types of structural systems. The rest of this entry is organized as follows. The fundamental problem of reliability analysis is formally defined and mathematically presented in section “[Structural Reliability Analysis](#).” The various sources of uncertainty are discussed, and the challenges involved in reliability calculation are outlined. Section “[Simulation-Based Methods](#)” discusses different types of simulation techniques for structural reliability analysis; while simulation techniques rely on high computational power, they are generally accurate and serve as benchmark solutions for comparison against other computational methods. Section “[First-Order Reliability Methods](#)” discusses a class of analytical techniques, popularly known as first-order reliability methods, to calculate structural reliability. While analytical methods are based approximations, they are computationally far cheaper than simulation-based approaches. The aforementioned methods are first illustrated using a structural beam in section “[Numerical Example I: Structural Beam](#)” and using a structural

frame in section “[Numerical Example 2: Structural Frame](#).” Section “[Advanced Concepts in Structural Reliability](#)” reviews a few advanced methods for reliability assessment, thereby providing an overview of the state of the art in the topic of structural reliability analysis.

Structural Reliability Analysis

The fundamental formulation of structural reliability analysis is based on the fact that a structural failure occurs when the loading demand on the structure is greater than its load-carrying capacity. Conventionally, the load-carrying capacity of the structure is equivalent to the resistance offered by the structure and is denoted by R . The loading applied on the structure is denoted by S . The classical formulation of reliability analysis is based on the following equation:

$$G \equiv R - S \quad (1)$$

If the load-carrying capacity is larger than the applied loading, $G = R - S > 0$, and the structure is considered to be safe. If the load-carrying capacity is smaller than the applied loading, $G = R - S < 0$, and the structure is considered to have failed. The equation $G = 0$ represents the condition when the applied loading is equal to the resistance offered, and this equation is popularly known as the limit state equation in structural reliability analysis. Further, the function G is also referred to as the performance function.

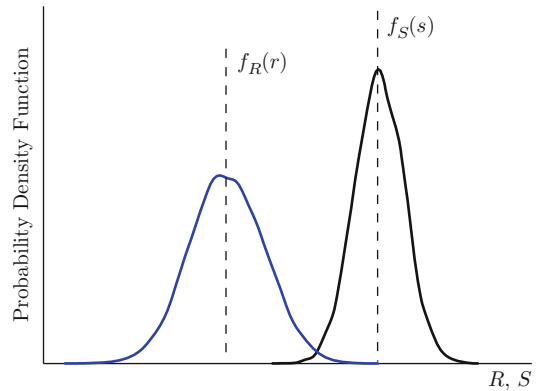
If the loading on the structure (S) were to be known precisely and if the load-carrying capacity (R) of a structure could be estimated precisely, then R and S would become completely deterministic. It would then be possible to design a structure whose load-carrying capacity would be greater than the anticipated loading, and design would become a trivial problem. However, since R and S are mostly uncertain, particularly in earthquake engineering, it becomes challenging to ascertain whether the structure is safe or not.

Sources of Uncertainty

It is important to understand the reason why R and S are uncertain in practical structural systems. The uncertainty in S , i.e., the applied/anticipated loading conditions, can be explained from the fact a given structure does not experience the same amount of loading at all possible time instants. In the context of earthquake engineering, the occurrences of earthquakes and their intensities are random, and therefore, the structural loading that is resultant of these earthquakes is also random.

The resistance offered by a structural member, i.e., R , is also random due to several reasons. The resistance offered by a structural member depends on several quantities, including its material and geometric properties. For example, the maximum bending moment that can be applied on a beam is a function of its yield stress (material property) and the sectional modulus (geometric property). There are two types of uncertainty that contribute to the overall uncertainty in R . First, the material and geometric properties that govern the value of R may themselves be uncertain due to natural variability across nominally identical members of the same type. Second, the functional dependence between these quantities and R may not be precisely known; though a physics-based model may be developed to represent this functional relationship, it may not be accurately representative of the actual carrying capacity of the structure. The latter issue is commonly known as model uncertainty and has been an important topic of research during the past few years.

Some researchers classify the different types of uncertainty into aleatory and epistemic (Der Kiureghian and Ditlevsen 2009). Aleatory uncertainty refers to those types of uncertainty that are inherent in nature and, therefore, irreducible by definition. For example, the variability in material properties, variability in earthquake loading, etc. are all irreducible in nature. Epistemic uncertainty refers to those types of uncertainty that may be reduced when more information is available. For example, when an improved model can be used to predict the response of a structure, then the uncertainty regarding the model prediction would decrease, thereby decreasing the overall uncertainty.



Structural Seismic Reliability Analysis, Fig. 1 Region of overlap between R and S

Nevertheless, it is clear as to why R and S need to be treated as uncertain variables. Therefore, they need to be represented as probability distributions. Let $f_R(r)$ and $f_S(s)$ denote the probability density functions (PDF) of R and S , respectively, as shown in Fig. 1. Note that r and s are generic realizations of the random variables R and S , respectively. While some realizations of R and S may render the structure safe (whenever $r > s$), some other realizations of R and S may correspond to structural failure. The most important aspect of structural reliability analysis is to calculate the probability of structural failure based on the probability distributions of R and S .

Calculation of Failure Probability

From the preceding discussion, it can be easily seen that the probability of the event $G < 0$ is directly equal to the probability of structural failure, and this can be mathematically written as

$$P_f = P(R < S) = P(G < 0) \quad (2)$$

Using principles of probability, Eq. 2 can be rewritten as (Ditlevsen and Madsen 1996):

$$P_f = \int_{R < S} f_R(r) f_S(s) ds dr \quad (3)$$

The reliability of the structure is the converse of failure probability and is equal to $1 - P_f$. Typically, good design practices ensure that

$\mu_R > \mu_S$, where μ_R and μ_S denote the means of R and S , respectively. However, due to the uncertainty in R and S , there is small region of overlap between the densities of R and S , and in this region of overlap, structural failure will occur because $R < S$. Therefore, the probability content in this region of overlap is equal to P_f , and in fact, this region is the domain of integration in Eq. 3. The goal in design is to minimize this region of overlap, and this may be accomplished either by moving μ_R and μ_S apart from each other or by reducing σ_R and σ_S , where σ_R and σ_S denote the standard deviations of R and S , respectively. There is a design cost related with each of these activities; while it may be cost-wise prohibitive to eliminate this region of overlap, good design practices advocate minimizing the probability of failure subject to certain cost constraints. That is why it is important to be able to accurately estimate the value of failure probability.

Analytical, closed-form expressions are available for P_f only in the special case when both R and S are normally distributed, i.e., when R and S follow Gaussian distributions as $N(\mu_R, \sigma_R)$ and $N(\mu_S, \sigma_S)$, respectively. In this case, G follows the distribution $N(\mu_R - \mu_S, \sqrt{\sigma_R^2 + \sigma_S^2})$, and P_f can be evaluated as

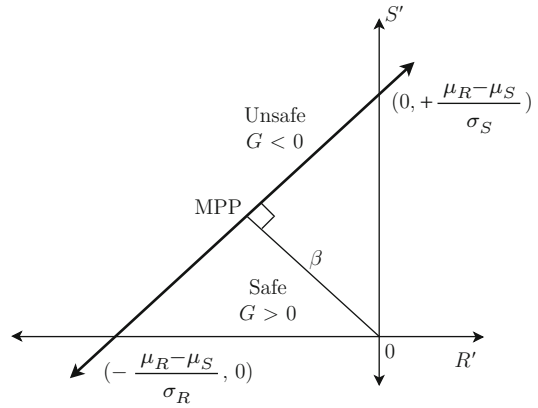
$$P_f = P(G < 0) = \Phi\left(-\frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}}\right) \quad (4)$$

In order to delve deeper into this computation, consider the transformation of the Gaussian variables R and S into standard normal distributions (with zero mean and unit standard deviation) as follows:

$$R' = \frac{R - \mu_R}{\sigma_R} \quad (5)$$

$$S' = \frac{S - \mu_S}{\sigma_S} \quad (6)$$

Now, both R' and S' follow the Gaussian distribution $N(0, 1)$. This coordinate space is referred to as the reduce coordinate system or the standard normal coordinate system. Rewrite the limit



Structural Seismic Reliability Analysis, Fig. 2 Linear performance function

state equation in Eq. 1 in terms of the transformed variables as

$$G \equiv \sigma_R R' - \sigma_S S' + \mu_R - \mu_S = 0 \quad (7)$$

Since the performance function G is linear, the limit state equation corresponds to a straight line in the $R' - S'$ coordinate axis, as shown in Fig. 2. This straight line divides the plane into two regions. The first region corresponds to $G > 0$ and is known as “safe region,” while the second region corresponds to $G < 0$ and is known as the “unsafe region.”

It is apparent from Fig. 2 that if the limit state line is closer to the origin in the reduced coordinate system, the failure region is larger, and if it is away from the origin, the failure region is smaller. Hence, it is intuitive that the distance of this straight line to the origin is of importance and this can be computed using basic trigonometry as

$$\beta = \frac{\mu_R - \mu_S}{\sqrt{\sigma_R^2 + \sigma_S^2}} \quad (8)$$

Comparing Eq. 8 and Eq. 4, it can be easily seen that

$$\beta = \Phi^{-1}(-P_f), \quad (9)$$

where Φ^{-1} denotes the inverse of the standard normal cumulative distribution function (Haldar and Mahadevan 2000).

The minimum distance point on the limit state is referred to as the checking point or the design point or the most probable point (MPP); since this coordinate axis consists of standard normal variables, the closer a point is to the origin, the higher is the probability of the maximum, and hence, the most probable point is called so. The minimum distance β is popularly known as the Hasofer-Lind Reliability Metric (Hasofer and Lind 1974). The lower the value of P_f , the higher is the value of β and vice versa. Sometimes, in design applications, a target reliability value is specified, and the component or the structure is designed to meet this target value.

Having discussed the computation of failure probability by considering R and S to be Gaussian variables, it is necessary to understand that the probability distributions of R and S may not be readily available, in many structural applications. In fact, they may not even follow Gaussian distributions. It may be necessary to estimate their probability distributions based on the probability distributions of those quantities that influence R and S . An alternate approach is to redefine the limit state directly in terms of these quantities, as explained in the following subsection.

Generalized Definition of the Limit State

Consider that the performance of a structure is defined by generic set of quantities, given by $\mathbf{X}$, where $\mathbf{X} = \{X_i, i = 1 \text{ to } n\}$. The loading on the structure (S) and load-carrying capacity (R) can be expressed in terms of these generic set of quantities, and without loss of generality, the limit state in Eq. 1 can be rewritten as

$$G(\mathbf{X}) = 0 \quad (10)$$

The structure is said to be safe when $G(\mathbf{X}) > 0$, and the structure is said to have failed when $G(\mathbf{X}) < 0$. Limit states in earthquake engineering are of different types and correspond to different types of failures. Limit state related to the serviceability of the structure is used to check whether the deflections of structural

components are within acceptable limits, whether the stresses are lower than yield stress, etc. Limit states related to the safety of the structure are also known as ultimate limit states and are used to check whether accelerations are smaller than acceptable thresholds, whether the stress on the structure is lower than the ultimate stress, etc. Depending upon the criteria of design, a single limit state or a combination of states may be chosen. The majority of this entry deals with situations consisting of single limit states. When multiple limit states need to be simultaneously evaluated, it is necessary to resort to system reliability techniques (Mahadevan et al. 2001; Mahadevan and Raghoeamachar 2000), as later explained in section “[Advanced Concepts in Structural Reliability](#)” of this entry.

Consider a limit state equation that is based on a specific functional performance of the structure. Using the corresponding limit state equation, the structural failure probability can be calculated as

$$P_f = \int_{G(\mathbf{X}) < 0} f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x}, \quad (11)$$

where $f_{\mathbf{X}}(\mathbf{x})$ denotes the joint probability density function of $\mathbf{X}$. If $\mathbf{X}$ consists of independent quantities, then the joint probability density function can be expressed as the product of the individual probability densities of X_i ($i = 1 \text{ to } m$).

Note that, in general, G may be nonlinear unlike Eq. 1, and therefore, the concepts explained in section “[Calculation of Failure Probability](#)” cannot be readily extended to this case. Further, underlying random variables, i.e., $\mathbf{X}$, may not follow Gaussian distributions at all. In order to overcome these challenges, researchers have developed a suite of computational methods to efficiently evaluate the integral in Eq. 11 and compute the failure probability. While some of these methods are based on simulation, other methods are based on linearizing G and, hence, popularly known as first-order reliability methods. These computational methods are discussed in sections “[Simulation-Based Methods](#)” and “[First-Order Reliability Methods](#),” respectively.

Simulation-Based Methods

The most intuitive approach to structural reliability analysis is to make use of simulation, i.e., repeatedly evaluate the performance function G for multiple realizations of the random variable $\mathbf{X}$, and use the results of these multiple evaluations to calculate the failure probability. Simulation techniques rely on the ability to generate pseudorandom numbers using computer programs and may require several thousand evaluations of G in order to produce meaningful results.

In any simulation-based approach, there are two steps. The first step is to randomly generate samples of $\mathbf{X}$, based on the joint probability density function $f_{\mathbf{X}}(\mathbf{x})$. The second step is to evaluate the function G for each of the randomly generated samples. While the former is a statistical problem, the latter is a modeling problem. Mathematical models (which may either be based on physics or which may be constructed using observed data) can be constructed to represent G , but it is necessary to resort to advanced statistical methods to generate random samples.

Generation of random samples is particularly easy when the various quantities in $\mathbf{X}$ are independent of each other. Then, random samples can be generated for each individual X_i . Several computer programs are capable of generating pseudorandom numbers that are uniformly distributed on the interval $[0, 1]$. Once a pseudorandom number (u_i) can be generated using a computer code, then the corresponding random sample of X_i may be generated by inverting the cumulative distribution function of X_i as

$$x_i = F_{X_i}^{-1}(u_i) \quad (12)$$

This procedure is repeated multiple times to obtain random samples of X_i . Then, the entire procedure is repeated for all the quantities in $\mathbf{X}$ to obtain multiple samples of $\mathbf{X}$.

When the quantities in $\mathbf{X}$ are dependent on one another, one approach is to use the full conditional distributions of the quantities in $\mathbf{X}$, i.e., probability densities such as $f_{X_i|X_j}(x_i|x_j)$, where $i \neq j$. Then, the conditional cumulative

distribution functions are used in Eq. 12 to successively generate one sample for each X_i . However, such information on conditional distributions may be difficult to obtain, in practice. Therefore, it is necessary to rely on correlation structure of $\mathbf{X}$ to generate correlated samples. The most popular approach to generate correlated samples using the covariance matrix of $\mathbf{X}$ is, perhaps, using the Nataf model (Liu and Der Kiureghian 1986). Alternatively, Cholesky decomposition (Haldar and Mahadevan 2000) or principal component analysis (Cox and Hinkley 1974) may also be used for such transformation. All of these methods are based on transforming the random variable $\mathbf{X}$ from the original, correlated coordinate space to an uncorrelated coordinate space using linear transformation techniques. An appropriate transformation matrix is computed, and using this matrix, it is possible to uniquely transform the variables from the correlated space to the uncorrelated space and vice versa. While generating correlated normal variables, it must be noted that it is challenging to accurately satisfy both the marginal distributions of the individual variables and the covariance matrix. In fact, it is not possible to uniquely define the joint probability density function using information on marginal distributions and the covariance matrix (Bickel and Doksum 1977). It must also be cautioned that correlation is able to capture only one type of dependence among variables, i.e., linear dependence. Other types of dependence cannot be captured using the aforementioned transformation techniques.

Having developed a mechanism to generate samples of $\mathbf{X}$, there are several types of simulation techniques that may be used to evaluate the failure probability. Some of the important techniques are briefly explained in this section.

Monte Carlo Sampling

Monte Carlo sampling (MCS) is the most straightforward approach to evaluate failure probability P_f . First, it is necessary to generate N random samples of $\mathbf{X}$, and then, the function G needs to be evaluated for every random sample $\mathbf{X}$. Recall that structural failure occurs when $G < 0$. Out of N evaluations of G , let N_f denote

the number of simulations in which $G < 0$. Then, an estimate of the probability of failure can be calculated as

$$P_f = \frac{N_f}{N} \quad (13)$$

Theoretically, an infinite number of samples are necessary to accurately estimate the probability of failure. Since only a finite number of samples are being used here, there is an uncertainty regarding the estimated value of P_f . This uncertainty can be expressed in terms of the variance of P_f as

$$\text{Variance of } P_f = \sqrt{\frac{P_f(1 - P_f)}{N}} \quad (14)$$

Thus, it can be seen that, for a given number of samples, the accuracy in the estimation of P_f increases proportional to the square root of the number of samples; for example, if the number of samples is quadrupled, then the accuracy is doubled.

The choice of the value of N is very critical in Monte Carlo analysis. As it can be seen from Eq. 14, the accuracy depends on the number of samples used. Further, in many engineering applications, the true probability of failure could be smaller than 10^{-5} . Therefore, on average, only 1 out of 100,000 samples would correspond to failure. Thus, at least 100,000 samples are necessary to observe failure. For a reliable estimate of failure probability, at least 10 times, this minimum (one million samples, in this case) is usually recommended. Sometimes, if G consists of a complex computer code (e.g., a finite element analysis), it may not be computationally feasible to evaluate G a million times. In such cases, it is necessary to resort to other alternative techniques for structural reliability estimation.

Importance Sampling

The basic idea of importance sampling is to deliberately select those samples of $\mathbf{X}$ that result in structural failure, instead of sampling them as per their original density $f_X(\mathbf{x})$. For this purpose, a new sampling density function $h_X(\mathbf{x})$ is defined so

that most samples derived out of this density function will lead to failure. Samples are drawn from $h_X(\mathbf{x})$ and corrected based on their original densities, while computing the failure probability. Therefore, $h_X(\mathbf{x})$ is popularly known as the proposal density function.

In order to mathematically derive the expression for failure probability, consider the definition of P_f as per Eq. 11. Note that the domain of integration is specifically over the failure region, defined by $G < 0$. Consider a mathematical indicator function as follows:

$$I_g(\mathbf{x}) = \begin{cases} 0, & \text{if } G(\mathbf{x}) > 0 \\ 1, & \text{if } G(\mathbf{x}) < 0 \end{cases} \quad (15)$$

Using this indicator function, the expression for failure probability can be rewritten as

$$P_f = \int I_g(\mathbf{x}) f_X(\mathbf{x}) d\mathbf{x} \quad (16)$$

Note that the domain of integration has now been explicitly rolled into the indicator function and need not be specified as integral limits. Now, multiplying and dividing by $h_X(\mathbf{x})$,

$$P_f = \int \left[I_g(\mathbf{x}) \frac{f_X(\mathbf{x})}{h_X(\mathbf{x})} \right] h_X(\mathbf{x}) d\mathbf{x} \quad (17)$$

If samples of $\mathbf{X}$ are drawn from $h_X(\mathbf{x})$, then the right-hand side of the above equation is simply the expectation of $I_g(\mathbf{x}) \frac{f_X(\mathbf{x})}{h_X(\mathbf{x})}$. Such an expectation over N samples can be computed numerically as

$$P_f = \frac{1}{N} \sum_{i=1}^N I_g(\mathbf{x}^i) \frac{f_X(\mathbf{x}^i)}{h_X(\mathbf{x}^i)} \quad (18)$$

The accuracy of the estimate of P_f obtained through importance sampling depends on the choice of $h_X(\mathbf{x})$. Several researchers have explored importance sampling in detail and studied its accuracy (Melchers 1989). In fact, the concept of importance sampling is also used in several other engineering applications such as particle filtering (Arulampalam et al. 2002), while tracking the behavior of dynamic systems.

Stratified Sampling

In this sampling approach, the overall domain of X is divided into multiple sub-domains, and samples are drawn from each sub-domain independently. The process of dividing the overall domain into multiple sub-domains is referred to as stratification. This method is applicable when subpopulations within the overall population are significantly different and when certain subpopulations contribute more to the failure probability than others.

Let the overall failure domain be divided into m mutually exclusive regions, denoted by Ω_i , where $i = 1$ to m . Each of these regions has its own probability, denoted by $P(\Omega_i)$. The unique feature of this method is that different number of samples may be chosen for each region; let N_i denote the number of samples in each region. Then, the failure probability can be expressed as

$$P_f = \sum_{i=1}^m \left[P(\Omega_i) \frac{1}{N} \sum_{j=1}^{N_i} I_g(\mathbf{x}^j) \right], \quad (19)$$

where $\mathbf{x}^j$ is the j th sample. An advantage of stratified sampling is that no particular region of the overall domain would be neglected during sampling.

Other Sampling Techniques

While it is almost impossible to explain all sampling approaches in detail, a few important techniques are briefly explained in this section, without delving into the details of implementation. Adequate references are provided for all of these techniques to instigate further reading on these topics.

While importance sampling requires far fewer number of samples than Monte Carlo sampling to estimate the failure probability, it is challenging to select an appropriate proposal density function because the region of “importance” (where more samples are likely to indicate structural failure) is usually unknown. **Adaptive sampling** is an advanced sampling technique where the efficiency of importance sampling is continuously improved by updating the proposal density function based on the information obtained after

evaluating G for a few samples. Two classes of adaptive sampling methods are multimodal sampling (Karamchandani et al. 1989) and curvature-based sampling (Wu 1992). It has been reported by researchers that adaptive sampling techniques can accurately estimate the failure probability using 100–400 samples, while traditional Monte Carlo techniques may require several hundreds of thousands of samples.

Recall that Eq. 14 provided an estimate for the variance P_f . Using the so-called variance reduction techniques, it is possible to reduce this variance and thereby obtain an improved estimate of P_f . Such techniques are called variance reduction techniques (Kalos and Whitlock 2008) and are commonly used while estimating P_f . One such technique is popularly called the **conditional expectation method**; in this method, a control variable is selected and the variance of P_f is reduced by removing the random fluctuations of this control variable which was not conditioned. In another technique, popularly known as the technique of **antithetic variates**, negative correlation is purposefully induced between successive samples to decrease the variance of the estimated mean value. It is also common to use the technique of antithetic variates in combination with the conditional expectation method (Haldar and Mahadevan 2000).

There are other sampling methods such as the Latin hypercube sampling (Iman 2008) method or the unscented transform sampling methods (Daigle and Sankararaman 2013). These sampling techniques are more well suited to compute the overall statistics of G , rather than to estimate $P_f = P(G < 0)$. While the Latin hypercube sampling approach focuses on covering the entire domain of X accurately, the unscented transform sampling method focuses on estimating the central moments of G .

Summary

This section discussed a variety of simulation techniques to estimate structural reliability. Topics such as Monte Carlo sampling, importance sampling, and stratified sampling were discussed in detail, and some other advanced topics such as adaptive sampling, variance

reduction techniques in sampling, etc. were briefly explained. The topic of structural reliability estimation using sampling has always been an important research topic, and researchers are constantly developing new methods to mitigate the computational challenges and estimate the failure probability as accurately as possible, using as fewer evaluations of G as possible.

First-Order Reliability Methods

This section discusses a class of methods known as the first-order reliability methods to compute the probability of failure of structural systems. These methods are based on the first-order Taylor's series expansion of the performance function $G(\mathbf{X})$. The first-method, known as the first-order second-moment (FOSM) method, focuses on approximating the mean and standard deviation of G and uses this information to compute P_f . Then, the FOSM method is extended to the advanced FOSM method in two steps: first, the methodology is developed for the case where all the variables in $\mathbf{X}$ are Gaussian (normal) and, second, the methodology is extended to the general case of non-normal variables.

First-Order Second-Moment Method

First, consider the generic performance function $G(\mathbf{X})$, and let $f_{\mathbf{X}}(\mathbf{x})$ denote the joint probability density function of $\mathbf{X}$. Recall $\mathbf{X} = \{X_i, i = 1 \text{ to } n\}$, and let μ_{X_i} and σ_{X_i} denote the mean and standard deviation of X_i , respectively. Further, the covariance of X_i and X_j is denoted by $\text{Cov}(X_i, X_j)$. The first-order second-moment (shortly, referred to as FOSM) method approximates G to be a Gaussian distribution, using only the mean ($\mu_{\mathbf{X}}$) and covariance of $\mathbf{X}$.

Consider the first-order Taylor series expansion of $G(\mathbf{X})$ around $\mu_{\mathbf{X}}$ as

$$G = G(\mu_{\mathbf{X}}) + \sum_{i=1}^{i=n} (X_i - \mu_{X_i}) \left(\frac{\partial G}{\partial X_i} \right)_{\mu_{\mathbf{X}}} \quad (20)$$

Note that G is now a linear function of $\mathbf{X}$ with the partial derivatives as coefficients, and

therefore, it is straightforward to approximate its mean and variance as

$$\mu_G = G(\mu_{\mathbf{X}}) \quad (21)$$

$$\sigma_G^2 = \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} \left(\frac{\partial G}{\partial X_i} \right)_{\mu_{\mathbf{X}}} \left(\frac{\partial G}{\partial X_j} \right)_{\mu_{\mathbf{X}}} \text{Cov}(X_i, X_j) \quad (22)$$

When the inputs to G are uncorrelated, then the expression for variance in Eq. 22 simplifies to

$$\sigma_G^2 = \sum_{i=1}^{i=n} \left(\frac{\partial G}{\partial X_i} \right)_{\mu_{\mathbf{X}}}^2 \sigma_{X_i}^2 \quad (23)$$

Once the mean and standard deviation of G is calculated, then G is approximated as a Gaussian distribution, with mean and standard deviation σ_G . Then, the failure probability can be calculated as

$$P_f = P(G < 0) = \Phi \left(-\frac{\mu_G}{\sigma_G} \right) \quad (24)$$

It is obvious that the failure probability calculated using the FOSM approach is accurate only in two cases: (1) when G is a linear function of $\mathbf{X}$ and the quantities in $\mathbf{X}$ are statistically independent normal random variables or (2) when G is a multiplicative function of $\mathbf{X}$ and the quantities in $\mathbf{X}$ are statistically independent lognormal random variables. In many practical examples, it is unlikely that all the variables are statistically independent normals or lognormals. Nor it is likely that G is an additive or multiplicative function of $\mathbf{X}$. In such cases, Eq. 24 is only approximate; however, it can be used to provide a rough idea of the level of risk or reliability.

Nevertheless, the FOSM approach has two important deficiencies. First, the method does not use the information regarding the variable $\mathbf{X}$ when available. The function G in Eq. 20 is linearized at the mean of $\mathbf{X}$. When G is nonlinear, significant error may be introduced by neglecting higher order terms. Second, and more importantly, the P_f calculated using Eq. 24 does not remain invariant for

different but mechanically equivalent formulations of the same performance function. For example, the following performance functions G_1 and G_2 are mechanically equivalent:

$$G_1 \equiv X_1 - X_2 \quad (25)$$

$$G_2 \equiv \frac{X_1}{X_2} < 1 \quad (26)$$

However, the FOSM approach, when implemented using these two performance functions, yields different values for $\frac{\mu_G}{\sigma_G}$ and hence different values of P_f . In other words, the same engineering problem can be formulated either in terms of strength or stress, but the FOSM approach should not lead to different results for different formulations. The lack of invariance is a serious problem and was recognized by the researchers in 1970s, and in order to overcome this problem, Hasofer and Lind (Hasofer and Lind 1974) proposed the advanced FOSM method.

Advanced FOSM for Normal Variables

First, consider the case where the quantities in $\mathbf{X}$ are independent normal variables. While it may be meaningful to use Taylor's series expansion to linearize G , there is no rationale behind choosing the point of linearization to be the mean of $\mathbf{X}$. It is necessary to choose a point of linearization such that the estimate P_f obtained after such linearization is directly related to the probability of failure of the structure.

Point of Linearization: Most Probable Point

Any point lying on the curve of demarcation would satisfy the equation $G(\mathbf{X}) = 0$. Since (1) this curve serves as the demarcation between the two zones given by $G > 0$ and $G < 0$ and (2) it is of interest to calculate the probability $P(G < 0)$, it is intuitive that it is important to identify a linear function which closely resembles the contour $G(\mathbf{x}) = 0$. Hence, the point of linearization must lie on this curve of demarcation; in other words, the point of linearization must satisfy the equation $G(\mathbf{x}) = 0$. This is

clearly different from the FOSM approach, where the mean μ_X was chosen as the point of linearization; in general, the mean μ_X will not satisfy this equation.

Therefore, the point of linearization should be located on the curve of demarcation. However, there are infinite points that satisfy this criterion, and it is important to select the appropriate one. Each of these infinite points has a likelihood of occurrence, and intuitively, the *point of maximum likelihood* is chosen as the point of linearization. This likelihood can be calculated using the probability density function of the underlying random variables. For a single normal random variable X with mean μ and standard deviation σ , the PDF is given by

$$f_X(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right] \quad (27)$$

For example, when $\mu = 10$ and $\sigma = 1$, $x = 10$ is 1.65 times more likely to occur than $x = 9$. The maximum value of the likelihood function occurs at $x = \mu$; therefore, the farther x is away from the mean μ , the lower the likelihood of occurrence of x . If there is only one random variable X instead of $\mathbf{X}$, the mean μ_X cannot be chosen as the point of linearization since $G(\mu_X) \neq 0$, in general. Therefore, if there is a single input variable, the point of linearization is chosen in such a way that it satisfies the equation $G(x) = 0$ and the value of $\|x - \mu_X\|$ is minimum.

However, in a general structural reliability analysis problem, the input to G is a vector, i.e., $\mathbf{X} = \{X_1, X_2, \dots, X_i, \dots, X_n\}$, and each X_i has its own mean μ_{X_i} and standard deviation σ_{X_i} . The objective is to identify the point of maximum likelihood, which can be calculated by maximizing the joint probability density function of all the input random variables. If the variables are independent, then the joint density function of $\mathbf{X}$ is expressed as (when the variables are independent)

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n \frac{1}{\sigma_i\sqrt{2\pi}} \exp \left[-\frac{(x_i - \mu_i)^2}{2\sigma_i^2} \right] \quad (28)$$

It can be verified (by taking the logarithm) that the maximizer of the above function simultaneously minimizes

$$\delta = \sum_{i=1}^{i=n} \left(\frac{x_i - \mu_i}{\sigma_i} \right)^2 \quad (29)$$

Eq. 29 can be rewritten as

$$\delta = \sum_{i=1}^{i=n} u_i^2, \quad (30)$$

where

$$u_i = \frac{x_i - \mu_i}{\sigma_i} \quad (31)$$

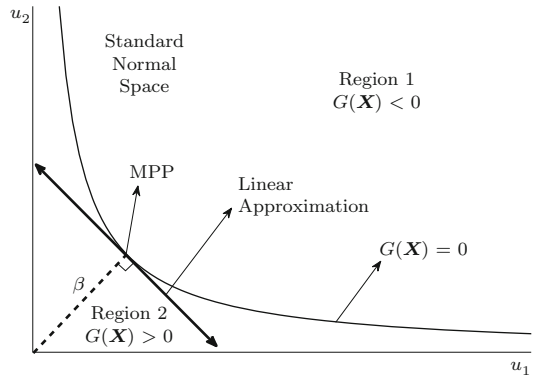
If the above computation were performed for every realization x_i of the random variable X_i , then the corresponding u_i 's would be realizations of the standard normal variable U_i , i.e., $U_i \sim N(0,1)$. Therefore, Eq. 31 is referred to as the standard normal transformation, similar to that in Eq. 5 and Eq. 6. In the space containing standard normal variables, maximizing the likelihood of occurrence is equivalent to minimizing Eq. 30, which implies that the point of linearization is that point on the curve of demarcation, whose distance (measured in the standard normal space) from the origin is minimum. Since the point of linearization has the maximum likelihood of occurrence, it is popularly known as the Maximum Probable Point (MPP), as indicated in Fig. 3.

Computing Failure Probability

The next task is to expand $G(\mathbf{X})$ around the MPP and make use of the linear approximation. Let $\mathbf{x}^* = \{x_i^*, i = 1 \text{ to } n\}$ denote the most probable point in the original coordinate space and $\mathbf{u}^* = \{u_i^*, i = 1 \text{ to } n\}$ denote the corresponding point in the standard normal space. Then,

$$G = G(\mathbf{x}^*) + \sum_{i=1}^{i=n} (X_i - x_i^*) \left(\frac{\partial G}{\partial X_i} \right)_{\mathbf{x}^*} \quad (32)$$

Now, G is a linear combination of Gaussian variable X and, therefore, can be approximated to



Structural Seismic Reliability Analysis,
Fig. 3 Estimating MPP in FORM

be Gaussian with mean μ_G^* and standard deviation σ_G^* . The mean can be calculated as

$$\mu_G^* = \sum_{i=1}^{i=n} (\mu_{X_i} - x_i^*) \left(\frac{\partial G}{\partial X_i} \right)_{\mathbf{x}^*} \quad (33)$$

By definition of the standard normal transformation, it can be easily seen that (for all $i = 1$ to n)

$$x_i^* = \mu_{X_i} + u_i^* \sigma_{X_i} \quad (34)$$

$$\frac{\partial u_i}{\partial x_i} = \frac{1}{\sigma_{X_i}} \quad (35)$$

$$\left(\frac{\partial G}{\partial X_i} \right)_{\mathbf{x}^*} = \left(\frac{\partial G}{\partial u_i} \right)_{\mathbf{u}^*} \frac{\partial u_i}{\partial x_i} = \frac{1}{\sigma_{X_i}} \left(\frac{\partial G}{\partial u_i} \right)_{\mathbf{u}^*} \quad (36)$$

Substituting Eq. 34 and 36 in Eq. 33,

$$\mu_G^* = \sum_{i=1}^{i=n} (\mu_{X_i} - x_i^*) \left(\frac{\partial G}{\partial X_i} \right)_{\mathbf{x}^*} \quad (37)$$

Substituting Eq. 34 and 36 in Eq. 37,

$$\mu_G^* = \sum_{i=1}^{i=n} -u_i^* \times \left(\frac{\partial G}{\partial u_i} \right)_{\mathbf{u}^*} \quad (38)$$

Note that $\left(\frac{\partial G}{\partial u_i} \right)_{\mathbf{u}^*}$ refers to the i th term of the gradient vector ($\boldsymbol{\alpha}^* = \{\alpha_i^* = i = 1 \text{ to } n\}$) of G in

the standard normal space. Therefore, Eq. 38 can be rewritten as:

$$\mu_G^* = -(\mathbf{u}^*)^T(\boldsymbol{\alpha}^*) \quad (39)$$

Note that the vectors represented by $\mathbf{u}^*$ and $\boldsymbol{\alpha}^*$ are both perpendicular to the limit state equation and hence parallel to each other. However, they are of opposite directions; while $\mathbf{u}^*$ is directed away from the origin, the gradient vector $\boldsymbol{\alpha}^*$ is directed towards the origin, i.e., in the direction of increasing value of G . Therefore, the above vector multiplication is basically a dot product of collinear but opposite vectors, and hence,

$$\mu_G^* = \|\mathbf{u}^*\| \times \|\boldsymbol{\alpha}^*\| \quad (40)$$

Similarly, based on Eq. 32, the variance of G , denoted by $(\sigma_G^*)^2$, is calculated as

$$(\sigma_G^*)^2 = \sum_{i=1}^n \sigma_{X_i}^2 \times \left[\left(\frac{\partial G}{\partial X_i} \right)_{\mathbf{x}^*} \right]^2 \quad (41)$$

Substituting Eq. 36 in Eq. 41,

$$\begin{aligned} (\sigma_G^*)^2 &= \sum_{i=1}^n \sigma_{X_i}^2 \times \left[\left(\frac{\partial G}{\partial X_i} \right)_{\mathbf{x}^*} \right]^2 \left[\frac{\partial u_i}{\partial x_i} \right]^2 \\ &= \sum_{i=1}^n \left[\left(\frac{\partial G}{\partial u_i} \right)_{\mathbf{x}^*} \right]^2 \end{aligned} \quad (42)$$

Therefore,

$$\sigma_G^* = \sqrt{\sum_{i=1}^n \left[\left(\frac{\partial G}{\partial u_i} \right)_{\mathbf{x}^*} \right]^2} = \|\boldsymbol{\alpha}^*\| \quad (43)$$

Having estimated μ_G^* and σ_G^* , the reliability index and the value of P_f can be calculated as

$$\beta = \frac{\mu_G^*}{\sigma_G^*} = \|\mathbf{u}^*\| \quad (44)$$

$$P_f = \Phi(-\beta) \quad (45)$$

Therefore, the reliability index is simply equal to the minimum distance measured from the

Given PDFs of $\mathbf{X}$
 Minimize $\beta = \mathbf{u}^T \mathbf{u}$
 such that $G(\mathbf{x}) = 0$
 where standard normal $\mathbf{u} = T(\mathbf{x})$
 $P(G < 0) = \Phi(-\beta)$

Structural Seismic Reliability Analysis,
Fig. 4 Optimization in advanced FOSM

origin to the curve represented by the limit state equation, in the standard normal space. Note that Eq. 39 and Eq. 44 are correct only when $\beta > 0$, i.e., $P_f < 0.5$. Otherwise, the right-hand sides of these two equations need to be negated. Since the failure probability of structures is usually less than 0.1, the above equations hold in normal circumstances. The most general expression (appropriate sign incorporated) for β can be obtained by dividing the right-hand side of Eq. 39 by that of Eq. 43.

Algorithm for Structural Reliability Analysis

Having discussed the theory behind the advanced FOSM, the method is now presented as an algorithm to aid practitioners. The key of the advanced FOSM method is to identify the MPP by solving an optimization problem, as shown in Fig. 4.

In this optimization problem, the focus is to select that point on the limit state equation that is closest to the origin, in the standard normal space. In Fig. 4, T represents the standard normal transformation function from the original space ($\mathbf{x}$) to the standard normal space ($\mathbf{u}$). This optimization is solved using the Rackwitz-Fiessler (Fiessler et al. 1979) algorithm, an iterative procedure, as follows:

1. Initialize counter $j = 0$ and start with an initial guess for the most probable point (MPP), i.e., $\mathbf{x}^j = \{x_1^j, x_2^j, \dots, x_i^j, \dots, x_n^j\}$, a column vector.
2. Transform into standard normal space and calculate $\mathbf{u}^j = \{u_1^j, u_2^j, \dots, u_i^j, \dots, u_n^j\}$ using Eq. 31, a column vector.

3. Compute the gradient vector in the standard normal space, i.e., $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$, another column vector where

$$\alpha_i = \frac{\partial G}{\partial u_i} = \frac{\partial G}{\partial x_i} \times \frac{\partial x_i}{\partial u_i} = \frac{\partial G}{\partial x_i} \times \sigma_i \quad (46)$$

4. In the iterative procedure, the next point u^{j+1} is calculated using the following equation:

$$u^{j+1} = \frac{1}{\|\alpha\|} [\alpha^T u^j - G(x^j)] \frac{\alpha}{\|\alpha\|} \quad (47)$$

5. Transform back into original space, i.e., compute x^{j+1} , and continue starting from Step 3 until the iterative procedure converges. Using tolerance limits δ_1 and δ_2 , convergence can be checked if the following two criteria are satisfied: (i) the point lies on the curve of demarcation, i.e., $|G(x^j)| \leq \delta_1$, and (2) the solution does not change between two iterations, i.e., $|x^{j+1} - x^j| \leq \delta_2$.

Note that, since this approach is gradient based, the gradient vector at MPP is an indicator of which sources of uncertainty are the strongest contributors to structural failure. The higher the magnitude of the gradient in a direction that corresponds to a particular uncertain variable, the more important is that variable in the context of structural safety.

Until now, the discussion did not account for statistical among variables. If there is any statistical dependence, then it is necessary to transform the variables into uncorrelated standard normal space. The same type of transformations discussed earlier in section “[Simulation-Based Methods](#)” may be used for this purpose.

Advanced FOSM for Non-normal Variables

Now consider the case where the inputs X_i ($i = 1$ to n) have arbitrary probability distributions given by their CDFs as $F_{X_i}(x_i)$ ($i = 1$ to n). Now that X_i is not normally distributed, Eq. 31 cannot be used for standard normal transformation. Therefore, it is necessary to calculate u_i from a given x_i meaningfully, so that u_i represents

a realization of the standard normal variable. The only difference from the implementation of the advanced FOSM algorithm is the transformation step and the gradient computation which is dependent on the choice of transformation.

One simple transformation is based on probability integral transform concept as

$$u_i = \Phi^{-1}(F_{X_i}(X_i = x_i)), \quad (48)$$

where $\Phi^{-1}(\cdot)$ refers to the inverse of the standard normal distribution function (Haldar and Mahadevan 2000). Now, the calculation of the gradient in the standard normal space is different from Eq. 46 and can be derived directly using Eq. 48. First, decompose Eq. 48 into two parts as

$$v_i = F_{X_i}(X_i = x_i) \quad (49)$$

$$u_i = \Phi^{-1}(v_i) \quad (50)$$

Then, each element of the gradient vector $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ can be calculated as

$$\alpha_i = \frac{\partial G}{\partial u_i} = \frac{\partial G}{\partial x_i} \times \frac{\partial x_i}{\partial v_i} \times \frac{\partial v_i}{\partial u_i} = \frac{\partial G}{\partial x_i} \times \frac{\phi(u_i)}{f_{X_i}(x_i)}, \quad (51)$$

where $\phi(\cdot)$ refers to the standard normal density function and $f_{X_i}(x_i)$ is the PDF of the i th input variable X_i .

In addition to the above procedure, there are also other transformation techniques. For example, a two-parameter transformation procedure estimates the mean μ_i and standard deviation σ_i of the normal distribution by equating the CDF and PDF values of the distribution of X and the normal distribution. Then, Eq. 31 can be used to calculate u_i from x_i . Note that the mean μ_{X_i} and standard deviation σ_{X_i} are dependent on the value of x_i . Similarly, Chen and Lind (Chen and Lind 1983) proposed a three-parameter transformation procedure by introducing a third parameter, a scale factor which is estimated by matching the slope of the probability density function in addition to the PDF and CDF values. Further, when

the inputs are correlated or statistically dependent, it is necessary to transform them to *uncorrelated* standard normal space. Haldar and Mahadevan (Haldar and Mahadevan 2000) describe methods for such transformation. It must be noted that any transformation must be accompanied by suitably computing the derivatives in the standard normal space, and Eq. 46 must be appropriately replaced.

Sometimes, the variable X may follow arbitrary distributions and be correlated. In that case, it is still necessary to transform these variables to uncorrelated standard normal space (Liu and Der Kiureghian 1986; Haldar and Mahadevan 2000). This is usually performed in two steps: first, the variables are converted to uncorrelated space and then transformed to standard normal space where optimization is performed to estimate MPP.

Summary

The section discussed the use of first-order reliability methods in order to estimate the reliability of structures. First, the first-order second-moment (FOSM) method was presented and then extended to the advanced FOSM method. The concept of most probable point (MPP) was introduced. It was derived that the distance from the origin to the MPP, in standard normal space, is equal to the safety index or reliability index, denoted by β . Information regarding the gradient at the MPP can be used to identify the sources of uncertainty that are significant contributors to the failure of the structure.

The inverse of the advanced FOSM algorithm is commonly employed in design. This procedure is known as inverse-FORM and is used to select a design parameter (which is input as unknown quantity to the performance function G) so that a target reliability index may be attained. Details of the inverse-FORM methodology can be found in several research articles (Der Kiureghian et al. 1994) and textbooks (Haldar and Mahadevan 2000).

Numerical Example I: Structural Beam

This section illustrates the estimation of structural reliability through an illustrative example that has

been extended from a case study discussed by Haldar and Mahadevan (Haldar and Mahadevan 2000). This numerical example consists of a steel beam that is suggested to carry an applied deterministic bending moment M that follows a Gaussian distribution whose mean is equal to 1,500 kip-in and standard deviation is equal to 75 kip-in. The yield stress (Y) of steel is considered to be a lognormal variable with mean equal to 38 ksi and standard deviation equal to 3.8 ksi. The plastic section modulus (Z) is known to be a normal random variable with mean equal to 54 in³ and standard deviation equal to 2.7 in³. The goal is to compute the structural reliability of this beam.

The resistance offered by this beam (maximum loading possible) can be expressed as the product of yield stress (Y) and plastic section modulus (Z). The structure will fail if the resistance is smaller than the applied bending moment (M). Therefore, the limit state equation is

$$G = YZ - M \quad (52)$$

The structure is said to fail when $G < 0$.

First, the problem is solved using Monte Carlo simulation. It is possible to directly generate samples of M and Z since they follow Gaussian distributions. However, in order to generate samples from the lognormally distributed Y , its distribution parameters (the mean and standard deviation of the corresponding normal distribution) need to be estimated. The location parameter λ_Y is calculated to be equal to 3.632611 and the scale parameter ζ_Y is calculated to be equal to 0.0997513. It is trivial to code Monte Carlo simulation in a programming environment such as MATLAB. For this numerical example, the MATLAB codes would be

1. $N = 50,000,000$
2. $M = \text{randn}(N,1)*75 + 1,500$
3. $Z = \text{randn}(N,1)*2.7 + 54$
4. $Y = \text{icdf}('logn', \text{rand}(N,1), 3.632611, 0.0997513)$
5. $G = Y.*Z - M$
6. $N_f = \text{length}(\text{find}(G < 0))$
7. $P_f = N_f/N$

Structural Seismic Reliability Analysis,
Table 1 Advanced FOSM method: implementation

Quantity	Iteration I	Iteration II	Iteration III
Y	38	30.6	30.81
Z	54	51.4	51.1
M	1,500	1,552.5	1,573.7
G	552	19.21	−0.08
μ_Y	37.81	37.07	37.11
μ_Z	54	54	54
μ_M	1,500	1,500	1,500
σ_Y	3.79	3.05	3.07
σ_Z	2.7	2.7	2.7
σ_M	75	75	75
u_Y	0.05	−2.13	−2.05
u_Z	0	−0.96	−1.08
u_M	0	0.7	0.98
α_Y	54	51.41	51.08
α_Z	38	30.57	30.81
α_M	−1	−1	−1
α_{u_Y}	204.69	156.78	156.97
α_{u_Z}	102.06	82.53	83.18
α_{u_M}	−75	−75	−75
New u_Y	−1.9	−2.05	−2.05
New u_Z	−0.96	−1.08	−1.09
New u_M	0.70	0.98	0.98
β	2.24	2.52	2.52

As seen from the above code, a very large number of samples have been used and the resultant P_f is calculated to be equal to 0.0059. Since the number of samples is high, this serves a benchmark solution to verify the solutions from FOSM and advanced FOSM methods.

The same numerical example can also be solved using the first-order second-moment method, by linearizing the limit state equation at the mean of the variables. The distribution information of the variables is not used. In this approach, $\mu_G = 38 \times 54 - 1,500 = 552$, and $\sigma_G^2 = (3.8 \times 54)^2 + (2.7 \times 38)^2 + (75)^2$ which yields $\sigma_G = 241.37$. Therefore, $\beta = 2.287$, and hence, $P_f = 0.011$. Evidently, this is extremely erroneous because of the inaccuracy of the FOSM method.

Finally, the numerical example is solved using the advanced FOSM method, by estimating the most probable point (MPP), as shown in Table 1. The initial guess for the MPP is $Y = 38$, $Z = 54$,

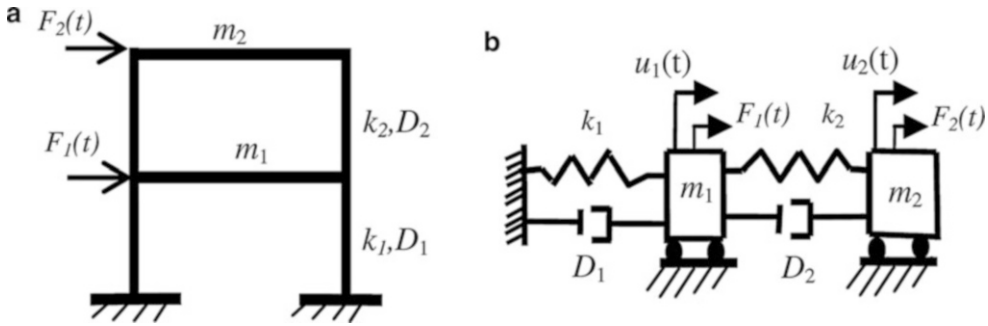
and $M = 1,500$. In this table, μ_Y , μ_Z , μ_M , and σ_Y , σ_Z , σ_M denote the equivalent normal mean and standard deviation in each iteration. Obviously, since Z and M are originally normal variables, their means and standard deviations do not change. The coordinate values in the standard normal space are denoted by u_Y , u_Z , and u_M . The derivatives in the original space are denoted by σ_Y , σ_Z , and σ_M , and these are multiplied by the equivalent normal standard deviation to calculate the derivatives in the standard normal space, denoted by σ_{u_Y} , σ_{u_Z} , and σ_{u_M} , respectively. The coordinate values after the derivative-based update is given by “New u_Y ,” “New u_Z ,” and “New u_M .” These are transformed back to the original space using the previously computed equivalent normal mean and standard deviation, and the new values of Y , Z , and M are the starting values for the next iteration. This procedure is continued until the optimization converges.

It is seen that, after three iterations, the value of G is close to zero, and there is no significant change in the value of $\beta = 2.52$ that corresponds to $P_f = 0.0059$, which is in excellent agreement with the Monte Carlo simulation approach. Note that the FOSM approach required only 12 evaluations of G to estimate this failure probability. As explained earlier, the gradient in the standard normal space is an indicator of those sources of uncertainty that are the strongest contributors to structural failure. In this case, the yield stress (Y) is found to be the most important contributor since it has the highest gradient.

Numerical Example 2: Structural Frame

This section illustrates the calculation of structural reliability using the model of a simple two-story frame, as shown in Fig. 5.

This two-story frame has six parameters, m_1 , m_2 , k_1 , k_2 , D_1 , D_2 , that represent the masses, stiffness, and damping parameters of the two stories of the structural frame. Except the masses ($m_1 = 136$ and $m_2 = 66$ kg), all of the other quantities are uncertain. Further, the inputs to the system are forces (lateral excitations) at the two levels given by $A_1 \sin(\omega_1 t)$ and $A_2 \sin(\omega_1 t)$.



Structural Seismic Reliability Analysis, Fig. 5 (a) Structural frame, (b) mechanical model

Structural Seismic Reliability Analysis, Table 2 Statistics of uncertain quantities

Parameter	Value	Std. Dev.	Unit
k_1	30,700	1,500	N/m
k_2	44,300	2,000	N/m
D_1	307	30	Ns/m
D_2	443	40	Ns/m
A_1	75	5	N
A_2	100	5	N
ω_1	9	0.5	s ⁻¹
ω_2	9	0.5	s ⁻¹

All the uncertain quantities are assumed to follow Gaussian distributions, and their statistics are provided in Table 2.

In this example, the maximum displacement of the second story during the first 5 s of loading is the quantity of interest, and the structure is said to have failed due to serviceability reasons if this maximum displacement exceeds 35 mm.

First, the response of the system (u_2 , in this case) needs to be computed based on the fundamental equations:

$$m_1 \ddot{u}_1 + (D_1 + D_2) \dot{u}_1 - D_2 \dot{u}_2 + (k_1 + k_2) u_1 - k_2 u_2 = A_1 \sin(\Omega_1 t) \quad (53)$$

$$m_2 \ddot{u}_2 - D_2 \dot{u}_1 + D_2 \dot{u}_2 - k_2 u_1 + k_2 u_2 = A_2 \sin(\Omega_2 t) \quad (54)$$

Once u_2 is computed, its maximum value is computed, and the limit state G is defined as

$$G = \max(u_2) - 0.035 \quad (55)$$

The probability of failure is directly evaluated using Monte Carlo sampling (10,000 samples), and P_f is observed to be equal to 0.0028.

Note that sinusoidal loading functions were used in this example so that the response of structural frame may be easily computed using well-known equations structural dynamics. In order to study the response of a structure to an earthquake, the appropriate earthquake loading needs to be used, and the equations of structural dynamics need to be solved using numerical techniques. Then the maximum deflection would be calculated and used to construct the limit state for reliability analysis. Since the scope of this article is to familiarize the readers with reliability methods, simpler numerical examples were considered for illustrative purposes.

Advanced Concepts in Structural Reliability

In addition to the first-order reliability methods and simulation-based techniques, there are other types of techniques and methods that have become popular for structural reliability analysis, over the past two decades. The purpose of this section is to provide an overview of some of these approaches and list appropriate references that would aid in-depth understanding of these methods.

Second-Order Reliability Methods

Recall that $G(X)$ is a generic function and the class of first-order reliability methods focus on linearizing G using the first-order Taylor's series expansion. Several researchers have developed computational methods to improve the estimates provided by the first-order reliability methods. In particular, second-order reliability methods focus on quadratic approximations of the limit state (Fiessler et al. 1979). Breitung (1984) estimated closed-form analytical expressions for P_f , using the principal curvatures of the limit state at the MPP. While Breitung's method was based on a parabolic approximation, Tvedt (Tvedt 1990) developed a generalized second-order approximation to the limit state to compute the failure probability. Der Kiureghian et al. (1987) approximated the limit state using two different semi-parabolas around the MPP and used the analytical expressions developed by Breitung (1984).

Subset Simulation

The topic of subset simulation addresses structural reliability calculation using sampling. Recall that, since failure probabilities are generally small, it is necessary to consider a large number of samples to estimate the failure probability accurately. The basic idea of subset simulation is to express the failure probability as a product of larger conditional failure probabilities by introducing intermediate failure events (Au and Beck 2001). As a result, the original problem of calculating a small failure probability, which is computationally demanding, is reduced to calculating a sequence of conditional probabilities, which can be readily and efficiently estimated by means of simulation. This approach has been applied to structural reliability analysis of frames subjected to seismic excitation (Au and Beck 2001).

Surrogate Modeling

Another class of methods for structural reliability analysis relies on approximating the performance function $G(X)$ using different types of mathematical tools. This class of methods is referred to as surrogate modeling techniques (sometimes, as

response surface methods) since they use a few evaluations of $G(X)$ (referred to as training points) to construct a mathematical function that approximates the original G . Obviously, the surrogate model will not be able to match the value of G at values of X , and therefore, this imparts additional uncertainty to the problem. Once a surrogate model is constructed, then Monte Carlo simulation may be used to compute the failure probability. Since the surrogate model is simple to evaluate, it is easy to use a million samples of X during Monte Carlo sampling.

Commonly used surrogate modeling approaches include regression techniques (Haldar and Mahadevan 2000), polynomial chaos expansion (Najm 2009), kriging (Stein 1999), etc. Each of these methods uses different types of basis functions, and one may approximate G better than the other, and it is necessary to choose a suitable surrogate model based on the application of interest.

Efficient Global Reliability Analysis

While conventional surrogate modeling approaches focus on approximating the performance function $G(X)$ over the entire domain of X , the technique of efficient global reliability analysis (Bichon et al. 2008) argues that it is not necessary for such approximation. It is sufficient to approximate the function $G(X)$ near the limit state equation, i.e., around the region where $G(X) = 0$. This method uses a few training points that lie near the curve represented by the limit state equation to construct a Gaussian process surrogate model (Rasmussen 2004) and continues updating this surrogate model using additional training points until the resultant surrogate model sufficiently approximates the performance function $G(X)$ around the curve represented by the limit state equation. Finally, Monte Carlo simulation can be used along with the final surrogate model; it is sufficient to know whether $G > 0$ or $G < 0$, and the numerical value of G is not significant. Since this method approximates G only near the limit state curve, it is well suited to find the sign of G and hence provides estimates of P_f with reasonable accuracy.

System Reliability Methods

Sometimes, a structural system may consist of multiple structural components, each of which has its own limit state. In a series system, the failure of any one component implies the failure of the entire structural system. In a parallel system, it is imperative that all the components fail individually to imply that the system has failed. The probability of system failure can be expressed as a union of component-level failures in the former case, while it is expressed as an intersection of component-level failures in the latter case. Methods for predicting system reliability have been studied by several researchers and documented in several research articles (Hohenbichler and Rackwitz 1983; Cruse et al. 1994) and textbooks (Ditlevsen and Madsen 1996; Haldrar and Mahadevan 2000). Sometimes, even a single structural component may have multiple limit states; methods for system reliability methods are applicable even to such situations since it is necessary to evaluate probability of union or intersection of different events that correspond to failure across multiple limit states.

Summary

This entry introduced the concept of structural reliability analysis, in the context of earthquake engineering, and reviewed several fundamental concepts that may be used to assess the safety of structural systems. There are several sources of uncertainty that affect the performance of structural systems, and therefore, the safety of structural systems is uncertain. It is important to constantly perform structural safety assessment for the purposes of analysis and design and estimate the probability that the structure may fail due to the applied loading. In general, structural failure occurs when the applied loading is greater than the load-carrying capacity (resistance) of the structure. Since both the load-carrying capacity and the actual loading may be uncertain, the safety of the structure also becomes uncertain. This concept was explored in detail by considering Gaussian distributions for the loading and the

resistance quantities, mathematical expressions for failure probability were derived, and the concepts of limit state and reliability index were introduced.

In many structural systems, the loading and the resistance may be functions of various quantities such as material and geometric properties, and this led to the explanation of the generalized limit state function. The quantification of failure probability was explained in detail using simulation-based approaches and first-order reliability methods. These methods were also illustrated using two different numerical examples, one consisting of a beam and the other consisting of a structural system. Finally, an overview of advanced reliability concepts such as the second-order reliability method, efficient global reliability analysis, system reliability techniques, etc. was provided, thereby explaining the state of the art in the field of structural reliability analysis.

Cross-References

- ▶ [Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach](#)
- ▶ [Reliability Estimation and Analysis](#)
- ▶ [Response Variability and Reliability of Structures](#)
- ▶ [Seismic Reliability Assessment, Alternative Methods for](#)
- ▶ [Structural Reliability Estimation for Seismic Loading](#)

References

- Arulampalam MS, Maskell S, Gordon N, Clapp T (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans Signal Process* 50(2):174–188
- Au S, Beck JL (2001) Estimation of small failure probabilities in high dimensions by subset simulation. *Probab Eng Mech* 16(4):263–277
- Bichon BJ, Eldred MS, Swiler LP, Mahadevan S, McFarland JM (2008) Efficient global reliability analysis for nonlinear implicit performance functions. *AIAA J* 46(10):2459–2468
- Bickel PJ, Doksum KA (1977) *Mathematical statistics: basic ideas and selected topics*. Holden Day, San Francisco

- Breitung K (1984) Asymptotic approximations for multinormal integrals. *J Eng Mech* 110(3):357–366
- Chen X, Lind NC (1983) Fast probability integration by three-parameter normal tail approximation. *Struct Saf* 1(4):269–276
- Cox DR, Hinkley DV (1974) *Theoretical statistics*. Chapman and Hall, London
- Cruse TA, Mahadevan S, Huang Q, Mehta S (1994) Mechanical system reliability and risk assessment. *AIAA J* 32(11):2249–2259
- Daigle M, Sankararaman S (2013) Advanced methods for determining prediction uncertainty in model-based prognostics with application to planetary rovers. In: *Annual conference of the Prognostics and Health Management Society*, New Orleans, pp 262–274
- Der Kiureghian A (1996) Structural reliability methods for seismic safety assessment: a review. *Eng Struct* 18(6):412–424
- Der Kiureghian A, Ditlevsen OD (2009) Aleatory or epistemic? Does it matter? *Struct Saf* 31(2):105–112
- Der Kiureghian A, Lin HZ, Hwang SJ (1987) Second-order reliability approximations. *J Eng Mech* 113(8):1208–1225
- Der Kiureghian A, Zhang Y, Li CC (1994) Inverse reliability problem. *J Eng Mech* 120(5):1154–1159
- Ditlevsen OD, Madsen HO (1996) *Structural reliability methods*. Wiley, Chichester
- Fiessler B, Rackwitz R, Neumann HJ (1979) Quadratic limit states in structural reliability. *J Eng Mech Div* 105(4):661–676
- Haldrup A, Mahadevan S (2000) *Probability, reliability, and statistical methods in engineering design*. Wiley, New York
- Hasofer AM, Lind NC (1974) Exact and invariant second-moment code format. *J Eng Mech Div* 100(1):111–121
- Hohenbichler M, Rackwitz R (1983) First-order concepts in system reliability. *Struct Saf* 1(3):177–188
- Iman RL (2008) *Latin hypercube sampling*. Wiley Online Library
- Kalos MH, Whitlock PA (2008) *Monte carlo methods*. Wiley, Weinheim
- Karamchandani A, Bjerager P, and Cornell, AC (1989) *Adaptive Importance Sampling*, Proceedings, International Conference on Structural Safety and Reliability (ICOSSAR), San Francisco, pp 855–862
- Liu P, Der Kiureghian A (1986) Multivariate distribution models with prescribed marginals and covariances. *Probab Eng Mech* 1(2):105–112
- Mahadevan S, Raghobhamachar P (2000) Adaptive simulation for system reliability analysis of large structures. *Comput Struct* 77(6):725–734
- Mahadevan S, Zhang R, Smith N (2001) Bayesian networks for system reliability reassessment. *Struct Saf* 23(3):231–251
- Melchers RE (1989) Importance sampling in structural systems. *Struct Saf* 6(1):3–10
- Najm HN (2009) Uncertainty quantification and polynomial chaos techniques in computational fluid dynamics. *Annu Rev Fluid Mech* 41:35–52
- Rasmussen CE (2004) Gaussian processes in machine learning. In: *Advanced lectures on machine learning*. Springer, pp 63–71
- Stein ML (1999) *Interpolation of spatial data: some theory for kriging*. Springer, New York
- Tvedt L (1990) Distribution of quadratic forms in normal space-application to structural reliability. *J Eng Mech* 116(6):1183–1197
- Wu, YT (1992) An adaptive importance sampling method for structural system reliability analysis. In: Cruse TA (ed) *Reliability technology 1992*, ASME winter annual meeting, vol AD-28, Anaheim, pp 217–231

Structures with Nonviscous Damping, Modeling, and Analysis

Sondipon Adhikari

College of Engineering, Swansea University,
Swansea, UK

Introduction

The role of damping is vitally important in predicting dynamic response of structures, such as building and bridges subjected to earthquake loads. Noise and vibration are not only uncomfortable to the users of these complex dynamical systems but also may lead to fatigue, fracture, and even failure of such systems. Increasing use of composite structural materials, active control, and damage-tolerant systems in the aerospace and automotive industries has led to renewed demand for energy absorbing and high damping materials. Effective applications of such materials in complex engineering dynamical systems require robust and efficient analytical and numerical methods. Due to the superior damping characteristics, the dynamics of viscoelastic materials and structures have received significant attention over the past two decades. This chapter is aimed at developing computationally efficient and physically insightful approximate numerical methods for linear dynamical systems with nonviscous damping.

A key feature of nonviscously damped systems is the incorporation of the time history of the state variables in the equation of motion. Here we use the Biot model (Biot 1958) which

allows one to incorporate a wide range of functions in the frequency domain by means of summation of simple “pole residue forms.” Several authors have considered this model due to its simplicity and generality (see, e.g., Adhikari 2013a, b; Muravyov 1998; Muravyov and Hutton 1997; Zhang and Zheng 2007). The equation of motion of an n -degree-of-freedom linear visously damped system can be expressed by coupled differential equations as

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \int_0^t \mathcal{G}(t-\tau)\dot{\mathbf{u}}(\tau) + \mathbf{K}\mathbf{u}(t) = \mathbf{f}(t). \quad (1)$$

Here $\mathbf{u}(t) \in \mathbb{R}^n$ is the displacement vector; $\mathbf{f}(t) \in \mathbb{R}^n$ is the forcing vector; $\mathbf{M}, \mathbf{K} \in \mathbb{R}^{n \times n}$ are respectively the mass matrix and stiffness; and $\mathcal{G}(t)$ is the matrix of damping kernel functions. In general $\mathbf{M}$ is a positive definite symmetric matrix and $\mathbf{K}$ is a nonnegative definite symmetric matrix. In the special case when $\mathcal{G}(t) = \mathbf{C}\delta(t)$, where $\delta(t)$ is the Dirac delta function, it reduces to the classical viscous damping case with a damping matrix $\mathbf{C}$. Therefore, Eq. 1 can be viewed as the generalization of the conventional visously damped systems.

The natural frequencies ($\omega_j \in \mathbb{R}$) and the mode shapes ($\mathbf{x}_j \in \mathbb{R}^n$) of the corresponding undamped system can be obtained (Meirovitch 1997) by solving the matrix eigenvalue problem

$$\mathbf{K}\mathbf{x}_j = \omega_j^2 \mathbf{M}\mathbf{x}_j, \quad \forall j = 1, 2, \dots, n. \quad (2)$$

The undamped eigenvectors satisfy an orthogonality relationship over the mass and stiffness matrices, that is,

$$\mathbf{x}_k^T \mathbf{M}\mathbf{x}_j = \delta_{kj} \quad (3)$$

and

$$\mathbf{x}_k^T \mathbf{K}\mathbf{x}_j = \omega_j^2 \delta_{kj}, \quad \forall k, j = 1, 2, \dots, n \quad (4)$$

where δ_{kj} is the Kronecker delta function. We construct the modal matrix

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^n. \quad (5)$$

The modal matrix can be used to diagonalize system (1) provided the damping matrix $\mathcal{G}(t)$ is simultaneously diagonalizable with $\mathbf{M}$ and $\mathbf{K}$. This condition, known as the proportional damping, originally introduced by Lord Rayleigh (1877) in 1877 in the context of viscous damping, is still in wide use today. The mathematical condition for proportional damping can be obtained from the commutative behavior of the system matrices (Adhikari 2001; Caughey and O’Kelly 1965). This can be expressed as $\mathcal{G}(t)\mathbf{M}^{-1}\mathbf{K} = \mathbf{K}\mathbf{M}^{-1}\mathcal{G}(t)$. The concern of this chapter is when this condition is not met, the most likely case for many practical applications. In particular, due to the recent developments in actively controlled structures and the increasing use of composite and smart materials, the need to consider general nonproportionally damped linear dynamic systems is more than ever before.

For nonproportionally damped systems, the nonviscous modal damping matrix

$$\mathcal{G}'(t) = \mathbf{X}^T \mathcal{G}(t) \mathbf{X} \quad (6)$$

is not a diagonal matrix. Such problems can be solved using a spectral approach similar to the undamped or proportionally damped system by transforming Eq. 1 into a state-space form (Wagner and Adhikari 2003). The state-space approach is not only computationally more expensive; it also lacks the physical insight provided by the classical normal mode-based approach. Therefore, many authors have developed approximate methods in the original space (Adhikari 1999a, b).

The eigenvalue problem corresponding to system (1) can be expressed as

$$[s_j^2 \mathbf{M} + s_j \mathbf{G}(s_j) + \mathbf{K}] \mathbf{u}_j = 0, \quad \forall j = 1, 2, \dots, m \quad (7)$$

where $s_j \in \mathbb{C}$ are the eigenvalues and $\mathbf{u}_j \in \mathbb{C}^n$ are the eigenvectors. The matrix $\mathbf{G}(s)$ is the Laplace transfer of $\mathcal{G}(t)$.

In general m is more than $2n$, that is, $m = 2n + p$; $p \geq 0$. Thus, although the system has n degrees of freedom, the number of eigenvalues is

more than $2n$. This is a major difference between the nonviscously damped systems and the viscously damped systems where the number of eigenvalues is exactly $2n$, including any multiplicities. When the eigenvalue s_j appears in complex conjugate pairs, $\mathbf{u}_j$ also appears in complex conjugate pairs, and when s_j is real, $\mathbf{u}_j$ is also real. Corresponding to the $2N$ complex conjugate pairs of eigenvalues, the n eigenvectors together with their complex conjugates will be called *elastic modes*. These modes are related to the n modes of vibration of the structural system. Physically, the assumption of “ $2N$ complex conjugate pairs of eigenvalues” implies that all the elastic modes are oscillatory in nature, that is, they are subcritically damped. The modes corresponding to the “additional” p eigenvalues will be called *nonviscous modes*. These modes are induced by the nonviscous effect of the damping mechanism. For stable passive systems, the nonviscous modes are overcritically damped (i.e., negative real eigenvalues) and not oscillatory in nature. Nonviscous modes, or similar to these, are known by different names in the literature of different subjects, for example, “wet modes” in the context of ship dynamics (Bishop and Price 1979) and “damping modes” in the context of viscoelastic structures (McTavish and Hughes 1993). Determination of the eigenvectors is considered next.

In this work we consider complex conjugate eigensolutions only as for stable systems such eigenvalues are of great practical importance. Using the eigensolutions, the frequency response function (FRF) can be obtained (see, for example, Adhikari (2002)) as

$$\mathbf{H}(i\omega) = \sum_{j=1}^n \left[\frac{\gamma_j \mathbf{u}_j \mathbf{u}_j^T}{i\omega - s_j} + \frac{\gamma_j^* \mathbf{u}_j^* \mathbf{u}_j^{*T}}{i\omega - s_j^*} \right] + \sum_{j=2n+1}^m \frac{\gamma_j \mathbf{u}_j \mathbf{u}_j^T}{i\omega - s_j} \quad (8)$$

where $\gamma_j = \frac{1}{\mathbf{u}_j^T [2s_j \mathbf{M} + d\mathbf{G}/ds|_{s=s_j}] \mathbf{u}_j}$.

Here $(\bullet)^*$ denotes complex conjugation, $(\bullet)^T$ denotes matrix transposition, and $(\bullet)'$ denotes differentiation with respect to s . This equation shows that if the complex eigensolutions s_j and $\mathbf{u}_j$ can be

obtained efficiently, the dynamic response can be obtained exactly using Eq. 8. In this chapter an iterative approach is developed to obtain the complex eigensolutions of nonproportionally damped systems from the undamped eigensolutions.

Iterative Approach for the Elastic Modes

Considering the proportional damping assumption, recently few methods (Adhikari and Pascual 2009, 2011) have been proposed to obtain the eigenvalues of nonviscously damped systems. So far, only perturbation type of approaches (Adhikari 2002) is available to obtain the eigenvectors for the general nonproportionally damped systems. This type of approaches may be suitable for the case on small nonproportionality. For the general case, only computationally expensive state-space approach (Wagner and Adhikari 2003) is currently available. Here a novel iterative method is proposed as an alternative to the state-space approach to obtain the elastic modes of general nonviscously damped systems.

For distinct undamped eigenvalues (ω_l^2) , $\mathbf{x}_l$, $\forall l = 1, \dots, n$, form a complete set of vectors. For this reason, $\mathbf{u}_j$ can be expanded as a complex linear combination of $\mathbf{x}_l$. Thus, an expansion of the form

$$\mathbf{u}_j = \sum_{l=1}^n \alpha_l^{(j)} \mathbf{x}_l \quad (9)$$

may be considered. Without any loss of generality, we can assume that $\alpha_j^{(j)} = 1$ (normalization) which leaves us to determine $\alpha_j^{(i)}$, $\forall i \neq j$. Substituting the expansion of $\mathbf{u}_j$ into the eigenvalue Eq. 7, one obtains the approximation error for the j -th mode as

$$\varepsilon_j = \sum_{l=1}^n s_j^2 \alpha_l^{(j)} \mathbf{M} \mathbf{x}_l + s_j \sigma_l^{(j)} \mathbf{G}(s_j) \mathbf{x}_l + \alpha_l^{(j)} \mathbf{K} \mathbf{x}_l. \quad (10)$$

We use a Galerkin approach to minimize this error by viewing the expansion (9) as a projection in the basis functions $\mathbf{x}_l \in \mathbb{R}^n$, $\forall l = 1, 2, \dots, n$.

Therefore, we make the error orthogonal to the basis functions, that is,

$$\varepsilon_j \perp \mathbf{x}_l \quad \text{or} \quad \mathbf{x}_k^T \varepsilon_j = 0 \quad \forall k = 1, 2, \dots, n. \quad (11)$$

Using the orthogonality property of the undamped eigenvectors described by (3) and (4), one obtains

$$s_j^2 \alpha_k^{(j)} + s_j \sum_{l=1}^n \alpha_l^{(j)} G'_{kl}(s_j) + \omega_k^2 \alpha_k^{(j)} = 0, \quad \forall k = 1, \dots, n \quad (12)$$

where $G'_{kl}(s_j) = \mathbf{x}_k^T \mathbf{G}(s_j) \mathbf{x}_l$ are the elements of the modal damping matrix $\mathbf{G}'(s_j)$ defined in Eq. 6. The j -th equation of this set obtained by setting $k = j$ can be written as

$$\left(s_j^2 + s_j G'_{jj}(s_j) + \omega_j^2 \right) \alpha_j^{(j)} + s_j \sum_{l \neq j}^n \alpha_l^{(j)} G'_{jl}(s_j) = 0. \quad (13)$$

Recalling that $\alpha_j^{(j)} = 1$ and $\mathbf{G}'(s_j)$ is a symmetric matrix, this equation can be rewritten as

$$s_j^2 + s_j \underbrace{\left(G'_{jj}(s_j) + \sum_{l \neq j}^n \alpha_l^{(j)} G'_{lj}(s_j) \right)}_{\gamma_j} + \omega_j^2 = 0 \quad (14)$$

$$\gamma_j = G'_{jj}(s_j) + \mathbf{b}_j^T \mathbf{a}_j \quad (15)$$

$$\mathbf{b}_j = \left\{ G'_{1j}(s_j), G'_{2j}(s_j), \dots, \{j\text{-th term deleted}\}, \dots, G'_{nj}(s_j) \right\}^T \in \mathbb{R}^{(n-1)} \quad (16)$$

and

$$\mathbf{a}_j = \left\{ \alpha_1^{(j)}, \alpha_2^{(j)}, \dots, \{j\text{-th term deleted}\}, \dots, \alpha_n^{(j)} \right\}^T \in \mathbb{C}^{(n-1)}. \quad (17)$$

The vector $\mathbf{a}_j$ is unknown and can be obtained by excluding the $j = k$ case in Eq. 12. Excluding this case, one has

$$s_j^2 \alpha_k^{(j)} + s_j \left(G'_{kj}(s_j) + \alpha_k^{(j)} G'_{kk}(s_j) + \sum_{l \neq k \neq j}^n \alpha_l^{(j)} G'_{kl}(s_j) \right) + \omega_k^2 \alpha_k^{(j)} = 0 \quad (18)$$

or $\left(s_j^2 + \omega_k^2 + G'_{kk}(s_j) \right) \alpha_k^{(j)} + s_j \sum_{l \neq k \neq j}^n G'_{kl}(s_j) \alpha_l^{(j)} = -s_j G'_{kj}(s_j), \quad \forall k = 1, \dots, n; k \neq j.$

These equations can be combined into a matrix form as

$$[\mathbf{P}_j - \mathbf{Q}_j] \mathbf{a}_j = \mathbf{b}_j. \quad (19)$$

In the above equation, the vectors $\mathbf{a}_j$ and $\mathbf{b}_j$ have been defined before. The matrices $\mathbf{P}_j$ and $\mathbf{Q}_j$ are defined as

$$\mathbf{P}_j = \text{diag} \left[\frac{s_j^2 + s_j G'_{11}(s_j) + \omega_1^2}{-s_j}, \dots, \{j\text{-th term deleted}\}, \dots, \frac{s_j^2 + s_j G'_{nn}(s_j) + \omega_n^2}{-s_j} \right] \in \mathbb{C}^{(n-1) \times (n-1)}, \quad (20)$$

and

$$\mathbf{Q}_j = \begin{bmatrix} 0 & G'_{12}(s_j) & \cdots & \{j\text{-th term deleted}\} & \cdots & G'_{1n}(s_j) \\ G'_{21}(s_j) & 0 & \vdots & \vdots & \vdots & G'_{2n}(s_j) \\ \vdots & \vdots & \vdots & \{j\text{-th term deleted}\} & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ G'_{n1}(s_j) & G'_{n2}(s_j) & \cdots & \{j\text{-th term deleted}\} & \cdots & 0 \end{bmatrix} \in \mathbb{R}^{(n-1) \times (n-1)}. \quad (21)$$

From Eq. 19, $\mathbf{a}_j$ should be obtained by solving the set of linear equations. Because $\mathbf{P}_j$ is a diagonal matrix, one way to do this is by using the Neumann expansion method (Adhikari 1999a). Using the Neumann expansion, we have

$$\begin{aligned} \mathbf{a}_j &= [\mathbf{I}_{n-1} - \mathbf{P}_j^{-1} \mathbf{Q}_j]^{-1} \{\mathbf{P}_j^{-1} \mathbf{b}_j\} \\ &= [\mathbf{I}_{n-1} + \mathbf{R}_j + \mathbf{R}_j^2 + \mathbf{R}_j^3 + \cdots] \mathbf{a}_{j0} \end{aligned} \quad (22)$$

where $\mathbf{I}_{n-1}$ is a $(n-1) \times (n-1)$ identity matrix:

$$\mathbf{R}_j = \mathbf{P}_j^{-1} \mathbf{Q}_j \in \mathbb{C}^{(n-1) \times (n-1)} \quad \text{and} \quad \mathbf{a}_{j0} = \mathbf{P}_j^{-1} \mathbf{b}_j \in \mathbb{C}^{(n-1)}. \quad (23)$$

Because $\mathbf{P}_j$ is a diagonal matrix, its inversion can be carried out analytically, and subsequently, the closed-form expressions of the elements of $\mathbf{a}_j$ can be obtained. Keeping one term in the series (22), the first-order expression of the elements of $\mathbf{a}_j$ can be obtained as

$$\mathbf{a}_j \equiv \{\alpha_k^{(j)}\}_{\forall k \neq j} = -\frac{s_j G'_{kj}(s_j)}{\omega_k^2 + s_j^2 + s_j G'_{kk}(s_j)}. \quad (24)$$

Similarly, the second-order expression of the elements of $\mathbf{a}_j$ can be obtained as

$$\begin{aligned} \mathbf{a}_j \equiv \{\alpha_k^{(j)}\}_{\forall k \neq j} &= -\frac{s_j G'_{kj}(s_j)}{\omega_k^2 + s_j^2 + s_j G'_{kk}(s_j)} \\ &+ \sum_{\substack{l=1 \\ l \neq j \neq k}}^n \frac{s_j^2 G'_{kl}(s_j) G'_{lj}(s_j)}{(\omega_k^2 + s_j^2 + s_j G'_{kk}(s_j)) (\omega_l^2 + s_j^2 + s_j G'_{ll}(s_j))}. \end{aligned} \quad (25)$$

The vector $\mathbf{a}_j$ obtained using this way can be substituted back in the expression of the eigenvalues in (14), which in turn can be solved for s_j as

$$s_j = -\left(\gamma_j \pm i\sqrt{4\omega_j^2 - \gamma_j^2}\right)/2. \quad (26)$$

However, the vectors $\mathbf{a}_j$ and $\mathbf{b}_j$ are also a function of s_j . As a result γ_j in Eq. 15 becomes a

function of s_j . This forms the basics of the iterative approach as from Eq. 26, one can write

$$\begin{aligned} s_j^{(r+1)} &= -\gamma_j(s_j^{(r)})/2 \pm i\sqrt{\omega_j^2 - \gamma_j^2(s_j^{(r)})}/4; \\ r &= 0, 1, 2, \dots \end{aligned} \quad (27)$$

For every iteration step, the vectors $\mathbf{a}_j$ and $\mathbf{b}_j$ get updated based on new values of s_j using

Eq. 24 or Eq. 25 depending on the order of terms retained in the series (22). The iteration can be started with the equivalent proportional viscous damping assumption (Udwadia 2009), namely,

$$s_j^{(0)} = -G'_{jj}(i\omega_j)/2 \pm i\sqrt{\omega_j^2 - G'^2_{jj}(i\omega_j)/4}. \quad (28)$$

The iteration can be stopped when the successive values of s_j or $\mathbf{a}_j$ do not change significantly. Once the final values of $\alpha_k^{(j)}, \forall k$ are obtained, the j -th complex mode $\mathbf{u}_j$ can be obtained from the series (9).

The necessary and sufficient conditions for the convergence of the proposed method are difficult to obtain. Below we give a sufficient condition.

Proposition 1 *A sufficient condition for the convergence of the proposed iterative method is that $|\mathbf{G}'(s_j)|$ is a diagonally dominant matrix $\forall j \in [1, 2n]$.*

Proof During the iteration process, the value of s_j changes for different iteration steps. We aim to derive the condition for the convergence of series (22) for an arbitrary value of s_j . This will guarantee the convergence of the iterative method, no matter what the value of s_j . The complex matrix power series (22) converges if, and only if, for all the eigenvalues $\sigma_l^{(j)}$ of the matrix $\mathbf{R}_j$, the inequality $|\sigma_l^{(j)}| < 1$ holds. Although this condition is both necessary and sufficient, checking convergence for all $j = 1, \dots, n$ is not feasible for every iteration step. So we look for a sufficient condition which is relatively easy to check and which ensures convergence for all $j = 1, \dots, n$.

For an arbitrary r -th iteration, let us denote the matrix $\mathbf{R}_j$ defined in Eq. 23 as $\mathbf{R}_j^{(r)}$. Suppose the value of s_j for the r -th iteration step is $s_j^{(r)}$. The kl -th element of the matrix $\mathbf{R}_j^{(r)}$ can be obtained as

$$R_{jkl}^{(r)} = \frac{-s_j^{(r)} G'_{kl}(s_j) (1 - \delta_{kl})}{\omega_k^2 + s_j^{(r)^2} + s_j^{(r)} G'_{kk}(s_j)}, \quad \forall k, l \neq j. \quad (29)$$

Since a matrix norm is always greater than or equal to its maximum eigenvalue, it follows from the inequality $|\sigma_l^{(j)}| < 1$ that the convergence of the series is guaranteed if $\|\mathbf{R}_j^{(r)}\| < 1$. Writing the sum of absolute values of entries of $\mathbf{R}_j^{(r)}$ results in the following inequality as the required sufficient condition for the convergence

$$\sum_{k=1}^n \sum_{l=1}^n \left| \frac{s_j^{(r)} G'_{kl}(s_j)}{\omega_k^2 + s_j^{(r)^2} + s_j^{(r)} G'_{kk}(s_j)} \right| (1 - \delta_{lk}) < 1. \quad (30)$$

Dividing both the numerator and denominator by $s_j^{(r)}$, the above inequality can be written as

$$\sum_{k=1}^n \sum_{l=1}^n \left| \frac{G'_{kl}(s_j)}{1/s_j^{(r)} (\omega_k^2 + s_j^{(r)^2}) + G'_{kk}(s_j)} \right| < 1. \quad (31)$$

Taking the maximum for all $k \neq j$, this condition can further be represented as

$$\max_{k \neq j} \frac{\sum_{l=1}^n |G'_{kl}(s_j)|}{\left| 1/s_j^{(r)} (\omega_k^2 + s_j^{(r)^2}) + G'_{kk}(s_j) \right|} < 1. \quad (32)$$

It is clear that (32) always holds if

$$\sum_{l=1}^n |G'_{kl}(s_j)| < |G'_{kk}(s_j)|, \quad \forall k \neq j \quad (33)$$

which in turn implies that for all $j = 1, \dots, n$, the inequality $\|\mathbf{R}_j^{(r)}\| < 1$ holds if $|\mathbf{G}'(s_j)|$ is a diagonally dominant matrix. It is important to note that the diagonal dominance of $|\mathbf{G}'(s_j)|$ is only a sufficient condition and the lack of it does not necessarily prevent convergence of the proposed iterative method.

Summary of the Algorithm

We give a simple iterative algorithm to implement the idea developed in the previous section. We select a tolerance between the differences of the successive values of s_j , denoted

by ε_m . A small value, say $\varepsilon_m = 0.001$, can be selected for numerical calculations. Considering that the undamped eigensolutions (ω_j and x_j) and the modal damping matrix $\mathbf{G}'(s_j)$ are known, the complex eigensolutions (s_j and $\mathbf{u}_j$) can be obtained using the following iterative algorithm:

```

for  $j = 1, 2, \dots, n$  do
  Initialize  $\varepsilon = 100, r = 0$ 
   $s_j^{(r)} = -G'_{jj}(i\omega_j)/2 \pm i\sqrt{\omega_j^2 - G'^2_{jj}(i\omega_j)}/4$ 
  while  $\varepsilon > \varepsilon_m$  do
     $\mathbf{b}_j = \{G'_{1j}(s_j^{(r)}), G'_{2j}(s_j^{(r)}), \dots, \{j\text{-th term deleted}\}, \dots, G'_{nj}(s_j^{(r)})\}^T$ 
     $\mathbf{a}_j \equiv \{\alpha_k^{(j)}\}_{\forall k \neq j} = -\frac{s_j^{(r)} G'_{kj}(s_j^{(r)})}{\omega_k^2 + s_j^{(r)2} + s_j^{(r)} G'_{kk}(s_j^{(r)})}$ 
     $\gamma_j = G'_{jj}(s_j) + \mathbf{b}_j^T \mathbf{a}_j$ 
     $s_j^{(r+1)} = -\gamma_j(s_j^{(r)})/2 \pm i\sqrt{\omega_j^2 - \gamma_j^2(s_j^{(r)})}/4$ 
     $\varepsilon = |s_j^{(r+1)} - s_j^{(r)}|/|s_j^{(r)}|$ 
     $r = r + 1$ 
  end while
   $\mathbf{u}_j = \sum_{k=1}^n \alpha_k^{(j)} \mathbf{x}_k$ 
end for

```

The algorithm is outlined for the first-order expression of $\alpha_k^{(j)}$ given by Eq. 24. However, the extension to the second- or higher-order expressions is straightforward. One simply needs to change the expression of $\mathbf{a}_j$ in this algorithm. If the higher-order terms are used, then less number of steps in the iteration is needed. Once the complex eigensolutions s_j and $\mathbf{u}_j$ are obtained using this method for all j , the dynamic response such as the frequency response function can be obtained exactly using Eq. 8. Next we illustrate this new method using a numerical example.

Numerical Illustration: A 3-DOF System with Exponential Nonviscous Damping

We consider a three-degree-of-freedom nonviscous and nonproportionally damped system from Adhikari (2013a). The system is shown in Fig. 1. Three masses, each of mass m_u , are connected by

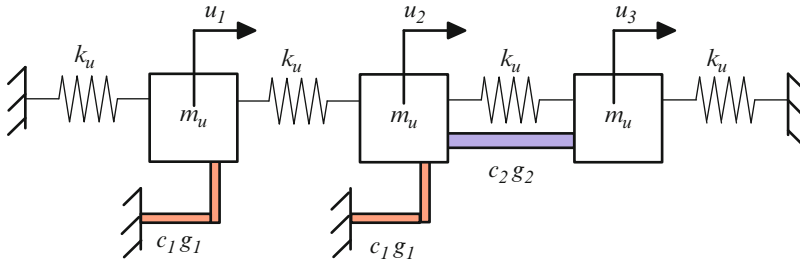
springs of stiffness k_u . The nonviscous damping elements of the system are shown in the figure. The equation of motion of this model system can be represented by Eq. 1. The mass and the stiffness matrices of the system are given by

$$\mathbf{M}_1 = \begin{bmatrix} m_u & 0 & 0 \\ 0 & 2m_u & 0 \\ 0 & 0 & m_u \end{bmatrix} \quad \text{and} \quad \mathbf{K} = \begin{bmatrix} 2k_u & -k_u & 0 \\ -k_u & 2k_u & -k_u \\ 0 & -k_u & 2k_u \end{bmatrix}. \quad (34)$$

The matrix of damping kernel functions can be expressed in the time domain as

$$\mathcal{G}(t) = C_1 \mu_1 e^{-\mu_1 t} + C_2 \mu_2 e^{-\mu_2 t}. \quad (35)$$

The coefficient matrices of this double-exponential model are given by



Structures with Nonviscous Damping, Modeling, and Analysis, Fig. 1 A three-DOF model system with nonviscous damping, the shaded bars represent the

nonviscous damping elements. $c_i, i = 1, 2$ are the damping constants and $g_i(t - \tau) = \mu_i e^{-\mu_i(t-\tau)}$, $i = 1, 2$ are the nonviscous damping functions

$$\mathbf{C}_1 = \begin{bmatrix} c_1 & 0 & 0 \\ 0 & c_1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \{\omega_1, \omega_2, \omega_3\} = \{0.4315, 1.1547, 1.2616\} \quad (36)$$

$$\mathbf{C}_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & c_2 & -c_2 \\ 0 & -c_2 & c_2 \end{bmatrix} \quad \text{and} \quad [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3] = \begin{bmatrix} 0.1750 & -0.4082 & -0.3688 \\ 0.3012 & -0.0000 & 0.1429 \\ 0.1750 & 0.4082 & -0.3688 \end{bmatrix}. \quad (42)$$

Both the matrices have rank deficiency because one can easily verify that

$$r_1 = \text{rank}(\mathbf{C}_1) = 2 \leq 3 \quad (37)$$

and

$$r_2 = \text{rank}(\mathbf{C}_2) = 1 \leq 3. \quad (38)$$

The order of the system, that is, the number of eigenvalues

$$m = 2N + \sum \text{rank}(\mathbf{C}_i) = 6 + 3 = 9 \quad (39)$$

For the numerical calculation, we have assumed $m_u = 3.0$ kg, $k_u = 2.0$ N/m, $c_1 = 0.6$ Ns/m, $c_2 = 0.2$ Ns/m, $\mu_1 = 1.0$ s⁻¹, and $\mu_2 = 5.0$ s⁻¹. The matrix $\mathbf{G}(s)$, necessary for the implementation of the iterative method, can be obtained from (35) as

$$\mathbf{G}(s) = \mathbf{C}_1 \frac{\mu_1}{s + \mu_1} + \mathbf{C}_2 \frac{\mu_2}{s + \mu_2}. \quad (40)$$

The undamped eigenvalues and eigenvectors are obtained as

Note that the last two undamped eigenvalues are very close and therefore one would expect significant modal coupling. The complex conjugate eigenvalues obtained using the proposed iterative method is compared with the exact state-space method and the first-order perturbation method in Table 1. We have used the first-order expression of $\alpha_k^{(j)}$ given by Eq. 24 and considered the error tolerance to be $\varepsilon_m = 0.001$. For all the three eigenvalues, 3 iterations are used. The first-order perturbation results are obtained from Eq. 28. The percentage errors are calculated with respect to the exact state-space results as

$$\varepsilon = 100 \times \frac{|\text{exact} - \text{approximate}|}{|\text{exact}|}. \quad (43)$$

Using the proposed iterative method, errors corresponding to all the three modes are reduced compared to the first-order perturbation results.

Next we consider the complex eigenvectors of the system. The exact eigenvectors obtained using the state-space approach (Wagner and Adhikari 2003) are given by

Structures with Nonviscous Damping, Modeling, and Analysis, Table 1 The complex eigenvalues of the system obtained using the proposed method are compared with the exact state-space method and the first-order perturbation method. The numbers in the parenthesis represent the percentage error

Eigenvalue number	State space (exact)	First-order perturbation	Proposed iterative (three iterations)	Method
1	$-0.0335 \pm 0.4453i$	$-0.0323 \pm 0.4448i$ (0.2918)	$-0.0335 \pm 0.4453i$	(0.0003)
2	$-0.0386 \pm 1.1806i$	$-0.0373 \pm 1.1831i$ (0.2445)	$-0.0383 \pm 1.1810i$	(0.0450)
3	$-0.0420 \pm 1.2941i$	$-0.0427 \pm 1.2907i$ (0.2695)	$-0.0424 \pm 1.2911i$	(0.2321)

$$\mathbf{U}_{\text{Exact}} = \begin{bmatrix} 0.1721 - 0.0073i - 0.3519 - 0.0115i - 0.4397 + 0.0163i \\ 0.3014 - 0.0002i - 0.0305 - 0.0066i \quad 0.1410 + 0.0006i \\ 0.1769 + 0.0078i \quad 0.4632 - 0.0096i - 0.3019 - 0.0124i \end{bmatrix}. \quad (44)$$

The eigenvectors calculated using the expression of the first-order perturbation in (24) are given by

$$\mathbf{U}_{\text{1st}} = \begin{bmatrix} 0.1721 - 0.0073i - 0.3519 - 0.0115i - 0.4397 + 0.0163i \\ 0.3014 - 0.0001i - 0.0309 - 0.0058i \quad 0.1420 + 0.0020i \\ 0.1770 + 0.0077i \quad 0.4646 - 0.0115i - 0.2989 - 0.0139i \end{bmatrix}. \quad (45)$$

Using the propose iterative approach, the matrix of complex eigenvectors can be obtained as

$$\mathbf{U}_{\text{iter}} = \begin{bmatrix} 0.1721 - 0.0073i - 0.3519 - 0.0115i - 0.4397 + 0.0163i \\ 0.3013 - 0.0000i - 0.0305 - 0.0062i \quad 0.1421 + 0.0019i \\ 0.1769 + 0.0078i \quad 0.4623 - 0.0094i - 0.2994 - 0.0135i \end{bmatrix}. \quad (46)$$

The eigenvectors are normalized such that the first element is identical for all the three computational method. Therefore, the eigenvectors only differ in the second and the third elements. The absolute value of the errors for the first-order perturbation method and the iterative method is given by

$$\begin{aligned} \varepsilon_{\text{1st}} &= \begin{bmatrix} 0.0593 & 3.0320 & 1.2494 \\ 0.0340 & 0.5051 & 1.1030 \end{bmatrix} \quad \text{and} \\ \varepsilon_{\text{iter}} &= \begin{bmatrix} 0.0756 & 1.2961 & 1.2448 \\ 0.0085 & 0.1864 & 0.9141 \end{bmatrix}. \end{aligned} \quad (47)$$

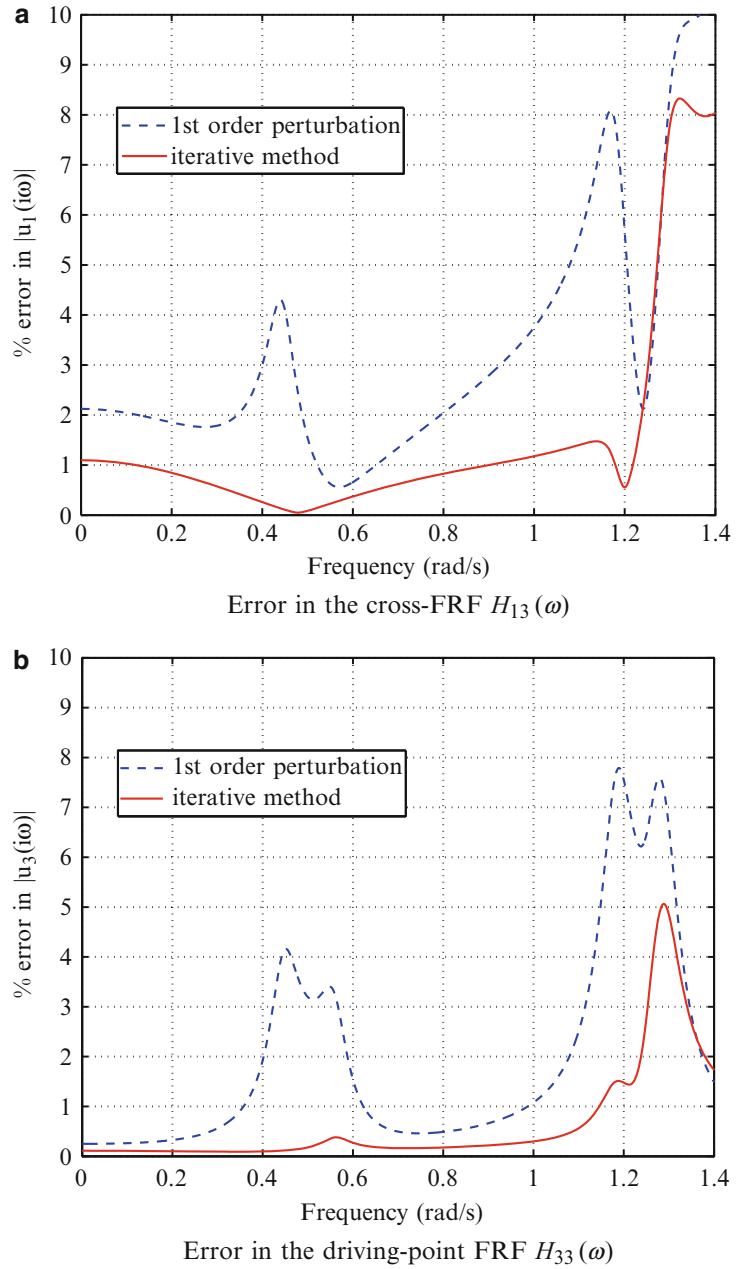
In general, the iterative method gives lower error in the eigenvectors. In Fig. 2, errors in two

typical frequency response functions of the system calculated from Eq. 8 using the first-order perturbation and the iterative method are shown. The first-order perturbation method performs poorly across the frequency range. From these results, the relative accuracy of the proposed iterative method can be observed.

Conclusions

Due to the recent focus in sustainable earthquake-resistant design, there is a renewed interest to consider general nonviscous and nonproportionally damped linear dynamic systems as new

Structures with Nonviscous Damping, Modeling, and Analysis, Fig. 2 Percentage errors with respect to the exact state-space eigensolutions in a cross-FRF and the driving-point FRF of the system. Results from the first-order perturbation method and the proposed iterative method are shown



generation of materials are being used. State-space-based methods were normally applied to address such problems. These methods are computationally more expensive and often do not give the physical insight compared to the classical normal mode-based method. In this work a new iterative method has been proposed to obtain

the complex eigensolutions of a general nonviscous nonproportionally damped system from the undamped eigensolutions. It is assumed that all the eigenvalues are distinct and are real or appear in complex conjugate pairs. The proposed method exploits a mathematical construction where complex eigenvalues and eigenvectors

can be updated from their previous values in an iterative manner. A sufficient condition for the convergence of the proposed iterative method is derived. A simple algorithm is proposed to implement this method.

The applicability of the proposed method is investigated using an example with two nonviscous damping kernels. Acceptable accuracy has been observed. Using the iterative method developed here, it is possible to obtain the eigenvalues, eigenvectors, and consequently the dynamic response of nonproportionally damped systems by post-processing of the undamped eigenvalues and eigenvectors, which in turn can be obtained using a general-purpose finite element software. Future work is necessary to extend this method to systems with repeated eigenvalues.

Acknowledgments The author gratefully acknowledges the financial support of the Royal Society of London through the Wolfson Research Merit Award.

References

- Adhikari S (1999a) Modal analysis of linear asymmetric non-conservative systems. *ASCE J Eng Mech* 125(12):1372–1379
- Adhikari S (1999b) Rates of change of eigenvalues and eigenvectors in damped dynamic systems. *AIAA J* 37(11):1452–1458
- Adhikari S (2001) Classical normal modes in non-viscously damped linear systems. *AIAA J* 39(5):978–980
- Adhikari S (2002) Dynamics of non-viscously damped linear systems. *ASCE J Eng Mech* 128(3):328–339
- Adhikari S (2013) Structural dynamic analysis with generalized damping models: analysis. Wiley ISTE, UK, 368 pp. <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1848215215.html>
- Adhikari S (2013) Structural dynamic analysis with generalized damping models: identification. Wiley ISTE, UK, 272 pp. <http://eu.wiley.com/WileyCDA/WileyTitle/productCd-184821670X.html>
- Adhikari S, Pascual B (2009) Eigenvalues of linear viscoelastic systems. *J Sound Vib* 325(4–5):1000–1011
- Adhikari S, Pascual B (2011) Iterative methods for eigenvalues of viscoelastic systems. *Trans ASME J Vib Acoust* 133(2):021002-1–7
- Biot MA (1958) Linear thermodynamics and the mechanics of solids. In: *Proceedings of the third U. S. National Congress on applied mechanics*. ASME, New York, pp 1–18
- Bishop RED, Price WG (1979) An investigation into the linear theory of ship response to waves. *J Sound Vib* 62(3):353–363
- Caughey TK, O’Kelly MEJ (1965) Classical normal modes in damped linear dynamic systems. *Trans ASME J Appl Mech* 32:583–588
- McTavish DJ, Hughes PC (1993) Modeling of linear viscoelastic space structures. *Trans ASME J Vib Acoust* 115:103–110
- Meirovitch L (1997) *Principles and techniques of vibrations*. Prentice-Hall International, New Jersey
- Muravyov A (1998) Forced vibration responses of a viscoelastic structure. *J Sound Vib* 218(5):892–907
- Muravyov A, Hutton SG (1997) Closed-form solutions and the eigenvalue problem for vibration of discrete viscoelastic systems. *Trans ASME J Appl Mech* 64:684–691
- Rayleigh JW (1877) *Theory of sound* (two volumes), 1945 re-issue, 2nd edn. Dover Publications, New York
- Udwadia FE (2009) A note on nonproportional damping. *J Eng Mech-ASCE* 135(11):1248–1256
- Wagner N, Adhikari S (2003) Symmetric state-space formulation for a class of non-viscously damped systems. *AIAA J* 41(5):951–956
- Zhang J, Zheng GT (2007) The biot model and its application in viscoelastic composite structures. *J Vib Acoust* 129:533–540

Subset Simulation Method for Rare Event Estimation: An Introduction

Konstantin M. Zuev

Department of Computing and Mathematical Sciences, California Institute of Technology, USA

Synonyms

Engineering reliability; Failure probability; Markov chain Monte Carlo; Monte Carlo simulation; Rare events; Subset Simulation

Introduction

This entry provides a detailed introductory description of Subset Simulation, an advanced stochastic simulation method for estimation of small probabilities of rare failure events. A simple and intuitive derivation of the method is given along with the discussion on its

implementation. The method is illustrated with several easy-to-understand examples. The reader is assumed to be familiar only with elementary probability theory and statistics.

Subset Simulation (SS) is an efficient and elegant method for simulating rare events and estimating the corresponding small tail probabilities. The method was originally developed by Siu-Kui Au and James Beck in the already classical paper (Au and Beck 2001a) for estimation of structural reliability of complex civil engineering systems such as tall buildings and bridges at risk from earthquakes. The method turned out to be so powerful and general that over the last decade, SS has been successfully applied to reliability problems in geotechnical, aerospace, fire, and nuclear engineering. Moreover, the idea of SS proved to be useful not only in reliability analysis but also in other problems associated with general engineering systems, such as sensitivity analysis, design optimization, and uncertainty quantification. As of October 2014, according to the Web of Science (ISI) database and Google Scholar, the original SS paper (Au and Beck 2001a) received 315 and 572 citations respectively, that indicates the high impact of the Subset Simulation method on the engineering research community.

Subset Simulation is essentially based on two different ideas: conceptual and technical. The conceptual idea is to decompose the rare event F into a sequence of progressively “less-rare” nested events,

$$F = F_m \subset F_{m-1} \subset \dots \subset F_1, \quad (1)$$

where F_1 is a relatively frequent event. For example, suppose that F represents the event of getting exactly m heads when flipping a fair coin m times. If m is large, then F is a rare event. To decompose F into a sequence (Eq. 1), let us define F_k to be the event of getting exactly k heads in the first k flips, where $k = 1, \dots, m$. The smaller the k , the less rare the corresponding event F_k , and F_1 – getting heads in the first flip – is relatively frequent.

Given a sequence of subsets (Eq. 1), the small probability $\mathbb{P}(F)$ of the rare event F can then be

represented as a product of larger probabilities as follows:

$$\begin{aligned} \mathbb{P}(F) &= \mathbb{P}(F_m) \\ &= \mathbb{P}(F_1) \frac{\mathbb{P}(F_2) \mathbb{P}(F_3)}{\mathbb{P}(F_1) \mathbb{P}(F_2)} \dots \frac{\mathbb{P}(F_{m-1}) \mathbb{P}(F_m)}{\mathbb{P}(F_{m-2}) \mathbb{P}(F_{m-1})} \\ &= \mathbb{P}(F_1) \cdot \mathbb{P}(F_2|F_1) \cdot \dots \cdot \mathbb{P}(F_m|F_{m-1}), \end{aligned} \quad (2)$$

where $\mathbb{P}(F_k|F_{k-1}) = \mathbb{P}(F_k)/\mathbb{P}(F_{k-1})$ denotes the conditional probability of event F_k given the occurrence of event F_{k-1} , for $k = 2, \dots, m$. In the coin example, $\mathbb{P}(F_1) = 1/2$, all conditional probabilities $\mathbb{P}(F_k|F_{k-1}) = 1/2$, and the probability of the rare event $\mathbb{P}(F) = 1/2^m$.

Unlike the coin example, in real applications, it is often not obvious how to decompose the rare event into a sequence in Eq. 1 and how to compute all conditional probabilities in Eq. 2. In Subset Simulation, the “sequencing” of the rare event is done adaptively as the algorithm proceeds. This is achieved by employing Markov chain Monte Carlo, an advanced simulation technique, which constitutes the second – technical – idea behind SS. Finally, all conditional probabilities are automatically obtained as a by-product of the adaptive sequencing.

The main goals of this entry are (a) to provide a detailed exposition of Subset Simulation at an introductory level, (b) to give a simple derivation of the method and discuss its implementation, and (c) to illustrate SS with intuitive examples. Although the scope of SS is much wider, in this entry the method is described in the context of engineering reliability estimation; the problem SS was originally developed for in Au and Beck (2001a).

The rest of the entry is organized as follows: section “[Engineering Reliability Problem](#)” describes the engineering reliability problem and explains why this problem is computationally challenging. Section “[The Direct Monte Carlo Method](#)” discusses how the Direct Monte Carlo method can be used for engineering reliability estimation and why it is often inefficient. In section “[Preprocessing: Transformation of Input Variables](#),” a necessary preprocessing step which is

often used by many reliability methods is briefly discussed. Section “[The Subset Simulation Method](#)” is the core of the entry, where the SS method is explained. Illustrative examples are considered in section “[Illustrative Examples](#).” For demonstration purposes, the MATLAB code for the considered examples is provided in section “[MATLAB code](#).” Section “[Summary](#)” concludes the entry with a brief summary.

Engineering Reliability Problem

One of the most important and computationally challenging problems in reliability engineering is to estimate the probability of failure for a system, that is, the probability of unacceptable system performance. The behavior of the system can be described by a *response variable* y , which may represent, for example, the roof displacement or the largest interstory drift. The response variable depends on *input variables* $x = (x_1, \dots, x_d)$, also called *basic variables*, which may represent geometry, material properties, and loads,

$$y = g(x_1, \dots, x_d), \quad (3)$$

where $g(x)$ is called the *performance function*. The performance of the system is measured by comparison of the response y with a specified critical value y^* : if $y \leq y^*$, then the system is safe; if $y > y^*$, then the system has failed. This failure criterion allows to define the *failure domain* F in the input x -space as follows:

$$F = \{x : g(x) > y^*\}. \quad (4)$$

In other words, the failure domain is a set of values of input variables that lead to unacceptable system performance, namely, to the exceedance of some prescribed critical threshold y^* , which may represent the maximum permissible roof displacement, maximum permissible interstory drift, etc.

Engineering systems are complex systems, where complexity, in particular, means that the information about the system (its geometric and material properties) and its environment (loads)

is never complete. Therefore, there are always uncertainties in the values of input variables x . To account for these uncertainties, the input variables are modeled as random variables whose marginal distributions are usually obtained from test data, from expert opinion, or from literature. Let $\pi(x)$ denote the joint probability density function (PDF) for x . The uncertainty in the input variables is propagated through Eq. 3 into the response variable y , which makes the *failure event* $\{x \in F\} = \{y > y^*\}$ also uncertain. The *engineering reliability problem* is then to compute the probability of failure p_F , given by the following expression:

$$p_F = \mathbb{P}(x \in F) = \int_F \pi(x) dx. \quad (5)$$

The behavior of complex systems, such as tall buildings and bridges, is represented by a complex model (3). In this context, complexity means that the performance function $g(x)$, which defines the integration region F in Eq. 5, is not explicitly known. The evaluation of $g(x)$ for any x is often time-consuming and usually done by the finite element method (FEM), one of the most important numerical tools for computation of the response of engineering systems. Thus, it is usually impossible to evaluate the integral in Eq. 5 analytically because the integration region, the failure domain F , is not known explicitly.

Moreover, traditional numerical integration is also generally not applicable. In this approach, the d -dimensional input x -space is partitioned into a union of disjoint hypercubes, $\square_1, \dots, \square_N$. For each hypercube $\square_i$, a “representative” point $x^{(i)}$ is chosen inside that hypercube, $x^{(i)} \in \square_i$. The integral in Eq. 5 is then approximated by the following sum:

$$p_F \approx \sum_{x^{(i)} \in F} \pi(x^{(i)}) \text{vol}(\square_i), \quad (6)$$

where $\text{vol}(\square_i)$ denotes the volume of $\square_i$ and summation is taken over all failure points $x^{(i)}$. Since it is not known in advance whether

a given point is a failure point or not (the failure domain F is not known explicitly), to compute the sum in Eq. 6, the failure criterion in Eq. 4 must be checked for all $x(i)$. Therefore, the approximation in Eq. 6 becomes

$$p_F \approx \sum_{i=1}^N I_F(x^{(i)}) \pi(x^{(i)}) \text{vol}(\square_i), \quad (7)$$

where $I_F(x)$ stands for the indicator function, i.e.,

$$I_F(x) = \begin{cases} 1, & \text{if } x \in F, \\ 0, & \text{if } x \notin F. \end{cases} \quad (8)$$

If n denotes the number of intervals each dimension of the input space is partitioned into, then the total number of terms in Eq. 7 is $N = n^d$. Therefore, the computational effort of numerical integration grows exponentially with the number of dimensions d . In engineering reliability problems, the dimension of the input space is typically very large (e.g., when the stochastic load time history is discretized in time). For example, $d \sim 10^3$ is not unusual in the reliability literature. This makes numerical integration computationally infeasible.

Over the past few decades, many different methods for solving the engineering reliability problem (5) have been developed. In general, the proposed reliability methods can be classified into three categories, namely:

- (a) *Analytic methods* are based on the Taylor-series expansion of the performance function, e.g., the first-order reliability method (FORM) and the second-order reliability method (SORM) (Ditlevsen and Madsen 1996; Madsen et al. 2006; Melchers 1999).
- (b) *Surrogate methods* are based on a functional surrogate of the performance function, e.g., the response surface method (RSM) (Faravelli 1989; Schuëller et al. 1989; Bucher 1990), Neural Networks (Papadrakakis et al. 1996), support vector machines (Hurtado and Alvarez 2003), and other methods (Hurtado 2004).

- (c) *Monte Carlo simulation methods*, among which are Importance Sampling (Engelund and Rackwitz 1993), Importance Sampling using Elementary Events (Au and Beck 2001b), Radial-based Importance Sampling (Grooteman 2008), Adaptive Linked Importance Sampling (Katafygiotis and Zuev 2007), Directional Simulation (Ditlevsen and Madsen 1996), Line Sampling (Koutsourelakis et al. 2004), Auxiliary Domain Method (Katafygiotis et al. 2007), Horseracing Simulation (Zuev and Katafygiotis 2011), and *Subset Simulation* (Au and Beck 2001a).

Subset Simulation is thus a reliability method which is based on (advanced) Monte Carlo simulation.

The Direct Monte Carlo Method

The Monte Carlo method, referred in this entry as *Direct Monte Carlo* (DMC), is a statistical sampling technique that has been originally developed by Stan Ulam, John von Neumann, Nick Metropolis (who actually suggested the name “Monte Carlo” (Metropolis 1987)), and their collaborators for solving the problem of neutron diffusion and other problems in mathematical physics (Metropolis and Ulam 1949). From a mathematical point of view, DMC allows to estimate the expected value of a quantity of interest. More specifically, suppose the goal is to evaluate $\mathbb{E}_\pi[h(x)]$, that is, an expectation of a function $h: \mathcal{X} \rightarrow \mathbb{R}$ with respect to the PDF $\pi(x)$,

$$\mathbb{E}_\pi[h(x)] = \int_{\mathcal{X}} h(x) \pi(x) dx. \quad (9)$$

The idea behind DMC is a straightforward application of the *law of large numbers* that states that if $x^{(1)}, x^{(2)}, \dots$ are i.i.d. (independent and identically distributed) from the PDF $\pi(x)$, then the empirical average $\frac{1}{N} \sum_{i=1}^N h(x^{(i)})$ converges to the true value $\mathbb{E}_\pi[h(x)]$ as N goes to $+\infty$. Therefore, if the number of samples N is large enough,

then $\mathbb{E}_\pi[h(x)]$ can be accurately estimated by the corresponding empirical average:

$$\mathbb{E}_\pi[h(x)] \approx \frac{1}{N} \sum_{i=1}^N h(x^{(i)}). \quad (10)$$

The relevance of DMC to the reliability problem (5) follows from a simple observation that the failure probability p_F can be written as an expectation of the indicator function (8), namely,

$$p_F = \int_F \pi(x) dx = \int_{\mathcal{X}} I_F(x) \pi(x) dx = \mathbb{E}_\pi[I_F(x)], \quad (11)$$

where $\mathcal{X}$ denotes the entire input x -space. Therefore, the failure probability can be estimated using the DMC method (10) as follows:

$$p_F \approx \hat{p}_F^{\text{DMC}} = \frac{1}{N} \sum_{i=1}^N I_F(x^{(i)}), \quad (12)$$

where $x^{(1)}, \dots, x^{(N)}$ are i.i.d. samples from $\pi(x)$.

The DMC estimate of p_F is thus just the ratio of the total number of *failure samples* $\sum_{i=1}^N I_F(x^{(i)})$, i.e., samples that produce system failure according to the system model, to the total number of samples, N . Note that $\hat{p}_F^{\text{DMC}}$ is an *unbiased* random estimate of the failure probability, that is, on average, $\hat{p}_F^{\text{DMC}}$ equals to p_F . Mathematically, this means that $\mathbb{E}[\hat{p}_F^{\text{DMC}}] = p_F$. Indeed, using the fact that $x^{(i)} \sim \pi(x)$ and in Eq. 11,

$$\begin{aligned} \mathbb{E}[\hat{p}_F^{\text{DMC}}] &= \mathbb{E}\left[\frac{1}{N} \sum_{i=1}^N I_F(x^{(i)})\right] \\ &= \frac{1}{N} \sum_{i=1}^N \mathbb{E}[I_F(x^{(i)})] \\ &= \frac{1}{N} \sum_{i=1}^N \mathbb{E}_\pi[I_F(x)] = p_F. \end{aligned} \quad (13)$$

The main advantage of DMC over numerical integration is that its accuracy does not depend on the dimension d of the input space. In reliability analysis, the standard measure of accuracy

of an unbiased estimate $\hat{p}_F$ of the failure probability is its *coefficient of variation* (c.o.v.) $\delta(\hat{p}_F)$, which is defined as the ratio of the standard deviation to the expected value of $\hat{p}_F$, i.e., $\delta(\hat{p}_F) = \sqrt{\mathbb{V}[\hat{p}_F] / \mathbb{E}[\hat{p}_F]}$, where $\mathbb{V}$ denotes the variance. The smaller the c.o.v. $\delta(\hat{p}_F)$, the more accurate the estimate $\hat{p}_F$ is. It is straightforward to calculate the variance of the DMC estimate:

$$\begin{aligned} \mathbb{V}[\hat{p}_F^{\text{DMC}}] &= \mathbb{V}\left[\frac{1}{N} \sum_{i=1}^N I_F(x^{(i)})\right] = \frac{1}{N^2} \sum_{i=1}^N \mathbb{V}[I_F(x^{(i)})] \\ &= \frac{1}{N^2} \sum_{i=1}^N \left(\mathbb{E}[I_F(x^{(i)})^2] - \mathbb{E}[I_F(x^{(i)})]^2 \right) \\ &= \frac{1}{N^2} \sum_{i=1}^N (p_F - p_F^2) = \frac{p_F(1 - p_F)}{N}. \end{aligned} \quad (14)$$

Here, the identity $I_F(x)^2 = I_F(x)$ was used. Using Eqs. 13 and 14, the c.o.v. of the DMC estimate can be calculated:

$$\delta(\hat{p}_F^{\text{DMC}}) = \frac{\sqrt{\mathbb{V}[\hat{p}_F^{\text{DMC}}]}}{\mathbb{E}[\hat{p}_F^{\text{DMC}}]} = \sqrt{\frac{1 - p_F}{N p_F}}. \quad (15)$$

This result shows that $\delta(\hat{p}_F^{\text{DMC}})$ depends only on the failure probability p_F and the total number of samples N and does not depend on the dimension d of the input space. Therefore, unlike numerical integration, the DMC method does not suffer from the “curse of dimensionality,” i.e., from an exponential increase in volume associated with adding extra dimensions, and is able to handle problems of high dimension.

Nevertheless, DMC has a serious drawback: it is inefficient in estimating small failure probabilities. For typical engineering reliability problems, the failure probability p_F is very small, $p_F \ll 1$. In other words, the system is usually assumed to be designed properly, so that its failure is a *rare event*. In the reliability literature, $p_F \sim 10^{-2} - 10^{-9}$ have been considered. If p_F is very small, then it follows from Eq. 15 that

$$\delta(\hat{p}_F^{\text{DMC}}) \approx \frac{1}{\sqrt{N p_F}}. \quad (16)$$

This means that the number of samples N needed to achieve an acceptable level of accuracy is inverse proportional to p_F , and therefore very large, $N \propto 1/p_F \gg 1$. For example, if $p_F = 10^{-4}$ and the c.o.v. of 10 % is desirable, then $N = 10^6$ samples are required. Note, however, that each evaluation of $I_F(x^{(i)})$, $i = 1, \dots, N$, in Eq. 12 requires a system analysis to be performed to check whether the sample $x^{(i)}$ is a failure sample. As it has been already mentioned in section “[Engineering Reliability Problem](#),” the computation effort for the system analysis, i.e., computation of the performance function $g(x)$, is significant (usually involves the FEM method). As a result, the DMC method becomes excessively costly and practically inapplicable for reliability analysis. This deficiency of DMC has motivated research to develop more advanced simulation algorithms for efficient estimation of small failure probabilities in high dimensions.

Remark 1 It is important to highlight, however, that even though DMC cannot be routinely used for reliability problems (too expensive), it is a very robust method, and it is often used as a check on other reliability methods.

Preprocessing: Transformation of Input Variables

Many reliability methods, including Subset Simulation, assume that the input variables x are independent. This assumption, however, is not a limitation, since in simulation one always starts from independent variables to generate the dependent input variables. Furthermore, for convenience, it is often assumed that x are i.i.d. Gaussian. If this is not the case, a “preprocessing” step that transforms x to i.i.d. Gaussian variables z must be undertaken. The transformation from x to z can be performed in several ways depending on the available information about the input variables. In the simplest case, when x are independent Gaussians, $x_k \sim \mathcal{N}(\cdot | \mu_k, \sigma_k^2)$, where μ_k and σ_k^2 are,

respectively, the mean and variance of x_k , the necessary transformation is standardization:

$$z_k = \frac{x_k - \mu_k}{\sigma_k}. \quad (17)$$

In other cases, more general techniques should be used, such as the Rosenblatt transformation (Rosenblatt 1952) and the Nataf transformation (Nataf 1962). To avoid introduction of additional notation, hereinafter, it is assumed without loss of generality that the vector x has been already transformed and it follows the standard multivariate Gaussian distribution,

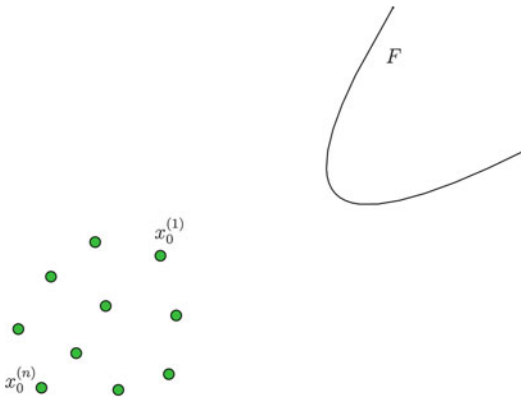
$$\pi(x_1, \dots, x_d) = \prod_{k=1}^d \phi(x_k), \quad (18)$$

where $\phi(\cdot)$ denotes the standard Gaussian PDF,

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}. \quad (19)$$

The Subset Simulation Method

Unlike Direct Monte Carlo, where all computational resources are directly spent on sampling the input space $x^{(1)}, \dots, x^{(N)} \sim \pi(\cdot)$ and computing the values of the performance function $g(x^{(1)}), \dots, g(x^{(N)})$, Subset Simulation first “probes” the input space $\mathcal{X}$ by generating a relatively small number of i.i.d samples $x_0^{(1)}, \dots, x_0^{(n)} \sim \pi(x)$, $n < N$, and computing the corresponding system responses $y_0^{(1)} = g(x_0^{(1)}), \dots, y_0^{(n)} = g(x_0^{(n)})$. Here, the subscript 0 indicates the 0th stage of the algorithm. Since F is a rare event and n is relatively small, it is very likely that none of the samples $x_0^{(1)}, \dots, x_0^{(n)}$ belongs to F , that is, $y_0^{(i)} < y^*$ for all $i = 1, \dots, n$. Nevertheless, these Monte Carlo samples contain some useful information about the failure domain that can be utilized. To keep the notation simple, assume that $y_0^{(1)}, \dots, y_0^{(n)}$ are arranged in the decreasing order, i.e., $y_0^{(1)} \geq \dots \geq y_0^{(n)}$ (it is always possible to achieve this by renumbering $x_0^{(1)}, \dots, x_0^{(n)}$ if needed). Then, $x_0^{(1)}$ and $x_0^{(n)}$ are, respectively, the closest to failure and the safest samples among $x_0^{(1)}, \dots, x_0^{(n)}$, since $y_0^{(1)}$ and $y_0^{(n)}$ are the largest and the smallest responses. In general,



Subset Simulation Method for Rare Event Estimation: An Introduction, Fig. 1 Monte Carlo samples $x_0^{(1)}, \dots, x_0^{(n)}$ and the failure domain F . $x_0^{(1)}$ and $x_0^{(n)}$ are, respectively, the closest to failure and the safest samples among $x_0^{(1)}, \dots, x_0^{(n)}$

the smaller the i , the closer to failure the sample $x_0^{(i)}$ is. This is shown schematically in Fig. 1.

Let $p \in (0, 1)$ be any number such that np is integer. By analogy with Eq. 4, define the *first intermediate failure domain* F_1 as follows:

$$F_1 = \{x : g(x) > y_1^*\}, \quad (20)$$

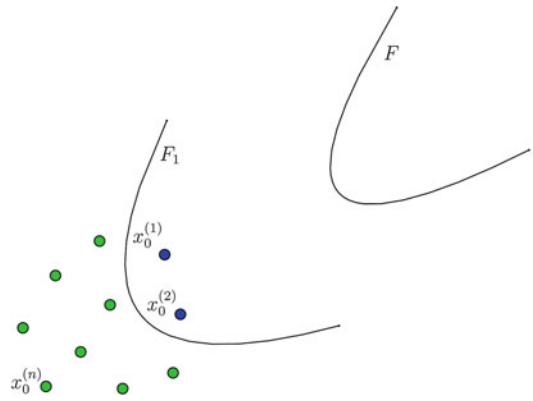
where

$$y_1^* = \frac{y_0^{(np)} + y_0^{(np+1)}}{2}. \quad (21)$$

In other words, F_1 is the set of inputs that lead to the exceedance of the *relaxed threshold* $y_1^* < y^*$. Note that by construction, samples $x_0^{(1)}, \dots, x_0^{(np)}$ belong to F_1 , while $x_0^{(np+1)}, \dots, x_0^{(n)}$ do not. As a consequence, the Direct Monte Carlo estimate for the probability of F_1 which is based on samples $x_0^{(1)}, \dots, x_0^{(n)}$ is automatically equal to p ,

$$\mathbb{P}(F_1) \approx \frac{1}{n} \sum_{i=1}^n I_{F_1}(x_0^{(i)}) = p. \quad (22)$$

The value $p = 0.1$ is often used in the literature, which makes F_1 a relatively frequent event. Figure 2 illustrates the definition of F_1 .



Subset Simulation Method for Rare Event Estimation: An Introduction, Fig. 2 The first intermediate failure domain F_1 . In this schematic illustration, $n = 10$, $p = 0.2$, so that there are exactly $np = 2$ Monte Carlo samples in F_1 , $x_0^{(1)}, x_0^{(2)} \in F_1$

The first intermediate failure domain F_1 can be viewed as a (very rough) conservative approximation to the target failure domain F . Since $F \subset F_1$, the failure probability p_F can be written as a product:

$$p_F = \mathbb{P}(F_1)\mathbb{P}(F|F_1), \quad (23)$$

where $\mathbb{P}(F|F_1)$ is the conditional probability of F given F_1 . Therefore, in view of Eq. 22, the problem of estimating p_F is reduced to estimating the conditional probability $\mathbb{P}(F|F_1)$.

In the next stage, instead of generating samples in the whole input space (like in DMC), the SS algorithm aims to populate F_1 . Specifically, the goal is to generate samples $x_1^{(1)}, \dots, x_1^{(n)}$ from the conditional distribution

$$\pi(x|F_1) = \frac{\pi(x)I_{F_1}(x)}{\mathbb{P}(F_1)} = \frac{I_{F_1}(x)}{\mathbb{P}(F_1)} \prod_{k=1}^d \phi(x_k). \quad (24)$$

First of all, note that samples $x_0^{(1)}, \dots, x_0^{(np)}$ not only belong to F_1 but are also distributed according to $\pi(\cdot|F_1)$. To generate the remaining $(n - np)$ samples from $\pi(\cdot|F_1)$, which, in general, is not a trivial task, Subset Simulation uses the so-called Modified Metropolis algorithm (MMA). MMA belongs to the class of *Markov chain Monte Carlo* (MCMC) algorithms (Liu 2001;

Robert and Casella 2004), which are techniques for sampling from complex probability distributions that cannot be sampled directly, at least not efficiently. MMA is based on the original Metropolis algorithm (Metropolis et al. 1953) and specifically tailored for sampling from the conditional distributions of the form (24).

Modified Metropolis Algorithm

Let $x \sim \pi(\cdot|F_1)$ be a sample from the conditional distribution $\pi(\cdot|F_1)$. The Modified Metropolis algorithm generates another sample $\tilde{x}$ from $\pi(\cdot|F_1)$ as follows:

1. Generate a “candidate” sample ξ : For each coordinate $k = 1, \dots, d$,
 - (a) Sample $\eta_k \sim q_k(\cdot|x_k)$, where $q_k(\cdot|x_k)$, called the *proposal distribution*, is a univariate PDF for η_k centered at x_k with the symmetry property $q_k(\eta_k|x_k) = q_k(x_k|\eta_k)$. For example, the proposal distribution can be a Gaussian PDF with mean x_k and variance σ_k^2 ,

$$q_k(\eta_k|x_k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(\eta_k - x_k)^2}{2\sigma_k^2}\right), \quad (25)$$

or it can be a uniform distribution over $[x_k - \alpha, x_k + \alpha]$, for some $\alpha \geq 0$.

- (b) Compute the acceptance ratio

$$r_k = \frac{\phi(\eta_k)}{\phi(x_k)}. \quad (26)$$

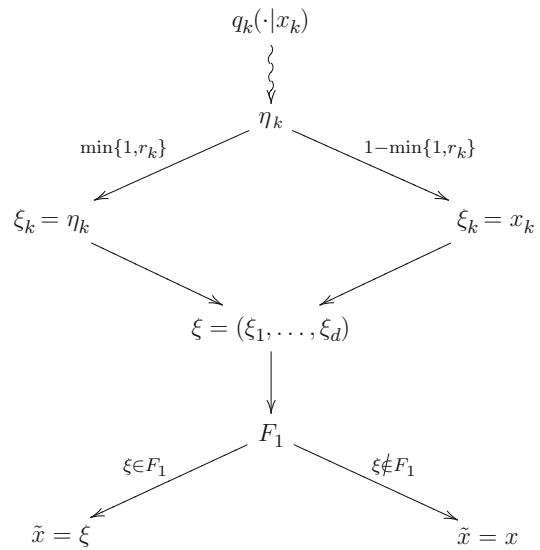
- (c) Define the k th coordinate of the candidate sample by accepting or rejecting η_k ,

$$\xi_k = \begin{cases} \eta_k, & \text{with probability } \min\{1, r_k\}, \\ x_k, & \text{with probability } 1 - \min\{1, r_k\}. \end{cases} \quad (27)$$

2. Accept or reject the candidate sample ξ by setting

$$\tilde{x} = \begin{cases} \xi, & \text{if } \xi \in F_1, \\ x, & \text{if } \xi \notin F_1. \end{cases} \quad (28)$$

The Modified Metropolis algorithm is schematically illustrated in Fig. 3.



Subset Simulation Method for Rare Event Estimation: An Introduction, Fig. 3 Modified Metropolis algorithm

It can be shown that the sample $\tilde{x}$ generated by MMA is indeed distributed according to $\pi(\cdot|F_1)$. If the candidate sample ξ is rejected in Eq. 28, then $\tilde{x} = x \sim \pi(\cdot|F_1)$ and there is nothing to prove. Suppose now that ξ is accepted, $\tilde{x} = \xi$, so that the move from x to $\tilde{x}$ is a proper transition between two distinct points in F_1 . Let $f(\cdot)$ denote the PDF of $\tilde{x}$ (the goal is to show that $f(\tilde{x}) = \pi(\tilde{x}|F_1)$). Then

$$f(\tilde{x}) = \int_{F_1} \pi(x|F_1) t(\tilde{x}|x) dx, \quad (29)$$

where $t(\tilde{x}|x)$ is the transition PDF from x to $\tilde{x} \neq x$. According to the first step of MMA, coordinates of $\tilde{x} = \xi$ are generated independently, and therefore $t(\tilde{x}|x)$ can be expressed as a product,

$$t(\tilde{x}|x) = \prod_{k=1}^d t_k(\tilde{x}_k|x_k), \quad (30)$$

where $t_k(\tilde{x}_k|x_k)$ is the transition PDF for the k th coordinate $\tilde{x}_k$. Combining Eqs. 24, 29, and 30 gives

$$\begin{aligned}
 f(\tilde{x}) &= \int_{F_1} \frac{I_{F_1}(x)}{\mathbb{P}(F_1)} \prod_{k=1}^d \phi(x_k) \prod_{k=1}^d t_k(\tilde{x}_k|x_k) dx \\
 &= \frac{1}{\mathbb{P}(F_1)} \int_{F_1} \prod_{k=1}^d \phi(x_k) t_k(\tilde{x}_k|x_k) dx. \quad (31)
 \end{aligned}$$

The key to the proof of $f(\tilde{x}) = \pi(\tilde{x}|F_1)$ is to demonstrate that $\phi(x_k)$ and $t_k(\tilde{x}_k|x_k)$ satisfy the so-called detailed balance equation,

$$\phi(x_k) t_k(\tilde{x}_k|x_k) = \phi(\tilde{x}_k) t_k(x_k|\tilde{x}_k). \quad (32)$$

If $\tilde{x}_k = x_k$, then Eq. 32 is trivial. Suppose that $\tilde{x}_k \neq x_k$, that is, $\tilde{x}_k = \xi_k = \eta_k$ in Eq. 27. The actual transition PDF $t_k(\tilde{x}_k|x_k)$ from x_k to $\tilde{x}_k \neq x_k$ differs from the proposal PDF $q_k(\tilde{x}_k|x_k)$ because the acceptance-rejection step in Eq. 27 is involved. To actually make the move from x_k to $\tilde{x}_k$, one needs not only to generate $\tilde{x}_k \sim q_k(\cdot|x_k)$ but also to accept it with probability $\min\left\{1, \frac{\phi(\tilde{x}_k)}{\phi(x_k)}\right\}$. Therefore,

$$t_k(\tilde{x}_k|x_k) = q_k(\tilde{x}_k|x_k) \min\left\{1, \frac{\phi(\tilde{x}_k)}{\phi(x_k)}\right\}, \quad \tilde{x}_k \neq x_k. \quad (33)$$

Using Eq. 33, the symmetry property of the proposal PDF, $q_k(\tilde{x}_k|x_k) = q_k(x_k|\tilde{x}_k)$, and the identity $a \min\left\{1, \frac{b}{a}\right\} = b \min\left\{1, \frac{a}{b}\right\}$ for any $a, b > 0$,

$$\begin{aligned}
 \phi(x_k) t_k(\tilde{x}_k|x_k) &= q_k(\tilde{x}_k|x_k) \phi(x_k) \min\left\{1, \frac{\phi(\tilde{x}_k)}{\phi(x_k)}\right\} \\
 &= q_k(x_k|\tilde{x}_k) \phi(\tilde{x}_k) \min\left\{1, \frac{\phi(x_k)}{\phi(\tilde{x}_k)}\right\} \\
 &= \phi(\tilde{x}_k) t_k(x_k|\tilde{x}_k), \quad (34)
 \end{aligned}$$

and the detailed balance in Eq. 32 is thus established. The rest is a straightforward calculation:

$$\begin{aligned}
 f(\tilde{x}) &= \frac{1}{\mathbb{P}(F_1)} \int_{F_1} \prod_{k=1}^d \phi(\tilde{x}_k) t_k(x_k|\tilde{x}_k) dx \\
 &= \frac{1}{\mathbb{P}(F_1)} \prod_{k=1}^d \phi(\tilde{x}_k) \int_{F_1} t(x|\tilde{x}) dx \\
 &= \pi(\tilde{x}|F_1), \quad (35)
 \end{aligned}$$

since the transition PDF $t(x|\tilde{x})$ integrates to 1, and $I_{F_1}(\tilde{x}) = 1$.

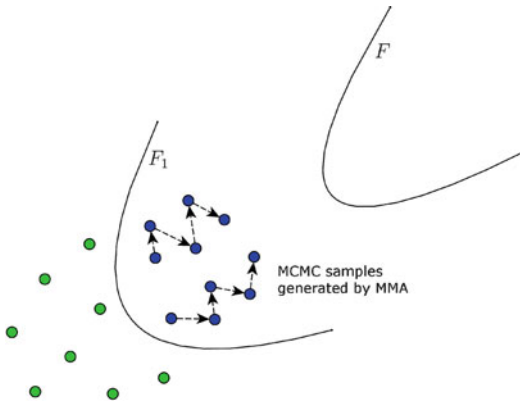
Remark 2 A mathematically more rigorous proof of the Modified Metropolis algorithm is given in (Zuev et al. 2012).

Remark 3 It is worth mentioning that although the independence of input variables is crucial for the applicability of MMA, and thus for Subset Simulation, they need not be identically distributed. In other words, instead of Eq. 18, the joint PDF $\pi(\cdot)$ can have a more general form, $\pi(x) = \prod_{k=1}^d \pi_k(x_k)$, where $\pi^k(\cdot)$ is the marginal distributions of x_k which is not necessarily Gaussian. In this case, the expression for the acceptance ratio in Eq. 26 must be replaced by $r_k = \frac{\pi_k(\eta_k)}{\pi_k(x_k)}$.

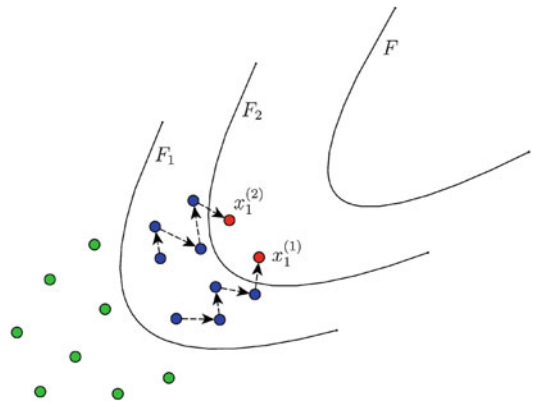
Subset Simulation at Higher Conditional Levels

Given $x_0^{(1)}, \dots, x_0^{(np)} \sim \pi(\cdot|F_1)$, it is clear now how to generate the remaining $(n - np)$ samples from $\pi(\cdot|F_1)$. Namely, starting from each $x_0^{(i)}$, $i = 1, \dots, np$, the SS algorithm generates a sequence of $\left(\frac{1}{p} - 1\right)$ new MCMC samples $x_0^{(i)} = x_{0,0}^{(i)} \mapsto x_{0,1}^{(i)} \mapsto \dots \mapsto x_{0,\frac{1}{p}-1}^{(i)}$ using the Modified Metropolis transition rule described above. Note that when $x_{0,j}^{(i)}$ is generated, the previous sample $x_{0,j-1}^{(i)}$ is used as an input for the transition rule. The sequence $x_{0,0}^{(i)}, x_{0,1}^{(i)}, \dots, x_{0,\frac{1}{p}-1}^{(i)}$ is called a *Markov chain* with the *stationary distribution* $\pi(\cdot|F_1)$, and $x_{0,0}^{(i)} = x_0^{(i)}$ is often referred to as the “seed” of the Markov chain.

To simplify the notation, denote samples $\{x_{0,j}^{(i)}\}_{j=0, \dots, \frac{1}{p}-1}^{i=1, \dots, np}$ by $\{x_1^{(1)}, \dots, x_1^{(n)}\}$. The subscript 1 indicates that the MCMC samples $x_1^{(1)}, \dots, x_1^{(n)} \sim \pi(\cdot|F_1)$ are generated at the first conditional level of the SS algorithm. These conditional samples are schematically shown in Fig. 4. Also assume that the corresponding system responses $y_1^{(1)} = g(x_1^{(1)}), \dots, y_1^{(n)} = g(x_1^{(n)})$ are arranged in the decreasing order, i.e., $y_1^{(1)} \geq \dots \geq y_1^{(n)}$. If the failure event F is rare enough, that is, if p_F is sufficiently small, then it is very likely that none of the samples $x_1^{(1)}, \dots, x_1^{(n)}$ belongs to F , i.e., $y_1^{(i)} < y^*$ for all $i = 1, \dots, n$. Nevertheless, these MCMC samples can be used in the similar way the Monte Carlo samples $x_0^{(1)}, \dots, x_0^{(n)}$ were used.



Subset Simulation Method for Rare Event Estimation: An Introduction, Fig. 4 MCMC samples generated by the Modified Metropolis algorithm at the first conditional level of Subset Simulation



Subset Simulation Method for Rare Event Estimation: An Introduction, Fig. 5 The second intermediate failure domain F_2 . In this schematic illustration, $n = 10$, $p = 0.2$, so that there are exactly $np = 2$ MCMC samples in F_2 , $x_1^{(1)}, x_1^{(2)} \in F_2$

By analogy with Eq. 20, define the *second intermediate failure domain* F_2 as follows:

$$F_2 = \{x : g(x) > y_2^*\}, \quad (36)$$

where

$$y_2^* = \frac{y_1^{(np)} + y_1^{(np+1)}}{2}. \quad (37)$$

Note that $y_2^* > y_1^*$ since $y_1^{(i)} > y_1^*$ for all $i = 1, \dots, n$. This means that $F \subset F_2 \subset F_1$, and therefore, F_2 can be viewed as a conservative approximation to F which is still rough, yet more accurate than F_1 . Figure 5 illustrates the definition of F_2 . By construction, samples $x_1^{(1)}, \dots, x_1^{(np)}$ belong to F_2 , while $x_1^{(np+1)}, \dots, x_1^{(n)}$ do not. As a result, the estimate for the conditional probability of F_2 given F_1 which is based on samples $x_1^{(1)}, \dots, x_1^{(n)} \sim \pi(\cdot|F_1)$ is automatically equal to p ,

$$\mathbb{P}(F_2|F_1) \approx \frac{1}{n} \sum_{i=1}^n I_{F_2}(x_1^{(i)}) = p. \quad (38)$$

Since $F \subset F_2 \subset F_1$, the conditional probability $\mathbb{P}(F|F_1)$ that appears in Eq. 23 can be expressed as a product:

$$\mathbb{P}(F|F_1) = \mathbb{P}(F_2|F_1)\mathbb{P}(F|F_2). \quad (39)$$

Combining Eqs. 23 and 39 gives the following expression for the failure probability:

$$p_F = \mathbb{P}(F_1)\mathbb{P}(F_2|F_1)\mathbb{P}(F|F_2). \quad (40)$$

Thus, in view of Eqs. 22 and 38, the problem of estimating p_F is now reduced to estimating the conditional probability $\mathbb{P}(F|F_2)$.

In the next step, as one may have already guessed, the Subset Simulation algorithm populates F_2 by generating MCMC samples $x_2^{(1)}, \dots, x_2^{(n)}$ from $\pi(\cdot|F_2)$ using the Modified Metropolis algorithm, defines the third intermediate failure domain $F_3 \subset F_2$ such that $\mathbb{P}(F_3|F_2) \approx \frac{1}{n} \sum_{i=1}^n I_{F_3}(x_2^{(i)}) = p$, and reduces the original problem of estimating the failure probability p_F to estimating the conditional probability $\mathbb{P}(F|F_3)$ by representing $p_F = \mathbb{P}(F_1)\mathbb{P}(F_2|F_1)\mathbb{P}(F_3|F_2)\mathbb{P}(F|F_3)$. The algorithm proceeds in this way until the target failure domain F has been sufficiently sampled so that the conditional probability $\mathbb{P}(F|F_L)$ can be accurately estimated by $\frac{1}{n} \sum_{i=1}^n I_F(x_L^{(i)})$, where F_L is the L th intermediate failure domain and $x_L^{(1)}, \dots, x_L^{(n)} \sim \pi(\cdot|F_L)$ are the MCMC samples generated at the L th conditional level. Subset Simulation can thus be viewed as a method that decomposes the rare failure event F into a sequence of

progressively “less-rare” nested events, $F \subset F_L \subset \dots \subset F_1$, where all intermediate failure events $F_1, \dots, F_L$ are constructed adaptively by appropriately relaxing the value of the critical threshold $y_1^* < \dots < y_L^* < y^*$.

Stopping Criterion

In what follows, the stopping criterion for Subset Simulation is described in detail. Let $n_F(l)$ denote the number of failure samples at the l^{th} level, that is,

$$n_F(l) = \sum_{i=1}^n I_F(x_l^{(i)}), \quad (41)$$

where $x_l^{(1)}, \dots, x_l^{(n)} \sim \pi(\cdot|F_l)$. Since F is a rare event, it is very likely that $n_F(l) = 0$ for the first few conditional levels. As l gets larger, however, $n_F(l)$ starts increasing since F_l , which approximates F “from above,” shrinks closer to F . In general, $n_F(l) \geq n_F(l-1)$, since $F \subset F_l \subset F_{l-1}$ and the np closest to F samples among $x_{l-1}^{(1)}, \dots, x_{l-1}^{(n)}$ are present among $x_l^{(1)}, \dots, x_l^{(n)}$. At conditional level l , the failure probability p_F is expressed as a product,

$$p_F = \mathbb{P}(F_1)\mathbb{P}(F_2|F_1) \dots \mathbb{P}(F_l|F_{l-1})\mathbb{P}(F|F_l). \quad (42)$$

Furthermore, the adaptive choice of intermediate critical thresholds $y_1^*, \dots, y_l^*$ guarantees that the first l factors in Eq. 42 approximately equal to p , and, thus,

$$p_F \approx p^l \cdot \mathbb{P}(F|F_l). \quad (43)$$

Since there are exactly $n_F(l)$ failure samples at the l^{th} level, the estimate of the last conditional probability in Eq. 42 which is based on samples $x_l^{(1)}, \dots, x_l^{(n)} \sim \pi(\cdot|F_l)$ is given by

$$\mathbb{P}(F|F_l) \approx \frac{1}{n} \sum_{i=1}^n I_F(x_l^{(i)}) = \frac{n_F(l)}{n}. \quad (44)$$

If $n_F(l)$ is sufficiently large, i.e., the conditional event $(F|F_l)$ is not rare, then the estimate in Eq. 44 is fairly accurate. This leads to the following stopping criterion:

- If $\frac{n_F(l)}{n} \geq p$, i.e., there are at least np failure samples among $x_l^{(1)}, \dots, x_l^{(n)}$, then Subset Simulation stops: the current conditional level l becomes the last level, $L = l$, and the failure probability estimate derived from Eqs. 43 and 44 is

$$p_F \approx \hat{p}_F^{\text{SS}} = p^L \frac{n_F(L)}{n}. \quad (45)$$

- If $\frac{n_F(l)}{n} < p$, i.e., there are less than np failure samples among $x_l^{(1)}, \dots, x_l^{(n)}$, then the algorithm proceeds by defining the next intermediate failure domain $F_{l+1} = \{x : g(x) > y_{l+1}^*\}$, where $y_{l+1}^* = (y_l^{(np)} + y_l^{(np+1)})/2$, and expressing $\mathbb{P}(F|F_l)$ as a product $\mathbb{P}(F|F_l) = \mathbb{P}(F_{l+1}|F_l)\mathbb{P}(F|F_{l+1}) \approx p \cdot \mathbb{P}(F|F_{l+1})$.

The described stopping criterion guarantees that the estimated values of all factors in the factorization $p_F = \mathbb{P}(F_1)\mathbb{P}(F_2|F_1) \dots \mathbb{P}(F_L|F_{L-1})\mathbb{P}(F|F_L)$ are not smaller than p . If p is relatively large ($p = 0.1$ is often used in applications), then it is likely that the estimates $\mathbb{P}(F_1) \approx p$, $\mathbb{P}(F_2|F_1) \approx p, \dots, \mathbb{P}(F_L|F_{L-1}) \approx p$, and $\mathbb{P}(F|F_L) \approx \frac{n_F(L)}{n} (\leq p)$ are accurate even when the sample size n is relatively small. As a result, the SS estimate in Eq. 45 is also accurate in this case. This provides an intuitive explanation as to why Subset Simulation is efficient in estimating small probabilities of rare events. For a detailed discussion of error estimation for the SS method, the reader is referred to Au and Wang (2014).

Implementation Details

In the rest of this section, the implementation details of Subset Simulation are discussed. The SS algorithm has two essential components that affect its efficiency: the parameter p and the set of univariate proposal PDFs $\{q_k\}$, $k = 1, \dots, d$.

Level Probability

The parameter p , called the *level probability* in Au and Wang (2014) and the *conditional failure probability* in Zuev et al. (2012), governs how many intermediate failure domains F_l are needed

to reach the target failure domain F . As it follows from Eq. 45, a small value of p leads to a fewer total number of conditional levels L . But at the same time, it results in a large number of samples n needed at each conditional level l for accurate determination of F_l (i.e., determination of y_l^*) that satisfies $\frac{1}{n} \sum_{i=1}^n I_{F_l}(x_{l-1}^{(i)}) = p$. In the extreme case when $p \leq p_F$, no levels are needed, $L = 0$, and Subset Simulation reduces to the Direct Monte Carlo method. On the other hand, increasing the value of p will mean that fewer samples are needed at each conditional level, but it will increase the total number of levels L . The choice of the level probability p is thus a trade-off between the total number of level L and the number of samples n at each level. In the original paper (Au and Beck 2001a), it has been found that the value $p = 0.1$ yields good efficiency. The latter studies (Au and Wang 2014; Zuev et al. 2012), where the c.o.v. of the SS estimate $\hat{p}_F^{SS}$ has been analyzed, confirmed that $p = 0.1$ is a nearly optimal value of the level probability.

Proposal Distributions

The efficiency and accuracy of Subset Simulation also depends on the set of univariate proposal PDFs $\{q_k\}$, $k = 1, \dots, d$, that are used within the Modified Metropolis algorithm for sampling from the conditional distributions $\pi(\cdot|F_l)$. To see this, note that in contrast to the Monte Carlo samples $x_0^{(1)}, \dots, x_0^{(n)} \sim \pi(\cdot)$ which are i.i.d., the MCMC samples $x_l^{(1)}, \dots, x_l^{(n)} \sim \pi(\cdot|F_l)$ are *not independent* for $l \geq 1$, since the MMA transition rule uses $x_l^{(i)} \sim \pi(\cdot|F_l)$ to generate $x_l^{(i+1)} \sim \pi(\cdot|F_l)$. This means that although these MCMC samples can be used for statistical averaging as if they were i.i.d., the efficiency of the averaging is reduced if compared with the i.i.d. case (Doob 1953). Namely, the more correlated $x_l^{(1)}, \dots, x_l^{(n)}$ are, the slower is the convergence of the estimate $P(F_{l+1}|F_l) \approx \frac{1}{n} \sum_{i=1}^n I_{F_{l+1}}(x_l^{(i)})$, and, therefore, the less efficient it is. The correlation between samples $x_l^{(1)}, \dots, x_l^{(n)}$ is due to proposal PDFs $\{q_k\}$, which govern the generation of the next sample $x_l^{(i+1)}$ from the current one $x_l^{(i)}$. Hence, the choice of $\{q_k\}$ is very important.

It was observed in Au and Beck (2001a) that the efficiency of MMA is not sensitive to the type of the proposal PDFs (Gaussian, uniform, etc.); however, it strongly depends on their *spread* (variance). Both small and large spreads tend to increase the correlation between successive samples. Large spreads may reduce the acceptance rate in Eq. 28, increasing the number of repeated MCMC samples. Small spreads, on the contrary, may lead to a reasonably high acceptance rate, but still produce very correlated samples due to their close proximity. As a rule of thumb, the spread of q_k , $k = 1, \dots, d$, can be taken of the same order as the spread of the corresponding marginal PDF π_k (Au and Wang 2014). For example, if π is given by Eq. 18, so that all marginal PDFs are standard Gaussian, $\pi_k(x) = \phi(x)$, then all proposal PDFs can also be Gaussian with unit variance, $q_k(x/xk) = \phi(x - xk)$. This choice is found to give a balance between efficiency and robustness.

The spread of proposal PDFs can also be chosen adaptively. In Zuev et al. (2012), where the problem of optimal scaling for the Modified Metropolis algorithm was studied in more detail, the following nearly optimal scaling strategy was proposed: at each conditional level, select the spread such that the corresponding acceptance rate in Eq. 28 is between 30 % and 50 %. In general, finding the optimal spread of proposal distributions is problem specific and a highly nontrivial task not only for MMA but also for almost all MCMC algorithms.

Illustrative Examples

To illustrate Subset Simulation and to demonstrate its efficiency in estimating small probabilities of rare failure events, two examples are considered in this section. As it has been discussed in section “[Engineering Reliability Problem](#),” in reliability problems, the dimension d of the input space $\mathcal{X}$ is usually very large. In spite of this, for visualization and educational purposes, a linear reliability problem in two dimensions ($d = 2$) is first considered in section “[Subset Simulation in 2D](#).” A more realistic

high-dimensional example ($d = 10^3$) is considered in the subsequent section “[Subset Simulation in High Dimensions](#).”

Subset Simulation in 2D

Suppose that $d = 2$, i.e., the response variable y depends only on two input variables x_1 and x_2 . Consider a linear performance function

$$g(x_1, x_2) = x_1 + x_2, \quad (46)$$

where x_1 and x_2 are independent standard Gaussian, $x_i \sim N(0, 1)$, $i = 1, 2$. The failure domain F is then a half-plane defined by

$$F = \{(x_1, x_2) : x_1 + x_2 > y^*\}. \quad (47)$$

In this example, the failure probability p_F can be calculated analytically. Indeed, since $x_1 + x_2 \sim N(0, 2)$ and, therefore, $\frac{x_1 + x_2}{\sqrt{2}} \sim N(0, 1)$,

$$\begin{aligned} p_F &= \mathbb{P}(x_1 + x_2 > y^*) = \mathbb{P}\left(\frac{x_1 + x_2}{\sqrt{2}} > \frac{y^*}{\sqrt{2}}\right) \\ &= 1 - \Phi\left(\frac{y^*}{\sqrt{2}}\right), \end{aligned} \quad (48)$$

where Φ is the standard Gaussian CDF. This expression for the failure probability can be used as a check on the SS estimate. Moreover, expressing y^* in terms of p_F ,

$$y^* = \sqrt{2}\Phi^{-1}(1 - p_F), \quad (49)$$

allows to solve the inverse problem, namely, to formulate a linear reliability problem with a given value of the failure probability. Suppose that $p_F = 10^{-10}$ is the target value. Then the corresponding value of the critical threshold is $y^* \approx 9$.

Subset Simulation was used to estimate the failure probability of the rare event in Eq. 47 with $y^* = 9$. The parameters of the algorithm were chosen as follows: the level probability $p = 0.1$, the proposal PDFs $q_k(x/x_k) = \phi(x - x_k)$, and the sample size $n = 10^3$ per each level. This implementation of SS led to $L = 9$ conditional

levels, making the total number of generated samples $N = n + L(n - np) = 9.1 \times 10^3$. The obtained SS estimate is $\hat{p}_F^{SS} = 1.58 \times 10^{-10}$ which is quite close to the true value $p_F = 10^{-10}$. Note that, in this example, it is hopeless to obtain an accurate estimate by the Direct Monte Carlo method since the DMC estimate in Eq. 12 based on $N = 9.1 \times 10^3$ samples is effectively zero: the rare event F is too rare.

Figure 6 shows the samples generated by the SS method. The dashed lines represent the boundaries of intermediate failure domains F_l , $l = 1, \dots, L = 9$. The solid line is the boundary of the target failure domain F . This illustrates how Subset Simulation pushes Monte Carlo samples (red) toward the failure region.

Subset Simulation in High Dimensions

It is straightforward to generalize the low-dimensional example considered in the previous section to high dimensions. Consider a linear performance function

$$g(x) = \sum_{i=1}^d x_i, \quad (50)$$

where $x_1, \dots, x_d$ are i.i.d. standard Gaussian. The failure domain is then a half-space defined by

$$F = \left\{x : \sum_{i=1}^d x_i > y^*\right\}. \quad (51)$$

In this example, $d = 10^3$ is considered; hence the input space $\mathbf{x} = \mathbb{R}^d$ is indeed high dimensional. As before, the failure probability can be calculated analytically:

$$\begin{aligned} p_F &= \mathbb{P}\left(\sum_{i=1}^d x_i > y^*\right) = \mathbb{P}\left(\frac{\sum_{i=1}^d x_i}{\sqrt{d}} > \frac{y^*}{\sqrt{d}}\right) \\ &= 1 - \Phi\left(\frac{y^*}{\sqrt{d}}\right). \end{aligned} \quad (52)$$

This expression will be used as a check on the SS estimate.

First, consider the following range of values for the critical threshold, $y^* \in [0, 200]$.

**Subset Simulation
Method for Rare Event
Estimation:
An Introduction,**

Fig. 6 Samples generated by Subset Simulation: red samples are Monte Carlo samples generated at the 0th unconditional level, purple samples are MCMC sample generated at the 1st conditional level, etc. The dashed lines represent the boundaries of intermediate failure domains F_l , $l = 1, \dots, L = 9$. The solid line is the boundary of the target failure domain F [Example 6.1]

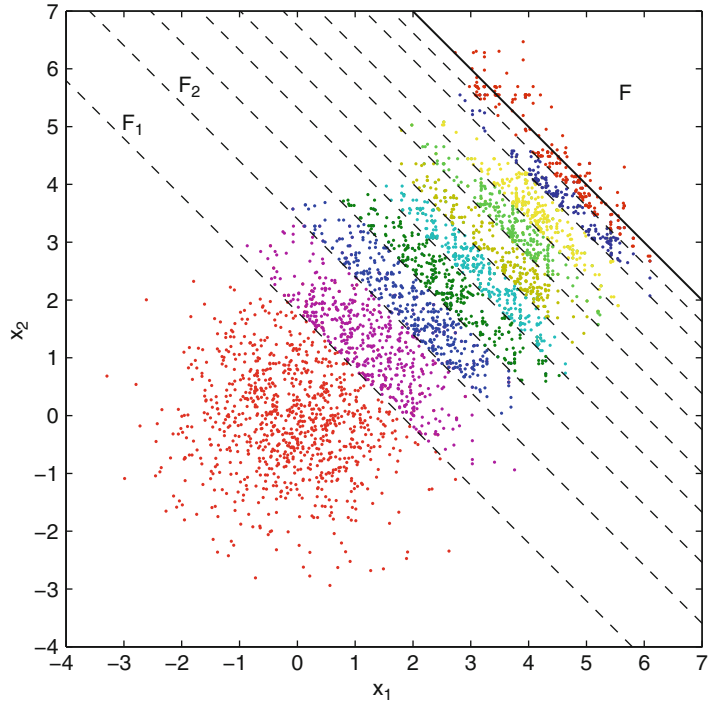
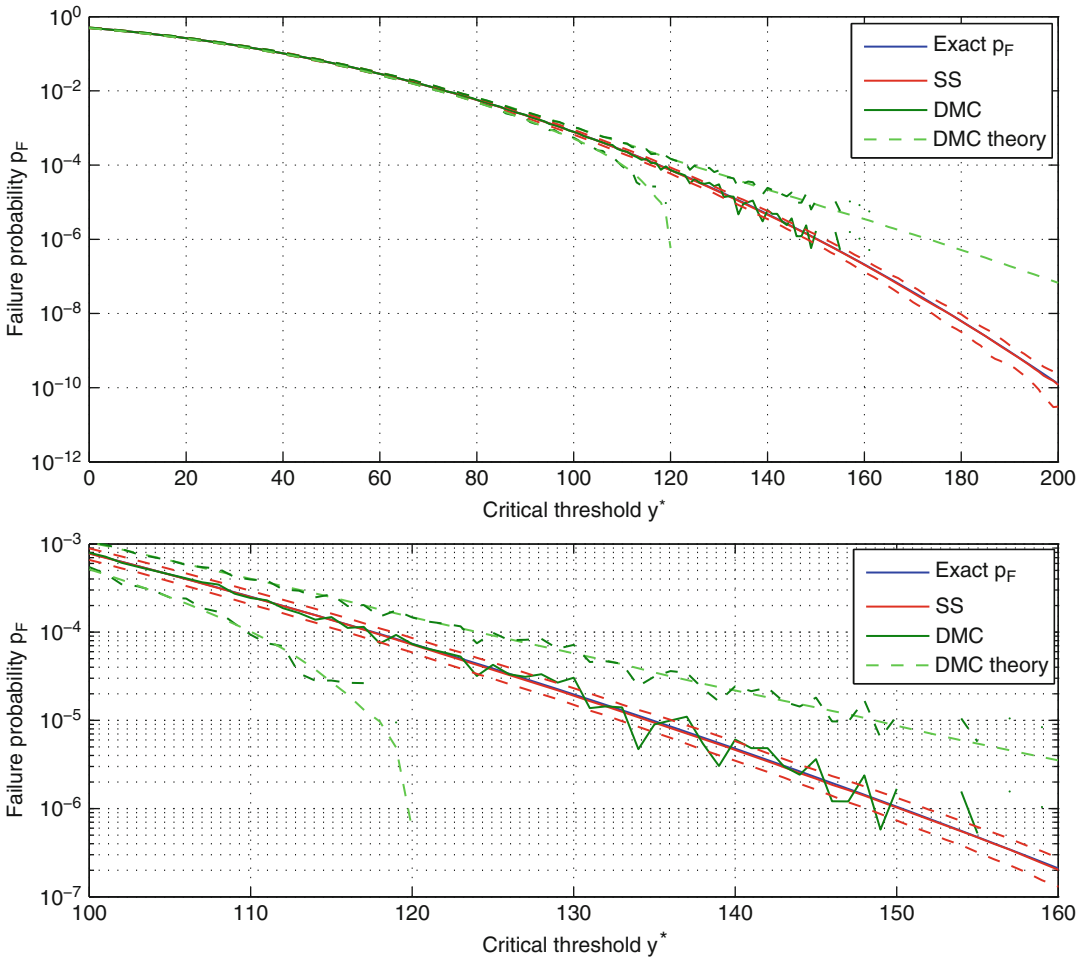


Figure 7 plots p_F versus y^* . The solid red curve corresponds to the sample mean of the SS estimates $\hat{p}_F^{SS}$ which is based on 100 independent runs of Subset Simulation. The two dashed red curves correspond to the sample mean $\pm$ one sample standard deviation. The SS parameters were set as follows: the level probability $p = 0.1$, the proposal PDFs $q_k(x/x_k) = \phi(x - x_k)$, and the sample size $n = 3 \times 10^3$ per each level. The solid blue curve (which almost coincides with the solid red curve) corresponds to the true values of p_F computed from Eq. 52. The dark green curves correspond to Direct Monte Carlo: the solid curve is the sample mean (based on 100 independent runs) of the DMC estimates $\hat{p}_F^{DMC}$ in Eq. 12, and the two dashed curves are the sample mean $\pm$ one sample standard deviation. The total number of samples N used in DMC equals to the average (based on 100 runs) total number of samples used in SS. Finally, the dashed light green curves show the theoretical performance of Direct Monte Carlo, namely, they correspond to the true value of p_F (52) $\pm$ one theoretical standard deviation obtained from Eq. 14. The bottom panel of Fig. 7

shows the zoomed-in region that corresponds to the values $y^* \in [100, 160]$ of the critical threshold. Note that for relatively large values of the failure probability, $p_F < 10^{-3}$, both DMC and SS produce accurate estimates of p_F . For smaller values however, $p_F < 10^{-5}$, the DMC estimate starts to degenerate, while SS still accurately estimates p_F . This can be seen especially well in the bottom panel of the figure.

The performances of Subset Simulation and Direct Monte Carlo can be also compared in terms of the coefficient of variation of the estimates $\hat{p}_F^{SS}$ and $\hat{p}_F^{DMC}$. This comparison is shown in Fig. 8. The red and dark green curves represent the sample c.o.v. for SS and DMC, respectively. The light green curve is the theoretical c.o.v. of $\hat{p}_F^{DMC}$ given by Eq. 15. When the critical threshold is relatively small $y^* < 60$, the performances of SS and DMC are comparable. As y^* gets large, the c.o.v. of $\hat{p}_F^{DMC}$ starts to grow much faster than that of $\hat{p}_F^{SS}$. In other words, SS starts to outperform DMC, and the larger the y^* , i.e., the more rare the failure event, the more significant the outperformance is.

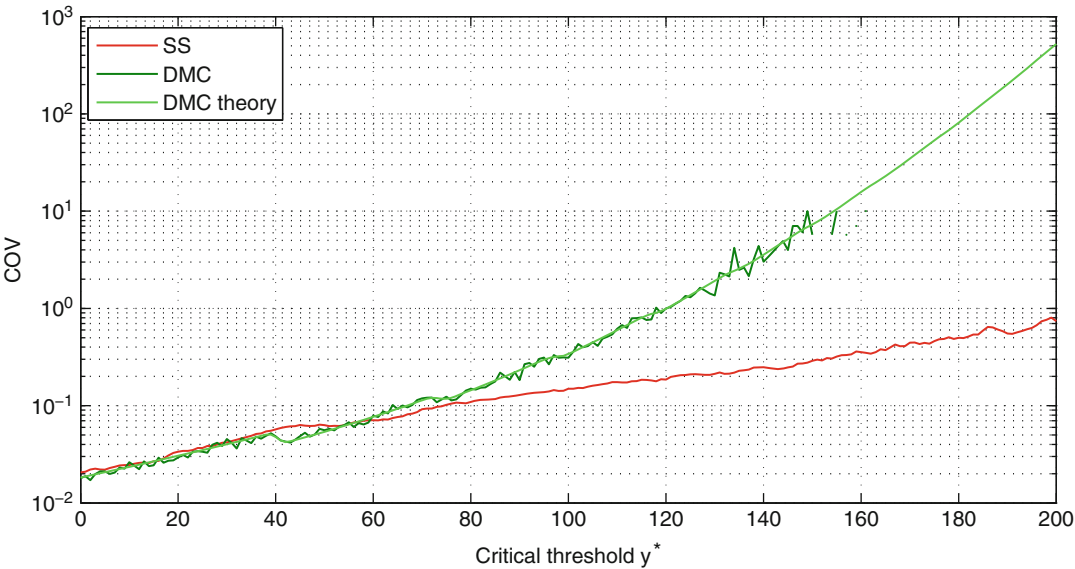


Subset Simulation Method for Rare Event Estimation: An Introduction, Fig. 7 Failure probability p_F versus the critical threshold y^* [Example 6.2]

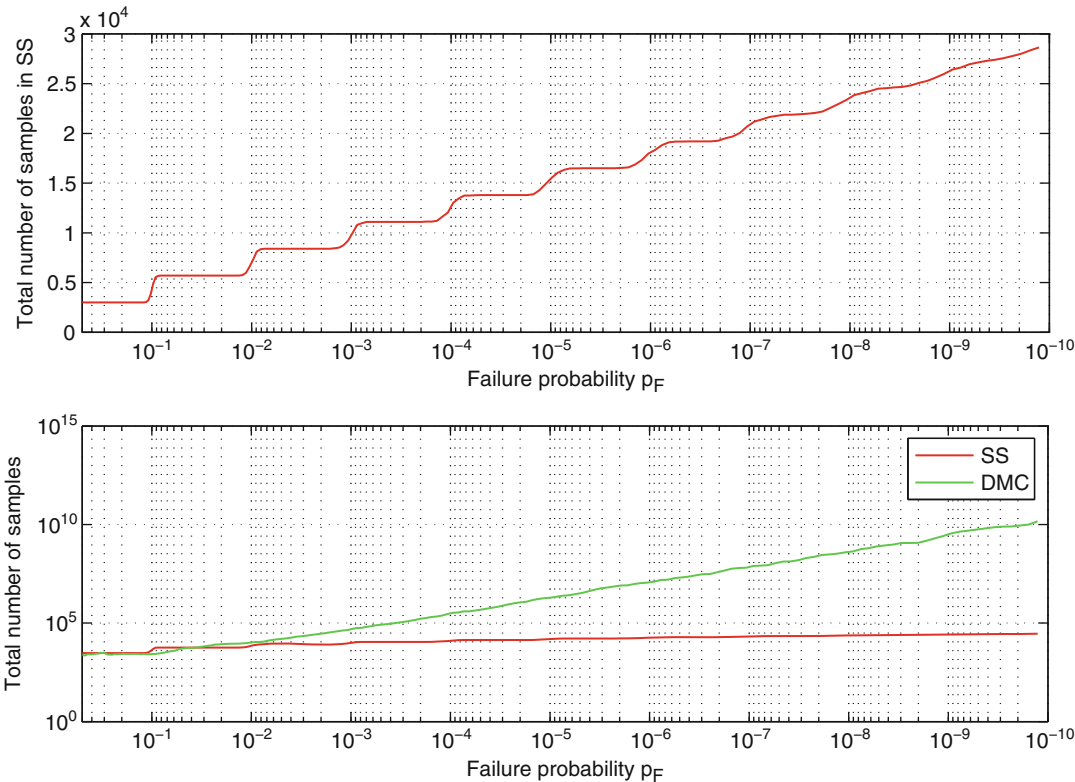
The average total number of samples used in Subset Simulation versus the corresponding values of failure probability is shown in the top panel of Fig. 9. The staircase nature of the plot is due to the fact that every time p_F crosses the value p^k by decreasing from $p^k + \epsilon$ to $p^k - \epsilon$, an additional conditional level is required. In this example, $p = 0.1$ is used, that is why the jumps occur at $p_F = 10^{-k}$, $k = 1, 2, \dots$. The jumps are more pronounced for larger values of p_F , where the SS estimate is more accurate. For smaller values of p_F , where the SS estimate is less accurate, the jumps are more smoothed out by averaging over independent runs.

In Fig. 8, where the c.o.v.'s of SS and DMC are compared, the total numbers of samples (computational efforts) used in the two methods are the same. The natural question is then the following: by how much should the total number of samples N used in DMC be increased to achieve the same c.o.v as in SS (so that the green curve in Fig. 8 coincides with the red curve)? The answer is given in the bottom panel of Fig. 9. For example, if $p_F = 10^{-10}$, then $N = 1010$, while the computational effort of SS is less than 105 samples.

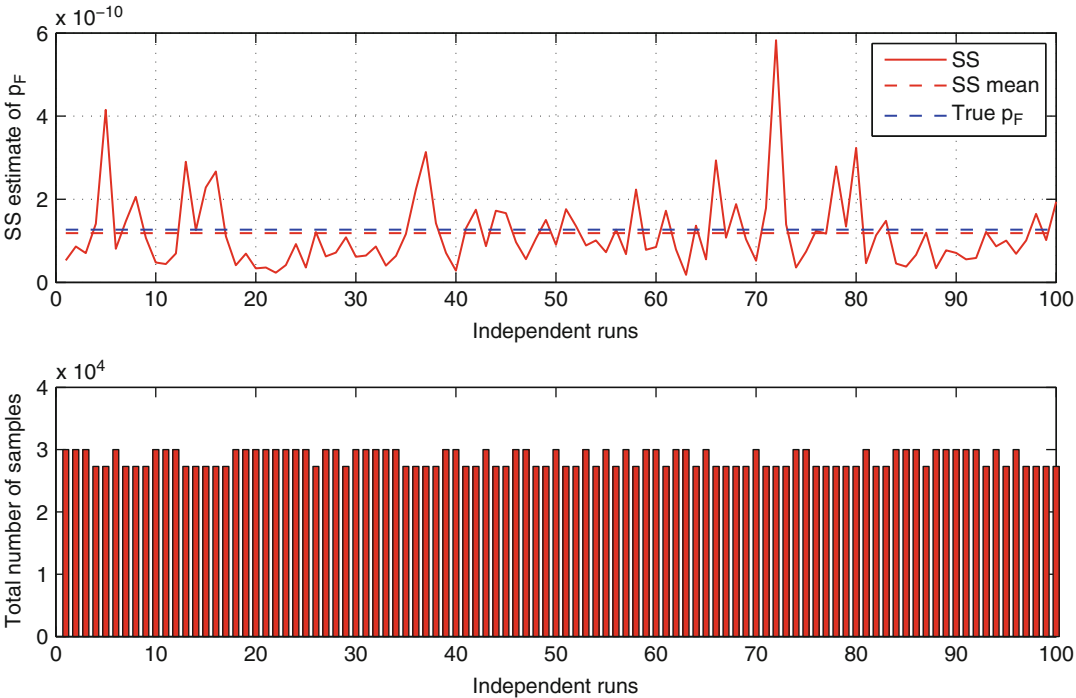
Simulation results presented in Figs. 7, 8, and 9 clearly indicate that (a) Subset Simulation



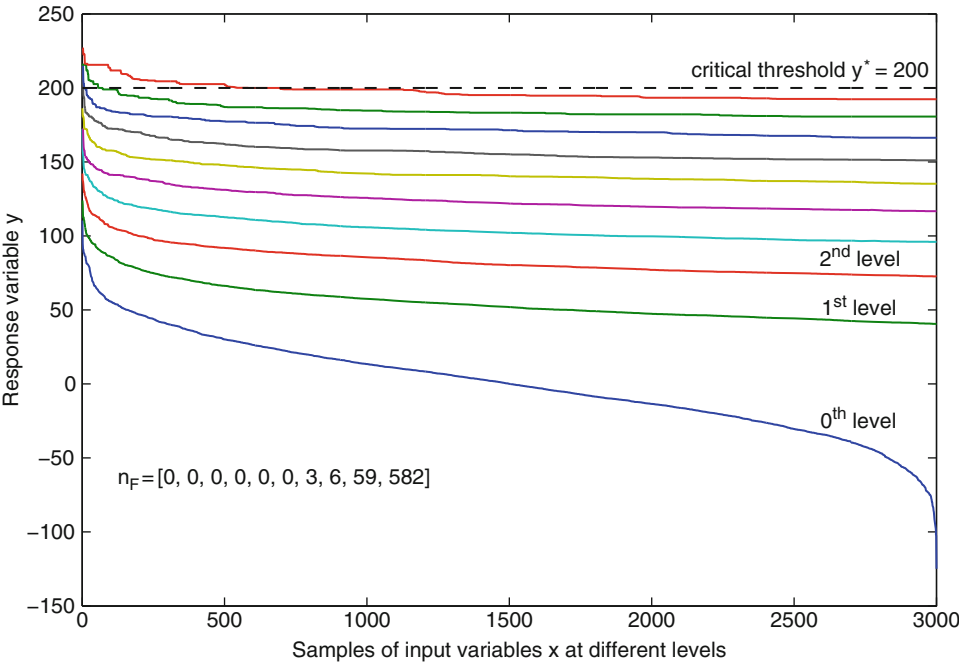
Subset Simulation Method for Rare Event Estimation: An Introduction, Fig. 8 C.o.v versus the critical threshold [Example 6.2]



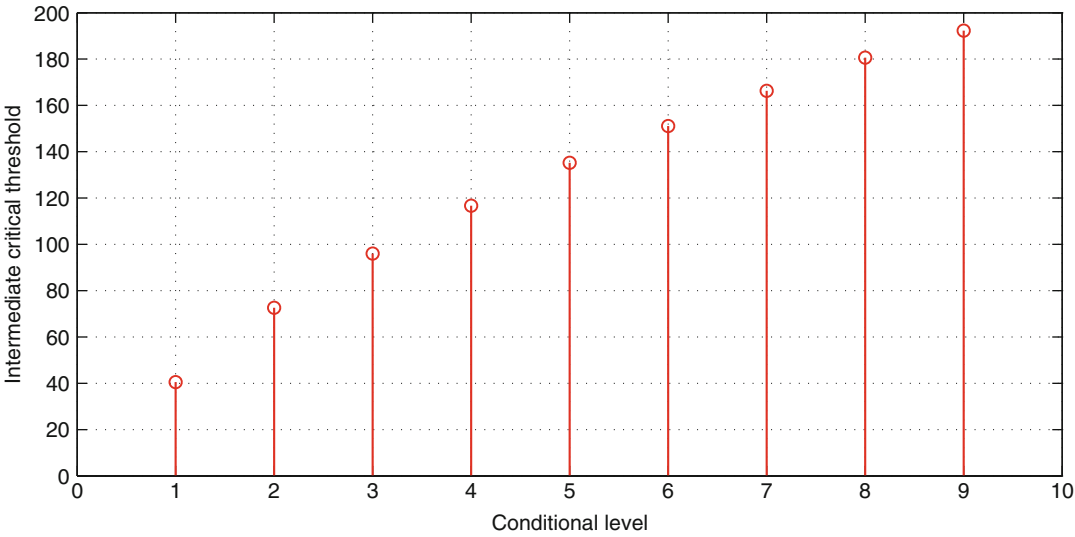
Subset Simulation Method for Rare Event Estimation: An Introduction, Fig. 9 Total number of samples versus the failure probability [Example 6.2]



Subset Simulation Method for Rare Event Estimation: An Introduction, Fig. 10 Performance of Subset Simulation for 100 independent runs. The critical threshold is $y^* = 200$, and the corresponding true value of the failure probability is $p_F = 1.27 \times 10^{-10}$ [Example 6.2]



Subset Simulation Method for Rare Event Estimation: An Introduction, Fig. 11 System responses $y_1^{(1)} \geq \dots \geq y_1^{(n)}$, $n = 3 \times 10^3$, for all levels, $l = 0, \dots, L = 9$, for a fixed simulation run [Example 6.2]



Subset Simulation Method for Rare Event Estimation: An Introduction, Fig. 12 Intermediate critical thresholds $y_1^*, \dots, y_L^*$, $L = 9$, at different conditional levels in a fixed simulation run [Example 6.2]

produces a relatively accurate estimate of the failure probability and (b) Subset Simulation drastically outperforms Direct Monte Carlo when estimating probabilities of rare events.

Let us now focus on a specific value of the critical threshold, $y^* = 200$, which corresponds to a very rare failure event in Eq. 51 with probability $p_F = 1.27 \times 10^{-10}$. Figure 10 demonstrates the performance of Subset Simulation for 100 independent runs. The top panel shows the obtained SS estimate $\hat{p}_F^{SS}$ for each run. Although $\hat{p}_F^{SS}$ varies significantly (its c.o.v. is $\delta(\hat{p}_F^{SS}) = 0.74$), its mean value $\overline{\hat{p}_F^{SS}} = 1.18 \times 10^{-10}$ (dashed red line) is close to the true value of the failure probability (dashed blue line). The bottom panel shows the total number of samples used in SS in each run. It is needless to say that the DMC estimate based on $N \sim 3 \times 10^4$ samples would almost certainly be zero.

Figure 11 shows the system responses $y_l^{(1)} \geq \dots \geq y_l^{(n)}$, $n = 3 \times 10^{-10}$ for all levels, $l = 0, \dots, L = 9$, for a fixed simulation run. As expected, for the first few levels (six levels in this case), the number of failure samples $n_F(l)$, i.e., samples $x_l^{(i)}$ with $y_l^{(i)} = g(x_l^{(i)}) > y^*$, is zero. As Subset Simulation starts pushing the samples toward the failure domain, $n_F(l)$ starts increasing

with $n_F(6) = 3$, $n_F(7) = 6$, $n_F(8) = 59$, and, finally, $n_F(9) = 582$, after which the algorithm stopped since $n_F(9)/n = 0.194$ which is large than $p = 0.1$. Finally, Fig. 12 plots the intermediate (relaxed) critical thresholds $y_1^*, \dots, y_L^*$ at different levels obtained in a fixed simulation run.

MATLAB Code

This section contains the MATLAB code for the examples considered in section “[Illustrative Examples](#).” For educational purposes, the code was written as readable as possible with numerous comments. As a result of this approach, the efficiency of the code was unavoidably scarified.

```
% Subset Simulation for Liner
Reliability Problem
% Performance function: g(x)=x1+...+xd
% Input variables x1,...,xd are i.i.d. N
(0,1)
% Written by K.M. Zuev, Institute of Risk
& Uncertainty, Uni of Liverpool
clear;
d=1000; % dimension of the
input space
```

(continued)

YF=200;	% critical threshold (failure $\leq \Rightarrow g(x) > YF$)
pF=1-normcdf(YF/sqrt(d));	% true value of the failure probability
n=3000;	% number of samples per level
p=0.1;	% level probability
nc=n*p;	% number of Markov chains
ns=(1-p)/p;	% number of states in each chain
L=0;	% current (unconditional) level
x=randn(d,n);	% Monte Carlo samples
nF=0;	% number of failure samples
for i=1:n	
y(i)=sum(x(:,i));	% system response $y=g(x)$
if y(i)>YF	% $y(i) > YF \Leftrightarrow x(:,i)$ is a failure sample
nF=nF+1;	
end	
end	
while nF(L+1)/n<p	% stopping criterion
L=L+1;	% next conditional level is needed
[y(L,:),ind]=sort(y(L,:), 'descend');	% renumbered responses
x(:, :, L)=x(:, ind(:, L));	% renumbered samples
Y(L)=(y(L,nc)+y(L,nc+1))/2;	% L th intermediate threshold
z(:, :, 1)=x(:, 1:nc, L);	% Markov chain "seeds"

% Modified Metropolis algorithm for sampling from $\pi(x / FL)$

```

for j=1:nc
    for m=1:ns
        % Step 1:
        for k=1:d
            a=z(k,j,m)+randn; % Step 1(a)
            r=min(1, normpdf(a)/normpdf(z(k,j,m))); % Step 1(b)
        end
        % Step 1(c):
        if rand<r

```

(continued)

```

q(k)=a;
else
    q(k)=z(k,j,m);
end
end
% Step 2:
if sum(q)>Y(L) % q belongs to FL
    z(:,j,m+1)=q;
else
    z(:,j,m+1)=z(:,j,m);
end
end
end
for j=1:nc
    for m=1:ns+1
        x(:, (j-1)*(ns+1)+m, L+1)=z(:,j,m); %
        samples from  $\pi(x / FL)$ 
    end
end
clear z;
nF(L+1)=0;
for i=1:n
    y(L+1,i)=sum(x(:,i,L+1)); % system response
    y=g(x)
    if y(L+1,i)>YF % then  $x(:,i,L+1)$  is a failure sample
        nF(L+1)=nF(L+1)+1; % number of failure samples at level L+1
    end
end
end
end
pF SS=p^(L)*nF(L+1)/n; % SS estimate
N=N+n*(1-p)*(L); % total number of samples

```

Summary

In this entry, a detailed exposition of Subset Simulation, an advanced stochastic simulation method for estimation of small probabilities of rare events, is provided at an introductory level. A simple step-by-step derivation of Subset Simulation is given, and important implementation details are discussed. The method is illustrated with a few intuitive examples.

After the original paper (Au and Beck 2001a) was published, various modifications of SS were proposed: SS with splitting (Ching et al. 2005a), hybrid SS (Ching et al. 2005b), and two-stage SS (Katafygiotis and Cheung 2005), to name but

a few. It is important to highlight, however, that none of these modifications offers a drastic improvement over the original algorithm. A Bayesian analog of SS was developed in Zuev et al. (2012). For further reading on Subset Simulation and its applications, a fundamental and very accessible monograph (Au and Wang 2014) is strongly recommended, where the method is presented from the CCDF (complementary cumulative distribution function) perspective and where the error estimation is discussed in detail.

Also, it is important to emphasize that Subset Simulation provides an efficient solution for general reliability problems without using any specific information about the dynamic system other than an input–output model. This independence of a system’s inherent properties makes Subset Simulation potentially useful for applications in different areas of science and engineering.

As a final remark, it is a pleasure to thank Professor Siu-Kui Au whose comments on the first draft of the entry were very helpful; Professor James Beck, who generously shared his knowledge of and experience with Subset Simulation and made important comments on the pre-final draft of the entry; and Professor Francis Bonahon for his general support and for creating a nice atmosphere at the Department of Mathematics of the University of Southern California, where the author started this work.

Cross-References

- [Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach](#)
- [Reliability Estimation and Analysis](#)
- [Response Variability and Reliability of Structures](#)
- [Seismic Reliability Assessment, Alternative Methods for](#)
- [Seismic Risk Assessment, Cascading Effects](#)
- [Site Response for Seismic Hazard Assessment](#)
- [Structural Reliability Estimation for Seismic Loading](#)
- [Structural Seismic Reliability Analysis](#)

References

- Au SK, Beck JL (2001a) Estimation of small failure probabilities in high dimensions by subset simulation. *Probabilist Eng Mech* 16(4):263–277
- Au SK, Beck JL (2001b) First-excursion probabilities for linear systems by very efficient importance sampling. *Probabilist Eng Mech* 16(3):193–207
- Au SK, Wang Y (2014) *Engineering risk assessment and design with subset simulation*. Wiley, Singapore, To appear
- Bucher C (1990) A fast and efficient response surface approach for structural reliability problem. *Struct Saf* 7:57–66
- Ching J, Au SK, Beck JL (2005a) Reliability estimation of dynamical systems subject to stochastic excitation using subset simulation with splitting. *Comput Method Appl Mech Eng* 194(12–16):1557–1579
- Ching J, Beck JL, Au SK (2005b) Hybrid subset simulation method for reliability estimation of dynamical systems subject to stochastic excitation. *Probabilist Eng Mech* 20(3):199–214
- Ditlevsen O, Madsen HO (1996) *Structural reliability methods*. Wiley, Chichester
- Doob JL (1953) *Stochastic processes*. Wiley, New York
- Engelund S, Rackwitz R (1993) A benchmark study on importance sampling techniques in structural reliability. *Struct Saf* 12(4):255–276
- Faravelli L (1989) Response-surface approach for reliability analysis. *J Eng Mech* 115:2763–2781
- Grooteman F (2008) Adaptive radial-based importance sampling method for structural reliability. *Struct Saf* 30(6):533–542
- Hurtado JE (2004) *Structural reliability. Statistical learning perspectives*. Springer, Heidelberg
- Hurtado JE, Alvarez DA (2003) A classification approach for reliability analysis with stochastic finite element modeling. *J Struct Eng* 129(8):1141–1149
- Katafygiotis LS, Cheung SH (2005) A two-stage subset simulation-based approach for calculating the reliability of inelastic structural systems subjected to Gaussian random excitations. *Comput Method Appl Mech Eng* 194(12–16):1581–1595
- Katafygiotis LS, Zuev KM (2007) Estimation of small failure probabilities in high dimensions by adaptive linked importance sampling. In: M. Papadrakakis, D.C. Charmpis, N.D. Lagaros, Y. Tsompanakis (Eds.), *ECCOMAS thematic conference on computational methods in structural dynamics and earthquake engineering (COMPDYN)*, Rethymno, Crete, Greece, June 2007
- Katafygiotis LS, Moan T, Cheung SH (2007) Auxiliary domain method for solving multi-objective dynamic reliability problems for nonlinear structures. *Struct Eng Mech* 25(3):347–363
- Koutsourelakis PS, Pradlwarter HJ, Schuëller GI (2004) Reliability of structures in high dimensions, part I: algorithms and applications. *Probabilist Eng Mech* 19(4):409–417

- Liu JS (2001) Monte Carlo strategies is scientific computing. Springer, New York
- Madsen HO, Krenk S, Lind NC (2006) Methods of structural safety. Dover, Mineola
- Melchers R (1999) Structural reliability analysis and prediction. Wiley, Chichester
- Metropolis N (1987) The beginning of the Monte Carlo method. *Los Alamos Science* 15:125–130
- Metropolis N, Ulam S (1949) The Monte Carlo method. *J Am Stat Assoc* 44:335–341
- Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) Equation of state calculations by fast computing machines. *J Chem Phys* 21(6): 1087–1092
- Nataf A (1962) Détermination des distributions de probabilité dont les marges sont données. *C R Acad Sci* 225:42–43
- Papadrakakis M, Papadopoulos V, Lagaros ND (1996) Structural reliability analysis of elastic–plastic structures using neural networks and Monte Carlo simulation. *Comput Method Appl Mech Eng* 136:145–163
- Robert CP, Casella G (2004) Monte Carlo statistical methods. Springer, New York
- Rosenblatt M (1952) Remarks on a multivariate transformation. *Ann Math Stat* 23:470–472
- Schuëller GI, Bucher CG, Bourgund U, Ouypornprasert W (1989) On efficient computational schemes to calculate structural failure probabilities. *Probabilist Eng Mech* 4(1):10–18
- Zuev KM, Katafygiotis LS (2011) Horseracing simulation algorithm for evaluation of small failure probabilities. *Probabilist Eng Mech* 26(2):157–164
- Zuev KM, Beck JL, Au SK, Katafygiotis LS (2012) Bayesian postprocessor and other enhancements of subset simulation for estimating failure probabilities in high dimensions. *Comput Struct* 92–93: 283–296

Substructuring Methods for Finite Element Analysis

S. Abhinav, Debraj Ghosh and C. S. Manohar
Department of Civil Engineering, Indian Institute of Science, Bangalore, Karnataka, India

Synonyms

Component mode synthesis; Domain decomposition; Hybrid simulations

Introduction

The motivations for employing substructuring in finite element modeling vary from reduction of computational time, modal synthesis using substructure modes, combining experimental and numerical modeling approaches, equitable sharing of resources in parallel computing environment, and treatment of global/local nonlinearities. The details of methods and tools accordingly also vary. An overview of related issues is presented in this entry.

Problems of computational structural mechanics of realistic systems involve inversion and eigenanalysis of large-size matrices and solutions of a large number of coupled ordinary differential equations or algebraic equations. These are computationally demanding tasks, and development of methods to reduce the computational efforts remains relevant notwithstanding advances in computational hardware. This is particularly true in problems of uncertainty quantification, reliability analysis, structural optimization, modeling of actively controlled systems, local/global sensitivity analysis, and problems of damage detection. Substructuring methods primarily serve to achieve reduction in computational effort in these types of problems. Here, the given structure is divided into a set of subsystems, each component is modeled separately, and the behavior of the built-up structure is inferred by synthesis of component behavior. The decomposition and synthesis steps are designed to achieve reduction in sizes of matrices to be handled and (or) reduction in time required to integrate equations of motion. In the context of computing using distributed memory multiprocessor computers, substructuring is used to distribute workload equally among all the processors. On a different note, the idea of substructuring is also attractive in experimental studies. Here again, the structure to be studied is divided into a set of subsystems, and one, more, or all of the subsystems can be studied experimentally in their uncoupled states, and a model for behavior of built-up structure is synthesized therefrom. In such studies, where both

computational and experimental studies are combined, the substructuring methods are termed as hybrid simulation methods. Furthermore, in treatment of transient dynamical problems, if the coupling between numerical substructures and experimental substructures is achieved in real time, then one gets real-time substructuring methods. Apart from computational advantages, the substructuring methods afford other benefits too: (a) structures made up of technologically diverse components having notably different dynamical characteristics can be separately studied and developed by different teams, (b) better insights can be gained on global behavior of built-up structures in terms of local behavior of components, and (c) a combination of computational and experimental tools can be brought to bear on study of large complex systems. The literature on substructuring in structural mechanics is vast, and comprehensive overviews can be found in the works of Hurty (1965), Craig (1995), Maia and de Silva (1997), Ewins (2000), Williams and Blakeborough (2001), de Klerk et al. (2008), Bursi and Wagg (2008), and Saouma and Sivaselvan (2008). This entry details a few select set of tools for substructuring and briefly touches upon a few other issues.

Problem Statement

The motivations for substructuring in finite element modeling vary and so do the details of the methods and criteria for assessing their success. The following is a list of questions which may be conceived in this context:

- (a) Given modal characteristics (natural frequencies, mode shapes, damping ratios, and participation factors), set of frequency response functions, or structural matrices and forcing vectors, for a set of N_S substructures in their uncoupled state, how to synthesize the dynamical characteristics and response of the built-up system? How to deal with incompleteness in spatial, modal, and frequency domains? How to best configure the substructures?
- (b) How to deal with presence of nonlinearities in one or more of the subsystems?
- (c) Given an N degree of freedom dynamical system, how to partition the degrees of freedom into N_S subsets so that efficient schemes can be evolved to integrate the equation of motion of the built-up system?
- (d) If some of the subsystems (linear or nonlinear) are studied experimentally and the remaining numerically, how to interface the subsystem responses? How to carry out this in real time for transient dynamic response analysis?
- (e) Given the response of a built-up system, how to infer behavior of component subsystems? (This is known as the problem of inverse substructuring.)
- (f) Given a computer with multiple processors (generally distributed memory), how to solve the governing equation by decomposing the domain followed by distributing the workload among the processors so that the execution cost reduces significantly? How to best maximize the usage of computational resources and use simultaneously all available processors equally well?
- (g) How to divide a large structure into N_S subsets without a corresponding increase in the number of interface coordinates involved in the component mode synthesis step?

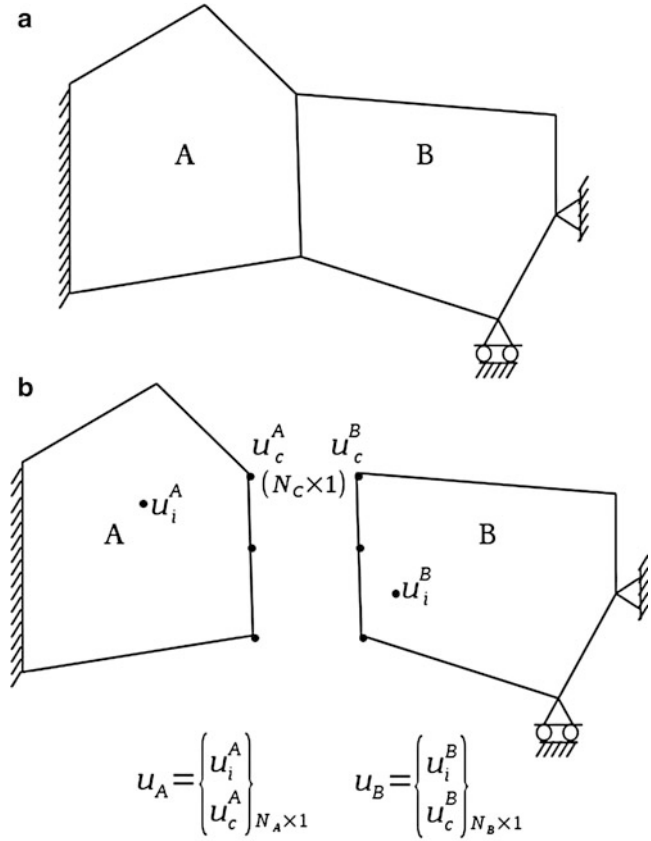
In the same vein, one could pose several other questions pertaining to uncertainty propagation, system identification, vibration energy flow modeling, and treatment of multiphysics problems (such as fluid–structure interactions, soil–structure interactions, primary–secondary structure interactions, etc.). The discussion in the following sections will focus on a few of these issues.

Fixed-Interface Modal Synthesis

The method is illustrated by considering the system shown in Fig. 1a (Maia and de Silva 1997). The given system is divided into two substructures labeled as A and B, and the two

Substructuring Methods for Finite Element Analysis, Fig. 1

(a) Built-up system with two substructures; (b) two substructures in uncoupled states



substructures in their uncoupled state are considered (Fig. 1b). The substructure degrees of freedom (dofs) are partitioned as shown into interior dofs (subscript i) and coupling dofs (subscript c). The equation for substructure A is written in the partitioned form as

$$\begin{bmatrix} M_{ii}^A & M_{ic}^A \\ M_{ci}^A & M_{cc}^A \end{bmatrix} \begin{Bmatrix} \ddot{u}_i^A \\ \ddot{u}_c^A \end{Bmatrix} + \begin{bmatrix} K_{ii}^A & K_{ic}^A \\ K_{ci}^A & K_{cc}^A \end{bmatrix} \begin{Bmatrix} u_i^A \\ u_c^A \end{Bmatrix} = \begin{Bmatrix} 0_i^A \\ f_c^A \end{Bmatrix} \quad (1)$$

Here $f_c^A(t)$ is the vector of interaction forces between A and B. Two separate analyses on this system are performed: first by assuming $u_c^A(t) = 0$ (fixed interface) and the second by taking $u_c^A(t) \neq 0$. In the first case, the governing equation is obtained as $M_{ii}^A \ddot{u}_i^A + K_{ii}^A u_i^A = 0_i^A$, and this equation is analyzed to determine the natural frequencies ω_r^A and modal vectors $\phi_r^A, r = 1, 2, \dots, n_i^A$. The system response is

further represented using a k -term modal expansion as $u_{i(\text{Fix})}^A(t) = [\Phi_{ik_A}^A] \{p_{k_A}^A(t)\}$; the subscript k_A is used here to denote that the expansion has been truncated at k_A modes. In the second analysis, it is assumed that the interior and coupling dofs are related through conditions valid only under static conditions. This leads to the solution $u_{i(\text{Free})}^A(t) = -[K_{ii}^A]^{-1} K_{ic}^A u_c^A(t)$. The solution vector in Eq. 1 is now represented as

$$\begin{Bmatrix} u_i^A \\ u_c^A \end{Bmatrix} = \begin{bmatrix} \Phi_{ik}^A & -[K_{ii}^A]^{-1} K_{ic}^A \\ 0 & I \end{bmatrix} \begin{Bmatrix} p_{k_A}^A(t) \\ u_c^A(t) \end{Bmatrix} = \Psi_k^A \begin{Bmatrix} p_{k_A}^A(t) \\ u_c^A(t) \end{Bmatrix} \quad (2)$$

Here I is the identity matrix. Substituting this in Eq. 1, premultiplying by $[\Psi_{k_A}^A]^T$, and simplifying, one gets equation of the form

$$\begin{aligned}
& \begin{bmatrix} \tilde{M}_{kk}^A & \tilde{M}_{kc}^A \\ \tilde{M}_{ck}^A & \tilde{M}_{cc}^A \end{bmatrix} \begin{Bmatrix} \ddot{p}_{k_A}^A(t) \\ \ddot{u}_c^A(t) \end{Bmatrix} + \begin{bmatrix} \tilde{K}_{kk}^A & \tilde{K}_{kc}^A \\ \tilde{K}_{ck}^A & \tilde{K}_{cc}^A \end{bmatrix} \begin{Bmatrix} p_{k_A}^A(t) \\ u_c^A(t) \end{Bmatrix} = [\Psi_{k_A}^A]^t \begin{Bmatrix} 0_i^A \\ f_c^A \end{Bmatrix} \\
& \text{with} \\
& \begin{bmatrix} \tilde{M}_{kk}^A & \tilde{M}_{kc}^A \\ \tilde{M}_{ck}^A & \tilde{M}_{cc}^A \end{bmatrix} = \begin{bmatrix} I & (\Phi_{ik}^A)^t (M_{ii}^A \alpha^A + M_{ic}^A) \\ (\alpha^A)^t M_{ii}^A \Phi_{ik}^A & (\alpha^A)^t (M_{ii}^A \alpha^A + M_{ic}^A) + (M_{ci}^A \alpha^A + M_{cc}^A) \end{bmatrix} \\
& \begin{bmatrix} \tilde{K}_{kk}^A & \tilde{K}_{kc}^A \\ \tilde{K}_{ck}^A & \tilde{K}_{cc}^A \end{bmatrix} = \begin{bmatrix} \Lambda^A & (\Phi_{ik}^A)^t (K_{ii}^A \alpha^A + K_{ic}^A) \\ (\alpha^A)^t K_{ii}^A \Phi_{ik}^A & (\alpha^A)^t (K_{ii}^A \alpha^A + K_{ic}^A) + (K_{ci}^A \alpha^A + K_{cc}^A) \end{bmatrix} \\
& \Lambda^A = \text{Diag}[(\omega_r^A)^2]; \alpha^A = -[K_{ii}^A]^{-1} K_{ic}^A
\end{aligned} \tag{3}$$

A similar analysis on substructure B leads to the equation

$$\begin{aligned}
& \begin{bmatrix} \tilde{M}_{kk}^B & \tilde{M}_{kc}^B \\ \tilde{M}_{ck}^B & \tilde{M}_{cc}^B \end{bmatrix} \begin{Bmatrix} \ddot{p}_{k_B}^B(t) \\ \ddot{u}_c^B(t) \end{Bmatrix} + \begin{bmatrix} \tilde{K}_{kk}^B & \tilde{K}_{kc}^B \\ \tilde{K}_{ck}^B & \tilde{K}_{cc}^B \end{bmatrix} \begin{Bmatrix} p_{k_B}^B(t) \\ u_c^B(t) \end{Bmatrix} \\
& = [\Psi_{k_B}^B]^t \begin{Bmatrix} 0_i^B \\ f_c^B \end{Bmatrix}
\end{aligned} \tag{4}$$

Returning to the built-up system in Fig. 1a and by imposing the conditions $u_c^A = u_c^B$ and $f_c^A + f_c^B = 0$ at the interfaces, one gets the governing equations for the built-up system as

$$\begin{aligned}
& \begin{bmatrix} I_{kk}^A & 0 & \tilde{M}_{kc}^A \\ 0 & I_{kk}^B & \tilde{M}_{kc}^B \\ \tilde{M}_{ck}^A & \tilde{M}_{ck}^B & \tilde{M}_{cc}^A + \tilde{M}_{cc}^B \end{bmatrix} \begin{Bmatrix} \ddot{p}_k^A(t) \\ \ddot{p}_k^B(t) \\ \ddot{u}_c(t) \end{Bmatrix} \\
& + \begin{bmatrix} (\omega_r^A)^2 & 0 & 0 \\ 0 & (\omega_r^B)^2 & 0 \\ 0 & 0 & \tilde{K}_{cc}^A + \tilde{K}_{cc}^B \end{bmatrix} \begin{Bmatrix} p_k^A(t) \\ p_k^B(t) \\ u_c(t) \end{Bmatrix} = 0
\end{aligned} \tag{5}$$

The dynamic characteristics for the built-up system can now be determined using this model. The total number of dofs here is equal to the total number of substructure modes included in representing the fixed-interface responses for

systems A and B plus the number of interface dofs. This number is expected to be substantially less than the dof that one would get if the built-up system were to be analyzed without substructuring. This method of substructuring is also known as component mode synthesis and is a widely studied method in the existing literature. Various generalizations to include more than two substructures, presence of damping, substructure coupling through flexible elements, and response to external excitations have been explored.

Free-Interface Modal Synthesis

The fixed-interface method discussed in the preceding section requires the knowledge of subsystem structural matrices, and for this reason, they are not suited if one or more of the subsystems are studied experimentally (Maia and de Silva 1997). Also, creating fixed interfaces for substructures in experimental work is generally not feasible. An alternative formulation to overcome these difficulties is as follows. Consider the substructure Eq. 1 and determine the natural frequencies and mode shapes by solving the eigenvalue problem $K^A \phi^A = \omega_A^2 M^A \phi^A$. This leads to the $N_A \times N_A$ matrices: $\Lambda^A = \text{diag}[\omega_{Ai}^2]$ and Φ^A such that $[\Phi^A]^t M^A \Phi^A = I$ and $[\Phi^A]^t K^A \Phi^A = \Lambda^A$. Now u^A is represented by a k_A term expansion $u^A(t) = \Phi_{k_A}^A p_{k_A}(t)$. The modal

matrix is further partitioned as $\Phi_{k_A}^A = [\Phi_{k_A i}^A \quad \Phi_{k_A c}^A]^t$, and using Eq. 1, one gets

$$I\ddot{p}_{k_A}(t) + \Lambda_{k_A}^A p_{k_A}(t) = [\Phi_{k_A c}^A]^t f_c^A(t) \quad (6)$$

A similar equation can also be obtained for subsystem B, and by combining these equations, one gets

$$\begin{aligned} [I] \begin{Bmatrix} \ddot{p}_{k_A}(t) \\ \ddot{p}_{k_B}(t) \end{Bmatrix} + \begin{bmatrix} \Lambda_{k_A}^A & 0 \\ 0 & \Lambda_{k_B}^B \end{bmatrix} \begin{Bmatrix} p_{k_A}(t) \\ p_{k_B}(t) \end{Bmatrix} \\ = \begin{bmatrix} [\Phi_{k_A c}^A]^t & 0 \\ 0 & [\Phi_{k_B c}^B]^t \end{bmatrix} \begin{Bmatrix} f_c^A(t) \\ f_c^B(t) \end{Bmatrix} \end{aligned} \quad (7)$$

The condition $u_c^A(t) = u_c^B(t)$ at the interface is expressed as $\Phi_{k_A c}^A p_{k_A}(t) = \Phi_{k_B c}^B p_{k_B}(t)$. This equation is rewritten as

$$\begin{aligned} [\Phi_{k_A c}^A \quad -\Phi_{k_B c}^B] \begin{Bmatrix} p_{k_A}(t) \\ p_{k_B}(t) \end{Bmatrix} = 0 \Rightarrow \\ Sp = 0 \text{ with } S = [\Phi_{k_A c}^A \quad -\Phi_{k_B c}^B] \end{aligned} \quad (8)$$

Now, the matrix S is partitioned as $S = [S_d \quad S_i]$ where S_d is a nonsingular square matrix and remaining part is S_i . This leads to

$$\begin{aligned} [S_d \quad S_i] \begin{Bmatrix} p_d \\ p_i \end{Bmatrix} = 0 \Rightarrow p_d = -S_d^{-1} S_i p_i \text{ and} \\ \begin{Bmatrix} p_{k_A}(t) \\ p_{k_B}(t) \end{Bmatrix} = \begin{Bmatrix} p_d \\ p_i \end{Bmatrix} = \begin{bmatrix} -S_d^{-1} S_i \\ I \end{bmatrix} p_i = \Psi \{p_i\} \end{aligned} \quad (9)$$

Furthermore, by noting that $f_c^A(t) + f_c^B(t) = 0$, one gets

$$I \begin{Bmatrix} \ddot{p}_{k_A}(t) \\ \ddot{p}_{k_B}(t) \end{Bmatrix} + \begin{bmatrix} \Lambda_{k_A}^A & 0 \\ 0 & \Lambda_{k_B}^B \end{bmatrix} \begin{Bmatrix} p_{k_A}(t) \\ p_{k_B}(t) \end{Bmatrix} = 0 \quad (10)$$

By using Eq. 9 in the above equation and premultiplying by Ψ^t , one gets

$$M\ddot{q} + Kq = 0 \text{ with } M = \Psi^t \Psi; K = \Psi^t \begin{bmatrix} \Lambda_{k_A}^A & 0 \\ 0 & \Lambda_{k_B}^B \end{bmatrix} \Psi \quad (11)$$

This equation can now be analyzed to deduce the built-up system natural frequencies and mode shapes. It is emphasized that the mass and stiffness matrices for the built-up system here are constructed in terms of the natural frequencies, and mode shapes of the subsystem in the uncoupled states and knowledge of substructure structural matrices are not needed in this formulation. This enables the introduction of experimentally studied substructures to be embedded into the modal synthesis of the built-up system.

Multilevel Substructuring

Consider an N -dof system being split into N_S substructures, such that dof of the largest substructure n satisfies the requirement $n \ll N$. If the value of N is large, the value of N_S also needs to be large in order to satisfy this requirement. This results in a large number of interface dofs appearing in the model thereby increasing the size of the synthesized system equation and consequent demands on computational efforts. The multilevel substructuring method addresses this issue by allowing for small substructures to be used without having to include a large number of interface dofs in the modal synthesis step.

The first step in multilevel substructuring is splitting the structure into multiple substructures, which form the first level of substructures. Then, each of these substructures is subdivided into multiple substructures, which form the second level. This process is continued until the substructures are as small as necessary. The substructures follow a tree topology, with the system to be analyzed at the top. At each level, each substructure has one parent one level above it and multiple-child substructures one level below. The equation of motion for each substructure (up to the penultimate level) can be

synthesized using its child substructures using component mode synthesis, and the interface elements belonging to those substructures alone appear in the synthesized system equation. Modal reduction is achieved in each of the substructures, and the interface dofs are retained. After the equations of motion of all substructures at a level are synthesized, the algorithm proceeds to the level above it. A fraction of the interface dofs at any level appears as interface dofs at the level above it, the rest turning into interior dofs, which can now be eliminated using modal reduction. Therefore, for the structure, the equation of motion can have, in the extreme case, only the interface dofs at the first level. This means that the method can be used to achieve a high degree of modal reduction. Additionally, at each level, the substructures can be analyzed in parallel, as in the case of single-level substructuring. At the higher levels, the number of substructures, and parallelizability, increases (see section “[Domain Decomposition Techniques and Their Parallel Implementation](#)” for issues related to parallelization).

The multilevel substructuring method has been used for linear and nonlinear structures with a large number of dofs (Papalukopoulos and Natsiavas 2007). Bennighof and Kaplan (1998) describe an adaptive multilevel substructuring procedure, where the contribution of each mode to the strain energy is used to decide whether or not to include the mode in the substructure analysis. This allows for a control on error due to modal reduction.

Nonlinear Substructures

Treatment of structural nonlinearities within the framework of ideas presented in the preceding sections offers several conceptual challenges. For problems involving linear substructures coupled through nonlinear elements, the fixed- and free-interface methods can be extended in a relatively easy manner. However, for problems involving globally distributed nonlinearities, the extensions

are not straightforward. An approach that employs fixed-interface component mode synthesis based on the idea of nonlinear normal modes has been developed by Apiwattanalungarn et al. (2005). In a conservative structure vibrating in its nonlinear normal mode, the motion would be periodic and all points on the structure reach their respective extrema simultaneously and pass through zeros simultaneously (see Kerschen et al. 2009 for a detailed introduction). A few details of this approach are presented here.

Consider a nonlinear system divided into two nonlinear substructures A and B , as shown in Fig. 1a. The equation for substructure A is written in partitioned form as

$$\begin{bmatrix} M_{ii}^A & M_{ic}^A \\ M_{ci}^A & M_{cc}^A \end{bmatrix} \begin{Bmatrix} \ddot{u}_i^A \\ \ddot{u}_c^A \end{Bmatrix} + \begin{bmatrix} K_{ii}^A & K_{ic}^A \\ K_{ci}^A & K_{cc}^A \end{bmatrix} \begin{Bmatrix} u_i^A \\ u_c^A \end{Bmatrix} + \begin{Bmatrix} G_i^A(u_i^A, u_c^A) \\ G_c^A(u_i^A, u_c^A) \end{Bmatrix} = \begin{Bmatrix} 0_i^A \\ f_c^A \end{Bmatrix} \quad (12)$$

Here, $G_i^A(u_i^A, u_c^A)$ and $G_c^A(u_i^A, u_c^A)$ are the nonlinear forces associated with the interior and coupling dofs, respectively, and $f_c^A(t)$ is the vector of interaction forces between A and B . The procedure followed here has much in common with the fixed-interface method described earlier for linear substructures (section “[Fixed-Interface Modal Synthesis](#)”). In order to relate u_c^A to u_i^A , two separate analyses are done. First, fixed-interface linear modes are computed for A by assuming $u_c^A = 0$ and ignoring the nonlinear terms in Eq. 11. The governing equation is obtained as $M_{ii}^A \ddot{u}_{i(Fix)}^A + K_{ii}^A u_{i(Fix)}^A = 0$, which has natural frequencies ω_r^A and modal vectors $\phi_k^A, r = 1, 2, \dots, n_i^A$. The system response represented by a modal expansion (truncated to k_A modes) is $u_{i(Fix)}^A = [\Phi_{i_{k_A}}^A] \{p_{k_A}^A(t)\}$. Next, a relationship is obtained between u_c^A and u_i^A under static conditions, again ignoring nonlinear terms in Eq. 11. This leads to the relation $u_{i(Free)}^A(t) = -[K_{ii}^A]^{-1} K_{ic}^A u_c^A(t)$. The solution vector in Eq. 11 is represented as

$$\begin{Bmatrix} u_i^A \\ u_c^A \end{Bmatrix} = \begin{bmatrix} \Phi_{ik_A}^A & -[K_{ii}^A]^{-1} K_{ic}^A \\ 0 & I \end{bmatrix} \begin{Bmatrix} p_{k_A}^A(t) \\ u_c^A(t) \end{Bmatrix} \quad (13)$$

Substituting Eq. 12 in Eq. 11 leads to

$$\begin{bmatrix} I & \tilde{M}_{ic}^A \\ \tilde{M}_{ci}^A & \tilde{M}_{cc}^A \end{bmatrix} \begin{Bmatrix} \ddot{p}_{k_A}^A \\ \ddot{u}_c^A \end{Bmatrix} + \begin{bmatrix} \Lambda^A & 0 \\ \tilde{K}_{ci}^A & \tilde{K}_{cc}^A \end{bmatrix} \begin{Bmatrix} p_{k_A}^A \\ u_c^A \end{Bmatrix} + \begin{Bmatrix} \tilde{G}_i^A(p_{k_A}^A, u_c^A) \\ \tilde{G}_c^A(p_{k_A}^A, u_c^A) \end{Bmatrix} = \begin{Bmatrix} 0_i^A \\ f_c^A \end{Bmatrix} \quad (14)$$

Here, $\tilde{M}_{ic}^A$, Λ^A , $\tilde{M}_{ci}^A$, $\tilde{M}_{cc}^A$, $\tilde{K}_{ci}^A$, and $\tilde{K}_{cc}^A$ are given in Eq. 3, $\tilde{G}_i^A(p_{k_A}^A, u_c^A) = (\Phi_{ik_A}^A)^T G_i^A(\Phi_{ik_A}^A p_{k_A}^A, u_c^A)$, and $\tilde{G}_c^A(p_{k_A}^A, u_c^A) = G_c^A(\Phi_{ik_A}^A p_{k_A}^A, u_c^A)$.

In Eq. 13, setting $u_c^A = 0$, one gets $\ddot{p}_{k_A}^A + \Lambda^A p_{k_A}^A + \tilde{G}_i^A(p_{k_A}^A, u_c^A) = 0_i^A$. Now, a nonlinear normal mode is constructed for this equation. The M^{th} coordinate of $p_{k_A}^A$, $p_{k_A,M}^A$, is taken as the master node, and all other dofs are related to it by the following constraints

$$p_{k_A,i}^A = X_i(p_{k_A,M}^A, p_{k_A,M}^A), i = 1, 2, \dots, k_A, i \neq M \quad (15)$$

$$\dot{p}_{k_A,i}^A = Y_i(p_{k_A,M}^A, \dot{p}_{k_A,M}^A), i = 1, 2, \dots, k_A, i \neq M \quad (16)$$

These constraint relations are used to reduce A to a $N_C + 1$ degree of freedom system with the equation of motion

$$\begin{bmatrix} \hat{M}_{ii}^A & \hat{M}_{ic}^A \\ \hat{M}_{ci}^A & \hat{M}_{cc}^A \end{bmatrix} \begin{Bmatrix} \ddot{p}_{k_A,M}^A \\ \ddot{u}_c^A \end{Bmatrix} + \begin{bmatrix} \hat{K}_{ii}^A & \hat{K}_{ic}^A \\ \hat{K}_{ci}^A & \hat{K}_{cc}^A \end{bmatrix} \begin{Bmatrix} p_{k_A,M}^A \\ u_c^A \end{Bmatrix} + \begin{Bmatrix} \hat{G}_i^A(p_{k_A,M}^A, \dot{p}_{k_A,M}^A, u_c^A) \\ \hat{G}_c^A(p_{k_A,M}^A, \dot{p}_{k_A,M}^A, u_c^A) \end{Bmatrix} = \begin{Bmatrix} 0 \\ f_c^A \end{Bmatrix} \quad (17)$$

Here, the terms $\hat{M}_{ii}^A$, $\hat{M}_{ic}^A$, $\hat{M}_{ci}^A$, $\hat{M}_{cc}^A$, $\hat{K}_{ii}^A$, $\hat{K}_{ic}^A$, $\hat{K}_{ci}^A$, $\hat{K}_{cc}^A$, $\hat{G}_i^A(p_{k_A,M}^A, \dot{p}_{k_A,M}^A, u_c^A)$, and $\hat{G}_c^A(p_{k_A,M}^A, \dot{p}_{k_A,M}^A, u_c^A)$ are obtained by substituting the constraints Eqs. 14 and 15 in Eq. 13. A similar procedure is followed for substructure B , and the equation of motion is obtained as

$$\begin{bmatrix} \hat{M}_{ii}^B & \hat{M}_{ic}^B \\ \hat{M}_{ci}^B & \hat{M}_{cc}^B \end{bmatrix} \begin{Bmatrix} \ddot{p}_{k_B,M}^B \\ \ddot{u}_c^B \end{Bmatrix} + \begin{bmatrix} \hat{K}_{ii}^B & \hat{K}_{ic}^B \\ \hat{K}_{ci}^B & \hat{K}_{cc}^B \end{bmatrix} \begin{Bmatrix} p_{k_B,M}^B \\ u_c^B \end{Bmatrix} + \begin{Bmatrix} \hat{G}_i^B(p_{k_B,M}^B, \dot{p}_{k_B,M}^B, u_c^B) \\ \hat{G}_c^B(p_{k_B,M}^B, \dot{p}_{k_B,M}^B, u_c^B) \end{Bmatrix} = \begin{Bmatrix} 0 \\ f_c^B \end{Bmatrix} \quad (18)$$

Applying the constraints $u_c^A = u_c^B = u_c$ and $f_c^A + f_c^B = 0$, Eqs. 16 and 17 combine to form the following equation of motion for the built-up system:

$$\begin{bmatrix} \hat{M}_{ii}^A & 0 & \hat{M}_{ic}^A \\ 0 & \hat{M}_{ii}^B & \hat{M}_{ic}^B \\ \hat{M}_{ci}^A & \hat{M}_{ci}^B & \hat{M}_{cc}^A + \hat{M}_{cc}^B \end{bmatrix} \begin{Bmatrix} \ddot{p}_{k_A,M}^A(t) \\ \ddot{p}_{k_B,M}^B(t) \\ \ddot{u}_c(t) \end{Bmatrix} + \begin{bmatrix} \hat{K}_{ii}^A & 0 & \hat{K}_{ic}^A \\ 0 & \hat{K}_{ii}^B & \hat{K}_{ic}^B \\ \hat{K}_{ci}^A & \hat{K}_{ci}^B & \hat{K}_{cc}^A + \hat{K}_{cc}^B \end{bmatrix} \begin{Bmatrix} p_{k_A,M}^A(t) \\ p_{k_B,M}^B(t) \\ u_c(t) \end{Bmatrix} + \begin{Bmatrix} \hat{G}_i^A(p_{k_A,M}^A, \dot{p}_{k_A,M}^A, u_c^A) \\ \hat{G}_i^B(p_{k_B,M}^B, \dot{p}_{k_B,M}^B, u_c^B) \\ \hat{G}_c^A(p_{k_A,M}^A, \dot{p}_{k_A,M}^A, u_c^A) + \hat{G}_c^B(p_{k_B,M}^B, \dot{p}_{k_B,M}^B, u_c^B) \end{Bmatrix} = 0 \quad (19)$$

The reduced system has $N_C + 2$ dofs. This procedure can be extended straightforwardly to more than two substructures.

Mesh Partitioning and Implicit–Explicit Schemes

Here time integration of equations of motion of nonlinear dynamical systems is considered. Methods to obtain numerical solutions can be either explicit or implicit in nature. In the context of linear dynamical systems, implicit methods of integration are preferred since they are unconditionally stable and permit the analyst to choose step size based on engineering judgment. On the other hand, explicit schemes for such systems are conditionally stable, and the step size is often controlled by the stability requirements. For nonlinear dynamical systems, implicit methods typically involve solution of nonlinear algebraic/transcendental equations at every time step. On the other hand, explicit schemes are non-iterative, and the computational effort needed at each time step is less than that for implicit schemes. Given this, in structural dynamic problems involving spatially localized nonlinearities, it becomes advantageous to bifurcate the degrees of freedom into those associated with linear and those with nonlinear parts of the system. Such a strategy has earlier been proposed by Hughes et al. (1979) and in more recent years has gained attention in the context of substructuring involving hybrid simulation (see section “Hybrid Simulations”).

The essence of this strategy can be explained by considering an N degree of freedom dynamical system governed by the equation $M\ddot{U} + F(U, \dot{U}) = G(t)$, $U(0) = U_0$, $\dot{U}(0) = \dot{U}_0$. The elements are bifurcated into an implicit and an explicit set, and the corresponding internal force vectors are $F^I(U, \dot{U})$ and $F^E(U, \dot{U})$, respectively. The equation of motion becomes $M\ddot{U} + F^I(U, \dot{U}) + F^E(U, \dot{U}) = G(t)$, $U(0) = U_0$, $\dot{U}(0) = \dot{U}_0$. This equation is solved using a predictor–corrector scheme. The prediction step provides the initial estimates of the values of

displacement and velocity vectors at the $(n + 1)^{th}$ time step, $n = 1, 2, \dots$, as $\tilde{U}_{n+1}$ and $\dot{\tilde{U}}_{n+1}$, respectively. An implicit scheme is then used for the corrector step to obtain U_{n+1} and $\dot{U}_{n+1}$. The value of the explicit force vector used in the corrector step is $F^E(\tilde{U}_{n+1}, \dot{\tilde{U}}_{n+1})$. If the implicit partition of the mesh is linear, $F^I(U, \dot{U})$ is of the form $C^I\dot{U} + K^IU$, where C^I and K^I are the implicit damping and mass matrices. The equation in the corrector step, $M\ddot{U}_{n+1} + C^I\dot{U}_{n+1} + K^IU_{n+1} = G(t_{n+1}) - F^E(\tilde{U}_{n+1}, \dot{\tilde{U}}_{n+1})$, is also linear. Hence, a non-iterative implicit scheme can be used to obtain U_{n+1} and $\dot{U}_{n+1}$. In this procedure, although mesh partitioning is done, the implicit and explicit dofs are evaluated together. On the other hand, in problems of fluid–structure interaction, although the flow equation and the equation of motion of the structure are coupled, they are usually solved separately. In the strong coupling method (Ahn and Kallinderis 2006) based on predictor–corrector steps, an explicit scheme is used for the predictor step. The flow equation is then solved in order to provide the input for solving the equation of the structure. The corrector step involves solving the structure and flow equations in that sequence iteratively. Here, the time step used in the equations for the fluid and the structure has to be equal. If the structure is linear, implicit schemes can be used with a large time step. Both implicit and explicit schemes can be applied in solving the flow equations, with implicit schemes allowing for larger time steps, but requiring an iterative solution.

Domain Decomposition Techniques and Their Parallel Implementation

Present-day computer processors possess multicore architecture, and the usual sequential solvers are unable to use all the cores simultaneously. Therefore, the resource remains underutilized. Furthermore, with the advent of computer technology, parallel computers (often referred to as clusters) can be easily and economically built from off-the-shelf components.

Therefore, to take advantage of these inexpensive technologies, the solvers need to be parallelized.

As the name domain decomposition (DD) suggests, in these techniques the computational domain is decomposed into a set of subdomains, and a divide-and-conquer strategy is developed that permits using multiple processors more efficiently than the traditional single-domain approaches. For parallelizing a time-dependent (deterministic) problem, two options can be explored: (i) parallelizing in space and (ii) parallelizing in time. While there have been attempts on parallelizing in time for a few applications, it is generally very difficult due to the inherent sequential nature in the temporal domain. Whereas the spatial dependence is not sequential, although the subdomains are coupled. Therefore, to achieve parallelization in space, in a typical (iterative) DD method, (a) the spatial domain is divided into a number of subdomains, (b) each of the processors independently performs the computation related to one or more subdomains, and (c) at the end of each iteration, the processors communicate relevant information among them. As the iteration grows, the method converges toward achieving the global (considering the entire spatial domain) equilibrium and compatibility. A few dominant methods are component mode synthesis (CMS), finite element tearing and interconnecting (FETI) and its variants, and Schur complement (this is a direct method). One major use of DD methods is development of preconditioners for linear systems. When the condition number of the coefficient matrix A in a linear system $Ax = b$ becomes high, iterative techniques such as conjugate gradient becomes very slow. In this situation, a preconditioner matrix P is often used where PA has a low condition number, and finally $PAx = Pb$ is solved. DD is often used in developing P (Ghosh et al. 2009).

FETI methods are iterative methods where the subdomains communicate with each other through a set of Lagrange multipliers defined at subdomain boundaries. In a static problem, the equilibrium of each subdomain is satisfied at each iteration whereas the continuity of the displacement field is achieved at convergence.

The estimates from a DD method must be in good agreement with the exact result, that is, the results from a direct solver using a single domain. The success of a DD method is measured by its convergence rate toward the exact solution and by its scalability. In terms of scalability of an iterative method when the iteration count does not depend strongly on the problem size, then it is called numerical scalability. On the other hand, when an m -times bigger problem can be solved using m -times bigger computer (in terms of number of processors) in similar time, it is called parallel scalability. The efficiency of a DD method is often measured in terms of the ratio H/h where H denotes a characteristic size of a subdomain and h denotes element size.

A linear structural dynamics problem can be solved in two ways, either using a modal reduction or using a direct integration. For the modal approach, the domain decomposition can be used in the eigenvalue computation level, as described earlier in this entry (section “[Fixed-Interface Modal Synthesis](#)”). For direct integration, two separate approaches can be taken. First, casting the vibration problem in a single domain and adopting an implicit time integration scheme and then using the domain decomposition in the linear system solving level. To this end, the methods developed for solving elliptic equations can be directly used. For instance, when the Newmark-beta time integration is used, the system of linear algebraic equations can be solved using DD at every time step. The second option in the direct integration is to cast the structural dynamics problem in multiple domains itself and develop a time integrator. For this purpose, the compatibility and equilibrium among the subdomains must be ensured for all time instants or a selected set of time instants when different time steps are considered at different subdomains. In a FETI-type scheme, a Lagrange multiplier is used. Both these approaches are outlined below.

To explain the first option, consider a linear static problem (which essentially should be viewed as a system of linear algebraic equations)

$$Ku = f \quad (20)$$

Now let the computational domain D be divided into N_s subdomains with the s -th subdomain denoted as $D^{(s)}$ and the boundary as $\partial D^{(s)}$. In the original formulation of FETI, these domains are completely unconnected; however, in a later development (FETI-DP), the corners of the subdomains are assumed to be connected. Following the FETI-DP formulation (Farhat et al. 2000), within the subdomain $D^{(s)}$, partition the displacement field $u^{(s)}$ as

$$u^{(s)} = \begin{bmatrix} u_{int}^{(s)} \\ u_{b_r}^{(s)} \\ u_{b_c}^{(s)} \end{bmatrix} = \begin{bmatrix} u_r^{(s)} \\ u_{b_c}^{(s)} \end{bmatrix} \text{ with } u_r^{(s)} = \begin{bmatrix} u_{int}^{(s)} \\ u_{b_r}^{(s)} \end{bmatrix} \quad (21)$$

where the subscript *int* means interior, b_c denotes the nodes at the corner, b_s denotes the nodes at the boundary but not on the corner, and r means the collection of *int* and b_r , that is, all the nodes except the corners – often referred to as *residual*. Accordingly, let the stiffness matrix and forcing vector be partitioned as

$$K^{(s)} = \begin{bmatrix} K_{rr}^{(s)} & K_{rc}^{(s)} \\ K_{cr}^{(s)} & K_{cc}^{(s)} \end{bmatrix}, f^{(s)} = \begin{bmatrix} f_r^{(s)} \\ f_{b_c}^{(s)} \end{bmatrix} \quad (22)$$

Let λ denote a vector-valued Lagrange multiplier defined globally over the subdomain interface dofs except the corner points, that is, on the dof denoted by b_r . Following these notations, the equilibrium for the interior and boundary (except the corner nodes) is written as

$$K_{rr}^{(s)} u_r^{(s)} + K_{rc}^{(s)} B_c^{(s)} u_c + B_r^{(s)T} \lambda = f_r^{(s)}; s = 1, 2, \dots, N_s \quad (23)$$

and the equilibrium for the corner dofs is written as

$$\begin{aligned} \sum_{s=1}^{N_s} B_c^{(s)T} K_{rc}^{(s)T} u_r^{(s)} + \sum_{s=1}^{N_s} B_c^{(s)T} K_{cc}^{(s)T} B_c^{(s)} u_c \\ = \sum_{s=1}^{N_s} B_c^{(s)T} f_{b_c}^{(s)} \end{aligned} \quad (24)$$

where $B_c^{(s)}$ denotes a Boolean matrix mapping the global vector u_c containing all corner DOF to $u_{b_c}^{(s)}$ and $B_r^{(s)}$ denotes another Boolean matrix mapping the residual dof to $u_{b_r}^{(s)}$. The Boolean matrices are (usually rectangular) matrices with the entries as zeros or one. They are used to describe the binary relationship among the elements of two vectors. For instance, consider the Boolean relationship $B_c^{(s)} u_c = u_{b_c}^{(s)}$

where the Boolean matrix $B_c^{(s)}$ operates on the vector u_c enlisting all the corner DOFs and produces a vector $u_{b_c}^{(s)}$ of lower dimension that enlists only the corner DOFs in the subdomain $\mathcal{D}^{(s)}$. To explain it further, consider the i^{th} row of the matrix $B_c^{(s)}$. The j^{th} column in this row will be 1 if the j^{th} element in the vector u_c corresponds to the i^{th} element in the vector $u_{b_c}^{(s)}$, with all other entries being zero. The compatibility condition at the subdomain interface is stated as

$$u_{b_c}^{(s)} - u_{b_c}^{(q)} = 0 \text{ on } \partial D^{(s)} \cap \partial D^{(q)} \quad (25)$$

Upon a few algebraic operations, Eqs. 22, 23, and 24 are combined to a single system of linear algebraic equations with λ as the unknown vector. This resulting system is solved using a preconditioned conjugate gradient (PCG) method. The parallelization is achieved by distributing the subdomains among the processors and thereby distributing the operations involved in (Eqs. 22, 23, and 24).

To apply this method in a direct integration scheme, consider an implicit scheme such as Newmark-beta. Here the linear system of equations is needed to be solved at every time step

$$\hat{K} u_{t+\Delta t} = \hat{f}_{t+\Delta t} \quad (26)$$

where the coefficient matrix $\hat{K}$ and the vector $\hat{f}_{t+\Delta t}$ depend upon stiffness, mass and damping matrices, the parameters of the integration scheme, and the time step. In this case, the aforementioned DD solver can be used at every time step. Instead of an iterative method such as FETI, direct methods such as Schur complement can also be used which can be parallelized.

For explaining the Schur complement, consider the structure with two subdomains in Fig. 1a. Accordingly, the stiffness matrix, displacement vector, and the force vector are partitioned, and the static equilibrium condition is written as

$$\begin{bmatrix} K_{ii}^A & 0 & K_{ic}^A \\ 0 & K_{ii}^B & K_{ic}^B \\ K_{ci}^A & K_{ci}^B & K_{cc}^A + K_{cc}^B \end{bmatrix} \begin{Bmatrix} u_i^A \\ u_i^B \\ u_c \end{Bmatrix} = \begin{Bmatrix} f_i^A \\ f_i^B \\ f_c \end{Bmatrix} \quad (27)$$

To solve this equation, first the displacement at the boundary, u_c , is computed by solving

$$\begin{aligned} \sum_{s=A,B} \left(K_{cc}^s - K_{ci}^s (K_{ii}^s)^{-1} K_{ic}^s \right) u_c \\ = \sum_{s=A,B} \left(f_c^s - K_{ci}^s (K_{ii}^s)^{-1} f_i^s \right) \end{aligned} \quad (28)$$

where the coefficient matrix is known as the Schur complement. Then the displacement at internal degrees of freedom is computed by solving

$$K_{ii}^s u_i^s = f_i^s - K_{ic}^s u_c; s = A, B \quad (29)$$

The computation in Eqs. 27 and 28 can be parallelized by performing the computation related to subdomains A and B in two different processors.

In the second approach (Prakash and Hjelmstad 2004), consider the equation of motion of a forced undamped motion as

$$M\ddot{u}(t) + ku(t) = f(t) \quad (30)$$

Now let the domain D be divided into N_s subdomains without the connectivity at the corner. Then the equation of motion for each subdomain can be written as

$$M^{(s)} \ddot{u}^{(s)}(t) + K^{(s)} u^{(s)}(t) + B^{(s)T} \lambda = f^{(s)}(t) \quad (31)$$

where $B^{(s)}$ denotes a Boolean matrix mapping from the subdomain s to a global vector

containing the dofs at all the inter-subdomain boundaries (interfaces) and λ is a Lagrange multiplier, this time with the units of velocity. Inter-subdomain compatibility is imposed in terms of velocity as

$$\sum_{s=1}^{N_s} B^{(s)} \dot{u}^{(s)}(t) = 0 \quad (32)$$

Once again, using a Newmark-beta scheme, the discretized equation

$$\begin{aligned} \mathcal{M}^{(s)} \ddot{\mathbf{u}}_{i+1}^{(s)}(t) + \mathcal{B}^{(s)} \lambda_{i+1} \\ = \mathbf{f}_{i+1}^{(s)}(t) - \mathcal{N}^{(s)} \mathbf{u}_i^{(s)}; s = 1, 2, \dots, N_s \end{aligned} \quad (33)$$

where the index i denotes the time step,

$$\mathbf{u}_i^{(s)} = \begin{Bmatrix} \ddot{u}_i^{(s)} \\ \dot{u}_i^{(s)} \\ u_i^{(s)} \end{Bmatrix}, \mathbf{f}_i^{(s)} = \begin{Bmatrix} f_i^{(s)} \\ \mathbf{0} \\ \mathbf{0} \end{Bmatrix}, \mathcal{B}^{(s)} = \begin{Bmatrix} B^{(s)T} \\ \mathbf{0} \\ \mathbf{0} \end{Bmatrix}, \quad (34)$$

and the matrices $\mathcal{M}^{(s)}$ and $\mathcal{N}^{(s)}$ follow from the Newmark-beta scheme. Equations 32 and 31 are then solved to find the time history of response. Parallelization follows from the data independence of the subdomains in Eq. 32 by distributing the subdomains among the processors. Note that this method can be further refined to accommodate different time steps in different subdomains. In that case, the compatibility condition Eq. 32 is not enforced at every time step, but after a few time steps.

In all the DD methods described in this section, no reduction in the total number of degrees of freedom is made. Therefore, the error associated with static condensation of component mode synthesis where a set of degrees of freedom is eliminated from the calculation does not appear here. However, approximations arise from usage of iterative solvers. The main goal here is to reduce the total computational cost by distributing the total computational burden among a

number of processors. To this end, two important criteria must be followed for computational efficiency. First, the computational load must be equally distributed among the processors; this is called load balancing. Second, the inter-processor communication should be minimal, as communication is slower than computation within a processor. In a distributed memory system, the Message Passing Interface (MPI) is used for parallelization. Once a parallel algorithm is developed, often the data structure needs to be reworked. Then the computer program (written in, for instance, Fortran/C/C++) is augmented by a set of MPI commands to distribute and manage the computation among the processors (Karniadakis and Kirby 2003). The parallel computing would help not only in solving very large and complicated problems, but also in the context of reliability analysis where numerous repeated analyses are required.

Hybrid Simulations

The idea of substructuring also has applications in laboratory testing-based performance assessment of engineering systems for dynamic loads such as those included due to earthquakes. The test hardware here consists of computer-controlled servo-hydraulic actuators driving either a shake table or serving as loading devices in reaction-wall-based systems. In conventional test methods, these actuators are either in displacement or a force control and aim to apply prescribed time variations of dynamic loads on the test structure. This strategy however suffers from two drawbacks, namely, the need to geometrically scale and structure the study (dictated by limitations on payload capacity of the shake table and actuator force ratings) and the neglect of possible dynamic interaction between structural subsystems being tested (e.g., heavy machinery like turbines, rotors, and pumps) and the structure in which these systems are housed. Newer testing protocols aimed at overcoming these limitations which employ substructuring

schemes have been developed in recent years (Saouma and Sivaselvan 2008; Bursi and Wagg 2008; Williams and Blakeborough 2001). Two such strategies, namely, pseudodynamic testing and real-time substructuring, are discussed in the following:

The pseudodynamic test is carried out on a reaction-wall-based system using servo-hydraulic actuators under displacement control. The test structure under study is modeled as

$$M\ddot{X} + C\dot{X} + R[X(\tau), 0 \leq \tau \leq t] \\ = -M\Gamma\ddot{x}_g(t), X(0) = X_0, \dot{X}(0) = \dot{X}_0 \quad (35)$$

Here, the term R represents nonlinearities due to inelastic behavior of the structure, Γ is the influence matrix, and $\ddot{x}_g(t)$ is the earthquake-induced ground acceleration. The basic premise of this approach is that the inertial, viscous damping and external force characteristics can be numerically modeled while the inelastic nonlinear term R is obtained experimentally. This would mean that the terms $M\ddot{X}$, $C\dot{X}$, and $-M\Gamma\ddot{x}_g(t)$ constitute the numerical component of the test while R represents the experimental component. Furthermore, the time variable t is scaled to slow down such that the test involves the application of only static actions. A partial finite element model of the structure which characterizes the terms $M\ddot{X}$, $C\dot{X}$, and $-M\Gamma\ddot{x}_g(t)$ is first formulated, and it is embedded into the software that commands the servo-hydraulic actuators. The actuators are kept under displacement control. Beginning with an initial guess on stiffness characteristics, the finite element equations of motion are integrated to determine displacement vector $X(t)$ at $t = \Delta t$. These displacements are further applied on the test structure experimentally, and the reactions transferred to the walls are measured through load cells. These measured reactions serve to establish the term R in the governing equation of motion, which in turn is used to advance the integration steps. Thus the integration of equation of motion and measurement of nonlinear stiffness characteristics of

the structure go hand in hand leading to determination of structural response to a specified $\ddot{x}_g(t)$. The works of Takanashi and Nakashima (1987), Severn et al. (1989), Nakashima (2001), and Williams and Blakeborough (2001) provide comprehensive overviews on related topics. It is important to note that the earthquake-induced structural displacements are evaluated computationally and applied to the structure in a static manner. This test clearly does not permit experimental evaluation of time dependant nonlinear behavior, rate dependant stress–strain laws, and behavior of active elements, if any.

The real-time substructure test does not scale the time as is done in pseudodynamic testing. The test structure here is spatially divided into two parts: one which permits reliable numerical modeling and the other that requires experimental testing. The finite element model for the numerical substructure is embedded into the control software that controls the actuators, and the experimental substructure is placed on the test rig. Clearly the numerical and experimental models are coupled, and the integration of governing equations for the numerical model and the testing of experimental substructure take place hand in hand in real time. This requires online exchange of interfacial forces/displacements between the two substructures which in turn places demands on accurate sensing, fast data acquisition, and efficient numerical integration. Several issues related to time delays, measurement noise, and choice of integration schemes become critical in this context. Examples of recent studies include those by Sajeeb et al. (2009), Chen and Ricles (2010, 2012), and Gao et al. (2013).

Summary

A wide range of contexts in which substructuring is used in finite element modeling is reviewed. The motivations for substructuring include desire to reduce the model size, desire to combine experimental and numerical modeling

approaches in an online or offline manner for steady-state/transient dynamics, and desire to equitably employ computational resources in a parallel computing environment. The challenges here pertain to choice of substructures, incompleteness in modal, spatial, and (or) frequency domains, dissipative mechanisms, local/global nonlinearities, and non-idealities in experimental work (such as measurement noise and time delays) and their role in selection of algorithmic parameters. The overview presented in this entry has endeavored to bring out currently available techniques and tools in the area of modeling.

Cross-References

- ▶ [Classically and Nonclassically Damped Multi-degree of Freedom \(MDOF\) Structural Systems, Dynamic Response Characterization of Modal Analysis](#)

References

- Ahn HT, Kallinderis Y (2006) Strongly coupled flow/structure interactions with a geometrically conservative ALE scheme on general hybrid meshes. *J Comput Phys* 219(2):671–696
- Apiwattanalungarn P, Shaw SW, Pierre C (2005) Component mode synthesis using nonlinear normal modes. *Nonlinear Dyn* 41(1–3):17–46
- Bennighof JK, Kaplan MF (1998) Frequency window implementation of adaptive multi-level substructuring. *J Vib Acoust* 120(2):409–418
- Bursi OS, Wagg D (eds) (2008) *Modern testing techniques for structural systems: dynamics and control*. Springer, New York
- Chen C, Ricles J (2010) Tracking error-based servohydraulic actuator adaptive compensation for real-time hybrid simulation. *J Struct Eng* 136(4):432–440
- Chen C, Ricles JM (2012) Large-scale real-time hybrid simulation involving multiple experimental substructures and adaptive actuator delay compensation. *Earthq Eng Struct Dyn* 41:549–569
- Craig RR Jr (1995) Substructure methods in vibration. *J Mech Des* 117:207
- Ewins DJ (2000) *Modal testing: theory, practice and application*. Research Studies Press, Baldock

- Farhat C, Lesoinne M, Pierson K (2000) A scalable dual–primal domain decomposition method. *Numer Linear Algebra Appl* 7:687–714
- Gao X, Castaneda N, Dyke SJ (2013) Real time hybrid simulation: from dynamic system, motion control to experimental error. *Earthq Eng Struct Dyn* 42:815–832
- Ghosh D, Avery P, Farhat C (2009) FETI-preconditioned conjugate gradient method for large-scale stochastic finite element problems. *Int J Numer Methods Eng* 80(6–7):914–931
- Hughes TJ, Pister KS, Taylor RL (1979) Implicit-explicit finite elements in nonlinear transient analysis. *Comput Methods Appl Mech Eng* 17:159–182
- Hurty WC (1965) Dynamic analysis of structural systems using component modes. *AIAA J* 3(4):678–685
- Karniadakis GE, Kirby RM II (2003) Parallel scientific computing in C++ and MPI: a seamless approach to parallel algorithms and their implementation. Cambridge University Press, New York
- Kerschen G, Peeters M, Golnival JC, Vakakis AF (2009) Nonlinear normal modes, part I: a useful framework for structural dynamicist. *Mech Syst Signal Process* 23:170–194
- Klerk DD, Rixen DJ, Voormeeren SN (2008) General framework for dynamic substructuring: history, review and classification of techniques. *AIAA J* 46(5):1169–1181
- Maia NMM, de Silva JMM (eds) (1997) Theoretical and experimental modal analysis. Research Studies Press, Taunton
- Nakashima M (2001) Development, potential, and limitations of real-time online (pseudo-dynamic) testing. *Philos Trans R Soc Lond Ser A Math Phys Eng Sci* 359(1786):1851–1867
- Papalukopoulos C, Natsiavas S (2007) Dynamics of large scale mechanical models using multilevel substructuring. *J Comput Nonlinear Dyn* 2(1):40–51
- Prakash A, Hjelmstad KD (2004) A FETI-based multi-time-step coupling method for Newmark schemes in structural dynamics. *Int J Numer Methods Eng* 61:2183–2204
- Sajeeb R, Manohar CS, Roy D (2009) A conditionally linearized Monte Carlo filter in nonlinear structural dynamics. *Int J Nonlinear Mech* 44:776–790
- Saouma VE, Sivaselvan MV (eds) (2008) Hybrid simulation: theory, implementation and applications. CRC Press, London
- Severn RT, Brownjohn JMW, Dumanoglu AA, Taylor CA (1989) A review of dynamic testing methods for civil engineering structures. In: *Proceedings of the conference on civil engineering dynamics*, University of Bristol, pp 1–24
- Takanashi K, Nakashima M (1987) Japanese activities on on-line testing. *J Eng Mech* 113(7):1014–1032
- Williams MS, Blakeborough A (2001) Laboratory testing of structures under dynamic loads: an introductory review. *Philos Trans R Soc Lond Ser A Math Phys Eng Sci* 359(1786):1651–1669

Sustained Earthquake Preparedness: Functional, Social, and Cultural Issues

Douglas Paton¹, Norio Okada², Julia Becker³ and Li-ju Jang⁴

¹School of Psychology, University of Tasmania, Launceston, TAS, Australia

²Graduate School of Science and Engineering, Kumamoto University, Kumamoto City, Japan

³GNS Science, Lower Hutt, New Zealand

⁴Department of Social Work, National Pingtung University of Science and Technology, Pingtung, Taiwan

Synonyms

Culture; Preparedness; Readiness; Risk communication

Introduction

When earthquakes occur, affected societies and their members suddenly find themselves having to deal with demands that differ considerably from anything they would encounter under normal conditions and in circumstances in which normal societal functions and resources are marked by their absence. The aftershock sequence that can accompany seismic events can prolong the period over which people have to deal with disruption. However, the degree of disruption and loss that people, communities, and societies experience is a function of the degree to which they have developed the knowledge, skills, and relationships required to anticipate, cope with, adapt to, and recover from earthquake consequences during both the initial event and the consequences they can encounter as they cycle through response and recovery processes with successive aftershocks. Furthermore, the fact that earthquakes occur without warning makes it imperative that people prepare prior to any event. Because it is impossible to predict when the next

earthquake will occur (it could be years or decades or longer into the future), preparedness strategies must ensure that once developed, preparedness is sustained by people and communities. The task of facilitating sustained preparedness occurs as part of a comprehensive risk management strategy.

Risk management offers ways in which societies and their members (individually and collectively) can make choices about mitigating risk and facilitating their preparedness to respond to seismic hazard events using a mix of mitigation and preparedness strategies. Mitigation and preparedness strategies play complementary roles in risk management.

Mitigation strategies describe ways societies can prepare themselves in advance of earthquakes occurring by undertaking activities designed to prevent or minimize the risk earthquakes pose to a society and its members. Mitigation encompasses, for example, land-use planning (e.g., precluding building in areas susceptible to liquefaction) and developing and implementing building codes and standards (e.g., specifying building codes to include, for instance, base isolation to increase building capacity to withstand ground shaking) and retrofitting existing buildings to facilitate their capacity to withstand the action of seismic hazards (at least up to a point). Building codes can reduce the adverse impacts of hazards and increase the likelihood of a building remaining available for people's use after the event. However, additional strategies may be required if it is possible to anticipate future earthquake events whose intensity or duration would exceed design parameters and retrofitting capability.

For example, a mitigation measure may function effectively when impacted by a 100-year event yet fail catastrophically if a 500-year (typically more intense) event is experienced. Thus, people could experience problems from events at the higher end of the spectrum of seismic intensities, magnitudes, and durations (e.g., continuing damage resulting from prolonged aftershocks such as what occurred in Christchurch) that could exceed the parameters of structural measures designed in part on

cost-benefit and political criteria (which reflect the level of risk mitigation a society is willing to pay for). This introduces into the risk management process a need for strategies designed to reduce the degree of loss and disruption experienced through encouraging personal and community preparedness.

The goal of preparedness strategies is to increase the likelihood of people and communities being in a position to be able to *respond* in planned and functional (resilient and adaptive) ways to the complex, challenging, and emergent challenges and demands that earthquakes create rather than having to *react* to them in ad hoc ways. For example, ensuring the physical integrity of the house and storing water and food helps people deal with the effects of ground shaking on their home and on the loss of utilities. Taking steps to ensure the structural integrity of the home not only reduces the risk of injury and death to its inhabitants, but it also increases the likelihood of people having shelter during their recovery, reduces demands on societal resources for temporary accommodation, and increases the probability that people will remain in an area and be able to participate in social (mutual aid, social support), economic, and environmental recovery activities. The effectiveness of the mutual aid and social support people within communities can provide for each other and the quality of the working relationship between community members and civic agencies during and after earthquake events is a function of the degree to which community members have developed the knowledge, skills, and relationships with neighbors, community members, and civic agencies necessary to expedite local response and recovery initiatives before earthquakes occur (Paton and McClure 2013).

Preparedness programs aim to facilitate the proactive development of household, community, and societal capabilities (e.g., develop emergency plans and resources, capacity for self-reliance, ability to work with others to confront local problems, etc.) in ways that increase people's ability to anticipate what they may have to contend with and develop their capacity to cope with, adapt to, and recover from the physical, personal, and social consequences that

earthquakes create (Paton and McClure 2013). Achieving this goal falls within the remit of comprehensive risk management (i.e., where responsibility is shared between community members and agencies in ways that increase the likelihood of their respective activities playing complementary roles) in ways that devolve (partial) responsibility for personal and local risk management to individuals and communities. Involving people in risk management is pursued using risk communication and community outreach strategies.

In order to design and implement risk communication and community outreach strategies, it is important to know what to communicate and engage people about. The content of risk communication tends to reflect the fact that, fundamentally, risk is a product of the likelihood of earthquakes occurring and the consequences that arise when they do occur. Communicating about both, and particularly about consequences, is essential. Communicating information about the earthquake likelihood is not a good predictor of preparedness (McClure et al. 1999; Mileti and Darlington 1995). Preparedness is more likely when people believe they are likely to suffer negative consequences from an earthquake if they do not prepare (Palm et al. 1990). This suggests that risk communication strategies profit from including information about earthquake consequences and encourage citizens to personalize this risk in ways that focus on how they can prepare in ways that can prevent or minimize adverse earthquake consequences (Paton and McClure 2013). To do so requires several pieces of information.

The first concerns identifying what people need to know and be able to do (individually and collectively) to cope with and adapt to the earthquake hazard consequences that emerge and evolve over the course of the response, recovery, and rebuilding phases of earthquake disasters. That is, identifying earthquake consequences and the physical and social demands they create for people and identifying the strategies required to prepare for these consequences and demands (i.e., what being comprehensively prepared looks like). It is also important to identify why some people prepare and others less so or not at all.

Knowledge of the latter informs how risk communication and community outreach strategies can be developed to facilitate sustained, comprehensive preparedness. Finally, because earthquakes occur in culturally diverse locations, a third issue is ascertaining whether or not preparedness theories and practices are applicable across cultures. The first issue to be tackled is identifying what comprehensive preparedness is.

Comprehensive Earthquake Preparedness

The goal of preparedness strategies is to facilitate people's ability to deal with the full spectrum of demands that unfold over prolonged periods of response and recovery. People need to be able to deal with direct (e.g., impact of ground shaking, liquefaction, and aftershocks on buildings and infrastructure) and secondary (e.g., loss of lifelines like water, power, and sewerage services and the consequent need for people to be self-reliant and able to continue to function as well as possible in the absence of normal services) hazard consequences. As physical and social rebuilding occurs, people also need to be prepared to contend with challenges emanating from recovery processes (e.g., dealing with government agencies, insurance companies, builders, etc.) that may persist for weeks, months, or years. Furthermore, they may, as was evident in Christchurch, New Zealand, following the 2011 earthquake, have to cycle through response and recovery several times as they deal with the implications of aftershock sequences that prolonged people's experience of the physical and social consequences of earthquake activity (Paton et al. 2014). Recognition of the diverse issues people could encounter led to the development of functional typologies of earthquake preparedness that maps the content of preparedness programs onto the demands and challenges people have to contend with over time.

Functional Typologies

Russell et al. (1995) developed a preparedness typology that comprised three factors: structural,

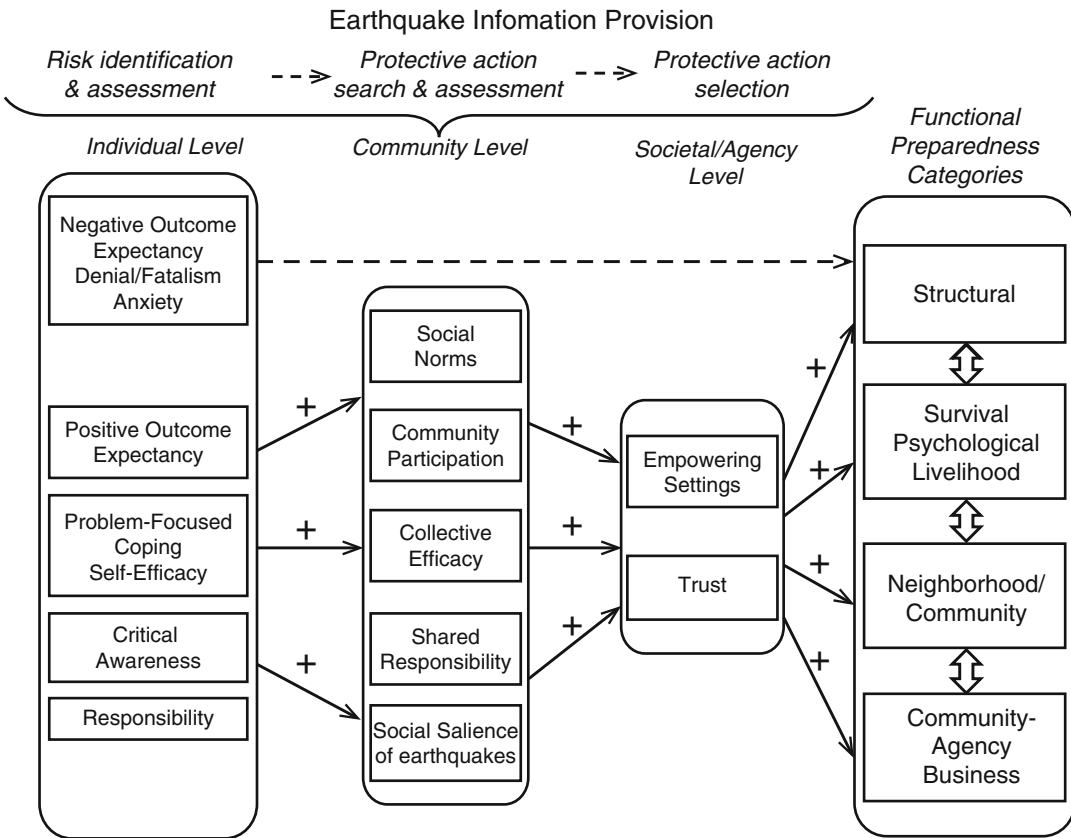
survival, and planning preparedness. Structural actions encompass activities that secure the house (e.g., secure house to foundations) and its contents (e.g., securing water heaters and tall furniture) to prevent contents from injuring inhabitants (e.g., from ground shaking). Survival actions facilitate people's capacity for self-reliance during periods of disruption (e.g., ensuring a supply of water/dehydrated or canned food for several days, having a radio with spare batteries, etc.). Finally, planning includes, for example, developing household hazard plans and attending meetings to learn about earthquakes and how to deal with their consequences. The latter introduces a social dimension into how preparedness is conceptualized. A subsequent factor analytic study (Lindell et al. 2009) proposed that preparedness comprise direct action (e.g., learn how to shut off utilities, have a 4-day supply of canned food, strap heavy objects, etc.) and capacity building (e.g., join an earthquake-related organization, attend meetings about earthquake hazards) factors. These studies identify a need for preparedness strategies to facilitate the development of structural, survival, planning, and capacity building (social) activities. The importance of including all these aspects within a typology of comprehensive earthquake preparedness has been reinforced by studies of what people affected by earthquakes identified as being required to increase their ability to deal with the consequences of earthquakes.

For example, in Christchurch, NZ, residents affected by the 2011 earthquake reinforced the importance of developing (pre-event) structural and survival preparedness and household emergency planning (Paton et al. 2014). Christchurch respondents reported how their lack of survival and planning preparedness made coping with the loss of essential societal utilities (e.g., loss of water, power, and sewerage services) more challenging than it would have been had they prepared before the earthquake struck. Respondents also called for preparedness programs to address dealing with loss of or disruption to people's livelihoods (both directly from damage to place of work and indirectly from being relocated or injured), developing psychological preparedness,

adapting to changes in living conditions and loss of and disruption to family and social relationships, and accommodating the fact that the period of loss and disruption people have to confront extends to months and years. A need for effort to be directed to developing social relationships and social competencies within a preparedness strategy was identified by those affected by the 2011 earthquake.

Respondents suggested that preparedness programs should develop the ability of neighborhood and community group members to collaborate in ways that facilitate their ability to effectively confront local physical and social demands (e.g., removing rubble, providing mutual support, setting up community meeting places, taking care of those with special needs, organizing local efforts to repair homes, identifying and meeting local needs) when responding under conditions in which normal societal resources and functions are absent. Respondents also called for effort to be directed to preparedness programs including content to develop, for example, community leadership, social inclusion, prioritizing problems, problem-solving and decision-making skills. The latter, in conjunction with developing community capacity to represent diverse community and local needs to agencies and government will facilitate the ability of communities to proactively secure the resources they need to take responsibility for their own recovery (Paton et al. 2014).

Taken together, this discussion highlights a need for preparedness to encompass structural, survival, planning, and community capacity capabilities. The Christchurch work introduced a need to complement these functional categories with those addressing livelihood and psychological and community-societal relationship (e.g., between communities and governments, NGOs, business, etc.) preparedness. These are summarized in Fig. 1 (right-hand column). While the Christchurch research discussed what people realized they needed during a disaster, the focus of preparedness research is on developing these capacities and relationships prior to the occurrence of a hazard event. A need to research the latter reflects the finding that comprehensive preparedness is the exception rather than the rule.



Sustained Earthquake Preparedness: Functional, Social, and Cultural Issues, Fig. 1 Summary of the relationship between protective action decision-making;

personal-, social-, and agency-level predictors; and the functional categories of earthquake preparedness

Despite the considerable time, effort, and expenditure invested in earthquake public education programs, levels of preparedness for earthquakes typically remain low (Lechlitter and Willis 1996; Lindell et al. 2009; Paton et al. 2014; Paton and McClure 2013). Finding considerable diversity in existing levels of preparedness introduces the second issue discussed in this chapter; the need to account for why some people prepare comprehensively, some much less so, and others not at all.

Accounting for Differences in Preparedness

When making decisions about uncertain, threatening events, several dispositional or individual-level predictors of preparedness have been identified. Some, such as fatalism, reduce the likelihood of preparing. Others influence how people

make judgments about who is responsible for their preparedness. Some increase the likelihood of people preparing.

People who are fatalistic about hazard activity are unlikely to prepare (Turner et al. 1986). Another dispositional characteristic that reduces preparedness is denial (Crozier et al. 2006). If people believe that they have no control over an earthquake and/or its activity, they can attempt to cope with this by denying the seriousness of the risk. Household and personal preparedness is also less likely if people believe that risk management agencies are responsible for all aspects of public safety and so transfer responsibility for community preparedness from themselves to agencies (Paton and McClure 2013). Certain characteristics of people's social relationships influence preparedness.

If people believe that their significant others (parents, spouses, friends, peer group, etc.) hold favorable attitudes toward preparedness or that performance of a specific behavior is likely to be interpreted favorably by significant others, they are more likely to perform these actions (Smith and Terry 2003). Prevailing social norms regarding preparedness expectations of significant others and personal motivation to act in ways consistent with these expectations increases the likelihood of preparing (McIvor and Paton 2007). The work on individual predictors has been further developed in several theories of earthquake preparedness.

Person-relative-to-event (PrE) theory (Duval and Mulilis 1995) describes how the degree to which people perceive themselves at risk from an earthquake interacts with person (self-efficacy, outcome efficacy) and event (severity of event and probability of occurrence of the event) variables to predict preparedness. The PrE model argues that someone who appraises their resources as sufficient in both quality and quantity *relative* to the perceived magnitude of a particular threat will engage in more problem-focused coping activities than one who appraises their personal resources as insufficient relative to the seriousness of the threat. This work introduces a need to include perceived vulnerability, severity, response efficacy (outcome expectancy), problem-focused coping, and anxiety (fear-arousing persuasive communication) as predictors of earthquake preparedness (see Fig. 1).

Building on ideas advanced in the PrE theory, Paton et al. (2005) developed a theory that identified how critical awareness or the relative social salience (i.e., the extent to which people think and talk about earthquakes and earthquake preparedness with others in their community) of earthquakes was a key motivator of preparedness. Paton et al.'s (2005) critical awareness theory demonstrated that the anxiety, outcome expectancy (response efficacy), self-efficacy, and problem-focused/action coping mediated the relationship between critical awareness and preparedness and so identified a need for these variables to be included as predictors of earthquake

preparedness (see Fig. 1). Other theories have focused on how people's information search and evaluation influences the degree to which they take protective action.

The Protective Action Decision Model (PADM) (e.g., Lindell and Hwang 2008) argues that people's decisions about earthquake preparedness reflects the judgments they make about three factors: the threat posed by an earthquake, identifying the actions available and capable of protecting them against earthquake consequences, and identifying the sources from whom information about hazard and protective measures can be obtained. The PADM proposes that this involves people proceeding through several stages of information seeking and evaluation: detection/warning, psychological preparation, logistical preparation, and protective action selection and implementation (see Fig. 1). The information required for people to act at each stage can be sourced from the physical environment and/or people's social context (e.g., other people in their community, emergency management, media, etc.).

Lindell's work discussed the fact that people do not take information from expert sources at face value and introduced a need to consider both information about earthquake risk and preparedness and the quality of the (communication) relationship between the sources and recipients of earthquake risk information. An important contextual element of this relationship is that it occurs under conditions of uncertainty. People may have no or limited experience of earthquakes, and the infrequent nature of earthquake activity makes it impossible for them to independently test the validity of the information about risk and preparedness. Under these circumstances, an important aspect of the relationship between source of information and its recipients is the degree to which recipients trust sources of information. Levels of risk acceptance and people's willingness to take responsibility for their own safety is increased, and decisions to take steps to actively manage their risk more likely, if people believe that their relationship with risk management agencies is fair and empowering (e.g., agencies are perceived as trustworthy, as

acting in the interest of community members) (e.g., Paton 2008).

Paton (2008) developed a community engagement theory that captured key elements of these relationships, empowerment and trust, and the personal beliefs and social competencies that influenced preparedness decision-making under uncertainty. This theory has demonstrated an ability to predict earthquake preparedness (Paton 2013).

Paton (2008) argued that if people hold negative outcome expectancy beliefs (belief that earthquakes are uncontrollable and their consequences are too catastrophic for personal action to make any difference to peoples' safety), their likelihood of preparing is reduced. In contrast, holding positive outcome expectancy beliefs (beliefs that personal actions can enhance personal safety and/or mitigate hazard consequences) motivates people to start to prepare. However, the infrequent and complex nature of earthquakes means that believing that preparing can be effective does not necessarily equate with knowing what to do or how to prepare. To reduce uncertainty and guide action, people look to other community members and to expert sources (Paton 2008). Paton's community engagement theory introduced a need to include outcome expectancy, community participation and collective efficacy (which serve to socially construct risk beliefs and identify how risk can be reduced through social action by articulating community members preparedness plans), empowerment, and trust as predictors of earthquake preparedness (see Fig. 1).

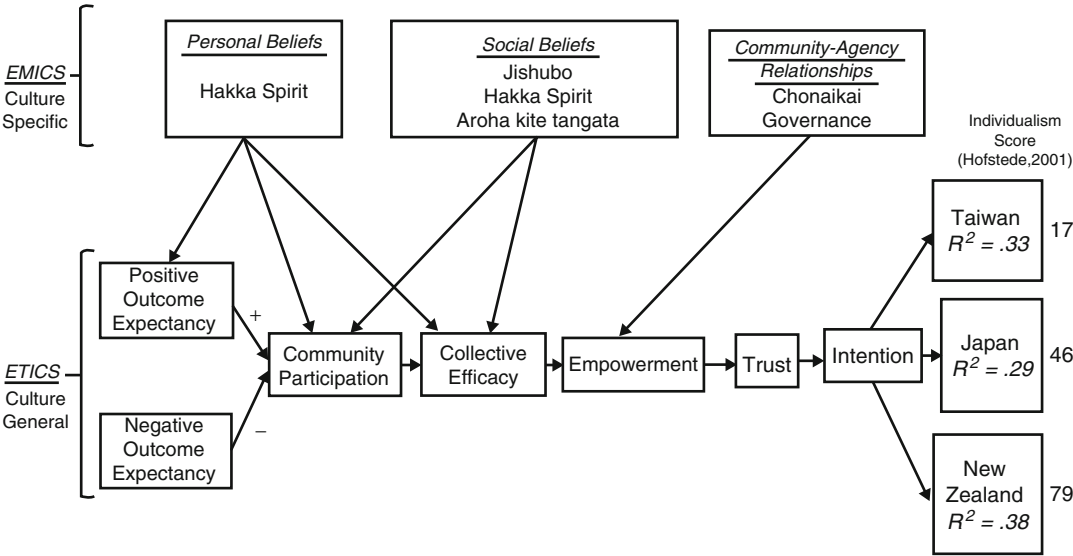
Figure 1 summarizes the several variables that were implicated as predictors in the theories discussed above. It also includes how the PADM suggestions about information search and action might relate to these other variables. Collectively, this depiction could serve as a framework for future research into earthquake preparedness. The theories introduced above have empirically demonstrated their ability to account for differences in levels of earthquake preparedness. This confers upon them a capacity to inform the development of risk communication and community outreach strategies (see below)

by identifying how information content and personal and social competencies and relationships complement one another in ways that facilitate the ability of people to make earthquake preparedness decisions under conditions of uncertainty. However, given the fact that earthquakes affect people in countries characterized by considerable social and cultural diversity, it cannot be automatically assumed that they will retain their predictive utility if applied in different cultural contexts. Developing a comprehensive understanding of earthquake preparedness requires considering this cultural dimension.

Cross-Cultural Issues in Earthquake Preparedness

One reason for exercising caution in this regard derives from the fact that the theories introduced above were developed and tested in countries that are culturally highly individualistic. The variables they tested reflected this. In contrast, Asian countries, where a disproportionate number of significant and damaging earthquakes occur, tend to fall at the collectivistic end of the cultural spectrum in which social processes play relatively more important roles in people's decisions and actions (Triandis 1995). This raises issues regarding the applicability of Western-developed theories in Asian cultures that typically lie at the collectivistic end of these cultural dimensions. Examining whether theory can be applied across the individualistic-collectivistic spectrum is important (Paton et al. 2013).

A preliminary investigation of cultural equivalence was undertaken by examining whether Paton's (2008) community engagement theory (see above) of preparedness could account for differences in earthquake preparedness in countries (New Zealand, Japan, and Taiwan) that differ with regard to their relative positions on the individualistic-collectivistic cultural spectrum (see Fig. 2). This theory was selected for this initial comparison because its inclusion of both individual-level (outcome expectancy) and community-/social-level process (community participation and collective efficacy) variables provides a useful starting point for studying cultures that



Sustained Earthquake Preparedness: Functional, Social, and Cultural Issues, Fig. 2 Summary of cross-cultural studies of earthquake preparedness and the

relative contribution of culture-general (etic) and culture-specific (emic) processes to earthquake preparedness

differ with regard to their implicit emphasis on individual (individualistic cultures) and social (collectivistic) characteristics. For example, the relative position of a country on the I-C dimension could influence the comparative importance of person-level versus group-level factors on preparedness decisions (Paton et al. 2013). Empirical analysis of the ability of this theory to predict earthquake preparedness in New Zealand, Japan, and Taiwan (Fig. 2) supports the view that Western theory can be used across cultural borders (Paton et al. 2013). In Fig. 2, the R^2 value refers to how well the model can explain differences in people's intention to prepare for earthquakes (e.g., in Taiwan, the model explained 33% (or .33) of these differences). Models that predict over 28% are regarded as offering a good explanation (Sheeran 2002).

Evidence for cross-cultural equivalence should not, however, be taken to imply that cultural factors can be ignored. Cultural studies distinguish between etic (culturally universal) and emic (culture-specific) processes. While the cross-cultural analyses discussed above identified how the same process (Fig. 2) could operate across different countries (i.e., it describes an etic process), this does not mean

that the process is enacted in the same way in different cultures. For example, community participation was important in each country. However, the origins of participatory activities and how they are sustained and enacted to influence people's risk beliefs and preparedness options differ from country to country. Thus, emic processes (i.e., how and why participation occurs in a specific culture) capable of affecting earthquake preparedness need to be identified. Doing so can provide valuable insights into the theoretical and practical aspects of earthquake preparedness (see below). This is illustrated here with reference to New Zealand Māori and to how Jishubo (in Japan) and the Hakka Spirit (in Taiwan) describe emics that affect earthquake preparedness.

The importance of accommodating emic characteristics was evident in New Zealand. Analysis of unique cultural characteristics capable of facilitating social preparedness was evident in Māori populations. The Māori cultural value of "aroha ki te tangata" (love to all people) represents an emic characteristic (Fig. 2) capable of informing understanding of how community participation influences preparedness in this population (Paton et al. 2014).

Jishu-bosai-soshiki (Jishubo) are neighborhood-based “autonomous organizations for disaster reduction” unique to Japan (Bajek et al. 2008). Jishubo is a resource available to local government that can be used to organize earthquake preparedness events within communities (e.g., drills, workshops) and facilitate local response and rescue activities (e.g., rescue residents, provide initial first aid, supply food and water to survivors) should an earthquake occur. Jishubo are organized and operate through Chonaikai (“community councils”), a traditional Japanese neighborhood governance system that includes within its functions the maintenance of safe and secure communities through implementing various community activities (e.g., disaster risk reduction and crime prevention). This example illustrates how culture-specific (emic) characteristics can influence how community (Jishubo) and community-agency (Chonaikai) relationships (i.e., they represent unique ways of enabling community participation and empowerment) inform earthquake preparedness in ways unique to Japanese society (Bajek et al. 2008) (see Fig. 2). Another example of how culture-specific social mechanisms can influence earthquake preparedness is found in Taiwan.

Jang and LaMendola (2006) discussed how the Hakka Spirit, a cultural characteristic of the Hakka people residing in Tung Shih in Taiwan (affected by the 921 earthquake that hit Taiwan in 1999) influences earthquake preparedness. The Hakka Spirit describes an emic process, a unique set of cultural beliefs and practices that encompasses qualities such as frugality, diligence, self-reliance, responsibility, and persistence that combine to predispose people to prepare for adversity. Jang and LaMendola discussed how farming practices that evolved to limit the impact of typhoons on people’s livelihoods generated enduring outcome expectancy beliefs (regarding the effectiveness of collective actions in response to hazard consequences). These were linked to increasing the likelihood that people would act to secure their homes from earthquake consequences. Culturally specific reciprocal support and collaborative problem-solving practices, developed over time as part of an adaptive response to meeting

the challenges imposed by agricultural life in remote mountain communities, represent the source of the community participation and collective efficacy described in the above analyses (Fig. 2).

Practical Implications

Being able to identify (Fig. 1) factors that facilitate and hinder earthquake preparedness can inform the development of practical risk communication and community outreach strategies. For example, fatalistic attitudes can be reduced by asking people what can be done to help specific vulnerable groups, such as people living in unsafe buildings or young children in schools (McClure et al. 2001; Turner et al. 1986). When people deliberated about preparedness in vulnerable groups, they became less fatalistic and were more likely to perceive preparedness actions as more manageable. Denial and negative outcome expectancy can be countered by providing information that focuses people’s attention specifically on how preparedness actions can mitigate the consequences of hazard activity and increase the degree of control people perceive themselves having over earthquake consequences (Lehman and Taylor 1987) and by sourcing information about preparedness and its effectiveness from comparable (to the target one) communities (Paton and McClure 2013).

Positive outcome expectancy beliefs can be developed by presenting people with a small number of items initially, starting with relatively easily adopted items and introducing progressively more complex tasks over time (Lindell and Perry 2000). By presenting specific explanations about how preparedness actions reduce risk and mitigate hazard consequences progressively over time, sustained adoption is more likely (Paton and McClure 2013). Social and relationship predictors (e.g., community participation, collective efficacy, social norms, etc.) can be enhanced by integrating community development and risk management strategies and by ensuring that community outreach adopted empowerment and community engagement principles (Paton and McClure 2013). This raises the possibility that one reason for the success of

Chonaikai is by integrating its disaster risk reduction activities with other community activities.

The work discussed above introduced how earthquake preparedness theory can possess etic quality that allows it to be used in a broad range of cultural contexts. Demonstrating that occidental theories retain their predictive utility in oriental contexts has several theoretical and practical benefits. For example, it would provide a common theoretical basis for collaborative learning and research across national borders. Demonstrating a degree of cross-cultural theoretical equivalence provides risk management agencies in different countries with access to a wider range of risk management options (which are implemented by accommodating local emic processes). It would make information about how to facilitate preparedness available to countries that lack the resources to undertake this work themselves and inform the development of humanitarian aid planning knowing that they have a robust framework that could be used to assist earthquake recovery irrespective of the country they find themselves being deployed at short notice. It remains important, however, to accommodate emic, culture-specific processes in the development of comprehensive earthquake preparedness theory and practice (e.g., by mapping local cultural practices (emic) onto this etic framework). Researching emic cultural influences on how earthquake preparedness and recovery processes are developed, sustained, and enacted also opens up new research and outreach opportunities for developing progressively comprehensive theories of earthquake preparedness.

Summary

The three countries that were specifically discussed in this chapter, New Zealand, Japan, and Taiwan, along with many others, all sit on the circum-Pacific seismic belt (the Pacific Ring of Fire). Approximately 90 % of the world's earthquakes and 81 % of its largest earthquakes occur along the Ring of Fire. Earthquakes will remain an inevitable aspect of the natural history of these countries. However, people can prepare in ways

that can reduce their risk of experiencing adverse earthquake consequences and recover more promptly and effectively should an earthquake occur. This chapter first identified a need for preparedness to cover structural, planning, psychological, livelihood, community, and community-agency relationship categories (Fig. 1). Despite the evident advantages of being prepared, the fact that some people prepare comprehensively and others less so or not at all identifies a need for robust theories of hazard preparedness to inform DRR practice. This chapter also recognized the need for preparedness theory and strategy to accommodate the social and cultural diversity inherent in countries whose citizens must live with seismic risk. By demonstrating how theory can predict earthquake preparedness in culturally diverse countries, the chapter illustrated the potential for collaborative learning, research, and practice across national borders. However, it is important to ensure that national work accommodates the unique cultural resources present in each country that can be used to facilitate sustained earthquake preparedness.

Cross-References

- ▶ [Building Disaster Resiliency Through Disaster Risk Management Master Planning](#)
- ▶ [Community Recovery Following Earthquake Disasters](#)
- ▶ [Earthquake Disaster Recovery: Leadership and Governance](#)
- ▶ [Interim Housing Provision Following Earthquake Disaster](#)
- ▶ [Resilience to Earthquake Disasters](#)
- ▶ [Site Response for Seismic Hazard Assessment](#)
- ▶ [Social Media Benefits and Risks in Earthquake Events](#)

References

- Bajek R, Matsuda Y, Okada N (2008) Japan's Jishu-bosai-soshiki community activities: analysis of its role in participatory community disaster risk management. *Nat Hazards* 44:281–292

- Crozier M, McClure J, Vercoe J, Wilson M (2006) The effects of land zoning information on judgments about earthquake damage. *Area* 38:143–152
- Duval TS, Mulilis J-P (1995) A person-relative-to-event (PrE) approach to negative threat appeals and earthquake preparedness: a field study. *J Appl Soc Psychol* 29:495–516
- Jang L, LaMendola W (2006) The Hakka spirit as a predictor of resilience. In: Paton D, Johnston D (eds) *Disaster resilience: an integrated approach*. Charles C. Thomas, Springfield, pp 174–189
- Lechlitter GJ, Willis FN (1996) Living with earthquakes: beliefs and information. *Psychol Rec* 46:391–396
- Lehman D, Taylor SE (1987) Date with an earthquake: coping with a probable, unpredictable disaster. *Pers Soc Psychol Bull* 13:546–555
- Lindell MK, Hwang SN (2008) Households' perceived personal risk and responses in a multi-hazard environment. *Risk Anal* 28:539–556
- Lindell MK, Perry RW (2000) Household adjustment to earthquake hazard: a review of research. *Environ Behav* 32:461–501
- Lindell MK, Arlikatti S, Prater CS (2009) Why people do what they do to protect against earthquakes risk: perceptions of hazard adjustment attributes. *Risk Anal* 29:1072–1088
- McClure J, Walkey F, Allen M (1999) When earthquake damage is seen as preventable: attributions, locus of control and attitudes to risk. *Appl Psychol* 48:239–256
- McClure J, Allen MW, Walkey F (2001) Countering fatalism: causal information in news reports affects judgments about earthquake damage. *Basic Appl Soc Psychol* 23:109–121
- McIvor D, Paton D (2007) Preparing for natural hazards: normative and attitudinal influences. *Disaster Prev Manag* 16:79–88
- Mileti DS, Darlington JD (1995) Societal response to revised earthquake probabilities in the San Francisco Bay area. *Int J Mass Emerg Disasters* 13:119–145
- Palm R, Hodgson M, Blanchard RD, Lyons D (1990) *Earthquake insurance in California*. Westview Press, Boulder
- Paton D (2008) Risk communication and natural hazard mitigation: how trust influences its effectiveness. *Int J Glob Environ Issues* 8:2–16
- Paton D (2013) Disaster resilient communities: developing and testing an all-hazards theory. *J Integr Disaster Risk Manag* 3:1–17
- Paton D, McClure J (2013) Preparing for disaster: building household and community capacity. Charles C. Thomas, Springfield
- Paton D, Smith L, Johnston D (2005) When good intentions turn bad: promoting natural hazard preparedness. *Aust J Emerg Manag* 20:25–30
- Paton D, Okada N, Sagala S (2013) Understanding preparedness for natural hazards: a cross cultural comparison. *J Integr Disaster Risk Manag* 3:18–35
- Paton D, Johnston D, Mamula-Seadon L, Kenney CM (2014) Recovery and development: perspectives from New Zealand and Australia. In: Kapucu N, Liou KT (eds) *Disaster & development: examining global issues and cases*. Springer, New York
- Russell LA, Goltz JD, Bourque LB (1995) Preparedness and hazard mitigation actions before and after two earthquakes. *Environ Behav* 27:744–770
- Sheeran P (2002) Intention-behaviour relations: A conceptual and empirical review. In: Stroebe W, Hewstone M (eds) *European review of social psychology*. Wiley, Chichester, pp 1–36
- Smith JR, Terry DJ (2003) Attitude-behaviour consistency: The role of group norms, attitude accessibility, and mode of behavioural decision-making. *European J of Soc Psychol* 33:591–608
- Triandis HC (1995) *Individualism and collectivism*. Westview, Boulder
- Turner RH, Nigg JM, Paz DH (1986) *Waiting for disaster: earthquake watch in California*. University of California Press, Los Angeles

Symmetric Triaxial Seismometers

Bruce Townsend

Nanometrics, Inc., Kanata, Ontario, Canada

Synonyms

Galperin configuration seismometer; Homogeneous triaxial seismometer

Introduction

Of the class of seismic instruments measuring ground motion known as triaxial seismometers, most provide three signal outputs that represent mutually orthogonal motions in the East, North, and Vertical (or X, Y, and Z) directions (see entry “► [Principles of Broadband Seismometry](#)”). Of these, some are designed with three independent internal sensors that are respectively sensitive to motions in the three XYZ directions, and so directly measure the single vertical and two horizontal motion degrees of freedom. Some triaxial seismometers use a different configuration known as a Galperin arrangement (Galperin 1955), also known as a “symmetric triaxial” or “homogeneous” design (Melton and Kirkpatrick 1970).

In the Galperin configuration, the sensing elements are also arranged to be mutually orthogonal, but instead of one axis being vertical, all three are inclined upwards from the horizontal at precisely the same angle, as if they were aligned with the edges of a cube balanced on a corner. The operational principles of a design based on the Galperin configuration have many similarities as well as some important differences relative to one based on the conventional XYZ arrangement. The Galperin configuration presents some significant benefits for the users, owners, and manufacturers of seismometers that use this topology, as well as implying some trade-offs users should be aware of.

Benefits of a symmetric triaxial seismometer include being able to more easily distinguish external noise sources from internal ones, that mass centering does not compromise mutual orthogonality of the three axes, assurance of well-matched responses of the three outputs, and ability to achieve higher performance in smaller packages. A trade-off made by the Galperin approach is that unlike conventional XYZ seismometers that can suffer the failure of one axis element while continuing to provide valid output signals for the remaining two directions, all elements in a symmetric triaxial system must function for any of its XYZ outputs to be valid.

The first modern broadband seismometer to be based on a Galperin configuration was the Streckeisen STS-2, an observatory-grade vault instrument introduced in 1990. Nanometrics introduced a symmetric triaxial seismometer with a 240 s lower corner period, the Trillium 240, in 2004. There are now a wide variety of symmetric triaxial seismometers available including models designed for borehole, posthole, ocean bottom, and vault installation and ranging from ultracompact to the large form factors.

Principles of Operation

Inertial Sensing Principles

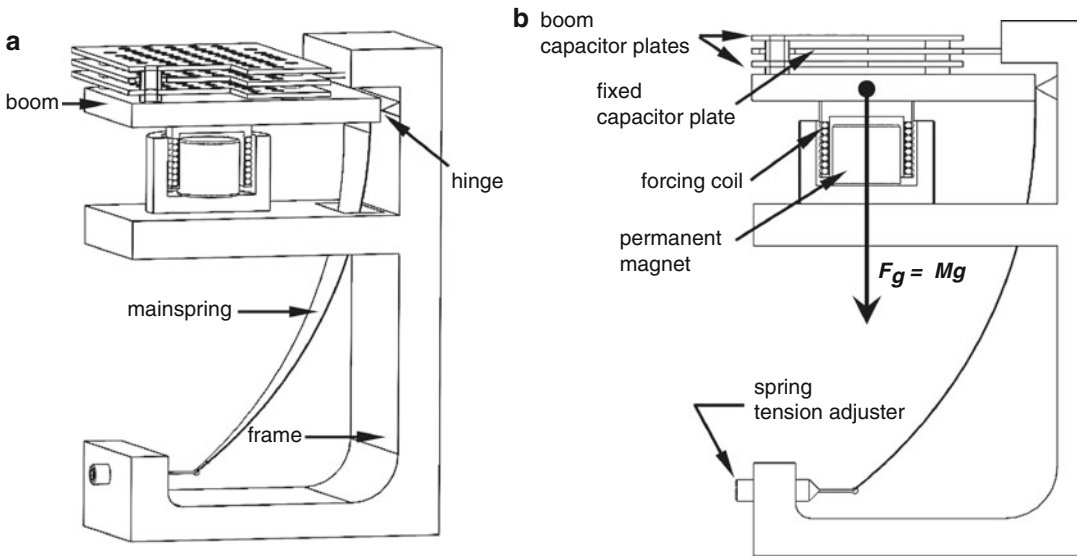
Measuring motion must always be done relative to some frame of reference. Seismometers by definition use an inertial reference; that is, motion

is sensed relative to a proof mass that is suspended in such a way that the proof mass tends to remain stationary while the seismometer that is coupled to the earth (or a structure) moves relative to it. The sensing element includes some means of measuring the motion of the sensing frame to its suspended proof mass. The proof mass in most broadband seismometers is a boom hinged at one end and balanced near a null point by a mainspring. A capacitive displacement transducer is often used to sense boom deflection.

A broadband seismometer achieves its wide response and exceptional linearity by means of a force-feedback electronic control loop that includes a forcing coil constantly working to keep the mass at dead center. The electric current required to counteract any deflection of the proof mass is proportional to the acceleration of the frame. In a sense, the forcing coil must at each instant apply exactly the same acceleration to the proof mass as the seismometer is experiencing for the proof mass to “keep up with” (stay stationary relative to) the seismometer. The seismometer output is derived from a signal within the control circuit that is proportional to the time integral of acceleration, providing a signal representative of velocity.

Broadband Axis Essential Elements

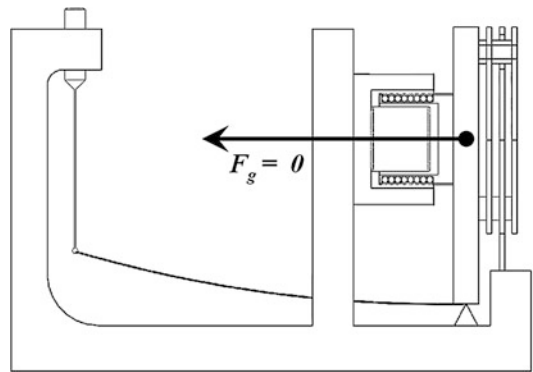
A typical arrangement for a broadband seismometer axis is to have a boom hinged at one end with two capacitor plates and a wire-wound coil mounted on it to serve as the displacement transducer and forcing coil, respectively. The mounting frame has one fixed capacitor plate arranged to sit between the two on the boom, and a permanent magnet and yoke fitting into and around the boom’s forcing coil. This arrangement is mounted within the pressure vessel of the seismometer in one attitude for the vertical axis (the boom and capacitor plates horizontal) but in the upright attitude for the horizontal axes (the boom and capacitor plates vertical with the hinge at the bottom). Usually, the electronic feedback circuit together with the power conditioning, output signals, controllers, and related circuitry is on printed circuit boards arranged above the three sensing elements.



Symmetric Triaxial Seismometers, Fig. 1 Force-feedback seismometer vertical axis construction. (a) Perspective view. (b) Side view (© Nanometrics)

Figure 1 shows conceptual mechanical diagrams of a typical broadband axis construction configured to sense vertical motion; both perspective and side views are shown. This and subsequent diagrams are highly simplified for the purpose of illustrating solely the essential functional elements. The hinged boom with its attached forcing coil (shown in cutaway) and two outer capacitor plates have a combined proof mass M and together with the mainspring make a mechanical oscillator. The electric forcing coil attached to the movable boom acts around a permanent magnet fixed to the frame and is driven by force-feedback control loop electronics to maintain the proof mass at its measurement null point, which is where the fixed center capacitor plate is equally centered between the two outer capacitor plates. The boom is pulled down by gravity with force F_g that must be counterbalanced by the mainspring applying torque at the hinge point.

Figure 2 shows the essential construction of a horizontal axis. The principles of the horizontal and vertical designs are evidently the same, but with the horizontal axis oriented so as to be sensitive to sideway acceleration and insensitive to vertical motion. A significant difference between



Symmetric Triaxial Seismometers, Fig. 2 Force-feedback seismometer horizontal axis construction (© Nanometrics)

the two designs is that the horizontal axis mainspring only supplies a restoring force when the boom deflects from its rest position and does not counterbalance any gravitational force ($F_g = 0$), whereas the vertical axis mainspring both supplies a restoring force and counterbalances the weight of the boom against gravity ($F_g = Mg$).

See entry “► [Principles of Broadband Seismometry](#)” for more information on the operational principles.

The XYZ Sensor Configuration

The conventional vertical/horizontal or XYZ arrangement within a triaxial seismometer has one sensing element arranged so that it is sensitive to vertical (Z) motion and two sensing elements that are sensitive to horizontal motions in the X and Y directions. This corresponds to the output signals preferred for most seismology purposes; one generally wants to record the vertical, East and North components. However, the design of the vertical sensing element is unavoidably distinct from that of the two horizontal elements. The constant gravitational acceleration of the Earth that acts on (in fact, defines) the up direction and is absent in the horizontal directions means the vertical and horizontal elements must be of different designs.

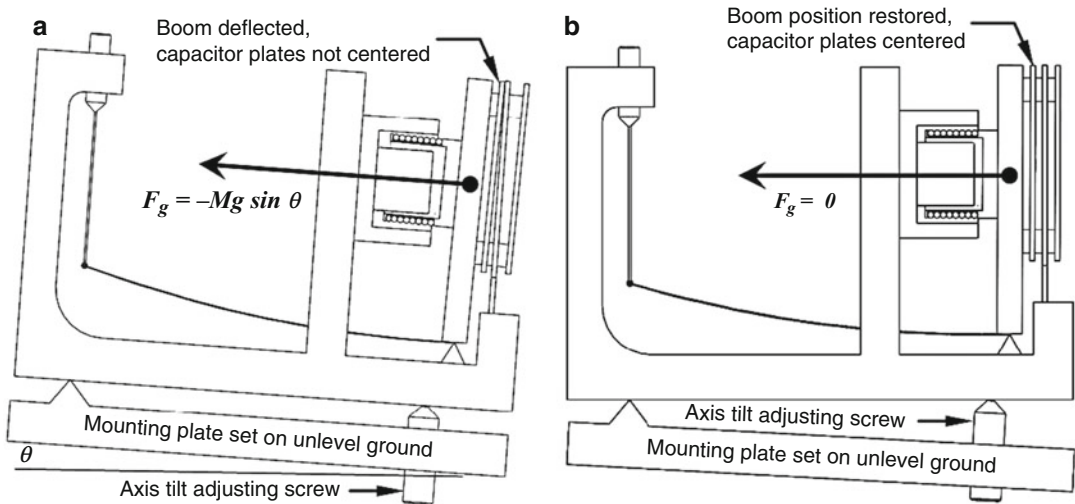
The essential differences forced on the design by the presence or absence of the constant $g_o \cong 9.81 \text{ m/s}^2$ acceleration of gravity are the spring suspending the boom and the mass centering mechanism if there is one. Other design differences may also arise from practical design considerations such as the physical orientation of the vertical axis sensor being different with respect to the seismometer case, common electronics circuit boards, connectors, and the like. The vertical axis will have a hinged boom lying in the horizontal plane needing a strong spring to suspend it against gravity, and will tend to be tall and narrow as in Figure 1. A horizontal axis, having an inverted pendulum oriented in the vertical plane, will tend to be wider and shorter as in Figure 2. This creates challenges for the designer when attempting to fit one vertical and two horizontal axis elements in a single enclosure and can result in larger enclosures and/or further asymmetries between the horizontal and vertical designs to optimize physical fit.

The primary function of the mainspring that suspends the proof mass is to apply a restoring force to the mass in the direction of its center position, so that when the mass is deflected by some acceleration acting on it, the spring acts in the opposite direction to restore the position of the mass. The spring constant (the force the spring applies per unit of deflection) is generally as weak as can be practically achieved, so as to

ensure the mass may freely move relative to the frame in which it is suspended. However, the mainspring of a vertical sensor must also balance the suspended proof mass against the acceleration of gravity, whereas the mainspring of a horizontal sensor does not. For example, a proof mass of $M = 200 \text{ g}$ in a vertical axis of the type shown in Fig. 1 experiences a downwards force due to gravity of $F_g = Mg = 1.96 \text{ N}$ that applies a torque at the hinge point, and the spring must apply the equivalent counteracting torque to keep the mass balanced at the center position.

Adjusting the sensing axis so that the boom is positioned at the center position of the displacement transducer is known as “mass centering.” This is a mechanical operation which is done for some seismometers by the operator turning an adjustment screw, and for others by microprocessor-controlled motors within the seismometer making the adjustments automatically, initiated by an external command or signal. On a horizontal axis, this is often done by tilting the axis (e.g., Guralp CMG-3T, Streckeisen STS-1) one way or the other within a range of a few degrees, which allows gravity to act sideways on the axis in proportion to the sine of the tilt angle (which for small angles is directly proportional to the tilt angle). If the mass is decentered either because the spring is applying a decentering force or because the seismometer is tilted off level, applying a small amount of gravity to one side or the other of the axis by tilting it in the opposite direction deflects the mass back to its center position. This method is depicted in Fig. 3, showing a horizontal axis on its internal mounting plate in an off-level situation. In Fig. 3a, gravity has pulled the boom off-center and its two capacitor plates are not centered about the middle capacitor plate that is fixed to the frame. In Fig. 3b, the axis has been tilted mechanically to re-center the boom mass. Note that the direction of sensitivity changes as the axis is tilted.

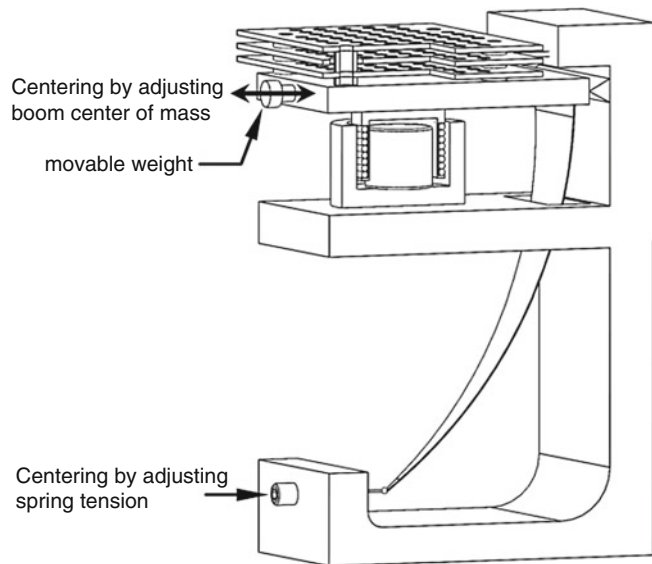
Tilting a vertical axis to center the mass is not practical because the effect of gravity on the axis is proportional to the cosine of the tilt angle, which for small angles has almost no effect: a vertical axis would have to be tilted by 10.7°



Symmetric Triaxial Seismometers, Fig. 3 Mass centering a horizontal axis by tilting. (a) Tilted axis on unlevel ground. (b) Axis internally tilted to re-center mass (© Nanometrics)

Symmetric Triaxial Seismometers,

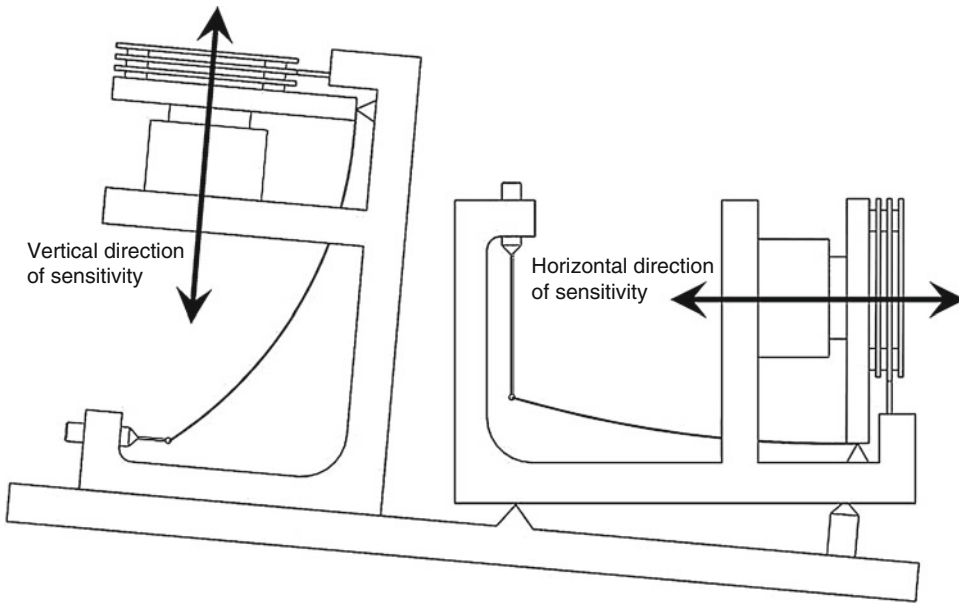
Fig. 4 Two methods for mass centering a vertical axis (© Nanometrics)



to counteract the same deflection as for a 1° tilt of the horizontal axis element. Instead, the mass centering for a vertical axis is done in one of two ways: by adjusting the tension of the main spring to change the force applied to the boom (e.g., Guralp CMG-3T) or by changing the position of the center of gravity of the hinged proof mass by moving an adjustable slug horizontally along the mass relative to the hinge point (Streckeisen STS-1). Figure 4 shows how both

methods work in principle. In practice the mechanisms are of course more elaborate and are often motorized. Note that, unlike the method of tilting the axis, the direction of sensitivity does not change as the spring tension is changed or the center of mass of the boom is adjusted.

A non-Galperin three-component broadband seismometer includes two horizontal axes and one vertical axis, usually mounted within a common pressure vessel. Because axis tilting



Symmetric Triaxial Seismometers, Fig. 5 Vertical and horizontal component non-orthogonality (© Nanometrics)

is used for mass centering the horizontal elements but not the vertical element, the mutual orthogonality of the XYZ components is compromised by the degree of tilt applied. This phenomenon is illustrated in Fig. 5, which shows one of the two horizontal axes and the vertical axis side by side on the same tilted base, the two axis elements having been centered by different methods resulting in non-orthogonal directions of sensitivity.

Practical design considerations also tend to encourage differences in implementation between the vertical and horizontal sensors. Asymmetries between the vertical and horizontal sensor elements arise for several practical reasons, despite the principles of operation being much the same. Besides the mainspring designs that must be different (one counters gravity, the other does not) and the significantly different mass centering mechanisms, the physical orientation and aspect height/width ratio of the two types of axis are asymmetrical. As the two horizontal axes and one vertical axis mount side by side on a common base plate (or are stacked vertically for a borehole configuration), such practical details, such as mounting features, wiring harness, and circuit board connectorization,

fit, and placement within the enclosure and other considerations generally lead to substantive design differences between the vertical and horizontal elements.

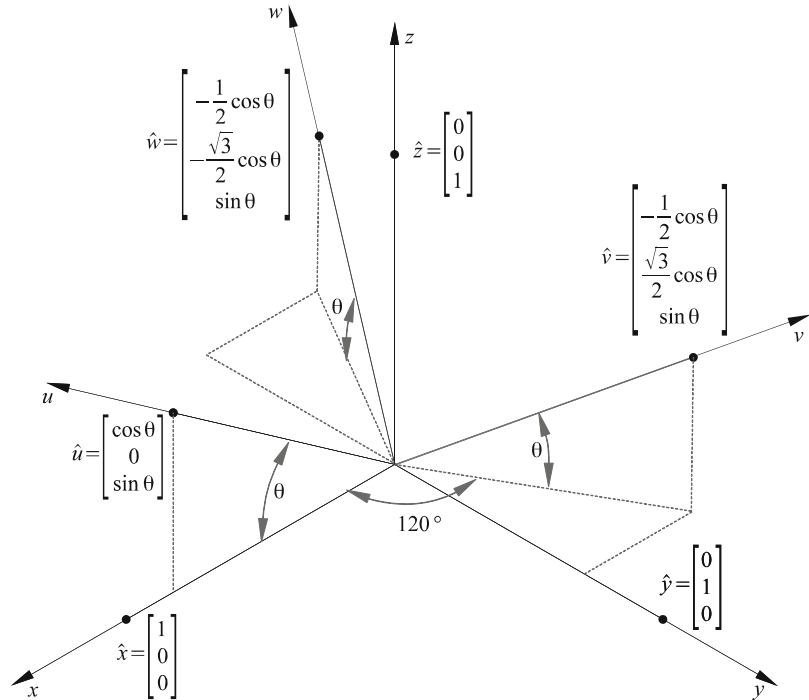
The Galperin Sensor Configuration

An alternative configuration to the conventional XYZ horizontal/vertical arrangement was proposed by Galperin, in which the three sensing elements of a triaxial seismometer are arranged in such a way that permits them to be identical in every respect. This is achieved by rotating the orthogonal sensing axes from the XYZ frame of reference to an orientation where each of the three (still mutually orthogonal) axes is inclined up from the horizontal plane at the same angle.

Figure 6 illustrates a symmetric triaxial arrangement. The directions of sensitivity of the new U, V, and W axes are shown overlaid on the XYW coordinate system. In this orientation, the edges (or directions of sensitivity, or axes) are often given the designations U, V, and W (Wielandt 2002), creating a new “UVW” coordinate system. The projections downwards of the three UVW axes onto the horizontal plane are lines radiating from the center equally spaced 120° apart. This arrangement is called symmetrical

Symmetric Triaxial Seismometers,

Fig. 6 Isometric view of UVW and XYZ coordinate systems (© Nanometrics)



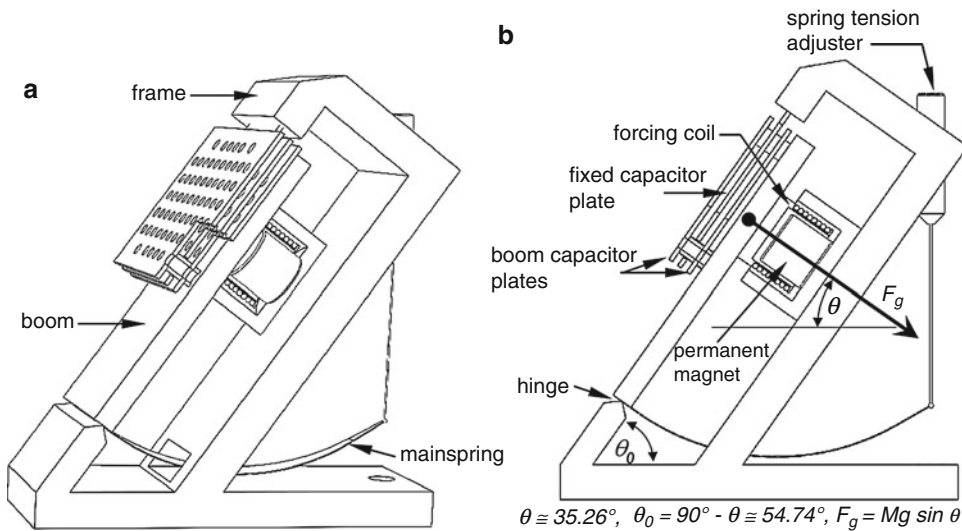
because each axis “sees” the same proportion of gravitational acceleration and this allows the axes to be constructed identically. It is possible to choose any upwards inclination angle θ for the three axes for this arrangement to be symmetrical (such as 45°), but setting $\theta = \tan^{-1}\left(\frac{1}{\sqrt{2}}\right) \cong 35.26^\circ$ also makes the UVW axes mutually orthogonal.

The UVW system is then a simple rotation from the XYZ. This can be visualized by thinking of the XYZ directions of sensitivity as the three edges of a cube radiating from one of its eight corners, where the cube is resting on a horizontal surface. The X and Y directions are along two perpendicular edges of the cube’s bottom face, and the Z direction is the vertical edge of the cube that joins the same corner where the X and Y edges meet. Now visualize the cube being tilted upwards with only the common corner resting on the flat surface, and balance it so that the topmost corner of the cube is directly above its bottom corner. The edges that meet at the bottom corner are still of course mutually orthogonal but now form the same angle upwards with respect to the flat surface base.

The trigonometry of this “balanced cube” arrangement results in the angle formed between any of the UVW axes and the horizontal plane being $\theta = \tan^{-1}\left(\frac{1}{\sqrt{2}}\right) \cong 35.26^\circ$. The complementary angle with respect to the vertical is then $\theta_0 = 90^\circ - \theta \cong 54.74^\circ$.

Figure 6 employs the convention used in Nanometrics Trillium seismometers whereby the projection of U in the XY plane points in the X direction, and the UVW system is right-handed like the XYZ coordinate system (Nanometrics Inc. 2013). The convention for Streckeisen seismometers is that the projection of U in the XY plane is antiparallel to X and that the UVW coordinate system is left-handed (Streckeisen and Messgeräte 1995).

Note that the term “symmetric triaxial” is not in itself sufficiently specific to fully define the Galperin configuration, as any triaxial orientation with horizontal projections spaced 120° apart and where each axis of sensitivity forms the same angle with respect to vertical (say, e.g., 45° instead of 54.74°) is a symmetric arrangement. Each axis still experiences gravity equally and they are arranged symmetrically with respect to



Symmetric Triaxial Seismometers, Fig. 7 Oblique axis construction suitable for a symmetric triaxial seismometer. (a) Perspective view. (b) Side view (© Nanometrics)

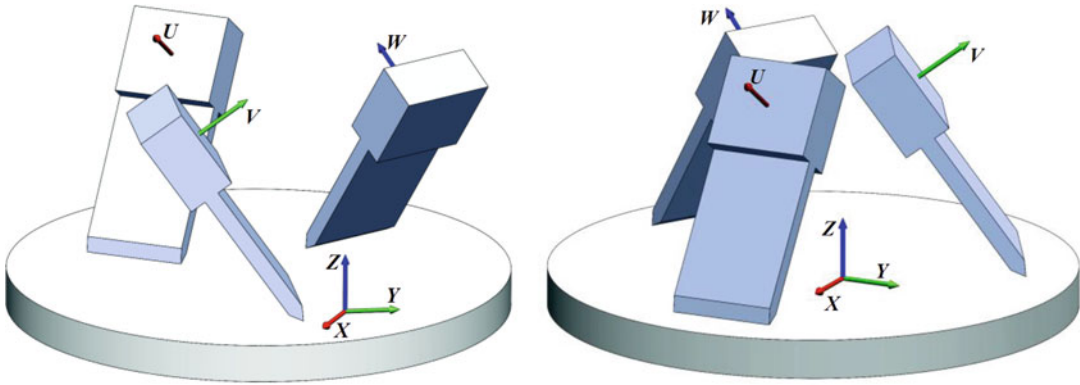
each other, but they no longer meet at right angles as the edges of a cube. The Galperin topology requires the three axes to be mutually orthogonal as well as having the same angle with respect to the vertical direction, which then fully constrains their mutual orientation. Nevertheless, the “symmetric triaxial” designation is usually assumed to refer to the more specific Galperin topology.

Figure 7 shows conceptual mechanical diagrams of the structure of an axis suitable for a symmetric triaxial seismometer, showing the boom set at an oblique angle, but otherwise based on the same principle as a vertical axis, with the mainspring suspending the weight of the boom against gravity ($F_g = Mg \sin \theta$). For the specific case of a Galperin axis set at a $\theta \cong 35.26^\circ$ angle, $\sin \theta = \frac{1}{\sqrt{3}}$. Because each axis experiences the same static acceleration due to gravity, $\frac{1}{\sqrt{3}}g_0$, the mainspring counterbalancing gravity is identical on all three axes. Because the direction of sensitivity of each axis is inclined by the same angle $\theta_0 = 90 - \theta$ with respect to the vertical, the physical design of the axis can be made identical for all three UVW component directions.

The directions of sensitivity with respect to the X or East direction are of course different for each axis – they are equally spaced 120° apart

as shown in Fig. 6 – but that is achieved by arranging the three identical sensing elements next to each other on the same horizontal base and pointing them in different but equally spaced directions as shown in Fig. 8. This is typically how vault seismometers are configured internally. Seismometers for borehole or posthole installations usually have the three axis elements stacked in a vertical column, one above the other pointing in different directions, but the principle is the same. The diagrams in Fig. 8 illustrate that there is more than one suitable way to dispose the three Galperin axis elements within a seismometer. Note that the relative directions of X, Y, Z, U, V, and W are the same in both arrangements; the three axis elements have just been translated to different positions on the base. The leftmost arrangement of axis elements is employed in the Nanometrics Trillium 120P vault instrument, for example, while the Streckeisen STS-2 vault instrument employs an arrangement similar to the one shown on the right.

Mass centering of a Galperin-type axis may be achieved by any of the methods used with horizontal or vertical axes: tilting the axis, adjusting the mainspring, or manipulating the center of gravity of the boom. Examples of each method



Symmetric Triaxial Seismometers, Fig. 8 Two Galperin-type axis arrangements for a vault instrument (© Nanometrics)

used on Galperin-type instruments are the Streckeisen STS-2 vault instrument that moves a slug on the boom, the various models of Nanometrics Trillium observatory-grade seismometers that adjust the tension on the main-spring, and the Geotech KS54000 borehole seismometer that tilts each axis in its vertical stack.

Mutually orthogonality UVW (and therefore XYZ) directions of sensitivity are maintained because the masses of all three axes are centered using the same technique, taking advantage of the symmetry of the design. Unlike in Fig. 5 that shows how a vertical and horizontal axis can become mutually non-orthogonal, the Galperin designs do not adjust one axis using a different technique than the others. For example, the direction of sensitivity of the Galperin axis depicted in Fig. 7 is fixed by its geometry to be at 90° to the face of the boom. Adjusting the tension of the spring to re-center a deflected mass does not alter the direction of sensitivity, and because the three axes are permanently fixed to a common mounting plate as in Fig. 8, the three UVW directions remain mutually orthogonal.

The commonly desired XYZ horizontal/vertical signals are readily derived from the UVW signals using a vector algebraic transformation:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} 2 & -1 & -1 \\ 0 & \sqrt{3} & -\sqrt{3} \\ \sqrt{2} & \sqrt{2} & \sqrt{2} \end{bmatrix} \begin{bmatrix} u \\ v \\ w \end{bmatrix}$$

The inverse transformation derives UVW from XYZ:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \frac{1}{\sqrt{6}} \begin{bmatrix} 2 & 0 & \sqrt{2} \\ -1 & \sqrt{3} & \sqrt{2} \\ -1 & -\sqrt{3} & \sqrt{2} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

The transformation to the conventional XYZ frame of reference may be done at any point, such as by seismic analysis processing software just before the signals are plotted or used as input to seismic analysis algorithms that assume the XYZ reference frame. However, as a convenience to the operators of seismic signal data centers and the users of the seismic data, most Galperin-type seismometers effect the signal transforms within the instrument by summing the UVW signals in the required proportions using a precision analog mixing circuit. Some seismometers (e.g., Nanometrics Trillium models) allow the operator to optionally configure the instrument to output raw UVW instead of the mixed XYZ if desired, though this is primarily used for instrument troubleshooting rather than for routine recording of seismic data.

Benefits and Drawbacks of the Galperin Configuration

The choice between a Galperin and a conventional configuration is usually regarded as a secondary consideration in specifying an

instrument, as the internal topology of the seismometer does not of itself dictate the performance or reliability of an instrument. However, while either configuration is capable of achieving reliable high performance, the symmetric triaxial configuration does offer some significant benefits to users. The two topologies have distinct characteristics that are useful to consider when selecting, installing, operating, or troubleshooting a broadband seismometer.

Commonly Cited Trade-Offs

Response Matching

Because the designs of a horizontal axis and a vertical axis differ significantly, it requires special effort in design and diligence in manufacture to match the transfer functions of the two different constructions, if they are to provide the same response to ground motion in terms of passband sensitivity, lower corner frequency/phase response, and the upper corner response. Indeed, there is no constraint other than designer or user preference that requires the response of the vertical to match the response of the horizontal component; the two could be markedly different with the manufacturer providing the specifications for each component by giving, for each axis, a frequency/phase response plot or a set of poles and zeros and passband sensitivity. In contrast, because the Galperin configuration requires its cardinal XYZ outputs to be derived as weighted sums of all the UVW components, the design and the manufacturing must ensure that all three components are very closely matched; otherwise the transfer functions of the XYZ outputs become complicated. The symmetrical nature of the Galperin topology helps produce matched responses because the three axes are constructed identically. Manufacturers of Galperin seismometers supplement this advantage with precision machining, trimming of electronics values, and other calibration measures to precisely match axis responses. The precision matching of U, V, and W components in a Galperin design provides assurance to the user that the X, Y, and Z transfer functions are likely to be consistent and the same as any of the U, V, or W transfer functions.

Response matching is easily tested on a Galperin instrument by providing a vertical signal on a high-quality vertical shake table or by injecting the same calibration signal into all three UVW components simultaneously (which is equivalent to providing a calibration input in the vertical direction). The output should be pure vertical; the extent to which signal is present on the X and Y outputs is a measure of response difference between the UVW elements. By sweeping the frequency of the input and recording the vertical and horizontal outputs, the quality of the response matching can be measured across the frequency band.

In contrast, subjecting a conventional XYZ design seismometer to a pure vertical stimulus tests only the vertical axis and does not provide any insight into how well the responses of the three axis components are matched.

Manufacturing Trade-Offs

Some aspects of the design and manufacture of Galperin instruments are simplified or facilitated, while others are made more complex. Obvious advantages to the manufacturer derive from having only one electromechanical axis design within a seismometer. That means there are fewer unique part types to manage, only one style of axis to build and test, and the identical axes can be readily swapped for troubleshooting or repair purposes. On the other hand, a symmetric triaxial seismometer requires a UVW-to-XYZ analog electronic mixing circuit that is not needed in an XYZ design, adding somewhat to the cost of manufacture.

Because the vertical output signal of a symmetric triaxial design is an equally weighted sum of the UVW components, a clean vertical signal provides complete assurance that all three components are well behaved. As the horizontal signals at most surface sites are usually noisier than the vertical due to tilt effects, especially at long periods, it is often feasible to test and qualify the vertical channel but more difficult to do so for the horizontals. This permits the manufacturers of Galperin designs to reliably test the performance of all three sensor elements by qualifying the vertical signal, providing

assurance to the users that all three outputs meet the published specifications. In contrast, the X and Y components of a conventional design cannot be assessed by examining the vertical channel signal.

Reliability Trade-Offs

An often-cited drawback relating to the Galperin configuration is that all three UVW components must be functional and well behaved for any of the XYZ outputs to work (Graizer 2009). It is held that this contrasts with a conventional seismometer where any one of the three components could fail without compromising the other two. This can be seen as a corollary of the fact that a good vertical output from a Galperin configuration provides assurance that all three sensor elements are good.

The spontaneous catastrophic failure of just an individual axis element within an operating seismometer is relatively uncommon. It is more likely for catastrophic failure to occur or be evident at installation time (in which case most operators would choose not to deploy even if there was a good vertical still operating) or for a failure to occur in a way that affects the entire system, such as an electronics fault.

Many seismometers with Galperin configurations provide the ability to remotely choose between UVW and XYZ outputs. In this case, if a failed axis is detected and a site visit is not feasible, high-quality biaxial data can still be recorded after the flip of a switch.

It is possible for an element to develop noise spontaneously, and in this case the conventional design has the advantage: a Galperin design would mix the noisy channel into its vertical output (where it is most easily detected), whereas the conventional XYZ design would only manifest the noise on the output associated with one degraded axis.

A manufacturer can minimize the likelihood of excessive channel noise by ensuring the seismometer self-noise is tested before leaving the factory and by designing the instrument for long-term operation to ensure performance is maintained through the instrument's operating life.

Other Benefits of a Symmetric Triaxial Design

Using the UVW Orientations to Discriminate Noise Sources

The symmetric triaxial configuration makes it possible to distinguish phenomena that occur within one axis element from those that occur independently of one axis element. The operator can use this to advantage, helping diagnose and resolve problems associated with the environment, installation, ancillary equipment, or the seismometer itself. This is discussed in greater detail in the section "Using the Galperin Topology to Help Discriminate Noise Sources" below.

Assuring the Mutual Orthogonality of Component Signals

The three identical and symmetrically arranged sensor elements of the Galperin configuration are fixed to be mutually orthogonal during manufacture. A mass centering operation, whether moving a slug to shift the boom center of gravity or adjusting mainspring tension to return the boom to center, does not alter the mutual orthogonality of the directions of sensitivity. In the first case the sensitivity direction of each axis is shifted by the same amount and so they remain mutually orthogonal. In the latter case, the tension of the main spring does not alter direction of sensitivity and so orthogonality is also preserved. However, a conventional XYZ design tilts the horizontal axes but adjusts the mainspring tension for the vertical, causing the horizontal directions of sensitivity to change relative to the Z axis. This causes the horizontal components to shift relative to the vertical by an angle equivalent to the tilt of the seismometer housing itself. Effectively, the vertical of an XYZ design remains aligned to the housing while the horizontals remain perpendicular to gravity.

Using the Galperin Topology to Help Discriminate Noise Sources

Using a seismometer that has a Galperin design can facilitate identifying noise sources by showing whether they are specific to one axis element

or not. This section discusses noise sources and some methods for troubleshooting noise artifacts that use the characteristics of the symmetric triaxial configuration to advantage.

Noise Sources

A noise artifact is usually apparent only once it is in the seismic record. This record is furthermore the sum of all effects applied by the entire system including the environment, site, seismometer mount, the seismometer itself, cables and connectors, the digitizer, and signal transmission and post-processing. [In this context the “seismometer mount” is a general term that refers to the structure the seismometer is mechanically coupled to. That may be a concrete pier, a gimbal mount such as for an ocean bottom system, surrounding granular substrate as in the case of a buried instrument, a metal bracket affixed to a building structure, and others.] All these components of the end-to-end system must be considered when investigating noise or other signal impairments.

Noise may be generated by processes within the seismometer or by external processes acting on the seismometer. Some noise is expected or unavoidable, such as the manufacturer’s characterized self-noise, or may be indicative of instrument defects such as persistent spurious transient noise event (“pops”) or installation deficiencies such as thermally driven noise or the seismometer shifting or tilting. It is first useful to review some typical sources of internal and external noise.

The sources of noise (defined as any type of additive unwanted signal) in a seismic record include inherent instrument noise (self-noise), environmental effects, excessive instrument noise, installation-driven noise, and ancillary equipment noise. A seismometer’s self-noise is usually specified by the manufacturer, typically as a power spectral density (PSD) plot, and this establishes a performance baseline for an installation. Good installation practice will include mitigating sources of environmental noise as much as is practical. Excessive instrument noise is defined as noise originating within the seismometer in addition to its published self-noise characteristic and may represent variability in the

manufacturing, defects, or failures. Installation-driven noise is defined as a non-seismic signal that originates from poor or defective installation practices. Lastly, noise may also be added by downstream ancillary equipment such as the digitizer.

Environmental and Installation-Driven Sources of Noise

Seismometers respond in varying degrees to many stimuli besides seismic motion, and while seismometer designers try to maximize the immunity or insensitivity of the instrument to everything except translation ground motion, this is difficult to achieve. The installation must also be designed to maximize the effectiveness of the seismometer. An inadequate installation of even the highest performance seismometers can yield very poor results, showing high noise levels and susceptibility to non-seismic environmental inputs. There is therefore significant benefit in tools and techniques that help diagnose installation deficiencies and distinguish them from instrument problems.

Environmental influences on a seismometer that may cause unwanted output signal (noise) include temperature changes, seismic wave-induced tilt, Earth’s magnetic field or fields generated from local equipment, wind-induced ground motion, atmospheric pressure changes, electrical interference induced on the output signal, and other effects. Many environmental influences can be mitigated, such as by providing a temperature stable vault and locating the installation far from vertical structures such as trees that conduct wind noise into the ground. Some are more difficult or even impossible to eliminate, such as apparent vertical accelerations induced by changes in atmospheric pressure, because the apparent acceleration due to gravitation cannot be distinguished from other kinds of acceleration and is relatively insensitive to the depth of the site (Zörn and Wielandt 2007).

Installation-driven noise sources are those that induce motion on the seismometer due to defects in the installation itself. A common example is poor mechanical coupling of the instrument to the ground, which may allow the seismometer to

move relative to the structure being monitored. This can be due to poor mounting, inadequate substrate, poor coupling to the ground, improperly locked mounting feet, or cables applying strain on the seismometer. Such problems most commonly produce tilt noise, but vertical noise is also possible, such as for a seismometer “bouncing” up and down on adjustable feet that have not been tightly locked. Another example is convection-driven air currents within a vault inducing low-frequency horizontal tilt noise. Such noise is produced by slow-moving air currents acting on the seismometer to expand or shrink the sides of its case unevenly and thus tilting it.

A seismometer exhibiting excessive response to environmental stimuli may be indicative of an instrument defect. For example, a leak in the pressure vessel that allows atmospheric air pressure changes to pump air in or out of the seismometer will manifest as a non-seismic vertical signal as the buoyancy of the proof masses responds to internal air density changes. This is most commonly seen in instruments that have access ports that an operator may open, such as to insert a screwdriver for manual mass centering, and that may not be properly closed or where the seals may have degraded.

Sources of Internally Generated Seismometer Noise

All seismometers generate noise internally, that is, they would produce a signal even if the ground was not moving at all. This inherent noise is characteristic of the design and is usually specified in some detail by the manufacturer, although in the case of passive seismometers, the manufacturer will instead usually provide a few parameters from which the self-noise can be estimated. Noise is inherent in electronics circuits due to a multitude of phenomena in conductors and semiconductors, such as thermal noise and flicker noise. The suspended mass and spring has inherent noise due to the Brownian motion of air molecules buffeting the mass, air damping the boom motion, as well as other energy-loss mechanisms. The summation of these effects produces a stationary noise spectrum that is commonly

referred to as the seismometer’s “self-noise,” often represented as a PSD graph and commonly shown together with the New Low Noise Model (NLNM) and New High Noise Model (NHNM) (Peterson 1993). Reputable manufacturers provide this information for each model and produce instruments with performance consistent with their published specifications.

Excessive instrument noise is internally generated noise that exceeds the self-noise specification of the seismometer. There are three broad classes of excessive internal noise:

1. Stationary broadband noise that significantly exceeds the instrument’s published self-noise specification
2. Nonstationary transient noise events, such as pops or spikes
3. Stationary narrowband noise, such as tones or oscillations

A common technique for measuring noise is frequency domain analysis, such as plotting the acceleration power spectral density (PSD) of the seismic data over some time period. While tones are readily distinguishable from broadband noise, nonstationary events such as pops have a broadband spectral characteristic (typically proportional to $1/f^2$) that can mislead the troubleshooter. However, small pops may not be readily identified by examining the time-domain time-series data as they may have amplitudes too small to be distinguished from the background seismic activity.

The Galperin Topology and Noise Source Identification

Being able to demonstrate that a noise phenomenon is peculiar to one particular axis element and not the other two, or conversely that a noise is registered by more than one axis, can be a powerful technique to help isolate and identify noise sources and causes. The Galperin topology facilitates this because its sensor reference coordinate system (UVW) is rotated from the vertical/horizontal (XYZ) coordinate system, providing a powerful way to distinguish instrument noise

Symmetric Triaxial Seismometers, Table 1 Channel characteristics of some possible noise sources (© Nanometrics)

Channels showing non-seismic noise	Potential noise sources						
	Axis mechanics or electronics	Tilt motion	XY oriented seismometer mount (e.g., gimbal)	UVW-XYZ conversion electronics, cable, digitizer	Varying temperature or pressure leaks	Varying magnetic fields	Power supply, electrical/magnetic interference
U, V or W only	Probable						
X or Y only		Possible	Possible	Possible			
Z only				Possible	Possible		
X and Y in some proportion but not Z		Probable					
Equally on all X/Y/Z channels				Possible			Possible
Unequally on X/Y/Z channels						Possible	Possible

sources from environmental or installation phenomena. Many external noise phenomena have a characteristic direction – often horizontal or vertical – and in any case are unlikely to align with one of the U, V, or W directions. Likewise, noise that is associated with a U, V, or W direction provides a clear indication that the noise source is internal and associated with that axis.

Table 1 provides general guidelines for narrowing down the sources of an observed noise signal and provides a starting point for troubleshooting. This presumes the operator is able to perform the essential basics of signal analysis: conversion from XYZ to UVW domains, plotting time series and frequency domain plots such as a PSD, removing instrument frequency response, and the like. Table 1 also presumes the seismometer has a Galperin topology.

Once a noise artifact is recognized and classified (e.g., stationary broadband, nonstationary transient, or stationary narrowband) and characterized as much as possible, the next step is to determine whether it consistently manifests on a specific X, Y, Z, U, V or W channel or a particularly suggestive combination of these channels.

This is equivalent to asking which direction the signal represents. Signals specific to X, Y, or Z suggest East/West, North/South, and up/down motions or phenomena that mimic these. Signals specific to U, V, or W suggest motions in those axis directions or, more likely, phenomena within the axis mimicking this direction. A signal that manifests on X and Y but not (or very little) on Z represents motion in a horizontal direction that is not East/West or North/South, or a phenomena such as tilt that mimics that. A signal that is equal on X, Y, and Z is more likely to be a common-mode electronics noise problem than a real motion in the direction represented.

Because it is unusual to have real seismic motion act exactly along any of these X, Y, Z, U, V, or W directions, signal artifacts with these characteristic directions are more likely to be of non-seismic origin. An exception could be a site expected to have motions exactly along a specific direction, such as a seismometer located near the top of a swaying structure, oriented so that its X channel (say) is aligned with the direction of sway.

Having narrowed the potential sources, it is usually possible to make further deductions from

the character of the signal artifact observed. The next sections discuss typical phenomena and their classifications according to possible sources listed in Table 1.

Axis Mechanics or Electronics: UVW Channel Noise

Noise artifacts apparent on one U, V, or W channel and not on the others point squarely to a noise source associated with that axis. Of course this can only be recognized by transforming the seismic data into the UVW domain to observe that a noise previously apparent on all XYZ channels is suddenly associated with just a single one of the Galperin axis elements. Noise that is vertical (Z only) or horizontal (some combination of X and Y) cannot have a source within a Galperin-type axis.

A noise event commonly attributed to highly sensitive broadband seismometers is that of pops, otherwise known as spikes or steps. These are sudden sharp excursions in the seismic waveform that appear as if there had been a sudden step in acceleration. This can be the result of spontaneous movements at a microscopic level within the axis mechanism or sudden shifting of the seismometer itself due to installation or mounting problems. If a pop is evident only on one U, V, or W axis, it is due to spontaneous movement within the axis, such as some mechanical stress or strain suddenly relaxing. Inadequate design, assembly flaws, component defects, or rough handling can give rise to excessive rate of pops in an axis. One indication of excessive axis pops is when the PSD is plotted for the U, V, and W signals and there is a well-defined $1/f^2$ spectrum in the lowest frequency band of one of the axes. It is well known that seismometers are more prone to pop noise when first manufactured, installed in a new location, or acclimating to a new environment as stresses are relieved.

Excessive broadband axis noise (specific to a U, V, or W axis) can be due to electronics noise, for example, excessive current noise due to components not meeting specification.

Excessive axis-specific narrow band noise is less common, but can occur if there is excessive

loop gain or insufficient phase margin in the control loop of one axis. This would generally suggest a manufacturing or design defect and could also be the consequence of component failure or rough handling.

A signal impairment specific to one U, V, or W axis such as zero signal or railed signal suggests an axis failure, such as failure to center, or a failure of the electronic circuitry associated with that axis.

Likewise, a non-seismic noise or other impairment (such as zero signal or railed signal) that is not limited to a U, V, or W channel but manifests as a horizontal or vertical signal cannot be an axis problem but must have some other origin, as indicated in Table 1. The following sections discuss these other potential noise origins.

Tilt Effects: XY Channel Noise

The most common sources of noise are installation deficiencies and environmental stimuli, often where inadequate installation fails to counter environmental drivers such as temperature changes or air currents. Some of these effects are quite subtle but can create significant noise artifacts. It is quite common to see horizontal noise levels for a seismic station be significantly greater than the vertical noise floor (Hanka 2002). While horizontal noise could be ascribed to noisier horizontal components in a conventional XYZ design, with a Galperin topology with identical UVW elements, noisier horizontal data can definitively be attributed to a process external to the seismometer.

Because tilt mixes gravitational acceleration into the horizontal channels in direct proportion to the tilt, and because gravity is significant compared with the small accelerations due to background seismicity, even sudden tilts of less than a millionth of a degree will be registered by broadband seismometers and appear as noise artifacts. The effect is far less significant for vertical because gravity mixes to the vertical in proportion to the cosine of the tilt angle, which is a near-zero effect for small angles.

Seismic waves traveling on the surface of the Earth that produce vertical translation motion (like a cork bobbing on the waves of an ocean)

also cause tilt: one can visualize a raft on the same ocean that tilts one way as it rises up one side of the wave and then tilts the other way as it falls. The rise/fall is vertical motion the seismometer will faithfully record, but the tilt induces a small varying sideways component of gravity producing a noise signal on the horizontal channels in addition to the real horizontal motion. Because tilt motion is inherent in seismic motion at the Earth's surface, horizontal tilt noise cannot always be mitigated. However, excessive tilt noise and in particular sudden discrete tilt events are usually indicative of installation problems that should be addressed.

Noise artifacts that appear on X and/or Y but not Z are horizontal and often represent the seismometer tilting in some manner. There are many potential causes and sources of tilt, and so it is useful to further characterize the noise. Sudden isolated transitions are suggestive of rapid discrete tilt events, oscillations can indicate rocking behavior, and broadband horizontal noise may indicate continual random tilting. Steps that occur with a consistent polarity suggest a series of tilts in the same direction, as may happen if the seismometer is slumping in discrete steps to one side. Steps that alternate polarity may indicate a back-and-forth tipping action. The direction of the tilt is indicated by the proportion of X-to-Y noise, and some combination of X and Y (with very little Z) noise is a further confirmation of probable tilt action.

Atmospheric pressure changes acting on a sealed vessel such as a seismometer will deform the vessel to some extent, and because tilts much less than 10^{-6} of a degree can generate noise, even miniscule deformations can matter. Seismometer designers generally take great care to ensure those deformations do not act to tilt the internal sensing elements; otherwise horizontal noise driven by weather changes would be inherent in the instrument (Widmer-Schmidrig and Kurrel 2006). In a well-designed instrument, pressure-driven tilt should not be evident. It is also difficult for a manufacturing or component defect to give rise to pressure-tilt sensitivity; the design either has sufficient pressure immunity or it does not. The inherent symmetry of a Galperin-

type topology also provides more opportunity to the designer to optimize pressure vessels with highly symmetric internal mounting arrangements that attenuate pressure vessel deformation effects.

XY-Oriented Seismometer Mount: X or Y Noise

For Galperin-type seismometers, tilt that is aligned strictly with either the X (East) or Y (North) direction is of special interest: while it is possible for tilt to coincide with the X or Y, there may be a specific reason for tilt to occur in a well-defined and aligned orientation. For example, an unlocked foot that is positioned in line with East or North could be suspected, or the seismometer may be mounted on some structure that may tend to tilt in a particular direction. With a conventional XYZ seismometer, X- or Y-aligned noise could either be related to the internal X or Y axes or be due to external tilting. A symmetric triaxial construction experiencing such noise clearly points to an external cause, and when the tilt direction is highly aligned with X or Y, or in some other specific direction in line with a particular mounting feature such as a foot, it provides a useful indication of the likely trouble spot.

A special instance of this is when a Galperin-type seismometer is mounted within a moveable platform, such as a leveling gimbal for ocean bottom deployment. If tilt events are clear in the data and strictly oriented in line with the potential movement of the platform, it is suggestive of gimbal platform movement.

Output Stage Electronics, Cable, and Digitizer: X, Y, Z, or XYZ Noise

Single X, Y, or Z Channel Noise

The only electronics internal to a typical symmetric triaxial seismometer which are specific to a single X, Y, or Z output channel are the UVW-to-XYZ coordinate-conversion circuits that combine the UVW signals to produce XYZ outputs, the signal output drivers, internal cabling, and external connector. It is rare to see noise problems originate from these areas, but it

is possible for the failure of an electronic component to cause a failure of one output channel (or one side of a differential output channel that then would manifest itself as a half-amplitude signal on one channel).

Because the signal is in the XYZ domain from the seismometer onwards, the cable and digitizer must be suspected when troubleshooting. Because the output stage electronics are relatively simple and usually reliable, it is more common for the cause of single-channel or common-mode XYZ failures or noise to be associated with the downstream digitizer or the cable connecting the seismometer to the digitizer.

It is useful to determine if noise apparent on a single output channel is predominantly or exclusively evident on that channel. If there is no attenuated version of that noise on the other channels, it is likely to represent a failure or deficiency of hardware associated with that channel, such as a connector pin or digitizer channel problem. Noise appearing predominantly on one output channel but also on the others in an attenuated form points to an effect predominantly aligned with that channel but not exclusive to it, such as tilt along the X direction.

Predominant But Not Exclusive Z Channel Noise

A special case of single-channel noise is Z channel noise. That is because the Z (vertical) channel is produced by an equally weighted sum of the UVW axis signals, and so any phenomena that affects all axis elements equally will manifest predominately on the Z channel. Therefore, Z channel noise could originate from within the seismometer due to a common-mode effect acting on all axis elements or from an external fault on the Z channel in the cable or digitizer. One can usually determine which by the nature of the noise or impairment and also by determining if the noise or impairment is exclusively or just predominately on the Z channel. A common-mode effect acting on all axis elements will predominately manifest on the Z channel, but usually there are slight differences in sensitivity to these effects from axis to axis, so some attenuated version of that noise on the X and

Y channels may be present. If the impairment is external, such as a noisy digitizer Z channel, it is much less likely to leak into the X and Y channels.

Common-Mode XYZ Channel Noise

A noise source or impairment that affects all output channels more or less equally (known as a common-mode characteristic) could be associated with the seismometer electronics that serve common functions such as power conversion and control circuits, but is more likely an external common-mode effect, such as noise induced on all input channels to the digitizer by electrical interference. Common-mode XYZ cannot be problems with any or all of the internal Galperin axis elements, because these would manifest in the U, V, W, or Z directions, respectively, never in the direction corresponding to equal XYZ signals.

Temperature Changes and Pressure Leaks: Z Channel Noise

The most common sources of predominately Z channel broadband noise are environmental effects acting on the seismometer, such as varying temperature or pressure leaks. Seismometers respond to temperature changes in part because thermoelastic effects act on the spring that suspends the proof mass against gravity, causing apparent changes in acceleration. These are slow effects and therefore affect low-frequency noise performance. The three axis elements will usually have a similar but not exactly equal sensitivity to temperature changes, manifesting as predominately Z channel noise with much smaller effects on X and Y.

A leak in the pressure vessel of the seismometer would allow air to pump in or out driven by atmospheric pressure changes, and this would cause the proof masses to rise or fall buoyed by changing air density inside the vessel. This creates a very pronounced noise that is strongly vertical. Pressure leaks mostly arise in seismometers that provide service access ports such as for manual mass centering that may not have been properly closed or where the seals have deteriorated.

Magnetic Field Effects: Unequal XYZ Noise

Broadband seismometers are designed to be as insensitive as possible to changes in external magnetic fields, but some sensitivity is almost inevitable (Forbriger et al. 2010). Springs designed to be relatively insensitive to temperature changes are often fabricated from magnetic materials and so have small but measurable forces applied by the Earth's magnetic field. These can manifest as low-amplitude noise at long periods. Each axis will be sensitive to the direction as well as the strength of the geomagnetic field. Because the direction of the Earth's magnetic field differs depending on location, the effect of changing magnetic fields is not equal on all axis elements, nor are the relative proportions among axis elements consistent or predictable. These are typically small effects and usually only of concern for very high-quality low-noise sites with very low-noise seismometers. Mitigation strategies can include augmenting magnetic shielding.

Electronic Interference: Equal or Unequal XYZ Noise

Noise that appears on all channels may arise from interference of nearby electrical equipment. A common-mode noise appearing equally (or nearly equally) on all channels could indicate a conducted or radiated noise source from a power source such as a battery charger cycling, a nearby motor, or similar problem. The interference is unlikely to be affecting the seismometer itself, as noise induced into the three Galperin axis elements manifests as predominately vertical (Z) channel signals. Induced noise affecting all channels is likely to suggest a cable or digitizer vulnerability or issue of some sort. Mitigation typically includes determining the source if possible, improving shielding or grounding, and changing cables, digitizers, or power sources if faults are suspected.

Summary

A seismometer that is designed using a Galperin arrangement has three identical sensing elements with their directions of sensitivity arranged to be mutually orthogonal, and inclined up from the

horizontal plane at an angle of $\theta_0 \cong 35.26^\circ$. This is a subclass of the more general “symmetric triaxial” configuration in which three axis elements are identical and symmetrically arranged with respect to each other. The signals from the three sensors, designated U, V, and W, are usually remixed by analog circuitry within the seismometer to provide a vertical and two horizontal outputs equivalent to a conventional XYZ seismometer. The benefits of the symmetric triaxial approach pertain to site and instrument troubleshooting, channel response matching, ease of manufacturing and testing, reliability, and channel orthogonality. The most frequently cited drawback is that all three axis elements must be functional for any of the three XYZ seismic signal outputs to be valid. The high reliability of the best seismometer designs largely mitigates this concern, and it is furthermore avoided by choosing a seismometer which allows the user to remotely select UVW signals to be output in place of the mixed XYZ in the event of a single axis failure. The second cited drawback inherent in the symmetric triaxial configuration is the requirement for an internal analog electronic mixing circuit to convert UVW to XYZ, although in a well-designed seismometer, this is invisible to the user.

A primary benefit of the Galperin approach is that, because the UVW coordinate systems are different from the XYZ horizontal/vertical reference system, noise sources and faults within an axis are distinguishable from problems that manifest in primarily vertical or horizontal directions. A corollary to this is that a properly functioning vertical channel on a Galperin instrument is a reliable indication that all three UVW elements are good, since the vertical is an equal sum of the three Galperin axes. Response matching is necessary for Galperin instruments to enable proper remixing to XYZ outputs, providing assurance of consistent response independent of the direction of motion. Galperin designs also generally assure rigid orthogonality of the XYZ outputs, whereas the tilting of the horizontal elements in the mass centering operation of conventional designs can produce deviations in X and Y orientation relative to the fixed vertical axis. Manufacturing benefits include having one

rather than two axis designs to build and improved ability to test and troubleshoot.

A review of noise sources and characteristics and how their directions align relative to the separate Galperin and horizontal/vertical coordinate systems demonstrates the powerful utility of using a symmetric triaxial configuration to help discriminate and diagnose potential noise problems with the site, installation, or seismometer.

In summary, the Galperin geometry has proven to offer significant benefits to seismometer users and manufacturers and through use and experience has proven to have significant additional benefits in terms of remotely troubleshooting station noise.

References

- Forbriger T, Widmer-Schmidrig R, Wielandt E, Hayman M, Ackerley N (2010) Magnetic field background variations can limit the resolution of seismic broad-band sensors. *Geophys J Int* 183(1):303–312
- Galperin EI (1955) Azimuthal method of seismic observations (in Russian). Gostoptechizdat 80
- Graizer V (2009) The response to complex ground motions of seismometers with galperin sensor configuration. *Bull Seismol Soc Am* 99(2B):1366–1377
- Hanka W (2002) Parameters which influence the very long-period performance of a seismological station: examples from the GEOFON Network, Section 7.4.4. In: Bormann P (ed) *New manual of seismological observatory practice*, vol 1. GeoForschungsZentrum, Potsdam, pp 64–74
- Melton BS, Kirkpatrick BM (1970) The symmetric triaxial seismometer – its design for application to long-period seismometry. *Bull Seismol Soc Am* 60(3):717–739
- Nanometrics Inc. (2013) Trillium posthole user guide. (17217R5). Nanometrics, Inc., Kanata
- Peterson J (1993) Observations and modeling of seismic background noise. Open-file report 93–322, U. S. Geological Survey, Albuquerque
- Streckeisen G, Messgeräte AG (1995) Portable very-broad-band tri-axial seismometer STS-2 manual, 50
- Widmer-Schmidrig R, Kurrle D (2006) Evaluation of installation methods for Streckeisen STS-2 seismometers. <http://www.geophys.uni-stuttgart.de/~widmer/ge2.pdf>. Retrieved 25 Oct 2013.
- Wielandt E (2002) Seismic sensors and their calibration, Chapter 5. In: Bormann P (ed) *New manual of seismological observatory practice*, vol 1. GeoForschungsZentrum, Potsdam
- Zürn W, Wielandt E (2007) On the minimum of vertical seismic noise near 3 mHz. *Geophys J Int* 168:647–658

System and Damage Identification of Civil Structures

Babak Moaveni¹ and Joel P. Conte²

¹Department of Civil and Environmental Engineering, Tufts University, Medford, MA, USA

²Department of Structural Engineering, University of California at San Diego, CA, USA

Synonyms

Damage identification; Nondestructive evaluation; Structural damage; Structural health monitoring; System identification

Introduction

Definitions

Structural damage	Changes in a structural system that adversely affect its performance
System identification	Process of extracting the properties (often dynamic) of a system from measured output or input-output data
Damage identification	Process of detecting, localizing, and quantifying damage in a system
Structural health monitoring	Process of global damage identification in a structural system
Nondestructive evaluation	Process of local damage identification in a structural system

Major structural failures in recent years have brought the public's attention to the urgent need for improved infrastructure monitoring and maintenance. In their latest report card for America's infrastructure (American Society of Civil Engineers 2009), the American Society of Civil

Engineers (ASCE) described an infrastructure that is poorly maintained, unable to meet current and future demands, and, in some cases, unsafe. Expanding and improving structural health monitoring [C1] (SHM) for damage assessment and maintenance is essential for establishing sustainable and resilient civil infrastructure systems and ensuring they can meet the needs of future users. SHM refers to the process of damage identification in a structural system. Damage identification can be performed at four levels: (1) detection, i.e., estimating the existence of damage; (2) localization; (3) quantification; and (4) prognosis, i.e., predicting the remaining useful life of the structure. System identification refers to the process of extracting a system's dynamic properties from the measured data. System identification can be performed to extract damage-sensitive features of a system from its measured data to be used for the estimation of damage (Aktan et al. 1997). It can also be used to “realize” (i.e., identify) a mathematical dynamic model that best predicts the measured output data from the measured input excitation or the statistical correlation structure of the measured output data in the case of output-only measurements. The realized model can then be used to predict the structural response to different deterministic input excitations or the response correlation matrix to different stochastic excitations. Recent advances in sensing and computational power provide an opportunity for improved system and damage identification of large-scale and complex civil structures. The ASCE technical report (Catbas et al. 2013) provides a comprehensive review of the current state-of-the-art approaches, methods, and technologies for effective practices of system and damage identification methods on civil structures. While many researchers have successfully applied various identification approaches to numerical and/or small-scale lab models of civil structures, the literature lacks successful applications to real-world structures in the real loading environment. A recent special issue of the ASCE Journal of Structural Engineering (Moaveni et al. 2013a) was devoted to help bridge this gap.

System Identification

As previously mentioned, system identification refers to the process of estimating the dynamic properties of a structural system from its measured data. System identification methods can be categorized into input–output versus output-only, parametric versus nonparametric, time-domain versus frequency-domain, or linear versus nonlinear methods. Input–output system identification methods also referred to as experimental modal analysis [C2] (EMA) methods extract the dynamic system properties based on measurement data of both the dynamic response and the input excitation. On the other hand, output-only methods or operational modal analysis (OMA) methods [C3] are used when only the measured (ambient) response of a structure is available. OMA methods are suitable when the input excitation is not measureable and can be assumed as a broadband random signal such as wind loads on a building or bridge or vehicular traffic on a bridge. These methods are very useful in situations involving large-scale civil structures which are difficult to excite experimentally.

A system identification method is considered parametric if a mathematical dynamic model (often formulated in state-space) is “realized” in a first step and the dynamic properties of the system estimated from the realized model in the second step. Nonparametric system identification methods directly estimate the dynamic parameters of a system from transformation of data, e.g., Fourier transform or power-spectral density estimation. Time-domain identification methods estimate the dynamic parameters of a system by directly using the measured response time histories, while frequency-domain methods use the Fourier transformation or power-spectral density estimation of the measured time histories. There is also a class of time-frequency methods such as the short-time Fourier transform and the wavelet transform. These methods are commonly used for identification of time-varying systems in which the dynamic properties are time-variant. Linear system identification methods are based on the assumption that the system behaves linearly and

the identified dynamic properties correspond to an underlying linear system. Nonlinear methods [C4] assume specific types of nonlinearities in the system and its dynamic response and provide parameter estimates of the considered nonlinear model class. Nonlinear system identification is a challenging task due to the complex nature of nonlinearities encountered in civil structures and is the subject of ongoing research by many investigators.

A large number of system identification methods have been used successfully for identification of civil structures. Some of the more commonly used system identification methods in the literature include the peak picking method (both input–output and output-only, nonparametric, frequency domain, linear); the frequency-domain decomposition and enhanced frequency-domain decomposition methods (output-only, nonparametric, frequency domain, linear); the class of prediction error methods such as the auto-regressive (AR), auto-regressive with moving average (ARMA), auto-regressive with exogenous input (ARX), ARMAX, ARIMA, CARIMA, output error, and Box-Jenkins methods (input–output except AR, parametric, time domain, linear); subspace-based methods such as the eigensystem realization algorithm (ERA), the natural excitation technique combined with ERA (NExT-ERA), stochastic subspace identification (SSI), and deterministic–stochastic subspace identification (DSI) (output-only except DSI, parametric, time domain, linear); time-frequency methods such as the short-time Fourier transform, the wavelet transform, Hilbert-Huang transform, and proper orthogonal decomposition in combination with Hilbert transform or the wavelet transform (output-only, nonparametric, time frequency domain, nonlinear); and time-varying prediction error methods such as TV-ARX and TV-ARMA (input–output, parametric, time domain, nonlinear).

Damage Identification

Damage identification methods for structural systems are generally categorized into two groups:

(1) global methods (also referred to as SHM methods) and (2) local methods (also referred to as nondestructive evaluation or NDE methods). The global methods are usually applied for macroscopic damage identification of large-scale structures when there is no a priori knowledge about the location of damage and the instrumentation of the monitored structure is relatively scarce. In this case, damage is generally detected as the equivalent or effective loss of stiffness in large segments of structure (i.e., substructures). The local methods are applied when the monitoring region is smaller in size (due to, e.g., knowledge about the potential damage location(s)) and more dense (spatially) sensor data are available. NDE methods are commonly applied for monitoring aerospace structures, pipelines, and train tracks among others. These methods are capable of providing more accurate detection, localization, and quantification of damage; however, they are not in general suitable for application to larger structural systems such as building and bridges. Among NDE methods, acoustic emission, infrared thermography, and ultrasound methods can be named. Among global damage identification methods, vibration-based methods [C5] have received increased attention in the civil engineering research community in the last decade. The basic concept behind vibration-based damage identification is that the dynamic parameters of a structure are functions of its physical properties (mass, damping, and stiffness). Therefore, in concept, changes in these physical properties due to structural damage are detectable through changes in the identified dynamic parameters. One of the most common classes of methods is finite element (FE) model updating (Friswell and Mottershead 1995) which is able to detect, locate, and quantify damage. This method consists of updating the physical parameters of a FE model of the structure by minimizing an objective function that expresses the discrepancy between FE-predicted and experimentally measured response or dynamic features extracted from the response that are sensitive to damage. Dynamic features that are most commonly used for damage identification are modal parameters (especially natural frequencies and mode shapes) and the



System and Damage Identification of Civil Structures, Fig. 1 Dowling hall footbridge (*left*) and placement of concrete blocks on the bridge (*right*)

process of extracting modal parameters from vibration data is referred to as modal identification or system identification.

Damage identification methods can also be categorized into model-based methods and model-free or data-driven methods. In model-based damage identification methods, a physics-based model of the structure such as a FE model is used in the process of damage identification. These methods account for the known information about the geometry and material properties of the structure and are usually capable of detecting, localizing, and quantifying damage (level 3). Data-driven methods directly make use of the measured data on a structure and are usually capable of detection and sometimes localization of damage (level 1 or level 2). Some of the damage identification methods in the literature include methods based on changes in modal parameters such as natural frequencies and mode shapes; methods based on changes in features derived from modal parameters such as curvature mode shapes, modal strain energy, and modal flexibility; methods based on signal processing such as principal component analysis, Kalman filtering, and Hilbert–Huang transformation; and finally methods based on FE model updating.

Sample Case Studies

In this section, applications of system and damage identification methods to four large-scale civil structures are reviewed. Note that this is

not a comprehensive list of successful applications by any means and only includes a few of the authors' past work based on which some of the conclusions are drawn.

Dowling Hall Footbridge

Dowling Hall Footbridge, a full-scale footbridge located on the Tufts University campus as shown in Fig. 1, is equipped with a continuous monitoring system that measures continuously ambient vibration and temperature of the bridge. Modal parameters of the footbridge are identified every hour using an automated SSI method (Moser and Moaveni 2011; Moaveni and Behmanesh 2012). Correlation of the identified natural frequencies with ambient temperatures was investigated and different models were proposed to capture the relationship. Effects of physical damage were physically simulated on this test-bed structure by loading a small segment of the footbridge deck with concrete blocks (Fig. 1, right) so as to locally modify the bridge inertia properties since no damage could be inflicted to the bridge to modify its stiffness properties. Deterministic and probabilistic FE model updating approaches were implemented for damage identification of the bridge (Behmanesh and Moaveni 2014). Both methods could successfully estimate the location and extent of damage (here change in mass). However, the probabilistic (Bayesian) method could also provide the level of confidence in the damage identification results by estimating the posterior probability distributions of the FE model updating parameters.

Lessons Learned: The maximum a posteriori (MAP) estimates of the updating model parameters match the values of the physically simulated damage (i.e., local change in mass) and are in good agreement with the optimum parameter values obtained from the deterministic FE model updating approach. Effects of the number of data sets used in the damage identification process (i.e., “value” of added data) were investigated by using different subsets of available data and it was found that the estimation uncertainty of the updating model parameters is significantly reduced by adding more data sets to the likelihood function, which yields more accurate model updating results. However, the reduction in estimation uncertainty becomes progressively less significant as the number of data sets keeps increasing. In the application of deterministic FE model updating, addition of more data sets does not necessarily reduce the estimation uncertainty of the model updating results. Thus, probabilistic FE model updating approaches based on multiple sets of measurement data are strongly recommended for structural damage identification purposes as they can provide a measure of confidence to the estimated damage values.

Seven-Story Shear Wall Building Section

A full-scale seven-story reinforced concrete (RC) shear wall building section, shown in Fig. 2, was tested on the UCSD-NEES shake table in the period October 2005–January 2006. The shake table tests were designed to damage the building progressively through several historical earthquake ground motions reproduced on the shake table. At several levels of damage, ambient vibration tests and low-amplitude white-noise base excitation tests were applied to the building which responded as a quasi-linear system with dynamic parameters evolving as a function of structural damage. Six different state-of-the-art system identification methods including three output-only and three input–output methods were used to estimate the modal parameters (natural frequencies, damping ratios, and mode shapes) of the building in its undamaged (baseline) and various damage states



System and Damage Identification of Civil Structures, Fig. 2 Seven-story shear wall test structure

(Moaveni et al. 2011). Deterministic and probabilistic finite element (FE) model updating strategies were applied for vibration-based damage identification of the test specimen (Moaveni et al. 2010; Simeon et al. 2013). Three damage identification cases are considered based on modal parameters identified using ambient vibration, 0.03 and 0.05 g white-noise base excitation test data, respectively. The damage identification results obtained for these three cases do not match exactly, but they are consistent with the actual damage observed in the building which shows a concentration of damage at the bottom two stories of the web wall. The ambient vibration data satisfy better the assumption of system linearity and therefore are more appropriate as input for linear FE model updating.

Lessons Learned: From the results of the system identification study, it was observed that the identified natural frequencies of the three longitudinal

(in the direction of shaking) vibration modes decrease with increasing level of damage, while the corresponding identified modal damping ratios do not follow a clear trend as a function of structural damage. The (effective) natural frequencies of the three longitudinal vibration modes identified based on higher amplitude structural response data are in general lower than their counterparts identified based on low-amplitude response data at all damage states considered, especially for the first longitudinal vibration mode. This is most likely due to the fact that the test structure was nonlinear (even at the relatively low levels of excitation considered in the system identification study) with effective modal parameters depending strongly on the amplitude of the excitation and therefore of the structural response. In general, higher modal damping ratios were identified for the three longitudinal vibration modes during the higher amplitude base excitation tests. This is due to the fact that the additional hysteretic energy dissipation (due to inelastic action of the material) at higher level of response nonlinearity is identified as equivalent viscous damping by the linear system identification methods used. It should be emphasized that these “inflated” identified equivalent viscous damping ratios are not to be used to represent the viscous damping component in a nonlinear FE model of a structure that explicitly accounts for the nonlinear material behavior. The damage identification results obtained using a linear FE model updating approach are sensitive to the amplitude of the white-noise base excitation to which the structure is assumed to respond quasi-linearly. With increasing level of base excitation and structural damage, the level of nonlinearity in the structural response increases. Therefore, the assumption that the structure behaves as a quasi-linear dynamic system is violated and a linear dynamic model is not strictly able to represent well the behavior of a damaged structure. The spatial distribution (i.e., relative amplitudes) of the identified damage factors however is not sensitive to the amplitude of base excitation and was found in good agreement with the

damage observed or inferred (from strain sensors) in the building specimen. The Bayesian FE model updating approach also succeeded in identifying the damage in the test structure and the results obtained were found in good agreement with their deterministic counterparts. The probabilistic model updating method also provided an estimate for the uncertainties in the identified damage values. It was found that the data used contained little information about the state of the top stories of the building, as shown by the fact that the uncertainty of the parameters representing this portion of the structure could not be reduced through Bayesian processing of the observed data.

Three-Story RC Frame with Masonry Infilled Walls

A deterministic linear FE model updating strategy was applied for vibration-based damage identification of a 2/3-scale, three-story, two-bay, infilled RC frame, tested on the UCSD-NEES outdoor shake table (Moaveni et al. 2013b). Figure 3 shows the test structure on the shake table. FE model updating was first used to calibrate the FE model at the undamaged state which served as the reference/baseline state and then applied for damage identification at a number of damage states of the structure. These damage states correspond to states of increasing damage of the physical specimen which was subjected to a sequence of earthquake excitations of increasing intensity. The damage identification results indicate that the severity of structural damage increases as the structure is exposed to stronger earthquake excitations. The damage identification method correctly identifies the spatial distribution of damage in the structure, with the most severe damage at the bottom story and the least damage at the top story. The method also captures the fact that the extent of damage is more significant in the infill walls than in the columns. The analytical modal parameters obtained from the updated FE models are in good agreement with their experimentally identified counterparts, an indication of the accuracy of the updated FE models.



System and Damage Identification of Civil Structures, Fig. 3 Front view (*left*) and side view of the three-story reinforced concrete frame with masonry infilled walls (*right*)

System and Damage Identification of Civil Structures, Fig. 4 Three-story precast concrete parking garage test structure



Lessons Learned: Comparison of the damage identification results and the seismic shake table test results shows that the level of damage identified may not reflect the loss of the structural strength, since the loss of stiffness (commonly defined as structural damage) is not well correlated to the loss of strength. This motivates the need for new research based on nonlinear FE model updating based on calibrated, mechanics-based, nonlinear FE models of structural systems able to capture stiffness degradation and strength deterioration phenomena.

Three-Story Precast Concrete Parking Garage

This study performed damage assessment through system identification of a three-story half-scale precast concrete building resembling a parking garage tested under earthquake type loading on the NEES-UCSD Shake Table in 2008 (Belleri et al. 2014). Figure 4 shows the parking garage test structure on the NEES-UCSD shake table. The system identification was performed using the DSI method. The effective modal parameters of the structure at different damage states show significant changes after

each of the high intensity earthquake tests. In general, the identified natural frequencies decrease and the identified damping ratios increase as the structure is exposed to base excitations of increasing intensity. Changes in the identified modal parameters are correlated with the observed damage in the structure. The analysis of the identified mode shapes allow to point to the location of damage. Specifically, localized changes in the first and second spatial derivatives of mode shapes, modal rotations, and curvatures indicate local losses of shear and bending stiffness, respectively. Local significant changes in modal rotations and curvatures allowed localizing both visually detected damage in the third floor mid-span joint and visually undetected damage in the second floor mid-span joint, thus providing a useful tool for precast floor damage detection.

Lessons Learned: The accuracy and spatial resolution of the damage identification results depend significantly on the accuracy and completeness of the identified modal parameters. The estimation variability (or uncertainty) of the modal parameters can be influenced by several factors such as the number and types of sensors, measurement noise, length of the data time windows used for system identification, and the system identification method used, in addition to changes in the environmental conditions such as the ambient air temperature and relative humidity during data collection. The variability in the identified modal parameters due to non-damage related factors needs to be smaller than the changes in these parameters due to damage or compensated for in order to identify the actual damage in the structure using vibration measurements.

Summary

This paper provides an overview of the concepts of system identification and damage identification of civil structures. First, the need for system and damage identification of civil structures is underlined and the current gap in the knowledge

is discussed. System identification and damage identification processes are defined and some of the popular methods for application to civil structures are mentioned. Finally, a few successful applications to large-scale civil structures are provided as illustrations and conclusions are drawn from these studies.

Cross-References

- [Modal Analysis](#)
- [Nonlinear System Identification: Particle-Based Methods](#)
- [Operational Modal Analysis in Civil Engineering: An Overview](#)
- [Vibration-Based Damage Identification: The Z24 Bridge Benchmark](#)

References

- Aktan A, Farhey D, Helmicki A, Brown D, Hunt V, Lee K, Levi A (1997) Structural identification for condition assessment: experimental arts. *J Struct Eng ASCE* 123(12):1674–1684
- American Society of Civil Engineers (ASCE) (2009) Report card for America's infrastructure. <http://www.asce.org/reportcard/2009>
- Behmanesh I, Moaveni, B (2014) Probabilistic identification of simulated damage on the Dowling Hall Footbridge through Bayesian FE model updating. *Structural Control and Health Monitoring*, in press
- Belleri A, Moaveni B, Restrepo JI (2014) Damage assessment through structural identification of a large-scale three-story precast concrete structure. *Earthq Eng Struct Dyn* 43(1):61–76
- Catbas FN, Kijewski-Correa TL, Aktan AE (eds) (2013) *Structural identification of constructed systems: approaches, methods and technologies for effective practice of St-Id*. Technical report, ASCE, Structural Engineering Institute, Reston
- Friswell MI, Mottershead JE (1995) *FE model updating in structural dynamics*. Kluwer, Boston
- Moaveni B, Behmanesh I (2012) Effects of changing ambient temperature on finite element model updating of the Dowling Hall Footbridge. *Eng Struct* 43:58–68
- Moaveni B, He X, Conte JP, Restrepo JI (2010) Damage identification study of a seven-story full-scale building slice tested on the UCSD-NEES shake table. *Struct Saf* 32(5):347–356

- Moaveni B, He X, Conte JP, Restrepo JJ, Panagiotou M (2011) System identification study of a seven-story full-scale building slice tested on the UCSD-NEES shake table. *J Struct Eng ASCE* 137(6):705–717
- Moaveni B, Hurlbaush S, Moon F (2013a) Editorial of special issue on real-world application of structural identification and health monitoring methodologies. *J Struct Eng ASCE* 139:1637–1638
- Moaveni B, Stavridis A, Lombaert G, Conte JP, Shing PB (2013b) Finite element model updating for assessment of progressive damage in a three-story infilled RC frame. *J Struct Eng ASCE* 139:1665–1674
- Moser P, Moaveni B (2011) Environmental effects on the identified natural frequencies of the Dowling Hall Footbridge. *Mech Syst Signal Process* 25(7):2336–2357
- Simeon E, Moaveni B, Conte JP, Lombaert G (2013) Uncertainty quantification in the assessment of progressive damage in a seven-story full-scale building slice. *J Eng Mech ASCE* 139(12):1818–1830

Timber Structures

Geoff Thomas

School of Architecture, Victoria University of Wellington, Wellington, New Zealand

Synonyms

Earthquake design; Earthquake resistance; Timber buildings; Wood; Wood buildings; Wood products

Introduction

As societies developed it was seen as desirable to build buildings in “permanent” materials such as stone and brick, rather than materials such as timber and thatch. This became more desirable as settlements increased in size and buildings were built closer together. This increases the risk of rapid fire spread between closely spaced buildings with timber claddings and thatch or other combustible roofs. Prior to the advent of modern fire fighting equipment and organizations, and automatic suppression systems in the late nineteenth century, noncombustible materials were the only way of preventing fire spread between and within buildings. This is reflected in the earliest building codes which predominantly dealt with designing buildings to prevent urban conflagrations, such as the Roman ordinances

reported by Klitze (1959). This desire for permanent materials, however, led to another risk, that of building failures in earthquakes.

Timber frame buildings have an inherent resistance to earthquakes. As the loads induced by earthquakes are directly proportional to building mass and increase with the stiffness of a structural system, timber buildings, with their light mass and low stiffness, result in reduced earthquake loads. Timber connections, in the main, are relatively flexible. This has the effect of reducing earthquake loads as movement in connections further reduces the stiffness of the structural system. In the event of failure, it is more survivable in a light weight structure than one with heavy elements. In earthquake-prone regions, the concept that permanent materials were better did not take hold, with no better example than Japan with its long history of both earthquakes and timber buildings. For the Kobe earthquake of 1995, the fatality rate in timber buildings was estimated at 2 % but 15 % in reinforced concrete buildings (Langham 1995). In New Zealand, a country with similar seismicity to Japan, timber buildings were common as European settlement increased in the nineteenth century, as timber was readily available. The predominantly British colonists however valued masonry buildings, but this attitude changed significantly after the magnitude 8+ Wairarapa Earthquake in 1855.

In the past the size of timber frame structures was limited by the available length and girth of

structural elements cut from timbers. In some significant buildings, very large timber elements were sourced to produce very large structures. Very long spans were achieved by joining separate timber elements into trusses or frames, for example, Westminster Hall in London and Kiyomizudera Temple in Kyoto. In the nineteenth- and twentieth-century, large timbers have been hard to source, and most timber frame structures were of small scale, until the advent of glue laminated timber (glulam), starting in the 1960s, and laminated veneer lumber (LVL) which became widely available in the 1990s.

As timber structures have tended to be low or medium rise, and the inherent mass and stiffness of timber structures is low, lateral loading is often governed by wind load rather than earthquake loads. For many buildings in many areas, design for wind loads was sufficient to provide adequate resistance to earthquake loads.

The design of timber frame structures has really split into two areas, light timber frame structures, most commonly used for low-rise residential dwellings and similar-sized buildings, and heavy timber frame buildings, most commonly as single-story long-span structures. Timber is suited to long-span roof structures because in roofs, most of the load is self-weight of the structural elements and timber has low self-weight. Light timber frame has been used for low- to medium-rise apartment complexes, a development originating in the Pacific Northwest of the United States and the western seaboard of Canada. More recent developments in connection details for glulam and LVL moment-resisting frame buildings have enhanced their suitability for multistory heavy timber frame buildings.

As concerns around sustainability grow, the use of timber in buildings is becoming more desirable as timber is grown rather than manufactured with low embodied energy (John et al. 2009). During the life of the structure, the carbon sequestered in the timber as it grew is stored and therefore not released to the atmosphere until after the building is demolished.

This entry will start with some background on timber properties and how they affect earthquake

resistance and then go into more detail on light and heavy timber frame buildings separately. For the purposes of this entry, light timber frames are defined as smaller timber elements that require panel bracing as secondary elements and heavy timber frames as other types of structures.

Timber Properties

Timber is a natural, not manufactured, material; therefore its properties cannot be determined by design. The strength of a piece of timber is predominantly determined by defects such as knots and sloping grain (grain that is running at an angle to the cut surfaces of the piece of timber). Strength and stiffness are variable even after timber has been graded visually or mechanically based on stiffness. Figure 1 shows the variation of strength and stiffness of the same species of timber within a grade.

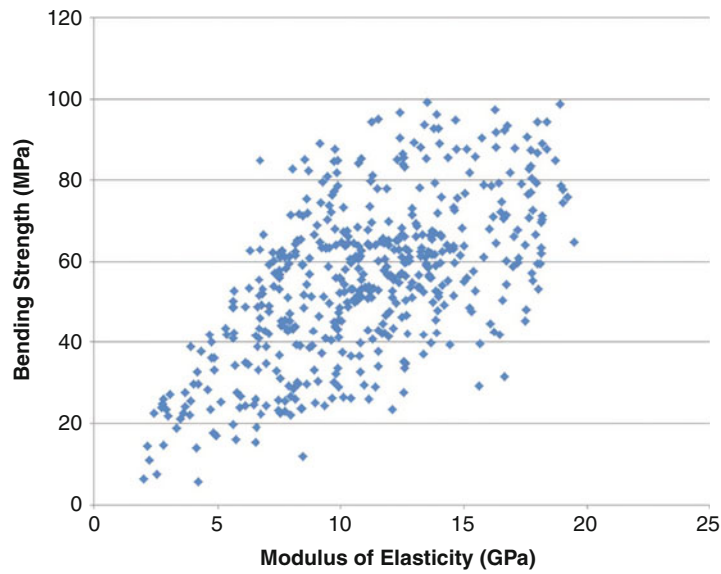
Mechanically stress-graded timber is graded based on stiffness, but as can be seen in Fig. 1, strength is not directly proportional to stiffness.

The strength and stiffness of timber are highly dependent on species, with values of 4.7 MPa and 460 MPa for the strength and stiffness of balsa, respectively, to 84 MPa and 21.5 GPa, respectively, for “box,” an Australian hardwood. In structural terms “hardwood” and softwood” can be misnomers, as they are a botanical classification, with hardwoods being crudely, flowering trees, and softwoods trees that do not flower. Technically balsa is a hardwood.

At a microscopic level, timber consists mostly of long cylindrical cells, orientated along the trunk or branches of a tree (Fig. 2), held together by lignin which acts as a “glue.” This arrangement explains the difference in strength perpendicular and parallel to the grain in timber (Fig. 3), as the cells are orientated to take the load along the trunk of the tree. It also partly explains the weaker strength in tension, as it is easier to break the lignin in tension in a brittle failure, whereas in compression the lignin can yield and the cell walls buckle in a ductile fashion. Hence timber is brittle in tension and ductile in compression. In bending, the timber will initially yield in a ductile

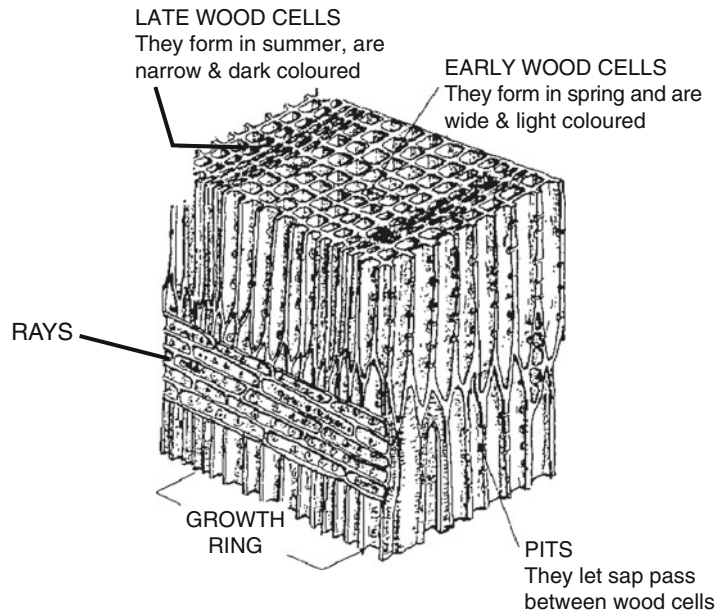
Timber Structures,

Fig. 1 Typical variation of strength and stiffness for timber



Timber Structures,

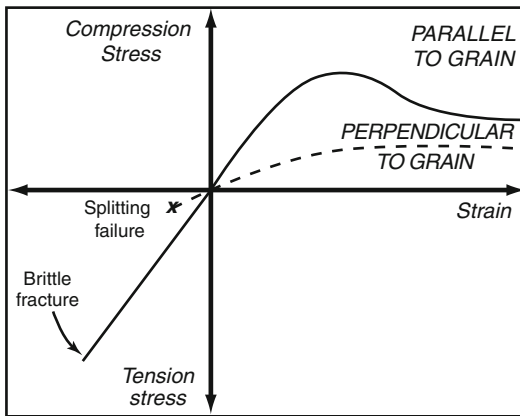
Fig. 2 Cellular structure of timber (Buchanan 2007)



manner in compression but will eventually fail with a brittle tension failure in the extreme tension fiber which then rapidly propagates through the rest of the member, as the effective section depth decreases, and hence stress in the extreme tension fiber increases. Defects such as knots and sloping grain also have more effect in tension than compression, further explaining the

difference between tension and compressions strength. A material such as timber that has different properties in different directions is known as "anisotropic." As the physical makeup of timber varies over its volume, it is also known as "nonhomogeneous" or literally not the same.

Timber behaves better in compression and bending, the way in which trees resist their



Timber Structures, Fig. 3 Strength in timber in compression and tension, perpendicular and parallel to the grain (Buchanan 2007)

natural loading, as trees have evolved to resist compression from its self-weight acting down the trunk and bending from wind loads.

Good earthquake design for timber requires that timber elements are loaded in the fashion they are best suited to, that is, in bending or compression parallel to grain, where timber is both stronger and more ductile. However few building failures are caused by failure of the main structural elements. More failures are caused by poor detailing of connections, both timber to timber connections and timber connections to other materials such as concrete foundations.

Connections

Good performance of timber structures requires good detailing of connections, but for timber as well as taking care to ensure that connections are not compromised by secondary moments that caused eccentricities of loads at joints, joint design must take into account the anisotropic nature of timber. In particular care must be taken that fasteners do not produce excessive loads in tension perpendicular to the grain, the worst way to load timber. Timber codes have limits on minimum fastener spacing to avoid this problem, but as with any joint design, secondary moments may cause additional loading on

fasteners at the edges of fastener groups which if not addressed can lead to progressive “unzipping” failures of connections, where the most highly stressed fastener at the edge of a group fails, leading to higher stress on the next fastener which then fails and so it continues until the joint fails.

Timber connections can be designed for ductility but a feature common to most types of timber connections is that the hysteresis loops under cyclic loading are “pinched” (Fig. 4).

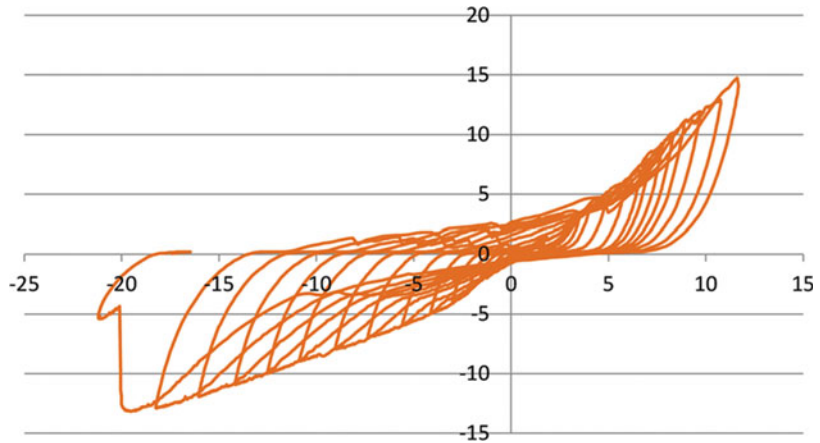
This “pinching” is caused by the system having less stiffness when unloaded than when loaded. The other feature to note is that the stiffness of the system decreases with each load cycle and is only recovered when the displacement exceeds the maximum displacement reached on the previous load cycle. The area within the hysteresis loop represents energy as it is a force times a distance, or work. This is the energy absorbed in each load cycle, which with a pinched hysteresis loop is less than that absorbed in other ductile materials with “fatter” loops so the vibration of the structure under earthquake loads has less damping. The decreased stiffness with subsequent load cycles will produce large displacements with a large number of cycles, which can eventually lead to a P-delta failure. With large displacements the line of force from gravity loads may be so far outside columns and load-bearing walls that they overturn or fail due to secondary bending moments caused by eccentric gravity loads.

Light Timber Framing

Light framing has been around for millennia, with the first structures likely to have been designed to resist wind load and hence earthquake load by cantilevering columns out of post-holes in the ground. The next innovation was almost certainly braced frames with diagonal braces cut-in between columns and beams acting only in compression as the connection need only act in bearing and not require a tensile connection. These developed into tension and compression braces as the technology developed with

Timber Structures,

Fig. 4 Hysteresis loop for timber connection; note that this right angle joint is stiffer in closing (right-hand upper quadrant) than opening

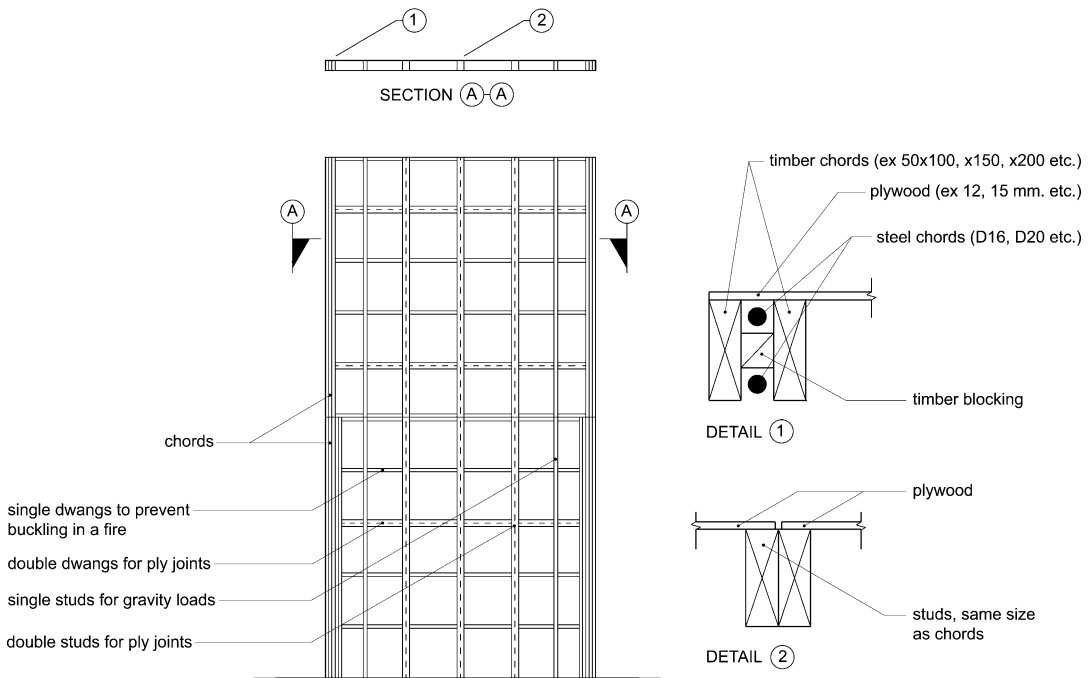


steel bolts and nails. These technologies, cantilever posts, and braces are still used for the foundations of many light timber frame structures such as single houses to brace the foundations against earthquake and wind loads.

Houses and other light timber frame structures built in the nineteenth and early twentieth centuries frequently had cut-in timber braces for lateral loads, but these only really held the frames together until claddings and linings were installed. Lateral load resistance was provided by claddings such as concrete plaster or bricks, which tend to have sufficient in-plane stiffness to resist moderate earthquakes, but are prone to out-of-plane failures. Timber weatherboard has some lateral load resistance, but it depends on the way they are nailed, and the resulting structure is very flexible. The widespread availability of plywood, and then reconstituted board products such as “particle-board” or “oriented strand board” (OSB), in the second half of the twentieth century led to their use as either bracing elements where high loadings occurred or, as is common north American practice, fully sheathing house framing with a panel product. In the 1920s and 1930s, interior linings of gypsum plasterboard started to become the norm in many countries. These linings are much stiffer than timber board linings and it was soon found they were much stiffer than cut-in braces or plywood bracing. Hence if a building is lined with gypsum plasterboard, then any other bracing element will only develop its full load after the gypsum plasterboard has failed. This has led to the

development of gypsum plasterboard as a bracing material, with some highly engineered products achieving high levels of bracing performance providing they are fixed in accordance with specifications (Winstones 2011). The approach of fully sheathing houses is probably more robust as it is less dependent on the quality of workmanship, but may result in more damage to internal linings in moderate earthquakes as these are not detailed to resist load, even though they will attract most load at low displacements due to their higher stiffness. Shear walls constructed as panel products nailed to timber framing can fail in four ways under lateral load (Thomas 1991), tension failure of tie-downs, panel shear, bending failure (as tensile or compressive failure of chords), or nail slip, allowing the panel to rotate on the framing. Nail slip is a ductile failure mode, with nails bending into an S shape between the panel and the framing, with localized crushing of both the timber framing and the panel material. The fastening size and spacing and the panel sheet thickness must be chosen so that the ductile nail-slip failure mode occurs and other details must be relatively oversized to prevent other failure modes. Although code requirements for bracing have significantly improved in the latter part of the twentieth century, performance of houses has not necessarily improved as the trend for more and larger windows in houses has compensated for this (Thomas et al. 2013).

The performance of small light timber frame structures such as houses has been demonstrated to be very good, with a low proportion of serious



Timber Structures, Fig. 5 Plywood shear wall and details (Thomas 1991)

failures and wholesale collapse being extremely rare (Graf 2008). Common cause of failure is poor bracing in older Japanese houses with heavy roofs (Langham 1995), poorly braced lower floors such as basement or ground floor levels used car parking (Graf 2008; Iqbal 2013), or poorly braced or poorly fixed foundations (Graf 2008; Thomas et al. 2013).

Most light timber frame structures in the world are low rise, but light timber frame is also suited to medium-rise residential buildings, which have relatively small room sizes. Their development originated in the Pacific Northwest of the United States and western Canada, areas with a long history of light timber frame structures. US practice has been codified by the American Wood Council (ANSI & AF&PA, 2008). Other relevant standards and information can be accessed via Woodworks (2013).

The development of medium-rise timber buildings required the development of lateral load-resisting elements that are sufficiently strong enough to resist earthquake and wind loads for taller buildings. The approach taken was to use

highly loaded plywood shear walls. In order for these to be successful, two shortcomings of light timber frame structures need to be addressed, that is, continuity between short elements and low tensile strength of timber elements.

The shear walls are designed to fail with nail slip allowing the plywood panels to rotate on the framing, a ductile failure mode (Deam 1997). The behavior of the nail-slip failure mode for shear walls under cyclic loading is similar to that in Fig. 4 but is symmetric in tension and compression. The tie-downs at the ends of the wall are oversized, as are the plywood sheet thickness and chord sizes. An over strength factor of 2.0 is normally used (Buchanan 2007). In order to provide for continuity over the length of the chords, which in the case of a five-story building may be 16 m or more long, the chords consist of timber studs interspersed with steel tension rods that can be coupled together to provide tensile strength and continuity and are anchored into the foundation (Fig. 5).

An example of this type of building is Martin Square (Fig. 6), with six floors of timber frame

Timber Structures,**Fig. 6** Martin Square Apartments, Wellington, New Zealand

construction with plywood shear walls, over a basement car park constructed from concrete. This building was completed in 2004, at which time was the highest timber building built in an earthquake-prone area (Wellington, New Zealand).

In light timber frame buildings, the floor diaphragms can be designed as simply supported between lateral load-resisting elements but would normally be designed as a continuous horizontal beam. The floor diaphragms are designed for elastic response. Unlike steel or concrete buildings where diaphragm forces are normally dealt with by typical construction and a check only is required, specific design is essential. With timber floor diaphragms, the connections to lateral load-resisting elements need to be designed to resist the shear transfer, and chords at the edges of the diaphragms need to be designed to take all the bending forces, as they act in tension and compression.

Heavy Timber Framing

Heavy timber frame buildings traditionally were built out of specially sourced large pieces of sawn

timber, usually for buildings of great significance. As stocks of virgin timber disappeared, it became harder and more expensive to source sufficiently large pieces of timber, and with the development of steel and reinforced concrete structures, heavy timber buildings became less common until the advent of glulam and LVL. Heavy timber buildings can really be divided into three types in terms of earthquake resistance; low-rise buildings which are dominated by wind load, structures with a lateral load-resisting system made of another material, such as concrete shear walls or braced steel frames, or more recently moment-resisting heavy timber frames with rigid beam-column connections and cross-laminated timber structures.

Wind Load Dominated

Heavy timber is ideally suited for long-span roof structures, as most of the load on a roof is self-weight, and timber is very light compared to other common structural materials. It can be more economical than steel for portal frame industrial buildings, especially when other implications

Timber Structures,

Fig. 7 Odlines building, Petone, showing concrete shear walls and glulam gravity frame (Photo: Peter Smith)



such as lighter foundation requirements are taken into account. Connections in portal frame buildings can be built using plywood or steel gussets or cross-lapped and glued or using steel brackets and epoxied steel rods. Plywood gussets can be fixed with nail guns, but steel plates need to be drilled and hand nailed. Nailing patterns tend to be hollow squares or rectangles which distribute loads into nails relatively evenly avoiding unzipping failures. As these structures tend to be dominated by wind load, they can be elastically responding in earthquakes or of low ductility, so detailing for ductility is either not required or relatively straightforward. Longitudinal loads can be taken by plywood shear walls or bracing in other materials, but the most economic option is usually steel tension bracing. The longitudinal bracing is usually designed for an elastic response or low ductility. With more complex arrangements such as trusses, very long spans can be achieved, for example, the Hamar Olympic Stadium built for the 1994 Winter Olympics with a span of 96.4 m.

Smaller buildings, building platforms for small buildings, and car-parking decks can be built using poles or naturally round timber, with the simple form of construction being a cantilever pole either driven into the ground or concreted into a bored posthole. Roof or floor elements can then be built up using solid

timber beams or trusses, joists, and flooring. Earthquake load resistance is provided by ensuring the poles have sufficient embedment to resist overturning.

Other Lateral Load Resisting Materials

Design and construction of gravity frames in heavy timber is relatively straightforward. In single-story structures, beams can rest upon columns with simple fastenings. In multistory structures, this is not desirable as perpendicular to the grain compression or bearing forces are developed in the beams. Timber is not as good at resisting such forces as compression parallel to the grain. Brackets can be bolted or sometimes nailed to the columns to support the beams which are then fixed to the brackets. The lateral loads are then taken by another structure, such as concrete shear walls ([► Reinforced Concrete Structures in Earthquake-Resistant Construction](#)) which may be around a service core containing the stairs and elevators. Odlines building (Fig. 7) is an example of this. Other options are steel tension bracing, cross bracing and K bracing, or eccentrically braced frames ([► Steel Structures](#)). In this type of building, the design of the lateral load-resisting structure is as it would be for a building made in its entirety of another material, but

Timber Structures,
Fig. 8 COCA building,
 Massey University,
 Wellington. The steel truss
 tower is temporary bracing



earthquakes loads may be lower due to the lighter mass of the timber structure.

Some timber buildings may have a timber moment-resisting frame (Moment Resisting Frames) in one direction and another have a structural system of another material in the other direction. The Massey COCA building (Fig. 8) is one such building, with concrete shear walls in the longitudinal direction. Other buildings may have other types of structural system in locations where timber structural systems would not work. An example is Martin Square (Fig. 6), which has three steel k braces in the outside walls of the building. Lateral load-resisting elements are therefore at the extremities of the building with a large separation between these elements to provide adequate resistance to torsion. It was not possible to install plywood shear walls on the external walls due to the large number of and small spacing between windows (Milburn and Banks 2005).

Moment-Resisting Frames

Designing and building a timber gravity frame structure is relatively straightforward, and light timber frame structures with plywood shear walls are restricted in terms of room size because of short floor plans and the requirement for lots of walls. Lateral load-resisting structures such as

concrete shear walls and steel bracing also restrict options in terms of building layouts. The only realistic way of building timber buildings as commercial space with long floor spans and large open areas is a moment-resisting frame structure. As timber does not perform well in cyclic loading outside the elastic range as it is brittle in tension, then ductility can only be achieved in the connections.

Nailed moment-resisting connections are designed to ensure nail slip as the steel nails yield, a ductile failure mode, rather than a brittle failure of the timber. These types of connections tend to have a pinched hysteresis loop similar to Fig. 4, with the same issues of poor energy absorption and degrading stiffness. These limit the height that buildings with nailed moment-resisting connections can be built to.

Better joint behavior can be achieved using steel rods running through columns and the ends of beams and epoxied into place. This system takes advantage of the ductile behavior of the steel rod which can yield not only in tension but also in compression as buckling is prevented by the surrounding timber. The cyclic load behavior is good, with characteristics similar to reinforced concrete.

The development of LVL which has better strength and stiffness properties than glulam, as well as being less expensive with a more automated production process, has resulted in LVL buildings being more economical than their

glulam equivalent and therefore more competitive against other materials.

Care needs to be taken in the design and connection of floor diaphragms as described in the section on light timber framing.

More recently timber “rocking frame” buildings have been designed based on a procedure developed at the University of Canterbury, New Zealand. This design is similar to the “PRESSS” system developed for reinforced concrete at the University of California San Diego in the 1990s (► [Reinforced Concrete Structures in Earthquake-Resistant Construction](#)) and is known as “Pres-Lam.” These types of systems use “damage avoidance design,” with systems that self-center after earthquakes and damage from earthquakes concentrated in easily replaceable energy-dissipating elements. In timber buildings frames or walls are posttensioned with steel tendons and fitted with external energy-dissipating devices. An experimental building with such frames in one direction and walls in the other has been tested at the University of Canterbury and after modification for use as an office building survived the 22 February 2011 Christchurch Earthquake. Buildings with posttensioned walls at Nelson Marlborough Institute of Technology and Massey University (Fig. 8), with LVL posttensioned frames in the transverse direction, have subsequently been built ([NZ Wood 2013](#)).

Cross-Laminated Timber Structures

Cross-laminated timber are large panels of glue-laminated timber, but the laminates run perpendicular to each other rather than parallel to form large panels. These panels are more suited to residential buildings with small room sizes and a large numbers of walls. The panels are relatively thick and very stiff and under earthquake loads rotate effectively as a rigid body, with any ductility coming from the connections (Pei et al. 2012). A seven-story experimental cross-laminated apartment building was successfully tested on the E-Defense shake table in Japan in 2009 (Merolla 2009).

Summary

Timber structures inherently have good earthquake resistance as they are lightweight and have low stiffness, but ductility can only really be achieved in the connections. For smaller and low-rise timber structures, design for wind load and good practice in detailing are sufficient to ensure adequate earthquake resistance. A desirability to use more timber in construction for environmental reasons has recently led to more interest in multistory timber buildings and innovative research in how to design these for earthquake resistance.

Cross-References

- [Reinforced Concrete Structures in Earthquake-Resistant Construction](#)
- [Steel Structures](#)

References

- ANSI & AF&PA (2008) Special design provisions for wind and seismic. American Forest and paper Association & American Wood Council. <http://www.awc.org/pdf/2008WindSeismic.pdf>. Accessed 13 Sept 2013
- Buchanan AH (2007) Timber design guide. New Zealand Timber Industry Federation, Wellington
- Deam BL (1997) The seismic design and behaviour of multi-storey plywood sheathed timber frame shearwalls. Civil Engineering Research Report 97/3. University of Canterbury, Christchurch. <http://ir.canterbury.ac.nz/handle/10092/5920>
- Graf W (2008) The shakeout scenario. Supplemental study. Wood frame buildings. USGS and CGS. <http://www.colorado.edu/hazards/shakeout/woodframe.pdf>. Accessed 13 Sept 2013
- Iqbal A (2013) Seismic retrofit solutions for light timber frame buildings with soft first story. In: Proceedings of the NZSEE conference, poster 51, Wellington. http://www.nzsee.org.nz/db/2013/Poster_51.pdf
- John S, Nebel B, Perez N, Buchanan AH (2009) Environmental impacts of multi-storey buildings using different construction materials. Civil Engineering Research Report 2008–02. University of Canterbury, Christchurch. <http://ir.canterbury.ac.nz/handle/10092/8359>
- Klitze RA (1959) Roman building ordinances relating to fire protection. *Am J Legal Hist* 3:173–187. Archived on <http://www.jstor.org>

- Langham B (1995) Lessons learnt from Kobe. NZ Wood. <http://www.nzwood.co.nz/learningcentre/lessons-learnt-from-kobe/>. Accessed 13 Sept 2013
- Merolla L (2009) Designing a (wooden) earthquake proof home. Popular Mechanics. <http://www.popularmechanics.com/science/environment/natural-disasters/4324941>. Accessed 11 Sept 2013
- Milburn J, Banks W (2005) Six-level timber apartment building in a high seismic zone. N Z Timber Des J 13(2):11–15. <http://www.timberdesign.org.nz/files/Six-Level%20Timber%20Apartment%20Building.pdf>
- Pei S, Popovski M, van de Lindt JW (2012) Seismic design of a Multi-Storey Timber Building based on Component Level Testing. In: Auckland proceedings of the world conference on timber engineering. <http://www.timberdesign.org.nz/files/00279%20Shiling%20Pei.pdf>
- Thomas GC (1991) The feasibility of multi-storey light timber frame buildings. Civil Engineering Research report 91/2. University of Canterbury, Christchurch
- Thomas GC, Kim B, Beattie GJ, Shelton RA, Sim DA (2013) Lessons from the performance of houses in the Canterbury earthquake sequence of 2010–11. In: Proceedings of the NZSEE conference, paper 21, Wellington. http://www.nzsee.org.nz/db/2013/Paper_21.pdf
- NZ Wood (2013) Post tensioned structural components. NZ Wood. <http://www.nzwood.co.nz/what-wood/structural-materials/post-tensioned-structural-components/>. Accessed 11 Sept 2013
- Winstones (2011) GIB Ezybrace® Systems. Winstone Wallboards Ltd <http://gib.co.nz/products/plaster-board/gib-braceline-and-gib-noiseline/>
- Woodworks (2013) <http://woodworks.org/>. Accessed 13 Sept 2013

Time History Seismic Analysis

Nikos D. Lagaros, Chara Ch. Mitropoulou and Manolis Papadrakakis
Institute of Structural Analysis & Antiseismic Research, Department of Structural Engineering, School of Civil Engineering, National Technical University of Athens, Athens, Greece

Synonyms

Artificial accelerograms; Direct integration methods; Inertia and damping forces; Linear and nonlinear analysis; Natural records; Seismic loading

Introduction

Most of the current seismic design codes belong to the category of prescriptive design procedures (or limit-state design procedures), where if a number of checks are satisfied, then the structure is considered safe since it fulfills the safety criterion against collapse. A typical limit-state-based design can be viewed as one- (i.e., ultimate strength) or two-limit-state approach (i.e., serviceability and ultimate strength). Existing seismic design procedures are based on the principle that a structure will avoid collapse if it is designed to absorb and dissipate the kinetic energy that is induced during a seismic excitation. Most of the modern seismic norms express the ability of the structure to absorb energy through inelastic deformation using a reduction or behavior factor that depends on the material and the construction type of the structure. The concept of performance-based design (PBD) was introduced a few decades ago, for designing structures subjected to seismic loading conditions. In PBD more accurate and time-consuming analysis procedures are employed, to estimate nonlinear structural response. The progress that takes place in the area of computational mechanics, as well as in computer technology, continuously expands the capabilities and the applicability of PBD procedures. The main objective of this kind of design procedures is to achieve more predictable and reliable levels of safety and operability against natural hazards. According to PBD procedures, the structures should be able to resist earthquakes in a quantifiable manner and to present specific target performance levels of possible damages. PBD criteria try to define certain levels of structural performance for various levels of seismic hazard.

In the construction industry, decision making for structural systems situated in seismically active regions requires consideration of damage cost and other losses resulting from earthquakes occurring during the life span of a structure. Thus, nowadays, life-cycle cost analysis (LCCA) becomes an essential component of the design process used to control the initial and the future

cost of building ownership. In the early 1960s, LCCA was applied in the commercial area and in particular in the design of products considering the total cost of developing, producing, using, and retiring. The introduction of LCCA in the construction industry was made in the field of infrastructures as an investment assessment tool. In particular, in the early 1980s, it was used in the USA as an appraisal tool for the total cost of ownership over the life span of an asset. Later, in view of large losses due to extreme hazards, like earthquakes and hurricanes, there was a need for new design procedures of facilities that could lead to life protection and reduction of damage and economical impact of such hazards to an acceptable level. In this context LCCA was introduced in the field of constructions as a complex investment appraisal tool incorporating a structural performance criterion.

Over the last decades, risk management of structural systems has gained the attention of various economic and technical decision centers in modern society. The optimal allocation of the public resources for a sustainable economy requires proper tools for estimating the consequences of natural hazardous events on the built environment. The risk management addresses this claim indicating the way for implementing optimal choices. Thus, the main purpose of the risk management process is to choose among different options relying on technical and economical considerations. Risk assessment and decision analysis are the main steps of the risk management concept. It is therefore essential to establish a reliable procedure for assessing the seismic risk of structural systems. Seismic fragility analysis, which provides a measure of the safety margin for the structural system, is considered as the core of the risk assessment framework.

The implementation of PBD, LCCA, or seismic risk procedures requires a reliable computing tool for estimating the capacity and the demand for any structural system. Some of these computing tools (Fajfar 2000; Chopra and Goel 2002; Vamvatsikos and Cornell 2002) are based on time history seismic analyses, and they are considered as analysis methods for obtaining good estimates

of the structural performance in the case of earthquake hazard, and therefore, they are considered as appropriate methods to be incorporated into such design and assessment frameworks. In this entry, we provide an overview of the time history seismic analysis applied for design and assessment purposes.

Ground Motion Excitation

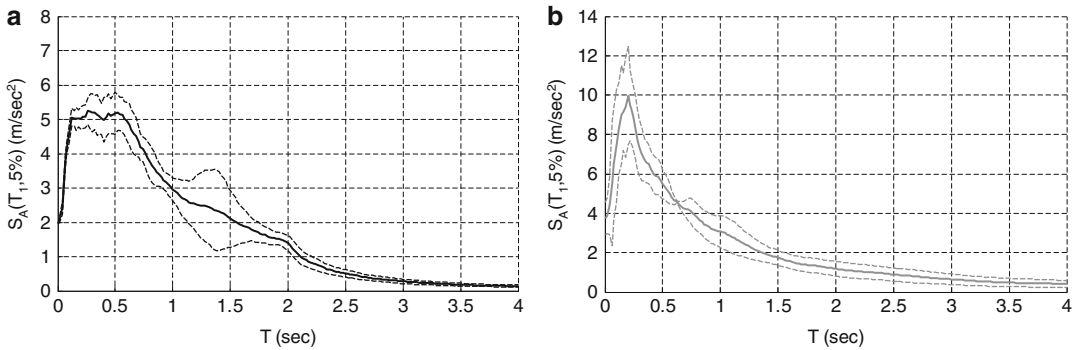
Selecting the seismic loading for design and/or assessment purposes is not an easy task due to the uncertainties involved in the very nature of seismic excitations. One possible approach for the treatment of the seismic loading is to assume that the structure is subjected to a set of records that are more likely to occur in the region where the structure is located.

Natural Records

For the implementation of dynamic seismic analysis, a scale factor is calculated for each hazard level considered and for each one of the records selected. In order to preserve the relative scale of the two components of the records in the longitudinal and transversal directions, the component of the record having the highest $S_A(T_1, 5\%)$ is scaled first, while the scaling factor that preserves their relative ratio is assigned to the second component.

Generation of Artificial Accelerograms

In order for the artificial accelerograms to be representative, they have to match some requirements of the seismic codes. The most essential one is that the accelerograms have to be compatible with the elastic design response spectrum of the region. Each accelerogram corresponds to a single response spectrum for a given damping ratio. On the other hand, on each response spectrum corresponds an infinite number of accelerograms. Gasparini and Vanmarke (1976) were the first to propose the creation of artificial accelerograms based on a specific response spectrum. The mean (μ) response spectrum is depicted in Fig. 1, along with its dispersion ($\mu + \sigma$ and $\mu - \sigma$ denoted with dotted lines) for the case of



Time History Seismic Analysis, Fig. 1 Mean response spectra of the records for the class of (a) artificial accelerograms and (b) natural records (*dotted lines* represent the $\mu + \sigma$ and $\mu - \sigma$ response spectra)

the natural class of records (NAT) selected by PEER (2012) and artificial accelerograms (ART) generated according to the procedure described below in this section. The response spectra for the NAT class of records were obtained after scaling the records to the same $S_A(T_1, 5\%)$, where $T_1 = 0.628$ s for the 40/50 hazard level, in accordance with the hazard curve of the city of San Diego, California. Comparing NAT and ART class of accelerograms, it can be seen that the dispersion of the response spectra for the ART class of accelerograms is lower than that of the NAT class.

Intensity Measures

Earthquake engineering is a scientific field strongly connected with the effect of strong ground motions on people and environment. Strong ground motions are extremely complicated and a lot of data is required for their full description. The definition of a number of ground motion parameters, namely, intensity measures (IMs), simplifies the description of a strong ground motion and links the seismic hazard with the structural data required for the solution of earthquake engineering problems. The most significant characteristics of a ground motion from an earthquake engineering point of view are the frequency content, the amplitude, and the motion duration. Some of the IMs are related to amplitude or to frequency content, to duration, or to more than one of the three essential ground motion characteristics. The IMs that are related

to the effect of more than one ground motion characteristics are considered more reliable for the description of a ground motion and are most suitable to reflect the potential damage that a ground motion can produce.

The most commonly used *amplitude* IMs derived from an accelerogram are the peak ground acceleration (high-frequency component), peak velocity (intermediate-frequency component), peak displacement (low-frequency component), sustained maximum acceleration and velocity, and the effective design acceleration. Amplitude IMs are used for the derivation of empirical attenuation relationships used in probabilistic hazard analysis, because their production is based on IMs' dependence, on the magnitude of the earthquake, and on the site-to-source distance. *Frequency content* IMs describe through different types of spectra how the amplitude of a ground motion is distributed among different frequencies. IMs related to frequency content are Fourier spectra and power spectra that correspond to the frequency content of the ground motion itself and the response spectra that correspond to the influence of the ground motion on structures with different first eigenperiods. Among these types of IMs are spectral parameters, like the predominant period, bandwidth, central frequency, shape factor, Kanai-Tajimi parameters, and the ratio $v_{\max}/a_{\max}$ which describes the frequency content of a ground motion. The most commonly used *duration* IM is the bracketed duration which is defined as the

time between the first and the last exceedance of a threshold acceleration, usually equal to 0.05 g.

Arias intensity (I_A), *characteristic intensity* (I_C), and *cumulative absolute velocity* (CAV) are three IMs that reflect the amplitude, the frequency content, and the duration of a strong ground motion, respectively, which correlate well with structural damage. *Arias intensity* is defined as the time integral of the square of the ground acceleration

$$I_A = \frac{\pi}{2g} \int_0^{\infty} [a(t)]^2 dt \quad (1)$$

where $a(t)$ is the ground acceleration and g is the acceleration due to gravity (9.81 m/s²) and is expressed in units of velocity, i.e., meter per second. The symbol of infinity in the time integration means that I_A is calculated over the entire duration and not over the duration of the strong ground motion T_d that is defined through specific methods.

Characteristic intensity is defined as

$$I_C = \alpha_{rms}^{1.5} T_d^{0.5} \quad (2)$$

where α_{rms} is the *rms acceleration* (root mean square acceleration) ground motion parameter and T_d is the duration of the strong motion.

Cumulative absolute velocity is defined as the integral of the absolute acceleration in a time history, and it is obtained through the following equation:

$$CAV = \int_0^{T_d} |a(t)| dt \quad (3)$$

where $|a(t)|$ is the absolute value of the acceleration time series at time t and T_d is the duration of the strong motion. CAV expresses the absolute area under the absolute accelerogram and corresponds to the cumulative absolute velocity that is well correlated with structural damage (Kramer 1996). These three IMs are used for investigating their dispersion with reference to the two classes

of seismic excitations considered. In particular, the *characteristic intensity* is a measure that is related to an index that quantifies the damage due to the maximum deformation and the absorbed hysteretic energy. In order to examine the influence of the class of seismic excitations considered on the values of the three IMs, the box plots depicted in Fig. 2 have been generated for the 40/50 hazard level (denoting 40 % probability of exceedance in 50 years).

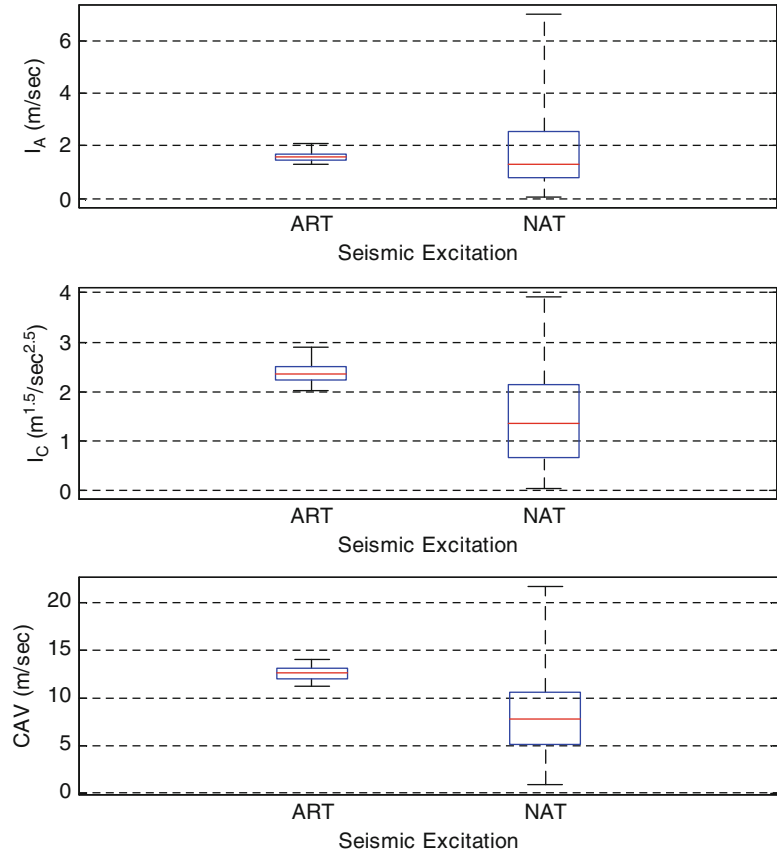
For each IM considered and for each class of seismic excitation, a box plot was created. On each box, the central mark denoted with a bold black line is the median value of the IM in question. The edges of the box represent the 25 % and 75 % percentiles, while the whiskers denoted with black lines extend to the most extreme data points, i.e., represent the range of the IM values. For all three IMs, ART class of accelerograms has the narrower range and confidence bounds. On the other hand, the median values of ART are close to each other while NAT varies significantly for the case of I_C and CAV. For the case of I_A , the three median values are quite close. For the 25 % and 75 % percentiles, I_C and CAV, NAT class of seismic excitations, have similar box sizes denoting that they are related to the length of the range of values.

Analysis Methods

Having an appropriate model for the structure and its earthquake loads, structural analysis is the next step. The choice of linear elastic or nonlinear modeling of the element force-deformation behavior and the consideration, or not, of geometric nonlinearities essentially dictate the selection of the appropriate analysis method. Modeling of structural mass distribution and the dynamic nature of earthquake loads allows the use of dynamic rather than static methods of analysis. Thus, broadly speaking, four categories of analysis methods are available, each with its own modeling and solution algorithm requirements: (a) nonlinear dynamic, (b) nonlinear static, (c) elastic dynamic, and (d) elastic static (ASCE 41-06 2007).

Time History Seismic

Analysis, Fig. 2 Mean and 25 % and 75 % confidence bounds for the values of I_A , I_C , and CAV obtained for the classes of records considered



Linear Time History Analysis Methods

Although simple elastic linear methods may seem antiquated by modern research standards, they define, by overwhelming majority, the current standard of practice for seismic design. Despite the many efforts to offer building a design methodology on a nonlinear static or dynamic basis (Fib 2012), the vast majority of practical earthquake engineering work is still done at the elastic level using either (a) equivalent linear static, whereby a fixed lateral load pattern is applied to an elastic structural model, or (b) linear time history analysis and (c) modal response spectrum analysis, where the modal responses are combined to estimate the peak MDOF response (e.g., EN1998 2005). Inelastic response is universally taken care of by dividing the elastic seismic loads (i.e., design spectral acceleration values) by the appropriate reduction R (or behavior q) factor that is meant to represent the ductility and

overstrength of a yielding system. Although little recent research has been directed in the way of elastic methods, recent advancements in nonlinear analysis have helped to shed some light into the premise of using elastic results to capture nonlinear behavior.

The equilibrium equations for a system in motion can be written as

$$\mathbf{M}\ddot{\mathbf{u}}(t) + \mathbf{C}\dot{\mathbf{u}}(t) + \mathbf{K}\mathbf{u}(t) = \mathbf{R}(t) \quad (4)$$

where $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ are the mass, damping, and stiffness matrices and $\mathbf{R}(t)$ is the external load vector, while $\mathbf{u}(t)$, $\dot{\mathbf{u}}(t)$, and $\ddot{\mathbf{u}}(t)$ are the displacement, velocity, and acceleration vectors of the finite element assemblage, respectively. The choice for a static or dynamic analysis (i.e., for including or neglecting velocity and acceleration-dependent forces in the analysis) is usually decided by engineering judgment.

Mathematically, Eq. 4 represents a system of linear differential equations of second order and the solution of this system can be obtained by standard procedures for the solution of differential equations. In practical finite element analysis, we are mainly interested in a few effective methods and we will concentrate in the next sections on the presentation of those techniques and in particular on the direct integration ones. In direct integration the system of linear differential equations in Eq. 4 is integrated using a numerical step-by-step procedure; the term “direct” means that no transformation of the equations is carried out prior to the numerical integration.

Central Difference Method

One procedure that can be very effective in the solution of certain problems is the central difference method where it is assumed that

$$\ddot{\mathbf{u}}(t) = \frac{1}{\Delta t^2} [\mathbf{u}(t - \Delta t) - 2\mathbf{u}(t) + \mathbf{u}(t + \Delta t)] \quad (5)$$

and the velocity expansion is defined by

$$\dot{\mathbf{u}}(t) = \frac{1}{2\Delta t} [\mathbf{u}(t + \Delta t) - \mathbf{u}(t - \Delta t)] \quad (6)$$

The displacement expression for time $t + \Delta t$ is obtained by substituting Eqs. 5 and 6 into Eq. 4:

$$\begin{aligned} \left(\frac{1}{\Delta t^2} \mathbf{M} + \frac{1}{2\Delta t} \mathbf{C} \right) \mathbf{u}(t + \Delta t) = \\ \mathbf{R}(t) - \left(\mathbf{K} - \frac{2}{\Delta t^2} \mathbf{M} \right) \mathbf{u}(t) \\ - \left(\frac{1}{\Delta t^2} \mathbf{M} - \frac{1}{2\Delta t} \mathbf{C} \right) \mathbf{u}(t - \Delta t) \end{aligned} \quad (7)$$

from which we can solve for $\mathbf{u}(t + \Delta t)$. It should be noted that the solution of $\mathbf{u}(t + \Delta t)$ is based on using the equilibrium conditions of

Eq. 4 at time t . For this reason the integration procedure is called an explicit integration method. On the other hand, the Houbolt, Wilson, and Newmark methods, considered in the next sections, use the equilibrium conditions at time $t + \Delta t$ and are called implicit integration methods.

Houbolt Method

The Houbolt integration scheme is somewhat related to the central difference method in the sense that standard finite difference expressions are used to approximate the acceleration and velocity components in terms of displacement. The following expansions are employed in the Houbolt (1950) integration method:

$$\ddot{\mathbf{u}}(t + \Delta t) = \frac{1}{\Delta t^2} [2\mathbf{u}(t + \Delta t) - 5\mathbf{u}(t) + 4\mathbf{u}(t - \Delta t) - \mathbf{u}(t - 2\Delta t)] \quad (8)$$

and

$$\dot{\mathbf{u}}(t + \Delta t) = \frac{1}{6\Delta t} [11\mathbf{u}(t + \Delta t) - 18\mathbf{u}(t) + 9\mathbf{u}(t - \Delta t) - 2\mathbf{u}(t - 2\Delta t)] \quad (9)$$

In order to obtain the displacement $\mathbf{u}$ at time $t + \Delta t$, we now consider Eq. 4 at time $t + \Delta t$, which gives

$$\mathbf{M}\ddot{\mathbf{u}}(t + \Delta t) + \mathbf{C}\dot{\mathbf{u}}(t + \Delta t) + \mathbf{K}\mathbf{u}(t + \Delta t) = \mathbf{R}(t + \Delta t) \quad (10)$$

Substituting Eqs. 8 and 9 into Eq. 10 and arranging all known vectors on the right-hand side of the equations, we obtain

$$\begin{aligned} \left(\frac{2}{\Delta t^2} \mathbf{M} + \frac{11}{6\Delta t} \mathbf{C} + \mathbf{K} \right) \mathbf{u}(t + \Delta t) = \mathbf{R}(t + \Delta t) + \left(\frac{5}{\Delta t^2} \mathbf{M} + \frac{3}{\Delta t} \mathbf{C} \right) \mathbf{u}(t) \\ - \left(\frac{4}{\Delta t^2} \mathbf{M} + \frac{3}{2\Delta t} \mathbf{C} \right) \mathbf{u}(t - \Delta t) + \left(\frac{1}{\Delta t^2} \mathbf{M} + \frac{1}{3\Delta t} \mathbf{C} \right) \mathbf{u}(t - 2\Delta t) \end{aligned} \quad (11)$$

As shown in Eq. 11, the solution of $\mathbf{u}(t + \Delta t)$ requires knowledge of $\mathbf{u}(t)$, $\mathbf{u}(t - \Delta t)$, and $\mathbf{u}(t - 2\Delta t)$.

Wilson- θ Method

The Wilson- θ method is essentially an extension of the linear acceleration method, where linear variation of the acceleration from time t to time $t + \Delta t$ is assumed. In the Wilson- θ method, the acceleration is assumed to be linear from time t to time $t + \theta\Delta t$, where $\theta > 1$ (Wilson et al. 1973). Let τ denote the increase in time, where $0 \leq \tau \leq \theta\Delta t$; then for the time interval t to $t + \theta\Delta t$, it is assumed that

$$\ddot{\mathbf{u}}(t + \tau) = \ddot{\mathbf{u}}(t) + \frac{\tau}{\theta\Delta t} [\ddot{\mathbf{u}}(t + \theta\Delta t) - \ddot{\mathbf{u}}(t)] \quad (12)$$

Integrating Eq. 12, we obtain

$$\dot{\mathbf{u}}(t + \tau) = \dot{\mathbf{u}}(t) + \ddot{\mathbf{u}}(t)\tau + \frac{\tau^2}{2\theta\Delta t} [\ddot{\mathbf{u}}(t + \theta\Delta t) - \ddot{\mathbf{u}}(t)] \quad (13)$$

and

$$\begin{aligned} \mathbf{u}(t + \tau) = & \mathbf{u}(t) + \dot{\mathbf{u}}(t)\tau + \frac{1}{2}\ddot{\mathbf{u}}(t)\tau^2 \\ & + \frac{\tau^3}{6\theta\Delta t} [\ddot{\mathbf{u}}(t + \theta\Delta t) - \ddot{\mathbf{u}}(t)] \end{aligned} \quad (14)$$

Using Eqs. 13 and 14, we obtain for the time $t + \theta\Delta t$

$$\dot{\mathbf{u}}(t + \theta\Delta t) = \dot{\mathbf{u}}(t) + \frac{\theta\Delta t}{2} [\ddot{\mathbf{u}}(t + \theta\Delta t) - \ddot{\mathbf{u}}(t)] \quad (15)$$

$$\begin{aligned} \mathbf{u}(t + \theta\Delta t) = & \mathbf{u}(t) + \dot{\mathbf{u}}(t)\theta\Delta t \\ & + \frac{\theta^2\Delta t^2}{6} [\ddot{\mathbf{u}}(t + \theta\Delta t) + 2\ddot{\mathbf{u}}(t)] \end{aligned} \quad (16)$$

from which we can solve for $\ddot{\mathbf{u}}(t + \theta\Delta t)$ and $\dot{\mathbf{u}}(t + \theta\Delta t)$ in terms of $\mathbf{u}(t + \theta\Delta t)$:

$$\dot{\mathbf{u}}(t + \theta\Delta t) = \frac{3}{\theta\Delta t} [\mathbf{u}(t + \theta\Delta t) - \mathbf{u}(t)] - 2\dot{\mathbf{u}}(t) - \frac{\theta\Delta t}{2}\ddot{\mathbf{u}}(t) \quad (17)$$

$$\begin{aligned} \ddot{\mathbf{u}}(t + \theta\Delta t) = & \frac{6}{\theta^2\Delta t^2} [\mathbf{u}(t + \theta\Delta t) - \mathbf{u}(t)] \\ & - \frac{6}{\theta\Delta t}\dot{\mathbf{u}}(t) - 2\ddot{\mathbf{u}}(t) \end{aligned} \quad (18)$$

To obtain the solution for the displacements, velocities, and accelerations at time $t + \Delta t$, the equilibrium equations given in Eq. 4 are considered at time $t + \theta\Delta t$. However, since accelerations are assumed to vary linearly, a linearly extrapolated load vector is used:

$$\begin{aligned} \mathbf{M}\ddot{\mathbf{u}}(t + \theta\Delta t) + \mathbf{C}\dot{\mathbf{u}}(t + \theta\Delta t) \\ + \mathbf{K}\mathbf{u}(t + \theta\Delta t) = \bar{\mathbf{R}}(t + \theta\Delta t) \end{aligned} \quad (19)$$

where

$$\bar{\mathbf{R}}(t + \theta\Delta t) = \mathbf{R}(t) + \theta[\mathbf{R}(t + \Delta t) - \mathbf{R}(t)] \quad (20)$$

Substituting Eqs. 17 and 18 into Eq. 19, an equation is obtained from which $\mathbf{u}(t + \theta\Delta t)$ can be solved. Then substituting $\mathbf{u}(t + \theta\Delta t)$ into Eq. 18, we obtain $\ddot{\mathbf{u}}(t + \theta\Delta t)$, which is used in Eqs. 12, 13, and 14, all evaluated at time $\tau = \Delta t$ to calculate $\ddot{\mathbf{u}}(t + \Delta t)$, $\dot{\mathbf{u}}(t + \Delta t)$ and $\mathbf{u}(t + \Delta t)$.

Newmark Method

The Newmark integration scheme is an extension of the linear acceleration method. Under this scheme the variations of velocity and displacement are given by (Newmark 1959)

$$\dot{\mathbf{u}}(t + \Delta t) = \dot{\mathbf{u}}(t) + [(1 - \delta)\ddot{\mathbf{u}}(t) + \delta\ddot{\mathbf{u}}(t + \Delta t)]\Delta t \quad (21)$$

$$\begin{aligned} \mathbf{u}(t + \Delta t) = & \mathbf{u}(t) + \dot{\mathbf{u}}(t)\Delta t \\ & + [(0.5 - \alpha)\ddot{\mathbf{u}}(t) + \alpha\ddot{\mathbf{u}}(t + \Delta t)]\Delta t^2 \end{aligned} \quad (22)$$

where α and δ are parameters that can be determined to obtain integration accuracy and stability. When $\delta = 1/2$ and $\alpha = 1/6$, Eqs. 21 and 22 correspond to the linear acceleration method. In addition to Eqs. 21 and 22, for calculating the displacements, velocities, and accelerations at

time $t + \Delta t$, the equilibrium Eq. 4 is also considered at time $t + \Delta t$:

$$\mathbf{M}\ddot{\mathbf{u}}(t + \Delta t) + \mathbf{C}\dot{\mathbf{u}}(t + \Delta t) + \mathbf{K}\mathbf{u}(t + \Delta t) = \mathbf{R}(t + \Delta t) \quad (23)$$

Solving from Eq. 22 for $\ddot{\mathbf{u}}(t + \Delta t)$ in terms of $\mathbf{u}(t + \Delta t)$ and then substituting for $\ddot{\mathbf{u}}(t + \Delta t)$ into Eq. 21, we obtain equations for $\ddot{\mathbf{u}}(t + \Delta t)$ and $\dot{\mathbf{u}}(t + \Delta t)$ each in terms of the unknown displacements $\mathbf{u}(t + \Delta t)$. These two relations for $\ddot{\mathbf{u}}(t + \Delta t)$ and $\dot{\mathbf{u}}(t + \Delta t)$ are substituted into Eq. 23 to solve for $\mathbf{u}(t + \Delta t)$ after using Eqs. 21 and 22; $\ddot{\mathbf{u}}(t + \Delta t)$ and $\dot{\mathbf{u}}(t + \Delta t)$ can also be calculated. As a result of this substitution, the following well-known equilibrium equation is obtained at each Δt :

$$\mathbf{K}_{\text{eff}}\mathbf{u}(t + \Delta t) = \mathbf{R}^{\text{eff}}(t + \Delta t) \quad (24)$$

where

$$\mathbf{K}_{\text{eff}} = \mathbf{K} + \alpha_0\mathbf{M} + \alpha_1\mathbf{C} \quad (25)$$

$$\begin{aligned} \mathbf{R}^{\text{eff}}(t + \Delta t) = & \mathbf{R}(t + \Delta t) \\ & + \mathbf{M}[\alpha_0\mathbf{u}(t) + \alpha_2\dot{\mathbf{u}}(t) + \alpha_3\ddot{\mathbf{u}}(t)] \\ & + \mathbf{C}[\alpha_1\mathbf{u}(t) + \alpha_4\dot{\mathbf{u}}(t) + \alpha_5\ddot{\mathbf{u}}(t)] \end{aligned} \quad (26)$$

with $\delta \geq 0.50$; $\delta \geq 0.25(0.5 + \delta)^2$, while

$$\begin{aligned} \alpha_0 &= \frac{1}{\alpha\Delta t^2}, \alpha_1 = \frac{\delta}{\alpha\Delta t}, \alpha_2 = \frac{1}{\alpha\Delta t}, \alpha_3 = \frac{1}{2\alpha} - 1, \\ \alpha_4 &= \frac{\delta}{\alpha} - 1, \\ \alpha_5 &= \frac{\Delta t}{2} \left(\frac{\delta}{\alpha} - 2 \right), \alpha_6 = \Delta t(1 - \delta), \alpha_7 = \delta\Delta t. \end{aligned}$$

Nonlinear Time History Analysis

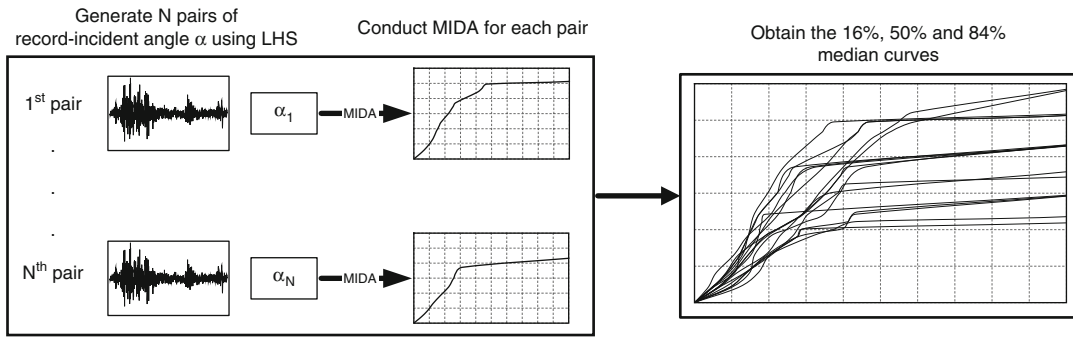
Nonlinear dynamic analysis is undoubtedly the most realistic and accurate analysis method available. It is also referred as “nonlinear time history analysis,” “nonlinear response history analysis,” or according to ASCE 41-06 (2007) as “nonlinear dynamic procedure” (NDP). Earthquake loading is taken into consideration as a natural or a synthetic ground motion on a structural model

that incorporates elements with inelastic (inelastic and nonlinear are typically used interchangeably for seismic applications) force-deformation relationships and at least a first-order approximation of geometric nonlinearities (P-Delta effects). The propagation of the ground motion throughout the structure generates complete response histories for any quantity of interest (e.g., displacements, stress resultants) leading to a wealth of data. While different levels of complexity are possible due to modeling choices, different ground motion records will produce demands that vary considerably. This record-to-record variation dominates the application of dynamic methods. Therefore, in order to get reliable response estimates, an appropriately selected set of several ground motion records is necessary, since a single time history analysis is of limited practical use. Furthermore, one or more levels of seismic intensity, typically corresponding to one or more levels of the IM, may need to be employed to investigate structural behavior at different regimes of response or damage (e.g., elastic, post-yield, or near-collapse). Thus, nonlinear dynamic analysis procedures may be categorized into the narrow- and broad-range assessment categories.

In most practical design/assessment situations, only a narrow, single-point estimate of structural response is required. This is consistent with current seismic codes that only provide a design hazard spectrum (typically at an exceedance probability of 10 % in 50 years) and require checking that a structure shall not sustain significant or life-threatening damage at such a level of intensity. Thus, seismic codes (e.g., ASCE 7-10 2010; EN1998 2005) prescribe using ground motion records that match or exceed the design spectrum in the period range of interest. If 3–6 records are employed, the overall maximum of the recorded peak responses is taken as the structural demand, while for 7 records and above, the mean of the peak responses can be employed.

Incremental Analysis Procedures

In the framework of seismic assessment of structures, a wide range of seismic records and more than one hazard levels should be considered in



Time History Seismic Analysis, Fig. 3 The MIDA procedure

order to take into account the uncertainties inherent in any seismic hazard for assessing the performance of a structure under seismic actions. The most reliable but also computationally intensive methodology for the performance-based assessment implements nonlinear dynamic analyses based on a single- or a multiple-hazard-level approach. In fragility analysis of structural systems or life-cycle cost analysis, the two most appropriate methods for performing this task are the multiple-stripe dynamic analysis and incremental dynamic analysis (Vamvatsikos and Cornell 2002). To be consistent with the current technical literature, IDA abbreviation is used for both methods. The main objective of an IDA study is to define a curve through the relation of the intensity level with the maximum seismic response of the structural system. The intensity level and the seismic response are described through an intensity measure (IM) and an engineering demand parameter (EDP), respectively.

Selecting IM and EDP is one of the most important steps of IDA. In Giovenale et al. (2004), the significance of selecting an efficient IM is discussed. The IM should be a monotonically scalable ground motion intensity measure, like the peak ground acceleration (PGA), peak ground velocity (PGV), the $\xi = 5\%$ damped spectral acceleration at the structure's first-mode period ($S_A(T_1, 5\%)$), as well as others. On the other hand, the damage may be quantified by using any of the EDPs whose values can be related to particular structural damage states. A number of available response-based EDPs were discussed and

critically evaluated in the past for their applicability in seismic damage evaluation. In the work by Ghobarah et al. (1999), the EDPs are classified into four categories based on maximum deformation, cumulative damage, and combinations of maximum deformation and cumulative damage. According to the multicomponent incremental dynamic analysis (MIDA) proposed by Lagaros (2010) which is an extension of IDA, a set of natural records, each one represented by its longitudinal and transverse components, are applied to the structure in order to account for the randomness on the seismic excitation. The difference of the MIDA framework from the original one-component version of the IDA stems from the fact that for each record, a number of MIDA representative curves can be defined depending on the incident angle selected, while in most cases of the one-component version of IDA, only one IDA representative curve is obtained. A schematic representation of MIDA procedure can be seen in Fig. 3.

Use of Time History Seismic Analyses

Structural Models and Numerical Simulation

Nonlinear analyses need detailed simulation of the structure in the regions where inelastic deformations are expected to develop. Either the plastic-hinge or the fiber approach can be adopted for this cause. Given that the plastic-hinge approach has limitations in terms of accuracy, the fiber beam-column elements are preferable (Fragiadakis and Papadrakakis 2008).

According to the fiber approach, every beam-column element has a number of integration sections, each divided into fibers. Each fiber in the section can be assigned concrete, structural steel, or reinforcing bar material properties. The sections are located either at the center of the element or at its Gaussian integration points. The main advantage of the fiber approach is that every fiber has a simple uniaxial material model allowing an easy and efficient implementation of the inelastic behavior. This approach is considered to be suitable for inelastic beam-column elements under dynamic loading and provides a reliable solution compared to other formulations. However, it results to higher computational demands in terms of memory storage and CPU time. When a displacement-based formulation is adopted, the discretization should be adaptive with a dense mesh at the joints and a single elastic element for the remaining part of the member. On the other hand, force-based fiber elements allow modeling a member with a single beam-column element.

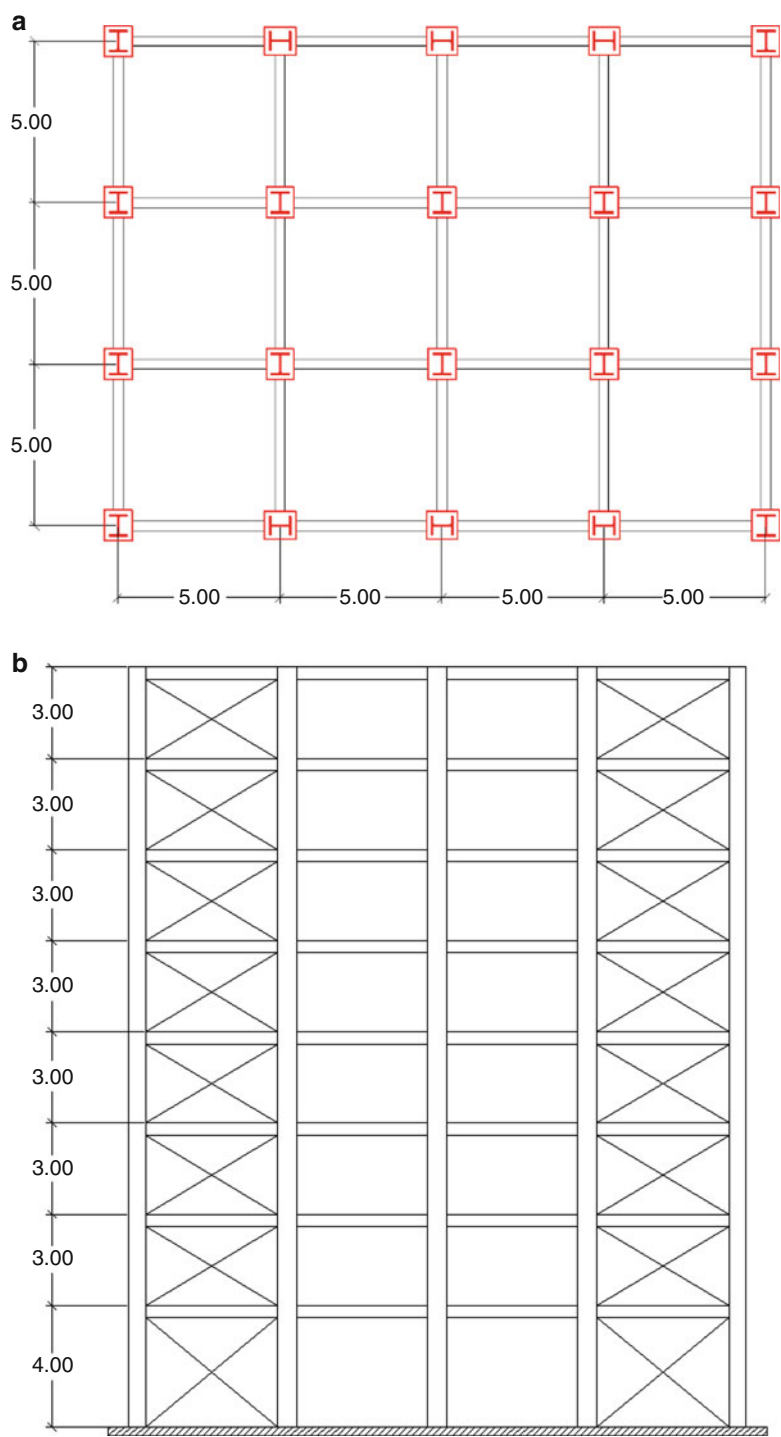
The two multistory buildings shown in Figs. 4 and 5 are considered for the numerical tests presented herein. The first one is an eight-story steel-RC composite building and the second test example is an eight-story RC building having symmetrical plan view, while the second one is a five-story RC building with nonsymmetrical plan view. The building has been designed for minimum initial cost (Mitropoulou et al. 2010). Concrete of class C20/25 (nominal cylindrical strength of 20 MPa) and steel of class S500 (nominal yield stress of 500 MPa) are assumed. Each member is modeled with a single force-based, fiber beam-column element. The modified Kent and Park (1971) models are employed for the simulation of the concrete fibers. This model was chosen because it allows for an accurate prediction of the structural demand for flexure-dominated RC members despite its relatively simple formulation. The inelastic behavior of the reinforcing bars was simulated with the Menegotto and Pinto (1973) model. A bilinear material model with pure kinematic hardening is adopted for the steel fibers, while geometric nonlinearity is explicitly taken into consideration.

Critical Orientation of the Seismic Incidence

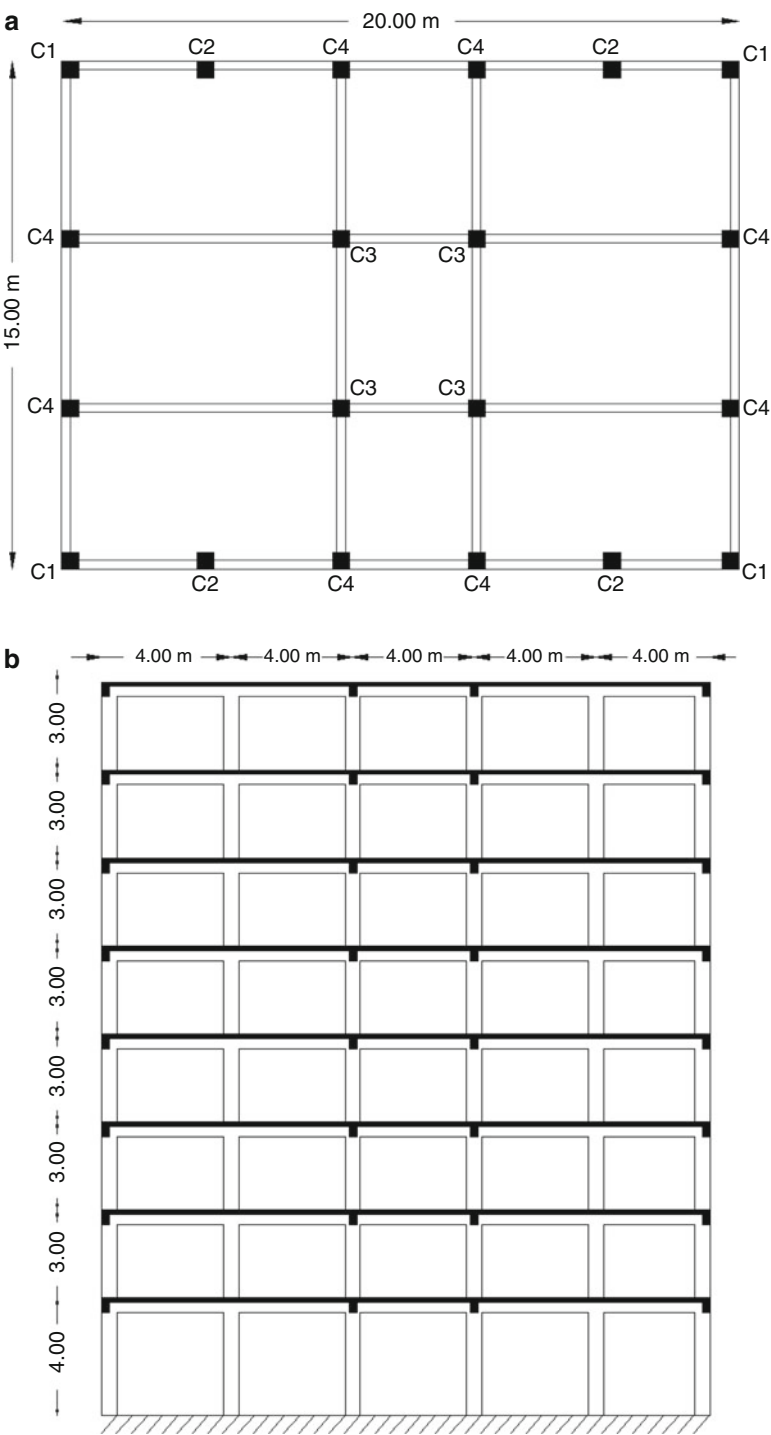
A structure subjected to the simultaneous action of two orthogonal horizontal ground accelerations along the directions Ow and Op is illustrated in Fig. 6. The orthogonal system $Oxyz$ defines the reference axes of the structure (structural axes). The angle defined with a counterclockwise rotation of the structural axis Ox to coincide with the ground motion axis Ow is called as incident angle of the record.

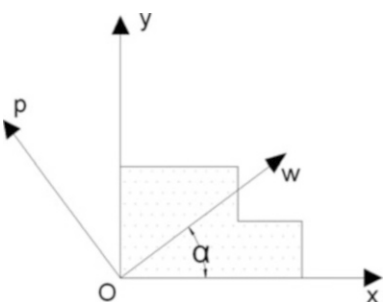
According to Penzien and Watabe (1975), the orthogonal directions of a ground motion can be considered uncorrelated in the principal directions of the structure. This finding was the basis for many researchers in order to define the orientation that yields the maximum response when the response spectrum dynamic analysis was applied. In the work by Wilson et al. (1995), an alternative code-approved method that results in structural designs which have equal resistance to seismic motions from all directions was proposed. Lopez and Torres (1997) proposed a simple method which can be employed by the seismic codes to determine the critical angle of seismic incidence and the corresponding peak response of structures subjected to two horizontal components applied along any arbitrary directions and to the vertical component of earthquake ground motion. The CQC3 response spectrum rule for combining the three orthogonal components of ground motion to the maximum value of a response quantity is presented in the work by Menun and Der Kiureghian (1998). In the work by Lopez et al. (2000), it is proposed an explicit formula, convenient for code applications, in order to calculate the critical value of the structural response to the two principal horizontal components acting along any incident angle with respect to the structural axes. In the work by Menun and Der Kiureghian (2000), a response spectrum-based procedure for predicting the envelope that bounds two or more responses in a linear structure is developed. In the work by Anastassiadis et al. (2002), a seismic design procedure for structures is proposed where the three translational ground motion components on a specific point of the ground are statistically noncorrelated.

Time History Seismic Analysis, Fig. 4 Eight-story steel-RC composite: (a) plan view and (b) front view



Time History Seismic Analysis, Fig. 5 Eight-story RC building: (a) plan view and (b) front view





Time History Seismic Analysis, Fig. 6 Definition of the incident angle α

There are only few studies in the literature where the case of the critical incident angle is examined when time history analysis is employed. In these studies it was found that in the most general case, where nonlinear behavior is encountered, it is not an easy task to define the critical angle. In the work by MacRae and Mattheis (2000), it is shown the ability of the 30 % SRSS rule and the sum of absolute values methods to assess building drifts for bidirectional shaking effects, while it is also shown that the response is dependent on the reference axes chosen. MacRae and Tagawa (2001) have found that design-level shaking caused the structure to exceed story yield drifts in both directions simultaneously and significant column yielding occurred above the base. Shaking a structure in the direction orthogonal to the main shaking direction increased drifts in the main shaking direction, indicating that 2D analyses would not estimate the 3D response well. Ghersi and Rossi (2001) examined the influence of bidirectional seismic excitations on the inelastic behavior of in-plan irregular systems having one symmetry axis where it was found that in most cases the adoption of Eurocode 8 provisions to combine the effects of the two seismic components allows the limitation of the orthogonal element ductility demand. In the work by Athanatopoulou (2005), analytical formulae were developed for determining the critical incident angle and the corresponding maximum value of a response quantity of structures subjected to three seismic

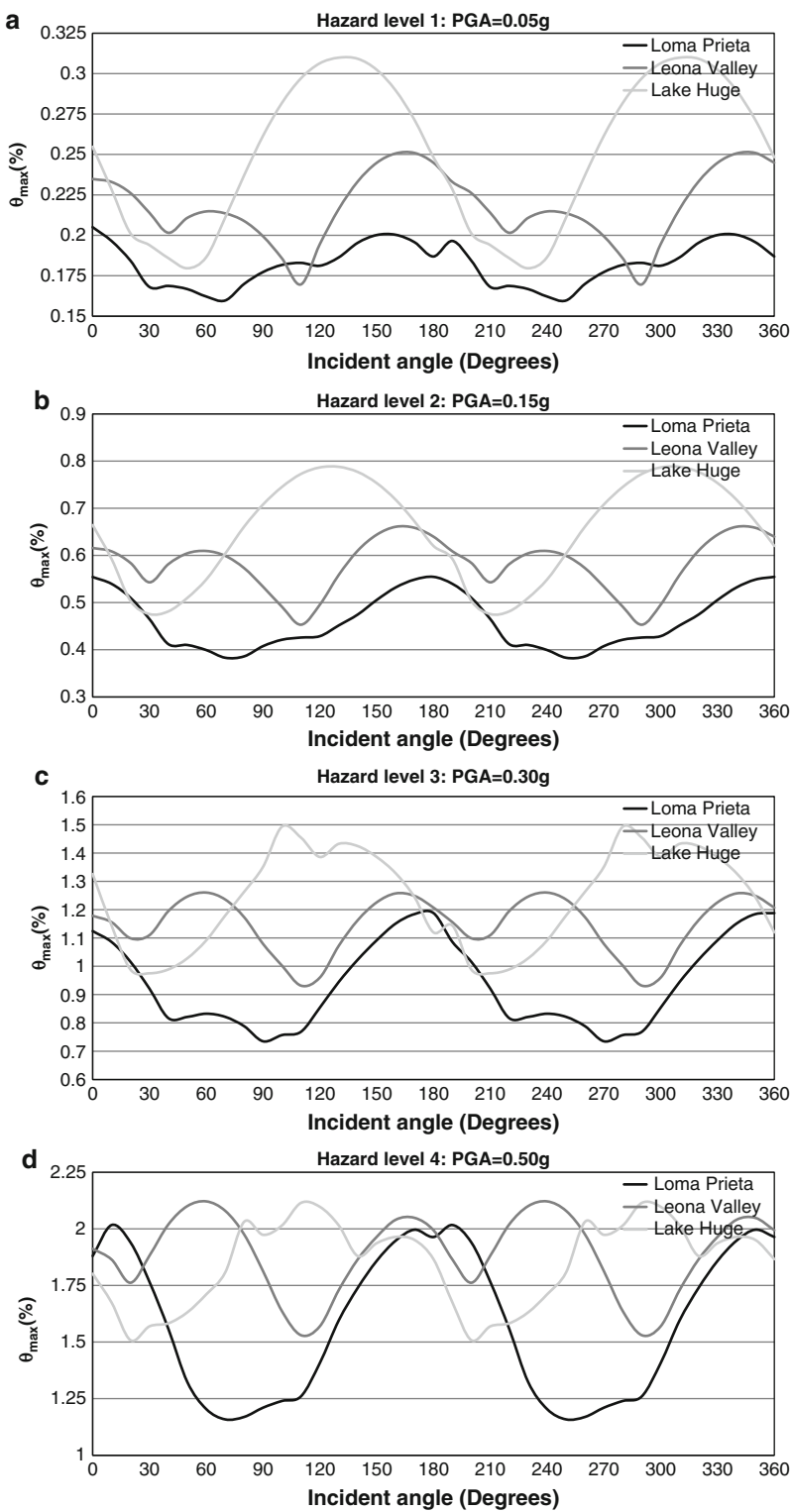
correlated components. Rigato and Medina (2007) studied a number of symmetrical and asymmetrical structures having fundamental periods ranging from 0.2 to 2.0 s where the influence that the incident angle of the ground motion has on several engineering demand parameters was examined.

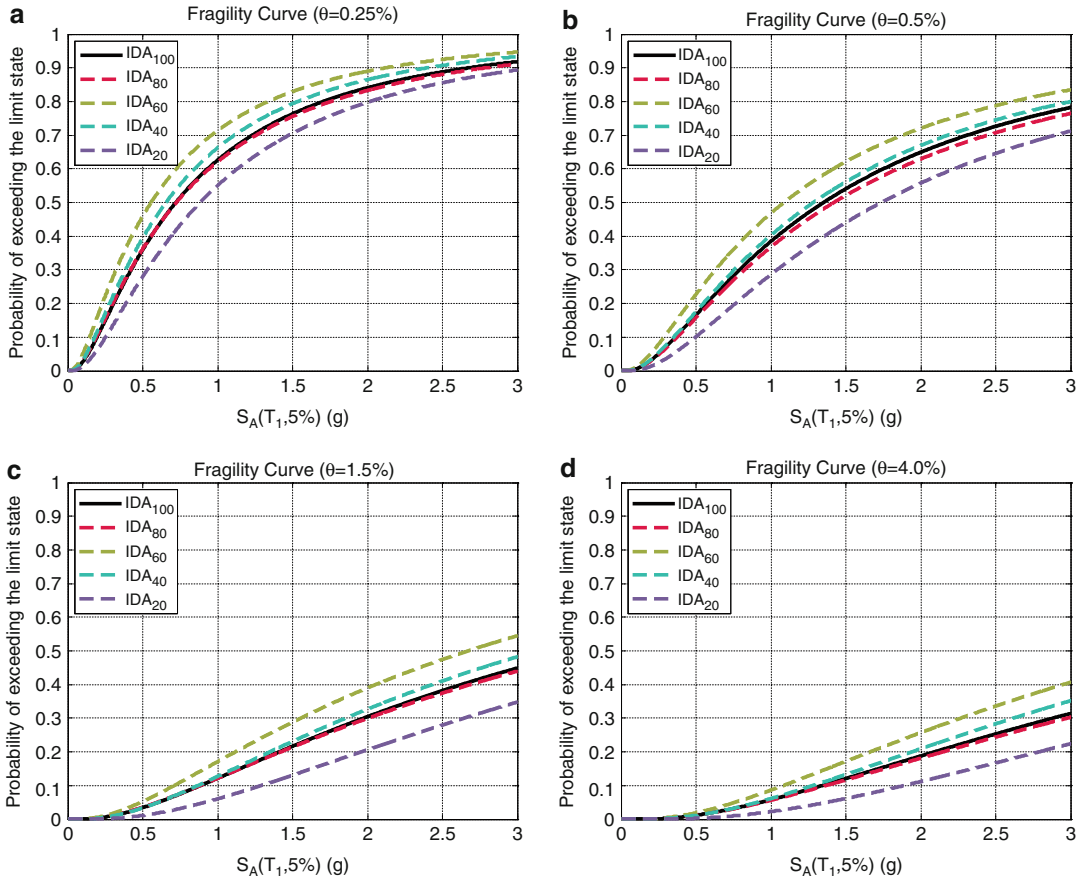
In order to examine the influence of the incident angle on the seismic response of the structure, three records have been selected at random and are applied to the steel-RC composite building of Fig. 4. The three records considered are the Loma Prieta, the Leona Valley, and the Lake Huger. The three records have been applied considering a varying incident angle in the range of 0–360° with a step of 5°. In order to examine the influence of the incident angle on the maximum interstory drift to different intensity levels, the three records have been scaled with respect to the 5 % damped spectral acceleration at the structure's first-mode period to 0.05, 0.15, 0.30, and 0.50 g, and the maximum interstory drift has been recorded for all the incident angles and the intensity levels considered. The variation of the maximum interstory drift with respect to the incident angle and the intensity level for the three records is depicted in Fig. 7. As it can be seen, the seismic response when the incident angle varies in the range of 0–180° almost coincides with the seismic response when the incident angle varies in the range of 185–360°. This is because the relative ratio of the two horizontal components of the records is close to one; thus, the two components are scaled to almost the same intensity level.

Fragility Analysis

In the second part of this numerical investigation, four-limit-state fragility curves are obtained for the RC building of Fig. 5. The limit states considered are defined by means of maximum drift values and cover the whole range of structural damage from serviceability, to life safety, and finally to the onset of collapse. The following $\theta_{\max}$ values are chosen, according to HAZUS (2003), for each of the four limit states: 0.25 %, 0.50 %, 1.50 %, and 4.0 %. For each limit state,

Time History Seismic Analysis, Fig. 7 Eight-story steel-RC composite test example – maximum interstory drift with respect to the incident angle of the record scaled to (a) 0.05 g, (b) 0.15 g, (c) 0.30 g, and (d) 0.50 g intensity levels





Time History Seismic Analysis, Fig. 8 Eight-story test example – fragility curves for four limit states using alternative number of records

the IDA and “vertical” statistics are implemented for computing the statistical parameters μ and β . “Vertical” statistics is performed for all four limit states described above. For the purposes of the present investigation, five cases have been examined for the calculation of the two parameters (μ_k , β_k , $k = 1, \dots, 4$), IDA_{20} , IDA_{40} , IDA_{60} , IDA_{80} , and IDA_{100} , where 20–100 records are implemented.

Figure 8 depicts the fragility curves corresponding to limit states that cover the whole range of structural damage developed for two buildings considered. The fragility curves corresponding to IDA_{100} are considered as the “correct” ones. For the case of the eight-story symmetric building, it can be seen that for all

limit states, IDA_{20} overestimate the structural capacity, while in the cases of IDA_{40} and IDA_{60} , the structural capacity is underestimated. Furthermore, it can be seen that 80 records provide a good estimate of the two parameters μ_k and β_k , since the fragility curves of IDA_{80} almost coincide with those of IDA_{100} case. Therefore, it can be said that IDA_{80} case is a good compromise between robustness and efficiency.

Some Remarks

The choice for a static or dynamic analysis is usually based on engineering judgment. The use of static analysis, however, should be justified

since otherwise the analysis results are meaningless. In particular, when nonlinearities are taken into account, the assumption of neglecting inertia and damping forces may be severe. In the last decade many efforts to formulate a design methodology based on a nonlinear static or dynamic basis have been done. In the seismic design and assessment of structures, a wide range of seismic records and more than one performance levels should be considered in order to take into account the uncertainties that the seismic hazard introduces into a performance-based seismic assessment or design problem. In the construction industry, decision making for structural systems situated in seismically active regions requires consideration of the cost of damage and other losses resulting from earthquakes which may occur during the life span of the structures. In this entry an overview of the time history seismic analysis procedures is provided along with indicative application of such procedures.

Summary

In the case of design or assessment purposes, the seismic capacity of the structural system should be computed for different seismic hazard levels by means of linear or nonlinear time history seismic analyses using a number of properly selected ground motions. The selection of the ground motions plays an important role in the efficiency and accuracy of the design or assessment procedure. In this entry, we provide an overview of the time history seismic analysis applied for design and assessment purposes.

Cross-References

- [Analytic Fragility and Limit States \[P\(EDPI IM\)\]: Nonlinear Dynamic Procedures](#)
- [Assessment of Existing Structures Using Response History Analysis](#)
- [Incremental Dynamic Analysis](#)
- [Nonlinear Dynamic Seismic Analysis](#)

- [Seismic Actions Due to Near-Fault Ground Motion](#)
- [Selection of Ground Motions for Response History Analysis](#)

References

- Anastassiadis K, Avramidis I, Panetsos P (2002) Concurrent design forces in structures under three-component orthotropic seismic excitation. *Earthq Spectra* 18:1–17
- ASCE (2007) Seismic rehabilitation of existing buildings. ASCE/SEI Standard 41-06. American Society of Civil Engineers, Reston
- ASCE/SEI (2010) Minimum design loads for buildings and other structures. ASCE/SEI Standard 07-10. American Society of Civil Engineers, Reston
- Athanatopoulou AM (2005) Critical orientation of three correlated seismic components. *Eng Struct* 27(2):301–312
- CEN (2005) Eurocode 8: design of structures for earthquake resistance. EN 1998. European Committee for Standardisation, Brussels
- Chopra AK, Goel RK (2002) A modal pushover analysis procedure for estimating seismic demands for buildings. *Earthq Eng Struct Dyn* 31(3):561–582
- Fajfar P (2000) A nonlinear analysis method for performance-based seismic design. *Earthq Spectra* 16(3):573–592
- FEMA (2003) HAZUS-MH MR1, Multi-hazard loss estimation methodology earthquake model. National Institute of Building Sciences, Washington, DC
- Fib (2012) Probabilistic performance-based seismic design. Bulletin 68. International Federation of Structural Concrete, Lausanne
- Fragiadakis M, Papadrakakis M (2008) Modelling, analysis and reliability of seismically excited structures: computational issues. *Int J Comput Methods* 5(4):483–511
- Gasparini DA, Vanmarke EH (1976) Simulated earthquake motions compatible with prescribed response spectra. Massachusetts Institute of Technology (MIT), Department of Civil Engineering, Publication No. R76-4, Cambridge, MA
- Ghera A, Rossi PP (2001) Influence of bi-directional ground motions on the inelastic response of one-storey in-plan irregular systems. *Eng Struct* 23(6):579–591
- Ghobarah A, Abou-Elfath H, Biddah A (1999) Response-based damage assessment of structures. *Earthq Eng Struct Dyn* 28(1):79–104
- Giovenale P, Cornell CA, Esteva L (2004) Comparing the adequacy of alternative ground motion intensity measures for the estimation of structural responses. *Earthq Eng Struct Dyn* 33(8):951–979
- Houbolt JC (1950) A recurrence matrix solution for the dynamic response of elastic aircraft. *J Aeronaut Sci* 17:540–550

- Kent DC, Park R (1971) Flexural members with confined concrete. *J Struct Div* 97(7):1969–1990
- Kramer SL (1996) Geotechnical earthquake engineering. Prentice Hall, Upper Saddle River
- Lagaros ND (2010) Multicomponent incremental dynamic analysis considering variable incident angle. *J Struct Infrastruct Eng* 6(1–2):77–94
- Lopez OA, Torres R (1997) The critical angle of seismic incidence and the maximum structural response. *Earthq Eng Struct Dyn* 26:881–894
- Lopez OA, Chopra AK, Hernandez JJ (2000) Critical response of structures to multicomponent earthquake excitation. *Earthq Eng Struct Dyn* 29:1759–1778
- MacRae GA, Mattheis J (2000) Three-dimensional steel building response to near-fault motions. *J Struct Eng* 126(1):117–126
- MacRae GA, Tagawa H (2001) Seismic behaviour of 3D steel moment frame with biaxial columns. *J Struct Eng* 127(5):490–497
- Menegotto M, Pinto PE (1973) Method of analysis for cyclically loaded reinforced concrete plane frames including changes in geometry and non-elastic behaviour of elements under combined normal force and bending. In: *Proceedings, IABSE symposium on resistance and ultimate deformability of structures acted on by well defined repeated loads*, Zurich, Switzerland, pp 15–22
- Menun C, Der Kiureghian A (1998) A replacement for the 30 %, 40 % and SRSS rules for multicomponent seismic analysis. *Earthq Spectra* 14(1):153–163
- Menun C, Der Kiureghian A (2000) Envelopes for seismic response vectors. I: theory. *J Struct Eng* 126:467–473
- Mitropoulou Ch Ch, Lagaros ND, Papadrakakis M (2010) Economic building design based on energy dissipation: a critical assessment. *Bull Earthq Eng* 8(6):1375–1396
- Newmark NM (1959) A method of computation for structural dynamics. *ASCE J Eng Mech Div* 85:67–94
- Pacific Earthquake Engineering Research (PEER) (2012) <http://peer.berkeley.edu/smcat/search.html/>. Accessed Nov 2012
- Penzien J, Watabe M (1975) Characteristics of 3-dimensional earthquake ground motions. *Earthq Eng Struct Dyn* 3(4):365–373
- Rigato AB, Medina RA (2007) Influence of angle of incidence on seismic demands for inelastic single-storey structures subjected to bi-directional ground motions. *Eng Struct* 29(10):2593–2601
- Vamvatsikos D, Cornell CA (2002) Incremental dynamic analysis. *Earthq Eng Struct Dyn* 31(3):491–514
- Wilson EL, Farhoomand I, Bathe KJ (1973) Nonlinear dynamic analysis of complex structures. *Earthq Eng Struct Dyn* 1:241–252
- Wilson EL, Suharwardy A, Habibullah A (1995) A clarification of the orthogonal effects in a three-dimensional seismic analysis. *Earthq Spectra* 11(4):659–666

Tracking Changes in Volcanic Systems with Seismic Interferometry

Matthew M. Haney¹, Alicia J. Hotovec-Ellis², Ninfa L. Bennington³, Silvio De Angelis⁴ and Clifford Thurber³

¹U.S. Geological Survey, Alaska Volcano Observatory, Anchorage, AK, USA

²Department of Earth and Space Sciences, University of Washington, Seattle, WA, USA

³Department of Geoscience, University of Wisconsin-Madison, Madison, WI, USA

⁴Earth, Ocean and Ecological Sciences, School of Environmental Sciences, University of Liverpool, Liverpool, UK

Synonyms

Ambient noise; Coda waves; Seasonal subsurface changes; Volcanic conduits; Volcano monitoring; Volcano seismology

Introduction

The detection and evaluation of time-dependent changes at volcanoes form the foundation upon which successful volcano monitoring is built. Temporal changes at volcanoes occur over all time scales and may be obvious (e.g., earthquake swarms) or subtle (e.g., a slow, steady increase in the level of tremor). Some of the most challenging types of time-dependent change to detect are subtle variations in material properties beneath active volcanoes. Although difficult to measure, such changes carry important information about stresses and fluids present within hydrothermal and magmatic systems. These changes are imprinted on seismic waves that propagate through volcanoes.

In recent years, there has been a quantum leap in the ability to detect subtle structural changes systematically at volcanoes with seismic waves. The new methodology is based on the idea that useful seismic signals can be generated “at will”

from seismic noise. This means signals can be measured any time, in contrast to the often irregular and unpredictable times of earthquakes. With seismic noise in the frequency band 0.1–1 Hz arising from the interaction of the ocean with the solid Earth known as microseisms, researchers have demonstrated that cross-correlations of passive seismic recordings between pairs of seismometers yield coherent signals (Campillo and Paul 2003; Shapiro and Campillo 2004). Based on this principle, coherent signals have been reconstructed from noise recordings in such diverse fields as helioseismology (Rickett and Claerbout 2000), ultrasound (Weaver and Lobkis 2001), ocean acoustic waves (Roux and Kuperman 2004), regional (Shapiro et al. 2005; Sabra et al. 2005; Bensen et al. 2007) and exploration (Draganov et al. 2007) seismology, atmospheric infrasound (Haney 2009), and studies of the cryosphere (Marsan et al. 2012).

Initial applications of ambient seismic noise were to regional surface wave tomography (Shapiro et al. 2005). Brenguier et al. (2007) were the first to use ambient noise tomography (ANT) to map the 3D structure of a volcanic interior (at Piton de la Fournaise). Subsequent studies have imaged volcanoes with ANT at Okmok (Masterlark et al. 2010), Toba (Stankiewicz et al. 2010), Katmai (Thurber et al. 2012), Asama (Nagaoka et al. 2012), Uturuncu (Jay et al. 2012), and Kilauea (Ballmer et al. 2013b). In addition, Ma et al. (2013) have imaged a scatterer in the volcanic region of southern Peru by applying array techniques to ambient noise correlations. Prior to and in tandem with the development of ANT, researchers discovered that repeating earthquakes, which often occur at volcanoes, could be used to monitor subtle time-dependent changes with a technique known as the doublet method or coda wave interferometry (CWI) (Poupinet et al. 1984; Roberts et al. 1992; Ratdomopurbo and Poupinet 1995; Snieder et al. 2002; Pandolfi et al. 2006; Wegler et al. 2006; Martini et al. 2009; Haney et al. 2009; De Angelis 2009; Nagaoka et al. 2010; Battaglia et al. 2012; Erdem and Waite 2005;

Hotovec-Ellis et al. 2014). Chaput et al. (2012) have also used scattered waves from Strombolian eruption coda at Erebus volcano to image the reflectivity of the volcanic interior with body wave interferometry. However, CWI in its original form was limited in that repeating earthquakes, or doublets, were not always guaranteed to occur. With the widespread use of noise correlations in seismology following the groundbreaking work by Campillo and Paul (2003) and Shapiro et al. (2005), it became evident that the nature of the ambient seismic field, due to its oceanic origin, enabled the continuous monitoring of subtle, time-dependent changes at both fault zones (Wegler and Sens-Schönfelder 2007; Brenguier et al. 2008b; Wegler et al. 2009; Sawazaki et al. 2009; Tatagi et al. 2012) and volcanoes (Sens-Schönfelder and Wegler 2006; Brenguier et al. 2008a) without the need for repeating earthquakes.

Seismic precursors to eruptions based on ambient noise were first detected at Piton de la Fournaise volcano on the island of Reunion (Brenguier et al. 2008a; Duputel et al. 2009). The studies at Piton de la Fournaise demonstrated the possibility of resolving small ($\sim 0.1\%$) decreases in seismic velocity in the weeks leading up to eruptions. Brenguier et al. (2008a) and Duputel et al. (2009) further showed how subtle spatial and temporal changes at the volcano could be mapped and used as a real-time tool for volcano monitoring and eruption forecasting. Traditionally, the forecasting ability of volcano seismology has rested on the assumption that volcanic unrest is preceded in advance by a significant increase in seismicity. However, some eruptions, such as Okmok in 2008 (Larsen et al. 2009), have begun with little or no precursory seismicity. For those eruptions, providing accurate and timely advance warnings is much more problematic, placing the public at risk of being exposed to the harmful effects of volcanic activity. Methods based on ambient noise have the potential to assist with forecasting of volcanic unrest in such cases, as well as for eruptions accompanied by ample seismicity and deformation.

Principles of Coda Wave Interferometry

CWI can in principle detect several types of temporal variations, among them changes in subsurface velocity, changes in the source location, changes in bulk scattering properties, and changes in the focal mechanism of earthquakes (Snieder 2006). The sensitivity to subsurface velocity changes initially led to the widespread adoption of CWI for applications at volcanoes. The sensitivity to subsurface velocity can be shown in simple terms with a model of a homogeneous half-space of velocity v with randomly distributed, small-scale scatterers. For this model, the travel time t for a particular scattering path is given simply as

$$t = \frac{d}{v} \quad (1)$$

where d is the total distance traveled for that path. Note that this distance is not necessarily along a straight-line path. To analyze variations in travel time, assume that the velocity changes by an amount Δv but the locations of the scatterers do not change. Since the locations do not change, the distance d traveled by each path stays the same. However, the change in velocity Δv causes there to be a resulting change in the travel time Δt , causing the relation in Eq. 1 to become

$$t + \Delta t = \frac{d}{v + \Delta v} \quad (2)$$

Assuming the changes are small and taking a Taylor series approximation of Eq. 2 yields

$$t + \Delta t = \frac{d}{v} \left(1 - \frac{\Delta v}{v} \right) \quad (3)$$

Combining Eqs. 1 and 3 finally gives the well-known result (Snieder 2006):

$$\frac{\Delta t}{t} = -\frac{\Delta v}{v} \quad (4)$$

Equation 4 is widely used in CWI, and it states that the fractional travel time change is equal to

the negative of the fractional velocity change. Thus, by measuring the fractional travel time change, the change in the subsurface velocity can be estimated. Three different methods have been proposed to measure the time-shifts in CWI: time-windowed cross-correlations (Snieder et al. 2002), the stretching method (Wegler and Sens-Schönfelder 2007), and the phase of the cross-spectrum (Poupinet et al. 1984). Note that Eq. 4 applies to direct waves as well as scattered or coda waves. However, for a given velocity change, the time-shift is greater for waves with longer travel times, i.e., scattered or coda waves. Thus, CWI, in contrast to many other seismic techniques, benefits from higher amounts of scattering since later-arriving waves have larger and more clearly identified time-shifts.

An alternative to the above derivation is for a changing 1D resonator of length d and with an internal propagation velocity v . Resonators have been discussed extensively in volcano seismology as models for sources acting in a crack or cavity (Chouet, 1996; Fee et al., 2010). This model is different from the model of randomly distributed scatterers but still produces a sequence of late-arriving waves. In this case, the derivation proceeds along the same lines as shown above, except that the length d is also allowed to vary, yielding the following expression for the travel time change:

$$\frac{\Delta t}{t} = \frac{\Delta d}{d} - \frac{\Delta v}{v} \quad (5)$$

Haney et al. (2009) have interpreted CWI delay times measured from repeating explosions at Pavlof volcano in the context of this resonator model. As a result, the observed travel time change could be interpreted as a change in the length of the resonator, a change in the velocity of the material inside the resonator, or a suitable combination of both types of changes (see discussion on page 173 of Garces and McNutt (1997)). The implication is that changing properties of the resonator, or conduit, control the changes observed in the coda of the repeating explosions at Pavlof. Landro and Stammeijer (2004) similarly make use of Eq. 5 for

interpreting time-lapse changes from a subsurface layer in an exploration seismic setting.

Overview of Coda Wave Interferometry at Volcanoes

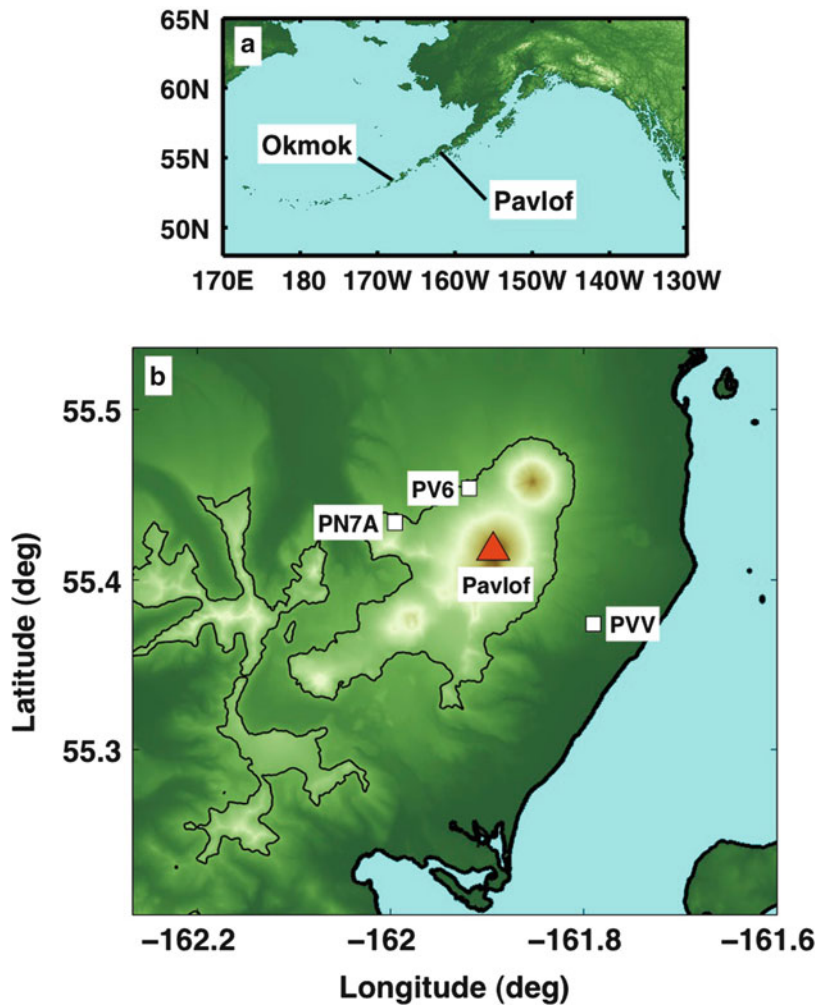
Coda waves of repeating earthquakes, or doublets, have been used to observe temporal changes in the subsurface for several decades (Poupinet et al. 1984). As described by Snieder et al. (2002), CWI takes advantage of subtle time-shifts in the coda of repeating seismic events. In CWI, the time lag between events as a function of recording time is determined via time-windowed cross-correlation of the traces. From the observed time lags, a corresponding change in velocity can be determined. The technique can detect small changes at volcanoes (e.g., Haney et al. 2009; Nagaoka et al. 2010; Hotovec-Ellis et al. 2014), with temporal resolution on the same order as the rate of the repeating events (e.g., hourly resolution in Hotovec-Ellis et al. (2013)). However, the requirement of repeating events precludes the use of CWI at many volcanoes where repeating events are infrequent or short lived. In contrast, ambient noise occurs continuously and, if the signal is highly repeatable, offers the ability to observe temporal changes at volcanoes at will (Sens-Schönfelder and Wegler 2006).

Since the field of interferometry is in its early stages, only a handful of studies (Sens-Schönfelder and Wegler 2006; Brenguier et al. 2008a; Duputel et al. 2009; Baptie 2010; Mordret et al. 2010; Anggono et al. 2012; Obermann et al. 2014) have employed ambient seismic noise to study changes at volcanoes. The existing studies have typically found velocity decreases at volcanoes due to or preceding activity: -0.5% by Baptie (2010), -2.3 to 3.3% by Anggono et al. (2012), -0.8% by Mordret et al. (2010), and -0.4% by Brenguier et al. (2008a). The changes detected with ambient noise to date include seasonal variations (Sens-Schönfelder and Wegler 2006), eruption precursors due to magma pressurization (Brenguier et al. 2008a; Duputel et al. 2009; Anggono et al. 2012), posteruption changes in the volcanic edifice due to dome collapse (Baptie 2010), and

changes in the hydrothermal system (Mordret et al. 2010).

The determination of subtle changes in velocity using ambient noise interferometry is carried out in a similar way to the process used in CWI with repeating earthquakes. Following Duputel et al. (2009), a long-time-period seismic correlation between station pairs is computed, and this correlation represents the reference correlation function (CF). The reference CF must be determined over a period of quiescence at the volcano (Duputel et al. 2009). The reference CF in Duputel et al. (2009) was generated during a two-month period when Piton de la Fournaise was relatively quiet. To identify temporal variations in velocity, they determined the current CF as the seismic correlation between a station pair over a smaller period of time than the reference CF by averaging over a time period of 10 days. Duputel et al. (2009) and Hadziioannou et al. (2009) demonstrated that the stretching technique is a stable method of determining the relative velocity change between station pairs showing significant time lags. The stretching method is an alternative to the traditional method of time-windowed correlations (Snieder et al. 2002). In the stretching method, the reference CF is stretched or compressed to best match the current CF. This stretched/compressed CF is calculated using an assumed relative change in velocity. Over a set of possible relative velocity changes, the relative velocity change that yields the best correlation between the stretched/compressed CF and the current CF is selected. Note that the stretching method assumes a uniform velocity change in the subsurface when measuring time delays between signals. This need not be the case, as demonstrated by Pacheco and Snieder (2006) in their study of time delays from localized velocity perturbations. There remains debate over the relative performance of the stretching method and time-windowed correlations (Zhan et al. 2013). A third method, known as the cross-spectrum method (Poupinet et al. 1984), is closely related to the time-windowed correlation method, since the two represent equivalent processes implemented in either the time or frequency domain.

Tracking Changes in Volcanic Systems with Seismic Interferometry, Fig. 1 (a) Regional map of the Aleutian Arc showing the two volcanoes discussed in this entry, Pavlof and Okmok. (b) Local map of Pavlof volcano and the seismic stations in the monitoring network analyzed in this entry



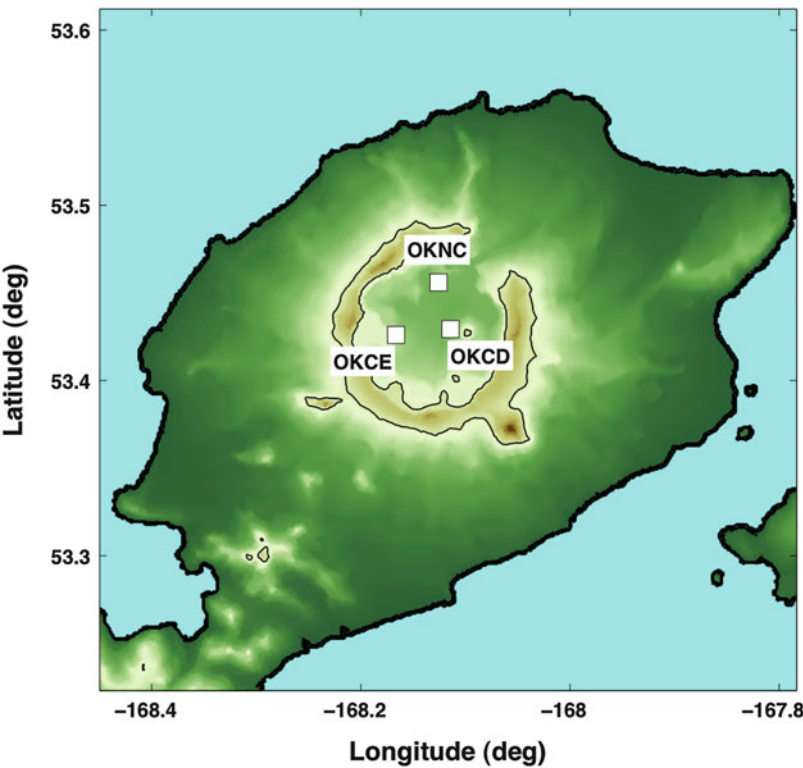
In the following sections, case studies of CWI are given that use ambient noise and repeating earthquakes at two volcanoes in the Aleutian Islands of Alaska: Okmok and Pavlof. The locations of these volcanoes within the Aleutian Arc are given in Fig. 1a. The seismic stations at Pavlof and Okmok discussed in the following sections are shown in Figs. 1b and 2, respectively. These two case studies illustrate the relative advantages and disadvantages of using ambient noise or repeating earthquakes to measure changes at volcanoes. Following these two case studies, new developments in coda wave interferometry are discussed that will influence future studies at volcanoes.

Seasonal Changes from Ambient Noise at Okmok Volcano, Alaska

Okmok volcano is one of the most active volcanoes in the Aleutian Arc, with an average of one eruption every decade over the past 100 years. Its broad shield structure is interrupted by a roughly 10-km-wide caldera, the remnant of two historical caldera-forming eruptions in the past 10,000 years (Larsen et al. 2009). The most recent eruption of Okmok in 2008 occurred with almost no warning, the only precursory activity being a short-lived earthquake swarm during the 5 h prior to the eruption (Larsen et al. 2009).

Tracking Changes in Volcanic Systems with Seismic Interferometry,

Fig. 2 Local map of Okmok volcano and the seismic stations in the monitoring network analyzed in this entry. The three stations all sit within the 10-km-wide caldera and are broadband installations

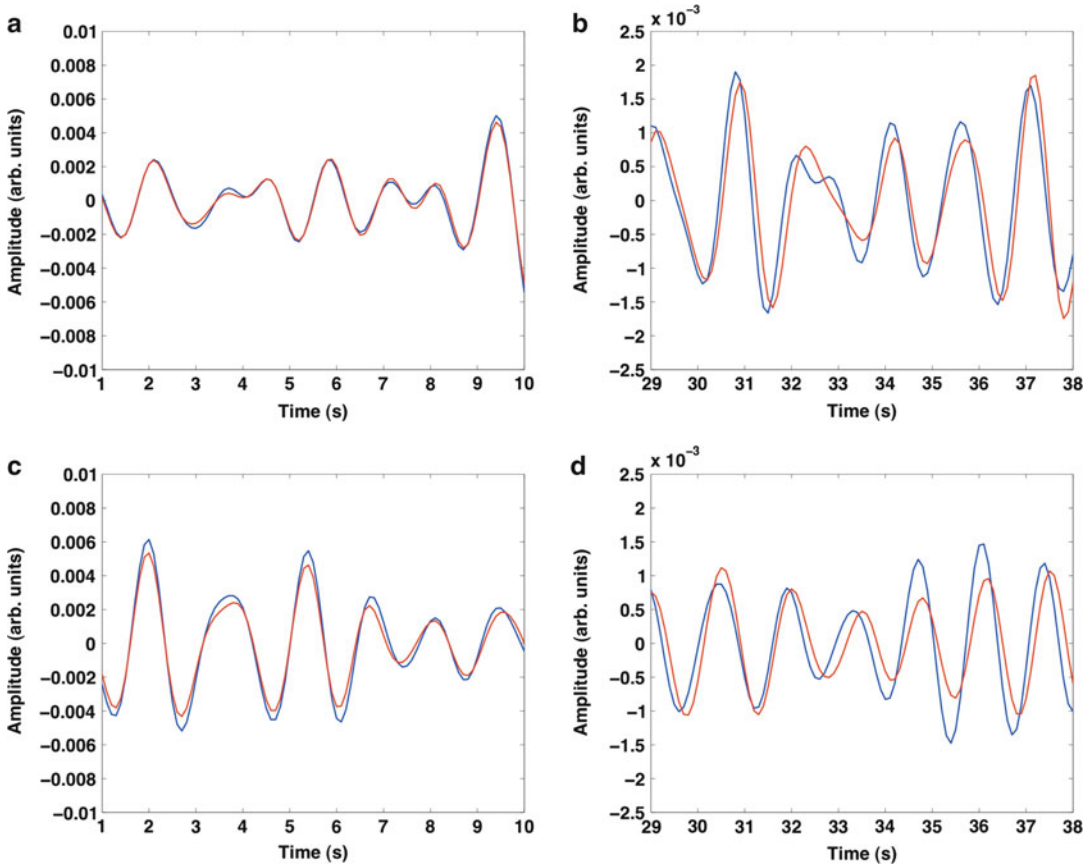


At Okmok, as in most of the Aleutian Islands, the majority of the seismic stations are short-period installations. However, several high-quality broadband seismic stations have existed at different times within the Okmok network. These stations represent some of the most remote broadbands in the network operated by the Alaska Volcano Observatory (AVO), a partnership between the University of Alaska Fairbanks Geophysical Institute, the Alaska Division of Geological and Geophysical Surveys, and the US Geological Survey. Three of the five broadbands at Okmok have been sited within the large caldera. The locations of these stations are shown in Fig. 2. The stations have operated over different time periods: OKNC (2010–present), OKCE (2003–present), and OKCD (2003–2008). Station OKCD was destroyed during the initial explosive phase of the 2008 eruption, which emanated from a nearby intracaldera cone. Here, ambient noise correlations are analyzed for station pair OKCE-OKNC during late 2012 and station pair OKCE-

Tracking Changes in Volcanic Systems with Seismic Interferometry, Table 1 Parameters for ambient noise-based coda wave interferometry at Okmok

Parameter	Value
Length of time window considered	5–60 s
Frequency passband	0.5–1 Hz
Moving window length	20 s
Min correlation to accept Δt measure	0.9
Temporal averaging window	+/- 4 days

OKCD during 2006. During both of these times, Okmok volcano was in a period of quiescence. In Table 1, several parameters related to the practical implementation of CWI at Okmok are given. The values of the parameters depend on the desired time scale of resolution, quality control of the correlations, and desired frequency band. The use of the frequency band from 0.5 to 1 Hz represents a trade-off between having a high level of ocean-related noise and being sufficiently high enough in frequency to ensure the generation of scattered waves in the subsurface over the length



Tracking Changes in Volcanic Systems with Seismic Interferometry, Fig. 3

Reference and current correlation functions (CFs) between station pair OKCE-OKNC in late 2012. The reference CF is obtained by averaging over the final 135 days of 2012. The current CF is centered on September 9, 2012, with the average including ± 4 days around this center time. Panels (a) and (c) depict early lag times in the reference (blue) and current

(red) CFs for both the positive (a) and negative (c) lag. At early times, between 1 and 10 s lag, the CFs virtually overlap. Panels (b) and (d) depict late lag times of the same CFs for positive (b) and negative (d) lag. At late lag times, between 29 and 38 s, the current CF (red) shows a subtle time delay relative to the reference correlation function (blue) for both positive and negative lags

scale of a volcano. Note that CWI is subject to statistical considerations related to random scattering paths, and therefore the moving window measurement in CWI must be sufficiently wide to average over scattering from subsurface heterogeneities (Snieder 2006). At Okmok, the coda has been examined using the time-windowed correlation method between time lags of 5–60 s using a moving window of 20 s length. The sensitivity of late-arriving coda waves to changes is shown in Fig. 3 between stations OKCE and OKNC. The late-arriving coda waves (Figures 3b and 3d) are more sensitive compared to the early-arriving

direct and forward-scattered waves (Figures 3a and 3c). Furthermore, as discussed below, the changes in Fig. 3 are consistently observed for both positive and negative lags, which constitutes a redundancy check for CWI based on ambient noise (Breguier et al. 2008a).

Additional CWI parameters concern the similarity between the reference and current CFs and temporal averaging. For the application at Okmok, the peak of the time-windowed correlation is required to exceed 0.9 in order for the associated time-lag measurement within the moving time window to be considered when

fitting the linear relationship in Eq. 4. Wegler et al. (2006) employed a similar criterion based on a minimum correlation coefficient in a CWI study at Merapi volcano, Indonesia. In addition to those parameters, the use of ambient noise requires that the nonstationarity of the ocean noise source be taken into account. Brenguier et al. (2008a) found that temporal averaging of the cross-correlations over a time interval on the order of 1 week renders ocean noise a sufficiently repeatable source. In the following application at Okmok, the current correlation functions (CFs) have been averaged over 8 days centered on the current time (Table 1). The reference CFs are the result of averaging over the entire time period under consideration, either 365 days for the 2006 data or 135 days for the 2012 data.

As mentioned briefly above, CWI based on ambient noise is inherently redundant, since changes should be independently observed for both the positive and negative lags for a correlation between a single station pair (Brenguier et al. 2008a). However, if the distance between a pair of stations is relatively small compared to the wavelength or if the changes can be assumed to be uniform in space, then two additional redundancies exist based on the autocorrelations of the two stations. This is because the autocorrelations can be viewed as ambient noise Green's functions for the case of a coincident source and receiver. For such a configuration, coda waves are still measured. This suggests that changes can be observed in four independent ways for stations that are in relatively close proximity. CWI best practices should exploit these redundancies to apply quality control to the observed changes.

For CWI to be meaningful, error estimates on the relative travel time change given in Eq. 4 must be provided. Following Brenguier et al. (2008a), the standard deviation of the time lag (i.e., the root-mean-square uncertainty of the linear fit) is given by

$$\sigma_{\Delta t} = \sqrt{\frac{\sum_{i=1}^N \left(\Delta t_i - \frac{\Delta v}{v} t_i \right)^2}{N}} \quad (6)$$

where N is the number of time-lag measurements. From Eq. 6, the standard deviation of the relative travel time change follows from Brenguier et al. (2008a) as

$$\sigma_{\Delta t/t} = \sqrt{\frac{\sum_{i=1}^N \left(\Delta t_i - \frac{\Delta v}{v} t_i \right)^2}{N \sum_{i=1}^N t_i^2}} \quad (7)$$

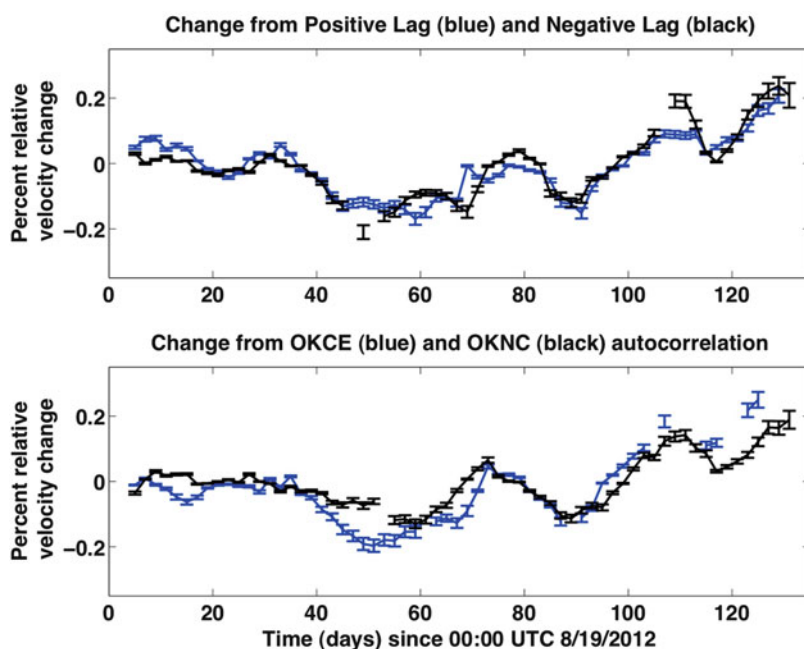
Equation 7 provides the formal error bars on estimates of the relative travel time and, through Eq. 4, velocity changes. However, in addition to formal error bars, goodness-of-fit criteria must be taken into account as well (Press et al. 1986). In the implementation of CWI at Okmok, two goodness-of-fit criteria are required to be met in order to accept an estimate of relative travel time, irrespective of the error estimate in Eq. 7. It must be that either (a) the standard deviation in Eq. 6 is less than one time sample or (b) the Pearson linear correlation coefficient given by

$$r_{xy} = \frac{N \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{N \sum x_i^2 - \left(\sum x_i \right)^2} \sqrt{N \sum y_i^2 - \left(\sum y_i \right)^2}} \quad (8)$$

is greater than 0.8, where, for CWI, the (x, y) variables in Eq. 8 are the travel time and the time lag $(t, \Delta t)$. Values of r_{xy} exceeding 0.8 are taken to indicate a strong linear relation. In practice, the first criterion means that measurements between similar waveforms in which the changes are not appreciable are automatically accepted. The second and possibly more important criterion requires that Δt and t have a strong linear correlation when the standard deviation is greater than one time sample, which in practice occurs when the changes are significant. Note that a linear relation between Δt and t may not resemble the measured Δt - t curve in general (Pacheco and Snieder 2006). However, if a linear model is being used, then it is at least consistent to require that the linear fit is good. Without considering the goodness-of-fit criterion in Eq. 8, it may be the

Tracking Changes in Volcanic Systems with Seismic Interferometry,

Fig. 4 Plotted in the *top* panel are the relative velocity changes computed from correlations of stations OKCE and OKNC in 2012 for positive lags (blue) and negative lags (black). The *bottom* panel shows the relative velocity changes for the OKCE (blue) and OKNC (black) autocorrelations. All four measurements are broadly in agreement, which they should be for velocity changes that are spatially extensive. This redundant property of ambient noise correlations can be used to measure the confidence of the velocity changes

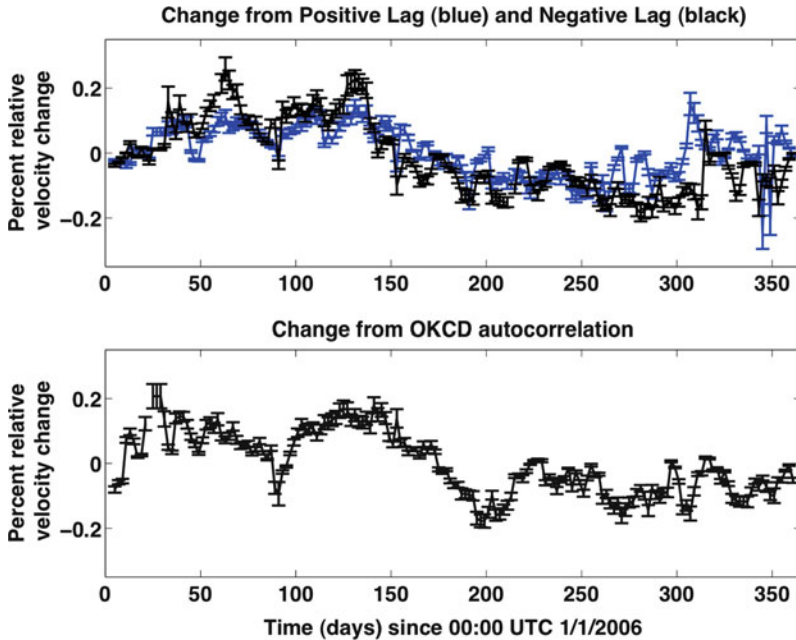


case that the formal error in Eq. 7 is acceptable even though the linear fit is not good. This emphasizes the need for a goodness-of-fit criterion, as discussed in Press et al. (1986).

Taking into account the above considerations, Fig. 4 shows the relative velocity changes computed from both positive and negative lag portions of cross-correlations and autocorrelations for Okmok stations OKCE and OKNC over the final 135 days of 2012. The formal uncertainties for the measurements are also provided in Fig. 4. Note that the various time series are discontinuous at certain points due to the CWI measurements failing to satisfy the goodness-of-fit criteria. The relative velocity changes in Fig. 4 are observed to vary between $\pm 0.2\%$. Moreover, the changes are internally consistent in that they are observed independently for both positive and negative lags. Consistent changes are also observed for the autocorrelations, indicating that the spatial distribution of the change in the subsurface is generally uniform inside of the caldera. The source of the change will be discussed after examining data from 2006, but in principle it could be the result of magmatic or seasonal variations in the subsurface. Seasonal changes

observed with CWI measurements have been studied previously by Sens-Schönfelder and Wegler (2006), Meier et al. (2010), Tsai (2011), and Hotovec-Ellis et al. (2014). Sources of the seasonal changes include thermoelastic and hydrologic effects and, at high elevations or high latitudes, the annual snow cycle.

In Fig. 5, relative velocity changes are shown between stations OKCE and OKCD for all of 2006. Positive and negative lag correlations are plotted, in addition to the autocorrelation of OKCD (the autocorrelation of OKCE is not plotted in Fig. 5 since in 2006 it suffered from a type of uncorrelated electronic noise). In spite of this noise, the three correlations plotted in Fig. 5 show consistent estimates of relative velocity change during 2006. The subsurface velocity is observed to be slightly higher in winter and spring (days 1–180 and 350–365) and lower in summer and fall (days 180–350). These annual variations are highly similar to the relative velocity changes observed by Hotovec-Ellis et al. (2014) at high-elevation sites on Mount St. Helens. Although the observations at Okmok are based on ambient noise, the observations by Hotovec-Ellis et al. (2014) were derived from long-term



Tracking Changes in Volcanic Systems with Seismic Interferometry, Fig. 5 Plotted in the *top* panel are the relative velocity changes computed from correlations of stations OKCE and OKCD in 2006 for positive lags (*blue*) and negative lags (*black*). The *bottom* panel shows the relative velocity change for the OKCD autocorrelation. Station OKCE suffered from electronic noise that was not present on OKCD, and thus its autocorrelation is not

shown. Since the noise only appeared on OKCE, it did not affect the OKCE-OKCD cross-correlation. All three measurements are broadly in agreement, with subtle velocity increases in the winter/spring and velocity decreases in summer/fall. Station OKCE suffered from electronic noise that was not present on OKCD, and thus its autocorrelation is not shown

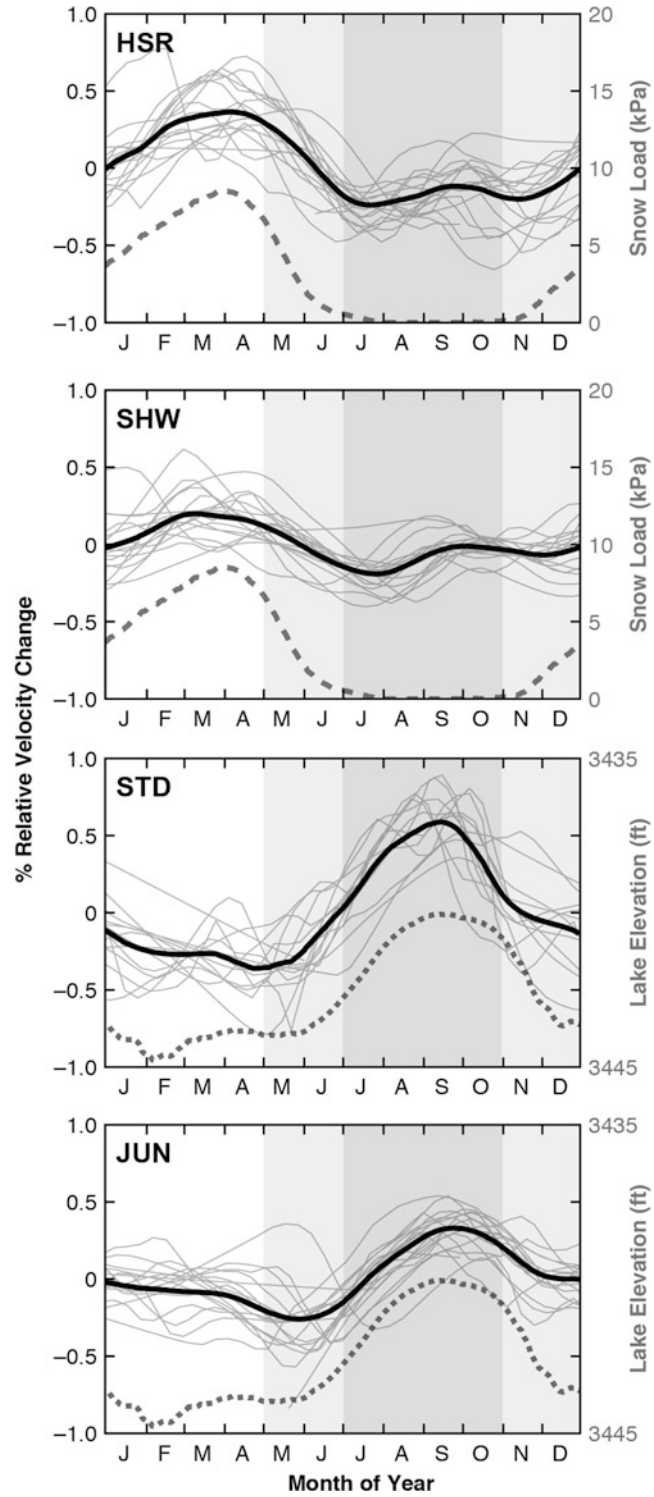
(decadal) records of repeating local earthquakes. Hotovec-Ellis et al. (2014) concluded that the annual variability at high-elevation sites on Mount St. Helens was controlled by the snow cycle and at lower-elevation sites by shallow fluid saturation (Fig. 6). The snow loading interpretation could apply at Okmok as well in spite of the low elevation of the caldera, since Okmok is located within the high-latitude chain of the Aleutian Islands. Returning to Fig. 4, the overall increase in velocity during the final 135 days of 2012 can be attributed to the onset of the snow-pack as fall transitioned into winter. Taken together, the relative velocity variations observed at Okmok during times of quiescence in 2006 and 2012 appear to be seasonal in nature and, due to the annual snow cycle, in agreement with the conclusions by Hotovec-Ellis et al. (2014) at high-elevation sites on Mount St. Helens.

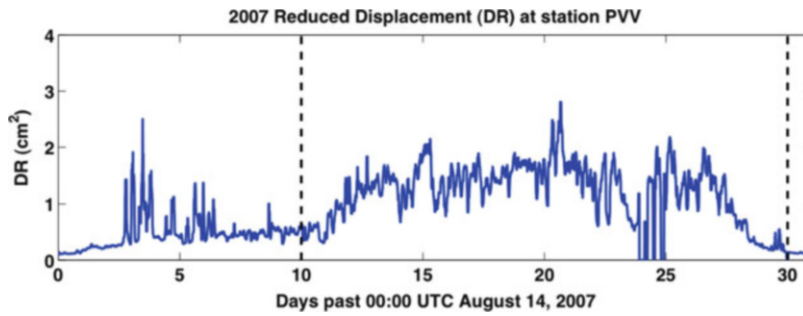
Conduit Changes from Repeating Explosions at Pavlof Volcano, Alaska

The 2007 eruption of Pavlof volcano was Strombolian in character and generated lava flows, lahars, and small explosions (Waythomas et al. 2007). Seismicity consisted of volcanic tremor and repeating long-period signals associated with the small explosions at the summit of the volcano. Haney et al. (2009) have previously applied CWI to the repeating explosions during the 2007 Pavlof eruption and detected subtle changes. In contrast to CWI based on ambient noise, as discussed in the previous section, CWI based on explosions at Pavlof takes advantage of the highly repetitive waveforms to observe small changes in the later portion of their signals. In this section, we summarize the study of Haney et al. (2009) and give some additional supporting observations.

Tracking Changes in Volcanic Systems with Seismic Interferometry,

Fig. 6 Velocity as a function of months of the year at Mt. St. Helens, with demeaned raw solutions in *light gray* and average in *thicker black*. *Dashed line* corresponds to average yearly snow load at an SNOTEL station in Sheep Canyon. *Dotted line* corresponds to average lake elevation at Spirit Lake, plotted with an inverted y-axis to better illustrate the anticorrelation and interpreted to correspond to changes in shallow fluid saturation. *Shaded area* denotes months of the year with increased shallow seismicity (Reproduced from Hotovec-Ellis et al. (2014))





Tracking Changes in Volcanic Systems with Seismic Interferometry, Fig. 7 Reduced displacement (D_R) of volcanic tremor over the course of the month-long eruption of Pavlof volcano in 2007. The plot of D_R shows that the eruption began with low-level tremor during the first 10 days, with pulses of higher D_R occurring during times of lahars. After 10 days, the tremor level increased and remained elevated for over 2 weeks before tapering off at

the end of the eruption in mid-September. The vertical dashed lines indicate the 20 days when repeating explosions occurred. The overall tremor D_R is indicative of a low-level Strombolian eruption of VEI between 1 and 2. Note data dropouts for this station between days 24 and 25. The vertical dashed lines indicate the 20 days presented in Fig. 8

To put the entire Pavlof eruption in context, Figure 7 shows the reduced displacement (D_R) computed at station PVV on the east side of the volcano (see Fig. 1b). D_R is a measure related to the energy radiated by volcanic tremor (Aki and Koyanagi 1981). As indicated in Fig. 7, the 2007 Pavlof eruption began on August 14, and, over the next 10 days, the tremor level was relatively low. Short periods of elevated D_R during the first 10 days of the eruption were not the result of elevated tremor but instead corresponded with lahars flowing close to PVV. On August 24, 10 days into the eruption, the tremor increased from a D_R of 0.5 cm² to a D_R of 1.5 cm² and stayed at that level until the end of the eruption. It was during this period of relatively higher tremor amplitude that repetitive explosions occurred at the summit of Pavlof, at the rate of approximately one explosion every 3–6 min. These Strombolian explosions represented a subtle increase in the overall intensity and explosivity of the eruption.

As described in Haney et al. (2009), the waveforms due to the repeating explosions over the entire eruption were identified using the master event matched filter technique described by Petersen (2007). Once the catalog of repeating events had been identified, Haney et al. (2009) stacked the repeating waveforms within successive 12-h time periods to increase the signal-to-

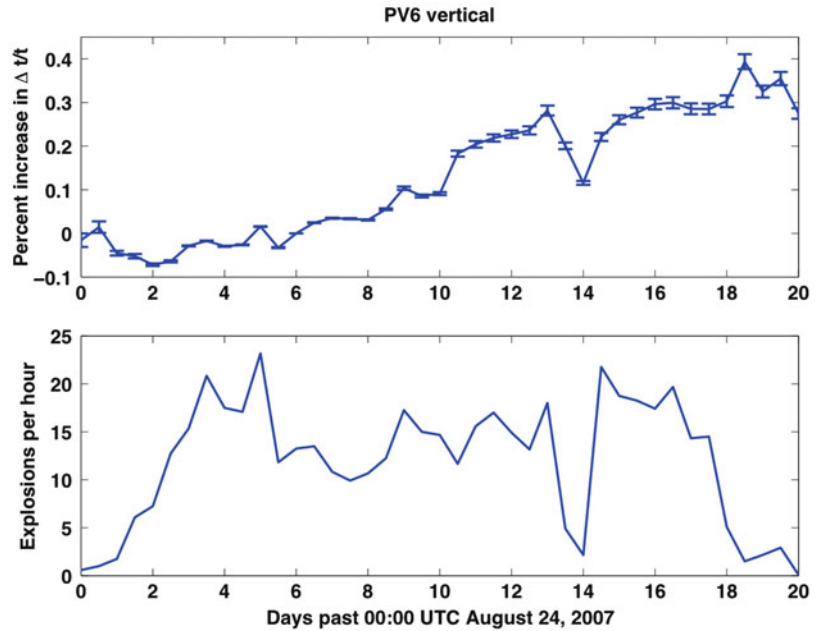
Tracking Changes in Volcanic Systems with Seismic Interferometry, Table 2 Parameters for explosion-based coda wave interferometry at Pavlof

Parameter	Value
Length of time window considered	0 20 s
Frequency passband	1–4 Hz
Moving window length	6 s
Min correlation to accept Δt measure	0.7
Temporal averaging window	+/- 6 h

noise ratio and applied CWI using the stack from the earliest time period as the reference. Table 2 gives additional parameters used for the application of CWI with the repeating explosions. Plotted in Fig. 8 is the relative travel time change over the final 20 days of the eruption, along with the average rate of explosions during each 12-h time period. An overall increase in the relative travel time of 0.4 % is observed as the eruption progressively came to an end over its final 20 days. The relative travel time change is plotted in Fig. 8 instead of relative velocity change, since Haney et al. (2009) interpreted the change in terms of the resonator model of Eq. 5. According to this model, the increase in travel time can be interpreted as an increase in the length of the resonator or conduit; a decrease in the internal propagation velocity of the conduit; or a suitable combination of both types of change.

Tracking Changes in Volcanic Systems with Seismic Interferometry,

Fig. 8 The *upper* panel shows the relative travel time change for stacks of repeating explosions over the course of 20 days beginning on August 24, 2007. The number of Strombolian explosions per hour is plotted in the *lower* panel for comparison. The explosions occurred during a time period of elevated reduced displacement, as seen in Fig. 7



Without additional information, the source of the change at Pavlof cannot be resolved further among these possible models. Although the relative traveltimes changes are only shown for station PV6 in Fig. 8, Haney et al. (2009) also observed similar changes at station PN7A located to the west of the summit (Fig. 1b).

Based on the frequency content of the explosions and common spectral peaks between stations PV6 and PN7A, Haney et al. (2009) concluded that the relative travel time changes observed in Fig. 8 were the result of changes within the volcanic conduit at Pavlof. The interpretation by Haney et al. (2009) is significant since the changes are therefore the result of a varying source effect instead of a varying path effect. This is in contrast to the conclusions of most CWI studies at volcanoes. A notable exception is the study by Erdem and Waite (2013) in which relative velocity changes were detected with CWI at Fuego volcano with repeating explosions over short time scales, in as little as hours. Erdem and Waite (2013) concluded that the variations were occurring within the conduit due to the differences in the apparent relative velocity changes measured on seismometers at different distances from the volcano.

To illustrate the interpretation by Erdem and Waite (2013) based on seismometers at different distances, the resonator model shown in Eq. 5 needs to be reconsidered. For this, consider the travel time of the n -th reverberation within the conduit, given by

$$t_n = t_E + \frac{nd}{v} \quad (9)$$

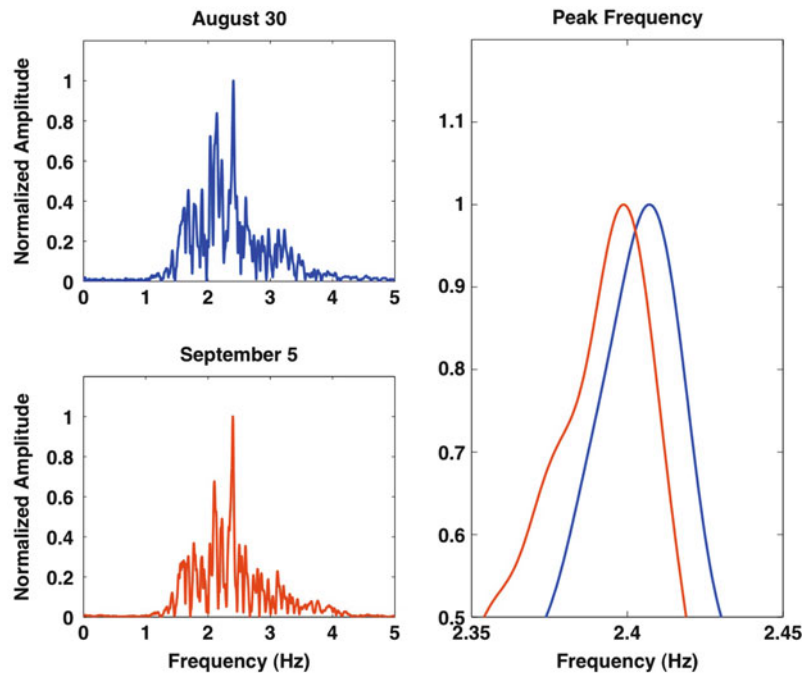
in which t_E is the traveltime needed to propagate through the Earth from the conduit to the seismometer. Assuming small changes in the conduit length d , velocity within the conduit v , and travel time of the n -th reverberation t_n , but no changes in the traveltime through the Earth t_E , a first-order perturbation of Eq. 9 yields

$$\Delta t_n = (t_n - t_E) \left[\frac{\Delta d}{d} - \frac{\Delta v}{v} \right] \quad (10)$$

Equation 5 is a special case of Eq. 10 when $t_E = 0$ and t_n is identified as the travel time. In addition, Eq. 5 can be viewed as an approximation that applies at times much later than the time needed to propagate through the Earth from the conduit to the seismometer ($t_n \gg t_E$). When these

Tracking Changes in Volcanic Systems with Seismic Interferometry,

Fig. 9 In *left upper* and *left lower* panels are amplitude spectra of stacks of repeating explosions on August 30 (*blue*) and September 5 (*red*) for station PV6. The spectra are dominated by a peak frequency at ~ 2.4 Hz. In the *right* panel is a zoom-in of the peaks on August 30 and September 5 plotted together. The peak is observed to have shifted from 2.407 to 2.399 Hz over the course of the 6 days, in agreement with the CWI measurements



conditions are not met, the use of Eq. 5 underpredicts the actual velocity change, as noted by Haney et al. (2009). Erdem and Waite (2013) astutely noted that the amount of underprediction varies for seismometers at different distances from the volcano and used that fact to diagnose the source of time-lapse changes at Fuego volcano.

An additional piece of evidence for conduit changes at Pavlof, not presented in Haney et al. (2009), follows from the spectral properties of the repeating explosions. For a simple 1D resonator, the change in the resonance frequency is related to changes in the propagation velocity inside the resonator and the length of the resonator as follows (e.g., Garces and McNutt 1997):

$$\frac{\Delta f}{f} = \frac{\Delta v}{v} - \frac{\Delta d}{d} \quad (11)$$

Note that the right-hand side of this equation is the negative of the right-hand side of Eq. 5. Thus, the relative changes in resonance frequency and traveltime for the model of a resonator have the same absolute value but are opposite in sign.

Whereas the traveltime, or residence time, of a 1D resonator increases when its internal velocity is decreased, its resonance frequency decreases. In Fig. 9, amplitude spectra of the repeating explosions are shown at station PV6 for time periods one week apart. The dominant resonant peak in both amplitude spectra is approximately 2.4 Hz with a slight shift toward lower frequencies later in the eruption. Garces and McNutt (1997) reported on a larger -18% change in the spectral peaks of tremor before and after an eruption at Mount Spurr volcano in 1992. The change at Pavlof, observed in the spectral domain, is consistent with the change derived from CWI in the time domain. This consistency in the spectral domain supports the interpretation by Haney et al. (2009) of conduit resonance and hints at deeper connections between the methods of interferometry and spectroscopy in physics (Zadler et al. 2005). In fact, the CWI experiment described in Snieder et al. (2002), involving a sample of Elberton granite, could have been equivalently analyzed in the frequency domain since the waves traversed the finite-sized sample many times over the time window of the experiment.

New Developments in Coda Wave Interferometry

Two emerging areas of research in the use of seismic interferometry for monitoring changes at volcanoes are highlighted in this section. The first example addresses the use of new sources besides repeating earthquakes and the oceanic microseism. The second example concerns the use of decorrelation between repeating signals in addition to traditional time-lag measurements.

The simple relationship in Eq. 1 holds for the direct wave portion of the ambient noise Green's function in a homogeneous medium with d the interstation distance. However, if the distribution of noise sources is not uniform, spurious arrivals can appear in ambient noise Green's functions. This issue is a problem for ANT, since tomography requires accurate interstation travel times. However, as will be shown here, this isn't an issue for CWI with ambient noise. Consider an extreme case in which the noise source comes from a single azimuth, as shown in Fig. 10. For this case, the interstation travel time for the direct wave is given by

$$t = \frac{d \cos \theta}{v} \quad (12)$$

where θ is the azimuthal angle between the interstation azimuth and the back azimuth of the plane wave source. Equation 12 shows that for this case the arrival is spurious since it doesn't correspond to a physical wave in the Green's function, which would arrive at $t = d/v$. Proceeding as in Eqs. 1, 2, 3, and 4, assume that the subsurface velocity changes by an amount Δv . The change in velocity Δv causes there to be a resulting change in the travel time Δt . In addition,

consider that the source azimuth changes by $\Delta\theta$, giving the following relation:

$$t + \Delta t = \frac{d \cos(\theta + \Delta\theta)}{v + \Delta v} \quad (13)$$

Assuming small changes and taking a first-order Taylor series approximation of Eq. 13 yields

$$t + \Delta t = \frac{d}{v} \left(\cos \theta - \Delta\theta \sin \theta - \frac{\Delta\theta^2}{2} \cos \theta \right) \left(1 - \frac{\Delta v}{v} \right) \quad (14)$$

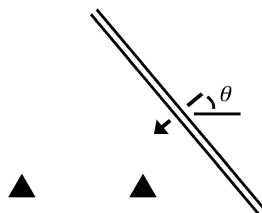
Combining Eqs. 12 and 14 yields

$$\frac{\Delta t}{t} = -\frac{\Delta v}{v} - \Delta\theta \tan \theta \left(1 - \frac{\Delta v}{v} \right) - \frac{\Delta\theta^2}{2} \quad (15)$$

An important outcome of this result is that, in the case of no change in the azimuth of the source ($\Delta\theta = 0$), Eq. 15 is the same as Eq. 4 in spite of the Green's function being incorrect due to the noise only coming from one direction. Equation 15 for a stable source ($\Delta\theta = 0$) shows the insensitivity of ambient-noise-based CWI to the direction of an oblique plane wave source. It is a simple demonstration that accurately detecting changes does not require the cross-correlations to be the true Green's function. Nonphysical arrivals in Green's functions constructed from ambient noise, so-called spurious arrivals (Snieder et al. 2008), still obey Eq. 4 when the source is stable. Hadzioannou et al. (2009) first took note of this property and concluded that time-lapse changes can be reliably estimated even when the Green's function is not properly reconstructed. Hadzioannou et al. (2009) further concluded that the only condition necessary for monitoring changes is the relative stability of the noise.

These results bring up the possibility of using unconventional, localized sources, that are nonetheless continuous and stable, to detect changes. An example of such a source is man-made cultural noise produced by machinery or traffic. Another example is stable volcanic tremor

Tracking Changes in Volcanic Systems with Seismic Interferometry,
Fig. 10 A plane wave source incident obliquely on a pair of seismometers



(Ballmer et al. 2013a), which at some volcanoes can persist for weeks, months, or even years. This also motivates the study of such noise sources to establish their stability prior to using them for detecting time-lapse changes. The wave fields from these localized noise sources do not even begin to satisfy the conditions necessary for imaging with a technique such as ANT however, the reduced requirement of relative source stability for monitoring permits a wider set of applicable noise sources.

A recent development by Obermann et al. (2014) addresses decorrelation instead of time-shifts in coda-wave measurements. In CWI, when two seismograms are cross-correlated over a small time window, the time delay is related to the time-lag of the peak. The value of the peak itself is the correlation coefficient C ; the decorrelation coefficient is simply defined as $D = 1 - C$. This approach offers a new type of measurement in addition to the widely used time delay in CWI. Obermann et al. (2014) applied a technique based on decorrelation at Piton de la Fournaise and found information about changes in the subsurface that was complementary to the information provided by conventional CWI based on time shifts. Specifically, whereas time-shift measurements detected velocity variations in the subsurface, decorrelation measurements detected changes in scattering. Newly created cracks in the subsurface related to magma intrusions were interpreted by Obermann et al. (2014) to be the source of the decorrelation signal. In the following, the expression for decorrelation due to a single scatterer presented in Obermann et al. (2014) is specialized for the case of randomly distributed scatterers in a homogeneous background medium. This expression forms the link between the decorrelation coefficient in Obermann et al. (2014) and similar expressions first derived in Snieder et al. (2002).

From Obermann et al. (2014), the decorrelation for a single scatterer with scattering cross section S in a background medium with wave speed v is expressed as

$$D(s_1, s_2, x_0, t) = \frac{vS}{2} K(s_1, s_2, x_0, t) \quad (16)$$

where s_1 and s_2 are the locations of the two stations, x_0 is the location of the single scatterer, and K is the CWI sensitivity kernel presented in Pacheco and Snieder (2005). The CWI sensitivity kernel expresses the spatial location of the sensitivity of a CWI measurement in a generally variable background medium. The derivation of Eq. 16 for the case of a single scatterer is quite complicated (Rossetto et al. 2011); however, as shown below, its generalization to a distribution of new scatterers leads to a simple and insightful expression (the terminology of “new scatterers” refers to scatterers that developed in the intervening time period between measurements taken at different times). Given N identical new scatterers in a small area dA , the effect of the individual new scatterers on the decorrelation can be assumed to be additive:

$$D(s_1, s_2, x_0, t) = \frac{NvS}{2} K(s_1, s_2, x_0, t) \quad (17)$$

Multiplying and dividing this expression by the small area dA leads to

$$D(s_1, s_2, x_0, t) = \frac{nvS}{2} K(s_1, s_2, x_0, t) dA \quad (18)$$

where $n = N/dA$ is the (2D) number density of the new scatterers. Integration of both sides of this equation over the whole spatial area considered gives the decorrelation due to a distribution of scatterers:

$$D(s_1, s_2, t) = \int \frac{nvS}{2} K(s_1, s_2, x_0, t) dA \quad (19)$$

which similarly assumes as before that the effects of the different areas are additive, although here the different areas may have variable scattering cross sections $S(x_0)$. The product of the number density and scattering cross-section is related to the inverse of the mean free path within the independent scattering approximation ($L = 1/nS$) (Obermann et al. 2014). Note that this mean free path is for the new scatterers not present in the background medium. Assuming the background velocity is constant finally gives

$$D(s_1, s_2, t) = \frac{v}{2} \int \frac{K(s_1, s_2, x_0, t)}{L} dA \quad (20)$$

where the mean free path L is taken to have a dependence on the spatial coordinate x_0 .

To further simplify the expression, a model is adopted in which the mean free path for the new scatterers is constant everywhere (globally). This is the opposite end member of a single scatterer, Eq. 16, among models. It is also similar to the global change in velocity that gives rise to Eq. 4. Given the following expression for travel time known as the Chapman-Kolmogorov equation (Pacheco and Snieder 2005)

$$t = \int K(s_1, s_2, x_0, t) dA \quad (21)$$

the decorrelation for a constant mean free path increases linearly with time according to

$$D(t) = \frac{v}{2L} \int K(s_1, s_2, x_0, t) dA = \frac{vt}{2L} \quad (22)$$

The linearly increasing decorrelation behavior with time is in fact the same type of behavior described by Snieder et al. (2002) for a model of moving scatterers, in which all scatterers are randomly perturbed by a distance Δ . This model is different from the model of Obermann et al. (2014) that consists of new scatterers appearing among otherwise stationary and unchanging background scatterers. From Snieder et al. (2002), the decorrelation for the moving scatterer model is given by

$$D(t) = k^2 \Delta^2 \frac{vt}{l} \quad (23)$$

where k is the wavenumber and l is the mean free path in the background medium. This result means that the two models – either the moving scatterers of Snieder et al. (2002) or the newly appearing scatterers of Obermann et al. (2014) – both predict a linear increase in decorrelation when the changes are global and are thus indistinguishable in that case. A possible way to discern these two models in the case of global

changes would be to carefully exploit the wavenumber dependence for the moving scatterer model in different frequency bands. It is not clear whether these models are indistinguishable for changes that are local, but the fact that they are indistinguishable for global changes does hint at some amount of intrinsic nonuniqueness in resolving the different models.

Equation 22, $D(t) = vt/2L$, can be rewritten as $D(d) = d/2L$, where d is the total distance traveled by the path. This relation clearly shows the mechanism of the decorrelation in the model of Obermann et al. (2014), which is scattering by the new scatterers. It is the fraction of the total distance divided by twice the scattering mean free path of the newly appearing scatterers. Note that this mean free path L is not necessarily close to the mean free path of the background medium. The advancements in the understanding of decorrelation reported in Obermann et al. (2014) should lead to increased use of this measurement in future studies of time-lapse changes at volcanoes.

Conclusion

As demonstrated by a flurry of research activity in recent years, tracking changes at volcanoes with seismic interferometry has the potential to become a powerful volcano monitoring tool. The ambient seismic noise arising from the interaction of the ocean with the solid Earth yields repeatable signals that can detect subtle time-dependent changes at volcanoes. In addition to the overview of previous studies discussed in this entry, two examples of detecting time-lapse changes were presented for Okmok and Pavlof volcanoes in Alaska. Interestingly, Newhall (2007) independently pointed out an approach similar to ambient noise CWI in a review of volcanology and volcano monitoring. Newhall (2007), when discussing the challenge of interpreting indirect geophysical measurements at volcanoes, suggested tracking changes at volcanoes in response to known, repeating natural signals. With the advent of ambient noise interferometry, the concept of using repeatable,

natural sources of energy from the ocean to probe volcanoes has been realized and offers a new opportunity to understand the fascinating and complex inner workings of magmatic systems worldwide.

Cross-References

- [Frequency-Magnitude Distribution of Seismicity in Volcanic Regions](#)
- [Noise-Based Seismic Imaging and Monitoring of Volcanoes](#)
- [Seismic Anisotropy in Volcanic Regions](#)
- [Seismic Monitoring of Volcanoes](#)
- [Seismic Noise](#)
- [Seismic Tomography of Volcanoes](#)
- [Very-Long-Period Seismicity at Active Volcanoes: Source Mechanisms](#)
- [Volcanic Eruptions, Real-Time Forecasting of](#)
- [Volcano-Tectonic Seismicity of Soufriere Hills Volcano, Montserrat](#)

References

- Aki K, Koyanagi RY (1981) Deep volcanic tremor and magma ascent mechanism under Kilauea, Hawaii. *J Geophys Res* 86:7095–7110
- Anggono T, Nishimura T, Sato H, Ueda H, Ukawa M (2012) Spatio-temporal changes in seismic velocity associated with the 2000 activity of Miyakejima volcano as inferred from cross-correlation analyses of ambient noise. *J Volcanol Geotherm Res* 247:93–107
- Ballmer S, Wolfe CJ, Okubo PG, Haney MM, Thurber CH (2013a) Ambient seismic noise interferometry in Hawaii reveals long-range observability of volcanic tremor. *Geophys J Int* doi:10.1093/gji/ggt112
- Ballmer S, Haney MM, Wolfe CJ, Okubo P, Thurber CH (2013b) Short-period Rayleigh wave tomography for Kilauea and Mauna Loa volcanoes, Hawaii, from ambient seismic noise, Abstract V34B-05 presented at 2013 Fall Meeting, AGU, San Francisco, pp 9–13
- Baptie BJ (2010) Lava dome collapse detected using passive seismic interferometry. *Geophys Res Lett* 37, L00E10. doi:10.1029/2010GL042489
- Battaglia J, Métaixian J-P, Garaebiti E (2012) Earthquake-volcano interaction imaged by coda-wave interferometry. *Geophys Res Lett* 39, L11309. doi:10.1029/2012GL052003
- Bensen GD, Ritzwoller MH, Barmin MP, Levshin AL, Lin F, Moschetti MP, Shapiro NM, Yang Y (2007) Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements. *Geophys J Int* 169:1239–1260
- Brenguier F, Shapiro NM, Campillo M, Nercessian A, Ferrazzini V (2007) 3D surface wave tomography of the Piton de la Fournaise volcano using seismic noise correlations. *Geophys Res Lett* 34(2), L02205
- Brenguier F, Shapiro NM, Campillo M, Ferrazzini V, Duputel Z, Coutant O, Nercessian A (2008a) Towards forecasting volcanic eruptions using seismic noise. *Nat Geosci* 1:126–130. doi:10.1038/ngeo104
- Brenguier F, Campillo M, Hadziioannou C, Shapiro NM, Nadeau RM, Larose E (2008b) Postseismic relaxation along the San Andreas fault at Parkfield from continuous seismological observations. *Science* 321:1478–1481. doi:10.1126/science.1160943
- Campillo M, Paul A (2003) Long range correlations in the diffuse seismic coda. *Science* 299:546–549
- Chaput J, Zandomenighi D, Aster R, Knox H, Kyle P (2012) Imaging of Erebus volcano using body wave seismic interferometry of Strombolian eruption coda. *Geophys Res Lett* 39, L07304. doi:10.1029/2012GL050956
- Chouet BA (1996) Long-period volcano seismicity: its source and use in eruption forecasting. *Nature* 380:309–316. doi:10.1038/380309a0
- De Angelis S (2009) Seismic source displacement by coda wave interferometry at Soufriere Hills Volcano, Montserrat, WI. *Nat Hazards Earth Syst Sci* 9:1341–1347
- Draganov D, Wapenaar K, Mulder W, Singer J, Verdel A (2007) Retrieval of reflections from seismic background-noise measurements. *Geophys Res Lett* 34, L04305. doi:10.1029/2006GL028735
- Duputel Z, Ferrazzini V, Brenguier F, Shapiro N, Campillo M, Nercessian A (2009) Real time monitoring of relative velocity changes using ambient seismic noise at the Piton de la Fournaise volcano (La Réunion) from January 2006 to June 2007. *J Volcanol Geotherm Res* 184:164–173. doi:10.1016/j.jvolgeores.2008.11.024
- Erdem JE, Waite GP (2005) Temporal changes in eruptive behavior identified with coda wave interferometry and seismo-acoustic observations at Fuego Volcano, Guatemala. *Geol Soc Am Spec Pap* 498:107–123
- Fee D, Garcés M, Patrick M, Chouet B, Dawson P, Swanson D (2010) Infrasonic harmonic tremor and degassing bursts from Halema'uma'u Crater, Kilauea Volcano, Hawaii. *J Geophys Res* 115, B11316. doi:10.1029/2010JB007642
- Garcés MA, McNutt SR (1997) Theory of the airborne sound field generated in a resonant magma conduit. *J Volcanol Geotherm Res* 78:155–178
- Hadziioannou C, Larose E, Coutant O, Roux P, Campillo M (2009) Stability of monitoring weak changes in multiply scattering media with ambient noise correlation: laboratory experiments. *J Acoust Soc Am* 125(6):3688–3695. doi:10.1121/1.3125345
- Haney MM (2009) Infrasonic ambient noise interferometry from correlations of microbaroms. *Geophys Res Lett* 36, L19808

- Haney MM, van Wijk K, Preston LA, Aldridge DF (2009) Observation and modeling of source effects in coda wave interferometry at Pavlof volcano. *Lead Edge* 28:554–560. doi:10.1190/1.3124930
- Hotovec-Ellis AJ, Vidale JE, Gombert JS, Moran SC, Thelen WA (2013) Changes in seismic velocity during the 2004–2008 eruption of Mount St. Helens Volcano, Abstract V43D-03 presented at 2013 Fall Meeting, AGU, San Francisco, pp 9–13
- Hotovec-Ellis AJ, Gombert JS, Vidale JE, Creager KC (2014) A continuous record of inter-eruption velocity change at Mount St. Helens from coda-wave interferometry. *J Geophys Res* 119:2199–2214
- Jay JA, Pritchard ME, West ME, Christensen D, Haney M, Minaya E, Sunagua M, McNutt SR, Zabala M (2012) Shallow seismicity, triggered seismicity, and ambient noise tomography at the long-dormant Uturuncu Volcano, Bolivia. *Bull Volcanol* 74(4): 817–837
- Landro M, Stammeijer J (2004) Quantitative estimation of compaction and velocity changes using 4D impedance and traveltimes changes. *Geophysics* 69:949–957
- Larsen J, Neal C, Webley P, Freymueller J, Haney M, McNutt S, Schneider D, Prejean S, Schaefer J, Wessels R (2009) Eruption of Alaska volcano breaks historic pattern. *Eos Trans AGU* 90:173–174
- Ma Y, Clayton RW, Tsai VC, Zhan Z (2013) Locating a scatterer in the active volcanic area of Southern Peru from ambient noise cross-correlation. *Geophys J Int* 192:1332–1341. doi:10.1093/gji/ggs103
- Marsan D, Weiss J, Larose E, Metaxian J-P (2012) Sea-ice thickness measurement based on the dispersion of ice swell. *J Acoust Soc Am* 131(1):80–91
- Martini F, Bean CJ, Saccorotti G, Viveiros F, Wallenstein N (2009) Seasonal cycles of seismic velocity variations detected using coda wave interferometry at Fogo volcano, Sao Miguel, Azores, during 2003–2004. *J Volcanol Geotherm Res* 181:231–246
- Masterlark T, Haney M, Dickinson H, Fournier T, Searcy C (2010) Rheologic and structural controls on the deformation of Okmok volcano, Alaska: FEMs, InSAR, and ambient noise tomography. *J Geophys Res* 115, B02409
- Meier U, Shapiro NM, Brenguier F (2010) Detecting seasonal variations in seismic velocities within Los Angeles basin from correlations of ambient seismic noise. *Geophys J Int* 181:985–996
- Mordret A, Jolly AD, Duputel Z, Fournier N (2010) Monitoring of phreatic eruptions using interferometry on retrieved cross-correlation function from ambient seismic noise: results from Mt. Ruapehu, New Zealand. *J Volcanol Geotherm Res* 191:46–59. doi:10.1016/j.jvolgeores.2010.01.010
- Nagaoka Y, Nishida K, Aoki Y, Takeo M (2010) Temporal change of phase velocity beneath Mt. Asama, Japan, inferred from coda wave interferometry. *Geophys Res Lett* 37, L22311. doi:10.1029/2010GL045289
- Nagaoka Y, Nishida K, Aoki Y, Takeo M, Ohminato T (2012) Seismic imaging of magma chamber beneath an active volcano. *Earth Planet Sci Lett* 333:1–8
- Newhall CG (2007) *Volcanology 101 for seismologists*. In: Kanamori H (ed) *Earthquake seismology*. Elsevier, Amsterdam
- Obermann A, Planès T, Larose E, Campillo M (2014) Imaging pre- and co-eruptive structural and mechanical changes of a volcano with ambient seismic noise. *J Geophys Res*. doi:10.1002/2013JB010399
- Pacheco C, Snieder R (2005) Time-lapse travel time change of multiply scattered acoustic waves. *J Acoust Soc Am* 118(3):1300–1310
- Pacheco C, Snieder R (2006) Time-lapse traveltime change of singly scattered acoustic waves. *Geophys J Int* 165:485–500
- Pandolfi D, Bean CJ, Saccorotti G (2006) Coda wave interferometric detection of seismic velocity changes associated with the 1999 M = 3.6 event at Mt. Vesuvius. *Geophys Res Lett* 33, L06306. doi:10.1029/2005GL025355
- Petersen T (2007) Swarms of repeating long-period earthquakes at Shishaldin Volcano, Alaska, 2001–2004. *J Volcanol Geotherm Res* 166:177–192. doi:10.1016/j.jvolgeores.2007.07.014
- Poupinet G, Ellsworth WL, Frechet J (1984) Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras Fault, California. *J Geophys Res* 89(B7):5719–5731
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1986) *Numerical recipes in FORTRAN: the art of scientific computing*. Cambridge University Press, Cambridge
- Ratdomopurbo A, Poupinet G (1995) Monitoring a temporal change of seismic velocity in a volcano: application to the 1992 eruption of Mt. Merapi (Indonesia). *Geophys Res Lett* 22(7):775–778
- Rickett JE, Claerbout JF (2000) Calculation of the Sun's impulse response by multi-dimensional factorization. *Sol Phys* 192:203–210
- Roberts PM, Phillips WS, Fehler MC (1992) Development of the active doublet method for measuring small velocity changes and attenuation in solids. *J Acoust Soc Am* 91:3291–3302
- Rossetto V, Margerin L, Planes T, Larose E (2011) Locating a weak change using diffusive waves: theoretical approach and inversion procedure. *J Appl Phys* 109:034903
- Roux P, Kuperman WA (2004) Extracting coherent wavefronts from acoustic ambient noise in the ocean. *J Acoust Soc Am* 116:1995–2003
- Sabra KG, Gerstoft P, Roux P, Kuperman WA, Fehler MC (2005) Extracting time-domain Green's function estimates from ambient seismic noise. *Geophys Res Lett* 32, L03310. doi:10.1029/2004GL021862
- Sawazaki K, Sato H, Nakahara H, Nishimura T (2009) Time-lapse changes of seismic velocity in the shallow ground caused by strong ground motion shock of the 2000 Western-Tottori earthquake, Japan, as revealed

- from coda deconvolution analysis. *Bull Seismol Soc Am* 99(1):352–366. doi:10.1785/0120080058
- Sens-Schönfelder C, Wegler U (2006) Passive image interferometry and seasonal variations of seismic velocities at Merapi Volcano, Indonesia. *Geophys Res Lett* 33, L21302. doi:10.1029/2006GL027797
- Shapiro NM, Campillo M (2004) Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophys Res Lett* 31, L07614. doi:10.1029/2004GL019491
- Shapiro NM, Campillo M, Stehly L, Ritzwoller M (2005) High-resolution surface wave tomography from ambient seismic noise. *Science* 29:1615–1617
- Snieder R (2006) The theory of coda wave interferometry. *Pure Appl Geophys* 163:455–473. doi:10.1007/s00024-005-0026-6
- Snieder R, Grêt A, Douma H, Scales J (2002) Coda wave interferometry for estimating nonlinear behavior in seismic velocity. *Science* 295:2253–2255
- Snieder R, van Wijk K, Haney M, Calvert R (2008) Cancellation of spurious arrivals in Green's function extraction and the generalized optical theorem. *Phys Rev E* 78:036606
- Stankiewicz J, Ryberg T, Haberland C, Fauzi C, Natawidjaja D (2010) Lake Toba volcano magma chamber imaged by ambient seismic noise tomography. *Geophys Res Lett* 37, L17306. doi:10.1029/2010GL044211
- Tatagi R, Okada T, Nakahara H, Umino N, Hasegawa A (2012) Coseismic velocity change in and around the focal region of the 2008 Iwate-Miyagi Nairiku earthquake. *J Geophys Res* 117, B06315. doi:10.1029/2012JB009252
- Thurber C, Murphy R, Prejean S, Haney M, Bennington N, Powell L, Paskievitch J (2012) Earthquake studies reveal the magmatic plumbing system of the Katmai volcanoes: in Winfree, Robert (project lead), Katmai science studies. *Alsk Park Sci J* 11(1):34–39
- Tsai VC (2011) A model for seasonal changes in GPS positions and seismic wave speeds due to thermoelastic and hydrologic variations. *J Geophys Res* 116, B04404. doi:10.1029/2010JB008156
- Waythomas C, McNutt S, Prejean S (2007) Alaska's Pavlof volcano ends 11-year repose. *Eos Trans AGU* 89(23):209–211
- Weaver RL, Lobkis OI (2001) Ultrasonics without a source: thermal fluctuation correlations at MHz frequencies. *Phys Rev Lett* 87:134301. doi:10.1103/PhysRevLett.87.134301
- Wegler U, Sens-Schönfelder C (2007) Fault zone monitoring with passive image interferometry. *Geophys J Int* 168:1029–1033
- Wegler U, Lühr B-G, Snieder R, Ratdomopurbo A (2006) Increase of shear wave velocity before the 1998 eruption of Merapi volcano (Indonesia). *Geophys Res Lett* 33, L09303. doi:10.1029/2006GL025928
- Wegler U, Nakahara H, Sens-Schönfelder C, Korn M, Shiomi K (2009) Sudden drop of seismic velocity after the 2004 Mw 6.6 mid-Niigata earthquake, Japan, observed with passive image interferometry. *J Geophys Res* 114, B06305. doi:10.1029/2008JB005869
- Zadler BJ, Grêt A, Scales JA (2005) Spectroscopy versus interferometry: resolving small changes. *Am J Phys* 73(9):837–844. doi:10.1119/1.1891172
- Zhan Z, Tsai VC, Clayton RW (2013) Spurious velocity changes caused by temporal variations in ambient noise frequency content. *Geophys J Int* 194:1574–1581. doi:10.1093/gji/ggt170

Tsunamis as Paleoseismic Indicators

Klaus Reicherter

Institute of Neotectonics and Natural Hazards,
Department of Geosciences and Geography,
RWTH Aachen University, Aachen, Germany

Synonyms

High-energy deposits; Tsunami deposits;
Tsunamigenic layers; Tsunamiites; Tsunamites

Introduction

Tsunami waves are the result of a large mass of water being displaced by different processes. Earthquake-induced tsunamis have small wave heights offshore and very long wavelengths up to several hundred kilometers (line source tsunami). This is in contrast to tsunamis caused by mass movements (subaerial or underwater origin), volcanic eruptions and explosions, glacier calvings, or meteorite impacts, which have very high waves and very short wavelengths close to the origin (point source tsunami).

More than 85 % of tsunamis have tectonic causes such as fault movements (5 % are caused by volcanic activity, 5 % by landslides, and 5 % by a combination of all other causes; impact tsunamis are very rare), and 80 % of these occur in the Pacific Ocean; other oceans and seas have also been affected by tsunamis like the Mediterranean Sea. Tsunami evidence has also been

found on the shorelines of large and deep lakes, such as Lake Geneva in Switzerland.

Extreme waves can also be induced by storms and become tsunamis when approaching the coastline due to shoaling or shallowing. This can result in the superposition of waves and a dramatic reduction in wave velocity; sometimes also infragravity waves with high velocity have been observed during cyclones, e.g., the Haiyanstorm on the Philippines in 2014.

Tsunami action is mirrored in geological records and archives as phases of erosion, reworking, and redeposition leaving unconformities and unusual sedimentary layers. Tsunamites or tsunami deposits are proof that past or prehistoric tsunami events have occurred along coastlines or lakeshores. This article aims to describe the main environments, deposits of earthquake-derived tsunamis, and the limitations on their use.

Tsunami Processes

To understand how tsunami deposits may be preserved in the rock record, the process of tsunami generation, propagation, and landfall must be first understood.

The speed of a tsunami wave is calculated by the square root of the product of acceleration by gravity (9.8 m/s^2) and the water depth; in deep water, the wavelength of the tsunami must also be taken into account. The plural form is “tsunamis” or “tsunami.”

A tsunami train is a series of waves, which are destructive when they make landfall (Fig. 1). Refraction of tsunami waves (bending) may occur and mainly depends on the morphology and shape of the seabed in front of the coast. Wave reflection also occurs which describes the wave bouncing back after striking the shoreline. The first wave in the tsunami wave train is often not the largest wave and therefore leaves almost no deposits (or only in favorable setting) as the deposits are usually reworked by following waves. The time interval between waves hitting the coastline varies but can be around 10 min or more. Generally, three to four major wave

landfalls can be documented in the sedimentary record. The ecology and the geomorphology of coastlines are severely affected and modified during earthquake-induced tsunami landfall.

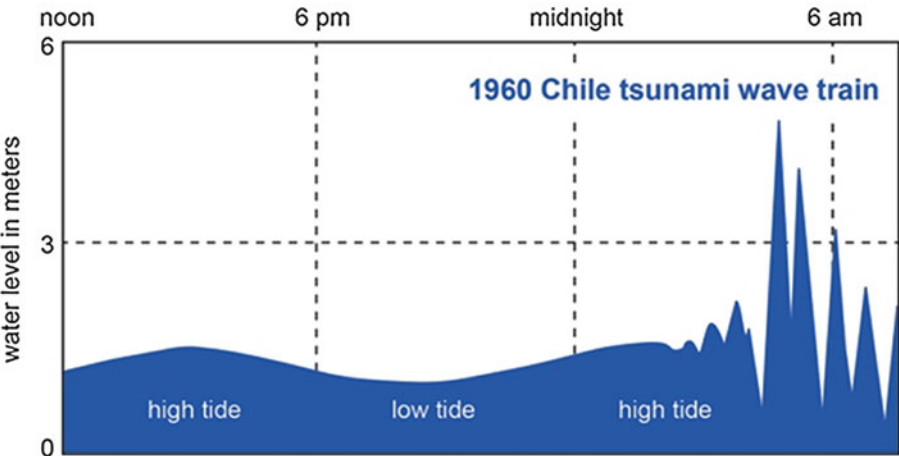
Several parameters obtained from recent or historical tsunamis are used in paleotsunami research, the most important of which are the inundation distance and run-up height (Fig. 2). The maximum run-up and inundation are delineated by the location of the wrack line or wrack line deposits. These are then used as basic parameters for coastal evacuation and emergency planning. Local effects of coastal morphology such as bays, fjords, and rias can amplify wave heights and inundations.

Coastal Environments for Paleotsunamis

Sedimentary evidence for tsunamis is found along the shorelines of marine and lacustrine environments. The shorelines of the Earth vary from place to place and can be simplified into two characteristic environments: flat sandy coasts or marshes and cliffed rocky coasts (Fig. 3a, b). The formation of a coast with sediments at the sea-shore depends on several parameters including the position of the sea level, the tidal influence, and the presence of river inlets (e.g., sculpted into bays, cusped forelands with mudflats, marshes, sandy beaches, or dunes). The geomorphology of a coast plays a major role in preserving tsunamigenic deposits.

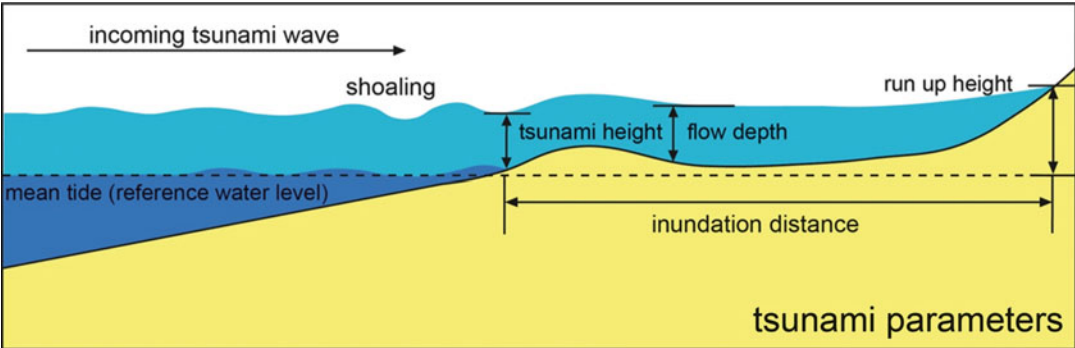
Soft Coasts

Paleotsunami research focuses on suitable archives along affected coastlines. To have a suitable archive, the deposits must firstly have had the potential to be preserved, which means long-term subsidence and sedimentation should prevail and not uplift/erosion; secondly, the deposits should be distinguishable from storm layers (tempestites); and thirdly, the deposits should yield datable material. Lagoons or marshy flatlands have been widely used as archives for investigating tsunami deposits (Fig. 3a). In these settings, tsunami landfall has a huge effect on the

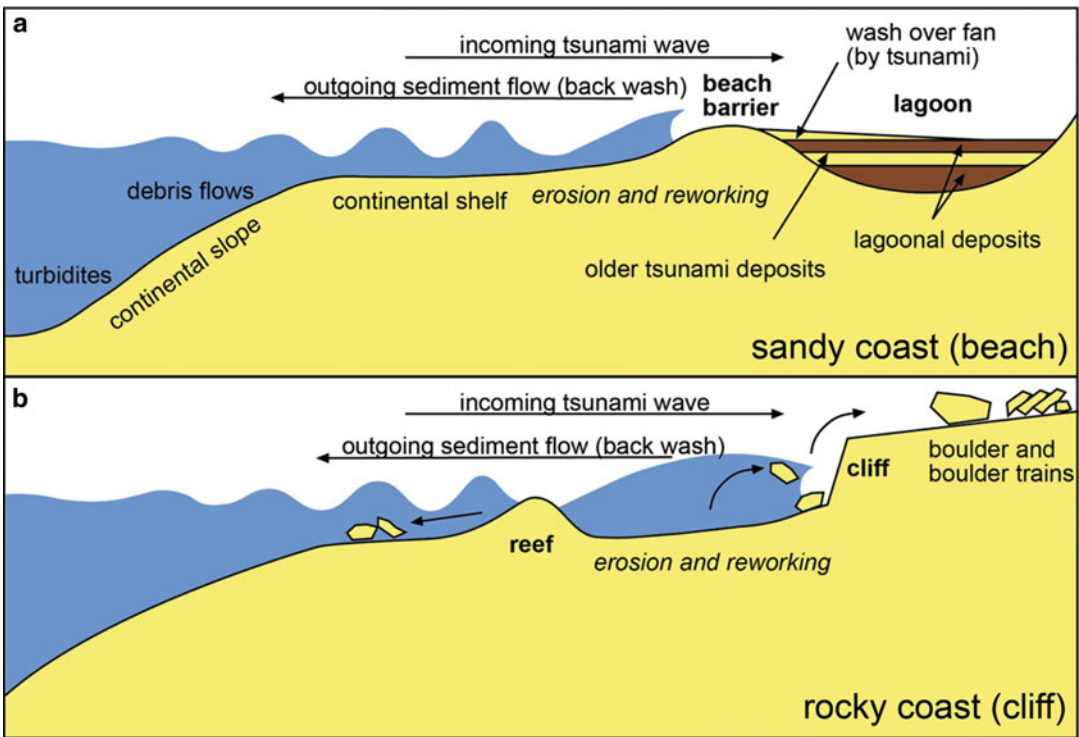


Tsunamis as Paleoseismic Indicators, Fig. 1 Tsunami wave train approaching the Japanese coastline after the 2011 Tohoku earthquake, note tsunami has already made landfall (*left*; Photo by Douglas Sprott is licensed under

CC BY-NC 2.0). Gauge data showing the wave train from the 1960 Chilean earthquake tsunami along the Japanese coast (Onagawa city) with tide data superimposed (*right*, Modified from Atwater et al. 1999)



Tsunamis as Paleoseismic Indicators, Fig. 2 General tsunami parameters



Tsunamis as Paleoseismic Indicators, Fig. 3 (a) Tsunami landfall and deposits along sandy shores, and various types of back wash deposits (*middle*). (b) Tsunami landfall at rocky coasts (*bottom*)

local sedimentary environment and ecology. Freshwater, brackish, or hypersaline lagoons will be affected by short durations of marine ingressions containing characteristic marine flora and fauna. Furthermore, coarse-grained deposits (sand sheets and washover fans) will be deposited in fine-grained lagoonal and organic clays or evaporitic sediments. After landfall, the lagoonal environment may then reestablish itself including sedimentation, and hence sand sheets can be preserved (Fig. 3a).

Hard Coasts

Along steep rocky shores, the locations of huge blocks and the arrangement of smaller blocks are used to delineate tsunami landfall. Tsunami boulders or “boulder trains” are chains of imbricated blocks interpreted as remnants of tsunami action (Fig. 3b). These “boulder trains” have been found all around the world (Scheffers and Kelletat 2003; Scheffers 2008).

Backwash Offshore

Offshore records of paleotsunamis are rare, and related deposits are ambiguous; therefore, they cannot be reliably associated with tsunami action. Backwash deposits into the sea/lake after tsunami landfall by the backflow of the water mass are also hard to identify. For relatively recent tsunamis, backwash deposits can be identified mainly by the presence of anthropogenic pollution as evidenced by some chemical proxies; however, older tsunamigenic deposits are complex and hard to recognize as they may not yield these characteristic constituents.

Due to Holocene sea level rise, many shorelines have developed relatively recently. Other shorelines have been affected by tectonic movements (uplift or subsidence). Due to the glacial sea level low stands, some tsunamis older than 10–15,000 years may not be preserved along the seashores as they are now drowned or have been reworked. However,

interglacial/-stadial tsunamigenic deposits (during a high stand, e.g., MIS 5) have the potential to be preserved.

Tsunami Sediments or Deposits (Tsunamiites, Tsunamiites)

Tsunami sedimentology has advanced considerably in the past 25 years since Brian Atwater in 1987 (Atwater 1987) associated sand layers on the Pacific Coast of Washington State with the occurrence of major tsunami-producing subduction earthquakes (Rhodes et al. 2006). The 2004 Indian Ocean and 2011 Tohoku-oki tsunamis and their deposits significantly changed our understanding of sedimentary features, sediment architecture, and the processes of tsunamigenic deposition, preservation, and postdepositional changes. Prior to 2004, the historical record and especially catastrophic events were too infrequent to be satisfactorily studied, and the identification of tsunami remains was equivocal. Paleotsunami sedimentology is complex, but more and more researchers have integrated sedimentological, geochemical, and paleontological analyses from post-tsunami field surveys and laboratory analyses. This multidisciplinary approach allows a datum to be assigned to the event layers using a combination of several dating methods. To some extent, a clear assignment of an event layer to tsunami action will always remain equivocal, but large storms have different wavelengths, have a major peak of wave action, and have no train of several major waves. Consequently, tsunami deposits should be different to those caused by storms. Some authors also refer to the recurrence periods of, e.g., tropical storms, which is generally an order of magnitude higher than that of major (paleo)tsunamis (e.g., Nanayama et al. 2003). The tsunami database has been expanded during the last decades and, in combination with an improved understanding of tsunami depositional mechanisms (Shiki et al. 2008; Goff et al. 2012), has helped to identify paleotsunami deposits in many parts of the world.

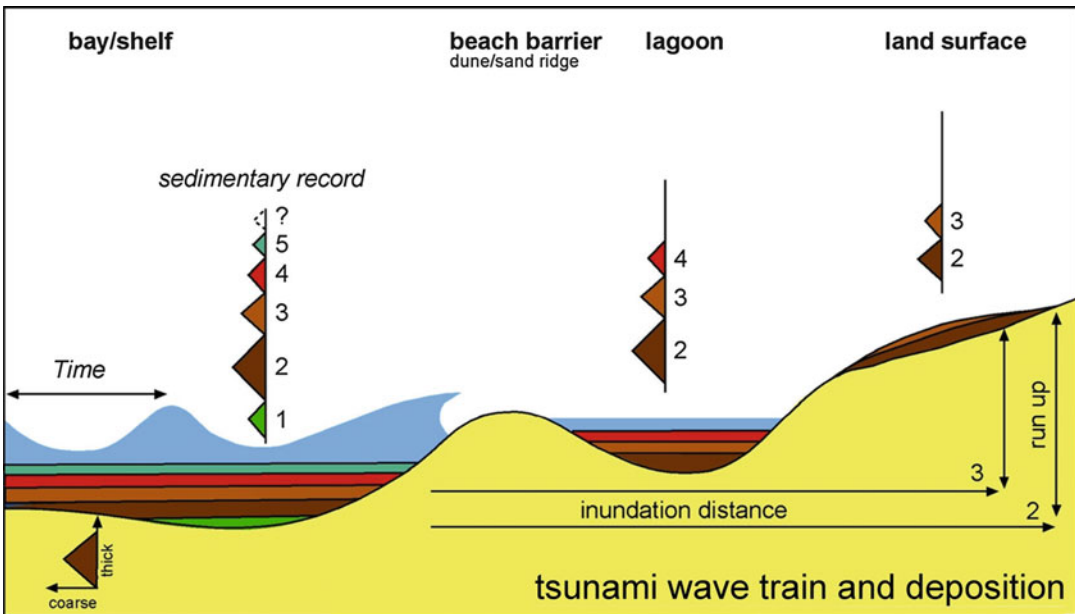
Identification

Recognizing and identifying paleotsunami deposits and assigning the magnitude and extent of (pre)historic tsunamis is an important part of paleoseismological studies. An outstanding example of this is from Japan, where tsunami deposits of a historical event in AD 869 were identified on the Sendai plain by Minoura et al. (2001). The 2011 Tohoku-oki tsunami deposits now cover the AD 869 deposits in the same area. The presence of the older deposits was used to determine/verify the occurrence of a previous large earthquake and tsunami event and proved that this area is susceptible to large-scale tsunami inundation. The world then realized that this part of Japan had experienced magnitude 9 earthquakes with associated tsunamis in the past. Sedimentary, geochemical (inorganic, organic, isotopic), and paleontological proxies are usually used for identification of tsunami deposits.

Despite the existence of a vast amount of publications on diagnostic features of paleotsunamis (e.g., Shiki et al. 2008), identification is not always easy in other parts of the world, and unambiguous evidence for tsunami deposit identification is not always present. In these cases, a range of features must be assessed in order to determine the origin of the deposits.

Soft Coasts

A tsunami is recorded sometimes only by a single sand layer of varying thickness. Sandy sheets of tsunami deposits in the sedimentary record are generally characterized by (a) a sheetlike geometry stretching in the inundation area; (b) a landward thinning of the sand layer(s) (wedging, tapering); (c) a landward fining of the sand sheet(s); (d) fining upward in the layer(s), normal grading (suspension sediment), partly reverse grading, and lamination (traction sediment; Fig. 5); (e) multiple layers of the sand sheet(s), showing all a fining and thinning upward (Fig. 4), representing the “wave train,” but sometimes multiple layers are not present or are overlain by mud in the landward limit (mud drape); (f) beach



Tsunamis as Paleoseismic Indicators, Fig. 4 Sedimentary record after tsunami landfall and variation of deposits along sandy shores, with respect to tsunami wave trains, run up and depositional environments. Different colors reflect deposits of differing ages and thickness. Note also the varying topographic height of deposits

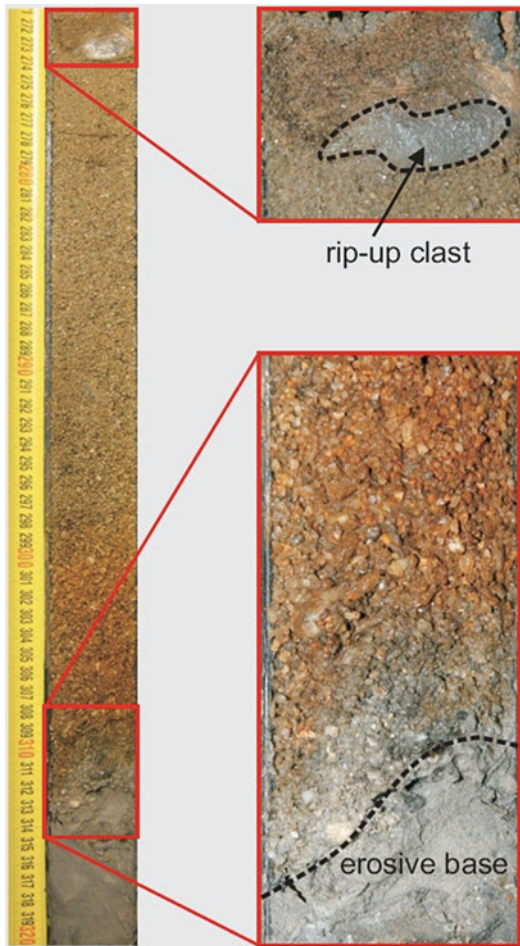
of the identical ages. Time reflects several minutes between the arrival of major waves. The record in the lagoon contains three deposits from three waves, with thinning-up of the layers and fining-up within the layers

sediments and biogenic material (shells, corals, aquatic plants) are incorporated in the sand sheets; (g) rip-up clasts of semilithified clayey or sandy materials and incorporated (Fig. 5), partly armored mud balls; and (h) the sediments show a marine geochemical signature, including microfossils (foraminifera, ostracods, diatoms, pollen, nannoliths, etc., but all of those may be reworked), but also this feature is not always present or preserved (Goff et al. 2012). Identification of a muddy tsunami deposit is still difficult at present; (inorganic and organic) geochemical, microfossil, and x-radiograph approaches are promising (Goff et al. 2012). The best identification is always given if extraordinary layers are intercalated in sediments of different grain size (i.e., sand sheets in fine-grained lagoonal deposits, see Figs. 4 and 5). Tsunami deposits can also consist of complex layers of mud, sand (dune sands), and boulders, containing abundant stratigraphic evidence for sediment reworking (rip-up clasts: eroded out from the original

surface or lagoon, Fig. 5) and redeposition, including anthropogenic materials (ceramics, concrete fragments artifacts, biogeochemical tracers, like oils, fats, metal-rich deposits, charcoal, and others).

Hard Coasts

Tsunami boulder deposits are identified around the world along many sandy and rocky coasts, and the findings are still increasing (Scheffers 2008; Shiki et al. 2008; and e.g., Pacific/Japan: Goto et al. 2007, 2009, 2010; Pacific/Indian Ocean: Etienne et al. 2011; Indian Ocean: Hoffmann et al. 2013a; Fig. 6). Some of the tsunami boulders have weights larger than 100 t. Boulder deposits provide useful information to improve our understanding of recurrence intervals and paleotsunami size, if dated correctly, by using boulder orientation and inland variation in boulder size, blocks, and boulders with pot holes (bioerosion, up-side down or rotated), paleonotches, adhesive growth of animals, like



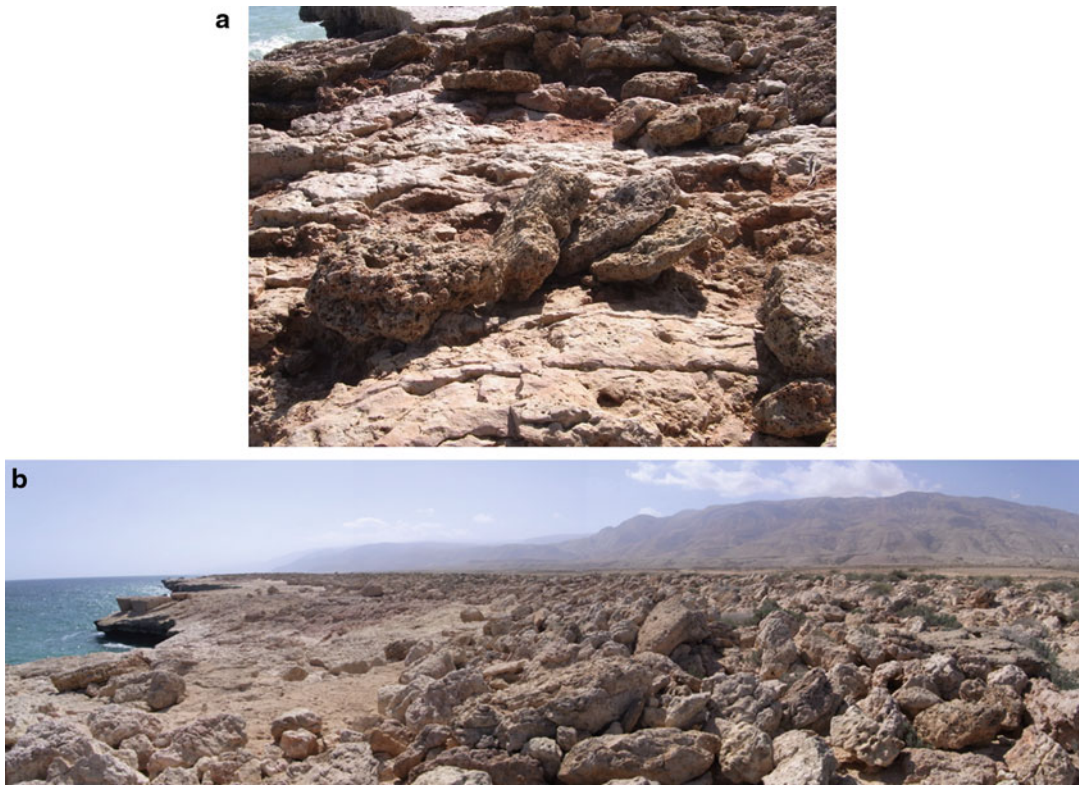
Tsunamis as Paleoseismic Indicators, Fig. 5 Close-up of a drill core from locality Sozopoli/northern Greece (From Reicherter et al. 2010), showing two cycles of the tsunami train. The lower layer has an erosive base at 3.15 m depth, and is fining-up. The second layer starts with coarse-grained sands and rip-up clasts at 2.74 m depth, and is generally finer-grained (thinning and fining up)

oysters, *Lithophaga* sp., bryozoans, corals, worms, and algae (providing carbonate for radiocarbon dating). The quarrying and subsequent transport of a block depends on several parameters such as weight (mass of block), shape, density (of water and block), porosity, water velocity, water depth, duration, roughness of platform, steepness of beach, direction, and rate of flow (Keating et al. 2011). Additionally, the height of uplifted blocks above present sea level, wave length, and the coefficients of drag and of

friction have to be taken into account and considered in calculations (Goto et al. 2009, Hoffmann et al. 2013a). Even for one of the simplest formulas, the “transport figure” (T_f) by Scheffers and Kelletat (2003) and Scheffers (2008), these parameters are needed to quantify the energy of displacement of a boulder by the equation $T_f = W \times D \times V$; where W is the weight of the boulder (in tons); D is the distance moved from the cliff line (m), and V is the vertical distance, i.e., cliff height (m). A value of 2,000 is considered the upper limit of storm wave transport energy (Scheffers 2008). The hydrodynamics of boulder and block transport during tsunami has also been discussed by Benner et al. (2010), who calculated wave velocities based on the conservation of energy; by Paris et al. (2010), who created a formula for the minimum wave velocity of a sliding boulder; and by Etienne et al. (2011), who chose an approach to calculate the minimum wave velocity necessary for overturning a boulder.

Mapping of Paleotsunami Deposits

Field surveys aim to map the three-dimensional distribution of a sand sheet and associated layers and to find evidence for a regional continuity of the layer. Indicative for washover fans are a tapering and fining landward, a common height above mean sea level, and the content of marine fossils. These observations can be easily achieved by shallow drillings (e.g., Wilson et al. 2014) and GPS control or ERT (electric resistivity) of the drill points for large strips along a coast. For boulder deposits, the orientation of the long axis of min. 50 boulders needs to be measured (where a is long axis, b is intermediate axis, and c is short axis) to obtain paleocurrent direction and boulder volume/weight information. Terrestrial laser scanning or LiDAR is applied to establish precisely the boulder volume (see Hoffmann et al. (2013a) for methodology). Post-tsunami field surveys are extremely useful that have been carried out for more than a decade now (Dominey-Howes et al. 2012). The aim of mapping is also focusing and aiming on return periods of events and magnitude of those.



Tsunamis as Paleoseismic Indicators, Fig. 6 (a) Imbricate boulders, boulder train (Fins, Sultanate of Oman) (b) Tsunami rampart and boulders in Fins, Sultanate of Oman (From Hoffmann et al. 2013a)

Preservation of Tsunami Deposits/ Postdeposition Changes

Postdepositional changes that tsunami deposits can undergo have to be considered. The processes that decrease preservation include physical and chemical weathering and bioturbation (Szczeniński 2012). These processes include (a) physical processes, like washing out the fine sediment at the surface, redeposition of sediment, leveling of the surface, disruption by frost heaving, and liquefaction. The (b) chemical processes include the removal of sea salt, the dissolution of carbonate materials, and the oxidation of metal-oids, whereas (c) biological processes cover bioturbation by roots and animals and human activity. Isotopic composition may be changed chemically by rain and freshwater, i.e., oxygen, carbon, nitrogen isotope ratios, Sr isotopes, Ca,

salt (by seawater intrusions), carbonate, organic components, metal-rich deposits, etc., and eventually terrigenous clays in these marine environments are enriched.

Deposits from the 2004 Indian Ocean and 2011 Tohoku-oki tsunamis provided opportunities to study the early stage of taphonomic processes for various types of tsunami deposits. For example, Nakamura et al. (2012) studied tsunami deposits along the Misawa coast in March 2011 and May 2012 and compared their thickness patterns. One year after the event and deposition, deposits were not significantly altered and retained their original structure and thickness. But thin tsunami deposits close to the inundation limit were hardly distinguishable as they had been eroded or mixed with organic soil; consequently, the geometry and inland extent of deposits changed significantly in only 1 year.

Dating Tsunami Deposits

Tsunami deposits can commonly be dated by radiocarbon ^{14}C , ^{210}Pb , ^{137}Cs , optically stimulated luminescence (OSL), and tephrochronological methods employing absolute dating methods. The most frequent methods are AMS radiocarbon and tephrochronology (i.e., volcanic ashes, dating with Ar/Ar, or K/Ar methods). For radiocarbon dating, the time of tsunami inundation is estimated based on the ages of the soil deposited just above and beneath the target tsunami deposit (the so-called bracket dating of the event layer), because the tsunami can have eroded significantly the original surface and again have redeposited the eroded soil within the tsunami layer. If a carbonate shell, preferably articulated, is encountered within the layer, it can also yield perfect ages. Tsunami deposits in peat lands show generally the best accuracy of the dating. Geochemically well-characterized and widely distributed tephra is also very useful to estimate the time of deposition, when tephra and tsunami deposits are close to each other in the stratigraphic record (e.g., in Japan or Italy). Boulder deposits should yield age clustering of the data; otherwise, no events can be modeled. In areas with a long human occupation, artifacts (e.g., ceramics) are excellent to date precisely the date of tsunami impact; furthermore, reconstruction periods can help in identifying tsunami-related structural damage. So archeology provides, besides historical reports, a significant tool in paleotsunami research.

Estimating Paleotsunami Size and Earthquake Magnitude from Tsunami Deposits

The assessment tsunami hazard at a certain coast is crucial; essentially, several parameters need to be known: the size of tsunamis (run-up) at each site, the maximum inundation, and the return periods. As discussed in the previous sections, the lateral extent and spatial distribution of tsunami deposits are significantly affected by the local topography of the coastal area and of the

inundated area and location and type of the source material. Only where a tsunami inundated flat areas and continuous deposits are preserved, the maximum inland extent of the deposits can be correlated and evaluated for obtaining indicators of the present coastal tsunami hazard (other parameters, like coastline retreat, sedimentation and subsidence rates, sea level changes, etc., have to be regarded and assessed as well). Along rocky coasts, (mega-)boulders may indicate past tsunamigenic action. For tsunamis and damage in historical times, reports and interviews of elder persons yield valuable information (Hoffmann et al. 2013b).

Several tsunami scales have been introduced in various parts of the world, both magnitude and intensity based (e.g., the Imamura-Iida Magnitude Scale, Soloviev-Imamura Tsunami Intensity Scale, or a scale by Papadopoulos and Imamura 2001; ITIS Integrated Tsunami Intensity Scale, Lekkas et al. 2013). These scales are of limited use in paleotsunami science, as most of the equations for tsunami magnitude need parameters like maximum or average or local tsunami height or the earthquake magnitude M_W . Parameters are hard to obtain in paleotsunami research, where only minimum parameters can be revealed. The ESI-2007 scale (Guerrieri and Vitori 2007; Guerrieri et al. 2007) refers to tsunamis as “Earthquake Environmental Effects” (EEE), which takes into account and refers to present-day tsunami action and describes and qualifies an intensity. However, unfortunately, the intensity scale lacks exact parameters for the classification (i.e., tsunami height, inundation extent and area, and more).

Summary

Tsunami deposits are highly variable in extent, sediments, and preservation through time. If tsunami deposits are used as paleoseismic indicators, other source mechanisms like mass wasting or volcanic eruption have to be ruled out. Also, in high seismic areas with more than one capable fault, it is difficult to assign tsunami generation to an individual fault rather than to an area.

As paleotsunamis are extremely variable and depending on many parameters to be assessed in the field or lab, also for exact dating, a highly experienced quaternary geologist or sedimentologist should always be consulted for dating tsunami-related sediments and features.

Cross-References

- [EEE Catalogue: A Global Database of Earthquake Environmental Effects](#)
- [Intensity Scale ESI 2007 for Assessing Earthquake Intensities](#)
- [Luminescence Dating in Paleoseismology](#)
- [Paleoseismology of Rocky Coasts](#)
- [Radiocarbon Dating in Paleoseismology](#)

References

- Atwater BF (1987) Evidence for Great Holocene earthquakes along the outer coast of Washington State. *Science* 236(942):944
- Atwater BF, Cisternas M, Bourgeois J, Dudley WC, Hendley JW II, Stauffer PH (1999) Surviving a tsunami – lessons from Chile, Hawaii and Japan. USGS Circular 1187. This report and any updates to it are available online at: <http://pubs.usgs.gov/circ/c1187/>
- Benner R, Browne T, Brückner H, Kelletat D, Scheffers A (2010) Boulder transport by waves: progress in physical modeling. *Zeit Geomorphol* 54(Suppl. 3):127–146. doi:10.1127/0372-8854/2010/0054s3-0022
- Dominey-Howes D, Dengler L, Dunbar P, Kong L, Fritz H, Imamura F, McAdoo B, Satake K, Yalciner A, Yamamoto M, Yulianto E, Koshimura S, Borrero J (2012) International tsunami survey team (ITST) post-tsunami survey field guide, 2nd edn. UNESCO-IOC, Paris
- Etienne S, Buckley M, Paris R, Nandasena AK, Clark K, Strotz L, Chagué-Goff C, Goff J, Richmond B (2011) The use of boulders for characterizing past tsunamis: lessons from the 2004 Indian Ocean and 2009 South Pacific tsunamis. *Earth-Sci Rev*. doi:10.1016/j.earscirev.2010.12.006
- Goff J, Chagué-Goff C, Nichol S, Jaffe BE, Dominey-Howes D (2012) Progress in paleotsunami research. *Sediment Geol* 243–244, 70–88. doi:10.1016/j.sedgeo.2011.11.002
- Goto K, Chavanich SA, Imamura F, Kunthasap P, Matsui T, Minoura K, Sugawara D, Yanagisawa H (2007) Distribution, origin and transport process of boulders deposited by the 2004 Indian Ocean tsunami at Pakarang Cape, Thailand. *Sediment Geol* 202(4):821–837. doi:10.1016/j.sedgeo.2007.09.004
- Goto K, Okada K, Imamura F (2009) Characteristics and hydrodynamics of boulders transported by storm waves at Kudaka Island, Japan. *Mar Geol* 262(1–4):14–24. doi:10.1016/j.margeo.2009.03.001
- Goto K, Miyagi K, Kawamata H, Imamura F (2010) Discrimination of boulders deposited by tsunamis and storm waves at Ishigaki Island, Japan. *Mar Geol* 269(1–2):34–45. doi:10.1016/j.margeo.2009.12.004
- Guerrieri L, Vitori E (eds) (2007) Environment seismic intensity scale 2007-ESI 2007. *MemoriaDescrittivedellacartaGeologicad'Italia*, vol 74, ServizioGeologicod'Italia-DipartimentoDifesa del Suolo. APAT, Roma, 54 pp
- Guerrieri L, Tatevossian R, Vittori E, Comerci V, Esposito E, Michetti A, Porfido S, Serva L (2007) Earthquake environmental effects (EEE) and intensity assessment: the INQUA scale project. *Boll Soc Geol It (Ital J Geosci)* 126(2):375–386
- Hoffmann G, Reicherter K, Wiatr T, Grützner C, Rausch T (2013a) Block and boulder accumulations along the coastline between Fins and Sur (Sultanate of Oman): tsunamigenic remains? *Nat Hazards* 65(1):851–873
- Hoffmann G, Rupprechter M, Balushi NA, Grützner C, Reicherter K (2013b) The impact of the 1945 Makran tsunami along the coastlines of the Arabian Sea (Northern Indian Ocean)—a review. *ZeitschriftfürGeomorphologie Supplementary Issues* 57(4):257–277
- Keating BH, Helsley CE, Wanink M, Walker D (2011) Tsunami deposit research: fidelity of the tsunami record, ephemeral nature, tsunami deposits characteristics, remobilization of sediment by water waves, and boulder movement. In: Möner NA (ed) *The tsunami threat – research and technology*, chapter 19, InTech, Croatia, pp 389–422
- Lekkas E, Andreadakis I, Kostaki I, Kapourani E (2013) A proposal for a new integrated tsunami intensity scale (ITIS-2012). *Bull Seismol Soc Am* 103(2B):1493–1502
- Minoura K, Imamura F, Kuran U, Kono Y, Iwashita T (2001) The 869 Jogan tsunami deposit and recurrence interval of large-scale tsunami on the Pacific coast of northeast Japan. *J Natural Disaster Sci* 23:83–88
- Nakamura Y, Nishimura Y, Putra PS (2012) Local variation of inundation, sedimentary characteristics, and mineral assemblages of the 2011 Tohoku-oki tsunami on the Misawa coast, Aomori, Japan. *Sediment Geol* 282:216–227
- Nanayama F, Satake K, Furukawa R, Shimokawa K, Atwater BF, Shigeno K, Yamaki S (2003) Unusually large earthquakes inferred from tsunami deposits along the Kuril trench. *Nature* 424:660–663
- Papadopoulos G, Imamura F (2001) A proposal for a new tsunami intensity scale. In: ITS 2001. Proceedings session 5:5–1
- Paris R, Fournier J, Poizot E, Etienne S, Morin J, Lavigne F, Wassmer P (2010) Boulder and fine sediment transport and deposition by the 2004 tsunami in

- LhokNga (western Banda Aceh, Sumatra, Indonesia): a coupled offshore-onshore model. *Mar Geol* 268(1–4):43–54. doi:10.1016/j.margeo.2009.10.011
- Reicherter K, Papanikolaou I, Roger J, Mathes-Schmidt-M, Papanikolaou D, Rössler S, Grützner C, Stamatidis G (2010P) Holocene tsunamigenic sediments and tsunami modeling in the Thermaikos Gulf area (northern Greece). *Zeitschrift für Geomorphologie NF Suppl* 54/3:99–126
- Rhodes B, Tuttle M, Horton B, Doner L, Kelsey H, Nelson A, Cisternas M (2006) Paleotsunami research. *EOS* 87(21):205
- Scheffers A (2008) Tsunami boulder deposits. In: Shiki T, Tsuji Y, Yamazaki T, Minoura K (eds) *Tsunamiites. Features and implications*. Elsevier, Amsterdam, pp 299–317
- Scheffers A, Kelletat D (2003) Sedimentologic and geomorphologic tsunami imprints worldwide – a review. *Earth-Sci Rev* 63:83–92. doi:10.1016/s0012-8252(03)00018-7
- Shiki T, Tsuji Y, Yamazaki T, Minoura K (2008) (eds) *Tsunamiites. Features and implications*. Elsevier, Amsterdam, 411 pp
- Szczuciński W (2012) The post-depositional changes of the onshore 2004 tsunami deposits on the Andaman Sea coast of Thailand. *Nat Hazards* 60:115–133
- Wilson R, Hemphill-Haley E, Jaffe B, Richmond B, Peters R, Graehl N, Kelsey H, Leeper R, Watt S, McGann M, Hoirup D, Chague-Goff C, Goff J, Caldwell D, Loofbourrow C (2014) The search for geologic evidence of distant source tsunamis using new field data in California, chap C of Ross SL, Jones LM eds *The SAFRR (Science Application for Risk Reduction) tsunami scenario: U.S. Geological Survey Open-File report 2013-1170-C*, 122 p. doi:10.3133/ofr20131170c

Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations

Daniil Yurchenko
Institute of Mechanical, Process and Energy
Engineering, Heriot-Watt University,
Edinburgh, UK

Synonyms

Autoparametric vibrations; Earthquake engineering; Parametric pendulum; Pendulum mass dampers; Random vibrations; Seismic excitations; Tuned mass dampers

Introduction

In the last hundred years, the height of the tallest buildings (skyscrapers) went up from 283 m (Woolworth Building, New York City, 1913) to 830 m (Burj Khalifa, Dubai, 2013). Building such structures presents various architectural and engineering challenges, including the development of plumbing and air conditioning systems, fast elevators, etc. Nevertheless, the main effort is always directed toward safety and reliability of buildings and occupants. These concerns are not immanent to tall buildings only but important as well for any structures, whether it is a bridge or nuclear plant. Earthquake and wind load are the major reasons for worries when designing structures or buildings, especially tall ones, since these loads are hard to predict.

Tuned mass damper (TMD) is a concept (Den Hartog 2013; Hunt 1979), which has appeared in the beginning of the twentieth century. The concept seemed to be so attractive that it was properly developed and one of the first TMDs were installed in CN Tower, Toronto, and John Hancock Building, Boston, in 1976. Since then the number of different types of implemented TMDs has been rapidly increasing, approaching one hundred (Soto and Adeli 2013), including those which are used to mitigate the motion of tall buildings (Kareem et al. 1999), bridges, pumps, structures, etc. The basic idea behind a TMD is that an oscillatory two-degree-of-freedom (TDOF) system may have such a response that one of the masses (primary mass) will be motionless. Thus, if the response of the primary mass itself was high due to resonance, adding a second mass will significantly reduce the primary mass response amplitude. However, the response attenuation does not happen by default, so the properties of the added mass-spring-damper system have to be tuned to deliver the best outcome; therefore, the name tuned mass damper is used to reflect this feature. A concern may be related to the fact that adding another degree of freedom to the system will result in two peaks of the response amplitude curve rather than one, characteristic for a single-degree-of-freedom (SDOF) system, but this issue can be

handled by proper tuning as well. It has also been shown that the tuning of a stochastic system (a system subjected to a random excitation) is different from that for a deterministic system.

There are different types of devices, which serve as passive TMDs: conventional TMD, liquid TMD (TLD) (Di Matteo et al. 2014 and references therein), and parametric pendulum TMD (PPD). An excellent review paper by Ibrahim (Ibrahim 2008) provides a full description of various passive vibration isolators. It has been argued that the passive TMDs are effective means of mitigating buildings' vibrations due to seismic loading. It became a motivation for the development of a theory of active and hybrid TMDs (Soong and Spencer 2002), which use different types of control to enhance the effectiveness of dampers (Wang and Lin 2007).

Pendulum TMDs have not been studied as thoroughly as conventional TMDs due to the fact that their analytical analysis is much more complicated. Besides being a nonlinear system due to a pendulum restoring force, it is also parametrically excited since the excitation transmitted to the pendulum through its suspension point. Therefore, it is required to study a nonlinear parametric system, which is possible only by using approximate methods.

The entry reviews the fundamental theory of passive tuned mass dampers and parametric pendulum dampers as well as the major developments of the theory in the last hundred years. The motion of the primary mass subjected to a deterministic or stochastic excitation is considered. The principle parameters of the system influencing the dynamics of the primary mass such as mass and frequency ratios are identified. Different formulas derived by a number of authors based on the various optimization criteria are presented for optimal tuning of a damper. The effectiveness of a parametric pendulum tuned mass damper, which interaction with the primary mass is completely different from a conventional tuned mass damper, is discussed. The optimal values of the PPD damping are found numerically for the primary system subjected to a narrow-band noise. Vibrations of the primary mass with a parametric pendulum

tuned mass damper subjected to the three historical earthquake records are presented.

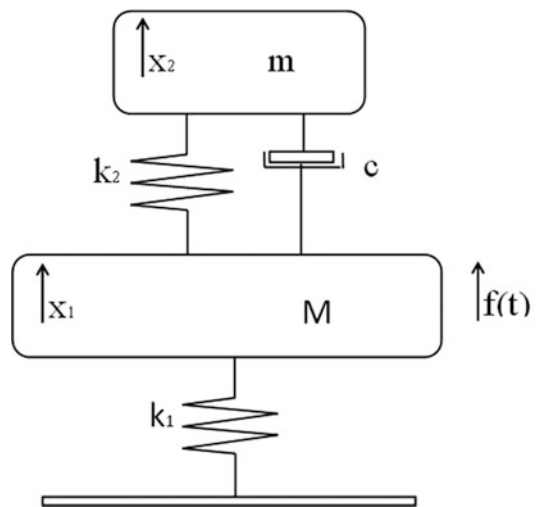
Deterministic Theory of Tuned Mass Dampers

In this section a classical deterministic theory of a tuned mass damper is revisited. This theory is based on the properties of a vibrating TDOF system. The first development of the TMD theory, besides the US patent no. 989958 registered by Frahm back in 1909, was proposed in Ormondroyd (1928) and latter extended to multi-degree-of-freedom (MDOF) systems, as well as nonlinear systems and nonlinear TMDs.

Undamped System

Consider a linear TDOF system with a primary mass subjected to a harmonic force 1, so that dynamics of the system can be described by the following equation of motion (Fig. 1):

$$\begin{aligned} M\ddot{x}_1 + (k_1 + k_2)x_1 - k_2x_2 &= A \sin \omega t \\ m\ddot{x}_2 - k_2x_1 + k_2x_2 &= 0 \end{aligned} \quad (1)$$



Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations, Fig. 1 A primary mass M with a damper of mass m

where M and k_1 are mass and stiffness coefficients of the primary system, whereas m and k_2 are the corresponding coefficients of the secondary mass, served as a damper. Assuming the response of the form

$$\begin{aligned} x_1 &= X_1 \sin \omega t \\ x_2 &= X_2 \sin \omega t \end{aligned} \quad (2)$$

and substituting it into Eq. 1, the following system of two linear algebraic equations is derived:

$$\begin{aligned} X_1[-M\omega^2 + k_1 + k_2] - k_2 X_2 &= A \\ X_2[-m\omega^2 + k_2] - k_2 X_1 &= 0 \end{aligned} \quad (3)$$

Dividing the first equation by k_1 and the second equation by k_2 results in:

$$X_1 \left[1 - \frac{\omega^2}{\Omega_1^2} + \frac{\Omega_2^2}{\Omega_1^2} \mu \right] - \frac{\Omega_2^2}{\Omega_1^2} \mu X_2 = \frac{A}{k_1} \quad (4)$$

$$X_2 \left[1 - \frac{\omega^2}{\Omega_2^2} \right] - X_1 = 0 \quad (5)$$

$$\Omega_1^2 = \frac{k_1}{M}, \Omega_2^2 = \frac{k_2}{m}, \mu = \frac{m}{M} \quad (6)$$

where μ is called the mass ratio and is typically much smaller than unity. Resolving Eqs. 4 and 5 with respect to X_1 and X_2 , one gets the response amplitudes of each system, so that the dynamic amplification factors (DAF) are:

$$\begin{aligned} \frac{X_1}{A/k_1} &= \frac{1 - \frac{\omega^2}{\Omega_2^2}}{\left(1 - \frac{\omega^2}{\Omega_2^2} \right) \left[1 - \frac{\omega^2}{\Omega_1^2} + \frac{\Omega_2^2}{\Omega_1^2} \mu \right] - \frac{\Omega_2^2}{\Omega_1^2} \mu} \\ \frac{X_2}{A/k_1} &= \frac{1}{\left(1 - \frac{\omega^2}{\Omega_2^2} \right) \left[1 - \frac{\omega^2}{\Omega_1^2} + \frac{\Omega_2^2}{\Omega_1^2} \mu \right] - \frac{\Omega_2^2}{\Omega_1^2} \mu} \end{aligned} \quad (7)$$

It can be deduced from the first equation in Eq. 7 that the response amplitude of the primary mass can be zero when the secondary mass tuned to the excitation frequency, i.e., $\omega = \Omega_2$. Thus, the secondary mass acts as a damper, basically absorbing all the input energy, leaving the primary mass motionless. This fundamental result has become the basis for the theory of tuned mass dampers. Figure 2 demonstrates the results of this funding, where the left and right plots display the DAF of the primary mass and damper correspondingly for different values of μ , $v = \omega/\Omega_1$ and $\Omega = \Omega_2/\Omega_1 = 1$.

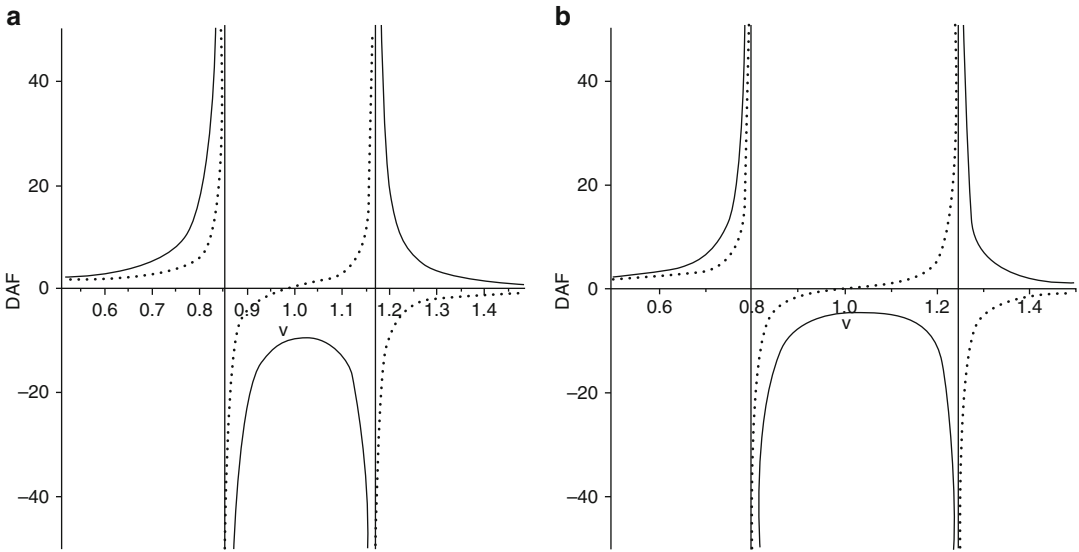
Note that it is not always the primary mass has zero displacement. Figure 3 demonstrates the influence of increasing frequency ratio onto the DAF of the primary mass for $\mu = 0.1$. It can be observed from Fig. 3 that the frequency value corresponding to the zero displacement of the primary mass has been shifted to the left and then vanished completely at $\Omega^2 = 1.5$.

Two natural frequencies of the system can be found from the corresponding characteristic equation, derived by equating the denominator of Eq. 7 to zero:

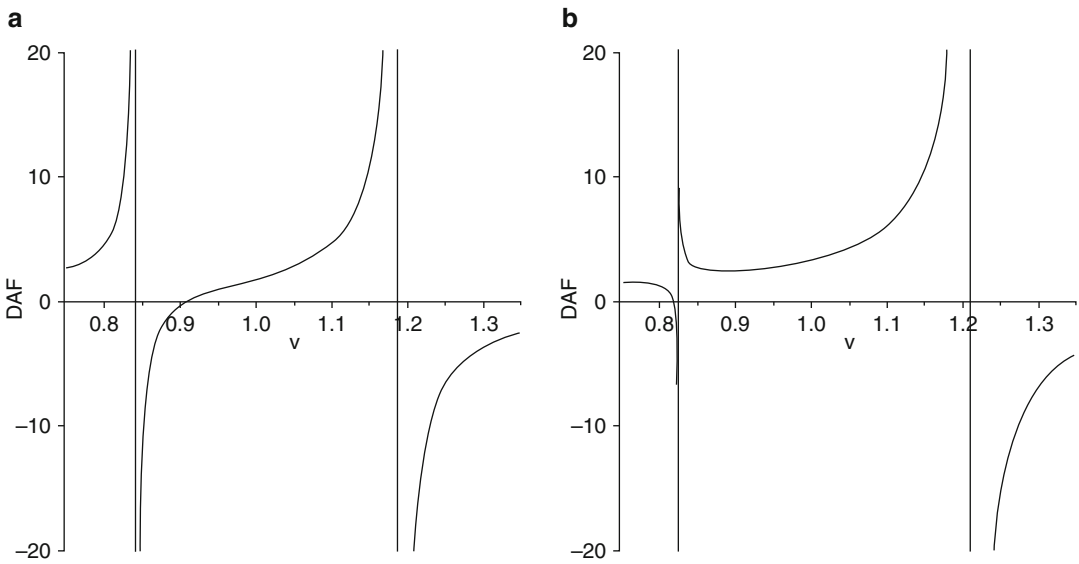
$$\left(1 - \frac{\omega^2}{\Omega_2^2} \right) \left[1 - \frac{\omega^2}{\Omega_1^2} + \frac{\Omega_2^2}{\Omega_1^2} \mu \right] - \frac{\Omega_2^2}{\Omega_1^2} \mu = 0, \quad (8)$$

so that:

$$\frac{\omega^2}{\Omega_2^2} = \frac{\Omega_1^2 + \Omega_2^2(1 + \mu) \pm \sqrt{[\Omega_2^2 \mu + (\Omega_1 + \Omega_2)^2][\Omega_2^2 \mu + (\Omega_1 - \Omega_2)^2]}}{2\Omega_2^2} \quad (9)$$



Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations, Fig. 2 Dynamic amplification factor of the primary mass (*dot*) and TMD (*solid*) from Eq. 7 versus v ratio, (a) $\mu = 0.1$; (b) $\mu = 0.2$



Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations, Fig. 3 Dynamic amplification factor of the primary mass from Eq. 7 versus v ratio, (a) $\Omega^2 = 1.2$; (b) $\Omega^2 = 1.5$

This result coincides with the classical one obtained by Den Hartog (Den Hartog 2013) for the case when both natural frequencies were equal $\Omega_1 = \Omega_2$:

$$\frac{\omega^2}{\Omega_2^2} = \left(1 + \frac{\mu}{2}\right) \pm \sqrt{\mu + \mu^2/4} \quad (10)$$

The assumption of $\Omega_1 = \Omega_2$ is very reasonable since the damper is needed when the primary mass is at resonance, whereas the maximum absorption is achieved, as it has been seen from Eq. 7, when $\omega = \Omega_2$. Expression 10 demonstrates the dependence of the natural

frequencies on the mass ratio, an increase of which broadens the distance between two frequencies.

Damped Absorber

Since any realistic structures and materials have some dissipation mechanisms, it seems reasonable to consider a damper with viscous damping. The equations of motion of the damped system will be similar to the one shown in section [Undamped System](#) with the additional damping term:

$$\begin{aligned} M\ddot{x}_1 + c(\dot{x}_1 - \dot{x}_2) + (k_1 + k_2)x_1 - k_2x_2 &= Ae^{i\omega t} \\ m\ddot{x}_2 + c(\dot{x}_2 - \dot{x}_1) - k_2x_1 + k_2x_2 &= 0 \end{aligned} \quad (11)$$

where i is the imaginary unit. Substituting $x_i(t) = X_i e^{i\omega t}$ into the above equations will lead to the following set:

$$\begin{aligned} X_1[-M\omega^2 + k_1 + k_2 + i\omega c] - (k_2 + i\omega c)X_2 &= A \\ X_2[-m\omega^2 + k_2 + i\omega c] - (k_2 + i\omega c)X_1 &= 0 \end{aligned} \quad (12)$$

Solution to these equations may be written as:

$$\begin{aligned} \frac{X_1}{A} &= \frac{k_2 - m\omega^2 + i\omega c}{(k_1 - M\omega^2)(k_2 - m\omega^2) - k_2m\omega^2 + i\omega c(k_1 - m\omega^2 - M\omega^2)} \\ \frac{X_2}{A} &= \frac{k_2 + i\omega c}{(k_1 - M\omega^2)(k_2 - m\omega^2) - k_2m\omega^2 + i\omega c(k_1 - m\omega^2 - M\omega^2)} \end{aligned} \quad (13)$$

Applying the complex analysis it is easy to derive the expression for the amplification factor of the primary mass:

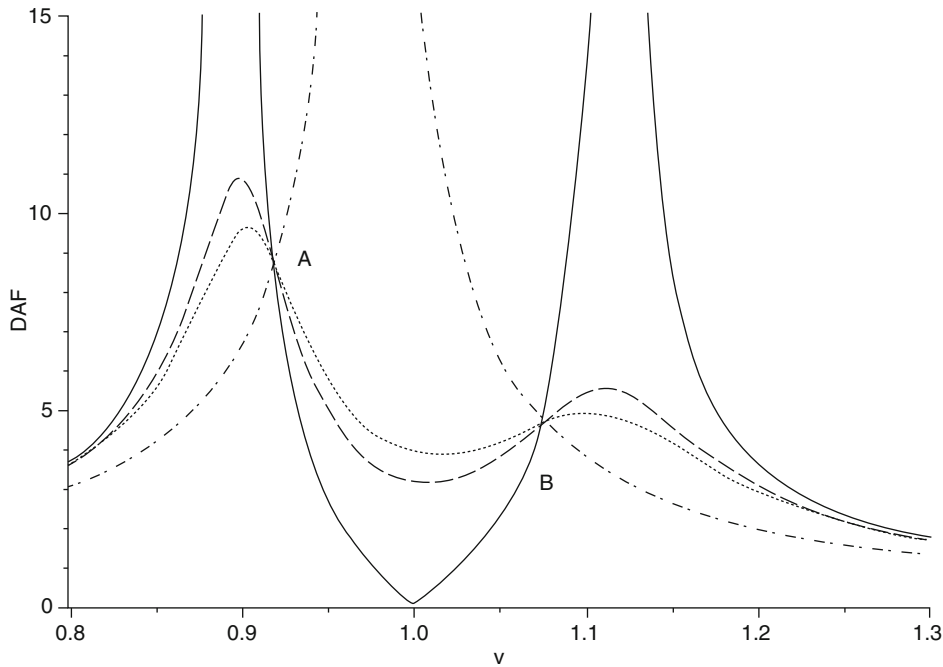
$$\frac{X_1^2}{A^2} = \frac{(k_2 - m\omega^2)^2 + \omega^2 c^2}{[(k_1 - M\omega^2)(k_2 - m\omega^2) - k_2m\omega^2]^2 + \omega^2 c^2(k_1 - m\omega^2 - M\omega^2)^2} \quad (14)$$

Coming back to variables $v = \omega/\Omega_1$ and $\Omega = \Omega_2/\Omega_1$ will reduce Eq. 14 to:

$$\frac{X_1}{A/k_1} = \sqrt{\frac{(v^2 - \Omega^2)^2 + (2\eta v)^2}{[(v^2 - \Omega^2)(v^2 - 1) - \mu v^2 \Omega^2]^2 + (2\eta v)^2(v^2 - 1 + \mu v^2)^2}} \quad (15)$$

where $\eta = c/(2m_2\Omega_1)$. In the limiting case of $\eta = 0$, the result from Eq. 15 tends to the one obtained in Eq. 9, whereas in the opposite case, when the damping is too big ($\eta \rightarrow \infty$), Eq. 15 will be identical to the response of a single-degree-of-freedom system with mass $M + m$ to a harmonic excitation. Thus, there exists an optimal value of the damping coefficient η , which provides the minimal response amplitude of the primary

mass. Figure 4 shows the dependence of the amplification factor Eq. 15 as a function of v for the mass ratio $\mu = 0.05$. Solid line and dash-dot lines correspond to the limiting cases of $\eta = 0$ and $\eta = \infty$, respectively, whereas the dash and dot lines correspond to $\eta = 0.08$ and $\eta = 0.1$. To find the optimal value of the damping coefficient, let us note that for a given value of μ and Ω , all curves independent of the damping coefficient



Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations, Fig. 4 Dynamic amplification factor of the primary mass for difference values of the TMD damping and $\mu = 0.05$

values intersect two points *A* and *B*. Den Hartog proposed (Den Hartog 2013) that the optimal value could be found when the ordinates of these two points were equal and the curve passing these two points should have the horizontal tangent line.

The values at points *A* and *B* can be found by taking into account that all curves can go through the same two points independent of the values of damping coefficient if:

$$\left[\frac{(v^2 - \Omega^2)}{(v^2 - \Omega^2)(v^2 - 1) - \mu v^2 \Omega^2} \right]^2 = \left[\frac{1}{v^2 - 1 + \mu v^2} \right]^2 \quad (16)$$

The nontrivial solution to this equation is:

$$v_A^2 = \frac{1 + (1 + \mu)\Omega^2 - \sqrt{\Omega^4(\mu^2 + 2\mu) + (1 - \Omega^2)^2}}{2 + \mu}$$

$$v_B^2 = \frac{1 + (1 + \mu)\Omega^2 + \sqrt{\Omega^4(\mu^2 + 2\mu) + (1 - \Omega^2)^2}}{2 + \mu} \quad (17)$$

The value of the amplitude at these points can be calculated by simply taking $\eta \rightarrow \infty$, which indicates that:

$$\left| \frac{X_1}{A/k_1} \right| = \frac{\pm 1}{1 - v^2 - \mu v^2} \quad (18)$$

or

$$\frac{1}{1 - v^2 - \mu v^2} = \frac{-1}{1 - v^2 - \mu v^2} \quad (19)$$

The latter results in the following equality:

$$v_A^2 + v_B^2 = \frac{2}{1 + \mu} \quad (20)$$

Having combined this result with a sum of $v_A^2 + v_B^2$ from Eq. 17, one arrives to the following expression:

$$\Omega = \frac{1}{1 + \mu} \quad (21)$$

For small values of the mass ratio, formula (21) indicates that the damper should be tuned to the natural frequency of the primary mass. To find the optimal value of the damping coefficient, Brock (Brock 1946) has differentiated expression (16) with respect to ν and equated it to zero at points A and B , correspondingly. By doing so he was able to recover two optimal values of the damping coefficient:

$$\eta_A^2 = \frac{\mu \left(3 - \sqrt{\frac{\mu}{\mu+2}} \right)}{8(1+\mu)^3} \quad (22)$$

$$\eta_B^2 = \frac{\mu \left(3 + \sqrt{\frac{\mu}{\mu+2}} \right)}{8(1+\mu)^3},$$

which resulted in the optimal mean value:

$$\eta_o^2 = \frac{3\mu}{8(1+\mu)^3} \quad (23)$$

The expression for the optimal value of the damping coefficient will be different when $\Omega = 1$:

$$\eta_c^2 = \frac{\mu(3+\mu) \left[1 + \sqrt{\mu/(2+\mu)} \right]}{8(1+\mu)} \approx \frac{3\mu}{8(1+\mu)} \quad (24)$$

Another approach for defining the optimal value of η was proposed by Krenk (2005), who suggested to equate the ordinates of three points A , B , and the point in the middle of the previous two points. By doing so he has derived a new value for the optimal damping coefficient, which is 15 % larger than the classical result (Eq. 24):

$$\eta_{Kr}^2 = \frac{\mu}{2(1+\mu)} \quad (25)$$

It has been shown that the above optimal set of parameters can be used to minimize the response of a continuous system, provided that the equivalent mass and stiffness are used and its natural frequencies are well separated. Warburton (1982) has investigated the influence of the optimal values of μ and Ω , derived in Den Hartog

(2013), onto the response of a TDOF primary system and the contribution of the higher modes to the overall response. A translation-rotation TDOF vibration absorber instead of an SDOF one was proposed in Jang et al. (2012), where the method of choosing the optimal set of parameters of the absorber was developed. Since the real structure has always some amount of damping, even when the latter is modeled as a viscous one, it becomes difficult to derive an analytical expression for design purposes; therefore, some methods have been developed to overcome this issue (Tsai and Lin 1993; Vakakis and Paipetis 1986; and references therein). Moreover, some experimental data, collected from the structures subjected to earthquakes, suggested that the energy dissipation mechanism was not linear but rather hysteretic. A system subjected to six historical earthquake records with a hysteresis damping model was investigated in Val and Segal (2005). It has been shown that the linear viscous damping model may not be accurate enough to describe the system behavior due to earthquake and can provide up to 30 % difference in the maximum displacement.

Stochastic Theory of Tuned Mass Dampers

The response of a linear system to a broadband excitation is different to one due to a harmonic excitation, since the former happens with the natural frequency(ies) of the system. In cases when the broadband excitation has an almost constant spectrum over the range of the natural frequencies, it is convenient to replace such an excitation with its limiting cases – a white noise. Without loss of generality consider a linear TDOF system with the primary mass subjected to a white noise excitation:

$$\begin{aligned} M\ddot{x}_1 + c_2(\dot{x}_1 - \dot{x}_2) + c_1\dot{x}_1 + (k_1 + k_2)x_1 - k_2x_2 &= M\ddot{\xi}(t) \\ m\ddot{x}_2 + c_2(\dot{x}_2 - \dot{x}_1) - k_2x_1 + k_2x_2 &= 0 \end{aligned} \quad (26)$$

Dividing both equations by the corresponding mass set (Eq. 26) can be written as the following:

$$\begin{aligned} \ddot{x}_1 + 2\zeta_2\Omega_2\mu(\dot{x}_1 - \dot{x}_2) + 2\zeta_1\Omega_1\dot{x}_1 \\ + (\Omega_1^2 + \Omega_2^2\mu)x_1 - \Omega_2^2\mu x_2 = \xi(t) \\ \ddot{x}_2 + 2\zeta_2\Omega_2(\dot{x}_2 - \dot{x}_1) - \Omega_2^2x_1 + \Omega_2^2x_2 = 0 \\ E[\xi(t)] = 0, E[\xi(t)\xi(t+\tau)] = D\delta(\tau) \end{aligned} \quad (27)$$

The first equation in Eq. 27 is different from Eq. 11 by the third term, which represents damping in the primary system, and $\xi(t)$ is zero-mean Gaussian delta-correlated white noise with intensity D . It is well known that the mean and the mean square response of an SDOF subjected to a white noise is ($m_2 = 0$, $k_2 = 0$, $c_2 = 0$):

$$E[x_1] = 0, E[x_1^2] = \frac{D}{4\zeta_1\Omega_1^3} \quad (28)$$

where $E[\cdot]$ denotes the averaging. Some methods for analyzing linear stochastic systems may be found in many textbooks, for instance, Crandall and Mark (1963). System (27) can be analyzed analytically using the method of moments or the spectral density and convolution integral approach. The latter has been applied by Crandall and Mark (Crandall and Mark 1963) to the similar system but excited through its base. Since the mean values of the responses are zero, it is required to calculate the second-order moments, which can be done by using the matrix approach

with $z_1 = x_1$, $z_2 = \dot{x}_1$, $z_3 = x_2$, $z_4 = \dot{x}_2$. Rewrite Eq. 27 in the form:

$$\begin{aligned} \dot{\mathbf{Z}} &= \mathbf{S}\mathbf{Z} + \mathbf{B}\xi(t) \\ \mathbf{Z} &= \begin{pmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ \mathbf{S} &= \begin{pmatrix} 0 & 1 & 0 & 0 \\ -\Omega_1^2 - \mu\Omega_2^2 & -2\zeta_2\Omega_2\mu - 2\zeta_1\Omega_1 & \mu\Omega_2^2 & 2\mu\zeta_2\Omega_2 \\ 0 & 0 & 0 & 1 \\ \Omega_2^2 & 2\zeta_2\Omega_2 & -\Omega_2^2 & -2\zeta_2\Omega_2 \end{pmatrix} \end{aligned} \quad (29)$$

The equation for the second-order moments is:

$$\dot{\mathbf{M}} = \mathbf{S}\mathbf{M} + \mathbf{M}\mathbf{S}^T + \mathbf{D}\mathbf{B}\mathbf{B}^T \quad (30)$$

where $\mathbf{M}$ is a symmetric matrix of moments:

$$\mathbf{M} = \begin{pmatrix} M_{11} & M_{12} & M_{13} & M_{14} \\ M_{12} & M_{22} & M_{23} & M_{24} \\ M_{13} & M_{23} & M_{33} & M_{34} \\ M_{14} & M_{24} & M_{34} & M_{44} \end{pmatrix} \quad (31)$$

with the matrix elements defined as $M_{ij} = E[z_i z_j]$. It is not difficult to find a steady-state ($\dot{\mathbf{M}} = 0$) solution of Eq. 30; however, only two moments M_{11} and M_{33} are of special interest:

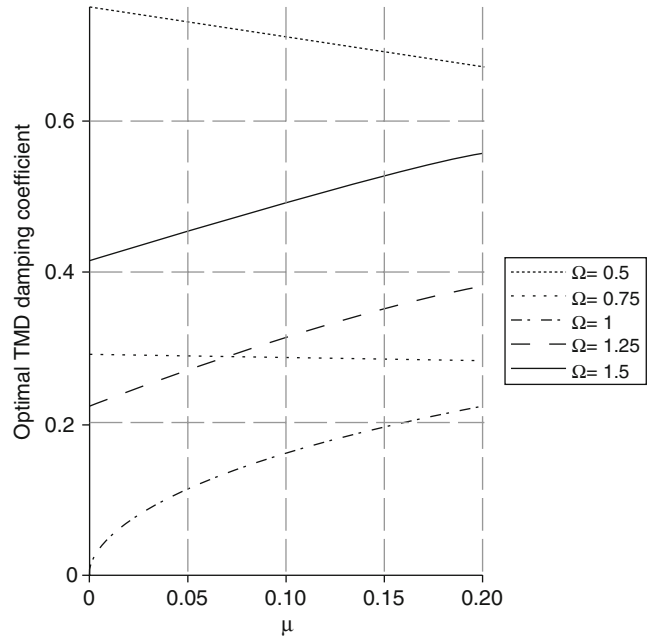
$$\begin{aligned} M_{11} = M_{x_1 x_1} &= \frac{D}{4\Omega_1^3} \frac{C_1 \zeta_1^2 + C_2 \zeta_1 + C_3}{D_1 \zeta_1^3 + D_2 \zeta_2^2 + D_3 \zeta_1 + D_4} \\ M_{33} = M_{x_2 x_2} &= \frac{D}{4\Omega_1^3} \frac{\Omega_2 [E_1 \zeta_1^2 + E_2 \zeta_1 + E_3]}{D_1 \zeta_1^3 + D_2 \zeta_2^2 + D_3 \zeta_1 + D_4} \\ C_1 &= 4\Omega_1^2 \Omega_2^2 \zeta_2, C_2 = 4\Omega_1 \Omega_2^3 \zeta_2^2 (1 + \mu) + 4\zeta_2^2 \Omega_1^3 \Omega_2 + \mu \Omega_1 \Omega_2^3 \\ C_3 &= \zeta_2 \left\{ (\Omega_1^2 - \Omega_2^2)^2 + \Omega_2^2 [\Omega_2^2 (2\mu + \mu^2) + 4\Omega_1^2 \zeta_2^2 (\mu + 1) - \mu \Omega_1^2] \right\} \\ E_1 &= 4\zeta_2 \Omega_1^2 \Omega_2, E_2 = \Omega_1^3 (1 + 4\zeta_2^2) + 4\zeta_2^2 \Omega_1 \Omega_2^2 (1 + \mu) + \mu \Omega_1 \Omega_2^2 \\ E_3 &= \zeta_2 \Omega_2 \left[(1 + \mu)^2 \Omega_2^2 + \Omega_1^2 (\mu + 4\mu \zeta_2^2 + 4\zeta_2^2) \right] \\ D_1 &= C_1, D_2 = C_2, D_3 = C_3 + \mu \zeta_2 \Omega_1^2 \Omega_2^2, D_4 = \mu \zeta_2^2 \Omega_2 \Omega_1^3 \end{aligned} \quad (32)$$

Since the above result is cumbersome, the optimal value of the damping coefficient ζ_2 is usually obtained numerically. It should be stressed that

the system response is finite when one of the damping coefficients is equal to zero. Putting $\zeta_1 = 0$ will reduce the expression for M_{11} to:

Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations, Fig. 5

The optimal damper value ζ_2 for different values of μ and Ω , while $\zeta_1 = 0$



$$M_{11} = \frac{D}{4\Omega_1^3} \frac{C_3}{D_4} \quad (33)$$

Looking for ζ_2^* which delivers the minimum value of M_{11} , one differentiates the latter with respect to ζ_2 and equates the expression to zero. This results in the optimal value of ζ_2 :

$$\zeta_2^* = \frac{\sqrt{(\mu+1) [\Omega^4(\mu^2+2\mu) - \mu\Omega^2 + (1-\Omega^2)^2]}}{2\Omega(\mu+1)} \quad (34)$$

where Ω is defined in section **Damped Absorber**. Figure 5 shows the dependence of the optimal damping coefficient on the mass ratio for different values of the frequency ratio Ω . It can be seen that the case of $\Omega = 1$ quantitatively agrees with one declared by Wiesner (1979) for any value of ζ_1 .

It should be emphasized that the optimal value of the damping coefficient depends significantly on the frequency ratio, which is observed in Fig. 5. For the values of $\Omega_2 < \Omega_1$, the optimal value of ζ_2 is much higher than the corresponding value for the case of $\Omega_2 > \Omega_1$ and decreases with

increase of μ . It can also be observed that the value of ζ_2 is almost a constant for $\Omega = 0.75$

In the case of $\Omega = 1$ this result reduces to:

$$\bar{\zeta}_2 = \frac{\sqrt{\mu}}{2} \quad (35)$$

The result from expression (35) is compared in Fig. 6 to the classical result derived by Den Hartog (Eq. 24) and the result recently obtained by Krenk (Eq. 25). In Fig. 7 one can see the difference in performance of the primary mass when ζ_2 is calculated based on formulas (34) and (35).

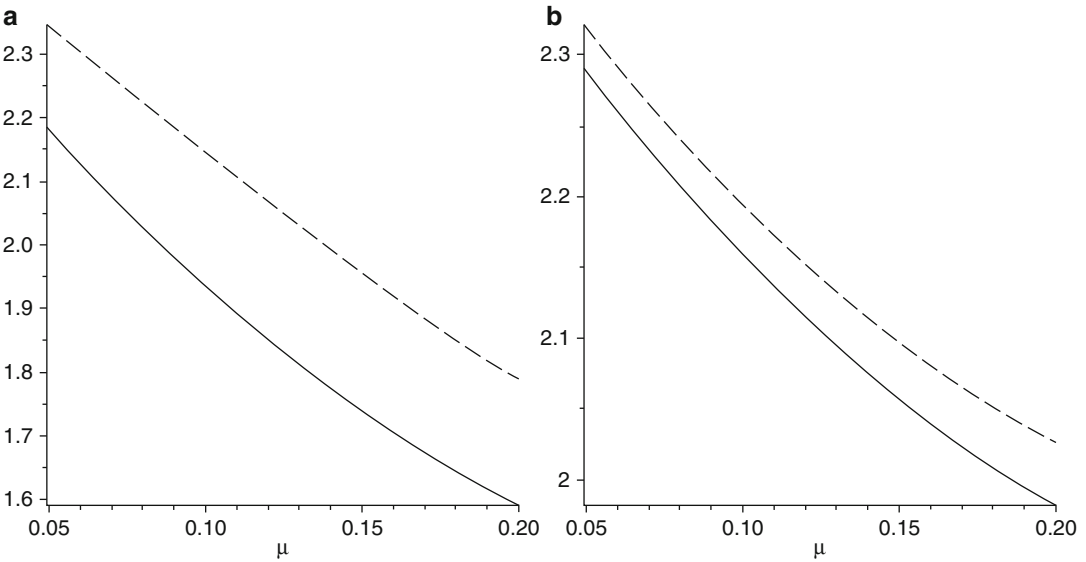
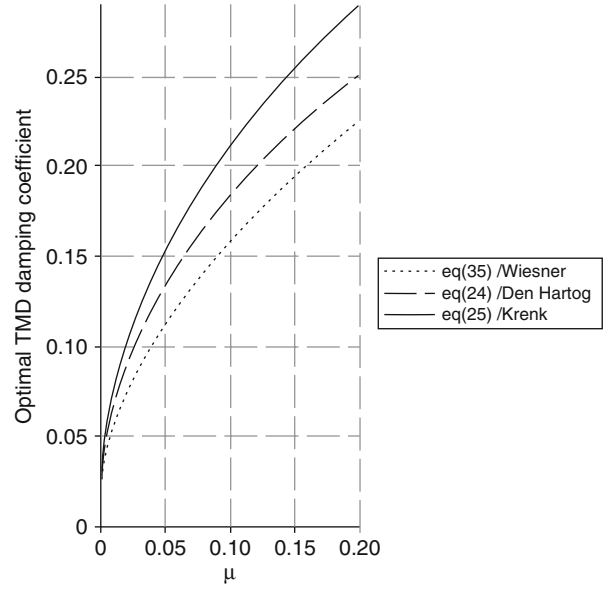
Warburton in 1981 reported that formula (35) provided the minimum mean square velocity of the primary mass, whereas ζ_2^D and ζ_2^A :

$$\begin{aligned} \zeta_2^D &= \sqrt{\frac{\mu(1+3\mu/4)}{4(1+\mu)(1+\mu/2)}} \\ \zeta_2^A &= \sqrt{\frac{\mu(1-\mu/4)}{4(1+\mu)(1-\mu/2)}} \end{aligned} \quad (36)$$

provide the minimum displacement and acceleration of the primary mass subjected to a white noise excitation. The recommendations for the optimal set of parameters of a TMD can also be

Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations,

Fig. 6 Comparison of different values of the optimal damper coefficient ζ_2 as a function of μ



Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations, Fig. 7 Comparison of M_{11} for ζ_2 calculated based on Eq. 34—solid and Eq. 35—dash, (a) $\Omega = 0.5$; (b) $\Omega = 1.5$

found in Warburton (1981) for the deterministic case, i.e., when the primary mass is under a harmonic excitation.

It is obvious from Fig. 7 that the mean square response is lower when ζ_2^* ; however, this difference diminishes quickly as Ω trends to unity and then reappears again as Ω increases above unity.

Similar observations and conclusions were reported in Krenk and Høgsberg (2008) for the case of $\zeta_1 = 0$, whereas for the case of $\zeta_1 \neq 0$, the final expression was different. Nevertheless, the observed difference was so small that the recommendation was given to disregard the structural damping when designing a TMD.

A similar analysis can be applied to a linear multi-degree-of-freedom system to derive the mean square response of the system; however, the expressions for the second-order moments become extremely lengthy and difficult to work with; therefore, such systems are investigated numerically or experimentally (see, e.g., Adam et al. (2003)). The systems with 2, 5 and 10 degrees of freedom under a Gaussian white noise were numerically investigated by Wirsching and Cambell (1973). Their primary interest was how a TMD, which was supposed to reduce a motion of only a certain degree of freedom, would perform in the presence of other modes. They concluded that a TMD was capable of reducing the structural motion of the system by up to 30 % when $\mu = 0.1$ and produced a curve for the optimal damping coefficient of the damper, which was very similar to Eq 35. A different expression of the optimal damping coefficient was reported for an SDOF system in the presence of structural damping (Sadek et al. 1997):

$$\zeta_2^R = \frac{\zeta_1}{1 + \mu} + \sqrt{\frac{\mu}{1 + \mu}} \quad (37)$$

The latter was numerically obtained when the primary mass was subjected to various historical earthquake records. Note that expression (37) with $\zeta_1 = 0$ has been recently proposed in Krenk and Høgsberg (2014) based on bifurcation analysis.

Although the theoretical results demonstrate the effectiveness of a TMD, in fact there is a certain concern whether this is a reliable mean of protecting buildings from seismic vibrations. Sladek and Klingner (Sladek and Klingner 1983) studied an 8-node linear elastic building model subjected to a component of a realistic El Centro earthquake which happened in 1940. They were among the firsts who found that the optimally tuned TMD had no effect on the maximum displacement of the tip of the building model. Chowdhury et al. (1987) have also investigated the influence of a TMD onto a higher mode response by considering the primary structure as a vertically oriented cantilever beam the top of which was fitted with a TMD. Six historical

earthquake records were used to assess the performance of the primary system. They found that the higher mode response actually increased when the TMD was tuned to the fundamental frequency of the primary mass. Moreover, they have observed that although the overall response with a TMD is reduced, the peak value was not affected at all. It has also been reported that if the primary mass peak occurs without a significant pre-peak motion, then the passive TMD cannot get enough excitation and would not be effective; therefore, the effectiveness of the TMD depends not only on the tuning frequency and mass ratio but also on the pre-peak motion of the primary mass, the magnitude, and the acceleration of the ground motion (Chowdhury et al. 1987). In other words a passive TMD cannot give response within a short period of time, which is the reason for discrepancies between the transient and steady-state results. Based on that a number of authors (Chowdhury et al. 1987; Kaynia et al. 1981; Sladek and Klingner 1983) expressed skepticism toward the unconditional usage of TMDs.

The higher mode behavior of a multi-degree-of-freedom system with a TMD implies that the latter helps to transfer the system energy from lower to higher modes. There are some ways of dealing with this problem. For instance, the TMD energy can be dissipated by introducing an impact damper (Dimentberg and Iourtchenko 2004; Ibrahim 2009) on the top of the TMD. It seems also expedient for improving the structural performance to introduce multiple TMDs (MTMD) in a parallel or series configuration. The performance of spatially distributed MTMD was originally reported in Bergman et al. (1989) and later, for instance, in Kareem and Kline (1995) and references therein. The reported results indicated that the structural behavior could be improved when the MTMD were tuned to the corresponding frequencies of the structure, keeping the overall mass ratio the same as for a TMD. It has been shown that the increase in a number of dampers results in flattening out the response curve over a wider range of frequencies (Debbarma and Hazari 2013; Kareem and Kline 1995). The performance comparison of the TMD

and MTMD with 21 units suggested that the values of the optimal damper damping of MTMD were lower than that of a TMD. Later a number of algorithms were proposed for selecting the optimal values of the MTMD parameters (see Zuo and Nayfeh (2006) and references therein). The optimal value of a TMD damping coefficient based on various optimization procedures, including the optimization of the mean square displacement, velocity, or acceleration of the response for a variety of loadings, may be found in Bisegna and Caruso (2012), Sadek et al. (1997) and Tigli (2012). Efforts have been made to develop a theory of a nonlinear TMD, where either nonlinear stiffness or damping was involved. Recently a tuned mass damper with nonlinear damping was investigated (Carpineto et al. 2014; Rüdinger 2007) to understand the influence of nonlinearities onto the performance of a primary mass. A very simple model of a nonlinear stiffness TMD is a parametric pendulum TMD (PPD), which is discussed next.

Pendulum Tuned Mass Damper

Traditionally a TMD requires significant space for installation, which may cause significant problems for architectural design. In this case a PPD seems to be a very reasonable alternative. It should be stressed that the horizontal/vertical floor motion due to seismic activity will excite the pendulum's suspension point, therefore creating a parametric rather than external excitation, which happens in a conventional TMD. Therefore, it becomes essential to study the properties of a parametrically excited system. To explain the basic properties of a parametrically excited lumped mass pendulum, let us write its equation of motion in the following form:

$$\ddot{\theta} + 2\alpha\dot{\theta} + \left[\frac{g}{l} + \frac{\ddot{f}_v}{l} \right] \sin \theta + \frac{\ddot{f}_h}{l} \cos \theta = 0 \quad (38)$$

where f_v, f_h are the vertical and horizontal components of the excitation, and α is the damping

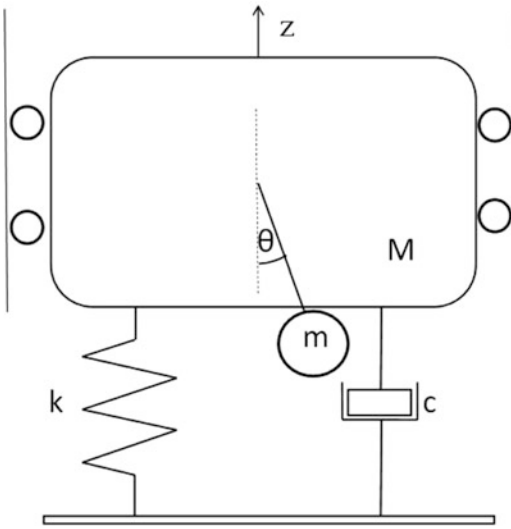
coefficient. It is well known that this system has a number of resonance frequencies defined as $\nu = 2/n$, $n = 1, 2, 3$, with the primary parametric resonance identified at the value of the doubled natural frequency. This system has been studied extensively for the deterministic and stochastic loadings (Bobryk and Chruszczak 2009; Xu and Wiercigroch 2007; Yurchenko et al. 2013; and references therein). Note that due to the nature of excitation entering the system parametrically, the pendulum has to be perturbed to start the oscillations.

One of the benefits using a PPD has already been mentioned and is related to the fact that it does not require much horizontal space. However, since the natural frequency of a lumped mass pendulum is inversely proportional to the square root of its length, a PPD may require excessive vertical space. To avoid this, there is a number of designs of small pendulums with very low natural frequencies (Ibrahim 2008; Yurchenko and Alevras 2013a). Another potential benefit of a PPD is that it can absorb the energy of multidirectional horizontal and vertical vibrations, whereas a conventional TMD can only be effective in a single direction. Until recently it has been common to assume that damages in buildings and structures were due to a horizontal relative motion of floors caused by a seismic event in a horizontal direction. However, field observations pointed toward damage patterns, which could only be caused by a vertical oscillations. Moreover, the vertical to horizontal spectral ratio was found to be above the overall accepted, and in some cases the vertical component was dominant (Elgamal and He 2004). It has also been found that higher frequencies (over 8Hz) can be expected from the vertical records rather than the horizontals. An overview on the vertical earthquake ground motion can be found in Elgamal and He (2004). Thus, it seems that a PPD has a certain benefits compared to its linear analogy.

Since the main goal is to investigate whether the primary mass response will be lower with a PPD, consider the dynamics of the following system with the primary mass subjected to a narrow-band excitation:

$$\begin{aligned}\ddot{z} + 2\alpha\dot{z} + \Omega_1^2 z &= f(t) + \gamma(\ddot{\theta} \sin \theta + \dot{\theta}^2 \cos \theta), \\ \ddot{\theta} + 2\beta\dot{\theta} + (\Omega_2^2 - \ddot{z}) \sin \theta &= 0, \\ \gamma &= \mu / (1 + \mu),\end{aligned}\quad (39)$$

Figure 8 shows the sketch of the system described by Eq. 39, which is qualified as an autoparametric system (Cartmell and Cartmell 1990; Tondl et al. 2000) and has been studied by a number of

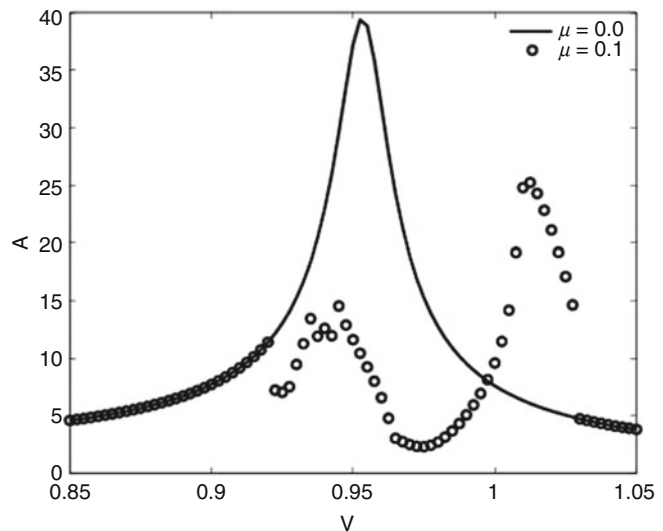


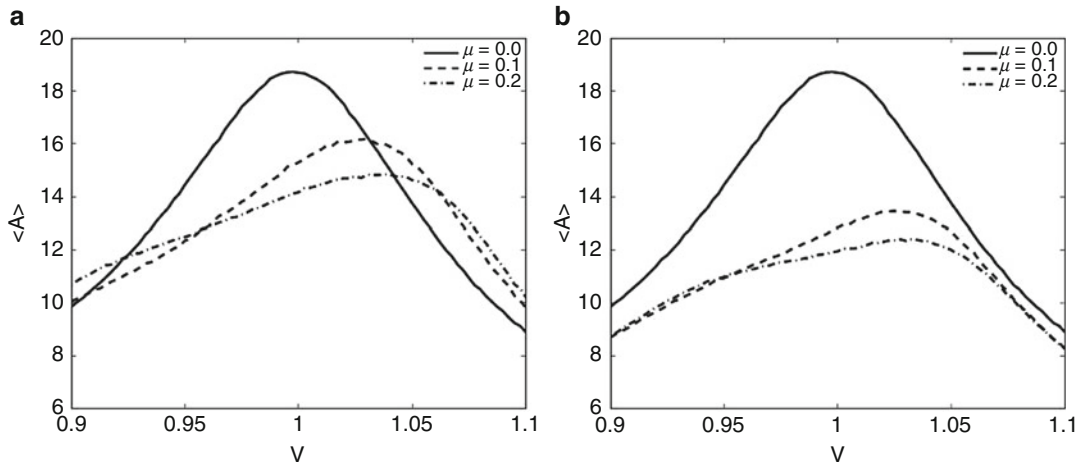
Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations, Fig. 8 A sketch of a primary system model with PPD

authors on the issue of stability of the pendulum. To analyze system Eq. 39 analytically, its inherent nonlinearity of *sine* and *cosine* type is either linearized or expressed in the Taylor series, keeping two or three first terms, thereby reducing to a polynomial nonlinearity. Deterministic response of system Eq. 39 ($f(t) = A \sin \omega t$) can be found in Song et al. (2003) and Warminski and Kecik (2009). The approximate analytical results for the third-order approximation derived in Song et al. (2003) agreed well with numerical results, presented in Fig. 9. The latter demonstrates the primary system deterministic response amplitude for mass ratio $\mu = 0$ (no PPD) and $\mu = 0.1$ versus ν with frequency ratio $\Omega = 2$. Clearly one can observe two peaks and the minimum value of the response amplitude in the case of $\mu = 0.1$, although the trough is shifted to the right from its original resonance frequency. Deterministic dynamics of a PPD comprising two independent pendulums was reported in Ikeda (2011), whereas the optimization of a friction PPD, implemented on Taipei 101 to reduce the wind load, was presented in Chung et al. (2013).

Stochastic response of Eq. 39 to a Gaussian white noise $f(t) = \xi(t)$ was investigated in Ibrahim and Roberts (1976). That paper considered vibrations of an autoparametric system consisting of a primary mass, subjected to a broadband excitation, and a vertically oriented cantilever beam.

Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations, Fig. 9 The deterministic response amplitude of the primary mass with ($\mu = 0.1$) and without ($\mu = 0.0$) PPD





Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations, Fig. 10 Comparison of the mean response amplitude of the primary mass for $\zeta_1 = 0.01$, $\Omega = 2.0$ and (a) $\zeta_2 = 0.00$; (b) $\zeta_2 = 0.01$

The dynamics of this system is very similar to the dynamics of the system, shown in Fig. 8.

Consider a linearized set of equations (Eq. 39) subjected to a zero-mean Gaussian white noise $\xi(t)$ in slightly another form:

$$\ddot{z} + 2\alpha\dot{z} + \Omega_1^2 z = \xi(t) + \varepsilon\gamma(\ddot{\theta}\theta + \dot{\theta}^2), \quad (40)$$

$$\ddot{\theta} + 2\beta\Omega\dot{\theta} + (\Omega^2 - \varepsilon\dot{z})\theta = 0$$

Eliminating the second-order derivatives in the right-hand side of the above equations, one arrives to:

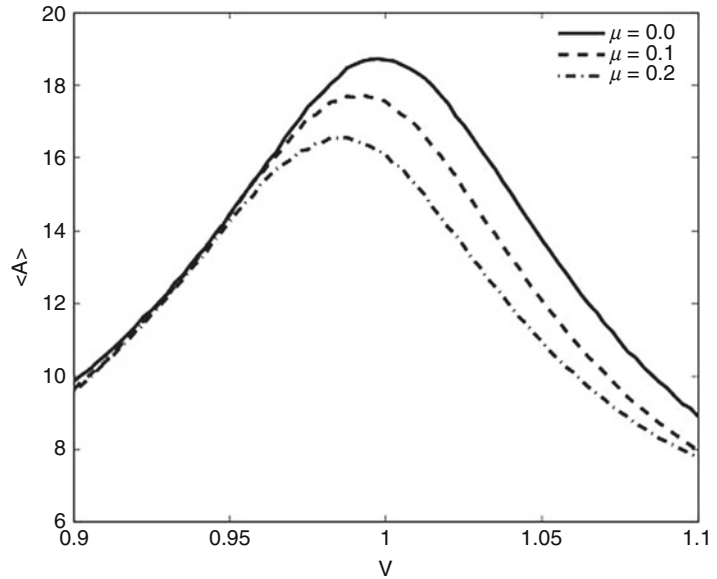
$$\begin{aligned} \dot{x}_1 &= x_3 \\ \dot{x}_2 &= x_4 \\ \ddot{x}_3 &= -2\alpha x_3 - x_1 + \xi(t) \\ &\quad + \varepsilon\gamma(2\Omega\beta x_2 x_4 + \Omega^2 x_2^2 - x_4^2) \\ \dot{x}_4 &= -2\Omega\beta x_4 - \Omega^2 x_2 + \varepsilon(x_2 \xi(t) \\ &\quad - x_1 x_2 - 2\alpha x_2 x_3) - \varepsilon^2 \gamma x_2 x_4 \end{aligned} \quad (41)$$

Similar to Ibrahim and Roberts (1976), this set can be studied by the method of moments and Gaussian closure. It should be stressed that none of the papers so far, including the abovementioned, were focused on the optimization of the system's parameters to achieve the minimal displacement or velocity of the primary mass.

The behavior of the system similar to Ibrahim and Roberts (1976) has been studied by the method of moments (Cho et al. 2003) when the primary mass was excited by a narrow-band noise $f(t) = \lambda \cos q(t)$, $\dot{q}(t) = \omega + \sigma\xi(t)$. Recently the stochastic nonlinear dynamics of the original system (39) subjected to a narrow-band excitation has been investigated with the purpose of identifying the possible rotational domains of the pendulum (Yurchenko and Alevras 2013b). The building response with a PPD was studied numerically and experimentally in Majcher and Wójcicki (2014). The reported results indicated that there was a process of energy exchange between the primary system and the PPD, as expected. It would seem logical to expect that the minimal displacement (energy) of the primary mass corresponds to the maximum pendulum energy; however, the latter will go to infinity (become unstable) for the linearized pendulum model, leading to the instability of the primary mass. Thus, the derived analytical solution to the linearized model, although it brings some light onto the system dynamics, does not provide complete information about it. The numerical results of the original nonlinear system (Eq. 39) are presented in Figs. 10 and 11 for noise intensity $\sigma^2 = 0.1$. Since the original nonlinearity allows the pendulum to have a rotational motion, Fig. 11 presents the results when the pendulum

Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations, Fig. 11

The mean amplitude of the primary mass with a PPD for different values of $\mu = 0.0$, the pendulum in 65 % rotational mode



has 65 % of the rotational motion, whereas in Fig. 10 the pendulum is in oscillatory mode all the time.

The overall effect of the bandwidth widening can be seen in Figs. 10 and 11 compare to Fig. 8, which is the result of random fluctuations of the phase. Furthermore, one can observe in Fig. 10 that the zero damping of the pendulum (Fig. 10 (left)) is not the optimal value of the damping ratio since $\zeta_2 = 0.01$ (Fig. 10 right) provides lower response amplitude. Although both Figs. 10 and 11 demonstrate the influence of a PPD onto the primary mass, the overall mean square response of the primary mass in Fig. 10 is smaller than that in Fig. 11 for the identical set of the parameters' values. Such a behavior can be explained by the centrifugal force, arising when the pendulum starts rotating, so that the latter pulls the system up and down.

The search for the optimal value of ζ_2 as a function of μ is conducted in this entry for nonlinear system (Eq. 35) subjected to a narrow-band excitation. It should be noted that for some combinations of ζ_2 and μ , the primary mass response has only a single peak, whereas for

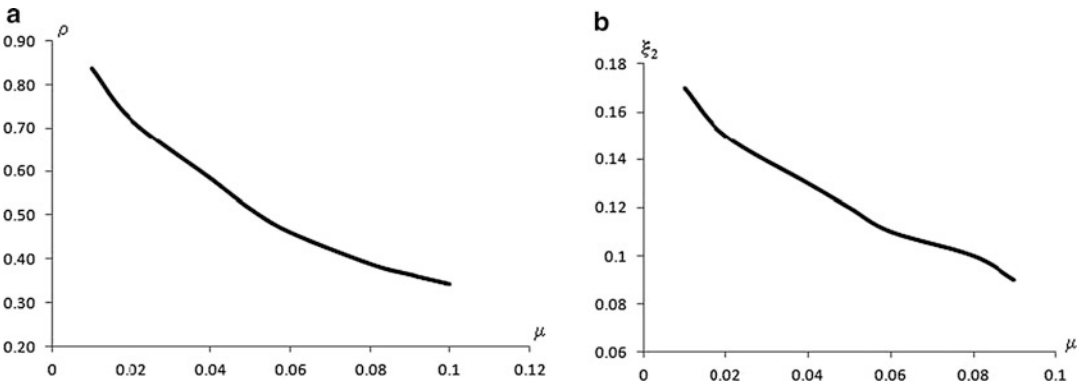
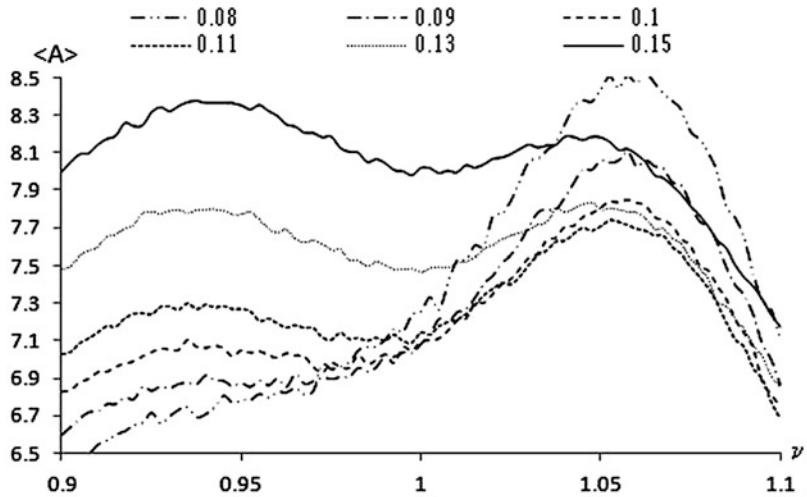
others it has two peaks. Figure 12 shows the transition from one to two peaks for $\mu = 0.08$ and six different values of ζ_2 . The overall results of the numerical Monte Carlo simulations are presented in Fig. 13.

Figure 13 (left) presents the results of the nondimensional amplitude ratio $\rho = \langle A \rangle / \langle A_0 \rangle$, where $\langle A_0 \rangle$ is the mean response amplitude of the primary mass without a PPD, as a function of the mass ratio μ . Amplitude attenuation up to 60 % can be observed in that figure for small values of μ . Figure 13 (right) shows the optimal values of ζ_2 as a function of the mass ratio. Note that contrary to the TMD cases, the optimal value of the PPD damping is reducing with increasing μ .

Responses of a PPD to three earthquake records are presented in Fig. 14. All figures demonstrate the results of numerical modeling for $\mu = 0.0$ (no PPD), $\mu = 0.01$ and $\mu = 0.1$. Figure 14a demonstrates the record of El Centro earthquake and Fig. 14b shows the corresponding response of a primary mass with a PPD. It can be seen that a PPD is not very effective neither for small nor for bigger value of the mass ratio.

Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations, Fig. 12

The mean amplitude of the primary mass with a PPD for different values of ζ_2 and $\mu = 0.08$

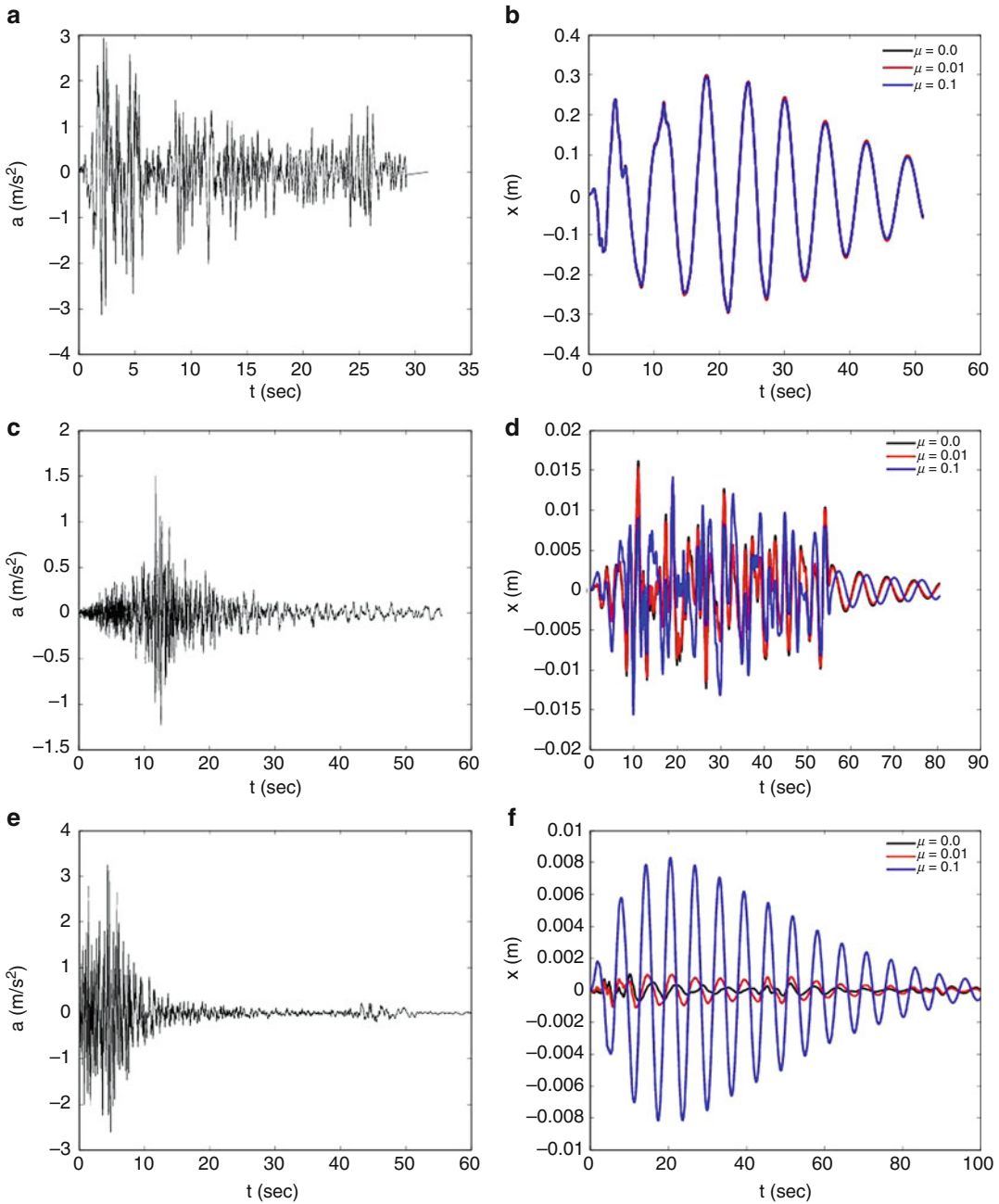


Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations, Fig. 13 Nondimensional amplitude ratio ρ versus μ (a) and optimal values of $\zeta_2 = 0.01$ versus μ (b) for $\sigma^2 = 0.1$ and $\zeta_1 = 0.05$

The next two plots in Fig. 14c and d present the results for the Nisqually earthquake. The overall response of the primary mass is reduced especially when $\mu = 0.1$. Finally, the last set of plots shows the results for the Northridge earthquake. It can be observed that the high value of μ provides worse result compared to the low value of the mass ratio. These simulations clearly indicate the problems mentioned above for conventional TMDs, namely, their inability to perform well in the transient interval.

Summary

The entry revisits the development of the classical theory of passive tuned mass dampers with a primary mass subjected to either deterministic or stochastic excitations. Various formulas for optimizing the performance of the damper as well as opposite points of view on the efficiency of dampers to mitigate the seismic vibrations are presented. Pendulum tuned mass dampers are presented and their characteristics are outlined.



Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations, Fig. 14 Earthquake records (left column) and the primary mass response (right

column) (b) for $\zeta_1 = \zeta_2 = 0.01$, $\Omega = 2.0$, and (a), (b) El Centro, (c), (d) Nisqually, and (e), (f) Northridge

The optimal values of the linear damping ratio of the PPD are expressed as a function of the mass ratio. It has been observed that the damping coefficient for a PPD decreases with the increase of

the mass ratio, which is opposite to what is observed for a conventional TMD. Due to the features inherent to a PPD system, it can be stated that the potential of PPDs has not been fully

investigated and comparison of their performance to a conventional TMD has not been made yet.

Cross-References

- [Stochastic Analysis of Linear Systems](#)
- [Stochastic Analysis of Nonlinear Systems](#)
- [Tuned Mass Dampers for Passive Control of Structures Under Earthquake Excitations](#)

References

- Adam C, Heuer R, Pirrotta A (2003) Experimental dynamic analysis of elastic-plastic shear frames with secondary structures. *Exper Mech* 43(2):124–130
- Bergman L, McFarland D, Hall J, Johnson E, Kareem A (1989) Optimal distribution of tuned mass dampers in wind-sensitive structures. In: *Structural safety and reliability*, ASCE, pp 95–102
- Bisegna P, Caruso G (2012) Closed-form formulas for the optimal pole-based design of tuned mass dampers. *J Sound Vib* 331(10):2291–2314
- Bobryk R, Chrzesczyk A (2009) Stability regions for mathieu equation with imperfect periodicity. *Phys Lett A* 373(39):3532–3535
- Brock JE (1946) A note on the damped vibration absorber. *J Appl Mech*, ASME 13, A–284
- Carpineto N, Lacarbonara W, Vestroni F (2014) Hysteretic tuned mass dampers for structural vibration mitigation. *J Sound Vib* 333(5):1302–1318
- Cartmell M, Cartmell M (1990) *Introduction to linear, parametric and nonlinear vibrations*. Chapman and Hall, London
- Cho DS, Mo CK, Ban GS, Lee KH (2003) Modal interactions in an autoparametric vibration absorber to narrow band random excitation. *KSME Int J* 17(1):97–104
- Chowdhury AH, Iwuchukwu MD, Garske JJ (1987) The past and future of seismic effectiveness of tuned mass dampers, *structural control* (Leipholz, H.H.E., ed.). Martinus Nijhoff Publishers, Dordrecht 105–127
- Chung L, Wu L, Lien K, Chen H, Huang H (2013) Optimal design of friction pendulum tuned mass damper with varying friction coefficient. *Struct Control Health Monitor* 20(4):544–559
- Crandall SH, Mark WD (1963) *Random vibration in mechanical systems*, vol 963. Academic, New York
- Debbarna R, Hazari S (2013) Mass distribution of multiple tuned mass dampers for vibration control of structures under earthquake load. *Int J Emerg Technol Adv Eng* 3(8):198–202
- Den Hartog JP (2013) *Mechanical vibrations*. Courier Dover Publications, New York
- Di Matteo A, Lo Iacono F, Navarra G, Pirrotta A (2014) Direct evaluation of the equivalent linear damping for tuned liquid column damper systems in random vibration for pre-design purpose. *Int J Non-Linear Mech* 63:19–30
- Dimentberg M, Iourtchenko D (2004) Random vibrations with impacts: a review. *Nonlinear Dynam* 36(2–4):229–254
- Elgamal A, He L (2004) Vertical earthquake ground motion records: an overview. *J Earthquake Eng* 8(5):663–697
- Hunt JB (1979) *Dynamic vibration absorbers*. Mechanical Engineering Publications, London
- Ibrahim R (2008) Recent advances in nonlinear passive vibration isolators. *J Sound Vib* 314(3):371–452
- Ibrahim RA (2009) *Vibro-impact dynamics: modeling, mapping and applications*, vol 43. Springer, Berlin
- Ibrahim R, Roberts J (1976) Broad band random excitation of a two-degree-of-freedom system with autoparametric coupling. *J Sound Vib* 44(3):335–348
- Ikeda T (2011) Nonlinear responses of dual-pendulum dynamic absorbers. *J Comput Nonlinear Dyn* 6(1):011012
- Jang S-J, Brennan M, Rustighi E, Jung H-J (2012) A simple method for choosing the parameters of a two degree-of-freedom tuned vibration absorber. *J Sound Vib* 331(21):4658–4667
- Kareem A, Kline S (1995) Performance of multiple mass dampers under random loading. *J Struct Eng* 121(2):348–361
- Kareem A, Kijewski T, Tamura Y (1999) Mitigation of motions of tall buildings with specific examples of recent applications. *Wind Struct* 2(3):201–251
- Kaynia AM, Biggs JM, Veneziano D (1981) Seismic effectiveness of tuned mass dampers. *J Struct Div* 107(8):1465–1484
- Krenk S (2005) Frequency analysis of the tuned mass damper. *J Appl Mech* 72(6):936–942
- Krenk S, Høgsberg J (2008) Tuned mass absorbers on damped structures under random load. *Probab Eng Mech* 23(4):408–415
- Krenk S, Høgsberg J (2014) Tuned mass absorber on a flexible structure. *J Sound Vib* 333(6):1577–1595
- Majcher K, Wójcicki Z (2014) Kinematically excited parametric vibration of a tall building model with a tmd. Part 1: numerical analyses. *Arch Civil Mech Eng* 14(1):204–217
- Ormondroyd J (1928) Theory of the dynamic vibration absorber. *Trans ASME* 50:9–22
- Rüdinger F (2007) Tuned mass damper with nonlinear viscous damping. *J Sound Vib* 300(3):932–948
- Sadek F, Mohraz B, Taylor AW, Chung RM (1997) A method of estimating the parameters of tuned mass dampers for seismic applications. *Earthquake Eng Struct Dynam* 26(6):617–636
- Sladek JR, Klingner RE (1983) Effect of tuned-mass dampers on seismic response. *J Struct Eng* 109(8):2004–2009

- Song Y, Sato H, Iwata Y, Komatsuzaki T (2003) The response of a dynamic vibration absorber system with a parametrically excited pendulum. *J Sound Vib* 259(4):747–759
- Soong T, Spencer B Jr (2002) Supplemental energy dissipation: state-of-the-art and state-of-the-practice. *Eng Struct* 24(3):243–259
- Soto MG, Adeli H (2013) Tuned mass dampers. *Arch Comput Meth Eng* 20(4):419–431
- Tigli OF (2012) Optimum vibration absorber (tuned mass damper) design for linear damped systems subjected to random loads. *J Sound Vib* 331(13):3035–3049
- Tondl A, Ruijgrok T, Verhulst F, Nabergoj R (2000) *Autoparametric resonance in mechanical systems*. Cambridge University Press, Cambridge
- Tsai H-C, Lin G-C (1993) Optimum tuned-mass dampers for minimizing steady-state response of support-excited and damped systems. *Earthquake Eng Struct Dyn* 22(11):957–973
- Vakakis A, Paipetis S (1986) The effect of a viscously damped dynamic absorber on a linear multi-degree-of-freedom system. *J Sound Vib* 105(1):49–60
- Val DV, Segal F (2005) Effect of damping model on pre-yielding earthquake response of structures. *Eng Struct* 27(14):1968–1980
- Wang A-P, Lin Y-H (2007) Vibration control of a tall building subjected to earthquake excitation. *J Sound Vib* 299(4):757–773
- Warburton G (1981) Optimum absorber parameters for minimizing vibration response. *Earthquake Eng Struct Dyn* 9(3):251–262
- Warburton G (1982) Optimum absorber parameters for various combinations of response and excitation parameters. *Earthquake Eng Struct Dyn* 10(3):381–401
- Warminski J, Kecik K (2009) Instabilities in the main parametric resonance area of a mechanical system with a pendulum. *J Sound Vib* 322(3):612–628
- Wiesner KB (1979) Tuned mass dampers to reduce building wind motion. In: *ASCE convention and exposition*, vol 3510. ASCE, Boston
- Wirsching PH, Campbell GW (1973) Minimal structural response under random excitation using the vibration absorber. *Earthquake Eng Struct Dyn* 2(4):303–312
- Xu X, Wiercigroch M (2007) Approximate analytical solutions for oscillatory and rotational motion of a parametric pendulum. *Nonlinear Dyn* 47(1–3):311–320
- Yurchenko D, Alevras P (2013a) Dynamics of the n-pendulum and its application to a wave energy converter concept. *Int J Dyn Control* 1(4):290–299
- Yurchenko D, Alevras P (2013b) Stochastic dynamics of a parametrically base excited rotating pendulum. In: *Procedia IUTAM* 6, pp 160–168
- Yurchenko D, Naess A, Alevras P (2013) Pendulum's rotational motion governed by a stochastic mathieu equation. *Probabilist Eng Mech* 31:12–18
- Zuo L, Nayfeh SA (2006) The two-degree-of-freedom tuned-mass damper for suppression of single-mode vibration under random and harmonic excitation. *J Vib Acoust* 128(1):56–65

Tuned Mass Dampers for Passive Control of Structures Under Earthquake Excitations

Rafael H. Lopez¹, Leandro F. F. Miguel¹ and André T. Beck²

¹Civil Engineering Department, Federal University of Santa Catarina, Florianópolis, SC, Brazil

²Department of Structural Engineering, São Carlos School of Engineering, University of São Paulo, São Carlos, SP, Brazil

Synonyms

Optimal design; Optimization under uncertainties; Robust optimization; Seismic loading; TMD

Introduction

The importance of reducing vibration amplitudes of structures under rare but severe loadings is well recognized. In the last decades, a fast increase in the development and application of passive energy dissipation devices, such as viscoelastic dampers, viscous fluid dampers, metallic yield dampers, friction dampers, and tuned mass dampers (TMDs), has been observed (Soong and Dargush 1997). One of the main applications of such devices is to reduce the dynamic response of structures subjected to strong earthquake ground motions.

To achieve the best structural performance, attention has been focused on determining the optimum parameters of TMDs in structures subjected to seismic excitation. Sadek et al. (1997) presented a methodology to find optimum parameters of TMDs for single degree of freedom (SDOF) and multiple degree of freedom (MDOF) structures subjected to a number of earthquake excitations. Joshi and Jangid (1997) carried out a study to determine the optimum parameters of MTMDs in SDOF structures

subjected to a base excitation, which was modeled as a stationary white noise random process. Results showed that the optimally designed MTMD system is more effective than the single TMD. Hadi and Arfiadi (1998) discussed the optimum design of single TMDs in MDOF structures subjected to earthquake excitation. The H2 norm was employed as the objective function, and genetic algorithms were applied to find the best TMD parameters. Chen and Wu (2001) presented a study focused on the optimal placement of MTMDs on structures subjected to seismic excitations. Because oscillators are placed one by one in sequence, their placement in a structure is selected from a restricted domain. Thus, the final results are indeed suboptimal. Hoang and Warnitchai (2005) presented a method that uses a numerical optimizer following a gradient-based nonlinear programming algorithm to search for optimal parameters of MTMDs. The developed method was applied to optimize MTMDs for a SDOF structure subjected to wideband excitation. Lee et al. (2006) presented an optimal design theory for structures implemented with TMDs. The proposed optimal design theory feasibility was verified by using a SDOF structure with a single TMD, a five-DOF structure with two TMDs, and a ten-DOF structure with a single TMD. Leung et al. (2008) employed the particle swarm optimization algorithm to obtain the optimum parameters including the optimum mass ratio, damper damping, and tuning frequency of the TMD system attached to a viscously damped single-degree-of-freedom main system, which is subjected to nonstationary excitation. In the sequence, Leung and Zhang (2009) employed again the particle swarm optimization algorithm to obtain the optimum parameters of a TMD system attached to a viscously damped SDOF main system under various combinations of different kinds of excitations. Lin et al. (2010) applied a two-stage optimum design procedure for MTMD to reduce structural dynamic responses with the limitation of MTMD's stroke. Shaking table tests of a large-scale three-story building with and without the MTMD under earthquake excitations were

conducted. Mohebbi et al. (2013) proposed a procedure for designing optimal MTMDs to mitigate the seismic response of MDOF structures. Genetic algorithms were used for solving the optimization problem.

Despite the importance of the studies mentioned in the previous entry, it is today widely acknowledged that deterministic optimization is not robust with respect to the uncertainties which affect the performance of engineering facilities. Deterministic design optimization (DDO) does not explicitly address the uncertainties inherently present in structural, especially environmental, loads, material strengths, and load effect and member strength models. Reliability-based design optimization (RBDO) has emerged as an alternative to model the safety-under-uncertainty part of the problem. However, both deterministic and reliability-based design optimizations suffer from the same drawback: these methods allow optimum structures to be designed but by considering conservative safety margins with respect to failure modes. Such methods are not appropriate for the optimal design of control structures such as TMD and MTMD. DDO is not appropriate because it is not robust with respect to the uncertainties. RBDO is not appropriate because control needs to be designed at the precise point (design point), and not within a conservative safety margin with respect to the design point. In fact, one can say that control is not control unless it is robust with respect to the uncertainties described above. Specifically, structural control should be made robust with respect to the uncertainties in the loading, mass, damping, and stiffness and in the inherent variability of materials properties, imperfect modeling of loads, load effects, system response, and so on (Taflanidis et al. 2007, 2009; Marano et al. 2010; Taflanidis 2011; Jensen 2005; Jensen et al. 2008; Jensen and Sepulveda 2011; Mohtat and Dehghan-Niri 2011). Thus, in the design and optimization of TMD passive control systems, it is of paramount importance to take uncertainties into account.

Within this context, this entry deals with the robust optimal design of TMD systems for structures subjected to seismic loading.

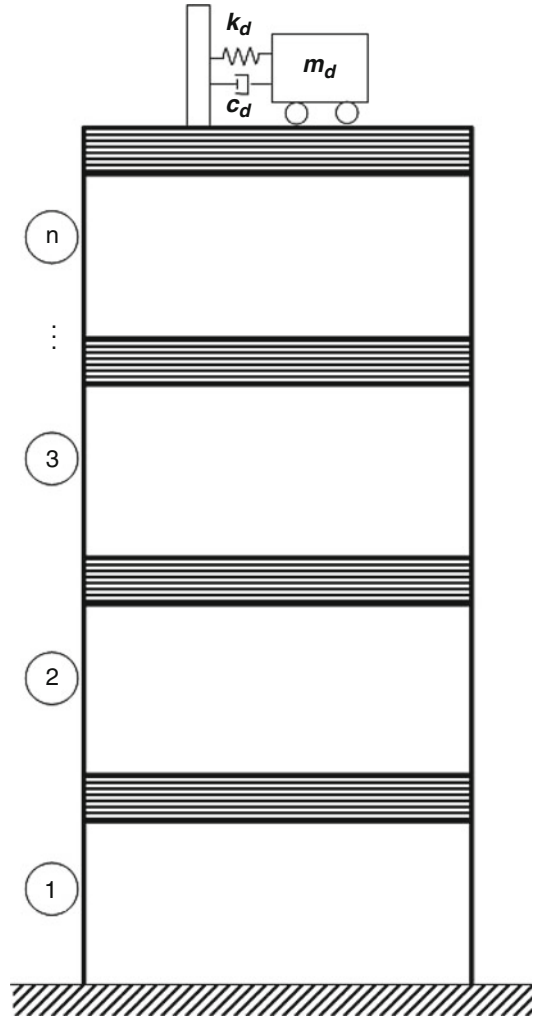
In section [Structural Model](#) the structural model is presented. Time-variant reliability analysis of oscillators is presented in section [Time-Variant Reliability of Oscillators](#). The robust optimization problem and a robustness measures are introduced in section [Robust Optimization Problem](#). Numerical examples are presented in section [Numerical Examples](#), in order to illustrate the importance of taking into account uncertainties in the optimum design of TMD systems. For the sake of simplicity and to focus on the formulation of the robust design optimization problem, the numerical analysis section deals with one TMD only. It should be remarked, however, that the extension to MTMD is straightforward. The chapter is finished with a Summary in section [Summary](#).

Structural Model

The differential equation that governs the motion of multi-degree-of-freedom systems, with *one* TMD located on the top floor (Fig. 1) and subjected to earthquake ground motions, may be written as

$$\mathbf{M}\ddot{\mathbf{x}}(t) + \mathbf{C}\dot{\mathbf{x}}(t) + \mathbf{K}\mathbf{x}(t) = -\mathbf{M}\mathbf{B}\ddot{\mathbf{y}}(t) \quad (1)$$

in which $\mathbf{M}$, $\mathbf{C}$, and $\mathbf{K}$ represent the $(n + 1) \times (n + 1)$ structural mass, inherent damping, and stiffness matrices, respectively, and n is the number of degrees of freedom of the model. $\mathbf{x}(t)$ is the $(n + 1)$ -dimensional relative displacement vector with respect to the base, and a dot over a symbol indicates differentiation with respect to time. $\mathbf{B}$ is a $(n + 1) \times d$ matrix of ground motion influence coefficients, i.e., this matrix contains the cosine directors of the angles formed between the base motion and the associated displacement direction with the considered degree of freedom. d is the number of considered ground motions (directions). $\ddot{\mathbf{y}}(t)$ is a d -dimensional vector representing the seismic excitation, i.e., the ground motion or base acceleration. For design purposes, assume that all relevant structural parameters (e.g., nominal values of mass, damping, and stiffness) can be



Tuned Mass Dampers for Passive Control of Structures Under Earthquake Excitations, Fig. 1 n -DOF building with single TMD on the *top* floor

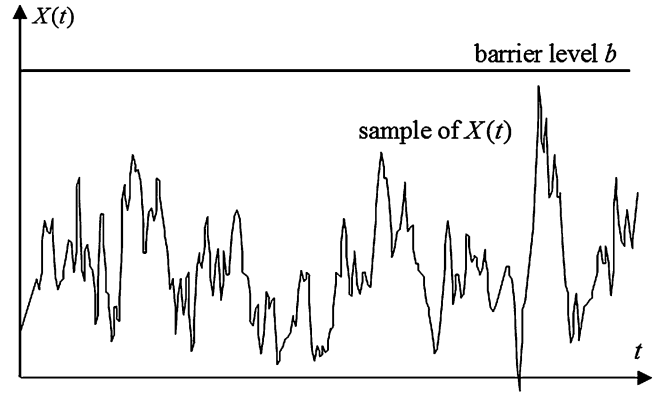
assembled in a parameter vector $\mathbf{p}$. In the TMD design optimization process, the design variables are considered to be its stiffness and damping, which are then grouped in the design vector $\mathbf{d}$ (further details are given in section [Robust Optimization Problem](#)).

Time-Variant Reliability of Oscillators

The classical reliability problem for an oscillator is depicted in Fig. 2. During an excitation event of

Tuned Mass Dampers for Passive Control of Structures Under Earthquake Excitations,

Fig. 2 Classical time-variant reliability problem



specified duration t_E , a specific, scalar displacement response of the oscillator $x(t)$ (e.g., relative displacement between floors, displacement of top floor, etc.) should not exceed $\pm b$, where b is a given barrier which corresponds to a critical response level. In this entry, the displacement of top floor is chosen as the relevant displacement response. The critical response may represent permanent damage like cracking of concrete, or it may represent ultimate failure due to loss of equilibrium and is further addressed later in this entry. For a given barrier level b and excitation duration t_E , the failure probability is given by the classical Poisson model (Melchers 1999):

$$P_f(b, t_E) = 1 - \exp\left(-2 \int_0^{t_E} v_x^+(b, t) dt\right) \quad (2)$$

where v_x^+ is the up-crossing rate. For stationary excitation and for a time-invariant barrier $b(t) = b$, the crossing rate is also constant and $\int_0^{t_E} v_x^+(b, t) dt = v_x^+(b) t_E$. For a linear system excited by a Gaussian process, the response is Gaussian and the crossing rate can be evaluated as:

$$v_x^+(b) = \frac{\sigma_{\dot{x}}}{\sigma_x} \frac{1}{2\pi} \exp\left(-\frac{b^2}{2\sigma_x^2}\right) \quad (3)$$

where σ_x and $\sigma_{\dot{x}}$ are the standard deviation of the displacement and of the velocity response, respectively.

Many types of environmental structural loading can be described by the arrival of an unknown number of events (winds, storms, sea waves, earthquakes). By modeling the arrival of events as a Poisson process, and for a design life t_D , the failure probability becomes

$$P_f(b, t_D) = \sum_{i=1}^{\infty} P_f(b, t_E|i) p_i(t_D) \quad (4)$$

where $P_f(b, t_E|i)$ is the conditional probability of failure, given the occurrence of exactly i events during the design life, and $p_i(t_D)$ is the probability of having exactly i events, given by the Poisson distribution:

$$p_i(t_D) = \frac{(vt_D)^i \exp(-vt_D)}{i!} \quad (5)$$

In Eq. 5, v is the arrival rate of events. It can be shown that the conditional failure probability, assuming independence, is given by (Melchers 1999):

$$P_f(b, t_E|i) = 1 - (1 - P_f(b, t_E))^i \quad (6)$$

For convenience, failure probabilities can be converted into reliability index (β) by means of the classical first-order reliability result:

$$\beta = -\Phi^{-1}(1 - P_f(b, t_D)) \quad (7)$$

where Φ is the standard Gaussian cumulative distribution function. Clearly, the reliability index is a function of the vector of design parameters as well

as the vector of structural parameters $\mathbf{p}$. Hence, once can write $\beta = \beta(\mathbf{d}, \mathbf{p})$.

Robust Optimization Problem

This entry aims at pursuing the TMD design optimization of a given structure subjected to seismic loading, taking into account the uncertainties present in the structural system and its loading. That is, an approach for the robust design optimization of TMD is presented. In order to take into account these uncertainties, some parameters of the structure are modeled as random variables and the seismic loading is modeled as a random process.

One classical objective function in the deterministic design of TMD is the minimization of some mean square response of the structure. However, it has been reported in the literature that such an objective function does not necessarily correspond to the optimal design in terms of reliability (Papadimitriou et al. 1997; Scruggs et al. 2006). Hence, the probability of failure of the structure is chosen as the objective function to be minimized in this optimization problem. Such an objective function has been adopted, for instance, in the works by Taflanidis et al. (2007) and Taflanidis (2011), to name just a few. This probability of failure is given by Eq. 4, and the design variables are the stiffness (k_d) and damping constant (c_d) of the TMD. These two design variables are grouped into the vector $\mathbf{d} = [k_d \ c_d]$ for notation convenience. As commented in the previous section, the reliability index is chosen as the reliability measure. Since the higher its value is, the lower the probability of failure is, the objective function is given by $-\beta(\mathbf{d}, \mathbf{p})$. Thus, the optimization problem can be posed as:

$$\begin{aligned} &\text{Find : } \mathbf{d} \\ &\text{which minimizes : } -\beta(\mathbf{d}, \mathbf{p}) \\ &\text{subject to : } d_i^{\min} \leq d_i \leq d_i^{\max}, \quad j = 1, 2 \end{aligned} \quad (8)$$

This formulation leads to an unconstrained optimization problem, where the feasible design set is limited only by the bounds of the design variables. In this context, several optimization

algorithms can be used to solve this problem, e.g., gradient-based algorithm, Nelder–Mead algorithm, among others. In the numerical analysis section, a sequential quadratic programming (SQP) algorithm, coupled with a restart procedure, is employed in order to ensure the convergence to a global optimum. This coupling has been tested in several papers and has successfully solved complex optimization problems (Luersen and Riche 2004; Ritto et al. 2011; Lopez et al. 2013). For the interested reader, a detailed explanation of the optimization algorithm is given in these references.

Optimal TMD parameters found by the above formulation are intrinsically more robust than TMD parameters found by any equivalent deterministic formulation. However, it is important to demonstrate that indeed results obtained by the robust optimization are less sensitive to changes in the estimated structural parameters $\mathbf{p}$: mass, stiffness, and damping of the structural elements and of the TMD. Hence, a perturbation is introduced in vector $\mathbf{p}_\Delta$, such as to verify what would be the actual reliability index of the structure if parameter values were not the nominal values used in designing the TMD. Then, the perturbed reliability index β_Δ can be compared with the original, and a measure of robustness is obtained: the smaller the relative difference, the more robust is the design:

$$\text{robustness} \sim \frac{\beta}{|\beta - \beta_\Delta|} \quad (9)$$

The robustness measure proposed in Eq. 9 is employed in the next section to demonstrate that results of the optimal design of TMD considering uncertainties in structural and TMD parameters are more robust than optimal TMD designs obtained by considering only uncertain loading.

Numerical Examples

In this section, an example optimization problem is solved: robust design of TMD of a structure subjected to seismic excitation.

Problem Definition

A classical example, previously studied by several researchers (Hadi and Arfiadi 1998; Lee et al 2006; Mohebbi et al. 2013), was selected from the literature. The structure is a ten-story shear frame as illustrated in Fig. 1. In the top floor of this structure, a TMD is to be installed. The mass of the TMD (m_d) is chosen a priori to be 3 % of the total mass of the structure. The parameters to be determined by the designer are mean values of the stiffness (k_d) and damping coefficient (c_d) of such a TMD. Two variants of the example are solved: (a) only the seismic excitation is considered random; (b) the seismic excitation is random, but uncertainties are also considered in the mass, stiffness, and damping coefficient of the structure and the TMD. These are then modeled as random variables whose characteristics are presented in Table 1. Table 1 also shows the perturbations considered in order to evaluate the robustness of the solutions, following Eq. 9. These perturbations are applied to the random variables of the structure. The critical combination of perturbations is given by a reduction of stiffness and damping and an increase of the mass (this explains the signs on the perturbations presented in Table 1).

To determine the optimum TMD parameters, a stationary earthquake excitation is assumed, which can be modeled as a white noise signal with constant spectral density, S_0 , filtered through the Kanai–Tajimi spectrum. The PSD function is given by:

$$s(\omega) = S_0 \left[\frac{\omega_g^4 + 4\omega_g^2 \zeta_g^2 \omega^2}{(\omega^2 - \omega_g^2)^2 + 4\omega_g^2 \zeta_g^2 \omega^2} \right], \quad (10)$$

$$S_0 = \frac{0.03 \zeta_g}{\pi \omega_g (4 \zeta_g^2 + 1)}$$

where ζ_g and ω_g are the ground damping and frequency, respectively. In all the situations analyzed in this paper, the filter parameters are modeled as random variables. Their mean values were adopted as $\zeta_g = 0.6$, $\omega_g = 37.3$ rad/s (Mohebbi et al. 2013) and peak ground acceleration (PGA) = 0.20 g. These random variables are considered independent; Gaussian and their coefficients of variation were set to 20 %. Similar procedures were adopted in Taflanidis et al. (2007). It should be remarked that for more realistic situations, a nonstationary excitation should be employed. This procedure is detailed, for instance, in Jensen and Sepulveda (2011).

The upper bound and the lower bound value of the stiffness (k_d) and damping coefficient (c_d) of the TMD are 0–4,000 kN/m and 0–1,000 kNs/m, respectively. In all cases, the Newmark method ($\gamma = 1/2$ and $\beta = 1/4$) was employed to solve the dynamic problem. The time step adopted was $\Delta t = 0.02$ s. The mechanical model was calibrated and it is in accordance with the results provided in the literature.

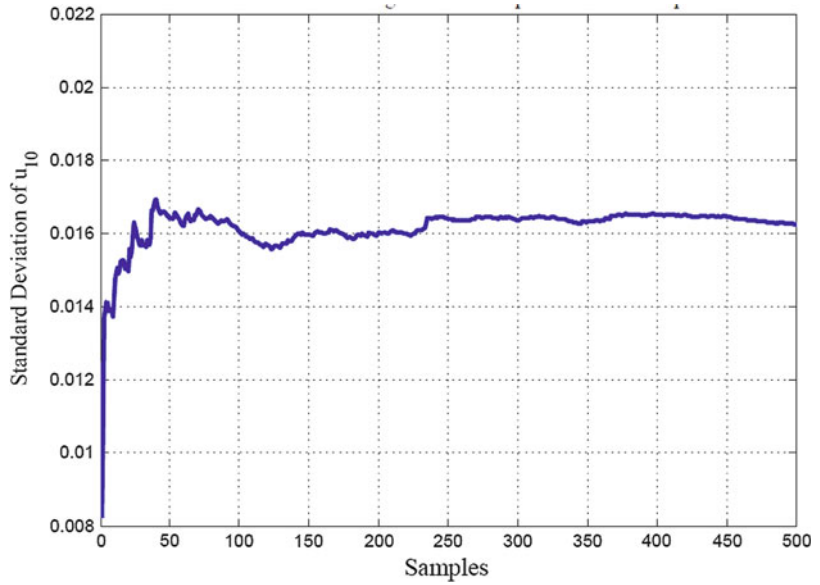
Convergence of σ_x and $\sigma_{\dot{x}}$

Since MCS is employed to evaluate the standard deviations σ_x and $\sigma_{\dot{x}}$ required in Eq. 3 to approximate the probability of failure of the structure, a proper sample size needs to be selected. Thus, a study of the convergence of these standard deviations is shown in this section. Results are shown in Figs. 3 and 4. One may see from these figures that around 250 samples are required to stabilize the convergence curves. Thus, a sample size of 250 is employed herein to evaluate σ_x and $\sigma_{\dot{x}}$.

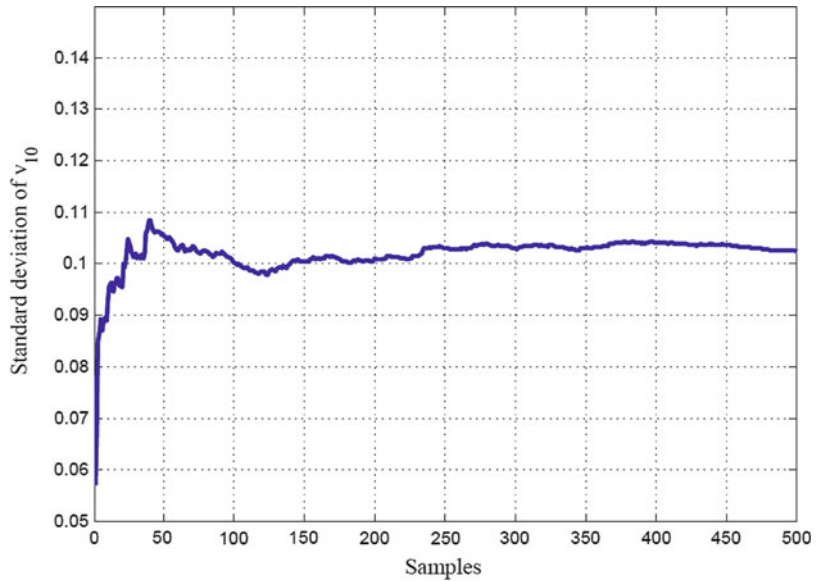
Tuned Mass Dampers for Passive Control of Structures Under Earthquake Excitations, Table 1 Statistical information about the random variables of the structure

		Mean value	C.o.V. [%]	Perturbations considered in robustness analysis
Stiffness (k_i) [N/m]	Story (i)	650.0×10^6	15.0	–5 %
	TMD	k_d	15.0	–
Mass (m_i) [kg]	Story (i)	360.0×10^3	10.0	+3 %
	TMD	108.0×10^3	10.0	–
Damping (c_i) [Ns/m]	Story (i)	620×10^6	25.0	–8 %
	TMD	c_d	25.0	–

Tuned Mass Dampers for Passive Control of Structures Under Earthquake Excitations, Fig. 3 Convergence of the standard deviation of top floor displacement (σ_x)



Tuned Mass Dampers for Passive Control of Structures Under Earthquake Excitations, Fig. 4 Convergence of the standard deviation of the top floor velocity ($\sigma_{\dot{x}}$)



Minimization of the Probability of Failure

The optimization problem presented in sections [Robust Optimization Problem](#) and [Problem Definition](#) is solved in this section. An SQP algorithm from the MATLAB optimization toolbox, coupled with a restart procedure, is employed for this purpose. Three variants of each analysis are presented, by varying parameter b in Eq. 2: (i) $b = h/300$, (i) $b = h/400$, and (iii) $b = h/500$, where h is the total height of the shear

frame. Results of the analyses considering only excitation uncertainty are shown in Table 2. Results of the robust optimal design, i.e., considering uncertainty in excitation and structural parameters, are presented in Table 3. It should be remarked that the reliability indexes presented in these tables were evaluated considering uncertainty in the excitation and in the structural parameters, in order to allow for comparisons between the different analyses. For instance,

Tuned Mass Dampers for Passive Control of Structures Under Earthquake Excitations, Table 2 Results of the TMD optimization considering only excitation uncertainty

Case	k_d (MN/m)	c_d (MNs/m)	$\beta_{\text{uncontrolled}}$	β	β_{Δ}	Robustness
(i) $b = h/300$	3.69	0.146	2.23	5.23	4.51	7.26
(ii) $b = h/400$	3.69	0.146	0.48	2.95	2.26	4.28
(iii) $b = h/500$	3.69	0.147	<i>fail</i>	1.09	0.25	1.30

Tuned Mass Dampers for Passive Control of Structures Under Earthquake Excitations, Table 3 Results of the robust TMD optimization

Case	k_d (MN/m)	c_d (MNs/m)	$\beta_{\text{uncontrolled}}$	β	β_{Δ}	Robustness
(i) $b = h/300$	3.29	0.161	2.23	5.33	4.72	8.74
(ii) $b = h/400$	3.29	0.161	0.48	3.05	2.46	5.17
(iii) $b = h/500$	3.29	0.161	<i>fail</i>	1.20	0.50	1.71

optimal values of k_d and c_d in Table 2 were obtained solving the optimization problem given by Eq. 8 and considering uncertainty in the excitation only; whereas reliability indexes shown in this table were evaluated also considering uncertainties in structural parameters in order to allow for direct comparison of the reliability indexes.

In Tables 2 and 3 one observes that reliability index of uncontrolled structures ($\beta_{\text{uncontrolled}}$) is significantly smaller than the reliability index of the TMD-controlled structures (β). This result shows the effectiveness of the TMD in reducing the probabilities of failure of the frame structures under seismic loading.

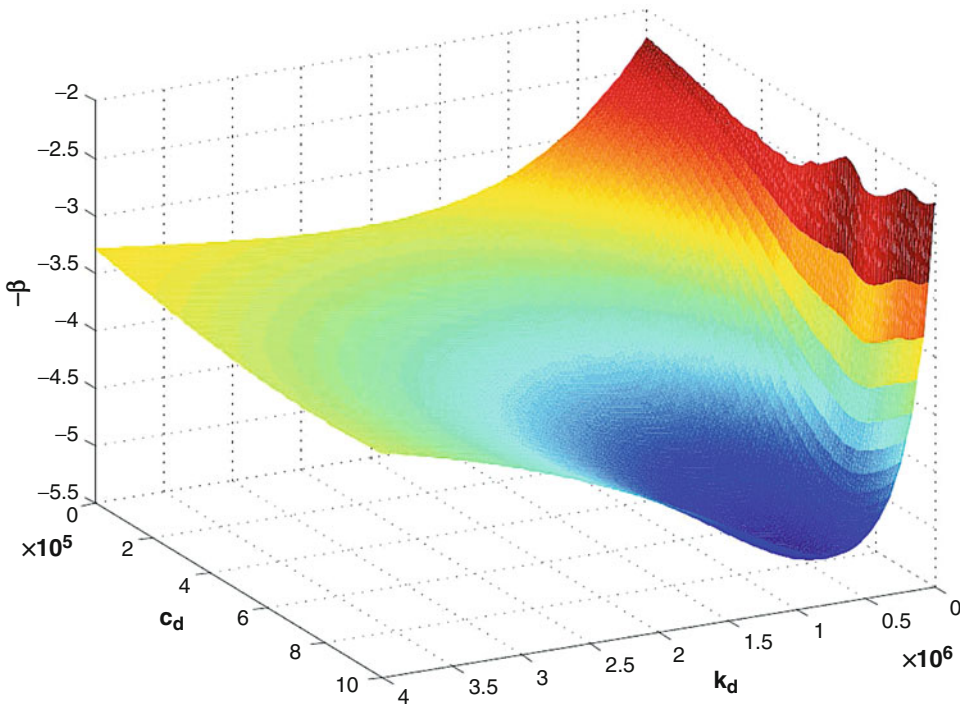
The reliability indexes of the TMD-controlled structures (β), obtained by considering uncertainty in the excitation only (Table 2), are smaller than the β 's obtained by the robust optimization (Table 3). This result is a first indication that the optimization discussed herein, considering uncertainties in structural parameters, leads to more robust control by the TMD. To further demonstrate this result, the robustness measure introduced in Eq. 9, and evaluated for the perturbations presented in Table 1, is also shown in Tables 2 and 3. One can see that the robust design discussed herein, considering uncertainties in structural parameters, leads to larger values of the robustness measure. Hence, structural control obtained by the robust optimal design of TMD becomes less sensitive to fluctuations or uncertainties in structural and TMD parameters.

The optimization algorithm converged to the same optimum design independently of the starting point chosen by the restart procedure. An approximation of the objective function of case (iii), for the more general case considering uncertain load and structural parameters, is given in Fig. 5. One can see from this figure that the objective function is fairly smooth and without local minima.

Limitations of the Models

The goal of this entry was to present an introduction to the robust optimal design of TMD systems for structures subjected to seismic excitation. Hence, a simple model was presented herein. The reader should be aware of its main limitations, which are:

- Approximation of the crossing rates by Eq. 3, which assumes a stationary Gaussian excitation and linear, hence also Gaussian response; a more elaborate excitation model could involve non-Gaussian excitations and nonlinear structural responses, hence non-Gaussian displacement responses.
- The use of a stationary process to model the earthquake excitation: seismic excitation is nonstationary in nature; hence, a more elaborated model can be employed in real applications. This procedure is detailed, for instance, in Jensen and Sepúlveda (2011).



Tuned Mass Dampers for Passive Control of Structures Under Earthquake Excitations, Fig. 5 Objective function in the design space for the robust optimal design of a TMD (case (iii))

- The use of a linear structural model: the reader must be aware that nonlinear responses are expected for earthquake excitations.

Summary

This entry dealt with the robust optimum design of TMD systems for structures subject to seismic loading. First, the basic concepts of uncertainty quantification were presented focusing on the evaluation of the probability of failure of a given structure subject to seismic loading. It was detailed that the computation of the standard deviation of the stochastic process that define the displacement and velocity of a given degree of freedom of the structure and the definition of a threshold level for it were necessary to evaluate this probability of failure. Here, the MCS was employed to obtain these standard deviations; however, it may be remarked that other methods may be employed. Then, the robust design of the

TMD was posed having as the objective function the minimization of the structural probability of failure. The design variables were chosen to be the stiffness and the damping of the TMD. It led to an optimization problem in which the only constraints were the bounds of the design variables. An SQP-based algorithm, coupled with a restart procedure, was suggested as the optimizer to ensure the convergence to a global optimum. Then, a numerical example was presented to illustrate the importance of taking into account uncertainties in the optimum design of TMD systems for structures subject to seismic loading. This was achieved by introducing perturbations, which create a distance between parameter values considered in the design and parameter values “observed” in the actual structure. From the results of the numerical analysis section, it was shown that the structural control obtained by the robust optimal design of TMD becomes less sensitive to fluctuations or uncertainties in structural parameters.

Cross-References

- [Actively and Semi-Actively Controlled Structures Under Seismic Actions: Modeling and Analysis](#)
- [Friction Dampers for Seismic Protections of Steel Buildings Subjected to Earthquakes: Emphasis on Structural Design](#)
- [Passive Control Techniques for Retrofitting of Existing Structures](#)
- [Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach](#)
- [Robust Control of Building Structures under Uncertain Conditions](#)
- [Robust Design Optimization for Earthquake Loads](#)
- [Structural Optimization Under Random Dynamic Seismic Excitation](#)
- [Structures with Nonviscous Damping, Modeling, and Analysis](#)

References

- Chen G, Wu J (2001) Optimal placement of multiple tune mass dampers for seismic structures. *J Struct Eng* 127:1054–1062
- Hadi MNS, Arfiadi Y (1998) Optimum design of absorber for MDOF structures. *J Struct Eng ASCE* 124:1272–1280
- Hoang N, Warnitchai P (2005) Design of multiple tuned mass dampers by using a numerical optimizer. *Earthq Eng Struct Dyn* 34:125–144
- Jensen HA (2005) Structural optimization of linear dynamical systems under stochastic excitation: a moving reliability database approach. *Comput Meth Appl Mech Eng* 194:1757–1778
- Jensen HA, Sepulveda JG (2011) Structural optimization of uncertain dynamical systems considering mixed-design variables. *Probab Eng Mech* 26:269–280
- Jensen HA, Valdebenito MA, Schueller GI (2008) An efficient reliability-based optimization scheme for uncertain linear systems subject to general Gaussian excitation. *Comput Meth Appl Mech Eng* 198:72–87
- Joshi AS, Jangid RS (1997) Optimum parameters of multiple tuned mass dampers for base-excited damped. *J Sound Vib* 202(5):657–667
- Lee CL, Chen YT, Chun LL, Wang YP (2006) Optimal design theories and applications of tuned mass dampers. *Eng Struct* 28:43–53
- Leung AYT, Zhang Z (2009) Particle swarm optimization of tuned mass dampers. *Eng Struct* 31:715–728
- Leung AYT, Zhang H, Cheng CC, Lee YY (2008) Particle swarm optimization of TMD by non-stationary base excitation during earthquake. *Earthq Eng Struct Dyn* 37:1223–1246
- Lin CC, Wang JF, Lien CH, Chiang HW, Lin CS (2010) Optimum design and experimental study of multiple tuned mass dampers with limited stroke. *Earthq Eng Struct Dyn* 39:1631–1651
- Lopez RH, Ritto TG, Sampaio R, Souza de Cursi JE (2013) A new algorithm for the robust optimization of rotor-bearing systems. *Optim Eng*. doi:10.1080/0305215X.2013.819095
- Luersen MA, Riche R (2004) Globalized Nelder–Mead method for engineering optimization. *Comput Struct* 82(23–26):2251–2260
- Marano GC, Greco R, Sgobba S (2010) A comparison between different robust optimum design approaches: application to tuned mass dampers. *Probab Eng Mech* 25:108–118
- Melchers R (1999) Structural reliability analysis and prediction. Wiley, New York
- Mohebbi M, Shakeri K, Ghanbarpour Y, Majzoub H (2013) Designing optimal multiple tuned mass dampers using genetic algorithms (GAs) for mitigating the seismic response of structures. *J Vib Control* 19(4):605–625
- Mohtat A, Dehghan-Niri E (2011) Generalized framework for robust design of tuned mass damper systems. *J Sound Vib* 330:902–922
- Papadimitriou C, Katafygiotis LS, Au SK (1997) Effects of structural uncertainties on TMD design: a reliability based approach. *J Struct Control* 4(1):65–88
- Ritto TG, Lopez RH, Sampaio R, Souza de Cursi JE (2011) Robust optimization of a flexible rotor-bearing system using the Campbell diagram. *Eng Optim* 43(1):77–96
- Sadek F, Mohraz B, Taylor AW, Chung RM (1997) A method of estimating the parameters of tuned mass dampers for seismic applications. *Earthq Eng Struct Dyn* 26:617–635
- Scruggs JT, Taflanidis AA, Beck JL (2006) Reliability-based control optimization for active base isolation systems. *J Struct Control* 2006(13):705–723
- Soong TT, Dargush GF (1997) Passive energy dissipation systems in structural engineering. Wiley, Chichester
- Taflanidis A (2011) Optimal probabilistic design of seismic dampers for the protection of isolated bridges against near-fault seismic excitations. *Eng Struct* 33:3496–3508
- Taflanidis A, Beck JL, Angelides DC (2007) Robust reliability-based design of liquid column mass dampers under earthquake excitation using an analytical reliability approximation. *Eng Struct* 29:3525–3537
- Taflanidis A, Angelides DC, Scruggs JT (2009) - Simulation-based robust design of mass dampers for response mitigation of tension leg platforms. *Eng Struct* 31:847–857

Uncertainty Quantification in Structural Health Monitoring

Shankar Sankararaman¹ and
Sankaran Mahadevan²

¹Intelligent Systems Division, SGT Inc., NASA
Ames Research Center, Moffett Field, CA, USA

²Department of Civil and Environmental
Engineering, Vanderbilt University, Nashville,
TN, USA

Synonyms

Damage; Detection; Estimation; Localization;
Structural frames; Structural health monitoring;
Uncertainty

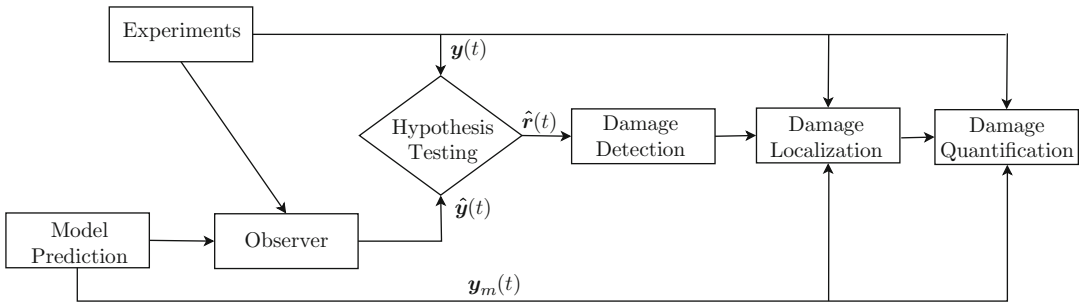
Introduction

Structural health monitoring (SHM) refers to the process of developing and implementing a damage identification strategy for structural applications in civil, aerospace, and mechanical domains (Farrar and Worden 2007). Structural health monitoring techniques play a vital role in earthquake engineering, because they can be used to monitor the performance of structures that have been subjected to earthquakes. This is established through the use of a combination of strategically placed array of sensors and a variety of mathematical methods and techniques, all of which

constitute the overall structural health monitoring system.

An ideal SHM system allows rapid assessment of structure, which is particularly relevant in the aftermath of an earthquake, and supports decisions regarding repairs and replacements that are necessary for maintenance, rehabilitation, and structural retrofitting. The most important functional aspect of such an SHM system is damage diagnosis. Damage diagnosis consists of three major steps: damage detection, damage localization (or isolation), and damage quantification. The first step of damage detection attempts to detect abnormalities in the structure and finds out if there has been any structural damage or deterioration. The second step of localization focuses on identifying that component(s) of the overall structure has been subjected to damage. The third step of damage quantification quantifies the extent of damage, for example, the reduction of stiffness in a structural member.

Several structural systems are complex in nature and consist of several members and components. Usually, only a few response quantities are monitored by the SHM system, but the number of prospective damage candidates is very high. In such cases, it is almost impossible to achieve precise damage diagnosis. Further, the inputs to the system, the physical parameters, the sensor measurements, etc. may all be uncertain and hence render the results of diagnosis uncertain. The quantification of uncertainty in damage diagnosis is an essential step to guide



Uncertainty Quantification in Structural Health Monitoring, Fig. 1 Model-based framework for structural health monitoring

decision-making regarding the functioning of the overall structural system.

In several research articles, the uncertainty in diagnosis (Achenbach 2000) has been previously addressed using nondestructive evaluation (NDE) techniques, by estimating the probability of detection (POD). In such POD calculations, nominally identical damage is introduced in a number of nominally identical specimens, and the number of successful detections is used to calculate the probability of detection. Such an approach is completely different from the goal in structural health monitoring where the focus is on one structural system that is being monitored. Further, it is not possible to perform repeated tests since the goal in SHM is to make use of monitoring data in order to diagnose damage that has been caused due to external sources.

This chapter discusses statistical methods based on classical statistics (Sankararaman and Mahadevan 2011) and Bayesian statistics (Sankararaman and Mahadevan 2013) for quantifying uncertainty in damage diagnosis in the context of structural health monitoring. For this purpose, a generic model-based framework, as shown in Fig. 1, is used.

In this framework, a mathematical model (which may be developed using first principles of physics or using data-driven techniques) is used to predict the output ($y_m(t)$) of the structural system, and this output is compared with the measurements ($y(t)$) collected using sensors, to aid structural damage diagnosis. In the model, damage is represented by change in value of a particular model parameter. It is assumed that

only one parameter is damaged at a given instant, and the focus of this chapter is to identify that damage parameter and quantify the extent damage by estimating how much the value of that damage parameter has changed.

Until damage is detected, a tracking method (such as Kalman filtering (Kalman and Bucy 1961) or particle filtering (Arulampalam et al. 2002)) is used to estimate the state and the output of the system; the output estimated by the tracking method is denoted by $\hat{y}(t)$. Damage detection is performed by comparing $\hat{y}(t)$ against $y(t)$, and the residuals ($\hat{r}(t) = \hat{y}(t) - y(t)$) are subjected to hypothesis testing. Then, the uncertainty in detection is computed using classical hypothesis testing and Bayesian hypothesis testing methods. In classical hypothesis testing (Sankararaman and Mahadevan 2011), the choice of the significance level of testing is used to calculate the uncertainty in detection. In the Bayesian hypothesis testing-based approach, the likelihoods of two scenarios, i.e., “no fault” and “fault,” are calculated, and the uncertainty in the detection is quantified.

After damage is detected, tracking is completely stopped; from here on, only the model prediction ($y_m(t)$) is compared against experimental data ($y(t)$). Localization is performed in two steps. First, qualitative cause-effect relationships between model parameters and sensor measurements are used to isolate a prospective set of candidate damage parameters (Mosterman and Biswas 1999). Second, statistical tools are used to estimate the uncertainty in localization, by considering this set of candidate

damage parameters. In the classical statistics-based approach, a heuristic measure based on the least squares corresponding to each candidate damage parameter is used to quantify the uncertainty in damage isolation. In the Bayesian statistics-based approach, the probability of a particular parameter being faulty is calculated directly.

The third step of damage quantification is performed for each candidate damage parameter, using least square minimization or likelihood maximization. It is possible to approach damage quantification using system identification techniques that focus on parameter estimation in dynamic systems. The uncertainty in damage quantification is calculated using statistical confidence intervals and confidence bounds (Beck 1989; Dalai et al. 2005) in the classical statistics-based approach and using Bayesian inference techniques (Peterka 1981; Ching et al. 2006) in the Bayesian approach. Note that the latter approach directly calculates the probability distribution of the damage parameter, while the former approach provides a simpler measure of uncertainty.

Finally, the overall uncertainty in diagnosis is calculated by combining the uncertainties in the three stages of diagnosis using the principle of total probability, and the overall probability distribution of the candidate damage parameters are calculated. The methods for uncertainty quantification in damage detection, localization, and quantification are developed in the sections “[Uncertainty in Damage Detection](#),” “[Uncertainty in Damage Localization](#),” and “[Uncertainty in Damage Estimation](#).” The method for overall uncertainty quantification is presented in section “[Overall Uncertainty in Diagnosis](#),” and finally, all of the discussed methods are illustrated using the example of a structural frame in section “[Illustration: Structural Frame](#).”

Uncertainty in Damage Detection

Detection is based on comparison between healthy system response (provided by the tracking method, as $\hat{y}(t)$) and the observed system

response ($y(t)$). “Significant” deviation of system response from the expected performance indicates the presence of damage. It is necessary to perform a hypothesis test of significance for each individual output quantity that is measured using sensors (some structural systems may have multiple quantities that are being measured), since a deviation in any one output quantity indicates the presence of damage. Therefore, the rest of this section is presented using scalar residual quantities.

For the purpose of hypothesis testing (Fisher 1930; Romano 2005), the residuals $\hat{r}(t)$ are averaged over a moving window and the mean (μ) is subjected to a hypothesis test at a chosen significance level α , as follows:

$$H_0 \rightarrow \mu = 0 \text{ (No damage)} \quad (1)$$

$$H_1 \rightarrow \mu \neq 0 \text{ (Damage)} \quad (2)$$

The null hypothesis (H_0) in Eq. 1 states that the residuals are not significantly different from zero (thereby indicating no presence of damage), while the alternate hypothesis (H_1) in Eq. 2 states that the residuals are significantly different from zero (thereby, suggesting the presence of damage).

Classical Statistics-Based Approach

The uncertainty in detection is calculated as the probability that the damage is true after having detected a fault. This is established using the choice of the significance level α in the hypothesis test. The probabilities of different types of errors in hypothesis testing are well established in literature. The error of rejecting a correct null hypothesis is known as Type-I error, while the error of not rejecting a false null hypothesis is known as Type-II error. The probability of Type-I error is equal to α and the probability of Type-II error is denoted by β . This information can be written as:

$$P(H_0 \text{ is decision} | H_0 \text{ is true}) = 1 - \alpha \quad (3)$$

$$P(H_1 \text{ is decision} | H_0 \text{ is true}) = \alpha \quad (4)$$

$$P(H_0 \text{ is decision} | H_1 \text{ is true}) = \beta \quad (5)$$

$$P(H_1 \text{ is decision} | H_1 \text{ is true}) = 1 - \beta \quad (6)$$

Suppose that the residuals calculated from the system have been subjected to statistical hypothesis testing at significance level α and damage has

been detected, i.e., in other words, the decision is H_1 and the hypothesis H_0 has been rejected. The confidence in damage detection can be calculated as the probability that the truth is H_1 , subject to having rejected the hypothesis H_0 , i.e., $P(H_1 \text{ is true} | H_1 \text{ is decision})$. Let P_d denote this probability and this can be calculated using Bayes' theorem as:

$$P_d = \frac{P(H_1 \text{ is decision} | H_1 \text{ is true}) \times P(H_1 \text{ is true})}{P(H_1 \text{ is decision} | H_1 \text{ is true}) \times P(H_1 \text{ is true}) + P(H_1 \text{ is decision} | H_0 \text{ is true}) \times P(H_0 \text{ is true})} \quad (7)$$

In the context of damage detection, equal probabilities may be assigned to the events corresponding to " H_0 is true" and " H_1 is true." Therefore, $P(H_0 \text{ is true})$ and $P(H_1 \text{ is true})$ are both equal to 0.5, and hence:

$$P_d = \frac{1 - \beta}{1 + \alpha - \beta} \quad (8)$$

Bayesian Statistics-Based Approach

In this method, residuals are assumed to be normally distributed, i.e., $r \sim N(\mu, \sigma^2)$, and the probability distribution of the mean μ is continuously updated with measurements. Let the prior distribution be $N(\rho, \tau^2)$. If no additional information is available, Migon and Gamerman (Gamerman and Migon 1999) suggest $\rho = 0$ and $\tau^2 = \sigma^2$. Then, using the observed residual values ($r_1, r_2, \dots, r_n$, calculated using the available data D), the posterior distribution of μ can be calculated to be normal with mean and variance as follows:

$$\text{Posterior mean of } \mu = \frac{\frac{\rho}{\tau^2} + \frac{r_1 + r_2 + \dots + r_n}{\sigma^2}}{\frac{1}{\tau^2} + \frac{n}{\sigma^2}} \quad (9)$$

$$\text{Posterior variance of } \mu = \left(\frac{1}{\tau^2} + \frac{n}{\sigma^2} \right)^{-1} \quad (10)$$

Damage detection can be achieved through the use of Bayes' factor (B) which is defined as the

ratio of likelihood (Jeffreys 1998) of the two scenarios: "damage" and "no damage" as follows:

$$B = \frac{P(D|H_1)}{P(D|H_0)} \quad (11)$$

Jiang and Mahadevan (Jiang and Mahadevan 2008) derived the expression for the natural logarithm of the Bayes' factor, as:

$$\log(B) = -\frac{1}{2} \log(n+1) + \frac{(n \times \mu_r)^2}{2(n+1)\sigma^2} \quad (12)$$

In Eq. 12, μ_r refers to the mean of the observed residuals ($r_1, r_2, \dots, r_n$). If the Bayes' factor B is greater than 1, it implies that the data favors the hypothesis H_1 and hence suggests that there is damage. The confidence in damage detection, i.e., the probability that there is true damage, can be calculated as $P(H_1|D)$. If the prior probabilities $P(H_0)$ and $P(H_1)$ are chosen to be equal (implying that there is equal knowledge whether damage is present or not), then,

$$P(H_1|D) = \frac{B}{B+1} \quad (13)$$

This metric gives a direct measure of the probability of damage and hence is easy to compute and interpret in comparison with the classical statistics-based metric. Further, the probabilities

$P(H_0|D)$ and $P(H_1|D)$ can be updated continuously with measurements and thereby helps in tracking the uncertainty with acquisition of more measurements.

Uncertainty in Damage Localization

Damage localization is based on the comparison between damage signatures calculated from the system model and the symbols calculated from the system measurements. First, the system model is used to derive cause-effect relationships between the model parameters (that correspond to damage parameters) and model outputs (that correspond to system measurements). These cause-effect relationships, collectively called as the damage signatures, describe the qualitative changes in the zeroth and first derivatives of a system output as a result of a change in a system parameter. Sometimes, advanced modeling techniques such as bond graphs and temporal causal graphs have been used to generate such signatures (Sankararaman and Mahadevan 2011; Moustafa et al. 2010). Alternatively, these signatures may also be generated through simpler finite differencing approaches using the model, as a means of approximating derivatives that are equivalent to fault signatures (Sankararaman and Mahadevan 2013).

However, practical engineering systems usually have a large number of parameters that could become faulty, but only a small set of measurements. It may be difficult to exclusively isolate one particular damage parameter if several candidates have the same set of damage signatures. Instead, a set of prospective candidate damage parameters, $\Theta_i (i = 1 \text{ to } m)$, may be suggested by the isolation procedure. This section associates a measure of uncertainty to each of the candidate damage parameters. Recall that only one of these parameters is the true damage parameter because of the assumption that multiple damages do not occur concurrently. Further, after the damage has occurred, the value of the damage parameter Θ_i has changed to a new unknown value that is denoted by θ_i .

In order to assess the uncertainty in damage localization, measurements of the system are necessary after damage has been detected. Let t_f denote the time at which damage was detected. Let N denote the number of measurements available post-damage detection. Now it is necessary to consider all output quantities at once, as a vector. While y denotes the measurements, $y_m(\theta_i)$ denotes the model prediction available from time t_f until time $t_f + N$. Hereon, the model prediction needs to be considered as a function of the damage parameter because the value of the damage parameter has changed.

Classical Statistics-Based Approach

The uncertainty in damage localization is calculated using a heuristic least square-based metric. Assuming that Θ_i is the correct damage parameter, all the other damage parameters are assumed to be healthy (and their values are chosen to be nominal), and the least square estimate of θ_i can be calculated by minimizing Eq. 14.

$$S(\theta_i) = \sum_{t=t_f}^{t_f+N} (y_m(\theta_i) - y)^T (y_m(\theta_i) - y) \quad (14)$$

Let the least square estimate be $\hat{\theta}_i$ and $S(\hat{\theta}_i)$ the corresponding least square value. Then the probability that Θ_i is the true damage parameter can be heuristically said to be inversely proportional to $S(\hat{\theta}_i)$, and the constant of proportionality can be evaluated by considering all prospective damage parameters and equating the total probability to unity. Thus, the uncertainty in localizing the damage to the damage parameter Θ_i can be expressed mathematically as:

$$P_{\Theta_i} = P(\Theta_i \text{ is the damage parameter}) \propto \frac{1}{S(\hat{\theta}_i)} \quad (15)$$

and the constant of proportionality can be evaluated by imposing the condition:

$$\sum_{i=1}^{i=m} P_{\Theta_i} = 1 \quad (16)$$

Bayesian Statistics-Based Approach

The Bayesian approach directly calculates the probability that a given damage parameter is faulty and provides an exact measure of uncertainty as against a heuristic measure provided by the classical statistics-based approach. Let A denote the event that Θ_i is the true damage parameter. The Bayesian

approach calculates the probability of event A after collecting measurements from time t_f until time $t_f + N$. A prior probability P'_{Θ_i} is first assumed for each damage parameter to be the true damage parameter. A suitable assumption for the prior probability P'_{Θ_i} would be equal to $1/m$ since any of the candidate damage parameters is equally likely to be the true damage parameter. Then, this prior probability is updated using Bayes' theorem to calculate a posterior probability P''_{Θ_i} based on the following likelihood function:

$$P(D/A) = P(D|\Theta_i \text{ is the true damage parameter}) = L(\Theta_i) \quad (17)$$

This probability is the likelihood of event A , denoted by $L(\Theta_i)$. The calculation of this probability is done in two steps.

Let B denote the event that θ_i is the new value of the damage parameter Θ_i , i.e., after damage has occurred. Then the likelihood function corresponding to the event $A \cap B$ (the probability of observing the data conditioned on (i) event A , i.e., Θ_i is the true damage parameter, and (ii) event B , i.e., the true damaged value is θ_i), can be calculated using the probability distribution of residuals. In the context of statistical parameter estimation, it is common to assume a joint Gaussian probability distribution for the residuals (with zero mean and Σ covariance). Therefore, the likelihood can be written as (Pawitan 2001):

$$P(D|A \cap B) = L(\Theta_i, \theta_i) \propto \prod_{t=t_f}^{t_f+N} \exp\left(-(\mathbf{y}_m(\theta_i) - \mathbf{y})^T \Sigma (\mathbf{y}_m(\theta_i) - \mathbf{y})\right) \quad (18)$$

Note that $\mathbf{y}_m(\theta_i)$ and $\mathbf{y}$ are column vectors whose length is equal to number of output quantities. The covariance matrix Σ is a square matrix whose size is equal to the number of output quantities. Hence, the right hand side of Eq. 18 is a scalar.

This likelihood function $L(\Theta_i, \theta_i)$ is then used to calculate $L(\Theta_i)$, which, in turn, can be used to

update P'_{Θ_i} . Let $f'(\theta_i|\Theta_i)$ denote the probability density function of the damage value conditioned on the event that Θ_i is the damage parameter. The key is to calculate the posterior probabilities P'_{Θ_i} and $f''(\theta_i|\Theta_i)$ using Bayes' theorem. While the former indicates the uncertainty in damage localization, the latter indicates the uncertainty in damage quantification. Section “[Uncertainty in Damage Estimation](#)” focuses on the latter topic, i.e., estimation of $f''(\theta_i|\Theta_i)$, while the present section deals with the estimation of P'_{Θ_i} . It can be easily proved that (Sankararaman and Mahadevan 2013):

$$P''_{\Theta_i} \propto P'_{\Theta_i} \times P(D|A) \\ \propto P'_{\Theta_i} \times \int L(\Theta_i, \theta_i) f'(\theta_i|\Theta_i) d\theta_i \quad (19)$$

Note that there is a proportionality constant in Eq. 19. This constant can be evaluated by imposing the constraint that the sum of the values of P''_{Θ_i} corresponding to the m different damage parameters should be equal to unity. This condition is similar to that in Eq. 16, with the difference being that now the posterior probabilities need to be normalized. Note that the notion of posterior probabilities was absent while Eq. 16 was presented in the classical statistics-based method. Therefore, the advantage of the Bayesian approach is that such posterior probabilities can

be updated continuously with the availability of measurements, and therefore, the uncertainty in damage localization can be tracked continuously as a function of time.

Uncertainty in Damage Estimation

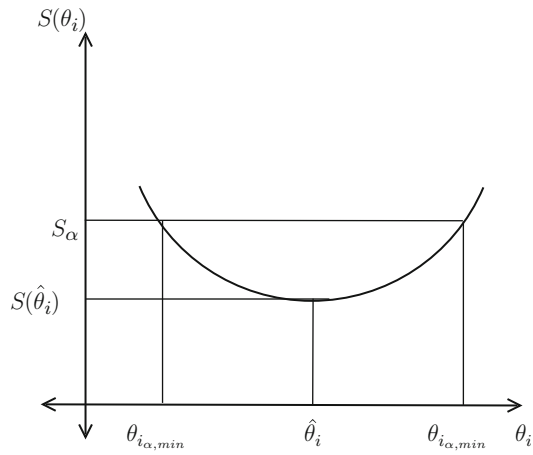
This section discusses the calculation of the uncertainty in the damage quantification stage, by estimating the uncertainty in the value of damage parameter. The classical statistics-based approach calculates statistical confidence intervals on the value of damage parameter, while the Bayesian statistics-based approach directly calculates the probability distribution of the value of the damage parameter.

Classical Statistics-Based Approach

Recall that the least square estimate of the damaged value of the damage parameter Θ_i was obtained by minimizing the sum of squares in Eq. 14. The least square estimate was denoted by $\hat{\theta}_i$ and the corresponding value of least squares was denoted by $S(\hat{\theta}_i)$. Using this information, it is possible to compute confidence bounds on the θ_i . The values of θ_i that satisfy the inequality in Eq. 20 belong to the α -level confidence bounds of θ_i . In several problems, there is an interval $([\theta_{i\alpha, \min}, \theta_{i\alpha, \max}])$ within which all values of θ_i satisfy this inequality, and therefore, this interval is referred to as the α -level confidence interval.

$$S(\theta_i) \leq S_\alpha = S(\hat{\theta}_i) \left(1 + \frac{p}{n-p} F_{p, N-p}^\alpha \right) \quad (20)$$

In Eq. 20, N denotes the number of observations over which the sum of square error was estimated and p denotes the length of the θ_i . In this case, only one parameter is assumed to be damaged at any given instant, and therefore, $p = 1$. Further, $F_{(p, N-p)}^\alpha$ with estimation (numerator) degrees of freedom p and residual (denominator) degrees of freedom $(N - p)$ at significance level α (Seber and Wild 1989).



Uncertainty Quantification in Structural Health Monitoring, Fig. 2 Confidence bounds in damage estimation

The confidence bounds of θ_i can be graphically illustrated using Fig. 2. The minimum $(\theta_{i\alpha, \min})$ and maximum $(\theta_{i\alpha, \max})$ values that satisfy Eq. 20 are also indicated in this figure, and they can be solved using optimization techniques (Sankararaman and Mahadevan 2011).

The minimum value $(\theta_{i\alpha, \min})$ can be estimated by minimizing the expression $(S(\theta_i) - S_\alpha)^2$ by considering values of θ_i that are strictly smaller than $\hat{\theta}_i$, i.e., $\theta_i < \hat{\theta}_i$. Similarly, the maximum value $(\theta_{i\alpha, \max})$ can be estimated by minimizing the expression $(S(\theta_i) - S_\alpha)^2$ by considering values of θ_i that are strictly larger than $\hat{\theta}_i$, i.e., $\theta_i > \hat{\theta}_i$.

The α -level confidence interval on θ_i , i.e., $[\theta_{i\alpha, \min}, \theta_{i\alpha, \max}]$, needs to be interpreted carefully. It acknowledges that the true value of the damage parameter Θ_i is unknown but deterministic; there is a $(1 - \alpha)\%$ probability that this interval will contain the true value of the damage parameter. This confidence interval can be calculated at multiple levels of α , but this is not equivalent to estimating the probability distribution of θ_i . In fact, in the classical statistics approach, θ_i is an unknown but deterministic quantity and hence cannot be assigned a probability distribution. However, with the availability of more data, the confidence interval can be reconstructed, and its

width may decrease with more acquisition of data, if the correct damage parameter is being estimated.

Bayesian Statistics-Based Approach

The Bayesian approach can assign probability distributions even to deterministic quantities (Calvetti and Somersalo 2007) since it assigns probabilities based on the extent of belief. Therefore, it calculates the entire probability distribution of θ_i , the value of the damage parameter Θ_i . The likelihood function $L(\Theta_i, \theta_i)$ earlier defined in Eq. 18 can be used in Bayes' theorem (Leonard and Hsu 1999) and the posterior probability distribution of θ_i can be calculated as:

$$f''(\theta_i|\Theta_i) = \frac{L(\Theta_i, \theta_i)f'(\theta_i|\Theta_i)}{\int L(\Theta_i, \theta_i)f'(\theta_i|\Theta_i)d\theta_i} \quad (21)$$

As explained earlier, $f'(\theta_i|\Theta_i)$ is the prior density function and represents the knowledge about the value of the damage parameter Θ_i before performing damage quantification. It is likely that there is no such "prior" knowledge, and a uniform distribution may be assumed, to begin with. Then, this distribution can be updated continuously with acquisition of more measurements. While Markov Chain Monte Carlo methods are commonly used to solve Eq. 21, advanced numerical integration techniques (Sankararaman and Mahadevan 2011; McKeeman 1962) may be used to provide quicker results. Such numerical techniques are particularly useful since there is only one damage parameter that is being estimated.

The probability distribution of the value θ_i of the damage parameter Θ_i can be calculated and updated continuously with acquisition of more measurements. For example, a first set of measurements can be used to update the uniform prior distribution to the corresponding posterior distribution. Then, this posterior distribution is used as a prior while updating using the second set of measurements. This procedure can be continued until a desired confidence level in damage quantification is attained. This convenience of continuously updating the uncertainty is a unique

feature of the Bayesian method and is not applicable when classical statistics-based techniques are used.

Overall Uncertainty in Diagnosis

Having estimated the uncertainty in each of the individual steps of damage diagnosis, i.e., detection, localization, and quantification, respectively, it is important to aggregate these three results to compute the overall uncertainty in damage diagnosis. In order to address this issue, consider the uncertainty estimated in the damage quantification step. The probability distribution of θ_i is, in fact, conditioned on two events: damage detection being correct and damage localization being correct. The goal in this section is to estimate the probability distribution of θ_i by accounting for the uncertainty in damage detection and damage localization, and the resultant distribution will be reflective of the overall uncertainty in diagnosis.

Recall that P_d denotes the confidence in damage detection, and P_{Θ_i} denotes the confidence in damage localization. While the former refers to the probability that the damage is true, the latter refers to the probability that Θ_i is the true damage parameter. The unconditional distribution of θ_i can be estimated by considering three events:

1. There is no damage, and hence, the structural health monitoring system triggered a false alarm. This event has a probability $(1 - P_d)$.
2. There is damage, but Θ_i is not the correct damage parameter. The probability for this event is equal to $P_d \times (1 - P_{\Theta_i})$.
3. There is damage and Θ_i is the correct damage parameter. The probability for this event is equal to $P_d \times P_{\Theta_i}$.

In the first two cases, the probability distribution of θ_i is still nominal (i.e., no damage). Let $f(\theta_i|\text{healthy})$ denote this probability distribution. In the third case, the probability distribution of θ_i corresponds to the distribution estimated in section "Uncertainty in Damage Estimation."

Using the tools of total probability, the unconditional distribution of θ_i can be calculated as:

$$\begin{aligned} f(\theta_i) = & (1 - P_d) \times f(\theta_i|\text{healthy}) + P_d \times (1 - P_i) \\ & \times f(\theta_i|\text{healthy}) + P_d \times P_i \times f''(\theta_i|\Theta_i) \end{aligned}$$

(22)

Since the above unconditional distribution accounts for the uncertainty in damage detection, damage localization, and damage quantification, it is indicative of the overall uncertainty in diagnosis. This distribution also can be continuously estimated in time, with the acquisition of measurements. If the true damage parameter is being correctly estimated, then it is expected that the variance of this distribution will decrease with acquisition of measurements, thus indicating increasing confidence in diagnosis.

Illustration: Structural Frame

Consider a simple two-storied structural frame, as shown in Fig. 3. This section illustrates the various methods for uncertainty quantification in structural health monitoring using this structural frame as a numerical example.

Description of the Structural Frame

This system has six parameters, m_1, m_2, k_1, k_2, D_1 , and D_2 , that represent the masses, stiffness, and damping parameters of the two stories of the structural frame. The inputs to the system are forces at the two levels and the measured outputs

are the accelerations at the two levels. The nominal values of these quantities are provided in Table 1.

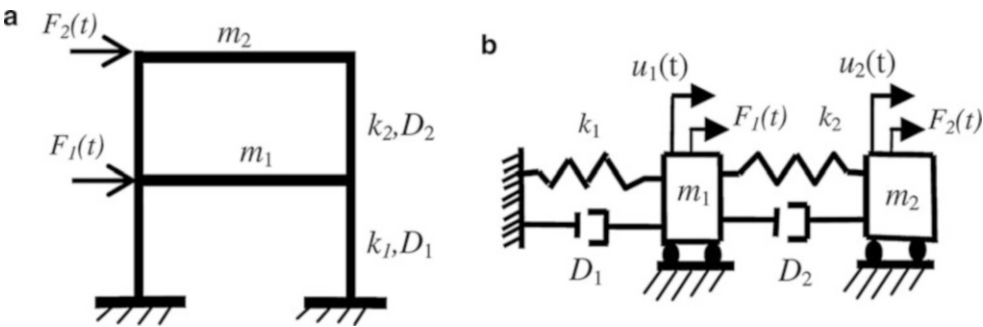
The masses of the first and second floors are assumed not to change. Hence, there are only four parameters that could be affected by damage. Using the damage signatures (cause-effect relationships between the model parameters and measurements), it is possible to qualitatively identify the story (floor) that is damaged, and hence the prospective damage parameters are the stiffness and damping of that particular story (floor).

Damage Simulation

In order to verify the effectiveness of the methodology, it is assumed that a known amount of damage is introduced during the normal operation of the structure. For the example of illustration, consider that the value of k_1 is reduced to 80 % of its nominal value at a particular time instant. The goal is to make use of the methods described in this chapter and diagnose this

Uncertainty Quantification in Structural Health Monitoring, Table 1 Two-story frame: numerical values

Parameter	Value	Unit
k_1	30,700	N/m
k_2	44,300	N/m
D_1	307	Ns/m
D_2	443	Ns/m
m_1	136	Ns ² /m
m_2	66	Ns ² /m
$F_1(t)$	75 sin(9 t)	N
$F_2(t)$	100 sin(9 t)	N



Uncertainty Quantification in Structural Health Monitoring, Fig. 3 (a) Structural frame, (b) mechanical model

damage. For this purpose, sensors are used to measure the accelerations of both the stories. The sensors are not accurate and the measurement noise of the sensors is accounted for, through inclusion in the tracking method.

Damage detection is done through hypothesis testing, as explained in section “[Uncertainty in Damage Detection](#).” Residuals are computed by differencing the observed sensor measurements and the output of tracking system. A moving window is used to calculate the mean of the residuals, and this mean is subjected to hypothesis testing to detect damage. The classical statistics-based hypothesis testing detects damage with 95 % confidence. Since the values of Type-I and Type-II errors in classical hypothesis testing do not change, the confidence in detection also does not change. In contrast, the Bayesian method facilitates quantification and updating of uncertainty from a 61 % confidence at 5 s to 99.98 % confidence at 5.05 s.

Without loss of generality, the time at which damage is detected is denoted as $t = 0$ s. Once damage has been detected, tracking is not used. The measurements and model predictions are directly used for the purpose of quantifying the uncertainty in damage localization and damage quantification. For this purpose, three sets of data are considered, with each set consisting of 10 s of acceleration data. The first set of data consists of measurements from $t = 0$ s until $t = 10$ s. The second set of data consists of measurements from $t = 10$ s until $t = 20$ s. Finally, the third set of data consists of measurements from $t = 20$ s until $t = 30$ s. Note that continuous updating of uncertainty through these three sets of data is possible only using the Bayesian method. Therefore, continuous updating results are provided below, only for the Bayesian method. On the other hand, the results using classical statistics-based method are provided by considering all the available data as a single set.

Damage localization, as explained in section “[Uncertainty in Damage Localization](#),” is performed in two steps. First, qualitative damage signatures are used to identify a list of candidate damage parameters. In this example, as stated earlier, it is possible to uniquely identify which

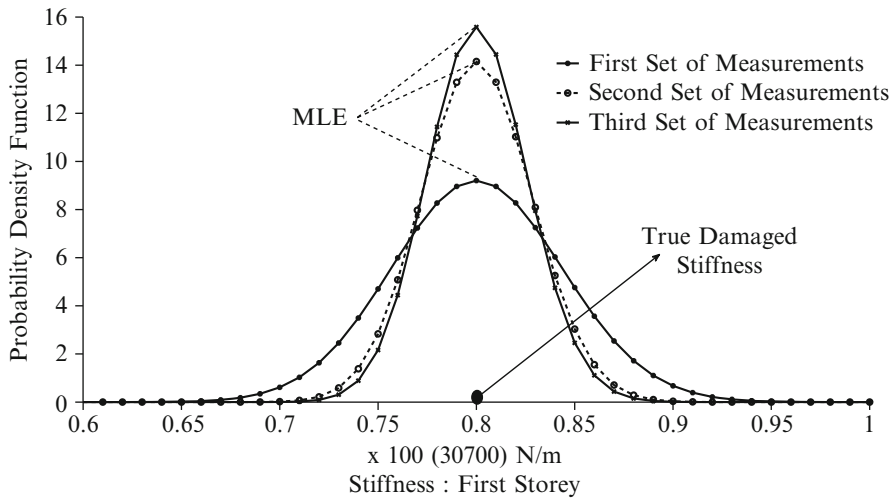
story (floor) is faulty. Therefore, the candidate damage parameters are k_1 and D_1 , respectively. The heuristic measures of uncertainty using classical statistics-based methods yield 98 % and 2 % probabilities for k_1 and D_1 to be the damage parameters, respectively. On the other hand, the uncertainty in localization of k_1 , using the Bayesian method, is calculated for the first data set as 99 % and updated using Bayes’ theorem for the following two sets to 100 % and 100 %.

Damage quantification using the classical statistics procedure yields a 90 % confidence interval of (24,509 N/m, 24,612 N/m). The Bayesian approach directly calculates the probability distribution of k_1 as explained in section “[Uncertainty in Damage Estimation](#).” The overall uncertainty in diagnosis can be calculated as in section “[Overall Uncertainty in Diagnosis](#),” and the corresponding probability density function is shown in Fig. 4. From Fig. 4, it can be seen that (i) the maximum likelihood estimate (MLE) of k_1 matches closely with the true damage value, and (ii) the uncertainty in the estimate of k_1 decreases with acquisition of more measurements.

The above diagnosis procedure was also extended to five-story frames, where qualitative localization is able to isolate to the floor of damage accurately. Different damage scenarios were considered for the five-story frame, and these were detected and isolated (localized) satisfactorily with the proposed method. The damage was quantified with high accuracy. Uncertainty quantification associated with each of these three processes was also successful, and both classical statistics-based and Bayesian methods were investigated. The easy extension to multistory frames clearly indicates that this diagnosis methodology is applicable to practical structures.

Summary

This chapter presented computational methods to quantify the uncertainty in structural health monitoring, by considering the three steps of damage diagnosis, i.e., damage detection, damage localization, and damage quantification. Classical statistics-based methods and Bayesian methods



Uncertainty Quantification in Structural Health Monitoring, Fig. 4 Continuous uncertainty quantification in structural health monitoring

were investigated for this purpose. Damage detection was based on statistical hypothesis testing. While the classical hypothesis testing procedure calculated the uncertainty in detection based on the significance level of testing, the Bayesian hypothesis testing procedure calculated the likelihood of “no-damage” scenario to “damage” scenario, thereby providing a robust Bayesian metric to assess the confidence in detection. The uncertainty in localization was quantified using heuristic measures in the classical statistics-based approach, while the Bayesian method directly quantified the probability that a given parameter is the true damage parameter. The uncertainty in damage quantification was expressed in statistical confidence intervals in the classical statistics-based approach, while the Bayesian method gives the entire probability distribution of the damage parameter. Finally, the uncertainties in the three different steps were combined using the principle of total probability to provide an estimate of the overall uncertainty in diagnosis. The classical statistics-based method and Bayesian method are fundamentally different in interpretation and implementation. However, it can not only be seen that the results from Bayesian method is easy to implement and interpret but also that the Bayesian method also facilitates easy updating of uncertainty with acquisition of measurements.

It must be noted that this chapter discussed practical situations where only one damage parameter is present at any given instant of time. In some cases, it may be necessary to consider multiple, concurrent damage parameters during structural health monitoring. Ongoing research is addressing such challenges and also studying the application of structural health monitoring methods to composites in detail.

Cross-References

- [Building Damage from Multi-resolution, Object-Based, Classification Techniques](#)
- [Damage to Buildings: Modeling](#)
- [Laser-Based Structural Health Monitoring](#)
- [Seismic Event Detection](#)
- [Spatial Filtering for Structural Health Monitoring](#)

References

- Achenbach JD (2000) Quantitative nondestructive evaluation. *Int J Solids Struct* 37(1):13–27
- Arulampalam MS, Maskell S, Gordon N, Clapp T (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans Signal Process* 50(2):174–188
- Beck, James L (1989) Statistical System Identification of Structures. In: *Proceedings of ICOSSAR '89*, the 5th

- International Conference on Structural Safety and Reliability, San Francisco, August 7–11, 1989. ASCE, New York, pp. 1395–1402
- Calvetti D, Somersalo E (2007) An introduction to Bayesian scientific computing: ten lectures on subjective computing, vol 2. Springer, New York
- Ching J, Beck JL, Porter KA (2006) Bayesian state and parameter estimation of uncertain dynamical systems. *Probab Eng Mech* 21(1):81–96
- Dalai M, Weyer E, Campi MC (2005) Parametric Identification of Nonlinear Systems: Guaranteed Confidence Regions. In: Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05. 44th IEEE Conference on, Seville, 12–15 Dec 2005 pp 6418–6423
- Farrar CR, Worden K (2007) An introduction to structural health monitoring. *Philos Trans R Soc A Math Phys Eng Sci* 365(1851):303–315
- Fisher RA (1930) Statistical methods for research workers 3rd edn. Oliver and Boyd, London
- Gamerman D, Migon HS (1999) Statistical inference: an integrated approach
- Jeffreys H (1998) The theory of probability. Oxford University Press, Oxford, UK
- Jiang X, Mahadevan S (2008) Bayesian wavelet methodology for structural damage detection. *Struct Control Health Monit* 15(7):974–991
- Kalman RE, Bucy RS (1961) New results in linear filtering and prediction theory. *J Basic Eng* 83(3):95–108
- Leonard T, Hsu JSJ (1999) Bayesian methods: an analysis for statisticians and interdisciplinary researchers, vol 5. Cambridge University Press, Cambridge
- McKeeman WM (1962) Algorithm 145: adaptive numerical integration by Simpson's rule. *Commun ACM* 5(12):604
- Mosterman PJ, Biswas G (1999) Diagnosis of continuous valued systems in transient operating regions. *IEEE Trans Syst Man Cybern Part A* 29(6):554–565
- Moustafa A, Mahadevan S, Daigle M, Biswas G (2010) Structural and sensor damage identification using the bond graph approach. *Struct Control Health Monit* 17(2):178–197
- Pawitan Y (2001) In all likelihood: statistical modelling and inference using likelihood. Oxford University Press, Oxford
- Peterka V (1981) Bayesian approach to system identification. Trends and progress in system identification, vol 1. Pergamon Press, Oxford, pp 239–304
- Romano JP (2005) Testing statistical hypotheses. Springer, New York
- Sankararaman S, Mahadevan S (2011) Uncertainty quantification in structural damage diagnosis. *Struct Control Health Monit* 18(8):807–824
- Sankararaman S, Mahadevan S (2013) Bayesian methodology for diagnosis uncertainty quantification and health monitoring. *Struct Control Health Monit* 20(1):88–106
- Seber GAF, Wild CJ (1989) Nonlinear regression. Wiley, New York

Uncertainty Theories: Overview

E. Elwood and R. Corotis

Department of Civil, Environmental and Architectural Engineering, University of Colorado, Boulder, CO, USA

Synonyms

Overview of Uncertainty in Earthquake Engineering; Uncertainty in Earthquake Engineering Background; Uncertainty in Earthquake Engineering Introduction; Uncertainty in Earthquake Engineering Overview

Introduction

Uncertainty is associated with many aspects of earthquake engineering. Considerable uncertainty resides in the prediction of earthquake occurrence, prediction of seismic forces, estimated capacity of a structural system to withstand such forces, and consequences associated with seismic events. This challenging problem lends itself to information/partial knowledge and uncertainty in many forms.

Traditionally, probabilistic methods have often been employed to systematically treat the inherent uncertainty in earthquake engineering applications. Notably, probability theory is used in seismic hazard analysis first outlined by Cornell (1968). Probabilistic methods also underlie load and resistance criteria for structural design. A probabilistic design methodology is based on initial work from Cornell (1969) and Ravindra and Galambos (1978) among others. These efforts form the basis of codified design of structural systems that are founded on structural reliability concepts (Ellingwood et al. 1980). Probabilistic concepts have long been utilized for loss estimation of potential earthquake events. Early studies by Whitman (1973) and Whitman et al. (1973) developed damage probability matrices that aim to characterize the effect of varying

levels of earthquake intensities. These matrices express the probability that a particular level of building damage will result from various seismic intensities. Other significant efforts that conceptualize loss estimates within a probabilistic framework include ATC-13 (1985) and HAZUS (NBIS 1997). In recent decades, beginning with FEMA 273 (1997) and FEMA 356 (2000) and more comprehensively in the Performance Based Earthquake Engineering (PBEE) approach outlined by the Pacific Earthquake Engineering Research Center (PEER) (Cornell and Krawinkler 2000; Porter 2003), performance-based measures include probabilistic concepts as a methodology that incorporates uncertainty associated with seismic hazard, vulnerabilities, and consequences related to the performance of structural systems. The PEER framework is presented in Eq. 1:

$$p(DV) = \int \int \int p(DV|DM)p(DM|EDP) \\ p(EDP|IM)p(IM)dIMdEDPdDM \quad (1)$$

Where DV is a decision variable, DM is a damage measure, EDP is an engineering demand parameter, and IM is an intensity measure. The notation $p(x)$ here represents the probability distribution function of x (and $p(x|y)$ the conditional probability distribution of x given a particular value for the variable y) and will be discussed in more detail in a later section.

This equation illustrates the types of terms that are involved in assessing the decision process in seismic design. The intensity measure (IM) consists of various characteristics of an earthquake, such as peak ground acceleration, maximum energy release, duration, etc. Each of these is dependent on the shaking mechanism at the epicenter and the geophysical conditions through which the seismic waves travel. As can be surmised, these factors are likely to exhibit different types of uncertainty, some of which include experience and expert judgment, while in current approaches, all are generally modeled

probabilistically. The engineering demand parameter (EDP) reflects the response quantities of a structure that are important to its ability to perform during the earthquake. Such quantities are typically taken as the maximum inter-story drift, the maximum displacement, or the maximum acceleration. The damage measure (DM) may include multiple failure modes of beams, columns, and floor slabs, as well as losses to nonstructural elements. As with the intensity measures, some of these quantities might involve expert judgment. Finally, the decision variable (DV) may reflect such issues as economic and social as well as structural loss to a community. Therefore, design decisions often incorporate the combination of physical and societal factors, for which the uncertainties exhibit very different characteristics of sources of knowledge and uncertainty.

Uncertainty can assume different forms and can arise from different sources. Uncertainty may arise from expert opinions/judgments, qualitative assessments (Vick 2002), measurement error (or measurement uncertainty), and/or the reliability of the source of information (Klir 2006). For example, probabilistic seismic hazard analysis combines information from geologic evidence and identification of fault locations with data from previous seismic events (Cornell 1968; Vick 2002). In loss estimations, expert opinions are often integrated with knowledge of previous earthquake events for the prediction of damage and losses of seismic scenarios that may contain little or no data (ATC 1985). This is particularly the case for large (high intensity), rare earthquake events.

Although probabilistic methods offer several tools for manipulating uncertainty and partial information, methods outside of classical probability theory, under the umbrella of *Generalized Information Theory* (GIT), help to mathematically expand information theory beyond traditional probability (Klir 2006). Prominent uncertainty theories include probability theory, possibility theory, and evidence theory. The following presents a historical perspective and mathematical overview of these uncertainty theories.

Probability Theory

Probability Theory: Historical Perspective

Probability theory initially emerged as a means of enhancing the treatment of uncertainty in the face of what is unknown or uncertain. Its inception, credited to Blaise Pascal and Pierre de Fermat in 1654, originated from games of chance (Hacking 2006). A probability came to be defined as the ratio of a particular outcome of interest to the total number of possible outcomes (Dubois and Prade 1988; Vick 2002). Characterized as a “calculus of chances,” this conception formed the frequency interpretation of probability (Dubois and Prade 1988, p. 4). This interpretation views probabilities as limits of observed frequencies of repeatable events (i.e., placing repeated bets), each event characterized by an uncertain but observable outcome. As the theory evolved, it became apparent that probabilities could assume more than one interpretation. The “subjectivists” offered another interpretation of probabilities as a “degree of belief” or commonly termed *Bayesian degrees of belief*. This class of probabilities originated from the work of Thomas Bayes, an English mathematician who established the well-known Bayes’ theorem. This interpretation views probabilities of uncertain events as a measure of one’s belief about its occurrence (Benjamin and Cornell 1970; Dubois and Prade 1988; Vick 2002).

Although a straightforward interpretation of probabilities has eluded the field since its beginnings, the relative frequency and subjective degree of belief are prominently recognized in applying probabilistic concepts (Hacking 2006; Vick 2002).

Probability Theory: Mathematical Overview

In probability theory, uncertainty is expressed in the form of a *probability distribution function* by

$$p : x \rightarrow [0, 1] \quad (2)$$

A variable, whose distribution is defined by $p(x)$, is usually referred to as a *random variable* (Ang and Tang 2007). Given a probability distribution function p , an event in the space of possibilities

can be specified by a range of values in the universe X (commonly termed the *sample space* in the context of probability theory) as a unique value. This value is termed a *probability measure*, here denoted $\text{Pro}(A)$, where A denotes an event within the sample space (S) of all possible events. A probability measure is defined by Eq. 3 (Klir 2006).

$$\text{Pro}(A) = \sum_{x \in A} p(x) \quad (3)$$

Equation 3 can be interpreted as “the probability of A .” This probability must conform to certain mathematical properties or *axioms*. Probability theory is based on the following axioms (Benjamin and Cornell 1970):

$$\begin{aligned} (\text{Pro1}) \quad & 0 \leq \text{Pro}(A) \leq 1 \\ (\text{Pro2}) \quad & \text{Pro}(S) = 1, \text{ where } S \text{ is the sample space} \\ & \text{(or set of all sample points)} \\ (\text{Pro3}) \quad & \text{Pro}(A \cap B) = \text{Pro}(A) + \text{Pro}(B) \\ & \text{when sets } A \text{ and } B \text{ are disjoint;} \\ & A \cap B = \emptyset \end{aligned} \quad (4)$$

Equation 4 indicates that probability measures are positive numbers, and as indicated by the third axiom (Pro3), a basic property of probability measures is that they are additive. This requirement is commonly referred to as the *additivity requirement*. This requirement ensures the sum of probabilities over any collectively exhaustive, mutually exclusive set of events is equal to 1 (unity). Considering a single event A , and its complement denoted $\bar{A}$, from (Eq. 4) it can be shown that

$$P(\bar{A}) = 1 - P(A) \quad (5)$$

For the special case when the probability of an event depends on the outcome (occurrence or nonoccurrence) of another event, such a probability is a *conditional probability* given by

$$\text{Pro}(A|B) = \frac{\text{Pro}(A \cap B)}{\text{Pro}(B)} \quad (6)$$

where the numerator, $\text{Pro}(A \cap B)$, is the probability of the *joint event* or the probability of the intersection of events A and B (i.e., $A \cap B$). Equation 6 can be interpreted as the probability of event A given that event B has occurred. Rearranging Eq. 6, the probability of the joint event $\text{Pro}(A \cap B)$ can be written as

$$\text{Pro}(A \cap B) = \text{Pro}(A|B)\text{Pro}(B) \quad (7)$$

Or equivalently,

$$\text{Pro}(A \cap B) = \text{Pro}(B|A)\text{Pro}(A) \quad (8)$$

From this, the “inverse” probability can be obtained by substituting the numerator in Eq. 6 with Eq. 8.

$$\text{Pro}(A|B) = \frac{\text{Pro}(B|A)\text{Pro}(A)}{\text{Pro}(B)} \quad (9)$$

Equation 9 is known as Bayes’ theorem or Bayes’ rule. Using the theorem of total probability given by

$$\text{Pro}(B) = \sum_{i=1}^n \text{Pro}(B|A_i)\text{Pro}(A_i) \quad (10)$$

where $i = 1, 2, \dots, n$, denote a set of mutually exclusive collectively exhaustive events $A_1, A_2, \dots, A_n$. Bayes’ theorem can be expanded such that

$$\text{Pro}(A_j|B) = \frac{\text{Pro}(B|A_j)\text{Pro}(A_j)}{\sum_{i=1}^n \text{Pro}(B|A_i)\text{Pro}(A_i)} \quad (11)$$

where (Eq. 11) is the conditional probability of A_j given that event B has occurred. The theorem of total probability given in Eq. 10 simply expresses the probability of B, $\text{Pro}(B)$, in terms of its conditional probabilities.

Equation 11 (Bayes’ theorem) provides a means for incorporating new information by updating *prior probabilities* or previous information expressed in the form of a probability (Benjamin and Cornell 1970). These updated

probabilities, or so-called *posterior probabilities*, lead to new numerical probabilistic values of the event of interest by considering additional knowledge about the relative likelihood of a particular event. Thus, probability theory provides a logical basis for manipulation of data in the form of relative frequencies and a mathematically permissible method for the continuous updating probabilistic values as a function of the current state of knowledge through the use of Bayes’ theorem.

Probability Theory: Interpretations

It is well established that probabilistic values can be characterized by two general classes of interpretation: relative frequencies of an observed outcome and Bayesian probabilities (or so-called subjective probabilities). The class of subjective probabilities allows for a broader context of probability theory. This interpretation proposes that probability can be justified not necessarily by the objective or frequentist basis (a frequency of occurrence among “trials”) but to single occurrence events in the form of a measure of one’s uncertainty about a particular event (Dubois and Prade 1988; Vick 2002). From this, Bayes’ theorem serves as a mathematical basis for manipulating relationships between prior and new probabilistic information. As such, the axioms of probability theory serve as a foundation for expressing uncertainty in multiple contexts.

Possibility Theory

Possibility Theory: Historical Perspective

Possibility theory gained much attention in the latter half of the twentieth century through the work of Zadeh (1978), which linked the concepts of fuzzy sets to the notions of possibilities. Possibility theory differs from probability theory in that it explicitly recognizes the case when evidence or judgments support the possibility of one event, but does not necessarily implicate evidence for the contrary event (Dubois and Prade 1988). In probability theory, uncertainty is represented by a single probability measure.

If either the probability of an event or the probability of its negation (or complement) is known, the additivity requirement guarantees that the probabilities of both are known. In possibility theory, by comparison, uncertainty is represented by dual measures, termed *possibility* and *necessity measures* (Dubois and Prade 1988; Klir 2006). These measures are founded on the basic concepts of *possibility theory*.

Possibility Theory: Mathematical Overview

Consider a set of mutually exclusive alternatives X , of which some alternatives may or may not be possible (Klir 2006). To determine truth (i.e., the correct alternative), suppose that relevant information is obtained through experimentation, testing, or observation. Presumably, this new knowledge will determine that some of the alternatives are supported by the new knowledge (or evidence E) while other alternatives are deemed impossible. The alternatives supported by the new evidence form the subset E of X . The basis of possibility theory builds on those alternatives contained in E , the subset of the remaining possible alternatives. Mathematically, this is expressed by a *basic possibility function* r_E , defined by

$$r_E(x) = \begin{cases} 1, & \text{when } x \in E \\ 0, & \text{when } x \notin E \end{cases} \quad (12)$$

Equation 12 states that alternatives supported by evidence are possible and functionally equal to one. Conversely, zero is assigned to those alternatives that are impossible.

As an extension to differentiating between possibility and impossibility in the binary sense, a generalization of classical possibility theory, a *theory of graded possibilities*, distinguishes grades (or degrees) of possibilities (Klir 2006). Instead of designating alternatives as possible (1) or impossible (0) in accordance with the characteristic possibility function in Eq. 12, alternatives are assigned a value on $[0, 1]$. Defined in Eq. 13, a graded possibility function maps the individual elements x in the universe X to the unit interval (Klir 2006):

$$r : X \rightarrow [0, 1] \quad (13)$$

Binary possibilities can be viewed as a special class of graded possibilities (i.e., when a particular alternative is fully possible $r_E(x) = 1$ or impossible $r_E(x) = 0$). For this reason, the term *graded* and subscript E can be omitted in the generalized possibilistic mappings. Thus, these functions are referred to as *basic possibility functions*, denoted $r(x)$ (Klir 2006).

In probability theory, probability measures (Pro) are determined by the values expressed in a probability distribution function $p(x)$. Similarly, in possibility theory, a basic possibility function $r(x)$ contains a description for each *singleton* (or individual element) that expresses the *possibility measure*, here denoted $\text{Pos}(A)$. A possibility measure is determined for a given event A by the following formula (Klir and Yuan 1995):

$$\text{Pos}(A) = \max_{x \in A} [r(x)] \quad (14)$$

where $A \in P(X)$, and $P(X)$ denotes the power set of X (the collection or group of all possible subsets).

Possibility theory assumes a reordering of the subsets of a possibility function such that the evidence is decreasing. This is expressed symbolically as $r(x_i) \geq r(x_{i+1})$. Reordered singletons can be aggregated into subsets A_n as follows:

A_1	x_1
A_2	$x_1 \cup x_2$
A_3	$x_1 \cup x_2 \cup x_3$
A_4	$x_1 \cup x_2 \cup x_3 \cup x_4$
$\vdots$	$\vdots$
A_n	$x_1 \cup x_2 \cup x_3 \cup x_4 \dots x_n$

Thus, the evidence from this reordering is such that $A_1 \subset A_2 \subset A_3 \subset \dots \subset A_n$. With this property, the sets are such that they are *nested* (Klir 2006). In the case of nested evidence, possibility measures are said to be *consonant*.

The preceding description introduces the concepts of possibility functions and possibility measures. These topics provide the foundation of a more complete measurement of evidence in

the form of dual measures of possibility and necessity.

In the context of a possibility measure, the alternatives not possible in the negation of the complement of A define a *necessity measure* $Nec(A)$. This is expressed by Eq. 15 (Klir and Yuan 1995).

$$Nec(A) = 1 - Pos(\overline{A}) \quad (15)$$

Thus, in comparison to probability theory, two measures of possibility and necessity fully characterize the uncertainty of an event A.

Similarly, possibility measures can be defined from necessity measures by Eq. 16 (Klir and Yuan 1995):

$$Pos(A) = 1 - Nec(\overline{A}) \quad (16)$$

Demonstrated in Eqs. 15 and 16, both possibility and necessity measures of A can be derived if one knows the complete possibility function (also termed a *possibility profile*).

Recall from Eq. 14, possibility measures can be determined from possibility values represented by a basic possibility function. Necessity measures do not share this property (Klir 2006). Therefore, much of the possibility calculus is defined in terms of possibility measures.

Possibility theory is based on the following mathematical properties or axioms:

$$\begin{aligned} (Pos1) \quad Pos(\emptyset) &= 0 \\ (Pos2) \quad Pos(X) &= 1 \\ (Pos3) \quad Pos(A \cup B) &= \max[Pos(A), Pos(B)] \\ &\text{for any sets } A, B \in P(X) \end{aligned} \quad (17)$$

Relating Eq. 14 and (Pos2) yields

$$Pos(X) = \max_{x \in X} [r(x)] = 1 \quad (18)$$

The relationship expressed in Eq. 18 guarantees that at least one alternative in the universe is fully possible (Klir 2006). From this, the second axiom (Pos2),

and the reordering property of elements, it is known that the first element in any ordered possibility function is always equal to unity or, mathematically, $r_1 = r(x_1) = 1$. This is the property of *possibilistic normalization* (Klir 2006).

Possibility theory is distinguished from probability theory by the third axiom (Pos3), the max-min requirement. Dubois and Prade (1988) suggest that the interpretation of the third axiom follows an intuitive notion of possibility. Achieving what is possible in $A \cup B$ realizes the lesser of what is possible for either A or B.

From (Eq. 17), it can be shown that possibility and necessity measures satisfy the following inequalities (Klir 2006):

$$\begin{aligned} Pos(A) + Pos(\overline{A}) &\geq 1 \\ Nec(A) + Nec(\overline{A}) &\leq 1 \end{aligned} \quad (19)$$

From Eqs. 19 and 15, it can be further shown that (Klir 2006)

$$Nec(A) \leq Pos(A) \quad (20)$$

Equation 20 states that events cannot be necessary unless they are possible (Dubois and Prade 1988). Stated another way by Zadeh (1986), necessity implies possibility.

Possibility Theory: Interpretations

There are two main perceptions of possibility theory. One view assumes an interpretation first introduced by Zadeh (1978), the *fuzzy set interpretation of possibility theory*. This interpretation will be discussed in the following section. The other interpretation sees possibility theory as a special case of evidence theory. In the latter case, possibility theory is obtained when a body of evidence in the context of evidence theory is nested (Shafer 1987). This interpretation will be discussed in more detail in section “Evidence Theory.”

As discussed by Dubois and Prade (1988), the fuzzy set interpretation of possibility theory associates a degree of possibility with how well an object corresponds to a category into which it may be placed. From this perspective, the

“categories” are crisp sets, and the associated degree of measure of a particular element expresses how well it more or less relates to a particular set of interest. Within the context of this interpretation, a possibility function (or possibility profile) is developed from information expressed in the form of a fuzzy set. This idea is represented mathematically by

$$r(x) = \mu_{\underline{A}}(x) \quad (21)$$

In this expression, $\underline{A}$ is a fuzzy set on a universe X , and the membership function $\mu_{\underline{A}}(x)$ (interpreted as the degree of membership of an element x in the fuzzy set $\underline{A}$) assigns a grade or degree of “possibility,” $r(x)$, that the element x is $\underline{A}$. Thus, the fuzzy set interpretation of possibility theory provides a mathematical framework linking the concepts of fuzzy sets with formal possibility measures.

Evidence Theory

Evidence Theory: Historical Perspective

In the mid-1960s Dempster introduced work in the area of statistical inference founded on the idea of upper and lower probabilities (Dempster 1967; Shafer 1976). Roughly a decade later, Shafer related Dempster’s work to the uncertain expressions of judgments and introduced his own developments (Shafer 1976). In his work, Shafer presented a *theory of evidence* in which *belief functions* could be formalized, arising not necessarily from bounded probabilities, but from a degree of belief based on available evidence (Shafer 1976). Shafer retitled Dempster’s original notions of upper and lower probabilities *beliefs* and *plausibilities* (Yager and Liu 2008). As the works became known to the artificial intelligence community, the theory fell under the name of the Dempster-Shafer Theory (DST) of evidence or, commonly, *evidence theory* (Shafer 2012).

Evidence theory is based on a measure of degree of belief, called a *belief measure* (Shafer 1976).

Belief measures (also commonly referred to as belief functions) aim to generalize the well-known interpretation of subjective probability theory – i.e., the Bayesian probability for a subjective measure of uncertainty – to a broader concept of “evidence.” Like possibility and necessity measures, evidence theory is developed from dual measures of belief and plausibility (Klir and Yuan 1995). These measures express “beliefs” or judgments formulated by available evidence (Yager and Liu 2008). Although Dempster’s original works were closely linked to classical probability theory, there are some significant distinctions between classical probability theory and evidence theory.

Importantly, evidence theory differentiates between lack of belief or evidence and disbelief (Shafer 1987). That is, evidence theory explicitly recognizes ignorance. Secondly, partial information in evidence theory can be expressed on sets, rather than information defined on single elements (i.e., singletons) (Yager and Liu 2008). In this case, evidence can be allocated to a set of “events,” without specifying how the evidence is distributed to the individual elements themselves (Dubois and Prade 1988). These concepts present a versatile framework for treating uncertainty within the realm of evidence theory.

Evidence Theory: Mathematical Overview

Belief Measures

A belief measure, denoted $\text{Bel}(A)$, expresses the degree of support or evidence (i.e., a degree of belief) that a particular element belongs to the set A (Klir and Yuan 1995). Formally, a belief measure is a function defined by

$$\text{Bel} : P(X) \rightarrow [0, 1] \quad (22)$$

In the context of evidence theory, the universal set X is commonly referred to as a *frame of discernment* or more simply as a *frame* (Shafer 1976).

Belief functions are based on the following axioms (Klir 2006):

$$(Bel1) \text{ Bel}(\emptyset) = 0$$

$$(Bel2) \text{ Bel}(X) = 1$$

$$(Bel3) \text{ Bel}(A_1 \cup A_2 \cup \dots \cup A_n) \geq \sum_i \text{ Bel}(A_i) - \sum_{i < j} \text{ Bel}(A_i \cap A_j) + \dots + (-1)^{n+1} \text{ Bel}(A_1 \cap A_2 \cap \dots \cap A_n) \quad (23)$$

for all integers, n , and all possible subsets $A_1, A_2, \dots, A_n$, of x .

Like possibility theory, belief measures replace additivity with a weaker requirement. The inequality in Eq. 23, (Bel3), requires that the union of the sets be greater than or equal to the sum of the degrees of belief of each individual set (Klir 2006). When $A_1, A_2, \dots, A_n$ are pairwise disjoint, ($A_i \cap A_j = \emptyset$) and (Bel3) becomes

$$(Bel3) \text{ Bel}(A_1 \cup A_2 \cup \dots \cup A_n) \geq \text{ Bel}(A_1) + \text{ Bel}(A_2) + \dots \text{ Bel}(A_n) \quad (24)$$

Rewritten for two sets, A and B , where $A \cap B = \emptyset$:

$$(Bel3) \text{ Bel}(A \cup B) \geq \text{ Bel}(A) + \text{ Bel}(B) \quad (25)$$

The axioms in Eq. 23 are motivated by the rule for combining evidence in evidence theory, referred to as *Dempster's rule of combination* (Shafer 1976). This rule is formally introduced later in section “[Dempster's Rule of Combination](#).”

Plausibility Measures

As mentioned previously, evidence theory is based on dual pairs of belief and plausibility measures. Plausibility measures can be defined through evidence expressed in the form of a belief measure or they can be defined independently. Defined in the context of a belief measure, plausibility measures, denoted $Pl(A)$, express the amount of belief not committed to $\bar{A}$ (Yager and Liu 2008):

$$Pl(A) = 1 - \text{ Bel}(\bar{A}) \quad (26)$$

Equation 26 can be understood as follows: for beliefs not committed to $\bar{A}$ (the contrary of A), evidence theory does not automatically commit evidence to A . It recognizes, however, that A may be more plausible from this evidence (Yager and Liu 2008). Summarized by Shafer (1987), the quantities denoted by $\text{ Bel}(A)$ and $Pl(A)$ express respectively, “our judgment of the extent to which the evidence we are considering supports A and our judgment of the extent to which this evidence falls short of refuting A ” (Shafer 1987, p. 62). Similar to Eq. 26, as dual measures, belief measures can be related to plausibility measures via the following equation (Klir 2006):

$$\text{ Bel}(A) = 1 - Pl(\bar{A}) \quad (27)$$

Alternatively, a plausibility measure can be defined independent of one's belief by a functional mapping on the unit interval given by

$$Pl : P(X) \rightarrow [0, 1] \quad (28)$$

The value assigned to $Pl(A)$ characterizes total evidence that is possible or “plausible.” Notice in Eqs. 22 and 28 that belief and plausibility measures are defined on the power set $P(X)$. They therefore can represent evidence for a collection of elements.

Due to the inequality of the third axiom (Bel3), it can be shown that possibility and necessity measures are characterized by (Klir 2006)

$$\begin{aligned} \text{ Bel}(A) + \text{ Bel}(\bar{A}) &\leq 1 \\ Pl(A) + Pl(\bar{A}) &\geq 1 \end{aligned} \quad (29)$$

Equation 29 describes the possibility that the belief measure for some set A combined with the belief measure of $\bar{A}$ may be less than 1 (unity). Conversely, plausibility measures are characterized by evidence for A and $\bar{A}$ which, when combined, may be greater than 1(unity).

Equations 26 and 29 yield the following relationship between beliefs and plausibilities (Klir and Yuan 1995):

$$Pl(A) \geq Bel(A) \quad (30)$$

This relationship expresses the idea that a measure of plausibility for the set A is at least as great as the belief committed to the set (Ross 2010). Mathematically, the plausibility measure $Pl(A)$ represents not only the evidence represented by the belief $Bel(A)$ but also the evidence associated with any sets that overlap with A . Hence, at a minimum, the plausibility will be as strong as indicated by a belief (Yager and Liu 2008).

As demonstrated by Eq. 27, a degree of belief in or evidential support of A , $Bel(A)$, does not implicate disbelief of A – expressed by $Bel(\bar{A})$ (Shafer 1976). Conversely, Eq. 26 shows that a positive value of $Pl(A)$ does not necessarily imply the belief that there is evidence in support of A . For this reason, evidence theory differs from classical probability theory in that it provides a natural framework for modeling ignorance (Shafer 1976).

If there is limited evidence available for A and no evidence available for $\bar{A}$, then the amount of ignorance is large (Ross 2010). Total ignorance, the special case where $Bel(A) = 0$ and $Pl(A) = 1$, is characterized by evidence that neither supports nor refutes A (Shafer 1987). *Vacuous belief functions* are those belief functions defined by complete ignorance (Shafer 1976). Conversely, in the special case of complete evidence (zero ignorance) such that the sum of the belief and the complement of the belief equal unity, the belief measure is a probability measure.

Basic Evidence Assignments

Information within the context of evidence theory is often expressed in terms of a *basic evidence assignment*, denoted by $m(A)$ (Ross 2010). Other literature have used the terms basic probability assignments (Klir 2006), basic probability numbers (or probability masses) (Shafer 1976; Yager and Liu 2008), or basic belief assignments (or basic belief masses) (Smets 1994). The quantity, $m(A)$, represents the degree of belief that a particular element x of the universe X belongs to the set A , but does not include any evidence assigned to the subsets of A (Ross 2010). In other words, the basic evidence assignment, $m(A)$,

expresses a degree of evidence in support of exactly A . Every set in which $m(A) > 0$, where $A \in P(X)$, is called a *focal element*, so named because the evidence focuses on these sets or the elements contained within them (Klir and Yuan 1995). In other words, any element of $P(X)$ supported by evidence is a focal element. If there is evidence that the element belongs to a subset of A , say B (i.e., $B \subset A$), this belief must be represented by another quantity $m(B)$ (Klir 2006).

Evidence expressed in the form of basic evidence assignments can be combined to form belief and plausibility measures using the following respective formulas (Klir and Yuan 1995):

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (31)$$

$$Pl(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad (32)$$

Equations 31 and 32 effectively aggregate individual “pieces” of evidence expressed in terms of basic evidence assignments (Klir 2006). $m(A)$ and $Bel(A)$ in Eq. 31 may be related as follows: $m(A)$ expresses the amount of evidence which supports the proposition that the element in question belongs to the set A alone. A belief measure, $Bel(A)$, meanwhile is an expression of the evidence that supports A or any of the subsets of A (Klir and Yuan 1995). The quantity $Pl(A)$ is the degree of evidence that the element in question belongs to the set A , to any subset of A , or to any set that overlaps (or intersects) with A (Klir and Yuan 1995).

Dempster's Rule of Combination

Another facet of evidence theory is the ability to combine information from multiple sources (Yager and Liu 2008). The standard method for combining information from independent sources (i.e., independent bodies of evidence) in evidence theory is given by (Shafer 1976)

$$m_{1,2}(A) = \frac{\sum_{B \cap C = A} m_1(B) \cdot m_2(C)}{1 - K} \quad (33)$$

where

$$K = \sum_{B \cap C \neq \emptyset} m_1(B) \cdot m_2(C) \quad (34)$$

Equation 33 is most commonly referred to as *Dempster's rule of combination*. In this expression, A is a focal element, and m_1 and m_2 represent independent bodies of evidence in the form of basic evidence assignments. The value $m_{1,2}$ can be thought of as a joint message or a *joint evidence assignment* of the two pieces of evidence (Shafer 1987; Klir and Yuan 1995).

Dempster's rule of combination can be conceptualized as a ratio of two quantities of evidence. In the numerator, it is the evidence that supports or is in agreement with the hypothesis A. In the denominator, it is the total evidence that does not conflict among the sets, where conflict is quantified by null intersections of A and B (i.e., disjoint focal elements). As suggested by Shafer (1976), empty intersections can be thought of as information that has no commonality or is conflicting (Shafer 1976).

Special Cases of Evidence Theory

At a fundamental level, belief functions are expressions of subjective degrees of belief (Shafer 1976). In this regard, subjective probabilities can be viewed as a special case of belief measures defined on singletons. When all available evidence is defined only on individual elements, belief and plausibility measures become probability measures (proof in Klir 2006). The evidence may therefore be described probabilistically (Ross 2010). Shafer (1976) refers to this special class of belief functions as *Bayesian Belief Functions*.

In the context of possibility theory, when a body of evidence is nested, that is, a body of evidence is defined by $A_1 \subset A_2 \subset A_3 \subset \dots \subset A_n$, belief and plausibility measures become necessity and possibility measures (Klir 2006). Accordingly, possibility measures are consonant plausibility functions, a special case of belief and plausibility measures (Shafer 1987). For nested bodies of evidence, it can be shown that

plausibilities reduce to the axioms of possibility measures (Klir and Yuan 1995):

$$Pl(A \cup B) = \max[Pl(A), Pl(B)] \text{ for all } A, B \in P(X) \quad (35)$$

Equation 35 shows the close connections between belief functions and possibility measures (Shafer 1987).

Evidence Theory: Interpretations

Evidence theory allows for the imprecise representation of data. Evidence can be assigned to a set of alternatives without specifying how the evidence is distributed to individual elements. This introduces a mathematical alternative for manipulating subjective information that can distinguish between lack of belief and disbelief (Shafer 1976). Unlike probability theory, evidence theory does not presuppose beliefs about the negation of an event through beliefs assigned to a given proposition. Hence, evidence theory can capture ignorance explicitly (Ross 2010). Shafer (1976) suggests this is of particular importance in the case of a set of possibilities that contains several events. As shown, special cases of belief and plausibility measures become probability and possibility/necessity measures. Evidence theory, however, is also applicable independent of other theories (Smets 1994; Yager and Liu 2008).

Within evidence theory, Dempster's rule of combination presents a unique mathematical framework for combining evidence from distinct sources (Smets 1994). Dempster's rule of combination intuitively combines evidence so that the combination of two beliefs – both supporting A – yields more support than either piece of evidence alone (Shafer 1987).

Types of Uncertainty

Sections “Probability Theory,” “Possibility Theory,” and “Evidence Theory” demonstrate that the formulations of other uncertainty theories were motivated by acknowledgment of different

forms of uncertainty. From these theories, distinct types of uncertainty commonly termed *conflict* and *nonspecificity* are identified (Klir and Yuan 1995).

Conflict is a form of uncertainty associated with classical probability theory. Probability theory expresses evidence in the form of a probability distribution, which distributes evidential assignments over multiple values. The probability assigned to any given element, however, is evidence that “conflicts” with the evidence assigned to other elements contained in the distribution (Klir and Smith 2001). In other words, evidence defined for some element x_i in the form of a probability distribution conflicts with evidence that supports other elements, the probabilities assigned to other elements, such as x_j and x_k . Similarly, the probability in support of the elements x_j and x_k indicates less support for the element x_i . These elements are in competition with one another so to speak, and hence, the form of uncertainty characterized by classical probability theory is commonly referred to as *conflict* (Klir and Smith 2001).

Possibility theory possesses a type of uncertainty called *nonspecificity* (Klir and Smith 2001). This uncertainty is concerned with information contained in sets of possible alternatives (predictions, diagnoses, etc.), “unspecific” to any particular alternative. In this regard, large sets are less specific (greater nonspecificity), while small sets are more specific (less nonspecificity). Because evidence is nested in possibility theory, it does not conflict (Klir and Smith 2001; Klir 2006). Possibility theory is therefore free of conflict, while probability theory is free of nonspecificity. As discussed in Section “Evidence Theory,” evidence theory has direct mathematical connections to both probability theory and possibility theory. As such, evidence theory can express both types of uncertainty (Klir and Smith 2001).

Summary

Broadly, different uncertainty theories expand the mathematical framework of conventional

methods and tools characterized by probability theory. There are mathematical and interpretational connections among the theories, but each theory differs fundamentally in its modeling of uncertain situations. Each theory is distinguished by its mathematical properties or axioms. From this basis, different uncertainty theories offer varying mathematical properties in the treatment of uncertainty, including (i) information at the set and subset levels, and (ii) incorporation of the other notions of uncertainty (e.g., ignorance) (Klir 2006).

There are multiple well-established interpretations of the various uncertainty theories. Instinctively, dual pairs of possibility/necessity and belief/plausibility measures are a mechanism for representing imprecise probabilities. This may be well suited for uncertainty that arises from partial “evidence” or a lack of information. This is separate from the uncertainty that may result from a random process or information modeled as purely probabilistic in nature (Smets 1994).

Cross-References

- ▶ Analytic Fragility and Limit States [P(EDPI IM)]: Nonlinear Dynamic Procedures
- ▶ Analytic Fragility and Limit States [P(EDPI IM)]: Nonlinear Static Procedures
- ▶ Bayesian Statistics: Applications to Earthquake Engineering
- ▶ Empirical Fragility
- ▶ Model-Form Uncertainty Quantification for Structural Design
- ▶ Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers
- ▶ Probabilistic Seismic Hazard Models
- ▶ Seismic Loss Assessment
- ▶ Seismic Risk Assessment, Cascading Effects
- ▶ Site Response for Seismic Hazard Assessment
- ▶ Structural Reliability Estimation for Seismic Loading
- ▶ Structural Seismic Reliability Analysis
- ▶ Uncertainty Quantification in Structural Health Monitoring

References

- Ang AH-S, Tang WH (2007) Probability concepts in engineering emphasis on applications to civil and environmental engineering, 2nd edn. Wiley, Hoboken
- Applied Technology Council (1985) Earthquake damage evaluation data for California (ATC 13). Applied Technology Council (ATC), Redwood City
- Benjamin JR, Cornell CA (1970) Probability, statistics, and decision for civil engineers. McGraw-Hill, New York
- Cornell AC (1968) Engineering seismic risk analysis. *Bull Seismol Soc Am* 58(5):1583–1606
- Cornell AC (1969) A probability-based structural code. *ACI J* 66:974–985
- Cornell AC, Krawinkler H (2000) Progress and challenges in seismic performance assessment. *PEER Center News*, vol 3, no. 2
- Dempster AP (1967) Upper and lower probabilities induced by a multivalued mapping. *Ann Math Stat* 38(2):325–359
- Dubois D, Prade H (1988) Possibility theory: an approach to computerized processing of uncertainty. Plenum, New York
- Ellingwood BR, Galambos TV, MacGregor JG, Cornell CA (1980) Development of a probability based load criterion for American National Standard A58. National Bureau of Standards, U.S. Department of Commerce, Washington, DC
- Federal Emergency Management Agency (FEMA) (1997) FEMA 273, NEHRP guidelines for the seismic rehabilitation of buildings. Federal Emergency Management Agency, Washington, DC
- Federal Emergency Management Agency (FEMA) (2000) FEMA 356, prestandard and commentary for the seismic rehabilitation of buildings. Federal Emergency Management Agency, Washington, DC
- Hacking I (2006) The emergence of probability: a philosophical study of early ideas about probability intuition and statistical inference, Second edn. Cambridge University Press, Cambridge, UK
- Klir GJ (2006) Uncertainty and information foundations of generalized information theory. Wiley, Hoboken
- Klir GJ, Smith RM (2001) On measuring uncertainty and uncertainty-based information: recent developments. *Ann Math Artif Intell* 32(1–4):5–33
- Klir GJ, Yuan B (1995) Fuzzy sets and fuzzy logic theory and applications. Prentice Hall, Upper Saddle River
- National Institute of Building Sciences (1997) HAZUS earthquake loss estimation methodology. NIBS Document No. 5201. Federal Emergency Management Agency, Washington, DC
- Porter K (2003) An overview of PEER's performance-based earthquake engineering methodology. Ninth international conference of applications of statistics and probability in civil engineering (ICASP9), San Francisco, 6–9 July 2003
- Ravindra MK, Galambos TV (1978) Load and resistance factor design for steel. *J Struct Div Am Soc Civil Eng (ASCE)* 104:1337–1353
- Ross TJ (2010) Fuzzy logic with engineering applications, 3rd edn. Wiley, West Sussex
- Shafer G (1976) A mathematical theory of evidence. Princeton University Press, Princeton
- Shafer G (1987) Belief functions and possibility measures. In: Bezdek JC (ed) *Analysis of fuzzy information*, vol 1, Mathematics and logic. CRC Press, Boca Raton
- Shafer G (2012) A mathematical theory of evidence. Glenn Shafer, <http://www.glennshafer.com/books/amte.html>. 7 Nov 2012
- Smets P (1994) What is Dempster-Shafer's model? *Advances in the Dempster-Shafer theory of evidence*. Wiley, New York, pp 5–34
- Vick SG (2002) Degrees of belief subjective probability and engineering judgment. American Society of Civil Engineers (ASCE) Press, Reston
- Whitman RV (1973) Damage probability matrices for prototype buildings, Massachusetts Institute of Technology Department of Civil Engineering, Research Report R73–57. MIT Press, Cambridge, MA
- Whitman RV, Hong S, Reed JW (1973) Damage Statistics for High-Rise Buildings in the Vicinity of the San Fernando Earthquake, Massachusetts Institute of Technology Department of Civil Engineering, Research Report R73–24. MIT Press, Cambridge, MA
- Yager RR, Liu L (2008) In: Kacprzyk J (ed) *Classic Works of the Dempster-Shafer Theory of Belief Functions: An Introduction*, vol 219, *Studies in fuzziness and soft computing*. Springer, New York
- Zadeh LA (1978) Fuzzy sets as a basis for a theory of possibility. *Fuzzy Set Syst* 1:3–28
- Zadeh LA (1986) A simple view of the Dempster-Shafer theory of evidence and its implication for the rule of combination. *AI Mag* 7(2):85–90

Urban Change Monitoring: Multi-temporal SAR Images

Fumio Yamazaki and Wen Liu
Department of Urban Environment Systems,
Chiba University, Chiba, Japan

Synonyms

Building; Change detection; Multi-temporal images; Remote sensing; SAR; Satellite; Urban area

Introduction

Urban areas grow and change rapidly in various parts of the world, and hence, monitoring and modeling of urban areas have become increasingly important. Regular and up-to-date information on urban changes is required for city planning, disaster management, environmental impact assessment, and appropriate allocation of services and infrastructures. Current approaches to urban monitoring generally include field surveys and interpretation of aerial photographs. Moreover, the recent advancements in remote sensing technology have made satellite imagery an efficient tool for collecting urban area's information. Compared with optical sensors, synthetic aperture radar (SAR) does not suffer from the limitation of sunlight and weather conditions. In addition, side-looking observation scheme of SAR is useful to assess the heights and exterior walls conditions of buildings and other structures. SAR images obtained from very-high-resolution (VHR) satellites, e.g., TerraSAR-X and COSMO-SkyMed, are especially useful for urban modeling and change detection as well as for damage mapping after disasters. These VHR SAR satellites are only recent vintages, but the imagery acquired has been used in various fields. In this entry, a basic method of urban change monitoring using multi-temporal SAR imagery data is introduced and several examples are presented.

Method of Change Detection from Multi-temporal SAR Intensity Images

SAR is one of the most powerful tools for monitoring changes in the earth's surface. SAR systems are capable of recording both the amplitude and phase of backscattering echoes from the objects on the earth's surface. SAR interferometric analyses using phase information successfully provided relative ground displacements due to earthquakes (Massonnet et al. 1993) and due to land subsidence in normal time (Bayuaji et al. 2010). Although SAR interferometry has

high sensitivity for gradual changes such as land subsidence and volcano's inflation, it is difficult to apply to sudden changes in urban areas such as new construction and demolition of buildings. Thus, SAR intensity (amplitude) data are often used to assess changes in urban built environment with low coherence (Yonezawa and Takeuchi 2001; Matsuoka and Yamazaki 2004).

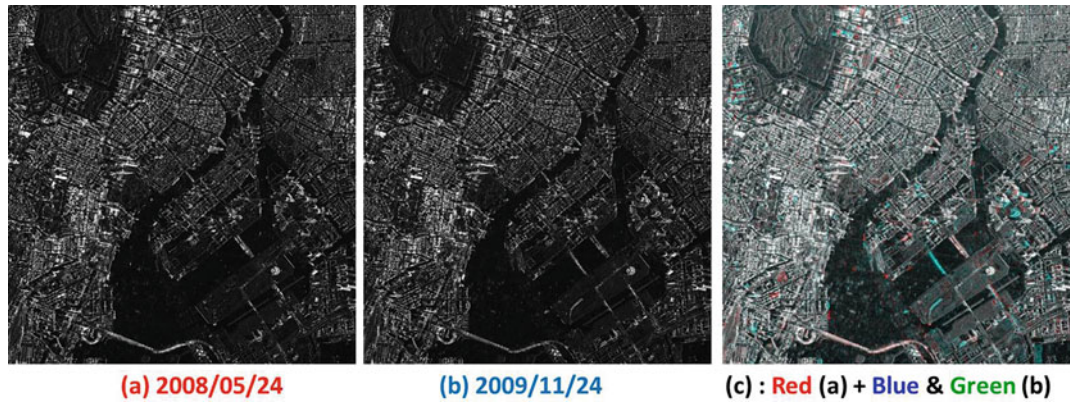
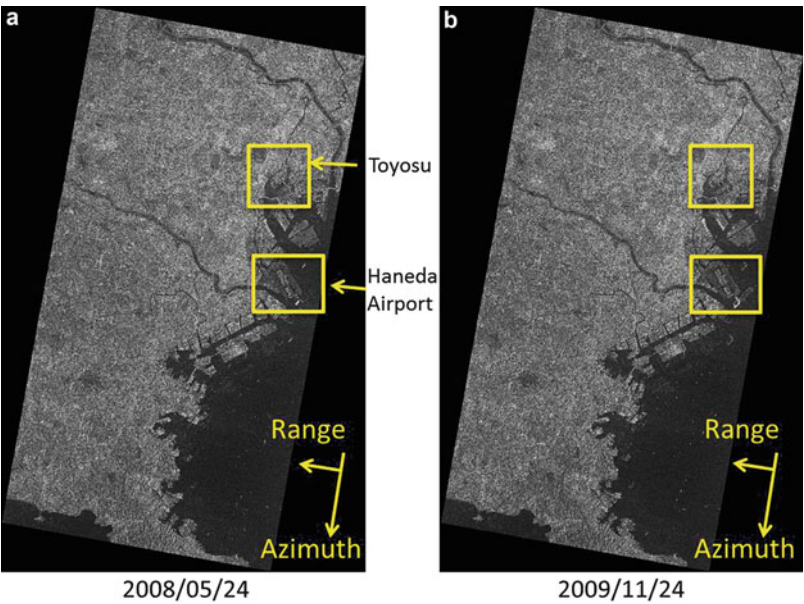
SAR intensity data acquired and processed in digital number (DN) are often converted to a physical quantity called the backscattering coefficient (σ_0) measured in decibel (dB) units. After this conversion, an adaptive filter (e.g., Lee 1980) is usually applied to original SAR images to reduce speckle noise, which makes the radiometric and textural aspects less efficient, and to improve the correlation coefficient between the two images.

Figure 1 shows an example of two-temporal TerraSAR-X backscattering coefficient images covering Tokyo, Japan. Both images were taken by HH polarization and from the descending path with 42.82° incidence angle at the center. Since they were taken in the stripmap mode, the azimuth resolution was about 3.3 m and the ground range resolution was about 1.2 m. After the enhanced ellipsoid correction (EEC), the two images were transformed into 1.25 m/pixel.

If two-temporal SAR intensity images for the same location in the same radar incidence conditions exist, the simplest comparison is the color composite of the two images. The color composite approach can be used for up to three-temporal images by assigning red (R), green (G), and blue (B) colors for the three images, respectively. If the color composite shows a gray color (between white and black), it means no change in backscatter among them. On the contrary, if the color composite exhibits a dominant color, it indicates a stronger backscatter for that colored SAR image. By this method, changes in buildings and structures in urban areas can be visually observed.

Figure 2 shows a close-up of the TerraSAR-X images of Toyosu district, shown in the upper square in Fig. 1. In the figure, cyan (blue + green)

Urban Change Monitoring: Multi-temporal SAR Images, Fig. 1 Two-temporal TerraSAR-X backscattering coefficient images acquired on (a) 24 May 2008 and (b) 24 November 2009, covering Tokyo, Japan. They were taken by HH polarization in stripmap mode, from the descending path



Urban Change Monitoring: Multi-temporal SAR Images, Fig. 2 Close-up of two TerraSAR-X images for Toyosu district in Fig. 1 and their color composite. Red color is assigned for the image (a) 24 May 2008 and

blue and green colors for the image (b) 24 November 2009 in the color composite image (c). Cyan (blue + green) color in (c) shows new structures and red demolished ones during this period

color shows new structures, such as bridges and buildings, and red color shows demolished ones during the 1.5 years period.

The change detection from two-temporal SAR intensity images can be evaluated

quantitatively by their difference value, d , and the correlation coefficient, r , by the following equations:

$$d = \bar{I}a_i - \bar{I}b_i \quad (1)$$

$$r = \frac{N \sum_{i=1}^N I a_i I b_i - \sum_{i=1}^N I a_i \sum_{i=1}^N I b_i}{\sqrt{\left(N \sum_{i=1}^N I a_i^2 - \left(\sum_{i=1}^N I a_i \right)^2 \right) \cdot \left(N \sum_{i=1}^N I b_i^2 - \left(\sum_{i=1}^N I b_i \right)^2 \right)}} \quad (2)$$

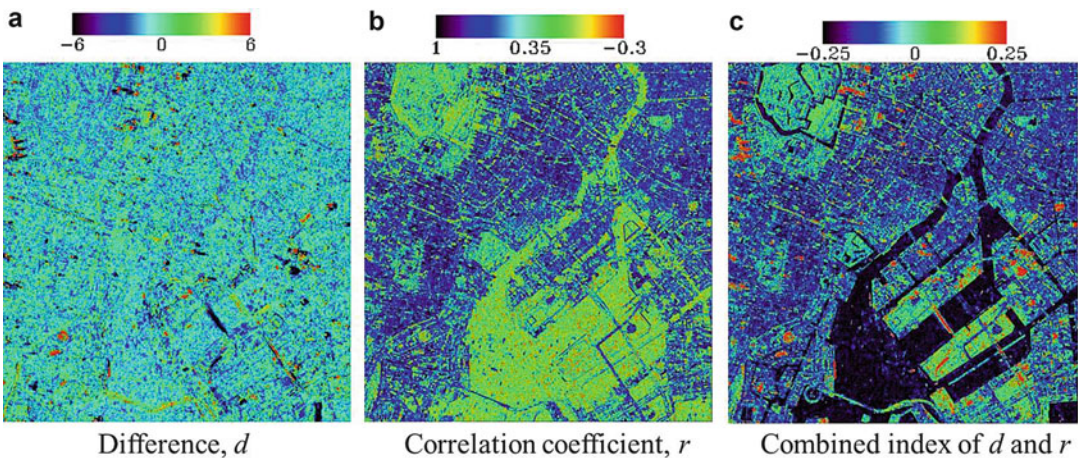
where i is the pixel number, $I a_i$ and $I b_i$ are the backscattering coefficient of the second and first images, and $\bar{I a}_i$ and $\bar{I b}_i$ are the corresponding averaged values over the $N (=k \times k)$ pixels window surrounding the i -th pixel. The window size, N , may be assigned by considering the spatial resolution of the SAR images and the size of target objects to extract.

Figure 3 plots the difference and correlation coefficient images corresponding to the SAR data in Fig. 2. Change detection in a period is usually carried out using one of these two indices or using a combined index of the two. For example, Matsuoka and Yamazaki (2004, 2010) proposed a linear combination of the two indices as an empirical discriminant formula between change and no change. A similar combined index, z , was also proposed by Liu and Yamazaki (2011) as

$$z = \frac{|d|}{\max(|d|)} - c r \quad (3)$$

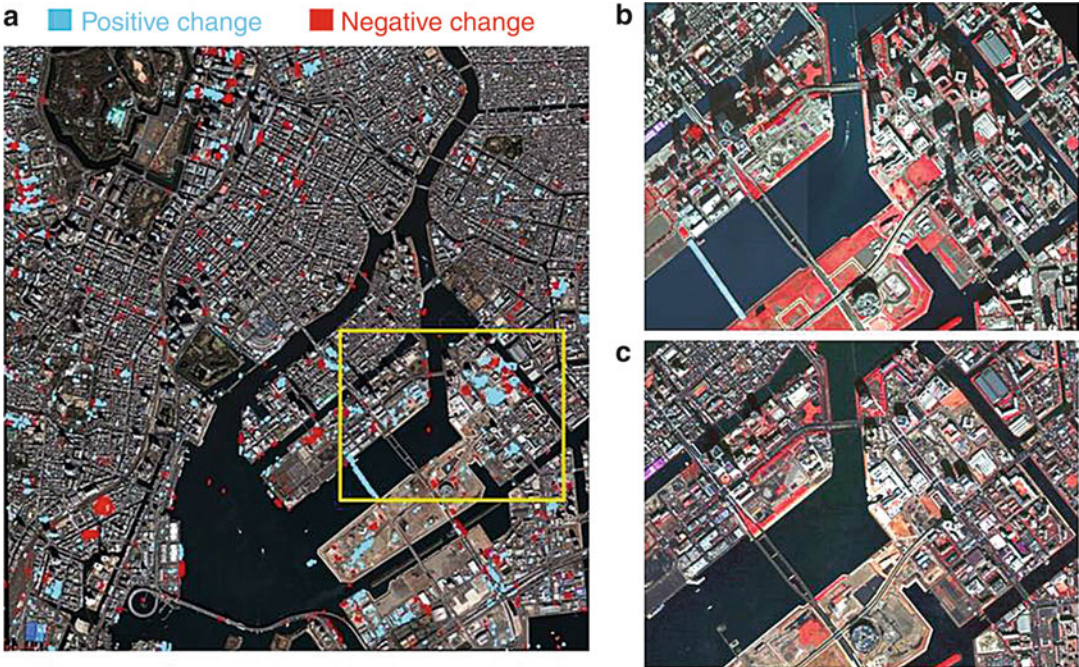
where $\max(|d|)$ is the maximum absolute value in the difference and c is a weighting constant of the two indices. The correlation coefficient is generally sensitive to subtle changes, and thus, it shows a low value even for a small change. On the contrary, the normalized absolute value of difference is relatively stable. Hence, in this example, the weight for difference was set as four times of that for correlation. An example of the combined index is shown in Fig. 3c. Using this plot, the threshold value for judging change or no change can be determined by comparing with optical satellite or airborne images or GIS data.

By setting the threshold as $z > \mu + 2\sigma$, with the mean value μ and standard deviation σ , Fig. 4



Urban Change Monitoring: Multi-temporal SAR Images, Fig. 3 Difference (a) and correlation coefficient (b) images corresponding to the SAR images in Fig. 2. The window size of 9×9 was used in calculation.

Combined index, z , in (c) was calculated by setting $c = 1/4$, and thus, it has the range between -0.25 to 1.25 . Possible changed areas are shown in red color (c), indicating new construction or demolition



Urban Change Monitoring: Multi-temporal SAR Images, Fig. 4 The extracted changed areas overlapping on the QuickBird true-color image on 20 May 2008 (a), showing new construction (*positive change*) and

demolition (*negative change*). Comparison of false-color QuickBird image on 20 May 2008 (b) and WorldView-2 image on 11 May 2010 (c) for the *yellow square* in (a)

shows the extracted changed areas overlapping on a QuickBird true-color image, where new construction (positive change of d) and demolition (negative change of d) are highlighted in cyan and red colors. Comparing with close-up high-resolution optical images of similar acquisition dates, newly constructed bridges and buildings can be confirmed. However, negative changes of backscatter were observed not only for demolished building areas but also for areas which were casted by radar shadow of new buildings. Hence, the extraction of change in SAR backscatter in dense urban areas is rather complex. For more accurate evaluation of radar shadow and layover, a 3D GIS urban model is necessary.

Example for Airport Runway Construction

As another example of urban change monitoring, the progress of the expansion construction in the

Tokyo Haneda International Airport, Japan, is considered. To satisfy increasing demand of domestic and international passengers, the fourth runway, named the D-runway, was constructed in the Haneda Airport on the Tokyo Bay. Figure 5 shows aerial photographs of the D-runway's construction site, showing the stage of construction. The new D-runway is composed of two parts: floating steel-jacket platforms (lower left in each photo) and land reclamation (lower right in each photo).

Three-temporal TerraSAR-X images taken in different dates in 2008 and 2009 are shown in Fig. 6 to monitor the D-runway's construction stage in the Haneda Airport. In these images, the construction stages of the new runway by combining steel-jacket platforms and land reclamation can be clearly observed. The color composite of the three SAR images highlights new constructions, such as the D-runway, the connection bridge, and the new international terminal building. Since the new runway was built on the



Urban Change Monitoring: Multi-temporal SAR Images, Fig. 5 Aerial photographs of the D-runway's construction site in the Tokyo Haneda International airport. In the photos of four dates, the lower left is floating steel-jacket platforms and the lower right is reclaimed

land. These two parts constitute the new D-runway. This new runway is connected to the existing part of the airport by the connection bridge (Haneda Airport Construction Office 2010)

Tokyo Bay, boats used for construction are seen in different locations for each image.

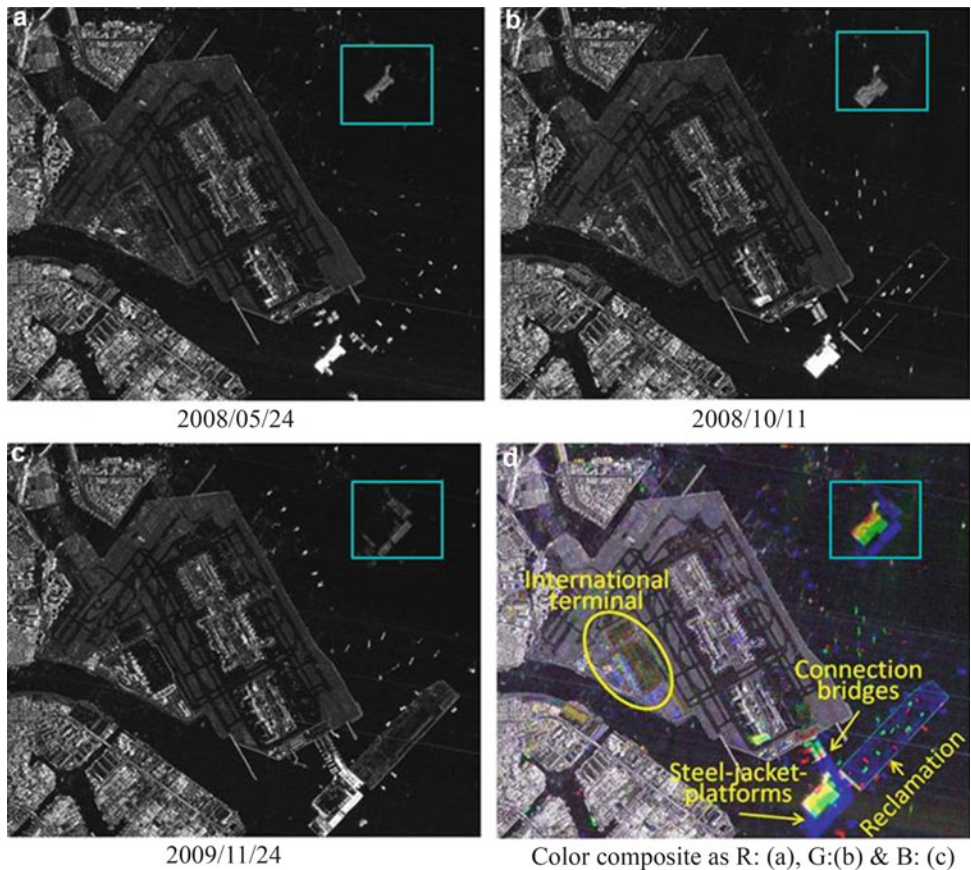
At the right upper part of the three TSX images, a shape that is similar to the steel-jacket platforms, a part of the D-runway, could be seen. This phenomenon, known as “ambiguity,” appears as the ground objects duplication on image along range or azimuth direction (Li and Johnson 1983). Such phenomena often appear and are caused by overlapping return signals from one main pulse with multiple Doppler frequencies. Similar azimuth ambiguities were presented by Brusch et al. (2011) in automatic ship detection using TerraSAR-X images.

The difference between the first (24 May 2008) and second (11 October 2008) images and their correlation coefficient were calculated as shown in Fig. 7. The change factor, z , defined

above was also calculated and shown in the figure. Using the histogram of the factor z , the positive and negative changed areas were extracted. Compared with the aerial photographs, the extracted results were found to be reasonable.

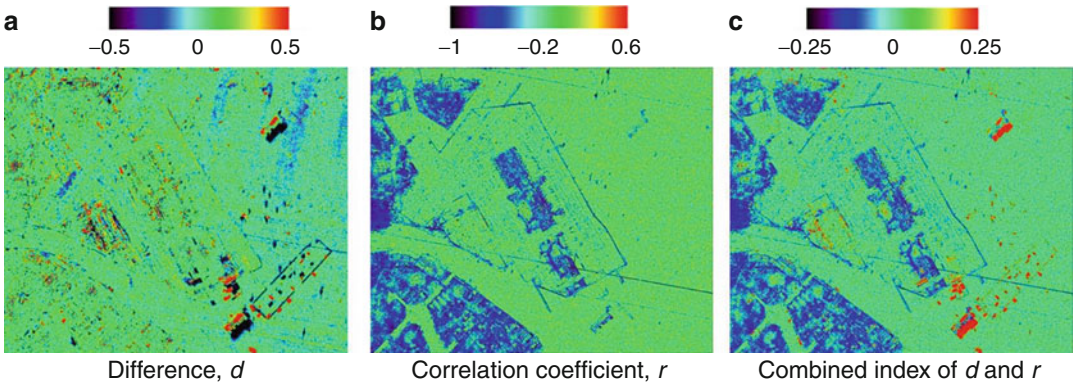
Example for Tsunami Damage to Building Exterior Walls

The Mw 9.0 Tohoku earthquake occurred on 11 March 2011, off the Pacific coast of northeastern (Tohoku) Japan, caused gigantic tsunamis, resulting in widespread devastation. Typical building damage in the 2011 Tohoku earthquake tsunami is shown in Fig. 8. It is noted that damage is concentrated to building exterior walls or lower parts, which is not the same with damage caused



Urban Change Monitoring: Multi-temporal SAR Images, Fig. 6 Three-temporal TerraSAR-X images of construction site in the Tokyo Haneda airport. Color composite image (d) highlights the stage of new constructions.

Azimuth ambiguities, which duplicate the real steel-jacket platforms, can be observed in the *blue square* of each image in the sea



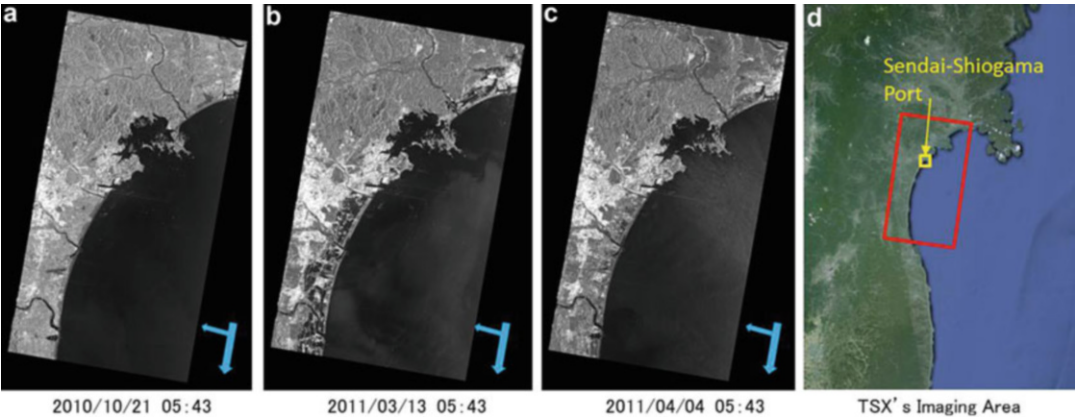
Urban Change Monitoring: Multi-temporal SAR Images, Fig. 7 Difference (a) and correlation coefficient (b) images corresponding to the first (24 May 2008) and second (11 October 2008) SAR images in Fig. 6.

Combined index z was calculated in (c), where high z values correspond to the new runway under construction and ships



Urban Change Monitoring: Multi-temporal SAR Images, Fig. 8 Typical tsunami damage to reinforced concrete building (*left*) and steel-frame building (*right*)

observed in the 2011 Tohoku, Japan, earthquake. Damage is concentrated to the exterior walls and lower stories of buildings (Yamazaki et al. 2013)



Urban Change Monitoring: Multi-temporal SAR Images, Fig. 9 The pre-event TSX image taken on 21 October 2010 (a), the post-event images taken on 13 March 2011 (b) and on 4 April 2011 (c), and the

TerraSAR-X's imaging area along the Pacific coast of Tohoku, Japan, including the Sendai-Shiogama Port area (d)

by strong earthquake shaking. This kind of damage to building exterior walls can be extracted with the aid of side-looking nature of SAR. The strong backscattering echoes from exterior walls and double bounce with the ground are often reduced after this type of structural damage.

SAR Data and Preprocessing

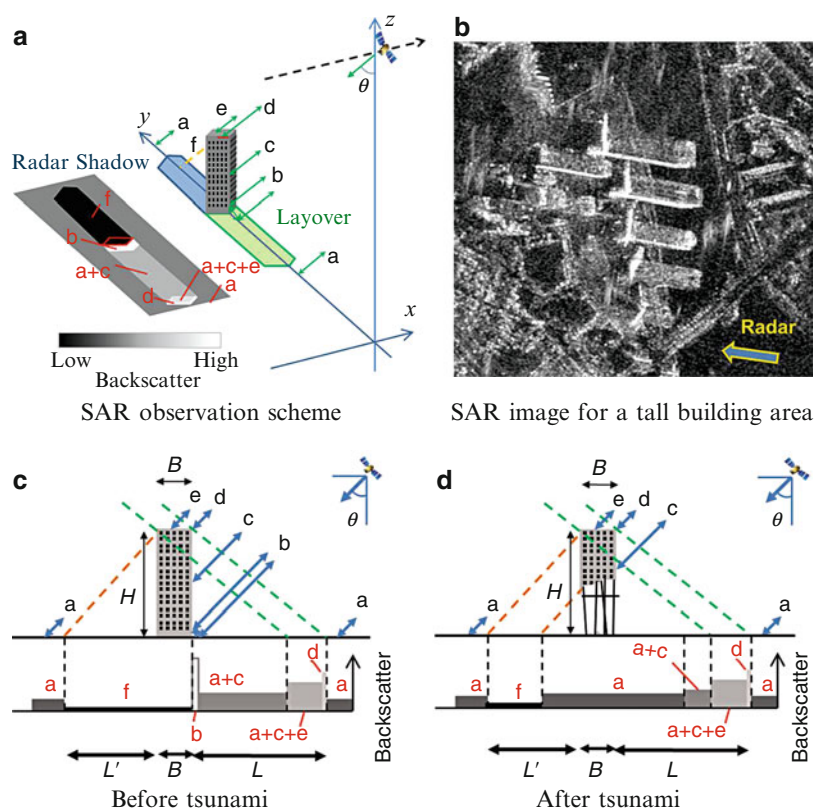
Three-temporal TerraSAR-X images taken before and after the earthquake are shown in Fig. 9. The pre-event image was taken on 21 October 2010 (local time) while the post-event ones were taken on 13 March 2011 (2 days after the

earthquake) and on 4 April 2011. The incident angle was 37.31° at the center of the images. The three images were captured with HH polarization from a descending path. The images were acquired in the stripmap mode with azimuth resolution about 3.3 m and ground range resolution about 1.2 m. The images were orthorectified multi-look corrected products (EEC) where the image distortion caused by variable terrain height was compensated by using a globally available DEM (SRTM). They have been resampled and projected to a WGS84 reference ellipsoid with a square pixel size of 1.25 m.

Two preprocessing were applied to the images. First, the three images were transformed

Urban Change Monitoring:
Multi-temporal SAR Images,

Fig. 10 Schematic figure of SAR observation for an ideal tall flat-roof building (Yamazaki et al. 2013). 3D layout (a), an actual TerraSAR-X intensity image for the central Tokyo (b), the amplitude of backscatter from an intact building (c), and the amplitude of backscatter from a damaged building due to tsunami (d)



to the backscattering coefficient (σ_0), which represents the radar reflectivity per unit area in the ground range. After the transformation, the values range between -30 and 30 dB. An enhanced Lee filter (Lopes et al. 1990) was then applied to the original SAR images to reduce the speckle noise. To minimize any loss of information included in the intensity images, the window size of the filter was set to 3×3 pixels.

It is known that large crustal movements (coseismic and post-seismic displacements) occurred in this area due to the Mw 9.0 event (Ozawa et al. 2011). The shift when the post-event TerraSAR-X images were taken compared with the pre-event one was estimated 3.11 m to the east and 0.55 m to the south in the Sendai area (Liu and Yamazaki 2013). Thus, the pre-event SAR image was manually shifted 2 pixels (2.5 m) to the east in order to match with the post-event SAR images.

SAR Backscatter Characteristics of Buildings

Backscattering echoes from buildings in VHR SAR images include the information on the three-dimensional shapes of buildings. Reconstruction of buildings from VHR SAR images were carried out by several researchers (e.g., Brunner et al. 2011). Due to the side-looking geometry of SAR sensors, the backscattering echoes of VHR SAR from a stand-alone building can be shown schematically in Fig. 10.

If a building stands alone on a flat ground, the backscatter from the building shows layover on the ground range. In a usual case, the fringe of the rooftop exhibits strong backscatter (blue arrow **d** in Fig. 10a, c) and it leans toward the direction of SAR illumination. The foot fringe of a building also exhibits strong backscatter (blue arrow **b**) due to the double bounce of radar from

the ground and the building's exterior wall. The backscatter from the building exterior wall (blue arrow **c**) together with the double-bounce component, the layover area corresponding to the building's exterior wall, is generally distinguishable in a ground range SAR intensity image. Thus, the height of the building, H , can be calculated from the length of layover area, L , as

$$H = L \tan \theta \quad (4)$$

where θ is the incidence angle of SAR. A similar geometrical relationship is obtained for a radar shadow area (area **f** behind the building corresponding to $L' + B$ with B = the building width). If the radar shadow is clearly observed as in Fig. 10b, the building height is also estimated by the length of radar shadow, L' , as

$$H = L' / \tan \theta. \quad (5)$$

In other words, if the height of a building is known, the areas of layover and radar shadow can be estimated on the ground range SAR intensity image. If the lower part of the building's exterior wall is damaged as represented in Fig. 10d, the area of layover due the exterior wall (**a** + **c**) will be reduced or the backscatter intensity of the area will get smaller. The area of radar shadow (**f**) will also become small for a damaged building.

Case Study for Sendai-Shiogama Port

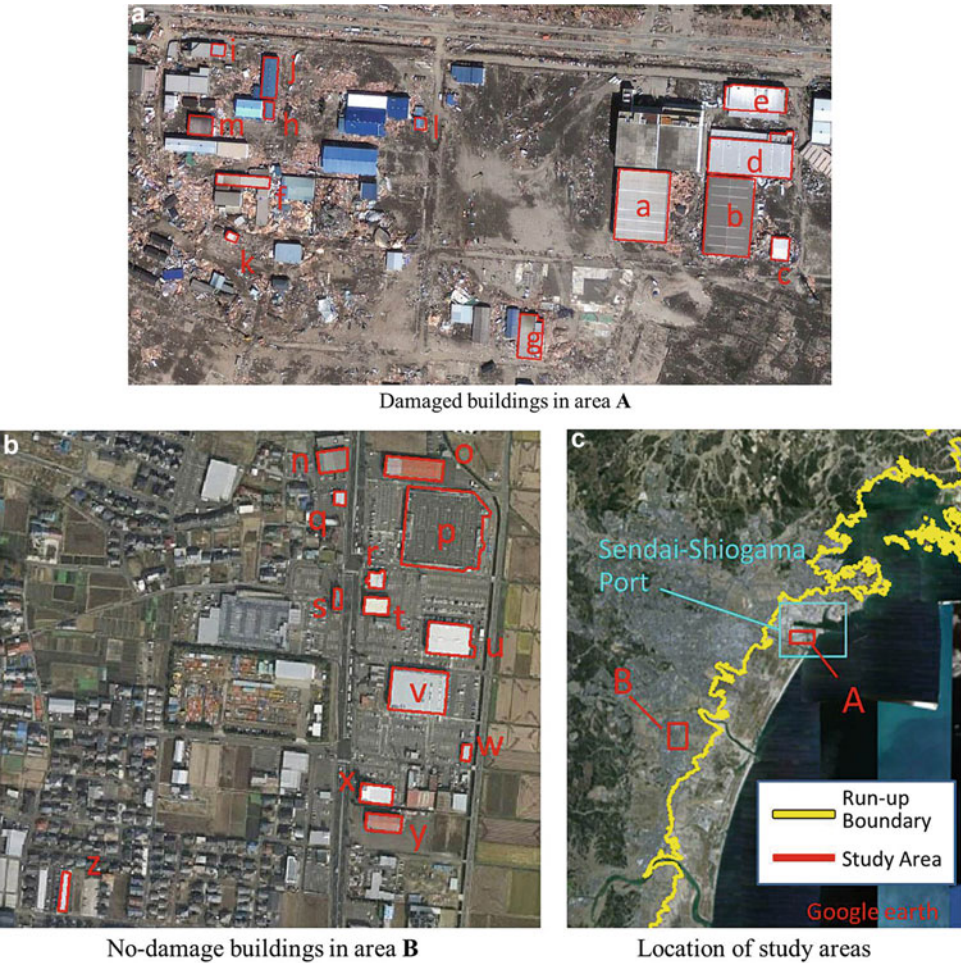
The Sendai-Shiogama Port, which was severely hit by the 2011 Tohoku earthquake tsunami, was selected as a case study site of building damage detection from SAR images since there were many large buildings and warehouses damaged by the tsunami attack. Figure 11 shows aerial images of the Sendai-Shiogama Port and its surroundings after the tsunami attack. Area **A** was located within the tsunami inundation zone and area **B** outside of the zone. Thirteen (13) damaged buildings were extracted from

area **A** and thirteen non-damaged ones were selected from area **B**.

In the study, building footprints were drawn manually using the optical images. Then the layover areas were determined by shifting the rooftop frames manually. In this stage, the building heights were estimated from the buildings' cast shadow lengths (Iwasaki and Yamazaki 2011) and the ground photos in Google Street View. The backscattering coefficients within the estimated layover area for each building were extracted for the pre- and post-event SAR images, and their values were used in change detection due to the tsunami attack.

The average value of the backscattering coefficient within the layover area was calculated for the 13 damaged and 13 non-damaged buildings for the three time instants. The difference of the average values of the backscattering coefficients within the layover area of each building before and after the tsunami attack is plotted in Fig. 12 for the 13 damaged buildings (a–m) in area **A** and 13 non-damaged buildings (n–z) in area **B**. It is observed that for the damaged buildings, the average backscattering coefficient within the layover area tends to be reduced in the post-tsunami SAR images while the average value is almost unchanged for the non-damaged buildings. The 13 March 2011 image was captured only two days after the tsunami attack, and hence, the effect of debris spread on the ground might raise the backscatter there and then canceled out the effect of exterior wall damage. In the 04 April 2011 image, some debris had been cleared, but vehicles and some other objects might exist in the layover area. Due to these possibilities, the difference values are not the same for the two time intervals.

Figure 13 compares the close-up of the TerraSAR-X images and the optical images of similar acquisition dates for a damaged building c. Building c suffered from severe tsunami attack on its exterior wall and the wall was lost almost completely. Due to this damage situation, the backscatter from the exterior wall as well as the double-bounce component reduced



Urban Change Monitoring: Multi-temporal SAR Images, Fig. 11 Aerial images of Sendai-Shiogama Port and its surroundings after the 2011 tsunami attack.

Area A within the tsunami inundation zone was severely affected and area B was located outside of the zone

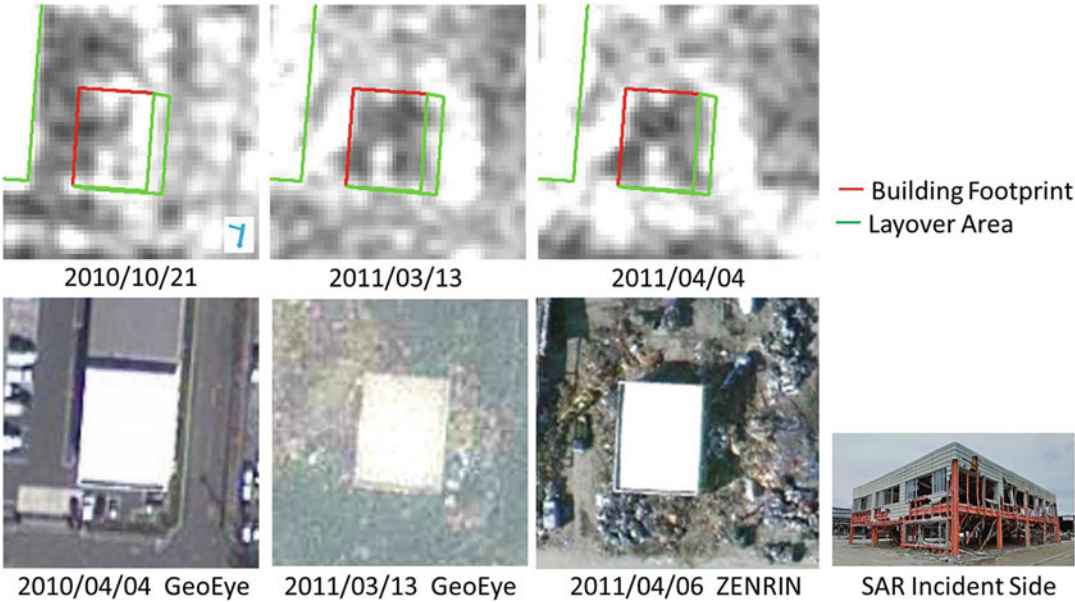
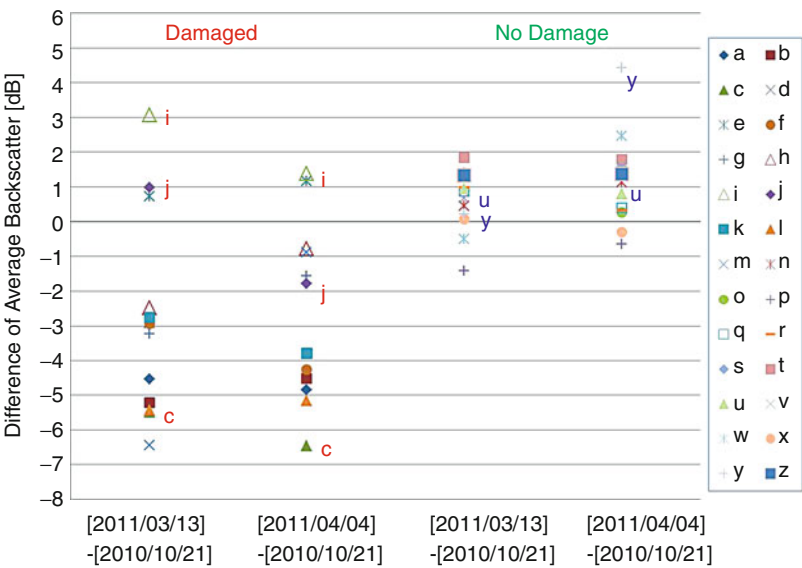
significantly. From the detailed observation of damaged and non-damaged buildings at their SAR reflection side faces, the backscatter coefficient in the layover area is considered to be a good indicator to judge the damage status of building exterior walls.

Summary

The use of multi-temporal synthetic aperture radar (SAR) imagery for urban change

monitoring is presented. Basically SAR does not suffer from the limitation of sunlight and weather conditions, and thus, it has wider acquisition opportunities than optical sensors. In addition, side-looking observation scheme of SAR is useful to acquire 3D configuration of buildings and other structures. In this entry, recent applications of very-high-resolution (VHR) SAR in urban change detection and damage assessment after natural disasters were presented. To access changes in urban areas, at least two SAR images from the same acquisition conditions are

Urban Change Monitoring: Multi-temporal SAR Images, Fig. 12 The difference of the average values of the backscattering coefficients within the layover area of each building before and after the tsunami attack for the 13 damaged buildings in area A and 13 non-damaged buildings in area B



Urban Change Monitoring: Multi-temporal SAR Images, Fig. 13 Close-up of the TerraSAR-X images at the three acquisition dates and the optical images of

similar acquisition times for damaged building c. The ground photo of SAR reflection side of the building extracted from Google Street View is also shown

necessary. The difference and correlation coefficient of the backscattering coefficients from multi-temporal SAR data are considered to be good indicators to extract changes on the earth

surface. The side-looking scheme of SAR is also useful in assessing the condition of the building's exterior wall, by comparing layover and radar shadow of multi-temporal images.

Cross-References

- [Earthquake Damage Assessment from Vhr Data: Case Studies](#)
- [Remote Sensing in Seismology: An Overview](#)
- [SAR Images, Interpretation of](#)
- [SAR Tomography for 3D Reconstruction and Monitoring](#)

References

- Bayuaji L, Sumantyo JTS, Kuze H (2010) ALOS PALSAR D-InSAR for land subsidence mapping in Jakarta, Indonesia. *Can J Remote Sens* 36(1):1–8
- Brunner D, Lemoine G, Bruzzone L, Greidanus H (2011) Building height retrieval from VHR SAR imagery based on an iterative simulation and matching technique. *IEEE Trans Geosci Remote Sens* 48(3):1487–1504
- Brusch S, Lehner S, Fritz T, Soccorsi M, Soloviev A, van Schie B (2011) Ship surveillance with TerraSAR-X. *IEEE Trans Geosci Remote Sens* 49:1092–1103
- Haneda Airport Construction Office (2010) D-runway news letter & report. http://www.pa.ktr.mlit.go.jp/haneda/haneda/haneda_saikaku/archives_d_run/newsletter/index.html. Accessed 13 July 2010
- Iwasaki Y, Yamazaki F (2011) Detection of building collapse from the shadow lengths in optical satellite images. In: *Proceedings of the 32nd Asian conference on remote sensing*, Taipei, Paper No. 202, CD-ROM, 6 p
- Lee JS (1980) Digital image enhancement and noise filtering by use of local statistics. *Trans Pattern Anal Mach Intell* 2:165–168
- Li FK, Johnson WTK (1983) Ambiguities in spaceborne synthetic aperture radar systems. *IEEE Trans Aerospace Electron Syst* 19:389–396
- Liu W, Yamazaki F (2011) Urban change monitoring from multi-temporal TerraSAR-X images. In: *Proceedings of 2011 joint urban remote sensing event*. IEEE, Munich, pp 277–280
- Liu W, Yamazaki F (2013) Detection of crustal movement from TerraSAR-X intensity image. *IEEE Geosci Remote Sens Lett* 10(1):199–203
- Lopes A, Touzi R, Nezry E (1990) Adaptive speckle filters and scene heterogeneity. *IEEE Trans Geosci Remote Sens* 28(6):992–1000
- Massonnet D, Rossi M, Carmona C, Adragna F, Peltzer G, Fiegl K, Rabaute T (1993) The displacement field of the Landers earthquake mapped by radar interferometry. *Nature* 364:138–142
- Matsuoka M, Yamazaki F (2004) Use of satellite SAR intensity imagery for detecting building areas damaged due to earthquakes. *Earthq Spectra* 20(3):975–994
- Matsuoka M, Yamazaki F (2010) Comparative analysis for detecting areas with building damage from several destructive earthquakes using satellite synthetic aperture radar images. *J Appl Remote Sens, SPIE* 4:041867
- Ozawa S, Nishimura TH, Suito H, Kobayashi T, Tobita M, Imakiire T (2011) Coseismic and postseismic slip of the 2011 magnitude-9 Tohoku-Oki earthquake. *Nature* 475(7356):373–377
- Yamazaki F, Iwasaki Y, Liu W, Nonaka T, Sasagawa T (2013) Detection of damage to building side-walls in the 2011 Tohoku, Japan earthquake using high-resolution TerraSAR-X images. In: *SPIE remote sensing 2013*, Dresden, SC05, pp 266–273
- Yonezawa C, Takeuchi S (2001) Decorrelation of SAR data by urban damages caused by the 1995 Hyogoken-Nanbu earthquake. *Int J Remote Sens* 22:1585–1600

Very-Long-Period Seismicity at Active Volcanoes: Source Mechanisms

Gregory P. Waite
Department of Geological and Mining
Engineering and Sciences, Michigan
Technological University, Houghton, MI, USA

Synonyms

Very low-frequency seismicity; VLP events; VLPs

Introduction

Deformation at active volcanoes spans a wide range of frequencies, from permanent changes up to cycles of 100s of Hz. Scattering and anelasticity attenuate the high-frequency signals rapidly, making their detection difficult. The dominant range of frequencies observed with seismological methods, between about 0.5 and 5 Hz, has been well recorded for many decades, beginning with analog recordings. Studies of discrete events and longer-duration tremor in this so-called long-period (LP) band have led to significant advances in understanding of volcanological processes, but studies of events with lower frequencies have only been possible since

the use of instruments that can record them has become widespread.

Very-long-period (VLP) seismicity with frequencies below the LP band has been observed at nearly every type of active volcano, including those with activity limited to hydrothermal systems. As with LP seismicity, VLP seismicity includes both discrete events and tremor. Because VLP events involve source mechanisms that are different from those of earthquakes that occur along tectonic faults, they are referred to as VLP events, rather than VLP earthquakes. Studies of VLP events have only been possible since permanent and temporary networks of broadband seismic recording stations at active volcanoes began in the 1990s. In this comparatively short amount of time, detailed investigations of VLP events have led to great advances in understanding of the geometries of shallow magma conduits, how magma moves through these systems, and how magma and magmatic gases interact with host rocks. These interactions are manifested as deformations, which result in elastic wave propagation. When these processes have timescales of 10s of seconds to minutes, VLP seismicity can result. Because of their lower frequency nature, VLP seismicity typically represents somewhat different fluid processes than those responsible for LP events. This entry describes the observations and principal models for VLP events. It begins with a brief discussion of the instrumentation used to record VLP seismicity, with a

particular emphasis on the limitations and complications that can arise. A summary of VLP observations leads to a detailed section on modeling and interpretation.

Instrumentation

Observations of VLP events require broadband seismometers installed in quiet locations. The use of active force feedback technology in sensors allows for recordings at low frequencies without requiring an unreasonably large mass ([► Principles of Broadband Seismometry](#)). Broadband sensors in use on volcanoes have a flat response to a wide range of periods, or frequencies, typically from about 120 to 0.02 s period (0.0083–50 Hz), although the lower limit on many of these sensors is a 60 or 30 s period. These narrower band sensors are sometimes called intermediate band rather than broadband sensors. This distinction is important, as the band of the instruments used limits observations of VLP events. The lower bound of the VLP band varies therefore, depending on the capability of the recording systems.

Below the low-frequency corner of the broadband seismometer, the response typically falls off at 40 dB per decade, but strong signals with frequencies below the corner can be recovered. However, at these lower frequencies, the signals are more likely to include both ground rotation (tilt) and translation. The horizontal components of broadband seismometers are sensitive to tilt through changes in gravitational acceleration (Rodgers 1968). In fact, the broadband signal is more susceptible to tilt with increasing period so that tilt dominates translation at periods below the sensor corner if the seismometer experiences substantial rotation. Because VLP recordings are typically made in the near field, the influence of tilt at frequencies below the corner cannot be ignored. Because tilt does not affect the vertical component to a measurable degree, it can be used to help discriminate between translational and rotational motion. Recent studies have taken advantage of the tilt response to incorporate it

into modeling of VLP band volcanic signals (Chouet and Dawson 2013; Maeda et al. 2011).

The tilt-dominated signals at frequencies below the instrument corner are in what is known as the ultra-long-period or ULP band (Johnson et al. 2009). Because of the importance of the instrument on the nature of the signal (i.e., rotation versus translation), the period that separates the VLP band from the ULP band might best be defined as the low corner of the seismometer used in the recording of the signals and therefore might vary from 30 to 120 s.

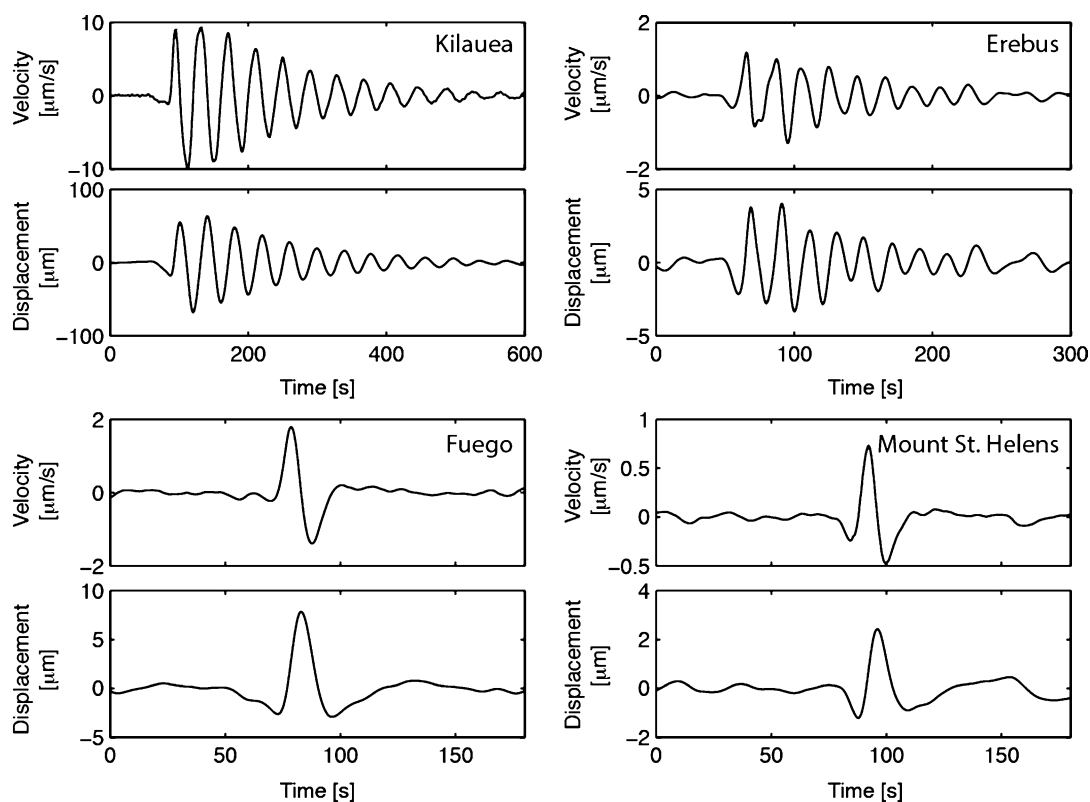
At the upper end of the VLP band from about 8 to 3 s period, the ubiquitous oceanic microseismic signal typically dominates. This signal, which is primarily due to interactions between ocean waves and the near-shore ocean floor, is observed everywhere on Earth but is particularly strong at ocean islands. Therefore the upper end of the observable VLP band is, in most cases, about 8 s.

Observations of VLP Seismicity

Unlike LP events, which typically have a long-duration, narrow-band coda of 10s of seconds that includes dozens of cycles, VLP events most often include only one cycle. Notable exceptions include events recorded at Mt. Erebus (Aster et al. 2008; Rowe et al. 1998) and Kilauea (Chouet and Dawson 2013; Chouet et al. 2010). VLP event locations are difficult to determine with respect to higher-frequency events, as discussed below. Where they have been determined, depths are usually within 1 km of the surface.

Although less common, VLP seismicity can also be found in the form of long-duration tremor. As with volcanic tremor in the LP band, VLP tremor signals may last for many minutes to days. Changes in tremor amplitude have been correlated with changes in surface activity (Kaneshima et al. 1996).

While some VLP seismicity can be clearly seen on broadband recordings, the amplitudes are typically much lower than those of any associated higher-frequency signal, including the



Very-Long-Period Seismicity at Active Volcanoes: Source Mechanisms, Fig. 1 Example VLP waveforms from four volcanoes with well-studied VLP seismicity.

All data were band-pass filtered between 60 and 12 s period, except Erebus which was filtered from 30 to 15 s period

ocean microseism, so seismograms must be filtered before the VLPs are visible. Figure 1 shows examples of VLP events from four different volcanoes.

VLP events may be associated with phreatic eruptions (Kawakatsu et al. 2000), strombolian eruptions (Chouet et al. 2008), vulcanian eruptions (Chouet et al. 2005), passive effusion (Waite et al. 2008), or without any surface expression at all (Hill et al. 2002).

VLP events have been observed at volcanoes ranging from low-viscosity, low-silica systems such as Stromboli and Kilauea to intermediate composition systems such as Augustine and Popocatepetl and to more viscous, silica-rich systems, such as Mount St. Helens. Chouet and Matoza (2013) have compiled a list of studies published since 1992, to which the reader is referred for a comprehensive review.

Modeling

For typical earthquakes, the most basic type of information determined includes the time of occurrence, the location, and the size of the event. VLP events do not lend themselves to simple location and magnitude determinations, so other techniques have been developed. The key challenge involves the uncertainty in the first arrival time. Because VLP events do not have a clear, impulsive onset, the uncertainty in choosing the onset of the event may be many seconds. Since seismic waves propagate on the order of several km per second, this translates into an unacceptably large uncertainty in the location and time of the origin. Furthermore, the magnitudes determined for typical tectonic earthquakes are based on the assumption that seismicity is generated by shear on a near-planar fault.

In the case of volcanic VLP events, this is almost certainly never the case.

Fortunately, because of the low frequency of VLP events, the wavelengths are 10s to 100s of km, which simplifies the modeling of wave field propagation. The waveforms are not distorted by shallow heterogeneities between the source and recording station, although topography does have an important affect (Neuberg and Pointer 2000).

Particle Motion Methods

The simplest techniques for modeling the source location of VLP events use the mapped particle motions. Assuming that most of the motion is in the radial-vertical plane, seismic recordings made on stations surrounding the surface projection of a VLP source can be used to identify that point. This first-order approach can also be used to estimate the VLP event depth but is complicated by surface topography that can significantly distort the particle motion.

A method based on the assumption of radial particle motions, called radial semblance, involves solving for the model that minimizes particle motion out of the radial-vertical plane (Kawakatsu et al. 2000). The method requires that the seismic velocity of the medium is known, but the choice of velocity does not greatly affect the location. A more important potential problem with the method is that it assumes an isotropic source. Because VLP recordings are typically made in the near field, they can have significant components of motion out of the radial plane for source geometries such as dikes or sills.

Full-Waveform Modeling

Most studies of VLPs in recent years have employed full-waveform modeling, as long as enough seismic stations are available to constrain the source mechanism. In these studies, a point source is typically chosen, and the wave field is modeled using a finite-difference approximation approach. For VLP events, the long wavelengths simplify the modeling in two important ways. First, the structural model used for the wave field modeling can be fairly simple. Variations in the elastic parameters of rocks on the order of km can be ignored for stations within a few km of

the source and wavelengths of 10s to 100s of km. This means the velocity-density structure can be assumed to be homogeneous. Choosing the wrong velocity will introduce some error in the source location but will not significantly alter the source mechanism.

The second way in which modeling of VLP data is simplified by the long wavelengths is that source finiteness can be ignored. In studies where the wavelength is on the order of the size of the source, for example, for a 100-m-long section of the conduit, the source dimensions affect the predicted waveforms at stations in the near field. However, the wavelength of VLP events is orders of magnitude larger than the source dimension, meaning a point source approximation is appropriate.

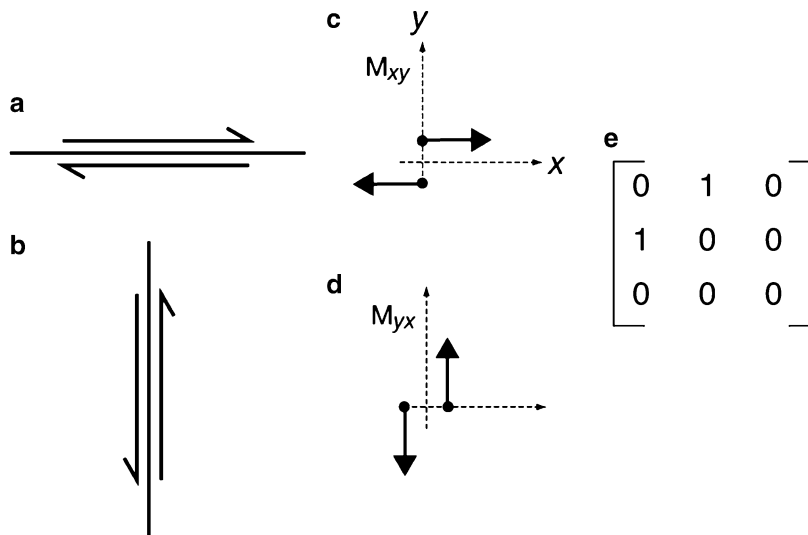
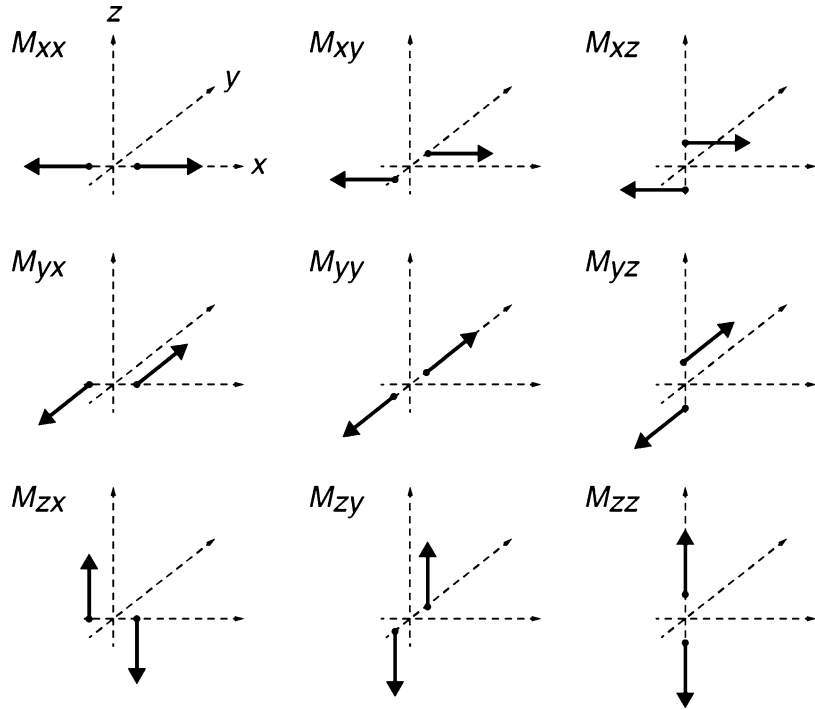
Moment Tensor Inversion

Waveform modeling approaches rely on source representation theory; that is, that there exists a set of equivalent forces that can represent those that caused the event (“► [Long-Period Moment-Tensor Inversion: The Global CMT Project](#)”, and “► [Non-Double-Couple Earthquakes](#)”). In modeling of tectonic earthquakes, the representative forces can be described by the symmetric second-order moment tensor, which has six independent components (Fig. 2). The components of the moment tensor are nine force couples, which represent either dipole or shear couples. A purely isotropic source would involve volume change, but not shear, and could be represented by a tensor with nonzero values for the dipole components on the diagonal and a nonzero trace. The off-diagonal terms describe torque and are nonzero when there is shear. Tectonic earthquakes, which involve slip on a near-planar fault, can adequately be described with shear and no volume change (the trace of the tensor is zero as in Fig. 3). On the other hand, VLP events may involve volume changes and no shear. Figure 4 provides an example of a source commonly found in VLP inversions, that of a tensile crack, with a nonzero trace.

Advective net mass transfer gives rise to single forces in addition to the moment components. Single forces are insignificant in tectonic

Very-Long-Period Seismicity at Active Volcanoes: Source Mechanisms,

Fig. 2 Force couples representing the full moment tensor. Dipole components are on the diagonal. The off-diagonal components represent shearing force couples

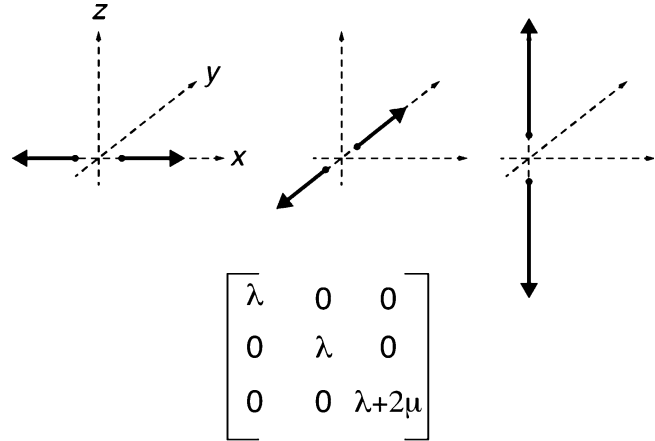


Very-Long-Period Seismicity at Active Volcanoes: Source Mechanisms, Fig. 3 Example of force couples representative of slip on a fault. For simplicity, the third dimension is not shown. For x and y in the horizontal plane, this would represent a vertical strike-slip fault. Right-lateral slip on the fault normal to the y direction,

represented by (a), or left-lateral slip on the fault normal to the x direction, represented by (b), would both be represented by the pair of force couples shown in (c) and (d), which is why this is called a double-couple solution. The dimensionless symmetric moment tensor for this set of force couples is also shown in (e)

Very-Long-Period Seismicity at Active Volcanoes: Source Mechanisms,

Fig. 4 Example of force couples representative of opening of a tensile crack lying in the x - y plane and opening in the z direction. Note that this can be represented with positive dipole components



earthquake modeling but may become important for VLP events associated with eruption or transfer of a large mass of material. Large landslides and glacier movements can also produce signals that are well modeled with single force solutions. In these cases, as the mass accelerates along the surface of the Earth, the response is a single force in the direction opposite the motion of the mass movement. Later, when the mass begins to slow, the reaction force is parallel to the direction of the slide. Similarly, the reaction force to a volcanic eruption is opposite the direction of motion of the erupted mass. Single forces can arise within the Earth as well. If the center of mass of a magma volume accelerates, the force can be observed seismically as long as there is a means for it to couple to the solid earth. Specific models are addressed below.

In order to model the source mechanism, Green functions that describe the response at one location (the station) to an impulse at another (the source) are computed. The displacement field $u_n(\vec{r}, t)$, generated by a source, may be written as:

$$u_n(\vec{r}, t) = F_p(t) * G_{np}(\vec{r}, t) + M_{pq}(t) * G_{np,q}(\vec{r}, t), \quad p, q = x, y, z, \quad (1)$$

where $u_n(\vec{r}, t)$ is the n -component of the displacement at time t and receiver position $\vec{r}$, $F_p(t)$ is the time history of the force applied in the p -direction, $M_{pq}(t)$ is the time history of the

pq -component of the moment tensor, and $G_{np}(\vec{r}, t)$ is the tensor of Green functions that relates the n -component of displacement at the receiver position, $\vec{r}$, with the p -component of impulsive force at the source position. Although it is not specified in Eq. 1, $G_{np}(\vec{r}, t)$ depends on the source position as well. The q notation indicates spatial differentiation with respect to the q -coordinate, and the symbol $*$ denotes convolution. Summation over repeated indices is implied.

With the matrix of Green functions, $\mathbf{G}$, the linear inverse problem, $\mathbf{d} = \mathbf{G}\mathbf{m}$, is solved where $\mathbf{d}$ is the data vector and $\mathbf{m}$ is the model vector, which consists of moment tensor components and single forces. Each model parameter has a time history. In the most general case, the inversion is allowed to vary freely so that a new model vector is computed for each corresponding time in the data vector. An analogous approach is to compute the inversions in the frequency domain, so that one frequency at a time is solved for each moment component. Other approaches to solving the inverse problem have also been used.

Because the number of model and data parameters involved in the inversions is small, the process is computationally inexpensive. This approach does not require that the ratio of the moment components is constant throughout the source time history, but in most cases the resulting ratios are stable for most of the source time history of the event.

Although the inversion can be constrained in various ways, for example, to force the trace of

the tensor to be zero in the case of tectonic earthquakes, unconstrained inversions for VLP sources are preferred. VLP source inversions are typically dominated by the isotropic components, suggesting volume changes in conduits as the primary source of deformation. Single forces are included in inversions to account for advection-related forces. By repeating the inversion over a large volume of points (Green functions) in a grid, the preferred source centroid location is chosen as that with the lowest misfit to the data.

Typical VLP Source Models

The time history of the moment tensor and single forces representing the VLP source must be interpreted in terms of a physical model. Although the modeling assumes a point source, which represents the centroid location of moment and force, the characteristics of the moment and force components have implications for the orientation of the source components. If the ratios of the moment components are stable for the duration of the event, the eigenvectors and eigenvalues (principal components) can easily be obtained and compared with simplified geometries. Assuming the host rock surrounding the source can be represented with as a homogeneous elastic material with elastic Lamé parameters λ and μ and the volume change is ΔV , a tensile crack can be represented with dipole components having magnitudes equal to $\Delta V\lambda$ for the in-plane components and $\Delta V(\lambda + 2\mu)$ for the component that is normal to the crack plane. For igneous rocks, possibly at high temperatures given their proximity to magmatic fluids, values of λ and μ may vary over several orders of magnitude. The ratio between them, however, is limited; λ ranges from 1 to 2 times μ . The ratios of the eigenvalues of a tension crack are then between 1:1:2 and 1:1:3. For a pipe, the dipole components are $\Delta V\lambda$ for the component parallel to the pipe and $\Delta V(\lambda + \mu)$ for those perpendicular to it leading to ratios of 1:2:2 to 2:3:3. For a purely spherical source volume change, the components are equal.

In some cases the moment tensor looks very much like one of these end member physical

models. In these cases, the source can be interpreted as volume change related to inflation and/or deflation of a portion of the conduit. In other cases, the interpretation is not straightforward but may involve a combination of sources. For example, for magma flowing from a sill into a dike, one would expect the total moment to be close to the combined volume decrease in the sill $-\Delta V\lambda[1:1:(1+2\mu/\lambda)]$ and increase in the dike, which might be $\Delta V\lambda[1:(1+2\mu/\lambda):1]$.

Single forces arise from net mass advection. Acceleration of the center of mass of a fluid will produce a reaction force in the Earth. The nature of the reaction force is related to the coupling, which can be separated into shearing and normal forces. Shearing forces along the conduit walls in a direction antiparallel to the flow direction couple as frictional forces. The efficiency of this mechanism is strongly related to the magma fluid viscosity. Normal forces couple to conduit geometrical features where there is a change in the shape or orientation of the conduit. Numerical and analog laboratory experiments demonstrate the efficacy of the coupling of single forces through this mechanism. A change in conduit geometry, for example, from a sill to a dike as above, can provide a location for the coupling of pressure changes resulting from mass movements to occur. It is important to recognize that steady flow, without acceleration, will not produce observable single forces.

VLP Events and Magma Type

VLP events have been identified and modeled at volcanoes with nearly all magma types. Styles of VLP event tend to be related to the fluid viscosity, which depends on magma composition and temperature. In general, low-silica magmas, such as basalts, have low viscosity, and high-silica magmas, such as rhyolite, have higher viscosity. In addition to the temperature, the presence of dissolved volatiles and the crystal content influence viscosity as well. Because of the importance of viscosity on interpreting the source mechanisms, the description of types of VLP sources is organized by fluid type.

VLP Events at Low-Viscosity Volcanoes

The nature of eruptions at low-viscosity volcanoes is linked to the movement of exsolved gas. Because gas bubbles can ascend buoyantly through the higher-density magma relatively easily, eruptions tend to be less explosive than at volcanoes with magmas that are capable of trapping bubbles under higher pressures. Typical eruptions include bursting of bubbles at the top of the magma column, although more explosive eruptions are possible.

Bubbles grow through diffusion of volatiles out of solution and through coalescence of smaller bubbles. Bubbles also grow as they ascend due to the lower magma pressures at shallow depths. Under favorable conditions bubbles can grow to a diameter as large as the conduit, and if the gas volume is large enough, the bubble will occupy an elongated region within the conduit. As this gas slug ascends, magma must descend around it through a relatively thin film. An abrupt acceleration of the center of mass occurs where the conduit changes shape, either by changing diameter or orientation. VLP seismicity can result from these accelerations, which produce the equivalent of single forces and moments.

A well-studied laboratory analog model involves the ascent of a gas slug in a cylindrical conduit that has an abrupt change from a smaller to larger diameter (James et al. 2006). The slug accelerates slowly as it ascends through the lower portion of the conduit due to the gradual decrease in pressure of the overlying fluid. When the slug enters the upper portion of the conduit, it expands and accelerates rapidly. The mass of the gas is low so that the acceleration of the slug does not impart a significant force, but the acceleration of the gas requires a downward acceleration of the fluid. The fluid is massive enough to produce an observable force which couples to the conduit at the point where it enlarges and, to a lesser degree, through the frictional forces along the conduit wall. This model provides context for interpreting VLP events in low-viscosity magmas where gas slug ascent is common.

VLP events from Mount Erebus, Etna, Kilauea, and Stromboli are among the best-studied examples from low-viscosity volcanoes. In each case, the migrations of gas slugs have been interpreted as the source of at least some of the VLP events. In most cases the VLP events are temporally linked to explosive eruptions. At Stromboli, for example, moment tensor inversions point to the inflation and deflation of crack-like portions of the conduit to accommodate the movement of gas slugs that lead to strombolian-style eruptions (Chouet et al. 2008). VLP events associated with eruptions for separate, closely spaced vents have distinct waveform characteristics, and the source mechanisms are consistent with changes in the conduit geometry leading to each of the vents.

At Kilauea, there are also multiple styles of VLP events, all of which are associated with differential motions of gas and magma. Again, the best coupling of the forces within the magma to the solid conduit occurs at places in the conduit where the geometry changes. One feature that emerges from the modeling is the transition from a near-horizontal crack to a pipe that leads to the surface (Chouet et al. 2010). This corner is at a depth of about 1 km below the surface, about two or three times deeper than the main VLP source centroid at Stromboli. At Erebus, similar types of eruptions are associated with the ascent of a gas slug from a position 400 hundred meters below and a several hundred meters to the northwest of the active lava lake (Aster et al. 2008).

VLP Events at High-Viscosity Volcanoes

In contrast to VLP seismicity at low-viscosity volcanoes, which tend to be associated with small-scale strombolian explosions, at higher-viscosity systems, VLP events are most commonly associated with more explosive vulcanian-style eruptions. These higher-viscosity volcanoes are typically andesitic and dacitic in composition, but less silicic magmas can have high viscosities in the shallow subsurface if their crystal content is high. These larger

eruptions may eject significant masses of rock, meaning the vertical reaction force is the dominant component of the VLP source mechanism (Ohminato et al. 2006).

In other cases, the source mechanisms of VLP events associated with vulcanian-style eruptions are dominated by crack-like mechanisms associated with changes in conduit geometry or shallow (1–2 km) magma reservoirs (Chouet et al. 2005). Inflation or deflation mechanisms are not likely to involve gas slugs in these cases, since the magmas are too viscous to allow slug ascent. Instead, deflation may follow the explosion and be related to the eruption of material from the conduit. Re-inflation of the conduit can be the result of a combination of the influx of new magma and pressurization associated with rapid gas exsolution (Nishimura 2004).

VLP events in silicic systems may also be associated with the passive extrusion of magma, as in the case of Mount St. Helens (Waite et al. 2008). The VLP events were the result of unsteady flow as the magma moved from a vertical to horizontal portion of the conduit just below the surface. The mechanism included inflation and deflation of cracks and also nearly horizontal single forces. The single forces were associated with acceleration and deceleration of magma as it moved through this portion of the conduit.

VLP Events at Hydrothermal Systems

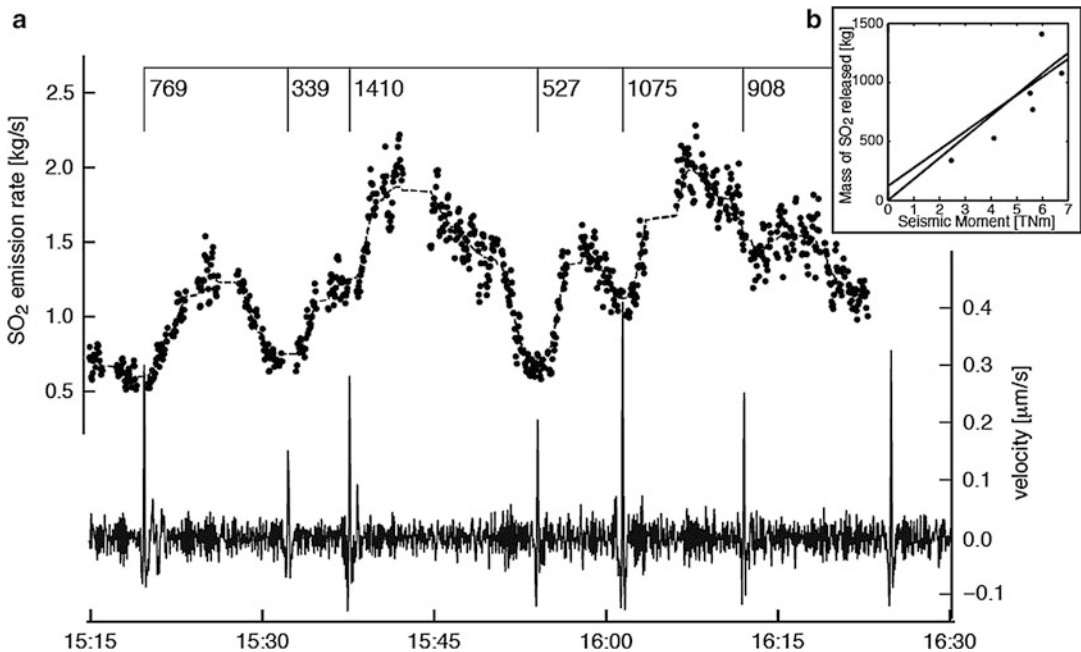
Hydrothermal cracks, filled with water, steam, or a combination of nonmagmatic fluids, may also be the source of VLP events. Unsteady flow of pressurized hydrothermal fluids through cracks (Nakamichi et al. 2009; Ohminato 2006) can produce VLP events that are very similar to those produced in magmatic systems. The low viscosity of water and steam relative to magma inhibits measureable flow-related frictional single forces, so mechanisms tend to be dominated by volume changes in cracks. At Aso volcano, nearly continuous VLP tremor is attributed to resonance in a horizontal crack (Kaneshima et al. 1996).

Long-Duration VLPs and VLP Tremor

VLP events tend to have short duration with respect to the period, having only one or two cycles. However, VLP events may have long durations indicative of a resonant source mechanism (Chouet et al. 2010; Kumagai et al. 2003; Rowe et al. 1998). Nearly continuous VLP tremor, with intermittent higher amplitudes, was recorded at Usu, but unlike Kilauea and Erebus, this signal was not associated with any observable surface activity (Yamamoto et al. 2002). In these low-viscosity systems, the ringing of the signal may be attributed to the oscillations of pressure waves, which are induced by the initial mass advection-related mechanism. At Usu, it was most likely related to unsteady flow of magma through a constricted portion of the conduit. Although VLP tremor is uncommon, it can be useful for constraining the extent of magma conduit components, as resonant properties of the signal depend on both the fluid and host rock properties and the size and shape of the resonator.

Relationship Between VLP Events and Gas Emissions

At Erebus, Kilauea, and Stromboli, VLP events are temporally linked to visual observations of explosions; the correlation is consistent with models for the VLP that involve ascent of a gas slug through the conduit to the surface. In a few cases, VLP events have been recorded coincident with high-resolution gas emission measurements. These cases provide a means for quantitative estimates of the relationship between outgassing and VLP events. At Asama (Kazahaya et al. 2011) and for certain types of VLP events at Fuego (Waite et al. 2013), the VLP event amplitudes correlate with the mass of SO₂ measured for each discrete gas emission (Fig. 5). The correlation is less clear at Etna but is present under certain conditions (Zuccarello et al. 2013). These correlations demonstrate the influence of unsteady gas flow on deformation in the shallow conduit. They can be used to infer the role of gases for VLP events even when quantitative measurements are not available.



Very-Long-Period Seismicity at Active Volcanoes: Source Mechanisms, Fig. 5 (a) Temporal correlation between VLP events (*solid line*) and SO₂ emission rate (*dots*) plotted for about 90 min on 15 January 2008 from Fuego. Masses computed for each of the six events are

listed above the SO₂ data. The *dashed line* is a weighted 50-point moving average of the SO₂ emission data. The linear regressions of SO₂ mass against seismic moment (b) show a positive correlation.

Summary

Very-long-period seismicity is usually the result of fluid transport processes in magmatic or hydrothermal systems. VLP events are most often related to inflation and deflation of elements of a conduit system, such as dikes, sills, and pipes; they tend to be dominated by isotropic elements of the moment tensor. Forces associated with mass advection can give rise to single force source components which are not found in models for typical earthquake sources. VLP events have been found in nearly every type of volcano yet they are not ubiquitous. If fluid flow is steady, there will not be large pressure fluctuations necessary to cause VLP seismicity. While their existence has only been known since the early 1990s, interpretations of VLP events have led to a far greater understanding of geometry and dynamics of fluids in shallow magmatic and hydrothermal systems.

References

- Aster R, Zandomenighi D, Mah S, McNamara S, Henderson DB, Knox H, Jones K (2008) Moment tensor inversion of very long period seismic signals from Strombolian eruptions of Erebus Volcano. *J Volcanol Geotherm Res* 177(3):635–647. doi:10.1016/j.jvolgeores.2008.08.013
- Chouet B, Dawson P (2013) Very long period conduit oscillations induced by rockfalls at Kilauea Volcano, Hawaii. *J Geophys Res Solid Earth* 118(10): 5352–5371. doi:10.1002/jgrb.50376
- Chouet BA, Matoza RS (2013) A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. *J Volcanol Geotherm Res* 252:108–175. doi:10.1016/j.jvolgeores.2012.11.013
- Chouet B, Dawson P, Arciniega-Ceballos A (2005) Source mechanism of Vulcanian degassing at Popocatepetl Volcano, Mexico, determined from waveform inversions of very long period signals. *J Geophys Res* 110:B07301. doi:10.1029/2004JB003524
- Chouet B, Dawson P, Martini M (2008) Shallow-conduit dynamics at Stromboli Volcano, Italy, imaged from waveform inversions. *Geol Soc Lond Spec Publ* 307(1):57–84. doi:10.1144/SP307.5

- Chouet BA, Dawson PB, James MR, Lane SJ (2010) Seismic source mechanism of degassing bursts at Kilauea Volcano, Hawaii: results from waveform inversion in the 10–50 s band. *J Geophys Res* 115(B9). doi:10.1029/2009jb006661
- Hill DP, Dawson PB, Johnston MJS, Pitt AM, Biasi GP, Smith K (2002) Very-long-period volcanic earthquakes beneath Mammoth Mountain, California. *Geophys Res Lett* 29(10):1370. doi:10.1029/2002GL014833
- James MR, Lane SJ, Chouet BA (2006) Gas slug ascent through changes in conduit diameter: laboratory insights into a volcano-seismic source process in low-viscosity magmas. *J Geophys Res* 111:B05201. doi:10.1029/2005JB003718.
- Johnson JB, Sanderson R, Lyons J, Escobar-Wolf R, Waite G, Lees JM (2009) Dissection of a composite volcanic earthquake at Santiaguito, Guatemala. *Geophys Res Lett* 36(16). doi:10.1029/2009gl039370
- Kaneshima S, Kawakatsu H, Matsubayashi H, Oikawa J, Takeo M, Iidaka T, Yamasato H (1996) Mechanism of phreatic eruptions at Aso volcano inferred from near-field broadband seismic observations. *Science* 273(5275):642–645. doi:10.1126/science.273.5275.643
- Kawakatsu H, Kaneshima S, Matsubayashi H, Ohminato T, Sudo Y, Tsutsui T, Legrand D (2000) Aso94: Aso seismic observation with broadband instruments. *J Volcanol Geotherm Res* 101(1–2):129–154
- Kazahaya R, Mori T, Takeo M, Ohminato T, Urabe T, Maeda Y (2011) Relation between single very-long-period pulses and volcanic gas emissions at Mt. Asama, Japan. *Geophys Res Lett* 38:L11307. doi:10.1029/2011GL047555
- Kumagai H, Miyakawa K, Negishi H, Inoue H, Obara K, Suetsugu D (2003) Magmatic dike resonances inferred from very-long-period seismic signals. *Science* 299(5615):2058–2061. doi:10.1126/science.1081195
- Maeda Y, Takeo M, Ohminato T (2011) A waveform inversion including tilt: method and simple tests. *Geophys J Int* 184(2):907–918. doi:10.1111/j.1365-246X.2010.04892.x
- Nakamichi H, Kumagai H, Nakano M, Okubo M, Kimata F, Ito Y, Obara K (2009) Source mechanism of a very-long-period event at Mt. Ontake, central Japan: response of a hydrothermal system to magma intrusion beneath the summit. *J Volcanol Geotherm Res* 187(3–4):167–177. doi:10.1016/j.jvolgeores.2009.09.006
- Neuberg J, Pointer T (2000) Effects of volcano topography on seismic broad band waveforms. *Geophys J Int* 143(1):239–248
- Nishimura T (2004) Pressure recovery in magma due to bubble growth. *Geophys Res Lett* 31(12):L12613. doi:10.1029/2004gl019810
- Ohminato T (2006) Characteristics and source modeling of broadband seismic signals associated with the hydrothermal system at Satsuma-Iwojima volcano, Japan. *J Volcanol Geotherm Res* 158(3–4):467–490. doi:10.1016/j.jvolgeores.2006.08.004
- Ohminato T, Takeo M, Kumagai H, Yamashima T, Oikawa J, Koyama E, Urabe T (2006) Vulcanian eruptions with dominant single force components observed during the Asama 2004 volcanic activity in Japan. *Earth Planets Space* 58:583–593
- Rodgers PW (1968) The response of the horizontal pendulum seismometer to Rayleigh and Love waves, tilt, and free oscillations of the Earth. *Bull Seism Soc Am* 58:1384–1406
- Rowe CA, Aster RC, Kyle RR, Schlue JW, Dibble RR (1998) Broadband recording of Strombolian explosions and associated very-long-period seismic signals on Mount Erebus volcano, Ross Island, Antarctica. *Geophys Res Lett* 25(13):2297–2300
- Waite GP, Chouet BA, Dawson PB (2008) Eruption dynamics at Mount St. Helens imaged from broadband seismic waveforms: interaction of the shallow magmatic and hydrothermal systems. *J Geophys Res* 113(B2):B02305. doi:10.1029/2007jb005259
- Waite GP, Lyons JJ, Nadeau PA (2013) Variability in eruption style and associated very-long-period earthquakes at Fuego volcano, Guatemala. *J Geophys Res* 118(4):1526–1533. doi:10.1002/jgrb.50075
- Yamamoto M, Kawakatsu H, Yomogida K, Koyama J (2002) Long period (12 sec) volcanic tremor observed at Usu 2000 eruption: seismological detection of a deep magma plumbing system. *Geophys Res Lett* 29(9):1329. doi:10.1029/2001GL013996
- Zuccarello L, Burton MR, Saccorotti G, Bean CJ, Patanè D (2013) The coupling between very long period seismic events, volcanic tremor, and degassing rates at Mount Etna volcano. *J Geophys Res* 118(9):4910–4921. doi:10.1002/jgrb.50363

Vibration-Based Damage Identification: The Z24 Bridge Benchmark

Edwin Reynders and Guido De Roeck
Department of Civil Engineering, KU Leuven,
Leuven, Belgium

Synonyms

Ambient vibration testing; Data normalization;
Progressive damage testing; Structural health
monitoring; System identification

Introduction

Structural health monitoring (SHM) relies on the repeated observation of features that are each sensitive to a certain type of structural damage. Among the many possible features, dynamic characteristics, in particular natural frequencies, are often selected as they depend on the global and the local stiffness of the structure of interest as well as its boundary conditions.

Monitoring the evolution of the features over time allows, in principle, to detect structural damage. In practice, however, this needs to be applied with care because of two reasons. Firstly, many features cannot be measured directly, but they have to be estimated from measured data using system identification techniques. Modal characteristics, for instance, can be estimated from vibration response data such as accelerations or strains, but this introduces estimation errors (Reynders et al. (2008); Reynders 2012). Secondly, nearly all features are not only sensitive to structural damage but also to changes in temperature, relative humidity, wind speed, operational loading, etc. This means that both the accuracy of the estimated features and the environmental and operational influences should be accounted for.

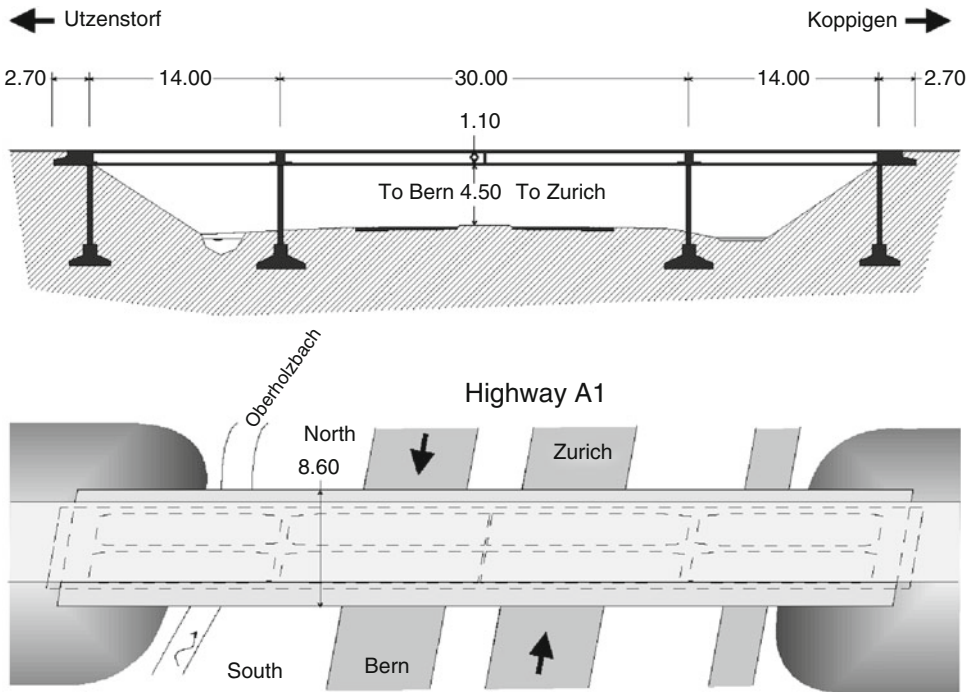
This entry presents a case study, where a full-sized structure has been monitored for almost a year before introducing realistic damage in a controlled way. The aim of the long-term continuous monitoring test was to quantify the environmental variability of the bridge dynamics, while the aim of the subsequent short-term progressive damage test was to investigate experimentally whether realistic damage has a measurable influence on the bridge's dynamic behavior. The combination of a long-term continuous monitoring test with realistic short-term progressive damage tests resulted in a unique data set that has aided many researchers in benchmarking new methods for system identification, operational modal analysis, damage identification, and structural health monitoring (Maeck and De Roeck 2003; Reynders and De Roeck 2009). In the next sections, the bridge and the performed tests are first described in more detail. Subsequently, the

identified modal characteristics are presented as well as their evolution during the monitoring period. Finally, results on structural health monitoring and damage identification are presented.

Description of the Structure and the Dynamic Tests

The Z24 bridge was located in the canton Bern near Solothurn, Switzerland. It was part of the road connection between the villages of Koppigen and Utzenstorf, overpassing the A1 highway between Bern and Zürich. It was a classical post-tensioned concrete two-cell box-girder bridge with a main span of 30 m and two side spans of 14 m (Fig. 1). The bridge was built as a freestanding frame with the approaches backfilled later. Both abutments consisted of triple concrete columns connected with concrete hinges to the girder. Both intermediate supports were concrete piers clamped into the girder. An extension of the bridge girder at the approaches provided a sliding slab. All supports were rotated with respect to the longitudinal axis, which yielded a skew bridge. The bridge, which dated from 1963, was demolished at the end of 1998, because a new railway adjacent to the highway required a new bridge with a larger side span.

To monitor the bridge dynamics over a 1-year time span, 16 accelerations were measured on the bridge at different points and in different directions. Every hour, a sequence of 65,536 acceleration samples, taken at the 16 sensors, was collected and stored to a hard disk after compression. The construction works at the new bridge that replaced the Z24 caused damage to one accelerometer. Although the type of accelerometer was specially designed for long-term use, some showed a considerable drift and a few of them failed during operation. Since the aim of the long-term monitoring test was to quantify the environmental variability of the bridge dynamics, all environmental variables that were considered of possible importance for the bridge dynamics have been monitored. Sensors to measure air temperature, air humidity, rain true or false, wind speed, and wind direction were installed at



Vibration-Based Damage Identification: The Z24 Bridge Benchmark, Fig. 1 Front view (top) and top view (bottom) of the Z24 bridge (Reproduced with

permission from: J. Maeck and G. De Roeck. Description of Z24 benchmark. *Mechanical Systems and Signal Processing*, 17(1):127–131, 2003)

the bridge, resulting in five sensors for the atmospheric conditions. A sensor consisting of two inductive loops was installed to detect the presence of vehicles on the bridge. Since temperature was known to have a key influence on the dynamics of civil engineering structures, the bridge's thermal state was monitored in detail. At the middle of the three spans, the temperature was measured at eight positions on the girder. The girder was a continuous beam with the intermediate piers clamped into it, and therefore the angular deflection of the girder at these piers and the elongation of the middle span were also measured. The soil temperature near each of the concrete columns at the approaches was monitored, as well as that near the north, central, and south parts of the intermediate piers (12 sensors in total). Finally, the temperature of the pavement was also measured at the middle of the three spans, since the drilling of the access holes for the installation of the temperature sensors on the girder revealed that the asphalt cover had a

thickness of 16–18 cm instead of the 5 cm as indicated on the blueprints.

In order for the subsequent progressive damage tests to be significant, it was made sure that they were relevant for the safety of the bridge and that the simulated damage occurred frequently, a condition that was checked in the literature and by questioning Swiss bridge owners. Since the A1 highway was never closed to traffic, some damage scenarios that meet these criteria could not be applied without reducing the safety of the traffic, which was considered of paramount importance. The traffic on the Z24 bridge was diverted to the A36 highway. Table 1 gives a complete overview of all progressive damage tests that were performed. Some of them are illustrated in Fig. 2. Before and after each applied damage scenario, the bridge was subjected to a forced and an ambient operational vibration test. With a measurement grid consisting of a regular 3×45 grid on top of the bridge deck and a 2×8 grid on each of the two pillars,

291 degrees of freedom have been measured: all displacements on the pillars and mainly vertical and lateral displacements on the bridge deck. Because the number of degrees of freedom to be measured exceeded the number of available accelerometers and acquisition channels, the data were collected in nine setups using five

Vibration-Based Damage Identification: The Z24 Bridge Benchmark, Table 1 Chronological overview of applied damage scenarios, indicating on which date a specific scenario was fully realized (Reproduced with permission from: E. Reynders, G. Wursten, and G. De Roeck. Output-only structural health monitoring in changing environmental conditions by means of nonlinear system identification. *Structural Health Monitoring*, 13(1):82–93, 2014)

Date (1998)	Scenario
4 August	Undamaged condition
9 August	Installation of pier settlement system
10 August	Lowering of pier, 20 mm
12 August	Lowering of pier, 40 mm
17 August	Lowering of pier, 80 mm
18 August	Lowering of pier, 95 mm
19 August	Lifting of pier, tilt of foundation
20 August	New reference condition
25 August	Spalling of concrete at soffit, 12 m ²
26 August	Spalling of concrete at soffit, 24 m ²
27 August	Landslide of 1 m at abutment
31 August	Failure of concrete hinge
2 September	Failure of 2 anchor heads
3 September	Failure of 4 anchor heads
7 September	Rupture of 2 out of 16 tendons
8 September	Rupture of 4 out of 16 tendons
9 September	Rupture of 6 out of 16 tendons

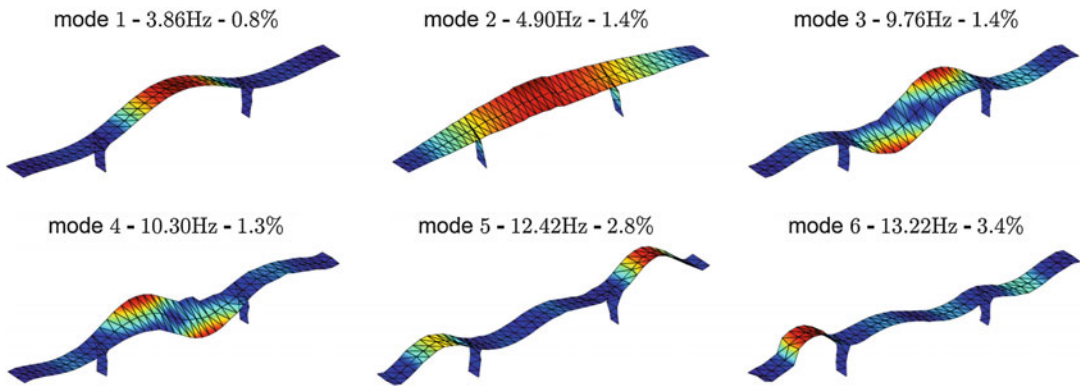
reference channels. The forced excitation was applied by two vertical shakers, placed on the bridge deck. A 1-kN shaker was placed on the middle span and a 0.5-kN shaker was placed at the Koppigen side span. The shaker input signals were generated using an inverse fast Fourier transform (FFT) algorithm, resulting in a fairly flat force spectrum between 3 and 30 Hz. After scenario 8, a drop weight test was also performed, using a device that allowed to drop a mass of up to 120 kg from a height of up to 1 m in a controlled way (Reynders et al. 2010). The applied shaker and drop weight forces were periodic with eight periods. A total of 65,536 samples were collected at a sampling rate of 100 Hz, using an antialiasing filter with a 30-Hz cutoff frequency.

Identification of Modal Characteristics

The ambient, shaker, and drop weight data from scenario 8 of the progressive damage test have been employed as benchmark data for system identification methods for operational modal analysis. Peeters and Ventura (2003) compare the modal parameter estimates obtained by seven different research teams in the framework of this benchmark. In addition, new modal parameter estimation techniques have been validated on the benchmark data. The best reported result was obtained by applying a subspace identification algorithm (Reynders and De Roeck 2008) and a maximum likelihood algorithm



Vibration-Based Damage Identification: The Z24 Bridge Benchmark, Fig. 2 Settlement system used for damage scenarios 3–6 (a) and cutting of tendons for scenarios 15–17 (b)



Vibration-Based Damage Identification: The Z24 Bridge Benchmark, Fig. 3 Natural frequencies, damping ratios, and mode shapes of the Z24 for damage scenario 8, identified with reference-based stochastic

subspace identification (Reproduced with permission from: E. Reynders et al. Fully automated (operational) modal analysis, *Mechanical Systems and Signal Processing*, 29:228–250, 2012)

(Parloo, Guillaume, and Cauberghe 2003) to the shaker data. In this way, 10 modes could be determined.

From the ambient vibration data, 6 modes can be determined with good quality (see Fig. 3). These modes are also predominantly present in the long-term monitoring data. Modes 1, 5, and 6 are vertical bending modes, while mode 2 is a lateral mode. Modes 3 and 4 are modes that show combined bending and torsion, which is caused by the skewness of the pillars with respect to the bridge deck. They have close natural frequencies and their mode shapes look similar, but in fact they show a low degree of correlation.

Peeters and De Roeck (2001) also identified the modal characteristics of the first four modes (see Fig. 3) for each of the 5,652 sets of hourly acceleration data, recorded during the continuous monitoring test. For these data, detailed mode shapes are not available due to the small number of accelerometers employed during the monitoring period. The evolution of the natural frequencies over time is plotted in Fig. 4. It can be noted in this figure that the monitoring system failed (was not active) during short periods of time.

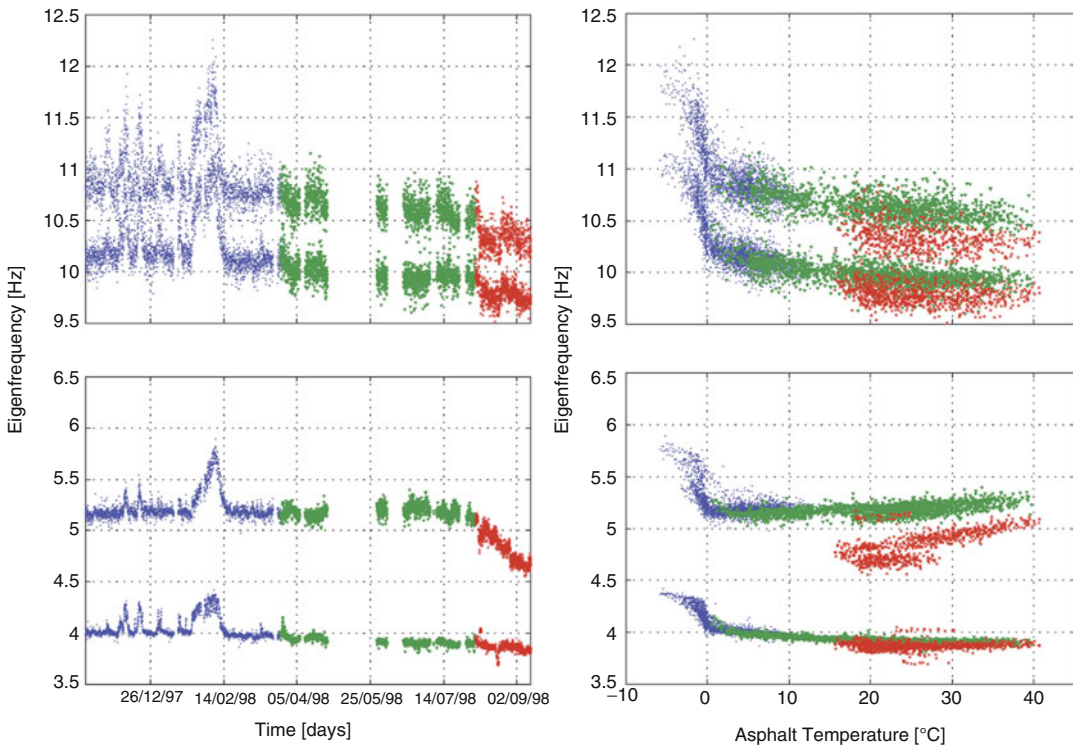
Of all environmental variables that had been recorded, the temperature was found to have the largest influence on the modal characteristics. A plot of the natural frequencies as a static function of the temperature of the surface asphalt

layer (Fig. 4) reveals the high influence of the temperature on the bridge dynamics, as well as its nonlinear nature. The most probable explanation lies in the large temperature dependency of the Young's modulus of the asphalt layer (Peeters and De Roeck 2001).

Structural Health Monitoring by Natural Frequency Tracking

As was observed in Fig. 4, the natural frequency variations due to regular environmental and operational variability may mask the influence of even severe damage completely. Therefore, when performing damage detection by monitoring natural frequencies, it is necessary to filter out the regular environmental variability.

This can be achieved by system identification. In this approach, a mathematical model that describes the relationship between regular environmental variations and variations in natural frequencies is fitted to training data that have been collected on the undamaged structure. Afterwards, in the monitoring phase, the mathematical model is employed for predicting the natural frequencies. When the differences between these predicted values and the observed ones grow large, the structure behaves differently as during the training phase. These differences are then most probably caused by damage,



Vibration-Based Damage Identification: The Z24 Bridge Benchmark, Fig. 4 Natural frequencies of the Z24 bridge: evolution as a function of time (*left*) and as a function of the temperature of the asphalt layer. *Blue dots*, training data (3,000 training samples); *green triangles*, monitoring data in undamaged condition; *red crosses*,

monitoring data in damaged condition (Reproduced with permission from: E. Reynders, G. Wursten, and G. De Roeck. Output-only structural health monitoring in changing environmental conditions by means of nonlinear system identification. *Structural Health Monitoring*, 13(1):82–93, 2014)

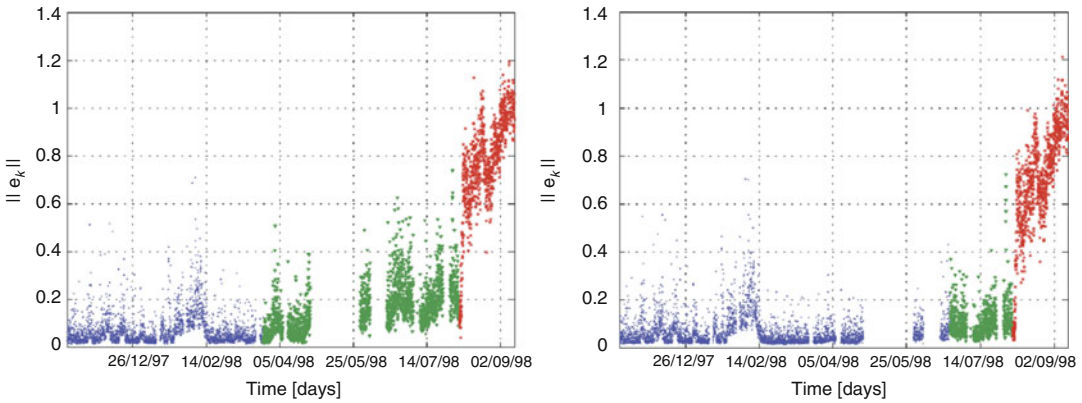
because the regular environmental variations are accounted for by the model.

Several system identification techniques have been proposed in this context (Xia et al. 2012). One of the most promising approaches is based on improved kernel principal component analysis or kPCA (Reynders et al. 2014). This approach has the advantages of (1) identifying a very general nonlinear model in a computationally cheap and robust way and (2) being output only, i.e., not requiring the measurement of the environmental and operational quantities that cause the regular variations in the modal characteristics. Figure 5 illustrates that kPCA can effectively predict the onset of damage (installation of the pier settlement system, see Table 1) by natural frequency monitoring. The predictions are better when more training data samples are used, because in this

way, the nonlinear model covers a larger portion of the regular environmental and operational variations of the natural frequencies. Alternative approaches that do need the environmental data as inputs also perform well on the present data set (see, e.g., Spiridonakos and Chatzi 2014), since a large amount of environmental variables was measured.

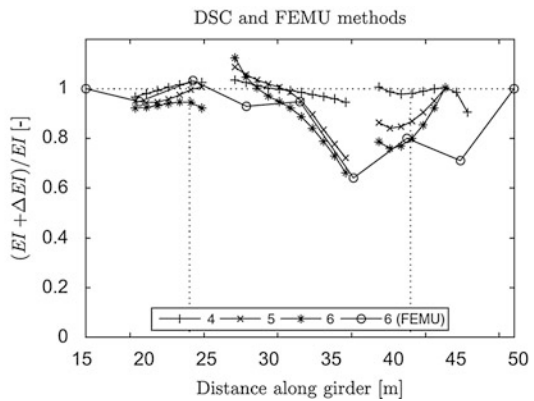
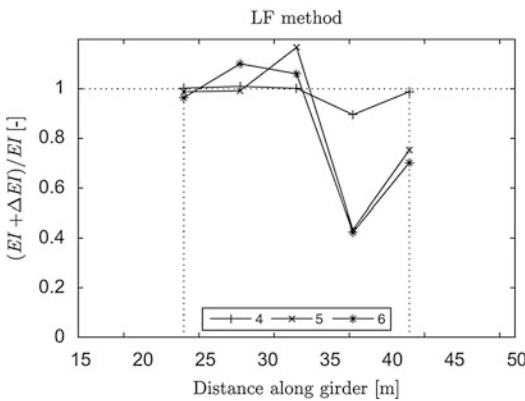
Damage Localization and Quantification

Modal characteristics cannot only be used for monitoring and damage detection but also for damage localization and quantification. The progressive damage test data from the Z24 bridge have served as benchmark data in this context, and Reynders and De Roeck (2009) present a



Vibration-Based Damage Identification: The Z24 Bridge Benchmark, Fig. 5 Evolution of the misfit of a nonlinear output-only model that predicts the evolution of the first four natural frequencies of the Z24 bridge. *Left*: misfit of a model trained with 3,000 data samples. *Right*: misfit of a model trained with 4,000 data samples. *Blue dots*, training data; *green triangles*, monitoring data in

undamaged condition; *red crosses*, monitoring data in damaged condition (Reproduced with permission from: E. Reynders, G. Wursten, and G. De Roeck. Output-only structural health monitoring in changing environmental conditions by means of nonlinear system identification. *Structural Health Monitoring*, 13(1):82–93, 2014)



Vibration-Based Damage Identification: The Z24 Bridge Benchmark, Fig. 6 Relative difference in bending stiffness of the Z24 bridge deck between damage scenarios 4–6 and damage scenario 2. *Left*, local flexibility (LF) method; *right*, direct stiffness calculation (DSC) and

finite element model updating (FEMU) (Reproduced with permission from: E. Reynders and G. De Roeck. A local flexibility method for vibration-based damage localization and quantification. *Journal of Sound and Vibration*, 329(12):2367–2383, 2010)

literature review of the benchmark results reported up to 2009.

Two major approaches for damage localization and quantification can be discriminated: model based and non-model based, also called parametric and nonparametric, respectively. Parametric methods are based on a (finite element) model of the structure, some parameters of which are adjusted using vibration measurements, by minimizing the difference between

the modal characteristics computed with the model and the ones that are identified from measurements. Nonparametric methods do not need a detailed model of the structure, but most of them are based on ad hoc damage indicators with a rather limited range of application; exceptions are the damage-locating vector method of Bernal (2002) and the local flexibility method of Reynders and De Roeck (2010).

For the Z24 bridge, the linear bending and torsion stiffness of the bridge deck have been most often employed for damage localization and quantification. The pier settlement of the bridge (damage scenarios 3–6 from Table 1) is the damage case that has the largest influence on these quantities. Figure 6 shows the estimated ratio between the bending stiffness in damaged and undamaged condition, as obtained from three different damage identification methods: one non-parametric (Reynders and De Roeck 2010) and two parametric methods (Maeck et al. 2001; Teughels and De Roeck 2004). All three methods yield qualitatively similar results: a large reduction in bending stiffness around the settled pier, which is in agreement with observed crack patterns in the bridge deck. Quantitative differences are probably related to the fact that the three methods use different assumptions and strategies to aggregate the local damage over a certain zone of the structure.

Summary

Structural health monitoring (SHM) relies on the repeated observation of features that are each sensitive to a certain type of structural damage. Among the many possible features, dynamic characteristics, in particular natural frequencies, are often selected as they depend on the global and the local stiffness of the structure of interest as well as its boundary conditions. This entry presents a case study, where a full-sized structure has been monitored for almost a year before introducing realistic damage in a controlled way. The combination of a long-term continuous monitoring test with realistic short-term progressive damage tests resulted in a unique data set that has aided many researchers in benchmarking new methods for system identification, operational modal analysis, damage identification, and structural health monitoring. The bridge and the performed tests are first described in detail. Subsequently, the identified modal characteristics are presented along with their evolution during the monitoring period. Finally, recent results on structural health monitoring and damage identification, obtained with the Z24 data, are presented.

Cross-References

- ▶ [Advances in Online Structural Identification](#)
- ▶ [Ambient Vibration Testing of Cultural Heritage Structures](#)
- ▶ [Operational Modal Analysis in Civil Engineering: An Overview](#)
- ▶ [Post-Earthquake Diagnosis of Partially Instrumented Building Structures](#)
- ▶ [Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring](#)
- ▶ [Stochastic Structural Identification from Vibrational and Environmental Data](#)
- ▶ [System and Damage Identification of Civil Structures](#)

References

- Bernal D (2002) Load vectors for damage localization. *ASCE J Eng Mech* 128(1):7–14
- Maeck J, De Roeck G (2003) Description of Z24 benchmark. *Mech Syst Signal Process* 17(1):127–131
- Maeck J, Peeters B, De Roeck G (2001) Damage identification on the Z24-bridge using vibration monitoring. *Smart Mater Struct* 10(3):512–517
- Parloo E, Guillaume P, Cauberghe B (2003) Maximum likelihood identification of non-stationary operational data. *J Sound Vib* 268(5):971–991
- Peeters B, De Roeck G (2001) One-year monitoring of the Z24-bridge: environmental effects versus damage events. *Earthq Eng Struct Dyn* 30(2):149–171
- Peeters B, Ventura C (2003) Comparative study of modal analysis techniques for bridge dynamic characteristics. *Mech Syst Signal Process* 17(5):965–988
- Reynders E (2012) System identification methods for (operational) modal analysis: review and comparison. *Arch Comput Meth Eng* 19(1):51–124
- Reynders E, De Roeck G (2008) Reference-based combined deterministic-stochastic subspace identification for experimental and operational modal analysis. *Mech Syst Signal Process* 22(3):617–637
- Reynders E, De Roeck G (2009) Continuous vibration monitoring and progressive damage testing on the Z24 bridge. In: Boller C, Chang FK, Fujino Y (eds) *Encyclopedia of structural health monitoring*. Wiley, New York, pp 2149–2158
- Reynders E, De Roeck G (2010) A local flexibility method for vibration-based damage localization and quantification. *J Sound Vib* 329(12):2367–2383
- Reynders E, Pintelon R, De Roeck G (2008) Uncertainty bounds on modal parameters obtained from Stochastic Subspace Identification. *Mech Syst Signal Process* 22(4):948–969

- Reynders E, Degrauwe D, De Roeck G, Magalhaes F, Caetano E (2010) Combined experimental-operational modal testing of footbridges. *ASCE J Eng Mech* 136(6):687–696
- Reynders E, Wursten G, De Roeck G (2014) Output-only structural health monitoring in changing environmental conditions by means of nonlinear system identification. *Structural Health Monitoring* 13(1):82–93
- Spiridonakos M, Chatzi E (2014) Polynomial chaos expansion models for SHM under environmental variability. In: Cunha Á, Caetano E, Ribeiro P, Müller G (eds) *Proceedings of the 9th international conference on structural dynamics, EURODDYN 2014, Porto*, pp 2393–2398, June–July 2014
- Teughels A, De Roeck G (2004) Structural damage identification of the highway bridge Z24 by FE model updating. *J Sound Vib* 278(3):589–610
- Xia Y, Chen B, Weng S, Ni Y-Q, Xu Y-L (2012) Temperature effect on vibration properties of civil structures: a literature review and case studies. *J Civ Struct Heal Monit* 2(1):29–46

Vibrations of Beams for Seismic Response Estimation

Vittorio Gusella and Federico Cluni
Department of Civil and Environmental
Engineering, University of Perugia, Perugia, Italy

Introduction

The last decades have seen a growing interest for design and construction of tall buildings. In order to accurately estimate the dynamic response to environmental loads of such complex constructions, structural models are usually characterized by a large number of degrees of freedom; as a consequence, structural optimization algorithms, used to reduce the cost and the energy required to construct such buildings while maintaining appropriate safety levels, might be time consuming and even not feasible given the large dimension of the problem. As far as the global response is concerned, it is of interest to develop reduced models that are capable of giving approximate optimal solutions by significantly reducing the computation time. Therefore, a reduced model based on a Timoshenko beam will be presented in this paper which can be used to estimate the

response of tall buildings subjected to environmental loads, such those induced by earthquakes.

Flexural Beam Models

In the present section, the main beam models are recalled, focusing mainly on the Timoshenko model, which is able to describe the flexural behavior of the beam taking into account, besides the flexural stiffness and inertia due to deflections, also the shear deformability and the rotational inertia. Nevertheless, before introducing the Timoshenko beam model, the two models that can be considered to be the ones on which it is based, the Euler-Bernoulli and the simple shear models, will be briefly recalled.

The Cartesian reference system consists of the principal axes of the section, x and y , and the axis of the beam (the line where centroids of the sections lie), z . The beam is loaded with a distributed load $q(z)$, acting in the direction y , and a distributed moment $m(z)$ whose axis is in x direction (see Fig. 1). The analysis is limited to the plane y - z , and the only displacements of interest are those in y direction, the deflection $v(z)$. In the following, since the dependence on z is obvious, it will be dropped.

In what follows, the balance equations, which relate the loads q and m to the bending moment M and shear force V , will be used:

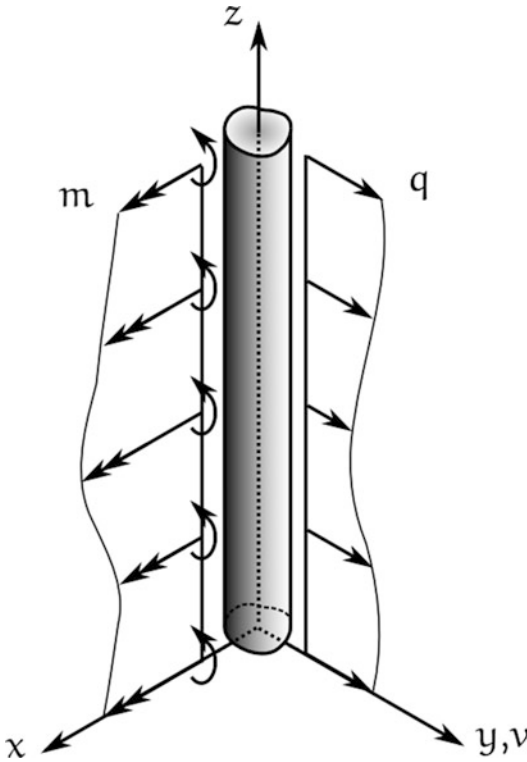
$$\frac{dM}{dz} = V - m \quad (1)$$

$$\frac{dV}{dz} = -q \quad (2)$$

The Euler-Bernoulli Beam

It is assumed that the deflection of the beam is due to the moment only, and therefore this behavior is known as “pure bending.” The constitutive equation which relate the bending moment M to the slope of deflection curve is

$$M = EI \frac{d\theta}{dz} \quad \text{with} \quad \theta = -\frac{dv}{dz} \quad (3)$$



Vibrations of Beams for Seismic Response Estimation, Fig. 1 Reference system of the beam model

where EI is the bending stiffness, the product of the Elastic (Young's) modulus, E , and the second moment of area (or moment of inertia) for rotation around x , I . The previous relation takes into account the fact that the sections of the beam remain plane in the deformed configuration and normal to the axis.

Using Eqs. 3 and 1, the following equation is obtained:

$$EI \frac{d^2 \theta}{dz^2} = V - m \Rightarrow -EI \frac{d^3 v}{dz^3} = V - m$$

Assuming $m = 0$, differentiating again, and using Eq. 2, the Euler-Bernoulli equation for the beam is obtained:

$$\frac{d^4 v}{dz^4} = \frac{q}{EI}$$

Moreover, the following are also valid:

$$M = -EI \frac{d^2 v}{dz^2}$$

$$V = -EI \frac{d^3 v}{dz^3}$$

The Euler-Bernoulli model can be extended to the dynamic case using the D'Alembert principle and adding to the external load q the load equivalent to inertia forces associated with deflections, q_i :

$$q_i = -\rho A \frac{\partial^2 v}{\partial t^2}$$

where ρ is the density of mass and A is the area of the section. Assuming ρ and A constant along the axis, the equation that rules the problem is

$$EI \frac{\partial^4 v}{\partial z^4} + \rho A \frac{\partial^2 v}{\partial t^2} = q$$

The free motion equation is

$$\frac{\partial^4 v}{\partial z^4} + \frac{\rho A}{EI} \frac{\partial^2 v}{\partial t^2} = 0$$

with solution that can be written through separation of variables as

$$v(z, t) = \tilde{v}(z) \cos(\omega t + \phi) \quad (4)$$

where

$$\tilde{v}(z) = D_1 \cos(kz) + D_2 \sin(kz) + D_3 \cosh(kz) + D_4 \sinh(kz)$$

and

$$k = \sqrt[4]{\frac{\omega^2 \rho A}{EI}}$$

The parameters D_1 , D_2 , D_3 , and D_4 can be obtained using the boundary conditions and determining ω to avoid the trivial solution. For example, in the case of cantilever beam clamped at $z = 0$ and free at $z = L$, the following boundary conditions hold

$$v(0,t) = 0, \quad \theta(0,t) = 0, \quad M(L,t) = 0, \quad V(L,t) = 0$$

and the values of ω are given by the following

$$\cos \sqrt[4]{\frac{\omega^2 \rho A}{EI}} L \cosh \sqrt[4]{\frac{\omega^2 \rho A}{EI}} L + 1 = 0$$

which has solutions

$$\omega_j = (a_j L)^2 \frac{1}{L^2} \sqrt{\frac{EI}{\rho A}}$$

with

$$a_1 L = 1.875, \quad a_2 L = 4.694, \quad a_3 L = 7.855, \\ a_j L \approx a_3 L + (j-3)2\pi \quad \text{for } j = 4, 5, \dots$$

The Simple Shear Beam

The simple shear beam model is based on the assumption that the beam undergoes only shear deformation. In this case, the constitutive equation which relates V to the slope of deflection curve is

$$V = GK\gamma \quad \text{with} \quad \gamma = \frac{dv}{dz} \quad (5)$$

where GK is the shear stiffness, the product of the tangential modulus, G , and shear area K . Moreover, the shear area is related to A by means of a *shear factor* κ , $K = \kappa A$, which depends on the shape of the section and is determined by means of energetic equivalence.

Using the preceding and Eq. 2, the following equation is obtained:

$$\frac{d^2 v}{dz^2} = -\frac{q}{GK}$$

In dynamics, by means of D'Alembert principle, the following equation of motion is obtained:

$$GK \frac{\partial^2 v}{\partial z^2} - \rho A \frac{\partial^2 v}{\partial t^2} = -q$$

and in free motion

$$\frac{\partial^2 v}{\partial z^2} - \frac{\rho}{GK} \frac{\partial^2 v}{\partial t^2} = 0$$

Adopting the separation of variables Eq. 4, the following expression for deflection is obtained:

$$\tilde{v}(z) = D_1 \cos(kz) + D_2 \sin(kz)$$

and

$$k = \sqrt{\frac{\omega^2 \rho}{GK}}$$

The parameters D_1 , D_2 can be obtained using the boundary conditions and determining ω to avoid the trivial solution. For example, in the case of cantilever beam clamped at $z = 0$ and free at $z = L$, the boundary conditions are

$$v(0,t) = 0, \quad V(L,t) = 0$$

and therefore

$$\omega_j = (2j-1) \frac{\pi}{2L} \sqrt{\frac{GK}{\rho}} \quad \text{for } j = 1, 2, \dots$$

The Timoshenko Beam Model

In the Timoshenko beam model (Timoshenko 1921, 1922), it is assumed that the slope of the deflection curve is the sum of those due to bending moment and those given by shear deformation, each one assumed to act alone

$$\frac{dv}{dz} = \gamma - \theta \quad (6)$$

The first part of both Eqs. 3 and 5 is unchanged, while the second parts are modified according to Eq. 6, and therefore

$$V = GK\gamma = GK \left(\frac{dv}{dz} + \theta \right), \quad M = EI \frac{d\theta}{dz}$$

Assuming EI and GK constant along the axis, the following equations are obtained:

$$\begin{cases} \text{GK} \left(\frac{d^2 v}{dz^2} + \frac{d\theta}{dz} \right) = -q \\ \text{EI} \frac{d^2 \theta}{dz^2} - \text{GK} \left(\frac{dv}{dz} + \theta \right) = -m \end{cases} \quad \begin{cases} \text{GK} \left(\frac{\partial^2 v}{\partial z^2} + \frac{\partial \theta}{\partial z} \right) - \rho A \frac{\partial^2 v}{\partial t^2} = 0 \\ \text{EI} \frac{\partial^2 \theta}{\partial z^2} - \text{GK} \left(\frac{\partial v}{\partial z} + \theta \right) - \rho I \frac{\partial^2 \theta}{\partial t^2} = 0 \end{cases}$$

In dynamics the D'Alembert principle is used again. Moreover, in the Timoshenko beam model, the inertial forces associated to rotations θ are also taken into account

$$m_i = -\rho I \frac{\partial^2 \theta}{\partial t^2}$$

and the equations of motion are therefore

$$\begin{cases} \text{GK} \left(\frac{\partial^2 v}{\partial z^2} + \frac{\partial \theta}{\partial z} \right) - \rho A \frac{\partial^2 v}{\partial t^2} = -q \\ \text{EI} \frac{\partial^2 \theta}{\partial z^2} - \text{GK} \left(\frac{\partial v}{\partial z} + \theta \right) - \rho I \frac{\partial^2 \theta}{\partial t^2} = -m \end{cases} \quad (7)$$

In free motion the following equations hold:

and considering that

$$\begin{aligned} \theta &= \gamma - \frac{\partial v}{\partial z} \Rightarrow \frac{\partial \theta}{\partial z} = \frac{\partial \gamma}{\partial z} - \frac{\partial^2 v}{\partial z^2} = \frac{1}{\text{GK}} \frac{\partial V}{\partial z} - \frac{\partial^2 v}{\partial z^2} \\ &= -\frac{\rho}{\text{GK}} \frac{\partial^2 v}{\partial t^2} - \frac{\partial^2 v}{\partial z^2} \end{aligned}$$

the following equation can be obtained:

$$\text{EI} \frac{\partial^4 v}{\partial z^4} + \rho A \frac{\partial^2 v}{\partial t^2} - \rho I \left(1 + \frac{E}{\text{GK}} \right) \frac{\partial^4 v}{\partial z^2 \partial t^2} + \frac{\rho^2 I}{\text{GK}} \frac{\partial^4 v}{\partial t^4} = 0$$

The solution is given by

$$v(z, t) = \tilde{v}(z) \cos(\omega t + \phi)$$

where

$$\tilde{v}(z) = \begin{cases} D_1 \cos(k_1 z) + D_2 \sin(k_1 z) + D_3 \cos h(k_2 z) + D_4 \sin h(k_2 z) & \omega < \omega_c \\ D_1 \cos(k_1 z) + D_2 \sin(k_1 z) + D_3 \cos(k_2 z) + D_4 \sin(k_2 z) & \omega > \omega_c \end{cases} \quad (8)$$

where ω_c is a cut-off frequency given by

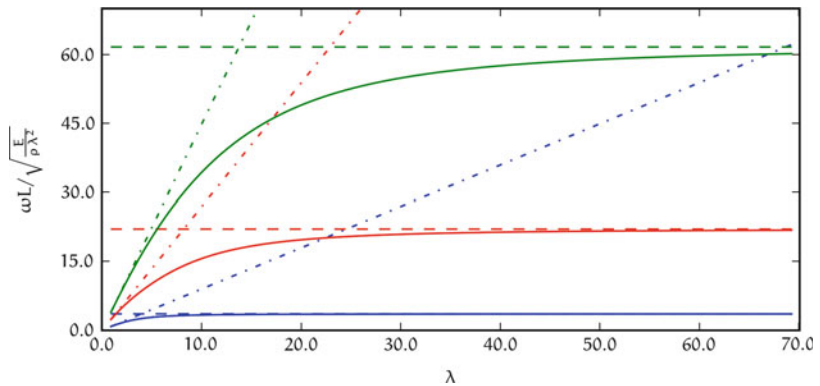
$$\omega_c = \frac{\text{GK}A}{\rho I}$$

As usual, ω is found by imposing the suitable boundary conditions.

It can be interesting to plot the values of ω in terms of variation of slenderness $\lambda = L/r$ where $r = \sqrt{I/A}$, as in Fig. 2 (the values have

Vibrations of Beams for Seismic Response Estimation,

Fig. 2 Pulsations of the first (blue), second (red), and third (green) mode of cantilever beam with $\kappa = 0.85$ and $E/G = 2.6$: (—) Timoshenko beam model, (— —) Euler-Bernoulli beam model, (— · —) simple shear model



been obtained assuming $\kappa = 0.85$ and $E/G = 2.6$). It can be noted that as λ increases, the values ω of the Euler-Bernoulli model are achieved

$$\frac{\omega_j L}{\sqrt{\frac{E}{\rho \lambda^2}}} = (a_j L)^2$$

while as λ decreases the values of the simple shear model are recovered

$$\frac{\omega_j L}{\sqrt{\frac{E}{\rho \lambda^2}}} = (2j - 1) \frac{\pi}{2} \sqrt{\frac{G}{E}} \kappa \cdot \lambda$$

It is worth noting that the convergence to the Euler-Bernoulli or simple shear beam tends to slow down as mode number increases. For intermediate values of slenderness, the natural frequencies ω depend on both the bending stiffness EI and the shear stiffness GK .

Example The case of a cantilever beam with the following characteristics:

$$\begin{array}{lll} A = 0.120 \text{ mm}^2, & I = 1.60 \cdot 10^{-3} \text{ mm}^4 & K = 0.102 \text{ mm}^2 \\ E = 30,000 \text{ N/mm}^2, & G = 11,538 \text{ N/mm}^2 & \\ \rho = 2,500 \text{ kg/m}^3 & L = 2.0 \text{ m} & \omega_c = 542.43 \text{ rad/s} \end{array}$$

is considered in what follows.

The first 12 modal shapes are shown in Fig. 3, and the parameters of Eq. 8 are shown in Tables 1 and 2. In the tables also the phase velocity, defined as the ratio between the angular frequency, ω , and the wavenumber, k_1 , is reported: as well known, the phase velocity is unbounded for the Euler-Bernoulli beam, while in the case of Timoshenko beam the two different families of waves present when $\omega > \omega_c$, with wavenumbers k_1 and k_2 , tend to the following values (Hagedorn and DasGupta 2007):

$$\begin{aligned} \lim_{j \rightarrow +\infty} \frac{\omega_j}{k_1} &= \sqrt{\frac{G\kappa}{\rho}} = 62.63 \text{ rad/s} \\ \lim_{j \rightarrow +\infty} \frac{\omega_j}{k_2} &= \sqrt{\frac{E}{\rho}} = 109.54 \text{ rad/s} \end{aligned}$$

where j is the mode number.

Deterministic Excitation-Response Relations

The unit impulse applied at the section $z = z_0$ at the instant $t = t_0$ is considered:

$$f(z, t) = \delta(z - z_0) \delta(t - t_0)$$

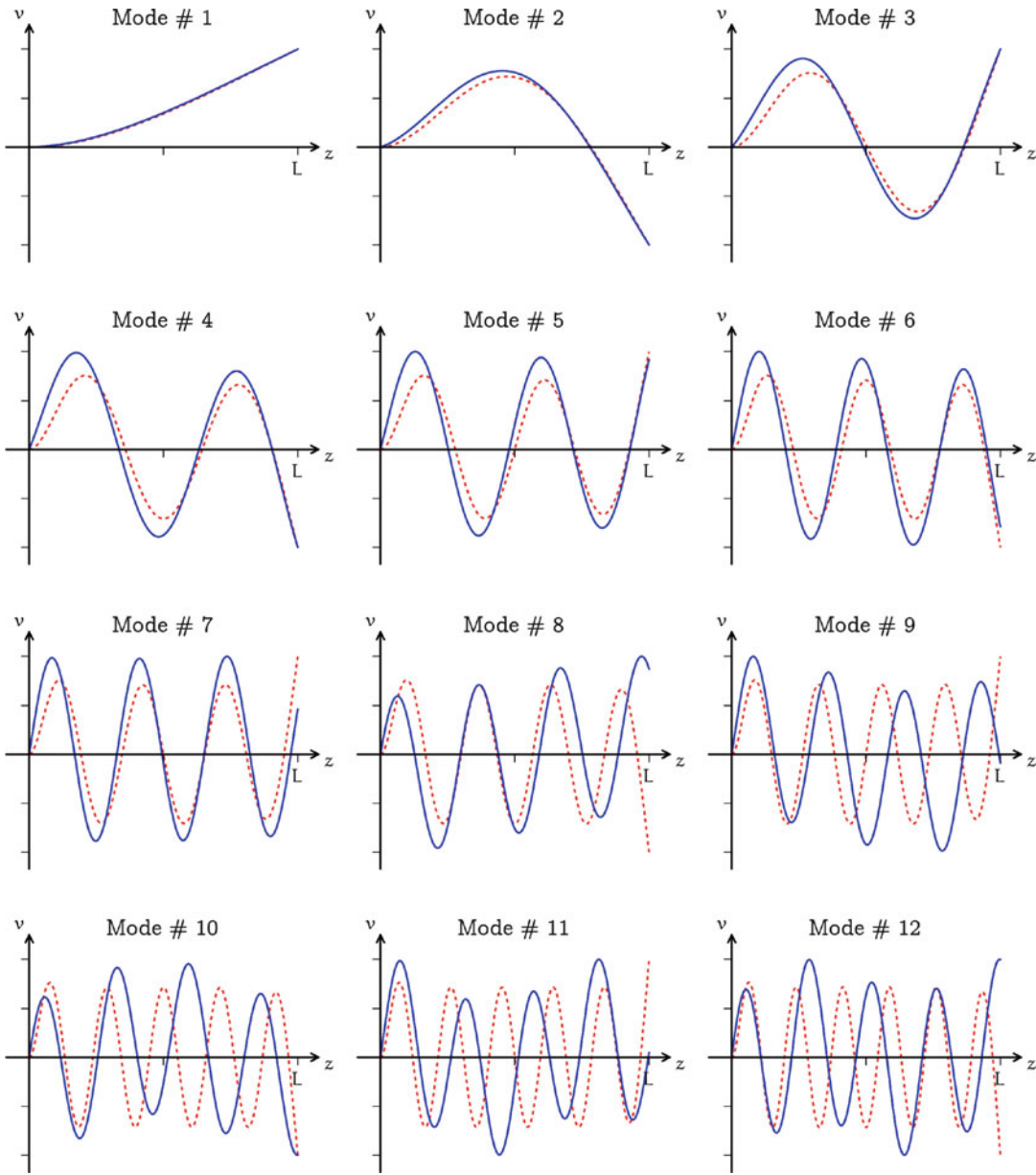
Assuming that prior of the unit impulse application the beam is at rest with zero displacement, the response can be expressed by (Robson et al. 1971; Crandal 1979; Lin 1967; Elishakoff 1988)

$$v(z, t) = h(z, t - t_0; z_0)$$

where $h(z, t - t_0; z_0)$ is the *impulse response function* or *Green's function*. A $f(z, t)$ arbitrary space-time action can be regarded as a superposition of unit impulses at the time τ in generic section $0 \leq \xi \leq L$

$$f(z, t) = \int_{-\infty}^{+\infty} d\tau \int_0^L f(\xi, t - \tau) \delta(z - \xi) \delta(\tau) d\xi$$

If the impulse response function is known, then the response at the time t in the generic section $0 \leq \xi \leq L$ can be obtained by the following convolution integral or superposition integral:



Vibrations of Beams for Seismic Response Estimation, Fig. 3 First 12 modal shapes of the Euler-Bernoulli beam (dashed red) and Timoshenko beam (solid blue)

$$v(z,t) = \int_{-\infty}^{+\infty} d\tau \int_0^L f(\xi,t-\tau)h(z,\tau;\xi)d\xi$$

The unit complex sinusoidal action with the fixed frequency ω applied at the section $z = z_o$ is

$$f(z,t) = \delta(z-z_o)e^{i\omega t}$$

Given the *complex frequency response function* $H(z,\omega;z_o)$, the response is given by

$$v(z,t) = H(z,\omega;z_o)e^{i\omega t}$$

Vibrations of Beams for Seismic Response Estimation, Table 1 Parameters of first 12 modes of Euler-Bernoulli beam

#	ω	k_1	k_2	D_1	D_2	D_3	D_4	ω/k_1
1	11.12	0.94	0.94	-5.00e-01	3.67e-01	5.00e-01	-3.67e-01	11.86
2	69.68	2.35	2.35	5.00e-01	-5.09e-01	-5.00e-01	5.09e-01	29.69
3	195.09	3.93	3.93	-5.00e-01	5.00e-01	5.00e-01	-5.00e-01	49.68
4	382.29	5.50	5.50	5.00e-01	-5.00e-01	-5.00e-01	5.00e-01	69.54
5	631.91	7.07	7.07	-5.00e-01	5.00e-01	5.00e-01	-5.00e-01	89.40
6	943.90	8.64	8.64	5.00e-01	-5.00e-01	-5.00e-01	5.00e-01	109.26
7	1318.23	10.21	10.21	-5.00e-01	5.00e-01	5.00e-01	-5.00e-01	129.11
8	1754.87	11.78	11.78	5.04e-01	-5.00e-01	-5.00e-01	5.00e-01	148.96
9	2253.79	13.35	13.35	-5.04e-01	5.00e-01	5.00e-01	-5.00e-01	168.80
10	2814.93	14.92	14.91	5.07e-01	-5.00e-01	-5.00e-01	5.00e-01	188.64
11	3438.27	16.49	16.48	-5.08e-01	5.01e-01	4.99e-01	-4.99e-01	208.47
12	4123.74	18.06	18.05	5.12e-01	-5.01e-01	-5.00e-01	5.00e-01	228.29

Vibrations of Beams for Seismic Response Estimation, Table 2 Parameters of first 12 modes of Timoshenko beam

#	ω	k_1	k_2	D_1	D_2	D_3	D_4	ω/k_1	ω/k_2
1	10.79	0.93	0.91	-4.97e-01	3.77e-01	4.97e-01	-3.60e-01	11.55	
2	58.20	2.28	2.01	4.59e-01	-6.12e-01	-4.59e-01	4.76e-01	25.56	-
3	138.73	3.79	2.80	-3.92e-01	7.10e-01	3.92e-01	-3.89e-01	36.63	-
4	230.57	5.27	3.13	3.02e-01	-8.39e-01	-3.02e-01	3.04e-01	43.75	-
5	328.14	6.76	3.05	-2.01e-01	8.86e-01	2.01e-01	-1.99e-01	48.52	-
6	426.14	8.24	2.53	-1.14e-01	9.27e-01	1.14e-01	-1.16e-01	51.74	-
7	520.37	9.65	1.20	-5.98e-02	9.30e-01	5.98e-02	-3.96e-02	53.95	-
8	568.55	10.37	1.36	2.13e-01	7.69e-01	-2.13e-01	3.14e-02	54.84	417.59
9	620.33	11.14	2.44	-1.73e-01	8.02e-01	1.73e-01	4.98e-02	55.66	254.07
10	654.90	11.66	3.00	-2.04e-01	-7.63e-01	2.04e-01	-5.23e-02	56.15	218.09
11	726.85	12.75	4.02	-2.34e-01	7.24e-01	2.34e-01	5.37e-02	57.01	180.80
12	762.49	13.29	4.48	1.72e-01	8.02e-01	-1.72e-01	5.99e-02	57.38	170.13

Let $F(z, \omega)$ be the Fourier transform of the action $f(z, t)$ which is assumed to be an integrable function, so that

$$F(z, \omega) = \int_{-\infty}^{+\infty} f(z, t) e^{-i\omega t} dt$$

and

$$f(z, t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(z, \omega) e^{+i\omega t} d\omega$$

The complex frequency response function is the Fourier transform of the impulse response function and

$$H(z, \omega; z_0) = \int_{-\infty}^{+\infty} h(z, \tau; z_0) e^{-i\omega \tau} d\tau$$

$$h(z, \tau; z_0) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} H(z, \omega; z_0) e^{i\omega \tau} d\omega$$

Introduced the Fourier transform of the response

$$V(z, \omega) = \int_{-\infty}^{+\infty} v(z, t) e^{-i\omega t} dt$$

$$v(z, t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} V(z, \omega) e^{+i\omega t} d\omega$$

the excitation-response relation for Fourier transform is

$$V(z, \omega) = \int_0^L H(z, \omega; \xi) F(\xi, \omega) d\xi$$

so that

$$v(z, t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{i\omega t} d\omega \int_0^L H(z, \omega; \xi) d\xi \int_{-\infty}^{+\infty} f(\xi, \tau) e^{-i\omega \tau} d\tau$$

and

$$V(z, \omega) = \int_{-\infty}^{+\infty} e^{-i\omega \tau} d\tau \int_0^L h(z, \tau; \xi) d\xi \frac{1}{2\pi} \int_{-\infty}^{+\infty} F(\xi, \omega) e^{+i\omega \tau} d\omega$$

Statistical Characteristics of the Excitation

Let $f(z, t)$ be a random space-time processes. In order to describe this processes, it is introduced the *ensemble average* or *expected value*

$$E[f(z, t)] = m_f(z, t)$$

the *space-time correlation*

$$E[f(z_1, t_1) f(z_2, t_2)] = R_f(z_1, z_2, t_1, t_2)$$

and the *space-time covariance*

$$\begin{aligned} E[(f(z_1, t_1) - m_f(z_1, t_1)) (f(z_2, t_2) - m_f(z_2, t_2))] \\ = \Gamma_f(z_1, z_2, t_1, t_2) \end{aligned}$$

It is assumed that the previous characteristics are sufficient to describe the excitation.

Considering the case of weak stationary process, the mean is independent on t , and

the space-time correlation depends on $\tau = t_2 - t_1$ so that

$$\begin{aligned} E[f(z, t)] &= m_f(z) \\ E[f(z_1, t) f(z_2, t + \tau)] &= R_f(z_1, z_2, \tau) \\ E[(f(z_1, t) - m_f(z_1)) (f(z_2, t + \tau) - m_f(z_2))] &= \\ = \Gamma_f(z_1, z_2, \tau) &= R_f(z_1, z_2, \tau) - m_f(z_1) m_f(z_2) \end{aligned}$$

If $z_1 = z_2 = z$, the *autocovariance function* is obtained

$$\Gamma_f(z, \tau) = R_f(z, \tau) - m_f^2(z)$$

that for $\tau = 0$ reduces to the *variance*

$$\sigma_f^2(x) = E[f^2(z)] - m_f^2(z)$$

where $R_f(z, \tau)|_{\tau=0} = E[f^2(z)]$ is the *mean-square value*.

Let $S_f(z_1, z_2, \omega)$ be the *space-time cross-spectral density function*; this function is connected to the cross-correlation function by the *Wiener-Khintchine relations*:

$$\begin{aligned} S_f(z_1, z_2, \omega) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} R_f(z_1, z_2, \tau) e^{-i\omega \tau} d\tau \\ R_f(z_1, z_2, \tau) &= \int_{-\infty}^{+\infty} S_f(z_1, z_2, \omega) e^{i\omega \tau} d\omega \end{aligned}$$

If $z_1 = z_2 = z$ and $\tau = 0$, the mean-square value can be expressed by the cross-spectral density function

$$E[f^2(z)] = \int_{-\infty}^{+\infty} S_f(z, z, \omega) d\omega$$

and for excitation with zero mean

$$\sigma_f^2(x) = \int_{-\infty}^{+\infty} S_f(z, z, \omega) d\omega$$

It is worth noting that earthquake-induced base accelerations are samples of non-stationary random process. Anyway, there is an interval,

called *strong motion phase*, where the base accelerations can be considered samples of stationary random process. The length of strong motion phase is T_S . It is commonly assumed that, if T_S is much greater than the period of the elementary oscillator, both the base accelerations and the response can be considered samples of stationary random process with zero mean.

Statistical Characteristics of the Response

The statistical characteristics of the response can be obtained from corresponding characteristics of the excitation by taking the averages of the previously introduced relations. The expected value of the response can be obtained by equation

$$E[v(z, t)] = E \left[\int_{-\infty}^{+\infty} d\tau \int_0^L f(\xi, t - \tau) h(z, \tau; \xi) d\xi \right]$$

and interchanging the integral operators and mean operator, the following equation is obtained:

$$E[v(z, t)] = \int_{-\infty}^{+\infty} d\tau \int_0^L E[f(\xi, t - \tau)] h(z, \tau; \xi) d\xi$$

where $E[f(z, t)]$ is the mean of the excitation. If the excitation is a stationary processes, its mean value is constant, $E[f(z, t)] = m_f(z)$, and also the mean value of the response is constant, $E[v(z, t)] = m_v(z)$, where

$$m_v(z) = \int_{-\infty}^{+\infty} d\tau \int_0^L m_f(\xi) h(z, \tau; \xi) d\xi$$

If in addition $m_f(z) = 0$, as for the seismic excitation, then $m_v(z) = 0$.

The space-time correlation function of the response is obtained by average of the product of $v(z_1, t)$ and $v(z_2, t + \tau)$, which are given by the impulse response function

$$E[v(z_1, t), v(z_2, t + \tau)] = E \left[\int_{-\infty}^{+\infty} d\theta_1 \int_0^L f(\xi_1, t - \theta_1) h(z_1, \theta_1; \xi_1) d\xi_1 \times \int_{-\infty}^{+\infty} d\theta_2 \int_0^L f(\xi_2, t + \tau - \theta_2) h(z_2, \theta_2; \xi_2) d\xi_2 \right]$$

and interchanging the integral operators and mean operator

$$E[v(z_1, t), v(z_2, t + \tau)] = \int_{-\infty}^{+\infty} d\theta_1 \int_{-\infty}^{+\infty} d\theta_2 \int_0^L d\xi_1 \times \int_0^L E[f(\xi_1, t - \theta_1) f(\xi_2, t + \tau - \theta_2)] h(z_1, \theta_1; \xi_1) h(z_2, \theta_2; \xi_2) d\xi_2$$

If the excitation is stationary, then its space-time correlation is independent of t and so is the response space-time correlation

$$R_v(z_1, z_2, \tau) = \int_{-\infty}^{+\infty} d\theta_1 \int_{-\infty}^{+\infty} d\theta_2 \int_0^L d\xi_1 \times \int_0^L R_f(\xi_1, \xi_2, \tau + \theta_1 - \theta_2) h(z_1, \theta_1; \xi_1) h(z_2, \theta_2; \xi_2) d\xi_2$$

Applying the Wiener-Khintchine relations, the space-time cross-spectral density of the response can be obtained by the following:

$$S_v(z_1, z_2, \omega) = \int_0^L d\xi_1 \int_0^L d\xi_2 \int_{-\infty}^{+\infty} h(z_1, \theta_1; \xi_1) d\theta_1 \int_{-\infty}^{+\infty} h(z_2, \theta_2; \xi_2) d\theta_2 \times \frac{1}{2\pi} \int_{-\infty}^{+\infty} R_f[\xi_1, \xi_2, \tau + \theta_1 - \theta_2] e^{-i\omega\tau} d\tau$$

Introducing in the previous integrals the unity products $e^{-i\omega\theta_1} e^{i\omega\theta_1}$ and $e^{-i\omega\theta_2} e^{i\omega\theta_2}$, by the frequency response function

$$H(z_1, -\omega; \xi_1) = \int_{-\infty}^{+\infty} h(z_1, \theta_1; \xi_1) e^{i\omega\theta_1} d\theta_1$$

$$H(z_2, \omega; \xi_2) = \int_{-\infty}^{+\infty} h(z_2, \theta_2; \xi_2) e^{-i\omega\theta_2} d\theta_2$$

and introducing the space-time cross-spectral density of the excitation,

$$S_f(\xi_1, \xi_2, \omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} R_f[\xi_1, \xi_2, \tau + \theta_1 - \theta_2] e^{-i\omega(\tau + \theta_1 - \theta_2)} d(\tau + \theta_1 - \theta_2)$$

the following expression to determine $S_v(z_1, z_2, \omega)$ is obtained:

$$S_v(z_1, z_2, \omega) = \int_0^L \int_0^L H(z_1, -\omega, \xi_1) S_f[\xi_1, \xi_2, \omega] H(z_2, \omega, \xi_2) d\xi_1 d\xi_2$$

Natural Modes and Frequencies

The dynamic proprieties of the systems can be described in terms of *natural modes*, $\tilde{v}$, and corresponding *natural frequencies*, ω . In what follows, the natural modes will be expressed in terms of both deflection, $v_j(z) = \tilde{v}(z)$, and rotation, $\theta_j(z)$, where j is the mode number. Once the natural modes have been found, the equations of motion Eq. 7 can be reduced to

$$m_j \frac{d^2 g_j}{dt^2} + k_j g_j = f_j$$

where m_j , k_j , and f_j are, respectively, the generalized mass, stiffness, and force corresponding to mode j , given by

$$\begin{cases} m_j = \int_0^L (\rho A v_j^2 + \rho I \theta_j^2) dz \\ k_j = \int_0^L \left(-EI \left(\frac{d\theta_j}{dz} \right)^2 + GK \left(\frac{d\gamma_j}{dz} \right)^2 \right) dz, \quad \text{with } k_j = \omega_j^2 m_j \\ f_j = \int_0^L (q v_j + m \theta_j) dz \end{cases} \quad (9)$$

and g_j is the j th principal coordinate. In the preceding, the following orthogonality properties of

the modes have been used (Bishop and Price 1977):

$$\begin{cases} \int_0^L (\rho A v_j v_k + \rho I \theta_j \theta_k) dz = a_{kj} \delta_{kj} \\ \int_0^L \left(-EI \frac{d\theta_j}{dz} \frac{d\theta_k}{dz} + GK \frac{d\gamma_j}{dz} \frac{d\gamma_k}{dz} \right) dz = a_{kj} \omega_k^2 \delta_{kj}, \quad \text{with } \delta_{kj} = \begin{cases} 1 & \text{if } k = j \\ 0 & \text{if } k \neq j \end{cases} \end{cases}$$

Once the g_j are known, the values of displacements are given by

$$h_j(t) = \begin{cases} 0 & t \leq 0 \\ \frac{1}{m\omega_j\sqrt{1-v_j^2}} e^{-v_j\omega_j t} \sin(\omega_j\sqrt{1-v_j^2} t) & t > 0 \end{cases}$$

$$v(z, t) = \sum_{j=1}^{\infty} v_j(z) g_j(t), \quad \theta(z, t) = \sum_{j=1}^{\infty} \theta_j(z) g_j(t) \quad \text{so that}$$

If a modal damping v_j is assumed, the equations of motion are

$$h(z, t; z_0) = \sum_{j=1}^{\infty} v_j(z) v_j(z_0) h_j(t)$$

$$m_j \frac{d^2 g_j}{dt^2} + 2v_j \sqrt{k_j m_j} \frac{dg_j}{dt} + k_j g_j = f_j$$

and using the Fourier transform relation, the complex frequency response is

that can also be written as

$$H(z, \omega; z_0) = \sum_{j=1}^{\infty} v_j(z) v_j(z_0) H_j(\omega)$$

$$\frac{d^2 g_j}{dt^2} + 2v_j \omega_j \frac{dg_j}{dt} + \omega_j^2 g_j = \frac{1}{m_j} f_j$$

where the single degree of freedom unit complex frequency response function is given by

Let $h_j(t)$ be the single degree of freedom unit impulse response function. With initial conditions $g_i = 0$ and $dg_j/dt = 0$, and assuming distributed load $q = \delta(z - z_0) \delta(t)$ and distributed moments $m = 0$, the solution is

$$H_j(\omega) = \frac{1}{m(\omega_j^2 - \omega^2 + 2v_j \omega_j \omega i)}$$

$$g_j(t; z_0) = h_j(t) v_j(z_0)$$

Inserting in the previous equations, the following expressions for space-time correlation and space-time cross-spectral density are obtained:

where

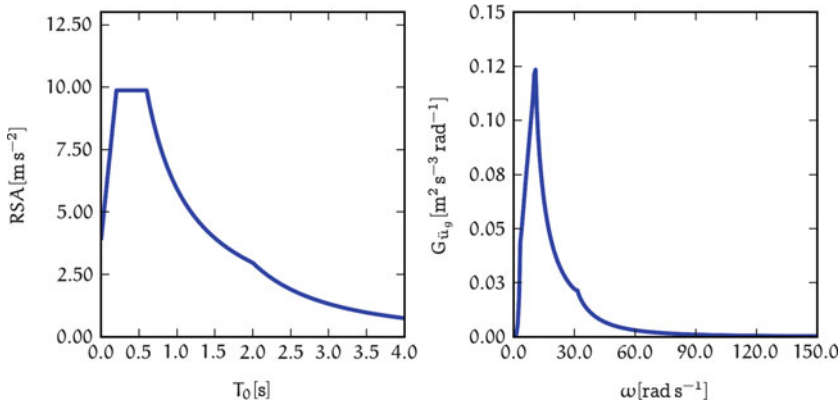
$$R_v(z_1, z_2, \tau) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} v_j(z_1) v_k(z_2) \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h_j(\theta_1) h_k(\theta_2) d\theta_1 d\theta_2 \times \int_0^L \int_0^L v_j(\xi_1) v_k(\xi_2) R_f[\xi_1, \xi_2, \tau + \theta_1 - \theta_2] d\xi_1 d\xi_2 \quad (10)$$

$$S_v(z_1, z_2, \omega) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} v_j(z_1) v_k(z_2) H_j(-\omega) H_k(+\omega) \times \int_0^L \int_0^L v_j(\xi_1) v_k(\xi_2) S_f[\xi_1, \xi_2, \omega] d\xi_1 d\xi_2 \quad (11)$$

Application to the Beam Under Seismic Load

The concepts illustrated in the previous sections are applied to the case of the cantilever beam seen previously in the example, subjected to a seismic

load. The seismic load is modeled by means of an elastic response spectrum according to the European standard for seismic actions (Eurocode 8 2004). It has been assumed a type 1 elastic response spectrum with a ground type C, with parameters:



Vibrations of Beams for Seismic Response Estimation, Fig. 4 Elastic response spectrum according to Eurocode 8 and corresponding power spectral density function

$$\begin{aligned} a_g &= 0.355 \text{ g (Peak ground acceleration)}, & S &= 1.15 \text{ (Soil factor)} \\ T_B &= 0.20 \text{ s}, & T_C &= 0.60 \text{ s}, & T_D &= 2.00 \text{ s} \end{aligned}$$

A 5 % damping, $\nu = 0.05$, was adopted. The response spectrum in terms of pseudo-accelerations (RSA) is shown in Fig. 4a.

In order to apply what presented in previous sections, the power spectral density function, $G_{\ddot{u}_g}$, corresponding to the assumed response spectrum is needed. The method proposed in (Cacciola et al. 2004) is used. The time observing window (coincident with the strong motion phase) T_S has been set to 10s, while the cut-off frequency is $\omega_u = 150 \text{ rad/s}$ and the frequency step is $\Delta\omega = 2\pi/T_S = 0.63 \text{ rad/s}$. An iterative scheme has been used as suggested in (Cacciola 2010), and the power spectral density function obtained is shown in Fig. 4b.

The generation of a time history of base acceleration is achieved by means of superposition of N_a harmonics with random phases Φ_i as follows:

$$\ddot{u}_g(t) = \varphi(t) \sum_{i=1}^{N_a} \sqrt{2G_{\ddot{u}_g(i\Delta\omega)}\Delta\omega} \cos(i\Delta\omega t + \Phi_i)$$

where $\varphi(t)$ is the modulating function (which changes the total energy but not its distribution among the frequencies; see Fig. 4b) (Cacciola 2010):

$$\varphi(t) = \begin{cases} \left(\frac{t}{t_1}\right)^2 & \text{for } t \leq t_1 \\ 1 & \text{for } t_1 < t \leq t_2 \\ e^{-\beta(t-t_2)} & \text{for } t > t_2 \end{cases}$$

with $t_2 = t_1 + T_S$. In the present case, it was assumed $t_1 = 10\text{s}$ and $\beta = 0.3 \text{ s}^{-1}$. An example of the generated $\ddot{u}_g(t)$ is shown in Fig. 5.

The problem of the beam subjected to the base acceleration given by $\ddot{u}_g(t)$ can be solved by means of modal analysis, as formulated in Eq. 9.

In the case of seismic action represented by base accelerations $\ddot{u}_g(t)$, the external loads are

$$\begin{cases} q = -\rho A \ddot{u}_g(t) \\ m = 0 \end{cases}$$

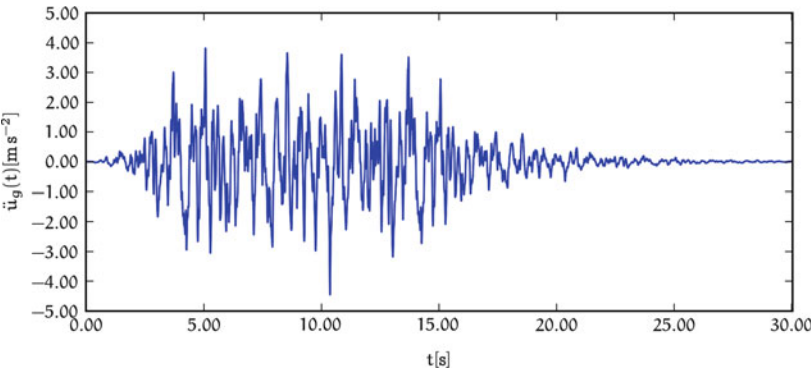
and therefore:

$$\begin{aligned} f_j &= -\int_0^L \rho A \ddot{u}_g(t) v_j dz = -\Gamma_j \ddot{u}_g(t), \text{ with} \\ \Gamma_j &= \int_0^L \rho A v_j dz \end{aligned}$$

Once the time histories of g_j have been determined, the solution is given by

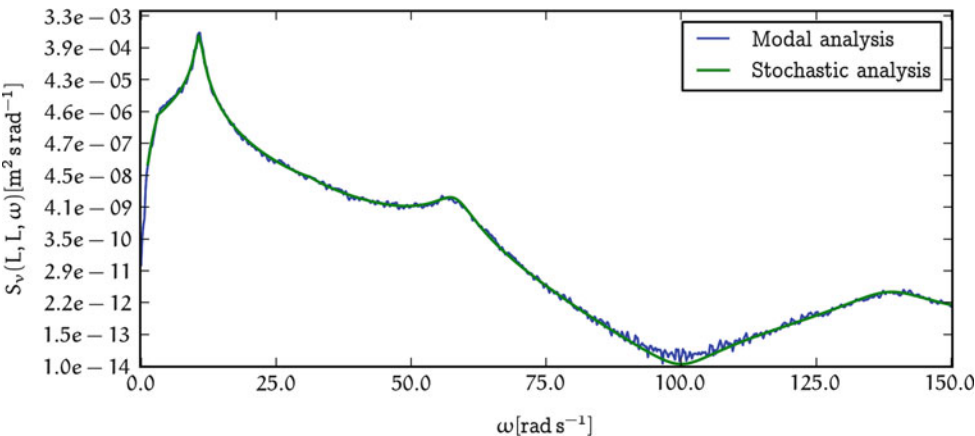
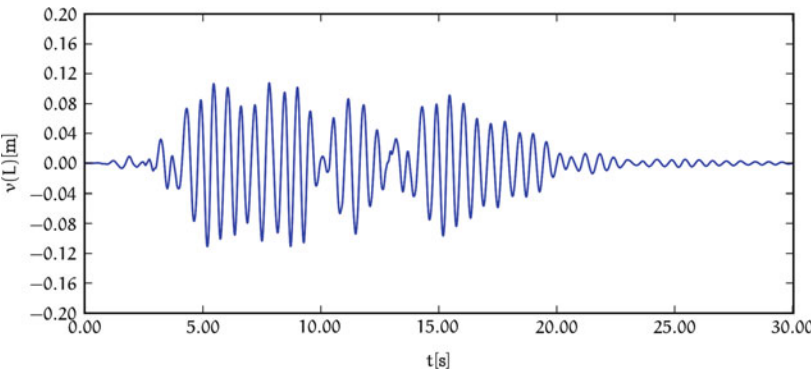
Vibrations of Beams for Seismic Response Estimation,

Fig. 5 Example of a time history of acceleration compatible to power spectral density function $G_{\ddot{u}_g}$



Vibrations of Beams for Seismic Response Estimation,

Fig. 6 Displacement at free end for base accelerations of Fig. 5



Vibrations of Beams for Seismic Response Estimation, Fig. 7 Power spectral density function of free end displacement by means of modal analysis with generated

spectrum compatible accelerograms and by means of stochastic analysis

$$v(z, t) = \sum_{j=1}^{N_p} v_j(z) g_j(t), \quad \theta(z, t) = \sum_{j=1}^{N_p} \theta_j(z) g_j(t)$$

where N_p is the number of modes considered in the solution. For example, the displacement at the free end, $v(L)$, for the base acceleration shown in Fig. 5 and assuming $N_p = 10$ (the modal contribution tends to be negligible quite rapidly) is shown in Fig. 6.

The power spectral density function of the displacement at free end, $S_v(L, L, \omega)$ is shown in Fig. 7, compared to that evaluated by mean of Eq. 11. As can be appreciated, the same results are obtained using both procedures.

Cross-References

- Random Process as Earthquake Motions
- Stochastic Analysis of Linear Systems
- Stochastic Ground Motion Simulation
- Time History Seismic Analysis

References

- Bishop R, Price W (1977) Coupled bending and twisting of a Timoshenko beam. *J Sound Vib* 50:469–477
- Cacciola P (2010) A stochastic approach for generating spectrum compatible fully nonstationary earthquakes. *Comput Struct* 88:889–901
- Cacciola P, Colajanni P, Muscolino G (2004) Combination of modal responses consistent with seismic input representation. *J Struct Eng* 130:47–55
- Crandal SH (1979) Random vibration of one- and two-dimensional structures. In: Krishnaiah PR (ed) *Developments in statistics*, vol 2. Academic, Waltham
- Elishakoff I (1988) Random vibration of discrete and continuous systems. In: Schiehlen W, Wedig W (eds) *Analysis and estimation of stochastic mechanical systems*, International centre for mechanical sciences – CISM, courses and lectures no. 303. Springer, Berlin
- Eurocode 8 (2004) Design of structures for earthquake resistance – part 1: general rules, seismic actions and rules for buildings. European Committee for Standardization (CEN), Bruxelles
- Hagedorn P, DasGupta A (2007) *Vibrations and waves in continuous mechanical systems*. Wiley, Chichester
- Lin YK (1967) *Probabilistic theory of structural dynamics*. McGraw-Hill, New York
- Robson DJ, Dodds CJ, Macvean DB, Paling VR (1971) Random vibration, International centre for mechanical sciences – CISM, courses and lectures no. 115. Springer, Berlin
- Timoshenko SP (1921) On the correction factor for shear of the differential equation for transverse vibrations of bars of uniform cross-section. *Philos Mag* 41:744–746
- Timoshenko SP (1922) On the transverse vibrations of bars of uniform cross-section. *Philos Mag* 43:125–131

Volcanic Eruptions, Real-Time Forecasting of

Andrew F. Bell¹, Christopher R. J. Kilburn² and Ian G. Main¹

¹School of GeoSciences, University of Edinburgh, Edinburgh, UK

²Department of Earth Sciences, Aon Benfield UCL Hazard Centre, University College London, London, UK

Synonyms

Eruption precursors; Event-tree eruption forecasts; Probabilistic and deterministic eruption forecasting

Introduction

One of the primary goals of volcanology is to develop methods to forecast future eruptive activity. Good forecasts have the potential to increase the effectiveness of hazard mitigation plans, allowing well-timed evacuations or improved management of a return to exclusion zones. Volcanic earthquakes and seismicity are key datasets that form the basis for eruption forecasts. Seismometers are amongst most widely installed volcano monitoring equipment. Increased seismicity is a primary indicator of volcanic unrest, and eruptions are frequently preceded by changes in the rate, location, magnitude, and type of earthquakes. In combination with geodetic data, they can provide indications as to the volumes and locations of magma.

Earthquakes provide an insight into otherwise hidden processes occurring at volcanoes and that may have a causal link to eruptive activity, such

as edifice deformation, magma migration, and magma failure. Consequently, changes in the seismic activity at a volcano can be expected to provide information on the state of unrest, whether unrest is likely to lead to an eruption, and, if so, when. It may also in future provide information about changes in the nature of eruptive activity and the style, size, and location of impending activity.

A forecast is a probabilistic estimate of future activity based on current observations and past behavior. It needs to include aspects of timing, as well as physical attributes of the activity itself. A forecasting model quantitatively links observations to these probabilistic estimates, by applying empirical statistics or physical and theoretical considerations. But, how well do different forecasting models perform? Many evaluations of forecast performance have been undertaken retrospectively. These types of analyses are important for developing and refining methods and for identifying the action of different processes. However, the true performance of forecasts can only be fully evaluated through real-time forecasting (i.e., with forecasts issued ahead of the eruption) and rigorous testing. This approach eliminates conscious or subconscious biases due to data, model, and parameter selection.

Forecasting Eruptive Activity Using Seismicity Data

What Is a Forecast?

Although the words “forecast” and “prediction” have similar dictionary definitions, the associated connotations are different in different disciplines of geophysics. In some instances, the words are used interchangeably. For example, the US National Weather Service has a “Weather Prediction Center” which issues “Short Range Forecasts” of weather conditions (<http://www.hpc.ncep.noaa.gov/>). This topic has been heavily debated in the earthquake community. The term “prediction” has come to mean a highly specific statement about the time, location, and magnitude of a future event, generally with very narrow

confidence bounds. In contrast, a “forecast” is a more probabilistic statement, with a greater (and, perhaps, better quantified) uncertainty. It is broadly accepted that reliable and accurate earthquake “prediction” is unlikely to be a possibility, except in exceptional circumstances (Jordan et al. 2011).

In volcanology, similar distinctions have been made between more specific, shorter-term predictions and more general, longer-term forecasts (Swanson et al. 1983). In some scenarios, highly confident and precise statements about future volcanic activity may well be justifiable. However, such statements can readily be included as one end-member of a broader, probabilistic framework of eruption “forecasting” that incorporates more common, but less confident, statements (Sparks 2003). Such a probabilistic framework implies that there is an effort to more fully account for uncertainties in data, models, and system behavior, however small or large this may be. It can also accommodate and aggregate information from empirical observations, expert opinion, and physics-based models.

Eruption forecasts can be categorized on the basis of data sources and timescales (Marzocchi and Bebbington 2012). Short-term eruption forecasting is predominantly based on the observation of precursors. A precursor is an observable phenomenon that correlates with an increase in the probability of occurrence of a certain type of eruptive activity. Short-term forecasting is usually associated with timescales of hours to weeks or months before the event and is the type of information associated with evacuations. Forecasts can include aspects of timing but also eruption style, size, and location. Seismic activity is an important eruption precursor and plays a key role in short-term forecasting. Long-term forecasting uses historical and geological data to establish probability distributions for eruption frequencies, magnitudes, and styles. Although seismology may play a role, for example, in characterizing the size and location of the volcanic plumbing system, it is relatively minor compared to other considerations for long-term forecasting purposes.

Short-term forecasts use properties of monitoring data and other physical observations to determine the time-dependent probability of different types of activity. These probabilities may be based on physical models or on empirical statistics derived from observations before previous eruptions. The form in which a forecast is presented has implications for its utility and verifiability, and, as yet, there is no established standard or protocol for how a forecast is formulated. There are important differences between, for example, statements such as “the eruption will occur in x days $\pm y$ days,” which assume that current unrest will certainly lead to an eruption, and those such as “there is an $x\%$ probability of an eruption within y days,” which do not. Probabilities of different phenomena based on certain models are likely to be conditional on other imprecisely known factors, such as the probability that the volcano is currently in “unrest” or that unrest will lead to an eruption. The resulting forecasts will be associated with confidence bounds reflecting the uncertainty in the process, statistics, and underlying model. Consequently, some forecasts will be associated with narrower confidence bounds than others.

Seismic Eruption Precursors

Numerous seismic phenomena have been observed to precede eruptions and changes in eruptive activity (Chouet and Matoza 2013). These include changes in the characteristics of earthquakes and in the seismic properties of the volcanic edifice and surrounding crust (e.g., seismic velocity or anisotropy). Earthquakes can indicate deformation and failure of the volcanic edifice (produced by increased magma pressure or flank instability) and the movement of magma and other fluids in the crust. In addition, earthquakes can be generated by deformation of magma itself, providing warnings, for example, of changes in the growth and failure of lava domes. Consequently, factors including changes in the rate, location, and magnitude of earthquakes, changes in the type of earthquakes, and changes in the focal mechanisms of earthquakes might contain information about changes in probability of eruptive activity,

though it can be hard to assign a unique underlying cause.

Earthquake Types

A variety of different earthquakes are observed at volcanoes, characterized by frequency content and duration. The appearance of different types of earthquakes can be sufficient to diagnose the activation of particular physical processes.

Seismic waves from volcano-tectonic (VT) earthquakes have a sharp onset and high-frequency content and indicate brittle failure of the volcanic edifice, underlying crust, or even the magma itself (McNutt 2005). Low-frequency (LF) or long-period earthquakes have frequencies that are generally lower than for VT earthquakes, with an absence of high frequencies, and originate from the movement of magmatic or hydrothermal fluids. Volcanic tremor has frequencies similar to LF events but sustains its amplitude for longer periods of time (from minutes to days). Hybrid events have a high-frequency onset, identical to VT events, but an increased low-frequency coda more like LF events. Hybrid events have been interpreted as fluid movement following an initial brittle failure event or by the scattering of high-frequency signals during wave propagation within the edifice.

Elevated Earthquake Rates

The most common indicator of volcanic unrest is an increase in the rate of earthquakes. Volcano-tectonic earthquakes are generated by stick-slip shear motion on faults within the volcanic edifice. Each event represents an increment of brittle failure and the accumulation of damage. Increased rates of these events indicate increased rates of deformation and can result from combinations of factors including (1) an increase in the stresses driving deformation, such as increasing magma pressure or flank deformation; (2) progressive material weakening (damage) of the volcanic edifice such that deformation rates increase at a constant driving stress; (3) an increased proportion of the brittle component of deformation (i.e., the seismic to aseismic ratio); or (4) thermal activation of the hydrothermal system. Although some eruptions are preceded by years of elevated rates of seismicity, other eruptions can occur with

little or no change in seismicity. Extended periods of seismic unrest are expected in situations where there is prolonged magma accumulation, the progressive formation of a fracture pathway, or slow evolution of edificial stresses. In contrast, little seismicity is expected where a magma pathway has been preserved from a previous eruption, where magma rapidly rises from depth to the surface, or where temperature and stress conditions promote aseismic deformation. The total strain inferred from earthquakes typically represents only a very small proportion of the total deformation observed at the surface and is biased toward large, more dynamic events occurring at shallow depth. Accordingly inferences made from the earthquake seismograms or population dynamics may preserve this bias.

The simplest model for the temporal occurrence of earthquakes is a Poisson process. In this model, the average rate of earthquakes is λ , and occurrence times are random and independent. If λ is constant with time, the model is called a “homogenous Poisson process.” In the more general case that $\lambda(t)$ varies with time, the model is called a “non-homogenous Poisson process”; increases in $\lambda(t)$ for volcanic earthquakes are potential precursors to eruptive activity. For a Poisson process, the times between successive earthquakes are described by an exponential distribution, and earthquake rates (the numbers of earthquake occurring within a given time window) are described by a Poisson distribution. Aside from perhaps the very largest events, tectonic earthquake occurrence has been shown to be non-Poissonian (i.e., the variations in earthquake rate are greater than would be expected by from a Poisson process), and this is also the case for volcanic earthquakes. More realistic point process earthquake occurrence models, such as the ETAS (epidemic-type aftershock sequence) model, incorporate earthquake interaction and triggering processes on top of an underlying Poisson process. However, even outside periods of eruptive activity, volcanic seismicity is characterized by swarms and other nonstationary phenomena, resulting in variations in earthquake rate that are greater than those that can be expected from a simple earthquake triggering

model. These variations mean that it can be difficult to identify anomalous increases in the underlying earthquake rate $\lambda(t)$ especially when the background behavior is poorly characterized.

Preruptive changes in earthquake rate can take different forms. At a volcano that has been quiescent for many years, simply the occurrence of detectable earthquakes may be indicative of the resupply of magma and the approach to eruption. At volcanoes where there is ongoing seismicity, a change in the system (e.g., a switch from deformation being primarily driven by tectonic process to deformation being driven by magma pressure) may be indicated by step changes in earthquake rate. Systematic increases in earthquake rate with time are frequently reported before eruptions. Such sequences can evolve over a wide range of timescales, from hours to years, and it has been argued that they may potentially allow more direct forecasts of the timing of an eruption.

Increases in the rate of earthquakes originating within a magma body are likely to be indicative of processes including magma pressurization, increased rates of magma movement, increased flow of non-magmatic fluids or gases, or changes in the rheology of the magma. Consequently, such changes are likely to correlate with changes in the probability of an eruption or a change in eruptive behavior.

Completeness Magnitude

Identifying and quantifying changes in earthquake rates relies on accurate determination of the earthquake catalogue completeness magnitude. The completeness magnitude is a lower magnitude threshold above which all earthquakes are recorded. As the frequency-magnitude relation is exponential for small magnitudes, small changes in the completeness magnitude can lead to apparently significant changes in rate if they are unaccounted for. Changes in completeness magnitude can arise from improvements in monitoring networks, seasonal noise changes, or hypocenter migration. Completeness magnitude is also correlated with earthquake rate; at times of high earthquake rate, a greater proportion of

small magnitude earthquakes will be undetected. Correct determination of the completeness magnitude is also important before any analysis of the frequency (F) versus magnitude (m) distribution. As in the case of tectonic earthquakes, this takes the form of a Gutenberg–Richter distribution, $\log(F) = a - bm$, where a and b are model parameters related, respectively, to total event rate and the relative proportion of large and small events.

There are a range of different quantitative methods to determine the completeness magnitude for an earthquake catalogue. The three most frequently used methods are the maximum-curvature method, the goodness-of-fit test, and the b -value stability method (Mignan and Woessner 2012). These methods are all based on the assumption that the earthquake magnitudes follow a Gutenberg–Richter distribution. In scenarios where the rate (and completeness magnitude) increases with time, it is likely that these methods systematically underestimate the completeness magnitude. This results in an underestimate of the total number of events above the apparent threshold, because several small events are not detected.

Earthquake Magnitudes

Earthquake magnitude is a measure of the size of the earthquake. Magnitude is proportional to the logarithm of the amount of energy released and depends on the amount of slip and the area of the source. Larger magnitude earthquakes reflect larger amounts of energy release, greater deformation and slip, and a greater increment of brittle damage. Events of magnitude 5–6 have been recorded in volcanic areas. However, volcanic earthquake catalogues are commonly dominated by events of magnitudes 1–3. An earthquake of magnitude 2 is associated with movements of about 10 mm over a distance of 100 m (McNutt 1999). If the b -value of the frequency–magnitude distribution remains constant, increases in the maximum magnitude in a given time window reflect an increase in earthquake rate and so are a proxy measure for rate, albeit with a much greater uncertainty in $m_{\max}$ than N . A decrease in the b -value of the frequency–magnitude

distribution can also result in proportionally more large events.

The frequency–magnitude distribution of earthquakes, and in particular the seismic b -value, describes the relative proportions of large and small events. Changes in b -value have been linked to changes in stress and to changes in the populations of faults and fractures in the volcanic edifice. These observations are supported by those from laboratory experiments and the predictions of theoretical models.

Large tectonic earthquakes have also been reported as precursors to eruptions. Global studies suggest a correlation between large events and the onset of eruptions or changes in eruptive style (Schmid and Grasso 2012). Earthquakes occurring within a few fault lengths of the volcano may influence the magmatic system by a static stress transfer mechanism, directly dilating or compressing the plumbing system. These effects are also observed for “silent” earthquakes in the flanks of volcanoes such as Kilauea, Hawaii, where slip occurs too slowly to generate seismic waves but does result in a considerable static stress change. However, triggering effects have been identified over much larger distances of hundreds of kilometers. At these distances, static stress changes are small, and so any interaction is likely to be due to dynamic effects as the earthquake waves pass through the volcano. Potential mechanisms for this effect include the perturbation of the magma chamber and diffusion of pore fluids in the surrounding crust.

Hypocenter Migration

Earthquakes occur where the local deviatoric stress overcomes rock strength and friction. The stress field in a volcano is a function of tectonic and edificial stresses and magmatic and fluid-driven stresses. Consequently, earthquakes are expected to occur either where magmatic and fluid stresses are the greatest or where background stresses have already brought the edifice close to failure. In some instances, earthquake locations may be expected to indicate magma migration. There are many good examples where the location of earthquake hypocenters closely correlates with independent observations

of the location of magma, such as surface deformation signals and eruptive vents. This evidence is particularly strong at volcanoes where there is lateral migration of magma, including Krafla in Iceland and Kilauea in Hawaii. At Kilauea, the lateral migration of epicenters has been observed into one or both of the rift zones that emanate from the volcano's summit region. The most accurate locations indicate that events occur along a narrow range of depths, suggesting the unzipping of crust along a preferred stress horizon above a deeper rift-zone reservoir. In general, however, it remains unclear as to what extent earthquake hypocenters can be used to map magma migration.

Upward migration of earthquake sources might be expected to be a commonly observed precursor, mapping the movement of magma from depth to the surface. However, it is rarely reported. The absence of clear upward migrating signals may partly be explained by large uncertainties in estimates of the hypocentral location estimation, but, even in the case of well-located data, such patterns are rare. Alternatively, the breaking of the crust that occurs in the vicinity of a propagating magmatic fracture may involve fault movements that are too small to be detected at the surface; in this case, the detected volcano-tectonic seismicity may reflect how stress changes across the deforming crust and volcanic edifice induce movements of existing, larger faults. Distributed deformation is also observed on a smaller scale in laboratory tests on natural rock samples in compression, where localization on or near the final macroscopic crack that controls bulk failure or fault often occurs only at the very last minute. In the case of a volcanic eruption, the final upward migration of magma through this cloud of seismicity is often rapid and accompanied by high levels of tremor and other seismic noise. These attributes place strong constraints on the reliability, accuracy, and timeliness of any eruption forecast based on inferences of physical processes from the seismicity.

Before many eruptions, elevated rates of earthquakes occur in several discrete clusters, kilometers across, located in different parts of the edifice. This behavior has been observed at a

variety of different types of volcano including Mt Pinatubo, Mount St Helens, and Kilauea, with clusters separated by several kilometers horizontally or vertically. These different clusters may indicate distinct magma bodies, magma movement, or non-magmatic deformation processes. At volcanoes with flank instability, such as Kilauea and Mt Etna, elevated rates of earthquakes in the flank are indicative of increased rates of flank movement. These swarms can occur concurrently with seismicity associated with the magma plumbing system and have been observed to directly precede changes in eruptive activity.

Earthquake Focal Mechanisms

Earthquake focal mechanisms describe the nature of deformation at the earthquake source. Most VT earthquakes are double-couple, indicative of shear motion on a fault plane. The orientation of focal mechanism describes the stress field at the source. Rotations of the focal mechanism with time from, for example, strike-slip to reverse, have been reported before some eruptions and interpreted in terms of magma pressurization changing the orientations of the principal stresses in the vicinity of the magma body (Roman and Cashman 2006). Non-double-couple earthquakes indicate a more complex source, including tensile failure and events associated with finite volume changes, such as explosions or collapses. In Iceland, for example, such sources have been associated with tensile failure during the opening of subsurface cracks.

Factors Influencing the Characteristics of Precursory Seismicity

For magma to reach the surface along a new pathway, the magma pressure must be sufficient (1) to raise the magma to the surface against gravity, (2) to drive magma flow against the magma viscosity, and (3) to open the conduit against the crustal stress and rock strength sufficiently wide for magma flow. Patterns of precursory seismicity before any given eruption will reflect the details of the physical processes that result in that specific volcanic system satisfying these conditions. The volcanic system may

already be close to the eruptive conditions at the start of unrest, either because the system has not relaxed since the preceding eruption or because non-magmatic processes (e.g., flank movement or regional earthquakes) bring the system to this state. In this scenario, only a small amount of preeruptive magma-driven deformation may be required, with short-lived VT seismicity. After long repose times or large eruptions, or in tectonic settings that inhibit upward magma movement, greater amounts of magma-driven deformation are likely to be required before the necessary eruptive stress conditions are reached. Consequently, prolonged VT seismicity may be expected. These may be combined with changes in the statistical properties of seismicity such as changes with time in the clustering of events and their frequency–magnitude distribution.

Presence of a Preexisting Magma Pathway

Magma requires a physical through-going pathway in order to migrate from depth to the surface. The formation and opening of a pathway are generally associated with VT seismicity and so the state of any preexisting pathway influences the nature of precursory seismicity. At volcanoes where a pathway already exists, the onset of eruption, or a change in style of an ongoing eruption, may be associated with little additional deformation of the edifice. In this scenario, the deformation and seismicity that occur result from pressure changes within the magma body, leading to earthquakes in the surrounding edifice and, potentially, within the magma body itself. Depending on the rheology of the magma, these processes may be associated with seismicity resulting from brittle failure of the magma or tremor-like seismicity associated with fluid movement. Where a magma pathway does not exist before the start of unrest, the formation of a new magma pathway is expected to be associated with increased rates of VT seismicity, owing to faulting in the crust around a pressurized magma body. However, the general lack of a close correlation between the location of dikes and earthquake hypocenters suggests that the earthquakes themselves are unlikely be directly related to the formation of a magmatic pathway.

Volcanic Stress Conditions

Magma requires a sufficiently wide pathway through which to flow. For a pathway to open, magma pressure must exceed the minimum principle compressive stress. This can be achieved by an increase in magma pressure or by a reduction in the stress. Increases in magma pressure can be caused by the supply of additional magma from depth or by the vesiculation of magma, as a result of processes such as crystallization or stimulation by dynamic stress changes produced by the passage of seismic waves. Reduction in edificial stress conditions can occur by processes including flank instability, regional tectonic earthquakes, and time-dependent deformation of the edifice.

Magma Rheology

Magma composition, crystallinity, and temperature are primary factors controlling magma rheology. Magma rheology determines how readily it will flow through a pathway and how flow might generate earthquakes. Basic magma is generally hotter and less viscous than silicic magma, meaning it can flow through narrower pathways and with a lower driving pressure.

Forecasting Eruptions Using Precursory Seismic Phenomena

The previous sections have set out different types of precursory seismic phenomena and the different broad factors that influence their occurrence. However, many of these factors are poorly known at the start of unrest, and the complexity of volcanic systems mean that precursory patterns may vary considerably between eruptions. It is therefore a challenge to quantitatively link precursory behavior to probabilities of eruptive activity. Three different approaches are (1) using empirical statistics based on previous data, (2) using physics-based models, and (3) using the opinion of a group of experts. The boundaries between these approaches are indistinct, and the application of one almost certainly requires elements of the others.

Empirical Data

Probabilistic forecasts of future activity can be constructed using the empirical statistics of datasets associated with past preeruptive behavior. This methodology relies on the robust characterization and classification of precursory phenomena (both in archive datasets and in real-time monitoring data). The probability of different types of activity, within different time windows, can be determined on the basis of past correlation with one or more metrics, such as the rate of local VT earthquakes, the amplitude of tremor, or the occurrence of large magnitude regional earthquakes. The probability of an eruption can then be estimated, for example, from the number of times previous eruptions have been observed to occur when the given metrics have exceeded particular values. The confidence in the forecasts depends heavily on the quality and quantity of past data. At frequently active volcanoes with long-running monitoring programs (such as Kilauea and Mt Etna), the data are sufficient to build volcano-specific statistical models for precursory seismicity. At less frequently active volcanoes, those with less well-established monitoring programs, or those where the style of activity is changing, it is necessary to combine data with those from analogous volcanoes. This process involves a degree of subjectivity – just what are the most likely analogues, and how can the statistics be relatively weighted? Even when datasets are combined, current data archives are small and inconsistent. Regional and global initiatives to integrate monitoring datasets, and to collate historical data, will improve the basis for empirical forecasts.

Physics-Based Models

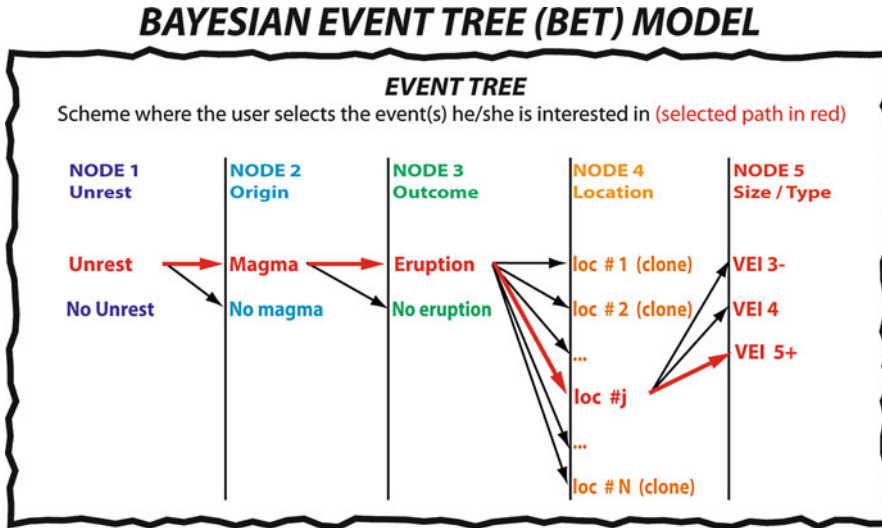
Physics-based models use an understanding of the physical process controlling the approach to eruption to quantitatively link aspects of seismic or other precursory phenomena to the properties of future eruptive activity. The biggest appeal of physics-based models is that they have the potential to be applied to eruptions at volcanoes where there is little or no archive data on which to base empirical models. The most widely applied model of this type is the material failure model,

which links rates of seismicity and deformation to the timing of eruption onset through theoretical and experimental understanding of rock mechanics, fracturing, and faulting. Additional models related to seismicity, such as earthquake focal mechanisms and the evolution and transfer of stress, also have the potential to be formulated to produce probabilistic eruption forecasts.

Expert Elicitation and Event Trees

Many monitoring datasets are too limited, and the understanding of physical processes is too incomplete, to produce completely objective eruption forecasts. In practice, the gap is bridged by expert opinion, where one or many individuals assess the likelihood of different scenarios given the empirical and physical information available. Expert elicitation can also be used to define thresholds for seismic precursor metrics that indicate likely changes in behavior (e.g., the rate of magnitude 2 and greater VT earthquakes exceeding 10 per day). Formal methodologies, such as the Delphi method, can assimilate opinions from groups of experts. These approaches may include weighting for individuals on the basis of past performance or performance on benchmark questions.

Several strategies have been developed to integrate information based on different precursory phenomena, including event trees (Sobradelo and Martí 2010; Marzocchi et al. 2008; Newhall and Hoblitt 2002) and Bayesian belief networks (Aspinall et al. 2003). Event trees provide a quantitative method to combine different forecasting approaches, including empirical statistics, the outputs of physics-based models, and expert opinion to provide probabilities of future volcanic activity. Event trees follow a hierarchical structure, with a series of increasingly specific levels mapping out all possible outcomes of unrest (Fig. 1). The probabilities at each node may be independent or may be conditional on those of preceding nodes. The probability of any given outcome can then be calculated by combining the probabilities associated with all the nodes along the path. These probabilities can be calculated using Bayesian statistics in a Bayesian framework that



Volcanic Eruptions, Real-Time Forecasting of,
Fig. 1 Bayesian event-tree scheme for forecasting volcanic eruptions (Marzocchi et al. 2008). The tree follows a

hierarchical structure, with the probability of each path being computed using the conditional probabilities of each node and, in this instance, Bayesian statistics

can account for both epistemic and aleatoric uncertainties. The event-tree approach allows many different sources of information to be combined and uncertainty to be formally quantified. It also provides a structured framework in which to consider the range of potential future activity, different eruptive scenarios, and the evidence required to distinguish between possible outcomes.

Seismic monitoring data are a primary input to event-tree forecasts. A key challenge is how to develop appropriate metrics to describe seismic characteristics and to establish the thresholds, and relative weightings, for different metrics. The record of monitoring data from volcanoes is currently too small and short to do this empirically, and instead, these values are usually provided by expert opinion.

Material Failure Models

Material failure models are physics-based forecasting methodologies that utilize the rates of preeruptive seismicity and deformation. Increasing rates of local earthquakes have been widely reported before a range of volcanic eruptions and

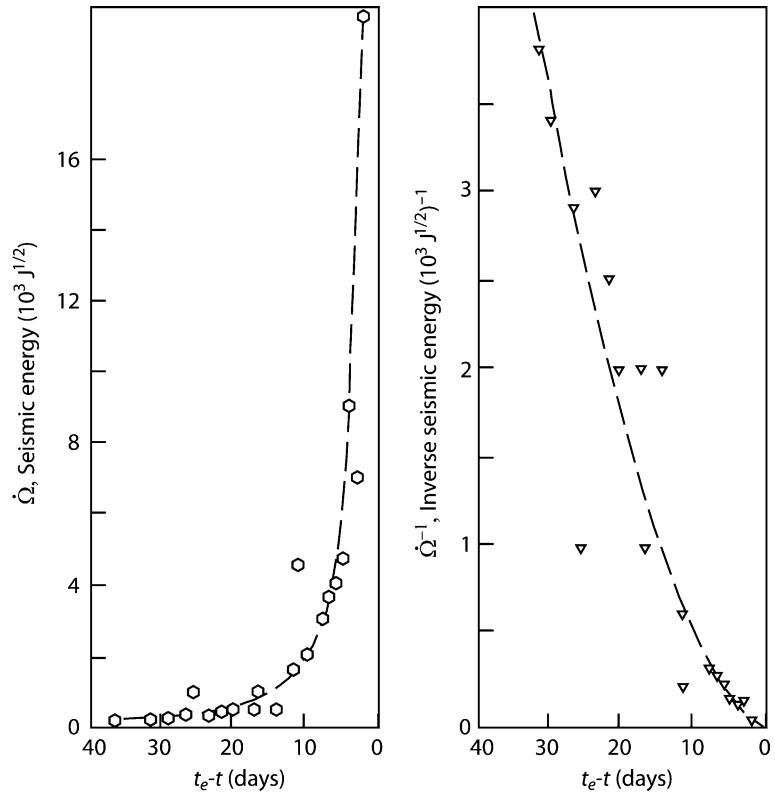
changes in volcanic activity. In some instances, the rate appears to increase systematically with time, leading to the development of models to describe preeruptive changes in earthquake rate and aiming to provide quantitative forecasts of the timing of eruption. In scenarios where baseline monitoring data is unavailable, or where empirical data from analogue volcanoes may not be reliable, physics-based forecasting models have the potential to provide quantitative forecasts of future activity.

The History of Material Failure Models

Several pioneering studies observed increases in rates of earthquakes and their energy release before eruptions and used these parameters to propose simple, empirical methods for forecasting the timing of eruptions. Thus, Omori (1920) and Minakami (1960) used examples from Japanese volcanoes to investigate the forecasting potential of increasing rates of occurrence of local earthquakes, or volcano-tectonic events, whereas Tokarev (1963, 1971) argued that the cumulative Benioff strain (which is proportional to the square root of the seismic energy released) followed a hyperbolic increase with time before selected andesitic eruptions in Russia and Japan.

Volcanic Eruptions, Real-Time Forecasting of,

Fig. 2 Cumulative seismic strain release rate $\dot{\Omega}$ and inverse rate $\dot{\Omega}^{-1}$ before the 12 April 1960 eruption at Bezymianny (After Voight 1988)



Voight (1988) recognized a similarity in the behavior of field precursors to eruptions and the rates of cracking monitored before the bulk failure of samples undergoing creep in the laboratory. He implicitly inferred that the similarity indicated the control of scale-independent deformation processes before larger-scale failure (from failure of a laboratory sample to breaking of the crust before an eruption) and proposed a “fundamental law for failing materials,” for which the acceleration ($\ddot{\Omega}$) and rate of change ($\dot{\Omega}$) of a precursor (Ω) are related by

$$\ddot{\Omega} = A\dot{\Omega}^\alpha$$

which describes a self-accelerating trend, such that as the rate increases, the acceleration becomes larger, increasing the rate still further; A is an empirical constant, and α , which usually lies between 1 and 2, describes the strength of the feedback relation that drives self-acceleration (Fig. 2). The limit $\alpha = 1$ corresponds to an exponential increase in rate with time,

whereas $\alpha > 1$ describes power-law increases that are faster than exponential and include a hyperbolic increase with time when $\alpha = 2$.

The power-law forms of Voight’s relation inherently define a singularity, associated with an infinite rate of occurrence of local earthquakes (and of their release of seismic energy). This singularity corresponds to an inverse-rate approaching zero and is inferred to indicate the formation of a magmatic pathway immediately before eruption. Forecasting the time of eruption therefore equates to extrapolating the precursory trend to the time at which the inverse rate becomes zero (or when the rate tends to an infinite value). In contrast, the exponential form of Voight’s relation $\alpha = 1$ does not contain a singularity and will yield the theoretical infinite rate only after an infinitely long time. In such cases, the end of the precursory sequence must be defined by an alternative threshold, such as a finite maximum rate or a maximum cumulative value for the precursor (e.g., a maximum rate or a maximum total number of local earthquakes).

Cornelius and Voight (1995) describe methods for preparing forecasts from trends with different values of α .

When first proposed, the Voight relation was presumed to apply to a range of precursors related to deformation and seismicity. The term Ω has thus been identified with ground deformation, numbers of seismic events, seismic energy release, and the amplitude of seismic signals, notably RSAM or real-time seismic amplitude measurements (Voight and Cornelius 1991). It was also assumed that accelerations in precursors followed a single trend, associated with a single value of α . However, field observations and data from rock-fracture experiments in which the applied stress increases with time (instead of the creep experiments, under constant stress, that were used to develop the original relation) show that seismic-related precursors can accelerate while the bulk deformation rate remains constant. In such cases, the Voight relation may be applicable to the seismic precursors, but not to bulk deformation. They also show that accelerations can evolve with time from exponential to faster-than-exponential trends. The selection of suitable precursors and their behavior with time may thus change according to time-dependent variations in the conditions of loading during the approach to bulk failure (Kilburn 2012).

Theoretical Underpinning of the Voight Relation: Rock Mechanics, Fracture Growth, Damage, and Lava Dome Growth/Failure

The Voight equation describes the type of self-accelerating behavior that is consistent with observations of precursors to eruptions. It also emphasizes the key role played by the rate of change a precursor, rather than the value of the precursor above a background amount. Given the strong connection of precursors to brittle failure, models to explain the Voight relation have focused on fracture or fault growth and linkage through a volcanic edifice and underlying crust. The models have adopted a broadly damage-mechanics approach and related the seismic precursors, in particular, to the behavior of a population of faults within a volume of

deforming crust that is much larger than the typical fault length (Kilburn 2003; Lengliné et al. 2008; Main 1999). Accelerating rates of earthquakes can also be interpreted in terms of rate-and-state friction acting on a population of faults (Segall 2013), and in addition, they can emerge under specific regimes within the Epidemic-Type Aftershock Sequence (ETAS) model (Amitrano and Helmstetter 2006).

All the models require the action of a time-dependent process to control rates of self-acceleration. For creep deformation under a constant applied stress, continued fracturing is associated with the weakening of rock around fracture tips. A preferred weakening mechanism is stress-enhanced chemical corrosion, in which fluids in the rock, notably water, chemically attack the silicon–oxygen bonds in the zones of high strain at the tips of fractures. For deformation under increasing stress, time dependence is provided by the bulk stress rate itself. For both loading conditions, the rate of redistribution of stress in the vicinity of growing fractures may appear as another controlling time-dependent process.

Accelerating rates of local seismicity have also been reported before changes in eruptive style at volcanoes that have already established a connection between a magma body and the surface. They are especially evident during the effusion of andesitic–dacitic lava domes, including those from Mount St Helens in the USA and Soufriere Hills on Montserrat, and appear to precede major increases in effusion rate or the sector collapse of domes. Current models suggest that the fracturing in these cases may occur as stresses increase within a feeding conduit or dike and cause cooler magma in contact with the wall rock to ascend by fracture, rather than by flow (Hammer and Neuberg 2009).

Model Application and Statistics

An important aspect of the development of forecasts on the basis of physical models is how to optimize models given observed data. Most physics-based forecasting models (including those underpinned by Voight's relation) take the

form of differential equations. Forecasts rely on estimating equation parameters and their uncertainty given some observed data and, perhaps, some prior belief. Consequently, the reliability of forecasts using this approach relies on the robustness of the methods used to estimate parameter values, as well as on the quality and quantity of the data.

Best-fit values for parameters have conventionally been estimated using least-squares regression. However, the most appropriate method for estimating parameters depends on the nature of the data and whether the data is aggregated into finite size bins. Continuous data (such as strain or tilt measurements) require different error distributions to discrete data (such as number of earthquakes per day). As a result, least-squares regression may not be a suitable method for estimating parameters. More robust estimates may instead be obtained using alternative fitting techniques, such as the generalized linear model and the maximum-likelihood method. Thus, the generalized linear model allows different error distributions to be accounted for on aggregated data and can be formulated for different simple forms of Voight's relation. In contrast, for the occurrence of discrete events (e.g., earthquake occurrence), the best model parameters can be estimated using the maximum-likelihood function (Bell et al. 2013).

Real-Time Forecasting and Testing

Neils Bohr is often quoted as saying that "prediction is difficult, especially about the future." In this context, fitting past data, i.e., with the benefit of hindsight, is always going to be much easier than fitting "future" data or indeed any data not used to formulate a hypothesis. Fitting past data is sometimes called "history matching." It is useful for developing insight into the processes at work and developing appropriate hypotheses to test, but cannot be used to assess predictability. After developing models on one dataset, the next step may be to fit a different set of past data and

see if the model still works or to break up a new dataset into two components, use the older data to forecast the younger using the same model, and then test the outcome. This is sometimes called "hind-casting" by the meteorological community. It is a step up from history matching, but it is still plagued by the fact that it is not possible to avoid biases creeping in from knowing the actual outcome, for example, retrospectively selecting a dataset containing events with a disproportional number of eventual eruptions or even in deciding what and when to publish. However hard humans try to be unbiased, it is an established fact that they cannot avoid unconscious biases of data, model, and parameter selection creeping in unconsciously. This is why medical trials are run as "double blind" tests, for which neither doctors nor patients know who is on the drug being tested and who is on the placebo. History matching and hind-casting are necessary, but not sufficient, steps in developing a forecasting model. Such testing may help rule out some models but cannot really be used to prove that a particular one will really work in true forecasting mode. The only way to test this properly is to make and test a forecast in real time, in true prospective mode, in a way that can be verified by the community. Moreover, to test a probabilistic forecast, it is also necessary to make many real-time forecasts and assess the accuracy of the outcomes for a representative statistical sample.

The Importance of Real-Time Testing: Lessons from the Earthquake and Meteorological Communities

Forecasting is used in two ways. In scientific terms, it is the "gold standard" hypothesis test and can be used to test multiple hypotheses in a manner that is as objective and unbiased a manner as possible. In practical terms, forecasting is an operational matter – when do we have enough evidence that activity is sufficient to justify some preparatory action, for example, a heightened alert or the evacuation of a vulnerable district? In the introduction, we distinguished between "forecasting" – a statement of probability – and "prediction," which has connotations with

determinism that is difficult to reconcile with what we know about the complexity, nonlinearity, and our incomplete and uncertain sampling of potential precursors to an eruption. This semantic distinction is old but important. Recently, Ken Mylne of the UK Met Office has written that “*uncertainty* is an inherent part of weather forecasting – the very word *forecast*, coined by Admiral Fitzroy, who founded the Met Office over 150 years ago, differentiates it from a *prediction* by implication of uncertainty” (our italics). This distinction has been taken up by the seismological community in the wake of the debate on earthquake predictability and was influenced by the failure of the Parkfield, California earthquake prediction experiment. Bakun et al. (2005) concluded “The 2004 Parkfield earthquake, with its lack of obvious precursors, demonstrates that reliable short-term earthquake prediction still is not achievable.”

Discussion in the earthquake seismology community has now moved on to the prospective forecasting of space, time, and magnitude-dependent probabilities for earthquake recurrence, informed by the experience in meteorology. To verify forecasting power, the forecasts are lodged in advance with, and made public by, the co-laboratory for the study of earthquake predictability (CSEP) and then tested by the community after a sufficient time has elapsed to distinguish between competing models. Models may be updated, but then they need to be re-tested by the same method. At present, this activity has concentrated on forecasting over a few months or years, where predictability is largely determined by the properties of a mixture of (spatially smoothed) random and triggered seismicity. Several hundreds of years may be needed to gauge the power of longer-term forecasts based on possible nonrandom recurrence, owing to a residual memory of past events and stress renewal. The predictability based on statistical forecasts for a potentially destructive tectonic earthquake is high in relative terms over a time-independent background rate. However, the absolute probabilities remain low at 1 % per day or less (Jordan et al. 2011). Given the much

clearer precursory signals prior to volcanic eruptions, we may expect a significantly larger predictability than this, though still well short of 100 %. At the time of writing, no similar globally coordinated forecasting experiment for testing predictability of volcanic eruptions (CSVP) has yet been constituted.

Testing Eruption Forecasting Models

The development of eruption forecasting models has usually involved retrospective application to example datasets. In many cases, these datasets have been specifically chosen to illustrate the potential viability of the model in question. However, a more systematic approach is recommended to establish the forecasting performance of a model, involving three stages of model testing and evaluation. Firstly, models are benchmarked against synthetic data. With modern computational methods, it is relatively simple to produce large numbers of realizations of synthetic earthquake or surface deformation data with some of properties that might be expected for real preeruptive sequences. For example, synthetic precursory data can be generated using the assumptions underlying the failure forecast model. The FFM can then be applied to the data as if they were real, with forecasts compared against the known input parameters. The performance of the model against these data will show a “best-case” outcome that will not be beaten in real applications. Secondly, models are validated against archive data. Datasets from past eruptions at the volcano of interest, or from analogue volcanoes, can be used for “retrospective testing.” Models are then applied as if the data were being collected in real-time, forecasts issued, and the results compared against the real observations. These tests can begin to establish true forecast uncertainty, and the extent to which the model is truly consistent with the observed data and trends. Finally, models are tested in real time on volcanic monitoring data and with forecasts issued ahead of eruptions. Such “prospective testing” establishes the true forecasting performance of different models, their relative performance, and their

performance compared to simpler null hypotheses: In this way there is no possibility for unfair selection of data or model parameters.

This approach requires formulating models to run automatically and independently, meaning that the forecasting model must be formalized and otherwise hidden sources of uncertainty are accounted for. For example, it must be determined when it is appropriate to apply any given model to data, what the model inputs and outputs are, and what are the values and uncertainties of any model constants. This approach also brings attention to questions such as: what constitutes a good forecasting performance?; and what are the most appropriate forecast metrics for both testability and utility for different potential end-users?

The Future of Eruption Forecasting

There is much that is still unknown about preeruptive processes, and so plenty of scope for improvement in eruption forecasting models. When does the inclusion of physics improve forecasting performance above simply empirical models? How can we best combine multi-parameter datasets, especially seismicity and deformation data? And can we develop reliable physical models for forecasting the style, size, and location of eruption as well as the timing? Key to answering these questions work is the establishment of consistent and well-integrated multi-parameter monitoring datasets from a wide range of volcanoes. These datasets should include information from apparently quiescent systems to establish baseline behavior; a continuous record from a single seismometer and GPS station at an otherwise unmonitored volcano would potentially provide valuable information when unrest was detected. International databases, data standards, and data sharing agreements, supported by new Informatics and computational facilities, should allow the most to be made of available data.

Forecast testing and evaluation will provide some answers to questions regarding forecast

performance and the physical processes controlling volcanic eruptions. However, these developments will also bring further challenges regarding forecast operation and outputs. It will be necessary to develop ensemble forecasting methodologies for integrating the outputs of different models and Bayesian methods for updating forecasts in real time as new data arrives. Ultimately, the aim of eruption forecasting is to help efforts to increase resilience to volcanic hazards. This will require the inclusion of potential end-users at early stages in the model development process to establish the best approach for formulating, communicating, and visualizing forecasts and their uncertainty.

Summary

Eruption forecasts are probabilistic assessments of future eruptive activity, and are important for effective mitigation of volcanic hazards. Changes in patterns of volcanic earthquakes are a key indicator of unrest, and seismic signals are one of the primary datasets used for short-term forecasting of volcanic activity. Forecasting models take information from seismic data and convert it into quantitative estimates of future activity and their uncertainty. These models can be purely empirical, based on statistics of precursory signals before previous events, or physical, based on theoretical concepts of preeruptive processes. A critical step to future improvements of forecasting methods is the development of independent real-time testing of forecasting models (i.e., before the eruption has happened).

Cross-References

- ▶ [Earthquake Swarms](#)
- ▶ [Frequency-Magnitude Distribution of Seismicity in Volcanic Regions](#)
- ▶ [Rockfall Seismicity Accompanying Dome-Building Eruptions](#)
- ▶ [Volcano-Tectonic Seismicity of Soufriere Hills Volcano, Montserrat](#)

References

- Amitrano D, Helmstetter A (2006) Brittle creep, damage, and time to failure in rocks. *J Geophys Res* 111:1–17
- Aspinall WP, Woo G, Voight B, Baxter PJ (2003) Evidence-based volcanology: application to eruption crises. *J Volcanol Geotherm Res* 128(1–3):273–285. doi:10.1016/S0377-0273(03)00260-9
- Bakun WH, Aagaard B, Dost B, Ellsworth WL, Hardebeck JL, Harris RA et al (2005) Implications for prediction and hazard assessment from the 2004 Parkfield earthquake. *Nature* 437(7061):969–974. doi:10.1038/nature04067
- Bell AF, Naylor M, Main IG (2013) The limits of predictability of volcanic eruptions from accelerating rates of earthquakes. *Geophys J Int.* doi:10.1093/gji/ggt191
- Chouet BA, Matoza RS (2013) A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. *J Volcanol Geotherm Res* 252:108–175. doi:10.1016/j.jvolgeores.2012.11.013
- Cornelius R, Voight B (1995) Graphical and PC-software analysis of volcano eruption precursors according to the Materials Failure Forecast Method (FFM). *J Volcanol Geotherm Res* 64(3–4):295–320. doi:10.1016/S0377-0273(94)00078-U
- Hammer C, Neuberg J (2009) On the dynamical behaviour of low-frequency earthquake swarms prior to a dome collapse of Soufrière Hill volcano, Montserrat. *Geophys Res Lett* 36:1–3
- Jordan T, Chen Y, Gasparini P, Madariaga R, Main I, Marzocchi W et al (2011) Operational earthquake forecasting: state of knowledge and guidelines for implementation. *Ann Geophys* 1–77. doi:10.4401/ag-5350
- Kilburn C (2003) Multiscale fracturing as a key to forecasting volcanic eruptions. *J Volcanol Geotherm Res* 125(3–4):271–289. doi:10.1016/S0377-0273(03)00117-3
- Kilburn C (2012) Precursory deformation and fracture before brittle rock failure and potential application to volcanic unrest. *J Geophys Res* 117(B2). doi:10.1029/2011JB008703
- Lengliné O, Marsan D, Got J, Pinel V, Ferrazzini V, Okubo P (2008) Seismicity and deformation induced by magma accumulation at three basaltic volcanoes. *J Geophys Res Solid Earth* (1978–2012) 113(B12305). doi:10.1029/2008JB005937
- Main I (1999) Applicability of time-to-failure analysis to accelerated strain before earthquakes and volcanic eruptions. *Geophys J Int* 139(3):F1–F6. doi:10.1046/j.1365-246x.1999.00004.x
- Marzocchi W, Bebbington M (2012) Probabilistic eruption forecasting at short and long time scales. *Bull Volcanol* 74(8):1777–1805. doi:10.1007/s00445-012-0633-x
- Marzocchi W, Sandri L, Selva J (2008) BET_EF: a probabilistic tool for long-and short-term eruption forecasting. *Bull Volcanol* 70(5):623–632
- McNutt SR (1999) Volcano seismicity. In: Sigurdsson H (ed) *Encyclopedia of volcanoes*. Academic, San Diego, pp 1015–1033
- McNutt SR (2005) Volcanic seismology. *Annu Rev Earth Planet Sci* 33(1):461–491. doi:10.1146/annurev.earth.33.092203.122459
- Mignan A, Woessner J (2012), Estimating the magnitude of completeness for earthquake catalogs. Community Online Resource for Statistical Seismicity Analysis, doi:10.5078/corssa-00180805. Available at <http://www.corssa.org>
- Minakami T (1960) 31. Fundamental research for predicting volcanic eruptions (Part 1): earthquakes and crustal deformations originating from volcanic activities. *Bull Earthq Res Inst* 38:497–544
- Newhall C, Hoblitt R (2002) Constructing event trees for volcanic crises. *Bull Volcanol* 64(1):3–20. doi:10.1007/s004450100173
- Omori F (1920) The Sakura-Jima eruptions and earthquakes. V. *Bull Imperial Earthq Investig Committee* 8(5):353–466
- Roman D, Cashman K (2006) The origin of volcano-tectonic earthquake swarms. *Geology* 34(6):457. doi:10.1130/G22269.1
- Schmid A, Grasso J (2012) Omori law for eruption foreshocks and aftershocks. *J Geophys Res Solid Earth* 117(B7):1–19. doi:10.1029/2011JB008975
- Segall P (2013) Volcano deformation and eruption forecasting. Geological Society, London, Special Publications. doi:10.1144/SP380.4
- Sobradelo R, Martí J (2010) Bayesian event tree for long-term volcanic hazard assessment: application to Teide – Pico Viejo stratovolcanoes, Tenerife, Canary Islands. *J Geophys Res* 115. doi:10.1029/2009JB006566
- Sparks RS (2003) Forecasting volcanic eruptions. *Earth Planet Sci Lett* 210(1–2):1–15. doi:10.1016/S0012-821X(03)00124-9
- Swanson DA, Casadevall T, Dzurisin D, Malone S, Newhall C, Weaver C (1983) Predicting eruptions at Mount St. Helens, June 1980 through December 1982. *Science* 221(4618):1369–1376
- Tokarev P (1963) On a possibility of forecasting of bezymianny volcano eruptions according to seismic data. *Bull Volcanol* 26(1):379–386. doi:10.1007/BF02597299
- Tokarev P (1971) Forecasting volcanic eruptions from seismic data. *Bull Volcanol* 35(1):243–250. doi:10.1007/BF02596821
- Voight B (1988) A method for prediction of volcanic eruptions. *Nature* 332(6160):125–130. doi:10.1038/332125a0
- Voight B, Cornelius R (1991) Prospects for eruption prediction in near real-time. *Nature* 350(6320):695–698. doi:10.1038/350695a0

Volcano-Tectonic Seismicity of Soufriere Hills Volcano, Montserrat

Patrick Smith

Montserrat Volcano Observatory, Flemmings,
Montserrat

Seismic Research Centre, University of the West
Indies, Trinidad and Tobago, West Indies

Synonyms

A-type volcanic earthquakes; HF earthquakes; High-frequency volcanic earthquakes; The eruption of Soufrière Hills Volcano; Volcano-tectonic earthquakes; VT earthquakes; Montserrat

Introduction

Introduction and Definitions

Volcano-tectonic (hereafter VT) earthquakes, sometimes also termed high-frequency (HF) earthquakes, are a category of volcanic seismic signal. They are local earthquakes that occur in a volcanic setting and possess clear impulsive *P*- and *S*-phases with energy between 1 and 20 Hz, an example of which is shown in Fig. 1.

VT earthquakes are interpreted as resulting from brittle failure or shear fracturing along a fault plane and are therefore indistinguishable from ordinary double-couple tectonic earthquakes. Since VT earthquakes possess the same characteristics as tectonic earthquakes, if they are recorded at sufficient stations, standard seismological methods and tools can be used to determine their hypocenters, magnitudes, and focal mechanisms.

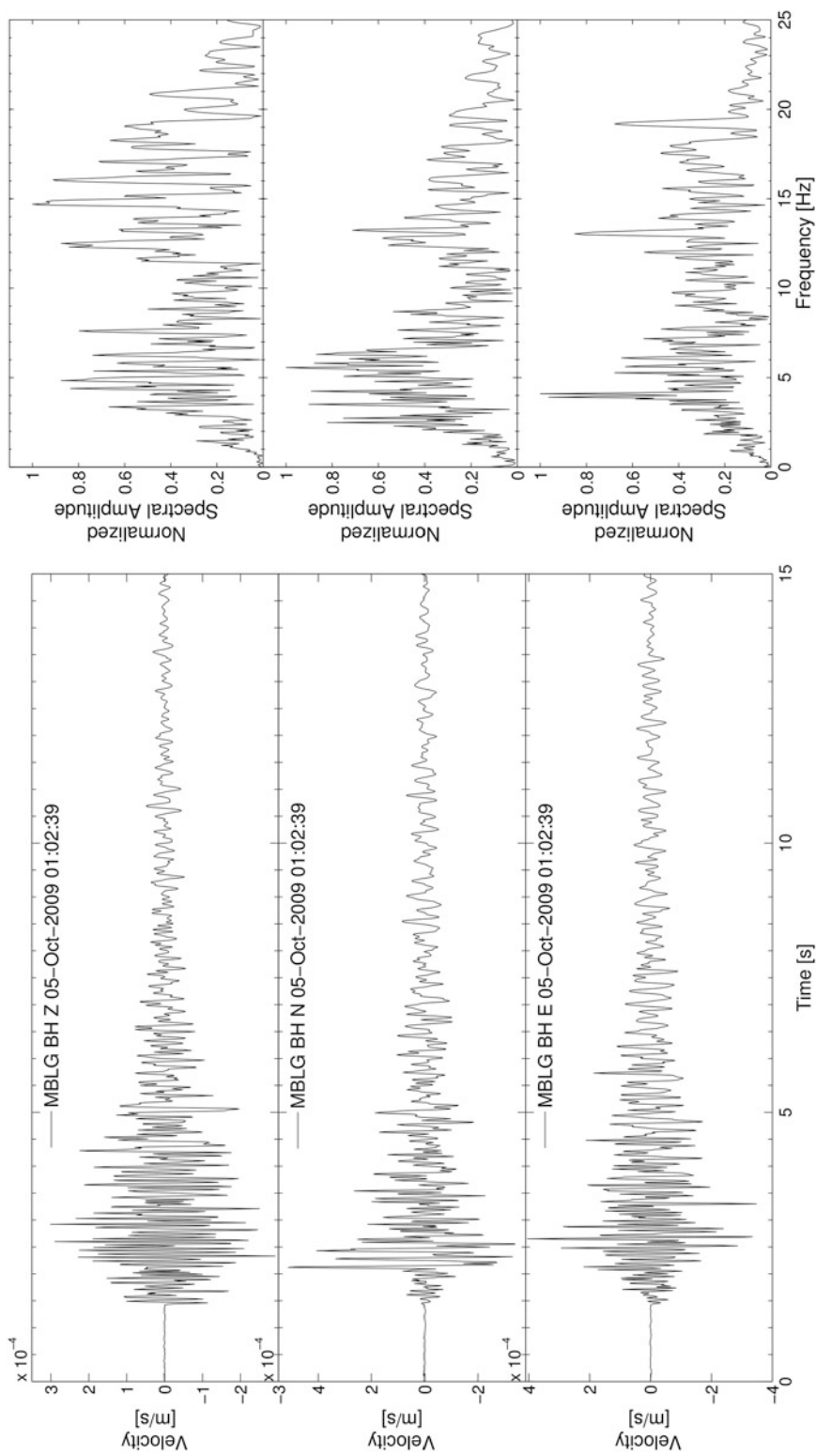
The local stress field in volcanic environments is a combination of the regional tectonic stress and that produced by magmatic and hydrothermal processes. The interaction of these two components leads to the generation of VT earthquakes, which are produced by stress changes, linked to the movement of magma or other fluids.

Therefore, VT earthquakes are often linked to magmatic intrusion and are generally regarded as one of the earliest precursors to increased volcanic activity or unrest (Roman and Cashman 2006). VT earthquakes rarely follow the typical mainshock-aftershock sequence seen for tectonic earthquakes, but instead often exhibit different patterns of swarm behavior, with events clustering in time and space without a single mainshock (McNutt 2005). An increase or acceleration in the rate of VT seismicity is often associated with volcanic eruptions, and theoretical models and concepts, such as material failure, have been used to explain these trends and even to produce short-term eruption forecasts.

Magnitudes of VT earthquakes are typically small, $0 < m_b < 3$, and are rarely higher than $m_b 4.5$ (Zobin 2012), but their generally shallow depth can produce locally high intensities close to the source. The relationship between earthquake magnitude and frequency of occurrence is described by the Gutenberg-Richter frequency-magnitude relation, generally expressed by the *b*-value, where *b* is the slope of linear relationship between magnitude and number when written as $\log_{10} N = a - bM$. In tectonically active areas the *b*-value is generally close to 1.0, but in volcanic areas swarms of VT earthquakes can have much higher *b*-values, with many smaller events and fewer larger ones.

Examples

VT earthquakes can occur before and during volcanic eruptions in a variety of tectonic settings. In basaltic systems such as Piton de La Fournaise volcano, La Reunion, VT earthquakes have been used to track the migration and propagation of magma in an intruding dyke (Battaglia et al. 2005). This behavior is uncommon, however, especially at more silicic volcanoes, and VT seismicity does not usually reveal the path of magma migrating towards the surface prior to eruption. A common precursory feature appears to be the presence of distinct distal clusters of VT earthquakes away from the main volcanic vent, which often disappear with the onset of magmatic activity.



Volcano-Tectonic Seismicity of Soufriere Hills Volcano, Montserrat, Fig. 1 VT earthquake recorded at station MBLG, approximately 3 km NE of Soufrière Hills volcano, Montserrat, on 5 October 2009
earthquake waveform showing *P*- and *S*-phases and broad amplitude spectrum. This example shows 15 s long three-component velocity seismograms from a M_L 2.8 VT

Several examples of this phenomenon have been seen preceding volcanic eruptions in recent years. Each of the last major eruptions at Augustine volcano, Alaska, was preceded by several months of increased and escalating VT seismicity. In the 8 months preceding the explosive eruption in 2006, a cluster of distal VT earthquakes occurred approximately 25 km northeast of Augustine volcano (Fisher et al. 2010). The 1990 eruption of Unzen volcano in southwest Japan was preceded by several years of VT seismicity distributed mainly more than 4 km to the west of the summit (Umakoshi et al. 2001). Two months of VT seismicity also preceded the 1991 eruption of Mount Pinatubo, Philippines. Again, the seismicity occurred in distinct clusters, at shallow depths below the summit and slightly deeper around 5 km to the northwest, with the rate of seismicity accelerating in the days before 7 June eruption. During the early stages of the eruption of Soufrière Hills volcano, Montserrat, in 1995, two main distal clusters of VT earthquakes several km northeast and northwest of the vent were also observed (Aspinall et al. 1998). These are discussed in more detail in the section “[Pre-1995 and Phase 1 VT Seismicity](#).”

Outline

This contribution comprises of a roughly chronological review of VT seismicity at Soufrière Hills volcano (hereafter SHV), Montserrat, chosen because the well-monitored and well-studied ongoing eruption of SHV offers an unparalleled example of the evolution of VT earthquakes during a long-lived multiphase dome-building volcanic eruption.

Section “[Soufrière Hills Volcano](#)” provides some background to the current eruption of SHV, before the VT seismicity is discussed in detail in the following sections. Section “[Pre-1995 and Phase 1 VT Seismicity](#)” discusses the preeruption and early-eruption seismicity, while the relationship between VT seismicity, stress changes, and magmatic intrusion throughout the eruption is examined in the section “[VT Earthquakes and Stress](#).” Some of the more recent VT seismicity at SHV is described in the section

“[Recent VT Seismicity at SHV](#),” primarily the short spasmodic bursts or swarms of VT seismicity that have been termed VT “strings.”

Soufrière Hills Volcano

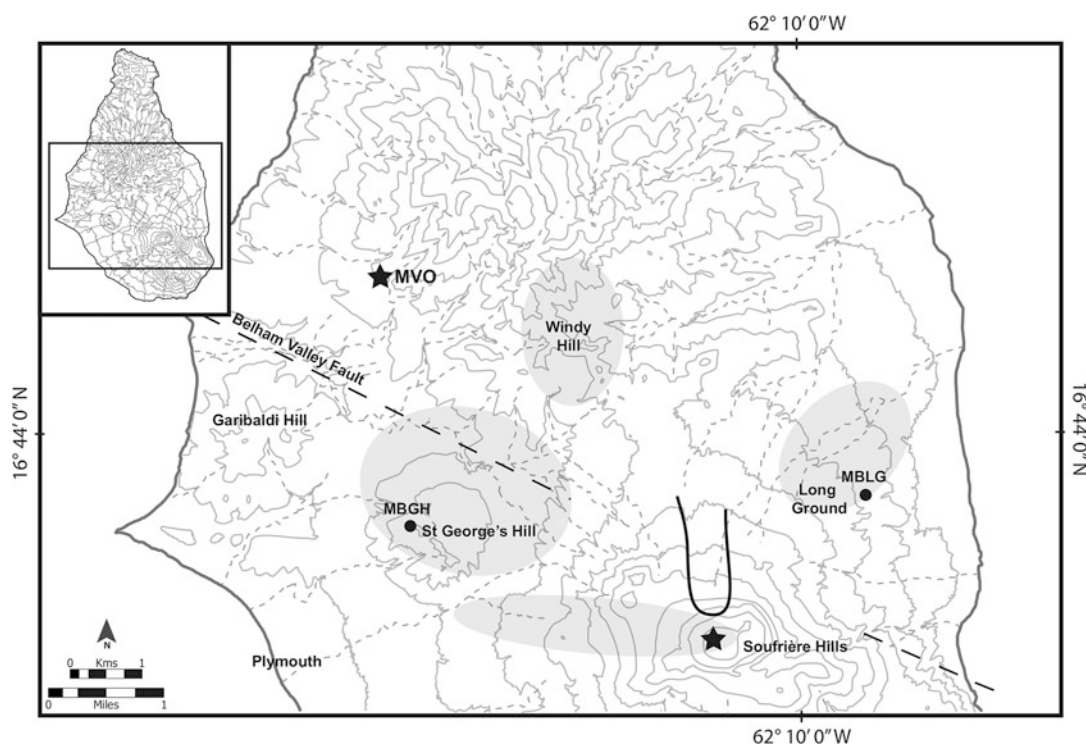
Geological Background and Eruptive History

Soufrière Hills volcano is an andesitic volcano occupying the southern portion of the island of Montserrat. The 16 × 10 km island is located in the northern part of the 800 km long Lesser Antilles volcanic arc, a result of the subduction of the North American plate beneath the Caribbean plate at a convergence rate of around 2 cm/year.

Montserrat is made up of three main volcanic complexes, Silver Hills (c.2.6-1.2 Ma), Centre Hills (950–550 ka), and South Soufrière Hills-Soufrière Hills (c.170 ka – present) (Fig. 2), which get progressively younger from north to south. SHV is the youngest and only currently active volcanic center and comprises of a core of several andesitic domes, surrounded by pyroclastic aprons and other volcanic sedimentary deposits. South Soufrière Hills is the result of a short-lived period of basaltic to basaltic-andesite volcanism that occurred approximately 120 ka.

The current eruption of SHV has consisted of five phases (Fig. 3) of lava extrusion which have erupted a total of more than 1 km³ of andesite in the 18 years since 1995. These extrusive dome-building phases have alternated with distinct pauses, during which no lava was extruded but geophysical monitoring data continued to indicate unrest.

Following a period of 3 years of elevated seismic activity (see the section “[Pre-1995 and Phase 1 VT Seismicity](#)”), the eruption of SHV began on 18 July 1995, with several months of phreatic activity followed by the first dome growth in November 1995 (Young et al. 1998). The first magmatic explosion occurred in September 1996 and activity escalated through 1996 and mid- to late 1997, culminating in the destructive lateral sector collapse event on 26 December (Boxing Day) 1997 with associated blast and debris avalanche. The only known



Volcano-Tectonic Seismicity of Soufriere Hills Volcano, Montserrat, Fig. 2 Map of southern Montserrat showing SHV and other key features. The location of MVO along with two seismic stations, MBGH and MBLG, are shown. The *black star* indicates the explosion crater formed during the 11 February 2010 partial dome collapse, close to the dome summit. The topography shown is pre-collapse, with the approximate location of the 11 February 2010 collapse scar shown in *black* to the

north of the dome. The regions of the four distal clusters of VT seismicity in 1995 are shown by the *shaded ellipses*: north of St. George's Hill (Aspinall et al. 1998; Gardner and White 2002; Miller et al. 2010), beneath Windy Hill (Gardner and White 2002; Miller et al. 2010), to the NE (Gardner and White 2002), and to the WNW (Miller et al. 2010). The approximate location of the Belham Valley fault (Feuillet et al. 2010) is also marked

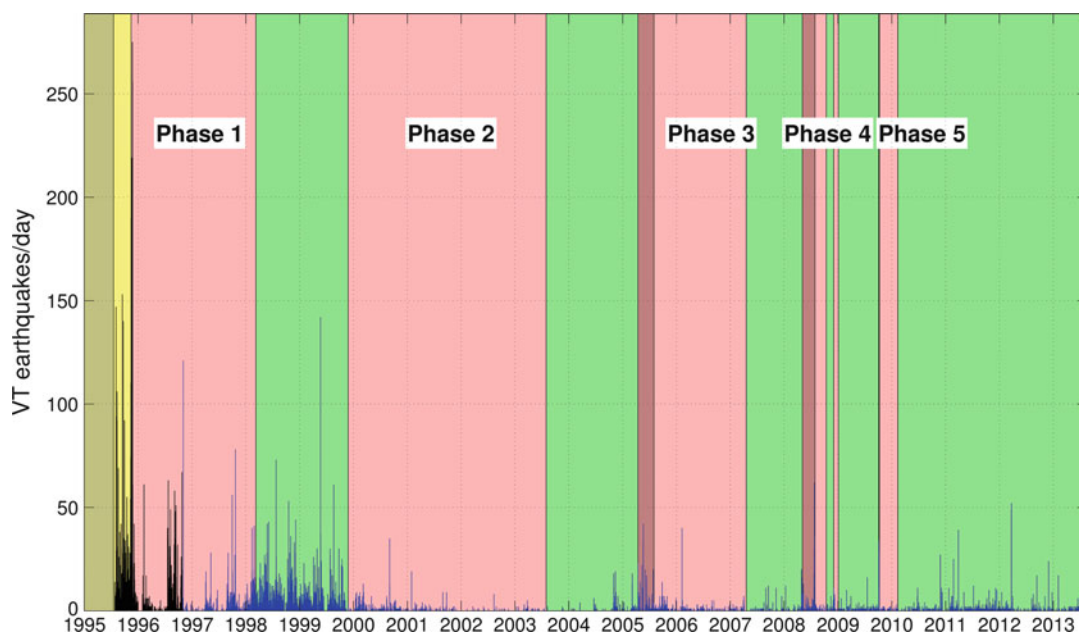
fatalities of the eruption occurred during this first phase of extrusion following a significant dome collapse event on 25 June 1997, whose associated pyroclastic flows and surges led to 19 deaths. Since Phase 1, several other whole or partial dome collapses have occurred throughout the eruption, notably in March 2000, July 2003 (the largest recorded collapse of $200 \times 10^6 \text{ m}^3$), and May 2006. The fifth, and most recent, phase of lava extrusion ended on 11 February 2010, with a partial collapse of the lava dome that involved $50 \times 10^6 \text{ m}^3$ of material.

The eruption has caused major social and economic disruption to the island. The former capital Plymouth and much of the island's major infrastructure were destroyed, with the southern

two-thirds of the island now a permanent exclusion zone. Many residents were forced to leave the island during the early part of the crisis, although the population has since stabilized. A 2011 census estimated the population at just below 5,000 people, slightly less than half of the preeruption population.

Seismic Crises

Prior to the activity of SHV in July 1995, unpublished evidence from radiocarbon dates of possible volcanic charcoal deposits suggests that the last eruption of SHV occurred during or just before the first European settlement of Montserrat in the early seventeenth century. In later historical times, in the century or so before the onset



Volcano-Tectonic Seismicity of Soufriere Hills Volcano, Montserrat, Fig. 3 Daily numbers of VT earthquakes recorded at SHV, Montserrat, from 1995 to present. Earthquake counts from the analogue network are shown in *black* and those from the MVO digital

broadband network in *blue*. The five extrusive phases are shown in *pink*, with pauses in extrusion in *green*. *Yellow* indicates the initial phreatic phase in 1995, and the *dark-pink* periods are subsequent transitional, precursory, or phreatic phases

of the current eruption, there were several volcanic seismic crises on Montserrat, recurring at approximately 30-year intervals. There is very little record of the first in 1897–1898, but more is known of the second, which occurred between 1933 and 1937. Several thousand felt earthquakes were reported, including a large tectonic (i.e., nonvolcanic) earthquake on 10 November 1935, followed by many aftershocks (Perret 1939). As a result of the Royal Society of London expedition to Montserrat to investigate the activity, Powell (1938) established a seismic network of up to eight stations, which operated until 1951. The locations of around 200 local volcanic earthquakes were reported between May 1937 and May 1938, showing several spatial clusters, notably at shallow depths below Soufrière Hills and St. George's Hill to the northwest (Fig. 2). The seismic crisis in 1966–1967 is also well documented. Following felt earthquake reports in early 1966, four seismic stations were installed by the Seismic Research Unit (SRU), Trinidad, and recorded over 700 local earthquakes between

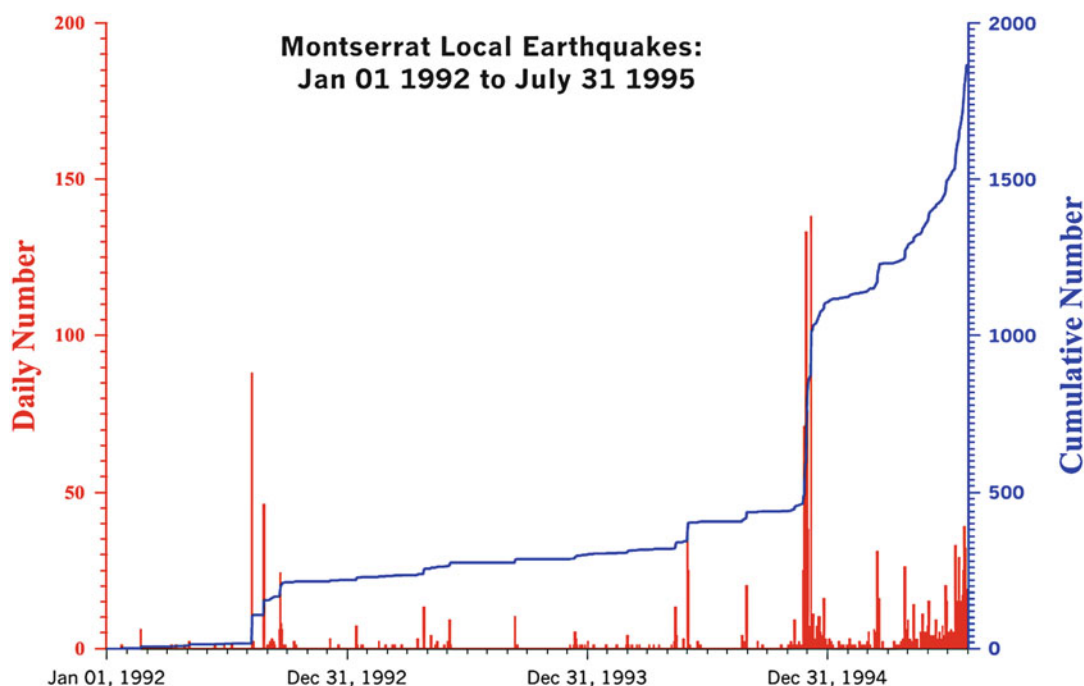
May 1966 and the end of 1967. Hypocenters of VT-type earthquakes were located across the south of the island and beneath Soufrière Hills at depths of less than 15 km (Shepherd et al. 1971).

Increased hydrothermal or fumarolic activity at soufrières surrounding the volcano was observed in association with each of these seismic crises, which have been interpreted as magmatic intrusions that failed to reach the surface.

Pre-1995 and Phase 1 VT Seismicity

VT Seismicity Prior to the 1995 Eruption

After volcanic seismicity on Montserrat waned in late 1967, seismic monitoring on the island was reduced to a single short-period vertical station until 1980, when a real-time telemetered station was installed at St. George's Hill. Between 1967 and 1980 data were recorded locally, and Shepherd et al. (2002) report an average of 3–4 volcanic earthquakes per month, with occasional bursts of



Volcano-Tectonic Seismicity of Soufriere Hills Volcano, Montserrat, Fig. 4 Daily and cumulative numbers of local volcanic earthquakes recorded on Montserrat between 1992 and July 1995 (Taken from Shepherd et al. 2002)

greater numbers. A $M_{6.2}$ tectonic earthquake occurred on 6 March 1985 approximately 20 km N of Montserrat, near the island of Redonda, followed by a large number of aftershocks in 1985–1986. During this time possible local earthquakes were recorded on Montserrat, and in order to distinguish them from tectonic aftershocks, two new seismic stations were established on Montserrat in 1989 (Shepherd et al. 2002).

Following the pattern of the earlier seismic crises of the twentieth century, seismic activity at SHV increased prior to the 1995 eruption. There is perhaps surprisingly little published literature describing this period, but there is consensus that elevated levels of seismicity on Montserrat were observed from January 1992 (Ambeh and Lynch 1996). However, the earlier seismic network was badly damaged by Hurricane Hugo which impacted the island in 1989, and the network was not restored until 1992 (Shepherd et al. 2002). Therefore, it is difficult to assert with certainty that the period of elevated seismicity began in 1992.

Figure 4 shows the volcanic seismicity on Montserrat from 1992 to the beginning of the eruption in mid-1995, with the average daily rate throughout this period higher than at any time since 1938. Little detail of the waveforms is recorded in the literature, but these events are assumed to be mostly high-frequency VT-type earthquakes, located below southern Montserrat at depths of less than 20 km. A total of 18 steadily increasing episodic swarms, lasting from hours to days, were recorded between January 1992 and 17 July 1995 (Ambeh and Lynch 1996). The sharp increase in mid- to late 1994, including many felt events, prompted the installation of four new seismic stations on the island. The rate began to increase further during 1995, and more than 300 events preceded the first phreatic activity in mid-July. The tenfold increase in the daily rate of seismicity before the first magma extrusion in November 1995 has been attributed to slow-rock fracturing as magma approached the surface (Kilburn and Voight 1998).

Distal VT Swarms

In 1995 seismic monitoring consisted of the short-period (analogue) network, comprising of the existing SRU stations supplemented by a network installed by the USGS Volcano Disaster Assistance Program (VDAP) (Aspinall et al. 1998; Gardner and White 2002). These were mostly 1Hz vertical component instruments, with a total of nine stations, although not all were concurrently active. From October 1996 onwards, a network of telemetered digital seismic stations was installed, eventually superseding the short-period network. This initially comprised of five three-component broadband stations and three short-period vertical stations, but has been incrementally upgraded, evolving into the current MVO seismic network of ten broadband and two short-period stations.

VT earthquakes were the dominant type of volcanic seismicity in both the lead up to the eruption and during the initial phreatic phase (Miller et al. 1998), but were replaced by low-frequency seismicity and rockfalls once dome growth began. They were also more abundant during the early part of the eruption (Fig. 3) and continued in high numbers during the first pause in lava extrusion between 1998 and 2000.

As discussed in the section “**Examples**,” distal clusters of VT earthquakes, away from the main volcanic vent, appear to be a relatively common precursory phenomenon, particularly in silicic systems. This was also the case during the eruption of SHV, Montserrat, in 1995. In terms of hypocentral locations, the vast majority of VT earthquakes at SHV have occurred in a main proximal cluster, occupying a relatively small seismogenic volume beneath the volcanic vent, at depths typically 1–3 km bsl. However, during the early phreatic stages of the eruption in 1995, there is consensus that at least two other short-lived spatial clusters of VT earthquakes occurred (Aspinall et al. 1998; Gardner and White 2002; Roman et al. 2008).

The first cluster occurred approximately 3 km to the northeast of the volcano, over a 2-day period between 5 and 6 August 1995, prior to the onset of lava extrusion. A second distal cluster, 4 km to the northwest of the volcano and

centered almost directly beneath St. George’s Hill (SGH), occurred between 12 and 14 August 1995, although occasional further events continued until early 1996. Earlier studies (Aspinall et al. 1998; Gardner and White 2002) found depth ranges of 1–6 km, but more recent work suggests a narrower depth range of 3–4 km bsl (Roman et al. 2008). Gardner and White (2002) also identify a third cluster of distal earthquakes, to the north of SHV below Windy Hill, and in more recent analysis Miller et al. (2010) identify a fourth, extending WNW from SHV (Fig. 2).

This spatial pattern of hypocenters, particularly the cluster beneath SGH to the NW, is qualitatively similar to that observed in the earlier seismic crises at SHV (Powell 1938; Shepherd et al. 1971). Recent work (Roman et al. 2008; Miller et al. 2010) has interpreted these distal clusters, particularly the cluster below SGH, as resulting from stress changes produced by an ascending magmatic dyke, with an NE or NNE trend. The explanation of the distal VT seismicity is that the intruding dyke altered the stress distribution in a weakened tectonic zone and oriented ESE across Montserrat, promoting localized fault movements and changes in pore fluid pressure. Geological evidence suggests that SGH is a result of tectonic uplift, rather than a parasitic cone, and the stress transfer hypothesis is preferred to a secondary magmatic intrusion below SGH, where the distal VT earthquakes would have been proximal to the secondary intrusion.

VT Earthquakes and Stress

VT Source Characteristics

Since VT earthquakes are produced by brittle failure or shear fracturing along a fault plane in the same way as tectonic earthquakes, analysis of their source characteristics – their focal mechanisms and moment tensors – can be used to infer information about the local stress conditions.

Full moment tensors describing the inelastic deformation in the source region can be obtained by waveform inversion, but data from VT earthquakes are often too sparse for such methods, so

they are often assumed to be a result of a pure double-couple source mechanism. The geometry of the faulting that occurs during an earthquake can be studied by determining the earthquake focal mechanism (or fault-plane solution), since it is the fault geometry that controls the seismic radiation pattern.

Focal mechanisms are usually determined from the polarity of *P*-wave first motions at various distances and azimuths but can also be further constrained by the amplitude ratio between *P*- and *S*-wave arrivals. Sets of focal mechanisms can be combined together in order to invert for a common stress tensor that completely describes the local stress regime. Alternatively, the orientation of the principal axes of the focal mechanism can be used as a proxy for the orientation of the stress field. These are the so-called pressure (*P*) and tension (*T*) axes, which represent the maximum and minimum compressive stress, respectively, and are determined by bisecting the dilatational and compressional quadrants.

Studies at several volcanoes have shown that the orientation of the axes of VT fault-plane solutions can undergo systematic changes in response to volcanic activity (e.g., Umakoshi et al. 2001; Roman et al. 2006) and, in particular, in association with changes in the pressurization of magmatic intrusions.

Regional Stress Conditions on Montserrat

Roman et al. (2008) cite evidence from the fault-plane solutions of nearby regional earthquakes and the orientation of dykes mapped on neighboring islands to suggest that the regional stress field around Montserrat is characterized by maximum compressive stress in an arc-normal, approximately NE-SW direction. Recent studies of the local tectonics, arising from offshore seismic reflection profiles, have suggested that, at a local level, the stress regime may in fact be more complex (Kenedi et al. 2010).

Montserrat is located in the northern part of the Lesser Antilles and is a result of oblique subduction of the North American plate beneath the Caribbean plate, which is accommodated by a combination of large-scale sinistral strike-slip faulting and smaller local normal faults

perpendicular to the arc. The dominant structural feature on the island is the extensional Montserrat-Havers fault system (MHFS) (Feuillet et al. 2010), along which the andesitic domes of the SHV complex and other uplifted features such as SGH and Garibaldi Hill are aligned. The MHFS includes an ESE-trending lineament that passes south of the Centre Hills and is interpreted as the Belham Valley fault (BVF). The MHFS also extends to the southeast towards Guadeloupe, where Montserrat lies at the northern end of the offshore Bouillante-Montserrat graben structure (BMF). This could imply a roughly E-W orientation of minimum compressive stress, approximately normal to the N-S regional extensional trend of the MHFS and BMF structures.

VTs and Stress at SHV

The source mechanisms of VT earthquakes at SHV have been analyzed in several studies in order to improve understanding of the local stress field and the nature and orientation of the magmatic intrusions supplying the eruption.

Roman et al. (2006) analyzed the focal mechanisms from a subset of 551 VT earthquakes at SHV with well-constrained double-couple fault-plane solutions, spanning October 1996 to July 2005, and observed temporal changes in the orientation of the *p*-axes. Several periods of a few months were identified where the dominant orientation of the *p*-axes switched from the predominant NE-SW to a NW-SE trend. This represents a 90° rotation from the assumed arc-normal regional maximum compressive stress. Analysis of seismic anisotropy through study of shear-wave splitting of regional earthquakes recorded on the MVO network has provided further evidence to support the presence of short-term variations in the local stress field at SHV.

Models of the interaction between regional and magma-induced stresses (Vargas-Bracamontes and Neuberg 2012) have shown that VT earthquakes on regionally aligned faults are more easily triggered by lower internal pressures of the intruding magma body. By contrast, pressures much higher than the regional stress field are needed to promote rotated VT

earthquakes, supporting observations of their short-lived episodic nature. The presence of such rotated events could therefore indicate significant pressurization of the volcanic system, and hence, the occurrence of these rotated VT events has been interpreted as evidence of stresses induced by the inflation of a NE-SW-oriented dyke beneath SHV. This is broadly supported by the analysis of distal VT focal mechanisms during the early phase of the eruption (see the section “[Distal VT Swarms](#)”).

This NE to NNE orientation is somewhat in disagreement with the NW- to NNW-oriented dyke proposed for SHV derived from geodetic or borehole strain data. Kenedi et al. (2010) propose an approximately N-S-oriented feeder dyke as a possible way to reconcile these competing models.

Recent VT Seismicity at SHV

Patterns of VT Occurrence at SHV

VT earthquakes were most numerous at SHV in the buildup to the eruption and during the initial phreatic phase before dome growth began. They also continued in relatively high numbers during the first pause in extrusion, occurring at a rate of around 5 per day between March 1998 and November 1999. Since this period the daily rate has been much lower and has remained roughly constant at an average rate of less than one per day from 2000 to present.

However, short-term increases in VT activity have been associated with the restarts in extrusion during the course of the eruption at SHV, often increasing notably during the short transitional or phreatic phases which preceded renewed lava extrusion. Diffuse episodes of VT earthquakes associated with ash venting and minor explosive activity preceded the onset of extrusion of both Phase 3 in mid-2005 and Phase 4 in mid-2008. A more intense swarm of 24 VT earthquakes on 5 October 2009 heralded the onset of several days of ash venting before the start of Phase 5 extrusion.

Since around the end of Phase 3 in 2007, the character of VT seismicity seen at SHV appears

to have changed. Fewer intense swarms have been observed, and instead the dominant feature has been the sporadic occurrence of short bursts of high-frequency VT events termed VT “strings.” A sequence of repeating VT earthquakes with similar waveforms was also observed in late 2008 and early 2009. The timing of this change in behavior roughly coincides with a departure from the longer extrusive periods of Phases 1–3 to the shorter more rapid extrusion of Phases 4 and 5.

Repeating VTs

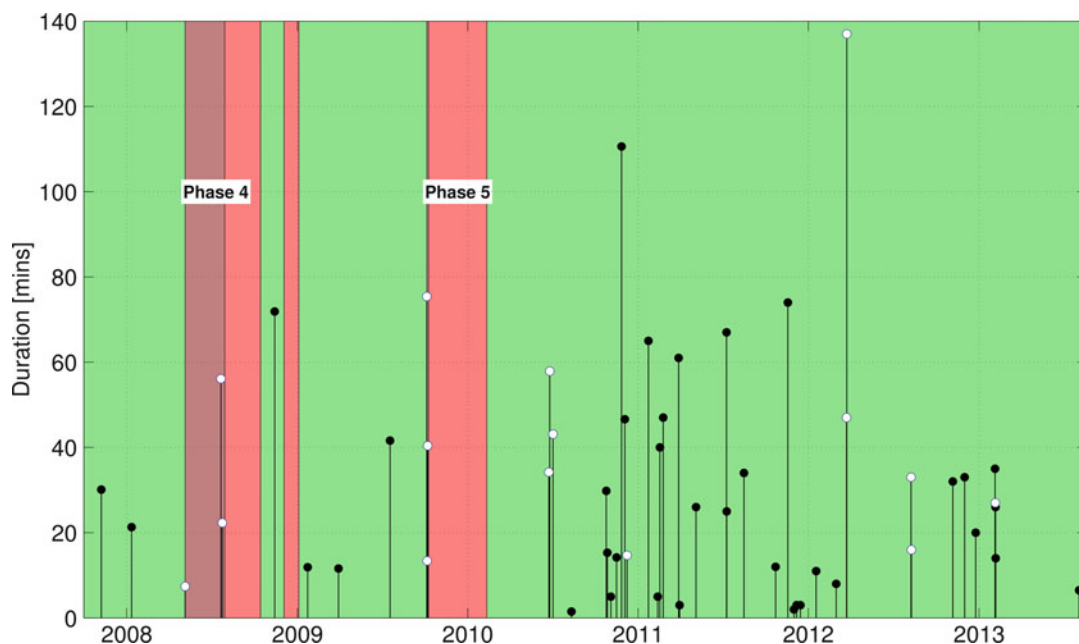
An unusual quasiperiodic sequence of seven large ($2.6 < M_L < 3.2$) repeating VT earthquakes was observed at SHV between October 2008 and May 2009, straddling the two brief periods of extrusion that comprised Phase 4. Analysis of arrival times, waveforms, and focal mechanisms suggests a common location and source process for six of the seven events, with inter-event cross-correlation coefficients of 0.75 and above. A composite focal mechanism revealed mostly strike-slip-style faulting, with a p -axis orientation of approximately 75° , which does not correspond to either the regional stress aligned or the 90° rotated solutions of Roman et al. (2006).

VT Strings

Description and Characteristics

VT strings, defined as a short, intense swarms of VT earthquakes, a phenomenon occasionally referred to elsewhere as “spasmodic bursts” (e.g., Hill et al. 2002), have become a feature of the eruption of SHV since late 2007. To date, 52 such earthquake sequences have been identified in the 69 months since the first in November 2007, continuing across Phases 4 and 5 of lava extrusion (Fig. 5).

The details of individual strings have varied, but certain common characteristics have been identified. The duration of these sequences has typically been less than an hour, with a mean of roughly 10 events occurring in 30 min, when averaged across all strings. The first event in the swarm is generally not the largest, and thus, VT



Volcano-Tectonic Seismicity of Soufriere Hills Volcano, Montserrat, Fig. 5 Duration of all VT strings recorded at SHV, first seen in November 2007. The white circles represent VT strings that were associated

with observable activity at the surface. Extrusive phases are shown in *pink* and pauses in extrusion in *green*, and *dark-pink* periods represent the transitional, precursory, or phreatic phases

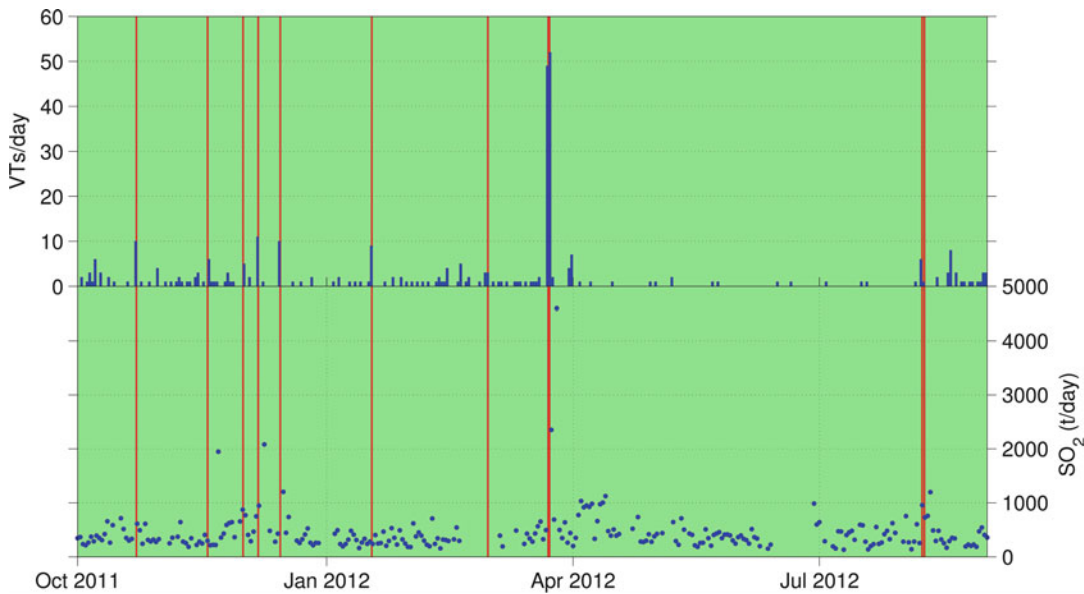
strings do not represent a mainshock-aftershock sequence, instead exhibiting swarm-like behavior, with events tightly clustered in time and space (see, e.g., Fig. 7). In terms of their waveforms, hypocenters, and focal mechanisms, analysis of individual earthquakes has revealed no clear differences between those occurring as part of strings and ordinary or background VT events.

Approximately a quarter of VT strings have been linked to observed surficial activity such as ash venting or visibly increasing degassing and fumarolic activity. This has led to some speculation as to their predictive capability in forecasting such activity or perhaps even renewed lava extrusion. This relationship is supported by SO_2 flux data from SHV, which show some degree of correlation with the seismicity, with often, though not in all cases, a spike in gas output several hours to days after the earthquakes occur (Fig. 6). This has led to tentative proposals that the seismicity is perhaps driven by gas release and/or hydrothermal

processes, although it is clear that this phenomenon is not well understood and that more work is needed to refine the generation mechanism. Similar models have been suggested elsewhere, such as Nishi et al. (1996), who propose a hydrothermal source for spasmodic bursts observed at White Island, New Zealand, based on a mechanism of brittle failure induced by rapid fluid pressure fluctuations in the shallow hydrothermal system.

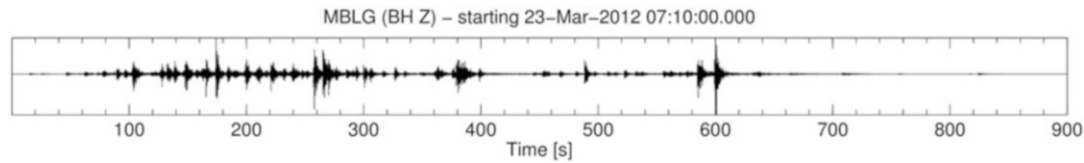
March 2012

Some of the most significant VT seismicity at SHV occurred on 22 and 23 March 2012, with two swarms of around 50 VT earthquakes each. The second swarm on 23 March was more energetic than the first, with local magnitudes of up to $M_L 3.9$: some of the largest VT earthquakes to have been recorded at SHV. The most intense phase of the VT swarm lasted only 15 min (Fig. 7) and was followed by three large hybrid earthquakes. A very-long-period (VLP) seismic signal (below 0.1 Hz or 10 s period) was also



Volcano-Tectonic Seismicity of Soufriere Hills Volcano, Montserrat, Fig. 6 Correlation between VT strings and SO₂ output at SHV for the period October 2011 to September 2012. The top panel shows daily VT

earthquake counts and the bottom the recorded SO₂ flux in tonnes per day. VT strings are indicated by the red vertical lines



Volcano-Tectonic Seismicity of Soufriere Hills Volcano, Montserrat, Fig. 7 Normalized waveform showing the most intense phase of the VT swarm that occurred

at SHV on 23 March 2012. Shown is a 15-min-long vertical component velocity seismogram recorded at station MBLG. The start time is given in UTC

observed across the MVO seismic network during this swarm, coincident with a large amplitude (≈ 250 nanostrain) strain signal recorded by several borehole strainmeters. The seismicity was followed by mild ash venting several hours later. As with the explosion and accompanying strain signal on 3 March 2004 (Linde et al. 2010), preliminary analysis of the strain data has suggested the involvement of a shallow dyke or fracture, although in this case there was no emission of magmatic material. The events of 22–23 March 2012 have some broad similarities with the VT strings that have occurred since at SHV since 2007, perhaps hinting at a common

driving process. These include the relatively short duration of the VT seismicity and its correlation with ash venting and increased degassing. There was a time lag between the occurrence of the seismicity and the SO₂ output which responded on 24 March (2,400 T) followed by the third highest value ever measured by the spectrometer network of 4,600 T on 26 March. The strong pulse in gas associated with this event is similar to the increase in emissions associated with several other VT strings. However, similar strain signals have not been identified with any other VT strings, although the magnitude of this event was clearly much larger.

Summary

Volcano-tectonic earthquakes are brittle failure events that occur in volcanic environments, sharing many of the properties of tectonic earthquakes. Understanding their characteristics and behavior is of critical importance in monitoring and forecasting efforts, particularly as volcano-tectonic earthquakes are often one of the first signs of volcanic unrest and potential future eruptions. The ongoing eruption of Soufrière Hills volcano, Montserrat, has offered an unparalleled opportunity to observe the evolution of volcano-tectonic seismicity throughout a long-term dome-building eruption.

Several volcano-seismic crises, of presumably mostly VT events, occurred on Montserrat at approximately 30-year intervals throughout the twentieth century, in 1933–1937 and 1966–1967. Volcano-tectonic seismicity was also the most prevalent precursor to the start of the current eruption, with increasing numbers seen in escalating episodic swarms in the 3 years preceding the onset of phreatic activity in 1995. During these initial phreatic stages of the eruption, distal swarms of VT earthquakes several kilometers away from the main summit vent were observed, particularly below St. George's Hill to the northwest of the volcano. This distal cluster is proposed to have resulted from stress transfer through a weakened tectonic zone, rather than a secondary magmatic intrusion. A rapid acceleration in the rate of VT earthquakes, linked to subcritical rock fracturing, also preceded the first magmatic activity and the onset of lava extrusion.

The source mechanisms of VT events can be used to infer the local stress conditions under which they are generated. Analysis of the orientation of the p -axes of focal mechanisms of VT earthquakes at SHV has shown systematic temporal changes in response to volcanically induced stresses, in particular the occurrence of events with a 90° rotation from the dominant trend. The presence of these “rotated” events has been interpreted as the result of pressure changes in an NNE- to NE-oriented shallow dyke beneath SHV overprinting the regional

stress regime and promoting failure on optimally aligned faults.

A different character of VT seismicity was seen at SHV after the end of the third phase of lava extrusion. VT earthquakes were the dominant seismicity during the short transitional phases that preceded restarts in lava extrusion which were associated with minor explosive activity and ash-venting episodes. The emergence of short spasmodic bursts of VT earthquakes termed VT “strings” since 2007 has also been linked to ash venting, increased fumarolic activity, and SO_2 output. An intense VT swarm on 23 March 2012 occurred coincident with a very-long-period (VLP) seismic signal and a large step in strain. This seismicity preceded mild ash venting and a large spike in SO_2 release, suggesting the processes driving this event and VT strings may be similar.

Cross-References

- [Earthquake Location](#)
- [Earthquake Magnitude Estimation](#)
- [Earthquake Mechanisms and Stress Field](#)
- [Frequency-Magnitude Distribution of Seismicity in Volcanic Regions](#)
- [Seismic Anisotropy in Volcanic Regions](#)
- [Seismic Monitoring of Volcanoes](#)
- [Volcanic Eruptions, Real-time Forecasting of](#)

References

- Ambenh W, Lynch L (1996) Seismicity preceding the current eruption of the Soufrière Hills volcano, Montserrat, West Indies. In: Ahmad R (ed) *Science, hazards and hazard management – the second Caribbean conference on natural hazards and disasters*, vol 1. Unit for Disaster Studies, The University of the West Indies, Kingston, p 30
- Aspinall W, Miller A, Lynch L, Latchman J, Stewart R, White R, Power J (1998) Soufrière Hills eruption, Montserrat, 1995–1997: volcanic earthquake locations and fault plane solutions. *Geophys Res Lett* 25:3397–3400
- Battaglia J, Ferrazzini V, Staudacher T, Aki K, Chemin J-L (2005) Pre-eruptive migration of earthquakes at the Piton de la Fournaise volcano (Réunion Island). *Geophys J Int* 161(2):549–558

- Feuillet N, Leclerc F, Tapponnier P, Beauducel F, Boudon G, Le Friant A, Deplus C, Lebrun J-F, Nercessian A, Saurel J-M, Clément V (2010) Active faulting induced by slip partitioning in Montserrat and link with volcanic activity: new insights from the 2009 GWADASEIS marine cruise data. *Geophys Res Lett* 37(19), L00E15, doi:10.1029/2010GL042556
- Fisher MA, Ruppert NA, White RA, Sliter RW, Wong FL (2010) Distal volcano-tectonic seismicity near Augustine Volcano. In: Power J, Coombs M, Freymueller J (eds) *The 2006 eruption of Augustine Volcano, Alaska*. U.S. Geological Survey, Reston, Virginia, professional paper 1769, pp 119–126
- Gardner CA, White RA (2002) Seismicity, gas emission and deformation from 18 July to 25 September 1995 during the initial phreatic phase of the eruption of Soufrière Hills Volcano, Montserrat. In: Druitt TH, Kokelaar BP (eds) *The eruption of Soufrière Hills Volcano, Montserrat from 1995 to 1999*, vol 21. Geological Society Memoirs, London, pp 567–581
- Hill DP, Dawson P, Johnston MJS, Pitt AM, Biasi G, Smith K (2002) Very-long-period volcanic earthquakes beneath Mammoth Mountain, California. *Geophys Res Lett* 29(10):8-1–8-4
- Kenedi CL, Sparks RSJ, Malin P, Voight B, Dean S, Minshull T, Paulatto M, Peirce C, Shalev E (2010) Contrasts in morphology and deformation offshore Montserrat: New insights from the SEA-CALIPSO marine cruise data. *Geophys Res Lett* 37(19), L00E25, doi:10.1029/2010GL043925
- Kilburn CRJ, Voight B (1998) Slow rock fracture as eruption precursor at Soufrière Hills volcano, Montserrat. *Geophys Res Lett* 25(19):3665–3668
- Linde AT, Sacks S, Hidayat D, Voight B, Clarke A, Elsworth D, Mattioli G, Malin P, Shalev E, Sparks S, Widiwijayanti C (2010) Vulcanian explosion at Soufrière Hills Volcano, Montserrat on March 2004 as revealed by strain data. *Geophys Res Lett* 37(19), L00E07, doi:10.1029/2009GL041988
- McNutt SR (2005) Volcanic seismology. *Ann Rev Earth Planet Sci* 33(1):461–491
- Miller A, Stewart R, White R, Luckett R, Baptie B, Aspinall W, Latchman J, Lynch L, Voight B (1998) Seismicity associated with dome growth and collapse at the Soufrière Hills Volcano, Montserrat. *Geophys Res Lett* 25:3401–3404
- Miller V, Voight B, Ammon CJ, Shalev E, Thompson G (2010) Seismic expression of magma-induced crustal strains and localized fluid pressures during initial eruptive stages, Soufrière Hills Volcano, Montserrat. *Geophys Res Lett* 37(19)
- Nishi Y, Sherburn S, Scott BJ, Sugihara M (1996) High-frequency earthquakes at White Island volcano, New Zealand: insights into the shallow structure of a volcano-hydrothermal system. *J Volcanol Geotherm Res* 72(34):183–197
- Perret FA (1939) *The volcano-seismic crisis at Montserrat, 1933–1937*. Carnegie Institution of Washington, Washington, DC
- Powell CF (1938) *The Royal Society expedition to Montserrat, BWI. Final report*. Philos Trans R Soc A 237:1–34
- Roman DC, Cashman KV (2006) The origin of volcano-tectonic earthquake swarms. *Geology* 34(6):457–460
- Roman D, Neuberg J, Luckett R (2006) Assessing the likelihood of volcanic eruption through analysis of volcanotectonic earthquake fault-plane solutions. *Earth Planet Sci Lett* 248:244–252
- Roman D, De Angelis S, Latchman J, White R (2008) Patterns of volcanotectonic seismicity and stress during the ongoing eruption of the Soufrière Hills Volcano Montserrat (1995–2007). *J Volcanol Geotherm Res* 173:230–244
- Shepherd JB, Tomblin JF, Woo DA (1971) Volcano-seismic crisis in Montserrat, West Indies, 1966–67. *Bull Volcanol* 35:143–163
- Shepherd JB, Robertson R, Latchman J, Lynch L (2002) Precursory activity to the 1995 eruption of the Soufrière Hills Volcano, Montserrat, Technical report. Seismic Research Unit, University of the West Indies
- Umakoshi K, Shimizu H, Matsuwo N (2001) Volcanotectonic seismicity at Unzen Volcano, Japan, 1985–1999. *J Volcanol Geotherm Res* 112(1–4): 117–131
- Vargas-Bracamontes D, Neuberg J (2012) Interaction between regional and magma-induced stresses and their impact on volcano-tectonic seismicity. *J Volcanol Geotherm Res* 243–244:91–96
- Young SR, Sparks RSJ, Aspinall W, Lynch L, Miller A, Robertson T, Shepherd J (1998) Overview of the eruption of Soufrière Hills Volcano, Montserrat, 18 July 1995 to December 1997. *Geophys Res Lett* 25:3389–3392
- Zobin VM (2012) *Introduction to volcanic seismology*. Elsevier, London

Waste Management Following Earthquake Disaster

Charlotte Brown
Christchurch Polytechnic Institute of
Technology, Christchurch, New Zealand

Synonyms

Construction and demolition waste management;
Debris management; Disaster debris; Disaster
waste management

Introduction

Earthquakes can create large volumes of debris and solid waste. Depending on the severity of the earthquake and the nature of the built environment, waste volumes can be the equivalent of many times the annual waste generated by an affected community. Improved standards for built infrastructure are decreasing the probable impact of earthquakes in many communities. However, increased urbanization and dependence on complex infrastructure networks increase a community's vulnerability to a disaster. This also increases the likely amount of waste produced.

Table 1 shows debris volumes generated from past earthquake (and tsunami) events. Not only can these large waste volumes overwhelm

existing solid waste management facilities and personnel, it can also affect both the response and long-term recovery of an earthquake-affected area. Immediately after an earthquake, debris can block roads, which in turn impedes rescuers and emergency services reaching survivors. In some cases, where hazardous materials are present, waste can pose a public and environmental health risk, for example, where hazardous materials (asbestos, chemicals, radioactive materials, etc.) are present (Brown et al. 2011).

In the longer term, poor management of debris can result in a slow and costly recovery. Rebuilding and repair cannot be carried out before the waste is removed. It is also a public indicator of the speed of recovery. So it is necessary to ensure safe and expedient removal of debris. If managed effectively, debris can become a valuable resource in the recovery and rebuilding process and can have a positive effect on social and economic recovery (Brown et al. 2011).

Historically, disaster waste management has been managed as a logistical exercise. The main focus was removing the material from the affected area with little thought to the longer-term environmental impacts. In L'Aquila, in 1703, a large earthquake severely damaged the town. The debris was deposited in an area behind the Basilica, which now forms the Botanic Gardens. This site has required ongoing remediation to stabilize the poorly compacted fill. In Napier, New Zealand, following the 1931 earthquake, the waste management solution involved dumping

Waste Management Following Earthquake Disaster, Table 1 Reported waste quantities from previous disasters

Year	Event	Estimated waste quantities	Data source
2011	Tohoku earthquake and tsunami	Over 23 million tonnes	(UNEP 2012)
2011	Christchurch earthquakes	8 million tonnes	(Brown 2012)
2010	Haiti earthquake	23–60 million tonnes	(Booth 2010)
2009	L'Aquila earthquake, Italy	1.5–3 million tonnes	(Di.Coma.C accessed 2010)
2008	Sichuan earthquake, China	20 million tonnes	(Taylor 2008)
2004	Indian Ocean earthquake and tsunami	10 million cubic meters (Indonesia alone)	(UNOCHA 2011)
1999	Marmara Earthquake, Turkey	13 million tonnes	(Baycan 2004)
1995	Great Hanshin-Awaji earthquake, Kobe, Japan	15 million tonnes	(Hirayama et al. 2009)

the unsorted debris firstly into a lagoon and secondly onto the beach adjacent to town. With increasing awareness and concern over potential environmental, social, and economic impacts of waste management strategies, as well as the growing size and complexities of disasters, modern disaster waste managers have to take a far more integrated and considered approach.

This chapter will look at the technical and operational aspects of earthquake waste management. The chapter will first consider: waste composition and quantity, waste handling and treatment, waste disposal, hazard management, strategic and operations management, funding, regulations, and planning. The chapter will mainly focus on earthquake-generated waste but will give examples from management of debris resulting from other hazard events.

Waste Composition and Quantity

The composition and quantity of earthquake debris will vary based on the nature of the built environment affected as well as the severity of the earthquake. For example, a community with predominantly timber buildings will have a different waste composition than a community that builds houses from adobe bricks. Or a city with poor building standards affected by a moderate earthquake may generate more debris than a city that has good building standards affected by a much larger earthquake. It is important to understand both the composition and the quantity of

waste generated as this will help authorities determine the most effective strategies for managing the waste.

Waste Composition

In terms of composition, earthquake waste is predominantly construction and demolition waste, that is, waste generated from the demolition of earthquake-affected structures and infrastructure. Waste materials may include metal, concrete, brick, timber, plasterboard, pipes, asphalt, etc. In some cases, where buildings collapse during the earthquake or where buildings are not safe to enter following the earthquake, the waste will include the contents of the building. This could include personal property (e.g., essential documents, money, mobile phones), carpet, furniture, electronic and electrical equipment, plastics, paper, whiteware, putrescible waste, and potentially hazardous materials stored on site (e.g., gas cylinders, oils, pesticides). The exact composition of this waste will depend on the type of building construction and nature of the earthquake impacts.

Earthquakes can also generate other types of waste. Vehicles are often damaged by falling building materials or collapsed buildings. In some earthquake events, liquefaction occurs. Wet silt material is ejected from the ground and accumulates in rivers and on the land surface and can flow onto and into properties. In cases where wastewater pipes have broken during the earthquake, silt can be contaminated. This silt needs to be removed to make buildings habitable and to

ensure stormwater infrastructure can operate properly. Earthquake-generated rockfall is another source of debris that may need to be managed. Solid waste managers may also find themselves dealing with indirect wastes resulting from the earthquake recovery. These wastes could include unwanted donations, health-care wastes, emergency relief food packaging, and rotten food resulting from power outages or damage to food storage facilities.

Earthquakes can also cause secondary hazards that may generate waste. Tsunamis and fires are both possible following an earthquake event. These hazards are unlikely to change the composition of the waste significantly; however, they will affect the nature of the waste. Tsunami debris will be more mixed and will likely be saturated with saline water. Fire debris will also be more mixed and charred. These factors will affect how the debris can be managed. This will be discussed later in the chapter.

On top of earthquake-generated waste, authorities must continue to manage municipal waste. If not collected, municipal waste may pose a public health risk or may become mixed with disaster debris making it more difficult to recycle.

Waste Quantity

Estimating earthquake waste quantities can be challenging. As noted above, both the composition and volume of waste changes significantly depending on the built environment and the severity and effects of the specific earthquake event. Some countries, such as the United States, have developed earthquake debris volume estimation tools for use both before and after an earthquake event. While useful in the United States, these may not be completely transferrable to different countries. One of the main factors restricting the development of estimation tools in other countries is the lack of waste volume data from previous earthquake (and other disaster) events.

There are a number of approaches that can be taken (both pre and post-event) to estimate waste volumes and/or weights. Estimates can be made based on total quantity of waste, or they can be broken down into quantities for different waste

types (timber, concrete, metal, etc.). These methods are briefly explained below.

Unit Estimates

Estimates can be built up based on how much waste is expected from a single type of property and the level of damage it sustained. For example, for residential areas, you could divide the affected properties into the level of damage (say, full demolition, partial repair, and minor repairs). Each level can then be divided further based on the type of construction of the property, its age, and size. For each category you can then estimate, using local information and experience, the quantity and composition of the waste. This strategy is illustrated in Fig. 1. The total waste will be the sum of the number of houses in each category (N) multiplied by the quantity of waste (W , in mass or volume as needed), as shown in Eq. 1. If desired, this technique could be further refined by not only estimating the total quantity of debris but for each subcategory estimating the composition of the waste (20 % concrete, 30 % brick, 10 % timber, etc.). In that case, Eq. 1 would be applied to each type of waste:

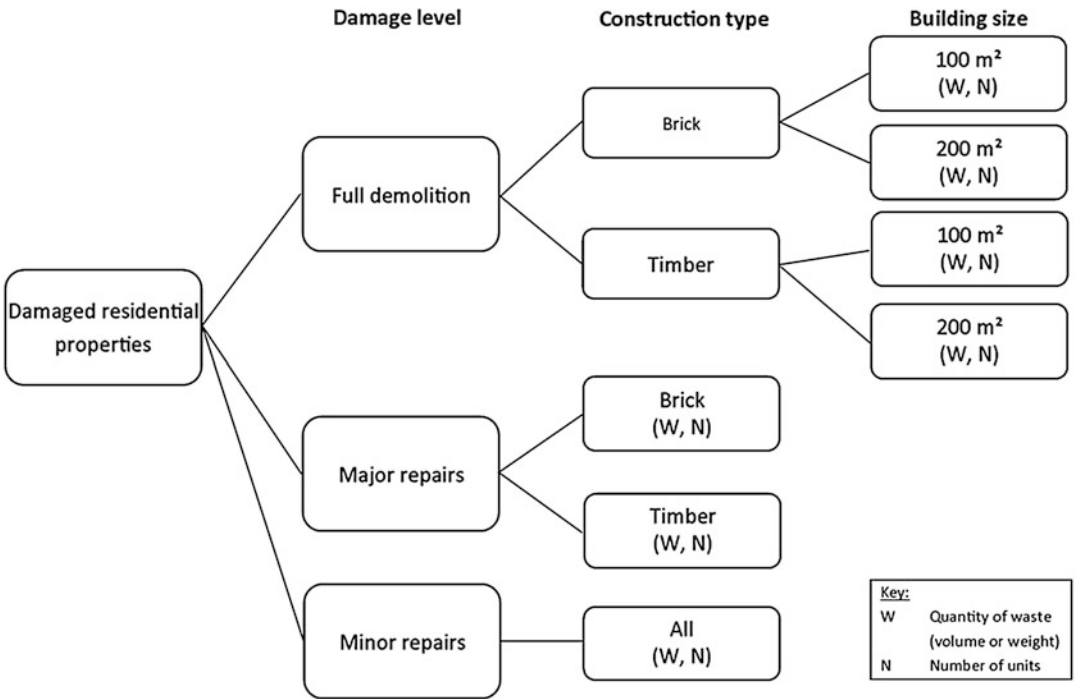
$$\text{Total waste} = \sum_{i=0}^n W_n N_n \quad (1)$$

The same technique can be used for commercial properties; however, it is slightly more involved due to the wide range of building sizes and types. If a unit estimation technique is not possible, then a volume estimate may be possible.

Volume Estimates

For variably shaped buildings, a volume estimation technique may be most appropriate. As a very rough rule of thumb, it can be assumed that the weight of debris in a commercial building is one third of the volume of the erect building, as shown in Eq. 2. Note that this varies widely and would be significantly less for lightweight buildings such as steel-framed warehouses:

$$\text{Total waste (in tonnes)} = \frac{1}{3} \text{ Building volume (m}^3\text{)} \quad (2)$$



Waste Management Following Earthquake Disaster, Fig. 1 Schematic for unit estimation technique of residential earthquake waste

This approximation could only be used for buildings being fully demolished and where the volume of the building is known. Building volumes could be obtained using LIDAR images or using local authority held building data (say building height or number of storeys and floor area). As for the above method, this may be further refined by splitting the buildings into construction type and assigning approximate waste compositions for each building type.

Area Estimates

Areal estimates are based on the amount of waste over a particular area, be it floor area of a building or land area.

For buildings, waste quantities per unit floor area (in m²) could be estimated based on damage level and building type. These areal estimates can then be multiplied by the total floor area of that building damage level and building type. This technique is similar to the unit estimates except that instead of estimates per building (N), the estimates are based on floor area (A); see Eq. 3.

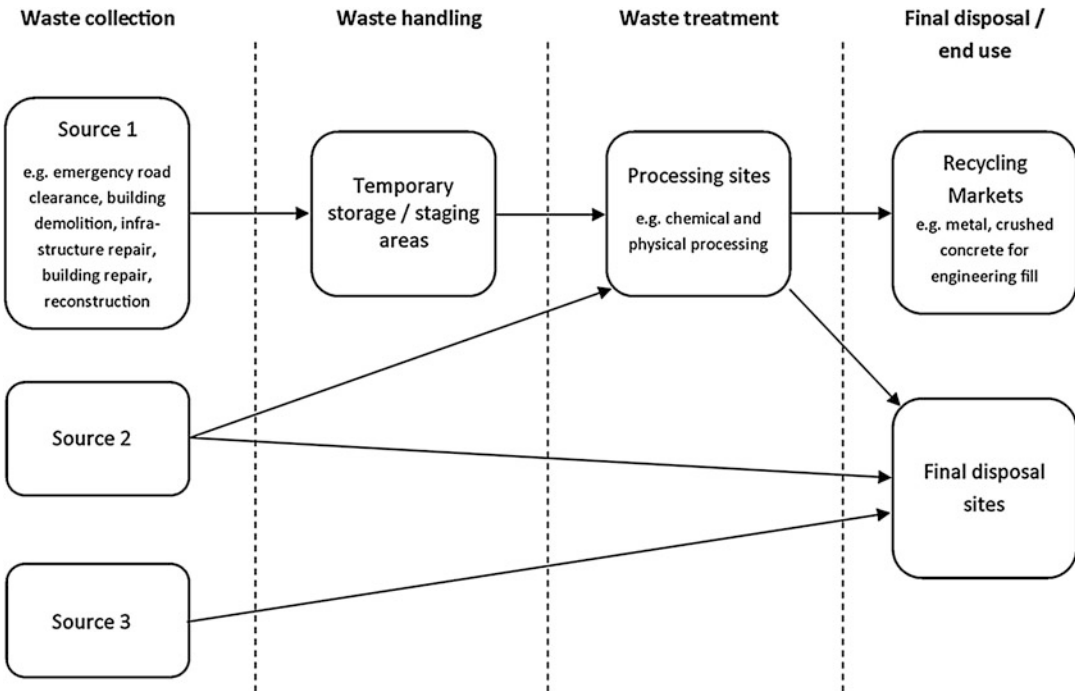
As with the previous methods, this technique allows for further refinement and splitting waste quantities into different types:

$$\text{Total waste} = \sum_{i=0}^n W_n A_n \tag{3}$$

Area estimates are especially useful for earthquake damage that is spread over a large area, such as waste resulting from liquefaction and tsunami damage. Aerial images can be used to determine the extent of the damage. If desired, the areas can be split into damage levels. For each area, unit (per m²) waste quantities (and composition, if desired) can be assigned. This technique was used in Japan following the devastating 2011 Tohoku tsunami.

Waste Management System Overview

Depending on the composition and quantity of waste, an appropriate cradle-to-grave waste management system will have to be designed.



Waste Management Following Earthquake Disaster, Fig. 2 Generic waste management system flow diagram

Generally, waste management systems have several stages. Broadly, these are waste collection, waste handling, waste treatment, and final disposal/end use. Figure 2 shows a generic sample waste flow through these stages.

All these stages, as well as transportation and hazard management, will be discussed in the following sections.

Waste Collection

During the emergency phases of an earthquake response, waste may need to be removed or collected to clear roads. Roads may need to be opened to allow emergency service, critical infrastructure, and rescue personnel access. Some emergency debris works may also be required to remove acute hazards such as unstable buildings and hazardous materials. Generally, this work will be done swiftly by first responders, and waste will often be mixed and taken to existing solid waste management facilities.

During the early recovery phase of an earthquake response, provision may need to be made for collection of minor waste from private properties. This would be particularly important if there was liquefaction- or tsunami-related debris that needed clearing. Authorities generally provide for this in one of the following two ways:

1. Curbside collection service

Residents are asked by the local authority to place material cleared from their property to the curbside (either in a pile or in bins or bags provided) for collection. The waste material can be mixed or separated into recyclables. A contractor or contractors are engaged to collect the waste and take it to a designated waste handling facility. Generally, this service is free for residents.

2. Provision of waste depots

Residents are instructed to take material cleared from their properties to a nominated location. These locations could be existing waste transfer stations or may be specifically

established facilities. The waste material can be mixed or separated. This service could be free or charged.

When determining the waste collection system, it is necessary to consider a number of factors, including: (1) the state of roads (a curbside collection may be more suitable if road movements need to be minimized), (2) how easy and safe it is to handle the waste (are there public or environmental health risks that need to be managed?), and (3) how easy it is for residents to separate different types of waste.

The majority of debris following an earthquake will be generated and collected during the repair and demolition stage of the recovery. Demolition and building contractors will generate waste as they demolish or repair building and horizontal infrastructure (roads and pipes). Generally, the contractors will then transport the waste to handling and treatment and/or disposal facilities (see the next section).

Demolition

Demolition works should be carried out by experienced demolition contractors. Demolition methods should follow best practice adjusted as necessary to account for potential building instability. Demolition techniques will not be discussed here, but special considerations around waste management are briefly discussed.

During the demolition process following an earthquake, there are some types of waste that have to be managed carefully. The first is material linked to fatalities. In many cases, authorities will want to retain building material linked to fatalities for cultural and/or investigative purposes. Separate collection and storage for this material may have to be provided. The next is material from heritage buildings. Where heritage buildings have to be demolished, it may be desired that some of the building materials are retained and/or specially documented. Archeological material may also be found and require preservation. Last, personal property salvaged from buildings needs to be considered. People are often forced to leave buildings quickly following an earthquake and sometimes leave valuable

personal items. Where possible, these should be returned to the owner.

Waste Handling and Treatment

Once waste has been collected, it will either go straight to a disposal site (see next section) or it will go to a handling and/or treatment facility to be processed. Waste handling facilities are generally where mixed waste goes to be separated into recyclable and nonrecyclable parts. Following an earthquake these are generally called temporary storage areas. Waste treatment facilities are where waste is physically or chemically changed in form, to make it suitable for recycling or alternative uses such as engineering fill. Waste may also be incinerated. Waste handling and treatment sites are commonly combined.

Temporary Staging Areas

Generally, because of the large volumes of waste, existing handling and treatment facilities cannot take all the waste immediately. Therefore, temporary staging areas are often established to receive mixed waste, separate them, and in some cases carry out treatment. As a result, temporary staging areas slow the rate and quantity of waste going to existing facilities (which otherwise would be overwhelmed). The downside to temporary storage areas is that waste has to be double handled, which has cost implications.

The size and nature of temporary staging areas will vary depending on the composition and quantity of waste; however, there are some general things to consider. As an approximation, the space required for a temporary staging area is 50 ha per 1,000,000 cubic meter of debris (FEMA 2007). Sites should be located to minimize environmental damage. Operational arrangements must consider items such as site layout, plant requirements, traffic management, operations management, recording/invoicing systems, operational costs, public health, and environmental health risk management procedures. It is vital that temporary storage facilities are aware of what waste is entering the site in order to identify and mitigate potential hazards.



Waste Management Following Earthquake Disaster, Fig. 3 Temporary staging area for managing waste following the 2009 L'Aquila earthquake, L'Aquila, Italy (Photo date: September 2010) (Photo credit: Charlotte Brown)

Good monitoring and record keeping is also essential for billing purposes. Where possible, temporary staging site locations should be identified before an earthquake event. Inappropriate location of temporary storage sites has led to environmental and public health effects in past events, for example, following the 2004 Indian Ocean tsunami (Basnayake et al. 2006; Pilapitiya et al. 2006; UNDP 2006) (Fig. 3).

Recycling

Many components of earthquake waste can be recycled. Commonly recycled materials include concrete, metal, untreated timber, vegetation, and plasterboard/gypsum. Recycling is often a desired option because it saves resources, reduces landfill space needed, can (not always) reduce waste management costs (including transportation), and can create job opportunities.

However, there are some disadvantages associated with recycling following an earthquake. First, depending on the value of recyclables and cost of land disposal, it can be more expensive than landfilling, particularly where there is no market available for the large volume of recyclable materials produced. Second, recycling is more time-consuming than landfilling and is resource intensive: both time and resources are often short during an earthquake recovery. Third, specialist equipment may not be available to process the large volumes of waste, and storage of materials for processing can be challenging. Fourth, hazardous material may have contaminated the waste and this will likely make recycling impossible. Last, if there has been liquefaction, tsunami, or fire damage, it is likely that the materials will be difficult to separate and recycle.

For recycling to be carried out, waste needs to be separated. Waste can either be source-separated (by private home owners or contractors) or it can be separated off-site at waste handling facilities such as the temporary staging areas discussed in the previous section. Generally, waste will be source-separated by demolition contractors. This option is more cost-effective as it reduces double handling and there is often a higher yield of recyclables. It also reduces the demand for waste handling facilities. However, source separation will be more time-consuming, meaning demolition work will take longer. If there is urgency in the demolition works (for safety or recovery purposes), then off-site recycling may be more desirable.

Land Reclamation and Engineering Fill

Demolition waste from earthquakes is often used as engineering fill for land reclamation or land remediation. Inert materials such as concrete, brick, and rock can be broken down into small material suitable for filling operations. This has been used following the 1995 Great Hanshin-Awaji earthquake, the 1999 Marmara earthquake in Turkey, and the 2011 Christchurch earthquakes.

In order for construction waste to make good engineering fill, suitable monitoring of the engineering and environmental quality of the material needs to be in place.

Incineration

Incineration (and waste to energy) is a potential treatment option following an earthquake. This technique has been used following the Indian Ocean tsunami and Great Hanshin-Awaji earthquake. However, there is only a small proportion of earthquake waste that is safely and efficiently combustible, particularly in areas where structural timber is treated. Incineration should only be carried out where appropriate environmental controls are in place for both air emissions and ash disposal.

Waste Disposal

All residual earthquake waste (material that cannot be recycled, used as engineering fill or incinerated) has to be disposed of. Most communities will have existing waste disposal facilities (landfills); however, often these disposal facilities will not be big enough to take all earthquake waste. Following some earthquakes, new disposal facilities have been constructed to increase disposal capacity. Often these landfills are constructed to a lesser degree of environmental protection than ordinary municipal waste landfills. If this approach is taken, authorities need to consider the potential for hazardous material and potential environmental contamination. Public and worker health and safety should also be considered.

As for temporary storage facilities, disposal sites must keep clear records of what waste is entering the site and monitoring systems need to be in place to ensure items that are too hazardous are not deposited.

Hazardous material must be appropriately handled and disposed in accordance with local hazardous material disposal standards.

Hazard Management

There are several public health and environmental hazards present in earthquake waste. The main hazards are rotten food, asbestos, fecal-contaminated material (from broken wastewater pipes), treated timber, industrial waste (chemicals, paints, etc.), and household hazardous wastes (oils, pesticides, etc.). It is important that these wastes are identified and appropriate management approaches are put in place to protect public and environmental health. As a minimum, workers should wear correct personal protective equipment (suitable for the hazard), and hazardous wastes should be removed from properties prior to demolition work to reduce the volume of contaminated waste. It is likely that

risks will be heightened during an earthquake recovery due to the speed and volume of work being carried out, likely by those who have not worked in the waste industry.

Management approaches for hazardous materials should be determined in consideration of existing hazardous waste handling and treatment regulations. In some cases, authorities will choose to change environmental regulations to help to speed up the recovery. This must be done with caution, appropriate mitigation measures in place, and with long-term impacts in mind.

Below is a brief description of the main hazards found in earthquake waste.

Rotten food Particularly in cases where there have been prolonged power outages, rotting food waste must be managed. Food waste from residential and commercial properties (supermarket warehouses, cool-stores, hospitality businesses) is possible. Both solid and liquid waste may need to be managed. Following Hurricane Katrina, extended power outages led to significant amounts of rotting food in domestic fridges. Residents were advised to tape their fridges shut and place the entire fridge on the curbside for collection and disposal. Authorities cleaned the fridges, removed the hazardous refrigerant chemicals, and recycled the metal.

Asbestos Asbestos is a fibrous material used in construction. It is highly regulated in many countries now due to its associated health risks. The risk lies in exposure to airborne fibers being inhaled, so the risk is high when asbestos material is being disturbed

during demolition. Asbestos is a serious risk, and typical mitigation measures include: identifying asbestos prior to demolition; removing asbestos where possible; wearing protective gloves, masks, and clothing (which should be disposed of safely); keeping asbestos surfaces damp; carrying out air monitoring; erecting hazard signs; and sealing the worksites where possible.

Asbestos-contaminated material should be separated on site from clean debris as far as possible to reduce the volume of contaminated material. Contaminated material should be disposed of at an appropriate facility and its disposal location marked to prevent further disturbance.

Fecal-contaminated material (from broken wastewater pipes) When there is damage to the wastewater network (pipes or treatment facility), there will likely be sewage-contaminated materials to manage: both during the repair work and from residents immediately after the event who may not have access to a toilet. Public health protection for workers and general public during collection and disposal must be managed appropriately.

Treated timber Treated timber may be present in the waste. Generally, treated timber cannot be recycled or incinerated due to the likely environmental effects; however, this will depend on local environmental regulations. If necessary, efforts should be made to separate

treated and untreated timber during demolition to reduce the quantities of wood that must be safely landfilled.

Household and industrial hazardous wastes (oils, chemicals, paints, pesticides, mercury, etc.) All properties are likely to contain some hazardous materials. The type and quantity of materials are likely to depend on the building use. The occupants are the ones most likely to know about the hazardous materials present and should be included in locating and managing these wastes, separate from general construction and demolition waste. Hazardous materials should be dealt with as they would during normal (non-disaster) times.

Waste Transportation

Throughout most stages of the waste management process, waste needs to be transported. During an earthquake response, it is likely that a variety of different trucks will be used to transport waste as there will not be enough traditional trucks available. Control and monitoring of trucking activities is important to protect public and environmental health. Measures need to be considered such as dust suppression (either by water or dust covers), wheel washes, truck weight limits, dedicated truck routes (e.g., to avoid dangerous roads), and truck travel hour limits.

Authorities sometimes implement a waste tracking system following earthquakes to help monitor waste movement (to prevent illegal dumping) and, where applicable, to help waste managers manage contract costs and billing. Waste tracking systems generally document where waste comes from and goes to, what and how much waste there is, who transports the waste, and when it is picked up and dropped off. Typical systems used are paper-based systems

(e.g., daywork sheets or dockets), barcode/scanner systems, or GPS tracking systems. Paper-based systems tend to be less expensive but can be subject to fraudulent activity and human error. The barcode and GPS systems are generally more expensive to establish but are less susceptible to tampering and errors.

Strategic and Operations Management

As well as the technical details of managing earthquake waste, it is necessary to consider the strategic and operational management aspects.

Strategic Management

There are many organizations that have a stake in earthquake waste. Typical organizations include:

- Disaster response and recovery authorities
- Waste management industry
- Environmental authorities
- Health and safety authorities
- Public health authorities
- Hazardous substance authorities and industry
- Lifeline (critical infrastructure) authorities
- Marine authorities (for tsunami events: debris in marine environment)
- Transportation authorities
- Heritage building and archeological authorities
- Nondomestic agencies (e.g., international governmental and nongovernmental groups)
- Local cultural groups
- Community representatives

When designing a waste management system, ideally all these entities should be consulted. They should also be involved in any pre-event planning. It is beneficial to appoint one organization as the lead organization to oversee the strategic management and operation as a whole.

Operational Management

Operational earthquake waste management programs are managed in many different ways. A common configuration is that community-wide

waste collection is carried out by publically engaged contractors, private property demolition is carried out by privately engaged contractors (either by a home owner or an insurer), and waste treatment sites are managed by public/or private operators, as per usual municipal waste management. For example, this is how the US Federal Emergency Management Agency-funded debris operations are generally carried out (FEMA 2007). However, other arrangements are possible depending on the earthquake event and what funding is available. For example, immediately after the 2011 earthquakes in Christchurch, New Zealand, the national (public) response agency “Civil Defence” carried out scores of emergency demolitions to manage public health and safety hazards and try to get city streets open quickly.

In the United States, there is a growing industry of specialist disaster waste contractors that are ready to respond to disaster events. In particular, they respond to frequent storm events. Contractors can be paid in a number of ways but typically contracts are either a lump sum (for a specific body of work) or time and cost. Generally these costs include for all aspects of waste management from demolition or collection through to final disposal.

For private property demolition contracts, it is important to be clear on who owns the waste collected from the property.

Record Keeping and Monitoring

Another important part of earthquake waste management is record keeping. Information and monitoring are invaluable when planning and operating a waste management system. Information will enable the bottlenecks and potential risks and hazards to be identified and monitoring acts as a deterrent to illegal or improper practices. Record keeping is also important for managing contracts and billing.

Often, information is difficult to track during a disaster response, particularly when there are many different organizations involved in the waste management system. Despite this, an effort, be it paper based or electronic, should be made to monitor waste movement from source to disposal.

Human Resourcing

Following disasters, resources will be stretched. Skilled waste management and emergency management personnel may have lost their lives, and it is likely that unskilled personnel may have to be engaged to assist with waste management. Earthquake waste management offers opportunities for livelihood. Job and business creation is possible during every aspect of the waste management system. It is important to train and adequately manage unskilled workers – particularly when working around hazardous materials and specialized heavy equipment.

Public Communication and Engagement

Public information and communication is essential to all aspects of disaster response and recovery and this includes waste management. Negative public reaction to waste management efforts can lead to delays and disruptions. For example, in 2010, communities in L’Aquila held protests over the slow demolition and waste management of their earthquake-affected town (Brown et al. 2010b). Following Hurricane Andrew in 1992, public protests against incineration of debris led to the abandonment of that treatment technique and authorities were forced to find an alternative management strategy (USEPA 1995).

The public needs to be kept informed of all waste management issues, including health and safety issues (and established mitigation measures), debris handling and disposal procedures, and normal municipal waste collection arrangements (where applicable). Public information can be disseminated in a number of ways including radios, local newspaper, website, television, letter drops, etc. Where possible, information should be prepared pre-disaster for ready dissemination when an earthquake strikes.

There also needs to be good communication between regulatory authorities and the waste management industry. There are many organizations involved in earthquake waste management, and it is important that they work collaboratively following a disaster event.

Funding

Managing earthquake waste can be costly. Funds generally come from both public and private sources. Typically, private property demolition (and subsequent waste management costs) will be funded by the property owner or through their insurance. Debris removed from public spaces (such as on roads and in waterways) will be funded publically – through government or donated funds. There are circumstances where public money will fund private demolition and waste management works. This may be the case where private owners have few financial resources (such as in Haiti in 2010) or where there is a public health and safety risk (e.g., following 2005 Hurricane Katrina) or where there is a public commitment to fund private earthquake recovery works (such as in L'Aquila in 2009). The type, source, and amount of funding will generally dictate what operational strategy will be used.

Regulations

Solid waste management, particularly in developed countries, is governed by diverse legislation and regulation to minimize the potentially harmful effects of waste on humans and the environment. There are public health and environmental regulations around demolition, waste transportation, waste handling and treatment, and waste disposal. In the wake of a disaster, these normal regulations can cause significant delays in the waste management process. Authorities following many earthquake (and other disaster) events have used legal waivers to reduce or circumvent regulatory requirements (Brown et al. 2010a). For example, in 2011 in the wake of the Christchurch, New Zealand, earthquakes, a temporary staging area and recycling facility was approved without going through the normal approval process to enable quick establishment of the facility (and processing of earthquake waste) (Brown 2012). If legal waivers are taken, mitigation measures must be put in place to reduce potential long-term effects.

As well as environmental and public health legislation, earthquake waste managers will also need to be aware of regulations around demolition (and associated rights to the waste) of structures. Where owners are not present or do not consent, more processes may need to be followed before demolition can take place. In some instances, authorities may wish to waive or streamline these regulatory requirements. This was done following the 2005 Hurricane Katrina because many residents were absent but houses had to be removed to enable the recovery.

Planning for Earthquake Waste

Planning for earthquake (and disaster waste generally) has only really started in the last 20 years (Brown et al. 2011). The United States was the first country to have established disaster debris management guidelines (FEMA 2007; USEPA 2008), largely because of the frequency of events in the United States. These guidelines offer technical and operational guidance and are written specific to the United States context and disaster management and funding arrangements. Recently the Joint United Nations Environment Programme/Office for Coordination of Humanitarian Affairs Environment Unit has prepared disaster waste management guidelines specifically for less developed countries (UNOCHA 2011). The guidelines cover many of the technical issues addressed in the USEPA guidelines, and management and implementation strategies are designed for countries with little or no existing infrastructure and/or waste management expertise. Opportunities for livelihood promotion and maximizing value from the resources are also emphasized in this document.

Ideally every community subject to disaster events should develop disaster waste management plans. Plans need to be developed collaboratively with all relevant stakeholders. These plans should be written to align with expected hazards and damage as well as local regulations and waste management practices. Plans should be written generically such that anyone can pick them up and use them and they should be

reviewed regularly. Plans should cover all the aspects discussed in this chapter and are summarized below.

Summary

Earthquakes generate significant volumes of waste. The waste needs to be managed in a coordinated manner to minimize impacts on public health, environmental health, and on the recovery of the community. To prepare for managing earthquake waste, communities should develop waste management plans. Plans should cover the items discussed in this chapter.

Waste composition and quantity	Earthquake waste predominantly consists of construction and demolition waste. The composition and quantity of the waste will depend on the nature of the built environment and the severity of the earthquake. There are three quantity estimation approaches: unit, volume, and area estimates
Waste collection	The majority of earthquake waste is collected by demolition and building contractors carrying out demolitions and repairs. If there is a lot of small debris, curbside collection services or debris drop-off centers can be useful
Waste handling and treatment	Waste handling facilities are often required to separate waste into recyclable and nonrecyclable material and to create a buffer to ease the pressure on existing waste facilities. These are called temporary staging areas. After waste has been sorted, it can be treated in a number of ways including recycling and incineration or used as land reclamation and engineering fill. If the waste is contaminated with hazardous materials or is wet, recycling or reuse may not be feasible
Waste disposal	Ultimately some or all of the waste has to be permanently disposed of in a landfill. Shortage of space in existing landfills may be a problem so alternative disposal sites may need to be considered

(continued)

Hazard management	Earthquake waste generally includes numerous environmental and public health hazards, including rotten food, asbestos, fecal-contaminated material, treated timber, and household and industrial hazardous materials. These hazards need to be assessed and risks mitigated as soon as possible
Waste transportation	Waste needs to be transported throughout the management process. Trucks not usually used for waste management will often be needed. Authorities need to consider and manage any potential risks associated with transportation
Strategic and operational management	There are many different authorities and industries involved in waste management. It is important that they are all included in the planning and implementation of the waste management system. Ideally one organization should be given responsibility to oversee the waste management process In terms of operations, contractors are typically engaged on a lump sum or time and cost basis. Contract templates should be drafted pre-event
Funding	Funding for management of earthquake waste may come from a range of private and public sources. It is important to consider funding when designing the waste management system
Regulations	Regulations govern many parts of earthquake waste management: environmental, public health, building, and waste ownership regulations. Legal waivers have and can be used to expedite disaster responses and recovery, but this must be done with caution

Cross-References

- [Community Recovery Following Earthquake Disasters](#)
- [Damage to Buildings: Modeling](#)
- [Damage to Infrastructure: Modeling](#)
- [Earthquake Risk Mitigation of Lifelines and Critical Facilities](#)
- [Earthquakes and Their Socio-economic Consequences](#)

- **Economic Recovery Following Earthquakes Disasters**
- **Emergency Response Coordination Within Earthquake Disasters**
- **Legislation Changes Following Earthquake Disasters**
- **Liquefaction: Performance of Building Foundation Systems**
- **Post-Earthquake Diagnosis of Partially Instrumented Building Structures**
- **Resiliency of Water, Wastewater, and Inundation Protection Systems**
- **Resourcing Issues Following Earthquake Disaster**

References

- Basnayake BFA, Chiemchaisri C, Visvanathan C (2006) Wastelands: clearing up after the tsunami in Sri Lanka and Thailand. *Waste Manag World* 2006:31–38
- Baycan F (2004). Emergency Planning for disaster waste: a proposal based on the experience of the Marmara earthquake in Turkey. Paper presented at the international conference and student competition on post-disaster reconstruction 'Planning for reconstruction', Coventry. <http://www.grif.umontreal.ca/pages/papers2004/Paper%20-%20Baycan%20F.pdf>
- Booth W (2010) Haiti faces colossal and costly cleanup before it can rebuild, *The Washington Post*. Retrieved from <http://www.washingtonpost.com/wp-dyn/content/article/2010/03/06/AR2010030602544.html?hpid=artslot>
- Brown C (2012) Disaster waste management: a systems approach. PhD, University of Canterbury, Christchurch. Retrieved from http://www.resorgs.org.nz/images/stories/pdfs/theses/brown_thesis_june12.pdf
- Brown C, Milke M, Seville E (2010a) Legislative Implications of managing disaster waste in New Zealand. *New Zeal J Environ Law* 14:261–308
- Brown C, Milke M, Seville E, Giovanazzi S (2010b). Disaster waste management on the road to recovery: L'Aquila earthquake case study. Paper presented at the 14 European conference on earthquake engineering, Ohrid
- Brown C, Milke M, Seville E (2011) Disaster waste management: a review article. *Waste Manag* 31:1085–1098
- Di Coma C (2010) Procedure e strumenti per la gestione delle macerie da crolli e demolizioni e da interventi di ristrutturazione immobiliare Direzione di Comando e Controllo funzione tutela ambientale. Il Capo del Dipartimento della Protezione Civile, accessed, Abruzzo, Italy
- FEMA (2007) Debris management guide public assistance. Federal Emergency Management Agency, United States
- Hirayama N, Kawata Y, Suzuki S, Harada K (2009) Estimation procedure for potential quantity of tsunami debris on tsunami earthquake disasters. Paper presented at the Sardinia 2009, twelfth international waste management and landfill symposium, S. Margherita di Pula, Cagliari
- Pilapitiya S, Vidanaarachchi C, Yuen S (2006) Effects of the tsunami on waste management in Sri Lanka. *Waste Manag* 26(2006):107–109
- Taylor A (2008) Sichuan's earthquake, six months later, *The Boston Globe*. Retrieved from http://www.boston.com/bigpicture/2008/11/sichuans_earthquake_six_months.html
- UNDP (2006) Tsunami recovery waste management programme multi donor fund for Aceh and Nias. UNDP, Indonesia
- UNEP (2012) Managing post-disaster debris: the Japan experience. United Nations Environmental Programme, Geneva
- UNOCHA (2011) Disaster waste management guidelines (vol. January 2011). Emergency Preparedness Section, Joint UNEP/OCHA Environment Unit, Geneva
- USEPA (1995) Planning for disaster debris. Retrieved from <http://www.epa.gov/osw/conservation/rrr/imr/cdm/pubs/disaster.htm>
- USEPA (2008) Planning for natural disaster debris. Office of Solid Waste and Emergency Response, Office of Solid Waste, USEPA, Washington, DC

Wavelet Domain Seismic Correction

A. A. Chanerley^{1,2} and N. A. Alexander³

¹School of Architecture, Computing and Engineering (ACE), University of East London, London, UK

²The University of Auckland, Auckland, New Zealand

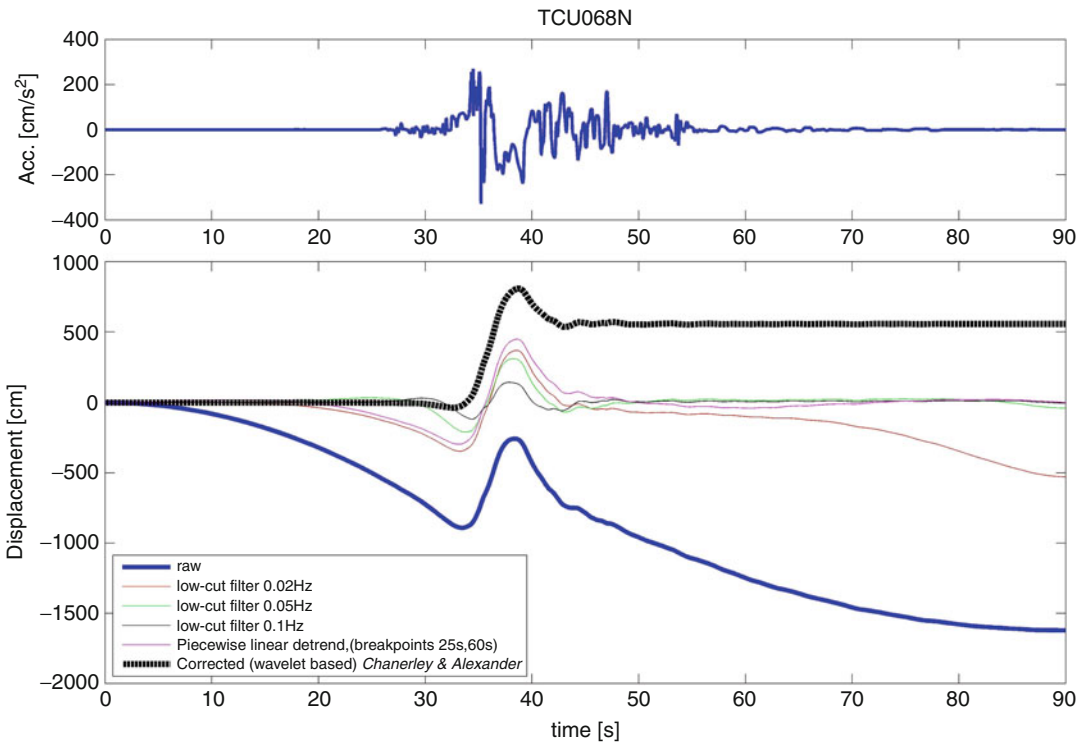
³Civil Engineering Department, University of Bristol, Bristol, UK

Synonyms

Correction fling; Decimation; De-noising; Filters; Interpolation; Seismic; Spectra; Wavelet

Introduction

Unfortunately, recovering displacements from strong-motion acceleration time histories is not



Wavelet Domain Seismic Correction, Fig. 1 Comparison of estimate displacement time series (for 1999 Chi-Chi event, station TCU068N) using various correction schemes

straightforward. Direct double integration does not yield a stable displacement time history as is shown in Fig. 1. The displacement time histories in Fig. 3 demonstrate the sort of trends obtained from double integrating after filtering with standard low-cut filters with differing cutoff frequencies.

The dashed-line displacement curve is the wavelet-corrected curve which shows a good estimate, as compared with GPS, of the displacement for station TCU068N. The other displacement curves show the difficulties usually encountered in obtaining reasonable estimates of displacement. The reason for this can be due to noise and/or baseline shift, the latter being a DC shift in the acceleration and velocity baseline, which prevents the velocity from reaching zero and therefore gives rise to an upward or downward linear/quadratic trend in the displacement. It is inferred that this is due to the instrument pitching, rolling, and yawing as a result of the turbulent ground motion.

Sophisticated methods exist for correcting baseline errors and obtaining stable double-time integration. Trifunac (1971, 1972), Graizer (1976, 2005) advocated a baseline correction procedure by obtaining and fitting a straight line to a segment of the velocity. Chiu (1997) high-pass filtered before integration. Iwan et al. (1985) removed pulses and steps by locating the time points which exceeded a predefined acceleration, later generalized by Boore (2001) and Boore and Bommer (2005), then added further time points, and made the time points t_1 and t_2 free of any acceleration thresholds, the accumulated effects of these baseline changes being represented by average offsets in the baseline. Wu and Wu (2007) also used a modified method due to Iwan et al. (1985) on the Chi-Chi event and defined t_1 at the beginning of the ground motion and defined t_2 on the basis of a flatness coefficient and defined a further parameter t_3 , the time at which the displacement had reached a final value. Wang et al. (2003) removed pulses and steps fitted

with amplitudes which gave the same areas as the slope of the displacement to achieve stable double integration. Pillet and Virieux (2007), on the other hand, using data from station TCU068 from the 1999 Chi-Chi earthquake, recovered a baseline error as an average acceleration from a linear trend in the velocity and removed it at the time points where the velocity crossed the zero axes. Graizer (2006) proposed a method for identifying the existence of tilts by comparing the sensitivity of the vertical and horizontal components to rotational ground motion by comparing their Fourier spectra, given that the vertical component is insensitive to tilts. It has been shown experimentally that tilt-induced acceleration has to be removed to recover estimates of the true ground motion. The aforementioned methods, however, require some expertise to implement. On the other hand, the algorithm in the wavelet domain Chanerley and Alexander (2010) and Chanerley et al. (2013) presented here is automated, gives immediate results, and is easy to apply. The algorithm uses the undecimated wavelet transform, with a de-noising scheme, which is the key to its success. The advantage of using the undecimated transform with de-noising is that it is, as mentioned, automated; does all the work and recovers the low-frequency “fling” time history; locates an acceleration transient, i.e., the baseline error; and permits stable integration to displacement.

The troublesome baseline error, which for some decades has made direct integration difficult if not impossible, is recovered and found to be embedded in the low-frequency “fling.” It manifests as an acceleration transient (“spike”) in all three translational components, which on integration shifts the DC level of the velocity and causes linear trends in displacement. It has been shown in Graizer (2005) and in Chanerley et al. (2013) that the vertical component is insensitive to tilts; therefore, any “spikes” in the vertical direction in the acceleration time history are usually attributed to instrument noise and indeed are usually small when compared with the “spikes” of the horizontal components, which are attributed to ground rotation. This use of the undecimated wavelet transform recovers the

residual displacement and extracts and then removes the acceleration transients and permits double-time integration to displacement. This shows its usefulness and flexibility in being able to provide not only displacements but in addition locate the baseline error and find information on the rotational acceleration transients hitherto always inferred, but never located. We begin this entry with the section on De-noising.

De-noising Overlapping Spectra Using a Threshold

De-noising is a nonlinear method of removing unwanted signals; it’s advantageous because the spectra of signal and noise can overlap. It works by applying a threshold to outlier amplitudes and has been applied together with the undecimated wavelet transform to seismic events when correcting records. De-noising is superior to filtering, the latter being a function of frequency in the sense that filtering may remove or attenuate those frequencies, which in fact we want to retain. In particular in this case, filtering low-frequency noise will also filter the low-frequency signal (i.e., the fling); therefore, filtering is to be avoided.

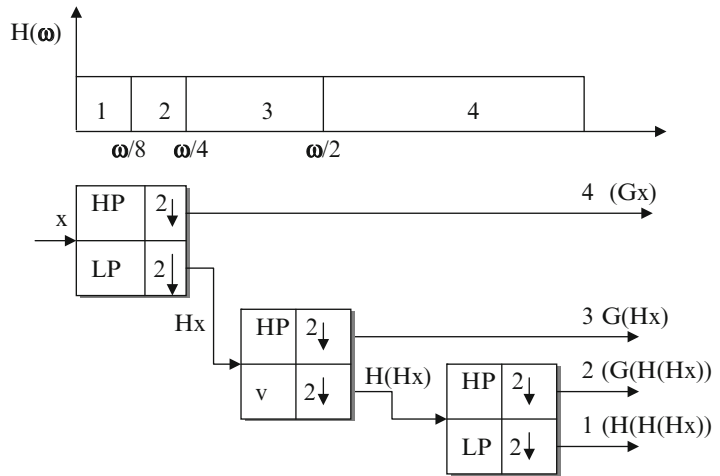
The application of a soft-de-noising scheme (Donoho 1995; Lang et al. 1996; Coifman and Donoho 1995) to the low-frequency sub-bands, after applying the wavelet filters, retained the detail without affecting the overall displacement result such that when overlaying the original time history on the corrected time history, it didn’t demonstrate any significant differences.

A soft threshold time series $\tilde{a}_i$, where $i \in \{1, 2, \dots, N\}$, is given by Eq. 1, where τ is a threshold value. Hence, the low-power components of x_i are de-noised (removed):

$$\tilde{a}_i = \begin{cases} (a_i - \tau)\text{sgn}(a_i) & |a_i| > \tau \\ 0 & |a_i| \leq \tau \end{cases} \quad (1)$$

Typically, the threshold τ is estimated from the standard deviation of the data σ , multiplied by Donoho’s “root two log N ” (Donoho 1995):

Wavelet Domain Seismic Correction, Fig. 2 A 4-channel analysis (decomposition) wavelet filter bank showing sub-band



$$\tau = \sigma \sqrt{2 \log N}, \sigma = \frac{MAD}{0.6745} \quad (2)$$

Though there are many other options for the selection of a threshold value, the critical component is the standard deviation σ of the noise in the data, which is unknown in a general case. Ideally, we would have recordings from the accelerographs for in situ cases where there is no earthquake to characterize the background noise levels and its statistics. However, this is not often available in practice. Furthermore, as we have indicated above, the presence of ground tilts introduces noise that is only present during the earthquake. Thus, we have an uncertain estimate of the statistics of the noise. In the case of accelerograms, we use the highest-frequency sub-band as an estimate of the noise for the purposes of computing the threshold. From this, we compute the median absolute deviation (MAD) that is a robust estimator and more resilient to outliers in a data set than the sample standard deviation. In order to use MAD as a consistent estimator for the standard deviation, it is divided by a scale factor that depends on the probability distribution of the noise (which is again unknown). For a normal distribution, this scale factor is $\Phi^{-1}(3/4) = 0.6745$, where Φ^{-1} is the inverse of the cumulative distribution function for the standard normal distribution, hence Eq. 2.

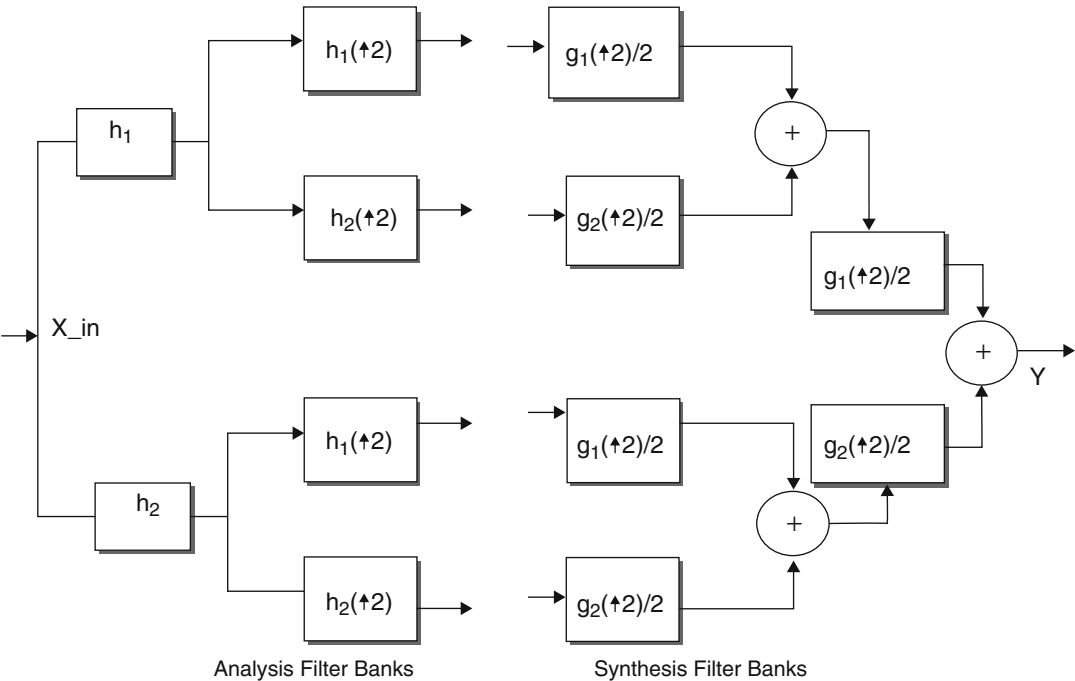
A Note on the Undecimated *à trous* Wavelet Transforms

From an engineering and seismological perspective, the easiest way to envisage the decimated/undecimated wavelet transform is in terms of octave digital filters, which are quite common in digital audio. Octave filters form the basis of the wavelet transform, whether the transform is decimated or undecimated. The wavelet transform (Daubechies 1992) comprises in this case well-designed filter banks. The decimated discrete wavelet transform (DWT) is applied in an octave-band filter bank, implementing successive low-pass and high-pass filtering and down-sampling by a factor of 2, so that every second sample is discarded. As an example, a four-channel filter bank scheme is shown in Fig. 2. The filters used in wavelet filter banks are digital finite impulse response (FIR) filters, also called non-recursive filters because the outputs depend only on the inputs and not on previous outputs.

The discrete convolution expression for a FIR filter is given by

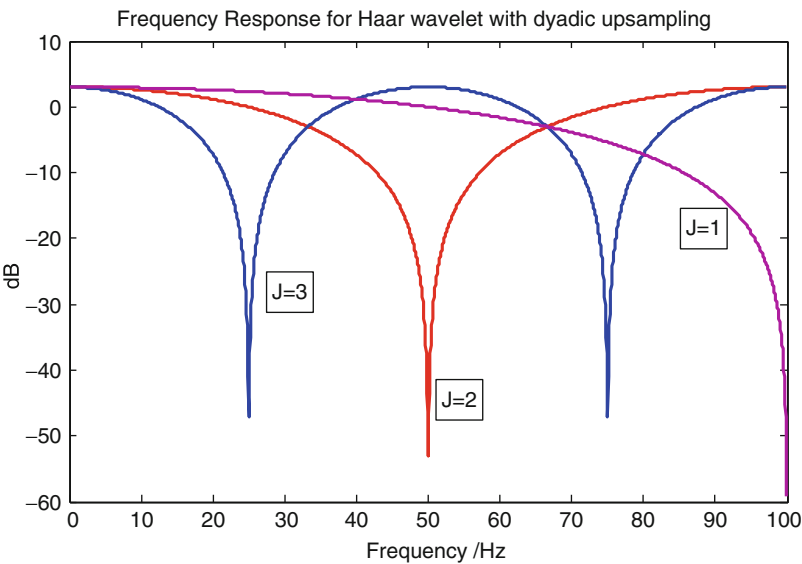
$$y(n) = \sum_{k=0}^{N-1} h(k)x(n-k) \quad (3)$$

where $h(k)$ are the filter coefficients and $x(n-k)$ are the time-shifted samples of data. Equation 3 lends itself to the multiply-accumulate data shift



Wavelet Domain Seismic Correction, Fig. 3 The undecimated or stationary wavelet transform (SWT) filter banks, showing the dyadic up-sampling of the filter coefficients, i.e., the pushing of zeros in between the coefficients

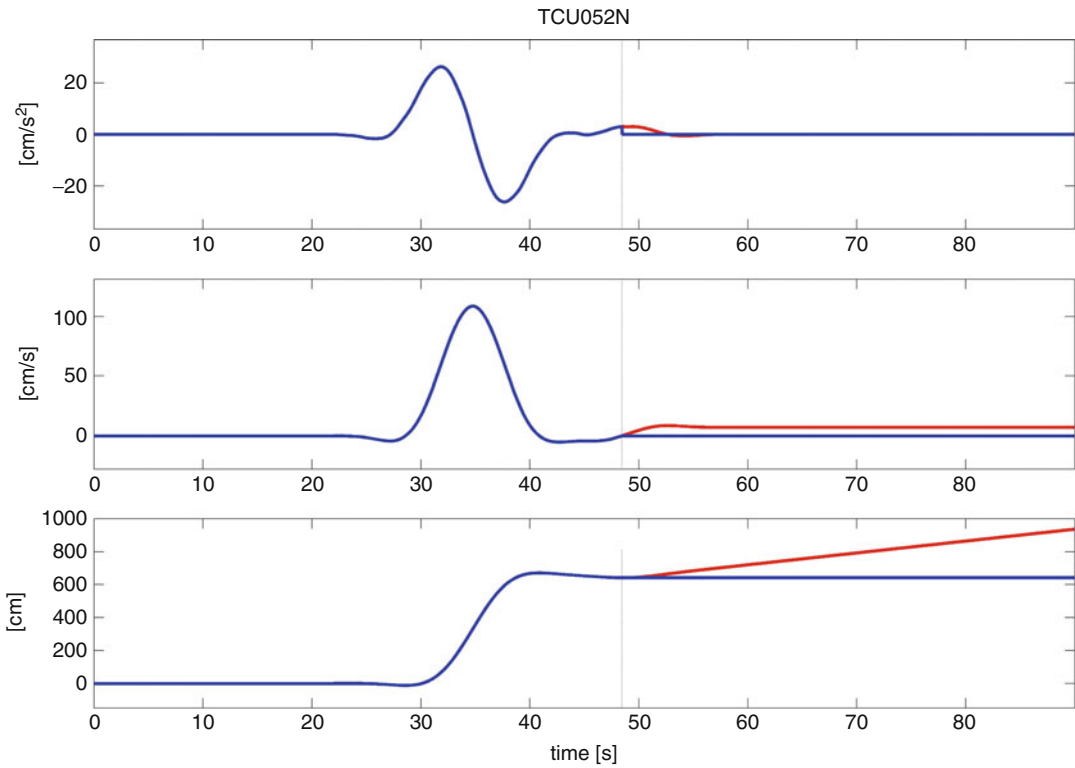
Wavelet Domain Seismic Correction, Fig. 4 The *à trous* frequency responses for the Haar wavelet at three levels of decomposition



architecture (MACD) of most modern processing engines.

Therefore, we proceed in octave sub-bands as per Fig. 2, where the filters H and G operate on the data vector x , as follows:

First octave low-frequency sub-band = Hx
Second octave low-frequency sub-band = $H(Hx)$
Third octave low-frequency sub-band = $H(H(Hx))$
First octave higher-frequency sub-band = Gx



Wavelet Domain Seismic Correction, Fig. 5 TCU052NS LFS fling, which shows results before (*red, green*) and after (*blue*) baseline correction; wavelet used is the *bior1.3*

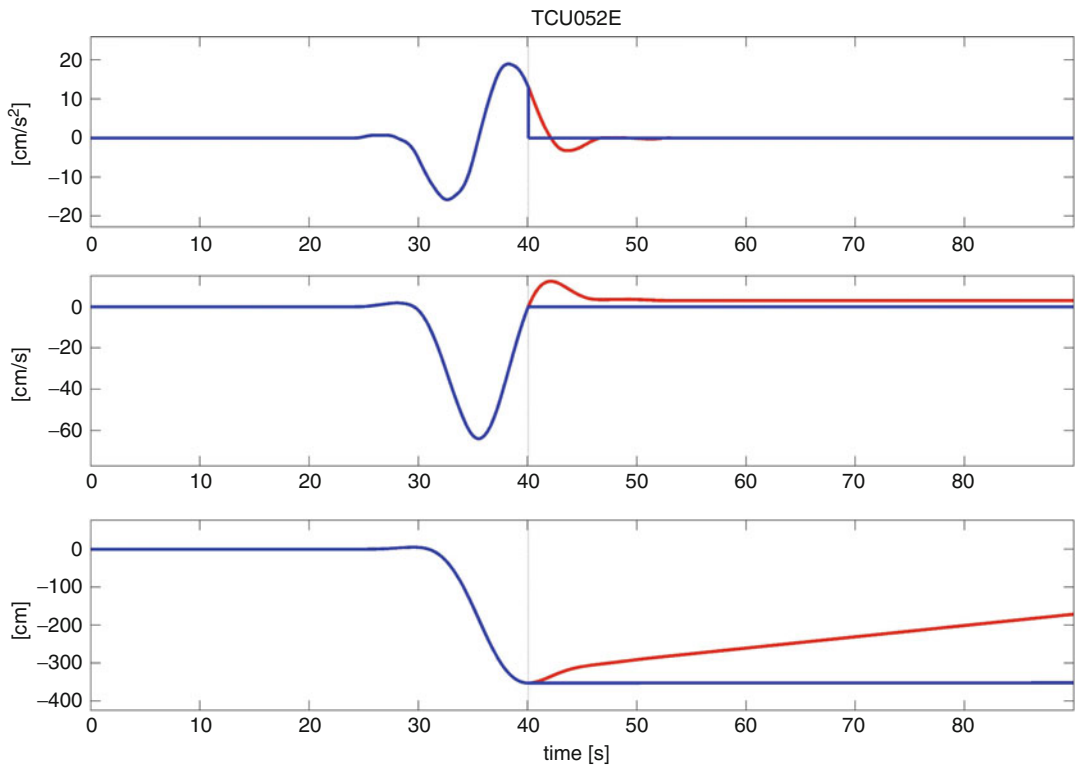
Second octave higher-frequency sub-band = $G(Hx)$

Third octave higher-frequency sub-band = $G(H(Hx))$

The behavior of the filters is to operate on a column vector of data by convoluting the data with the filter coefficients and to down-sample by 2, i.e., discard half the values.

However, there is a problem with the DWT in that due to down-sampling the DWT, it is not shift invariant. Aliasing can occur between the sub-bands, which is undesirable if the application needs de-noising as in this case. Therefore, the easiest way to overcome this is not to down-sample so that the length of the signal remains constant and not as in the decimated DWT, and so we use a generalization of the DWT, which is the undecimated wavelet transform or stationary wavelet transform (SWT), which is shift invariant.

Since we don't decimate, then instead we have to interpolate by pushing zeros into each level of the transform (i.e., between the filter coefficients), i.e., dyadically "up-sampling"; this is the *à trous* algorithm (Shensa 1992) (*trous* = holes). It was also shown that the SWT de-noises with a lower root-mean-square error than that with the standard DWT and de-noising rather than filtering are a key requirement for this application. The *à trous* algorithm imitates the subsampling (decimation) of the filtered signal as shown in Fig. 3, by up-sampling the low-pass and high-pass filters. Their impulse responses (i.e., filter coefficients) are up-sampled versions of the filter coefficients from the previous level. Figure 3 shows the undecimated case for both the analysis and synthesis of filter banks; it should be noted that data remains at the same length, but the filter coefficients are dyadically interpolated. The *à trous* transform is applied as circular convolutions and multiplications with zero elements,



Wavelet Domain Seismic Correction, Fig. 6 TCU052EW low-frequency sub-band fling, which shows results before (red) and after baseline correction (blue). The wavelet used is the *bior1.3*

which can be omitted. The frequency response for the Haar wavelet in Fig. 4 shows how the frequency band is divided by two (50Hz, 25Hz, etc.) after each up-sampling by adding zeros in powers of two. In this particular case, the low-pass filter coefficients are given by:

L1 = [0.7071 0.7071]; j = 1 (Haar/dB1)
L2 = [0 0.7071 0 0.7071 0]; j = 2
L3 = [0 0 0 0.7071 0 0 0 0.7071 0 0 0]; j = 3 etc.

The original Haar coefficients are given in L, and then the zeros are increased as $(2^{J-1} - 1)$, where J is the next level of decomposition.

In the synthesis case, the inverse *à trous* transform averages the estimates at each level again minimizing the noise, so overall the SWT is a better transform to apply.

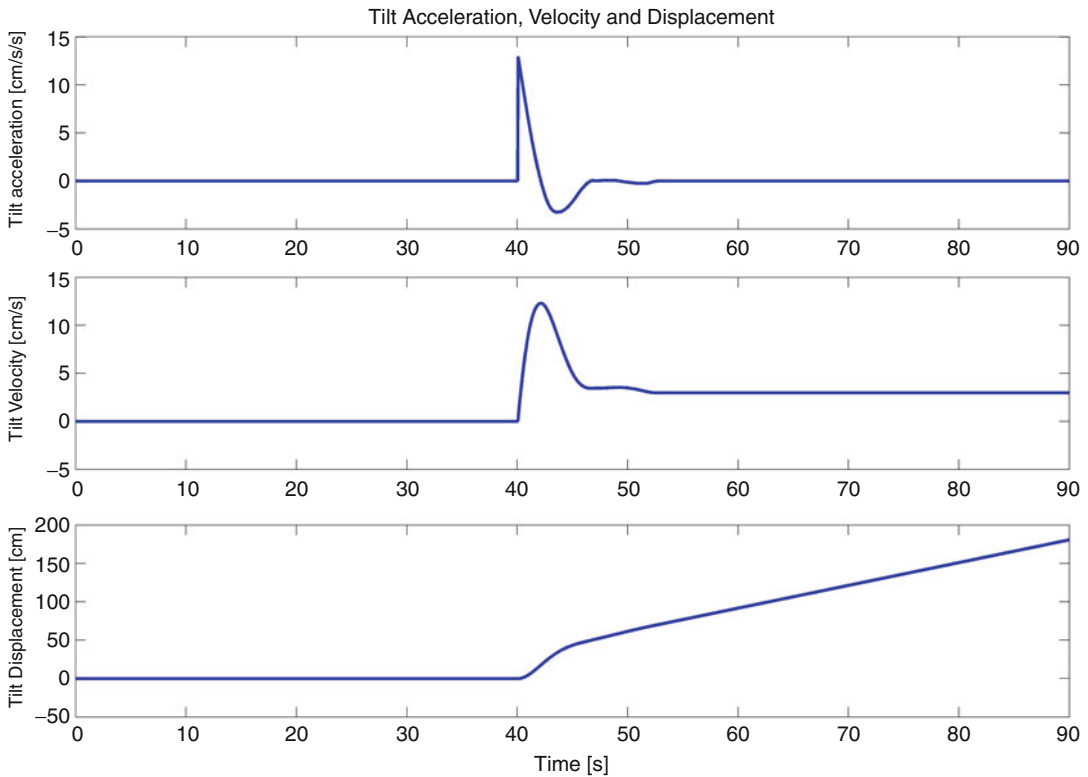
Time domain filter banks in Fig. 11 show the SWT; there are also the iSWT synthesis filter banks for reconstruction. In the analysis filter

banks, the easiest thing to do is to push zeros (*à trous*) in between the filter coefficients and keep filtering even and odd samples from every band. In the synthesis filter banks, they are averaged at each level. The data of course is also de-noised between the analysis phase and the synthesis phase using the iSWT.

Results

Some Results from the Chi-Chi (1999) Event in Taiwan: Stations TCU052, TCU068, and TCU129

The TCU052NS station from the 1999 Chi-Chi event shows the low-frequency, fling profile time history of Fig. 5, which hides an acceleration tilt transient of 3.015 cm/s/s peak, at 48.45 s. The acceleration transient is recovered by subtracting the corrected and uncorrected “fling” time history. The recovered tilt acceleration transient



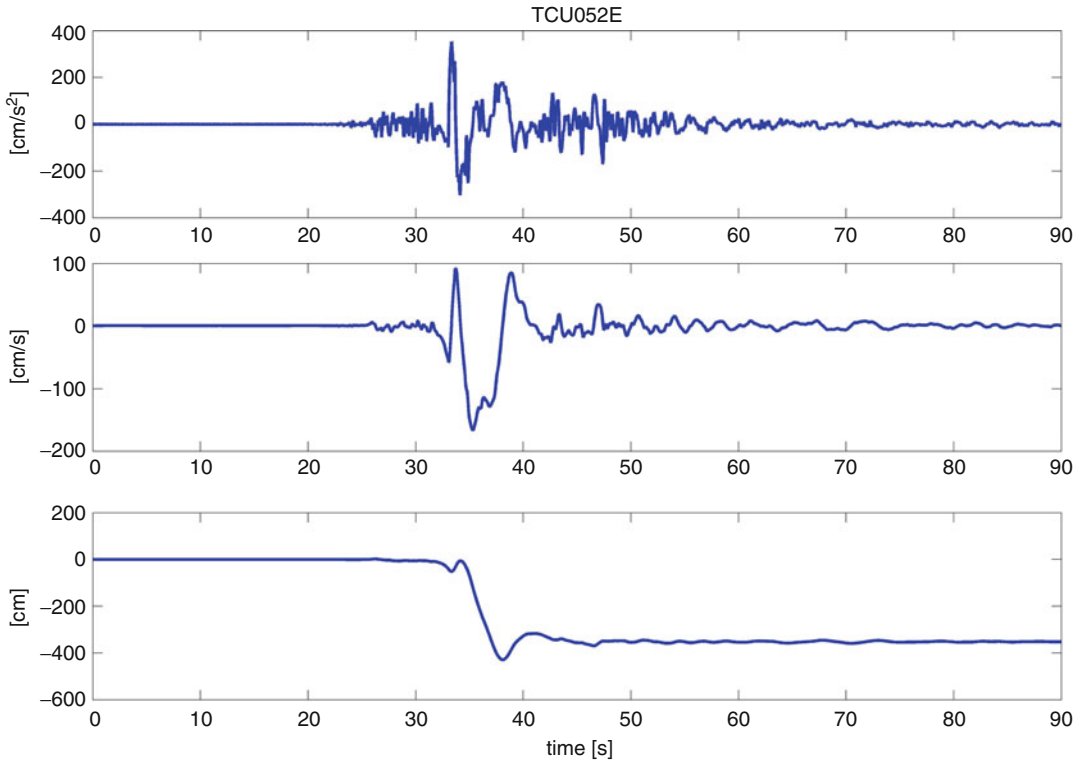
Wavelet Domain Seismic Correction, Fig. 7 Recovered tilt $g\theta$ acceleration, velocity, and displacement response of instrument A900 for the TCU052 component

leads to a velocity offset and displacement trend after double-time integration of the transient. These are precisely the error offsets observed in the velocity and displacement time histories after double-time integrating the (almost) sinusoidal acceleration time history of Fig. 5. It is inferred that acceleration transient of 3.015 cm/s/s is in fact a $g\theta$ tilt transient. The area of the acceleration tilt profile is 7.21 cm/s, which is the velocity offset. Similarly, the area under the velocity step is 294.7 cm, the maximum displacement offset as shown in Fig. 5. The corrected displacement is 641 cm, which compares with those of other researchers; the GPS reading is 845 cm. The point to note is that the area of the acceleration transient is exactly equal to the constant velocity offset after integrating the acceleration transient and observed in the velocity sub-band. Then, after integrating the constant velocity offset, then the result matches exactly the linear trend in the displacement before correction. The

TCU052EW low-frequency, fling time histories are shown in Fig. 6. In this case, the acceleration tilt transient occurs at 40.06 s, approximately 8 s earlier than the NS component. The displacement is -352 cm (GPS -342 cm); the instantaneous tilt acceleration $g\theta$, shown in Fig. 7, is 12.96 cm/s/s; and θ , the instantaneous tilt angle, is 13.2 mrad. The area of the acceleration tilt profile is 2.98 cm/s, the velocity offset.

Similarly, the area of the velocity step is 180.7 cm, which is also the maximum displacement offset as shown in Fig. 7, and Fig. 8 shows the resulting, corrected time histories. Moreover, as for the NS component, then on integrating the extracted acceleration transient, it yields the velocity DC shift and displacement offset of the velocity and displacement profile after integrating the acceleration time history as shown in Fig. 8.

Similar results are obtained for TCU068NS in Figs. 9 and 10; using the undecimated wavelet



Wavelet Domain Seismic Correction, Fig. 8 The resultant plots for TCU052EW after the corrected low-frequency sub-bands (LFS) are added to the high-frequency sub-band (HFS)

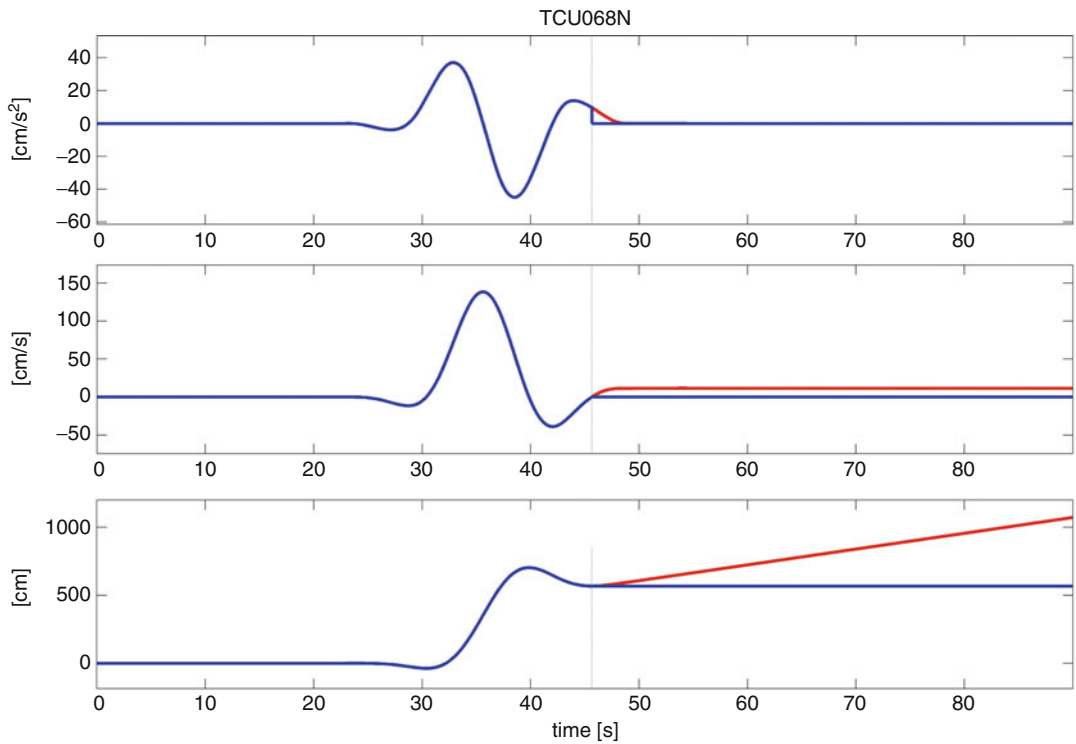
transform method, the peak instantaneous tilt angles calculated from the peak acceleration transient are calculated as -6.9mrad (-6.75 cm/s^2) (EW) and 9.8mrad (9.57 cm/s^2) (NS), and the resulting dynamic tilt amplitude obtained is 12.09mrad (11.86 cm/s^2). It is these peak acceleration transients, which are inferred as due to instantaneous tilt/rotation angles, that cause the DC shift in velocity in the latter portion of the time history and prevent stable double-time integration. These results are similar to those in (Pillet and Virieux, 2007) using a different method, with which to estimate the tilt angles. The estimates of displacement are shown in Table 1.

Station TCU129, Chi-Chi Event (1999)

Figure 11 shows the low-frequency fling for record TCU129EW after the application of the undecimated wavelet transform, before and after baseline correction. There is clear post-fling

distortion in the acceleration. At 30s (Fig. 12), there is a δ pulse-like jump in the acceleration of $g\theta = -4.8\text{ cm/s/s}$ and so an instantaneous tilt angle $\theta = 4.9\text{mrad}$ s, similar to that discussed for stations TCU052 and TCU068.

However, in addition, there is another set of distortions at 40 s to 55 s. This has led to a series of downward steps in velocity giving rise to a velocity shift of -7.5 cm/s/s from zero. The profile of these changes in the acceleration, in particular at $t > 30\text{ s}$, with a spike at 30s shown in Fig. 12 giving a step in velocity, followed by further velocity steps at $t \geq 40\text{ s}$. Of course there isn't any certitude that these distortions are due to tilts/rotations; there are other effects that could have caused these. However, there is some justification for proposing that these distortions may be due to tilts/rotation. The net permanent displacement is shown in Figs. 11 and 13 as 78.48 cm . The GPS station AF11 at 2.3 km away measured a permanent displacement of



Wavelet Domain Seismic Correction, Fig. 9 TCU068NS low-frequency sub-band fling, which shows results before (*red*) and after (*blue*) baseline correction. The triangular area in the acceleration is the

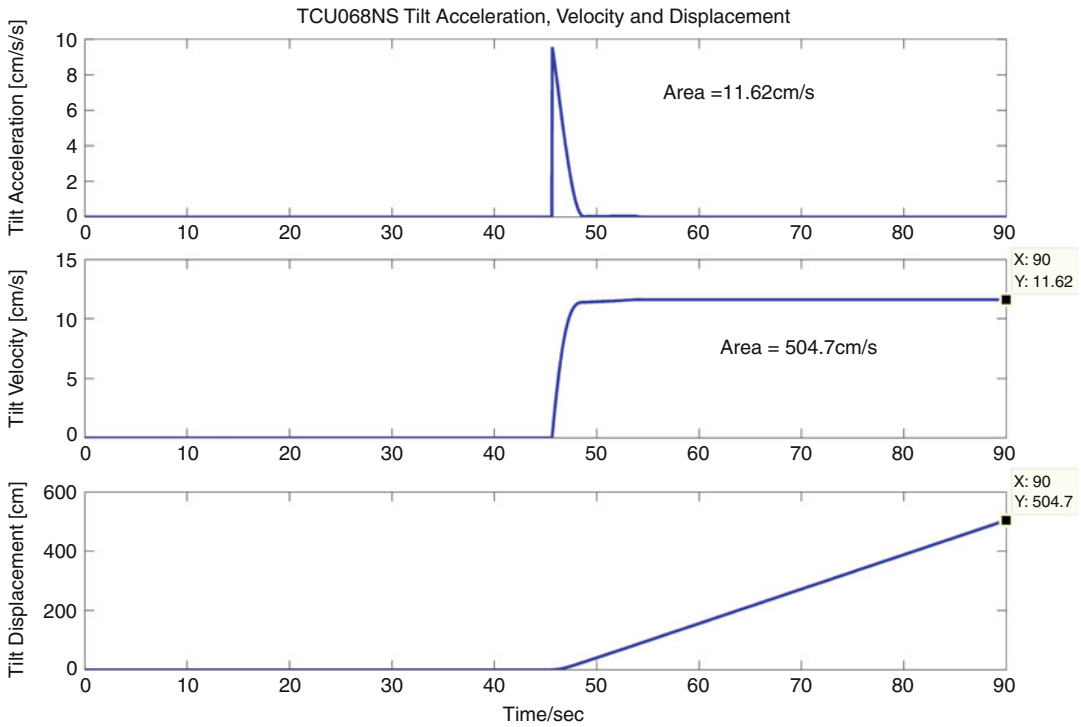
results of an acceleration transient $g\theta$ at 45.65 s. The resulting constant velocity DC shift of 11.61 cm/s is shown in light gray as a post-shaking, flat velocity time history

Wavelet Domain Seismic Correction, Table 1 Lists of the component acceleration transients and displacements and the time at which the transients occurred at which baseline corrections were applied

Greendale Station			
Component	Tilt acceleration cm/s/s	Time/s	Displacement (cm)
N55W	2.33	26.55	−177.8 (level 10)
S35W	3.44	28.56	−47.85 (level 9)
UP	0.49	27.2	−66.5 (level 10)

approximately 100 cm. The time history profiles for TCU129NS are shown in Fig. 21, and in this case the *bior2.6* wavelet was used at a level 9 of decomposition. At 28.1 s, the low-frequency fling showed an acceleration transient of magnitude $g\theta = 9.93$ cm/s/s, giving a DC shift to the latter part of the velocity time history of 7.65 cm/s. The velocity offset after integration then gave a linear trend with a DC offset of 435 cm at 90s. These sorts of baseline errors make double-time integration impossible without removing the error.

Therefore, the undecimated wavelet algorithm then zeroes out the acceleration from 28.1 s and reintegrates to give the corrected time histories shown in Fig. 14. The GPS displacement for the NS component was −32.1 cm and that obtained using the presented wavelet transform method is −26.8 cm. The TCU129V component is interesting because it suggests either at least 9 more tilts or just a very noisy instrument. However, in Chanerley et al. (2013), there is a suggestion that the location of the instrument (A900) on a



Wavelet Domain Seismic Correction, Fig. 10 Acceleration tilt transient for TCU068NS at 45.65 s. The resulting constant velocity DC shift of 11.61 cm/s is a clear example of a post-shaking, flat velocity time history

concrete pier may have been the cause of the oscillatory ringing. Certainly, there are significant oscillatory effects as can be seen from the time history in Fig. 15.

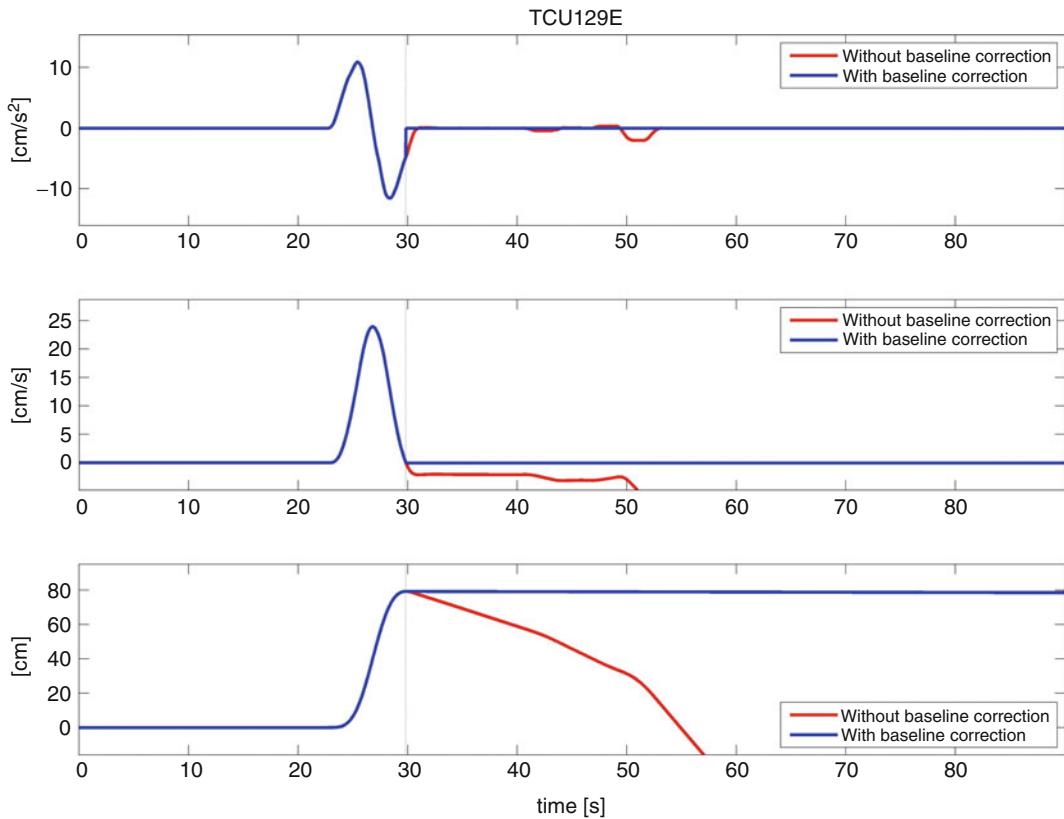
Nevertheless, the fling pulse in TCU129V is clearly visible both in the acceleration and velocity time history, and we take the zero-velocity cross-over point at 33.97 s immediately after the fling pulse in velocity zeroes the acceleration from that point and reintegrates. The first acceleration transient occurs at 33.97 s, with a magnitude of 2.12 cm/s/s. The displacement shown in Fig. 15 and Fig. 16 is -18.34 cm and that given by GPS is -17.7 cm.

At the time point where the velocity fling pulse crosses the zero axis, the acceleration is zeroed to the end of the record and reintegrated, giving a displacement of -18.34 cm; the GPS reading for that component is -17.7 cm. The acceleration transient, though not explicitly shown, is small

in magnitude (2.12 cm/s/s) compared with that of other components, suggesting its origin may be from instrument noise.

New Zealand M_w 7.1 Darfield Event, 4 September 2010

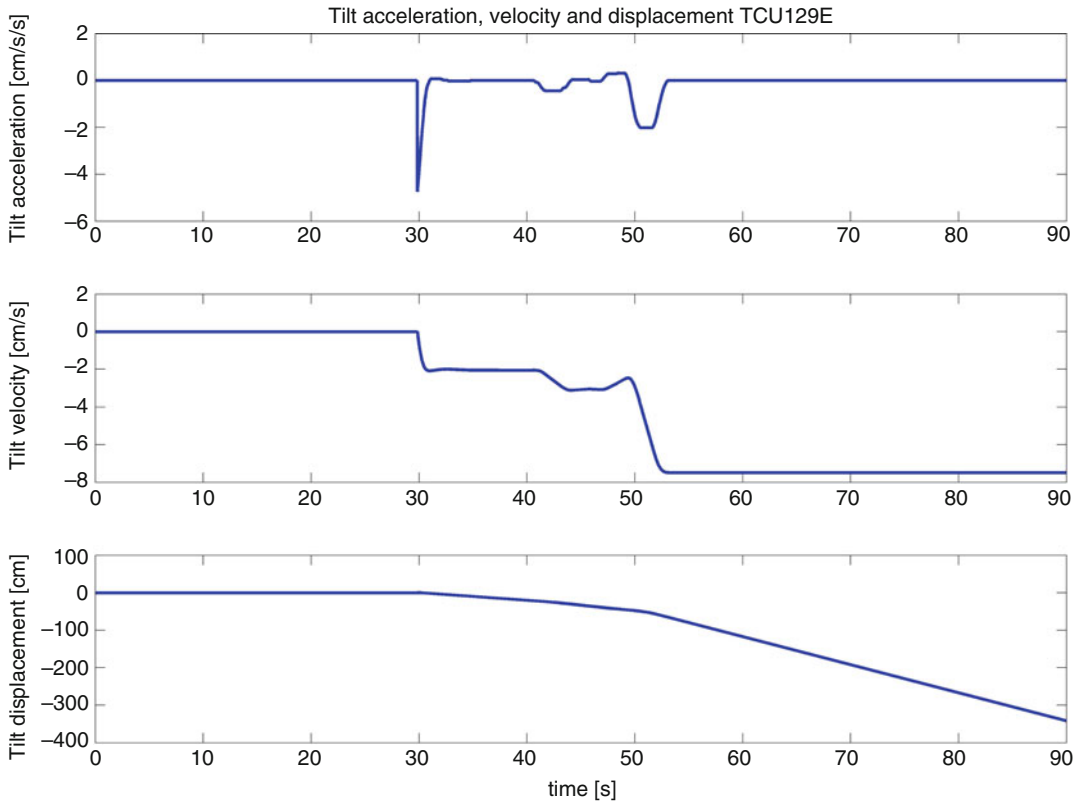
The Canterbury region on the South Island of New Zealand experienced a magnitude M_w 7.1 earthquake on the 4 September 2010. The epicenter was located about 10 km SE of the town of Darfield and about 40 km west of the city of Christchurch. The seismic event produced a right-lateral, strike-slip surface rupture ≥ 28 km long, approximately west-east. Much damage occurred to residential buildings as well as a lot of liquefaction and lateral spreading. The situation however deteriorated further after a second earthquake of M_w 6.3 hit, damaging the Christchurch city center with considerable loss of life.



Wavelet Domain Seismic Correction, Fig. 11 TCU129EW, low-frequency sub-band fling before (*red*) and after (*blue*) baseline correction showing tilt/rotation in acceleration and baseline DC shift in velocity

The fling pulse for N55W is presented in Fig. 17, and the acceleration transient and displacement results are given in Table 1 for all components from the Darfield earthquake as recorded on a 14-bit CUSP-3B Avery et al. (2004) at the Greendale station, using the wavelet transform method. More displacement results obtained using the wavelet transform method are in Chanerley and Alexander (2010), Chanerley et al. (2013), and Berrill et al. (2011). The horizontal displacements give a resultant of 183.6 cm, which compares well with GPS at 189.79 cm (Berrill et al. 2011), 1.6 km away. Table 1 also gives estimates of the tilt acceleration transients taken at the fling zero-velocity time point, at which the correction is applied.

Figure 17 shows the low-frequency, sub-band fling for the N55W component, at level 10 decomposition, and Fig. 18 shows the net acceleration, velocity, and displacement after correction, as well as the individual higher-frequency components. Table 1 also demonstrates that the vertical acceleration transient is small compared to the two horizontal components, as would be expected (Chanerley and Alexander 2010), and may be a measure of the noise floor. It should be noted that as for the Chi-Chi station TCU068, level 10 decomposition has removed any post-fling perturbations in the fling acceleration and produced an almost sinusoidal time history. The zero-velocity, cross-over time point is taken at time $T_i = 26.55$ s at the end of the fling, at which the correction is applied. The acceleration



Wavelet Domain Seismic Correction, Fig. 12 Recovered tilt $g\theta$ acceleration, velocity, and displacement response of instrument (A900) for TCU129EW component

transient at that zero-velocity, cross-over point is 2.33 cm/s^2 .

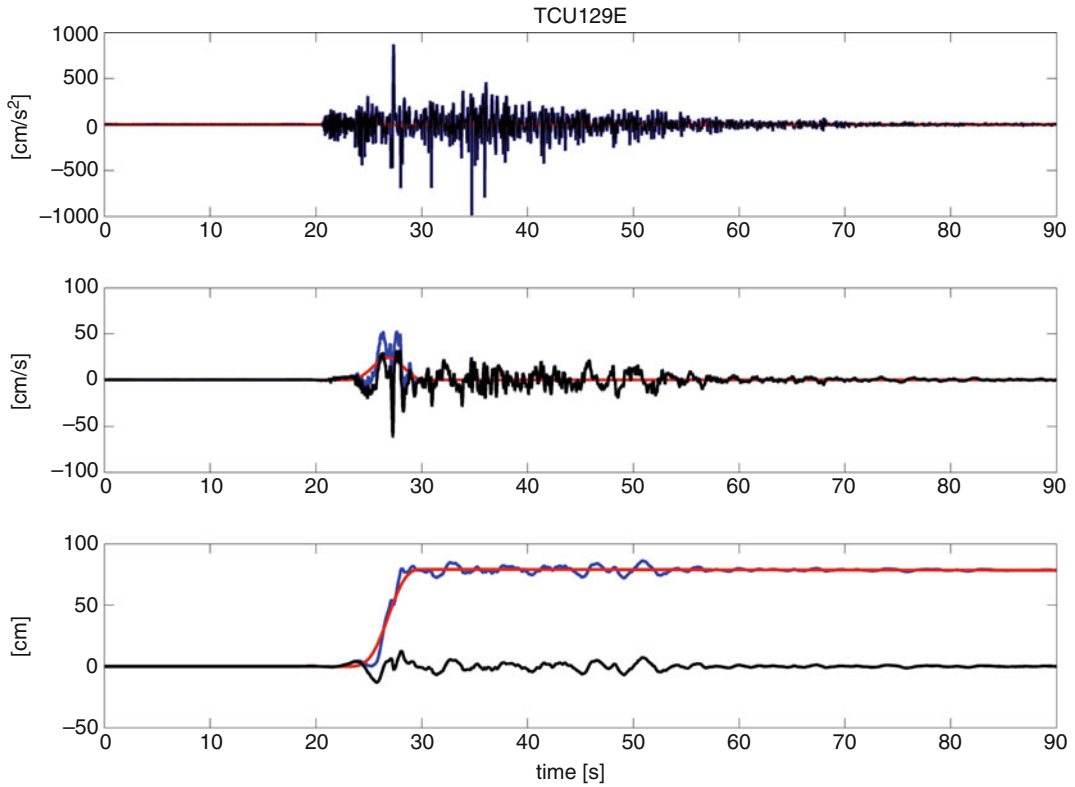
Acceleration Transients for the ICEARRAY After the M_w 6.3 Ölfus Earthquake (29th of May 2008) in the South Iceland Seismic Zone (SISZ)

The ICEARRAY is the first small-aperture strong-motion array in Iceland, which has yielded some really interesting results. The ICEARRAY has been in operation since October 2007 and was installed in the village of Hveragerdi on the western edge of the SISZ (Halldorsson and Sigbjornsson 2009). The final layout of the array comprises 14 ICEARRAY stations with an aperture of approximately 1.9 km and an inter-

element distance of 50 m (Halldorsson et al. 2009; Halldorsson and Avery 2009).

The recording system at each ICEARRAY station consists of a CUSP-3C1p strong-motion accelerograph unit manufactured by Canterbury University Seismic Project in New Zealand, for the Canterbury Network in New Zealand. The units are equipped with 24-bit, triaxial, low-noise ($\sim 70 \mu\text{g rms}$) microelectromechanical (MEM) accelerometers with a high maximum range ($\pm 2.5 \text{ g}$) and a wide-frequency passband (0-80Hz at 200 Hz sampling frequency).

The ICEARRAY yielded a useful set of data from the 11 CUSP stations, and the 33 component results for the displacements using the wavelet transform were compiled in May 2009

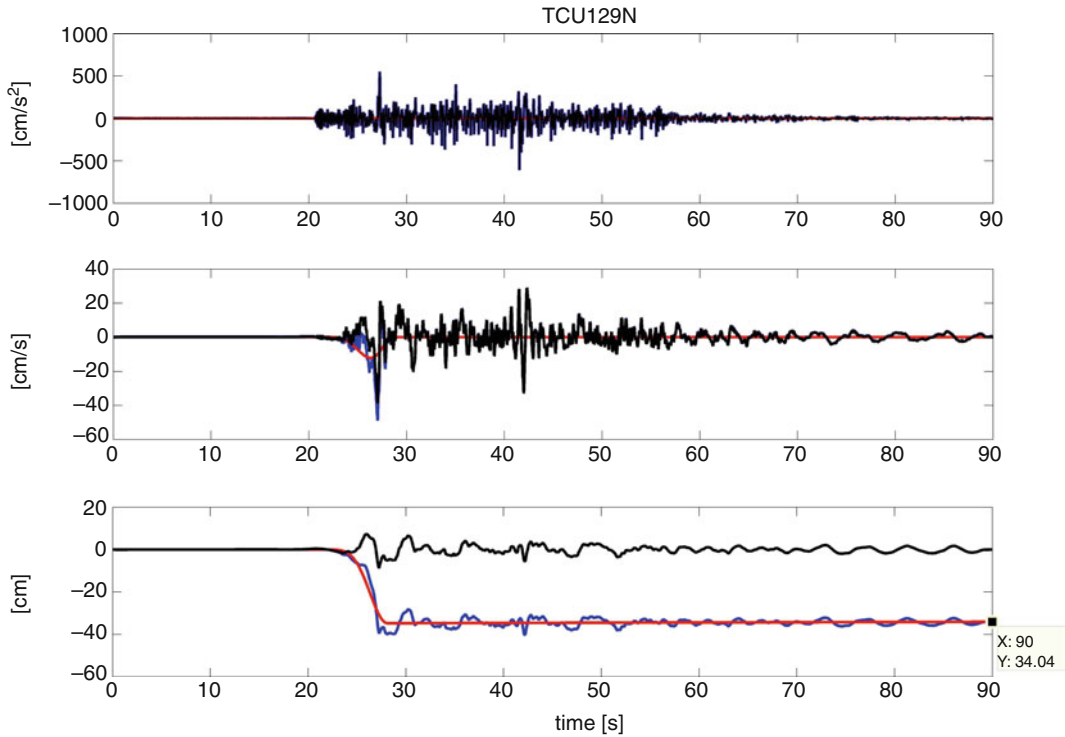


Wavelet Domain Seismic Correction, Fig. 13 Low-frequency sub-bands (*red*), high-frequency sub-bands (*black*), corrected (*blue*) using *bior1.3* for TCU129EW component obtained at level 9

(Rupakhety et al. 2010). We present some time series plots for the purpose of demonstrating the wavelet transform method for obtaining displacement estimates as well as the inferred tilt acceleration transients. Also included are the horizontal displacements obtained using the wavelet transform method in Table 2, which lists the NS and EW permanent displacements obtained for the 11 CUSP stations. A comparison of the displacement is given in Table 3 for station IS609, showing a resultant of 16.81 cm NW using the wavelet transform method, which gives reasonable correlation with a GPS station 1.6 km away at 17 cm NW, operated by the Icelandic Meteorological Office.

Figure 19 shows the corrected acceleration, velocity, and displacement for station IS604 EW as an example. The decomposition level is 7 with

only one threshold iteration and in this case using *bior1.3*. This is a somewhat lower decomposition level than the previous, probably due to a lower-magnitude earthquake compared, for example, to that of the Chi-Chi event. Level 7 produced a clear fling from which the zero-velocity, cross-over time point, $T_i = 20.21$ s, is used to remove the acceleration transient at $t \geq T_i$ and reintegrate the fling, sub-band acceleration to give a sub-band velocity without a DC shift and a displacement without any linear/quadratic errors as shown in Fig. 20. The acceleration transient is taken at the fling at $T_i = 20.21$ s and has a value of 18.6 cm/s^2 . As for the Chi-Chi events above, the higher-frequency sub-bands didn't require baseline correction and double-time integrated normally. Level 7 includes a band of frequencies up to 0.78 Hz, higher than for Chi-Chi TCU068, for



Wavelet Domain Seismic Correction, Fig. 14 Low-frequency sub-bands (*red*), high-frequency sub-bands (*black*), corrected (*blue*) using *bior2.6* for TCU129NS component obtained at level 9

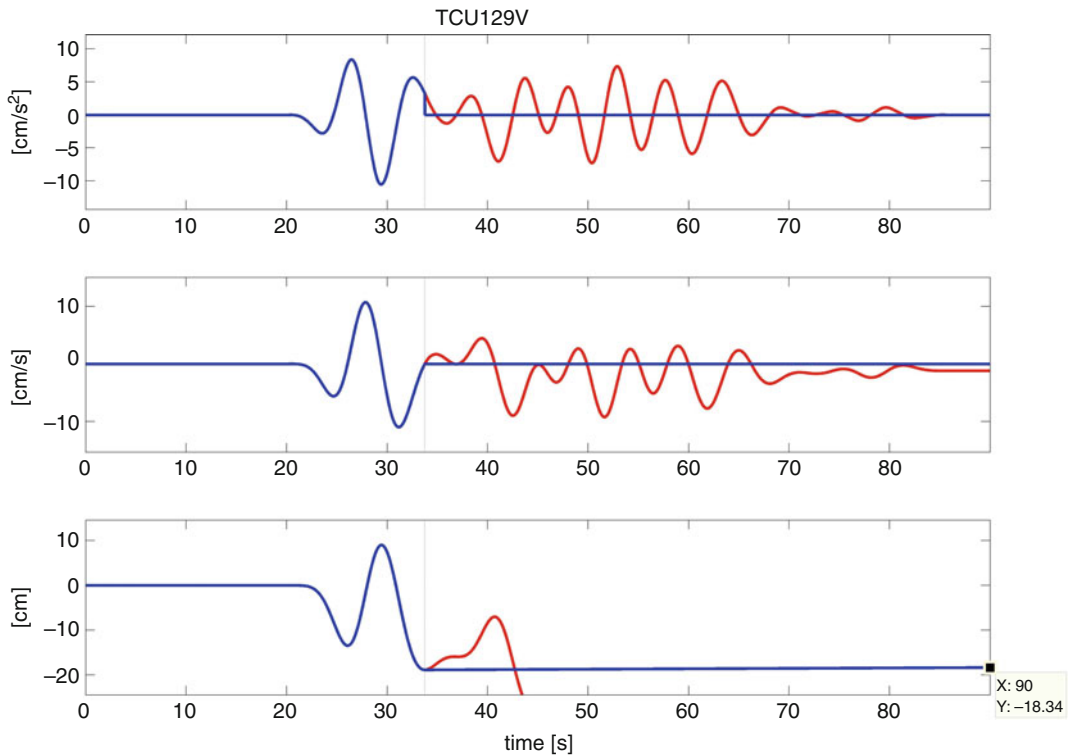
example, but then the displacements are much smaller.

Other stations for the EW component give similar results all at level 7 decomposition, and all time points, T_i , were measured at the optimal fling, and all acceleration transients were removed at that point. This gave some consistency in analysis. The focus here is on the acceleration transient at the low-frequency fling, since it is from there that the correction for baseline shift is applied and the permanent displacement obtained. Table 2 gives the complete list of acceleration transients obtained from each of the ICEARRAY stations. Table 2 shows that the EW components have the largest initial acceleration transients, though most stations had several velocity zero-crossings.

The EW component of most stations had several zero-crossing points at level 7 with smaller acceleration transients further along the time

history in much the same way as station IS604. Therefore, any one of those could have been used by the algorithm which also gave sensible displacement values. However, following Graizer (2005), a clear-cut fling profile of the velocity pulse is used by the algorithm at which time point the acceleration transient removed. The results also show that the acceleration transients and their T_i times correlate well for each component for each station, but that the NS transients are much lower in magnitude and similar to those of the vertical component and so can be attributed to noise. The much larger acceleration transients for the EW component may be indicative of tilts/rotations.

The results are consistent with the fact that the aperture of the ICEARRAY is, as mentioned, small at 1.9 km; therefore, perhaps it is right to expect some degree of correlation from such a close-knit array of sensors. This consistency in



Wavelet Domain Seismic Correction, Fig. 15 Low-frequency sub-band (red, before correction), LFS (blue, after correction) of the vertical component TCU129V

results does lend some weight to the fact that certainly in the EW component, these transient accelerations may be due to tilts/rotations.

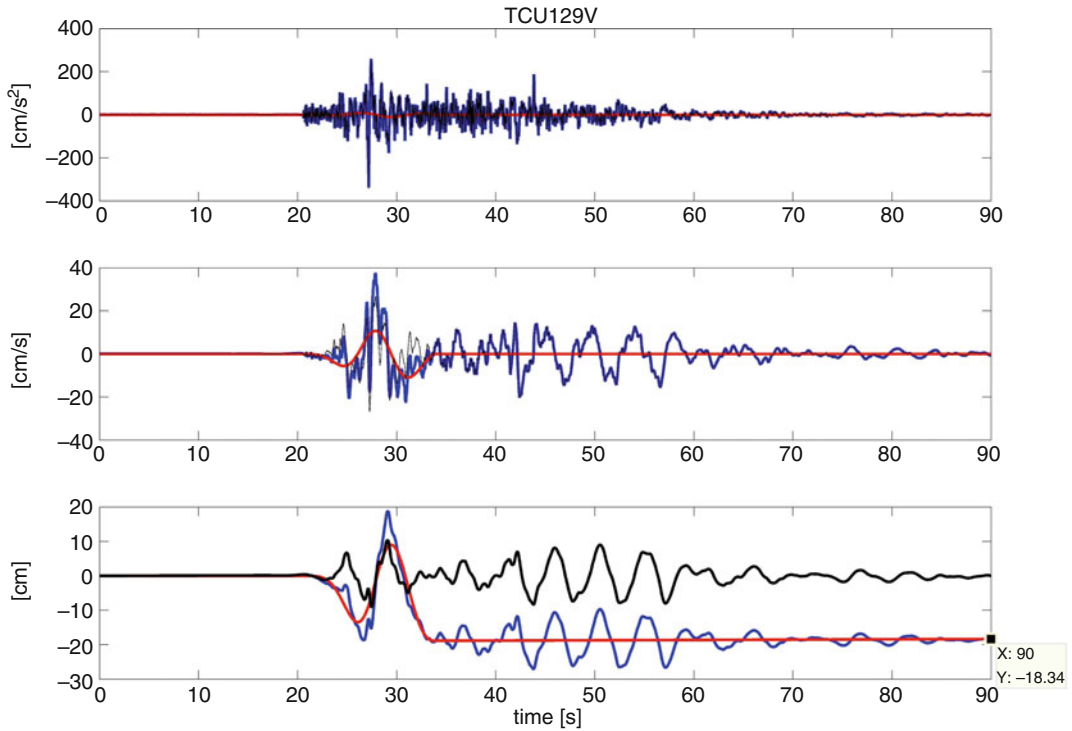
Some Permanent Displacement and Tilt Estimate Summaries

The results show some consistency in the occurrence of the acceleration transients in TCU129 and the other events displayed in Table 4. Generally, the horizontal component “tilt” acceleration transients are larger than that of the vertical component; in particular, that for TCU129NS, TCU068NS, and TCU102N are very much larger suggesting strong ground rotation.

However, it could also be argued that the component TCU129EW is too close in magnitude to that of the vertical and therefore could be attributed to noise. The results for TCU052 exhibit a large acceleration transient or “spike” for the

TCU052EW component (suggesting ground rotation), but demonstrate small acceleration transients for the TCU052NS and TCU052V, suggesting these are just noise. The disparity in the times of occurrence of the acceleration transients, for each component, also suggests that instrument noise plays a part. The results in Table 4 for TCU068 and TCU102 are consistent with predicted behavior, where the vertical components are either not responding to tilts or responding to small amounts of noise, while the horizontal components are responding to ground rotations, given their large magnitudes. In all cases for the above events, the estimates of residual displacements show good correlation with those from GPS.

The collated results in Tables 5, 6, and 7 show that the wavelet transform method (Boore and Bommer 2005; Boore 2001) produces a good



Wavelet Domain Seismic Correction, Fig. 16 LFS (red), HFS (black), corrected (blue) for TCU129V

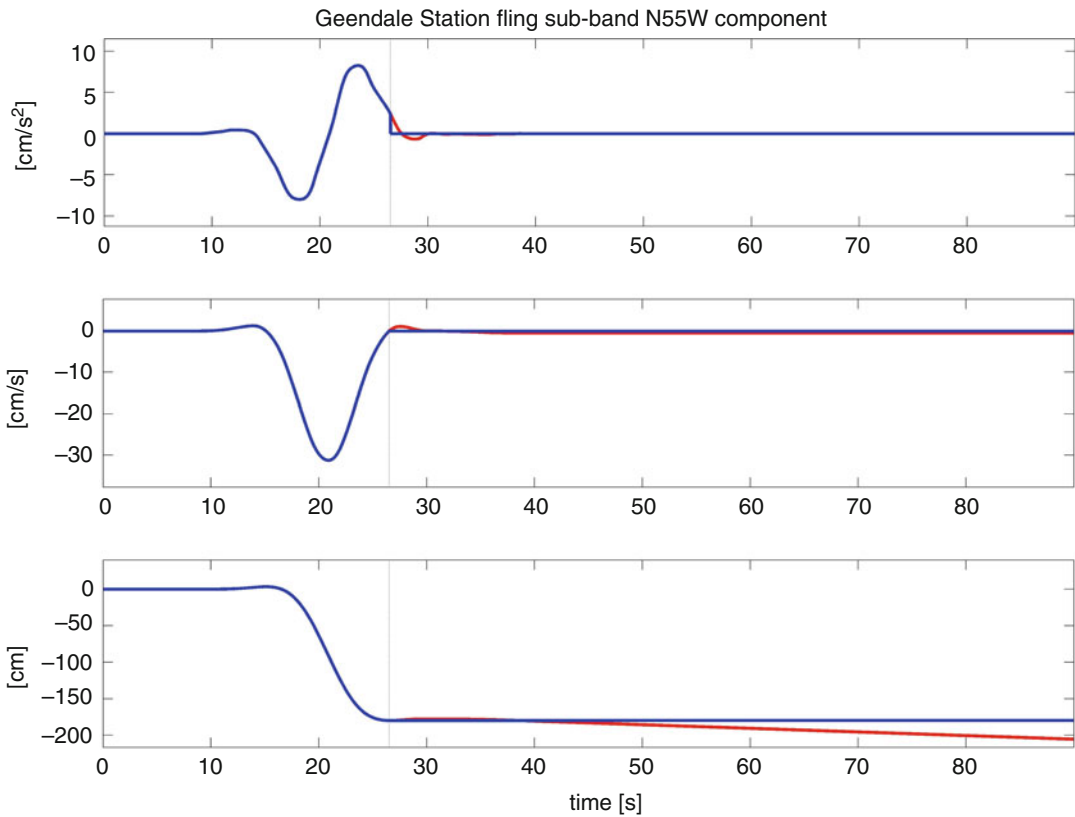
estimate of residual ground displacements and estimates of the baseline error in the form of δ -like acceleration transients (“spikes”) caused by ground rotations. There are many more results from events in New Zealand (Berrill et al. 2011), Iceland, and Taiwan in the literature published by the author and listed in the references.

Results from the Nepal 2015 event recorded at Lamjung Station 2015 06:11:26 UTC

The recent 25 April 2015 event in Nepal occurred at a depth of approximately 15 km, with the Indian plate pushing under the Eurasian plate, the result of thrust faulting on or near the main thrust interface between the subduction of the Indian plate and the Eurasian plate to the north. The earthquake activity in Nepal is caused by the

ongoing continental collision between India and Asia, which formed the Himalayan Mountains and the Tibetan Plateau. At the location of this earthquake, the Indian plate is converging with Eurasia at a rate of 45 mm/year, toward the north-northeast, driving the uplift of the Himalayas and the Tibetan Plateau. The earthquake was damaging in this case because the capital Kathmandu is located in a basin filled with soft sediment. These basins of sediment exert a large effect on the ground motion above, because they amplify the amplitude of the earthquake wave above as the wave slows down at the rock–sediment interface.

The results presented are from the Lamjung Station in Lamjung, Nepal, which lies about 180 km from Kathmandu and approximately 31 km ESE from the epicenter. The results show some consistency with radar measurements (Fig. 21) from NASA/USGS and from the results obtained by Professor Roger Bilham from the University of Boulder, Colorado, who predicted such an



Wavelet Domain Seismic Correction, Fig. 17 N55W component time series, *bior1.3* wavelet, at decomposition level 10, showing low-frequency fling with an (*tilt*)

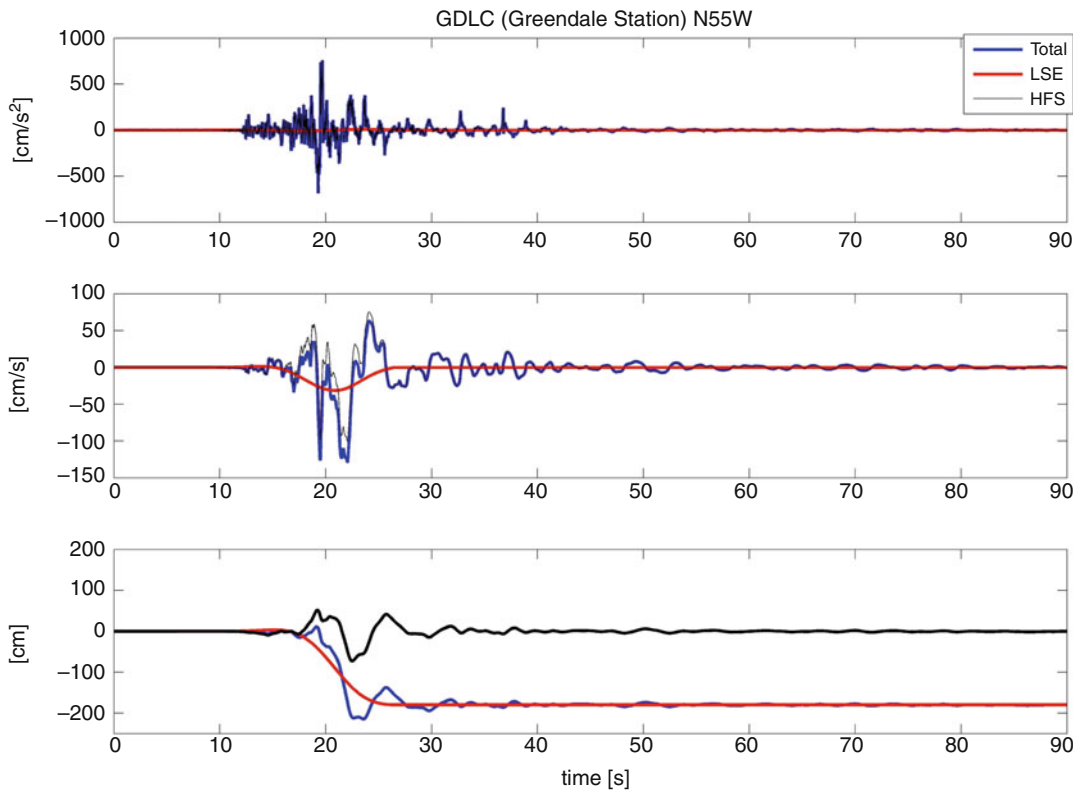
acceleration transient of 2.33 cm/s/s at 26.55 s giving a velocity DC shift of -0.5 cm/s and a displacement offset of 25.8 cm. Instrument is CUSP 3B

event in the 1990s (Bilham 1994; Bilham et al. 1995).

The results of using the wavelet transform method are shown in Figs. 22, 23, and 24.

Professor Roger Bilham describes from Kathmandu that “the rupture front travelled 90 km from the epicentre at 2.5 km/s towards Kathmandu, arriving at the western outskirts of the valley 35 s after the mainshock. There it slowed to 1 km/s as it ruptured 10 km beneath Kathmandu eastwards. Strong motion data and GPS high rate sampled data (5 Hz) in the Kathmandu Valley indicate that in less than 5 s the city was heaved upwards by 60 cm, and to the SSW by 1.5 m, at velocities of up to 50 cm/s.” The Lamjung station results, using the wavelet transform

method (Chanerley and Alexander 2010; Chanerley et al. 2013), are largely vindicated by Professor Bilham’s findings. The Lamjung station vertical results from the USGS raw data are as Professor Bilham describes; the graphs below show that indeed for the Lamjung station UP profiles, from the start of the fling step at 40 s, the displacement upward is 60.07 cm at 45.5 s and continues to then leave a residual displacement of 1.44 m up at the end of the fling step, which is consistent with information from NASA/USGS that radar shows an uplift of 1.5 cm. The maximum velocity of 55.36 cm/s for Lamjung station UP is consistent with the velocity as stated of 50 cm/s, and the maximum fling velocity is 23.16 cm/s, the latter giving rise to the



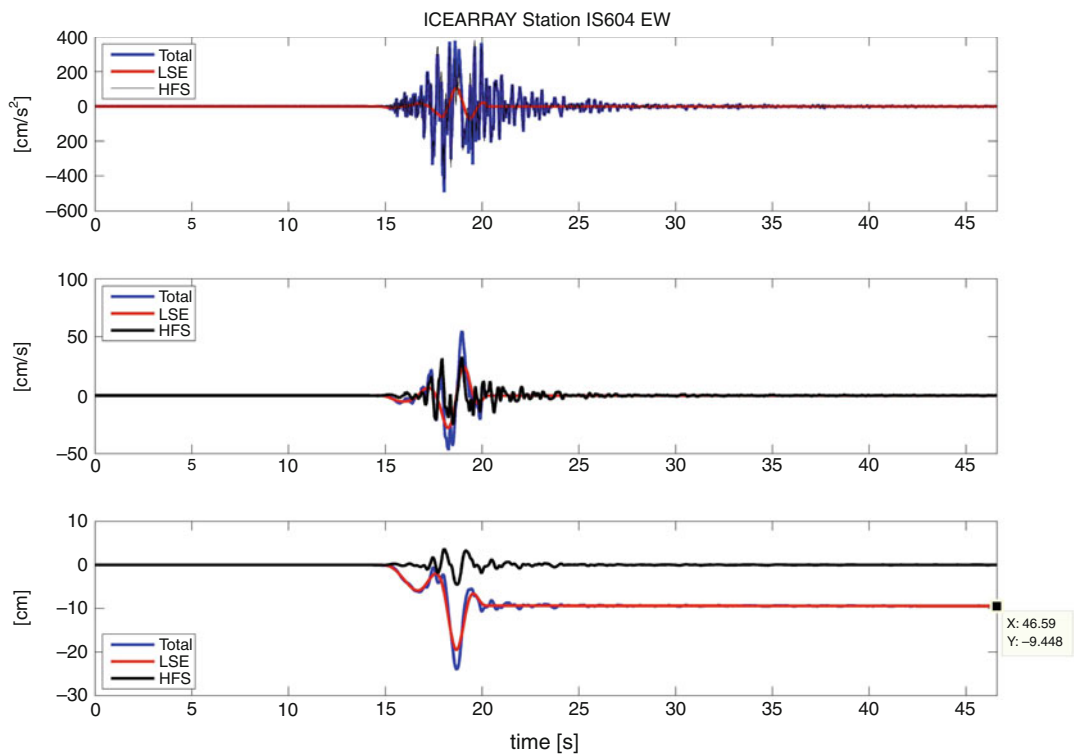
Wavelet Domain Seismic Correction, Fig. 18 GDLC (N55W) results for first and second integration of superimposed “fling” component (red), higher-frequency component (black), and resultant sum of low-frequency fling and higher-frequency component (blue). Permanent displacement shows -177.8 cm

Wavelet Domain Seismic Correction, Table 2 The estimates of acceleration transients and time of occurrence for each of the 11 CUSP stations and for each component

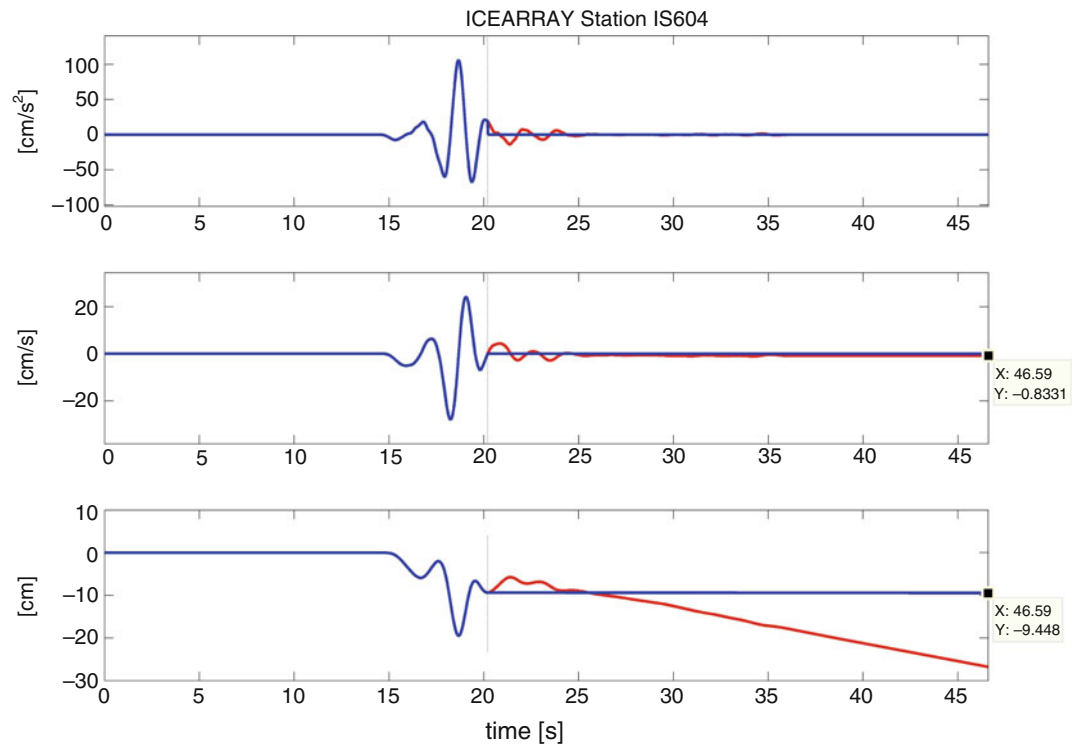
Stations	NS components		EW components		Vertical components	
	Tilt acceleration	Time	Tilt acceleration	Time	Tilt acceleration	Time
	cm/s/s	s	cm/s/s	s	cm/s/s	sec
IS601	-0.8894	19.54	14.71	20.26	0.8263	20.87
IS602	-3.194	18.95	26.14	20.29	-0.66	22.67
IS603	-3.014	19.14	20.6	20.29	-2.22	20.16
IS604	-3.503	19.04	18.6	20.21	-2.064	20.13
IS605	-2.861	19.21	15.52	20.29	-5.559	19.8
IS607	-2.804	19.27	23.24	20.28	-8.99	19.56
IS608	-2.758	19.45	16.42	20.30	-2.317	19.94
IS688	-2.936	19.34	15.72	20.30	-1.761	20
IS609	-2.052	19.00	9.18	20.23	-1.328	20.53
IS610	-1.119	19.15	2.8	22.27	-1.76	20.36
IS611	-2.456	19.38	12.9	20.32	-2.001	20.36
Average	-2.51	19.23	15.99	20.46	-2.54	20.4
Standard deviation	0.84		6.74		2.63	

Wavelet Domain Seismic Correction, Table 3 Estimates of displacement results for the ICEARRAY Icelandic event of the 28 May 2008

Station		NS cm	EW cm	Resultant cm	W(degrees)N
1	IS601	12.88	−9.98	16.3	53
2	IS602	14.93	−17.37	20.71	41
3	IS603	15.38	−13.62	20.55	49
4	IS604	15.58	−9.448	18.23	59
5	IS605	14.48	−11.61	18.56	52
6	IS607	15.36	−10.67	18.71	56
7	IS608	19.48	−9.73	21.78	64
8	IS688	17.36	−11.41	20.78	57
9	IS609	10.93	−12.76	16.81	41
10	IS610	9.173	−8.68	12.63	47
11	IS611	16.57	−14.77	22.2	49
Average		14.73845455	−11.822	19.04	



Wavelet Domain Seismic Correction, Fig. 19 Station IS604: corrected low-frequency sub-band (LFS) fling and higher-frequency sub-bands (HFS) and the resulting total using *bior1.3* for ICEARRAY station 4, EW component obtained at level 7, with a displacement of −9.448 cm, with a time point at 20.21 s. Instrument is the CUSP-3C1p



Wavelet Domain Seismic Correction, Fig. 20 ICEARRAY station IS604 EW, low-frequency sub-band (LFS): showing results before and after baseline correction. The triangular area in the acceleration is the results of $g\varphi$ acceleration tilt transient taken at 20.21 s. The resulting constant velocity DC shift (-0.833 cm/s) is

shown and the linear displacement offset. The velocity time series shows 6 zero-velocity, crossover points, suggesting 6 tilt acceleration transients in the low-frequency, fling space; these are given in Table 8. Instrument is the CUSP-3C1p

Wavelet Domain Seismic Correction, Table 4 Summary of zero-velocity points employed and estimated peak tilt acceleration

Chi-Chi station	NS components		EW components		Vertical components	
	Tilt acceleration	Time	Tilt acceleration	Time	Tilt acceleration	Time
	cm/s/s	s	cm/s/s	s	cm/s/s	sec
TCU129	9.93	28.1	-4.8	30	2.12	33.97
TCU052	3.015	48.45	12.96	40.06	3.15	44.23
TCU068	9.565	45.65	-6.75	45.83	n/a	n/a
TCU102	10.63	38.39	4.04	41.2	2.174	42.93

Wavelet Domain Seismic Correction, Table 5 Comparison of residual displacement (wavelet method) vs. GPS

Level	TCU052NS	TCU052EW	TCU052V
	[cm]	[cm]	[cm]
10	671	-352	369
GPS	845.1	-342.3	397

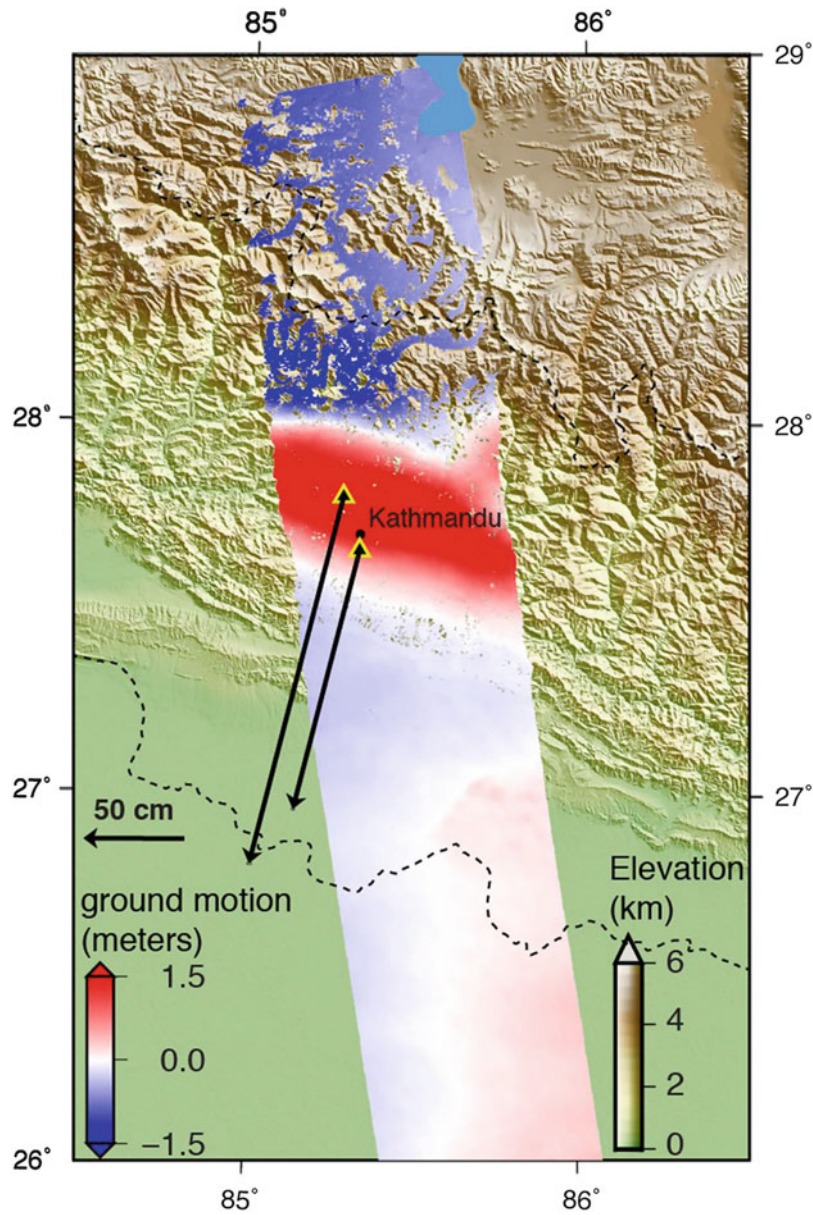
Wavelet Domain Seismic Correction, Table 6 Comparison of residual displacement (wavelet method) vs. GPS

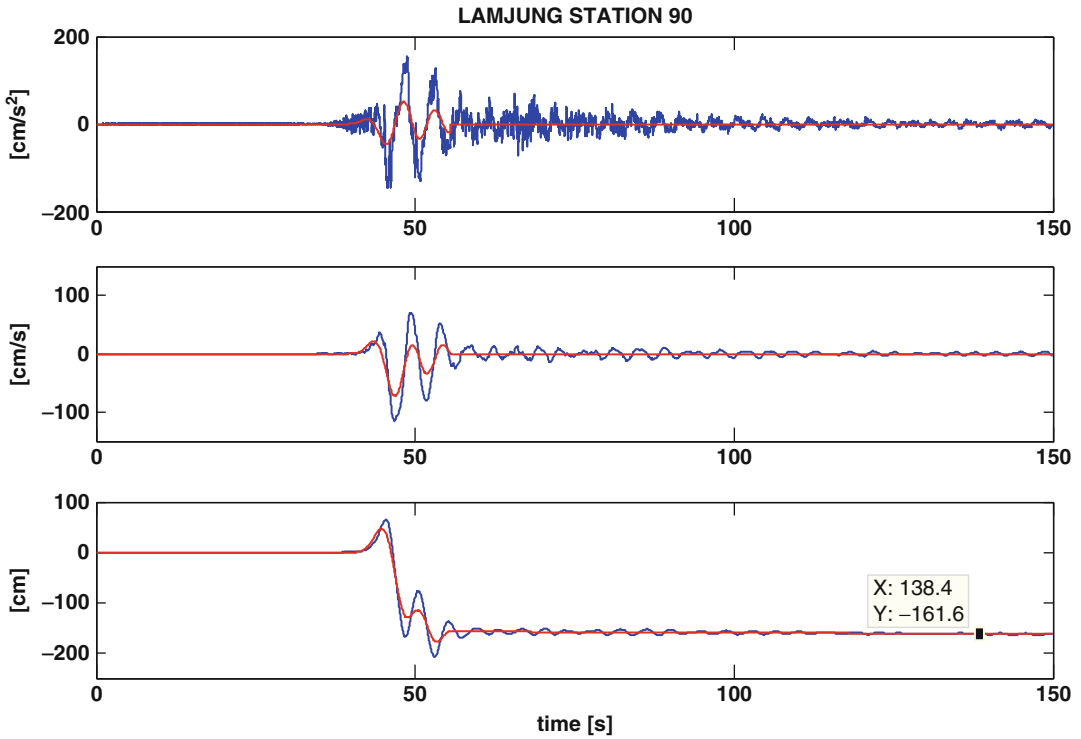
Level	TCU129EW	TCU129NS	TCU129V
	[cm]	[cm]	[cm]
9	78.46	-26.8	-12.26
GPS	88.2	-32.1	-17.7

Wavelet Domain Seismic Correction, Table 7 Displacements for station TCU068, Chi-Chi event 1999

		Baseline point	Residual displacement
	i	T_i	$x_i(t_{\text{end}})$
Wavelet <i>bior1.3</i>		[s]	[cm]
TCU068EW	1	45.7	-731
TCU068NS	2	45.67	555
TCU068V	3	56.84	300

Wavelet Domain Seismic Correction, Fig. 21 Radar shows Kathmandu area uplifted 5 ft by the Gorkha, Nepal, Earthquake (Courtesy of NASA/USGS)





Wavelet Domain Seismic Correction,
Fig. 22 LAMJUNG STATION 90: (a) *Top trace*: acceleration fling pulse (red), max 50.44 cm/s/s, overlaid with the high-frequency signal (b) *Middle trace*: velocity fling pulse, max velocity -72.86 cm/s, overlaid with the high-

frequency signal (c) *Bottom trace*: fling step with a residual displacement of -1.61 m, overlaid with the high-frequency signal. The analyzing wavelet is the *bior1.3* at level 9

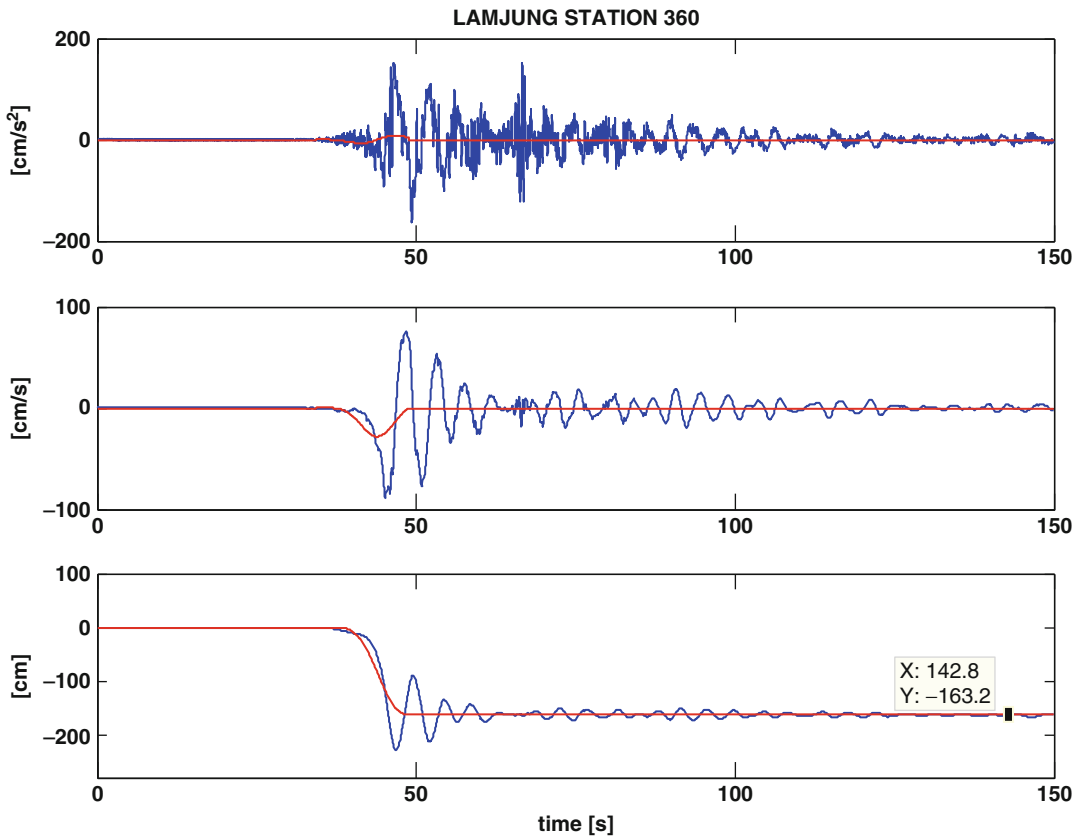
fling step. The horizontal displacement shows a shift south by approximately 1.6 m.

This recent event in Nepal demonstrates the utility of the wavelet transform method as published by Chanerley and Alexander (2010) and Chanerley et al. (2013). The method recovers the residual displacements with a good degree of accuracy, but moreover, the method recovers the fling pulse(s) and the fling step.

Summary

In this entry, some of the state-of-the-art techniques for mitigating noise/error in accelerogram time series have been reviewed. Older analogue accelerographs clearly present a greater challenge due to their design. Nevertheless, this legacy

recording is still an important source for structural and geotechnical earthquake engineering designers/analysts. Modern, digital accelerographs are a great improvement in many ways as they correct design problems of earlier instruments such as higher dynamic range, higher sampling rate, automatic digitization and triggering, flat instrument response down to DC, etc. However, the problem of ground rotation (tilt) noise corruption of low-frequency components of accelerograms is still present, because all strong-motion instruments are 3 (translational)-axis instruments rather than full 6-axis instruments. This may be partially due to the cost of designing and building a 6-axis instrument, but it is also the case that rotation accelerometers (such as gyroscopic transducers) currently may have lower performance specifications than translational ones.



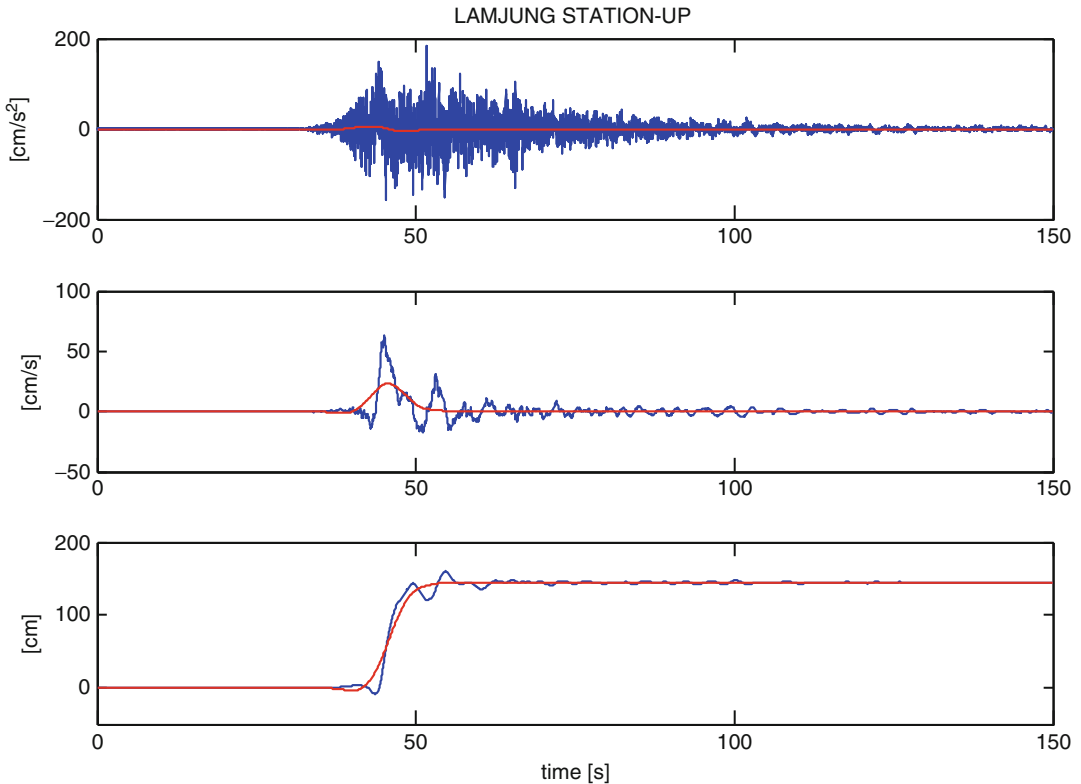
Wavelet Domain Seismic Correction, Fig. 23 LAMJUNG STATION 360°: (a) *Top trace*: acceleration fling pulse (red), max 7.768 cm/s/s, overlaid with the high-frequency signal (b) *Middle trace*: velocity fling pulse, max value 28.15 cm/s, overlaid with the

high-frequency signal (c) *Bottom trace*: displacement fling step, with max displacement -227 cm, residual displacement -163 cm, overlaid with the high-frequency signal. The analyzing wavelet is *bior1.3* at level 10

One solution may be to use multiple 3-axis instruments spaced on a rigid structure such that differences in translational accelerometers (in different instruments) can be used to obtain rotational acceleration estimates. This is often done on bridges for evaluating torsional (rotation) deck modes and could also be applied to ground motion. In this case, the classical 3-axis instrument is not redesigned; however, it would have to be strategically positioned in situ very close to other 3-axis instruments. Another solution may be to have 2 sets of 3-axis sensors in one instrument, with one set fixed and the other on a 6DOF platform with the difference in the time series

giving an estimate of the rotational ground motion.

Nevertheless, it is clear that modern, state-of-the-art signal processing techniques such as the undecimated wavelet transform and de-noising algorithm presented have been applied to a range of seismic events and facilitate meaningful double-time integration and recover the fling pulse, the displacement fling step, and to some extent the rotational (tilt) acceleration transients that are generated by ground rotations. This goes some way toward the recovery of credible ground motion velocity and displacement time series and therefore better-quality boundary conditions in the design of structural artifacts.



Wavelet Domain Seismic Correction, Fig. 24 LAMJUNG STATION-UP: (a) Top trace: acceleration fling pulse (red), max 5.64 cm/s/s, overlaid with the high-frequency signal (b) Middle trace: velocity fling pulse, max value 23.16 cm/s, overlaid with the

high-frequency signal whose max value is 55.36 cm/s (c) Bottom trace: displacement fling step, with residual displacement 144.4 cm, overlaid with the high-frequency signal. The analyzing wavelet is *bior1.3* at level 10

Cross-References

- Downhole Seismometers
- Electrochemical Seismometers of Linear and Angular Motion
- Engineering Characterization of Earthquake Ground Motions
- Historical Seismometer
- Nonlinear System Identification: Particle-Based Methods
- Ocean-Bottom Seismometer
- Physics-Based Ground-Motion Simulation
- Principles of Broadband Seismometry
- Remote Sensing in Seismology: An Overview
- Seismic Accelerometers
- Seismic Network and Data Quality
- Seismic Noise
- Seismometer Arrays

- Seismometer Self-Noise and Measuring Methods
- Seismometer, Extended Response
- Sensors, Calibration of
- Stochastic Ground Motion Simulation
- Time History Seismic Analysis
- Very-Long-Period Seismicity at Active Volcanoes: Source Mechanisms

References

- Avery HR et al (2004) The Canterbury University strong motion recording project. In: Proceedings 13th world conference of earthquake engineering, Canada paper No. 1335, Vancouver
- Berrill J et al (2011) The Canterbury Accelerograph Network (CanNet) and some results from the September 2010, M7.1 Darfield earthquake. In: Proceedings of the

- 9th conference on earthquake engineering, Auckland, 14–16 Apr 2011
- Bilham R (1994) The next great earthquake in the Nepal Himalaya. *Himal Mag* 7(3):26–30
- Bilham R, Bodin P, Jackson M (1995) Entertaining a great earthquake in Western Nepal: Historical activity and geodetic tests for the development of strain. *J Geol Soc Nepal* 11:73–88
- Boore DM (2001) Effect of baseline corrections on displacements and response spectra for several recordings of the 1999 chi-chi, taiwan, earthquake. *Bull Seismol Soc Am* 91:1199–1211
- Boore DM, Bommer JJ (2005) Processing of strong-motion accelerograms: needs, options and consequences. *Soil Dyn Earthq Eng* 25:93–115
- Chanerley A, Alexander NA (2010) Obtaining estimates of the low-frequency “fling”, instrument tilts and displacement timeseries using wavelet decomposition. *Bull Earthq Eng* 8:231–255
- Chanerley AA, Alexander NA, Berrill J, Avery H, Halldorsson B, Sigbjornsson R (2013) Concerning baseline errors in the form of acceleration transients when recovering displacements from strong motion records using the undecimated wavelet transform. *Bull Seismol Soc Am* 103:283–295. doi:10.1785/0120110352
- Chiu H-C (1997) Stable baseline correction of digital strong-motion data. *Bull Seismol Soc Am* 87:932–944
- Coifman R, Donoho DL (1995) Translation invariant de-noising, wavelets and statistics, lecture notes in statistics, vol 103. Springer, New York, pp 125–150
- Daubechies I (1992) Ten lectures on wavelets. Society for Industrial and Applied Mathematics, Philadelphia
- Donoho DL (1995) De-noising by soft-thresholding. *IEEE Trans Inf Theory* 41:613–627
- Graizer VM (2005) Effect of tilt on strong motion data processing. *Soil Dyn Earthq Eng* 25:197–204
- Graizer V (2006) Tilts in strong ground motion. *Bull Seismol Soc Am* 96:2090–2102
- Halldorsson B, Avery H (2009) Converting Strong-motion Networks to Arrays via Common Triggering. *Seismol Res Lett* 80:572–578
- Halldorsson B, Sigbjornsson R (2009) The M_w 6.3 Olfus Earthquake at 15:45 UTC on 29 May 2008 in South Iceland: ICEARRAY Strong-Motion Recordings. *Soil Dyn Earthq Eng* 29:1073–1083
- Halldorsson B, Sigbjornsson R, Schweitzer J (2009) ICEARRAY: the first small aperture, strong-motion array in Iceland. *J Seismol* 13(1):173–178
- Iwan WD, Moser MA, Peng CY (1985) Some observations on strong-motion earthquake measurement using a digital accelerograph. *Bull Seismol Soc Am* 75:1225–1246
- Lang M, Guo H, Odegard JE, Burrus CS, Wells RO Jr (1996) Noise reduction using an undecimated discrete wavelet transform. *IEEE Signal Process Lett* 3:10–12
- Pillet R, Virieux J (2007) The effects of seismic rotations on inertial sensors. *Geophys J Int* 171:1314–1323
- Rupakhety R et al (2010) Estimating coseismic deformations from Near Source Strong Motion Records: Methods and Case Studies. *Bull Earthq Eng* 8:787–811
- Shensa M (1992) The discrete wavelet transform: wedding the a trous and mallat algorithms. *IEEE Trans Signal Process* 40:2464–2482
- Trifunac MD (1971) Zero baseline correction of strong-motion accelerograms. *Bull Seismol Soc Am* 61:1201–1211
- Trifunac MD (1972) A note on correction of strong-motion accelerograms for instrument response. *Bull Seismol Soc Am* 62:401–409
- Wang G-Q, Boore DM, Igel H, Zhou X-Y (2003) Some observations on colocated and closely spaced strong ground-motion records of the 1999 chi-chi, taiwan, earthquake. *Bull Seismol Soc Am* 93:674–693
- Wu Y-M, Wu C-F (2007) Approximate recovery of coseismic deformation from taiwan strong-motion records. *J Seismol* 11:159–170

List of Entries

Artificial and Other Sources and Mechanisms

- Artificial Offshore Seismic Sources
- Seismic Sources from Landslides and Glaciers

Aseismic Design

- Buckling-Restrained Braces and Their Implementation in Structural Design of Steel Buildings
- Cast Fuse Braces: Design and Implementation
- Code-Based Design: Seismic Isolation of Buildings and Nuclear Facilities
- Design of Cable-Supported Bridges: Control Strategies
- Design, Testing, and Evaluation of the Web Plastifying Damper for the Aseismic Protection of Buildings
- Friction Dampers for Seismic Protections of Steel Buildings Subjected to Earthquakes: Emphasis on Structural Design
- Lead-Rubber Bearings with Emphasis on Their Implementation to Structural Design
- Performance-Based Design Procedure for Structures with Magneto-Rheological Dampers
- Rubber Shock Absorbers as a Mitigation Technique for Earthquake-Induced Pounding
- Steel Posttensioned Connections with Web Hourglass Pins: Toward Earthquake Resilient Steel Structures

Case Histories

- Drift Issues of Tall Buildings in 2011 Tohoku Earthquake
- Estimation of Potential Seismic Damage in Urban Areas
- Reconstruction in Indonesia Post-2004 Tsunami: Lessons Learnt

Computational and Sensing

- Early Earthquake Warning (EEW) System: Overview
- Global Navigation Satellite System (GNSS) in Earthquake Engineering, Usage of
- Graphics Processing Unit (GPU) Technology in Earthquake Engineering, Application of
- Integrated Earthquake Simulation
- MEMS Sensors for Measurement of Structure Seismic Response and Their Application
- Nonlinear Analysis and Collapse Simulation Using Serial Computation
- Nonlinear Seismic Ground Response Analysis of Local Site Effects with Three-Dimensional High-Fidelity Model

Computational Rock Mechanics

- Earthquake Early Warning System in Taiwan
- Earthquake Recurrence Law and the Weibull Distribution
- Probability Seismic Hazard Mapping of Taiwan

- Review and Implications of Inputs for Seismic Hazard Analysis
- Seismic Design of Earth-Retaining Structures

- Sustained Earthquake Preparedness: Functional, Social, and Cultural Issues
- Waste Management Following Earthquake Disaster

Construction Techniques

- Masonry Box Behavior
- Masonry Components
- Masonry Macro-block Analysis
- Masonry Modeling
- Masonry Structures: Overview
- Reinforced Concrete Structures in Earthquake-Resistant Construction
- Steel Structures
- Timber Structures

Disaster Recovery and Reconstruction and Loss Modelling

- “Build Back Better” Principles for Reconstruction
- Community Recovery Following Earthquake Disasters
- Damage to Buildings: Modeling
- Damage to Infrastructure: Modeling
- Earthquake Disaster Recovery: Leadership and Governance
- Economic Impact of Seismic Events: Modeling
- Economic Recovery Following Earthquake Disasters
- Emergency Response Coordination Within Earthquake Disasters
- Interim Housing Provision Following Earthquake Disaster
- Land Use Planning Following an Earthquake Disaster
- Learning from Earthquake Disasters
- Legislation Changes Following Earthquake Disasters
- Probabilistic Seismic Hazard Models
- Reconstruction Following Earthquake Disasters
- Resilience to Earthquake Disasters
- Resourcing Issues Following Earthquake Disaster

Geotechnical Engineering

- Analysis and Design Issues of Geotechnical Systems: Flexible Walls
- Analysis and Design Issues of Geotechnical Systems: Rigid Walls
- Bridge Foundations
- Design of Mechanically Stabilized Walls (MSWs)
- Dynamic Soil Properties: In Situ Characterization Using Penetration Tests
- Geotechnical Earthquake Engineering: Damage Mechanism Observed
- Liquefaction: Countermeasures to Mitigate Risk
- Liquefaction: Performance of Building Foundation Systems
- Residual Strength of Liquefied Soils
- Safety Assessment of Piled Buildings in Liquefiable Soils: Mathematical Tools
- Seismic Design of Dams
- Seismic Design of Pipelines
- Seismic Design of Tunnels
- Seismic Design of Waste Containment Systems
- Site Response: 1-D Time Domain Analyses
- Site Response: Comparison Between Theory and Observation

Mathematical Tools

- Bayesian Statistics: Applications to Earthquake Engineering
- Model-Form Uncertainty Quantification for Structural Design
- Reliability Analysis of Nonlinear Vibrating Systems-Spectral Approach
- Reliability Estimation and Analysis for Dynamical Systems
- Seismic Fragility of Aging Offshore Platforms
- Seismic Response Prediction of Degrading Structures

- Spectral Finite Element Approach for Structural Dynamics
- Structural Reliability Estimation for Seismic Loading
- Structural Seismic Reliability Analysis
- Structures with Nonviscous Damping, Modeling, and Analysis
- Substructuring Methods for Finite Element Analysis
- Uncertainty Quantification in Structural Health Monitoring
- Wavelet Domain Seismic Correction

Paleoseismology

- Archeoseismology
- Damage to Ancient Buildings from Earthquakes
- EEE Catalogue: A Global Database of Earthquake Environmental Effects
- Intensity Scale ESI 2007 for Assessing Earthquake Intensities
- Luminescence Dating in Paleoseismology
- Paleoseismology of Glaciated Terrain
- Paleoseismic Trenching
- Paleoseismology
- Paleoseismology and Landslides
- Paleoseismology of Rocky Coasts
- Paleoseismology: Integration with Seismic Hazard
- Radiocarbon Dating in Paleoseismology
- Tsunamis as Paleoseismic Indicators

Random Vibration

- Complex Fractional Moments and Their Use in Earthquake Engineering
- Non-Poisson Impulse Processes
- Probability Density Evolution Method in Stochastic Dynamics
- Random Process as Earthquake Motions
- Response-Spectrum-Compatible Ground Motion Processes
- Seismic Analysis of Wind Energy Converters
- Stochastic Analysis of Linear Systems
- Stochastic Analysis of Nonlinear Systems

- Stochastic Finite Elements
- Structural Optimization Under Random Dynamic Seismic Excitation
- Tuned Mass and Parametric Pendulum Dampers Under Seismic Vibrations
- Vibrations of Beams for Seismic Response Estimation

Reliability and Robustness

- Base-Isolated Systems, Reliability-Based Characterization of
- Classically and Nonclassically Damped Multi-degree of Freedom (MDOF) Structural Systems, Dynamic Response Characterization of
- Reliability Estimation and Analysis
- Response Variability and Reliability of Structures
- Robust Control of Building Structures Under Uncertain Conditions
- Robust Design Optimization for Earthquake Loads
- Seismic Robustness Analysis of Nuclear Power Plants
- Sensitivity of First-Excursion Probabilities for Nonlinear Stochastic Dynamical Systems
- Subset Simulation Method for Rare Event Estimation: An Introduction
- Tuned Mass Dampers for Passive Control of Structures Under Earthquake Excitations
- Uncertainty Theories: Overview

Remote Sensing

- Building Damage from Multi-resolution, Object-Based Classification Techniques
- Building Monitoring Using a Ground-Based Radar
- Damage Detection in Built-Up Areas Using SAR Images
- Earthquake Damage Assessment from VHR Data: Case Studies
- Features Extraction from Satellite Data
- Hyperspectral Data in Urban Areas

- InSAR and A-InSAR: Theory
- Remote Sensing in Seismology: An Overview
- SAR Images, Interpretation of
- SAR Tomography for 3D Reconstruction and Monitoring
- Urban Change Monitoring: Multi-temporal SAR Images

Retrofitting and Strengthening

- Assessment and Strengthening of Partitions in Buildings
- Assessment of Existing Structures Using Inelastic Static Analysis
- Assessment of Existing Structures Using Response History Analysis
- Passive Control Techniques for Retrofitting of Existing Structures
- Retrofitting and Strengthening Masonries of Heritage Structures: Materials Used
- Retrofitting and Strengthening Measures: Liability and Quality Assurance
- Retrofitting and Strengthening of Contemporary Structures: Materials Used
- Retrofitting and Strengthening of Structures: Basic Principles of Structural Interventions
- Seismic Strengthening Strategies for Existing (Code-Deficient) Ordinary Structures
- Seismic Strengthening Strategies for Heritage Structures
- Strengthened Structural Members and Structures: Analytical Assessment
- Strengthening Techniques: Bridges
- Strengthening Techniques: Code-Deficient R/C Buildings
- Strengthening Techniques: Code-Deficient Steel Buildings
- Strengthening Techniques: Masonry and Heritage Structures

Risk Management: Decision Analysis

- Seismic Fragility Analysis

Risk Mitigation Policies and Approaches

- Building Codes and Standards
- Building Disaster Resiliency Through Disaster Risk Management Master Planning
- Earthquake Mitigation Strategies Through Insurance
- Earthquake Protection of Essential Facilities
- Earthquake Risk Mitigation of Lifelines and Critical Facilities
- Global View of Seismic Code and Building Practice Factors
- Resiliency of Water, Wastewater, and Inundation Protection Systems
- School Seismic Safety and Risk Mitigation
- School Seismic Safety: Case Studies
- Social Media Benefits and Risks in Earthquake Events

Seismic Risk Assessment

- Analytic Fragility and Limit States [P(EDP|IM)]: Nonlinear Dynamic Procedures
- Analytic Fragility and Limit States [P(EDP|IM)]: Nonlinear Static Procedures
- Beginner's Guide to Fragility, Vulnerability, and Risk
- Conditional Spectra
- Earthquake Recurrence
- Earthquakes and Their Socio-economic Consequences
- Empirical Fragility
- Engineering Characterization of Earthquake Ground Motions
- Insurance and Reinsurance Models for Earthquake
- Physics-Based Ground-Motion Simulation
- Seismic Collapse Assessment
- Seismic Loss Assessment
- Seismic Reliability Assessment, Alternative Methods for
- Seismic Resilience
- Seismic Risk Assessment, Cascading Effects

- Seismic Vulnerability Assessment: Lifelines
- Seismic Vulnerability Assessment: Masonry Structures
- Seismic Vulnerability Assessment: Reinforced Concrete Structures
- Site Response for Seismic Hazard Assessment
- Source Characterization for Earthquake Early Warning
- Stochastic Ground Motion Simulation

Sensors and Sensor Systems

- Downhole Seismometers
- Earthquake Location
- Earthquake Magnitude Estimation
- Electrochemical Seismometers of Linear and Angular Motion
- Historical Seismometer
- Ocean-Bottom Seismometer
- Passive Seismometers
- Principles of Broadband Seismometry
- Recording Seismic Signals
- Seismic Accelerometers
- Seismic Event Detection
- Seismic Instrument Response, Correction for
- Seismic Network and Data Quality
- Seismic Noise
- Seismometer Arrays
- Seismometer Self-Noise and Measuring Methods
- Seismometer, Extended Response
- Sensors, Calibration of
- Symmetric Triaxial Seismometers
- Buildings and Bridges Equipped with Passive Dampers Under Seismic Actions: Modeling and Analysis
- Earthquake Response Spectra and Design Spectra
- Earthquake Return Period and Its Incorporation into Seismic Actions
- Equivalent Static Analysis of Structures Subjected to Seismic Actions
- European Structural Design Codes: Seismic Actions
- Incremental Dynamic Analysis
- Liquid Storage Tanks: Seismic Analysis
- Mixed In-Height Concrete-Steel Buildings Under Seismic Actions: Modeling and Analysis
- Modal Analysis
- Nonlinear Dynamic Seismic Analysis
- Nonlinear Finite Element Analysis
- Numerical Modeling of Masonry Infilled Reinforced Concrete Frame Buildings
- Plastic Hinge and Plastic Zone Seismic Analysis of Frames
- Response Spectrum Analysis of Structures Subjected to Seismic Actions
- Secondary Structures and Attachments Under Seismic Actions: Modeling and Analysis
- Seismic Actions Due to Near-Fault Ground Motion
- Seismic Analysis of Concrete Bridges: Numerical Modeling
- Seismic Analysis of Masonry Buildings: Numerical Modeling
- Seismic Analysis of Steel and Composite Bridges: Numerical Modeling
- Seismic Analysis of Steel Buildings: Numerical Modeling
- Seismic Analysis of Steel-Concrete Composite Buildings: Numerical Modeling
- Selection of Ground Motions for Response History Analysis
- Soil-Structure Interaction
- Spatial Variability of Ground Motion: Seismic Analysis
- Structural Design Codes of Australia and New Zealand: Seismic Actions
- Time History Seismic Analysis

Structural Analysis for Earthquake-Resistant Design

- Actively and Semi-actively Controlled Structures Under Seismic Actions: Modeling and Analysis
- Ancient Monuments Under Seismic Actions: Modeling and Analysis
- Behavior Factor and Ductility

Structural Health Monitoring

- Advances in Online Structural Identification
- Ambient Vibration Testing of Cultural Heritage Structures
- Bayesian Operational Modal Analysis
- Blind Identification of Output-Only Systems and Structural Damage via Sparse Representations
- Laser-Based Structural Health Monitoring
- Model Class Selection for Prediction Error Estimation
- Nonlinear System Identification: Particle-Based Methods
- Novel Bio-Inspired Sensor Network for Condition Assessment
- Online Response Estimation in Structural Dynamics
- Operational Modal Analysis in Civil Engineering: An Overview
- Parametric Nonstationary Random Vibration Modeling with SHM Applications
- Post-Earthquake Diagnosis of Partially Instrumented Building Structures
- Seismic Behavior of Ancient Monuments: From Collapse Observation to Permanent Monitoring
- Spatial Filtering for Structural Health Monitoring
- Stochastic Structural Identification from Vibrational and Environmental Data
- System and Damage Identification of Civil Structures
- Vibration-Based Damage Identification: The Z24 Bridge Benchmark

Tectonic Sources and Mechanisms

- Earthquake Mechanism and Seafloor Deformation for Tsunami Generation

- Earthquake Mechanism Description and Inversion
- Earthquake Mechanisms and Stress Field
- Earthquake Mechanisms and Tectonics
- Earthquake Swarms
- Long-Period Moment-Tensor Inversion: The Global CMT Project
- Mechanisms of Earthquakes in Aegean
- Mechanisms of Earthquakes in Iceland
- Mechanisms of Earthquakes in Vrancea
- Moment Tensors: Decomposition and Visualization
- Non-Double-Couple Earthquakes
- Regional Moment Tensor Review: An Example from the European–Mediterranean Region
- Reliable Moment Tensor Inversion for Regional- to Local-Distance Earthquakes

Volcanic Seismology

- Frequency-Magnitude Distribution of Seismicity in Volcanic Regions
- Noise-Based Seismic Imaging and Monitoring of Volcanoes
- Rockfall Seismicity Accompanying Dome-Building Eruptions
- Seismic Anisotropy in Volcanic Regions
- Seismic Monitoring of Volcanoes
- Seismic Tomography of Volcanoes
- Tracking Changes in Volcanic Systems with Seismic Interferometry
- Very-Long-Period Seismicity at Active Volcanoes: Source Mechanisms
- Volcanic Eruptions, Real-Time Forecasting of
- Volcano-Tectonic Seismicity of Soufriere Hills Volcano, Montserrat